Cell, Vol. 114, 419–429, August 22, 2003, Copyright ©2003 by Cell Press

The Gene *csd* Is the Primary Signal for Sexual Development in the Honeybee and Encodes an SR-Type Protein

Martin Beye,[1,*] Martin Hasselmann,[1] M. Kim Fondrk,[2] Robert E. Page, Jr.,[2] and Stig W. Omholt[3]

[1]Martin-Luther-Universität Halle/Wittenberg
Institut für Zoologie
Biozentrum
Weinberg Weg 22
06120 Halle
Germany

[2]Department of Entomology
University of California, Davis
Davis, California 95616

[3]Agricultural University of Norway
Department of Animal Science
P.O. Box 5025
1432 AAS
Norway

Summary

Haplodiploid organisms comprise about 20% of animals. Males develop from unfertilized eggs while females are derived from fertilized eggs. The underlying mechanisms of sex determination, however, appear to be diverse and are poorly understood. We have dissected the *complementary sex determiner* (*csd*) locus in the honeybee to understand its molecular basis. In this species, *csd* acts as the primary sex-determining signal with several alleles segregating in populations. Males are hemizygous and females are heterozygous at this locus; nonreproducing diploid males occur when the locus is homozygous. We have characterized *csd* by positional cloning and repression analysis. *csd* alleles are highly variable and no transcription differences were found between sexes. These results establish *csd* as a primary signal that governs sexual development by its allelic composition. Structural similarity of *csd* with *tra* genes of Dipteran insects suggests some functional relation of what would otherwise appear to be unrelated sex-determination mechanisms.

Introduction

Sexual differentiation is a fundamental process of life. In higher organisms, it not only includes the sexual differentiation of reproductive organs, but also affects almost every aspect of an organism such as behavior, physiology, and morphology. The phenotypic differences between the sexes are specified by differential gene activities. The necessary sex-specific information is given by a primary sex-determining signal at the beginning of a regulatory cascade. Organisms, however, have evolved a variety of apparently different sex-determining systems to generate the two sexes (Schutt and Nothiger, 2000; Marin and Baker, 1998). Primary genetic signals of sex determination are dissected to their molecular and genetic components for only a few organisms having sex chromosomes, such as the fruit fly *Drosophila melanogaster*, the nematode *Caenorhabditis elegans* which both use the X:A ratio (genic balance) (Cline and Meyer, 1996), and a few mammalian species that use *Sry*, a single dominant male factor on the Y chromosome (Goodfellow and Lovell-Badge, 1993). No molecular correspondence has been found for the primary signals among these phylogenetically distant groups.

A different mode of sex determination is found in haplodiploids, in which males are derived from unfertilized, haploid eggs while females develop from fertilized eggs that are diploid. This mode of reproduction was first described in the honeybee (Dzierzon, 1845), and this was more than 50 years before the discovery of sex chromosomes (Wilson, 1905). Today studies showed that about 20% of animal species are haplodiploid (Bull, 1983; Bell, 1982) comprising the ticks and mites (Acarina), white flies (Aleyroidea), some scale insects (Coccoidea, Margarodidae), thrips (Thysanoptera), bark beetles (Scolytidae), rotifera (Monogononta) and the entire insect order Hymenoptera that contains over 200,000 species. This mode of reproduction holds a specific position in sex determination because uniparental males inherit a random half of the maternal genome, while females inherit both the maternal and paternal genes rendering any chromosomal-based sex-determining system impossible (Bull, 1983). The genetic basis of sex determination in haplodiploids is poorly understood and appears to be diverse (Bell, 1982; Cook, 1993). So far no molecular primary signal has been identified or cloned in a haplodiploid organism. The best-studied genetic example of sex determination in haplodiploids is found in some hymenopteran insects (bees, wasps, ants) in which a single locus with several alleles directs sexual development (Cook, 1993), the so-called complementary sex-determining mechanism (see Figure 2A). This mechanism was proposed based on the finding of diploid males in inbred crosses of the wasp *Barcon hebetor* (Whiting, 1943). In these crosses 50% diploid males arise from fertilized eggs suggesting that the complementary composition of alleles at a single locus and not the process of fertilization is the primary signal of sex determination. Diploid males arise when the alleles at the sex-determining locus are the same (homozygous composition). Females develop when the alleles are different (heterozygous composition). Unfertilized eggs develop into males because they are hemizygous at this locus. Subsequent analysis of a mutant marker allele that was genetically linked to the sex-determining locus confirmed the finding of homozygous and heterozygous allelic composition in males and females, respectively. The diploid males, however, do not reproduce. In some species diploid males are fertile, but produce triploid, sterile offspring (Oishi et al., 1993; Cook, 1993). For the honeybee, *Apis mellifera*, the estimated numbers of sex-determining alleles segregating in populations range from 11 to 19 (Mackensen, 1955; Laidlaw et al., 1956; Adams et al., 1977).

*Correspondence: beye@zoologie.uni-halle.de

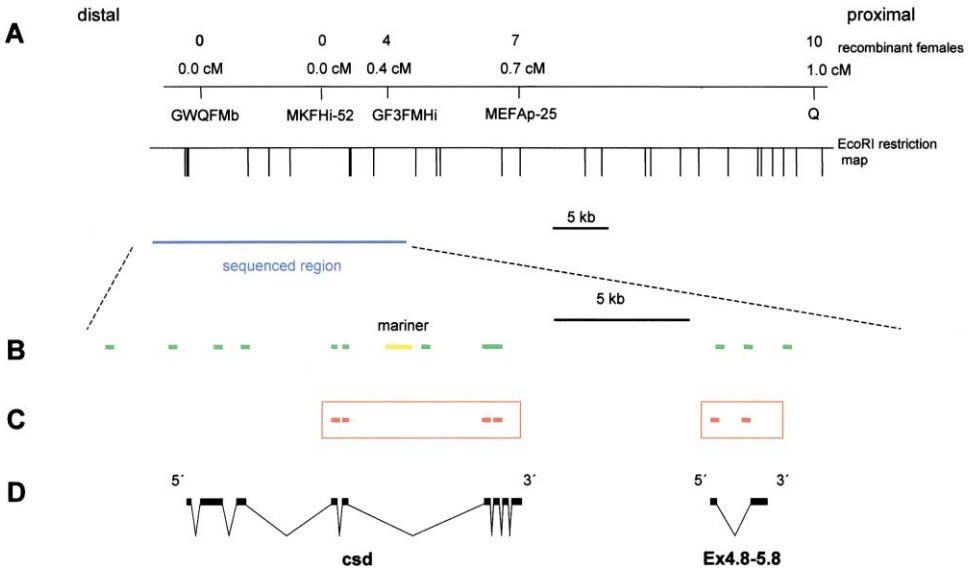


Figure 1. Genomic Region of the *complementary sex determiner*

(A) Genomic map of recombinants and EcoRI restriction map of the sex locus region covered by chromosomal walk. Marker Q was used to start the walk. A contiguous stretch of genomic DNA was assembled from a genomic partial lambda library, and in a few cases, from GenomeWalk kit (Clontech) fragments. The length and restriction pattern of clones were determined by restriction mapping and overlaps between clones were inferred by hybridization experiments and by PCR analysis using primers specific for the ends of clones. Clone ends were tested for correct position in the physical map of the sex-determining region (Beye et al., 1999), and chimeric clones were excluded from analyses. Recombinant mapping of females that are heterozygous for the sex-determining locus placed the sex-determining gene distal to marker MKFHi-52 (52–53 kb proximal to Q). No further recombination was found toward the next distal marker GWQFMb (62–64 kb proximal to Q) in 1000 females. This demonstrated that at least part of the sex-determining locus lies between the markers with an estimated precision of 5 kb +/− 2.5 kb (95% confidence interval). No further distal clone has been isolated by various approaches. A region of 24 kb covering the markers GWQFMb and MKFHi-52 was subjected to shotgun sequencing and assembled using STADEN PACKAGE software (Staden et al., 2000).

(B) Exons were predicted by various algorithms MZEF, Grail, Xpound, Genscan, and Fgene that are included in the RUMMAGE (http://gen100.imb-jena.de/rummage/ref.html) and Gene Machine (http://genemachine.nhgri.nih.gov/) annotation software using the human or *Drosophila* model organism option predicting exons (green). Blast searches detected one similarity to mariner amino acid sequences of various insects (shown in yellow).

(C) Predicted exons that produce transcripts in the early embryos. The exons belong to two transcripts as determined by RT-PCR (red boxes).

