Article

Current Biology

# The evolutionary trajectory of drosophilid walking

## Highlights

- Unconstrained locomotor behavior measured in more than one thousand individual flies
- Drosophilid walking can be captured by a universal low-dimensional behavior space
- Hierarchical evolution and varying dynamics among behavioral components
- Behaviors evolve more rapidly than other traits within *Drosophila*

## Authors

Ryan A. York, Luke E. Brezovec, Jenn Coughlan, ..., Daniel R. Matute, John C. Tuthill, Thomas R. Clandinin

## Correspondence

ryanayork@gmail.com (R.A.Y.), trc@stanford.edu (T.R.C.)

## In brief

York et al. study the evolutionary dynamics of the *Drosophila* walking repertoire over 40 million years. They find that the kinematic structure of walking is deeply conserved, although the occurrence and sequencing of specific movement patterns have evolved rapidly. This study opens new pathways for the macroevolutionary study of behavioral traits.

York et al., 2022, Current Biology *32*, 3005–3015
July 25, 2022 
https://doi.org/10.1016/j.cub.2022.05.039

CellPress

Current Biology

CellPress

Article

# The evolutionary trajectory of drosophilid walking

Ryan A. York,[1,*] Luke E. Brezovec,[1] Jenn Coughlan,[2] Steven Herbst,[1] Avery Krieger,[1] Su-Yee Lee,[3] Brandon Pratt,[3] Ashley D. Smart,[1] Eugene Song,[1] Anton Suvorov,[2] Daniel R. Matute,[2] John C. Tuthill,[3] and Thomas R. Clandinin[1,4,*]

[1]Department of Neurobiology, Stanford University, Stanford, CA 94305, USA

[2]Biology Department, University of North Carolina, Chapel Hill, NC 27599, USA

[3]Department of Physiology and Biophysics, University of Washington, Seattle, WA 98195, USA

[4]Lead contact

*Correspondence: ryanayork@gmail.com (R.A.Y.), trc@stanford.edu (T.R.C.)

https://doi.org/10.1016/j.cub.2022.05.039

## SUMMARY

Neural circuits must both execute the behavioral repertoire of individuals and account for behavioral variation across species. Understanding how this variation emerges over evolutionary time requires large-scale phylogenetic comparisons of behavioral repertoires. Here, we describe the evolution of walking in fruit flies by capturing high-resolution, unconstrained movement from 13 species and 15 strains of drosophilids. We find that walking can be captured in a universal behavior space, the structure of which is evolutionarily conserved. However, the occurrence of and transitions between specific movements have evolved rapidly, resulting in repeated convergent evolution in the temporal structure of locomotion. Moreover, a meta-analysis demonstrates that many behaviors evolve more rapidly than other traits. Thus, the architecture and physiology of locomotor circuits can execute precise individual movements in one species and simultaneously support rapid evolutionary changes in the temporal ordering of these modular elements across clades.

## INTRODUCTION

A central goal of neuroscience is to understand how circuits shape behavior. Using a relatively small number of model systems, diverse efforts in this field have greatly advanced our understanding of neural mechanisms . However, how the structure and function of neural circuits can support the diversification behavioral repertoires over evolutionary time is incompletely understood. To understand how circuits can evolve requires establishing model clades and groups of related species displaying both behavioral diversity and having the requisite tools to dissect circuit and computational mechanisms.[1–4] New methods in behavioral measurement[5,6] and statistical analysis[7,8] have made it possible to collect and study many aspects of behavior at a large scale in many clades. Similarly, broadly sampled and time-resolved molecular phylogenies have become increasingly available for many groups of animals.[9] Here, we demonstrate that combining high-resolution, quantitative behavioral analyses with robust phylogenetics can reveal rich patterns of behavioral evolution comparable with those seen for other traits.

We focus on the evolution of walking, a critical element of many behaviors, in fruit flies of the genus *Drosophila*. Fruit flies are nearly unique in being both a genetic and an evolutionary model system, representing a broad range of life histories and ecological contexts that provide strong bases for inter- and intra-specific comparisons.[10–13] Moreover, the structure of fruit fly walking has been well resolved in the model species *D. melanogaster*,[14–16] and recent work has characterized the functions of a number of critical neuron types that determine the initiation of walking, turning, and stopping.[17–21] Here, we develop an approach to quantitatively compare the structure of walking across fly species and strains. We find that this well-constrained example of motor control can evolve surprisingly rapidly, with closely related strains diverging and distantly related species converging on similar temporal patterns of locomotor movements. These results demonstrate that behavioral variation can emerge from changing the temporal sequence of individual, modular movements, thereby identifying potential neural mechanisms of locomotor evolution.

## RESULTS

### High-throughput measurement of drosophilid walking

To analyze the structure of naturalistic walking in fruit flies, we developed the Coliseum, a novel apparatus for behavioral experiments in *Drosophila* (Figure 1A), and flyvr software (STAR Methods) for the acquisition and analysis of data from the Coliseum (Figure 1B). The Coliseum uses a fly-initiated dispenser to introduce flies onto a darkened ~1 m × 1 m platform under infrared illumination (Figure 1A; STAR Methods). As flies freely explore the apparatus, a camera attached to a stepper-motor system tracks and updates the camera position with animal movement (100 Hz), yielding high-quality video and positional data in real time (Figure 1B). We used the Coliseum to measure the trajectories of 1,030 individuals from 13 globally distributed *Drosophila* species with diverse distributions ranging from cosmopolitan (e.g., *D. melanogaster*, *D. simulans*), temperate (e.g., *D. virilis*, *D. persimilis*), tropical (e.g., *D. yakuba*, *D. mauritiana*), and desert (e.g., *D. arizonae*) habitats[13]; in addition, to detect traits that can undergo very rapid evolutionary changes, we analyzed 15 wild-derived strains of *D. melanogaster* from its ancestral distribution in Africa (Figure 1C).

