Insights into the Regulatory Relationships Implementing Sexual Development in the Honeybee

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Björn Schmitt
aus Heinsberg

Düsseldorf, April 2014

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

aus dem Institut für Evolutionsgenetik
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Martin Beye
Korreferentin: Prof. Dr. Laura Rose

Tag der mündlichen Prüfung: 25. Juni 2014

Eidesstattliche Erklärung

Ich versichere an Eides statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist. Die Dissertation wurde in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Björn Schmitt Düsseldorf, den 7. April 2014

Inhaltsverzeichnis

I. Allgemeine Einleitung

Die geschlechtliche Entwicklung ist ein grundlegender biologischer Prozess, der die verschiedensten Eigenschaften sich sexuell reproduzierender Organismen beeinflusst und somit für die erfolgreiche Reproduktion entscheidend ist. Die binäre Entscheidung, ob sich ein Männchen oder Weibchen entwickelt, bestimmt unter anderem die dimorphe Entwicklung der Gonaden, das geschlechtsspezifische Verhalten und das morphologische Erscheinungsbild. Verschiedene primäre Signale sind für die Geschlechtsdetermination in den unterschiedlichen Lebewesen verantwortlich. Diese Signale umfassen chromosomale Faktoren, wie beispielsweise das X:A-System von *Drosophila melanogaster* (CLINE & MEYER 1996; ERICKSON & QUINTERO 2007) oder das Y-Chromosom vieler Säugetiere, zum Beispiel bei Mensch oder Maus (WILHELM *et al.* 2007), maternale Faktoren, wie bei *Ceratitis capitata* oder *Musca domestica* (GABRIELI *et al.* 2010; HEDIGER *et al.* 2010), Imprinting, zum Beispiel bei *Nasonia vitripennis* (VERHULST *et al.* 2010), oder auch temperaturabhängige Geschlechtsbestimmung, wie sie bei Reptilien vorkommt (BULL 1980). Darüber hinaus findet man die haplodiploide Geschlechtsbestimmung bei etwa 20 % aller Tierarten, darunter die Insektenordnung Hymenoptera zu der auch die Honigbiene (*Apis mellifera*) zählt. Bei dieser Art der Geschlechtsdetermination schlüpfen die Weibchen aus befruchteten Eiern und besitzen einen diploiden Chromosomensatz, wohingegen die Männchen haploid sind und sich aus unbefruchteten Eiern entwickeln (BELL 1982; BULL 1983).

Studien an der Maus sowie an verschiedenen Insekten, wie der Taufliege, der Honigbiene und anderen (BELL *et al.* 1991; PANE *et al.* 2002; SEKIDO & LOVELL-BADGE 2008; GEMPE *et al.* 2009; HEDIGER *et al.* 2010; VERHULST *et al.* 2010; SHUKLA & PALLI 2012), legen nahe, dass die initialen geschlechtsbestimmenden Signale nur dafür benötigt werden, um das Geschlecht in einem kritischen Zeitfenster in der frühen Entwicklung zu determinieren. Diese Information wird dann von einem anderen Mechanismus während der Entwicklung aufrechterhalten. In der Maus kontrolliert das Gen *Sox9* die männliche Entwicklung mittels komplexer regulatorischer Wechselwirkungen (SEKIDO & LOVELL-BADGE 2008; SEKIDO & LOVELL-BADGE 2009). In holometabolen Insekten ist das Gen *transformer* (*tra*) ein konservierter Regulator, der die geschlechtliche Entwicklung implementiert und aufrechterhält (PANE *et al.* 2002; LAGOS *et al.* 2007; RUIZ *et al.* 2007b; CONCHA & SCOTT 2009; GEMPE *et al.* 2009;

HEDIGER *et al.* 2010; VERHULST *et al.* 2010; SCHETELIG *et al.* 2012; SHUKLA & PALLI 2012).

Des Weiteren unterscheidet man grundsätzlich zwei Möglichkeiten, um die geschlechtliche Entwicklung im gesamten Organismus zu etablieren. Die eine stellt dabei die Verwendung von Sexualhormonen dar, die die Ausbildung männlicher oder weiblicher Merkmale steuern. Bei Säugetieren entscheidet daher das Geschlecht der Gonaden über das Geschlecht des ganzen Organismus (CAMERINO *et al.* 2006; WILHELM *et al.* 2007).

Im Gegensatz dazu erfolgt die Entscheidung für weibliche oder männliche Entwicklung bei vielen Insekten unabhängig in jeder Zelle (zellautonom). Einen solchen Mechanismus findet man unter anderem bei der Taufliege und der Honigbiene (DRESCHER & ROTHENBUHLER 1964; SANCHEZ & NOTHIGER 1982; CLINE 1993; GEMPE *et al.* 2009). Bei diesen Organismen muss die zellautonome Regulation des Geschlechts sehr robust erfolgen, da sonst sexuelle Mosaike aus männlichen und weiblichen Zellen entstehen können. So konnten beispielsweise in der Taufliege und der Honigbiene Intersexe induziert werden, die sowohl weibliche als auch männliche Merkmale aufweisen (CLINE 1984; GEMPE *et al.* 2009).

I.I Die komplementäre Geschlechtsbestimmung der Honigbiene

Die Honigbiene (*Apis mellifera*) verfügt mit der komplementären Geschlechtsbestimmung über eine spezielle Form der haplodiploiden Geschlechtsbestimmung, die innerhalb der Hymenopteren weit verbreitet ist (COOK 1993; HEIMPEL & DE BOER 2008). Hierbei entscheidet weder der Ploidiegrad noch ein Befruchtungsereignis über das Geschlecht, sondern der Genotyp an einem bestimmten Locus. In der Honigbiene wurde der zu Grunde liegende geschlechtsbestimmende Faktor, das Gen *complementary sex determiner* (*csd*), identifiziert (BEYE *et al.* 2003), von dem mehr als 53 Allele auf Artebene segregieren (LECHNER *et al.* 2013).

Es entwickeln sich nur Tiere zu Weibchen (Arbeiterinnen oder Königinnen), die heterozygot für *csd* sind (Abb. 1). Die Entscheidung, ob aus einer weiblichen Larve eine Königin entsteht, wird dabei durch die Arbeiterinnen gefällt, da Königinnenlarven von diesen ein spezielles Futter erhalten (KUCHARSKI *et al.* 2008; KAMAKURA 2011). Haploide oder diploide Tiere, die einen hemi- bzw. homozygoten

csd-Genotyp besitzen (Abb. 1), entwickeln sich dagegen zu Männchen (Drohnen), wobei die diploiden Drohnen bereits kurz nach dem Schlupf von den Arbeiterinnen aufgefressen werden (WOYKE 1963).

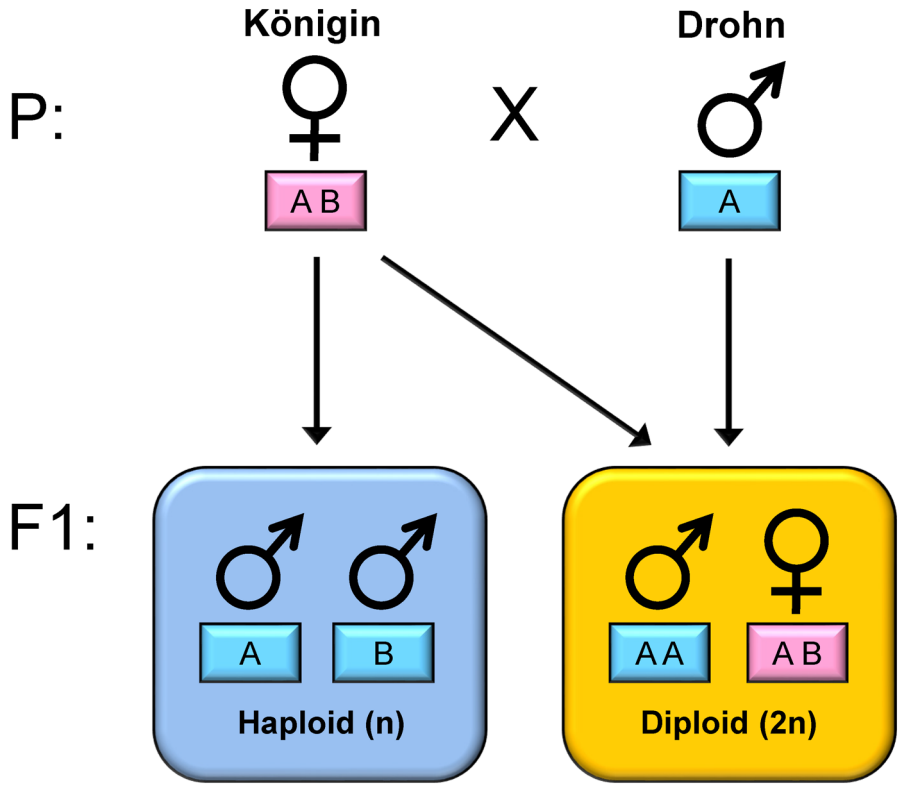

Abbildung 1: Komplementäre Geschlechtsbestimmung der Honigbiene. Die allelische Kombination des *csd*-Gens entscheidet über das Geschlecht der Honigbiene. Individuen entwickeln sich nur dann zu Weibchen, wenn sie heterozygot für *csd* sind (AB). Dagegen entwickeln sich Individuen, die hemizygot (A) oder homozygot (AA) für *csd* sind, zu Männchen.

Es konnte mittels RNA-Interferenz (RNAi)-Experimenten gezeigt werden, dass der heterozygote Zustand die aktive Form von *csd* darstellt. So führte ein Ausschalten der *csd*-Genfunktion zu männlicher Entwicklung von genetisch weiblichen Tieren, wohingegen die Entwicklung genetischer Männchen nicht beeinflusst wurde (BEYE *et al.* 2003).

Die von den einzelnen *csd*-Allelen kodierten Csd-Proteine gehören zur Klasse der SR-Typ-Proteine und unterscheiden sich durchschnittlich in ca. 3 % ihrer Aminosäuren (HASSELMANN & BEYE 2004). Ein Großteil der allelischen Diversität resultiert aus Polymorphismen im Bereich der potentiell spezifitätsgebenden Domäne (PSD), die unter anderem eine hypervariable Region (HVR) umfasst (HASSELMANN & BEYE 2004; HASSELMANN *et al.* 2008a; BEYE *et al.* 2013). Die HVR weist zwischen den einzelnen Allelen Längenpolymorphismen in Form von Asparagin-Tyrosin-Wiederholungen auf (BEYE *et al.* 2003; HASSELMANN *et al.* 2008a). Die PSD umfasst

zudem noch eine arginin-/serinreiche (RS)-Domäne sowie einen Teil der prolinreichen (PR)-Domäne. Während die PR-Domäne und die RS-Domäne sowie ein im 5‘-Bereich der RS-Domäne liegendes CC-Motiv (coiled-coil, eine spezielle Form der Doppelhelix) mögliche Kandidaten für Protein-Protein-Interaktionen sind (BEYE *et al.* 2003; BEYE 2004; HASSELMANN *et al.* 2008a), konnte dem relativ invariablen N-Terminus bisher keine Funktion zugeordnet werden.

Das *csd*-Gen ist durch Genduplikation aus einem Vorläufer des Gens *feminizer* (*fem*) hervorgegangen und diesem funktional übergeordnet (HASSELMANN *et al.* 2008b; GEMPE *et al.* 2009). Aktive Csd-Proteine induzieren in Weibchen die Bildung von weiblich gespleißten Transkripten des *fem*-Gens. Diese kodieren für das Fem-Protein, das ebenfalls zur Familie der SR-Typ-Proteine gehört, über eine PR-Domäne sowie zwei RS-Domänen verfügt (HASSELMANN *et al.* 2008b) und ein Ortholog des *D. melanogaster* Tra-Proteins ist (GEMPE *et al.* 2009). Die Fem-Proteine kontrollieren das Spleißen weiblicher mRNA des Gens *Am-dsx* (*Apis mellifera doublesex*) und implementieren die weibliche Entwicklung. RNAi-Experimente haben gezeigt, dass Csd-Proteine die weibliche Entwicklung nur innerhalb eines kritischen Zeitfensters während der frühen Entwicklung induzieren können (GEMPE *et al.* 2009). Danach erhalten die Fem-Proteine die weibliche Entwicklung aufrecht und kontrollieren ihre eigene Expression durch die autoregulative Prozessierung weiterer weiblicher *fem*-Transkripte (GEMPE *et al.* 2009). In Männchen führt die Abwesenheit aktiver Csd-Proteine zum Spleißen männlicher *fem*-Transkripte, die ein vorzeitiges Stoppkodon enthalten und somit für kein funktionales Protein kodieren (GEMPE *et al.* 2009). Ohne Fem-Proteine kommt es zur Bildung männlich gespleißter *Am-dsx*-mRNAs und männlicher Entwicklung.

Sowohl die weiblichen als auch die männlichen Transkripte des *Am-dsx*-Gens kodieren für Transkriptionsfaktoren vom DM-Typ, die eine charakteristische zinkfingerartige Domäne für DNA-Interaktionen besitzen und jeweils über ein spezifisches C-terminales Ende verfügen (DEARDEN *et al.* 2006; CHO *et al.* 2007).

Ein weiteres Gen, das in der Geschlechtsdeterminationskaskade der Honigbiene eine wichtige Rolle spielt, ist das *Am-tra2*-Gen. Es konnte mittels RNAi-Experimenten gezeigt werden, dass die *Am-tra2*-Transkripte sowohl für die Bildung weiblich gespleißter *Am-dsx*-Transkripte als auch für die Prozessierung weiblicher und männlicher *fem*-Transkripte benötigt werden (NISSEN *et al.* 2012a). Die Transkripte des *Am-tra2*-Gens werden nicht geschlechtsspezifisch gespleißt und

kodieren für sechs Isoformen eines SR-Typ Spleißfaktors, der über eine RNA-Bindedomäne (RBD) und zwei RS-Domänen verfügt, wie Tra2-Orthologe in anderen Spezies (NISSEN *et al.* 2012a).

I.II Konservierte Elemente in den Geschlechtsdeterminationskaskaden verschiedener Insekten

Obwohl sich die primären Signale in *A. mellifera*, *C. capitata*, *M. domestica*, *N. vitripennis* und *T. castaneum* unterscheiden (PANE *et al.* 2002; GEMPE *et al.* 2009; HEDIGER *et al.* 2010; VERHULST *et al.* 2010; SHUKLA & PALLI 2012), weisen die Geschlechtsdeterminationskaskaden dieser Insekten deutliche Ähnlichkeiten auf (Abb. 2). Eine zentrale Rolle in der Geschlechtsdeterminationskaskade holometaboler Insekten nimmt das konservierte Schaltergen *fem*/*tra* ein. Die *tra*-Transkripte werden, wie die orthologen *fem*-Transkripte in der Honigbiene, geschlechtsspezifisch gespleißt und nur die weiblichen *fem/tra*-Transkripte kodieren für funktionelle Fem/Tra-Proteine. Diese implementieren die weibliche Entwicklung durch die Bildung weiblich gespleißter *dsx*-mRNA und halten diese durch das autoregulative Spleißen weiterer weiblicher *fem/tra*-mRNAs aufrecht (PANE *et al.* 2002; GEMPE *et al.* 2009; HEDIGER *et al.* 2010; VERHULST *et al.* 2010; SHUKLA & PALLI 2012). Sind keine Fem/Tra-Proteine vorhanden, wird männlich gespleißte *dsx*-mRNA gebildet und es erfolgt eine männliche Entwicklung. Die geschlechtsregulierende Funktion der *dsx*-Transkripte konnte in vielen Spezies nachgewiesen werden (BURTIS & BAKER 1989; COSCHIGANO & WENSINK 1993; SUZUKI *et al.* 2003; HEDIGER *et al.* 2004; WILLIAMS & CARROLL 2009). Eine konservierte Funktion haben in diesen Spezies zudem die Tra2-Proteine, die für die Prozessierung weiblicher *tra*- und *dsx*-Transkripte benötigt werden (NISSEN *et al.* 2012a; SHUKLA & PALLI 2013).

Ein ähnliches und gut charakterisiertes regulatives Netzwerk (NAGOSHI *et al.* 1988; BURTIS & BAKER 1989; VALCARCEL *et al.* 1993; CLINE & MEYER 1996), bestehend aus *tra*, *tra2* und *dsx*, findet sich auch in *D. melanogaster* (Abb. 2). Hier wird die weibliche Entwicklung jedoch vom Gen *Sex-lethal* (*Sxl*) aufrechterhalten, da die Sxl-Proteine das Spleißen weiblicher *tra*-mRNAs und ihrer eigenen Transkripte kontrollieren (BELL *et al.* 1991; KEYES *et al.* 1992).

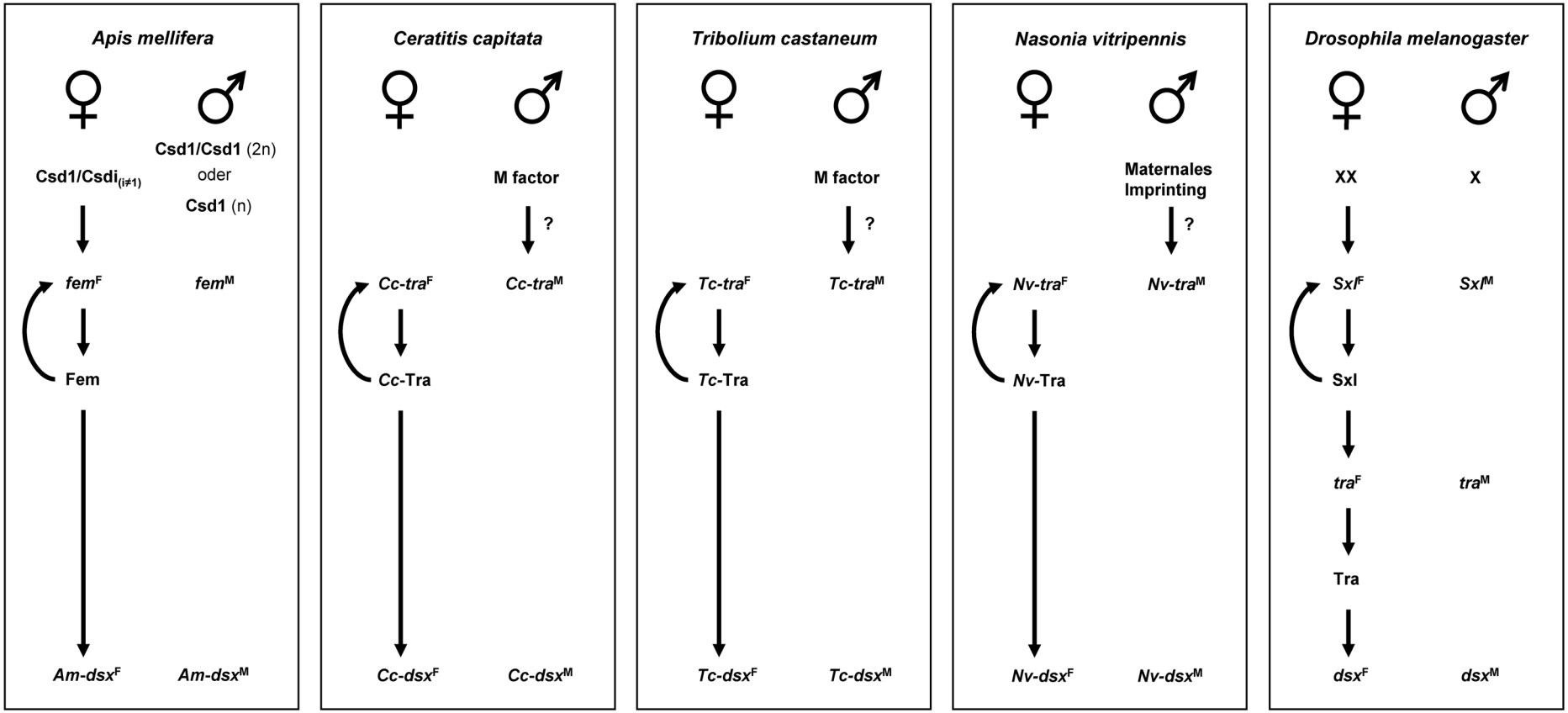

Abbildung 2: Vergleichende Darstellung der Geschlechtsdeterminationskaskaden verschiedener Insekten. Obwohl sich die primären Signale der Geschlechtsdetermination der dargestellten Organismen unterscheiden, teilen sie doch wesentliche regulatorische Elemente. So ist das Schaltergen *tra/fem* in allen gezeigten Spezies für die Implementierung des weiblichen Entwicklungsweges zuständig. Die Tra/Fem-Proteine kontrollieren die Prozessierung weiblicher *dsx*-Transkritpe. Sind keine Tra/Fem-Proteine vorhanden, werden männliche *dsx*-Transkripte produziert und es erfolgt eine männliche Entwicklung. Die Tra/Fem-Proteine halten außerdem den weiblich determinierten Status aufrecht, indem sie autoregulativ das Spleißen ihrer eigenen Transkripte kontrollieren. Die einzige Ausnahme stellt die Taufliege (*Drosophila melanogaster*) dar. Hier hat das Sxl-Protein die autoregulative Funkltion zur Aufrechterhaltung des weiblichen Zustands übernommen. Abbildung modifiziert nach GEMPE & BEYE (2011).

An diesen Beispielen ist ersichtlich, dass die regulatorischen Prinzipien der Geschlechtsdetermination sowie die involvierten Gene zwischen Vertretern verschiedener Insektenordnungen konserviert sind, während sich die initialen Signale unterscheiden. Dies wird auch als „Bottom-Up“ (Von-unten-nach-oben) Evolution der Geschlechtsdeterminationskaskade bezeichnet, da sich die primären Signale der Kaskade verändern, die untergeordneten Komponenten jedoch erhalten bleiben (WILKINS 1995; BEYE 2004).

I.III Genduplikation spielt eine wichtige Rolle bei der Evolution von Geschlechtsdeterminationskaskaden

Während das Schaltergen *fem*/*tra*, der Spleißregulator Tra2 und der geschlechtsspezifische Transkriptionsfaktor Dsx eine konservierte Funktion in der Geschlechtsbestimmung holometaboler Insekten haben, wurden andere Gene im Laufe der Evolution neu rekrutiert. Es konnte gezeigt werden, dass Genduplikation entscheidend zur Evolution neuer Komponenten der Geschlechts-determinationskaskaden beigetragen hat. Ein Beispiel hierfür ist das *Sxl*-Gen der Taufliege, das wahrscheinlich zusammen mit seinem Schwestergen *ssx* (*sister-of-Sex-lethal*) aus einem gemeinsamen Vorläufer in der Familie der Drosophilidae durch Duplikation entstanden ist und danach seine Funktion als geschlechtsbestimmendes Schaltergen von *Drosophila* erlangt hat (TRAUT *et al.* 2006; CLINE *et al.* 2010). Ein weiteres Beispiel ist das *csd*-Gen der Honigbiene, das durch Genduplikation aus einem Vorläufer des *fem*-Gens hervorgegangen und zum primären Signal der Geschlechtsdetermination evolviert ist (BEYE *et al.* 2003; HASSELMANN *et al.* 2008b). Es wurden zudem kürzlich Duplikate des *fem*/*tra*-Gens in den Genomen von Hummeln und Ameisen gefunden, die möglicherweise das primäre Signal der komplementären Geschlechtsbestimmung in diesen Arten darstellen, was bisher jedoch nicht funktional getestet wurde (SCHMIEDER *et al.* 2012).

Die Duplikation einzelner Gene, Chromosomenabschnitte oder ganzer Genome (Polyploidisierung) stellt eine potentielle Quelle für die Evolution neuer Genfunktionen dar (OHNO 1970; HOLLAND *et al.* 1994; SIDOW 1996). Bei der Genduplikation kommt es beispielsweise in Folge eines ungleichen Crossing-overs zur Verdopplung eines Chromosomenabschnitts. Verschiedene Studien an Genen, die in mehreren Kopien im Genom vorliegen, haben gezeigt, dass neue Kopien

konstant durch Genduplikation entstehen (NEI *et al.* 1997; PIONTKIVSKA *et al.* 2002; EIRIN-LOPEZ *et al.* 2004; NIIMURA & NEI 2005; ROBERTSON & WANNER 2006; ROBERTSON *et al.* 2010). Es wurde geschätzt, dass die Häufigkeit von Genduplikationen bei Eukaryoten 0,01 pro Gen und Millionen Jahre beträgt (LYNCH & CONERY 2000; LIPINSKI *et al.* 2011).

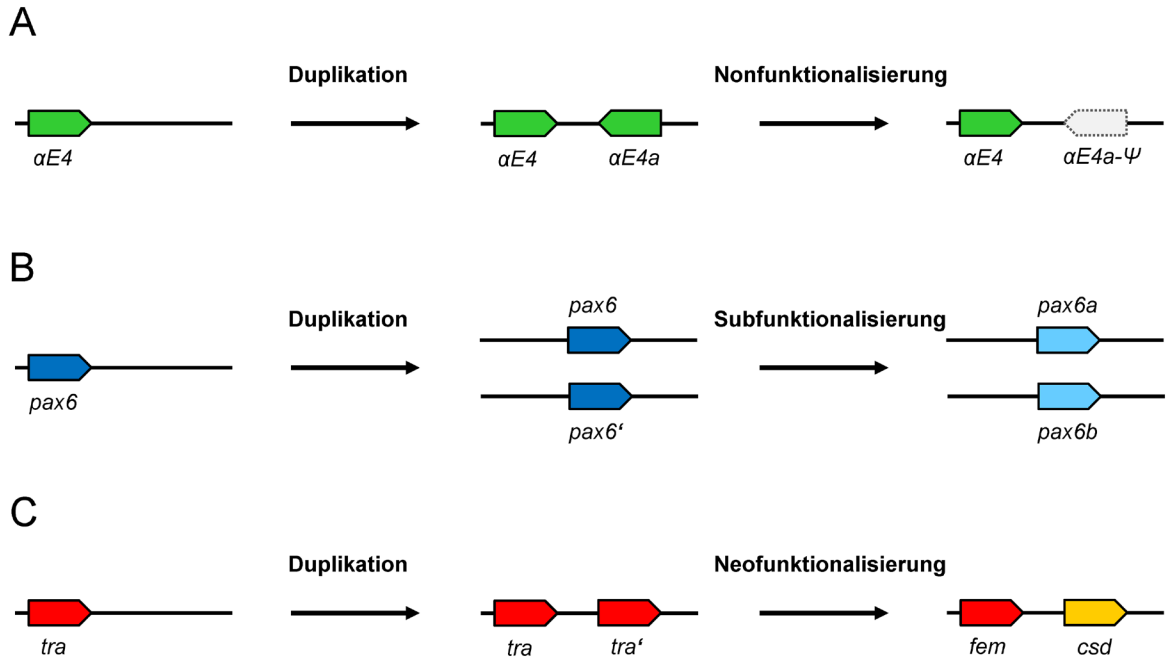

Abbildung 3: Die Evolution duplizierter Gene. Es werden drei Szenarien unterschieden, die das evolutionäre Schicksal duplizierter Gene beschreiben. Bei der Nonfunktionalisierung (A) wird eine Genkopie in Folge der Akkumulation degenerativer Mutationen ausgeschaltet. Dies ist am Beispiel des Pseudogens *αE4a-Ψ* der Taufliege dargestellt (ROBIN *et al.* 2000). Bei der Subfunktionalisierung (B) werden beide Genkopien durch Mutationen teilweise kompromittiert, so dass beide Kopien nötig sind, um die vollständige Genfunktion des anzestralen Vorläufers zu gewährleisten. Dies ist bei den beiden *pax6*-Genen des Zebrafischs der Fall (*pax6a* und *pax6b*), die auf Grund des Verlusts *cis*-regulatorischer Elemente in unterschiedlichen Geweben exprimiert werden (KLEINJAN *et al.* 2008). Im Zuge der Neofunktionalisierung (C) erlangt eine Genkopie eine neue Funktion, während die andere Kopie die ursprüngliche Funktion beibehält. So evolvierte das *csd*-Gen der Honigbiene zum initialen Signal der Geschlechtsdetermination, während das *fem*-Gen die anzestrale geschlechtsregulierende Funktion beibehielt (BEYE *et al.* 2003; HASSELMANN *et al.* 2008b).

Die Genkopien können nach der Duplikation auf unterschiedliche Weise evolvieren (LYNCH & CONERY 2000). Da die beiden Kopien zunächst funktional redundant sind, führt die Akkumulation degenerativer Mutationen in den meisten Fällen zur Inaktivierung (Nonfunktionalisierung) einer Kopie innerhalb weniger Millionen Jahre (WATTERSON 1983; LYNCH & FORCE 2000). Ein Beispiel für ein Gen, das erst vor relativ kurzer Zeit seine Funktion verloren hat, ist das Pseudogen *αE4a-Ψ* (*α-Esterase 4a-Ψ*) von *D. melanogaster* (Abb. 3A). Dieses ist durch

Genduplikation vor etwa 30-40 Millionen Jahren aus dem *αE4*-Gen entstanden (ROBIN *et al.* 2000) und wurde innerhalb der letzten 2,5 Millionen Jahre in der *D.-melanogaster*-Linie inaktiviert (ROBIN *et al.* 2000).

Es gibt jedoch auch den Fall, dass beide Genkopien durch Mutationen teilweise kompromittiert werden (Subfunktionalisierung) und nur die kombinierte Aktivität der Duplikate ausreichend ist, die ursprüngliche Genfunktion zu gewährleisten (LYNCH & FORCE 2000). So verfügt der Zebrafisch (*Danio rerio*) über zwei Kopien des *pax6*-Gens, *pax6a* und *pax6b* (Abb. 3B), die in Folge einer Genomduplikation in einem Vorfahren der Knochenfische vor 350 Millionen Jahren entstanden sind (KLEINJAN *et al.* 2008). Die *pax6a*- und *pax6b*-Gene des Zebrafischs werden bedingt durch den Verlust *cis*-regulatorischer Elemente in unterschiedlichen Geweben exprimiert, während ihre Funktion in den meisten Vertebraten von einem einzigen *PAX6*-Gen erfüllt wird (KLEINJAN *et al.* 2008).

Schließlich ist es auch möglich, dass eine Genkopie eine neue Funktion erlangt (Neofunktionalisierung), während die andere Kopie die ursprüngliche Funktion beibehält (WALSH 1995). Dieser Prozess wird dadurch begünstigt, dass nach der Duplikation auf Grund ihrer funktionalen Redundanz ein geringerer Selektionsdruck auf die Genkopien wirkt (LYNCH & CONERY 2000).

Ein gut untersuchtes Beispiel für eine solche Neofunktionalisierung ist die Evolution des *csd*-Gens der Honigbiene (Abb. 3C), das durch Genduplikation aus einem Vorläufer des *fem*-Gens vor etwa 10 bis 70 Millionen Jahren entstanden ist (HASSELMANN *et al.* 2008b). Die paralogen Csd- und Fem-Proteine weisen eine Ähnlichkeit von ungefähr 70 % auf Aminosäureebene auf (HASSELMANN *et al.* 2008b) und beide sind essentiell für die Geschlechtsbestimmung der Honigbiene (BEYE *et al.* 2003; GEMPE *et al.* 2009). Dennoch unterscheiden sie sich in wesentlichen Punkten. Die *csd*-prä-mRNA wird nicht geschlechtsspezifisch gespleißt und die kodierten Csd-Proteine führen, abhängig von ihrer allelischen Komposition, zu männlicher oder weiblicher Entwicklung (BEYE *et al.* 2003). Demgegenüber werden die *fem*-Transkripte geschlechtsspezifisch gespleißt; nur die weibliche *fem*-mRNA kodiert für das Fem-Protein, das die weibliche Determination implementiert und aufrechterhält (GEMPE *et al.* 2009). Während das *fem*-Gen somit durch negative Selektion seine konservierte Funktion als Schaltergen in der Geschlechtsdeterminationskaskade beibehalten hat (Abb. 4), kam es beim *csd*-Gen in Folge positiver (gerichteter)

Selektion zur Neofunktionalisierung als primäres Signal der Honigbienengeschlechtsdetermination (HASSELMANN *et al.* 2008b).

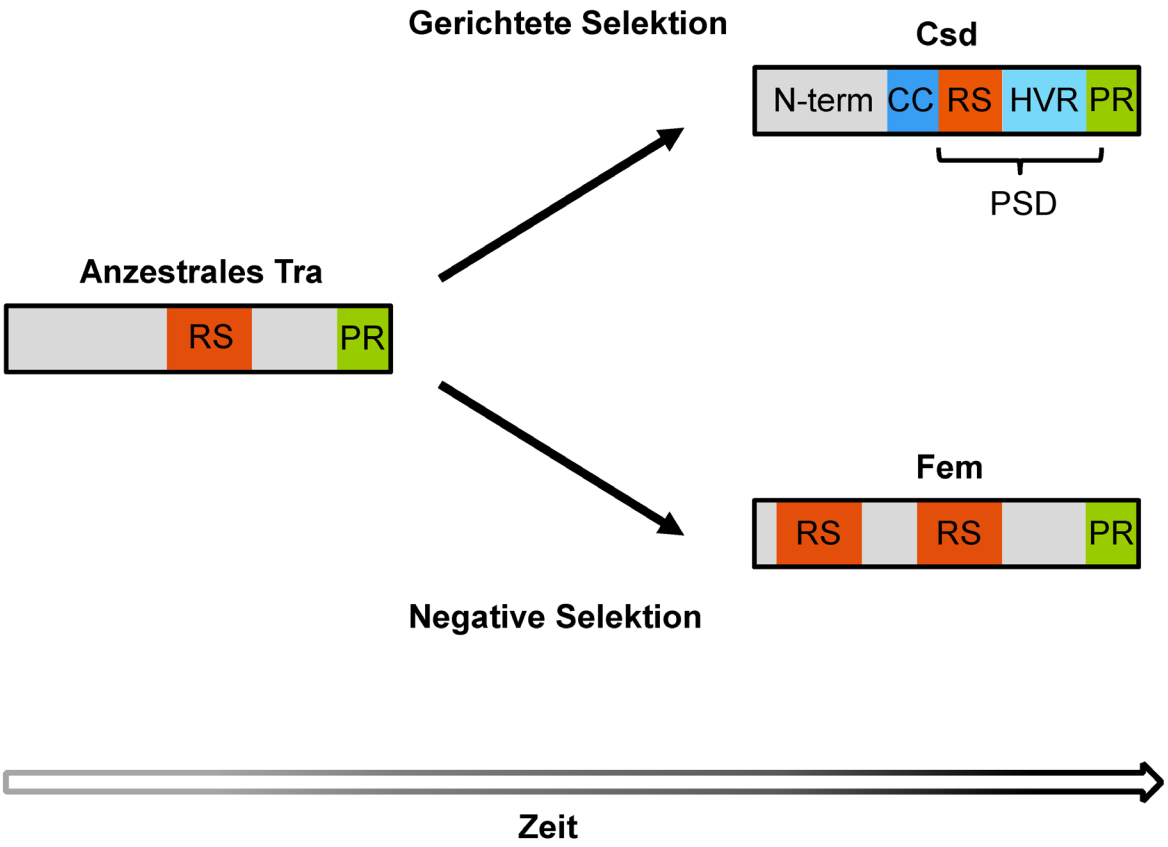

Abbildung 4: Die Entstehung von *csd* durch Genduplikation aus einem Vorläufer des *fem* Gens in der Honigbienenlinie. Die kodierten Proteine sind schematisch dargestellt. Die Csd- und Fem-Proteine gehen auf einen gemeinsamen Vorfahren (anzestrales Tra-Protein) zurück, mit dem sie den grundlegenden Aufbau bestehend aus einer arginin-/serinreichen (RS)-Domäne und einer prolinreichen (PR)-Domäne teilen. Während das *fem*-Gen unter negativer Selektion evolvierte und die anzestrale geschlechtsregulierende Funktion beibehielt, evolvierte das paraloge *csd*-Gen im Zuge der Neofunktionalisierung als primäres Signal durch gerichtete Selektion u.a. eine hypervariable Region (HVR) und eine putative CC-(Coiled-Coil)-Domäne. Abbildung modifiziert nach GEMPE & BEYE (2011).

In jüngster Zeit konnten weitere Duplikate des *fem/tra*-Gens in einigen Vertretern der Hummeln und Ameisen identifiziert werden, deren Funktion bisher unbekannt ist (SCHMIEDER *et al.* 2012). Es ist zudem unklar, ob diese Gene mehrfach unabhängig durch Genduplikation im Verlauf der Evolution der Hymenopteren entstanden sind, oder ob sie auf ein einziges Duplikationsereignis zurückzuführen und konzertiert evolviert sind. Bei der konzertierten Evolution werden paraloge Sequenzen in Folge des wiederholten Auftretens ungleicher Crossing-over oder durch Genkonversion einander angeglichen (LIAO 1999; NEI & ROONEY 2005; CHEN *et al.* 2007; INNAN 2009). Deshalb erscheinen konzertiert evolvierte paraloge Gene innerhalb einer Art ähnlicher zueinander als zu den jeweiligen orthologen Genen in

anderen Spezies. Studien haben gezeigt, dass Genduplikation ein häufiger Prozess ist, da ein Gen im Schnitt alle 100 Millionen Jahre dupliziert wird (LYNCH & CONERY 2000; LIPINSKI *et al.* 2011). Dagegen war Genkonversion in einer anderen Studie nur in 2 % aller duplizierten Gene nachweisbar und erforderte zudem einen geringen Abstand zwischen den Genkopien (weniger als 9 Kilobasenpaare) (SEMPLE & WOLFE 1999).

I.IV Die Geschlechtsbestimmung der Insekten wird über alternatives Spleißen reguliert

Verschiedene Gene, die in den Geschlechtsdeterminationskaskaden diverser Insekten involviert sind, werden alternativ gespleißt. Dabei kann durch geschlechtsspezifisches alternatives Spleißen die Genfunktion aus- oder eingeschaltet werden. So kodieren beispielsweise nur die weiblichen Transkripte der Schaltergene *Sxl* und *fem/tra* für ein funktionales Protein, das die weibliche Entwicklung implementiert. Die männlichen *Sxl*- bzw. *fem/tra*-Transkripte enthalten dagegen ein vorzeitiges Stoppkodon und kodieren für kein funktionales Protein. Außerdem können durch alternatives Spleißen Transkripte erzeugt werden, die für unterschiedliche Proteinisoformen kodieren. Ein Beispiel hierfür sind die geschlechtsspezifischen Transkripte des *dsx*-Gens, die für Varianten eines Transkriptionsfaktors kodieren; diese weibchen- bzw. männchenspezifischen Dsx-Proteine kontrollieren wiederum die geschlechtliche Entwicklung vieler Insekten. Das alternative Spleißen dieser Gene ist somit essentiell für eine korrekte männliche oder weibliche Entwicklung vieler Organismen.

I.IV.I Der Spleißprozess eukaryotischer prä-mRNA

Beim Spleißen werden die nicht-kodierenden Introns aus dem Primärtranskript (prä-mRNA) eines Gens entfernt und die Exons zu einer kodierenden Sequenz verbunden. Die Spleißreaktion, bei der es sich um zwei aufeinanderfolgende Transesterreaktionen handelt, wird von einem Ribonukleoprotein-Komplex, dem Spleißosom, im Zellkern katalysiert. Der Komplex besteht aus fünf kleinen nuklearen RNAs (snRNAs = small nuclear RNAs) und einer größeren Zahl assoziierter Proteine (REED 2000; STEVENS *et al.* 2002; ZHOU *et al.* 2002). Das Spleißosom erkennt die 5'-Spleißstelle und die 3'-Spleißstelle, die das Intron flankieren (MANIATIS & TASIC 2002).

Dabei kommt es zunächst zu einer Duplexbildung zwischen der 5‘-Spleißstelle und dem freien Ende der U1-snRNA (KAMMLER *et al.* 2001; FREUND *et al.* 2003), die durch die anderen Komponenten des U1-snRNPs (small nuclear ribonucleoproteins) stabilisiert wird (DU & ROSBASH 2002). Danach bindet das SF1-(splice factor1) Protein an die Verzweigungsstelle, während U2AF65 und U2AF35, die beiden Untereinheiten des U2AF (U2 auxiliary factor) Proteins, an den Polypyrimidintrakt bzw. an die 3‘-Spleißstelle binden (GRAVELEY 2000; MANIATIS & TASIC 2002; BLACK 2003; TURUNEN *et al.* 2013). Im Folgenden wird das SF1-Protein durch das U2-snRNP ersetzt, wonach das Spleißosom durch das U4/U6.U5-tri-snRNP vervollständigt wird (GRAVELEY 2000; MANIATIS & TASIC 2002; BLACK 2003; TURUNEN *et al.* 2013). Schließlich kommt es zu einer Reorganisation des Spleißosoms, die zur Bildung des aktiven Spleißkomplexes führt (GRAVELEY 2000; BLACK 2003; TURUNEN *et al.* 2013). Dieser katalysiert die Freisetzung des Introns (als Lariatstruktur) sowie die Verknüpfung der beiden freien Exonenden (MANIATIS & TASIC 2002; BLACK 2003).

Durch alternatives Spleißen können aus einer prä-mRNA verschiedene Transkripte generiert werden, die für unterschiedliche Proteine kodieren (MANIATIS & TASIC 2002; BLACK 2003) oder auch vorzeitige Stoppkodons enthalten und degradiert werden (LEWIS *et al.* 2003; LAREAU *et al.* 2004). Dies wird durch die differentielle Nutzung von Spleißstellen erreicht, wodurch es zur Verlängerung bzw. Verkürzung oder Entfernung von Exons sowie zum Einschluss von Introns in der mRNA kommen kann (CARTEGNI *et al.* 2002).

Eine entscheidende Voraussetzung sowohl für das konstitutive als auch für das alternative Spleißen einer prä-mRNA ist die Definition der Exons durch die Erkennung der 5‘- bzw. 3‘-Spleißstellen (MANIATIS & TASIC 2002; BLACK 2003; TURUNEN *et al.* 2013). Man unterscheidet kanonische und nicht-kanonische Spleißstellen. Bei den kanonischen Spleißstellen, die mit schätzungsweise 99 % den Großteil der Spleißstellen ausmachen, beginnt das Intron mit dem Dinukleotid GT und endet mit AG (BURSET *et al.* 2000). Bei den restlichen 1 %, den nicht-kanonischen Spleißstellen, markieren andere Dinukleotide den Beginn und das Ende eines Introns (BURSET *et al.* 2000).