(D) Genomic organization of *csd* and *Ex4.8–5.8* transcripts that were established by various cDNA clones and 5′/3′-RACE experiments. The *csd* transcript was placed between markers GWQFMb and MKFHi-52 where no recombinants were found. Final analysis showed that MKFHi-52 starts 748 bp proximal to the 3′ end of *csd* and that GWQFMb encompass the first exon of *csd*. *csd* consists of nine exons that span a genomic sequence of 9279 bp and encompasses a mariner sequence between exons 5 and 6. The 3′ end of the second transcript *Ex4.8–5.8* is placed 1586 bp proximal of the starting sequence of marker GF3FMHi (approximately 0.4 cM from *csd*). The nucleotide sequence of *Ex4.8–5.8* is 598 bp long and consists of two exons that span a genomic sequence of 1446 bp. Note that we cannot exclude the possibility that the 5′ end of *Ex4.8–5.8* was not fully obtained because no 5′ untranslated region has been detected in the putative open reading frame (ORF).

The phenomenon of complementary sex determination raises the question of how the polymorphic signal of 19 different sex-determining alleles is transformed into the binary information of male and female development. Several molecular explanations have been proposed for the allelic mechanism although there is no biochemical or genetic evidence for any of them. All models are based on active heteromeric molecules (RNA or proteins) that are formed in the heterozygous allelic composition, thus initiating female development (Crozier, 1971; Cook, 1993).

The population dynamics of sex-determining alleles have some notable parallels to the self-incompatibility loci in plants and fungi. When an allele is shared between pollen and pistil or between fungi with the same mating type then an incompatibility response will follow (Charlesworth, 2002; Casselton, 2002). Typically, these systems are controlled by a single genetic locus having multiple allelic versions or specificities. Investigations into the molecular bases have revealed diversity in the mode of recognition and rejection in these systems.

To understand the molecular nature of the complementary sex-determining mechanism, its putative function, and allelic structure, we have positionally cloned and analyzed the *complementary sex determiner* (*csd*) in the economically important honeybee, *Apis mellifera*. Honeybees are organized in colonies with thousands of individuals. Colonies consist of a few hundred males (drones), thousands of workers, and a single queen. Female eggs develop either into sterile workers or fertile queens depending on the food they are fed by worker bees during larval development.

Results

Positional Cloning of *csd*

To obtain entry points for a chromosomal walk to *csd*, genetic markers were identified that cosegregate with

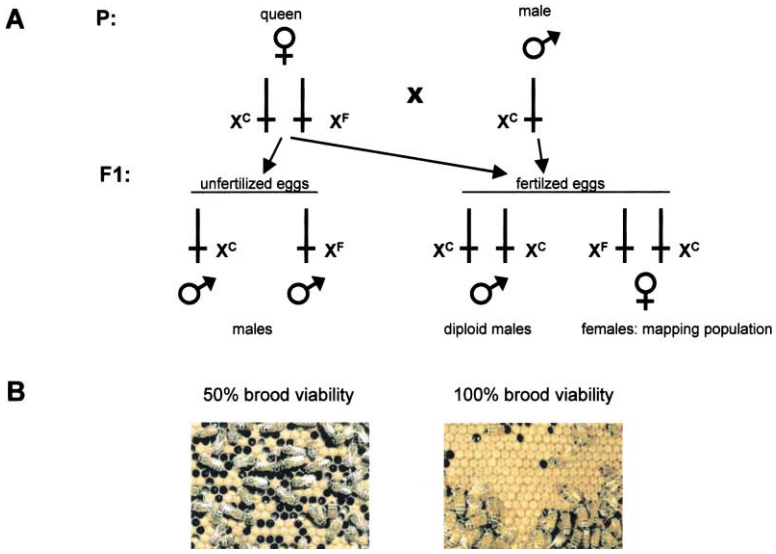


Figure 2. Colony Phenotype, Sex, and Genotypes of Animals of the Inbred Cross

(A) Sex and the genotypes of sex-determining alleles in the inbred cross. Half of the fertilized eggs are homozygous for the sex-determining locus and develop into diploid males that are eaten by worker bees (see B). Heterozygotes develop into females that served as the mapping population. Hemizygotes derived from unfertilized eggs develop into males.

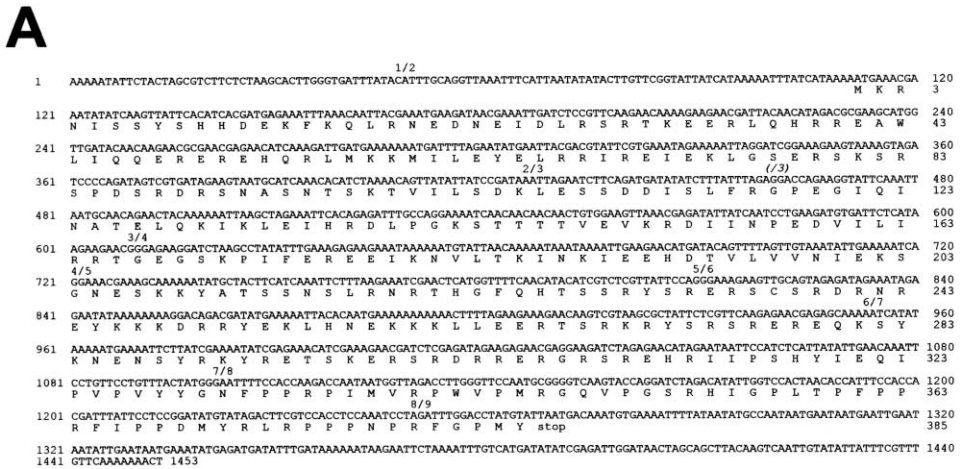
(B) Colony phenotype as the result of inbreeding. Diploid male larvae are consumed by workers shortly after they hatch, resulting in empty cells on brood combs (left). This condition is known as "shot brood." The right panel shows the brood comb from a colony that does not produce diploid males. Males derived from unfertilized eggs are laid on separate combs and cells and are not presented.

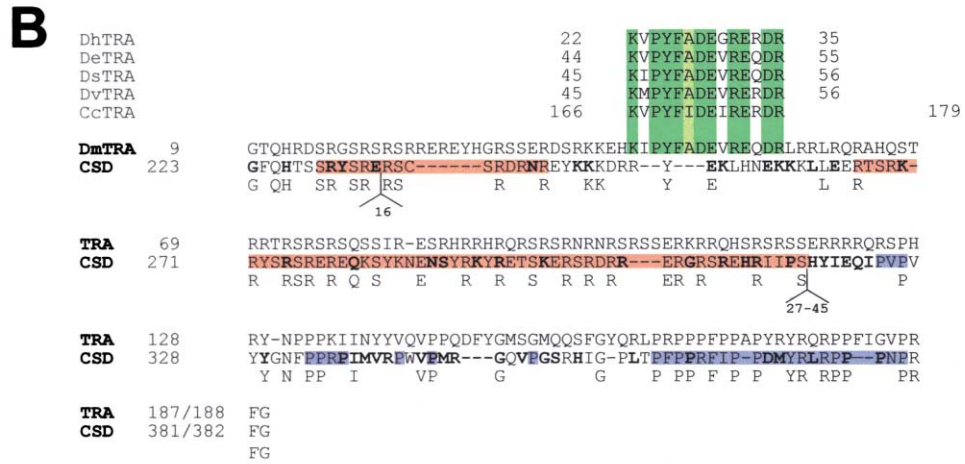
diploid male and female development. One RAPD based marker (Q, Hunt and Page, 1994) and one marker obtained by bulk segregation analysis using multilocus fingerprinting (Z, Beye et al., 1994) were shown to flank the sex-determining locus (Beye et al., 1999) at a genetic distance of 1 and 7 cM +/− 0.85 cM (95% confidence intervals), respectively. We embarked on a genomic walk to isolate the *csd* region using the Q-marker as a starting point. More than 70 kb of contiguous DNA containing the sex-determining region was isolated (Figure 1A). Progress of the chromosome walk was monitored with recombinant fine-scale mapping using polymorphisms identified as cleaved amplified polymorphic sequence (CAPS) (Hasselmann et al., 2001). Mapping was conducted on female progeny derived from a cross that yielded 50% diploid males and 50% females derived from fertilized eggs (Figures 2A and 2B). Diploid male larvae are eaten by adult female worker honeybees (Figure 2B), therefore mapping was performed only on surviving females that were heterozygous for the sex-determining locus (designated sex-determining alleles X^C/X^F). To fine scale map the sex-determining region genetic makers (CAPS) were established based on the sequence information obtained by the chromosomal walk. Analyses of 1000 meioses identified a 13 kb genomic region between marker GWQFMb and MKFHi-52 that was always heterozygous (Figure 1A). This established that at least part of the sex-determining locus that is always heterozygous in females was located between this pair of markers with an estimated precision of 5 kb +/− 2.5 kb. To identify potential exons in this region, we selected a 24 kb segment for shotgun and primer walking sequencing. Exon finding algorithms predicted several exons (Figure 1B), which were individually tested by PCR using specific primers on cDNA of 0–30 hr old embryos. Positive-tested exons were assembled to single transcripts (Figure 1C). One transcript, named *csd*, was identified in the region between markers that showed no recombinants in 1000 meioses (Figure 1D). A second transcript, named *Ex4.8–5.8*, was detected in the region with recombinants (Figure 1D). No other transcripts were detected, suggesting that this region harbors only two genes that are active in early embryogenesis. We hypothesized that the distal transcript expressed in embryos is the *complementary sex determiner csd* based on the high-resolution mapping.

