

RESEARCH ARTICLE

The genetic basis of mimicry in the snowy bumble bee (*Bombus niveatus*) in Anatolia with insights from a color polymorphic gynandromorph

Tunç Dabak¹, Çiğdem Özenirler², Ece Kamalak², Cecil Smith¹, Seçil Aytekin², Ahmet Murat Aytekin², Heather M. Hines^{1*}

1 Department of Biology, The Pennsylvania University, University Park, Pennsylvania, United States of America, **2** Department of Biology, Hacettepe University, Beytepe, Ankara, Türkiye

* hmh19@psu.edu



OPEN ACCESS

Citation: Dabak T, Özenirler Ç, Kamalak E, Smith C, Aytekin S, Aytekin AM, et al. (2026) The genetic basis of mimicry in the snowy bumble bee (*Bombus niveatus*) in Anatolia with insights from a color polymorphic gynandromorph. *PLoS Genet* 22(3): e1012060. <https://doi.org/10.1371/journal.pgen.1012060>

Editor: Artyom Kopp, University of California Davis, UNITED STATES OF AMERICA

Received: November 14, 2025

Accepted: February 16, 2026

Published: March 11, 2026

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pgen.1012060>

Copyright: © 2026 Dabak et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution,

Abstract

Bumble bees (*Bombus spp.*) display remarkable color pattern diversity and convergence driven largely by Müllerian mimicry. In Anatolia, bumble bees mimic each other by converting ancestral yellow anterior setal body color to white in multiple independent lineages. Here, we investigate the genetic basis of white–yellow mimetic color dimorphism in the snowy bumble bee *Bombus niveatus*, separated into two subspecies based on coloration: the white *Bombus niveatus niveatus* and the yellow *Bombus niveatus vorticosus*. Using a genome-wide association study (GWAS) of males sampled across dimorphic populations, we identify a strong association peak linked to white–yellow variation in the cis-regulatory region of the homeobox gene *BarH*, a gene previously implicated in driving spatial patterning of epidermal projections and pigmentation. This locus, coined the *snowy* locus, involves a derived tandem duplication unique to the white form that likely increases the number of transcription factor binding sites. Comparative sequencing of *snowy* indicates co-mimicking species use different variants for their white-yellow convergent transitions. Additionally, we describe and genetically analyze a largely bilateral mosaic gynandromorph of *B. niveatus* with a mix of both color forms across its body. This was determined to be generated by a mosaic of at least two separate haploid sources with different *snowy* alleles, and diploid tissue heterozygous for the color locus. This supports the genetic basis for the color polymorphism and reinforces the conspecific status of the two forms. Together, these findings expand our understanding of the genetic basis of mimetic color pattern convergence in this phenotypic radiation.

and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All raw genomic data generated for *B. niveatus* samples and tissue-specific genomic sequences have been deposited in the NCBI Sequence Read Archive under BioProject accession PRJNA1358947. Scripts for read processing, alignments, and downstream data analysis and visualizations are available in Figshare (<https://doi.org/10.6084/m9.figshare.30601733>).

Funding: Research and graduate stipend support (TD) were provided by the US National Science Foundation DEB Grant #2126417 (to HMM). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

This study examines the genetic basis of mimetic convergence in Anatolian bumble bees, which repeatedly acquired a distinct white color pattern from ancestral yellow color forms. We sequenced genomes of males of both color forms in the snowy bumble bee, *Bombus niveatus*, and performed a genotype-phenotype association analysis to identify the genomic variant driving this color variation to a regulatory mutation in the developmental gene *BarH*. This gene is implicated in pterin pigment and sensory bristle patterning in other insects, two functions it also serves in these bees. In this case the white form evolved a novel duplication in the regulatory region that likely modifies transcriptional activity to prevent yellow pigment deposition. This variant does not explain parallel evolution of the same white color from yellow patterns in several other comimetic species from this region, revealing convergent mechanisms drive these mimicry patterns. Analysis of a rare gynandromorph showing both color types provides further support for this genetic mechanism and for the conspecificity of *B. niveatus* color forms.

Introduction

Bumble bees (*Bombus spp.* Latreille, 1802) are known for their ecological importance for pollination across much of the globe [1–3] but also are recognized for their striking diversity of setal color patterns [4]. Bumble bees exhibit ~427 color patterns across the ~265 species, comprised predominantly of different segmental combinations of yellow, black, red, and white setal pile. This exceptional diversity, while likely influenced by thermoregulation, is thought to be generated largely by Müllerian mimicry [5], where co-occurring noxious species converge on shared warning signals to deter predators. Convergence of these bees onto over 24 local mimicry patterns globally has resulted in numerous intraspecific polymorphisms as species converge onto local complexes while promoting cross-species color diversity [5].

Such mimicry systems offer an excellent opportunity to explore the genetic basis of phenotypic convergence and evolutionary adaptation [6–8]. Genetic studies have begun to unravel the mechanisms underlying these patterns, revealing genomic regions and pathways that are responsible for the adaptive mimetic shifts in their color patterns [8–10]. For example, in the North American bumble bee mimicry system, species converge onto red or black mid-abdominal coloration in Rocky Mountain and Pacific Coastal regions, respectively. The pigments that are responsible for red and black coloration in these bees were found to be pheomelanin and melanin [11]. By studying red and black polymorphic species crossing these mimicry zones, it was revealed that *Abdominal-B (Abd-B)* – the Hox gene driving morphology of the most posterior abdominal segments – was independently targeted across species to upregulate red by expressing in an atypically anterior location [10]. For each set of mimics, this upstream developmental Hox gene triggers diverse genes in the melanin

pathway including primarily *ebony*, but also *yellow*, *pale*, and a suite of other melanin genes, that control the eumelanin/pheomelanin ratio to determine the setal color. The *Abd-B* regulatory region was also implicated in red vs. black posterior tail phenotypes in distantly related *B. breviceps* in South and South-East China [12], suggesting *Abd-B* may ancestrally regulate red color in its typical posterior domain and has shifted its expression spatially along the body to alter colors of other segments [10]. These studies show that a Hox gene responsible for segmental fate is a genomic hotspot for color pattern evolution in bumble bees and that a conserved pathway connecting *Abd-B* to pigmentation genes may be operating to generate this diversity.

The genetics of white and yellow color in bumble bees and the involved pigments are still poorly resolved. White coloration is considered to be a depigmentation, as it fails to yield any pigment signatures upon extraction [13] and yellow coloration is thought to involve some pheomelanin production [14] but they also contain an unidentified pigment with properties akin to pterins [13,15]. A mid-body yellow variant generated *de novo* from black color in a lab colony of *Bombus terrestris* (L. 1758) was generated by a protein-coding mutation in the major developmental transcription factor *cut*, although this showed pleiotropic effects given the protein is targeted [16]. This gene, however, has not yet been implicated in natural variations of the same type. Understanding the genetics of yellow is especially important considering that black and yellow are the most common colors in bumble bees [4]. Revealing genes driving yellow coloration will add additional genes to the toolkit generating color diversity in these bees.

To better understand the genomic basis of yellow color patterns and mimicry in general in bumble bees, we study the genetic basis of yellow and white color switches in the Anatolian mimicry complex. In Anatolia, ~15 species converge on the same white mimetic pattern involving a white thorax with a black band and white first and second metasomal segments, followed by a black band and red tail (Fig 1). Many of these species obtain this pattern by shifting to white from an ancestral yellow color pattern (Fig 1). This yellow pattern is the most frequent color form in bumble bees outside this region [5]. The result is multiple repeated instances of white and yellow color switches, fruitful for deciphering the genetic basis of these colors and of mimetic convergence.

To understand the genetics of this system, we examine a white and yellow dimorphic species from Central Anatolia - *Bombus niveatus*. Historically *B. niveatus* was considered to be two distinct species, *B. niveatus* and *B. vorticosus*, that differ in their setal coloration: white in the former and yellow in the latter (Figs 1 and S1 Fig). This lineage was later recognized as conspecific based on lack of distinction using mandibular gland secretions [18], geometric morphometrics [19], and genetic evidence from a few nuclear and mitochondrial genes [20], thus these color differences represent an intraspecific dimorphism. These color variants likely evolved in response to mimetic selection for convergence onto the white regional color pattern: the white-banded (*ssp. niveatus*; *B. n. niveatus*) form is primarily found in eastern Anatolia, while the yellowish-banded (*ssp. vorticosus*; *B. n. vorticosus*) form extends to the Balkan Peninsula, with both subspecies co-occurring in central Anatolia [21] across altitudes ranging from 800 to 3000 meters [18] (Fig 2A).

Our study identifies the genetic basis of the yellow-white color switch in *B. niveatus* utilizing a genome wide association study of bees from dimorphic populations from Türkiye. We seek to determine whether this divergence arises from modifications to upstream developmental genes, as in previous studies, or involves certain downstream pigment pathways. Additionally, we investigate the allelic relationships of this dimorphic coloration by using a bilateral gynandromorph specimen of *B. niveatus*, which harbors both color forms with respective coinciding sex-specific characters. Phenotypically, female-specific characters coincide with the white hairs while the yellow portions encapsulate male-specific characters. This presents a unique opportunity to explore both the genetic basis of coloration and uncover the origins of this gynandromorph.

Results

Mimicry in anatolia

A distribution map of white forms of bumble bees in the Anatolian region, circumscribed based on Atlas Hymenoptera (atlashymenoptera.net), iNaturalist observations, and personal collections, shows the heart of the distribution of the white

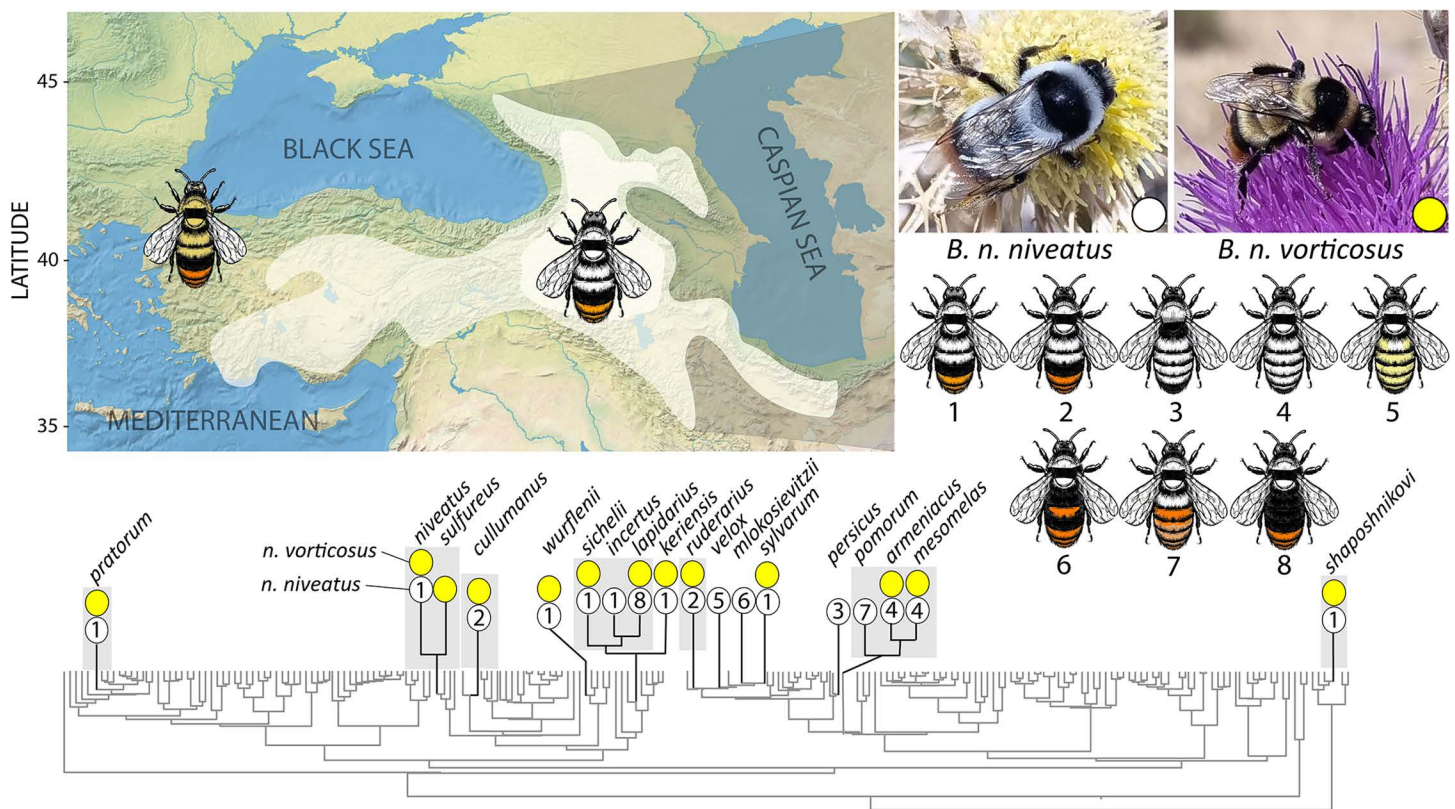


Fig 1. The distribution and convergent evolution of white forms of the Anatolian mimicry complex. Bees in this mimicry complex have evolved white forms from typically yellow ancestral patterns. Photos are shown of bees with white and yellow color forms featuring the two color variants of the species studied here, *B. n. niveatus* and *B. n. vorticusus* (Photo Credit: Çiğdem Özenirler). The more transparent white is the general range while the more opaque white represents the heart of the mimicry complex with the highest fidelity. Yellow patterns occur outside this zone. At right are color pattern diagrams representing the variants of the sympatric white mimicry form, which are assigned to species in the phylogenetic tree using numeric coding. Yellow forms of these same species typically have the same pattern but with yellow instead of white. The phylogeny of all bumble bees is at bottom, showing their high frequency of repeated yellow-white polymorphisms indicated by yellow and white circles. Comparison groups sequenced in this study are highlighted by gray boxes. The phylogeny is adapted from [17]. Map layers were made with Natural Earth (@naturalearthdata.com) <https://www.naturalearthdata.com/downloads/10m-raster-data/10m-cross-blend-hypsos/>.

