Nucleotide diversity based on *csd* gene of the black giant honey bee, *Apis laboriosa* (Hymenoptera: Apidae)

Huan WANG, Zi-Long WANG, Zhi-Jiang ZENG, Xiao-Bo WU and Wei-Yu YAN*

Honeybee Research Institute, Jiangxi Agricultural University, Nanchang 330045, P.R. China;
e-mails: HW: wanghuan111986@163.com; ZLW: wzlcqbb@126.com; ZJZ: bees1965@sina.com;
XBW: wuxiaobo21@163.com; WYY: ywygood-0216@163.com

Key words. Hymenoptera, Apidae, *Apis laboriosa*, complementary sex determination, *csd* haplotype, nucleotide diversity

Abstract. The mechanism of sex determination is common for all honeybee species (*Apis* spp.) by the complementary sex determination (*csd*) gene. The *csd* gene has been studied in the Western honey bee (*Apis mellifera* L.), the Eastern honey bee (*Apis cerana* F.) and the giant honey bee (*Apis dorsata* F.), whereas no studies had been conducted on the high altitude Himalayan or black giant honey bee, *Apis laboriosa* Smith. In the present study, we cloned the genomic exon 6 to exon 9 region of the *A. laboriosa csd* gene, and identified 13 *csd* haplotypes. The data was analyzed and compared with the other aforementioned three honeybee species. The results showed that, as with the other three *Apis* species, region 3 of the *csd* gene contains an RS domain at the N terminal, a proline-rich domain at the C terminal, and a hypervariable region in the middle. A phylogenetic tree showed that the *csd* haplotypes from *A. laboriosa* fell into one clade with those from *A. dorsata*, and were separated from those from the other two species, *A. mellifera* and *A. cerana*. The network map also showed that the *csd* haplotypes from *A. laboriosa* and *A. dorsata* are well mixed among each other, and do not form two separate branches. Pairwise *Fst* analysis revealed that the value between *A. laboriosa* and *A. dorsata* was very low (0.098), confirming a close relationship to exist between them.

INTRODUCTION

In the honeybee, sex is determined by heterozygosity at a single locus (the Sex Determination Locus, SDL) which carries the complementary sex determiner (*csd*) gene (Gempe et al., 2009). Bees with a heterozygous *csd* combination are female, those with a hemizygous copy (haploid, unfertilized eggs) or with a homozygous combination are male. When *csd* genes are homozygous, a honeybee will develop into a diploid male, which is consumed at the larval stage by workers a few hours after it hatches (Woyke, 1963). Therefore, honeybees are under great selective pressure to become heterozygous (Liu et al., 2011). This complementary sex determination theory remained a hypothesis for a long time. About a decade ago, Beye et al. (2003) cloned the *csd* gene from *Apis mellifera* by positional cloning. They found that no transcription differences existed between the two sexes, but suppression of *csd* in females with double stranded RNA for *csd* resulted in male phenotypes. The exon 6 to exon 9 region of the *csd* gene has an R (arginine)- and S (serine)-rich domain in the middle and a P (proline)-rich domain at its C terminus, which is a potential splicing factor. Between these two domains is a hypervariable region that differs significantly between alleles, and has a variable region of short repetitive sequences (Beye et al., 2003; Cho et al., 2006; Wang et al., 2012).

Previous studies have shown that the *csd* gene has evolved under balancing selection in three honeybee species (*A. mellifera* L., *A. cerana* F. and *A. dorsata* F.), and several parts of the coding region are possible targets of selection (Charlesworth, 2004; Hasselmann & Beye, 2004, 2006; Cho et al., 2006; Hasselmann et al., 2008). In addition, the level of polymorphism is approximately seven times higher in the *csd* coding region than in the neutral regions (Hasselmann et al., 2006).

More recently, Liu et al. (2011) analyzed the polymorphism of the *csd* gene in the dwarf honey bee, *A. florea* F., and found a higher level of polymorphism of the *csd* gene in this species than in the other *Apis* species so far examined. Moreover, they found that the *A. dorsata csd* gene showed no founder effect between geographical population groupings from the Guangxi and Hainan provinces of China (Liu et al., 2012).