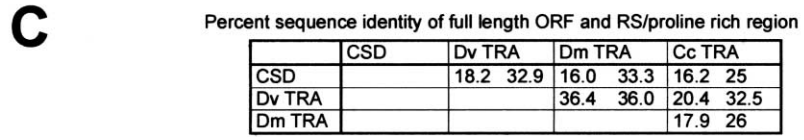
csd Encodes an Arginine (R) Serine (S)-Rich Protein

The 1453 base long *csd* sequence consists of nine exons and contains an open reading frame (ORF) of 385 amino acid residues (Figure 3A). We have found a less abundant alternative splice form of *csd* that corresponds to the deletion of the first 11 amino acids of exon 3. cDNA and genomic sequencing of *csd* alleles in males and females demonstrated that this is a rare splicing variant that is not sex- or allele-specific (Figure 3A). Sequence comparisons indicated that CSD is a member of an arginine-serine-rich (RS) protein. Highest similarity was found for its C terminus to TRA (*transformer*) protein (Boggs et al., 1987; Butler et al., 1986) (Figure 3B), to several proteins that have attributed function in mRNA processing and RNA binding (CG7971, CG7185, LD23634, GH01073) (Lasko, 2000; Mount and Salz, 2000), and to a lesser extent with well-characterized RNA binding proteins snRNP70K and Srp54 (Mancebo et al., 1990; Kennedy et al., 1998; Mount and Salz, 2000). The fact that RS-rich regions are found in functionally diverse proteins with similar protein identities (21%–33%), suggest that they are not orthologs, but distinct members of a gene family that have RS domains. CSD, like TRA and some other RS-type proteins, lacks the RNA recognition motif (RRM) in the N terminus (Graveley, 2000; Blencowe et al., 1999). Similarity to TRA extends to a proline-rich region at the C terminus (Figure 3B) with an unknown function, but proline-rich sequences are also found in other RS-type proteins (Blencowe et al., 1998). No similarities to known proteins were found for the N terminus that is composed of fewer reiterated motifs, supporting the view that *csd* codes for a RS-type protein.

We compared CSD and different TRA ortholog proteins from two *Drosophila* subgenera, and the medfly *Ceratitis* to find conserved structures or motifs among SR-type proteins that are involved in sexual regulation. Pairwise alignments of CSD to TRA sequences showed







Percent sequence identity of full length ORF and RS/proline rich region

	CSD	Dv TRA	Dm TRA	Cc TRA
CSD		18.2 32.9	16.0 33.3	16.2 25
Dv TRA			36.4 36.0	20.4 32.5
Dm TRA				17.9 26

Figure 3. Sequence of *csd* cDNA and Sequence Comparisons of CSD Protein

(A) Composite sequence and exon boundaries (e.g., 1/2) were determined by analyzing cDNAs and the genomic region. The 1453 bp consensus sequence was assembled by sequencing at least three cDNA clones and 5′/3′ RACE fragments. The assembled fragments encode a 385 amino acid protein. Several polymorphisms in the 5′ and 3′ untranslated regions (data not shown) and protein-coding regions were detected corresponding to different alleles of the same *csd* transcript as shown by subsequent analysis (Figure 5). The consensus sequences thus provide the most abundant nucleotide sequences derived from different cDNAs in these experiments. For exon 3, a rare alternative splice form of *csd* exists that corresponds to the deletion of the first 11 amino acids (shown in italics and brackets). This rare splice form was determined by cloning the same *csd* allele and was validated by genomic analysis. The reiterated amino acids arginine (R) and serine (S) are located in the C-terminal end (between amino acids 228 and 317) and show homology to various RS-protein domains in *Drosophila*. Note that the RS domain is composed of different exons. A proline-rich region is located in the C terminus between amino acids 324 and 383.

(B) Structural similarity of CSD to TRA. Example of pairwise alignments of CSD (sequence [position 223–382] is obtained from Figure 3A) and TRA protein sequences. Bold residues indicate polymorphic amino acids and arrows mark insertions (with the number of inserted amino acids), which were found in the different CSD sequences of Figure 5. Red boxes indicate the reiterated arginine (R) serine (S) domain and blue boxes highlight the proline-rich motif at the C terminus that is common to CSD and TRA sequences, indicating that similarity mainly results from reiterated RS- and proline-rich motifs found in both proteins. The green box shows an 11 amino acid motif that is highly conserved among species of *Drosophila*, *D. hydei* (Dh), *D. erecta* (De), *D. simulans* (Ds), *D. virilis* (Dv), and *D. melanogaster* (Dm), and *Ceratitis capitata* (Cc). Green, yellow, and white bars within this box indicate amino acid identity among 6/6, 5/6, and the majority of (less than 4/6) proteins, respectively. Note that this highly conserved motif is not present in CSD.

(C) Sequence identity for the entire ORF and for the RS/proline-rich region deduced from pairwise alignments. The analysis includes TRA sequences of *D. melanogaster*, *D. virilis* from the *virilis/hydei* complex, *C. ceratitis*, and CSD from Figure 3A. Sequence comparison and alignment is given for the RS domain and proline-rich region which follows the conserved TRA specific domain (green box) as defined in Figure 3B.

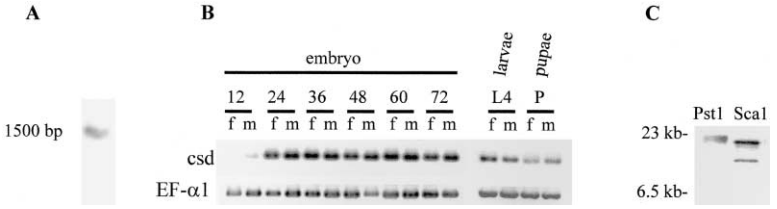


Figure 4. *csd* Transcription Profile and Single Copy Genomic Organization

(A) Northern blot analysis of total RNA from 0–30 hr old female embryos using cDNA of *csd* as a probe. A single transcript of about 1500 bases was detected.

(B) Example of transcription profile between sexes and throughout development using cDNA-PCR. After 12 hr (early blastoderm), *csd* transcription was detected in female (f) and male (m) embryos (12–72 hr), larvae (L4), and pupae (3 days before hatching from comb). The amplified 400 bp *csd* fragment encompasses exon 4–7. Other primer combinations encompassing the full ORF produced the same transcription profile throughout development and between the sexes (data not shown). A 700 bp fragment of *elongation factor 1* α (*EF-α1*) served as a positive control of transcription throughout development and between sexes. The negative control showed no amplification (data not shown).

