Current Biology, Vol. 14, R568–R569, July 27, 2004, ©2004 Elsevier Ltd. All rights reserved. DOI 10.1016/j.cub.2004.07.014

Sex Determination: Balancing Selection in the Honey Bee

Dispatch

Deborah Charlesworth

Sequences of alleles of the honey bee's primary sex-determining gene have extremely high diversity, with many amino acid variants, suggesting that different alleles of this gene have been maintained in populations for very long evolutionary times.

Selection can act to maintain allelic variants within a species or population — this is known as balancing selection. Some cases involve 'overdominance', where heterozygous individuals have a fitness advantage over homozygotes. The best known example of this is the human sickle cell polymorphism, a single amino acid difference in β globin which is maintained in malarial regions of the world because *S*/*A* heterozygotes are more resistant to malaria than non-S (*A*/*A*) homozygotes [1], while *S*/*S* homozygotes suffer from sickle-cell anaemia. In contrast, different alleles at self-incompatibility loci in plants are maintained by 'frequency-dependent selection' — pollen grains carrying rare alleles have an advantage as they are compatible with flowers of a majority of the population, rejection occurring only when a pollen grain lands on a stigma of a plant carrying the same allele [2]. New alleles can thus increase in frequency in populations, and come to stable equilibrium frequencies at which all alleles have equal success.

Progress has recently been made in understanding 'balancing' selection of this kind. New cases have been discovered, and the functional basis of the selection is sometimes becoming clear. For instance, the high variation of the histocompatibility (MHC) loci in humans, mice and other vertebrates suggests that balancing selection takes place at these loci, and the fact that non-synonymous variability in the codons encoding the peptide-binding amino acids is disproportionately high, relative to synonymous site diversity, rules out a higher than normal mutation rate and suggests that peptide recognition is involved in the selective process. This variability cannot be due to an elevated mutation rate, as at the codons encoding the peptide-binding amino acids, non-synonymous variability is disproportionately high relative to synonymous site diversity [3,4].

Sequence data alone cannot tell us what is maintaining polymorphism — it is not yet clear how MHC polymorphisms are maintained — but they do tell us that some polymorphisms involve long-term maintenance of variation. This is true of the MHC [5] and plant self-incompatibility [6] loci, whereas the human sickle cell globin polymorphism is clearly of recent origin, as *S* type allele sequences, but not *A* alleles, have the same genetic background for several hundred nucleotides of the surrounding genome region, implying there has been too little time for recombination with other alleles or the introduction of new variants by mutation [7].

Sex-determination systems also involve long-term balancing selection, although they are not often thought of in this way. With chromosomal sex-determination, such as the XX female XY male system of mammals, every mating is an XX by XY cross. The lack of recombination throughout most of the XY pair in males means that genes on the Y are isolated from their X counterparts. Y-linked genes and other sequences thus diverge from the X [8]; in other words, these loci are extraordinarily polymorphic.

Sex-determination in honeybees works quite differently from this. In this haplodiploid system, heterozygotes for a single sex-determining gene develop as females, while males develop when unfertilised, haploid eggs are laid. Rare alleles at the sex-determining gene will be homozygous less often than commoner ones. When a homozygote is formed in a mating, the heterozygosity required for female development is absent, and the egg develops as a diploid male — these are eaten by workers in the colony. Thus, common alleles have lower transmission to the progeny generation than rare ones and there is frequency-dependent selection in this recognition system. There is also heterozygote advantage, as homozygosity is lethal, but it is the frequency-dependent selection that prevents rare alleles from being lost and maintains the allelic polymorphism.

Hasselmann and Beye [9] have now sequenced alleles of the honeybee's primary sex-determining gene, *csd*, which last year [10] was genetically mapped to a small region of the bee genome and definitively identified by functional tests. The gene has nine exons (Figure 1), encoding a protein of around 400 amino acids (lengths differ between alleles) with an arginine-serine-rich (RS) domain similar to those known in some other proteins [10]. The initial study included sequences from only a few individuals, but high diversity was apparent. Now diversity has been studied with a much larger sample, and analysed to test for selection.

Thirty coding sequences were obtained for *csd* cDNAs derived from eggs of a couple of bee colonies each from four countries: Brazil, Berlin, Germany, USA and South Africa. Because queen bees mate with multiple males, each colony will include at least the two alleles of its queen — assuming that all progeny are sons and daughters of the resident queen — plus alleles of some or all of her mates. All the sequences were different, with fifteen recognisable lineages, divided into two clusters: a highly varied set of fourteen 'type I' lineages, found in all four countries, and a very different, and much less variable 'type II' lineage, found in three countries, but not South Africa (probably just a chance effect, as only five sequences were obtained from this population). The co-existence of different type I alleles within

Institute of Cell, Animal and Population Biology, University of Edinburgh, Ashworth Laboratory, King's Buildings, West Mains Road, Edinburgh EH9 3JT, UK.