Although much evolutionary research has been conducted on the *csd* gene in *A. mellifera*, *A. cerana*, *A. dorsata* and *A. florea*, no research had hitherto been performed on the *csd* gene of the Himalayan or Black Giant honeybee, *A. laboriosa* Smith, possibly a high altitude subspecies of *A. dorsata* sensu stricto (s.s.), and with which it may share little or no contemporary gene flow (Engel, 1999).

In the present study, we analyzed the polymorphism of the *csd* gene of *A. laboriosa*, which has the biggest body size among the nine *Apis* species. This species *A. laboriosa* is widespread, mainly in Nepal, Bhutan, North-East of India, Southern Tibet, Western and Southern of Yunnan province in China. In light of the dearth of molecular biological information on *A. laboriosa* and its sex determination system, we believed it to be worth investigating whether the *csd* gene is also polymorphic in

---

* Corresponding author.

TABLE 1. Nucleotide diversity (π) and nucleotide polymorphism (θ) of the coding regions of *csd* alleles in four *Apis* species.

Species	N[1]	L[2]	Mean ± SD % of π[3]	Mean ± SD % of θ[4]
A. laboriosa	13	503	5.669 ± 0.739 A	5.702 ± 2.204 A
A. dorsata	31	407	6.475 ± 0.338 Ac	7.626 ± 0.685 Ac
A. mellifera	69	431	4.747 ± 0.135 Ab	7.389 ± 1.984 A
A. cerana	49	446	3.546 ± 0.170 B	4.978 ± 0.500 Ab

Data for *A. laboriosa* was from this study, while data for species other than *A. laboriosa* were derived from Cho et al. (2006) and Hasselmann et al. (2008). [1]N– the sequence number used for analysis; [2]L– the sequence length excluding alignment gaps; [3]π– in this column, means followed by different uppercase letters differ significantly at $P < 0.01$ by two-tailed Z-test; [4]θ in this column, means followed by different lowercase letters differ significantly at $P < 0.05$ by two-tailed Z-test.

this species, as it appears to be in the other three *Apis* species previously tested. In *A. mellifera*, the *csd* gene contains nine exons, which form three clusters separated by two large introns (Beye et al., 2003). The genomic region of the third cluster (from exon 6 to 9, region 3) has the highest level of polymorphism compared with the other two regions. Therefore, we chose this region to study in terms of potential polymorphism.

MATERIAL AND METHODS

Sample collection

A. laboriosa samples were collected from Yunnan province, China. A total of 30 workers from five different colonies were sampled, with each colony providing six workers. The samples were initially collected into 95% ethanol and then stored deep frozen at –70°C until molecular DNA analysis.

DNA extraction

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

PCR and sequencing

The primers used in this study for amplifying region 3 of the *A. laboriosa csd* gene were the same as reported by Cho et al. (2006). The high fidelity LA Tag DNA polymerase (BEST ALL-HEAL LLC, NY, USA) were 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 50°C for 30 s and extension at 72°C for 2 min, with a final extension at 72°C for 10 min. PCR products were purified using DNA GEL EXTRACTION kits (Sangon, Shanghai, China) and cloned into the pEASY-T3 vector (Transgene, Beijing, China). To obtain as many *csd* alleles as possible, the genomic region 3 of the *csd* gene was cloned from cephalothorax of each sampled worker bee, and 1–3 clones of each cloned fragment were subjected to double-strand sequencing. Single-sequencing reads were assembled using the Seqman program in the DNAstar software (Burland, 2000).

Sequence analysis

After assembly, the vector sequences were removed from the assembled raw sequences, whereafter all the *csd* sequences obtained were compared with each other by sequence alignment to remove repeated sequences, the remaining sequences being different haplotypes of genomic region 3 of *csd* gene, which were then used for further analysis. The exons, introns and coding regions of these sequences were determined by consulting those previously determined for the genomic region 3 of the *csd* gene of *A. mellifera*, *A. cerana* and *A. dorsata* as reported by Cho (Cho et al., 2006) and cDNA sequences for the

