



Article

Discovery of SNP Molecular Markers and Candidate Genes Associated with Sacbrood Virus Resistance in *Apis cerana cerana* Larvae by Whole-Genome Resequencing

Aqai Kalan Hassanyar ^{1,2}, Hongyi Nie ², Zhiguo Li ², Yan Lin ², Jingnan Huang ², Samuel Tareke Woldegiorgis ¹, Mubasher Hussain ³, Wangjiang Feng ², Zhaonan Zhang ², Kejun Yu ² and Songkun Su ^{2,*}

¹ College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 2191914004@fafu.edu.cn (A.K.H.)

² College of Animal Sciences (College of Bee Science), Fujian Agriculture and Forestry University, Fuzhou 350002, China

³ College of Plant Protection, Fujian Agriculture and Forestry University, Fuzhou 350002, China

* Correspondence: susongkun@zju.edu.cn; Tel.: +86-18105039938

Abstract: Sacbrood virus (SBV) is a significant problem that impedes brood development in both eastern and western honeybees. Whole-genome sequencing has become an important tool in researching population genetic variations. Numerous studies have been conducted using multiple techniques to suppress SBV infection in honeybees, but the genetic markers and molecular mechanisms underlying SBV resistance have not been identified. To explore single nucleotide polymorphisms (SNPs), insertions, deletions (Indels), and genes at the DNA level related to SBV resistance, we conducted whole-genome resequencing on 90 *Apis cerana cerana* larvae raised in vitro and challenged with SBV. After filtering, a total of 337.47 gigabytes of clean data and 31,000,613 high-quality SNP loci were detected in three populations. We used ten databases to annotate 9359 predicted genes. By combining population differentiation index (F_{ST}) and nucleotide polymorphisms (π), we examined genome variants between resistant (R) and susceptible (S) larvae, focusing on site integrity ($INT < 0.5$) and minor allele frequency ($MAF < 0.05$). A selective sweep analysis with the top 1% and top 5% was used to identify significant regions. Two SNPs on the 15th chromosome with GenBank KZ288474.1_322717 (Guanine > Cytosine) and KZ288479.1_95621 (Cytosine > Thiamine) were found to be significantly associated with SBV resistance based on their associated allele frequencies after SNP validation. Each SNP was authenticated in 926 and 1022 samples, respectively. The enrichment and functional annotation pathways from significantly predicted genes to SBV resistance revealed immune response processes, signal transduction mechanisms, endocytosis, peroxisomes, phagosomes, and regulation of autophagy, which may be significant in SBV resistance. This study presents novel and useful SNP molecular markers that can be utilized as assisted molecular markers to select honeybees resistant to SBV for breeding and that can be used as a biocontrol technique to protect honeybees from SBV.

Keywords: sacbrood virus; single nucleotide polymorphism; larva; *Apis cerana cerana*; population differentiation index; nucleotide polymorphisms; variants; sequencing



Citation: Hassanyar, A.K.; Nie, H.; Li, Z.; Lin, Y.; Huang, J.; Woldegiorgis, S.T.; Hussain, M.; Feng, W.; Zhang, Z.; Yu, K.; et al. Discovery of SNP Molecular Markers and Candidate Genes Associated with Sacbrood Virus Resistance in *Apis cerana cerana* Larvae by Whole-Genome Resequencing. *Int. J. Mol. Sci.* **2023**, *24*, 6238. <https://doi.org/10.3390/ijms24076238>

Received: 29 January 2023

Revised: 5 March 2023

Accepted: 7 March 2023

Published: 25 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Both *Apis cerana cerana* (*A. c. cerana*) and *Apis mellifera* (*A. mellifera*) are widely used model species for apiculture and crop pollination, offering significant economic benefits [1,2]. However, concerns about honeybee health and its impact on the global economy have increased in recent years [2,3]. Sacbrood virus (SBV) was first reported as the most prevalent viral infection in honeybees in 1913, with White from the United States initially describing this group of diseases in *A. mellifera* [4,5]. SBV can infect larva, prepupa, and adult bees, but larvae are particularly sensitive to infection [6–8]. Due to its

high replication rate, SBV can cause significant morphological changes in infected larvae, leading to their death and it becoming a widely distributed virus [9,10]. The disease was initially referred to as sacbrood due to the inability of infected larvae to pupate and the development of sac fluid, which is rich in SBV particles, beneath their undigested sack. SBV has been found to infect not only honeybees but also non-bee insects, such as bumblebees [11,12]. The diseased larvae exhibit a scale-like dark brown color resembling a ship, particularly in the mouth and digestive systems, which contain high amounts of external viruses [13]. Latent infection is significant for SBV propagation because it accumulates in the head and hypopharyngeal glands of infected nursing bees and is responsible for larval feeding through infected glandular secretions [14]. Honeybee diseases pose a significant threat to the health and well-being of honeybees, which researchers and beekeepers are seriously concerned about [15]. Honeybees, like other insects, have evolved innate immunity to protect themselves from external invaders. The immune system of *A. cerana* comprises 144 genes and 12 pathways [16]. SBV is a significant brood disease with seasonal prevalence and is much more lethal in *A. cerana* than in *A. mellifera* [17].

To develop effective disease control strategies for preventing SBV development, it is essential to gain a deeper understanding of bee resistance and the molecular mechanisms involved in resistant bee larvae. Various methods have been explored for managing SBV infection in honeybees, such as RNA interference (RNAi). However, large-scale RNAi isolation and purification are necessary to prevent SBV infection in bees [18]. The use of silver ions to protect bees from SBV is still challenging and not cost-effective [19]. Administering bee syrup mixed with herbal medicine may not be very effective in preventing SBV infection [20]. Requeening and feeding sugar syrup and pollen to the infected colony may improve the ability of the colony to manage the virus [21]. The application of immunoglobulin Y (IgY) to enhance immunization with a deactivated SBV vaccine for controlling Chinese sacbrood virus (CSBV) has proven useful but it requires large-scale purification [22]. Polypore mushroom mycelia have been shown to reduce viruses in honeybees, suggesting that fungi and their antibacterial compounds may add to bee SBV resistance. Silencing-related genes were examined in *A. cerana* using exogenous double-stranded RNA (dsRNA) produced by the pathogenic insect bacterium *Bacillus thuringiensis* (Bt), demonstrating the potential for using Bt as a platform for effective dsRNA production to control viral infections in host insects. Additionally, *Radix isatidis* extract has been shown to have antiviral properties against SBV [23].

Molecular markers and genomic DNA sequencing technologies are commonly used in studying honeybees from various perspectives. Genome-wide association studies (GWAS) in *A. cerana* have reported the usefulness of SNPs and simple sequence repeats (SSRs) as markers [24]. The *A. c. cerana* genome is smaller than the *A. mellifera* genome due to length inconsistency and has been sequenced to investigate adaptation and diversity in different climates [25]. Studies on disease resistance using SNP molecular markers in *A. mellifera* are more widespread than those in *A. cerana* [26,27]. The genomes of both species have been sequenced and completed [28]. The first draft genome of *A. cerana* was generated in 2015 by Park et al., with an estimated size of 228.32 Mb, 2430 scaffolds, and a 152.0 N50 genome coverage [29]. In 2018, Diao et al. reported the genome assembly of the Chinese honeybee species *A. c. cerana*, with a genome size of 228.79 Mb [16]. *A. c. cerana* is suitable for studying population differentiation index (F_{ST}), gene variations, and genetic structure. Illumina sequencing is a precise platform that can generate millions of short-read sequences [30]. DNA resequencing using SNPs with F_{ST} and population differentiation index (π) has been used to assess the high royal jelly production of the honeybee *A. mellifera* [31]. SNP molecular markers have also been used to examine genetic variations in honeybees that have undergone Africanization for ancestry assignments. F_{ST} comparison pools and individual sequencing have been reported for studying the populations, structure, and diversity of European honeybees [32]. SNP markers are a contemporary approach to molecular breeding programs, especially in disease control and genomic research. More studies on *A. c. cerana* are required to identify which SNPs alter

cellular function and contribute to disease to cover any protection gaps for this bee species. SNP molecular markers are commonly used in genetic research in various organisms [33].

All studies have shown that identifying SNPs and molecular mechanisms of larval resistance to SBV is crucial for disease control in honeybees. However, the exact mechanism by which SNPs and alleles affect honeybee susceptibility or resistance to *A. c. cerana* remains unknown. To address this, we conducted experiments to explore the molecular mechanism of larval bee resistance to SBV. We raised larvae in vitro and challenged them with SBV, and the results are shown in Figure 1. Genomic DNA was extracted from 90 individual R and S larvae using next-generation high-throughput DNA sequencing, and an integrated analysis was conducted based on a comparison of R and S larvae to screen SNP molecular markers and their target genes related to SBV resistance. The analysis included resequencing of 90 samples, data evaluation, genome comparison, mutation detection and annotation, and genetic evolutionary analysis based on variant detection results. We analyzed differences in SNPs in honeybee larvae and their target genes and investigated the distribution of allele frequencies between R and S in the larvae. We found several regions with SNPs and their target genes that were significantly associated with SBV resistance, and a summary of the results is shown in Table 1. A minor allele frequency (MAF < 0.05) analysis was conducted between R and S to assess variations compared to the reference genome, and a high percentage of genotypes was selected to verify the candidate SNPs. Among them are two SNP loci located on the 15th chromosome that were significantly associated with SBV resistance based on their allele frequencies. We validated each SNP with multiple confirmations including (at the same colony level, field validations, comparisons of colony resistance (CR) and colony susceptibility (CS), long-distance geographical locations, and tested on resistant queens).

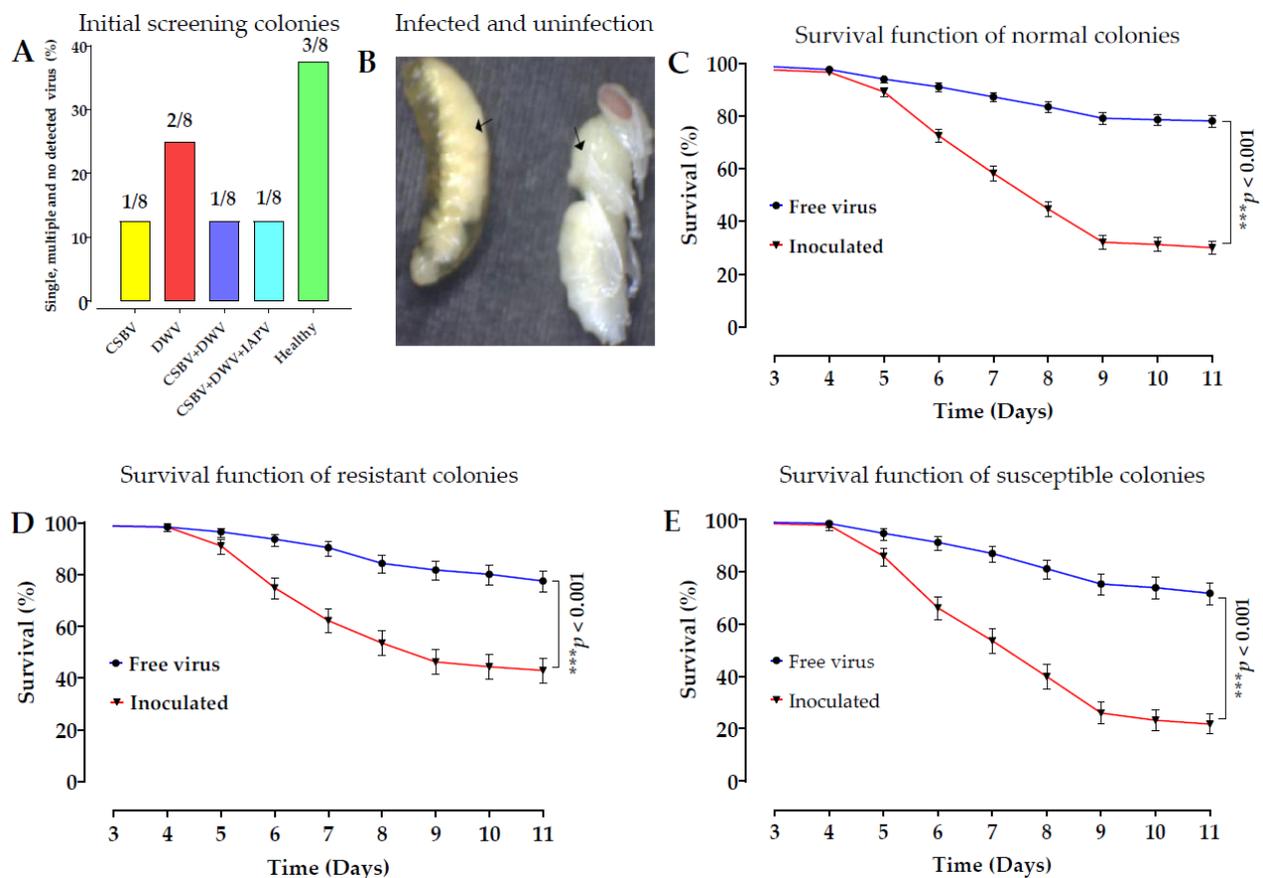


Figure 1. Screening colonies, infected and uninfected larva, and probability survival function. (A) Screening candidate colonies against seven common bee viruses. Based on RT-PCR detection

of viruses, three out of eight colonies were infected with a single virus (CSBV and DWV), two were infected with multiple viruses (CSBV, DWV, and IAPV) in the three colonies, and no virus was detected. (B) Infected larvae on the left ← and resistant larvae (pupae) on the right side → after inoculation with SBV. (C) To assess the efficient infectiousness of the purified virus, a comparison of the survival function of normal colonies with the free virus group with the treatment groups was conducted using a log-rank (Mantel–Cox) test with 95% CI ratio, where a p value of < 0.01 was defined as significantly different. $\chi^2 = 595$, $df = 1$, $n = 3$ colonies. Each colony comprised 144 larvae with three replicates of the control and treatment groups of nine months. A total of 1296 samples from two groups were included in the data analysis. (D) Comparison of the survival function of the colony-resistant control with that of the treatment groups. $\chi^2 = 115.2$, $df = 1$, $n = 3$ colonies. A total of 864 samples with three replicates from the control and treatment groups were included in the data analysis. (E) Comparison of the survival function of susceptible colonies control with the treatment groups. $\chi^2 = 229.5$, $df = 1$, $n = 3$ colonies. The data from these comparisons between the control and treatment groups, and between resistant and susceptible colonies and with the control groups, also showed significant differences ($p < 0.001$). However, comparison analyses of the control groups with the control groups of all colonies were not significantly different ($p > 0.01$).

Table 1. An overview of the sequenced data, identified SNPs, indels, and annotations of 90 samples in *A. c. cerana*.

Data Sequences		SNP and Indel Detection		Annotated SNP, Indels, and Genes	
Base sum	3.37×10^{11}	Raw filter SNP	31,000,613	Raw SNP annotation VCF	1,797,078
Read sum	1,126,922,588	Transition	25,586,296	Filtered SNP annotation GATK	1,048,575
Total reads	2,253,845,176	Trans-version	5,414,317	Indels genotype VCF	424,131
Raw reads	1,127,774,116	Heterozygosity	18,163,8808	Raw filter Indels annotation GATK	424,131
Clean reads	1,126,922,588	Homozygosity	12,836,805	Gene with Indels	91,114
Ave. GC	34.16%	SNP genotype info cloud	872,228	Genes with nonsynonymous SNP	308,336
Ave. coverage 14X	99.04%	DEG SNP	1,769,990	Annotated genes of 90 samples	9359
Ave. mapped	87.92%	Final SNP genotype	696,352	300 ≤ length < 1000	2997
Ave. properly mapped	82%	Private S SNPs	286,552	Length ≥ 1000	6135
Genome: total	11,036,568	Private R SNPs	275,422	Significance regions entire genome at scaffold level	457

VCF = Variant Call Format, GATK, Genome Analysis Toolkit.

The primary aim of this study was to identify SNPs linked to SBV resistance in honeybees and to explore genomic DNA variations between R and S larvae using whole-genome resequencing. Additionally, this study aimed to identify SNP loci, genotypes, and associated genes to gain a better understanding of the molecular mechanisms underlying SBV resistance in honeybee (*A. c. cerana*) larvae. The results of this study may be useful for accelerating molecular breeding, assisted selective breeding to develop SBV resistance, and as a biocontrol method to manage SBV infection in honeybees.

