Current Biology 23, 2559–2564, December 16, 2013 ©2013 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2013.10.070

Report

Gradual Molecular Evolution of a Sex Determination Switch through Incomplete Penetrance of Femaleness

Martin Beye,[1,*] Christine Seelmann,[1] Tanja Gempe,[1] Martin Hasselmann,[2] Xavier Vekemans,[3] M. Kim Fondrk,[4] and Robert E. Page, Jr.[4]

[1]Institute of Evolutionary Genetics, Heinrich Heine University Duesseldorf, Universitaetsstrasse 1, 40225 Duesseldorf, Germany

[2]Institute of Genetics, University of Cologne, Zuelpicher Straße 47, 50674 Koeln, Germany

[3]Laboratoire de Génétique et Evolution des Populations Végétales, UMR 8198, CNRS, Université Lille 1, 59655 Villeneuve d'Ascq, France

[4]School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA

Summary

Some genes regulate phenotypes that are either present or absent. They are often important regulators of developmental switches and are involved in morphological evolution. We have little understanding of the molecular mechanisms by which these absence/presence gene functions have evolved, because the phenotype and fitness of molecular intermediate forms are unknown. Here, we studied the sex-determining switch of 14 natural sequence variants of the *csd* gene among 76 genotypes of the honeybee (*Apis mellifera*). Heterozygous genotypes (different specificities) of the *csd* gene determine femaleness, while hemizygous genotypes (single specificity) determine maleness. Homozygous genotypes of the *csd* gene (same specificity) are lethal [1–6]. We found that at least five amino acid differences and length variation between Csd specificities in the specifying domain (PSD) were sufficient to regularly induce femaleness. We estimated that, on average, six pairwise amino acid differences evolved under positive selection [7–9]. We also identified a natural evolutionary intermediate that showed only three amino acid length differences in the PSD relative to its parental allele. This genotype showed an intermediate fitness because it implemented lethality regularly and induced femaleness infrequently (i.e., incomplete penetrance). We suggest incomplete penetrance as a mechanism through which new molecular switches can gradually and adaptively evolve.

Results

The origin of adaptive gene functions in organisms is predicted to evolve through a series of single mutations. This trajectory imposes some constraints on the molecular origin because molecular intermediates must be functional and must follow mutational paths of equal or increased fitness [10–13]. Details of the molecular evolution of enzymatic gene functions have been systematically followed in the laboratory by induced mutations. Some mutational paths have been identified that associated with changes of enzymatic activity and followed predicted fitness gains [12, 14]. Genes can also regulate phenotypes that are either present or absent that are often important regulators of development and morphological evolution. Hox genes can, for instance, control the presence and absence of entire body segments, while sex determination genes can induce the entire male or female morphology. We have little understanding of the molecular mechanisms by which these absence/presence gene functions have gradually evolved molecularly because the phenotype and fitness gains of the molecular intermediate forms are unknown.

Here, we studied the function and followed the molecular evolution of 14 natural variants among 76 genotypes of the *complementary sex determiner* (*csd*) gene. The *csd* gene controls the switch between male and female development in the honeybee (*Apis mellifera*) [4–6] (Figure 1A). If the *csd* gene is heterozygous, the Csd proteins derived from different *csd* specificities are activating the female developmental pathway. If the *csd* gene is hemizygous (in unfertilized eggs), the Csd proteins derive from the same specificity and have no function in activating the male differentiation pathway. If the *csd* gene is homozygous in diploids, the Csd proteins derive also from the same specificity resulting in male development. However, the diploid male development induces lethality in a colony because the larvae are eaten by the worker bees [2, 3]. Selection against homozygotes provides an evolutionary advantage for newly originated specificities because they can form fewer lethal homozygotes and more vital heterozygotes than specificities that are common within populations [15, 16]. Hence, multiple sex-determining specificities [17–19]) of the *csd* gene have arisen naturally in honeybee populations [16, 20], allowing us to study multiple evolutionary outcomes of the same gene. A pair of *csd* alleles with differences in the nucleotide sequence belong to the same specificity if it induces maleness.

We randomly sampled 15 gene copies of *csd* in a population around Davis, CA. Fourteen of the 15 copies showed differences in the nucleotide (Figure S1A available online) and deduced amino acid sequences (Figure S1B). We generated 76 genotype variants by combining these 14 *csd* sequences in the diploid offspring of the third generation in our crosses (Figure S2A). We studied which of these genotypes induced female development and which implemented maleness and lethality.

