



MOLEKULARE EVOLUTION VON GENEN DER KOMPLEMENTÄREN GESCHLECHTSBESTIMMUNGSKASKADE BEI *APIS*- UND *BOMBUS*-ARTEN

Inauguraldissertation

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

vorgelegt von

Sarah Tatjana Lechner
aus Reutlingen

Düsseldorf, August 2011

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

Korreferent: Prof. Dr. Laura Rose

Tag der mündlichen Prüfung: 19.09.2011

NOTHING IN BIOLOGY
MAKES SENSE EXCEPT IN
THE LIGHT OF EVOLUTION

THEODOSIUS DOBZHANSKY

Inhaltsverzeichnis

1	Einleitung	1
1.1	Das Prinzip der komplementären Geschlechtsdetermination am Beispiel der Honigbiene <i>Apis mellifera</i>	2
1.2	Die Geschlechtsdetermination und Paarungsbiologie in der Honigbiene <i>Apis mellifera</i>	3
1.3	Die Evolution von <i>complementary sex determiner</i> , dem initialen Faktor der Geschlechtsdetermination in <i>A. mellifera</i>	4
1.4	Das <i>csd</i> -Paralog <i>feminizer</i> in der Honigbiene	6
1.5	Die Geschlechtsdetermination und Paarungsbiologie der Erdhummel <i>Bombus terrestris</i>	8
1.6	Fragestellung und Ziele der Arbeit	8
2	Material und Methoden	11
2.1	Materialien	11
2.1.1	Enzyme	11
2.1.2	Kitsysteme	11
2.1.3	Chemikalien	11
2.1.4	Oligonukleotide	12
2.1.5	Verwendete Proben	14
2.1.6	Verwendete Nukleotidsequenzen	15
2.1.7	Verwendete Programme	15
2.2	Methoden	17
2.2.1	Isolierung von genomischer DNA	17
2.2.2	Isolierung von RNA	18
2.2.3	cDNA-Synthese	19
2.2.4	Polymerase-Kettenreaktion (PCR)	21
2.2.5	Rapid Amplification of cDNA Ends (RACE)	21
2.2.6	Sequenzierung von genomischen <i>csd</i> -Allelfragmenten	21
2.2.7	Schematische Darstellung der Sequenzierung verschiedener mRNA-Sequenzen	22
2.2.7.1	Schematische Darstellung der Sequenzierung der <i>csd</i> -mRNA-Sequenzen aus <i>A. florea</i>	22
2.2.7.2	Schematische Darstellung der Sequenzierung der <i>feminizer</i> -mRNA-Sequenzen aus <i>B. lapidarius</i> und <i>B. hypnorum</i>	23

2.2.7.3	Schematische Darstellung der Sequenzierung der <i>feminizer</i> -mRNA aus <i>E. hemichlorea</i>	24
2.2.7.4	Schematische Darstellung der Sequenzierung der <i>twin of feminizer</i> -mRNA aus <i>B. terrestris</i>	24
2.2.8	Computeranalysen	25
2.2.9	Berechnung von Koaleszenzzeit und effektiver Langzeitpopulationsgröße	26
3	Ergebnisse	27
3.1	Das multiallelische Gen <i>complementary sex determiner</i>	27
3.1.1	Analyse der Sequenzvariation des multiallelischen Gens <i>complementary sex determiner</i>	27
3.1.1.1	Analyse der Nukleotiddiversität in den <i>csd</i> -Sequenzen der Honigbienenpopulation	29
3.1.1.2	Analyse der genetischen Differenzierung der <i>csd</i> -Sequenzen zwischen den verschiedenen <i>A. mellifera</i> -Subpopulationen	35
3.1.2	Struktur und Evolution der hypervariablen Region von <i>csd</i>	37
3.1.2.1	Die Länge der hypervariablen Region korreliert nicht mit dem Alter der <i>csd</i> -Allele	37
3.1.2.2	Die hypervariable Region evolviert unabhängig von den angrenzenden Genregionen	39
3.1.3	Berechnung der in der <i>A. mellifera</i> -Population repräsentierten Zahl an <i>csd</i> -Allelen	41
3.1.4	Neue Berechnung der Koaleszenzzeit der <i>csd</i> -Allele	43
3.2	Das Gen <i>complementary sex determiner</i> existiert auch in der asiatischen Zwerghonigbiene <i>Apis florea</i>	45
3.3	Die Nukleotidsequenzevolution des Gens <i>feminizer</i> in verschiedenen Hymenopteren	57
3.3.1	Vergleich der Sequenzen des Gens <i>feminizer</i> in verschiedenen Hymenopteren	57
3.3.1.1	Sequenzierung der <i>fem</i> -mRNA aus verschiedenen Hymenopteren	57
3.3.1.2	Analyse der <i>fem</i> -Sequenzen verschiedener Hymenopteren	59
3.3.1.3	Sequenzanalyse des männchenspezifischen Exons von <i>fem</i>	61

3.3.2	Die Evolution von <i>feminizer</i> unter Einfluss seines Schwestergens <i>complementary sex determiner</i>	63
3.3.2.1	Csd hat einen funktionellen Einfluss auf die Sequenzevolution seines Paralogs <i>fem</i>	63
3.3.2.2	Die Evolution von <i>fem</i> wird durch die Nachbarschaft zu <i>csd</i> beeinflusst	65
3.4	Die <i>fem</i> -Duplikation <i>twin of feminizer</i> in der Erdhummel <i>Bombus terrestris</i>	67
3.4.1	Das Gen <i>twin of feminizer</i> ist eine Duplikation von <i>feminizer</i>	67
3.4.2	Das Gen <i>twin of feminizer</i> ist in der <i>Bombus</i> -Linie entstanden	71
3.4.3	Das Gen <i>tof</i> ist ein potentieller allelischer Faktor für die komplementäre Geschlechtsdetermination	74
4	Diskussion	75
4.1	Das multiallelische Gen <i>complementary sex determiner</i> in <i>Apis mellifera</i>	75
4.1.1	Das unter balancierender Selektion stehende Gen <i>csd</i> zeigt kaum genetische Differenzierung zwischen verschiedenen Subpopulationen	75
4.1.2	Die hypervariable Region des <i>csd</i> -Gens evolviert schnell und durch Sequenzduplikation	77
4.1.3	Die Zahl der <i>csd</i> -Allele in den <i>A. mellifera</i> -Populationen ist um das vier- bis sechsfache höher und die Koaleszenzzeit niedriger als bislang postuliert	78
4.2	Das Gen <i>complementary sex determiner</i> der asiatischen Zwerghonigbiene <i>A. florea</i> zeigt geringe allelische Diversität	82
4.3	Das Gen <i>feminizer</i> in <i>Apis</i> und anderen Hymenopteren	85
4.3.1	Das Gen <i>fem</i> ist in der Gruppe der corbiculaten Apidae konserviert	85
4.3.2	Die Sequenzevolution des Gens <i>fem</i> wird in der <i>Apis</i> -Linie durch sein Paralog <i>csd</i> beeinflusst	86
4.4	Das Gen <i>twin of feminizer</i> ist eine <i>fem</i> -Duplikation in der Hummel	88
5	Zusammenfassung	93
5.1	Summary in English	94
6	Anhang	95
6.1	HVR-Sequenzen funktionell verschiedener <i>csd</i> -Allele	95
6.1.1	Genealogien ausgewählter <i>A. mellifera</i> - <i>csd</i> -Sequenzen	99
6.2	Sequenzen des Gens <i>csd</i> verschiedener <i>Apis</i> -Arten	101
6.2.1	Sequenz des <i>csd</i> -Gens aus <i>Apis florea</i>	101
6.2.2	Alignment der Proteinsequenz von Csd verschiedener <i>Apis</i> -Arten	102

6.3	Sequenzen des <i>feminizer</i> -Gens verschiedener Hymenopteren	103
6.3.1	Sequenz des Gens <i>fem</i> aus der asiatischen Zwerghonigbiene <i>Apis florea</i>	103
6.3.2	Sequenz des Gens <i>fem</i> aus der Baumhummel <i>Bombus hypnorum</i> . .	104
6.3.3	Sequenz des Gens <i>fem</i> aus der Steinhummel <i>Bombus lapidarius</i> . .	105
6.3.4	Sequenz des Gens <i>fem</i> aus der nordamerikanischen Hummel <i>Bom-</i> <i>bus impatiens</i>	106
6.3.5	Sequenz des Gens <i>fem</i> aus der Erdhummel <i>Bombus terrestris</i>	107
6.3.6	Sequenz des Gens <i>fem</i> aus der Prachtbiene <i>Euglossa hemichlora</i> . .	108
6.3.7	Sequenz des Gens <i>fem</i> aus der parasitischen Wespe <i>Nasonia vitripennis</i>	109
6.3.8	Alignment der Fem-Sequenzen verschiedener Apidae	110
6.4	Sequenzen des Gens <i>tof</i> aus <i>Bombus terrestris</i> und <i>B. impatiens</i>	113
6.4.1	Sequenz der mRNA des <i>twin of feminizer</i> (<i>tof</i>) aus <i>B. terrestris</i> . .	113
6.4.2	Sequenz des Gens <i>tof</i> aus <i>Bombus impatiens</i>	114
6.4.3	Alignment von <i>Tof</i> und <i>Fem</i> aus <i>B. terrestris</i>	115
6.4.4	Alignment von <i>Tof</i> aus <i>B. terrestris</i> und <i>B. impatiens</i>	116

1 Einleitung

Die Geschlechtsdifferenzierung ist ein essentieller Prozess in der Ontogenese aller zweigeschlechtlicher Lebewesen. Die Entscheidung über die männliche oder weibliche Geschlechtsentwicklung erfolgt jedoch nicht durch einen konservierten Mechanismus, sondern durch ein breites Spektrum verschiedener Determinationsfaktoren, welche sich im Laufe der Evolution entwickelt haben. In den meisten Fällen erfolgt die Geschlechtsbestimmung durch einen genetischen Faktor. In manchen Tierarten haben sich aber auch Systeme evolviert, bei denen Umwelteinflüsse als Determinationsfaktoren wirken, beispielsweise die Temperatur bei der Geschlechtsbestimmung in Reptilien (Bull 1980). Genetische Determinanten reichen von allelischer Variation eines einzelnen Lokus bis hin zu Geschlechtschromosomen (Graves 2008). Ein gut untersuchtes Beispiel für ein geschlechtschromosomengekoppeltes System ist das auf dem Y-Chromosom der Säugetiere lokalisierte *sry*-Gen (Graves 1995). Andere auf Geschlechtschromosomen basierende Systeme finden sich z.B. bei Vögeln und einigen Reptilien (Schartl 2004) sowie der Taufliege *Drosophila melanogaster* und dem Fadenwurm *Caenorhabditis elegans* (Erickson und Quintero 2007, Cline und Meyer 1996). Ein Beispiel für Geschlechtsbestimmung durch allelische Diversität findet sich z.B. in der Honigbiene *Apis mellifera* in Form der komplementären Geschlechtsdetermination („complementary sex determination“, CSD; Beye et al. 2003). Wie bei allen Hymenopteren und insgesamt etwa 20 % aller Tierarten (Bull 1983) sind bei der Honigbiene Männchen haploid, während Weibchen diploid sind. Haplo-Diploidie resultiert häufig aus Arrhenotokie, d.h. Weibchen entstehen aus befruchteten Eiern, Männchen entwickeln sich dagegen aus unbefruchteten Eiern (Petrunkewitsch 1901). Auch bei der Arrhenotokie wird die Entscheidung über das Geschlecht nicht durch ein einheitliches System getroffen, sondern es haben sich eine Reihe verschiedener genetischer, zytoplasmatischer und verhaltensbasierender Mechanismen entwickelt (Heimpel und de Boer 2008). Ein Beispiel eines genetischen Mechanismus ist die schon genannte komplementäre Geschlechtsdetermination (Whiting 1933, 1943, Cook 1993), welche unter anderem in der Honigbiene *Apis mellifera* (1.1) und der Erdhummel *Bombus terrestris* (1.5) vorkommt. Auch bei CSD zeigt sich eine Vielfalt an Geschlechtsdeterminationsmechanismen, die im Laufe der Evolution entstanden sind. So wird selbst in den eng verwandten Gattungen *Apis* und *Bombus* das Geschlecht durch unterschiedliche initiale Faktoren determiniert. In *Bombus terrestris* ist der CSD-Faktor noch nicht bekannt, in *Apis mellifera* handelt es sich um ein in der *Apis*-Linie neu aus einer Genduplikation entstandenes Gen, *complementary sex determinier* (*csd*; Beye et al. 2003, Hasselmann et al. 2008a; 1.3).

Während sich im Laufe der Evolution vielfältige Faktoren als initiales Signal der Geschlechtsbestimmung entwickelt haben, gibt es Hinweise darauf, dass die darunter liegende Kaskade zwischen den Arten konserviert ist. So konnte gezeigt werden, dass die orthologen Gene *fem* und *transformer* (*tra*) und das von ihnen aktivierte Gen *doublesex* (*dsx*) zwischen Hymenopteren und Dipteren konserviert sind und möglicherweise schon über 300. Mio Jahre existieren (Gempe und Beye 2010).

1.1 Das Prinzip der komplementären Geschlechtsdetermination am Beispiel der Honigbiene *Apis mellifera*

Die komplementäre Geschlechtsdetermination findet sich in mehr als 60 Hymenopterenarten aus mindestens vier verschiedenen Familien (Heimpel und de Boer 2008). Sie beruht auf dem Prinzip eines allelischen Faktors, der im heterozygoten Zustand den weiblichen Entwicklungsweg initiiert (Whiting 1933).

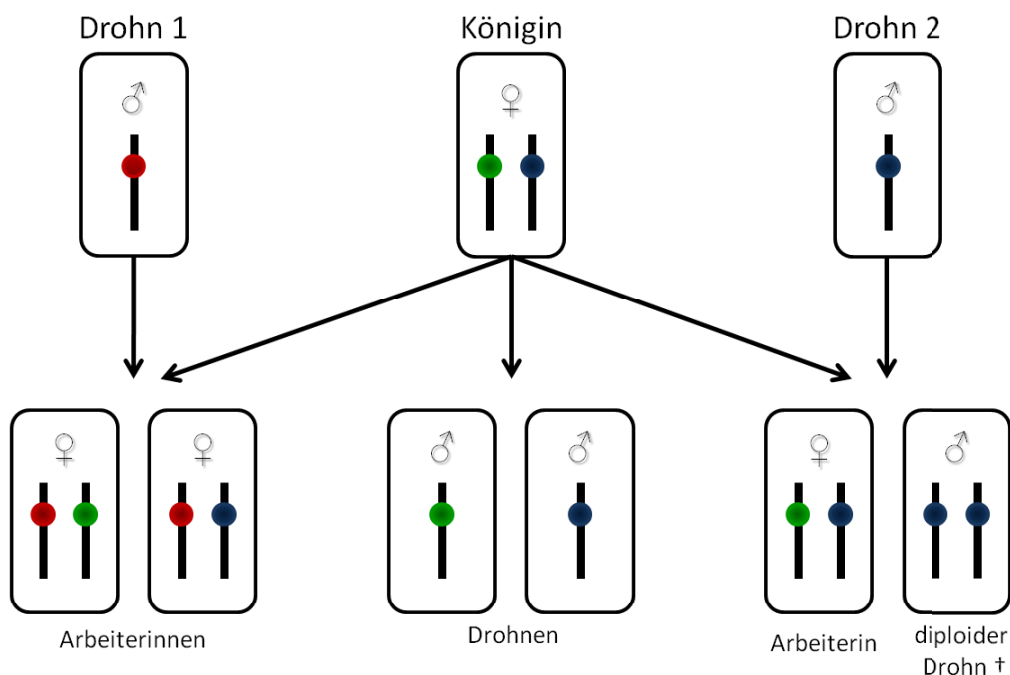


Abbildung 1: Prinzip der komplementären Geschlechtsdetermination in *A. mellifera*: Aus unbefruchteten Eiern, die nur ein Allel tragen, entwickeln sich haploide Drohnen. Aus befruchteten Eiern, die zwei verschiedene Allele tragen (durch zwei verschieden farbige Punkte dargestellt), entwickeln sich Arbeiterinnen oder Königinnen. Tragen Königin und Drohn das gleiche Allel, entstehen aus befruchteten Eiern diploide Drohnen, welche jedoch im frühen Larvenstadium von Arbeiterinnen aufgefressen werden.

Im homo- oder hemizygoten Zustand ist der Faktor nicht aktiv und der voreingestellte, männchenspezifische Entwicklungsweg wird eingeschlagen (Gempe et al. 2009). In der Honigbiene *A. mellifera* handelt es sich bei diesem Faktor um einen einzelnen Locus (Beye et al. 2003). In Abbildung 1 ist das Prinzip der auf einem einzelnen Locus basierenden, komplementären Geschlechtsdetermination dargestellt.

1.2 Die Geschlechtsdetermination und Paarungsbiologie in der Honigbiene *Apis mellifera*

In der Honigbiene *A. mellifera* ist der initiale Faktor das multiallelische Gen *complementary sex determiner* (*csd*), welches im heterozygoten Zustand zur Aktivierung des weiblichen Entwicklungswegs führt (Beye 2004). Es reguliert das geschlechtsspezifische Spleißen von *feminizer* (*fem*), dem zweiten Gen der Geschlechtsdeterminationskaskade (Hasselman et al. 2008a). So entwickeln sich aus diploiden, für *csd* heterozygoten Eiern Weibchen und aus haploiden Eiern, in denen nur ein *csd*-Allel vorliegt, Männchen. Trägt ein befruchtetes Ei zwei funktionell identische *csd*-Allele, so führt dies zur Entwicklung eines diploiden Männchens. Da die Drohnen der Honigbiene keine meiotische Teilung durchführen können, würden diploide Drohnen triploide, infertile Nachkommen erzeugen (Cook 1993). Bei Honigbienen werden diploide Drohnen jedoch kurz nach dem Schlupf von Arbeiterinnen aufgefressen (Woyke 1963). Durch Inzucht kann es so zum Verlust eines Teils der Brut kommen (Lückenbrut, Woyke 1962). In natürlichen Völkern kommt dies selten vor, da in *A. mellifera*-Populationen eine große Anzahl funktionell verschiedener Allele segregieren (Adams et al. 1977, Hasselman und Beye 2004), so dass die Wahrscheinlichkeit eines Zusammentreffens zweier funktionell identischer Allele dadurch verringert wird. Zur Paarung fliegen Honigbienenköniginnen zu einem so genannten Drohnensammelplatz („drone congregation area“, DCA). Dort sammeln sich Drohnen aus bis zu 240 Völkern (Baudry et al. 1998). Dies ermöglicht einen Austausch der in den verschiedenen Völkern repräsentierten Allele. Am DCA paart sich eine Honigbienenkönigin mit mehreren Drohnen (ca. 14, Palmer und Oldroyd 2000), so dass sich auch im Falle einer Paarung mit einem das gleiche *csd*-Allel tragenden Drohn nur ein kleiner Teil der Brut zu diploiden Männchen entwickelt (Adams et al. 1977).

1.3 Die Evolution von *complementary sex determiner*, dem initialen Faktor der Geschlechtsdetermination in *A. mellifera*

Das Gen *complementary sex determiner* ist in der *Apis*-Linie neu aus einer Duplikation des Gens *feminizer* entstanden (Hasselmann et al. 2008a, Abb. 2). Im Gegensatz zu seinem konservierten Paralog *fem* ist die *csd*-Sequenz variabel und es existieren eine große Anzahl funktionell verschiedener Allele in der Honigbienenpopulation. Berechnungen gehen von 14 (Hasselmann et al. 2008b) bis 19 Allelen (Adams et al. 1977) aus. Balancierende Selektion (negativ-frequenzabhängige Selektion) bewirkt die Erhaltung dieser

Allelzahl über einen langen Zeitraum. Homozygotie am *csd*-Lokus ist indirekt letal, da diploide Drohnen nach dem Schlupf gefressen werden. Dies führt dazu, dass häufige Allele einen selektiven Nachteil haben, da sie öfter homozygot auftreten (Zayed und Packer 2005). Dementsprechend haben seltene Allele einen selektiven Vorteil. Dies führt dazu, dass häufige Allele in ihrer Frequenz abnehmen, während die Frequenz seltener Allele ansteigt und sie deshalb weniger häufig durch genetische Drift verloren gehen als dies für Allele eines neutralen Lokus erwartet werden würde. Ihre Koaleszenzzeit ist hierdurch deutlich länger als die neutraler Allele (Takahata 1990). Da neu entstandene Allele ebenfalls einen selektiven Vorteil haben, führt die balancierende Selektion zu einer häufigen Fixierung von Polymorphismen, die für eine neue Allelspezifität kodieren. So entstehen immer wieder neue Allele, die den Verlust durch genetische Drift ausgleichen. Steht ein Gen unter Selektion, so hat dies auch Einfluss auf benachbarte Regionen. Der so genannte Hitchhiking-Effekt bewirkt, dass bei der Fixierung eines vorteilhaften Polymorphismus auch gekoppelte, meist neutrale Polymorphismen fixiert werden können (Fay und Wu 2000). In *csd* ist die Zahl der Polymorphismen im genomweiten Vergleich um das 10-13fache erhöht (Hasselmann und Beye 2006). Besonders hoch ist die Variation in den Exons 6 und 7, welche deshalb als das Ziel der balancierenden Selektion und potentielle Spezifizierungsdomäne (PSD) angenommen werden (Hasselmann et al. 2008b). In diesem Bereich befindet sich eine arginin-serinreiche (RS)

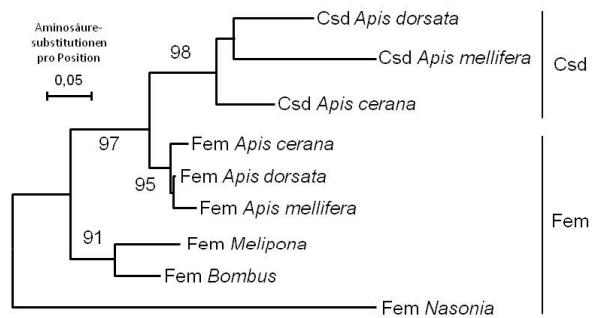


Abbildung 2: Genealogie der *Apis-csd*-Sequenzen sowie *fem*-Sequenzen aus *Apis*, *M. compressipes*, *B. terrestris* und *N. vitripennis*, aus Hasselmann et al. (2008a)

Domäne sowie die hypervariable Region (HVR). Die HVR ist in *csd* im Verlauf der Genduplikation entstanden (Abb. 3). Es handelt sich dabei um einen stark in seiner Länge variierenden Bereich, der hauptsächlich aus einer unterschiedlichen Anzahl an Wiederholungen eines N_1-5Y -Motivs besteht. Hasselmann et al. (2008b) postulierten, dass dieser Bereich entscheidend zur Allelspezifität beiträgt. Von der Sequenzvariation in der ebenfalls zur PSD gehörenden RS-Domäne wird angenommen, dass es sich dabei zum Teil um funktionell relevante Polymorphismen handelt (Hasselmann et al. 2008b).

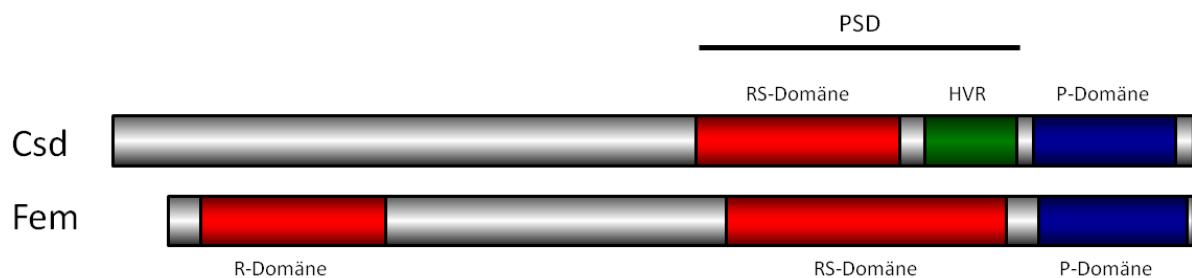


Abbildung 3: Schematische Darstellung der paralogen Proteine Fem und Csd. Beide Proteine weisen eine arginin-serinreiche (RS) und eine prolinreiche (P) Domäne auf. In Fem findet sich zudem eine N-terminale argininreiche (R) Domäne und in Csd die hypervariable Region (HVR). PSD steht für potentielle Spezifizierungsdomäne. Nach Hasselmann et al. 2008a und Hasselmann et al. 2008b.

Hasselmann et al. (2008a) postulieren, dass das Gen *complementary sex determiner* nur in der Gattung der Honigbienen existiert. Von den zwölf *Apis*-Arten (Lo et al. 2010) konnte es bisher in der westlichen Honigbiene *A. mellifera* sowie der östlichen Honigbiene *A. cerana* und der Riesenhonigbiene *A. dorsata* nachgewiesen werden (Hasselmann et al. 2008a). Nicht gefunden wurde es in der stachellosen Biene *Melipona compressipes* und der Erdhummel *Bombus terrestris*, welche Vertreter der am nächsten mit der Honigbiene verwandten Gattungen sind. Zudem gruppieren sich die *Apis-csd*-Sequenzen in einer Genealogie zu den *Apis-fem*-Sequenzen und separieren sich deutlich von *fem* aus anderen Hymenopterenarten. Daraus wird abgeleitet, dass *csd* erst in der *Apis*-Linie entstanden ist. Ob *csd* neben den drei oben genannten *Apis*-Arten auch in den anderen neun Arten dieser Gattung existiert, ist nicht bekannt. Hasselmann et al. (2008a) postulieren, dass *csd* nach der Trennung von *Apis*, *Melipona* und *Bombus*, jedoch vor Aufspaltung der *Apis*-Arten entstanden ist (Abb. 4).

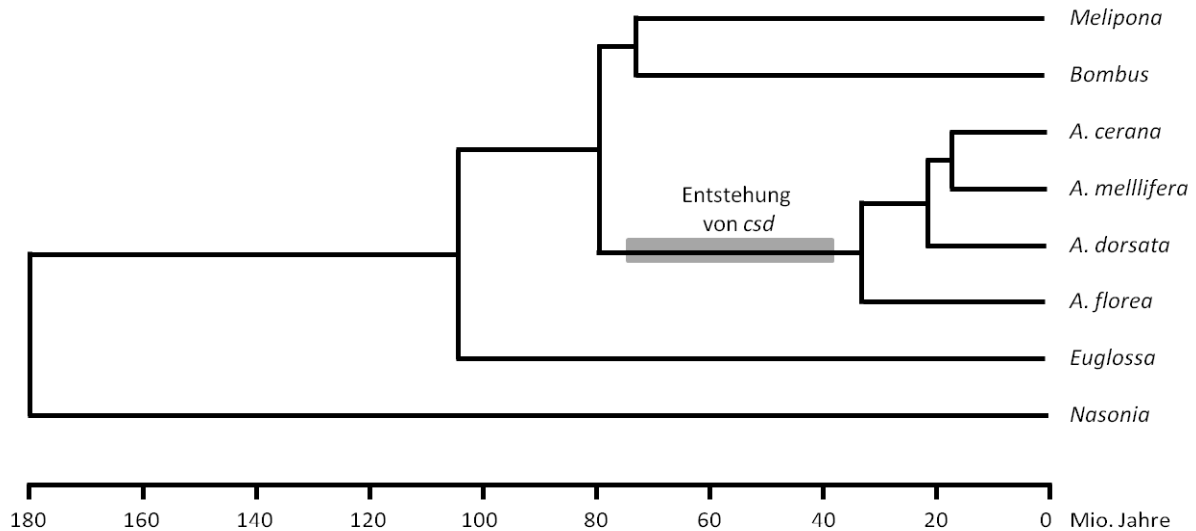


Abbildung 4: Genealogie der Hymenopterengattungen *Apis*, *Bombus*, *Melipona*, *Euglossa* (Apidae) und *Nasonia* (Pteromalidae) nach Ramírez et al. (2010) und Hasselmann et al. (2008a). Der postulierte Entstehungszeitraum des Gens *csd* in der *Apis*-Linie ist angegeben (grauer Balken).

Die asiatische Zwerghonigbiene *Apis florea* hat eine basale Stellung in der *Apis*-Gattung (Alexander 1991). Sofern sich die *fem*-Duplikation, durch welche *csd* entstand, vor der Trennung von *A. florea* von den anderen *Apis*-Arten ereignete, ist die Existenz von *csd* auch hier anzunehmen. Der Nachweis des Gens *csd* in der basalen Art *A. florea* würde es ermöglichen, die Frage nach dem Entstehungszeitpunkt des Gens genauer zu beantworten.

1.4 Das *csd*-Paralog *feminizer* in der Honigbiene

Das Gen *feminizer* ist der zweite Faktor in der Geschlechtsdeterminationskaskade der Honigbiene *Apis mellifera*. Es existieren zwei verschiedene, geschlechtsspezifische Spleißformen des Gens. In Männchen kommt es durch ein Stoppkodon im 3'-terminalen Bereich des Exons 3 zu einem Translationsabbruch, so dass kein funktionelles Protein gebildet werden kann (Hasselmann et al. 2008a). In diesem Fall wird der voreingestellte, männchenspezifische Entwicklungsweg eingeschlagen. Sind jedoch in einem diploiden Ei zwei verschiedene Allele des *csd*-Gens und somit ein aktives Signal vorhanden, initiiert dies das alternative Spleißen der *fem*-mRNA (Abb. 5). Es wird der 3'-terminale Bereich des Exons 3, welcher ein Stoppkodon enthält, sowie die Exons 4 und 5 herausgespleißt.

Hierdurch entsteht ein längerer Leserahmen, der in ein aktives Protein translatiert werden kann.

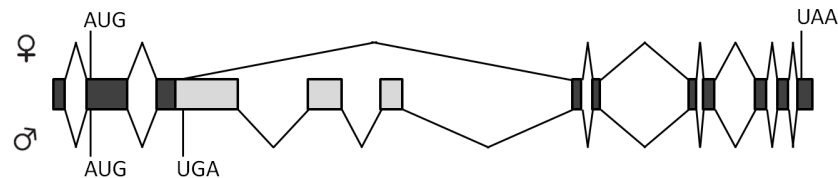


Abbildung 5: Spleißmuster des Gens *fem* in der Honigbiene *A. mellifera* (nach Hasselmann et al. 2008a). Die männchenspezifische Spleißform (unten) enthält sämtliche Exons, bei Weibchen (oben) werden der 3'-Bereich des Exons 3 sowie die Exons 4 und 5 herausgespleißt.

Fem vermittelt das weibchenspezifische Spleißen der mRNA des nächsten Gens der Kaskade, *doublesex* (*Am-dsx*). Hierdurch wirkt *fem* als Schaltergen für die Entscheidung zwischen dem männlichen und dem weiblichen Entwicklungsweg (Gempe et al. 2009). Bei *fem* findet sich im Gegensatz zu *csd* keine allelische Variation und seine Existenz wird in allen Hymenopteren vermutet. Es konnte bereits in der stachellosen Biene *M. compressipes*, sowie der Erdhummel *B. terrestris* nachgewiesen werden (Hasselmann et al. 2008a). Die Paraloge *csd* und *fem* befinden sich beide im geschlechtsdeterminierenden Locus (sex determination locus, SDL; Abb. 6). Der geschlechtsdeterminierende Locus ist definiert als die genomische Region, die in Arbeiterinnen immer heterozygot und in diploiden Drohnen immer homozygot vorliegt. Es ist eine etwa 47 kb große Region, die mindestens fünf Gene enthält (Hasselmann et al. 2008b, Gempe et al. 2009).

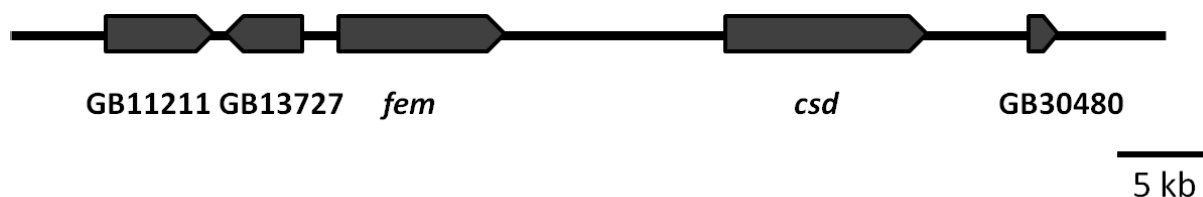


Abbildung 6: Der geschlechtsdeterminierende Locus der Honigbiene *Apis mellifera* enthält fünf Gene (nach Hasselmann et al. 2008a). Er wird definiert als die in Weibchen immer heterozygot vorliegende Region (Hasselmann und Beye 2006).

Das Gen *fem* und das in *Apis* neu entstandene *csd*-Gen liegen etwa 12 kb voneinander getrennt. Dies bedeutet für *fem* die direkte Nachbarschaft zu einem Gen unter

balancierender Selektion, welche sich durch den Hitchhiking-Effekt möglicherweise auch auf das konservierte *fem* auswirkt. Ob dies sowie die strukturelle Ähnlichkeit zu *csd* möglicherweise die Evolution von *fem* in *Apis* beeinflusst, ist bislang nicht bekannt.

1.5 Die Geschlechtsdetermination und Paarungsbiologie der Erdhummel *Bombus terrestris*

In der Erdhummel *Bombus terrestris* wird das Geschlecht ebenfalls durch komplementäre Geschlechtsdetermination bestimmt (Cook und Crozier 1995). Wie auch bei *A. mellifera* treten bei *B. terrestris* vermehrt diploide Drohnen als Folge von Inzuchtkreuzungen auf (Duchateau et al. 1994). Dies ist ein klares Indiz für die Existenz eines multiallelischen Geschlechtsdeterminationsfaktors, welcher im heterozygoten Zustand die Entwicklung von Weibchen initiiert. Gadau et al. (2001) konnten zeigen, dass in *B. terrestris* ebenfalls ein einzelner geschlechtsbestimmender Locus existiert. Der SDL konnte auf eine ca. 24 cM große Region eingegrenzt werden. Im Gegensatz zu Honigbienenköniginnen paaren sich Hummelköniginnen nur mit einem Drohn (Röseler 1973). Auch werden diploide Drohnen in Erdhummelvölkern nicht getötet und entwickeln sich zu normalen, wenn auch kleineren, adulten Tieren (Duchateau et al. 1994). Sie haben kleinere Testikel und eine reduzierte Fitness, sind aber fertil und können sich erfolgreich paaren, wobei jedoch triploide Nachkommen entstehen (Duchateau und Mariën 1995).

Bisher gibt es keine Hinweise auf die Identität des geschlechtsdeterminierenden Faktors in der Hummel. Wie in Abschnitt 1.3 beschrieben, berechneten Hasselmann et al. (2008a) die Entstehung von *csd* auf einen Zeitpunkt nach der Aufspaltung der *Apis*- und *Bombus*-Linie. Dies spricht gegen die Existenz des *csd*-Gens in *Bombus* und lässt die Frage nach dem CSD-Faktor bislang ungeklärt.

1.6 Fragestellung und Ziele der Arbeit

In der Gattung der Honigbienen (*Apis*) ist durch eine Genduplikation ein neuer initialer Faktor für die Geschlechtsbestimmung entstanden. Das Gen *complementary sex determiner* (*csd*) aktiviert im heterozygoten Zustand den weiblichen Entwicklungsweg. Die hierfür notwendige allelische Diversität ist durch positive Selektion auf Grund des Heterozygotenvorteils entstanden und wird durch balancierende Selektion in der Population erhalten. Balancierende Selektion fördert zudem die Migration der Allele.

Ein Ziel dieser Arbeit ist es, anhand von *csd*-Sequenzen, welche aus verschiedenen weltweiten *A. mellifera*-Stichproben gewonnen wurden, zu untersuchen, in wie weit sich diese Aspekte auf die genetische Diversität am *csd*-Lokus in einzelnen Lokalitäten auswirken. Eine bisher noch offene Frage ist, ob sich bei diesem unter balancierender Selektion stehenden Gen genetische Differenzierung zwischen natürlichen Subpopulationen findet. Um diese Frage zu beantworten, wird die Nukleotidvariation von *csd*-Sequenzen kenianischer *A. mellifera*-Arbeiterinnen analysiert, welche aus weitgehend vom Menschen unbeeinflussten Völkern stammen.

Die hypervariable Region des *csd*-Gens unterscheidet sich in ihrer Struktur deutlich von der restlichen Gensequenz. Eine Hypothese ist, dass sie zudem anders evolviert als die benachbarten Regionen. Diese Frage soll hier geklärt werden. Es wird davon ausgegangen, dass sie entscheidend zur Allelspezifität beiträgt. Durch den Vergleich funktionell verschiedener Allelpaaire sollen Rückschlüsse darauf gezogen werden, wie viele Unterschiede für eine neue Allelspezifität notwendig sind. Mit dieser Information soll die Zahl der in der gesamten Honigbienenpopulation segregierenden Allele neu bestimmt werden. Dies ermöglicht zudem eine genauere Berechnung der Koaleszenzzeit der *csd*-Allele.

Es ist bislang nicht bekannt, ob *csd* in allen *Apis*-Arten existiert, doch wird vermutet, dass es vor der Aufspaltung der *Apis*-Arten entstanden ist. Es konnte in *A. mellifera*, *A. cerana* und *A. dorsata* nachgewiesen werden. In dieser Arbeit soll versucht werden, das Gen in der phylogenetisch basal stehenden Art *A. florea* nachzuweisen. Hieraus ergeben sich weitere Informationen über den Zeitpunkt der Genduplikation. Sofern *csd* in *A. florea* existiert, können neue Erkenntnisse über die im *csd*-Gen verschiedener *Apis*-Arten vorhandene Variation gewonnen werden.

Das Gen *feminizer*, der zweite Faktor der Geschlechtsdeterminationskaskade in *Apis* ist im Gegensatz zu seinem Paralog *csd* hoch konserviert. Ein Ziel dieser Arbeit ist die Detektion dieses Gens in der Prachtbiene *Euglossa hemichlora* und der parasitischen Wespe *Nasonia vitripennis*. Dies ermöglicht es, den Vergleich von *fem* verschiedener Arten auf zwei mit *Apis* entfernter verwandte Spezies auszuweiten.

Durch die Entstehung von *csd* befindet sich *fem* im Genom der Honigbiene in direkter Nachbarschaft zu einem Gen unter balancierender Selektion. Zudem interagiert es mit einem Gen, das eine starke strukturelle Ähnlichkeit zu *fem* aufweist. Eine offene Frage ist, ob diese beiden Faktoren die Evolution von *fem* in *Apis* beeinflusst haben. Um dies zu überprüfen, werden Sequenzdaten von *fem* verschiedener Hymenopteren mit Daten anderer Gene im geschlechtsbestimmenden Locus verglichen. Ob durch die

Genduplikation und die Entstehung eines Gens mit einer neuen Funktion (*csd*) die Evolution seines Vorläufergens (*fem*) beeinflusst wird, soll durch den Vergleich von Nukleotidsubstitutionen in *fem* Sequenzen von *Apis*- und Nicht-*Apis*-Arten beantwortet werden.

In der Hummel wird ebenfalls ein allelischer Faktor für die komplementäre Geschlechtsdetermination vermutet, der stellvertretend in der Erdhummel *Bombus terrestris* nachgewiesen werden soll. Eine Hypothese ist, dass es sich bei dem allelischen Faktor ebenfalls um eine *fem*-Duplikation handelt. Das Genom der Erdhummel wird auf solche potentiellen Kandidatengene hin überprüft. Mittels Sequenzanalysen kann dann die Frage geklärt werden, ob es sich dabei um den gesuchten allelischen Faktor handeln könnte.

2 Material und Methoden

2.1 Materialien

2.1.1 Enzyme

Liste der verwendeten Enzyme. Verwendung nach Herstellerangaben, soweit unter 2.2 nicht anders angegeben.

- DNase I, RNase free, Fermentas GmbH, St. Leon Rot, Deutschland
- DNA Polymerase I, Fermentas GmbH, St. Leon Rot, Deutschland
- GoTaq® DNA Polymerase, Promega Corporation, Madison, USA
- Phusion™ High-Fidelity DNA Polymerase, Finnzymes OY, Espoo, Finnland
- Restriktionsenzyme, Fermentas GmbH, St. Leon Rot, Deutschland
- Ribonuclease H, Fermentas GmbH, St. Leon Rot, Deutschland
- T4-DNA-Ligase, Fermentas GmbH, St. Leon Rot, Deutschland

2.1.2 Kitsysteme

Liste der verwendeten Kitsysteme. Verwendung nach Herstellerangaben.

- Dynabeads® Oligo (dT)₂₅, Invitrogen Dynal AS, Oslo, Norwegen
- E.Z.N.A. Cycle Pure Kit, Omega Bio-Tek, Inc., Norcross, USA
- FirstChoice™ RLM-RACE Kit, Ambion, Austin, USA
- peqGOLD Gel Extraction Kit, PEQLAB Biotechnologie GMBH, Erlangen, Deutschland
- pGEM®-T Vector, Promega Corporation, Madison, USA
- GenElute™ Plasmid Miniprep Kit, Sigma Aldrich, St. Louis, USA
- RevertAid® First Strand cDNA Synthesis Kit, Fermentas GmbH, St. Leon Rot, Deutschland

2.1.3 Chemikalien

Chemikalien wurden in der Qualität „pro analysis“ verwendet und hauptsächlich von den folgenden Firmen bezogen: Becton Dickinson, Heidelberg; Merck, Darmstadt; Roth, Karlsruhe; Sigma-Aldrich, Steinheim; BioAmerica Inc., Miami, USA.

2.1.4 Oligonukleotide

Die verwendeten Oligonukleotide wurden von Eurofins MWG Operon, Ebersberg, Deutschland synthetisiert. Sämtliche Oligonukleotidsequenzen sind in 5'-3'-Richtung dargestellt. Enthält ein Oligonukleotid eine Restriktionsschnittstelle so ist diese durch kleine Buchstaben angezeigt. Das entsprechende Restriktionsenzym ist dem Oligonukleotidnamen angefügt.

Tabelle 1: Liste der verwendeten Oligonukleotide. Die markierten Oligonukleotide wurden freundlicherweise von folgenden Personen zur Verfügung gestellt: * Christina Schulte, ** Martin Hasselmann, *** Tanja Gempe.

Name	Oligonukleotidsequenz
<u><i>Apis mellifera csd</i></u>	
genoRfw **	AGACRATATGAAAAATTACACAATGA
genoRfw ApaI	GATCgggcccAGACRATATGAAAAATTACACAATGA
Cons csd rev **	TCATCTCATWTTTCATTATTCAAT
Cons csd rev SalI	GATCgtcgacTCATCTCATWTTTCATTATTCAAT
S-39	TATAATGAAAAAGAAAAATTTTGTAGAAG
S-39 ApaI	GATCgggcccTATAATGAAAAAGAAAAATTTTGTAGAAG
S-40	ACTATGTGCATCAATATAAATTC
S-40 SalI	GATCgtcgacACTATGTGCATCAATATAAATTC
<u><i>Apis mellifera fem</i></u>	
csd II siRNA 714for ***	AGATTCAAGACATGAAGACAG
GenoRNAi IIRT rev ***	TATCTGGAGGAATAAATCGTGGTG
<u><i>Apis mellifera GB13727</i></u>	
GB13727 fw5' **	CGTTTGGTGCCATTTTGGATG
GB13727 Aug07 rev **	CTATCATCAGGTAGCACTGATG
S-119	GATAAATGTCCACCAAATTTTACAG
S-121	CTATATTAACATATGATTGTG
<u><i>Apis mellifera GB11211</i></u>	
S-57	CGTTACTTTTTCGATCATAGTGACGATG
S-58	CCATTGTCGGCGTACACCTATTGG

Apis florea csd

S-136	GTATCATCCATTCTTCGCCATCAGC
S-137	CTTGTTGTATCATCCATTCTTCG
S-138	CATACATCRTCTCGTTATTCCAGC
S-139	GGACCAATATACCTAAATCTTGG
S-140	CCAAGATTTAGGTATATTGGTCCA
S-141	CGTCCACCTCTAAATCCTAG
S-143	CAGAAGAACGATTACGACGGAGACGCG
S-144	GTATGATCTTTGCTCTCGTTCTCTC
S-156	CTCCCGTTCTTCTTTTGATTATCACATT
S-158	GTCAAGGCTGAGTAATAGTATTAA
S-159	CATACATCGTCTCGTTATTCCAGGG
S-166	CGGTTTCTCTAAGCATATAGGTGA
S-176 (ApaI)	GATCggtcgacCATTTGACCCGCTAGTTGTCCAATCTCG
S-177 (SalI)	GATCggggcccGTTGCAGTAGAGATAGAAATAGAGG

Euglossa hemichlora fem

Euglfw3	TCTACCTAAAAGAAGATTATCTAG
Euglrev4	ATGATACAAAAAGAGCGGGAACGA
S-2	GGTAAACCCTTTTGACGAGCACGTTG
S-3	CCGGGTAAAACAACAACCTGGTGAAC
S-4	GATGGTACAACACCCTTATTTAAAGG
S-9	CTTCAACGAGATATTGTCAATCCTG
S-12	GAGGAACATCGTACAGTTGTAGCAAT
S-15	CGATCTCTTTCCCTATCATTTCTG
S-16	TTTTCTATTCCTTTGTAAATTTCTG
S-23	ATGAAACGAAATATATCAAGTCATTCTG
S-179	GTGAGACTGGTTCGCGCGCAAGAACAG

Nasonia vitripennis fem

S-18	ATGAGGAGAAGAAGTTCCTTAGATG
S-19	TTCAATACATCTGTCCAAATCGAGG

Bombus terrestris fem

#10 3race28I *	GCCAGGTATGAGGATACAAAATATG
#11 3race28II *	GGAAGAAATCACAGTAGAGGTAGAAG
#16 5RACEfemI *	CAGCGGTAAACCTTTCTCAAGAGCACG
#23 femBom_rev2 *	CTATCGGCACCATTCCCTTCCAAC
#36 FemBom_st2 *	AACTACCCTAACTTCCTCGATTTC
#61 fem_gen_fw *	GAAATACGTCGTGCTCTTGAGAAAG
#62 fem_gen_rev *	CTGTCTCTGTTATCATATTTTGAATCC
S-184	GCGACAACCTGATGATCTCCAACGAG
S-185	CTCGTTGGAGATCATCAGTTGTCTGC
S-186	CATCGCATTATATCGAGCAGATTCCTG

S-187	CAGGAATCTGCTCGATATAATGCGATG
S-190	CACGCAGATGCTCTTCCTCTGCAGTG
S-191	CGAAGCCCCAGTCCTCGACATAGATCG
S-201	CTACCATACTGCATCTATGAATGCATTTCTAATGCG
S-202	GGAATTGTATGGAAGAGTGTTACG
S-203	GGTCTGTCACAACTGCTAATACTGC
S-204	CCTTTAGTACTTCTTCACTGGATGACTG

Bombus terrestris tof

S-65	GAGACAGAGATCATAGTAGAGG
S-66	ACATTGGCCTAAATTAGGATTTGG
S-67	GCACATATTACATTATAATCGG
S-72	CCAAAGATTGTCACAACAGCAACAAA
S-73	CACTGCCTTCATCACATTATATTGAACCC
S-91	CTGTATAAAACCAGTGTATATTCTG
S-101	CACGCTTACGTCGTAATCTTCTCTCG
S-102	GAGAGCACGGCGTATTTTCATATTCTAAG
S-106	GCTAGTAATGTAAGCATTATTCGTC

2.1.5 Verwendete Proben

Für die Analysen standen Individuen verschiedener Honigbienen- und anderer Hymenopterenarten zur Verfügung. Diese stammten weitgehend aus wildlebenden Völkern, mit Ausnahme der *Apis mellifera*-Drohnen, welche aus verschiedenen domestizierten Völkern gesammelt wurden, sowie den *Nasonia vitripennis*-Individuen aus Zucht.

- *Apis mellifera*-Drohnen aus verschiedenen Völkern, zur Verfügung gestellt von:
 - Katherine Aronstein, US Department of Agriculture, Texas, USA
 - Christina Grozinger, North Carolina State University, North Carolina, USA
 - Kim Fondrik, Arizona State University, Arizona, USA
 - Ryszard Maleszka, Australian National University, Canberra, Australia
 - Guy Bloch, Hebrew University of Jerusalem, Jerusalem, Israel
- *Apis mellifera scutellata* und *monticola*-Arbeiterinnen aus verschiedenen Völkern in Kenia, von Caspar Schöning, Länderinstitut für Bienenkunde Hohen Neuendorf e.V., Hohen Neuendorf, Deutschland
- Afrikanisierte *Apis mellifera*-Drohnen und -Arbeiterinnen aus Brasilien, von Klaus Hartfelder, Faculdade de Medicina de Ribeirão Preto, Brasilien
- *Apis florea*-Arbeiterinnen aus Samut Songkhram, Thailand, von Jochen Pflugfelder, Agroscope Liebefeld-Posieux ALP, Bern, Schweiz

- *Bombus terrestris*-Drohnen aus verschiedenen Völkern, zur Verfügung gestellt von:
 - Andrew F.G. Bourke, School of Biological Sciences, University of East Anglia, Norwich, United Kingdom
 - Matthias Schindler, Institut für Nutzpflanzenwissenschaften und Ressourcenschutz (INRES), Universität Bonn, Deutschland
- *Bombus terrestris*-Völker von Syngenta Bioline, Bergen op Zoom, Niederlande
- *Euglossa hemichlora* gesammelt in Costa Rica, von Thomas Elz, Heinrich-Heine-Universität Düsseldorf, Deutschland
- *Nasonia vitripennis* aus Zucht von Joachim Ruther, FU Berlin / Uni Regensburg, Deutschland

2.1.6 Verwendete Nukleotidsequenzen

Für die Sequenzanalysen in dieser Arbeit fanden die folgenden in der NCBI-Datenbank veröffentlichten Sequenzen Verwendung:

- *Apis cerana csd*, NCBI-Nummern: EU100900.1, EU100901.1, EU100904.1, EU100908.1, EU100909.1, EU100910.1, EU100911.1, EU100914.1, EU100915.1, EU100916.1
- *Apis dorsata csd*, NCBI-Nummern: EU100917.1, EU100918.1, EU100919.1, EU100925.1, EU100926.1, EU100928.1, EU100931.1, EU100932.1, EU100934.1, EU100935.1,
- *Apis mellifera fem*, NCBI-Nummern: AY569715.1
- *Apis cerana fem*, NCBI-Nummern: EU100936.1
- *Apis dorsata fem*, NCBI-Nummern: EU100939.1
- *Bombus terrestris fem*, NCBI-Nummer: EU288185.1
- *Melipona compressipes fem*, NCBI-Nummer: EU139305.1

2.1.7 Verwendete Programme

Liste der verwendeten Programme. Sind Adressen von Internetseiten angegeben, so handelt sich dabei um Onlineprogramme.

