GBE

ERGA-BGE Genomes of *Culex laticinctus*, *Culex modestus*, *Culex perexiguus*, and *Culex theileri*: Unveiling the Genomes of the Key Vectors of West Nile Virus in the Mediterranean Basin

María José Ruiz-López [1,2,*], Sonia Cebrián-Camisón [1,2], Iker Irisarri [3], Sergio Magallanes [1,2], Josué Martínez-de la Puente [1,2], Jordi Figuerola [1,2], Robert M. Waterhouse [4], Wellcome Sanger Institute Tree of Life Management: Samples and Laboratory Team[5], Wellcome Sanger Institute Scientific Operations: Sequencing Operations[5], Wellcome Sanger Institute Tree of Life Core Informatics Team[5], Leanne Haggerty [6], Fergal Martin [6], Tom Brown [7,8]

[1]Departamento de Biología de la Conservación y Cambio Global, Estación Biológica de Doñana (EBD), CSIC, Seville 41092, Spain

[2]CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid 28029, Spain

[3]Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales-CSIC, Madrid 28006, Spain

[4]Environmental Bioinformatics Group, SIB Swiss Institute of Bioinformatics, Batiment Amphipole, UNIL-Sorge, Lausanne 1015, Switzerland

[5]Tree of Life, Wellcome Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK

[6]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK

[7]Department of Evolutionary Genetics, Leibniz Institute for Zoo and Wildlife Research, Berlin 10315, Germany

[8]Berlin Center for Genomics in Biodiversity Research (BeGenDiv), Berlin 14195, Germany

*Corresponding author: Email: mjruiz@ebd.csic.es.

Accepted: December 08, 2025

Abstract

The reference genomes of *Culex laticinctus*, *Culex modestus*, *Culex perexiguus*, and *Culex theileri* offer crucial resources for exploring the genetic basis of vector competence for different pathogens, including the emerging West Nile virus. Each genome was assembled into three contiguous chromosomal pseudomolecules. These chromosome-level assemblies encompass 834, 698, 436, and 934 Mb, respectively. They are composed of 3,284, 4,220, 1,917, and 2,704 contigs and 312, 979, 475, and 709 scaffolds, with contig N50 values of 434, 313, 473, and 770 kb, and scaffold N50 values of 292, 246, 150, and 281 Mb, respectively. These genomes offer an essential genomic resource for advancing our understanding of *Culex* evolution and population genomic analyses, which will help to implement sustainable mosquito control efforts and disease prevention.

Key words: *Culex*, Culicidae, mosquito genome, European reference genome atlas, biodiversity genomics Europe, Earth BioGenome project.

Significance Statement

Despite their medical and ecological importance, *Culex* mosquitoes are underrepresented in genomic resources. Here we present high-quality reference genomes for four new species, *Culex laticinctus*, *Culex perexiguus*, *Culex modestus*, and *Culex theileri*. These genomes will support research on mosquito evolution, vector competence, behavior, and insecticide resistance. They also provide a foundation for population-level genomic studies that will help unravel local adaptation patterns or inform sustainable control strategies. Expanding genomic coverage across *Culex* species is essential for uncovering evolutionary adaptations and enhancing long-term efforts in vector surveillance and management.

Introduction

The genus *Culex* comprises key disease vectors occurring worldwide. However, despite their relevance, the genomes of mosquito species from this genus are underrepresented in public databases. Here we present high-quality genomes of four species never sequenced before: *Culex (Culex) laticinctus*, *Culex (Barraudius) modestus*, *Culex (Culex) perexiguus*, and *Culex (Culex) theileri*.

Cx. (Culex) laticinctus (Edwards 1913) occurs mainly in Mediterranean and Afrotropical regions, with its range extending from Madeira and the Canary Islands east through the Mediterranean. The larvae inhabit mainly freshwater sources, both natural and anthropogenic (pools, wells, drums, fountains, troughs, etc.). *Cx. laticinctus* is more frequently found during the spring and summer months (April to October). Adult females mainly feed on birds and are not known to bite humans (Harbach 1988).

Cx. (Barraudius) modestus (Ficalbi, 1890) presents a Palearctic distribution. In Europe, it is commonly found in central and southern countries (Becker et al. 2010). The larvae occur in open areas such as meadows, vegetated freshwater, or slightly saline habitats, including rice fields and marshes in southern Europe; it is predominantly associated with saltwater marshes (Ribeiro et al. 1988) and rice fields. The larvae can be found from early spring to late autumn. Adults frequently feed on birds (Muñoz et al. 2012), and they rarely enter buildings, mainly biting humans outdoors.

