ISSN 1022-7954, Russian Journal of Genetics, 2019, Vol. 55, No. 7, pp. 815–824. © Pleiades Publishing, Inc., 2019.
Russian Text © The Author(s), 2019, published in Genetika, 2019, Vol. 55, No. 7, pp. 778–787.

REVIEWS
AND THEORETICAL ARTICLES

The Role of Whole-Genome Studies in the Investigation of Honey Bee Biology

U. B. Yunusbaev[a, b], M. D. Kaskinova[a, *], R. A. Ilyasov[a, b], L. R. Gaifullina[a], E. S. Saltykova[a], and A. G. Nikolenko[a]

[a] *Institute of Biochemistry and Genetics, Ufa Federal Research Center, Russian Academy of Sciences, Ufa, 450054 Russia*
[b] *Incheon National University, South Korea, Yeonsu-gu, Incheon, 22012 South Korea*
**e-mail: kaskinovamilyausha@mail.ru*

Received October 25, 2018; revised December 7, 2018; accepted January 15, 2019

Abstract—Given is an overview of original publications devoted to the investigation of the honey bee genome. The history of honey bee genome studies and its characteristics are described. The results of genetic studies of honey bees using genome-wide data are presented. A special focus is put on the search for alleles associated with economically valuable, adaptive, and other important honey bee traits.

Keywords: Apis mellifera, honey bee, genome, whole-genome data analysis, allele-associated traits, SNP
DOI: 10.1134/S102279541906019X

INTRODUCTION

Decoding the genome of *Apis mellifera* L., 1758 has opened up new possibilities for studying the biology of honey bees and increasing the effectiveness of applied research for beekeeping. Whole-genome studies made it possible to clarify and to a large extent complement the previously obtained data on the evolution, distribution, and genetic diversity of honey bees [1–4]. These studies clarified the long-standing scientific disputes on the genealogy of different honey bee subspecies. It should be noted that the aforementioned fundamental studies are the basis of breeding work for the development of environmentally and economically sustainable highly productive beekeeping. In addition, whole-genome studies open up new opportunities for the identification and conservation of rare honey bee populations with unique gene pools.

Since the publication of the complete genome sequence of honey bee in 2006, a considerable amount of genome-wide studies devoted to the study of the honey bee biology has been accumulated in the world literature. At present, there is an obvious need for systematization and synthesis of scientific information collected in this area. This review is devoted to the current state of genomic studies of honey bee and the determination of its promising areas.

CHARACTERISTICS OF THE HONEY BEE GENOME

According to the latest published data, the complete genome sequences were obtained for the following bee species: honey bee *Apis mellifera* (GenBank accession number GCA_003254395.2) [5], eastern honey bee *A. cerana* Fabricius, 1793 (GCA_002290385.1) [6], giant honey bee *A. dorsata* Fabricius, 1793 (GCA_000469605.1) [7], and dwarf honey bee *A. florea* Fabricius, 1787 (GCA_000184785.1) [8, 9]. For science and beekeeping, information about the genome of *A. mellifera* is of particular interest. In this review, we focused only on the papers devoted to the study of the *A. mellifera* genome.

According to research data, the haploid genome of *A. mellifera* includes 16 chromosomes, which consist of 225178604 base pairs (bp) [5]. The honey bee nuclear DNA contains 15314 annotated genes [10]. The main features of this DNA should be mentioned. The researchers note [2] that nuclear DNA of the honey bee evolved more slowly compared to the DNA sequences of *Drosophila melanogaster* Meigen, 1830 and *Anopheles gambiae*, Giles 1902. This is indicated by the higher number of introns in the honey bee genome, as well as the presence of genes lost in fly and mosquito. Slower evolution of honey bees can be explained by the fact that most honey bee genes are located in AT-rich regions [2]. AT-rich regions, in contrast to GC-rich regions, are characterized by low recombination frequency [11, 12].

The mean recombination frequency in honey bee is 19–37 cM/1000 bp [11, 13]. This is more than five crossover events per each chromosome pair in each meiotic division. At the same time, in most sexually reproducing eukaryotes, the mean recombination frequency rarely exceeds one crossover event per one chromosome arm [12]. As in some other taxa [14], in

honey bee, a high degree of correlation between the recombination frequency and the GC content was demonstrated.

In the course of evolution, a number of honey bee genes reoriented functions in association with the transition to eusociality. In particular, gene families encoding royal jelly protein [2] and vitellogenin [15] began to perform functions associated with honey bee caste differentiation. An increase in the number of olfactory receptor genes [2] and switching of the *CYP450* genes to hormone metabolism [16] ensured the communication of social insects. An important role in the honey bee genome is played by miRNAs that affect age- and caste-specific gene expression [2].

The mitochondrial genome of honey bee consists of 16463 bp, including 13 coding genes, 22 tRNA genes, and 2 rRNA genes [5]. The honey bee mitochondrial DNA, in comparison with the similar *Drosophila* molecule, has longer intergenic noncoding sequences. In addition, it was demonstrated that 84.9% of the honey bee mtDNA comprises AT-rich regions [17].

THE HISTORY OF THE HONEY BEE GENOMIC STUDIES

The history of the honeybee genome sequencing began with the reads of individual mitochondrial DNA fragments. In particular, in 1989, the sequences of the intergenic locus between the *COI* and *COII* genes and several mtDNA tRNA genes were published [18]. In 1993, the honey bee mitochondrial genome of 16343 bp was completely sequenced using specimens of working bees *A. m. ligustica* [17]. In this study, in the honey bee mitochondrial genome, 13 protein-coding genes, 22 tRNA genes, and 2 rRNA genes were annotated. The mitochondrial genome of *Drosophila yakuba* was used as reference genome.

The earliest studies focused on the investigation of the honey bee nuclear DNA contain information on nucleotide sequences of rRNA subunits [19], microsatellite loci [20, 21], royal jelly protein-encoding genes [22], and other protein-coding genes [23, 24].

In 2005, the Honeybee Genome Sequencing Consortium (HGSC) placed the complete honey bee genome assembly, Amel_4.0, in the GenBank database (GCF_000002195.3). It should be noted that all previous versions, from Amel_1.0 to Amel_3.0, were intermediate and were not considered in subsequent publications of the Consortium.

