Full-title

Lack of parent-of-origin effects in Nasonia jewel wasp: a replication and extension study

Short-title

Nasonia parent-of-origin effects

Authors

Kimberly C. Olney[1,2], Joshua D. Gibson[3], Heini M. Natri[2], Avery Underwood[1,2], Juergen Gadau[4], Melissa A. Wilson[1,2,5*]

Affiliations

1. School of Life Sciences, Arizona State University, Tempe AZ 85281 USA
2. Center for Evolution and Medicine, Arizona State University, Tempe AZ 85281 USA
3. Department of Biology, Georgia Southern University, Statesboro GA 30460 USA
4. Institut fuer Evolution and Biodiversity, University of Muenster, Muenster, Germany 48149
5. Center for Mechanisms of Evolution, The Biodesign Institute, Arizona State University, Tempe AZ 85281 USA

***Corresponding author**

Melissa A. Wilson

School of Life Sciences | Arizona State University | PO Box 874501 | Tempe, AZ 85287-4501

mwilsons@asu.edu

Author contributions

KCO: Supervision, Formal Analysis, Investigation, Visualization, Writing - Original Draft Preparation, Writing - Review and Editing

JDG: Conceptualization, Supervision, Formal Analysis, Investigation, Resources, Experimental Procedures, Writing - Original Draft Preparation, Writing - Review and Editing, Visualization

HMN: Supervision, Formal Analysis, Investigation, Visualization, Writing - Original Draft Preparation, Writing - Review and Editing

AU: Formal Analysis, Investigation, Writing - Review and Editing

JG: Conceptualization, Funding acquisition, Supervision, Resources, Writing - Review and Editing

MAW: Conceptualization, Supervision, Visualization, Resources, Project Administration, Writing - Original Draft Preparation, Writing - Review and Editing, Funding Acquisition

Abbreviations

Allele-specific expression (ASE)

Differential expression (DE)

Differentially expressed genes (DEGs)

False discovery rate (FDR)

$\log_2$ fold-change ($\log_2$FC)

log counts per million (logCPM)

*Nasonia girault*i maternal and paternal (GG)

Nasonia vitripennis maternal and paternal (VV)

*N. girault*i maternal, *N. vitripennis paternal* (GV) F_1GV

N. vitripennis maternal, *N. girault*i *paternal* (VG) F_1VG

No difference (ND)

RNA sequence (RNAseq)

Standard deviation (SD)

Variant call format (VCF)

Abstract

In diploid cells, the paternal and maternal alleles are, on average, equally expressed. There are exceptions from this: a small number of genes express the maternal or paternal allele copy exclusively. This phenomenon, known as genomic imprinting, is common among eutherian mammals and some plant species; however, genomic imprinting in species with haplodiploid sex determination is not well characterized. Previous work reported no parent-of-origin effects in the hybrids of closely related haplodiploid *Nasonia vitripennis* and *Nasonia giraulti* jewel wasps, suggesting a lack of epigenetic reprogramming during embryogenesis in these species. Here, we replicate the gene expression dataset and observations using different individuals and sequencing technology, as well as reproduce these findings using the previously published RNA sequence data following our data analysis strategy. The major difference from the previous dataset is that they used an introgression strain as one of the parents and we found several loci that resisted introgression in that strain. Our results from both datasets demonstrate a species-of-origin effect, rather than a parent-of-origin effect. We present a reproducible workflow that others may use for replicating the results. Overall, we reproduced the original report of no parent-of-origin effects in the haplodiploid *Nasonia* using the original data with our new processing and analysis pipeline and replicated these results with our newly generated data.

Introduction

Parent-of-origin effects occur when there is a biased expression (or completely monoallelic expression) of alleles inherited from the two parents [1,2]. Monoallelic gene expression in the offspring is hypothesized to be primarily the result of genetic conflict between parents over resource allocation in the offspring [3,4]. In mammals, the mechanism of these parent-of-origin effects occurs via inherited methylation of one allele [1,5]. In insects, the relationship between methylation of genomic DNA and the expression of the gene that it encodes is not as well

characterized but studies of social insects showed that there is a positive correlation of DNA methylation of gene bodies and gene expression [6].

Honey bees have been a focal group for investigation of parent-of-origin effects in insects due to differences in the kinship between queens, males, and workers [7,8]. Multiple mating by queens results in low paternal relatedness between workers and should lead to intragenomic conflict over worker reproduction (laying unfertilized eggs to produce males), and ultimately should favor the biased expression of paternal alleles that promote worker reproduction [9]. Utilizing a cross between European (*Apis mellifera ligustica*) and Africanized honey bees, Galbraith et al. 2016 identified genes exhibiting a pattern of biased paternal allele overexpression in worker reproductive tissue from colonies that were queenless and broodless, a colony condition that promotes worker reproduction [9]. Smith et al. 2020 found a similar pattern of paternal allele overexpression in diploid (worker-destined) eggs in a cross between two African subspecies, *A.m. scutellate* and *A.m. capensis* [10]. In reciprocal crosses of European (*A.m. ligustica* and *A.m. carnica*) and Africanized honey bees reared in colonies containing both brood and a queen, Kocher et al. 2015 instead found parent-of-origin effects in gene expression that were largely overexpressing the maternal allele in both directions of the cross [11]. These studies provide evidence for parent-of-origin effects in the honey bee, a eusocial Hymenoptera. The Kocher et al. 2016 dataset also exhibited asymmetric maternal allelic bias in which the paternal allele was silenced, but only in hybrids with Africanized fathers [12]. This set of biased genes was enriched for mitochondrial-localizing proteins and is overrepresented in loci associated with aggressive behavior in previous studies [13,14]. Interestingly, these same crosses exhibit high aggression in the direction of the cross with the Africanized father but not in the reciprocal cross [15], and aggression and brain oxidative metabolic rate appears to be linked in honey bees [16]. This study points toward a potential role of allelic bias and nuclear-mitochondrial genetic interactions in wide crosses of honey bees.

The parasitoid wasp genus *Nasonia* has emerged as an excellent model for studying genomic imprinting in Hymenoptera. Like honey bees and all Hymenoptera, *Nasonia* has a haplodiploid sex-determination system in which females are diploid, developing from fertilized eggs, and males are haploid, developing from unfertilized eggs. However, it serves as a strong contrast to studying parent-of-origin effects in the eusocial Hymenoptera as *Nasonia* is solitary and singly-mated, which should result in less genomic conflict and therefore less selective pressure for genomic imprinting based on kinship. By studying allelic expression biases in this system, we can better assess genomic imprinting in the absence of kin selection and the potential contribution of nuclear-mitochondrial interactions to biased allelic expression. *Nasonia* is well-suited for these kinds of studies as two closely related species of *Nasonia* - *N. vitripennis* and *N. giraulti* - that diverged ~1 million years ago (Mya) and show a synonymous coding divergence of ~3% [17], can still produce viable and fertile offspring [18]. Highly inbred laboratory populations of *N. vitripennis* and *N. giraulti* with reduced polymorphism provide an ideal system for identifying parent-of-origin effects in hybrid offspring [19]. However, the species do show genetic variation and incompatibilities, such that recombinant F2 males (from unfertilized eggs of F1 hybrid females) suffer asymmetric hybrid breakdown in which 50% to 80% of the offspring die during development [18]. The mortality is dependent on the direction of the cross and those with *N. giraulti* maternity (cytoplasm) have the highest level of mortality. Nuclear-mitochondrial incompatibilities have been implicated in this and candidate loci have been identified [20–22]. Despite this high level of mortality in F2 males, there is no obvious difference in mortality of the F1 mothers of these males and non-hybrid females, further highlighting this as an excellent system in which to test the potential role of allelic expression bias in mitigating hybrid dysfunction.