(C) Copy number of *csd* in the honeybee genome by Southern blot hybridization analysis. To reduce the likelihood of observing allelic restriction polymorphisms in the Southern blot analysis, DNA from animals of the inbred cross were used and restriction enzymes were chosen as such that they lie outside of the coding region that showed the high nucleotide polymorphism (see Figure 5). Detected bands correspond to restriction sites predicted by *csd* genomic sequence. Sca1 was predicted to cut between exon 3 and 4, and 5 kb distal of the 5′*csd* end that corresponds to the observed 12 kb band. Both Pst1 sites were predicted to flank the *csd*-coding region (2109 bp distal of the *csd* 3′ end and 3032 bp proximal to the *csd* 5′ end) that corresponds to the 15 kb band detected in the hybridization experiment.

that all sequences consist of an RS domain that is followed by a proline-rich domain (Figure 3B and data not shown). The sequence identities were very low over the full-length ORF (16%–18%), but increased when only the RS/proline-rich regions were compared (25%–33%) (Figure 3C). Multiple alignments showed no conserved amino acid sequence motifs among sequences. TRA sequences alone show low identity values for full-length ORF (17.9%–36.5%) and the RS/proline-region (26%–36%) suggesting *tra* is a rapidly evolving gene (McAllister and McVean, 2000). Multiple alignment of TRA sequences revealed a short, but highly conserved domain of 11 amino acids in the N terminus (green box in Figure 3B; Pane et al., 2002) that is not found in the CSD protein of Figure 3A and among allelic CSD proteins (Figure 5).

csd Transcript Is the Same in Both Sexes and Is Present after Cell Formation throughout Development

csd transcription in early development was confirmed by Northern blot analysis of total RNA derived from 1–30 hr old female embryos (Figure 4A). Only one transcript of about 1.5 kb was observed. Transcription profiles showed no obvious differences (Figure 4B) between male and female 0–72 hr old embryos, pupae, and larvae. *csd* appears to be transcriptionally inactive in 0–12 hr old embryos, the syncytial stage of honeybee development. After 12 hr, the beginning of blastoderm, *csd* transcripts are continuously present throughout development. To test for splicing variants between males and females, *csd* allelic transcripts were cloned from male and female embryos and sequenced. To exclude the possibility that putative sequence differences between the sexes belong to transcripts derived from different alleles, crosses were established which obtained the same three sex alleles in males and females ("one source crosses"; see Experimental Procedures). Two allelic transcripts were identified by cloning and subsequent sequence analysis in these crosses (Figure 5, *csd* A and *csd* C). No nucleotide differences of the same alleles were found between males and females, suggesting that *csd* transcripts are the same in both sexes. We failed to isolate the predicted third allelic transcript, probably due to differences in primer access to the designated primer binding sites.

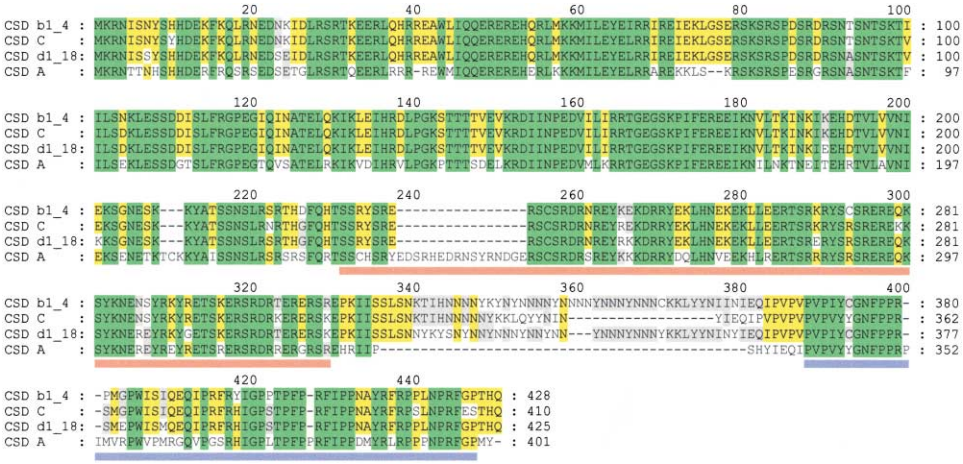


Figure 5. Alignment of Deduced Amino Acid Sequences of *csd* Alleles

Green, yellow, and gray boxes indicate amino acid identity among all, 3/4, and half of the sequenced allelic CSD polypeptides, respectively. Dashes denote gaps that are introduced to maximize homology among alleles. Alleles CSD C and CSD A were isolates from the "one source crosses" (see Experimental Procedures) and CSD b1–4 and CSD d1–18 were obtained from a natural colony. Note that the red box indicates the RS domain and the blue box marks the proline-rich region.

Table 1. *csd* Genotypes and Gonad Development in the *csd* Repression Analysis of Genetic Males and Female Embryos

		Gonad Development		
Eggs	*csd* Genotypes	*csd* dsRNA	Non-*csd* dsRNA	Control
Female	csd^{G}/csd^{H} or csd^{G}/csd^{I}	36 Testis, 3 Ovary	33 Ovary	23 Ovary
Male	csd^{H} or csd^{J}	21 Testis	ND	22 Testis

Genetic female (derived from fertilized eggs) and genetic male (derived from unfertilized eggs) embryos were injected with 900 pg dsRNA or were not injected (Control). Larvae were dissected to determine gonad development (Figure 6). Only full testis or ovary development was observed in these experiments. Exon 6–9 genomic sequences of *csd* alleles in the crosses were determined. The sequence was used to predict restriction sites that could distinguish the *csd* alleles in the crosses. Animal *csd* allelic genotypes were determined by CAPS analysis according to these restriction sites (see Experimental Procedures). The genotype analysis demonstrated that all genetic females were heterozygous for *csd* while genetic males show two hemizygous *csd* genotypes. ND, not determined.

The length polymorphic sequences were also found in the genomic sequence (genotypes of Table 1) further supporting the allelic origin of transcripts.

Analyses of *csd* Alleles Reveal Extreme Amino Acid Variation

We cloned allelic transcripts to determine how different *csd* alleles and the products they encode are related to each other (see Experimental Procedures). Support for the hypothesis that allelic transcripts are encoded from one genomic locus comes from Southern blot analysis. Hybridization of *csd* cDNA to restricted genomic DNA revealed the same bands that were expected from the genomic sequence data supporting the view that alleles originate from the same genomic locus (Figure 4C). The finding that *csd* allelic composition differed between crosses (see *csd* genotypes of Table 1) further supports a single locus origin. Nucleotide sequences of allelic transcripts were determined and the corresponding amino acid sequences were aligned (Figure 5). *csd* A and *csd* C were isolates from the "one source crosses" (see above) and *csd* b1–4 and *csd* d1–18 were obtained from embryos of a naturally mated queen. The comparisons among alleles revealed several striking features. All alleles have a single ORF of 428, 410, 425, or 401 amino acids encoding polypeptides, which have molecular weights of 51461.55 (CSD b1–4), 49305.23 (CSD i-C), 51205.01 (CSD d1–18), and 48361.97 (CSD i-A). If one accounts for conservative amino acid positions based on alignments shown by the green boxes in Figure 5, 70.3% of the N-terminal amino acids are identical (amino acid positions 1–229), while the degree of identity in the RS domain and proline-rich region is only 40.7%, including several gaps. The allelic differences in the N-terminal part are distributed evenly, but appear in short clusters separated by conserved blocks, the largest being 18 amino acids long (amino acid position 163–181). The same pattern of conserved blocks separated by short clusters of sequence differences is found for the first 1/3 of the RS and proline-rich region (amino acid position 254–322). Major differences were found in a hypervariable region located between amino acid positions 339 and 383 that consist mainly of reiterated asparagine (N) and tyrosine (Y) amino acids detected in three out of four CSD alleles. This region separates the RS domain from the proline-rich region at the C terminus. Several short, conserved, proline-rich blocks in the C terminus were found (amino acid position 399–450). Gaps were introduced in the alignment to maximize homology, the largest ones were found at the beginning of the RS-rich region (amino acid position 237–253) and another was located in the hypervariable region. Pairwise comparisons revealed that CSD A differs most from CSD C, CSD d1–18, and CSD b1–4 with only 59%–61% amino acid identities. CSD C is more similar to alleles CSD d1–18 and CSD b1–4 (86%–90%) suggesting that relationships among CSD alleles can vary. Despite the differences in amino acid composition, the predicted isoelectric points of the four alleles are extremely similar: 10.25 (CSD b1–4), 10.42 (CSD C), 10.15 (CSD d1–18), and 10.83 (CSD A). All predicted polypeptides are very basic and thus have similar physiochemical properties.

csd Repression by RNA Interference Results in Male Gonad Differentiation in Genetic Females

