 doi:10.20944/preprints202209.0434.v1

Original paper

Genetic relationships and signatures of adaptation to the climatic conditions in populations of *Apis cerana* based on the polymorphism of the gene Vitellogenin

Rustem A. Ilyasov [1, 2, 5, *], Slađan Rašić [3], Junichi Takahashi [4], Valery N. Danilenko [5], Maxim Y. Proshchalykin [6], Arkady S. Lelej [6], Vener N. Sattarov [7], Pham Hong Thai [8], Rika Raffiudin [9], Hyung Wook Kwon [1, *]

[1] Department of Life Sciences and Convergence Research Center for Insect Vectors, Incheon National University, 119 Academy-ro, Yeonsu-gu, Incheon 22012, Republic of Korea, E-mail: hwkwon@inu.ac.kr
[2] Bashkir State Agrarian University, Scientific and Educational Center, 50-letiya Oktyabrya str. 34, Ufa 450001, Russia, E-mail: apismell@hotmail.com
[3] Educons University, Faculty of Ecological Agriculture, Sremska Kamenica, Vojvode Putnika 87, 21208, Serbia, E-mail: rasic.sladjan@gmail.com
[4] Faculty of Life Sciences, Kyoto Sangyo University, Kamigamo Motoyama, Kita Ward, Kyoto, 603-8555, Japan, E-mail: jit@cc.kyoto-su.ac.jp
[5] Department of Genetics and Biotechnology, Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Gubkina str. 3, 119333, Russia, E-mail: valerid@vigg.ru
[6] Federal Scientific Center of the East Asia Terrestrial Biodiversity, Prospekt 100-let Vladivostoka, 159, Vladivostok, 690022, Russia, E-mail: lelej@biosoil.ru
[7] Department of Genetics, Bashkir State Pedagogical University named after M. Akmulla, 3a October Revolution street, Ufa, 450008, Russia, E-mail: wener5791@yandex.ru
[8] Vietnam national University of Agriculture and Research Center for Tropical Bees and Beekeeping, 32, Alley 125/2 Trung Kinh, Hanoi, Vietnam, E-mail: phthaihua@gmail.com
[9] Department of Biology, Bogor Agricultural University, Darmaga, Bogor, 16680, Indonesia, E-mail: rraffiudin@gmail.com

*Ilyasov, R.A. and Kwon H.W. are the co-corresponding authors

Abstract

Apis cerana and *Apis mellifera*, are very important honey species for agriculture in Asian countries. In recent decades, *A. cerana* populations have sharply declined in all Asian countries as a result of Sacbrood Virus infection and have now recovered to their original size. It can change the genetic structure of local populations of *A. cerana*. We used the nuclear gene Vitellogenin *VG* to assess the genetic structure of local populations of *A. cerana* and the signature of adaptive selection. We performed a population genetic analysis of the honey bees *A. cerana* from South Korea in comparison with *A. cerana* samples from Russia, Japan, Nepal, and China. The sequences of the gene *VG* of a closely related honey bee species, *A. mellifera*, from India and Poland were used as outgroup samples. A comparative analysis of northern and southern *A. cerana* populations was performed. The signatures of positive adaptive selection were found in the local population of *A. cerana*. We performed the Tajima's neutrality D test for *A. cerana* populations from different local populations based on the gene *VG* exon sequences. All *A. cerana* populations showed signs of population size expansion following the possible recent decline in population sizes. The local populations of *A. c. koreana* were subdivided according to their geographical distribution into southern, northern, and central Korean clusters. The gene *VG* exon sequences can be used as informative markers for monitoring the changes in genetic structure and adaptation to the environment processes in *A. cerana* populations.

Keywords: *Apis cerana*; *A. c. koreana*; population; gene vitellogenin; genetic structure; nucleotide polymorphism; adaptation

 doi:10.20944/preprints202209.0434.v1

Introduction

The Eastern honey bee, *Apis cerana*, is one of the nine *Apis* species that inhabit Asia [1]. The *A. cerana* is an important and famous pollinator of crops in Asia and a producer of honey, wax, royal jelly, and bee pollen [2]. *A. cerana* has a high potential for further genetic improvement by selective breeding based on molecular markers [3]. The range of *A. cerana* is distributed from northern to southern Asia [4]. *Apis cerana* occurs across southern and southeastern Asia up to Russia in northern Asia and extends from Japan in the east to Afghanistan in the west. It occupies a large range of climatic conditions, from cool regions in higher latitudes and altitudes to dry semi-desert environments and tropical climates. *A. cerana* has high genetic and phenotypic variations occurring across a range of spatial scales and adaptations to various climatic conditions, which led to the division into several ecotypes in the geographic cline of the north to south [3, 5, 6]. The most northern population of *A. cerana* is feral and occurs in Russia on the territories of Primorsky Krai and Khabarovsky Krai to 47° 54’ N [7]. In Korea, the *A. cerana* population is formed by the subspecies *A. c. koreana* [8, 9]. This population of *A. c. koreana* is the most interesting and important for practical and scientific studies due to its distribution. It has a transitional position between the northern and southern populations of *A. cerana* in Asia, and most of the honey bee migrations take place through the Korean peninsula [8, 9].

Recently, *A. cerana* is threatened across its native area due to the spread of the Korean Sacbrood Virus (kSBV) and the importing of *A. mellifera* [4, 10, 11]. Molecular genetic studies allow the development of basic strategies for *A. cerana* conservation. Within its native range, *A. cerana* requires conservation efforts. Deforestation, loss of nest sites, pathogens, intraspecific hybridization, and increased pesticide use have contributed to a steady decline of the *A. cerana* population [3, 8, 9]. Furthermore, the replacement of *A. cerana* management by *A. mellifera* management in many areas affects the native flora in addition to the bee population. While the *A. mellifera* colonies may be more profitable for honey production, they are not equivalent pollinators of native plants [3; 4, 12].

Previously, the sequences of the gene Vitellogenin (*VG*) were used in investigations of the *Apis mellifera* population [13]. The signatures of positive selection and evolutionary adaptation in the *A. mellifera* population were found by analyzing nucleotide sequences of the gene *VG* [14]. The gene *VG* is responsible for the development and caste differentiation in the honey bees through interaction with juvenile hormone (JH) and insulin-like growth factor (IGF) [15, 16, 17]. Hence, *VG* may likely be useful in the study of the evolutionary adaptation and structure of *A. cerana* populations. We studied populations of *A. cerana* from different Asian

 doi:10.20944/preprints202209.0434.v1

countries, using the *A. mellifera* species as an outgroup species based on a comparative analysis of the gene *VG* exon sequences. The genetic structure of populations and signatures of positive adaptive selection to the local climatic conditions were studied in *A. cerana* populations from Russia, Japan, Korea, China, and Nepal.