- Arlequin 3.11 (Excoffier et al. 2005)
- BioEdit Sequence Alignment Editor (Hall 1999)
- Bn-Bs: Berechnung von Nukleotidsubstitutionen pro Position (Zhang et al. 1998)
- DnaSP v5 (Librado und Rozas 2009)

- EBI: Alinierung zweier Sequenzen, <http://www.ebi.ac.uk/Tools/emboss/align/>
- FGENES: Detektion von Leserahmen, <http://linux1.softberry.com/berry.phtml?topic=fgenes&group=programs&subgroup=gfind>
- GeneMark: Detektion von Leserahmen, <http://exon.gatech.edu/GeneMark/eukhmm.cgi>
- Guidance: Programm zur Überprüfung von Alingments, <http://guidance.tau.ac.il/>
- InterProScan: Vorhersage von Proteindomänen, <http://www.ebi.ac.uk/Tools/pfa/iprscan/>
- MEGA 5 (Kumar et al. 2008, Tamura et al. 2007)
- Jalview version 2 (Waterhouse et al. 2009)
- Mesquite 2.74 (Maddison und Maddison 2010)
- MrBayes 3.1 (<http://mrbayes.csit.fsu.edu/index.php>)
- NCBI: BLAST-Suche in Sequenzdatenbanken, <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
- PASW Statistics Version 18.0, Polar Engineering and Consulting.
- Recombination Detection Program v.3.41 RDP3 (Martin et al. 2005)
- SITES (Hey und Wakeley 1997)
- ProSite: Vorhersage von Proteindomänen, <http://expasy.org/prosite/>, <http://www.genome.jp/tools/motif/>

2.2 Methoden

Die angewandten molekularbiologischen Standardmethoden wie die Ligation von DNA-Fragmenten mit Hilfe der T4-Ligase, Restriktion von PCR-Amplifikaten mit Hilfe von Restriktionsenzymen, Gelelektrophorese, photometrische Mengenbestimmung von Nukleinsäuren und Bakterien sowie die Herstellung von chemisch kompetenten *E. coli*-Zellen werden in Sambrook et al. (1989) beschrieben und hier nicht weiter ausgeführt.

Kit-Systeme sind unter 2.1.2 aufgelistet und wurden nach Herstellerangaben verwendet.

2.2.1 Isolierung von genomischer DNA

Für die DNA-Isolation wurde von adulten Bienen lediglich der Thorax verwendet. In Ethanol konservierte Bienen wurden zuerst 1-2 Stunden in Wasser gewaschen.

Protokoll von Saghai-Marooft et al. (1984)

- Probe kurz in flüssigen Stickstoff tauchen, anschließend mit einem Pistill in einem Reaktionsgefäß zermörsern.
- 350 µl 2xCTAB zugeben und vollständig zermörsern
- Zugabe weiterer 350 µl 2xCTAB sowie 10 µl Proteinase K
- 1,5 bis 2 h bei 55 °C inkubieren
- Zugabe von 650 µl Chloroform:Isoamylalkohol (24:1), mehrmals invertieren (15-20 mal, nicht vortexen!)
- 15 min bei 13.000 rpm zentrifugieren
- 180 µl der oberen Phase in ein neues Reaktionsgefäß überführen
- Zugaben von 1 µl RNase A, Inkubation bei 37 °C für 30 min
- Zugabe von 650 µl Phenol:Chloroform:Isoamylalkohol (25:24:1), mehrmals invertieren
- 15 min bei 13.000 rpm zentrifugieren
- Obere Phase in ein neues Reaktionsgefäß überführen
- Zugabe von 650 µl Chloroform:Isoamylalkohol (24:1) und 15 min bei 13.000 rpm zentrifugieren
- Obere Phase in ein neues Reaktionsgefäß überführen (ca. 400 µl)
- Zugabe von 1000 µl kaltem 100 % Ethanol und 40 µl 3M Natriumacetat, 10-20 mal invertieren

- Zur Fällung der DNA über Nacht bei -20 °C inkubieren
- 30 min bei 4 °C und 13.000 rpm zentrifugieren
- Proben auf Eis stellen und Überstand abnehmen
- Mit 1000 µl kaltem 80 % Ethanol waschen
- 20 min bei 4 °C waschen
- Proben auf Eis stellen und Überstand vollständig abnehmen
- 3-5 min in einer Vakuumzentrifuge zentrifugieren
- Pellet in 50 µl H₂O über Nacht bei 4 °C resuspendieren

Lösungen:

2xCTAB-Puffer	0,1 M Tris-HCl (pH 8,0) 1,4 M NaCl 0,02 M EDTA 2% CTAB
---------------	---

2-Mercaptoethanol jedes Mal frisch zum 2xCTAB-Puffer
zugeben (2 µl pro 1 ml 2xCTAB)

Proteinase K	1 mg / µl
--------------	-----------

RNase A	10 mg / µl
---------	------------

2.2.2 Isolierung von RNA

TRIZOL™-Methode zur RNA-Isolierung

Für die RNA-Isolation wurde von adulten Bienen lediglich der Thorax verwendet. Larven wurden vollständig eingesetzt. Bei der Isolation von RNA aus Eiern und kleinen Larven (ca. L1- und L2-Stadium) wurden geringere Mengen der Lösungen eingesetzt, welche jeweils in Klammer angegeben sind. Die Proben wurden bei -70 °C aufbewahrt.

- Direkt nach dem Auftauen der Proben etwa 500 µl (50 µl) TRIZOL™ zugeben und mit Pistill zermörsern
- Zugabe von 100 µl (10 µl) Chloroform, Invertieren der Probe und Inkubation bei Raumtemperatur für 2-3 min

- Zentrifugation für 15 min bei 13000 rpm
- Obere Phase in ein neues Reaktionsgefäß überführen
- 300 µl (20 µl) Isopropanol zugeben, vorsichtiges Invertieren
- 10 min bei Raumtemperatur inkubieren
- Zentrifugation bei 4 °C für 10 min bei 13000 rpm
- Abnahme des Überstandes
- Waschen mit 500 µl (100 µl) 70 % Ethanol
- Zentrifugation bei 4 °C für 10 min bei 13000 rpm
- Trocknen des Pellets bei Raumtemperatur
- Resuspension des Pellets in 20-30 µl DEPC-H₂O

Lösungen:

TRIZOL™-Reagenz	0,8 M Guanidinthiocyanat
	0,4 M Ammoniumthiocyanat
	0,1 M Natriumacetat (pH 5)
	5 % Glycerol
	38% Phenol (z.B. Aqua Roti Phenol)
	DEPC-H ₂ O

2.2.3 cDNA-Synthese

Aufreinigung der RNA

Vor der cDNA-Synthese ist es notwendig die Gesamt-RNA entweder einer DNase-Behandlung zu unterziehen um die genomische DNA zu entfernen oder aber mit Hilfe von Dynabeads® die mRNA zu isolieren.

Die Isolation der mRNA durch Dynabeads® wurde wie im Protokoll des Kitsystems beschrieben durchgeführt. Die DNase-Behandlung wurde ausgeführt wie folgt.

DNase-Behandlung von Gesamt-RNA

- Zu 17 µl RNA 2 µl DNase-Puffer und 1 µl DNase gegeben
- 15 min bei 37 °C inkubieren

- Zugabe von 1 µl DNase
- 15 min bei 37 °C inkubieren
- Zugabe von 1 µl 25 mM EDTA
- 10 min bei 65 °C inkubieren zur Hitzeinaktivierung der DNase
- Zugabe von 22 µl Isopropanol, vorsichtiges Invertieren
- 10 min bei Raumtemperatur inkubieren
- Zentrifugation bei 4 °C für 10 min bei 13000 rpm
- Abnahme des Überstandes
- Waschen mit 100 µl 70 % Ethanol
- Zentrifugation bei 4 °C für 10 min bei 13000 rpm
- Trocknen des Pellets bei Raumtemperatur
- Resuspension des Pellets in 11 µl DEPC-H₂O für die anschließende cDNA-Synthese
- Photometrische Messung der RNA

cDNA-Erststrangsynthese

Die cDNA-Erststrang-Synthese wurde mit dem RevertAid® First Strand cDNA Synthesis Kit von Fermentas nach Herstellerangaben durchgeführt.

cDNA-Zweitstrangsynthese

Bei Bedarf wurde zudem eine cDNA-Zweitstrangsynthese durchgeführt. Hierfür wurde die DNA-Polymerase I von Fermentas verwendet, womit folgender Reaktionsmix angesetzt und für 2 h bei 16 °C inkubiert wurde.

- 20 µl Reaktionsmix aus der cDNA-Einzelstrangsynthese
- 10 µl 10x DNA-Polymerase I-Reaktionspuffer
- 4 µl DNA-Polymerase I (40 Units)
- 0,32 µl Ribonuclease H
- mit H₂O auf 100 µl auffüllen

Anschließend wurde die cDNA mit dem E.Z.N.A Cycle Pure Kit aufgereinigt.

2.2.4 Polymerase-Kettenreaktion (PCR)

Die Polymerase-Kettenreaktion wurde in den meisten Fällen mit einer selbst hergestellten Taq-Polymerase und nach dem Protokoll von Sambrook et al. (1989) durchgeführt. Für PCR-Produkte, die direkt sequenziert wurden, wurde die PhusionTM High Fidelity DNA Polymerase verwendet. Ebenfalls wurde sie verwendet, wenn PCR-Produkte mit Hilfe von Oligonukleotiden mit Restriktionsschnittstellen in einen Vektor ligiert wurden. Die Ligation in einen Vektor ermöglicht ein besseres Sequenzierungsergebnis. Für PCR-Produkte, die für die Sequenzierung in einen pGEM-T-Vektor ligiert wurden, wurde die GoTaq-Polymerase verwendet um die notwendigen Poly-A-Überhänge zu erzeugen. In beiden Fällen wurde die PCR wie vom Hersteller empfohlen durchgeführt.

2.2.5 Rapid Amplification of cDNA Ends (RACE)

Die RACE-Methode dient dazu bei cDNA-Fragmenten, bei denen lediglich ein Teil der Sequenz bekannt ist, die fehlenden 5'- und 3'-Bereiche zu amplifizieren. Hierfür werden bei der cDNA-Synthese Nukleotidfragmente spezifischer Sequenz als Adapter an das 5'- oder 3'-Ende der cDNA angefügt. Ein an diesen Adapter spezifisch bindendes Oligonukleotid ermöglicht es dann in Kombination mit einem im bekannten Sequenzbereich spezifisch bindenden Oligonukleotid den unbekannten Bereich zu amplifizieren. Üblicherweise erfolgt dies mit zwei nacheinander folgenden, verschachtelten PCRs, wobei bei der zweiten PCR die Oligonukleotide strangaufwärts von den in der ersten PCR verwendeten Oligonukleotiden liegen. Dieses Vorgehen erhöht die Spezifität der RACE-PCR. Die Synthese der RACE-Banken erfolgte mit dem FirstChoiceTM RLM-RACE Kit nach Herstellerangaben.

2.2.6 Sequenzierung von genomischen *csd*-Allelfragmenten

Sowohl aus Drohnen wie auch aus Arbeiterinnen wurde die genomische DNA wie in 2.2.1 beschrieben isoliert.

Sofern bei den Drohnen mehrere Tiere aus einem Volk zur Verfügung standen, wurde die DNA aus mindestens fünf Individuen isoliert. Mit *csd*-spezifischen Oligonukleotiden (siehe 2.1.4) wurde der Bereich von Exon 6-9, der auch die hypervariable Region enthält, amplifiziert. Auf Grund der Variation in den *csd*-Allelen war es notwendig zwei unterschiedliche Oligonukleotidpaare zu verwenden (genoRfw/Cons *csd* rev oder S-39/S-40). Um die beiden Königinnenallele zu detektieren wurde das PCR-Produkt der verschiedenen Individuen mit den Restriktionsenzymen XapI und SspI restringiert und auf einem SynergelTM aufgetrennt.

Im Falle der Arbeiterinnen war es notwendig das PCR-Produkt zuerst in einen Vektor zu

ligieren, um die beiden Allele unterscheiden zu können. Hierfür wurden die PCR-Produkte, die mit den Oligonukleotidpaaren genoRfw ApaI / Cons csd rev SalI oder S-39 ApaI / S-40 SalI amplifiziert wurden, sowie der Vektor (rezirkularisierter pGEM®-T-Vektor) mit den Enzymen ApaI und SalI restringiert und mit einer T4-Ligase ligiert. Die Unterscheidung der beiden Allele fand hier durch eine Restriktion der Amplifikate der Kolonie-PCR statt.

Zur Kontrolle wurden einige Allele mehrfach sequenziert um auszuschließen, dass Unterschiede zwischen Sequenzen auf Fehler durch die Polymerase bei der Amplifikation zurückzuführen sind.

2.2.7 Schematische Darstellung der Sequenzierung verschiedener mRNA-Sequenzen

2.2.7.1 Schematische Darstellung der Sequenzierung der *csd*-mRNA-Sequenzen aus *A. florea*

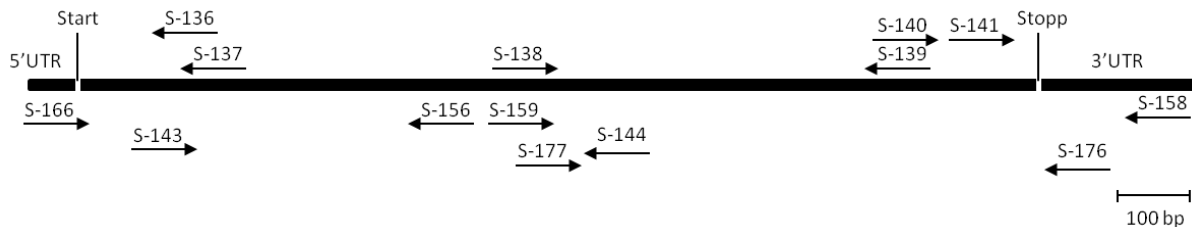


Abbildung 7: Schematische Darstellung der mRNA-Sequenz von *Af-csd* und der für ihre Sequenzierung verwendeten Oligonukleotide: 1. PCR mit den für die in *Af-fem* gefundene Variation spezifischen Oligonukleotiden S-138 und S-139, 5'-RACE mit S-137 und S-136, 3'-RACE mit S-140 und S-141. Anschließende Amplifikation des gesamten Bereichs in drei Teilstücken: S-166 + S-156, S-143 + S-144 und S-159 + S-158. Für die Amplifikation aus genomischer DNA wurden die Oligonukleotidpaare S-177 + S-176 sowie S-143 und S-156 verwendet. Die hier dargestellte mRNA-Sequenz ist 1608 bp lang. Die Oligonukleotidsequenzen finden sich in Tabelle 1.

2.2.7.2 Schematische Darstellung der Sequenzierung der *feminizer*-mRNA-Sequenzen aus *B. lapidarius* und *B. hypnorum*

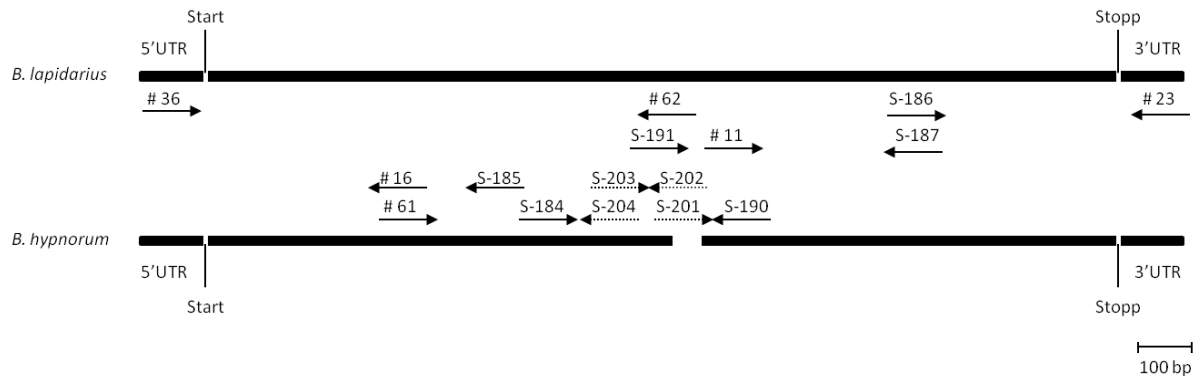


Abbildung 8: Schematische Darstellung der ca. 1,4 kb langen mRNA-Sequenzen (so weit bekannt) des Gens *feminizer* in *Bombus lapidarius* und *B. hypnorum* sowie der Bindungsstellen der für die Sequenzierung verwendeten Oligonukleotide. Auf Grund der cDNA-Qualität musste die gesamte Sequenz in mehreren Abschnitten sequenziert werden. Bei *B. lapidarius* wurde dies in drei PCRs mit den Oligonukleotidkombinationen #36 + #62, S-191 + S-187, S-186 + #23 durchgeführt. Bei *B. hypnorum* erfolgte eine zusätzliche Amplifikationen von genomischer DNA mit den Oligonukleotidkombinationen S-184 + S-204, S-203 + S-202 und S-201 + S-190. Mit S-201 + S-190 gelang keine spezifische Amplifikation. Der fehlende Sequenzbereich ist entsprechend als Lücke dargestellt. Die gepunkteten Pfeile stehen für im Intron bindende Oligonukleotide. Der 5'- und 3'-Bereich vom *B. hypnorum-fem* wurde jeweils in zwei PCRs mit folgenden Kombination amplifiziert: #36 + #16, #61 + S-185 und #11 + S-187, S-186 + #23. Start- und Stoppkodon sind als weiße Striche dargestellt. Die Oligonukleotidsequenzen finden sich in Tabelle 1.

2.2.7.3 Schematische Darstellung der Sequenzierung der *feminizer*-mRNA aus *E. hemichlorea*

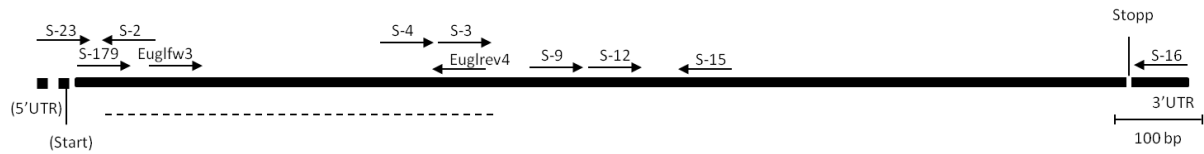


Abbildung 9: Schematische Darstellung der weiblichen Spleißform der *fem*-mRNA-Sequenz in *Euglossa hemichlora* und der Position der für ihre Sequenzierung verwendeten Oligonukleotide. Ausgehend von einem bereits sequenzierten Fragment (gestrichelte Linie), konnte der 3'-terminale Bereich in einer ersten RACE-PCR mit den Oligonukleotiden S-4 und S-3 und anschließend in einer weiteren RACE-PCR mit S-9 und S-12 amplifiziert werden. Mit den in *A. cerana* spezifisch am Startkodon bindenden Oligonukleotid S-23 und dem für *Eh-fem* spezifischen Oligonukleotid S-2 konnte ein Teil der 5'-terminalen Sequenz amplifiziert werden. Am Startkodon sind ca. 34 Nukleotide unbekannt, was durch den gestrichelten Bereich zu Beginn der Sequenz angedeutet wird. Mit den Kombinationen S-179 + S-15 und S-9 + S-16 wurde die gesamte bekannte Sequenz amplifiziert.

2.2.7.4 Schematische Darstellung der Sequenzierung der *twin of feminizer*-mRNA aus *B. terrestris*

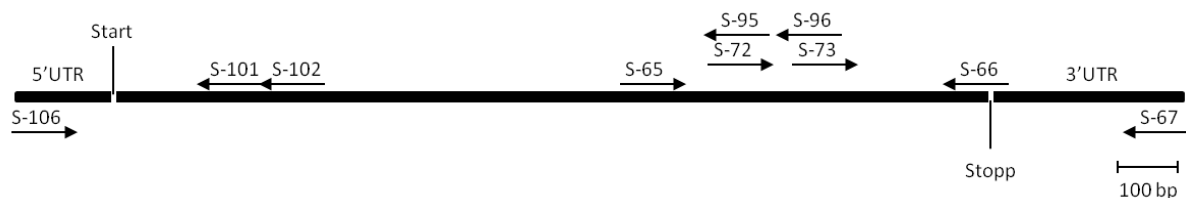


Abbildung 10: Schematische Darstellung der in der EST-Bank unter der Nummer *terrestris.fna.Contig17498* gefundene Sequenz und der für ihre vollständige Sequenzierung verwendeten Oligonukleotide. Das gefundene Fragment konnte mit den Oligonukleotiden S-65 und S-66 erneut amplifiziert werden. Die Sequenz wurde am 3'-Ende mit den Oligonukleotiden S-72 und S-73 vervollständigt. Zur Vervollständigung des 5'-terminalen Bereichs wurde eine erste RACE-PCR mit den Oligonukleotiden S-96 und S-95 durchgeführt, von der hier generierten Sequenz ausgehend eine weitere RACE-PCR mit S-102 und S-101. Mit den Oligonukleotiden S-106 und S-67 konnte die gesamte *tof*-Sequenz amplifiziert werden. Die Sequenz ist 1913 bp lang. Die Oligonukleotidsequenzen finden sich in Tabelle 1. Die Oligonukleotide S-65 und S-67 wurden des Weiteren verwendet um ein Fragment aus genomischer DNA zu amplifizieren (3.4.3).

2.2.8 Computeranalysen

Im folgenden Abschnitt sind die verschiedenen Programme und Parameter beschrieben, die zur Berechnung der Werte in dieser Arbeit verwendet wurde. Eine vollständige Liste aller genutzten Programme findet sich unter 2.1.7.

- **Sequenzalignments** wurden mit den Algorithmen CLUSTAL W (Thompson et al. 1994), MAFFT (Katoh et al. 2005) sowie PRANK (Löytynoja und Goldman 2008) durchgeführt. Bei allen Alignments aus mindestens vier Sequenzen wurde zudem mit Hilfe der Bewertung durch Guidance das optimalste Alignment ausgewählt (<http://guidance.tau.ac.il/>, Penn et al. 2010a, Penn et al. 2010b). Notwendige Korrekturen, wie das Positionieren einzelner Basen zu vollständigen Triplets, wurden per Hand durchgeführt.
- Die Berechnung der **Identität zwischen zwei Sequenzen** wurden mit den Algorithmus „needle“ (Needleman und Wunsch 1970 und Kruskal 1983) durch paarweisen Vergleich auf Nukleotid- oder Aminosäureebene auf der Seite <http://www.ebi.ac.uk/Tools/emboss/align/> durchgeführt.
- Die **Nukleotiddiversität** pro Position π wurde im Programm MEGA 5 (Tamura et al. 2007) mit der Nei-Gojobori-Methode (nach Nei und Gojobori 1986 und Nei und Kumar 2000) berechnet. Die folgenden Parameter wurden dafür verwendet: complete deletion, Jukes-Cantor-Modell für synonyme und nichtsynonyme Polymorphismen.
- Die **genetische Diversität** (d_n / d_s) wurde mit dem Programm MEGA berechnet. Paarweise Analyse aller Sequenzen nach Nei-Gojobori (Nei und Gojobori 1986 und Tamura et al. 2007). Parameter: complete deletion, Jukes-Cantor-Modell für synonyme und nichtsynonyme Polymorphismen.
- Die **Fst-Werte** wurden mit Hilfe von Arlequin 3.11 (Excoffier et al. 2005) nach Nei und Kumar (2000), Nei und Li (1979) und Tamura et al. (2007) berechnet. Die zugehörigen Signifikanzwerte beziehen sich auf die Nullhypothese, die von keiner Differenzierung zwischen den Populationen ausgeht.
- **Genealogien** wurden mit dem Programm MEGA 5 (Tamura et al. 2007) erstellt. Für Genealogien, die auf Basis von Aminosäure- oder Nukleotidunterschieden erstellt wurden, wurde die Maximum-Likelihood Methode verwendet. Genealogien die auf Basis von synonymen Nukleotidpolymorphismen erstellt wurden, wurden mit der Neighbor-Joining-Methode erstellt. Der Bootstrap-Test wurde mit 500 Wiederholungen durchgeführt (Felsenstein 1985). Die evolutionäre Distanz wurde bei Nukleotidssequenzen nach Jukes-Cantor-Modell (Jukes und Cantor 1969) und bei Aminosäuresequenzen nach dem JTT matrixbasierenden Modell (Jones et al. 1992) berechnet, Lücken wurden von der Analyse ausgenommen.

- **Statistische Tests**, die Erstellung von Regressionsgeraden sowie des Box-Whisker-Blots wurden mit dem Programm PASW durchgeführt.

2.2.9 Berechnung von Koaleszenzzeit und effektiver Langzeitpopulationsgröße

Die Zahl der pro synonymen Position akkumulierenden Polymorphismen (d_S) wurden durch den Vergleich der Gene *fem*, *csd*, *Elongation factor-alpha 1*, *Actin*, *Calmodulin-Kinase II* und *RNA-Polymerase II* auf 0,0900 zwischen *A. mellifera* und *A. cerana* und auf 0,0927 für *A. mellifera* und *A. dorsata* berechnet. Hieraus kann nach der Formel $d_S = 2g\mu$, unter der Annahme einer Generationszeit (g) von einem Jahr, die Mutationsrate (μ) auf $2,47 \times 10^{-9}$ - $1,97 \times 10^{-9}$ berechnet werden. Die Entstehungsrate neuer Allele u ergibt sich aus der Divergenzzeit der Allele und der Zahl nichtsynonymer Substitutionen zwischen den Sequenzen und beträgt 2×10^{-6} bis 2×10^{-7} . Nach der Formel

$$F = \sqrt{\frac{-3}{4N} \ln(u\sqrt{12\pi N})}$$

kann die effektive historische Populationsgröße bestimmt werden (Yokoyama und Nei 1979). F entspricht hierbei auch $1/n$, wobei n die Zahl der in der Population segregierenden Allele ist.

3 Ergebnisse

3.1 Das multiallelische Gen *complementary sex determiner*

3.1.1 Analyse der Sequenzvariation des multiallelischen Gens *complementary sex determiner*

Das Gen *complementary sex determiner* (*csd*), welches als allelischer Faktor die Geschlechtsbestimmung in der Honigbiene initiiert, zeigt einen im genomweiten Vergleich um das 10-13fach erhöhten Wert an Nukleotidpolymorphismen (Hasselman und Beye 2006). Die meiste Variation findet sich im Bereich der Exons 6 und 7, die für die arginin-serinreiche (RS) Domäne und die hypervariable Region (HVR) kodieren. Dies weist darauf hin, dass das Ziel der balancierenden Selektion dieser Bereich des Gens ist, weshalb er als potentielle Spezifizierungsdomäne (PSD) angesehen wird (Hasselman et al. 2008b, Abb. 11).

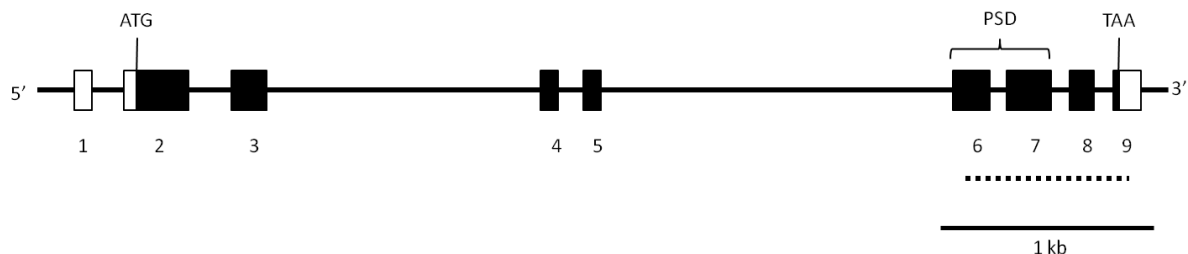


Abbildung 11: Schematische Darstellung der Exon-Intron-Struktur des Gens *csd* aus *Apis mellifera*. Die Exons sind als Kästchen dargestellt, wobei schwarz für die kodierenden Bereiche und weiß für die 5'- und 3'-UTR steht. Die Introns sind als Linie dargestellt. Der gestrichelte Balken markiert den für die hier beschriebenen Analysen amplifizierten und sequenzierten Bereich. PSD = potentielle Spezifizierungsdomäne. Nach Hasselman und Beye (2006) und Hasselman et al. (2008b).

Bei den bisherigen Studien wurde der Fokus auf die Analyse der Variation in den verschiedenen Regionen des *csd*-Gens (Hasselman und Beye 2006) sowie auf Vergleiche der *csd*-Sequenzen verschiedener *Apis*-Arten gelegt (Hasselman et al. 2008b). In der vorliegenden Arbeit soll nun die Sequenzvariation zwischen den Allelen innerhalb der

Apis mellifera-Population analysiert werden. Hierzu stehen Drohnen und Arbeiterinnen zur Verfügung, die aus verschiedenen *A. mellifera*-Völkern weltweit gesammelt wurden. Von der genomischen DNA dieser Individuen wurden die Exons 6-9 amplifiziert, wobei der Anfang von Exon 6 (ca. 100 bp) fehlt (Abb. 11, 2.2.6).

Bei den Stichproben aus den USA, Australien, Israel und Brasilien handelt es sich um Drohnen, die aus verschiedenen Völkern einer geographischen Region gesammelt wurden. Da diese Völker stark vom Menschen beeinflusst und keine natürlichen Subpopulationen sind, werden die Proben anhand der Regionen, in denen sie gesammelt wurden, zusammengefasst und im folgenden als Lokalitäten bezeichnet. Im Gegensatz dazu wurden die kenianischen *A. mellifera*-Arbeiterinnen aus sechs natürlichen Subpopulationen gesammelt, welche drei Berg- und drei Savannenregionen in Kenia besiedeln. Somit werden sie auch in dieser Arbeit als Subpopulationen bezeichnet.

Zur Identifikation beider Allele einer Arbeiterin wurden die amplifizierten Fragmente in einen Vektor ligiert und anschließend durch eine RFLP-Analyse unterschieden (2.2.6). Je ein Klon wurde für jedes der beiden Restriktionsmuster, durch die sich die beiden Allele einer Arbeiterin unterscheiden, ausgewählt und für die Sequenzierung verwendet. Da Drohnen haploid sind, konnte bei ihnen das *csd*-Fragment direkt sequenziert werden. Sie tragen jeweils eines der beiden Allele der Königin. Bei den Proben aus den USA, Australien und Brasilien standen mehrere Drohnen aus jedem Volk zur Verfügung, so dass meist beide Königinnenallele nachgewiesen werden konnten. Die Unterscheidung erfolgte ebenfalls durch eine RFLP-Analyse der *csd*-Amplifikate mehrerer Individuen.

Insgesamt wurden 193 *csd*-Fragmente sequenziert. 167 dieser Sequenzen unterschieden sich an mindestens einer Aminosäureposition. Für die Sequenzanalysen wurden doppelte, d.h. auf Aminosäureebene identische Sequenzen bei den jeweiligen untersuchten Gruppen aussortiert (Tab. 2).

Es ist bisher nicht bekannt, welche Aminosäureunterschiede funktionell relevant sind und für eine neue Allelspezifität kodieren (siehe auch 3.1.2.2). Im Folgenden wird deshalb von Sequenzen gesprochen, sofern nicht explizit verschiedene Allelspezifitäten gemeint sind.

Tabelle 2: Liste der verwendeten *Apis mellifera*-Stichproben. Angegeben ist der Sammelort, die Zahl der verwendeten Individuen, die Anzahl der amplifizierten und sequenzierten *csd*-Fragmente (Exon 6-9) sowie die Sequenzen, die sich an mindestens einer Aminosäureposition unterscheiden.

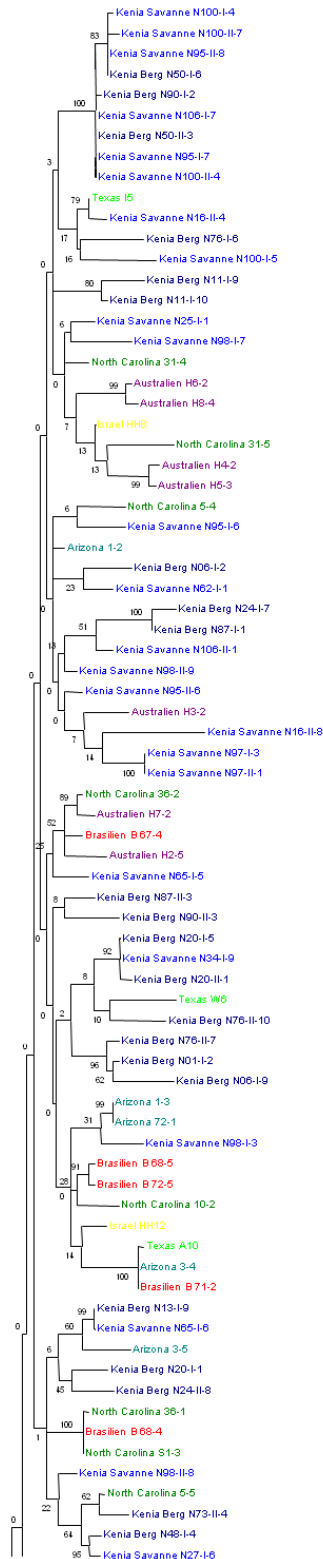
Unterart	Herkunft	Zahl der Individuen	Sequenzen pro Lokalität	Zahl der verschiedene Sequenzen
<i>A. m. ligustica</i>	Texas, USA	12 ♂	12	12
<i>A. m. ligustica</i>	North Carolina, USA	20 ♂	20	19
<i>A. m. afrikanisiert</i>	Arizona, USA	12 ♂	11	9
<i>A. m. ligustica</i>	Australien	17 ♂	17	16
<i>A. m. ligustica</i>	Israel	5 ♂	5	5
<i>A. m. afrikanisiert</i>	Brasilien	17 ♂ + 2 ♀	19	18
<i>A. m. monticola</i>	Kenia	60 ♀	54	49
<i>A. m. scutellata</i>	Kenia	60 ♀	55	51

3.1.1.1 Analyse der Nukleotiddiversität in den *csd*-Sequenzen der Honigbienenpopulation

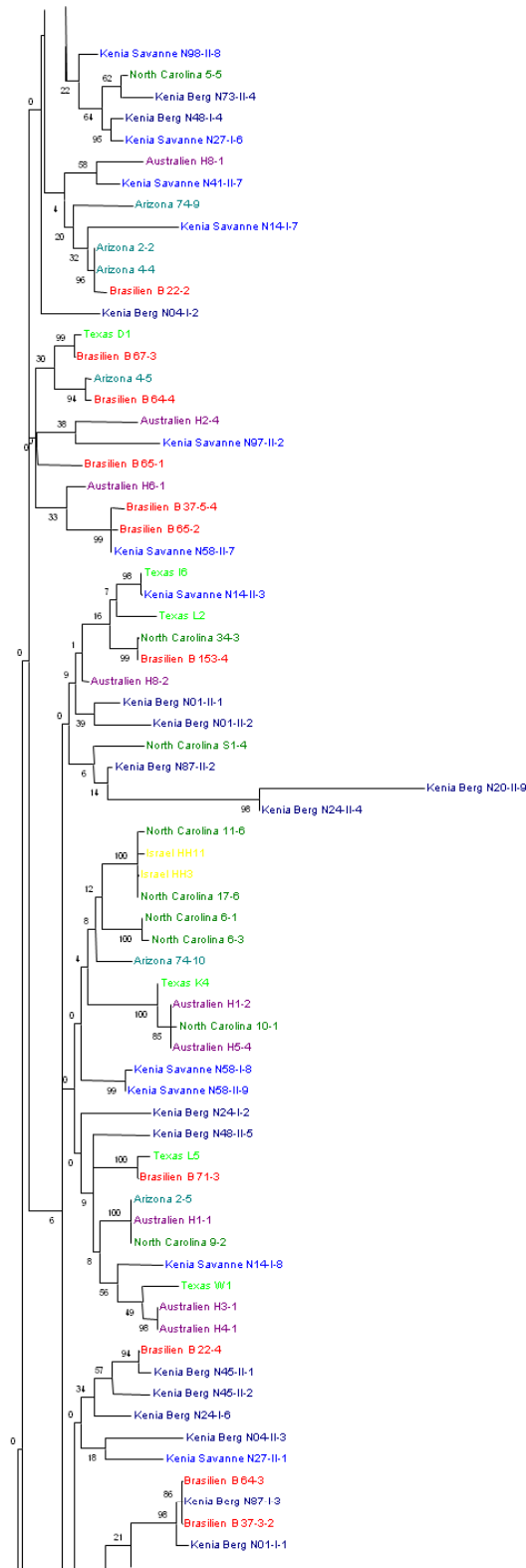
Die Paarungsbiologie der Honigbiene (Palmer und Oldroyd 2000) impliziert, dass sich *csd*-Allele in einer Population verbreiten können. Neu in eine Population migrierende Allele haben zudem auf Grund ihrer Seltenheit einen selektiven Vorteil und werden durch die balancierende Selektion erhalten. Die durchschnittliche Koaleszenzzeit, welche bisher auf mindestens 6 Mio. Jahre berechnet wurde (Hasselmann et al. 2008b, siehe auch 3.1.4), lässt annehmen, dass sich die Allele über die Zeit in der gesamten Population verteilt haben. Bei der Stichprobe aus Kenia wird deshalb erwartet, dass sich ein vergleichbares Maß an genetischer Diversität bei den *csd*-Sequenzen aller Subpopulationen zeigt. Die Honigbiene hat ihre natürliche Verbreitung in Afrika, Europa und Zentralasien (Whitfield et al. 2006). Von hier aus wurde sie ab dem 17. Jahrhundert nach Amerika und später nach Australien gebracht. Auf Grund der langen Koaleszenzzeit der *csd*-Allele ist davon auszugehen, dass sich in Amerika und Australien nur Allele aus der alten Welt finden und keine neuen entstanden sind. Bedingt durch den Heterozygotenvorteil und die lange Erhaltung der Allele durch die balancierende Selektion ist zu vermu-

ERGEBNISSE

ten, dass der durch den Export von Honigbienenenvölkern nach Amerika und Australien verursachte genetische Flaschenhals keine Auswirkungen auf die genetische Diversität hat.



ERGEBNISSE



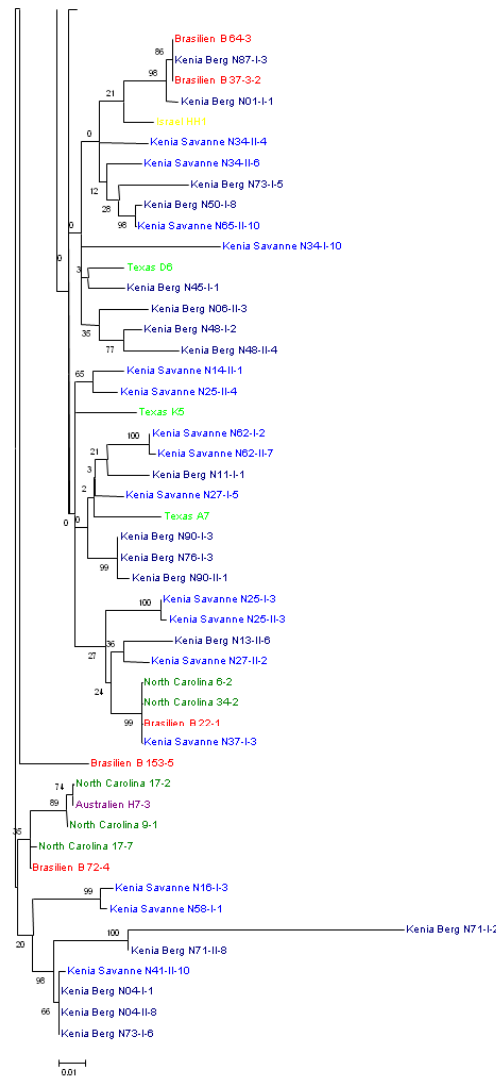


Abbildung 12: Genealogie der im Rahmen dieser Arbeit sequenzierten 193 *csd*-Fragmente aus den weltweiten Stichproben. Die Genealogie wurde anhand der Aminosäuresequenzen und unter Verwendung der Maximum-Likelihood-Methode und des JTT matrixbasierenden Modells erstellt.

Eine Möglichkeit, die Verteilung der Allele in der Population zu untersuchen ist die Erstellung einer Genealogie. In Abbildung 12 ist eine Genealogie sämtlicher im Rahmen dieser Arbeit sequenzierten *csd*-Fragmente dargestellt. Sie wurde auf Basis der Aminosäuresequenzen erstellt, da sich hier die höchsten Bootstrapwerte ergeben. Die Bootstrapwerte liegen dennoch in der Mehrzahl der Fälle unter der Signifikanzgrenze. Lediglich Sequenzen mit sehr wenigen Nukleotidunterschieden lassen sich zuverlässig anordnen (Bootstrapwerte > 85 %). Da Genealogien kleiner Sequenzdatensätze ähnlich niedrige Bootstrapwerte aufweisen (Abb. 36 und 37 im Anhang) kann ausgeschlossen werden, dass die große Zahl der Sequenzen die Ursache für die niedrigen Bootstrapwerte

ist. Vermutlich sind sie durch die große Variation bedingt, d.h. dass ancestrale Muster durch neu akkumulierende Polymorphismen überdeckt werden, so dass Verwandtschaften nur zwischen sehr eng verwandten Sequenzen erkannt werden können. In diesen Gruppen bestätigt sich die Annahme, dass sich die gleichen Allele auf verschiedenen Kontinenten finden.

Zur Analyse der genetischen Diversität wurde für die Sequenzen der verschiedenen Lokalisationen und Subpopulationen die Nukleotiddiversität pro Position für nichtsynonyme und synonyme Polymorphismen (π_n - und π_s -Werte) im Bereich der Exons 6-9 berechnet (Tab. 3). Die Berechnung wurde für jeden Sammelort separat durchgeführt. Die kenianischen Proben stammen von sechs verschiedenen Subpopulationen, die drei Bergregionen (Mount Kenya Forest, Nyambene Hills Forest und Mau Forest) sowie drei Savannenregionen (Mount Kenya Savannah, Nyambene Hills Savannah und Mau Forest Savannah) besiedeln. Für den Vergleich mit den weltweiten Stichproben wurden die Proben der aus den Bergregionen stammenden *A. m. monticola* und der in der Savanne verbreiteten *A. m. scutellata* entsprechend in die Gruppen „Kenia Berg“ und „Kenia Savanne“ zusammengefasst. Für die Untersuchung möglicher Unterschiede der Diversität in den kenianischen Subpopulationen wurden π_n und π_s zudem für jede dieser Gruppen separat berechnet.

Die Variation in den *csd*-Sequenzen führt dazu, dass in einem Alignment Lücken eingefügt werden. Positionen, die in einem Teil der Sequenzen Lücken aufweisen, wurden für die Analyse nicht verwendet. Die exakte Zahl der für diese und folgende Berechnungen verwendeten Nukleotidpositionen ist in Tabelle 3 für die jeweiligen Gruppen angegeben. Die HVR, die aus einer variablen Anzahl von Wiederholungen eines N₁₋₅Y-Motivs besteht, kann nicht vertrauenswürdig aligniert werden. Es ist nicht möglich, ihre Variation mit den bekannten Algorithmen zu analysieren, welche einzelne Positionen miteinander vergleichen. Sie kann deshalb nicht in die folgenden Analysen mit einbezogen werden.

Wie erwartet variieren die π_n - und π_s -Werte der weltweiten Lokalisationen wie auch der kenianischen Subpopulationen kaum (Tab. 3). Es finden sich in keinem Fall signifikante Unterschiede (Zweiseitiger Z-Test, $P > 0,05$) zwischen den einzelnen π_n - bzw. π_s -Werten. Auch die Größe der Stichprobe hat keinen Einfluss auf die gefundene Nukleotiddiversität. Selbst bei der nur aus fünf Sequenzen bestehenden Stichprobe aus Israel finden sich keine niedrigeren π -Werte. Grund hierfür ist vermutlich, dass durch das hohe Maß an Sequenzvariation am *csd*-Lokus auch wenige, zufällig gewählte, funktionell verschiedene Allele mit großer Wahrscheinlichkeit ein ähnliches Maß an genetischer Diversität aufweisen, wie dies bei einer großen Zahl an Allelen der Fall ist.

Tabelle 3: π_n - und π_s -Werte, berechnet für die von den einzelnen Lokalitäten und Subpopulationen stammenden *csd*-Sequenzen. Hierfür wurden die synonymen und nichtsynonymen Substitutionen pro Positionen im Bereich der Exons 6-9 (ohne HVR) des *csd*-Gens analysiert (2.2.8)). Der Standardfehler ist in Klammern angegeben. Die Berechnung erfolgte zum einen für die weltweiten Stichproben, zum anderen separat für die kenianischen Proben.

Herkunft	Nukleotid- positionen	analysierte Sequenzen	π_n	π_s
Weltweite Stichproben				
Texas	362	12	0,051 ($\pm 0,008$)	0,060 ($\pm 0,013$)
North Carolina	353	19	0,045 ($\pm 0,008$)	0,057 ($\pm 0,011$)
Arizona	359	9	0,043 ($\pm 0,008$)	0,060 ($\pm 0,014$)
Australien	353	16	0,049 ($\pm 0,008$)	0,053 ($\pm 0,014$)
Israel	365	5	0,046 ($\pm 0,009$)	0,055 ($\pm 0,016$)
Brasilien	356	18	0,042 ($\pm 0,007$)	0,055 ($\pm 0,012$)
Kenia Berg	338	49	0,058 ($\pm 0,008$)	0,066 ($\pm 0,012$)
Kenia Savanne	341	51	0,049 ($\pm 0,008$)	0,053 ($\pm 0,013$)
Kenianische Stichproben				
Mount Kenya Forest	356	16	0,051 ($\pm 0,008$)	0,059 ($\pm 0,014$)
Mount Kenya Savannah	362	15	0,050 ($\pm 0,008$)	0,049 ($\pm 0,013$)
Nyambene Hills Forest	350	17	0,053 ($\pm 0,008$)	0,064 ($\pm 0,012$)
Nyambene Hills Savannah	359	17	0,053 ($\pm 0,008$)	0,058 ($\pm 0,013$)
Mau Forest	353	15	0,067 ($\pm 0,008$)	0,071 ($\pm 0,014$)
Mau Forest Savannah	350	14	0,035 ($\pm 0,006$)	0,053 ($\pm 0,013$)

3.1.1.2 Analyse der genetischen Differenzierung der *csd*-Sequenzen zwischen den verschiedenen *A. mellifera*-Subpopulationen

Die kenianischen Honigbienen stammen aus drei Berg- und drei Savannenregionen. Die Bergregionen werden von der Unterart *A. m. monticola* und die Savanne von *A. m. scutellata* besiedelt (Meixner et al. 2000). Hiervon leitet sich die Frage ab, ob zwischen den Subpopulationen eine genetische Differenzierung existiert. Es ist anzunehmen, dass sich für *csd* keine Differenzierung zwischen den verschiedenen Subpopulationen zeigt, da sich diese, wie schon in Abschnitt 3.1.1.1 erwähnt, durch den selektiven Vorteil neu in eine Population migrierender Allele in der gesamten Population verbreiten. Muirhead (2001) beschrieb für die Allele des ebenfalls unter balancierenden Selektion stehenden S-Lokus der Pflanzen, dass sich diese in einer ausreichend großen Population panmiktisch verhalten. Dies ist auch für *csd* anzunehmen und wird durch die in Abschnitt 3.1.1.1 berechneten π -Werte unterstützt. Um auf eine mögliche genetische Differenzierung zu testen, wurde für die verschiedenen Subpopulationen der Fixierungsindex K_{ST} berechnet (Tab. 4, 2.2.8), der ein Maß für die durchschnittliche genetische Differenzierung darstellt (Charlesworth und Charlesworth 2010). Er wird anhand der Nukleotiddiversität (π) in den Sequenzen der gesamten Population (T) und der Subpopulation (S) berechnet.

$$K_{ST} = \frac{\pi_T - \pi_S}{\pi_T}$$

Die dazugehörigen Signifikanzwerte finden sich in Tabelle 5. Die zugrundeliegende Nullhypothese bezieht sich auf die Annahme, dass es keine genetische Differenzierung zwischen den Subpopulationen gibt. Die in Tabelle 4 mit einem Sternchen markierten Werte weichen signifikant von null ab, d.h. die *csd*-Sequenzen der beiden verglichenen Subpopulationen zeigen eine Differenzierung voneinander ($P < 0,05$, nicht-parametrischer Permutationstest).