Wang et al. 2016 used genome-wide DNA methylation and transcriptome-wide gene expression data from 11 individuals to test whether differences in DNA methylation drive the differences in

gene expression between *N. vitripennis* and *N. giraulti,* and whether there are any parent-of-origin effects (parental imprinting and allele-specific expression) [19]. They used reciprocal crosses of these two species and found no parent-of-origin effects, suggesting a lack of genomic imprinting. Unlike the work in honey bees, however, there have not been multiple independent investigations of evidence for parent-of-origin effects in *Nasonia*.

Reproducibility is a major concern in science, particularly for the biological and medical sciences [23,24]. To replicate is to make an exact copy. To reproduce is to make something similar to something else. Reports have shown that significant factors contributing to irreproducible research include selective reporting, unavailable code and methods, low statistical power, poor experimental design, and raw data not available from the original lab [23,25,26]. In RNAseq experiments, raw counts are transformed into gene or isoform counts, which requires an *in silico* bioinformatics pipeline [27]. These pipelines are modular and parameterized according to the experimental setup [27]. The choice of software, parameters used, and biological references can alter the results. In RNAseq, filters can also improve the robustness of differential expression calls and consistency across sites and platforms [28]. There is no, and there may never be, a defined optimal RNAseq processing pipeline from raw sequencing files to meaningful gene or isoform counts. Thus, the same data can be processed in a multitude of ways by the choice of software, parameters, and references used [27]. Given the exact same inputs, software, and parameters, one can reproduce the analysis if the authors provide this documentation and make explicit the information related to the data transformation used to the RNAseq data [27]. In the case of the Wang et al. 2016, the methods and experimental design were exceptionally well documented, and the authors made available their raw data [19].

To address whether the Wang et al. 2016 findings of lack of parent-of-origin effects in *Nasonia* may be replicated and reproduced, we conducted two sets of analyses. We first downloaded the

153 raw data from 11 individuals [19] and replicated differential expression (DE) and allele-specific
154 expression (ASE) analyses. This allowed us to characterize species differences in gene
155 expression, hybrid effects relative to each maternal and paternal line, and possible parent-of-
156 origin effects using new alignment methods and software. Second, we reproduced the
157 experimental setup with new individuals, generated transcriptome-wide expression levels of 12
158 *Nasonia* individuals (parental strains and reciprocal hybrids), named here as the Wilson data
159 using similar, but not identical strains as the Wang et al. 2016 samples, which we named as the
160 R16A Clark data. The Wilson data, reported here, used the standard *N. giraulti* strain (RV2Xu).
161 The R16A Clark *N. giraulti* differs from the RV2Xu strain in that it has a nuclear *N. giraulti* genome
162 introgressed into a *N. vitripennis* cytoplasm which harbour *N. vitripennis* mitochondria. Both
163 studies used the same highly inbred standard *N. vitripennis* strain, ASymCx. We completed the
164 above analyses to test for robust reproducibility in biased allele and parent-of-origin effects in
165 *Nasonia*. In this analysis, we processed both the R16A Clark and Wilson data using the same
166 software and thresholds, starting with the raw FASTQ files. While we detect some differences in
167 the specific differentially expressed genes between the two datasets, our study reproduces and
168 confirms the main conclusions of the Wang et al. 2016 study: we observe similar trends in the DE
169 and ASE genes, and we detect no parent-of-origin effects in *Nasonia* hybrids, indicating a
170 validation of the lack of epigenetic reprogramming during embryogenesis in this taxa [19]. We
171 make available the bioinformatics processing and analysis pipeline used for both the R16A Clark
172 and Wilson datasets for easily replicating the results reported here:
173 https://github.com/SexChrLab/Nasonia. Finally, during the process of reproducing these results,
174 we extend them to show potential interactions between the mtDNA and autosomal genome that
175 were not apparent in the original study.

Results

Samples cluster by species and hybrid in R16A Clark and Wilson datasets

We used Principal Component Analysis (PCA) of gene expression data to explore the overall structure of the two datasets, R16A Clark and Wilson. Although the reciprocal hybrids from the two datasets are slightly different **Fig 1B,** in both sets, samples from the two species (strains) form separate clusters, with the clustering of the hybrid samples between them **Fig 2A**. The first PC explains most of the gene expression variation in both datasets, with proportions of variance explained 58.17% in R16A Clark and 61.69% in the Wilson data. Further, despite differences in experimental protocols, the transcriptome-wide gene expression measurements across the different crosses and species are highly correlated between the R16A Clark and Wilson dataset, **Fig 3**. There is a difference in the mean RNAseq library size between the two datasets. The mean RNAseq library size for the R16A Clark samples is 48,893,872 base pairs (bp) (SD=11,603,536) and the Wilson samples is 16,518,955 bp (SD=3,205,303), **S1 Table**. Overall, we observe that most of the variation in the data is explained by species and hybrids.

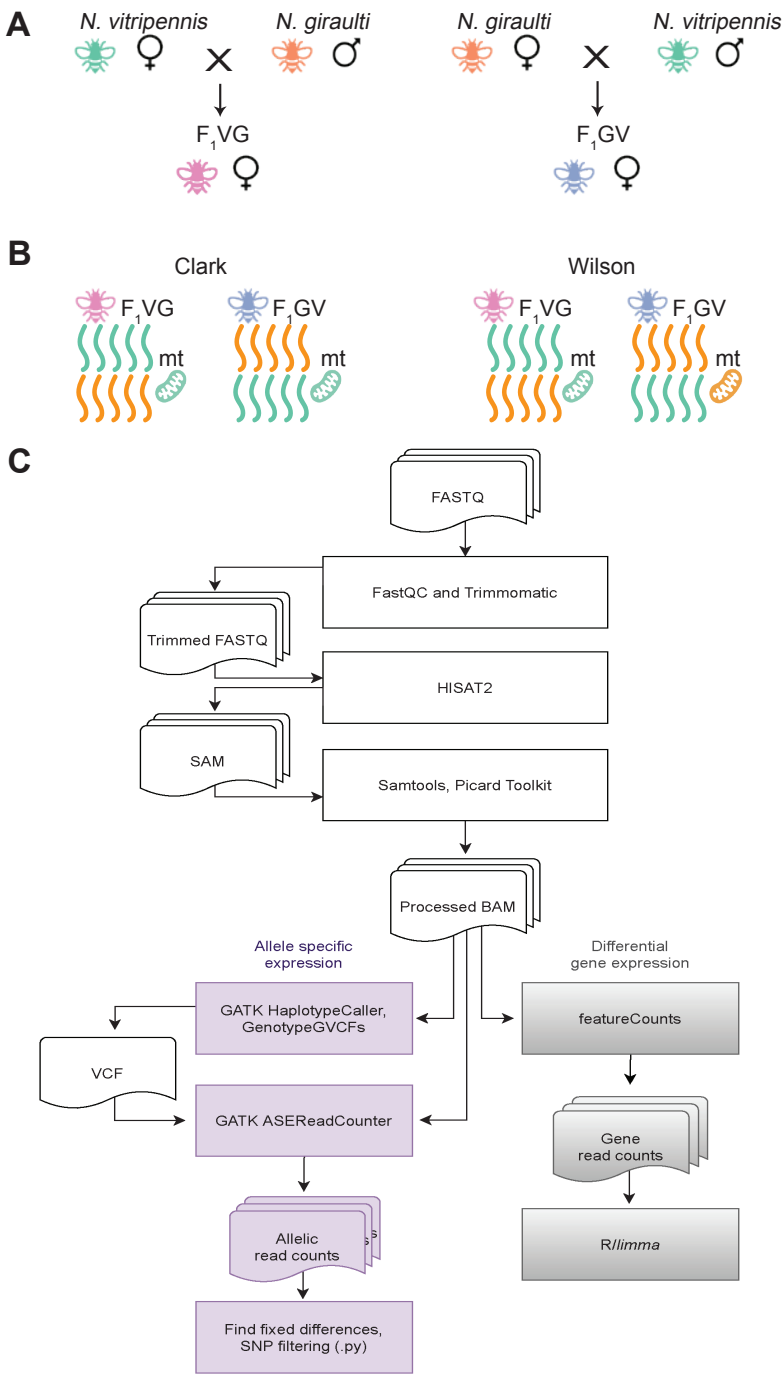
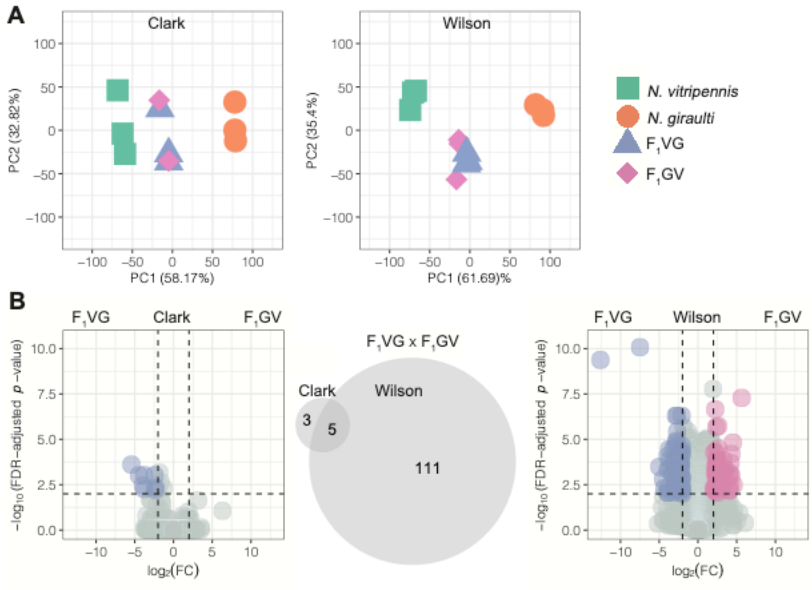