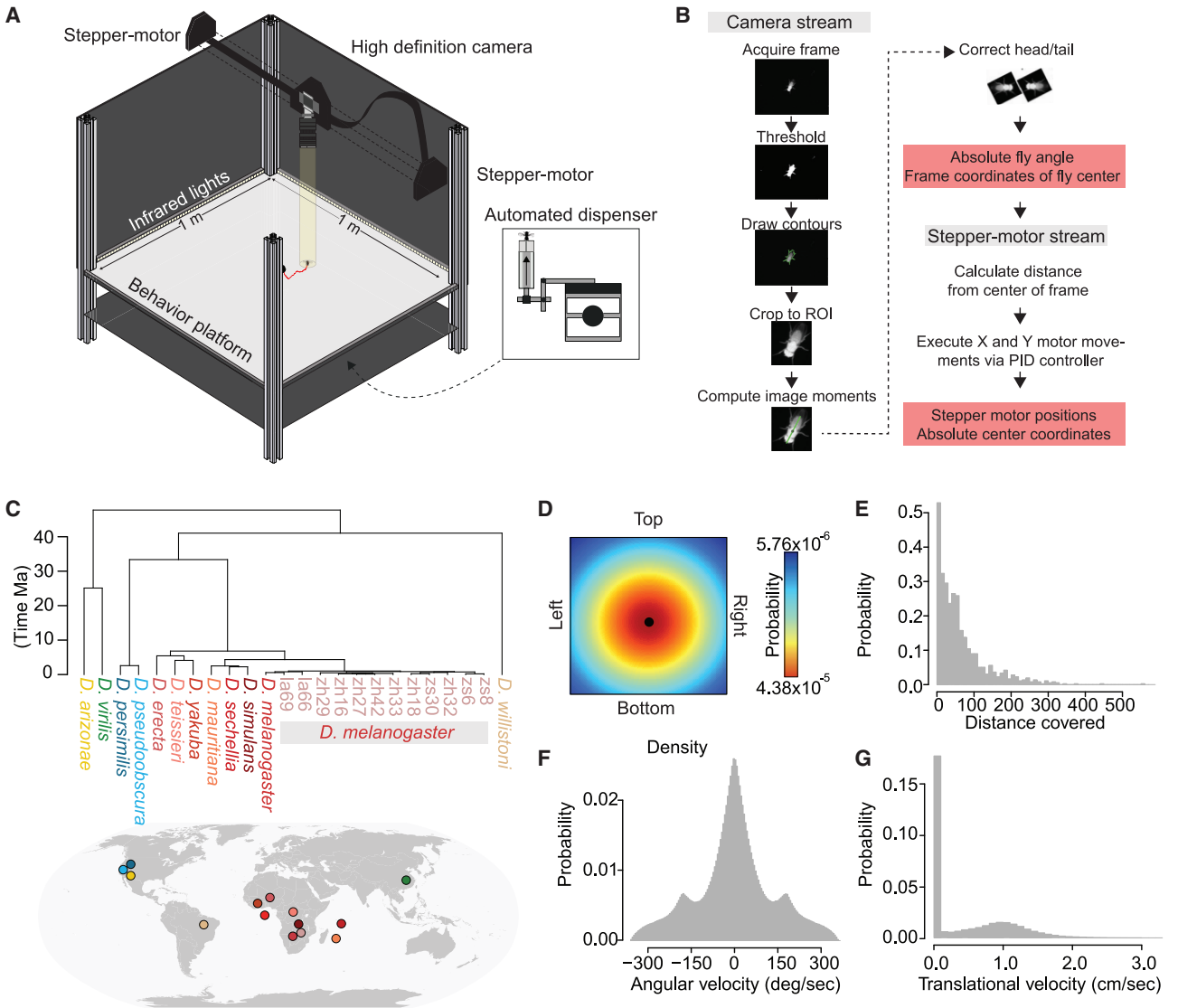


**Figure 1. High-throughput measurement of locomotion across *Drosophila***

(A) Schematic of the Coliseum. Flies are introduced onto an IR illuminated behavior platform by an automated dispenser. A high-definition camera is suspended above on a stepper-motor system that updates the camera's position as the fly moves. The visual field of the camera is indicated by the yellow cylinder tracking the fly's trajectory (marked by the red path).

(B) Workflow for acquiring fruit fly position in the Coliseum, extracting orientation, and updating camera position via stepper motors.

(C) (Above) Time-calibrated phylogeny of the species analyzed. (Below) The approximate global locations of the ancestral populations for each species or strain represented in the phylogeny (coded by color).

(D) Probability density function of fly position in the Coliseum, computed using the positions of all trials in the data set. The dispenser hole location is indicated by the black dot.

(E–G) The distributions of distance covered (cm, E), angular velocity (deg/s, F), and translational velocity (cm/s, G) encompassed by the full locomotor data set. See also Figure S1 and Data S2.

Comparing their relationships using a fossil-calibrated phylogeny established that these species and strains represent ~40 million years of evolutionary history (Figure 1C). A variety of species and clade-level relationships were examined, including recent evolutionary diversifications in the *D. melanogaster* clade (crown age = 41 mya; minimum branch length = 0.12 mya; Figure 1C), facilitating a phylogenetically thorough exploration of the evolution of walking. Overall, flies uniformly explored the Coliseum (Figure 1D) and displayed a variety of trajectory lengths and durations (Figure 1E) within the typical kinematic range of adult *Drosophila* (Figures 1F and 1G).[15,16,22] These measurements varied as a function of species and strain (Figures S1A–S1C), highlighting diversity in these first order statistics across genotypes. Furthermore, the species and strains varied broadly in the average time spent moving per trial (Figure S1D), suggesting that the structure and sequencing of behavior may vary as a function of genotype. Moreover, we found little of evidence of thigmotaxis (boundary preference) in these data—a common

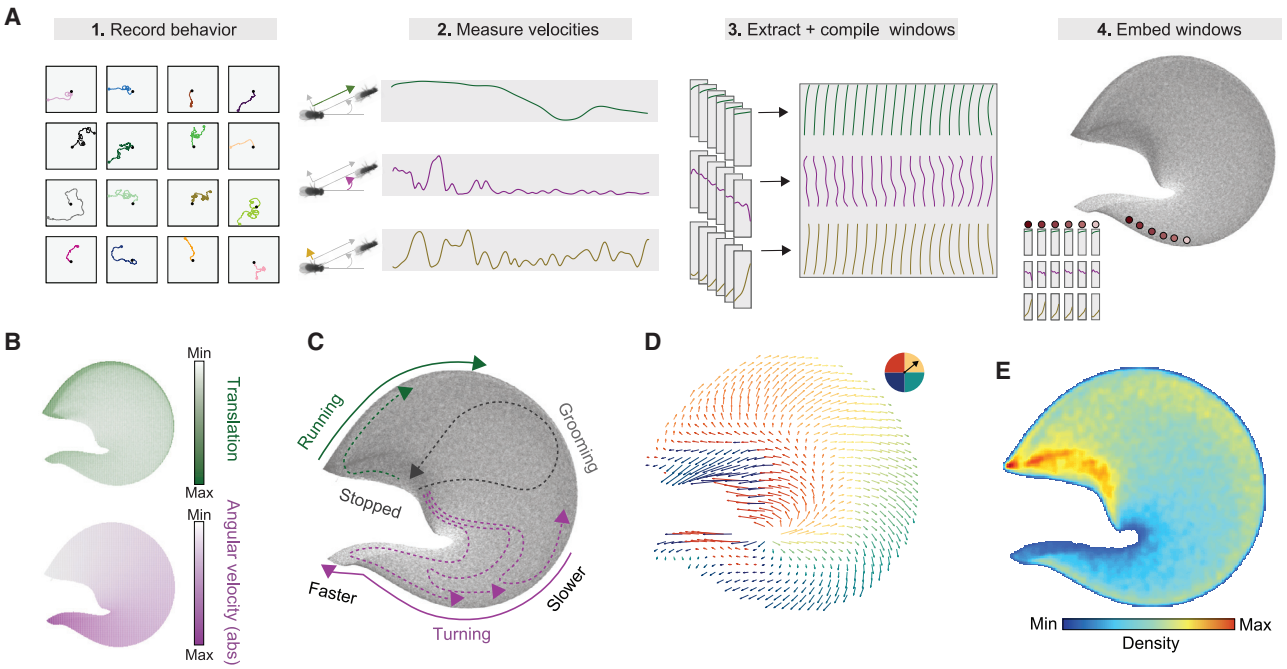


**Figure 2. Defining a universal walking behavior space**

(A) The workflow for the TREBLE framework.

(B) The distributions of translational (top) and angular (bottom) velocities as a function of position in behavior space.

