VIDWAN Journal of Medical and Pharmaceutical Sciences

OPEN ACCESS

PEER-REVIEWED - OPEN ACCESS Vol. 01 · Issue 01 · 2026 · pp. 33-37| ISSN: APPLIED (Online) | INAGURAL EDITION

PSEUDOBASE A COMPARATIVE GENOMICS PLATFORM FOR THE DROSOPHILA PSEUDOOBSCURA SUBGROUP

Dr. D. Uday Kumar, *Department Of Pharmacology,*
Joginpally B.R. Pharmacy College, Hyderabad

ABSTRACT

PseudoBase, a comparative genomics platform, studies the genetic diversity of Drosophila pseudoobscura, D. persimilis, D. miranda, and D. lowei. Using high-quality whole-genome sequencing data and a conventional variant-calling algorithm, the platform finds SNPs and indels in numerous species and strains. PseudoBase's simple web interface lets you search by gene or region, use an interactive JBrowse genome browser, check sequence alignment, download datasets, and do multiple queries. These methods let scientists compare genomic regions, find species-specific variations, and study species evolution and adaptability. Comparative genomics, evolutionary biology, population genetics, and functional genomics studies on Drosophila pseudoobscura are possible with its genomic data and rich visualization and analytic tools.

Keywords: *PseudoBase; Comparative genomics; Drosophila pseudoobscura subgroup; Whole-genome sequencing; Single nucleotide polymorphisms (SNPs); Insertions and deletions (Indels).*

Article history: Received 8 July 2026 · Revised 20 July 2026 · Accepted 29 July 2026 · Published 7 August 2026

1. INTRODUCTION

Biology is studied through evolutionary links, genetic diversity, and gene function. Genome structure, gene organization, sequence conservation, and species evolution can be studied using next-generation sequencing data. These similarities highlight genomic evolution, gene roles, speciation, chromosome rearrangements, and adaptive molecular pathways. Analysis of large datasets requires genetic databases and comparative analytic methods.

Drosophila pseudoobscura has numerous well-characterized genes, therefore evolutionary genetics studies it. The species in this subgroup have taught us about genome evolution, natural selection, reproductive isolation, and chromosomal inversions. By investigating these closely related species' sequence conservation and divergence trends, scientists might learn how genes develop and adapt. High-quality genome assemblies and annotations enable thorough comparative genomic studies in this subgroup.

Historical Development of Comparative Genomics

Early Development

Classical geneticists investigated genes and chromosomes to understand evolution. Here, comparative genomics began. The initial study used cytogenetics, genetic mapping, and gene sequencing without genome data. Since the Human Genome Project and DNA sequencing have evolved, it is now possible to compare species' genomes to detect gene differences.

Advancement of Sequencing Technologies

NGS has boosted genomic research by making DNA sequencing cheaper and producing more data. Oxford Nanopore, PacBio, and Illumina sequenced several genomes. Comparative genetics is simplified. These discoveries make comparative genomics useful for studying genetic diversity, adaptability, and genome evolution.

Development of Genomic Databases

For genomic data, GenBank, ENA, and DDBJ were formed. Organism-specific services like FlyBase, Ensembl, and the UCSC Genome Browser simplify genomic data access with genome assemblies, gene labels, and visualization tools. These approaches mostly compare taxonomic groups, not evolutionary groups.

Emergence of PseudoBase

PseudoBase, a Drosophila pseudoobscura comparative genomics platform, addresses these concerns. The database includes SNPs, indels, gene annotations, standardized genome alignments, and an interactive genome viewer. This specialized database evaluates closely related species' genetic variants for evolutionary, demographic, and functional genomics studies.

2.METHODOLOGY

Data Collection

From NCBI Short Read Archive publications, we got whole-genome paired-end Illumina sequencing files for Drosophila pseudoobscura, persimilis, miranda, and lowei The best and most recent sequencing data were used for strains having different data. Cross genomes created in labs were rejected.

Genome Alignment and Variant Detection

BWA v0.7.17 matched sequencing reads with Dpse_3.04, its D. pseudoobscura reference genome. Picard's tools deleted sequencing reads and adapters. GATK 4.1.1 found SNPs and indels. QD, FS, SOR, MQ, MQRankSum, and ReadPosRankSum were employed to find suitable genetic variants.