<https://doi.org/10.1371/journal.pgen.1012060.g001>

form is in the eastern Anatolian region (Fig 1). Yellow forms increase outside this region, as exemplified by the distribution of the color forms of *Bombus niveatus* (Fig 2A). Phylogenetic mapping of the white mimetic species and their sister taxa onto a time-calibrated phylogeny (Hines, 2008b) shows how white color forms in Anatolia are largely phylogenetically convergent (Fig 1), as they are scattered across multiple different lineages. They frequently are acquired within species from ancestral yellow color forms, leading to multiple species (n~15) with intraspecific yellow-white polymorphisms.

Population clustering, GWAS and linkage analysis

We performed whole genome sequencing of 18 white and 18 yellow males of *B. niveatus* from 5 collection sites within the Central Anatolian Region transition zone (Fig 2A). Males are used as they are haploid, thus eliminating complications from dominance effects and heterozygosity. Each sampling site contained both color forms except the most northwestern site which only contained the yellow form (VOR4). PCA plots of these samples revealed some site-specific separation among samples, but yellow and white forms are otherwise genomically admixed, supporting no evidence for population or species-level distinction between color forms (Fig 2B).

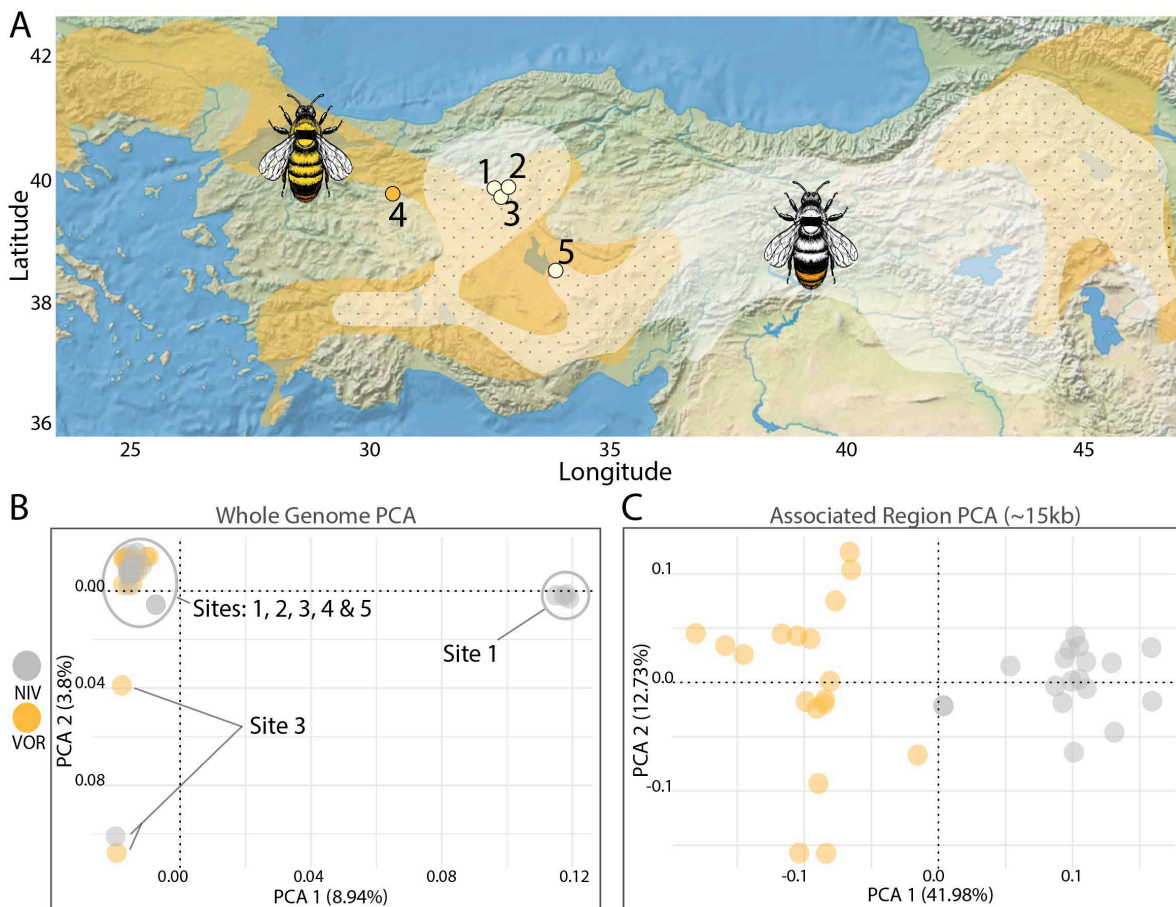


Fig 2. *B. niveatus* specimens sampled for genomic sequencing relative to the greater distributions of the subspecies (A), and PCA of samples inferred from whole genome (B) and color locus (C) sequences. Broader ranges of each color form are demarcated with white, orange-yellow, and dotted (overlapping transition zones) regions, and the five sampling locations for specimens are shown. The PCA plot shows the first two PC axes of each genomic sample with dots colored by species. The color locus includes the 15kb region around the color associated locus. Map layers were made with Natural Earth (@naturalearthdata.com) <https://www.naturalearthdata.com/downloads/10m-raster-data/10m-cross-blend-hypso/>.

<https://doi.org/10.1371/journal.pgen.1012060.g002>

Genome-wide association analysis of these samples by color, aligned against the *B. n. niveatus* reference genome, identified a single 232-bp peak that contains five SNPs fixed by phenotype: the *B. n. niveatus* haplotype (A, A, T, T, G at SNPs 1–5) and the *B. n. vorticosus* haplotype (G, G, A, C, A at SNPs 1–5). These are located in the cis-regulatory region approximately 7.5 kb upstream of the homeobox gene *BarH* (Fig 3A and 3B). In Diptera, *BarH* exists as two paralogs (*BarH1* and *BarH2*) due to a lineage-specific duplication; however, our Blast searches revealed that Hymenoptera and all other lineages other than Brachycera flies (*Anastrepha* and *Drosophila* sampled here) have only a single copy. Phylogenetic analysis using both protein and CDS sequences confirms that the paralog of *BarH* evolved in the flies after the phylogenetic split from *Anopheles* (Fig 3F).

Manual inspection in IGV revealed that, in addition to these SNPs, there was a drop in read coverage in this region in yellow *B. n. vorticosus* but not in white *B. n. niveatus*, suggesting either a large deletion or substantial base variation could be involved. In this region there is a 409 bp tandem duplication that is present in the *B. n. niveatus* reference genome (white form) (Fig 3C). Read coverage loss occurs in one of these copies, suggesting the *B. n. vorticosus* (yellow form) lacks this tandem duplication. This coverage loss in the duplicated region is consistent across *B. n. vorticosus* samples,

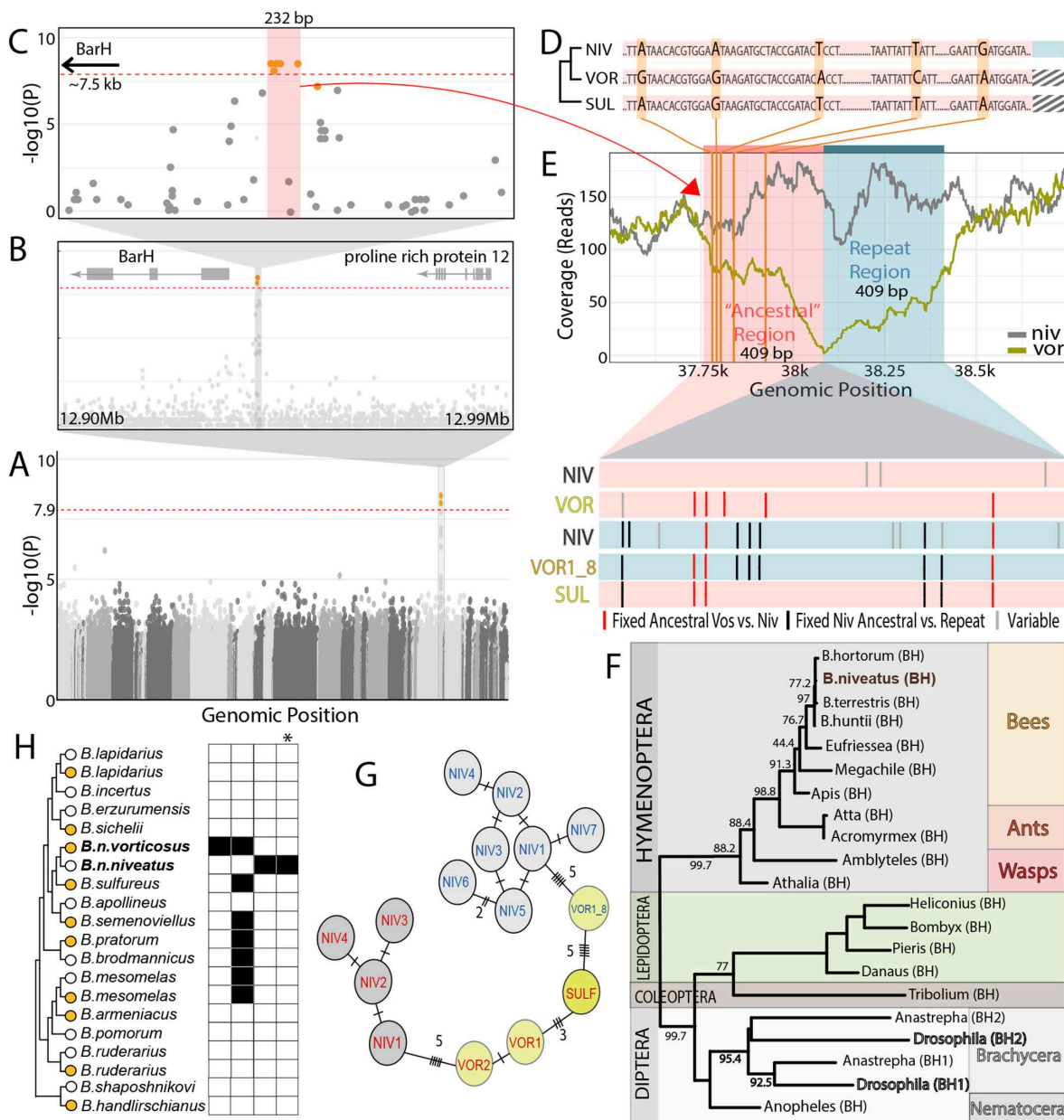


Fig 3. Genetic association and genomic variation among yellow and white *B. niveatus* morphs. (A) Manhattan plot from GWAS across the genome with the peak region highlighted and the genome-wide significance threshold indicated with a horizontal red dashed line. The orange dots represent variants fixed by phenotype; (B) Contig-level association showing nearby annotated genes; (C) Close-up of the most associated SNPs; (D) Sequences showing the variants of the five fixed sites from the GWAS between *B. n. niveatus* (NIV), *B. n. vorticosus* (VOR), and *B. sulfureus* (SUL). The orange connectors indicate exact positions of the variants relative to the ancestral/repeat region in panel (E); (E) (Top) The local total read-depth profile change across the associated region by color form, indicating the repeated genomic region for the white form (*B. n. niveatus*) in blue and the ancestral region in pink. (Bottom) Alignment with indicated variants relative to the ancestral *B. n. niveatus* reference including comparisons between the ancestral and repeat regions of the consensus sequence of *B. n. niveatus* (NIV), the consensus sequence in this region of *B. n. vorticosus* (VOR), the singular *B. n. vorticosus* sequence VOR1_8, and the sequence of yellow sister lineage *B. sulfureus* (SUL). Variants that are fixed within a color form are in black, with fixed SNPs from the GWAS in red, while variants that are not fixed within a color form are in gray; (F) Phylogenetic analysis for CDS sequences of *BarH* of different insect taxa showing the orthology of the *BarH* gene and its paralogs across the phylogeny; (G) Haplotype network of Ancestral and Repeat region sequences, with each region treated separately. Gray and yellow colored nodes indicate the white and the yellow phenotypes, respectively. Red and blue colored labels indicate the Ancestral and the Repeat region sequences, respectively; (H) Evolutionary variants for 3 of the 5 fixed GWAS SNPs that remain fixed after considering VOR1_8 variants, as well as the duplication (shown with an asterisk), showing the alleles of these variants across the white/yellow phylogenetic mimicry pairs. Ancestral variants are indicated with white squares and derived variants with black squares.