same three honeybee species as reported by Hasselmann et al. (2008). Since the presently determined sequences for *A. laboriosa* are genome sequences containing exons and introns, it was easy to identify these regions by aligning our sequences with the cDNA sequences for the *csd* gene reported by Hasselmann et al. (2008). Coding sequences of *A. mellifera*, *A. cerana* and *A. dorsata csd* genomic region 3 (n = 228 sequences) published by Cho et al. (2006) and Hasselmann et al. (2008) were downloaded from Genbank under accession numbers DQ324946-DQ325026, DQ325038-DQ325133 and EU100885-EU100935. Nucleotide sequence alignments were performed using Clustal X version 1.8 (Thompson et al., 1997), and alignment results were adjusted manually for obvious alignment errors.

DAMBE 4.1.19 (Xia & Xie, 2001) was used to identify haplotypes. 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) was calculated by using the DNAsp 5.0 program (Librado & Rozas, 2009). Pair-wise *Fst* distances were computed using ARLEQUIN 3.1 software (Excoffier et al., 2005), whilst phylogenetic trees were constructed using MEGA version 4.0 (Tamura et al., 2007). Two tailed Z test was adopted to detect significant difference between two π values. Kimura's 2-parameter genetic distance was calculated by MEGA 4.0. A Median joining (MJ) network was drawn using the program Network 4.5 (Bandelt et al., 1999); briefly, an "rdf" file based on all the haplotypes was generated using DNAsp 5.0, it was then used for network calculation using Network 4.5 with the median joining method, and the final network map was drawn based on the calculated results.

RESULTS

Polymorphism of *A. laboriosa csd* haplotypes

Following cloning and sequencing of the genomic region 3 of the *csd* gene from *A. laboriosa*, we obtained 13 haplotypes (GeneBank accession number: JX393059-JX393071). We compared the nucleotide diversity (π) of the *csd* coding region of *A. laboriosa* with that of other three *Apis* species, including *A. mellifera*, *A. cerana*, *A. dorsata* s.s. As shown in Table 1, the *csd* gene has a high level of polymorphism in all three true species as well as the putative subspecies *A. laboriosa* as here studied. The polymorphism of the *A. laboriosa csd* gene was significantly higher than the polymorphism of *A. cerana csd* gene (two-tailed Z-test, $P < 0.01$), but failed to be significant compared with π values of the other two full species, including *A. dorsata* s.s.

```
Alcsd-1    SSLSNNYNYK-NYNN-YNNYN--NKHYN-----NYNNKHYNNKHYKQYYNINYIE
Alcsd-9    SSLSNNYNYK-NYN---------NKHYN-----KHYNKHYNN---KQYYNINYIE
Alcsd-6    SSLSNNYNYK-NYNNNYNNY---NKHYN-----KHYNKHYN-HYKQYYNINYIE
Alcsd-4    SSLSNNYNYK-NYNNNYNNY---NKHYNNKHYNKHYNKHYN-HYKQYYNINYIE
Alcsd-7    SSLSNNYNYK-NYNNNYNNY---NKHYN-----KYYNKHYNK-HYKQYYNINYIE
Alcsd-12   SSLSNNYNYK-NYNSNYNNY---NKHYN-----KHYNKHYN-HYKQYYNINYIE
Alcsd-3    SSLSNNYNYNNNYNN--NNYN--NNNYK------HYNKHYNK---HLYYNINYIE
Alcsd-10   SSLSNNYNCN-NNNYNNNNYEHYYKHYN------HYNKH-NHYNKHLYYNINYIE
Alcsd-5    SSLSNNYSCN-NNNY--EHYYKHYNHYN------HYNKHYNHYNK-LYYNINYIE
```

Fig. 1. Amino acid sequence alignments of *A. laboriosa* csd alleles. "−" indicates alignment gaps.

Hypervariable region of *A. laboriosa csd* proteins

We analyzed the amino acid sequences of the *csd* haplotypes in *A. laboriosa*, since this region contains a hypervariable region critical for determining the specificity of *csd* alleles. The exons, introns and coding sequences on the *csd* haplotypes obtained were determined by consulting sequences of the *A. mellifera csd* gene as reported by Cho et al. (2006) and Hasselmann et al. (2008). The hypervariable region is rich in asparagine (N) and tyrosine (Y), and form basic $(N)_{1–4}Y$ repeats. Besides $(N)_{1–4}Y$ repeats, the $(KHYN)_{1–5}$ repeats are also an important type of motif existing in all the haplotypes studied (Fig. 1).