Database Development

In MySQL and Django, PseudoBase contains strain, gene, and indexed sequencing data. Indexing genome alignments with chromosomes found gene sequences and genomic areas quickly. VCF files add database data.

Genome Visualization

PseudoBase displays genes, SNPs, indels, and genome annotations with an interactive JBrowse genome browser. A simple web tool that works with most browsers lets users compare strains, discover genes or genomic areas, generate FASTA alignments, and analyze genomic variation.

Data Integration

All genomes were aligned to the reference genome using bioinformatics to simplify comparisons. The final PseudoBase has 61 genomes from four Drosophila species and was designed for sequence and annotation updates.

3. MATERIALS AND METHODS

Whole-genome paired-end Illumina sequencing samples from Drosophila pseudoobscura subgroup members D. pseudoobscura, D. bogotana, D. persimilis, D. miranda, and D. lowei are in Pseudo The NCBI Short Read Archive provides raw sequencing data and metadata. When there were several sequencing datasets for the same strain, the latest approach with the best coverage was chosen. Genomes from lab strain hybrids were eliminated to unify samples and avoid bias in comparison analysis.

Sequence Alignment and Variant Calling

Common bioinformatics ensured all datasets had the same sequencing read processing. The Drosophila pseudoobscura reference genome assembly (Dpse_3.04; GCA_000001765.2) was matched using BWA 0.7.17. After matching, Picard command-line tools eliminated duplicate reads and sequencing adapters. GATK 4.1.1 sorted differences. GATK's hard-filtering separated SNPs and indels. List of quality by depth, Fisher Strand Bias, Strand Odds Ratio, Mapping Quality, and Read Position Rank Sum. This method developed standardized high-confidence variation databases for all genomes.

Database Architecture

The genomic data storing and organization system PseudoBase was constructed using Django and MySQL. Aligned or unaligned genomic sequences for each gene or region are easy to access in indexed, chromosome-specific files. The database speeds searches with strain information, genomic coordinates, and indexed sequence files. SNP and indel variant call format (VCF) files simplify strain addition. The database grows as genetic resources are found.

Genome Visualization Platform

All major browsers support PseudoBase. The built-in JBrowse genome browser shows genomic areas, annotations, SNPs, and indels from multiple species and strains. VCFTabix and HTMLVariant tracks simplify comparative genomic research by showing genetic differences. NeatHTMLFeatures, HierarchicalCheckBoxPlugin, and HideTrackLabels ease track administration, interface visibility, and navigation. These traits allow non-bioinformaticists to explore genetic variation interactively.

4. EVALUATION RESULTS AND DISCUSSION

Interface and Key Features

PseudoBase (http://pseudobase.biology.duke.edu/) simplifies Drosophila pseudoobscura genomic variation exploration. Website sections include Home, Browse, Info, Links, Updates, and Contact Us. Use genomic region or gene name to search page. Acronyms include Gene, GA, GL, CG, and FBgn. View genomic regions with JBrowse or FASTA alignments.

JBrowse lets users pick or deselect strains to compare the genomes of D. pseudoobscura pseudoobscura, D. pseudoobscura bogotana, D. persimilis, D. miranda, and D. lowei. The browser plots genome-wide FlyBase gene annotations, SNPs, and indels. Clicking variants shows allele identification and sequencing depth, simplifying comparative genomic analysis.

