A comparative analysis of highly conserved sex-determining genes between *Apis mellifera* and *Drosophila melanogaster*

Alexandre S. Cristino[1], Adriana Mendes do Nascimento[2], Luciano da Fontoura Costa[3] and Zilá Luz Paulino Simões[2,4]

[1]Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, SP, Brasil
[2]Departamento de Biologia, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brasil
[3]Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP, Brasil
[4]Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Ribeirão Preto, SP, Brasil
Corresponding author: A.S. Cristino
E-mail: alexsc@rge.fmrp.usp.br

ABSTRACT. A comparison of the most conserved sex-determining genes between the fruit fly, *Drosophila melanogaster*, and the honey bee, *Apis mellifera*, was performed with bioinformatics tools developed for computational molecular biology. An initial set of protein sequences already described in the fruit fly as participants of the sex-determining cascade was retrieved from the Gene Ontology database (http://www.geneontology.org/) and aligned against a database of protein sequences predicted from the honey bee genome. The *doublesex* (*dsx*) gene is considered one of the most conserved sex-determining genes among metazoans, and a male-specific partial cDNA of putative *A. mellifera dsx* gene (*Amdsx*) was identified experimentally. The theoretical predictions were developed in the context of sequence similarity. Experimental evidence indicates that *dsx* is present in embryos and larvae, and that it

encodes a transcription factor widely conserved in metazoans, containing a DM DNA-binding domain implicated in the regulation of the expression of genes involved in sexual phenotype formation.

Key words: Sex determination, Sexual development, *Apis mellifera*, *doublesex* gene, Bioinformatics, Transcription factors, Evolution of genes, Molecular genetics

INTRODUCTION

Most metazoans express two sexual phenotypes, and the choice of sexual cell fate is a developmental process similar to the choice between becoming an epidermal cell or a neuron. One of the most completely characterized genetic regulatory hierarchies is that regulating sexual differentiation. Sex determination during development in *Drosophila melanogaster* (Diptera) is transmitted through a cascade of regulatory genes to the terminal differentiation genes, and their products are responsible for the sexually dimorphic characteristics of the adult (Baker and Ridge, 1980; Baker et al., 1987; Baker, 1989; Burtis, 1993). The initial signal at the top of the regulatory cascade shows surprising forms of solutions to switch the genetic program to produce two alternative developmental fates, a male or a female. In organisms with heteromorphic sex chromosomes, males and females can be either heterogametic (i.e., XY) or homogametic (i.e., XX), independent of the phylogenetic distance. Furthermore, some organisms have no distinct sex chromosomes; males and females are produced by a specific composition in the alleles of a single sex-determining locus or combination of loci (i.e., Hymenopteran haploid-diploid system) (White, 1973; Bull, 1983; Marín and Baker, 1998). Despite the considerable variation in the initial signal at the top of the regulatory hierarchy, sex determination shares some general features at the bottom of the pathway in all species studied elsewhere (Marín and Baker, 1998; Raymond et al., 1998; Graham et al., 2003).

The gene *doublesex* (*dsx*) at the bottom of sex-determination hierarchy of the insects *D. melanogaster* (Burtis and Baker, 1989), *Bombyx mori* (Ohbayashi et al., 2000), *Megaselia scalaris* (Kuhn et al., 2000), *Musca domestica* (Hediger et al., 2004), *Bactrocera tryoni* (Shearman and Frommer, 1998), and *Ceratitis capitata* (Graham et al., 2003) is alternatively spliced to encode a sex-specific transcription factor with a DM DNA-binding motif. Genes that encode proteins containing a DM domain are required in male development in the nematode *Caenorhabditis elegans* (*male abnormal-3*, *mab-3*) and humans (*dmrt1*), suggesting that at least some aspects of sexual regulation have a common evolutionary origin (Erdman and Burtis, 1993; Cline and Meyer, 1996; Raymond et al., 1998; Raymond et al., 2000). The regulation of yolk protein gene expression in *D. melanogaster* is the best-characterized DSX function at the molecular level. The binding of DSX-M (male-specific protein) to the fat body enhancer in the promoter region of yolk proteins (*yp-1* and *yp-2*) represses the expression of these genes, whereas binding of DSX-F (female-specific protein) to the same sequences cooperates with

other factors (BZIP-1) to activate transcription of yolk proteins in the fat body (An and Wensink, 1995a,b). Genitalia formation in flies also requires the integration of several pathways: 1) *decapentaplegic* (*dpp*), *hedgehog* (*hh*) and *wingless* (*wg*) to inform position in the genital disc, 2) the homeotic gene *abdominal-B* (*abd-B*) to define differences between genital disc segments, and 3) *dsx* to play the major role in the formation of genitalia and analia (Sanchez et al., 2001; Sánchez and Guerrero, 2001; Keisman and Baker, 2001; Keisman et al., 2001; Estrada et al., 2003; DeFalco et al., 2004). Other functional roles of *dsx* are related to the target genes responsible for sex-specific abdominal pigmentation (*bric-a-brac*, *bab*) and male courtship behavior (*takeout* and *fruitless*) (Kopp et al., 2000; Dauwalder et al., 2002; Kyriacou, 2005).

When compared to *D. melanogaster*, little is known about the molecular genetic basis of haplodiploid sex determination. About 20% of animal species are haplodiploid; in these species, unfertilized haploid eggs develop into males and fertilized diploid eggs into females. The honey bee *Apis mellifera* (Hymenoptera) is an emergent model organism, and recently a simplified model of sex determination in honey bees has been published to explain the initial signal that depends on one allele or two different alleles of a single gene, the *complementary sex determiner* (*csd*). Heterozygosity generates an active protein that initiates female development, while hemizygosity/homozygosity results in a non-active CSD protein and default male development. An orthologous *dsx* transcript has been reported to occur in honey bees, but this study has not been published yet (Beye et al., 2003; Beye, 2004).

