Materials and Methods

Culex pipiens Population Genomics Project

This study is one of two flagship studies associated with the Culex pipiens Population Genomic Project, also known as PipPop. Both studies make use of 840 individual whole-genome sequences of Culex pipiens complex mosquitoes (Cx. pipiens sensu lato) and outgroups (Table S1). Within the complex, we specifically targeted Cx. pipiens s. s. Linnaeus, 1758 and hybrids (n=688), but also sequenced smaller numbers of Cx. quinquefasciatus Say, 1823 (n=101), Cx. pallens Coquillet, 1898 (n=33), and Cx. australicus Dobrotworsky & Drummond, 1953 (n=5). Cx. torrentium Martini, 1925 (n=9) was included as an outgroup, and a handful of sequenced mosquitoes were inferred to belong to more distant, unknown taxa (n=4; Table S1). A total of 790 genomes were sequenced for PipPop, while 50 were previously published (40 from (61) and 10 from (62)). Full details on sampling of the 790 PipPop genomes, as well as variant calling and sample filtering for the full dataset, are provided in the companion study and summarized briefly here.

Mosquito collection and sequencing. We collected and sequenced 790 mosquitoes from 163 populations spread across 44 countries in the Americas, Europe, Africa, Asia, and Australia, targeting n ~ 5 individuals per population. 752 mosquitoes (95%) were collected from 2014–2021, and the remaining 38 (5%) were collected from 2003–2012. Mosquitoes were mostly sampled as adults and identified as Cx. pipiens s. l. or Cx. torrentium using standard metrics, but a subset came from dense larval pools. They were collected from both aboveground (87%) and belowground (13%) sites. Belowground sites included basements of residential buildings, manholes, stormwater drains, cesspits, subway systems, and underground floors of a parking garage. Aboveground sites spanned a variety of habitats, from dense urban environments to residential areas to natural parks. Detailed sample metadata, including individual and population IDs, GPS coordinates, collection date, life stage, sex, and trapping method can be found in Table S1.

Genomic DNA was extracted using the NucleoSpin 96 DNA RapidLyse kit (Macherey-Nagel, Germany). We confirmed that samples belonged to the Cx. pipiens complex or Cx. torrentium (outgroup) via a multiplex PCR targeting the ace-2 locus (63) and visual inspection of amplicon sizes on a gel. Samples with no bands or unexpected band sizes were excluded. DNA sequencing libraries were prepared using Illumina DNA Prep Kits (Illumina, USA) with custom dual-unique barcodes. Approximately 80 barcoded libraries were pooled and sequenced on individual S4 lanes of a Novaseq 6000 PE150 sequencer (Illumina, USA), with a target genome-wide coverage of 10-15X. One pool including Mediterranean and Middle Eastern mosquitoes was sequenced across four lanes (a full S4 flow cell) to achieve higher coverage (~60X) for use in cross-coalescence analyses.

Read processing and mapping. Raw reads were assessed for quality using FastQC v.0.11.8 (64), and low-quality bases and adapters were trimmed using Trimmomatic (65). Trimmed reads were mapped onto the recently updated, chromosome-scale CpipJ5 assembly (62). We used BWA-MEM v.0.7.17 (66) to map the reads with default settings and identified and removed optical and PCR duplicates with Picard MarkDuplicates v.2.20.2 (67). We then used GATK v.3.8 (68) to perform local realignment around small insertions and deletions. We calculated genome-wide coverage after deduplication using Mosdepth v.0.3.3 (69). We used the deduplicated, realigned reads for all the analyses below.
Accessible regions and variant calling. We used 100 high-coverage individuals (50 Cx. pipiens s. s. with >20X coverage and 50 Cx. quinquefasciatus with >10X coverage) to characterize variation in coverage across the genome and mask ‘inaccessible’ regions for variant calling. More specifically, we masked sites with <0.5X or >1.5X normalized coverage in either species. We also masked putative repeat elements regardless of coverage (62). This left us with ~131 million ‘accessible’ sites or approximately 23% of the 559 Mb genome. We then called single nucleotide variants in all 840 individuals using BCFtools v1.13 (70). Variant calling was parallelized across multiple 20 Mb chunks of the genome. In addition to masking the inaccessible sites and repeat elements described above, we also masked multiallelic SNPs, indels, and SNPs falling within 5 bp of indels. We calculated key statistics for each SNP and further removed those with QUAL<50, MQ<50, >10% individuals with missing genotypes, average mean depth across all samples of <10X or >30X, and alleles of GQ<20. These cutoffs were chosen after visual inspection of the distribution of each statistic, following GATK hard-filtering best practices for non-model species (71). After filtering, we were left with 30.6M high-quality, accessible, biallelic SNPs (of ~131M total accessible sites). This full SNP set was used for all analyses except where otherwise specified.

Individual filtering. We removed two samples from Raleigh, USA with < 2X coverage and >50% genotype missingness (RAL5, RAL6). We also filtered the full sample set for kin based on pairwise KING kinship coefficients computed in NgsRelate v.2.0 (72). The vast majority of pairs showed low relatedness as expected (mean kinship coefficient 0.00026). However, a subset of pairs showed higher values, including a subset of mosquitoes collected as larvae in the same pools. We identified all pairs with kinship > 0.09 and excluded the individual with lower coverage. We additionally excluded 3 individuals (PAR4, OSJi4, KAV5) that showed unexpectedly high relatedness to many individuals from other populations and one individual from Malaysia (MEL5) that clustered with North American samples. The unexpected relatedness of PAR4, OSJi4, and KAV5 to many other individuals could not be explained by their position in the 96-well plates used to process samples nor by low sequence coverage. While the Malaysian sample could conceivably be a migrant, we chose to remove it out of an abundance of caution.

