The Fruitless gene in Nasonia displays complex sex-specific splicing and contains new zinc finger domains
Bertossa, Rinaldo; Jacobus Mgn Van De Zande, Louis; Beukeboom, Leonardus; Uyenoyama, Marcy

Published in:
Molecular Biology and Evolution

DOI:
10.1093/molbev/msp067

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:
2009

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.
The transcription factor Fruitless exerts a broad range of functions during Drosophila development, the most apparent of which is the determination of sexual behavior in males. Although fruitless sequences are found in other insect orders, little is known about fruitless structure and function outside Diptera. We have performed a thorough analysis of fruitless transcripts in the haplo-diploid wasp Nasonia vitripennis and found both sex-specific and non–sex-specific transcripts similar to those found in Drosophila. In Nasonia, however, a novel, large fruitless transcript is present in females only. Putative binding sites for sex-specific splicing factors found in Nasonia fruitless and doublesex as well as Apis mellifera doublesex transcripts were sufficient to identify a corresponding female-specific fruitless exon in A. mellifera, suggesting that similar factors in both hymenopteran species could be responsible for sex-specific splicing of both genes. Furthermore, new C2H2 zinc finger domains found in Nasonia fruitless transcripts were also identified in the fruitless locus of major holometabolous insect species but not in drosophilids. Conservation of important domains and sex-specific splicing in Diptera and Hymenoptera support the hypothesis that fruitless is an ancient gene and has conserved functions in insects. Considerable divergences in other parts of the gene are expected to underlie species-specific differences and may help to explain diversity observed in insect sexual behaviors.

**Introduction**

Fruitless (Fru) is a transcription factor that exerts multiple functions during Drosophila development. Because of its unique role in the determination of male sexual behavior, Fru is becoming an epitome for understanding the genetic determination of complex behaviors (Baker et al. 2001; Billeter et al. 2002; Billeter, Rideout, et al. 2006; Yu and Dickson 2006; Yamamoto 2007). The fru locus in Drosophila produces several transcript variants through the use of six different promoters and extensive alternative splicing (Davis and Hiroki 2001). Most transcripts begin with a sequence coding for the broad complex, tramtrack, bric-a-brac (BTB) domain, a highly conserved protein–protein interaction module in Fru proteins (Stogios et al. 2005), and end with either one of four zinc fingers coding exons (Ito et al. 1996; Ryner et al. 1996; Goodwin et al. 2000; Usui-Aoki et al. 2000; Anand et al. 2001; Billeter, Villella, et al. 2006). Although comparative analyses of fru may be crucial to understand the evolution of sexual behavior in insects, its analysis in other species is only incipient (Bertossa 2005).

In Drosophila melanogaster, functions inherent to male sexual behavior are mediated by transcripts derived from the most distal fru promoter (P1). In females, Transformer (Tra) and Transformer-2 (Tra-2)—two factors belonging to the sex-determining cascade in Drosophila (Billeter, Rideout, et al. 2006)—splice P1 derived primary transcripts into female-specific transcripts and, in addition, prevent their translation (Ryner et al. 1996; Heinrichs et al. 1998; Usui-Aoki et al. 2000). In males, due to the absence of Tra, default splicing produces male fru transcripts (fruM), each ending with either one of three different C2H2 zinc finger pairs, translated into equivalent male Fru protein isoforms (collectively referred to as FruM). As a consequence, mutations in sex-specific ffru transcripts affect nearly all aspects of male sexual behavior but have no effect on female behavior (Ito et al. 1996; Ryner et al. 1996; Villella et al. 1997; Anand et al. 2001; Lee and Hall 2001; Lee et al. 2001; Demir and Dickson 2005). FruM isoforms are also necessary for the development of the muscle of Lawrence (MOL), an abdominal muscle present only in wild-type Drosophila males (Villella et al. 1997; Usui-Aoki et al. 2000; Anand et al. 2001; Billeter, Villella, et al. 2006), and for the development of aggressive behavior (Vrontou et al. 2006).

Transcripts derived from other studied promoters (P2–P4) are present in both sexes (Anand et al. 2001; Song et al. 2002; Dornan et al. 2005). Some of them are detected already in early embryonic stages (Song et al. 2002; Dornan et al. 2005) and mediate the correct development of neuronal tissues (Anand et al. 2001; Song et al. 2002). Other transcripts appear later in time and are necessary for the differentiation of imaginal-disc derivatives, such as legs and wings (Anand et al. 2001; Dornan et al. 2005).

This brief overview gives an idea of the complexity of frru transcriptional regulation and the functions it mediates, the most intriguing of which is probably the determination of courtship behavior in adult Drosophila males. As other transcription factors containing BTB and C2H2 zinc finger domains (henceforth referred to as BTB–C2H2 genes) in insects (Spokony and Restifo 2007), frru was likely present, albeit possibly in a simpler form, before the radiation of major insect orders. It is hence possible that some frru functions may be conserved in other insects, as found for fru in the mosquito Anopheles gambiae (Gailey et al. 2006). Fru BTB domains and some zinc fingers have been identified in several insect orders (Davis et al. 2000; Ustinova and Mayer 2006), but frru transcripts architecture and sex-specific splicing have not been investigated. The parasitoid wasp Nasonia vitripennis (Hymenoptera: Pteromalidae) is quickly emerging as a model system in many areas of biology (Beukeboom and Desplan 2003; Pultz and Leaf 2003; Lynch et al. 2006). Its haplo-diploid genetics (diploid females arise from fertilized and haploid males from unfertilized eggs), sequenced genome, and complex stereotyped male courtship behavior (e.g., Beukeboom and van den Assem 2001), make this organism particularly suited for...
studying the evolution of sex-determining mechanisms and behavior in insects (Beukeboom et al. 2007). For these reasons, we set out to analyze fru transcripts expressed in male and female N. vitripennis wasps. cDNA libraries specifically constructed for this purpose revealed the presence of more than 20 different fru transcript classes in the heads of adult N. vitripennis wasps. Conservation in Nasonia fru transcripts architecture and sex-specific splicing and conservation of Fru domains in major holometabolous insect orders suggest that also fru functions are likely conserved in most holometabolous insects and provide a framework to understand fru’s role in the evolution of sexual behavior in insects.