Currently, the honey bee genome is in the fourth release, Amel_v3.0 (http://www.hgsc.bcm.tmc.edu/projects/honeybee), and it opens a new possibility for a comparative approach to functional genomics between different insect species. This can be fruitful in exploring conserved and variable genetic elements in a biological process among different species (i.e., *D. melanogaster* (*Dm*) and *A. mellifera* (*Am*)). We identified in the *A. mellifera* genome 13 genes homologous to *D. melanogaster* genes that participate in the sex determination and differentiation process. We focused preferentially on the *dsx* sequence because it is one of the most conserved genes involved in the sex determination mechanism in *D. melanogaster* and other insects. *Amdsx* encodes a putative protein with a DNA-binding motif named DM domain. Comparisons between DM-related gene family members in the fruitfly and honey bee indicated the most probable *dsx* ortholog in *A. mellifera*. Multiple sequence comparisons among available insect *dsx* genes pointed to the most probable *Amdsx* candidate. Finally, a male-specific partial *Amdsx* gene transcript was found in total RNA of the honey bee larva.

MATERIAL AND METHODS

Computational analysis

A list of annotated genes involved in sex determination (GO:0007530) and sex differentiation (GO:0007548) in *D. melanogaster* was retrieved from Gene Ontology (GO) database (http:// www.geneontology.org/) without redundant UniProt and isoform sequences. Similar *A. mellifera* sequences were searched on the predicted protein official set (ftp://beeftp: analysis@ftp.beegenome.hgsc.bcm.tmc.edu/GenePredictions/) at the Human Genome Sequencing Center of Baylor College of Medicine (HGSC-BCM) using BLASTP (with ex-

pected value less than 1e-18 and soft masking options). Only the alignments with more than 40% identity are reported in this manuscript. Version 3.0 of the whole genome shotgun sequence assembly (ftp://ftp.hgsc.bcm.tmc.edu/pub/data/Amellifera) was used to infer the partial genomic architecture of the putative *Amdsx* gene, as well as the genomic location of other hypothetical conserved genes. Insect *dsx* gene sequences were extracted from GenBank (http://www.ncbi.nlm.nih.gov/): *Bactrocera oleae* (CAD67987), *Anastrepha obliqua* (AAY25167), *B. mori* (BAB13472), *D. melanogaster* (AAF54169), *M. scalaris* (AAK38832), *M. domestica* (AAR23813), *Anopheles gambiae* (AAX48939), and *C. capitata* (AAN63597).

A pipeline was designed using Python programming tools (http://www.python.org/) and Biopython modules (http://www.biopython.org/) to identify the most closely related sequences in different proteomes of *D. melanogaster* and *A. mellifera*. The method used for identifying a group of related (gene family) *dsx* sequences between these two species was based on reciprocal best hit of alignments (Chervitz et al., 1998); only matched sequences with a very conservative statistical value (e-value < 1*e*-18) were considered.

Multiple protein sequence comparisons were generated using CLUSTALW (Thompson et al., 1994) with default parameters (gap opening 10.0, gap extension penalty 0.1). Protein distance-matrix analysis was performed with PHYLIP, version 3.64 (Felsenstein, 1989). One thousand bootstrap replications were carried out using the program SEQBOOT. Distance matrix was created using PROTDIST with PAM matrix (Dayhoff, 1979) to construct neighbor-joining (Saitou and Nei, 1987) trees using the NEIGHBOR program. The consensus tree was calculated for each set of trees produced in the program NEIGHBOR.

The DSX protein-conserved domains were predicted with Pfam Hidden Markov Models (HMM) for "DM DNA-binding domain" (Pfam accession: PF00751) and for "DMRTA motif" (Pfam accession: PF03474) (http://www.sanger.ac.uk/Software/Pfam/) using the program HMMER (the current release is 2.3.2, see http://hmmer.wustl.edu/). Only the very conserved motifs were considered, with expected values less than 1e-10.

DNA isolation, cloning, PCR amplification, and DNA sequencing

First-strand cDNA was synthesized by RT-PCR (SuperScript II; Invitrogen) from 10 µg total RNA isolated from female and male embryos and larvae. Aliquots of first-strand cDNA products were employed in PCR reactions using PCR master mix (Promega). The thermal cycling program consisted of 1 min at 95°C, followed by 35 cycles of 1 min at 95°C, 1 min at 59°C, 1 min at 72°C, and a final extension step at 72°C for 10 min. Three different primers were used in different combinations for PCR amplification, where one set was used to amplify a non-sex-specific region of the transcript (P1: 5'-TGCGAGAAGTGTAAGATCAC-3' and P2: 5'-GTGCTCCAATAGAATTTCCAC-3') and another one was used to amplify a male-specific region of the transcript (P1: 5'-TGCGAGAAGTGTAAGATCAC-3' and P3: 5'-GCACGACTA GGTTGGGACAT-3'). The amplification products were analyzed by electrophoresis on 1% agarose gels. The non-sex-specific fragments of about 443 bp and the male-specific fragments of about 560 bp corresponding to *dsx* gene partial sequence were purified and subcloned into the *Eco*RI site of pGEM-T easy plasmid (Promega). Insert-containing plasmids were subjected to sequencing reactions using the M13-reverse and M13-forward universal primers.

─────────────────────

RESULTS

Computational localization of conserved honey bee genes involved in sex determination and sex differentiation processes

Table 1 shows a total of 13 protein sequences of 29 non-redundant (only Flybase sequences) initial sequences that were retrieved from GO database annotated to be involved in sex determination (GO:0007530) and sex differentiation (GO:0007548) of the fruit fly *D. melanogaster*. All 29 sequences were aligned against the predicted honey bee protein database (Official set of predicted protein sequences from Baylor College of Medicine), and only the very conserved genes were selected (at least 40% of identity and e-value $< 1e$-18). There are three gene products commonly participating in both biological processes in the fruifly, and they are highly conserved in the honey bee genome, namely *sexlethal* (fruitfly *sxl* FBgn0003659 is 68% identical to the honey bee GB13127-PA), *intersex* (fruitfly *ix* FBgn0001276 is 43% identical to honey bee GB19364-PA) and *dsx* (fruifly *dsx* FBgn0000504 is 51% identical to honey bee GB18426-PA). Among the others, 10 genes are important elements responsible for establishment of complete sexual dimorphism, but none one of them is as well studied and known as the *dsx* gene. In the following, we describe the main functional activity of the most conserved gene, *dsx*, already studied in other insects as regulators of sexual phenotype.