2. Results

2.1. Confirmation of the SBV Purity and Larval Mortalities

The purity of SBV was validated through multiple methods, including reverse transcription-polymerase chain reaction (RT-PCR), SBV detection, transmission electron microscopy (TEM) characterization, sequencing, and phylogenetic analysis. The results revealed that SBV was the only infection detected in the pure viral samples, and none of the six common bee viruses (acute bee paralysis virus (ABPV), Kashmir bee virus (KBV), black queen cell virus (BQCV), Israeli acute paralysis virus (IAPV), and chronic bee paralysis virus (CBPV)) were detected. Colony screening for common bee viruses showed single virus infection, multiple virus infections, and no infection (Figure 1A). The q-PCR analysis of viral particles in the purified samples yielded 6.32×10^{12} copies/ μL , and the generated standard curve detected SBV. The phylogenetic analyses are shown in Figure S1A–D. Inoculated larvae

clearly exhibited SBV symptoms, while pupae were uninfected (Figure 1B). To assess the efficient infectiousness of the purified virus, a comparison of the survival time between the treated (inoculated with SBV) and control groups (virus-free) in three colony larvae were conducted. The results showed a significant difference (Figure 1C); however, the comparisons of the control within the group were not significantly different ($p > 0.01$). The comparison of survival function between the treatment and control groups of resistant (CR) and susceptible (CS) colonies showed a difference. However, the mortality rates of the control groups between CR and CS were not significantly different (Figure 1D).

2.2. Resequencing and Data Quality Statistics Compared with the Reference Genome

The resequencing of 90 individual samples from *A. c. cerana* S and R larvae was completed, and the results are presented in Tables S1 and S2. The amount of data analyzed was 337.47 gigabytes, with an average Q 30 score of 92.73%. The average mapping ratio of the samples to the reference genome was 87.92%, with an average coverage depth of 14 x and genome coverage of 99.04%. The statistics for data evaluation, sequencing, data volume, sequencing data quality, GC content, genomic comparison, comparison rate, genomic coverage, and genome coverage depth are displayed in Tables S3–S6. The whole-genome resequenced clean reads of R and S samples were compared using the same species, and the average mapping ratio was relatively high, with an average of 90 samples at 87.92%, Table S7. Comparisons between R and S and to the reference genome of *A. c. cerana*, all variants of SNP loci, and Indels are summarized.

2.3. Identified SNP and Indels

The pipeline was utilized to identify all variations through data processing, bioinformatics analysis, and functional annotation, as shown in Figure S2. After sequence alignment and a comprehensive analysis, a total of 31,000,613 SNPs and Indels (Table S8) were detected in the three populations, consisting of 15,596,070 SNPs in R and 15,404,543 SNPs in S. The DNA mutations were categorized into transition and transversion (Ti/Tv). In the R group, the Ti and Tv of SNP were 12,872,420 and 2,723,650, respectively, composed of 9,292,227 heterozygosity and 6,303,843 homozygosity, with an average of 59.56% heterozygosity. In the S group, the Ti and Tv of SNP were 12,713,876 and 2,690,667, respectively, including 8,871,581 heterozygosity and 6,532,962 homozygosity, with an average of 56.88% heterozygosity (Appendix B, Tables A1 and A2). These results indicated differences in SNP numbers and mutations detected in the two groups. To ensure the reliability of the sample SNPs, the number of reads supporting the detected SNPs and the cumulative statistical distribution of the distance of adjacent SNPs are shown in Figure S3. Genome-wide SNP mutations showed that two types, C: G > T: and T: A > C: G, had the highest SNP mutations, while two other types, T: A > G: C and C: G > A: T, had the lowest. The other two mutation types, T: A > G: C and C: G > A: T, had almost the same number of SNP mutations between the 90 samples, as shown in Figure S4. Based on the comparison results of the samples with the reference genome, all the different variants between the samples were summarized, and the DEG SNP list file format is shown in Table S9. The unique SNP genotypes after filtering are shown in Table S10. Private SNPs in the R group contained 275,421 SNPs (Table S11), and 286,551 SNPs were detected in the S group (Table S12).

2.4. SNP and Indels Annotation

Based on the location of the mutation site in the reference genome and the location of the gene in the reference genome, we can identify the region of the genome, including intergenic regions, coding sequence regions (CDS), and the impact of the mutation, such as synonymous and nonsynonymous mutations. Using the VCF format file as both input and output, we can assess various fields in the info column of the VCF file, including EFF = Effect (effect impact, functional class, codon change, amino acid change, amino acid, length gene name, transcript, biotype gene coding, transcript ID, exon rank, genotype number, and error warning), which are added as part of the output. All 1,048,575 SNPs

and 424,131 indels identified were annotated, and gene information was obtained from the reference genome to detect and annotate the location of small indels in CDS regions. The results show that most of the SNPs and indels were annotated in intergenic regions. Figure 2 displays the annotation statistics for all SNPs and indels, and Tables S13 and S14 provide detailed annotations of the SNPs and indels across CDS and the genome for the 90 samples.

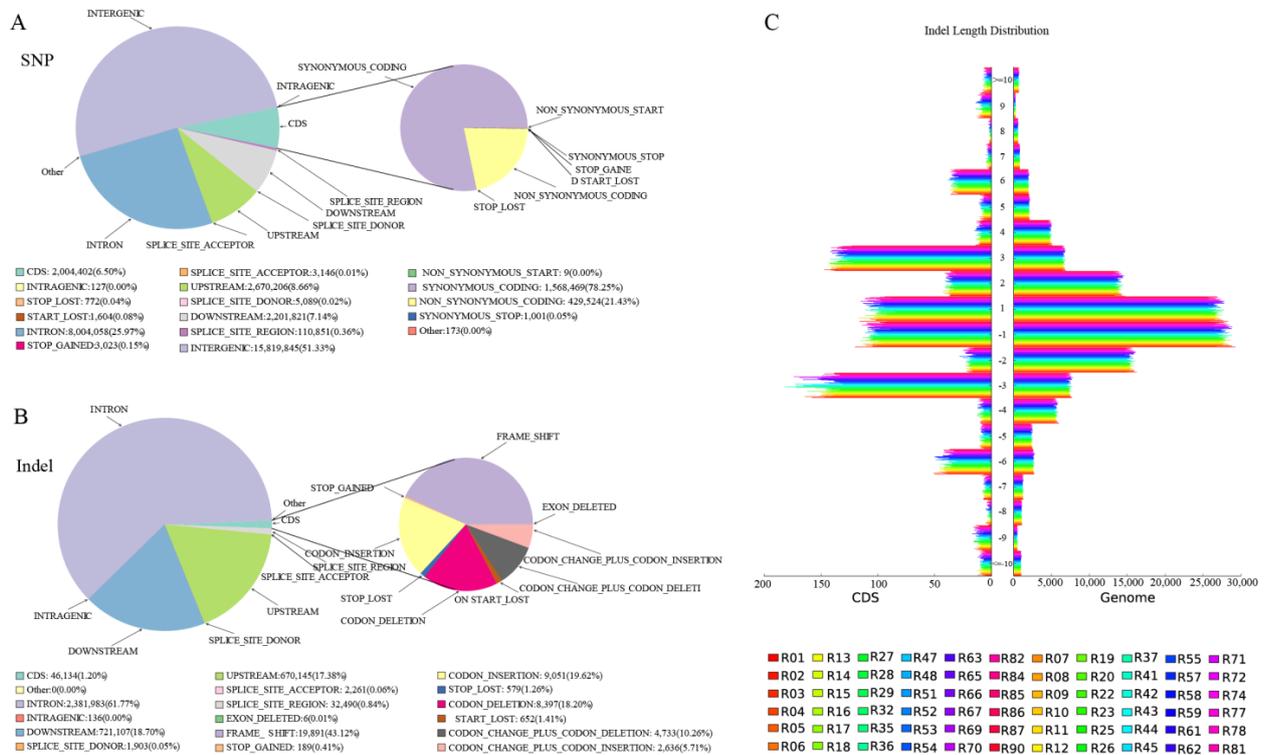


Figure 2. Genome variations, annotation, and indel length distribution map in CDS, condign area, and the whole genome. (A) A plot of all SNP annotations. (B) A plot of all indel annotation distributions. (C) Indel length distribution in CDS and the genome. The longitudinal Y-axis coordinates are indel lengths (within 10 bp); greater than 0 is an insertion and less than 0 is a deletion. The horizontal X-axis coordinates are the corresponding quantities. R represents resistance and S represents the susceptibility of 90 samples in the three populations of *A. c. cerana*.

2.5. Variation in the Gene at DNA Level Analysis

Variations in the CDS may lead to changes in gene function and genes with functional differences between the sample and the reference genome. We can find genes that may have functional differences by looking for CDS and variant region (SV) nonsynonymous mutated genes between the reference genome and the sample. The differences in genes between the R and S individuals of 90 samples compared with the reference genome are shown in Appendix B, Tables A3 and A4. Significant regions and predicted regions were detected based on the combination of F_{ST} and compared with the reference genome, and $INT < 0.5$ and $MAF < 0.05$ with their significant region genes are shown (Table S15).

2.6. Annotation of Variant Genes

The Basic Local Alignment Search Tool (BLAST) was used to compare mutant genes with functional databases [34] such as SwissProt [35], the Gene Ontology (GO) database [36], Clusters of Orthologous Groups of proteins (COG) [37], and the Kyoto Encyclopedia of Genes and Genomes (KEGG) [38]. The annotations of the genes were obtained from these public databases to analyze their functions. The annotation results of all genes with a comparison of the mentioned databases and the integrated functional analysis for all genes

are shown in Table S16. Functional and enrichment analysis of significant SNPs and their significant target genes based on the combination of F_{ST} and π from COG, KOG, GO, and KEGG and the results of the classification are shown in Tables S17 and S18 and Figure S5.

2.7. Genetic Evolution and Diversity

To investigate the evolutionary relationships between R and S from the same species, a phylogenetic tree was constructed based on a neighbor-joining (NJ) tree composed of 45 R and 45 S samples from three populations of all informative SNPs used in the phylogenetic analysis. The results showed that the group samples from three colonies were well differentiated in clades; however, the R and S from the sample colony were still slightly mixed together and differentiated from other colonies. Therefore, the three populations were the same species and in the same location, and there was no distance between the sampled geographic areas. However, the phylogenetic tree of SNP validation in the field from different geographic locations and a comparison of RC and SC showed a distinct evolution between R and S larvae, which were well differentiated into separate clades and branches (Appendix A, Figures A1 and A2).

2.8. Population Structure

A population structure analysis can quantify the number of ancestors of the studied populations and infer each sample's blood origin. It is currently a more widely used group cluster analysis method that helps us understand material evolutionary processes. The number of subgroups (K-value) was set from 1 to 10 for the clustering of the study population. The optimal number of groups was determined based on the valley value of the cross-validation error rate in the admixture, and the individual ancestries from SNP genotypes were investigated. The clustering results of K-values from 1 to 10, the cross-validation error rate corresponding to each K value, and the model K = 6 obtained from 90 samples in the two groups is shown in (Figure 3C,D).

2.9. Principal Components Analysis

Through a principal component analysis (PCA), it is possible to determine which samples are relatively close and which samples are relatively distant, which can assist in understanding the evolution between traits. The results showed that total of 20.76% variance between R and S. PCA 1, PCA 2, and PCA 3 showed some differences. Most of the SNPs in R were relatively separately clustered; however, simultaneously, the clusters were close to S in distance. A similar result was observed in the phylogenetic tree PCA 3D plot (Figure 3E). To estimate the relative kinship between two individuals in a natural population we use the affinity metric. The affinity itself is a relative value that defines the genetic similarity between two specific materials and the genetic similarity between any material, so when the kinship value between the two materials is less than 0, it is directly defined as 0. The frequency distribution of the kinship value is shown in Figure S6.

2.10. Selective Sweep and Selection Signature

In the selective sweep analysis, due to the selection-specific site (natural selection/manual selection), the frequency of the neutral linkage relationship of neutral mutation increased. In contrast, the frequency of no linkage neutral site mutations decreased, resulting in a decrease in polymorphism in this region. The method for detecting selective sweep regions in the genome-wide range is to calculate the population genetic indicators of all SNP sites within a sliding window (e.g., 100 kb). A specific step size (e.g., 10 kb) was chosen, such as F_{ST} and θ_{π} based on the calculation results, and a particular region was selected (Table S19). The samples were divided into two groups for selective sweep analysis, and the combination of F_{ST} and θ_{π} from the selective sweep analysis in the two groups is shown in Figure 3F. Tables 2 and 3 show the distributions of G versus C and C versus T allele frequencies of SNP KZ288474.1_322717 and KZ288479.1_95621, which were selected and validated, and colonies were selected to validate these two SNPs (Table 4).

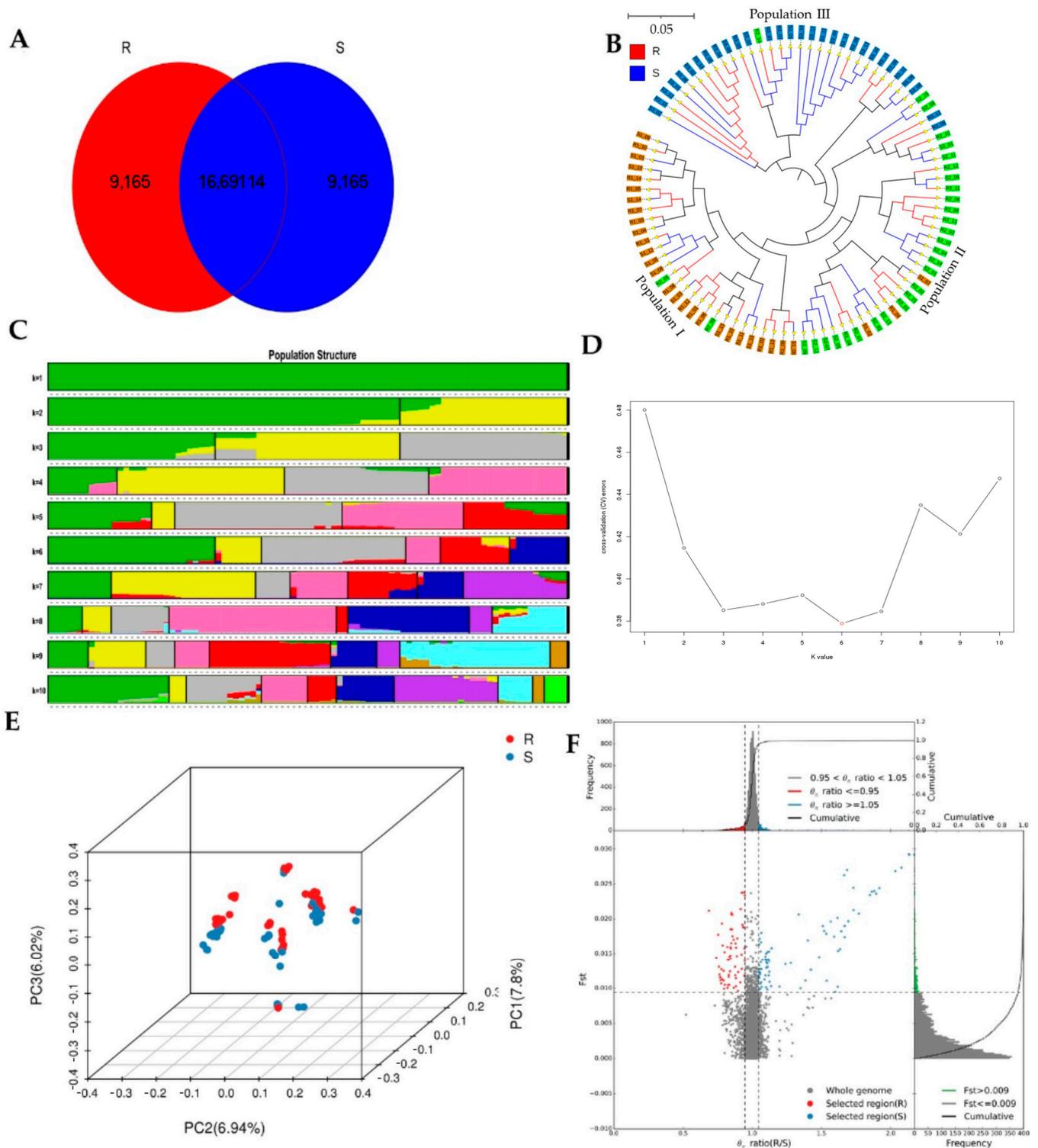


Figure 3. Venn diagram of unique SNPs, phylogenetic tree, population structure, principle component analysis, and selection signature for resistance to SBV. **(A)** Venn diagram of unique SNPs and shared SNP genotypes between R and S. By using 1,797,078 VCFs and considering the 50% optimal genotype in the R group, we obtained 9165 unique SNPs in R and S, and 669,114 SNP genotypes were not found to be the cause of 118,801 SNP loci in at least one subgroup that was under 50% optimal genotypes. Such loci were not involved in analyzing the group. **(B)** A phylogenetic neighbor-joining tree was constructed using 696,352 SNPs with 1000 bootstrap replicates of three populations. **(C)** Population structure and admixture of 90 samples between R and S from three populations and the cross-validation

error rate. The x -axis represents the K values from 1 to 10 and the y -axis represents the cross-validation (CV) errors between 0.38 and 0.48, where the best K was 6. The accessions were divided into ten subgroups (there was a minimum K -value when $K = 6$); within each subgroup, the accessions were ordered according to the genetic component, and each line gives the subgroup value. Each accession shown as a vertical line partitioned into K colored components represents inferred membership in K genetic clusters. (D) A 3D principle component analysis (PCA) plot was constructed using SNP genotypes of the two groups (R and S) with a 3D plot. (E) Distribution of selective sweep and population differentiation index (F_{ST}) with nucleotide polymorphism (θ_π) ratio selection regions of genes and SNPs for the top 5% R (resistant) and S (susceptible) *A. c. cerana* larvae and (F) The horizontal coordinate represents the ratio of θ_π (π ratio); the vertical coordinate represents the F_{ST} value, corresponding to the frequency distribution diagram above and the frequency distribution diagram on the right, respectively; and the dot plot in the middle represents the corresponding ratio of F_{ST} to θ_π in different windows. Gray represents the whole genome; the red and blue areas at the top are the top 5% regions selected by θ_π , the green area is the top 5% region selected by F_{ST} , and the blue and red areas in the middle are the intersection of F_{ST} and θ_π , which are the candidate sites. The distributions of threshold line $F_{ST} < 0.95$, $\theta_\pi < 1.05$, F_{ST} , and $\theta_\pi < 0.95$ ratios of the top 5% strong signal region analysis on genome selection of genes were calculated in 10 kb windows sliding in 1 kb steps in R and S, respectively.