The differences among allelic CSD proteins are unusual for genes involved in sexual differentiation, but support the function of *csd* as the primary signal whose allelic combination governs sexual development. The function of *csd* was tested with RNA interference (RNAi) (Kennerdell and Carthew, 1998; Hunter, 1999) by injecting dsRNA in genetic males and females. To ensure that RNAi represses the variable allelic transcripts, dsRNA was generated from the conserved N-terminal regions of the alleles that differed most (Figure 5). The sources of dsRNA demonstrated more than 98% sequence identity to the *csd* alleles in the crosses (see Experimental Procedures), which is sufficient to repress transcription (Maine, 2001). Sexual differentiation was assayed by larval gonad morphology (Snodgrass, 1956; Beye et al., 1994) and *csd* genotypes of animals were determined (see Experimental Procedures). *csd* RNAi resulted in more than 92% (n = 36) of genetic females developing male gonads (Table 1 and Figure 6D). The few female phenotypes detected in the *csd* RNAi experiments (8%, Table 1 and Figure 6E) are most likely due to incomplete penetrance of RNAi, an effect that is common for RNAi (Kennerdell and Carthew, 1998; Beye et al., 2002). Effects on gonad development were very uniform and resulted in a full developmental switch from morphologically female to male gonads. No obvious intersexes were observed with both ovary and testis (Figures 6D and 6E). 100% of genetic female controls that were either not injected (n = 23, Figure 6A) or injected with non-*csd* control dsRNA (n = 33, Figure 6B) showed female gonad differentiation (Table 1). Analysis of *csd* genotypes established that treated and control larvae was heterozygous for *csd* (assigned *csd* genotypes were csd^{G}/csd^{H}

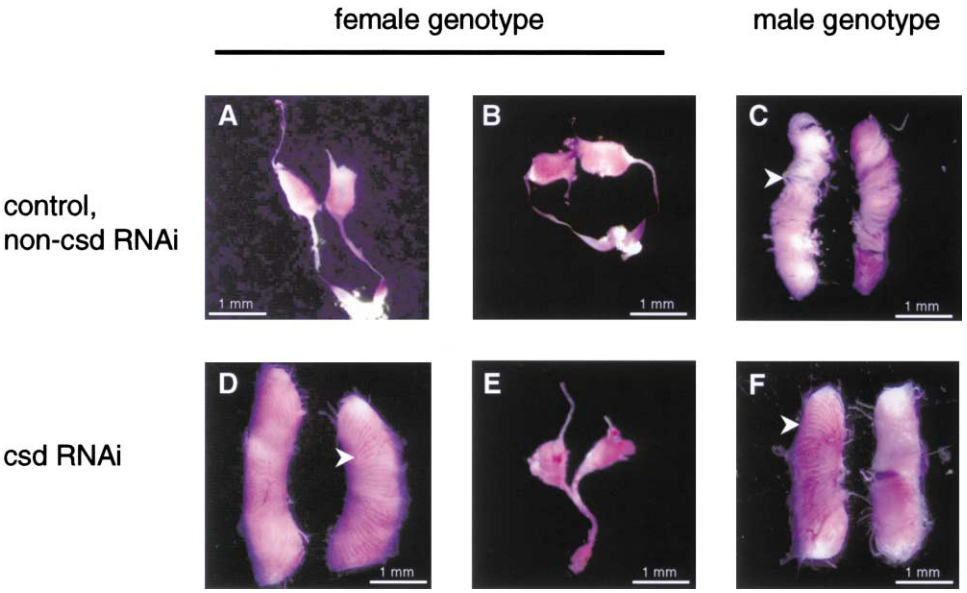


Figure 6. Gonad Development in the *csd* Repression Analysis of Genetic Male and Female Embryos

Fertilized eggs (genetic females) and unfertilized eggs (genetic males) were injected (0–4 hr after deposition) with dsRNAs or were not injected. Embryos were reared until prepupal stage to examine gonad development. For clarification, gonads were dissected from the larvae and examined independently. Slight purple color of gonads result from acetoorcein staining that enhances detection of larval anatomical structures in the large amount of larval fat body. Gonad development of noninjected genetic females (A) and genetic males (C). Genetic female embryos injected with *csd* dsRNA show full testis (D) or full ovary development (E). (B) Non-*csd* dsRNA injected genetic females. (F) Genetic (haploid) males injected with *csd* dsRNA. Note that single testioles are visible upon inspection of male gonads (arrows in C and F demonstrating normal testis development in the genetic females of D).

or csd^G/csd^I, Table 1). We found no effects of *csd* RNAi in the haploid male larvae, all developed testes (Table 1, n = 21; Figure 6F), which was also observed in the noninjected male controls (Figure 6C, n = 22). *csd* genotype analysis (see Experimental Procedures) established that the larvae had a hemizygous *csd* composition with the assigned genotypes csd^H or csd^J. Taken together these results suggest that *csd* function is required in females but not in males. *csd* encodes a functional product when *csd* transcripts are derived from different alleles (heterozygous *csd* composition) resulting in female gonad development. *csd* encodes a nonfunctional product when *csd* transcripts are derived from just one allele resulting in male gonad development.

Discussion

csd Is a Primary Signal Initiating Sexual Differentiation by Its Allelic Composition

We have identified a gene that is always heterozygous in females and encodes polypeptides with some 60% of their amino acids being different in the variable carboxy-terminal region. No differences in transcription or splicing of *csd* are detected between males and females that could serve as sex-specific information. Our functional analysis with RNAi establishes that *csd* encodes a functional product in *csd* heterozygotes resulting in female differentiation and a nonfunctional product in hemizygous males.

Comparison between the *csd* gene and other primary signals reveals differences and similarities. The X:A signal in *Drosophila* and *Caenorhabditis* (Cline and Meyer, 1996), *Sry* in mammals (Goodfellow and Lovell-Badge, 1993), *M* in *Musca* (Dubendorfer et al., 2002) and *Ceratitis* (Willhoeft and Franz, 1996) represent genetic factors that are differentially present in the two sexes. For example, when the male-determining factor *M* is present in *Musca* and *Ceratitis*, male development results; when absent, female development ensues (Hilfiker-Kleiner et al., 1993; Willhoeft and Franz, 1996). In *Drosophila*, the dose of X-linked numerator genes determines whether the master gene *Sxl* is activated (in XX females) or remains inactive (in XY males) (Cline and Meyer, 1996). Unlike the signals mentioned above, the primary genetic factor *csd* in *Apis* is present in males and females with the same alleles. The decisive difference is that females carry two different alleles whereas males have only one allele or two identical alleles.

The primary signals studied so far are only needed during a critical phase of development (Cline and Meyer, 1996; Schutt and Nothiger, 2000; Hilfiker-Kleiner et al., 1993; Goodfellow and Lovell-Badge, 1993) when the sexual pathway is established. Later and throughout development, the sexual state is maintained by a different mechanism. In *Drosophila,* once *Sxl* is activated in females, the gene maintains its own activity by a posttranscriptional positive feedback loop via alternative splicing (for review see Schutt and Nothiger, 2000). This raises the question if and how the sexual state in *Apis* is maintained during development. Transcription of *csd* begins after cellularization at the early blastoderm stage and continues throughout development (Figure 4B); no differences in the mRNAs of females and males are observed. Thus, the active or inactive allelic compo-

sition of *csd* could serve as a continuous source of information about the sexual state of the cells and may keep them on the male or female pathway. Since sexual differentiation in *Apis*, as in *Drosophila*, is a cell-autonomous process (Drescher and Rothenbuhler, 1964), the hypothesis can be tested by generating *csd* homozygous cell clones in heterozygous females at different times of development. The experiment will show whether *csd* function is required to maintain the cells on the female pathway or just to initiate it.

Similarity of *csd* to SR-Type Proteins Involved in Splicing

Previous studies have suggested that proteins with an arginine-serine (RS) rich domain are involved in protein-protein interactions and have a dominant role in constitutive and regulated pre-mRNA splicing and metabolism (Hastings and Krainer, 2001; Graveley, 2000). Proteins with RS domains fall into two distinct groups: those with an RNA recognition motif (RRM) at its N terminus that binds directly to RNA, the SR proteins (Hastings and Krainer, 2001), and those that are structurally distinct from the SR proteins, the so-called SR-related proteins (Blencowe et al., 1999). The *csd* gene constitutes a divergent member of the SR-related protein family that does not contain any RNA binding domain. CSD shows some sequence identity to various RS domains that include some conserved SR protein members involved in pre-mRNA splicing and the specialized female TRA protein. No homology, however, is found for the N terminus that is supposed to be conserved among RS protein orthologs (Blencowe et al., 1999; Dauwalder et al., 1996), supporting its function in the specialized process of complementary sex determination. SR-related proteins can generate multimeric proteins that govern development by specialized splicing (for review see Graveley, 2000; Hastings and Krainer, 2001).

