Identification and characterization of the *doublesex* gene of *Nasonia*

D.C.S.G. Oliveira  
J.H. Werren  
E.C. Verhulst  
J.D. Giebel  
A. Kamping  
L.W. Beukeboom  
L. van de Zande

Abstract
The *doublesex* (*dsx*) gene of the parasitic wasp *Nasonia* is described and characterized. Differential splicing of *dsx* transcripts has been shown to induce somatic sexual differentiation in Diptera and Lepidoptera, but not yet in other insect orders. Two splice forms of *Nasonia dsx* mRNA are differentially expressed in males and females. In addition, in a gynandromorphic line that produces haploids (normally males) with full female phenotypes, these individuals show the female spliceform, providing the first demonstration of a direct association of *dsx* with somatic sex differentiation in Hymenoptera. Finally, the DM domain of *Nasonia* DSX clusters phylogenetically with DSX from other insects, and *Nasonia dsx* shows microsynteny with *dsx* of *Apis*, further supporting identification of the *dsx* ortholog in *Nasonia*.

Chapter 2

Introduction

It has been known for a long time that Nasonia sex determination is not governed by complementary sex determiner (CSD) (Whiting, 1967; Skinner and Werren, 1980), but until recently little progress had been made in elucidating its mode of sex determination (Beukeboom, 1995). Recently, several studies reported on the genetics of sex determination in Nasonia (Beukeboom and Kamping, 2006; Trent et al., 2006), including the characterization of genetic strains that produce gynandromorphic phenotypes (Beukeboom et al., 2007a; Kamping et al., 2007). See also Chapter 1.

In all insects examined to date is doublesex the master switch gene that is responsible for male or female development (Saccone et al., 1996; Shearman and Frommer, 1998; Kuhn et al., 2000; Hediger et al., 2004; Lagos et al., 2005; Ruiz et al., 2005; Scali et al., 2005). This conservation of dsx genes is in accordance with the “bottom up evolution of sex determining pathways” hypothesis of Wilkins (Wilkins, 1995) and can be utilized to start with uncovering the sex determining mechanism of Nasonia. The recent advent of the Nasonia genome sequence (Werren et al., 2010) now provides avenues for identification of genes involved in sex determination, and to investigate their potential role in Nasonia sex determining pathways and the evolution of genetic sex determining mechanisms in general. In this Chapter the identification and characterization of a dsx ortholog in Nasonia is reported. It is shown that Nasonia dsx clusters phylogenetically with dsx in other insects, is differentially spliced in males and females, and is associated with sex determination in haploid individuals from a gynandromorphic producing strain of Nasonia. Also, the presence and sequence divergence of dsx in related species of Nasonia is investigated.

Results

Computational detection of the Nasonia dsx ortholog

A gene containing a DM-domain coding sequence that showed striking homology to dsx of other insects was originally detected during efforts to clone genes involved in sex specific wing size differences between Nasonia species. For this study, all wing size and gene expression experiments used the minimal-introggression ws1gV strain ws1gV.40kb, produced by backcrossing and selection for recombinants between wing size 1 locus (ws1) and linked visible and lethal mutants. This strain contains 40Kb of introgressed N. giraulti DNA containing ws1g in a N. vitripennis genetic background. It was constructed by backcrossing males from a minimal-recombinant strain wm114 into AsymCX for 10 generations to produce a homogeneous genetic background. An AFLP marker termed “AF1” was identified and found to be tightly linked to ws1 (Loehlin et al.,
The putative coding region of the gene containing the DM-domain coding sequence, provisionally termed \textit{Dm1}, was represented in a fully sequenced BAC of \textit{N. giraulti} (GenBank accession number AC185330). Subsequently, the \textit{N. vitripennis} genome was bioinformatically screened with the DM-domain coding sequence and, in addition to finding the \textit{N. vitripennis Dm1}, three additional DM-domain genes were detected, termed \textit{Dm2}, \textit{Dm3} and \textit{Dm4}. These four DM-domain containing genes were found in the following scaffolds: \textit{Dm1}, scaffold 23; \textit{Dm2}, scaffold 53; \textit{Dm3}, scaffold 62; \textit{Dm4}, scaffold 6 (\textit{N. vitripennis} genome version 1.0; HGSC at Baylor College of Medicine). Besides giving the strongest blast match to any known insect \textit{dsx} gene, the \textit{Dm1} gene also contains a second conserved coding sequence characteristic of DSX proteins: the \textit{dsx} dimerization domain (\textit{dsx} dimer). Only one gene containing the \textit{dsx} dimer was found in each of the five insect genomes searched (see below). Further analysis of scaffold 23, involving the immediate flanking sequences of \textit{Dm1} revealed that \textit{Dm1} is flanked upstream by a \textit{prospero} ortholog and downstream by an \textit{elongase} gene. Microsynteny of \textit{prospero} – \textit{dsx} gene order is also present in the genomes of \textit{Apis mellifera}, \textit{Tribolium castaneum}, and \textit{Anopheles gambiae}, but not in \textit{Drosophila melanogaster}.