Table 2. Distributions of G and C allele frequencies of SNP KZ288474.1_322717.

Category	Colony	Sample Size	Frequency of Genotypes			Frequency of Alleles ^c						Remark
			G/G	G/C	C/C	G	P_G ^a	Mean \pm SD	C	P_C ^b	Mean \pm SD	
R	R1	48	26	20	2	72	0.750	0.843 \pm 0.066	24	0.250	0.156 \pm 0.066	First and second validation
	R2	48	37	11	0	85	0.885		11	0.115		
	R3	48	38	10	0	86	0.896		10	0.104		
S	S1	48	4	40	4	48	0.500	0.580 \pm 0.064	48	0.500	0.420 \pm 0.064	
	S2	48	8	40	0	56	0.583		40	0.417		
	S3	48	15	33	0	63	0.656		33	0.344		
CR	CR1	64	53	11	0	117	0.914	0.898 \pm 0.017	11	0.086	0.102 \pm 0.017	Comparative validation
	CR2	64	48	16	0	112	0.875		16	0.125		
	CR3	64	52	12	0	116	0.906		12	0.094		
CS	CS1	64	28	31	5	87	0.680	0.685 \pm 0.026	41	0.320	0.315 \pm 0.026	
	CS2	64	36	20	8	92	0.719		36	0.281		
	CS3	64	25	34	5	84	0.656		44	0.344		
R	R1	162	104	54	4	262	0.804	0.862 \pm 0.043	62	0.196	0.138 \pm 0.043	Further validation
	R2	32	26	6	0	58	0.906		6	0.094		
	R3	32	25	6	1	56	0.875		8	0.125		
S	S1	112	45	45	22	135	0.603	0.648 \pm 0.059	89	0.397	0.352 \pm 0.059	
	S2	32	12	14	6	38	0.731		26	0.269		
	S3	32	9	21	2	39	0.609		25	0.391		

R, resistant S, susceptible CR, resistant colony CS, susceptible colony, ^a frequency of the G allele (P_G), ^b frequency of the C allele (P_C), and ^c the differences between C and T allele frequencies and colony category in exact Chi-square test at a p -value of <0.05 was defined as a significant difference. The overall results showed a significant difference between the resistant allele (G) in R and susceptible allele (C) in S, $\chi^2 = 15.595$, $df = 1$, $p = <0.000078$.

2.11. Significant Regions between Resistance and Susceptibility

Based on the comparison results between R and S using multiple tests of F_{ST} and θ_π , independent and combined F_{ST} , and θ_π genome-wide analysis of each candidate site showed that R and S are considerably different from the reference genome in the 457 scaffolds region of the chromosome-level selective sweep analysis. Comparison results of each chromosomal region were extracted at 19,551 positions. A selective sweep analysis showed that the window and step size statistics intersection in 6,464 scaffolds between R and S were significantly different (Figure 4). Functional and enrichment analysis results of the selected regions are shown in Figure 5.

Table 3. Distributions of C and T allele frequencies of SNP KZ288479.1_95621.

Category	Colony	Sample Size	Frequency of Genotypes			Frequency of Alleles ^c						Remark
			C/C	C/T	T/T	C	P_C^a	Mean \pm SD	T	P_T^b	Mean \pm SD	
R	R1	48	25	23	0	73	0.760		23	0.240		First validation and second validation
	R2	48	22	26	0	70	0.729		26	0.271		
	R3	48	37	11	0	85	0.885	0.792 \pm 0.068	11	0.115	0.208 \pm 0.068	
S	S1	48	4	38	6	46	0.479		50	0.521		
	S2	48	4	40	4	48	0.50	0.542 \pm 0.074	48	0.50	0.458 \pm 0.074	
	S3	48	14	34	0	62	0.646		34	0.354		
CR	CR1	64	46	18	0	110	0.859		18	0.141		Comparative validation
	CR2	64	38	24	2	100	0.781	0.757 \pm 0.094	28	0.219	0.242 \pm 0.093	
	CR3	64	23	35	6	81	0.633		47	0.367		
CS	CS1	64	35	18	11	88	0.688	0.614 \pm 0.071	40	0.313	0.385 \pm 0.077	
	CS2	64	28	27	9	83	0.648		45	0.352		
	CS3	64	1	63	0	65	0.508		63	0.492		
R	R1	162	113	47	2	273	0.843		51	0.157		Further validation
	R2	32	8	16	8	32	0.50		32	0.50		
	R3	32	25	7	0	57	0.891		7	0.109		
	R4	32	21	10	2	52	0.797	0.781 \pm 0.144	14	0.203	0.219 \pm 0.144	
	R5	16	13	2	1	28	0.875		4	0.125		
S	S1	112	50	41	21	141	0.629		83	0.371		
	S2	32	5	17	10	27	0.422		37	0.578		
	S3	32	15	17	0	47	0.734	0.570 \pm 0.105	17	0.266	0.430 \pm 0.105	
	S4	32	8	18	6	34	0.531		30	0.469		
	S5	16	4	9	3	17	0.531		15	0.469		

R, resistant S, susceptible CR, resistant colony, CS, susceptible colony, ^afrequency of C allele (P_C), ^b frequency of T allele (P_T), and ^c the differences between C and T allele frequencies and colony category in exact Chi-square test at a p -value of < 0.05 was defined as a significant difference. The overall results showed a significant difference between the resistant allele (C) in R and the susceptible allele (T) in S ($\chi^2 = 8.931$, $df = 1$, $p < 0.0028$).

Box plots of F_{ST} and π R and S selected genomic regions with selective sweep signals versus the whole genome results are shown in Figure 4. The box represents the range of quartiles (IQR) between the first and third quartiles (25th and 75th percentiles, respectively), and the line in between represents the median. Whiskers represent the lowest and highest values within the first and third quartiles ($1.5 \times IQR$). The outliers beyond whiskers are shown as hollow dots. A Mann–Whitney U test was used to calculate statistical significance, with a p -value of < 0.05 considered significantly different.

2.12. Flanking SNP between Resistance and Susceptibility

Due to the large number of SNPs, we considered narrow significant regions, and a high percentage of allele frequency was selected. Additional statistical analyses were conducted for the best difference between R and S compared with the reference genome. F_{ST} and π S divided the top 1% R 84 SNP positions, and F_{ST} π R divided the top 1% S 735 SNP positions (Table S19). When F_{ST} π S was divided by R and F_{ST} π R was divided by the S top 1% or > 0.99 , the best difference of 10 and 19 SNPs, respectively, was significantly different.

2.13. Verified SNPs

To verify the candidate SNPs, we selected 10 SNPs based on the combined F_{ST} and π S divided R top 1%, and 10 SNPs based on the combined F_{ST} and π R divided S top 1% with a high percentage genotype. After verifying the first candidate SNPs, none were found to be significantly different. We then performed SNP validation for the second group of candidate SNPs. Among them, two SNPs (GenBank IDs and positions: KZ288474.1_322717, $G > C$, and KZ288479.1_95621, $C > T$) and their corresponding alleles (based on their allele frequencies) were significantly different in the three colonies with the same colony samples used for DNA sequencing. Further SNP validation was performed to ensure the reliability of these two SNPs with comparisons of R and S larvae from different geographic locations

and populations. The results showed a highly significant difference between the R and S larvae.

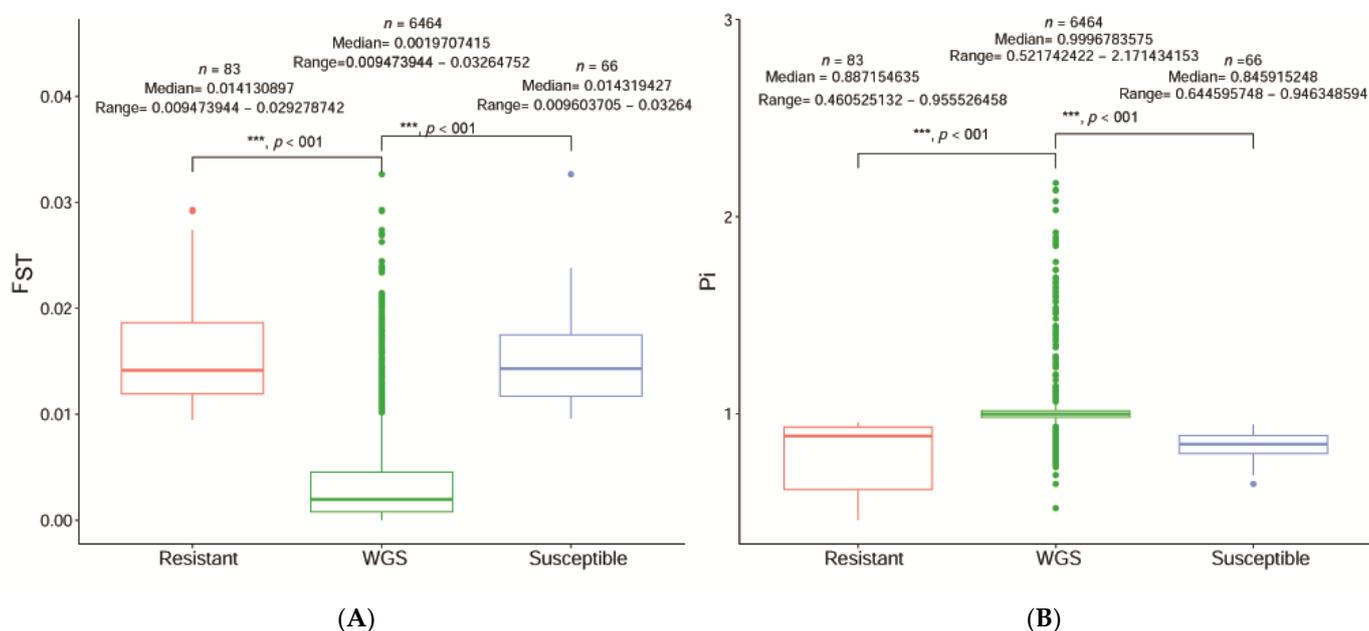


Figure 4. Distributions of selective sweep analysis results of significant regions of intersection F_{ST} and π ratios between R and S in *A. c. cerana* larvae that had undergone positive selection the top 5% of the whole genome. (A,B) Comparison analysis of selected regions based on population differentiation index (F_{ST}) and nucleotide polymorphisms (π) of the significant intersecting regions (top 5%) of genome variants between R (resistant) larvae and S (susceptible) larvae of *A. c. cerana*. Data were reanalyzed and plotted from analyzed significance regions.

According to SNP annotation, the first SNP (KZ288474.1_322717) was an upstream effect transcript ID = RNA 9701 with a gene close to the *APCC_06899* gene at 1804 bp distance. The second SNP was an intergenic SNP relatively far from the *APCC_01833* gene with a 53,591 bp distance.

The comparison results of multiple SNP validations from the same individual sample and different samples from different geographic locations were consistent between two SNPs in the R, S, and genotyped queens conformed to each other with a slight change.

In addition, in the R samples, the frequencies of G allele at the SNP KZ288474.1_322717 position and C allele at the SNP KZ288479.1_95621 position were high percentage allele frequencies. On the other hand, in the S samples, high percentage allele frequencies of C and low percentage allele frequencies of T were observed; however, in the R samples, C and T at two SNPs position were very low and even were not observed in some colonies. A comparison of two SNPs with same sample was conducted, which indicated the reliability of each SNP and the accuracy of the Sanger sequencing method and supported the reliability of the first SNP detection and validation results.

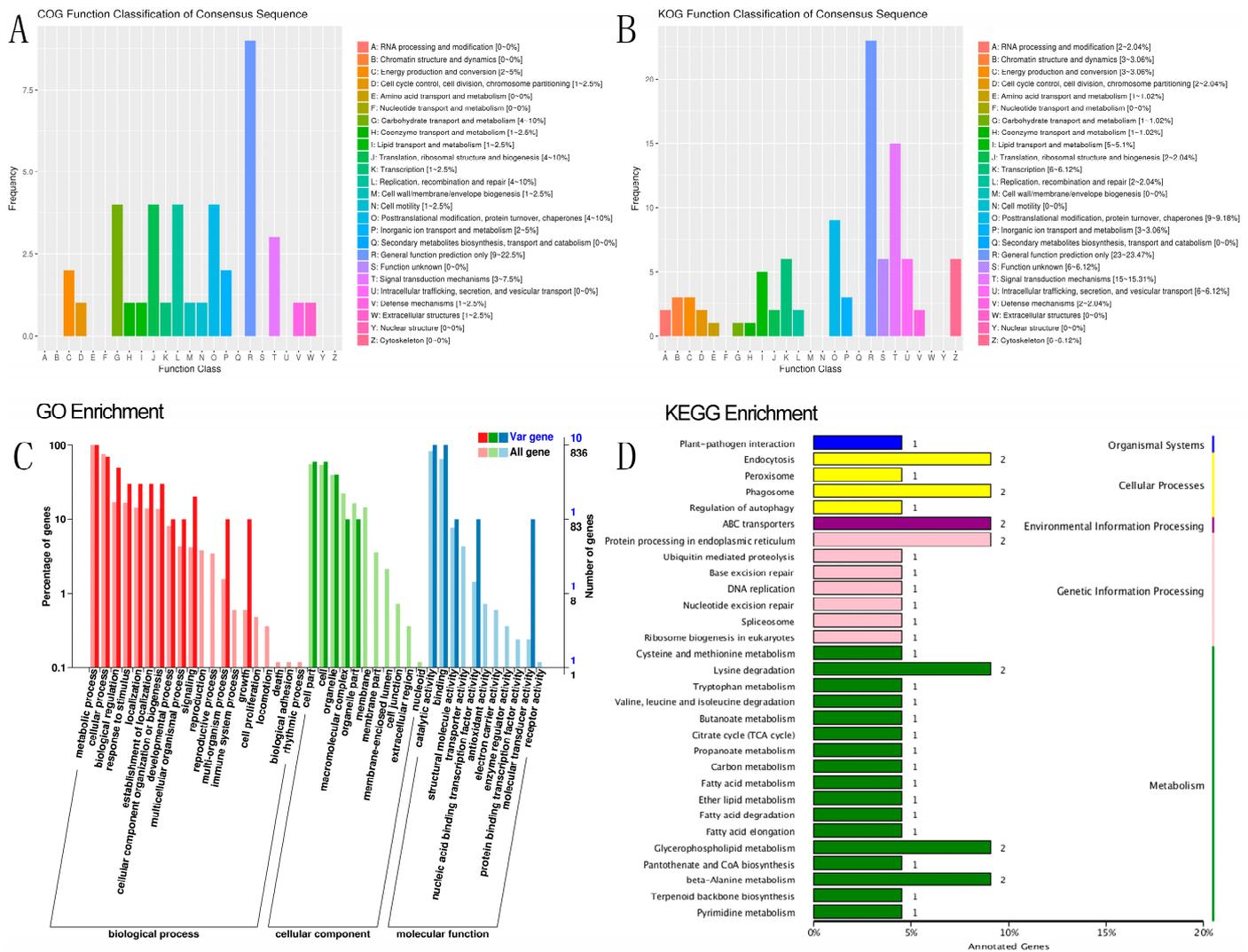


Figure 5. (A) Clusters of Orthologous Genes (COG); (B) Gene Ontology (GO); (C) KOG (Eukaryotic Orthologous Groups), and (D) Kyoto Encyclopedia of Genes and Genomes (KEGG) of top 5% population differentiation index (F_{ST}) and nucleotide polymorphisms (π) in *A. c. cerana*.

