A Search for Parent-of-origin Effects

in the Parasitoid Jewel Wasp *Nasonia vitripennis*

by

Avery Elizabeth Underwood

A Thesis Presented in Partial Fulfilment
of the Requirements for the Degree
Master of Science

Approved April 2019 by the
Graduate Supervisory Committee

Melissa Wilson, Chair
Kenneth Buetow
Gillian Gile

ARIZONA STATE UNIVERSITY

May 2019

ABSTRACT

In most diploid cells, autosomal genes are equally expressed from the paternal and maternal alleles resulting in biallelic expression. However, as an exception, there exists a small number of genes that show a pattern of monoallelic or biased-allele expression based on the allele's parent-of-origin. This phenomenon is termed genomic imprinting and is an evolutionary paradox. The best explanation for imprinting is David Haig's kinship theory, which hypothesizes that monoallelic gene expression is largely the result of evolutionary conflict between males and females over maternal involvement in their offspring. One previous RNAseq study has investigated the presence of parent-of-origin effects, or imprinting, in the parasitic jewel wasp *Nasonia vitripennis* (*N. vitripennis*) and its sister species *Nasonia giraulti* (*N. giraulti*) to test the predictions of kinship theory in a non-eusocial species for comparison to a eusocial one. In order to continue to tease apart the connection between social and eusocial Hymenoptera, this study proposed a similar RNAseq study that attempted to reproduce these results in unique samples of reciprocal F1 *Nasonia* hybrids. Building a pseudo *N. giraulti* reference genome, differences were observed when aligning RNAseq reads to a *N. vitripennis* reference genome compared to aligning reads to a pseudo *N. giraulti* reference. As well, no evidence for parent-of-origin or imprinting patterns in adult *Nasonia* were found. These results demonstrated a species-of-origin effect. Importantly, the study continued to build a repository of support with the aim to elucidate the mechanisms behind imprinting in an excellent epigenetic model species, as it can also help with understanding the phenomenon of imprinting in complex human diseases.

DEDICATION

To quote Abraham Lincoln: “All that I am, or hope to be, I owe to my mother.”

ACKNOWLEGEMENTS

I would first like to thank my thesis advisor, Dr. Melissa Wilson, who supported me in completing my master's degree and thesis. She is an inspirational leader who is always there for her students. I am appreciative of her insightful mentorship along the way, as well as the many opportunities that working in the Wilson Lab has afforded me. I would also like to acknowledge Dr. Kenneth Buetow and Dr. Gillian Gile as committee members, and as second and third readers of this thesis. I am grateful for their expertise in the field, and for their generous time and feedback. Furthermore, I would like to thank Dr. Juergen Gadau and his lab at Arizona State University for providing the sequencing data, as well as for his collaboration on this project.

I would also like to take the time to thank Wilson Lab members for their support. I would especially like to acknowledge Dr. Wilson's doctoral student, Kimberly Olney, for her valuable mentoring and guidance on this project and thesis.

Finally, thank you to the School of Life Sciences and the Graduate College at Arizona State University for the help and support. This would not have been possible without all the amazing faculty and staff who helped me achieve my academic goals.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

1. Imprinting: Parent-of-origin Specific Silencing of Alleles

In most diploid cells, autosomal genes are equally expressed from the paternal and maternal alleles resulting in biallelic expression. However, as an exception, there exists a small number of genes that show a pattern of monoallelic or biased-allele expression based on the allele's parent-of-origin (Ishida and Moore, 2013). This phenomenon is termed genomic imprinting and was first coined by the cytogeneticist Helen Crouse in 1960 to describe the programmed elimination of paternally derived X chromosomes in sciarid flies (Crouse, 1960). For most genes under sexual reproduction, mammals inherit two working copies that are functional in their cells: one copy from mom, and one copy from dad. When a gene is imprinted, though, only one of the two copies are functional in the offspring, and the copy that is active is dependent on which parent it was inherited from. In other words, for a maternally imprinted gene, the gene copy from the mother is always turned "off" and is not expressed, whereas the copy from the father is always turned "on" and is expressed. The inverse holds true of a paternally imprinted gene. Typically, the inactive copy is epigenetically silenced through the process of DNA methylation during gametogenesis, although other mechanisms such as histone deacetylation can also result in uniparental gene expression (Reik & Walter, 2001). If the parental copy of a gene is to be imprinted during the formation of the gamete, then methyl groups are added to cytosine nucleotides in CG dinucleotides to reduce gene expression in that region (Barlow and Bartolomei, 2014). This acts as a type of label that the cell then recognizes and knows not to express or transcribe that copy of the gene. Importantly this methylation is maintained during DNA replication and in this

way, the marking is not lost when the cells divide (Wood and Oakley, 2006). Genomic imprinting is therefore one type of transcription regulation in higher eukaryotes as it regulates whether a gene is on or off. It is also an example of epigenetic alteration in DNA since there is no change in the DNA sequence of the allele, only a change in the allele's expression state depending on whether it is transmitted via a female or male gamete (Köhler et al., 2012). Thus, genomic imprinting is characterized as an epigenetic parent-of-origin effect that is reset after gametogenesis in every new generation.

Since the discovery of mammalian genomic imprinting in the 1980's with a set of experiments involving the nuclear transfer of mouse embryos (Surani et al., 1984), attempts to quantify and define the mechanisms for imprinting in mammals have been made. According to the Harwell and University of Otago online database, over 100 imprinted genes in mice have been censused to date, with approximately half of those same genes also classified as imprinted in humans (Morison et al., 2001). General characteristics about the DNA sequence of imprinted genes have also been identified in an attempt to distinguish them from non-imprinted genes (Neumann et al., 1995). Studies have shown that the methylated regions in imprinted genes typically overlap with CpG islands, and clustered tandem repeats are often found near or within those islands, theorized to be involved in the regulation of imprinting (Dindot et al., 2009; Hutter et al., 2006).

In addition, much of our understanding of the phenotypic consequences of imprinting has come from studying complex human disorders (Reik & Walter, 2001). To illustrate how heritable disease can also be caused by mistakes in epigenetics, consider how errors in imprinting can lead to Prader-Willi syndrome (PWS) and Angelman

syndrome (AS), two devastating diseases where imprinting is responsible for their unique presentation. Each phenotypically very different disorders, PWS is a debilitating disease that causes symptoms of cognitive deficiencies as well as obesity and excessive hunger where patients can almost "eat themselves to the point of death" (Butler, 2011). Described in the literature as "Happy Puppet syndrome," Angelman disorder is a disease that causes symptoms such as sleep, mental, and developmental deficiencies as well as uncontrollable laughter (Clayton-Smith & Pembrey, 1992; Sarkar et al., 2011). Both diseases can be caused by deletions in the region of chromosome 15q11 to q13 (Glenn et al., 1997), but they can also be caused by mistakes in imprinting. Since a gene that is imprinted is silenced or unable to be transcribed, it is functionally equivalent to being deleted. Whether it is deleted or silenced due to imprinting, the gene is not expressed either way. As such, these two disorders are commonly characterized together because they both involve a gene located on human chromosome 15 called UBE3A that codes for ubiquitin ligase (Kishino et al., 1997). In a healthy individual, the maternal copy of this gene is normally expressed or on, while the paternal copy is imprinted and therefore silenced or off. However, if the maternal copy of the gene mistakenly gets imprinted during development, then without a functional copy of UBE3A from the maternal chromosome, a person develops Angelman syndrome (Nicholls et al., 1998). In the same region on chromosome 15, there is a section of multiple genes that are normally imprinted on the maternal chromosome—the opposite scenario as in Angelman syndrome (Knoll et al., 1989). The paternal chromosome is the one that contributes to gene expression as the maternal region is silenced or imprinted. However, if the paternal copy is imprinted by mistake, or if there is a deletion or mutation in this region, then the result

is Prader-Willi syndrome. These sister disorders demonstrate how genomic imprinting is an important process of non-Mendelian inheritance, and how further study is still needed to better understand the phenomenon, especially in multisystem diseases such as PWS and AS.

2. Kinship Theory

Genomic imprinting is an evolutionary paradox (Ferguson-Smith, 2011). In theory, natural selection is expected to favor expression of both alleles from each parent to protect against recessive mutations that could possibly lead to a loss of function in an affected gene (Kondrashov and Crow, 1991). What is the benefit then of silencing one copy of a gene, rendering the organism haploid instead of diploid at that locus? Clearly, the evolutionary benefits must outbalance or outweigh the vulnerability that comes from functional haploidy.

Several explanations have been proposed to investigate the evolutionary origin behind genomic imprinting, the most supported being David Haig's kinship theory, also referred to as the parental conflict theory, or conflict hypothesis (Moore & Haig, 1991). The theory hypothesizes that monoallelic gene expression is largely the result of evolutionary conflict between males and females over maternal involvement in their offspring, and it predicts that paternally-derived genes will favor greater demands on mothers than maternally-derived genes. In other words, some paternally expressed genes are selected to maximize the survival of the offspring, at the expense of the mother and gestated offspring of different fathers, whereas maternally expressed genes are selected to ration resources to ensure the mother’s survival and equal allocation of nutrients among her offspring (Haig, 2000).

In humans, one of the best studied imprinted regions on the human genome, chromosome 11p15.5, supports the predictions outlined in Haig's theory. The most well-known gene on this chromosome is insulin-like growth factor 2 (IGF2), a fetal growth promoter, and it shows an interesting pattern of expression: the maternal copy of the gene is silenced, whereas the paternal copy is expressed and promotes fetal growth that continues into post-natal developmental (Barlow et al., 1991; DiChiara et al., 1991). Found also on chromosome 11p15.5 is the gene CDKN1C, cycle-dependent kinase inhibitor 1C, that acts as a negative regulator of cell proliferation as it inhibits progression through the cell cycle. This gene also shows a pattern of imprinting, however, in the opposite direction: the paternal copy of the gene in this case is silenced, whereas the maternal copy is instead expressed and is growth-limiting (Haig, 2004). An explanation for this simultaneous parental tug-of-war between mom and dad over offspring growth could be explained by kinship theory. On one hand, a gene expressed on dad's chromosome is accelerating growth in an attempt to selfishly increase the fitness of his offspring, while at the same time, a gene on mom's corresponding chromosome is attempting to slam the brakes on offspring growth in her best interest and in the best interest of subsequent progeny (Bartolomei & Tilghman, 1997).

