

RESEARCH ARTICLE

The Sex Determination Gene Shows No Founder Effect in the Giant Honey Bee, *Apis dorsata*

Zhi Yong Liu^{1,2#}, Zi Long Wang^{1#}, Wei Yu Yan¹, Xiao Bo Wu¹, Zhi Jiang Zeng^{1*}, Zachary Y. Huang^{3,4*}

1 Honeybee Research Institute, Jiangxi Agricultural University, Nanchang, China, **2** Experimental Animal Center, Institute of Occupational Disease Prevention, Nanchang, China, **3** Department of Entomology, Michigan State University, East Lansing, Michigan, United States of America, **4** Ecology, Evolutionary Biology and Behavior Program, Michigan State University, East Lansing, Michigan, United States of America

Abstract

Background

All honey bee species (*Apis spp*) share the same sex determination mechanism using the complementary sex determination (*csd*) gene. Only individuals heterozygous at the *csd* allele develop into females, and the homozygous develop into diploid males, which do not survive. The honeybees are therefore under selection pressure to generate new *csd* alleles. Previous studies have shown that the *csd* gene is under balancing selection. We hypothesize that due to the long separation from the mainland of Hainan Island, China, that the giant honey bees (*Apis dorsata*) should show a founder effect for the *csd* gene, with many different alleles clustered together, and these would be absent on the mainland.

Methodology/Principal Findings

We sampled *A. dorsata* workers from both Hainan and Guangxi Provinces and then cloned and sequenced region 3 of the *csd* gene and constructed phylogenetic trees. We failed to find any clustering of the *csd* alleles according to their geographical origin, i.e. the Hainan and Guangxi samples did not form separate clades. Further analysis by including previously published *csd* sequences also failed to show any clade-forming in both the Philippines and Malaysia.

Conclusions/Significance

Results from this study and those from previous studies did not support the expectations of a founder effect. We conclude that because of the extremely high mating frequency of *A. dorsata* queens, a founder effect does not apply in this species.

Citation: Liu ZY, Wang ZL, Yan WY, Wu XB, Zeng ZJ, et al. (2012) The Sex Determination Gene Shows No Founder Effect in the Giant Honey Bee, *Apis dorsata*. PLoS ONE 7(4): e34436. doi:10.1371/journal.pone.0034436

Editor: Guy Smagghe, Ghent University, Belgium

Received: April 26, 2011; **Accepted:** March 5, 2012; **Published:** April 12, 2012

Copyright: © 2012 Liu et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by China Agriculture Research System (No. CARS-45), the National Natural Science Foundation of China (No. 31060327), the China Postdoctoral Science Foundation (20100481191) and Michigan Agricultural Experiment Station at the Michigan State University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: bees1965@sina.com; bees@msu.edu (ZYH)

These authors contributed equally to this work.

INTRODUCTION

In honey bees (genus *Apis*), as in other hymenopteran insects, sex is controlled by a mechanism named complementary sex determination (*csd*) [1], [2]. Diploid heterozygotes become females, while haploids develop into males. When the two copies of *csd* alleles are homozygous, the individual develops into a diploid male, which is eaten by workers [3]. Therefore, honey bees are under great selective pressure to become heterozygous. This complementary sex determination theory remained a hypothesis until recently. Beye et al. [4] cloned the *csd* gene by positional cloning and RNA interference. They found that this gene maintains many highly variable alleles in a group and no transcription difference exist between the sexes. *Csd* encodes an arginine serine-rich (SR) type protein, which contains an RS domain in the middle and a proline-rich region at its C terminus, between these two domains is a hypervariable region that differs highly among alleles and has variable numbers of asparagines/tyrosine repeats. The honey bee *csd* is homologous to *Drosophila tra*, which is involved in *Drosophila* sex determination.

Previous studies have shown that *csd* is evolving under balancing selection, and several parts of the coding region are possible targets of selection [5]–[9]. Moreover, the polymorphic level is approximately seven times higher in *csd* than in the neutral regions [8]. Although a previous study found no “bottleneck” effect in the *csd* gene of *A. dorsata* and *A. mellifera*, but suggestive of such an effect in *A. cerana* [9], the Hainan Island in China provides a more specific opportunity to test whether the *csd* gene has a founder effect. Hainan Island in China was thought to have been separated from the mainland about 20 million years ago [10]. The width of the Qiongzhou Strait separating Hainan Island from the mainland is about 18–33 km (Fig. 1). The maximum distance of *Apis mellifera* queen mating flight is about 15 km [11], but most of the queens mated at a distance of 7.5 km or less.