\textbf{Figure 2.1:} DM domain gene tree. The \textit{Nasonia dsx} clusters with \textit{dsx} of other insects. The most parsimonious tree (consensus of 20 trees with 222 steps) and a neighbor-joining tree obtained using the amino acid sequences of the DM domain region (67 total characters, 42 that are parsimony-informative) are shown with bootstrap supports above the branches. Many of the DM domain genes used in these analyses are predicted gene models.
Four different DM-domain genes were also found in the genomes of D. melano-
gaster and A. mellifera, while only three were found in the genomes of T. cast-
taneum and A. gambiae. Phylogenetic analysis using amino acid sequences of
the DM domain showed four gene clusters, each containing a single gene from
each species (Fig 2.1). These phylogenetic clusters are consistent with the overall
gene architecture, as indicated by other identifiable structural markers, e.g. the
dsx dimer in the dsx cluster and the DMRTA motif (pfam 03474) in the dsx-mab
93B cluster (data not shown). The Nasonia Dm1 groups with dsx of the other
species with strong support. A basal position of the Nasonia Dm1 DM domain
is caused by a number of derived amino acid replacements that are conserved
in other species, including the honeybee (Fig 2.2). In addition to dsx, all five
species share genes homologous to dsx-mab 93B and a dsx-mab 99B genes (Dm2
and Dm4 respectively for Nasonia), the latter being the less divergent gene, in-
dicating stronger selective pressure to maintain the amino acid sequence. The
fourth gene dsx-mab 11E (DM3 for Nasonia) was not found in the genome of T.
castaneum and A. gambiae. Since most insect genomes are still in draft stage, at
this point it is difficult to be sure whether dsx-mab 11E has not been sequenced
or it was really lost during T. castaneum and A. gambiae evolution. However, it
appears that four DM domain genes is the ancestral condition in holometabolous
insects.

Gene architecture, genomic microsynteny, and phylogenetic analysis of con-
served regions of established dsx genes lead us to the conclusion that Dm1 is the
Nasonia ortholog of dsx.

Sex differential splicing of Nasonia dsx mRNA

Male and female forms of dsx were detected using primers designed for the pre-
dicted 3'UTR and the DM-domain exon. Results of these experiments show that
from female mRNA an approximately 500 bp fragment is predominantly ampli-
fied, whereas in males the amplicon using the same primer set is approximately
110 bp longer (Fig. 2.3).

The male and female splice forms were confirmed by sequencing of the 3'
RACE products generated for N. vitripennis (AsymC and HiCD12 strains) and
N. giraulti. Full length transcripts were determined for the N. vitripennis strain
HiCD12 by sequencing of the 5’ RACE products. For the second N. vitripennis
strain, AsymC, and for N. giraulti several RT-PCR reactions with different primer
combinations confirmed all intron-exon boundaries including the most 5’ exon
(Fig. 2.4).

