

Alternative Starts of Transcription, Several Paralogues, and Almost-Fixed Interspecific Differences of the Gene *fruitless* in a Hemimetabolous Insect

Jana Ustinova, Frieder Mayer

Department of Zoology, University of Erlangen, Staudtstr. 5, 91058 Erlangen, Germany

Received: 30 September 2005 / Accepted: 12 June 2006 [Reviewing Editor: Dr. Willie Swanson]

Abstract. In fruit flies, the gene *fruitless* (*fru*) governs the establishment of the potential for male sexual behavior. We partially cloned *fru* from a hemimetabolous insect for the first time and we compared *fru* among three closely related and acoustically communicating grasshopper species: *Chorthippus biguttulus*, *C. brunneus*, and *C. mollis*. The *fru* of grasshoppers is organized similarly to *fru* of holometabolous insects, with a BTB and Zn-finger domains separated by a nonconserved repetitive linker. As in *Drosophila*, several transcripts of *fru* are found in grasshoppers. We also present evidence for the coexistence of several copies of *fru* in the grasshopper genome. Within species these copies are almost identical and carry almost-fixed species-specific differences. This suggests that the paralogous copies of *fru* in grasshoppers do not evolve independently from each other.

Key words: *fruitless* — Concerted evolution — Birth-and-death evolution — Orthoptera — *Chorthippus*

Introduction

Understanding the genetics of complex and evolutionary important traits, such as sexual behavior, is one of the most challenging tasks in evolutionary

genetics. Compared to vertebrates, insects have nervous systems of relatively low complexity with a much smaller number of neurons. Their sexual behavior often is rather simple and stereotyped. Therefore, it is not surprising that most research on the genetics of sexual behavior focussed on the fruit fly *Drosophila*. The male courtship behavior of *Drosophila* consists of a sequence of fixed action patterns: the male orients toward the female, follows her, taps her with his forelegs, produces a courtship song by opening and vibrating one wing it, licks her genitalia, and, finally, attempts copulation (reviewed in, e.g., Greenspan and Ferveur 2000; Yamamoto and Nakano 1999). In recent years it has been shown that almost all aspects of this behavior are governed by a genetic hierarchy, which is headed by the regulatory gene *fruitless* (*fru*).

Originally, *fru* was identified on basis of the aberrant courtship behavior of mutant males, which were not able to discriminate between sexes and courted males and females at equal rates. In addition, these mutants lacked a male-specific abdominal muscle of Lawrence (MOL [Ito et al. 1996; Ryner et al. 1996; Villella et al. 1997]). Other *fru* mutants showed behavioral abnormalities of different severity, from almost-complete loss of all steps of male courtship (Anand et al. 2001; Goodwin et al. 2000; Villella et al. 1997) to mild changes of particular components of courtship including courtship songs (Ryner et al. 1996; Villella et al. 1997). In contrast, female courtship behavior as well as general locomotion function and wing usage of males was not affected by mutations of *fru* (Ryner et al. 1996; Villella et al. 1997). This led to a hypothesis that *fru*

Correspondence to: Frieder Mayer; email: fmayer@biologie.uni-erlangen.de

regulates sex identity of the central nervous system (CNS) and builds the potential for male courtship behavior of fruit flies on a sex determination pathway (reviewed in, e.g., Baker et al. 2001).

The gene *fru* is one of the largest and most complex of a fruit fly. It spans about 130 kb and codes for a family of transcription factors with the BTB (Broad-Complex, Tramtrack, and Bric à brac [Zollman et al. 1994]) and zinc-finger domains. Numerous transcripts are generated by the use of four promoters (P1–P4) and alternative splicing at both the 5' and the 3' ends. Only transcripts from the most distal promoter P1 are spliced in a sex-specific manner: in males per default and in females under control of the sex-determination regulatory proteins Transformer (Tra) and Tra2 (Heinrichs et al. 1998; Ryner et al. 1996). Female-specific transcripts are not translated in either sex of *D. melanogaster* and—with very rare exceptions—all other *Drosophila* species (Lee et al. 2000; Usui-Aoki et al. 2000; Yamamoto and Nakano 1999; Yamamoto et al. 2004). Recently, Demir and Dickson (2005) and Manoli et al. (2005) experimentally proved that the male-specific splicing of the P1-derived transcripts is necessary for generation of male behavior in males and, more importantly, is sufficient for the generation of male behavior in otherwise unaltered females. Transcripts from other promoters (P2–P4) do not have sex-specific functions but are essential for viability of flies, as they are involved in many aspects of neuronal and nonneuronal development (Anand et al. 2001; Goodwin et al. 2000; Lee et al. 2000; Ryner et al. 1996; Song et al. 2002; Song and Taylor 2003).

Male-specific proteins are expressed in about 2% of all neurons in the CNS, particularly in neurons innervating abdominal organs directly relevant to *fru* function, specifically, the MOL and the male internal reproductive organs (Billeter and Goodwin 2004; Lee et al. 2000). Males lacking the P1-derived Fru proteins show no other defects except a complete absence of sexual behavior and a loss of the MOL (Anand et al. 2001). It has been shown that in *D. melanogaster* the MOL is formed only if the innervating motoneurons are of male origin (Lawrence and Johnston 1986; Usui-Aoki et al. 2000).

Thus, the gene *fru* regulates a wide variety of functions in *D. melanogaster*, ranging from male courtship behavior and formation of the MOL to vital steps in the development of both embryonic and adult nervous systems and some external structures. These functions are segregated between different transcripts, which are generated from the *fru* transcription unit.

The gene *fru* is conserved within fruit flies (Davis et al. 2000a, 2000b; Gailey et al. 2000). Data for other insects are scarce. *fru* was partially sequenced for a damselfly *Ischnura asiatica* (Gailey et al. 2000). For *Anopheles gambiae* and *Apis mellifera* complete se-

quences of *fru* homologues are available due to genome sequencing projects. Interestingly, in *A. gambiae* *fru* is also spliced in a sex-specific manner (GenBank accessions AAU50567–AAU50568, AAV52864–AAV52865 [Gailey et al. 2006]). Findings for *fru* in different insect orders (Diptera, Odonata, and Hymenoptera) suggest that *fru* may be present in all insects. However, *fru* has not yet been found in a hemimetabolous insect.

Here we report on the results of partial cloning and sequencing of *fru* homologues of three closely related species of hemimetabolous insects belonging to the genus *Chorthippus* (Orthoptera, Acrididae, Gomphocerinae). Gomphocerine grasshoppers produce calling and courtship songs to attract and find mating partners, and it is believed that their complex bidirectional acoustic communication system led to rapid radiation by the evolution of premating hybridization barriers (reviewed by von Helversen and von Helversen 1994). The three species studied, *Chorthippus biguttulus*, *C. brunneus*, and *C. mollis*, are closely related and occur sympatrically and often syntopically. They are morphologically and genetically similar, but can be readily identified by innate species-specific male calling songs (von Helversen and von Helversen 1994; Mason et al. 1995; Ragge 1987). *fru* is regarded as a candidate gene for song production in acoustically communicating grasshoppers, because it is consistently considered an important gene for song generation in *Drosophila* (reviewed, e.g., in Kyriacou 2002).

Materials and Methods

Animals

Individuals of *C. biguttulus biguttulus*, *C. biguttulus eisentrauti*, *C. brunneus*, *C. mollis*, and *C. parallelus* were collected at locations shown in Table 1. The species affiliation of adult males was determined using their individual calling songs; that of females, according to their receptivity responsiveness toward males.