Entgegen der Erwartung zeigt sich für die *csd*-Sequenzen der aus Mau Forest Savannah stammenden *A. m. scutellata*-Subpopulationen eine genetische Differenzierung zu den anderen Subpopulationen. Zudem findet sich ein signifikanter Unterschied zwischen Mount Kenya Forest und Mau Forest ($P < 0,05$). Der Fixierungsindex wurde nicht für die weltweiten Lokalitäten berechnet, da zwischen den verschiedenen Kontinenten kein natürlicher Genfluss möglich ist.

Tabelle 4: Paarweise K_{ST} -Werte der *csd*-Sequenzen der kenianischen Subpopulationen. K_{ST} -Werte wurden mit Arlequin berechnet (2.2.8). Signifikant erhöhte K_{ST} -Werte ($P < 0,05$) sind mit Sternchen markiert (Tab. 5).

	MKF	MKS	NHF	NHS	MF	MFS
Mount Kenya Forest (MKF)	0,00000					
Mount Kenya Savannah (MKS)	0,00759	0,00000				
Nyambene Hills Forest (NHF)	0,01540	0,00893	0,00000			
Nyambene Hills Savannah (NHF)	0,00615	0,00107	-0,000069	0,00000		
Mau Forest (MF)	0,00863	0,04910*	0,00906	0,03097	0,00000	
Mau Forest Savannah (MFS)	0,07615*	0,05058*	0,07557*	0,08955*	0,07540*	0,00000*

Tabelle 5: Signifikanzwerte P für die paarweisen K_{ST} -Werte der *csd*-Sequenzen der verschiedenen kenianischen Subpopulationen (Tab. 4). Die Signifikanzwerte wurden mit Hilfe eines nicht-parametrischen Permutationstest (Excoffier et al. 1992) beziehen sich auf die Nullhypothese, die von einer nicht vorhandenen Differenzierung ausgeht.

	MKF	MKS	NHF	NHS	MF	MFS
Mount Kenya Forest	*					
Mount Kenya Savannah	0,21622 $\pm$ 0,0528	*				
Nyambene Hills Forest	0,15315 $\pm$ 0,0273	0,19820 $\pm$ 0,0503	*			
Nyambene Hills Savannah	0,32432 $\pm$ 0,0624	0,40541 $\pm$ 0,0546	0,54955 $\pm$ 0,0598	*		
Mau Forest	0,22523 $\pm$ 0,0546	0,00901 $\pm$ 0,0091	0,27928 $\pm$ 0,0533	0,06306 $\pm$ 0,0194	*	
Mau Forest Savannah	0,00901 $\pm$ 0,0091	0,01802 $\pm$ 0,0121	0,00000 $\pm$ 0,0000	0,00000 $\pm$ 0,0000	0,00000 $\pm$ 0,0000	*

3.1.2 Struktur und Evolution der hypervariablen Region von *csd*

Die hypervariable Region von *csd* besteht aus einer zwischen den verschiedenen Allelen sehr unterschiedlichen Zahl von Wiederholungen eines $N_{1-5}Y$ -Motivs. Sie variiert stark in ihrer Länge und Nukleotidkomposition (Abb. 13). Hasselmann et al. (2008b) postulierten, dass sie entscheidend zur Allelspezifität beiträgt.

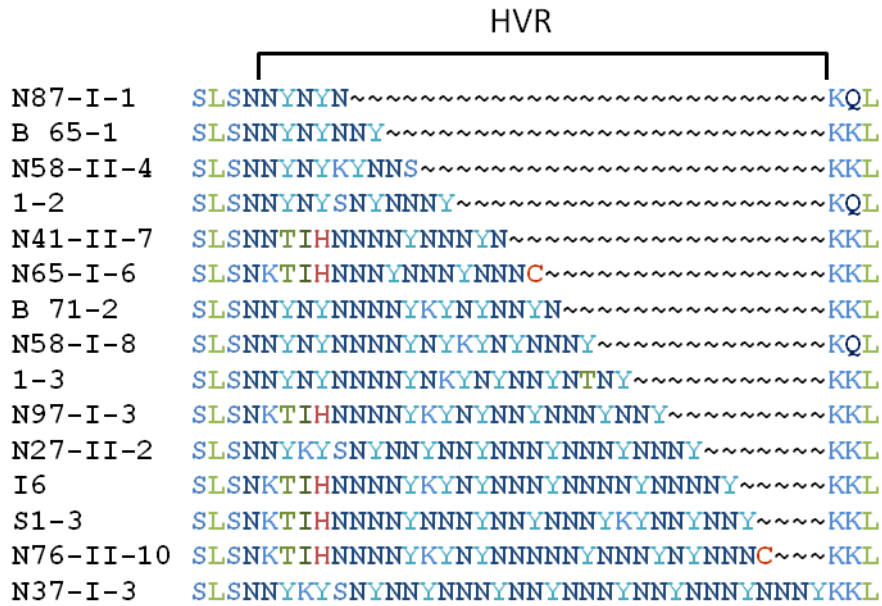


Abbildung 13: Exemplarische Darstellung einiger ausgewählter HVR-Aminosäuresequenzen verschiedener Allele des *csd*-Gens aus *A. mellifera*.

3.1.2.1 Die Länge der hypervariablen Region korreliert nicht mit dem Alter der *csd*-Allele

Die Länge der hypervariablen Region variiert bei den bislang analysierten *csd*-Sequenzen zwischen fünf und 32 Aminosäuren. Eine Hypothese ist, dass die Länge der HVR mit dem Alter eines Allels assoziiert ist, d.h. dass beispielsweise alte Allele eine längere HVR haben als junge. Um dies zu überprüfen wurden die Sequenzen abhängig von der Länge ihrer hypervariablen Region (5-13 AS, 14-23 AS, 24-32 AS) in drei Gruppen unterteilt. Für jede dieser Gruppen wurde im sequenzierten Bereich der Exons 6-9 (ohne HVR) die Nukleotiddiversität pro Position für die synonymen Polymorphismen (π_s -Werte) berechnet (2.2.8), welche als neutral gelten und daher ein Maß für das Alter einer Sequenz sind. In Tabelle 6

sind diese sowie zum Vergleich die Werte für die nichtsynonymen Polymorphismen (π_n) dargestellt.

Tabelle 6: Nukleotiddiversität pro Position für synonyme (π_s) und nichtsynonyme (π_n) Polymorphismen im kodierenden Bereich der Exons 6-9 der sequenzierten *csd*-Sequenzen (2.2.8). Die Sequenzen wurden für die Analyse abhängig von der Länge ihrer HVR in drei Gruppen unterteilt.

HVR-Länge (AS)	Zahl der analysierten Sequenzen	π_s	π_n
5-13	36	0,069 ($\pm 0,012$)	0,070 ($\pm 0,009$)
14-23	104	0,069 ($\pm 0,012$)	0,072 ($\pm 0,009$)
24-32	25	0,056 ($\pm 0,011$)	0,070 ($\pm 0,008$)

Die Werte für die verschiedenen Gruppen in Tabelle 6 weichen nicht signifikant voneinander ab ($P > 0,05$; zweiseitiger Z-Test). Es ist somit nicht davon auszugehen, dass die Länge der HVR mit dem Alter der Sequenzen korreliert.

Abbildung 14 zeigt eine grafische Darstellung der Verteilung der paarweisen π_s -Werte zwischen den Sequenzen innerhalb der drei Gruppen.

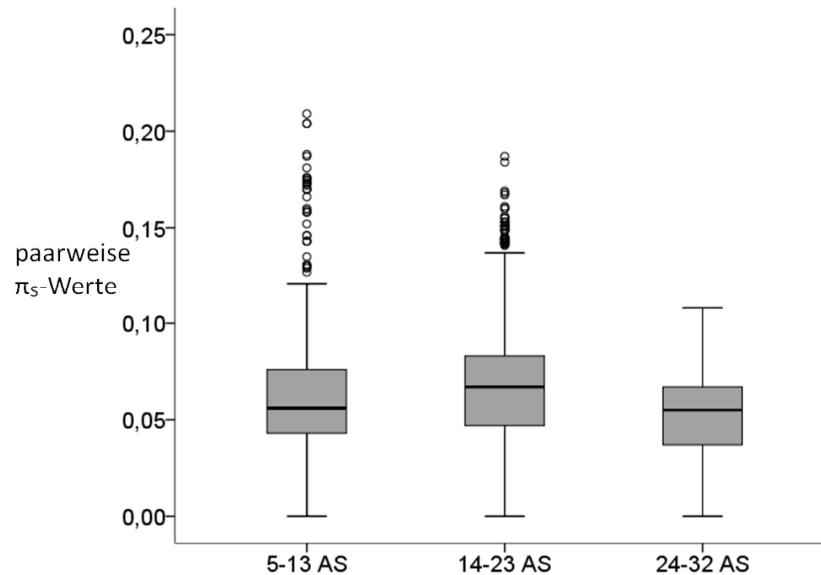


Abbildung 14: Grafische Darstellung der Verteilung der paarweisen π_s -Werte. Die Sequenzen wurden für die Analyse abhängig von der Länge ihrer HVR in drei Gruppen unterteilt. Die Berechnung der π_s -Werte erfolgte mit MEGA, paarweise Vergleich der Sequenzen nach Nei-Gojobori (Nei und Gojobori 1986). Die Box-Whisker-Plots wurden mit Hilfe von PASW erstellt. Alle drei Mittelwerte weichen signifikant voneinander ab ($P < 0,001$; ANOVA nach Kruskal-Wallis).

fität entsteht, ist nicht klar, ob es sich beispielsweise bei den Allelen H3-1 und H4-1 schon um funktionell verschiedene Allele handelt oder nicht.

L5	SLS D KTIHNNNNYKYNNNNNNNYS
B 71-3	SLSN K TIHNNNNYKYNNNNNNNYS
9-1	SLSNKT I HNNNNYNNNNYNNNNYNNY
17-2	SLSNKT I HNNNNYNNNNYNNNNYNNY
H3-1	SLSNKT I HNNNNYK NY NNY
H4-1	SLSNKT I HNNNNYK NY NNYK NY NNYK NY NNY
N73-II-4	SLSNKT I HNNNNYNNNNYNNNNYNNY
N106-I-1	SLSNKT I HNNNNYNNNNYNNNNYNNYNNY
34-2	SLSNNYKYSNNNNYNNNNYNNY D NNYNNNY
6-2	SLSNNYKYSNNNNYNNNNYNNYNNYNNYNNNY
B22-1	SLSNNYKYSNNNNYNNNNYNNYNNYNNYNNNY
N37-I-3	SLSNNYKYSNNNNYNNNNYNNYNNYNNYNNNY
B153-4	SLSNKT I HNNNNYKYNNNNYNNNNYNNNNY
34-3	SLSNKT I HNNNNYKYNNNNYNNNNYNNNNYNNNNY
H4-2	SLSNNYK NY NNNNYNNNNYNNY
H5-3	SLSNNYK Y SNNNNYNNY
B68-4	SLSNKT I HNNNNYNNNNYNNNNYNNY
S1-3	SLSNKT I HNNNNYNNNNYNNNNYK NY NNYNNY

Abbildung 16: Darstellung der HVR-Sequenzen (Aminosäuren) von Allelen, die lediglich in der HVR, jedoch nicht im sequenzierten Bereich der Exons 6-9 Sequenzunterschiede aufweisen. Durch farbige Striche sind duplizierte und inserierte bzw. in der jeweils anderen Sequenz deletierte Bereiche markiert. Mit Pfeilen sind Basensubstitutionen oder -insertionen gekennzeichnet.

Es finden sich jedoch auch Sequenzen mit identischer HVR und Unterschieden in den benachbarten Regionen. Beispielsweise weisen die beiden Allele 17-7 und H7-3, welche eine identische HVR-Sequenz haben, einen d_n -Wert von 0,036 und einen d_s -Wert von 0,028 auf. Zwischen N34-II-4 und N87-II-3 ist der d_n -Wert 0,036 und d_s 0,069, die HVR-Sequenz jedoch ebenfalls identisch.

Es zeigt sich somit, dass die HVR unabhängig vom umliegenden Sequenzbereich evolviert. Da duplizierte Bereiche auch ebenso wieder deletiert werden können, ist es möglich,

dass Unterschiede in der HVR der zuvor genannten Sequenzen lediglich nicht mehr zu erkennen sind. Dies würde bedeuten, dass die HVR nicht nur unabhängig, sondern möglicherweise auch deutlich schneller als die umliegenden Bereiche evolviert.

3.1.3 Berechnung der in der *A. mellifera*-Population repräsentierten Zahl an *csd*-Allelen

Bisher ist die Frage nach der Zahl der in der gesamten *A. mellifera*-Population segregierenden *csd*-Allele nicht eindeutig geklärt. Laidlaw et al. (1956) schätzten die Zahl auf 9-11 Allele, Adams et al. (1977) ging von 19 Allelen aus und 2008 wurde von Hasselmann et al. die Existenz von 14 Allelen nachgewiesen. Diese Berechnungen wurden bisher jeweils nur an einzelnen Regionen durchgeführt. Mit dem im Rahmen dieser Arbeit erstellten *csd*-Sequenzdatensatz ist es zum ersten Mal möglich, diese Frage anhand einer größeren Zahl von Stichproben aus verschiedenen Regionen der Welt zu beantworten. Es zeigt sich, dass sich an den verschiedenen Lokalitäten und Subpopulationen deutlich unterschiedliche Spektren an Allelen finden. Dies deutet darauf hin, dass in den einzelnen Regionen nur ein Teil der weltweit vorhandenen repräsentiert sind und somit von der Analyse einer Lokalität keine Rückschlüsse auf die in der gesamten Honigbienenpopulation vorhandenen Allele gezogen werden können.

Um die Zahl der *csd*-Allele exakt bestimmen zu können, ist es notwendig, den Unterschied zwischen Allelen, d.h. funktionell verschiedenen Sequenzen, und den funktionell identischen Replikaten zu kennen. Bisher ist nicht bekannt, welche Polymorphismen zur Allelspezifität beitragen und welche funktionell neutral sind und sich auf Grund des Hitchhiking-Effekts durch die auf *csd* wirkende balancierende Selektion akkumuliert haben.

Bei einem Großteil der für diese Arbeit verwendeten Arbeiterinnen war es möglich, beide *csd*-Allele zu sequenzieren. Zudem standen (mit Ausnahme der israelischen Stichprobe) immer mehrere Drohnen aus den verschiedenen Völkern zu Verfügung. Soweit es gelang beide von der Königin vererbten und in den Drohnen ihres Volks repräsentierten Allele zu identifizieren, stehen auch diese Allelpaaire zur Verfügung. Insgesamt konnten so 78 Allelpaaire sequenziert werden, die Informationen darüber liefern können, welche Unterschiede notwendig sind, um eine neue Allelspezifität zu determinieren. Besonderer Fokus wird hierbei auf die hypervariable Region gelegt, von der angenommen wird, dass sie entscheidend zur Allelspezifität beiträgt (Hasselmann et al. (2008b)). Die Aminosäuresequenzen der hypervariablen Regionen dieser Paare sind im Anhang im Abschnitt 6.1 aufgelistet.

Alle 78 Allelpaare weisen in ihrer HVR deutliche Sequenzunterschiede sowie mindestens eine Aminosäure Längenunterschied auf. Ein Anhaltspunkt für ein Kriterium, nach welchem funktionell verschiedene Sequenzen unterschieden werden können, ist ein Allelpaar, das bei nur einer Aminosäure Längendifferenz sechs Unterschiede in der Sequenzkomposition aufweist (2-2 und 2-5, 6.1). Es gibt auch Hinweise darauf, dass die hypervariablen Regionen zweier Allele gleich lang sein können (C. Seelmann, pers. Mitteilung). In diesem Fall finden sich in der HVR elf Aminosäureunterschiede. Bisher nicht bekannt ist, ob es sich hierbei um das mögliche Minimum an Unterschieden handelt. Geht man davon aus, dass die HVR-Sequenzen von Allelen sich in der Länge oder aber an mindestens sechs bis elf Aminosäurepositionen unterscheiden müssen, und wendet dieses Kriterium auf den gesamten hier analysierten Datensatz an, so finden sich 50-90 Allele, die zueinander die entsprechenden Unterschiede aufweisen. Auch wenn es sich dabei nur um eine ungefähre Annahme handelt, zeigt es, dass die Zahl der Allele deutlich höher zu sein scheint, als bisher in der Literatur angenommen.

Ein weiteres Kriterium sind die nichtsynonymen Polymorphismen in der potentiellen Spezifizierungsdomäne (PSD). Hierfür wurde π_n und π_s in der PSD (Exons 6 und 7) für jedes Allelpaar berechnet (ohne HVR, 6.1). Dabei zeigte sich, dass junge Allelpaare, d.h. Allelpaare mit niedrigem π_s -Wert, eine im Vergleich zu älteren Allelen deutlich erhöhte Rate an nichtsynonymen zu synonymen Substitutionen haben (Abb. 17). Dies deutet darauf hin, dass es in neu entstandenen Allelen im Vergleich zu alten eine höhere evolutionäre Rate gibt. Die Tatsache, dass diese nichtsynonymen Substitutionen in der PSD schnell fixiert wurden, ist ein Hinweis darauf, dass sie funktionell relevant sind und **zusätzlich** zu den Unterschieden in der HVR zur Allelspezifität beitragen.

Der **niedrigste** π_n -Wert zwischen den potentiellen Spezifizierungsdomänen (PSD) findet sich mit 0,025 ($\cong$ 4 Aminosäureunterschieden) bei dem Allelpaar N01-II (6.1). Betrachtet man dies als niedrigsten möglichen π_n -Wert zwischen zwei Allelen und wendet dieses Kriterium auf den gesamten Datensatz an, so finden sich im gesamten Sequenzdatensatz 59 Allele, die zueinander mindestens einen π_n -Wert von 0,025 aufweisen und zwischen deren hypervariablen Regionen sich mindestens sechs Aminosäureunterschiede finden. Diese Zahl ist ebenfalls deutlich höher als die in der Literatur postulierten Allelzahlen.

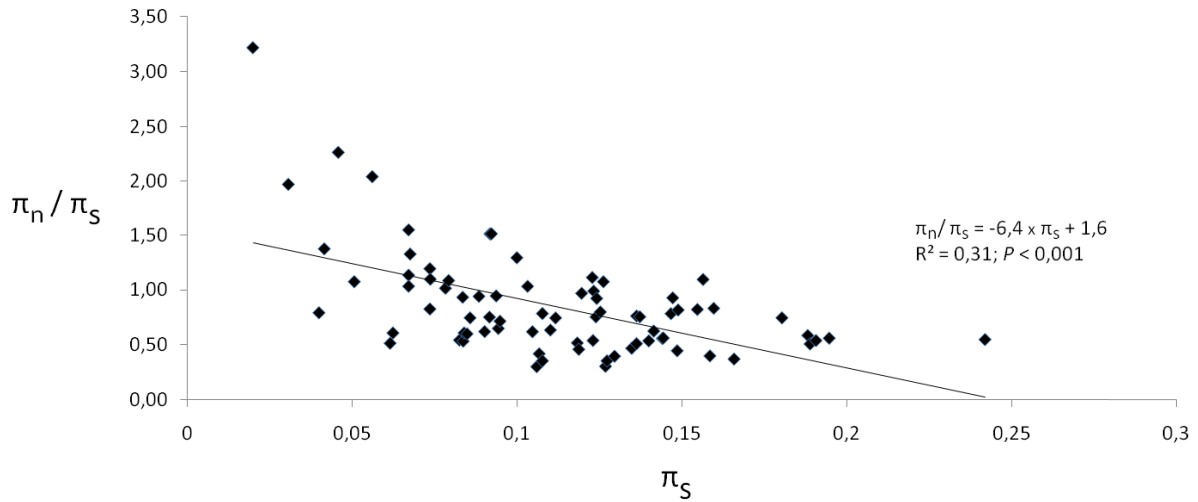


Abbildung 17: Verhältnis von π_n zu π_s in Abhängigkeit zu den π_s -Werten der 78 Allelpaare. Es zeigt sich eine lineare Regression ($\pi_n/\pi_s = -6,4 \times \pi_s + 1,6$) mit einem Regressionskoeffizient von $R^2 = 0,31$. Die Regressionsgerade zeigt einen hoch signifikanten Trend ($P < 0,001$). Junge Allele (niedriger π_s -Wert) weisen ein höheres π_n/π_s -Verhältnis auf als Allelpaare mit höherem π_s -Wert.

3.1.4 Neue Berechnung der Koaleszenzzeit der *csd*-Allele

Allele eines Gens, das unter balancierender Selektion steht, haben im Vergleich zu neutralen Loci eine deutlich längere Koaleszenzzeit (Takahata 1990). Für *csd*-Allele wurde die durchschnittliche Koaleszenzzeit innerhalb dreier *Apis*-Arten von Hasselmann et al. (2008b) auf ungefähr 6 Mio. Jahre berechnet. Diese Kalkulation basierte auf der durchschnittlichen synonymen Divergenz zwischen *A. mellifera* und *A. dorsata* unter Verwendung der molekularen Uhr. Diese wurde unter anderem anhand der Divergenzzeitpunkte zwischen *A. mellifera* und *A. cerana* vor 6 - 8 und *A. dorsata* vor 10 Mio. Jahren berechnet. In der 2010 erschienenen Publikation von Ramírez et al. wurden die Divergenzzeitpunkte erneut bestimmt. In der Studie wurde postuliert, dass die Aufspaltung zwischen *A. mellifera* und *A. cerana* schon vor 18,2 Mio. Jahre stattfand und sich *A. dorsata* vor 23,5 Mio. Jahre von der Linie der beiden anderen *Apis*-Arten trennte. Zudem stehen mittlerweile zusätzliche Sequenzinformationen weiterer Referenzgene aus den drei *Apis*-Arten zur Verfügung (2.2.9), welche eine exaktere Bestimmung der pro Position akkumulierenden synonymen Polymorphismen (d_s) zwischen *A. mellifera* und *A. cerana* bzw. *A. dorsata* ermöglichen (*Am-Ac*: $d_s = 0,09$, *Am-Ad*: $d_s = 0,0927$). Auf Basis dieser veränderten Parameter wurde die genomische Mutationsrate (μ) erneut bestimmt (2.2.9). Sie beträgt unter Verwendung der *A. mellifera*- und *A. cerana*-Daten $2,47 \times 10^{-9}$ und unter Verwendung der *A. mellifera*- und *A. dorsata*-Daten $1,97 \times 10^{-9}$ (2.2.9). Zwischen den Ad-

und *Am-csd*-Sequenzen (2.1.6, 3.1.1.1) wurde ein d_s -Wert von 0,071 berechnet, woraus sich mit den veränderten Parametern eine durchschnittliche Koaleszenzzeit von 14,2 - 17,8 Mio. Jahren ergibt. Zwar liegen diese Werte deutlich über der in Hasselmann et al. (2008b) postulierten durchschnittlichen Koaleszenzzeit von 6 Mio. Jahren, im Verhältnis zum Divergenzzeitpunkt ist der Wert jedoch gleich geblieben. Für die *csd*-Allele von *Apis mellifera* ergibt sich bei einem π_s -Wert von 0,064 eine durchschnittliche Koaleszenzzeit von 13 - 16,2 Mio. Jahren. Der höchste π_s -Wert liegt für *Am-csd* bei 0,24, woraus sich ein Höchstalter von 48 - 61 Mio. Jahren für die Allele errechnet.

Die Zahl der in einer Population segregierenden Allele steht in Abhängigkeit zur effektiven historischen Populationsgröße (N_e) und der Entstehungsrate neuer Allele, welche sich aus der Mutationsrate (μ) und der Anzahl an nichtsynonymen Positionen in der PSD des *csd*-Gens ergibt. Die Entstehungsrate (u) wurde mit 2×10^{-6} bis 2×10^{-7} Allele pro Gen und pro Jahr bestimmt (2.2.9). Hieraus errechnet sich bei einer Zahl von 50 - 90 Allelen eine effektive historische Populationsgröße (N_e) zwischen 12.368 und 41.929 (Tab. 7, 2.2.9).

Tabelle 7: Parameter für die Berechnung der durchschnittlichen Koaleszenzzeit (T_d) der *csd*-Allele (2.2.9).

Entstehungsrate u	Allelzahl n	Populationsgröße N_e
2×10^{-6}	50	12.368
	90	36.762
2×10^{-7}	50	13.971
	90	41.929

3.2 Das Gen *complementary sex determiner* existiert auch in der asiatischen Zwerghonigbiene *Apis florea*

Für die asiatische Zwerghonigbiene *Apis florea* kann die Existenz des Gens *complementary sex determiner*, welches in den *Apis*-Arten *A. mellifera*, *A. cerana* und *A. dorsata* bekannt ist, ebenfalls postuliert werden. Grundlage hierfür ist die Berechnung, dass *csd* zwar nach der Aufspaltung der Gattungen der stachellosen Bienen, Hummeln und Honigbienen, jedoch vor der Aufspaltung der verschiedenen *Apis*-Arten entstanden ist (Hasselmann et al. 2008a). Die Existenz von *csd* konnte in *A. florea* bisher jedoch nicht nachgewiesen werden. In vorangegangenen Arbeiten wurde *feminizer* von *A. florea* sequenziert (M. Hasselmann, unveröffentl. Daten; 6.3.1), wobei sich zusätzliche, zu *fem* variierende Sequenzstücke fanden. Ausgehend von diesen konnte zunächst ein weiteres von *fem* abweichendes Sequenzfragment amplifiziert und die cDNA-Sequenz anschließend mittels 5'- und 3'-RACE-PCR vervollständigt werden (2.2.7.1). Die Sequenz ist im Anhang unter 6.2.1 dargestellt. Das sequenzierte Gen zeigt deutliche Homologien zu *csd* in den anderen *Apis*-Arten, so dass davon ausgegangen wird, dass es sich hierbei um *csd* der asiatischen Zwerghonigbiene *A. florea* handelt.

Im Anhang ist unter 6.2.2 ein Alignment der *csd*-Sequenzen aus den vier untersuchten *Apis*-Arten *A. florea*, *A. mellifera*, *A. cerana* und *A. dorsata* dargestellt. In Abbildung 18 ist eine Genealogie der *csd*-Sequenzen der vier *Apis*-Arten und von *fem* von *Apis* und weiteren Hymenopteren gezeigt. Dort gruppiert sich *Af-csd* zu den *csd*-Sequenzen der anderen *Apis*-Arten.

Ein Abgleich mit der Genomsequenz (GenBank-Nummer AAEKZ000000000.1 auf NCBI) von *A. florea* ermöglichte die Analyse der Genstruktur von *Af-csd* sowie auch *Af-fem*. Die Genomsequenzen liegen bisher als teilweise assemblierte Konsensussequenzen (Contigs) vor und wurden noch nicht annotiert. Es wurde ein Contig gefunden, auf dem beide Gene lokalisiert sind. Während *Am-fem* 12 kb strangaufwärts von *Am-csd* liegt (Hasselmann et al. 2008a), beträgt die Distanz zwischen den beiden Genen in *A. florea* lediglich ca. 3,5 kb. Die Exon-Intron-Struktur stimmt bei *csd* und *fem* von *A. florea* und *A. mellifera* überein (Abb. 19, auch Abb. 5 und 11). Länge und Sequenz der intronischen Bereiche unterscheiden sich dagegen stark. Weitere *fem*- oder *csd*-ähnliche Sequenzen fanden sich in der Genomsequenz von *A. florea* nicht. Auch bei einem Abgleich der *csd*- und *fem*-Sequenzen mit *A. florea*-Transkriptomdaten in der NCBI-Datenbank fanden sich keine weiteren *csd*- oder *fem*-ähnlichen Sequenzen.

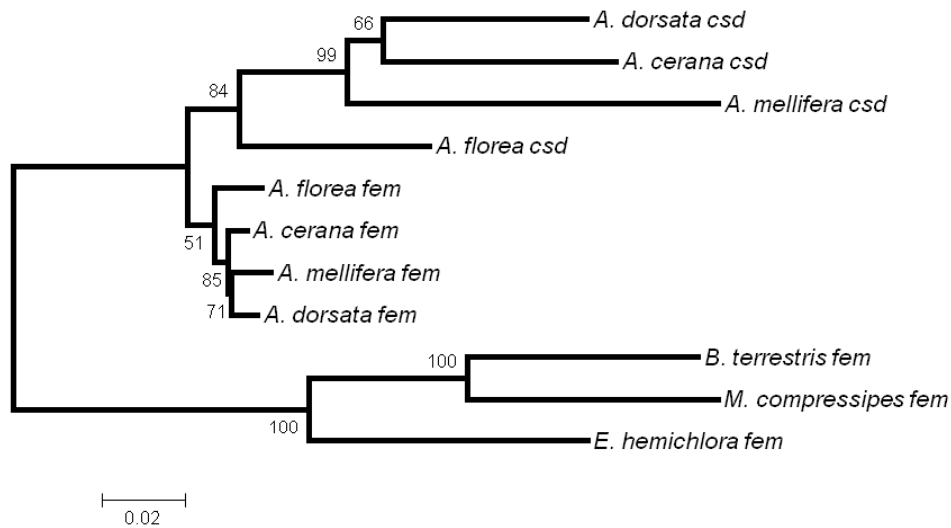


Abbildung 18: Genealogie auf Basis der Nukleotidsubstitutionen der *csd*- und *fem*-Sequenzen aus *A. mellifera*, *A. cerana*, *A. dorsata* und *A. florea* sowie *fem* aus *B. terrestris*, *M. compressipes* und *E. hemichlora* nach der Maximum-Likelihood-Methode. Die Astlängen stellen die Anzahl an Nukleotidsubstitutionen pro Position dar, die auf Basis des Tamura-Nei-Modells berechnet wurden. Die Bootstrap-Werte (500 Wiederholungen) sind an den inneren Ästen angegeben. Die Länge der Äste entspricht den Nukleotidsubstitutionen pro Position.

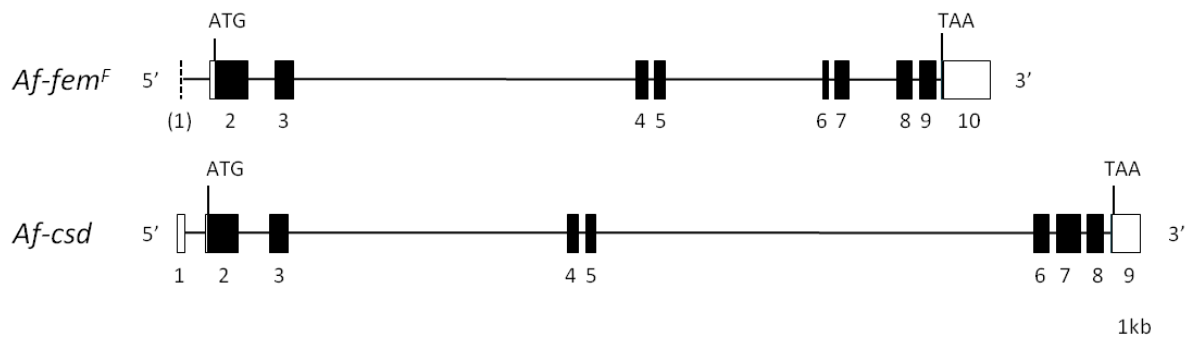


Abbildung 19: Struktur des *Af-csd*-Gens und der in der weiblichen Spleißform von *Af-fem* vorhandenen Exons in 5'-3'-Richtung. Exons sind als Kästchen, Introns als Linie dargestellt. Die Position und Größe der männchenspezifischen Exons in *fem* sind bisher nicht bekannt. Zudem fehlt in *fem* Sequenzinformation über die 5'-terminale untranslatierte Region (5'-UTR), ein vermutetes Exon ist angedeutet. *fem* ist ca. 7 kb groß, *csd* 8,7 kb. *fem* liegt im Genom der Zwerghonigbiene *A. florea* ca. 3,5 kb strangaufwärts von *csd*.

Af-csd besteht wie *Am-csd* aus neun Exons, das Startkodon befindet sich in Exon 2, das Stoppkodon in Exon 9. Von *Af-fem* wurde die 5'-UTR nicht sequenziert, weshalb die Zahl der Exons nicht vollständig bestimmt werden kann. Ein Vergleich mit der Struktur von *Am-fem*, die im bekannten Genbereich übereinstimmt, lässt vermuten, dass noch ein weiteres Exon am 5'-Ende existiert. Somit würde das weibliche *Af-fem* aus zehn Exons bestehen, von denen Exon 2 das Start- und Exon 10 das Stoppkodon enthält.

Bei den *Apis*-Arten *A. mellifera*, *A. cerana* und *A. dorsata* findet sich in *csd* ein hohes Maß an Sequenzvariation (Hasselmann et al. 2008b). Um zu testen, ob sich im *csd*-Gen von *A. florea* ein ähnliches Maß an Variation findet, wurden aus der genomischen DNA von zwanzig Arbeiterinnen die Exons 6-9 amplifiziert und sequenziert (2.2.7.1 und 2.2.6). Diese Arbeiterinnen wurden aus einem *A. florea*-Volk in Thailand gesammelt. Dies bedeutet, dass alle Individuen von der gleichen Königin abstammen und somit eines der beiden von ihr vererbten Allele tragen (Königinnenallele) sowie ein vom Vater stammendes Drohnenallel, das eine andere Spezifität hat. Auch bei *A. florea* paart sich die Königin mit mehreren Drohnen (Palmer und Oldroyd 2001), so dass davon auszugehen ist, dass in einem Volk mehrere Drohnenallele repräsentiert sind. Insgesamt konnten elf in ihrer HVR variierende Sequenzen identifiziert werden (Abb. 20). Dabei zeigte sich, dass sich die für *csd* charakteristische HVR in *A. florea* im Vergleich zu den anderen *Apis*-Arten deutlich unterscheidet. Während dort die HVR aus Wiederholungen eines $N_{1-5}Y$ -Motivs besteht, dem in *A. cerana* und *A. dorsata* $(KHYN)_{1-4}KH$ -Motive folgen (Hasselmann et al. 2008b), besteht die HVR in *A. florea* fast ausschließlich aus Asparagin (N), unterbrochen durch ein bis zwei Serine (S) sowie einem Tyrosin (Y) am Beginn einiger Sequenzen (Abb. 20). Lediglich in einer der elf Sequenzen findet sich zusätzlich ein einzelnes Histidin (H, Sequenz 21-4) und in einer weiteren ein Isoleucin (I, Sequenz 1-9). Die Länge der HVR der gefundenen Sequenzen variiert um bis zu 18 Aminosäuren. Vergleichend dazu wurden in *A. mellifera* bisher bis zu 27 Aminosäuren Längenunterschied zwischen den hypervariablen Regionen der verschiedenen Sequenzen gefunden.

Von sechs Arbeiterinnen konnten beide *csd*-Allele sequenziert werden. Zudem konnten die beiden Königinnenallele des Volks identifiziert werden (K1 und K2, Abb. 21). Ein erster Test, ob in *Af-csd* eine ähnliche Nukleotiddiversität vorhanden ist wie in *Am-csd*, soll anhand dieser sieben Allele durchgeführt werden. In Tabelle 8 sind die π_n - und π_s -Werte der Exons 6-9 dargestellt, die für die jeweiligen Allelpaare berechnet wurden. In Abbildung 21 sind die HVR-Sequenzen dieser Paare gezeigt.



Abbildung 20: Hypervariable Region von *csd*-Allelen aus verschiedenen *A. florea*-Individuen.

Tabelle 8: Rate an synonymen sowie nicht-synonymen Polymorphismen pro Position zwischen den Exons 6-9 (ohne HVR) der verschiedenen Allele, sowie die Zahl an Aminosäureunterschieden. Berechnung der Werte mit MEGA nach dem Nei-Gojobori-Modell und der Jukes-Cantor-Methode.

	π_n	π_s	AS-Untersch.
2-4 / 2-5	0,018	0,000	8
4-1 / 4-2	0,012	0,008	6
7-2 / 7-4	0,000	0,000	0
8-1 / 8-2	0,008	0,000	3
11-1 / 11-2	0,012	0,009	4
17-3 / 17-4	0,013	0,000	2
K1 / K2	0,010	0,000	4

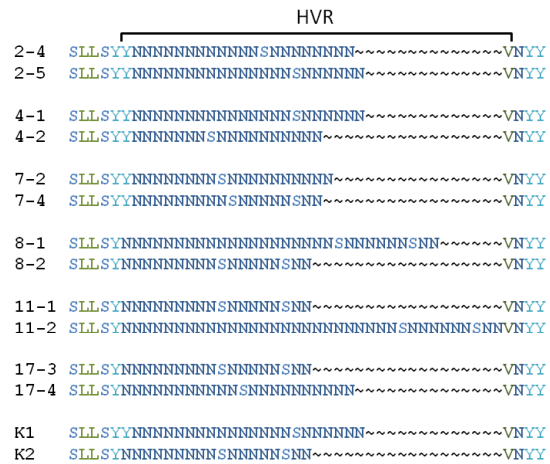


Abbildung 21: Funktionell verschiedene *csd*-Allelpaare aus *A. florea*. Mit K1 und K2 sind die beiden von der Königin des Volks stammenden Allele bezeichnet. Sie konnten aus den Sequenzen der Arbeiterinnen identifiziert werden und stellen ebenfalls ein Allelpaar da.

Die Diversität zwischen den funktionell verschiedenen Allelpaaren ist in *A. florea* deutlich geringer als in *A. mellifera*. Alle π_n - und π_s -Werte von *Af-csd* befinden sich unterhalb des niedrigsten für ein *A. mellifera*-Allelpaar gefundenen Werts (π_n und π_s jeweils 0,021). Auch im Durchschnitt liegen sie deutlich unter den Werten von *A. mellifera*: π_n : 0,010 (*Af-csd*) zu 0,048 (*Am-csd*), π_s : 0,001 (*Af-csd*) zu 0,057 (*Am-csd*). Die hypervariablen Regionen der Allelpaare zeigen in jedem Fall Längenunterschiede. Zudem ist die Position

der einzelnen Serine (S) innerhalb des Poly-N-Bereichs der HVR zwischen den beiden Allelen in allen sieben Fällen unterschiedlich.

Ein direkter Vergleich der Diversität zwischen den *csd*-Sequenzen der *A. florea*- und *A. mellifera*-Population ist auf Grund der unterschiedlichen Stichproben nicht möglich. Während die *A. florea*-Proben aus einem Volk stammen, wurden die *A. mellifera*-Proben aus vielen verschiedenen Völkern gesammelt. Zum Vergleich der Diversität in den beiden Arten ist es notwendig, vergleichbare Stichproben zu erheben, die dann mit einem Rangsummentest überprüft werden können. Für diese Stichproben wurden immer vier Sequenzen von zwei Arbeiterinnen des gleichen Volks verwendet und die π -Werte berechnet. Bei den *A. florea*-Proben wurden π_n und π_s jeweils für zwei Allelpaare aus Tabelle 8 berechnet. Die hierbei berechneten Werte sind in Tabelle 9 dargestellt. Für *A. mellifera* stehen aus Kenia die Sequenzen von jeweils zwei aus dem gleichen Volk stammenden Arbeiterinnen zur Verfügung. Es wurden die Sequenzen der zwei Arbeiterinnen eines Volkes verglichen, bei denen alle vier Allele sequenziert werden konnten (Tab. 10). So können die π_n - und π_s -Werte, die jeweils zwischen zwei Arbeiterinnen eines Volkes berechnet wurden, zwischen *A. florea* und *A. mellifera* verglichen werden. Der Unterschied zwischen der Diversität in beiden Arten ist signifikant ($P < 0,001$, zweiseitiger Mann-Whitney-U-Test). Auch in diesem Vergleich zeigte sich, dass die Diversität zwischen den *csd*-Allelen von *A. florea* deutlich niedriger ist als zwischen den *csd*-Allelen von *A. mellifera*.

Tabelle 9: π_n - und π_s -Werte zwischen den *csd*-Allelen (Bereich Exon 6-9) jeweils zweier *A. florea*-Arbeiterinnen.

	A 2		A 4		A 7		A 8		A 11	
	π_n	π_s	π_n	π_s	π_n	π_s	π_n	π_s	π_n	π_s
A 4	0,011	0,004								
A 7	0,009	0,000	0,006	0,004						
A 8	0,010	0,000	0,007	0,004	0,001	0,000				
A 11	0,014	0,004	0,011	0,009	0,006	0,004	0,007	0,004		
A 17	0,011	0,000	0,008	0,004	0,003	0,000	0,004	0,000	0,008	0,004

Tabelle 10: π_n - und π_s -Werte zwischen den *csd*-Allelen (Bereich Exon 6-9) von je zwei aus dem gleichen Volk stammenden *A. mellifera*-Arbeiterinnen. Sämtliche hier verwendeten Proben stammen aus Kenia.

Volk	π_n	π_s	Volk	π_n	π_s
N01	0,048	0,059	N48	0,057	0,056
N04	0,036	0,061	N58	0,046	0,042
N06	0,046	0,057	N62	0,053	0,061
N14	0,037	0,038	N76	0,063	0,078
N16	0,028	0,053	N87	0,046	0,049
N20	0,028	0,053	N90	0,056	0,047
N25	0,034	0,055	N95	0,025	0,037
N27	0,045	0,061	N98	0,029	0,056
N34	0,077	0,072	N106	0,043	0,061
N41	0,049	0,057			

Dass sich im *csd*-Gen von *A. florea* im Vergleich zu *Am-csd* so wenige Polymorphismen akkumuliert haben, deutet darauf hin, dass die *Af-csd*-Allele noch sehr jung sind. Das Alter eines Allels kann anhand der Mutationsrate und der synonymen Polymorphismen (π_s) berechnet werden. Für *A. florea* ist die Mutationsrate noch nicht bekannt, doch in der Annahme, dass diese ähnlich ist wie in den anderen *Apis*-Arten ($1,97 \times 10^{-9}$ - $2,47 \times 10^{-9}$; 3.1.4), errechnet sich für die *Af-csd*-Allele ein durchschnittliches Alter von 404.858 bis 507.641 Jahren (Höchstalter 2,2-2,8 Mio. Jahre). Das Alter der *Af-csd*-Allele ist somit deutlich niedriger als das der *Am-csd*-Allele (13 - 16 Mio. Jahre, 3.1.4).

Die niedrige genetische Diversität und das junge Alter entsprechen nicht der Erwartung, unter der Annahme, dass *Af-csd* wie die *csd*-Gene der anderen drei *Apis*-Arten unter balancierender Selektion steht. Um dies zu testen, soll die Variation innerhalb des *Af-csd*-Gens getestet werden. Hasselmann et al. (2008b) zeigten, dass in den *csd*-Genen von *A. mellifera*, *A. cerana* und *A. dorsata* das Maß der Sequenzvariation innerhalb des Gens schwankt. Die höchste Variation findet sich in den Exons 6 und 7, was bedeutet, dass diese Region das Ziel der balancierenden Selektion ist. Wirkt sie auch auf *Af-csd*, so wäre zu erwarten, dass sich hier ebenfalls Unterschiede in der Variation innerhalb des Gens finden. Um diese detektieren zu können, wurden zusätzlich zum Bereich der Exons 6-9 die Exons 2-3 aus acht Arbeiterinnen sequenziert. Durch den Vergleich der π -Werte der zwei Bereiche (Tab. 11) können Unterschiede in der Nukleotiddiversität innerhalb des Gens detektiert werden und somit Hinweise auf eine potentielle Zielregion der balancierenden Selektion gewonnen werden.

Tabelle 11: π_n - und π_s -Werte für die Bereiche von Exon 2-3 und Exon 6-9 des *csd*-Gens von *A. florea*. Die Analyse der synonymen und nichtsynonymen Substitutionen pro Position erfolgte mit MEGA nach dem Nei-Gojobori-Methode nach der Jukes-Cantor-Modell (2.2.8).

	π_n	π_s
Exon 2-3	0,000 ($\pm 0,001$)	0,009 ($\pm 0,006$)
Exon 6-9	0,007 ($\pm 0,002$)	0,002 ($\pm 0,002$)

Es findet sich kein Unterschied in den π_s -Werten des vorderen und des hinteren Genbereichs ($P > 0,1$; zweiseitiger Z-Test). Der π_n -Wert in den Exons 6-9 ist dagegen signifikant höher als in den Exons 2-3 ($P < 0,001$). Wie beim *csd*-Gen der anderen *Apis*-Arten (Hasselman et al. 2008a) zeigen sich bei *Af-csd* ebenfalls Diversitätsunterschiede innerhalb des Gens. Dies ist ein Hinweis darauf, dass *csd* auch in *A. florea* unter balancierender Selektion steht.

Um weitere Einblicke zu gewinnen, wodurch die Unterschiede zwischen *Af-csd* und den anderen *csd*-Genen zu erklären sind, wurde eine Genealogie der *csd*-Sequenzen erstellt (Abb. 22). Die Anordnung der Sequenzen und der Vergleich der Astlängen kann Hinweise auf mögliche Unterschiede bei den selektiven Bedingungen in den verschiedenen Arten geben. In Abbildung 22 ist eine Genealogie aus jeweils zehn *csd*-Sequenzen sowie einer *fem*-Sequenz der vier *Apis*-Arten gezeigt. Ein Vergleich der Länge der inneren Äste ergab für *Af-csd* deutlich reduzierte Werte im Vergleich zu den anderen drei *csd*-Sequenzgruppen ($P < 0,001$, zweiseitiger Mann-Whitney-U-Test). Dies deutet darauf hin, dass die *Af-csd*-Allele deutlich schneller entstehen und eine kürzere Koaleszenzzeit haben als für ein Gen unter balancierender Selektion angenommen wird (Takahata 1990) und wie es sich bei den anderen *Apis-csd*-Sequenzen auch findet (Hasselman et al. 2008b). Eine alternative Erklärung hierfür wäre genetischer Flaschenhals in der *A. florea*-Population (Richman et al. 1996), der ebenfalls zu einer Reduktion der Sequenzvariation und des Alters der Allele führen kann.

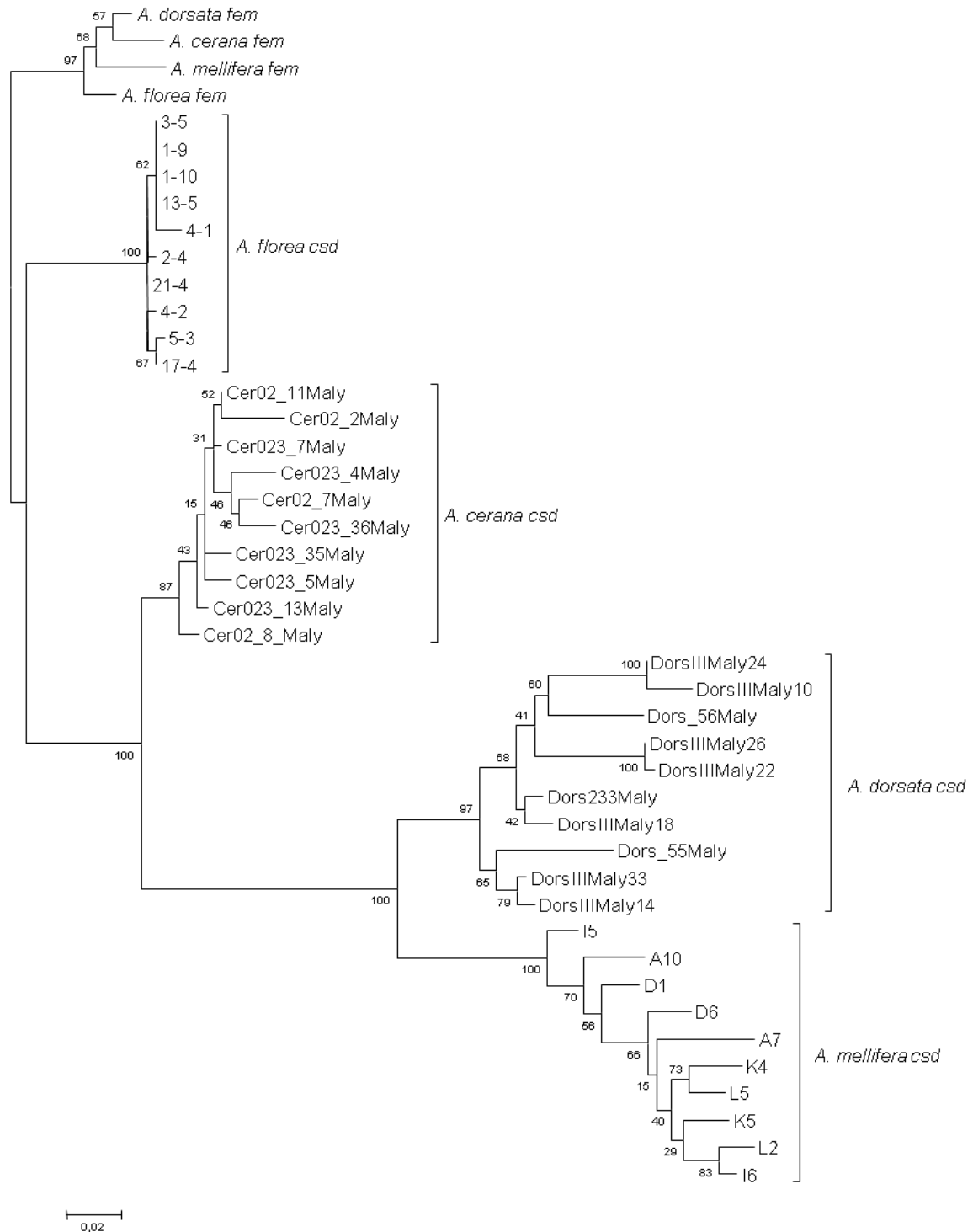


Abbildung 22: Genealogie verschiedener *csd*-Nukleotidsequenzen aus *A. mellifera*, *A. dorsata*, *A. cerana* und *A. florea* sowie einer *fem*-Sequenz der jeweiligen Arten (2.1.6). Die Genealogie ist erstellt auf Basis der Nukleotidunterschiede nach der Neighbor-Joining-Methode und dem Jukes-Cantor-Modell (2.2.8). Die Bootstrapwerte beziehen sich auf 500 Wiederholungen und sind an den inneren Ästen angegeben. Die Länge der Äste entspricht den Nukleotidsubstitutionen pro Position.