Seventy of the 76 different genotypes regularly induced female development (as measured by morphology at the pupal stage) at frequencies that did not deviate from the expected frequency (Figures 1B and S2B). Among these 70 different genotypes, we also found eight genotypes in which the sexual origin of the two specificities forming the same genotype were switched (Figures 1B and S2B). The switched sexual origin did not affect the induction of female development (for instance, due to genomic imprinting). We found one genotype (Y2/G1) that induced femaleness only in some embryos (see the results reported below). Thus, after one sequence of the Y2/G1 pair was excluded, our results showed that each of the 13 *csd* sequences induced femaleness regularly when combined with one of the six to 12 other sequences. We conclude that we have characterized 13 fully separated

*Correspondence: martin.beye@hhu.de

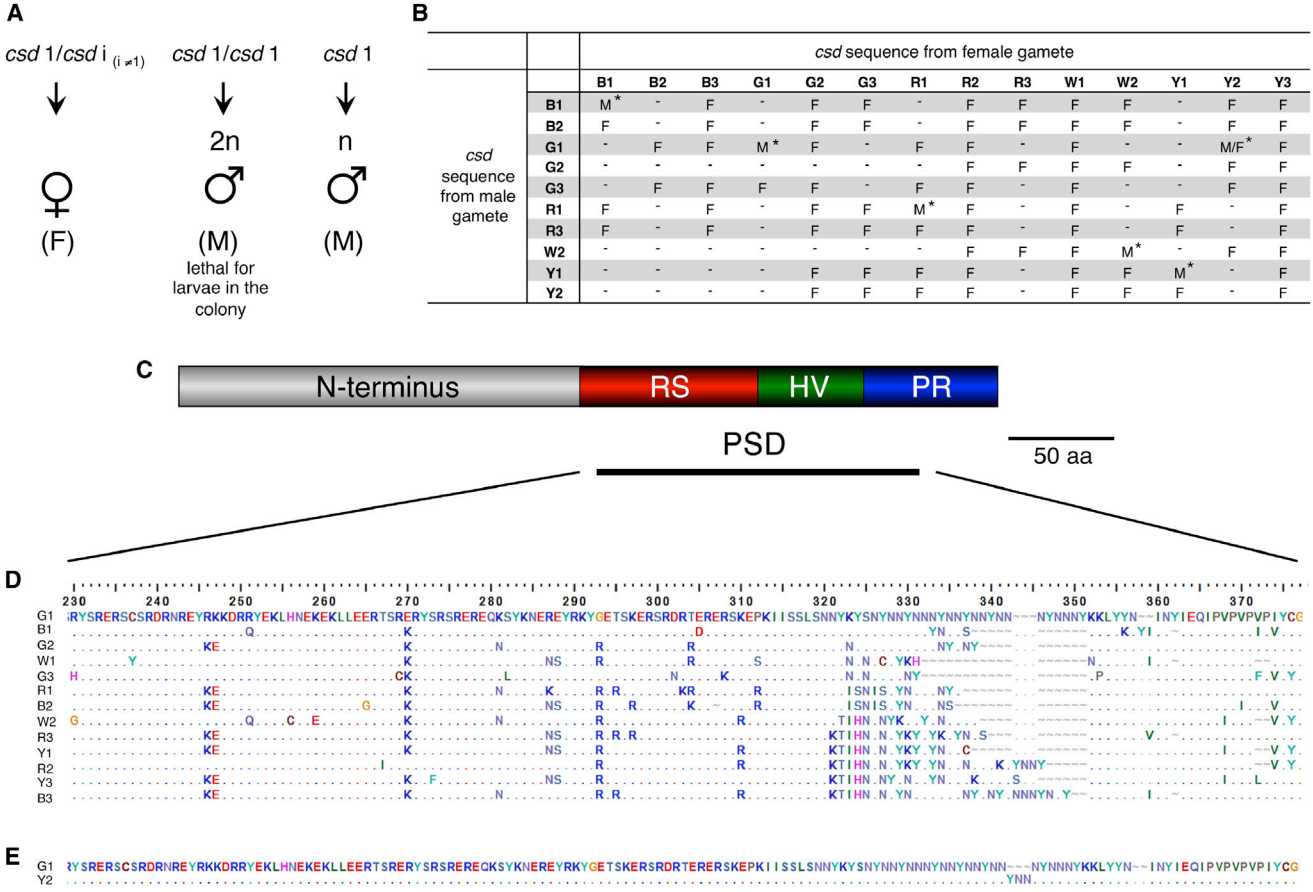

Figure 1. Natural Variants of the *csd* Gene and Sexual Development

(A) *csd* genotypes and the resulting sexual development.