In 2006, on the basis of the analysis of the Amel_4.0 assembly, HGSC published in the journal *Nature* one of the most important studies devoted to the analysis of the honey bee genome [2]. The authors first published information on the nuclear genome size of *A. mellifera*, which, according to their estimates, amounted to 267504338 bp and included 10157 genes. In the study, the genomes of 175 specimens of *A. mellifera* collected from different geographical ranges of the species in Africa, Europe, and Asia were examined. The sample was represented by the following subspecies: *A. m. mellifera* L., 1758 ($N = 20$), *A. m. iberiensis* Engel., 1999 (11), *A. m. ligustica* Spinola, 1806 (18), *A. m. carnica* Pollmann, 1879 (16), *A. m. anatoliaca* Maa, 1953 (18), *A. m. caucasica* Pollmann, 1889 (14), *A. m. syriaca* Skorikov, 1829 (9), *A. m. scutellata* Lepeletier, 1836 (21), *A. m. lamarckii* Cockerell, 1906 (19), and *A. m. intermissa* Maa, 1953 (19). In addition, the genomes of 13 specimens of other honey bee species (*A. cerana*, *A. dorsata*, and *A. florea*) were analyzed. All specimens were represented by worker bees. The sample was described in detail earlier [1]. For the genomes of these subspecies, 1536 single nucleotide polymorphisms (SNP) were typed.

In the next study, HGSC provides an improved assembly of the *A. mellifera* genome (Amel_4.5) [10]. The improved assembly of *A. mellifera* genome consists of 250287000 bp and includes 15314 genes. In studies of recent years, the latest version, the Amel_4.5 assembly, is used [4, 25, 26].

In addition to the complete genome assemblies reported by HGSC, there are other assemblies of the honey bee genome. For instance, in 2015, the *A. m. intermissa* genome was deposited in the GenBank database under the accession number GCA_000819425.1 [27]. In 2018, the INRA (French National Institute for Agricultural Research) added the genome of another honey bee subspecies, *A. m. mellifera*, to the GenBank database (GCA_003314205.1). In the last two papers, previously published data on the honey bee genome were confirmed.

Today, *A. mellifera* genome assembly Amel_HAv3 (GCA_003254395.1), obtained on the basis of four sequencing and mapping technologies (PacBio, 10× Chromium, BioNano, and Hi-Chere), is of the greatest interest [5]. The Amel_HAv3 assembly is much more informative and complete compared to Amel_4.5. The improvement was mainly achieved by the inclusion of repeated sequences that were not previously mapped. As mentioned above, the size of this genome assembly is 225178604 bp [5].

OPEN-ACCESS DATABASES WITH THE HONEY BEE GENOME SEQUENCES

Publication of the results of genomic research is usually accompanied by the deposition of genomic data in open-access electronic databases. Below is an overview of the electronic resources containing honey bee genomes available for download.

The Hymenoptera Genome Database (HGD) is a specialized resource containing genomic data for Hymenoptera. HGD consists of three divisions, BeeBase, containing information about the honey bee

genome; NasoniaBase, containing information about the genome of parasitic wasp *Nasonia vitripennis* Walker, 1836; and Ant Genomes Portal, a section with the genomic data of different ant species [28]. The HGD database contains FASTA-formatted Amel_2.0, Amel_3.0, Amel_4.0, and Amel_4.5 assemblies, as well as the official amel_OGSv1.0, amel_OGSv1.1, and OGSv3.2 gene sets.

The *A. mellifera* genome sequences are represented in the database resource of the National Center for Biotechnology Information (NCBI), the European nucleotide archive, and the Dryad Digital Repository.

The NCBI open-access resource (https://www.ncbi.nlm.nih.gov) contains eight assemblies of the *A. mellifera* genome, as well as more than 1000 full genome data sets in Sequence Read Archive (SRA). SRA is a set of archived data containing FASTQ- or BAM-formatted nucleotide sequences. GenBank is loaded with RNA and DNA sequences from both published and unpublished studies.

The European nucleotide archive is a resource containing annotated DNA and RNA sequences (https://www.ebi.ac.uk/ena). Most resource data duplicate those of NCBI.

The Dryad Digital Repository was developed by a group of scientific journals for posting additional information attached to scientific publications (https://datadryad.org). The resource contains the following genome-wide samples: 148 bee specimens for articles [3, 29], 244 specimens for article [30], and 176 specimens for article [31].

WHOLE-GENOME STUDIES IN THE INVESTIGATION OF THE HONEY BEE EVOLUTION

Analysis of whole-genome sequencing data from different honey bee populations enables reconstruction of the stages and time of their filiation. For instance, Whitfield et al. [1] on the basis of genome-wide analysis of 1136 SNPs of four honey bee species (*A. cerana*, *A. dorsata*, *A. florea*, and *A. mellifera*), as well as ten subspecies of *A. mellifera*, hypothesized that the present-day subspecies of *A. mellifera* arose in Africa and migrated to Eurasia. One migration wave penetrated to Europe through the Iberian Peninsula and then spread to the territory of Central Europe and Russia (M lineage); the second wave penetrated to Asia and Eastern Europe (lineages O and C, respectively).

Han et al. [32] reanalyzed data obtained earlier [1] by testing different methods of constructing phylogenetic trees and removing individuals with potentially hybrid origin from the sample. The obtained data ambiguously indicated African origin of the honey bee subspecies. The repeated analysis confirmed the high level of divergence between western and eastern populations of *A. mellifera*, which suggests that they arose as a result of two separate migrations. However, it remains unclear whether they migrated from Africa or Asia.

Wallberg et al. [3] analyzed 140 honey bee genomes from 14 populations using 8.3 million SNPs and also did not find evidence of the African origin of *A. mellifera*. However, this study shed light on the genetic basis of adaptations of the honey bee subspecies to habitat conditions. In particular, they showed that the sizes of the honey bee populations fluctuated considerably, reflecting changes in climatic conditions. At the same time, high genetic diversity of modern populations points to the absence of a bottleneck stage. The authors identified the signatures of local adaptations in the honey bee genomes [3]. The latter are expressed in the fact that genes associated with the immune system and sperm motility are enriched in SNPs that are under positive selection. Differences in the expression of these genes may underlie geographical differences in reproduction, dispersal, and disease resistance.

