Parental and endosymbiont effects on sex determination in haplodiploid wasps
Geuverink, Elzemiek

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 4

Alternative splicing and maternal provision of sex determination genes in the parasitoid *Asobara tabida*

Elzemiek Geuverink
Pieter Neerincx
Marloes van Leussen
Bart A. Pannebakker
Leo W. Beukeboom
Louis van de Zande
Eveline C. Verhulst

To be submitted to Insect Molecular Biology
ABSTRACT

Sex determination cascades evolve at a rapid pace causing variation in sex determination mechanisms in many organismal groups. Particularly insects use multiple mechanisms and show fast diverging sex determination genes. Little, however, is known about how this divergence relates to the functioning and evolution of sex determination cascades. In haplodiploid species of the insect order Hymenoptera only two mechanisms have been studied thus far: Complementary Sex Determination (CSD) and maternal effect genomic imprinting sex determination (MEGISD). The distribution of these mechanisms and conservation of the underlying genes are still to be determined. The parasitoid wasp Asobara tabida has tested negative for a CSD mechanism, yet the molecular and functional details of its sex determination system are still unknown. This study presents the characterization of genes associated with sex determination in A. tabida, based on a draft genome assembly. Sex determination genes doublesex (dsx), transformer (tra) and transformer-2 (tra2) are identified in A. tabida with the first two showing sex specific splicing suggesting their conserved function in sex determination. Tra and tra2 mRNA are maternally provided to embryos, but the maternal input of tra mRNA is restricted to unique non-sex-specific transcripts contrary to other insects. A. tabida sex determination differs from the MEGISD mechanism in its absence of sex-specific tra provision and may constitute a hitherto unknown haplodiploid sex determination mechanism.
INTRODUCTION

The genetic basis of developmental pathways is presumed to be well conserved due to the functional necessity of these basic processes in any organism. One of these basic processes is sex determination, a developmental pathway present in almost all eukaryotes, which governs the sexual differentiation leading to the development of female and male individuals or structures. Despite its universality within the eukaryotic domain, sex determination comprises a wide variety of fast evolving mechanisms. Although different genes are involved in the genetic pathway of sex determination, the overall structure is shared between the different mechanisms. A primary signal starts the cascade and is passed through a range of sex determination genes towards the downstream master switch, which regulates sexual differentiation genes (Herpin & Schartl, 2015). The downstream genes appear more conserved than the upstream signals as the cascade evolves from the bottom upwards (Wilkins, 1995). The master switch at the bottom of the sex determination cascade in insects is doublesex (dsx), which has been characterized in a range of insect species (Shukla & Nagaraju, 2010; Verhulst & van de Zande, 2015). It belongs to a group of DNA-binding motif (DM) encoding genes which are present among Metazoa and is potentially involved in sex determination of other invertebrates and vertebrates (Matson & Zarkower, 2012). Dsx is spliced in sex specific variants containing a second oligomerization domain (OD2), which leads to the production of male and female specific DSX proteins.

The gene controlling the splicing of dsx is transformer/feminizer (tra/fem). It is not functionally conserved outside insects, and even within insects, it is absent in Lepidoptera and basal lineages of Diptera (Geuverink & Beukeboom, 2014; Kiuchi et al., 2014). Tra belongs to a class of SR-type proteins that contain high frequencies of arginine (R) and serine (S) residues. Most insects have an order-specific domain in tra, e.g. the hymenopteran domain (HYM) in Hymenoptera and the dipteran domain (DIP) in Diptera (Verhulst et al., 2010b). The Ceritatis-Apis-Musca (CAM) domain is present in all tra homologs, excluding drosophilids, but only in female splice variants, and absent from male specific splice variants of tra (Hediger et al., 2010). Sex-specific tra RNA is produced by alternative splicing and only the female-specific splice form yields a functional protein that can splice premature tra mRNA in the female-specific form, thereby creating an autoregulatory loop. The male specific splice variants only result in truncated proteins. The activation of tra and the signal-relaying pathway of tra switching on dsx is remarkably conserved despite the diversity in mechanisms (Bopp et al., 2014). The functional female TRA protein forms a complex with the Transformer-2 protein (TRA2) to enforce female splicing of dsx (Amrein et al., 1990; Hedley & Maniatis, 1991; Inoue et al., 1992). The structure of TRA2 is highly conserved and contains a RNA binding domain (RDB) flanked by two arginine/serine rich regions in all species documented thus far. Studies in various dipterans, the coleopteran Tribolium castaneum and in the hymenopteran Apis mellifera have shown that TRA2 is involved in female-specific tra splicing (Burghardt et al., 2005; Concha & Scott, 2009;
Haplodiploid insects, comprising all Hymenoptera and Thysanoptera, but also several branches of Coleoptera and Hemiptera, are of special interest for sex determination research, because they lack sex chromosomes, i.e. chromosomes that inherit in a sex-specific manner. Males and females share all chromosomes but differ in ploidy level (haploid males, diploid females). Genetic studies of sex determination in haplodiploids have so far only been performed within the Hymenoptera. Model organisms are the honey bee A. mellifera (Apoidea) and the parasitoid wasp Nasonia vitripennis (Chalcidoidea). Information obtained from these species provides a framework for studies in other haplodiploid systems. A. mellifera has complementary sex determination (CSD) in which sex is determined by the allelic state of the complementary sex determiner (csd) locus, a paralog of tra (Beye et al., 2003; Hasselmann et al., 2008a). An individual that is homozygous or hemizygous at csd becomes male, while heterozygous individuals develop into females.

Dsx and tra are conserved in N. vitripennis (Oliveira et al., 2009; Verhulst et al., 2010a), but no paralogs of tra (i.e. csd) have been detected (Hasselmann et al., 2008a; Werren et al., 2010). A sex determination mechanism combining maternal effects and genomic imprinting (MEGISD) was described for N. vitripennis (Beukeboom & Kamping, 2006; Verhulst et al., 2010a). Females provide their eggs with female specific tra mRNA to start the autoregulatory loop which is necessary for female development (Verhulst et al., 2010a). To maintain the autoregulatory loop in fertilized embryos, the presence of the non-silenced paternal womanizer (wom) (Verhulst et al., 2013) allele, in addition to the silenced maternal wom allele, activates tra expression. In unfertilized embryos, only the silenced maternal wom allele is present resulting in no zygotic tra expression, leading to haploid male development. This form of parental imprinting sex determination has not yet been documented in any other hymenopteran.