The *dsx* gene encodes a highly conserved zinc-finger transcription factor with a DM DNA domain. This domain is remarkable for a novel pattern of cysteines and histidines that bind to the DNA minor groove (Erdman and Burtis, 1993; Zhu et al., 2000). The DM domain was named on the basis of its occurence in DSX (*D. melanogaster*) and MAB-3 (*C. elegans*) (Raymond et al., 1998). The DM motif is conserved in metazoan sex determination. In humans, these genes are involved in testis differentiation, where *dmrt1* and *dmrt2* deletions are associated with XY sex reversal, even given the integrity of the male-determining gene, *sex-determining region Y* (*sry*) (Sinclair et al., 1990). In zebrafish and mammals, a DM gene, *terra*, appears to function in mesodermal patterning in both sexes (Meng et al., 1999; Volff et al., 2003). The *dsx* homolog may be the most ancient one at the bottom of the sex-determining cascade in metazoans, according to the "bottom-up" hypothesis (Wilkins, 1995).

The 13 genes reported in Table 1 are involved in three different levels of molecular regulation: alternative splicing, gene transcription and signal transduction. The six transcription factors are: *dsx* (Zn-finger with a DM-binding domain), *ix* (protein domain without DNA binding), *fruitless* (*fru*; Zn-finger with protein binding), *deadpan* (*dpn*; basic helix-loop-helix dimerization region, bHLH), *dissatisfaction* (*dsf*; Zn-finger with steroid hormone receptor activity), *runt* (*run*; RNA polymerase II activity), *bab* (helix-turn-helix), and *sex combs reduced* (*scr*; RNA polymerase II activity). Three genes participate in the splicing mechanism, they are: *sexlethal* (*sxl*), *transformer2* (*tra2*), and *female lethal* (*fl*). Two genes participate in the signal transduction pathway, they are: *hopscotch* (*hop*) and *protein kinase 61C* (*pk61C*) (FlyBase - http://flybase.bio.indiana.edu/ and GO - http://www.geneontology.org/).

Except for the *ix* gene, all other genes turned out to be similar to more than one gene in the honey bee genome. It can be inferred that these homologous genes are members of gene families with at least one conserved shared domain. Four genes, *fru*, *dsf*, *bab*, and *scr*, have a remarkably large number of homologs, leading us to infer that they are members of transcription factor families that are very common in *A. mellifera* and *D. melanogaster*. Two different *dsx-*

Table 1. Conserved genes involved in two biological processes (sex determination, GO:0007530; sex differentiation, GO:0007548) between *Drosophila melanogaster* and *Apis mellifera* (at least 40% identity and e-value < 1e-18). **doublesex is the most important gene in regulating sexual development in metazoans. *genes involved in both processes.

Sex determination (GO:0007530)

Name	FlybaseID	OfficialSetID	%Identity	e-value
dpn	FBgn0010109	GB14857-PA	45.73	4e-46
dsf	FBgn0015381	GB14217-PA	46.91	4e-58
		GB14217-PA	71.79	6e-47
		GB10077-PA	43.00	7e-42
		GB10077-PA	64.52	1e-29
		GB20053-PA	74.74	1e-37
		GB17775-PA	53.64	6e-28
		GB16648-PA	47.06	9e-21
		GB18358-PA	54.32	2e-21
dsx**	FBgn0000504	GB15791-PA	44.70	4e-20
		GB18426-PA	51.06	1e-19
fl	FBgn0000662	GB13927-PA	55.64	3e-77
fru	FBgn0004652	GB17617-PA	81.20	7e-52
		GB15064-PA	51.13	2e-37
		GB11420-PA	57.98	2e-36
		GB18625-PA	41.33	2e-36
		GB14194-PA	41.62	1e-33
		GB14649-PA	43.04	1e-32
		GB16756-PA	53.85	1e-31
		GB12094-PA	50.43	2e-31
		GB18588-PA	46.21	1e-30
		GB10633-PA	52.21	1e-30
		GB14070-PA	47.33	3e-30
		GB14243-PA	49.57	2e-28
		GB19568-PA	50.00	9e-28
		GB11400-PA	42.45	9e-28
		GB19033-PA	43.23	1e-27

Sex differentiation (GO:0007548)

Name	FlybaseID	OfficialSetID	%Identity	e-value
bab	FBgn0004870	GB13762-PA	62.81	2e-62
		GB15064-PA	41.72	4e-54
		GB16756-PA	52.73	2e-42
		GB17640-PA	50.66	3e-40
		GB14194-PA	53.85	9e-39
		GB18625-PA	58.12	1e-37
		GB19033-PA	45.95	3e-37
		GB10633-PA	57.26	9e-36
		GB16366-PA	56.90	3e-35
		GB15346-PA	56.41	1e-34
		GB14649-PA	52.07	1e-33
		GB12094-PA	54.87	2e-33
		GB17617-PA	43.54	5e-31
		GB18588-PA	45.22	6e-30
		GB14319-PA	45.22	2e-25
		GB11337-PA	40.60	3e-25
		GB14696-PA	42.37	7e-22
dsx**	FBgn0000504	GB15791-PA	44.70	4e-20
		GB18426-PA	51.06	1e-19
ix*	FBgn0001276	GB19364-PA	43.23	1e-36
pk61C	FBgn0020386	GB15780-PA	51.41	6e-161
scr	FBgn0003339	GB13491-PA	47.36	1e-85
		GB13409-PA	69.41	5e-30
		GB13409-PA	82.54	5e-26
		GB18813-PA	72.62	8e-30
		GB19738-PA	67.47	3e-24
		GB11524-PA	84.48	4e-23

Table 1. Continued.