Possible reasons for the lack of an unequivocal answer to the question of the origin of *A. mellifera* are the small sample size and the use of individuals of hybrid origin. Taking into account these factors, researchers [4] conducted a phylogenetic analysis of honey bees from a combined sample which included a set of data obtained by other authors [3, 33] and their personal whole-genome data for *A. mellifera*. The resulting sample contained honey bees from five evolutionary lineages of *A. mellifera* (A, C, M, O, and Y) and covered a wide geographic range, including Africa, Europe, and the Middle East. Their data favored the origin of *A. mellifera* in the Middle East or Northeast Africa, with the ancestral population the closest to the A and Y evolutionary lineages. According to the supposition of Cridland et al. [4], the dispersal of ancestral honey bees from Africa to the Middle East and Europe gave rise to other evolutionary lineages. There was a secondary contact between the diverged Middle Eastern populations, which led to the emergence of new subspecies. Thus, this study showed that the evolution of honey bees was a complex process consisting of several migration waves, divergence, and secondary contacts of once segregated populations.

Genome-wide data are also used to establish the evolution of eusociality. Comparing the genomes of ten bee species with different social organization, Kapheim et al. [34] showed that eusociality had several independent pathways of origin, but it was always accompanied by the increasing complexity of the gene regulation network.

Kent et al. [35] in their study on the honey bee social behavior came to three main conclusions. First, they confirmed the correlation between the recombination frequency and the GC content. Second, a high degree of correlation was found between the recombination frequency and the rate of molecular evolution, as well as between the GC content and the rate of

molecular evolution. Third, a large number of genes associated with behavior and genes in GC-enriched regions specifically expressed in the brain of worker bees were identified.

THE ROLE OF WHOLE-GENOME STUDIES IN THE INVESTIGATION AND CONSERVATION OF THE HONEY BEE GENETIC DIVERSITY

The species *A. mellifera* is divided into 31 subspecies, each of which is adapted to the peculiarities of its habitat. The spread of package beekeeping and the uncontrolled movement of bee colonies led to the hybridization of the aboriginal subspecies and the loss of their adaptive potential [36, 37]. There was a need for a reliable method for identifying the honey bee subspecies. The first study using genome-wide data for differentiation of the honey bee subspecies was published in 2006 [1]. Using 1136 SNPs, the authors analyzed a sample of 328 individuals, including 175 individuals from the natural distribution range of the honey bee subspecies (14 subspecies from Europe, Africa, and Asia) and 153 individuals from the introduced populations in North and South America. Principal component analysis showed the division of this sample into four evolutionary lineages, the A lineage, which included subspecies from Africa; M lineage, represented by the subspecies from Western and Northern Europe; C lineage, comprising subspecies from Eastern Europe; and O lineage, which included subspecies from the Middle East and Central Asia. The authors also identified 19 SNPs differentiating the European and African subspecies of honey bee. This study determined two themes for further research on the genetic structure of honey bee populations. The first theme is the search for the remaining purebred populations of *A. mellifera*. The second theme is the identification of Africanized honey bees on the American continents that are extremely dangerous to humans and domestic animals.

The main part of the studies aimed at the search for preserved purebred populations is focused on the analysis of honey bees belonging to the M evolutionary lineage, *A. m. mellifera* and *A. m. iberiensis*. This is associated with the fact that these subspecies suffered more or may suffer from hybridization. For instance, in 2014, the study of Pinto et al. [36] was published, in which the SNP set from study [1] in combination with the analysis of the *tRNAleu–cox2* mtDNA intergenic locus was used to assess the genetic diversity and the level of hybridization in populations of the European dark honey bee *A. m. mellifera* from protected and unprotected areas of Northern Europe. Analysis of the honey bee mtDNA from protected areas showed that only one colony had a foreign haplotype, whereas SNP genotyping showed different levels of introgression, varying from almost zero in Norway and Scotland to 14% in Denmark. The introgression was higher in unprotected areas (30%) than in protected areas (8%).

In the next paper [38], an attempt was made to reduce the initial SNP set [1] for routine identification of *A. m. mellifera* populations. On the basis of this set, the authors developed sets of 48, 96, 144, 192, and 384 SNPs. The set of 384 SNPs gave more statistically significant results. At the same time, panels with lower SNP numbers were also able to determine the origin and estimate the level of introgression in populations of the European dark honey bee.

In the study of Parejo et al. [39], whole-genome sequencing data (3.375 million SNPs) were used for 120 drones of *A. m. mellifera*, *A. m. carnica*, and the hybrid Bukfast breed, collected in Switzerland and France. The sample of *A. m. mellifera* was composed of honey bees from protected and unprotected areas. The authors tested the sets of 1000, 500, 100, and 50 SNPs, selected in three different ways (FST, PCA, and random selection). As a result, a set of 50 SNPs was shown to effectively differentiate the subspecies under study. In Northern Europe, hybridization proceeded more intensely outside the protected areas [36], while in Western European countries, the introgression of the foreign gene pool was recorded in all protected populations of *A. m. mellifera*. In Western Europe, outside the protected areas, low introgression proportions were identified, pointing to the possibility of purebred breeding also without taking measures to isolate the native population. The authors explain this in terms that beekeepers in Western Europe are interested in breeding of local honey bee. They bring their queen bees to play flights to specialized stations of controlled mating or purchase them from certified beekeepers.

Spanish honey bee *A. m. iberiensis*, compared to *A. m. mellifera*, was less affected by hybridization with Eastern European and African subspecies [40]. This is facilitated by beekeepers, because they prefer to breed local honey bees. However, this favorable situation may change owing to climate change and the availability of commercial package bees. In this regard, Henriques et al. [31] performed a genome-wide analysis of *A. m. iberiensis* ($N = 117$) and the two most popular commercial subspecies, *A. m. carnica* ($N = 28$) and *A. m. ligustica* ($N = 31$). A total of 11091 polymorphic SNPs that differentiated Spanish honey bee from these two subspecies belonging to the C evolutionary lineage were identified.