The hymenopteran superfamily of Ichneumonoidea consists of species displaying a wide range of reproductive modes and life histories. Both CSD and non-CSD mechanism are found in this superfamily, even within a single genus (Heimpel & de Boer, 2008). Amongst Hymenoptera, a CSD mechanism can be inferred through inbreeding tests as its presence will then result in diploid males. However, the molecular mechanism of CSD has only been elucidated in A. mellifera (Beye et al., 2003; Hasselmann et al., 2008a; Gempe et al., 2009) and may consists of different genetic elements in other branches of Hymenoptera. It is unknown if MEGISD is a widespread sex determination mechanism in non-CSD systems and no molecular data is available for species lacking CSD to infer a (novel) sex determination mechanism. The detection of MEGISD is much harder than that of CSD as it relies on molecular assessment of sex determination genes, which is only feasible with transcriptomic or genomic data.

Parasitoids of the Asobara genus (Braconidae) are a group of well-studied ichneumonoid wasps. They occur worldwide and use Drosophila larvae as hosts (Carton et al., 1986). The sex determination mechanism of Asobara tabida has tested negative for single locus...
CSD through inbreeding crosses (Beukeboom et al., 2000) and subsequent inbreeding tests for multi locus CSD rejected any models for up to ten sex determining loci in A. tabida, A. pleuralis, A. citri and A. japonica (Ma et al., 2013). In absence of CSD, a mechanism similar to the MEGISD model of N. vitripennis could be present in Asobara wasps. In this study we characterize the genes and their role in the sex determination cascade of A. tabida. We identify the central axis genes dsx, tra2 and tra and examine whether these genes are sex-specifically spliced. We describe the expression and splicing of these sex determination genes during embryo development. By measuring the expression in the early stage embryos we assess the maternal provision of sex-specific tra mRNA to examine if MEGISD could be the sex determination mechanism in A. tabida, similar to Nasonia. To facilitate the identification of these and future candidate genes we present the first draft of the A. tabida genome sequence.

**MATERIAL AND METHODS**

**Insect culturing**

The highly inbred TMS strain (Ma et al., 2013), that originates from strain SOS (Sospel, France) and Italy (Pisa, Italy), was used as genomic source material and in subsequent experiments. The wasps were cultured on second instar Drosophila melanogaster host larvae (72 hours of development) at 20°C under constant light.

**Asobara genome and transcriptome assembly**

A whole-genome next-generation sequencing approach was used to sequence the genome of 20 pooled adult females from the TMS strain. Two libraries were constructed with an insert size of 380bp and 780bp, and paired-end sequenced in two lanes on an Illumina HiSeq2000 resulting in, respectively, ~78M reads and ~68M reads of 100bp, giving an approximate genome coverage of 60x.

De novo genome assembly was performed using the parallel version ABySS 1.5.2 (Simpson et al., 2009) with default settings. First, a k-mer sweep was done to find the optimal k-mer for assembly and with a k-mer of 56 the final assembly was done. To optimize subsequent scaffolding in ABySS, the male and female RNAseq assemblies (see below) were used in the final re-scaffolding step. Genome assembly statistics were determined using Quast 4.1 (Gurevich et al., 2013).

An adult male and female transcriptome were separately sequenced from a pool of 11 females and 25 males. Females and males were stored in RNAlater in batches of 3 or 5 specimens 24-48h after emergence. RNA extractions were performed according to manufacturer’s protocol with TriZol (Invitrogen, Carlsbad, California, USA) followed by an additional ethanol wash. The constructed libraries were barcoded and single-end sequenced in one lane on an Illumina HiSeq2000 resulting in ~36M reads for the male library, and ~23M reads for the female library. Read length was 96bp after removal of the barcodes.
De novo transcriptome assembly was performed using Trinity version r2011-11-26 (Grabherr et al., 2011) with default settings. Transcriptome statistics were determined using TrinityStats, included in the Trinity version 2.2.0 package.

Ortholog identification

Orthologs of sex determination and methylation genes were identified using translated BLAST (Altschul et al., 1997) against the *A. tabida* genomic assembly. DSX and TRA protein sequences of *Apis mellifera* (ABW99105, NP_001128300) and *Nasonia vitripennis* (ACJ65507, NP_001128299) and the TRA2 sequence of *A. mellifera* (NP_001252514) were used as queries. Prior to availability of the genomic assembly, a small fragment of *transformer* was detected in an *A. tabida* EST dataset (Kremer et al., 2012) using translated BLAST. This fragment was used for initial primer development.