Phylogeny of all the *csd* haplotypes

A genealogical tree was constructed based on all the haplotypes from *A. cerana*, *A. dorsata* and *A. mellifera* and *A. laboriosa*. As shown in Fig. 2, all the *csd* haplotypes form three clades, haplotypes from *A. mellifera* and *A. cerana* form separate clades respectively, while sequences from *A. dorsata* and *A. laboriosa* do not form two branches but rather, the two groups are well mixed with each other.

A median joining network was constructed based on the 162 haplotypes from the four honeybee species to investigate the possible relationships among them (Fig. 3). The network again showed that haplotypes from *A. mellifera* and *A. cerana* independently form a clade, whereas those from *A. dorsata* s.s. and *A. laboriosa* were intermixed.

Genetic differentiation between *A. laboriosa* and other three honeybee species

We calculated the pairwise *Fst* Genetic distance between *A. laboriosa* and the other three honeybee species. Value indicates the level of genetic differentiation between two species. In general, *Fst* < 0.05 indicates a low level of genetic differentiation, 0.05 < *Fst* < 0.15 a middle genetic differentiation, 0.15 < *Fst* < 0.25 a high level of genetic differentiation, and *Fst* > 0.25 an extremely high level of genetic differentiation. The *Fst* value between *A. laboriosa* and *A. cerana* were found to be the highest (0.539), while that between *A. dorsata* and *A. laboriosa* were the lowest (0.0987). The *Fst* value between *A. laboriosa* and *A. mellifera* was 0.515.

DISCUSSION

Previous studies have shown that the *csd* genes in *A. mellifera*, *A. cerana* and *A. dorsata* s.s. have a high level of polymorphism (Cho et al., 2006; Hasselmann et al., 2008). In our study, we found that the *csd* gene in *A.*
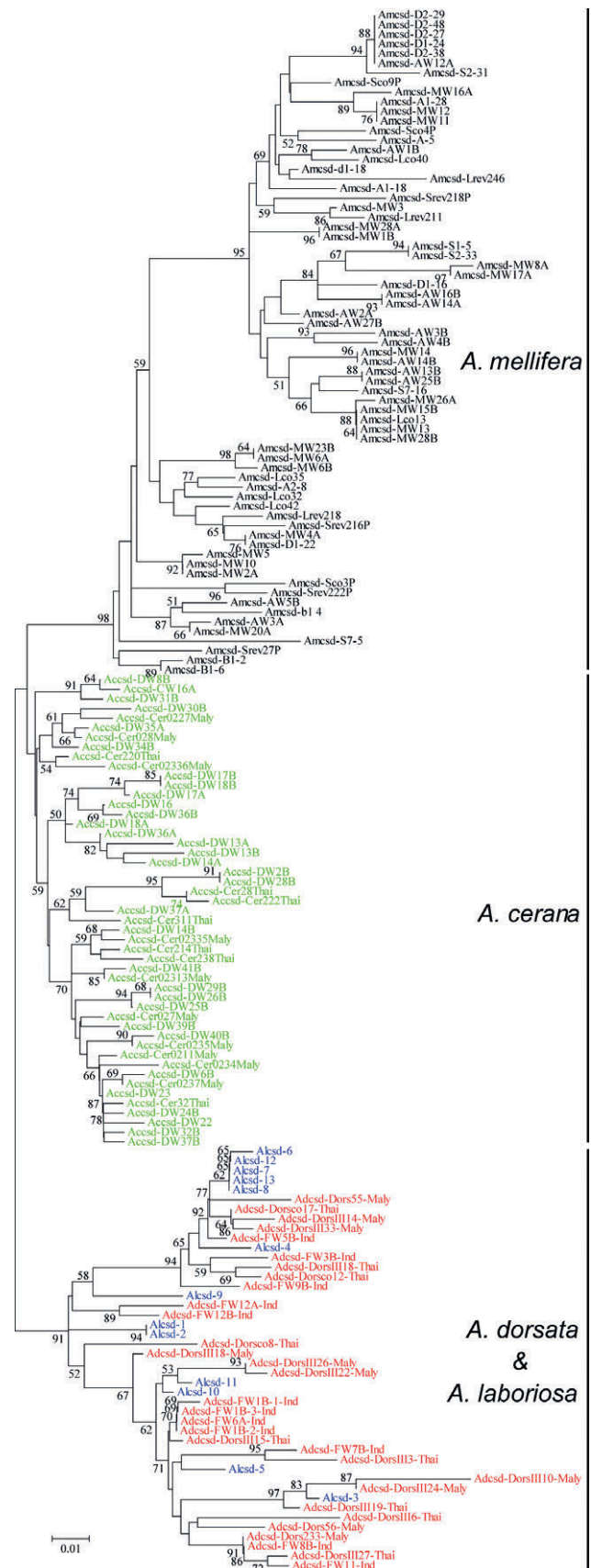