Final sample set. After filtering low quality individuals and kin, we were left with data for 743 unrelated mosquitoes. A few analyses presented here address the full global sample. However, unless otherwise specified, this study focuses on the subset of 357 individuals collected in the Western Palearctic (Europe, North Africa, and western Asia).

Analysis of population structure

We conducted a principal component analysis (PCA) of variation among Western Palearctic individuals (n = 357; Fig. 2 and fig. S1). As excessive linkage disequilibrium (LD) among genetic markers can lead to PC(s) of LD structure rather than population structure (73), we used Plink v.1.90 (74) to select a subset of 503,921 unlinked SNPs (--indep-pairwise 200 20 0.2). We then used PCAngsd v.1.10 (75) to estimate a covariance matrix and the princomp function in the R package stats v.3.6.2 to conduct the PCA (76).

Sequencing and analysis of historical specimens from London

To understand the relationship between historical and contemporary molestus populations, we extracted genomic DNA from 22 pinned Culex specimens in the National History Museum, London
(Table S2) using a recently published, minimally destructive protocol (33). Briefly, pinned specimens were removed from the main label pins and put in a styrofoam box filled with wet paper towels for rehydration at 37 °C for 3 hours. Each rehydrated sample was then dipped in 200 µl of Lysis Buffer C (200 mM Tris, 25 mM EDTA, 0.05% Tween-20, and 0.4 mg/ml Proteinase K) and incubated at 37 °C for 2 hours. Genomic DNA in the lysis buffer was then purified using a modified MinElute (Qiagen) silica column approach. After extraction, intact mosquito specimens were rinsed in increasing percentages of ethanol (30% and 50%) and sent back to the museum for critical point drying. Libraries of the purified genomic DNA were created using NEB Next Ultra II DNA Library Prep Kit (New England Biolabs) with no shearing and then purified using 2.2x SPRI (Beckman Coulter Agencourt AMPure XP) beads post library ligation and two times 1x SPRI post PCR amplification using a KAPA HiFi HotStart Uracil+ ReadyMix PCR Kit. The final libraries were sequenced on one lane of NovaSeq PE75 (Illumina).

Raw reads were run though the ancient DNA pipeline EAGER (77), with the following processing parameters: trimming adapter sequence, trimming bases of quality score < 20, removing sequences shorter than 30 bp, merging overlapping paired reads (with default minimum 11 bp overlap), aligning to the CpipJ5 assembly (62) using BWA-MEM, removing PCR duplicates and unaligned reads for final BAM files, and performing DamageProfiler to summarize ancient DNA characteristics (50 C > T and 30 G > A substitutions, read length in base pairs). We calculated genome-wide coverage after deduplication using Mosdepth (69) (mean = 5.77X, range = 1.05–9.22X). We used ANGSD v0.936 (78) to call genotype likelihoods (angsd -GL 1, SAMtools model) for the historical samples at the subset of 503,921 unlinked, biallelic SNPs used for PCA of contemporary genomes (see above). We then merged these samples with the contemporary Western Palearctic sample and conducted a joint PCA as described above (PCAngsd followed by princomp).

**Analysis of latitudinal gradient**

We modeled each *pipiens* population in the Western Palearctic as a mix of northern *pipiens* and molestus populations using genome-wide f3 statistics (Fig. 3A–C). Specifically, we used the threepop function in Treemix v1.13 (79) to calculate $f3(X; pipiens, molestus)$, where $X$ represents a focal *pipiens* population, and the *pipiens* and molestus reference populations came from Sweden (SWE) and Belgium (BVR), respectively. We used a block jackknife approach to obtain the standard error and compute Z-scores, dividing the genome into blocks of 500 SNPs (-k 500). A Z-score of −3 was used as a significance threshold (79).

We also estimated the number of derived alleles shared by focal *pipiens* populations with the same northern *pipiens* and molestus reference populations using Dsuite Dtrios (80), with *Cx. torrentium* as an outgroup. We specifically calculated the number of derived alleles shared with molestus as a fraction of those shared with either *pipiens* or molestus: $n(ABBA) / (n(ABBA) + n(BBAA))$ where A represents the ancestral allele and B represents the derived allele as in the tree shown in Fig. 3D.

**Distance (Dxy) tree inference**

We inferred a distance tree for *pipiens* and molestus mosquitoes with >10X genome wide coverage from the full global sample based on the number of pairwise nucleotide differences (Dxy). Sequenced mosquitoes from early branching *Cx. pipiens* lineages native to southern Africa (named ‘juppi’ and ‘mada’), as well as *Cx. torrentium* were included as outgroups, but *Cx. quinquefasciatus* was excluded.
As hybridization can confound relationships in distance trees, we used a variety of methods to identify and exclude populations or individuals that showed signs of admixture. Using f3 tests we found that no molestus populations were well modeled as a mixture of pипiens and molestus, suggesting that introgression from pипiens into molestus is generally rare (fig. S3). However, a more sensitive 4-population test (Patterson’s D) found small yet significant signs of introgression into some Mediterranean molestus populations (fig. S4), which we then excluded from the tree. Identification of pипiens populations that have received genetic input from molestus is more challenging because introgression is confounded by the ancestral genetic gradient (Fig. 3). To overcome this, we used the F-branch statistics (43) as presented in Fig. 5A (see Quantifying gene flow from molestus into pипiens for details) and then excluded all pипiens populations that showed non-zero introgression. Finally, we excluded any pипiens or molestus individual with >2% inferred ancestry from sibling species Cx. quinquiesquis based on an NGSadmix (75) analysis presented in the companion study (Table S1 column X). Such introgression is rare in the Western Palearctic (81), but extremely common in the Americas, leading to exclusion of most American samples. After filtering, we moved forward with Dxy tree inference for 99 molestus, 96 pипiens, and 10 individuals from outgroups.