3. Discussion

In this study, we challenged larvae of honeybees (*A. c. cerana*) with SBV from 90 individual samples from three populations in Fuzhou, China, then resequenced and analyzed them to screen SNP genetic markers associated with SBV resistance [28,29,32]. The samples were resequenced using an Illumina Nova His Seq X-Ten for sequencing (150 bp PE), and 31,000,613 SNPs were identified, including 18,163,808 (42%) homozygous SNPs, 12,836,805 (58%) heterozygous SNPs, 5,414,317 transversion SNPs, and 25,586,296 transition SNPs, and 9359 predicted genes from 90 individuals were annotated [39]. To identify genetic markers associated with SBV resistance, we compared the VCF list of 179,7078 SNPs between R and S, which was identified by GATK [40], with a comparison of the reference genome $INT < 0.5$ and $MAF < 0.05$ using multiple tests, especially F_{ST} and π , independently and in combination using Popgenome [40,41]. We determined significant regions of each SNP and their target genes at the level of the entire genome chromosome and the relevant sections at each chromosomal level. After identifying 6464 relevant significant regions at the scaffold level between R and S, we combined the F_{ST} and π ratios. All regions were significantly different for segregation sites and the INT was 0.05 and the MAF was < 0.05 for R and S. The significant regions under artificial selection with F_{ST} and π 0.99

and 0.95 (top 1% and top 5%) were then selected by integrating the F_{ST} and π statistics. We cataloged 84 unique SNPs with seven predicted genes and 735 unique SNPs, with ten anticipated genes based on combined F_{ST} and π . F_{ST} and π R divided S and F_{ST} and π S divided R were significantly detected in 90 samples.

To identify and validate the SNPs of interest, we selected the top ten SNPs with the highest proportion of genotypes based on the F_{ST} and π analyses, specifically the F_{ST} and π R divided S top 1% and the F_{ST} and π S divided R top 1%. Among these, we found two significant SNPs, KZ288474.1_322717 (G > C) and KZ288479.1_95621 (C > T), located on the 15th chromosome with GenBank IDs, which were associated with SBV resistance based on their allele frequencies.

The validation of the SNPs began with 16 R and 16 S samples from three colonies, followed by validation with 32 R and 32 S samples from the same colonies using DNA sequencing. Out of the 20 candidate SNPs, three were not well amplified, and 15 were not significantly different, shown in Table S21. For field validation, one SNP (KZ288474.1_322717) was selected from 32 R and 32 S regions of three asymptomatic colonies in Fuzhou. During larval rearing, we predicted that the three colonies might be resistant to SBV, as only a few larvae were infected with the virus at the same dose. The experiment was repeated with a double dose of the virus, and after sequencing, evaluation, and analysis, no significant difference was observed between R and S in allele frequency ($p > 0.05$), indicating that the three colonies were resistant to SBV. For the second SNP (KZ288479.1_95621), three symptomatic colonies were selected in Minhou after larval rearing, sequencing, and analysis. The allele frequency was significantly different. During the validation process, it was discovered that the honeybee colony might have evolved naturally to become resistant to SBV and not carry the susceptible allele C/C at the SNP KZ288474.1_322717 position. Each colony or population exhibited different levels of resistance and susceptibility [27].

Therefore, to further validate the SNPs and assess their association with SBV resistance across different geographic locations, we selected additional colonies from various regions. The SNP validation data showed significant variations in allele frequency between 32 R and 32 S samples in each symptomatic colony. Asymptomatic colonies had high frequencies of G/C and C/C alleles but very low frequencies of G/G alleles, while symptomatic colonies had low allele frequencies. In addition, asymptomatic colonies had a high C/C allele frequency at both SNPs, whereas symptomatic colonies had a high T/T allele frequency at the second SNP (C/T). To investigate whether these genotypic and phenotypic characteristics were consistent across different geographic regions, we obtained symptomatic samples from three colonies in Yunnan Province, located 2306.5 km away from Fuzhou City. Although these colonies were infected with DWV and did not show visible SBV symptoms, we selected R samples (pupae) for analysis. Screening and analysis of both SNPs revealed that most of the pupae samples had G/G and C/C alleles, and susceptible alleles (C/C and T/T) were not detected. These results provide further evidence of the association between these SNPs and SBV resistance, regardless of geographic location.

Additional evidence was gathered from 14 resistant queens from Nanping and three apiaries, as well as one apiary in Fuzhou. Genotyping revealed that only one queen carried the G/C and C/T genotypes for the two SNPs, while the remaining queens had G/G and C/C genotypes. These results, combined with those of the genotyped queen, suggest that honeybee colonies with homozygous resistant G/G alleles at the SNP position KZ288474.1_322717 and C/C alleles at the SNP position KZ288479.1_95621 are likely to be strongly resistant to SBV. Conversely, colonies with homozygous susceptible alleles (C/C and T/T) are expected to be extremely susceptible to SBV, while those with heterozygous susceptible alleles (G/C and C/T) may only be moderately susceptible. These findings underscore the importance of choosing colonies and queens for breeding programs.

The fact that the alleles generating SBV resistance in both SNPs were found to be correlated with one another in the 90 samples of genotype sequences and SNP validation highlights the accuracy of these SNPs. To further ensure their accuracy, we compared the results of 210 samples for the two SNPs using the same sample, as well as 48 samples

from two different sequencing companies. We aimed to verify whether the G/G allele at SNP position KZ288474.1_322717 and C/C allele at SNP position KZ288479.1_95621 corresponded with one another, and vice versa. We found that the results from both groups were nearly identical after peak assessment, indicating that sequencing errors were minimized. We further validated the successful SNPs in a large number of samples from various geographic locations to confirm their reliability. The use of these SNPs as valuable genetic markers in selective breeding programs for improving multiple honeybee traits, including SBV resistance, has been suggested in previous studies that found the same corresponding alleles related to SBV resistance, high royal jelly production, and chalk-brood resistance in *A. mellifera* [26,27,31].

Late SBV infections were observed in the prepupae during the larval-rearing experiment. The larvae and prepupae were not identified as SBV-resistant until the larvae had fully transformed into pupae at the 12–14 day instar and were then individually sampled as R [42]. The mortality rate of larvae in the susceptible colony was high (39%), in contrast to the low mortality rate in the resistant colony (28.12%) after inoculation. The survival rates of the control groups in the CS and RC were relatively high, at 85% and 88%, respectively [43].

Population Genetics and Variations

The clean reads were mapped to the reference genome to obtain the mapped reads. A high sequencing read ratio (~70%) indicates appropriate reference genome selection and the absence of contamination during the experiment. The assembly of the reference genome is important for the quality of sequencing reads and the locatability of reads in the reference genome, which is directly proportional to the degree of similarity between the species. On average, the mapping rate of the sequence data was 87.92% [44]. Compared to SNP variation, indel variation was generally less frequent. Most SNPs/Indels were annotated as intergenic variations, with fewer SNPs found in introns. A detailed annotation of SNP/Indels is provided in [45].

Whole-genome sequencing data were utilized to investigate the evolutionary relationships among three populations using various analyses, including phylogenetic tree, population structure and admixture, PCA, and kinship analyses. These analyses were conducted to describe the taxonomic and evolutionary relationships between the traits that were assumed to have a common ancestor. The three populations showed clear differentiation in clades with some mixing within groups. Additionally, some distance was observed in the ancestries between the three populations, and the best $K = 6$ was obtained in admixture analysis [39].

The mRNA level genes were evaluated using Illumina sequencing data. Selective sweep, $INT < 0.5$, and $MAF < 0.05$ were applied to obtain highly consistent genes. All genes within 100 kb of significant SNP regions were extracted, and their annotations were considered. All predicted gene directions were upregulated. The top 1% of genes based on combined F_{ST} and π were selected for SNP validation, considering the narrow range of significant regions and the high allele frequency. The SNP genotypes were present in the R group, and they differed from the selected verified reference genome alleles in the S group. However, only a few genes were annotated, which could not explain the more positive effects; therefore, we considered the top 5% annotation results from the same site where we detected two SNPs [31].

SNP and indel annotation, as well as a prediction of different genes, were performed on individual samples from R and S, comprising 90 annotated samples, and on the specific region detected by the combined F_{ST} and π top 5%. The functional annotation of 90 individual samples was conducted across 10 databases of 9359 different genes, including integrated functional annotation statistics, GO gene list 1399, function, 1812 in the biological process (BP), 820 in cellular component (CC), and 10,583 unique genes in molecular function (MF). In the 116 KEGG pathways, 2,908 and 3842 unique genes with positive effects were annotated, most of which were involved in hypothetical protein signal transduction mechanisms [45,46]. Some of the most enriched pathways in KEGG were spliceosome (111

genes), RNA transport (105 genes), purine metabolism (103 genes), ribosome (98 genes), endocytosis (97 genes), carbon metabolism and protein processing in endoplasmic reticulum (96 genes), ubiquitin-mediated proteolysis (82 genes), and phagosome (55 genes). The other genes involved in different pathways totaled than 50 genes. Some of these pathways have a significant effect on disease response [47–50]. Based on the functional annotation of 90 individual samples compared to the reference genome of *A. c. cerana*, 9359 different predicted genes against ten databases were annotated, which is close to the previously identified and annotated genes in the *A. cerana* genome (9627) [51]

Based on the F_{ST} and π values, R and S were divided into the top 5% gene annotations, resulting in 119 genes, of which 116 were annotated in various databases, such as COG, GO, Nr, Nt, Pfam, SwissProt, and TrEMBL (Tables S17 and S18). Out of the 844 genes annotated for biological processes, 49 were related to single-organism cellular processes (GO:0044763), 28 were associated with transport (GO:0006810), and 51 were involved in metabolic processes (GO:0008152), with the first two categories being statistically significant ($p < 0.05$). Notably, *APICC_03984* was found to be involved in the activation of MAPKK activity, pathogenesis, cellular hyperosmotic response, and signal transduction involved in filamentous growth in the biological process category. The curated GO BP annotation dataset shows that 63 genes participate in the activation of MAPKK biological processes, such as HRAS, JAK2, and BRAF. In the cellular component category, 79 out of 520 genes were related to the cytoplasm (GO:0005737), with the most enriched genes involved in metabolic processes (10), catalytic activity (10), cellular process (7), binding (10), biological regulation (5), response to stimulus (3), and signaling (2). The other genes were involved in different positive effects, and no genes were involved in the death pathway. *APICC_00122* was found to be involved in exopeptidase activity. While the insect midgut has been primarily associated with digestion and detoxification, endopeptidases such as serine proteases (trypsin and chymotrypsin-like) and exopeptidases with varying terminal amino acid specificity (aminopeptidases and carboxypeptidases) are believed to play a key role in protein hydrolysis [52].

In the GO molecular function category, two genes were significantly enriched in ATPase activity (GO:0016887). Among the 90 genes in KOG, 46 were annotated in KEGG, with ABC transporters (*APICC_00398* and *APICC_04707*) having the highest enrichment factor (14.77) and a q-value of 0.052. Other enriched pathways included beta-alanine metabolism (*APICC_01290* and *APICC_10032*), lysine degradation, glycerophospholipid metabolism, phagosome (*APICC_03257* and *APICC_08196*), fatty acid metabolism (*APICC_10032*), protein processing in the endoplasmic reticulum, and endocytosis (*APICC_01294* and *APICC_10111*) [53,54].

The genome of *A. cerana* has been previously reported to have a de novo assembly with 10,182 predicted genes [29]. Another study reported the genome sequencing and assembly of *A. cerana* with a data size of 228 Mbp and 879 scaffolds [17]. The *A. c. cerana* genome still needs further sequencing and improvement. Although the genomic size of *A. cerana* is slightly smaller than that of *A. mellifera*, more evidence is required to explain the significant differences between the two species [24].

Bee viruses are usually asymptomatic but can significantly affect honeybee health and their lifespan under certain conditions. While SBV symptoms in larvae are evident, they are not clear in adult bees [55,56]. Worker bees with hygienic behavior can help prevent pathogen dispersal by removing infected larvae, but this behavior can cause infections in other bees, leading to virus transmission through feeding into the larva [57]. Although technical bee management methods, such as requeening, replacing the old brood comb, and using herbal medicines, can help control SBV, they are not very efficient. It is necessary to manage bee colonies to reduce the levels of Nosema disease and control Varroa mites throughout the year [19]. The SBV isolated in this study is valuable for future research on honeybees. Researchers can use pure viruses to expand the study of SBV resistance, study gene function and characterization, and apply CRISPR–Cas9 to edit genes closely related to SBV resistance. Other potential projects include transcriptome analysis of larvae using

purified virus, genotyping using identified SNPs to inoculate the larvae, RNA sequencing, applying SBV to different bee species, and studying the behavior, physiology, and other aspects of honeybees.

4. Materials and Methods

4.1. Colony Selection

Experimental colonies were established at the experimental apiary at the College of Animal Sciences (College of Bee Science), Fujian Agriculture and Forestry University, to conduct experiments. Some samples were directly collected from apiaries. The samples and colonies were selected from Fujian Province and Yunnan Province. The details of sample collection and the number of colonies used in this study are summarized in Table 4.

Table 4. Experimental colony details used in this study.

Date	Apiary Location	Purpose	No. of Colonies	SNP 322717	No. of Colonies	SNP 95621
2017	Yongtai, Fuzhou	Virus purification	4	250	4	1000
	Minhou	Healthy colony selection	8	15	8	120
	Minhou	Larval rearing and genomic DNA sequencing	3	90	3	90
2021	Fujian and Yunnan Provinces	SNP validation	32	926	34	1022

To purify SBV from diseased larvae and to select healthy colonies, this study was conducted in different geographic locations in Fujian Province. We selected three apiaries in Yongtai and one apiary in Fuzhou. A symptomatic colony was sampled from each apiary. Three frames with obvious SBV symptoms were selected from each colony and extracted from the hives. The frames were marked and immediately placed on dry ice. The samples were quickly transferred to the laboratory. From each colony, 250 old, infected larvae were collected in pooled samples. All the samples were directly stored at -80°C for isolation and purification of SBV.

Three out of eight healthy colonies were selected for larval rearing after screening against seven common bee viruses, see the details in the virus isolation and purification section. After larval rearing from three colonies challenged with SBV 90 samples, each colony was composed of 30 samples (15 S larvae between 7–10th instars and 15 R larvae (pupae) at 12–14th instars) randomly selected for genomic DNA extraction and genome sequencing. The first SNP validation was conducted using the same colony samples that were previously used for genomic DNA sequencing, which were composed of 16 R and 16 S. The second SNP validation of each colony was composed of 32 R and 32 S from the same three colonies for SNP 95621 and SNP 322717; two colonies from the same colonies and one extra colony from Fuzhou were used.

Three asymptomatic bee colonies from the Fuzhou area were selected for in vitro culture to verify the effectiveness of these two SNPs in the field. Each colony consisted of 64 samples (32 R and 32 S). By SNP verification, three asymptomatic colonies in Fuzhou and three symptomatic colonies in Minhou were compared. The SNP 95621 was validated in two symptomatic colonies in Putian, one asymptomatic colony in Fuzhou, and three symptomatic colonies in Minhou County. Each population consisted of 64 larval and pupal samples (32 R and 32 S) of SNP 322717.

To verify whether these two SNPs function reliably in different geographical locations, we conducted further SNP verification in three symptomatic colonies (Nanping, Putian, and Fuzhou) and selected SNP 322717. Another SNP (95621) was identified in two symptomatic colonies in Putian, one colony in Nanping, and two colonies in Fuzhou. Each colony consisted of 32 R and 32 S. Samples in frame form, one symptomatic and two asymptomatic from Yunnan Province, were obtained by express, and 48 R samples (pupae at the white eye stage) were randomly sampled from combs and analyzed.

Further SNP validation in resistant queens was conducted on 14 queens from four apiaries (three apiaries in Nanping: twelve queens and two apiaries in Fuzhou: two queens), which were directly collected from the hives, kept in queen cages, and transferred to the laboratory. In addition to comparing the genotypes, we investigated whether the allele frequencies of the two SNPs corresponded to each other and whether the Sanger sequencing was accurate. Of 210 samples from R and S, the queen samples included in SNP validation were compared separately. Overall, a total of 926 samples for SNP 322717 and 1022 samples for SNP 95621 were used for SNP verification in this study.