Fig. 2. The gene genealogy of csd haplotypes in region 3 of four species. The minimum evolution method and Kimura's two parameter distances are used to construct the tree. Bootstrap percentages are shown on internal branches. The scale bar represents the number of nucleotide changes per site. Accsd – *A. cerana*; alcsd – *A. laboriosa*; adcsd – *A. dorsata*; amcsd – *A. mellifera*.
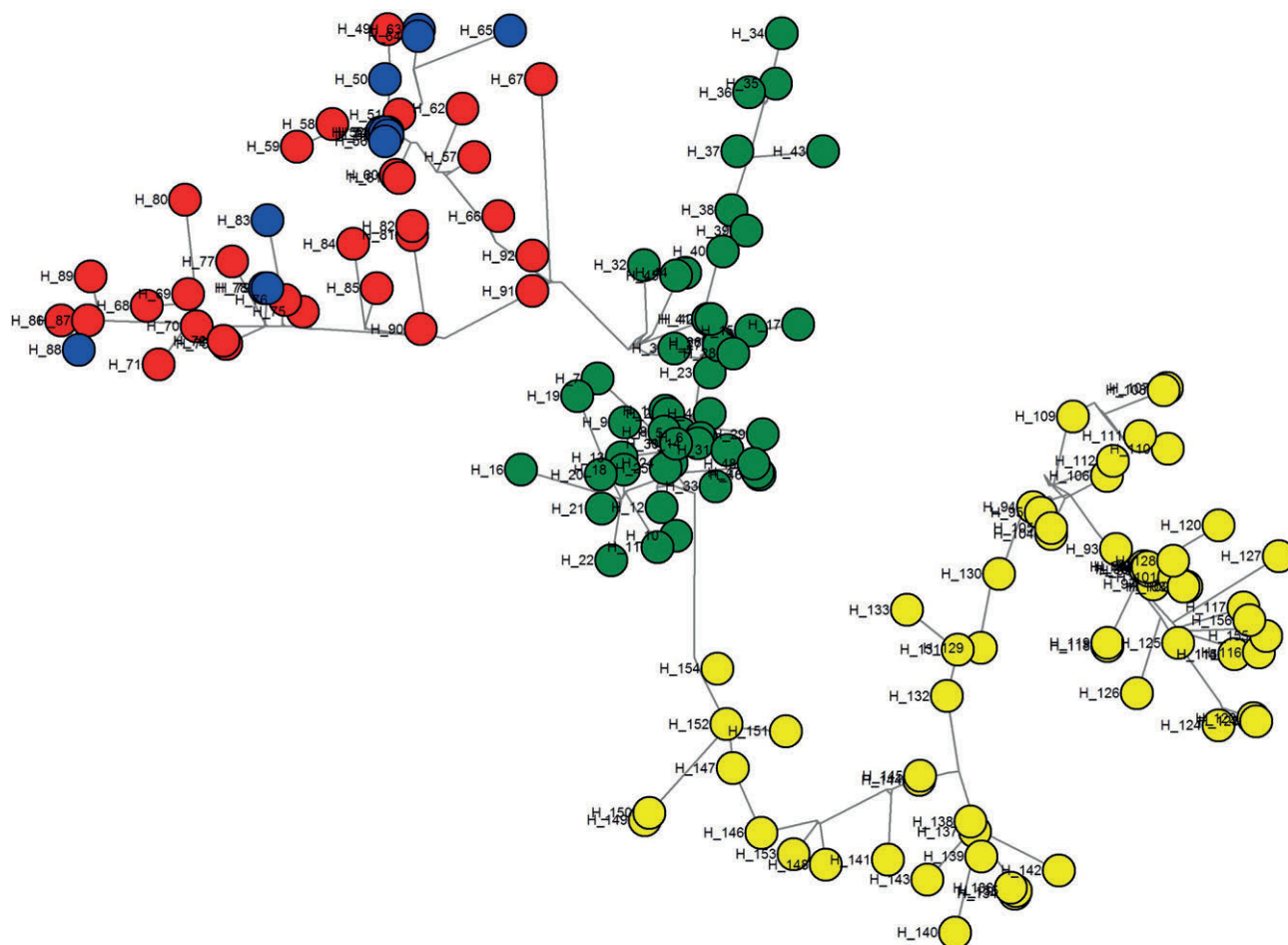
Fig. 3. The median joining network of 162 haplotypes from four species. Circles denote haplotypes unique to *A. laboriosa* (blue), *A. mellifera* (yellow), *A. cerana* (green) or unique to *A. dorsata* (red).