Cx. (Culex) perexiguus (Theobald, 1903) belongs to the *Cx. univittatus* complex. Despite the morphological similarities between these species, they can be distinguished using DNA barcoding (Bravo-Barriga et al. 2023) or by the larvae and male genitalia (Becker et al. 2010). *Cx. perexiguus* distribution includes arid regions of northern and eastern Africa, south-west Asia and several countries of the Mediterranean basin (Harbach 1988, Snow and Ramsdale 1999). Although its larvae are occasionally found in anthropic habitats, they are more abundant in natural waterlogged areas with emergent vegetation that are clean to moderately polluted. The species is common from spring to autumn, although there might be some variation depending on the habitat. For example, in some areas like the lower Guadalquivir in southern Spain, the adult phase can be found throughout the year and be very abundant from July to October in association with rice cultivation (Figuerola et al. 2022). The adult females feed primarily on birds (Harbach 1988, Muñoz et al. 2012).

Cx. (Culex) theileri (Theobald 1903) is an Afrotropical and Palearctic species, present in the southern half of Europe, Africa, the Middle East, and parts of Asia (Becker et al. 2010). This polycyclic species occurs across a wide range of elevations, from lowlands to 1,000 to 3,000 meters in the Himalayas (Sirivanakarn 1976). It can breed in freshwater or slightly saline waters (Pires et al. 1982). Adult females are mammophilic and bite humans occasionally, mostly outdoors, but can sometimes enter buildings in large numbers (Gutsevich et al. 1974; Ramos et al. 1977).

All four species play important roles in transmitting wildlife and zoonotic diseases. *Cx. laticinctus*, traditionally considered of limited public health relevance (Ribeiro et al. 1988; Schaffner et al. 2001), has recently been found infected with West Nile virus (Figuerola unpublished data) and Sindbis virus (Gutiérrez-López et al. 2025). *Cx. modestus* is a known vector of Bunyaviruses (Lundström 1994), West Nile virus (Soto and Delang 2023), and Usutu virus (Soto et al. 2024) and has also been found naturally infected with Sindbis virus (Gutiérrez-López et al. 2025) and tularemia (Gutsevich et al. 1974). *Cx. perexiguus* transmits several flaviviruses, including West Nile and Usutu viruses, with a key role in virus amplification in southwestern Spain (Figuerola et al. 2022). It has also been found infected with Sindbis virus (Gutiérrez-López et al. 2025). Ornithophilic *Culex* species, including these, are frequently infected with avian malaria parasites such as *Plasmodium relictum*, one of the world's 100 most invasive species (Martínez-de la Puente et al. 2021). *Cx. theileri* has been found carrying West Nile and Sindbis viruses in South Africa (McIntosh et al. 1975), and its competence for West Nile virus has been recently confirmed (Burgas-Pau et al. 2025). It also transmits Rift Valley fever virus and *Dirofilaria* spp. in North Africa and Portugal (Smith 1973; Ribeiro et al. 1988).

Assembling high-quality reference genomes for *Cx. laticinctus*, *Cx. modestus*, *Cx. perexiguus*, and *Cx. theileri* is essential for advancing our understanding of these species' evolution and ecological role, aiding in implementing sustainable mosquito control efforts and disease prevention.