Molecular Cloning of the BTB Domain

Total RNA was isolated from heads and thoraxes using the RNeasy Minikit (Qiagen). Degenerated primers ER348–ER351 (Table 2) were designed on the basis of multiple alignments of the *fru* genes of all insect species available from GenBank and used for reverse transcription and amplification (RT-PCR). RT-PCR was performed using the Access RT-PCR system (Promega). Twenty-five-microliter reactions contained 5 μ l of *Amv/Tfl* 5 \times buffer, 0.5 μ l of 10 mM dNTP mix, 1 μ l of 25 mM MgSO₄, 0.5 μ l each of *Amv* reverse transcriptase and *Tfl* polymerase, 1 μ l each of 25 μ M primers, 0.5 μ l of RNasin, and 100 ng of total RNA. Reverse transcription was performed according to the protocol recommended by the manufacturer. Cycling conditions were 94°C for 30 s, 56°C for 1 min, 68°C for 2 min (40 cycles), then 68°C for 7 min. RT-PCR products were cleaned using Qiaquick or Minelute PCR and gel-purification systems (Qiagen), ligated into the pGEM-T or

Table 1. Number of individuals from different locations used in this study

	<i>Chorthippus brunneus</i>	<i>C. mollis</i>	<i>C. biguttulus biguttulus</i>	<i>C. biguttulus eisentrauti</i>	<i>C. parallelus</i>
Erlangen, Germany	9	7	5	—	1
Haid, Germany	2	6	5	—	—
Sampzon, France	1	—	2	—	—
Puschino, Russia	—	—	1	—	1
Crnotice, Slovenia	—	1	—	—	—
Carinthia, Austria	—	—	—	2	—
Total	12	14	13	2	2

Table 2. Primers used for cloning of the *fru* gene of grasshoppers

	5'-to-3' sequence	Designed for	Used in
ER348 (F)	TGGAAYAAAYCAYCCSACNAAYYTRAC	Insects	RT-PCR
ER349 (F)	GCSGTYTTGAGRAACATSGGCARYGA	Insects	RT-PCR
ER350 (R)	CTNTGYGAYGTSACNCTNGCNTGYGA	Insects	RT-PCR
ER351 (R)	CCSACRTTKACYTCGCCYTKRTACATRAA	Insects	RT-PCR
ER354 (F)	TCCTACAGAATACTCATCCTCATCCAA	Grasshoppers	3'-RACE
ER361 (F)	CACAGAAATGAGAGCTTTATTGCAGTT	Grasshoppers	3'-nRACE
ER380 (R)	TACCAGCACTACTCCGTTCTTCAA	Grasshoppers	5'-RACE, PCR
ER379 (R)	AACATCGGCAACAACTTTGACTAA	Grasshoppers	5'-nRACE
ER462 (F)	GGCATCACATAAGGAGGAAAGAAAGT	<i>Chorthippus</i>	PCR, RT-PCR
ER464 (F)	AAATACGAAGAAAGCCAAGGAAACACT	<i>Chorthippus</i>	PCR, RT-PCR
ER478 (F)	CATAAGGRGGAAAGAAAGTCA	<i>Chorthippus</i>	Sequencing
ER479 (R)	ACTACTCCGTCTTCAATTG	<i>Chorthippus</i>	Sequencing
ER489 (R)	AAGTGGTGCCANCKRTTGGTSACCT	Insects	RT-PCR

Note. F, forward primer; R, reverse primer.

pGEM-T easy vectors (Promega), and propagated in *E. coli* JM109 competent cells (Promega). Plasmid DNA was purified by different methods, and inserts were cycle sequenced in both directions using the Thermo Sequenase DYEnamic Direct Cycle Sequencing Kit (Amersham Biosciences) and the automated IR2 Long Reader 4200 DNA sequencer (LI-COR). Sequences were manually aligned and analyzed with the help of the macDNASIS V 3.5 software (Hitachi Software Engineering).

Rapid Amplification of cDNA ends (RACE)

Total RNA (600 ng–1 µg) was used for the first-strand cDNA synthesis. The 5'-RACE-ready first strand was synthesized using the SMART RACE cDNA amplification kit (Clontech), and the 3'-RACE-ready first strand was synthesized using the SMART technology as recommended in the protocol or the ImProm-II reverse transcriptase (Promega). In the latter case, about 600 ng of RNA was incubated in a 5-µl reaction with 1 µl of the 10 µM 3'-CDS primer (Clontech) at 70°C for 5 min, and the mixture snap-chilled on ice. Then the mixture was combined in a 20-µl reaction with 4 µl of 5× buffer, 2.4 µl of 25 mM MgCl₂, 1 µl of 10 mM dNTP mix, 0.5 µl of RNasin (Promega), and 1 µl of the ImProm-II, and subsequently incubated at 25°C for 5 min, and at 42°C for 1 h 30 min. The mixture was diluted with 100 µl of Tricine-EDTA buffer (pH 8.5) and incubated at 70°C for 15 min. 5'- and 3'-RACE reactions were performed using Phusion high-fidelity DNA polymerase (Finnzymes). Twenty-five-microliter reactions contained 5 µl of the HF buffer, 2.5 µl of 10 mM dNTP mix, 1.25 µl of the 5'- or 3'-RACE-ready cDNA, 0.5 µl of a 10 µM gene-specific primer (Table 2), and 2.5 µl of the 10× Universal Primer Mix (Clontech). Cycling conditions were 98°C for 30 s; 98°C for 10 s, 61°–68°C (depending on the primer) for 30 s, 72°C for 4 min (35 cycles); then 72°C for 10 min. We routinely performed nested RACE (nRACE)

reactions in order to confirm the origin of bands and to obtain A-overhangs needed for TA cloning. Therefore, RACE reactions diluted 50× with Tricine-EDTA buffer or undiluted were used as templates in subsequent PCR reactions. Twenty-five-microliter nRACE reactions typically contained 2.5 µl of 10× buffer, 2.5 µl of 25 mM MgCl₂, 2.5 µl of 10 mM dNTP mix, 0.1 U of Taq DNA polymerase (Promega), 1 µl of 10 µM Nested Universal Primer (Clontech), 1 µl of 10 µM nested gene-specific primer (Table 2), and 1 µl of the RACE reaction. Cycling conditions were 95°C for 1 min; 95°C for 20 s, 57°–68°C for 30 s, 72°C for 5–8 min (35 cycles); then 72°C for 10–30 min. nRACE products were cloned and sequenced as described above.

PCR

Genomic DNA was isolated from different tissues according to standard protocols. A 960- to 980-bp-long fragment containing the 5'UTR (untranslated region) and the BTB domain of the *fru* was amplified in individuals listed in Table 1. Therefore, we used either the Taq DNA polymerase (Promega) or the proofreading Phusion DNA polymerase (Finnzymes). When Taq was used, 25-µl reactions were set as follows: 2.5 µl 10× buffer, 2.5 µl of 25 mM MgCl₂, 2.5 µl of 10 mM dNTP mix, 0.1 U of Taq, 1 µl each of the 10 µM primers ER462 and ER380, and about 50 ng of DNA. Cycling conditions were 95°C for 2 min; 95°C for 30 s, 63.5°C for 20 s; 72°C for 1.5 min (35 cycles); then 72°C for 5 min. When Phusion was used, 25-µl reactions contained 5 µl of HF buffer (Finnzymes), 2.5 µl of 10 mM dNTP mix, 0.6 µl each of 10 µM primers ER462 and ER380, 0.25 µl of Phusion, and 25–50 ng of DNA. Cycling conditions were 98°C for 30 s; 98°C for 10 s, 63.5°C for 20 s; 72°C for 1 min (35 cycles); then 72°C for 1 min. Taq products were cleaned and AT-cloned as described above. Phusion products were A-tailed prior to cloning as described in

the manual for pGEM vectors (Promega). Seven to twenty clones per individual were picked and sequenced as described above or using the sequencing service of Qiagen. Alternatively, PCR products were directly sequenced omitting the cloning step using primers ER478 and ER479. For studying the genomic organization of the 5'-UTR and the BTB domain of *fru*, we amplified a corresponding fragment of cDNA of several individuals (Table 1). Total RNA was isolated as described above. RT-PCR was performed with 50–200 ng of total RNA using the Access RT-PCR system (Promega) or the One-Step RT-PCR system (Qiagen) as described above. Consensus species-specific genomic sequences of the studied fragment and mRNA sequences of the additional Zn-finger containing fragment, recovered by RT-PCR with one gene-specific and one degenerated primer (ER 489; Table 2), were deposited in GenBank under accession numbers DQ424928–DQ424933.

Counting the Number of *fru* Haplotypes per Individual

All singletons (substitutions found in only one clone) were treated as PCR mistakes. All substitutions found in at least two clones and substitutions found in directly sequenced PCR products were treated as actual polymorphisms. Reliability of sequencing of clones was confirmed by occurrence of the same haplotypes in both cDNA and genomic DNA-derived PCR products for several individuals. In addition, it was confirmed in a pedigree of grasshoppers (parents and three progenies), in which parental haplotypes were found in progenies. For that, PCR was performed with individual genomic DNA using a proofreading DNA polymerase Phusion (Finnzymes), as described above. PCR products were cloned and 15–21 clones per individual were sequenced.