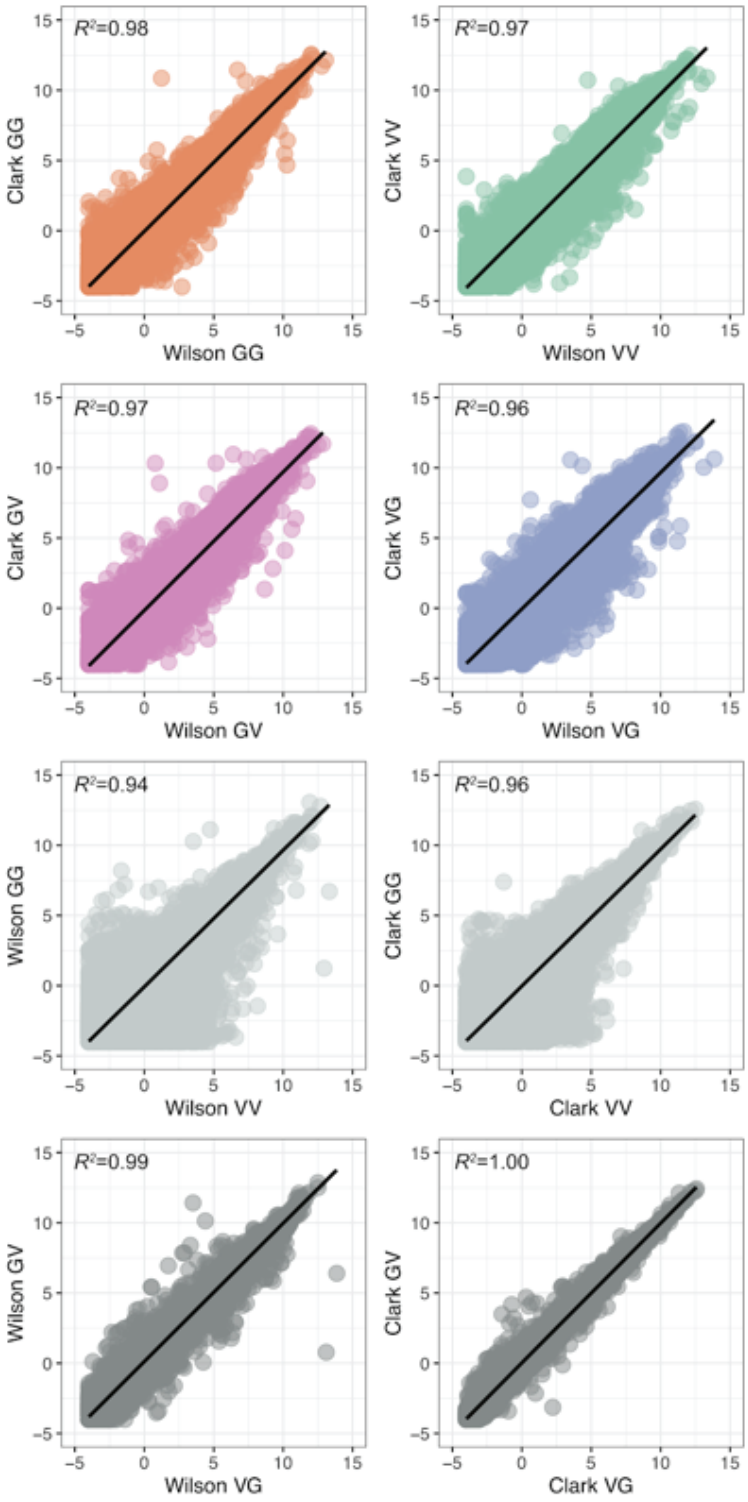
Fig 1. **A:** A schematic illustration of the reciprocal F_1 crosses. **B:** Schematic illustration of the hybrids nuclear and mitochondrial genomic make up. All hybrids are heterozygous at every nuclear locus for their two parent's alleles. The R16A Clark hybrids have *N. vitripennis* mitochondria, regardless of maternal species. The Wilson hybrids have their maternal species mitochondria. **C:** Overview of the data processing and analysis workflow.

196



197

198 **Fig 2. A:** Gene expression PCA based on all expressed genes (mean FPKM ≥ 0.5 across three
199 biological replicates in at least one sample group) in the R16A Clark and Wilson datasets when
200 taking the average between the *N. vitripennis* and pseudo *N. girault* reference genomes. **B:**
201 Volcano plots of differentially expressed genes between the two reciprocal hybrids in the R16A
202 Clark and Wilson datasets. Significance thresholds of an FDR-adjusted *p*-value ≤ 0.01 and an
203 absolute $\log_2$FC ≥ 2 are indicated. A Venn diagram shows the overlap of the significant DEGs.



204

205 **Fig 3.** Gene expression correlation between the Wilson and R16A Clark datasets, as well as

206 between species and between reciprocal hybrids within each dataset. Mean logCPM expression

207 of each quantified gene in each cross and dataset is shown. Pearson's correlation R^2 is indicated.

Species and hybrid differences in gene expression between closely related *N. vitripennis* and *N. girault*i

We detect more differentially expressed genes (DEGs) in the Wilson dataset despite the smaller library sizes, particularly in the comparison involving the hybrid samples (**Fig 2B**). We called DEGs, FDR ≤ 0.01, and absolute $\log_2$ fold change ≥ 2, between the different species and crosses within both datasets (**Fig 2B and S1 Fig**). In the *N. vitripennis* (VV) x *N. girault*i (GG) comparison, we identify 799 and 1,001 DEGs in the R16A Clark and Wilson datasets, respectively. We observe a 45.5% overlap of these DEGs between the datasets (**S1 Fig)**. As expected, we detect fewer DEGs in the comparisons involving the hybrids (**Fig 1B**). We detect only small differences in the numbers of DEGs called in the R16A Clark and Wilson datasets when examining hybrid effects relative to each maternal line (**S1 Fig)**. However, these DEGs show little overlap between the datasets, with the proportions of overlapping DEGs in VVxVG, VVxGV, GGxVG, and GGxGV, comparisons being 24.1%, 16.2%, 39%, and 31.6%, respectively.