We used pixy v.1.2.7 (82) to estimate pairwise genome-wide Dxy among the remaining samples. We included invariant accessible sites in addition to the full set of 30.6M biallelic SNPs as exclusion of invariant sites is known to generate bias (82). We bootstrapped genome-wide Dxy estimates 100 times by sampling 1Mb windows with replacement. We then built the genome-wide neighbor-joining tree as well as bootstrapped trees based on the resulting matrices of Dxy values using the R packages ape v.5.6.2 (83) and ggtree v3.6.2 (84).

We annotated populations in the tree based on microhabitat of origin—aboveground, belowground, or “suspected belowground”. Suspected belowground populations included one Belgian population (BVR) and one Chinese population (BEJ). The Belgian individuals were collected aboveground in a heavily industrialized zone and suspected of having escaped from a nearby tire factory. The Chinese individuals were collected trying to bite the collector inside a residential building in Tangshan, near Beijing.

Genetic diversity (π)

We calculated genome-wide nucleotide diversity (π) for all molestus populations included in the Dxy tree analysis using pixy v.1.2.7 (82). A potential concern in doing so was that the eastern Mediterranean molestus populations, including key populations from Egypt and Israel, might have experienced introgression from Cx. quinquiesquis below the 2% threshold we used for exclusion from the tree (81). Even a small amount of introgression from the divergent Cx. quinquiesquis could inflate diversity estimates. To identify putatively introgressed genomic regions, we used Dsuite Dtribos to calculate f4 admixture ratios in non-overlapping windows of fixed size (50kb, 150kb, 250kb, 500kb, 1Mb) using the following tree: (((pipiens, X), quinquiesquis), outgroup). Reference pipiens and quinquiesquis populations came from Sweden (SWE) and Saudi Arabia (JED). Cx. torrentium was used as the outgroup. The vast majority of windows in most individuals showed 0 introgression, but we observed a minor peak at f4 ~ 0.5 in some samples (fig. S6), likely representing the heterozygous state for introgressed haplotypes. Homozygous quinquiesquis haplotypes (f4 ~ 1) were also sometimes present, but extremely rare. After comparing signal to noise ratios, we settled on a window size of 150kb and an f4 cutoff of 0.2 for calling introgression (fig. S6). When computing diversity (π), we
masked every 150kb locus for which any of the 99 molestus individuals showed significant introgression from quinquefasciatus. In total, we masked ~5% of all 30.6M accessible sites.

**Cross-coalescent analysis of pipiens-molestus split time**

To estimate the divergence time between pipiens and molestus, we carried out cross-coalescent analyses using MSMC2 following published best practices (39, 85). As MSMC2 requires phased genomes, we assembled a genome phasing panel using 551 individuals with >10X coverage that represent all major geographic regions where pipiens and molestus occur (mean coverage = 19.6X, range = 10–87.3X). We considered the full set of 30.6M biallelic SNPs but further filtered out genotypes with DP < 8. We first individually phased nearby heterozygous sites based on information present in sequencing reads using HAPCUT2 (86). This read-based phasing alone was able to phase up to ~90% of variants in the highest-coverage samples (range = 0.8–90.2%, median = 22.9%). We then carried out statistical phasing with the pre-phased variants across all individuals using SHAPEIT4 v2.2 (87) with a phase set error rate of 0.0001. To increase accuracy, we increased the MCMC iterations in SHAPEIT4 from the default value of 15 to 27 (--mcmc-iterations 10b + 1p + 1b + 1p + 1b + 1p + 1b + 1p + 10m), and we increased PBWT depth from the default value of 4 to 8. We phased variants on each chromosome separately.

We selected two high-coverage individuals from an Egyptian molestus population (ADR, 47.5X and 56X coverage) and another two from a Moroccan pipiens population (MAK, 56.1X and 67.5X). We first extracted phased genomes of focal individuals using BCFtools and generated chromosome-specific masks based on average coverage using bamCaller.py (85). We also masked every 150kb locus at which the individuals showed signs of introgression from quinquefasciatus (see above, f4 > 0.2; fig. S6). We then ran MSMC2 to characterize rates of cross-coalescence within and between the two populations. The time at which the relative rate of cross-coalescence exceeded 50% was used as a point estimate of the split time (39). We bootstrapped MSMC2 analyses using 100 replicates of three 200 Mb ‘chromosomes’, each composed of resampled blocks of 10 Mb. To explore the robustness of our results to sample selection, we reran the analysis with an alternative Mediterranean pipiens population for which high-coverage genomes were available (MEG, Armenia; 50.8X and 20.7X) (fig. S7).

MSMC2 generates split time estimates in coalescent units (85), which can be converted to years given a taxon-specific mutation rate ($\mu$) and generation time ($g$). Since $\mu$ and $g$ have not been directly measured in natural Cx. pipiens populations, we used plausible, literature-based, ‘best-guess’ values, as well as biologically reasonable minima and maxima. For $\mu$, we considered published data from mosquitoes and other insects and set the reasonable range at 1.0–8.0×10$^{-9}$ (88–90). Our best-guess of $\mu$ was 4.85×10$^{-9}$, taken from a recent estimate in Ae. aegypti, a well-studied mosquito from the same subfamily (55). For $g$, our best-guess was 20 days, based on a study of an autogenous molestus lab colony (20–21.3 days) (91). However, lab conditions are often better than those found in nature (e.g., unlimited food) and pipiens mosquitoes might be delayed in finding bloodmeals. We therefore extended the reasonable range up to 30 days. Taken together, we used the following combinations of parameters for conversion of coalescent units to our best-guess, minimum, and maximum chronological split times (Fig. 4D): $\mu = 4.85 \times 10^{-9}$ and $g = 20$ (best-guess split time), $\mu = 8 \times 10^{-9}$ and $g = 20$ (minimum split time), $\mu = 1 \times 10^{-9}$ and $g = 30$ (maximum split time).
Quantifying gene flow from *molestus* into *pipiens*