Subspecies identification is important not only for the preservation of native honey bee subspecies but also to restrict the dispersal of invasive subspecies and their hybrids. In 1956, the Africanized honey bee was produced in Brazil by crossing-breeding of the African honey bee *A. m. scutellata* with the European populations of *A. mellifera*. The resulting hybrids were distinguished by increased aggressiveness and viability. These honey bees spread throughout the continent,

replacing local bee populations. They are known as "killer bees" because they are easily provoked to attack, moreover, they attack in a large swarm and can chase the victim for a long distance. An increase in the amount of the genome-wide data for subspecies from Africa initiated the solution of the sensitive issue with Africanized bees. On the basis of previous studies [1, 3], Chapman et al. [41] developed a set of 95 SNPs allowing for the differentiation between the subspecies from Africa (*A. m. scutellata*) and Europe (*A. m. ligustica*, *A. m. carnica*, *A. m. mellifera*, and *A. m. iberiensis*). Testing of this kit showed that the highest proportions of introgressed African alleles were in populations from Brazil (80.5%) and the United States (62.5%). In honey bee populations from Canada and Australia, this index had small values (3.4–4.9% on average). The authors suggested that the introgression level of African alleles, which was higher than 15%, was the reason for the ban on the import of the honey bee colonies. A new data set for the Africanized honey bees was obtained by Kadri et al. [42]. They carried out whole-genome sequencing of 360 worker bees from 30 Africanized colonies. The data obtained can be used to develop the SNP panel for routine identification of the Africanized honey bees.

Genome-wide studies were also used to establish the origin of honey bee populations from the New World and Australia, where originally the honey bee was not found. For the Australian [43] and Canadian [44] populations, it was demonstrated that they originated from the European subspecies belonging to phylogenetic lineages M and C and also contained a small proportion of African alleles.

To obtain information on the population genetic structure from a set of whole-genome sequencing data, a number of methods and software tools have been developed. These include principal component analysis (PCA), cluster analysis of populations using the Admixture and Structure software programs [45, 46], and also the evaluation of *F* statistics [47].

A GENOME-WIDE SEARCH FOR ALLELES ASSOCIATED WITH THE HONEY BEE TRAITS

Genome-wide association study (GWAS) is a new approach in genetic research that is aimed at the identification of associations between genetic variants and phenotypic traits. In some cases, a certain genetic variant directly affects the phenotype through gene regulation. In other cases, there may be the usual correlation between the gene variant and the trait. Studies of honey bees showed that most of the economically valuable traits and disease resistance were under polygenic control. For example, quantitative trait loci associated with queen fertility [48, 49], resistance to chalkbrood [50] and varroosis [51], and also foraging [52, 53], hygienic [54], and guarding [55] types of behavior were identified. However, no specific genetic markers that could be used in bee breeding were proposed.

Alleles Associated with Resistance to Pathogens and Parasites

To date, three papers have been published aimed at finding variants associated with resistance to pathogens and parasites. The first such investigation was paper [56]. The authors analyzed three samples of honey bees using 70000 SNPs to search for alleles associated with the resistance to the parasitic mite *Varroa destructor* Anderson & Trueman, 2000. Resistance here means the ability of honey bees to detect a mite-infested brood and uncap and remove it. This is so-called hygienic behavior. The first sample included honey bees showing high levels of hygienic behavior. The second sample included honey bees that were the offspring of the same queens as honey bees from the first group, but showing no hygienic behavior. The third sample consisted of honey bees showing no hygienic behavior and that were not related to honey bees from the first group. Most of the polymorphic SNPs were located at previously identified quantitative trait loci [51, 54] associated with hygienic behavior. On the basis of these 70000 SNPs, a panel of 44000 SNPs was developed that differentiated between individuals with hygienic and non-hygienic behavior. In 2016, this group of researchers conducted another study using this SNP panel [28]. This time, two samples of honey bees were analyzed, each consisting of 122 individuals. One of the samples belonged to the line of honey bees with pronounced hygienic behavior; the second, control, sample belonged to the line of honey bees not performing such behavior. After FDR correction of the *p*-values, 6 SNPs associated with mite tolerance were identified: AMB-00457689 (T → C), AMB-00386078 (C → T), AMB-00573174 (G → A), AMB-00913945 (A → G), AMB-01079196 (A → G), and AMB-00745078 (T → C). The candidate genes selected for four SNPs were as follows: *Adenosine receptor*, *Cyclin-dependent kinase 5 activator*, *Octopamine receptor beta-2R,* and *Odorant binding protein 1.* The *Varroa* mite causes substantial damage to beekeeping not only by its direct influence on honey bees but also by the fact that it is the vector of many viral diseases. In addition, the use of acaricides, to which mites develop resistance over time, makes honey products unsuitable for use and export, as they accumulate acaricide residuals. The way out of the situation is the selection of honey bees that can cope with the mite themselves.

Another common disease of honey bees is chalkbrood, caused by the fungus *Ascosphaera apis* Spiltoir and Olive, 1955. The chalkbrood resistance is also determined by hygienic behavior. Liu et al. [57] conducted an analysis of SNPs located on chromosomes 2 and 11 of chalkbrood-susceptible and chalkbrood-resistant honey bees. These chromosomes were not

chosen by chance. It was previously established [50] that these chromosomes carried quantitative trait loci associated with chalkbrood resistance. The authors of the study identified an association between the *C2587245T* locus *C* allele and chalkbrood resistance. This locus is located in the *Mrjp5* gene encoding royal jelly protein.

Alleles Associated with Social Behavior

Studies aimed at establishing the genetic architecture of social behavior are mainly devoted to age-related polyethism of honey bees and caste differences. Southey et al. [58] carried out an analysis of whole-genome sequencing data of 44 honey bee specimens to search for genetic variants associated with the behavior of scout bees. Among 1412705 polymorphic variants, 212 were statistically significantly associated with scouting behavior ($p < 0.0001$). Most of the variants were mapped to chromosomes 5, 12, and 15. The most frequent binding sites corresponded to putative transcription factors *Broad-complex 4* (18 positions), *Broad-complex 1* (16 positions), *Hunchback* (13 positions), and *chorion factor 2* (13 positions). Three variants associated with scouting behavior were located within noncoding RNA (ncRNA). Statistically significant variants were found on the 5'-untranslated region of the *membrin* gene and the 3'-untranslated regions of the *laccase 2* and *diacylglycerol kinase theta* genes. The 60 significant variants were located within introns of 39 genes, most of which were more than 1000 bp apart from each other. Functional categories of genes containing significant variants included nervous system, exoskeleton formation, immune response, salivary gland development, and enzymatic food processing.