There is a notable difference in the number of DEGs called between VG and GV hybrids between the R16A Clark and Wilson datasets. The R16A Clark data used an introgression strain of *N. girault*i, R16A, that has a nuclear genome derived from *N. girault*i but maintains *N. vitripennis* mitochondria, therefore the R16A Clark hybrids all have the same genetic makeup whereas the Wilson reciprocal hybrids have the same nuclear genome but different cytoplasms; yet, we do see eight genes called as differentially expressed between the VG and GV hybrids in the R16A Clark data. Three of the eight genes in the R16A Clark data (LOC116416025, LOC116416106, LOC116417553) were only called as differentially expressed between the VG and GV hybrids in the R16A Clark dataset and weren't called as differentially expressed in the Wilson dataset. The other five genes (LOC107981401, LOC100114950, LOC116415892, LOC103317241, LOC107981942) were called as differentially expressed between the VG and GV in both datasets.

In the Wilson data, we called 116 DEGs, 111 of which are uniquely to the Wilson data set. The original Wang et al. 2016 publication did not investigate differential expression between the hybrids [19]. Here we report a new way of looking at the data, and despite the same genetic makeup between the hybrids in the R16A Clark data, we do observe differential expression between the hybrids, and five of those eight genes are also called as differentially expressed in the Wilson data.

Four (LOC107981401, LOC100114950, LOC116415892, and LOC103317241) out of the five DEGs shared between the data sets are uncharacterized proteins located on Chr 1, Chr 2, and Chr 4. To gain insight into the possible functions of these genes, we used NCBIs BLASTp excluding *Nasonia [29,30]* to find regions of similarity between these sequences and characterized sequences. We observe several significant hits to different insects including *Drosophila* suggesting that these proteins have at least some conservation in insects over > 300 million years. The fifth shared DEG, LOC107981942, located on chromosome 1, is annotated as a zinc finger BED domain-containing protein 1. An NCBI Conserved Domain Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) using these protein sequences uncovered no significant hits with LOC100114950, LOC116415892, and LOC103317241. However, LOC107981401 and LOC107981942 show significant hits for transposase domain superfamilies cl24015 and cl04853, respectively. The role of these proteins in *Nasonia* remains unclear.

Lack of Parent-of-Origin effects in *Nasonia* hybrids

We used allele-specific expression (ASE) analyses to detect parent-of-origin effects — indicated by allelic bias — in *Nasonia* hybrids. The inference of genomic imprinting for each dataset was limited to those sites that meet our filtering criteria (see Methods). We find 107,206 and 115,490 sites to be fixed and different between VV and GG samples, in the R16A Clark and Wilson datasets, respectively. Limiting the analysis to only fixed and different sites, there are 6,377 and 7,164 genes with at least 2 informative SNPs in the reciprocal hybrids in the R16A Clark data set

and Wilson datasets, respectively. Using this approach, we find no evidence of genomic imprinting in whole adult female samples of *Nasonia* in the R16A Clark data (**Fig 4A**). But for the Wilson data we found two genes that show a pattern of expression consistent with genomic imprinting: CPR35 and LOC103315494. In the VG hybrid, CPR35 shows a bias towards the paternally inherited *N. giraulti* allele at an allele ratio of 65.3% and in the GV hybrid towards the paternally inherited *N. vitripennis* allele, with an allele ratio of 62% (**S2 Table).** CPR35 is a cuticular protein in the RR family member 35. Similarly, LOC103315494 shows bias towards the paternally inherited allele with allele ratios of 65.26% and 61.58% in VG and GV, respectively (**S2 Table)**. Although both imprinted genes, *CPR35* and *LOC103315494,* fall below the mean depth and average number of SNPs per gene, both genes are above the thresholds applied here (**S3 Table**).

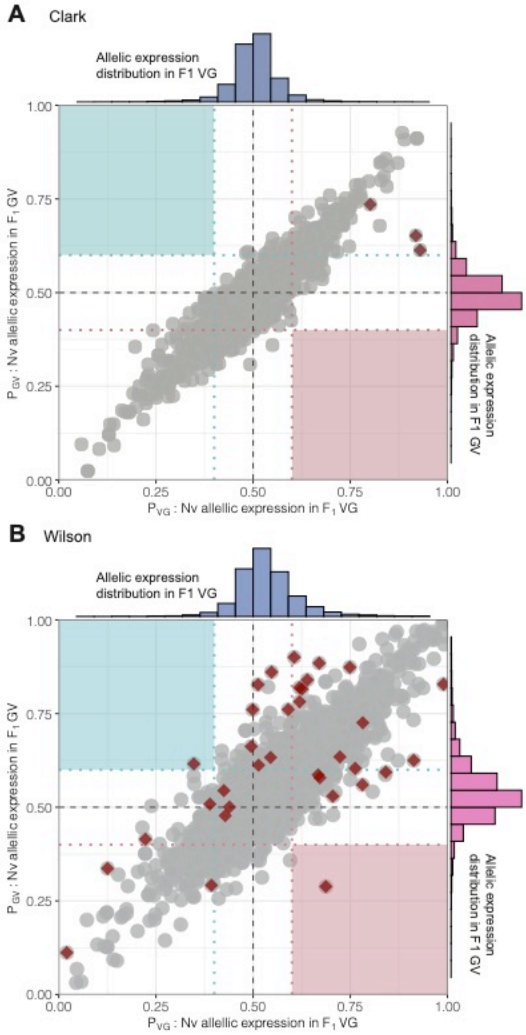


Fig 4. Scatterplots of the expression of the *N. vitripennis* alleles in the two reciprocal hybrids, VG (x-axis) and GV (y-axis), in the R16A Clark (**A**) and Wilson (**B**) datasets. Genes with at least two informative SNPs with a minimum depth of 30 were used (R16A Clark = 6,377, Wilson = 7,164). Genes exhibiting a significant difference in allelic bias between the hybrids (Fisher's exact test, FDR-adj. $p<0.05$) are highlighted in red. Paternally imprinted genes are expected to appear in the upper left corner (light blue box), and maternally imprinted genes in the lower right corner (light pink box). Histograms of the *N. vitripennis* allele expression are shown for VG (blue) and GV (pink).

Allele-specific expression differences in *Nasonia* hybrids

We find three genes with higher expression of the *N. vitripennis* allele in both hybrids, in both datasets, indicative of *cis*-regulatory effects. The genes LOC100123729, LOC100123734, and LOC100113683 show consistent differences in allelic expression between VG and GV hybrids (FDR-*p* ≤0.05) in both datasets, but the ratio of the *N. vitripennis* allele differs between the hybrids (**S2 Table)**. In the R16A Clark dataset: LOC100123729 in the VG hybrids the *N. vitripennis* allele accounts for 93% of the reads, whereas in the GV hybrids this ratio is 61%. In the Wilson dataset, both hybrids showed higher expression of the *N. vitripennis* allele. In the Wilson data, the *N. vitripennis* allele ratio was 61% in VG and 90% in GV. LOC100123729 is located on chromosome 2 and encodes the protein Nasonin-3, which plays a role in inhibiting host insect melanization [31]. Also on chromosome 2 is LOC100123734, annotated as cadherin-23, which is involved in cell attachment by interacting with other proteins in the cell membrane. Both hybrids in both datasets show a higher expression for the *N. vitripennis* allele for LOC100123734. In the R16A Clark data, the ratio of the *N. vitripennis* allele in VG was 92% and in GV 65%. In the Wilson data, the VG hybrids showed less expression for the *N. vitripennis* allele than the GV hybrids, at a ratio of 64% and 84% of the reads, respectively. Finally, LOC100113683, which is located on chromosome 4, and is annotated as a general odorant-binding protein 56d also shows more expression for the *N. vitripennis* allele in both datasets and both hybrids (80.13% and 73.54% for VG and GV in R16A Clark, 78.22% and 72.57% in Wilson). Odorant binding proteins are thought to be involved in the stimulation of the odorant receptors by binding and transporting odorants which activate the olfactory signal transduction pathway [32].