(C) The structure of behavior space annotated with behavioral states and pathways in between.

(D) The mean vector field of behavior space. Arrow angle and length indicate the direction and magnitude, respectively, of probabilistic movement between points. The angle degree is also denoted by color, corresponding to the circle plotted above.

(E) The 2D probability density function representing the frequency of occurrence across behavior space for the full walking dataset.

See also Figure S2.

pattern seen in open field tests of animal behavior—reflecting the fact that animals rarely reach the edge of the Coliseum. The Coliseum is therefore a powerful tool for measuring naturalistic walking and can be used to obtain high-resolution data from diverse strains and species.

## A common space captures drosophilid walking

The first order statistics of walking display substantial temporal structure on the timescale of hundreds of milliseconds.[14–16,22] To capture this structure, we employed the TREBLE framework (Figure 2A)[8] to create a common behavior space encompassing all individual flies from all strains and species (Figure S1J). The representation of behavior provided by TREBLE enables a number of powerful analyses. Intuitively, an animal's behavioral state is represented by a coordinate in behavior space. There coordinates are defined by unique combinations of behavioral parameters, here measured as a set of velocities over a defined time interval (discussed below). As these velocities change, the animal moves along a temporally continuous pathway in this space that allows for interrogations of the sequencing of movement over a variety of timescales. As a result, statistical models of the frequency of, and transition probabilities between, specific movements can be decoded. Furthermore, aspects of the overall structure of behavior can be assayed by measuring the frequency and timescale over which an animal revisits specific regions of behavior space. These facets together enable one to produce a representation of behavior that is temporally resolved, captures variation across individuals and groups, and can interrogate both local and global behavioral structure across timescales.

To construct such a representation here, kinematic parameters (Figure 2Aii) were measured from the raw fly trajectories (Figure 2Aiv), binned into windows that reflected the temporal structure of walking (see STAR Methods; Figures 2Aiii and S1E–S1I), and embedded into a 2D behavior space via a nonlinear dimensionality reduction (Figure 2Aiv).[23] The resulting space separated kinematic parameters along identifiable axes (Figure 2B) corresponding to recognizable behaviors (Figures 2C and S1L–S1N). To identify stereotyped movement through the space, we calculated the mean path of all flies through every point in behavior space, producing a vector map in which large vectors correspond to similar movements performed by many animals. This analysis revealed common pathways linking specific behaviors that varied in their frequency of occurrence (Figures 2D and 2E). TREBLE therefore produced a stereotyped, continuous behavior space encompassing the breadth of drosophilid walking kinematics.

Distinct gait patterns are associated with variation in kinematics over time in *D. melanogaster*.[22,24–28] To explore whether other species displayed similar patterns, we identified the positions of all six limbs for each species and strain and extracted the phase and swing-stance state at each frame (Figure 3A; STAR Methods). As expected from *D. melanogaster*,[22,24–28] we found that the number of legs in stance was speed dependent

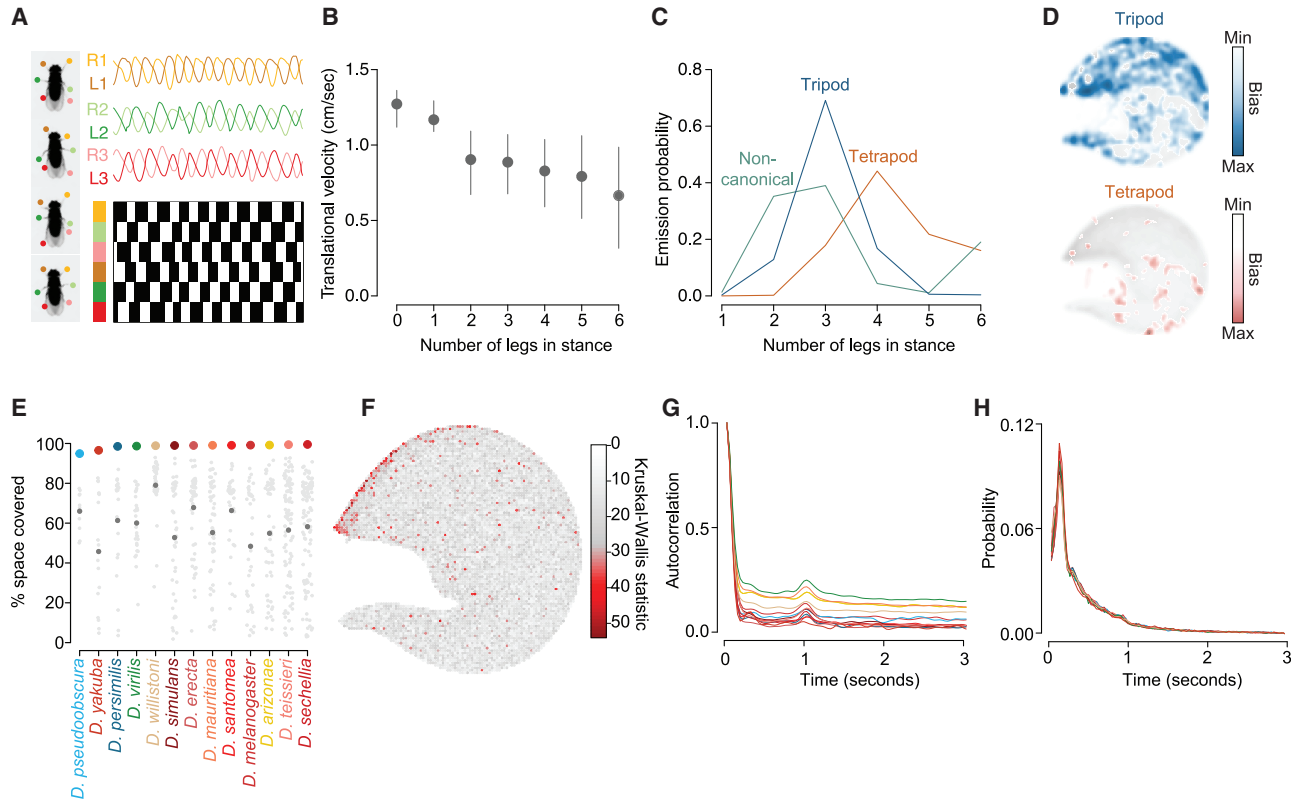


**Figure 3. Biomechanical and temporal characteristics of walking behavior space**

(A) The measurement of gait parameters from videos of freely walking flies. Positions of the fly's 6 tarsi are acquired for each frame (fly images on left) and are then egocentrically aligned and converted to phase (panel top right) from which swing-stance estimates are made (panel bottom right). Sample distributions are from *D. melanogaster*.

(B) The distribution of translational velocity as a function of the number of legs in stance across all genotypes.

(C) Emission probabilities corresponding to the number of legs in stance as a function of HMM state.

(D) The distributions of tripod- and tetrapod-biased densities in behavior space. Darker color corresponds to more bias toward the given gait type.

(E) The percent of behavior space covered by pure species in the dataset. Individual trial percentages are denoted per species as gray points, the mean of which is represented by the larger dark gray point.

(F) The distribution of per-bin significance in variance of occurrence across species (measured by Kruskal-Wallis test statistics calculated across all species). Darker colors correspond to increasing Kruskal-Wallis statistics, representing substantial variation in occurrence across species.