RNA extraction, cDNA synthesis and splice-variant detection

Adult females and males were individually collected 24-48h after their emergence from the *D. melanogaster* pupae. RNA extraction was performed according to manufacturer’s protocol with TriZol (Invitrogen, Carlsbad, California, USA). For 3'RACE (Rapid Amplification of cDNA Ends), RNA was reverse-transcribed with the RevertAidTM H Minus First Strand cDNA Synthesis Kit (Fermentas, Hanover, MD, USA) using 25µM 3'RACE adapter (5'-GCG AGC ACA GAA TTA ATA CGA CTC ACT ATA GGT 12VN-3') from FirstChoice RLM-RACE kit (Ambion, Austin, TX, USA). For 5'RACE, RNA was processed according to manufacturer’s instructions (FirstChoice RLM-RACE kit, Ambion, Austin, TX, USA) and reverse-transcribed using the RevertAidTM H Minus First Strand cDNA Synthesis Kit (Fermentas, Hanover, MD, USA). To assess the *At-tra* splice variants present in adult males and females 5'RACE-PCR was performed with outer primer Attra5RACEout (5’-CCATTCTGAAGTCGATCTGC-3’) and inner primer Attra5RACEin (5’-CTTCGTGGACTTGATTCTCCTCT-3’) in a reaction at 94°C for 3 minutes, 35 cycles of 94°C for 30 seconds, 58°C for 30 seconds and 72°C for 2 minutes, with a final extension of 10 minutes at 72°C. Outer primer Attra3RACEout (5’-CAGGAGAAGGCTCGAACC-3’) and inner primer Attra3RACEin (5’-GCAGGAAGGCTCGAACC-3’) were used in 3’RACE-PCR at an annealing temperature of 55°C and otherwise identical conditions. 5’RACE-PCR of *At-dsx* was performed with outer primer Atdsx5RACEout (5’-ATCACTTTTCTGTATTCGATCGT-3’) and inner primer Atdsx5RACEin (5’-CCTGGATATCTTCTTCTTCGATC-3’) in a reaction at 94°C for 3 minutes, 40 cycles of 94°C for 30 seconds, 56°C for 30 seconds and 72°C for 60 seconds, with a final extension of 7 minutes at 72°C. Outer primer Atdsx3RACEout (5’-GGGACACAAGAGATATTGCA-3’) and inner primer Atdsx3RACEin (5’-GGGACACAAGAGATATTGCA-3’) were used in 3’RACE-PCR at an extension time of 2 minutes and otherwise identical conditions. 5’RACE-PCR of *At-tra2* did not yield any fragments after various attempts and was eliminated from the tests. RNAseq isotigs and the genomic contig yielded a putative *tra2* homolog including the 5’UTR (Table 4.1). 3’RACE-PCR of *At-tra2* was performed with outer primer Attra23RACEout (5’-AGGAGCAGGTCTTTACATT-3’)
and inner primers Attra23RACEin1/Attra23RACEin2 (5’-TAGGAGTCCAATGTCATCAAGAAGG-3’/5’-TCAATGATGAAAGACTGGGAG-3’) in a reaction at 94°C for 3 minutes, 40 cycles of 94°C for 30 seconds, 56°C for 30 seconds and 72°C for 60 seconds, with a final extension of 7 minutes at 72°C.

Reverse transcription (RT)-PCRs to confirm the splicing variation of each gene were performed with primers AttraF (5’AAACAGTGAGAATAGAC-3’) and AttraSSR (5’-CTGTATGTGAAGGTTGAAACGA-3’) for At-tra, primers AtdsxFrontF (5’-CTCCACCCGTTACAAGTGTG-3’) and AtdsxbackR (5’-GGTAGAGCTCAGCCTCTGAC-3’) for At-dsx, and primers Attra2FrontF(5’-CCACGAGAACTCGGCAG-3’) and Attra2exon4R (5’-CCTCAAAGTTCCCTCTATCC-3’) for At-tra2. All RACE-PCR and RT-PCR products were ligated into pGEM-T vector (Promega, Madison, WI, USA) after purification using GeneJET Gel Extraction Kit (Fermentas, Hanover, MD, USA). Ligation reactions were used to transform competent JM-109 Escherichia coli (Promega, Madison, WI, USA). Colony-PCR was conducted by use of pGEM-T primers (5’-GTAAACGAGCAGGTC-3’ and 5’-GGAGAAAGCTATGACCATG-3’) at 94°C for 3 minutes, 30 cycles of 94°C for 30 seconds, 55°C for 30 seconds and 72°C for 2 minutes, with a final extension of 7 minutes at 72°C. Both strands were sequenced and fragments were aligned to one another and to the assembled genomic contigs to inspect the splicing variation. The structure of the genes was visualized with Exon-Intron Graphic Maker (http://wormweb.org/exonintron). Alignments of TRA, TRA-CAM, TRA2 and DSX were produced with Geneious8 (Biomatters Ltd) and gene trees were constructed using a Maximum Likelihood algorithm in MEGA7 (Kumar et al., 2016).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Scaffolds</th>
<th>Conservation (E-value compared to N. vitripennis amino acid sequence)</th>
</tr>
</thead>
<tbody>
<tr>
<td>transformer</td>
<td>1743548, 1858287</td>
<td>9e-005</td>
</tr>
<tr>
<td>doublesex</td>
<td>1858048, 1801331, 1850441, 1804618, 641499</td>
<td>4e-014</td>
</tr>
<tr>
<td>transformer-2</td>
<td>1856765, 35458, 1827793</td>
<td>4e-026</td>
</tr>
</tbody>
</table>

Embryo collection for At-tra expression and splice variation

Hosts containing TMS strain wasps in the pupal stage were separated individually in tubes to prevent females from mating. A batch of mated females was collected from mass culture bottles. Groups of three virgin or mated females were allowed to parasitize hosts for 2 hours to improve host detection. Following this pre-treatment they were kept for 2 days at 12°C and provided with honey for feeding. Next, each group of wasps was provided with 30 hosts for 2 hours every third day, alternated with a period of 2 days without hosts but honey-feeding. This allowed for collection of embryos of life stages up to 24-26 hours. As differences in development and chances of encapsulation by the host increase in later developmental stages.
of the hosts, the wasps were allowed more time to parasitize during the collection points of 48-52 hours (4 hours), 72-76 hours (4 hours) and 120-144 hours (24 hours). Petri dishes containing parasitized hosts were kept under constant light at 20°C. After the allotted development time the petri dishes were rinsed with water and the host larvae collected and stored crushed in TriZol at -80°C. A subset of larvae from each group and time-point was left to develop into adults. This served as control for the virginity of the unmated females, who indeed only produced male offspring. The progenies from mated females yielded information about offspring sex ratio, which was on average 0.38 (proportion male). Of each group and time-point, 6 tubes with each 10 parasitized larvae were used for RNA extraction, as described for adult tissue above. All total RNA was used for cDNA synthesis, as each sample consisted only for a small part of the parasitoid RNA. Reverse transcription was performed with a mixture of 1:6 random oligo-dT:random hexamers from the RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas, Hanover, MD, USA). Controls of unparasitized *D. melanogaster* larvae of similar age groups were included.