A member of the SR-related protein class is TRA, which achieves female-specific splicing of *dsx* transcripts in the sex-determining cascade of *Drosophila* (Tian and Maniatis, 1993; McKeown et al., 1988). Unlike *csd* transcripts, however, *tra* transcripts are sex-specifically spliced such that a functional product is only made in females (Cline and Meyer, 1996; McKeown et al., 1987). Detailed sequence comparison of CSD and TRA orthologous polypeptides shows the lack of the prominent conserved TRA-specific amino acid motif at the beginning of the RS domain in CSD (Figure 3B; Pane et al., 2002). As a further difference, *tra* consists of 3 exons encompassing 3 kb of genomic sequence while *csd* is composed of 9 exons encompassing 9 kb of genomic sequence (Figure 1D).

In *Drosophila*, TRA, together with TRA-2, an SR protein with an RNA binding domain, binds to the pre-mRNA of *dsx* where it achieves female-specific splicing (Hoshijima et al., 1991; Tian and Maniatis, 1993). As CSD has no RNA binding domain, we propose that a factor with RNA binding function exists in the honeybee that cooperatively binds with CSD and mediates RNA splicing. This could be a TRA-2 homolog or an RS protein with a similar RNA binding function.

Active Versus Inactive Allelic Composition

A challenging problem is to determine how the polymorphic signal of various allelic combinations of *csd* is transformed into the binary switch initiating male and female development. Our results with RNAi and our fine-scale mapping show that a heterozygous allelic composition encodes a functional product that initiates female development. The hemi- or homozygous allelic composition produces a nonfunctional product that regulates male development by default. With 19 sex-determining alleles segregating in populations, 171 heterozygous combinations giving rise to female development are possible, and 19 homozygous combinations will result in male development. The full ORF deduced in four alleles studied suggests that all of the 19 alleles encode polypeptides with large differences in the variable carboxy-terminal region, although we have no further proof which of these differences have functional significance. We propose that only the combination of polypetides from different alleles yields an active heteromeric protein complex. Most putative amino acid differences that characterize the various alleles are found in the very variable C terminus, i.e., in the RS domain, in the proline-rich region, and in a hypervariable region between them. RS domains and proline-rich regions in general have protein binding abilities (Graveley, 2000; Kay et al., 2000), suggesting that differences in the amino acid sequence may result in modification of protein-protein interactions that are critical in the regulation of splicing. Indeed, analyses of other RS domains have indicated that protein interactions mediated by RS have distinct sequence requirements, which influence the splicing efficiency of SR-related proteins (Dauwalder and Mattox, 1998).

Implications for the Evolution of Sex-Determining Cascades

Sex-determining systems of the Hymenoptera and Diptera have undergone 270 million years of specialized, separate evolution (Kristensen, 1995). In the Diptera, genetic factors are present differently in the two sexes, which is apparent in sex-specific chromosomes in most species, even though the molecular nature of these factors appears to be diverse (Schutt and Nothiger, 2000; Pane et al., 2002; Dubendorfer et al., 2002). In hymenopteran insects, there are no heteromorphic sex chromosomes (Bull, 1983). Instead, males originate from unfertilized eggs, and females from fertilized eggs. The underlying genetic mechanism is the composition of alleles providing the primary signal of sex determination. However, exceptions from this mechanism apparently occur in some habitually inbreeding species (Cook, 1993).

Despite the fundamental differences in the primary mechanisms of sex determination, CSD in *Apis* shows structural similarity to TRA. TRA sequences of different *Drosophila* species show the same structure of an RS domain and proline rich motif, however, they are highly diverged in sequence (McAllister and McVean, 2000). Thus, it is conceivable that *csd*, although diverged in sequence, functionally corresponds to *tra* and sex-specifically splices *dsx* transcripts (Tian and Maniatis, 1993;

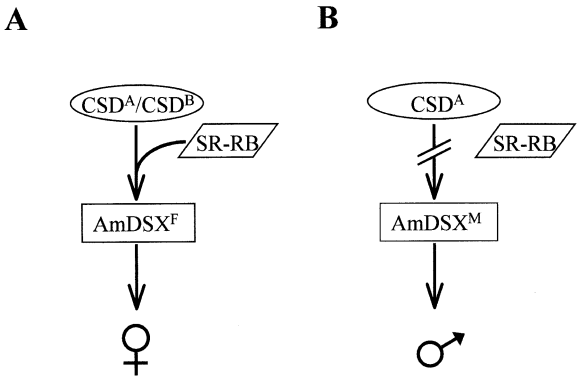


Figure 7. Model for Complementary Sex Determination in *Apis*

(A) With heterozygous *csd* composition in females the heteromeric CSD protein is functional through cooperatively binding of a SR-type protein containing an RNA recognition motif (SR-RB protein) that splices *Apis dsx* pre-mRNA. AmDSX^F protein is produced that induces female development.

(B) With hemi- or homozygous *csd* composition in males the one CSD protein present is nonfunctional. *Amdsx* is expressed by default to produce the AmDSX^M isoform, which induces male development.

Hoshijima et al., 1991) (Figure 7). As in *Drosophila* (Burtis and Baker, 1989), a homolog of *dsx* is present in the honeybee and encodes male- and female-specific proteins by sex-specific splicing (R. Becher and M.B., unpublished data). Therefore, the flow of sex-specific information to the bifunctional switch gene *dsx* via an SR-type protein may be conserved in both species. The demonstrated similarity supports the "bottom-up" hypothesis of Wilkins for the evolution of sex-determining pathways (Wilkins, 1995), which suggests that the most ancient genes operate at the bottom of the cascade, and that during evolution new regulatory elements are recruited upstream. In *Apis*, the primary female signal is most likely controlled on the posttranslational level when two CSD proteins are produced from two different alleles (Figure 7A). Male development follows when only one allelic CSD protein is present which is inactive (Figure 7B). In *Drosophila*, *tra* is regulated at the posttranscriptional level by alternative splicing and needs SXL as an upstream regulator to express its female-determining function (Inoue et al., 1990). SXL protein binds to *tra* pre-mRNA and enforces female-specific splicing thus generating a full ORF (Sosnowski et al., 1989; Belote et al., 1989). Activation of *Sxl* is controlled upstream by the polygenic X:A signal (Cline and Meyer, 1996). Taken together, these results suggest that the sex cascade in *Apis* is limited to only one level above *dsx* while in *Drosophila,* there are three levels of sex-specific instructional flow: *tra*, *Sxl* and the polygenic X:A signal. In two Dipteran species, *Musca* and *Ceratitis*, the only other insects for which sex determination mechanisms have been studied in more detail, there are at least two levels of instructional flow above *dsx* (Dubendorfer et al., 2002; Pane et al., 2002). Therefore, the *Apis* system appears to be the most ancient and the simplest system studied so far.

Female functional SR-related proteins are achieved differently in the sex-determining systems of *Apis* and *Drosophila*. CSD is active if the polypeptides are derived from two different alleles, while functional TRA protein is obtained if the *tra* transcripts are spliced by the SXL protein. In *Ceratitis,* active female TRA protein is proposed to be continuously provided via a positive feedback loop and is needed to splice its own transcript to obtain a full ORF (Pane et al., 2002). In males, the positive feedback loop is probably repressed by the primary signal of the male determiner *M*, thus providing a nonfunctional splice variant of *tra* transcripts. Taken together, these results suggest that different mechanisms have evolved to achieve the active state of female specific SR-related protein. The structural similarity among these proteins, however, indicates a common conserved downstream target gene, *dsx*, in sexual regulation. The conserved informational flow to *dsx* via an SR-related protein and the diversity of upstream signals suggest that during the course of evolution different sex-specific signals have been recruited resulting in the regulative changes of SR-related proteins found today. The observed regulative plasticity of SR-related proteins involved in sexual regulation can thus help to explain the diversity of sex-determining systems among insects.

Experimental Procedures

Crosses of *A. mellifera*

Inbred crosses were made to produce female (worker) honeybees (Figure 2) that were used for fine-scaled mapping. Virgin queens were crossed with single brother drones (males) (Laidlaw and Page, 1997; Hasselmann et al., 2001). Queens that produced 50% diploid males (homozygous sex allelic composition) from their fertilized eggs were chosen based on observing 50% larval survival (Beye et al., 1999). Surviving larvae were females and inherited the same two sex-determining alleles (designated sex alleles X^C/X^F). A second set of crosses was made ("one source crosses") to obtain the same sex alleles in haploid males and diploid female (workers) offspring. Sister virgin queens were crossed with single brother drones with a 50% probability of sharing a common sex allele. As a consequence half of these queens are expected to produce 50% diploid males. These diploid male-producing queens were excluded, leaving four queens as sources of 100% female embryos (designated sex alleles X^A/X^C, X^F/X^C, and X^A/X^F). Five additional unmated sister queens were treated with CO_2, which stimulated the laying of, unfertilized eggs (Laidlaw and Page, 1997) and were sources of male embryos (designated sex alleles X^A or X^C or X^F). cDNA used for library generation and transcription analysis was prepared from 1–30 hr old embryos, which were derived from a naturally mated queen. Genetic female and male embryos used in the RNAi analyses were derived from single-drone inseminated queens, which have only female offspring, and from additional unmated sister queens that were treated with CO_2.