Among the 31 honey bee subspecies, there is one whose social organization goes beyond the traditional framework of eusociality. Cape honey bee *A. m. capensis* Eschscholtz, 1822, which inhabits South Africa, differs from other subspecies in that its working individuals are capable of producing diploid viable offspring from unfertilized eggs by means of thelytoky. This phenomenon is called social parasitism.

To identify variants associated with this trait, Wallberg et al. [59] analyzed the genomes of Cape honey bees ($N = 10$) from South Africa, *A. m. scutellata* ($N = 10$) from Pretoria, and *A. m. adansonii* ($N = 10$) from Nigeria. They showed that most of the variants associated with social parasitism were located in hormone signaling genes and genes controlling the process of chromosome segregation. The former may cause ovarian activation in worker bees, while the latter may contribute to switching from normal meiosis to thelytoky. In addition, the authors did not find evidence that the *thelytoky* locus, located on chromosome 13 [60], was associated with social parasitism in the Cape honey bee.

Alleles Associated with Adaptive Traits

Honey bee is fascinating model system for studying local adaptations, as it originated in a tropical/subtropical climate and gradually spread to regions with a temperate and extreme continental climate. In addition, honey bees of the same subspecies inhabiting different geographic zones form local ecotypes that differ in physiological and biochemical traits.

Chen et al. [25] resequenced the whole genomes of ten individual bees of a newly discovered subspecies belonging to the M evolutionary lineage, *A. m. sinisxinyuan*, living in a temperate region of China. Comparative analysis of the genomes of populations from regions with temperate and tropical climates revealed several candidate genes that regulated the fat body synthesis and the *Hippo* signaling pathway. The preparedness of honey bees for the winter period depends on the degree of fat body development. The *Hippo* signaling pathway controls body size by regulating cell proliferation and apoptosis.

Fuller et al. [61] analyzed whole-genome sequences with the coverage of 3.6 million SNPs of 11 individual bees from Kenya collected from different ecological zones (savanna and desert populations). Despite its small size, the sample showed the division into two groups corresponding to two ecological zones. A number of genes and gene regions associated with adaptive evolution were identified. Many of these genes are involved in metabolism regulation (*Foxo*, *NDUFB2*, *takeout*), development of neuroplasticity and nervous system (*Neuroligin 3*, *RpA70*, *ZFYVE26*), resistance to parasites (*Foxo*, *peptidoglycan recognition protein-LC*, *relish*, *helicase 25E*, *hemolectin*), reproduction (*armitage* and *dunce*), and gland development and secretion (*Api M6*, *MRJP1*, *thickveins*).

In the next study [29], whole-genome sequencing of two populations of Kenyan honey bees was performed. The genome sequencing data with the coverage of 8.6 million SNPs of 39 worker bees *A. m. monticola* from the highlands and plains were analyzed. Despite the very low level of genetic differentiation between these populations, the authors identified two loci on chromosomes 7 and 9 which were fixed for the highland and lowland populations. The patterns of polymorphism of genetic variants suggest that recombination between these loci is suppressed, indicating that they comprise independent structural variants. The locus on chromosome 7 contains almost all octopamine receptor genes.

Henriques et al. [62] analyzed whole-genome sequencing data of 87 honey bees from the Iberian Peninsula and identified 670 genetic variants associated with precipitation, latitude, and longitude. Over 88.7% of these variants were located outside the exons. The *in silico* protein modeling showed that a number of nonsynonymous SNPs were likely targets for selection, since they led to amino acid substitutions in functionally important protein regions. The authors

revealed genomic signatures of local adaptations in 140 genes, many of which were associated with reproduction, immunity, olfaction, lipid synthesis, and circadian clock.

The following software programs were developed to search for associations between the genotype and adaptation to habitat conditions: PCAdapt [63], Samβada [64], and LFMM [65].

CONCLUSIONS

This review article presents an overview of publications focused on whole-genome research of honey bee. Some of these studies are aimed at establishing the phylogeny and evolutionary history of *Apis mellifera*. According to the latest data, the possible honey bee center of origin is the territory of the Middle East or Northeast Africa. The hypothetical ancestral honey bee population, belonging to either the A or Y evolutionary lineage, gave rise to 31 subspecies, which populated Eurasia and Africa and developed specific adaptations to different climatic zones.

In the genome-wide studies, special attention is paid to the ecological adaptations of different honey bee subspecies, since they are the basis for breeding practices. Another part of the genome-wide studies is aimed at finding the genetic variants associated with these adaptations. Such studies are still not widespread, but genetic variants associated with resistance to chalkbrood and varroosis, honey bee social structure, and adaptations to environmental conditions have been already identified.

Among the issues that remain to be resolved, phylogenetic analysis of whole-genome sequencing data of all known honey bee subspecies should be mentioned. Previous studies were limited to only ten subspecies [1, 3, 4, 33]. Another problem is the use of the obtained markers in practice. The implementation of the obtained genetic markers for breeding practices is hindered by specific features of the honey bee biology (polyandry, high frequency of recombination, complementary sex determination).

FUNDING

We are grateful to the anonymous reviewer for comments and valuable recommendations. Work of the 2nd and 5th authors was financially supported by the Russian Foundation for Basic Research and the Academy of Sciences of the Republic of Bashkortostan (project no. 17-44-020648); work of the 4th and 6th authors, by the State Assignment (no. AAAA-A16-116020350026-0); and work of the 1st and 3rd authors, by the postdoctoral fellowships of Incheon National University, Incheon, South Korea..

COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that they have no conflict of interest. This article does not contain any studies involving animals or human participants performed by any of the authors.