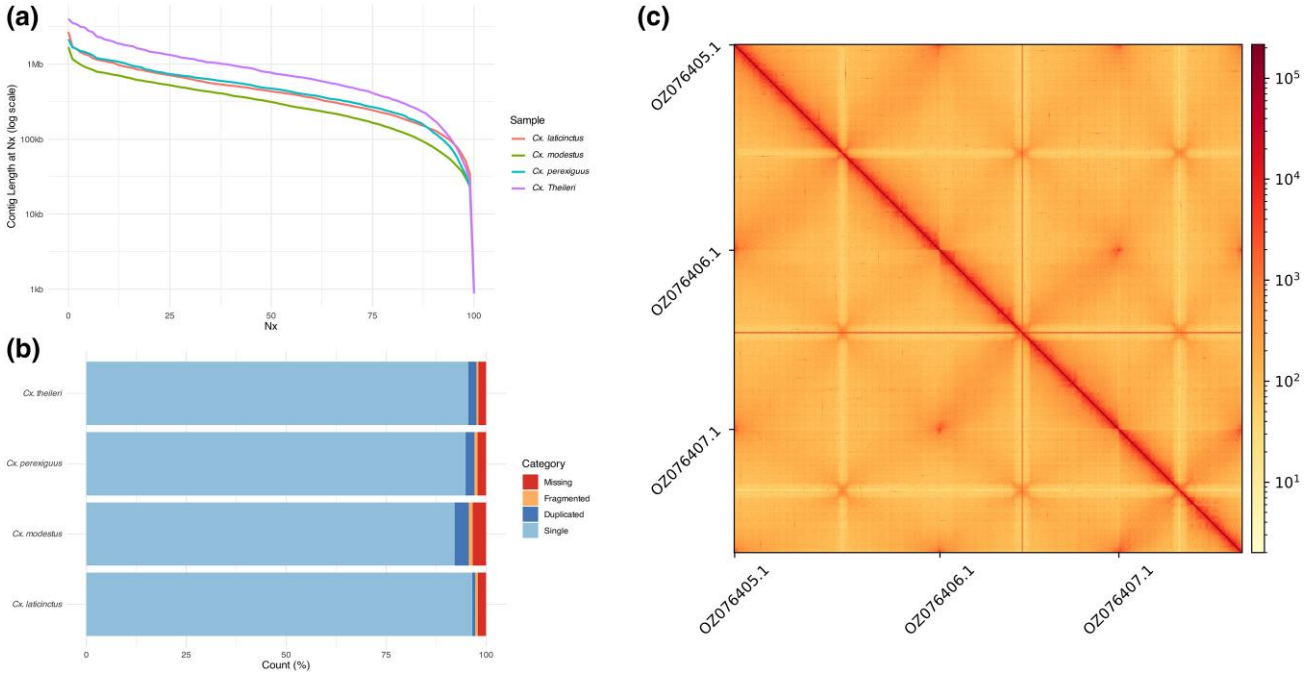

Fig. 1. Genome assembly quality. a) Contiguity of each *Culex* assembly. Shown are the lengths of each contig sorted from smallest to largest, displaying the proportion of the genome (Nx) made up of contigs of at least this length. b) Genome completeness as assessed by BUSCO using the diptera_odb10 database. c) Representative Hi-C map showing 3-dimensional interactions in the *Cx. laticinctus* genome and the chromosomal territories. The diagonal corresponds to intrachromosomal contacts, depicting chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Hi-C matrix bins were merged into a 25 kb bin size for plotting.

These genomic resources will assist in informing more effective control measures by (i) facilitating the identification of genes that influence vector competence; (ii) improving our understanding of mosquito behavior, such as feeding and breeding preferences; (iii) being the basis for population genomic studies which provide insights into mosquito dispersal, gene flow, and adaptation, including the evolution of insecticide resistance. Furthermore, these high-quality reference genomes will help study the evolutionary adaptations of mosquitoes to changing environments, helping predict shifts in disease transmission patterns and informing long-term strategies.

This genome was generated as part of European Reference Genome Atlas (ERGA, Mazzoni et al. 2023) initiative's Biodiversity Genomics Europe (BGE) project.

Results and Discussion

To generate the assemblies, 39 Gb (47X) PacBio HiFi, with mean length of 8.4 kb, and 163 Gb (196X) HiC data were sequenced for *Cx. Laticinctus*; 21 Gb (30X) PacBio HiFi, mean length 10.3 kb, and 115 Gb (164X) HiC data for *Cx. Modestus*; 19 Gb (43X) PacBio HiFi, mean length 10.5 kb, and 82 Gb (186X) HiC data for *Cx. Perexiguus*; and 62 Gb (75X) PacBio HiFi, mean length 7.8 kb, and 119 Gb (143X) HiC data for *Cx. theileri*. For genome annotation, 32.5 million read pairs of RNA-seq data were sequenced for *Cx. laticinctus*, with a mapping rate of 95.35%; 33.4 million read pairs for *Cx. modestus*, with a mapping rate of 87.9%; 32.9 million read pairs for *Cx. perexiguus*, with a mapping rate of 85.7%; and 35.5 million read pairs for *Cx. theileri*, with a mapping rate of 92.64% as given by Hisat2 v2.2.1.

Genome Assemblies

The genome assemblies for all four species ranged in size from 436,197,037 bp—833,828,062 bp with contig N50 s between 312 and 770 kb and scaffold N50s between 150 and 291 Mb. The genome assemblies contained between 95.6% and 97.6% single-copy orthologs from the BUSCO diptera odb10 database (Manni et al. 2021); all assemblies are chromosome-level with between 95% and 6%, 99.2% of each assembly assigned to the three chromosomes; and all assemblies were highly accurate, with quality values (QVs) between 54.9% and 57.0% as calculated by Merqury (Rhie et al. 2020 , Fig. 1, Table S1). The mitochondrial sequences were all recovered and are included in the submitted genome assemblies, with consistent lengths around 15.6 kb.