(B) The sex determination function of 14 natural *csd* sequence variants in 76 genotypes. The diploid offspring either develop into females (F) or males (M). The diploid males were eaten by worker bees at larval stage. "*" denotes genotypes that deviate from the expected frequency under the model of 100% female development detected at pupal stage, with the Bonferroni-corrected significance level of $\alpha' = 6 \times 10^{-4}$ for multiple genotype testing and with p = 0.25 (the expected frequency of the genotypes in a cross), $p < 6 \times 10^{-4}$ for the homozygous genotypes, and $p < 5 \times 10^{-5}$ for the Y2/G1 genotype. "–" denotes genotypes that were not tested.

(C) The domains of the Csd proteins. The arginine/serine (RS)-rich domain, the hypervariable (HVR) domain consisting of asparagine- and tyrosine-enriched tandem repeats, the proline-rich (PR) domain, and the potential specifying domain (PSD) are shown. The scale shows the number of amino acids (aa).

(D) The amino acid sequences of the PSD for the 13 fully separated sex-determining specificities.

(E) The amino acid sequences of the PSD for the two not fully separated specificities that induce female development only in a fraction of embryos.

See also Figures S1 and S2.

sex-determining specificities. We studied the molecular differences between the 13 specificities (Table 1 and Figures 1C and 1D). The 70 protein pairs that induced femaleness showed, on average, 30 pairwise amino acid differences (minimum [min.] 12 and maximum [max.] 53). The sequence of the potential specifying domain (PSD), which has been identified as the target of balancing selection [16, 21], shows an average of 27 (min. 5 and max. 65) pairwise amino acid differences in the arginine/serine (RS) domain and part of the proline-rich (PR) domain. The hypervariable (HVR) domain of the PSD showed pairwise length differences of 1 to 21 amino acids in the asparagine- and tyrosine-enriched tandem repeat sequences. The sequence encoding the N-terminal domain had an average of six (min. 0 and max. 14) pairwise amino acid differences, while the genotypes G2/R1 and G1/G3 showed no pairwise differences (G2/R1 have a single-amino-acid indel). Hence, the female-inducing protein variants demonstrate that multiple amino acid differences in the PSD are sufficient to produce separated specificities.

Five genotypes in five crosses never produced females in the pupal stage (Figures 1B and S2B). However, if these genotypes induce male development, we expected to find embryos with the genotypes. We found them in the embryos at the expected frequency (Figure S2C; p > 0.05, with p = 0.25 being the expected frequency of this genotype in that cross). This result suggests that these genotypes induced maleness and lethality. Sequence analysis showed that these genotypes corresponded to five different types of homozygous genotypes. Each homozygous genotype had identical *csd* coding sequences that were derived from a different, single parental specificity in our crosses. These identical sequences result in male development as previously shown by crossing and larval rearing experiments [2–6].

The Y2/G1 genotype induced femaleness only in some embryos (Figures 1B and S2B). We detected only one pupal individual with this genotype (n = 45) that had differentiated female reproductive organs and a female abdomen (Figure 2A), which is below the expected frequency. We genotyped embryos of

Table 1. Pairwise Amino Acid and Synonymous Nucleotide Differences for the Different Functional Classes of *csd* Genotypes

			Sequence of the Entire Protein[a]			Sequence of the N-Terminal Domain			Sequence of PSD[a]			Sequence of HVR
Genotype	Function	Sample Size (n)	aa Total	aa %	K_S	aa Total	aa %	K_S	aa Total	aa %	K_S	aa Length Differences
Homozygous	Male determining	5	0	0	0	0	0	0	0	0	0	0
Intermediate	Female determining at low frequency	1	6	0.01	0.017	6	0.03	0.030	0	0	0	3
Heterozygous	Female determining	70	30 (12–53)	0.07 (0.03–0.13)	0.032 (0.013–0.054)	6 (0–14)	0.03 (0–0.06)	0.016 (0–0.030)	27 (5–65)	0.22 (0.04–0.54)	0.069 (0.024–0.138)	5 (1–21)

Average (minimum–maximum) pairwise amino acid (aa) differences and pairwise synonymous nucleotide sequence differences per site (K_S) for the different functional classes of *csd* genotypes. See also Figures S4 and S5.
[a]Excluding the HVR.