Materials and Methods

Animal Stocks and Rearing

The laboratory inbred strain N. vitripennis AsymC was used for all experiments. This wild-type line collected in The Netherlands was cured from Wolbachia infection and maintained in the laboratory since 1971 (van den Assem and Jachmann 1999). Wasps were reared in mass culture vials (70 × 20 mm) at 25 °C, constant light, and around 45% relative humidity. To maintain the line, about 25–30 wasps (females with some males) were transferred to new vials containing about 50 Calliphora sp. fly pupae, on which N. vitripennis females parasitize. After 14 days (at 25 °C), the eclosing progeny was rehosted on fresh pupae. Calliphora flies were obtained as last instar larvae from a commercial manufacturer (Kreikamp & zn, Hoevelaken, The Netherlands). After pupation at room temperature in the laboratory, fly pupae were maintained at 4 °C and used within 3–4 weeks. Apis mellifera bees were received as adults from W. Boot and J. Calis at Wageningen University, The Netherlands.

DNA/RNA Extraction, cDNA Synthesis, and Northern Blots

Animals were frozen in liquid nitrogen and stored at −80 °C prior to RNA extraction. Where necessary (as for the preparation of the cDNA libraries from heads), individual body parts were dissected on liquid nitrogen to prevent RNA degradation. Total RNA was extracted with RNAeasy kit (Qiagen, Hilden, Germany). PolyA+ RNA was purified from total RNA with Poly(A)Purist MAG kit (Ambion, Austin TX). RNA and mRNA were stored at −80 °C either in RNase-free MilliQ water or THE RNA Storage Solution (Ambion) until use. Typically, 100–500 ng polyA+ RNA or 1–3 μg total RNA was used for cDNA synthesis. cDNA was prepared with the SuperScriptII Reverse Transcriptase kit (Invitrogen, Carlsbad, CA). Before use in polymerase chain reaction (PCR) reactions, cDNA was treated with 2 units Ribonuclease H (Fermentas, Burlington, Canada) and diluted 1:10 in MilliQ water. For northern blots, poly-A+ RNA purified from heads was size fractionated on a 1.5% agarose gel containing 6.5% formaldehyde, blotted onto Hybond-N nylon membranes (Amersham Biosciences, Piscataway, NJ) and hybridized with 32P-labeled DNA probes prepared according to standard protocols (Ausubel et al. 2003). DNA was extracted from animals following standard protocols (Ausubel et al. 2003).

Construction and Screening of Sex-Specific cDNA Libraries

cDNA libraries were prepared with the SuperScript Choice System for cDNA Synthesis kit (Invitrogen) following manufacturer’s instructions. PolyA+ RNA purified from heads of 1- to 2-day-old adult N. vitripennis males and females was reverse transcribed with oligo(dT) and random hexamer primers in separate reactions (2.5 μg polyA+ RNA for each reaction) and pooled after second strand synthesis. Double-stranded cDNAs were cloned into pSMART-cDNA plasmids and electroporated into Lucigen’s E. cloni cells following the manufacturer’s instructions (Lucigen, Middleton, WI). cDNA libraries were screened according to standard techniques (Ausubel et al. 2003).

Quantitative Real-Time PCR (qPCR)

qPCR reactions were based on SYBR green chemistry and performed with the qPCR GreenMaster with ROX kit (Jena Bioscience, Jena, Germany) on the Applied Biosystems 7300 Real-Time PCR System (Applied Biosystems). Three independent template cDNAs per sex were analyzed. Each cDNA was synthesized from total RNA extracted (see above) from 20 individually collected heads of males or female wasps (1–2 days old). After reverse transcription (see above), cDNA was treated with Ribonuclease H and diluted 1:10 in MilliQ water prior to use in qPCR reactions. Primers used, designed on different exons (or exon–exon junction) to prevent genomic DNA amplification, are listed in supplementary table 1, Supplementary Material online. Primer combinations producing a single dissociation product were used. PCR efficiency was determined for each sample and gene by LinRegPCR v7.5 as described in Ramakers et al. (2003). Mean amplicon efficiency and individual Ct values were used to calculate relative expressions with the formula “expression = (mean efficiency)^−Ct,” and divided by the
P1-f male average expression (calibrator). Expression levels were normalized by normalization factors calculated by GeNorm v3.5 according to Vandesompele et al. (2002). Data were analyzed by analysis of variance with SPSS v14.0, and for each gene, a t-test was used to assay significant expression differences between the sexes (nominal \( P < 0.05 \)).

Phylogenetic Trees, Sequence Management, and Short Repeats Search

Sequences corresponding to fru domains and domains of other BTB–C2H2 family members presented in phylogenetic trees were obtained from publicly available databases (see supplementary table 2, Supplementary Material online, for details) by searching directly for annotated genes or through TBlastN and BlastP searches (Altschul et al. 1990). In cases in which no or partial sequences were annotated, missing domains were predicted using FGENESH (see supplementary table 2, Supplementary Material online, for details) using already known protein sequences as template. Accession numbers and genomic coordinates of each sequence are given in supplementary table 2, Supplementary Material online. Phylogenetic trees, based on nucleotide sequences aligned on their coded amino acid sequences through MEGA 4.0 (Tamura et al. 2007), were obtained using Bayesian (MrBayes 3.1.2, Ronquist and Huelsenbeck 2003) and parsimony (PAUP 4.0b2, Wilgenbusch and Swofford 2003) frameworks. The optimal model for sequence evolution for the Bayesian analyses, determined with Modeltest 3.7 (Posada and Crandall 1998) under the Akaike information criterion, was general time reversible model plus gamma. Two parallel runs and four Monte Carlo Markov Chains searches per run were performed for \( 5 \times 10^6 \) generations. Burn-in, determined visually by plotting likelihood versus generation, was set at \( 10^6 \) generations. Support for parsimony trees was calculated by bootstrap analysis (1,000 replicates).

Sequence data management was based on a free version of Vector NTI Advance 10 (Invitrogen). Alignments in supplementary figure 3, Supplementary Material online, were performed with ClustalW at http://www.ebi.ac.uk/Tools/clustalw/index.html. Direct repeat search was performed by Dot Plot Matrix analysis on Vector NTI Advance 10.

