

RESEARCH ARTICLE

Comparative single-cell transcriptomic atlases of drosophilid brains suggest glial evolution during ecological adaptation

Daehan Lee^{1,2*}, Michael P. Shahandeh^{1,3}, Liliane Abuin¹, Richard Benton^{1*}

1 Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland, **2** Department of Biological Sciences, College of Natural Sciences, Sungkyunkwan University, Suwon, Republic of Korea, **3** Department of Biology, Hofstra University, Hempstead, New York, United States of America

* Richard.Benton@unil.ch (RB); lee.daehan@skku.edu (DL)



OPEN ACCESS

Citation: Lee D, Shahandeh MP, Abuin L, Benton R (2025) Comparative single-cell transcriptomic atlases of drosophilid brains suggest glial evolution during ecological adaptation. PLoS Biol 23(4): e3003120. <https://doi.org/10.1371/journal.pbio.3003120>

Academic Editor: Chris D Jiggins, University of Cambridge, UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND

Received: December 15, 2023

Accepted: March 17, 2025

Published: April 29, 2025

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pbio.3003120>

Copyright: © 2025 Lee et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution,

Abstract

To explore how brains change upon species evolution, we generated single-cell transcriptomic atlases of the central brains of three closely related but ecologically distinct drosophilids: the generalists *Drosophila melanogaster* and *Drosophila simulans*, and the noni fruit specialist *Drosophila sechellia*. The global cellular composition of these species' brains is well-conserved, but we predicted a few cell types with different frequencies, notably perineurial glia of the blood–brain barrier, which we validate *in vivo*. Gene expression analysis revealed that distinct cell types evolve at different rates and patterns, with glial populations exhibiting the greatest divergence between species. Compared to the *D. melanogaster* brain, cellular composition and gene expression patterns are more divergent in *D. sechellia* than in *D. simulans*—despite their similar phylogenetic distance from *D. melanogaster*—indicating that the specialization of *D. sechellia* is reflected in the structure and function of its brain. Expression changes in *D. sechellia* include several metabolic signaling genes, suggestive of adaptations to its novel source of nutrition. Additional single-cell transcriptomic analysis on *D. sechellia* revealed genes and cell types responsive to dietary supplement with noni, pointing to glia as sites for both physiological and genetic adaptation to this fruit. Our atlases represent the first comparative datasets for “whole” central brains and provide a comprehensive foundation for studying the evolvability of nervous systems in a well-defined phylogenetic and ecological framework.

Introduction

Animal nervous systems contain hundreds to billions of cells. Encompassing neurons and glia, these cells can be categorized into a large number of different types—based upon developmental, anatomical, molecular, and functional properties [1,2]—with diverse roles. For example, neurons act in sensory detection, information processing,

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

Data availability statement: Raw sequencing data are archived in NCBI Gene Expression Omnibus (GEO) under accession code GSE247965. All bioinformatic and experiments datasets (including raw confocal images) for generating the figures and tables are available from Zenodo (<https://zenodo.org/records/15016454>). Processed single-cell transcriptomic atlases are available from SCoPE (https://scope.aertslab.org/#/DmelDsimDsec_brains).

Funding: DL is supported by the National Research Foundation of Korea (NRF) grants funded by the Ministry of Science and ICT under Project Numbers RS-2023-00211007 and RS-2023-00218602, and by the Ministry of Education under Project Number NRF-2019R1A6A1A10073079. Research in RB's laboratory is supported by the University of Lausanne, an ERC Advanced Grant (833548) and the Swiss National Science Foundation (310030B_185377). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

Abbreviations: ANOVA, analysis of variance; BBB, blood–brain barrier; DEG, differentially expressed gene; FDR, false discovery rate; GO, Gene Ontology; GPCR, G-protein-coupled receptor; Ms, Myosuppressin; OPN, olfactory projection neuron; PCA, principal component analysis; RPCA, reciprocal principal component analysis; snRNA-seq, single-nucleus RNA sequencing.

and locomotor control, while glia generally have support functions, including as structural and insulating scaffolds, in nutrient supply, and through removal of cellular debris and toxins [3,4]. The complement of cells in the nervous system of an extant species arises from ongoing evolutionary processes, where external selection pressures can lead to the emergence of new (or modified) cell types that fulfill functions conferring a fitness advantage (e.g., detecting a novel pertinent sensory stimulus, fine-tuning a motor action, or for modulating energy homeostasis). Conversely, cells whose function no longer contributes to organismal fitness might be lost (or repurposed). Understanding the genetic mechanisms and selective pressures underlying the gain, loss, and modification of cell types can reveal how and why nervous systems change over evolutionary timescales, as well as basic insights into the development and function of neural circuits [5,6].

Until recently, our understanding of nervous system evolution relied heavily on correlations of differences in the macroscopic (neuro)anatomy and behavior of different species [5–8], limiting our mechanistic (i.e., genetic) understanding of how evolutionary changes occur. Single-cell transcriptomic approaches have enormous potential to advance this knowledge, by enabling cataloging of neurons and glia and their molecular relationships in various species to suggest hypotheses for how and why divergence in cellular composition has occurred. For example, profiles of cerebellar output neurons in mice, chickens, and humans suggested that developmental duplications of subsets of these cells underlie the large expansion of the human cerebellum [9], while comparisons of cell type-specific transcriptomes in reptiles, amphibians, and other vertebrates have provided insights into both ancient cell types and mammalian-specific innovations of the cortex and other brain regions [10–12]. Such surveys can also relate gene and cell type evolution; for example, a comparison of hypothalamic neuronal populations in fish revealed that species-specific cell types are often characterized by the expression of species-specific gene paralogs [13].

A limitation of studying nervous system evolution in vertebrates is the very large number of cells in their brains: a sampling of >3 million cells from diverse sites in the human brain is a minuscule fraction of the estimated 100–200 billion cells of this organ [14], while a “whole”-brain scRNA-seq atlas of the mouse *Mus musculus* profiled ~7 million neurons [15], but this still represents only about 10% of the total [16]. Moreover, relating structural differences to ecologically relevant functional differences is often challenging. In this context, flies of the *Drosophila* genus define an excellent model clade for investigating nervous system evolution for several reasons. First, these species have relatively compact central brains, comprising ~43,000 cells (of which ~90% are neurons) in *Drosophila melanogaster* [17]. Second, *D. melanogaster* has been the focus of a wealth of molecular, anatomical, and functional studies relating brain structure to function over several decades [18–20], including the generation of single-cell transcriptomic atlases [21–23]. Third, different drosophilid species have adapted to diverse ecological niches—with different food sources, climate conditions, competitors, pathogens, and predators—and display numerous species-specific behaviors, including sensory responses to environmental stimuli and motor actions, such as courtship song production [24,25].

Amongst drosophilids, the trio of *D. melanogaster*, *Drosophila simulans*, and *Drosophila sechellia* present a particularly interesting set of species for comparative neurobiology. These species diverged from a common ancestor 3–5 million years ago, with *D. simulans* and *D. sechellia* diverging much more recently (100–250,000 years ago) [26,27] (Fig 1A). While *D. melanogaster* and *D. simulans* are cosmopolitan generalists, with overlapping geographic ranges and similar broad use of fermenting vegetal substrates for feeding and breeding, *D. sechellia* is endemic to the Seychelles archipelago and has evolved extreme ecological specialization, spending most or all of its life cycle exclusively on the “noni” fruit of the shrub *Morinda citrifolia* [28]. As noni fruit is toxic for other drosophilids, including *D. simulans* inhabiting the Seychelles [29,30], this niche specialization might alleviate interspecific competition, and possibly has essential nutritional benefits [31]. Commensurate with its unique ecology, *D. sechellia* displays many behaviors that are distinct from its generalist relatives, including olfactory and gustatory preferences, circadian plasticity, and certain reproductive behaviors [32–37]. Some of these behaviors have been linked to structural and/or functional changes in the peripheral nervous system. For example, several populations of olfactory sensory neurons display increased sensitivity to noni-derived odors in *D. sechellia* compared to *D. melanogaster* and *D. simulans*, due to coding mutations in specific olfactory receptors [32,33,38,39]. Furthermore, a number of olfactory sensory neuron populations are larger (or smaller) in *D. sechellia*, presumably due to changes in the developmental patterning of the olfactory organs [32,33,40–42], which can affect adaptation properties in the circuitry [42].

Despite these advances in defining causal, or at least correlative, relationships between peripheral sensory neuron changes and species-specific behaviors and ecologies of these flies, we know essentially nothing about if and how their central brains have evolved. Here, we performed comparative single-cell transcriptomic analyses of the central brains of this drosophilid trio to produce, to our knowledge, the first comparative whole central brain atlases in any species. Our results reveal conserved and divergent features of the cellular composition and gene expression patterns in these drosophilids’ central brains as well as signals and patterns of brain evolution linked to niche specialization. This study provides a valuable dataset for future studies to investigate the genetic and cellular basis of brain evolution.

Results

Identification of diverse conserved cell types in drosophilid central brains

We generated central brain comparative single-cell transcriptomic atlases of *D. melanogaster*, *D. simulans*, and *D. sechellia* through single-nucleus RNA sequencing (snRNA-seq) (Fig 1A and 1B, Methods). In brief, we dissected the brains of 5-day-old, mated female adults cultured on standard food medium, and removed the optic lobes. All steps—tissue dissection, nuclear isolation, RNA extraction, library preparation, and sequencing—were performed in parallel for the three species for six biological replicates, each consisting of 20 brains per species. During the processing of snRNA-seq data with the Cell Ranger software (ver. 7.1.0) [43], sequence reads from *D. melanogaster*, *D. simulans*, and *D. sechellia* were mapped to their respective genomes. In total, the number of nuclei sequenced and analyzed per species (*D. melanogaster* = 49,830; *D. simulans* = 43,689; *D. sechellia* = 47,503) all exceed the estimated cell number in the central brain of *D. melanogaster* (~43,000 [17]). Our transcriptomic atlases therefore comprise ~1× cell coverage of their central brains. The median number of genes expressed per cell were 950, 923, and 861 for *D. melanogaster*, *D. simulans*, and *D. sechellia*, respectively, corresponding to 2001, 1841, and 1612 Unique Molecular Identifiers. Detailed metrics for the snRNA-seq results can be found in S1 Table.

To integrate and cluster the single-cell transcriptomic atlases of these three species, we identified 13,124 one-to-one gene orthologs among *D. melanogaster*, *D. simulans*, and *D. sechellia*. These genes were used to establish anchors for reciprocal principal component analysis (RPCA)-based integration across datasets from all three species (Fig 1C, Methods). Our dataset reproduced the previously described global structure of a brain single-cell transcriptome atlas in *D. melanogaster* [21], where cells are differentiated by their expression of developmental patterning genes (notably, *pros* and *Imp*) (S1A and S1B Fig). Glial cells (*repo*+) form several clusters, while neurotransmitter markers could

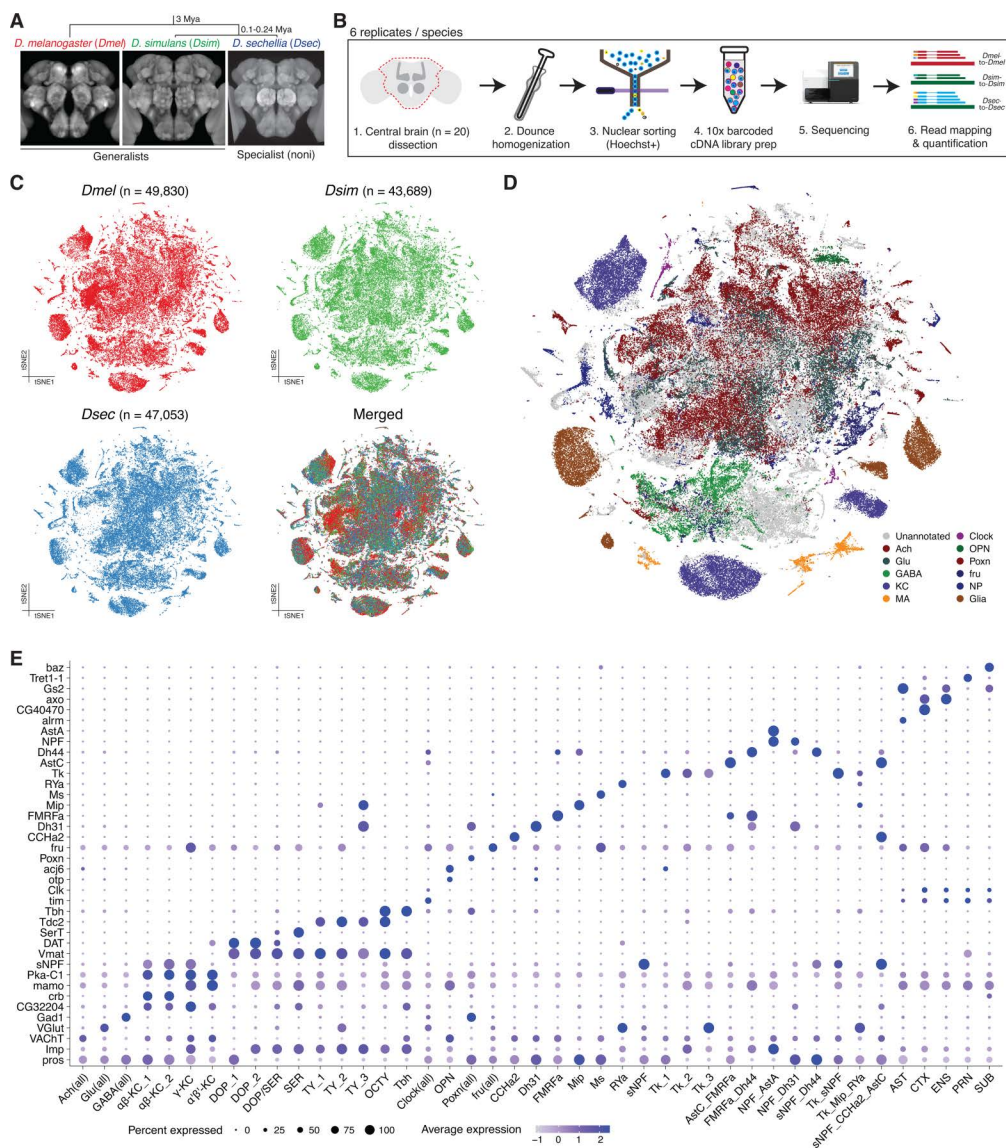


Fig 1. Single-nucleus transcriptomic atlases of *D. melanogaster*, *D. simulans*, and *D. sechellia* central brains. (A) Top: phylogeny of the drosophilid species studied in this work. Mya = Million years ago. Bottom: images of reference central brains for these species (all female; source: *D. melanogaster* [74]; *D. simulans* [75]; *D. sechellia* [76]). (B) Workflow of the single-nucleus RNA-sequencing of drosophilid central brains. (C) tSNE plots of *D. melanogaster* (red), *D. simulans* (green) and *D. sechellia* (blue) central brain cells from an integrated dataset after RPCA integration. In the bottom right plot, all cells from the three species are merged. (D) tSNE plot of the integrated and annotated datasets. Cells are colored by the 11 annotation groups. Unannotated cells are colored gray. (E) Dot plot summarizing the expression of genes used for the annotation of 107 cell types. Cholinergic, Glutamatergic, GABAergic, Clock, Poxn, and fru cell sub-types are merged into Ach (all), Glu (all), GABA (all), Clock (all), Poxn (all), and fru (all), respectively. Gene expression in this and other similar plots is shown as SCTransform-normalized Pearson residuals. The data underlying this figure can be found in <https://zenodo.org/records/15016454>.