laboriosa also shows a high level of polymorphism. This result further confirmed that the complementary sex determination mechanism is common for all the honeybee species so far examined (Cook, 1993).

Previous studies have shown that the $(N)_{1-4}Y$ repeats and $(KHYN)_{1-5}$ repeats are two types of important repeat sequences in the hypervariable region of the *csd* protein, which most likely plays a key role in determining the specificity of alleles. Interestingly, the $(KHYN)_{1-5}$ motifs exist in all the species investigated, except in *A. mellifera* (Hasselmann et al., 2008). Thus whilst this kind of repeat still exists in the ancient *csd* alleles, they have apparently been lost in *A. mellifera* at sometime during its evolution, perhaps during its intense domestication over the past 7000 years or so (Bloch et al., 2010).

A. laboriosa was named by F. Smith in 1871, according to morphological differences observed between *A. dorsata* s.s. and this putative species. In 1980, Sakagami found that these two species have significant difference in more than 100 morphological features. Arias & Sheppard (2005) have divided ten known honeybee species into three major clusters based on *ND2* mitochondrial gene and *EF1*-α intron. These are giant bees *A. dorsata*, *A. laboriosa* and *A. binghami* (author), dwarf bees *A. andreniformis* and *A. florea*, and the cavity-nesting bees

A. mellifera, *A. cerana*, *A. koschevnikovi* (author), *A. nuluensis* (author), and *A. nigrocincta* (author) (Arias & Sheppard, 2005). Recent studies of phylogenetic relationships within the genus *Apis* also suggest that *A. dorsata* s.s. and *A. laboriosa* are in fact two separate species (Arias & Sheppard, 2005; Raffiudin & Crozier, 2007; Lo et al., 2010), although other researchers have failed to find obvious differences between the genitalia of drones in these two groups (McEvoy & Underwood, 1988). Thus, many scholars have questioned the classification status of *A. laboriosa*. In our study, the phylogenetic tree and the network of the *A. dorsata* s.s. and *A. laboriosa* *csd* haplotypes clearly reveal that the latter form is not a separate species, but rather, is probably an ecologically-divergent (altitude-related) subspecies, more especially because the genetic differentiation between these groups is very low (these results). Probably the split of *A. dorsata* s.s. and *A. laboriosa* is a very recent event, while the divergence time of the *csd* gene is older than the divergence time between these two groups, leading to a trans-specific polymorphism of the *csd* gene. To date, trans-specific polymorphism of alleles has been reported in several other organisms, including the major histocompatibility complex (MHC) of jawed vertebrates (Klein, 1987; Hughes & Nei, 1988; Takahata & Nei, 1990; Taka-