Abgesehen von dieser grundsätzlichen Einteilung in kanonische und nicht-kanonische Spleißstellen, weisen die Sequenzen der 5‘- und 3‘-Spleißstellen mehr oder weniger Ähnlichkeit zu der jeweiligen Konsensussequenz auf. Es konnte für 5‘-Spleißstellen gezeigt werden, dass eine hohe Komplementarität zur U1-snRNA

ausreichend ist, um die Spleißstelle zu aktivieren und das Spleißosom zu rekrutieren, während 5‘-Spleißstellen mit einer niedrigen Komplementarität inaktiv bleiben (KAMMLER *et al.* 2001; FREUND *et al.* 2003).

Es gibt auch noch den Sonderfall, dass Introns nicht von den U1/U2-snRNAs erkannt werden. Bei dieser Art des Spleißens, bei der ein gesonderter Ribonukleoproteinkomplex (das Minor Spleißosom) bestehend aus vier spezifischen snRNAs die Spleißreaktion katalysiert, erkennen die U11- bzw. U12-snRNA die 5‘-Spleißstelle bzw. die Verzweigungsstelle (TURUNEN *et al.* 2013). Zwar werden weniger als 0,05 % aller Introns vom Minor Spleißosom gespleißt, diese scheinen jedoch eine funktionelle Bedeutung zu besitzen (TURUNEN *et al.* 2013). So sind die 5‘-Spleißstelle und die Verzweigungsstelle dieser Introns stärker konserviert als in den meisten anderen Introns (BURGE *et al.* 1998), zudem fehlt ihnen ein erkennbarer Polypyrimidintrakt (TURUNEN *et al.* 2013).

Unabhängig davon, welches Spleißosom die prä-mRNA prozessiert, kann die Aktivität der Spleißstellen durch regulative Elemente modifiziert werden (GRAVELEY 2000; MANIATIS & TASIC 2002; TURUNEN *et al.* 2013).

I.IV.II Die Regulation des alternativen Spleißens

Es gibt sowohl im Exon als auch im Intron aktivierende (ESE/ISE, exonic/intronic splicing enhancers) und reprimierende Elemente (ESS/ISS, exonic/intronic splicing silencers) (GRAVELEY 2000; MANIATIS & TASIC 2002; BLACK 2003). Diese Sequenzelemente dienen als Bindestellen von Spleißfaktoren wie den spleißregulatorischen SR-Proteinen oder heterogenen Ribonukleoproteinen (hnRNP = heterogeneous ribonucleoprotein), die nicht Bestandteil des normalen Spleißosoms sind. Je nachdem, ob sie im Exon oder im Intron binden, wirken die Vertreter dieser Proteinfamilien aktivierend oder reprimierend auf die regulierte Spleißstelle. Dabei wirken SR-Proteine generell aktivierend, wenn sie im Exon, und reprimierend, wenn sie im Intron binden; dies ist bei den hnRNP-Proteinen genau umgekehrt (GRAVELEY 2000; ERKELENZ *et al.* 2013a).

Die SR-Proteine haben einen charakteristischen strukturellen Aufbau, bestehend aus ein bis zwei N-terminalen RNA-Bindedomänen (RBD) und einer C-terminalen RS-Domäne von variabler Länge (GRAVELEY 2000). Eine Gruppe weiterer Proteine, die über eine RS-Domäne verfügen, sich jedoch von den SR-Proteinen

unterscheiden, sind die SR-ähnlichen oder SR-Typ-Proteine (GRAVELEY 2000). Einige von ihnen, wie die Tra2-Proteine (BURTIS & BAKER 1989; HEDLEY & MANIATIS 1991; RYNER & BAKER 1991; INOUE *et al.* 1992; HOFMANN *et al.* 2000; SCIABICA & HERTEL 2006; QI *et al.* 2007), können direkt an die RNA binden, während andere keine RBD besitzen und ihre Funktion zusammen mit anderen Spleißfaktoren ausüben. Beispiele hierfür sind die Tra/Fem-Proteine verschiedener Insekten (BURTIS & BAKER 1989; PANE *et al.* 2002; SCIABICA & HERTEL 2006; LAGOS *et al.* 2007; RUIZ *et al.* 2007b; CONCHA & SCOTT 2009; GEMPE *et al.* 2009; HEDIGER *et al.* 2010; VERHULST *et al.* 2010; SCHETELIG *et al.* 2012; SHUKLA & PALLI 2012) sowie die Csd-Proteine der Honigbiene (BEYE *et al.* 2003).

Die molekulare Funktion der Tra-Proteine wurde detailliert in *D. melanogaster* untersucht. Hier binden die Proteine Tra und Tra2 zusammen mit dem SR-Protein 9G8 an das *dsx*RE (*dsx* repeat element), ein aus sechs Wiederholungen eines 13-nt-langen ESE Motivs bestehendes regulatives Element, um das Spleißen weiblicher *dsx*-mRNA zu induzieren (BURTIS & BAKER 1989; HEDLEY & MANIATIS 1991; RYNER & BAKER 1991; INOUE *et al.* 1992; SCIABICA & HERTEL 2006). Die ESE Sequenzen kontrollieren dabei die Aktivierung der 3‘ Spleißstelle des Exons 4 der *dsx*-prä-mRNA (BURTIS & BAKER 1989; HEDLEY & MANIATIS 1991; INOUE *et al.* 1992; TIAN & MANIATIS 1993; LYNCH & MANIATIS 1995; LYNCH & MANIATIS 1996; SCIABICA & HERTEL 2006). Ein ähnliches Sequenzmotiv, das *fru*RE (*fru* repeat element), reguliert in der Taufliege das Tra/Tra2-abhängige Spleißen weiblicher Transkripte des Gens *fru* (*fruitless*). Das Tra2-Protein kontrolliert auch die eigene Expression in den Testes männlicher Taufliegen durch negative Autoregulation, ein Vorgang, der für die Spermatogenese essentiell ist (MATTOX & BAKER 1991; MATTOX *et al.* 1996; MCGUFFIN *et al.* 1998). Die Tra2-Proteine binden dabei an ein regulatorisches Element, das aus fünf 5-nt-langen Wiederholungen eines ISS-Motivs besteht (QI *et al.* 2007). Dies führt zum Einschluss von Intron M1 und zur Bildung von *tra2*-mRNA, die für kein funktionales Protein kodiert (MCGUFFIN *et al.* 1998).

Sequenzmotive, die dem *dsx*RE ähnlich sind, wurden in den *dsx*-Genen verschiedener weiterer Dipteren identifiziert (HEDIGER *et al.* 2004; LAGOS *et al.* 2005; RUIZ *et al.* 2007a; SACCONE *et al.* 2008), weshalb diese Form des *dsx*-Spleißens möglicherweise in der ganzen Ordnung Diptera verbreitet ist. Es wurden zudem *dsx*RE-ähnliche Sequenzen in den *tra*-Genen der Dipteren *A. suspense, B. oleae, C. capitata, M. domestica* gefunden, weshalb das autoregulatorische Spleißen

weiblicher *tra*-mRNA möglicherweise ebenfalls durch Bindung von Tra- und Tra2-Proteinen an die *tra*-prä-mRNA induziert wird (LAGOS *et al.* 2007; SALVEMINI *et al.* 2009; HEDIGER *et al.* 2010; SCHETELIG *et al.* 2012).

Bisher wurden keine Sequenzmotive in der Honigbiene identifiziert, die dem *dsx*RE ähneln (NISSEN *et al.* 2012a). Es konnte aber gezeigt werden, dass die Proteine Fem und *Am*-Tra2 das Spleißen weiblicher *Am-dsx*-Transkripte induzieren (NISSEN *et al.* 2012b). Hierfür scheint eine ESE-Sequenz benötigt zu werden, die vor der weibchenspezifischen 5‘ Spleißstelle der *dsx*-prä-mRNA liegt (NISSEN *et al.* 2012b). Dagegen ist über die Regulation des geschlechtsspezifischen Spleißens der *fem*-prä-mRNA bisher nur wenig bekannt. Zwar wurde mittels RNAi-Experimenten gezeigt, dass die Proteine Csd, Fem und *Am*-Tra2 das geschlechtsspezifische Spleißen der *fem*-Transkripte kontrollieren; die zu Grunde liegenden molekularen Mechanismen, die zur Prozessierung weiblicher bzw. männlicher *fem*-mRNA führen und so die geschlechtliche Entwicklung implementieren, wurden bisher nicht identifiziert.

I.V Zielsetzung

In holometabolen Insekten wird die Entscheidung für männliche oder weibliche Entwicklung durch das geschlechtsspezifische Spleißen der prä-mRNA des konservierten Schaltergens *fem*/*tra* implementiert. In der Honigbiene bestimmt die allelische Komposition des *csd*-Gens, welche geschlechtsspezifischen *fem*-Transkripte gebildet werden. Das *csd*-Gen ist durch Genduplikation aus einem Vorläufer des *fem*-Gens in der Linie der Honigbienen entstanden und evolvierte in Folge positiver Selektion zum primären Signal der komplementären Geschlechtsbestimmung. In jüngster Zeit konnten Duplikate des *fem*-Gens (*fem1* genannt) in den Genomen verschiedener Hummel- und Ameisenarten nachgewiesen werden; es ist jedoch unklar, ob diese Duplikate in den letzten 120 Millionen Jahren mehrmals unabhängig voneinander entstanden oder konzertiert evolviert sind. Die *fem1*-Gene stellen möglicherweise die primären Signale der komplementären Geschlechtsdetermination in diesen Spezies dar, es fehlen hierzu jedoch funktionale Daten. Funktionell wurde dies bisher nur für das *csd*-Gen der Honigbiene nachgewiesen (BEYE *et al.* 2003).

Die Csd-Proteine sind nur im heterozygoten Zustand aktiv und induzieren die Prozessierung weiblicher *fem*-mRNAs. Diese kodieren für die funktionalen Fem Proteine, die das Spleißen weiblicher *Am-dsx*-Transkripte kontrollieren und die weibliche Entwicklung implementieren. Die Fem-Proteine halten auch den weiblichen Entwicklungsweg aufrecht, indem sie das Spleißen weiterer weiblicher *fem*-mRNA autoregulativ kontrollieren. Im homo-/hemizygoten Zustand sind die Csd-Proteine inaktiv, weshalb ausschließlich die nicht für ein Protein kodierenden männlichen *fem*-Transkripte gebildet werden. Die Abwesenheit von Fem-Proteinen führt wiederum zur Prozessierung männlicher *Am-dsx*-Transkripte und männlicher Entwicklung. Über die molekulare Regulation des *fem*-Spleißens ist bisher nur wenig bekannt. Es konnte jedoch in einer früheren Studie gezeigt werden, dass der Spleißfaktor *Am*-Tra2 für die Prozessierung weiblicher *Am-dsx*-Transkripte, sowie männlicher und weiblicher *fem*-Transkripte benötigt wird.

In dieser Arbeit soll die molekulare Regulation der weiblichen bzw. männlichen Entwicklung der Honigbiene untersucht werden, die durch die geschlechtsspezifische Prozessierung der *fem*-Transkripte implementiert wird. Dafür soll zunächst, analog zu früheren Experimenten, die weibliche Entwicklung in männlichen Tieren induziert

werden. Hierzu werden Csd- bzw. Fem-Proteine transient in männlichen Embryonen exprimiert. Der Nachweis weiblicher Transkripte des *fem*- bzw. des *Am-dsx*-Gens dient als Indikator für eine weibliche Entwicklung. Zudem soll der Mechanismus des geschlechtsspezifischen *fem*-Spleißens untersucht werden. Hierzu soll anhand eines *fem*-Minigens *in vivo* in *Sf*21-Insektenzellen untersucht werden, welchen Einfluss die Expression von Csd-, Fem- und *Am*-Tra2-Proteinen auf die Regulation der weiblichen bzw. männlichen Spleißstellen der *fem*-prä-mRNA hat. Weiterhin soll untersucht werden, ob das *Am*-Tra2-Protein ein möglicher Kofaktor von Csd-Proteinen und/oder Fem-Proteinen bei der Regulation des geschlechtsspezifischen *fem*-Spleißens ist.

Ziel der Arbeit ist es zudem zu klären, ob die *csd/fem1*-Gene der Hummeln, Honigbienen und Ameisen mehrmals unabhängig voneinander entstanden sind, oder ob sie auf ein gemeinsames ursprüngliches Duplikationsereignis zurückgehen und durch konzertierte Evolution den jeweiligen *fem*-Genen angeglichen wurden. Es soll daher untersucht werden, ob Aminosäuresubstitutionen in den anzestralen Sequenzen des letzten gemeinsamen Vorfahren der Ameisen und Bienen spezifisch für die jeweiligen Csd/Fem1- oder Fem-Proteine sind, was einen Rückschluss darauf ermöglicht, wie lange diese Gene unabhängig voneinander evolviert sind. Außerdem soll mittels verschiedener Sequenzanalyseprogramme nach Spuren von Sequenzaustauschen in Folge konzertierter Evolution zwischen den *csd/fem1*- und den paralogen *fem*-Genen gesucht werden. Die Suche nach weiteren Duplikaten der *fem*- und *fem1*/*csd*-Gene in den Genomen von Bienen und Ameisen soll zudem Hinweise liefern, ob wiederholte Genduplikation ein häufiger Prozess in der Evolution dieser Gene ist.

I.VI Thesen

Folgende Thesen können im Rahmen dieser Arbeit aufgestellt werden:

- Die transiente Expression der Proteine Csd oder Fem induziert temporär das Spleißen weiblicher *Am-dsx*- bzw. *fem*-Transkripte in Männchen, führt jedoch nicht zur weiblichen Entwicklung. Hieraus folgt, dass die Fem-Autoregulation ein langsamer Prozess ist und durch ein kurzes weibliches Signal nicht vollständig aktiviert werden kann.

- Das *Am*-Tra2-Protein ist ausreichend, um männchenspezifisches Spleißen der prä-mRNA eines *fem*-Minigenkonstrukts in *Sf*21-Insektenzellen zu induzieren. Das *Am-tra2*-Gen kontrolliert somit die Bildung männlicher *fem*-Transkripte.

- Die pRR1-Sequenz in der *fem*-prä-mRNA fungiert als *Am*-Tra2-abhängiges ESE und dient als putative Bindestelle der *Am*-Tra2-Proteine bei der Regulation des männchenspezifischen *fem*-Spleißens.

- *Am*-Tra2-Proteine interagieren mit Fem-Proteinen, aber nicht mit Csd-Proteinen. Hieraus folgt, dass *Am*-Tra2-Proteine putative Kofaktoren der spleißregulierenden Fem-Proteine sind, nicht jedoch der geschlechts-bestimmenden Csd-Proteine.

- *Am*-Tra2-Proteine haben eine duale Funktion bei der Regulation des geschlechtsspezifischen Spleißens der *fem*-Transkripte, da sie direkt das Spleißen der männlichen *fem*-mRNA kontrollieren und als Kofaktoren der Fem-Proteine an deren autoregulatorischer Produktion beteiligt sind. Eine vergleichbare duale Funktion ist von Orthologen des *Am*-Tra2-Proteins in anderen Organismen nicht bekannt und könnte in der Honigbiene neu entstanden sein.

- Die Wechselwirkung zwischen der Fem-Autoregulation und der dualen Funktion des *Am*-Tra2-Proteins könnte als Filter fungieren, um eine Aktivierung der Fem-Produktion in männlichen Tieren zu verhindern.

- Konzertierte Evolution (gegenseitiger Sequenzaustausch) zwischen den paralogen Genpaaren *fem1*/*csd* und *fem* hat keinen nachweisbaren Beitrag zur Evolution dieser Gene geleistet.

- Die *fem1*/*csd*-Gene der Hummeln, Honigbienen und Ameisen sind im Laufe der Evolution durch Genduplikation mehrmals unabhängig voneinander entstanden. Folglich wird die in den Hymenopteren weit verbreitete komplementäre Geschlechtsbestimmung von unterschiedlichen Genen kontrolliert.

II. Manuscripts

Manuscript I

Regulation of Honeybee Sex-Determination: The Dual Role of *Am*-Tra2 Proteins

Björn D. Schmitt, Inga Nissen, Marianne Otte and Martin Beye

Institute of Evolutionary Genetics, Heinrich-Heine University Duesseldorf, Duesseldorf, Germany

Manuscript under review in Developmental Cell

Abstract

Despite a variable initial signal, sex-determination systems in insects follow certain general regulatory principles. Sexual development is implemented through the sex-specific splicing of the *fem/tra* transcripts, which leads to the production of male and female *dsx* transcripts and sexual differentiation.

In the honeybee, the allelic composition of the primary *csd* gene is responsible for the production of the sex-specific transcripts of the *fem* gene. The female *fem* mRNA encodes the functional Fem protein that implements and maintains female development and controls its own production in a positive feedback loop. Male development occurs by default if just the non-productive male *fem* transcripts are produced.

Here, we show that transient expression of Csd or Fem proteins induces a temporary female response in males; however, this response is not sufficient to fully induce female development. Therefore, we conclude that the *fem* feedback loop acts slowly and requires the prolonged presence of the Csd or Fem proteins for full activation. Furthermore, we demonstrate that the *Am*-Tra2 proteins, which do not exhibit sex-specific expression, control the production of male *fem* transcripts, most likely by binding to ESE sequences in the *fem* pre-mRNA. We show that the *Am*-Tra2 protein interacts with Fem but not with Csd proteins and therefore has a dual role in the production of sex-specific *fem* transcripts. We propose that the functional relationship of the slow feedback loop and the dual role of the *Am*-Tra2 protein serves to evaluate the commitment of the *csd* gene, which is not a present/absent signal but has different activities depending on its allelic composition. We propose that this evaluation filters out putative female background activity resulting from hemi-/homozygous Csd proteins and therefore robustly establishes stable sex-determined states in every cell in females and males.

Introduction

Sex-determination mechanisms and the molecules involved differ substantially between species. Studies on the mouse (*Mus musculus*), fruit fly (*Drosophila melanogaster*), honeybee (*Apis mellifera*) and other insects (BELL *et al.* 1991; PANE *et al.* 2002; SEKIDO & LOVELL-BADGE 2008; GEMPE *et al.* 2009; HEDIGER *et al.* 2010; VERHULST *et al.* 2010; SHUKLA & PALLI 2012) suggest that various sex-determination mechanisms generate a primary sex-determining signal that is only required for a short period of time during development and that this sex-specific information is then maintained by a different mechanism throughout development.

In the mouse, male development is initiated by the *Sry* gene. However, the gene *Sox9* is thought to maintain male development in a complex network of positive feedback and feedforward loops that drive its own expression throughout male development (SEKIDO & LOVELL-BADGE 2008; SEKIDO & LOVELL-BADGE 2009).

In holometabolous insects, the binary switch gene *transformer* (*tra*) is a common regulator that implements and maintains sexual development (PANE *et al.* 2002; LAGOS *et al.* 2007; RUIZ *et al.* 2007b; CONCHA & SCOTT 2009; GEMPE *et al.* 2009; HEDIGER *et al.* 2010; VERHULST *et al.* 2010; SCHETELIG *et al.* 2012; SHUKLA & PALLI 2012). Female-specifically spliced *tra* transcripts encode the functional Tra proteins whose production is ensured by a positive feedback loop. Through this feedback loop, the *tra* switch is locked in a female- or male-determined state and provides a continuous signal that maintains sexual identity. A primary target of the Tra protein is the *doublesex* (*dsx*) gene (CLINE & MEYER 1996; PANE *et al.* 2002; HEDIGER *et al.* 2004; LAGOS *et al.* 2007; HASSELMANN *et al.* 2008b; ALVAREZ *et al.* 2009; CONCHA & SCOTT 2009; GEMPE *et al.* 2009; SALVEMINI *et al.* 2009; HEDIGER *et al.* 2010; VERHULST *et al.* 2010; SHUKLA & PALLI 2012) whose role in sex-determination has been functionally shown in several species (BURTIS & BAKER 1989; COSCHIGANO & WENSINK 1993; SUZUKI *et al.* 2003; HEDIGER *et al.* 2004; WILLIAMS & CARROLL 2009). The *dsx* transcripts are sex-specifically spliced, and they encode DM-type transcription factors with sex-specific carboxy-terminal ends (CLINE & MEYER 1996; DEARDEN *et al.* 2006; CHO *et al.* 2007).

The entire insect order of hymenoptera has haplodiploid sex-determination (BULL 1983); haploid males develop from unfertilized and diploid females from fertilized eggs. However, in the honeybee, *Apis mellifera*, the allelic composition of

the *csd* gene determines sexual fate (BEYE *et al.* 2003). Female development depends on the activity of the Csd proteins that belong to the SR-type protein family and segregate in at least 15 major allelic variants in honeybee populations (BEYE 2004; HASSELMANN *et al.* 2008b) that differ on average by ~3 % of their amino acid residues (HASSELMANN & BEYE 2004). The Csd proteins are only active if they are derived from a heterozygous genotype. Consequently, only diploid individuals that are heterozygous for *csd* will develop into female workers and queens, whereas haploid hemizygous or diploid homozygous individuals develop into males.

The Csd proteins can induce femaleness only in a critical period during early development after which sexual identity is maintained by the *feminizer* (*fem*) gene (GEMPE *et al.* 2009). In females, the Csd proteins derived from a heterozygous genotype direct the splicing of pre-messenger RNAs (pre-mRNAs) of the *fem* gene into the female form that encodes the Fem protein, an SR-type protein, which is an ortholog of the *D. melanogaster* Tra protein (HASSELMANN *et al.* 2008b; GEMPE *et al.* 2009). When active Csd proteins are absent in the hemi- and homozygous *csd* genotype, *fem* transcripts are spliced into the non-productive male form that contains a premature termination codon (PTC) in exon 3 and thus does not yield a functional protein (HASSELMANN *et al.* 2008b).

The Fem protein implements female development and is required for the production of female *Am-dsx* transcripts (GEMPE *et al.* 2009). Absence of the Fem protein results in male development and the default production of male *Am-dsx* transcripts (GEMPE *et al.* 2009). Similar to the transcripts of the *Drosophila melanogaster* ortholog *dsx*, both sex-specific *Am-dsx* transcripts encode a DM-type transcription factor with a sex-specific carboxy-terminal end (CLINE & MEYER 1996; DEARDEN *et al.* 2006; CHO *et al.* 2007).

The Fem proteins also maintain their own expression by a positive feedback loop in which they promote the splicing of *fem* transcripts into the productive female form (GEMPE *et al.* 2009). Thus, Fem proteins maintain the sex-determined state independently in every cell as it was shown by the occurrence of sexual mosaics in *fem*-repressed females (GEMPE *et al.* 2009).

RNAi studies showed that the *Am*-Tra2 protein, which shares a common structure of two RS domains and a conserved RNA-binding domain (RBD) with its *Drosophila* ortholog, is required for the production of female *Am-dsx* transcripts and both female and male *fem* transcripts (NISSEN *et al.* 2012a). However, sequence

motifs with homology to the *dsx* repeat elements (*dsx*REs) of *Drosophila* are absent in the *fem* and *Am-dsx* genes, and it was suggested that the binding sites of the *Am*-Tra2 protein might have evolved in the honeybee (NISSEN *et al.* 2012a).

In the well-studied dipteran *D. melanogaster*, the number of X chromosomes determines the sex (CLINE & MEYER 1996; ERICKSON & QUINTERO 2007). The presence of 2 X chromosomes leads to the production of functional Sex-lethal (Sxl) proteins that in turn implement the female state by promoting the splicing of functional *tra* transcripts through the repression of the male 3' splice site in exon 2 (VALCARCEL *et al.* 1993). The continuous supply of functional Tra proteins, which in turn ensures the production of female Dsx proteins (NAGOSHI *et al.* 1988; BURTIS & BAKER 1989), is established in the fruitfly by a positive feedback loop at the level of the Sxl protein that maintains its own expression throughout female development (BELL *et al.* 1991; KEYES *et al.* 1992). In the presence of 1 X chromosome, no functional Sxl proteins and subsequently, no functional Tra proteins are produced, which results in the production of male Dsx proteins and male development.

The molecular function of Tra proteins has been thoroughly studied in *D. melanogaster dsx* transcript splicing, where the Tra protein acts together with the Tra2 protein to facilitate splicing of female *dsx* transcripts by binding to the *dsx*RE, a set of six 13-nt-long repeat elements in the *dsx* pre-mRNA (BURTIS & BAKER 1989; HEDLEY & MANIATIS 1991; RYNER & BAKER 1991; INOUE *et al.* 1992; SCIABICA & HERTEL 2006). These repeat elements act as exonic splice enhancers (ESEs) to direct the activation of the 3' splice site of *dsx* exon 4 (BURTIS & BAKER 1989; HEDLEY & MANIATIS 1991; INOUE *et al.* 1992; TIAN & MANIATIS 1993; LYNCH & MANIATIS 1995; LYNCH & MANIATIS 1996; SCIABICA & HERTEL 2006). This mode of *dsx* control appears to be conserved in the order diptera because motifs similar to the *dsx*RE ESEs have been found in the *dsx* orthologs of *Bactrocera oleae, Ceratitis capitata, Musca domestica* and the *Anastrepha* species (HEDIGER *et al.* 2004; LAGOS *et al.* 2005; RUIZ *et al.* 2007a; SACCONE *et al.* 2008).

In *D. melanogaster*, the Tra2 protein is not required for somatic male-specific regulation but controls the sex-specific splicing of transcripts of the genes *exuperantia* (*exu*) and *alternative testis transcripts* (*att*) in the male germline (HAZELRIGG & TU 1994; MADIGAN *et al.* 1996; MATTOX *et al.* 1996). The Tra2 protein also regulates its own expression in the testes of male fruit flies by a negative feedback loop and thus limits the amount of available Tra2 protein, which is critical

for proper spermatogenesis (MATTOX & BAKER 1991; MATTOX *et al.* 1996; MCGUFFIN *et al.* 1998). In the testes, the Tra2 protein binds to five 5-nt-long repeat elements in the *tra2* pre-mRNA that act as intronic splicing silencers (ISS) to repress the adjacent 5' and 3' splice sites (QI *et al.* 2007), which leads to the inclusion of intron M1 and consequently a nonfunctional product (MCGUFFIN *et al.* 1998).

Other dipterans such as *C. capitata* and *M. domestica* use a different mechanism to establish and maintain stably determined sexual states. It was shown that the female *tra* transcripts are supplied maternally in *Musca* (HEDIGER *et al.* 2010), which has also been suggested in *Ceratitis* (GABRIELI *et al.* 2010). The maternally provided Tra proteins implement female development by directing the female splicing of *dsx* transcripts (PANE *et al.* 2002; HEDIGER *et al.* 2010). Male development occurs if a dominant male-determining factor (called the M factor) is present that interferes with the maternally provided *tra* signal leading to the production of male *dsx* transcripts (PANE *et al.* 2002; HEDIGER *et al.* 2010). In *C. capitata* and *M. domestica* females, the Tra proteins, like the Fem proteins in the honeybee, stably establish and maintain female development by acting together with Tra2 proteins in a positive feedback loop that facilitates the female-specific splicing of the zygotically transcribed *tra* transcripts (PANE *et al.* 2002; HEDIGER *et al.* 2010). It was suggested that the Tra and Tra2 proteins cooperatively bind to sequence motifs in the *tra* pre-mRNA similar to the *dsx*REs that have been found in the *tra* genes of several dipteran insects including *A. suspense, B. oleae, C. capitata, M. domestica*, and thus direct the autoregulatory splicing of *tra* transcripts in the female mode (LAGOS *et al.* 2007; SALVEMINI *et al.* 2009; HEDIGER *et al.* 2010; SCHETELIG *et al.* 2012).

It was recently proposed that in the coleopteran *Tribolium castaneum*, a similar mechanism involving maternally provided female *tra* transcripts, Tra/Tra2 protein-dependent *tra* autoregulation and a male-determining M factor is responsible for establishing and maintaining sexual development through an ortholog of the *dsx* gene (SHUKLA & PALLI 2012; SHUKLA & PALLI 2013). Functional data indicate that the Tra2 protein is required to suppress production of the female *tra* transcripts in males (SHUKLA & PALLI 2013).

In the hymenopteran *Nasonia vitripennis*, the mechanism that implements and maintains female development is similar to that of *C. capitata* and *M. domestica* but differs in the respect that the activation of the *tra* feedback loop in *N. vitripennis* requires the presence of a paternal genome, which suggests that imprinting of the

maternal genome prevents female development in haploid males (VERHULST *et al.* 2010).

Although the genes involved in the regulation of sex-determination diverged and have been newly recruited during evolution, these examples from various insects belonging to different orders suggest the existence of some general regulatory principles that underlie sex-determination systems in insects.

The key players of the primary sex-determination cascade of the honeybee have been identified, but little is known of the molecular mechanisms that regulate the sex-determined state that is stably implemented through sex-specific splicing of the *fem* transcripts.

In this study, we showed that the transient provision of female activity in males by expressing the Csd and Fem proteins is sufficient to induce a temporary partial switch of the splicing of *Am-dsx* transcripts into the female mode, and when Fem proteins are provided also of endogenous female *fem* transcripts. However, we observed that providing female activity for a short period during early embryonic development is not sufficient to stably induce the female-determined state in males, which suggests that the *fem* feedback loop acts relatively slowly and requires a longer feminizing signal to be fully activated. Furthermore, we showed that the *Am*-Tra2 protein is sufficient to direct splicing of male-specific *fem* transcripts, which is most likely achieved by binding to putative ESE motifs in the *fem* pre-mRNA. Together with the observation that the *Am*-Tra2 protein interacts with the Fem protein but not with the Csd proteins, these data suggest that the *Am*-Tra2 protein has a dual role in male and female sex-determination. This dual involvement of the *Am*-Tra2 protein may help to establish stable sex-determined states by preventing female activity in males and assisting in the activation of the feedback loop of the Fem protein that maintains the sexual state in females.

Materials & Methods

Plasmid construction

We constructed plasmids using the pDO vector (Fig. S6), which is based on the PIZ/V5-His vector (Invitrogen, Carlsbad, CA, USA), for the expression of the *fem* minigene, hGH, *Am*-Tra2, Csd and Fem proteins. Construction of the plasmid pDO, which contains two expression cassettes driven by the viral OpIE2 promoter, and cloning of pDO *Am*-Tra2 and pDO hGH were previously described (NISSEN *et al.* 2012b). The full-length ORF sequences of the Csd alleles D1-16, A1-18 and D2-27 (HASSELMANN & BEYE 2004) and Fem allele S2-38 (HASSELMANN *et al.* 2008a) were amplified by PCR (Tab. S5) and inserted into plasmid pDO by cloning the ORFs into the *Bam*HI and *Mlu*I or *Avr*II and *Afl*II sites, respectively, generating the plasmids pDO Csd^{16}+Csd^{18}, pDO 2xCsd^{18}, pDO 2xCsd^{27}, pDO Csd^{18}+Csd^{27} and pDO Fem.

For FRET analysis, we generated the plasmids PIZ/V5-His Csd^{25}-Cerulean, PIZ/V5-His Csd^{27}-Cerulean and PIZ/V5-His Fem-Cerulean by inserting the ORFs of the Csd alleles B2-25, and D2-27 or Fem allele S2-38 (HASSELMANN *et al.* 2008a) using *Eco*RI and *Not*I, the Cerulean fluorescent protein (RIZZO *et al.* 2004) using *Xba*I and *Sac*II and a spacer sequence using *Not*I and *Xba*I into plasmid PIZ/V5-His (Invitrogen, Carlsbad, CA, USA). The plasmid PIZ/V5-His *Am*-Tra2-YFP was generated by inserting the ORFs of *Am*-Tra2 (NISSEN *et al.* 2012a) and the yellow fluorescent protein (YFP) (DAUBNER *et al.* 1987) using the *Eco*RI/*Not*I and *Xba*I/*Sac*II restriction sites, respectively. The plasmid pFB Csd^{25}-Cerulean was constructed by removing the His tag from plasmid pFastBac (Invitrogen, Carlsbad, CA, USA) by PCR (Tab. S5) followed by restriction with *Ase*I and *Eco*RI and subsequent ligation with the expression cassette from PIZ/V5-His Csd^{25}-Cerulean that was amplified by PCR (Tab. S5) followed by restriction with *Eco*RI and *Sal*I. To generate the plasmid pFB Csd^{27}-Cerulean+Csd^{25}, the expression cassette in pFB Csd^{25}-Cerulean was exchanged with the cassette from PIZ/V5-His Csd^{27}-Cerulean using the *Eco*RI and *Sal*I restriction sites, and the ORF of Csd allele B2-25 was amplified from pFB Csd^{25}-Cerulean by PCR (Tab. S5) and inserted using the *Pst*I and *Hind*III restriction sites.

To generate mRNAs encoding the Fem, Csd and Rubia proteins, we modified the plasmid $pfem^{csd\text{-}UTR}$ (GEMPE *et al.* 2009). To generate the plasmids pGEMT Myc-Csd^{25} and pGEMT Myc-Csd^{29}, we exchanged the Fem ORF with the ORFs of the Csd alleles B2-25 and D2-29 (HASSELMANN & BEYE 2004) by PCR (Tab. S5) and

restriction with *Nco*I and *Spe*I. To generate the plasmid pGEMT Fem, we removed the Myc tag by inserting the ORF of the Fem allele S2-38 (HASSELMANN *et al.* 2008a) by PCR amplification (Tab. S5) followed by restriction with *Nco*I and *Spe*I. To generate the plasmids pGEMT Rubia, pGEMT Csd^{25} and pGEMT Csd^{27}, we exchanged the Fem ORF with the ORFs of Rubia (SCHULTE *et al.* 2013) and the Csd alleles B2-25 and D2-27 (HASSELMANN & BEYE 2004), respectively by PCR (Tab. S5) followed by restriction with *Nco*I and *Spe*I.

To test whether sequences in *fem* exon 3 function as *Am*-Tra2 protein-dependent ESE sequences, we inserted the putative regulatory regions (pRR) (Fig. 6A, B) into the splicing reporter pOpI D1-*env*/eGFP that is based on SV D1-*env*/eGFP kindly provided by Dr. Heiner Schaal, Heinrich-Heine University Duesseldorf. We generated the plasmids pOpI D1-*env*/eGFP-pRR1-5 and pOpI D1-*env*/eGFP-pRR1-mut1-6 by restricting the plasmid pOpI D1-*env*/eGFP with the enzymes *Eco*RI and *Sac*I and inserting the pRR fragments using linkers. The construction of pOpI D1-*env*/eGFP was previously described (NISSEN *et al.* 2012b).

Construction of the *fem* minigene

We cloned the *fem* minigene by amplifying four fragments of the *fem* gene by PCR (Tab. S5) using genomic DNA. The first fragment contained the *fem* exon 3 and the first 124 bp of intron 3 and was inserted into the pGEM-T cloning vector (Promega, Fitchburg, WI, USA). The second fragment contained the last 102 bp of intron 3 and exon 4 and was inserted into the plasmid containing fragment 1 using the restriction enzymes *Nco*I and *Not*I. The third fragment contained exon 5 and the first 103 bp of intron 5 and was inserted using *Not*I and *Sal*I. The fourth fragment encompassing the last 100 bp of intron 5 and exon 6 was inserted using *Sal*I and *Sac*I. The resulting fem^{mg} sequence was then amplified by PCR (Tab. S5) and inserted into the plasmid pDO replacing the *Avr*II/*Xho*I fragment downstream of the second OpIE2 promoter. We then generated the plasmid pDO fem^{mg} by inserting upstream and downstream spacer sequences.

To generate the mutant minigene pDO $fem^{mg(G257->A)}$, we generated a G to A mutation at nucleotide position 257 in the 5' splice site $D3^{C1}$. The mutant sequence was synthesized and cloned into the plasmid pMCR $fem^{mg(G257->A)}$ (Mr. Gene) and then inserted into pDO fem^{mg} by restriction with *Avr*II and *Pvu*I.

We cloned the plasmid pDO *fem*$^{mg\text{-}pRR1\text{-}mut2}$ by amplifying a fragment of the minigene by PCR (Tab. S5) using mutant primers and inserting the fragment into pDO *fem*^{mg} using the restriction enzymes *Avr*II and *Afl*II.

Identification of ESE sequences

We used ImageJ (SCHNEIDER *et al.* 2012) to semiquantitatively analyze the intensities of gel bands corresponding to the spliced reporter transcripts and spliced hGH transcripts that were amplified by PCR (Tab. S5). We determined the *Am*-Tra2 protein-induced increase in reporter splicing as the ratio of the measured band intensities in the presence and absence of the *Am*-Tra2 protein. To adjust differences in transfection efficiencies among samples, the band intensity values corresponding to the spliced reporter transcripts were normalized using the hGH band intensity values of the respective samples.

RNA extraction, cDNA and mRNA synthesis

The total RNA was extracted using the TRIzol method (GIBCO BRL, Life Technologies, Darmstadt, Germany). The first strand complementary DNA (cDNA) was transcribed from mRNAs using the RevertAid Reverse Transcription Kit (Fermentas, Thermo Fisher Scientific, Waltham, MA, USA) and oligo(dT)$_{18}$ primers according to the manufacturer's instructions. The mRNAs were generated by *in vitro* transcription using the mMessage mMachine T7 Kit (Ambion, Austin, TX, USA) followed by polyadenylation using the Poly(A) Tailing Kit (Ambion, Austin, TX, USA). The mRNAs were purified using the MEGAclear Kit (Ambion, Austin, TX, USA). We injected mRNAs with a Myc tag (with the exception of Rubia) into haploid embryos and mRNAs without a Myc tag into diploid embryos as described previously (GEMPE *et al.* 2009).

Bee sources

Diploid female and male eggs were collected from queens that were derived from brother-sister inbred crosses and produced 50 % female and 50 % diploid male offspring. We identified female and diploid male offspring by determining whether the *csd* genotype was homo- or heterozygous (BEYE *et al.* 2003). Haploid male eggs

were collected from unmated virgin queens. These unmated queens produced unfertilized male eggs after repeated CO_2 treatment (LAIDLAW & PAGE 1997). The hatched larvae were reared as described previously (GEMPE *et al.* 2009). Dissections of 4th instar larvae were performed using a stereomicroscope (S8APO, Leica Microsystems, Wetzlar, Germany) and documented using a microscope-coupled camera (DFC280, Leica Microsystems, Wetzlar, Germany).

FRET-FLIM analysis

We performed Multiparameter Fluorescence Image Spectroscopy (MFIS) in *Sf*21 lepidopteran cells transiently expressing Cerulean-tagged proteins using a multiparameter fluorescence detection setup with the corresponding fluorescence lifetime analyses as described previously (KUDRYAVTSEV *et al.* 2007; WEIDTKAMP-PETERS *et al.* 2009). We used a CLSM (confocal laser scanning microscope, FV1000, IX81 inverted microscope, Olympus, Shinjuku, Tokio, Japan) equipped with a single-photon counting device with picosecond time resolution (Hydra Harp 400, PicoQuant, Berlin, Germany). The Cerulean fluorescent protein was excited at 440 nm using a linearly polarized, pulsed (40 MHz) diode laser (PicoQuant, Berlin, Germany) at 1 µW using a 60x water immersion objective (UPlanSApo NA 1.2, diffraction limited focus, Olympus, Shinjuku, Tokio, Japan). The fluorescence was then detected by SPADs (single-photon avalanche diodes, PD5CTC, MPD, Bolzano, Italy) in a narrow range of the Cerulean emission spectrum (band-pass filter: HC480/40, AHF, Tübingen, Germany). The images were obtained with 10 µs pixel time and zoom10 at 256x256 pixels. A series of 60 frames was merged into one image and further analyzed (WIDENGREN *et al.* 2006).

The Cerulean fluorescence lifetime (τ) was fitted pixel-wise with a mono-exponential model (WEIDTKAMP-PETERS *et al.* 2009) and then analyzed using the software tools “AnI-3SF” and “Margarita” developed by Prof. C.A.M Seidel group [Software Package for Multiparameter Fluorescence Spectroscopy, Full Correlation and Multiparameter Fluorescence Imaging (http://www.mpc.uni-duesseldorf.de/seidel)].

Expression of the fusion proteins was visually confirmed using the internal detectors of the CLSM.

Cell culture and transfection

*Sf*21 cells were maintained according to the supplier's protocols (Invitrogen, Carlsbad, CA, USA) in Spodopan medium (PAN-Biotech, Aidenbach, Germany) that was supplemented with 10 µg/ml of gentamycin (Carl Roth, Karlsruhe, Germany). The transfections were performed in six-well plates ($1x10^6$ cells/well) using the Roti-Insectofect reagent (Carl Roth, Karlsruhe, Germany) according to the manufacturer's instructions. Cells were harvested 72 h after transfection, and the total RNA was extracted and transcribed into cDNA as described above. Each transfection experiment was repeated with at least three biological replicates. The Bac-to-Bac Baculovirus Expression System (Invitrogen, Carlsbad, CA, USA) was used for baculovirus-based protein expression in *Sf*21 cells according to the manufacturer's instructions. For the FRET analysis, cells were examined 48 h after transfection.

Bioinformatic analysis

We used the HBond Score web interface (www.uni-duesseldorf.de/rna) to calculate the strength of the 5' splice sites of the *fem* gene (Table S1) based on their complementarity to the U1 snRNA (KAMMLER *et al.* 2001; FREUND *et al.* 2003).

Statistical analysis

The SPSS 20 software (IBM, Armonk, NY, USA) was used for statistical analysis of FRET data using *t* tests (Student's and Welch's) and Levene's test for equality of variances. We tested whether the measured values were distributed normally using the Kolmogorov-Smirnov test.

Results

The transcripts of the *fem* minigene (*fem*^{mg}) are not sex-specifically spliced in *Sf*21 cells

We cloned a *fem* minigene (*fem*^{mg}) to mimic the sexual splicing of *fem* transcripts in *Sf*21 cells (Fig. 1A, B). The minigene is 2054 bp long and contains exon 3 with the male and female 5' splice sites, the male-specific exons 4 and 5 and the commonly used exon 6. We fused exon 4 to exon 5 and shortened introns 3 and 5, retaining at least 100 bp of intronic sequence on the 5' and 3' ends of each exon. Because *cis* elements that regulate splicing are typically found near the splice sites (GRAVELEY *et al.* 1998; SCIABICA & HERTEL 2006; SHEN & GREEN 2006; HERTEL 2008), the minigene (*fem*^{mg}) should contains the sequence elements required for the regulation of the sex-specifically activated splice sites.

