

OP-054 [Bee Biology]

Genetic diversity and population structure of *Apis cerana* complementary sex determiner (*csd*) genes in China

Chao Chen, Yufei Zou

State Key Laboratory of Resource Insects, Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing, China

In honeybees, fertilized eggs with homozygous complementary sex determiner (*csd*) alleles develop into sterile diploid males, which are eliminated by workers during the larval stage. A decrease in *csd* allele diversity leads to increased inbreeding and a higher occurrence of diploid males, significantly decreasing the population growth rate and effective population size. Therefore, understanding and maintaining *csd* gene diversity is crucial. This study aims to analyze the genetic diversity and population structure of *Apis cerana* *csd* genes in China. We sequenced the exon 6-8 region of the *csd* gene in 250 workers across 13 geographical populations, identifying 234 nucleotide haplotypes and 141 unique amino acid sequences, 89 of which were newly discovered. The *csd* gene in these populations shows high genetic diversity comparable to that of *A. mellifera*, however, distribution of alleles did not show clusters based on geographic regions. Analyses of genetic differentiation showed variable genetic distance among populations. Our results demonstrate high diversity of *csd* genes of *Apis cerana* populations across China, suggesting low risk of inbreeding in the populations.



Genetic diversity and population structure of *Apis cerana* complementary sex determiner (*csd*) genes in China

