LETTERS

Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees

Martin Hasselmann[1]*, Tanja Gempe[1]*, Morten Schiøtt[1,2], Carlos Gustavo Nunes-Silva[3], Marianne Otte[1] & Martin Beye[1]

Sex determination in honeybees (*Apis mellifera*) is governed by heterozygosity at a single locus harbouring the *complementary sex determiner* (*csd*) gene[1], in contrast to the well-studied sex chromosome system of *Drosophila melanogaster*[2]. Bees heterozygous at *csd* are females, whereas homozygotes and hemizygotes (haploid individuals) are males. Although at least 15 different *csd* alleles are known among natural bee populations[3], the mechanisms linking allelic interactions to switching of the sexual development programme are still obscure. Here we report a new component of the sex-determining pathway in honeybees, encoded 12 kilobases upstream of *csd*. The gene *feminizer* (*fem*) is the ancestrally conserved progenitor gene from which *csd* arose and encodes an SR-type protein, harbouring an Arg/Ser-rich domain. Fem shares the same arrangement of Arg/Ser- and proline-rich-domain with the *Drosophila* principal sex-determining gene *transformer* (*tra*), but lacks conserved motifs except for a 30-amino-acid motif that Fem shares only with Tra of another fly, *Ceratitis capitata*[4]. Like *tra*, the *fem* transcript is alternatively spliced. The male-specific splice variant contains a premature stop codon and yields no functional product, whereas the female-specific splice variant encodes the functional protein. We show that RNA interference (RNAi)-induced knockdowns of the female-specific *fem* splice variant result in male bees, indicating that the *fem* product is required for entire female development. Furthermore, RNAi-induced knockdowns of female allelic *csd* transcripts result in the male-specific *fem* splice variant, suggesting that the *fem* gene implements the switch of developmental pathways controlled by heterozygosity at *csd*. Comparative analysis of *fem* and *csd* coding sequences from five bee species indicates a recent origin of *csd* in the honeybee lineage from the *fem* progenitor and provides evidence for positive selection at *csd* accompanied by purifying selection at *fem*. The *fem* locus in bees uncovers gene duplication and positive selection as evolutionary mechanisms underlying the origin of a novel sex determination pathway.

We have identified a second switch gene, *fem*, in the honeybee that, besides *csd*, also localizes to the sex determination locus (SDL; Fig. 1a). The SDL defines the genomic region that is always heterozygous in females[5] and thus possibly harbours extra genes involved in sex determination. We isolated a new part of the SDL (26 kilobases (kb)) by assembling sequences of a previously analysed region (21 kb)[1], fragments from shotgun cloning, contigs from the honeybee genome project[6] and site-specific amplicons. In our previous study[1] we failed to isolate more parts of the SDL because SDL sequences are AT-rich and are under-represented in our various cloning and shotgun sequencing strategies[1,6]. We now report three extra genes within the SDL. We tested all five genes located within the SDL (Fig. 1a) for

sex-determining function by RNAi knockdown experiments. Only *csd* and the new *fem* gene, located 12 kb upstream of *csd*, have sex determination function (Fig. 1b). RNAi-induced knockdowns of *fem* in females result in a developmental switch to entire male head differentiation (Fig. 1b), whereas knockdowns in males do not affect head development. Repressing the function of *csd* by RNAi results in apparently the same male-like development in females, but again does not affect head differentiation in males (bottom right panel of Fig. 1b). These findings indicate that *fem* is the second binary switch gene of the sex determination pathway that, when active, regulates the entire developmental programme of females but not that of males. The *fem* gene encodes a protein that has a carboxy-terminal Arg/Ser-rich and Pro-rich domain with a high degree of sequence identity to the Csd protein (>70% identical amino acid residues, Fig. 1c), but no similarity to other proteins in the database by applying BLAST program searches. Fem has an extra Arg/Ser-domain in its amino terminus, but lacks the hypervariable region of Csd (Fig. 1c)[3]. Thus, both genes are evidently paralogues coding for SR-type proteins, which are thought to be involved generally in the regulation of RNA splicing[7].