REFERENCES

1. Whitfield, C.W., Behura, S.K., Berlocher, S.H., et al., Thrice out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera, Science*, 2006, vol. 314, no. 5799, pp. 642—645. https://doi.org/10.1126/science.1132772
2. Honeybee Genome Sequencing Consortium, Insights into social insects from the genome of the honeybee *Apis mellifera, Nature*, 2006, vol. 443, no. 7114, pp. 931—949. https://doi.org/10.1038/nature05260
3. Wallberg, A., Han, F., Wellhagen, G., et al., A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera, Nat. Genet.*, 2014., V. 46, no. 10, pp. 1081—1088. https://doi.org/10.1038/ng.3077
4. Cridland, J.M., Tsutsui, N.D., and Ramírez, S.R., The complex demographic history and evolutionary origin of the western honey bee, *Apis mellifera, Genome Biol. EV*, 2017, vol. 9, no. 2, pp. 457—472. https://doi.org/10.1093/gbe/evx009
5. Wallberg, A., Bunikis, I., Pettersson, O.V., and Mosbech, M.B., A hybrid *de novo* genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. http://dx.doi.org/10.1101/361469
6. Park, D., Jung, J.W., Choi, B.-S., et al., Uncovering the novel characteristics of Asian honey bee, *Apis cerana*, by whole genome sequencing, *BMC Genomics*, 2015, vol. 16, p. 1. https://doi.org/10.1186/1471-2164-16-1
7. Takahashi, J.-I., Deowanish, S., and Okuyama, H., Analysis of the complete mitochondrial genome of the giant honeybee, *Apis dorsata* (Hymenoptera: Apidae) in Thailand, *Conserv. Genet. Resour.,* 2017. https://doi.org/10.1007/s12686-017-0942-7
8. Woodard, S.H., Fischman, B.J., Venkat, A., et al., Genes involved in convergent evolution of eusociality in bees, *Proc. Natl. Acad. Sci. U.S.A.*, 2011, vol. 108, no. 18, pp. 7472—7477. https://doi.org/10.1073/pnas.1103457108
9. Wang, A.R., Kim, M.J., Park, J.S., et al., Complete mitochondrial genome of the dwarf honeybee, *Apis florea* (Hymenoptera: Apidae), *Mitochondrial DNA*, 2013, vol. 24, no. 3, pp. 208—210. https://doi.org/10.3109/19401736.2012.744986
10. Elsik, C.G., Worley, K.C., Bennett, A.K., et al., Finding the missing honey bee genes: lessons learned from a genome upgrade, *BMC Genomics*, 2014, vol. 15, p. 86. https://doi.org/10.1186/1471-2164-15-86
11. Beye, M., Gattermeier, I., Hasselmann, M., et al., Exceptionally high levels of recombination across the honey bee genome, *Genome Res.*, 2006, vol. 16, no. 11,

pp. 1339—1344.
https://doi.org/10.1101/gr.5680406

12. Wallberg, A., Glémin, S., and Webster, M.T., Extreme recombination frequencies shape genome variation and evolution in the honeybee, *Apis mellifera, PLoS Genet.*, 2015, vol. 11, no. 4. e1005189.
https://doi.org/10.1371/journal.pgen.1005189
13. Liu, H., Zhang, X., Huang, J., et al., Causes and consequences of crossing-over evidenced via a high-resolution recombinational landscape of the honey bee, *Genome Biol.*, 2015, vol. 16, p. 15.
https://doi.org/10.1186/s13059-014-0566-0
14. Duret, L. and Galtier, N., Biased gene conversion and the evolution of mammalian genomic landscapes, *Annu. Rev. Genomics Hum. Genet.*, 2009, vol. 10, pp. 285—311.
https://doi.org/10.1146/annurev-genom-082908-150001
15. Nelson, C.M., Ihle, K.E., Fondrk, M.K., et al., The gene vitellogenin has multiple coordinating effects on social organization, *PLoS Biol.*, 2007, vol. 5, no. 3. e62.
https://doi.org/10.1371/journal.pbio.0050062
16. Wu, Y.Q., Zheng, H.Q., Corona, M., et al., Comparative transcriptome analysis on the synthesis pathway of honey bee (*Apis mellifera*) mandibular gland secretions, *Sci. Rep.*, 2017, no. 7(4530).
https://doi.org/10.1038/s41598-017-04879-z
17. Crozier, R.H. and Crozier, Y.C., The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization, *Genetics*, 1993, vol. 133, no. 1, pp. 97—117.
18. Crozier, R.H., Crozier, Y.C., and Mackinlay, A.G., The *CO-I* and *CO-II* region of honeybee mitochondrial DNA: evidence for variation in insect mitochondrial evolutionary rates, *Mol. Biol. EV*, 1989, vol. 6, no. 4, pp. 399—411.
https://doi.org/10.1093/oxfordjournals.molbev.a040553
19. Bigot, Y., Lutcher, F., Hamelin, M.H., and Periquet, G., The 28S ribosomal RNA-encoding gene of Hymenoptera: inserted sequences in the retrotransposon-rich regions, *Gene*, 1992, vol. 121, no. 2, pp. 347—352.
https://doi.org/10.1016/0378-1119(92)90142-C
20. Estoup, A., Solignac, M., Harry, M., and Cornuet, J.M., Characterization of (GT)n and (CT)n microsatellites in two insect species: *Apis mellifera* and *Bombus terrestris, Nucleic Acids Res.*, 1993, vol. 21, no. 6, pp. 1427—1431.
21. Estoup, A., Garnery, L., Solignac, M., and Cornuet, J.M., Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models, *Genetics*, 1995, vol. 140, no. 2, pp. 679—695.
22. Schmitzova, J., Klaudiny, J., Albert, S., et al., A family of major royal jelly proteins of the honeybee *Apis mellifera* L., *Cell. Mol. Life Sci.*, 1998, vol. 54, no. 9, pp. 1020—1030.
https://doi.org/10.1007/s000180050229
23. Wilanowski, T.M. and Gibson, J.B., *sn-Glycerol-3-phosphate dehydrogenase* in the honey bee *Apis mellifera*—an unusual phenotype associated with the loss of introns1GenBank accession No.: AF023666.12 Nucleotide symbol combinations: H = A/C/T; N = any nucleotide; R = A/G; Y = C/T.2, *Gene*, 1998, vol. 209, no. 1, pp. 71—76.
https://doi.org/10.1016/S0378-1119(98)00016-X
24. Walldorf, U., Binner, P., and Fleig, R., Hox genes in the honey bee *Apis mellifera, Dev. Genes EV*, 2000, vol. 210, no. 10, pp. 483—492.
https://doi.org/10.1007/s004270050337
25. Chen, C., Liu, Z., Pan, Q., et al., Genomic analyses reveal demographic history and temperate adaptation of the newly discovered honey bee subspecies *Apis mellifera sinisxinyuan* n. ssp., *Mol. Biol. EV*, 2016, vol. 33, no. 5, pp. 1337—1348.
https://doi.org/10.1093/molbev/msw017
26. Wragg, D., Techer, M.A., Canale-Tabet, K., et al., Autosomal and mitochondrial adaptation following Admixture: a case study on the honeybees of reunion island, *Genome Biol. EV*, 2018, vol. 10, no. 1, pp. 220—238.
https://doi.org/10.1093/gbe/evx247
27. Haddad, N.J., Loucif-Ayad, W., Adjlane, N., et al., Draft genome sequence of the Algerian bee *Apis mellifera intermissa, Genome Data*, 2015, vol. 4, pp. 24—25.
https://doi.org/10.1016/j.gdata.2015.01.011
28. Elsik, C.G., Tayal, A., Diesh, C.M., et al., Hymenoptera Genome Database: integrating genome annotations in HymenopteraMine, *Nucleic Acids Res.*, 2016, vol. 44, no. D1, pp. D793—D800.
https://doi.org/10.1093/nar/gkv1208
29. Wallberg, A., Schöning, C., Webster, M.T., and Hasselmann, M., Two extended haplotype blocks are associated with adaptation to high altitude habitats in East African honey bees, *PLoS Genet.*, 2017, vol. 13, no. 5. e1006792.
https://doi.org/10.1371/journal.pgen.1006792
30. Spötter, A., Gupta, P., Mayer, M., et al., Genome-Wide association study of a varroa-specific defense behavior in honeybees (*Apis mellifera*), *J. Hered.*, 2016, vol. 107, no. 3, pp. 220—227.
https://doi.org/10.1093/jhered/esw005
31. Henriques, D., Parejo, M., Vignal, A., et al., Developing reduced SNP assays from whole-genome sequence data to estimate introgression in an organism with complex genetic patterns, the Iberian honeybee (*Apis mellifera iberiensis*), *Evol. Appl.*, 2018, no. 11, pp. 1270—1282.
https://doi.org/10.1111/eva.12623
32. Han, F., Wallberg, A., and Webster, M.T., From where did the Western honeybee (*Apis mellifera*) originate?, *Ecol. Evol.*, 2012, vol. 2, no. 8, pp. 1949—1957.
https://doi.org/10.1002/ece3.312
33. Han, F., Wallberg, A., and Webster, M.T., Population genomics of the honey bee reveals strong signatures of positive selection on worker traits, *Proc. Natl. Acad. Sci. U.S.A.*, 2014, vol. 111, no. 7, pp. 2614—2619.
https://doi.org/10.1073/pnas.1315506111
34. Kapheim, K.M., Pan, H., Li, C., et al., Genomic signatures of evolutionary transitions from solitary to group living, *Science*, 2015, vol. 348, no. 6239,