Institute of Apiculture, Chinese Academy of Agricultural Sciences

Chao Chen

9/2025

CONTENTS

1 Introduction

2 Technical Route

3 Results

4 Future research direction

01 Introduction

Background

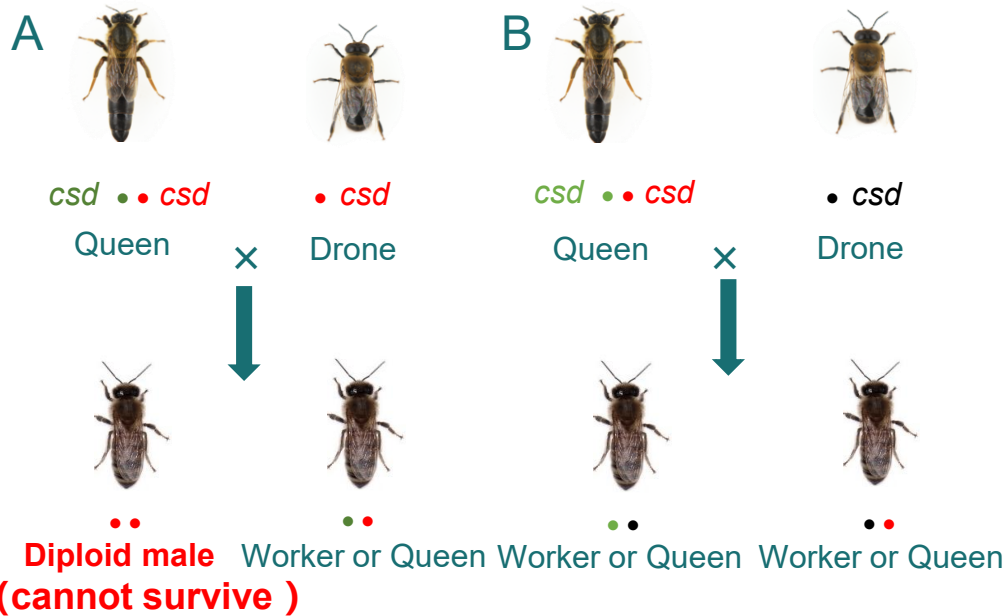


Figure 1. Diagram of sex determination in honeybees

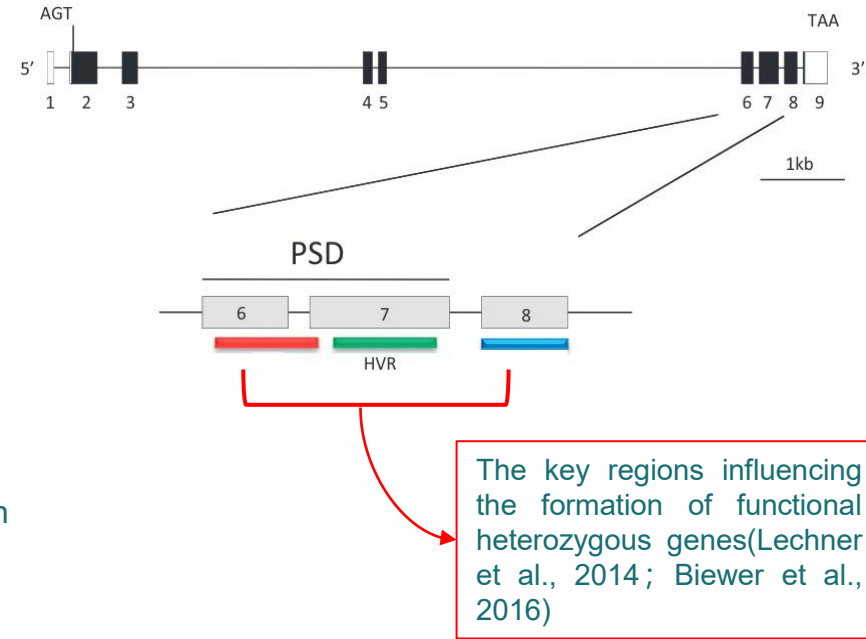
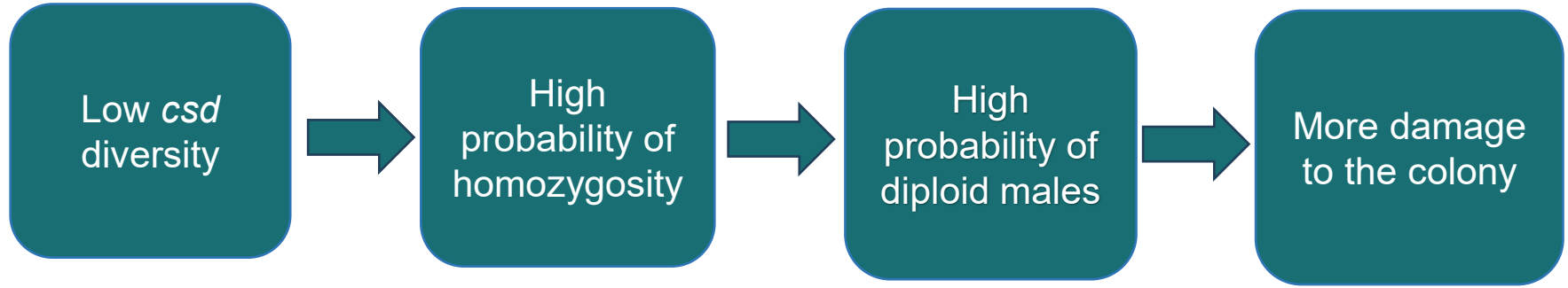


Figure 2. Structure of the *csd* gene
(Cited from Lechner et al., 2014)

- The *csd* gene is the sex determining gene in honeybees. When the *csd* locus is heterozygous, the individual develops as female; **if the *csd* locus is homozygous, it develops into a diploid male**, which cannot survive. Additionally, **only functionally heterozygous *csd* alleles can trigger female development**.

Background



- Therefore, it is necessary to monitor and manage the *csd* gene alleles

Research status

The study of *csd* gene diversity in *A. mellifera* is relatively well-developed.

Genetic diversity of *Apis mellifera* complementary sex determiner (*csd*) in four Australian breeding populations

Erica Y. Y. Mo¹, Patsavee UTAIPANON^{1,2}, Tiffane BATES³, John DAVIES⁴, Catherine E. GRUEBER⁵, Julianne LIM¹, Elizabeth A. FROST^{2,6}, and Nadine C. CHAPMAN¹

¹ School of Life and Environmental Sciences, University of Sydney, Life, Earth and Environmental Science Building F22, Sydney, NSW 2006, Australia

² NSW Department of Primary Industries, Paterson, NSW 2421, Australia

³ School of Molecular Sciences, University of Western Australia, Crawley, WA 6009, Australia

⁴ Better Bees WA Inc, Baldvis, WA 6171, Australia

⁵ School of Life and Environmental Sciences, University of Sydney, Heydon-Laurence Building A08, Sydney, NSW 2006, Australia

⁶ Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 2350, Australia

Received 26 June 2023 – Revised 17 April 2024 – Accepted 3 May 2024

Abstract – Honey bee (*Apis mellifera*) breeding has intensified as populations experience increasing stress and

according to their subspecies origin, which suggested either an

Genetic diversity of *Apis mellifera* complementary sex determiner (*csd*) in four Australian breeding populations

Genetic diversity and population genetic structure analysis of *Apis mellifera* subspecies in Algeria and Europe based on complementary sex determiner (*CSD*) gene

Tiffane BATES³, John DAVIES⁴, Elizabeth A. FROST^{2,6}, and Nadine C. CHAPMAN¹

Life, Earth and Environmental Science Building F22, University of Sydney, Sydney, NSW 2006, Australia

Guillaume GREGORI
CANALE-T

Department of Genetics
Boudiaf, BP 150,
SNV, University
recherche pour l'A
anie, Toulouse, F
(AP), UMT PrAD
riment, UMT
rf, F-Engels-Str
ia and University
Maribor, Slovenia

ed 14 January 21

etermination de
lelic diversity. I
variable region
estry bees. We
ies) and 183 fr
ividuals of unk
o 119 protein ha
ls were compar
ersity at the loc

RESEARCH ARTICLE

New insights into the criteria of functional heterozygosity of the *Apis mellifera* complementary sex determining gene – Discovery of a functional allele pair differing by a single amino acid

Robert Mroczek¹, Agnieszka Laszkiewicz¹, Pawel Blazej², Kinga Adamczyk-Weglarzy¹, Joanna Niedbalska-Tarnowska¹, Malgorzata Cebrat^{1,*}