Hinweise auf einen genetischen Flaschenhals können durch das Modell von Takahata et al. (1992) gewonnen werden. Hierbei wird für die *csd*-Allele jeweils die höchste Zahl an synonymen und nichtsynonymen Unterschieden im paarweisen Vergleich mit der durchschnittlichen Anzahl an entsprechenden Polymorphismen verglichen (Tab. 12). Dieser Analyse liegt die Theorie zu Grunde, dass *csd*-Allele alle die gleiche Fitness haben und deshalb alle mit der gleichen Wahrscheinlichkeit erhalten bleiben oder verloren gehen. Deshalb wird erwartet, dass sie in ähnlichem In diesem Fall würde das Verhältnis der höchsten Zahl an Unterschieden zum Durchschnittswert in etwa $2(1-1/i)$ entsprechen (Takahata et al. 1992), wobei i die Zahl der analysierten Sequenzen ist.

Tabelle 12: Die durchschnittliche und größte Zahl an paarweisen synonymen (S) und nicht-synonymen (N) Unterschieden für je zehn *csd*-Sequenzen der jeweiligen Art. Die Ratio ist der entsprechende höchste Wert im Verhältnis zum Mittelwert. Die erwartete Ratio wird nach dem Koaleszenzmodell nach $2(1-1/i)$ berechnet (Takahata et al. 1992), wobei i die Zahl der analysierten Sequenzen ist, d.h. für diese Analyse ist $i = 10$.

		S	N
<i>A. mellifera</i>	Mittelwert	$4,8 \pm 1,06$	$14,6 \pm 2,39$
	höchster Wert	$7,8 \pm 3,17$	$24,5 \pm 5,85$
	Ratio	1,6	1,7
	erwartete Ratio	1,8	1,8
<i>A. cerana</i>	Mittelwert	$3,4 \pm 1,15$	$6,6 \pm 1,32$
	höchster Wert	$6,4 \pm 2,74$	$11,3 \pm 3,16$
	Ratio	1,9	1,7
	erwartete Ratio	1,8	1,8
<i>A. dorsata</i>	Mittelwert	$4,0 \pm 1,13$	$18,2 \pm 2,46$
	höchster Wert	$9,1 \pm 3,16$	$31,5 \pm 6,64$
	Ratio	2,3	1,7
	erwartete Ratio	1,8	1,8
<i>A. florea</i>	Mittelwert	$0,2 \pm 0,18$	$2,0 \pm 0,76$
	höchster Wert	$1,0 \pm 1,07$	$6,1 \pm 2,53$
	Ratio	5,0	3,1
	erwartete Ratio	1,8	1,8

Es zeigt sich, dass dies, wie schon von Hasselmann et al. (2008b) beschrieben, für *Am-csd*, *Ad-csd* und *Ac-csd* der Fall ist. Für *Af-csd* ist das tatsächliche Verhältnis mit 5,0 bei den synonymen und 3,1 bei den nichtsynonymen Polymorphismen deutlich größer als die

erwartete Ratio von 1,8. Dies könnte ein Hinweis auf einen genetischen Flaschenhals und eine anschließende Entstehung neuer Allele seine, wodurch die *Af-csd*-Allele ein anderes genealogisches Verwandtschaftsverhältnis aufweisen würden als neutrale Allele.

Unterschiede in der evolutionären Geschichte der *csd*-Gene können anhand der synonymen und nichtsynonymen Substitutionen (b_n / b_s) entlang der Äste einer Genealogie analysiert werden. Die Genealogie in Abbildung 23 wurde auf Basis der Nukleotidsubstitutionen aus Abbildung 18 erstellt. Neben den relativen Werten wurden zudem für jeden Ast die absoluten Substitutionswerte (a_n / a_s) berechnet. Das Verhältnis von synonymen zu nichtsynonymen Substitutionen an den einzelnen Ästen der Genealogie gibt Hinweise darauf, welche Art von Selektion zum jeweiligen Zeitpunkt auf ein Gen gewirkt hat.

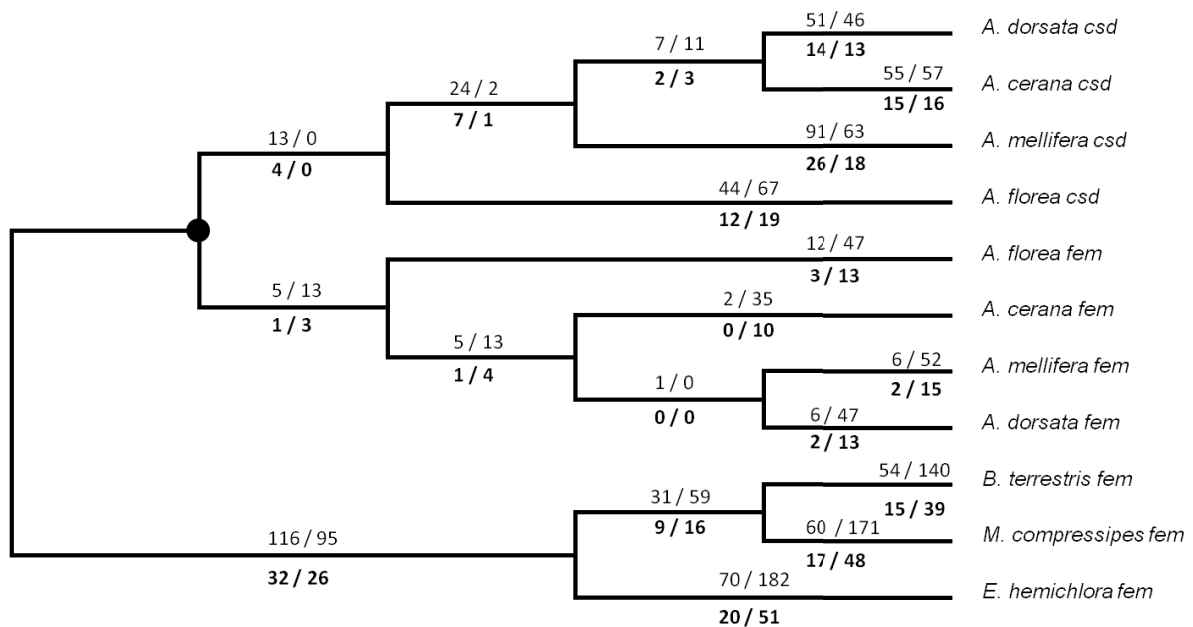


Abbildung 23: Genealogie auf Basis der Nukleotidsubstitutionen, abgeleitet von Abbildung 18. Für jeden Ast wurden die nichtsynonymen und synonymen Substitutionen pro Position (b_n / b_s) berechnet (Programm Bn-Bs). Die Werte sind oberhalb des jeweiligen Astes angegeben. Unterhalb jedes Astes ist die absolute Zahl an Substitutionen (fett gedruckt) angegeben (a_n / a_s). Der Punkt stellt den Zeitpunkt der Genduplikation dar.

Hasselmann et al. (2008a) zeigten, dass *csd* in *A. mellifera*, *A. cerana* und *A. dorsata* während seiner Neofunktionalisierung unter positiver Selektion stand. Diese zeigt sich durch eine erhöhte Rate von nichtsynonymen zu synonymen Substitutionen, da positive Selektion zur Fixierung von nichtsynonymen Aminosäureunterschieden beiträgt. Nach der Entstehung von *csd* (Abb. 23, durch Punkt markiert) findet sich ein a_n/a_s -Verhältnis von 4/0. Dieser Wert ist zu niedrig um statistisch ausgewertet werden zu können. Nimmt man den gemeinsamen Ast von *Am-csd*, *Ac-csd* und *Ad-csd* hinzu, zeigt sich jedoch, dass das hier gefundene Verhältnis (7/1 bzw. $4/0 + 7/1 = 11/1$) keine erhöhte Rate an nichtsynonymen Substitutionen zeigt ($P < 0,1$; zweiseitiger Fisher's-Exact-Test). Vermutlich entsteht diese Diskrepanz zu den in Hasselmann et al. (2008a) publizierten Werten durch die Hinzunahme von *csd* aus der basal stehenden Art *A. florea*. In einer Genealogie ohne *A. florea* konnte die auf *csd* wirkende, positive Selektion von Hasselmann et al. (2008a) eindeutig nachgewiesen werden.

Bei den Werten für die einzelnen *csd*-Sequenzen finden sich in Abbildung 23 keine signifikanten Unterschiede: *Ad-csd* = 27/17, *Ac-csd* = 28/20, *Am-csd* = 37/19 und *Af-csd* = 16/19 ($P > 0,05$, zweiseitiger Fisher's-Exact-Test). Dies deutet auf eine ähnliche evolutionäre Geschichte für alle *csd*-Gene hin und spricht somit dafür, dass *Af-csd* ebenfalls nach der Neofunktionalisierung unter positiver Selektion stand. Unterstützt wird dies durch den Vergleich zu *fem*. Auch hier stellt sich das Problem, dass die Werte, die sich an den Ästen direkt nach der Duplikation finden, zu niedrig sind, um statistisch ausgewertet werden zu können (4/0 : 1/3). Nimmt man den gemeinsamen Ast von *csd* bzw. *fem* von *A. mellifera*, *A. cerana* und *A. dorsata* hinzu ($4/0 + 7/1 : 1/3 + 1/4$), zeigt sich ein signifikanter Unterschied ($P < 0,01$; zweiseitiger Fisher's-Exact-Test). Dies ist ein deutlicher Hinweis auf eine unterschiedliche evolutionäre Geschichte von *csd* und *fem*, welches unter negativer Selektion steht (3.3.2.1 und Hasselmann et al. 2010). Die Ergebnisse dieser Analyse unterstützen die Hypothese, dass *Af-csd* nach der Neofunktionalisierung unter positiver Selektion stand.

3.3 Die Nukleotidsequenzrevolution des Gens *feminizer* in verschiedenen Hymenopteren

3.3.1 Vergleich der Sequenzen des Gens *feminizer* in verschiedenen Hymenopteren

Das Gen *feminizer* (*fem*) ist das Schaltergen der Geschlechtsdeterminationskaskade in *A. mellifera*. In Weibchen vermittelt es das alternative Spleißen von *doublesex* und aktiviert so die weibchenspezifische Geschlechtsdifferenzierung (Gempe et al. 2009). In Männchen findet sich in der mRNA von *fem* ein geschlechtsspezifisches Exon, das ein Stoppkodon enthält. Dies führt zu einem verfrühten Translationsabbruch (Hasselman et al. 2008a). So liegt in Männchen kein aktives Fem-Protein vor und der voreingestellte männchenspezifische Entwicklungsweg wird initiiert. Während das Gen *complementary sex determiner* in der *Apis*-Linie neu entstanden ist (Hasselman et al. 2008a), ist sein Vorläufergen *feminizer* konserviert und ein Ortholog des in *Drosophila melanogaster* gefundenen *transformer*-Gens. Die Homologie zwischen *fem* und *tra* lässt sich über einen Zeitraum von 300 Mio. Jahren nachweisen (Gempe et al. 2009). Bisher konnte *fem* neben der Honigbiene auch in der stachellosen Biene *Melipona compressipes* und der Erdhummel *Bombus terrestris* nachgewiesen werden (Hasselman et al. 2008a, Hasselman et al. 2010).

Um weitere Einblicke in die Sequenzrevolution von *fem* in anderen Hymenopteren-Arten zu gewinnen, wurde die mRNA-Sequenz des Gens im Rahmen dieser Arbeit in zwei weiteren Hummelarten (*Bombus lapidarius* und *B. hypnorum*) sowie in der Prachtbiene *Euglossa hemichlora* und der parasitischen Wespe *Nasonia vitripennis* sequenziert. Zudem konnte die Sequenz des Gens aus der Genomsequenz von *B. impatiens* abgeleitet werden.

3.3.1.1 Sequenzierung der *fem*-mRNA aus verschiedenen Hymenopteren

Das Gen *feminizer* wurde bislang nur in einer Hummelart, der Erdhummel *Bombus terrestris*, sequenziert (Hasselman et al. 2010). Von dieser Sequenz ausgehend wurden Oligonukleotide entworfen, mit welchen es aus cDNA der beiden Arten *Bombus lapidarius* (Steinhummel) und *Bombus hypnorum* (Baumhummel) sequenziert werden konnte (2.2.7.2).

Für *B. hypnorum* gelang die Sequenzierung des *fem*-Gens auf Grund der schlechten cDNA-Qualität nicht vollständig (2.2.7.2). Auf genomischer Ebene konnte die Sequenz weitgehend vervollständigt werden, es fehlt aber noch ein vermutlich 48 bp langer Bereich des Exons 9. Der fehlende Sequenzbereich wird entsprechend auch bei den anderen Sequenzen nicht in die weiteren Analysen mit einbezogen. Für *B. lapidarius* gelang es, den vollständigen Leserahmen zu sequenzieren.

Die *fem*-Sequenz der nordamerikanischen Hummel *B. impatiens* konnte anhand der in der NCBI-Datenbank veröffentlichten Genomsequenz (NCBI-GenBank-Nummer AE-QM000000000.1) identifiziert werden. Ausgehend von einem Alignment der cDNA-Sequenz von *Bt-fem* mit der *Bi*-Genomsequenz war es möglich, die Exonsequenzen des Gens von *B. impatiens* abzuleiten (6.3.4).

Zudem konnten mittels der Genomsequenzen von *B. terrestris* (AELG000000000.1) und *B. impatiens* die Exon- und Intronstruktur des *fem*-Gens für beide Arten nachvollzogen werden. Sie stimmt in beiden Arten überein und ist in Abbildung 25 schematisch dargestellt. In *Bombus* besteht das Gen *fem* aus 13 Exons, wovon die Exons 4, 5, 6 und der 5'-Bereich von Exon 7 männchenspezifisch sind. Exon 1 und der 5'-Bereich von Exon 2 enthalten die 5'-UTR, der 3'-Bereich des Exons 13 die 3'-UTR. Bei der Genomsequenz der beiden Hummelarten handelt es sich um nur teilweise assemblierte Konsensussequenzen (Contigs). Die Sequenz des *fem*-Gens war bei beiden Arten auf zwei verschiedenen Contigs lokalisiert, die jeweils einen Bereich des Introns 4 enthalten. Somit ist die Länge dieses Introns nicht bekannt, was in der Abbildung durch Schrägstriche angedeutet ist.

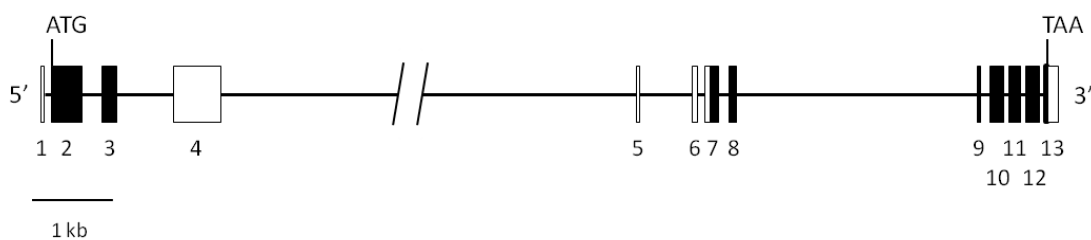


Abbildung 25: Struktur des Gens *fem* aus der Erdhummel *B. terrestris* und der nordamerikanischen Hummel *B. impatiens*. Die Exons sind als Kästchen dargestellt, die Introns als Linie. Die geschlechtsunspezifischen Exons sind in schwarz dargestellt, die männchenspezifischen in weiß. Die 5'- und 3'-UTRs sind ebenfalls in weiß dargestellt. Die Länge des Introns 4 ist nicht genau bekannt. Die Gesamtlänge des Gens ist aus diesem Grund ebenfalls nicht bekannt, der hier dargestellte Bereich umfasst ca. 12,3 kb.

Die Sequenzierung von *fem* in *E. hemichlora* erfolgte ausgehend von einem Sequenzfragment (K. Gruber Gonzales, pers. Mitteilung) wie unter 2.2.7.3 dargestellt. Die Ver-

vollständigung des 5'-terminalen Bereichs gelang trotz mehrfacher RACE-Experimente nicht. Mit einem in *A. cerana-fem* am Startkodon bindenden Oligonukleotid konnte der vordere Bereich des Leserahmens sequenziert werden. Ein Vergleich von *Eh-fem* mit den anderen *fem*-Sequenzen zeigt, dass vermutlich die ersten 34 Nukleotide des Leserahmens nicht sequenziert wurden (6.3.8). Zudem ist die 5'UTR noch nicht bekannt.

Neben verschiedenen corbiculaten Apidae konnte das *fem*-Gen auch in der parasitischen Wespe *Nasonia vitripennis* (Pteromalidae) sequenziert werden. Ausgehend von einer als „*csd-like*“ beschriebenen, unvollständigen Sequenz in der NCBI-Datenbank gelang mit den Oligonukleotiden S-18 und S-19 (Tab. 1) die Amplifikation des vollständigen Leserahmens. Es konnte sowohl die weibliche als auch die männliche Spleißform sequenziert werden, welche im Anhang unter Abschnitt 6.3.7 dargestellt ist.

In Abbildung 26 ist die Exon-Intron-Struktur des Gens dargestellt. Sie wurde von der in der NCBI-Datenbank veröffentlichten Genomsequenz abgeleitet (The Nasonia Genome Working Group 2010). Da nur der Leserahmen sequenziert wurde, sind nur die kodierenden Exons sowie das männchenspezifische Exon dargestellt, die untranslatierten Bereiche sind nicht bekannt. Der hier dargestellte genomische Bereich ist 6,5 kb lang.

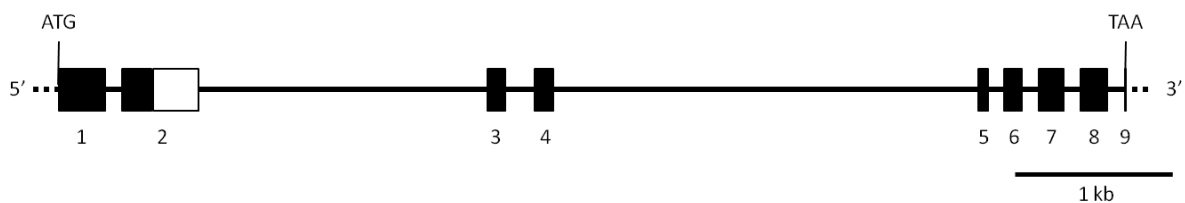


Abbildung 26: Struktur des Gens *fem* aus *Nasonia vitripennis*. Die Exons sind dargestellt soweit bekannt, die UTR-Sequenzen sind bislang unbekannt. In schwarz sind die geschlechtsunspezifischen, kodierenden Exons dargestellt, in weiß das männchenspezifische Exon. Der dargestellte Bereich ist 6,5 kb lang.

3.3.1.2 Analyse der *fem*-Sequenzen verschiedener Hymenopteren

Das Gen *fem* ist in der *Apis*-Gattung hoch konserviert und zeigt eine große Sequenzhomologie zu *fem* in *B. terrestris* und *M. compressipes* (Hasselman et al. 2008a). Dies lässt die Annahme zu, dass seine Sequenz innerhalb der Apidae konserviert ist. Anhand der im Rahmen dieser Arbeit sequenzierten *fem*-mRNA-Sequenzen soll dies genauer untersucht werden. Mit den Aminosäuresequenzen von Fem der Arten *A. mellifera*, *A. cerana*,

A. dorsata, *A. florea*, *B. terrestris*, *B. impatiens*, *B. lapidarius*, *B. hypnorum*, *M. compressipes* und *E. hemichlora* wurde ein Alignment erstellt (6.3.8 im Anhang). Hierbei zeigt sich, dass die hohe Sequenzhomologie ein zuverlässiges Alignment ermöglicht.

Der hohe Grad der Konservierung von *fem* innerhalb der Apidae lässt auch die Vermutung zu, dass sich die verschiedenen *fem*-Sequenzen in einer Genealogie entsprechend der Verwandtschaft der Arten anordnen. Die Genealogie in Abb. 27 wurde auf Basis der synonymen Nukleotidunterschiede erstellt, welche ein Maß für die Verwandtschaft zwischen Sequenzen sind. Wie angenommen, geben auch die synonymen Substitutionen zwischen den *fem*-Sequenzen die Verwandtschaft zwischen den Arten wieder. Deutlicher stellt sich dies noch in Abbildung 27 dar, wo durch *Nv-fem* eine Außengruppe hinzugekommen ist.

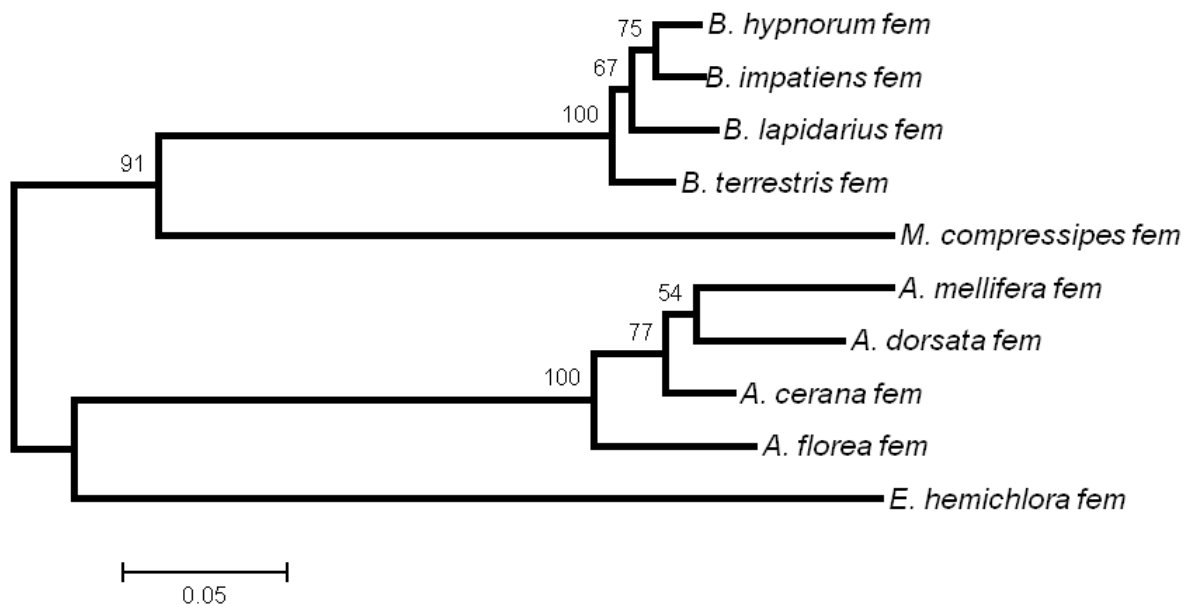


Abbildung 27: Genealogie der *fem*-Nukleotidsequenzen der corbiculaten Apidae *A. mellifera*, *A. cerana*, *A. dorsata* und *A. florea*, *B. terrestris*, *B. impatiens*, *B. lapidarius*, *B. hypnorum*, *M. compressipes* und *E. hemichlora*. Die phylogenetische Analyse erfolgte nach der Neighbor-Joining-Methode und dem Jukes-Cantor-Modell anhand der synonymen Substitution; die Astlänge gibt die synonymen Substitutionen pro Position an. Die Bootstrapwerte beziehen sich auf 500 Wiederholungen.

Die parasitische Wespe *Nasonia vitripennis* gehört in die Familie der Pteromalidae. Diese hat sich schon vor ca. 180 Mio. Jahren von der Linie der Apidae getrennt. In Folge dieses langen evolutionären Zeitraums wird erwartet, dass *Nv-fem* deutlich divergent zu den Apidae-*fem* ist. Verhulst et al. (2010a) postulierten, dass bei *fem* primär die funktionellen Domänen konserviert werden. Entsprechend würde zwischen *Nv-fem* und den Apidae-*fem*

die größte Sequenzhomologie im für die RS- und P-Domäne kodierenden 3'-terminalen Bereich erwartet.

Dies bestätigt sich auch im Alignment der Fem-Sequenzen. Auf Grund der Sequenzunterschiede ist ein Alignment zwischen der *Nasonia*-Sequenz und den Apidae-Sequenzen nur auf Aminosäureebene und nur im Bereich der RS- und P-Domäne zuverlässig möglich. Die Sequenzdivergenz wird zusätzlich auch in einer auf Aminosäureebene und für den C-terminalen Bereich erstellten Genealogie deutlich. *Nv*-Fem separiert sich, wie auf Grund der evolutionären Zeiträume erwartet, deutlich von den Apidae-Fem-Sequenzen.

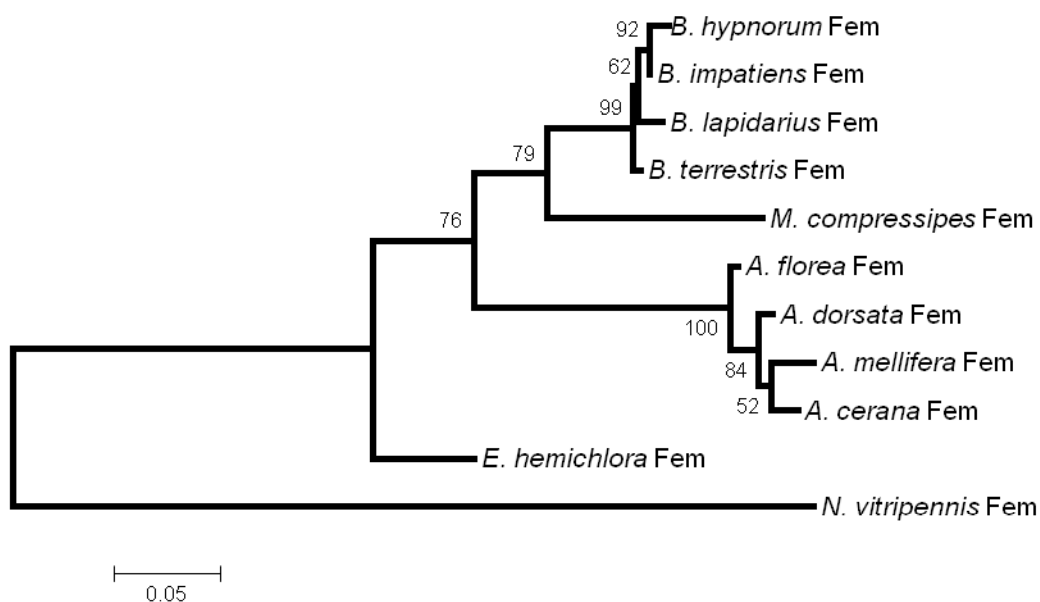


Abbildung 28: Genealogie der *fem*-Aminosäuresequenzen verschiedener corbiculaten Apidae und von *Nasonia vitripennis* (Pteromalidae). Es wurde ein ca. 140 AS langes Alignment des C-Terminus verwendet. Die Genealogie wurde nach der Maximum-Likelihood-Methode und dem JJT matrixbasierenden Modell erstellt. Die Astlänge gibt die Aminosäuresubstitutionen pro Position an, die Bootstrapwerte beziehen sich auf 500 Wiederholungen.

3.3.1.3 Sequenzanalyse des männchenspezifischen Exons von *fem*

In *A. mellifera-fem* existiert im 3'-Bereich des Exons 3 ein Stoppkodon, welches zu einem frühen Translationsstopp führt. In Weibchen wird dieser Bereich des Exons 3 sowie die Exons 4 und 5 herausgespleißt, wodurch ein längerer Leserahmen entsteht und ein funktionelles Protein translatiert werden kann (Hasselman et al. 2008a). Alternatives

Spleißen wird auch für das *fem*-Ortholog *transformer* (*tra*) in *Drosophila melanogaster* beschrieben (Boggs et al. 1987, Sosnowski et al. 1989). Auch hier wird die mRNA in Weibchen alternativ gespleißt, wodurch ein frühes Stoppkodon entfernt wird und ein langer Leserahmen entsteht. Dies lässt die Vermutung zu, dass das geschlechtsspezifische Spleißen bei *fem* konserviert ist. Schulte (2008) wies eine männchenspezifische Spleißform mit einem verfrühten Stoppkodon in *B. terrestris* nach (6.3.5). Wie in Abschnitt 3.3.1.1 beschrieben, konnten im Rahmen dieser Arbeit auch bei *N. vitripennis* entsprechende geschlechtsspezifische Spleißformen sequenziert werden. Somit konnte nicht nur gezeigt werden, dass *fem* zwischen den Hymenopterenfamilien Apidae und Pteromalidae in seiner Sequenz und seiner Domänenstruktur konserviert ist (3.3.1.2), sondern auch, dass die Existenz des männchenspezifischen Exons in beiden Familien erhalten wurde. Auf Grund der Homologie in den in beiden Geschlechtern vorkommenden Sequenzbereichen lässt sich die Hypothese aufstellen, dass auch die männchenspezifischen Exons in ihrer Sequenz konserviert sind. Um dies zu untersuchen, wurde die Sequenzidentität dieses Bereichs zwischen den drei Arten untersucht. Hierbei zeigt sich jedoch, dass es keine Sequenzübereinstimmung gibt. Auch variiert die Länge dieses Bereichs zwischen den einzelnen Arten um einige hundert Basenpaare. Während er in *A. mellifera* 1309 bp lang ist, sind es in *B. terrestris* nur 875 bp und in *N. vitripennis* lediglich 286 bp. Im Gegensatz zu den in ihrer Nukleotidsequenz zwischen den Arten konservierten geschlechtsunspezifischen Exons, wurde die Sequenz der männchenspezifischen Exons nicht erhalten. Die männchenspezifischen Exons sind durch die Stoppkodons nicht kodierend. Somit unterliegen sie möglicherweise einem deutlich geringeren selektiven Druck und können schneller evolvieren als die umliegenden Genbereiche, so dass Ähnlichkeiten zwischen den Arten hier verschwunden sind.

3.3.2 Die Evolution von *feminizer* unter Einfluss seines Schwestergens *complementary sex determiner*

Einige der im Rahmen dieser Arbeit generierten Daten wurden bereits publiziert:

ORIGIN OF A FUNCTION BY TANDEM GENE DUPLICATION LIMITS THE EVOLUTIONARY CAPABILITY OF ITS SISTER COPY

Martin Hasselmann, Sarah Lechner, Christina Schulte, and Martin Beye

Proc. Natl. Acad. Sci. USA, 2010, Vol. 107 No. 30, S. 13378-13383

Die Funktion des Gens *feminizer* als Faktor der Geschlechtsdeterminationskaskade in Hymenopteren hat sich seit über 300 Mio. Jahren erhalten (Gempe et al. 2009). Die *fem*-prä-mRNA wird geschlechtsspezifisch gespleißt, woraufhin das Fem-Protein das nächste Gen der Kaskade, *doublesex* aktivieren kann (Hasselmann et al. (2008a)). In der Honigbiene haben sich die selektiven Zwänge, unter denen *fem* steht, durch die Neuentstehung von *csd* deutlich verändert. Es bedeutet die direkte Nachbarschaft zu einem Gen unter balancierender Selektion sowie einen weiteren Interaktionspartner in der Geschlechtsdifferenzierungskaskade. Wie stark sich der Hitchhiking-Effekt durch die auf das benachbarte *csd* wirkenden balancierenden Selektion die Evolution von *fem* auswirkt und welchen funktionellen Einfluss *csd* auf *fem* hat, wurde in der Publikation von Hasselmann et al. (2010) untersucht und wird im folgenden dargestellt und erläutert.

3.3.2.1 Csd hat einen funktionellen Einfluss auf die Sequenzevolution seines Paralog *fem*

In *Apis* ist mit Csd ein weiterer Interaktionspartner für Fem hinzugekommen, weshalb Fem in *Apis* im Vergleich zu den Nicht-*Apis*-Arten vermutlich unter einem stärkeren Selektionsdruck steht. Makino et al. (2006) postulieren eine langsamere Evolutionsrate von Proteinen mit vielen Interaktionspartnern im Vergleich zu solchen mit weniger Partnern. Die Analyse der nichtsynonymen und synonymen Substitutionen entlang der Genealogie von *csd* in *A. mellifera*, *A. cerana*, *A. dorsata* sowie *fem* in diesen *Apis*-Arten und *B. terrestris*, *M. compressipes* und *E. hemichlora* zeigte, dass *fem* in *Apis* sowohl vor als auch nach der Entstehung von *csd* unter negativer Selektion stand. Jedoch ist in *Apis-fem*

nach der Duplikation eine niedrigere n/s-Ratio zu finden, was auf eine Intensivierung des Selektionsdrucks hindeutet (Hasselmann et al. (2010)).

Ein weiterer Hinweis auf die Verstärkung des Selektionsdrucks findet sich beim Vergleich von *fem* mit den beiden benachbarten Genen GB13727 und GB11211 (Abb. 29A). Hierfür wurden im Rahmen der vorliegenden Arbeit genomische Sequenzen der drei Gene aus mehreren Individuen sequenziert (Tab. 2.1.4). Für jedes dieser Gene wurde die Zahl der akkumulierten synonymen und nichtsynonymen Polymorphismen berechnet (Abb. 29B). Die Gene GB13727 und GB11211 zeigen im Gegensatz zu *fem* in *Apis* kein reduziertes Verhältnis von nichtsynonymen und synonymen Substitutionen im Vergleich zu den Nicht-*Apis*-Arten. Dies bedeutet, dass es nicht alleine der Hitchhiking-Effekt durch die auf *csd* wirkende, balancierende Selektion sein kann, der in *fem* zu einer Reduktion der nichtsynonymen Substitutionen im Verhältnis zu den synonymen führt (Abb. 29B). Eine stärkere negative Selektion könnte auch hier die Ursache für die reduzierte n/s-Ratio in *fem* sein. Es wurden im Rahmen der vorliegenden Arbeit zudem Fragmente von *fem*, GB13727 und GB11211 aus cDNA von Nicht-*Apis*-Arten

(*M. compressipes*, *B. terrestris* und *E. hemichlora*) amplifiziert und sequenziert. Ein Vergleich mit den entsprechenden Sequenzen aus *Apis* zeigt in Übereinstimmung mit den oben beschriebenen Ergebnissen, dass *fem* in den Nicht-*Apis*-Arten unter einem

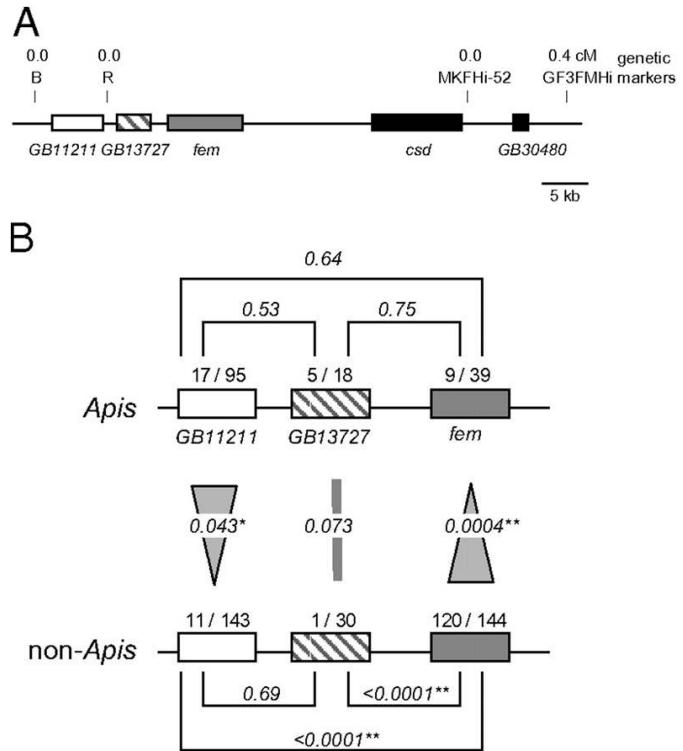


Abbildung 29: Anordnung der Gene GB13737, GB11211, *fem* und *csd* in der Honigbiene *A. mellifera* und ihre Evolution auf Ebene der Nukleotidpolymorphismen in *Apis* und anderen corbiculaten Apidae. (A) Schematische Darstellung des SDL. GB13727 liegt 3,5 kb und GB11211 10,3 kb strangaufwärts von *fem* (Startkodon). Keines der beiden Gene ist in die Geschlechtsdeterminierung involviert. (B) Zahl der akkumulierten nichtsynonymen (n) und synonymen (s) Veränderungen in der Evolution der Gene GB13727, GB11211 und *fem* in *Apis* und Nicht-*Apis*-Arten. Die errechneten n/s-Verhältnisse sind über den Kästchen gezeigt und die Gennamen sind unterhalb angegeben. Die n/s-Verhältnisse wurden durch den zweiseitigen Fisher's-Exact-Test verglichen und die Wahrscheinlichkeiten sind als kursive Zahlen über den Linien angegeben, die den jeweiligen Vergleich darstellen. (** $P < 0,01$; * $P < 0,05$). Aus Hasselmann et al. (2010).

schwächeren Selektionsdruck steht als in *Apis*, so dass sich mehr nichtsynonyme Mutationen im Verhältnis zu synonymen Substitutionen akkumuliert haben.

3.3.2.2 Die Evolution von *fem* wird durch die Nachbarschaft zu *csd* beeinflusst

Die Selektion, die auf ein Gen wirkt, beeinflusst indirekt die benachbarten Regionen. Rekombination kann diesen Effekt abschwächen. Die Rekombination ist in *A. mellifera* jedoch innerhalb des SDL unterdrückt, weshalb *fem* stärker an *csd* gekoppelt ist (Hasselmann und Beye 2006). Im Rahmen der vorliegenden Arbeit wurden Fragmente von *Am-fem*, *Am-csd* und *Bt-fem* amplifiziert und sequenziert. Die Untersuchung der Nukleotiddiversität in diesen Sequenzen ist in Tabelle 13 dargestellt.

Tabelle 13: Exon- und Intronpolymorphismenanalyse der Gene *fem* und *csd*. ND = nicht determiniert. Nukleotiddiversität der *fem*- und *csd*-Gene aus denselben *Apis*-Haplotypen ($\pm$ Standardfehler) wurde für die genomische Region der Exons 8-11 in *fem* und Exons 6-9 in *csd* berechnet. Aus Hasselmann et al. 2010

	<i>A. mellifera fem</i>	<i>B. terrestris fem</i>	<i>A. mellifera csd</i>
Exon			
Anzahl der Sequenzen	30	23	28
Anzahl der Positionen (ohne Lücken)	439	492	330
Anzahl der Polymorphismen	15	6	93
Anzahl der syn. Substitutionen	7	3	22
Anzahl der nichtsyn. Substitutionen	7	3	56
π_{total}	$0,0034 \pm 0,0010$	$0,0036 \pm 0,0016$	$0,0536 \pm 0,0066$
π_{syn}	$0,0084 \pm 0,0036$	$0,0059 \pm 0,0036$	$0,0646 \pm 0,0142$
$\pi_{nichtsxn}$	$0,0014 \pm 0,0050$	$0,0029 \pm 0,0016$	$0,0508 \pm 0,0087$
$\pi_{nichtsxn}/\pi_{syn}$	0,16	0,49	0,79
Intron			
Anzahl der Positionen (ohne Lücken)	539	262	162
Anzahl der Polymorphismen	48	5	39
π	$0,0151 \pm 0,0026$	$0,0023 \pm 0,0017$	$0,0380 \pm 0,0072$
genomweite Durchschnittsdiversität	$0,0055 \pm 0,0012$	ND	

Das Gen *csd* zeigt eine 16-fach erhöhte Nukleotiddiversität im Vergleich zum genomweiten Durchschnitt. Die π_{syn} -Werte von *Apis-fem* zeigen dagegen keinen Unterschied zum genomweiten Durchschnitt ($Z = 0,76$; $P > 0,08$) und sind im Vergleich zu *csd*

signifikant reduziert ($Z = 3,8$; $P < 0,001$, zweiseitiger Z-Test). Es finden sich jedoch signifikant erhöhte π -Werte in den *fem*-Introns im Vergleich zum genomweiten Durchschnitt ($Z = 3,2$; $P < 0,0001$). Vermutlich ist die Zahl der synonymen Positionen in den Exons zu gering, um Unterschiede in der Zahl der Nukleotidpolymorphismen messen zu können. In den intronischen Bereichen, in denen eine etwa viermal höhere Zahl an neutralen Polymorphismen analysiert werden kann, zeigen sich deutlich die Spuren des Hitchhiking-Effekts.

3.4 Die *fem*-Duplikation *twin of feminizer* in der Erdhummel *Bombus terrestris*

Die Geschlechtsdetermination bei der Erdhummel *Bombus terrestris* erfolgt wie in der Honigbiene nach dem Mechanismus der komplementären Geschlechtsbestimmung (Cook und Crozier 1995). Während das Ortholog des Gens *feminizer* in der Erdhummel sowohl in der weiblichen wie auch der männlichen Spleißform nachgewiesen wurde (Schulte 2008), ist der multiallelische Faktor, der im heterozygoten Zustand die weibliche Entwicklung initiiert, bisher unbekannt. Es handelt sich hierbei wie bei *csd* in *A. mellifera* um einen einzelnen Locus (Duchateau et al. 1994, Gadau et al. 2001). In Vorarbeiten ergab sich bei Southern-Blot-Hybridisierungen mit *fem*-spezifischen Sonden eine zweite, schwache Bande. Im Folgenden soll untersucht werden, ob in *Bombus* ebenfalls eine *fem*-Duplikation existiert und ob es sich bei dieser, wie im Fall von *Am-fem*-Duplikation *csd* um den gesuchten, komplementären Geschlechtsdeterminationsfaktor handeln könnte.

3.4.1 Das Gen *twin of feminizer* ist eine Duplikation von *feminizer*

In Vorarbeiten (C. Schulte) wurde ein 867 bp langes Sequenzfragment gefunden, das 50 % Nukleotidhomologie zu *Bt-fem* aufweist. Die mRNA-Sequenz des Gens konnte nie vollständig sequenziert und das Fragment nur in einer einzelnen cDNA-Probe nachgewiesen werden, so dass davon ausgegangen werden muss, dass es sich hierbei nicht um ein funktionelles Gen handelt. Ausgehend von dieser Sequenz, wurde in einer *Bombus-terrestris*-EST-Datenbank (Expressed Sequence Tags, Gene Robinson, Brielle Fischman) ein weiteres Fragment gefunden, das 85 % Nukleotidsequenzhomologie zu *Bt-fem* aufweist. Mit den Oligonukleotiden S-65 und S-66 gelang es, das unter der Nummer *terrestris.fna.Contig17498* gefundene Fragment aus cDNA-Proben zu amplifizieren und es anschließend durch eine 5'- sowie eine 3'-RACE-PCR vollständig zu sequenzieren (2.2.7.4).

Die mRNA des Gens ist 1915 bp lang und kodiert für ein 440 AS langes Protein. Seine Nukleotidsequenz stimmt innerhalb des Leserahmens mit der von *Bt-fem* zu 84 % überein, die Aminosäuresequenz zu 74 % (6.4.3). Das Gen, bei dem es sich offensichtlich um eine *fem*-Duplikation handelt, wird deshalb im Weiteren als *twin of feminizer* (*tof*) bezeichnet. Ein Vergleich der Struktur beider Gene auf Nukleotid- und Proteinebene ist in Abbildung 30 dargestellt. Wie für *Fem* konnte auch für *Tof* eine N-terminale, argininreiche Domäne sowie C-terminal eine arginin-serinreiche und

eine prolinreiche Domäne vorhergesagt werden (ProSite, InterProScan, 2.1.7). Die arginin-serinreiche Domäne ist ein charakteristisches Merkmal für ein SR-Typ-Protein. Diese Proteine spielen eine wichtige Rolle bei der Erkennung von Spleißstellen und der Spleißregulation von prä-mRNA (Blencowe et al. 1999). Möglicherweise nimmt Tof mit diesen Eigenschaften wie Fem eine Rolle in der Geschlechtsdeterminationskaskade ein.

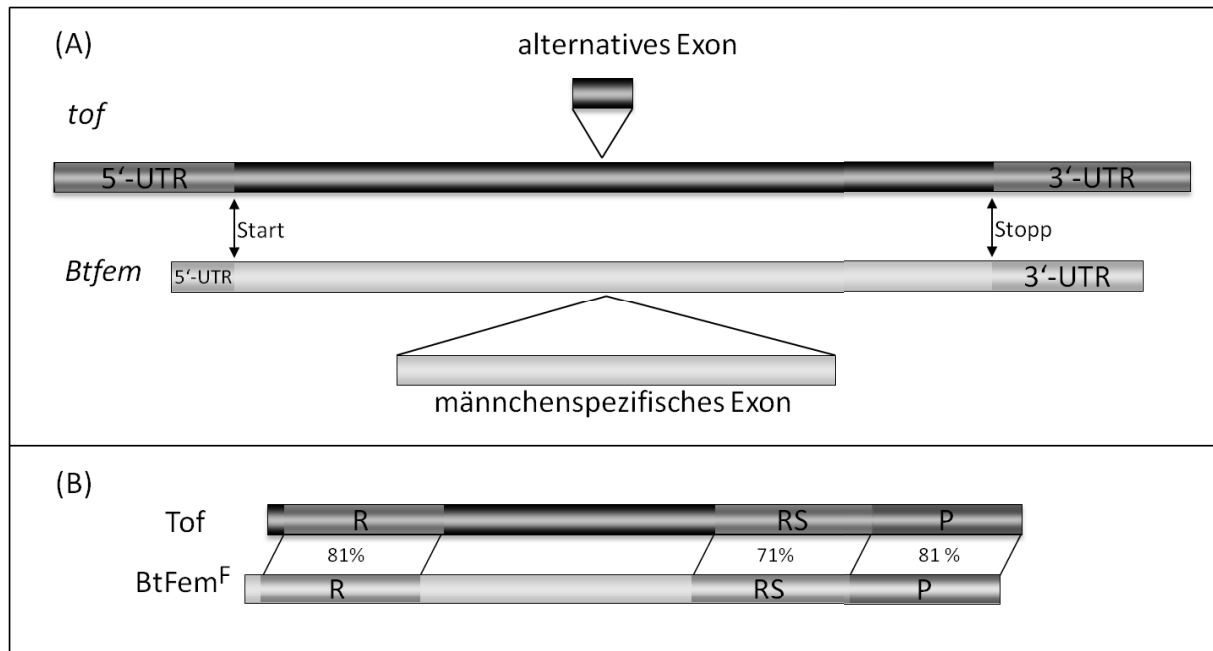


Abbildung 30: (A) Schematische Darstellung der mRNA-Struktur des Gens *tof* im Vergleich zur Struktur von *Bt-fem*. Die Position des alternativen Exons in *tof* und des männchenspezifischen *fem*-Exons (Seite 69) sind markiert. (B) Schematische Darstellung der Proteinstruktur von Tof und Bt-Fem. R = argininreiche Domäne, RS = arginin-serinreiche Domäne, P = prolinreiche Domäne. Die Übereinstimmung der Aminosäurepositionen der Domänen ist angegeben.

Mit 1317 bp bei *tof* und 1203 bp bei *Bt-fem* weisen die Leserahmen der beiden Gene eine ähnliche Länge auf (6.4.3 im Anhang), die UTR des *tof*-Gens ist jedoch sowohl 5' als auch 3' deutlich länger als die von *Bt-fem* (305 bp im Gegensatz zu 76 bp in der 5'-UTR, 326 bp gegen 119 bp in der 3'-UTR) und es finden sich keine Sequenzübereinstimmungen zwischen den UTRs der beiden Gene.

In der Gendatenbank von NCBI steht unter der Nummer AELG00000000.1 die Genomsequenz der Erdhummel *B. terrestris* zur Verfügung. Bisher handelt es sich hierbei um nicht annotierte, nur teilweise assemblierte Konsensussequenzen (Contigs). Mit dem BLAST-Algorithmus konnte die Sequenz des *tof*-Gens auf einem Contig detektiert werden. Auch für die nordamerikanische Hummelart *Bombus impatiens* stehen in der Datenbank assem-

blierte Konsensussequenzen zur Verfügung (Genbanknummer AEQM00000000.1). Hierdurch ergab sich die Möglichkeit herauszufinden, ob das Gen *tof* auch in einer anderen Hummelart existiert. Mit Hilfe des Blast-Algorithmus konnte *tof* auch in *B. impatiens* gefunden werden. Ausgehend von der *Bt-tof*-mRNA-Sequenz konnte die Exon-Intron-Struktur des Gens in beiden Arten ermittelt werden. Sie ist in Abbildung 31 dargestellt. *tof* besteht aus zehn Exons, von denen Exon 1 die 5'- und Exon 10 die 3'-UTR enthält. Die 3'-UTRs der beiden Gene weichen in ihrer Sequenz voneinander ab, so dass diese für *B. impatiens* nicht vollständig ermittelt werden konnte. Die Sequenz von *Bt-tof* liegt auf zwei verschiedenen, sich nicht überschneidenden Contigs, so dass die Länge des fünften Introns in *B. terrestris* nicht genau bekannt ist. Die Nukleotidsequenzen der beiden Gene sind innerhalb des Leserahmens zu 93,3 % identisch, die Aminosäuresequenzen zu 87,8 % (6.4.4). Anhand der Konsensussequenz konnte nicht geklärt werden, ob *tof* im Genom der Hummel in der Nähe seines Paralogs *fem* liegt. Es konnte jedoch ausgeschlossen werden, dass es direkt strangabwärts von *fem* liegt, wie dies beim *fem*-Paralog *csd* in *Apis* der Fall ist.