R16A strain retains *N. vitripennis* alleles

R16A is a strain produced by backcrossing an *N. vitripennis* female to an *N. giraulti* male and repeating that for 16 generations [18]. This should give a complete *N. giraulti* nuclear genome

with *N. vitripennis* mitochondria. However, we identified two regions in the R16A strain that still show *N. vitripennis* alleles and named them R16A non-introgressed locus 1 and R16A non-introgressed locus 2 (**S4 Table).** Each region is identified by a single marker that retains the *N. vitripennis* allele. Locus 1 contains 44 genes and Locus 2 contains 14 genes. Both of these regions are found on Chromosome 1, and Locus 2 lies within the confidence intervals of the mortality locus for *N. vitripennis* maternity hybrids identified by Niehuis et al. 2008 [21] (i.e., F2 recombinant hybrids with a *N. vitripennis* cytoplasm showed a significant transmission ratio distortion at this region favoring the *N. vitripennis* allele). R16 A non-introgressed locus 1 harbors a mitochondrial ribosomal gene (39 S ribosomal protein 38) which is a good candidate gene for causing its retention in R16A despite intensive introgression. It would also explain the observed nuc-cytoplasmic effect in F2 recombinant males in a vitripennis cytoplasm, despite the fact that R16A was used as a girault parental line in Gadau et al. (1999) [20]. Gadau et al. interestingly also mapped one of the nuc-cytoplasmic incompatibility loci to chromosome 1 (called LG1 in the manuscript) [20]. Mutations in mitochondrial ribosomal proteins in humans have severe effects [33].

Expression of genes in regions associated with hybrid mortality or nuclear-mitochondrial incompatibility

We compared the location of genes with either significant differential gene expression or significant differences in allele-specific expression between VG and GV hybrids to the location of previously identified mortality-associated loci. Three of the five genes that were called as differentially expressed between VG and GV hybrids in both the R16A Clark and Wilson data sets (**S5 Table**) are located within mortality associated loci. LOC103317241 is located within a locus on Chr 2 that is associated with mortality in VG hybrids, and LOC107981401 and LOC100114950 are within a locus on Chr 4 that is associated with mortality in GV hybrids. Moreover, two of the

three genes showing consistent allele-specific expression in the two data sets are located near one another in the mortality-associated locus on Chr. 2 (LOC100123729 and LOC100123734). None of the genes that are differentially expressed or that exhibit allele-specific expression are located within the 2 loci that retain the *N. vitripennis* genotype in the R16A Clark strain, nor did we find any overlap of these gene sets with either the oxidative phosphorylation or the mitochondrial ribosomal proteins.

Discussion

We successfully replicate the findings from Wang et al. 2016, showing a lack of parent-of-origin effects in *Nasonia* transcriptomes [19]. This replication occurs independently in a different laboratory, with different *Nasonia* individuals derived from a slightly different cross, different bioinformatic pipelines, and sequencing technology. Our results from both the reanalyzed R16A Clark and Wilson datasets could only demonstrate a species-of-origin effect but no parent-of-origin effect within *Nasonia* F1 female hybrids, which may have explained the lack of mortality in the F1 females relative to the F2 recombinant hybrid males. The larger number of differentially expressed genes between the two parental species in our study relative to the Wang et al (2016) [19] (1001 vs 799) is most likely the result of using a standard *N. giraulti* strain (RV2Xu) rather than an introgression strain (R16A) where the nuclear genome of *N. giraulti* was introgressed into a *N. vitripennis* cytoplasm. Additionally, we found genomic regions that resisted introgression in the R16A *Nasonia* strains utilized by Wang et al. 2016 [19]. Furthermore, we present a reproducible workflow for processing raw RNA sequence samples to call differential expression and allele-specific expression openly available on the GitHub page: https://github.com/SexChrLab/Nasonia.

Differences between the R16A Clark and Wilson datasets

The primary difference between the R16A Clark cross and the Wilson cross is the *N. giraulti* strain choice **Fig 1B**. The new crosses presented here used the strain Rv2X(u), which is a pure *N. giraulti* strain that was used for sequencing the genome [17]. Wang et al. 2016 used an introgression strain, R16A, which has a largely *N. giraulti* nuclear genome with an *N. vitripennis* cytoplasm [19]. This strain was produced by mating an *N. vitripennis* female with an *N. giraulti* male, and then repeatedly backcrossing the strain to *N. giraulti* males for a further 15 generations [18]. Hence, both sets of hybrids should be heterozygous at every nuclear locus for species specific markers (though see above for two non-introgressed regions); however, both reciprocal R16A Clark hybrids have *N. vitripennis* mitochondria while the new hybrids have their maternal species' mitochondria. This means that in addition to looking at parent-of-origin effects, our new crosses are uniquely suited to investigate allelic expression biases in the context of nuclear-mitochondrial incompatibility and hybrid dysfunction.

Observed differences in hybrids between data sets

We observe substantially more DEGs between the hybrids, VG and GV, in the Wilson data set compared to the R16A Clark data set. The smaller number of DEGs detected in the R16A Clark data in this particular comparison is likely partially due to the one excluded F_1GV sample (see Materials and methods). Another likely contributing factor is the differences in one parental strain between the Wilson and R16A Clark data sets. The Wilson data presented here consist of inbred parental *N. vitripennis* (strain AsymCX) VV and *N. giraulti* (strain RV2Xu) GG lines, and reciprocal F1 crosses. This cross differs from the R16A Clark data, which used the same *N. vitripennis* strain but rather than a normal *N. giraulti* strain they used the introgression strain, R16A, that has a nuclear genome derived from *N. giraulti* and a cytoplasm/mitochondria derived from *N. vitripennis* (see R16A section). Despite these differences, of the eight genes that are differentially expressed between the VG and GV hybrids. five are shared between both data sets. Although we were not specifically looking for this, we found that three of the five genes showing differential expression

374 in both data sets as well as two of the three genes showing allele (species)-specific expression in
375 both data sets are located in previously identified loci that are associated with the observed F2
376 recombinant male hybrid breakdown from the same crosses [20,21]. These findings point towards
377 an involvement of cis regulatory elements in the genetic architecture of the F2-hybrid male
378 breakdown in *Nasonia*. The finding that, despite using different strains of wasps, we are still able
379 to identify genes associated with these hybrid defects, which bolsters our confidence in further
380 pursuing these genes in our investigation of the genetic architecture of hybrid barriers in *Nasonia*.

381 **The choice of reference and tools does not alter main findings**

382 The authors of the Wang et al. 2016 paper used different computational tools for trimming and
383 alignment than the current study [19]. Additionally, in Wang et al. 2016, the RNAseq reads were
384 aligned to both an *N. vitripennis* and *N. giraulti* reference genome [19]; whereas here, we created
385 a pseudo *N. giraulti* reference genome from the fixed and differentiated sites between the inbred
386 *N. vitripennis* and *N. giraulti* parental lines. Often, different tools and statistical approaches result
387 in different findings [34,35]; however, despite different approaches, we observe the same pattern
388 as what was originally reported in Wang et al. 2016 [19], a lack of parent-of-origin expression in
389 *Nasonia.*

390 **A reproducible workflow for investigating genomic imprinting**

391 Significant factors contributing to irreproducible research include selective reporting, unavailable
392 code and methods, low statistical power, poor experimental design, and raw data not available
393 from the original lab [23]. We replicate a robust experimental design (current study) initially
394 presented in the Wang et al. (2016) [19] and present a new workflow for calling DE and ASE in
395 those two independent but analog *Nasonia* datasets. Both datasets are publicly available for
396 download on the short read archive (SRA) PRJNA260391 and PRJNA613065, respectively. In
397 our analyses of the Wilson data and reanalysis of the R16A data, we corroborated the original

398 findings from Wang et al. 2016 [19]. There are no parent-of-origin effects in *Nasonia*. All
399 dependencies for data processing are provided as a Conda environment, allowing for seamless
400 replication. All code is openly available on GitHub https://github.com/SexChrLab/Nasonia.