¹ Laboratory of Molecular and Cellular Immunology, Ludwik Hirsztfeld Institute of Immunology and Experimental Therapy Polish Academy of Sciences, Wrocław, Poland, ² Department of Bioinformatics and Genomics, Faculty of Biotechnology, Wrocław University, Wrocław, Poland

* malgorzata.cebrat@hirsztfeld.pl

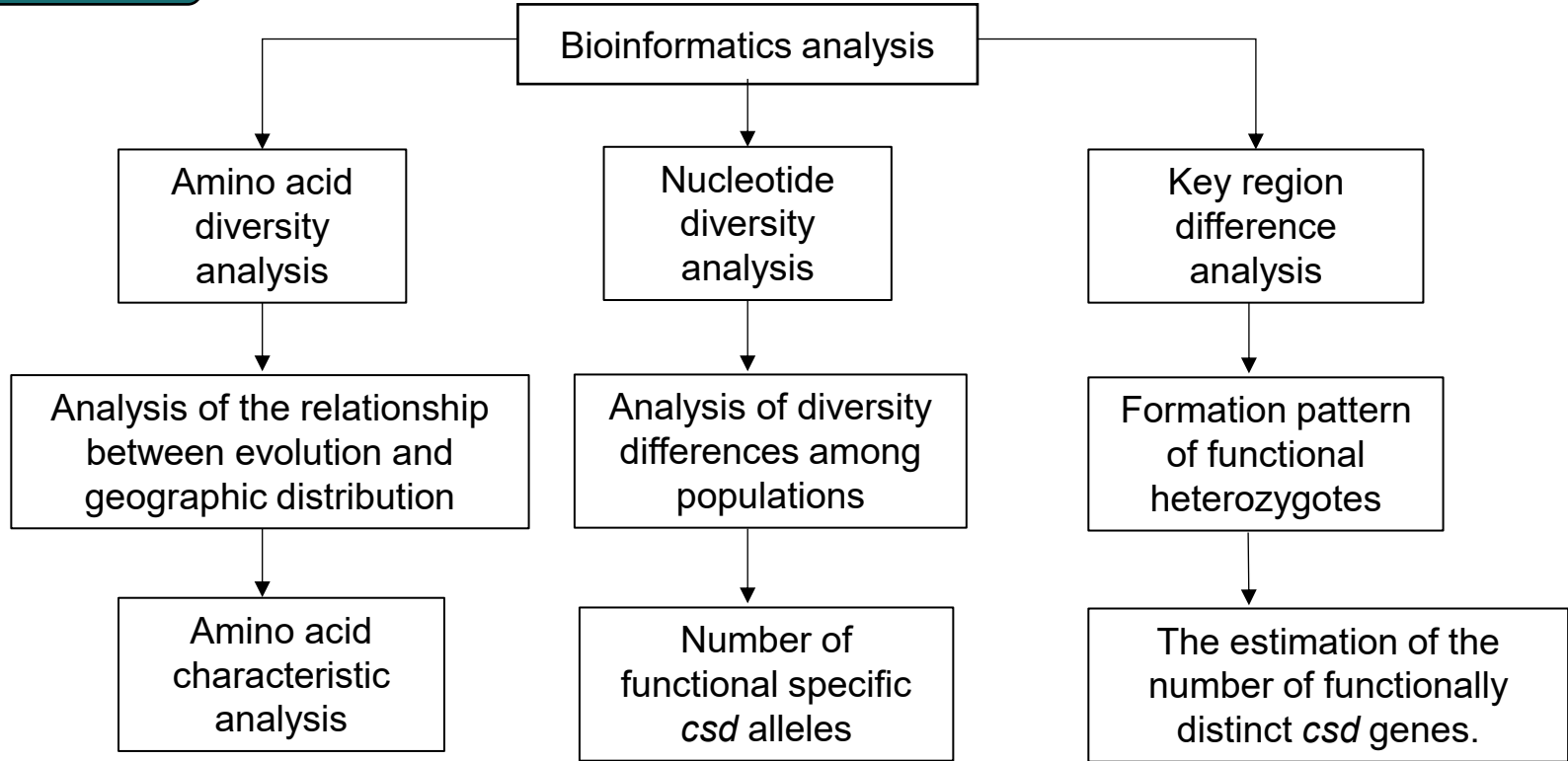
Abstract

The complementary sex determiner (*csd*) gene is responsible for controlling the sex-determination molecular switch in western honey bees (*Apis mellifera*): bees that are heterozygous for *csd* develop into females, whereas bees that are hemizygous or homozygous develop into males. The homozygous diploid males are destroyed at an early stage of the development. It has been proposed that the minimal number of amino acid differences

The study of *csd* gene diversity in *A. cerana* is relatively limited and has been confined to a few regions. A comprehensive understanding of the *csd* gene diversity in *A. cerana* across China has yet to be achieved. Based on this, we conducted the present study.