The mature Nasonia male dsx mRNA is composed of five exons, with a single
large exon 5M of 1082 bp (hereafter sizes are for N. vitripennis). This last male
exon includes the C-terminal coding sequence, starting from the dsx dimer region
and does not undergo further splicing. The first four exons of Nasonia female
mature dsx mRNA are identical to the male isoform. The Nasonia female dsx
mRNA, however, is derived from six exons. Exons 5F and 6F are derived from the
same exact genomic region as the last male exon (exon 5M) but are interrupted
Chapter 2

Figure 2.2: Amino acid sequence alignment of the Nasonia dsx and other insects. Only the two conserved regions are presented in this Figure: the DNA binding (DM) and the DSX dimer. The zinc chelating residues in the DM domain are in bold in the reference pfam sequence (Finn et al., 2006). Putative conserved residues that distinguish the dsx DM domain from the DM domain of other proteins are shown in dotted boxes. The female isoforms were used for the dsx dimer region alignment. The predicted amino acid sequences of the 3’ region of the male isoform are shown for Nasonia and Trichomalopsis.

by an intron of 108 bp, that is not spliced out in the male mRNA, resulting in two exons of 147 bp (exon 5F) and 827 bp (exon 6F), respectively. Sex specific differences in DSX protein sequence are caused by female specific splicing of the last exon resulting in a different stop codon (see below).

To summarize, in Nasonia females an intron of 108 bp is spliced out of the primary transcript, while in males no splicing of this intron occurs (Fig. 2.4). This process of generating the male specific dsx isoform has not been described for other insects. In other insects, including A. mellifera (Cho et al., 2007), D. melanogaster (Baker and Wolfner, 1988), and B. mori (Ohbayashi et al., 2001), the male isoform is generated by an alternative splice event that completely skips the female specific exon (see Fig. 2.4).

RT-PCR conducted for different life stages (embryos, larvae, pupae, and adults), and for different adult body parts (head, thorax, and abdomen) showed that both sexes predominantly express their sex-specific form in all stages and body parts tested (Fig. 2.3B shows adults). The female form was often detected in independent experiments as a weaker band in males, whereas the male form was rarely seen in females. The presence of the female form of dsx in males was also noted in A. mellifera (Cho et al., 2007).
Interestingly, for *N. vitripennis* in both males and females a 564 bases longer 3’ UTR transcript was also identified (but not for *N. giraulti*). Apparently a second poly(A) site can be used in both males and females to generate longer *dsx* mRNA’s. Since these longer forms are observed in spliced mRNA, i.e. they do not include intronic sequences, it is assumed that these products are not the result of genomic contamination, but represent a genuine *dsx* mRNA. However, comparison of these 3’ UTR sequences to the genomic sequence of *N. vitripennis* shows that a stretch of 25 As is present in the *N. vitripennis* genomic sequence. Therefore, the actual alternative poly(A) site could be even more downstream, but it was not detected due to the annealing of the oligo-d(T) to the internal A-tract upon performing 3’ RACE. In addition a smaller transcript that skips exon 2 (Fig. 2.4), i.e. a splice-form that joins directly exons 1 and 3, was consistently observed as a clear second band in RT-PCR in both sexes. This is puzzling because exon 2 codes for the DM domain and it also has the start codon (the alternative start codon maintains the frame). Obviously, these alternative transcripts will need further characterization. It is worth noting that evidence for more than two splicing variants of *dsx* have been shown previously, e.g. for *D. melanogaster* (Baker and Wolfner, 1988) and *A. mellifera* (Cho et al., 2007).
dsx mRNA splice variants in gynandromorphs

Gynandromorphism (individuals with both male and female structures) has been previously described in Nasonia. A line (HiCD12) has been characterized that produces gynandromorphs from unfertilized (haploid) eggs which range from single female structures (e.g. antennal segment) in an otherwise male morphology to haploids with a complete female morphology (Beukeboom et al., 2007a; Kamping et al., 2007). To evaluate the involvement of dsx in hymenopteran sex determination, haploid individuals from the HiCD12 line that were either completely morphological male or female, were analyzed for the presence of male and female specific dsx splice variants. Morphological male gynandromorphs always express the male specific splice variant at a considerable level, sometimes together with the female-specific variant, whereas haploid females express mostly or only the female transcript (Fig. 2.3A). Presence of the female specific spliceform in morphological males of the gynandromorphic line may be due to presence of female-like structures in these individuals that were not apparent from their outer morphology. As stated above, however, haploid males from the non-gynandromorphic strain N. vitripennis AsymC can also contain low levels of female-specific splice variants of dsx mRNA (see Fig. 2.3B). In contrast, females of either strain show a clear predominant female-specific spliceform, confirming the female specificity of this splice variant.