Material and methods

The workers of the Eastern honey bee, *Apis cerana*, were collected from the five Asian countries: South Korea (provinces Gyeonggido, Gangwondo, Chungcheongbukdo, Gyeongsangbukdo, Jeollabukdo, Gyeongsangnamdo), Japan (prefectures Hiroshima, Hokkaido), Taiwan (Central Taiwan region), Nepal (province Bagmati Pradesh), Russia (Primorsky Territory) (Fig. 1, Tab. S1). Worker bees were captured and directly displaced in liquid nitrogen before being stored at -800 degrees Celsius. The genomic DNA was isolated from the thoraxes of honey bees using the G-spin Total DNA Extraction Kit (iNtRON Biotechnology, South Korea, Cat No. 17045), according to the manufacturer's instructions at the Division of Life Sciences of Incheon National University. The species of all collected honey bee samples have been determined morphologically [5].

The polymerase chain reaction (PCR) using primers for the gene *VG* exons (Tab. 1) performed in Agilent SureCycler 8800 (Agilent Technologies, Santa Clara, CA) in 12.5 µl volume included 0.15 µl TaKaRa Ex-Taq, 1.25 µl 10X Ex-Taq buffer, 1 µl 2.5 mM dNTP mixture, 1 µl 10 pmol of each primer (Tab. 1), 0.5 µl template DNA, and 8.6 µl nuclease free water under following conditions: 95 °C for 3 min; 35 cycles: 95 °C for 30 s, 60 °C for 30 s, 72 °C for 30 s ; and final elongation 72 °C for 10 min.

All PCR products were purified by the QIAquick PCR Purification Kit (250) (QIAGEN, Hilden, Germany) following the instructions of the manufacturer. The PCR products were sequenced on both strands in Macrogen Inc. (Seoul, Korea) using the ABI 3730xl 96-capillary DNA analyzer (Applied Biosystems, Foster City, CA) and the ABI BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). All sequences have been analyzed and submitted to the GenBank (the National Center for Biotechnology Information, NCBI) (Tab. S1). In total, 210 sequences of six exons of 35 honey bee samples (*A. cerana* and *A. mellifera*) were deposited to GenBank (Tab. S1).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 September 2022 doi:10.20944/preprints202209.0434.v1

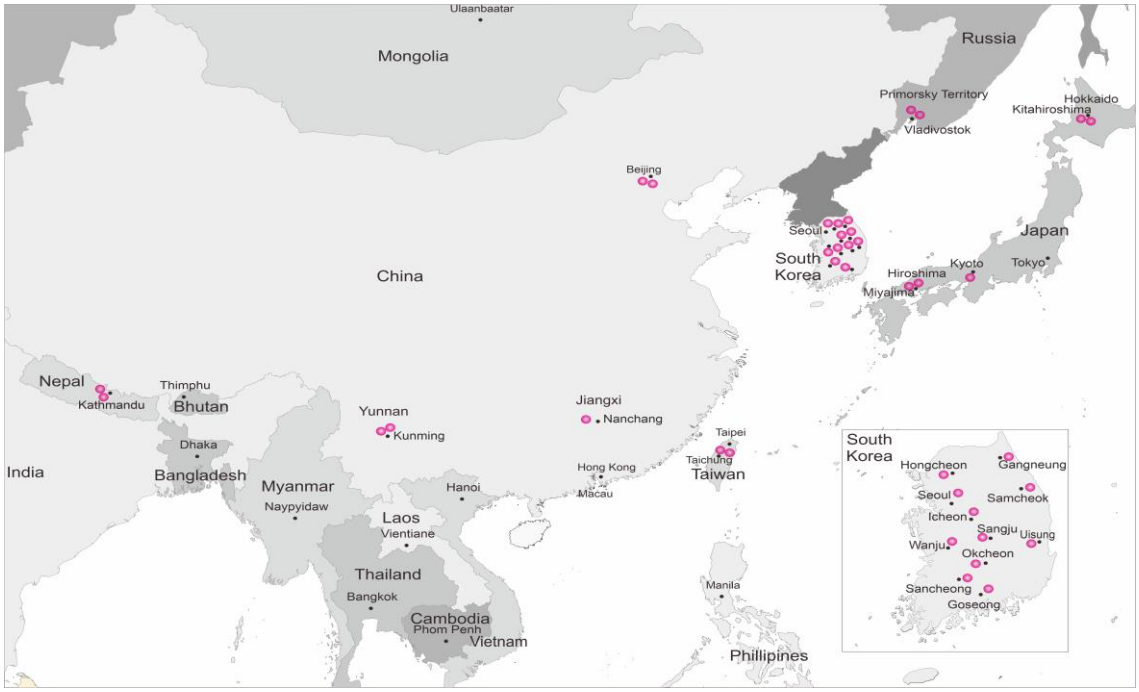

Figure 1. Spatial distribution of collected *A. cerana* honey bee samples.

Comparative analysis of the gene Vitellogenin exons of honey bees *A. cerana* has been performed using Clustal W methods of alignment of nucleotide sequences of honey bee samples retrieved from GenBank in MEGA v. 10.0.5 at the Division of Life Sciences of Incheon National University [18] (Table S1). In total, 48 sequences of six exons of eight honey bee samples (*A. cerana* and *A. mellifera*) were retrieved from GenBank. Alignments of the nucleotide sequences of the gene Vitellogenin exons (Table S1).

Pairwise nucleotide sequence divergences were estimated using Unipro UGENE 1.28 (UNIPRO, Russia) and CLC Genomics Workbench 11 (CLCbio, Denmark) based on complete mtDNA sequences with the Jukes-Cantor model [19]. Pairwise Euclidean distances between *A. cerana* populations based on morphology data were calculated using Statistica 8.0. Based on pairwise alignments, amino acid identity (%) was calculated for homologous genes. The significance of Tajima's D test was determined using coalescent simulations with 10,000 runs as implemented in DNAsp [20]. Phylogenetic trees were constructed using the Neighbor-Joining method [21] based on the Jukes-Cantor model with 1000 bootstrap replications [21]. The estimation of evolutionary divergence over sequence pairs between groups, the estimation of the number of base differences (transitions +transversions) per sequence from averaging over all sequence pairs between groups, the estimation of codon-based Z-test of selection and evolutionary divergence, the ratio of non-synonymous (dN) to synonymous (dS) substitutions dN/dS and Standard Errors were conducted in MEGA X based on gene *VG* exon sequences

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 September 2022 doi:10.20944/preprints202209.0434.v1

using the Nei-Gojobori method [22].