that cross and found the Y2/G1 genotype in 28 of 213 embryos. Hence, the Y2/G1 genotype is underrepresented in the pupal compared to the embryonic sample (Fisher's exact test; $p < 0.02$). This result suggests that this genotype can induce femaleness, but also can frequently induce maleness and lethality. To further determine the sex-determining activity of this genotype, we studied the alternatively spliced transcripts of the *Am-dsx* (*doublesex*) gene in embryos. The *dsx* gene is a key component of insects' sex-determining pathway. The alternatively spliced transcripts of the *dsx* gene translate sex-specific proteins, which control many aspects of morphological differentiation in insects [22–25]. If *csd* genotypes are heterozygous, the Csd proteins direct the female-specific splicing of *Am-dsx* transcripts via the Fem proteins [6]. If *csd* genotypes are hemizygous or homozygous, the male-specific splicing of *Am-dsx* transcripts is mediated [6]. We amplified the *dsx* transcripts in embryos using RT-PCR (Figure 2B). We found in 12 Y2/G1 embryos one embryo (8%) that only contained the female splice form, five embryos that contained both the male and female splice forms (42%), and six embryos that only contained the male splice form (50%) with a female splice form mostly near the background of detection. In the control female embryos and in the female embryos with genotype Y3/G1 (n = 6) that derived from the Y2/G1 genotype containing cross, we detected only the female *Am-dsx* splice form (examples show in Figure 2B). The results on the *Am-dsx* transcripts showed that the Y2/G1 genotype can induce femaleness in a fraction of embryos while it can regulate maleness in another fraction of embryos. We conclude that the Y2 and G1 sequences are not fully separated specificities because the Y2/G1 genotype shows incomplete penetrance of femaleness. The female-determining activity is not complete in most of the feminized embryos because we found the female and male transcripts in those embryos. Next we studied the differences and evolutionary relationship of the Y2 and G1 sequences. The sequences differ in a NNY (asparagine [N]/tyrosine[Y]) repeat in the HVR and in six amino acid residues of the N-terminal domain (Table 1 and Figure 1E). We studied the phylogenetic relationships of sequences encoding either the PSD or the N-terminal domain by generating phylogenetic trees of the 14 *csd* nucleotide sequences (Figure S3). The Y2 and G1 PSD sequences form one phylogenetic cluster (i.e., they showed no differences at nonsynonymous and synonymous nucleotide sites; Table 1), while the Y2 and G1 N-terminal domain sequences are found in separate phylogenetic clusters; this result is supported by the bootstrap test. The phylogenetic relationships of the PSD and the N-terminal domain suggest distinct evolutionary histories due to meiotic recombination. As previously reported [16, 21], meiotic recombination has operated repeatedly in the large fifth intron that separates the exons encoding the N-terminal domain and PSD. Altogether, our results suggest that the single NNY repeat difference in the PSD and the six amino acid differences in the N-terminal domain are not sufficient to establish fully separated sex-determining specificities.

We studied signatures of adaptive evolution in the fully separated specificities. To estimate α, we compared the pairwise nonsynonymous and synonymous polymorphisms in the 70 female-determining genotypes (Figure S4) relative to the net divergence of *csd* sequences between the western (*A. mellifera*) and eastern honeybee species (*Apis cerana*) [26]. We define α as the proportion of pairwise nonsynonymous differences between specificities that were driven by positive selection; this is an extension of the McDonald-Kreitman approach [7, 8]. In this study, we used α to quantify the relative increase in polymorphism relative to divergence [9]. For the entire coding sequence, we obtained an average $\alpha = 0.09$. For the PSD sequence, we obtained an average $\alpha = 0.29$. Relating these values to the number of sites gave an average of 4.7 pairwise nonsynonymous differences in the entire coding sequence (95% confidence interval [CI]: 2.4–7.2) and six in the sequence of PSD (95% CI: 4.5–7.4). We conclude that, on average, approximately five to six of the pairwise amino acid differences between specificities were driven by positive selection and these differences are mainly found in the PSD.

If adaptive evolution in the PSD shaped the rise of new sex-determining specificities, we expect that the α values would be larger for the newly diverged specificities. We used K_S values of the genetically linked exon 8 and 9 sequences to approximate the relative divergence times of the PSD sequences. This provided some support for our hypothesis: PSD α values negatively correlated with the K_S exon 8/9 values (Spearman's rho = −0.24 [95% CI: −0.45 to −0.038]; one-tailed $p = 0.028$). We therefore suggest that positive selection may be more important among the presumably more recently diverged specificities and that the rise of a new specificity was shaped by adaptive evolution.

To study the effect of multiple-step, molecular evolution on the origin of specificities, we compared the shape of the genealogies based on nonsynonymous and synonymous sites. The theory for systems evolving under balancing selection predicts that the shape of the genealogy (a tree representing the phylogenetic relationships among the nucleotide sequences of specificities) is similar to a neutral coalescent but over a longer time scale [27–29] if the specificities can split instantaneously