Results

Architecture of fru Locus and Transcripts in Nasonia is Conserved

Degenerate PCR on DNA extracted from adult Nasonia wasps was used to obtain the sequence of the BTB coding exon of an N. vitripennis fruitless (fru) ortholog. This sequence was used to screen male and female cDNA libraries prepared from polyA+ RNA purified from heads of adult Nasonia wasps. 5’ and 3’ RACE reactions were carried out to identify 5’ and 3’ transcripts ends. All N. vitripennis fru transcript sequences are stored at NCBI (http://www.ncbi.nlm.nih.gov/).

Of 65 fru positive clones obtained from each library, 42 female and 21 male clones contained inserts coding for full-length Fru open reading frames (ORFs) (table 1). The architecture of the fru locus in N. vitripennis (fig. 1) was determined by aligning transcript sequences to the genome sequence of N. vitripennis (http://www.hgsc.bcm.tmc.edu/projects/nasonia/). Nasonia fru transcripts isolated from the cDNA libraries derive from at least six different promoters. All begin with exons P that are spliced to the BTB exon and, through connecting exons C1 and C2, end with either one of five different exons each coding for a different C2H2 zinc finger pair (fig. 1).

The BTB domain found in N. vitripennis fru transcripts (supplementary fig. 1A, Supplementary Material online) is encoded by a single exon, in contrast to D. melanogaster where it is encoded on three different exons (Davis and Hiroki 2001). Connecting exons C1 and C2 are homologous to corresponding connecting exons in Drosophila (Davis and Hiroki 2001) but have a diverged sequence (supplementary fig. 2A, Supplementary Material online). The three C2H2 zinc finger domains2 A, B, and C found in D. melanogaster’s Fru proteins (e.g., Anand et al. 2001)—elsewhere referred to as A, E, and B (Usui-Aoki et al. 2000)—are also present in translated Nasonia fru transcripts (figs. 1 and 2). The highest homology (98% similarity and identity) is found for zinc finger (Znf) B, followed by C and A, with 94% and, respectively, 63% similarity to corresponding D. melanogaster Fru domains (supplementary fig. 3, Supplementary Material online). Importantly, two new exons termed F and G and coding for C2H2 zinc finger domains as well as the Znf D coding exon (Dorman et al. 2005) were identified in the Nasonia fru locus (figs. 1 and 2). However, only Znf F was present in fru transcripts isolated from cDNA libraries (table 1).

Because the “zinc finger domain” in BTB–C2H2 genes is generally composed of two consecutive C2H2 zinc fingers, if not otherwise indicated, for simplicity the notion “zinc finger” or zinc finger domain referred to Fru zinc fingers (e.g., Znf B) and C2H2 zinc fingers, in general, indicates always the sequence encompassing both zinc fingers.
Gain and Loss of fru Zinc Finger Domains during Evolution

The presence of two new C2H2 zinc finger domains in the Nasonia fru locus but missing in D. melanogaster called for a deeper analysis to possibly reveal their evolutionary origin. To that end, a phylogenetic analysis with fru zinc finger nt sequences collected from different databases was performed. Only C2H2 zinc fingers in proximity (within 50 kb) of putative fru BTB domains, consistent with both being part of a unique transcriptional unit, and based on the present phylogenetic reconstruction, were considered as true fru zinc fingers. BTB sequences, on the other hand, were considered as bona fide fru BTB domains based on a phylogenetic analysis containing also BTB sequences of other BTB–C2H2 insect genes (supplementary fig. 1B, Supplementary Material online). This search led to the identification of putative fru loci in species representing most important holometabolous orders as well as in a hemipteran, the human body louse (Pediculus humanus corporis), and of a fru BTB domain (but no zinc finger) in the crustacean Daphnia pulex.

Bayesian and maximum parsimony trees were based on nt sequences aligned on their coded amino acids (around 50 aa) encompassing the double C2H2 zinc finger sequence (supplementary fig. 3, Supplementary Material online). Because a similar analysis indicated that Fru zinc fingers are highly related to corresponding domains in other BTB–C2H2 gene families (Spokony and Restifo 2007), a representative zinc finger sequence from each one of three other BTB–C2H2 genes (broad complex [br], tramtrack [ttk], and longitudinalis lacking [lola]) from different insect species were initially included in the analysis to serve as outgroups. This first analysis (data not shown) indicated that fru and ttk C2H2 zinc fingers are related, as shown by Spokony and Restifo (2007). However, a unique tree topology accommodating all zinc fingers, in particular Znf D (which is not of C2H2...
type), could not be inferred. This was caused in particular by two zinc fingers, one from *A. mellifera* and one from *P. humanus corporis*, which were both called H because of their similarity (supplementary fig. 3, Supplementary Material online). Znf H clustered differently in every analysis and disturbed the formation of other clusters as well. For this reason, in order to resolve the relationship of major *fru* zinc fingers among themselves, these two zinc fingers and zinc fingers of other BTB-C2H2 gene families were omitted, whereas Znf D was used as outgroup.

**FIG. 2.** *fru* zinc fingers phylogenetic tree and *fru* locus architectures. (A) Maximum parsimony tree of nucleotide sequences of *fru* zinc fingers from various insect taxa. Bayesian posterior probabilities and nonparametric bootstrap proportions (1,000 replicates) are shown above and below the branches, respectively. Znf D is set as outgroup. Support for Znf G clade as a sister branch to Znfs B and F is missing; instead, the Bayesian analysis puts Znf G clearly together with Znfs A and C (see main text). Znfs H are omitted.