We characterized sex-specific transcripts of *fem* and identified male and female *fem* transcripts with the same 5′ untranslated region (UTR) sequence, but differences in their downstream exon composition (Fig. 2a). Male-specific *fem* transcripts retain a full exon 3, which contains a stop codon, so that translation terminates prematurely. In females, this part of exon 3 plus exons 4 and 5 are spliced out, leading to a complete open reading frame that translates into a protein of 403 amino acids. Consistent with the allelic mode of *csd* activation[1], we isolated *csd* transcripts that differ in their sequence composition, but not in their combination of exons (Fig. 1c). To test whether *fem* in females is regulated by the activity of the *csd* gene, we repressed *csd* in early embryogenesis by RNAi and studied *fem* transcripts in fourth instar larvae. These females have a predominant transcript of the male composition (Fig. 2b), establishing that splicing of *fem* is regulated in response to the function of *csd*. Next, we examined whether transcriptional levels of *csd* are regulated in response to *fem* function. We repressed *fem* in early embryogenesis and measured *csd* messenger RNA levels at the middle stage of embryogenesis. We observed no differences in *csd* mRNA levels between *fem* knockdown and mock short-interfering-RNA-treated (siRNA-treated) embryos ($P > 0.1$, *t*-test; Fig. 2c). Taken together, these results indicate that the binary switch of the sex determination pathway is implemented by alternative splicing of the *fem* transcript in response to heterozygosity at *csd*.

We compared our findings with the *D. melanogaster* pathway. The genes *fem* and *tra* seem to have equivalent functions in sex determination, belong to the same family of SR-type proteins, share the

[1]Department of Genetics, Heinrich Heine University Duesseldorf, Universitaetsstrasse 1, 40225 Duesseldorf, Germany. [2]Centre for Social Evolution, Department of Biology, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark. [3]Grupo de Pesquisas em Abelhas (GPA), Instituto Nacional de Pesquisas da Amazônia (INPA) Avenida André Araújo 2936, 69060-001 Manaus, AM, Brazil.
*These authors contributed equally to this work.

same arrangement of regions enriched with arginine and serine (Arg/Ser domain) and proline (Pro-rich region), but harbour no significant identity in sequence motifs (Fig. 1c). Such identify may not be expected given the rapid divergence of this sequence between different dipteran species[4,8]. The *tra* gene of *D. melanogaster* functions as a switch gene, is regulated sex-specifically by alternative splicing and is necessary for the entire development of females[9,10]. It is also part of the sex determination cascade that communicates the upstream X:A (ratio of X chromosomes to autosome sets) and *Sex lethal* (*Sxl*) signal to the downstream gene *doublesex* (*dsx*), which controls the activity of the final target genes necessary for sexual differentiation[11–13]. When we compared Fem with the orthologue of Tra from the

mediterranean fly *C. capitata*[4], in addition to the same arrangement of domains (Fig. 1c), we identified a 30-amino-acid motif in which 15 residues are identical (yellow and framed box in Fig. 1c). On the basis of equivalent function and regulation of sex determination, the same arrangement of Arg/Ser- and Pro-enriched domains and the conserved sequence motif, we conclude that *fem* and *tra* have a common evolutionary origin. These homologies establish a common ancestral pathway of sexual regulation at the level of the *tra* gene across insect orders and ~300 million years (Myr) of independent evolution.

We compared the coding sequences of the honeybee paralogues *fem* and *csd* with their orthologues from two related Asian honeybee species (*Apis cerana*, *Apis dorsata*), as well as *fem* sequences from the stingless bee (*Melipona compressipes*) and the bumble bee (*Bombus terrestris*)—both of which are members of the major sister branches of honeybees (corbiculate bees)—and the jewel wasp (*Nasonia vitripennis*), to obtain information on the evolutionary relationship and functional divergence of these genes.

Significantly, the *csd* gene is unique to the honeybee lineage. The gene duplication event can be placed on the phylogeny by comparing the gene family tree with the phylogeny of the bee species (Fig. 3a). The *fem* and *csd* sequences of honeybees form a single clade in the gene tree irrespective of whether we analyse synonymous (Fig. 3a) or amino acid substitutions (Supplementary Fig. 1). We next addressed whether an accelerated substitution rate in the stingless bee sequence offers an alternative explanation to the clustering of honeybee *fem* and *csd* sequences. The stingless bee sequence had a lower relative substitution rate than honeybees (relative rate test, Supplementary Table 1), suggesting that the *fem* and *csd* clade in the gene tree is the consequence of a duplication event within the honeybee lineage. Given some uncertainty in the current phylogenetic relationships among the major sister branches of corbiculate bees, we included