401 # Materials and methods

402 ## *Nasonia vitripennis and Nasonia giraulti* inbred and reciprocal F1 hybrid datasets

403 RNA sequence (RNAseq) samples for 4 female samples each from parental species, *N.*
404 *vitripennis* (VV) and *N. giraulti* (GG), and from each reciprocal F1 cross (F_1VG, female hybrids
405 with *N. vitripennis* mothers, and F_1GV, female hybrids with *N. giraulti* mothers), as shown in **Fig**
406 **1A**, were obtained from a 2016 publication [19] from SRA PRJNA299670. We refer to the data
407 from [19] as R16A Clark. One F_1GV RNAseq sample from the R16A Clark dataset (SRR2773798)
408 was excluded due to low quality, as in the original publication [19].
409
410 The newly generated crosses consisted of 12 RNAseq samples of inbred isofemale lines of
411 parental *N. vitripennis* (strain AsymCX) VV and *N. giraulti* (strain RV2Xu) GG lines, and reciprocal
412 F1 crosses F_1VG, and F_1GV. (**Fig 1A**). Whole transcriptome for these samples is available on
413 SRA PRJNA613065. This cross differs from the R16A Clark data, which used the same *N.*
414 *vitripennis* strain but rather than a standard *N. giraulti* strain used an introgression strain, R16A,
415 that has a nuclear genome derived from *N. giraulti* and a cytoplasm/mitochondria derived from *N.*
416 *vitripennis* (see R16A section below) **Fig 1B**. Total RNA was extracted from a pool of four 48 hour
417 post-eclosion adult females using a Qiagen RNeasy Plus Mini kit (Qiagen, CA). RNA-seq libraries
418 were prepared with 2µg of total RNA using the Illumina Stranded mRNA library prep kit and were
419 sequenced on a HiSeq2500 instrument following standard Illumina protocols. Three biological
420 replicates were generated for each parent and hybrid, with 100-bp paired-end reads per replicate.

Sample IDs, parent cross information, and SRA bioproject accession numbers for R16A Clark and Wilson datasets are listed in **S1 Table**.

Quality control

Raw sequence data from both datasets were processed and analyzed according to the workflow presented in **Fig 1C**. The quality of the FASTQ files was assessed before and after trimming using FastQC v0.11 [36] and MultiQC v1.0 [37]. Reads were trimmed to remove bases with a quality score less than 10 for the leading and trailing stand, applying a sliding window of 4 with a minimum mean PHRED quality of 15 in the window and a minimum read length of 80 bases, and adapters were removed using Trimmomatic v0.36 [38]. Pre- and post-trimming multiQC reports for the R16A Clark and Wilson datasets are available on the GitHub page: https://github.com/SexChrLab/Nasonia.

Variant calling

For variant calling, BAM files were preprocessed by adding read groups with Picard's AddOrReplaceReadGroups and by marking duplicates with Picard's MarkDuplicates (https://github.com/broadinstitute/picard). Variants were called using GATK [39–41] and the scatter-gather approach: Sample genotype likelihoods were called with HaplotypeCaller minimum base quality of 2. The resulting gVCFs were merged with CombineGVCFs, and joint genotyping across all samples was carried out with GenotypeGVCFs with a minimum confidence threshold of 10.

Pseudo *N. girault*i reference genome assembly

To create a pseudo *N. girault*i reference genome, fixed differences in the homozygous *N. girault*i and *N. vitripennis* variant call file (VCF) files were identified using a custom Python script,

443 available on the GitHub page: https://github.com/SexChrLab/Nasonia. Briefly, a site was
444 considered to be fixed and different if it was homozygous for the *N. vitripennis* reference allele
445 among all three of the biological VV samples and homozygous alternate among all three of the
446 biological GG samples. Only homozygous sites were included, as the *N. giraulti* and *N. vitripennis*
447 lines are highly inbred. The filtered sites were then used to create a pseudo *N. giraulti* reference
448 sequence with the FastaAlternateReferenceMaker function in GATK version 3.8 (available at:
449 http://www.broadinstitute.org/gatk/). Reference bases in the *N. vitripennis* genome were replaced
450 with the alternate SNP base at variant positions. Following a similar protocol for comparison, we
451 now aligned reads in each sample to the pseudo *N. giraulti* genome reference with HISAT2
452 version 2.1.0, and performed identical preprocessing steps prior to variant calling with GATK
453 version 3.8 HaplotypeCaller.

454 **RNAseq alignment and gene expression level quantification**

455 Trimmed sequence reads were mapped to the NCBI *N. vitripennis* reference genome (assembly
456 accession GCF_009193385.2), as well as the pseudo *N. giraulti* reference using HISAT2 [42].
457 The resulting SAM sequence alignment files were converted to BAM, and coordinates were sorted
458 and indexed with samtools 1.8 [43]. RNAseq read counts were quantified from the *N. vitripennis*
459 as well as the custom *N. giraulti* alignments using Subread featureCounts [44] with the *N.*
460 *vitripennis* gene annotation.

461 **Inference of differential gene expression**

462 Differential expression (DE) analyses were carried out by linear modeling as implemented in the
463 R package *limma* [45]. An average of the reads mapped to each gene in the *N. vitripennis* and
464 the pseudo *N. giraulti* genome references were used in the DE analyses. Counts were filtered to
465 remove lowly expressed genes by retaining genes with a mean FPKM ≤ 0.5 in at least one sample
466 group (VV, GG, VG, or GV). Normalization of expression estimates was accomplished by

calculating the trimmed mean of M-values (TMM) with edgeR [46]. The voom method [47] was then employed to normalize expression intensities by generating a weight for each observation. Gene expression is then reported as log counts per million (logCPM). Gene expression correlation between datasets and between species within each dataset was assessed using Pearson's correlation of mean logCPM values of each gene. Dimensionality reduction of the filtered and normalized gene expression data was carried out using scaled and centered PCA with the *prcomp*() function of base R. Differential expression analysis with voom was carried out for each pairwise comparison between strains (VV, GG, VG, and GV) for each data set. We identified genes that exhibited significant expression differences with an adjusted *p*-value of ≤ 0.01 and an absolute $\log_2$ fold-change ($\log_2$FC) ≤ 2.

Analysis of allele-specific expression in reciprocal F1 hybrids

Allele-specific expression (ASE) levels were obtained using GATK ASEReadCounter [40] with a minimum mapping quality of 10, minimum base quality of 2, and a minimum depth of 30. Only sites with a fixed difference between inbred VV and GG for both R16A Clark and Wilson datasets were used for downstream analysis of allele-specific expression. Allele counts obtained from GATK ASEReadCounter were intersected with the *N. vitripennis* gene annotation file using bedtools version 2.24.0 [48]; the resulting output contained allele counts for each SNP and corresponding gene information. The F1 hybrids' allele counts with gene information was read into R and then filtered to only include genes with at least two SNPs with minimum depth of 30. We counted the number of allele-counts for the reference allele (*N. vitripennis*) and alternative (*N. girault*i) allele at polymorphic SNP positions. We quantified the number of SNPs in each hybrid replicate that 1) showed a bias towards the allele that came from the *N. vitripennis* parent, 2) showed a bias towards the allele that came from the *N. girault*i parent, and 3) showed no difference (ND) in an expression of its parental alleles. The significance of allelic bias was determined using Fisher's exact test. Significant genes were selected using a Benjamini-

492 Hochberg false discovery rate FDR-adjusted *p*-value threshold of 0.05. As *Nasonia* are
493 haplodiploid, all ASE analyses were carried out on the diploid female hybrids.