(G) Autocorrelation functions of behavior space position over a 3 s span as species means.

(H) Distributions of the average time taken to return to a point in behavior space, calculated per species.

See also Figure S2 and Table S1.

across all genotypes wherein velocity decreased as the number of legs in stance increased (Figure 3B). We next fit hidden Markov models (HMMs) predicting the number of legs in stance, given a varying number of hidden states (STAR Methods; Table S1). Notably, a model with three hidden states representing tripod, tetrapod, and noncanonical gaits fit the data the best, again matching *D. melanogaster*[22,26,27] (Figure 3C). Although each state displayed a characteristic distribution across behavior space (Figure 3D), tripod gait was associated with a much broader distribution in comparison with the other states, corroborating findings that tripod gait can be employed broadly across walking speeds.[28] These observations suggest that the relationship between gait and kinematics across *Drosophila* is similar to that seen in *D. melanogaster*. Moreover, position in behavior space maps onto identifiable gait patterns.

Although the above observations suggest conservation of the overall structure of behavior space, it is possible that locomotor evolution led to the development of unique kinematic combinations in subsets of lineages. Such changes would result in species- or clade-specific clustering within behavior space. To test this possibility, we analyzed the amount of behavior space covered by each species and observed that each of them explored at least 95% of the space (Figure 3E). However, there was significant variation in the amount of space covered by individual flies from each species (Kruskal-Wallis test; $p = 4.2 \times 10^{-6}$; Figure 3E), reflecting variation in the amount of time spent in specific regions (Figure 3F) that largely corresponded to fast running, stopping, and fast turning. Furthermore, variation in these traits was greater between groups than within, indicating that these differences arise from distinct behavioral differences between genotypes rather than context (Figures S2A and S2B; STAR Methods). In addition, the behavior of each species displayed stationarity (Figure S2C). This finding argues that during each trial, each animal behaves as if in a single "state."

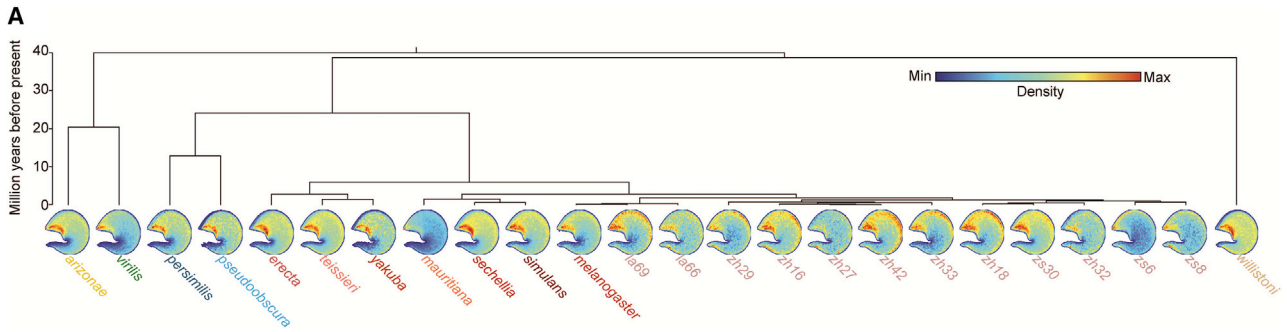


**Figure 4. The phylogenetic distribution of drosophild walking**

(A) The phylogenetic distribution of behavior space frequency maps across all species/strains. Frequency is represented by a heatmap of color ranging from blue to red. Blue regions are less frequently visited; red regions are more frequently visited.

Are the differences in frequency observed above mirrored by changes in the sequencing of movements? Since position in behavior space represents an animal's current movement at a given time point, we reasoned that the autocorrelation of position could be used as a proxy for measuring the temporal structure of behavior. Intuitively, this autocorrelation corresponds to the timescale over which an animal returns to a given region of behavior space. We found a common pattern across all flies in which movements were tightly correlated over several hundred milliseconds, declined rapidly, and rebounded briefly around a second (Figure 3G). However, although each individual species showed this same autocorrelation pattern, their baseline levels of autocorrelation varied in a fashion that mirrored overall differences in behavioral frequency: more active species (e.g., *D. virilis*, *D. mauritiana*) showed higher baseline autocorrelations, reflecting their tendency to perform longer and more frequent bouts of movement (Figure 3G; see below). Interestingly, the distributions of the average time taken to return to specific points in behavior space also showed similar patterns across species but peaked over shorter timescales (Figure 3H). These lines of evidence suggest that the frequency and sequencing of individual movements, but not the range of kinematic parameter combinations, have seen increased differentiation over evolutionary time.

## Tempo and dynamics of the evolution of walking

To explore the evolutionary dynamics of these patterns, we leveraged a fossil-calibrated genome-scale phylogeny of *Drosophila* (STAR Methods)[29] to perform a series of comparative phylogenetic analyses (Figure 4A). In phylogenetic studies of traits such as morphology, the evolutionary diversification of multiple components of a phenotype can be mapped, providing insights into patterns of developmental constraint, trait hierarchy, and phenotypic correlations.[30,31] In these frameworks, multidimensional traits are often represented in a lower dimensional morphological space ("morphospace") that represents the range of trait values possible across the species analyzed. Given that the TREBLE behavior space captures the full range of a behavioral trait (walking kinematics), we wondered whether the structure and dynamics of this space might be studied in a fashion analogous to morphological traits.

We considered three behavioral traits: structure (% of behavior space explored), frequency (probability of occurring in specific regions), and transitions (probability of sequencing between specific regions; STAR Methods) (Figures 5A–5I). This hierarchical description of behavior represents, in order, the movements that are possible, how often they are performed, and how often certain movements precede others. We created a "behavioral morphospace" for these three traits using principal component analysis (PCA) and examined the patterns of variation accounted for by each PC by analyzing their associated eigenvectors (Figure 5; STAR Methods). The first two components of the morphospace for structure separated species by the amount of behavior space covered (Figures 2J and 5A), the variation of which was associated with specific, structured regions of the space (Figures 5B and 5C). On the other hand, frequency and transitions morphospaces yielded striking results. The first two components of the frequency morphospace together explained 72% of variation in the trait (Figure 5D). Comparing the distribution in morphospace with the phylogenetic patterns of space occupancy and their corresponding eigenvectors (Figures 5D–5F) demonstrated that PC1 separated species and strains along specific components of high translational and angular velocities (Figures 5D and 5E; 51.73% variance explained). Conversely, PC2 separated *D. melanogaster* strains and *D. melanogaster*-like species from all others and was associated with variation in stopped/grooming behavior (Figures 5D and 5F; 20.23% variance explained). Notably, both frequency PCs were correlated with the mean translation velocity of each strain and species (PC1: $r = -0.77$; PC2: $r = -0.55$; Pearson correlation), further demonstrating that the mirroring of behavioral space occupancy and kinematic parameters. The first two components of the transitions morphospace explained 81% of variation (Figure 5G). PC1 (66.97% variance explained) was associated with variation in transitions leading to bouts of running (Figures 5G and 5H), whereas PC2 (13.75% variance explained) represented behavioral differences associated with turning sequences (Figures 5G and 5I). Taken together, these results indicate that locomotor diversity across the *Drosophila* phylogeny is driven by variation in specific components of movement, largely related to the occurrence and sequence of running and turning. Furthermore, there is a substantial degree of trait variation not associated with phylogenetic relationships, providing initial evidence of both behavioral convergence across distantly related species and divergence among closely related strains.