<https://doi.org/10.1371/journal.pbio.3003120.g001>

distinguish GABAergic (*Gad1*+), monoaminergic (*Vmat*+), cholinergic (*VaChT*+), and glutamatergic (*VGlut*+ neurons (S1C and S1D Fig).

To characterize cell type diversity in the central brain, we performed iterative clustering and marker-based annotation (S2 and S3 Figs). After this iterative process, we classified clusters into eight groups (glia, Kenyon cells, monoaminergic neurons, Clock neurons, olfactory projection neurons [OPN], Poxn neurons, fru neurons, neuropeptidergic neurons);

the remaining clusters were grouped based on their neurotransmitter type (cholinergic, glutamatergic, GABAergic) (Fig 1D). For each group, we subclustered the cells and manually annotated subclusters based on marker gene expression, except the OPN that could not be further subclustered. In parallel, we disintegrated the dataset into three species-specific datasets to validate that clusters and their marker gene expression patterns are consistent across all datasets (Methods). Through this process, we identified 107 annotated cell types that are conserved across all three species (Fig 1E and S4 Fig). These cell types include five glial cell types, four Kenyon cell types, nine monoaminergic cell types, three Clock cell types, one OPN cell type, six Poxn cell types, 15 fru cell types, 18 neuropeptidergic cell types, 17 cholinergic cell types, 15 glutamatergic cell types, and 14 GABAergic cell types. In addition to previously characterized cell types [21,22], the larger size of our combined datasets provided power to identify novel, rare cell types. For example, we identified a cluster expressing the Myosuppressin (Ms) neuropeptide, which comprises less than 0.08% of central brain cells in all three species (S2 Table). These annotated cell types cover 63.8% of the total cells in the dataset, which were used for the downstream analysis. Our single-cell transcriptomic atlases show that the cellular composition of the central brain is globally conserved across *D. melanogaster*, *D. simulans*, and *D. sechellia*, consistent with their overall similarity in gross neuroanatomy (Fig 1A) [43]. Although we did not identify any examples of annotated cell types that are unique to any species, or absent in only one species, we cannot exclude that some examples might exist in the currently unannotated cells.

Conserved gene expression patterns

We first exploited our atlases to identify genes with conserved expression patterns—beyond the markers used in cell type annotation—which we reasoned would reveal the core molecules with essential roles in global brain organization and cellular functions. To identify such genes, we performed correlation analyses of cell type-specific gene expression levels across three pairwise species comparisons: *D. melanogaster*-*D. simulans* (*Dmel-Dsim*), *D. melanogaster*-*D. sechellia* (*Dmel-Dsec*), and *D. simulans*-*D. sechellia* (*Dsim-Dsec*). We focused on the 2,405 genes that are expressed in at least 5% of *D. melanogaster* central brain cells. The majority of these genes are broadly expressed in 5 to all 107 annotated cell types (on average, ~97 cell types)—where expression is counted if a gene is detected in at least 5% of the cells of a given type. For each of these genes, we computed the average expression level across 107 annotated cell types and compared these cell type-wide expression patterns across species, subsequently measuring the pairwise correlations (S5A Fig). This analysis identified 413 genes displaying strongly correlated expression patterns (Spearman's $p > 0.7$) across all three species pairings (S5B Fig and S3 Table). Gene Ontology (GO) term analysis revealed a notable enrichment in genes coding for membrane proteins, including adhesion molecules, ion channels, and G-protein-coupled receptors (GPCRs) (S5C Fig).

Next, we refined our analysis to genes that are more specifically expressed, positing that these are likely to govern cell type identity or underlie cell type-specific functions. Within each cell type, we identified genes that are consistently expressed in more than 30% of cells across all three species. Of these, we focused on the genes that are restricted to only 1–21 of the 107 annotated cell types. With this approach, we cataloged 925 genes exhibiting both conserved and specific expression patterns (S4 Table). This list of genes serves as a valuable resource for linking the functional roles of distinct brain cell types to their unique gene expression profiles. We note that although the chosen thresholds of gene expression specificity were arbitrary in this analysis, these datasets (S3 and S4 Tables) can be further mined with other thresholds.

The strong conservation of gene expression patterns across species implies the existence of shared gene regulatory mechanisms. While our profiling of mature adult brains is likely to limit our ability to identify important developmental genes, we reasoned that our datasets should still capture information on the expression patterns of terminal selector transcription factors, which establish and maintain the identity of post-mitotic neurons [44,45]. Of the 925 genes displaying specific and conserved expression patterns described above, 104 encode known or predicted transcription factors (S6 Fig and S4 Table). This group includes key developmental regulators, such as a POU domain transcription factor (*acj6*) and a paired-like homeobox transcription factor (*ey*), essential for the development of OPN and Kenyon cells, respectively).

[46–48]. These examples suggest the potential of other genes in this set to play critical roles in regulating cell type identity. For example, another paired-like homeobox transcription factor (*Ptx1*), which is expressed in a few cell types including *Poxn_2* and several *fru*⁺ cell types (S6 Fig), might be a terminal selector for these cell types, similar to its known role for enteroendocrine cell specification in the gut [49]. Furthermore, expression patterns of these transcription factors group cell types with shared functions and/or developmental origins (e.g., glia, Kenyon cells) (S6 Fig). Together, these analyses define a unique molecular fingerprint for each cell type's terminal identity and offer a resource for identifying previously unknown terminal selectors specific to each cell type (S4 Table).

Divergence in the frequencies of homologous cell types

Having determined conserved types and molecular properties of likely homologous cell populations in the drosophilid trio, we next explored if and how these species' brains have diverged. Taking advantage of the statistical power afforded by having six biological replicates per species, we first analyzed interspecific variation in the representation of the 107 cell types. The majority of cell types displayed similar frequencies across the three species, including Kenyon cells, Clock neurons, and *fru* neurons (which play critical roles in learning and memory, circadian rhythms, and sexual behaviors, respectively) (Fig 2A and 2B and S7 Fig). From the comparison of *D. sechellia* with the other species, three cell types displayed significant variation (Fig 2A–C): there are significantly fewer perineurial glial cells (PRN)—which comprise the blood–brain barrier (BBB)—while one cholinergic cell type (*Ach_1*) and one *Poxn*-producing cell type (*Poxn_1*) are expanded. From the comparison of *D. simulans* with the other species, we found only one cell type, expressing the neuropeptide Tachykinin (*Tk_1*), that was significantly reduced (Fig 2A and 2B), while no cell types were reduced or expanded specifically in *D. melanogaster*.

To assess the observed differences in the cell composition of the *D. sechellia* brain atlases *in vivo*, we focused on PRNs, as the location of the *Ach_1* and *Poxn_1* subtypes of neurons amongst the much larger population of all *Ach* and *Poxn* neurons is unknown. We performed hybridization chain reaction fluorescence in situ hybridization (HCR FISH) experiments for *Tret1-1*, a specific marker for PRNs [50]. *Tret1-1* transcripts are broadly distributed across the PRN cell layer (likely because the transcripts are present throughout the cytoplasm of these large cells), which prevented direct quantification of PRN population size. However, as PRN *Tret1-1* RNA levels are not different between species (S8 Fig), we reasoned the total *Tret1-1* signal in the central brain should be a measure of PRN cell number. In *D. sechellia*, we found the global *Tret1-1* RNA signal is significantly lower, consistent with a reduction in PRN population size when compared to both *D. melanogaster* and *D. simulans* (Fig 2D and S9 Fig). To determine whether this reflects a reproducible interspecific difference, we performed *Tret1-1* RNA FISH in an independent strain of each of these species (not used in the snRNA-seq), observing the same lower signal in *D. sechellia* (Fig 2E). Finally, while PRNs were the only glial cell population with a significant decrease in representation in *D. sechellia*, we noted a similar trend for other types of glia (ensheathing [ENS] and cortex [CTX]) (S7 Fig), as well as all glia considered together (Fig 2B). We validated the latter prediction *in vivo* through immunofluorescence for the pan-glial nuclear marker, Repo, which revealed a lower number of glial cells in *D. sechellia* (Fig 2F and S9 Fig). These results suggest that our comparative single-cell atlases can provide reliable predictions for the interspecific variation in cellular composition of the central brains (as well as differences in cell population size within a species); in this context, we note that a snRNA-seq atlas of the *D. melanogaster* antenna that we generated in an independent study is remarkably representative of neuronal population size [51]. Our results suggest that the cellular composition of the central brain in *D. sechellia* is highly conserved with its generalist relatives, but a few cell types might have changed in number during ecological specialization.

Transcriptomic divergence in homologous cell types

Given the global conservation of the types and frequencies of cells in these drosophilids' brains, we reasoned that phenotypic differences between these species might be reflected more prominently in the divergence of the transcriptomic profile of homologous cell types. We identified the 50 most abundantly expressed genes from each of the 107 cell types in