We transiently transfected lepidopteran *Sf*21 cells with the plasmid pDO *fem*^{mg} to express the minigene pre-mRNA. We cotransfected plasmid pDO hGH to express the constitutively spliced transcript of the human growth hormone gene (hGH, (SELDEN *et al.* 1986)) that we used to semiquantitatively adjust differences in the transfection efficiencies among samples. We studied the processed mRNAs of *fem*^{mg} using semiquantitative RT-PCRs.

We did not amplify fragments of processed transcripts that were spliced using male or female splice sites (Fig. 1C, lane 1). This result suggests that neither the male nor the female 5' splice site in exon 3 ($D3^M$ and $D3^F$, see Fig. 1B for explanation) were activated. However, we amplified fragments of another processed transcript (named *fem*$^{mg\text{-}C1}$) that consists of the anterior part of exon 3 including the premature stop codon of the male transcript, the retained shortened intron 5 and exon 6. This transcript was spliced using the alternative 5' splice site $D3^{C1}$ and the downstream 3' splice site $A5^{C1}$. $D3^{C1}$ is located in exon 3, 71 bp downstream of the female 5' splice site ($D3^F$) and 640 bp upstream of the male 5' splice site ($D3^M$), whereas $A5^{C1}$ is located in exon 5 at the same position as the 5' splice site D5 of exon 5.

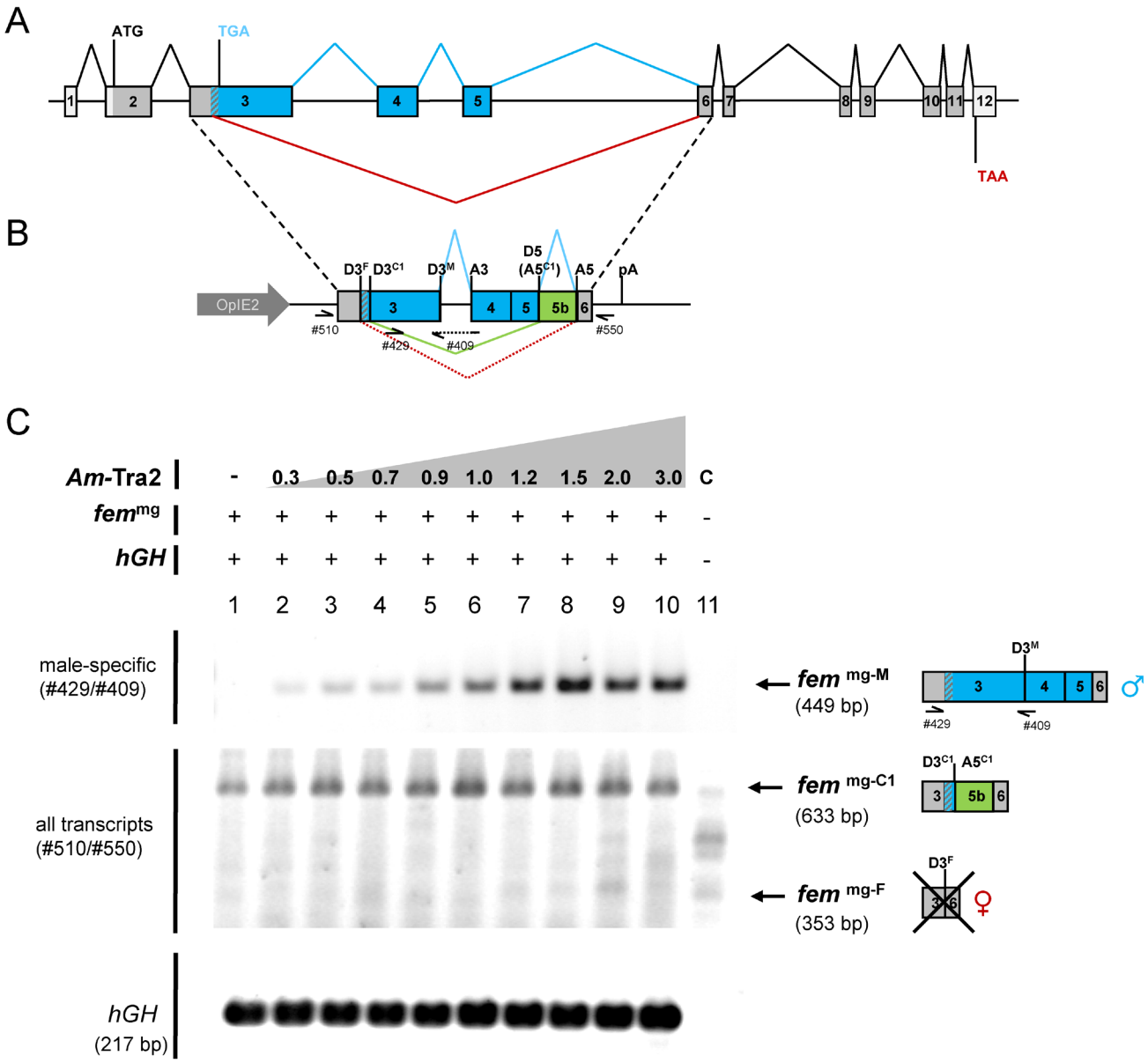

Figure 1. The processing of *fem* minigene (*fem* ^{mg}) transcripts without splicing activators from the honeybee and in response to *Am*-Tra2 protein expression in *Sf*21 cells.
(A) Organization of the *fem* gene in the honeybee genome. The male exons 4 and 5 and the male part of exon 3 are shown as blue boxes, and the common male and female exons and the common part of exon 3 are shown as gray boxes. The connecting lines between the exons indicate splicing of the alternative transcripts: black lines indicate common splice events, the blue lines represent male and the red line female splice events. The common translational start codon is shown in black. The male and female translational stop codons are shown in blue and red, respectively. The light gray boxes represent the 5' and 3' UTRs. (B) Intron and exon structure of the *fem* minigene (*fem* ^{mg}). The minigene driven by the baculovirus-derived OpIE2 promoter was cloned into the plasmid pDO. pA indicates the OpIE2 polyadenylation site. The sequences of introns 3 and 5 were shortened, and at least 100 bp of intron sequence were retained at the 5' and 3' ends of each exon. The colors of the boxes and connecting lines indicate the putative sex-specifically processed transcripts as shown in (A). The green box and line denote a previously unreported alternative splicing event. The 5' splice sites are labeled with a "D" and the 3' splice sites are labeled with an "A", whereas the number denotes the corresponding exon. (C) RT-PCR analysis of the *fem* ^{mg} transcripts without splicing

activators from the honeybee and in response to *Am*-Tra2 protein expression in *Sf*21 cells. 1x10^6 *Sf*21 cells were not transfected (marked with a "C") or transiently transfected with 1 µg of pDO *fem* ^{mg} and 1 µg pDO hGH as indicated by a "+". Increasing amounts of pDO *Am*-Tra2 (0.3 to 3 µg) were cotransfected. Fragments corresponding to the male (449 bp) and an alternatively processed transcript (633 bp) were amplified by RT-PCRs. The positions of the used oligonucleotide primers are indicated by arrows in (B). We used RT-PCR fragments from the mRNA of the cotransfected pDO hGH plasmid to semiquantitatively adjust transfection efficiency across the samples. The fragments were resolved by agarose gel electrophoresis and stained with ethidium bromide (black and white reversed). The processed transcripts were characterized by sequencing. The structure of the transcripts is shown schematically to the right of the gel image.To examine whether splicing from D3^{C1} to A5^{C1} occurs naturally in the honeybee, we used RT-PCRs to amplify fragments of mRNAs encompassing this splice junction from male 4th instar larvae, tissues of the female gonads and female 1st instar larvae.

In all the tissue samples analyzed, we detected fragments of mRNAs that were spliced from D3^{C1} to A5^{C1} (Fig. S1), suggesting that this is a novel natural splice variant of the honeybee.

Together, these data indicate that the sex-specific splice sites of the minigene transcripts are not used by default in *Sf*21 cells, suggesting that additional activating factors are required for sexual splicing of the ***fem*^{mg}** pre-mRNA. Instead, we discovered a non-sexual alternative splice product of the minigene (*fem*$^{mg\text{-}C1}$) and identified its equivalent in the honeybee.

Bioinformatic analysis of *fem* 5' splice sites

To understand why neither the male nor the female 5' splice site was activated in the ***fem*^{mg}** pre-mRNA, we studied the strength of the sex-specifically and constitutively used 5' splice sites using a bioinformatic prediction program (Tab. S1). In this program, the strength of a 5' splice site is calculated based on its complementarity to the free 5' end of the U1 snRNA (KAMMLER *et al.* 2001; FREUND *et al.* 2003). High complementarity between a 5' splice site and the U1 snRNA leads to recruitment of the spliceosome and results in the activation of the splice site, whereas the activation of 5' splice sites with low complementarity depends on the assistance of splicing activators.

We observed that the alternative 5' splice site in exon 3 (D3^{C1}), which was used by default in our minigene assay, is predicated to be relatively strong (H-bond score of 18.8) because of its high complementarity to the U1 snRNA (Tab. S1). This indicates that D3^{C1} can be used without the assistance of splicing activators.

We found that the female and male 5' splice sites ($D3^F$ and $D3^M$) have a low complementarity to the U1 snRNA (H-bond score of 12.4 and 12.6, respectively), resulting in predicted strengths below the threshold of 13 hydrogen bonds that is sufficient to recruit the U1 snRNA (FREUND *et al.* 2003). These predictions suggests that the male $D3^M$ and female $D3^F$ 5' splice sites require splicing enhancers to be activated and used in the processed transcripts. We suggest that additional activating factors from the honeybee are required to mimic the sexual splicing of *fem* transcripts in *Sf*21 cells.

The *Am*-Tra2 protein is sufficient to direct male-specific splicing of the *fem* minigene transcripts

Knockdown studies on the *Am-tra2* gene in honeybees suggested that the *Am*-Tra2 protein is required for male splicing of *fem* transcripts (NISSEN *et al.* 2012a). We tested whether we can induce the use of the male 5' splice site ($D3^M$) and the male exons 4/5 together with the commonly used exon 6 by coexpressing the *Am*-Tra2 protein and the fem^{mg}. We transfected *Sf*21 cells with the plasmid pDO fem^{mg} and increasing amounts of the plasmid pDO *Am*-Tra2 and studied the fem^{mg}-derived transcripts using RT-PCRs (Fig. 1C, lanes 2-10). The cells were cotransfected with the plasmid pDO hGH to adjust differences in transfection efficiency.

The RT-PCR data revealed transcripts ($fem^{mg\text{-}\mathbf{M}}$) that used the male 5' splice site $D3^M$ in exon 3 and the 3' splice site A3 in exon 4, suggesting that this splicing process mimics the splicing of male *fem* transcripts in the honeybee. We observed that the amount of this product semi-quantitatively increased with the level of plasmid pDO *Am*-Tra2 used for transfection (Fig. 1C, lanes 2-10). These results suggest that the *Am*-Tra2 protein mediates the use of the male splice site $D3^M$ and the inclusion of the male exons 4/5, thus directing the production of the male *fem* mRNAs.

Furthermore, we found fragments corresponding to the $fem^{mg\text{-}\mathbf{C1}}$ transcript. The semi-quantitative analysis suggests that the production of this splice variant is not affected by the amount of pDO *Am*-Tra2 plasmid used for transfection (Fig. 1C, lanes 1-10), suggesting that the 5' splice site $D3^{C1}$ is regulated independently of *Am*-Tra2 proteins.

The Csd and Fem proteins are not sufficient to activate the female 5' splice site $D3^F$ in *Sf*21 cells

We studied whether the proteins Csd, Fem and *Am*-Tra2 are sufficient to activate the female 5' splice site $D3^F$ of the *fem* minigene and induce splicing to the 3' splice site A5 of the commonly used exon 6. The female-specific splicing of the *fem* transcripts is a critical step in honeybee sex-determination that implements and maintains the female-determined state throughout development (GEMPE *et al.* 2009). Csd proteins derived from different alleles are required to mediate the female-specific splicing of the *fem* transcripts that initiates female development in the honeybee (GEMPE *et al.* 2009). The female-determined state is then maintained through a positive feedback loop in which the Fem protein directs splicing of the *fem* pre-mRNA into the productive female form (GEMPE *et al.* 2009). The female-specific splicing of *fem* transcripts also involves the *Am*-Tra2 protein that is required for the female- and male-specific processing of the *fem* transcripts as shown by RNAi knockdown experiments (NISSEN *et al.* 2012a).

We studied whether expression of the Csd proteins from three different alleles (Csd^{16}, Csd^{18}, Csd^{27}) together with the *Am*-Tra2 protein is sufficient to activate the female 5' splice site $D3^F$ (Fig. 2). We transfected *Sf*21 cells with plasmids expressing the *fem* minigene and the *Am*-Tra2 protein (pDO fem^{mg} and pDO *Am*-Tra2, respectively). We cotransfected plasmid pDO hGH (to adjust differences in transfection efficiency) and increasing amounts (0.5 µg to 3 µg) of plasmids expressing Csd proteins from one allele (plasmids with two copies of the same *csd* allele: pDO $2xCsd^{18}$ and pDO $2xCsd^{27}$) or from two different alleles (plasmids pDO Csd^{16}+Csd^{18} and pDO Csd^{18}+Csd^{27}) and studied the effect on the activation of the female 5' splice site $D3^F$.

When only one Csd allele was expressed (homozygous expression), we detected no transcripts that used the female 5' splice site $D3^F$, irrespective of the used plasmid concentration (Fig. 2, lanes 11-13 and 17-19). This observation suggests that proteins from a single Csd allele cannot activate the female splice site, which is consistent with our expectation.

However, we found fragments corresponding to the $fem^{mg\text{-}\mathbf{C1}}$ and $fem^{mg\text{-}\mathbf{M}}$ transcripts. We observed an inverse quantitative correlation between the level of the male-like splice variant $fem^{mg\text{-}\mathbf{M}}$ and the amount of the Csd-expressing plasmid used

for transfection. This tendency may indicate that the Csd proteins can inhibit activation of the male-specific splicing that is induced by the *Am*-Tra2 proteins.

When cells expressed two combinations of two different Csd alleles in increasing concentrations (heterozygous expression), we observed the same transcripts (*fem*$^{mg\text{-}C1}$ and *fem*$^{mg\text{-}M}$) that we detected when only one Csd allele was expressed (Fig. 2, lanes 8-10 and 14-16). We found no splice variants that used the female 5' splice site D3^{F}, suggesting that the expression of two different Csd alleles together with the *Am*-Tra2 protein is not sufficient to activate the female 5' splice site of the *fem* minigene in *Sf*21 cells.

In the honeybee, the Fem protein direct the splicing of *fem* transcripts into the female-specific form. We studied whether coexpression of the Fem and *Am*-Tra2 proteins can activate the female 5' splice site D3^{F} (Fig. 2, lanes 3-7). We transfected cells with the plasmids pDO *fem*^{mg} and pDO *Am*-Tra2 and increasing amounts of the pDO Fem plasmid (0.3 µg to 3.5 µg). We observed that expression of the Fem protein did not lead to the use of the female splice site, suggesting that coexpression of the Fem and *Am*-Tra2 proteins is not sufficient to activate the female 5' splice site D3^{F} of the minigene in *Sf*21 cells. Instead, we found the *fem*$^{mg\text{-}C1}$ and *fem*$^{mg\text{-}M}$ transcripts, and as described above, the amount of male-specific splicing appeared to decrease with increasing amounts of the transfected plasmid pDO Fem. This suggests that the Fem protein, like the Csd proteins, reduces the *Am*-Tra2-derived activation of the male splice sites.

Next, we investigated whether the female 5' splice site D3^{F} was not activated in the experiments above in *Sf*21 cells because of the frequent use of the 'strong' 5' splice site D3^{C1}. To test this hypothesis, we inactivated the D3^{C1} 5' splice site by replacing the canonical GT dinucleotide at the U1 snRNA-binding site with AT (Fig. S2A). However, we could not detect transcripts that were spliced using the female 5' splice site D3^{F} when we transfected *Sf*21 cells with plasmid pDO *fem*$^{mg(G257->A)}$ that expresses the minigene in which the 5' splice site D3^{C1} was inactivated, cotransfected plasmid pDO *Am*-Tra2 and plasmids expressing the Fem proteins (Fig. S2B), homoallelic Csd proteins (pDO 2xCsd^{25} and pDO 2xCsd^{27}, data not shown) or heteroallelic Csd proteins (pDO Csd^{25}+Csd^{27}, data not shown). Instead, we found fragments that used the male 5' splice site D3^{M} and fragments that were spliced using a pair of cryptic splice sites (D3^{C2} and A5^{C2}), which were used instead of the inactivated splice site.

Together, these results suggest that the Csd and Fem proteins alone or together with the *Am*-Tra2 protein are not sufficient to activate the female 5' splice site $D3^F$ of the *fem* minigene in *Sf*21 cells even when the potentially competing 'strong' 5' splice site $D3^{C1}$ is non-functional. We propose that additional regulatory factors from the honeybee may be necessary to promote activation of the female 5' splice site $D3^F$.

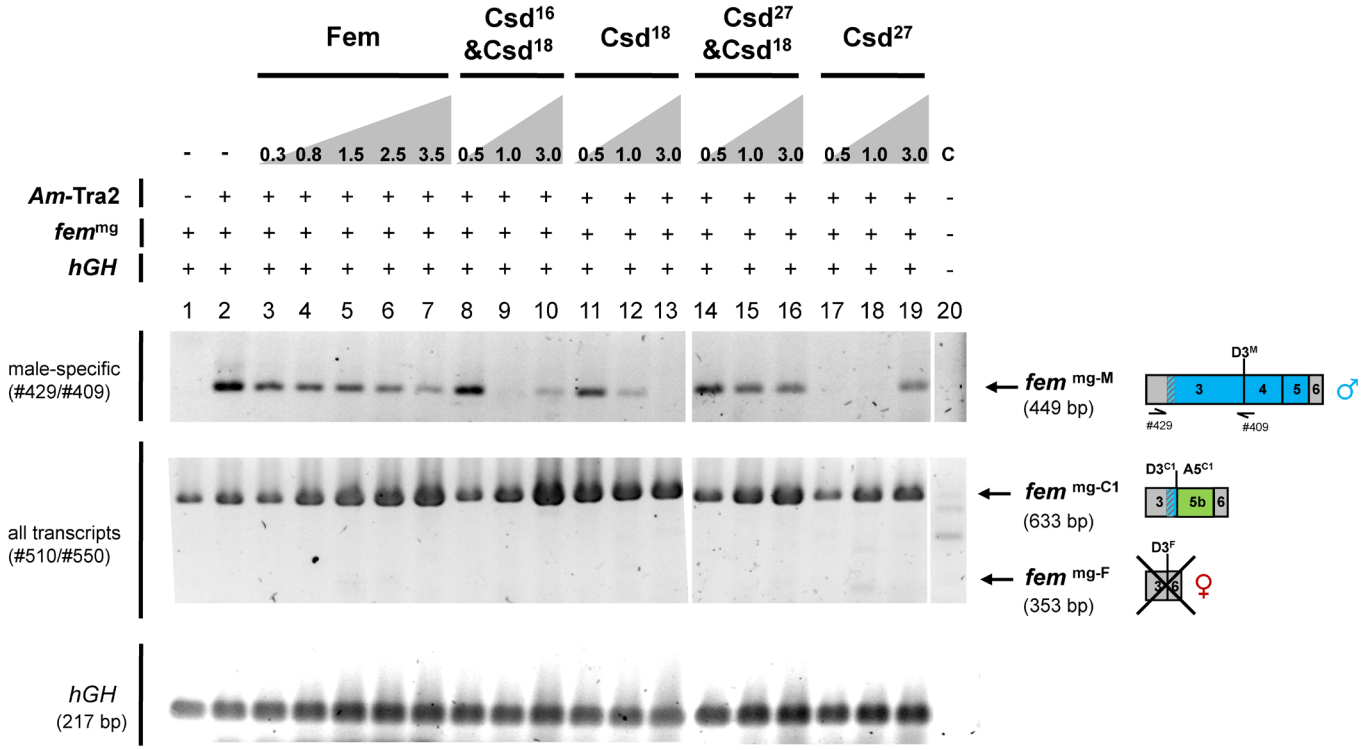

Figure 2. RT-PCR analysis of the *fem* ^{mg} transcripts in response to Fem and Csd protein expression in *Sf*21 cells. $1x10^6$ *Sf*21 cells were not transfected (marked with a "C") or transiently transfected with 1 µg of pDO *fem* ^{mg} and 1 µg of pDO hGH as indicated by "+". 1 µg of pDO *Am*-Tra2 was cotransfected with increasing amounts of pDO Fem (0.3 to 3.5 µg) or plasmids expressing Csd (0.5 to 3 µg), which encode either one Csd allele (pDO $2xCsd^{18}$ and pDO $2xCsd^{27}$) or two different alleles (pDO Csd^{16}+Csd^{18} and pDO Csd^{27}+Csd^{18}). Fragments corresponding to the male (449 bp) and an alternatively processed *fem* minigene splicing variant (633 bp) were amplified by RT-PCRs. The positions of the used oligonucleotide primers are indicated by arrows in (Fig. 1B). We used the RT-PCR fragments from the mRNA of the cotransfected pDO hGH plasmid to semiquantitatively adjust the transfection efficiency across samples. The fragments were resolved by agarose gel electrophoresis and stained with ethidium bromide (black and white reversed). The processed transcripts were characterized by sequencing. The structure of the transcripts is shown schematically to the right of the gel image.

The Fem and *Am*-Tra2 proteins interact in living cells

In *Drosophila melanogaster*, the Fem ortholog Tra interacts with Tra2 to activate female development (BURTIS & BAKER 1989; HEDLEY & MANIATIS 1991; INOUE

et al. 1992; SCIABICA & HERTEL 2006). Therefore, the observed reduction in male splicing could be an indirect indication that the Fem or Csd proteins directly interact with the *Am*-Tra2 protein, which was proposed as a key step to direct and maintain female development in the honeybee (NISSEN *et al.* 2012a).

We studied the interaction of the Fem and *Am*-Tra2 proteins in *Sf*21 cells using Förster Resonance Energy Transfer (FRET) as a measurement for protein-protein interaction. We fused Fem with the Cerulean fluorescent protein (a monomeric CFP variant (RIZZO *et al.* 2004)) and *Am*-Tra2 with the yellow fluorescent protein (YFP (DAUBNER *et al.* 1987)) and expressed these proteins in *Sf*21 cells. The Cerulean fluorophore is excited using a laser pulse and emits light for a certain amount of time (the fluorescence lifetime). If the Cerulean and YFP fusion proteins are at a very close distance < 10 nm (BORST & VISSER 2010), the fluorescence of the Cerulean fluorophore is quenched by YFP, which we measured as a reduction of the Cerulean fluorescence lifetime using FRET-FLIM (Fluorescence Lifetime Imaging Microscopy (WEIDTKAMP-PETERS *et al.* 2009; BORST & VISSER 2010)).

We measured the fluorescence lifetime under conditions where the Fem-Cerulean protein was expressed alone or together with the *Am*-Tra2-YFP protein (Fig. 3A). We observed that the fluorescence lifetime of the excited Fem-Cerulean protein was significantly lower when it was coexpressed with the *Am*-Tra2-YFP protein (Student's *t*-test, 2-tailed, $P < 0.001$). This result suggests that the Fem and *Am*-Tra2 proteins interact in living cells.

Next, we studied whether the Csd proteins derived from a single allele (mimicking the homozygous male genotype) interact with *Am*-Tra2 proteins. The fluorescence lifetime of cells expressing the Csd^{25}-Cerulean protein did not differ significantly from those expressing the Csd^{25}-Cerulean and *Am*-Tra2-YFP proteins (Fig. 3B) (Welch's *t*-test, 2-tailed, $P = 0.073$). This result suggests that there is no interaction between the Csd and *Am*-Tra2 proteins in the homozygous male condition.

We investigated whether the Csd proteins derived from two allelic variants (mimicking the heterozygous female genotype) interact with *Am*-Tra2 proteins. We coexpressed the Csd^{27}-Cerulean and Csd^{25} proteins in the presence and absence of the *Am*-Tra2-YFP protein (Fig. 3C). The Cerulean fluorescence lifetime was not affected by the presence or absence of coexpressed *Am*-Tra2-YFP protein (Student's *t*-test, 2-tailed, $P = 0.417$). This result suggests that even under the heterozygous

female condition, there is no interaction between the two Csd protein variants and *Am*-Tra2 proteins.

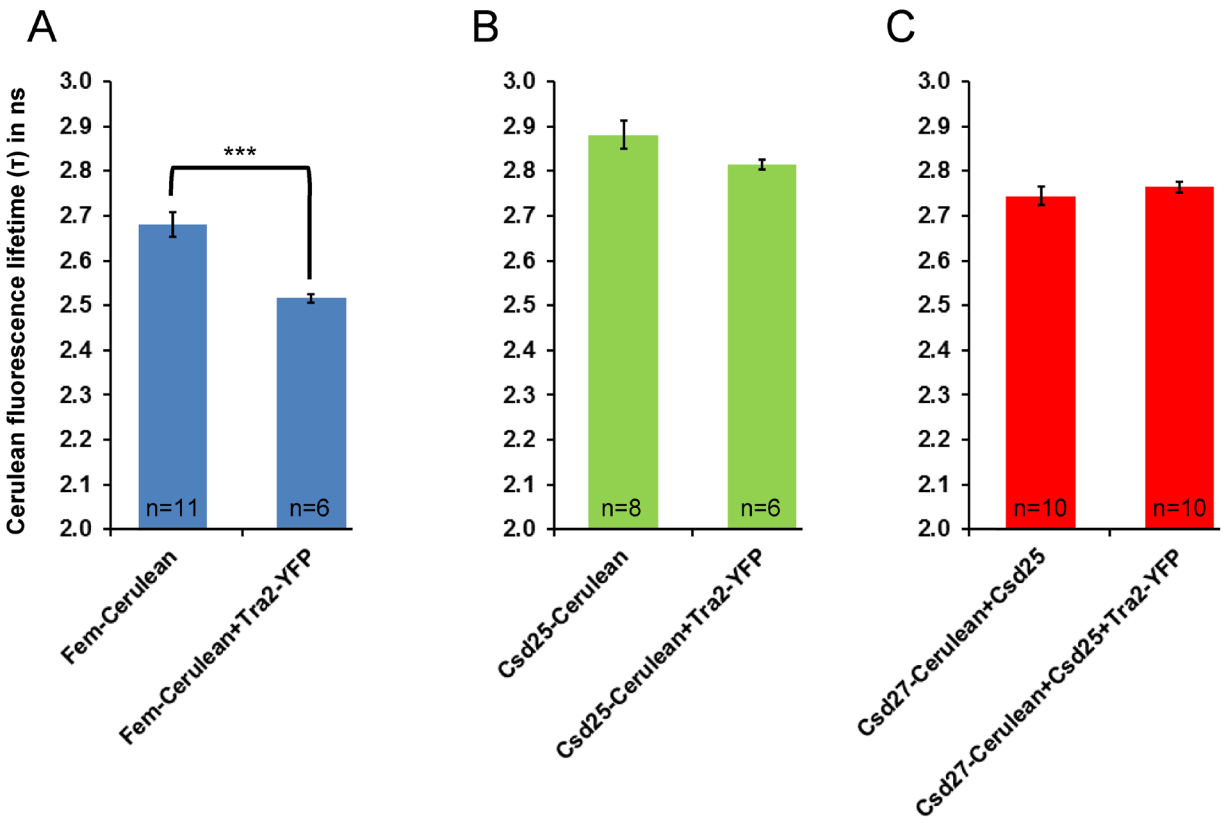

Figure 3. Fluorescence lifetime analysis of the *Am*-Tra2/Fem proteins and the *Am*-Tra2/Csd proteins in *Sf*21 lepidopteran cells using FLIM-FRET. Fem-Cerulean, Csd-Cerulean and *Am*-Tra2-YFP fusion proteins were expressed in *Sf*21 cells. The following conditions were tested in the presence and absence of *Am*-Tra2-YFP proteins: (A) Fem-Cerulean protein. (B) Csd^{25}-Cerulean protein (homozygous male condition). (C) Csd^{27}-Cerulean and Csd^{25} proteins (heterozygous female condition). The Cerulean fluorescence lifetime (τ) was measured using FLIM-FRET (Fluorescence Lifetime Imaging Microscopy - Förster Resonance Energy Transfer) and compared for each condition in the presence and absence of coexpressed *Am*-Tra2-YFP protein using the 2-tailed Student's *t*-test (A and C) or Welch's *t*-test (B). n denotes the number of cells that were analyzed for each condition. 2-tailed *t*-test (Student and Welch): *** indicates $P < 0.001$.

Transient expression of the Csd or Fem protein can only mediate a temporary splicing of female *Am-dsx* and *fem* transcripts in males

We tested whether the expression of another Csd protein or Fem protein can induce femaleness in haploid males. Previous studies showed that *Am-dsx* acts downstream of *fem,* and the activity of Csd proteins from different alleles and of the Fem protein is required to mediate the female splicing of *Am-dsx* transcripts (GEMPE *et al.* 2009). If the transient expression of Fem or additional Csd proteins is sufficient

to induce femaleness in males, we would expect to detect the female transcript of the *Am-dsx* gene (*Am-dsx*^{F}) in the treated male individuals. We injected syncytial haploid embryos with different amounts of *in vitro*-synthesized mRNA encoding the proteins of one Csd allele (Csd^{25}, 210 pg, 820 pg and 1.2 ng) or two alleles (Csd^{25} and Csd^{29}, 1.5 ng each) or 1 ng of mRNA encoding the Fem protein. After 48-72 h, we studied the effect of these injections on *Am-dsx*^{F} splicing using RT-PCR. Control embryos were injected with mRNA (210 pg, 820 pg and 1 ng) encoding the red fluorescent protein Rubia (SCHULTE *et al.* 2013).

We amplified fragments corresponding to *Am-dsx*^{F} from 21 % of the *csd*- and 90 % of the *fem*-injected haploid individuals (Fig. 4A, lanes 12-21 and 22-31; Tab. 1) but not from the male control embryos (Fig. 4A, lanes 2-11). We obtained strong and weak fragments above the detection level, suggesting that the *csd* and *fem* mRNAs had weak effects on many embryos; however, the effect of the *fem* mRNA was more consistent because only one *fem*-injected individual did not yield a fragment corresponding to *Am-dsx*^{F}.

Apart from the female variant, we detected the transcript variant *Am-dsx*^{252} in the treated embryos, which has not been described before. This new transcript variant contains exons 3, 4 and 5, like the female transcript, but in exon 5, an alternative 3' splice site that is 140 nt downstream of the female 3' splice site is used, resulting in an ORF that is 75 nt shorter.

These results suggest that expression of the Fem protein consistently mediates female splicing of *Am-dsx* transcripts, whereas expression of other Csd proteins only sporadically directed female splicing in male embryos.

We also tested whether the mRNA injections induced splicing of female *fem* transcripts. It was previously shown that injection of the *fem* mRNA induces a partial activation of female *fem* splicing (GEMPE *et al.* 2009), but this has not been tested in the case of *csd*. We detected weak fragments of endogenous female *fem* transcripts in 70 % of the embryos injected with the *fem* mRNA (Fig. 4B, lanes 22-31; Tab. 1). The embryos that were injected with the *csd* mRNA did not yield fragments corresponding to female *fem* mRNA (Fig. 4B, lanes 12-21; Tab. 1) like the male controls (Fig. 4B, lines 2-11). This observation suggests that unlike Fem, expression of additional Csd proteins does not cause splicing of detectable amounts of female *fem* transcripts.

Table 1. Production of Am-dsxF and femF transcripts in 48-72 h old haploid males in response to embryonic treatment with fem, csd or rubia mRNA

Treatment	Number of Embryos	Number of Embryos with transcript *fem*^{F}		Number of Embryos with transcript *Am-dsx*^{F}	
rubia mRNA	14	0	0%	0	0%
fem mRNA	10	7	70%	9	90%
csd mRNA	28	0	0%	6	21%

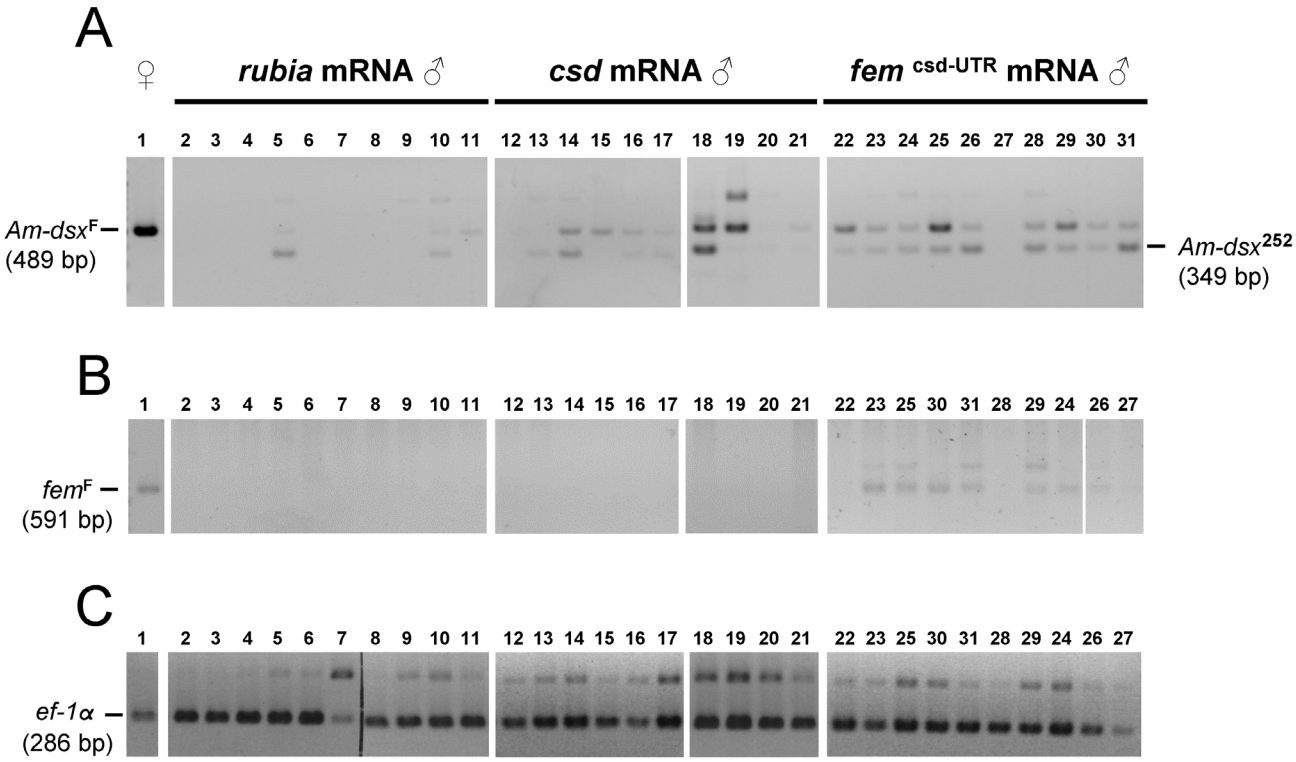

Figure 4. Splicing of *Am-dsx*^{F} and *fem*^{F} transcripts in 48-72 h old haploid male embryos in response to embryonic treatment with *fem, csd* or *rubia* mRNA. 0-3 h old haploid embryos were injected with 820 pg (lanes 12-13) and 1.2 ng (lanes 14-21) of *csd* mRNA or 1 ng (lanes 22-31) *fem* $^{csd\text{-}UTR}$ mRNA. Embryos injected with 210 pg (lanes 2-7) or 1 ng (lanes 8-11) of *rubia* mRNA and an untreated female (lane 1) served as controls. The same set of samples was used in (A-C), as indicated by the numbers. Fragments corresponding to (A) the female (489 bp) and a previously unreported (349 bp) *Am-dsx* transcript and the female *fem* transcript (591 bp) (B) were amplified by RT-PCR, resolved by agarose gel electrophoresis and stained with ethidium bromide (black and white reversed). Fragments corresponding to (C) the mRNA of the *ef-1α* gene were amplified as a positive control for reverse transcription.

Next, we examined whether the induction of female splicing is maintained until the larval stage. In this experiment, we injected two different amounts of mRNAs (100 pg and 500 pg) encoding either the Csd protein of one allele (Csd^{25} or Csd^{27}) or the Fem protein. Control embryos were injected with 500 pg of mRNA encoding the Rubia protein. We injected the mRNA into syncytial embryos derived from crosses between brothers and sisters that produced 50 % female and 50 % diploid male offspring with a homozygous *csd* genotype. We reared the larvae to the 1st and 4th instar and studied the effect of mRNA injection on *Am-dsx* and *fem* splicing. We used diploid instead of haploid males to exclude the possibility that the onset of female development may lead to lethality due to the possible accumulation of deleterious recessive mutations in female-specific genes in the haploid genome. In addition, diploid embryos showed higher survival rates than haploid embryos after microinjection in previous studies (GEMPE *et al.* 2009; NISSEN *et al.* 2012a).

Using RT-PCRs, we detected no fragments corresponding to *Am-dsx*^{F} or *Am-dsx*^{252} in the male controls or the male larvae that were injected with the *csd* or *fem* mRNA (Fig. 5A, Tab. 2). However, we amplified fragments corresponding to *Am-dsx*^{M} in all these individuals, suggesting that the sex-specific splicing of *Am-dsx* transcripts was not affected by the mRNA injections.

Next, we examined the *fem* splicing pattern, and we did not detect female-processed *fem* in the diploid males that were treated with the *csd* or *fem* mRNA, except for one individual that was injected with the *fem* mRNA (Fig. 5B, Tab. 2). Male *fem* was detected in all samples. Combined with the results obtained from the haploid males that showed female transcripts, this suggests that female *Am-dsx* splicing is induced only temporarily by injection of the *csd* or *fem* mRNA. Furthermore, injection of the *fem* mRNA appeared to only temporarily mediate female *fem* splicing. Consistent with these results, the *fem*-treated male 4th instar larvae showed normally developed testes (n=5, Tab. S4, Fig. S3C).

We also analyzed the *Am-dsx* splicing pattern in the treated female 1st and 4th instar larvae. We expected that the female controls injected with the *csd* or *fem* mRNA would only contain the *Am-dsx*^{F} transcript variant, like the females that were treated with the *rubia* mRNA (Fig. 5A, Tab. 2). However, only 26 % of the *csd* mRNA-injected females yielded fragments corresponding only to *Am-dsx*^{F}. Another 37 % contained both splice variants *Am*-dsx^{M} and *Am-dsx*^{F}. The remaining 37 % yielded fragments of *Am-dsx*^{M} only.

Table 2. Production of male and female *fem* and *Am-dsx* transcripts in early diploid male and female larvae (1st and 4th instar) in response to embryonic treatment with *fem, csd* or *rubia* mRNA

Treatment	Genotypic Sex	Number of Larvae	Number of Larvae with transcript									
			feminizer				*Am-doublesex*					
			Only fem^M		fem^M & fem^F		Only dsx^M		Only dsx^F		dsx^M & dsx^F	
rubia mRNA	males	11	11	100%	0	0%	11	100%	0	0%	0	0%
	females	8	0	0%	8	100%	0	0%	8	100%	0	0%
fem mRNA	males	10	9	90%	1	10%	10	100%	0	0%	0	0%
	females	17	4	24%	13	76%	2	12%	8	47%	7	41%
csd mRNA	males	16	16	100%	0	0%	16	100%	0	0%	0	0%
	females	19	11	58%	8	42%	7	37%	5	26%	7	37%

Similarly, 47 % of the *fem* mRNA-injected females yielded fragments of *Am-dsx*^{F} only (Fig. 5A, Tab. 2). Fragments corresponding to both *Am-dsx*^{M} and *Am-dsx*^{F} were detected in 41 %, and fragments corresponding to *Am*-dsx^M only were found in 12 % of the females treated with *fem* mRNA. The *Am-dsx*^{252} variant was not detected in any of the individuals that were treated with the *csd* or *fem* mRNA.

We also tested the effect on *fem* splicing in these individuals. Male *fem* was found in all samples, whereas female *fem* was detected in 76 % of the *fem* mRNA-treated females and only 42 % of the *csd* mRNA-treated females (Fig. 5B, Tab. 2). The injection of *fem* mRNA also affected gonad development in the female larvae as we found testis-like gonads in the treated 4th instar larvae (n=6, Tab. S4, Fig. S3D); this result was unexpected because *fem* was previously shown to provide female but not male activity (GEMPE *et al.* 2009). These results suggest that female development was compromised in the *csd* and *fem* mRNA-treated females.

We propose that the development of male-like gonads, the lack of female *fem* splicing and the induction of male *Am-dsx* transcripts might result from overexpression of the Fem and Csd proteins. Similar induction of maleness was observed after ectopic expression of the Tra protein in *Drosophila melanogaster* females (KEISMAN & BAKER 2001).