Figure 1 | Sex-determining genes within the SDL genomic region.
a, Diagram of the identified genes within the SDL genomic region[5]. Genes are orientated 5′ to 3′ according to the direction of the arrows; the names of previously analysed genes[1] are underlined. **b**, Head development of males and females treated with *fem* and *csd* siRNAs in early embryogenesis. Frontal view of female (workers, control on the left, $n = 8$) and diploid male (control on the left, $n = 18$) heads are shown. Seventy-eight per cent ($n = 17$) of females treated with *fem* siRNAs and 75% ($n = 27$) of females treated with *csd* siRNAs develop the entire head structures of males. Diploid males treated with *fem* ($n = 9$) or *csd* ($n = 19$) siRNAs have normally developed male heads. Scale bars, 1 mm. **c**, Domain diagrams of Csd, Fem and Tra proteins. Arg/Ser domains are indicated by red, the hypervariable region by green, and the common Pro-rich region by blue boxes. The sequence motif shared across insect orders (Fem and *C. capitata* Tra (Cc-Tra[4])) and its schematic location (yellow boxes) are shown. The percentage of amino acid identities of Fem and Csd domains are indicated. aa, amino acids.

Figure 2 | Structure of *fem* splice variants and the functional relationship of *fem* and *csd* genes. a, Female and male splicing diagram of the *fem* gene. Common exons are marked in white, and male-specific exons and exon extension are in grey. Translational start and stop sites as well as the poly(A) addition sites are indicated. **b**, The processing of *fem* transcripts in response to the repression of *csd* function by RNAi. Fragments corresponding to transcripts of female (~350 bp) and male (~1.6 kb) composition were amplified by RT–PCR reactions. **c**, *csd* mRNA amounts in response to repression of *fem* by RNAi in early embryogenesis. Relative transcription amounts of *csd* in *fem* knockdown embryos ($n = 6$) and mock siRNA-treated control embryos ($n = 9$). Error bars represent s.d.

the *N. vitripennis* protein sequence as a further distant out-group (~120 Myr of divergence) in our gene tree analysis (Fig. 3b) and again found a clustering of honeybee Csd and Fem proteins and no evidence for relative differences in substitution rates (relative rate test, Supplementary Table 2). In support of a recent duplication event in the honeybee lineage, we detected only a single copy of this gene family in the bumble bee (*B. terrestris*) and the jewel wasp (*N. vitripennis*). Our Southern blot hybridization of *fem* DNA to *B. terrestris* genomic DNA showed single bands in four different restrictions, supporting the presence of a single genomic locus (Supplementary Fig. 2). By searching the sequenced genome of *N. vitripennis*, we identified only a single gene with homologies to *fem*. The close evolutionary relationship of honeybee *csd* and *fem* genes thus strongly suggests that gene duplication occurred after the split of the stingless bee, bumble bee and honeybee (~70 Myr ago)[14] but before honeybee divergence (~10 Myr ago). Consequently, the *csd* gene is not the universal molecular basis of complementary sex determination in a variety of hymenopteran insects (bees, ants and wasps)[15,16], suggesting that other unknown molecular signals are the primary sex determiners in these species.

Further analysis suggests that the *csd*-based sex determination was shaped by positive selection. An excess of non-synonymous to synonymous substitutions in a branch of a phylogeny indicates that positive darwinian selection has operated in enhancing the fixation of amino acid changes. A significant excess of non-synonymous over synonymous substitutions is observed in the branch immediately after gene duplication, indicating the action of positive selection in the rise of *csd* (Fig. 3c; branch point i–ii, $P = 10^{-4}$, one-tailed Fisher's exact test). This finding is consistent whether we used different honeybee divergence scenarios or included varying numbers of *csd* alleles in the analysis. Six substitutions that are fixed in the *csd* gene are components of a coiled-coil motif which encodes protein-binding properties[17] (Fig. 3d)—a prerequisite for the function of *csd*-based sex determination[18]. In addition to branch i–ii, an excess of non-synonymous over synonymous changes is also found in the branch ii–iii ($P = 0.007$, one-tailed Fisher's exact test) leading to the western honeybee (*A. mellifera*). This latter excess is, however, significantly lower than the former ($P = 0.016$, one-tailed Fisher's exact test) suggesting that stronger positive selection operated during the early formation of the *csd* gene than during lineage-specific divergence. The *fem* gene—the progenitor of *csd*—is under purifying selection, showing an excess of synonymous to non-synonymous substitutions in branches of *fem* ($P < 0.05$, one-tailed Fisher's exact test; Fig. 3c). This low evolutionary divergence of Fem proteins is consistent with our previous result indicating an ancestral sex-determining function for *fem*.