*Quantitative real-time PCR*

For quantitative real-time PCR (qPCR) both embryological and adult samples were diluted 1:50. 5µL of diluted cDNA was combined with 10µL PerfeCTa SYBR Green Fast Mix (Quanta BioSciences, Gaithersburg, USA) and 200nM of primers. Two different primer sets were used to differentiate between the *At-tra* splice variants: *AttraSSqF* (5’-GAACGTGAAGAACTGAAGGTTGAG-3’) plus *AttraSSqR* (5’-CTTGACCTCCCGTCATGCCT-3’) and *AttraNSSqF* (5’-GAAGAGAAGGAGAAGGCTCG-3’) plus *AttraNSSqR* (5’-GAAGGGTGGGATGAATAAGG-3’). The first set measures the expression of both female-specific and male-specific splice variants, combined noted as SS (sex-specific). The latter set amplifies the non-sex-specific (NSS) splice variants of *transformer* in *A. tabida*, which were chosen to be combined as their amino acid sequence shows very strong similarity. The primer set for *At-dsx*, consisting of *AtdsxqF* (5’-TTCAGCAATGTACCAATCGGTG-3’) and *AtdsxqR* (5’-TACCAGAATTCCAGAAGTTAGC-3’), amplified all splice variants. Primers *Attra2qF* (5’-TAACCGAAGCAGACATAACAC-3’) and *Attra2qR* (5’-GAGCTTTCTCTACGCTG-3’) were used to amplify the single mRNA transcript of *At-tra2*. *Elongation factor 1 alpha* was used as the reference gene with primers *EF1αF* (5’-TCACCGCTCAGGATTATGCT-3’) and *EF1αR* (5’-GGCACAAGGCGAGAAGTTAGC-3’). All primer sets were used at 95°C for 15 minutes, 45 cycles of 95°C for 15 seconds, 56 °C for 15 seconds, 56 (At-dsx and At-tra2) or 58 (EF1α, At-traSS and At-traNSS) °C for 30 seconds and 72°C for 30 seconds. Dissociation curves were produced to check for non-specific amplification. An amplified product of each primer set was cloned, according to the protocol above, to confirm the sequence identity of the amplicon. Negative control samples of unparasitized *D. melanogaster* larvae were tested under the same qPCR conditions and did not show any amplification after 45 cycles.
Relative mRNA levels were determined by dividing At-tra2, At-traSS, At-traNSS and At-dsx N0 values by At-EF1α N0. At-tra2, At-traNSS and At-dsx were tested in a general linear model with categorical factors fertilization and time-point. A Kruskal Wallis test was used to compare the relative mRNA levels of At-traSS between fertilized and unfertilized eggs sorted by time-point. Relative mRNA levels of At-traSS across development differences between females and males were compared with a Mann-Whitney U test.

**Splice variant presence during development**

RT-PCRs with primers AttraRTF (5'-TCTTCGTGACTATCAAATCC-3') and AttraRTR (5'-TTCTCAACCTCAGTTCTTAC-3') were performed on embryo and adult cDNA samples at 94°C for 3 minutes, 45 cycles of 94°C for 30 seconds, 54°C for 30 seconds and 72°C for 2 minutes, with a final extension of 7 minutes at 72°C. The non-sex-specific At-tra transcripts were amplified with the same primers used in the qPCR AttraNSSF (5'-GAAGAGGAAGGAGAAGGCTCG-3') and AttraNSSR (5'-GAAGGGTGGGATGAAATAAGG-3'). Reactions were performed at 94°C for 3 minutes, 45 cycles of 95°C for 30 seconds, 58°C for 30 seconds and 72°C for 30 seconds, with a final extension of 7 minutes at 72°C. At-dsx splicing was amplified using AtdsxRTF (5'-GGAGCAATTCTGATTTATGGGGTGGGATGAAATAAGG-3') and AtdsxRTR (5'-CCTGATTTATGGGTGCATATTGTG-3') at 94°C for 3 min, 40 cycles of 94°C for 30 sec, 55°C for 30 seconds, 72°C for 1 minute, with a final extension of 7 minutes at 72°C. Products were run and visualized on an ethidium bromide-containing 1.5% agarose gel.

**RESULTS**

**Draft genome of A. tabida**

De novo assembly of two illumina paired-end read libraries resulted in a draft genome of the A. tabida inbred TMS strain with 925,296 contigs and a total length of 271 Mb. Subsequent scaffolding using the male and female RNAseq data resulted in 48,930 scaffolds bigger than 500bp having a total length of 202 Mb with half of the bases residing in contigs larger than 8291bp (N50). The largest contig is 258 kb and 245 contigs are larger than 50 kb. As A. tabida is suggested to possess 12-17 chromosomes (G. Massimo, pers. comm.) and a genome size of 286Mb (J.G. de Boer, pers. comm.), we assembled 71% of the genome in contigs larger than 500 bp. We assessed the completeness of the assembly using BUSCO (Simão et al., 2015). Of the 2675 conserved Arthropod orthologs tested, 78.6% were present in complete form, of which 4.8% were found multiple times. 14.8% of the orthologs were found to be fragmented and 6.7% could not be detected in the assembly.

De novo assembly of the RNAseq libraries resulted in two draft transcriptomes of A. tabida adult males and females. The male assembly contains 32,484 genes coding for 44,403 transcripts, and the female assembly contains 23,968 genes coding for 30,259 transcripts. The transcriptomic assemblies were assessed with BUSCO for the presence of the 2675 conserved
Arthropod orthologs. The male and female datasets revealed respectively 77.8% and 77.2% complete orthologs, of which 23.6% and 17.9% were present multiple times. 8.9% and 6.6% of the orthologs were fragmented and 13.2% and 16.2% could not be detected in the transcriptomic assemblies. The two assemblies were used to re-scaffold the genome assembly yielding longer contigs.