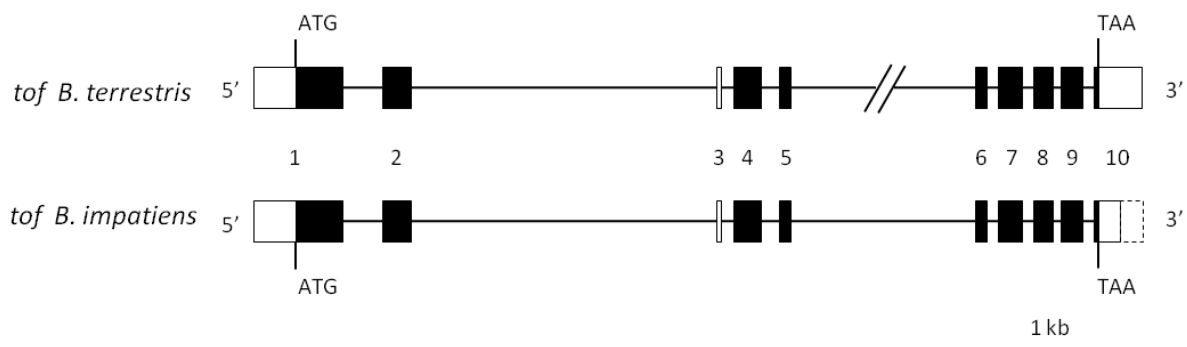


Abbildung 31: Struktur des Gens *tof* aus *B. terrestris* und *B. impatiens*. Die kodierenden Exons sind als schwarze, die UTRs und das alternative Exon (Exon 3, s.u.) als weiße Kästchen dargestellt, die Introns als Linie. *tof* hat zehn Exons. Die Länge von Intron 5 ist für *B. terrestris* nicht bekannt, da es sich auf zwei verschiedenen, nicht überschneidenden Contigs befindet. Bei *B. impatiens* ist die Länge von Exon 10 nicht bekannt, da die 3'-UTR-Sequenzen in den beiden Arten deutlich voneinander abweichen. Dieser unbekannte Bereich ist bei *B. impatiens* mit gestrichelten Linien dargestellt. Die Länge der genomischen Sequenz des Gens ist für beide Arten nicht exakt bekannt, der hier dargestellte Bereich ist ca. 7,5 kb lang.

An der Position, an der sich in *Bt-fem* das männchenspezifische Exon befindet, existiert auch in *tof* ein zusätzliches Exon (Abb. 30). Es handelt sich hierbei um das 42 bp lange Exon 3 (Abb. 31) und enthält wie das männchenspezifische Exon in *Bt-fem* Stoppkodons. Es würde somit auch zu einem verfrühten Translationsabbruch führen.

Ob dieses zusätzliche Exon in *tof* ebenfalls geschlechtsspezifisch ist, konnte bislang noch nicht nachgewiesen werden. Beide Varianten wurden bei adulten Tieren sowohl in Männchen als auch in Weibchen gefunden. Möglicherweise wird *tof*, wie dies auch bei *Am-fem* der Fall ist (Gempe et al. 2009), nur im Embryonal- und Larvenstadium spezifisch gespleißt. Ein Nachweis geschlechtsspezifischer Spleißformen wäre somit nur in diesen frühen Stadien und nicht in adulten Tieren möglich. Im Gegensatz zu Honigbienen ist es bei Hummeln nicht möglich, das Geschlecht eines Eis auf Grund seines Legeorts zu bestimmen. Molekulare Marker, die eine Geschlechtsbestimmung ermöglichen, konnten bislang nicht etabliert werden.

Eine Möglichkeit, die Funktion des Gens *tof* zu testen, sind RNAi-Experimente. Die Injektion von dsRNA führt durch DNA-Interferenz zum spezifischen Abbau der mRNA des Zielgens (Fire et al. 1998, Hannon 2002). Idealerweise entsteht hierbei ein „Knockdown“-Phänotyp, der eindeutige Rückschlüsse auf die Funktion des Gens zulässt. Sofern *tof* Teil der Geschlechtsdeterminationskaskade ist, würde durch Ausschalten seiner aktiven Form die Ausbildung von für das andere Geschlecht typischen Merkmalen erwartet werden, wie dies für *Am-fem* beschrieben wurde (Gempe et al. 2009). Die RNAi-Injektion ist für die Honigbiene etabliert (Beye et al. 2002), konnte aber für die Erdhummel im Rahmen dieser Arbeit nicht realisiert werden. Schon bei Vorversuchen ergaben sich Probleme bei der Entnahme der Eier aus dem Nest. Hummelköniginnen legen ihre Eier in geschlossene Brutnäpfchen aus Wachs ab (Heinrich 1979), bei deren Entnahme aus dem Nest und Öffnung es leicht zu einer Beschädigung der Eier kommen kann. Eine weitere Problematik ergab sich bei der Aufzucht der Larven, welche in Anlehnung an das für Honigbienenlarven verwendete Protokoll durchgeführt wurde (Osman Kaftanoglu, pers. Mitteilung). Zum einen gelang die Optimierung des Futters für die Erdhummellarven bislang nicht, was in einer hohen Sterblichkeitsrate der Larven resultierte. Zum anderen ergab sich ein Problem daraus, dass sich mehrere Eier in einem Brutnäpfchen befinden, die über mehrere Stunden gelegt worden sein können, so dass eine standardisierte Altersbestimmung der Eier nicht möglich ist und der Schlupfzeitpunkt, welcher für die Aufzucht sehr relevant ist, nicht bestimmt werden konnte.

3.4.2 Das Gen *twin of feminizer* ist in der *Bombus*-Linie entstanden

Die Existenz des Gens in *B. terrestris* wie auch *B. impatiens* wirft die Frage auf, ob *tof* in allen *Bombus*-Arten existiert und ob es auch in anderen, verwandten Gattungen zu finden ist. Zur Untersuchung der ersten Frage wurde eine Genealogie der *tof*-Sequenzen und der bekannten Apidae-*fem*-Sequenzen erstellt, welche Informationen über den Zeitpunkt der Entstehung von *tof* liefern kann (Abb. 32).

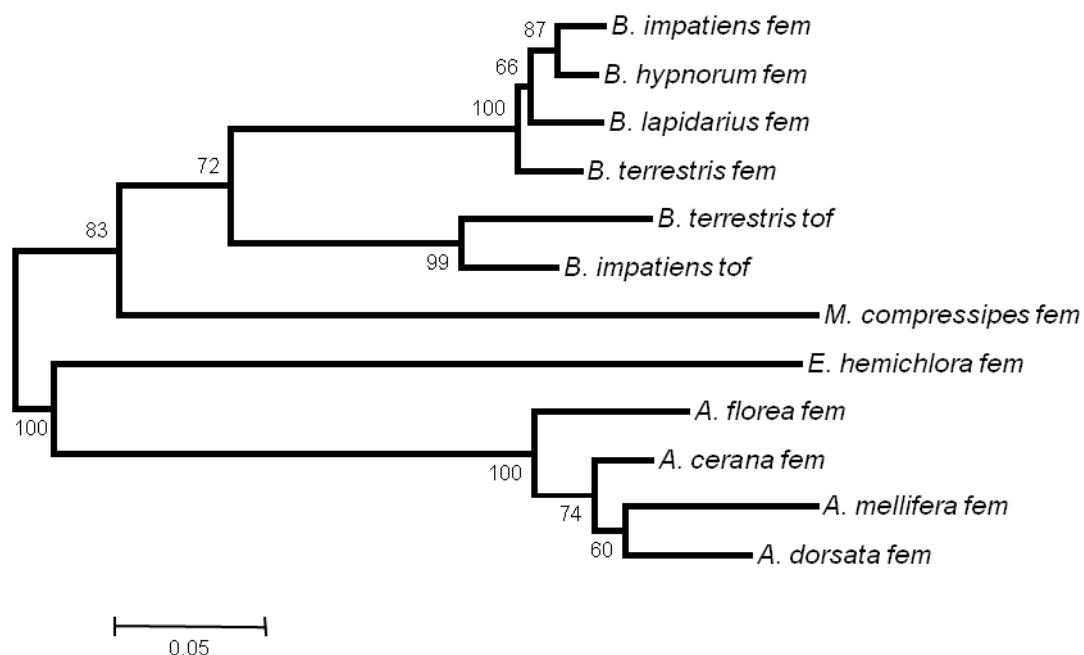


Abbildung 32: Genealogie der *tof*-Nukleotidsequenzen aus *B. terrestris* und *B. impatiens* sowie *fem*-Sequenzen aus den Apidae-Gattungen *Bombus*, *Apis*, *Melipona* und *Euglossa*. Die Genealogie wurde auf Basis der synonymen Substitutionen nach der Neighbor-Joining-Methode und dem Jukes-Cantor-Modell erstellt. Die Astlänge gibt die synonymen Substitutionen pro Position an.

Es zeigt sich, dass sich die *tof*-Sequenzen an einem separaten Ast anordnen. Die Bootstrapwerte an der Verzweigung zwischen den *Bombus*-*fem* und *tof*-Sequenzen betragen nur 72, was vermutlich darauf zurück zu führen ist, dass die Genealogie auf Basis der synonymen Substitution (d_s) erstellt wurde. Eine Genealogie, die mit den gesamten Nukleotidsequenzen erstellt wurde und somit auf einer größeren Anzahl an Nukleotidpositionen basiert, ergibt eine gleiche Anordnung und bestätigt die Separierung von *tof* und den *Bombus*-*fem*-Sequenzen (Bootstrap = 99 %). Dies ist ein klarer Hinweis darauf, dass die Duplikation vor der Aufspaltung der *Bombus*-Arten stattfand und

spricht dafür, dass *tof* in allen Hummelarten existiert.

Unterstützt wird die Anordnung in der Genealogie auch durch die Analyse der synonymen Divergenzwerte (d_s), die zwischen *tof* und den *fem*-Sequenzen berechnet wurden (Tab. 14). Zwischen den beiden *tof*-Sequenzen ist d_s im einseitigen Z-Test signifikant niedriger als zwischen *tof* und den *Bombus-fem*-Sequenzen ($P < 0,05$ für *Bt-tof*, $P \leq 0,05$ für *Bi-tof*). Die engere Verwandtschaft der *tof*-Gene zueinander als zu *Bombus-fem* bestätigt die Separierung der *tof*-Sequenzen in der Genealogie und unterstützt die Hypothese, dass der Zeitpunkt der Genduplikation vor der Aufspaltung der *Bombus*-Arten liegt.

Tabelle 14: Genetische Divergenz berechnet für die synonymen Substitutionen (d_s) zwischen *B. terrestris-tof* bzw. *B. impatiens-tof* und den verschiedenen *Bombus-fem*-Sequenzen. Berechnet mit MEGA nach Nei-Gojobori (2.2.8). Die Standardfehler sind in Klammern angegeben.

	<i>B. terrestris-tof</i>	<i>B. impatiens-tof</i>
<i>B. terrestris-tof</i>		0,084 ($\pm 0,020$)
<i>B. terrestris-fem</i>	0,195 ($\pm 0,028$)	0,168 ($\pm 0,027$)
<i>B. impatiens-fem</i>	0,198 ($\pm 0,028$)	0,176 ($\pm 0,027$)
<i>B. hypnorum-fem</i>	0,208 ($\pm 0,029$)	0,174 ($\pm 0,028$)
<i>B. lapidarius-fem</i>	0,208 ($\pm 0,030$)	0,183 ($\pm 0,029$)

Aus der Hypothese, dass *tof* vor der Aufspaltung der *Apis*-Arten entstand, ergibt sich als weitere Hypothese, dass es auch in anderen Apidae existieren könnte. In *Apis* fanden sich weder im Rahmen dieser noch früherer Arbeiten (z.B. Beye et al. 2003, Hasselmann et al. 2008a) Hinweise auf eine weitere *fem*-Duplikation. Bei den in der NCBI-Datenbank verfügbaren Transkriptomdaten von *Melipona quadrifasciata* und *Euglossa cordata* findet sich kein Hinweis auf die Existenz des Gens *tof* oder einer anderen *fem*-Duplikation.

Einen weiteren Hinweis über die Exklusivität von *tof* in der *Bombus*-Linie kann durch die Berechnung des Alters der *tof*-Sequenzen gewonnen werden. Zwar ist die Mutationsrate für *Bombus* nicht bekannt, doch lässt die Verwendung der Mutationsrate von *Apis* ($2,47 \times 10^{-9}$ syn. Sub./Position/Jahr, Abschnitt 3.1.4) eine Annäherung bei der Bestimmung des Entstehungszeitpunkts von *tof* zu. Unter Verwendung der synonymen Substitutionsrate als neutrales Maß errechnet sich bei einem d_s -Wert von 0,01 zwischen den *tof*-Sequenzen ein Zeitraum vor etwa 40 Mio. Jahren für die Genduplikation. Ramírez et al. (2010) datierten die Trennung der *Bombus*-Linie von den anderen Apidae auf einen Zeitpunkt vor ungefähr 80 Mio. Jahren. Diese große Differenz unterstützt die Hypothese, dass *tof* vor der Aufspaltung der *Bombus*-Arten entstand, jedoch nach der Trennung von *Bombus* von den anderen Apidae-Arten.

Um die evolutionäre Geschichte des Gens *tof* genauer zu untersuchen, d.h. welche Form der Selektion auf das Gen gewirkt hat, wurden die nichtsynonymen und synonymen Substitutionen pro entsprechender Position entlang der einzelnen Äste ($b_n/b_s \times 1000$) der Genealogie aus Abbildung 32 berechnet (Abb. 33). Außerdem wurde die absolute Zahl an entsprechenden Substitutionen berechnet (a_n/a_s).

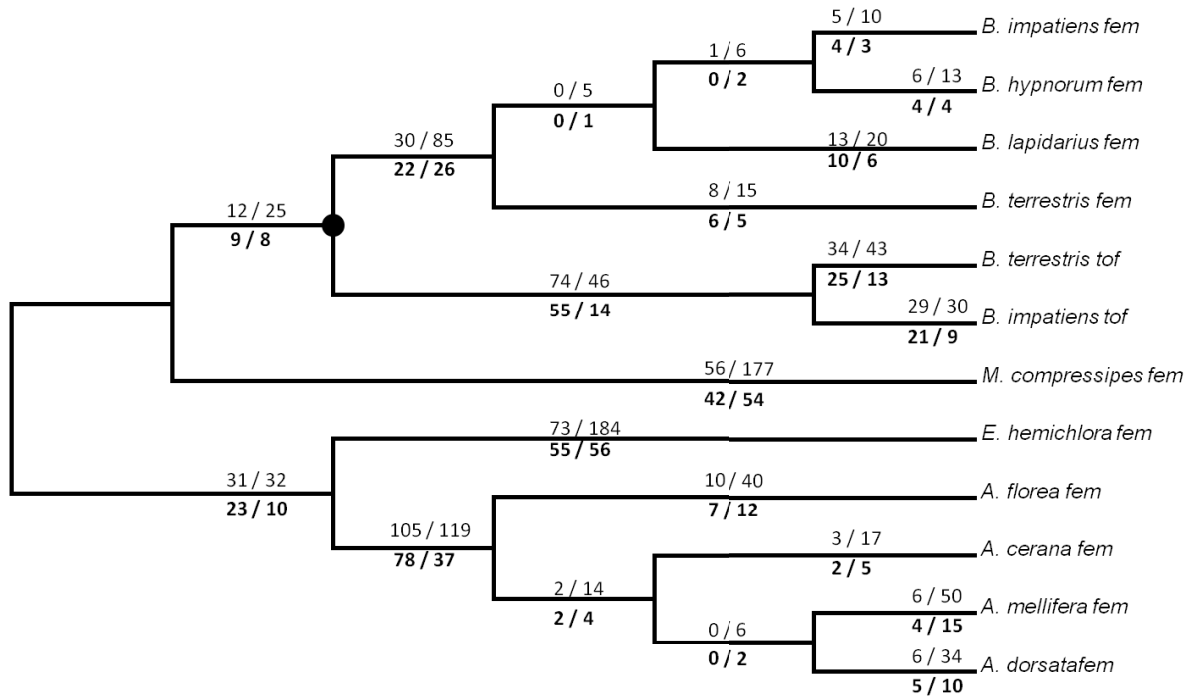


Abbildung 33: Genealogie auf Basis der Nukleotidsubstitutionen, abgeleitet von Abbildung 32. Für jeden Ast wurde die Rate der nichtsynonymen und synonymen Substitutionen pro Position an den einzelnen Ästen (b_n/b_s , Werte $\times 1000$) berechnet (Programm Bn-Bs, 2.2.8), die Werte sind oberhalb des jeweiligen Astes angegeben. Unterhalb des jeweiligen Astes ist die absolute Zahl an Substitutionen (a_n/a_s , fett gedruckt) angegeben. Der Punkt markiert das Ereignis der *fem*-Duplikation und Entstehung von *tof*.

Die *Bombus-fem*-Sequenzen zeigen deutliche Spuren von negativer Selektion in Form einer hochsignifikant reduzierten Anzahl an nichtsynonymen Substitutionen im Vergleich zu synonymen Substitutionen ($P < 0,0001$, Fisher's Exact-Test). Bei *tof* ist die Zahl an nichtsynonymen Substitutionen im Vergleich zu *fem* der gleichen Art um das dreifache erhöht ($P < 0,0001$). In *B. terrestris* findet sich für *tof* eine Zahl von 80 nichtsynonymen zu 27 synonymen Substitutionen entlang der Äste (ab der Duplikation, $55/14 + 25/13$), während sich bei *fem* ein Verhältnis von 28 zu 31 findet. Ähnlich sieht dies in *B. impatiens* mit 76 zu 23 bei *tof* und 26 zu 32 bei *fem* aus. Dies ist ein Hinweis auf positive Selektion und eine schnelle Evolution des Gens, was wiederum auf eine Neofunktionalisierung der *fem*-Duplikation *tof* hin deutet. Auffällig ist außerdem, dass sich auch nach der

Artenaufspaltung eine hohe Zahl an nichtsynonymen im Vergleich zu synonymen Substitutionen findet (25 zu 13 bei *B. terrestris*, 21 zu 9 bei *B. impatiens*). Dies zeigt, dass *tof* weiterhin schnell evolviert. Eine mögliche Erklärung hierfür wäre, dass *tof* unter balancierender Selektion steht, welche zu einer Akkumulation von Polymorphismen führen würde und so die hohe Zahl an nichtsynonymen Substitution zwischen den beiden Arten erklären könnte. Dies würde bedeuten, dass es sich bei *tof* tatsächlich um den gesuchten komplementären Geschlechtsdeterminationsfaktor handeln könnte.

3.4.3 Das Gen *tof* ist ein potentieller allelischer Faktor für die komplementäre Geschlechtsdetermination

Um die Hypothese zu untersuchen, dass es sich bei *tof* um den gesuchten multiallelischen und initialen Faktor der komplementären Geschlechtsdetermination in der Erdhummel handelt, wurde die Nukleotiddiversität von *tof* untersucht. Mit den Oligonukleotiden S-65 und S-67 wurde aus der genomischen DNA von zwölf *B. terrestris*-Drohnen ein ca. 420 bp langes Fragment des *tof*-Gens amplifiziert und sequenziert (Abb. 10 und Tab. 1). Im Vergleich der Sequenzen fällt auf, dass sie auf Aminosäureebene keinen Polymorphismus aufweisen. Dies ist ein Indiz, dass es sich bei *tof* nicht um einen allelischen Faktor handelt. Zur genaueren Analyse der Variation in *tof* wurde für die generierten Sequenzen die Nukleotiddiversität pro Position (π , 2.2.8) berechnet. Zum Vergleich wurde die Berechnung auch für den korrespondierenden Bereich der genomischen *Bt-fem*-Sequenzen durchgeführt.

Für die nichtsynonymen Substitutionen pro Position (π_n) ergab sich für *tof* ein Wert von 0 und für *fem* von 0,003. Der Unterschied zwischen den beiden Werten ist nicht signifikant ($P > 0,05$, zweiseitiger Z-Test). Auch für die synonymen Substitutionen zeigt sich bei einem π_s -Wert von 0,001 bei *tof* und 0,008 bei *fem* kein signifikanter Unterschied ($P > 0,05$). Diese Werte unterstützen die Hypothese, dass *tof* nicht der gesuchte, multiallelische, komplementäre Geschlechtsdeterminationsfaktor ist. Bestätigt werden kann dies zusätzlich durch den Vergleich der hier analysierten Sequenzen mit der aus einem Eierpool anderer Herkunft amplifizierten cDNA-Sequenz, der genomischen Sequenz aus der NCBI-Datenbank und der Ausgangssequenz aus der EST-Datenbank. Zu keiner dieser Sequenzen fanden sich Unterschiede, die auf eine allelische Diversität in *tof* hindeuten.

4 Diskussion

4.1 Das multiallelische Gen *complementary sex determiner* in *Apis mellifera*

Das Gen *complementary sex determiner* ist das initiale Gen der Geschlechtsdeterminationskaskade in der Honigbiene *A. mellifera* und aktiviert im heterozygoten Zustand den weiblichen Entwicklungsweg. Homozygotie am *csd*-Lokus ist jedoch letal, was zu einem starken Heterozygotenvorteil führt. Das *csd*-Gen ist hoch variabel und in den Honigbienenpopulationen segregieren eine große Anzahl verschiedener Allele.

Im Rahmen dieser Arbeit wurden *csd*-Sequenzen von *A. mellifera*-Arbeiterinnen und -Drohnen untersucht, die in verschiedenen Regionen der Welt gesammelt wurden (3.1.1). Hierbei wurde der Fokus auf die Analyse der Sequenzvariation in der potentiellen Spezifizierungsdomäne (PSD) und der in ihr enthaltenen hypervariablen Region (HVR) gelegt.

4.1.1 Das unter balancierender Selektion stehende Gen *csd* zeigt kaum genetische Differenzierung zwischen verschiedenen Subpopulationen

Balancierende Selektion erhält die *csd*-Allele über einen langen Zeitraum, da seltene Allele einen selektiven Vorteil haben und dadurch weniger häufig durch genetische Drift verloren gehen als bei neutralen Loci (Takahata 1990). Von Schierup et al. (2000b) wurde zudem postuliert, dass seltene Allele eines Locus unter balancierender Selektion schneller migrieren, da sie in Subpopulationen, in denen sie noch nicht vorhanden sind, einen selektiven Vorteil auf Grund ihrer neuen Spezifität haben. Diese beiden Aspekte führen zu der Annahme, dass sich zwischen verschiedenen natürlichen *A. mellifera*-Subpopulationen keine genetische Differenzierung am *csd*-Lokus findet.

Um dies zu untersuchen, wurden *csd*-Sequenzen von *A. mellifera*-Arbeiterinnen aus Kenia untersucht. Diese Proben wurden in drei Berg- sowie drei Savannenregionen gesammelt und stammen von Völkern, die kaum menschlichem Einfluss ausgesetzt sind. Die Bergregionen werden von *A. mellifera monticola* besiedelt und die Savannenregionen von *A. m. scutellata* (Meixner et al. 2000). Bei neutralen Loci (Mikrosatelliten) zeigt sich eine geringe, jedoch signifikante genetische Differenzierung (K_{ST}) zwischen den Subpopulatio-

nen (K. Gruber Gonzales, unveröffentl. Daten). Es wird angenommen, dass dies durch die geografischen Gegebenheiten begründet ist, die zu einer Reduktion des Genflusses zwischen den verschiedenen Subpopulationen führen. Am *csd*-Lokus findet sich dagegen entsprechend der Erwartung zwischen fünf der sechs untersuchten Subpopulationen keine genetische Differenzierung (Tab. 4). Lediglich die Sequenzen aus Mau Forest Savannah zeigen im Vergleich zu den anderen Subpopulationen eine signifikante genetische Differenzierung ($P < 0,05$). Dies könnte jedoch in der Größe der Stichprobe begründet liegen (14 Sequenzen, Tab. 3) und ein Artefakt darstellen. Die kaum vorhandene genetische Differenzierung steht in Übereinstimmung mit Analysen am S-Lokus. Glémin et al. (2005) zeigten am S-Lokus von *Brassica insularis* eine um die Hälfte niedrigere genetische Struktur im Vergleich zu neutralen Loci, Stoeckel et al. (2008) fanden bei *Prunus avium* eine um das dreifache reduzierte genetische Differenzierung am S-Lokus. Verantwortlich dafür ist die balancierende Selektion, die die genetische Struktur zwischen Populationen reduziert (Charlesworth et al. 1997, Schierup et al. 2000b, Muirhead 2001), da Individuen mit einem seltenen Allel einen selektiven Vorteil haben und es so zu einer höheren „effektiven Migrationsrate“ kommt (Barton und Bengtsson 1986, Schierup et al. 2000a). Muirhead (2001) postulierte für den S-Lokus, dass sich ein Locus unter balancierender Selektion ab einer bestimmten Migrationsrate wie in einer panmiktischen Population verhält. Dieses Level ist ab einer Migrationsrate (m) von 10^{-3} migrierenden Gameten in der gesamten Population pro Generation erreicht und scheint im Fall der *csd*-Allele zuzutreffen. Ist die Migrationsrate näher an der Mutationsrate, wie dies für neutrale Loci der Fall ist, ist eine genetische Differenzierung zu erwarten. Das notwendige Migrationslevel wird vermutlich durch die Paarung an Drohnensammelplätzen erreicht. Dort treffen sich Drohnen und Arbeiterinnen aus bis zu 240 Völkern aus einem Umkreis von bis zu 5 km (Baudry et al. 1998). Dies ermöglicht vermutlich den Austausch der Allele zwischen verschiedenen Subpopulationen.

Ein weiterer Faktor, der eine genetische Differenzierung am *csd*-Lokus verhindert, ist vermutlich die Koaleszenzzeit der Allele, die im Rahmen dieser Arbeit auf 13 - 16,2 Mio. Jahre berechnet wurde (3.1.4). Das Alter der Allele führt dazu, dass sich die Allele in verschiedenen Subpopulationen über lange Zeit erhalten und in Folge keine genetische Differenzierung zu detektieren ist. Die Genealogie in Abbildung 12 zeigt, dass sich die gleichen Allele in verschiedenen Subpopulationen finden, während vermutlich auf Grund des hohen Alters der Allele und der Migration keine regionalen Gruppen zu erkennen sind.

4.1.2 Die hypervariable Region des *csd*-Gens evolviert schnell und durch Sequenzduplikation

Die hypervariable Region zeigt das höchste Maß an Variation innerhalb des *csd*-Gens. Im hier analysierten Datensatz aus 193 Sequenzen fand sich bei 172 Sequenzen ein Unterschied von mindestens einer Aminosäure in der HVR. Analysen der *csd*-Sequenzen in dieser Arbeit sowie in vorangegangenen Publikationen (z.B. Hasselmann et al. 2008b, Hasselmann und Beye 2006) wurden ohne diesen Sequenzbereich durchgeführt. Auf Grund ihrer Struktur aus einer variierenden Zahl von Motiv-Wiederholungen unterschiedlicher Länge, ist dieser Bereich nicht auf zuverlässige Weise alinierbar, d.h. Verwandtschaften zwischen den einzelnen Positionen können dadurch nicht erkannt werden. Im Rahmen dieser Arbeit wurde eine Kodierung der HVR in Betracht gezogen, um die hier enthaltene Information ebenfalls in die Erstellung einer Genealogie mit einzubeziehen. Hierfür die Länge der HVR oder die Zahl der Motiv-Wiederholungen zu verwenden, musste verworfen werden. Wie in Abschnitt 3.1.2.1 dargestellt, korreliert die Länge der HVR nicht mit dem Alter der Allele. Unterstützt wird diese Erkenntnis dadurch, dass die HVR offenbar durch Duplikationen und Insertionen von Kodons verlängert, aber ebenso durch Deletionen verkürzt werden kann, wie dies auch für Mikrosatelliten beschrieben wird, die eine vergleichbare, repetitive Struktur aufweisen (Ellegren 2004). Ein Indiz hierfür ist auch in der Tatsache zu sehen, dass vermehrt HVR-Sequenzen mittlerer Länge zu finden sind (Tab. 6). Bedingt durch dieses Mutationsmuster aus Duplikationen und Deletionen würde beispielsweise das in Abbildung 16 dargestellte Paar H3-1 und H4-1 nicht als eng verwandt erkannt werden, obwohl sich in den Exons 6-9 zwischen beiden Sequenzen keine Nukleotidpolymorphismen finden. Dies ist ein Hinweis darauf, dass die Mutationsrate in der HVR höher ist und sie dadurch deutlich schneller evolviert als die benachbarten Regionen.

Ein weiteres Problem für eine Kodierung stellt die Identifizierung der repetitiven Motive innerhalb der HVR dar. Sowohl in dieser Arbeit (3.1.2) als auch bei Hasselmann et al. (2008b) wurde die Struktur der HVR als Wiederholungen eines $N_{1-5}Y$ -Motivs beschreiben. Wie Mikrosatelliten scheint auch die HVR gleichzeitig durch Duplikation und Deletion von Sequenzmotiven wie auch durch Basensubstitutionen zu mutieren (Kruglyak et al. 1998). Die in der HVR vorkommenden Kodons für Asparagin (AAT) und Tyrosin (TAT) unterscheiden sich lediglich an einer Basenposition. So kann innerhalb eines Replikationszykluses das Motiv NNNNNY zu NNYNNY mutieren, oder aber durch eine Duplikation zu NNNNNYNNNNNY. NNYNNY dagegen kann ebenso aus einer Duplikation eines NNY-Motivs entstanden sein wie durch eine Mutation aus NNNNNY (Abb. 34).

Dies bedeutet, dass anhand der HVR-Sequenzen nicht zu erkennen ist, ob es sich bei NNYNNY um ein oder zwei Motiv-Wiederholungen handelt. Gleichmaßen kann eine Verwandtschaft zwischen NNNNNYNNNNNY und NNYNNY nicht erkannt werden. Des Weiteren sind auch sämtliche Mutationen in die entgegengesetzte Richtung möglich, eine Y zu N-Mutation oder die Deletion einer Motivwiederholung. Aus diesen Überlegungen

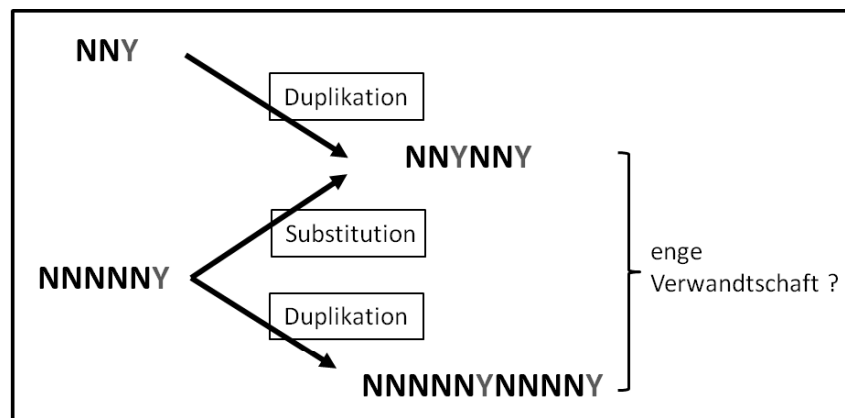


Abbildung 34: Darstellung der möglichen Modifikationen und somit der Verwandtschaft zwischen HVR-Sequenzmotiven. (N = Asparagin, Y = Tyrosin).

muss der Schluss gezogen werden, dass die HVR wegen ihrer schnellen Evolution, der Mutation durch Insertion und Deletion von Sequenzmotiven sowie der Tatsache, dass Insertion und Substitution teilweise nicht unterschieden werden können, keine Information enthält, wie diese für die Erstellung einer Genealogie verwendet werden kann.

Bei dieser Analyse zeigt sich zudem, dass das *csd*-Gen zwei unabhängig von einander evolvierende Bereiche hat, die beide zur Allelspezifität beitragen. Dies erhöht die Entstehungsrate neuer Allele, da Mutationen in beiden Bereichen möglicherweise unabhängig zu neuen Allelspezifitäten führen können.

4.1.3 Die Zahl der *csd*-Allele in den *A. mellifera*-Populationen ist um das vier- bis sechsfache höher und die Koaleszenzzeit niedriger als bislang postuliert

Im Rahmen der vorliegenden Arbeit wurde die Anzahl der in der gesamten Honigbienenpopulation vorkommenden *csd*-Allele erneut bestimmt. Die Bestimmung der Allelzahl erfolgte mittels Kriterien, die anhand der Unterschiede zwischen 78 Paaren funktionell verschiedener Allele entwickelt wurden. Diese Kriterien wurden dann auf den gesamten Datensatz von 193 Sequenzen angewandt und so 50-90 verschiedene Allele identifiziert.

Dieser Wert liegt deutlich über den bislang in der Literatur postulierten Allelzahlen. Laidlaw et al. (1956) berechneten anhand von Bruder-Schwesterkreuzungen eine Zahl von 9-11 Allelen. Adams et al. (1977) bestimmten die Zahl der in Völkern mit natürlich gepaarter Königin auftretenden diploiden Drohnenlarven und errechneten daraus eine Zahl von 19 Allelen. Hasselmann et al. (2008b) unterschieden bei *csd*-Sequenzen aus Stichproben, welche an zwei Lokalitäten gesammelt wurden, 14 verschiedene Allele.

Es ist davon auszugehen, dass in der HVR und der benachbarten, potentiellen Spezifizierungsdomäne (PSD) nicht nur eine bestimmte Anzahl an Aminosäureunterschieden notwendig ist, sondern dass diese Unterschiede funktionellen Kriterien genügen müssen. Da der Mechanismus der Allelspezifität bisher nicht bekannt ist, existieren bislang nur Hypothesen zu den für die Allelspezifität notwendigen Kriterien. Eine Hypothese hierzu ist, dass *csd*-Allele Heteromere bilden, indem sie über die arginin-serinreiche und die prolinreiche Domäne Proteinbindungen eingehen (M. Otte, pers. Mitteilung). Bei identischen Allelen könnte es dabei zu einer Abstoßung zwischen den hypervariablen Regionen kommen, wodurch die Bindung verhindert würde. Die Vermutung ist, dass es die negative Partialladung der Tyrosine ist, welche zur Abstoßungsreaktion führt. Durch die Längenunterschiede kann es zu einer Verschiebung der hypervariablen Regionen gegeneinander kommen, die dazu führt, dass die Tyrosine nicht mehr einander gegenüber positioniert sind. So könnten sich die HV-Regionen tolerieren und die Proteinbindung wäre möglich. Es zeigt sich, dass dieses Prinzip auf alle 78 Allelpaare anwendbar ist.

Es werden weitere Analysen sowie funktionelle Tests notwendig sein, um den Mechanismus genauer zu untersuchen und zu testen, welche Art von Unterschieden zwischen zwei verschiedenen Allelen relevant sind, um die Allelzahl genauer bestimmen zu können.

Auch der Selbstinkompatibilitätsloкус (S-Lokus) der Angiospermen und der Haupthistokompatibilitätsloкус (MHC-Lokus) der Vertebraten stehen unter balancierender Selektion und weisen eine Vielzahl von Allelen auf (Vekemans und Slatkin 1994, Aguilar et al. 2004). Während es sich jedoch bei *csd* um ein einzelnes Gen handelt, besteht der S-Lokus aus zwei gekoppelten und koevolvierenden Genen, die die gleiche Spezifität aufweisen. Es handelt sich dabei um die S-Rezeptorkinase (SRK), die die Spezifität im Pistill determiniert, und das S-cysteinreiche Protein (SCR), welches die Spezifität im Pollen bestimmt (Schopfer et al. 1999, Takasaki et al. 2000). Exprimieren Pollen und Pistill die gleiche Spezifität so wird die Befruchtung verhindert. Die Zahl der am S-Lokus vorkommenden Allele schwankt sehr zwischen den einzelnen Arten und es wurden Allelzahlen von 13 (*Solanum carolinense*, Richman et al. 1995), 40-70 (z.B. *Ipomoea trifida*, Kowyama et al. 1994; *Papaver rhoeas*, Lane und Lawrence 1993) bis hin zu 140-190 in *Trifolium pratense* (Williams und Williams 1947, Lane und Lawrence 1993) berechnet. Hierbei ist zu beachten, dass die Bestimmung der Allelzahl in fast

allen Fällen durch Kreuzungsexperimente erfolgte, bei denen die Zahl der auf Grund identischer Allelspezifitäten inkompatiblen Individuen erhoben wurde. Lediglich in der Studie von Richman et al. (1995) bildeten wie in der vorliegenden Arbeit Sequenzdaten die Grundlage für die Berechnung der Allelzahl. Die Bestimmungsmethode könnte Einfluss auf die kalkulierten Allelzahlen haben, was bei deren Vergleich zu beachten ist. Die große Variation bei den Allelzahlen am S-Lokus der verschiedenen Arten (Lawrence 2000) erschwert zudem den Vergleich zwischen dem *csd*-Gen und dem S-Lokus.

Ein Vergleich der Allelzahlen am MHC-Lokus und des *csd*-Gens ist nicht möglich. Grund hierfür ist zum einen, dass der MHC-Lokus unter einem Selektionsregime steht, das sich deutlich zu dem unterscheidet, unter welchem das *csd*-Gen steht. Am MHC-Lokus ist im Gegensatz zu *csd* und dem S-Lokus Homozygotie möglich. Bei *csd* ist Homozygotie indirekt letal, da diploide Männchen im frühen Larvenstadium aufgefressen werden (Woyke 1963). Beim S-Lokus verhindert eine identische Allelspezifität bei Pollen und Pistill die Befruchtung (Dzelzkalns et al. 1992). Am MHC-Lokus führt Homozygotie zwar zu reduzierter Fitness, ist jedoch nicht letal. Ein weiterer Aspekt, der den Vergleich nicht möglich macht, ist, dass beim MHC-Lokus lediglich die Variation in der proteinbindenden Furche des Proteins relevant ist, jedoch keine funktionell verschiedenen Allele existieren (Potts und Wakeland 1993).

Die Zahl der in einer Population segregierenden Allele eines Gens unter balancierender Selektion ist von der effektiven Populationsgröße und der Entstehungsrate neuer Allele abhängig. Der Nachweis einer deutlich größeren Zahl an *csd*-Allelen in der vorliegenden Arbeit im Vergleich zur Literatur deutet darauf hin, dass auch diese beiden Faktoren größer sind, als bislang angenommen. Die Neuberechnung der Entstehungsrate zeigt, dass diese mit 2×10^{-6} bis 2×10^{-7} pro Gen und Jahr um etwa eine Zehnerpotenz höher ist als die in Hasselmann et al. (2008b) berechnete Rate von 10^{-7} - 10^{-8} . Grund hierfür ist die Neukalkulation der molekularen Uhr auf Basis neuer Sequenzdaten und der Neubestimmten Divergenzzeitpunkte zwischen den *Apis*-Arten (Ramírez et al. 2010) sowie die Etablierung von Kriterien, die eine Bestimmung der Mindestanzahl an nichtsynonymen Substitutionen in funktionell verschiedenen Allelen ermöglichte. Die Berechnungen der Parameter, die im Koaleszenzprozess eine Rolle spielen zeigen zudem, dass die effektive historische Populationsgröße mit 12.000 - 42.000 ebenfalls um etwa eine Zehnerpotenz größer ist als die in Hasselmann et al. (2008b) berechnete Populationsgröße von 1500 - 4000. Wie in dieser Studie sowie Tabelle 7 zu sehen ist, haben schon Änderungen einzelner Parameter (z.B. die Entstehungsrate der Allele um eine Zehnerpotenz) einen ausgeprägten Einfluss auf die anderen Parameter. Auch wenn die hier aufgeführten Parameter lediglich als Annäherung zu betrachten sind, wird deutlich, dass von einer substantiell erhöhten Anzahl von Allelen und historischen

Populationsgröße auszugehen ist.

Das hohe Alter von Allelen eines Gens unter balancierender Selektion führt zu der Annahme, dass so genannte transspezifische Allele existieren. Diese Allele sind älter als der Divergenzzeitpunkt der Arten und in Folge sind sie näher zu einem Allel der anderen Art verwandt als zu den meisten Allelen der eigenen Art. Sowohl für den S-Lokus als auch den MHC-Lokus wurde die Existenz transspezifischer Allele beschrieben (Castric et al. 2008, Takahata 1993). In der Studie von Hasselmann et al. (2008b) konnten zwischen *A. mellifera*, *A. cerana* und *A. dorsata* jedoch keine transspezifischen *csd*-Allele detektiert werden. Hier ergab sich ein Widerspruch zwischen dem in der Studie berechneten Höchstalter der Allele, welches für *A. mellifera* bei 14,6 Mio. Jahren lag und den Divergenzzeitpunkt zwischen den Arten (6 - 8 Mio Jahre), auf denen diese Berechnung beruhte. In der Studie wurde postuliert, dass dies möglicherweise durch die Zahl der analysierten Allele bedingt war und potentielle transspezifische Allele nicht in den Stichproben beider Arten repräsentiert waren.

Bei den in der vorliegenden Arbeit analysierten Sequenzen finden sich ebenfalls keine Hinweise auf transspezifische Allele. Auch nach den aktuellen Berechnungen steht dies im Widerspruch zum Höchstalter der Allele von 48 - 61 Jahren und den Divergenzzeitpunkten der drei *Apis*-Arten vor 18,2 bzw. 23,5 Mio. Jahre. Zwar ist die Wahrscheinlichkeit sehr gering, dass potentielle transspezifische Allele in dem hier analysierten Datensatz aus 193 Sequenzen nicht repräsentiert sind, jedoch wurden für die Analyse die gleichen *A. cerana*- und *A. dorsata-csd*-Sequenzen verwendet wie in der Studie von Hasselmann et al. (2008b). Somit ist es möglich, dass transspezifische Allele zwar bei den *A. mellifera*-Sequenzen jedoch nicht bei den Sequenzen der anderen beiden Arten vertreten sind. Eine alternative Erklärung wäre, dass die Allele, welche älter sind als die Artenaufspaltung, jeweils nur in einer der Arten erhalten wurden. Eine weitere Erklärung findet sich in der Studie von Charlesworth (2006), in der postuliert wird, dass die Akkumulation neuer Polymorphismen, welche die ursprünglichen Muster überdecken, dazu führen können, dass keine transspezifischen Allele mehr detektiert werden können.

4.2 Das Gen *complementary sex determiner* der asiatischen Zwerghonigbiene *A. florea* zeigt geringe allelische Diversität

Im Rahmen dieser Arbeit konnte das Gen *csd* in der asiatischen Zwerghonigbiene *Apis florea* identifiziert und sequenziert werden. Die Identifizierung des Gens in dieser innerhalb der Gattung *Apis* basal stehenden Art bestätigt die von Hasselmann et al. (2008a) postulierte Entstehung von *csd* vor der Aufspaltung der *Apis*-Arten.

Im Gegensatz zu den bisher bekannten *csd*-Genen der westlichen Honigbiene *A. mellifera* sowie der asiatischen Honigbienen *A. cerana* und *A. dorsata* kodiert *Af-csd* für eine fast ausschließlich aus Asparagin bestehende hypervariable Region (Abb. 20). Das für die *csd*-Gene der anderen *Apis*-Arten vorhergesagte Coiled-Coil-Motiv (Hasselmann et al. 2008a) ist in *Af-csd* nicht nachweisbar (Abb. 24). Stattdessen wird ein N-terminales Coiled-Coil-Motiv vorhergesagt, dass sich in den anderen *csd*-Genen nicht findet. Die Sequenzvariation zwischen den Allelen bei *Af-csd* ist sehr gering (Tab. 9 und 11), was darauf hindeutet, dass die *Af-csd*-Allele noch sehr jung sind. Dies weist auf Unterschiede in der evolutionären Geschichte von *Af-csd* und den *csd*-Genen der anderen *Apis*-Arten hin.

Die Analysen in der vorliegenden Arbeit (3.1.2) sowie in der Studie von Hasselmann et al. (2008b) deuten darauf hin, dass bei *Am-csd* ein großes Maß an Sequenzvariation in der HVR und auch einige Unterschiede in der restlichen PSD notwendig sind, um allelische Diversität zu vermitteln. Gleiches gilt vermutlich für *csd* in *A. cerana* und *A. dorsata*. Bei *Af-csd* scheint dies nicht der Fall zu sein. Die Sequenzanalysen funktionell verschiedener Allelpaaire weisen darauf hin, dass hier kleine Längenunterschiede von zum Teil nur einer einzelnen Aminosäure und möglicherweise die Position der einzelnen Serine in der HVR notwendig und zugleich hinreichend für den funktionellen Unterschied zwischen zwei Allelen sind (Abb. 21). Bei *Af-csd* findet sich ein Paar funktionell verschiedener Allele, das in den zur HVR benachbarten Bereichen, welche bei *Am-csd* als PSD definiert wurden, keinen Sequenzunterschied aufweist (Tab. 8). Dies ist ein Indiz dafür, dass bei *Af-csd* möglicherweise ausschließlich die HVR für die Allelspezifität relevant ist. Eine alternative Erklärung wäre jedoch, dass bei *Af-csd* Sequenzunterschiede in anderen Genbereichen zur Allelspezifität beitragen, worüber die Sequenzierung und Analyse der entsprechenden Genabschnitte aus verschiedenen Individuen Aufschluss geben könnte. In Anbetracht der geringen Sequenzvariation in der HVR und dem für *Af-csd* an einer anderen Position vorhergesagten Coiled-Coil-Motiv ergibt sich die Frage, ob sich möglicherweise Unterschiede im Mechanismus der Proteininteraktion im Vergleich zu *csd* der anderen *Apis*-Arten finden.

Zur Bestätigung der Hypothese, dass eine geringe Sequenzvariation in *Af-csd* für die allelische Diversität ausreichend ist, können Sequenzierungen der beiden *csd*-Allele weiterer *A. florea*-Arbeiterinnen beitragen, anhand derer die Nukleotidsequenzvariation zwischen funktionell verschiedenen Allelen genauer untersucht werden kann.

Die Unterschiede zwischen den Merkmalen des *csd*-Gens in *A. florea* und in anderen *Apis*-Arten lässt die Annahme zu, dass in der *Apis*-Gattung zwei verschiedene Typen des Gens existieren. Eine sich hieraus ergebende Hypothese ist, dass sich nach der Genduplikation und während der Neofunktionalisierung zwei Varianten des *csd*-Gens evolviert haben. Während sich die eine Variante der Linie von *A. mellifera*, *A. dorsata* und *A. cerana* sehr variabel ist, hat sich die andere, weniger variable Variante in *A. florea* durchgesetzt.

Die niedrige Sequenzvariation in *Af-csd* und die kurzen, inneren Äste in der Genealogie (Abb. 22), welche für eine kurze Koaleszenzzeit der *csd*-Allele sprechen, deuten auf einen genetischen Flaschenhals oder eine hohe Durchsatzrate („turnover rate“) als Ursache hin. Für einen genetischen Flaschenhals spricht das Ergebnis des Tests nach dem Modell von Takahata et al. (1992). Das Verhältnis zwischen der höchsten Zahl an paarweisen synonymen und nichtsynonymen Unterschieden zwischen den *csd*-Sequenzen und dem entsprechenden Durchschnittswert liegt für *Af-csd* mit 5,0 und 3,1 deutlich über der Erwartung von 1,8 (Tab. 12). Für Allele unter balancierender Selektion wird erwartet, dass sie in einem ähnlichen Maß miteinander verwandt sind wie neutrale Allele. Dies bestätigt sich für die *csd*-Allele von *A. mellifera*, *A. cerana* und *A. dorsata* (Tab. 12, Hasselmann et al. 2008b). Es wurden auch für *Af-csd* Hinweise auf balancierende Selektion gefunden, da das d_n/d_s -Verhältnis bei > 1 liegt und sich signifikante Unterschiede in der Diversität innerhalb des Gens (Tab. 11) finden. Somit könnten die von der Erwartung abweichenden Werte im Takahata-Test durch einen genetischen Flaschenhals bedingt sein.

Ein alternatives Erklärungsmodell für eine geringe Koaleszenzzeit ist eine hohe Durchsatzrate („turnover rate“). Diese wird hauptsächlich durch die effektive Populationsgröße und die Entstehungsrate neuer Allele bestimmt. Die effektive Populationsgröße ist für *A. florea* nicht bekannt. Ist sie jedoch gering, könnte hierin die Ursache für die kurze Koaleszenzzeit liegen, da in diesem Fall die Allele stärker dem Prozess der genetischen Drift ausgesetzt sind. Verhindert eine kleine effektive Populationsgröße die Akkumulation einer größeren Anzahl von Polymorphismen in *Af-csd*, könnte dies eine Erklärung sein, weshalb sich in *A. florea* eine Variante des *csd*-Gens evolviert hat, für dessen allelische Spezifität wenige Aminosäureunterschiede hinreichend sind.

Die Ursache für die hohe Entstehungsrate neuer Allele könnte neben der geringen effektiven Populationsgröße auch im Mechanismus der Allelespezifität liegen, welcher mit einer geringen Sequenzvariation auszukommen scheint. Sind bei *Af-csd* einzelne

Aminosäureunterschiede für eine neue Allelspezifität hinreichend, würde dies die Entstehungsrate neuer Allele im Vergleich zu *Am-csd* erhöhen. In Folge würden Allele auch häufiger durch genetische Drift wieder verloren gehen und damit zu einer hohen Durchsatzrate führen. Dies würde wiederum in einer niedrigen Koaleszenzzeit resultieren. Die Sequenzierung des Gens in weiteren Individuen aus verschiedenen Regionen und die Ermittlung der Zahl der in der *A. florea*-Population segregierenden *csd*-Allele wird dazu beitragen, die Frage nach der Ursache für das junge Alter der *Af-csd*-Allele zu beantworten.