We wondered whether morphological variation, particularly body size, might contribute to these striking patterns of

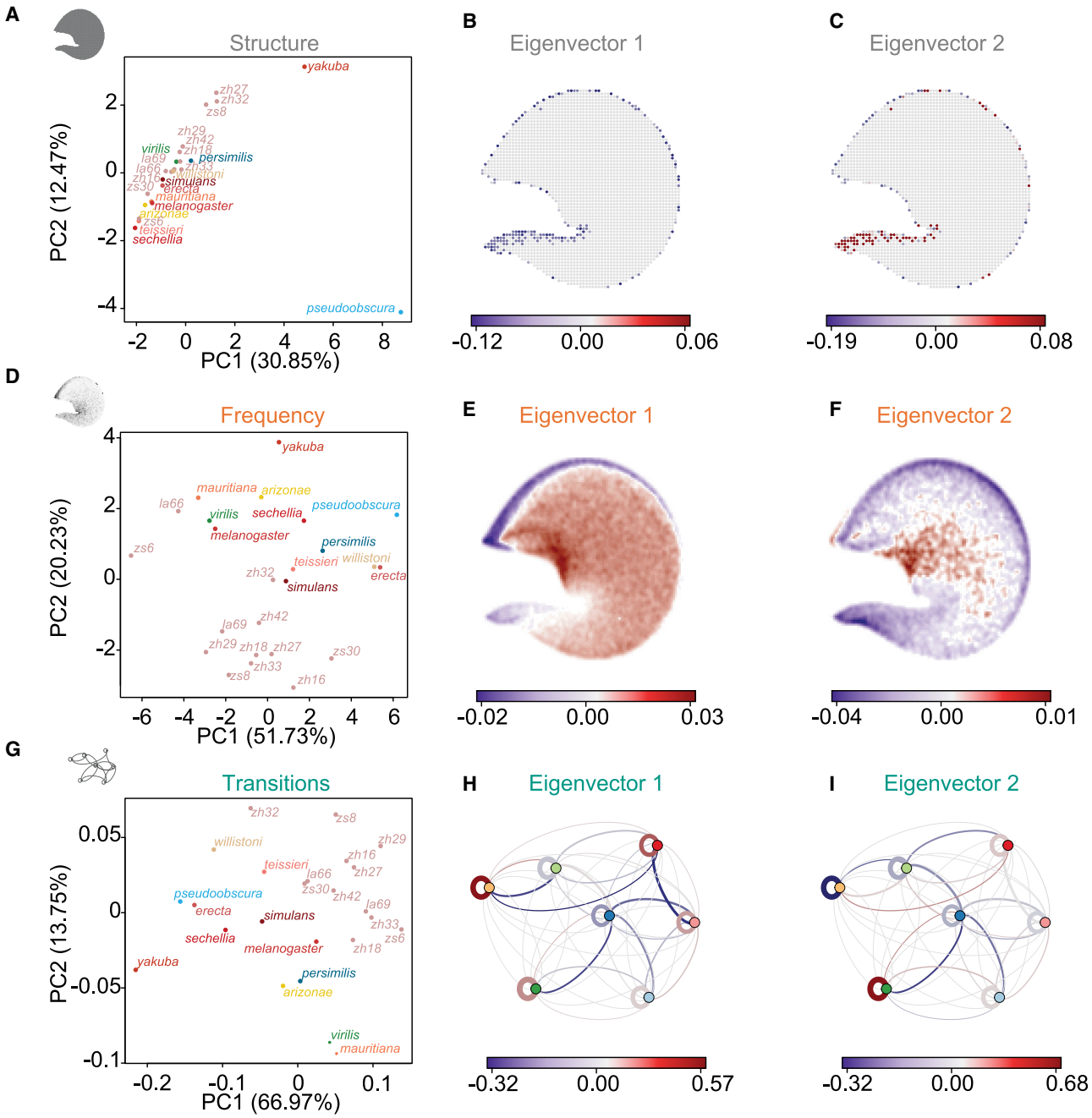


**Figure 5. Morphospace representations of behavioral variables**

(A) Structure morphospace. Variance explained by the first two PCs are denoted in the axes.

(B and C) The 2D representations of the first 2 eigenvectors associated with the structure morphospace. Eigenvector values correspond to the blue-red heatmap.

(D) Frequency morphospace.

(E and F) Eigenvector representations of the frequency morphospace.

(G) Transitions morphospace.

(H and I) Eigenvector representations of the transitions morphospace.

See also Figures S3 and S4.

diversification as our ensemble of species displayed a range of sizes (Figure S4A). To what extent might this variation be predictive of behavior? Although we found mild to moderate correlations between body size and structure, frequency, and transitions (Figures S4B–S4D), species of origin was a significantly better predictor of these behavioral traits than body size (Figures S4E–S4G). We also explored whether measuring translational movement as a fraction of body length, as opposed to

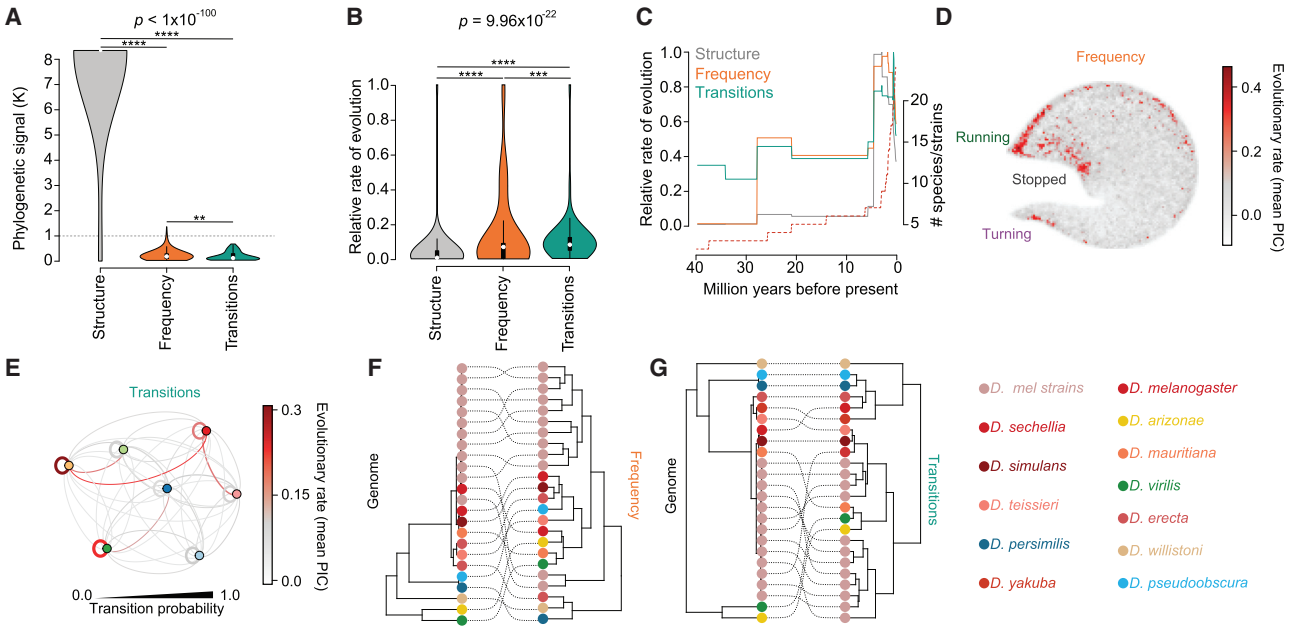


**Figure 6. The dynamics of locomotor evolution across *Drosophila***

(A) Violin plot of the distribution of phylogenetic signal across behavior space for structure, frequency, and transitions (see STAR Methods; Kruskal-Wallis test followed by post-hoc Dunn's test) ****p < 0.0001, ***p < 0.001.