(B) Genomic *fru* locus architecture in representative species of important neopteran insect orders. On the left, the phylogenetic relationship among species is indicated (species tree reproduced and modified with permission from Savard et al., @ 2006, Genome Research, Cold Spring Harbor Laboratory Press). In the center, the locations of zinc finger coding exons within 60 kb of genomic sequence are indicated by the corresponding letters (BTB coding exon/s is/are indicated with “ ’ ”). In *Chorthippus brunneus*, the genomic sequence between the BTB and Znf C coding exons is unknown. When known, the presence of *fru* sex-specific splicing is indicated (to the right). Full species names are *Chorthippus brunneus*, *Pediculus humanus corporis*, *Apis mellifera*, *Nasonia vitripennis*, *Tribolium castaneum*, *Bombyx mori*, and *Anopheles gambiae*. Drosophilids are *Drosophila melanogaster*, *Drosophila simulans*, *Drosophila polychaeta*, *Drosophila ananassae*, *Drosophila grimshawi*, *Drosophila virilis*, and *Drosophila erecta*.
In conclusion, the phylogenetic analyses indicate that the ancestral *fru* locus could have contained Znf D and two other C2H2 zinc finger types. These latter duplicated very early to give origin to B and F one, and to ancestral A and C the other. An additional early duplication of the ancestral A could have originated G and the present A. Conservation of highly similar zinc fingers but loss of otherwise ancestral zinc fingers suggests that some zinc finger classes may have more vital and conserved functions, whereas other zinc fingers are less constrained and so could acquire new functions or even disappear.

Diverse *fru* Sex-Specific Splicing in *Nasonia*

Sexually dimorphic transcripts isolated from *Nasonia* heads begin with exons P0 and P1. In P1 derived male transcripts (fig. 1, P1-m), exon P1 is spliced 418 nt downstream of the transcription start site (+1) and connects to the BTB exon (fig. 3). Conceptual translation of P1-m transcripts in frame with the Fru ORF results in the addition of 16 aa N-terminally to the first Fru start codon (methionine) on the BTB exon (supplementary fig. 1A, Supplementary Material online). In P1 derived female transcripts (fig. 1, P1-f), exon P1 is spliced 2,241 nt downstream of the male splice site and connects also to the BTB exon. Conceptual translation of P1-f in frame with the Fru ORF results in truncated protein sequences due to stop codons present in the sequence (fig. 3). The longest Fru ORF begins with the first methionine on the BTB exon (supplementary fig. 1A, Supplementary Material online). Based on the sequences isolated from the cDNA libraries, P1-f transcripts end with either exon A, B, or C, P1-m transcripts end with either exon B or C (table 1).

Sex-specific splicing of P1 derived transcripts in *Nasonia* was confirmed by RT-PCR and northern blot analyses (fig. 4). RT-PCR on cDNA synthesized with polyA+ RNA purified from *Nasonia* heads with primer combinations b and j (fig. 1) resulted in a 470-bp product in males and a 2.7-kb band in females (fig. 4Aa). The longer product in females is due to the longer P1 sequence in P1-f transcripts (figs. 1 and 3). This was further confirmed by RT-PCR using primer combinations c and j (figs. 1 and 4Ab). Northern blot of polyA+ RNA extracted from heads of *Nasonia* adults and hybridized with a P1-m derived probe (fig. 3, P1-m) resulted in one predominant band in both males and females (fig. 4B, P1-m). The bands likely correspond to same-sized bands appearing on a northern blot of the same polyA+ RNA hybridized with a probe targeting the BTB exon (fig. 4B, BTB: corresponding bands in BTB and P1-m blots are indicated by black arrowheads). Similar northern blot assays in *D. melanogaster* show a clear distinction in size between the three different sex-specific transcript classes according to the zinc finger exon they contain (cf. fig. 1D/P1 in Billeter, Villella, et al. 2006). The fact that in *Nasonia* the same assay resulted in one prevailing band in both sexes (fig. 4B, P1-m) could be explained by transcripts of different transcript classes having nearly the same size or simply because some transcript classes are more abundantly expressed than others in the head of *Nasonia* wasps. Male and female P1 transcripts on the northern blot of figure 4B could hence represent different transcript classes (e.g., P1-m transcripts ending with exon C and P1-f transcripts ending with exon B). This would also explain the

In figure 2A, one of two highly similar most parsimonious trees obtained is shown together with Bayesian posterior probability and nonparametric bootstrap values. Minimal differences between the trees concern only the relationship among some drosophilids within the Znfs B cluster. The analysis shows that virtually all *fru* zinc fingers (D, G, A, B, F, and C) were likely present in a common ancestor of all holometabolous insects. In fact, all these zinc finger types are present in major neopteran taxa (fig. 2B), form very well-supported clusters, but do not seem to be derived from one another. On the contrary, there is clear indication for only three or two ancestral *fru* C2H2 zinc finger domains. In fact, Znfs A and C likely originate from a common ancestor as do Znfs F and B.

Besides *Nasonia*, Znf G was additionally identified in *A. mellifera* and *Bombyx mori* and together form a well-supported cluster. Although in the tree of figure 2A Znf G is related to B and F, support for a unique cluster containing Znfs B, F, and G is missing from both phylogenetic analyses. Instead, the Bayesian analysis puts Znf G, always as a distinct branch, located to B and F, and G is missing from both phylogenetic analyses. In other species, other species have either Znf G or A, suggesting that these two zinc fingers may have redundant functions. In synthesis, therefore, there is clear evidence for only three or two ancestral *fru* C2H2 zinc finger domains. In fact, Znfs A and C likely originate from a common ancestor as do Znfs F and B.

The origin of Znf D, which is of the TTF type (smart00597, Marchler-Bauer et al. 2007) and not of the C2H2 type, could not be established. Because it is present in almost all insect species analyzed and in *Nasonia* as well as *D. melanogaster*, it is coded by a homologous exon located between both *fru* connecting exons (fig. 1); Znf D arose probably very early in the *fru* locus and has possibly a more basic function. This hypothesis, however, remains tempered by the finding that no Znf D is found in the *fru* locus of *A. mori*.

A final important note is that C2H2 zinc finger domains found in *fru* loci have all the same consensus sequence CX3CX2HXHX2CX3CX4-RXXDXXHXXH, where C and H are cysteines and histidines, respectively, typical of C2H2 zinc finger domains, whereas X are nonconserved amino acids. If both zinc finger H sequences are excluded, three additional residues (underlined) result conserved in all Fru zinc fingers: CX3CX2XXVXXHXXHXX3CX4-RXXDXXHXXKH. As for other zinc finger domains in BTB-C2H2 insect genes, however, only the second C2H2 zinc finger matches the classical consensus sequence CX1-5CX12HX14-H/C (Marchler-Bauer et al. 2007). The remarkable conservation among all Fru C2H2 zinc fingers adds support to the idea that all probably originated from fewer zinc finger domains with the same consensus sequence.