4.3 Das Gen *feminizer* in *Apis* und anderen Hymenopteren

4.3.1 Das Gen *fem* ist in der Gruppe der corbiculaten Apidae konserviert

Im Rahmen dieser Arbeit konnte die Existenz des Gens *feminizer* in den drei Hummelarten *Bombus impatiens*, *B. hypnorum* und *B. lapidarius*, in der Prachtbiene *Euglossa hemichlora* und einer Nicht-Apidae-Art, der parasitischen Wespe *Nasonia vitripennis*, nachgewiesen werden (3.3.1). Bislange war *fem* neben *Apis mellifera* in den Honigbienenarten *A. cerana*, *A. dorsata* und *A. florea*, sowie in der stachellosen Biene *Melipona compressipes* und der Erdhummel *B. terrestris* bekannt (Hasselman et al. 2008a).

Das Gen *fem* ist der zweite Faktor in der Geschlechtsdeterminationskaskade der Honigbiene *Apis mellifera* und im Gegensatz zu seinem Paralog *csd* zeigt es keine Sequenzvariation. Wie bei *transformer* (*tra*) der Dipteren finden sich auch bei *fem* geschlechtsspezifische Spleißformen (Hasselman et al. 2008a) und in seiner weibchenspezifischen, aktiven Form initiiert Fem ebenfalls das alternative Spleißen von *doublesex* (*dsx*, Gempe et al. 2009). Auf Grund dieser funktionellen Übereinstimmungen sowie Homologien in konservierten Domänen postulierten Hasselman et al. (2008a), dass es sich bei *fem* und *tra* um orthologe Gene handelt. Dies bedeutet, dass die funktionell relevanten Domänen in Fem seit über 300 Mio. Jahren konserviert wurden.

Die Ergebnisse dieser Arbeit gehen mit den in der Literatur beschriebenen Eigenschaften von *fem* konform. Sequenzanalysen zeigen, dass das Gen *fem* innerhalb der Apidae konserviert ist.

Für die Erdhummel *B. terrestris* konnte in der Arbeit von Schulte (2008) ebenfalls eine männchenspezifische Spleißform sequenziert und die Existenz von *dsx* nachgewiesen werden. Dies spricht dafür, dass *Bt-Fem* die gleiche Funktion hat wie *Am-Fem*.

Zudem konnten in *N. vitripennis*, die kein Vertreter der Apidae ist, ebenfalls zwei geschlechtsspezifische Spleißformen nachgewiesen werden (3.3.1.3). Verhulst et al. (2010a) wiesen für *Nv-Fem* (in der Publikation als *Nv-Tra* bezeichnet) die von Pane et al. (2005) für *Tra* (*Fem*) postulierte, konservierte Funktion der Aktivierung des geschlechtsspezifischen Spleißens der *dsx*-prä-mRNA nach. Diese Ergebnisse sowie die Sequenzhomologie zwischen den verschiedenen Apidae-*fem* unterstützen die Hypothese, dass Fem in allen Hymenopteren die gleiche Funktion innehat.

4.3.2 Die Sequenzevolution des Gens *fem* wird in der *Apis*-Linie durch sein Paralog *csd* beeinflusst

Ein Teil der im Rahmen dieser Arbeit generierten Daten wurde schon in der Publikation von Hasselmann et al. (2010) analysiert und publiziert. Hierbei konnte gezeigt werden, dass die Evolution von *fem* in *Apis* durch die Nähe zu *csd* im Genom, sowie durch die strukturelle Ähnlichkeit zu seinem Paralog, mit welchem es interagiert, beeinflusst wird. Verhulst et al. (2010b) postulierten, dass Fem/Tra eine zentrale Position in der Evolution der Geschlechtsdetermination der Insekten hat. Die im vorangegangenen Abschnitt (4.3.1) beschriebenen Hinweise sprechen dafür, dass *fem* innerhalb der Hymenopteren verbreitet ist und unterstützten die Hypothese, dass die Aktivierung des weiblichen Entwicklungswegs über das geschlechtsspezifische Spleißen von *fem* konserviert ist (Gempe und Beye 2010). Der Mechanismus, der dieses alternative Spleißen initiiert, ist jedoch nicht konserviert (Heimpel und de Boer 2008) und es scheint eine Vielzahl an Faktoren zu existieren, die als initiales Signal dienen. In *Apis* wird das Geschlecht wie in vielen Hymenopteren über komplementäre Geschlechtsdetermination bestimmt (Cook 1993) und es handelt sich bei dem multiallelischen, initialen Faktor der Geschlechtsdetermination um eine Duplikation des *fem*-Gens, welche eine neue Funktion entwickelt hat. Dies bedeutet für *Apis*-Fem eine funktionelle Interaktion mit einem strukturell sehr ähnlichen Protein (Abb. 3). Dadurch steht *fem* in *Apis* unter stärkeren funktionellen Zwängen als in anderen Gattungen. Entwickelt Fem eine zu große strukturelle Ähnlichkeit zu Csd, könnte es möglicherweise mit diesem interagieren und so zur Aktivierung des weiblichen Entwicklungswegs in einem haploiden Embryo führen. Diese funktionellen Zwänge zeigen sich durch eine reduzierte Rate von nichtsynonymen zu synonymen Substitutionen in *Apis-fem* im Vergleich zu *fem* anderer Apidae, was auf eine stärkere negative Selektion hinweist (Hasselmann et al. 2010, 3.3.2.1).

Balancierende Selektion führt durch den Hitchhiking-Effekt zu einer Akkumulation von Polymorphismen in den zum Zielgen benachbarten Regionen (Hudson und Kaplan 1988, Strobeck 1983). Wie in 3.3.2.2 beschrieben, konnte dieser Effekt von Hasselmann et al. (2010) auch bei *fem* und *csd* nachgewiesen werden.

Csd und Fem sind ein gutes Beispiel für die von Fraser et al. (2002, 2003) postulierte und von Makino et al. (2006) anhand von duplizierten Genen untersuchte, verlangsamte Evolution eines Proteins mit mehr Protein-Protein-Interaktionspartnern im Vergleich zu einem Protein mit weniger Interaktionspartnern. Nach der Genduplikation in *Apis* hat das eine Protein (Fem) seine Funktion beibehalten. Dies erlaubte dem anderen Protein (Csd) schneller und unabhängiger zu evolvieren, da es keinen funktionellen Zwängen

mehr unterliegt. So konnte sich im Fem-Paralog die HVR und das Coiled-Coil-Motiv in der RS-Domäne entwickeln (Abb. 3 und 24). Bei *fem* zeigt sich eine deutlich langsamere Evolution als bei seinem Paralog *csd*.

In der Erdhummel *Bombus terrestris* existiert ebenfalls ein Fem-Paralog (Tof). Falls Tof mit *Bt*-Fem interagiert, bietet sich hier neben *fem* und *csd* in der Honigbiene ein weiteres Paar duplizierter und möglicherweise interagierender Gene, anhand derer der Einfluss eines neu entstandenen Gens auf sein Paralog untersucht werden kann.

4.4 Das Gen *twin of feminizer* ist eine *fem*-Duplikation in der Hummel

Klassische Kreuzungsexperimente von Duchateau et al. (1994) zeigten, dass es sich beim initialen Signal der komplementären Geschlechtsdetermination in der Erdhummel um einen multiallelischen Faktor handelt. Diese Studie sowie eine von Gadau et al. (2001) durchgeführte genetische Kopplungskartierung weisen darauf hin, dass es sich dabei um einen einzelnen geschlechtsbestimmenden Locus handelt. Die Identität dieses Locus ist bisher nicht bekannt. Hasselmann et al. (2008a) postulieren, dass es sich bei dem initialen Signal in der Hummel nicht um das Gen *csd* handelt, da es Hinweise dafür gibt, dass es erst nach der Trennung von *Bombus*, *Apis* und *Melipona* in der *Apis*-Linie entstanden ist. In Vorarbeiten wurden im Rahmen der Sequenzierung des Gens *feminizer* (*fem*) in *B. terrestris* Hinweise auf die Existenz einer *fem*-ähnlichen Sequenz gefunden. Dies führte zu der Hypothese, dass es sich dabei um den gesuchten CSD-Faktor handeln könnte.

Im Rahmen dieser Arbeit konnte diese *fem*-Duplikation in *B. terrestris* identifiziert werden. Es handelt sich um das Gen *twin of feminizer* (*tof*). Es zeigt nicht nur eine hohe Sequenzhomologie zu *fem*, sondern es werden für *tof* auch die gleichen funktionellen Domänen vorhergesagt wie beim Fem-Protein. Wie bei *fem* treten bei *tof* mindestens zwei verschiedene Spleißformen auf. Eine Spleißform kodiert für einen langen Leserahmen, während es in der anderen durch ein Stoppkodon in einem alternativ gespleißten Exon zu einem frühen Translationsabbruch kommt (Abb. 30). Das alternative Exon befindet sich an einer zum männchenspezifischen Exon in *fem* homologen Position. Ob es in *tof* jedoch ebenfalls geschlechtsspezifisch gespleißt wird, ist bislang nicht bekannt. Bei der Analyse von Nukleotidsequenzen des *tof*-Gens aus verschiedenen Drohnen mehrerer Populationen zeigte sich, dass *tof* kaum Nukleotidsequenzvariation aufweist und es fanden sich Hinweise, dass es sehr schnell evolviert. Die Bestimmung des Entstehungszeitpunkts und die Stellung von *tof* und *Bombus-fem* in einer Genealogie weisen darauf hin, dass *tof* vor der Aufspaltung von *B. terrestris* und *B. impatiens* entstanden ist. Gleichzeitig finden sich keine Hinweise darauf, dass *tof* auch in anderen Gattungen der Apidae existiert.

Untersuchungen der synonymen und nichtsynonymen Substitutionen entlang der Äste einer Genealogie der *tof*- und *Bombus-fem*-Sequenzen (Abb. 33) zeigen für *tof* ein im Vergleich zu *Bombus-fem* um das zwei- bis dreifache erhöhte Verhältnis an nichtsynonymen zu synonymen Substitutionen. Dies kann entweder eine Folge von positiver, balancierender oder sehr schwacher negativer Selektion sein und ist ein Hinweis auf eine schnelle Evolution des Gens. Auffällig ist, dass die Zahl der nichtsynonymen Substitutionen nicht nur nach der Duplikation des Gens erhöht ist, sondern auch noch nach der Trennung von *B. terrestris* und *B. impatiens*. Nach der Genduplikation ist die

hohe Zahl an nichtsynonymen Substitutionen vermutlich durch positive Selektion im Rahmen einer Neofunktionalisierung bedingt. Nach der Aufspaltung der Arten müssen alle drei Selektionsregimes als mögliche Szenarien in Betracht gezogen werden.

Eine erste Hypothese ist, dass *tof*, wie auch die *fem*-Duplikation *csd* in *Apis*, unter balancierender Selektion steht. Diese führt zu einer Akkumulation von Polymorphismen, was sich auch in Form einer erhöhten Zahl an nichtsynonymen Substitutionen zwischen verschiedenen Arten zeigen würde. Steht *tof* unter balancierender Selektion, könnte es sich bei dem Gen um den gesuchten multiallelischen, komplementären Geschlechtsdeterminationsfaktor handeln. Ein multiallelisches, unter balancierender Selektion stehendes Gen zeigt jedoch Sequenzvariation, welche bei *tof* bislang nicht detektiert werden konnte. Die untersuchten, genomischen DNA-Sequenzen stammen von Drohnen aus mehreren Völkern. Sofern *tof* ein CSD-Faktor ist, der im heterozygoten Zustand aktiv ist, müssten bei den Drohnen eines Volkes zwei verschiedene Allele auftreten, bei welchen es sich um die beiden von der Königin stammenden, funktionell verschiedenen Allele handelt. Die *tof*-Sequenz konnte in allen zwölf untersuchten Drohnen amplifiziert werden, es konnten jedoch keine Nukleotidsequenzunterschiede gefunden werden. Zudem konnte im Vergleich der genomischen Sequenzen mit der sequenzierten mRNA-Sequenz keine Variation detektiert werden. Auch zwischen der in der EST-Datenbank gefundenen Sequenz und der genomischen Sequenz aus der NCBI-Datenbank konnten keine Nukleotidunterschiede festgestellt werden. Die Herkunft der verschiedenen Sequenzen ist unbekannt. Es ist unwahrscheinlich, dass in allen Fällen - sofern unterschiedliche Allele existieren - immer dasselbe Allel gefunden wurde. Dies spricht klar gegen *tof* als allelischen Faktor und gegen balancierende Selektion als Ursache für die hohe Zahl an nichtsynonymen Substitutionen. Eine alternative Hypothese ist eine sehr schwache negative Selektion. Zwar unterstützt die Berechnung von Tajima's *D* mit einem Wert von -0,028, welcher auf neutrale Evolution hindeutet, dieses Erklärungsmodell. Der Wert ist jedoch nicht signifikant ($P > 0,1$). Zudem würde es bei einem schwachen selektiven Druck ebenfalls zur Akkumulation von Polymorphismen im Gen kommen. Da keine Sequenzvariation bei *tof* auftritt, muss auch diese Hypothese abgelehnt werden.

Die dritte Hypothese ist, dass die vielen Substitutionen zwischen den *tof*-Genen der beiden Arten durch positive Selektion bedingt sind. Dies könnte bedeuten, dass das Gen in den beiden Arten unterschiedlich evolviert, wie dies beispielsweise auch für das *Odysseus*-Gen in *Drosophila* beschrieben wird (Ting et al. 2000). *Odysseus* hat eine Funktion bei der Spermatogenese und seine divergente Evolution in den verschiedenen Arten führt zu Hybridsterilität (Sun et al. 2004). Positive Selektion kann sowohl die hohe Zahl an nichtsynonymen Substitutionen in *tof* nach der Artenspaltung von *B. terrestris* und *B. impatiens* erklären, als auch die geringe Nukleotidsequenzvariation

innerhalb des Gens, und kann deshalb als wahrscheinlichste Erklärung angesehen werden. Unterstützung für diese Hypothese könnte möglicherweise durch die Sequenzierung des *tof*-Gens in weiteren Hummelarten gewonnen werden.

Die Funktion von Tof ist bislang unbekannt. Hinweise darauf könnte das alternativ gespleißte Exon liefern (Abb. 30A). Wird das Exon wie in *fem* geschlechtsspezifisch gespleißt, wäre auch für *tof* eine Funktion in der geschlechtsbestimmenden Kaskade anzunehmen. Um zu untersuchen, ob geschlechtsspezifische Spleißformen existieren, müssen die Transkripte zum Zeitpunkt der Geschlechtsdifferenzierung untersucht werden. Hierfür ist es notwendig, die mRNA im Embryonal- oder frühen Larvenstadium zu amplifizieren. Es ist jedoch nicht möglich, bei der Erdhummel haploide von diploiden Eier anhand des Legeorts zu unterscheiden, wie dies bei der Honigbiene möglich ist. Eine Bestimmung der Ploidie der Eier ist durch die Verwendung von Mikrosatelliten (Stolle et al. 2009) oder durch Durchflusszytometrie (Nagamitsu und Yamagishi 2009) möglich, konnte jedoch in dieser Arbeit nicht realisiert werden. Im Rahmen einer weiteren Doktorarbeit in der Arbeitsgruppe ist die Etablierung von Mikrosatelliten und die Verwendung von Durchflusszytometrie zur Bestimmung der Haplo- oder Diploidie von Hummeleiern geplant. Neben der Bestimmung der Ploidie ist es auch notwendig, zwischen am geschlechtsbestimmenden Locus (SDL) hetero- und homozygoten Individuen zu unterscheiden. So kann verhindert werden, dass die Untersuchung diploider männlicher Eier oder Larven anstelle von weiblichen zu einer Verfälschung der Daten führt. Hierfür ist die Kenntnis des SDL in *B. terrestris* notwendig, dessen Kartierung unter Verwendung einer etablierten Inzucht Kreuzung und der Annotierung des *B. terrestris*-Genoms bereits geplant ist. Hierbei können zudem Informationen über die Identität des komplementären Geschlechtsdeterminationsfaktors gewonnen werden.

Einen weiteren Indiz darauf, dass Tof eine Funktion in der Geschlechtsdeterminationskaskade haben könnte ergibt sich aus der Ähnlichkeit zu seinem Paralog Fem. Wie Fem weist auch Tof die für ein SR-Typ-Protein charakteristische RS-Domäne auf. Im Gegensatz zu SR-Proteinen weisen SR-Typ-Proteine keine RNA-Bindemotive auf (Blencowe et al. 1999), wie es auch bei Fem und Tof der Fall ist. SR-Typ-Proteine sind an der prä-mRNA-Prozessierung beteiligt, sie vermitteln bei der Spleißstellenerkennung sowie der Spleißregulierung.

Die Funktion von Fem, bzw. dessen Ortholog Tra, bei der Spleißregulation von *dsx* ist in den Hymenopteren und Dipteren konserviert (Pane et al. 2005, Gempe und Beye 2010). Die Vermutung liegt daher nahe, dass dieser Teil der Geschlechtsdeterminationskaskade auch in der Erdhummel konserviert ist und *Bt*-Fem ebenfalls das geschlechtsspezifische Spleißen von *dsx* vermittelt. In Anlehnung an die von Wilkins (1995) und Nöthiger & Steinmann-Zwicky (1985) beschriebene „Bottom-Up“-Hypothese könnte Tof, wenn es

ebenfalls Teil dieser Kaskade ist, eine Position über Fem einnehmen. Dort könnte es z.B. für das alternative Spleißen der *Bt-fem*-prä-mRNA zuständig sein.

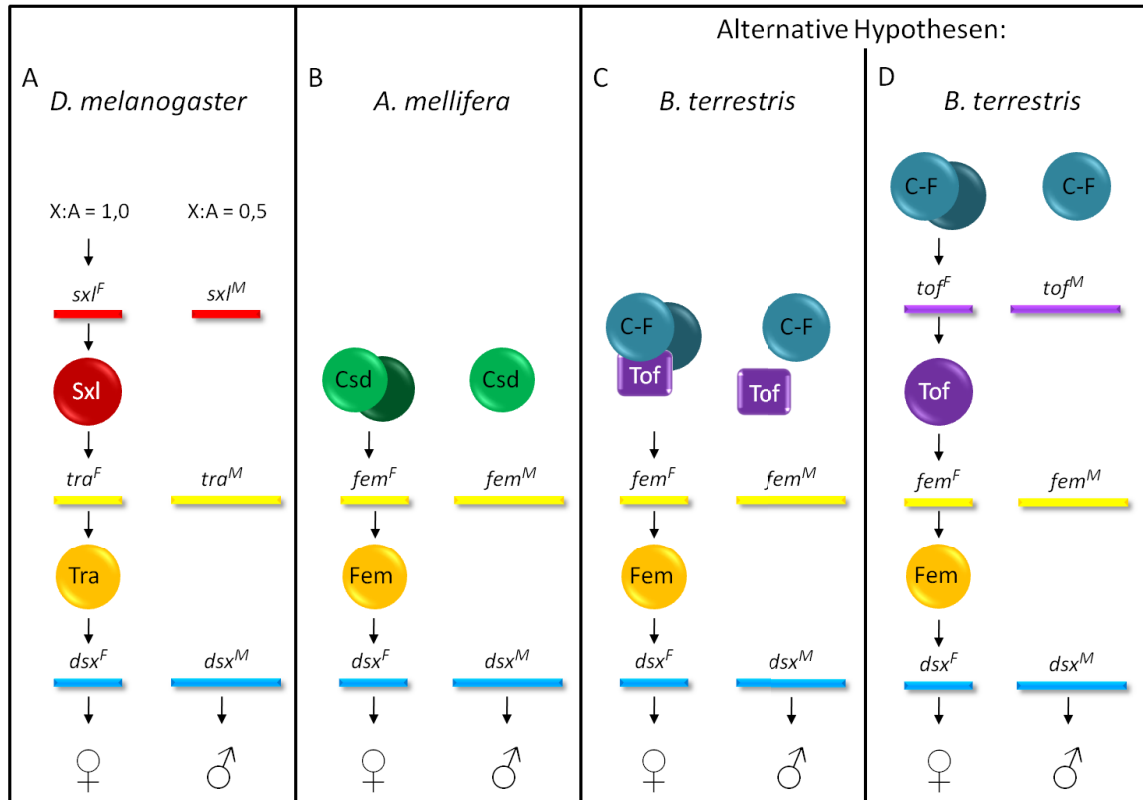


Abbildung 35: Schematische Darstellung der Geschlechtsdeterminationskaskaden in *D. melanogaster* und *A. mellifera* sowie zwei alternativer, hypothetischer Kaskaden für *B. terrestris*. Sowohl in *D. melanogaster* als auch in *A. mellifera* ist der männliche Entwicklungsweg voreingestellt. Auf Grund der engen Verwandtschaft zu *A. mellifera* wird dieses System auch bei *B. terrestris* angenommen. (A) In *D. melanogaster* wird das Geschlecht durch das Verhältnis von Geschlechtschromosomen zu Autosomen determiniert (Cline und Meyer 1996, Erickson und Quintero 2007). Bei einem Verhältnis von 1:1 wird *sexlethal* aktiviert. Es initiiert das geschlechtsspezifische Spleißen von *tra*, welches dadurch ebenfalls als aktives Protein vorliegt und so das alternative Spleißen von *dsx* vermittelt, wodurch der weibliche Entwicklungsweg eingeschlagen wird. (B) In *A. mellifera* erfolgt die Geschlechtsdetermination in Weibchen durch zwei Csd-Proteine unterschiedlicher Spezifität, welche das geschlechtsspezifische Spleißen von *fem* vermitteln und es dadurch aktivieren. Wie sein Ortholog Tra aktiviert auch Fem *dsx*, wodurch auch in *A. mellifera* der weibliche Entwicklungsweg aktiviert wird (Gempe et al. 2009). (C) und (D) Hypothetische Kaskaden in *B. terrestris*. Es wird davon ausgegangen, dass die Kaskade ab dem Gen *fem* konserviert ist und in *B. terrestris* nach dem gleichen System funktioniert wie in *A. mellifera*. (C) Zwei CSD-Faktoren (C-F) unterschiedlicher Spezifität interagieren mit ToF und können so die Aktivierung von *fem* vermitteln. (D) Zwei CSD-Faktoren (C-F) unterschiedlicher Spezifität initiieren das geschlechtsspezifische Spleißen von *tof* wodurch ein aktives ToF-Protein entsteht, welches *fem* aktivieren kann.

In Abbildung 35 sind zwei mögliche Szenarien im Vergleich zu den geschlechtsdeterminierenden Kaskaden von *D. melanogaster* und *A. mellifera* dargestellt. Dabei wird davon ausgegangen, dass wie bei diesen auch bei *B. terrestris* die männliche Entwicklung voreingestellt ist, während die weibliche Entwicklung durch Heterozygotie am geschlechtsbestimmenden Locus aktiviert wird (Erickson und Quintero 2007, Gempe et al. 2009). Auf Grund der fehlenden allelischen Diversität kann *Tof* nicht als komplementärer Geschlechtsdeterminator funktionieren, es könnte jedoch in Kombination mit einem solchen wirken (Abb. 35C). Eine zweite Hypothese gründet auf der Annahme, dass das alternative Exon geschlechtsspezifisch gespleißt wird. In diesem Fall wäre es möglich, dass *Tof* eine ähnliche Funktion wie *Fem* hat (Abb. 35D). So könnte es sich in der Kaskade zwischen einem allelischen Faktor, von dem es aktiviert wird, und *fem* befinden, dessen geschlechtsspezifisches Spleißen es vermittelt.

Die Analyse weiterer *tof*-Sequenzen kann zu genaueren Erkenntnissen über eine potentielle allelische Diversität und damit über eine mögliche Funktion führen. RNAi-Experimente könnten den Nachweis erbringen, ob *Tof* am geschlechtsspezifischen Spleißen von *fem* beteiligt ist. Unter der Voraussetzung, dass auch in *Bombus* der männliche Entwicklungsweg voreingestellt ist, würde die Injektion von *tof*-dsRNA in diesem Fall dazu führen, dass *fem* nur noch in seiner männchenspezifischen Spleißform vorliegt und sich der Embryo in Folge dessen in ein Männchen entwickelt.

5 Zusammenfassung

Das Gen *complementary sex determiner* (*csd*) und sein Paralog *feminizer* (*fem*) stehen am Beginn der Geschlechtsdeterminationskaskade der Honigbiene *Apis mellifera*. Das multiallelische *csd* ist das initiale Signal und aktiviert im heterozygoten Zustand den weiblichen Entwicklungsweg durch geschlechtsspezifisches Spleißen von *fem*. Um Heterozygotie in Weibchen zu realisieren, segregieren in der gesamten Honigbienenpopulation etwa 50-90 *csd*-Allele. Sie unterscheiden sich durch starke Sequenzvariation sowie ihre hypervariable Region (HVR), einem in seiner Länge und Aminosäurekomposition stark variierenden Bereich, dessen repetitive Sequenz unabhängig von der restlichen Gensequenz evolviert.

Das Gen *csd* steht unter balancierender Selektion. Hierdurch bedingt, finden sich keine Unterschiede in der Diversität der *csd*-Allele verschiedener Subpopulationen sowie auch kaum genetische Differenzierung zwischen diesen.

Die Existenz von *csd* konnte nun auch in der basal stehenden *Apis*-Art *A. florea* nachgewiesen werden. *Af-csd* zeigt jedoch in der für das *csd*-Gen charakteristischen HVR sowie in seiner Sequenzvariation deutliche Unterschiede zu den bisher bekannten *csd*-Genen.

Das Gen *csd* ist in der HonigbienenGattung aus einer Duplikation des Gens *fem* entstanden. Im Gegensatz zu *csd* ist *fem* konserviert und vermutlich in allen Hymenopteren-Gattungen vorhanden: Es konnte in einigen Vertretern der Apidae-Gattungen *Apis*, *Bombus*, *Melipona* und *Euglossa* sowie der parasitischen Wespe *Nasonia vitripennis* (Pteromalidae) nachgewiesen werden. In *Apis* ist *fem* durch die Existenz von *csd* anderen selektiven Bedingungen ausgesetzt als in Nicht-*Apis*-Arten, in denen *csd* fehlt. Zum einen wird *fem* durch die direkte Nachbarschaft zu dem unter balancierender Selektion stehenden Gen *csd* beeinflusst. Zum anderen bedeutet die strukturelle Ähnlichkeit zu *csd* gleichzeitig einen verstärkten funktionellen Zwang, der die Evolution von *fem* in *Apis* stärker einschränkt als in Nicht-*Apis*-Arten.

Bei der Erdhummel *Bombus terrestris* initiiert ein noch unbekannter komplementärer Geschlechtsdeterminationsfaktor die Geschlechtsdifferenzierung. Es konnte hier ebenfalls eine *fem*-Duplikation nachgewiesen werden, welche als *twin of feminizer* (*tof*) benannt wurde. Das Gen weist jedoch kaum Sequenzvariation auf und somit ist unwahrscheinlich, dass es einen komplementären Geschlechtsdeterminationsfaktor darstellt.

5.1 Summary in English

The gene *complementary sex determiner* (*csd*) and its paralogous gene *feminizer* (*fem*) are the first two genes in the sex determination cascade of the honey bee *Apis mellifera*. The multiallelic *csd* is the initial signal and activates the female development path in the heterozygous state through sex-specific splicing of *fem*. To realize heterozygosity in females, about 50-90 *csd*-alleles segregate in the entire honey bee population. They are marked by a high degree of sequence variation and differences in their hypervariable region (HVR). The HVR is a region that varies considerably in length and amino acid composition. Its repetitive sequence evolves independently from neighboring regions.

The *csd* gene is under balancing selection. As a result, there are no differences in sequence diversity between *csd* alleles of separate subpopulations and little genetic differentiation is found between them.

The existence of *csd* could now also be proven in the most basal *Apis* species, *A. florea*. However, *Af-csd* shows clear differences in the structure of the HVR, which is characteristic for *csd* genes, and in its sequence variation as compared to other *csd* genes known to date.

The *csd* gene originated from a duplication of the *fem* gene in the honey bee genus. In contrast to *csd*, *fem* is conserved and probably exists in all Hymenoptera genera: it could be detected in some representatives of the Apidae-genera *Apis*, *Bombus*, *Melipona* and *Euglossa* as well as in the parasitic wasp *Nasonia vitripennis* (Pteromalidae). Due to the presence of *csd* in *Apis* *fem* is subject to different selective constraints than in non-*Apis*-species in which this gene is missing. On the one hand, *fem* is influenced by its direct proximity to a gene under balancing selection, the *csd* gene. On the other hand, its structural similarity to *csd* increases the functional constraints on *fem*, which limit its evolution in *Apis* more strongly than in non-*Apis* species.

In the buff-tailed bumblebee *Bombus terrestris* sex determination is initiated by an still unknown complementary sex determination factor. In this species, the existence of a *fem* duplicate could also be shown; this duplicate was named *twin of feminizer* (*tof*). However, this gene shows only little sequence variation and is therefore unlikely to be a complementary sex determination factor.

6 Anhang

6.1 HVR-Sequenzen funktionell verschiedener *csd*-Allele

Tabelle 15: Die Sequenzen der hypervariablen Region von funktionell verschiedenen Allelpaaren. Zudem sind die d_s - und d_n -Werte für den sequenzierten Bereich der potentiellen Spezifizierungsdomäne (ohne HVR) angegeben.

Allelpaar	HVR-Sequenz	d_s	d_n
A7	SLSNNYNNNNNYKYNYYNN	0,120	0,116
A10	SLSNNYNSNNNNNNNNNNNNNN		
I5	SLSNNTIHNNNNYKYNNNNNNN	0,074	0,081
I6	SLSNKTIHNNNNYKYNNNNNNNNN		
D1	SLSNNTIHNNNNYNNNN	0,067	0,104
D6	SLSNNYKYSNNNNNNNNNNNNNN		
L2	NNYNYKYNNNNNNN	0,136	0,104
L5	SLSDKTIHNNNNYKYNNNNNNNYS		
K4	SLSNHYNNNNNKYNNNNDY	0,108	0,084
K5	NNNSLSNNYNNNNYNNNNKHN		
W1	SLSNNTIHNNNYKYNHNNNNY	0,094	0,061
W6	SLSNKTIHNNNNNNNNNNNNNNNS		
S1-2	SLSNKTIHNNNNYKYNNNNNNNNN	0,144	0,081
S1-3	SLSNKTIHNNNNYNNNNNNNNNNNN		
1-2	SLSNNYNSNNNN	0,127	0,038
1-3	SLSNNYNNNNNNNNNNNNNNNN		
2-2	SLSNNTIHNNNNYNNNN	0,166	0,061
2-5	SLSNKTIHNNNNNNNNNN		
3-4	SLSNNYNNNNNNYKYNNNNN	0,110	0,070
3-5	SLSNKTIHNNNNYKYN		
4-4	SLSNNTIHNNNNYNNNN	0,106	0,031
4-5	SLSNKTIHNNNNYKYNNNNNNNNNNNYS		
5-4	SLLNNTIHNNNN	0,084	0,051
5-5	SLSNKTIHNNNNNNNNNNNNNNNN		
6-1	SLSNNYNNNNNNYKKLYCNY	0,086	0,064
6-2	SLSNNYKYSNNNNNNNNNNNNNNNN		
9-1	SLSNKTIHNNNNNNNNNNNNNNNN	0,095	0,068
9-2	SLSNKTIHNNNNNNNNNN		

10-1	SLSNHYNNNNKYNNYNDY	0,155	0,127
10-2	SLSNNYKYSNNYNNNNNNNNYS		
31-4	SLSNNYNNNNKYNNNNYN	0,085	0,051
31-5	SLSSSCNYSNNYNYNNN		
34-2	SLSNNYKYSNNYNNNNNNYDNNYNNY	0,078	0,079
34-3	SLSNKTIHNNNNYKYNNNNNNNNNY		
36-1	SLSNKTIHNNNNYNNNNYNNNNYKYNNNY	0,092	0,069
36-2	SLSNNYNNNNNNNNNNNNNNNNYN		
74-9	SLSKNTIHNNNNYKYNNNNNNNNNY	0,124	0,114
74-10	SLSNNYNSNSYNNNNNNYNN		
H1-1	SLSNKTIHNNNNNNNNNN	0,137	0,103
H1-2	SLSNHYNNNNKYNNYNDY		
H2-4	SLSNKTIHNNNNY	0,040	0,032
H2-5	SLSNKTIHNNNNYKYNNNNNNNNNYS		
H3-1	SLSNKTIHNNNNYKYNNNY	0,147	0,115
H3-2	SLSNNYNYSNNNNNNNNNY		
H4-1	SLSNKTIHNNNNYKYNNNYKYNNNYKYNNNY	0,160	0,133
H4-2	SLSNNYKYNNNYNNNNNNNNYN		
H5-3	SLSNNYKYSNNYNNNNYKLYS	0,180	0,134
H5-4	SLSNHYNNNNKYNNYNDY		
H6-1	SLSNKTIHNNNNYNNNNNNNNY	0,103	0,106
H6-2	SLSNNYNNNCNYKH		
H7-2	SLSNNYNNNNNNNNNNY	0,112	0,083
H7-3	SLSNKTIHNNNNYNNNNNNNNY		
B 64-3	SLSNKTIHNNNNYNNNNYNNNN	0,119	0,054
B 64-4	SLSNKTIHNNNNYKYNNNNNNNNNNNYS		
B 65-1	SLSNNYNNNNY	0,042	0,057
B 65-2	SLSNNYNYKYNNNS		
B 67-3	SLSNNTIHNNNNYNNNN	0,136	0,069
B 67-4	SLSNNYNNNNNNNNNNNNNNNNNNYN		
B 68-4	SLSNKTIHNNNNYNNNNYNNNNY	0,125	0,100
B 68-5	SLSNNYKYSNNYNNNNNNKYNNNNYN		
B 71-2	SLSNNYNNNNNNYKYNNNYN	0,123	0,137
B 71-3	SLSNKTIHNNNNYKYNNNNNNNNNYS		
B 72-4	SLSNKTIHNNNNYNNNNNNNNY	0,144	0,080
B 72-5	SLSNNYKYSNNYNNNNNNKYNNNNYN		
B 153-4	SLSNKTIHNNNNYKYNNNNNNNNNY	0,031	0,061
B 153-5	SLSNKTIHNNNNYNNNNNNNC		

N01-I-1	SLSNKT I HNNNNYNNYNNYNN	0,149	0,121
N01-I-2	NNNSLSNNYNNNNYNNYNNYNNKHNY		
N01-II-1	SLSNNT I HNNNYNNKNNY	0,000	0,025
N01-II-2	SLSNNT I HNNNYKYNNNNYNNY		
N04-I-1	SLSNKT I HNNNNYNN	0,062	0,038
N04-I-2	SLSKNT I HNNNNYKYNNYNNNYS		
N04-II-3	SLSNNYKYSNNYNNNNYNNNNY	0,242	0,132
N04-II-8	SLSNKT I HNNNNYNN		
N06-I-2	SLSNNYKYSNNYNNNNYNNNTNY	0,189	0,095
N06-I-9	NNNSLSNNYNNNNYNNYNNYNNKHNY		
N06-II-3	SLSNNYSNNYNNYNNSSNNNYF	0,195	0,108
N06-II-6	NNNSLSNNYNNNNYNNYNNYNNKHNY		
N14-I-7	SLSNNT I HNNNYKYNNNNYNNY	0,108	0,038
N14-I-8	SLSNKT I HNNNNYKYNNNNHYNNSN		
N14-II-1	SLSNNYSNNYNNNNYNNNNYNN	0,090	0,056
N14-II-3	SLSNKT I HNNNNYKYNNNNYNNNNYNNNY		
N16-II-4	SLSNNT I HNNNNYKYNNNNYNNYNNYNN	0,124	0,093
N16-II-6	SLSNNYKYSNNYNNNNYNNNNYNNYNN		
N20-I-1	SLSNKT I HNNNNYNNNYKNYNNNNYNNYNTNY	0,135	0,063
N20-I-5	SLSNNYNNNNNNYNNYKYNNYNNY		
N20-II-1	SLSNNYNNNNNNYNNYKYNNYNNY	0,158	0,063
N20-II-9	SLSNNT I HNNNYKNYNNNS		
N24-II-4	SLLNNT I HNNNYNDY	0,118	0,061
N24-II-8	SLSNKT I HNNNNYNNYNNNY		
N25-I-1	SLSNNYNNNNKYNNYNN	0,140	0,075
N25-I-3	SLSNNYNSNNYNNNNNNYNNNNYNNY		
N25-II-3	SLSNNYNSNNYNNNNNNYNNNNYNNY	0,130	0,051
N25-II-4	SLSNNYNNNNNNYNNKYNNYNNYNN		
N27-I-5	SLSNNYNNNNYNN	0,100	0,129
N27-I-6	SLSNKT I HNNNNYNNY		
N27-II-1	SLSNKT I HNNNNYNNYNNYNN	0,149	0,066
N27-II-2	SLSNNYKYSNNYNNNNYNNNNYNNYNNY		
N34-I-9	SLSNNYNNNNNNYNNYKYNNYNNY	0,156	0,171
N34-I-10	SLSNNYNSNNYNNYNNKYNYKYNN		
N34-II-4	SLSNNYNNNNNNYNNNYKCNYN	0,051	0,054
N34-II-6	SLSNNYNNNNNNYNNYKYNNYNNYNN		
N37-I-2	SLSNNYNNNNYNNKYNY	0,020	0,064
N37-I-3	SLSNNYKYSNNYNNNNYNNYNNNNYNNYNNNNYNNY		

N41-I-2	SLSNNYNNNNYKYN	0,147	0,136
N41-I-3	SLSNKTIHNNNNYDN		
N41-II-7	SLSNNTIHNNNNYNN	0,107	0,044
N41-II-10	SLSNKTIHNNNNYNN		
N45-II-1	SLSNNYNSNNNNYNNKYNNS	0,062	0,031
N45-II-2	SLSNNYNSNNNNYNN		
N48-I-2	SLSNNYNCNNNNNNYNNNSKKYC	0,126	0,135
N48-I-4	SLSNKTIHNNNNYNNY		
N48-II-4	SLSNNYNCNNNNNNYNNNSKKYC	0,092	0,139
N48-II-5	SLSNKTIHNNNNYNNNNYNN		
N50-I-6	NNYNYKYNNNYNNNDY	0,056	0,115
N50-I-8	SLSNNTIFYNNNN		
N58-I-1	SLSNSYNNNYKPKF	0,141	0,088
N58-I-8	SLSNNYNNNNNNYNYKYNNNNY		
N58-II	SLSNNYNYKYNNS	0,074	0,088
N58-II-9	SLSNNYNNNNNNYNYKYNNNNY		
N62-I-1	SLSNNYNYKYNNNNNYNNY	0,188	0,110
N62-I-2	SLSNNYNNNNNNYNN		
N62-II-3	SLSNNYNNNNNNYNYKYNNNNY	0,105	0,064
N62-II-7	SLSNNYNNNNNNYNN		
N65-I-5	SLSNNYNNNNNNYNNY	0,067	0,076
N65-I-6	SLSNKTIHNNNNYNNNNNC		
N73-I-5	SLSNKTIHNNNNLIHNNKNYN	0,074	0,061
N73-I-6	PLSNKTIHNNNNYNN		
N76-I-3	SLSNNYNNNNYKYN	0,092	0,139
N76-I-6	SLSNNYTCNNYNK		
N76-II-7	NNNSLSNNYNNNNYNNNNYNNKHN	0,068	0,090
N76-II-10	SLSNKTIHNNNNYKYNNNNNNYNNNNYNNNC		
N87-I-1	SLSNNYNN	0,123	0,122
N87-I-3	SLSNKTIHNNNNYNNNNYNNNN		
N87-II-2	SLLNNTIHNNNNYNDY	0,079	0,086
N87-II-3	SLSNNYNNNNNNYNNNYKCNYN		
N90-I-2	NNYNYKYNNNYNNNDY	0,094	0,088
N90-I-3	SLSNNYNNNNYKYN		
N90-II-1	SLSNNYNNNNYKYN	0,191	0,102
N90-II-3	NNYNYKYNNS		
N95-I-6	SLLNKTIHNNNNYNNNNYNNNNNNYNNY	0,088	0,083
N95-I-7	NNYNYKYNNNYNNNDY		

N95-II-6	SLSNNYISNNYNNYNNNNYKNYDY	0,083	0,045
N95-II-8	NNYNYKYNNNYYNNDY		
N97-II-1	SLSNKTIHNNNNYKYNYSNNNNYNNY	0,084	0,044
N97-II-2	SLSDKTIHNNNNYKYNYNNNYNNNNYNN		
N98-I-3	SLSNKTIHNNNKYNYNKYNNHYNNNS	0,123	0,066
N98-I-7	SLSNNYKYSNNYNNNNYNN		
N98-II-8	SLSNKTIHNNNNYKYNYNNNHYNNNNN	0,067	0,069
N98-II-9	SLSNNYNNNNYNNNNYNN		
N100-I-4	NNYNYKYNNNYYNNDY	0,084	0,078
N100-I-5	SLSTNNYNNNNNNI		
N106-I-1	SLSNKTIHNNNNYNNNNYNNNNYNNY	0,046	0,104
N106-I-7	NNYNYKYNNNYYNNDY		
N106-II-1	SLSNNYNSNNNNNNY	0,127	0,045
N106-II-8	SLSNNYKYSNFNNNNNNNNNNYS		

6.1.1 Genealogien ausgewählter *A. mellifera*-csd-Sequenzen

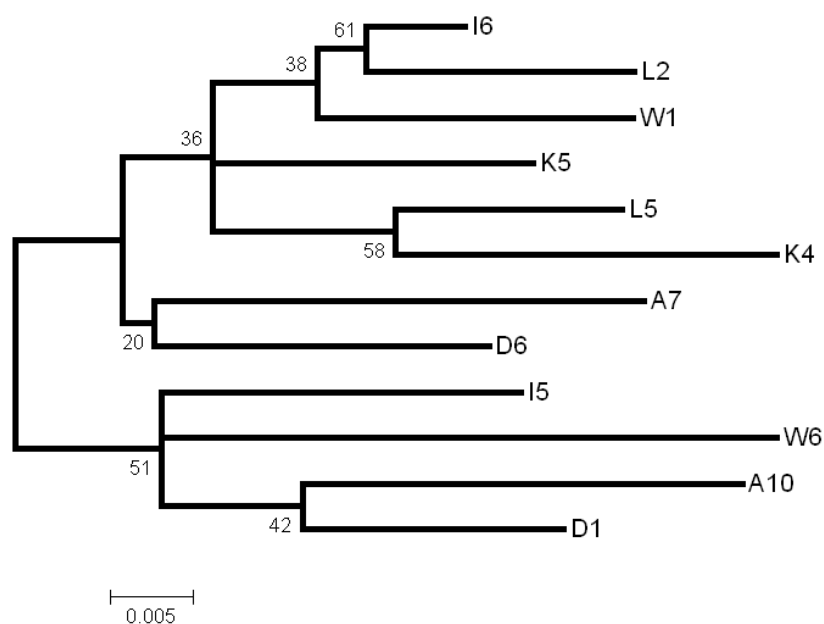


Abbildung 36: Genealogie der *A. mellifera*-csd-Sequenzen aus Texas, USA. Erstellt mit MEGA nach der Maximum-Likelihood-Methode und dem Tamura-Nei-Modell auf Basis der Nukleotidsubstitutionen.

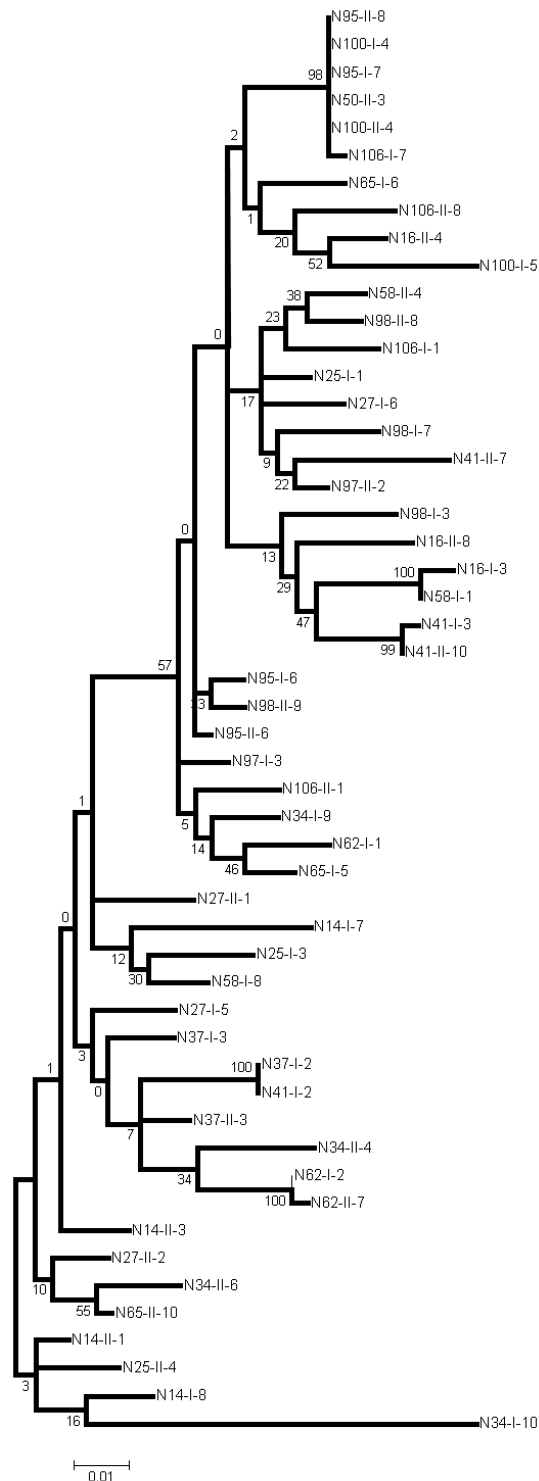


Abbildung 37: Genealogie der *A. mellifera*-*csd*-Sequenzen aus den kenianischen Savannenregionen. Erstellt mit MEGA nach der Maximum-Likelihood-Methode und dem Tamura-Nei-Modell auf Basis der Nukleotidsubstitutionen.

6.2 Sequenzen des Gens *csd* verschiedener *Apis*-Arten

6.2.1 Sequenz des *csd*-Gens aus *Apis florea*

Im Folgenden ist die mRNA-Sequenz des Gens *complementary sex determiner* aus *Apis florea* in 5'-3'-Richtung dargestellt. Das Start- und Stoppkodon ist in dunkelgrau, die 5'- sowie 3'UTR jeweils in hellgrau markiert.

```

1  AAAATGAAGT TAGCAATATA AATATAGCAT TCTCAAAGTA TTCCACTAAC GGTTTCTCTA
61  AGCATATAGG TGATTTATAG TATTTATTAT AAAAATGAAA CGAAATATAT CAAGTTATTC
121 GCATTGCGAT GAGAGATTTA GACAATCACG CAGTGAAGAT AGCGAAATTG GTCTGCGTTC
181 AAGAACAGAA GAACGATTAC GACGGAGACG CGAAGAATGG ATGATACAAC AAGAACTAGA
241 ACGAGAACAC GAAAGATTGA AGAAAAAGAT GATTTTAGAA TATGAACGTG CTCGTGAGAA
301 GAAACTAGCA TTATCAAAAA GAAGTAA AAC TAAAGTAGA TCCCCAGAAA GCCGAAATAA
361 AAATAATGCA CCAAATATAT CTGAAACATT TATATTATCC GAAAAATTAA AATCTTCAGA
421 TGGTACATCT TTATTTAGAG GACCAGAAGG TACTCAAGTT AGTGAAACAG AACTACGAAA
481 AATTAAGGTA GATATTCACA GAATTTTGCC AGGAAAAACA ACAACAAC TG ATGAACTTAA
541 ACGAGATATT ATCAATCCTG AAAATGTGAT AATCAAAAAGA AGAACGGGAG AAGGATCTAA
601 GCCTATATTT GAAAGAGAAG AAATAAAAAA TGTATTAACA AAAACAAATG AAATTACAGA
661 ACATCGTACA GTTTTAGCTG TAAATATTGA AAAATCAGAA AATGAAATTA AAACATGCAA
721 AAAATATGCT GTTTCATCGA ATTCTATCAG AAGTCGAAGT CATAATTTTC AACATACATC
781 GTCTCGTTAT TCCAGCGAAA AAAGTTGCAG TAGAGATAGA AATAGAGGAT ATAAAAAGAA
841 GGATAGGCAA TATGAAAAAT TATACAATGT AAAAGAAAAA CTCTTTACGA GAAAAACAA
901 GTCTTAAGCG TTATTCTCGT TCGAGAGAAC GAGAGCAAAG ATCATACAAA AATGAAAGTC
961 CTTATCGAGA ATATCGAGAA ACATCGAGAG AACGATCTCG AGATAGAAGA GAACGAGAAA
1021 GATCTAGAGA ACGTACAATA ATTTTCATCGA ATTCTTTATT GAGTTATTAT AATAATAATA
1081 ATAATAATAA TAATAATAAT AGTAATAATA ATAATGTTAA TTATTATATC GAACAAATTC
1141 CTGTTCTGTG TTATTATGGG ATTTTCCAC CAAGGCCAAT AATGGTTAGA TCTTGGGTTC
1201 CAATGAGGGG TCAAGTACCA GGATTTAGAC ATATTGGTCC ATTACCACCA TTTCCACCAC
1261 GATTTATTGC TCCAGATATG TGTAGATTTC GTCCTCCAAA TCCTAGATTT GGACGTATGT
1321 ATTAATGAAA AATGTGAAAA TCTTATATAA TATAAAGTTA TAATATGCGA ATAATGAATA
1381 ATGAAATATG AGATGACATT TGACAAAAAA TAAGAATCTA AAATCTGTCA TAGTATATAT
1441 CGAGATTGGA CAACTAGCGG GTCAAATGAT TTGTATACCT TTTCTTTGT TCAAAAGAAC
1501 CGTAAAAAAA TGGCTCTTTT CATATCTTTA AAAATTTTAG ATGACACAGG AATCTTTTCA
1561 CCATTAATAA ACCATATAAT ATATTTTTTT ATAAGTAAAA AAAAAAAA

```

6.2.2 Alignment der Proteinsequenz von Csd verschiedener *Apis*-Arten

Dieses Alignment zeigt die mRNA-Sequenzen der *csd*-Gene aus den vier *Apis*-Arten *A. mellifera*, *A. cerana*, *A. dorsata* und *A. florea*. Die übereinstimmenden Nukleotide sind in grau unterlegt, Unterschiede mit weiß. Lücken werden durch Punkte dargestellt.