To quantify gene flow from *molestus* into *pipiens* across the Western Palearctic while accounting for the ancestral genetic gradient, we used Dsuite Fbranch to calculate branch-specific f4 admixture ratios (43) (Fig. 5). We specified the tree shown in Fig. 5B and estimated gene flow into focal *pipiens* populations from the three other branches. Latitudinally varying gene flow from a northern *pipiens* population (SWE, Sweden, arrow 1) accounted for the ancestral gradient. Gene flow from a Middle Eastern *molestus* population (ADR, Egypt, arrow 2) and a northern *molestus* population (BVR, Belgium, arrow 3) allowed us to isolate genetic input from *molestus* subsequent to the split with *pipiens*.

We used a linear modeling framework to explore a potential association between *molestus* gene flow (Fig. 5D) and human population density (a proxy for urbanization). We first downloaded 30-second resolution population density data from the Gridded Population of the World v4 (92) and compared the effect of density on introgression when averaging density within circles of the following radii (centered around collection sites): 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, and 100 km. In a simple linear regression excluding three outlier populations (PAR, LND, FON, Cook’s distance > 4), human density had a significant effect using radii of 1–10 km, but not across larger distances (Fig. 5E). The model with human density averaged across a 3 km buffer explained the most variance ($R^2 = 0.21$) and was used in the analysis shown in Fig. 5F. We also asked whether climate could explain additional variance in *molestus* introgression across populations by adding WorldClim2 bioclimatic variables (Bio1–19) (93) to the human density only model one at a time, again in a linear modeling framework. None of the bioclimatic variables significantly improved the model. Bio8 (mean temperature of the wettest quarter) was the only variable that had a marginal effect (linear model $P = 0.09$).
**Supplementary Figures**

**A**

Fig. S1. Location of phenotyped and historical samples in PCA analyses. (A) Individuals with known phenotypes highlighted in PC1 x latitude plot for contemporary Western Palearctic samples (same analysis as Fig. 2C; n = 357). The autogenous samples (able to lay eggs without a blood meal) correspond to lab strains derived from belowground sites in Amsterdam (~52 N°) and Athens (~38 N°), as well as field-collected samples from a flooded area on the belowground floor of a parking complex in Paris (~49 N°). The anautogenous samples (require a blood meal for egg development) came from various outdoor locations in Paris (~49 N°; cemetery, school, and hospital). The human-biting mosquitoes were collected trying to bite the collector in a residence in Rome (~42 N°). (B) Combined
analysis of genetic variation across contemporary Western Palearctic samples ($n = 357$) and historical London samples ($n = 22$, see also Fig. 2C inset). Historical samples were collected between 1940–1985 and archived in the Natural History Museum in London. See Table S1 and S2 for detailed sample metadata.
Figure S2. Second major axis of genetic variation across Western Palearctic samples is correlated with longitude. 

(A) Variance explained by the first 10 principal components in PCA of all Western Palearctic samples (Fig. 2B; \( n = 357 \) individuals). 

(B) Plot of PC2 vs Longitude showing a strong and significant correlation (Pearson’s correlation test, \( R = 0.80, P = 7.8 \times 10^{-82} \)). North African samples are separated from others at the same longitude, most likely due to limited gene flow across the Mediterranean Sea.
Figure S3. *f3* statistics for *molestus* populations modeled as a mixture of *pipiens* and *molestus* reference populations from the north. Each population was modeled as a mixture of Swedish *pipiens* (SWE) and Belgian *molestus* (BVR). Y axis labels list country, 3-letter population code (see table S1), and sample size (number of individuals). Error bars indicate 95% block-jackknife confidence intervals. None of the populations showed significant signs of mixture (*f3* < 0). This analysis includes all global *molestus* samples that showed <2% *quinquefasciatus* ancestry, inferred via NGSadmix in a companion paper (Table S1, column X). Only a few samples outside the Western Palearctic met this criterion.
Figure S4. Patterson’s $D$ statistics testing *molestus* populations for signs of introgression from *pipiens*. Each population was modeled as sister to Belgian *molestus* (BVR) and receiving potential genetic input from Swedish *pipiens* (SWE). Y axis labels list country, 3-letter population code (see Table S1), and sample size (number of individuals). Error bars indicate 95% block-jackknife confidence intervals. Blue outlines highlight populations showing small, but significant, introgression from *pipiens* ($D>0$). These were excluded from the neighbor-joining Dxy tree in Fig. 4A and S5. This analysis includes all Western Palearctic *molestus* samples as well as any global *molestus* sample that showed <2% *quinquefasciatus* ancestry inferred via NGSadmix in a companion paper (Table S1, column X). Only a few samples outside the Western Palearctic met this criterion.
Figure S5. Neighbor-Joining tree based on whole genome $D_{xy}$, related to Figure 4A. Analysis only included individuals of ‘pure’ ancestry and >10X genome-wide coverage. $\text{pipiens}$ and $\text{molestus}$ individuals were deemed ‘pure’ if they showed no signs of introgression from either $\text{Cx. quinquefasciatus}$ or each other (see Materials and Methods for details). A zoom version of the $\text{molestus}$ clade appears in Figure 4A. $\text{Cx. torrentium}$ was used as an outgroup, with branch cut short for visualization. Each tip is labeled with country of origin and population code (table S1).
Figure S6. Identification of local *Cx. quinquefasciatus* introgression using *f4* ratio. We looked for stretches of introgression in the genomes of select *molestus* and *pipiens* samples so that they could be masked in analyses of genetic diversity (Fig. 4B) and cross-coalescence (Fig. 4C and S7). The figure shows example data from chromosome 2 of an Egyptian *molestus* individual (ADR1). *f4* was calculated in sliding windows of between 50kb (bottom) and 1MB (top). A, Full results for each window size. B, Summary histograms. Elevated *f4* values indicate *quinquefasciatus* introgression. Red outline and dashed lines show the chosen window size (150kb) and threshold (*f4* ratio = 0.2) chosen for masking.
Figure S7. Relative cross-coalescence (rCC) rate for Armenian *pipiens* (MEG) and Egyptian *molestus* (ADR) inferred by MSMC2, related to Figure 4D. The black line indicates the genome-wide result and gray lines indicate 100 bootstrap results. Chronological time along the x-axis corresponds to a scaling factor derived from best-guess values of the *de novo* mutation rate and generation time for *Cx. pipiens* (32). This scaling factor gives a split time (rCC = 50%) of ~2.2K years (red dashed line and arrowhead). Minimum and maximum scaling factors (derived from biologically reasonable bounds on the mutation rate and generation time (32) give split times of ~1.3K and 16.1K years (grey arrowheads).
Figure S8. Sources of introgression into *pipiens*, related to Figure 5. Full results from *f*-branch analysis used to assess three sources of introgression into focal *pipiens* populations: (A) northern *pipiens* (Sweden, SWE), (B) northern *molestus* (Belgium, BVR), and (C) Egyptian *molestus* (ADR). *pipiens* populations are ordered along the x-axis by latitude from south (left) to north (right).
### Table S1. Sample metadata. (Excel file)