**Figure 4.1.** Exon-intron structure of the female-specific, male-specific and non-sex-specific (NSS) splice variants of *tra* in *A. tabida*. White boxes represent the 5’ and 3’UTR’s, the black boxes contain the coding sequence. The scale bars on the right represent 100bp. Primer positions for qPCR are depicted marked qF (forward) and qR (reverse). Primers used to detect splice variation by RT-PCR are shown at rtF (forward) and rtR (reverse). Two different primer sets are used to test for sex-specific (SS) and non-sex-specific (NSS) splicing. Incomplete intron lengths are marked by breaks in the line.

**Identification of sex determination genes**

Orthologs of *dsx*, *tra2* and *tra* were identified in the *A. tabida* genomic assembly, extending this conserved part of sex determination genes to the superfamily of Ichneumonoidea within the Hymenoptera (Table 4.1). Many species of Hymenoptera possess a paralog of *tra* (Schmieder et al., 2012; Privman et al., 2013; Jia et al., 2016), but no paralogs of *tra* were detected in *A.
Alternative splicing and maternal provision of sex determination genes in the parasitoid A. tabida. Single homologs of tra, tra2 and dsx are present, which we will describe based on their position in the sex determination cascade.

**Sex-specific and non-sex-specific splicing of tra**

Identification of tra splice variants in adult A. tabida revealed a characteristic female and a characteristic male splice variant (Figure 4.1 and Figure S4.1a). The female splice variant (At-tra\(^F\)) translates into an amino acid sequence contains all conserved TRA domains: the Hymenoptera domain (HYM), the CAM domain, an arginine-serine domain (RS-domain) and a proline-rich region. It is placed in the Hymenoptera phylogeny more basal to the derived Aculeata species (ants and bees) (Figure S4.1b). The male splice variants (At-tra\(^{M1}\), At-tra\(^{M2}\) and At-tra\(^{M3}\)), on the other hand, contain additional exons, which change their open reading frames (ORF), resulting in a truncated protein with only the HYM domain present towards the 5’ start of the gene. The inclusion of male-specific exons is characteristic of the sex-specific splicing of tra, where the CAM domain is typically absent in males.

Two additional At-tra splice variants (At-tra\(^{NSS1}\) and At-tra\(^{NSS2}\)) are present in both sexes and thus noted as non-sex-specific (NSS) (Figure 4.1). The two At-tra\(^{NSS}\) splice variants are comparable, only At-tra\(^{NSS2}\) retains the last intron, close to the stop-codons in the ORF. Both splice variants contain a putative second CAM domain, followed by a shortened RS-rich region compared to the one present in At-tra\(^{F}\) (Figure 4.1). The two CAM regions (each 75bp) show high similarity at the amino acid level, but are more diverged on the nucleotide level (Figure 4.2).

**Figure 4.2.** Alignment of the two putative CAM domains of At-tra against hymenopteran, dipteran and coleopteran sequences. Nucleotide sequences and their relative conservation in grey scale are shown at the top with the amino acid sequence noted directly below. The first 31 identical nucleotides from At-tra\(^{F}\) and At-tra\(^{NSS}\) are transcribed from the same exon 3, from nucleotide 32 onwards the At-tra\(^{F}\) sequence has continued on exon 7 and the At-tra\(^{NSS}\) sequence on exon 4.
No alternative splicing of tra2

A single splice form of tra2 was identified in A. tabida (Figure 4.3). At-tra2 is highly conserved and its translated amino acid sequence contains the characteristic RNA-binding domain (RBD) with a large concentration of arginine and serine amino acids on either side (Figure S4.2A and S4.2B). Alternative splice forms were neither detected in the two sexes nor at different stages of development. Many species only transcribe a single splice form of tra2 (Burghardt et al., 2005; Concha & Scott, 2009; Salvemini et al., 2009; Sarno et al., 2010; Schetelig et al., 2012; Liu et al., 2015). A. mellifera (Nissen et al., 2012) and N. vitripennis (Chapter 3) do possess alternative splice variants of tra2, however these do not display any variation in presence across sexes and life stages.

Figure 4.3. Exon-intron structure of tra2 in A. tabida. White boxes represent the 5’ and 3’ UTR’s, the black boxes contain the coding sequence. The RBD is plotted in grey on the exons. The scale bars on the right represent 100bp. Primer positions for qPCR are depicted marked qF (forward) and qR (reverse). Incomplete intron lengths are marked by breaks in the line.

Figure 4.4. Exon-intron structure of the female-specific and male-specific splice variants of dsx in A. tabida. White boxes represent the 5’ and 3’ UTR’s, the black boxes contain the coding sequence. The scale bars on the right represent 100bp. Primer positions for qPCR are depicted marked qF (forward) and qR (reverse). Primers used to detect splice variation by RT-PCR are shown at rtF (forward) and rtR (reverse). The alternative splice variant present in 12-14h old embryos is noted by the grey arrow and bar in front of exon 5. Incomplete intron lengths are marked by breaks in the line.
Sex specific splicing of *dsx*

The structure of AT-DSX, including the DM domain and the OD2 domain, is conserved (Figure S4.3a) and phylogenetically clusters with other hymenopteran DSX orthologs (Figure S4.3b). At-\textit{dsx} is sex-specifically spliced in adult \textit{A. tabida}, consisting of one female-specific and one male-specific splice form (Figure 4.4). The female splice variant includes exon 4, which results in a female-specific OD2 domain and a shorter translated sequence compared to the male splice variant which splices out exon 4, resulting in a male-specific OD2 domain. The presence of sex-specific \textit{dsx} splicing suggests functional conservation of the bottom of the sex determination cascade where \textit{dsx} regulates sexual differentiation.