	Sex determination (GO:0007530)					Sex differentiation (GO:0007548)			
Name	FlybaseID	OfficialSetID	%Identity	e-value	Name	FlybaseID	OfficialSetID	%Identity	e-value
		GB16366-PA	41.03	1e-27	sxl*	FBgn0003659	GB18792-PA	41.50	5e-19
		GB17640-PA	46.49	7e-27			GB13127-PA	68.18	4e-70
		GB11337-PA	48.70	4e-26			GB18785-PA	46.67	2e-45
		GB17083-PA	44.35	8e-26					
		GB14319-PA	44.72	1e-22					
hop	FBgn0004864	GB16422-PA	45.42	9e-59					
		GB17556-PA	41.09	5e-19					
ix*	FBgn0001276	GB19364-PA	43.23	1e-36					
run	FBgn0003300	GB11654-PA	40.25	5e-64					
		GB16431-PA	46.69	4e-50					
		GB15836-PA	50.00	7e-50					
sxl*	FBgn0003659	GB13127-PA	68.18	4e-70					
		GB18785-PA	46.67	2e-45					
tra2	FBgn0003742	GB11130-PA	45.81	3e-47					

related genes were reported in this first analysis, and a comparison between all possible homologs in the genome of these two insects is necessary to predict the best candidate for the *dsx* ortholog.

Alignments of the DM-related family of proteins between *D. melanogaster* and *A. mellifera*

A list of eight DM-related protein sequences found in the *D. melanogaster* and *A. mellifera* databases was generated by reciprocal best hit of alignments, considering only matched sequences with a very conservative statistical value (e-value < 1*e*-18) (Chervitz et al., 1998). The phylogenetic relationship among the eight aligned products encoded by DM-related genes revealed four clusters of paralogs, in agreement with a previous study on the phylogeny of *dsx*-like genes and the evolution of sex determination (Ottolenghi et al., 2002) (Figure 1A). All related genes encode proteins with the DM domain (Pfam accession: PF00751; Figure 1B, black boxes) at the N-terminal region, but two groups of paralogs showed a conserved motif outside the DM domain at the C-terminal region termed DMRTA domain (Pfam accession: PF03474; Figure 1B, grey boxes). There are three forms of evidence that consider GB18426-PA and Dmdsx (AAF54169) as orthologs instead of GB15791-PA and Dmdsx. First, the amino acid sequences encoded by the orthologs GB18426-PA and Dmdsx (AAF54169) showed 51% identity, while the comparison between GB15791-PA and Dmdsx revealed 44.7% identity. Second, a second significant local alignment with 34% identity can be observed only between GB18426-PA and Dmdsx (not shown in Table 1 because only the best high score pair is shown). Third, only the DM domain was found in GB18426-PA and Dmdsx at the N-terminal region (e-value < 1*e*-10); two domains are present in GB15791-PA (DM and DMRTA), but the DMRTA domain is not present in the Dmdsx gene (Figure 1B). The most probable orthologs, GB18426-PA and Dmdsx, are located on chromosomes 5 and 3R, respectively. GB18260-PA and Dmdmrt93B (AAF55843) showed 44% identity and two domains (DM and DMRTA), and they are located on chromosomes 8 and 3R, respectively (Figure 1B). The most conserved orthologs, GB12040-PA and Dmdmrt11E (AAF48261), showed 78% identity and are located on chromosomes 5 and X, respectively (Figure 1B). Finally, the last clusters, GB15791-PA and Dmdmrt99B (AAF56919), are the least conserved orthologs with 39% identity. They share the DM and DMRTA domains and are located on chromosomes 1 and 3R, respectively. No synteny seems to exist between these DM genes in these insects.

Alignments with the DSX protein of insects

The orthologs of *dsx* genes in other insects were searched against the nr database at GenBank using the *Dmdsx* male isoform (AAF54169) as query. Nine insect *dsx* orthologs were aligned: *B. oleae* (CAD67987), *A. obliqua* (AAY25167), *B. mori* (BAB13472), *D. melanogaster* (AAF54169), *M. scalaris* (AAK38832), *M. domestica* (AAR23813), *A. gambiae* (AAX48939), *C. capitata* (AAN63597), and *A. mellifera* (GB18426-PA). Two highly conserved regions were evident in the alignments (Figure 2), and they have been experimentally described as functionally essential to the DSX proteins. The N-terminal conserved region includes the distinct class of an intertwined Zn-finger DNA-binding domain (DBD), that binds in
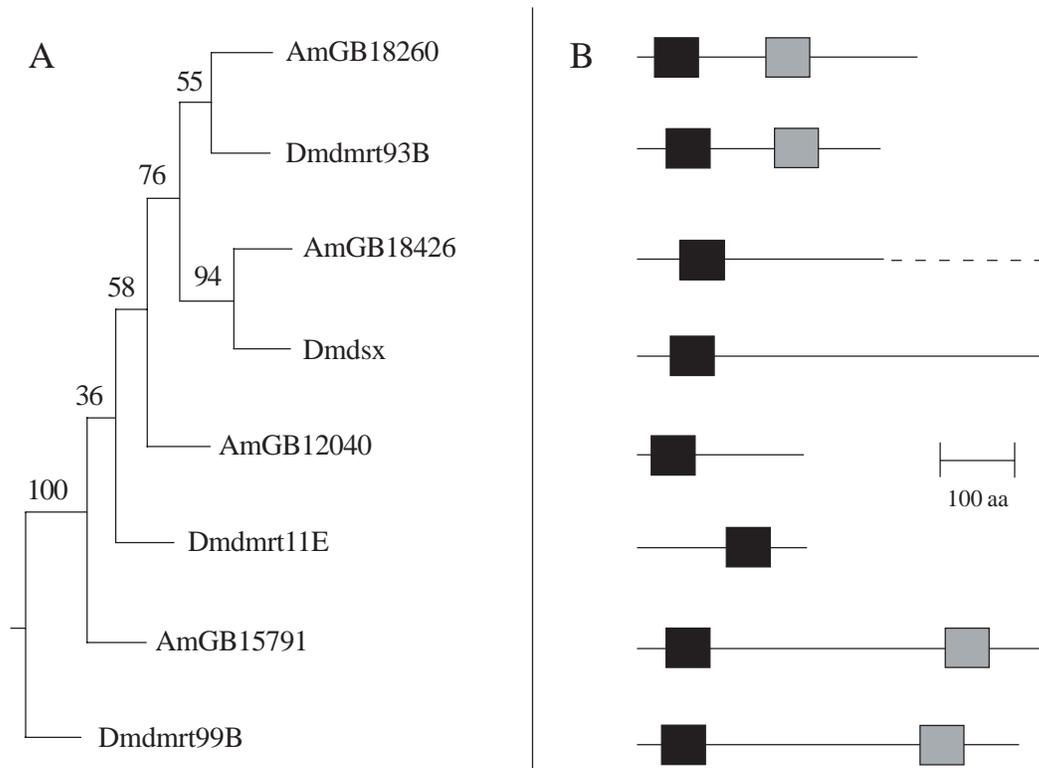
Figure 1. A. Midpoint rooted tree constructed according to the neighbor-joining method showing phylogenetic relationships of all putative homologs of DM-related proteins of *Apis mellifera* (Am) and *Drosophila melanogaster* (Dm). The bootstrap values are shown above the nodes. **B.** Schematic view of the DM-related proteins with predicted domains. Black boxes, DM domain (Pfam accession: PF00751); grey boxes, DMRTA domain (Pfam accession: PF03474); Dmdmrt99B (AAF56919), Dmdmrt11E (AAF48261), Dmdmrt93B (AAF55843), Dmdsx (AAF54169); broken line in GB18426-PA is the unknown sex-specific segment.