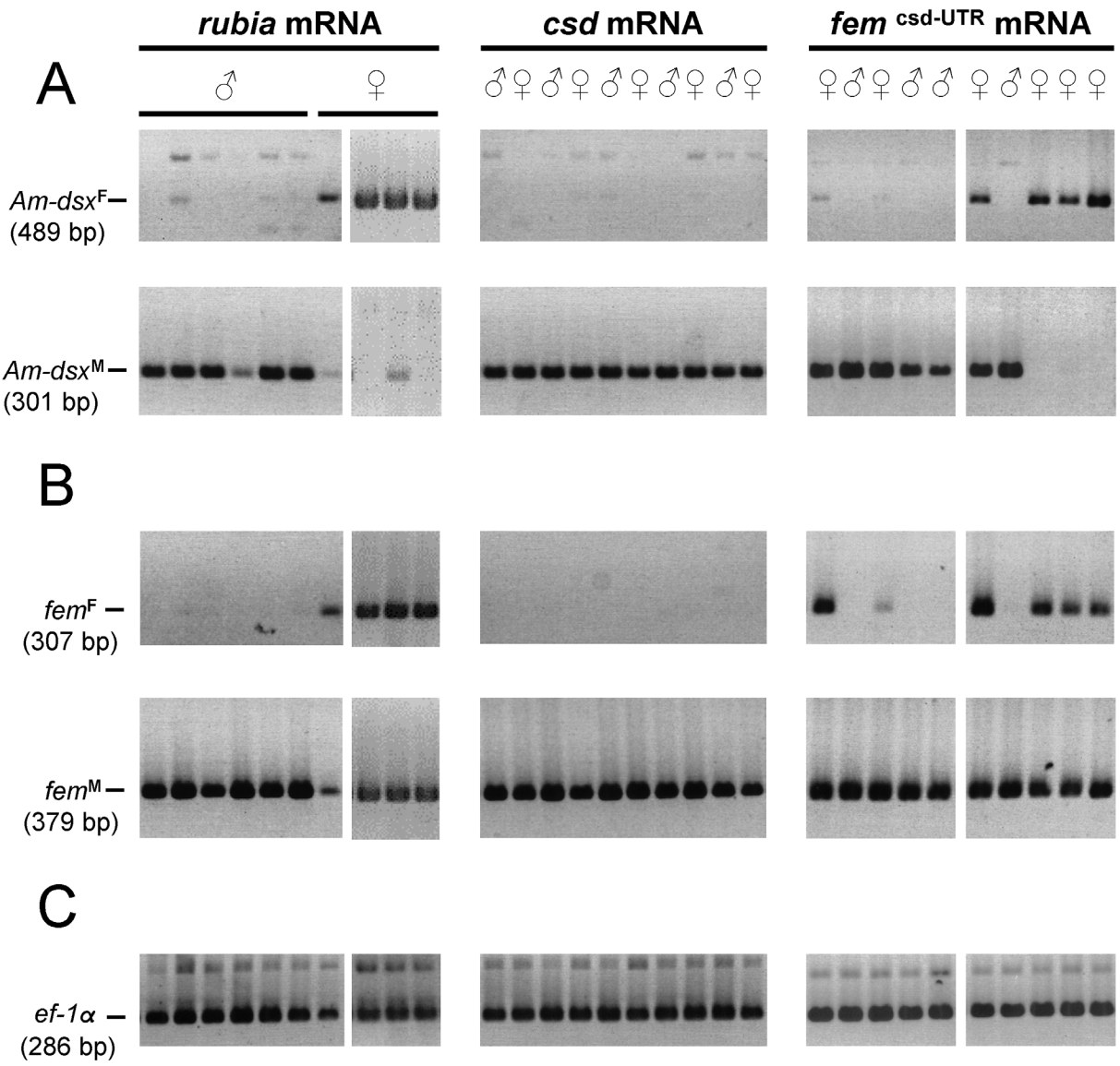

Figure 5. Sex-specific splicing of *Am-dsx* and *fem* transcripts in diploid 1st instar larvae in response to embryonic treatment with *fem, csd* or *rubia* mRNA. 0-3 h old diploid embryos were injected with 500 pg of *csd* mRNA or 100 pg of *fem* $^{csd\text{-UTR}}$ mRNA. Embryos injected with 500 pg of *rubia* mRNA served as controls. The same set of samples was used in the same order in (A-C). Fragments corresponding to (A) the female (489 bp) and male (301 bp) *Am-dsx* transcripts and (B) the female (307 bp) and male (379 bp) *fem* transcripts were amplified by RT-PCR, resolved by agarose gel electrophoresis and stained with ethidium bromide (black and white reversed). Fragments corresponding to (C) the mRNA of the *ef-1α* gene were amplified as a positive control for reverse transcription.

Putative ESE sequences in the *fem* male exon 3 promote reporter splicing in the presence of the *Am*-Tra2 protein

We found that the male 5' splice site $D3^M$ in exon 3 has low complementarity to the U1 snRNA, indicating that the assistance of a splicing activator is necessary for use of this splice site. We also showed that the *Am*-Tra2 protein directs splicing of the male *fem* minigene transcripts through activation of the male 5' splice site $D3^M$.

Next, we studied whether the *Am*-Tra2 protein could direct male splicing by interacting with an ESE sequence. In *Drosophila melanogaster*, the Tra2, Tra and 9G8 proteins bind to the six ESE sequences of the *dsx* repeat element (*dsx*RE) to promote female splicing of *dsx* transcripts (BURTIS & BAKER 1989; HEDLEY & MANIATIS 1991; INOUE *et al.* 1992; TIAN & MANIATIS 1993; LYNCH & MANIATIS 1995; LYNCH & MANIATIS 1996; SCIABICA & HERTEL 2006). Similar repeat elements (*fru*RE) regulate the female-specific splicing of transcripts of the *D. melanogaster fru* gene (LAM *et al.* 2003). Sequence motifs similar to the *dsx*RE have also been identified in the *dsx* and *tra* genes of several other dipteran insects (HEDIGER *et al.* 2004; LAGOS *et al.* 2005; RUIZ *et al.* 2007a; SACCONE *et al.* 2008) but have not been detected in the *fem* and *Am-dsx* genes of the honeybee.

We screened the 155-nt-long sequence upstream of the male 5' splice site $D3^M$ for putative *Am*-Tra2-dependent ESE sequences. We focussed our analysis on this sequence because regulatory elements are typically located in the vicinity of exon/intron junctions (TIAN & MANIATIS 1994; GRAVELEY *et al.* 1998; SHEN & GREEN 2006; HERTEL 2008), and there is evidence suggesting that the regulatory elements closest to the 5' splice site have the greatest influence on splicing decisions (ERKELENZ *et al.* 2013a). Furthermore, SR proteins in general are known to act as splicing enhancers through ESE sequences (GRAVELEY 2000; ERKELENZ *et al.* 2013a), which has been shown for Tra2 protein orthologs in several examples (HOFMANN *et al.* 2000; LAM *et al.* 2003; SCIABICA & HERTEL 2006; GRELLSCHEID *et al.* 2011; ERKELENZ *et al.* 2013b).

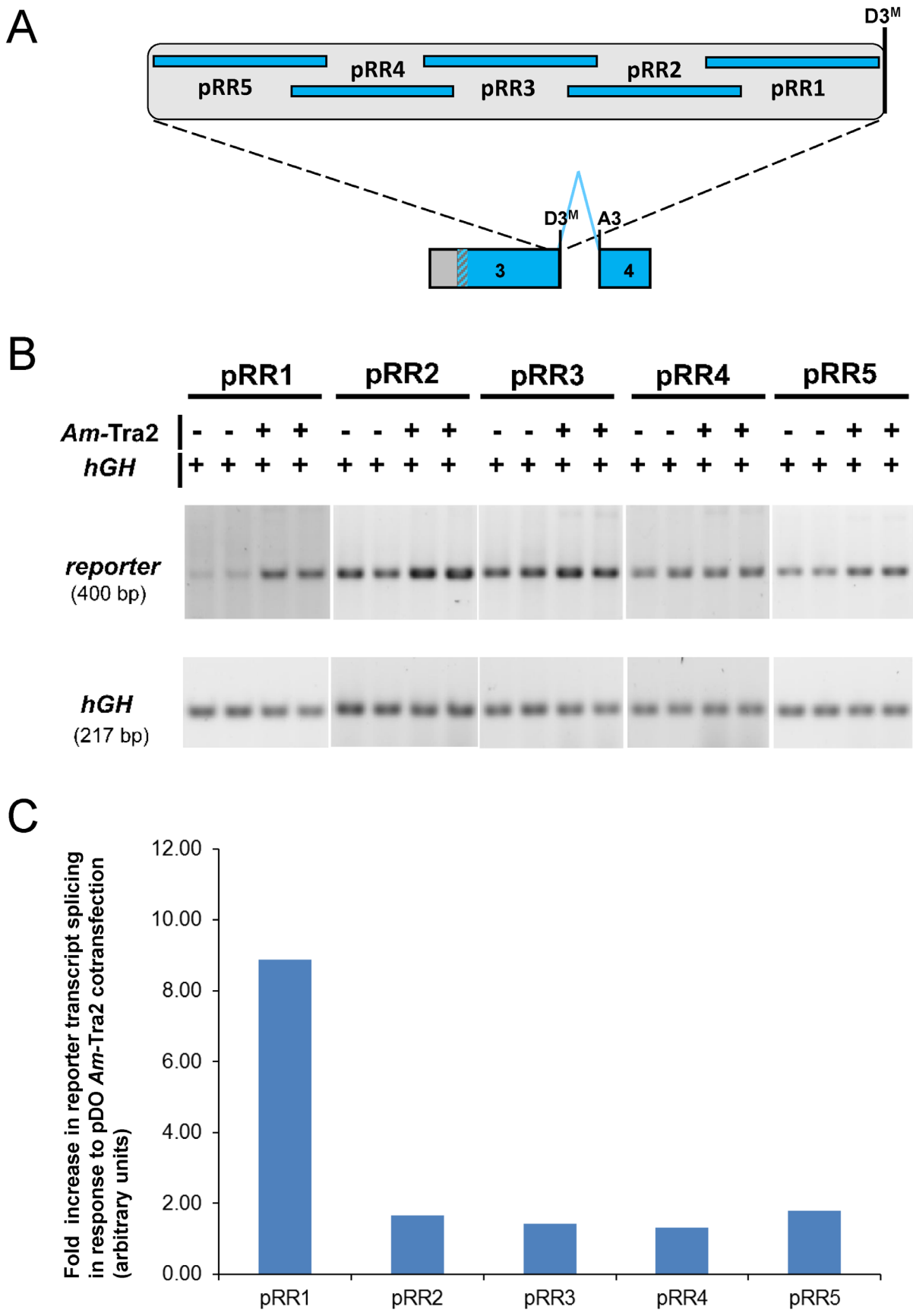

Figure 6. Identification of putative *Am*-Tra2 protein-dependent ESEs upstream of the male 5' splice site D3^{M} in the male-specific part of exon 3. (A) Positions of pRR sequences within exon 3. The positions of the putative regulatory regions (pRR 1 to 5) that were tested using the splicing reporter construct D1-env/eGFP within the 155 bp sequence upstream of the male 5' splice site D3^{M}

are shown. The male exon 4 and the male part of exon 3 are shown as blue boxes. The common part of exon 3 is shown as a gray box. The blue line indicates splicing from the male 5' splice site $D3^M$ to the male 3' splice site A3. (B) The role of pRR sequences in *Am*-Tra2 protein-dependent activation of the reporter construct 5' splice site D1. $1x10^6$ *Sf*21 cells were transiently transfected with 1 µg of the respective pRR reporter clone (plasmids pOpI D1-*env*/eGFP-pRR1-5) and 1 µg of plasmid pDO hGH. Cells cotransfected with 1 µg of plasmid pDO *Am*-Tra2 are indicated by a "+". We amplified fragments corresponding to the spliced reporter transcripts using RT-PCR, which we semiquantitatively analyzed across samples using transcripts of the cotransfected plasmid pDO hGH as relative internal control. The fragments were resolved by agarose gel electrophoresis and stained with ethidium bromide (black and white reversed). The identity of the spliced transcripts was verified by sequencing. (C) Diagram showing the increase in reporter 5' splice site activation relative to hGH transcripts in response to pDO *Am*-Tra2 cotransfection. We determined the increase in reporter 5' splice site activation that was induced by expression of the *Am*-Tra2 protein; the increase was measured as the ratio of the band intensities corresponding to spliced reporter transcripts in the presence and absence of *Am*-Tra2 protein relative to the band intensities of the respective hGH transcripts. We measured the fold changes in four biological replicates and analyzed each biological replicate using three independent PCR reactions.To identify putative *Am*-Tra2-dependent ESE sequences, we cloned five 33- to 40-nt-long overlapping fragments spanning the 155-nt-long sequence upstream of $D3^M$ (pRR fragments, Fig. 6A, Tab. S2) into a splice reporter construct (pOpI D1-*env*/eGFP, Fig. S4; (NISSEN *et al.* 2012b)) that was modified from (VOELKER *et al.* 2012) for use in *Sf*21 cells. This splice reporter contains a part of the HIV-1 *env* gene including the suboptimal 5' splice site D1 whose activation requires the presence of an ESE sequence (VOELKER *et al.* 2012). Therefore, if the tested sequences contain ESE sequences, the reporter 5' splice site D1 will be activated.

We transiently transfected *Sf*21 cells with plasmids containing the five different reporter constructs (pOpI D1-*env*/eGFP-pRR plasmids) and plasmid pDO hGH. We used RT-PCR to amplify fragments corresponding to the spliced reporter transcripts to semiquantitatively estimate the amount of reporter transcript splicing relative to the hGH transcript (Fig. 6B). Next, we determined the amount of spliced reporter transcript when cells were cotransfected with the *Am*-Tra2 protein expressing plasmid pDO *Am*-Tra2.

For fragment pRR1, we observed that the level of amplification substantially increased if the *Am*-Tra2 protein was coexpressed (Fig. 6C), suggesting strong *Am*-Tra2 protein-dependent activation of the reporter 5' splice site. For fragments pRR2 to pRR5, we observed only a low increase, suggesting that expression of the *Am*-Tra2 protein had very little effect on the activation of the 5' splice site.

From these results, we conclude that *Am*-Tra2-dependent ESE sequences are located in the pRR1 fragment. Notably, we found three sequence motifs (CAATR) with similarity to known Tra2-binding sites (CAAGR) in the *tra2* transcripts of *Drosophila* (QI *et al.* 2007). To identify those putative ESE sequences, we introduced a series of double point mutations in which we replaced four adenines with cytosines to silence putative ESEs (fragments pRR1-mut 1 to pRR1-mut5). Furthermore, we created one fragment (pRR1-mut6) that contained all tested mutations. The six

mutant sequence variants were also tested for *Am*-Tra2 protein-dependent 5' splice site activation (Fig. 7A). We cotransfected cells with the different reporter constructs, pDO hGH and increasing amounts (0.5, 1 and 2 µg) of the pDO *Am*-Tra2 plasmid. We again semiquantitatively estimated the activation of the reporter 5' splice site by RT-PCRs relative to the hGH transcripts (Fig. 7B, C).

All the introduced mutations negatively affected *Am*-Tra2 protein-dependent activation of the reporter 5' spice site (Fig. 7B, C). The reduction was strongest for the fragments pRR1-mut2 and pRR1-mut3. The fragments pRR1-mut1 and pRR1-mut5 had a weaker effect on the activation of the reporter 5' splice site. The fragment pRR1-mut4 had the least effect on splicing activation, and the amount of activated reporter splicing was comparable to the unmutated control when 2 µg of plasmid pDO *Am*-Tra2 was used to cotransfect the cells. Notably, the different mutations did not seem to have additive effects because splicing activation was not further decreased when all tested mutations were combined in the fragment pRR1-mut6. However, we observed a much stronger reduction in Tra2-dependent reporter splicing when the first CAATR motif was mutated compared to when the last two motifs were mutated (pRR1-mut1 and pRR1-mut4), which indicates that other parts of sequence pRR1 are important for the binding of the Tra2 protein. These results suggest that the entire sequence of pRR1 may be required for the *Am*-Tra2-dependent ESE function, with the exception of the most downstream nucleotides that were mutated in sequence pRR1-mut4.

Next, we tested whether the putative ESE sequences in pRR1 are necessary to direct *Am*-Tra2 protein-mediated male splicing of the *fem* minigene transcripts. We introduced the mutations from fragment pRR1-mut2, which showed the least ESE function in the splice reporter assay, into the *fem* minigene and studied the activation of the male 5' splice site ($D3^M$) in response to *Am*-Tra2 protein expression (Fig. S5). The levels of spliced transcripts were not detectably different with the mutated and non-mutated minigene. This observation suggests that silencing the splicing-enhancing function of the pRR1 sequence had no effect on the *Am*-Tra2 protein-dependent activation of the male splice site $D3^M$. This result indicates that other *Am*-Tra2-dependent ESEs may be involved in regulation of the male 5' splice site and splicing of male *fem* transcripts.

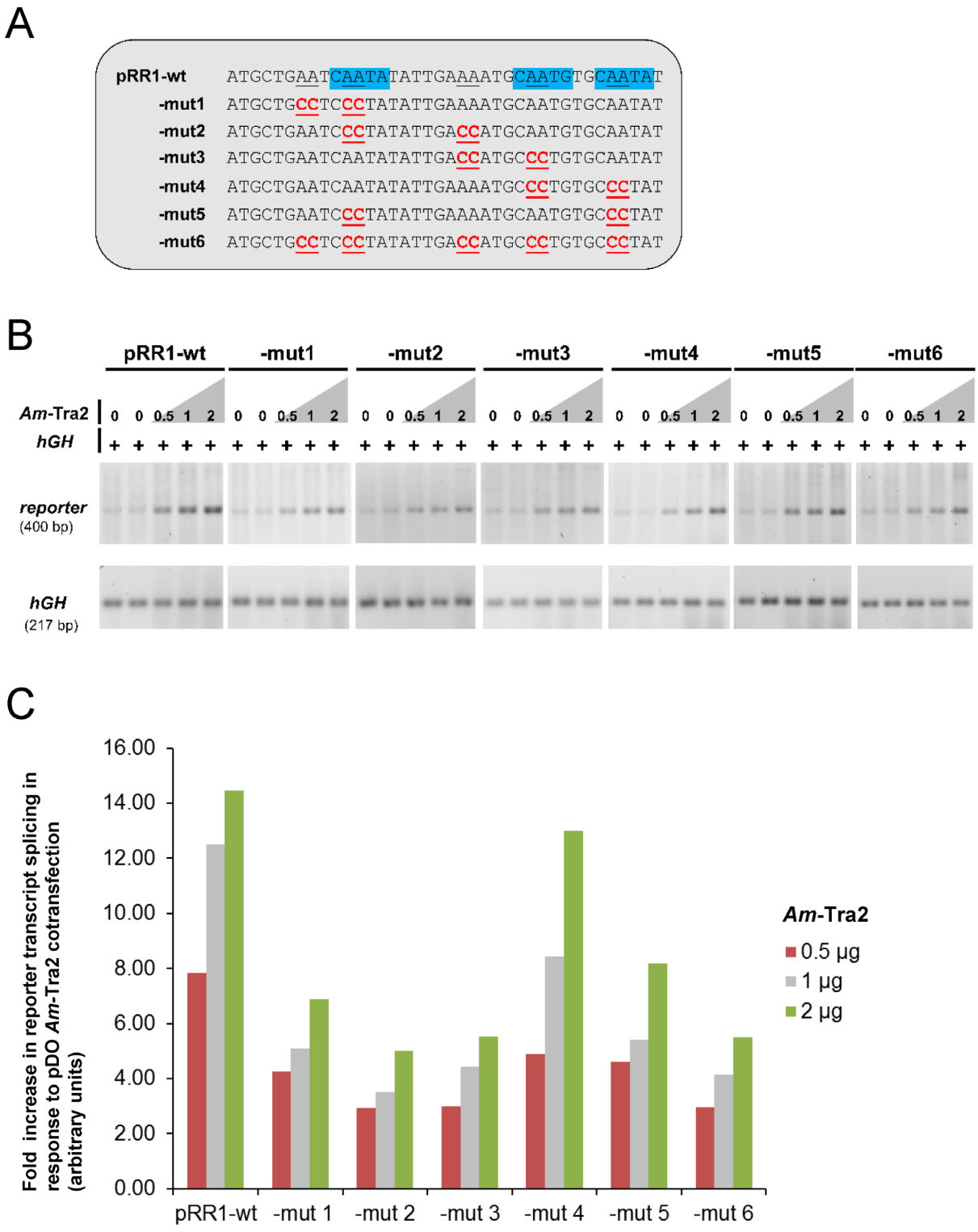

Figure 7. Identification of putative *Am*-Tra2 protein-dependent ESE sequences in the pRR1 fragment using point mutations. (A) Point mutations that were introduced into the pRR1 fragments. The nucleotide changes compared to the wild type sequence are underlined and colored in red. Blue boxes indicate predicted *Am*-Tra2 protein-dependent ESEs. (B) The effect of mutations on the *Am*-Tra2 protein-dependent activation of the reporter 5' splice site D1. $1x10^6$ *Sf*21 cells were transiently transfected with 1 µg of the respective mutant pRR1 reporter clone (plasmids pOpl D1-*env*/eGFP-pRR1-mut1-6) and 1 µg of plasmid pDO hGH. Cells were cotransfected with increasing amounts of plasmid pDO *Am*-Tra2 (0, 0.5, 1 and 2 µg) as indicated. Using RT-PCR, we amplified fragments corresponding to the spliced reporter transcripts, which we semiquantitatively analyzed across

samples using transcripts of the cotransfected plasmid pDO hGH as relative internal controls. The fragments were resolved by agarose gel electrophoresis and stained with ethidium bromide (black and white reversed). The identity of the spliced transcripts was verified by sequencing. (C) Diagram showing the increase in reporter 5' splice site activation relative to hGH transcripts in response to pDO *Am*-Tra2 cotransfection. We determined the increase in reporter 5' splice site activation that was induced by expression of the *Am*-Tra2 protein; the increase was measured as the ratio of the band intensities corresponding to spliced reporter transcripts in the presence and absence of *Am*-Tra2 protein relative to the band intensities of the respective hGH transcripts. We measured the fold changes in three biological replicates and analyzed each biological replicate using three independent PCR reactions.

Discussion

We studied the induction, implementation and maintenance of honeybee sexual development that is induced by the allelic composition of the primary signal *csd* and established through the processing of the male and female *fem* transcripts. The Fem protein implements female development by regulating the female splicing of *Am-dsx* pre-mRNAs. Fem proteins also maintain the female-determined state through an autoregulatory feedback loop (GEMPE *et al.* 2009) because it was previously shown that *csd* loses its ability to activate female regulation during larval development (GEMPE *et al.* 2009). In the absence of the Fem protein, male *Am-dsx* splicing and male development are activated by default (GEMPE *et al.* 2009).

First, we showed that expression of the Csd and Fem proteins induces the production of female *Am-dsx* and, in the case of Fem protein expression, female *fem* transcripts in males. However, because this is a temporary and incomplete switch, we conclude that the *fem* feedback loop acts relatively slowly and that a female signal (provided by the active Csd or Fem proteins) must be present for a longer period to fully activate the loop, which cannot be achieved using embryonic mRNA injections. Furthermore, we found evidence that unknown honeybee-specific factors are involved in the Csd and Fem protein-directed splicing of female *fem* transcripts. We showed that the male-specific splicing of *fem* transcripts is controlled by the *Am*-Tra2 protein most likely by binding to putative ESE motifs in the *fem* pre-mRNA. We detected an interaction between the *Am*-Tra2 and Fem proteins but not between the *Am*-Tra2 and Csd proteins.

We conclude that the *Am*-Tra2 protein serves a dual role in the processing of *fem* transcripts because it directly controls male *fem* splicing and is an important cofactor of the Fem protein in the autoregulatory production of female *fem* transcripts through the *fem* feedback loop that implements and maintains female development independently in every cell. We propose that the weak feedback loop together with the dual role of the *Am*-Tra2 protein acts as a filter to measure and exclude weak female signals that might result from hemi-/homozygous Csd proteins. This filter prevents the erroneous activation of female development in some or all cells in male individuals.

The *fem* feedback loop has low responsiveness to female signals

We first found evidence that female activity provided by Csd or Fem proteins is required for a longer period of time during embryogenesis to fully induce and implement the female development at the level of *fem* and *Am-dsx* transcript splicing.

We demonstrated this in haploid male embryos by introducing *fem* and *csd* mRNAs in early embryogenesis during the syncytial stage; this temporarily induced the production of female *Am-dsx* transcripts, which reverted back into the male mode at later stages. Our data are consistent with a previous study in which *fem* mRNA caused the production of endogenous female *fem* transcripts (GEMPE *et al.* 2009), and we demonstrated that the female signal is transmitted to the next level of *Am-dsx* splicing. However, when *csd* mRNAs were provided, we found a partially female *Am-dsx* splicing pattern but no induction of female *fem* mRNAs. It is possible that the provided Csd activity directly induced the production of female *Am-dsx* transcripts. However, it was shown in a previous study that *fem* activity is required for the production of female *Am-dsx* transcripts (GEMPE *et al.* 2009).

We propose that the female activity provided by the injection of *csd* mRNA had a weaker effect on *fem* transcript splicing than the injection of *fem* mRNA, and therefore, the production of female *fem* mRNAs had completely reverted back to the default male state when the embryos were analyzed (48-72 h after injection), whereas the downstream female *Am-dsx* signal was still partially in the female mode.

From these results, we conclude that (i) both *csd* and *fem* can direct the processing of female *Am-dsx* mRNAs, (ii) the period/activity is not sufficient to completely switch processing to the female mode, (iii) the positive feedback loop does not respond to a female signal at a level that is sufficient to reach a self-sustainable state of female mRNA production once low amounts of female *fem* mRNAs are produced.

We exclude the possibility that we did not provide sufficient female activity because we observed that increasing the amount of injected *fem* and *csd* mRNA inhibited female development and resulted in male development in females. This was not due to malfunction of the protein encoded by the respective mRNAs because we observed a female-determining activity (at the level of splicing) of these mRNAs in haploid male embryos. It was also shown in *Drosophila* that overexpression of female-determining genes can have a contrary, sex-reversing effect in females.

Temperature-induced overexpression of Tra proteins driven by a heat-shock promoter induced male development in females in addition to feminizing males (KEISMAN & BAKER 2001). Weaker Tra expression at a lower temperature did not compromise female development in females and still switched males into females. Therefore, we propose the alternative explanation that the period for which the injected *csd* or *fem* mRNAs provided female activity was not sufficient to fully activate female development. This conclusion is supported by the finding that no female *fem* or *Am-dsx* transcripts were detected later in development in 1st or 4th instar diploid male larvae that were treated with embryonic injection of *csd* or *fem* mRNA.

However, the period during which continuous Csd activity has to be provided does not extend beyond the larval stage. This was concluded in a previous study in which male *fem* and *Am-dsx* transcripts were predominantly found in *fem*-RNAi treated females under conditions of resumed *fem* transcription, suggesting that the Csd proteins had lost the ability to switch *fem* and *Am-dsx* transcript splicing into the female mode (GEMPE *et al.* 2009). We propose that the female activity has to be provided constantly for a specific period during larval development, which cannot be achieved by mRNA injections. We propose that a heat-shock promoter-based expression technique (SCHULTE *et al.* 2013) might provide Csd / Fem proteins long enough to fully activate the *fem* feedback loop and induce female development.

The *Am*-Tra2 protein directly regulates male *fem* splicing

We demonstrated that the *Am*-Tra2 protein is sufficient to direct the processing of male *fem* transcripts and that the production of female *fem* transcripts requires beside *Am*-Tra2, Fem and Csd other thus far not identified honeybee-specific factors.

We demonstrated by using a *fem* minigene that both the male and female splicing of *fem* requires factors not present in lepidopteran *Sf*21 insect cells. Without the addition of honeybee-specific factors, we only found transcripts that were processed via an alternative ‘strong’ 5’ splice site, which contained the male premature stop codon. Furthermore, we identified a corresponding splice variant in the honeybee and showed that this variant is not sexually regulated.

The alternative splice site might have a role in the regulation of sex-specific *fem* splicing. In *D. melanogaster*, there was evidence indicating that the regulated

activation of the female-specific 5' splice site of *fruitless* (*fru*) transcripts requires a competing non-sex-specific 5' splice site (LAM *et al.* 2003).

We demonstrated that the *Am*-Tra2 protein is a key component directing the male processing of *fem* transcripts. We observed that expression of the *Am*-Tra2 protein directs the activation of the male 5' splice site of exon 3 in the minigene pre-mRNA. We identified a sequence (pRR1) upstream of the 5' splice site that is sufficient for *Am*-Tra2 protein-dependent activation of the 5' splice site of a reporter construct, suggesting that the *Am*-Tra2 protein directly interacts with putative ESE sequences in the pRR1 fragment.

pRR1 contains putative ESE motifs (CAATR) with some similarity to known Tra2 protein binding sites in *D. melanogaster* (QI *et al.* 2007). Mutation of the pRR1 sequence greatly reduced the splicing-enhancing function in the reporter transcripts. However, we observed a much stronger reduction in Tra2-dependent reporter splicing when the first CAATR motif was mutated compared to when the last two motifs were mutated, which indicates that other sequence elements are important for Tra2-dependent splicing activation. Further EMSA (electrophoretic mobility shift assay) experiments will enable the identification of the binding sites of the *Am*-Tra2 protein within the pRR1 sequence.

However, the introduction of mutations did not affect the male splicing of the minigene transcripts. This might have occurred because i) the introduced mutations did not sufficiently inactivate the function of the pRR1 ESEs or ii) there are additional ESEs within the minigene that can compensate for the loss of the pRR1 ESE sequences. A similar mechanism in which several functionally redundant ESE sequences control sex-specific splicing was described in the *D. melanogaster fru* gene. It was shown that the presence of any one of three similar ESE sequences is sufficient to activate the regulated splice site (HEINRICHS *et al.* 1998; LAM *et al.* 2003).

Together, our results suggest that the *Am*-Tra2 protein is directly involved in the processing of the male *fem* transcript. Tra2 protein is required for the female splicing of *tra* and *dsx* transcripts in different species including the honeybee (BURGHARDT *et al.* 2005; SCIABICA & HERTEL 2006; SALVEMINI *et al.* 2009; SARNO *et al.* 2010; NISSEN *et al.* 2012a; SCHETELIG *et al.* 2012; SHUKLA & PALLI 2013). Apart from the honeybee, the involvement of Tra2 proteins in the regulation of male splicing has only been reported in the coleopteran *T. castaneum* and in the germline of *D. melanogaster.* However, the role of the *Tc*-Tra2 protein in controlling male *Tc-tra*

mRNA splicing is indirect because *Tc-tra2* is not required for the production of male *tra* transcripts. Knockdown of *Tc-tra2* function by RNAi led to the production of both male and female *Tc-tra* transcripts in males and females (SHUKLA & PALLI 2013), suggesting that the *Tc*-Tra2 protein represses female *Tc-tra* splicing in males instead of directly activating male-specific splice sites. In *D. melanogaster,* the role of *tra2* is also indirect because the Tra2 protein limits its own expression in a negative feedback loop, which is required for proper spermatogenesis (MATTOX & BAKER 1991; MATTOX *et al.* 1996; MCGUFFIN *et al.* 1998).

Therefore, we propose that the function of the *Am*-Tra2 protein in controlling male *fem* splicing newly evolved in the honeybee, in addition to its role in female *fem* mRNA production (NISSEN *et al.* 2012a). Therefore, the *Am*-Tra2 protein has a dual role in the regulation of male and female *fem* splicing.

We found that expression of the *Am*-Tra2, Fem and Csd proteins is not sufficient to mimic female splicing in lepidopteran *Sf*21 cells. However, the induction of a partial feminization by providing Fem or Csd activity by injecting mRNA of *fem* or a second *csd* allele in haploid males showed that both proteins are critical for establishing the female pathway. These observations suggest that additional honeybee-specific components are required for female *fem* splicing, which are not provided by the general splicing machinery in *Sf*21 cells.

We cannot exclude the possibility that regulatory elements were absent in the minigene because of shortening of the intronic sequences or removal of the flanking exons or introns. However, such elements are typically located close to the regulated splice sites (GRAVELEY *et al.* 1998; SCIABICA & HERTEL 2006; SHEN & GREEN 2006; HERTEL 2008), and we were unable to activate the female splice sites even when we used a minigene that contained intron 4 (data not shown).

The *Am*-Tra2 protein is a cofactor of Fem but not Csd

RNAi-mediated knockdown of the *csd*, *fem* or *Am-tra2* gene resulted in the loss of female *fem* mRNAs in females (GEMPE *et al.* 2009; NISSEN *et al.* 2012a), suggesting that all these components are essential for female-specific *fem* splicing. Our observation (Fig. 3, Tab. S3) that the Fem and *Am*-Tra2 proteins physically interact in cells, whereas homo- or heteroallelic Csd and *Am*-Tra2 proteins do not interact suggests that the Fem and *Am*-Tra2 proteins are cofactors that function

together to mediate the autoregulatory splicing of female *fem* transcripts. However, another honeybee-specific (thus far unidentified) factor is required to mimic female splicing in *Sf*21 cells. The *Am*-Tra2 protein contains an RNA-binding domain (RBD) that is supposed to directly interact with the RNA (NISSEN *et al.* 2012a), suggesting a similar mechanism as it is found for sex-specific regulation of *dsx* and *tra2* transcript splicing in *Drosophila* (BURTIS & BAKER 1989; HEDLEY & MANIATIS 1991; RYNER & BAKER 1991; INOUE *et al.* 1992; SCIABICA & HERTEL 2006; QI *et al.* 2007). Because Csd proteins lack a RBD and the *Am*-Tra2 protein does not bind to Csd, we propose that Csd operates in concert with another RBD-containing protein.

The *fem* feedback loop and the dual role of the *Am*-Tra2 protein act as a filter to distinguish between male and female activity states

The binary switch gene *tra*, which is called *fem* in the honeybee, is a common regulator in holometabolous insects that implements and maintains sexual development (PANE *et al.* 2002; LAGOS *et al.* 2007; RUIZ *et al.* 2007b; CONCHA & SCOTT 2009; GEMPE *et al.* 2009; HEDIGER *et al.* 2010; VERHULST *et al.* 2010; SCHETELIG *et al.* 2012; SHUKLA & PALLI 2012). Functional Fem/Tra proteins are encoded by the female *fem/tra* transcripts whose continuous supply is ensured by a positive feedback loop. This feedback loop provides a continuous signal that maintains sexual identity by locking the *fem/tra* switch in a female- or male-determined state. Fem/Tra proteins in turn regulate the sex-specific splicing of *dsx* transcripts (CLINE & MEYER 1996; PANE *et al.* 2002; HEDIGER *et al.* 2004; LAGOS *et al.* 2007; HASSELMANN *et al.* 2008b; ALVAREZ *et al.* 2009; CONCHA & SCOTT 2009; GEMPE *et al.* 2009; SALVEMINI *et al.* 2009; HEDIGER *et al.* 2010; VERHULST *et al.* 2010; SHUKLA & PALLI 2012). These sex-specifically spliced *dsx* transcripts encode distinct variants of DM-type transcription factors with sex-specific carboxy-terminal ends (CLINE & MEYER 1996; DEARDEN *et al.* 2006; CHO *et al.* 2007) that have a functional role in sex determination in several species (BURTIS & BAKER 1989; COSCHIGANO & WENSINK 1993; SUZUKI *et al.* 2003; HEDIGER *et al.* 2004; WILLIAMS & CARROLL 2009).

Sex-determination in the honeybee and several other insects follows these regulatory principles. Previous studies have shown that the heterozygous *csd* genotype robustly induces the female pathway in individual cells and in the progeny (BEYE *et al.* 2003; GEMPE *et al.* 2009). Sex is determined repeatedly in different cells

because this process is cell-autonomous (DRESCHER & ROTHENBUHLER 1964) and it is stably implemented at the level of the *fem* gene (GEMPE *et al.* 2009). Multiple *csd* alleles consistently induced femaleness when they were combined in various heterozygous genotypes (BEYE *et al.* 2013), and more than 53 alleles segregate at the species level (LECHNER *et al.* 2013).

Previous studies have shown that *fem* transcripts are sexually spliced, which is controlled by the *csd* gene (BEYE *et al.* 2003; GEMPE *et al.* 2009). Female *fem* transcripts encode the functional Fem protein that maintains its own production and implements female *Am-dsx* transcript splicing and female development. The production of non-productive male *fem* transcripts results in male *Am-dsx* transcript splicing and male development. Consequently, male development is the default state and female development is actively maintained through the autoregulation of the Fem protein.

Therefore, the finding that the positive *fem* feedback loop only weakly amplifies the production of female *fem* mRNAs and that the *Am*-Tra2 protein is involved in the male and the female pathway were unexpected, especially because male *fem* transcripts have no role in male sex-determination (GEMPE *et al.* 2009). We propose that the regulatory relationship of the weak feedback loop and the dual role of the *Am*-Tra2 protein functions as a filter to accurately distinguish between heterozygous and hemi-/homozygous states of the primary signal *csd.* The primary sex-determining signal of the honeybee is not a presence/absence signal such as the dominant female *tra* allele or the male M factor in *Musca* (HEDIGER *et al.* 2010), the presence of a paternal genome in *Nasonia* (VERHULST *et al.* 2010) or the *Sry* gene in the mouse (SEKIDO & LOVELL-BADGE 2008). However, the primary *csd* gene of the honeybee provides different activity states encoded by as little as five amino acid differences and length variation in the Potential Specifying Domain (PSD) between two Csd alleles (BEYE *et al.* 2013) and more than 53 alleles segregate at the species level (LECHNER *et al.* 2013).

The simulation of the decision process leading to male or female development using different scenarios provided evidence for this assumption. A positive feedback loop involving the *Am*-Tra2 protein in both male and female *fem* transcript splicing required more female activity for a longer period of time to induce female development compared to a scenario where the *Am*-Tra2 protein was only involved in female splicing (personal communication from Dr. M. Kollmann and N. Heramvand).

We propose that the combined actions of the *fem* feedback loop and the dual role of the *Am*-Tra2 protein establish a threshold that (i) filters out female background activity of hemi-/homozygous Csd proteins in males and (ii) allows the repeated and accurate implementation of the female pathway in females. Therefore, this mechanism robustly and cell-autonomously generates stable sex-determined states from numerous possible allelic combinations of the sex-determining signal Csd. A regulatory network with only a strong positive feedback loop would not have the ability to filter out incorrect female activity resulting from female background splicing.

Evolutionary implications for the honeybee sex-determination pathway

The question remains whether the role of the *Am*-Tra2 protein in male *fem* splicing newly evolved in the honeybee or whether this was the ancestral state. We propose that the critical role of the *Am*-Tra2 protein in male *fem* splicing evolved in the honeybee because the other two known examples where Tra2 proteins are involved in male-specific processes are the repression of female *Tc-tra* splicing in males of *T. castaneum* (SHUKLA & PALLI 2013) and negative autoregulation in the germline of male fruit flies (MATTOX & BAKER 1991; MATTOX *et al.* 1996; MCGUFFIN *et al.* 1998).

The finding that the *Am*-Tra2 protein is a cofactor of the Fem protein but not the Csd proteins is notable because *csd* and *fem* arose from a common ancestor, and the Csd and Fem proteins still share approximately 70 % amino acid identity (HASSELMANN *et al.* 2008b); furthermore, interaction with Tra2 proteins is a conserved function of Fem/Tra proteins, which was shown in *Drosophila* (BURTIS & BAKER 1989; HEDLEY & MANIATIS 1991; INOUE *et al.* 1992; TIAN & MANIATIS 1993; LYNCH & MANIATIS 1995; LYNCH & MANIATIS 1996; SCIABICA & HERTEL 2006) and has been proposed in several insect species (SALVEMINI *et al.* 2009; HEDIGER *et al.* 2010; NISSEN *et al.* 2012a; SCHETELIG *et al.* 2012; SHUKLA & PALLI 2013). We propose that the Csd proteins lost their ability to interact with the *Am*-Tra2 protein during the evolution of *csd* as the primary signal of honeybee sex-determination.

The results obtained in this study enable us to refine the model of sex-determination in honeybees (Fig. 8). We found that the *fem* positive feedback loop only slowly responds to Fem proteins, and we did not detect a direct response to Csd proteins, suggesting that the loop is weak. The *Am*-Tra2 protein is involved in both

the female and the male pathway. In females, the *Am*-Tra2 protein directly interacts with the Fem protein and serves as a cofactor in the positive feedback loop of female *fem* transcripts and the production of female *Am-dsx* mRNA. In males, the *Am*-Tra2 protein controls the splicing of non-productive male *fem* transcripts. Considering the enormous allelic diversity of the primary *csd* gene, we propose that this dual role of *Am-tra2* and the weak positive feedback activity might help to set the switch robustly in each cell and every offspring. The primary *csd* signal induces the regulatory relationship of the downstream components that operate as a filter to amplify and quality-check the primary signal to implement the female pathway only in individuals that are heterozygous for *csd*.

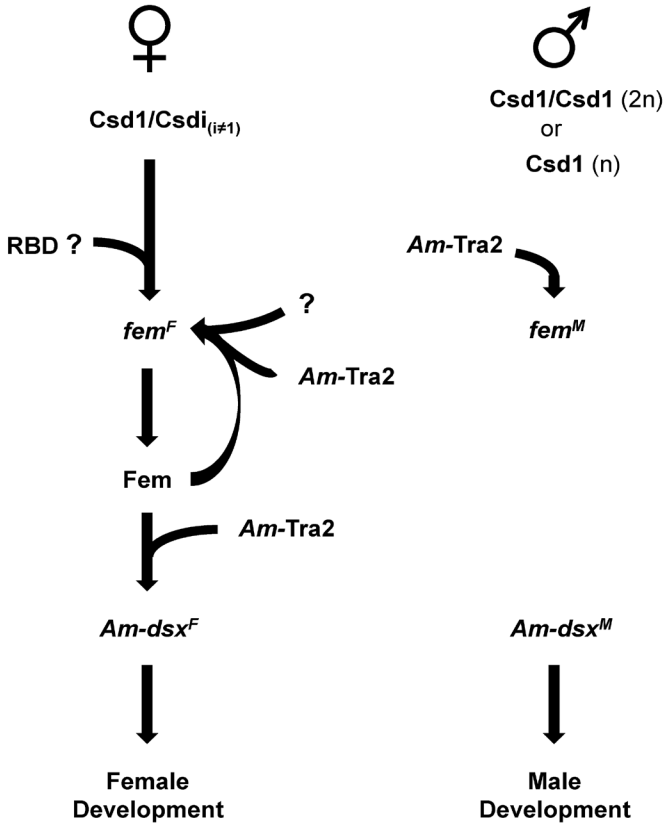

Figure 8. Model of the honeybee sex-determination pathway. In females, heterozygous Csd proteins (Csd1/Csdi) act together with an unknown RBD-containing protein to direct the production of female *fem* transcripts that encode the Fem protein. The Fem protein implements female development by promoting together with the *Am*-Tra2 protein the splicing of female *Am-dsx* transcripts. Female development is maintained by the Fem proteins that control the production of further female *fem* transcripts in a positive feedback loop involving the *Am*-Tra2 protein and an unknown factor. In males, the *Am*-Tra2 protein directs the splicing of non-productive male *fem* transcripts in the absence of functional Csd proteins (homo-/hemiallelic genotype, Csd1/Csd1 or Csd1).