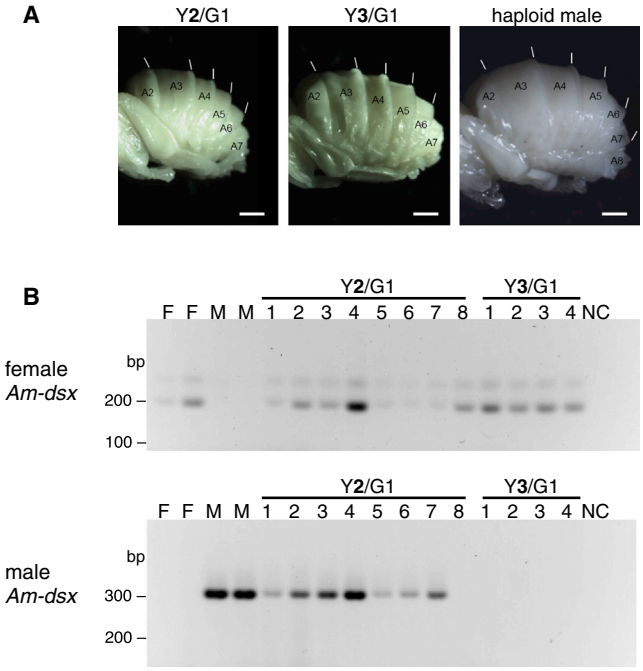

Figure 2. Sexual Phenotype of Pupa and Embryos of Genotype Y2/G1

(A) The abdomen of individual animals are shown with the different numbers of visible abdominal tergits in females (A2 to A7) and males (A2 to A8). Individuals with genotype Y2/G1; a female pupa with genotype Y3/G3; and a haploid male pupa with unknown genotype that were reared in a worker cell are shown. The diploid individuals were derived from the same cross and colony. The Y2/G1 genotype was verified by nucleotide sequence analysis. Scale bars, 1 mm.

(B) Example of sexual splicing of the *Am-dsx* transcripts in single embryos. The RT-PCR fragments from embryos of genotype Y2/G1, of female-determining genotype Y3/G1 and of female (F) and male (M) controls are shown (embryos were older than 48 hr after egg deposition). The female and male *Am-dsx* mRNAs were amplified from embryos using RT-PCRs, resolved by agarose gel electrophoresis, stained with ethidium bromide, and visualized; images are shown in inverted black-and-white mode. The fragments of the female transcripts (with an expected band size of 188 bp) were amplified by using oligonucleotides *Am-dsx*417 and *Am-dsx*419. The PCRs produced also in the control female embryos only a faint band suggesting low amounts of female transcripts. The fragments of the male transcripts (with an expected band size of 301 bp) were amplified using oligonucleotides *dsx*J26 and *dsx*J27. PCRs were repeated at least three times. The identity of fragments as verified by sequencing in at least three control and Y2/G1 embryos. The genotype of single embryos was identified by length polymorphism and the Y2/G1 genotype was confirmed by nucleotide sequencing in at least three embryos. We included a negative control PCR, to which no cDNA was added (last lane, NC).

See also Figure S3.

(for instance, due to a single mutation). We detected that the nucleotide sequences of the PSD have longer terminal branches than would be expected under the neutral coalescent (Figure S5) using two tree-shape statistics in which values close to 1 are expected under the neutral coalescent (R_{ST} and R_{SD}, from [28]; R_{SD} = 6.30, $p < 0.001$ and R_{ST} = 2.44, $p < 0.05$ for synonymous sites; R_{SD} = 5.78, $p < 0.001$ and R_{ST} = 2.33, $p < 0.05$ for nonsynonymous sites).

Discussion

The functional testing of 14 natural *csd* sequence variants in 76 genotypes showed that multiple pairwise amino acid differences in the specifying domain (PSD) are required to induce femaleness (Figure 1). We found 13 fully separated specificities that consistently demonstrated at least five pairwise amino acid and length differences in PSD. Two pairs of fully separated specificities showed no amino acid differences in the N-terminal sequence, suggesting that N-terminal amino acid differences are in general not essential to determine specificity. The partly separate specificity pair Y2/G1 showed in PSD only a length difference located in the HVR, suggesting that such length difference alone is not sufficient to induce the full female-determining activity. We further showed that the N-terminal amino acid differences in the Y2/G1 pair are generated through recombination (Figure S3) and linkage to PSD [16, 21].

We propose that the absence or presence of amino acid differences in the PSD, which comprises the arginine/serine (RS), proline-rich (PR), and the HVR domains, affect binding between Csd proteins [30]. Hence, proteins derived from different specificities have other conformations than proteins derived from the same specificity. This conformational differences in turn lead to functional changes of the proteins that either can or cannot direct the female splicing of downstream target transcripts inducing female development [6].