hata et al., 1992), the self-incompatibility *S* locus of plants (Ioerger et al., 1990; Richman et al., 1996; Charlesworth & Awadalla, 1998), and the *het-c* heterokaryon incompatibility locus of fungi (Wu et al., 1998; May et al., 1999). The *MHC* and *S* alleles have been used to analyze population genetic structure and infer relative coalescence times and population history in many organisms (Takahata, 1990; Yuhki & O'Brien, 1990; Richman et al., 1996; Miller & Withler, 1997; Richman & Kohn, 1999; Bos et al., 2008). If trans-specific polymorphism of the *csd* gene really exists between the *A. dorsata* s.s. and *A. laboriosa*, the *csd* gene can be used in the future to determine the historical population size and study the speciation process under eusocial organization.

In conclusion, we found a high level of polymorphism of the *csd* gene in the black giant honeybee, *A. laboriosa* following molecular biological analysis. However, subsequent phylogenetic tree and network analysis of the sequences obtained revealed that *A. dorsata* s.s. and *A. laboriosa* are mixed together and are not separable by such means, thereby providing further convincing evidence, here molecular biological, that the latter is indeed a subspecies of the former.

REFERENCES

ARIAS M.C. & SHEPPARD W.S. 2005: Phylogenetic relationships of honey bees (Hymenoptera: Apinae: Apini) inferred from nuclear and mitochondrial DNA sequence date. — *Mol. Phylogenet. Evol.* **37**: 25–35.

BANDELT H.J., FORSTER P. & RÖHL A. 1999: Median-joining networks for inferring intraspecific phylogenies. — *Mol. Biol. Evol.* **16**: 37–48.

BEYE M., HASSELMANN M., FONDRK M.K., PAGE R.E. & 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.

BLOCH G., FRANCOY T.M., WACHTEL I., PANITZ-COHEN N., FUCHS S. & MAZAR A. 2010: Industrial apiculture in the Jordan valley during Biblical times with Anatolian honeybees. — *Proc. Natl. Acad Sci. USA* **107**: 11240–11244.

BOS D.H., GOPURENKO D., WILLIAMS R.N. & DEWOODY J.A. 2008: Inferring population history and demography using microsatellites, mitochondrial DNA, and major histocompatibility complex (MHC) genes. — *Evolution* **62**: 1458–1468.

BURLAND T.G. 2000: DNASTAR's Lasergene sequence analysis software. — *Methods Mol. Biol.* **132**: 71–91.

CHARLESWORTH D. 2004: Sex determination: Balancing selection in the honey bee. — *Curr. Biol.* **14**: R568–R569.

CHARLESWORTH D. & AWADALLA P. 1998: Flowering plant selfincompatibility: the molecular population genetics of *Brassica* S-loci. — *Heredity* **81**: 1–9.

CHO S., HUANG Z.Y., GREEN D.R., SMITH D.R. & ZHANG J. 2006: Evolution of the complementary sex-determination gene of honeybees: Balancing selection and trans-species polymorphisms. — *Genome Res.* **16**: 1366–1375.

COOK J.M. 1993: Sex determination in the Hymenoptera: a review of models and evidence. — *Heredity* **71**: 421–435.

ENGEL M.S. 1999: The taxonomy of recent and fossil honey bees (Hymenoptera: Apidae: *Apis*). — *J. Hymenopt. Res.* **8**: 165–196.

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.

GEMPE T., HASSELMANN M., SCHIØTT M., HAUSE G., OTTE M. & BEYE M. 2009: Sex determination in honeybees: two separate mechanisms induce and maintain the female pathway. — *PLoS Biol.* **7**: e1000222.

HASSELMANN M. & BEYE M. 2004: Signatures of selection among sex-determining alleles of the honey bee. — *Proc. Natl. Acad. Sci. USA* **101**: 4888.

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.

HASSELMANN M., VEKEMANS X., PFLUGFELDER J., KOENIGER N., KOENIGER G., TINGEK S. & 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.

HUGHES A.L. & NEI M. 1988: Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. — *Nature* **335**: 167–170.

IOERGER T.R., CLARK A.G. & KAO T.H. 1990: Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. — *Proc. Natl. Acad. Sci. USA* **87**: 9732–9735.