4.2. Isolation and Purification of Sacbrood Virus

To isolate SBV, all necessary materials were autoclaved before virus purification. Prior to purification, 15 infected larvae were randomly selected. Additionally, 15 uncapped larvae were sampled from eight candidate colonies. Total RNA was extracted, cDNA was synthesized, and seven common bee viruses were screened using the RT-PCR method, as described by Hassanyar et al. [58]. After screening, two colony samples that were positive for SBV and negative for the other viruses were used for SBV purification. We detected other viruses (DWV and IAPV) in the two samples, including SBV, and these samples were discarded. To purify SBV, we followed the protocol described by Maori et al. with some modifications [59]. Briefly, all the samples were separately ground into powder using liquid nitrogen with a mortar and pestle and then homogenized in 0.01 M Na-phosphate (pH 7.6) and 0.2% Na-deoxycholate (pH 7.2) (Sangon Biotech, Shanghai, China). The samples were transferred into centrifuge tubes (Axygen, USA) and centrifuged at $800\times g$ for 20 min. The supernatants were collected in new clean tubes and the pellets were discarded. The high-speed centrifuge (HITACHI ultracentrifuge CPN 80X, Japan) was turned on for a while to cool to 4 °C. The supernatant was transferred into ultracentrifuge tubes and balanced before centrifugation at $27,000\times g$ for 4 h. The supernatants were discarded, the pellets were dissolved using 0.4% sodium deoxycholate acid (Na-DCA), and 4% Brij 58 (Sangon Biotech, Shanghai, China) was added, vortexed, and centrifuged again at $800\times g$ for 15 min for clarification. The supernatants were carefully transferred into the same tubes we used before, then 5.5 g cesium chloride (CsCl) was added to each tube, well-mixed, then balanced, and transferred into the ultracentrifuge again and centrifuged at $27,000\times g$ for 24 h at 18 °C. Two typical cloudy bands were formed in the supernatant and in the middle of the tubes. The bands were carefully collected separately in fresh tubes. To remove CsCl from the solution, three dialysis cycles were performed in double-distilled water (ddH₂O) at 4 °C, and the water was replaced with fresh water after eight hours. Then, the solution was prepared in different vials (250 mL, 500 mL, and 1000 mL) and kept at –80 °C for viral load quantification and inoculation.

4.3. Confirmation and Verification of SBV Purity

The first observation of SBV was conducted using TEM, following the protocol developed by Zhou et al. [60]. Briefly, 20 µL of the two purified samples was placed in separate fresh tubes, kept on ice, and transferred to the College of Plant Protection, Fujian Agriculture and Forestry University, for observation under TEM. The grids were negatively stained with 2% sodium phosphotungstate at pH 6.8 for 2 min. To prepare the standard staining solution, we used 1% phosphotungstic acid (PTA) at pH 7.2. A drop of the viral suspension was applied onto a formvar- or collodium-covered electron microscope grid for 30 s, and the liquid was absorbed using filter paper. A drop of the staining solution was used after a few seconds, then absorbed and dried. The specimens were placed into an electron beam and then transferred to the TEM room for characterization. The TEM (model HITACHI-H 7650 Japan) was used to observe the particle characteristics of SBV in the two samples. The particle size of SBV was measured and imaged at 100 nm. RT-PCR was used to conduct the second verification of SBV with specific primer pairs [61], and the third verification was conducted by Sanger dideoxy sequencing and phylogenetic analysis.

4.4. Quantitative PCR Assays of Purified Virus

qPCR was performed to measure the exact copy number of the virion particles [62]. Purified SBV was quantified on a high-speed cryogenic centrifuge (5840R; Sigma) using TransStart® Top Green qPCR SuperMix + Dye II (Beijing Qianjin, China). The cDNA was diluted 5 and 10 times, and templates of plasmid (Chinese sacbrood virus) CSBV DNA were used to create a standard curve and serially diluted from 10^{-2} to 10^{-8} . cDNA and plasmid CSBV were used as templates for qPCR. The qPCR solutions were prepared according to the following components, and the total reaction system was 20 μ L: cDNA purified virus 2 μ L, Chinese Sacbrood virus (CSBV) forward primer 0.4 μ L, reverse primer 0.4 μ L, 2 \times TransStart® Top Green qPCR SuperMix 10 μ L, and ddH₂O 7.2 μ L. qPCR amplification consisted of 40 cycles of 95 °C for 30 s, 95 °C for 5 s, 58 °C for 15 s, 72 °C for 10 s, and 95 °C for 0.5 s. Data were analyzed based on the molecular weight and the CSBV DNA plasmid sequence length [63].

4.5. Larval Rearing In Vitro

Larval rearing was conducted between Mar and Jun 2018 and continued until Nov 2018. This was based on the protocol developed by Crailsheim et al. [64]. Larvae were reared in a 48-well cell culture plate (clusters of sterile non-pyrogenic polystyrene; USA). To collect same-age larvae, an empty comb was placed in the hive for 24 h to allow the worker bees to clean up. The comb was inserted into the queen cage and the queen was confined in the cage for 24 h and placed in the middle of the hive (the queen bee and larvae could be better fed by worker bees). After 24 h, the eggs produced by the queen bees in the comb were examined. After 96 h, the comb with the second-instar larvae was removed from the hive and returned to the laboratory for larval grafting. The larval diet was prepared according to the protocol developed by Aupinel et al. [65]. The larval diet consisted of 6 g d-glucose, 6 g d-fructose, 1 g yeast extract, 50 g royal jelly, and 37 mL double-distilled water (ddh₂O). All the materials were weighed one by one, poured into a 100 mL beaker containing 37 mL of ddh₂O, and mixed well. Royal jelly was stored at -20 °C and placed in a water bath at $+34.5$ °C for 10 min, then 50 g of royal jelly was poured off and placed into a 100 mL beaker and some of the nonessential particles were removed. Then, the larval diet was pureed and placed into a 20 mL test tube, and the date was marked. We fed this diet to larvae for three days, and the larval diet was prepared using the same method for subsequent experiments.

4.6. Larval Grafting and Inoculation

Before grafting, the larvae were placed in cell culture plates, and the food and plates were prewarmed at ± 34 °C for 15 min. Then, 20 μ L of food was added to the culture plates, and second instar larvae (within 48 h of hatching) were gently grafted by Chinese grafting tools [66]. The grafting tools were placed in ethyl alcohol for 15 min and then washed with distilled water. To prevent cross-contamination by pathogenic microorganisms after grafting 16 larvae, the grafting tools were replaced with a new disinfected tool. Six 48-well cell culture plates (three in the control group and three in the treatment group) were considered for each colony. The culture plates were placed in an incubator (Ningbo Haishu, Saifu Experimental Instrument, Ningbo, China) at ± 34 °C with relative humidity (RH) of 95–99% inside a plastic box saturated with Potassium Sulfate (K₂SO₄) 360 g/L water to prevent dehydration, which was prepared in a plastic box. The larvae were fed with a pipette and modified tips, and the remaining food was removed daily. Inoculation tests were conducted based on the method described by Hu et al. [43]. The 3rd instars larvae with a single dose of the virus preparation (Fuzhou SBV at 6.32×10^6 copies/1 μ L of the purified virus) were mixed well with 10 μ L of food and placed into cell culture plates using a micropipette with a modified tip. The larvae were then allowed to thoroughly consume the food. After 24 h, in treatment groups 1, 2, and 3 and control groups 4, 5, and 6, no virus was added to the diet in all subsequent portions of the larval food. At 7 days, the larvae began defecating, determined by glycol uric acid crystals and fibrous substances on the

well. The larvae were gently transferred by plastic forceps to the new cell culture plates, which were lined with two layers of sterilized filter paper (autoclaved and dried). Culture plate larvae were cleaned with sterilized tissue paper before being transferring to new cells. The plates were placed in incubators at ± 34 °C and 75–80% RH (360 g/L saturated salt water) until pupation. The larvae were reared for an additional six days.

4.7. Observation and Sample Collections

Mortality was recorded daily, and larvae were observed under a microscope for immobility, swelling, tissue breakdown, and infection with SBV. The larvae that showed symptoms of SBV were individually collected in a 1.5 mL Eppendorf tube and marked. When the larvae showed discontinued respiration and reduced flexibility, developed edema, and exhibited a color change to brownish, they were removed from the cell culture plates every day and recorded. Therefore, they could be contaminated and decomposed by bacteria and fungi. Larvae infected with SBV were distinguished, collected during observation, recorded, and stored at -80 °C for subsequent experiments [67].

4.8. Sample Preparation and DNA Isolation

After four months of larval rearing, we collected a sufficient number of samples. We randomly selected 90 samples (45 S at the 7–10th instar larva stage and 45 R at the 12–14th pupa stage) from three *A. c. cerana* colonies with three biological replicates. Each colony was composed of 15 R and 15 S for genomic DNA extraction. Genomic DNA was extracted from whole larvae and pupae using the Cetyl Trimethyl Ammonium Bromide (CTAB) method of Paterson et al. [68]. The samples were ground into a powder with liquid nitrogen and quickly transferred into a 1.5 mL Ep tube, where 800 μ L of CTAB extraction DNA buffer was added and mixed well. CTAB was preheated at 65 °C in a water bath, gently oscillated several times every 5 min, centrifuged at 12,000 rpm for 20 min, carefully centrifuged for 15 min, and the supernatant was transferred to a fresh Eppendorf tube. An equal volume of phenol-chloroform 800 μ L solution was added, well mixed, and then centrifuged at 12,000 rpm for 10 min at 4 °C. The supernatant was carefully pipetted, and an equal volume of chloroform was added and centrifuged at 12,000 rpm for 10 min at 4 °C. Step 4 was repeated twice. The supernatant was precipitated at -20 °C for 1 h and then centrifuged at 12,000 rpm for 10 min at 4 °C. The supernatant was discarded and the pellet was washed twice with 70% ethanol. After drying at room temperature (generally 5–15 min), the samples were dissolved in 30 μ L DEPC water and stored at -20 °C for later use. The concentration and quality of DNA were measured by measuring its absorbance at OD_{260/288}–OD_{260/230} nm using a Nanodrop (Thermo Scientific 2000/2000 Waltham, MA, USA). A spectrophotometer was used to measure 1 μ L of the DNA sample, and it was recorded directly. The DNA quality was visually monitored by electrophoresis on 1.0% agarose gel. Qubit 4.0 (Thermo Fisher Scientific, Waltham, MA, USA) was used to quantify DNA using a fluorometer and diluted to 50 ng/ μ L for sequencing. The amount of DNA used for whole-genome sequencing was 2 μ g per sample. For SNP validation, larvae were reared using the same method described previously. The samples were removed at -80 °C to a refrigerator (Haier, Kyoto, Japan) and dissected with sterilized scissors. The thorax from the pupae and half of the larvae were used for DNA extraction.

4.9. Sample Sequencing and Data Processing

Data sequencing and bioinformatics analyses were conducted based on the pipeline. The experiments were conducted according to the standard procedure developed by Illumina (San Diego, CA, USA). Qualified DNA samples were resequenced by whole-genome sequencing using a commercial Illumina NOVA HisSeq X-Ten Biomarker Technologies Corporation (Beijing, China). Sample quality detection, genomic DNA library construction, library quality detection, and sequencing were also included. After the genomic DNA test, qualified DNA samples were fragmented by mechanical interruption (ultrasound), fragmented purification, end pair, 3' ends plus A, and connection sequencing joints; elec-

trophoresis on 1.0% agarose gels was used to select fragment sizes. Polymerase chain reaction (PCR) amplification was performed to construct a sequencing library. DNA libraries were constructed using the DNA Library Prep Reference Guide with an insert size of 150 bp and were constructed based on Illumina, following the manufacturer's instructions (Illumina Genome Analyzer II), and paired-end reads were generated using Illumina. The DNA library was first built to conduct a quality inspection, and a qualified library was used for sequencing. The quality of the first reads (double-end sequence) was obtained by sequencing, evaluation, and filtering. Clean reads were used for subsequent bioinformatics analyses. Clean reads were compared with reference genome sequences, and variations in SNPs, indels, and other variants were detected and annotated based on the comparison results. Subsequent genetic evolution analyses were performed based on the detection of variation.

4.10. Sequence Data Filter, Quality Control, and Data Assessment

Quality control (QC) processed low-quality data by filtering the sequenced reads or raw reads (Table S4). To ensure data quality, raw reads were filtered and clean reads were obtained and used for subsequent bioinformatics analyses before aligning them to the reference genomes of *A. c. cerana*. Low-quality reads containing adaptor sequences, duplicate reads, and reads containing adaptor sequences were filtered based on the following three rules: the main steps of data filtering were the removal of adapter reads, removal of >10% Ns reads, and removal of >50% reads with low bases (base quality scores <10). The raw data or raw reads and the results were in the FASTQ format. They were stored in fq file format, containing sequence information for the sequencing reads and corresponding sequencing quality information. Clean sequencing data were used for SNP calling.

4.11. Mapping Variant Detection and SNP Calling

The clean reads were calculated using the SAMTOOLS flag stat command [69]. Filtered clean reads were compared with reference genome sequence reads that were mapped to the reference genome of *A. c. cerana* (APICC1.0, GenBank, Assembly GCA_002290385.1) using the 'mem' algorithm of Burrows–Wheeler Aligner (BWA version 0.7.5a-r405) [44], with the default parameters. The mapping results were processed by sorting and duplicate marking using SAMTOOLS (version 0.1.19–44428 cd), and variants were detected [69]. Clean data were used to call both SNPs and Indels using PICARD tools (<http://broadinstitute.github.io/picard/>) (accessed on 26 December 2018) and local realignment around Indels. The Indels-Realigner in GATK Haplotype Caller algorithm in GATK software version 3.8 was used for SNP SNPs and small Indels mutations calling across the 90 samples with default parameters [40]. Raw variances, including SNPs and indels with low quality (QUAL < 30, QD < 2.0, FS > 60.0, MQ < 40.0), were filtered. Distribution of base sequencing quality (Base-calling analysis, Illumina Casava Software version 1.8) was used with sequencing parameters (paired-end PE; sequenced read length, 150 bp). The average sequencing depth was 14x coverage. Low-quality reads were filtered based on the following four rules: If one end of a pair-end read had > 5% Ns bases, the paired-end read was removed. Each pair-end read was removed if it had an average base quality of Q < 20 (Phred-like score). Each read was trimmed to its three bases if their quality scores were Q < 13. Trimming was stopped at the base with a quality score of ≥ 13 . If the number of remaining bases was <40, then the paired-end reads were removed. Duplicates of paired-end reads were removed.

4.12. Variation Detection, Quality Control, and Annotation

SNPs and small Indels were detected using Genome Analysis Toolkit (GATK) software [40]. Based on the positioning results of the clean reads in the reference genome, PICARD tools (Picard: <http://sourceforge.net/projects/picard/>) (accessed on 26 December 2018) (Picard) were used to filter redundant reads (mark duplicates) to ensure the accuracy of the test results. The variation detection of SNPs and indels was carried out using the GATK haplotype caller (local monomer assembly) algorithm, and each sample was first

generated in genome variant call format (gVCF). Then, the group joint-genotyped all VCF SNPs (Table S11). Finally, the filter was set on the final mutation sites using the variant call format (VCF) in the R private SNP VCF R private SNPs. The annotation line contained the message of the file data row and the interpretation of the meaning of the various identifiers was used in the format column. In contrast, the header and data row contain the variation detection results of each sample. The variation results were strictly filtered to ensure reliability; the main filtration parameters were as follows: based on the subroutine in BCF tools, vcfutils. Pl (var filter-w 5-w 10), the SNP and adjacent Indels in 5 bp near Indels were filtered out within 10 bp; cluster size 2 cluster window size 5 bp, indicating that the number of variations in the 5 bp window should not be exceeded. The quality value of the $Q < 30$ Phred format demonstrates that there is a possibility of divergence at that point. Suppose that the mass value is >30 , which was filtered out. In this case, $QD < 2.0$, the variant mass value (quality) was divided by the ratio of coverage depth, and the covering depth was the sum of the depth of coverage of all samples containing mutant bases at this point. $QD < 2.0$ were filtered out. If the $MQ < 40$, all ratios were read on the bit point to the mean square root of the mass value. If the $MQ > 40$, it was filtered out. If the $FS > 60$, the value was converted through the p value < 0.05 of the Fisher exact test, which described whether there was a significant positive or negative chain specificity for the reads that contained only variations and reads that had only reference sequence bases during sequencing or comparison. There are no chain-specific comparison results, and the FS should be close to zero. An FS above 60 was filtered out, and other variation filtration parameters were processed using the default values specified by GATK.

4.13. Annotation of Mutated Genes at the DNA Level

Annotation was conducted to classify SNPs and Indels as genetic variants using the SnpEff toolbox [70], which was used to annotate the effect impact, function class, codon change distance, amino acid change, gene name, transcript biotype, gene coding, transcript ID, exon, and intron rank warning error for each SNP. An annotation analysis of differentially expressed genes was the basis for further explanation of gene function. The KEGG database is the primary public database for this pathway. The presence of genes in specific regions was studied using pathway enrichment analysis [71].

4.14. Population Genetics Analysis

To study the relationships and evolution between SNPs and traits, a phylogenetic analysis of all detected SNP genotypes after filtering was used to construct the phylogenetic tree SNPs with low MAF (< 0.05). A low genotyping rate ($p \leq 0.5$) was used from 90 samples using the neighbor-joining (N-J) method under the Kimura 2-parameter model implemented by M_{EGA} X Software, with bootstrap replicates of 1000 [72]. In addition, to study the relationship and evolution between R and S in SNP validation, phylogenetic trees were constructed from R and S from different SNP validation data. Phylogenetic tree visualization and editing assignments were performed using ITOL (<http://itol.embl.de/>) (accessed on 13 March 2019).