Phylogenetic and Molecular Evolutionary Analyses

Phylogenetic trees were constructed using the software PAUP*, version 4.0b10 (Swofford 2000) or MEGA version 3.1 (Kumar et al. 2004). Molecular evolutionary analyses were conducted using MEGA version 3.1.

Results

Molecular Cloning of the fruitless Homologue of Grasshoppers

Using degenerated primers we amplified and cloned a cDNA fragment, which corresponds to the BTB domain of the gene *fruitless* (*fru*). This sequence was used to design nondegenerated primers for 5'-RACE reactions. Two types of transcripts with the same coding sequence, carrying a start codon and an open reading frame (ORF), but with different 5'-UTRs (leaders), were amplified in 5'-RACE from cDNA of a single adult male (Fig. 1). The presence of both types of transcripts in several adult males was confirmed by RT-PCR. To date, nothing is known about differences in *fru* transcripts between males and females.

PCR products generated from genomic DNA revealed that both leaders found in RACE products are arranged head to tail and are separated from each

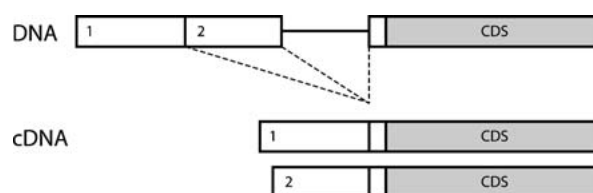


Fig. 1. Schematic representation of genomic organization of the cloned portion of *fru* (upper line) and of different splice variants of *fru* (two lower lines) found in grasshoppers. Gray and white boxes represent coding and noncoding regions, respectively. A thin line represents an intron. Dashed lines show splicing patterns. Numbers correspond to the two alternative leaders 1 and 2.

other by only two nucleotides (Fig. 1). Thus, the two types of 5' leaders, which both are spliced to a common site 26 bp upstream of the start codon, seem to be generated via alternative starts of transcription and alternative splicing of one transcriptional unit. A 114-bp intron containing a poly(A) track of variable length is excised from transcripts of the second type. No other introns were found within the cloned fragment of the coding sequence (CDS). In contrast, *fru* in *Drosophila* species harbors two introns within the homologous region. We did not find any known transcription factor binding sites or other regulatory signals in the whole 5' noncoding region of *fru* in grasshoppers.

3'-RACE reactions were repeated several times and always resulted in early-terminated products due to nonspecific binding of the poly(dT) primer to an A-rich track of nucleotides within the *fru* gene. However, an additional 500 bp of cDNA was recovered in RT-PCR reactions with one grasshopper-specific and one degenerated primer targeted to a Zn-finger domain B (Ryner et al. 1996; Gailey et al. 2006), which is conserved in several *fru* transcripts of *D. melanogaster*, *A. mellifera*, and *A. gambiae*.

In total, we cloned 1511 bp of the *fru* gene covering a large part of the CDS and the 5' flanking region (540 bp). Conceptual translation of the longest ORF (323 amino acids) most closely resembles the *fruitless* gene of *D. melanogaster* and the locus XP392552 of *A. mellifera*, as revealed by the translating BLAST (Altschul et al. 1997). The ORF in grasshoppers starts at the same position as in the female-specific and some non-sex-specific transcripts of *D. melanogaster*. The BTB domain is highly conserved between grasshoppers and other insects (Fig. 2). At the protein level, the BTB domain of grasshoppers is to 86% and 80% identical to the BTB domains of *A. mellifera* and *D. melanogaster*, respectively. The sequence downstream of the BTB domain is most similar to the locus XP392552 of *A. mellifera*, but the level of identity is much lower (31%). The 3' distal end of the obtained fragment contains a portion of a Zn-finger domain that is similar to the *fru* Zn-finger B domains of other in-

<i>C. mollis</i>	101	MDQQFCLRW	<u>NNHPTNLTDV</u>	LSNLLHROAL	VDVTLACEGE	TFKAHOTLLS	150
<i>C. brunneus</i>	101T.....	150
<i>A. mellifera</i>	101	K....Y....A.....	..S..A.E..	C.....V..	150
<i>D. melanogaster</i>	101	A.....G..	.TS..Q.E..	C.....D..	.V.....	150
<i>C. capitata</i>	101G..	.TS..Q.E..	C.....D..	.V.....	150
<i>A. gambiae</i>	101	I....Y....	...QS...T..	.TT..QDEK.	C.....KG	MV...A..	150
<i>I. asiatica</i>	101A..	..S..Q.E..D.Q	..R.....	150
<i>B. cucurbitae</i>	101G..	.TS..Q.E..	C.....D..	.V.....	150
<i>C. mollis</i>	151	<u>ACSPYFERIF</u>	<u>LONTHPHPTI</u>	<u>FIRDVHYTEM</u>	<u>RALLOFMYKG</u>	<u>EVNVSOSLLP</u>	200
<i>C. brunneus</i>	151	200
<i>A. mellifera</i>	151S..K..NE..	K...H....H..	200
<i>D. melanogaster</i>	151T..	...Q.....	Y.K..R.S..	.S..D....	...G..S..	200
<i>C. capitata</i>	151T..	...R.....	Y.K..R.S..	.S..D....	...G..S..	200
<i>A. gambiae</i>	151Q..	VE.K.L...	Y...EVN..	...D...Q.	...G.HN.Q	200
<i>I. asiatica</i>	151SL.	I..H....V	I..K..N...N..	200
<i>B. cucurbitae</i>	151T..	...R.....	Y.K..R.S..	.S..D....	...G..S..	200
<i>C. mollis</i>	201	<u>MFLKTAFALE</u>	<u>ISGLTQGAVK</u>	KPGDRTSSSP	AASPARTTIE	ERSSAGSPPP	250
<i>C. brunneus</i>	201	250
<i>A. mellifera</i>	201Q	.R...DNS.N	NKTEEK.P..	EPETQTGIRH	TE.PNLQ...	250
<i>D. melanogaster</i>	201S.Q	VR...DNNNL	NYRSDCKLR	DSAASSPTGR	GP.NYTGGLG	250
<i>C. capitata</i>	201S.Q	VR...DNNNL	NYR---TEHR	DSPVSSPTGR	TPY.SS-GLG	250
<i>A. gambiae</i>	201	N.....S.K	VR.....	-----ES.A	DRYS.D.DSK	L..ERIRDSR	250
<i>I. asiatica</i>	201	250
<i>B. cucurbitae</i>	201	250

Fig. 2. Comparison of the cloned part of the Fru protein of grasshoppers and other insects. The BTB domain is underlined. Only the first 35 amino acids of the highly variable part of the protein are shown. Dots indicate identity with the sequence of *C. mollis*. Dashes indicate gaps in the alignment. Numbers correspond to the position in the complete alignment of Fru proteins, starting

with the male-specific transcripts of *D. melanogaster*. Accession numbers of used sequences: *C. mollis*, DQ424930; *C. brunneus*, DQ424929; *Apis mellifera*, XP392552; *D. melanogaster*, NP732349; *Ceratitis capitata*, AAF22477 and AAF22527; *Anopheles gambiae*, AAU50567; *Ischnura asiatica*, AAF22481; *Bactrocera cucurbitae*, AAF22479.

sects. In contrast, the 5' noncoding region does not have significant similarity to any accession in GenBank. However, it is conserved within the genus *Chorthippus*, as it was also amplified from the distantly related grasshopper species *C. parallelus*.