Each row represents a mosquito sequenced for this study (n=790) or previously published (n=50, see column AA). Columns A–S contain basic collection information. Columns T and S list average sequence coverage (T) and whether or not the sample was included in the final set of 743 unrelated mosquitoes (U). Column V lists the species identity, which was usually inferred from genomic analysis but also cross-checked with information on male genitalia when available. Mosquitoes with ancestry from both *Cx. pipiens* and *Cx. quinquefasciatus* are listed as ‘hybrids’ if they showed 11–89% *Cx. quinquefasciatus* ancestry inferred using NGSadmix (column X, see below) and as one or the other ‘pure’ species otherwise. *Cx. pallens* mosquitoes from East Asia are listed as hybrids with either *Cx. quinquefasciatus* (or *Cx. pipiens*) if they showed more (or less) *Cx. quinquefasciatus* ancestry than expected given the species’ hybrid origin (cutoffs: >40% or <20% respectively). Column W lists the intraspecific *Cx. pipiens* lineage (*pipiens*, *molestus*, *juppi*, or *mada*) to which any *Cx. pipiens* or *Cx. pipiens* hybrid belongs. This inference is based on PCA and is only available for samples that passed the NGSrelate filter listed in column U. Columns X–Z show the results of an NGSadmix analysis of samples from outside sub-Saharan Africa with k=3 clusters. Note that the third cluster (Mediterranean *pipiens/molestus*) combines signatures of both *pipiens* and *molestus* ancestry and does not easily translate into assignment of a mosquito to the *pipiens* vs *molestus* lineage.

### Table S2. Museum specimens collected in London from 1940–1985. (Excel file)

Each row represents one mosquito individual. Column A shows sample ID. Column B shows collection year. Column C shows locality within London, if known. Columns D and E show GPS coordinates (Latitude and Longitude). Note the coordinates are approximate, inferred from the location names on the collection note. Column F shows additional collection notes, if any. Column G shows mean genome-wide coverage after de-duplication. Column H shows inferred ecotype of each sample based on the PCA with contemporary Western Palearctic samples (fig. S1B).
1. M. T. J. Johnson, J. Munshi-South, Evolution of life in urban environments. *Science* **358**, eaam8327 (2017).

2. H. Ritchie, V. Samborska, M. Roser, Urbanization. *Our world in data* (2024).

3. E. B. Vinogradova, *Culex Pipiens Pipiens Mosquitoes: Taxonomy, Distribution, Ecology, Physiology, Genetics, Applied Importance and Control* (Pensoft, Sofia, Bulgaria, 2000).

4. A. Farajollahi, D. M. Fonseca, L. D. Kramer, A. Marm Kilpatrick, “Bird biting” mosquitoes and human disease: a review of the role of Culex pipiens complex mosquitoes in epidemiology. *Infect. Genet. Evol.* **11**, 1577–1585 (2011).

5. R. Harbach, Culex pipiens: Species versus species complex – taxonomic history and perspective. *J. Am. Mosq. Control Assoc.* **28**, 10–23 (2012).

6. K. Byrne, R. A. Nichols, Culex pipiens in London Underground tunnels: differentiation between surface and subterranean populations. *Heredity (Edinb.)* **82** (Pt 1), 7–15 (1999).

7. A. P. Hendry, K. M. Gotanda, E. I. Svensson, Human influences on evolution, and the ecological and societal consequences. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**, 20160028 (2017).

8. K. A. Thompson, L. H. Rieseberg, D. Schluter, Speciation and the City. *Trends Ecol. Evol.* **33**, 815–826 (2018).

9. S. P. Otto, Adaptation, speciation and extinction in the Anthropocene. *Proc. Biol. Sci.* **285**, 20182047 (2018).

10. L. R. Rivkin, J. S. Santangelo, M. Alberti, M. F. J. Aronson, C. W. de Keyzer, S. E. Diamond, M.-J. Fortin, L. J. Frazee, A. J. Gorton, A. P. Hendry, Y. Liu, J. B. Losos, J. S. Maclvor, R. A. Martin, M. J. McDonnell, L. S. Miles, J. Munshi-South, R. W. Ness, A. E. M. Newman, M. R. Stothart, P. Theodorou, K. A. Thompson, B. C. Verrelli, A. Whitehead, K. M. Winchell, M. T. J. Johnson, A roadmap for urban evolutionary ecology. *Evol. Appl.* **12**, 384–398 (2019).