The EIGENSOFT software package was used to conduct PCAs based on SNP data characteristics to obtain sample clustering. The PCs were used to visualize the group of samples, which could explain the percentage of the total variation based on SNP data to carry out PCA and obtain the clustering of samples [73]. An admixture analysis in the haploid model, Admixture 1.3, was used to filter SNPs with multiple K values (i.e., the number of putative populations) ranging from 1 to 10, and the CV error was used to select the best K-value [74]. SPAGeDi (Spatial Pattern Analysis of Genetic Diversity) software was used to estimate the relative kinship between two individuals in a natural population [75].

4.15. Identification of Selective Sweep-between Traits

The F_{ST} and π ratios were calculated to detect differentiation and reveal the significant regions using PopGenome [41]. To detect genomic regions that were potentially differ-

entiated for R and S, the F_{ST} and π rates in the top 1% and top 5% were calculated for 100-kb sliding windows with a step size of 10 kb using VCFTOOLS. Sliding windows with F_{ST} values of > 95% or > 99% and π ratios of > 95% or > 99% of genome-wide F_{ST} values were selected and regarded as significantly different windows. Overlapping significance windows were merged into single fragments, and the fragments were regarded as highly divergent regions across groups. To identify genes putatively under selective sweep between the two populations, SnpSift software was used to screen each SNP site. Based on the SNP results of the mutation detection, filtering, and selective sweep, an INT of < 0.5 and an MAF of < 0.05 were used to obtain highly consistent SNP loci in the genome-wide empirical distributions. F_{ST} : Estimation of gene flow levels from the DNA sequence data. π : Statistical test for detecting geographic subdivisions [45].

4.16. Functional Annotation and Enrichment Analysis

Functional and annotation analyses of 90 individual samples were conducted, and their gene changes were detected and annotated against ten databases. After detecting the significant regions by selective sweep between traits, the genes and SNPs were annotated. The enrichment factor was used to analyze the enrichment degree of the pathways, and Fisher's exact test was performed and the false discovery rate (FDR) was determined to calculate the significance of enrichment with a p value of < 0.05.

4.17. SNP Validation and Sanger Sequencing

To verify the candidate SNPs, we first conducted an (MAF < 0.05) analysis to select high-consistency SNPs based on the combination of F_{ST} and π S divided R top 1% and F_{ST} and π R divided S top 1% with a high percentage genotype. For significant SNPs, 40 pairwise primers were designed to amplify the desired SNP for SNP validation (Table S20). In the PCR, the total reaction volume was 25 μ L, containing the following components: 2 μ L DNA template, 0.4 μ L of each primer (forward and reverse), 9.7 μ L ddH₂O, and 12.5 μ L 2X Fine Easy Taq Super Mix Trans solution (TransGene, Beijing China). The PCR amplification was conducted at 94 °C for 3 min, followed by 35 cycles of 94 °C for 30 s, a temperature of between 57 °C and 58 °C for 30 s, 72 °C for 1 min, and a final step at 72 °C for 5 min. To ensure that the PCRs were running correctly and to prevent contamination, the *Apis*- β -actin primer pair (181 bp) was used as an internal positive control [76]. A negative control, ddH₂O, was used for all PCR amplification assays. Amplified PCR products were analyzed using 5 μ L of 2% agarose gel electrophoresis and visualized under UV light. The lengths were compared to that of a standard Trans 2 K molecular weight DNA ladder (Invitrogen, Waltham, MA, USA). After visualization, the PCR product of 20 μ L of amplified target DNA was purified, positive strands (forward primer) were used in di-deoxy Sanger Sequencing, and the ABI file data were analyzed.

4.18. Statistical Analysis

GraphPad Prism version 9.00 (121) for Windows (Graph Pad Software, San Diego, CA, USA) was used to compare the control group with the treatment group to evaluate the survival function analysis. The Kaplan–Meier chi-square log-rank (Mantel–Cox) test was performed, where a p value of < 0.01 was considered significantly different. Additionally, the survival functions of CR and CS were compared. For SNP flanking validation, specific primer pairs were designed using the default parameters of the BatchPrimer3 v1.0 online software. Chromas, BioEdit sequence Alignment Editor, and DNASTAR Lasergen SeqMan Pro were used to manually evaluate each sample peak and compare the variance in the genotype differences. The distributions of allele frequencies between R and S were analyzed using the chi-square test in Hardy–Weinberg equilibrium, where a p value of < 0.05 was considered significantly different. The ggplot2 package in R was used to plot the significant region graphs.

5. Conclusions

SBV is prevalent in the genus *Apis*, including bumblebees, and effective methods to control this viral disease are urgently needed. Genome sequences are crucial for basic biological research on bees, and Asian honeybees provide an excellent model for understanding the genetic basis of SBV resistance at the genomic level. To catalog the genetic variations and nucleotide polymorphisms associated with SBV resistance in the *A. c. cerana* genome, bee larvae were challenged with SBV and raised in vitro, and 90 individual samples from three populations were resequenced. This study identified 31,000,613 high-quality SNPs with 696,352 individual genotypes compared to the reference genome. Two SNP molecular markers related to SBV resistance were identified and successfully validated by multiple verifications. The results showed that the two SNPs were significantly associated with SBV resistance, and these significant SNP loci were found on the 15th chromosome, composed of five putative candidate genes. The favorable G/G and C/C allele frequencies in the R samples at the two SNP loci revealed the molecular mechanism of SBV resistance by validation and analysis of the allele frequencies. Functional and enrichment analyses of predicted genes showed that genes involved in cellular processes might play a significant role in SBV resistance. The findings of this study may support and improve the honeybee *A. c. cerana* genome and may be employed in breeding to enhance SBV resistance and decrease SBV infection. Hence, SNP molecular markers can quickly and accurately identify SBV-resistant colonies, and further investigation is needed to determine if the two discovered SNPs are also responsive to other closely related species, such as *A. mellifera*.

6. Patents

This study received two patents from the China Innovation Center which have already been published. Patent. Issue No. 2021030301132160. Publication No. CN 112430675 A, SNP labeling at KZ288474.1_322717. Another Invention Issue No. 2021100418633. Publication No. CN 112430675 A, SNP labeling at KZ288479.1_95621. Invention method for identifying SNP molecular markers related to sacbrood virus resistance.

Supplementary Materials: The supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms24076238/s1>.

Author Contributions: Conceptualization, S.S., A.K.H. and H.N.; methodology, S.S., A.K.H. and Z.L.; software, H.N., A.K.H. and J.H.; validation, A.K.H., H.N. and H.N.; formal analysis, S.S., A.K.H. and H.N.; investigation, A.K.H.; resources, S.S.; data curation, S.S., A.K.H., W.F. and K.Y.; writing—original draft preparation, A.K.H.; writing—review and editing, S.S., A.K.H., H.N. and Y.L.; visualization, S.T.W., Z.Z. and M.H.; supervision, S.S.; project administration, S.S. and H.N.; funding acquisition, S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (no. 31772684), the China Agriculture Research System of MOF, and MARA (no. CARS-44-KXJ4).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data from 90 sequenced individuals were deposited in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under the accession number BioProject: PRJNA815733.

Acknowledgments: We thank the local beekeepers, especially Zhang Zongmu, for their cooperation in sample collection for virus purification.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

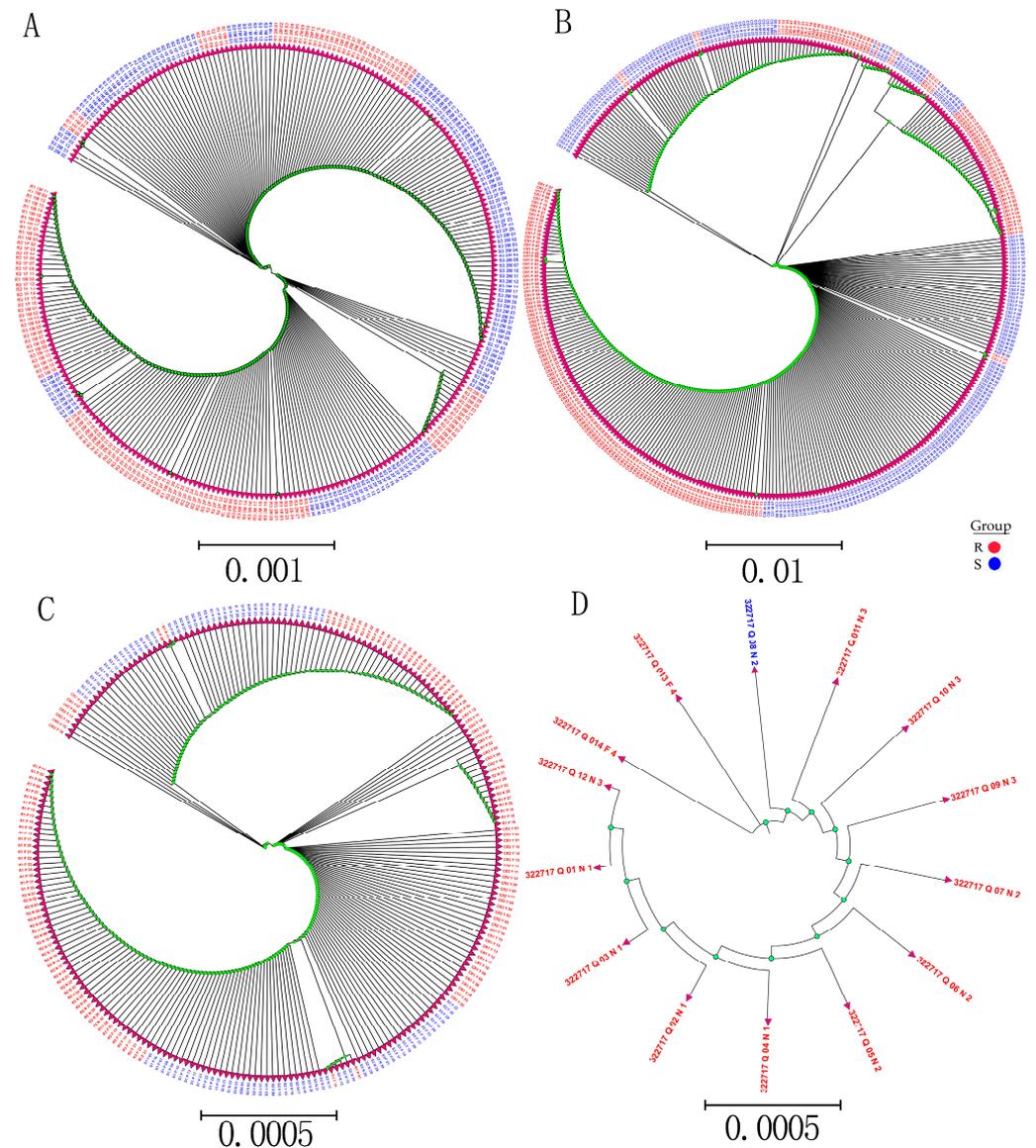


Figure A1. Phylogenetic trees of short DNA Sanger Sequencing of SNP KZ288474_322717 from 926 samples in different categories. **(A)** Phylogenetic trees of first and second SNP validation of R (resistant) and S (susceptible) of 288 samples from three *Apis ceranacarana* colonies in Minhou (M) (1 and 2 M) and F (Fuzhou) (1 and 2 F); **(B)** comparison of SNP validation of 384 samples from six colonies (three colonies resistant (CR) in Fuzhou and three colonies susceptible (CS) in Minho); **(C)** further SNP validation of 240 samples from six colonies and different geographic locations in P (Putian) (one colony), N (Nanping) (one colony), Fuzhou (one colony), and in Y (Yunnan Province) (three colonies); **(D)** SNP validation at the queen level of 14 colonies (12 colonies in Nanping and 2 colonies in Fuzhou). Neighbor-joining trees were constructed using MEGA 11 software with 1000-fold bootstrap resampling. The numbers at the bottom of each tree represent the level of evolution relationship difference between R and S samples after the trimmed end.

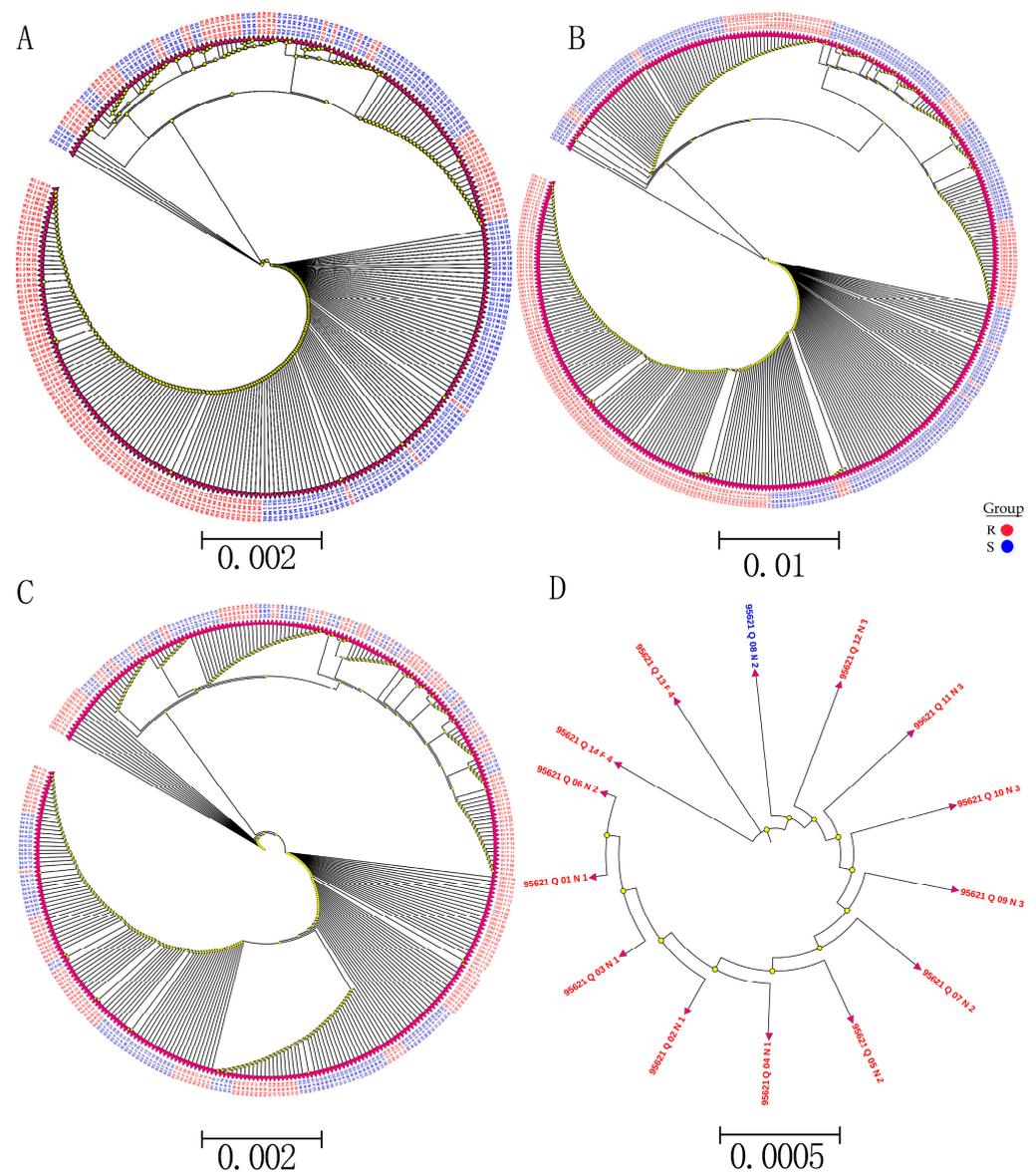


Figure A2. Phylogenetic trees of short DNA Sanger Sequencing of SNP KZ288479_95621 from 1022 samples in different categories. **(A)** Phylogenetic tree of first and second SNP validation of R (resistant) and S (susceptible) of 288 samples from three *Apis ceranacerana* colonies in Minhou (1 and 2 M); **(B)** comparison of SNP validation of 384 samples from six colonies (three RC (resistant colonies) in Fuzhou and three CS (colonies susceptible) in Minho). **(C)** Further SNP validation of 336 samples from different geographic locations from eight colonies (one colony in Putian (P), one colony in Minhou (M), two colonies in Fuzhou (F), one colony in Nanping (N), and three resistant colonies (CR) in Yunnan Province (Y)); and **(D)** SNP validation at the queen level from 14 colonies (12 colonies in Nanping from three apiaries and two colonies in Fuzhou from one apiary). Neighbor-joining trees were constructed using MEGA 11 software with 1000-fold bootstrap resampling. After trimming, the numbers at the bottom of each tree represent the different levels of evolutionary relationships between the R and S samples.

Appendix B

Table A1. The number of SNPs and mutation statistics detected in 45 resistant individual samples in *A. c. cerana*.