A key question is how new specificities can adaptively evolve through a series of single mutations if the advantageous phenotype, femaleness, is either present or absent. We suggest that incomplete penetrance of femaleness is an evolutionary mechanism through which a molecular intermediate can possess a fitness advantage. The Y2/G1 genotype contained a natural molecular intermediate sequence, because this sequence pair showed (1) no neutral divergence (K_S = 0) in the sequence encoding the PSD, (2) no amino acid divergence in the PSD outside the hypervariable region (HVR), (3) induction of femaleness in only a fraction of individuals (incomplete penetrance), and (4) no full female-determining activity in those few embryos that were feminized. Because this pair induced femaleness instead of lethality in a fraction of offspring, the molecular intermediate possesses an evolutionary advantage. We suggest a model in which an arising new specificity follows a mutational path of increased fitness through increasing the penetrance of femaleness (Figure 3). This model does not require molecular intermediate forms that evolved through selectively neutral mutations [13, 31]. Consistent with a series of adaptively evolved mutations giving rise to new specificities, we found an average five to six pairwise amino acid differences that evolved under positive selection between specificities which are mainly found in the PSD. We propose that the fast-evolving length differences at the HVR [16] are the initial evolutionary step of separation that is followed by single site mutations in the PSD.

In this study, we showed that the PSD encodes amino acid polymorphisms that determine the specificity of the *csd* gene (Figure 1) and thereby confirm that this part of the gene is the target of balancing selection conferring a rare allele advantage [16, 21]. Previous investigations of variation in *csd* observed ten times higher levels of polymorphism (nucleotide diversity π = 0.09) in the nucleotide sequence encoding the PSD than other parts of the gene and the genome-wide average estimate (π = 0.006) [16, 21]. The estimates of the functionally characterized specificities in this study also showed a peak of polymorphism in the PSD sequence within the *csd* gene (Table1; the average nonsynonymous nucleotide differences per site for PSD is $K_{a:}$ = 0.06).

Although the PSD sequences evolved under strong balancing selection, the genealogy analysis showed that the

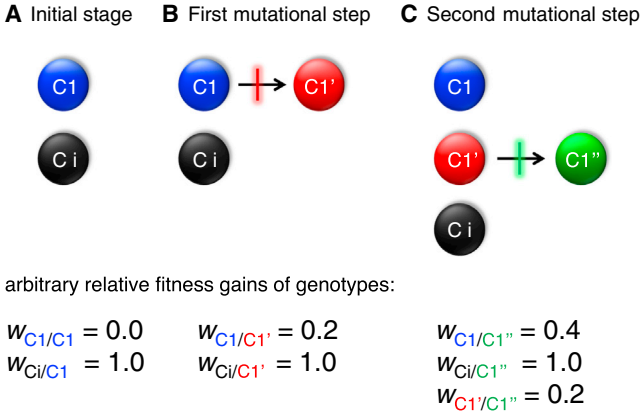

Figure 3. A Scenario Depicting the Gradual Adaptive Evolution of a New Specificity through Incomplete Penetrance

(A) At the initial stage, the parental allele specificity C1 coexists with n − 1 other allele specificities, of which only one allele, the Ci, is represented. The relative fitness of C1/C1 homozygotes is 0 ($w_{C1/C1} = 0$), whereas that of Ci/C1 heterozygotes is 1 ($w_{Ci/C1} = 1$).

(B) An initial mutation introduces the mutant allele C1′. The mutation is initially rare and produces fewer lethal larvae and more females in the offspring than the parental alleles from which the new sequence derived; C1′ is favored over C1 until they reach equal frequencies. The genotype C1/C1′ shows incomplete penetrance of the female-determining activity, producing an intermediate fitness ($w_{C1/C1'} = 0.2 > w_{C1/C1}$).

(C) A second mutation introduces the mutant allele C1″, which is selectively favored over C1′ because it further increases the penetrance of femaleness that will also drive the first mutation C1′ to extinction ($w_{C1/C1''} = 0.4 > w_{C1'/C1''}$). This selection process will continue on a series of single mutations until the penetrance is complete (all offspring of this genotype will develop into females) and the new and parental specificities are fully separated. Vertical colored lines indicate mutation events leading either to the C1′ or the C1″ allele.

topology substantially deviated from those expected under a simple "neutral" coalescence model [15, 16]. The neutral coalescence model of allelic genealogies assumes that specificities have equal fitness and can split instantly (i.e., by a single nonsynonymous mutation), producing genealogies similar to neutral ones but with substantially expanded coalescence time [27, 32]. We suggest that the additional time that is required to produce separate specificities through a series of single mutations will produce the observed longer-than-expected terminal branches. We cannot exclude that meiotic recombination between specificities [33, 34], the sampling of only some specificities, or the population structure might also contribute to generating longer-than-expected terminal branches. However, there has been little evidence for the effects of sampling and population structure in empirical and numerical simulation studies [27, 28, 35–38]. Additionally, we previously found no signature of meiotic recombination events within the PSD [16, 21].