Supplementary Information

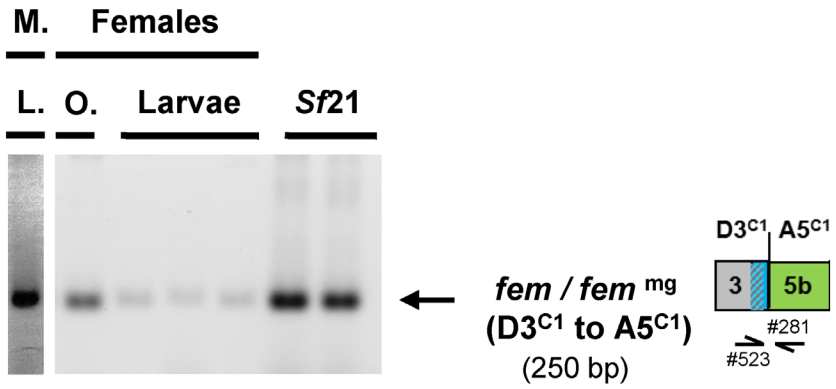

Figure S1. Use of the alternative 5' and 3' splice sites ($D3^{C1}$ and $A5^{C1}$) in female and male honeybee larvae and in lepidopteran *Sf*21 cells transfected with pDO *fem*^{mg}. Untreated honeybee larvae were reared to the 1st or 4th instar stage. As a positive control, 1x10^6 *Sf*21 cells were transiently transfected with 1 µg of pDO *fem* ^{mg}, 1 µg of pDO hGH and 1 µg of pDO *Am*-Tra2. The fragments corresponding to *fem* and *fem*^{mg} transcripts in which the splice sites $D3^{C1}$ and $A5^{C1}$ were used (250 bp) were amplified by RT-PCR, resolved by agarose gel electrophoresis and then stained with ethidium bromide (black and white reversed). The positions of the used oligonucleotide primers are indicated by arrows in the schematic transcripts shown to the right of the gel image. The identity of the spliced transcripts was verified by sequencing.

M = Male; L = whole male 4th instar larvae; O = ovaries dissected from female 4th instar larvae; Larvae = whole female 1st instar larvae; Sf21 = 1x10^6 *Sf*21 cells

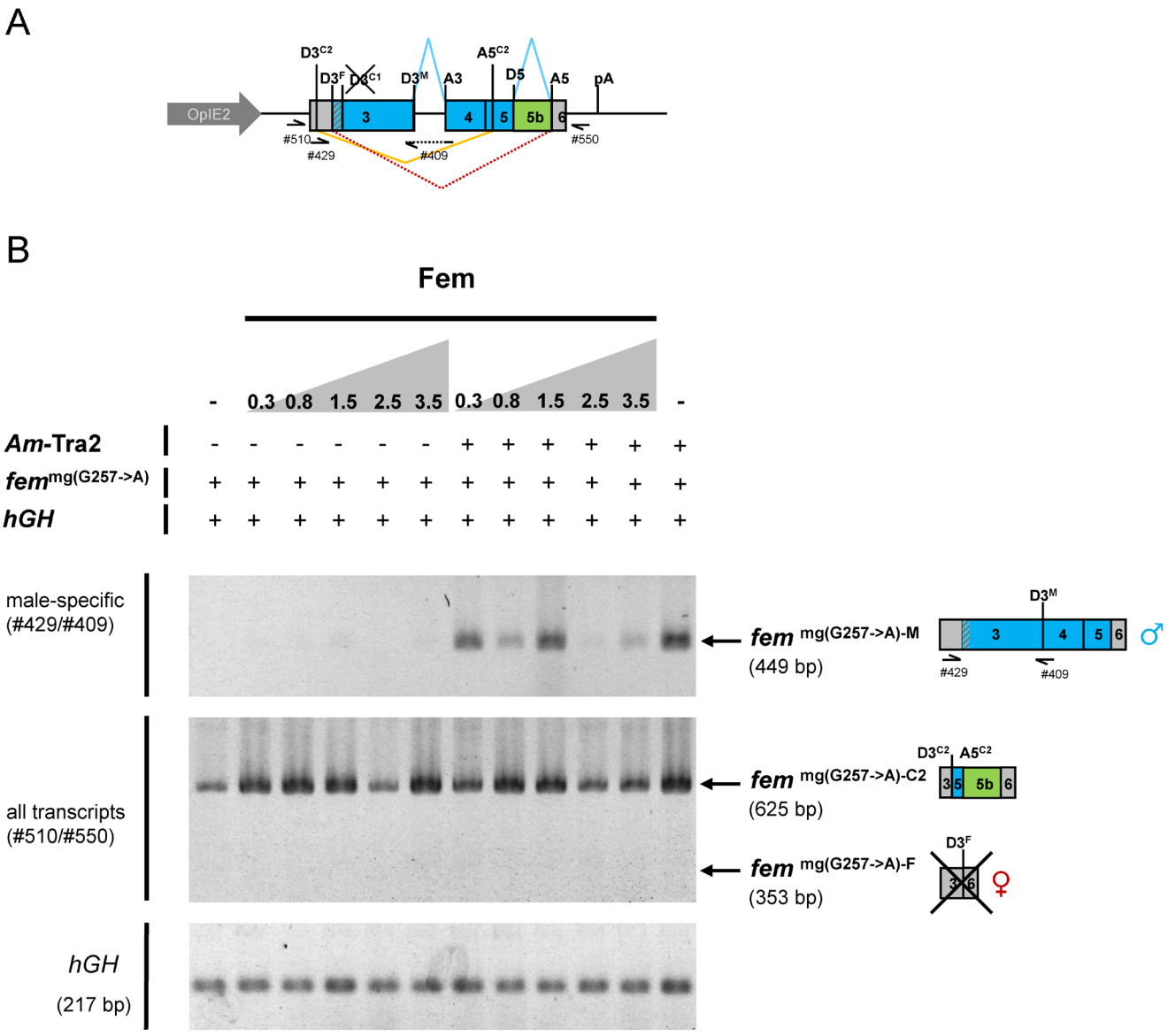

Figure S2. Processing of the mutated *fem* minigene (*fem* $^{mg(G257->A)}$) transcripts in *Sf*21 cells. (A) Intron and exon structure of the mutated minigene *fem* $^{mg(G->A)}$. The mutated minigene is identical to the unmutated minigene described in Fig. 1B, except for a single nucleotide substitution in the sequence of the 5' splice site $D3^{C1}$ (see Material & Methods). Common and male exons are indicated as in (Fig. 1A). The colors of the boxes and connecting lines indicate the putative sex-specifically processed transcripts as shown in (Fig. 1A). We did not detect use of the mutated 5' splice site $D3^{C1}$. However, the cryptic splice sites $D3^{C2}$ and $A5^{C2}$ were constitutively used as indicated by the yellow line. The 5' splice sites are labeled with a "D" and the 3' splice sites are labeled with an "A", whereas the number denotes the corresponding exon. **(B) RT-PCR analysis of the *fem* $^{mg(G->A)}$ transcripts in response to Fem protein expression.** $1x10^6$ *Sf*21 cells were transiently transfected with 1 µg of pDO *fem* $^{mg(G257->A)}$, 1 µg of pDO hGH and 1 µg of pDO *Am*-Tra2 as indicated by a "+". Increasing amounts of pDO Fem (0.3 to 3.5) were cotransfected. Fragments corresponding to the male (449 bp) and an alternatively processed transcript (625 bp) were amplified by RT-PCRs. The positions of the used oligonucleotide primers are indicated by arrows in (Fig. 1B). We used RT-PCR fragments from the mRNAs of the cotransfected pDO hGH plasmid to semiquantitatively adjust the transfection efficiency across the samples. The fragments were resolved by agarose gel electrophoresis and stained with ethidium bromide (black and white reversed). The processed transcripts were characterized by sequencing. The structure of the transcripts is schematically shown to the right of the gel picture.

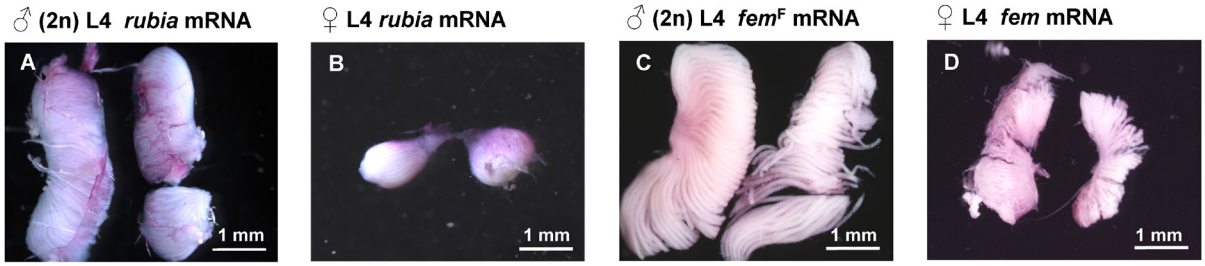

Figure S3. Gonad development of diploid 4th instar larvae treated with embryonic mRNA injection. (A) a pair of normally differentiated testes from a diploid male and (B) a pair of normally differentiated ovaries from a female both treated with embryonic injection of *rubia* mRNA. (C) a pair of normally differentiated testes from a diploid male and (D) a pair of testis-like differentiated gonads from a female both treated with an embryonic injection of *fem* mRNA. The reddish coloring is due to staining with aceto-orcein during dissection. Scale bars, 1 mm.

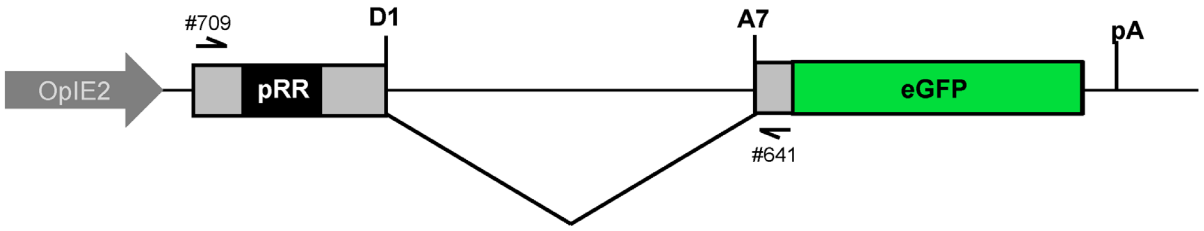

Figure S4. Schematic representation of the pOpI D1-env/eGFP splicing reporter construct. This splicing reporter designed by NISSEN *et al.* (2012b) is a modified version of the splicing reporter SV D1-env/eGFP of VOELKER *et al.* (2012) in which the SV40 promoter was replaced with the baculovirus-derived OpIE2 promoter (THEILMANN & STEWART 1992). The black box indicates the position at which the pRR sequences were inserted. Activation of the 5' splice site D1 requires an ESE sequence. eGFP = coding sequence (CDS) of the enhanced Green Fluorescent Protein that is expressed from the spliced mRNA. Arrows indicate the primer binding sites. pA = polyadenylation signal.

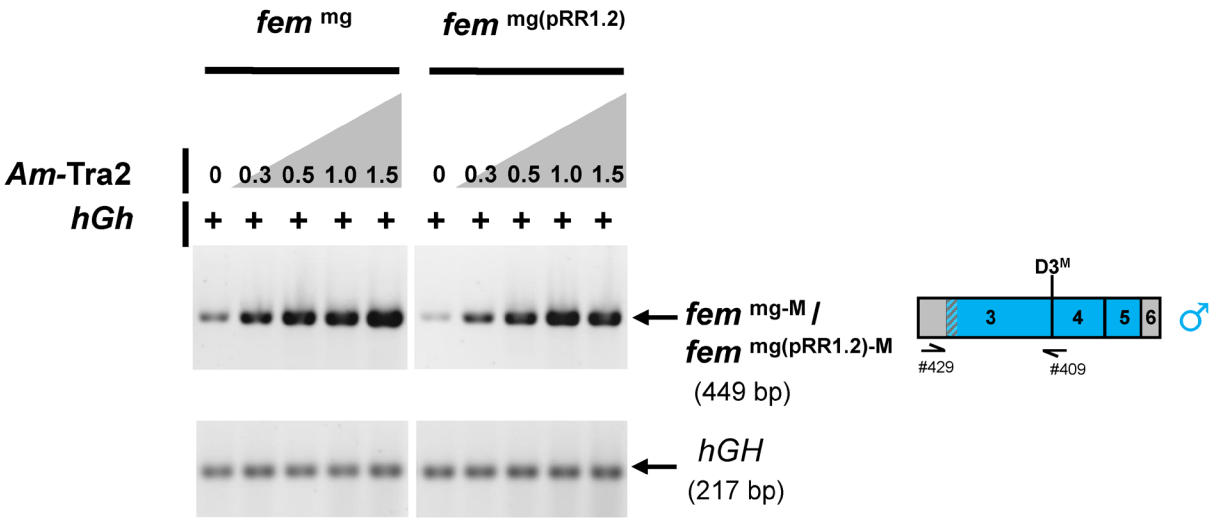

Figure S5. RT-PCR analysis of the *fem* mg and *fem* mg(pRR1.2) transcripts in response to *Am*-Tra2 protein expression in *Sf*21 cells. 1x10^6 *Sf*21 cells were transiently transfected with 1 µg of plasmid pDO *fem* mg or 1 µg of plasmid pDO *fem* mg(pRR1.2) and 1 µg of plasmid pDO hGH as denoted by a "+". Increasing amounts of plasmid pDO *Am*-Tra2 (0 to 1.5 µg) were cotransfected. Fragments

representing the male *fem* minigene splicing variant (449 bp) were amplified by RT-PCR, which we semiquantitatively analyzed across samples using transcripts of the cotransfected plasmid pDO hGH as relative internal control. The positions of the used oligonucleotide primers are indicated by arrows in the schematic representations of the male transcript to the right of the gel image. The fragments were resolved by agarose gel electrophoresis and stained with ethidium bromide (black and white reversed). The processed transcripts were characterized by sequencing.

Figure S6. Complete sequence of plasmid pDO.

GGATCATGATGATAAACAATGTATGGTGCTAATGTTGCTTCAACAACAATTCTGTTGAACTGTGTT
TTCATGTTTGCCAACAAGCACCTTTATACTCGGTGGCCTCCCCACCACCAACTTTTTTGCACTGCA
AAAAAACACGCTTTTGCACGCGGGCCCATACATAGTACAAACTCTACGTTTCGTAGACTATTTTAC
ATAAATAGTCTACACCGTTGTATACGCTCCAAATACACTACCACACATTGAACCTTTTTGCAGTGC
AAAAAAGTACGTGTCGGCAGTCACGTAGGCCGGCCTTATCGGGTCGCGTCCTGTCACGTACGAA
TCACATTATCGGACCGGACGAGTGTTGTCTTATCGTGACAGGACGCCAGCTTCCTGTGTTGCTAA
CCGCAGCCGGACGCAACTCCTTATCGGAACAGGACGCGCCTCCATATCAGCCGCGCGTTATCTC
ATGCGCGTGACCGGACACGAGGCGCCCGTCCCGCTTATCGCGCCTATAAATACAGCCCGCAAC
GATCTGGTAAACACAGTTGAACAGCATCTGTTCGAATTTAAAGCTTGGTACCGAGCTCGGATCCA
CTAGTCCAGTGTGGTGGAATTCATGGCCTCCTCCGAGGATGTCATCAAAGAGTTTATGAGATTTA
AGGTCAAGATGGAGGGAAGCGTCAACGGACACGAGTTCGAGATTGAGGGAGAAGGAGAAGGCC
GGCCTTACGAGGGCACACAAACCGCTAAGCTCAAGGTCACAAAAGGAGGACCCCTCCCCTTCTC
CTGGGATATTCTGAGCCCTCAGTTCCAGTACGGAAGCAAAGCCTATGTTAAACACCCTGCCGACA
TCCCTGACTATCTGAAGCTCTCCTTCCCTGAAGGCTTCAAGTGGGAGAGATTCATGAACTTCGAG
GACGGAGGCGTGGTGACAGTCACACAAGATAGCACCCTCCAGGACGGAGAGTTTATTTATAAGG
TGAAACTCAGAGGAACCAACTTCCCCTCCGATGGCCCTGTCATGCAAAAAAAAACAATGGGATGG
GAAGCCTCCACCGAGAGAATGTATCCTGAGGATGGCGCTCTGAAAGGCGAAATTAAAATGAGAC
TGAAACTCAAAGACGGAGGACACTACGATGCCGAGGTCAAAACAACCTACAAGGCCAAGAAACA
AGTGCAGCTGCCTGGCGCCTACATGACTGATATTAAACTCGACATTATCAGCCATAATGGGGACT
ACACCATCGTGGAACAATATGAGAGAGCTGAGGGCAGACATAGCACAGGCGCTGGATCCGCGG
CCGCGAGCTCACTAGTCATATGTTCTAGAATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGT
GGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGA
GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCT
GCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTA
CCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
CGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGC
GACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTG
GGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGA
ACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGA
CCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCT
GAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGA
GTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAACCGCGGTTCGAA
GGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCATCACCATCA
CCATTGAGTTTATCTGACTAAATCTTAGTTTGTATTGTCATGTTTTAATACAATATGTTATGTTTAAA
TATGTTTTTAATAAATTTTATAAAATAATTTCAACTTTTATTGTAACAACATTGTCCATTTACACACTC
CTTTCAAGCGCGTGGGATCGATGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGG
GATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCG
CGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGT
CAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCG
TGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAG
CGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGC
TGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCT
TGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGC
AGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTA
GAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGC
TCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTAC
GCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGA

ACGAAAACTCACGTTAAGGGATTTTGGTCATGATGATAAACAATGTATGGTGCTAATGTTGCTTCA
ACAACAATTCTGTTGAACTGTGTTTTCATGTTTGCCAACAAGCACCTTTATACTCGGTGGCCTCCC
CACCACCAACTTTTTTGCACTGCAAAAAAACACGCTTTTGCACGCGGGCCCATACATAGTACAAA
CTCTACGTTTCGTAGACTATTTTACATAAATAGTCTACACCGTTGTATACGCTCCAAATACACTACC
ACACATTGAACCTTTTTGCAGTGCAAAAAAGTACGTGTCGGCAGTCACGTAGGCCGGCCTTATCG
GGTCGCGTCCTGTCACGTACGAATCACATTATCGGACCGGACGAGTGTTGTCTTATCGTGACAG
GACGCCAGCTTCCTGTGTTGCTAACCGCAGCCGGACGCAACTCCTTATCGGAACAGGACGCGCC
TCCATATCAGCCGCGCGTTATCTCATGCGCGTGACCGGACACGAGGCGCCCGTCCCGCTTATCG
CGCCTATAAATACAGCCCGCAACGATCTGGTAAACACAGTTGAACAGCATCTGTTCGAATTCGTT
CAGCACCTAGGAGCAGCCTTAAGAGCAGCCTCGAGATCTTAGTTTGTATTGTCATGTTTTAATACA
ATATGTTATGTTTAAATATGTTTTTAATAAATTTTATAAAATAATTTCAACTTTTATTGTAACAACATT
GTCCATTTACACACTCCTTTCAAGCGCGTGGCCGGGCTGCAGCACGTGTTGACAATTAATCATCG
GCATAGTATATCGGCATAGTATAATACGACTCACTATAGGAGGGCCACCATGGCCAAGTTGACCA
GTGCCGTTCCGGTGCTCACCGCGCGCGACGTCGCCGGAGCGGTCGAGTTCTGGACCGACCGG
CTCGGGTTCTCCCGGGACTTCGTGGAGGACGACTTCGCCGGTGTGGTCCGGGACGACGTGACC
CTGTTCATCAGCGCGGTCCAGGACCAGGTGGTGCCGGACAACACCCTGGCCTGGGTGTGGGTG
CGCGGCCTGGACGAGCTGTACGCCGAGTGGTCGGAGGTCGTGTCCACGAACTTCCGGGACGCC
TCCGGGCCGGCCATGACCGAGATCGGCGAGCAGCCGTGGGGGCGGGAGTTCGCCCTGCGCGA
CCCGGCCGGCAACTGCGTGCACTTCGTGGCCGAGGAGCAGGACTGACCGACGCCGACCAACAC
CGCCGGTCCGACGGCGGCCCACGGGTCCCAGGGGGGTCGACCTCGAAACTTGTTTATTGCAGC
TTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATT
CTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCT

Table S1. Sequence and hydrogen bond score of the 5' splice sites of the fem gene. Exonic positions +1 and +2 are shown in bold. Nucleotides identical to the honeybee U1 snRNA are shown in upper case.

5' splice site	U1 snRNA binding site	H-bond score
D1	tAt**GT**AAGTAT	17.40
D2	aAa**GT**AtGTAT	14.10
$D3^{F}$	CgG**GT**AAaatg	12.40
$D3^{M}$	tAt**GT**AAGcAc	12.60
$D3^{C1}$	aAG**GT**AtGTAT	18.80
$D3^{C1}$(G->A)	aagatatgtat	-*
$D3^{C2}$	aAG**GT**ActcAa	12.30
D4	aAa**GT**AtGTAT	14.10
D5	aAG**GT**AtGTAT	18.80
D6	CAG**GT**AtaTtc	14.00
D7	CAG**GT**ttGTtT	13.70
D8	tgG**GT**AAcata	12.00
D9	aAa**GT**AcGTAa	12.80
D10	ggG**GT**AAGTAT	20.00
D11	Cta**GT**AtGTtT	11.20
Ideal	CAG**GT**AAGTAT	23.80

* H-bond score can only be calculated for splice sites with a **GT** in positions +1 to +2

Table S2. Sequences of pRR 1 to 5.

pRR1	ATGCTGAATCAATATATTGAAAATGCAATGTGCAATAT
pRR2	TTTAATTCATCAATGAAGAAAATTTGGACAATGCTGAA
pRR3	GTGAAAATATTTTATGAAGTTAATAACATATTTTTAATTC
pRR4	TTATTGTCATATATCTTTTATAAGTGTGAAAAT
pRR5	TAAATATAAGTTTCTACAAACAAGAAAAATTTATTGTC

Table S3. FLIM-FRET measurement data and statistical analyses.

Donor	**Csd^{25}-Cerulean**		**Csd^{27}-Cerulean + Csd^{25}**		**Fem-Cerulean**	
Acceptor	**-**	**+ Tra2-YFP**	**-**	**+ Tra2-YFP**	**-**	**+ Tra2-YFP**
Donor Fluorescence Lifetime (τ) in ns	3.002	2.813	2.772	2.814	2.526	2.533
	2.942	2.808	2.729	2.750	2.518	2.497
	2.900	2.857	2.805	2.766	2.773	2.516
	2.851	2.772	2.658	2.806	2.685	2.527
	2.846	2.811	2.776	2.703	2.670	2.546
	2.947	2.821	2.777	2.764	2.654	2.475
	2.849		2.782	2.800	2.773	
	2.712		2.746	2.761	2.722	
			2.611	2.692	2.714	
			2.788	2.786	2.700	
					2.757	
n	8	6	10	10	11	6
Mean (τ) in ns	2.881	2.813	3.404	3.422	3.374	2.516
SE	0.031	0.011	0.020	0.013	0.026	0.011
Levene's test		4.792		1.364		3.651
***P*-value**		0.049		0.258		0.075
***t*-test (2-tailed)**		2.038		-0.831		4.448
***P*-value**		0.073		0.417		0.000

Table S4. Gonad development of diploid 4th instar larvae treated with embryonic mRNA injection. A sample of 18 *rubia* mRNA treated and 11 *fem* mRNA treated diploid 4th instar larvae were sexed according to their *csd* genotype and dissected to determine gonad differentiation.

Treatment	**Number of Diploid Males with**		**Number of Females with**	
	Testes	Ovaries	Testes	Ovaries
rubia mRNA	6	0	0	12
fem mRNA	5	0	6	0

Table S5. Oligonucleotide sequences.

cloning of pRR sequences	
pRR1_fw	AATTCATGCTGAATCAATATATTGAAAATGCAATGTGCAATATGAGCT
pRR1_rev	CATATTGCACATTGCATTTTCAATATATTGATTCAGCATG
pRR2_fw	AATTCTTTAATTCATCAATGAAGAAAATTTGGACAATGCTGAAGAGCT
pRR2_rev	CTTCAGCATTGTCCAAATTTTCTTCATTGATGAATTAAAG
pRR3_fw	AATTCGTGAAAATATTTTATGAAGTTAATAACATATTTTTAATTCGAGCT
pRR3_rev	CGAATTAAAAATATGTTATTAACTTCATAAAATATTTTCACG
pRR4_fw	AATTCTTATTGTCATATATCTTTTATAAGTGTGAAAATGAGCT
pRR4_rev	CATTTTCACACTTATAAAAGATATATGACAATAAG
pRR5_fw	AATTCTAAATATAAGTTTCTACAAACAAGAAAAATTTATTGTCGAGCT
pRR5_rev	CGACAATAAATTTTTCTTGTTTGTAGAAACTTATATTTAG
pRR1.1_fw	AATTCATGCTGCCTCCCTATATTGAAAATGCAATGTGCAATATGAGCT
pRR1.1_rev	CATATTGCACATTGCATTTTCAATATAGGGAGGCAGCATG
pRR1.2_fw	AATTCATGCTGAATCCCTATATTGACCATGCAATGTGCAATATGAGCT
pRR1.2_rev	CATATTGCACATTGCATGGTCAATATAGGGATTCAGCATG
pRR1.3_fw	AATTCATGCTGAATCAATATATTGACCATGCCCTGTGCAATATGAGCT
pRR1.3_rev	CATATTGCACAGGGCATGGTCAATATATTGATTCAGCATG
pRR1.4_fw	AATTCATGCTGAATCAATATATTGAAAATGCCCTGTGCCCTATGAGCT
pRR1.4_rev	CATAGGGCACAGGGCATTTTCAATATATTGATTCAGCATG
pRR1.5_fw	AATTCATGCTGAATCCCTATATTGAAAATGCAATGTGCCCTATGAGCT
pRR1.5_rev	CATAGGGCACATTGCATTTTCAATATAGGGATTCAGCATG
pRR1.6_fw	AATTCATGCTGCCTCCCTATATTGACCATGCCCTGTGCCCTATGAGCT
pRR1.6_rev	CATAGGGCACAGGGCATGGTCAATATAGGGAGGCAGCATG
construction of *fem* minigene	
276: AfemEx3_AatIIf	GATCGACGTCATTAGAATCTTCAGATGGTACATCTTTATTTAGA
277: Afemln3_NcoIr	GATCCCATGGCAGCAATAACAATATAATTATTTTTATATTATTAAT
278: Bfemln3_NcoIf	GATCCCATGGATAATTTTTGTGTATTGTTTTATATATTGATATG
279: BfemEx4_NotIr	GATCGCGGCCGCTTTTCGGATAATATAAATGTTTTAGATATGTTTG
280: CfemEx5_NotIf	GATCGCGGCCGCATTAGAATCTTCAGATGGTACATCTTTATTTAG
281: Cfemln5_SalIr	GATCGTCGACCACAATAATTTTGATATTGCGTCAATCTTCAAT
282: Dfemln5_SalIf	GATCGTCGACCTGGCTTACAATTGTTAAGAAATAGTCAAT
283: DfemEx6_SacIr	GATCGAGCTCCTGATTTTTCAATATTTACAGCTAAAACTGT
309: FemEx3_AvrII_fw	GATCCCTAGGATTAGAATCTTCAGATGGTACATCTT
312: FemEx6_XhoI_rv	GATCCTCGAGCTGATTTTTCAATATTTACAGCTAAAACT
500: 5-XmnAvr-Link1	CGTTCTGCAGGCGCGCCCCCAAAGAGCATGCGGGTTTCTC
501: 3-XmnAvr-Link1	CTAGGAGAAACCCGCATGCTCTTTGGGGGCGCGCCTGCAGAACG
530: Fem-HpaI-fw	TTGATTGTTAACAATTTAAAAATTTATCATCA
531: Fem-Link-XhoI_r	CTCGAGGCCGAACCCGGAACTGATTTTTCAATATTTACAGCTAAAACT
001: fem-mg-pRR1.2_AvrII_fw	GATCCCTAGGATTAGAATCTTCAG
002: fem-mg-pRR1.2_AflII_rv	GATCCTTAAGATAAATGATATTGTGCTTACATATTGCACATTGCATGGTCA
	ATATAGGGATTCAGCATTGTCCAAATT

cloning of plasmids for FRET experiments	
011: 5-tra2-EcoRI_fw	GATCGAATTCATGAGTGACATTGAGCGAAGTA
012: 3-tra2_oe-NotI_rv	GATCGCGGCCGCATATCGACGTGGTGAATAAGAG
019:5-1: Allel 16 25 N-term	CCGAATTCAAACGAAATATATCAAATTATTCAC
066: csd25Flu_rev	GATCGCGGCCGCTTGATACGTAGGTCCAAATCTAGGATTT
020:5-1: Allel 27 29 N-term	CCGAATTCAAACGAAATATATCAAGTTATTCAC
067: csd27Flu_rev	GATCGCGGCCGCTTGATGCGTAGATCCAAATCTAGGATTTA
249: 5' femJco43	AGTCGAATTCATGAAACGGAATACAACAAATCATTC
602: fem rev NotI	AGTCGCGGCCGCCATAGGTCCAAATCTAGGATTTGG
009: SpacerFlu_for	GGCCGCGAGCTCACTAGTCATATGTT
010: SpacerFlu_rev	CTAGAACATATGACTAGTGAGCTCGC
069: EGFPFlu_for	GATCTCTAGAATGGTGAGCAAGGGCGAGGAG
070: EGFPFlu_rev	GATCCCGCGGTTACTTGTACAGCTCGTCCATGCC
330: 5`Bac His-	GATCATTAATAGATCATGGAGATAATTAAAATGATAACC
335: 3`Bac His-	GATCGTCGACCCCGGGGCCGGCGAATTCGGTTTCGGACCGAGATCCGC
154: 3-KpnIEcoRILink	AATTCGGCGGCCTTAAGGGCGGCTGATCAGGCGGCGGTAC
436: PH-Pro+Fragfor mit PST I	GATCCTGCAGCGGAATATTAATAGATCATGGAGAT
013: 3-csd25TGA_HindIII_rv	GATCAAGCTTTCATTGATACGTAGGTCCAAATCTAG
cloning of plasmids for *in vitro* mRNA synthesis	
147: 5`Allel25ATG+TGA	CATGGAATTCATGAAACGAAATATATCAAATTATTCACATCACG
174: 5-csd25NcoI	GATCCCATGGGGAAACGAAATATATCAAATTATTCACATCACG
175: 3-csd25TGASpeI	GATCACTAGTTCATTGATACGTAGGTCCAAATCTAGGATTTAG
149: 5`Allel27ATG+TGA	CATGGAATTCATGAAACGAAATATATCAAGTTATTCACATCACG
177: 5-csd2729NcoI	GATCCCATGGGGAAACGAAATATATCAAGTTATTCACATCACG
171: 3-csd2729TGASpeI	GATCACTAGTTCATTGATGCGTAGATCCAAATCTAGGATTT
431: 5'femS2-38_Nco	GATCCCATGGGGAAACGGAATACAACAAATCATTCGCATC
331: 3'fem SpeI	GATCACTAGTTTAATACATAGGTCCAAATCTAGGATTTG
084: Mars_fw_NcoI	GATCCCATGGCCTCCTCCGAGGATGTCATCA
133: 3'mars2	GATCACTAGTGGATCCAGCGCCTGTGCT
construction of pDO plasmids	
686: FEM_fw_BamHI_kurz	GATCGGATCCATGAAACGGAATACAACA
687: FEM_TAT_rv_MluI	GATCACGCGTTTAATACATAGGTCCAAATCTAGGATTTGG
471: CSD27_fw_BamHI	GATCGGATCCATGAAACGAAATATATCAAGTTATTCACATC
474: CSD27TGA_rv_MluI	GATCACGCGTTCATTGATGCGTAGATCCAAATCTA
361: csd27_AvrII_fw	GATCCCTAGGATGAAACGAAATATATCAAGTTATTCACATC
362: csd27_AflII_rv	GATCCTTAAGTCATTGATGCGTAGATCCAAATCTA
003: Csd 16 For BamHI	GATCGGATCCATGAAACGAAATATATCAAATTATTCACAT
004: CSD 16 REV MLUI	GATCACGCGTTCATTGATGTGTAGGTCCAAATCTAGG
005: CSD 18 FOR AVRLL	AGTCCTAGGATGAAACGAAATATATCAAGTTATTCACATC
006: CSD18REVAFLII2	AGCTCTTAAGTCATTGATGCGTAGGTTCAAATCTAGGA
007: CSD 18 FOR BAMH	AGTGGATCCATGAAACGAAATATATCAAGTTATTCACATC
008: CSD18REVMLUI2	AGCTACGCGTTCATTGATGCGTAGGTTCAAATCTAGGA
471: CSD27_fw_BamHI	GATCGGATCCATGAAACGAAATATATCAAGTTATTCACATC
474: CSD27TGA_rv_MluI	GATCACGCGTTCATTGATGCGTAGATCCAAATCTA

to amplify *elongation factor-1α*	
033: RT_EF566_for	CGTTCGTACCGATCTCCGGATG
034: EF-1aRev	GCTGCTGGAGCGAATGTTAC
to analyze *fem* splicing	
048: TypeII_W2	CAACATCTGATGAACTTAAACGG
045: Race_typeII_rev2	CTATTTCTGTCTTCATGTCTTGAA
053: tra-likeM_781_rev	GGATTCAAATCATCTGAAGTTAC
325: 5-fem_UTR	TTTGCAGATTAAATTTCATAAATATATATA
326: 3-fem_Ex6	TTCTCTTTCAAATATAGGCTTAGATCCTTCT
to analyze *Am-dsx* splicing	
026: DSX6FW	TGGTCACCCATTTGCCACAGAC
027: dsxrev_M	TCGTATGTCGGAGGTCCCGTTG
028: dsxrev_F	TGTTGTTGCCAATTTCGTTATCCA
to amplify hGH	
597: Hgh-SK2_fw	AGGTTTGGGGCTTCTGAATAG
465: hGh_MluI_rv	GATCACGCGTCTAGAAGCCACAGCTGCCCTCC
to analyze *fem* minigene splicing	
510: Fem-M-L_fw	GCGCGCCCCCAAAGAGCA
550: Fem-mg_NW_rv2	GAGTTCCGGGTTCGGCGC
429: fem_mini_fw3m	GTAACTTCAGACGATTTGAATCC
409: Fem_minig_ex34_r2	AGAGAAATATATGAATATATGTAAAATTTAATATATTG
523: Fem_SP8.4_fw	CAACATCTGATGAACTTAAACGG
281: CfemIn5_SalIr	GATCGTCGACCACAATAATTTTGATATTGCGTCAATCTTCAAT
to amplify splice reporter transcripts	
709: splice reporter fw	CAGCATCTGTTCGAATTTGAATTC
641: splice reporter rv	AGCACTGCACGCCGTAGGTC

Author's Contribution

Regulation of Honeybee Sex-Determination: The Dual Role of *Am*-Tra2 Proteins

Journal: Developmental Cell (under review)

Impact factor: 12.861

1st author

Author's contribution: 80 %

- Injection of mRNAs into haploid and diploid honeybee embryos
- Analysis of embryonic and larval sex-specific *fem* and *Am-dsx* splicing patterns
- Rearing of honeybee larvae
- Larval dissections
- Genotyping of honeybee larvae
- Screening for putative ESE sequences in *fem* exon 3
- Mutational analysis of putative ESE sequence pRR1
- Bioinformatic analyses
- Statistical analyses
- Authoring the manuscript
- FRET experiments (together with Stefanie Weidtkamp-Peters)
- *fem* minigene experiments (together with Inga Nissen)
- Vector construction and cloning (together with Inga Nissen and Marianne Otte)
- Experimental design (together with Inga Nissen and Marianne Otte)

Manuscript II

Independent Evolutionary Origin of *fem* Paralogous Genes and Complementary Sex Determination in Hymenopteran Insects

Vasco Koch, Inga Nissen, Björn D. Schmitt and Martin Beye

Institute of Evolutionary Genetics, Heinrich-Heine University Duesseldorf,

Duesseldorf, Germany

Manuscript in press in PLOS ONE

Abstract

The primary signal of sex determination in the honeybee, the *complementary sex determiner* (*csd*) gene, evolved from a gene duplication event from an ancestral copy of the *fem* gene. Recently, other paralogs of the *fem* gene have been identified in several ant and bumblebee genomes. This discovery and the close phylogenetic relationship of the paralogous gene sequences led to the hypothesis of a single ancestry of the *csd* genetic system of complementary sex determination in the Hymenopteran insects, in which the *fem* and *csd* gene copies evolved as a unit in concert with the mutual transfers of sequences (concerted evolution). Here, we show that the paralogous gene copies evolved repeatedly through independent gene duplication events in the honeybee, bumblebee and ant lineage. We detected no sequence tracts that would indicate a DNA transfer between the *fem* and the *fem1/csd* genes between different ant and bee species. Instead, we found tracts of duplication events in other genomic locations, suggesting that gene duplication was a frequent event in the evolution of these genes. These and other evidences suggest that the *fem1/csd* gene originated repeatedly through gene duplications in the bumblebee, honeybee and ant lineages in the last 100 million years. Signatures of concerted evolution were not detectable, implicating that the gene tree based on neutral synonymous sites represents the phylogenetic relationships and origins of the *fem* and *fem1/csd* genes. Our results further imply that the *fem1* and *csd* gene in bumblebees, honeybees and ants are not orthologs, because they originated independently from the *fem* gene. Hence, the widely shared and conserved complementary sex determination mechanism in Hymenopteran insects is controlled by different genes and molecular processes. These findings highlight the limits of comparative genomics and emphasize the requirement to study gene functions in different species and major hymenopteran lineages.

Introduction

Complementary sex determination, in which the heterozygous genotype at a certain locus determines femaleness, is widely shared in hymenopteran insects and has a deep ancestry (COOK 1993; HEIMPEL & DE BOER 2008). Thus far, the underlying gene *complementary sex determiner* (*csd*) has been identified in the western honeybee (*Apis mellifera*) by positional cloning and knockdown studies (BEYE *et al.* 2003; HASSELMANN *et al.* 2008b). The *csd* gene encodes an SR-type protein. Csd proteins derived from the heterozygous *csd* genotype induce the female sex pathway by directing the female splicing of the primary transcripts of the *fem* gene (HASSELMANN *et al.* 2008b; GEMPE *et al.* 2009). The resulting female mRNAs subsequently encode the functional Fem proteins. Csd proteins derived from the hemizygous or homozygous genotypes are not required for sex determination. The male splicing of the *fem* transcripts results by default. The male-specific exons contain a translational stop codon to prematurely stop translation. The absence of functional Fem proteins leads to the development of maleness (GEMPE *et al.* 2009). More than 14 *csd* alleles have been identified in local honeybee populations, which show an average of 3% pairwise difference in their entire amino acid encoding sequence (HASSELMANN & BEYE 2004; HASSELMANN *et al.* 2008a).

The low divergence of the honeybee *csd* and *fem* genes at synonymous sites compared to bumblebee and stingless bee sequences suggests that the *csd* gene was derived from a gene duplication event of an ancestral copy of the *fem* gene in the honeybee lineage (HASSELMANN *et al.* 2008b). The *csd* gene was shaped by positive selection shortly after it originated (HASSELMANN *et al.* 2008b; HASSELMANN *et al.* 2010). *fem* is the putative ortholog of the *transformer* (*tra*) gene (HASSELMANN *et al.* 2008b), a key sex-determining gene in *Drosophila melanogaster.* However, unambiguous homology relies on identities in a 30-amino-acid motif deduced from another dipteran ortholog of the *tra* gene from *Ceratitis capitata* (PANE *et al.* 2002; GEMPE & BEYE 2011).

A recent study found repeated duplicates of the *fem* gene in four ant and two bumblebee genomes (SCHMIEDER *et al.* 2012). The transcripts of these *fem* genes are sex-specifically spliced, suggesting a conserved sex-determining role of this gene. The function of the duplicated copies are thus far unknown (SCHMIEDER *et al.* 2012). The wasp *Nasonia vitripennis*, however, lacks a sister copy of the *fem*/*tra* gene

(VERHULST *et al.* 2010). In this study, we named the other copies of the *fem* gene *fem1*. This is because we have no functional information as to whether these genes control the complementary sex determination process as in honeybees.

The phylogenetic relationships deduced from coding nucleotide sequences (SCHMIEDER *et al.* 2012) showed that the paralogous gene pairs of *fem* and *csd/fem1* are more closely related in four ant species, the bumblebee and, as previously shown, the honeybee lineage. Figure S1 shows the sequence relationship of the genes for the neutral synonymous sites. Two recent studies (SCHMIEDER *et al.* 2012; PRIVMAN *et al.* 2013) have proposed that in contrast to a model of independent gene duplications, the most parsimonious explanation of the close relationship between the *fem* and *fem1/csd* sequences is that concerted evolution (either due to repeated unequal crossing-over or gene conversion) homogenized the duplicated copies in the different lineages (Fig. 1). One or a few ancestral duplication events gave rise to the *csd* gene and complementary sex determination observed in the Hymenoptera order (SCHMIEDER *et al.* 2012; PRIVMAN *et al.* 2013). The process of concerted evolution between the *fem* and *fem1/csd* genes repeatedly homogenized the two loci, producing the low divergence in the gene pair that we find today.

Here, we readdress the question of whether the *fem1/csd* copies repeatedly evolved through gene duplication (Fig. 1a) or whether the *fem* and *fem1/csd* gene pairs evolved through concerted evolution (Fig. 1b). The clarification of this question will provide fruitful insight into the evolution of paralogous genes and the evolution of a complementary sex determination system.

The arguments given below prompted us to further investigate this question.

1) Studies at the genome-wide scale showed that concerted evolution only affects 2% of the paralogous gene pairs (SEMPLE & WOLFE 1999), suggesting that this process rarely acts as a homogenizing force between paralogs.

2) The rate for the rise of new paralogous gene copies is 0.01 per gene per million years (LYNCH & CONERY 2000; LIPINSKI *et al.* 2011). This suggests that new duplicates of the *fem* gene can repeatedly originate in different hymenopteran lineages, which have an evolutionary history of more than 120 million years (GRIMALDI & ENGEL 2005; WILSON & HOLLDOBLER 2005; BRADY *et al.* 2006; RAMIREZ *et al.* 2010).