Evidently, *fem* has retained its sex determination function whereas its recent duplicate, *csd*, evolved an allelic mode of SR-type protein activation through positive selection. The *csd*-based sex determination system controls sex-specific splicing of its progenitor transcript, thus implementing the switch of male and female pathways. These findings suggest that gene duplication of an existing major sex
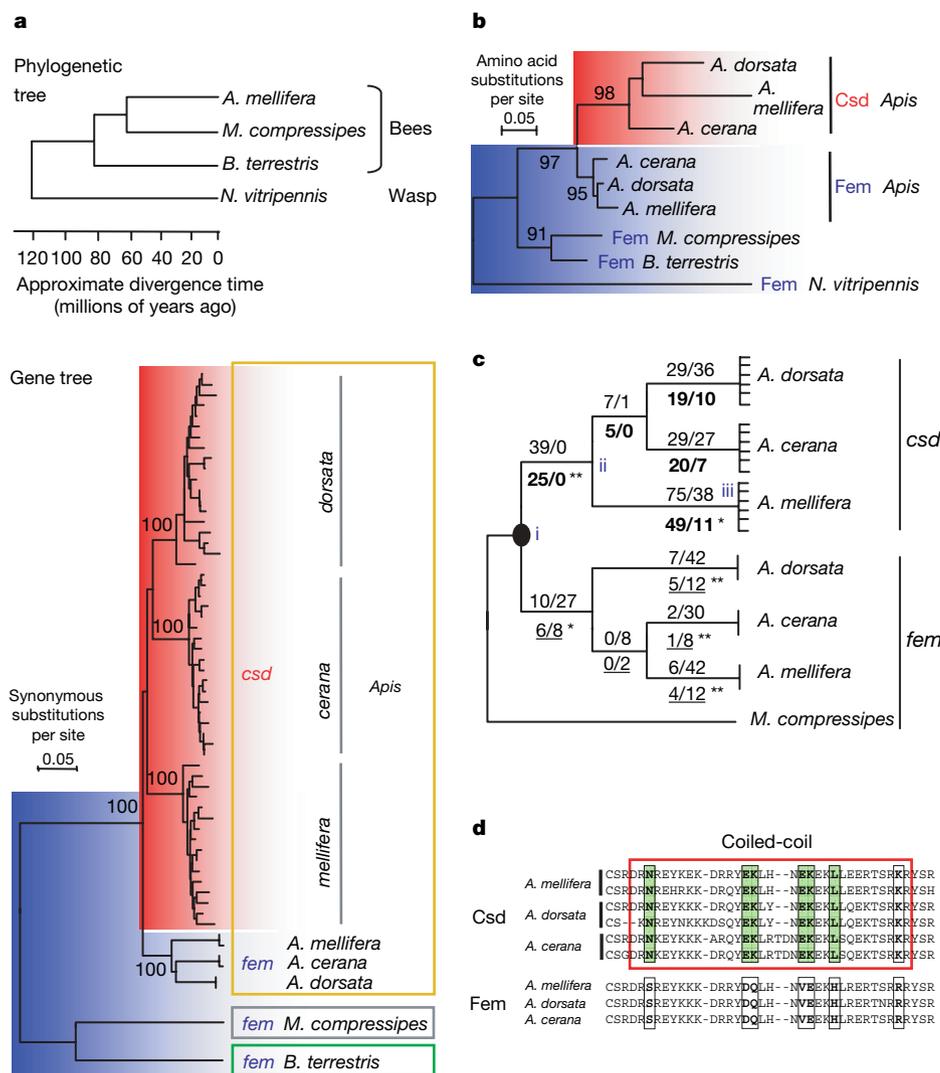
Figure 3 | Nucleotide and amino acid substitutions in the evolution of *csd* and *fem* genes. a, Comparison of the phylogenetic[14,30] (top panel) and gene (bottom panel) tree. Branches derived from *csd* and *fem* sequences are indicated by red and blue boxes, respectively. Species sources are marked by coloured frames. Numbers represent bootstrap values >80%. **b**, Gene tree of Fem and Csd protein sequences including the wasp *N. vitripennis*. Numbers denote bootstrap values >80%. **c**, Numbers of non-synonymous (N) and synonymous (S) substitutions along branches of the gene tree. Non-synonymous to synonymous substitution values per branch (a_N/a_S) and per branch and site (b_N/b_S) ×1000 are shown below and on the lines, respectively. Level of significance of Fisher's exact tests for positive (bold numbers) and for purifying (underlined numbers) selection[29]. Single asterisk, $P < 0.05$; double asterisk, $P < 0.01$. i, ii and iii (coloured in blue) define branches between nodes of the *csd* genealogy. **d**, Comparison of fixed amino acid substitutions between Csd and Fem proteins involved in coiled-coil formation. Fixed differences are marked by boxes; the ones in Csd that are components of the coiled-coil motif (red frame) are coloured in green.