Am_csd/1-412	MKRNIISNYSHHDEKFKQLRNEDNKIDLRSRKEERLQHRREVWLI	45
Ac_csd/1-464	MKRNTSSYLHRDERFKQLRSESESGLRRLTEDERLQHRREEWMI	45
Ad_csd/1-450	MKRNTSSYSYHDERFRQLRNEDNEIGLRSRKEERLQHKREEWMI	45
Af_csd/1-409	MKRNISSYSYHCDERFRQSRSESEIGLRSRTEERLRRRR.EEWMI	44
Am_csd/1-412	QQREREHERLMKKMILEYELRRIREIEKLGSEERSK..SRSPDSR	88
Ac_csd/1-464	QQREREHEKLLKKKMILEYELRRAREKKLALSKRSKTKSRSPETIQ	90
Ad_csd/1-450	QQREREYEKLLKKKRILEYELRRAREKK..LSKKSKTKSRSPENR	88
Af_csd/1-409	QQELEREHERLKKKMILEYE..RAREKKLALSKRSKTKSRSPESR	87
Am_csd/1-412	DRSNTSNTSKTVILSDKLESSDDISLFRGPEGIQTINATELQKIKL	133
Ac_csd/1-464	DRNNASNTSKTFILFEKSKSSDDTSLFRGPEGTQISATELRKIKV	135
Ad_csd/1-450	D...TSNISKTFILSKKLESSNDTSLFRGPESTQINATELRKIKI	130
Af_csd/1-409	NKNNAPNISETFILSEKLKSSDGTSLFRGPEGTQVSETELKIKV	132
Am_csd/1-412	EIHRDLPGKSTTTTVEVKRDIINPEDVILIRRTGEGSKPIFEREE	178
Ac_csd/1-464	DIHRVLP GKPTTTTDEVKRDIIINPEDVMVKRRRTGEGSKPIFEREE	180
Ad_csd/1-450	DIHRILPGKPTTTTDELKRDIIINPEDVMLKRRRTGEGSKPIFEREE	175
Af_csd/1-409	DIHRILPGK.TTTTDELKRDIIINPENVIKRRRTGEGSKPIFEREE	176
Am_csd/1-412	IKNVLTINKIKEHDTVLVVNIEKSGNE...S.KKYATSSNSLRN	219
Ac_csd/1-464	IKNVFTQITENTHGTVLVLNIEKSENETVTCKKYVAPS.NFLRN	224
Ad_csd/1-450	IKDIL.KTNEITKHRTVLVNVNIEKSENET.TCKNH.ATSSNSLRN	217
Af_csd/1-409	IKNVLTKTNEITEHRTVLAVNIEKSENETKTCKKYAV.SSNSIRN	220
Am_csd/1-412	RTHGFQHTSSRYSR.....	233
Ac_csd/1-464	QTHSFQHTSSRYSSFASQEIQEMAGLNKNICHLLCQHLYLPVQKEK	269
Ad_csd/1-450	RSRSFQHTSSSHSSLASQEIQEITGLNKNICHLLCQHLYLPVQKKN	262
Af_csd/1-409	RSHNFQHTSSRYSS.....	234
Am_csd/1-412	ERSCSRDRNREYKEKDRRYEKL.H.NEKEKLLERTSRKRY	272
Ac_csd/1-464	GVLRESCSGDRNKEYKKKDRQYEKLRTDNEKEKHSQEKTSRKRY	314
Ad_csd/1-450	AISRERSCSRNREYS.KKKDRQYEKL.Y.NEKEKLLQEKTSRKRY	304
Af_csd/1-409	EKSCSRDRNRGYKKKDRQYEKL.Y.NVKEKLLREKTSRKRY	273
Am_csd/1-412	SRSREREQNSYKNEREYRKYRETSKERSRDRRERERSKEPKIISS	317
Ac_csd/1-464	SRSREQEKSQYKNENSYREYRETSRERSRDRKERERSKEHKIISS	359
Ad_csd/1-450	SRSREREQKPHKNENSYREYRETSRERSRDRRERER..ECKIISS	347
Af_csd/1-409	SRSREREQRSYKNESPYREYRETSRERSRDRRERERSRERTIISS	318

Am_csd/1-412	LSNNYNYSNYNNYNNNNYNNYKKL.....YYNINYIEQI	350
Ac_csd/1-464	LSNNYNNYNNNNYNNNNYNNYNNKHYNNKHYNNKHLYYNINYIEQI	404
Ad_csd/1-450	LSNNYNYKNYNNNNYNNKHYNNKHYNNKHYNNKHYK..YYNINYIEQI	390
Af_csd/1-409	.NS.LLSYYNNNNNNNNNNNSNNNN.....VNYYYIEQI	348
Am_csd/1-412	PVPVPVPIIYC GNFP R P M . G . P W I S M Q E Q I P R F R Y I G P P T P F P . R	392
Ac_csd/1-464	PVPV....YYGNFPNPIMVRPWVPMQEQIPRFRYIGPLTSFPFR	445
Ad_csd/1-450	PVPV....YYGNFPSEPI TVK P W V S M Q K Q V P R F R Y I G P L T P F P . Q	430
Af_csd/1-409	PVPV....YYGIFPPRPIMVRSWVPMRGQVPGRHIGPLPPFPFR	389
Am_csd/1-412	FIPPNAYRFRPPLNPRFGPT	412
Ac_csd/1-464	FIPPNMYR.RPPLNPRFRPM	464
Ad_csd/1-450	FIPPN IYRPRPPLNPRFGPM	450
Af_csd/1-409	FIAPD MCRFRPPNPRFGRMY	409

6.3 Sequenzen des *feminizer*-Gens verschiedener Hymenopteren

6.3.1 Sequenz des Gens *fem* aus der asiatischen Zwerghonigbiene *Apis florea*

Die mRNA-Sequenz des Gens *feminizer* aus *A. florea* wurde in Vorarbeiten von Martin Hasselmann sequenziert. Start- und Stoppkodon sind in dunkelgrau, die 5'- und die 3'-UTR in hellgrau markiert. Die 5'-UTR wurde nicht vollständig sequenziert und ist angegeben soweit bekannt.

```

1 ATTTGCAGGT TAAATTTTCAT AAATATACAT ATACTTCTTC GGTGTTTATT ATAAAAATGA
61 AACGAAATAC ATCAAGTCAT TCGCATCGCG ATGAGAGATT TAGACAATCA CGCAGTGAAG
121 ATAGCGAAAC TGGTCTACGT TCAAGAACAG AAGAAGAACG AATACGACGT AGACGCGAAT
181 GGATGATACA ACAAGAACGA GAACGAGAAC ACGAAAGATT GAAGAAAAAG ATGATTTTAG
241 AATATGAACG TGCTCGTGAG AAGAACTAG CATTATCAAA AAGAAGTAAA ACTAAAAGTA
301 GATCCCCAGA AAGCCGAAAT AGAAGTAATG CACCAAATAC ATCTAAAACA TTTATATTAT
361 CCGAAAAATT AGAATCTTCC GATGGTACAT CTTTATTTAG AGGACCAGAA GGTACTCAAG
421 TTAGTGCAAC AGAACTACGA AAAATTAAGG TAGATATTCA CAGAATTTTG CCAGGAAAAA
481 CAACAACAAC AACTGATGAA CTTAAACGAG ATATTATCAA TCCTGAAGAT GTGATGCTCA
541 AAAGAAGAAC GGGAGAAGGA TCTAAGCCTA TATTTGAAAG AGAAGAAATA AAAAATGTAT
601 TAACAAAAAC AAATGAAATT ACAGAACATC GTACAGTTT AGCTGTAAAT ATTGAAAAAT
661 CAGAAAACGA AACTAAAACA TGCAAAAAAT ATGCAATTTT ATCAAATTCT ATCAGAAGTC
721 GAAGTCGTAG TTTTCAACGT ACATCATCCT GTCATTCCAG ATACGAGGAT TCAAGACATG
781 ACGACAGAAA TAGTTATAGA AATGATAGGG AAAGAAGTTG CAGTAGAGAT AGAAGTAGAG
841 AGTACAAAAA AAACAGACGA TATGAACCAT TGCACAATGT AGAAGAAAAG CATTTACGAG
901 AAAGAACAAG TCGTAGACGT TATTCACGTT CAAGAGAACG AGAACAAAAA TCATATAAAA
961 ATGAAAGAGA ATATCGAGAA TATCGAGAAA CATCGAGAGA ACGATCTCGA GATAGAAGAG
1021 AACGAGAAAG ATCTAGAGAA CGTAGAATAA TTCCATCTCA CTATATCGAA CAAATTCCTG
1081 TTCCTGTTTA TATTGGGAAT TTTCCATCAG AACCAATAAT GGTTAGACCT TGGGTTCCAA
1141 TGCGGGGTCA AGTACCAGGA TCTAGGCATA TTGGTCCATT AACACCATT CCACCACGAT
1201 TTATACCTCC AGATATGTAT AGACTTCGTC CACCTCCAAA TCCTAGATTT GGACCTATGT
1261 ATTAATGAAA AATGTGAAA TCTTATATAA TAGAAAGTTA TAATATGCAA ATAATGAATA

```

1321 ATGAATTGAA TAATATTGAA TAATGAGATG ACATTGACA AAAAAATAAA ATGTTAAAAT
 1381 CTATCATGAT ATATGTCAAG ATTGGACAAC TAGCTTATGA ATCAACTGTT TTTCGTTGTT
 1441 CAAAAAACC ATAAAAAAT GTTCCTTTCA TATCTCTAAA AAAAAACGTC TTCTGACTCA
 1501 ATAATCCTTC TTCCATTAAT AGATTACATA ATATATTTT TTAATCATT CGTTTTTCAT
 1561 AATCATATTT CATTTTGTT TACATATGTA ATAATGATTA ATATATTCAT GCTGAAAGAA
 1621 TCAATTAATT AAGAGTAAAA GTCATACAGT TCCTCGAAAA AAACGAGACG AACTATTTAA
 1681 TCGAGAATCA GAGAGCTCAC AAAAAAAAAA A

6.3.2 Sequenz des Gens *fem* aus der Baumhummel *Bombus hypnorum*

Hier dargestellt ist die Nukleotidsequenz des Gens *feminizer* aus *B. hypnorum* soweit bekannt. Sowohl die 5'- als auch die 3'-UTR wurde nicht vollständig sequenziert. Ebenso fehlen vermutlich 48bp innerhalb des Leserahmens, die nicht erfolgreich sequenziert werden konnten. Dieser Bereich ist mit Strichen dargestellt. Das Start- und Stoppkodon ist in dunkelgrau, die 5'- sowie 3'UTR jeweils in hellgrau markiert.

1 AACTACCCTA ACTTCCTCGA TTTCCAGCAA ATTTTTFACT TGTAGGTAA CTTCCACGCA
 61 TACGAAATTT TTCTTCCTTG GAAAAATGAAA CGAAACGTGT CGAATCATT CCATCGTGAT
 121 GAAAAGTTGA GATTATCACG CAGTGGAAAC TACGAGGCTG GTCTACGAAG AACAGAAGAA
 181 GAACGATTAC GGCGTAGGCG CGAATGGATG ATAGAACAAC AAAGGCTACG AGAACATGAA
 241 AAGTTGAAGG AGAAGAAGAT TATAGAATAT GAAATACGTC GTGCTCTTGA GAAAGGTTTA
 301 CCGCTGCCCCA GAGGAAGATT TTCACATCAC AGCCGCAGTG AGAGTAAAG CAAATCCACG
 361 GAAAATCGAC AAAGAACTGT TACATCAAAT ACATCTAATA CATCTATATT ATCTAAAAAA
 421 TTAGAACCGT CAGATGGTAC AACACCTTTA TTAAAGGAC CAGAAGGCAC ACAGGTTAGC
 481 GCTGCAGAAC TACGAAGAAT TAAAGTAGAT ATTCACAGGA ATGTACCAGG CAAAGCGACA
 541 ACTGATGATC TCCAACGAGA TATTATCAAT CCTGAAGATG TTCTGGTCAA AAGAAGAGAG
 601 GGAGAAGGAT CGAAACCTAT ATTTGAAAGA GAAGAAATAA AAGGTACAGC AGACAAAACA
 661 GATGAAATTG AGGAACGTCG TACAGTTGTT gttttaaata atgaaaattt agAAAACAAG
 721 TCGAAAGTGA ATAAAAAATA TACTACTTCT TCAAGAAATC GAAGCCCCAG TCCTCGACAT
 781 AGATCGTCCC GCCATGCCAG G----- -----
 841 -----G GAAGAAATCA CAGTAGAGGT AGAAGTAGAG AGTATAAAGA AACAGTCAGA
 901 CGACATAGTC AATTAAACAC TGCAGAGGAA GAGCATCTGC GTGAAAGAAA AGGTCATAGG
 961 GATCATTCAC GTTCAAGGGA AAGAGAACAA AGAGCATGGG TCAGAGATTC TCATCATAGG
 1021 TCATCTAAAG ATGAAAGATC TTATCGAGAA AGATCAAGAG AACGGTCCCG AGAAAGAAGA
 1081 GAAAGAGATA GGTCAAGAGA ACATAGGATA CCTCCATCGC ATTATATCGA GCAGATTCTT
 1141 GTTCCTGTTT ATTATGGAGA TTTTCCTCCG AGACCAATAA TGGTTGGATC CTTGGTTCCA
 1201 ATTCGAGGTC AGGTACCACC TTTAGGAGGA ACTAGACATA TGATTGGTCC ATTAAGACCA
 1261 TTTCCACCAC GATTTATTCC GCCAGATATG TATAGATTAC GTCCGCCTCT AAATCCTAGA
 1321 TTTAGGCCAA TGTTATTAATA GACTGGACGA GGAACGAGGA GAGAATGGTC GGTCCCAGGG
 1381 CCACCTCTCT TTTGGTAAGT GCGACCGAGG ACCAATTAAA AG

Darstellung der Aminosäuresequenz von Fem aus *B. hypnorum*. Der fehlende Sequenzbereich ist mit Strichen dargestellt.

MKRNVSNSHRDEKLRLSRSGNYEAGLRRTTEERLRRRREWMIEQQRLREHEKLKEKKII
EYEIRRALEKGLPLPRGRFSHHSRSESKSKSTENRQRTVTSNTSNTSILSKKLEPSDGTT
PLFKGPEGTVQSAAELRRIKVDIHRNVPKGATDDLQTDIINPEDVLVKRREGESKPIF
EREEIKGTADKTDEIEERRTVVVLNNENLENKSKVNKKYTTSSRNRSPPRHRSSRHAR-
-----GRNHSRGRSREYKETVRRHSQNTAEELRERKGRDHSRSRER
EQRAWVRDSSHRSKDESYRERSRERSRERRERDRSREHRIPPSHYIEQIPVPVYGD
PPRPIMVGLVPIRGQVPLGGTRHMIGPLRPFPPRFIPDMYRLRPPLNPRFRPM

6.3.3 Sequenz des Gens *fem* aus der Steinhummel *Bombus lapidarius*

Hier dargestellt ist die Nukleotidsequenz des Gens *feminizer* aus *B. lapidarius* soweit bekannt. Sowohl die 5'- als auch die 3'-UTR wurde nicht vollständig sequenziert. Das Start- und Stoppkodon ist in dunkelgrau, die 5'- sowie 3'UTR jeweils in hellgrau markiert

1	AACTACCCTA	CTTCCTCGAT	TTCCAGCAAA	TTTTTTACTT	GTAGGTAAAC	TTCCACGCAT
61	ACGAAGTTTT	TCTTCCTTGG	AAAATGAAAC	GAAACGTGTC	GAATCATTCC	CATCGTGATG
121	AAAAGTTGAG	ATTATCACGC	AATGGAAATT	ACGAGACTGG	ACTACGAAGA	ACAGAAGAAG
181	AACGATTGCG	GCGTAGGCGC	GAATGGATGA	TAGAACAACA	AAGGCTACTA	GAACATGAAA
241	AGTTGAAGGA	GAAGAAGATT	TTAGAATATG	AAATACGTCG	TGCTCTTGAG	AAAGGTTTAC
301	CGCTGCCCAG	AGGAAGATTT	TCACATCATA	GCCGTAGTAA	GAGTAAAAGC	AAATCCACGG
361	AAAATCGACA	TAGAACTGTT	ACATCAAATA	CATCTAATAC	ATCTATATTA	TCTAAAAAAT
421	TAGAACCATC	TGATGGTACA	ACACCTTTAT	TTAAAGGACC	AGAAGGCACA	CAGGTTAGCG
481	CTGCAGAACT	ACGAAGAATT	AAAGTAGATA	TTCACAGGAA	TGTACCAGGC	AAAGCTACAA
541	CTGATGATCT	CCAACGAGAT	ATTATCAATC	CTGAAGATGC	TCTGGTCAAA	AGAAGAGAGG
601	GAGAAGGATC	GAAACCTATA	TTTGAAAGAG	AAGAAATAAA	AGGCACAACA	GTCAAAACAG
661	ATGAAATTGT	GGAACGTCGT	ACAGCTGTTG	TTTTAAATAA	TGAAAATTTA	GAAAGCAAGT
721	CGAAAGTGAA	TAAAAACTAT	ACTACTTCTT	CAAGAAATCG	AAGCCCCAGT	CCTCGACATA
781	GATCGTCCCG	CCATGCCAGG	TATGAGGATT	CAAAATATGA	TAACAGAGGC	AGTTATAGAA
841	GTGATAGAGG	AAGAAATCAC	AGTAGAGGTA	GAAGTAGAGA	GTATAAAGAA	ACAGTCAGTC
901	GACATAGTCA	ATTAAACACT	GCAGAGGAAG	AGCATCTTCG	TGAAAGAAAA	GGTCATAGGG
961	ATCATTACAG	TTCAAGGGAA	CGAGAACAAA	GAGCATGGGT	CAGAGATTCT	CATCATAGGT
1021	CATCTAAAGA	TGAAAGATCT	TATCGAGAAA	GATCAAGAGA	ACGGTCCCGA	GAAAGAAGAG
1081	AAAGAGATAG	GTCTAGAGTA	CATAGGATAC	CTCCATCGCA	TTATATCGAG	CAGATTCTTG
1141	TTCTGTGTTA	TTATGGAGAT	TTTCCTCCGA	GACCAATAAT	GGTTGGACCC	TCCGTCCCAA
1201	TTGAGGTCA	GGTACCACCT	TTAGGAGGAA	CTAGGCATAT	GATGGGTCCA	TTAAGACCAT
1261	TTCCACCACG	ATTTATTCCG	CCAGATATGT	ATAGATTACG	TCCGCCTCTA	AATCCTAGAT
1321	TTAGGCCAAT	GTATTAATAG	ACTGGACGAG	GAACGAGGAG	AGGATGGTCC	GTCCAGGGC
1381	CACCTCTCTT	TCCGTAAGTG	CGACCGGGGA	CTAATTAAAA	G	

Im Folgenden ist die Aminosäuresequenz von Fem aus *B. lapidarius* dargestellt.

MKRNVSNSHRDEKLRLSRNGNYETGLRRTEERLRRRREWMIEQQRLLLEHEKLKEKKILE
YEIRRALEKGLPLPRGRFSHRSKSKSKSTENRHRTVTSNTSNTSILSKLEPSDGTTT
LFGKPEGTVQVSAAEELRIKVDIHRNVPGKATTDLDLQDIINPEDALVKRREGESKPIFE
REEIKGTTVKTDIVERRTAVVLNNENLESKSKVNKNYTTSSRNRSPPRHRSSRHARYE
DSKYDNRGSYRSRGRNHSRGRSREYKETVSRHSQNLNTAEELRLERKGRDHSRERERE
QRAWVRDSSHRSKDESYRERSRERSRERRERDRSRVHRIPPSHYIEQIPVPVYYGDFP
PRPIMVGPSVPPIRGQVPLGGTRHMMGPLRPFPPRFIPPDMYRLRPPLNPRFRPM

6.3.4 Sequenz des Gens *fem* aus der nordamerikanischen Hummel *Bombus impatiens*

Hier gezeigt ist die Nukleotidsequenz des Gens *fem* aus *B. impatiens* soweit bekannt. Die Sequenz wurde im Abgleich mit der *B. terrestris-fem*-Sequenz aus genomischer DNA-Sequenz abgeleitet (GenBank-Nummer AEQM00000000.1 bei NCBI). Die 5'- sowie 3'-UTR konnte nicht vollständig abgeleitet werden, sie ist dargestellt soweit bekannt. Das Start- und Stoppkodon ist in dunkelgrau, die 5'- sowie 3'UTR jeweils in hellgrau markiert.

1 AACTACCCTA ACTTCCTCAA CTTCCAGCAA ATTTTTTGTT AACTTCCACG CATACGAAAT
61 TTTTCTTCCT TGGAAAATGA AACGAAACGT GTCGAATCAT TCCCATCGTG ATGAAAAGTT
121 GAGATTATCA CGCAATGGAA ACTACGAGGC TGGTCTACGA AGAACAGAAG AAGAACGATT
181 ACGGCGTAGG CGCGAATGGA TGATAGAACA ACAAAGGCTA CGAGAACATG AAAAGTTGAA
241 GGAGAAGAAG ATTTTAGAAT ATGAAATACG TCGTGCTATT GAGAAAGGTT TACCACTGCC
301 CAGAGGAAGA TTTTCACATC ACAGCCGCAG TAAGAGTAAA AGCAAATCCA CGGAAAATCG
361 ACATAGAACT GTTACATCAA ATACATCTAA TACATCTATA TTATCTAAAA AATTAGAACC
421 ATCAGATGGT ACAACACCTT TATTTAAAGG ACCAGAAGGC ACACAGGTTA GCGCTGCAGA
481 ACTACGAAGA ATTAAAGTAG ATATTCACAG GAATGTACCA GGCAAAGCGA CAACTGACGA
541 TCTCCAACGA GATATTATCA ATCCTGAAGA TGTTCTGGTC AAAAGAAGAG AGGGAGAAGG
601 ATCGAAACCT ATATTTGAAA GAGAAGAAAT AAAAGGTACA ACAGTCAAAA CAGATGAAAT
661 TGAGGAACGT CGTACAGTTG TTGTTTTAAA TAATGAAAAT TTAGAAAGCA AGTCGAAAAT
721 GAATAAAAAA TATACTACTT CTTCAAGAAA TCGAAGCCCC AGCCCTCGAC ATAGATCATC
781 CCGCCATGCC AGGTTTGAGG ATTCAAAATA TGATAACAGG GACAGTTATA GAAGTGATAG
841 AGGAAGAAAT CACAGTAGAG GTAGAAGTAG AGAGTATAAA GAAACAGTTA GACGACATAG
901 TCAATTAAAC ACTGCAGAGG AAGAGCATCT GCGTGAAAGA AAAGGTCATA GGGATCATTC
961 ACGTTCAAGG GAACGAGAAC AAAGAGCATG GGTGAGAGAT TCTCATCATA GGTCATCTAA
1021 AGATGAAAGA TCTTATCGAG AAAGATCAAG AGAACGGTCC CGAGAAAGAA GAGAAAGAGA
1081 TAGGTCTAGA GAACATAGGA TACCTCCATC GCATTATATC GAGCAGATTC CTGTTCTGT
1141 TTATTATGGA GATTTTCCTC CGAGACCAAT AATGGTTGGA CCCTTGTTTC CAATTGAGG
1201 TCAGGTACCA CCTTTAGGAG GAACTAGACA TATGATTGGT CCATTAAGAC CATTTCCACC
1261 ACGATTTATT CCGCCAGATA TGTATAGATT ACGTCCGCCT CTAAATCCTA GATTTAGGCC
1321 AATGTATTAG TAGACTGGAC GAGGAACGAA GAGAGAATGG TCAGTCCCAG GGCCACCTCT
1381 CTTTCGGTAA GTGCGACCGG GGACCAATTA AAAGCGTTGG AGGGGGGTGG TGTCGATAGG
1441 ACATT

Darstellung der Proteinsequenz von Fem aus *B. impatiens*.

MKRNVSNSHRDEKLRLSRNGNYEAGLRRTEERLRRRREWMIEQQRLREHEKLKEKKIL
EYEIRRAIEKGLPLPRGRFSHHSRSKSKSKSTENRHRTVTSNTSNTSILSKKLEPSDGTT
PLFKGPEGTVQVSAELRRIKVDIHRNVPKGATDDDLQRDIINPEDVLVKRREGESKPIF
EREEIKGTTVKTDIEERRTVVVLNNENLESKSKMNKYYTTSSRNRSPPRHRSSRHARF
EDSKYDNRDSYRSDRGRNHSRGRSREYKETVRRHSQNLNTAEHLRERKGRHDHSRSRER
EQRAWVRDSSHRSKDESYRERSRERSRERRERDRSREHRIPPSHYIEQIPVPVYGGDF
PPRPIMVGPLVPIRGQVPLGGTRHMIGPLRPFPPRFIPDMYRLRPPLNPRFRPM

6.3.5 Sequenz des Gens *fem* aus der Erdhummel *Bombus terrestris*

Im folgenden ist die mRNA-Sequenz des Gens *feminizer* aus *B. terrestris* inklusive des männchenspezifischen Exons angegeben, welches in hellgrau markiert ist. Start- und Stoppkodon sind in dunkelgrau, die 5'- und die 3'-UTR in mittelgrau markiert. Diese Sequenz wurde in Vorarbeiten von Christina Schulte sequenziert. Die weibchenspezifische Spleißform ist in der NCBI-Datenbank unter der Nummer EU288185.1 veröffentlicht.

1	AACTACCCTA	ACTTCCTCGA	TTTCCAGCAA	ATTTTTTGT	AACTTCCACG	CATATGAAAT
61	TTTTCTTCT	TGGAATGAA	AACGAAACGT	GTGGAATCAT	TCCCATCGTG	ATGAAAAGTT
121	GAGATTATCA	CGCAATGGAA	ATTACGAGGC	TGGTCTACGA	AGAACAGAAG	AAGAACGATT
181	ACGGCGTAGG	CGCGAATGGA	TGATAGGACA	ACAAAGGCTA	CGAGAACATG	AAAAGTTGAA
241	GGAGAAGAAG	ATTTTAGAAT	ATGAAATACG	TCGTGCTCTT	GAGAAAGGTT	TACCGCTGCC
301	CAGAGGAAGA	TTTTCACGTC	ACAGCCGCAG	TAAGAGTAAA	AGCAAATCCA	CGGAAAATCG
361	ACATAGAACT	GTTACATCAA	ATACATCTAA	TACATCTATA	TTATCCAAAA	AATTAGAACC
421	ATCAGATGGT	ACAACACCTT	TATTTAAAGG	ACCAGAAGGC	ACACAGGTTA	GCGCTGCAGA
481	ACTACGAAGA	ATTAAAGTAG	ATATTACAG	GAATGTACCA	GGCAAAGCGA	CAACTGATGA
541	TCTCCAACGA	GATATTATCA	ATCCTGAAGA	TGTTCTGGTC	AAAAGAAGAG	AGGGACGAAG
601	AAAATAGGAG	AAGGATCGAA	ACCCATATTT	GAATAAGGAG	AAATAAAAAG	TACAACAGTT
661	AAAAACTGTA	GAAATTGTGG	AATGTGATAC	AGTCGTTGCT	ATAAATAATA	AAAATTTAGG
721	TATATATCAA	CAAAGTAAAG	GACCAAAAGG	TACATTAGTT	TGTGCTGTGA	AACTACGAAG
781	AATCAAATGG	TTATTGGGCA	GAAATTTGTC	TAGAAAAGTG	ATAACTAATA	ATCTTCAACA
841	AGATATGTAT	CTAGAAGATG	TAACAGTCAG	AAGAGGAAGG	GGTAAAATGT	TTTTTTTCTA
901	CAATTATCTA	ACAACCTTCA	GTCATCCAAT	GAAGAAATAC	TTAAGGAATA	TATGAATTCA
961	ACTTGAAAGA	TTATCCTTTT	GGAATTTTTC	GAAATCATA	ATTTATAAGT	TCAAGTTCCA
1021	ATGGATTGTA	TGCTTTGTAA	TGTAAAATTG	AGAACCACAT	GTTATATTTT	TCCCTTTTTC
1081	TTTAATGTGG	TTAAAATTTG	ATTAAGAATA	GCAAAATTTT	CACTTTATTA	AATTCATCAG
1141	CTGAAGAATG	ATGACATTAT	GTGGAAGTTT	CTACAGACAT	GAAGATAATG	GATTTATCAT
1201	CATATTGTTT	CCTAGAGCAT	TCAACACAAT	TTACACAATG	TTGAAGTCAA	TAATGTGTTT
1261	TTGATTACCC	AAGGATGAAG	AAAATTTGGA	ATACAACGTT	GGATTTGAAT	ATTTGGCTGA
1321	GACCTGGACG	TCCTTCGACA	CTCAAATGCA	CTCTCACACC	ATGTTTACCA	TTTCATTAC
1381	TTCATACATC	TATTATAAAA	TGTCCTGGCT	TACAATCGTT	CAGAGTACGT	TAGATGTTTA
1441	TTTATAAATA	AATTTATTTA	ACCATAGGAG	AAGGATCAAA	ACCTATATTT	GAAAGAGAAG
1501	AAATAAAAAG	TACAACAGTC	AAAGCAGATG	AAATTGTGGA	ACGTCGCACA	GTTGTTGTTT
1561	TAAATAATGA	AAATTTAGAA	AGCAAGTCGA	AAGTGAATAA	AAAATATACT	ACTTCTTCAA
1621	GAAATCGAAG	CCCCAGTCCT	CGACATAGAT	CGTCCCGCCA	TGCCAGGTAT	GAGGATTCAA

1681 AATATGATAA CAGAGACAGT TATAGAAGTG ATAGAGGAAG AAATCACAGT AGAGATAGAA
 1741 GTAGAGAGTA TAAAGAAACA GTCAAACGAC ATAGTCAATT AAACACTGCA GAGGAAGAGC
 1801 ATCTGCGTGA AAGAAAAGGT CATAGGGATC ATTCACGTTT AAGGGAACGA GAACAAAGAG
 1861 CATGGGTCAG AGATTCTAAT CATAGGTCAT CCAAAGATGA AAGATCTTAT CGAGAAAGAT
 1921 CAAGAGAACG GTCCCGAGAA AGGAGAGAAA GAGATAGGTC TAGAGAACAT AGGATACCTC
 1981 CACCGCATT TATCGAGCAG ATTCCTGTTT CTGTTTATTA TGGAGATTTT CCTCCGAGAC
 2041 CAATAATGGT TGGACCCTTG GTTCCAATTC GAGGTCAAGT ACCACCTTTG GGAGGAACCTA
 2101 GGCATATGAT GGGTCCATT AAGACATTTT CACCACGATT TATTCCGCCA GATATGTATA
 2161 GATTACGTCC GCCTCTAAAT CCTAGATTTA GGCCAATGTA TTAATAGACT GGATGAGGAA
 2221 TGAAGAGAGG ATGGTTCGGT CCAGGGCTAC CTCTCTTTTCG ATGAGTGCGA CCGGGGACCA
 2281 ATTAAGAGAG TTGGAAGGGA ATGGTGCCGA TAGAATCG

6.3.6 Sequenz des Gens *fem* aus der Prachtbiene *Euglossa hemichlora*

Die Nukleotidsequenz ist angegeben soweit bekannt. Die 5'-UTR sowie ein Fragment von 34 Nukleotide zu Beginn des Leserahmens fehlen (davon ausgehend, dass dieser Bereich in *Eh-fem* im ähnlichen Maße konserviert ist wie in den anderen bekannten *fem*-Sequenzen). Die 3'-UTR (hellgrau) und das Stoppkodon (dunkelgrau) sind markiert. Die Sequenz ist in der NCBI-Datenbank unter der Nummer HM584944.1 veröffentlicht.

1 ATGAGAAGTG TAAATCATCA CGCAGTGAAG ATTGTGAGAC TGGTTCGCGC GCAAGAACAG
 61 AAGAAGAACG ATTACGACGT AGACGCGAAT GGATGATACA AAAAGAGCGG GAACGAGAAC
 121 ACGAAAAATT GAAGGAAAAA ATGATTTTAG AATATGAAAT ACAACGTGCT CGTCAAGAGG
 181 GTTTACCTCT ACCTAAAAGA AGATTATCTA GTCAAAGTCG TAGTAAGAAT GAAAGCAAGT
 241 CTCCAGAAAA TCGACATAGA AGAACAGCTG TATCAAATAC GTGTAAAACA TCTGTATTAT
 301 CTGAAAAAGT AGAATCATCA GATGGTACAA CACCCCTTATT TAAAGGCCCA GAAGGTACAC
 361 AAGTTAGTGC TACAGAACTG CGAAGAATAA AAGTAGATAT TCATAGAAAT ATACCGGGTA
 421 AAACAACAAC TGGTGAACCT CAACGAGATA TTGTCAATCC TGAAGATGTG GTGCTGAAAA
 481 GAAGAACGGG AGAGGGATCA AAACCTATAT TTGAAAGAGA AGAAATAAAA GGTGCAACTA
 541 AAACAGAAGA AATTGAGGAA CATCGTACAG TTGTAGCAAT AAATAATGAA AATTTAGAAA
 601 ATAAATCGAA AATATGTAAA AAATTTGCTA CTTCAAGTCC TTTAAGAAGT CGAAGTCGTA
 661 GTCCTCGACG TACAGCATCT CATCATTCCA GGTATGACAA TTCAAACAT AATAAGGATA
 721 GTTACAGAAA TGATAGGGAA AGAGATCGTA GTAGAGGTAA AAGCAAAGAG TATCAAGAAA
 781 AAGATAGAAG ACACGGTGAA TCGCGCGATA TAGAGGAAGA TCACTTACGA GAAAGAAAAA
 841 GTCATCGAGA TCATTCACAT TCAAGGGAGC GAGAAGAAAAG AAAATGGGAT AGAGATTCTT
 901 CCCATCATAG GTTTAAAGAC GAAAGATTTT ATCGAGAATA TCAAGAAAGA TCGAGAGGAC
 961 GATCTCGTGA AAGAAGAGAC AGAGATAGGT CTAGAGAACG TGGGCTACCG CCATCACACT
 1021 TCATCGAACA GATTCCTGTT CCTTTTTATT ACGGTAACCT CCCTCCGAGA CCAATAATGG
 1081 TTGGACCTTT GGTCCAATA CGAGGACAGG TTCTCTAAG TAGTAGACAT ATGATGGGTT
 1141 CATTAAAGACC GTTTCCACCA CGATTTATTC CTCCAGATAT GTATAGATTA CGTCCACCAC
 1201 CAAATCCTAG ATTTGGGCCA ATGTATTAAT GGAAGACTTG TTATAAAATC ATATATGATG
 1261 TTGATTTATT TATTTAAACG AAATTTACAA AGGAATAGAA AATAAAGAAA TAATTCTGTC
 1321 AAAAAAAAAA AAAA

Die Proteinsequenz ist gezeigt soweit bekannt. Geschätzte elf Aminosäuren fehlen zu Beginn der Sequenz.

EKCKSSRSEDCE TGSRARTEEEERLRRRREWMIQKEREREHEKLKEK MILEYEIQRARQEG
LPLPKRRLSSQSRSKNESKSPENRHRRTAVSNTCKTSVLSEKVESSDGTTP LFKGPEGTQ
VSATELRRIKVDIHRNIPGKTTT GELQRDIVNPEDVVLKRRTGEGSKPIFEREEIKGATK
TEEIEEHRTVVAINNENLENKSKICKKFATSSPLRSRSRSPRRTASHHSRYDNSKH NKDS
YRNDRERDRSRGKSKEYQEKDRRHGESRDIEEDHLRERKSHRDHSHSREREERK WDRDSS
HHRFKDERFYREYQERSRGRSRERRDRDRSRERGLPPSHFIEQIPVPFYGNFPPRPIMV
GPLVPIRGQVPLSSRHMMGSLRPFPPRFIPDMYRLRPPPNPRFGPMY

6.3.7 Sequenz des Gens *fem* aus der parasitischen Wespe *Nasonia vitripennis*

Hier ist die Nukleotidsequenz der *fem*-mRNA-Sequenz aus der parasitischen Wespe *Nasonia vitripennis* gezeigt. Das männchenspezifische Exon, das in Weibchen durch alternatives Spleißen entfernt wird, ist hier ebenfalls angegeben und in hellgrau hinterlegt. Das Start- und Stoppkodon ist in dunkelgrau, die 5'- sowie 3'UTR jeweils in mittelgrau markiert. Die Sequenzen sind in der NCBI-Datenbank unter den Nummern NM001134827 (weibchenspezifische Spleißform) und FJ159104 (männchenspezifische Spleißform) veröffentlicht.

1	GCAGGCGCGT	CGCAATACTC	GTGGGTTATA	TATTCAAATA	ACACTCCGAC	AGACTAACTG
61	ATAAATAGAG	GTTCCGGATG	CATAGATTGT	CATTTCGAGTA	GAGATTGTCT	TAGAGTAAAC
121	ATTGGGCTGC	TCCACTTTGC	ACTGTTCCACC	TGTTTTTTGCT	CGCGTTTTTG	TTTAAATGCG
181	CTCGCGCTGC	GATTCTAACC	TCAAACCTAA	TGAAAATAGC	AAAAAGCTTC	GAGTTGCTAC
241	AGTCATAGCT	TTTTGCAGTG	TGTTTTTTGTT	CCTCGTGTGT	GATACATTGC	AATTATGAGG
301	AGAAGAAGTT	CCTTAGATGA	GCTGGAGAGA	GAAAGAGAGA	AGGAAAAGCG	AAGGATTGCT
361	TGGATGGTAC	AGCAAGAGAA	GGAGCGAGAG	CATGAGAGAC	TCAAGCGAAA	GAAGATAGAA
421	GAGTACGAGA	GGAAAAGAGC	TGAGCAGCTG	GGCCTGAACC	GCAAGCGAAG	ATCACCACAC
481	AGAAGCCATA	GTGAAAGCAG	AAGCCGTAGT	AGGTCTGAAG	AATCGCATCA	TAGAAGCCGA
541	GCCAAGCATA	GAAGCAGCAG	ATCTCAAGTC	ATGTCAGAAA	AACTCGATTTC	ATCAGAAGGA
601	AACAAACCTT	TTTTCAATGG	ACCAAAAAGAG	GCACCAAAAC	TAGATGAGAT	CGAACTGCGA
661	CAAGTTGTAG	TGAACATCCA	TAGAAAAATT	CCTGCATCTG	CAACTGAAAC	TAGTGAAATT
721	CGACGTGATA	TTGTGAATCA	TGAGGAAATA	ACCCTGAAGA	GAAGAGACGG	TAAC TTGGTG
781	CTTCCTTATT	TTTTTTTTTAA	TCTACGATAC	AGGGTGACAG	AAGTTCAAGA	TTCTGTGCAC
841	TGAAGAAATA	CAGATGCGAC	TCAAATCTA	ACGAGAAGAT	ACATTACATC	ATGTAGAGGA
901	CATCACATTT	TTTGAAGATT	ATATCCTTGA	TCAATGTCAT	TAATCGATAT	TTAACCATGC
961	GATAGTATAA	TTTTCTTTAT	ATAGGTTTCT	TCATCTTACA	GAAGAAAGTA	TTAAGAAGAT
1021	AGTATTAAGA	AGATGGTACG	CAAAGTCATC	ACTGTGAGAA	GGAGCAAAAC	CTATATTTGA
1081	TCGAGAGGAA	CTAAAACAGT	TTGAAGCTAA	GTGATTGAA	GTGGAAGAGC	GCCGCACAAT
1141	CGAGGCTGTC	GGCAAAAATA	GTTTAGTTAA	GAGATCGTAT	GCAAAGCGCC	GTTCTAAGTC
1201	ATTGAGTCCT	CGCAGGCATC	GGAGTCCTAG	TCCTCACGCT	CACGGTCACT	CTGGTTCGCG
1261	TTTATACACA	CATCGATCGA	GGCACGACGA	GTA CTGAGA	CATCAGTTAG	AAGATCGACA
1321	TTACAAAGAT	CACAAAAACG	ATAGGCATTA	TTCCGATTTA	AGAGATACTA	GTCGAGATGA
1381	GTCAAGAAGA	CGTCGGTCAT	CTCGAGGAAG	ATCGCACTCT	AGAGAACGGG	AGTACCACAG

1441 AGATGAGCAT GACTATTAC GATCAAGCAG AAGTGAAAAG GAATACCGAG AATACCGAGG
 1501 CCGATCAAAA GATCGATCTT ACGACAGACG AGATAGAAGA GATTCCTCCA GGAACGACG
 1561 ACTACCCCCC GTCCAATACA TAGAACAAGT GCCATTCCCC ATTTACTATA GTGCGGGATT
 1621 CGCCCCAAGA CCAATGCTAG TCGGACCCGC CATTCTCTCC ATCCGAGGGC CGCTGGCCGG
 1681 CAGAGGCCGC CCGCTCATGG CACCCATTCTG ACCACCTTC CCGCCACGTT TCATAGGACC
 1741 ACAAGACATG TACAGGGCTG GGCTACCTTC CGATCCTCGA TTTGGACAGA TGTATTGAAT
 1801 CAAATTATTT GAATGAATTG AATCTATTGT ATGAGATACT TTATAAAAGG TCGACGTAGG
 1861 ACTATAAAAC GTGAAACGTG GAAGCTTCAA CGACGATGAA ATAACGTTAG TCTTTAATGT
 1921 TTGCTACTTA AAGGACGCGC TTTTAATGAT GATGAATTCT TGTGATCGAA TCAAACGATT
 1981 TTTTAATGTC AACCGCGAAG TGATATAATT GATTAGAGTC ATTATAACGA TTTTTTGT
 2041 TGAAAAAACT GTTGCTATTT CTTTAAGATT ATGCAGACAA GTTTGTATTA ATTACAAAAA
 2101 TATTATATAT GTAAATATAA CTATTTACTT TACCATTTTT ATACTCTA

Proteinsequenz des weiblichen, aktiven Fem aus *Nasonia vitripennis*

MRRRSSLDELEREREKEKRRIAWMVQKEKEREHERLKRKKIEEYERKRAEQLGLNRKRRS
 PHRSHSESRSRSRSEESHRSRAKHRSSRSQVMSEKLDSEGNKPFNGPKEAPKLDEIE
 LRQVVVNIHRKIPASATETSEIRRDIVNHEEITLKRRDGEGAKPIFDREELKQFEAKSIE
 VEERTIEAVGKNSLVKRSYAKRRSKSLSPRRHRSPSPHAHGHSRSLYTHRSRHDEYSR
 HQLEDHRHYKDHKNDRHYSDLRDTSRDESRRRRSSRGRSHSREREYHRDEHDYSRSSRSEK
 EYREYRGRSKDRSYDRDRDRDSSRERRLPVQYIEQVPFPIIYSAGFAPRMLVGPAIPP
 IRGPLAGRGRPLMAPIRPPFPFRFIGPQDMYRAGLPSDPRFGQMY

6.3.8 Alignment der Fem-Sequenzen verschiedener Apidae

Alignment der Fem-Aminosäuresequenzen von *Apis mellifera*, *A. cerana*, *A. dorsata*, *A. florea*, *Bombus terrestris*, *B. impatiens*, *B. hypnorum*, *B. lapidarius*, *Melipona compressipes* (GenBank-Nummer: EU139305) und *Euglossa hemichlora*. Die übereinstimmenden Aminosäuren sind in grau unterlegt, Unterschiede mit weiß. Lücken werden durch Punkte dargestellt.