Sex-specific At-tra mRNA is not maternally provided

\textit{At-tra}^{ \text{F} } and \textit{At-tra}^{ \text{M} } mRNA are absent in 0-2h old embryos (Figure 4.5a, measured combined as \textit{At-tra}^{ \text{SS} }), indicating that the mother does not provide sex-specific \textit{tra} (\textit{At-tra}^{ \text{SS} } ) mRNA to her offspring. At 12-14 hours of development expression of \textit{At-tra}^{ \text{M} } has started in unfertilized (male) embryos (Figure 4.5a and 4.6a). At this point, no splice variant can be detected in the fertilized eggs using this set of primers (Figure 4.6), however expression levels are equal to that of the unfertilized embryos (Figure 4.5). The RT-PCR splicing results fluctuate in 12-14h old embryos as zygotic transcription is just starting and split between all four \textit{At-tra}^{ \text{SS} } splice variants (\textit{At-tra}^{ \text{F} }, \textit{At-tra}^{ \text{M1} }, \textit{At-tra}^{ \text{M2} } and \textit{At-tra}^{ \text{M3} }), whereas the qPCR that amplifies the common region of the \textit{At-tra}^{ \text{SS} } splice variants is more sensitive. Amplicons obtained in 12-14h old embryos with a range of \textit{At-tra}^{ \text{SS} } or \textit{At-tra}^{ \text{M} } specific primers are shown in Supplementary figure 4.4. After the start of zygotic \textit{At-tra}^{ \text{SS} } transcription higher levels of \textit{At-tra}^{ \text{SS} } mRNA are present across female embryonic development (Z=2.48, p=0.013).

At 24-26 hours \textit{At-tra}^{ \text{F} } splicing has initiated in fertilized eggs (Figure 4.6). \textit{At-tra}^{ \text{F} } is not present at any point in the development of unfertilized eggs (Figure 4.6) and the relative mRNA levels shown in Figure 4.4a for unfertilized eggs consist solely of \textit{At-tra}^{ \text{M} } transcripts. The \textit{At-tra}^{ \text{SS} } expression of fertilized eggs shows a similar pattern (Figure 4.5a), but here the transcripts consists mostly, but not exclusively, of \textit{At-tra}^{ \text{F} } mRNA. Both haploid male (from unfertilized eggs) and diploid female (from fertilized eggs) development peaks in its sex-specific \textit{tra} expression after 48-52 hours of development (Figure 4.5a). The peak of expression in unfertilized eggs would be surprising in systems with maternal provision of \textit{tra}^{ \text{F} }, as only the female specific splice variant of \textit{tra} is expected to undergo autoregulation. This is the case in \textit{N. vitripennis} where \textit{tra} expression peaks at the start of zygotic transcription in fertilized eggs only (Verhulst et al., 2010a).

Expression of \textit{At-tra}^{ \text{NSS} } variants does not only occur in adults but is also observed in embryos of all developmental stages and in both sexes (Figure 4.5b and Figure 4.6). These variants are present in embryos less than 2 hours of age, indicating that the mother provides these mRNAs to her eggs. The expression of the combined \textit{At-tra}^{ \text{NSS} } variants follows a pattern similar to the \textit{At-tra}^{ \text{SS} } expression in the later stages of development (Figure 4.5b). The relative
levels of $At-tra^{NSS}$ mRNA do not differ between female and male embryonic development ($F_{(1, 57)}=1.2975, p=0.26$).

Figure 4.5. Relative expression of sex determination genes during development of diploid female (fertilized eggs) and haploid male (unfertilized eggs) offspring. Note that the RNA pool of diploid embryos also contains haploids as mated females lay a mixture of fertilized and unfertilized eggs. Relative mRNA levels of (a) combined $At-tra^{CSS}$ female and male splice variants, (b) combined $At-tra^{NSS}$ splice variants, (c) $At-tra2$ and (d) $At-dsx$. Error-bars in all figures display the standard error per category.

Maternal provision of $At-tra2$

$At-tra2$ mRNA is maternally provided to embryos (Figure 4.5c). It remains present in all stages of both male and female development and shows no peak of mRNA levels. $At-tra2$ shows more fluctuation in males and is overall higher expressed during male development than during female development ($F_{(1, 52)}=19.592, p<0.005$).

Expression of $At-dsx$ during development

Low $At-dsx$ mRNA levels are visible in the earliest stages of development (Figure 4.5d), indicating maternal provision. This is not visible on the RT-PCR gels (Figure 4.6), but this presumptive absence of bands may be a technical artifact as $At-dsx$ is split into three splice variants each of
low abundance. The third splice variant of $At-dsx$ is a slight modification of the $At-dsx^M$ (Figure 4.4) and is only observed in some samples of both haploid and diploid embryos at the 12-14h time point. This $At-dsx^{12h}$ transcript has a small 54bp addition at the 5’end of exon 5 and this leads to a shortened ORF. At 12-14h the $At-dsx^F$ splice variant is not yet present in diploid eggs. Only after sufficient $At-tra^F$ is present, $At-dsx$ is spliced into the female mode at 24-26h of development (Figure 4.6).

After 12-14h the haploid embryos exclusively display the male-specific splice variant of $At-dsx$ (Figure 4.6) and this splicing pattern continues in adult samples. However, in diploid embryos and in adult females, a mixture of $At-dsx^F$ and $At-dsx^M$ is present in all stages after 12-14h. This pattern of $dsx^M$ leakage in females matches the patterns observed in $A. mellifera$ and $N. vitripennis$, where the females are the sex containing both splice variants (Verhulst et al., 2010a; Nissen et al., 2012) (Chapter 4). This is consistent with male development as the default state of the sex determination cascade.

![Figure 4.6](image)

**Figure 4.6.** Splice variants of $At-tra$ and $At-dsx$ during development. Splicing of $At-tra$ is depicted in diploid female (fertilized eggs) and haploid male (unfertilized eggs) offspring depicted for splice variants $At-tra^F$/$At-tra^M$ (a) and $At-tra^{NOS}$ (b). Splicing of $At-dsx$ in diploid female (fertilized eggs) and haploid male (unfertilized eggs) offspring is shown in (c). Adults of $A. tabida$ are included for female and male splicing patterns. $D. melanogaster$ larvae and no template reactions are included as negative controls.
DISCUSSION

Conservation of the sex determination cascade in A. tabida

The key insect sex determination genes tra, tra2 and dsx are conserved in A. tabida. Though the sequence of At-tra is strongly diverged, its role in the sex determination cascade appears conserved, signified by its sex-specific splicing patterns appearing at 12-14h of development followed by the activation of sex-specific splicing in At-dsx at 24-26h. AT-TRA\textsuperscript{F} possesses all known domains, including the HYM domain. This suggests that the HYM domain, previously documented in the derived Aculeata and N. vitripennis (Verhulst et al., 2010b) (plotted in Figure S4.1a), is conserved in all Hymenoptera. TRA2 displays the distinctive lack of sex-specific splicing and shows strong conservation of the RBD with flanking RS regions. The structure of DSX is similarly conserved with the presence of the OD2 domain. This suggests that the central axis of tra-dsx and start of sexual differentiation are highly conserved in the A. tabida sex determination cascade, even though At-tra regulation appears to deviate from the known insect mechanisms.