<https://doi.org/10.1371/journal.pgen.1012060.g003>

although one *B. n. vorticosus* (VOR1_8) has coverage loss in the other copy (S2A Fig; see below for explanation). To investigate patterns in this interval further, we generated a *de novo* genome assembly for *B. n. vorticosus*. As expected, it assembled just one copy of the tandem duplication. This assembled fragments that spanned the full linkage region (described below) 10kb upstream and 2kb downstream of the duplicated region. In this 12kb region, we identified only one 101 bp indel with the *B. n. vorticosus* assembled sequence that was not previously detected using the alignments against the *B. n. niveatus* genome, but IGV inspection revealed this was not associated with phenotype. There were no new SNP variants identified, thus our original analysis found all relevant fixed variants in this region. Hereafter we refer to the region including the original region with SNP variants plus the tandem duplication as the *snowy* locus, in reference to its control of white phenotypes in the Snowy Bumble Bee.

A linkage-disequilibrium analysis shows that among the ~14.84 Mbp contig containing this peak, there is very little LD aside from a single large LD block ($r^2 > 0.75$) confined to an ~1.6 kb block spanning the peak region (S3 Fig). Only the *BarH* cis-regulatory region falls in this linkage block. There is a disjunct block of sites with linkage to this region that fall in a *BarH* exon, suggesting some degree of linkage between the cis-regulatory domain and the protein-coding sequence. This exon includes 15 linked SNPs 8.2kb away from *snowy*, with the strongest association involving allele fixation for the white form and 69% of samples with the alternative allele in the yellow form. This 5' UTR variant (12929476) and the *snowy* region (12938203–12938236) are highly linked with an r^2 greater than 0.70. A PCA plot of the 15kb region around the color associated locus sorts samples by color (Fig 2C).

Cross-taxon sequence validation of the locus

We performed PCRs and Sanger sequenced our genomic samples for an ~1kb fragment spanning the *snowy* region to confirm and better characterize their sequence. This supported the bioinformatic results, revealing all white-form *B. n. niveatus* individuals contain a duplicated 409 bp sequence absent in all sequenced yellow form individuals (Fig 4). The 5 fixed SNP differences identified by bioinformatics are present in one of the two duplicated copies in *B. n. niveatus* where it best aligns with the same copy region in *B. n. vorticosus* (“ancestral region”, Fig 3D). The other duplicated copy (“repeat region”), the one unique to *B. n. niveatus*, has 9 bps different from both of the ancestral *B. n. niveatus* and *B. n. vorticosus* sequences (Fig 3E).

Sanger sequencing revealed that one of the yellow form bees (VOR1_8) has only one copy of the region, similar to the other *B. n. vorticosus*, but instead of having the typical *B. n. vorticosus* “ancestral” sequence, it has a sequence that is more similar to the “repeat region” (Fig 3E), of *B. n. niveatus* (Fig 3G). This variant only shares 3 of the 5 fixed SNPs with the other *B. n. vorticosus* and the coverage plot of reads in the region revealed that it, unlike the other *B. n. vorticosus*, aligned with the repeat region (S2A Fig). It thus did not contribute to the SNP calls in the GWAS of the ancestral sequence and ultimately supports just 3 SNPs plus the deletion in *snowy* being fixed by phenotype.

We also sequenced the closest sister lineage *B. sulfureus* - a monomorphic yellow colored species (most recent common ancestry with *B. niveatus* of ~1.5 MYA, (Hines, 2008b)). Like yellow *B. n. vorticosus*, *B. sulfureus* is missing the divergent “repeat” region. A haplotype network and the SNP comparison between the ancestral region and repeat regions across samples (Fig 3G), shows *B. sulfureus* shares more bases with the yellow *B. n. vorticosus*, including close relationships with both VOR1_8 and the remaining *B. n. vorticosus*. An unusual pattern was supported where white *B. n. niveatus* copies appear to be derived independently from each *B. n. vorticosus* branch, with the typical *B. n. vorticosus* leading to the ancestral *B. n. niveatus* sequence, and the VOR1_8 leading to the repeat region (Fig 3H). These data raise the possibility that the white form may have evolved by merging two divergent yellow form haplotypes, such as through a tandem duplication generated by alignment errors during meiosis.

To assess whether the variants in *snowy* are shared across other closely related *Bombus* lineages with white/yellow variation, we Sanger-sequenced the same cis-regulatory region across the co-mimicking lineages (7 yellow and 8 white

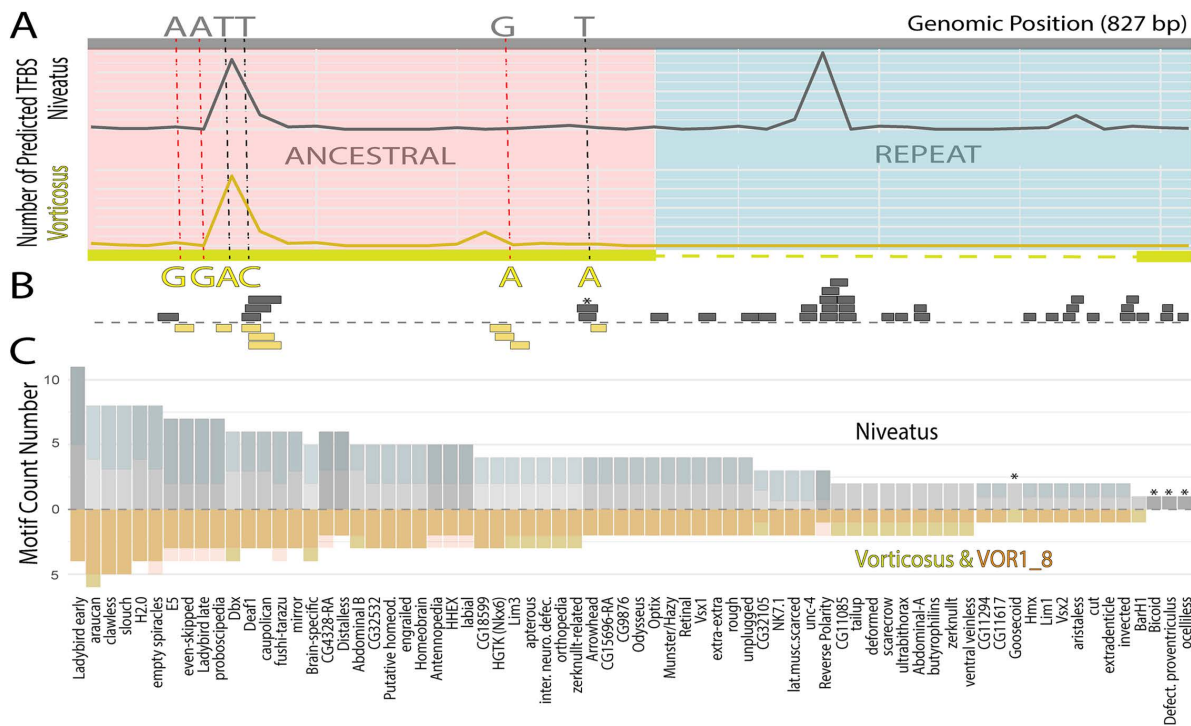


Fig 4. Predicted transcription factor binding sites across the associated genomic interval. (A) The frequency of binding sites across the ancestral and repeat regions for each *B. niveatus* color form by genomic position. GWAS fixed sites are shown, with red dashed lines representing the three fixed sites after considering VOR1-8. Ancestral (red area) and repeated (blue area) genomic regions are indicated.; (B) TF binding motifs that are different between *B. n. vorticosis* and *B. n. niveatus* are shown by genomic position, aligned with part A; (C) TF-specific motif counts for *B. n. vorticosis* and *B. n. niveatus* are shown. On the top half (*B. n. niveatus* side) of the graphic, translucent blue highlighted regions are sites gained by the duplication. On the bottom half, translucent red highlighted bars indicate the amount of TF in the VOR1_8 haplotype, while the main *B. n. vorticosis* haplotype is represented by solid yellow. Asterisks (*) indicate the novel TF bindings sites due to non-fixed SNPs.

<https://doi.org/10.1371/journal.pgen.1012060.g004>

form species) in Eastern Anatolia including species similarly polymorphic for white/yellow forms (species highlighted in gray in Fig 1C). The duplication observed in *B. n. niveatus* was not detected in any other species regardless of color. No other large indels in the fragment differed between other white or yellow pairs. There was also no clear pattern of shared variants in these species to the 3 fixed SNPs differing in *B. n. niveatus* and *B. n. vorticosis*, implying that the causal mutation(s) for mimicry in those lineages differ (Fig 3H). Only one lineage (*B. cullumanus* white form (= *B. apollineus*) and yellow form sister species *B. semenoviellus*) had a shared variant (second fixed SNP [A -> G] substitution) that also differed in *B. n. niveatus* compared to *B. n. vorticosis* and *B. sulfureus*, with variants matching the same respective colors.

Allelic dominance of the color locus

Phenotypic assessment across subspecies indicates a predominantly discrete white–yellow trait, with some variation in color intensity among individuals (S1 Fig). As bees used for this study were destructively sampled for other work, we were unable to compare allelic correlation to yellow intensity *post hoc*. PCR (gel band) and Sanger sequencing assays were performed on two available *B. n. niveatus* workers and one *B. n. vorticosis* worker. The yellow worker was homozygous for the short amplicon (*a/a* - no duplication), whereas both white workers were heterozygous for the duplication (*A/a*) (S4 Fig). These data suggest that the white-allele is dominant (*A*) over the yellow allele (*a*).

Transcription factor binding variation

TFBS prediction was performed on the three Sanger-sequenced haplotypes corresponding to the white form (Haplotype 1; white; *B. n. niveatus*), yellow form (Haplotype 2; yellow; *B. n. vorticosus*) and the yellow form with deviating haplotype (Haplotype 3; yellow; VOR1_8). The repeat region along with the associated ancestral region in the white haplotype has an increased number of transcription factors (296 sites) compared to the yellow haplotype without the duplication (167 and 157 sites for Haplotype 2 and 3, respectively) (Fig 4). The white-form specific repeat region alone contributes 139 TFBS while retaining the same repertoire of 70 motif families found in the shorter yellow haplotypes.

Only three JASPAR motifs are unique to the white haplotype (MA0212.1 “*bcd*”, MA0234.1 “*oc*”, and MA0915.1 “*dve*”), however, these are not introduced by any of the fixed SNPs, but rather a SNP present in most white individuals. Two of the three fixed SNPs (GWAS fixed SNP positions 1 and 5) were predicted to alter TF binding-site configuration between *B. n. niveatus* (Haplotype 1) and *B. n. vorticosus* (Haplotype 2) (S2 Fig). Although these could implicate TF loss in the white form, in all cases the repeat region adds these TFs back so that the white form has a larger number of copies than the yellow form. None of these fixed SNP sites were found to be consistently differentiating TF binding by color pattern in other yellow-white mimetic lineages, as the only shared SNP (SNP2) does not have any distinct TF binding sites (S2B and S2C Fig).

Examination of motif abundance reveals that TFBS for a set of developmental/ homeodomain factors are substantially increased in the white haplotype, with no binding sites being greater in number in the yellow form. The top factors showing the largest relative increases in the white form include: Ladybird early (*lbe*), Distal-less (*Dll*), **CG4328-RA**, Antennapedia (*Antp*), Hematopoietically-expressed homeobox protein (*HHEX*), Labial (*lab*), and Reversed polarity (*repo*) (Fig 4).

Gynandromorph

General bilateral mosaicism

The gynandromorph specimen was collected from the Mamak region of Ankara, Türkiye (39.957778, 33.113056) on *Onopordum* sp. in July 2024. The specimen exhibits a pronounced, though imperfect, bilateral split: the right half of the body from the dorsal perspective displays more female-typical morphology and white setal coloration, whereas the left half shows more male-typical morphology with yellow setae. However, multiple anatomical regions reveal intermediate or patchy mosaics. For example, the scutellum is white on both sides with a small yellow spot, and on T1 there is an inter-mixed wave of yellow and white in the medial portion (Fig 5A-5C).