Grasshoppers Have Several Paralogues of fru

We amplified and cloned a 980-bp-long genomic fragment of the *fru* gene (including alignment gaps) of 16 individuals representing three closely related species and one subspecies: *C. biguttulus biguttulus* ($N = 3$), *C. biguttulus eisentrauti* ($N = 2$), *C. brunneus* ($N = 8$), and *C. mollis* ($N = 3$). The cloned genomic fragment included the 5' noncoding region, the complete BTB domain, and about 120 bp downstream of it. Sequencing of clones showed that there were 2 to 13 (mean, 4.9) haplotypes per individual, when singletons (substitutions occurring only once in an alignment of sequences of clones) were treated as PCR artifacts. In addition, we performed a strong simplification of the observed pattern of DNA polymorphism assuming that (a) substitutions shared by otherwise different haplotypes, which were sequenced only once in two different individuals, and (b) differences in length of the poly(A) stretch within the intron in otherwise identical haplotypes were also PCR artifacts. Under these strong assumptions, the number of haplotypes varied from 1 to 5 (mean, 2.8; e.g., Fig. 3). Obviously, several haplotypes obtained from genomic DNA represented pseudogenes, as they carried corrupted start codons or in-frame stop co-

ons in the coding sequence. If only RNA-derived data were considered, the mean number of haplotypes per individual was 4.25 or, under the strong simplification described above, 1.75. However, even in the latter case, at least one individual carried three haplotypes of transcribed *fru*, which cannot be explained by a single locus. Taken together, our data suggest that the genome of grasshoppers contains more than one copy of *fru*.

Intraspecific and Interspecific Variation

Under the strongest simplification of the pattern of DNA polymorphism mentioned above we found 37 distinct haplotypes in 16 individuals of the three species studied. Three haplotypes occurred in more than one animal. One of them was *C. brunneus*-specific and was found in five of the eight cloned individuals. The other two haplotypes were shared by *C. biguttulus eisentrauti* and *C. biguttulus biguttulus* or *C. mollis*. Surprisingly, in 9 of 16 cloned individuals haplotypes could be assigned to one or more sets, in which haplotypes resembled each other more strongly than haplotypes belonging to another individual or belonging to another set within the same individual. Each haplotype set was characterized by at least one set-specific substitution and haplotypes within a haplotype set differed in one to five nucleotides (0.1%–0.5%; e.g., individual bru1487 in Figs. 3 and 4). Two individuals of *C. parallelus* carried *C. parallelus*-specific haplotypes characterized by 15 substitutions and two indels within the 5'UTR.

Position	00111112222233333344444444445555666667777888888999
in the	3500233244578022566003333677858890123994779223467024
alignment	8257038215859328612691234739053513350479453182270460
	eeeeeeeeeeeeeeeeeeeeiiiiiiiiiiiiiCCCCCCCCCCCCCCCCCCCC
consensus bru	ACAAGGGTCGAAGCCATCTTGAGTTCAA-AGATTGGTCCACAGGATATAATA
consensus mol	R.....C.....
consensus big	G.....G...A...T...C.....GT.....
bru1240_1	...A..C.....-...C.....G.....G.. (4)
bru1240_2-..... (2)
bru1240_3G... (2)
bru1241_1TC..A.....-G.....G (2)
bru1241_2	..G.....-...A..... (4)
bru1241_3	..G.....T.....-...G.A.....C..... (2)
bru1242_1-..... (2)
bru1242_2T..... (7)
bru1242_3G.....-...G.....TT..T..... (2)
bru1243_1-..... (7)
bru1243_2G.....-...A.....A..... (3)
bru1244_1-.....G...C. (2)
bru1244_2-..... (1)
bru1244_3-...A..... (6)
bru1274_1	.T.....G.....AC.....A.....A..... (3)
bru1274_2-.....C...A... (2)
bru1274_3-..... (2)
bru1487_1	G..G..T.TA.....--CC.....C..... (2)
bru1487_2	G.....T.TA.....-CC..... (5)
bru1487_3T.....A.....-...A..... (2)
bru1487_4A.....-...A..... (3)

Fig. 3. Alignment of cloned *fru* haplotypes found in seven individuals of *Chorthippus brunneus*. Only polymorph sites are shown. Consensus sequences of three species—*C. brunneus*, *C. mollis*, and *C. biguttulus*—are shown above the haplotypes. Haplotypes are named according to the species, the number of the respective individual, and the number of the haplotype found within each individual. Dots indicate identity with the consensus sequence of *C. brunneus*. Dashes indicate indels. Numbers in the three upper lines, read top to bottom, show the positions of polymorph sites in the alignment of the studied fragment of *fru*. Letters in the fourth line indicate untranslated exon (e), intron (i), or coding sequence (c). Numbers in parentheses following each haplotype are the numbers of clones of a particular haplotype. The figure is based on the alignment obtained under the strongest simplification of the pattern of DNA polymorphism, as described in the text.

Assuming that duplication of *fru* took place in a common ancestor of singing grasshoppers and that since duplication each locus had its independent evolutionary history, we expected that orthologous haplotypes should cluster together in a phylogenetic tree. To test this hypothesis, we performed phylogenetic analyses with *C. parallelus*-specific haplotypes as an outgroup (Fig. 4). In contrast to our expectations, haplotypes could not be assigned to distinct loci. Instead, they built species-specific clusters for *C. biguttulus* and *C. brunneus*. Within these clusters the phylogeny was poorly resolved. The haplotypes of *C. mollis* did not cluster in a statistically supported group but diverged at the base of the tree. Generally, bootstrap support values were low, except for the *C. biguttulus* clade. Nevertheless, the obtained tree suggests that paralogous haplotypes within one species are more closely related to each other than to their orthologues in other species. This indicates that copies of *fru* in the grasshopper genomes did not evolve independently of each other.

Clustering of haplotypes in species-specific groups was caused by several indicative positions (Fig. 3, Table 3): *Chorthippus brunneus* and *C. mollis* differed from each other at one site (site 585 in the alignment of *fru* of grasshoppers), while *C. biguttulus* differed from both of them at five sites (sites 328, 409, 467, 694, and 697). This observation was confirmed by direct sequencing of PCR products without cloning

for an additional 4, 11, and 10 individuals of *C. brunneus*, *C. mollis*, and *C. biguttulus*, respectively.

The sole coding substitution was a C-to-A replacement in *C. brunneus* at site 585, which changes asparagine (N) to threonine (T). At this site, asparagine is absolutely conserved in all insect Fru proteins known to date (BLASTp search; see Fig. 2). However, from the three-dimensional structure of the BTB domains it can be concluded that N does not constitute the catalytic core of the BTB domain (Conserved Domain Database [CDD] search [Marchler-Bauer et al. 2005]), and the site is variable in BTB domains of different families of transcription factors (CDD search). This suggests that a replacement at this site should not significantly affect the function of the BTB domain but, rather, might modulate it, e.g., by playing some role in recognition or regulation of transcription of downstream specific targets of Fru. In fruit flies, two targets of Fru have been identified so far, *yellow* and *takeout* (Dauwalder et al. 2002; Drapeau et al. 2003). *yellow* reportedly plays a role in the development of adult male wing extension during courtship (Drapeau et al. 2003).

Another substitution segregating *C. biguttulus* from two other species, which could potentially play some role in expression of Fru, is G to A at position 409, because it lies within the 5' splice junction of the intron harbored within leader 2. It is unclear whether this substitution affects the splicing of the intron in