Figure 1. A map showing the strait between Hainan Island and Guangxi Province and the geographical distribution of the collected samples in this study.

doi:10.1371/journal.pone.0034436.g001

It was suggested that *Apis dorsata* drones have “significantly shorter drone flight distance” compared to *A. mellifera* [12]. Regardless, the 18 km distance of the narrowest distance in the straight should be far enough to prevent any gene exchange between the *A. dorsata* populations in the island and the mainland. It is believed that the *csd* gene is under balancing selection, that is, it is under positive selection to create more alleles to avoid becoming homozygous, but at the same time alleles are lost due to sampling effect, creating a balancing force. If the *A. dorsata* population on Hainan Island has been separated from these on the mainland for more than 20 million years, we expect that due to the founder effect [13], the island should have many different alleles of *csd* compared to the mainland due to the initial sampling and then the expansion of the initial population. In addition, the island population should be under stronger selection force due to a smaller initial subset of the mainland *csd* alleles. This can be manifested by showing a higher nucleotide diversity value, due to either a higher level of polymorphism or a higher mutation rate. In this study we tested whether island *A. dorsata* populations experience the founder effect and stronger positive selection by comparing *csd* genes in workers sampled on the island and those on the mainland.

METHODS

Samples collection

A. dorsata samples were collected from two provinces of China, Hainan Province, and Guangxi Province. The Hainan and Guangxi provinces are geographically separated by the Qiongzhou Strait. In each province, three *A. dorsata* colonies were collected, with 30 adult workers from each colony. In Hainan, two colonies were sampled from Qiongzhou County and one colony from Baoting County; and in Guangxi, two colonies were sampled from Chongzuo County and one from Wuming County. The *A. dorsata* in China are feral and usually nest on tall trees, adult workers were collected from colonies nested in trees using a net after climbing up the tree. The samples were first collected into 95% ethanol and then stored at -70°C until use.

DNA extraction

Total genomic DNA was extracted from the head and thorax of each sampled bee according to the protocol of the Animal Genomic DNA Extraction Kit (BEST ALL-HEAL LLC, NY, USA)

PCR and sequencing

The primers used for amplifying region 3 of *A. dorsata csd* gene in this study were designed by consulting those reported by [8] with some modifications. The primers are 5' AATTGGATTTATTAATATAATTTATTATTCAGG 3' (forward) and 5' ATYTCATTATTCAATATGTTNGCATCA 3' (reverse). High Fidelity LA Tag DNA polymerase (BEST ALL-HEAL LLC, NY, USA) was used for all PCR reactions. PCR conditions were denaturation at 94°C for 3 min, followed by 30 cycles of 94°C for 30 s, annealing at 54°C for 30 s and extension at 72°C for 2 min, with a final extension at 72°C for 7 min. PCR products were purified by using DNA GEL EXTRACTION kits (BEST ALL-HEAL LLC, NY, USA) and cloned into pEASY-T3 vectors (Transgene, Beijing, China). To obtain as many *csd* alleles as possible, the genomic

region 3 of *csd* gene was cloned from each sample, and 1–3 clones of each cloned fragment were subjected to double-strand sequencing. Single-sequencing reads were assembled by using Seqman program in DNASTAR software [14].

Definition of *csd* alleles

To avoid the effect of neutral mutation on our results, we adopted the method used by Hasselmann et al. [9], we classified those sequence variants that form a bush-like structure in the tree as neutral variants of a single allele. In each neutral cluster, only one sequence was chosen as an allele for further analysis, while the remaining sequences of the same cluster were excluded. We identified these clusters by π estimates and Z test on the whole genome sequences as well as the coding sequences of these alleles. Similar to Hasselmann and Beye [9], we found that using this method, sequences classified as neutral variants had the same repeat structure in the hypervariable region, while those identified as alleles were unique in that region.

Sequence analysis

Sequences of *A. dorsata csd* genomic region 3 (N = 42 sequences) published by other research groups were downloaded from Genbank under accession numbers DQ325038 - DQ325060 and EU100917- EU100935, repetitive sequences were removed and the remaining sequences (N = 31) were used for phylogenetic analysis together with our sequences. Exons, introns and coding regions of these sequences were determined by consulting those sequences of *csd* genomic region 3 reported by [8]. Nucleotide sequence alignments were performed with Clustal X version 1.8 [15], the alignment results were adjusted manually for obvious alignment errors.