02 Technical Route

Technical Route



03 Results

Sampling Locations

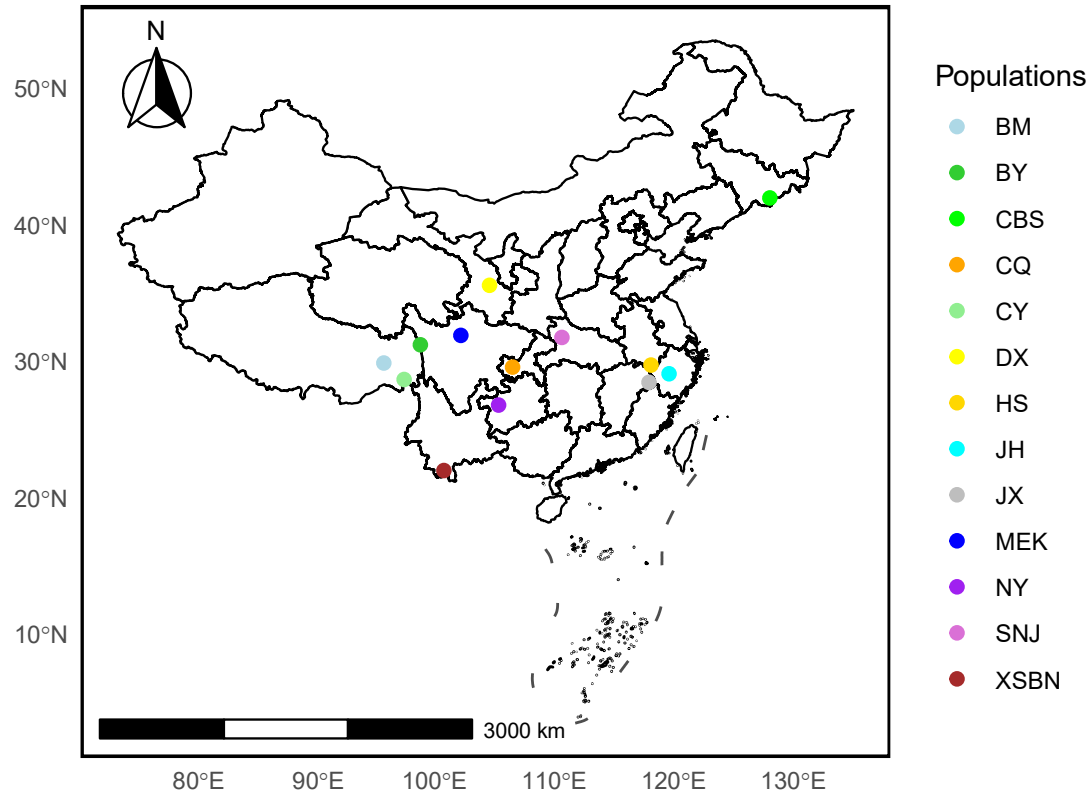


Figure 3. Location of *A. cerana* sampling sites.

Number of *csd* amino acid sequences.

A total of **234**
unique *csd*
haplotypes and
141 **distinct**
amino acid
sequences
were obtained.

It indicates that natural populations
possess **rich *csd* allele diversity**

Structural features of the *csd* amino acid sequence.

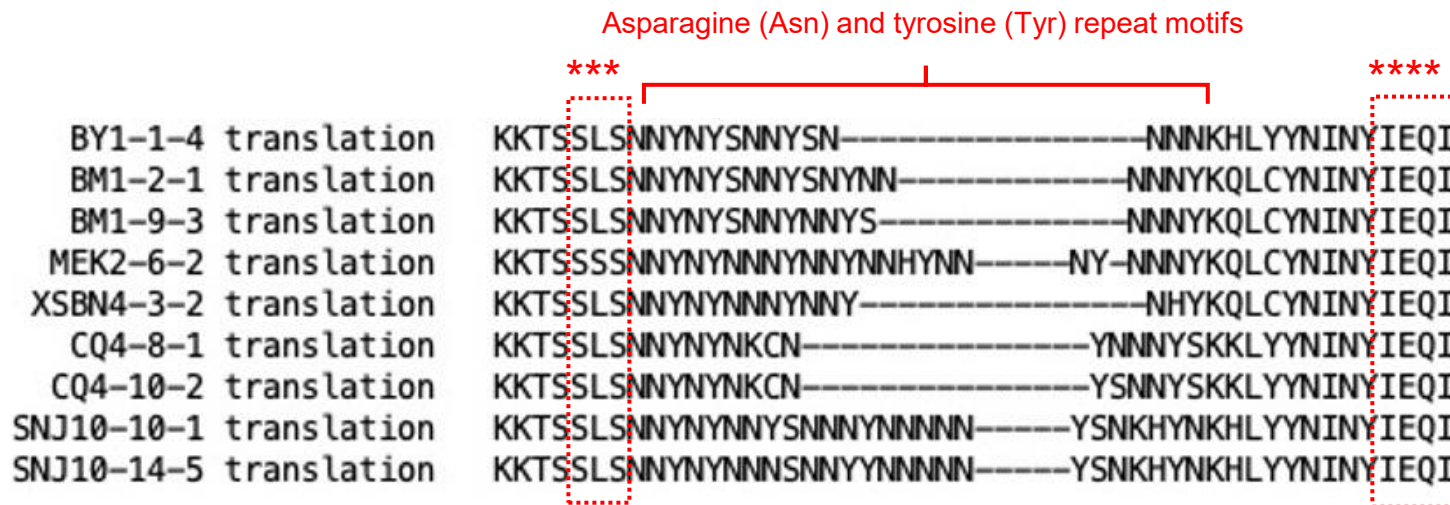


Figure 4. Representative alignment of amino acid sequences in the hypervariable region (HVR) of the *csd* genes.