Sample ID	No. of SNP	Transition	Transversion	Ti/Tv	Heterozygosity	Homozygosity	Het Ratio
R06	346,467	286,114	60,353	4.74	206,120	140,347	59.49%
R07	345,299	284,870	60,429	4.71	203,124	142,175	58.82%
R08	353,382	291,547	61,835	4.71	214,270	139,112	60.63%
R09	350,981	289,928	61,053	4.74	210,430	140,551	59.95%
R10	347,816	287,252	60,564	4.74	209,045	138,771	60.10%
R11	344,800	284,389	60,411	4.7	205,847	138,953	59.70%
R12	345,705	284,968	60,737	4.69	205,264	140,441	59.37%
R22	348,408	287,704	60,704	4.73	207,154	141,254	59.45%
R23	348,326	287,457	60,869	4.72	208,242	140,084	59.78%
R24	348,908	288,187	60,721	4.74	207,666	141,242	59.51%
R25	345,513	285,599	59,914	4.76	204,780	140,733	59.26%
R26	348,576	287,820	60,756	4.73	208,581	139,995	59.83%
R27	345,892	285,626	60,266	4.73	204,947	140,945	59.25%
R28	341,848	282,230	59,618	4.73	200,333	141,515	58.60%
R29	355,902	293,915	61,987	4.74	216,081	139,821	60.71%
R30	347,819	287,083	60,736	4.72	208,182	139,637	59.85%
R31	340,012	280,665	59,347	4.72	199,581	140,431	58.69%
R41	347,917	286,989	60,928	4.71	207,136	140,781	59.53%
R42	347,204	286,561	60,643	4.72	206,977	140,227	59.61%
R43	341,145	281,566	59,579	4.72	198,869	142,276	58.29%
R44	342,377	282,576	59,801	4.72	200,863	141,514	58.66%
R45	334,200	275,518	58,682	4.69	195,815	138,385	58.59%
R46	335,413	276,843	58,570	4.72	192,745	142,668	57.46%
R47	352,276	290,750	61,526	4.72	213,213	139,063	60.52%
R48	340,675	280,736	59,939	4.68	202,063	138,612	59.31%
R52	345,422	284,834	60,588	4.7	205,244	140,178	59.41%
R53	342,904	283,176	59,728	4.74	203,543	139,361	59.35%
R54	348,712	287,678	61,034	4.71	206,656	142,056	59.26%
R55	348,240	287,605	60,635	4.74	205,989	142,251	59.15%
R56	349,040	288172	60,868	4.73	208,621	140,419	59.76%
R57	355,008	293,026	61,982	4.72	224,797	130,211	63.32%
R58	344,805	284,702	60,103	4.73	202,387	142,418	58.69%
R59	343,380	283,340	60,040	4.71	204,508	138,872	59.55%
R63	356,973	294,757	62,216	4.73	219,730	137,243	61.55%
R64	352,856	291,321	61,535	4.73	214,712	138,144	60.84%
R65	354,640	292,673	61,967	4.72	219,396	135,244	61.86%
R66	349,637	288,462	61,175	4.71	211,251	138,386	60.42%
R70	359644	296,777	62,867	4.72	221,633	138,011	61.62%
R71	351,631	290,182	61,449	4.72	205,210	146,421	58.35%
R72	352,141	290,845	61,296	4.74	213,769	138,372	60.70%
R73	341,271	281,588	59,683	4.71	198,924	142,347	58.28%
R74	329,709	272,070	57,639	4.72	186,937	142,772	56.69%
R75	342,487	282,836	59,651	4.74	199,585	142,902	58.27%
R76	347,504	286,752	60,752	4.72	208,380	139,124	59.96%
R87	333,205	274,731	58,474	4.69	193,627	139,578	58.11%

A description of the meaning of each column; Sample ID, the uniform number of the sample; No. of SNPs, the number of SNPs detected in each sample; Transition, the number of SNPs of the conversion type; Transversion, the number of SNPs of the transversion type; Ti/Tv, conversion vs. transversion type ratio; Heterozygosity, the number of SNPs of heterozygous types; Homozygosity, the number of SNPs of homozygous type; and Het ratio, the SNP ratio of heterozygous types.

Table A2. The number of SNPs and mutation statistics detected in of 45 susceptible individual samples in *A. c. cerana*.

Sample ID	SNP number	Transition	Transversion	Ti/Tv	Heterozygosity	Homozygosity	Het Ratio
R01	360,756	297,704	63,052	4.72	221,997	138,759	61.53%
R02	336,887	278,127	58,760	4.73	193,460	143,427	57.42%
R03	354,788	292,912	61,876	4.73	215,591	139,197	60.76%
R04	352,611	291,035	61,576	4.72	212,184	140,427	60.17%
R05	355,590	293,461	62,129	4.72	215,604	139,986	60.63%
R13	341,666	282,181	59,485	4.74	205,640	136,026	60.18%
R14	336,601	277,936	58,665	4.73	201,407	135,194	59.83%
R15	346,514	286,123	60,391	4.73	207,671	138,843	59.93%
R16	345,868	285,642	60,226	4.74	206,487	139,381	59.70%
R17	346,979	286,493	60,486	4.73	205,656	141,323	59.27%
R18	354,831	292,875	61,956	4.72	215,909	138,922	60.84%
R19	345,792	285,265	60,527	4.71	207,135	138,657	59.90%
R20	340,020	280,665	59,355	4.72	198,541	141,479	58.39%
R21	346,264	285,605	60,659	4.7	205,851	140,413	59.44%
R32	334,725	276,436	58,289	4.74	191,350	143,375	57.16%
R33	350,660	289,435	61,225	4.72	210,324	140,336	59.97%
R34	347,674	286,833	60,841	4.71	203,579	144,095	58.55%
R35	340,552	281,218	59,334	4.73	199,362	141,190	58.54%
R36	350,405	289,239	61,166	4.72	210,779	139,626	60.15%
R37	358,373	295,811	62,562	4.72	219,819	138,554	61.33%
R38	253,796	209,592	44,204	4.74	1,746	252,050	0.68%
R39	350,910	289,589	61,321	4.72	206,651	144,259	58.89%
R40	251,387	207,638	43,749	4.74	1,429	249,958	0.56%
R49	344,807	284,517	60,290	4.71	205,470	139,337	59.58%
R50	345,906	285,478	60,428	4.72	206,208	139,698	59.61%
R51	353,940	292,179	61,761	4.73	217,536	136,404	61.46%
R60	331,995	273,687	58,308	4.69	191,343	140,652	57.63%
R61	332,638	27,4187	58,451	4.69	192,547	140,091	57.88%
R62	339,924	280,323	59,601	4.7	200,128	139,796	58.87%
R67	354,742	292,874	61,868	4.73	210,996	143,746	59.47%
R68	352,692	291,158	61,534	4.73	209,041	143,651	59.27%
R69	353,033	291,535	61,498	4.74	213,767	139,266	60.55%
R77	351,051	289,647	61,404	4.71	210,795	140,256	60.04%
R78	353,312	291,465	61,847	4.71	213,500	139,812	60.42%
R79	347,885	287,117	60,768	4.72	207,516	140,369	59.65%
R80	344,110	283,941	60,169	4.71	201,594	142,516	58.58%
R81	346,030	285,648	60,382	4.73	205,082	140,948	59.26%
R82	345,596	284,925	60,671	4.69	205,384	140,212	59.42%
R83	351,269	290,070	61,199	4.73	211,295	139,974	60.15%
R84	332,133	273,877	58,256	4.7	191,850	140,283	57.76%
R85	341,059	281,394	59,665	4.71	200,612	140,447	58.82%
R86	352,406	291,046	61,360	4.74	212,467	139,939	60.29%
R88	338,837	279,396	59,441	4.7	197,781	141,056	58.37%
R89	344,094	283,909	60,185	4.71	205,304	138,790	59.66%
R90	343,435	283,688	59,747	4.74	203,193	140,242	59.16%

Table A3. Classification statistics of differentially expressed genes produced by 45 resistant and 45 susceptible samples in *A. c. cerana*.

Sample ID	Genes with Non-synonymous	Genes with Indels	Sample ID	Genes with Non-synonymous	Genes with Indels
R06 R	3468	1029	R01-S	3569	1105
R07	3465	1001	R02	3372	1010
R08	3520	1022	R03	3505	1024
R09	3453	1024	R04	3526	1058
R10	3488	1021	R05	3527	1047
R11	3456	1007	R13	3422	1026
R12	3465	1007	R14	3489	1049
R22	3449	1035	R15	3411	1043
R23	3432	1001	R16	3422	1016
R24	3434	1020	R17	3445	1031
R25	3442	974	R18	3486	1058
R26	3463	1035	R19	3443	1023
R27	3451	1035	R20	3445	1027
R28	3376	1001	R21	3457	1027
R29	3493	1046	R32	3349	984
R30	3420	1031	R33	3458	999
R31	3408	989	R34	3431	1002
R41	3417	1017	R35	3388	1005
R42	3424	1015	R36	3458	1016
R43	3365	981	R37	3486	1063
R44	3409	997	R38	2766	815
R45	3388	969	R39	3459	1000
R46	3340	983	R40	2739	787
R47	3454	1010	R49	3453	986
R48	3392	1019	R50	3468	1007
R52	3482	977	R51	3537	1056
R53	3414	1017	R60	3379	988
R54	3454	1014	R61	3359	979
R55	3490	1036	R62	3360	1013
R56	3477	1022	R67	3457	1010
R57	3526	1007	R68	3442	998
R58	3442	994	R69	3459	1017
R59	3429	985	R77	3467	1038
R63	3498	1074	R78	3516	1066
R64	3467	1071	R79	3434	998
R65	3470	1020	R80	3450	1007
R66	3470	1052	R81	3396	988
R70	3473	1071	R82	3475	988
R71	3457	1014	R83	3482	1035
R72	3435	1031	R84	3348	1014
R73	3372	1031	R85	3369	973
R74	3355	987	R86	3481	1065
R75	3403	989	R88	3378	993
R76	3461	1017	R89	3415	1002
R87 R	3339	999	R90 S	3388	1009

Table A4. Statistical results of variant genes after annotation of 90 individual samples in *A. c. cerana*.

Sample ID	Gene Number	Sample	Gene Number	Sample	Gene Number
R01	3765	R31	3579	R61	3542
R02	3564	R32	3545	R62	3581
R03	3695	R33	3621	R63	3735
R04	3715	R34	3642	R64	3658
R05	3716	R35	3567	R65	3651
R06	3674	R36	3638	R66	3687
R07	3664	R37	3703	R67	3634
R08	3699	R38	2985	R68	3643
R09	3652	R39	3643	R69	3645
R10	3677	R40	2962	R70	3696
R11	3634	R41	3596	R71	3641
R12	3665	R42	3617	R72	3616
R13	3627	R43	3562	R73	3588
R14	3702	R44	3611	R74	3565
R15	3624	R45	3570	R75	3582
R16	3623	R46	3545	R76	3657
R17	3660	R47	3631	R77	3659
R18	3689	R48	3595	R78	3726
R19	3647	R49	3633	R79	3615
R20	3638	R50	3653	R88	3640
R21	3662	R51	3716	R81	3595
R22	3642	R52	3649	R82	3636
R23	3619	R53	3606	R83	3673
R24	3604	R54	3639	R84	3550
R25	3627	R55	3676	R85	3562
R26	3673	R56	3667	R86	3692
R27	3654	R57	3702	R87	3524
R28	3581	R58	3627	R88	3560
R29	3701	R59	3609	R89	3612
R30	3632	R60	3560	R90	3605

References

- Klatt, B.K.; Holzschuh, A.; Westphal, C.; Clough, Y.; Smit, I.; Pawelzik, E.; Tschamntke, T. Bee Pollination Improves Crop Quality, Shelf Life and Commercial Value. *Proc. R. Soc. B Biol. Sci.* **2013**, *281*, 20132440. [[CrossRef](#)] [[PubMed](#)]
- Potts, S.G.; Biesmeijer, J.C.; Kremen, C.; Neumann, P.; Schweiger, O.; Kunin, W.E. Global Pollinator Declines: Trends, Impacts and Drivers. *Trends Ecol. Evol.* **2010**, *25*, 345–353. [[CrossRef](#)]
- McMenamin, A.J.; Genersch, E. Honey Bee Colony Losses and Associated Viruses. *Curr. Opin. Insect Sci.* **2015**, *8*, 121–129. [[CrossRef](#)] [[PubMed](#)]
- Bailey, L.; Gibbs, A.J.; Woods, R.D. Sacbrood Virus of the Larval Honey Bee (*Apis mellifera* Linnaeus). *Virology* **1964**, *23*, 425–429. [[CrossRef](#)]
- Ellis, J.D.; Munn, P.A. The Worldwide Health Status of Honey Bees. *Bee World* **2005**, *86*, 88–101. [[CrossRef](#)]
- Bailey, L. The Multiplication and Spread of Sacbrood Virus of Bees. *Ann. Appl. Biol.* **1969**, *63*, 483–491. [[CrossRef](#)] [[PubMed](#)]
- Reynaldi, F.J.; Sguazza, G.H.; Albicoro, F.J.; Pecoraro, M.R.; Galosi, C.M. First Molecular Detection of Co-Infection of Honey Bee Viruses in Asymptomatic *Bombus atratus* in South America. *Braz. J. Biol.* **2013**, *73*, 797–800. [[CrossRef](#)]
- Freiberg, M.; De Jong, D.; Message, D.; Cox-Foster, D. First Report of Sacbrood Virus in Honey Bee (*Apis mellifera*) Colonies in Brazil. *Genet. Mol. Res.* **2012**, *11*, 3310–3314. [[CrossRef](#)] [[PubMed](#)]
- Berényi, O.; Bakonyi, T.; Derakhshifar, I.; Köglberger, H.; Nowotny, N.; Bere, O.; Derakhshifar, I. Occurrence of Six Honeybee Viruses in Diseased Austrian Apiaries. *Appl. Environ. Microbiol.* **2006**, *72*, 2414–2420. [[CrossRef](#)]
- Choe, S.E.; Nguyen, L.T.K.; Noh, J.H.; Koh, H.B.; Jean, Y.H.; Kweon, C.H.; Kang, S.W. Prevalence and Distribution of Six Bee Viruses in Korean *Apis cerana* Populations. *J. Invertebr. Pathol.* **2012**, *109*, 330–333. [[CrossRef](#)]
- McMahon, D.P.; Fürst, M.A.; Caspar, J.; Theodorou, P.; Brown, M.J.F.; Paxton, R.J. A Sting in the Spit: Widespread Cross-Infection of Multiple RNA Viruses across Wild and Managed Bees. *J. Anim. Ecol.* **2015**, *84*, 615–624. [[CrossRef](#)]
- Singh, R.; Levitt, A.L.; Rajotte, E.G.; Holmes, E.C.; Ostiguy, N.; Vanengelsdorp, D.; Lipkin, W.I.; Depamphilis, C.W.; Toth, A.L.; Cox-Foster, D.L. RNA Viruses in Hymenopteran Pollinators: Evidence of Inter-Taxa Virus Transmission via Pollen and Potential Impact on Non-*Apis* Hymenopteran Species. *PLoS ONE* **2010**, *5*, e14357. [[CrossRef](#)] [[PubMed](#)]
- Chen, Y.P.; Pettis, J.S.; Collins, A.; Feldlaufer, M.F. Prevalence and Transmission of Honeybee Viruses. *Appl. Environ. Microbiol.* **2006**, *72*, 606–611. [[CrossRef](#)] [[PubMed](#)]