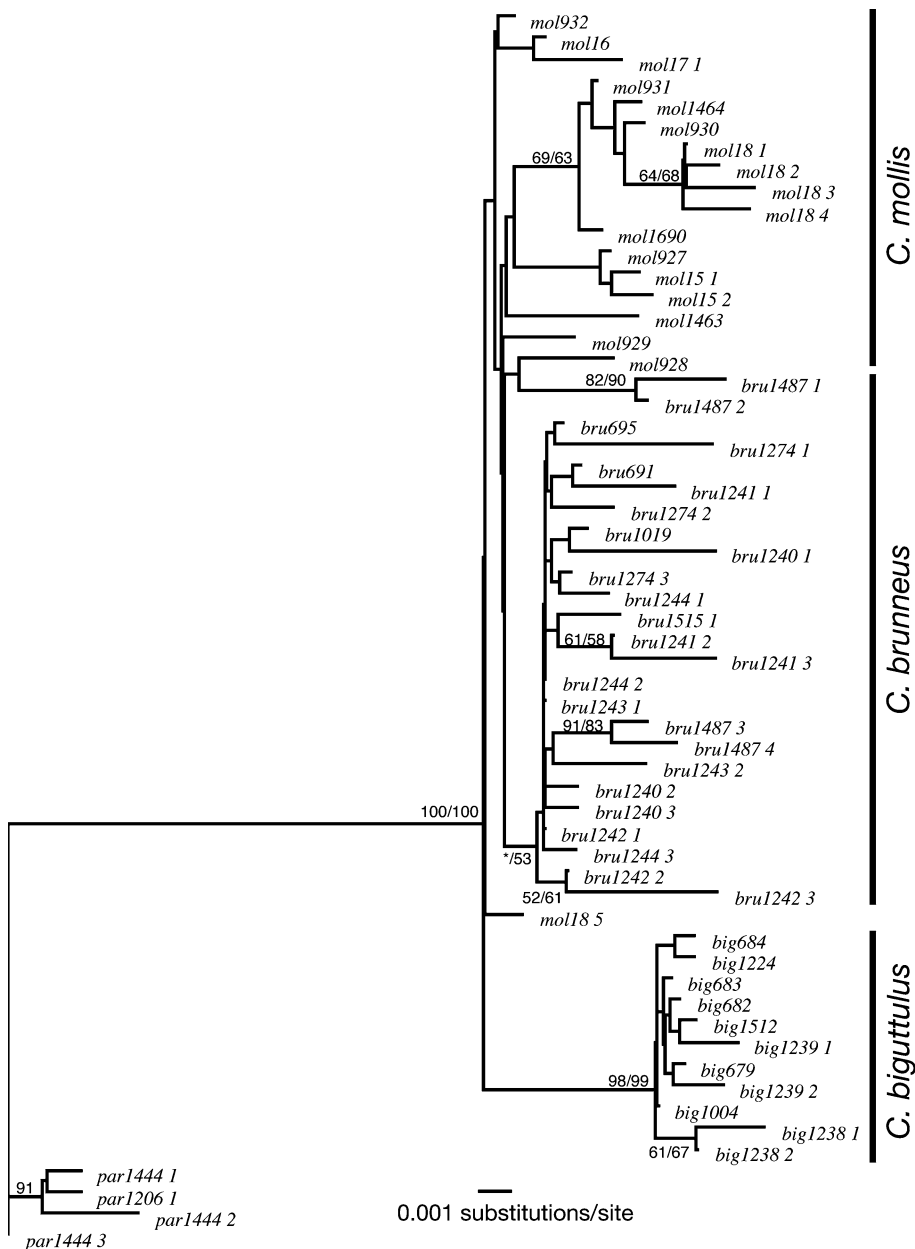


Fig. 4. Neighbor-joining (NJ) tree based on the Kimura two-parameter distances between cloned haplotypes and direct sequences of *fru* from representatives of *Chorthippus*. Each terminal node corresponds to one haplotype or one direct sequence. Each haplotype is named according to the species in which it was found (*big*, *C. biguttulus*; *bru*, *C. brunneus*; *mol*, *C. mollis*; *par*, *C. parallelus*), our internal number for an individual, and a consecutive

number for each haplotype. Incomplete sequences, primers, gapped sites, and singletons are excluded from the analyses, resulting in 897 bp. Numbers near bifurcations show bootstrap values for NJ and maximum parsimony trees, respectively (1000 replications each). Only values >50% are shown. An asterisk shows that the bootstrap support for the bifurcation was < 50% in the NJ tree.

C. biguttulus, but at this position G is conserved in 80% of all introns of insects and mammals (Mount et al. 1992) and thus might be crucial for the 5' splice site recognition.

In addition to five fixed segregating sites, *C. biguttulus* clearly differs from *C. brunneus* and *C. mollis* in the length of the poly(A) track within the intron. It was (A)₉ in 76% of all cloned haplotypes and direct sequences of *C. biguttulus* and (A)₈ in 71% and 73% of all cloned haplotypes and direct sequences of *C. brunneus* and *C. mollis*, respectively.

Variation Within the *fru* Gene Complex

As haplotypes could not be assigned to distinct loci, DNA polymorphism can be described only for the whole *fru* gene complex. Variation was found at 99 sites of the 980-bp sequence after removing of singletons. After strong simplification of the observed pattern of DNA polymorphism (see above), the total number of variable sites decreased to 88. In addition, 37 sites represented four indels, which were located exclusively in the 5' noncoding region. Remarkably, 69

Table 3. State of characters at each almost-alternatively fixed segregating site as percentage of total number of cloned haplotypes and directly sequenced PCR products of *n* individuals

Site	State of character	<i>C. brunneus</i> (<i>n</i> = 12)	<i>C. mollis</i> (<i>n</i> = 14)	<i>C. biguttulus</i> (<i>n</i> = 13)
328	A	97.7	100	
	G	2.3		100
409	G	95.4	100	
	A	2.3		100
	T	2.3		
467	C	100	100	4.8
	T			95.2
585	C	13.3	100	100
	A	82.3		
	G	2.2		
	M ^a	2.2		
694	C	97.8	97.5	
	G	2.2	2.5	100
697	C	97.8	97.5	
	T		2.5	100
	Y ^a	2.2		

^aY, T/C; M, A/C.

of the 88 polymorph sites (78%) were observed in only one individual and 45 sites (51%) were found in only one haplotype. There were more variable sites found in clones per individual in *C. mollis* than in *C. biguttulus* or *C. brunneus* (8.7, 2.3, and 5.6, respectively) and more segregating sites per directly sequenced individual (1.46, 0.1, and 0.25, respectively).

The highest variability was found at synonymous positions within the coding sequence in all species except *C. biguttulus* (Table 4). There was a two- to sevenfold excess of synonymous over nonsynonymous mutations in the three species. Leader 2 and the intron were more variable than leader 1, as they carried twice as many mutations per site and three of four indels. As leader 2 with the intron is excised from transcripts of the first type, the function of the leader 2 might be restricted to a small subset of loci. In the majority of cases it might be recognized as an intron and thus might be less constrained by selection. In contrast, leader 1 accumulated 2.4 times fewer mutations per site than the synonymous portion of the coding sequence. This suggests that leader 1 contains some regulatory signals, which are subject to a selection against deleterious mutations.

Discussion

Structure of *fru* in Grasshoppers

The *fru* of hemimetabolous grasshoppers has many features in common with *fru* of fruit flies and other holometabolous insects. The cloned fragment codes for both domains involved in transcription factor function of Fru proteins (Ryner et al. 1996). These

are the BTB domain at the N-terminus, which is strongly conserved between distantly related insect species (Fig. 2), and at least one Zn-finger domain at the C-terminus. The linker between both domains is much less conserved but resembles *fru* of *Apis mellifera* (accession XP392552), which ends with a Zn-finger domain similar to splice forms B of *fru* of *Drosophila*. The linker in grasshoppers is highly repetitive, thus resembling corresponding linkers in other insect species. This suggests that the domain architecture of Fru proteins is similar in holo- and hemimetabolous insects.

Fru transcripts of grasshoppers are alternatively spliced at the 5' end (or alternatively transcribed from different promoters) (Fig. 1). The usage of different promoters and alternative splicing at the 5' end creates the large variety of transcripts which is crucial for the different functions of *fru* in *Drosophila* (Anand et al. 2001; Goodwin et al. 2000; Lee et al. 2000; Ryner et al. 1996). Only male-specific transcripts from the distal P1 promoter are translated into mature Fru proteins with courtship-relevant functions in many species of *Drosophila* (Lee et al. 2000; Usui-Aoki et al. 2000). Male-specific transcripts encode about 110 more amino acids than female-specific transcripts (Song et al. 2002). Different lengths of male- and female-specific transcripts were also found for *Anopheles gambiae* (GenBank accessions AAU50567–AAU50568, AAV52864–AAV52865 [Gailey et al. 2006]). In contrast to *Drosophila* and *Anopheles*, in grasshoppers alternative starts of transcription and alternative splicing at the 5' end generate two types of transcripts, which differ only in their first noncoding exons. Coding sequences of these transcripts are identical and start at the same site as P4 transcripts of *D. melanogaster* immediately upstream of the BTB domain. The finding of transcripts with different UTRs suggests that their transcription can be differently regulated and that the resulting proteins may have different functions or localization.