- The structural features are highly consistent with those of the *A.mellifera*.

Geographic distribution of *csd* gene

Population-specific *csd* alleles.

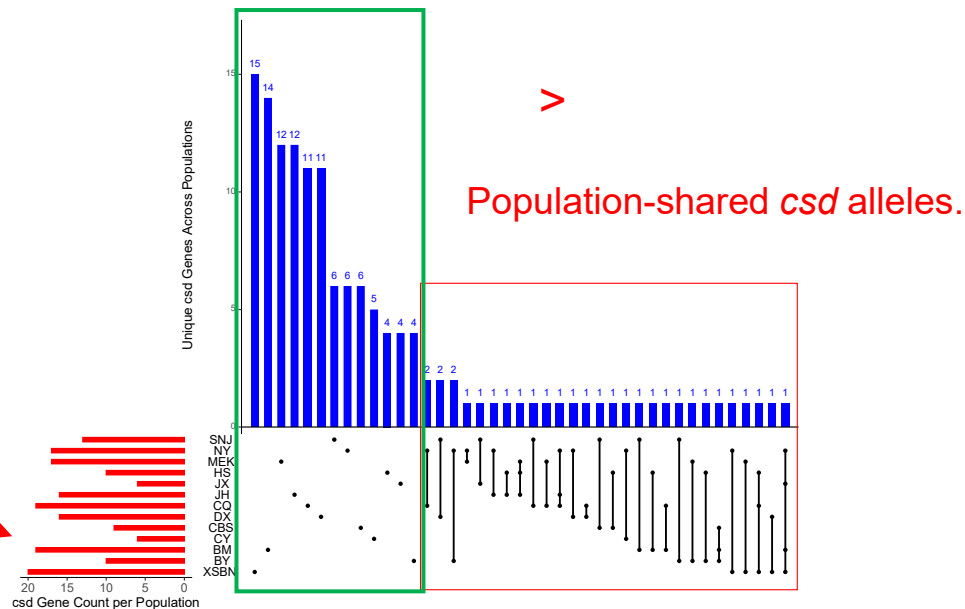
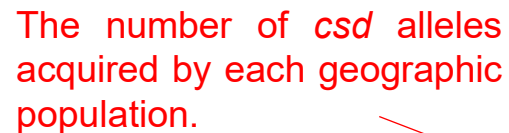


Figure 5. UpSet plot of *csd* alleles from 13 geographical populations.

Several *csd* alleles are shared among multiple populations, but most alleles are region-specific, with overall low overlap among populations.

Geographic distribution of *csd* gene

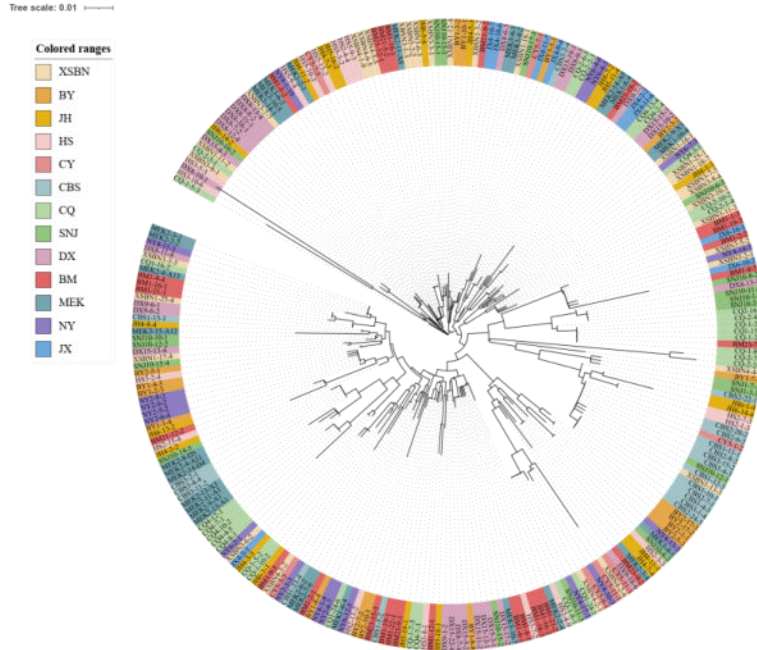


Figure 6. Maximum likelihood phylogenetic tree of *csd* amino acid sequence haplotypes