In this study, we molecularly characterized for the first time multiple sex-determining specificities of honeybees. These 13 specificities show an average of 6% pairwise nonsynonymous differences in the PSD. Previous studies based on frequency counts of the number of lethal, diploid drones estimated that 11 to 19 sex-determining specificities are present in local honeybee populations [17–19], suggesting that we have sampled a large fraction of specificities of a local breeding population. Similarly, in previous studies, we identified 14 to 15 substantially diverged gene copies of *csd* within local populations [16, 20]. These copies showed an average of 5% to 6% pairwise nonsynonymous differences in the PSD, but we could not assess whether they determine distinct specificities. Sequence studies of the most variable part of *csd* suggest a substantially larger number of putative specificities at the species level [39].

Using the sequence of information of 13 specificities, we can now predict lethal *csd* genotypes in honeybee breeding programs. Together with low-cost genotyping methodology [40], this will support selection programs in this economically important species which is an essential pollinator of crop plants and wild flowers.

Accession Numbers

The GenBank (NCBI) accession numbers for the *csd* sequence reported in this paper are KF741286–KF741299.

Supplemental Information

Supplemental Information includes Supplemental Experimental Procedures and five figures and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2013.10.070.

Acknowledgments

We thank Eva Theilenberg and Marion Müller-Borg for their splendid technical support. This work was supported by grants from the Deutsche Forschungsgemeinschaft.

Received: July 23, 2013
Revised: September 30, 2013
Accepted: October 29, 2013
Published: December 5, 2013

References

1. Mackensen, O. (1951). Viability and sex determination in the honey bee (*Apis mellifera* L.). Genetics *36*, 500–509.
2. Woyke, J. (1963). What happens to diploid drone larvae in a honey bee colony. J. Apic. Res. *2*, 73–75.
3. Woyke, J. (1965). Genetic proof of the origin of drones from fertilized eggs of the honeybee. J. Apic. Res. *4*, 7–11.
4. Beye, M., Hasselmann, M., Fondrk, M.K., Page, R.E., and Omholt, S.W. (2003). The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein. Cell *114*, 419–429.
5. Hasselmann, M., Gempe, T., Schiøtt, M., Nunes-Silva, C.G., Otte, M., and Beye, M. (2008). Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. Nature *454*, 519–522.
6. Gempe, T., Hasselmann, M., Schiøtt, M., Hause, G., Otte, M., and Beye, M. (2009). Sex determination in honeybees: two separate mechanisms induce and maintain the female pathway. PLoS Biol. *7*, e1000222.
7. Fay, J.C., Wyckoff, G.J., and Wu, C.I. (2002). Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. Nature *415*, 1024–1026.
8. Smith, N.G., and Eyre-Walker, A. (2002). Adaptive protein evolution in *Drosophila*. Nature *415*, 1022–1024.
9. Rose, L.E., Bittner-Eddy, P.D., Langley, C.H., Holub, E.B., Michelmore, R.W., and Beynon, J.L. (2004). The maintenance of extreme amino acid diversity at the disease resistance gene, RPP13, in *Arabidopsis thaliana*. Genetics *166*, 1517–1527.
10. Smith, J.M. (1970). Natural selection and the concept of a protein space. Nature *225*, 563–564.
11. Poelwijk, F.J., Kiviet, D.J., Weinreich, D.M., and Tans, S.J. (2007). Empirical fitness landscapes reveal accessible evolutionary paths. Nature *445*, 383–386.
12. Weinreich, D.M., Delaney, N.F., Depristo, M.A., and Hartl, D.L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. Science *312*, 111–114.
13. Soskine, M., and Tawfik, D.S. (2010). Mutational effects and the evolution of new protein functions. Nat. Rev. Genet. *11*, 572–582.