In conclusion, the phylogenetic analyses indicate that the ancestral *fru* locus could have contained Znf D and two other C2H2 zinc finger types. These latter duplicated very early to give origin to B and F one, and to ancestral A and C the other. An additional early duplication of the ancestral A could have originated G and the present A. Conservation of highly similar zinc fingers but loss of otherwise ancestral zinc fingers suggests that some zinc finger classes may have more vital and conserved functions, whereas other zinc fingers are less constrained and so could acquire new functions or even disappear.

Diverse *fru* Sex-Specific Splicing in *Nasonia*

Sexually dimorphic transcripts isolated from *Nasonia* heads begin with exons P0 and P1. In P1 derived male transcripts (fig. 1, P1-m), exon P1 is spliced 418 nt downstream of the transcription start site (+1) and connects to the BTB exon (fig. 3). Conceptual translation of P1-m transcripts in frame with the Fru ORF results in the addition of 16 aa N-terminally to the first Fru start codon (methionine) on the BTB exon (supplementary fig. 1A, Supplementary Material online). In P1 derived female transcripts (fig. 1, P1-f), exon P1 is spliced 2,241 nt downstream of the male splice site and connects also to the BTB exon. Conceptual translation of P1-f in frame with the Fru ORF results in truncated protein sequences due to stop codons present in the sequence (fig. 3). The longest Fru ORF begins with the first methionine on the BTB exon (supplementary fig. 1A, Supplementary Material online). Based on the sequences isolated from the cDNA libraries, P1-f transcripts end with either exon A, B, or C, P1-m transcripts end with either exon B or C (table 1).

Sex-specific splicing of P1 derived transcripts in *Nasonia* was confirmed by RT-PCR and northern blot analyses (fig. 4). RT-PCR on cDNA synthesized with polyA+ RNA purified from *Nasonia* heads with primer combinations b and j (fig. 1) resulted in a 470-bp product in males and a 2.7-kb band in females (fig. 4Aa). The longer product in females is due to the longer P1 sequence in P1-f transcripts (figs. 1 and 3). This was further confirmed by RT-PCR using primer combinations c and j (figs. 1 and 4Ab). Northern blot of polyA+ RNA extracted from heads of *Nasonia* adults and hybridized with a P1-m derived probe (fig. 3, P1-m) resulted in one predominant band in both males and females (fig. 4B, P1-m). The bands likely correspond to same-sized bands appearing on a northern blot of the same polyA+ RNA hybridized with a probe targeting the BTB exon (fig. 4B, BTB: corresponding bands in BTB and P1-m blots are indicated by black arrowheads). Similar northern blot assays in *D. melanogaster* show a clear distinction in size between the three different sex-specific transcript classes according to the zinc finger exon they contain (cf. fig. 1D/P1 in Billeter, Villella, et al. 2006). The fact that in *Nasonia* the same assay resulted in one prevailing band in both sexes (fig. 4B, P1-m) could be explained by transcripts of different transcript classes having nearly the same size or simply because some transcript classes are more abundantly expressed than others in the head of *Nasonia* wasps. Male and female P1 transcripts on the northern blot of figure 4B could hence represent different transcript classes (e.g., P1-m transcripts ending with exon C and P1-f transcripts ending with exon B). This would also explain the
larger band in females, because the additional 2,241 nt of the female P1 sequence in P1-f transcripts cannot alone explain the size difference between male and female P1 transcripts observed in the P1-m northern blot.

Three female cDNA clones contained fru transcripts coding for a full-length Fru ORF starting on P0 exons and ending with exon F (fig. 1, and supplementary fig. 2B, Supplementary Material online). Female specificity

Repeats found also in N. vitripennis doublesex as well as Apis mellifera doublesex and fruitless (bold and underlined) female exons are boxed (cf. fig. 6).

The sequence corresponding to a northern blot probe (underlined and labeled) is indicated (cf. fig. 4B).

Three female cDNA clones contained fru transcripts coding for a full-length Fru ORF starting on P0 exons and ending with exon F (fig. 1, and supplementary fig. 2B, Supplementary Material online). Female specificity
of P0 derived fru transcripts (termed P0-fru) was further confirmed by RT-PCR and northern blot assays (fig. 4). A northern blot of Nasonia head polyA+ RNA hybridized with a P0 derived probe resulted in a 1.6-kb band in both females and males, and a weak 5.9-kb band in females (fig. 4B, P0). The 1.6-kb band likely corresponds to a transcript coding for a predicted 1.2-kb Nasonia ORF (accession nr. XP_001602446) containing most of P0 sequences. The existence of this transcript was further confirmed by 5’ and 3’ RACE experiments (supplementary fig. 2B, Supplementary Material online). The 1.6-kb band was termed Nvu1 (for N. vitripennis unknown 1) and used as control for RNA amount in other northern blots (fig. 4B). The 5.9-kb band in females corresponds likely to P0-fru transcripts, confirming the cDNA library isolates and RT-PCR results (fig. 4A, a–i). Conceptual translation of P0-fru transcripts in frame with the Fru ORF results in the addition of 371 aa N-terminally to the first methionine of the Fru ORF starting on the BTB exon (supplementary fig. 1A, Supplementary Material online).

Non–Sex-Specific fru Transcripts

In addition to the sex-specific transcripts starting with exons P0 and P1, fru transcripts common to both males and females were isolated from Nasonia heads, as known for fru transcripts in Drosophila (Lee et al. 2000; Anand et al. 2001; Song et al. 2002; Dornan et al. 2005). These transcripts start with exons P2, P3, P4, P5, and P6, respectively (table 1 and fig. 1). The Fru ORF start codon is usually the first methionine on the BTB exon (cf. supplementary table 3, Supplementary Material online), except in P6 transcripts (fig. 1) in which the Fru ORF begins on exon P6 with 17 aa (MCFRLLLGHALVHTSSTQ) and continues on the BTB exon with the sequence GSGLEINKMDQ. . ., where M is the first methionine on the N. vitripennis BTB sequence (supplementary fig. 1A, Supplementary Material online).