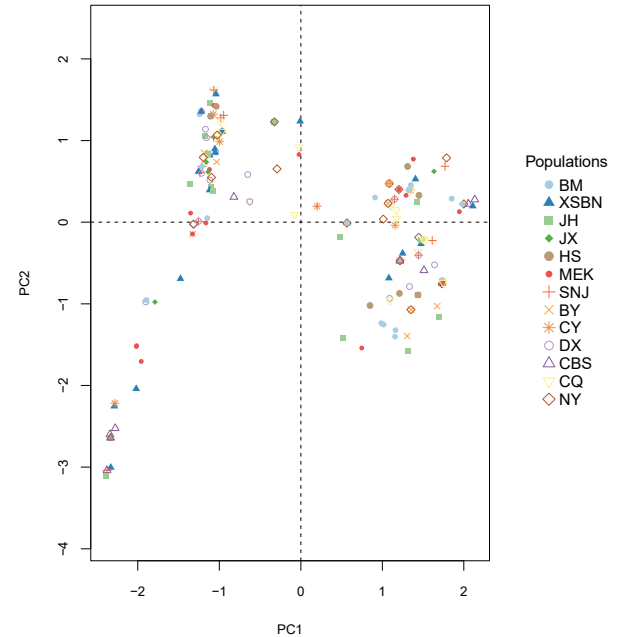


Figure 7. Principal component analysis(PCA) of *csd* amino acid haplotypes

The *csd* alleles from different geographic regions do not form a clear distribution or clustering based on geographic origin; instead, they are intermixed.

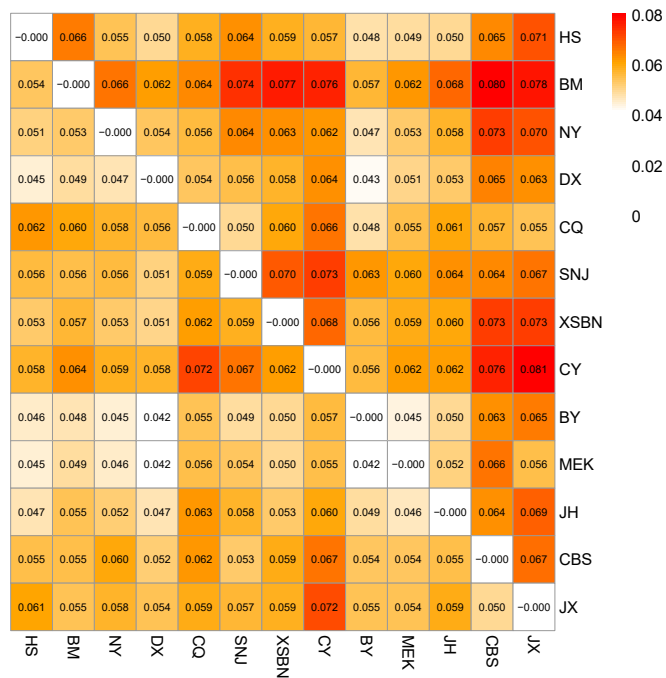
Nucleotide diversity of the *csd* gene

Table 1. Nucleotide polymorphism of *csd* in the exon regions 6-8 from 13 *A.cerana* geographic populations.

种群	单倍型	单倍型多样性	可变位点	核苷酸多样性	平均核苷酸差异
Populatio	Halotypes	hd	Variable sites	Nucleotide polymorphism	Average Number of Nucleotide Differences
XSBN	29	0.991	146	0.0663	34.85
JX	11	0.974	80	0.0608	33.82
CQ	28	0.976	123	0.0546	27.19
BM	28	0.979	120	0.0721	29.35
DX	27	0.985	137	0.0456	23.19
HS	17	0.974	112	0.0540	25.37
SNJ	19	0.975	112	0.0626	32.38
BY	17	1.000	95	0.0482	24.95
MEK	31	0.995	187	0.0517	28.11
JH	20	0.982	126	0.0549	29.09
NY	20	0.995	116	0.0568	29.61
CY	7	1.000	79	0.0649	34.43
CBS	18	0.981	67	0.0400	22.26
ALL	234	0.997	187	0.0509	17.81

- This indicates that the overall genetic diversity is high, with a low risk of inbreeding.

Genetic distance and nucleotide divergence.



- The genetic distance ranges from 0.042 to 0.072.
- The greatest genetic distance occurs between CQ and CY, as well as between JX and CY.
- The smallest genetic distance occurs between BY and DX, BY and MEK, and DX and MEK.
- The nucleotide divergence (Dxy) ranges from 0.043 to 0.081.
- The smallest nucleotide divergence occurs between BY and DX, while the greatest nucleotide divergence occurs between CY and JX.

Figure 8. Genetic distance (Ds) and nucleotide divergence (Dxy) heatmap among geographic populations of *A. cerana*

Note: The bottom left indicates genetic distance (Ds), and the top right indicates nucleotide divergence (Dxy).

Summary

Summary

- The *csd* gene of the *A.cerana* in China exhibits high genetic diversity and a low risk of inbreeding.
- The evolutionary process of the *csd* gene is not significantly influenced by geographical location.
- There is a high degree of genetic differentiation of the *csd* gene among different populations.

The dataset used in this section of the study.