Table 1. The PCR primers for the gene *VG* exons of honey bees *Apis cerana*

Name	Sequence 5' - 3'	TM, ^{0}C	GC, %	Size, bp
Ac*VG*E2-F	TCTTGTTCGTTCCAGGTTCC	58.4	50	677
Ac*VG*E2-R	GACAGTTTCAGCCGACTTCC	60.5	55	
Ac*VG*E3-F	CCTTTCGATCCATTCCTTGA	56.4	45	679
Ac*VG*E3-R	GTCAAAACGGATTGGTGCTT	56.4	45	
Ac*VG*E4-F	TCGAAGGGGAAGAATTTCAA	54.3	40	840
Ac*VG*E4-R	ACGAGCAATTCCTCAACACC	58.4	50	
Ac*VG*E5-F	GTCGGACAATTTCACGTCCT	58.4	50	1177
Ac*VG*E5-R	GTTCGAGCATCGACACTTCA	58.4	50	
Ac*VG*E6-F	AGAGCCAGGGATACGTCAAA	58.4	50	406
Ac*VG*E6-R	GAGTCATCTCGAGGCTCACC	62.5	60	
Ac*VG*E7-F	TTCTGGCTGAGGTCAGGATT	58.4	50	445
Ac*VG*E7-R	AATTTCGACCACGACTCGAC	58.4	50	

Results

The exons of gene *VG* of thirty-three samples of *A. cerana* collected from Korea, Russia, Japan, Nepal, and China with an average size of 4,125 bp were sequenced and aligned with retrieved from GenBank *VG* sequences of seven samples of *A. cerana* from China, Japan, Korea (Tab. 3, Fig. S1). The exons of gene *VG* of two samples of *A. mellifera* from India with an average size of 4,128 bp were sequenced and aligned with retrieved from the GenBank *VG* sequence of *A. mellifera* from Poland (Tab. S1). The exons of gene *VG* of *A. mellifera* were used as an outgroup for *A. cerana* samples. Both honey bee species *A. cerana* and *A. mellifera* have a similar number of exons of *VG* and almost similar size of their sequences (Fig. S1).

All samples of *A. cerana* from Korea, Russia, Japan, Nepal, and China were pooled into twenty populations according to their geographical origin. All three samples of *A. mellifera* from India and Poland were pooled into one outgroup. The population genetic parameters, which included the number of nucleotide differences (transitions + transversions), the ratio of non-synonymous to synonymous substitutions (dN/dS), and Jukes-Cantor genetic distances, were assessed on twenty-one populations (twenty populations of *A. cerana* and one outgroup population of *A. mellifera*) (Tab. S2, S3, S4).

The number of nucleotide differences including transitions + transversions varied from

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 September 2022 doi:10.20944/preprints202209.0434.v1

10 to 61 between *A. cerana* populations and varied from 249 to 270 between *A. cerana* and *A. mellifera* populations. The minimal number of nucleotide differences including transitions + transversions, which varied from 10 to 12 were found between Korean *A. cerana* populations (Goseong-Hongcheon, Goseong-Okcheon, Goseong-Uisung, Gangneung-Samcheok, Samcheok-Sancheong, Samcheok-Okcheon). The maximal number of nucleotide differences including transitions + transversions, which varied from 50 to 61 were found between Russian and Chinese (Beijing Taiwan Yunnan), Japanese Kyoto, Nepali Kathmandu *A. cerana* populations.

The dN/dS ratio varied from 0.236 to 0.999 between *A. cerana* populations and from 0.496 to 0.538 between *A. cerana* and *A. mellifera* populations. The minimal dN/dS ratio, which varied from 0.236 to 0.295, was found between Korean (Samcheok-Goseong, Samcheok-Hongcheon) and Japanese (Samcheok-Hiroshima, Goseong-Hiroshima) *A. cerana* populations. The maximal dN/dS ratio, which varied from 0.711 to 0.999, was found between Korean (Hongcheon-Uisung, Samcheok-Wanju), Korean and Japanese (Hongcheon-Kyoto), and Korean and Russian (Samcheok-Vladivostok) *A. cerana* populations.

The Jukes-Cantor genetic distances varied from 0.002 to 0.015 between *A. cerana* populations and from 0.065 to 0.070 between *A. cerana* and *A. mellifera* populations. The minimal Jukes-Cantor genetic distances, which varied from 0.002 to 0.003 were found between Korean (Goseong-Hongcheon, Goseong-Samcheok, Goseong-Sancheong, Goseong-Okcheon, Goseong-Uisung, Gangneung-Samcheok, Hongcheon-Samcheok, Hongcheon-Uisung, Samcheok-Okcheon, Gangneung-Sancheong, Hongcheon-Okcheon, Gangneung-Wanju, Sancheong-Wanju, Seoul-Wanju) *A. cerana* populations. The maximal Jukes-Cantor genetic distances, which varied from 0.012 to 0.015 were found between Russian and Korean (Vladivostok-Uisung, Vladivostok-Goseong), Russian and Japanese (Vladivostok-Kyoto), Russian and Chinese (Vladivostok-Jiangxi, Vladivostok-Yunnan), Russian and Nepali (Vladivostok-Kathmandu) *A. cerana* populations.

The phylogenetic tree was constructed based on Jukes-Cantor genetic distances using the Neighbor Joining method of clustering (Fig. 2). Korean *A. cerana* populations were subdivided into three clusters. Japanese *A. cerana* populations were distantly joined to the Korean *A. cerana* group. All Chinese and Nepali *A. cerana* populations were clustered together. Russian *A. cerana* population was located apart from all *A. cerana* populations (Fig. 2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 September 2022 doi:10.20944/preprints202209.0434.v1