3) The evidence for concerted evolution between the paralogs provided thus far, namely, (i) the alternative tree topologies of the *fem* and *fem1*/*csd* nucleotide sequences (PRIVMAN *et al.* 2013) and (ii) the putative gene conversion tracts in the nucleotide sequence (SCHMIEDER *et al.* 2012; PRIVMAN *et al.* 2013), could also result from a heterogeneity in the sequence divergence, a recombination event between the *csd* alleles, methodological problems or homoplasic (convergent) nucleotide changes (HASSELMANN & BEYE 2004; HASSELMANN & BEYE 2006; HASSELMANN *et al.* 2008b; HASSELMANN *et al.* 2010).

Here, we present evidence suggesting that the paralogous gene copies *fem1*/*csd* in ants, bumblebees and honeybees evolved independently and repeatedly through a series of gene duplication events (Fig. 1a).

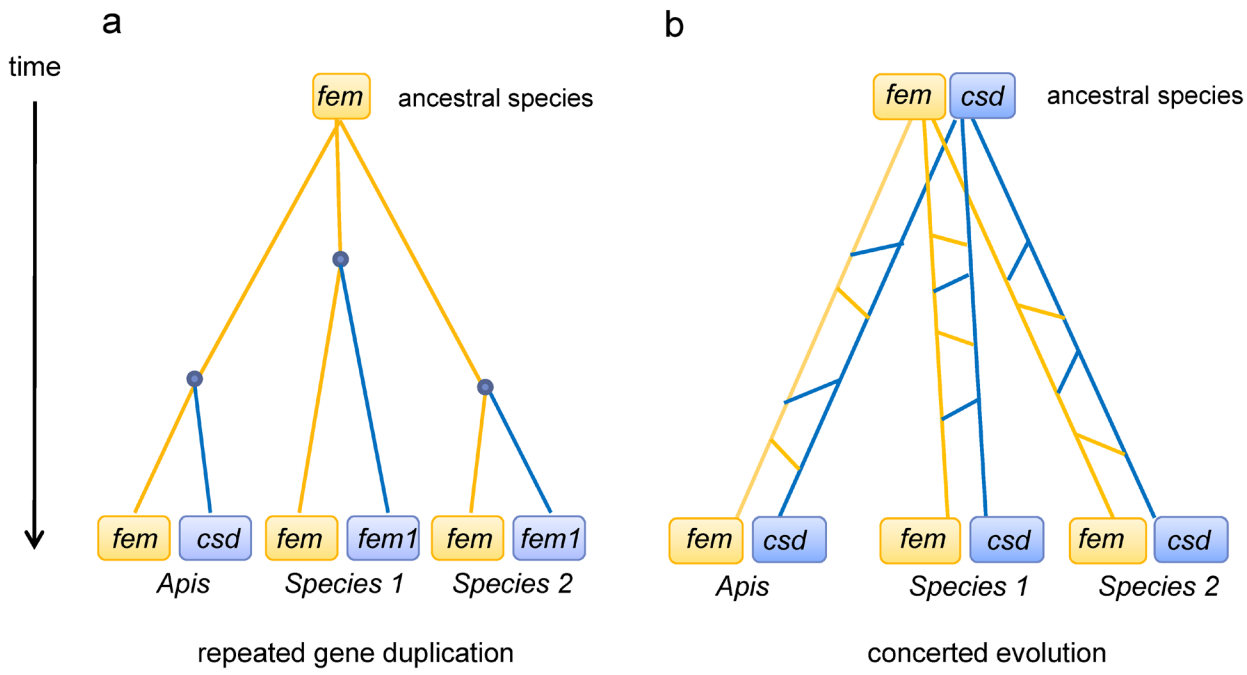

Figure 1. **Two models for the evolutionary history of *fem* paralogous genes in ants and bees:** (**a**) repeated gene duplication and (**b**) concerted evolution. Points in (**a**) denote gene duplication events giving rise to two gene copies. Connecting lines in (**b**) between branches indicate concerted evolution events resulting from unequal crossing over and/or gene conversion.

Materials and Methods

The *fem* paralogous sequences defined in this study as *fem1* (in ants and bumblebees) or *csd* (in honeybees) were taken from Schmieder et al. (SCHMIEDER *et al.* 2012) and were provided by Schmieder, S. and Poirie, M.. The genomic locus of the *fem*/*fem1* gene of *Solenopsis invicta* was not accessible in public libraries. The honeybee *fem* and *csd* sequences that we used to detect tracts of DNA transfers were kindly provided by Privman, E.. The coding sequences were aligned based on the deduced amino acids, assuming a standard genetic code table. We used either the Clustal or the Muscle program that was implemented in the MEGA5 program suite (TAMURA *et al.* 2011) to align the coding nucleotide sequences according to the deduced amino acid sequences. The alignments were edited manually. Nucleotide and amino acid substitution analyses were conducted using the MEGA5 program. Maximum likelihood fits of the substitution models with the lowest Bayesian information criterion (BIC) score were used to choose a substitution model for amino acids and nucleotides when possible. Pairwise gaps in the alignment were deleted.

Ancestral sequences

Ancestral amino acid sequences were inferred using the maximum likelihood method (NEI & KUMAR 2000) under the Jones-Taylor-Thornton (JTT) matrix-based model (JONES *et al.* 1992). The rates between sites were treated as a gamma distribution. The ancestral sequences were inferred from separate *fem* and *fem1*/*csd* sequence alignments, and Figure S2 shows the gene tree used. All informative changes used had a probability $P > 0.5$. We had to exclude some sites where the homology between sites of the Fem protein and the Csd/Fem1 alignment were ambiguous. In three cases of the bumblebee analysis, we changed the ancestral state we obtained from the MEGA analysis, because it contradicted the parsimony evolution of the sequence in the Fem and Fem1 tree (Fig. S5 and Fig. S6).

Evolutionary divergence between sequences

The d_N/d_S ratios of the interspecies comparisons between the ants, bumblebees and honeybees were inferred from sequence alignments that included either the bee or ant sequences, which greatly improved the number of identical

positions in the alignments. The evolutionary distances between the ants and bees suggested that the less degenerated synonymous sites were saturated, making the d_S estimates between the bee and ant sequences less reliable. Analyses were conducted using the Nei-Gojobori model to estimate either the nonsynonymous and synonymous substitutions per site or the absolute numbers. A χ^2-square test was used to test the ratio differences in terms of absolute numbers. We tested the equality of the evolutionary rate at the most degenerate third codon position (which is presumably largely synonymous) by using Tajima's relative rate (TAJIMA 1993), which we performed using MEGA5 software (TAMURA *et al.* 2011). We tested equality among the *Apis csd* and *fem* sequences by using the *Bombus fem* sequence as the outgroup. Similarly, we tested equality among *Bombus fem1* and *fem* sequences by using the *Apis fem* sequence as an outgroup.

RT-PCR analysis

We performed RT-PCR experiments on embryonic and larval RNA (GEMPE *et al.* 2009) to identify possible splice products of mRNA in regions of the identified pseudogenes. PCR fragments were sequenced and compared with the genomic data.

Phylogenetic relationships

The phylogenetic relationship of the *fem* and *fem1/csd* sequences was determined based on presumably neutral synonymous differences by excluding the *tra* sequences of *Ceratitis capitata* and *Nasonia vitripennis* (because d_S could not be estimated for these species) using the neighbor-joining method. The confidence probability (multiplied by 100) that the interior branch length is greater than 0 was estimated using the bootstrap test with 10,000 replicates. The evolutionary distances were computed using the Pamilo-Bianchi-Li method and are displayed in the units of the number of synonymous substitutions per synonymous site. For each sequence pair, all ambiguous positions were removed.

Sequence tracts of DNA transfers

Tracts of DNA transfers in the honeybee *fem* and *csd* sequences were identified using the RDP 3.44 software program (MARTIN *et al.* 2010). We used the following

tests: RDP (MARTIN & RYBICKI 2000) with internal and external references, GENECONV (PADIDAM *et al.* 1999), BootScan (MARTIN *et al.* 2005), MaxChi (SMITH 1992), Chimaera (POSADA & CRANDALL 2001), SiScan (GIBBS *et al.* 2000) and 3Seq (BONI *et al.* 2007). The methods implemented in the RDP 3.44 software program relied on the identification of recombinant sequences and the parental sequences from which these recombinant sequences were derived, which is facilitated by having a large set of sequences (MARTIN *et al.* 2010). We thus generated a single alignment. Our alignment included 36 *A. mellifera*, 16 *A. cerana* and 19 *A. dorsata csd* coding nucleotide sequences, the single f*em* nucleotide sequence from each *Apis* species, and the *fem* nucleotide sequence of *B. terrestris* as an outgroup reference sequence. We used the same program settings as described in a previous study (PRIVMAN *et al.* 2013).We removed sequence GenBank accession #: AY352276 from the analysis, as this is a chimeric sequence of the *fem* and the *csd* gene (the entry has now been deleted). We also updated the sequence (GenBank accession #: AY350616) because it is not a *csd* but a *fem*-derived sequence. To evaluate the effect of the number of sequences in the alignment, we generated 20 sequence alignments consisting of three sequences. Each alignment included two randomly chosen *csd* sequences from a single *Apis* species as well as the *fem* sequence from *B. terrestris.* We classified the detected transfer events as (i) concerted evolution events if they occurred between paralogous genes, (ii) recombination events if they occurred between the same gene and (iii) falsely discovered events (FDE) if they were biologically implausible. Such biologically implausible events are events in which the sequence transfer between orthologous genes should give rise to a paralogous recombinant sequence or events in which the sequence of the outgroup reference species was involved. We confirmed falsely detection by inferring the clustering of the sequence tracts into the gene cluster in phylogenetic tree analyses.

Results

Amino acid changes in the MRCA ancestral sequences of bees and ants are shared between the Fem and Csd/Fem1 proteins

To find further support for either the repeated gene duplication model or the concerted evolution model, we followed the evolutionary trajectory of substitutions that led to amino acid changes in the ancestral sequences of the most recent common ancestor (MRCA) of bees and ants (Fig. 2a). This evolutionary window predates the timing of the different gene duplication events under the repeated duplication model and can therefore provide unique information about the evolutionary history of the sister copies. Under the concerted evolution model, we would expect to find unique substitutions in the ancestral sequences of the MRCA of ants and bees, which are confined to the *fem* or the *csd/fem1* gene (Fig. 2a). This pattern would arise because the two sister copies originated only once in Hymenoptera (SCHMIEDER *et al.* 2012)and accumulated substitutions separately due to their separate evolutionary history, which predates the MRCA of bees and ants (SCHMIEDER *et al.* 2012). Concerted evolution, the exchange of sequences between evolutionary old paralogous genes, would partly homogenize the sister copy genes, which would thus appear as to have more recent common ancestry in the phylogenetic tree (Fig. S1, S2). Under the repeated gene duplication model, the ancestral sequence in the evolutionary time window that predates the different duplication events should be the same for the *fem* and *csd*/*fem1* genes because at this time point, only a single copy of the gene existed (Fig. 2a).

We generated separate phylogenetic trees using the amino acid sequence of the Fem and Fem1/Csd proteins, which allowed us to trace the putative separate evolutionary history of these sister copy genes. We inferred the ancestral amino acid sequences of the MRCAs of bees and of ants (Fig. 2, S2) using the maximum likelihood method (NEI & KUMAR 2000). These nodes had high statistical support and represented evolutionary time windows before the putative repeated gene duplication events. For the large evolutionary distances between ants and bees, we analyzed amino acid changes instead of synonymous substitutions, which were saturated, at least for the less degenerate sites. We identified changes in the Fem protein of the MRCA amino acid sequence of bees and ants by comparing the MRCA sequences of bees and of ants. We found 7 changes in the MRCA sequence of bees and 9 in the

MRCA sequence of ants. In ants, we found the same 9 amino acid changes in the sister copy sequence of the paralogous Fem1 protein (Fig. 2b, Fig. S3). In bees, we found the same 7 amino acid changes in the MRCA sequence of the paralogous Csd/Fem1 protein (Fig. 2c, Fig. S4). These informative changes in the ant and bee sequences are found in different parts of the protein, suggesting that we obtained information that covered the entire protein. Our study found no amino acid changes that were confined to only one of the sister copies, which would indicate a deeper ancestry of the gene duplication that predated the MRCA of bees and of ants. Therefore, the sequences harbor no information that can provide evidence for a separate history of the *fem* and *csd/fem1* genes that predates those of the ant and bee lineages, or for a single gene duplication event in Hymenoptera.

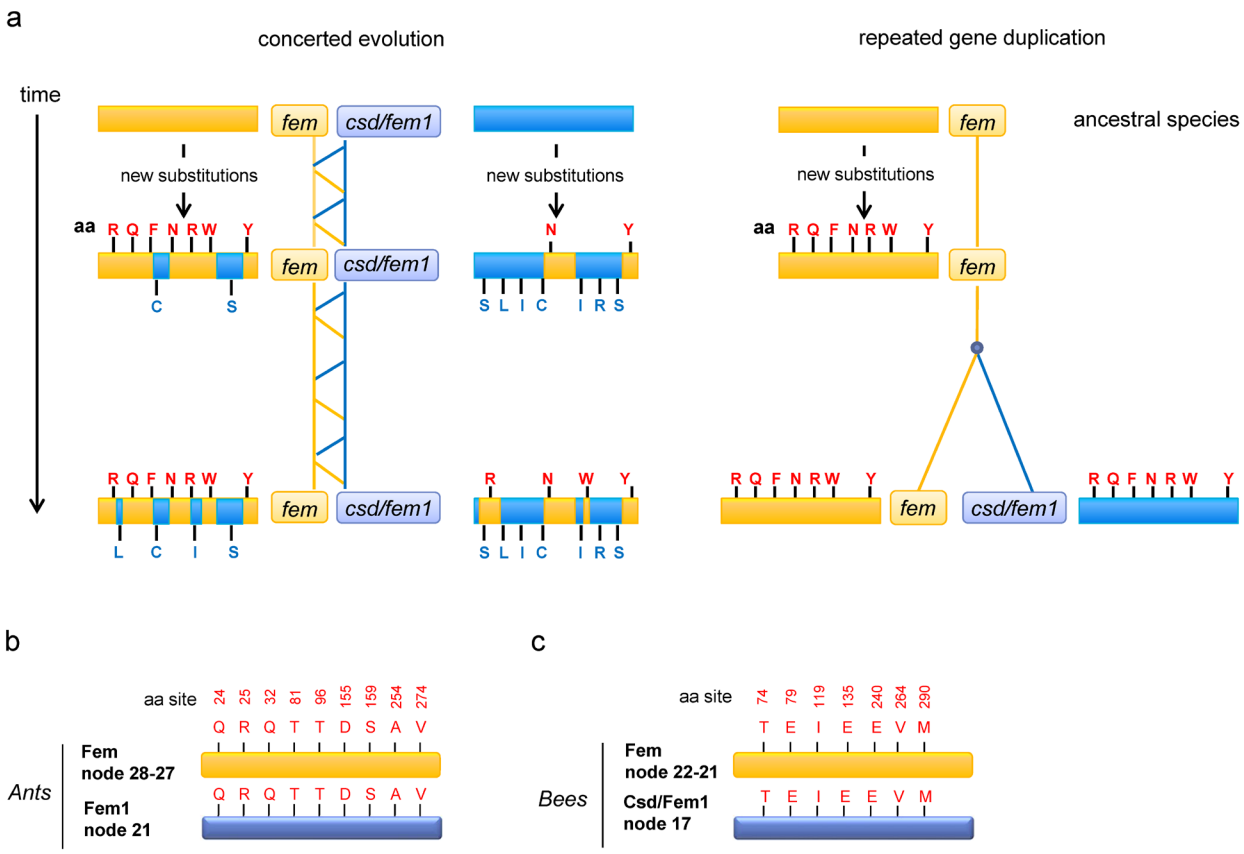

Figure 2. The evolutionary fate of *fem* gene substitutions in an evolutionary window predating the putative repeated gene duplications. (a) The expected evolutionary fate of *fem* substitutions in the paralogous genes *fem* and *csd/fem1* under the model of concerted evolution and repeated gene duplication. (**b, c**) The letters above the yellow boxes show the inferred amino acid changes in the Fem protein tree that evolved during the evolutionary window of the MRCA of ants and bees and the MRCAs of ants (**b**) and of bees (**c**). Letters above the blue boxes indicate the amino acid residues that are found at the homologous sites in the ancestral Csd/Fem1 protein sequence of the MRCA of ants (**b**) and of bees (**c**). Numbers above the letters designate the homologous sites in the Fem amino acid sequence alignment. Numbers before the boxes indicate the nodes (Fig. S2) used to infer the ancestral sequence information. aa denotes amino acid.

No evidence of concerted evolution is found in bumblebee sequences

We studied whether we can detect sequence tracts that would indicate a transfer of sequences between paralogous genes (concerted evolution) in the bumblebee lineage. The phylogenetic clustering of the *fem* and *fem1* sequences in the bumblebee lineage (Fig. S1) suggests that concerted evolutionary events should also occur in the bumblebee lineage. We used seven methods (RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan and 3Seq) designed to detect tracts of recombination events in the nucleotide sequences, which are included in the RDP 3.44 software package (MARTIN *et al.* 2010). We applied this method to a single sequence alignment, which included the *fem* and *fem*1 sequences of the two *Bombus* species and the *fem* sequence from *A. mellifera* as an outgroup reference. We classified the detected transfer events as either (i) concerted evolution events if they occurred between paralogous genes, (ii) recombination events if they occurred between the same gene, and (iii) falsely discovered events (FDE) if these events are biologically implausible. Such biologically implausible events are events in which sequence transfer between orthologous genes should give rise to a paralogous recombinant sequence or are events in which the sequence of the outgroup reference species is involved. The results of the analysis are shown in Figure 3.

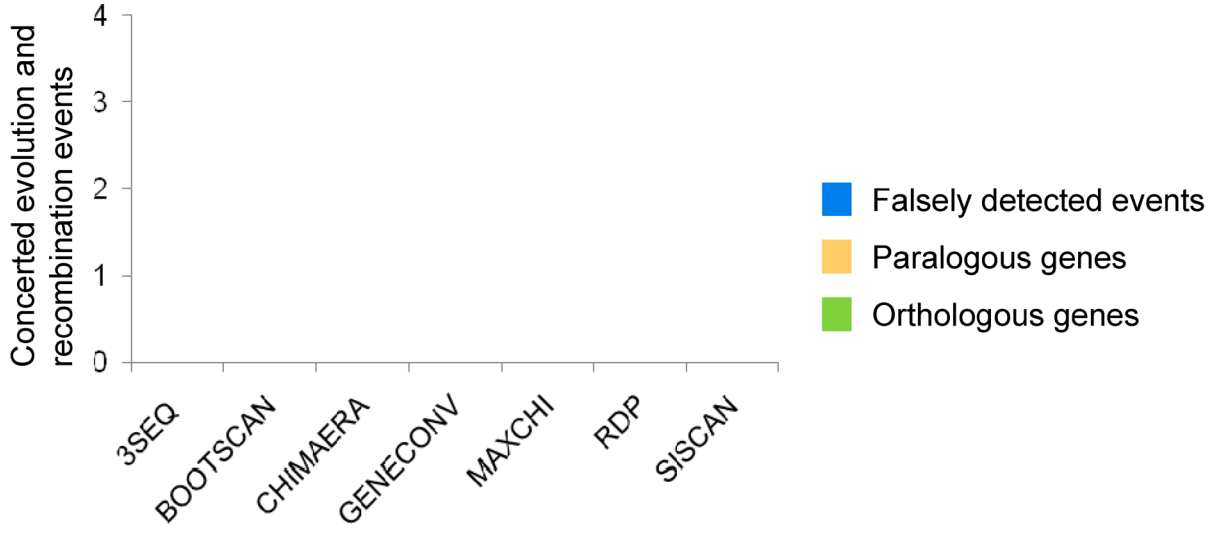

Figure 3. Number of gene conversion and recombination events in *B. terrestris* and *B. impatiens* sequences. Tracts of putatively recombined sequence were detected by the 7 methods as shown on the *x*-axis and using RDP 3.44 software program. The analysis was run on a single alignment of the *fem* and *fem1* sequences of the *B. terrestris*, *B. impatiens* and *A. mellifera fem* sequence. Gene conversion events refer to DNA transfers between the paralogous genes *fem* and *fem*1. Recombination events indicate transfer events between sequences of the same gene. Falsely detected events (FDE) refer to biologically implausible events (see Materials and Methods).

We found no sequence tracts that were transferred between paralogous genes, suggesting that concerted evolution played no role in the evolution of the paralogous genes of the *Bombus* lineage.

We next studied whether we can identify signatures of concerted evolution by confining our analysis to single amino acid substitutions. We tested whether some *fem* or *fem1* substitutions that newly evolved in each bumblebee species were transferred to its paralogous sister gene (Fig. 4a). Such shared evolved states between the two paralogous genes within each species would indicate a transfer of the corresponding nucleotide sequence by concerted evolution. We determined the ancestral sequences of the Fem and Fem1 protein of the MRCA of *Bombus terrestris* and *Bombus impatiens*, identified the evolutionary changes and studied whether these changes were also present in the paralogous sister copy (Fig. 4a). We found 4 newly evolved amino acid changes in the Fem protein of *B. impatiens* and 2 in that of *B. terrestris.* All of these newly evolved changes were not present in the sister Fem1 proteins (Fig. 4b, Fig. S5). We detected 8 newly evolved amino acid changes in the Fem1 protein of *B. impatiens* and 11 in *B. terrestris,* and these newly evolved changes were not present in the sister Fem protein (Fig. 4b, Fig. S6). Taken together, we found that all 12 newly evolved amino acid changes in *B. impatiens* and all 13 newly evolved amino acid changes in *B. terrestris* were absent in their corresponding sister copies. This survey covered different parts of the approximately 400 amino acid (aa)-long protein, providing evidence that concerted evolution events were absent in the *Bombus fem* and *fem1* sequences.

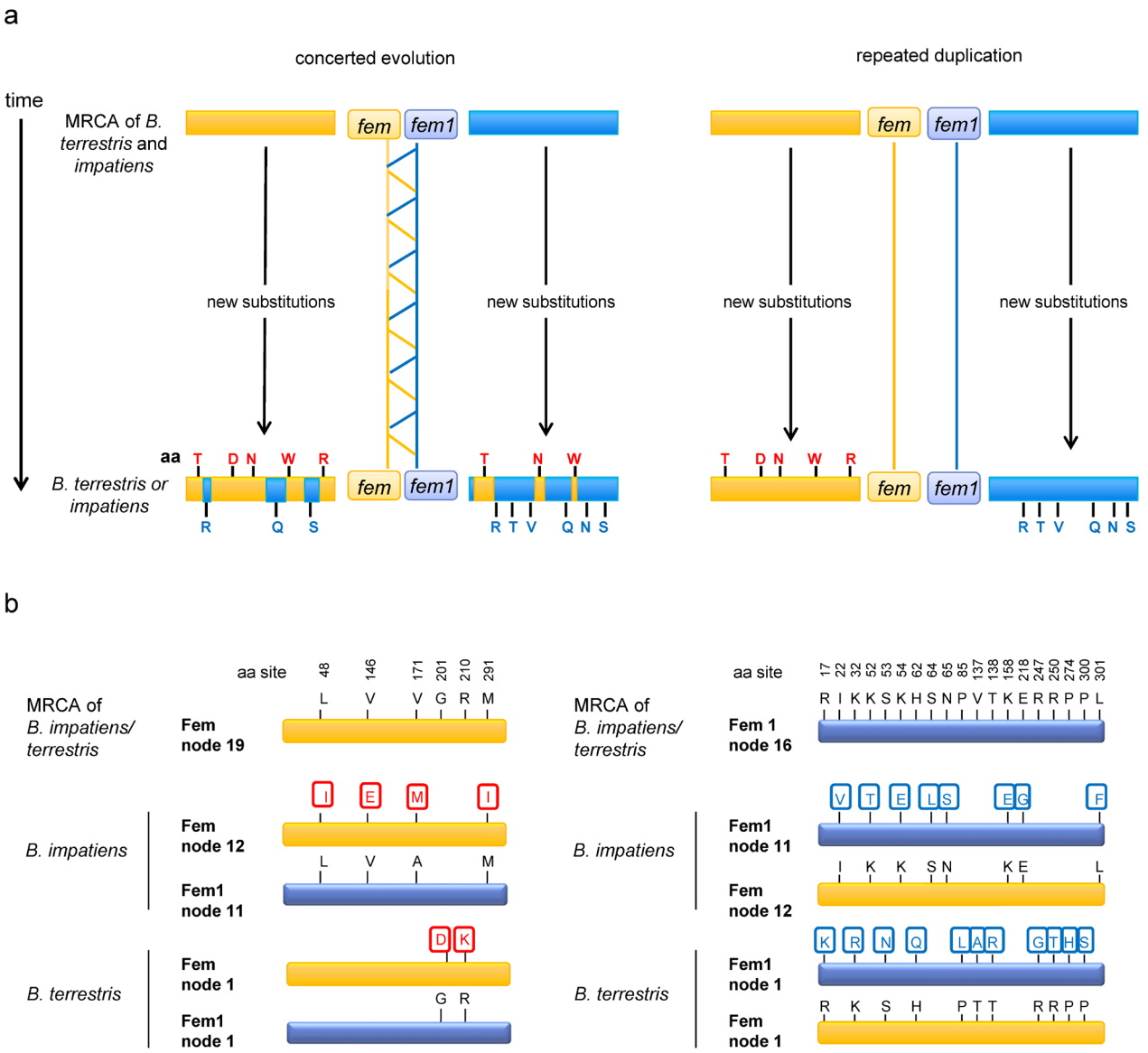

Figure 4. The evolutionary trajectory of *fem* gene substitutions in the evolutionary window that follows the putative gene duplication event in the *Bombus* lineage. (a) The expected evolutionary fate of *fem* substitutions in the paralogous genes *fem* and *fem1* under the models of concerted evolution and repeated gene duplication. (**b**) Deduced amino acid changes. The yellow box indicates the Fem protein, and the blue box indicates the Fem1 protein. Black letters above the boxes indicate the ancestral state of the amino acid residues found in the MRCA of *B. terrestris* and *B. impatiens*. Red letters in the red frame indicate the amino acid residues that evolved since the MRCA of *B. terrestris* and *B. impatiens* in the Fem protein. Blue letters in the blue frame indicate the amino acid residues that evolved since the MRCA of *B. terrestris* and *B. impatiens* in the Fem1 protein. Numbers before the boxes indicate the nodes (Fig. S2) used to infer the ancestral sequence information. aa denotes amino acid.

No evidence of concerted evolution in sequences of the honeybee lineage

Next we applied the RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan and 3Seq methods to identify sequence tracts of concerted evolution between the paralogous *fem* and *csd* nucleotide sequences of the honeybee. We included the same coding nucleotide sequences as in a previous study (PRIVMAN *et al.* 2013), comprised of 36 *csd* and 1 *fem A. mellifera*, 16 *csd* and 1 *fem Apis cerana csd*, 19 *csd* and 1 *fem Apis dorsata* sequence and a *fem B. terrestris* sequence as an outgroup reference.

In this study, we applied the methods to a single sequence alignment that included all *fem* and *csd* sequences. In the previous study from Privman et al., 100 alignments were used, each consisting of the *fem* sequences and a randomly chosen *csd* allele sequence from each *Apis* species and a *fem* sequence from *B. terrestris*. The rationale behind our altered experimental design was that the detection methods used in the RDP 3.44 software program are designed for large datasets, to identify the recombinant sequence and the two sequences from which the recombinant sequence was derived (MARTIN *et al.* 2010). We removed a sequence (GenBank accession #: AY352276) from the analysis because it was a chimeric sequence of the *csd* and *fem* gene. This sequence resulted from a misassembly of cDNA sequences derived from the *fem* and *csd* gene at a point in time when we had no knowledge about a second gene in the genome and the nature of allelic variation (BEYE *et al.* 2003). This sequence entry has now been removed from GenBank. We also updated the sequence (GenBank accession #: AY350616), which is a *fem* and not a *csd* sequence. The results of this sequence analysis are shown in (Fig. 5).

We found no tracts of gene conversion (Fig. 5a), suggesting that a DNA transfer between the *fem* and *csd* gene did not occur in the honeybee lineage. We observed 3 events in which recombination between the *Apis csd* ortholog sequences gave rise to a paralogous *Apis fem* sequence, suggesting falsely detected events (FDE) that were identified by the program. We confirmed this falsely detection by demonstrating that the putatively transferred fragment is indeed *fem* derived which we showed by the clustering into the *fem* gene cluster in the phylogenetic tree analysis. However, we detected multiple recombination events between the *csd* sequences (alleles) derived from the same and different *Apis* species (Fig. 5b). This

suggests that recombination is a regular process between alleles of the *csd* gene, consistent with previous reports (HASSELMANN & BEYE 2006; HASSELMANN *et al.* 2008a).

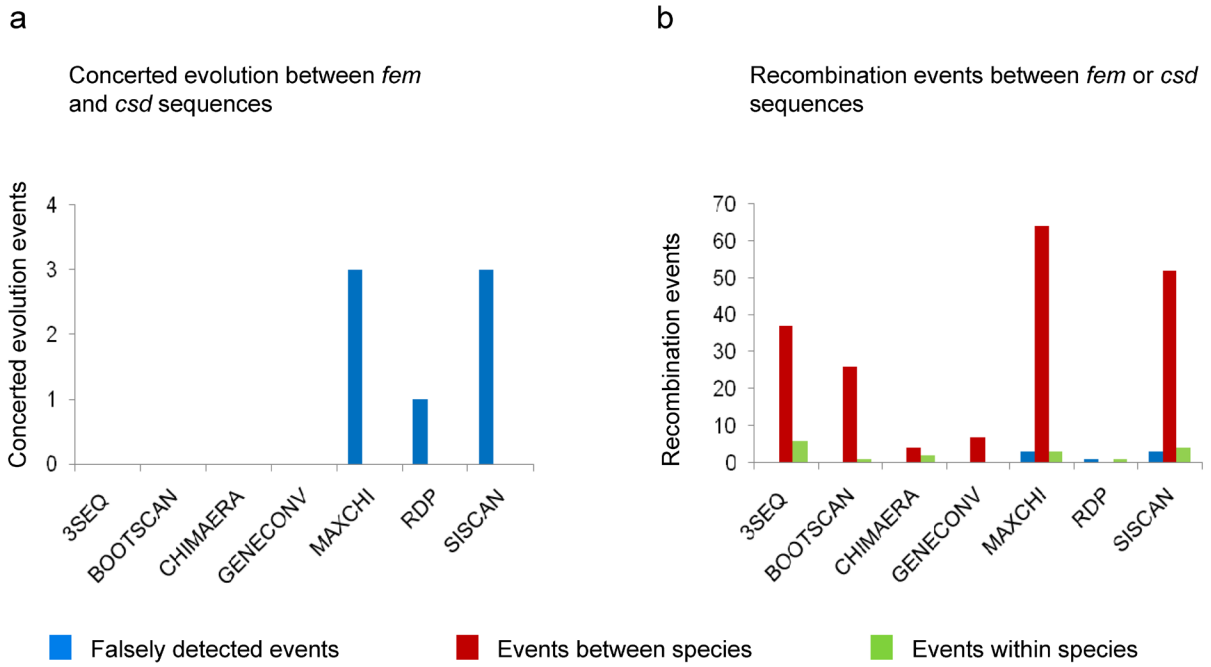

Figure 5. Events of gene conversion and recombination in the *fem* and *csd* sequences of *A. mellifera*, *A. dorsata* and *A. cerana*. Tracts of putatively recombined sequences were detected by seven different methods as indicated on the *x*-axis using the RDP 3.44 software program. The analysis was run on a single alignment of 71 *csd* and 4 *fem* sequences. (a) The number of concerted evolution events that refer to the DNA transfer between paralogous genes *fem* and *csd*. (b) The number of recombination events that identify events between sequences of the same gene. Falsely detected events (FDE) in (a) and (b) refer to biologically implausible events. The outgroup reference sequence, *B. terrestris fem,* was never involved in one of the detected events.

We next evaluated how the number of sequences affects the false detection of events. We generated 20 sequence alignments, each consisting of two randomly chosen *csd* sequences from a single *Apis* species and one *fem* sequence from *B. terrestris,* which served as the outgroup reference sequence for the alignment above. For these alignments, the GENECONV method detected 8, the MaxChi method detected 12, and the Chimaera method detected 9 tracts of sequences in which DNA was putatively transferred between the *csd* sequences of the *Apis* and *fem* gene of the *B. terrestris* sequences (Fig. S7). However, these events are biologically implausible, because the nucleotide differences we observe today in the *csd* alleles evolved after the split into different *Apis* species (HASSELMANN *et al.* 2008a). Hence, the *Bombus* sequence cannot have contributed through concerted evolutionary

events to the *csd* polymorphism. We confirmed this falsely detection in a sample of detected events by demonstrating that the putatively transferred fragment is not derived from the *Bombus* sequence, which we showed by the clustering into the *csd* gene cluster in the phylogenetic tree analysis. We detected no such transfer between the *Apis csd* alleles and the *B. terrestris fem* sequences if all sequences are included in a single alignment (Fig. 5) suggesting that having fewer sequences in an alignment increases the rate of falsely detecting DNA transfer between paralogs.

We did not perform single amino acid substitution analysis in the honeybee as we did for *Bombus* because we have not robustly identified enough newly evolved sites.

d_N/d_S ratio differences between paralogous genes suggest a directional DNA transfer process

We consistently observed that the d_N/d_S ratios (nonsynonymous (d_N) to synonymous (d_S) per site substitutions) of the *fem1*/*csd* sequence pairs were higher compared to those of the *fem* sequence pairs ((Table 1); χ^2-test,$P< 0.05$ for all comparisons), suggesting that selection operates differently on the *fem* and the *fem1/csd* genes. The ratio of the differences is most pronounced for the *Apis* sequences (d_N/d_{Sfem}=0.1-0.2 versus d_N/d_{Scsd}=0.8-1), consistent with previous findings (HASSELMANN *et al.* 2008b; HASSELMANN *et al.* 2010), and for the bumblebee sequences (d_N/d_{Sfem} = 0.16 versus d_N/d_{Scsd}= 0.56). The difference in the d_N/d_S ratios is less pronounced in the ants.

We further evaluated how differences in the d_N/d_S ratios are compatible with the mutual transfer of DNA and concerted evolution. The mutual transfer between the paralogous genes would also transfer the differences in the d_N/d_S ratios between paralogous genes. We assume that the entire sequences are in equilibrium of homogenization through concerted evolution and divergence. This is consistent with the model that concerted evolution is a random mutational and ongoing process that occurs through hymenopteran phylogeny. At this equilibrium, the gene-wide d_N/d_S values are good approximations for DNA fragments that are, on average, transferred between paralogs.

First, we showed that there are not different rates at synonymous sites in the honeybee and the bumblebee clade (Tajima's relative rate test, P > 0.05). This result

suggests that the differences in the d_N/d_S ratios between the genes reflect substitution rate differences at nonsynonymous sites (d_N).

DNA fragments transferred from the *fem1*/*csd* can only reach the lower d_N/d_S ratios in the *fem* gene as evolutionary time progresses if new mutations occur at the neutral synonymous sites along with purifying selection at the nonsynonymous sites. We approximated the mean number of neutral pairwise substitutions per site (d_S) that is required to reach the lower d_N/d_S ratios. The *csd* sequences in honeybees show an average ratio of $d_N/d_{S\,csd}$ = 0.9 (Table 1), suggesting that during the separation time of the two paralogous genes (in terms of $d_{Scsd/fem}$= 0.18, Table S1), approximately $d_{N\,csd}$ = 0.16 pairwise substitutions in the *csd* gene have accumulated. We next estimated $d_{S,}$ which has accumulated so that a transferred *csd* fragment (that has on average a ratio $d_N/d_{S\,csd}$ = 0.9) can reach the observed d_N/d_S ratio of the *fem* gene (in which $d_N/d_{S\,fem}$ = 0.17). We assume the most conservative model, in which all newly arising nonsynonymous mutations were removed by purifying selection and only new synonymous mutations became fixed. A *csd* fragment can only adjust for the *fem*'s d_N/d_S ratio if $d_{S\,x}$ = 0.77 additional synonymous substitutions have accumulated. This result suggests that a transferred *csd* DNA requires on average d_S = 0.77 synonymous substitutions to observe the low d_N/d_S ratio of the *fem* gene.

Similarly, we approximated $d_{S\ x}$ for bumblebee sequences. *fem1*-derived sequences in the *fem* sequence ($d_{S\ fem1/fem}$ =0.22, $d_N/d_{S\ fem1}$ = 0.56, $d_{N\ fem1}$ = 0.12, $d_N/d_{S\ fem}$ = 0.16) would require $d_{S\ x}$ = 0.53 pairwise synonymous differences to accumulate in order for the sequence to reach the same $d_N/d_{S\ fem}$ ratio.

$d_{S\ x}$ = 0.77 in honeybees and $d_{S\ x}$ = 0.53 in bumblebees, as required for the transferred *csd* DNA to reach the observed d_N/d_S ratios of the *fem* gene are largely exceeding d_S = 0.39 that have accumulated between the bumblebee and honeybee species. This suggests that these transfers should predate the MRCA of bees, which is inconsistent with our previous result that there would be an absence of such transfers in this evolutionary window (Fig. 2). In addition, $d_{S\ x}$ largely exceeds the divergence between paralogous genes in the bumblebee ($d_{S\ fem1/fem}$ =0.22) and the honeybee ($d_{S\ fem/csd}$ = 0.18), suggesting that the divergence between paralogs is too low to be compatible with such transfers and ratios. The results of this simple transfer model imply that the transfer processes cannot be bidirectional between paralogous genes, as predicted under a model of concerted evolution in which sequences are mutual exchanged and can become fixed through positive selection and genetic drift.

Only a directional transfer from *fem* to *csd/fem1* would be compatible with data suggesting that relaxed or positive selection could substantially increase the d_N/d_S ratio. This directional transfer process is consistent with the gene duplication model, in which a new gene copy becomes neofunctionalized (HASSELMANN *et al.* 2008b).

Table 1. The d_N and d_S values and ratios for the interspecies comparisons of the *fem* and *fem1/csd* genes.

Clade	Species		*fem* gene			*fem1/csd* gene			$d_N/d_{Sfem} < d_N/d_{Sfem\ 1/csd}$ *
			dN (SE)	dS (SE)	dN/dS	dN (SE)	dS (SE)	dN/dS	
Apis	*Ador*	*Amel*	0.02 (0.004)	0.09 (0.02)	0.22	0.13 (0.01)	0.13 (0.02)	1.00	*P*< 0.0001
	Amel	*Acer*	0.01 (0.003)	0.08 (0.02)	0.13	0.12 (0.01)	0.15 (0.02)	0.8	*P*< 0.0001
Bombus	*Bter*	*Bimp*	0.008 (0.003)	0.051 (0.013)	0.16	0.05 (0.007)	0.09 (0.02)	0.56	*P*< 0.01
Ants	*Cflo*	*Hsal*	0.26 (0.01)	0.51 (0.029)	0.51	0.35 (0.02)	0.5 (0.03)	0.7	*P*< 0.01
	Cflo	*Pbar*	0.19 (0.01)	0.44 (0.03)	0.43	0.3 (0.01)	0.43 (0.03)	0.7	*P*<0.01
	Hsal	*Pbar*	0.24 (0.01)	0.53 (0.03)	0.45	0.26 (0.01)	0.46 (0.03)	0.57	*P*< 0.02
	Acep	*Pbar*	0.11 (0.01)	0.37 (0.03)	0.30	0.16 (0.01)	0.41 (0.03)	0.39	*P*< 0.05
	Cflo	*Acep*	0.21 (0.01)	0.46 (0.03)	0.46	0.32 (0.01)	0.49 (0.03)	0.65	*P*< 0.01
	Acep	*Hsal*	0.23 (0.01)	0.5 (0.03)	0.46	0.28 (0.01)	0.44 (0.03)	0.64	P < 0.01

Species names: *Amel*, *A. mellifera*; *Ador, A. dorsata; Acer, A. cerana*; *Bimp, B. impatiens*; *Bter*, *B. terrestris*; *Hsal*, *H. saltator*; *Pbar*, *P. barbatus*; *Acep*, *Atta cephalotes*; *Cflo*, *Camponotus floridanus*. SE: standard error.
* A one-tailed χ^2 - test was conducted using the absolute number of synonymous and nonsynonymous differences.

Additional sequence copies are repeatedly found at other genomic loci in bees and ants

To find further evidence for the repeated duplication model, we searched for other gene duplication tracts of the *fem* and *fem1/csd* genes. Using BlastN searches, we found genomic sequences with a high similarity to the coding nucleotide sequences of the *fem* or *csd* genes in the western honeybee, two bumblebee species and two ant species at other genomic loci (Fig. 6a). These duplicate genes exist as inactive genes with no complete open reading frame (ORF), suggesting that they are pseudogenes (*ps*). For the *ps1 csd* gene of *A. mellifera*, we confirmed by RT-PCR that this gene is transcriptionally inactive. We estimated the synonymous pairwise divergence (d_S) of the pseudogenes and of the *fem* or *fem1/csd* genes within each species (Fig. 6b), which we related to the MRCA events by providing the *fem* d_S values of different species. For *A. mellifera*, we found two pseudogenes, one derived from the *csd* gene and one derived from the *fem* gene (Fig. 6). The latter of these gave rise to a new female intron sequence in the *fem* gene (GEMPE *et al.* 2009). The d_S divergence between pseudogenes and the *fem* or *csd* gene is smaller compared to the d_S between the *csd* and *fem* and the d_S between the *fem* of *A. mellifera* and *A. dorsata* (Fig. 6b), suggesting that both sequences were duplicated recently in the *A. mellifera* lineage. The d_S divergence in the bumblebee genomes suggests that the origin of these duplications predates the split between the *B. terrestris* and *B. impatiens* species (Fig. 6b) but that it originated after the split of the current functional gene copies of *fem* and *fem1*. In the ant *Harpegnathos saltator*, we observed one pseudogene that originated after and one that arose during the split between the functional *fem* and *fem1/csd* gene copies. However, in the ant *Pogonomyrmex barbatus*, we also found a pseudogene with a much deeper ancestry (*Pbarps1 fem*; d_S= 0.27) than the functional gene pairs. Our results demonstrate that other gene duplication events occurred throughout the phylogeny and even within the *A. mellifera* lineage.