To date, genomic imprinting has been consistently found in placental mammals and marsupials, however, it has not yet been observed in egg-laying mammals, birds, fish, or reptiles (Renfree et al., 2009). Outside the animal kingdom, imprinting has independently evolved in the endosperm (a tissue that supports embryo development) of flowing plants, such as in maize and in angiosperms like Arabidopsis (Kinoshita et al.,

1999; Kermicle & Alleman, 1990). This close association between imprinting and genes with placental-like function has continued to lend support to kinship theory.

3. Imprinting and Kinship in Hymenoptera

The third largest insect order, Hymenopteran, consists of approximately 115,000 species and includes some of the most notable eusocial insects—ants, bees, and wasps (Ayasse et al., 2001). Hymenoptera has long been recognized as an excellent candidate for the study of genomic imprinting due to its close colony interactions and resource allocation, as well as relatedness asymmetry from haplodiploidy (Dobata & Tsuji, 2012). In haplodiploidy sex determination, males come from eggs that are unfertilized, so they are haploid with only one set of chromosomes. Females, however, come from fertilized eggs so they are diploid and possess two sets of genetic material. This means that fathers are unique in that they only make a genetic contribution to their daughters, as haploid males have no father and only receive chromosomes from their mother. Furthermore, a large number of social systems that are present in Hymenoptera, such as sex allocation, division of labor and resources, and brood rearing may lead to potential reproductive and actual social conflict (Ratnieks et al., 2006). It has therefore been noted that the kinship theory of genomic imprinting might particularly apply to the evolution of social behavior in ants, bees, and wasps, since the theory predicts that conflicts of interest between parents can result in opposed patterns of maternally and paternally expressed alleles in the offspring (Haig, 1992; Queller, 2003; Kronauer, 2008; Wild & West, 2009). Although more independent tests are still needed, recent studies have discovered evidence for the presence of genomic imprinting in two bumble bee genes (Amarasinghe et al., 2015) and paternal effects on worker defensive behavior in honey bees (Galbraith et al., 2016).

However, one study has raised concern regarding the reliability of methylation estimates (Remnant, 2016) as it has been shown that DNA methylation varies substantially among social Hymenoptera (Kronforst et al., 2008).

4. The Parasitoid Jewel Wasp: *N. vitripennis*

The jewel wasp *Nasonia vitripennis* (*N. vitripennis*) is known as the "*Drosophila melanogaster* of the Hymenoptera order," or the "lab rat" of parasitic wasp species (Pultz & Leaf, 2013). The genus *Nasonia* (Hymenoptera: Pteromalidae) consists of four closely related species of non-eusocial parasitoid wasps: *N. vitripennis* found globally, *N. giraulti* found in eastern North America, *N. longicornis* found in western North America, and most recently discovered in New York State, *N. oneida* (Darling and Werren 1990; Raychoudhury et al., 2010). *N. vitripennis* split first ~1 million years ago and the other three sister species diverged later, ~0.2 million years ago (Campbell et al., 1994). All four species are naturally reproductively isolated from each other as a result of a *Wolbachia* bacterial infection—an endosymbiont that causes egg-sperm incompatibility. Nevertheless, after antibiotic treatment in lab, viable hybrid offspring can be produced between any of the four species (Bordenstein et al., 2001). Easy to rear in lab with approximately 500 offspring per female, the 2–3 mm female parasitoid wasp lays 30-50 eggs at a time in fly pupae, such as in houseflies or fleshflies like blowflies and Sarcophaga, ultimately killing the host two weeks later after eclosion (Desjardins et al., 2010).

Nasonia has been studied since the 1950's (Whiting, 1967) and has rapidly emerged in the last decade as a fantastic model for epigenetic studies due to its haplodiploidy form of sex determination and ability to inbreed nearly isogenic lines

(Gadau et al., 2008). Under haplodiploidy, females will develop from fertilized eggs and be diploid with two sets of chromosomes, contributing 50% of her genes to each offspring. In contrast, males will develop from unfertilized eggs and be haploid with only one set of chromosomes, contributing genetically identical sperm to his female offspring. As such, a female that has not mated will produce male offspring containing only her genetic material, since haploid males have no father. Haplodiploidy therefore helps facilitate the detection of inherited traits from parent to offspring, especially recessive phenotypes that are always expressed in males (Beukeboom & van de Zande, 2010; Breed & Moore, 2016). Additionally, since *Nasonia* does not have single-locus complementary sex determination, fully homozygous strains for study can be inbred (i.e., brother-sister and mother-son matings) for many generations without any observable defects such as diploid males (Beukeboom & Desplan, 2003).

In the past decade, newly developed genome resources have helped to advance *Nasonia* as an emerging model organism for genetic research (Shuker et al., 2003). For example, genetic markers are known from all five chromosomes, dense linkage maps are available, and whole genome sequences of *N. vitripennis* (6x Sanger coverage) and two interfertile species, *N. giraulti* and *N. longicornis* (both 1x Sanger coverage), are now available as well (Werren et al., 2010). Numerous laboratories, for instance, have already utilized the expanding genetic toolbox of *Nasonia* to investigate a variety of questions (Werren & Loehlin, 2009). To cite a few notable examples: hybrid courtship behavior (Beukeboom & van den Assem, 2001), sex ratio control (Shuker & West, 2004), embryo development (Lynch et al., 2006), incompatible nuclear-mitochondrial interactions

(Ellison et al., 2008), maternal control in sex determination (Verhulst et al., 2010), and evolution of sex pheromones (Niehuis et al., 2013).

5. A Search for Parent-of-origin Effects in *Nasonia*: Study Design

One previous RNAseq study has investigated the presence of parent-of-origin effects in *Nasonia* in order to test the predictions of kinship theory in a non-eusocial species for comparison to a eusocial one. For over 8,000 genes, the study quantified gene expression in the transcriptome of *N. vitripennis* and *N. giraulti*, as well as allele-specific expression in the reciprocal F1 hybrid daughters of the two sister species. The concluding results found no support for genomic imprinting in adult *Nasonia* as no parent-of-origin patterns were identified (Wang et al., 2016). In order to continue to tease apart the connection between social and eusocial Hymenoptera, we propose a similar RNAseq study that attempts to reproduce these results in our own samples of reciprocal F1 *Nasonia* hybrids. We also seek to investigate whether there is significant reference bias when aligning F1 hybrids to a *N. vitripennis* or *N. giraulti* reference genome. To avoid using a low coverage *N. giraulti* genome (1x Sanger sequencing), we propose to construct a "pseudo" *N. giraulti* reference genome. It is important to continue to build a repository of support with the aim to elucidate the mechanisms behind imprinting in this excellent epigenetic model species, as it can also help us better understand the phenomenon of imprinting in complex human diseases.

METHODS

1. Sampling Preparation and Sequencing

Courtesy of Dr. Juergen Gadau's lab at Arizona State University: RNAseq samples consisted of reciprocal F1 crosses (F1 $V_{paternal}G_{maternal}$ and F1 $G_{paternal}V_{maternal}$) of sufficiently inbred *N. vitripennis* and *N. giraulti* lines, shown below in Figure 1 and Figure 2. The parents and hybrids were sequenced on an Illumina instrument following standard Illumina RNAseq protocols. Three biological replicates were performed for each hybrid and parent, with 100-bp paired-end short reads per replicate.

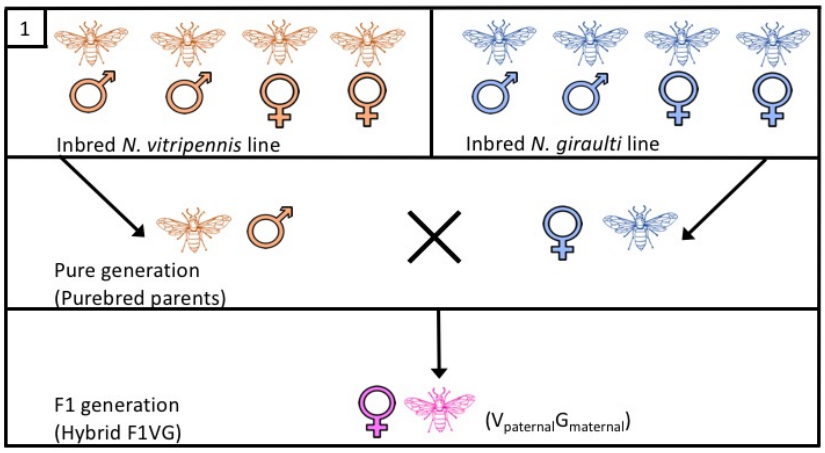


Figure 1. Reciprocal F1 Cross for Hybrid $V_{paternal}G_{maternal}$.

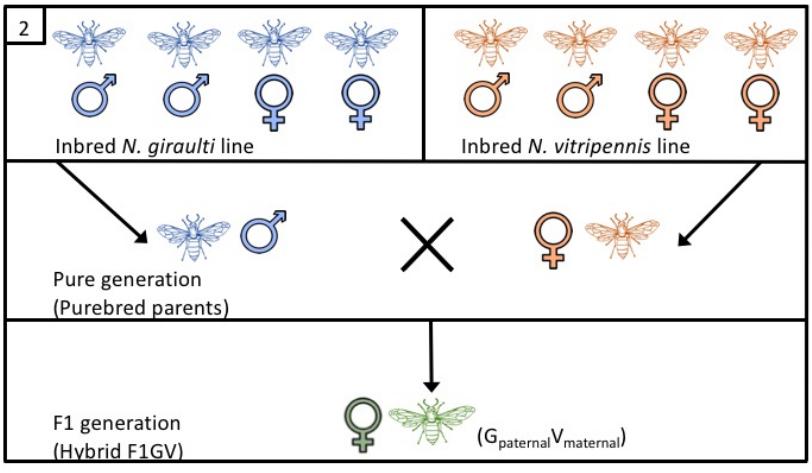


Figure 2. Reciprocal F1 Cross for Hybrid $G_{paternal}V_{maternal}$.