determining gene, followed by positive selection in one of the duplicates, favoured the origination of a new upstream signal and, thereby, the creation of a novel sex determination pathway.

The upwards growth of this pathway is consistent with theory[19,20] predicting that new signals are co-opted upstream of a cascade during the course of evolution. Furthermore, it has been proposed that the origin of alternative sex determination signals involve a selective advantage, such as the possibility to modify sex ratios[21] or to improve the quality of signals[21,22]. Our findings provide direct evidence for a role of strong positive selection in the formation of a new sex determination system and are thus consistent with previous suggestions. We propose that in our case the reduction of recombination at the sex determination locus[5] (Fig. 1a) results in a gradual loss of the levels of adaptation of the gene[23,24], which would facilitate the evolution of alternative initial signals of complementary sex determination. This process may thus relate to the evolution of chromosomal systems in which the cessation of recombination[25,26] results in a degradation of genes[27] and the extinction of the sex chromosomes[28]. In either case, our findings provide strong support for the role of positive selection in shaping the growth of a developmental pathway.

METHODS SUMMARY

Queens producing 50% female and 50% diploid males were derived from brother–sister crosses (inbred crosses). Male-producing queens laying exclusively unfertilized, haploid eggs were obtained from non-mated, CO_2-treated queens. Female-producing queens were obtained from queens inseminated by semen of a single male. The sequence reads were assembled using Staden Package software. Potential exons and genes within the SDL were detected by gene prediction programs as described previously[1]. Transcriptionally active genes in embryogenesis were identified by reverse transcription PCR (RT–PCR) of embryonic complementary DNA. The full sequences of transcripts were obtained by 5′ and 3′ RACE experiments. siRNA (50–100 pg per embryo) and double-stranded RNA[1] were injected into embryos at the syncytial stage. Fragments corresponding to sex-specific *fem* transcripts were amplified by RT–PCR reactions from individual fourth instar larvae and resolved by agarose gel electrophoresis. Freshly hatched larvae were reared *in vitro* at 34.5 °C and saturated humidity. Relative transcriptional levels of *csd* were calculated by comparing cycle thresholds to the reference gene, *elongation factor 1-alpha*. Fifty-one *csd* ($n = 15$, *A. mellifera*; $n = 17$, *A. cerana*; $n = 19$, *A. dorsata*) and ten *fem* coding sequences ($n = 4$, *A. mellifera*; $n = 2$, *A. cerana*; $n = 2$, *A. dorsata*; $n = 1$, *M. compressipes*; $n = 1$, *B. terrestris*) were isolated by high-fidelity PCR amplifications of cDNAs derived from RNA preparations of embryos. Gene trees were obtained by the minimum evolution method by applying the Nei–Gojobori distance with Jukes–Cantor correction for nucleotide and Poisson-corrected distances for amino acid differences. The non-synonymous and synonymous substitution values for each branch of the gene tree were obtained by least-squares method. Fisher's exact test for selection was performed on the numbers of changed and unchanged non-synonymous and synonymous sites[29].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Beye, M., Hasselmann, M., Fondrk, M. K., Page, R. E. & Omholt, S. W. The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein. *Cell* **114**, 419–429 (2003).
2. Cline, T. W. & Meyer, B. J. Vive la différence: males vs females in flies vs worms. *Annu. Rev. Genet.* **30**, 637–702 (1996).
3. Hasselmann, M. & Beye, M. Signatures of selection among sex-determining alleles of the honey bee. *Proc. Natl Acad. Sci. USA* **101**, 4888–4893 (2004).