RT-PCR reactions confirmed that transcripts containing the sequence spanning either one of exons P2, P3, P4, P5, or P6 and the BTB exon are present in both sexes (fig. 4A.b). Connections between BTB and zinc finger containing exons in fru transcripts of both sexes were also confirmed by RT-PCR assays (fig. 4A.c). No other transcript type—in particular those containing Znf D or G—was found among isolated N. vitripennis fru cDNA clones (table 1). Figure 1 provides an overview of all fru transcript classes expressed in Nasonia heads.

Variable Expression Levels Characterize Nasonia fru Sequences

The number of transcript classes isolated from the cDNA libraries displayed considerable variation. In addition, differences in the use of particular zinc fingers were apparent, especially when males and females were compared (table 1). To assess whether these values reflected real promoter-dependent expression differences rather than variation due to cDNA screening procedures, fru transcripts levels were determined by qPCR (fig. 5).

P0-fru and P1-f transcripts are expressed in females about 100 times more than P1-f in males (set arbitrarily as 1), whereas P0-fru in males are virtually absent. The level of P1-m transcripts in males is similar to P0-fru and P1-f transcripts in females, whereas in females, P1-m transcripts are 500 times less expressed than in males (fig. 5A), fru transcript classes common to both sexes are less abundant, with P2 transcripts (P2-2 and P2-3) displaying the highest expression (about 23-fold), followed by P4, P6, and P3 transcripts (fig. 5A). Apart from P0 and P1 derived transcripts, a significant difference in expression between sexes was observed for transcript class P2-3 (males higher than females).

The level of fru transcripts containing a specific zinc finger exon was also quantified (fig. 5B). In this case, the expression level is determined by all fru transcript classes
containing that exon, independently from which promoter they derive. Consistent with the cDNA isolates (table 1), exon B is the most abundant zinc finger in *Nasonia fru* transcripts, followed by C, F, A, G, and D (fig. 5B). Although particular zinc fingers seem to be more represented in male or female isolated full-length *fru* transcripts (table 1), sex-specific expression differences of different zinc finger exons seem marginal. However, for a definitive assessment of sex-biased expression of zinc finger domains, it will be important to determine also to which particular transcript class (i.e., from which driving promoter) specific zinc finger domains belong.

Conservation of Potative Binding Sites for Sex-Specific Splicing Factors in Hymenoptera

In *D. melanogaster*, female-specific splicing of *fru* as well as *doublesex* (dsx) primary transcripts is mediated by Tra and Tra2 (Hedley and Maniatis 1991; Heinrichs et al. 1998), which bind to dsxRE elements (repeats of the sequence (T/A)(C/T/A)/(T/A)(C/A/G)TAATCAACA) present on female-specific exons. *dsx* is also sex specifically spliced in *Nasonia* as well as *A. mellifera*, another hymenopteran species (Cho et al. 2007; Oliveira et al. 2009). In *Nasonia, fru* sex-specific splicing in the head of adult gynandromorphs (Kamping et al. 2007) follows the sex of the tissue (data not shown), indicating that also in *N. vitripennis fru* is likely processed by factors of the sex-determining cascade. Despite that, no Tra–Tra2 binding sites could be found on exon P1 (fig. 3). However, comparative analyses of *N. vitripennis fru* and *dsx* as well as *A. mellifera dsx* sequences revealed the presence of (T/G)GAAGAT(T/A) repeats clustered on a short sequence stretch in female exons of both genes (fig. 3 and 6A). Six of these 8-nt sequences were found spread (but not clustered) on around 30 kb of genomic sequence upstream of the predicted *fru* BTB exon in *A. mellifera*, a region that has completely diverged from the corresponding sequence in *N. vitripennis*. For one of these 8-nt sites, however, the conservation between *N. vitripennis* and *A. mellifera* extended to flanking sequences up to a total of 14 nt (CCCTGAAGATTTGC, cf. figs. 3 and 6A), strongly indicating a homologous site in both species. An RT-PCR reaction on cDNA synthesized with *A. mellifera* head polyA+ RNA purified from *A. mellifera* heads (fig. 6B) confirmed the 14-nt sequence to be also part of a *fru* female-specific exon in *A. mellifera* (fig. 6B). Notably, the distance from the conserved 14-nt sequence to the downstream splice site is exactly the same in both species. Taken together, these data suggest that repeats of the sequence (U/G)GAAGAU(U/A), alone or in conjunction with other sites, may constitute recognition sites for factors responsible for the activation of female-specific splicing in *Nasonia, Apis* and possibly other hymenopteran species.

Discussion

Fruitless plays important functions throughout *Drosophila* development, from specification of axons growth in embryos, differentiation of imaginal-disc derivatives during metamorphosis, to determination of male sexual behavior in adults. Although some Fru sequences are found in most insect orders, nothing is known about the architecture of *fru* transcripts and, in particular, whether *fru* sex-specific splicing is conserved outside Diptera. Here, we have presented a detailed analysis of *fru* transcripts expressed in heads of the haplo-diploid wasp *N. vitripennis*.

Conservation of *fru* Sequences in Evolution

Several lines of evidence indicate that *fru* has very conserved features shared probably by most holometabolous insect orders. As other BTB-C2H3 family members (Spokony and Restifo 2007), *fru* was already present in the last common ancestor of all holometabolous and possibly all neopteran species and evolved in a short period of time all zinc finger domains found in today *fru* loci. The BTB and most zinc finger domains are present in all insect orders analyzed, and some display a striking degree of
amino acid identity. \textit{fru} transcript architecture in \textit{Nasonia} and \textit{Drosophila} is essentially identical; \textit{fru} transcripts begin mainly with an exon linked to the BTB coding exon and, through two connecting exons, end up with zinc finger domains coding exons (cf. fig. 1 and, e.g., Billeter, Villella, et al. 2006). In both species, Fru ORFs begin in most transcripts on the BTB exon, and transcripts common to both sexes arise from promoters most proximal to the BTB exon, whereas sex-specific transcripts derive from promoters more 5’ upstream in the \textit{fru} locus. Also sex-specific splicing of P1 derived transcripts is very similar in both \textit{Nasonia} and \textit{Drosophila}: In males, the Fru ORF starts on exon P1 and is interrupted by in-frame stop codons when the female sequence in exon P1 is included (fig. 3). Another similarity is observed in the expression patterns of specific \textit{fru} transcript classes and the genomic location of the promoters from which they derive. Most expressed transcripts in heads are driven by P1 and P2 promoters in both \textit{Nasonia} and \textit{Drosophila}, assuming that P2 promoters are homologous in both species. As in \textit{Drosophila} (Dornan et al. 2005), P3 derived transcripts are weakly or not expressed in heads of adult \textit{Nasonia}, assuming homology of P3 promoters in both species. Finally, \textit{fru} transcripts containing Znf D are weakly or not expressed in heads of both \textit{Nasonia} and \textit{Drosophila} adults (Billeter, Villella, et al. 2006).