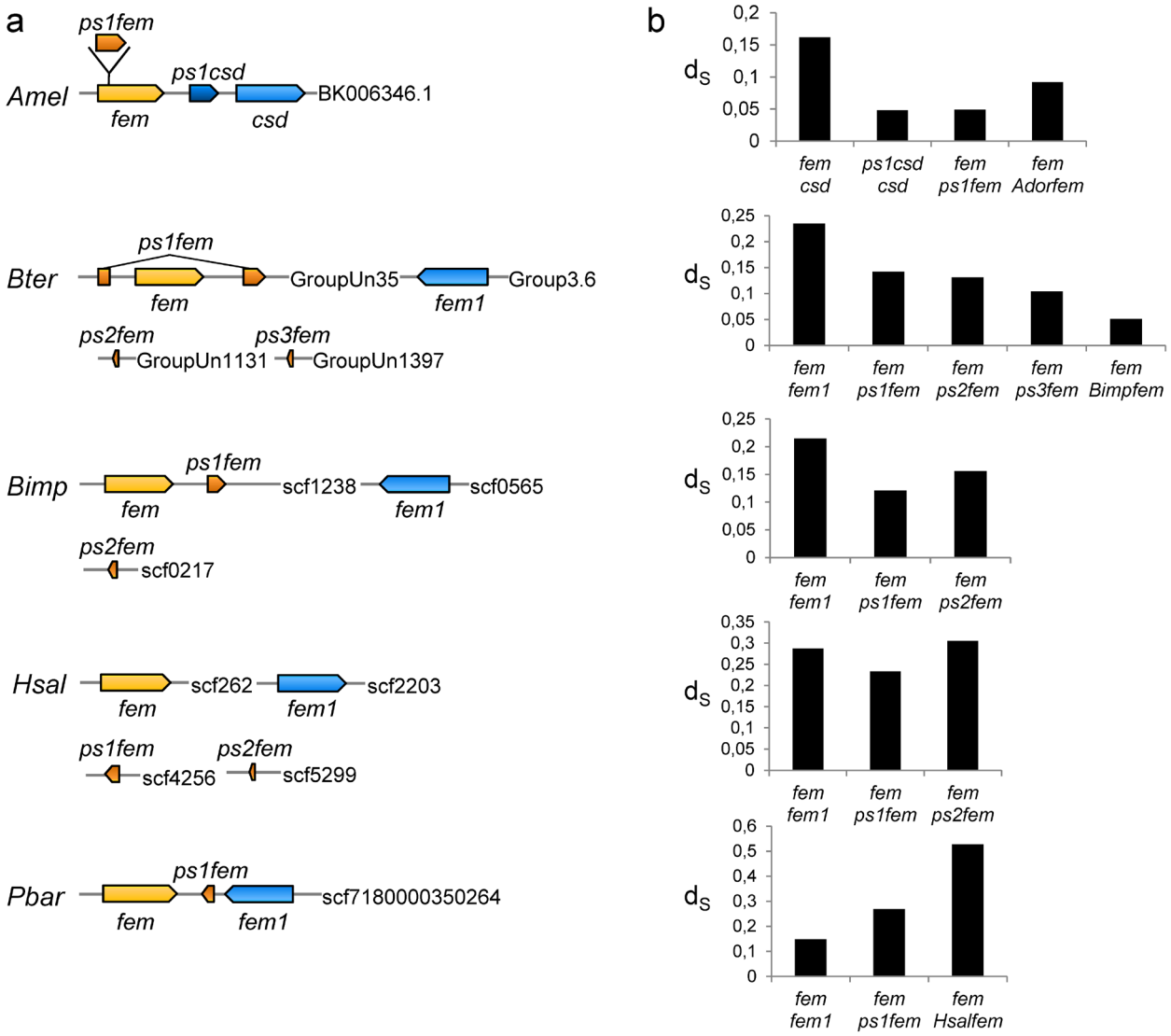

Figure 6. Pseudogenes (*ps*) of the *fem* and *csd* genes in the ant, bumblebee and western honeybee genomes. (a) The orientation and location of the pseudogenes (*psfem*, *pscsd*). Boxes denote the genes or pseudogenes. The box length of pseudogenes indicates the relative degree of homology to the coding nucleotide sequences of the *fem* or *csd* genes. The phylogenetic relationship assignments are based on the lowest d_S estimates or the ancestral state. Numbers behind the bars indicate the genomic scaffold, linkage group or the GenBank accession number. **(b)** Evolutionary distance between duplicated *fem* and *fem1/csd* gene copies are presented in terms of pairwise synonymous divergence per synonymous site (d_S). Abbreviations*: Amel*, *A. mellifera*; *Bimp, B. impatiens*; *Bter*, *B. terrestris*; *Hsal*, *H. saltator*; *Pbar*, *P. barbatus.*

Discussion

Our study presents several lines of evidence that support the repeated gene duplication model, but reject the concerted evolution model in which the low divergence of paralogs resulted from homogenization. We studied *fem* and *csd*/*fem1* paralogous genes in several bee and ant species, representing 120 million years of evolution (GRIMALDI & ENGEL 2005; WILSON & HOLLDOBLER 2005; BRADY *et al.* 2006; RAMIREZ *et al.* 2010). We first showed that there were no unique changes in the Fem or Csd/Fem1 proteins at a point in time that would indicate a separate history of the two gene copies, predating the MRCAs of bees and of ants. We detected no sequence tracts that would indicate a DNA transfer between the paralogs in two bumblebee and three honeybee species by using different methods. We also identified other tracts of duplicated copies of the *fem* and *fem1*/*csd* gene at other genomic loci in different ant and bee species, suggesting that repeated gene duplication is a frequent process in the evolution of these genes. Finally, we showed that the major differences in the d_N/d_S ratio between the *fem* and *fem1*/*csd* genes in bees exclude a mutual transfer of sequences but suggest a directional transfer from *fem* to the *fem1*/*csd* gene, which is consistent with gene duplication and a neofunctionalization model (HASSELMANN *et al.* 2008b) and not with a mutual exchange of sequences under concerted evolution. We conclude from these results that the *fem1*/*csd* genes repeatedly originated through gene duplication in the bumblebee, honeybee and ant lineages. Concerted evolution played no detectable role in the evolution of these genes, suggesting that the phylogenetic relationship of the paralogs is represented by a gene tree based on neutral synonymous sites (Fig. S1).

Our finding is consistent with frequency estimates of gene duplication and concerted evolution events. Previous studies estimated that a gene will, on average, duplicate every 100 million years (LYNCH & CONERY 2000; LIPINSKI *et al.* 2011), which is consistent with our finding of repeated gene duplications of the *fem* gene in the phylogeny of ants and bees which split approximately 120 million years ago (GRIMALDI & ENGEL 2005; WILSON & HOLLDOBLER 2005; BRADY *et al.* 2006; RAMIREZ *et al.* 2010). Another study showed that gene conversion is a rare event, detectable in only 2% of duplicated genes, and that this process requires physical distances smaller than 9 kb (SEMPLE & WOLFE 1999). Contrary to the latter requirement, *fem*

and *csd* gene are separated by more than 12 kb in the honeybee (*A. mellifera*). Studies of genes that have multiple copies in the genome have demonstrated that new copies constantly originate by gene duplication (OTA & NEI 1994; OTA & NEI 1995; NEI *et al.* 1997; SU & NEI 2001; PIONTKIVSKA *et al.* 2002; ROONEY *et al.* 2002; NIIMURA & NEI 2003; EIRIN-LOPEZ *et al.* 2004; NEI & ROONEY 2005; NIIMURA & NEI 2005; ROBERTSON & WANNER 2006; ROBERTSON *et al.* 2010). Some of the duplicated copies are maintained in the genome for an extended period of time, while other copies were deleted or became nonfunctional through the accumulation of deleterious mutations.

Previous studies (SCHMIEDER *et al.* 2012; PRIVMAN *et al.* 2013) proposed that concerted evolution produced the low divergence of the *fem* and *fem1*/*csd* genes. The authors suggested that (i) alternative tree topologies of the nucleotide sequences (PRIVMAN *et al.* 2013) and (ii) putative gene conversion tracts (SCHMIEDER *et al.* 2012; PRIVMAN *et al.* 2013) are evidence for concerted evolution. Schmieder et al. identified gene conversion tracts in the genomic sequences of the paralogs. Our results showed that *de novo* duplications of the *fem* and *fem1*/*csd* genes generated such tracts. Privman et al. counted more than 100 recombination events in the honeybee *fem* and *csd* sequences, which they take as evidence for concerted evolution. In our reanalysis of the same sequences using the same methods (Fig. 5), we distinguished whether the transfer of DNA occurred between paralogous genes (concerted evolution) or between alleles of the same (orthologous) gene (recombination events). We found no tracts of gene conversion events between the *fem* and *csd* sequences (Fig. 5a), suggesting the absence of concerted evolution in honeybees. However, we found repeated transfers between alleles of the same (*csd*) gene (Fig. 5b), a finding which has been repeatedly reported (HASSELMANN & BEYE 2006; HASSELMANN *et al.* 2008a).

In our reanalysis, we used the entire 75-sequence data set in a single alignment, in contrast to the Privman et al. study, which used 100 alignments of 7 *fem* and *csd* sequences with different sets of *csd* alleles chosen from each species. We inspected some of the results using the Privman et al. alignments and repeatedly found tracts that suggested recombination events between the *csd* alleles. We also demonstrated in this study that, as in the work of Privman et al., small alignments of only a few sequences (Fig. S7) can generate an increase of the number of falsely detected events.

Privman et al. also proposed that the differences in the phylogenetic relationships of recombinant and non-recombinant regions provide further evidence for concerted evolution. Because these falsely detected “recombinant” sequence tracts are sequences from the same gene (*csd*) (see Fig. 5) these inconsistencies in the phylogenetic relationships are no further evidence of concerted evolution events. These “inconsistencies” have been previously reported for *csd* alleles. The combined forces of meiotic recombination and balancing selection generate a heterogeneity of divergence across the *csd* gene (HASSELMANN & BEYE 2004; HASSELMANN & BEYE 2006; HASSELMANN *et al.* 2008a). Recombination redistributes a small subset of variants of the 5´ region with multiple, highly diverged 3´ variants, which generates inconsistencies in the resulting phylogenetic relationships as previously shown (HASSELMANN & BEYE 2006; HASSELMANN *et al.* 2008a; BEYE *et al.* 2013).

Privman et al. found *fem*-specific substitutions for *csd* allele AY352276. However, this sequence is actually a chimeric sequence that was generated by the misassembly of the *fem* and *csd* cDNA sequences. This sequence was generated at a point in time when we searched for a third *csd* allele and had no knowledge of a second gene or the nature of the allelic diversity (BEYE *et al.* 2003). This sequence entry has been removed from GenBank. Privman et al. also suggested that alternative tree topologies of *fem* and *fem1* sequences in the ants indicate gene conversion events (PRIVMAN *et al.* 2013). We argue that the divergence of these sequences is too high (d_S= 0.5 for most species pairs) to exclude the possibility that ambiguous trees (and split phylogenetic networks) resulted from homoplasic (convergent) substitutions in the sequences.

Our results imply that the *fem1* and *csd* genes in the ant, bumblebee and honeybee species are not orthologs because they originated independently through gene duplications (Fig. S1). The *csd* gene originated in the honeybee lineage (HASSELMANN *et al.* 2008b; GEMPE *et al.* 2009). Hence, complementary sex determination in bumblebee and ant species (COOK 1993; HEIMPEL & DE BOER 2008) is regulated by other genes and not by the orthologs of the *csd* gene. Consistent with this conclusion, the *csd* alleles of honeybees share a hypervariable region of asparagine- and tyrosine-enriched repeats (HASSELMANN *et al.* 2008a) that are consistently absent in the *fem1* genes of ant and bumblebee species.

Our results show that a new gene for complementary sex determination originated in honeybees (HASSELMANN *et al.* 2008b), while the phylogenetic

distribution of complementary sex determination indicates a deep ancestry in Hymenopteran insects (COOK 1993; HEIMPEL & DE BOER 2008). One explanation for the replacement of a complementary sex determination gene is that ancestral complementary sex determiner genes degenerate over evolutionary time (HASSELMANN *et al.* 2008b; GEMPE & BEYE 2011). This is because meiotic recombination is suppressed at the sex determiner gene locus (HASSELMANN & BEYE 2006), allowing more deleterious mutations to accumulate over time (CHARLESWORTH & CHARLESWORTH 1978; CHARLESWORTH *et al.* 2005; GRAVES 2006). This process could generate an adaptive advantage for evolving new sex determination genes that would eventually replace the older, malfunctioning, complementary sex determination gene. Such a degeneration process has been proposed for sex chromosomal systems (BULL 1983; MARIN & BAKER 1998; SCHUTT & NOTHIGER 2000; MARSHALL GRAVES 2008; SANCHEZ 2008; MANK & AVISE 2009) and may also explain the rapid evolution of the complementary sex determination system.

Characterizing the gene functions of the *fem1* genes and the molecular basis of complementary sex determination in ants and bumblebees would provide interesting insights into the evolution of this sex determination system. Our results imply that the conserved phenotype (complementary sex determination) is only loosely evolutionary associated with the controlling molecular process. These findings highlight the limits of comparative genomics and emphasize the requirement to study gene functions in different species and major hymenopteran lineages.

Supplementary Information

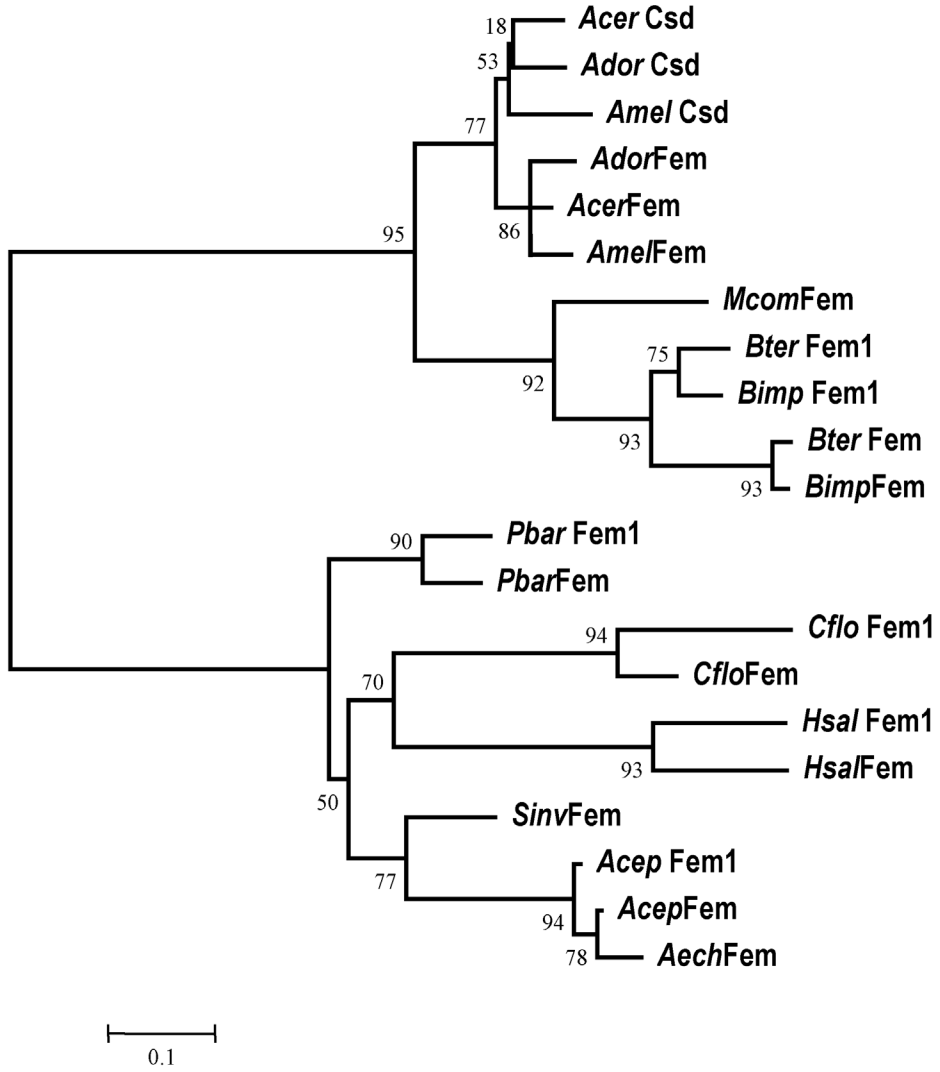

Figure S1. Gene tree of the *fem* and *fem1/csd* sister copies in ants and bees, which were inferred from synonymous differences. The evolutionary history was inferred using the neighbor-joining method. The confidence probability (multiplied by 100) that the interior branch length is greater than 0 was estimated using the bootstrap test (10000 replicates are shown next to the branches). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Pamilo-Bianchi-Li method (PAMILO & BIANCHI 1993) and are in the units of the number of synonymous substitutions per synonymous site. All ambiguous positions were removed for each sequence pair. There were a total of 575 positions in the final dataset. Evolutionary analyses were conducted in MEGA 5 (TAMURA *et al.* 2011). The sequences of *Nasonia* and *Ceratitis* were excludeto estimate d_S. Abbreviations: *Acep*, *Atta cephalotes*; *Acer, Apis cerana*; *Aech*, *Acromyrmex echinatior*; *Ador, Apis dorsata*; *Amel*, *Apis mellifera*; *Bimp, Bombus impatiens*; *Bter*, *Bombus terrestris*; *Cflo*, *Camponotus floridanus; Hsal*, *Harpegnathos saltator*; *Mcom, Melipona compressipes*; *Pbar*, *Pogonomyrmex barbatus*; *Sinv, Solenopsis invicta*.

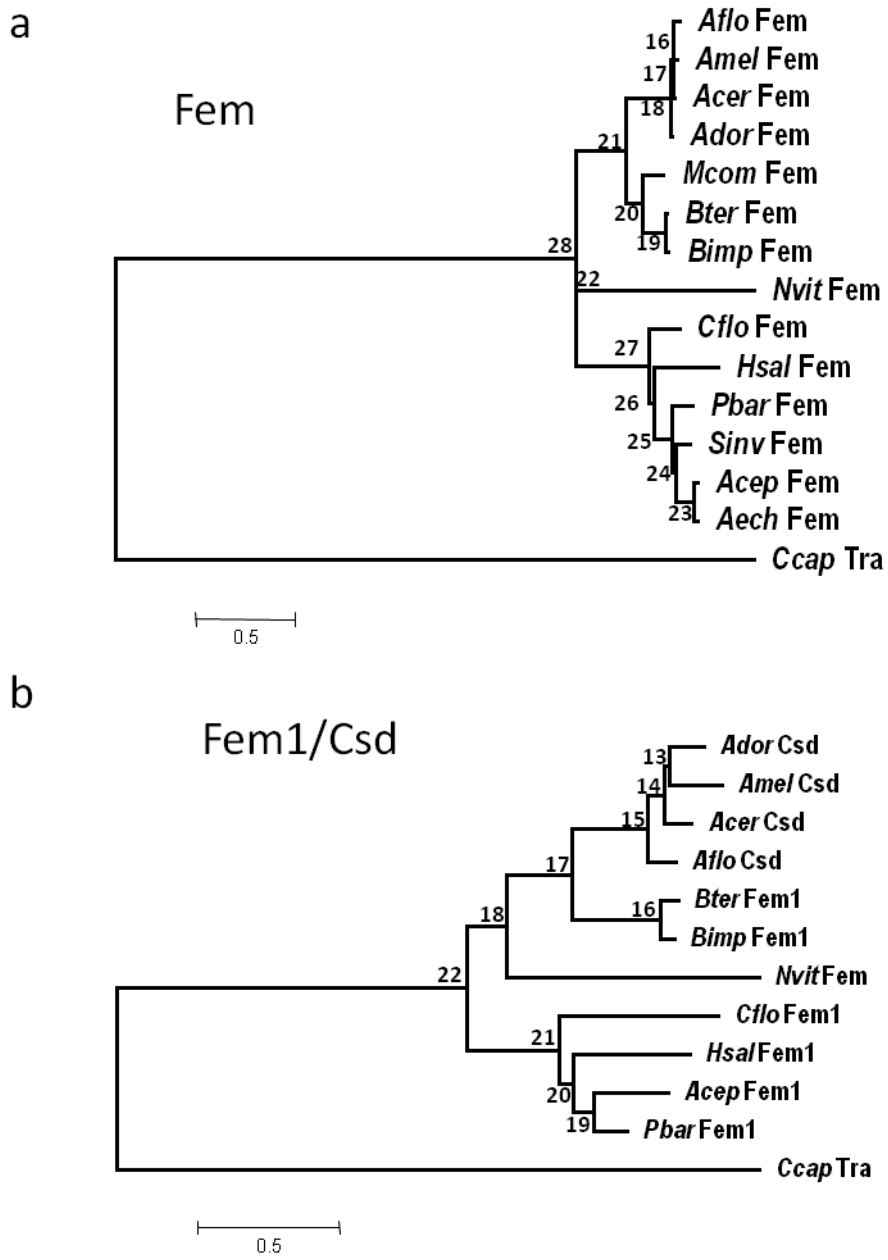

Figure S2. The initial tree of the Fem (a) and Fem1/Csd (b) proteins that were used to infer the ancestral amino acid sequences. The evolutionary history was inferred by using the maximum likelihood method based on the JFF model (JONES *et al.* 1992). Numbers in the tree assign the different nodes for which the ancestral sequence were obtained. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Abbreviations: Acep, Atta cephalotes; Acer, Apis cerana; Aech, Acromyrmex echinatior; Ador, Apis dorsata; Aflo, Apis florea; Amel, Apis mellifera; Bimp, Bombus impatiens; Bter, Bombus terrestris; Ccap, Ceratitis capitata; Cflo, Camponotus floridanus; Hsal, Harpegnathos saltator; Mcom, Melipona compressipes; Nvit, Nasonia vitripennis; Pbar, Pogonomyrmex barbatus; Sinv, Solenopsis invicta.

Figure S3. The informative substitutions found in the ant lineage that were used in Figure 2. The identity of the different species and nodes of the Fem and Fem1 protein tree is shown. Site number (#) indicates the positions in the Fem and in the Csd/Fem1 protein sequence alignment.

Node 27 – Fem	**Node 28 – Fem**	**Node21 – Fem1**
Q	**M**	**Q**

Position in Fem protein sequence alignment: # 24							
Site in Fem tree	*C.flo.* **Q**	*A.cep.* **Q**	*H.sal.* **R**	*A.ech.* **Q**	*S.inv.* **Q**	*P.bar.* **Q**	*N.vit.* **M**
	B.ter. **M**	*B.imp.* **M**	*A.cer.* **M**	*A.dor.* **M**	*A.flo.* **M**	*A.mel.* **M**	*M.com.* **M**

Position in Csd/Fem1 protein sequence alignment: # 21				
Site in Csd / Fem1 tree	*C.flo.* **Q**	*A.cep.* **Q**	*H.sal.* **I**	*P.bar.* **Q**

Node 27 – Fem	**Node 28 – Fem**	**Node21 – Fem1**
R	**I**	**R**

Position in Fem protein sequence alignment: # 25							
Site in Fem tree	*C.flo.* **R**	*A.cep.* **R**	*H.sal.* **H**	*A.ech.* **R**	*S.inv.* **R**	*P.bar.* **R**	*N.vit.* **V**
	B.ter. **I**	*B.imp.* **I**	*A.cer.* **I**	*A.dor.* **I**	*A.flo.* **I**	*A.mel.* **I**	*M.com.* **I**

Position in Csd/Fem1 protein sequence alignment: # 22				
Site in Csd / Fem1 tree	*C.flo.* **Q**	*A.cep.* **R**	*H.sal.* **R**	*P.bar.* **R**

Node 27 – Fem	**Node 28 – Fem**	**Node21 – Fem1**
Q	**E**	**Q**

Position in Fem protein sequence alignment: # 32							
Site in Fem tree	*C.flo.* **Q**	*A.cep.* **Q**	*H.sal.* **Q**	*A.ech.* **Q**	*S.inv.* **Q**	*P.bar.* **Q**	*N.vit.* **E**
	B.ter. **E**	*B.imp.* **E**	*A.cer.* **E**	*A.dor.* **E**	*A.flo.* **E**	*A.mel.* **E**	*M.com.* **E**

Position in Csd/Fem1 protein sequence alignment: # 29				
Site in Csd / Fem1 tree	*C.flo.* **Q**	*A.cep.* **Q**	*H.sal.* **Q**	*P.bar.* **E**

Node 27 – Fem	**Node 28 – Fem**	**Node21 – Fem1**
T	**S**	**T**

Position in Fem protein sequence alignment: # 81							
Site in Fem tree	*C.flo.* **T**	*A.cep.* **T**	*H.sal.* **A**	*A.ech.* **T**	*S.inv.* **T**	*P.bar.* **T**	*N.vit.* **S**
	B.ter. **S**	*B.imp.* **S**	*A.cer.* **S**	*A.dor.* **S**	*A.flo.* **S**	*A.mel.* **S**	*M.com.* **S**

Position in Csd/Fem1 protein sequence alignment: # 77				
Site in Csd / Fem1 tree	*C.flo.* **T**	*A.cep.* **T**	*H.sal.* **A**	*P.bar.* **T**

Node 27 – Fem	**Node 28 – Fem**	**Node21 – Fem1**
T	**A**	**T**

Position in Fem protein sequence alignment: # 96							
Site in Fem tree	*C.flo.* **T**	*A.cep.* **T**	*H.sal.* **T**	*A.ech.* **T**	*S.inv.* **T**	*P.bar.* **T**	*N.vit.* **E**
	B.ter. **A**	*B.imp.* **A**	*A.cer.* **A**	*A.dor.* **A**	*A.flo.* **A**	*A.mel.* **A**	*M.com.* **I**

Position in Csd/Fem1 protein sequence alignment: # 91				
Site in Csd / Fem1 tree	*C.flo.* **T**	*A.cep.* **T**	*H.sal.* **T**	*P.bar.* **T**

Node 27 – Fem	**Node 28 – Fem**	**Node21 – Fem1**
T	**A**	**T**

Position in Fem protein sequence alignment: # 155							
Site in Fem tree	*C.flo.* **D**	*A.cep.* **D**	*H.sal.* **D**	*A.ech.* **D**	*S.inv.* **D**	*P.bar.* **D**	*N.vit.* **G**
	B.ter. **N**	*B.imp.* **N**	*A.cer.* **N**	*A.dor.* **N**	*A.flo.* **N**	*A.mel.* **N**	*M.com.* **N**

Position in Csd/Fem1 protein sequence alignment: # 150				
Site in Csd / Fem1 tree	*C.flo.* **G**	*A.cep.* **N**	*H.sal.* **D**	*P.bar.* **D**

Node 27 – Fem	**Node 28 – Fem**	**Node21 – Fem1**
S	**L**	**S**

Position in Fem protein sequence alignment: # 159							
Site in Fem tree	*C.flo.* **Q**	*A.cep.* **S**	*H.sal.* **S**	*A.ech.* **S**	*S.inv.* **S**	*P.bar.* **S**	*N.vit.* **L**
	B.ter. **L**	*B.imp.* **L**	*A.cer.* **S**	*A.dor.* **P**	*A.flo.* **S**	*A.mel.* **S**	*M.com.* **L**

Position in Csd/Fem1 protein sequence alignment: # 154				
Site in Csd / Fem1 tree	*C.flo.* **S**	*A.cep.* **S**	*H.sal.* **S**	*P.bar.* **S**

Node 27 – Fem	**Node 28 – Fem**	**Node21 – Fem1**
A	**P**	**A**

Position in Fem protein sequence alignment: # 254							
Site in Fem tree	*C.flo.* **A**	*A.cep.* **A**	*H.sal.* **A**	*A.ech.* **T**	*S.inv.* **A**	*P.bar.* **T**	*N.vit.* **P**
	B.ter. **P**	*B.imp.* **P**	*A.cer.* **P**	*A.dor.* **P**	*A.flo.* **P**	*A.mel.* **P**	*M.com.* **S**

Position in Csd/Fem1 protein sequence alignment: # 253				
Site in Csd / Fem1 tree	*C.flo.* **A**	*A.cep.* **A**	*H.sal.* **M**	*P.bar.* **T**

Node 27 – Fem **V**	**Node 28 – Fem** **M**	**Node21 – Fem1** **V**

Position in Fem protein sequence alignment: # 274							
Site in Fem tree	*C.flo.* **V**	*A.cep.* **V**	*H.sal.* **M**	*A.ech.* **V**	*S.inv.* **V**	*P.bar.* **V**	*N.vit.* **L**
	B.ter. **M**	*B.imp.* **M**	*A.cer.* **M**	*A.dor.* **M**	*A.flo.* **M**	*A.mel.* **M**	*M.com.* **M**

Position in Csd/Fem1 protein sequence alignment: # 273				
Site in Csd / Fem1 tree	*C.flo.* **V**	*A.cep.* **V**	*H.sal.* **M**	*P.bar.* **V**

Figure S4. The informative substitutions found in the bee lineage that were used in Figure 2. The identity of the different species and nodes of the Fem and Csd/Fem1 protein tree is shown. Site number (#) indicates the positions in the Fem and in the Csd/Fem1 protein sequence alignment.

Node 21 – Fem **T**	**Node 22 – Fem** **S**	**Node17 – Csd / Fem1** **T**

Position in Fem protein sequence alignment: # 74							
Site in Fem tree	*C.flo.* **S**	*A.cep.* **S**	*H.sal.* **S**	*A.ech.* **S**	*S.inv.* **S**	*P.bar.* **S**	*N.vit.* **S**
	B.ter. **T**	*B.imp.* **T**	*A.cer.* **T**	*A.dor.* **T**	*A.flo.* **T**	*A.mel.* **T**	*M.com.* **T**

Position in Csd/Fem1 protein sequence alignment: # 69						
Site in Csd /Fem1 tree	*B.ter.* **T**	*B.imp.* **T**	*A.cer.* **T**	*A.dor.* **I**	*A.flo.* **T**	*A.mel.* **T**

Node 21 – Fem **E**	**Node 22 – Fem** **D**	**Node17 – Csd / Fem1** **E**

Position in Fem protein sequence alignment: # 79							
Site in Fem tree	*C.flo.* **E**	*A.cep.* **D**	*H.sal.* **D**	*A.ech.* **D**	*S.inv.* **D**	*P.bar.* **D**	*N.vit.* **D**
	B.ter. **E**	*B.imp.* **E**	*A.cer.* **E**	*A.dor.* **E**	*A.flo.* **E**	*A.mel.* **E**	*M.com.* **E**

Position in Csd/Fem1 protein sequence alignment: # 74						
Site in Csd /Fem1 tree	*B.ter.* **E**	*B.imp.* **E**	*A.cer.* **K**	*A.dor.* **E**	*A.flo.* **K**	*A.mel.* **E**

Node 21 – Fem **I**	**Node 22 – Fem** **V**	**Node17 – Csd / Fem1** **I**

Position in Fem protein sequence alignment: # 119							
Site in Fem tree	*C.flo.* **V**	*A.cep.* **L**	*H.sal.* **L**	*A.ech.* **L**	*S.inv.* **L**	*P.bar.* **L**	*N.vit.* **V**
	B.ter. **I**	*B.imp.* **I**	*A.cer.* **I**	*A.dor.* **I**	*A.flo.* **I**	*A.mel.* **I**	*M.com.* **V**

Position in Csd/Fem1 protein sequence alignment: # 116						
Site in Csd /Fem1 tree	*B.ter.* **I**	*B.imp.* **I**	*A.cer.* **I**	*A.dor.* **I**	*A.flo.* **I**	*A.mel.* **I**

Node 21 – Fem	**Node 22 – Fem**	**Node17 – Csd / Fem1**
E	**D**	**E**

Position in Fem protein sequence alignment: # 135							
Site in Fem tree	*C.flo.* **D**	*A.cep.* **D**	*H.sal.* **E**	*A.ech.* **D**	*S.inv.* **D**	*P.bar.* **D**	*N.vit.* **D**
	B.ter. **E**	*B.imp.* **E**	*A.cer.* **E**	*A.dor.* **E**	*A.flo.* **E**	*A.mel.* **E**	*M.com.* **E**

Position in Csd/Fem1 protein sequence alignment: # 130						
Site in Csd /Fem1 tree	*B.ter.* **E**	*B.imp.* **E**	*A.cer.* **E**	*A.dor.* **E**	*A.flo.* **E**	*A.mel.* **E**

Node 21 – Fem	**Node 22 – Fem**	**Node17 – Csd / Fem1**
E	**D**	**E**

Position in Fem protein sequence alignment: # 240							
Site in Fem tree	*C.flo.* **D**	*A.cep.* **D**	*H.sal.* **D**	*A.ech.* **D**	*S.inv.* **D**	*P.bar.* **E**	*N.vit.* **D**
	B.ter. **E**	*B.imp.* **E**	*A.cer.* **E**	*A.dor.* **E**	*A.flo.* **E**	*A.mel.* **E**	*M.com.* **E**

Position in Csd/Fem1 protein sequence alignment: # 235						
Site in Csd /Fem1 tree	*B.ter.* **E**	*B.imp.* **E**	*A.cer.* **E**	*A.dor.* **E**	*A.flo.* **E**	*A.mel.* **E**

Node 21 – Fem	**Node 22 – Fem**	**Node17 – Csd / Fem1**
V	**I**	**V**

Position in Fem protein sequence alignment: # 264							
Site in Fem tree	*C.flo.* **I**	*A.cep.* **I**	*H.sal.* **I**	*A.ech.* **I**	*S.inv.* **I**	*P.bar.* **I**	*N.vit.* **I**
	B.ter. **V**	*B.imp.* **V**	*A.cer.* **V**	*A.dor.* **V**	*A.flo.* **V**	*A.mel.* **V**	*M.com.* **F**

Position in Csd/Fem1 protein sequence alignment: # 263						
Site in Csd /Fem1 tree	*B.ter.* **V**	*B.imp.* **V**	*A.cer.* **V**	*A.dor.* **V**	*A.flo.* **V**	*A.mel.* **V**

Node 21 – Fem	**Node 22 – Fem**	**Node17 – Csd / Fem1**
M	**L**	**M**

Position in Fem protein sequence alignment: # 290							
Site in Fem tree	*C.flo.* **L**	*A.cep.* **L**	*H.sal.* **L**	*A.ech.* **L**	*S.inv.* **L**	*P.bar.* **L**	*N.vit.* **L**
	B.ter. **M**	*B.imp.* **M**	*A.cer.* **I**	*A.dor.* **I**	*A.flo.* **I**	*A.mel.* **I**	*M.com.* **M**

Position in Csd/Fem1 protein sequence alignment: # 287						
Site in Csd /Fem1 tree	*B.ter.* **M**	*B.imp.* **M**	*A.cer.* **I**	*A.dor.* **I**	*A.flo.* **I**	*A.mel.* **I**

Figure S5. The informative substitutions found in the *Bombus* lineage (Fem tree) that were used in Figure 3. The identity of the different species and nodes of the Fem and Fem1 protein tree is shown. Site number (#) indicates the position in the alignment of the Fem and of the Csd/Fem1 protein sequences.

Node 19-Fem **L**

Position in Fem protein sequence alignment: # 48					
Site in Fem tree	Node 12 (*B.imp.*)-Fem **I**		Node 1 (*B.ter.*)-Fem **L**		
	A.mel.-Fem **R**	*A.dor.*-Fem **R**	*A.cer.*-Fem **R**	*A.flo*-Fem **R**	*M.com.*-Fem **R**

Position in Csd/Fem1 protein sequence alignment: # 44		
Site in Csd /Fem1 tree	Node 11 (*B.imp*)-Fem 1 **L**	Node 1 (*B.ter*)-Fem 1 **L**

Node 19-Fem **V**

Position in Fem protein sequence alignment: # 146					
Site in Fem tree	Node 12 (*B.imp.*)-Fem **E**		Node 1 (*B.ter.*)-Fem **V**		
	A.mel.-Fem **T**	*A.dor.*-Fem **T**	*A.cer.*-Fem **T**	*A.flo*-Fem **T**	*M.com.*-Fem **V**

Position in Csd/Fem1 protein sequence alignment: # 141		
Site in Csd /Fem1 tree	Node 11 (*B.imp*)-Fem 1 **V**	Node 1 (*B.ter*)-Fem 1 **V**

Node 19-Fem **V**

Position in Fem protein sequence alignment: # 164					
Site in Fem tree	Node 12 (*B.imp.*)-Fem **M**		Node 1 (*B.ter.*)-Fem **V**		
	A.mel.-Fem **T**	*A.dor.*-Fem **T**	*A.cer.*-Fem **T**	*A.flo*-Fem **T**	*M.com.*-Fem **V**

Site was deleted in the Fem tree analysis due to gaps (site # 244 in the overall alignment)		
Site in Csd /Fem1 tree	Node 11 (*B.imp*)-Fem 1 **A**	Node 1 (*B.ter*)-Fem 1 **A**

Node 19-Fem **G**

Position in Fem protein sequence alignment: # 201					
Site in Fem tree	Node 12 (*B.imp.*)-Fem **G**		Node 1 (*B.ter.*)-Fem **D**		
	A.mel.-Fem **D**	*A.dor.*-Fem **D**	*A.cer.*-Fem **D**	*A.flo*-Fem **D**	*M.com.*-Fem **G**

Position in Csd/Fem1 protein sequence alignment: # 186		
Site in Csd /Fem1 tree	Node 11 (*B.imp*)-Fem 1 **G**	Node 1 (*B.ter*)-Fem 1 **G**

Node 19-Fem **R**

Position in Fem protein sequence alignment: # 210					
Site in Fem tree	Node 12 (*B.imp.*)-Fem **R**		Node 1 (*B.ter.*)-Fem **K**		
	A.mel.-Fem **R**	*A.dor.*-Fem **R**	*A.cer.*-Fem **R**	*A.flo*-Fem **R**	*M.com.*-Fem **R**

Position in Csd/Fem1 protein sequence alignment: # 195		
Site in Csd /Fem1 tree	Node 11 (*B.imp*)-Fem 1 **R**	Node 1 (*B.ter*)-Fem 1 **R**

Node 19-Fem **M**

Position in Fem protein sequence alignment: # 291					
Site in Fem tree	Node 12 (*B.imp.*)-Fem **I**		Node 1 (*B.ter.*)-Fem **M**		
	A.mel.-Fem **G**	*A.dor.*-Fem **G**	*A.cer.*-Fem **G**	*A.flo*-Fem **G**	*M.com.*-Fem **M**

Position in Csd/Fem1 protein sequence alignment: # 288		
Site in Csd /Fem1 tree	Node 11 (*B.imp*)-Fem 1 **M**	Node 1 (*B.ter*)-Fem 1 **M**

Figure S6:The informative substitutions found in the *Bombus* lineage (Fem1 tree) that were used in Figure 3. The identity of the different species and nodes of the Fem and Fem1 protein tree is shown. Site number (#) indicate the position in the alignment of the Fem and of the Csd/Fem1 protein sequences.