2. Quality Control

Raw sequence reads were assessed for quality using FastQC version 0.11.6 (available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc). With Trimmomatic version 0.38 (Bolger et al. 2014), Illumina TruSeq adapter content was removed, and the RNAseq reads were further trimmed for quality using a sliding window of 4 bases to clip reads when the average quality per base dropped below a PHRED-scaled threshold quality of 30 (i.e., the chances that a base is called incorrectly are 1 in 1000). Trimmed RNAseq reads were checked for improved quality using MultiQC version 1.6 (Ewels et al. 2016) after initial trimming and filtering.

3. Obtaining the *N. vitripennis* Reference Genome

We downloaded the *N. vitripennis* reference genome (total length of 239.8 Mb) and gene annotation files from the National Center for Biotechnology Information (NCBI) (available at: https://www.ncbi.nlm.nih.gov/genome?term=nasonia%20vitripennis) to be used for aligning RNAseq reads. We obtained the Nvit_2.1 genome assembly version of all regions, including reference chromosomes (1-5), unplaced scaffolds, and contigs (Werren et al. 2010). The NCBI *N. vitripennis* Annotation Release 102 was also downloaded for access to the comprehensive gene annotation records (available at: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Nasonia_vitripennis/102/). Using Picard tools version 1.119 (available at: http://broadinstitute.github.io/picard/), a dictionary of the contig names and sizes was generated from the reference genome, as well as an index file with SAMtools version 1.7 (Li et al. 2009). These allowed for efficient random access to the reference bases during down-stream analysis and mapping.

4. Sequence Alignment

For each sample, sequence reads were aligned to the *N. vitripennis* assembly using HISAT2 version 2.1.0 (Kim et al. 2015). The output SAM (Sequence Alignment/Map) files were converted into BAM (Binary Alignment/Map) files using SAMtools version 1.7 (Li et al. 2009). In order to prepare the BAM files for variant calling, we followed the GATK (Genome Analysis Tool Kit) preprocessing steps (DePristo et al. 2011). Duplicates were marked and reads in each sample were assigned to a single new read-group with MarkDuplicates and AddOrReplaceReadGroups, respectively, using Picard tools version 1.119 (available at: http://broadinstitute.github.io/picard/). BAM files were then sorted and indexed using BamTools version 2.5.1 (Barnett et al. 2011) for down-stream analysis.

5. Variant Calling

Aligned RNAseq files were processed to remove duplicates using Picard tools version 1.119 to avoid propagation of sequencing error during variant calling (DePristo et al. 2011). As recommended by the GATK Best Practices, variants were called with GATK HaplotypeCaller version 3.8 (McKenna et al. 2010) and VCF (Variant Call Format) files were merged and joint genotyped with combineGVCFs and GenotypeGVCFs utilities in GATK.

6. Pseudo *N. giraulti* Reference Genome Assembly

Comparing homozygous *N. giraulti* and homozygous *N. vitripennis* individuals, we identified sites between them that were homozygous for a different allele. We defined these sites as fixed differences. The fixed and different sites were then used to create a pseudo *N. giraulti* reference sequence with the FastaAlternateReferenceMaker function in

GATK version 3.8 (available at: http://www.broadinstitute.org/gatk/). Reference bases in the *N. vitripennis* genome were replaced with the alternate base at variant positions using this tool. Following a similar protocol for comparison, we aligned raw RNAseq reads to the pseudo *N. giraulti* genome reference using HISAT2 version 2.1.0, and performed identical preprocessing steps prior to variant calling with GATK version 3.8 HaplotypeCaller.

7. Testing for Reference Bias

In order to test for reference bias, the proportion of *N. giraulti* alleles (from the *N. giraulti* parent) in the hybrids was compared when aligned to a *N. vitripennis* reference and when aligned to our pseudo *N. giraulti* reference. Hybrid replicates (3 of $V_{paternal}G_{maternal}$ and 3 of $G_{paternal}V_{maternal}$) were merged to give a single hybrid representative of each genotype. The same 1,792 shared genes (filtered for fixed differences and a read depth of 100) in each mapping protocol group (i.e., aligned to *N. vitripennis* or aligned to *N. giraulti*) were used when comparing reference genomes. To determine whether the data were normally distributed, a Shapiro-Wilk test was done. The *p*-value given by this test must be above 0.05 to be considered normally distributed and suitable for a paired t-test (Razali & Wah, 2011). In the case of data that are not normally distributed, a nonparametric alternative for a paired t-test is the Wilcoxon signed-rank test (or the paired samples Wilcoxon test) (Gehan, 1965).

8. Allele-specific Expression Analysis

Allele-specific expression (ASE) analysis was considered for reads aligned to the *N. vitripennis* reference, and for reads aligned to the pseudo *N. giraulti* reference. To identify allele bias, only sites that were heterozygous in the hybrids, but homozygous in

the pure *N. vitripennis* and pure *N. giraulti* individuals were selected for analysis. Allele counts of reads were obtained using GATK ASEReadCounter version 3.8 (Castel et al. 2015) and a parameter of counts greater than 10 was applied to ensure adequate coverage (Skelly et al. 2011). For the analysis of multiple single nucleotide polymorphisms (SNPs) within a gene, a gene was labeled "TRUE" if the SNPs showed bias towards the same allele. A gene was labeled "FALSE" if the SNPs did not show bias towards the same allele. Labeling and quantifying gene direction (TRUE or FALSE) was repeated several times at various read depths (RD = 10, 20, 30, 50, 100) to eliminate problematic SNPs (Wang & Clark, 2014).

9. Preliminary Differential Expression Analysis

Aligned RNAseq reads were given as input to Subread version 1.6.2 (Liao et al. 2014) using the featureCounts function to generate counts of reads uniquely mapped to annotated genes in the *N. vitripennis* Generic Feature Format (GFF) file. Using edgeR version 3.22.5 (Robinson et al. 2010), raw counts were transformed to counts per million (CPM). A CPM value of 1 was used in our analysis to separate expressed genes from unexpressed genes. This means, for example, that for a library size of ~10 million reads, there are at least 10 counts per gene in that sample. Pairwise contrasts between hybrids and parents were then generated using the limma makecontrasts function (Law et al. 2014). We identified genes that exhibited significant expression differences with an adjusted *p*-value cutoff of less than 0.05 (5%) to account for multiple testing in pairwise comparisons (Storey, 2003).

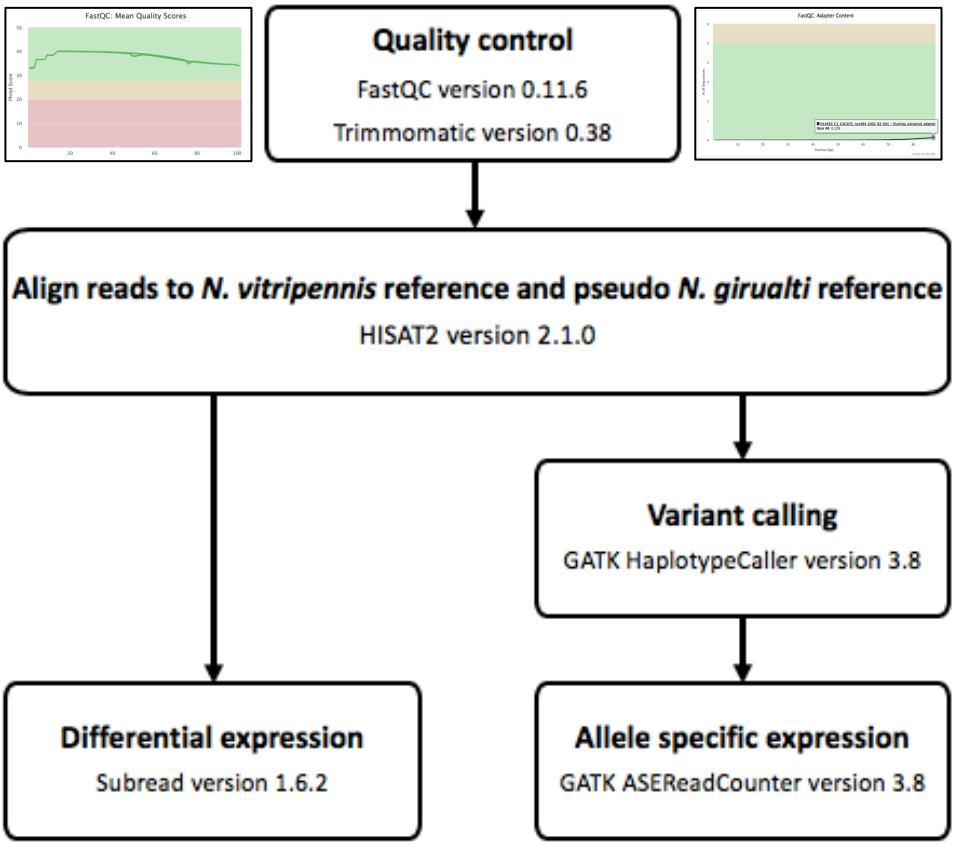


Figure 3. Methods Workflow Summary.

RESULTS

1. Search for Reference Bias

After read alignment to the *N. vitripennis* reference and our pseudo *N. giraulti* reference, number of mapped reads per sample showed some variability between the two references on a sample-wide basis, Table 1. The hybrids had more reads align to the pseudo *N. giraulti* compared to the *N. vitripennis* reference. The average depth of coverage using the *N. vitripennis* reference was ~16x, whereas the average depth using our pseudo *N. giraulti* reference was ~15x.

Table 1. Reads Mapped to *N. vitripennis* and Pseudo *N. giraulti* References. Presented below are number of reads mapped when aligned to the *N. vitripennis* reference, and to our pseudo *N. giraulti* reference. In the difference column (Nv–Ng), a positive value indicates more reads mapped to the *N. vitripennis* reference, and a negative value indicates more reads mapped to the pseudo *N. giraulti* reference.