Genome Annotations

Gene model annotations for all species resulted in detection of 12,534 to 14,175 genes. The protein-coding

Table 1 Statistics from assembled gene models including proteome completeness as evaluated by BUSCO using the diptera_odb10 database and proteome completeness and consistency as evaluated by OMArk using the Culicidae HOGs

	Culex laticinctus	*Culex modestus*	*Culex perexiguus*	*Culex theileri*
# Genes	13,725	14,175	12,534	12,627
# Transcripts	20,830	21,650	18,231	18,558
BUSCO Complete single	3,126 (95.2%)	2,975 (90.6%)	3,024 (92.1%)	3,066 (93.3%)
BUSCO complete duplicated	34 (1.0%)	135 (4.1%)	91 (2.8%)	66 (2.0%)
BUSCO fragmented	27 (0.8%)	27 (0.8%)	35 (1.1%)	31 (0.9%)
BUSCO missing	98 (3.0%)	148 (4.5%)	135 (4.0%)	122 (3.8%)
OMArk Complete single	6,323 (89.94%)	6,059 (86.19%)	6,026 (85.72%)	6,098 (86.74%)
OMArk complete duplicated	313 (4.45%)	509 (7.24%)	399 (5.68%)	395 (5.62%)
OMark missing	394 (5.60%)	462 (6.57%)	605 (8.61%)	537 (7.64%)
OMArk consistent	11,914 (86.81%)	12,270 (86.56%)	11,390 (90.87%)	11,516 (91.20%)
OMArk inconsistent	428 (3.12%)	473 (3.34%)	382 (3.05%)	357 (2.83%)
OMArk unknown	1,383 (10.08%)	1,432 (10.10%)	762 (6.08%)	754 (5.97%)

annotations show complete and consistent detection, with BUSCO completeness scores ranging from 94.7% to 96.2% based on the diptera_odb10 database and OMArk completeness scores ranging from 91.39% to 94.4% and consistency scores ranging from 86.56% to 91.2% based on the *Culicidae* Hierarchical Ortholog Groups (Table 1).

Materials and Methods

Sample and Sampling Information

Mosquitoes were collected from Andalusia, Spain (Table S1), using BG-Sentinel traps and identified using MosKeyTool (Gunay et al. 2018). Specimens were euthanized by freezing in dry ice after identification and all samples were then preserved at −80 °C until DNA and RNA extraction. Individuals of *Cx. laticinctus*, *Cx. modestus*, *Cx. perexiguus*, and *Cx. theileri* were selected and barcoding was performed following Folmer et al. (1994). Sampling was performed under permission of article 9, law 8/2003, issued on 28/06/2022 by the Consejería de Agricultura, Ganadería, Pesca y Desarrollo Sostenible, Junta de Andalucía. See Supplementary Material for vouchering data.

DNA/RNA Processing

Whole *Culex* mosquitoes were homogenized using a PowerMasher II (Denton et al. 2023). HMW DNA was extracted using the Automated MagAttract v2 (Oatley et al. 2023a). For Ultra-Low Input (ULI) PacBio sequencing, DNA was fragmented using a Covaris g-TUBE (Oatley et al. 2023c). Sheared DNA was purified using AMPure PB to eliminate shorter fragments and concentrate the DNA (Oatley et al. 2023b). RNA was extracted from whole-organism tissues of all species using the Automated MagMax mirVana kit (do Amaral et al. 2023).

Tissues from whole *Culex* mosquitoes were fixed, and the DNA crosslinked using 22% formaldehyde. After crosslinking, the tissues were homogenized with a Diagnocine Power Masher-II/BioMasher-II. Samples were processed using the Arima v2 Hi-C library preparation kit (see Supplementary Material for further details).