4. Pane, A., Salvemini, M., Bovi, P. D., Polito, C. & Saccone, G. The *transformer* gene in *Ceratitis capitata* provides a genetic basis for selecting and remembering the sexual fate. *Development* **129**, 3715–3725 (2002).
5. Hasselmann, M. & Beye, M. Pronounced differences of recombination activity at the sex determination locus (SDL) of the honey bee, a locus under strong balancing selection. *Genetics* **174**, 1469–1480 (2006).
6. Honeybee Genome Sequencing Consortium.. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* **443**, 931–949 (2006).
7. Blencowe, B. J., Bowman, J. A., McCracken, S. & Rosonina, E. SR-related proteins and the processing of messenger RNA precursors. *Biochem. Cell Biol.* **77**, 277–291 (1999).
8. Kulathinal, R. J., Skwarek, L., Morton, R. A. & Singh, R. S. Rapid evolution of the sex-determining gene, *transformer*: structural diversity and rate heterogeneity among sibling species of *Drosophila*. *Mol. Biol. Evol.* **20**, 441–452 (2003).
9. Butler, B., Pirrotta, V., Irminger-Finger, I. & Nothiger, R. The sex-determining gene *tra* of *Drosophila*: molecular cloning and transformation studies. *EMBO J.* **5**, 3607–3613 (1986).
10. Boggs, R. T., Gregor, P., Idriss, S., Belote, J. M. & McKeown, M. Regulation of sexual differentiation in *D. melanogaster* via alternative splicing of RNA from the *transformer* gene. *Cell* **50**, 739–747 (1987).
11. Bell, L. R., Maine, E. M., Schedl, P. & Cline, T. W. *Sex-lethal*, a *Drosophila* sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. *Cell* **55**, 1037–1046 (1988).
12. Keyes, L. N., Cline, T. W. & Schedl, P. The primary sex determination signal of *Drosophila* acts at the level of transcription. *Cell* **68**, 933–943 (1992).
13. Hoshijima, K., Inoue, K., Higuchi, I., Sakamoto, H. & Shimura, Y. Control of *doublesex* alternative splicing by *transformer* and *transformer-2* in *Drosophila*. *Science* **252**, 833–836 (1991).
14. Grimaldi, D. & Engel, M. S. *Evolution of the Insects* 454–467 (Cambridge Univ. Press, UK, 2005).
15. Kerr, W. E. Sex determination in bees. XXI. Number of XO-heteroalleles in a natural population of *Melipona compressipes fasciculata* (Apidae). *Insectes Soc.* **34**, 274–279 (1987).
16. Cook, J. M. Sex determination in the Hymenoptera: a review of models and evidence. *Heredity* **71**, 421–435 (1993).
17. Burkhard, P., Stetefeld, J. & Strelkov, S. V. Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol.* **11**, 82–88 (2001).
18. Beye, M. 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 (2004).
19. Wilkins, A. S. Moving up the hierarchy: A hypothesis on the evolution of a genetic sex determination pathway. *Bioessays* **17**, 71–77 (1995).
20. Nöthiger, R. & Steinemann-Zwicky, M. A single principle for sex determination in insects. *Cold Spring Harb. Symp. Quant. Biol.* **50**, 615–621 (1985).
21. Bull, J. J. *Evolution of Sex Determining Mechanisms* (Benjamin/Cummings Publishing Company, Menlo Park, California, 1983).
22. Pomiankowski, A., Nothiger, R. & Wilkins, A. The evolution of the *Drosophila* sex-determination pathway. *Genetics* **166**, 1761–1773 (2004).
23. Hill, W. G. & Robertson, A. The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
24. Otto, S. P. & Lenormand, T. Resolving the paradox of sex and recombination. *Nature Rev. Genet.* **3**, 252–261 (2002).
25. Lewis, D. The evolution of sex in flowering plants. *Biol. Rev.* **17**, 46–67 (1942).
26. Charlesworth, B. & Charlesworth, D. A model for the evolution of dioecy and gynodioecy. *Am. Nat.* **112**, 975–997 (1978).
27. Charlesworth, D., Charlesworth, B. & Marais, G. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95**, 118–128 (2005).
28. Graves, J. A. Sex chromosome specialization and degeneration in mammals. *Cell* **124**, 901–914 (2006).
29. Nei, M. & Kumar, S. *Molecular Evolution and Phylogenetics* 216–221 (Oxford Univ. Press, New York, 2000).
30. Engel, M. S. Monophyly and extensive extinction of advanced eusocial bees: insights from an unexpected Eocene diversity. *Proc. Natl Acad. Sci. USA* **98**, 1661–1664 (2001).