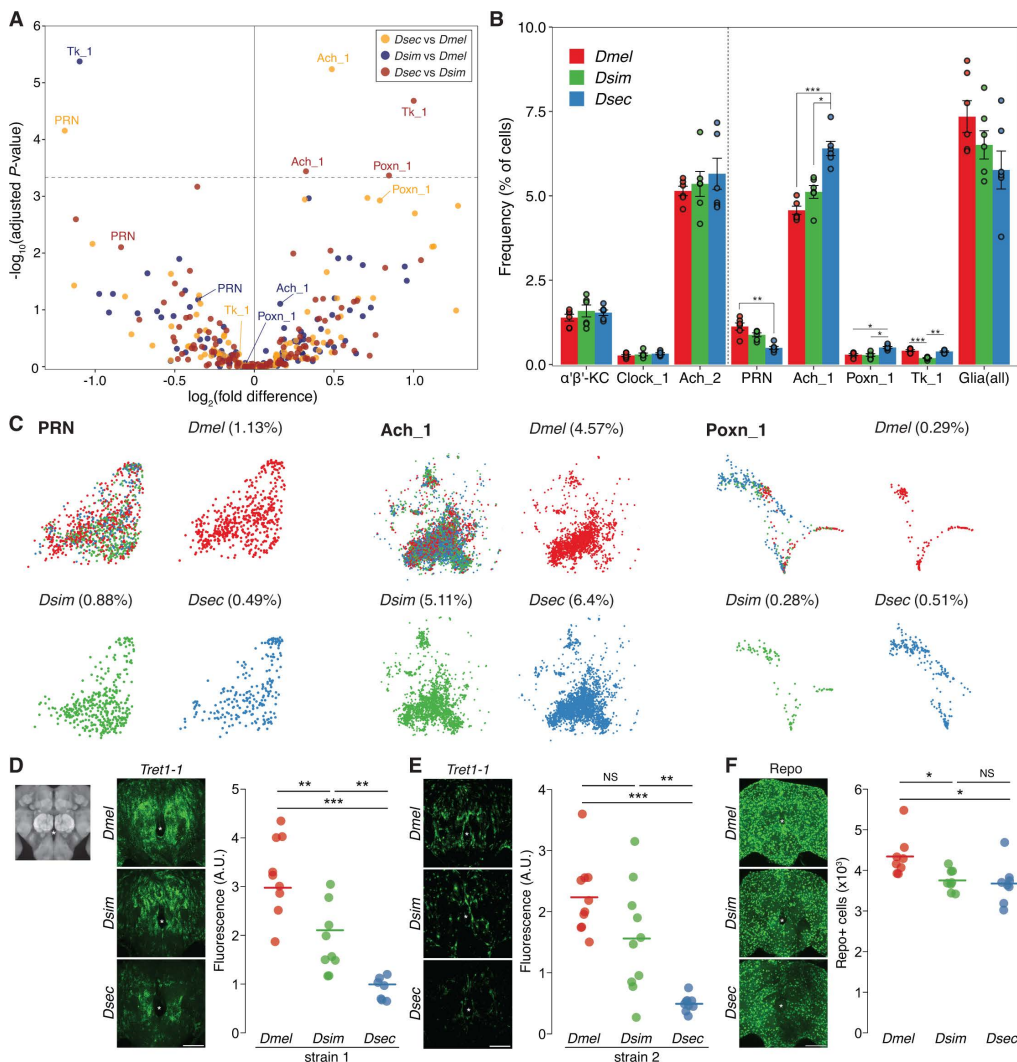


Fig 2. Species divergence in the frequencies of central brain cell types. (A) Volcano plot of the pairwise comparisons of cell type frequencies in the central brains of the three drosophilid species. The dashed horizontal line denotes the Bonferroni-corrected *P*-value threshold. (B) Bar plot illustrating selected cell type frequency comparisons. To the left of the dashed vertical line are three cell types with no significant differences across species. To the right, frequencies of four cell types exhibiting significant differences in at least one pairwise species comparison; far right, combined frequencies of all glial cell types. Each point corresponds to one of the six biological replicates. Statistical significance was calculated by one-way ANOVA with Tukey's HSD post-hoc test. *P*-values were adjusted using false discovery rate (FDR) for multiple comparisons. ****P* < 0.001; ***P* < 0.01; **P* < 0.05. (C) tSNE plots of PRN, Ach_1 and Poxn_1 populations. Cells are colored by their species of origin. The frequency of each cell type within its respective species is indicated in parentheses. (D) Left, representative images of *Tret1-1* HCR FISH on whole-mount brains of *Dmel*CS, *Dsim*04, and *Dsec*07 (see Methods for genotypes). The field of view of the brain in these and all other confocal images is shown on the far left (a cropped version of the *D. sechellia* reference brain from Fig 1A); the asterisk indicates where the esophagus would run, used as a central anatomical landmark. Right, quantifications of *Tret1-1* signals for each species. (E) Left, representative images of *Tret1-1* HCR FISH on *Dmel*OR, *Dsim*196, and *Dsec*28. Right, quantifications of *Tret1-1* signals for each species. (F) Left, representative images of Repo immunofluorescence on *Dmel*CS, *Dsim*04, and *Dsec*07. Right, quantifications of Repo+ cell counts for each species. The data underlying this figure can be found in <https://zenodo.org/records/15016454>.

<https://doi.org/10.1371/journal.pbio.3003120.g002>

D. melanogaster and examined the transcriptomic similarity between species for homologous cell types using correlation analysis (Fig 3A–3D). The similarity of gene expression profiles varies across cell types. For example, gene expression levels are highly similar in a Clock cell type (Clock_3) (*Dmel*-*Dsim*: *p* = 0.670, *Dmel*-*Dsec*: *p* = 0.75, *Dsim*-*Dsec*: *p* = 0.86) compared to OPN (*Dmel*-*Dsim*: *p* = 0.43, *Dmel*-*Dsec*: *p* = 0.39, *Dsim*-*Dsec*: *p* = 0.61). Notably, we found that many cell

types show reduced similarity in the *Dmel-Dsec* pair compared to the *Dmel-Dsim* pair (Fig 3B–3D), suggesting that gene expression profiles of these cell types have diverged more in the *D. sechellia* lineage than the *D. simulans* lineage.

To examine transcriptomic divergence at higher resolution, we characterized differentially expressed genes (DEGs, fold-change threshold >1.5, adjusted *P*-value < 0.05) within each cell type across the three species (*Dmel-Dsec*: *n* = 593, *Dmel-Dsim*: *n* = 487, *Dsim-Dsec*: *n* = 335) (S5 Table). Different cell types exhibited different DEGs, as illustrated for PRN and Poxn_1 cells (Fig 4A and 4B). Importantly, the vast majority of DEGs were identified as divergently expressed from only 1–3 cell types (Fig 4C). For example, among 593 DEGs from the *Dmel-Dsec* pair, 290 genes (48.9%) were identified from a single cell type, and 459 genes (77.4%) were identified from no more than three cell types, even though the majority of these genes are broadly expressed (on average, expressed in 79/107 cell types in *D. melanogaster*, minimum percent expression threshold: 5%). We next examined the cell type identity for each DEG (Fig 4D). Notably, in species comparisons with *D. sechellia* (*Dmel-Dsec* and *Dsim-Dsec*), four glial cell types (CTX, PRN, ENS, and astrocyte-like [AST]) consistently exhibited the highest number of DEGs. The higher DEG counts observed in glial cell types are not driven solely by their relatively large cell numbers (S10 Fig), suggesting that glial cell types display more divergent gene expression profiles than neurons in *D. sechellia*. Among the annotated neuronal cell types, a Poxn cell type (Poxn_3) expressing the neuropeptide Allostatin A (AstA) displayed the greatest number of DEGs in all three species comparisons. The observed cross-cell type and cross-species variation in transcriptomic divergence suggests that each cell type has undergone unique changes during the evolution of these species.

Unique gene expression changes in *D. sechellia*

To investigate gene expression changes potentially related to the ecological specialization of *D. sechellia*, we analyzed the overlap of DEGs from different species comparisons. Sixty percent of DEGs were identified in more than one comparison (Fig 4E and 4F). For example, a set of 195 DEGs appeared in both *Dmel-Dsec* and *Dmel-Dsim*, but not in *Dsim-Dsec*; these shared DEGs might reflect expression changes originating either in the *D. melanogaster* lineage or in the common ancestor of *D. simulans* and *D. sechellia* (Fig 4G). Importantly, we observed 22% more DEGs from *Dmel-Dsec* (*n* = 593) than *Dmel-Dsim* (*n* = 487), indicating that the *D. sechellia* lineage has gained more gene expression changes (Fig 4E). When we inferred lineage-specific gene expression changes by investigating overlaps among DEGs (Methods), we found that the estimated number of *D. sechellia* DEGs (*n* = 96–258) far exceeded *D. simulans* DEGs (*n* = 48–152) (Fig 4G). Given that a similar number of DEGs would be expected if expression differences were solely due to divergence time (“neutral changes”), this observation suggests that the majority of DEGs in the *Dsim-Dsec* pair result from expression changes specific to the *D. sechellia* lineage.

We next focused on the 96 genes with expression changes specific to the *D. sechellia* lineage (S6 Table). Although we performed GO enrichment analysis on these 96 genes [52], the resulting terms were often supported by only one or two genes, limiting interpretability. Therefore, we relied primarily on manual curation to evaluate the putative or known functions of these genes. In neuronal cell types, we identified 54 *D. sechellia* DEGs including those encoding potassium channels (*SK*, *Slo2*), a vesicular glutamate transporter (*VGlut*), a calmodulin (*Cam*), an ecdysone-induced EF-hand (putative Ca²⁺-binding) protein (*Eip63F-1*), an adenylate cyclase (*Ac3*) and at least four cell adhesion molecules for synapse assembly (*Cals*, *dpr1*, *dpr9*, and *side-IV*). One fru cell type (fru_5) and one Poxn cell type (Poxn_3) show the largest number of DEGs (*n* = 9), followed by an *NPF*-/*AstA*-positive cell type (*NPF_AstA*, *n* = 8) (Fig 5A).

Among glial cell types, we identified 42 *D. sechellia* DEGs in four glial cell types (CTX, AST, ENS, and PRN) (Fig 5A), consistent with the pronounced interspecific gene expression variation in glia (Fig 4D). These glial DEGs are enriched with various metabolic genes (Fig 5B and S11 Fig), including catalase (*Cat*), triglyceride lipase (*dob*), NAD(+) hydrolase (*sarm*), L-amino acid transmembrane transporter (CG4991), glycerophosphocholine phosphodiesterases (CG9394 and CG18135), and carbonic anhydrases (*CAH2* and *CAH3*). Intriguingly, for both of the latter two paralog pairs, we noted opposing expression changes in *D. sechellia*: CG9394 is downregulated but CG18135 is upregulated in AST (S11 Fig), and *CAH2* is upregulated and

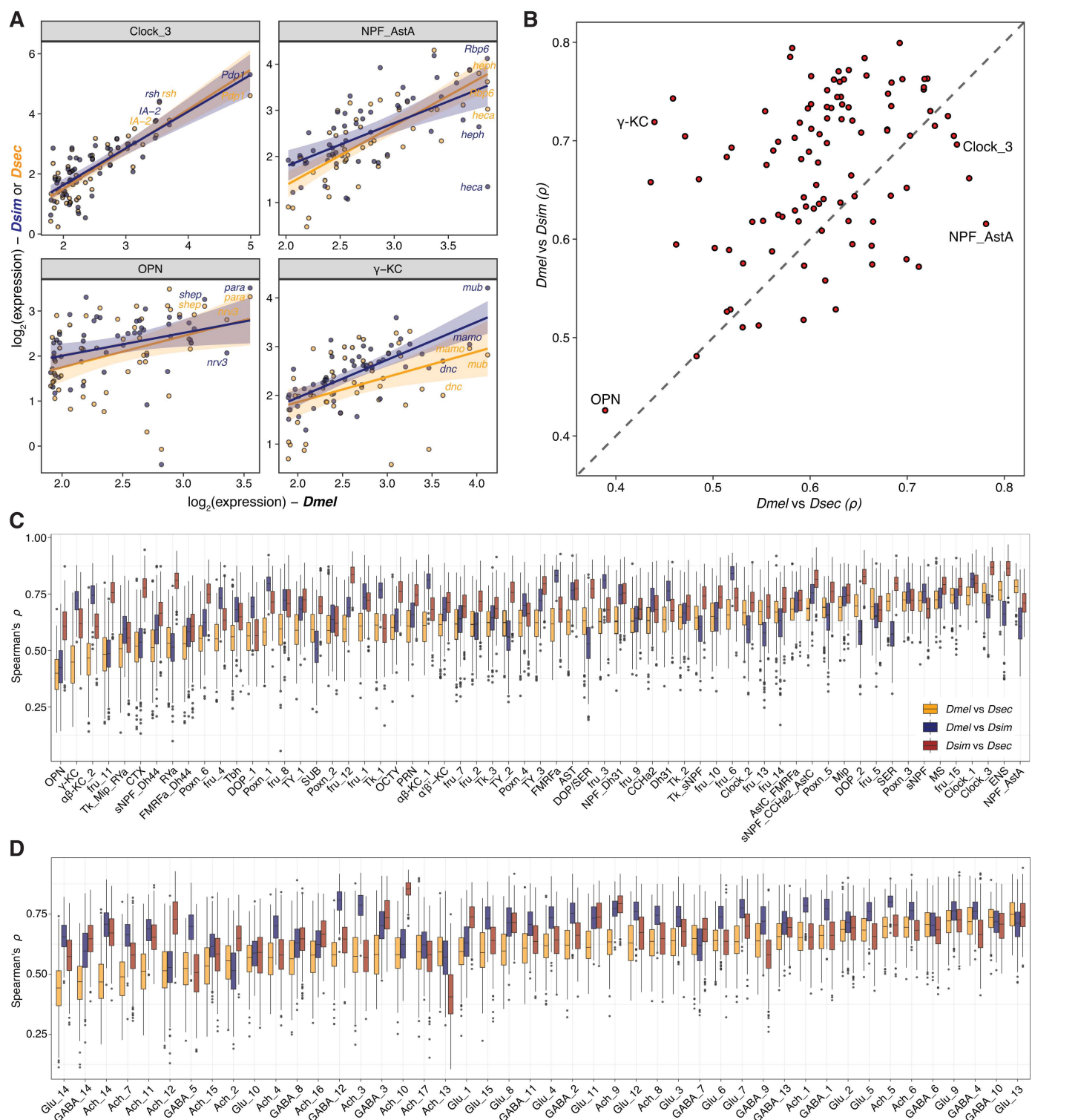


Fig 3. Gene expression divergence of conserved central brain cell types across three drosophilid species. (A) Scatter plots illustrating the interspecific gene expression variation across four distinct cell types: Clock_3, NPF_AstA (NPF-/AstA-positive cells), OPN (olfactory projection neurons), and γ -KC (Kenyon cells). The analysis features the 50 genes (each represented by a dot) with the highest expression levels in the pseudobulk transcriptome of *D. melanogaster* for each cell type. The plots display expression levels of these genes for three drosophilid species: *D. melanogaster* on the x-axis, and *D. simulans* (navy) and *D. sechellia* (orange) on the y-axis. Smoothed lines are linear model fits. (B) Scatter plot illustrating the transcriptomic

similarities among 107 annotated cell types between *D. melanogaster* and *D. sechellia* (x-axis), and *D. melanogaster* and *D. simulans* (y-axis). Each point represents an individual cell type. Spearman correlation coefficients (ρ) are derived from the expression level similarities of the top 50 highly expressed genes in the pseudobulk transcriptome of *D. melanogaster* for each respective cell type. (C,D) Tukey box plots presenting transcriptomic similarities among 107 annotated cell types (glia, Kenyon cells, monoaminergic neurons, Clock neurons, OPN, Poxn neurons, fru neurons and neuro-peptidergic neurons (C); cholinergic, glutamatergic, and GABAergic neurons (D)) in pairwise comparisons of three drosophilid species. The Spearman correlation coefficients (ρ) are calculated based on the expression similarities of 30 randomly selected genes (from 400 permutations) among the top 50 highly expressed genes in *D. melanogaster*'s pseudobulk transcriptome for each cell type. The horizontal line in the middle of each box is the median; box edges denote the 25th and 75th quantiles of the data; and whiskers represent 1.5 $\times$ the interquartile range. The data underlying this figure can be found in <https://zenodo.org/records/15016454>.