Library Preparation and Sequencing

Ultra-low input libraries were prepared using the PacBio SMRTbell Express Template Prep 2.0 and PacBio SMRTbell gDNA Sample Amplification kits, respectively. Samples were normalized to 20 ng of DNA and processed following manufacturer's instructions until the SMRTbell gDNA Sample Amplification Kit adapters were ligated. A 0.85X pre-PCR clean-up was performed using the Promega ProNex (Promega, Madison, Wisconsin, USA) system and the sample was divided into two. Both PCR reactions followed the PCR programs and were pooled ensuring the total mass was ≥500 ng in 47.4 μl. The pooled DNA sample was processed again to repair DNA damage, end repair/A-tailing, and additional hairpin adapter ligation. After a 1X clean-up and DNA quantification, size selection was performed to target fragments between 4,000 and 9,000 bp. Size-selected libraries then had a final 1.0× clean-up. Prepared libraries were normalized to 2 nM and 15 μL were used for making complexes. Circularized complexes were created according to the manufacturer's instructions. The complexes were purified with the 1.2× clean-up with magnetic beads (SMRTbell). The purified complexes were then diluted to the PacBio Revio loading concentration, between 200 and 300 pM, spiked with a Revio sequencing internal control and sequences using the Revio system on Revio 25 M SMRT cells. (See Supplementary Material for further details.)

For Hi-C library preparation, DNA was fragmented to a size of 400 to 600 bp using the Covaris E220 sonicator (Woburn, Massachusetts, USA). The DNA was then enriched, barcoded, and amplified using the NEBNext Ultra II DNA Library Prep Kit (NEB, Ipswich, Massachusetts,

USA) following manufacturers' instructions. The Hi-C sequencing was performed using paired-end sequencing with a read length of 150 bp on an Illumina NovaSeq X instrument.

Poly(A) RNA-Seq libraries were constructed using the Ultra II RNA Library Prep ki (NEB, Ipswich, Massachusetts, USA), following the manufacturer's instructions. RNA sequencing was performed on the Illumina NovaSeq X instrument.

Genome Assembly Methods

The HiFi reads were first assembled using Hifiasm (Cheng et al. 2021) with the –primary option. The Hi-C reads were mapped to the primary contigs using bwa-mem2 (Vasimuddin et al. 2019). The contigs were further scaffolded using the provided Hi-C data (Rao et al. 2014) in YaHS (Zhou et al. 2023) using the –break option.

The mitochondrial genomes were assembled using MitoHiFi (Uliano-Silva et al. 2023) in default mode, with the reference sequence from *Culex quinquefasciatus* (NC_014574.1) used to find mitochondrial reads in the PacBio HiFi datasets.

The assembly was decontaminated using the Assembly Screen for Cobionts and Contaminants (ASCC) pipeline. Manual curation was performed with PretextView, JBrowse2 (Diesh et al. 2023), HiGlass (Kerpedjiev et al. 2018), and the TreeVal pipeline (Pointon et al. 2023). Scaffolds were visually inspected and corrected as described by Howe et al. (2021), removing contamination and duplicate sequences and correcting missed- and misjoins. Summary analyses of the released assemblies were performed using the ERGA-BGE Genome Report ASM Galaxy workflow (https://doi.org/10.48546/workflowhub.workflow.1104.1).

For *Cx. theileri*, 435 Mb of contaminant sequence was removed, with the majority (97%) identified as Acinetobacter. We applied 181 scaffold breaks and 198 joins to correct the assembly and removed 61 erroneous haplotypic duplications. The *Cx. laticinctus* assembly was separated from 990 Mb of Wolbachia group B sequence. It required two scaffold breaks and 285 joins to correct the assembly structure, plus three removals of erroneous haplotypic duplication. *Cx. modestus* showed no contamination; however 564 scaffold breaks, 54 joins, and 423 removals were necessary. Lastly, *Cx. perexiguus* was separated from 84 Mb of mitochondrial sequence and required 142 scaffold breaks, 167 joins, and 342 removals.

Genome Annotation Methods

A gene set was generated using the Ensembl Gene Annotation system (Aken et al. 2016), primarily by aligning short-read RNA-seq data to the genomes, which were generated in this study. Gaps in the annotation were filled via protein-to-genome alignments of clade-specific proteins from UniProt ("UniProt: A Worldwide Hub of Protein Knowledge," 2019), with experimental evidence at the protein or transcript level. Models were aggregated and consolidated, prioritizing those from RNA-seq data, giving final gene model sets and associated nonredundant transcripts. Low-quality transcript models, such as those showing evidence of fragmented ORFs, were removed. In cases where RNA-seq data were fragmented or absent, homology data were prioritized, favoring longer transcripts with strong intron support from short-read data. The resulting gene models were classified as protein-coding and long noncoding (lncRNA). Models that did not overlap protein-coding genes and were constructed from transcriptomic data were considered potential lncRNA and were further filtered to remove single-exon loci. Putative microRNAs were predicted by mapping miRBase (Kozomara et al. 2019) against the genome using BLAST, followed by RNAfold analysis (Gruber et al. 2008). Other small noncoding loci were identified using the Rfam database (Kalvari et al. 2018) and passing the results through Infernal (Nawrocki and Eddy 2013). Summary analysis of the released annotation was carried out using the ERGA-BGE Genome Report ANNOT Galaxy workflow (https://doi.org/10.48546/workflowhub.workflow.1096.1).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We express our gratitude to Álvaro Solís for help during the fieldwork and the Junta de Andalucía for granting sampling permits. We acknowledge the assembly reviewers, Jéssica Gòmez Garrido and Fernando Cruz from the Centro Nacional de Análisis Genómico (CNAG). The authors acknowledge the support of the Freiburg Galaxy Team: Saim Momin and Björn Grüning, Bioinformatics, University of Freiburg (Germany), funded by the German Federal Ministry of Education and Research BMBF grant 031 A538A de.NBI-RBC and the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) within the framework of LIBIS/de.NBI Freiburg.