Sample ; Genotype	**Aligned to *N. vitripennis* (Nv) reference**	**Aligned to pseudo *N. giraulti* (Ng) reference**	**Nv – Ng**
014444 ; VV	27,943,376	26,969,855	+973,521
014445 ; VV	27,512,845	26,590,076	+922,769
014446 ; VV	21,516,143	20,780,758	+735,385
014447 ; GG	24,208,360	26,388,459	−2,180,099
014448 ; GG	19,997,753	21,856,369	−1,858,616
014449 ; GG	27,691,029	30,303,087	−2,612,058
014450 ; GV	42,043,452	42,644,497	−601,045
014451 ; GV	27,678,243	28,067,754	−389,511
014452 ; GV	31,890,582	32,358,349	−467,767
014453 ; VG	19,896,457	20,479,554	−583,097
014454 ; VG	16,727,114	17,175,905	−448,791
014455 ; VG	29,394,376	30,140,949	−746,573

After identifying only genes with fixed differences for analysis, we performed a sanity check to confirm that our pseudo reference was aligning in a non-random manner. For a subset of genes with fixed differences, we plotted the proportion of *N. vitripennis* allele (from the *N. vitripennis* parent) and the proportion of *N. giraulti* allele (from the *N. giraulti* parent) in each hybrid. All hybrid comparisons showed directionality agreement between the two references. Figure 4 depicts an example of this agreement in one hybrid sample, 014450 ($G_{paternal}V_{maternal}$), for a subset of genes.

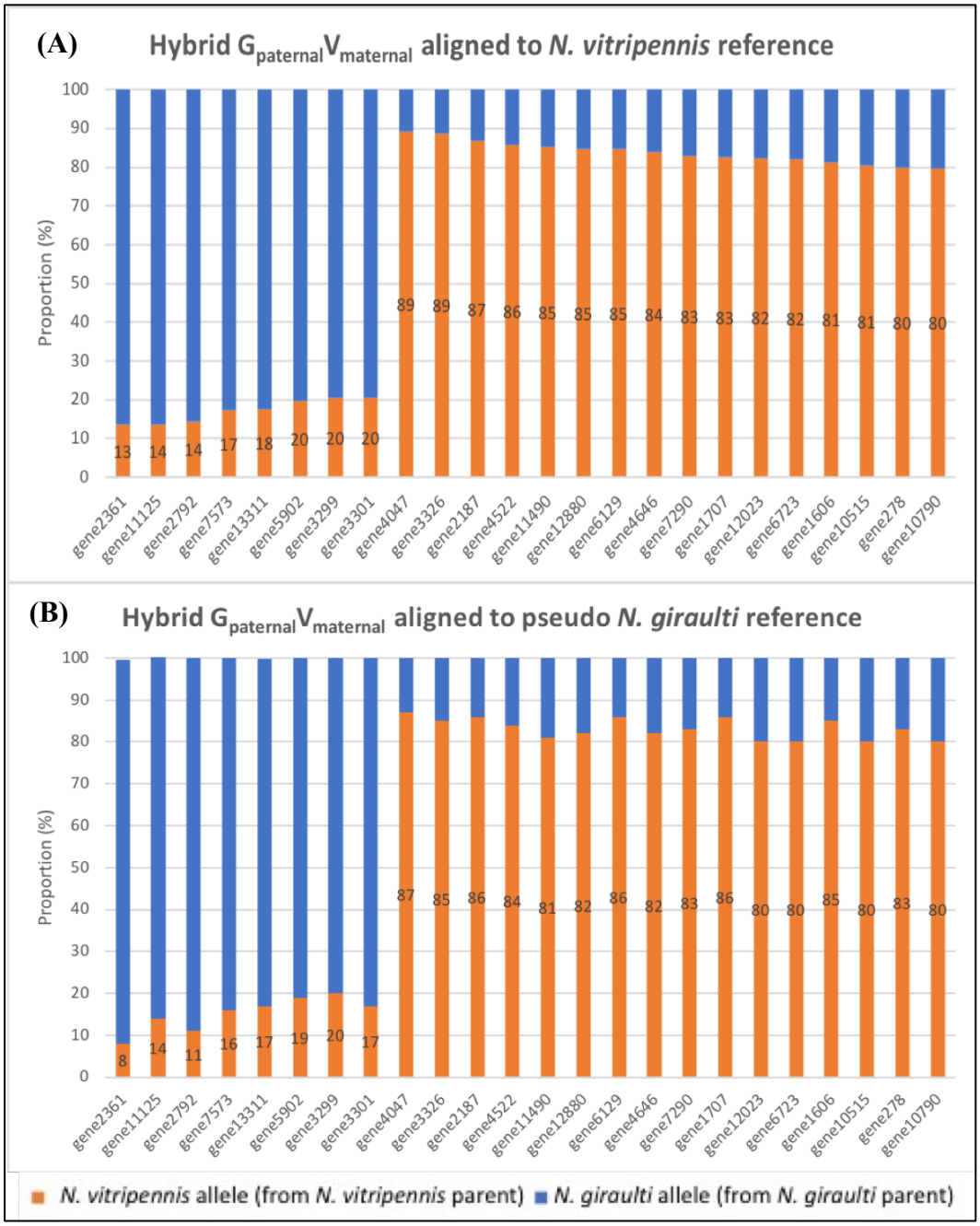


Figure 4. Pseudo *N. giraulti* Reference Sanity Check. In hybrid sample 014450 ($G_{paternal}V_{maternal}$), double bar graphs showing the proportion of *N. vitripennis* allele (from *N. vitripennis* parent) (in orange) and *N. giraulti* allele (from *N. giraulti* parent) (in blue). **(A)** is alignment to the *N. vitripennis* reference, and **(B)** is alignment to our pseudo *N. giraulti* reference.

In order to identify reference bias, we directly compared the proportion of *N. giraulti* allele in the hybrids when aligned to the *N. vitripennis* reference, to the proportion of *N. giraulti* allele in the hybrids when aligned to our pseudo *N. giraulti* reference. We contrasted the proportion of *N. giraulti* allele in hybrid samples for genes that were previously identified as being fixed and different between homozygous *N. vitripennis* and *N. giraulti* individuals. We selected only for fixed and different genes with a read depth of 100 that were shared among the hybrids, Figure 5. A Shapiro test was performed on the data to test for a normal distribution. The data were not normally distributed (p-value < 2.2e-16). The data was then $\log_{10}$ transformed and tested again for normal distribution, and after $\log_{10}$ transformation the data were still not evenly distributed. Thus, to test for differences in proportion of *N. giraulti* allele when aligned to *N. vitripennis* versus aligned to pseudo *N. giraulti* we used a nonparametic test, the Wilcoxon paired, and confirmed a significant difference between mapping protocol (p-value < 2.2e-16).

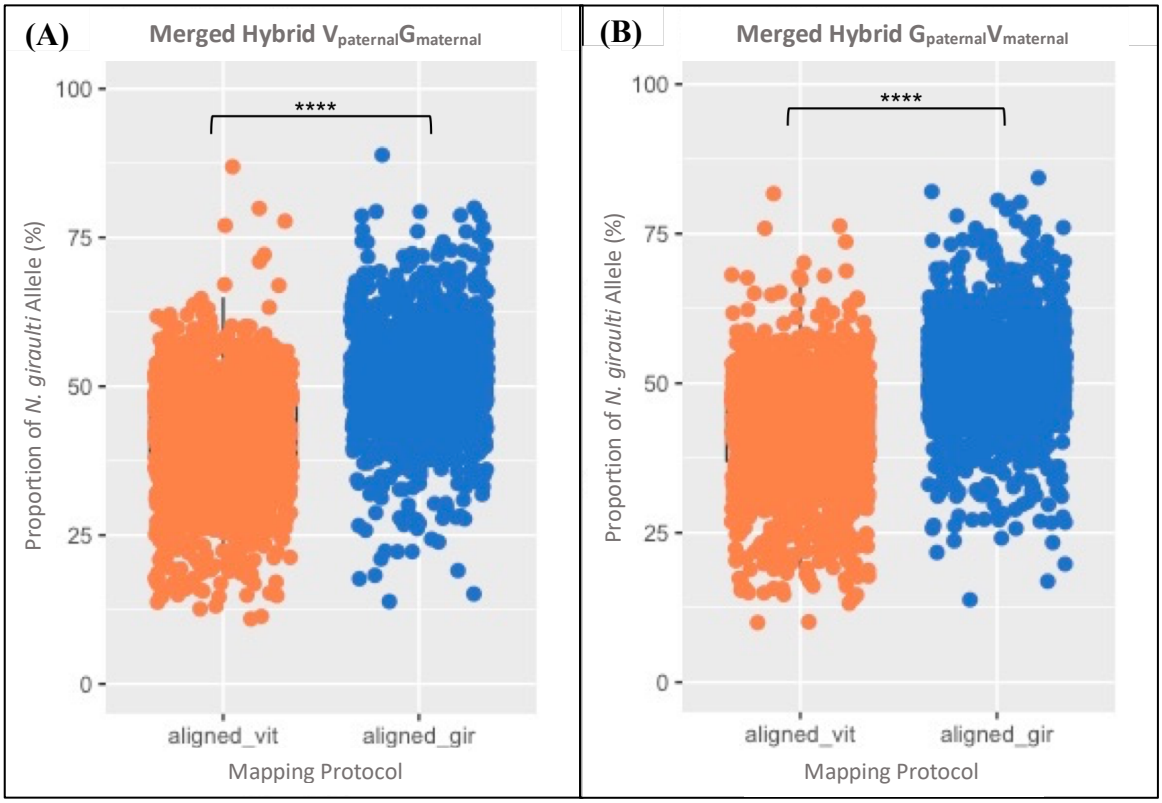


Figure 5. Reference Bias Jitter Plots. (A) is proportion of *N. giraulti* allele in merged hybrid $V_{paternal}G_{maternal}$ when aligned to the *N. vitripennis* reference (in orange) compared to our pseudo *N. giraulti* reference (in blue). There is significant difference in proportion of *N. giraulti* allele between the two reference genomes used for aligning reads; *p*-value ≤ 0.0001. **(B)** is proportion of *N. giraulti* allele in merged hybrid $G_{paternal}V_{maternal}$ when aligned to the *N. vitripennis* reference (in orange) compared to our pseudo *N. giraulti* reference (in blue).