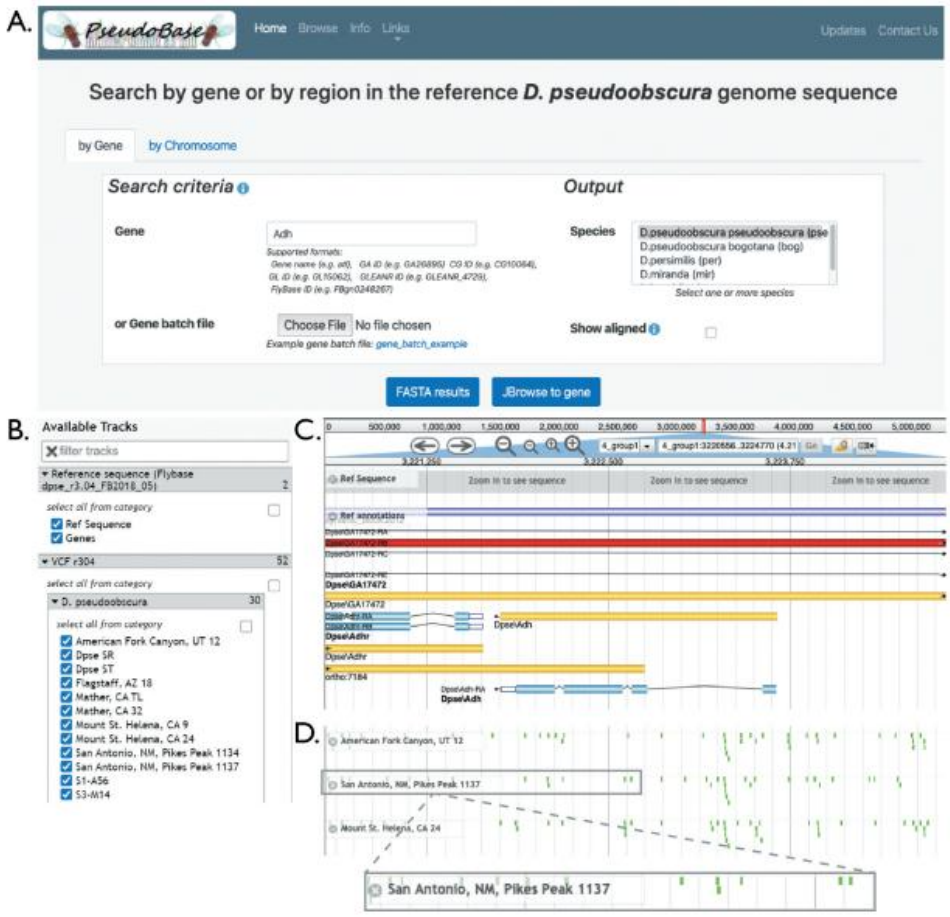

Figure1. PseudoBase Search and Genome Browser Interface

Additional tabs include contact information, genetic resource links, documentation, and release notes. Discussed include strains, variant-calling, pseudoBase, and documentation. Updates list upcoming program upgrades, while Links lists related resources.

Data Content and Data Types

The sixty-one sequenced genomes in PseudoBase 2.0 include 31 D. pseudoobscura pseudoobscura, five D. bogotana, thirteen D. persimilis, eleven D. miranda, and one D. lowei. All genomes were aligned and analyzed using the D. pseudoobscura reference genome to ensure correct SNP and indels across strains and species.

We created a genome sequencing and variation database. FlyBase no longer contains non-melanogaster species, but it has gene annotations. The authors recommend GenBank. Because of PseudoBase's static genome assembly and annotation, updates are feasible.

Applications

The authors say PseudoBase's simplicity works. No genetic information, genome assembly, or variant-calling analysis is needed to access the database. The integrated JBrowse interface lets users construct FASTA-formatted gene alignments, download them for future research, and explore genomic variation. With little computing knowledge, it speeds comparative genomics and marker creation.

Another PseudoBase benefit is education. Students can practice, compare datasets, and extract genetic variation with this application. Student research projects and beginning genetics laboratories have used the site, proving its research and teaching potential. Population genetic analysis and marker development without bioinformatics are easy on the platform.

Comparison with Other Genomic Databases

Genomic databases are beneficial for collecting sequences and annotating genomes, yet they are not successful for comparing Drosophila pseudoobscura subgroups. PseudoBase uses sequence alignment for closely related Drosophila species, comparative genome display, and standardized variation databases to overcome this restriction. PseudoBase allows strain-to-strain genetic variation comparison while preserving analytical methods, unlike other databases.