Author Contributions

M.J.R.L. coordinated the project; S.C.C. collected and identified the species; M.J.R.L. and S.C.C. sampled and preserved biological material and provided metadata; J.F., J.M.P., S.M. and I.I. provided sampling and metadata support and management; R.M.W. contributed to logistics and project coordination; Members of the Wellcome Sanger Institute Tree of Life Management, Samples and

Laboratory team extracted DNA, prepared libraries, and performed sequencing; Members of the Wellcome Sanger Institute Tree of Life Core Informatics team performed genome assembly and curation under the supervision of M.B.; L.H. performed genome annotation under the supervision of F.M.; and T.B. generated the analysis and report. All authors contributed to the writing, review, and editing of this genome report and read and approved the final version.

Funding

This project received funding from Horizon Europe under the Biodiversity, Circular Economy and Environment (REA.B.3), cofunded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 22.00173 and 24.00054 and by the UK Research and Innovation (UKRI) under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme. M.J.R.L. was funded by the Spanish State Research Agency (project PID2020-118921RJ-100/AEI/10.130939/501100011033). I.I. was funded by the Spanish State Research Agency (AEI) MICIU/AEI/10.13039/501100011033, ESF+, and ERDF/EU (Grants RyC2022-038245-I and PID2023-152168NB-I00).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

Cx. laticinctus, *Cx. modestus*, *Cx. Perexiguus*, and *Cx. theileri* and the related genomic studies were assigned to Tree of Life IDs (ToLIDs) idCulLati1, idCulMode1, idCulPerx1, and idCulThei1, respectively, and all sample, sequence, and assembly information are available under the umbrella BioProjects PRJEB75414, PRJEB75163, PRJEB75165, and PRJEB76570. The sample information is available at the following BioSample accessions: *Cx. laticinctus*, SAMEA 114402094, SAMEA114402091, SAMEA114402090, and SAMEA114402071; *Cx. modestus*, SAMEA114402081, SAMEA114402100, SAMEA114402102, and SAMEA11 4402103; *Cx. perexiguus*, SAMEA114402066, SAMEA11 4402085, SAMEA114402086, and SAMEA114402089; and *Cx. theileri*, SAMEA114402076, SAMEA114402095, SAMEA114402096, and SAMEA114402099. The genome assemblies are accessible from GenBank under accession numbers *Cx. laticinctus*, GCA_964187845.1; *Cx. modestus*, GCA_964213205.1; *Cx. perexiguus*, GCA_9642 43045.1; and *Cx. theileri*, GCA_964340515.1, and the annotated genomes are available via Ensembl (https://beta.ensembl.org/). Raw sequencing data produced as part of this project are available from SRA at the following accessions: *Cx. laticinctus*, ERX12405204, ERX12405205, ERX12433627, and ERX12519568; *Cx. modestus*, ERX12326303, ERX12317845, and ERX12519565; *Cx. perexiguus*, ERX12326304, ERX12317846, and ERX12519566; and *Cx. theileri*, ERX12671962, ERX12671963, and ERX13020690. Barcoding sequences used for species identification are available at the following BioProject accessions: *Cx. laticinctus*, PRJEB90329; *Cx. modestus*, PRJEB90328; *Cx. perexiguus*, PRJEB90330; and *Cx. theileri*, PRJEB90331. Supporting data used to generate the statistics, tables, and figures in this report are available at the following repository: https://doi.org/10.5281/zenodo.15655399. Documentation related to the genome assembly and curation can be found in the ERGA Assembly Report (EAR) documents available at the following locations: https://github.com/ERGA-consortium/EARs/tree/main/Assembly_Reports/Culex_laticinctus/idCulLati1 https://github.com/ERGA-consortium/EARs/tree/main/Assembly_Reports/Culex_modestus/idCulMode1 https://github.com/ERGA-consortium/EARs/tree/main/Assembly_Reports/Culex_perexiguus/idCulPerx1 https://github.com/ERGA-consortium/EARs/tree/main/Assembly_Reports/Culex_theileri/idCulThei1. Further details and data about the project are hosted on the ERGA portal at the following locations: https://portal.erga-biodiversity.eu/data_portal/1464561, https://portal.erga-biodiversity.eu/data_portal/545656, https://portal.erga-biodiversity.eu/data_portal/943103 and https://portal.erga-biodiversity.eu/data_portal/5612.