<https://doi.org/10.1371/journal.pbio.3003120.g003>

CAH3 downregulated in PRN (Fig 5B). These observations imply potential gene expression dosage compensation between paralogs. From the glial DEGs, we also found the ecdysone-inducible gene *E1* (*ImpE1*), which encodes a protein with low-density lipoprotein receptor type A repeats (Fig 5B). These results suggest that metabolic genes might have played a substantial role in the genetic adaptation to novel conditions in the *D. sechellia* lineage. To validate predicted gene expression differences *in vivo*, we performed HCR FISH for *CAH2* and *ImpE1*. For *andc*, we confirmed, in two independent strains, that gene expression levels are higher in *D. sechellia* compared to *D. melanogaster* or *D. simulans* (Fig 5C–5F and S12 Fig).

Gene expression plasticity in the specialist brain

While these comparative atlases were intentionally generated from flies grown on the same food medium, we considered the possibility that the composition and transcriptome of the *D. sechellia* brain might be influenced by the presence of nutrients in noni. Supplementing food with noni juice paste greatly improved this species' fitness under laboratory conditions, as assessed by egg number and development to the pupal stage (Fig 6A). In parallel with the datasets described above, we also generated a central brain cell atlas from *D. sechellia* that had been reared on standard medium supplemented with noni paste. Comparisons of the snRNA-seq central brain transcriptomes of *D. sechellia* grown with or without noni supplement revealed essentially no effect on the cellular composition of the *D. sechellia* central brain (S13 Fig and S2 Table). Moreover, the examination of gene expression differences using the same threshold that yielded hundreds of interspecific DEGs, we identified only two genes: CG5151, which encodes a protein of unknown function, is more highly expressed in the brains of flies grown on noni; *Mkp3*, which encodes a mitogen-activated protein kinase phosphatase 3, is reduced upon noni supplement (Fig 6B and 6C). Interestingly, CG5151 is broadly expressed but differentially expressed specifically in PRN (S14 Fig), which is one of the cell types with the most significantly changed frequency and gene expression between species (Figs 2A–2E and 4D). To validate predicted gene expression plasticity upon noni supplement, we chose the most significant DEG, CG5151, for HCR FISH experiments. We confirmed that CG5151 is indeed upregulated, albeit at somewhat variable levels, in the central brains of two *D. sechellia* strains fed with noni paste (Fig 6D and S15 Fig).

When we lowered the fold-change threshold (to >1.2), we identified four more noni DEGs (Fig 6B). Three of these (*Cipc*, *SNF4A γ* , and *Treh*) (S16 Fig) are differentially expressed in glial cell types, suggesting that glia are not only diverged in their gene expression between *D. sechellia* and two generalist *Drosophila* species but are also the most responsive to environmental conditions that drove the evolution of *D. sechellia* lineage. For example, expression of a trehalase (*Treh*) gene is higher in AST glia in *D. sechellia* than *D. melanogaster* or *D. simulans* when grown in the same standard medium; but *Treh* expression is decreased in *D. sechellia* grown on noni paste. The other noni DEG, *Vha100-1*, encoding a Vacuolar H⁺ ATPase subunit, is differentially expressed in Poxn_3 neurons (S16 Fig), which is one of the neuronal types with the largest numbers of *D. sechellia* DEGs (Fig 5A).

Discussion

Taking advantage of the small brains of closely related drosophilid species, our work provides an unprecedented, comprehensive view of the molecular and cellular conservation and divergence of animal brains in a well-defined phylogenetic

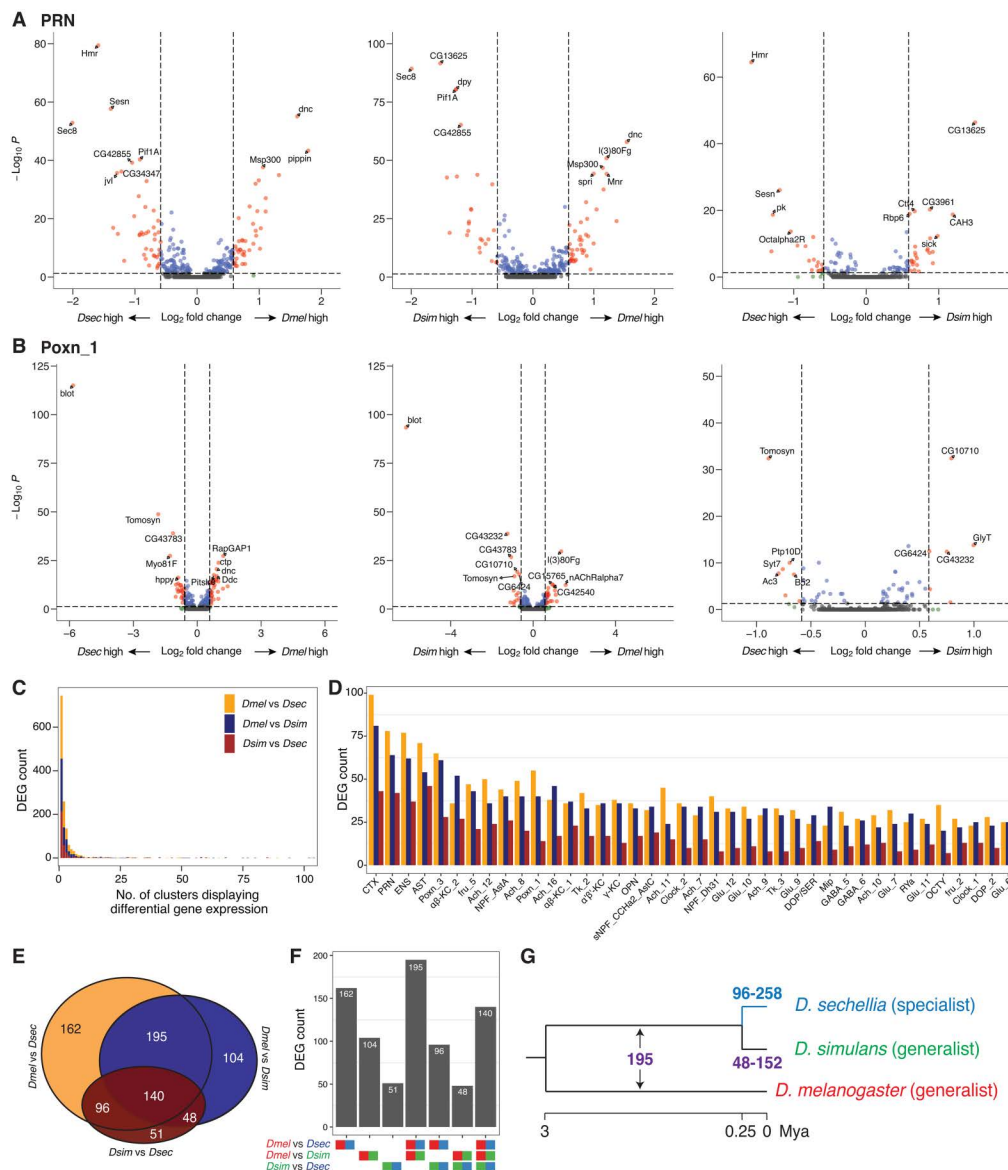
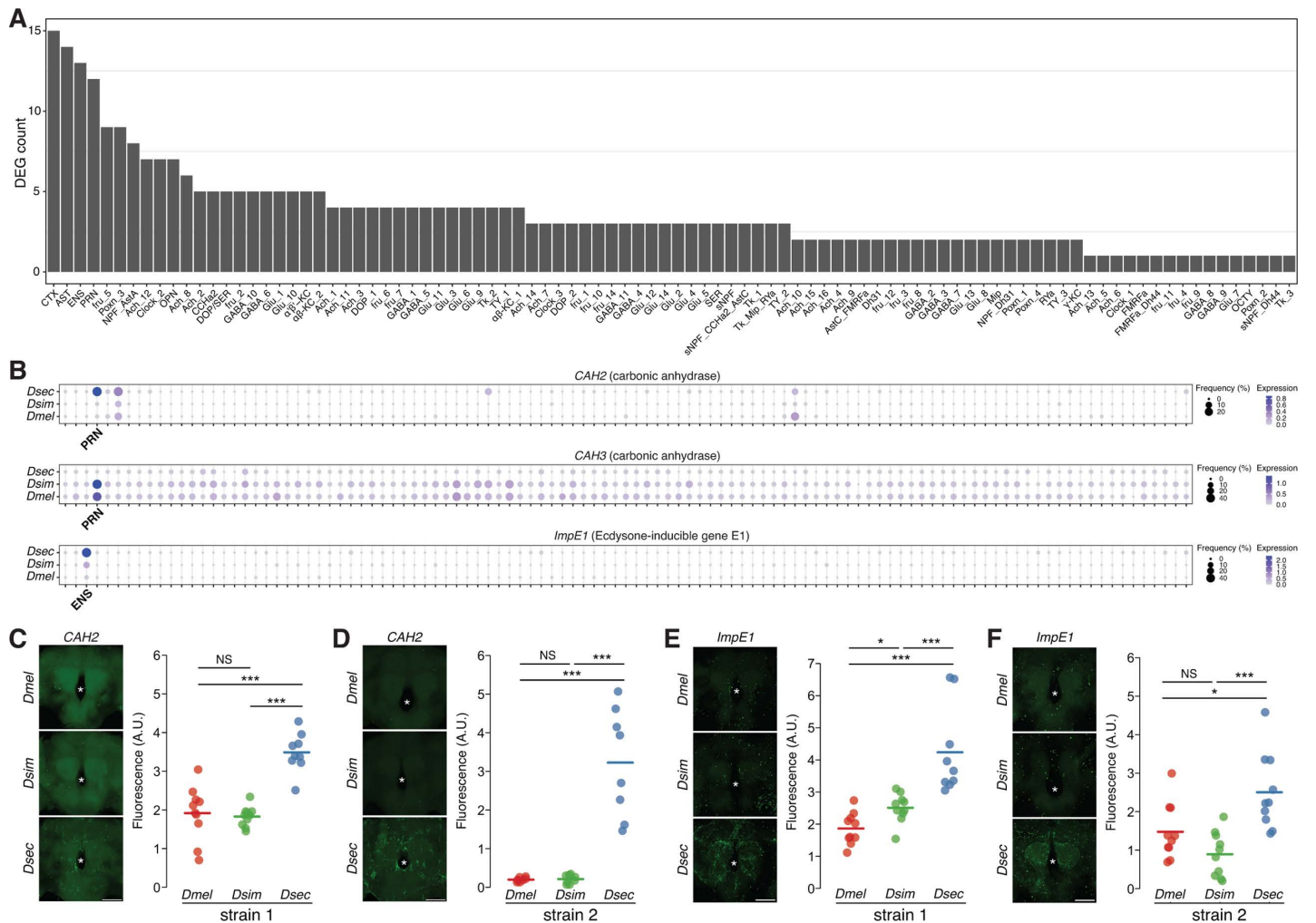


Fig 4. Characterization of differentially expressed genes (DEGs) in conserved cell types across drosophilid central brains. (A,B) Volcano plots showing DEGs (red) in two specific cell types, PRN (A) and Poxn_1 (B), across three pairwise species comparisons. For comparisons with more than 10 DEGs, only the top 10 most significant genes are labeled. (C) Histogram displaying the frequency of DEGs sorted by the total number of cell types/clusters in which they are identified as DEGs. Stacked bars with distinctive colors represent DEGs identified in pairwise species comparisons. (D) A bar plot (color-coded as in (C)) showing the frequency of DEGs across the 40 cell types with the largest number of DEGs from all pairwise comparisons among the three species. Cell types are arranged in order of the total number of DEGs across three pairwise comparisons. (E,F) A Venn diagram (E) (color-coded as in (C)) and a bar plot (F) illustrating the intersections of DEGs identified in three pairwise species comparisons. (G) Hypothetical numbers of genes that underwent expression changes are shown across the lineages of *D. melanogaster*, *D. simulans*, and *D. sechellia*. Mya = Million years ago. The data underlying this figure can be found in <https://zenodo.org/records/15016454>.