Regulation of female-specific tra and dsx

From the 24h developmental stage onwards, a peak of female-specific At-tra expression appears in diploid, fertilized eggs and At-tra splices into the female mode. In haploid embryos from unfertilized eggs, At-tra\textsuperscript{SS} expression peaks around 48h (Figure 4.5a) but the male splice variant is already detectable from 12h onwards (Figure 4.6). It has to be noted that expression patterns in fertilized eggs are less pronounced, because of the experimentally unavoidable inclusion of unfertilized male embryos (i.e. under our experimental set-up mated females lay 38% unfertilized and 62% fertilized eggs). Downstream of At-tra, At-dsx is spliced into sex-specific variants, consistent with its conserved role in insect sex determination. As At-dsx is not spliced into the female-specific variant before At-tra\textsuperscript{F} is present, it appears that as in all other studied insects, AT-TRA regulates At-dsx splicing.

Tra2 is present in early embryos, similar to its early expression in A. mellifera (Nissen et al., 2012) and N. vitripennis (Chapter 4). The potential complex with the female specific TRA protein can however only be formed at a later stage in development, when At-tra\textsuperscript{F} is actively transcribed in the zygote. Non-sex specific At-tra\textsuperscript{NSS} is present prior to zygotic transcription and most likely maternally provided, but it is not yet known whether TRA\textsuperscript{NSS} products could form a complex with TRA2, or if only female-specific TRA is involved in this interaction in A. tabida.

Functionality of alternative tra splice variants

The presence of multiple putative CAM domains in At-tra could have its origin in a recombination event with a subsequently lost copy of tra. If any signal of such a duplication remains it must be a very degenerate pseudogene, as no remaining sequence of this paralog can be detected in the genomic assembly. Duplications of tra occur widely in Apoidea and Vespoidea.
Alternative splicing and maternal provision of sex determination genes in the parasitoid *A. tabida* (Schmieder *et al*., 2012; Privman *et al*., 2013), and appear to be rapidly gained and lost (Koch *et al*., 2014). *Tra* duplications seem a specific feature of the Hymenoptera, as these paralogs have thus far not been identified in other insect orders (Geuverink & Beukeboom, 2014). The functional significance of these extra *tra* copies is unclear, except for the honey bee, in which the *tra* paralog, *csd*, has opted a sex determination function. Additional pseudogenes have also been found in five of the species that also possess *tra* paralogs (Koch *et al*., 2014). The alternative CAM domain present in the *At-tra<sup>NSS</sup>* transcripts is shorter than the CAM domain in *At-tra<sup>F</sup>* , and is also followed by a shorter arginine-serine rich region and an absence of a proline rich region. It is unknown whether these *At-tra<sup>NSS</sup>* splice variants in *A. tabida*, which have not been detected in species outside the *Asobara* genus (Chapter 6), could have a similar functionality in starting autoregulation as female-specific *tra* in other species. Additionally, non-sex-specific patterning and high expression in each developmental stage may indicate a role in overall development of the embryo and thus represent a neo-functionalization. Its provision to early embryos might also point to a role as splicing factor in starting the female-specific sex determination cascade. This would however only constitute as a part of such a mechanism, as *At-tra<sup>NSS</sup>* is provided to both fertilized and unfertilized eggs. It would require an activator on the paternal chromosome set to start the female developmental pathway in fertilized diploid eggs.

Figure 4.7. Characteristics of the sex determination mechanism in *Nasonia vitripennis*, *Apis mellifera* and *Asobara tabida*. Male haploid development is plotted on the left side of each box and female diploid development on the right side.
Absence of traF maternal provision

As no maternal At-traF mRNA is provided to the eggs, a yet unidentified factor must be invoked to initiate the tra autoregulatory loop. This resembles the start of the A. mellifera sex determination cascade where csd initiates the female specific splicing of fem (tra), which then maintains its own female specific splicing. A. mellifera mothers do not provide femF mRNA to their offspring and zygotic transcription of femF is only detected after 24 hours of development (Gempe et al., 2009). However, A. tabida does not contain a CSD mechanism (Beukeboom et al., 2000; Ma et al., 2013) and lacks any paralogs of tra, thus the unidentified activator is part of a different type of mechanism.