Author Information

Members of the Wellcome Sanger Institute Tree of Life Management, Samples and Laboratory team are listed here: https://doi.org/10.5281/zenodo.13960213.

Members of Wellcome Sanger Institute Scientific Operations: Sequencing Operations are listed here: https://doi.org/10.5281/zenodo.14870789.

Members of the Wellcome Sanger Institute Tree of Life Core Informatics team are listed here: https://doi.org/10.5281/zenodo.15440189.

Literature Cited

Aken BL, et al. The Ensembl gene annotation system. Database. 2016:2016:baw093. https://doi.org/10.1093/database/baw093.

Becker N, et al. Mosquitoes and their control. 2nd ed. Springer; 2010. https://doi.org/10.1007/978-3-540-92874-4.

Bravo-Barriga D, et al. Identification of Usutu virus Africa 3 lineage in a survey of mosquitoes and birds from urban areas of western Spain. Transbound Emerg Dis. 2023:2023:6893677. https://doi.org/10.1155/2023/6893677.

Burgas-Pau A, et al. Laboratory evidence on the vector competence of European field-captured *Culex theileri* for circulating west Nile virus lineages 1 and 2. Parasit Vectors. 2025:18:132. https://doi.org/10.1186/s13071-025-06763-6.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021:2021:170–175. https://doi.org/10.1038/s41592-020-01056-5.

Denton A, Oatley G, Cornwell C, Quail M, Howard C. Sanger tree of life sample homogenisation: PowerMash. protocols.io. 2023. https://doi.org/10.17504/protocols.io.5qpvo3r19v4o/v1.

Diesh C, et al. JBrowse 2: a modular genome browser with views of synteny and structural variation. Genome Biol. 2023:24:74. https://doi.org/10.1186/s13059-023-02914-z.

do Amaral RJV, et al. Sanger tree of life RNA extraction: Automated MagMax mirVana. protocols.io. 2023. https://doi.org/10.17504/protocols.io.6qpvr36n3vmk/v1.

Figuerola J, et al. A one health view of the west nile virus outbreak in Andalusia (Spain) in 2020. Emerg Microbes Infect. 2022:11:2570–2578. https://doi.org/10.1080/22221751.2022.2134055.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol. 1994:3:294–299.

Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA websuite. Nucleic Acids Res. 2008:36:W70–W74. https://doi.org/10.1093/nar/gkn188.

Gunay F, Picard M, Robert V. MosKeyTool, an interactive identification key for mosquitoes of Euro-Mediterranean. Version 2.1. 2018. Available from: https://www.onehealthsecure.com/entomology-tools-0/moskeytool.

Gutiérrez-López R, et al. First isolation of the Sindbis virus in mosquitoes from southwestern Spain reveals a new recent introduction from Africa. One Health. 2025:20:100947. https://doi.org/10.1016/j.onehlt.2024.100947.

Gutsevich AV, Monchadskii AS, Shtakel'berg AA. Fauna SSSR Vol 3(4), family Culicidae, 384pp. Akad Nauk SSSR Zool Inst; 1974.

Harbach RE. The mosquitoes of the subgenus *Culex* in southwestern Asia and Egypt (Diptera: Culicidae). Contrib Am Ent Inst. 1988:24:1–240.

Howe K, et al. Significantly improving the quality of genome assemblies through curation. GigaScience. 2021:10:giaa153. https://doi.org/10.1093/gigascience/giaa153.

Kalvari I, et al. Non-coding RNA analysis using the Rfam database. Curr Protoc Bioinformatics. 2018:62:e51. https://doi.org/10.1002/cpbi.51.

Kerpedjiev P, et al. Higlass: web-based visual exploration and analysis of genome interaction maps. Genome Biol. 2018:19:125. https://doi.org/10.1186/s13059-018-1486-1.

Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. Nucleic Acids Res. 2019:47:D155–D162. https://doi.org/10.1093/nar/gky1141.