DAMBE 4.1.19 [16] was used to identify haplotypes. Phylogenetic trees were constructed using MEGA version 4.0 program [17]. The minimal evolution (ME) method and Kimura's 2-parameter distances were adopted to obtain an unrooted tree with 2,000 bootstrap replications. Nucleotide diversity (π) and gene flow value (Nm) were calculated by using DNAsp5.0 program [18]. Two tailed Z test was used to detect significant differences between π values of Guangxi and Hainan populations. Pair-wise F_{st} distances were computed using ARLEQUIN 3.1 software [19]. Jost D test was performed at the website based program SMOGD with 200 bootstrap replicates [20]. Median joining (MJ) network was drawn using the program Network 4.5 [21].

RESULTS

csd variation in Guangxi and Hainan populations

For the *csd* gene region 3, we obtained a total of 63 and 55 different sequence variants from Guangxi and Hainan populations, respectively. These 63 and 55 sequence variants were respectively aligned and genealogy trees were constructed. After removing neutral variants, 17 different alleles from Guangxi and 18 different alleles from Hainan province were chosen for further analysis. Of them, 5 alleles were shared by both Guangxi and Hainan populations, resulting in a total of 30 haplotypes. To verify that our sequences were indeed *csd* gene, we compared our sequences with those *Apis dorsata csd* allele sequences reported by Hasselmann *et al.* [9] through amino acid sequence alignment (Figure S1). The

alignment results showed that our sequences were highly similar to those *csd* alleles reported by Hasselmann et al. [9]. For example, each of our alleles contained a hypervariable region between the arginine/serine-rich (RS) domain and the proline rich domain, just as reported in [9]. The hypervariable region was rich in asparagine (N) and tyrosine (Y) that form (N)₁₋₄Y repeats and (KHYN)₁₋₄ repeats in each allele.

The nucleotide diversity values (π) of *csd* in Guangxi and Hainan populations were 0.06123 ± 0.00576 and 0.05271 ± 0.00631 , respectively, without a significant difference between the two (two tailed Z test, $Z = 0.997$, $P > 0.05$, Table 1). Moreover, the nucleotide diversity values of coding region, exon region and intron region of *csd* alleles also showed no significant differences between these two locations (Table 1).

Table 1. Nucleotide diversity (π) (mean \pm SD) of *csd* alleles in Hainan and Guangxi populations.

doi:10.1371/journal.pone.0034436.t001

Genetic differentiation between Guangxi and Hainan populations

The population pairwise F_{st} genetic distance between Guangxi and Hainan populations was 0.00236. The gene flow value (Nm) between these two populations was 10.71. The harmonic mean of D_{est} value between them was -0.03 .

Phylogeny of *Apis dorsata csd* alleles

When all the alleles from Guangxi and Hainan populations were used to construct a phylogenetic tree by using the ME method, *csd* alleles did not form two branches reflecting the Guangxi and Hainan samples. Rather, the two groups were well mixed among each other (Fig. 2). A median joining network was constructed based on the 30 haplotypes to investigate the possible relationships among these haplotypes (Fig. 3). The network suggests that the haplotypes from the same population varied greatly, and haplotypes from different geographical populations were intermixed. We then constructed another tree using ME method, using the coding region of all the available *csd* data for *A. dorsata*, including these obtained in our current study, and that from [8] and [9] (Fig. 4). Again, it showed very little or no clustering according to geographical origins of each *csd* allele.



Figure 2. The gene genealogy of *csd* alleles in region 3 from Guangxi and Hainan *A. dorsata* populations.

The minimum evolution method and Kimura's two parameter distances were used to construct the tree. Bootstrap percentages are shown on internal branches. The scale bar represents the number of nucleotide changes per site. Geographical origin is indicated by letters: HN, Hainan; GX, Guangxi.

doi:10.1371/journal.pone.0034436.g002



Figure 3. The median joining network of 30 haplotypes in Guangxi and Hainan



populations.

Circles denote haplotypes unique to Guangxi (yellow), or unique to Hainan (blue), or in both Guangxi and Hainan (half yellow and half blue). The sizes of circles are proportional to haplotype frequencies.

doi:10.1371/journal.pone.0034436.g003



Figure 4. The gene genealogy of region 3 of the *A. dorsata csd* alleles from this study and those from Cho et al. 2006 [8] and Hasselmann et al. 2008 [9].