Cloning, Sequencing, and Sequence Analysis

Filters with arrayed BAC (kindly provided by G. Hunt, 5× genomic coverage) and cosmid (RZPD, Germany, (Beye et al., 1998), 20× genomic coverage) clones were screened with ^{32}P random labeled marker DNA (Beye et al., 1998). A partial lambda library was generated from genomic DNA enriched for the Sma1 sex locus specific fragment mapped previously by pulsed-field gel electrophoresis (PFGE) (Beye et al., 1999). The enriched Sma1 fragment was cloned into DashII lambda arms (Stratagene). 120,000 pfu (plaque forming units) were screened using end-specific probes of clones. In some cases, extensions of individual clones were obtained by GenomeWalker Kit (Clontech). Shotgun libraries were generated from ten μg hydrosheared DNA that were cloned into M13 cloning vector with average inserts of 1.3–2 kb. cDNA library was generated following the protocol of the ZAPII vector system (Stratagene). PCR reactions were performed under standard conditions and *Pwo* DNA polymerase (PeqLab) were added if the PCR fragment was subject to sequencing. PCR fragments were cloned into T-overhang vector pGEM-T (Promega). Three or more clones from each PCR product

were subject to sequencing. Nucleotide sequences were determined on both strands by automated cycle sequencing. Single reads were assembled using the STADEN PACKAGE software (Staden et al., 2000). The nucleotide sequence of the 24 kb sex-determining region was determined by either primer walking or shotgun approaches. For shotgun approaches, a 12 times coverage was obtained. Potential exons in sequences were predicted by different methods: MZEF, Grail, Xpound, Genscan, BlastX, BlastN, Fgene that are included in the RUMMAGE (http://gen100.imb-jena.de/rummage/ref.html) and Gene Machine (http://genemachine.nhgri.nih.gov/) annotation software using the human or *Drosophila* model organism option. Sequence analysis, alignments, and database searches (FlyBase, GenBank) were performed using the following algorithms: BLAST (Altschul et al., 1990) (BLOSUM 62 scoring matrix for database searches, BLOSUM 30 matrix for TRA/CSD alignment), computational Biology WorkBench version 3.2 (http://biowb.sdsc.edu/CGI/BW.cgi), and multiple sequence alignment ClustalX 1.8 using the guided tree option (Thompson et al., 1997). Searches for conserved protein domains were performed using the integrative protein sequence analysis tool PANAL (http://mgd.ahc.umn.edu/panal/run_panal.html). Alignment of two sequences and comparisons were performed with the BLAST-2-sequences program (http://www.ncbi.nlm.nih.gov/gorf/bl2.html) or the EMBOSS sequence alignment option (http://www.ebi.ac.uk/emboss/align/).

Transcript Analyses

Standard procedures for Southern and Northern blotting on Hybond N^+ membranes (Amersham Pharmacia Biotech) were employed (Sambrook et al., 1989). Total RNA was isolated following the TRIZOL protocol (Gibco BRL) and mRNA was isolated from total RNA using Dynabeads OligoT (Dynal). OligoT-primed cDNAs were made using a RevertAid cDNA Synthesis Kit (Fermentas). 5′ and 3′ ends of *csd* were determined by RACE (rapid amplification of cDNA ends) experiments following the protocol of the RLM-RACE kit (Ambion). Transcription analysis of *csd* throughout development was semi-quantitatively compared to *A. mellifera elongation factor 1* α (*EF-*α*1*) (Walldorf and Hovemann, 1990). *csd* allelic transcripts were amplified from cDNA by PCR using primer combinations fwcsd/rev1csd or rev2csd that were cloned. Differences between *csd* clones were identified by restriction polymorphisms of amplified inserts using enzymes Apo1 and Mse1 (Hasselmann et al., 2001).

Repression Analyses by RNAi

Eggs were collected from special hives (Omholt et al., 1995) and were microinjected with 900 pg dsRNA (Beye et al., 2002). *csd* dsRNAs encompassing the conserved region from position 5 to 840 were generated from the two most different allelic cDNA sequences (Figure 5). The *csd* alleles in the crosses demonstrated more than 98% similarity to the source of dsRNA used for repression, which was verified by sequence analyses of genotypes (Table 1) and allelic transcripts (data not shown). A control dsRNA (non-*csd* dsRNA) was generated following the procedure described elsewhere (Beye et al., 2002). Eggs were incubated at 35°C in a moist chamber containing 16% H_2SO_4 (to inhibit fungus growth) that was replaced with H_2O before hatching. Hatched larvae were fed (Peng et al., 1992) until larvae developed into the prepupal stage (10–12 days after injection). Larval dissections were performed in PBS buffer and stained with acetoorcein (1 g orcein [Sigma], 45 ml acetoacid, and 55 ml dH_2O). Morphological analyses of larval gonad development were conducted as described previously (Snodgrass, 1956; Beye et al., 1994). The polymorphic genomic region of exon 6–9 (primer combination GenoRNAi_Ib/ConsCSD_ rev) were amplified, cloned, and sequenced to determine *csd* genotypes of these crosses. Ssp1 restriction sites were predicted to identify the *csd* alleles of these crosses. Genomic DNA of injected and noninjected larvae were isolated and used to amplify exon 6–9 genomic sequences that were subject to Ssp1 restriction (CAPS analysis; Hasselmann et al., 2001). Observed restriction fragments were assigned to sequenced *csd* alleles.

Supplemental Data

Primer oligonucleotide sequences are available as Supplemental Data at http://www.cell.com/cgi/content/full/114/4/419/DC1.

Acknowledgments

We thank R. Nöthiger, D. Bopp, A. Wilkins, and three anonymous reviewers for critical review of the manuscript. The authors are very grateful to G. Hunt and M. Solignac for access to BAC libraries; B. Polaczek and B. Schricker for providing eggs; A. Peterson for sequencing; T. Gempe for help with cDNA synthesis; and G. Amdam and R. Becher for assisting with bee management. This research was funded by grants from the Deutsche Forschungsgemeinschaft (2194/3) and the Kultusminister of Sachsen-Anhalt (3266A/0020L) to M.B. and grants from the U.S. National Science Foundation to REP (IBN 0090482 and IBN 0076811). The collaboration with R.E.P. was made possible by an Alexander von Humboldt Foundation Feodor-Lynen-Fellowship to M.B.

Received: May 13, 2003
Revised: July 9, 2003
Accepted: July 18, 2003
Published: August 21, 2003

References

Adams, J., Rothman, E.D., Kerr, W.E., and Paulino, Z.L. (1977). Estimation of the number of sex alleles and queen maitings from diploid male frequencies in a population of *Apis mellifera*. Genetics *86*, 583–596.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Bell, G. (1982). The Masterpiece of Nature: The Evolution and Genetics of Sexuality. (Los Angeles: University of California Press).

Belote, J.M., McKeown, M., Boggs, R.T., Ohkawa, R., and Sosnowski, B.A. (1989). Molecular genetics of *transformer*, a genetic switch controlling sexual differentiation in *Drosophila*. Dev. Genet. *10*, 143–154.

Beye, M., Moritz, R.F.A., and Epplen, C. (1994). Sex linkage in the honeybee *Apis mellifera* detected by multilocus DNA fingerprinting. Naturwissenschaften *81*, 460–462.

Beye, M., Poch, A., Burgtorf, C., Moritz, R.F.A., and Lehrach, H. (1998). A gridded genomic library of the honeybee (*Apis mellifera*): a reference library system for basic and comparative genetic studies of a Hymenopteran genome. Genomics *49*, 317–320.

Beye, M., Hunt, G.J., Page, R.E., Fondrk, M.K., Grohmann, L., and Moritz, R.F.A. (1999). Unusually high recombination rate detected in the sex locus region of the honey bee (*Apis mellifera*). Genetics *153*, 1701–1708.

Beye, M., Härtel, S., Hagen, A., Hasselmann, M., and 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., Issner, R., Nickerson, J.A., and Sharp, P.A. (1998). A coactivator of pre-mRNA splicing. Genes Dev. *12*, 996–1009.