Head, eye, and antennal morphology

Cranial features possess a clear bilateral split with some intermediacy in these features. *B. niveatus* has big eyed males where male eyes are larger than females, showing a substantially greater width than length (Fig 5D and 5E). The left (male) eye of the gynandromorph matches the size and shape of a typical male eye, whereas the right eye is smaller but is intermediate in size and shape between a typical male and female eye. Other features that match typical male and female features in the face include the length of the malar space, the mandible morphology, and the color and density of facial setae, suggesting a distinctive mostly bilateral split (Fig 5D).

Antennal segmentation further corroborates the bilateral split. The left (yellow side) antenna comprises longer flagellomeres which match male reference specimens, while the right (white) side has shorter segments, as in females (Fig 5E). The only deviation from typical males vs. females is in scape length on the female side, which is slightly larger than typical, suggesting some proximal mixed-sex somatic lineages.

Sternite and leg traits

Ventral sternites on anterior metasomal segments predominantly resemble female morphology as they have the female-typical black hairs more than the male typical long pale hairs, even on the left side, although there are more pale

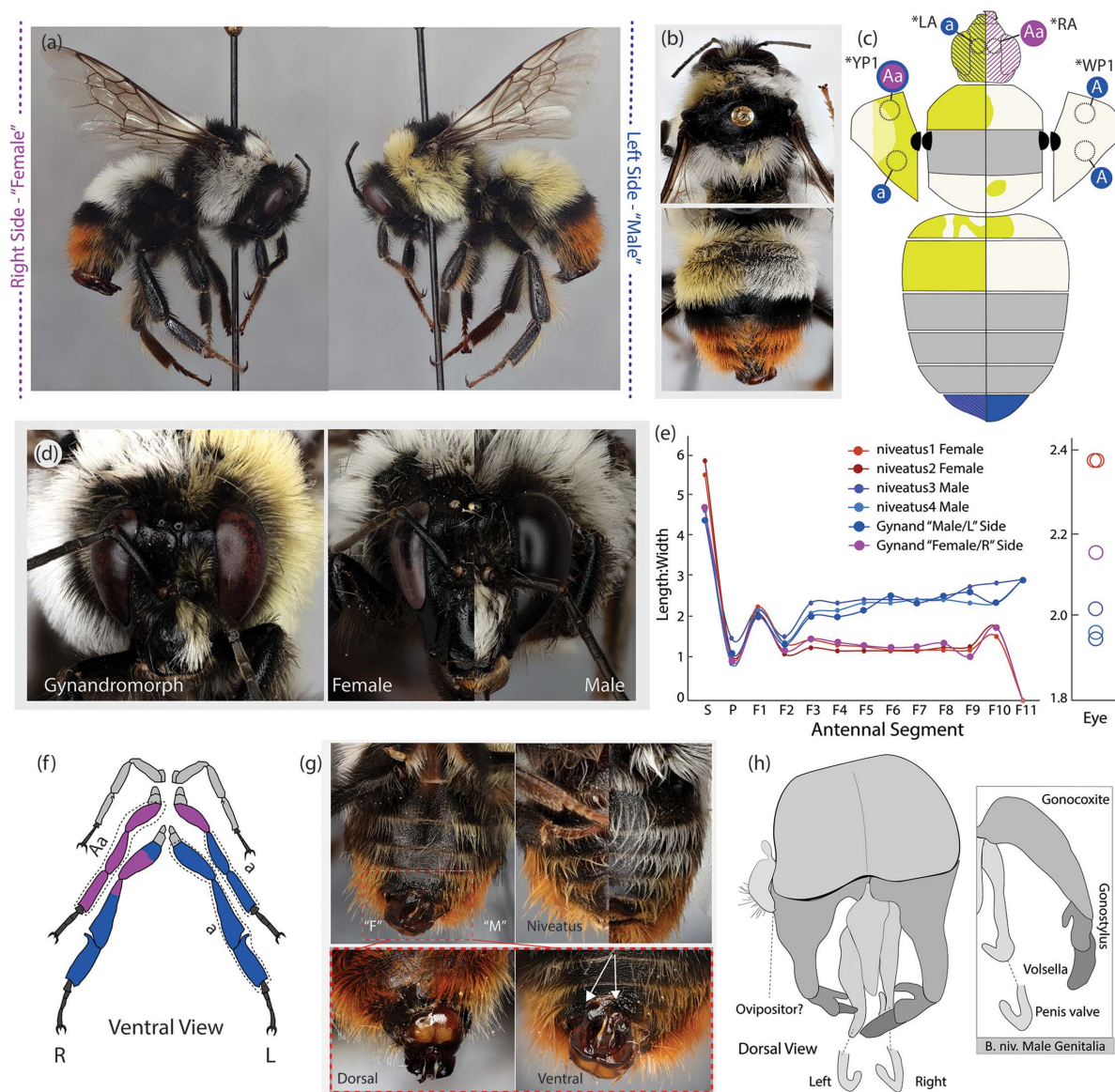


Fig 5. Anatomical features of the gynandromorph sample, with comparison to worker and male *B. niveatus* individuals. (A) Lateral image of the gynandromorph showing the right ("female" and white - matching *B. n. niveatus*) and the left ("male" and yellow - matching *B. n. vorticosus*) side of the bee; **(B)** Dorsal image of the bee where the thorax (top) and the abdominal segments are featured; **(C)** Illustrative map of the dorsal view showing phenotypes of each respective region. The white-yellow color phenotypes are shown with yellow and white color, regions where sex could be inferred are indicated with purple (female) and blue (male) (the left segment with genitalia is mixed sex), and remaining regions of uncertain sex are in light gray. Tissue samples analyzed genetically are indicated by black dashed circles and the inferred haplotypes are shown, with A=*B. n. niveatus* haplotype, a=*B. n. vorticosus* haplotype, each representing haploid male tissues, and Aa representing heterozygous female tissue, or in the case of the left pleuron, potential mixed A and a haploid tissue. Samples used for genomic sequencing are labeled by name and indicated with an asterisk. **(D)** Facial image of the gynandromorph (left) contrasted with the images of the faces of female (worker) and male *B. niveatus* individuals placed side-by-side (right); **(E)** Line graphs indicating the antennal segment lengths, comparing the male, female and the gynandromorph "male and female" antennae (left) and a dot plot for eye size comparison by sample (right); **(F)** Illustration of the ventral view of the legs of the gynandromorph sample with the sampled tissues indicated with dashed areas and their genotype shown. Purple and blue colors indicate female and male sex, respectively, while gray represents regions where sex could not be reliably determined; **(G)** Ventral view of posterior abdominal segments of the gynandromorph (top left) and the merged image of worker and male of *B. niveatus* (top right) along with the dorsal and ventral view of the most posterior segments of the gynandromorph sample (bottom). White arrows indicate the reduced sternites; **(H)** Illustration of genitalia features of the gynandromorph from the dorsal view with a diagram of one side of a normal male genitalia inset.

<https://doi.org/10.1371/journal.pgen.1012060.g005>

hairs on the left side. Males typically have an extra posterior pre-genitalic segment (sternite 6) which instead transitions to a stinger in females. We found this segment on the yellow (male) side but the female side is more like a female in narrowing in the sixth segment and lacking the extra segment, although this shows some intermediacy ([Figs 5E](#) and [S5](#)).

Legs can be sexed by the presence of light longer setae in males and differences in shape for the corbicular tibial hind-leg segment. These distinctions suggest legs on both sides are a mix of male and female, with more female setation on the female-side, however, the female side has a male corbicular morphology in the hind tibia with some intermediacy ([S5 Fig](#)).

Posterior segments and genitalia

The genital capsule as a whole matches to male anatomy ([Fig 5G](#) and [5H](#)). Yet each genital module on the left side is subtly reduced: the gonocoxite is shorter and the gonostylus and volsella are smaller and there is a small structure that looks to be an ovipositor outside the male genitalic structure. The penis-valve hook is more typical on the right side but appears rounded rather than the typical large “V” shape on the left. Intriguingly, this reduced morphology closely resembles *Pyrobombus* spp., which exhibits less pronounced sexual dimorphism and smaller genitalic modules.

Genetic basis of gynandromorphy

Phenotypically, areas of the body containing female traits tend to have white color and areas with male traits tend to be yellow. For example, the more male side of the face is yellow and the female side white, and longer hairs on the sternites and the legs, typical of males, are yellow. Since in Hymenoptera females are diploid (2n) and males are haploid (n) the simplest phenotype-based explanation is somatic loss of one chromosome set in a subset of cells, producing haploid (male) cells on the yellow side with the deletion allele, and diploid (female) cells on the white side that are heterozygous for the color locus.

Genomically, of the four samples tested (right (RA) and left (LA) antennae, left yellow pleuron (YP1) and right white pleuron (WP1), [Fig 5C](#)) each produced signatures of haploidy vs. diploidy based on levels of heterozygosity. For each sample, most contigs were assigned to the same ploidy across the full genome ([S6A Fig](#)) suggesting loss of ploidy spanned the genome rather than select chromosomes. Genomic data supported the right antenna being composed predominantly of diploid cells (female) and the left antenna as haploid (male) ([S6A Fig](#) and [5](#)). These results were also supported by PCR bands and Sanger sequencing. As expected, the female antenna is heterozygous for *snowy* alleles (white form, *A/a*) while the male side is haploid for the yellow allele (*a*). PCR data of male and female legs similarly match expectations, with the male leg haploid for the yellow allele (*a*), and the female leg heterozygous (*A/a*) ([S4 Fig](#) and [5F](#)).

Unexpectedly, however, the white pleuron was inferred to be haploid across the genome using genomic data and with *snowy* PCR sequences, and it contains only the white allele (*A*) ([Figs 5](#), [S4](#), and [S6F](#)). This suggests some of the white tissue is haploid male. The genomic sample from the yellow pleuron was unexpectedly diploid and heterozygous ([S4](#) and [S6A Figs](#)), which does not align with the inferred dominance of the white allele. A second sample taken from the more prominently yellow region ([Fig 5A](#) and [5C](#)) of the left pleuron with PCR was largely haploid *B. n. vorticosus* ([S4 Fig](#)) but sequences show some of the *B. n. niveatus* allele was amplified. Thus, the genetics of the YP1 is less certain, but this region is likely a cellular mosaic of the two alleles ([Fig 5C](#)).

Haploid and diploid cell lineage comparisons

Using heterozygous sites in the diploid RA tissue as a reference, we compared the haploid LA and WP1 samples to the RA diploid sample and found that the majority of heterozygous sites present in the diploid individual differed between haploids (98.25% of 33,187), thus these mostly represent a split of the chromosomal variants of the diploid ([S6C Fig](#)). However, 12.7% of the sites that differ between the haploids (n=4915) are not in the diploid, raising the possibility that the parental origin of the haploids is not simply the diploid tissue ([S6D Fig](#)). Of the alleles differing between haploids that are homozygous in the diploid, only haploid LA differs from the diploid. These unique haploid alleles were distributed fairly

evenly across the contigs (S6C Fig), a pattern not easily explained by most scenarios for gynandromorph formation. Manual inspection of reads confirmed most of these haploid-specific alleles to be accurately called variants, however, when comparing the number of reads with each haploid allele in the “diploid” RA tissue, we found RA to have a consistently higher frequency (~75%) of alleles of WP1 across heterozygous sites (S6E Fig). This suggests that RA is mostly diploid tissue but contains some haploid male tissue with the WP1 “white” haplotype. This would create a bioinformatic bias towards unique LA alleles only, as this can result in RA being called only for the WP1 allele, especially when read counts are low. That the head may represent a chimera of cells is not unexpected, as the flagellomere data suggests the antennal scape of the right antennae is more similar to a male in length and the right eye appears to have intermediacy in size between males and females (Fig 3). Low read coverage in gynandromorph samples (mean/median: 11/6, 9.8/6, 9.2/4 for RA, LA and WP1, respectively) could have also led to unreliable detection of alleles, especially with regard to the ability to accurately call heterozygotes.

Discussion

This study examines the genetic basis of traits in one of the highest fidelity mimicry complexes in bumble bees globally, developing this system for understanding the evolution of similar phenotypic traits and their potential genomic targets. From an ancestral “ground-plan” of yellow coloration [5], the Anatolian mimicry complex involves repeated regional convergence onto white-banded patterns, thus providing numerous replicated natural experiments to understand mimicry and revealing genes both for yellow-white transitions and anterior body patterning in these bees. Using a genome-wide association in the snowy bumble bee *Bombus niveatus*, we localize the white-yellow polymorphism to a narrowly delimited interval that we coined the *snowy* locus near the homeobox *BarH* gene. This gene has well-documented roles in thoracic, pigment, and sensory patterning in *Drosophila*. The locus is narrowed to few fixed SNPs and a 409 bp lineage-specific genomic duplication present only in the white haplotype.