(B) Violin plot of the comparison of relative rate of evolution measurements across all nodes in the phylogeny for all three traits (Kruskal-Wallis test followed by post-hoc Dunn's test) ****p < 0.0001, ** p < 0.001.

(C) Mean relative rates of evolution over time for each trait (computed in 0.1 million year windows). Species accumulation was calculated by summing the number of extant species per 0.1 million year window (inferred by the time-calibrated phylogeny) and is indicated by the dotted red line.

(D) The evolutionary landscape of frequency. Darker colors correspond to greater evolutionary rates (measured by phylogenetic independent contrasts [PICs]).

(E) The evolutionary landscape of transitions. Transition probability is represented by the thickness of the lines connecting nodes (corresponding to Louvain clusters; see STAR Methods). Evolutionary rate corresponds to darkness of color (PICs, phylogenetic independent contrasts).

(F) Cophyloplot comparing the fossil-calibrated whole-genome phylogeny (left) to a phylogeny made from variation in frequency (right). Species are represented by colored nodes at the terminal tips of both phylogenies, with their corresponding positions indicated by dotted lines.

(G) Cophyloplot comparing the fossil-calibrated whole-genome phylogeny (left) to a phylogeny made from variation in transitions (right). Color legend of species labels are plotted to the right.

See also Figures S3–S5.

translational velocity, would affect the observed behavioral distributions. We found that doing so had little effect on the resulting behavior spaces and did not change the overall relationship of species (Figure S4H). Thus, the evolutionary patterns present in our data are not driven by variation in body size.

Next, the evolutionary lability of the behavioral traits was estimated by calculating the distribution of phylogenetic signal for each (STAR Methods). Phylogenetic signal varied as a function of the behavioral hierarchy (Figure 6A). Structure was extremely conserved (mean Blomberg's $K$ = 8.32, SD = 2.14), whereas the mean values of both frequency (mean Blomberg's $K$ = 0.24, SD = 0.14) and transitions (mean Blomberg's $K$ = 0.16, SD = 0.11) were well below 1, suggesting rapid evolution of these traits independent of phylogenetic relationships. Furthermore, the phylogenetic signal distributions across all three traits varied significantly (Kruskal-Wallis test: $p < 1 \times 10^{-100}$) with transitions having a distribution significantly lower than frequency (Dunn's test: p = 0.0088; $z$ = 2.37), suggesting a particularly rapid tempo of evolution in the sequencing of behavior across the *Drosophila* tree.

To assess the temporal dynamics of these patterns, we estimated ancestral states with variable-rate models of evolution and examined variation in morphospace occupation over time. Across the full phylogeny, transitions evolved at the fastest rate (mean normalized relative evolutionary rate = 0.14; STAR Methods) followed by frequency (rate = 0.11) and structure (rate = 0.05). Each trait had a distinct distribution of rates across the phylogeny (Figure 6B) that varied significantly from the others and corroborated the phylogenetic signal analyses. Accordingly, the rate of evolution of structure largely mirrored the accumulation of species (Figures 6C and S5C), increasing in rate in parallel with the diversification of the *D. melanogaster* clade over ~6.7 Ma. On the other hand, frequency and transitions displayed multiple epochs of increases and decreases in rate across the phylogeny (Figures 6C and S5C).

To study the rate of behavioral trait diversification, we examined how each morphospace was populated over evolutionary time (Figures S5A–S5D; STAR Methods). We found that the full range of structure morphospace was sparsely sampled (Figure S5B), covering just 19% of phenotypic values and occurring at a rate slower than species accumulation (Figures S5C and S5D). The frequency morphospace was explored to a greater degree (37%; Figures S5B–S5D) that evolved in tandem with species accumulation until an increase in variation emerged

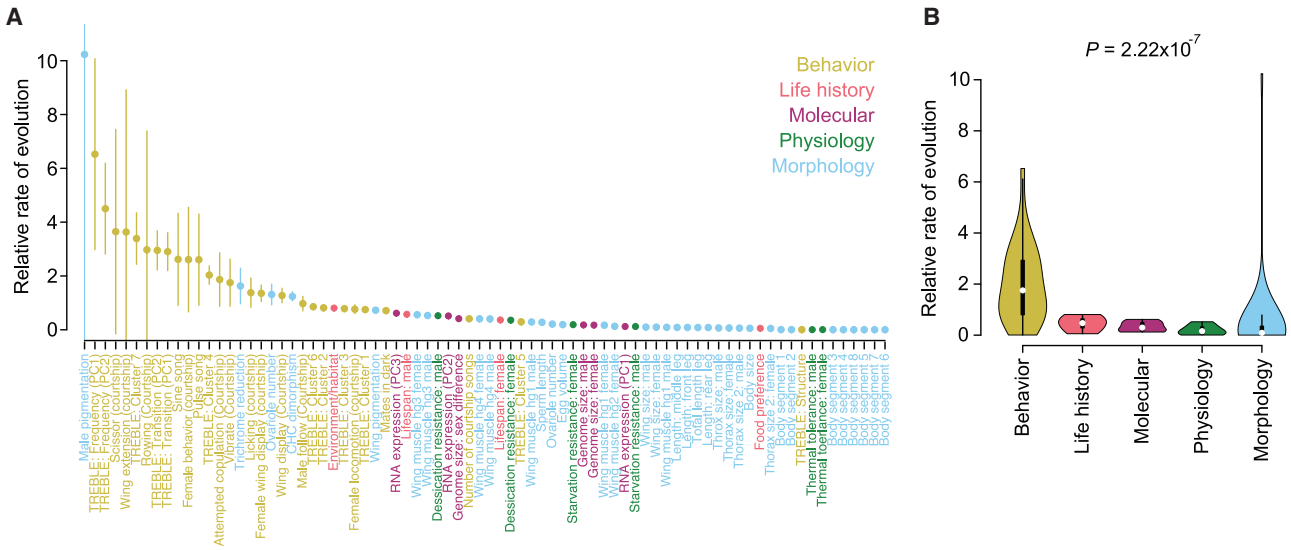


**Figure 7. Rapid, specific, and convergent behavioral evolution in *Drosophila***

(A) Distribution of relative rate of evolution across five categories of *Drosophila* traits: behavior, life history, molecular, physiology, and morphology. Points correspond to median values, bars represent standard error.

(B) Violin plot of relative rates of evolution given *Drosophila* trait type (Kruskal-Wallis test p value).

See also Figure S6, Table S2, and Data S1.

~1.5 Ma (Figures S5C and S5D). Finally, although the transitions morphospace was slightly less explored than frequency (30% covered; Figure S5C), it showed multiple pulses of diversification that outpaced species accumulation (Figures S5C and S5D) and filled earlier than the other traits (Figure S5D).