The difference between a normal female and a phenotypic female from the gynandromorphic line is that the former are diploid whereas the latter are derived from haploid unfertilized eggs, yet develop somatically into females. These data clearly support an association for dsx in Nasonia sex specific phenotypic differentiation, independent of ploidy.
Predicted DSX proteins in *Nasonia* and other insects

In *Drosophila* and other organisms, DSX has two characteristic domains: a DNA binding domain (DM or OD1) and an oligomerization domain (DSX dimer or OD2). The genomic region of all coding exons of *dsx* of two additional *Nasonia* species was sequenced as well as of two *Trichomalopsis* species, a closely related pteromalid wasp. The two conserved regions of the predicted DSX proteins were aligned with known *dsx* sequences from eight other species: the hymenopteran *A. mellifera*, the dipterans *Anastrepha brisgata*, *D. melanogaster*; *An. gambiae*, *M. domestica* and *Megaselia scalaris*, the lepidopteran *B. mori*, and the coleopteran *T. castaneum* (Fig. 2.2). Parsimony and neighbor-joining dendrograms of this alignment shows that the combined protein sequence groups *Nasonia* DSX with *A. mellifera* DSX (Fig. 2.5), as expected based on phylogenetic relationships. *Apis* and *Nasonia* did not cluster together when only the DM domain was used (Fig. 2.1), reflecting the amino acid sequence divergence of this domain in *Nasonia*. The Hymenoptera combined DSX domains show longer branch lengths than found in other insects (Fig. 2.5), probably caused by an elevated divergence rate.

**Figure 2.5:** Insect species trees based on the conserved *dsx* amino acid sequences. The most parsimonious tree (consensus of 20 trees with 335 steps) and a neighbor-joining tree obtained using the combined amino acid sequences of the DM-domain region and the DSX dimer region (138 total characters, 89 that are parsimony-informative) are shown with bootstrap support above the branches.
Examination of the DM domain consensus (pfam00751, Fig. 2.2) reveals amino acids that are conserved across the DM domains shown, relative to the consensus for DM domains. Most of these conserved amino acid residues are not DSX specific, but are common to other DM domain containing proteins and are important for zinc binding (Zhu et al., 2000). However, examination of the DM domain sequences also reveals two highly conserved amino acids that are shared among all insect DSX. An insect DSX shared threonine (T) occurs four amino acids upstream of a glutamine (Q) in the second half of the DM domain (Fig. 2.2). A third highly conserved glutamic acid (E) occurs five amino acids upstream of the conserved T. Only Trichomalopsis shows a variant at this position, with arginine (R) in place of E. Both are polar amino acids with strong hydropathy indices, whereas glutamic acid is acidic and arginine is strongly basic. The clustering of these conserved amino acids suggests that they may be important in DSX protein specificity in sexual differentiation, relative to other DM domain proteins.

Figure 2.2 also shows alignment of the DSX dimer domain in Nasonia, relatives and other insects, compared to the consensus pfam domain sequence for this region. As can be seen, there is considerable variation in the terminus of this domain, in both length and sequence. Because the DSX dimer is an alpha-helical motif (Bayrer et al., 2005), structure conservation rather than amino acid conservation is expected. Nevertheless, Nasonia and relatives retain a nine amino acid sequence in the middle of the DSX dimer region (MLYVILKDA) that is also mostly conserved in Diptera and Tribolium, but not in Apis or Bombyx. D. melanogaster shows the greatest similarity to the pfam sequence, but this likely reflects the contribution of Drosophila and other Diptera to determination of this consensus.