The coding sequences of all the alleles were used. The minimum evolution method and Kimura's two parameter distances were used to construct the tree. Bootstrap percentages are shown on internal branches. The scale bar represents the number of nucleotide changes per site. Amcsd-AW4B and Amcsd-AW5A, both from the Western honey bee *Apis mellifera*, were used as outgroup sequences.

doi:10.1371/journal.pone.0034436.g004

DISCUSSION

Contrary to our original hypothesis, we did not find an expansion of *csd* alleles in *Apis dorsata* at the Hainan province, which would manifest as a different clade in a phylogenetic tree (Fig. 2). Furthermore, after we combined sequence data from our study and previous studies [8], [9], we failed to find any clustering of *A. dorsata csd* alleles from other islands such as the Philippines and Malaysia. The separation times of these two regions from the mainland were believed to be 23 million and 40 million years, respectively [22]. Our results are consistent with a previous study by Hasselmann et al. [9] showing that the *csd* genes of *A. mellifera*, and *A. dorsata* not showing a bottleneck effect. Further, their *csd* genealogy tree (Fig. 2) showed a similar intermix of *csd* alleles originating from Malaysia and Thailand.

The F_{st} distance between the Hainan and Guangxi *A. dorsata* populations was 0.236%, that is, 99.764% of genetic variation was within populations and 0.236% was among populations. In general, $F_{st} < 5\%$ indicates a low level of genetic differentiation [23]. There was also a high gene flow value ($Nm = 10.71$) between these two populations, suggesting frequent exchange of genetic material between these two populations in history. All these results demonstrated that the genetic differentiation between these two populations was very low.

The lack of geographical clustering of *A. dorsata* alleles and the small genetic difference between the Hainan and Guangxi populations can be explained by two possible hypotheses. One is that our original assumption of the separation time between Hainan and the mainland was not accurate. Indeed, the separation time of Hainan from the mainland remains a mystery. While [10] cited a separation time of 20 million years, others claim the separation was not stable and the straight was dried and refilled many times since then [24]. The most recent separation was believed only to be 15,000 years ago. If the second estimate is correct, then we do not expect any *csd* changes in such a short time frame.

The alternative hypothesis is that even if the separation time is indeed 20 million years, one still would not expect a founder effect in the *csd* gene due to the particularity of honey bee biology. This is due to the

highly polyandrous nature of honey bees, especially for *A. dorsata*. Previous reports for the queen mating frequency is 20–26 times for *A. dorsata* [25], [26], but a more recent study reported that the maximum number of subfamilies found per colony is 102 and the effective mating frequency is 88.5 [27]. The total number of *csd* alleles world wide is probably well under 100, and in a particular region is more likely to be about 20–30 ([5]–[7], this study). Therefore, even if only one colony of *A. dorsata* was “sampled” to be on the Hainan island during the separation (which is a highly unlikely low number), there would still be enough *csd* alleles preserved in the island to prevent the occurrence of any founder effect. This is confirmed by results of analysis after including the *csd* data from other island like countries, such as the Philippines and Malaysia (Fig. 4).

Our current study therefore suggests that due to the highly polyandrous nature of honey bees, the founder effect does not apply to the *csd* gene in *Apis dorsata*, perhaps also to any other gene, to most *Apis* species. Consistent with this idea, a previous study [28] on Africanized bees in Costa Rica also showed that “that Africanized honey bee populations are very representative of the genetic variation of African bees from the Transvaal region of South Africa.” In contrast, in social insects that have no multiple matings, such as the fire ant *Solenopsis invicta*, a founder effect was found in their sex determination system [29].

SUPPORTING INFORMATION

Figure S1.

Amino acid sequence alignment of region 3 of the *Apis dorsata csd* alleles from this research (those started with Adcsd) and those reported by Hasselmann *et al.* 2008 [9] (those started with Dors). A total of 54 predicted amino acid sequences were used. Alignment was done by Clustal X version 1.8 software. The alignment suggests that our sequences from DNA are truly *csd* alleles because they are structurally highly similar to those of [9] which were from cDNA. Both our sequences and those from Hasselmann *et al.* [9] contain the arginine/serine-rich (RS) domain, the hypervariable region, and the proline rich domain. Furthermore, number 53 (this study) and number 54 [9] are identical in all amino acid sequences. Others, such as 29, 30, 31 only show small differences (2 in RS domain, 1 in proline rich domain), other than the differences in the hypervariable region.

(EPS)

ACKNOWLEDGMENTS

We thank Junjun Hu from Guangxi and Tianbing Wang from Hainan for sampling honey bees; Soochin Cho, Melissa Huang, Katrina Klett and Matthew Lundquist for improving this manuscript.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: ZYH ZJZ. Performed the experiments: ZYL ZLW. Analyzed the data: ZLW. Contributed reagents/materials/analysis tools: WYY XBW. Wrote the paper: ZYH ZYL ZJZ.