To what extent might specific components of these traits—movements or the transitions between them—be evolving at different rates to account for these patterns? Strikingly, we found that frequencies of specific movements did not evolve uniformly but rather varied >4-fold across behavior space with especially rapid rates in regions associated with high translational velocity, turning at high speed, and stopping (Figure 6D). Similarly, transitions between specific movements varied in rate disproportionately (Figure 6E). The recurrence of high translational velocity and fast turning movements, and the transitions between them, evolved most rapidly (Figure 6E). Furthermore, comparing species based on overall variation in these patterns indicated evolutionary convergence. Phylogenies produced from variation in frequency and transitions regrouped *D. virilis*, *D. arizonae*, and *D. mauritiana* with a subset of the *D. melanogaster* clade and strains (Figures 6F and 6G).

Taken together, these results imply a hierarchy in the evolution of drosophilid locomotion: transitions are most malleable, frequency evolves slightly less rapidly but with greater variance than transitions, and structure is largely conserved. We note that recent work described a similar relationship between frequency and structure using different methods and a smaller set of species but did not resolve transitions.[32] Furthermore, these findings suggest that the rapid evolution of frequency and transitions resulted in convergence of the walking repertoires of multiple species and strains. Finally, these findings reveal that walking is organized around two fundamentally different strategies, one that is built on longer sequences of high velocity runs and turns and the other on shorter sequences of lower-velocity movements.

## Behavior evolves more rapidly than other traits

Given the hierarchical and rapid evolution described above, we wondered how the diversification of walking might compare with that of other traits. To do so, we conducted a meta-analysis of the evolution of 78 behavioral, morphological, molecular, physiological, and life history traits (Table S2). We normalized trait measurements and used the fossil-calibrated genome tree to calculate evolutionary rate and phylogenetic signal for each (STAR Methods). For these traits, the relative rate of evolutionary change varied over an order of magnitude (Figure 7A). Strikingly, of the 25 most conserved traits, only one was behavioral and, as expected, was structure. Of the 25 most rapidly evolving traits, 21 were behavioral including aspects of frequency, transitions, and a variety of courtship behaviors. Importantly, the remaining 4 rapidly evolving traits were morphological and tightly associated with courtship that is known to evolve rapidly in drosophilids.[13,33,34] Reflecting this pattern, behavioral traits showed the greatest rates of evolution (Figure 7B).

## DISCUSSION

Here, we dissected the evolution of walking in a phylogenetically diverse clade, fruit flies of the genus *Drosophila*. A universal behavior space captured the full range of drosophilid walking kinematics. Variation in hierarchical aspects of the walking behavior space—structure, frequency, and transitions—was captured by individual morphospaces, each with a unique tempo and pattern of evolutionary diversification. Although structure was conserved, frequency and transitions displayed multiple

pulses of phenotypic evolution, resulting in species and strains that converge on common movement patterns. Furthermore, behaviors were associated with significantly faster rates of evolution than morphology, life history, physiology, and molecular traits, results that are consistent with previous studies.[35,36] Whether this pattern holds across other systems will be of great interest. Such comparative work could resolve the long-standing debate of whether behavior is a facilitator, or inhibitor, of phenotypic variation during organismal evolution.[36–42] In particular, notwithstanding the need to perform comparable analyses using other behaviors and clades, by revealing relatively rapid evolutionary change in the behaviors of one model clade, our work supports the idea that behavior can indeed facilitate the emergence of phenotypic variation.

Our data demonstrate that sequences of motor actions, rather than individual movements, have evolved rapidly. Furthermore, divergent species and strains have converged on similar behavioral patterns multiple times, whereas closely related strains that live in comparable natural environments can be highly divergent behaviorally. Thus, the evolutionary patterns we observed appear robust to the specifics of our behavioral assay. In some cases, behavioral convergence has occurred independent of life history and ecological background. For example, species from tropical (*D. mauritiana*) and temperate (*D. virilis*) habitats have converged on similar behavioral repertoires to those seen in the cosmopolitan *D. melanogaster*. In the same vein, very closely related *D. melanogaster* strains isolated from a single subregion in Africa can also diverge markedly in their walking behavioral repertoires, spanning distances in behavior space comparable with that seen in divergent species pairs. Future work should explore the evolutionary and coevolutionary relationships between behavioral structure and traits such as physiological regulation and anatomical differences across *Drosophila* species. Furthermore, it will be especially interesting to if these intraspecific patterns are present among other species populations (e.g., North American populations of *D. melanogaster* or *D. persimilis*) and if such variation arises from identifiable evolutionary processes such as adaptation or drift.

Studies of a wide range of both social and nonsocial behaviors, in many animals, have identified discrete, modular elements that can be arranged in temporal sequences of varying flexibility.[14–16,43–52] Here, we compare the structure of a common behavioral repertoire across a densely sampled set of species and strains to show that the apparent focus of rapid evolutionary change lies in changing the temporal sequences of individual movements, with substantially less variation in the fine structure of the movements themselves. Thus, the modularity of behavior seen in individual species in fact appears to reflect the structure of evolutionary change. These observations, in tandem with comparisons with other traits, suggest that the diversification of walking arises first via changes in neural control, as opposed to biomechanical or morphological mechanisms that would alter the fine structure of individual movements. In this case, behavioral diversification may be tightly coupled with neural evolution and labile over even very short evolutionary timescales, such as those seen within a species. Thus, the architecture and function of neural circuits appears to both enable complex behavior and facilitate its rapid diversification.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Fly strains
- METHOD DETAILS
  - Behavior
  - Generating a universal behavior space with TREBLE
  - Gait analysis
  - Phylogenetic analyses

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2022.05.039.

## ACKNOWLEDGMENTS

We thank members of the Clandinin, Tuthill, and Matute labs and members of the NIH U19 FlyLoops consortium for useful discussion. This work was supported by the NIH (U19NS104655: T.R.C. and J.C.T.; R01GM121750: D.R.M.), the Simons Foundation (T.R.C.), the Stanford School of Medicine (R.A.Y.), the New York Stem Cell Foundation (J.C.T.), a Searle Scholar Award (J.C.T.), a Klingenstein-Simons Fellowship (J.C.T.), a Pew Biomedical Scholar Award (J.C.T.), a Sloan Research Fellowship (J.C.T.), and NSF Dimensions of Biodiversity award 1737752 (D.R.M.). J.C.T. is a New York Stem Cell Foundation—Robertson Investigator. Finally, we want to acknowledge that our laboratory at the Stanford University stands on the ancestral lands of the Muwekma Ohlone Tribe (http://muwekma.org/home.html).

## AUTHOR CONTRIBUTIONS

R.A.Y. and T.R.C. conceived the study. L.E.B., S.H., A.K., A.D.S., E.S., and R.A.Y. developed the Coliseum and FlyVR software. R.A.Y. collected the walking data set. S.-Y.L., B.P., and J.C.T. performed gait tracking. J.C., A.D.S., and D.R.M. compiled and prepared the whole-genome phylogeny. R.A.Y. curated the data and performed statistical analysis. R.A.Y. and T.R.C. prepared the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 6, 2021
Revised: March 3, 2022
Accepted: May 13, 2022
Published: June 3, 2022

## REFERENCES

1. Katz, P.S. (2019). The conservative bias of life scientists. Curr. Biol. *29*, R666–R667.
2. Gallant, J.R., and O'Connell, L.A. (2020). Studying convergent evolution to relate genotype to behavioral phenotype. J. Exp. Biol. *223*.
3. Niepoth, N., and Bendesky, A. (2020). How natural genetic variation shapes behavior. Annu. Rev. Genomics Hum. Genet. *21*, 437–463.