11. D. N. Reznick, J. Losos, J. Travis, From low to high gear: there has been a paradigm shift in our understanding of evolution. *Ecol. Lett.* **22**, 233–244 (2019).

12. P. G. Shute, Culex molestus. *Trans. R. Entomol. Soc. Lond.* **102**, 380–382 (1951).

13. C. Wesenberg-Lund, *Contributions to the Biology of the Danish Culicidae* (AF Høst & søn, 1920) vol. 8.

14. J. Legendre, Le moustique cavernicole ou l’adaptation de “Culex pipiens” a l’urbanisme moderne. *Bull. Acad. Med.* **106**, 86–89 (1931).

15. O. Hecht, Experimentelle beiträge zur biologie der stechmücken II. *Z. Angew. Entomol.* **19**, 579–607 (1932).

16. D. N. Reznick, *The “Origin” Then and Now: An Interpretive Guide to the “Origin of Species”* (Princeton University Press, Princeton, NJ, 2012; https://play.google.com/store/books/details?id=vBB6db03MrkC).

17. B. Nye, “Chapter 18: Mosquito in the Tube” in *Undeniable: Evolution and the Science of Creation*.
18. J. W. Bull, Humans artificially drive evolution of new species, *EurekAlert!* (2016). https://www.eurekalert.org/pub_releases/2016-06/nhmo-had062716.php.

19. E. Blakemore, The London Underground Has Its Own Mosquito Subspecies, *Smithsonian Magazine* (2016). https://www.smithsonianmag.com/smart-news/london-underground-has-its-own-mosquito-subspecies-180958566/.

20. K. Silver, Earth, *BBC* (2016). https://www.bbc.com/earth/story/20160323-the-unique-mosquito-that-lives-in-the-london-underground.

21. M. Schilthuizen, *Darwin Comes to Town: How the Urban Jungle Drives Evolution* (St Martin’s Press, New York, NY, 2019).

22. D. M. Fonseca, N. Keyghobadi, C. A. Malcolm, C. Mehmet, F. Schaffner, M. Mogi, R. C. Fleischer, R. C. Wilkerson, Emerging vectors in the Culex pipiens complex. *Science* **303**, 1535–1538 (2004).

23. Y. Haba, L. McBride, Origin and status of Culex pipiens mosquito ecotypes. *Curr. Biol.* **32**, R237–R246 (2022).

24. P. Forskål, *Descriptiones animalium, avium, amphibiorum, piscium, insectorum, vermium: quae in itinere orientali observavit* (ex officina Mölleri, 1775).

25. E. F. Germar, *Reise nach Dalmatien und in das Gebiet von Ragusa* (F. A. Brockhaus, 1817).

26. E. Ficalbi, Notizie preventive sulle zanzare Italiane. *Boll. Soc. Entomol. Ital.* **21**, 124–131 (1890).

27. K. M. Winchell, J. B. Losos, B. C. Verrelli, Urban evolutionary ecology brings exaptation back into focus. *Trends Ecol. Evol.* **38**, 719–726 (2023).

28. F. Villani, S. Urbanelli, A. Gad, S. Nudelman, L. Bullini, Electrophoretic variation of Culex pipiens from Egypt and Israel. *Biol. J. Linn. Soc. Lond.* **29**, 49–62 (1986).

29. L. R. Petersen, A. C. Brault, R. S. Nasci, West Nile virus: review of the literature. *JAMA* **310**, 308–315 (2013).

30. C. Giesen, Z. Herrador, B. Fernandez-Martinez, J. Figuerola, L. Gangoso, A. Vazquez, D. Gómez-Barroso, A systematic review of environmental factors related to WNV circulation in European and Mediterranean countries. *One Health* **16**, 100478 (2023).

31. M. L. Aardema, S. K. Olatunji, D. M. Fonseca, The enigmatic *Culex pipiens* (Diptera: Culicidae) species complex: Phylogenetic challenges and opportunities from a notoriously tricky mosquito group. *Ann. Entomol. Soc. Am.* **115**, 95–104 (2022).

32. Materials and methods are available as supplementary materials.

33. P. Korlević, E. McAlister, M. Mayho, A. Makunin, P. Flicek, M. K. N. Lawniczak, A minimally morphologically destructive approach for DNA retrieval and whole-genome shotgun sequencing of pinned historic dipteran vector species. *Genome Biol. Evol.* **13** (2021).

34. B. Gomes, C. A. Sousa, M. T. Novo, F. B. Freitas, R. Alves, A. R. Córte-Real, P. Salgueiro, M. J. Donnelly, A. P. G. Almeida, J. Pinto, Asymmetric introgression between sympatric molestus and pipiens forms of Culex pipiens (Diptera: Culicidae) in the Comporta region, Portugal. *BMC Evol.*
35. S. Urbanelli, R. Cianchi, V. Petrarca, M. Sabatinelli, M. Coluzzi, L. Bullini, “Adattamento all’ambiente urbano nella zanzara Culex pipiens (Diptera, Culicidae)” in *Ecologia Atti I Congressi Nazionali S. It. E.*, Moroni, A., Ravera, O. and Anelli, A., Ed. (Zara, 1981), pp. 305–316.

36. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).

37. S. Nudelman, R. Galun, U. Kitron, A. Spielman, Physiological characteristics of Culex pipiens populations in the Middle East. *Med. Vet. Entomol.* **2**, 161–169 (1988).

38. O. Olsson, “First states I: Mesopotamia and Egypt” in *Paleoeconomics* (Springer Nature Switzerland, Cham, 2024), pp. 285–308.

39. S. Schiffels, K. Wang, MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. *Methods Mol. Biol.* **2090**, 147–166 (2020).