2. Allele-specific Expression Analysis

Allelic expression (allele-specific expression or allelic imbalance) quantifies expression variation between two haplotypes of a diploid (2n) individual by the heterozygous sites (Castel et al., 2015). Thus, we began by identifying sites that were heterozygous in the hybrids, but homozygous in the pure *N. vitripennis* and pure *N. giraulti* samples for analysis—a total of 283,324 sites. Allele counts were obtained for reads aligned to the *N. vitripennis* reference, as well as for reads aligned to the pseudo *N. giraulti* reference. By directly counting the number of reference and alternative allele-containing reads at polymorphic SNP positions, we quantified the number of SNPs in each hybrid replicate that 1) showed bias towards the allele that came from the *N. vitripennis* parent, 2) showed bias towards the allele that came from the *N. giraulti* parent, and 3) showed no difference (ND) in expression of its parental alleles (Appendix A). Since our pseudo *N. giraulti* reference was created from fixed and different sites (i.e., sites between homozygous *N. giraulti* and homozygous *N. vitripennis* individuals that are homozygous for a different allele), we could definitely say that the reference allele at a polymorphic SNP position in a hybrid aligned to the *N. vitripennis* reference was inherited from the *N. vitripennis* parent, and the alternative allele was inherited from the *N. giraulti* parent. Similarly, for a hybrid aligned to our pseudo *N. giraulti* reference genome, the reference allele at a polymorphic SNP position was inherited from the *N. giraulti* parent, and the alternative allele from the *N. vitripennis* parent.

As such, multiple SNPs per gene existed for the 9,119 genes identified. For the analysis of multiple SNPs within a gene, a gene was labeled “TRUE” if all SNPs within a gene showed the same bias towards either the *N. vitripennis* or *N. giraulti* allele. A gene

was labeled “FALSE” if the SNPs did not show bias in the same direction of either the *N. vitripennis* or *N. giraulti* allele. Labeling and quantifying gene direction (TRUE or FALSE) was repeated several times at various read depths (RD = 10, 20, 30, 50, 100) to eliminate problematic or inconsistent SNPs (Wang & Clark, 2014) (Appendix B). For down-stream analysis, we decided to only focus on TRUE genes—genes with SNPs showing bias for either the *N. vitripennis* or *N. giraulti* allele. As well, we decided to use the most stringent filter of RD = 100. Even with these parameters, there were still on average more FALSE genes than TRUE genes, Figure 6.

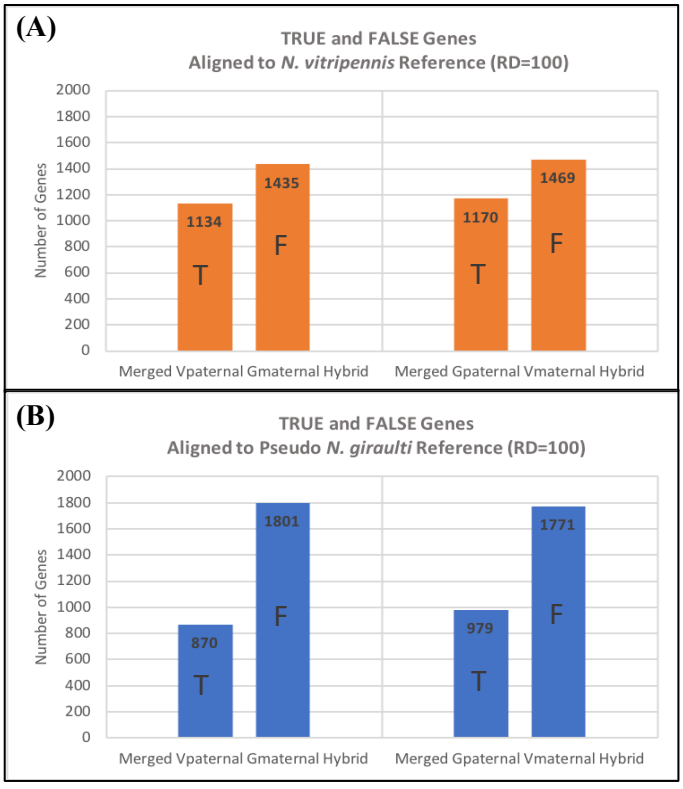


Figure 6. Bar Graphs of TRUE (T) and FALSE (F) Genes. At read depth 100, **(A)** shows the average number of T/F genes when aligned to the *N. vitripennis* reference compared to **(B)** the average number of T/F genes when aligned to our pseudo *N. giraulti* reference.

To evaluate allelic imbalance, we compared the proportion of *N. vitripennis* allele in the $G_{paternal}V_{maternal}$ and $V_{paternal}G_{maternal}$ hybrids when aligned to the *N. vitripennis* reference and when aligned to our pseudo *N. giraulti* reference. For this analysis, informative SNPs included the same 60 genes with SNP direction agreement (i.e., TRUE genes) at a read depth of 100. The R-squared regression value for the hybrid comparison aligned to the *N. vitripennis* reference was $R^2 = 0.9474$, and the regression value for the hybrid comparison aligned to our pseudo *N. giraulti* reference was $R^2 = 0.9529$, Figure 7.

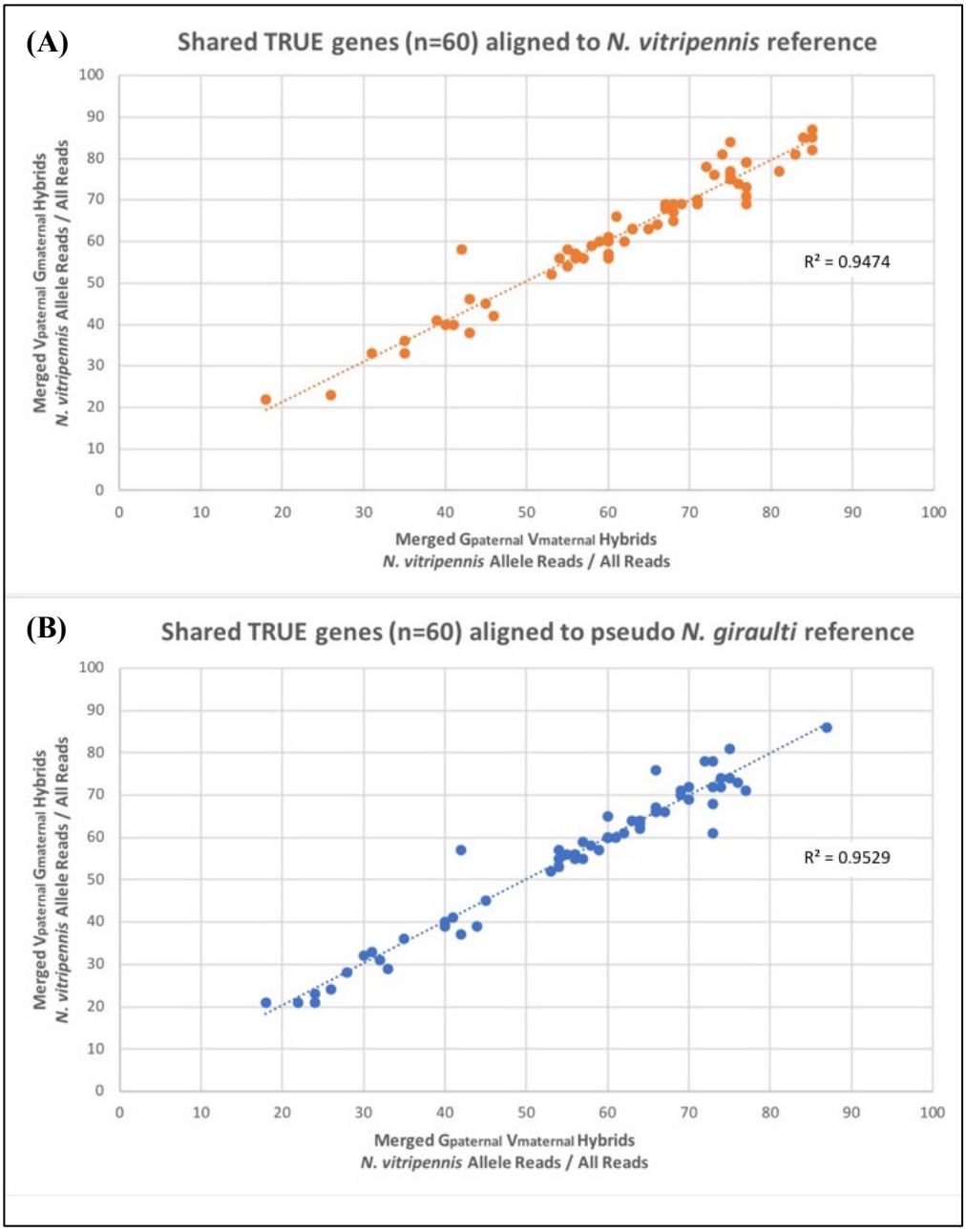


Figure 7. Species-of-origin Results. For 60 shared genes at a read depth of 100, scatterplot of allelic expression for proportion (%) of *N. vitripennis* allele between merged $G_{paternal}V_{maternal}$ (x-axis) and $V_{paternal}G_{maternal}$ (y-axis) hybrids. **(A)** is alignment to the *N. vitripennis* reference ($R^2 = 0.9474$), and **(B)** is alignment to our pseudo *N. giraulti* reference ($R^2 = 0.9529$). Informative SNPs include only genes with SNP direction agreement (i.e., TRUE genes).

DISCUSSION

1. Reference Bias

With our pseudo *N. giraulti* reference genome, we identified a difference in total mapped reads and a difference in allele counts compared to the *N. vitripennis* reference genome. The hybrids overall mapped more to our pseudo *N. giraulti* reference, Figure 5. Although we initially assumed that the difference between the two was insignificant, statistical testing showed otherwise (*p*-value < 2.2e-16). Given the sister species *N. vitripennis* and *N. giraulti* are ~1 million years diverged, and the synonymous coding divergence is ~3% (Werren et al., 2010), it is not unlikely that there is a difference.

2. Allele-specific Expression

The allele-specific expression levels from the reciprocal hybrids can be used to assess the impact of allelic variation and parent-of-origin effects (maternal versus paternal). If the two alleles are equivalently expressed, we would expect to observe a 1:1 ratio. In cases where one allele (*N. vitripennis* or *N. giraulti*) is preferentially expressed, it is expected that expression would be biased toward the same allele in both reciprocal hybrids regardless of the parental origin of the allele (referred to here as a species-of-origin effects). In contrast, examples of parent-of-origin effects would be expected to exhibit a bias toward one allele in one hybrid and bias toward the other allele in the reciprocal hybrid as the parents are reversed in the hybrid samples (i.e., $G_{paternal}V_{maternal}$ and $V_{paternal}G_{maternal}$), Figure 8.