Table 3-1 *csd* alleles used in this study

<i>csd</i> alleles	Locations of samples	Number of alleles	References
DW13A, DW13B, DW14A, DW14B, DW16, DW36A, DW36B, DW23, DW15, DW17A, DW17B, DW18A, DW18B	Philippines	13	
DW22,DW24A,DW24B,DW37A,DW37 B,DW39B,DW40B,DW41B	Bangalore, India	8	
DW8B,CW16A	Southern China	2	(Cho et al., 2006)
DW6B, DW25B, DW26B, DW2B, DW28B, DW29B	Japan	6	
DW30B, DW31B, DW32B, DW34B, DW35A	Thailand	5	
IND1-IND15	Flores Island, Indonesia	15	
IN1-IN25	Karnataka, India	25	(Ding, 2017)
THA1-THA30	Thailand	30	

Table 3-1 *csd* alleles used in this study(continued)

<i>csd</i> alleles	Locations of samples	Number of alleles	References
BJ1-BJ19, SD1-SD27, SC1-SC22, PNG1-PNG7,A1-A7	China Vanimo, New Guinea, Papua Makila Island in the Solomon Islands	68 14 8	(Ding, 2017)
A1-A7	Cairns, Australia	7	
MEK1-MEK19	Maerkang, Sichuan, China	19	
BY1-BY14	Baiyu, Sichuan, China	12	
BM1-BM17	Bomi, Tibet, China	17	
CY1-CY6	Chayu, Tibet, China	6	
CBS1-CBS11	Changbai Mountain, China	11	
CQ1-CQ21	Chongqing, China	21	
DX1-DX20	Dingxi, Gansu, China	19	(this study)
HS1-HS17	Huangshan, Anhui, China	15	
JH1-JH19	Jinhua, Zhejiang, China	19	
JX1-JX6	Shangrao, Jiangxi, China	6	
NY1-NY20	Nayong, Guizhou, China	20	
SNJ1-SNJ13	Shennongjia, Hubei, China	11	
XSBN1-XSBN24	Xishuangbanna, Yunnan, China	24	
Cer02336Maly, Cer02335Maly, Cer02313Maly, Cer0237Maly, Cer0235Maly, Cer0234Maly, Cer0227Maly, Cer0211Maly, Cer028Maly, Cer027Maly	Darum, Borneo	10	(Hasselmann et al., 2008b)
Cer311Thai, Cer238Thai, Cer222Thai, Cer220Thai, Cer214Thai, Cer32Thai, Cer28Thai	Samut Songkhram, Thailand	7	

- Integrated existing data from the database with the data from this study.

The formation patterns of functional heterozygous *csd* allele pairs.

Functional heterozygous <i>csd</i> allele pairs	Hypervariable region amino acid differences. d_{HVR}	Flanking regions outside the hypervariable region in the PSD. d_{PSD}	Exon 8 region d_{e8}
PNG3/PNG4	6	2	1
PNG4/PNG6	18	6	0
PNG4/PNG7	9	6	0
PNG6/PNG7	12	9	0
IND1/IND2	9	5	0
IND1/IND4	18	8	0
IND1/IND6	12	9	2
IND1/IND10	11	11	2
IND2/IND12	13	6	1
IND3/IND4	19	10	2
IND3/IND5	12	1	0
IND3/IND7	3	1	0
IND3/IND8	11	7	3
IND3/IND9	10	2	0
IND3/IND11	8	10	2
IND9/IND12	6	7	3
IND12/IND13	12	10	2
IND2/IND13	12	4	2

The difference is the greatest. It is the main factor inducing female development (Lechner et al., 2014).

The difference is not significant, and similar reports have been documented in other studies.

The minimum difference in the HVR region is 3; therefore, $d_{HVR} \geq 3$ is used as the basic criterion for functionally heterozygous *csd* allele pairs.

Figure 9. Analysis of amino acid sequence differences among 208 pairs of functionally heterozygous individuals in the dataset.

Note: The 208 pairs of functionally heterozygous individuals were derived from publicly available databases and *csd* gene sequences of workers collected in this study. Only a portion of them is shown here.

The formation patterns of functional heterozygous *csd* allele pairs.

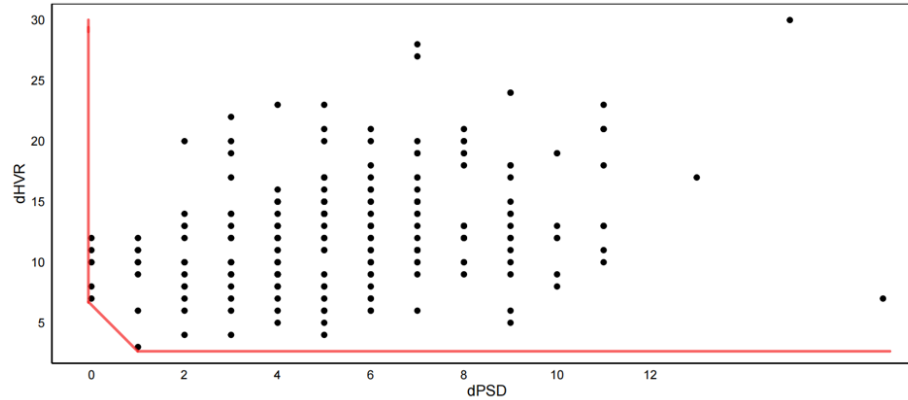


Figure 10. Plot of pairwise amino acid mismatches between d_{HVR} and d_{PSD}

- Using the lower limit of $d_{\text{HVR}} = 3$ as the basic criterion for functionally heterozygous genes.
- By analyzing the distribution of d_{HVR} and d_{PSD} and calculating their minimal boundary equation, the following critical rule encompassing all functionally heterozygous pairs sufficient to induce female development was derived:

$$d_{\text{HVR}} \geq 3 \text{ and } d_{\text{HVR}} + 4d_{\text{PSD}} \geq 7$$

Number of functionally specific *csd* alleles in the dataset.