Author Contributions M.H. performed the evolutionary nucleotide analysis, isolated gene sequences from different species and supervised some experiments; T.G. performed the gene studies; M.S. assembled SDL sequences and identified genes; C.G.N.-S. characterized *M. compressipes* sequences; M.O. analysed domain structures; and M.B. performed the experimental design, supervised the research project and wrote the manuscript.

Author Information The sequences generated in this study are available from GenBank under the accession numbers EU101387 (*GB11211*), EU101388 (*fem*^F), EU101389 (*fem*^M), EU101390 (*csd*), EU101391 (*GB30480*), EU101392 (*GB13727*), EU139305 (*M. compressipes fem*), EU288185 (*B. terrestris fem*), and EU100885–EU100941 (*Apis fem* and *csd* sequences from populations and different species). SDL assembly and sequence annotation data are available in the Third Party Annotation Section of the DDBJ/EMBL/GenBank databases under the accession number TPA: BK006346. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.B. (martin.beye@uni-duesseldorf.de).

METHODS

Bee material. Queens producing 50% female and 50% diploid males were derived from brother–sister crosses (inbred crosses). Male-producing queens were obtained from non-mated, CO_2 treated queens. Female-producing queens were obtained from queens that were inseminated with the semen of a single male. Samples of honeybee species for evolutionary nucleotide analyses were collected in Borneo (Tenom, *A. cerana* and *A. dorsata*), Thailand (Samut Songkram, *A. cerana*; Wanmanaow, *A. dorsata*), Slovenia (Litija, *A. mellifera*) and South Africa (Pretoria, *A. mellifera*). Samples of stingless bees (*Melipona compressipes*) were collected in Brazil (Manaus). Samples of bumble bees (*B. terrestris*) were obtained from the Koppert Company.

Sequence analysis. Sequences of the SDL (AADG05006532, AADG05006533, DQ681226, AY352277 (GenBank accession numbers), and 250636813, 165754571, 566358507, 566314258, 173514040, 566864637, 240540892 and 160908628 (NCBI trace archive numbers)) were assembled using the Staden Package software. Ordering of genomic sequences was verified by cDNA sequences. Potential exons were predicted, and transcriptionally active genes were identified by RT–PCR of embryonic cDNA. The first-strand cDNA was generated by reverse transcription with oligo(dT) primer or random hexamers. RACE experiments were performed to isolate the 5′ and 3′ ends of genes. Sequences were obtained from high-fidelity PCR amplifications of embryonic cDNA and at least three independent clones. We used BLAST programs and a low complexity filter option to compare genes with the database. Domains were analysed by PROSITE database comparisons (http://www.expasy.org/prosite/). The coiled-coil region was predicted by COILS program (http://www.ch.embnet.org/software/COILS_form.html) with a probability of one. Sequences of the coiled-coil motif comparison (Fig. 3d) were derived from GenBank accession numbers: EU100885, EU100893, EU100921, EU100928, EU100904, EU100902, EU100940, EU100938, EU100936. Fem and Csd and Tra proteins (*N. vitripennis* Fem, XM_001604744 and EU780924; *C. capitata* Tra, AF434936; and *D. melanogaster* Tra, P11596) were compared by BLAST2 sequence program (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi).