Table 1. Comparison of PseudoBase with Existing Genomic Databases

Feature	PseudoBase	FlyBase	Ensembl Metazoa	UCSC Genome Browser
Species-specific database	✓	✗	✗	✗
Comparative SNP analysis	✓	Limited	Limited	Limited
Multiple strain comparison	✓	Limited	✓	Limited
Interactive genome browser	✓	✓	✓	✓
Gene annotation	✓	✓	✓	✓
Sequence alignment download	✓	Partial	✓	Partial
Standardized variant pipeline	✓	✗	✗	✗
Focus on D. pseudoobscura subgroup	✓	✗	✗	✗

5. CONCLUSION

PseudoBase is a sophisticated comparative genomics platform for Drosophila pseudoobscura research. An integrated platform for genomic variation data, visualization, gene annotation, and sequence analysis lets researchers accurately compare disparate species. The interface simplifies genomic research for experienced researchers and trainees, and standardized genetic data processing ensures consistency and reliability. The platform's scalable architecture ensures its lifespan for population, evolutionary, and comparative genomics research by accommodating future genome and annotation updates. PseudoBase simplifies genomic data and explains Drosophila pseudoobscura subgroup genetic diversity and evolution.

REFERENCES

[1] Sturtevant AH, Dobzhansky T. Inversions in the third chromosome of wild races of Drosophila pseudoobscura, and their use in the study of the history of the species. Proc Nat Acad Sci. 1936;22:448–450.

[2] Dobzhansky T. Genetics of natural populations IX. Temporal changes in the composition of populations of Drosophila pseudoobscura. Genetics. 1943;28 (2):162–186.
[3] Dobzhansky T, Sturtevant AH. Inversions in the chromosomes of Drosophila pseudoobscura. Genetics. 1938;23(1):28–64.
[4] Lewontin R, Hubby J. A molecular approach to the study of genic heterozygosity in natural populations II. Amount of variation and degree of heterozygosity in natural populations of Drosophila pseudobscura. Genetics. 1966;54(2):565–609.
[5] Fuller ZL, Koury SA, Phadnis N, et al. How chromosomal rearrangements shape adaptation and speciation: case studies in Drosophila pseudoobscura and its sibling species Drosophila persimilis. Mol Ecol. 2018;28 (6):1283–1301.
[6] Fishman L, Stathos A, Beardsley PM, et al. Chromosomal rearrangements and the genetics of reproductive barriers in Mimulus (monkey flowers). Evolution. 2013;67(9):2547–2560.
[7] Hoffmann AA, Rieseberg LH. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? Annu Rev Ecol Evol Syst. 2008;39:21–42.
[8] Kirkpatrick M, Barrett B. Chromosome inversions, adaptive cassettes and the evolution of species' ranges. Mol Ecol. 2015;24(9):2046–2055.
[9] Kirkpatrick M, Barton N. Chromosome inversions, local adaptation and speciation. Genetics. 2006;173 (1):419–434.
[10] Noor MA, Grams KL, Bertucci LA, et al. Chromosomal inversions and the reproductive isolation of species. Proc Nat Acad Sci. 2001;98 (21):12084–12088.
[11] Noor MAF, Grams KL, Bertucci LA, et al. The genetics of reproductive isolation and the potential for gene exchange between Drosophila pseudoobscura and D. persimilis via backcross hybrid males. Evolution. 2001;55:512–521.
[12] Brown KM, Burk LM, Henagan LM, et al. A test of the chromosomal rearrangement model of speciation in Drosophila pseudoobscura. Evolution. 2004;58 (8):1856–1860.
[13] Chang AS, Noor MAF. The genetics of hybrid male sterility between the allopatric species pair Drosophila persimilis and D. pseudoobscura bogotana: dominant sterility alleles in collinear autosomal regions. Genetics. 2007;176(1):343–349.
[14] Kulathinal RJ, Stevison LS, Noor MAF. The genomics of speciation in Drosophila: diversity, divergence, and introgression estimated using low-coverage genome sequencing. PLoS Genet. 2009;5(7):e1000550.
[15] Phadnis N, Orr H. A single gene causes both male sterility and segregation distortion in Drosophila hybrids. Science. 2009;323(5912):376–379.