All *csd* alleles
in the
database (317
in total)

According to the rule we inferred,
functionally equivalent alleles were
gradually removed.



207
functionally
specific *csd*
alleles

Number of functionally specific *csd* alleles in *A. cerana* worldwide.

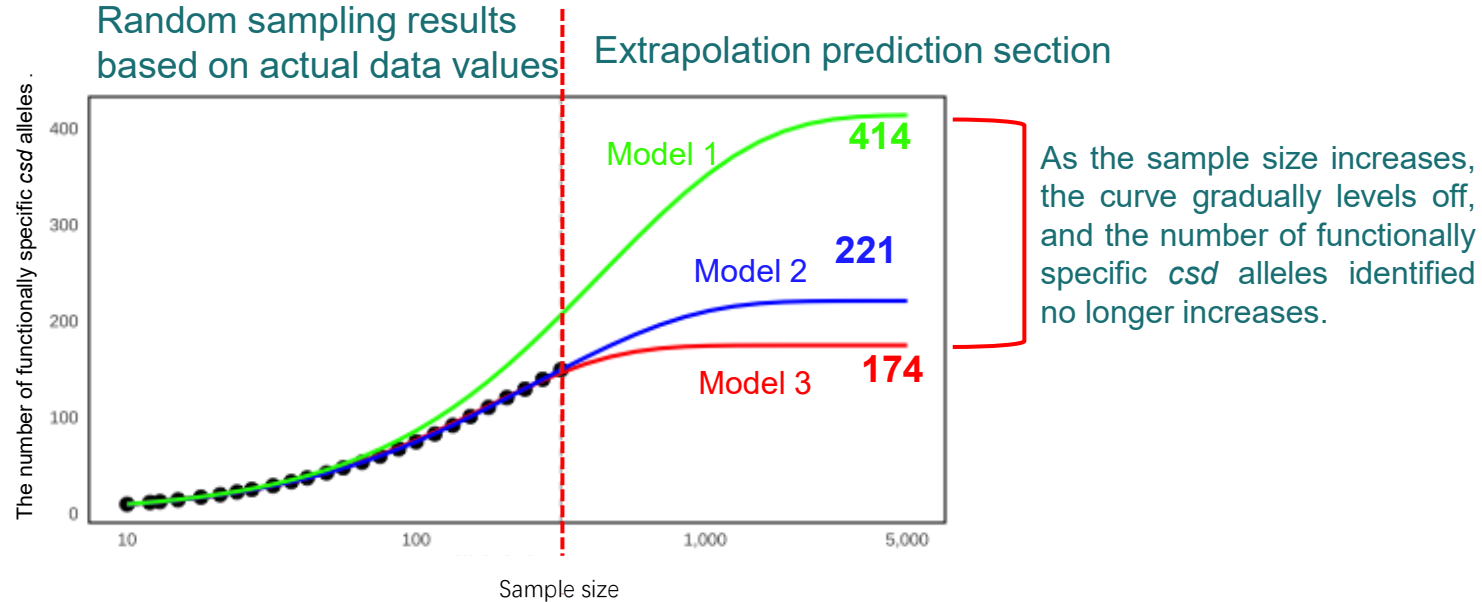


Figure 11. Plot of the mean number of functionally distinct *csd* alleles as the function of sampled *csd* alleles

- This estimate is slightly higher than the 170 to 307 reported in previous studies.
- The discrepancy may be due to the more comprehensive dataset integrated in the present study.

Summary

Summary

- We hypothesize that functionally heterozygous *csd* alleles conform to the following rule: $d_{\text{HVR}} \geq 3$ and $d_{\text{HVR}} + 4d_{\text{PSD}} \geq 7$.
- We estimate that there may be 174 to 414 functionally distinct *csd* alleles in natural populations of the *A.cerana* worldwide.

Future research direction

Future research direction

- Due to China's complex terrain, many regions have *A. cerana* populations that have not been covered. Therefore, future research should expand to more geographical areas to further reveal the adaptive changes of the *csd* gene in different environments and the gene flow between populations.
- This study has not yet verified the relationship between the functional pairing of the CSD protein and the specific regions of HVR, PSD, and exon 8 at the molecular level. This requires an in-depth exploration of the key role of the CSD protein in triggering female development, combining protein structure and function analysis.



THANK YOU

Institute of Apiculture Research, Chinese Academy of Agricultural Sciences

Chao Chen