the minor groove of DNA (Figure 2, DBD site I and site II) (Erdman and Burtis, 1993; Zhu et al., 2000) and a non-sex-specific oligomerization domain (Figure 2, OD1) that forms dimers that bind to regulatory sites of target genes. The C-terminal region corresponds to the oligomerization domain 2 (Figure 2, OD2).

Molecular cloning of *A. mellifera dsx* cDNA

A 443-bp cDNA fragment of *Amdsx* (GenBank accession: AY375535) was first isolated from female embryonic RNA using the combination of P1 and P2 primers designed on the two most highly conserved regions of the *dsx* gene, DBD/OD1 and OD2 (Figure 3A and D). This partial sequence of *Amdsx* represents the common region of the three first exons in both sexes. A male-specific cDNA fragment of 560 bp was obtained from male larval RNA using the combination of P1 and P3 primers (Figure 3A and D). P3 primer was designed based on an exon transcribed only in males (Figure 3A and D). The putative male-specific *Amdsx* sequence was aligned against the *dsx* family of the fruit fly, and the most similar sequence was *Dmdsx* (AAF54169). Furthermore, when comparing this partial sequence against the predicted honey

DBD/OD1

site I

site II

OD2

OD2

Figure 2. Alignment of the putative DSX male protein sequences from *Bactrocera oleae* (CAD67987), *Anastrepha obliqua* (AAY25167), *Bombyx mori* (BAB13472), *Drosophila melanogaster* (AAF54169), *Megaselia scalaris* (AAK38832), *Musca domestica* (AAR23813), *Anopholes gambiae* (AAX48939), *Ceratitis capitata* (AAN63597), and *Apis mellifera* (GB18426-PA). The N-terminal conserved region, including the intertwined Zn-finger DNA-binding domain (DBD with site I, dark grey boxes and site II, light grey boxes) and a non-sex-specific oligomerization domain (OD1). The highly conserved C-terminal region corresponds to oligomerization domain 2 (OD2) and is common to DSX-F and DSX-M. The OD2 has both sex-specific and non-sex-specific sequences, which are necessary for oligomerization (only male sequences were analyzed in this study).

bee official database set, the most similar sequence was GB18426-PA (Figure 3B) as we previously reported in the phylogenetic analysis (Figure 1A). The *Amdsx* is located in Group5.6 (chromosome 5, BCM-HGSP) and has at least three non-sex-specific exons. The first exon contains the DM motif (DBD/OD1 domain), and the third exon contains the non-sex-specific segment of the OD2 domain (Figure 3C). The fourth exon was predicted by computational methods and was experimentally inferred as male-specific. This study corroborates the sources of evidence of alternative splicing mechanisms in all insects studied so far. This sex-specific alternative splicing hypothesis is strengthened by the presence of an alanine (A) in the first position of the fourth exon at the male complement of the OD2 domain (Figure 2, black arrow) while glycine (G) is the most common residue at the first position in the female-specific exon (not shown in this paper). Beye et al. (2003) have mentioned that an ortholog *dsx* transcript is present in the honey bee and encodes a putative male- and female-specific protein by alternative splicing, but no experimental evidence had been shown to date.

DISCUSSION

We made a search of 29 nonredundant genes annotated in the GO database (Ashburner et al., 2000), which pointed out 13 highly conserved genes (at least 40% identity and e-value < 1e-18) involved in sex determination (GO:0007530) and sex differentiation (GO:0007548) in *A. mellifera*. Eight of 13 genes of the conserved genes encode a variety of transcription factors (*dsx*, *ix*, *fru*, *dpn*, *dsf*, *run*, *bab*, and *scr*, see Table 1). Three of 13 conserved genes, encoding proteins that participate in the initial signal (*sxl*, *tra2* and *fl*) and are involved in the splicing mechanisms to direct the choice of sexual fate, were found to be conserved in the *A. mellifera* genome. Despite the conservation in structure, no evidence has been found concerning the functional conservation of these initial signal genes, even inside a single fly genus (*Musca*) (Meise et al., 1998). The last two conserved genes are related to cellular communication by means of signal transduction pathways (*hop* and *pk61C*). These signal transduction pathways are considered very important for the formation of genitalia and analia by means of *dpp*, *hh* and *wg* pathways (Estrada et al., 2003). All these regulators are important in sexual development in *D. melanogaster*, and some of them have been described as highly conserved in structure and they function even in distant species (Cline and Meyer, 1996; Marín and Baker, 1998; Graham et al., 2003).