Node 16-Fem1 **R**

Position in Csd/Fem1 protein sequence alignment: # 17				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **R**		Node 1 (*B.ter.*) **K**	
	A.mel.-Csd **R**	*A.dor.*-Csd **K**	*A.cer.*-Csd **R**	*A.flo.*-Csd **R**

Position in Fem protein sequence alignment: # 20		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **R**	Node 1 (*B.ter*)-Fem 1 **R**

Node 16-Fem1 **I**

Position in Csd/Fem1 protein sequence alignment: # 22				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **V**		Node 1 (*B.ter.*) **I**	
	A.mel.-Csd **V**	*A.dor.*-Csd **I**	*A.cer.*-Csd **I**	*A.flo.*-Csd **I**

Position in Fem protein sequence alignment: # 25		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **I**	Node 1 (*B.ter*)-Fem 1 **I**

Node 16-Fem1 **K**

Position in Csd/Fem1 protein sequence alignment: # 32				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **K**		Node 1 (*B.ter.*) **R**	
	A.mel.-Csd **K**	*A.dor.*-Csd **K**	*A.cer.*-Csd **K**	*A.flo.*-Csd **R**

Position in Fem protein sequence alignment: # 35		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **K**	Node 1 (*B.ter*)-Fem 1 **K**

Node 16-Fem1 **K**

Position in Csd/Fem1 protein sequence alignment: # 52				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **T**		Node 1 (*B.ter.*) **K**	
	A.mel.-Csd **K**	*A.dor.*-Csd **K**	*A.cer.*-Csd **K**	*A.flo.*-Csd **K**

Position in Fem protein sequence alignment: # 57		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **K**	Node 1 (*B.ter*)-Fem 1 **K**

Node 16-Fem1 **S**

Position in Csd/Fem1 protein sequence alignment: # 53				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **S**		Node 1 (*B.ter.*) **N**	
	A.mel.-Csd **T**	*A.dor.*-Csd **T**	*A.cer.*-Csd **T**	*A.flo.*-Csd **T**

Position in Fem protein sequence alignment: # 58		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **S**	Node 1 (*B.ter*)-Fem 1 **S**

Node 16-Fem1 **K**

Position in Csd/Fem1 protein sequence alignment: # 54				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **E**		Node 1 (*B.ter.*) **K**	
	A.mel.-Csd **K**	*A.dor.*-Csd **K**	*A.cer.*-Csd **K**	*A.flo.*-Csd **K**

Position in Fem protein sequence alignment: # 59		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **K**	Node 1 (*B.ter*)-Fem 1 **K**

Node 16-Fem1 **H**

Position in Csd/Fem1 protein sequence alignment: # 62				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **H**		Node 1 (*B.ter.*) **Q**	
	A.mel.-Csd **D**	*A.dor.*-Csd **D**	*A.cer.*-Csd **D**	*A.flo.*-Csd **K**

Position in Fem protein sequence alignment: # 64		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **H**	Node 1 (*B.ter*)-Fem 1 **H**

Node 16-Fem1 **S**

Position in Csd/Fem1 protein sequence alignment: # 64				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **L**		Node 1 (*B.ter.*) **S**	
	A.mel.-Csd **S**	*A.dor.*-Csd **S**	*A.cer.*-Csd **S**	*A.flo.*-Csd **P**

Position in Fem protein sequence alignment: # 69		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **S**	Node 1 (*B.ter*)-Fem 1 **S**

Node 16-Fem1 **N**

Position in Csd/Fem1 protein sequence alignment: # 65				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **S**		Node 1 (*B.ter.*) **N**	
	A.mel.-Csd **N**	*A.dor.*-Csd **N**	*A.cer.*-Csd **N**	*A.flo.*-Csd **N**

Position in Fem protein sequence alignment: # 70		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **N**	Node 1 (*B.ter*)-Fem 1 **N**

Node 16-Fem1 **P**

Position in Csd/Fem1 protein sequence alignment: # 85				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **P**		Node 1 (*B.ter.*) **L**	
	A.mel.-Csd **P**	*A.dor.*-Csd **S**	*A.cer.*-Csd **P**	*A.flo.*-Csd **P**

Position in Fem protein sequence alignment: # 90		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **P**	Node 1 (*B.ter*)-Fem 1 **P**

Node 16-Fem1 **V**

Position in Csd/Fem1 protein sequence alignment: # 137				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **V**		Node 1 (*B.ter.*) **A**	
	A.mel.-Csd **V**	*A.dor.*-Csd **V**	*A.cer.*-Csd **V**	*A.flo.*-Csd **V**

Position in Fem protein sequence alignment: # 142		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **T**	Node 1 (*B.ter*)-Fem 1 **T**

Node 16-Fem1 **T**

Position in Csd/Fem1 protein sequence alignment: # 138				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **T**		Node 1 (*B.ter.*) **R**	
	A.mel.-Csd **L**	*A.dor.*-Csd **L**	*A.cer.*-Csd **F**	*A.flo.*-Csd **L**

Position in Fem protein sequence alignment: # 143		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **T**	Node 1 (*B.ter*)-Fem 1 **T**

Node 16-Fem1 **K**

Position in Csd/Fem1 protein sequence alignment: # 158				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **E**		Node 1 (*B.ter.*) **K**	
	A.mel.-Csd **K**	*A.dor.*-Csd **N**	*A.cer.*-Csd **K**	*A.flo.*-Csd **k**

Position in Fem protein sequence alignment: # 167		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **K**	Node 1 (*B.ter*)-Fem 1 **K**

Node 16-Fem1 **E**

Position in Csd/Fem1 protein sequence alignment: # 218				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **G**		Node 1 (*B.ter.*) **E**	
	A.mel.-Csd **E**	*A.dor.*-Csd **E**	*A.cer.*-Csd **E**	*A.flo.*-Csd **E**

Position in Fem protein sequence alignment: # 227		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **E**	Node 1 (*B.ter*)-Fem 1 **E**

Node 16-Fem1 **R**

Position in Csd/Fem1 protein sequence alignment: # 247				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **R**		Node 1 (*B.ter.*) **G**	
	A.mel.-Csd **K**	*A.dor.*-Csd **R**	*A.cer.*-Csd **R**	*A.flo.*-Csd **R**

Position in Fem protein sequence alignment: # 249		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **R**	Node 1 (*B.ter*)-Fem 1 **R**

Node 16-Fem1 **R**

Position in Csd/Fem1 protein sequence alignment: # 250				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **R**		Node 1 (*B.ter.*) **T**	
	A.mel.-Csd **K**	*A.dor.*-Csd **R**	*A.cer.*-Csd **K**	*A.flo.*-Csd **T**

Position in Fem protein sequence alignment: # 251		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **R**	Node 1 (*B.ter*)-Fem 1 **R**

Node 16-Fem1 **P**

Position in Csd/Fem1 protein sequence alignment: # 274				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **P**		Node 1 (*B.ter.*) **H**	
	A.mel.-Csd **P**	*A.dor.*-Csd **P**	*A.cer.*-Csd **P**	*A.flo.*-Csd **S**

Position in Fem protein sequence alignment: # 277		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **P**	Node 1 (*B.ter*)-Fem 1 **P**

Node 16-Fem1 **P**

Position in Csd/Fem1 protein sequence alignment: # 300				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **P**		Node 1 (*B.ter.*) **S**	
	A.mel.-Csd **P**	*A.dor.*-Csd **P**	*A.cer.*-Csd **P**	*A.flo.*-Csd **P**

Position in Fem protein sequence alignment: # 309		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **P**	Node 1 (*B.ter*)-Fem 1 **P**

Node 16-Fem1 **L**

Position in Csd/Fem1 protein sequence alignment: # 301				
Site in Csd /Fem1 tree	Node 11 (*B.imp.*)-Fem1 **F**		Node 1 (*B.ter.*) **L**	
	A.mel.-Csd **P**	*A.dor.*-Csd **P**	*A.cer.*-Csd **P**	*A.flo.*-Csd **P**

Position in Fem protein sequence alignment: # 310		
Site in Fem tree	Node 11 (*B.imp*)-Fem 1 **P**	Node 1 (*B.ter*)-Fem 1 **P**

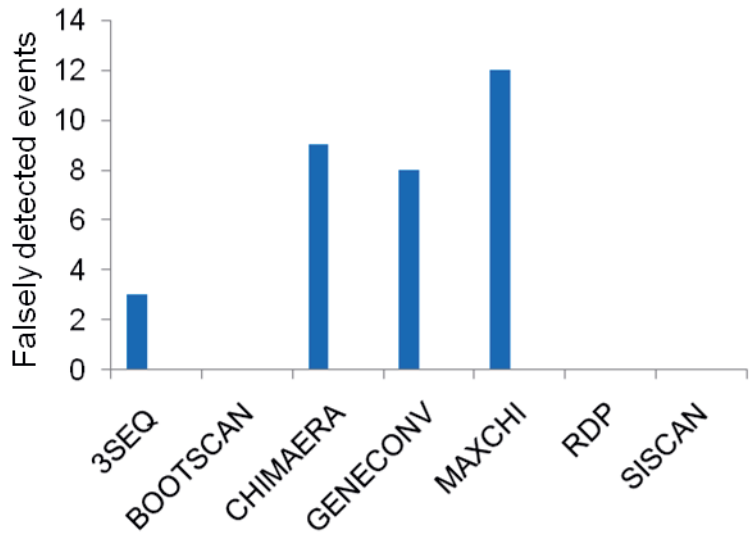

Figure S7: The number of falsely detected events (FDE) using the methods as indicated on the X axis. These programs were implemented in the RDP 3.44 software program. The methods were run on each of the 20 alignments which consisted of two randomly chosen *csd* sequences from a single *Apis* species and one *fem* sequence from *B. terrestris*.These events are falsely detected as this transfer involve the outgroup *Bombus* sequence and the polymorphism between csd alleles which newly evolved in the different *Apis* species.

Table S1. The $d_{S\,i}$ values for the paralogous gene pairs *fem* and *csd*/*fem*1 within each species (*Ador, Apis dorsata*; *Amel*, *Apis mellifera*; *Bimp, Bombus impatiens*; *Bter*, *Bombus terrestris*).

Species	$d_{S\,i}$ *between fem* and *csd/fem*1
A.mel	0.18
A.dor	0.18
B.ter	0.24
B.imp	0.22

Author's Contribution

Independent Evolutionary Origin of *fem* Paralogous Genes and Complementary Sex Determination in Hymenopteran Insects

Journal: PLOS ONE (in press)

Impact factor: 3.730

3rd author

Author's contribution: 15 %

- Identification of duplication tracts (together with Vasco Koch)
- Confirmation of duplicate gene expression by RT-PCR in *A. mellifera* (together with Vasco Koch)
- Analysis of dS divergence (together with Vasco Koch)
- Revising the manuscript

III. Zusammenfassung

Die Entscheidung für männliche oder weibliche Entwicklung wird in holometabolen Insekten durch das konservierte Schaltergen *fem*/*tra* implementiert, dessen Transkripte geschlechtsspezifisch gespleißt werden. In der Honigbiene bestimmt die allelische Komposition des *csd*-Gens, das in Folge einer Genduplikation aus dem Vorläufer des paralogen *fem*-Gens evolviert ist, welche geschlechtsspezifischen *fem*-Transkripte gebildet werden. Das von der weiblichen *fem*-mRNA kodierte Fem-Protein implementiert die weibliche Entwicklung und hält diese durch die autoregulatorische Prozession weiterer weiblicher *fem*-Transkripte aufrecht. Männliche Entwicklung erfolgt in Abwesenheit weiblicher Signale, wenn nur das nicht proteinkodierende männliche *fem*-Transkript gebildet wird.

Im Rahmen dieser Arbeit wurde gezeigt, dass die transiente Expression von Csd- oder Fem-Proteinen eine zeitlich begrenzte weibliche Regulation in Männchen induziert, die jedoch nicht ausreichend ist, um eine vollständige weibliche Entwicklung herbeizuführen. Dies legt den Schluss nahe, dass die Fem-Autoregulation ein langsamer Prozess ist, der nur durch ein länger wirkendes weibliches Csd- oder Fem-Signal vollständig aktiviert werden kann. Es wurde zudem gezeigt, dass das *Am*-Tra2-Protein eine duale Funktion erfüllt, da es sowohl direkt das Spleißen des männlichen *fem*-Transkripts kontrolliert als auch ein potentieller Kofaktor des autoregulativen Fem-Proteins ist. Es wird postuliert, dass die funktionelle Wechselwirkung zwischen der Fem-Autoregulation und der dualen Rolle des *Am*-Tra2-Proteins dazu dient das allelische *csd*-Signal zu überprüfen, um eine mögliche weibliche Hintergrundaktivität hemi-/homozygoter Csd-Proteine herauszufiltern und so robust das Geschlecht in allen Zellen von Weibchen und Männchen zu bestimmen.

In jüngster Zeit wurden weitere Paraloge des *fem*-Gens (genannt *fem1*) in den Genomen verschiedener Ameisen- und Hummelarten identifiziert, die möglicherweise das primäre Signal in anderen Hymenopteren darstellen. Es konnte gezeigt werden, dass die *csd/fem1*-Gene wiederholt unabhängig in den unterschiedlichen Hymenopterenlinien im Verlauf der letzten 100 Millionen Jahre entstanden sind. Dies impliziert, dass die weit verbreitete komplementäre Geschlechtsbestimmung von unterschiedlichen Genen reguliert wird. Darüber hinaus wurde gezeigt, dass eine konzertierte Evolution zwischen den paralogen Genpaaren *csd/fem1* und *fem* keine nachweisbare Rolle bei der Evolution dieser Gene gespielt hat.

IV. Summary

In holometabolous insects, sexual development is implemented through the production of sex-specific transcripts of the conserved *fem*/*tra* gene, which leads to the production of male and female *dsx* transcripts and sexual differentiation. In the honeybee, the allelic composition of the underlying gene *csd*, which evolved through a gene duplication event from the ancestral copy of the paralogous *fem* gene, is responsible for the production of the sex-specific transcripts of the *fem* gene.

The female *fem* mRNA encodes the functional Fem protein that implements and maintains female development and controls its own production in a positive feedback loop. Male development is set by default if only the non-productive male *fem* transcript is produced.

Here, it is shown that the transient expression of Csd or Fem proteins can induce a temporary female response in males that is, however, not sufficient to fully induce female development. This finding leads to the conclusion that the *fem* feedback loop acts slowly and requires a longer commitment by Csd or Fem proteins to fully activate. It is demonstrated that Am-Tra2 proteins serve a dual role, as they directly control male *fem* splicing and are potential cofactors of Fem proteins in the autoregulative production of female *fem* transcripts. It is proposed that the functional relationship of the slow feedback loop and the dual role of the *Am*-Tra2 protein serves to evaluate the commitment of the *csd* gene. This evaluation might filter out putative female background activity resulting from hemi-/homozygous Csd proteins, and thus robustly establishes stably sex-determined states in every cell of females and males.

Recently, other paralogs of the *fem* gene, called *fem1*, have been identified in several ant and bumblebee genomes, which might constitute the primary complementary signal in these hymenopterans.

In the present work, it is shown that the *csd/fem1* genes originated repeatedly and independently in the different hymenopteran lineages over the last 100 million years, which implicates that the widely shared and conserved complementary sex-determination mechanism is actually controlled by different genes. Furthermore, it is shown that concerted evolution between the paralogous gene pairs *csd/fem1* and *fem* played no detectable role in the evolution of these genes.

V. Literaturverzeichnis

Alvarez, M., Ruiz, M. F. & Sanchez, L., 2009 Effect of the gene *doublesex* of *Anastrepha* on the somatic sexual development of *Drosophila*. PLoS One 4**:** e5141.

Bell, G., 1982 The masterpiece of nature: The evolution and genetics of sexuality. Croom Helm.

Bell, L. R., Horabin, J. I., Schedl, P. & Cline, T. W., 1991 Positive autoregulation of *Sex-lethal* by alternative splicing maintains the female determined state in *Drosophila*. Cell 65**:** 229-239.

Beye, M., Hasselmann, M., Fondrk, M. K., Page, R. E. & Omholt, S. W., 2003 The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein. Cell 114**:** 419-429.

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

Beye, M., Seelmann, C., Gempe, T., Hasselmann, M., Vekemans, X. *et al.*, 2013 Gradual molecular evolution of a sex determination switch through incomplete penetrance of femaleness. Curr Biol 23: 2559-2564.

Black, D. L., 2003 Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem 72**:** 291-336.

Boni, M. F., Posada, D. & Feldman, M. W., 2007 An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics 176**:** 1035-1047.

Borst, J. W., & Visser, A. J. W. G., 2010 Fluorescence lifetime imaging microscopy in life sciences. Meas Sci Technol 21**:** 102002.

Brady, S. G., Schultz, T. R., Fisher, B. L. & Ward, P. S., 2006 Evaluating alternative hypotheses for the early evolution and diversification of ants. Proc Natl Acad Sci U S A 103**:** 18172-18177.

Bull, J. J., 1980 Sex determination in reptiles. Q Rev Biol 55**:** 3-21.

Bull, J. J., 1983 Evolution of sex determining mechanisms. Benjamin/Cummings Pub. Co., Advanced Book Program, Menlo Park, Calif.

Burge, C. B., Padgett, R. A. & Sharp, P. A., 1998 Evolutionary fates and origins of U12-type introns. Mol Cell 2**:** 773-785.

Burghardt, G., Hediger, M., Siegenthaler, C., Moser, M., Dubendorfer, A. *et al.*, 2005 The *transformer2* gene in *Musca domestica* is required for selecting and maintaining the female pathway of development. Dev Genes Evol 215**:** 165-176.

Burset, M., Seledtsov, I. A. & Solovyev, V. V., 2000 Analysis of canonical and non-canonical splice sites in mammalian genomes. Nucleic Acids Res 28**:** 4364-4375.

Burtis, K. C., & Baker, B. S., 1989 *Drosophila doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. Cell 56**:** 997-1010.

Camerino, G., Parma, P., Radi, O. & Valentini, S., 2006 Sex determination and sex reversal. Curr Opin Genet Dev 16**:** 289-292.

Cartegni, L., Chew, S. L. & Krainer, A. R., 2002 Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3**:** 285-298.

Charlesworth, B., & Charlesworth, D., 1978 A model for the evolution of dioecy and gynodioecy. Am Nat 112**:** 975-997.

Charlesworth, D., Charlesworth, B. & Marais, G., 2005 Steps in the evolution of heteromorphic sex chromosomes. Heredity (Edinb) 95**:** 118-128.

Chen, J. M., Cooper, D. N., Chuzhanova, N., Ferec, C. & Patrinos, G. P., 2007 Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet 8**:** 762-775.

Cho, S., Huang, Z. Y. & Zhang, J., 2007 Sex-specific splicing of the honeybee *doublesex* gene reveals 300 million years of evolution at the bottom of the insect sex-determination pathway. Genetics 177**:** 1733-1741.

Cline, T. W., 1984 Autoregulatory functioning of a *Drosophila* gene product that establishes and maintains the sexually determined state. Genetics 107**:** 231-277.

Cline, T. W., 1993 The *Drosophila* sex determination signal: how do flies count to two? Trends Genet 9**:** 385-390.

Cline, T. W., & Meyer, B. J., 1996 Vive la difference: males vs females in flies vs worms. Annu Rev Genet 30**:** 637-702.

Cline, T. W., Dorsett, M., Sun, S., Harrison, M. M., Dines, J. *et al.*, 2010 Evolution of the *Drosophila* feminizing switch gene *Sex-lethal*. Genetics 186**:** 1321-1336.

Concha, C., & Scott, M. J., 2009 Sexual development in *Lucilia cuprina* (Diptera, Calliphoridae) is controlled by the transformer gene. Genetics 182**:** 785-798.

Cook, J. M., 1993 Sex determination in the hymenoptera - a review of models and evidence. Heredity 71**:** 421-435.

Coschigano, K. T., & Wensink, P. C., 1993 Sex-specific transcriptional regulation by the male and female *doublesex* proteins of *Drosophila*. Genes Dev 7**:** 42-54.

Daubner, S. C., Astorga, A. M., Leisman, G. B. & Baldwin, T. O., 1987 Yellow light emission of *Vibrio fischeri* strain Y-1: purification and characterization of the energy-accepting yellow fluorescent protein. Proc Natl Acad Sci U S A 84**:** 8912-8916.

Dearden, P. K., Wilson, M. J., Sablan, L., Osborne, P. W., Havler, M. *et al.*, 2006 Patterns of conservation and change in honey bee developmental genes. Genome Res 16**:** 1376-1384.

Drescher, W., & Rothenbuhler, W. C., 1964 Sex Determination in the Honey Bee. J Hered 55**:** 91-96.

Du, H., & Rosbash, M., 2002 The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. Nature 419**:** 86-90.

Eirin-Lopez, J. M., Gonzalez-Tizon, A. M., Martinez, A. & Mendez, J., 2004 Birth-and-death evolution with strong purifying selection in the histone H1 multigene family and the origin of *orphon* H1 genes. Mol Biol Evol 21**:** 1992-2003.

Erickson, J. W., & Quintero, J. J., 2007 Indirect effects of ploidy suggest X chromosome dose, not the X:A ratio, signals sex in *Drosophila*. PLoS Biol 5**:** e332.

Erkelenz, S., Mueller, W. F., Evans, M. S., Busch, A., Schoneweis, K. *et al.*, 2013a Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. RNA 19**:** 96-102.

Erkelenz, S., Poschmann, G., Theiss, S., Stefanski, A., Hillebrand, F. *et al.*, 2013b Tra2-mediated recognition of HIV-1 5' splice site D3 as a key factor in the processing of *vpr* mRNA. J Virol 87**:** 2721-2734.

Freund, M., Asang, C., Kammler, S., Konermann, C., Krummheuer, J. *et al.*, 2003 A novel approach to describe a U1 snRNA binding site. Nucleic Acids Res 31**:** 6963-6975.

Gabrieli, P., Falaguerra, A., Siciliano, P., Gomulski, L. M., Scolari, F. *et al.*, 2010 Sex and the single embryo: early deveiopment in the Mediterranean fruit fly, *Ceratitis capitata*. BMC Dev Biol 10**:** 12.

Gempe, T., Hasselmann, M., Schiott, M., Hause, G., Otte, M. *et al.*, 2009 Sex determination in honeybees: two separate mechanisms induce and maintain the female pathway. PLoS Biol 7**:** e1000222.

Gempe, T., & Beye, M., 2011 Function and evolution of sex determination mechanisms, genes and pathways in insects. Bioessays 33**:** 52-60.

Gibbs, M. J., Armstrong, J. S. & Gibbs, A. J., 2000 Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16**:** 573-582.

Graveley, B. R., Hertel, K. J. & Maniatis, T., 1998 A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. EMBO J 17**:** 6747-6756.

Graveley, B. R., 2000 Sorting out the complexity of SR protein functions. RNA 6**:** 1197-1211.

Graves, J. A., 2006 Sex chromosome specialization and degeneration in mammals. Cell 124**:** 901-914.

Grellscheid, S. N., Dalgliesh, C., Rozanska, A., Grellscheid, D., Bourgeois, C. F. *et al.*, 2011 Molecular design of a splicing switch responsive to the RNA binding protein Tra2beta. Nucleic Acids Res 39**:** 8092-8104.

Grimaldi, D., & Engel, M. S., 2005 Evolution of the insects. University Press Cambridge, UK, Cambridge.

Hasselmann, M., & Beye, M., 2004 Signatures of selection among sex-determining alleles of the honey bee. Proc Natl Acad Sci U S A 101**:** 4888-4893.

Hasselmann, M., & Beye, M., 2006 Pronounced differences of recombination activity at the sex determination locus of the honeybee, a locus under strong balancing selection. Genetics 174**:** 1469-1480.

Hasselmann, M., Vekemans, X., Pflugfelder, J., Koeniger, N., Koeniger, G. *et al.*, 2008a Evidence for convergent nucleotide evolution and high allelic turnover rates at the *complementary sex determiner* gene of Western and Asian honeybees. Mol Biol Evol 25**:** 696-708.

Hasselmann, M., Gempe, T., Schiott, M., Nunes-Silva, C. G., Otte, M. *et al.*, 2008b Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. Nature 454**:** 519-522.

Hasselmann, M., Lechner, S., Schulte, C. & Beye, M., 2010 Origin of a function by tandem gene duplication limits the evolutionary capability of its sister copy. Proc Natl Acad Sci U S A 107**:** 13378-13383.

Hazelrigg, T., & Tu, C., 1994 Sex-specific processing of the *Drosophila* exuperantia transcript is regulated in male germ cells by the *tra-2* gene. Proc Natl Acad Sci U S A 91**:** 10752-10756.

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

Hediger, M., Henggeler, C., Meier, N., Perez, R., Saccone, G. *et al.*, 2010 Molecular characterization of the key switch F provides a basis for understanding the rapid divergence of the sex-determining pathway in the housefly. Genetics 184**:** 155-170.

Hedley, M. L., & Maniatis, T., 1991 Sex-specific splicing and polyadenylation of *dsx* pre-mRNA requires a sequence that binds specifically to *tra-2* protein in vitro. Cell 65**:** 579-586.

Heimpel, G. E., & de Boer, J. G., 2008 Sex determination in the hymenoptera. Annu Rev Entomol 53**:** 209-230.

Heinrichs, V., Ryner, L. C. & Baker, B. S., 1998 Regulation of sex-specific selection of fruitless 5' splice sites by *transformer* and *transformer-2*. Mol Cell Biol 18**:** 450-458.

Hertel, K. J., 2008 Combinatorial control of exon recognition. J Biol Chem 283**:** 1211-1215.

Hofmann, Y., Lorson, C. L., Stamm, S., Androphy, E. J. & Wirth, B., 2000 *Htra2-beta 1* stimulates an exonic splicing enhancer and can restore full-length *SMN* expression to *survival motor neuron 2* (*SMN2*). Proc Natl Acad Sci U S A 97**:** 9618-9623.

Holland, P. W., Garcia-Fernandez, J., Williams, N. A. & Sidow, A., 1994 Gene duplications and the origins of vertebrate development. Dev Suppl**:** 125-133.

Innan, H., 2009 Population genetic models of duplicated genes. Genetica 137**:** 19-37.

Inoue, K., Hoshijima, K., Higuchi, I., Sakamoto, H. & Shimura, Y., 1992 Binding of the *Drosophila* transformer and transformer-2 proteins to the regulatory elements of *doublesex* primary transcript for sex-specific RNA processing. Proc Natl Acad Sci U S A 89**:** 8092-8096.

Jones, D. T., Taylor, W. R. & Thornton, J. M., 1992 The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 8**:** 275-282.

Kamakura, M., 2011 Royalactin induces queen differentiation in honeybees. Nature 473**:** 478-483.

Kammler, S., Leurs, C., Freund, M., Krummheuer, J., Seidel, K. *et al.*, 2001 The sequence complementarity between HIV-1 5' splice site SD4 and U1 snRNA determines the steady-state level of an unstable *env* pre-mRNA. RNA 7**:** 421-434.

Keisman, E. L., & Baker, B. S., 2001 The *Drosophila* sex determination hierarchy modulates *wingless* and *decapentaplegic* signaling to deploy *dachshund* sex-specifically in the genital imaginal disc. Development 128**:** 1643-1656.

Keyes, L. N., Cline, T. W. & Schedl, P., 1992 The primary sex determination signal of *Drosophila* acts at the level of transcription. Cell 68**:** 933-943.

Kleinjan, D. A., Bancewicz, R. M., Gautier, P., Dahm, R., Schonthaler, H. B. *et al.*, 2008 Subfunctionalization of duplicated zebrafish *pax6* genes by *cis*-regulatory divergence. PLoS Genet 4**:** e29.

Kucharski, R., Maleszka, J., Foret, S. & Maleszka, R., 2008 Nutritional control of reproductive status in honeybees via DNA methylation. Science 319**:** 1827-1830.

Kudryavtsev, V., Felekyan, S., Wozniak, A. K., Konig, M., Sandhagen, C. *et al.*, 2007 Monitoring dynamic systems with multiparameter fluorescence imaging. Anal Bioanal Chem 387**:** 71-82.

Lagos, D., Ruiz, M. F., Sanchez, L. & Komitopoulou, K., 2005 Isolation and characterization of the *Bactrocera oleae* genes orthologous to the sex determining *Sex-lethal* and *doublesex* genes of *Drosophila melanogaster*. Gene 348**:** 111-121.

Lagos, D., Koukidou, M., Savakis, C. & Komitopoulou, K., 2007 The *transformer* gene in *Bactrocera oleae*: the genetic switch that determines its sex fate. Insect Mol Biol 16**:** 221-230.

Laidlaw, H. H., & Page, R. E., 1997 Queen rearing and bee breeding. Wicwas Press, Cheshire, Conn.

Lam, B. J., Bakshi, A., Ekinci, F. Y., Webb, J., Graveley, B. R. *et al.*, 2003 Enhancer-dependent 5'-splice site control of *fruitless* pre-mRNA splicing. J Biol Chem 278**:** 22740-22747.

Lareau, L. F., Green, R. E., Bhatnagar, R. S. & Brenner, S. E., 2004 The evolving roles of alternative splicing. Curr Opin Struct Biol 14**:** 273-282.

Lechner, S., Ferretti, L., Schoning, C., Kinuthia, W., Willemsen, D. *et al.*, 2013 Nucleotide variability at its Limit? Insights into the number and evolutionary dynamics of the sex-determining specificities of the honey bee *Apis mellifera*. Mol Biol Evol 31: 272-287.

Lewis, B. P., Green, R. E. & Brenner, S. E., 2003 Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci U S A 100**:** 189-192.

Liao, D., 1999 Concerted evolution: molecular mechanism and biological implications. Am J Hum Genet 64**:** 24-30.

Lipinski, K. J., Farslow, J. C., Fitzpatrick, K. A., Lynch, M., Katju, V. *et al.*, 2011 High spontaneous rate of gene duplication in *Caenorhabditis elegans*. Curr Biol 21**:** 306-310.

Lynch, K. W., & Maniatis, T., 1995 Synergistic interactions between two distinct elements of a regulated splicing enhancer. Genes Dev 9**:** 284-293.

Lynch, K. W., & Maniatis, T., 1996 Assembly of specific SR protein complexes on distinct regulatory elements of the *Drosophila doublesex* splicing enhancer. Genes Dev 10**:** 2089-2101.

Lynch, M., & Conery, J. S., 2000 The evolutionary fate and consequences of duplicate genes. Science 290**:** 1151-1155.

Lynch, M., & Force, A., 2000 The probability of duplicate gene preservation by subfunctionalization. Genetics 154**:** 459-473.

Madigan, S. J., Edeen, P., Esnayra, J. & McKeown, M., 1996 *att*, a target for regulation by *tra2* in the testes of *Drosophila melanogaster*, encodes alternative RNAs and alternative proteins. Mol Cell Biol 16**:** 4222-4230.

Maniatis, T., & Tasic, B., 2002 Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature 418**:** 236-243.

Mank, J. E., & Avise, J. C., 2009 Evolutionary diversity and turn-over of sex determination in teleost fishes. Sex Dev 3**:** 60-67.

Marin, I., & Baker, B. S., 1998 The evolutionary dynamics of sex determination. Science 281**:** 1990-1994.

Marshall Graves, J. A., 2008 Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. Annu.Rev.Genet. 42:565-86.**:** 565-586.

Martin, D., & Rybicki, E., 2000 RDP: detection of recombination amongst aligned sequences. Bioinformatics 16**:** 562-563.

Martin, D. P., Posada, D., Crandall, K. A. & Williamson, C., 2005 A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res Hum Retroviruses 21**:** 98-102.

Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D. *et al.*, 2010 RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26**:** 2462-2463.

Mattox, W., & Baker, B. S., 1991 Autoregulation of the splicing of transcripts from the *transformer-2* gene of *Drosophila*. Genes Dev 5**:** 786-796.

Mattox, W., McGuffin, M. E. & Baker, B. S., 1996 A negative feedback mechanism revealed by functional analysis of the alternative isoforms of the *Drosophila* splicing regulator *transformer-2*. Genetics 143**:** 303-314.

McGuffin, M. E., Chandler, D., Somaiya, D., Dauwalder, B. & Mattox, W., 1998 Autoregulation of *transformer-2* alternative splicing is necessary for normal male fertility in *Drosophila*. Genetics 149**:** 1477-1486.

Nagoshi, R. N., McKeown, M., Burtis, K. C., Belote, J. M. & Baker, B. S., 1988 The control of alternative splicing at genes regulating sexual differentiation in *D. melanogaster*. Cell 53**:** 229-236.

Nei, M., Gu, X. & Sitnikova, T., 1997 Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci U S A 94**:** 7799-7806.

Nei, M., & Kumar, S., 2000 Molecular evolution and phylogenetics. Oxford University Press, New York.

Nei, M., & Rooney, A. P., 2005 Concerted and birth-and-death evolution of multigene families. Annu Rev Genet 39**:** 121-152.

Niimura, Y., & Nei, M., 2003 Evolution of olfactory receptor genes in the human genome. Proc Natl Acad Sci U S A 100**:** 12235-12240.

Niimura, Y., & Nei, M., 2005 Evolutionary changes of the number of olfactory receptor genes in the human and mouse lineages. Gene 346**:** 23-28.

Nissen, I., Muller, M. & Beye, M., 2012a The *Am-tra2* gene is an essential regulator of female splice regulation at two levels of the sex determination hierarchy of the honeybee. Genetics 192**:** 1015-1026.

Nissen, I., Beye, M. & Rose, L., 2012b The control of the honeybee sex determination pathway by alternative splicing, Universitäts- und Landesbibliothek der Heinrich-Heine-Universität Düsseldorf, Düsseldorf.

Ohno, S., 1970 Evolution by gene duplication. Springer-Verlag, New York.

Ota, T., & Nei, M., 1994 Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. Mol Biol Evol 11**:** 469-482.

Ota, T., & Nei, M., 1995 Evolution of immunoglobulin VH pseudogenes in chickens. Mol Biol Evol 12**:** 94-102.

Padidam, M., Sawyer, S. & Fauquet, C. M., 1999 Possible emergence of new geminiviruses by frequent recombination. Virology 265**:** 218-225.

Pamilo, P., & Bianchi, N. O., 1993 Evolution of the *Zfx* and *Zfy* genes: rates and interdependence between the genes. Mol Biol Evol 10**:** 271-281.

Pane, A., Salvemini, M., Delli Bovi, P., Polito, C. & Saccone, G., 2002 The *transformer* gene in *Ceratitis capitata* provides a genetic basis for selecting and remembering the sexual fate. Development 129**:** 3715-3725.

Piontkivska, H., Rooney, A. P. & Nei, M., 2002 Purifying selection and birth-and-death evolution in the histone H4 gene family. Mol Biol Evol 19**:** 689-697.

Posada, D., & Crandall, K. A., 2001 Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc Natl Acad Sci U S A 98**:** 13757-13762.

Privman, E., Wurm, Y. & Keller, L., 2013 Duplication and concerted evolution in a master sex determiner under balancing selection. Proc Biol Sci 280**:** 20122968.

Qi, J., Su, S. & Mattox, W., 2007 The *doublesex* splicing enhancer components Tra2 and Rbp1 also repress splicing through an intronic silencer. Mol Cell Biol 27**:** 699-708.

Ramirez, S. R., Nieh, J. C., Quental, T. B., Roubik, D. W., Imperatriz-Fonseca, V. L. *et al.*, 2010 A molecular phylogeny of the stingless bee genus *Melipona* (Hymenoptera: Apidae). Mol Phylogenet Evol 56**:** 519-525.

Reed, R., 2000 Mechanisms of fidelity in pre-mRNA splicing. Curr Opin Cell Biol 12**:** 340-345.

Rizzo, M. A., Springer, G. H., Granada, B. & Piston, D. W., 2004 An improved cyan fluorescent protein variant useful for FRET. Nat Biotechnol 22**:** 445-449.

Robertson, H. M., & Wanner, K. W., 2006 The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. Genome Res 16**:** 1395-1403.

Robertson, H. M., Gadau, J. & Wanner, K. W., 2010 The insect chemoreceptor superfamily of the parasitoid jewel wasp *Nasonia vitripennis*. Insect Mol Biol 19 Suppl 1**:** 121-136.

Robin, G. C., Russell, R. J., Cutler, D. J. & Oakeshott, J. G., 2000 The evolution of an alpha-esterase pseudogene inactivated in the *Drosophila melanogaster* lineage. Mol Biol Evol 17**:** 563-575.

Rooney, A. P., Piontkivska, H. & Nei, M., 2002 Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. Mol Biol Evol 19**:** 68-75.

Ruiz, M. F., Eirin-Lopez, J. M., Stefani, R. N., Perondini, A. L., Selivon, D. *et al.*, 2007a The gene *doublesex* of *Anastrepha* fruit flies (Diptera, Tephritidae) and its evolution in insects. Dev Genes Evol 217**:** 725-731.

Ruiz, M. F., Milano, A., Salvemini, M., Eirin-Lopez, J. M., Perondini, A. L. *et al.*, 2007b The gene *transformer* of *Anastrepha* fruit flies (Diptera, tephritidae) and its evolution in insects. PLoS One 2**:** e1239.

Ryner, L. C., & Baker, B. S., 1991 Regulation of *doublesex* pre-mRNA processing occurs by 3'-splice site activation. Genes Dev 5**:** 2071-2085.

Saccone, G., Salvemini, M., Pane, A. & Polito, L. C., 2008 Masculinization of XX *Drosophila* transgenic flies expressing the *Ceratitis capitata* DoublesexM isoform. Int J Dev Biol 52**:** 1051-1057.

Salvemini, M., Robertson, M., Aronson, B., Atkinson, P., Polito, L. C. *et al.*, 2009 *Ceratitis capitata transformer-2* gene is required to establish and maintain the autoregulation of *Cctra*, the master gene for female sex determination. Int J Dev Biol 53**:** 109-120.

Sanchez, L., & Nothiger, R., 1982 Clonal analysis of *Sex-lethal*, a gene needed for female sexual development in *Drosophila melanogaster*. Rouxs Arch Dev Biol 191**:** 211-214.

Sanchez, L., 2008 Sex-determining mechanisms in insects. Int J Dev Biol 52**:** 837-856.

Sarno, F., Ruiz, M. F., Eirin-Lopez, J. M., Perondini, A. L., Selivon, D. *et al.*, 2010 The gene *transformer-2* of *Anastrepha* fruit flies (Diptera, Tephritidae) and its evolution in insects. BMC Evol Biol 10**:** 140.

Schetelig, M. F., Milano, A., Saccone, G. & Handler, A. M., 2012 Male only progeny in *Anastrepha suspensa* by RNAi-induced sex reversion of chromosomal females. Insect Biochem Mol Biol 42**:** 51-57.

Schmieder, S., Colinet, D. & Poirie, M., 2012 Tracing back the nascence of a new sex-determination pathway to the ancestor of bees and ants. Nat Commun 3**:** 895.

Schneider, C. A., Rasband, W. S. & Eliceiri, K. W., 2012 NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9**:** 671-675.

Schulte, C., Leboulle, G., Otte, M., Grunewald, B., Gehne, N. *et al.*, 2013 Honey bee promoter sequences for targeted gene expression. Insect Mol Biol 22**:** 399-410.

Schutt, C., & Nothiger, R., 2000 Structure, function and evolution of sex-determining systems in Dipteran insects. Development 127**:** 667-677.

Sciabica, K. S., & Hertel, K. J., 2006 The splicing regulators Tra and Tra2 are unusually potent activators of pre-mRNA splicing. Nucleic Acids Res 34**:** 6612-6620.

Sekido, R., & Lovell-Badge, R., 2008 Sex determination involves synergistic action of SRY and SF1 on a specific *Sox9* enhancer. Nature 453**:** 930-934.

Sekido, R., & Lovell-Badge, R., 2009 Sex determination and SRY: down to a wink and a nudge? Trends Genet 25**:** 19-29.

Selden, R. F., Howie, K. B., Rowe, M. E., Goodman, H. M. & Moore, D. D., 1986 Human growth hormone as a reporter gene in regulation studies employing transient gene expression. Mol Cell Biol 6**:** 3173-3179.

Semple, C., & Wolfe, K. H., 1999 Gene duplication and gene conversion in the *Caenorhabditis elegans* genome. J Mol Evol 48**:** 555-564.

Shen, H., & Green, M. R., 2006 RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. Genes Dev 20**:** 1755-1765.

Shukla, J. N., & Palli, S. R., 2012 Sex determination in beetles: production of all male progeny by parental RNAi knockdown of *transformer*. Sci Rep 2**:** 602.

Shukla, J. N., & Palli, S. R., 2013 *Tribolium castaneum* Transformer-2 regulates sex determination and development in both males and females. Insect Biochem Mol Biol 43: 1125-1132.

Sidow, A., 1996 Gen(om)e duplications in the evolution of early vertebrates. Curr Opin Genet Dev 6**:** 715-722.

Smith, J. M., 1992 Analyzing the mosaic structure of genes. J Mol Evol 34**:** 126-129.

Stevens, S. W., Ryan, D. E., Ge, H. Y., Moore, R. E., Young, M. K. *et al.*, 2002 Composition and functional characterization of the yeast spliceosomal penta-snRNP. Mol Cell 9**:** 31-44.

Su, C., & Nei, M., 2001 Evolutionary dynamics of the T-cell receptor VB gene family as inferred from the human and mouse genomic sequences. Mol Biol Evol 18**:** 503-513.

Suzuki, M. G., Funaguma, S., Kanda, T., Tamura, T. & Shimada, T., 2003 Analysis of the biological functions of a *doublesex* homologue in *Bombyx mori*. Dev Genes Evol 213**:** 345-354.

Tajima, F., 1993 Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135**:** 599-607.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. *et al.*, 2011 MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28**:** 2731-2739.

Theilmann, D. A., & Stewart, S., 1992 Molecular analysis of the trans-activating *IE-2* gene of *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus. Virology 187**:** 84-96.

Tian, M., & Maniatis, T., 1993 A splicing enhancer complex controls alternative splicing of *doublesex* pre-mRNA. Cell 74**:** 105-114.

Tian, M., & Maniatis, T., 1994 A splicing enhancer exhibits both constitutive and regulated activities. Genes Dev 8**:** 1703-1712.

Traut, W., Niimi, T., Ikeo, K. & Sahara, K., 2006 Phylogeny of the sex-determining gene *Sex-lethal* in insects. Genome 49**:** 254-262.

Turunen, J. J., Niemela, E. H., Verma, B. & Frilander, M. J., 2013 The significant other: splicing by the minor spliceosome. Wiley Interdiscip Rev RNA 4**:** 61-76.

Valcarcel, J., Singh, R., Zamore, P. D. & Green, M. R., 1993 The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of *transformer* pre-mRNA. Nature 362**:** 171-175.

Verhulst, E. C., Beukeboom, L. W. & van de Zande, L., 2010 Maternal control of haplodiploid sex determination in the wasp *Nasonia*. Science 328**:** 620-623.

Voelker, R. B., Erkelenz, S., Reynoso, V., Schaal, H. & Berglund, J. A., 2012 Frequent gain and loss of intronic splicing regulatory elements during the evolution of vertebrates. Genome Biol Evol 4**:** 659-674.

Walsh, J. B., 1995 How often do duplicated genes evolve new functions? Genetics 139**:** 421-428.

Watterson, G. A., 1983 On the time for gene silencing at duplicate Loci. Genetics 105**:** 745-766.

Weidtkamp-Peters, S., Felekyan, S., Bleckmann, A., Simon, R., Becker, W. *et al.*, 2009 Multiparameter fluorescence image spectroscopy to study molecular interactions. Photochem Photobiol Sci 8**:** 470-480.

Widengren, J., Kudryavtsev, V., Antonik, M., Berger, S., Gerken, M. *et al.*, 2006 Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. Anal Chem 78**:** 2039-2050.

Wilhelm, D., Palmer, S. & Koopman, P., 2007 Sex determination and gonadal development in mammals. Physiol Rev 87**:** 1-28.

Wilkins, A. S., 1995 Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. Bioessays 17**:** 71-77.

Williams, T. M., & Carroll, S. B., 2009 Genetic and molecular insights into the development and evolution of sexual dimorphism. Nat Rev Genet 10**:** 797-804.

Wilson, E. O., & Holldobler, B., 2005 The rise of the ants: a phylogenetic and ecological explanation. Proc Natl Acad Sci U S A 102**:** 7411-7414.

Woyke, J., 1963 Drone larvae from fertilized eggs of the honeybee. J Apic Res 2**:** 19-24.

Zhou, Z., Licklider, L. J., Gygi, S. P. & Reed, R., 2002 Comprehensive proteomic analysis of the human spliceosome. Nature 419**:** 182-185.

Danksagung

Ich möchte mich herzlich bedanken bei:

Prof. Dr. Martin Beye für die Betreuung dieser Arbeit, die vielen hilfreichen Diskussionen und die Unterstützung bei der experimentellen Planung

Prof. Dr. Laura Rose für die Übernahme des Korreferats

meinen Eltern Aggi und Wolfgang Schmitt für die unermüdliche Unterstützung in allen Phasen meiner Arbeit

Dr. Marianne Otte für das Einweihen in die hohe Kunst der Laborarbeit sowie viele konstruktive Diskussionen und Ratschläge

Dr. Tanja Gempe für Tipps & Tricks bei Larvenaufzucht und Sektion, viele konstruktive Diskussionen und Gespräche sowie den ein oder anderen Kaffee

Eva Theilenberg und Marion Müller-Borg für ihre Unterstützung im Labor und bei den Bienen

Dr. Jessica Hilbig, Vasco Koch, Dr. Inga Nissen und Dr. Christina Schulte fürs Korrekturlesen

Sabrina Koschack für ihre Unterstützung in einer kritischen Phase meiner Arbeit

den Bienen dafür, dass sie mich (fast) nie gestochen haben