40. S. Huang, G. L. Hamer, G. Molaei, E. D. Walker, T. L. Goldberg, U. D. Kitron, T. G. Andreadis, Genetic variation associated with mammalian feeding in Culex pipiens from a West Nile virus epidemic region in Chicago, Illinois. *Vector Borne Zoonotic Dis.* **9**, 637–642 (2009).

41. A. M. Kilpatrick, L. D. Kramer, M. J. Jones, P. P. Marra, P. Daszak, D. M. Fonseca, Genetic influences on mosquito feeding behavior and the emergence of zoonotic pathogens. *Am. J. Trop. Med. Hyg.* **77**, 667–671 (2007).

42. M. L. Fritz, E. D. Walker, J. R. Miller, D. W. Severson, I. Dworkin, Divergent host preferences of above- and below-ground Culex pipiens mosquitoes and their hybrid offspring. *Med. Vet. Entomol.* **29**, 115–123 (2015).

43. M. Malinsky, H. Svardal, A. M. Tyers, E. A. Miska, M. J. Genner, G. F. Turner, R. Durbin, Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat. Ecol. Evol.* **2**, 1940–1955 (2018).

44. M. Di Luca, L. Toma, D. Boccolini, F. Severini, G. La Rosa, G. Minelli, G. Bongiorno, F. Montarsi, D. Arnoldi, G. Capelli, A. Rizzoli, R. Romi, Ecological distribution and CQ11 genetic structure of Culex pipiens complex (Diptera: Culicidae) in Italy. *PLoS One* **11**, e0146476 (2016).

45. H. C. Osório, L. Zé-Zé, F. Amaro, A. Nunes, M. J. Alves, Sympatric occurrence of Culex pipiens (Diptera, Culicidae) biotypes pipiens, molestus and their hybrids in Portugal, Western Europe: feeding patterns and habitat determinants. *Med. Vet. Entomol.* **28**, 103–109 (2014).

46. J. Martínez-de la Puente, M. Ferraguti, S. Ruiz, D. Roiz, R. C. Soriguer, J. Figuerola, Culex pipiens forms and urbanization: effects on blood feeding sources and transmission of avian Plasmodium. *Malar. J.* **15**, 589 (2016).

47. Eurostat, Glossary: Urban centre (2018). https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Urban_centre.

48. O. Duron, J. Lagnel, M. Raymond, K. Bourtzis, P. Fort, M. Weill, Transposable element polymorphism of Wolbachia in the mosquito Culex pipiens: evidence of genetic diversity, superinfection and recombination. *Mol. Ecol.* **14**, 1561–1573 (2005).

49. E. Dumas, C. M. Atyame, P. Milesi, D. M. Fonseca, E. V. Shaivechiv, S. Unal, P. Makoundou, M. Weill, O. Duron, Population structure of Wolbachia and cytoplasmic introgression in a complex of mosquito species. *BMC Evol. Biol.* **13**, 181 (2013).
50. K. Knight, A. Malek, A morphological and biological study of Culex pipiens in the Cairo area of Egypt (Diptera-culicidae). *Bulletin de la Société Fouad 1er d'entomologie* **35**, 175–185 (1951).

51. A. Harpak, N. Garud, N. A. Rosenberg, D. A. Petrov, M. Combs, P. S. Pennings, J. Munshi-South, Genetic adaptation in New York City rats. *Genome Biol. Evol.* **13**, evaa247 (2021).

52. L. Weissbrod, F. B. Marshall, F. R. Valla, H. Khalaily, G. Bar-Oz, J.-C. Auffray, J.-D. Vigne, T. Cucchi, Origins of house mice in ecological niches created by settled hunter-gatherers in the Levant 15,000 y ago. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 4099–4104 (2017).

53. A. Wada-Katsumata, J. Silverman, C. Schal, Changes in taste neurons support the emergence of an adaptive behavior in cockroaches. *Science* **340**, 972–975 (2013).

54. M. Ravinet, T. O. Elgvin, C. Trier, M. Aliabadian, A. Gavrilo, G.-P. Sætre, Signatures of human-commensalism in the house sparrow genome. *Proc. Biol. Sci.* **285**, 20181246 (2018).

55. N. H. Rose, A. Badolo, M. Sylla, J. Akorli, S. Otoo, A. Gloria-Soria, J. R. Powell, B. J. White, J. E. Crawford, C. S. McBride, Dating the origin and spread of specialization on human hosts in Aedes aegypti mosquitoes. *Elife* **12** (2023).

56. J. A. Rioux, J. M. Pech, Le biotype autogene de Culex pipiens L. ne doit pas etre nomme Culex molestus Forskal (Diptera Culicidae). *Cahiers Des Nat.* **15**, 115–117 (1959).

57. C. M. Bahnck, D. M. Fonseca, Rapid assay to identify the two genetic forms of Culex (Culex) pipiens L. (Diptera: Culicidae) and hybrid populations. *Am. J. Trop. Med. Hyg.* **75**, 251–255 (2006).

58. E. Roubaud, Le pouvoir autogène chez le biotype nord-africain du moustique commun Culex pipiens (L.). *Bull. Soc. Pathol. Exot.* **36**, 172–175 (1939).

59. M. Naddaf, Mosquito-borne diseases are surging in Europe - how worried are scientists? *Nature* **633**, 749 (2024).

60. CDC, West Nile Virus Current Year Data, 2024, **CDC** (2024). https://www.cdc.gov/west-nile-virus/data-maps/current-year-data.html.

61. A. A. Yurchenko, R. A. Masri, N. V. Khrabrova, A. K. Sibataev, M. L. Fritz, M. V. Sharakhova, Genomic differentiation and intercontinental population structure of mosquito vectors Culex pipiens pipiens and Culex pipiens molestus. *Sci. Rep.* **10**, 7504 (2020).