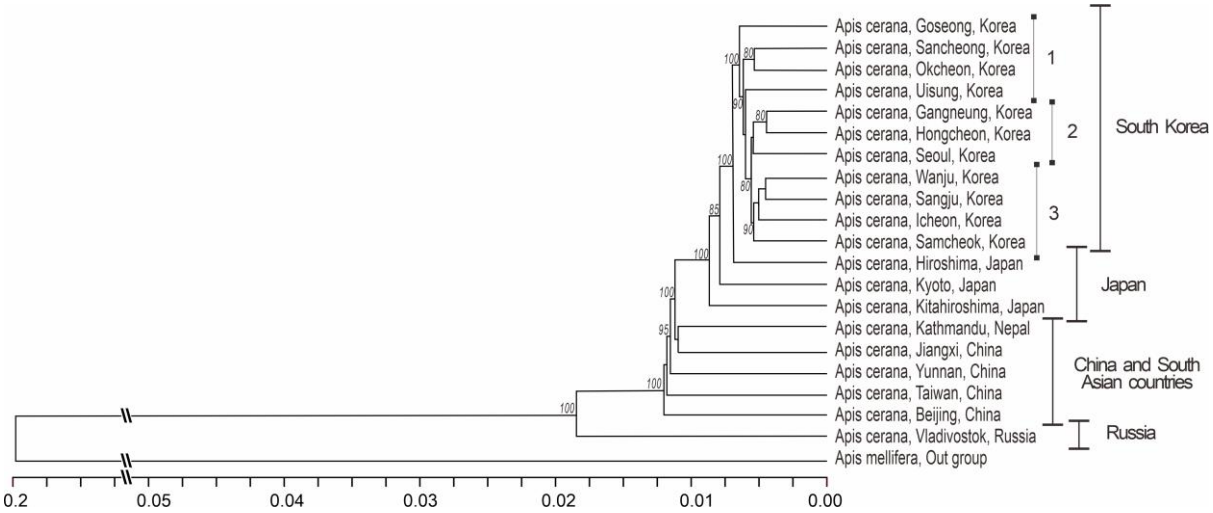

Figure 2. A neighbor-joining phylogenetic tree of *Apis cerana* samples was constructed based on the gene *VG* nucleotide sequences. The Korean *A. cerana* populations were subdivided into three groups according to their geographical distribution: 1. the southern region; 2. the northern region; and 3. the central region.

Further, all forty samples of *A. cerana* from Korea, Russia, Japan, Nepal, and China were pooled into five populations according to their country distributions. All three samples of *A. mellifera* from India and Poland were pooled into one outgroup. The population genetic parameters, which included the number of nucleotide differences (transitions plus transversions), the ratio of non-synonymous to synonymous substitutions (dN/dS), and Jukes-Cantor genetic distances, were assessed on six populations (five populations of *A. cerana* and one outgroup population of *A. mellifera*) (Tab. 2).

The average number of nucleotide differences, including transitions and transversions, varied from 20.6 to 54.3 between *A. cerana* populations and from 255.1 to 269.8 between *A. cerana* and *A. mellifera* populations. The minimal number of nucleotide differences, including transitions and transversions, which varied from 20.6 to 28.8, was found between Japanese-Korean and Japanese-Nepali *A. cerana* populations. The maximal number of nucleotide differences, including transitions and transversions, which varied from 51.5 to 54.3, were found between Russian-Chinese and Russian-Nepali *A. cerana* populations (Tab. 2).

The average dN/dS ratio varied from 236 to 999 between *A. cerana* populations and from 496 to 538 between *A. cerana* and *A. mellifera* populations. The minimal dN/dS ratio, which varied from 236 to 295, was found between Korean (Samcheok-Goseong, Samcheok-Hongcheon) and Japanese (Samcheok-Hiroshima, Goseong-Hiroshima) *A. cerana* populations. The maximal dN/dS ratio, which varied from 711 to 999, was found between Korean (Hongcheon-Uisung, Samcheok-Wanju), Korean and Japanese (Hongcheon-Kyoto), and Korean

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 September 2022 doi:10.20944/preprints202209.0434.v1

and Russian (Samcheok-Vladivostok) *A. cerana* populations.

The average number of nucleotide differences based on the gene *VG* exons, dN/dS ratio between all *A. cerana* and the *A. mellifera* outgroup populations varied from 20.6 to 269.8 (Tab. 2, Tab. S3). The number of nucleotide differences between *A. cerana* populations from different countries varied from 20.6 to 28.8. A maximal number of nucleotide differences (255.1–269.8) was found between *A. cerana* and *A. mellifera* outgroup populations. A minimal number of nucleotide differences (20.6) were observed between Korean and Japanese populations of *A. cerana* (Tab. 2, Tab. S2).

Table 2. The Jukes-Cantor genetic distances above the diagonal, the number of nucleotide differences (transitions + transversions), and the ratio of non-synonymous to synonymous substitutions (dN/dS) below the diagonal between local populations of *A. cerana* and outgroup *A. mellifera* based on the exon sequences of the gene *VG* and their standard error S.E.

Populations		Korea	China	Japan	Nepal	Russia	*A. mellifera* outgroup
	Jukes Cantor genetic distances / S.E.						
Korea	Number of nucleotide differences (tr + tv) / S.E. [dN/dS ratio / S.E.]		0.008 / 0.001	0.005 / 0.001	0.007 / 0.001	0.010 / 0.001	0.066 / 0.004
China		31.5 / 3.8 [0.49/0.001]		0.009 / 0.001	0.008 / 0.001	0.014 / 0.002	0.066 / 0.004
Japan		20.6 / 3.1 [0.39 / 0.001]	34.0 / 4.1 [0.42 / 0.001]		0.007 / 0.001	0.011 / 0.001	0.067 / 0.004
Nepal		28.8 / 4.7 [0.52 / 0.001]	31.0 / 4.2 [0.76 / 0.001]	28.8 / 4.8 [0.50 / 0.001]		0.013 / 0.002	0.068 / 0.004
Russia		40.7 / 5.4 [0.68 / 0.002]	54.3 / 5.35 [0.58 / 0.002]	42.2 / 5.6 [0.63 / 0.002]	51.5 / 6.7 [0.70 / 0.002]		0.070 / 0.005
A. mellifera outgroup		255.9 / 5.4 [0.52 / 0.005]	255.1 / 5.3 [0.51 / 0.004]	256.7 / 5.6 [0.51 / 0.005]	261.3 / 6.0 [0.52 / 0.005]	269.8 / 6.6 [0.50 / 0.005]	

Notes: S.E. - Standard error.