Variation of sex determination mechanisms in Hymenoptera

The sex determination mechanism of A. tabida shows considerable differences with that of N. vitripennis and A. mellifera (Figure 4.7), in that it respectively lacks maternal provision of female-specific tra and tra paralogs. This is consistent with the fast evolution of sex determination mechanisms and their underlying cascade of genes, even within insect orders. A. tabida does not determine sex by a CSD mechanism and the absence of a tra paralog is consistent with a non-CSD system. The mechanism of A. tabida could work similar to N. vitripennis, but apparently would function without maternal provision of female-specific tra mRNA. The N. vitripennis MEGISD mechanism consists of maternal provision of female-specific tra mRNA and activation of tra expression by a maternal imprinting or paternal signal to initiate and maintain tra autoregulation. As mentioned, we did not detect any female-specific mRNA provision of any sex determination gene. At-traNSS and At-tra2 mRNA are provided by the mother to the embryos, but these transcripts are not distinct female splice variants. At-tra2 is highly conserved and its maternal provision would suggest conservation of the canonical TRA/TRA2 complex. The translation of the At-traNSS transcripts leads to peptides with similar domains (CAM-like domain followed by a short arginine-serine rich region) to the female-specific TRA isoform. This could suggest an interaction between AT-TRANSS and AT-TRA2, possibly leading to the activation of the zygotic female pathway. This activation would however need to contain a sex specific activation step, which we have not identified here and the mechanism that initiates At-tra autoregulation in fertilized eggs only is therefore unclear. Initially, At-dsx mRNA seems to be spliced in the male mode in both fertilized and unfertilized embryos at 12-14h of development. At this time point both At-traNSS and At-tra2 mRNA are already present as a result of the maternal provision, but apparently not capable of directly activating At-traF splicing, which is required for At-dsxF splicing. Female At-tra splicing is clearly present at 24-26h, which coincides with At-dsx switching to female-specific splicing in diploid embryos. This suggests that a certain threshold of At-tra mRNA is required before the female sex determination cascade can be activated. The function of the abundant maternal provision of non-sex specific tra mRNA, and the putative duplicate CAM domain in these splice variants are currently unclear and further study is required to determine their role in the functioning of sex determination in A. tabida. Functional
studies of the $At-tra^{NSS}$ splice variants may provide the next important clue regarding the underlying sex determination mechanism, which deviates from thus far identified mechanisms. The absence of maternally provided $At-tra^{F}$ signifies an important distinction between the MEGISD system of $N. vitripennis$ and a possible MEGISD-like system in $A. tabida$. The differences in mechanism between the three thus far examined species illustrate the wide radiation of sex determination mechanisms in the haplodiploid Hymenoptera.

**Acknowledgements**

We thank Ammerins de Haan for performing DNA extractions for the genome samples and RNA extractions for the embryos series, Rogier Houwerzijl and Peter Hes for assistance with wasp and fly culturing, Jetske de Boer for providing flowcytometry estimates of $A. tabida$ genome size, Sean de Graaf for sequencing the first fragments of $tra$ and Morris Swertz for facilitating the bioinformatic work.
Supplementary figures

**Figure S4.1.** a.) Alignment of TRA female-specific amino acid sequences. Conservation of sites is shown in grey scale. b.) Gene tree of female-specific TRA using all sites by Maximum Likelihood method based on the Jones *et al.* w/freq. model (Jones *et al.*, 1992). Bootstrap values (1000 replicates) are shown on the branches. The scale bar shows the number of substitutions per site.

**Figure S4.2.** a.) Alignment of TRA2 amino acid sequences. Conservation of sites is shown in grey scale. b.) Gene tree of TRA2 using all sites by Maximum Likelihood method based on the Le Gascuel 2008 model (Le & Gascuel, 2008). Bootstrap values (1000 replicates) are shown on the branches. The scale bar shows the number of substitutions per site.

**Figure S4.3.** a) Alignment of DSX female-specific amino acid sequences. Conservation of sites is shown in grey scale. b.) Gene tree of female-specific DSX using all sites by Maximum Likelihood method based on the JTT matrix-based model (Jones *et al.*, 1992). Bootstrap values (1000 replicates) are shown on the branches. The scale bar shows the number of substitutions per site.

**Figure S4.4.** Splice variants of At-tra in 12-14h old fertilized and unfertilized eggs.
Figure S4.1a
Figure S4.1b
Figure S4.2a

<table>
<thead>
<tr>
<th>Genus</th>
<th>Species</th>
<th>Genus</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antherina</td>
<td>tabida</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Nasonia</td>
<td>viridifacies</td>
<td>Musca</td>
<td>domestica</td>
</tr>
<tr>
<td>Arrenurus</td>
<td>echinata</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Musca</td>
<td>domestica</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Nasonia</td>
<td>viridifacies</td>
<td>Musca</td>
<td>domestica</td>
</tr>
<tr>
<td>Arrenurus</td>
<td>echinata</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Musca</td>
<td>domestica</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Nasonia</td>
<td>viridifacies</td>
<td>Musca</td>
<td>domestica</td>
</tr>
<tr>
<td>Arrenurus</td>
<td>echinata</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Musca</td>
<td>domestica</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Nasonia</td>
<td>viridifacies</td>
<td>Musca</td>
<td>domestica</td>
</tr>
<tr>
<td>Arrenurus</td>
<td>echinata</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Musca</td>
<td>domestica</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Nasonia</td>
<td>viridifacies</td>
<td>Musca</td>
<td>domestica</td>
</tr>
<tr>
<td>Arrenurus</td>
<td>echinata</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Musca</td>
<td>domestica</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Nasonia</td>
<td>viridifacies</td>
<td>Musca</td>
<td>domestica</td>
</tr>
<tr>
<td>Arrenurus</td>
<td>echinata</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Musca</td>
<td>domestica</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Nasonia</td>
<td>viridifacies</td>
<td>Musca</td>
<td>domestica</td>
</tr>
<tr>
<td>Arrenurus</td>
<td>echinata</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Musca</td>
<td>domestica</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Nasonia</td>
<td>viridifacies</td>
<td>Musca</td>
<td>domestica</td>
</tr>
<tr>
<td>Arrenurus</td>
<td>echinata</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Musca</td>
<td>domestica</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Nasonia</td>
<td>viridifacies</td>
<td>Musca</td>
<td>domestica</td>
</tr>
<tr>
<td>Arrenurus</td>
<td>echinata</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Musca</td>
<td>domestica</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Nasonia</td>
<td>viridifacies</td>
<td>Musca</td>
<td>domestica</td>
</tr>
<tr>
<td>Arrenurus</td>
<td>echinata</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Musca</td>
<td>domestica</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Nasonia</td>
<td>viridifacies</td>
<td>Musca</td>
<td>domestica</td>
</tr>
<tr>
<td>Arrenurus</td>
<td>echinata</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
<tr>
<td>Musca</td>
<td>domestica</td>
<td>Antherina</td>
<td>tabida</td>
</tr>
</tbody>
</table>

Figure S4.2a

- Graphical representation of genotypic variation across different species and genuses.
Figure S4.2b
Figure S4.3a
Figure S4.3b

![Dendrogram showing evolutionary relationships between different species.](image)

Figure S4.4

![SVG: Gel electrophoresis showing bands for fertilized and unfertilized samples.](image)