<https://doi.org/10.1371/journal.pbio.3003120.g004>

and ecological framework. Previous studies on how the nervous system of *D. sechellia* differs from *D. melanogaster* and *D. simulans* have primarily focused on changes in peripheral chemosensory pathways [28,32–36,40,42], leaving knowledge of potential adaptations in the brain almost completely unexplored. Despite the very different ecology and behaviors of the equatorial island-endemic specialist *D. sechellia* from its cosmopolitan, generalist relatives, the overall



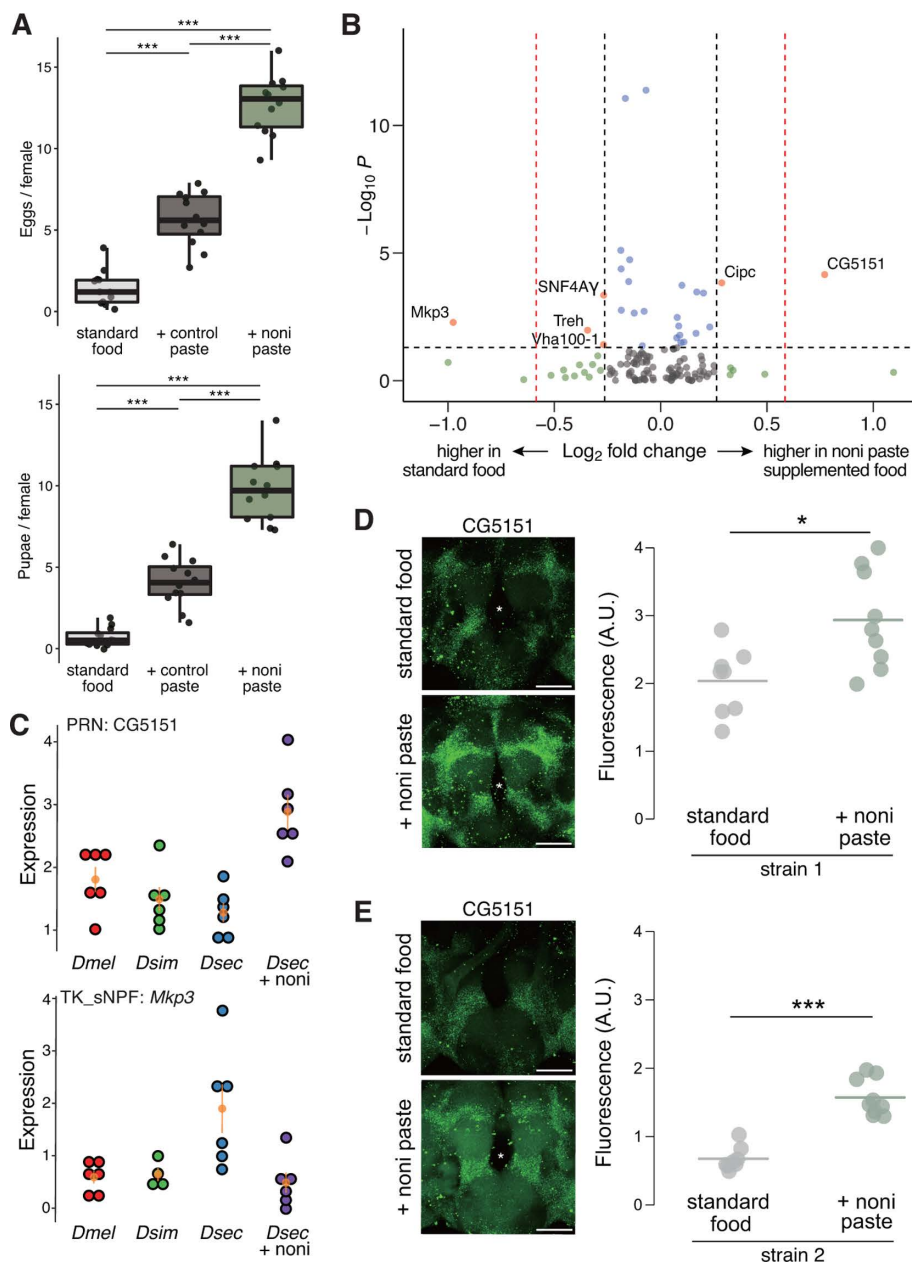


Fig 6. Noni diet influences on *D. sechellia* brain gene expression. (A) Tukey box plots for brood size (left) and pupal count (right) per mated *D. sechellia* female grown in three conditions. Each point corresponds to a biological replicate. The horizontal line in the middle of each box is the median; box edges denote the 25th and 75th quantiles of the data; and whiskers represent 1.5× the interquartile range. Statistical significance was calculated by one-way ANOVA with Tukey's HSD post-hoc test. ****P* < 0.001. (B) Volcano plot highlighting differences in gene expression of *D. sechellia* central brains from animals grown without and with added noni paste. All cell type-specific DEG analyses are integrated into this single representation, and a reduced fold-change threshold (from 50% to 20%) is applied for the identification of DEGs. (C) Dot plots illustrating expression levels of CG5151 in PRN and *Mkp3* in TK_sNPF in the indicated species and growth conditions. (D) Left, representative images of CG5151 HCR FISH on *Dsec07* grown under the indicated conditions. Right, quantifications of CG5151 signals. (E) Left, representative images of CG5151 HCR FISH on *Dsec28* grown under the indicated conditions. Right, quantifications of CG5151 signals. The data underlying this figure can be found in <https://zenodo.org/records/15016454>.

<https://doi.org/10.1371/journal.pbio.3003120.g006>

In light of this conservation, consideration of the central brain cell populations that do differ between species is of interest. The most pronounced difference in *D. sechellia*'s central brain composition was found not in neurons but in a specific glial type, PRN, which forms a diffusion barrier around the nervous system as part of the BBB. In *D. melanogaster*, PRNs have an important role in sugar uptake into the brain [50]. Interestingly, noni fruit has a much lower carbohydrate (sugar):protein ratio than most other fruits, and *D. sechellia* cannot develop on carbohydrate-rich diets, apparently due to defects in carbohydrate-induced metabolic and gene expression changes [54]. We speculate that the reduction in PRNs in this species reflects a reduced requirement for sugar uptake by the brain. While it is unclear whether this loss is adaptive in *D. sechellia* (i.e., increasing fitness of this species in another, unknown, way), it might be a phenotype that constrains this species to this niche (i.e., representing an “evolutionary dead end” [55]). Only two neuron populations are expanded in *D. sechellia*—subtypes of cholinergic neurons and Poxn neurons—but as their roles even in *D. melanogaster* are unknown, it is difficult to interpret the significance of such expansions. Future exploration of these species differences *in vivo* might provide insights into as-yet poorly characterized central brain types in drosophilids.

Within homologous cell types, global patterns of gene expression—spanning those encoding known (or newly predicted) terminal transcription factors to signaling effectors, such as neurotransmitters/peptides—are, as expected, conserved. However, we observed many examples of species-specific divergence in gene expression, which was different in each cell type in terms of gene identity and magnitude of change. In particular, many broadly expressed genes differ in their expression between species only in a subset of specific cell types. Such properties reinforce the idea of cell types being independent evolutionary units [1]. Of interest is our observation that the majority of glial cell types display the greatest expression divergence. These properties might reflect lower selective pressures on glia to maintain precise structure and function compared to neurons that act within stringently defined circuitry. However, the enrichment of gene expression changes in glia might also reflect adaptive changes within these cell types, which have numerous known or presumed supportive roles in the brain, such as in homeostatic regulation and filtering of the external milieu for energy supply to neurons and waste removal [56].

It is important not to interpret every change in gene expression as indicative of adaptive evolution: after species divergence, neutral genetic polymorphisms accumulate, raising the possibility that gene expression also undergoes neutral evolution [57,58]. The drosophilid trio offers a useful model system to distinguish neutral and potentially non-neutral gene expression evolution. Under neutral conditions, the expectation is for *D. simulans* and *D. sechellia* to exhibit a similar degree of gene expression divergence when compared to *D. melanogaster*. However, our data revealed more pronounced shifts in brain gene expression in *D. sechellia*, suggesting that this lineage underwent non-neutral gene expression changes, possibly due to adaptation to its unique ecological niche. Glia stand out as the cell types displaying the most potential adaptive expression alterations, reinforcing the notion that they have an important contribution to brain evolution. Among the genes displaying altered expression patterns, those associated with metabolism are particularly enriched. The metabolic gene regulatory network might have undergone remodeling to adapt to the new nutritional conditions of the species' unique niche. An alternative hypothesis, however, is that *D. sechellia*'s smaller effective population size [59] reduces the efficacy of purifying selection, allowing neutral or slightly deleterious mutations to accumulate and greater expression divergence. Furthermore, the observed inverse expression changes in certain paralog pairs in *D. sechellia* (e.g., CG9394/CG18135 and *CAH2/CAH3*) suggest potential compensatory mechanisms that might further exacerbate divergence in gene expression profiles. Future analyses of signatures of selection on gene expression with broader phylogenetic sampling will be key to distinguishing adaptive and non-adaptive scenarios.

One challenge of comparing transcriptomes of ecologically diverse species is distinguishing between genetic and environmental effects on gene expression. To minimize environmental influences, our initial comparative atlases were of the drosophilid species grown on an identical food medium. However, as *D. sechellia* is thought to have metabolic defects [31,54], we also generated a brain atlas from flies grown with noni juice. Although this medium substantially increased *D. sechellia*'s survival and fecundity, there was remarkably little influence upon the cellular composition or gene expression

patterns in the brain of this species. These results support the idea that the interspecific differences observed are largely genetically determined. In contrast to our findings, a previous bulk RNA-seq analysis of whole flies identified several hundred DEGs upon noni exposure [60], suggesting that the brain might be “spared” [61] from nutrient-dependent gene expression changes compared to other tissues. Of the few observed environmental-dependent gene expression changes in the brain, the majority occur in various types of glia, emphasizing the plasticity of these cell types both over evolutionary timescales and in response to environmental changes. It is most noteworthy that the gene displaying the largest differential expression in *D. sechellia* cultured with and without noni (CG5151) marks the main cell type (PRN) displaying differential population size between species. These observations hint at the possibility that the cellular nodes in the brain that initially responded to the environmental presence of noni in the *D. sechellia* ancestor eventually have been reshaped by natural selection during species-specific niche adaptation [62].

Finally, we acknowledge that our atlases almost certainly do not capture the full extent of interspecific differences in gene expression. Although several lab-based studies of the peripheral olfactory system identified several olfactory sensory neuron populations that are increased or decreased in *D. sechellia* compared to the other species [32,33,40,42], it is possible that some interspecific variation in the brain remains cryptic under these controlled conditions. Additional experiments under more natural ecological contexts could reveal other brain-related adaptations. Furthermore, because our primary goal was to focus on species-specific differences and minimize variation from other biological factors, including sex, we analyzed only females in this study. The choice of sex was arbitrary, but in part guided by our focus on females in many of our prior investigations (e.g., noni attraction [33,42] and oviposition [34]). In the future, comparative studies on males will reveal sex-specific species differences (or those common to both sexes) in drosophilid central brains. In addition, as technical limitations of scRNA-seq currently constrain our ability to detect changes in rare cell types or genes with low expression, future deeper sequencing efforts and larger-scale single-cell atlases might reveal further differences. Beyond these possible extensions to the current work, broader sampling of diverse drosophilid species will be essential to situate changes in the *D. sechellia* lineage within a more expansive comparative framework. Such an integrated approach will advance our understanding of how neural circuits adapt to novel ecological niches and elucidate the fundamental mechanisms driving brain evolution.

Methods

Drosophila strains and culture

Flies used in this study were *D. melanogaster* (Canton-S (*DmelCS*), Oregon-R (*DmelOR*)), *D. simulans* (*Drosophila* Species Stock Center [DSSC] 14021-0251.004 (*Dsim04*), 14021-0251.196 (*Dsim196*)), and *D. sechellia* (DSSC 14021-0248.07 (*Dsec07*), 14021-0248.28 (*Dsec28*)), which were all cultured on standard wheat flour–yeast–fruit juice food. *DmelCS*, *Dsim04*, and *Dsec07* were used to generate the atlases. For the animals analyzed in Fig 6, *Dsec07* was also grown on standard food supplemented on top with a paste of Formula 4–24 Instant *Drosophila* medium (blue, Carolina Biological Supply) mixed 1:5 weight:volume in noni juice (Raab Vitalfood).

Single-nucleus RNA sequencing (snRNA-seq)

Thirty to fifty newly eclosed male and female flies were collected and placed together in food vials for mating, followed by sorting by sex on day 5. The central brains of 20 mated female flies were dissected and collected in 100 µl Schneider’s medium, flash-frozen in liquid nitrogen, and stored at –80°C for nuclear extraction. Sample homogenization, single-nucleus suspension, and nuclear sorting were performed using the protocol described for the Fly Cell Atlas [23]. To obtain sequencing data from 10,000 nuclei, 15–20,000 nuclei were collected and loaded onto the Chromium Next GEM Chip (10x Genomics). Sequencing libraries were prepared with the Chromium Single Cell 3’ reagent kit v3.1 dual index, strictly following the manufacturer’s recommendations. Libraries were quantified by a fluorimetric method, and their quality was assessed on a Fragment Analyzer (Agilent Technologies). Cluster generation was performed with 0.8–1.0 nM of

an equimolar pool from the resulting libraries using the Illumina HiSeq 3000/4000 PE Cluster Kit reagents. Sequencing was performed on the Illumina HiSeq 4000 using HiSeq 3000/4000 SBS Kit reagents according to 10× Genomics' recommendations (28 cycles read1, 8 cycles i7 index read, 8 cycles i5 index, and 91 cycles read2). Sequencing data were demultiplexed using the bcl2fastq2 Conversion Software (v2.20, Illumina). We performed six biological replicates for *D. melanogaster*, *D. simulans*, *D. sechellia* grown on standard food, and *D. sechellia* grown on noni-supplemented food (with each replicate for the four genotypes/conditions processed in parallel) resulting in snRNA-seq data from a total of 120 central brains per genotype/condition.

Generation and integration of single-cell central brain transcriptomic atlases

Raw snRNA-seq data was first processed through Cell Ranger (v7.1.0, default parameters except `--include-introns`) [43]. The *D. melanogaster* reference genome and transcriptome from FlyBase (release 6.55) were used for all three species (*Dmel-to-Dmel*, *Dsim-to-Dmel*, *Dsec-to-Dmel*); sequence reads from *D. simulans* and *D. sechellia* were also processed with their own reference genomes (*D. simulans*: Prin_Dsim_3.1, *D. sechellia*: ASM438219v2) (*Dsim-to-Dsim*, *Dsec-to-Dsec*). Gene expression matrices from *Dsim-to-Dmel* and *Dsec-to-Dmel* processing were used for gene expression analysis (see below). All gene expression matrices were processed with SoupX (v1.6.2, default parameters) [63] to remove ambient RNA contamination. Decontaminated *Dmel-to-Dmel*, *Dsim-to-Dsim*, and *Dsec-to-Dsec* datasets were then normalized through SCT normalization [64], and putative doublets were filtered out using DoubletFinder (v2.0.4) [65]. Decontaminated and filtered datasets were subsetted for 13,124 one-to-one-to-one orthologs among reference genomes of *D. melanogaster*, *D. simulans*, and *D. sechellia*, which were inferred using Orthofinder [66] and manual curation of reciprocal best blastp hits; subsetted datasets were then integrated through the RPCA based integration method implemented in Seurat (v4.4.0, functions `SelectIntegrationFeatures` (nfeatures = 3000), `PrepSCTIntegration`, `FindIntegrationAnchors` (reference = *D. melanogaster*), and `IntegrateData` [67].