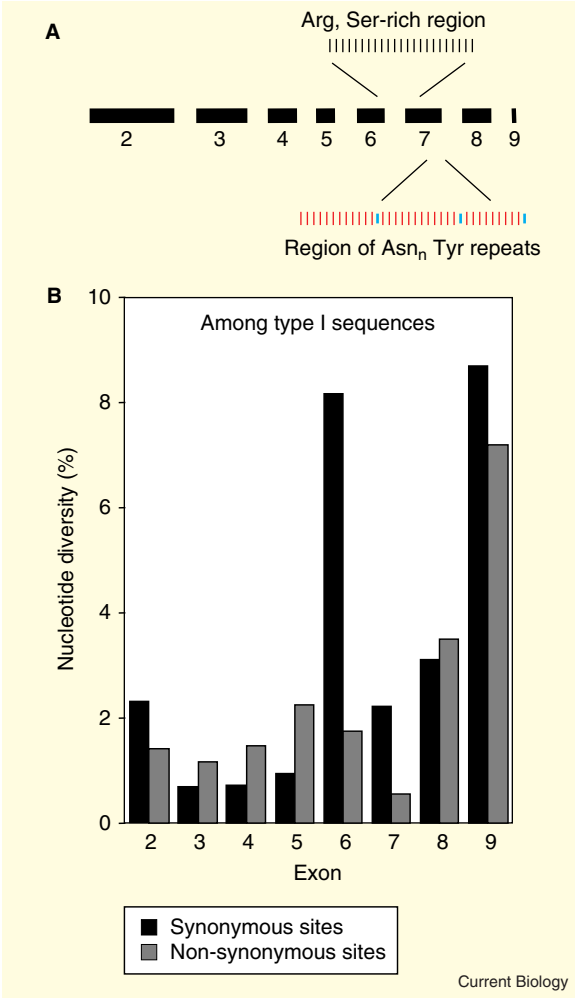

Figure 1. Diversity among different type I sequences of the honey-bee sex-determining gene *csd*.

(A) The exon structure of the *csd* gene. (B) The diversity is expressed on a per site basis, in order to compare the diversity levels for synonymous and non-synonymous sites (reviewed in [12]), and given as values multiplied by 100.

populations suggests that balancing selection maintains these variants.

Another sign of balancing selection is that part of the 3′ region of the type I sequences is hypervariable: different versions have different numbers of repeats of a motif of one to four asparagines, followed by tyrosine; this region is missing from the type II sequences. Even excluding this region, diversity among the type I sequences is generally very high, especially in exons 6–9 (Figure 1). In exons 4–5, non-synonymous diversity is 20% higher than the synonymous value. Although Hasselmann *et al.* [9] did not compare these values with those of other honeybee genes, sequences encoding functional proteins generally have much lower non-synonymous than synonymous diversity, and values as high as these suggest diversifying selection (provided that the sequences really are alleles at a single locus, rather than duplicate genes).

Differences between type I and type II sequences were used to obtain the expected numbers of synonymous and non-synonymous differences, showing that the diversity differences between regions within type I are statistically significant, with the highest non-synonymous variation close to the hypervariable region. There is also very high synonymous diversity in this region, consistent with long-term maintenance of different alleles, so that synonymous, as well as non-synonymous differences have accumulated between them [11]. It will be interesting to study the introns, as these should also have high diversity if the honeybee *csd* gene really is yet another case of long-term balancing selection.

The lineages cannot at present be assigned to their different functional specificities, but the numbers of lineages are consistent with the deduced numbers of up to nineteen alleles at this locus. Assigning alleles to different 'types' may help us to understand which amino acids determine specificity differences. If the alleles recombine, differences between lineages may be restricted to just those sites and adjoining ones, while sequence regions in between may be exchanged between alleles. This case may thus resemble MHC alleles, where the codons for the peptide binding amino acids have many non-synonymous differences [3,4], but other parts of the protein are much less variable. Knowing the sequences also opens the way to genetic studies. Females transmit their two alleles to different progeny, so looking for sequences present in exclusive sets of offspring will identify the queen's alleles, which must represent two different functional types, and then those of her mates can be inferred. Tests will then be possible of whether types I and II are alleles at one locus. Thus the present data lead to hopes for rapid progress in the future.

References

1. Hill, A.V., Allsopp, C.E., Kwiatowski, D., Anstey, N.M., Twumasi, P., Rowe, P.A., Bennett, S., Brewster, D., McMichael, A.J., and Greenwood, B.M. (1991). Common West African HLA antigens are associated with protection from malaria. Nature *352*, 595–600.
2. Nasrallah, J.B. (2002). Recognition and rejection of self in plant reproduction. Science *296*, 305–308.
3. Hughes, A., and Nei, M. (1989). Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. USA *86*, 958–962.
4. Hughes, A.L., Ota, T., and Nei, M. (1990). Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. Mol. Biol. Evol. *76*, 515–524.
5. Klein, J., Sato, A., and Nagl, S. (1998). Molecular trans-specific polymorphism. ARES *29*, 1–21.
6. Ioerger, T.R., Clark, A.G., and Kao, T.-H. (1990). Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. Proc. Natl. Acad. Sci. USA *87*, 9732–9735.
7. Currat, M., Trabuchet, G., Rees, D., Perrin, P., Harding, R.M., Clegg, J.B., Langaney, A., and Excoffier, L. (2002). Molecular analysis of the beta-globin gene cluster in the Niokholo Mandenka population reveals a recent origin of the beta(S) Senegal mutation. Am. J. Hum. Genet. *70*, 207–223.
8. Lahn, B.T., and Page, D.C. (1999). Four evolutionary strata on the human X chromosome. Science *286*, 964–967.
9. 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.
10. 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.
11. Rosenberg, N.A., and Nordborg, M. (2002). Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nat. Rev. Genet. *3*, 380–390.
12. Hurst, L.D. (2002). The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet. *18*, 486–487.