Lundström JO. Vector competence of western European mosquitoes for arboviruses: a review of field and experimental studies. Bull Soc Vector Ecol. 1994:19:23–36.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of genomes. Mol Biol Evol. 2021:38:4647–4654. https://doi.org/10.1093/molbev/msab199.

Martínez-de la Puente J, Santiago-Alarcon D, Palinauskas V, Bensch S. *Plasmodium relictum*. Trends Parasitol. 2021:37:355–356. https://doi.org/10.1016/j.pt.2020.06.004.

Mazzoni C, Ciofi C, Waterhouse R. Biodiversity: an atlas of European reference genomes. Nature. 2023:619:252. https://doi.org/10.1038/d41586-023-02229-w.

McIntosh BM, Jupp PG, Dos Santos I, Meenehan GM. Vector studies on Sindbis and West Nile viruses in South Africa. S Afr Med J. 1975:48:2074–2078.

Muñoz J, et al. Feeding patterns of potential West Nile virus vectors in South-West Spain. PLoS One. 2012:7:e39549. https://doi.org/10.1371/journal.pone.0039549.

Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013:29:2933–2935. https://doi.org/10.1093/bioinformatics/btt509.

Oatley G, Denton A, Howard C. Sanger tree of life HMW DNA extraction: automated MagAttract v2. Protocols.io. 2023a. https://doi.org/10.17504/protocols.io.kxygx3y4dg8j/v1.

Oatley G, Sampaio F, Howard C. Sanger tree of life fragmented DNA clean up: Automated SPRI. protocols.io. 2023b. https://doi.org/10.17504/protocols.io.q26g7p1wkgwz/v1.

Oatley G, Sampaio F, Kitchin L, do Amaral RJV, Howard C. Sanger tree of life HMW DNA fragmentation: Covaris g-TUBE for ULI PacBio. protocols.io. 2023c. https://doi.org/10.17504/protocols.io.q26g7pm81gwz/v1.

Pires CA, Ribeiro H, Capela RA, Ramos HC. Research on the mosquitoes of Portugal VI—the mosquitoes of Alentejo. Ann Inst Hig Med Trop. 1982:8:79–102.

Pointon DL, Eagles W, Sims Y, Muffato M, Surana P. sanger-tol/treeval v1.0.0—Ancient Atlantis. Zenodo, 2023. https://doi.org/10.5281/zenodo.10047653.

Ramos HC, Ribeiro H, Pires CA, Capela RA. Research on the mosquitoes of Portugal, II—the mosquitoes of Algarve, 1977/78. Ann Inst Hig Med Trop. 1977:5:236–256.

Rao SSP, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014:159:1665–1680. https://doi.org/10.1016/j.cell.2014.11.021.

Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 2020:21:245. https://doi.org/10.1186/s13059-020-02134-9.

Ribeiro H, Ramos HC, Pires CA, Capela RA. An annotated checklist of the mosquitoes of continental Portugal (Diptera: Culicidae). Actas III Congr Iber Ent Granada. 1988:233–253.

Schaffner F, et al. The mosquitoes of Europe: an identification and training programme. IRD Éditions; 2001.

Sirivanakarn S. Medical entomology studies–III. A revision of the subgenus *Culex* in the oriental region (Diptera: Culicidae). Contrib Am Ent Inst. 1976:12:1–272.

Smith KGV. Insects and other arthropods of medical significance. British Museum (Natural History; 1973.

Snow RW, Ramsdale CD. Distribution chart for European mosquitoes. Eur Mosq Bull. 1999:3:14–31.

Soto A, Delang L. *Culex modestus*: the overlooked mosquito vector. Parasit Vectors. 2023:16:373. https://doi.org/10.1186/s13071-023-05997-6.

Soto A, Wauters L, Delang L. Is *Culex modestus* a new Usutu virus vector? Parasit Vectors. 2024:17:285. https://doi.org/10.1186/s13071-024-06360-z.

Uliano-Silva M, et al. Mitohifi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. BMC Bioinformatics. 2023:24:288. https://doi.org/10.1186/s12859-023-05385-y.

UniProt Consortium. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019:47:D506–D515. https://doi.org/10.1093/nar/gky1049.

Vasimuddin M, Misra S, Li H, Aluru S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: *Proc IEEE int parallel distrib process symp (IPDPS)*; 2019. pp. 314–324. https://ieeexplore.ieee.org/document/8820962.

Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. Bioinformatics. 2023:39:btac808. https://doi.org/10.1093/bioinformatics/btac808.

Associate editor: Christopher Wheat