The average Jukes-Cantor genetic distances between local populations of *A. cerana* and *A. mellifera* varied from 0.005 to 0.070 (Tab. 2, Tab. S4). The genetic distances between *A. cerana* populations varied from 0.005 to 0.014. The maximal genetic distances varied from 0.066 to 0.070 and were observed between *A. cerana* and *A. mellifera* outgroup populations (Tab. 2, Tab. S4). A minimal genetic distance of 0.005 was observed between the Korean and Japanese populations of *A. cerana*. The maximal genetic distances varied from 0.010 to 0.014 and were observed between the Chinese, Nepali, and Russian populations of *A. cerana* (Tab. 2, Tab. S4).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 September 2022 doi:10.20944/preprints202209.0434.v1

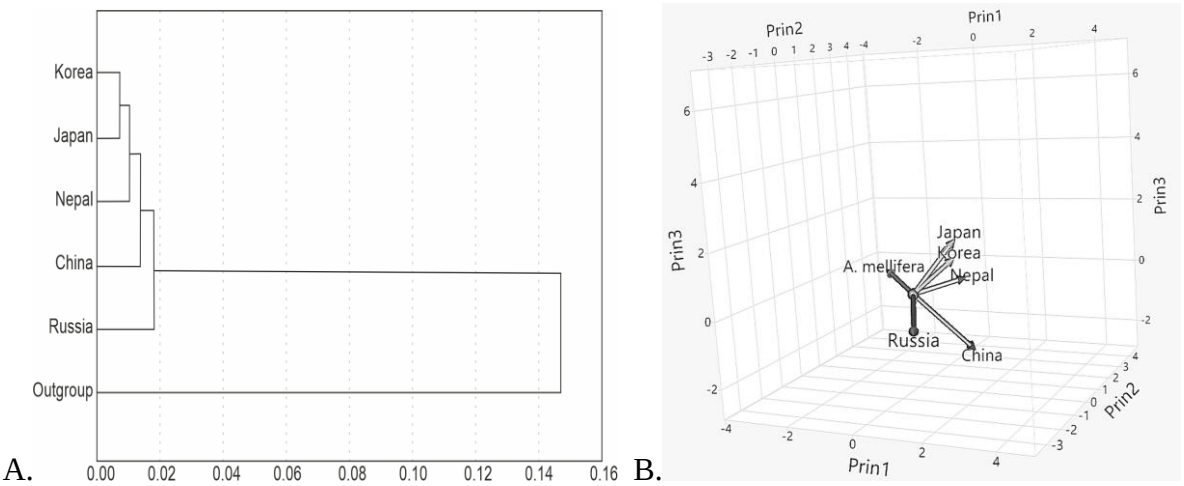

Figure 3. (A) Neighbor-Joining phylogenetic tree and (B) Principal Component Analysis of *Apis cerana* populations from different countries were constructed based on the gene *VG* nucleotide sequences. An outgroup sample of *A. mellifera*.

The average phylogenetic tree between *A. cerana* populations from different countries was constructed using the neighbor-joining method based on Jukes-Cantor genetic distances (Fig. 3). *A. cerana* populations from Korea and Japan were clustered together and separately from populations from Russia, Nepal, and China (Fig. 4).

The Tajima's neutrality test was assessed for all *A. cerana* populations based on the exon sequences of the gene *VG* (Tab. 3). Also, the number of segregating sites, nucleotide diversity, number of synonymous and nonsynonymous sites, and the number of nucleotide differences are counted (Tab. 3).

Table 3. Tajima's neutrality test for *Apis cerana* populations from different local populations based on the gene *VG* exon sequences

Population	S	Ps	Θ	π	D	dN	dS	dN/dS	d	S.E.
Korea	53	0,013	0,003	0,004	-0,685	0,003	0,007	0,448	16,692	0.003
Russia	61	0,015	0,010	0,010	-0,023	0,002	0,005	0,440	12,000	0.003
China	80	0,019	0,008	0,008	-0,252	0,006	0,013	0,462	36,667	0.003
Japan	46	0,011	0,005	0,005	-0,041	0,004	0,009	0,408	22,200	0.001
Nepal	32	0,008	0,005	0,005	-0,042	0,003	0,002	1,263	10,000	0.004

Abbreviations: S = number of segregating sites, Ps = S/n, Θ = Ps/a1, π - nucleotide diversity, D-the Tajima test statistic, dS– synonymous sites, dN– nonsynonymous sites, d-the number of nucleotide differences, S.E.– standard error.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 September 2022 doi:10.20944/preprints202209.0434.v1

Discussion

The genetic analysis of the population structure of oriental honey bees, *A. cerana* subspecies *A. c. koreana* in Korea has not yet been provided, despite the fact that all beekeepers and scientists are interested in this issue. Beekeepers have little knowledge of the type of *A. cerana* they breed and how they are related to the subspecies of *A. cerana* found in Japan, Russia, China, and other South Asian countries. Little is known about how the Asian *A. cerana* bees evolved and how they spread across Asia from the north to the south of the continent. It is possible to answer these questions by analyzing the polymorphism of genes involved in bee development, such as the gene *VG*. The gene *VG* is a major reproductive protein in insects in general and a proposed endocrine factor in honeybees. The RNAi interference of *VG* gene activity paces the onset of foraging behavior, primes bees for specialized foraging tasks, and influences worker longevity. Thus, the gene *VG* influences the geographical distribution and population structure of *A. cerana* [23]. A nucleotide polymorphism of the gene *VG* should be useful to study the population genetic structure of *A. cerana* in the Korean peninsula, similar to previous studies on the population genetic structure of *A. mellifera* [13]. The population genetic analysis based on gene *VG*, which was used in previous population genetic and phylogenetic studies of western honey bees, *A. mellifera* [13], promises to give interesting results on the relationship and distribution of *A. c. koreana* local populations in Korea since there is no consensus about the origin and evolution of honey bees from different parts of the country. The Jukes-Cantor genetic distances, cluster analysis, and Tajima's Neutrality Test are acceptable tools for the initial characterization of local *A. c. koreana* populations in Korea.

The population genetic structure of the *A. cerana* population in the Korean peninsula had not been studied previously. Korea has various climates in different altitude localities, ranging from warm to cold. The Korean Peninsula extends southward from the northeastern part of the Asian mainland. South Korea has no extensive plains; its lowlands are the result of mountain erosion. South Korea's lowlands cover approximately 30% of its total land area, with the remainder made up of uplands and mountains. The great majority of the lowland area lies along the coasts, particularly the west coast, and along the major rivers. A narrow littoral plain extends along the east coast. South Korea, part of the East Asian Monsoon region, has a temperate climate with four distinct seasons. The movement of air masses from the Asian continent exerts a greater influence on South Korea's weather than does air movement from the Pacific Ocean. Winters are usually long, cold, and dry, whereas summers are short, hot, and humid. Spring and autumn are pleasant but short-lived. Mean temperatures in Seoul range from minus 5°C in January to plus 25°C in July, while those in Jeju range from 2.5°C in January to 25°C in July.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 September 2022 doi:10.20944/preprints202209.0434.v1

The coldest regions of South Korea are located in the northeastern part of the mountain region. For the most part, rainfall is over 1,000 millimeters. About 50% of the annual precipitation occurs between June and September. Serious droughts occur about once every eight years in the southwestern part of the country [24].