Cis-regulation of *BarH* is implicated in color variation

The location of *snowy* suggests cis-regulatory alteration of *BarH* is the most likely candidate influencing the development of pigmented setae in these bumble bees. It is the closest gene, the variants occur in the more-typical 5' regulatory region, and its 5'UTR is also linked to the regulatory mutations.

Although functional links are speculative, *BarH* has three previous known functions that link it to the role it plays in these bees. As a selector homeobox gene, *BarH2* is known to influence the location of patterning of the thorax in *Drosophila* [22], thus is involved in spatial patterning similar to these bumble bees. In particular it patterns macrochaetae on the notum in *Drosophila* and interacts with *achaete–scute*–dependent sensillum development [22], highlighting its broader role in the patterning of epidermal projections. In bumble bees this gene also regulates epidermal projections (the setae). *BarH2* cooperates with *BarH1* to influence pterin-based eye pigmentation in *Drosophila* [23–24]. In a remarkably parallel case, white and yellow pterin variants of *Colias* butterflies were also found to be driven by an enhancer of *BarH1* [25]. Our analysis shows that this is the only copy in *Colias* and it is orthologous to the bumble bee *BarH*. In these butterflies *BarH* upregulation generates the white form by preventing the formation of pigment granules in butterfly scales, which are homologous to setae. Similarly, in bumble bees, the yellow vs. white setal coloration is likely driven by shifts in the presence of pterins [13]. Changes that alter pigment chemistry, transport, or deposition could produce the white-yellow switch observed here [15]. Apart from *BarH*, the closest alternative genes *proline-rich protein 12-like* and *sex-regulated protein janus-A* do not have functional relevance to coloration, bristle or setal patterning, but rather exhibit nervous system [26] and somatic sex differentiation [27] functions, respectively.

Linkage disequilibrium analysis reveals a block of especially high r^2 at the association peak. Localized LD can suggest a recent selective sweep [28], perhaps in this case from strong selection on white morphs due to fitness advantages of mimicry. The duplication, which generates structural variation, could also be the cause of reduced

recombination [29–30] and may therefore be a mechanism to help stabilize linked variants [31–32]. The linkage between this regulator region and the *BarH* 5'UTR exon could suggest epistasis between regulatory elements of the transcript and their expression [33].

This study fits a broader pattern in bumble bee mimicry genetics, in which major developmental homeobox genes tend to be responsible for color patterning. Regulatory changes at posterior abdominal Hox gene *Abd-B* have been implicated as the main driver of red and black mid-abdominal variation in the North American mimicry complex and this locus was found to be a hotspot for color evolution [10]. Our finding that a different homeobox locus (*BarH*) is implicated in coloration further highlights how developmental genes tend to be co-opted to drive major and localized changes in external adult patterning. In bumble bees, pale as opposed to melanic coloration is triggered in late pupation, while yellow is deposited in hairs after adult eclosion [15], thus this gene may operate to control this late deposition of pterin pigments in the anterior body region.

Evolution and potential function of the *BarH* color locus

The *B. niveatus* case suggests that similar phenotypes generated by mimicry arise through independent genetic mutations. Broad comparative Sanger sequencing across comimicking lineages showed that the tandem duplication is unique to *B. n. niveatus*. In terms of the allelic states of the three SNPs; most of these are variable across comimics and do not form a diagnostic cross taxon white haplotype. The second SNP showed consistent variation to phenotype between the *B. niveatus* clade and in the yellow-white paired taxa aligned to *B. cullumanus*, but this variant did not match color phenotypes in other lineages. Together these patterns indicate that the duplication and involved SNPs are mostly unique to *B. niveatus* and that other unidentified variants, either in this gene or other loci, explain parallel changes in comimicking bumble bees.

Our *in silico* regulatory analysis points to the possibility that alteration of transcription factor binding in *snowy* could contribute to yellow/white color shifts. The white-specific duplication amplifies existing TFBS counts, drastically increasing local binding potential for several developmental TFs and thereby plausibly modulating regulatory output through shifts in binding site number and threshold responses. By contrast, the fixed SNPs we detect only subtly alter predicted binding configurations. Our data point to differences in the number of binding sites being more important in driving phenotype as fixed differences do not create new transcription factor binding sites.

Small-scale copy number changes can drive large changes in expression by altering binding site density or enhancer strength, creating threshold effects in regulatory outputs [34]. Moreover, duplicate copies of regulatory sequence may diverge in function, either by partitioning ancestral roles (subfunctionalization) or by acquiring new regulatory or coding functions (neofunctionalization). Local duplications followed by point mutations were found to be a common way to generate new regulatory sites in *Drosophila* [35]. How ubiquitous these are in generating natural variants beyond the present study requires more accumulated information on specific sites regulating natural variants [36].

Our data support a classic Mendelian inheritance pattern for this locus, whereby yellow individuals are homozygous for the non-duplicated allele and white individuals are either homozygous or heterozygous for the duplication, consistent with the white insertion allele being dominant. A mechanistic explanation for this is that by amplifying transcription factor binding sites, even a single allele for the duplicated region may elevate regulatory activity past the threshold required for a certain phenotype [37].

Genetics of a yellow-white gynandromorph

The dissected bilateral gynandromorph provides a within-individual test of how color alleles translate to phenotype while also revealing the mechanisms whereby this gynandromorph was generated. While we tried to keep destruction of the sample to a minimum, the molecular and genomic data obtained from the small set of sampled tissues are informative and internally consistent with somatic mosaicism.

PCR and Sanger assays on the sampled tissues show correspondence between duplication genotype and tissue color. The white pleuron is haploid for the white insertion. The yellow pleuron yielded heterozygous signatures, but most likely this was a mosaic of white and yellow tissue given admixed hair colors in this side and a greater proportion of yellow *B. n. vorticatus* signal towards the more yellow dorsal side, thus we suggest yellow is likely a haploid with the *B. n. vorticatus* alleles. Conversely, most parts of the body that are female are diploid for both color alleles, while male tissues are haploid for the yellow *B. n. vorticatus* deletion. Together these validated assays indicate the specimen contains three somatic cell-lineages: diploid (2n) female cells, haploid (n) male cells associated with yellow setae, and a second class of haploid (n) male cells associated with white setae.

In Hymenoptera, gynandromorphs are categorized as of three main types: (i) bilateral if sex characteristics are distributed equally and symmetric, (ii) transversal if these characters are distributed by major regions but not symmetrically and (iii) mosaic if sex characters are distributed randomly in sections across the body [38]. Our specimen does not fit neatly into any single category and is best described as a partial bilateral mosaic. Across bees (Apidae), gynandromorphy has been documented in ~142 species [38–41] including 13 bumble bees, of which three were bilateral, six were mosaics, and four were transversal gynandromorphs (reviewed by [38,42]). These studies hypothesize that independent chromosomal eliminations or variation in Complementary Sex Determination (CSD) gene expression during embryogenesis are involved [43]. Early chromosome loss can partition cell lineages into distinct developmental territories, producing bilateral or mosaic outcomes depending on the stage and extent of chromosome elimination ([44], *Drosophila*).

The only molecular study in a bumble bee gynandromorph was on *B. ignitus* [42,45], which showed sex-determining genes are expressed as expected given the sex of the tissue. Our genomic data show fairly consistent ploidy across chromosomes, indicating that gynandromorphy in this specimen cannot be explained solely by perturbation of the sex-determining locus. The coexistence of two haploid lineages alongside diploid cells is most likely explained by typical fertilization of male and female gametes followed by splitting of the diploid chromosomes into two haploid sets/cells, given the haploids are variable in mostly the same sites as the heterozygote across chromosomes. While deviation in LA makes it seem that another gamete may be involved, these rare deviations could be explained by sampled tissues being partial mosaics (RA=Diploid+WP1 haploid cells), combined with bioinformatic errors inherent with alignment and variant calling (S6B Fig). Note that bumble bees are singly mated [46], making haploid male gametes clonal and multiple male gametes an unlikely source of this variation. A more thorough understanding of mechanisms of gynandromorphy would benefit from high coverage and long-read sequencing of fewer cells per tissue.

The problem with using color to delimit species

B. niveatus and *B. vorticatus* were historically defined only by their coloration and only recently have been synonymized [18–20]. Our genomic data shows both color forms are part of an admixed population, and the gynandromorph supports conspecificity of these species as both color forms occur in the same individual. Mullerian mimicry decouples color variation from species boundaries, as regional selection favors intraspecific polymorphisms and cross-species convergence that confounds diagnosis. This case presents another example of why color, often used as a key diagnostic feature in bumble bees, is unreliable as an exclusive criterion for species delimitation [47–50]. Our data highlights why color is so unreliable, as the white-yellow polymorphism in *B. niveatus* is controlled by a single locus with simple Mendelian inheritance, thus can easily be gained or lost. Such simple patterns of inheritance for coloration have been found in other bumble bee systems where major color polymorphisms are controlled by either a single locus, such as the red-black variation in *B. melanopygus* [8] and *B. breviceps* [12], or a few loci in *B. rufocinctus* and *B. vancouverensis* [10,51].

Conclusions

Using a combination of association mapping, comparative sequencing, and validation with a gynandromorph sample, we reveal that the *B. niveatus* white mimicry morph is a derived condition most likely generated by a lineage-specific

cis-regulatory duplication in *BarH*. While compelling, these data would benefit from follow-up assays to determine allelic function, such as transcriptomics to determine whether this leads to differential gene expression of *BarH*, gene editing (e.g., remove the repeat region) or knockdown approaches, and assays like ATAC-seq to assess transcription factor binding. In parallel, pigment-biochemistry assays should confirm whether the white morph represents active repression of pterin (or other) pigment pathways or instead results from a structural failure of pigment deposition. Determining whether *BarH* reflects an optimal genetic target that is repeatedly employed or a lineage-specific opportunity will require comparative functional work across comimics in this mimicry complex.

Larger structural variants, such as the *BarH* cis-regulatory duplication, are challenging to detect using standard short-read sequencing approaches. In this case the duplication was only detected because of fixed differences in linked SNPs. When short-read sequencing is used for GWAS, approaches that evaluate read coverage drops or other signatures of copy number variation should be considered to capture the full spectrum of causal variants.

Finally, the Anatolian mimicry complex is an outstanding comparative system deserving further research. The geographic region exhibits repeated convergence on white-banded phenotypes including ample intraspecific polymorphism across lineages amenable for assessing the genetic basis and evolution of convergence across this mimicry system. By combining comparative genetics, regulatory predictions, and future expression and functional experiments, this system can illuminate both the proximate mechanisms of color evolution and the broader evolutionary rules that govern this mimicry.

Materials and methods

Sample preparation and sequencing

DNA was extracted from two legs and thoracic muscles for each sample with a Zymo Quick-DNA Miniprep kit using standard protocols. DNA libraries were prepared with Illumina DNA kits followed by paired-end 100 bp sequencing on the NextSeq 2000 XLEAP P4 at ~ 18-22X coverage at the Penn State Huck Genomics Core Facility (University Park, PA) (S1 Table). Specimen vouchers are retained at the Hines Lab at the Pennsylvania State University and will ultimately be deposited at the Frost Entomology Museum at Pennsylvania State University.

Genome-wide association and principal component analyses

The quality assessment of each sequencing file was conducted using FastQC v.0.11.9 [52] and MultiQC v.1.21 [53]. Low quality reads, poly-g tails and adapters were removed using trimmomatic v0.39 (SLIDINGWINDOW:4:25 ILLUMINACLIP:-adapters.fa:2:30:5 MINLEN:36 LEADING:3 TRAILING:3) [54] (Quality Phred score cutoff < 25). High quality reads were then aligned to the *Bombus niveatus* genome assembly [55] using bwa-mem [56] using the strict alignment option (-T 40) where read alignments with scores lower than 40 were not allowed. Read group information and indexing steps were completed using Picard tools v.2.23.9 (available from: broadinstitute.github.io/picard/index.html) and sambamba v1.0.0 [57], respectively. Duplicate reads were removed and variant calling steps were conducted using GATK v4.4.0.0 [58] with haploid sample specific settings (--ploidy 1 -ERC BP_RESOLUTION), to generate Genomic VCF (GVCF) files. GVCF files generated for each sample were combined into a genomic database using the *GenomicsDBImport* function in GATK. Joint genotyping was then performed on this database with the *GenotypeGVCFs* function to produce a multi-sample raw VCF file. From this file, indels and SNPs were separated using *SelectVariants* and subjected to quality filtering with *VariantFiltration* following standard bootstrap filtering steps in Best Practices with GATK guideline [59–60]. For SNPs, variants were filtered out if they showed low quality by depth (QD < 2.0), strand bias (FS > 60.0, SOR > 4.0), poor mapping quality (MQ < 40.0), or evidence of biased mapping or read positioning (MQRankSum < -12.5; ReadPosRankSum < -8.0). For indels, variants were filtered out using thresholds of QD < 2.0, FS > 60.0, and SOR > 4.0. These criteria remove spurious variants arising from low sequencing quality, alignment artifacts, or sequencing errors, retaining only high-confidence sites.