The sequence of N. vitripennis and its sister species, N. giraulti, N. longicornis, and N. oneida, as well as two closely related wasps, Trichomalopsis sarcophagae and Trichomalopsis dubius were also compared. The species differ in length of the C terminus for the predicted female specific protein. A 16 amino acids long extension in the female specific protein appears to be the ancestral condition, as it is shared among two Nasonia species, N. vitripennis and N. longicornis, and Trichomalopsis (Fig. 2.2). In N. giraulti, the female transcript contains a stop codon that truncates the protein 12 amino acids relative to the ancestral form. A single deletion in N. oneida creates a frameshift in this tail, resulting in a 32 amino acids long extension in the females that has a completely different amino acid composition (Fig. 2.2). These results indicate accelerated evolution in the female dsx in these wasps and low purifying constraints. It is in agreement with recent work by Yang et al. (2008) showing that the C-terminal “tail” of the DSX protein in D. melanogaster does not play an important role in dimer formation or Intersex (IX) binding of the female protein, even though it is known that IX and the female form of DSX function together to repress male sexual differentiation in Drosophila (Baker and Ridge, 1980).
Discussion

In this Chapter the *Nasonia dsx*, an ortholog of the *Drosophila* transcription factor gene *doublesex* is described. Evidence that *Nasonia dsx* is the *Drosophila dsx* ortholog are (a) *Nasonia* DSX has an amino-terminus DNA binding domain (DM) that clusters phylogenetically with highly conserved DSX DM domains from other insect species, as well as an oligomerization domain (DSX dimer), (b) *Nasonia dsx* is expressed as male and female specific splice forms, (c) expression of *Nasonia dsx* in phenotypically variable gynandromorphic mutants matches their sexual phenotype differentiation, consistent with a role of *Nasonia dsx* in sex determination. Research on the sex-determining mechanism of *D. melanogaster* has shown that the transcription factor coding gene *dsx* is the major switch at the base of the sex determination cascade, which subsequently determines somatic sex determination (Baker and Wolfner, 1988; Raymond et al., 1998). It is confirmed that in *Nasonia* the *dsx* gene is conserved, both in functional domains and in sex-specific splicing, indicating a similar major switch role in *Nasonia* sex determination.

The identified *dsx* gene appears to be involved in the evolution of the morphological wing differences between male and female *N. vitripennis* (Loehlin et al., 2010). Males of *N. vitripennis* have small wings and cannot fly, while males of *N. giraulti* have large wings and do fly. The study of Loehlin et al. (2010) strongly suggests that the level of *dsx* expression dependent on its 5′ cis-acting regions is the evolutionary mechanism behind these morphological differences. Although the role of this *dsx cis*-regulation in wing size is not yet fully confirmed, it seems likely that *Nasonia dsx* is involved in sex specific morphology differences between two closely related species.

In *D. melanogaster*, males can be considered the default sex, as the male specific variant of *dsx* splicing requires no active upstream gene activity. This male default situation also seems to apply to the *Nasonia* sex determining system, because the female specific variant of *Nasonia dsx* requires an additional splicing event. However, in the lepidopteran *Bombyx*, which has female heterogamy, the female splice variant is the default due to suppression of the male splice variant (Suzuki et al., 2001). Therefore, without knowledge of upstream regulatory factors, the default sex in *Nasonia* remains unknown.

In *Drosophila*, the active specification of the sex determining cascade ultimately leading to sex-specific splicing of *dsx* has long been considered the ratio of X-chromosomes to autosomes (X:A). However, recent evidence indicates that the X-chromosome dosage rather than the X:A ratio is the key factor in sex determination in *Drosophila* (Erickson and Quintero, 2007). This dosage effect is exemplified through the timing of blastoderm formation: haploid embryos undergo an extra nuclear division cycle that prolongs the period in which X-encoded signal element proteins are expressed, whereas triploid embryos cellularize one cycle earlier than diploids. These authors recognize that sex determination in haplodiploids may work in a similar way. Crozier (1977) proposed that a chromosomal-cytoplasmatic (maternal effect) balance could be the distinguishing signal between haploid and diploid embryos. Kamping et al. (2007) found
evidence for a strong maternal effect on Nasonia sex determination. If haploid embryos develop slower than diploid embryos, a differential dosage effect of the chromosomal constitution of the embryo may tilt the balance towards males.