The sex-determination pathways have evolved from the bottom up, as hypothesized by Wilkins (1995) and expanded by others (Marín and Baker, 1998; Schütt and Nöthiger, 2000). According to this hypothesis, the *dsx* and *ix* genes are highly conserved genes at the bottom of the sex determination hierarchy; we found that they are also conserved in the honey bee genome, at least in structure. Functional conservation for these two genes has been reported for metazoans (Raymond et al., 1998; Suzuki et al., 2003; Hediger et al., 2004; Siegal and Baker, 2005). The *ix* product can act together with DSX-F in a complex to achieve the same type of functionality as DSX-M alone, repressing or activating target genes in a female-specific manner (Siegal and Baker, 2005). DSX-F and IX form a complex that binds to the regulatory region of *yp-1* and *yp-2* to control the transcription of these female-specific genes (Coschigano and Wensink, 1993; Garrett-Engele et al., 2002).

In honey bees, the initial signal is not dependent on an X:A ratio or any sex chromosome-linked gene, but on the complementation of allelic products. In complementary sex deter-
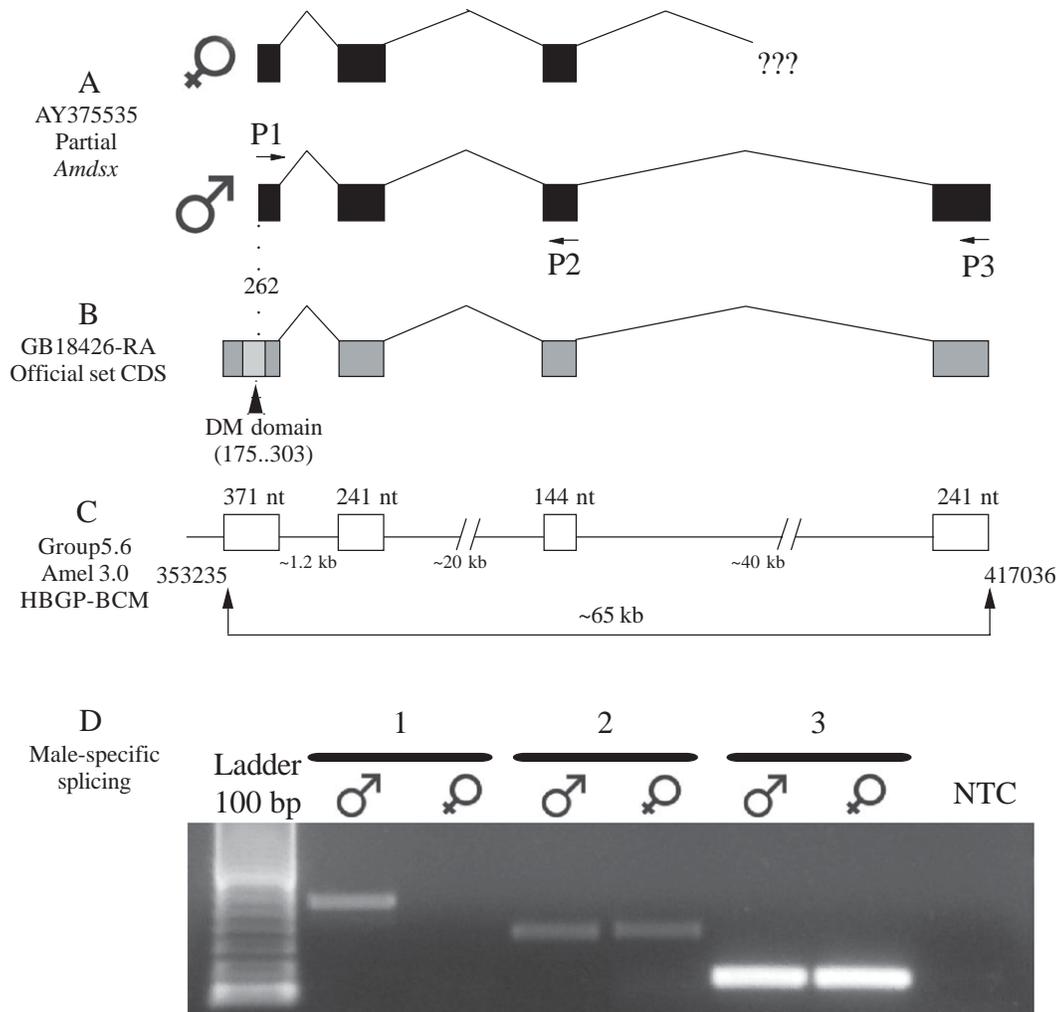
Figure 3. Schematic structure of *Apis mellifera dsx* (*Amdsx*) and the experimental evidence of male-specific splicing. **A.** A 443-bp cDNA fragment common to both sexes *Amdsx* was isolated (AY375535), cloned and sequenced from honey bee embryos. The presence of *Amdsx* female-specific exon is unknown. A 560-bp male-specific transcript form was experimentally found as predicted at GB18426-RA sequence. **B.** The prediction of DSX nucleotide sequence (GB18426-RA) showed a fourth exon that was inferred as the male-specific exon. **C.** *Amdsx* is located in Group5.6 (chromosome 5) with three common exons between sexes and the hypothetical male-specific exon. **D.** Agarose gel (1%) of RT-PCR cDNA fragment obtained by the combination of the primers: **1.** P1 and P3 as male-specific *Amdsx* splicing; **2.** P1 and P2 as positive control of *Amdsx* transcription between sexes; **3.** *EF*-a1 primers as positive control of transcription between sexes. NTC = negative control showed no DNA contamination; CDS = coding sequence; HBGP-BCM = Human Genome Sequencing Center of Baylor College of Medicine.

mination, the single allelic (hemizygous or homozygous) proteins are non-functional and males develop by default, while the combination of different alleles (heterozygosity) results in an active protein, triggering the female sexual program (Whiting, 1943; Beye et al., 2003). The *csd* gene encodes an SR-protein that acts at the top of the sex-determination hierarchy to regulate the formation of sex-specific *Amdsx* transcripts in a way very similar to that of TRA in *D. melanogaster* and *C. capitata (*Beye et al., 2003; Beye, 2004).