62. S. S. Ryazansky, C. Chen, M. Potters, A. N. Naumenko, V. Lukyanovich, R. A. Masri, I. I. Brusentsov, D. A. Karagodin, A. A. Yurchenko, V. L. Dos Anjos, Y. Haba, N. H. Rose, J. Hoffman, R. Guo, T. Menna, M. Kelley, E. Ferrill, K. E. Schultz, Y. Qi, A. Sharma, S. Deschamps, V. Llaca, C. Mao, T. D. Murphy, E. M. Baricheva, S. Emrich, M. L. Fritz, J. B. Benoit, I. V. Sharakhova, C. S. McBride, Z. Tu, M. V. Sharakhova, The chromosome-scale genome assembly for the West Nile vector Culex quinquefasciatus uncovers patterns of genome evolution in mosquitoes. *BMC Biol.* **22**, 16 (2024).

63. J. L. Smith, D. M. Fonseca, Rapid assays for identification of members of the Culex (Culex) pipiens complex, their hybrids, and other sibling species (Diptera: culicidae). *Am. J. Trop. Med. Hyg.* **70**, 339–345 (2004).

64. S. Andrews, Others, FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom [Preprint] (2010).

65. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
66. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

67. Picard. http://broadinstitute.github.io/picard/.

68. G. van der Auwera, B. D. O'Connor, *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (O’Reilly Media, Sebastopol, CA, ed. 1, 2020).

69. B. S. Pedersen, A. R. Quinlan, Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).

70. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, Twelve years of SAMtools and BCFTools. *Gigascience* **10** (2021).

71. Hard-filtering germline short variants, GATK. https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants.

72. K. Hanghøj, I. Moltke, P. A. Andersen, A. Manica, T. S. Korneliussen, Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *Gigascience* **8** (2019).

73. A. Abdellaoui, J.-J. Hottenga, P. de Knijff, M. G. Nivard, X. Xiao, P. Scheet, A. Brooks, E. A. Ehli, Y. Hu, G. E. Davies, J. J. Hudziak, P. F. Sullivan, T. van Beijsterveldt, G. Willemsen, E. J. de Geus, B. W. J. H. Penninx, D. I. Boomsma, Population structure, migration, and diversifying selection in the Netherlands. *Eur. J. Hum. Genet.* **21**, 1277–1285 (2013).

74. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

75. L. Skotte, T. S. Korneliussen, A. Albrechtsen, Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).

76. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2020; https://www.R-project.org/).

77. J. A. Fellows Yates, T. C. Lamnidis, M. Borry, A. Andrades Valtueña, Z. Fagernäs, S. Clayton, M. U. Garcia, J. Neukamm, A. Peltzer, Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager. *PeerJ* **9**, e10947 (2021).

78. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).

79. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).

80. M. Malinsky, M. Matschiner, H. Svartdal, Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* **21**, 584–595 (2021).

81. E. V. Shaiekevich, E. B. Vinogradova, A. Bouattour, A. P. Gouveia de Almeida, Genetic diversity of Culex pipiens mosquitoes in distinct populations from Europe: contribution of Cx. quinquefasciatus in Mediterranean populations. *Parasit. Vectors* **9**, 47 (2016).

82. K. L. Korunes, K. Samuk, pixy: Unbiased estimation of nucleotide diversity and divergence in the
presence of missing data. *Mol. Ecol. Resour.* **21**, 1359–1368 (2021).

83. E. Paradis, K. Schliep, ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).

84. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T.-Y. Lam, Ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).

85. K. Wang, I. Mathieson, J. O’Connell, S. Schifflers, Tracking human population structure through time from whole genome sequences. *PLoS Genet.* **16**, e1008552 (2020).

86. P. Edge, V. Bafna, V. Bansal, HapCUT2: robust and accurate haplotype assembly for diverse sequencing technologies. *Genome Res.* **27**, 801–812 (2017).

87. O. Delaneau, J.-F. Zagury, M. R. Robinson, J. L. Marchini, E. T. Dermitzakis, Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* **10**, 5436 (2019).

88. P. D. Keightley, R. W. Ness, D. L. Halligan, P. R. Haddrill, Estimation of the spontaneous mutation rate per nucleotide site in a Drosophila melanogaster full-sib family. *Genetics* **196**, 313–320 (2014).

89. P. D. Keightley, A. Pinharanda, R. W. Ness, F. Simpson, K. K. Dasmahapatra, J. Mallet, J. W. Davey, C. D. Jiggins, Estimation of the spontaneous mutation rate in Heliconius melpomene. *Mol. Biol. Evol.* **32**, 239–243 (2015).

90. I. Rashid, M. Campos, T. Collier, M. Crepeau, A. Weakley, H. Gripkey, Y. Lee, H. Schmidt, G. C. Lanzaro, Spontaneous mutation rate estimates for the principal malaria vectors Anopheles coluzzii and Anopheles stephensi. *Sci. Rep.* **12**, 226 (2022).

91. Q. Gao, F. Su, Y.-B. Zhou, W. Chu, H. Cao, L.-L. Song, J.-J. Zhou, P.-E. Leng, Autogeny, Fecundity, and Other Life History Traits of Culex pipiens molestus (Diptera: Culicidae) in Shanghai, China. *J. Med. Entomol.* **56**, 656–664 (2019).

92. Center for International Earth Science Information Network-CIESIN - Columbia University, Gridded Population of the World, Version 4 (GPWv4): Population Density, Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC) (2016); https://doi.org/10.7927/H4NP22DQ.

93. S. E. Fick, R. J. Hijmans, WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas: NEW CLIMATE SURFACES FOR GLOBAL LAND AREAS. *Int. J. Climatol.* **37**, 4302–4315 (2017).