Sex determining mechanisms vary greatly at the chromosomal level, ranging from male heterogamety (XX-XY) to female heterogamety (ZZ-ZW) to haplodiploidy without heteromorphic sex chromosomes. The underlying genes for sex determination are organized in cascades and can also vary between closely related species or even within species (Werren and Beukeboom, 1998). Nevertheless, there is now ample evidence for evolutionary conservation of \( dsx \) at the base of these cascades in holometabolous insects. During evolution, genes appear to be added to the top of the cascade (Wilkins, 1995), but the responsible selective forces for this pathway evolution are not well understood.

It is not yet known which genes regulate alternative splicing of \( dsx \) in Nasonia, although transformer (tra), or a hymenopteran homolog, is a likely candidate. The Drosophila TRA and Transformer-2 (TRA-2) proteins bind to regulatory elements, a 13-nucleotide sequence repeated 6 times, to activate the female specific splicing of the \( dsx \) primary transcripts (Inoue et al., 1992). These regulatory elements in Nasonia have not been identified, and they seem not present in Bombyx (Ohbayashi et al., 2001). Since sex determination in Nasonia does not depend on heterozygosity, a system with the \( csd \) gene as the primary regulator of \( dsx \) can be ruled out in this species. Nevertheless involvement of a \( tra \) homolog seems probable. In the next Chapter a possible involvement of a Nasonia \( tra \) homolog in the sex determining pathway will be analyzed.

Many models for sex determination in Nasonia have been proposed (reviewed in Cook, 1993; and Beukeboom, 1995, see also Chapter 1). Sex determination in Nasonia is currently viewed as an interaction of maternal effects, chromosome dosage and genomic imprinting (Beukeboom et al., 2007b). Identification of the upstream genes to \( dsx \) in Nasonia is needed for a better understanding of its sex determination. This will also provide more insights in the evolution of sex determination in insects in general and that of Hymenoptera in particular.

**Recent developments in \( dsx \) research**

In the past few years a number of additional studies have shown the function of DSX protein in different insect species by either RNAi knockdown experiments or over-expression in a transgenic background. C. capitata male specific \( dsx \) was expressed in D. melanogaster XX flies and resulted in partial masculinization of both somatic and germline tissues (Saccone et al., 2008). A similar study was done in which Anastrepha female or male specific DSX protein was expressed in D. melanogaster intersexes lacking endogenous \( dsx \) function (Alvarez et al., 2009). Both Anastrepha sex specific isoforms led to incomplete female and male development in D. melanogaster and activated or repressed the Drosophila yolk protein genes, respectively as do the DSX proteins of Drosophila itself. In Bactrocera dorsalis RNAi knockdown of female specific \( dsx \) repressed the expression of yolk protein genes and affected the reproductive ability of the females (Chen et al.,
2008). In both cases dsx appears to regulate the expression of yolk proteins in a sex specific manner.

Recently, dsx has also been characterized in two silk moth species, Antheraea assama and A. mylitta (Shukla and Nagaraju, 2010 submitted). A. assama is considered the progenitor species of silk moths and lacks the W chromosome in females (XO), while A. mylitta has ZW females and ZZ males. Knockdown of dsx in A. assama led to an abolishment of vitellogenin and hexamerin expression in females. All these new data confirm the conservation of dsx at the bottom of insect sex determination.

Acknowledgments

We are grateful to Rachael Edwards for assistance with the laboratory work. We thank anonymous reviewers for valuable comments. The work in the John H. Werren laboratory was funded by a National Institute of Health grant, 5R01 GM070026. Work in the Leo W. Beukeboom laboratory was funded by Pioneer grant, ALW 833.02.003, from the Netherlands Organization for Scientific Research.