REFERENCES

1. Cook JM
(1993) Sex determination in the hymenoptera: a review of models and evidence. *Heredity* 71: 421–435.
2. Heimpel GE, De Boer JD
(2008) Sex determination in the hymenoptera. *Annu Rev Entomol* 53: 209–230.
3. Woyke J
(1963) What happens to diploid drone larvae in a honeybee colony. *J Apicult Res* 2: 73–76.
4. Beye M, Hasselmann M, Fondrk MK, Page RE, Omholt SW
(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, Beye M
(2004) Signatures of selection among sex-determining alleles of the honey bee. *Proc Natl Acad Sci USA* 101: 4888–4893.
6. Hasselmann M, 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.
7. Charlesworth D
(2004) Sex determination: balancing selection in the honey bee. *Curr Biol* 14: 568–569.
8. Cho S, Huang ZY, Green DR, Smith DR, Zhang J
(2006) Evolution of the complementary sex-determination gene of honey bees: Balancing selection and trans-species polymorphisms. *Genome Res* 16: 1366–1375.
9. Hasselmann M, Vekemans X, Pflugfelder J, Koeniger N, Koeniger G, et al.
(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.
10. Huang FS
(2002) *Forest insects in Hainan*. Beijing: Science Press. 1064 p.
11. Jensen AB, Palmer KA, Chaline N, Raine NE, Tofilski, et al.
(2005) Quantifying honey bee mating range and isolation in semi-isolated valleys by DNA microsatellite paternity analysis. *Conservation Genetics* 6: 527–537.
12. Koeniger N, Koeniger G, Tingek S, Kalitu A, Mardan M
(1994) Drones of *Apis dorsata* (Frabircius 1793) congregate under the canopy of tall emergent trees in Borneo. *Apidologie* 25: 249–264.
13. Provine WB
(2004) Ernst Mayr: Genetics and speciation. *Genetics* 167(3): 1041–1046.
14. Burland TG
(2000) DNASTAR's Lasergene sequence analysis software. *Methods Mol Biol* 132: 71–91.
15. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG
(1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality

- analysis tools. *Nucleic Acids Res* 25: 4876–4882.
16. Xia X, Xie Z
(2001) DAMBE: Software package for data analysis in molecular biology and evolution. *J Hered* 92: 371–373.
 17. Tamura K, Dudley J, Nei M, Kumar S
(2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol and Evol* 24: 1596–1599.
 18. Librado P, Rozas J
(2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
 19. Excoffier L, Laval G, Schneider S
(2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform* 1: 47–50.
 20. Crawford NG
(2010) Smogd: software for the measurement of genetic diversity. *Mol Ecol Resour* 10: 556–557.
 21. Bandelt H, Forster P, Rohl A
(1999) Median joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16: 37–48.
 22. Soulier-Perkins A
(2000) A phylogenetic and geotectonic scenario to explain the biogeography of the Lophopidae (Hemiptera, Fulgoromorpha). *Palaeogeogr Palaeoclimatol* 160(3–4): 239–254.
 23. Hartl DL, Clark AG
(1997) *Principles of Population Genetics*, 3 ed. Sunderland: Sinauer Associates Press. 542 p.
 24. South China Sea Institute of Oceanology
(1978) *Quaternary Littoral Geology of South China*. Beijing: Science Press. pp. 15–54.
 25. Moritz RFA, Kryger P, Koeniger G, Estoup A, Tingek S, et al.
(1995) High degree of polyandry in *Apis dorsata* queens detected by DNA microsatellite variability. *Behav Ecol Sociobiol* 37: 357–363.
 26. Oldroyd BP, Smolenski AJ, Cornuet JM, Wongsiri S, Estoup A, et al.
(1996) Levels of polyandry and intracolony genetic relationships in *Apis dorsata* (Hymenoptera Apidae). *Ann Entomol Soc Am* 89: 276–283.
 27. Wattanachaiyingcharoen W, Oldroyd BP, Wongsiri S, Palmer K, Paar J
(2003) A scientific note on the mating frequency of *Apis dorsata*. *Apidologie* 34: 85–86.
 28. Segura JAL
(2000) Highly polymorphic DNA markers in an Africanized honey bee population in Costa Rica. *Genetics and Molecular Biology* 23: 317–322.
 29. Ross KG, Vargo EL, Keller L, Tragers JC
(1993) Effect of a founder event on variation in the genetic sex-determining system of the fire ant *Solenopsis invicta*. *Genetics* 135: 843–854.

All site content, except where otherwise noted, is licensed under a Creative Commons Attribution License.