It remains an open question whether grasshoppers have other promoters, and whether *fru* is differently spliced in both sexes. Sex-specific splicing of *fru* has recently been shown to be conserved in the dipterans *D. melanogaster* and *A. gambiae*, which diverged about 250 million years ago (Gailey et al. 2006). Therefore, there also is the possibility that sex-specific transcription from more distal promoters occur in grasshoppers. However, sex-specific splicing of *fru* is not universal even within the genus *Drosophila*. First, proteins with male-specific extensions are also translated in females of *D. suzukii* (Yamamoto et al. 2004). Second, the absence of the muscle of Lawrence (MOL), which in *D. melanogaster* is usually induced by loss of male-specific function of *fru*, was found to be “normal” in 67 of 95 *Drosophila* species (Gailey et al. 1997). It was proposed that those *fru* functions

Table 4. Distribution of segregating sites along the cloned fragment of *fru*: Gapped sites are not considered

	<i>n</i>	<i>All species</i>		<i>Chorthippus brunneus</i>		<i>C. mollis</i>		<i>C. biguttulus biguttulus</i>		<i>C. biguttulus eisentrauti</i>	
		<i>s</i>	<i>s/n</i>	<i>s</i>	<i>s/n</i>	<i>s</i>	<i>s/n</i>	<i>s</i>	<i>s/n</i>	<i>s</i>	<i>s/n</i>
Total	942	88	0.093	46	0.049	35	0.037	8	0.008	16	0.017
Leader 1	179	11	0.061	6	0.034	6	0.034	—	—	2	0.011
Leader 2	195	24	0.123	13	0.067	10	0.051	2	0.010	2	0.010
Intron	108	14	0.130	4	0.037	7	0.065	2	0.019	4	0.037
Leader	25	1	0.040	—	—	—	—	—	—	1	0.040
CDS	435	38	0.087	23	0.053	12	0.028	4	0.009	7	0.016
Sense	337	21	0.062	15	0.045	4	0.012	2	0.006	—	—
Silent	98	17	0.173	8	0.082	8	0.082	2	0.02	7	0.071

s, number of segregating sites; *n*, number of sites in a corresponding genomic region; *s/n*, proportion of segregating sites in a corresponding region. The number of sense and silent sites within the coding sequence (CDS) was calculated using the DnaSP 4.0 software.

Table 5. Mean uncorrected *p*-distances (×100) between all haplotypes and direct sequences within species (above diagonal) and between species (below diagonal)

	Noncoding (<i>p</i>)			Nonsynonymous (<i>p</i> N)			Synonymous (<i>p</i> S)		
	<i>big</i>	<i>bru</i>	<i>mol</i>	<i>Big</i>	<i>Bru</i>	<i>mol</i>	<i>big</i>	<i>bru</i>	<i>mol</i>
<i>C. biguttulus</i>	0.18			0.18			0.12		
<i>C. brunneus</i>	1.07	0.49		0.62	0.47		3.23	0.96	
<i>C. mollis</i>	1.31	0.83	0.83	0.14	0.57	0.11	3.57	1.44	1.44

that govern the formation of the MOL are somehow altered in species without the MOL (Yamamoto et al. 2004). In the Hawaiian species *D. heteroneura* and *D. silvestris*, which lack the MOL, male-specific transcripts were not found (Davis et al. 2000a, b). This suggested that sex-specific splicing might not occur in these species and that the loss of the MOL may be caused by the loss of additional male-specific elements upstream of the BTB domain. It was hypothesized that the loss of male-specific elements in the *fru* transcripts of Hawaiian *Drosophila* species was not severe enough to prevent sexual behavior but might have fostered evolution of male–male aggression behavior typical for these species (Davis et al. 2000b). This hypothesis is in agreement with the observation of *D. melanogaster fru* mutants, in which mutations were manifested in aggression-like head interactions between males (Lee and Hall 2000). Interestingly, male rivalry is typical for *Chorthippus* species, where males do not interact physically but sing alternated rival songs, which resemble courtship songs (Jacobs 1953).

Grasshoppers Have Several Paralogues of *fru*

The finding of more than two haplotypes of *fru* per individual suggests that there are several paralogues of *fru* in genomes of grasshoppers. Analyses of cDNA also suggested that more than one locus is tran-

scribed. The number of haplotypes per individual varied substantially, and haplotypes shared by two individuals were scarce. This observation implies that there could be even more *fru* copies per genome than suggested by the observed number of haplotypes per individual. If there were only a limited number of copies, with the low number of sequenced clones (up to 20) we would expect to collect most or all haplotypes from each individual. In this case, alleles at each locus would be sampled according to their intra-population frequencies, and the proportion of haplotypes shared between individuals would be much higher than observed. However, in the case of many similar loci, 10 to 20 sequenced clones per individual would sample only a portion of haplotypes, which would be mostly not allelic.

In the neighbor joining (NJ) tree (Fig. 4) haplotypes of *C. biguttulus* and *C. brunneus* grouped in species-specific clades, primarily due to almost-fixed interspecific differences. Traditionally, concerted evolution was widely used to explain apparent similarities between members of multigene families when paralogous genes are more similar to each other than to their orthologues in other species (reviewed in Nei and Rooney 2005). Under the concerted evolution model (Brown et al. 1972; Zimmer et al. 1980), paralogous loci evolve nonindependently by exchanging genetic information via gene conversion or unequal crossing-over. As concerted evolution can lead to rapid fixations of random nondeleterious mutations

in all duplicates (Innan 2004), there is no need to assume divergent selection creating interspecific differences in *fru* genes of grasshoppers. However, the “birth-and-death” model of evolution of duplicated genes (reviewed in Nei and Rooney 2005) can lead to similar patterns, and the widely accepted concerted evolution hypothesis was called into question recently for many multigene families (e.g., Piontkivska and Nei 2003; Piontkivska et al. 2002; Nei and Rooney 2005; Rooney 2004; Rooney and Ward 2005; Zhang et al. 2003).

According to the birth-and-death model, different haplotypes represent outcomes of different duplication events. Young duplicates are very similar, while older ones degrade by accumulation of point mutations. Fixed interspecific differences can be explained by alternative fixation of different lineages of duplicated loci in each species due to stochastic processes (gene sorting). Alternatively, under the assumption that *fru* influences calling songs in grasshoppers, haplotypes could be subject to divergent sexual selection and purifying selection once acoustic pre-mating isolation among the three species emerged.

Two distinct features are characteristic of multigene families evolving by the birth-and-death process. The first is the presence of a large number of pseudogenes, which represent “dying” copies (e.g., Zhang et al. 2000; Rooney and Ward 2005). For example, genomes of some rodents have more eosinophil-associated RNaseA pseudogenes than functional copies, and orthologous genes, which retained their activity in one species, became pseudogenes in another due to gene sorting (Zhang et al. 2000). On the other hand, under the concerted evolution hypothesis, pseudogenes could arise after a homogenization event, if a single mutation disrupts an ORF.

Although 3 of 37 *fru* haplotypes found in grasshoppers apparently represented pseudogenes, the number of pseudogenes was not large, which is difficult to explain from the birth-and-death point of view, in particular, because the elimination of junk DNA from genomes of grasshoppers is extremely slow (Petrov et al. 2000; Bensasson et al. 2001). We reason that pseudogenes representing old decaying duplicates should have accumulated significantly more point mutations than functional genes. In a NJ tree, such pseudogenes are expected to be on long branches. In contrast, if copies are “young” due to recent birth or concerted evolution, the number of such point mutations should not differ significantly among pseudogenes and functional copies.

Two of the three putative pseudogenes carried two point mutations compared to the species consensus sequence, which is close to the average difference between intact haplotype sequences and consensus sequences. The third pseudogene belonged to a set of

similar haplotypes found in the same individual and differed from the set consensus sequence at two sites. The number of sense mutations in the coding sequence was also not higher in the three pseudogenes than in other haplotypes, and pseudogenes did not occur on long branches in the NJ tree. Therefore, we conclude that pseudogenes did not evolve for a longer period or faster than the rest of haplotypes. Some of the “intact” haplotypes might, however, also be pseudogenes, which are undetected because mutations lie downstream of the sequenced part of *fru*. Based on the pattern found for this first feature it is thus not possible to reject the rapid birth-and-death or the recent homogenization scenario.