14. Afriat-Jurnou, L., Jackson, C.J., and Tawfik, D.S. (2012). Reconstructing a missing link in the evolution of a recently diverged phosphotriesterase by active-site loop remodeling. Biochemistry *51*, 6047–6055.
15. Yokoyama, S., and Nei, M. (1979). Population dynamics of sex-determining alleles in honey bees and self-incompatibility alleles in plants. Genetics *91*, 609–626.
16. Hasselmann, M., Vekemans, X., Pflugfelder, J., Koeniger, N., Koeniger, G., Tingek, S., and Beye, M. (2008). Evidence for convergent nucleotide evolution and high allelic turnover rates at the *complementary sex determiner* gene of Western and Asian honeybees. Mol. Biol. Evol. *25*, 696–708.
17. Mackensen, O. (1955). Further studies on a lethal series in the honeybee. J. Hered. *46*, 72–74.
18. Laidlaw, H.H., Gomes, F.P., and Kerr, W.E. (1956). Estimation of the number of lethal alleles in a panmitic population of *Apis mellifera* L. Genetics *41*, 179–188.
19. Adams, J., Rothman, E.D., Kerr, W.E., and Paulino, Z.L. (1977). Estimation of the number of sex alleles and queen matings from diploid male frequencies in a population of *Apis mellifera*. Genetics *86*, 583–596.
20. Hasselmann, M., and Beye, M. (2004). Signatures of selection among sex-determining alleles of the honey bee. Proc. Natl. Acad. Sci. USA *101*, 4888–4893.
21. Hasselmann, M., and Beye, M. (2006). Pronounced differences of recombination activity at the sex determination locus of the honeybee, a locus under strong balancing selection. Genetics *174*, 1469–1480.
22. 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.
23. Suzuki, M.G., Funaguma, S., Kanda, T., Tamura, T., and Shimada, T. (2003). Analysis of the biological functions of a *doublesex* homologue in *Bombyx mori*. Dev. Genes Evol. *213*, 345–354.
24. Hediger, M., Burghardt, G., Siegenthaler, C., Buser, N., Hilfiker-Kleiner, D., Dübendorfer, A., and Bopp, D. (2004). Sex determination in *Drosophila melanogaster* and *Musca domestica* converges at the level of the terminal regulator *doublesex*. Dev. Genes Evol. *214*, 29–42.
25. Shukla, J.N., and Palli, S.R. (2012). *Doublesex* target genes in the red flour beetle, *Tribolium castaneum*. Sci. Rep. *2*, 948.
26. Hasselmann, M., Lechner, S., Schulte, C., and Beye, M. (2010). Origin of a function by tandem gene duplication limits the evolutionary capability of its sister copy. Proc. Natl. Acad. Sci. USA *107*, 13378–13383.
27. Takahata, N. (1990). A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. Proc. Natl. Acad. Sci. USA *87*, 2419–2423.
28. Uyenoyama, M.K. (1997). Genealogical structure among alleles regulating self-incompatibility in natural populations of flowering plants. Genetics *147*, 1389–1400.
29. Schierup, M.H., Vekemans, X., and Christiansen, F.B. (1998). Allelic genealogies in sporophytic self-incompatibility systems in plants. Genetics *150*, 1187–1198.
30. Beye, M. (2004). The dice of fate: the *csd* gene and how its allelic composition regulates sexual development in the honey bee, *Apis mellifera*. Bioessays *26*, 1131–1139.
31. Kimura, M., and Ota, T. (1974). On some principles governing molecular evolution. Proc. Natl. Acad. Sci. USA *71*, 2848–2852.
32. Uyenoyama, M.K. (2003). Genealogy-dependent variation in viability among self-incompatibility genotypes. Theor. Popul. Biol. *63*, 281–293.
33. Schierup, M.H., Mikkelsen, A.M., and Hein, J. (2001). Recombination, balancing selection and phylogenies in MHC and self-incompatibility genes. Genetics *159*, 1833–1844.
34. Charlesworth, D. (2004). Sex determination: balancing selection in the honey bee. Curr. Biol. *14*, R568–R569.
35. Schierup, M.H., Vekemans, X., and Charlesworth, D. (2000). The effect of subdivision on variation at multi-allelic loci under balancing selection. Genet. Res. *76*, 51–62.
36. Muirhead, C.A. (2001). Consequences of population structure on genes under balancing selection. Evolution *55*, 1532–1541.
37. Glémin, S., Gaude, T., Guillemin, M.L., Lourmas, M., Olivieri, I., and Mignot, A. (2005). Balancing selection in the wild: testing population genetics theory of self-incompatibility in the rare species *Brassica insularis*. Genetics *171*, 279–289.
38. Schierup, M.H., Bechsgaard, J.S., and Christiansen, F.B. (2008). Selection at work in self-incompatible *Arabidopsis lyrata*. II. Spatial distribution of S haplotypes in Iceland. Genetics *180*, 1051–1059.
39. Lechner, S., Ferretti, L., Schöning, C., Kinuthia, W., Willemsen, D., and Hasselmann, M. (2013). Nucleotide variability at its limit? Insights into the number and evolutionary dynamics of the sex-determining specificities of the honey bee *Apis mellifera*. Mol. Biol. Evol. Published online November 12, 2013. http://dx.doi.org/10.1093/molbev/mst207.
40. Hyink, O., Laas, F., and Dearden, P.K. (2013). Genetic tests for alleles of *complementary sex determiner* to support honeybee breeding progammes. Apidologie *44*, 306–313.