Functional analysis of genes. RNAi knockdowns were induced in early embryogenesis at the syncytial stage (0–4 h after egg deposition) in haploid and diploid males and females. dsRNAs were generated from cloned cDNAs of genes GB11211, GB13727 and GB30480, and from a DNA marker sequence which we used as an RNAi control. The *fem* and *csd* siRNAs were synthesized (MWG BioTech) and injected at a concentration of 50–100 pg per embryo. Sequences for the nonsense siRNAs used in the control experiments were obtained by scrambling the nucleotide composition of *fem* and *csd* siRNA sequences. Injection of these mock siRNAs did not affect sexual differentiation (data not shown). Individuals derived from inbred crosses (producing diploid males and females) were subject to *csd* genotyping after phenotype analysis at the adult and pupal stage. Head development differs substantially between females (workers) and males (Fig. 1b). Females (workers) have a triangular shaped head, narrow oval-shaped compound eyes, a long proboscis and 12 antennal segments, whereas haploid and diploid males have a round-shaped head, large oval-shaped compound eyes that nearly join in a medio–dorsal position, a short proboscis and 13 antennal segments. Sex-specific splicing of *fem* transcripts were analysed by RT–PCR reactions with RNA prepared from the fourth instar larvae. Amplified fragments were composed of exons 3–6–7–8 (size ~350 bp) and exons 3–4–5–6–7–8 (size ~1.6 kb) corresponding to the female and male transcripts, respectively. Hatched larvae were reared *in vitro*. To quantify the mRNA levels with a BioRad Chromo4, aliquots of first stranded cDNA were amplified, and real-time fluorimetric intensity of SYBR green was monitored. Each sample was run twice in triple replicates. To determine relative transcriptional levels of *csd*, 2 was raised to the power of ΔCts values that were obtained by comparing cycle thresholds (Cts) to those of the reference gene, *elongation factor 1-alpha* (ΔCts = Cts$_{control}$ − Cts$_{target}$). Individual *fem* knockdown embryos were identified by testing individual transcription levels against the distribution of mock-siRNA-treated control samples using *t*-test statistics. Multiple comparisons against the distribution were adjusted using the Bonferroni procedure. Statistical analysis was carried out using the SPSS 15.0 software. Sequences of oligonucleotide primers and siRNAs are listed in the Supplementary Information.

Phylogenetic and molecular evolutionary analysis of nucleotide and amino acid substitutions. The proposed phylogenetic tree shows the evolutionary relationship and divergence times of species under study and is based on amber fossils and morphological data. Sequences were obtained from cloned high-fidelity PCR fragments of cDNAs prepared from embryonic mRNA. Extra primer sets to amplify *csd* from *A. cerana* and *A. dorsata* samples were: csd_forCer2/csd_rev4CIII/; csd_IIIfor/ csd_IIIrev3 (Supplementary Information). The *csd* (*A. mellifera*, n = 15; *A. cerana*, n = 17; *A. dorsata*, n = 19) and *fem* sequences (lacking the Asn/Tyr repeat (hypervariable region); *A. mellifera*, n = 4; *A. cerana*, n = 2; *A. dorsata*, n = 2; *M. compressipes*, n = 1; *B. terrestris*, n = 1) were aligned by the BioEdit program and edited manually to improve the conformity with the open reading frame. Gaps in the sequence alignments were completely deleted in the evolutionary analyses. Trees of coding sequences were obtained by the minimum evolution method. We applied the Nei–Gojobori distance with Jukes–Cantor correction for synonymous differences and Poisson-corrected distances for amino acid differences which are implemented in the MEGA Version 3.1 program. Gene tree analysis including the wasp *N. vitripennis* was confined to the C- and N-terminal parts of the Fem protein, the region for which we detected sufficient homology (amino acid positions 1–54 and 301–404, XP_001604794). Relative rate tests for substitution rate differences were analysed using the RRTree version 1.1 program on synonymous and amino acid differences using either the *B. terrestris* or the *N. vitripennis fem* sequences as an out-group. Non-synonymous and synonymous substitution analysis of branches of the gene tree were obtained by least square method implemented in the b$_N$ − b$_S$ program. The analysis included up to six *csd* alleles per honeybee species. Fisher's exact test for selection was performed on the numbers of changed and unchanged non-synonymous and synonymous sites. The number of potential non-synonymous and synonymous sites in the branch analysis are 647 and 277, respectively. For Southern blotting onto Hybond N+ membranes standard procedures were used.