The second feature, which is indicative of birth-and-death evolution of multigene protein families, in which copies evolve under strong purifying selection maintaining apparent protein identity, is significant prevalence of synonymous mutations per synonymous site (p_S) over nonsynonymous mutations per nonsynonymous site (p_N) in coding sequences (reviewed in Nei and Rooney 2005). For example, the birth-and-death evolution with strong purifying selection was detected in several gene families, such as histone and ubiquitin genes, in which initially concerted evolution was assumed (Nei et al. 2000; Piontkivska et al. 2002; Rooney et al. 2002). Phylogenies based on p_S showed that in these protein families different copies evolved independently from each other and, as a rule, were not more similar within species than between species (e.g., Piontkivska et al. 2002; Rooney et al. 2002). In contrast, concerted evolution should equally affect synonymous and nonsynonymous sites, and thus the p_S and the p_N should not differ. In phylogenetic trees copies should be always more similar within species than between species (Brown et al. 1972; reviewed in Nei and Rooney 2005).

Because pairwise locus-per-locus comparisons within and between species were not possible, we analyzed mean p -distances for noncoding, synonymous, and nonsynonymous sites (Table 5). For *C. biguttulus* the mean p -distance was equally large at all sites, and there was less synonymous polymorphism than in the two other species. This is in agreement with a rather recent birth of copies or a gene conversion event in the past, followed by accumulation of random mutations. For *C. brunneus* and *C. mollis* the p_S was about twice as large as p_N or p at noncoding sites. However, p_S values were by far not as large as was reported for histone and ubiquitin genes, in which synonymous substitutions supposedly have reached the saturation level (Nei et al. 2000; Piontkivska et al. 2002; Rooney et al. 2002). In pairwise comparisons, copies in *C. brunneus* and *C. mollis* were not more similar within species than between species. In NJ phylogenies based on p_N and

p_S (trees not shown), haplotypes mostly radiated from a base except for a *C. brunneus* clade in a p_N tree and a *C. biguttulus* clade in a p_S tree, which were due to fixed substitutions. Apart from these fixed substitutions, haplotypes were not more similar intraspecifically than interspecifically. This can be interpreted in favor of the birth-and-death and gene sorting hypothesis.

Additionally in support of the birth-and-death hypothesis, we found several individuals with up to two different sets of similar haplotypes. We propose that such pattern might arise, if in the grasshopper genome there are several independent sources of duplications, which are differently located, e.g., on different chromosomes. In contrast, concerted evolution should not lead to such a pattern, because it homogenizes haplotypes also between chromosomes. However, it was proposed that significantly differing rates of homogenization among and within chromosomes explain the evolution of ribosomal intergenic transcribed sequences in the grasshopper *C. parallelus* (Parkin and Butlin 2004). Under this assumption, sets of similar haplotypes could evolve also under the concerted evolution model. Taken together, our data do not unequivocally support either the concerted evolution or the birth-and-death scenario. Yet there are several weak lines of evidence in favor of the birth-and-death model of evolution of *fru* loci in grasshoppers. We also cannot rule out the possibility that the *fru* complex was subject to a mixed process of birth-and-death and concerted evolution.

The fru Genes Show Almost-Fixed Differences Among the Three Species

fru is the first genetic marker, which supports differentiation of the three species at the molecular level. The interspecific differences were consistently found in all directly sequenced PCR products, and deviations were observed in only a few cloned haplotypes. As the sampled grasshopper individuals came from different and distantly located populations, we conclude that the alternatively fixed nucleotides are indicative for the whole species. However, without an extensive comparison with another nuclear marker, it is difficult to say whether such fixations are common across genomes of the three species. To date, we can compare our finding only with DNA polymorphism at a noncoding nuclear locus, *Cpnl-1* (Cooper and Hewitt 1993), which we sequenced in 68 representatives of the three species (unpublished data). At least 22 of 320 sites were variable, but not a single site was fixed or nearly fixed in a species-specific manner. This is in agreement with the only published molecular study of DNA polymorphism in the “*C. biguttulus* group of species” (Mason et al. 1995), which likewise

found a lack of significant differentiation in the mtDNA of the three species of grasshoppers.

A Possible Role of Duplication of fru for Speciation of Grasshoppers

The most likely fate of duplicated genes is relatively quick silencing and degradation (Lynch and Conery 2000), although in some instances duplicates can either take over different functions of the gene or acquire new functions (Ohno 1970). As stated first by Ohno (1970), these two possible fates of duplicated genes make them a primary source for adaptive genome evolution. In *D. melanogaster* alternative splicing and use of alternative promoters generate numerous transcripts with sex-specific and non-sex-specific functions from the *fru* transcription unit (Anand et al. 2001; Goodwin et al. 2000; Lee et al. 2000; Ryner et al. 1996; Song et al. 2002; Song and Taylor 2003). In grasshoppers we did not find a similarly large variety of 5' splice transcripts. There were only two different types of transcripts, although we cannot fully exclude the possibility that further transcripts might be found in future. Instead, we present evidence for the existence of several closely related *fru* paralogues—a remarkable feature not known from other insects. Thus, it could be that different functions of *fru* are adopted by different paralogues in grasshoppers, whereas in fruit flies they are regulated by alternative transcripts of the *fru* gene. If one of the functions of *fru* in grasshoppers is the production of male calling songs, it is possible that duplication of *fru* could have participated in a rapid speciation that is based on divergence of acoustic communication signals without disruption of other functions of *fru*. A rapid birth-and-death process (but not the concerted evolution) could provide a mechanism for the generation of acoustic differences between species by creating copies which can acquire new functions in a short time. However, this scenario is purely speculative at the moment.

Acknowledgments. We thank Andrea Ross for her assistance in the laboratory. We are grateful to Marc Holderied for a critical reading of the manuscript and helpful advice. We also thank two anonymous reviewers for valuable comments that significantly improved the manuscript. This work was supported by a HWP grant to J. U. and by the Deutsche Forschungsgemeinschaft (DFG) grant within the priority program “Radiations—Origins of Biological Diversity” to F.M.