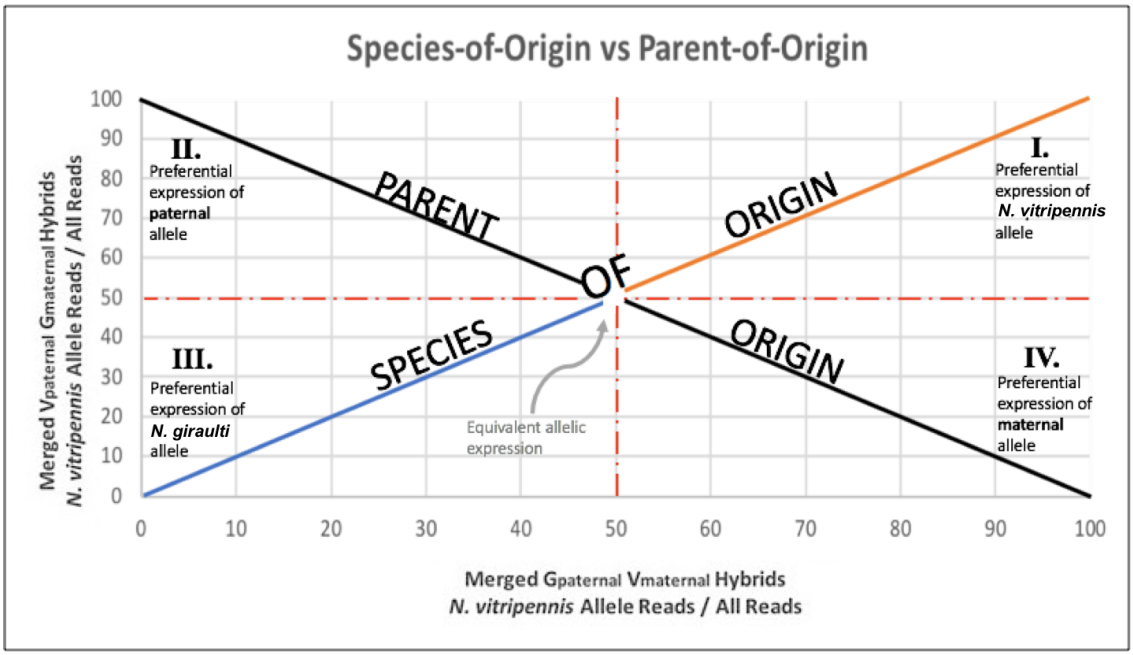


Figure 7. Species-of-origin Versus Parent-of-origin Expectations.

Our data in Figure 7 showed that most genes across the *Nasonia* genome are distributed along the species-of-origin curve seen above. Interestingly, the interpretation that we do not observe parent-of-origin expression patterns in adult *Nasonia* does not change depending on which aligner is used (i.e., $R^2 = 0.9474$ for alignment to the *N. vitripennis* reference and $R^2 = 0.9529$ for alignment to our pseudo *N. giraulti* reference).

UNEXPECTED SURPRISES

We had initially assumed that our homozygous *N. giraulti* and homozygous *N. vitripennis* individuals were sufficiently inbred. However, down-stream analysis raised red flags that eventually led us to double checking this assumption. As a result, we identified ~0.3% of the genome in the inbred lines showed heterozygous sites. It is important to note this as an example of how evolution occurs and should always be considered. Another unexpected challenge was the process of filtering out hundreds of problematic or inconsistent SNPs in order to analyze more genes with SNPs in the same direction. We hypothesize that the high number of mismatching SNPs observed could be from technical variation inherent in RNAseq data, such as artifacts or over dispersion of reads. To correct for technical variation, we filtered by read depths of various levels. However, it would also be beneficial to select genes from the FALSE category (i.e., genes whose SNPs are not consistent in direction) using a filter of 70:30 and 80:20 allele ratios. These ratios would represent high enough proportions of the reference *N. vitripennis* allele that we would consider them preferentially expressed.

FUTURE DIRECTIONS

In this project we searched for parent-of-origin effects in adult *Nasonia* across the genome on a gene-wide basis. Our concluding results found no support for genomic imprinting as no parent-of-origin patterns were identified. However, one limitation in our study was that without exome data we were unable to detect allelic imbalance in the heterozygous sites that showed equal bias towards both the *N. vitripennis* allele and the *N. giraulti* allele. Although our findings are in line with the previous Clark paper, it is still possible that we missed some genes that are actually imprinted. If we were to

observe imprinted genes in *Nasonia*, one theory is that it may have to do with the parasitic nature of the wasp as it competes for hosts.

As described in the methods section, some preliminary differential expression has been visualized with volcano plots thus far (Appendix C). However, patterns in up and down regulated genes have not yet been considered. As such, the next step is to investigate whether our species-of-origin genes are the same 178 differentially expressed genes identified in the previous Clark paper (Wang et al., 2016).

CONCLUSIONS

There are observed differences when aligning RNAseq reads to a *N. vitripennis* reference genome compared to aligning reads to our pseudo *N. giraulti* reference. However, the interpretation that we do not observe parent-of-origin expression patterns in adult *Nasonia* does not change based on which reference genome is used to align reads. Thus, our preliminary findings so far support the previous Clark paper (Wang et al., 2016).

REFERENCES

Introduction: Chapter 1

Barlow, D. P., & Bartolomei, M. S. (2014). Genomic Imprinting in Mammals. *Cold Spring Harbor Perspectives in Biology*, *6*(2), a018382. https://doi.org/10.1101/cshperspect.a018382

Butler, M. G. (2011). Prader-Willi Syndrome: Obesity due to Genomic Imprinting. *Current Genomics*, *12*(3), 204–215. https://doi.org/10.2174/138920211795677877

Clayton-Smith, J., & Pembrey, M. E. (1992). Angelman syndrome. *Journal of Medical Genetics*, *29*(6), 412–415.

Crouse, H. V. (1960). The Controlling Element in Sex Chromosome Behavior in Sciara. *Genetics*, *45*(10), 1429–1443.

Dindot, S. V., Person, R., Strivens, M., Garcia, R., & Beaudet, A. L. (2009). Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions. *Genome Research*, *19*(8), 1374–1383. https://doi.org/10.1101/gr.089185.108

Glenn, C. C., Driscoll, D. J., Yang, T. P., & Nicholls, R. D. (1997). Genomic imprinting: Potential function and mechanisms revealed by the Prader-Willi and Angelman syndromes. *Molecular Human Reproduction*, *3*(4), 321–332.

Hutter, B., Helms, V., & Paulsen, M. (2006). Tandem repeats in the CpG islands of imprinted genes. *Genomics*, *88*(3), 323–332. https://doi.org/10.1016/j.ygeno.2006.03.019

Ishida, M., & Moore, G. E. (2013). The role of imprinted genes in humans. *Molecular Aspects of Medicine*, *34*(4), 826–840. https://doi.org/10.1016/j.mam.2012.06.009

Kishino, T., Lalande, M., & Wagstaff, J. (1997). UBE3A/E6-AP mutations cause Angelman syndrome. *Nature Genetics*, *15*(1), 70–73. https://doi.org/10.1038/ng0197-70

Knoll, J. H., Nicholls, R. D., Magenis, R. E., Graham, J. M., Lalande, M., & Latt, S. A. (1989). Angelman and Prader-Willi syndromes share a common chromosome 15 deletion but differ in parental origin of the deletion. *American Journal of Medical Genetics*, *32*(2), 285–290. https://doi.org/10.1002/ajmg.1320320235

Köhler, C., Wolff, P., & Spillane, C. (2012). Epigenetic Mechanisms Underlying Genomic Imprinting in Plants. *Annual Review of Plant Biology*, *63*(1), 331–352. https://doi.org/10.1146/annurev-arplant-042811-105514

Morison, I. M., Ramsay, J. P., & Spencer, H. G. (2005). A census of mammalian imprinting. *Trends in Genetics*, *21*(8), 457–465. https://doi.org/10.1016/j.tig.2005.06.008

Neumann, B., Kubicka, P., & Barlow, D. P. (1995). Characteristics of imprinted genes. *Nature Genetics*, *9*(1), 12. https://doi.org/10.1038/ng0195-12

Nicholls, R. D., Saitoh, S., & Horsthemke, B. (1998). Imprinting in Prader–Willi and Angelman syndromes. *Trends in Genetics*, *14*(5), 194–200. https://doi.org/10.1016/S0168-9525(98)01432-2

Reik, W., & Walter, J. (2001). Genomic imprinting: parental influence on the genome. *Nature Reviews Genetics*, *2*(1), 21–32. https://doi.org/10.1038/35047554

Sarkar, P. A., Shigli, A., & Patidar, C. (2011). Happy Puppet syndrome. *BMJ Case Reports*, *2011*. https://doi.org/10.1136/bcr.09.2011.4747

Surani, M. A. H., Barton, S. C., & Norris, M. L. (n.d.). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis | Nature. Retrieved February 18, 2019, from https://www.nature.com/articles/308548a0

Wood, A. J., & Oakey, R. J. (2006). Genomic Imprinting in Mammals: Emerging Themes and Established Theories. *PLOS Genetics*, *2*(11), e147. https://doi.org/10.1371/journal.pgen.0020147

Introduction: Chapter 2

Barlow, D. P., Stöger, R., Herrmann, B. G., Saito, K., & Schweifer, N. (1991). The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature*, *349*(6304), 84. https://doi.org/10.1038/349084a0

Bartolomei, M. S., & Tilghman, S. M. (1997). Genomic Imprinting in Mammals. *Annual Review of Genetics*, *31*(1), 493–525. https://doi.org/10.1146/annurev.genet.31.1.493

DeChiara, T. M., Robertson, E. J., & Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. *Cell*, *64*(4), 849–859.

Ferguson-Smith, A. C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. *Nature Reviews Genetics*, *12*(8), 565–575. https://doi.org/10.1038/nrg3032

Haig, D. (2000). The Kinship Theory of Genomic Imprinting. *Annual Review of Ecology and Systematics*, *31*(1), 9–32. https://doi.org/10.1146/annurev.ecolsys.31.1.9

Haig, D. (2004). GENOMIC IMPRINTING AND KINSHIP: How Good is the Evidence? *Annual Review of Genetics*, *38*, 553–585. https://doi.org/10.1146/annurev.genet.37.110801.142741

Kermicle, J. L., & Alleman, M. (1990). Gametic imprinting in maize in relation to the angiosperm life cycle. *Development (Cambridge, England). Supplement*, 9–14.