In this study, we collected *A. cerana* honey bees in a wide range of Korea, from South to North, from East to West, to find genetic signatures of adaptations to the local environments based on polymorphism of the nuclear gene *VG* in comparison with *A. cerana* from Northern (Russia, Japan) and Southern (China, Nepal) countries and outgroup samples of *A. mellifera*. Both bee species have similar exon-intron structures and differ in nucleotide sequences (Fig. 2). It makes it possible to provide comparative analysis samples from different regions with adaptation to cold and hot climates. Due to the *VG* gene's involvement in adaptation, its sequences can reflect local adaptations to various climates. Samples of honey bee *A. cerana* adapted to surviving in similar environmental conditions should have less genetic variation in gene *VG* sequences than samples adapted to different environmental conditions. It is predictable that, according to the results of the analysis of the gene *VG* polymorphism, populations of *A. cerana* bees of northern Asian countries, such as China, Japan, Korea, and Russia, will be grouped together and separately from populations of *A. cerana* bees of southern Asian countries.

The genetic analysis was provided with 40 samples of *A. cerana* from Korea, Russia, Japan, Nepal, and China from 20 geographical populations and outgroup *A. mellifera* samples from India and Poland. The average number of nucleotide differences between the gene *VG* exons, including transitions and transversions, varied from 20.6 to 28.8 between *A. cerana* populations from different countries, which evaluates the level of genetic divergence. The minimal number of nucleotide differences (20.6) was observed between Korean and Japanese populations of *A. cerana* (Tab. S2), which proves their genetic relationship and early stage of divergence due to the recent time of their isolation and gene flow between populations as a result of human activities. The nucleotide differences between the *A. cerana* and *A. mellifera* outgroup populations were highest (255.1–269.8), which demonstrates their divergence at a significantly early period of their evolution. We counted the number of nucleotide differences between the gene *VG* exons within the northern populations of *A. cerana*, which varied from 10 to 52 (the average is 21.9). Also, we counted the number of nucleotide differences in the gene *VG* exons between the northern and southern populations of *A. cerana*, which varied from 22 to 61 (the average is 32.4) (Tab. S2). Here, the majority of the nucleotide differences of the gene *VG* exons between the northern and southern populations of *A. cerana* were higher than within the northern populations of *A. cerana*. The number of nucleotide substitutions is likely to reflect

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 September 2022 doi:10.20944/preprints202209.0434.v1

the magnitude of local adaptations, which are minimal between populations from similar climatic conditions.

We evaluated the dN/dS ratio between all *A. cerana* samples. The dN/dS ratio remains one of the most popular and reliable measures of evolutionary pressures on protein-coding regions. The dN/dS ratio provides information about the evolutionary forces operating on a particular gene [25]. The dN/dS ratio quantifies the mode and strength of selection by comparing synonymous substitution rates (dS) - assumed to be neutral - with nonsynonymous substitution rates (dN), which are exposed to selection as they change the amino acid composition of a protein. Under neutrality, dN/dS = 1. For genes that are subject to functional constraint such that non-synonymous amino acid substitutions are deleterious and purged from the population, i.e., genes under negative selection, dN/dS 1. For the positively selected genes, dN/dS > 1 [25].

The dN/dS ratio between all *A. cerana* samples varied from 0.158 to 0.999, which demonstrates the functional significance of the gene *VG* such that non-synonymous amino acid substitutions are deleterious and purged from the population. Indeed, the gene *VG* affects multiple aspects of social insect life histories and plays a key role in caste differentiation and maintenance of honey bees. The dN/dS ratio between the northern populations of *A. cerana* varied from 0.158 to 0.999 (the average is 0.46), which indicates a strong negative selection for the gene *VG*, which characterizes the high significance of the gene *VG*'s conservativeness in the adaptation of *A. cerana* to the cold northern climate. The dN/dS ratio between the northern and southern populations of *A. cerana* varied from 0.289 to 0.947 (the average is 0.49) (Tab. S3), which indicates a positive selection of the gene *VG* in evolution and the high significance of the gene *VG* polymorphism in the adaptation of *A. cerana* to the different climates of southern and northern Asia. In comparison, the average dN/dS ratio for 3256 nuclear genes was 0.1068 between 3 species of Formicidae, was 0.1033 between 3 species of Polistinae and Vespinae wasps, and was 0.1086 between 10 species of *Apis*, *Bombus*, and *Tetragonula*. It was 0.1025 between 3 species of Siricoidea and 0.1005 between 5 species of Cynipoidea, which characterized the positive selection of these genes in evolution and their role in adaptation to the local climate [26]. Thus, the gene *VG* provides the strong adaptation of *A. cerana* to the local north and south climates.

The Jukes-Cantor genetic distance is the genetic relatedness between two nucleotide sequences calculated as the sum of the substitutions that have accumulated between them since they diverged from their common ancestor during evolution, assessed using the Jukes-Cantor model. The Jukes-Cantor genetic distances based on the gene *VG* exon sequences between *A.*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 September 2022 doi:10.20944/preprints202209.0434.v1

cerana populations varied from 0.002 to 0.015, and between *A. cerana* and *A. mellifera* populations varied from 0.065 to 0.070. The Jukes-Cantor genetic distances based on the gene *VG* exon sequences between the northern populations of *A. cerana* varied from 0.002 to 0.013 (the average is 0.005). The Jukes-Cantor genetic distances based on the gene *VG* exon sequences between the northern and southern populations of *A. cerana* varied from 0.005 to 0.015 (the average is 0.010) (Tab. S4). In comparison, the Jukes-Cantor genetic distance was 0.018 between subspecies of *A. mellifera* (*A. m. carnica*, *A. m. ligustica*, *A. m. sicula*, *A. m. iberica*, *A. m. adami*, *A. m. macedonica*, *A. m. anatoliaca*, *A. m. syriaca*, *A. m. intermissa*) based on *COX1* gene of mtDNA [27], was 0.17 between 2 honeybee species (*A. mellifera*, *A. cerana*) based on complete mtDNA [28], varied from 0.001 to 0.014 between subspecies of *A. cerana* and varied from 0.075 to 0.081 between species *A. mellifera* and *A. cerana* based on gene *VG* of nDNA [9], was 0.033 between 15 species of Anastrepha based on *COX1* gene of mtDNA [29], varied from 0.002 to 0.039 within species and from 0.074 to 0.131 between five species of *Apis* based on *CYTB* gene of mtDNA [30]. Thus, the Jukes-Cantor genetic distances based on the gene *VG* exon sequences between *A. cerana* populations from similar climatic conditions is twice less than between *A. cerana* populations from different climatic conditions. It can be explained the important role of the gene *VG* in the adaptation of *A. cerana* populations to different climatic conditions from northern to southern Asia. The phylogenetic tree constructed using the neighbor-joining method with the Jukes-Cantor genetic distances based on the gene *VG* exon sequences (Fig. 3) shows the informativeness of the gene in phylogenetic and ecological studies of *A. cerana*. Also, we can see the genetic relationship between populations on a dendrogram based on cluster analysis and in a plot based on principal component analysis. Moreover, there is a genetic differentiation based on the gene *VG* exon sequences between Korean populations of *A. c. koreana* which is related to the geographic localization and climatic conditions of southern, central, and northern provinces of Korea.