Iterative subclustering

Principal component analysis (PCA) was performed on the integrated dataset; the first 50 PCs were used for clustering of cells in the integrated dataset into 36 clusters (functions `FindNeighbors` and `FindClusters`, default parameters except for resolution = 0.2). These clusters were further subjected to iterative subclustering. First, cells from each initial cluster were re-integrated at the cluster level. Subsequently, putative doublet clusters were identified and removed. Doublet clusters were defined as those exceeding 50% of the standard deviation of feature RNA counts, as these likely represent artifacts with an abnormally high transcript count. Following doublet filtering, the sizes of the resulting subclusters were evaluated. If a parent cluster did not segregate into at least two subclusters with frequencies >0.1% of the total cell population, iterative subclustering was terminated for that cluster. If there were at least two subclusters with frequencies >0.1%, the following steps were performed: (1) subclusters with frequencies <0.1% were reassigned to their parent cluster, (2) for subclusters with 0.1%–0.5% frequencies, the cells were assigned to the subcluster but not subjected to further subclustering, and (3) subclusters with >0.5% frequencies were subjected to an additional round of subclustering. This iterative subclustering process, summarized in S2 Fig, led to the definition of 240 clusters.

Cell type annotation and quantification

Marker genes for 240 clusters obtained through iterative subclustering were identified using the `FindMarkers` function in Seurat (only.pos = TRUE, min.pct = 0.15, logfc.threshold = 0.25, test.use = "MAST"). These marker genes were used to classify the clusters. Cluster identities were primarily determined by comparing these marker genes with those identified in previous studies for *D. melanogaster* [21,22]. Any clusters and subclusters that remained unannotated following this step were then identified based on marker genes involved in neurotransmission and neuromodulation [68]. The annotation

process involved sequentially assessing the expression of marker genes associated with specific cell types (S3 Fig), including glial cells, Kenyon cells, monoaminergic neurons, Clock neurons, OPN, Poxn neurons, and fru neurons. For the remaining unclassified clusters, neuropeptide gene expression was examined. Fourteen neuropeptide genes expressed in fewer than 5% of clusters were identified, and clusters expressing these genes were classified as neuropeptidergic neurons. The remaining clusters were further categorized into three groups based on their neurotransmitter production: cholinergic, glutamatergic, or GABAergic. Following this classification, same class clusters were grouped, and cells within each group were re-integrated and subsequently segregated into individual species datasets. PCA and clustering were re-performed for these datasets. Gene expression profiles were then analyzed to identify marker genes specific to the resulting clusters. Only clusters with clear and distinct marker gene signatures were retained and annotated as specific cell types. With this workflow (S2 Fig), we obtained 107 annotated cell types. 63.8% cells belong to these annotated cell types; we only included these cells and annotated cell clusters for the downstream analysis. The frequency of each cell type was determined by quantifying the proportion of cells assigned to a specific cell type relative to the total number of cells in each scRNA-seq experiment, with six replicates per species/condition. The significance of interspecific variation in cell type frequency was calculated by a one-way analysis of variance (ANOVA) test, followed by Tukey's post hoc test for multiple comparisons.

Conserved gene expression analysis

To mitigate artifacts stemming from the use of different reference genomes, we selected genes that exhibited similar expression levels when aligned to multiple genomes. Specifically, reads from *D. simulans* and *D. sechellia* were aligned to both the *D. melanogaster* genome and their own genomes. Subsequently, the percent expression for every gene was calculated across all cells and ranks between the two alignments (one to the *D. melanogaster* genome and the other to their own genomes) were compared. Of 7,926 genes that exhibited less than a 5% rank difference for both *D. simulans* and *D. sechellia* samples, we retained 2,405 genes that were expressed in at least 5% of *D. melanogaster* central brain cells for the analysis. The AverageExpression function in Seurat was employed to calculate the average gene expression of these selected genes for each cell type. Expression levels of these genes across 107 cell types were then compared across species, with the similarity between species being assessed through correlation analysis using Spearman's ρ . FlyEnrichr (<https://maayanlab.cloud/FlyEnrichr/>) was used for the GO analysis on 413 genes with conserved expression patterns [52,69].

Transcriptomic comparisons in homologous cell types

Before comparing gene expression between homologous cell types, we selected genes with reliable expression level information for every cell type by excluding genes whose expression levels in *D. simulans* or *D. sechellia* showed discrepancies (5% rank difference) when their transcripts were aligned to either *D. melanogaster* or their own reference genomes. With these criteria, we omitted 1146 genes (13.8% of analyzed genes). Subsequently, we identified the top 50 most abundantly expressed genes within each of the 107 annotated cell types in *D. melanogaster*. Transcriptomic similarities between species for homologous cell types were assessed through correlation analysis.

Differential gene expression analysis

For cell type-specific DEG analysis, cells were subsetted based on their cell type membership and grouped by species and growth conditions (four groups: *Dmel*, *Dsim*, *Dsec*, and *DsecNoni*). DEGs were identified using the FindMarkers function in Seurat (test.use = "MAST", min.pct = 0.05). To reduce false discoveries of DEGs that might arise from discrepancies or errors in the genome annotation of different species, sequence reads from *D. simulans* and *D. sechellia* were mapped to their own reference genome and to that of *D. melanogaster*. A gene was only designated as differentially

expressed if the mapping of the gene to both genomes consistently indicated differential expression (adjusted P -value < 0.05, fold-difference >50%) between the species or condition. For gene expression plasticity analysis, a lower fold-difference threshold (>20%) was also examined. The gene *Hr38* was excluded from the plasticity analysis because differential expression of this gene was driven by a single replicate, leading to unreliable results.

Lineage-specific gene expression changes

To infer gene expression changes that occurred after the divergence of *D. simulans* and *D. sechellia* lineages, two methodologies were implemented. The initial approach comprised of identifying non-overlapping DEGs between *Dmel-Dsim* DEGs and *Dmel-Dsec* DEGs, under the assumption that these DEGs do not encompass gene expression differences between the *Dmel* lineage and *Dsim/Dsec* lineage before their divergence. This method established the highest counts for *D. simulans* and *D. sechellia* DEGs. The second approach identified the intersection between DEGs of *Dsim-Dsec* and *Dmel-Dsim* or *Dmel-Dsec*. This approach assumed that any *Dsim* or *Dsec* lineage-specific changes would be captured by overlap between *Dsim-Dsec* and *Dmel-Dsim* (*D. simulans* DEGs) or *Dsim-Dsec* and *Dmel-Dsec* (*D. sechellia* DEGs), respectively. This approach established minimum counts for *D. simulans* and *D. sechellia* DEGs.

HCR FISH

Probe libraries for HCR FISH were synthesized by Molecular Instruments, based on the *D. melanogaster* gene sequences (sequence identity was >96% for all genes across all three species). 3–5-day-old females were fixed in 4% paraformaldehyde for 2h prior to brain dissection and HCR FISH was performed essentially as described [70]. Brains were imaged using an inverted confocal microscope (Zeiss LSM 880) equipped with a 40× objective, using fixed settings to maximize the comparability of images. The 40× objective permitted a region of interest containing the entire midbrain but excluding the optic lobes to facilitate direct comparison to the snRNA-seq data (which did not include the optic lobes). Images were captured using Microsoft ZEN 2.3 SP1 software. We quantified total fluorescence in the midbrain as a proxy for RNA abundance; images were processed in Fiji by first creating a maximum-intensity projection z-stack of the midbrain. Relative fluorescence was calculated as the integrated density of pixel intensity. We compared total fluorescence between strains using a Wilcoxon rank-sum test followed by *post hoc* correction for multiple tests. To ensure repeatability of the expression differences we observed, we replicated each experiment twice in our focal strains (*Dmel*CS, *Dsim*04, and *Dsec*07—“strain 1”), as well as another set of wild-type strains of the same species (*Dmel*OR, *Dsim*196, and *Dsec*28—“strain 2”). We did not compare between, or pool, replicates because these flies were dissected, labeled, and imaged separately. Beyond the examples illustrated in this work, transcripts for several other genes were not reliably detected in any species, likely because they are expressed below the detection threshold of this methodology.

Immunofluorescence

To compare total glial cell numbers between species, we performed immunofluorescence for the glial cell marker Repo [71]. 3–5-day-old brains of female flies were fixed, dissected, and stained essentially as described [72], using mouse-anti-Repo (1:20) (8D12) (Developmental Studies Hybridoma Bank AB_528448) and Cy5 goat anti-mouse secondary antibody (1:500) (Molecular Probes, Jackson ImmunoResearch 115-175-166). Images of brains were obtained using an inverted confocal microscope (Zeiss LSM 880) equipped with a 40× objective, using fixed settings. Here, to best approximate the snRNA sequencing protocol, the optic lobes were masked before quantification. Images were processed in Fiji and Repo-positive cells in the midbrain were counted using the Fiji macro, RS-FISH [73]. No signal was detected outside the brain. The Repo-positive cell counts were compared between strains using a Wilcoxon rank-sum test followed by *post hoc* correction for multiple tests. We repeated this experiment twice to ensure replicability of the overall pattern. We did not pool, or compare between, replicates because these flies were dissected, stained, and imaged separately.

Fitness assay

The fitness of *D. sechellia* was measured under three conditions: standard food, standard food supplemented with Formula 4–24 Instant *Drosophila* Medium blue (Carolina Biological Supply) that was hydrated with distilled water at a 1:5 weight-to-volume ratio (“control paste” in Fig 6A), or with noni paste (described above; “noni paste” in Fig 6A). For each condition, 10 mated, 5-day-old females were introduced into a culture vial. Brood sizes were measured after 24 h, followed by counting the number of pupae on day 8–9.

Supporting information

S1 Table. Summary of the metrics from the snRNA-seq experiments.

(XLSX)

S2 Table. Frequencies of 107 cell types in *D. melanogaster*, *D. simulans*, and *D. sechellia*.

(XLSX)

S3 Table. 413 broadly expressed genes (>5% of central brain cells in *D. melanogaster*) with conserved expression patterns (Spearman’s $p > 0.7$).

(XLSX)

S4 Table. 925 genes with specific expression (<22/107 cell types) and conserved expression patterns (expressed in >30% of cells across all three species). The subset encoding predicted/known transcription factors (TFs)

(GO:0000981) are shown separately on the second worksheet.

(XLSX)

S5 Table. Differentially expressed genes identified from pairwise comparisons of three drosophilid species in a cell type-specific manner.

(XLSX)

S6 Table. 96 genes with putative expression changes specific to the *D. sechellia* lineage.

(XLSX)

S1 Fig. Integrated single-cell transcriptomic atlases of *D. melanogaster*, *D. simulans*, and *D. sechellia*. (A,B) tSNE (A) and UMAP (B) plots of the integrated dataset after RPCA integration of single-cell transcriptomes of *D. melanogaster*, *D. simulans*, and *D. sechellia* central brain cells colored by the expression of *pros* or *Imp*. (C,D) tSNE (C) and UMAP (D) plots with cells colored by their predicted molecular identities. Neurons are identified through markers for neurotransmitter production, color-coded according to GABAergic (*Gad1*), monoaminergic (*Vmat*), glutamatergic (*Vglut*), and cholinergic (*VACHT*) designations. Glia, labeled by *repo* expression, are colored pink. Cells failing to meet expression thresholds for these markers are colored gray. The data underlying this figure can be found in <https://zenodo.org/records/15016454>.

(EPS)

S2 Fig. Workflow for snRNA-seq data processing. (A) Workflow of the generation of comparative single-cell transcriptomic atlases (see Methods). (B) Workflow of the iterative subclustering (ISub) (see Methods).

(EPS)

S3 Fig. Marker gene-based cluster classification. Workflow for classifying post-ISub clusters into 11 groups (glia, Kenyon cells, monoaminergic neurons, Clock neurons, olfactory projection neurons (OPN), Poxn neurons, fru neurons, neuropeptidergic neurons, cholinergic neurons, glutamatergic neurons, and GABAergic neurons).