Kinoshita, T., Yadegari, R., Harada, J. J., Goldberg, R. B., & Fischer, R. L. (1999). Imprinting of the MEDEA Polycomb Gene in the Arabidopsis Endosperm. *The Plant Cell*, *11*(10), 1945–1952. https://doi.org/10.1105/tpc.11.10.1945

Kondrashov, A. S., & Crow, J. F. (1991). Haploidy or diploidy: which is better? *Nature*, *351*(6324), 314. https://doi.org/10.1038/351314a0

Moore, T., & Haig, D. (1991). Genomic imprinting in mammalian development: a parental tug-of-war. *Trends in Genetics*, *7*(2), 45–49. https://doi.org/10.1016/0168-9525(91)90230-N

Renfree, M. B., Hore, T. A., Shaw, G., Marshall Graves, J. A., & Pask, A. J. (2009). Evolution of Genomic Imprinting: Insights from Marsupials and Monotremes. *Annual Review of Genomics and Human Genetics*, *10*(1), 241–262. https://doi.org/10.1146/annurev-genom-082908-150026

Introduction: Chapter 3

Amarasinghe, H. E., Toghill, B. J., Nathanael, D., & Mallon, E. B. (2015). Allele specific expression in worker reproduction genes in the bumblebee Bombus terrestris. *PeerJ*, *3*, e1079. https://doi.org/10.7717/peerj.1079

Ayasse, M., Paxton, R. J., & Tengö, J. (2001). Mating Behavior and Chemical Communication in the Order Hymenoptera. *Annual Review of Entomology*, *46*(1), 31–78. https://doi.org/10.1146/annurev.ento.46.1.31

Galbraith, D. A., Kocher, S. D., Glenn, T., Albert, I., Hunt, G. J., Strassmann, J. E., … Grozinger, C. M. (2016). Testing the kinship theory of intragenomic conflict in honey bees (Apis mellifera). *Proceedings of the National Academy of Sciences*, *113*(4), 1020–1025. https://doi.org/10.1073/pnas.1516636113

Haig, D. (1992). Intragenomic conflict and the evolution of eusociality. *Journal of Theoretical Biology*, *156*(3), 401–403. https://doi.org/10.1016/S0022-5193(05)80683-6

Kronauer, D. J. C. (2008). Genomic imprinting and kinship in the social Hymenoptera: What are the predictions? *Journal of Theoretical Biology*, *254*(4), 737–740. https://doi.org/10.1016/j.jtbi.2008.06.019

Kronforst, M. R., Gilley, D. C., Strassmann, J. E., & Queller, D. C. (2008). DNA methylation is widespread across social Hymenoptera. *Current Biology*, *18*(7), R287–R288. https://doi.org/10.1016/j.cub.2008.02.015

Queller, D. C. (2003). Theory of genomic imprinting conflict in social insects. *BMC Evolutionary Biology*, *3*, 15. https://doi.org/10.1186/1471-2148-3-15

Ratnieks, F. L. W., Foster, K. R., & Wenseleers, T. (2006). Conflict resolution in insect societies. *Annual Review of Entomology*, *51*, 581–608. https://doi.org/10.1146/annurev.ento.51.110104.151003

Remnant, E. J., Ashe, A., Young, P. E., Buchmann, G., Beekman, M., Allsopp, M. H., … Oldroyd, B. P. (2016). Parent-of-origin effects on genome-wide DNA methylation in the Cape honey bee (Apis mellifera capensis) may be confounded by allele-specific methylation. *BMC Genomics*, *17*(1), 226. https://doi.org/10.1186/s12864-016-2506-8

Shigeto, D., & Tsuji Kazuki. (2012). Intragenomic conflict over queen determination favours genomic imprinting in eusocial Hymenoptera. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1738), 2553–2560. https://doi.org/10.1098/rspb.2011.2673

Wild, G., West, S. A., Weissing, A. E. F. J., & Whitlock, E. M. C. (2009). Genomic Imprinting and Sex Allocation. *The American Naturalist*, *173*(1), E1–E14. https://doi.org/10.1086/593305

Introduction: Chapter 4

Beukeboom, L., & Desplan, C. (2003). Nasonia. *Current Biology: CB*, *13*(22), R860.

Beukeboom, L. W., & van den Assem, J. (2001). Courtship and Mating Behaviour of Interspecific Nasonia Hybrids (Hymenoptera, Pteromalidae): A Grandfather Effect. *Behavior Genetics*, *31*(2), 167–177. https://doi.org/10.1023/A:1010201427204

Beukeboom, Leo W., & van de Zande, L. (2010). Genetics of sex determination in the haplodiploid wasp Nasonia vitripennis (Hymenoptera: Chalcidoidea). *Journal of Genetics*, *89*(3), 333–339.

Bordenstein, S. R., O'Hara, F. P., & Werren, J. H. (2001). *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. *Nature*, *409*(6821), 707–710. https://doi.org/10.1038/35055543

Breed, M. D., & Moore, J. (2016). Chapter 13 - Social Behavior, Cooperation, and Kinship. In M. D. Breed & J. Moore (Eds.), *Animal Behavior (Second Edition)* (pp. 423–458). San Diego: Academic Press. https://doi.org/10.1016/B978-0-12-801532-2.00013-1

Campbell, B. C., Steffen-Campbell, J. D., & Werren, J. H. (1994). Phylogeny of the Nasonia species complex (Hymenoptera: Pteromalidae) inferred from an internal transcribed spacer (ITS2) and 28S rDNA sequences. *Insect Molecular Biology*, *2*(4), 225–237. https://doi.org/10.1111/j.1365-2583.1994.tb00142.x

Darling, D. C., & Werren, J. H. (1990). Biosystematics of Nasonia (Hymenoptera: Pteromalidae): Two New Species Reared from Birds' Nests in North America. *Annals of the Entomological Society of America*, *83*(3), 352–370. https://doi.org/10.1093/aesa/83.3.352

Desjardins, C. A., Perfectti, F., Bartos, J. D., Enders, L. S., & Werren, J. H. (2010). The genetic basis of interspecies host preference differences in the model parasitoid *Nasonia*. *Heredity*, *104*(3), 270–277. https://doi.org/10.1038/hdy.2009.145

Ellison, C. K., Niehuis, O., & Gadau, J. (2008). Hybrid breakdown and mitochondrial dysfunction in hybrids of Nasonia parasitoid wasps. *Journal of Evolutionary Biology*, *21*(6), 1844–1851. https://doi.org/10.1111/j.1420-9101.2008.01608.x

Gadau, J., Niehuis, O., Peire, A., Werren, J. H., Baudry, E., & Beukeboom, L. W. (2008). The Jewel Wasp – Nasonia. In W. Hunter & C. Kole (Eds.), *Genome Mapping and Genomics in Arthropods* (pp. 27–41). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-73833-6_3

Lynch, J. A., Brent, A. E., Leaf, D. S., Anne Pultz, M., & Desplan, C. (2006). Localized maternal *orthodenticle* patterns anterior and posterior in the long germ wasp *Nasonia*. *Nature*, *439*(7077), 728–732. https://doi.org/10.1038/nature04445

Niehuis, O., Buellesbach, J., Gibson, J. D., Pothmann, D., Hanner, C., Mutti, N. S., … Schmitt, T. (2013). Behavioural and genetic analyses of *Nasonia* shed light on the evolution of sex pheromones. *Nature*, *494*(7437), 345–348. https://doi.org/10.1038/nature11838

Pultz, M. A., & Leaf, D. S. (2003). The jewel wasp Nasonia: querying the genome with haplo-diploid genetics. *Genesis (New York, N.Y.: 2000)*, *35*(3), 185–191. https://doi.org/10.1002/gene.10189

Raychoudhury, R., Desjardins, C. A., Buellesbach, J., Loehlin, D. W., Grillenberger, B. K., Beukeboom, L., … Werren, J. H. (2010). Behavioral and genetic characteristics of a new species of *Nasonia*. *Heredity*, *104*(3), 278–288. https://doi.org/10.1038/hdy.2009.147

Shuker, D., Lynch, J., & Morais, A. P. (2003). Moving from model to non-model organisms? Lessons from Nasonia wasps. *BioEssays*, *25*(12), 1247–1248. https://doi.org/10.1002/bies.10367

Shuker, D. M., & West, S. A. (2004). Information constraints and the precision of adaptation: Sex ratio manipulation in wasps. *Proceedings of the National Academy of Sciences*, *101*(28), 10363–10367. https://doi.org/10.1073/pnas.0308034101

Verhulst, E. C., Beukeboom, L. W., & Zande, L. van de. (2010). Maternal Control of Haplodiploid Sex Determination in the Wasp Nasonia. *Science*, *328*(5978), 620–623. https://doi.org/10.1126/science.1185805

Werren, J. H., & Loehlin, D. W. (2009). The Parasitoid Wasp Nasonia: An Emerging Model System with Haploid Male Genetics. *Cold Spring Harbor Protocols*, *2009*(10), pdb.emo134. https://doi.org/10.1101/pdb.emo134

Werren, J. H., Richards, S., Desjardins, C. A., Niehuis, O., Gadau, J., Colbourne, J. K., … Gibbs, R. A. (2010). Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. *Science (New York, N.Y.)*, *327*(5963), 343–348. https://doi.org/10.1126/science.1178028

Whiting, A. R. (1967). The Biology of the Parasitic Wasp Mormoniella vitripennis [=Nasonia brevicornis] (Walker). *The Quarterly Review of Biology*, *42*(3), 333–406.