We performed the Tajima's neutrality D test for *A. cerana* populations from different local populations based on the gene *VG* exon sequences (Tab. 3). Tajima's D test is a statistical method for testing the neutral mutation hypothesis by DNA polymorphism [31]. Tajima's D is a statistical method for testing the neutral mutation hypothesis by DNA polymorphism [31]. The purpose of Tajima's D test is to distinguish between a DNA sequence evolving neutrally without selection and non-neutrally under selection. A neutrally evolving DNA sequence contains mutations with no effect on the fitness and survival of an organism. Tajima's D test computes a standardized measure of the total number of segregating sites in the sampled DNA and the average number of mutations between pairs in the sample. A negative Tajima's D signifies an

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 September 2022 doi:10.20944/preprints202209.0434.v1

excess of low-frequency polymorphisms relative to expectation, indicating population size expansion after a bottleneck or a selective sweep. A positive Tajima's D signifies low levels of both low and high-frequency polymorphisms, indicating a decrease in population size and/or balancing selection [31].

The results of Tajima's D test have been counted for each population of *A. cerana*. Almost all of the estimated D in *A. cerana* bees is not significantly different from zero, implying that *A. cerana* honey bees are in a state of mutation drift equilibrium [18]. The number of segregating sites (S) was lowest in the Nepali population of *A. cerana* and highest in the Chinese population of *A. cerana*. This may be due to the high genetic diversity that should be observed in the largest Chinese population of *A. cerana*. The nucleotide diversity was highest in the Russian population of *A. cerana* and lowest in the Korean population of *A. cerana*. It can be explained by the fact that the *A. cerana* population in Russia is feral under natural selection. There are no managed *A. cerana* populations in Russia (Tab. 3).

For all *A. cerana* populations, the value of Tajima's D test is statistically significant. Only the Korean population of *A. cerana* has the biggest negative value (-0,685), which means that this population can be after a bottleneck, a selective sweep, purifying selection, and positive selection. Indeed, the Korean population of *A. c. cerana* experienced an extreme population decline with the kSBV virus that killed 95% of the local population [4, 10, 11, 32]. All other populations of *A. cerana* also have the signatures of the population size expansion after a selective sweep in the past (Tab. 3).

Conclusions

We performed a population genetic analysis of the honey bees *A. cerana* from South Korea in comparison with *A. cerana* samples from Russia, Japan, Nepal, and China and outgroup samples of *A. mellifera*. Based on the gene *VG* sequences, we assessed the positive selection in *A. cerana* populations to the local climatic conditions. The cluster and principal component analysis based on the sequences of the gene *VG* divided all local populations of *A. c. koreana* in Korea according to their climatic and geographic distribution. The Jukes-Cantor genetic distances based on the gene *VG* sequences showed the *A. cerana* populations from northern countries such as Korea, Russia, and Japan are more closely related to each other than the southern populations of *A. cerana*. The signatures of positive adaptive selection were found in the local population of *A. cerana*. All *A. cerana* populations showed signs of population size expansion following recent declines. The local populations of *A. c. koreana* were subdivided according to their geographical distribution into southern, northern, and central Korean clusters.

The gene *VG* exon sequences can be used as informative markers for monitoring the changes in genetic structure and adaptation processes in *A. cerana* populations.

Acknowledgements

This research was supported by the Russian Government Contract 2021–2023 for RI (Registration No. AAAA-A21-121011990120-7), by grants from the Russian Foundation for Basic Research for NA (RFBR) (Grant No. 19-54-70002 e-Asia_t), and by the Post-Doctoral Research Programs at Incheon National University for RI (2017–2019).

References

1. Koeniger, N.; Koeniger, G.; Tingek, S. *Honey bees of Borneo: Exploring the centre of Apis diversity*. Natural History Publications: Borneo, 2010; p 262.

2. Behura, S.K. Analysis of nuclear copies of mitochondrial sequences in honeybee (*Apis mellifera*) genome. *Molecular Biology and Evolution* **2007**, *24*, 1492-1505. doi: 10.1093/molbev/msm068

3. Ilyasov, R.A.; Han, G.Y.; Lee, M.L.; Kim, K.W.; Proshchalykin, M.Y.; Lelej, A.S.; Park, J.H.; Takahashi, J.I.; Kwon, H.W.; Nikolenko, A.G. Genetic properties and evolution of Asian honey bee *Apis cerana ussuriensis* from primorsky krai, Russia. *Russian Journal of Genetics* **2021**, *57*, 568-581. doi: 10.1134/S1022795421050033

4. Koetz, A.H. Ecology, behaviour and control of *Apis cerana* with a focus on relevance to the Australian incursion. *Insects* **2013**, *4*, 558-592. doi: 10.3390/insects4040558

5. Ruttner, F. *Biogeography and taxonomy of honeybees*. Springer-Verlag: Berlin, Heidelberg, Germany, 1988; p 288.

6. Radloff, S.E.; Hepburn, C.; Hepburn, H.R.; Fuchs, S.; Hadisoesilo, S.; Tan, K.; Engel, M.S.;

7. Proshchalykin, M.Y.; Novomodnyi, E.V.; Bezborodov, V.G.; Koshkin, E.S. First records of wax bee *Apis cerana* Fabricius, 1793 (Hymenoptera, Apidae) in khabarovskii krai. *Euroasian Entomological Journal* **2014**, *13*, 295-298. doi:

8. Ilyasov, R.A.; Park, J.; Takahashi, J.; Kwon, H.W. Phylogenetic uniqueness of honeybee *Apis cerana* from the Korean peninsula inferred from the mitochondrial, nuclear, and morphological data. *Journal of Apicultural Science* **2018**, *62*, 189-214. doi: 10.2478/jas-2018-0018

9. Ilyasov, R.A.; Han, G.Y.; Lee, M.-l.; Kim, K.W.; Proshchalykin, M.Y.; Lelej, A.S.; Takahashi, J.-i.; Kwon, H.W. Phylogenetic relationships of russian far-east *Apis cerana* with other north Asian populations. *Journal of Apicultural Science* **2019**, *63*, 289-314. doi: 10.2478/jas-2019-0024

10. Choi, Y.S.; Lee, M.Y.; Hong, I.P.; Kim, N.S.; Kim, H.K.; Lee, K.G.; Lee, M.L. Occurrence of sacbrood virus in Korean apiaries from *Apis cerana* (Hymenoptera: Apidae). *Journal of Apiculture* **2010**, *25*, 187-191. doi: 10.17519/apiculture.2017.11.32.4.345

11. Vung, N.N.; Lee, M.-L.; Lee, M.-Y.; Kim, H.K.; Kang, E.J.; Kim, J.E.; Choi, Y.-S. Breeding and selection for resistance to sacbrood virus for *Apis cerana*. *Journal of Apiculture* **2017**, *32*, 345-352. doi: 10.17519/apiculture.2017.11.32.4.345

12. Grant, K.J.; DeVetter L.; Melathopoulos A. Honey bee (*Apis mellifera*) colony strength and its effects on pollination and yield in highbush blueberries (*Vaccinium corymbosum*). *PeerJ*. 2021, 27, 9: e11634. doi: 10.7717/peerj.11634

13. Ilyasov, R.A.; Poskryakov, A.V.; Nikolenko, A.G. New SNP markers of the honeybee Vitellogenin gene (*VG*) used for identification of subspecies *Apis mellifera mellifera* L. *Russian Journal of Genetics* **2015**, *51*, 194-199. doi: 10.7868/S0016675815020083

14. Kent, C.F.; Issa, A.; Bunting, A.C.; Zayed, A. Adaptive evolution of a key gene affecting queen and worker traits in the honey bee, *Apis mellifera*. *Molecular Ecology* **2011**, *20*, 5226-5235. doi: 10.1111/j.1365-294X.2011.05299.x

15. Corona, M.; Velarde, R.A.; Remolina, S.; Moran-Lauter, A.; Wang, Y.; Hughes, K.A.; Robinson, G.E. Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. *Proc Natl Acad Sci U S A* **2007**, *104*, 7128-7133. doi: 10.1073/pnas.0701909104

16. Amdam, G.V. Social context, stress, and plasticity of aging. *Aging Cell* **2011**, *10*, 18-27. doi: 10.1111/j.1474-9726.2010.00647.x

17. Bomtorin, A.D.; Mackert, A.; Rosa, G.C.; Moda, L.M.; Martins, J.R.; Bitondi, M.M.; Hartfelder, K.; Simoes, Z.L. Juvenile hormone biosynthesis gene expression in the corpora allata of honey bee (*Apis mellifera* L.) female castes. *PLOS One* **2014**, *9*, e86923. doi: 10.1371/journal.pone.0086923

18. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. Mega X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **2018**, *35*, 1547-1549. doi: 10.1093/molbev/msy096

19. Jukes, T.H.; Cantor, C.R. Evolution of protein molecules. In *Mammalian protein metabolism*, Munro, H.N., Ed. Academic Press: New York, USA, 1969; pp 21-132.

20. Librado, P.; Rozas, J. Dnasp v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451-1452. doi: 10.1093/bioinformatics/btp187

21. Saitou, N.; Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **1987**, *4*, 406-425. doi: 10.1093/oxfordjournals.molbev.a040454

22. Nei, M.; Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **1986**, *3*, 418-426. doi: 10.1093/oxfordjournals.molbev.a040410

23. Nelson, C.M.; Ihle, K.E.; Fondrk, M.K.; Page, R.E.; Amdam, G.V. The gene Vitellogenin has multiple coordinating effects on social organization. *PLoS Biol* **2007**, *5*, e62. doi: 10.1371/journal.pbio.0050062

24. Peel, M.C.; Finlayson, B.L.; McMahon, T.A. Updated world map of the köppen-geiger climate classification. *Hydrology and Earth System Sciences* **2007**, *11*, 1633-1644. doi: 10.5194/hess-11-1633-2007

25. Kryazhimskiy, S.; Plotkin, J.B. The population genetics of DN/DS. *PLoS Genetics* **2008**, *4*, e1000304. doi: 10.1371/journal.pgen.1000304

26. Weyna, A.; Romiguier, J. Relaxation of purifying selection suggests low effective population size in eusocial Hymenoptera and solitary pollinating bees. *bioRxiv - Evolutionary Biology* **2021**, *1*. doi: 10.1101/2020.04.14.038893

27. Ashokan, K.V. Molecular phylogenetic study on *Apis mellifera* subspecies inferred from cytochrome oxidase i. *Indian Journal of Fundamental and Applied Life Sciences* **2011**, *1*, 193-202. doi:

28. Tan, H.W.; Liu, G.H.; Dong, X.; Lin, R.Q.; Song, H.Q.; Huang, S.Y.; Yuan, Z.G.; Zhao, G.H.; Zhu, X.Q. The complete mitochondrial genome of the Asiatic cavity-nesting honeybee *Apis cerana* (Hymenoptera: Apidae). *PLOS One* **2011**, *6*, e23008. doi: 10.1371/journal.pone.0023008

29. Smith-Caldas, R.B.M.; Mcpheron, A.B.; Silva, G.J.; Zucchi, A.R. Phylogenetic relationships among species of the fraterculus group (Anastrepha: Diptera: Tephritidae) inferred from DNA sequences of mitochondrial Cytochrome Oxidase I. *Neotropical Entomology* **2001**, *30*, 565-557. doi: 10.1590/s1519-566x2001000400009

31. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **1989**, *123*, 585-595. doi: 10.1093/genetics/123.3.585

32. Wang, S.; Chen, G.; Lin, Z.; Wu, Y.; Hu, F.; Zheng, H. Occurrence of multiple honeybee viruses in the ectoparasitic mites *Varroa* spp. in *Apis cerana* colonies. *Journal of Invertebrate Pathology* **2019**, *166*, 107225. doi: 10.1016/j.jip.2019.107225