pp. 1139—1143.
https://doi.org/10.1126/science.aaa4788

35. Kent, C.F., Minaei, S., Harpur, B.A., and Zayed, A., Recombination is associated with the evolution of genome structure and worker behavior in honey bees, *Proc. Natl. Acad. Sci. U.S.A.*, 2012, vol. 109, no. 44, pp. 18012—18017.
https://doi.org/10.1073/pnas.1208094109

36. Pinto, M.A., Henriques, D., Chávez-Galarza, J., et al., Genetic integrity of the Dark European honey bee (*Apis mellifera mellifera*) from protected populations: a genome-wide assessment using SNPs and mtDNA sequence data, *J. Apic. Res.*, 2014, vol. 53, no. 2, pp. 269—278.
https://doi.org/10.3896/IBRA.1.53.2.08

37. Jensen, A.B., Palmer, K.A., Boomsma, J.J., and Pedersen, B.V., Varying degrees of *Apis mellifera ligustica* introgression in protected populations of the black honeybee, *Apis mellifera mellifera*, in northwest Europe, *Mol. Ecol.*, 2005, vol. 14, pp. 93—106.

38. Muñoz, I., Henriques, D., Johnston, J.S., et al., Reduced SNP panels for genetic identification and introgression analysis in the dark honey bee (*Apis mellifera mellifera*), *PLoS One*, 2015, vol. 10, no. 4. e0124365.
https://doi.org/10.1371/journal.pone.0124365

39. Parejo, M., Wragg, D., Gauthier, L., et al., Using whole-genome sequence information to foster conservation efforts for the European dark honey bee, *Apis mellifera mellifera, Front. Ecol. Evol.*, 2016, vol. 4, p. 583.
https://doi.org/10.3389/fevo.2016.00140

40. Chávez-Galarza, J., Henriques, D., Johnston, J.S., et al., Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms, *Mol. Ecol.*, 2013, vol. 22, no. 23, pp. 5890—5907.
https://doi.org/10.1111/mec.12537

41. Chapman, N.C., Harpur, B.A., Lim, J., et al., A SNP test to identify Africanized honeybees via proportion of "African" ancestry, *Mol. Ecol. Resour.*, 2015, vol. 15, no. 6, pp. 1346—1355.
https://doi.org/10.1111/1755-0998.12411

42. Kadri, S.M., Harpur, B.A., Orsi, R.O., and Zayed, A., A variant reference data set for the Africanized honeybee, *Apis mellifera, Sci. Data*, 2016, vol. 3, p. 160097.
https://doi.org/10.1038/sdata.2016.97

43. Chapman, N.C., Harpur, B.A., Lim, J., et al., Hybrid origins of Australian honeybees (*Apis mellifera*), *Apidologie*, 2016, vol. 47, no. 1, pp. 26—34.
https://doi.org/10.1007/s13592-015-0371-0

44. Harpur, B.A., Chapman, N.C., Krimus, L., et al., Assessing patterns of admixture and ancestry in Canadian honey bees, *Insectes Soc.*, 2015, vol. 62, no. 4, pp. 479—489.
https://doi.org/10.1007/s00040-015-0427-1

45. Falush, D., Stephens, M., and Pritchard, J.K., Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies, *Genetics*, 2003, vol. 164, no. 4, pp. 1567—1587.