14. Wang, T.-S.; Shi, T.-F.; Fang, L.; Lin-Sheng, Y.U.; Lei, Q.I.; Xiang-Jin, M. Occurrence and Distribution of Seven Bee Viruses in *Apis mellifera* and *Apis cerana* in Anhui Province, China. *Chin. J. Appl. Entomol.* **2015**, *52*, 324–332.
15. McMenamin, A.J.; Flenniken, M.L. Recently Identified Bee Viruses and Their Impact on Bee Pollinators. *Curr. Opin. Insect Sci.* **2018**, *26*, 120–129. [[CrossRef](#)]
16. Diao, Q.; Sun, L.; Zheng, H.H.H.; Zeng, Z.; Wang, S.; Xu, S.; Zheng, H.H.H.; Chen, Y.; Shi, Y.; Wang, Y.; et al. Genomic and Transcriptomic Analysis of the Asian Honeybee *Apis cerana* Provides Novel Insights into Honeybee Biology. *Sci. Rep.* **2018**, *8*, 822. [[CrossRef](#)] [[PubMed](#)]
17. Tsai, W.-S.S.; Chen, Y.-W.W.; Nai, Y.-S.S.; Hsu, M.-H.H.; Hsu, P.-S.S.; Sung, I.-H.H.; Ko, C.-Y.Y.; Hsu, P.-S.S.; Tsai, W.-S.S.; Chen, Y.-W.W.; et al. The Seasonal Detection of AcSBV (*Apis cerana* Sacbrood Virus) Prevalence in Taiwan. *J. Asia. Pac. Entomol.* **2018**, *21*, 417–422.
18. Liu, X.; Zhang, Y.; Yan, X.; Han, R. Prevention of Chinese Sacbrood Virus Infection in *Apis cerana* Using Rna Interference. *Curr. Microbiol.* **2010**, *61*, 422–428. [[CrossRef](#)] [[PubMed](#)]
19. Ahn, A.J.; Ahn, K.S.; Suh, G.H.; Noh, J.H.; Kim, Y.H.; Yoo, M.S.; Kang, S.W.; Shin, S.S. Efficacy of Silver Ions against Sacbrood Virus Infection in the Eastern Honey Bee *Apis Cerana*. *J. Vet. Sci.* **2015**, *16*, 289–295. [[CrossRef](#)]
20. Wegener, J.; Ruhnke, H.; Milchreit, K.; Kleebaum, K.; Franke, M.; Mispagel, S.; Bischoff, G.; Kamp, G.; Bienefeld, K. Ecotoxicology and Environmental Safety Secondary Biomarkers of Insecticide-Induced Stress of Honey Bee Colonies and Their Relevance for Overwintering Strength. *Ecotoxicol. Environ. Saf.* **2016**, *132*, 379–389. [[CrossRef](#)]
21. Tan, K.; Wang, Y.; Dong, S.; Liu, X.; Zhuang, D.; Chen, W.; Oldroyd, B.P. Associations between Reproduction and Work in Workers of the Asian Hive Bee *Apis Cerana*. *J. Insect Physiol.* **2015**, *82*, 33–37. [[CrossRef](#)] [[PubMed](#)]
22. Sun, L.; Li, M.; Fei, D.; Diao, Q.; Wang, J.; Li, L.; Ma, M. Preparation and Application of Egg Yolk Antibodies Against Chinese Sacbrood Virus Infection. *Front. Microbiol.* **2018**, *9*, 1814. [[CrossRef](#)] [[PubMed](#)]
23. Sun, L.; Zhang, X.; Xu, S.; Hou, C.; Xu, J.; Zhao, D.; Chen, Y. Antiviral Activities of a Medicinal Plant Extract Against Sacbrood Virus in Honeybees. *Virol. J.* **2021**, *18*, 83. [[CrossRef](#)]
24. Liu, L.; Qin, M.; Yang, L.; Song, Z.; Luo, L.; Bao, H.; Ma, Z.; Zhou, Z.; Xu, J. A Genome-Wide Analysis of Simple Sequence Repeats in *Apis cerana* and Its Development as Polymorphism Markers. *Gene* **2017**, *599*, 53–59. [[CrossRef](#)]
25. Chen, C.; Wang, H.; Liu, Z.; Chen, X.; Tang, J.; Meng, F.; Shi, W. Population Genomics Provide Insights into the Evolution and Adaptation of the Eastern Honey Bee (*Apis Cerana*). *Mol. Biol. Evol.* **2018**, *35*, 2260–2271. [[CrossRef](#)]
26. Holloway, B.; Allen Sylvester, H.; Bourgeois, L.; Rinderer, T.E. Association of Single Nucleotide Polymorphisms to Resistance to Chalkbrood in *Apis mellifera*. *J. Apic. Res.* **2012**, *51*, 154–163. [[CrossRef](#)]
27. Liu, Y.; Yan, L.; Li, Z.; Huang, W.F.; Pokhrel, S.; Liu, X.; Su, S. Larva-Mediated Chalkbrood Resistance-Associated Single Nucleotide Polymorphism Markers in the Honey Bee *Apis mellifera*. *Insect Mol. Biol.* **2016**, *25*, 239–250. [[CrossRef](#)]
28. DIBt Insights into Social Insects from the Genome of the Honeybee *Apis mellifera*. *Nature* **2006**, *443*, 931–949. [[CrossRef](#)]
29. Park, D.; Jung, J.W.; Choi, B.-S.; Jayakodi, M.; Lee, J.; Lim, J.; Yu, Y.; Choi, Y.-S.; Lee, M.-L.; Park, Y.; et al. Uncovering the Novel Characteristics of Asian Honey Bee, *Apis cerana* by Whole Genome Sequencing. *BMC Genom.* **2015**, *16*, 1. [[CrossRef](#)]
30. Nelson, J.C.; Wang, S.; Wu, Y.; Li, X.; Antony, G.; White, F.F.; Yu, J. Single-Nucleotide Polymorphism Discovery by High-Throughput Sequencing in Sorghum. *BMC Genom.* **2011**, *12*, 352. [[CrossRef](#)]
31. Rizwan, M.; Liang, P.; Ali, H.; Li, Z.; Nie, H.; Saqib, A.; Fiaz, S.; Fahad, M.; Kalan, A.; Niu, Q.; et al. Population Genomics of Honey Bees Reveals a Selection Signature Indispensable for Royal Jelly Production. *Mol. Cell. Probes* **2020**, *52*, 101542. [[CrossRef](#)]
32. Chen, C.; Parejo, M.; Momeni, J.; Langa, J.; Nielsen, R.O.; Shi, W.; Vingborg, R.; Kryger, P.; Bouga, M.; Estonba, A.; et al. Population structure and diversity in European honey bees (*Apis mellifera* L.)—An empirical comparison of pool and individual whole-genome sequencing. *Genes* **2022**, *13*, 182. [[CrossRef](#)] [[PubMed](#)]
33. Kim, S.; Misra, A. SNP Genotyping: Technologies and Biomedical Applications. *Annu. Rev. Biomed. Eng.* **2007**, *9*, 289–320. [[CrossRef](#)]
34. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST : A New Generation of Protein Database Search Programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [[CrossRef](#)]
35. Localization, I. Integrated Nr Database in Protein Annotation System and Its Localization. *Comput. Eng.* **2006**, *32*, 71–74.
36. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology : Tool for the Unification of Biology.The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)]
37. Tatusov, R.L.; Galperin, M.Y.; Natale, D.A.; Koonin, E.V. The COG Database : A Tool for Genome-Scale Analysis of Protein Functions and Evolution. *Nucleic Acids Res.* **2000**, *28*, 33–36. [[CrossRef](#)] [[PubMed](#)]
38. Kanehisa, M.; Goto, S.; Kawashima, S.; Okuno, Y. The KEGG Resource for Deciphering the Genome. *Nucleic Acids Res.* **2004**, *32*, D277–D280. [[CrossRef](#)]
39. Wragg, D.; Marti-Marimon, M.; Basso, B.; Bidanel, J.P.; Labarthe, E.; Bouchez, O.; Le Conte, Y.; Vignal, A. Whole-Genome Resequencing of Honeybee Drones to Detect Genomic Selection in a Population Managed for Royal Jelly. *Sci. Rep.* **2016**, *6*, 27168. [[CrossRef](#)]
40. Mckenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit : A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data. *Genome Res.* **2010**, *20*, 1297–1303. [[CrossRef](#)]

41. Pfeifer, B.; Wittelsbürger, U.; Ramos-Onsins, S.E.; Lercher, M.J. PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R. *Mol. Biol. Evol.* **2014**, *31*, 1929–1936. [[CrossRef](#)]
42. Thi, N.; Nguyen, B.; Le, T.H. Complete Genome Sequence of Sacbrood Virus Strain SBM2, Isolated from the Honeybee *Apis cerana* in Vietnam. *Genome Announc.* **2013**, *1*, 76–88.
43. Hu, Y.; Fei, D.; Jiang, L.; Wei, D.; Li, F.; Diao, Q.; Ma, M. A Comparison of Biological Characteristics of Three Strains of Chinese Sacbrood Virus in *Apis Cerana*. *Sci. Rep.* **2016**, *6*, 37424. [[CrossRef](#)]
44. Li, H.; Durbin, R. Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)]
45. Clément, J.A.J.; Toulza, E.; Gautier, M.; Parrinello, H.; Roquis, D.; Boissier, J.; Rognon, A.; Moné, H.; Mouahid, G.; Buard, J.; et al. Private Selective Sweeps Identified from Next-Generation Pool-Sequencing Reveal Convergent Pathways under Selection in Two Inbred *Schistosoma Mansoni* Strains. *PLoS Negl. Trop. Dis.* **2013**, *7*, e2591. [[CrossRef](#)]
46. Han, B.; Zhang, L.; Feng, M.; Fang, Y.; Li, J. An Integrated Proteomics Reveals Pathological Mechanism of Honeybee (*Apis Cerana*) Sacbrood Disease. *J. Proteome Res.* **2013**, *12*, 1881–1897. [[CrossRef](#)]
47. Su, Y.L.; Li, J.M.; Li, M.; Luan, J.B.; Ye, X.D.; Wang, X.W.; Liu, S.S. Transcriptomic Analysis of the Salivary Glands of an Invasive Whitefly. *PLoS ONE* **2012**, *7*, e39303. [[CrossRef](#)]
48. Ma, L.; Ma, Q.; Li, X.; Cheng, L.; Li, K.; Li, S. Transcriptomic Analysis of Differentially Expressed Genes in the Ras1CA-Overexpressed and Wildtype Posterior Silk Glands. *BMC Genom.* **2014**, *15*, 182. [[CrossRef](#)]
49. Hu, H.; Bezabih, G.; Feng, M.; Wei, Q.; Zhang, X.; Wu, F.; Meng, L.; Fang, Y.; Han, B.; Ma, C.; et al. In-Depth Proteome of the Hypopharyngeal Glands of Honeybee Workers Reveals Highly Activated Protein and Energy Metabolism in Priming the Secretion of Royal Jelly. *Mol. Cell. Proteom.* **2019**, *18*, 606–621. [[CrossRef](#)]
50. Ubiparip, Z.; Procházková, M.; Moravcová, J.; Plevka, P.; Škubník, K.; Přidal, A.; Füzik, T.; Škubník, K.; Moravcová, J.; Ubiparip, Z.; et al. Virion Structure and Genome Delivery Mechanism of Sacbrood Honeybee Virus. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 7759–7764.
51. Wang, Z.L.; Zhu, Y.Q.; Yan, Q.; Yan, W.Y.; Zheng, H.J.; Zeng, Z.J. A chromosome-scale assembly of the asian honeybee *apis cerana* Genome. *Front. Genet.* **2020**, *11*, 279. [[CrossRef](#)]
52. de Celorio-Mancera, M.I.P.; Courtiade, J.; Muck, A.; Heckel, D.G.; Musser, R.O.; Vogel, H. Sialome of a Generalist Lepidopteran Herbivore: Identification of Transcripts and Proteins from *Helicoverpa Armigera* Labial Salivary Glands. *PLoS ONE* **2011**, *6*, e26676. [[CrossRef](#)]
53. Fan, P.; Han, B.; Feng, M.; Fang, Y.; Zhang, L.; Hu, H.; Hao, Y.; Qi, Y.; Zhang, X.; Li, J. Functional and Proteomic Investigations Reveal Major Royal Jelly Protein 1 Associated with Anti-Hypertension Activity in Mouse Vascular Smooth Muscle Cells. *Sci. Rep.* **2016**, *6*, 30230. [[CrossRef](#)] [[PubMed](#)]
54. Holman, L.; Helanterä, H.; Tronetti, K.; Mikheyev, A.S. Comparative Transcriptomics of Social Insect Queen Pheromones. *Nat. Commun.* **2019**, *10*, 1593. [[CrossRef](#)]
55. Tantillo, G.; Bottaro, M.; Di Pinto, A.; Martella, V.; Di Pinto, P.; Terio, V. Virus Infections of Honeybees *Apis mellifera*. *Ital. J. Food Saf.* **2015**, *4*, 157–168. [[CrossRef](#)]
56. Sun, L.; Li, M.; Fei, D.; Hu, Y.; Ma, M. Chinese Sacbrood Virus Infection in *Apis mellifera*, Shandong, China, 2016. *Virus Res.* **2017**, *242*, 96–99. [[CrossRef](#)]
57. Abdullah, I.; Gary, S.R.; Marla, S. Field Trial of Honey Bee Colonies Bred for Mechanisms of Resistance against Varroa Destructor. *Apidologie* **2007**, *38*, 67–76.
58. Hassanyar, A.K.A.K.; Huang, S.; Li, Z.; Rizwan, M.; Mehmood, S.; Raza, M.F.M.F.; Qasim, M.; Hussain, M.; Su, S. Prevalence of Bee Viruses in *Apis cerana* Cerana Populations from Different Locations in the Fujian Province of China. *Microbiologyopen* **2019**, *8*, e830. [[CrossRef](#)] [[PubMed](#)]
59. Maori, E.; Lavi, S.; Mozes-Koch, R.; Gantman, Y.; Peretz, Y.; Edelbaum, O.; Tanne, E.; Sela, I. Isolation and Characterization of Israeli Acute Paralysis Virus, a Dicistrovirus Affecting Honeybees in Israel: Evidence for Diversity Due to Intra- and Inter-Species Recombination. *J. Gen. Virol.* **2007**, *88*, 3428–3438. [[CrossRef](#)] [[PubMed](#)]
60. Zhou, Q.; Qi, Y.; Yang, F. Application of Spectrophotometry to Evaluate the Concentration of Purified White Spot Syndrome Virus. *J. Virol. Methods* **2007**, *146*, 288–292. [[CrossRef](#)]
61. Grabensteiner, E.; Ritter, W.; Carter, M.J.; Davison, S.; Pechhacker, H.; Kolodziejek, J.; Boecking, O.; Derakhshifar, I.; Moosbeckhofer, R.; Licek, E.; et al. Sacbrood Virus of the Honeybee (*Apis mellifera*): Rapid Identification and Phylogenetic Analysis Using Reverse Transcription-PCR. *Clin. Vaccine Immunol.* **2001**, *8*, 93–104. [[CrossRef](#)]
62. Gauthier, L.; Tentcheva, D.; Tournaire, M.; Dainat, B.; Cousserans, F.; Colin, M.E.; Bergoin, M.; Dainat, B.; Tournaire, M.; Colin, M.E.; et al. Viral Load Estimation in Asymptomatic Honey Bee Colonies Using the Quantitative RT-PCR Technique. *Apidologie* **2007**, *38*, 426–435. [[CrossRef](#)]
63. Huang, W.F.; Zhang, Y.; Mehmood, S.; Wang, Z.; Hou, C.; Li, Z. Updating Sacbrood Virus Quantification Pcr Method Using a Taqman-Mgb Probe. *Vet. Sci.* **2021**, *8*, 63. [[CrossRef](#)]
64. Crailsheim, K.; Brodschneider, R.; Aupinel, P.; Behrens, D.; Genersch, E.; Vollmann, J.; Riessberger-Gallé, U. Standard Methods for Artificial Rearing of *Apis mellifera* Larvae. *J. Apic. Res.* **2015**, *52*, 1–16. [[CrossRef](#)]
65. Aupinel, P.; Fortini, D.; Dufour, H.; Tasei, J.N.; Michaud, B.; Odoux, J.F.; Pham-Delègue, M.H. Improvement of Artificial Feeding in a Standard in Vitro Method for Rearing *Apis mellifera* Larvae. *Bull. Insectology* **2005**, *58*, 107–111.

66. Van Toan, T.; Lee, M.L.; Sim, H.S.; Kim, H.K.; Byuon, G.H.; Choi, Y.S. Initial Results of Rearing Honey Bee *Apis cerana* in Vitro. *J. Apic.* **2014**, *29*, 193–197. [[CrossRef](#)]
67. Schmehl, D.R.; Tomé, H.V.V.; Mortensen, A.N.; Martins, G.F.; Ellis, J.D. Protocol for the in Vitro Rearing of Honey Bee (*Apis mellifera* L.) Workers. *J. Apic. Res.* **2016**, *55*, 113–129. [[CrossRef](#)]
68. Paterson, A.H.; Brubaker, C.L.; Wendel, J.F. A Rapid Method for Extraction of Cotton (*Gossypium* Spp.) Genomic DNA Suitable for RFLP or PCR Analysis. *Plant Mol. Biol. Rep.* **1993**, *11*, 122–127. [[CrossRef](#)]
69. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Data, G.P.; et al. The Sequence Alignment / Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)] [[PubMed](#)]
70. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M.; Cingolani, P.; et al. A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff. *Fly* **2012**, *6*, 80–92. [[CrossRef](#)]
71. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. *Nucleic Acids Res.* **2017**, *45*, D353–D361. [[CrossRef](#)]
72. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X : Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [[CrossRef](#)] [[PubMed](#)]
73. Price, A.L.; Patterson, N.J.; Plenge, R.M.; Weinblatt, M.E.; Shadick, N.A.; Reich, D. Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies. *Nat Genet.* **2006**, *38*, 904–909. [[CrossRef](#)]
74. Alexander, D.H.; Novembre, J.; Lange, K. Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome Res.* **2009**, *19*, 1655–1664. [[CrossRef](#)]
75. Vekemans, O.J.H.; Vekemans, X. Spagedi: A Versatile Computer Program to Analyse Spatial Genetic Structure at the Individual or Population Levels. *Mol. Ecol. Notes* **2002**, *2*, 618–620.
76. Li, J.; Qin, H.; Wu, J.; Sadd, B.M.; Wang, X.; Evans, J.D.; Peng, W.; Chen, Y. The Prevalence of Parasites and Pathogens in Asian Honeybees *Apis cerana* in China. *PLoS ONE* **2012**, *7*, e47955. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.