KLEIN J. 1987: Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. — *Hum. Immunol.* **19**: 155–162.

LIBRADO P. & ROZAS J. 2009: DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. — *Bioinformatics* **25**: 1451–1452.

LIU Z.Y., WANG Z.L., WU X.B., YAN W.Y. & ZENG Z.J. 2011: *csd* alleles in the red dwarf honey bee (*Apis florea*, Hymenoptera: Apidae) show exceptionally high nucleotide diversity. — *Insect Sci.* **18**: 645–651.

LIU Z.Y., WANG Z.L., YAN W.Y., WU X.B., ZENG Z.J. & HUANG Z.Y. 2012: The sex determination gene shows no founder effect in the giant honey bee, *Apis dorsata*. — *PloS One* **7**: e34436.

LO N., GLOAG R.S., ANDERSON D.L. & OLDROYD B.P. 2010: A molecular phylogeny of the genus *Apis* suggests that the giant honey bee of the Philippines, *A. breviligula* Maa, and the plains honey bee of southern India, *A. indica* Fabricius, are valid species. — *Syst. Entomol.* **35**: 226–233.

MAY G., SHAW F., BADRANE H. & VEKEMANS X. 1999: The signature of balancing selection: fungal mating compatibility gene evolution. — *Proc. Natl. Acad. Sci. USA* **96**: 9172–9177.

MCEVOY M.V. & UNDERWOOD B.A. 1988: The drone and species status of the Himalayan honey bee, *Apis laboriosa* (Hymenoptera: Apidae). — *J. Kans. Entomol. Soc.* **61**: 246–249.

MILLER K.M. & WITHLER R.E. 1997: Mhc diversity in Pacific salmon: population structure and trans-species allelism. — *Hereditas* **127**: 83–95.

RAFFIUDIN R. & CROZIER R.H. 2007: Phylogenetic analysis of honey bee behavioral evolution. — *Mol. Phylogenet. Evol.* **43**: 543–552.

RICHMAN A.D. & KOHN J.R. 1999: Self-incompatibility alleles from *Physalis*: implications for historical inference from balanced genetic polymorphisms. — *Proc. Natl. Acad. Sci. USA* **96**: 168–172.

RICHMAN A.D., UYENOYAMA M.K. & KOHN J.R. 1996: Allelic diversity and gene genealogy at the self-incompatibility locus in the Solanaceae. — *Science* **273**: 1212–1216.

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.

TAKAHATA N. & NEI M. 1990: Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. — *Genetics* **124**: 967–978.

TAKAHATA N., SATTA Y. & KLEIN J. 1992: Polymorphism and balancing selection at major histocompatibility complex loci. — *Genetics* **130**: 925–938.

TAMURA K., DUDLEY J., NEI M. & KUMAR S. 2007: MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. — *Mol. Biol. Evol.* **24**: 1596–1599.

THOMPSON J.D., GIBSON T.J., PLEWNIAK F., JEANMOUGIN F. & HIGGINS D.G. 1997: The CLUS-TAL_X windows interface: exible strategies for multiple sequence alignment aided by quality analysis tools. — *Nucl. Acid. Res.* **25**: 4876–4882.

WANG Z.L., LIU Z.Y., WU X.B., YAN W.Y. & ZENG Z.J. 2012: Polymorphism analysis of *csd* gene in six *Apis mellifera* subspecies. — *Mol. Biol. Rep.* **39**: 3067–3071.

WOYKE J. 1963: What happens to diploid drone larvae in a honeybee colony. — *J. Apic. Res.* **2**: 73–75.

WU J., SAUPE S.J. & GLASS N.L. 1998: Evidence for balancing selection operating at the het-c heterokaryon incompatibility locus in a group of filamentous fungi. — *Proc. Natl. Acad. Sci. USA* **95**: 12398–12403.

XIA X. & XIE Z. 2001: DAMBE: software package for data analysis in molecular biology and evolution. — *J. Hered.* **92**: 371–373.

YUHKI N. & O'BRIEN S.J. 1990: DNA variation of the mammalian major histocompatibility complex reflects genomic diversity and population history. — *Proc. Natl. Acad. Sci. USA.* **87**: 836–840.