(EPS)

S4 Fig. Cell type annotations in central brains of *D. melanogaster*, *D. simulans*, and *D. sechellia*. (A) Glial cell types. Left, UMAP plots of the glial cell type annotation for integrated and disintegrated single-cell transcriptomic atlases of *D. melanogaster*, *D. simulans*, and *D. sechellia*. Right, expression levels of marker genes distinguishing the glial cell types. (B) Kenyon cell types. Left, UMAP plots of the Kenyon cell type annotation for integrated and disintegrated single-cell transcriptomic atlases of *D. melanogaster*, *D. simulans*, and *D. sechellia*. Right, expression levels of marker genes distinguishing the Kenyon cell types. (C) Monoaminergic cell types. Left, UMAP plots of the monoaminergic cell type annotation for integrated and disintegrated single-cell transcriptomic atlases of *D. melanogaster*, *D. simulans*, and *D. sechellia*. Right, expression levels of marker genes distinguishing the monoaminergic cell types. (D) Clock cell types. Left, UMAP plots of the Clock cell type annotation for integrated and disintegrated single-cell transcriptomic atlases of *D. melanogaster*, *D. simulans*, and *D. sechellia*. Right, expression levels of marker genes distinguishing the Clock cell types. (E) Poxn cell types. Left, UMAP plots of the Poxn cell type annotation for integrated and disintegrated single-cell transcriptomic atlases of *D. melanogaster*, *D. simulans*, and *D. sechellia*. Right, expression levels of marker genes distinguishing the Poxn cell types. (F) fru cell types. Left, UMAP plots of the fru cell type annotation for integrated and disintegrated single-cell transcriptomic atlases of *D. melanogaster*, *D. simulans*, and *D. sechellia*. Right, expression levels of marker genes distinguishing the fru cell types. (G) Neuropeptidergic cell types. Left, UMAP plots of the neuropeptidergic cell type annotation for integrated and disintegrated single-cell transcriptomic atlases of *D. melanogaster*, *D. simulans*, and *D. sechellia*. Right, expression levels of marker genes distinguishing the neuropeptidergic cell types. (H) Cholinergic cell types. Left, UMAP plots of the cholinergic cell type annotation for integrated and disintegrated single-cell transcriptomic atlases of *D. melanogaster*, *D. simulans*, and *D. sechellia*. Right, expression levels of marker genes distinguishing the cholinergic cell types. (I) Glutamatergic cell types. Left, UMAP plots of the glutamatergic cell type annotation for integrated and disintegrated single-cell transcriptomic atlases of *D. melanogaster*, *D. simulans*, and *D. sechellia*. Right, expression levels of marker genes distinguishing the glutamatergic cell types. (J) GABAergic cell types. Left, UMAP plots of the GABAergic cell type annotation for integrated and disintegrated single-cell transcriptomic atlases of *D. melanogaster*, *D. simulans*, and *D. sechellia*. Right, expression levels of marker genes distinguishing the GABAergic cell types. The data underlying this figure can be found in <https://zenodo.org/records/15016454>. (PDF)

S5 Fig. Genes with conserved expression patterns across the central brains of *D. melanogaster*, *D. simulans*, and *D. sechellia*. (A) Scatter plots illustrating the expression pattern comparisons of two genes, *bru3* (top) and *gw* (bottom), between *D. melanogaster* and *D. sechellia*. Each point corresponds to one of the 107 cell types. Average expression levels of *D. melanogaster* are shown on the x-axis, and *D. sechellia* expression levels are shown on the y-axis. Smoothed lines depict generalized linear model fits. (B) Top: two-dimensional density plots of gene expression pattern similarity for three pairwise comparisons. Each point in the plot corresponds to one of 2,405 analyzed genes. The y-axis shows the expression pattern similarity, which is estimated by the correlation value (Spearman's ρ). Bottom: histograms of gene count with highly conserved expression patterns ($\rho > 0.7$) are shown. The x-axis shows the percentage of central brain cells expressing each gene. (C) GO enrichment analysis for 413 genes with highly conserved expression patterns ($\rho > 0.7$) across all three pairwise comparisons. The data underlying this figure can be found in <https://zenodo.org/records/15016454>. (EPS)

S6 Fig. Transcription factor fingerprints for central brain cell types. A heatmap illustrating the percentage of cells expressing 104 transcription factor genes across 107 cell types, derived from the *D. melanogaster* dataset. Both genes and cell types are clustered through hierarchical clustering. Labels for glial cell types and Kenyon cell types are colored blue and red, respectively. The data underlying this figure can be found in <https://zenodo.org/records/15016454>. (EPS)

S7 Fig. Frequency comparisons across homologous cell types. Bar plots showing comparisons across the drosophilid species of the frequencies of 107 central brain cell types. Each point corresponds to one of six replicate snRNA-seq datasets. The data underlying this figure can be found in <https://zenodo.org/records/15016454>. (EPS)

S8 Fig. *Tret1-1* expression levels in *Drosophila* species. Dot plots summarizing the expression levels of *Tret1-1* across 107 cell types of central brains of *D. melanogaster*, *D. simulans*, and *D. sechellia*. The data underlying this figure can be found in <https://zenodo.org/records/15016454>. (EPS)

S9 Fig. Differences in glial cell types among three *Drosophila* species. (A) Left, representative images of *Tret1-1* HCR FISH on *DmelCS*, *Dsim04*, and *Dsec07*. Right, quantifications of *Tret1-1* signals for each species. Images and data were obtained from independent experiments from those shown in Fig 2D. (B) Left, representative images of Repo immunofluorescence on *DmelCS*, *Dsim04*, and *Dsec07*. Right, quantifications of Repo+ cell counts for each species. Images and data were obtained from independent experiments from those shown in Fig 2F. The data underlying this figure can be found in <https://zenodo.org/records/15016454>. (EPS)

S10 Fig. Correlation between cell type frequencies and DEG counts. Scatter plots illustrating percent of cell type frequencies (x-axis) and mean DEG counts from the three comparisons (*Dmel-Dsec*, *Dmel-Dsim*, and *Dsim-Dsec*) (y-axis). The data underlying this figure can be found in <https://zenodo.org/records/15016454>. (EPS)

S11 Fig. Predicted divergent gene expression in *D. sechellia*. Dot plots illustrating expression levels and frequencies of 10 *D. sechellia*-specific DEGs in *D. melanogaster*, *D. simulans*, and *D. sechellia* across 107 cell types. The data underlying this figure can be found in <https://zenodo.org/records/15016454>. (EPS)

S12 Fig. In vivo validation of divergent gene expression in *D. sechellia*. (A) Left, representative images of *CAH2* HCR FISH on *DmelCS*, *Dsim04*, and *Dsec07*. Right, quantifications of *CAH2* signals for each species. Images and data were obtained from independent experiments distinct from those shown in Fig 5C. (B) Left, representative images of *ImpE1* HCR FISH on *DmelCS*, *Dsim04*, and *Dsec07*. Right, quantifications of *ImpE1* signals for each species. Images and data were obtained from independent experiments distinct from those shown in Fig 5E. The data underlying this figure can be found in <https://zenodo.org/records/15016454>. (EPS)

S13 Fig. Impact of noni paste supplement on cell type frequencies in the *D. sechellia* central brain. A dot plot showing comparisons of brain cell type frequencies between *D. sechellia* grown on standard medium (blue) and noni paste-supplemented medium (purple). Each point corresponds to one of six replicates. The error bars represent the standard error of the mean. A paired *t* test was conducted to assess statistical significance, with *P*-values adjusted using false discovery rate. No significant differences in cell type frequencies were observed between conditions. The data underlying this figure can be found in <https://zenodo.org/records/15016454>. (EPS)

S14 Fig. Expression changes of CG5151 in *D. sechellia* upon noni-paste supplement. Dot plots summarizing the predicted expression changes of CG5151 across 107 cell types of brains of *D. sechellia* grown on standard food or with noni paste supplement. In the latter condition, CG5151 is upregulated in PRN and SUB, but this

increase is statistically significant only in PRN. The data underlying this figure can be found in <https://zenodo.org/records/15016454>.

(EPS)

S15 Fig. In vivo validation of upregulation of CG5151 in *D. sechellia* upon noni-paste supplement Left, representative images of CG5151 HCR FISH on *Dsec07* grown under the indicated conditions. Right, quantifications of CG5151 signals. Images and data were obtained from independent experiments from those shown in Fig 6D. The data underlying this figure can be found in <https://zenodo.org/records/15016454>.

(EPS)

S16 Fig. Changes in *D. sechellia* brain gene expression upon noni supplement. Dot plots illustrating expression levels of four DEGs in the indicated cell types in response to noni treatment. The data underlying this figure can be found in <https://zenodo.org/records/15016454>.

(EPS)

Code availability

All code for generating the figures and tables are available from Zenodo (<https://zenodo.org/records/15016454>).

Acknowledgments

We thank the Flow Cytometry Facility (D. Labes) and Lausanne Genomic Technologies Facility (J. Marquis, C. Peter, R. Sermier, and K. Bojkowska) of the University of Lausanne for assistance with cell preparation and sequencing. We thank for B.-Y. Lee for bioinformatic support on single-cell analysis. We are grateful to members of the Benton laboratory for discussions and comments on the manuscript.

Author contributions

Conceptualization: Daehan Lee, Richard Benton.

Data curation: Daehan Lee, Michael P. Shahandeh.

Formal analysis: Daehan Lee, Michael P. Shahandeh.

Funding acquisition: Daehan Lee, Richard Benton.

Investigation: Daehan Lee, Michael P. Shahandeh, Liliane Abuin.

Methodology: Daehan Lee, Michael P. Shahandeh.

Project administration: Daehan Lee, Richard Benton.

Resources: Daehan Lee, Richard Benton.

Software: Daehan Lee.

Supervision: Daehan Lee, Michael P. Shahandeh, Richard Benton.

Validation: Daehan Lee.

Visualization: Daehan Lee, Michael P. Shahandeh.

Writing – original draft: Daehan Lee, Richard Benton.

Writing – review & editing: Daehan Lee, Michael P. Shahandeh, Liliane Abuin, Richard Benton.

References

1. Arendt D, Musser JM, Baker CVH, Bergman A, Cepko C, Erwin DH, et al. The origin and evolution of cell types. *Nat Rev Genet.* 2016;17(12):744–57. <https://doi.org/10.1038/nrg.2016.127> PMID: [27818507](https://pubmed.ncbi.nlm.nih.gov/27818507/)

2. Zeng H. What is a cell type and how to define it? *Cell*. 2022;185(15):2739–55. <https://doi.org/10.1016/j.cell.2022.06.031> PMID: 35868277
3. Wilton DK, Dissing-Olesen L, Stevens B. Neuron-glia signaling in synapse elimination. *Annu Rev Neurosci*. 2019;42:107–27. <https://doi.org/10.1146/annurev-neuro-070918-050306> PMID: 31283900
4. Kofuji P, Araque A. Astrocytes and behavior. *Annu Rev Neurosci*. 2021;44:49–67. <https://doi.org/10.1146/annurev-neuro-101920-112225> PMID: 33406370
5. Roberts RJV, Pop S, Prieto-Godino LL. Evolution of central neural circuits: state of the art and perspectives. *Nat Rev Neurosci*. 2022;23(12):725–43. <https://doi.org/10.1038/s41583-022-00644-y> PMID: 36289403
6. Tosches MA. Developmental and genetic mechanisms of neural circuit evolution. *Dev Biol*. 2017;431:16–25.
7. Barsotti E, Correia A, Cardona A. Neural architectures in the light of comparative connectomics. *Curr Opin Neurobiol*. 2021;71:139–49. <https://doi.org/10.1016/j.conb.2021.10.006> PMID: 34837731
8. Herculano-Houzel S. Life history changes accompany increased numbers of cortical neurons: a new framework for understanding human brain evolution. *Prog Brain Res*. 2019;250:179–216. <https://doi.org/10.1016/bs.pbr.2019.06.001> PMID: 31703901
9. Kobschull JM, Richman EB, Ringach N, Friedmann D, Albarran E, Kolluru SS, et al. Cerebellar nuclei evolved by repeatedly duplicating a conserved cell-type set. *Science*. 2020;370(6523):eabd5059. <https://doi.org/10.1126/science.abd5059> PMID: 33335034
10. Hain D, Gallego-Flores T, Klinkmann M, Macias A, Ciirdeaeva E, Arends A, et al. Molecular diversity and evolution of neuron types in the amniote brain. *Science*. 2022;377(6610):eabp8202. <https://doi.org/10.1126/science.abp8202> PMID: 36048944
11. Tosches MA, Yamawaki TM, Naumann RK, Jacobi AA, Tushev G, Laurent G. Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. *Science*. 2018;360(6391):881–8. <https://doi.org/10.1126/science.aar4237> PMID: 29724907
12. Woych J, Ortega Gurrola A, Deryckere A, Jaeger ECB, Gummit E, Merello G, et al. Cell-type profiling in salamanders identifies innovations in vertebrate forebrain evolution. *Science*. 2022;377(6610):eabp9186. <https://doi.org/10.1126/science.abp9186> PMID: 36048957
13. Shafer MER, Sawh AN, Schier AF. Gene family evolution underlies cell-type diversification in the hypothalamus of teleosts. *Nat Ecol Evol*. 2022;6(1):63–76. <https://doi.org/10.1038/s41559-021-01580-3> PMID: 34824389
14. Siletti K, Hodge R, Mossi Albiach A, Lee KW, Ding S-L, Hu L, et al. Transcriptomic diversity of cell types across the adult human brain. *Science*. 2023;382(6667):eadd7046. <https://doi.org/10.1126/science.add7046> PMID: 37824663
15. Yao Z, van Velthoven C, Kunst M, Zhang M, McMillen D, Lee C. A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature*. 2023;624:317–32.
16. Herculano-Houzel S, Mota B, Lent R. Cellular scaling rules for rodent brains. *Proc Natl Acad Sci USA*. 2006;103:12138–43.
17. Mu S, Yu S, Turner NL, McKellar CE, Dorkenwald S, Collman F, et al. 3D reconstruction of cell nuclei in a full *Drosophila* brain. 2021. <https://doi.org/10.1101/2021.11.04.467197>
18. Vosshall LB. Into the mind of a fly. *Nature*. 2007;450(7167):193–7. <https://doi.org/10.1038/nature06335> PMID: 17994085
19. Venken KJT, Simpson JH, Bellen HJ. Genetic manipulation of genes and cells in the nervous system of the fruit fly. *Neuron*. 2011;72(2):202–30. <https://doi.org/10.1016/j.neuron.2011.09.021> PMID: 22017985
20. Scheffer LK, Meinertzhagen IA. The fly brain atlas. *Annu Rev Cell Dev Biol*. 2019;35:637–53. <https://doi.org/10.1146/annurev-cell-bio-100818-125444> PMID: 31283380
21. Davie K, Janssens J, Koldere D, De Waegeneer M, Pech U, Kreft L, et al. A single-cell transcriptome atlas of the aging *Drosophila* brain. *Cell*. 2018;174(4):982–998.e20. <https://doi.org/10.1016/j.cell.2018.05.057> PMID: 29909982
22. Croset V, Treiber CD, Waddell S. Cellular diversity in the *Drosophila* midbrain revealed by single-cell transcriptomics. *Elife*. 2018;7:e34550. <https://doi.org/10.7554/eLife.34550> PMID: 29671739
23. Li H, Janssens J, De Waegeneer M, Kolluru SS, Davie K, Gardeux V, et al. Fly Cell Atlas: a single-nucleus transcriptomic atlas of the adult fruit fly. *Science*. 2022;375(6584):eabk2432. <https://doi.org/10.1126/science.abk2432> PMID: 35239393
24. Markow TA. Host use and host shifts in *Drosophila*. *Curr Opin Insect Sci*. 2019;31:139–45. <https://doi.org/10.1016/j.cois.2019.01.006> PMID: 31109667
25. Auer T, Benton R. Sexual circuitry in *Drosophila*. *Current Opinion in Neurobiology*. 2016;38:18–26.
26. Garrigan D, Kingan S, Geneva A, Andolfatto P, Clark A, Thornton K, et al. Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. *Genome Research*. 2012;22:1499–511.
27. Schrider DR, Ayroles J, Matute DR, Kern AD. Supervised machine learning reveals introgressed loci in the genomes of *Drosophila simulans* and *D. sechellia*. *PLoS Genet*. 2018;14(4):e1007341. <https://doi.org/10.1371/journal.pgen.1007341> PMID: 29684059
28. Auer TO, Shahandeh MP, Benton R. *Drosophila sechellia*: a genetic model for behavioral evolution and neuroecology. *Annu Rev Genet*. 2021;55:527–54. <https://doi.org/10.1146/annurev-genet-071719-020719> PMID: 34530638
29. Farine J-P, Legal L, Moreteau B, Le Quere J-L. Volatile components of ripe fruits of *Morinda citrifolia* and their effects on *Drosophila*. *Phytochemistry*. 1996;41:433–8.
30. Matute DR, Ayroles JF. Hybridization occurs between *Drosophila simulans* and *D. sechellia* in the Seychelles archipelago. *J Evol Biol*. 2014;27:1057–1068.