Introduction: Chapter 5

Wang, X., Werren, J. H., & Clark, A. G. (2016). Allele-Specific Transcriptome and Methylome Analysis Reveals Stable Inheritance and Cis-Regulation of DNA Methylation in Nasonia. *PLOS Biology*, *14*(7), e1002500. https://doi.org/10.1371/journal.pbio.1002500

Methods

FastQC. A quality control tool for high throughput sequence data. Babraham Bioinformatics Web site. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, *52*(1–2), 203–224. https://doi.org/10.1093/biomet/52.1-2.203

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., & Marth, G. T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics (Oxford, England)*, *27*(12), 1691–1692. https://doi.org/10.1093/bioinformatics/btr174

Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E., & Lappalainen, T. (2015). Tools and best practices for data processing in allelic expression analysis. *Genome Biology*, *16*(1). https://doi.org/10.1186/s13059-015-0762-6

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics (Oxford, England)*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/nmeth.3317

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, *15*(2), R29. https://doi.org/10.1186/gb-2014-15-2-r29

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Institute, B. (2016). Picard is a set of command line tools for manipulating high-throughput sequencing (HTS) data and formats such as SAM/BAM/CRAM and VCF. https://broadinstitute.github.io/picard/

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk , Kolmogorov-Smirnov , Lilliefors and Anderson-Darling tests.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Skelly, D. A., Johansson, M., Madeoy, J., Wakefield, J., & Akey, J. M. (2011). A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Research*, *21*(10), 1728–1737. https://doi.org/10.1101/gr.119784.110

Storey, J. D. (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. *The Annals of Statistics*, *31*(6), 2013–2035. https://doi.org/10.1214/aos/1074290335

Wang, X., & Clark, A. G. (2014). Using next-generation RNA sequencing to identify imprinted genes. *Heredity*, *113*(2), 156–166. https://doi.org/10.1038/hdy.2014.18

Werren, J. H., Richards, S., Desjardins, C. A., Niehuis, O., Gadau, J., Colbourne, J. K., … Gibbs, R. A. (2010). Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. *Science (New York, N.Y.)*, *327*(5963), 343–348. https://doi.org/10.1126/science.1178028

Results

Castel, S. E., Levy-Moonshine, A., Mohammadi, P., Banks, E., & Lappalainen, T. (2015). Tools and best practices for data processing in allelic expression analysis. *Genome Biology*, *16*(1). https://doi.org/10.1186/s13059-015-0762-6

Wang, X., & Clark, A. G. (2014). Using next-generation RNA sequencing to identify imprinted genes. *Heredity*, *113*(2), 156–166. https://doi.org/10.1038/hdy.2014.18

Discussion

Werren, J. H., Richards, S., Desjardins, C. A., Niehuis, O., Gadau, J., Colbourne, J. K., … Gibbs, R. A. (2010). Functional and evolutionary insights from the genomes of three parasitoid Nasonia species. *Science (New York, N.Y.)*, *327*(5963), 343–348. https://doi.org/10.1126/science.1178028

Conclusion

Wang, X., Werren, J. H., & Clark, A. G. (2016). Allele-Specific Transcriptome and Methylome Analysis Reveals Stable Inheritance and Cis-Regulation of DNA Methylation in Nasonia. *PLOS Biology*, *14*(7), e1002500. https://doi.org/10.1371/journal.pbio.1002500

APPENDIX A

MULTIPLE SNP ANALYSIS

Multiple SNP Analysis—Samples Aligned to *N. vitripennis* Reference. After filtering for read depth of 100, presented below are the number of single nucleotide polymorphisms (SNPs) in each hybrid replicate that 1) show bias towards the allele that came from the *N. vitripennis* parent, 2) show bias towards the allele that came from the *N. giraulti* parent, and 3) show no difference (ND) in expression of its parental alleles.

Sample ID ; genotype	SNPs in VIT direction	SNPs in GIR direction	SNPs with no difference
014450 ; GV	12,119	4,267	115
014451 ; GV	22,836	6,905	314
014452 ; GV	26,879	8,441	356
014453 ; VG	18,716	5,614	249
014454 ; VG	13,785	4,043	213
014455 ; VG	31,512	8,922	414

Multiple SNP Analysis—Samples Aligned to Pseudo *N. giraulti* Reference. After filtering for read depth of 100, presented below are the number of single nucleotide polymorphisms (SNPs) in each hybrid replicate that 1) show bias towards the allele that came from the *N. vitripennis* parent, 2) show bias towards the allele that came from the *N. giraulti* parent, and 3) show no difference (ND) in expression of its parental alleles.

Sample ID ; genotype	SNPs in VIT direction	SNPs in GIR direction	SNPs with no difference
014450 ; GV	9,753	7,898	234
014451 ; GV	18,121	14,191	470
014452 ; GV	21,785	16,906	518
014453 ; VG	14,759	11,940	313
014454 ; VG	10,681	8,764	302
014455 ; VG	24,613	19,646	575

APPENDIX B

GENE ANALYSIS

Gene Analysis for Multiple SNPs—Samples Aligned to *N. vitripennis* Reference. After filtering for read depth of 100, number of TRUE genes (SNPs showing bias towards the same allele) and number of FALSE genes (SNPs not showing bias towards the same allele) are presented below for each hybrid replicate.

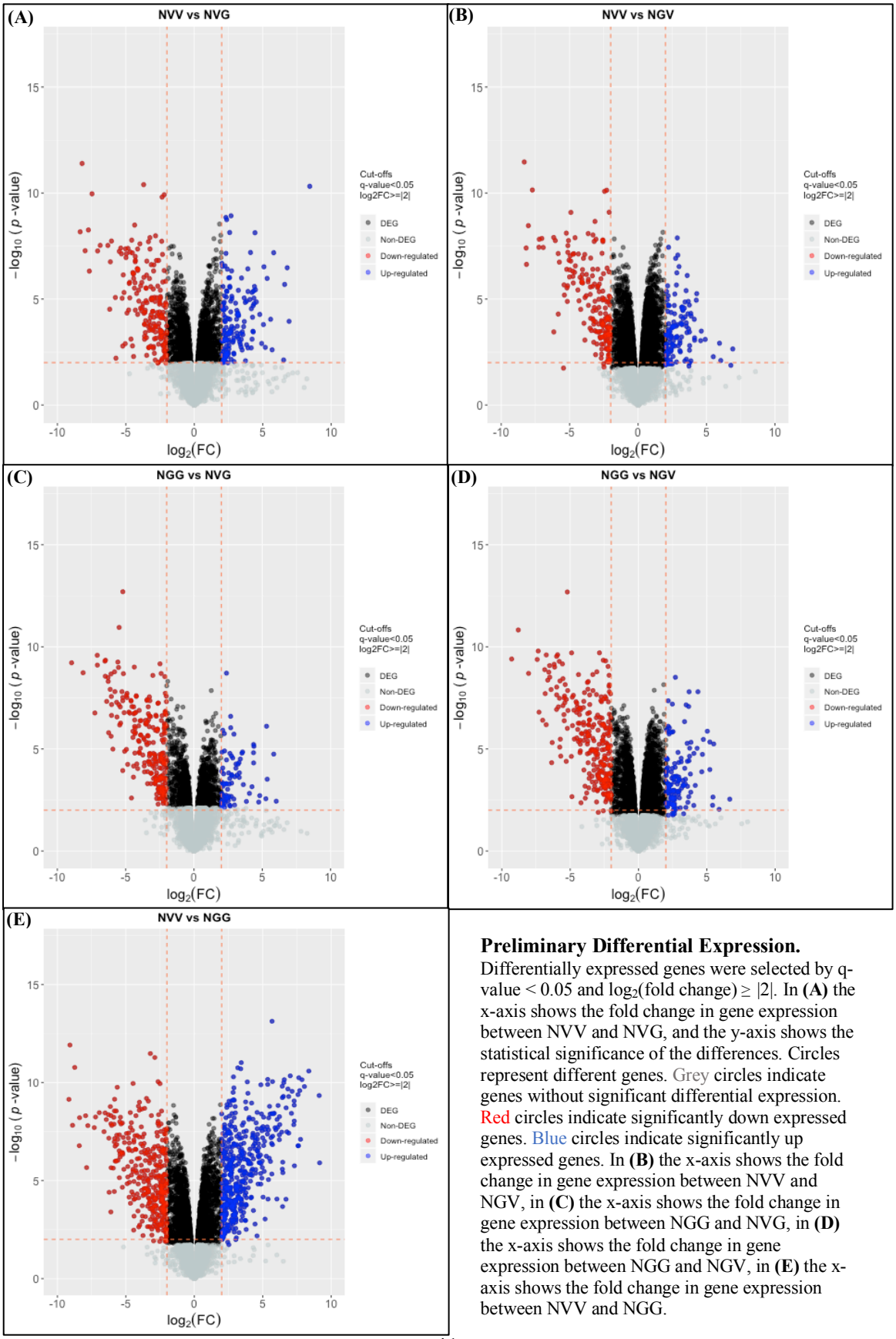
Sample ; Genotype	TRUE genes	FALSE genes	Total genes
014450 ; GV	884	993	1,877
014451 ; GV	1,278	1,564	2,842
014452 ; GV	1,348	1,849	3,197
014453 ; VG	1,065	1,302	2,097
014454 ; VG	806	999	1,805
014455 ; VG	1,531	2,003	3,534

Gene Analysis of Multiple SNPs—Samples Aligned to Pseudo *N. giraulti* Reference. After filtering for read depth of 100, number of TRUE genes (SNPs showing bias towards the same allele) and number of FALSE genes (SNPs not showing bias towards the same allele) are presented below for each hybrid replicate.

Sample ; Genotype	TRUE genes	FALSE genes	Total genes
014450 ; GV	800	1,158	1,958
014451 ; GV	1,013	1,937	2,950
014452 ; GV	1,123	2,218	3,341
014453 ; VG	849	1,640	2,489
014454 ; VG	637	1,253	1,890
014455 ; VG	1,125	2,511	3,636

APPENDIX C

VOLCANO PLOTS



Preliminary Differential Expression.
Differentially expressed genes were selected by q-value < 0.05 and $\log_2$(fold change) ≥ |2|. In **(A)** the x-axis shows the fold change in gene expression between NVV and NVG, and the y-axis shows the statistical significance of the differences. Circles represent different genes. Grey circles indicate genes without significant differential expression. Red circles indicate significantly down expressed genes. Blue circles indicate significantly up expressed genes. In **(B)** the x-axis shows the fold change in gene expression between NVV and NGV, in **(C)** the x-axis shows the fold change in gene expression between NGG and NVG, in **(D)** the x-axis shows the fold change in gene expression between NGG and NGV, in **(E)** the x-axis shows the fold change in gene expression between NVV and NGG.