Conservation in many Fru sequences across species and conservation in \textit{Nasonia} and \textit{Drosophila} \textit{fru} transcripts architecture and expression, suggest that also part of \textit{fru}’s functions may be conserved in most insect orders. Most conserved functions may be those required early during development and/or those necessary for viability. For instance, one of the \textit{fru} functions observed most early in development, and likely mediated by P3 derived \textit{fru} transcripts, is the regulation of axons growth during embryonic development (Song et al. 2002). 

\textit{Fru} functions required for viability, on the other hand, also mediated by P3 derived transcripts, provide for the correct differentiation of imaginal-disc derivatives during metamorphosis (Anand et al. 2001). Among zinc finger domains that could mediate most conserved and possibly vital functions are likely those that display the most conservation across species. These are Znfs B and C. B in particular, displays a striking degree of amino acid identity throughout its entire length (supplementary fig. 3, Supplementary Material online).

Conservation of Sex-Specific Splicing

Conserved P1 transcript architecture between \textit{Nasonia} and \textit{Drosophila} suggests that also sex-specific splicing was a \textit{fru} trait already present before the radiation of holometabolous insects. In \textit{Drosophila}, sex-specific splicing factors bind to dsxRE sites, repetitive sequence elements shared by \textit{fru} and \textit{dsx} primary transcripts (Hedley and Maniatis 1991; Heinrichs et al. 1998). Similar sites have been identified in \textit{A. gambiae}, suggesting that the use of such repeats is a common mechanism in Diptera for regulating \textit{fru} and \textit{dsx} sex-specific splicing. If we consider the ancient origin of \textit{fru} sex-specific splicing (Hymenoptera is the most basal of major holometabolous orders, Savard et al. 2006) and the striking conservation in the architecture of \textit{fru} sex-specific transcripts in \textit{Nasonia} and \textit{Drosophila}, it may not be surprising to find also similar repeats conserved in homologous exons of homologous genes also in two hymenopteran species. Interestingly, the distance from the downstream female splice sites to the upstream conserved sequences in \textit{Nasonia} and \textit{A. mellifera} and the 5’ most dsxRE element in \textit{Drosophila fru} transcripts, is exactly the same (240 bp). Because of the high degree of conservation between species and genes and their concentration at analogous sites in homologous \textit{Drosophila} genes, we anticipate that the elements identified may be important sites, if not the only ones, used possibly by sex-specific splice regulators. If this will be confirmed, factors responsible for sex-specific splicing may not only be the same for \textit{fru} and \textit{dsx} genes within one species but also be similar in both hymenopteran species, suggesting similar sex-specific splicing mechanisms in \textit{N. vitripennis} and \textit{A. mellifera}.

Although conservation of genes at the bottom of the sex-determining cascade in insects was hypothesized (Wilkins 1995), because \textit{N. vitripennis} and \textit{A. mellifera} differ in their primary sex-determining signal (Beye et al. 2003; Beukeboom et al. 2007), a difference is expected at the next upstream level or in the modulation of the sex-specific splicing mechanism.

\textbf{Nasonia-Specific fru Features}

Compared with corresponding \textit{Drosophila} and \textit{Anophelines} sequences, \textit{Nasonia} \textit{fru} transcripts display differences at two expected sites. Sequences connecting the BTB to zinc finger domains are different for all three species. However, because a transgene expressing the \textit{A. gambiae} \textit{Fru}^{MC} isoform was able to substitute for the \textit{D. melanogaster} ortholog (Gailey et al. 2006), the composition of sequences connecting BTB and zinc finger domains is not expected to play species-specific roles. The second difference concerns the N-terminal extension of male Fru proteins that, although present in \textit{N. vitripennis} as well as \textit{D. melanogaster} and \textit{A. gambiae} \textit{Fru}^{M} proteins, are different in composition and length for all three species. Expression of \textit{Fru}^{M} proteins missing the N-terminal extension were still able to promote the development of the MOL in \textit{Drosophila fru}^{AM} mutants (Usui-Aoki et al. 2000), whereas males (but not females) of a P1-gal4 transgenic \textit{Drosophila} line expressing \textit{Fru}^{M} isoforms lacking the male N-terminus displayed vigorous male courting behavior (Ferri et al. 2008). From these reports, the male N-terminal extension seems to be dispensable for some but not all functions controlled by \textit{Fru}^{M} proteins. Because in BTB-C_{2}H_{2} family members, protein–protein contacts involving the N-terminal extension form a significant fraction of the residues involved in protein–protein interaction (Stogios et al. 2005) and because there is no homology between the male Fru N-terminal extension of the three different species, for a conclusive answer on the role of these sequences in Fru proteins further analyses are required.

\textit{P0-fru} are the first \textit{fru} transcripts coding for a large Fru protein exclusively in females. Although its functionality still needs to be determined, this could be the first identified Fru female-specific protein. The sequence encoded by \textit{P0} exons does not show any similarity with known proteins. Because \textit{P0} sequences are highly expressed in both males and females (cf. fig. 4B, P0 blot: NvuI band), an alternative hypothesis is that \textit{P0-fru}
transcripts are by-products of an imperfect splicing of transcripts coding for protein XP_001602446. Whichever the explanation, the expression of PO-fru transcripts in Nasonia females is symptomatic for a gene that, already displaying a dynamic evolutionary history, seems to be still in active evolution.