The *dsx* gene encodes a transcription factor containing an intertwined CCHC and HCCC Zn-finger DNA-binding domain at the N-terminal region of the protein, termed DM domain (Erdman and Burtis, 1993; Raymond et al., 1998; Zhu et al., 2000). DM-related genes are involved in sexual development and in somite development in very distant metazoan phyla (Ottolenghi et al., 2002; Volff et al., 2003). In insects, *dsx* regulates somatic sexual differentiation. It encodes two functional products (DSX-F and DSX-M) produced by alternative splicing to yield a male or female pleiotropic factor acting on several independent target genes in a sex-specific manner (Burtis and Baker, 1989; Marín and Baker, 1998). The DSX proteins bind to a DNA palindromic sequence at the regulatory region of target genes as dimers. The DNA-binding affinities of female and male DSX are indistinguishable when considering only the DBD/OD1 domain, but the sex-specific OD2 domain makes crucial contributions to form dimeric structures necessary for the repression or activation of sex-specific target genes (Erdman et al., 1996; Cho and Wensink, 1998). The sex-specific sequence of OD2 domain is related to the sex specificity of DSX interaction with the transcriptional machinery or to DNA-binding cooperativity (An et al., 1996).

In our comparative approach, four homologs of *dsx* genes were found in the *A. mellifera* genome, and these are highly conserved when compared with the fruit fly (*D. melanogaster*). The clusters of paralogs indicated the best candidate for the putative *Amdsx*, GB18426-PA. Multiple alignment of *dsx* sequences of insects showed that two main conserved domains (DM or DBD/OD1 and OD2) are present in the predicted sequence of the honey bee *dsx*. A male-specific cDNA fragment (560 bp) and a non-sex-specific cDNA fragment (443 bp) were isolated from larval and embryonic transcripts, cloned and sequenced to confirm the presence of this *dsx* ortholog in the honey bee, but no functional analysis was conducted. Since mutant experiments in honey bees are rather impracticable, knockdown by double-strand RNA (dsRNA) or RNA interference (RNAi), and *in situ* hybridization can be a reasonable solution to test the functional role of the orthologous *Amdsx (*Amdam et al., 2003; Xavier-Neto and Behringer, 2005). Choosing a set of good phenotypic markers is a very important prerequisite, and the computational methods described in this study could be a simple and efficient solution to predict an initial set for good candidates for *Amdsx* target genes through an evolutionary approach.

Of particular interest to help understand the evolution of developmental pathways are the aspects of the regulatory network topology. Considering the high degree of conservation of the DNA-binding domain in the DSX transcription factor, conserved pleiotropic effects of this protein can act as a powerful force against evolutionary change, which implies some degree of conservation at the regulatory region of essential target genes responsible for sexual development. Currently, bioinformatics has turned out to be an efficient and powerful tool to make abstractions and to formalize theoretical concepts in biology, allowing more precise predictions based on mathematical evidence and not only on descriptive diagrams.

ACKNOWLEDGMENTS

REFERENCES

Amdam GV, Simoes ZL, Guidugli KR, Norberg K, et al. (2003). Disruption of vitellogenin gene function in adult honeybees by intra-abdominal injection of double-stranded RNA. *BMC Biotechnol*. 20: 1-3.

An W and Wensink PC (1995a). Integrating sex- and tissue-specific regulation within a single *Drosophila* enhancer. *Genes Dev*. 9: 256-266.

An W and Wensink PC (1995b). Three protein binding sites form an enhancer that regulates sex- and fat body-specific transcription of *Drosophila* yolk protein genes. *EMBO J*. 14: 1221-1230.

An W, Cho S, Ishii H and Wensink PC (1996). Sex-specific and non-sex-specific oligomerization domains in both of the doublesex transcription factors from *Drosophila melanogaster*. *Mol*. *Cell*. *Biol*. 16: 3106-3111.

Ashburner M, Ball CA, Blake JA, Botstein D, et al. (2000). Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat*. *Genet*. 25: 25-29.

Baker BS (1989). Sex in flies: the splice of life. *Nature* 340: 521-524.

Baker BS and Ridge KA (1980). Sex and the single cell. I. On the action of major loci affecting sex determination in *Drosophila melanogaster*. *Genetics* 94: 383-423.

Baker BS, Nagoshi RN and Burtis KC (1987). Molecular genetic aspects of sex determination in *Drosophila*. *Bioessays* 6: 66-70.

Beye M (2004). The dice of fate: the *csd* gene and how its allelic composition regulates sexual development in the honey bee, *Apis mellifera*. *Bioessays* 26: 1131-1139.

Beye M, Hasselmann M, Fondrk MK, Page RE, et al. (2003). The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein. *Cell* 114: 419-429.

Bull JJ (1983). Evolution of sex determining mechanisms. Benjamim Cummings, Menlo Park, CA, USA.

Burtis KC (1993). The regulation of sex determination and sexually dimorphic differentiation in *Drosophila*. *Curr*. *Opin*. *Cell Biol*. 5: 1006-1014.

Burtis KC and Baker BS (1989). *Drosophila* doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell* 56: 997-1010.

Chervitz SA, Aravind L, Sherlock G, Ball CA, et al. (1998). Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science* 282: 2022-2028.

Cho S and Wensink PC (1998). Linkage between oligomerization and DNA binding in *Drosophila* doublesex proteins. *Biochemistry* 37: 11301-11308.

Cline TW and Meyer BJ (1996). Vive la difference: males *vs* females in flies *vs* worms. *Annu*. *Rev*. *Genet*. 30: 637-702.

Coschigano KT and Wensink PC (1993). Sex-specific transcriptional regulation by the male and female doublesex proteins of *Drosophila*. *Genes Dev*. 7: 42-54.

Dauwalder B, Tsujimoto S, Moss J and Mattox W (2002). The *Drosophila* takeout gene is regulated by the somatic sex-determination pathway and affects male courtship behavior. *Genes Dev*. 16: 2879-2892.