The resulting high confidence variant sets were used to recalibrate base quality scores with *BaseRecalibrator* and applied back to each alignment file using *ApplyBQSR*. Finally, the recalibrated BAM files were used for variant calling across all samples with *bcftools* v1.16 [61] via the *mpileup* and *call* functions, using haploid ploidy settings (--ploidy 1). The final variant call file was then used for running Fisher's exact test and GWAS using case vs. control phenotypes (white vs yellow) with PLINK v1.90b6.21 [62]. A Manhattan plot was generated using qqman R package v0.1.9 [63] and the significance threshold was determined using a Bonferroni correction to control the family-wise error rate. Variants that are above this threshold were considered candidates for further annotation and downstream analyses.

A principal components analysis (PCA) was conducted using the finalized vcf file with the following parameters in PLINK: --vcf-half-call haploid --allow-extra-chr --allow-no-sex --distance-matrix --double-id --no-parents. The resulting pairwise distance matrix and corresponding sample identifiers were used to visualize the PCA plot via the tidyverse v2.0.0 [64] R v4.2.2 package. A linkage disequilibrium (LD) analysis on the contig of the most associated region was also conducted using the following parameters: --allow-no-sex --double-id --vcf-half-call haploid --r2 --ld-window-r2 0 --chr contig_67 --allow-extra-chr, with varying sliding window sizes (--ld-window-kb) of 4 kb, 10 kb, 30 kb, and 100 kb to assess how LD decays at different physical scales. A similar methodology described by [65] was followed to track the linked genes outside of our most associated site and to potentially detect any epistasis effect in our *BarH* exon. High linkage ($r^2 > 0.7$) positions with their corresponding minor allele frequency (MAF) and the level of fixation were detected and their sequence changes in exon regions were predicted using snpEff v5.2 [66].

Analysis of candidate loci

The genome of *B. niveatus* does not have an annotation and thus, to confirm the identity of resulting GWAS peaks, we generated the annotation file (.gtf) using the LiftOff standalone tool [67] using the *Bombus terrestris* genome (GCF_910591885.1) and its associated annotation files as a reference. Results were confirmed by generating a local blast database and using nucleotide blast on manually selected genomic regions [68]. The candidate variant sites were inspected using Integrated Genome Viewer (IGV) [69].

PCR primers were designed to span ~1 kb including all identified fixed variants in our association peak (Forward: 5'-ACAGTAATGCTCGAGGTCGTG-3' & Reverse: 5'-ATCTGTACCTCGTAGCCGGT-3' ~ 52.5 °C annealing temperature) using Primer3web v4.1.0 [70–72]. The same primer pairs were kept for every lineage other than the *B. shaposhnikovii* sister species pair which required a different reverse primer (Reverse: 5'-CATCTGTAACCGTGCCGGT-3'). PCR amplified products were purified with ExoSap and Sanger sequenced by the PSU Huck Genomics Sequencing Center (University Park, PA). Sequences were manually aligned and edited to confirm the variants using Geneious (available from: <https://www.geneious.com>).

To assess the impacts of reference bias during alignment, we generated a *de novo* genome assembly for *B. n. vorticocosus* (yellow form - VOR3_5) using SPAdes v4.2.0 [73]. Matching assembled sequences to the locus were compared against both Sanger data and the *B. n. niveatus* reference sequence for any variants not detected with the analysis using the *B. n. niveatus* genome.

Candidate locus phylogeny

Protein and coding-sequence (CDS) candidates for *BarH* were obtained from OrthoDB for a taxonomically broad sampling of insects (Hymenoptera, Lepidoptera, Coleoptera, Diptera) as well as our *B. n. niveatus* genome. In addition, to check paralog identity (*BarH1* vs *BarH2*), we performed reciprocal BLAST searches (tblastn/blastp, NCBI blast+) using *Drosophila melanogaster BarH1* and *BarH2* as queries against each target proteome. When both *BarH1* and *BarH2* top hits mapped to the same locus (reciprocal best hits) in a given taxon, we treated that taxon as carrying a single BarH-type ortholog relative to *Drosophila*. We also confirmed using BLAST that no additional paralogs in other lineages were close matches.

Protein and CDS nucleotide sequences were aligned separately, and poorly aligned sequence was removed with trimAl v1.5 [74]. Maximum-likelihood phylogenies were inferred from these alignments using IQ-TREE [75].

Transcription Factor Binding Site Analysis (TFBS)

The TFBS analysis was conducted in R v4.2.2 using Biostrings v2.66 [76], TFBSTools v1.36 [77] and JASPAR2020 v0.99.10 [78] packages. Insect-specific position-weight matrices (PWMs) were obtained from the JASPAR CORE collection via JASPAR2020 and converted to probabilistic PWMs with TFBSTools using *Bombus*-optimized background nucleotide frequencies (A=0.31, C=0.19, G=0.19, T=0.31) [79]. Each sequence was filtered at a 90% relative-score threshold (only matches scoring ≥ 0.9 of the maximal PWM score were retained), returning for each hit the JASPAR motif ID, genomic start/end, strand, raw and relative score, and matched subsequence. Hits from both haplotypes were combined into a single data frame and downstream summary statistics and visualizations were produced with dplyr v1.14.0 [80], ggplot2 v3.5.1 [81], and tidyr v1.3.0 [82] packages, including total site counts, top-motif frequencies per haplotype and motif densities (20 bp windows) along the genomic coordinates.

Haplotype network

For the network analysis, the haplotypes were generated from Sanger sequences and, for *B. n. niveatus*, they were further separated by the Ancestral and Repeat region sequences. The copies were overlaid into a single copy alignment, allowing assessment of the allelic origin of the duplication. Haplotype networks were generated using the TCS algorithm in the Hapsolutely tool [83].

Gynandromorph sampling and phenotypic analysis

The gynandromorph specimen was photographed using a Macroscopic Solutions microkit imaging station (Tolland, CT) and stacked via Zerene Stacker LLC software. Detailed examination under an Olympus SZ61 stereo microscope was conducted to identify sex- and form-specific phenotypic traits. Measurements of lengths of each antennal segment and the broadest width and height of the eyes were determined using an ocular micrometer, comparing both eyes and antennae of the gynandromorph as well as of control male and female *B. niveatus* specimens (Türkiye:Aksaray (38.454722, 34.157306) 2/8/2002, Hines). Tissues were collected from the body regions that had distinct coloration; white pleuron 1 (WP1), yellow pleuron 1 (YP1), and sex-specific characters, including the right antenna (RA) and left antenna (LA). DNA was extracted using the Omega E.Z.N.A. Tissue DNA Kit (BIO-TEK) kit and the libraries were prepared using the NEB-Next UltraExpress FS DNA Library Prep Kit followed by the same sequencing technique as described above. Legs with only one sex-specific trait (Male Midleg (MM), Male Hindleg (MH) and Female Midleg (FM)) and additional thoracic tissues (yellow pleuron 2), and white pleuron 2 (WP2) were also extracted, amplified using PCR (same primers as above), and Sanger sequenced to infer which copies and alleles were present.

Gynandromorph genetic and ploidy analysis

To assess the somatic ploidy state of each sex- and color-specific tissue in our gynandromorph, we called variants, filtered them stringently, and then quantified heterozygosity per contig. First, for each of the four genomic samples, we performed all-site variant calling (*BPRESOLUTION*) and joint genotyping (GATK v4.4.0.0 *GenotypeGVCFs*) against the *B. n. niveatus* reference genome under the diploid setting (*--ploidy 2*) which generated a diploid VCF file for each tissue. Next, we retained only sites with depth ≥ 6 , mapping quality ≥ 50 , and genotype quality ≥ 20 to ensure a high-confidence call set. Then, to focus our analysis on the most contiguous portions of the assembly, we selected all contigs ≥ 250 kb and computed per-contig heterozygosity on this high-confidence call set using the vcfR v1.15.0 package [84] in R v4.2.2. Finally, we assigned ploidy to each tissue based on rates of heterozygosity.

Heterozygosity-based cell lineage analysis

To investigate the origin of tissues from the gynandromorph specimen, we compared variants among three samples with clear ploidy status (RA, LA and WP1; excludes YP1 which is likely a mixed sample). Variant genotypes were jointly called for each sample using GATK *GenotypeGVCFs* retaining invariant positions (*--include-non-variant-sites*). All downstream genotype comparisons were restricted to sites that passed per sample depth filters (diploid RA: DP 6–16; haploid LA and WP1: DP 4–14). RA heterozygotes were defined by $0.25 \leq (\text{minor allele balance} = \text{MAB}) \leq 0.75$; RA heterozygous sites with $\text{MAB} < 0.25$ or $\text{MAB} > 0.75$ were removed from analysis as they were considered less reliably called (3.5% removed). For the haploid tissues (LA, WP1) we required sites to be fixed to avoid calling misaligned sites: we used samtools *mpileup* to compute per-position reference and alternate read counts and removed any haploid positions with any intermediate allele fraction (0.005% of sites). From this we produced a set of all common sites that passed depth and genotype quality filters in all three samples and had a variant in at least one of the samples. For RA-heterozygous sites we calculated proportions of matched and mismatched alleles in the haploid samples and we determined for RA-homozygous sites whether LA or WP1 contained the unique allele. To place site level similarity results in chromosomal context, we mapped contigs to chromosomes using BLAST alignments against the *Bombus huntii* reference genome (GCF_024542735.1) and inspected the chromosomal distribution of matched/mismatched/unique allele sites by position relative to these chromosome sets. A subset of randomly selected informative SNPs from each comparison category was also inspected in IGV to confirm calls. All downstream processing, comparisons and plotting was performed in R using *dplyr*, *tidyr* and *ggplot2* packages.

Supporting information

S1 Fig. Pinned male *B. niveatus* individuals featuring both subspecies. White form - 2 individuals from left; yellow forms - 4 individuals from right. Note the differing intensities of yellow setal coloration. (Photo Credit: Çiğdem Özenirler). (TIF)

S2 Fig. Features of the ancestral and repeat region of the color locus, including read coverage of each sample and TF binding localities of each haplotype across the region. (A) Illumina read coverages of each individual sample across the color locus (NIV - *B. n. niveatus*, VOR - *B. n. vorticosus*). The orange line highlights the *B. n. vorticosus* sample VOR1_8; (B) Genomic blocks featuring the TF binding site coverage by morph and across both ancestral and repeat genomic regions, with fixed/mostly-fixed SNPs indicated with red/black dashed lines, respectively (C) Motif differences across each yellow/white sister lineage pair and between two haplotypes of the yellow form are shown by genomic position. Asterisks (*) indicate the novel TF bindings sites due to mostly-fixed (as opposed to fully fixed) SNPs and bolded SNPs indicate the matching variants between *B. niveatus* and *B. cullumanus* clades. From top to bottom; NIV/VOR: *B.n.niveatus* vs *B.n.vorticosus*; VOR/VOR1_8: *B.n.vorticosus* (Haplotype 1) vs *B.n.vorticosus* (Haplotype 2); BRO/PRA: *B.brodmannicus* vs *B.pratorum*; APO/SEM: *B.apollineus* vs *B.semenoviellus*; INC/LAP: *B.incertus* vs *B.lapidarius*; RUD/RUD: *B.ruderarius* white vs yellow form; POM/RUD: *B.pomorum* vs *B.ruderarius* yellow form. (TIF)

S3 Fig. Linkage disequilibrium block showcasing the linkage between variations within the 30 kb (left) and 3.5 kb region (right) around the associated peak. An exon of *BarH* closest to the regulatory region is indicated with a solid blue bar and the intron region is indicated with a dashed blue line while the arrow indicates the direction of the unmapped portion of the gene. The black box with dashed black line indicates the 3.5kb zoomed in region on the right. (TIF)

S4 Fig. Gel images of the snowy locus PCR products. This includes gynandromorph tissues and workers as well as a few representative males of each color form to show how these bands appear for haploid alleles. From left to right for the

gynandromorph: Female Mid-Leg, Male Mid-Leg, Male Hind-Leg, Yellow Pleuron 1, White Pleuron 2, White Pleuron 1, Yellow Pleuron 2, Right Antenna, Left Antenna, *B. n. niveatus* male (Niv1_4 and Niv1_5), *B. n. vorticosus* male (Vor2_6 and Vor3_1), *B. n. niveatus* worker 1, *B. n. niveatus* worker 2, *B. n. vorticosus* worker 1. For the workers, three PCRs were run for each sample at, from left to right for each sample, 50, 51, and 52°C PCR annealing temperatures respectively. Additional gels were also prepared for the remaining male samples used for genomic sequencing and are available in the online data file. Gel images from separately run gels are placed together and aligned based on the ladder for presentation but were not otherwise altered from original images.