A Possible Evolutionary History of Fruitless in Insects

Besides conservation of many domains, important differences between major insect orders and even within them are apparent. The loss of important zinc fingers, such as D, A, and Fin B. mori, or the loss of Znf Fin drosophilids requires an explanation. As discussed in the Results section, most zinc finger domains were present before the radiation of holometabolous species. However, the ancestral fru locus had probably only three or two C2H2 zinc fingers, which might have been the common ancestors of A and C and of B and F. It could be then possible that A and C, and B and F, respectively, have partially redundant functions, retained possibly from the common ancestral zinc finger. In line with that, in some experimental settings, FruM transgenes expressing either FruMA or FruMC isoforms but not FruMB were able to rescue fru mutations affecting embryonic central nervous system axonal fasciculation (Song et al. 2002) and the formation of the MOL in fruM mutants (Usui-Aoki et al. 2000). On the other hand, Billeter, villella, et al. (2006) found that only a transgene expressing FruMC but not FruMA or FruMB was able to restore male fertility in a mutant lacking ZnfC in male-specific fru transcripts. Moreover, FruMC but not FruMA or FruMB rescued formation of the MOL in both FruMC- and Fru-2-null males. Notwithstanding differences among the different experimental setups used in the discussed experiments (e.g., different expression drivers, different developmental stages tested), which makes it difficult to compare apparently inconsistent results, it seems that, for some functions, Znf A and C are functionally redundant, whereas for others they are not.

Empirical and theoretical evidence indicates that duplicated genes are usually maintained because ancestral functions are partitioned among duplicates and/or because at least one duplicated gene evolves new functions (Force et al. 1999). In line with the above observations, the same processes occur also with duplicated gene domains. Here, alternative splicing becomes an important element for regulating subfunctionalization (Kopelman et al. 2005). Postulating that the present fru locus is derived from an important ancestral gene, some ancestral (and possibly vital) functions may have been maintained and now be mediated by the most conserved Znfs B and C (Znf B shows high relatedness to ttk Znf 69, Spokony and Restifo 2007), whereas A and F may have retained only part of the ancestral activity (for which functional redundancy could be a sign) and have acquired novel functions. Where the latter was not the case, zinc fingers have been eliminated. Although still conceivable for the gain and loss of C2H2 Fru zinc fingers, this hypothesis does not account for the loss of Znf D in B. mori. Either the function of Znf D in B. mori is taken over by other fru zinc fingers (or even other genes) or Znf D has a dispensable role, a hypothesis easily testable in D. melanogaster.

To gain better insight into fru’s evolutionary origin, one idea could be knocking out more derived zinc fingers, such as F, in species that still have them. A reciprocal experiment would be the expression of specific zinc fingers in species that lack them, such as Znf D in B. mori or Znf F in D. melanogaster. If these zinc fingers were really present in the ancestral fru locus, their expression or inactivation in present species could reveal ancestral features or show whether, once activated, they control the expression of novel phenotypes in species lacking them.

The picture of zinc fingers appearing and disappearing from the fru locus during evolution is not unusual but common to other families of genes containing C2H2 zinc finger domains. Tadepally et al. (2008) studied the evolution of C2H2 gene families in mammals and found that, besides species-specific loss and duplication of entire gene clusters, variation in zinc finger domain numbers in orthologous genes is a common phenomenon.

Conclusions

If the discovery of functional sex-specific fru transcripts in another dipteran insect was remarkable (Gailey et al. 2006), it is even more striking to observe conservation in fru locus architecture and transcript splicing pattern in a haplo-diploid insect. This finding indicates that fru sex-specific splicing evolved either prior to the split between Hymenoptera and Diptera (250–300 Ma) or that it has been acquired independently in both lineages. If we consider additionally that Hymenoptera constitute the basal order among holometabolous insects (Savard et al. 2006), that the major representatives of which have all conserved domains at the fru locus, and that core Fru domains are found in hemimetabolous insects (Davis et al. 2000; Ustinova and Mayer 2006), fru as a multifunctional gene in insects likely predates the split between hol- and hemimetabolous insects. It is then conceivable that, albeit in a primitive form, it belonged to the genetic toolkit of the insects’ last common ancestor. Further studies will reveal how many fru functions are conserved in insects and what significance the observed species-specific differences have. In particular, if also in other insect species fru is necessary for male sexual behavior, it will be interesting to understand how interspecific differences in this trait are produced. In that respect, Nasonia species—being cross-fertile, having distinct sexual behaviors (van den Assem and Werren 1994) and sequenced genomes—will provide an enormous opportunity for addressing questions pertaining to the evolution of sexual behavior in insects.

Supplementary Material

Supplementary tables 1–3 and supplementary figures 1–3 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Accession numbers of Nasonia fru sequences will be deposited in the NCBI sequence database.

Note

A new work by Salvemini et al. (2009) shows that fru is also sex-specifically spliced in the Medfly, Ceratitis capitata.
Acknowledgments

We thank T. E. Wallaart and G. Paterson for precious help in the construction of cDNA libraries; W. T. Stam for assistance in the phylogenetic analyses; R. Breitling and A. Kamping for help with statistical analyses. We gratefully acknowledge the Nasonia Genome Sequencing Consortium, the Honey Bee Genome Sequencing Consortium (Honeybee Genome Sequencing Consortium 2006), and the Tribolium castaneum Genome Sequencing Consortium (Tribolium Genome Sequencing Consortium et al. 2008), for making their data publicly available, and the BCM-HGSC for providing access to the data; The Silkworm Genome Sequencing Consortium (Xia et al. 2004) for making their data publicly available at http://silkworm.genomics.org.cn/; The Anopheles gambiae Genome Sequencing Consortium (Holt et al. 2002) and the Pediculus humanus corporis Genome Sequencing Consortium for making their data publicly available at http://www.vectorbase.org/index.php; The Daphnia Genomics Consortium for making their data publicly available at http://wfleabase.org/; The Drosophila 12 Genomes Consortium (Drosophila 12 Genomes Consortium et al. 2007) for making their data publicly available. R.C.B. was supported by a grant from the Swiss National Science Foundation (SNF) for “prospective researchers” and currently by an Aard- en Levenswetenschappen (ALW) grant (nr. 817.02.020) from the Netherlands Organization for Scientific Research (NWO). L.W.B. is supported by a “Pioneer” ALW fellowship (nr. 833.02.003) from NWO. This work was performed at Evolutionary Genetics, Centre for Ecological and Evolutionary Studies.

Literature Cited


Marcy Uyenoyama, Associate Editor

Accepted March 14, 2009