494 **Identifying loci associated with hybrid mortality**

495 *Nasonia* recombinant F2 hybrid males (haploid sons of F1 female hybrids) suffer mortality during
496 development that differs between VG and GV hybrids [18]. Niehuis et al. 2008 identified four
497 genomic regions associated with this mortality (i.e., regions in which one parent species' alleles
498 are underrepresented due to mortality during development); three are associated with mortality in
499 hybrids with *N. vitripennis* maternity and one is associated with hybrids with *N. giraulti* maternity
500 [21]. Gibson et al. 2013 later identified a second locus related to mortality in the hybrids with *N.*
501 *giraulti* maternity [22]. Given that the F2 hybrid females analyzed here experience far less
502 mortality than their haploid male offspring, we hypothesized that these diploid females may use
503 biased allelic expression to rescue themselves from the mortality. To compare our results with
504 these previous studies, we had to map the previous loci to the latest *Nasonia* assembly (PSR1.1,
505 [49]). Niehuis et al. 2008 defined their candidate loci based on the genetic distance along the
506 chromosome (centimorgans) [21]. The physical locations of the markers along the chromosomes
507 were later identified by Niehius et al. 2010 [50]. Using the genetic distances between these
508 markers in both the 2008 and 2010 Niehuis *et al*. studies [21,50], we calculated the conversion
509 ratio between the genetic distances in these two studies (**S6 Table**). We then converted those
510 2008 genetic distances that correspond to the 95% Confidence Intervals for these loci to the
511 genetic distances reported by Niehuis et al. 2010 [50], which used an Illumina Goldengate
512 Genotyping Array (Illumina Inc., San Diego, USA) to produce a more complete and much higher
513 resolution genetic map of *Nasonia*. This array uses Single Nucleotide Polymorphisms (SNPs) to
514 genotype samples at ~1500 loci, which allowed us to identify SNP markers that closely bound the
515 mortality loci from the 2008 study. Gibson et al. 2013 used the same genotyping array, so this
516 conversion was unnecessary for converting the second mortality locus in *N. giraulti* maternity

hybrids [22]. We used the 100bp of sequence flanking each SNP marker to perform a BLAST search of the PSR1.1 assembly and to identify their positions. We then used all of the PSR1.1 annotated genes within these loci to look for enrichment of genes showing biased expression. Mortality loci and genomic location are reported in **S4 Table**.

Additional gene categories of interest

Previous work has identified potential classes of genes that may be involved in nuclear-mitochondrial incompatibilities in *Nasonia,* the oxidative phosphorylation genes [51] and the mitochondrial ribosomal proteins [52]. We used the annotated gene sets from these studies to test for enrichment of genes with biased allelic expression. Lists of the genes of interest and their genomic location is reported in **S4 Table**.

Analysis of R16A strain

In order to assess whether the introgression of the *N. giraulti* nuclear genome into the R16A Clark strain is complete, we analyzed two samples of the R16A strain using the Illumina Goldengate Genotyping Array used in Niehuis *et al*. 2010 [50]. We searched for SNP markers that retained the *N. vitripennis* allele and only considered markers that consistently identified the proper allele in both parent species controls and that were consistent across both R16A samples, leaving 1378 markers. We defined a locus as all of the sequences between the two markers that flank a marker showing the *N. vitripennis* allele (**S2 Table**). As above, we performed a BLAST search of the PSR1.1 assembly to identify the positions of these markers. We identified all genes from the PSR1.1 assembly that lie between the flanking markers and further analyzed their expression patterns.

Scripts and gene lists used to analyze these data are publicly available on GitHub, https://github.com/SexChrLab/Nasonia.

Acknowledgments

The authors acknowledge Research Computing at Arizona State University for providing HPC resources that have contributed to the research results reported within this paper.

Funding Disclosure

ARCS Spetzler Scholar additionally supported KCO. HMN was supported by an ASU Center for Evolution and Medicine postdoctoral fellowship. MAW was supported by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) grant R35GM124827. JDG was supported by the Division of Integrative Organismal Systems (IOS) of the National Science Foundation (NSF) grant 1145509 and by research funds provided by Georgia Southern University. JG was in part supported by a grant from the German Research Foundation (DFG) to N.S. (281125614/GRK2220, Project B7). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Supplemental Data

S1 Fig. **Volcano plots for differential expression and venn diagram of DEGs between the datasets when taking the average of the counts when aligned to *N. vitripennis* and to pseudo *N. giraulti* reference genome.** Volcano plots of DEGs detected between the different comparisons involving *N. vitripennis*, *N. giraulti*, and the two reciprocal F_1 hybrids in the R16A Clark (left side) and Wilson (right side) datasets. Venn diagrams of the overlap of significant DEGs in each comparison is shown.

S1 Table. Sample identifiers. The samples for each dataset used in the project are provided here. Samples from this study are uploaded at https://www.ncbi.nlm.nih.gov/sra/PRJNA613065.

S2 Table. Allele-specific expression differences between hybrids. The number of allele-counts for the reference allele (*N. vitripennis*) and alternative (*N. giraulti*) allele at polymorphic SNPs within a gene. Minimum of two SNPs for a gene to be included. The significance of allelic bias was determined using Fisher's exact test. Significant genes were selected using a Benjamini-Hochberg false discovery rate FDR-adjusted *p*-value threshold of 0.05.

S3 Table. Mean and median allele and gene depth for Wilson dataset. Mean and median allele and gene depth for each GV and VG sample in the Wilson data set. Number of SNPs for all genes, *CPR35,* and *LOC103315494.*

S4 Table. Genomic location of mortality loci and gene sets of interest. Previously reported loci associated with mortality in *Nasonia* hybrids. 95% Confidence Intervals of loci identified in Niehuis et al. 2008 were converted to genetic distances along the chromosomes and the closest SNP markers from Niehius et al. 2010 were identified [21,50]. SNP markers for the locus identified

579 in Gibson et al. 2013 were used directly [22]. The SNP marker locations in the PSR1.1 assembly
580 were found via BLAST and all genes within the bounds of these markers are included. The two
581 non-introgressed regions from the R16A strain are included as well as genes from two
582 mitochondria-associated pathways, the oxidative phosphorylation pathway [51] and the
583 mitochondrial ribosomal proteins [52].
584
585 **S5 Table. Directional bias of differentially expressed genes between VG and GV in Clark**
586 **and Wilson datasets.** Five genes that were called as differentially expressed between VG and
587 GV hybrids in both the Clark and Wilson data sets.
588
589 **S6 Table. Locus conversion calculations.** Calculations for converting the genetic map
590 positions (centimorgan, cM) of mortality loci identified by Niehuis et al. 2008 to the physical
591 chromosomal positions of the latest genome assembly (PSR1.1) [21].
592

References

1. Reik W, Walter J. Genomic imprinting: parental influence on the genome. Nature Reviews Genetics. 2001. pp. 21–32. doi:10.1038/35047554

2. Ishida M, Moore GE. The role of imprinted genes in humans. Mol Aspects Med. 2013;34: 826–840.

3. Isles AR, Davies W, Wilkinson LS. Genomic imprinting and the social brain. Philos Trans R Soc Lond B Biol Sci. 2006;361: 2229–2237.

4. Moore T, Haig D. Genomic imprinting in mammalian development: a parental tug-of-war. Trends Genet. 1991;7: 45–49.

5. Lawson HA, Cheverud JM, Wolf JB. Genomic imprinting and parent-of-origin effects on complex traits. Nat Rev Genet. 2013;14: 609–617.

6. Yan H, Bonasio R, Simola DF, Liebig J, Berger SL, Reinberg D. DNA methylation in social insects: how epigenetics can control behavior and longevity. Annu Rev Entomol. 2015;60: 435–452.

7. Queller DC. Theory of genomic imprinting conflict in social insects. BMC Evol Biol. 2003;3: 15.

8. Haig D. Intragenomic conflict and the evolution of eusociality. J Theor Biol. 1992;156: 401–403.

9. Galbraith DA, Kocher SD, Glenn T, Albert I, Hunt GJ, Strassmann JE, et al. Testing the kinship theory of intragenomic conflict in honey bees (Apis mellifera). Proc Natl Acad Sci U S A. 2016;113: 1020–1025.