(TIF)

S5 Fig. Phenotypic comparisons of the gynandromorph, worker and male leg characteristics. (A) Comparative image panel showcasing the gynandromorph “female” and “male” side legs; (B) Worker mid- and hind-legs; (C) Male mid- and hind-legs of *B. niveatus*.

(TIF)

S6 Fig. Ploidy and cell-lineage analysis across different tissues of the gynandromorph. (A) For each of the four tissue samples, the proportion of variant sites with homozygous alternate (1/1) genotypes compared to the reference, which is 1-(% heterozygous sites) under a diploid model. Only contigs longer than 250 kb are shown (each red bar represents a contig). Contigs with a high proportion of homozygous alternate calls indicate that most variable sites on a contig lack heterozygous (0/1) genotypes, suggesting haploidy, whereas the low proportions indicate diploidy. This pattern was observed across broad sets of contigs rather than being confined to isolated loci. This discordance was not explainable by simple coverage artifacts (read depth distributions were examined and do not globally account for the low frequency of heterozygotes in WP1 and LA); (B) A schematic of the potential origin of the gynandromorph whereby fertilization is followed by the cell splitting into two new haploid cells and both the diploid and two haploids are retained in the tissues; (C) Top: Similarity assessment of mismatching alleles between haploid tissues (LA vs WP1) to the alleles in diploid RA, whereby “Match Het Variants” are sites that have a difference between haploids and are heterozygous for the same variants for RA, and in gray are alleles only present in haploid LA, thus the haploid WP1 matches the homozygous diploid RA. No alleles are only present in WP1. Bottom: For all heterozygous sites in RA, red indicates sites where two haploid tissues (LA and WP1) carry different alleles, thus split the alleles of the heterozygote, while blue indicates where both haploids carry the same allele at these heterozygous sites; (D) A Venn diagram of the overlap between heterozygous sites and sites different between the haploids showing that these sites are mostly overlapping but there is a low frequency of unique sites among the haploids and some unique heterozygous alleles in the diploid not present in the haploids. (E) Histogram of proportion of WP1 haploid alleles compared to LA alleles (WP1/(LA+WP1)) across the variable sites. A true heterozygote should have 50% of each allele but most sites are skewed towards WP1; (F) Alignment with indicated variants relative to the ancestral *B. n. niveatus* reference including comparisons between only the ancestral regions of the consensus sequence of *B. n. niveatus* (NIV), the consensus sequence in this region of *B. n. vorticosus* (VOR), white-pleuron 1 tissue of gynandromorph (WP1), left-antenna (LA) and right-antenna (RA). Variants that are fixed for a given sample are in black, fixed SNPs from the GWAS are in red, while variants that are not fixed in a color form are in gray. The presence and absence of the repeat region are indicated with blue and crossed-out boxes, respectively. The half bar for RA indicates heterozygosity. For part C, sites with <25% minor alleles would not have been called as heterozygote and were removed from the data.

(TIF)

S1 Table. *B. niveatus* genomic sequencing samples, their sex, phenotypes, localities, read pairs, and read coverage after alignments.

(DOCX)

Acknowledgments

We thank Burcu Daşer Özgüşi and Kurtuluş Özgüşi for their contribution in *B.niveatus vorticosus* samples from the Eskişehir region.

Author contributions

Conceptualization: Tunç Dabak, Heather M. Hines.

Data curation: Tunç Dabak, Seçil Aytekin, Ahmet Murat Aytekin, Heather M. Hines.

Formal analysis: Tunç Dabak, Cecil Smith.

Funding acquisition: Heather M. Hines.

Investigation: Tunç Dabak, Çiğdem Özenirler, Ece Kamalak, Cecil Smith, Heather M. Hines.

Methodology: Tunç Dabak, Heather M. Hines.

Project administration: Heather M. Hines.

Resources: Çiğdem Özenirler, Ece Kamalak, Heather M. Hines.

Supervision: Heather M. Hines.

Validation: Ahmet Murat Aytekin.

Visualization: Tunç Dabak, Çiğdem Özenirler, Ece Kamalak, Cecil Smith, Heather M. Hines.

Writing – original draft: Tunç Dabak, Heather M. Hines.

Writing – review & editing: Tunç Dabak, Çiğdem Özenirler, Ece Kamalak, Cecil Smith, Seçil Aytekin, Ahmet Murat Aytekin, Heather M. Hines.

References

1. Heinrich B. Bumblebee economics. Cambridge, Massachusetts: Harvard University Press. 1979.
2. Goulson D. Bumblebees: behaviour, ecology, and conservation. 2nd ed. Oxford University Press. 2009.
3. Mola JM, Hemberger J, Kochanski J, Richardson LL, Pearse IS. The Importance of Forests in Bumble Bee Biology and Conservation. *BioScience*. 2021;71(12):1234–48. <https://doi.org/10.1093/biosci/biab121>
4. Rapti Z, Duennes MA, Cameron SA. Defining the colour pattern phenotype in bumble bees (*Bombus*): a new model for evo devo. *Biol J Linn Soc Lond*. 2014;113(2):384–404. <https://doi.org/10.1111/bij.12356>
5. Williams P. The distribution of bumblebee colour patterns worldwide: possible significance for thermoregulation, crypsis, and warning mimicry. *Biological Journal of the Linnean Society*. 2007;92(1):97–118. <https://doi.org/10.1111/j.1095-8312.2007.00878.x>
6. Mallet J, Joron M. Evolution of diversity in warning color and mimicry: polymorphisms, shifting balance, and speciation. *Annu Rev Ecol Syst*. 1999;30:201–33.
7. Heliconius Genome Consortium. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*. 2012;487(7405):94–8. <https://doi.org/10.1038/nature11041> PMID: 22722851
8. Tian L, Rahman SR, Ezray BD, Franzini L, Strange JP, Lhomme P, et al. A homeotic shift late in development drives mimetic color variation in a bumble bee. *Proc Natl Acad Sci U S A*. 2019;116(24):11857–65. <https://doi.org/10.1073/pnas.1900365116> PMID: 31043564
9. Rahman SR, Terranova T, Tian L, Hines HM. Developmental Transcriptomics Reveals a Gene Network Driving Mimetic Color Variation in a Bumble Bee. *Genome Biol Evol*. 2021;13(6):evab080. <https://doi.org/10.1093/gbe/evab080> PMID: 33881508
10. Hines HM, Dabak T, Rahman SR, Terranova T, Tian L, Smith C, et al. Similar Genetic Routes Are Independently Targeted for Mimetic Color Convergence in Bumble Bees. *Mol Biol Evol*. 2025;42(9):msaf187. <https://doi.org/10.1093/molbev/msaf187> PMID: 40795243
11. Hines HM, Witkowski P, Wilson JS, Wakamatsu K. Melanic variation underlies aposematic color variation in two hymenopteran mimicry systems. *PLoS One*. 2017;12(7):e0182135. <https://doi.org/10.1371/journal.pone.0182135> PMID: 28753659
12. Yang W, Cui J, Chen Y, Wang C, Yin Y, Zhang W, et al. Genetic Modification of a Hox Locus Drives Mimetic Color Pattern Variation in a Highly Polymorphic Bumble Bee. *Mol Biol Evol*. 2023;40(12):msad261. <https://doi.org/10.1093/molbev/msad261> PMID: 38039153
13. Hines HM. Bumble bees (*Apidae*: *Bombus*) through the ages: Historical biogeography and the evolution of color diversity. Champaign, Illinois: University of Illinois at Urbana-Champaign. 2008.

14. Polidori C, Jorge A, Ormosa C. Eumelanin and pheomelanin are predominant pigments in bumblebee (Apidae: Bombus) pubescence. *PeerJ*. 2017;5:e3300. <https://doi.org/10.7717/peerj.3300>
15. Hines HM, Kilpatrick SK, Mikó I, Snellings D, López-Urbe MM, Tian L. The diversity, evolution, and development of setal morphologies in bumble bees (Hymenoptera: Apidae: Bombus spp.). *PeerJ*. 2022;10:e14555. <https://doi.org/10.7717/peerj.14555> PMID: [36573237](https://pubmed.ncbi.nlm.nih.gov/36573237/)
16. Rahman SR, Cnaani J, Kinch LN, Grishin NV, Hines HM. A combined RAD-Seq and WGS approach reveals the genomic basis of yellow color variation in bumble bee *Bombus terrestris*. *Sci Rep*. 2021;11(1):7996. <https://doi.org/10.1038/s41598-021-87194-y> PMID: [33846496](https://pubmed.ncbi.nlm.nih.gov/33846496/)
17. Hines HM. Historical biogeography, divergence times, and diversification patterns of bumble bees (Hymenoptera: Apidae: Bombus). *Syst Biol*. 2008;57(1):58–75. <https://doi.org/10.1080/10635150801898912> PMID: [18275002](https://pubmed.ncbi.nlm.nih.gov/18275002/)
18. Rasmont P, Terzo M, Aytekin AM, Hines H, Urbanova K, Cahlikova L, et al. Cephalic secretions of the bumblebee subgenus *Sibiricobombus* Vogt suggest *Bombus niveatus* Kriechbaumer and *Bombus vorticosus* Gerstaecker are conspecific (Hymenoptera, Apidae, Bombus). *Apidologie*. 2005;36(4):571–84. <https://doi.org/10.1051/apido:2005047>
19. Aytekin MA, Terzo M, Rasmont P, Çağatay N. Landmark based geometric morphometric analysis of wing shape in *Sibiricobombus* Vogt (Hymenoptera: Apidae: Bombus Latreille). *Annales de la Société entomologique de France (NS)*. 2007;43(1):95–102. <https://doi.org/10.1080/00379271.2007.10697499>
20. Cameron SA, Hines HM, Williams PH. A comprehensive phylogeny of the bumble bees (*Bombus*). *Biological Journal of the Linnean Society*. 2007;91(1):161–88. <https://doi.org/10.1111/j.1095-8312.2007.00784.x>
21. Rasmont P, Aytekin AM, Kaftanoğlu O, Flagothier D. The bumblebees of Turkey. Mons, Gembloux: Atlas Hymenoptera. 2009.
22. Sato M, Kojima T, Michiue T, Saigo K. Bar homeobox genes are latitudinal prepatter genes in the developing *Drosophila notum* whose expression is regulated by the concerted functions of decapentaplegic and wingless. *Development*. 1999;126(7):1457–66. <https://doi.org/10.1242/dev.126.7.1457> PMID: [10068639](https://pubmed.ncbi.nlm.nih.gov/10068639/)
23. Higashijima S, Kojima T, Michiue T, Ishimaru S, Emori Y, Saigo K. Dual Bar homeo box genes of *Drosophila* required in two photoreceptor cells, R1 and R6, and primary pigment cells for normal eye development. *Genes Dev*. 1992;6(1):50–60. <https://doi.org/10.1101/gad.6.1.50> PMID: [1346120](https://pubmed.ncbi.nlm.nih.gov/1346120/)
24. Hayashi T, Kojima T, Saigo K. Specification of primary pigment cell and outer photoreceptor fates by BarH1 homeobox gene in the developing *Drosophila* eye. *Dev Biol*. 1998;200(2):131–45. <https://doi.org/10.1006/dbio.1998.8959> PMID: [9705222](https://pubmed.ncbi.nlm.nih.gov/9705222/)
25. Woronik A, Tunström K, Perry MW, Neethiraj R, Stefanescu C, Celorio-Mancera MP et al. A transposable element insertion is associated with an alternative life-history strategy. *Nat Commun*. 2019;10:5757. <https://doi.org/10.1038/s41467-019-13596-2>
26. Córdova-Fletes C, Domínguez MG, Delint-Ramirez I, Martínez-Rodríguez HG, Rivas-Estilla AM, Barros-Núñez P, et al. A de novo t(10;19)(q22.3;q13.33) leads to ZMIZ1/PRR12 reciprocal fusion transcripts in a girl with intellectual disability and neuropsychiatric alterations. *Neurogenetics*. 2015;16(4):287–98. <https://doi.org/10.1007/s10048-015-0452-2> PMID: [26163108](https://pubmed.ncbi.nlm.nih.gov/26163108/)
27. Yanicostas C, Vincent A, Lepesant JA. Transcriptional and posttranscriptional regulation contributes to the sex-regulated expression of two sequence-related genes at the janus locus of *Drosophila melanogaster*. *Mol Cell Biol*. 1989;9(6):2526–35. <https://doi.org/10.1128/mcb.9.6.2526-2535.1989> PMID: [2503707](https://pubmed.ncbi.nlm.nih.gov/2503707/)
28. Slatkin M. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nat Rev Genet*. 2008;9(6):477–85. <https://doi.org/10.1038/nrg2361>
29. Völker M, Backström N, Skinner BM, Langley EJ, Bunzey SK, Ellegren H, et al. Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. *Genome Res*. 2010;20(4):503–11. <https://doi.org/10.1101/gr.103663.109> PMID: [20357050](https://pubmed.ncbi.nlm.nih.gov/20357050/)
30. Morgan AP, Gatti DM, Najarian ML, Keane TM, Galante RJ, Pack AI, et al. Structural Variation Shapes the Landscape of Recombination in Mouse. *Genetics*. 2017;206(2):603–19. <https://doi.org/10.1534/genetics.116.197988> PMID: [28592499](https://pubmed.ncbi.nlm.nih.gov/28592499/)
31. Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. *Genet Res*. 1974;23(1):23–35. <https://doi.org/10.1017/s0016672300014634> PMID: [4407212](https://pubmed.ncbi.nlm.nih.gov/4407212/)
32. Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK. Recombination: an underappreciated factor in the evolution of plant genomes. *Nat Rev Genet*. 2007;8(1):77–84. <https://doi.org/10.1038/nrg1970> PMID: [17173059](https://pubmed.ncbi.nlm.nih.gov/17173059/)
33. Cisneros AF, Gagnon-Arsenault I, Dubé AK, Després PC, Kumar P, Lafontaine K, et al. Epistasis between promoter activity and coding mutations shapes gene evolvability. *Sci Adv*. 2023;9(5):eadd9109. <https://doi.org/10.1126/sciadv.add9109> PMID: [36735790](https://pubmed.ncbi.nlm.nih.gov/36735790/)
34. Mileyko Y, Joh RI, Weitz JS. Small-scale copy number variation and large-scale changes in gene expression. *Proc Natl Acad Sci U S A*. 2008;105(43):16659–64. <https://doi.org/10.1073/pnas.0806239105> PMID: [18946033](https://pubmed.ncbi.nlm.nih.gov/18946033/)
35. Nourmohammad A, Lässig M. Formation of regulatory modules by local sequence duplication. *PLoS Comput Biol*. 2011;7(10):e1002167. <https://doi.org/10.1371/journal.pcbi.1002167> PMID: [21998564](https://pubmed.ncbi.nlm.nih.gov/21998564/)
36. Martin A, Orgogozo V. The Loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution*. 2013;67(5):1235–50. <https://doi.org/10.1111/evo.12081> PMID: [23617905](https://pubmed.ncbi.nlm.nih.gov/23617905/)
37. Voordeckers K, Pougach K, Verstrepen KJ. How do regulatory networks evolve and expand throughout evolution?. *Curr Opin Biotechnol*. 2015;34:180–8. <https://doi.org/10.1016/j.copbio.2015.02.001> PMID: [25723843](https://pubmed.ncbi.nlm.nih.gov/25723843/)
38. Michez D, Rasmont P, Terzo M, Vereecken NJ. A synthesis of gynandromorphy among wild bees (Hymenoptera: Apoidea), with an annotated description of several new cases. *Annales de la Société entomologique de France (NS)*. 2009;45(3):365–75. <https://doi.org/10.1080/00379271.2009.10697621>