References

- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new

- generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Anand A, Vilella A, Ryner LC, Carlo T, Goodwin SF, Song HJ, Gailey DA, Morales A, Hall JC, Baker BS, Taylor BJ (2001) Molecular genetic dissection of the sex-specific and vital functions of the *Drosophila melanogaster* sex determination gene *fruitless*. *Genetics* 158:1569–1595
- Baker BS, Taylor BJ, Hall JC (2001) Are complex behaviors specified by dedicated regulatory genes? Reasoning from *Drosophila*. *Cell* 105:13–24
- Bensasson D, Petrov DA, Zhang DX, Hartl DL, Hewitt GM (2001) Genomic gigantism: DNA loss is slow in mountain grasshoppers. *Mol Biol Evol* 8:246–253
- Billeter JC, Goodwin SF (2004) Characterization of *Drosophila fruitless-gal4* transgenes reveals expression in male-specific *fruitless* neurons and innervation of male reproductive structures. *J Comp Neurol* 475:270–287
- Brown DD, Wensink PC, Jordan E (1972) *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. *J Mol Biol* 63:57–73
- Cooper SJ, Hewitt GM (1993) Nuclear DNA sequence divergence between parapatric subspecies of the grasshopper *Chorthippus parallelus*. *Insect Mol Biol* 2:185–194
- Dauwalder B, Tsujimoto S, Moss J, 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
- Davis T, Kurihara J, Yamamoto D (2000a) Genomic organisation and characterisation of the neural sex-determination gene *fruitless (fru)* in the Hawaiian species *Drosophila heteroneura*. *Gene* 246:143–149
- Davis T, Kurihara J, Yoshino E, Yamamoto D (2000b) Genomic organisation of the neural sex determination gene *fruitless (fru)* in the Hawaiian species *Drosophila silvestris* and the conservation of the *fru* BTB protein-protein-binding domain throughout evolution. *Hereditas* 132:67–78
- Demir E, Dickson BJ (2005) *fruitless* splicing specifies male courtship behavior in *Drosophila*. *Cell* 121:785–794
- Drapeau MD, Radovic A, Wittkopp PJ, Long AD (2003) A gene necessary for normal male courtship, *yellow*, acts downstream of *fruitless* in the *Drosophila melanogaster* larval brain. *J Neurobiol* 55:53–72
- Gailey DA, Ohshima S, Santiago SJ, Montez JM, Arellano AR, Robillo J, Villarimo CA, Roberts L, Fine E, Vilella A, Hall JC (1997) The muscle of Lawrence in *Drosophila*: a case of repeated evolutionary loss. *Proc Natl Acad Sci USA* 94:4543–4547
- Gailey DA, Ho SK, Ohshima S, Liu JH, Eyassu M, Washington MA, Yamamoto D, Davis T (2000) A phylogeny of the Drosophilidae using the sex-behaviour gene *fruitless*. *Hereditas* 133:81–83
- Gailey DA, Billeter JC, Liu JH, Bauzon F, Allendorfer JB, Goodwin SF (2006) Functional conservation of the *fruitless* male sex determination gene across 250 million years of insect evolution. *Mol Biol Evol* 23(3):633–643
- Goodwin SF, Taylor BJ, Vilella A, Foss M, Ryner LC, Baker BS, Hall JC (2000) Aberrant splicing and altered spatial expression patterns in *fruitless* mutants of *Drosophila melanogaster*. *Genetics* 154:725–745
- Greenspan RJ, Ferreur JF (2000) Courtship in *Drosophila*. *Annu Rev Genet* 34:205–232
- Heinrichs V, Ryner LC, Baker BS (1998) Regulation of sex-specific selection of *fruitless* 5' splice sites by *transformer* and *transformer-2*. *Mol Cell Biol* 18:450–458
- Helversen Ov, Helversen Dv (1994) Forces driving coevolution of song and song recognition in grasshoppers. In: Schildberger K, Elsner N (eds) *Neural basis of behavioural adaptations*. Gustav Fischer Verlag, Stuttgart, pp 253–284
- Innan H (2004) Theories for analyzing polymorphism data in duplicated genes. *Genes Genet Syst* 79:65–75
- Ito H, Fujitani K, Usui K, Shimizu-Nishikawa K, Tanaka S, Yamamoto D (1996) Sexual orientation in *Drosophila* is altered by the satori mutation in the sex-determination gene *fruitless* that encodes a zinc finger protein with a BTB domain. *Proc Natl Acad Sci USA* 93:9687–9692
- Jacobs W (1953) Verhaltensbiologische Studien an Feldheuschrecken. *Z Tierpsychol Beiheft* 1:1–228
- Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 5:150–163
- Kyriacou CP (2002) Single gene mutations in *Drosophila*: What can they tell us about the evolution of sexual behaviour? *Genetica* 116:197–203
- Lawrence PA, Johnston P (1986) The muscle pattern of a segment of *Drosophila* may be determined by neurons and not by contributing myoblasts. *Cell* 45:505–513
- Lee G, Hall JC (2000) A newly uncovered phenotype associated with the *fruitless* gene of *Drosophila melanogaster*: aggression-like head interactions between mutant males. *Behav Genet* 30:263–275
- Lee G, Foss M, Goodwin SF, Carlo T, Taylor BJ, Hall JC (2000) Spatial, temporal, and sexually dimorphic expression patterns of the *fruitless* gene in the *Drosophila* central nervous system. *J Neurobiol* 43:404–426
- Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155
- Manoli DS, Foss M, Vilella A, Taylor BJ, Hall JC, Baker BS (2005) Male-specific *fruitless* specifies the neural substrates of *Drosophila* courtship behaviour. *Nature* 436:395–400
- Marchler-Bauer A, Anderson JB, Cherukuri PF, DeWeese-Scott C, Geer LY, Gwadz M, He S, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Liebert CA, Liu C, Lu F, Marchler GH, Mullokandov M, Shoemaker BA, Simonyan V, Song JS, Thiessen PA, Yamashita RA, Yin JJ, Zhang D, Bryant SH (2005) CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res* 33:D192–D196
- Mason DJ, Butlin RK, Gacesa P (1995) An unusual mitochondrial DNA polymorphism in the *Chorthippus biguttulus* species group (Orthoptera: Acrididae). *Mol Ecol* 4:121–126
- Mount SM, Burks C, Hertz G, Stormo GD, White O, Fields C (1992) Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. *Nucleic Acids Res* 20:4255–4262
- Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* 39:121–152
- Nei M, Rogozin IB, Piontkivska H (2000) Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc Natl Acad Sci USA* 97:10866–10871
- Parkin EJ, Butlin RK (2004) Within- and between-individual sequence variation among ITS1 copies in the meadow grasshopper *Chorthippus parallelus* indicates frequent intrachromosomal gene conversion. *Mol Biol Evol* 21:1595–1601
- Petrov DA, Sangster TA, Johnston JS, Hartl DL, Shaw KL (2000) Evidence for DNA loss as a determinant of genome size. *Science* 287:1060–1062
- Piontkivska H, Nei M (2003) Birth-and-death evolution in primate MHC class I genes: divergence time estimates. *Mol Biol Evol* 20:601–609
- Piontkivska H, Rooney AP, Nei M (2002) Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol Biol Evol* 19:689–697
- Ragge DR (1987) Speciation and biogeography of some southern European Orthoptera, as revealed by their songs. In: Baccetti B (ed) *Evolutionary biology of orthopteroid insects*. Ellis Horwood, Chichester, UK, pp 418–426

- Rooney AP (2004) Mechanisms underlying the evolution and maintenance of functionally heterogeneous 18S rRNA genes in Apicomplexans. *Mol Biol Evol* 21:1704–1711
- Rooney AP, Ward TJ (2005) Evolution of a large ribosomal RNA multigene family in filamentous fungi: birth and death of a concerted evolution paradigm. *Proc Natl Acad Sci USA* 102:5084–5089
- Rooney AP, Piontkivska H, Nei M (2002) Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. *Mol Biol Evol* 19:68–75
- Ryner LC, Goodwin SF, Castrillon DH, Anand A, Vellella A, Baker BS, Hall JC, Taylor BJ, Wasserman SA (1996) Control of male sexual behavior and sexual orientation in *Drosophila* by the *fruitless* gene. *Cell* 87:1079–1089
- Song HJ, Taylor BJ (2003) *fruitless* gene is required to maintain neuronal identity in evenskipped-expressing neurons in the embryonic CNS of *Drosophila*. *J Neurobiol* 55:115–133
- Song HJ, Billeter JC, Reynaud E, Carlo T, Spana EP, Perrimon N, Goodwin SF, Baker BS, Taylor BJ (2002) The *fruitless* gene is required for the proper formation of axonal tracts in the embryonic central nervous system of *Drosophila*. *Genetics* 162:1703–1724
- Swofford DL (2000) PAUP*: Phylogenetic Analysis Using Parsimony (*and other methods). Sinauer, Sunderland: MA
- Usui-Aoki K, Ito H, Ui-Tei K, Takahashi K, Lukacsovich T, Awano W, Nakata H, Piao ZF, Nilsson EE, Tomida J, Yamamoto D (2000) Formation of the male-specific muscle in female *Drosophila* by ectopic *fruitless* expression. *Nat Cell Biol* 2:500–506
- Vellella A, Gailey DA, Berwald B, Ohshima S, Barnes PT, Hall JC (1997) Extended reproductive roles of the *fruitless* gene in *Drosophila melanogaster* revealed by behavioral analysis of new fru mutants. *Genetics* 147:1107–1130
- Yamamoto D, Nakano Y (1999) Sexual behavior mutants revisited: molecular and cellular basis of *Drosophila* mating. *Cell Mol Life Sci* 56:634–646
- Yamamoto D, Usui-Aoki K, Shima S (2004) Male-specific expression of the Fruitless protein is not common to all *Drosophila* species. *Genetica* 120:267–272
- Zhang J, Dyer KD, Rosenberg HF (2000) Evolution of the rodent eosinophil-associated RNase gene family by rapid gene sorting and positive selection. *Proc Natl Acad Sci USA* 97:4701–4706
- Zimmer EA, Martin SL, Beverley SM, Kan YW, Wilson AC (1980) Rapid duplication and loss of genes coding for the chains of hemoglobin. *Proc Natl Acad Sci USA* 77:2158–2162
- Zollman S, Godt D, Privé GG, Couderc JL, Laski FA (1994) The BTB domain, found primarily in zinc finger proteins, defines an evolutionarily conserved family that includes several developmentally regulated genes in *Drosophila*. *Proc Natl Acad Sci USA* 91:10717–10721