10. Smith NMA, Yagound B, Remnant EJ, Foster CSP, Buchmann G, Allsopp MH, et al. Paternally-biased gene expression follows kin-selected predictions in female honey bee embryos. Mol Ecol. 2020;29: 1523–1533.

11. Kocher SD, Tsuruda JM, Gibson JD, Emore CM, Arechavaleta-Velasco ME, Queller DC, et al. A Search for Parent-of-Origin Effects on Honey Bee Gene Expression. G3 . 2015;5: 1657–1662.

12. Gibson JD, Arechavaleta-Velasco ME, Tsuruda JM, Hunt GJ. Biased Allele Expression and Aggression in Hybrid Honeybees may be Influenced by Inappropriate Nuclear-Cytoplasmic Signaling. Frontiers in Genetics. 2015. doi:10.3389/fgene.2015.00343

13. Hunt GJ, Guzmán-Novoa E, Fondrk MK, Page RE Jr. Quantitative trait loci for honey bee stinging behavior and body size. Genetics. 1998;148: 1203–1213.

14. Hunt GJ. Flight and fight: A comparative view of the neurophysiology and genetics of honey bee defensive behavior. J Insect Physiol. 2007;53: 399–410.

15. Shorter JR, Arechavaleta-Velasco M, Robles-Rios C, Hunt GJ. A Genetic Analysis of the Stinging and Guarding Behaviors of the Honey Bee. Behavior Genetics. 2012. pp. 663–

674. doi:10.1007/s10519-012-9530-5

16. Alaux C, Sinha S, Hasadsri L, Hunt GJ, Guzmán-Novoa E, DeGrandi-Hoffman G, et al. Honey bee aggression supports a link between gene regulation and behavioral evolution. Proc Natl Acad Sci U S A. 2009;106: 15400–15405.

17. Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, et al. Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. Science. 2010;327: 343–348.

18. Breeuwer JAJ, Werren JH. HYBRID BREAKDOWN BETWEEN TWO HAPLODIPLOID SPECIES: THE ROLE OF NUCLEAR AND CYTOPLASMIC GENES. Evolution. 1995;49: 705–717.

19. Wang X, Werren JH, Clark AG. Allele-Specific Transcriptome and Methylome Analysis Reveals Stable Inheritance and Cis-Regulation of DNA Methylation in Nasonia. PLOS Biology. 2016. p. e1002500. doi:10.1371/journal.pbio.1002500

20. Gadau J, Page RE Jr, Werren JH. Mapping of hybrid incompatibility loci in Nasonia. Genetics. 1999;153: 1731–1741.

21. Niehuis O, Judson AK, Gadau J. Cytonuclear genic incompatibilities cause increased mortality in male F2 hybrids of Nasonia giraulti and N. vitripennis. Genetics. 2008;178: 413–426.

22. Gibson JD, Niehuis O, Peirson BRE, Cash EI, Gadau J. Genetic and developmental basis of F2 hybrid breakdown in Nasonia parasitoid wasps. Evolution. 2013;67: 2124–2132.

23. Baker M. 1,500 scientists lift the lid on reproducibility. Nature. 2016. pp. 452–454. doi:10.1038/533452a

24. Casadevall A, Fang FC. Reproducible science. Infect Immun. 2010;78: 4972–4975.

25. Baker M. Irreproducible biology research costs put at $28 billion per year. Nature. 2015;533. Available: http://www.target-biomed.de/resources/Irreproducible-biology-research.pdf

26. Freedman LP, Inglese J. The Increasing Urgency for Standards in Basic Biologic Research. Cancer Research. 2014. pp. 4024–4029. doi:10.1158/0008-5472.can-14-0925

27. Simoneau J, Dumontier S, Gosselin R, Scott MS. Current RNA-seq methodology reporting limits reproducibility. Brief Bioinform. 2019. doi:10.1093/bib/bbz124

28. Su Z, Łabaj PP, Li S, Thierry-Mieg J, Thierry-Mieg D, Shi W, et al. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nat Biotechnol. 2014;32: 903–914.

29. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Research. 2008. pp. W5–W9. doi:10.1093/nar/gkn201

30. NCBI Resource Coordinators, NCBI Resource Coordinators. Database Resources of the National Center for Biotechnology Information. Nucleic Acids Research. 2017. pp. D12–D17. doi:10.1093/nar/gkw1071

667 31. Tian C, Wang L, Ye G, Zhu S. Inhibition of melanization by a Nasonia defensin-like peptide:
668 implications for host immune suppression. J Insect Physiol. 2010;56: 1857–1862.

669 32. He Y, Wang K, Zeng Y, Guo Z, Zhang Y, Wu Q, et al. Analysis of the antennal
670 transcriptome and odorant-binding protein expression profiles of the parasitoid wasp
671 Encarsia formosa. Genomics. 2020. pp. 2291–2301. doi:10.1016/j.ygeno.2019.12.025

672 33. Sylvester JE, Fischel-Ghodsian N, Mougey EB, O'Brien TW. Mitochondrial ribosomal
673 proteins: candidate genes for mitochondrial disease. Genet Med. 2004;6: 73–80.

674 34. Schaarschmidt S, Fischer A, Zuther E, Hincha DK. Evaluation of Seven Different RNA-Seq
675 Alignment Tools Based on Experimental Data from the Model Plant Arabidopsis thaliana.
676 Int J Mol Sci. 2020;21. doi:10.3390/ijms21051720

677 35. Del Fabbro C, Scalabrin S, Morgante M, Giorgi FM. An extensive evaluation of read
678 trimming effects on Illumina NGS data analysis. PLoS One. 2013;8: e85024.

679 36. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online].
680 Online. 2010. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

681 37. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for
682 multiple tools and samples in a single report. Bioinformatics. 2016;32: 3047–3048.

683 38. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
684 Bioinformatics. 2014;30: 2114–2120.

685 39. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et
686 al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best
687 practices pipeline. Curr Protoc Bioinformatics. 2013;43: 11.10.1–33.

688 40. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The
689 Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA
690 sequencing data. Genome Res. 2010;20: 1297–1303.

691 41. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for
692 variation discovery and genotyping using next-generation DNA sequencing data. Nat
693 Genet. 2011;43: 491–498.

694 42. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory
695 requirements. Nat Methods. 2015;12: 357–360.

696 43. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
697 Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–2079.

698 44. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for
699 assigning sequence reads to genomic features. Bioinformatics. 2014;30: 923–930.

700 45. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential
701 expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res.
702 2015;43: e47.

703 46. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential
704 expression analysis of digital gene expression data. Bioinformatics. 2010;26: 139–140.

705 47. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis
706 tools for RNA-seq read counts. Genome Biol. 2014;15: R29.

707 48. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features.
708 Bioinformatics. 2010. pp. 841–842. doi:10.1093/bioinformatics/btq033

709 49. Benetta ED, Antoshechkin I, Yang T, Nguyen HQM, Ferree PM, Akbari OS. Genome
710 Elimination Mediated by Gene Expression from a Selfish Chromosome.
711 doi:10.1101/793273

712 50. Niehuis O, Gibson JD, Rosenberg MS, Pannebakker BA, Koevoets T, Judson AK, et al.
713 Recombination and its impact on the genome of the haplodiploid parasitoid wasp Nasonia.
714 PLoS One. 2010;5: e8597.

715 51. Gibson JD, Niehuis O, Verrelli BC, Gadau J. Contrasting patterns of selective constraints in
716 nuclear-encoded genes of the oxidative phosphorylation pathway in holometabolous
717 insects and their possible role in hybrid breakdown in Nasonia. Heredity . 2010;104: 310–
718 317.

719 52. Burton RS, Barreto FS. A disproportionate role for mtDNA in Dobzhansky-Muller
720 incompatibilities? Molecular Ecology. 2012. pp. 4942–4957. doi:10.1111/mec.12006