39. Lucia M, Gonzalez VH. A New Gynandromorph of *Xylocopa frontalis* with a Review of Gynandromorphism in *Xylocopa* (Hymenoptera: Apidae: Xylocopini). *Annals of the Entomological Society of America*. 2013;106(6):853–6. <https://doi.org/10.1603/an13085>
40. Almeida RPS, Leite LAR, Ramos KDS. Two new records of Gynandromorphs in *Xylocopa* (Hymenoptera, Apidae s.l.). *Pap Avulsos Zool*. 2018;58:17. <https://doi.org/10.11606/1807-0205/2018.58.17>
41. Kasperek M, Bonforte M, Catania R. Male or Female? New Cases of Gynandromorphism in Wool Carder Bees of the Tribe Anthidiini (Hymenoptera: Megachilidae). *Sociobiology*. 2025;72(3):e11525. <https://doi.org/10.13102/sociobiology.v72i3.11525>
42. Ugajin A, Matsuo K, Kubo R, Sasaki T, Ono M. Expression profile of the sex determination gene *doublesex* in a gynandromorph of bumblebee, *Bombus ignitus*. *Naturwissenschaften*. 2016;103(3–4):17. <https://doi.org/10.1007/s00114-016-1342-7> PMID: 26868001
43. Cook JM. Sex determination in the Hymenoptera: a review of models and evidence. *Heredity*. 1993;71(4):421–35. <https://doi.org/10.1038/hdy.1993.157>
44. Sturtevant AH. The claret mutant type of *Drosophila simulans*: A study of chromosome elimination and of cell-lineage. *Z wiss Zool*. 1929;135:323–56. <https://doi.org/10.1093/bioinformatics/btw024> PMID: 26794315
45. Matsuo K, Kubo R, Sasaki T, Ono M, Ugajin A. Scientific note on interrupted sexual behavior to virgin queens and expression of male courtship-related gene fruitless in a gynandromorph of bumblebee, *Bombus ignitus*. *Apidologie*. 2018;49(3):411–4. <https://doi.org/10.1007/s13592-018-0568-0>
46. Bird SA, Pope NS, McGrady CM, Fleischer SJ, López-Urbe MM. Mating frequency estimation and its importance for colony abundance analyses in eusocial pollinators: a case study of *Bombus impatiens* (Hymenoptera: Apidae). *J Econ Entomol*. 2024;117(5):1712–22. <https://doi.org/10.1093/jee/toae178> PMID: 39137237
47. Carolan JC, Murray TE, Fitzpatrick Ú, Crossley J, Schmidt H, Cederberg B, et al. Colour patterns do not diagnose species: quantitative evaluation of a DNA barcoded cryptic bumblebee complex. *PLoS One*. 2012;7(1):e29251. <https://doi.org/10.1371/journal.pone.0029251> PMID: 22238595
48. Hines HM, Williams PH. Mimetic colour pattern evolution in the highly polymorphic *Bombus trifasciatus* (Hymenoptera: Apidae) species complex and its comimics. *Zool J Linn Soc*. 2012;166(4):805–26.
49. Koch JB, Rodriguez J, Pitts JP, Strange JP. Phylogeny and population genetic analyses reveals cryptic speciation in the *Bombus fervidus* species complex (Hymenoptera: Apidae). *PLoS One*. 2018;13(11):e0207080. <https://doi.org/10.1371/journal.pone.0207080> PMID: 30462683
50. Ghisbain G, Lozier JD, Rahman SR, Ezray BD, Tian L, Ulmer JM, et al. Substantial genetic divergence and lack of recent gene flow support cryptic speciation in a colour polymorphic bumble bee (*Bombus bifarius*) species complex. *Systematic Entomology*. 2020;45(3):635–52. <https://doi.org/10.1111/syen.12419>
51. Owen RE, Plowright RC. Inheritance of metasomal pile colour variation in the bumble bee *Bombus rufocinctus* Cresson (Hymenoptera: Apidae). *Can J Zool*. 1988;66(5):1172–8. <https://doi.org/10.1139/z88-171>
52. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010.
53. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047–8. <https://doi.org/10.1093/bioinformatics/btw354> PMID: 27312411
54. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
55. Eldem V, Çınar YU, Çay SB, Obut O, Kuralay SC, Balci MA, et al. De novo genome assembly and annotations of *Bombus lapidarius* and *Bombus niveatus* provide insights into the environmental adaptability. *Apidologie*. 2025;56(1). <https://doi.org/10.1007/s13592-024-01140-1>
56. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013. <https://doi.org/10.48550/arXiv.1303.3997>
57. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics*. 2015;31(12):2032–4. <https://doi.org/10.1093/bioinformatics/btv098> PMID: 25697820
58. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303. <https://doi.org/10.1101/gr.107524.110> PMID: 20644199
59. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43(1110):11.10.1-11.10.33. <https://doi.org/10.1002/0471250953.bi1110s43> PMID: 25431634
60. Van der Auwera GA, O'Connor BD. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. 1st ed. O'Reilly Media. 2020.
61. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987–93. <https://doi.org/10.1093/bioinformatics/btr509> PMID: 21903627
62. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75. <https://doi.org/10.1086/519795> PMID: 17701901
63. Turner SD. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *Biorxiv*. 2014. <https://doi.org/10.1101/005165>
64. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. *JOSS*. 2019;4(43):1686. <https://doi.org/10.21105/joss.01686>
65. Dimas AS, Stranger BE, Beazley C, Finn RD, Ingle CE, Forrest MS, et al. Modifier effects between regulatory and protein-coding variation. *PLoS Genet*. 2008;4(10):e1000244. <https://doi.org/10.1371/journal.pgen.1000244> PMID: 18974877

66. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6(2):80–92. <https://doi.org/10.4161/fly.19695> PMID: [22728672](https://pubmed.ncbi.nlm.nih.gov/22728672/)
67. Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinformatics*. 2021;37(12):1639–43. <https://doi.org/10.1093/bioinformatics/btaa1016> PMID: [33320174](https://pubmed.ncbi.nlm.nih.gov/33320174/)
68. Sayers EW, Beck J, Bolton EE, Brister JR, Chan J, Connor R, et al. Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Res*. 2025;53(D1):D20–9. <https://doi.org/10.1093/nar/gkae979> PMID: [39526373](https://pubmed.ncbi.nlm.nih.gov/39526373/)
69. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant Review with the Integrative Genomics Viewer. *Cancer Res*. 2017;77(21):e31–4. <https://doi.org/10.1158/0008-5472.CAN-17-0337> PMID: [29092934](https://pubmed.ncbi.nlm.nih.gov/29092934/)
70. Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. *Bioinformatics*. 2007;23(10):1289–91. <https://doi.org/10.1093/bioinformatics/btm091> PMID: [17379693](https://pubmed.ncbi.nlm.nih.gov/17379693/)
71. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3--new capabilities and interfaces. *Nucleic Acids Res*. 2012;40(15):e115. <https://doi.org/10.1093/nar/gks596> PMID: [22730293](https://pubmed.ncbi.nlm.nih.gov/22730293/)
72. Kõressaar T, Lepamets M, Kaplinski L, Raime K, Andreson R, Remm M. Primer3_masker: integrating masking of template sequence with primer design software. *Bioinformatics*. 2018;34(11):1937–8. <https://doi.org/10.1093/bioinformatics/bty036> PMID: [29360956](https://pubmed.ncbi.nlm.nih.gov/29360956/)
73. Pribelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. *Curr Protoc Bioinformatics*. 2020;70(1):e102. <https://doi.org/10.1002/cpbi.102> PMID: [32559359](https://pubmed.ncbi.nlm.nih.gov/32559359/)
74. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25(15):1972–3. <https://doi.org/10.1093/bioinformatics/btp348> PMID: [19505945](https://pubmed.ncbi.nlm.nih.gov/19505945/)
75. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268–74. <https://doi.org/10.1093/molbev/msu300> PMID: [25371430](https://pubmed.ncbi.nlm.nih.gov/25371430/)
76. Pagès H, Aboyou P, Gentleman R, DebRoy S. Biostrings: Efficient manipulation of biological strings. *Bioconductor*. 2022.
77. Tan G, Lenhard B. TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinformatics*. 2016;32(10):1555–6. <https://doi.org/10.1093/bioinformatics/btw024> PMID: [26794315](https://pubmed.ncbi.nlm.nih.gov/26794315/)
78. Baranasic D. JASPAR2020: Data package for JASPAR database (version 2020). R package. 2020. <http://jaspar.genereg.net/>
79. Crowley LM, Sivell O, Sivell D, University of Oxford and Wytham Woods Genome Acquisition Lab, Natural History Museum Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, et al. The genome sequence of the Buff-tailed Bumblebee, *Bombus terrestris* (Linnaeus, 1758). *Wellcome Open Res*. 2023;8:161. <https://doi.org/10.12688/wellcomeopenres.19248.1> PMID: [38283327](https://pubmed.ncbi.nlm.nih.gov/38283327/)
80. Wickham H, François R, Henry L, Müller K, Vaughan D. dplyr: A grammar of data manipulation. 2023.
81. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag. 2016.
82. Wickham H, Vaughan D, Girlich M. tidy: Tidy messy data. 2025.
83. Vences M, Patmanidis S, Schmidt J-C, Matschiner M, Miralles A, Renner SS. Hapsolutely: a user-friendly tool integrating haplotype phasing, network construction, and haploweb calculation. *Bioinform Adv*. 2024;4(1):vbae083. <https://doi.org/10.1093/bioadv/vbae083> PMID: [38895561](https://pubmed.ncbi.nlm.nih.gov/38895561/)
84. Knaus BJ, Grünwald NJ. vcfr: a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour*. 2017;17(1):44–53. <https://doi.org/10.1111/1755-0998.12549> PMID: [27401132](https://pubmed.ncbi.nlm.nih.gov/27401132/)