Dayhoff MO (1979). Atlas of protein sequence and structure. National Biomedical Research Foundation, Washington, DC, USA.

DeFalco T, Le BS and Van DM (2004). Abdominal-B is essential for proper sexually dimorphic development of the *Drosophila* gonad. *Mech*. *Dev*. 121: 1323-1333.

Erdman SE and Burtis KC (1993). The *Drosophila* doublesex proteins share a novel zinc finger related DNA binding domain. *EMBO J*. 12: 527-535.

Erdman SE, Chen HJ and Burtis KC (1996). Functional and genetic characterization of the oligomerization and DNA binding properties of the *Drosophila* doublesex proteins. *Genetics* 144: 1639-1652.

Estrada B, Casares F and Sanchez-Herrero E (2003). Development of the genitalia in *Drosophila melanogaster*. *Differentiation* 71: 299-310.

Felsenstein J (1989). PHYLIP: phylogeny inference package. Version 3.2. *Cladistics* 5: 164-166.

Garrett-Engele CM, Siegal ML, Manoli DS, Williams BC, et al. (2002). Intersex, a gene required for female sexual development in *Drosophila*, is expressed in both sexes and functions together with doublesex to regulate terminal differentiation. *Development* 129: 4661-4675.

Graham P, Penn JK and Schedl P (2003). Masters change, slaves remain. *Bioessays* 25: 1-4.

Hediger M, Burghardt G, Siegenthaler C, Buser N, et al. (2004). Sex determination in *Drosophila melanogaster* and *Musca domestica* converges at the level of the terminal regulator doublesex. *Dev*. *Genes Evol*. 214: 29-42.

Keisman EL and Baker BS (2001). The *Drosophila* sex determination hierarchy modulates wingless and decapentaplegic signaling to deploy dachshund sex-specifically in the genital imaginal disc. *Deve-*

lopment 128: 1643-1656.

Keisman EL, Christiansen AE and Baker BS (2001). The sex determination gene doublesex regulates the A/P organizer to direct sex-specific patterns of growth in the *Drosophila* genital imaginal disc. *Dev. Cell* 1: 215-225.

Kopp A, Duncan I, Godt D and Carroll SB (2000). Genetic control and evolution of sexually dimorphic characters in *Drosophila*. *Nature* 408: 553-559.

Kuhn S, Sievert V and Traut W (2000). The sex-determining gene doublesex in the fly *Megaselia scalaris*: conserved structure and sex-specific splicing. *Genome* 43: 1011-1020.

Kyriacou CP (2005). Behavioural genetics: sex in fruitflies is fruitless. *Nature* 436: 334-335.

Marin I and Baker BS (1998). The evolutionary dynamics of sex determination. *Science* 281: 1990-1994.

Meise M, Hilfiker-Kleiner D, Dubendorfer A, Brunner C, et al. (1998). Sex-lethal, the master sex-determining gene in *Drosophila*, is not sex-specifically regulated in *Musca domestica*. *Development* 125: 1487-1494.

Meng A, Moore B, Tang H, Yuan B, et al. (1999). A *Drosophila* doublesex-related gene, terra, is involved in somitogenesis in vertebrates. *Development* 126: 1259-1268.

Ohbayashi F, Suzuki MG, Mita K, Okano K, et al. (2000). A homologue of the *Drosophila* doublesex gene is transcribed into sex-specific mRNA isoforms in the silkworm, *Bombyx mori. Comp. Biochem. Physiol. B* 128: 145-158.

Ottolenghi C, Fellous M, Barbieri M and McElreavey K (2002). Novel paralogy relations among human chromosomes support a link between the phylogeny of doublesex-related genes and the evolution of sex determination. *Genomics* 79: 333-343.

Raymond CS, Shamu CE, Shen MM, Seifert KJ, et al. (1998). Evidence for evolutionary conservation of sex-determining genes. *Nature* 391: 691-695.

Raymond CS, Murphy MW, O'Sullivan MG, Bardwell VJ, et al. (2000). Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes Dev.* 14: 2587-2595.

Saitou N and Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.

Sanchez L and Guerrero I (2001). The development of the *Drosophila* genital disc. *Bioessays* 23: 698-707.

Sanchez L, Gorfinkiel N and Guerrero I (2001). Sex determination genes control the development of the *Drosophila* genital disc, modulating the response to Hedgehog, Wingless and Decapentaplegic signals. *Development* 128: 1033-1043.

Schütt C and Nöthiger R (2000). Structure, function and evolution of sex-determining systems in *Dipteran* insects. *Development* 127: 667-677.

Shearman DC and Frommer M (1998). The *Bactrocera tryoni* homologue of the *Drosophila melanogaster* sex-determination gene doublesex. *Insect Mol. Biol.* 7: 355-366.

Siegal ML and Baker BS (2005). Functional conservation and divergence of intersex, a gene required for female differentiation in *Drosophila melanogaster. Dev. Genes Evol.* 215: 1-12.

Sinclair AH, Berta P, Palmer MS, Hawkins JR, et al. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346: 240-244.

Suzuki MG, Funaguma S, Kanda T, Tamura T, et al. (2003). Analysis of the biological functions of a doublesex homologue in *Bombyx mori. Dev. Genes Evol.* 213: 345-354.

Thompson JD, Higgins DG and Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.

Volff JN, Zarkower D, Bardwell VJ and Schartl M (2003). Evolutionary dynamics of the DM domain gene family in metazoans. *J. Mol. Evol.* 57: S241-S249.

White M (1973). Animal cytology and evolution. Cambridge University Press, Cambridge, UK.

Whiting PW (1943). Multiple alleles in complementary sex determination of habrobracon. *Genetics* 28: 365-382.

Wilkins AS (1995). Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. *Bioessays* 17: 71-77.

Xavier-Neto J and Behringer RR (2005). Developmental biology takes on a Latin American rhythm. *Cell* 122: 1-6.

Zhu L, Wilken J, Phillips NB, Narendra U, et al. (2000). Sexual dimorphism in diverse metazoans is regulated by a novel class of intertwined zinc fingers. *Genes Dev.* 14: 1750-1764.

_____