46. Alexander, D.H., Novembre, J., and Lange, K., Fast model-based estimation of ancestry in unrelated individuals, *Genome Res.*, 2009, vol. 19, no. 9, pp. 1655—1664.
https://doi.org/10.1101/gr.094052.109

47. Weir, B.S. and Cockerham, C.C., Estimating *F*-statistics for the analysis of population structure, *Evolution*, 1984, vol. 38, no. 6, pp. 1358—1370.
https://doi.org/10.2307/2408641

48. Graham, A.M., Munday, M.D., Kaftanoglu, O., et al., Support for the reproductive ground plan hypothesis of social evolution and major QTL for ovary traits of Africanized worker honey bees (*Apis mellifera* L.), *BMC Evol. Biol.*, 2011, vol. 11, p. 95.
https://doi.org/10.1186/1471-2148-11-95

49. Rueppell, O., Metheny, J.D., Linksvayer, T., et al., Genetic architecture of ovary size and asymmetry in European honeybee workers, *Heredity*, 2011, vol. 106, no. 5, pp. 894—903.
https://doi.org/10.1038/hdy.2010.138

50. Holloway, B., Sylvester, H.A., Bourgeois, L., and Rinderer, T.E., Association of single nucleotide polymorphisms to resistance to chalkbrood in *Apis mellifera, J. Apic. Res.*, 2012, vol. 51, no. 2, pp. 154—163.
https://doi.org/10.3896/IBRA.1.51.2.02

51. Tsuruda, J.M., Harris, J.W., Bourgeois, L., et al., High-resolution linkage analyses to identify genes that influence Varroa sensitive hygiene behavior in honey bees, *PLoS One*, 2012, vol. 7, no. 11. e48276.
https://doi.org/10.1371/journal.pone.0048276

52. Hunt, G.J., Page, R.E., Jr., Fondrk, M.K., and Dullum, C.J., Major quantitative trait loci affecting honey bee foraging behavior, *Genetics*, 1995, vol. 141, no. 4, pp. 1537—1545.

53. Rüppell, O., Pankiw, T., and Page, R.E., Jr., Pleiotropy, epistasis and new QTL: the genetic architecture of honey bee foraging behavior, *J. Hered.*, 2004, vol. 95, no. 6, pp. 481—491.
https://doi.org/10.1093/jhered/esh072

54. Lapidge, K.L., Oldroyd, B.P., and Spivak, M., Seven suggestive quantitative trait loci influence hygienic behavior of honey bees, *Naturwissenschaften*, 2002, vol. 89, no. 12, pp. 565—568.
https://doi.org/10.1007/s00114-002-0371-6

55. Shorter, J.R., Arechavaleta-Velasco, M., Robles-Rios, C., and Hunt, G.J., A genetic analysis of the stinging and guarding behaviors of the honey bee, *Behav. Genet.*, 2012, vol. 42, no. 4, pp. 663—674.
https://doi.org/10.1007/s10519-012-9530-5

56. Spötter, A., Gupta, P., Nürnberg, G., et al., Development of a 44K SNP assay focusing on the analysis of a varroa-specific defense behaviour in honey bees (*Apis mellifera carnica*), *Mol. Ecol. Resour.*, 2012, vol. 12, no. 2, pp. 323—332.
https://doi.org/10.1111/j.1755-0998.2011.03106.x

57. Liu, Y., Yan, L., Li, Z., et al., Larva-mediated chalkbrood resistance-associated single nucleotide polymor-

phism markers in the honey bee *Apis mellifera, Insect Mol. Biol.*, 2016, vol. 25, no. 3, pp. 239—250. https://doi.org/10.1111/imb.12216

58. Southey, B.R., Zhu, P., Carr-Markell, M.K., et al., Characterization of genomic variants associated with scout and recruit behavioral castes in honey bees using whole-genome sequencing, *PLoS One*, 2016, vol. 11, no. 1. e0146430. https://doi.org/10.1371/journal.pone.0146430
59. Wallberg, A., Pirk, C.W., Allsopp, M.H., and Webster, M.T., Identification of multiple loci associated with social parasitism in Honeybees, *PLoS Genet.*, 2016, vol. 12, no. 6. e1006097. https://doi.org/10.1371/journal.pgen.1006097
60. Lattorff, H.M.G., Moritz, R.F.A., Crewe, R.M., and Solignac, M., Control of reproductive dominance by the *thelytoky* gene in honeybees, *Biol. Lett.*, 2007, vol. 3, no. 3, pp. 292—295. https://doi.org/10.1098/rsbl.2007.0083
61. Fuller, Z.L., Niño, E.L., Patch, H.M., et al., Genome-wide analysis of signatures of selection in populations of African honey bees (*Apis mellifera*) using new web-based tools, *BMC Genomics*, 2015, vol. 16, p. 518. https://doi.org/10.1186/s12864-015-1712-0
62. Henriques, D., Wallberg, A., Chávez-Galarza, J., et al., Whole genome SNP-associated signatures of local adaptation in honeybees of the Iberian Peninsula, *Sci. Rep.*, 2018, vol. 8, no. 1, p. 11145. https://doi.org/10.1038/s41598-018-29469-5
63. Luu, K., Bazin, E., and Blum, M.G.B., PCAdapt: an R package to perform genome scans for selection based on principal component analysis, *Mol. Ecol. Resour.*, 2017, vol. 17, no. 1, pp. 67—77. https://doi.org/10.1111/1755-0998.12592
64. Stucki, S., Orozco-terWengel, P., Forester, B.R., et al., High performance computation of landscape genomic models including local indicators of spatial association, *Mol. Ecol. Resour.*, 2017, vol. 17, no. 5, pp. 1072—1089. https://doi.org/10.1111/1755-0998.12629
65. Frichot, E., Schoville, S.D., Bouchard, G., and François, O., Testing for associations between loci and environmental gradients using latent factor mixed models, *Mol. Biol. Evol.*, 2013, vol. 30, no. 7, pp. 1687—1699. https://doi.org/10.1093/molbev/mst063

Translated by N. Maleeva