Am_fem/1-400	MKRNTTNHSHHDERFRQSRSEDS	SETGLRSRTQEERLRRRREWMIQ	45	
Ac_fem/1-404	MKRNISNLSHRNERFRQSRSEDS	SETGLRSRTEEERLRRRREWMIQ	45	
Ad_fem/1-404	MKRNISSHSHRDERFRQSRSEDS	SETGLRSRTEEERLRRRREWMIQ	45	
Af_fem/1-401	MKRNTSSSHSRNERFRQSRSEDS	SETGLRSRTEEERIRRRREWMIQ	45	
Bt_fem/1-416	MKRNVSNHSHRDERLRLSRNGNYEAGLR	.RTEEERLRRRREWMIE	44	
Bi_fem/1-416	MKRNVSNHSHRDEKLRLSRNGNYEAGLR	.RTEEERLRRRREWMIE	44	
Bh_fem/1-400	MKRNVSNHSHRDEKLRLSRNGNYEAGLR	.RTEEERLRRRREWMIE	44	
Bl_fem/1-416	MKRNVSNHSHRDEKLRLSRNGNYETGLR	.RTEEERLRRRREWMIE	44	
Mc_fem/1-410	MKRIMS	SHSHRDESRISRS	GDFETGLRSKTEEERLRRRREWMIE	45
Eh_fem/1-407		EKCKSSRSEDCETGSRARTEEERLRRRREWMIQ	33	

ANHANG

Am_fem/1-400	QEREREHERLK KK MILEYELRRAREKK ..LSKRS ..KS	79
Ac_fem/1-404	QEREREHEKLK KK MILEYELRRAREKK LALSKRSK TKS	83
Ad_fem/1-404	QEREREHERLK KK MILEYELRRAREKK LALSKRSK TKS	83
Af_fem/1-401	QEREREHERLK KK MILEYE .. RAREKK LALSKRSK TKS	81
Bt_fem/1-416	QQRLREHEKLKEK KK ILEYEIRRALEKGLPLPRGRFSRHRSRKS KS	89
Bi_fem/1-416	QQRLREHEKLKEK KK ILEYEIRRAIEKGLPLPRGRFSHHRSRKS KS	89
Bh_fem/1-400	QQRLREHEKLKEK KK ILEYEIRRALEKGLPLPRGRFSHHRSR ESKS	89
Bl_fem/1-416	QQRLLEHEKLKEK KK ILEYEIRRALEKGLPLPRGRFSHHRSRKS KS	89
Mc_fem/1-410	QEK L REHEKLK AK MILEYEIRRALEKGLPPPKGRFSHHRSRKS KS	90
Eh_fem/1-407	KEREREHEKLKEKMILEYEIQ RARQE GLPLPK RRLSSQ SRSK NES	78

Am_fem/1-400	RSPES ..RGR .SNASNTSKTFILSEKLESSDG ..TSLFRG PEGTQVS	121
Ac_fem/1-404	RSPES ..RGR .SNASNTSKTFILSEKLESSDG ..TSLFRG PEGTQVS	125
Ad_fem/1-404	RSPESR GGR .SNTSNTSKTFILSEKLESSDG ..TSLFRG PEGTQVS	126
Af_fem/1-401	RSPES ..RNR .SNAPNTSKTFILSEKLESSDG ..TSLFRG PEGTQVS	123
Bt_fem/1-416	KSTEN ..RHR .TVTSNTSNTSILSKKLEPSDGTTPLFKGPEGTQVS	132
Bi_fem/1-416	KSTEN ..RHR .TVTSNTSNTSILSKKLEPSDGTTPLFKGPEGTQVS	132
Bh_fem/1-400	KSTEN ..RQR .TVTSNTSNTSILSKKLEPSDGTTPLFKGPEGTQVS	132
Bl_fem/1-416	KSTEN ..RHR .TVTSNTSNTSILSKKLEPSDGTTPLFKGPEGTQVS	132
Mc_fem/1-410	KSPES ..RHR .ITTSSTSKTAILSKKLEPSDGTTPLFKGPEGTQVS	133
Eh_fem/1-407	KSPEN ..RHRRT AVSNTCKTSVLSEKVESDGTTPLFKGPEGTQVS	122

Am_fem/1-400	ATELRKIKVDIHRVLP GGKPT TTSD ELK RDIINPEDV ML KRR T GEG	166
Ac_fem/1-404	ATELRKIKVDIHRVLP GGKPT TTTDEL K RDIINPEDV ML KRR T GEG	170
Ad_fem/1-404	ATELRKIKVDIHRVLP GGKPT TTTDEL K RDIINPEDV ML KRR T GEG	171
Af_fem/1-401	ATELRKIKVDIHRIL P GGKTTTTDEL K RDIINPEDV ML KRR T GEG	168
Bt_fem/1-416	AAELRRIKVDIHRNV PGK .. ATTDDL QRDIINPEDV LV KRGE GEG	175
Bi_fem/1-416	AAELRRIKVDIHRNV PGK .. ATTDDL QRDIINPEDV LV KRRE GEG	175
Bh_fem/1-400	AAELRRIKVDIHRNV PGK .. ATTDDL QRDIINPEDV LV KRRE GEG	175
Bl_fem/1-416	AAELRRIKVDIHRNV PGK .. ATTDDL QRDIINPEDA LV KRRE GEG	175
Mc_fem/1-410	IVELRSIKVDIHRNV PGK .. TTTDEL QRAIVNPEDV QVM RRE GEG	176
Eh_fem/1-407	ATELRRIKVDIHRN IP GGK.. TTTGEL QRDIVNPEDV V LKRR T GEG	165

Am_fem/1-400	SKPIFEREEIK NILN KTNEITEHRTVLAVNIEKSEN ET KTCK KYA	211
Ac_fem/1-404	SKPIFEREEIK NVLT KTNEITEHRTVLAVNIEKSEN ET KTCK KYA	215
Ad_fem/1-404	SKPIFEREEIK NVLS KTNEITEHRTVLAVNIEK PEN ETKTCK KYA	216
Af_fem/1-401	SKPIFEREEIK NVLT KTNEITEHRTVLAVNIEKSEN ET KTCK KYA	213
Bt_fem/1-416	SKPIFEREEIKGTT VK ADEIVERRTVVVLNNENLES SKSVN KKYT	220
Bi_fem/1-416	SKPIFEREEIKGTT VKT DEIEERTVVVLNNENLES SKSMN KKYT	220
Bh_fem/1-400	SKPIFEREEIKGTAD KT DEIEERTVVVLNNENLEN SKSVN KKYT	220
Bl_fem/1-416	SKPIFEREEIKGTT VKT DEIVERRTAVVLNNENLES SKSVN K NYT	220
Mc_fem/1-410	SKPIFEREEIKSAT VKT KEIVEHRTVVAINNNENLEN SKSVN KKYT	221
Eh_fem/1-407	SKPIFEREEIKGAT ..KTE EIEEHRTVVAINNNENLEN SKSVN K ICKK FA	209

ANHANG

Am_fem/1-400	ISSNSLRSRSRSFQRTSSCHSRYEDSRHEDRNSYRNDGERSCSR	256
Ac_fem/1-404	ISSNSLRSRSRSFQRTSSCHSRYEDSRHEDRNSYRNDRERSCSR	260
Ad_fem/1-404	ISSNSLRSRSRSFQRTTSSCHSRYEDSRHENRNSYRNDRERSCSR	261
Af_fem/1-401	ISSNSLRSRSRSFQRTSSCHSRYEDSRHDDRNSYRNDRERSCSR	258
Bt_fem/1-416	TSS...RNRSPSPRHRSSRHARYEDSKYDNRDSYRSDRGRNH	262
Bi_fem/1-416	TSS...RNRSPSPRHRSSRHARFEDSKYDNRDSYRSDRGRNH	262
Bh_fem/1-400	TSS...RNRSPSPRHRSSRHAR.....GRNH	246
Bl_fem/1-416	TSS...RNRSPSPRHRSSRHARYEDSKYDNRGYSYRSDRGRNH	262
Mc_fem/1-410	.TS...RSQSCSPRRRLSHYTRYEDSKRDGRNNYRNDRERDRSR	262
Eh_fem/1-407	TSS.PLRSRSRSPRRTASHHSRYDNSKHND.SYRNDRERDRSR	252

Am_fem/1-400	RSREYKKKDRRYDQLHNVEEKHLRERTSRRRYSRSREREQ.....	296
Ac_fem/1-404	RSREYKKKDRRYDQLHNVEEKHLRERTSRRRYSRSREREQ.....	300
Ad_fem/1-404	RSREYKKKDRRYDQLHNVEEKHLRERTNRRRYSRSREREQ.....	301
Af_fem/1-401	RSREYKK.NRRYEPLHNVEEKHLRERTSRRRYSRSREREQ.....	297
Bt_fem/1-416	RSREYKETVKRHSQNLNTAEELRLERKGRDHRSREREQRAWVR	307
Bi_fem/1-416	RSREYKETVRRHSQNLNTAEELRLERKGRDHRSREREQRAWVR	307
Bh_fem/1-400	RSREYKETVRRHSQNLNTAEELRLERKGRDHRSREREQRAWVR	291
Bl_fem/1-416	RSREYKETVSRHSQNLNTAEELRLERKGRDHRSREREQRAWVR	307
Mc_fem/1-410	RSREYKEKDRRHSQNLHTTRE....ERKGRDHRSREREQRTWSR	303
Eh_fem/1-407	KKEYQEKDRRHGESRDIIEEDHLRERKSHRDHSHSREREERKWD	297

Am_fem/1-400	KSYKNEREYREYRETSRERSRDRRERGRSREHRIIPSHYIE	337
Ac_fem/1-404	KSYKNEREYREYRETSRERSRDRRERERSREHRIIPSHYIE	341
Ad_fem/1-404	KSYKNEREYREYRETSRERSRDRRERERSREHRIIPS.YIE	341
Af_fem/1-401	KSYKNEREYREYRETSRERSRDRRERERSRERRIIPSHYIE	338
Bt_fem/1-416	DSNHRSSKDESYR...ERSRERSRERRERDRSREHRIPPSHYIE	349
Bi_fem/1-416	DSHHRSSKDESYR...ERSRERSRERRERDRSREHRIPPSHYIE	349
Bh_fem/1-400	DSHHRSSKDESYR...ERSRERSRERRERDRSREHRIPPSHYIE	333
Bl_fem/1-416	DSHHRSSKDESYR...ERSRERSRERRERDRSRVHRIPPSHYIE	349
Mc_fem/1-410	ESHKSFKDESHR...ERSRERSRERSERDRSREHRTSSHYIE	345
Eh_fem/1-407	DSSHHRFKDERFYREYQERSRGRSRERRDRDRSRERGLPPSHFIE	342

Am_fem/1-400	QIPVPVYYGNFPPRPIMVRPWVPMRGQVP...GSRH.IGPLTPFP	378
Ac_fem/1-404	QIPVPVYYGNFPPRPITVRPWVPMRGQVP...GSRH.IGPLTPFP	382
Ad_fem/1-404	QIPVPVYYGNFPPRPIMVRPWVPMRGQVP...GSRH.IGPLTPFP	382
Af_fem/1-401	QIPVPVYYGNFPPRPIMVRPWVPMRGQVP...GSRH.IGPLTPFP	379
Bt_fem/1-416	QIPVPVYYGDFPPRPIMVGPLVPIRGQVPPLGGTRHMMGPLRPF	394
Bi_fem/1-416	QIPVPVYYGDFPPRPIMVGPLVPIRGQVPPLGGTRHMMGPLRPF	394
Bh_fem/1-400	QIPVPVYYGDFPPRPIMVGLVPIRGQVPPLGGTRHMMGPLRPF	378
Bl_fem/1-416	QIPVPVYYGDFPPRPIMVGPSVPIRGQVPPLGGTRHMMGPLRPF	394
Mc_fem/1-410	QIPVPFYYGDFPAPRIMVGPLVPIRGQVP.L.GPRHMMGPLRPF	388
Eh_fem/1-407	QIPVPFYYGNFPPRPIMVGPLVPIRGQVP.L.SSRHMMGSLRPF	385

Am_fem/1-400	PRFIPPDYRLRPPPNPRFGPM	400
Ac_fem/1-404	PRFIPPDYRLRPPPNPRFGPM	404
Ad_fem/1-404	PRFIPPDYRLRPPPNPRFGPM	404
Af_fem/1-401	PRFIPPDYRLRPPPNPRFGPM	401
Bt_fem/1-416	PRFIPPDYRLRPPPNPRFRPM	416
Bi_fem/1-416	PRFIPPDYRLRPPPNPRFRPM	416
Bh_fem/1-400	PRFIPPDYRLRPPPNPRFRPM	400
Bl_fem/1-416	PRFIPPDYRLRPPPNPRFRPM	416
Mc_fem/1-410	PRFIPPDYRLRPPPNPRFGPM	410
Eh_fem/1-407	PRFIPPDYRLRPPPNPRFGPM	407

6.4 Sequenzen des Gens *tof* aus *Bombus terrestris* und *B. impatiens*

6.4.1 Sequenz der mRNA des *twin of feminizer* (*tof*) aus *B. terrestris*

Hier ist die mRNA-Sequenz des Gens *tof* aus *Bombus terrestris* in 5'-3'-Richtung gezeigt. Das Start- und das Stoppkodon sind in dunkelgrau hinterlegt. Die 5'- und 3'-UTR sind in mittelgrau gekennzeichnet, das alternative Exon (siehe 3.4.1) ist in hellgrau markiert.

```

1  TTTTTCGTAT ATTGCTAGTA ATGTAAGCAT TTATTCGTCA AATTTTTCGCA GTTTATGAGC
61 TATTATATTA GTGTGTGTGA AATTTATTTT TTCGCGTTTA ACACGATAAC TTCACTACTC
121 ACATAATTTT TATCATTTTA TTTGATGTAC TTATCTTCTA TTTTGGAAAT TTTTCTGTGT
181 AATAACTAGA TATTACTTTG TATATTGTTT GGATCTAACG TACATATCTG AATCTTAAGT
241 TACCTTACCT TAAAAAAGAC TGCAGATTAG CCTAATTGCA TATGAAACTT TTCTTCTTTG
301 GGAAATGAAA CGAAACGTGT CGAATCATT CTTATCGTGAT GAAAAATTAA GATCATCACG
361 CAGAGAAAAT TATGAGACTG GTTTACGAAG AACAGACGAG AGAAGATTAC GACGTAAGCG
421 TGAAAAATGG ATTATAGAAC AACAAATGCT ACGAGAACAT GAAAGGTTGA AGGAAAAGAA
481 GATCTTAGAA TATGAAATAC GCCGTGCTCT CGAGAAAAGG TTACCACTGC CCATGGGTAG
541 ATTTTCACAT CACAGTAGCA ATAAGAATAA AAGTAAATCC CCGGAAAGTC AACAGAGAAC
601 TGTTGCTACA TCAAATACAT CTAATACATC TATATTATCT AAAAAATTAG AGCCATCAGA
661 TGGTACAACA CCATTATTTA AAGGACTAGA AGGCATACAA ATTAATGCAG CAGAACTACG
721 AAGAATCAAA GTATGTATT CACAGAAATTC ATTCACAGAA GATACTCAGA GAAATGTACC
781 AGACGAAGCA ACAACTGATG ATCTACAACG AGATATTATC AATTTTGAAG ATGTTTTGGT
841 CAAAAGAAGA GAGGAAGCGG TCAGTTTATT TACATATCAA CCTTAAGATT TATAAGGAGA
901 GGGATCCAAA TCTATATTTG AGAGAGAAGA AATAAAAAGC GCAAGAGTTA AAACAGAAGA
961 AATTGTGGAA CATCGTACAG TTGTTGCTAT AAATAATGAA AATTTGGAAA ACAAGCCGGA
1021 AGCAAATAAA AAATATACTA CTTCTTCAAG GAGTCGAAGT CCTCGATATA GATCATCCCA
1081 TCATACCAGG TATAAAGATT CAAAACATAA TAACAGGGAT AGTTATAGAA ATGATAGAGA
1141 CAGAGATCAT AGTAGAGGTG GAAGTAGAGA GTATAAAGAA AAAGTCAGGC CACATAGTCA
1201 ATTGCACACT GCAGAGGAAA AGCATCTGCG AAAAAGAAAA AGTCATAGAG ATAATTCACA
1261 TTCAAGGGAA GGAGAACAAA GAGCATGGGC CAAAGATTGT CACAACAGCA ACAAATCATC
1321 TAAACACGAA AGACCTTATC GAGAAAGATC GAGAGAGCGA TTCCGAGAAA GAAGAGAAAG
1381 GGATAAGTCT GGAGAACGTA CACTGCCTTC ATCACATTAT ATTGAACCCA TTCCTGTTCC
1441 TGTTTATTAT GGAGGTTTTT CTCCGAGACC AATAATGGTT GGACATTGTT TTCCAATTCC

```

1501 AAGGCAAGTA CCACCTTTAG GAGGAATTAG GCATATGATG AGTCCATTAA GACCATTTC
 1561 ACCACGATTT ATTACGCCAG ATATATATAG ATTTTGTTC CTTCCAAATC CTAATTTAGG
 1621 CCAATGTATT AATGGAGGAT TTATGATGAA TGGAAAATTG CCGCATTTAA TATCAGACAT
 1681 TTTACAAAAA TTAATAATTC TTCGAAGTAT CAGTGTGAAT TATCTTTAAT TCAAATTATC
 1741 TTTAATTCAA AATATTGTTA TAAATGTATG AGATATATGT TTTATATTTT TTTATCAGCC
 1801 ATAAGAATTT TTTAAAAATC GTGAAGGCAA AGAAATCTG TTTTGGACCT TTGATTTTTT
 1861 AAATACCGAT TATAATGTAA TATGTGCATT TTTAAGCATA TTTAATATTA TTTCCAAAAA
 1921 CTTTTAAAAA TTTTTTTATA TTTGATGTAT TTGTATA

Im Folgenden ist die Aminosäuresequenz von *Bt-Tof* ohne das alternative Exon dargestellt, in welchem sich Stoppkodons befinden, die zu einem verfrühten Translationsabbruch führen.

MKRNVSNSYRDEKLRRSSRRENYETGLRRTDERRRLRRKREKWIIEQQMLREHERLKEKKI
 LEYEIRRALEKGLPLPMGRFSHHSSNKNKSKSPESQRTVATSNTSNTSILSKKLEPSDG
 TTPLFKGLEGIQINAAELRRIKVCIHNSFTEDTQRNVPDEATTDDLQRDIINFEDVLVK
 RREGGSKSIFEREEIKSARVKTEEIVEHRTVVAINNENLENKPEANKKYTTSSRSRSPR
 YRSSHHTRYKDSKHNNRDSYRNDRRDRHSGGSREYKEKVRPHSQLHTAEEKHLRKRKSH
 RDNSHSREGEQRAWAKDCHNSNKSSKHERPYRERSRERFRERRERDKSGERTLPSSHYIE
 PIPVPVYYGGFPPRPIMVGHLPVIRRVPLGGIRHMSPLRPFPPRFIQPDIYRFCSLP
 NPNLGQCINGGFMNGKLP

6.4.2 Sequenz des Gens *tof* aus *Bombus impatiens*

Hier ist die mRNA-Sequenz des Gens *twin of feminizer (tof)* aus *B. impatiens* in 5'-3'-Richtung gezeigt. Das Start- und das Stoppkodon sind in dunkelgrau hinterlegt. Die 5'- und 3'-UTR sind in mittelgrau gekennzeichnet, das alternative Exon (siehe 3.4.1) ist in hellgrau markiert.

1 TTTCTTCTAT ATTCTATATA TAGCATTAT TCGTCAAATT TTTGCAGTTT GTGAGATATT
 61 ATATTAGTGT ATGTGAAATT TATTTCTTCG CGTTTAACAC GATAACTTTA CTATTCACAT
 121 AATTTTTATA ATTTTATTTG CTGTACTTAT CTTCTATTTT GGAAATTTTT CTATGTAATA
 181 ACTAGATATT TGTATATTGT TTGGATCTAA CGTACATATT TGAATCTTAA GTTACCTTAC
 241 CTTAAAAAAG ACTGCAGACT AGCCTAATTG CATATGAAAC TTTTCTTCTT TGGGAAATGA
 301 AACGAAACGT GTCGAATCAT TCTTATCGTG ATGAAAAATT GAGATCATCA CACAGTGAAA
 361 ATTATGAGAC TGATTTACGA AGAACAGAAG AAGAAAGATT ACGGCGTAGG CGCGAATGGA
 421 TTGTAGAACA ACAAATGCTA CGAGAACATG AAAAAATTGAA GAAGAAGAAG ATCTTAGAAT
 481 ATGAAATACG TCGTGCTCTC GAGAAAGGTT TATCACTGCC CAAAGAAAGA TTTTCACATC
 541 ACAGTAGCAA TACGAGTGAA AGTAAATCCC CGGAAAGTCA GCACAGAACT GTTGCTACAT
 601 TAAGTACATC TAATACATCT ATATTATCTA AAAAAATTAGA GGCATCAGAT GGTACAACAC
 661 CTTTATTTAA AGGACCAGAA GGCATACAAG TTAAGTCTGC AGAACTACGA AGAATCAAAG
 721 TATATATTCA CAGAAATTCA TTCACAGAAG ATATTCACAG AAATGTATCA GACGAAGCAA
 781 CCGATGATCT ACAACGAGAT ATTATCAATT TTGAAGATGT TTTGGTCAA AGAAGAGAGG
 841 AAGCGGTCAG TTTATTTACA TATTAACCTT AAGATTTCTA AGGAGAGGGA TCCAAATCTA
 901 TATTTGAGAG AGAAGAAATA AAAAGTGTA CAGCTAAAAC AGAAGAAATT GTGGAACATC
 961 GTACAGTTGT TGCTATAAAT AATGAAATT TGGAAAACAA GCCAGAAGCA AATAAAGAAT

1021 ATACTACTTC TTCAAGAAGT CGAAGTCCTC CATGTAGATC GTCCCATCAT ACCAGGTATA
 1081 AAGATTCAAA ACATAATAAC AGGGATAGTT ACAGAAATGA TAGAGACAGA GATCATAGTA
 1141 GAGGTGGAAG TAGAGAGTAT AAAGAAAAAG TCAGGCTACA TAGTCAATTG CACACTGCAG
 1201 AGGAAAAACA CCTGCGAGAA AGAAAAAGTA ATAGAGATAA TTCACATTCA AAGGGAGGAG
 1261 AACAAAGAGC ATGGGTCAAA GATTCTCATA ACAGCAACAA GTCATCTAAA CACGAAAGAC
 1321 CTTATCGAGA AAAATCGAGA GAGCGATTCC GAGAAAGAAG AGAAAGAGAT AAGTCTAGAG
 1381 AACGTAGGCT ACCTCCATCA CATTATATTG AACCCATTCC TGTTCCCTGTT TATTATGGAG
 1441 GTTTTCCTCC GAGACCAATA ATGGTTGGAC CATTGGTTCC AATTCCAAGG CAAGTTCCTT
 1501 TAAAAGGAAT TAGGCATATG ATGAGTCCAT TAAGACCATA TCCACCACGA TTTATTCAGC
 1561 CAGATATATA TAGATTTTGT CCATTTCCAA ATCCTAATTT AGGCCAATGT ATTAATGGAG
 1621 GATTTATGAT GAATGGAAAA TTGCTGCATT AAATATCAGA CATTTTACAA AAATTAATAA
 1681 TTTTTCGAAG TATCATAGTG AAGTAGATTA TCTTTAATTC AAAATATTGT TATGATTGTA
 1741 TGAGATATAT ACATGTTTTA TATTTTTTTA GCAGCCATAA GAATTTTTTA TGATTTAGAT
 1801 TAAAACTGT ATGAATTAGT TCATTACAGC ACTGGCATGC ATATAAGAAT TGATGTTTCA
 1861 TCGATTTTAC GATCTCGCTA TATTATAAAC ATTATTATAT TAATACAATC GCTTTTTAAC
 1921 CGTTTGAGTC CTAAGCAGCG CGATCACGCT GTTTGAATTT TTTGTTGA

Im Folgenden ist die Aminosäuresequenz von *Bi-Tof* ohne das alternative Exon dargestellt, in welchem sich Stoppkodons befinden, die zu einem verfrühten Translationsabbruch führen.

MKRNVSNSYRDEKLRSSHSENYETDLRRTEERLRRRREWIVEQQMLREHEKLEKLLKIL
 EYEIRRALEKGLSLPKERFSHSSNTSESKSPESQHRVATLSTSNLSILSKKLEASDGT
 TPLFKGPEGIQVTAELRRIKVYIHRNSFTEDIHRNVSEATDDLQRDIINFEDVLVKRR
 EEEGSKSIFEREIISVTAKTEEIVEHRTVVAINNENLENKPEANKEYTTSSRSRSPPCR
 SSHHTRYKDSKHNNRDSYRNDRDRDHSRGSREYKEKVRLLHSLHTAEKHLRERKSNRD
 NSHSGGGEQRAWVKDSHNSNKSSKHERPYREKSRRERFRERRERDKSRERRLPPSHYIEPI
 PVPVYYGGFPPRPIMVGPLVPIPRQVPLKGIHRMMSPLRPYPFRFIQPDYRFCFPNPN
 LGQCINGGFMMNGKLLH

6.4.3 Alignment von Tof und Fem aus *B. terrestris*

Dieses Alignment zeigt die Proteinsequenz von BtFem und Tof. Die übereinstimmenden Aminosäuren sind in hellgrau unterlegt, Unterschiede in weiß. Lücken werden durch Punkte dargestellt.

fem/1-416	MKRNVSNSH	HRDERLRLSR	NGNYEAGLR	RTEERLRRR	RE.WMIE	44
tof/1-440	MKRNVSNSH	YRDEKLRSS	SRRENYET	GLRRTDERR	LLRKREKWIIE	45
fem/1-416	QQR	LRHEK	LKEKKILE	YEIRRALEKGLPL	PRGRFSRHSRSKSKS	89
tof/1-440	QQM	LRHEH	LKEKKILE	YEIRRALEKGLPL	PMGRFSHSSNKNKS	90
fem/1-416	KSTEN	RHRTV.TS	NTSNTSIL	SKKLEPSD	GTTPLFKGPEGTQVSA	133
tof/1-440	KSPES	QQRTVAT	STSNTSIL	SKKLEPSD	GTTPLFKGLEGIQINA	135

fem/1-416	AELRRIKVDIHRN.....VPGKATTDDLQRDIINPEDVLVK	169
tof/1-440	AELRRIKVCIHRRNSFTEDTQRNVPEATTDDLQRDIINFEDVLVK	180
fem/1-416	RGELEGSKPIFEREEIKGTTVKADEIVERRTVVVLNNENLESKSK	214
tof/1-440	RRELEGSKSIFEREEIKSARVKTEEIVEHRTVVAINNNENLENKPE	225
fem/1-416	VNKKYTTSSRNRS.PSPRHRSSRHARYEDSKYDNRDSYRSRGRN	258
tof/1-440	ANKKYTTSSRSR.S..PRYRSSHHTRYKDSKHNNRDSYRNDRDRD	267
fem/1-416	HSRDRSREYKETVKRHSQNLNTAEEHLRERKGRDHSSRSREREQR	303
tof/1-440	HSRGGREYKEKVRPHSQLHTAEEKHLRKRKSHRDNSHSREGEQR	312
fem/1-416	AWVRDSNH..R.SSKDERSYRERSRERSRERRERDRSREHRIPPS	345
tof/1-440	AWAKDCHNSN.KSSKHERPYRERSRERFRERRERDKSGERTLPSS	356
fem/1-416	HYIEQIPVPVYYGDFPPRPIMVGPLVPIRGQVPPLGGTRHMMGPL	390
tof/1-440	HYIEPIPVVPVYYGDFPPRPIMVGHVPIRQVPPLGGIRHMMSP	401
fem/1-416	RPFPFRFIPPDMYRLRPPLNPR.F.R.P.M.....	416
tof/1-440	RPFPFRFIQPDIIYRFCSLPNP.N.L.G.Q.CINGGFMMNGKLP	440

6.4.4 Alignment von Tof aus *B. terrestris* und *B. impatiens*

Alignment der Tof-Proteinsequenzen aus den Hummeln *B. terrestris* und *B. impatiens*. Die übereinstimmenden Aminosäuren sind in grau unterlegt, Unterschiede mit weiß. Lücken werden durch Punkte dargestellt.

Bt_tof/1-427	MKRNVSNSHSYRDEKLRSSRRENYETGLRRTDERRLRRKREKWIIE	45
Bi_tof/1-424	MKRNVSNSHSYRDEKLRSSHSYNYETDLRRTTEERLRRRR.EWIVE	44
Bt_tof/1-427	QQMLREHERLKEKKILEYEIRRALEKGLPLPMGRFSHHSSNKNKS	90
Bi_tof/1-424	QQMLREHEKLLKKKILEYEIRRALEKGLSLPKERFSHHSSNTSES	89
Bt_tof/1-427	KSPESQQRRTVATSNSTNTSILSKKLEPSDGTTPLFKGLEGIQINA	135
Bi_tof/1-424	KSPESQHRTVATLSTNTSILSKKLEASDGTTPLFKGPEGIQVTA	134
Bt_tof/1-427	AELRRIKVCIHRRNSFTEDTQRNVPEATTDDLQRDIINFEDVLVK	180
Bi_tof/1-424	AELRRIKVYIHRNSFTEDIHRNVSDA.TDDLQRDIINFEDVLVK	178
Bt_tof/1-427	RRELEGSKSIFEREEIKSARVKTEEIVEHRTVVAINNNENLENKPE	225
Bi_tof/1-424	RRELEGSKSIFEREEIKSVTAKTEEIVEHRTVVAINNNENLENKPE	223
Bt_tof/1-427	ANKKYTTSSRSRSPRYRSSHHTRYKDSKHNNRDSYRNDRDRDHSR	270
Bi_tof/1-424	ANKKYTTSSRSRSPPCRSSHHTRYKDSKHNNRDSYRNDRDRDHSR	268
Bt_tof/1-427	GGREYKEKVRPHSQLHTAEEKHLRKRKSHRDNSHSREGEQRAWA	315
Bi_tof/1-424	GGREYKEKVRPHSQLHTAEEKHLREKSNRDNNSHSGGEQRAWV	313

ANHANG

Bt_tof/1-427	KDCHNSNKSSKHERPYRERSRERFRERRERDKS	GERTLPSSHYIE	360
Bi_tof/1-424	KDCHNSNKSSKHERPYREKSRERFRERRERDKS	RERRLPSSHYIE	358
Bt_tof/1-427	PIVPVYYGGFPPRPIMVGH	LVPIRRQVPPLGGIRHMSPLRFP	405
Bi_tof/1-424	PIVPVYYGGFPPRPIMVG	PLVPIPRQVP.LKGIRHMSPLRYP	402
Bt_tof/1-427	PRFIQPDYRFC	SLPNPNLGQC	427
Bi_tof/1-424	PRFIQPDYRFC	PFPNPNLGQC	424

Literatur

- ADAMS, J., ROTHMAN, E. D., KERR, W. E., UND PAULINO, Z. L. 1977. Estimation of the Number of Sex Alleles and Queen Matings from Diploid Male Frequencies in a Population of *Apis mellifera*. *Genetics* 86:583–596.
- AGUILAR, A., ROEMER, G., DEBENHAM, S., BINNS, M., GARCELON, D., UND WAYNE, R. K. 2004. High MHC Diversity Maintained by Balancing Selection in an Otherwise Genetically Monomorphic Mammal. *Proc Natl Acad Sci U S A* 101:3490–3494.
- ALEXANDER, B. 1991. Phylogenetic Analysis of the Genus *Apis* (Hymenoptera: Apidae). *Ann. Entomol. Soc. Am.* 84:137–149.
- BARTON, N. UND BENGTTSSON, B. O. 1986. The Barrier to Genetic Exchange Between Hybridising Populations. *Heredity* 57 (Pt 3):357–376.
- BAUDRY, E., SOLIGNAC, M., GARNERY, L., GRIES, M., CORNUET, J.-M., UND KOENIGER, N. 1998. Relatedness among Honeybees (*Apis mellifera*) of a Drone Congregation. *Proc Biol Sci.* 265:2009–2014.
- BEYE, M. 2004. The Dice of Fate: the *csd* Gene and how its Allelic Composition Regulates Sexual Development in the Honey Bee, *Apis mellifera*. *Bioessays* 26:1131–1139.
- BEYE, M., HASSELMANN, M., FONDRK, M. K., PAGE, R. E., UND 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., HÄRTEL, S., HAGEN, A., HASSELMANN, M., UND OMHOLT, S. W. 2002. Specific Developmental Gene Silencing in the Honey Bee Using a Homeobox Motif. *Insect Mol Biol* 11:527–532.
- BLENCOWE, B. J., BOWMAN, J. A., MCCracken, S., UND ROSONINA, E. 1999. SR-Related Proteins and the Processing of Messenger RNA Precursors. *Biochem Cell Biol* 77:277–291.
- BOGGS, R., GREGOR, P., IDRIS, S., BELOTE, J., UND MCKEOWN, M. 1987. Regulation of Sexual Differentiation in *D. melanogaster* via Alternative Splicing of RNA from the *transformer* Gene. *Cell* 50:739–747.
- BULL, J. 1983. Evolution of Sex Determining Mechanisms. Benjamin / Cummings Publishing Company, Menlo Park, CA.

- BULL, J. J. 1980. Sex Determination in Reptiles, volume 55. The University of Chicago Press.
- CASTRIC, V., BECHSGAARD, J., SCHIERUP, M. H., UND VEKEMANS, X. 2008. Repeated Adaptive Introgression at a Gene Under Multiallelic Balancing Selection. *PLoS Genet* 4:e1000168.
- CHARLESWORTH, B. UND CHARLESWORTH, D. 2010. Elements of Evolutionary Genetics. Roberts & Company, USA.
- CHARLESWORTH, B., NORDBORG, M., UND CHARLESWORTH, D. 1997. The Effects of Local Selection, Balanced Polymorphism and Background Selection on Equilibrium Patterns of Genetic Diversity in Subdivided Populations. *Genet Res* 70:155–174.
- CHARLESWORTH, D. 2006. Balancing Selection and its Effects on Sequences in Nearby Genome Regions. *PLoS Genet* 2:e64.
- CLINE, T. W. UND MEYER, B. J. 1996. Vive la Différence: Males vs Females in Flies vs Worms. *Annu. Rev. Genet.* 30:637–702.
- COOK, J. M. 1993. Sex Determination in the Hymenoptera: a Review of Models and Evidence. *Heredity* 71:421–435.
- COOK, J. M. UND CROZIER, R. H. 1995. Sex Determination and Population Biology in the Hymenoptera. *Trends Ecol Evol* 10:281–286.
- DUCHATEAU, M., HOSHIBA, H., UND VELTHUIS, H. 1994. Diploid Males in the Bumble Bee *Bombus terrestris*. Sex Determination, Sex Alleles and Viability. *Entomol. exp. appl.* 71:263–269.
- DUCHATEAU, M. J. UND MARIËN, J. 1995. Sexual Biology of Haploid and Diploid Males in the Bumble Bee *Bombus terrestris*. *Ins. Soc.* 42:255–266.
- DZELZKALNS, V. A., NASRALLAH, J. B., UND NASRALLAH, M. E. 1992. Cell-Cell Communication in Plants: Self-Incompatibility in Flower Development. *Dev Biol* 153:70–82.
- ELLEGREN, H. 2004. Microsatellites: Simple Sequences with Complex Evolution. *Nat Rev Genet* 5:435–445.
- ERICKSON, J. W. UND 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.

- EXCOFFIER, L., LAVAL, G., UND SCHNEIDER, S. 2005. Arlequin (version 3.0): an Integrated Software Package for Population Genetics Data Analysis. *Evol Bioinform Online* 1:47–50.
- EXCOFFIER, L., SMOUSE, P. E., UND QUATTRO, J. M. 1992. Analysis of Molecular Variance Inferred from Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics* 131:479–491.
- FAY, J. C. UND WU, C. I. 2000. Hitchhiking under positive darwinian selection. *Genetics* 155:1405–1413.
- FELSENSTEIN, J. 1985. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* 39:783–791.
- FIRE, A., XU, S., MONTGOMERY, M. K., KOSTAS, S. A., DRIVER, S. E., UND MELLO, C. C. 1998. Potent and Specific Genetic Interference by Double-Stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806–811.
- FRASER, H. B., HIRSH, A. E., STEINMETZ, L. M., SCHARFE, C., UND FELDMAN, M. W. 2002. Evolutionary Rate in the Protein Interaction Network. *Science* 296:750–752.
- FRASER, H. B., WALL, D. P., UND HIRSH, A. E. 2003. A Simple Dependence Between Protein Evolution Rate and the Number of Protein-Protein Interactions. *BMC Evol Biol* 3:11.
- GADAU, J., GERLOFF, C. U., KRÜGER, N., CHAN, H., SCHMID-HEMPEL, P., WILLE, A., UND PAGE, R. E. 2001. A Linkage Analysis of Sex Determination in *Bombus terrestris* (L.) (Hymenoptera: Apidae). *Heredity* 87:234–242.
- GEMPE, T. UND BEYE, M. 2010. Function and Evolution of Sex Determination Mechanisms, Genes and Pathways in Insects. *Bioessays* 33:52–60.
- GEMPE, T., HASSELMANN, M., SCHIØTT, M., HAUSE, G., OTTE, M., UND BEYE, M. 2009. Sex Determination in Honeybees: Two Separate Mechanisms Induce and Maintain the Female Pathway. *PLoS Biol* 7(10):e1000222.
- GLÉMIN, S., GAUDE, T., GUILLEMIN, M.-L., LOURMAS, M., OLIVIERI, I., UND MIGNOT, A. 2005. Balancing Selection in the Wild: Testing Population Genetics Theory of Self-Incompatibility in the Rare Species *Brassica insularis*. *Genetics* 171:279–289.

- GRAVES, J. A. 1995. The Origin and Function of the Mammalian Y Chromosome and Y-Borne Genes - an Evolving Understanding. *Bioessays* 17:311–320.
- GRAVES, J. A. M. 2008. Weird Animal Genomes and the Evolution of Vertebrate Sex and Sex Chromosomes. *Annu Rev Genet* 42:565–586.
- HALL, T. A. 1999. BioEdit: a User-friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41:95–98.
- HANNON, G. J. 2002. RNA Interference. *Nature* 418:244–251.
- HASSELMANN, M. UND 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. UND 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., GEMPE, T., SCHIØTT, M., NUNES-SILVA, C. G., OTTE, M., UND BEYE, M. 2008a. Evidence for the Evolutionary Nascence of a Novel Sex Determination Pathway in Honeybees. *Nature* 454:519–522.
- HASSELMANN, M., LECHNER, S., SCHULTE, C., UND 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.
- HASSELMANN, M., VEKEMANS, X., PFLUGFELDER, J., KOENIGER, N., KOENIGER, G., TINGEK, S., UND BEYE, M. 2008b. 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.
- HEIMPEL, G. E. UND DE BOER, J. G. 2008. Sex Determination in the Hymenoptera. *Annu Rev Entomol* 53:209–230.
- HEINRICH, B. 1979. Der Hummelstaat. Paul List Verlag, München, Leipzig.
- HEY, J. UND WAKELEY, J. 1997. A Coalescent Estimator of the Population Recombination Rate. *Genetics* 145:833–846.
- HICKS, M. R., HOLBERTON, D. V., KOWALCZYK, C., UND WOOLFSON, D. N. 1997. Coiled-Coil Assembly by Peptides with Non-Heptad Sequence Motifs. *Fold Des* 2:149–158.

- HUDSON, R. R. UND KAPLAN, N. L. 1988. The Coalescent Process in Models with Selection and Recombination. *Genetics* 120:831–840.
- JONES, D. T., TAYLOR, W. R., UND THORNTON, J. M. 1992. The Rapid Generation of Mutation Data Matrices from Protein Sequences. *Comput Appl Biosci* 8:275–282.
- JUKES, T. UND CANTOR, C. 1969. Evolution of Protein Molecules. Mammalian Protein Metabolism.
- KATOH, K., ICHI KUMA, K., TOH, H., UND MIYATA, T. 2005. MAFFT Version 5: Improvement in Accuracy of Multiple Sequence Alignment. *Nucleic Acids Res* 33:511–518.
- KOWYAMA, Y., TAKAHASI, H., MURAOKA, K., TANI, T., HARA, K., UND SHIOTANI, I. 1994. Number, Frequency & Dominance Relationships of *S*-Alleles in Diploid *Ipomoea trifida*. *Heredity* 73:275–283.
- KRUGLYAK, S., DURRETT, R. T., SCHUG, M. D., UND AQUADRO, C. F. 1998. Equilibrium Distributions of Microsatellite Repeat Length Resulting from a Balance Between Slippage Events and Point Mutations. *Proc Natl Acad Sci U S A* 95:10774–10778.
- KRUSKAL, J. B. 1983. An Overview of Sequence Comparison In D. Sankoff and J. B. Kruskal, (ed.), Time Warps, String Edits and Macromolecules: the Theory and Practice of Sequence Comparison. Addison Wesley.
- KUMAR, S., NEI, M., DUDLEY, J., UND TAMURA, K. 2008. MEGA: a Biologist-Centric Software for Evolutionary Analysis of DNA and Protein Sequences. *Brief Bioinform* 9:299–306.
- LAIDLAW, H. H., GOMES, F. P., UND KERR, W. E. 1956. Estimation of the Number of Lethal Alleles in a Panmitic Population of *Apis Mellifera* L. *Genetics* 41:179–188.
- LANE, M. D. UND LAWRENCE, M. J. 1993. The Population Genetics of the Self-Incompatibility Polymorphism in *Papaver rhoeas*. VII. The Number of *S*-Alleles in the Species. *Heredity* 71:596–602.
- LAWRENCE, M. J. 2000. Population Genetics of the Homomorphic Self-incompatibility Polymorphisms in Flowering Plants. *Ann Bot* 85:221–226.
- LIBRADO, P. UND ROZAS, J. 2009. DnaSP v5: a Software for Comprehensive Analysis of DNA Polymorphism Data. *Bioinformatics* 25:1451–1452.

- LO, N., GLOAG, R. S., ANDERSON, D. L., UND OLDDROYD, B. P. 2010. A Molecular Phylogeny of the Genus *Apis* suggests that the Giant Honey Bee of the Philippines, *A breviligula* Maa, and the Plains Honey Bee of Southern India, *A indica* Fabricius, are Valid Species. *Sys. Entomol.* 35:226–233.
- LUPAS, A., DYKE, M. V., UND STOCK, J. 1991. Predicting Coiled Coils from Protein Sequences. *Science* 252:1162–1164.
- LÖYTYNOJA, A. UND GOLDMAN, N. 2008. Phylogeny-Aware Gap Placement Prevents Errors in Sequence Alignment and Evolutionary Analysis. *Science* 320:1632–1635.
- MADDISON, W. P. UND MADDISON, D. 2010. Mesquite: a Modular System for Evolutionary Analysis.
- MAKINO, T., SUZUKI, Y., UND GOJOBORI, T. 2006. Differential Evolutionary Rates of Duplicated Genes in Protein Interaction Network. *Gene* 385:57–63.
- MARTIN, D. P., WILLIAMSON, C., UND POSADA, D. 2005. RDP2: Recombination Detection and Analysis from Sequence Alignments. *Bioinformatics* 21:260–262.
- MEIXNER, M. D., ARIAS, M. C., UND SHEPPARD, W. S. 2000. Mitochondrial DNA Polymorphisms in Honey Bees Subspecies from Kenya. *Apidologie* 31:181–190.
- MICHENER, C. D. 1999. The Corbiculae of Bees. *Apidologie* 30:67–74.
- MUIRHEAD, C. A. 2001. Consequences of Population Structure on Genes under Balancing Selection. *Evolution* 55:1532–1541.
- NAGAMITSU, T. UND YAMAGISHI, H. 2009. Nest Density, Genetic Structure, and Triploid Workers in Exotic *Bombus terrestris* Populations Colonized Japan. *Apidologie* 40:429 – 440.
- NEEDLEMAN, S. B. UND WUNSCH, C. D. 1970. A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. *J Mol Biol* 48:443–453.
- NEI, M. UND GOJOBORI, T. 1986. Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions. *Mol Biol Evol* 3:418–426.
- NEI, M. UND KUMAR, S. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

- NEI, M. UND LI, W. H. 1979. Mathematical Model for Studying Genetic Variation in Terms of Restriction Endonucleases. *Proc Natl Acad Sci U S A* 76:5269–5273.
- NÖTHIGER, R. UND STEINMANN-ZWICKY, M. 1985. A Single Principle for Sex Determination in Insects. *Cold Spring Harb Symp Quant Biol* 50:615–621.
- PALMER, K. UND OLDROYD, B. 2001. Mating Frequency in *Apis florea* Revisited (Hymenoptera, Apidae). *Ins. Soc.* 48:40–43.
- PALMER, K. A. UND OLDROYD, B. 2000. Evolution of Multiple Mating in the Genus *Apis*. *Apidologie* 31:235–248.
- PANE, A., SIMONE, A. D., SACCONE, G., UND POLITO, C. 2005. Evolutionary Conservation of *Ceratitis capitata transformer* Gene Function. *Genetics* 171:615–624.
- PENN, O., PRIVMAN, E., ASHKENAZY, H., LANDAN, G., GRAUR, D., UND PUPKO, T. 2010a. GUIDANCE: a Web Server for Assessing Alignment Confidence Scores. *Nucleic Acids Res* 38:W23–W28.
- PENN, O., PRIVMAN, E., LANDAN, G., GRAUR, D., UND PUPKO, T. 2010b. An Alignment Confidence Score Capturing Robustness to Guide Tree Uncertainty. *Mol Biol Evol* 27:1759–1767.
- PETRUNKEWITSCH, A. 1901. Die Richtungskörper und ihr Schicksal im befruchteten und unbefruchteten Bienenei. *Zool. Jahrb.* 14:573–608.
- POTTS, W. K. UND WAKELAND, E. K. 1993. Evolution of MHC Genetic Diversity: a Tale of Incest, Pestilence and Sexual Preference. *Trends Genet* 9:408–412.
- RAMÍREZ, S. R., NIEH, J. C., QUENTAL, T. B., ROUBIK, D. W., IMPERATRIZ-FONSECA, V. L., UND PIERCE, N. E. 2010. A Molecular Phylogeny of the Stingless Bee Genus *Melipona* (Hymenoptera: Apidae). *Mol Phylogenet Evol* 56:519–525.
- RICHMAN, A. D., KAO, T. H., SCHAEFFER, S. W., UND UYENOYAMA, M. K. 1995. S-Allele Sequence Diversity in Natural Populations of *Solanum carolinense* (Horsenettle). *Heredity* 75 (Pt 4):405–415.
- RICHMAN, A. D., UYENOYAMA, M. K., UND KOHN, J. R. 1996. Allelic Diversity and Gene Genealogy at the Self-Incompatibility Locus in the Solanaceae. *Science* 273:1212–1216.
- RÖSELER, P. 1973. Die Anzahl der Spermien im Receptaculum seminis von Hummelköniginnen (Hym., Apoidea, Bombinae). *Apidologie* 4:267–274.

- SAGHAI-MAROOF, M. A., SOLIMAN, K. M., JORGENSEN, R. A., UND ALLARD, R. W. 1984. Ribosomal DNA Spacer-length Polymorphisms in Barley: Mendelian Inheritance, Chromosomal Location, and Population Dynamics. *Proc Natl Acad Sci U S A* 81:8014–8018.
- SAMBROOK, J., FRITSCH, E. F., UND MANIATIS, T. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, NY, 2nd ed. edition.
- SCHARTL, M. 2004. Sex Chromosome Evolution in Non-Mammalian Vertebrates. *Curr Opin Genet Dev* 14:634–641.
- SCHIERUP, M. H., CHARLESWORTH, D., UND VEKEMANS, X. 2000a. The Effect of Hitch-Hiking on Genes Linked to a Balanced Polymorphism in a Subdivided Population. *Genet Res* 76:63–73.
- SCHIERUP, M. H., VEKEMANS, X., UND CHARLESWORTH, D. 2000b. The Effect of Subdivision on Variation at Multi-Allelic Loci under Balancing Selection. *Genet Res* 76:51–62.
- SCHOPFER, C. R., NASRALLAH, M. E., UND NASRALLAH, J. B. 1999. The Male Determinant of Self-Incompatibility in *Brassica*. *Science* 286:1697–1700.
- SCHULTE, C. 2008. Molekulare Determination der Geschlechtsbestimmung der Erdhummel *Bombus terrestris*. Master's thesis, Heinrich-Heine-Universität Düsseldorf.
- SOSNOWSKI, B. A., BELOTE, J. M., UND MCKEOWN, M. 1989. Sex-Specific Alternative Splicing of RNA from the *transformer* Gene Results from Sequence-Dependent Splice Site Blockage. *Cell* 58:449–459.
- STOECKEL, S., CASTRIC, V., MARIETTE, S., UND VEKEMANS, X. 2008. Unequal Allelic Frequencies at the Self-Incompatibility Locus within Local Populations of *Prunus avium* L.: an Effect of Population Structure? *J Evol Biol* 21:889–899.
- STOLLE, E., ROHDE, M., VAUTRIN, D., SOLIGNAC, M., SCHMID-HEMPEL, P., SCHMID-HEMPEL, R., UND MORITZ, R. F. A. 2009. Novel Microsatellite DNA Loci for *Bombus terrestris* (Linnaeus, 1758). *Mol Ecol Resour* 9:1345–1352.
- STROBECK, C. 1983. Expected Linkage Disequilibrium for a Neutral Locus Linked to a Chromosomal Arrangement. *Genetics* 103:545–555.
- SUN, S., TING, C.-T., UND WU, C.-I. 2004. The Normal Function of a Speciation Gene, *Odysseus*, and its Hybrid Sterility Effect. *Science* 305:81–83.

- TAKAHATA, N. 1990. A Simple Genealogical Structure of Strongly Balanced Allelic Lines and Trans-Species Evolution of Polymorphism. *Proc Natl Acad Sci U S A* 87:2419–2423.
- TAKAHATA, N. 1993. Allelic Genealogy and Human Evolution. *Mol Biol Evol* 10:2–22.
- TAKAHATA, N., SATTA, Y., UND KLEIN, J. 1992. Polymorphism and Balancing Selection at Major Histocompatibility Complex Loci. *Genetics* 130:925–938.
- TAKASAKI, T., HATAKEYAMA, K., SUZUKI, G., WATANABE, M., ISOGAI, A., UND HINATA, K. 2000. The *S* Receptor Kinase Determines Self-Incompatibility in *Brassica stigma*. *Nature* 403:913–916.
- TAMURA, K., DUDLEY, J., NEI, M., UND KUMAR, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Mol Biol Evol* 24:1596–1599.
- THE NASONIA GENOME WORKING GROUP, N. G. W. 2010. Functional and Evolutionary Insights from the Genomes of Three Parasitoid *Nasonia* Species. *Science* 327:343–348.
- THOMPSON, J. D., HIGGINS, D. G., UND GIBSON, T. J. 1994. CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment Through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Res* 22:4673–4680.
- TING, C. T., TSAUR, S. C., UND WU, C. I. 2000. The Phylogeny of Closely Related Species as Revealed by the Genealogy of a Speciation Gene, *Odysseus*. *Proc Natl Acad Sci U S A* 97:5313–5316.
- VEKEMANS, X. UND SLATKIN, M. 1994. Gene and Allelic Genealogies at a Gametophytic Self-Incompatibility Locus. *Genetics* 137:1157–1165.
- VERHULST, E. C., BEUKEBOOM, L. W., UND VAN DE ZANDE, L. 2010a. Maternal Control of Haplodiploid Sex Determination in the Wasp *Nasonia*. *Science* 328:620–623.
- VERHULST, E. C., VAN DE ZANDE, L., UND BEUKEBOOM, L. W. 2010b. Insect Sex Determination: it all Evolves Around *transformer*. *Curr Opin Genet Dev* 20:376–383.
- WATERHOUSE, A. M., PROCTER, J. B., MARTIN, D. M. A., CLAMP, M., UND BARTON, G. J. 2009. Jalview Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinformatics* 25:1189–1191.

- WHITFIELD, C. W., BEHURA, S. K., BERLOCHER, S. H., CLARK, A. G., JOHNSTON, J. S., SHEPPARD, W. S., SMITH, D. R., SUAREZ, A. V., WEAVER, D., UND TSUTSUI, N. D. 2006. Thrice out of Africa: Ancient and Recent Expansions of the Honey Bee, *Apis mellifera*. *Science* 314:642–645.
- WHITING, P. 1933. Selective Fertilization and Sex-Determination in Hymenoptera. *Science* 78:537–538.
- WILKINS, A. S. 1995. Moving up the Hierarchy: a Hypothesis on the Evolution of a Genetic Sex Determination Pathway. *Bioessays* 17:71–77.
- WILLIAMS, R. UND WILLIAMS, W. 1947. Genetics of Red Clover (*Trifolium pratense* L. Compatibility. III. The Frequency of Incompatibility S-Alleles in Two Non-Pedigree Populations of Red Clover. *Genetics* 48:67–79.
- WOYKE, J. 1962. The Hatchability of 'Lethal' Eggs in a Two-Sex-Allele Fraternity in Honeybees. *J. Apic Res.* 1:6–13.
- WOYKE, J. 1963. What Happens to Diploid Drone Larvae in a Honey Bee Colony. *J. Apic Res* 2:73–75.
- YOKOYAMA, S. UND NEI, M. 1979. Population Dynamics of Sex-Determining Alleles in Honey Bees and Self-Incompatibility Alleles in Plants. *Genetics* 91:609–626.
- ZAYED, A. UND PACKER, L. 2005. Complementary Sex Determination Substantially Increases Extinction Proneness of Haplodiploid Populations. *Proc Natl Acad Sci U S A* 102:10742–10746.
- ZHANG, J., ROSENBERG, H. F., UND NEI, M. 1998. Positive Darwinian Selection after Gene Duplication in Primate Ribonuclease Genes. *Proc Natl Acad Sci U S A* 95:3708–3713.

Danksagung

Mein herzlicher Dank gilt allen, die zum Gelingen dieser Arbeit beigetragen haben:

- *Dr. Martin Hasselmann für die Bereitstellung und Finanzierung meines Promotionsprojekts sowie die gute Betreuung, insbesondere während der Schreibphase*
- *Prof. Martin Beye für die Betreuung dieser Arbeit als Doktorvater*
- *Prof. Laura Rose für die bereitwillige Übernahme des Korreferats*
- *der ganzen Arbeitsgruppe für die angenehme Arbeitsatmosphäre und die gute Zusammenarbeit*
- *Dr. Tanja Gempe, Christina Schulte, Eva Theilenberg, Jessica Klein, Inga Nissen, Marion Müller-Borg, Dr. Marianne Otte, Vasco Koch und Laura Winter fürs Korrekturlesen*
- *allen, die mir in den letzten vier Jahren mit Rat und Tat zur Seite standen*
- *Prof. Xavier Vekemans für eine konstruktive Diskussion über csd in A. florea*
- *Yvonne Zimmermann, die immer noch irgendwo eine Prachtbiene für mich aufgetrieben hat*
- *meiner Familie für die große Unterstützung während meines Studiums und der Promotion*
- *und allen lieben Freunden, die in dieser Zeit einfach für mich da waren.*

Erklärung

Die hier vorgelegte Dissertation habe ich eigenständig und ohne unerlaubte Hilfsmittel angefertigt. Die Dissertation wurde in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Düsseldorf, den 04.08.11

Sarah Tatjana Lechner