31. Lavista-Llanos S, Svatoš A, Kai M, Riemensperger T, Birman S, Stensmyr MC, et al. Dopamine drives *Drosophila sechellia* adaptation to its toxic host. *Elife*. 2014;3:e03785. <https://doi.org/10.7554/eLife.03785> PMID: 25487989
32. Prieto-Godino LL, Rytz R, Cruchet S, Bargeton B, Abuin L, Silbering AF, et al. Evolution of acid-sensing olfactory circuits in Drosophilids. *Neuron*. 2017;93(3):661–676.e6. <https://doi.org/10.1016/j.neuron.2016.12.024> PMID: 28111079
33. Auer TO, Khallaf MA, Silbering AF, Zappia G, Ellis K, Álvarez-Ocaña R, et al. Olfactory receptor and circuit evolution promote host specialization. *Nature*. 2020;579(7799):402–8. <https://doi.org/10.1038/s41586-020-2073-7> PMID: 32132713
34. Álvarez-Ocaña R, Shahandeh MP, Ray V, Auer TO, Gompel N, Benton R. Odor-regulated oviposition behavior in an ecological specialist. *Nat Commun*. 2023;14(1):3041. <https://doi.org/10.1038/s41467-023-38722-z> PMID: 37236992
35. Reisenman CE, Wong J, Vedagarbha N, Lively C, Scott K. Taste adaptations associated with host specialization in the specialist *Drosophila sechellia*. *J Exp Biol*. 2023;226(3):jeb244641. <https://doi.org/10.1242/jeb.244641> PMID: 36637369
36. Dey M, Brown E, Charlu S, Keene A, Dahanukar A. Evolution of fatty acid taste in drosophilids. *Cell Rep*. 2023;42(10):113297. <https://doi.org/10.1016/j.celrep.2023.113297> PMID: 37864792
37. Shahandeh MP, Abuin L, Lescuyer De Decker L, Cergneux J, Koch R, Nagoshi E, et al. Circadian plasticity evolves through regulatory changes in a neuropeptide gene. *Nature*. 2024;635(8040):951–9. <https://doi.org/10.1038/s41586-024-08056-x> PMID: 39415010
38. Prieto-Godino LL, Rytz R, Bargeton B, Abuin L, Arguello JR, Peraro MD, et al. Olfactory receptor pseudo-pseudogenes. *Nature*. 2016;539(7627):93–7. <https://doi.org/10.1038/nature19824> PMID: 27776356
39. Prieto-Godino LL, Schmidt HR, Benton R. Molecular reconstruction of recurrent evolutionary switching in olfactory receptor specificity. *Elife*. 2021;10:e69732. <https://doi.org/10.7554/eLife.69732> PMID: 34677122
40. Dekker T, Ibba I, Siju KP, Stensmyr MC, Hansson BS. Olfactory shifts parallel superspecialism for toxic fruit in *Drosophila melanogaster* sibling, *D. sechellia*. *Curr Biol*. 2006;16(1):101–9. <https://doi.org/10.1016/j.cub.2005.11.075> PMID: 16401429
41. Ibba I, Angioy AM, Hansson BS, Dekker T. Macroglomeruli for fruit odors change blend preference in *Drosophila*. *Naturwissenschaften*. 2010;97(12):1059–66. <https://doi.org/10.1007/s00114-010-0727-2> PMID: 20972770
42. Takagi S, Sancer G, Abuin L, Stupski SD, Roman Arguello J, Prieto-Godino LL, et al. Olfactory sensory neuron population expansions influence projection neuron adaptation and enhance odour tracking. *Nat Commun*. 2024;15(1):7041. <https://doi.org/10.1038/s41467-024-50808-w> PMID: 39147786
43. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun*. 2017;8:14049.
44. Hobert O. A map of terminal regulators of neuronal identity in *Caenorhabditis elegans*. *Wiley Interdiscip Rev Dev Biol*. 2016;5:474–98.
45. Özel MN, Gibbs CS, Holguera I, Soliman M, Bonneau R, Desplan C. Coordinated control of neuronal differentiation and wiring by sustained transcription factors. *Science*. 2022;378(6626):eadd1884. <https://doi.org/10.1126/science.add1884> PMID: 36480601
46. Bai L, Goldman AL, Carlson JR. Positive and negative regulation of odor receptor gene choice in *Drosophila* by acj6. *J Neurosci*. 2009;29(41):12940–7. <https://doi.org/10.1523/JNEUROSCI.3525-09.2009> PMID: 19828808
47. Jafari S, Alkhori L, Schleiffer A, Brochtrup A, Hummel T, Alenius M. Combinatorial activation and repression by seven transcription factors specify *Drosophila* odorant receptor expression. *PLoS Biol*. 2012;10(3):e1001280. <https://doi.org/10.1371/journal.pbio.1001280> PMID: 22427741
48. Kurusu M, Nagao T, Walldorf U, Flister S, Gehring WJ, Furukubo-Tokunaga K. Genetic control of development of the mushroom bodies, the associative learning centers in the *Drosophila* brain, by the *eyeless*, *twins of eyeless*, and *Dachshund* genes. *Proc Natl Acad Sci USA*. 2000;97(5):2140–4. <https://doi.org/10.1073/pnas.040564497> PMID: 10681433
49. Guo X, Zhang Y, Huang H, Xi R. A hierarchical transcription factor cascade regulates enteroendocrine cell diversity and plasticity in *Drosophila*. *Nat Commun*. 2022;13(1):6525. <https://doi.org/10.1038/s41467-022-34270-0> PMID: 36316343
50. Volkenhoff A, Weiler A, Letzel M, Stehling M, Klämbt C, Schirmeier S. Glial glycolysis is essential for neuronal survival in *Drosophila*. *Cell Metab*. 2015;22(3):437–47. <https://doi.org/10.1016/j.cmet.2015.07.006> PMID: 26235423
51. Mermet J, Cruchet S, Borbora A, Lee D, Chai P, Jang A, et al. Multilayer regulation underlies the functional precision and evolvability of the olfactory system. *bioRxiv*. 2025. <https://doi.org/10.1101/2025.01.16.632932>
52. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44(W1):W90–7. <https://doi.org/10.1093/nar/gkw377> PMID: 27141961
53. Schulz D, Goillard J-M, Marder E. Quantitative expression profiling of identified neurons reveals cell-specific constraints on highly variable levels of gene expression. *Proc Natl Acad Sci USA*. 2007;104:13187–91.
54. Watanabe K, Kanaoka Y, Mizutani S, Uchiyama H, Yajima S, Watada M. Interspecies comparative analyses reveal distinct carbohydrate-responsive systems among *Drosophila* species. *Cell Rep*. 2019;28(7):2594–2607.e7. <https://doi.org/10.1016/j.celrep.2019.02.020>
55. Day EH, Hua X, Bromham L. Is specialization an evolutionary dead end? Testing for differences in speciation, extinction and trait transition rates across diverse phylogenies of specialists and generalists. *J Evol Biol*. 2016;29(6):1257–67. <https://doi.org/10.1111/jeb.12867> PMID: 26998589
56. Bittern J, Pogodalla N, Ohm H, Brüser L, Kottmeier R, Schirmeier S, et al. Neuron-glia interaction in the *Drosophila* nervous system. *Dev Neurobiol*. 2021;81(5):438–52. <https://doi.org/10.1002/dneu.22737> PMID: 32096904

57. Rifkin SA, Kim J, White KP. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat Genet.* 2003;33(2):138–44. <https://doi.org/10.1038/ng1086> PMID: [12548287](#)
58. Khaitovich P, Weiss G, Lachmann M, Hellmann I, Enard W, Muetzel B, et al. A neutral model of transcriptome evolution. *PLoS Biol.* 2004;2(5):E132. <https://doi.org/10.1371/journal.pbio.0020132> PMID: [15138501](#)
59. Legrand D, Tenaillon MI, Matyot P, Gerlach J, Lachaise D, Cariou M-L. Species-wide genetic variation and demographic history of *Drosophila sechellia*, a species lacking population structure. *Genetics.* 2009;182(4):1197–206. <https://doi.org/10.1534/genetics.108.092080> PMID: [19506309](#)
60. Drum Z, Lanno S, Gregory SM, Shimshak S, Barr W, Gatesman A, et al. Genomics analysis of *Drosophila sechellia* response to *Morinda citrifolia* fruit diet. *G3 (Bethesda).* 2022;12(10):jkac153. <https://doi.org/10.1093/g3journal/jkac153> PMID: [35736356](#)
61. Lanet E, Maurange C. Building a brain under nutritional restriction: insights on sparing and plasticity from *Drosophila* studies. *Front Physiol.* 2014;5:117. <https://doi.org/10.3389/fphys.2014.00117> PMID: [24723892](#)
62. Turner B. Epigenetic responses to environmental change and their evolutionary implications. *Philos Trans R Soc Lond B Biol Sci.* 2009;364:3403–18.
63. Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience.* 2020;9(12):giaa151. <https://doi.org/10.1093/gigascience/giaa151> PMID: [33367645](#)
64. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* 2019;20(1):296. <https://doi.org/10.1186/s13059-019-1874-1> PMID: [31870423](#)
65. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* 2019;8(4):329–337.e4. <https://doi.org/10.1016/j.cels.2019.03.003> PMID: [30954475](#)
66. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20(1):238. <https://doi.org/10.1186/s13059-019-1832-y> PMID: [31727128](#)
67. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell.* 2021;184(13):3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048> PMID: [34062119](#)
68. Deng B, Li Q, Liu X, Cao Y, Li B, Qian Y. Chemoconnectomics: mapping chemical transmission in *Drosophila*. *Neuron.* n.d.;101:876–93.
69. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics.* 2013;14:128. <https://doi.org/10.1186/1471-2105-14-128> PMID: [23586463](#)
70. Ferreira AAG, Sieriebriennikov B, Whitbeck H. HCR RNA-FISH protocol for the whole-mount brains of *Drosophila* and other insects. *protocols.io.* 2021 [cited 23 Jan 2025]. Available from: <https://www.protocols.io/view/hcr-rna-fish-protocol-for-the-whole-mount-brains-o-bzh5p386.pdf>
71. Yuasa Y, Okabe M, Yoshikawa S, Tabuchi K, Xiong W-C, Hiromi Y, et al. *Drosophila* homeodomain protein REPO controls glial differentiation by cooperating with ETS and BTB transcription factors. *Development.* 2003;130(11):2419–28. <https://doi.org/10.1242/dev.00468> PMID: [12702656](#)
72. Ostrovsky A, Cachero S, Jefferis G. Clonal analysis of olfaction in *Drosophila*: immunochemistry and imaging of fly brains. *Cold Spring Harb Protoc.* 2013;2013(4):342–6. <https://doi.org/10.1101/pdb.prot071720> PMID: [23547149](#)
73. Bahry E, Breimann L, Zouinkhi M, Epstein L, Kolyvanov K, Mamrak N, et al. RS-FISH: precise, interactive, fast, and scalable FISH spot detection. *Nat Methods.* 2022;19(12):1563–7. <https://doi.org/10.1038/s41592-022-01669-y> PMID: [36396787](#)
74. Ostrovsky AD, Goetz L, Jefferis GSXE. *Drosophila melanogaster* template brains. 2014. <https://doi.org/10.5281/zenodo.10591>
75. Ostrovsky A, Goetz L, Jefferis G. *Drosophila simulans* template brains. 2014. <https://doi.org/10.5281/zenodo.10594>
76. Jefferis GSXE, Benton R. DsecF *Drosophila sechellia* female template brain. 2019. <https://doi.org/10.5281/zenodo.2562141>