Blencowe, B.J., Bowman, J.A., McCracken, S., and Rosonina, E. (1999). SR-related proteins and the processing of messenger RNA precursors. Biochem. Cell Biol. *77*, 277–291.

Boggs, R.T., Gregor, P., Idriss, S., Belote, J.M., and 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.J. (1983). Evolution of Sex Determining Mechanisms. (Menlo Park, CA.: Benjamin/Cummings Publishing Company).

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

Butler, B., Pirrotta, V., Irminger-Finger, I., and Nothiger, R. (1986). The sex-determining gene *tra* of *Drosophila*: molecular cloning and transformation studies. EMBO J. *5*, 3607–3613.

Casselton, L.A. (2002). Mate recognition in fungi. Heredity *88*, 142–147.

Charlesworth, D. (2002). Self-incompatibility: how to stay incompatible. Curr. Biol. *12*, R424–R426.

Cline, T.W., and Meyer, B.J. (1996). Vive la difference: 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.

Crozier, R.H. (1971). Heterozygosity and sex determination in haplodiploidy. Am. Nat. *105*, 399–412.

Dauwalder, B., and Mattox, W. (1998). Analysis of the functional specificity of RS domains *in vivo*. EMBO J. *17*, 6049–6060.

Dauwalder, B., Amaya-Manzanares, F., and Mattox, W. (1996). A human homologue of the *Drosophila* sex determination factor *transformer-2* has conserved splicing regulatory functions. Proc. Natl. Acad. Sci. USA *93*, 9004–9009.

Drescher, W., and Rothenbuhler, W.C. (1964). Sex determination in the honey bee. J. Hered. *55*, 90–96.

Dubendorfer, A., Hediger, M., Burghardt, G., and Bopp, D. (2002). *Musca domestica*, a window on the evolution of sex-determining mechanisms in insects. Int. J. Dev. Biol. *46*, 75–79.

Dzierzon, J. (1845). Gutachten über die von Herrn Direktor Stöhr im ersten und zweiten Kapitel des General-Gutachtens aufgestellten Fragen. Eichstädter Bienenzeitung *1*, 109–113, 119–121.

Goodfellow, P.N., and Lovell-Badge, R. (1993). *Sry* and sex determination in mammals. Annu. Rev. Genet. *27*, 71–92.

Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. RNA *6*, 1197–1211.

Hasselmann, M., Fondrk, M.K., Page, R.E., and Beye, M. (2001). Fine scale mapping in the sex locus region of the honey bee (*Apis mellifera*). Insect Mol. Biol. *10*, 605–608.

Hastings, M.L., and Krainer, A.R. (2001). Pre-mRNA splicing in the new millennium. Curr. Opin. Cell Biol. *13*, 302–309.

Hilfiker-Kleiner, D., Dubendorfer, A., Hilfiker, A., and Nothiger, R. (1993). Developmental analysis of two sex-determining genes, *M* and *F*, in the housefly, *Musca domestica*. Genetics *134*, 1187–1194.

Hoshijima, K., Inoue, K., Higuchi, I., Sakamoto, H., and Shimura, Y. (1991). Control of *doublesex* alternative splicing by *transformer* and *transformer-2* in *Drosophila*. Science *252*, 833–836.

Hunt, G.J., and Page, R.E. (1994). Linkage analysis of sex determination in the honey bee (*Apis mellifera*). Mol. Gen. Genet. *244*, 512–518.

Hunter, C.P. (1999). Genetics: a touch of elegance with RNAi. Curr. Biol. *9*, R440–R442.

Inoue, K., Hoshijima, K., Sakamoto, H., and Shimura, Y. (1990). Binding of the *Drosophila Sex-lethal* gene product to the alternative splice site of *transformer* primary transcript. Nature *344*, 461–463.

Kay, B.K., Williamson, M.P., and Sudol, M. (2000). The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. FASEB J. *14*, 231–241.

Kennedy, C.F., Kramer, A., and Berget, S.M. (1998). A role for SRp54 during intron bridging of small introns with pyrimidine tracts upstream of the branch point. Mol. Cell. Biol. *18*, 5425–5434.

Kennerdell, J.R., and Carthew, R.W. (1998). Use of dsRNA-mediated genetic interference to demonstrate that *frizzled* and *frizzled 2* act in the *wingless* pathway. Cell *95*, 1017–1026.

Kristensen, N.P. (1995). Forty yearś insect phylogenetic systematics. Zool. Beitr. *36*, 83–124.

Laidlaw, H.H., and Page, R.E.J. (1997). Queen Rearing and Bee Breeding. (Cheshire, CT: Wicwas Press).

Laidlaw, H.H., Gomes, F.P., and Kerr, W.E. (1956). Estimations of the number of lethal alleles in a panmictic population of *Apis mellifera*. Genetics *41*, 179–188.

Lasko, P. (2000). The *Drosophila melanogaster* genome: translation factors and RNA binding proteins. J. Cell Biol. *150*, F51–F56.

Mackensen, O. (1955). Further studies on a lethal series in the honeybee. J. Hered. *46*, 72–74.

Maine, E.M. (2001). RNAi as a tool for understanding germline development in *Caenorhabditis elegans*: uses and cautions. Dev. Biol. *239*, 177–189.

Mancebo, R., Lo, P.C., and Mount, S.M. (1990). Structure and expression of the *Drosophila melanogaster* gene for the U1 small nuclear ribonucleoprotein particle 70K protein. Mol. Cell. Biol. *10*, 2492–2502.

Marin, I., and Baker, B.S. (1998). The evolutionary dynamics of sex determination. Science *281*, 1990–1994.

McAllister, B.F., and McVean, G.A. (2000). Neutral evolution of the sex-determining gene *transformer* in *Drosophila*. Genetics *154*, 1711–1720.

McKeown, M., Belote, J.M., and Baker, B.S. (1987). A molecular analysis of *transformer*, a gene in *Drosophila melanogaster* that controls female sexual differentiation. Cell *48*, 489–499.

McKeown, M., Belote, J.M., and Boggs, R.T. (1988). Ectopic expression of the female *transformer* gene product leads to female differentiation of chromosomally male *Drosophila*. Cell *53*, 887–895.

Mount, S.M., and Salz, H.K. (2000). Pre-messenger RNA processing factors in the *Drosophila* genome. J. Cell Biol. *150*, F37–F44.

Oishi, K., Sawa, M., Hatakeyama, M., and Kageyama, Y. (1993). Genetics and biology of the sawfly, *Athalia rosae* (Hymenoptera). Genetica *88*, 119–127.

Omholt, S.W., Hagen, A., Elmholdt, O., and Rishovd, S. (1995). A laboratory hive for frequent collection honeybee eggs. Apidologie (Celle) *26*, 297–304.

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

Peng, C.Y., Mussen, E., Fong, A., Montague, M.A., and Tyler, T. (1992). Effects of chlortetracycline of honeybee worker larvae reared in vitro. J. Invertebr. Pathol. *60*, 127–133.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

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

Snodgrass, R.E. (1956). The Anatomy of the Honeybee. (Ithaca, NY: Cornell University Press).

Sosnowski, B.A., Belote, J.M., and McKeown, M. (1989). Sex-specific alternative splicing of RNA from the *transformer* gene results from sequence-dependent splice site blockage. Cell *58*, 449–459.

Staden, R., Beal, K.F., and Bonfield, J.K. (2000). The Staden package, 1998. Methods Mol. Biol. *132*, 115–130.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. *25*, 4876–4882.

Tian, M., and Maniatis, T. (1993). A splicing enhancer complex controls alternative splicing of *doublesex* pre-mRNA. Cell *74*, 105–114.

Walldorf, U., and Hovemann, B.T. (1990). *Apis mellifera* cytoplasmic *elongation factor 1 alpha* (*EF-1 alpha*) is closely related to *Drosophila melanogaster EF-1 alpha*. FEBS Lett. *267*, 245–249.

Whiting, P.W. (1943). Multiple alleles in complementary sex determination of *Habrobracon*. Genetics *28*, 365–382.

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

Willhoeft, U., and Franz, G. (1996). Identification of the sex-determining region of the *Ceratitis capitata* Y chromosome by deletion mapping. Genetics *144*, 737–745.

Wilson, E.B. (1905). The chromosomes in relation to determination of sex in insects. Science *22*, 500–502.

Accession Numbers

The accession numbers for the sequences reported in this paper are AY350615, AY350616, AY350617, AY350618, AY352276, and AY352277.