4. Jourjine, N., and Hoekstra, H.E. (2021). Expanding evolutionary neuroscience: insights from comparing variation in behavior. Neuron *109*, 1084–1099.
5. Pereira, T.D., Shaevitz, J.W., and Murthy, M. (2020). Quantifying behavior to understand the brain. Nat. Neurosci. *23*, 1537–1549.
6. Mathis, A., Schneider, S., Lauer, J., and Mathis, M.W. (2020). A primer on motion capture with deep learning: principles, pitfalls, and perspectives. Neuron *108*, 44–65.
7. von Ziegler, L., Sturman, O., and Bohacek, J. (2021). Big behavior: challenges and opportunities in a new era of deep behavior profiling. Neuropsychopharmacology *46*, 33–44.
8. York, R.A., Carreira-Rosario, A., Giocomo, L.M., and Clandinin, T.R. (2022). Flexible analysis of animal behavior via time-resolved manifold embedding. Preprint at bioRxiv. https://doi.org/10.1101/2020.09.30.321406.
9. Piel, W.H., Chan, L., Dominus, M.J., Ruan, J., Vos, R.A., and Tannen, V. (2009). TreeBASE v. 2: A Database of Phylogenetic Knowledge. (e-BioSphere).
10. Markow, T.A., and O'Grady, P.M. (2005). Evolutionary genetics of reproductive behavior in *Drosophila*: connecting the dots. Annu. Rev. Genet. *39*, 263–291.
11. Markow, T.A. (2015). The secret lives of *Drosophila* flies. eLife *4*, e06793.
12. Hales, K.G., Korey, C.A., Larracuente, A.M., and Roberts, D.M. (2015). Genetics on the fly: a primer on the *Drosophila* model system. Genetics *201*, 815–842.
13. Jezovit, J.A., Levine, J.D., and Schneider, J. (2017). Phylogeny, environment and sexual communication across the *Drosophila* genus. J. Exp. Biol. *220*, 42–52.
14. Berman, G.J., Choi, D.M., Bialek, W., and Shaevitz, J.W. (2014). Mapping the stereotyped behaviour of freely moving fruit flies. J. R. Soc. Interface *11*, 20140672.
15. Katsov, A.Y., Freifeld, L., Horowitz, M., Kuehn, S., and Clandinin, T.R. (2017). Dynamic structure of locomotor behavior in walking fruit flies. eLife *6*, e26410.
16. Tao, L., Ozarkar, S., Beck, J.M., and Bhandawat, V. (2019). Statistical structure of locomotion and its modulation by odors. eLife *8*, e41235.
17. Bidaye, S.S., Machacek, C., Wu, Y., and Dickson, B.J. (2014). Neuronal control of *Drosophila* walking direction. Science *344*, 97–101.
18. Carreira-Rosario, A., Zarin, A.A., Clark, M.Q., Manning, L., Fetter, R.D., Cardona, A., and Doe, C.Q. (2018). MDN brain descending neurons coordinately activate backward and inhibit forward locomotion. eLife *7*, e38554.
19. Namiki, S., Dickinson, M.H., Wong, A.M., Korff, W., and Card, G.M. (2018). The functional organization of descending sensory-motor pathways in *Drosophila*. eLife *7*, e34272.
20. Cande, J., Namiki, S., Qiu, J., Korff, W., Card, G.M., Shaevitz, J.W., Stern, D.L., and Berman, G.J. (2018). Optogenetic dissection of descending behavioral control in *Drosophila*. eLife *7*, e34275.
21. Feng, K., Sen, R., Minegishi, R., Dübbert, M., Bockemühl, T., Büschges, A., and Dickson, B.J. (2020). Distributed control of motor circuits for backward walking in *Drosophila*. Nat. Commun. *11*, 6166.
22. DeAngelis, B.D., Zavatone-Veth, J.A., and Clark, D.A. (2019). The manifold structure of limb coordination in walking *Drosophila*. eLife *8*, e46409.
23. McInnes, L., Healy, J., and Melville, J. (2018). UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at arXiv. https://doi.org/10.48550/arXiv.1802.03426.
24. Mendes, C.S., Bartos, I., Akay, T., Márka, S., and Mann, R.S. (2013). Quantification of gait parameters in freely walking wild type and sensory deprived *Drosophila melanogaster*. eLife *2*, e00231.
25. Wosnitza, A., Bockemühl, T., Dübbert, M., Scholz, H., and Büschges, A. (2013). Inter-leg coordination in the control of walking speed in *Drosophila*. J. Exp. Biol. *216*, 480–491.
26. Isakov, A., Buchanan, S.M., Sullivan, B., Ramachandran, A., Chapman, J.K., Lu, E.S., Mahadevan, L., and de Bivort, B. (2016). Recovery of locomotion after injury in *Drosophila melanogaster* depends on proprioception. J. Exp. Biol. *219*, 1760–1771.
27. Pereira, T.D., Aldarondo, D.E., Willmore, L., Kislin, M., Wang, S.S., Murthy, M., and Shaevitz, J.W. (2019). Fast animal pose estimation using deep neural networks. Nat. Methods *16*, 117–125.
28. Chun, C., Biswas, T., and Bhandawat, V. (2021). *Drosophila* uses a tripod gait across all walking speeds, and the geometry of the tripod is important for speed control. eLife *10*, e65878.
29. Suvorov, A., Kim, B.Y., Wang, J., Armstrong, E.E., Peede, D., D'Agostino, E.R.R., Price, D.K., Waddell, P., Lang, M., Courtier-Orgogozo, V., et al. (2022). Widespread introgression across a phylogeny of 155 *Drosophila* genomes. Curr. Biol. *32*, 111. e5–123.e5.
30. Adams, D.C., and Collyer, M.L. (2009). A general framework for the analysis of phenotypic trajectories in evolutionary studies. Evolution *63*, 1143–1154.
31. Felice, R.N., Randau, M., and Goswami, A. (2018). A fly in a tube: macroevolutionary expectations for integrated phenotypes. Evolution *72*, 2580–2594.
32. Hernández, D.G., Rivera, C., Cande, J., Zhou, B., Stern, D.L., and Berman, G.J. (2021). A framework for studying behavioral evolution by reconstructing ancestral repertoires. eLife *10*, e61806.
33. Spieth, H.T. (1952). Mating behavior within the genus *Drosophila* (Diptera). Bull. Am. Mus. Nat. Hist. *99*, 395–474.
34. O'Grady, P.M., and Markow, T.A. (2012). Rapid morphological, behavioral, and ecological evolution in *Drosophila*: comparisons between the endemic Hawaiian Drosophila and the cactophilic repleta species group. In Rapidly Evolving Genes and Genetic Systems (Oxford University Press).
35. Gittleman, J.L., Anderson, C.G., Kot, M., Luk, H.-K., and Martins, E.P. (1996). Phylogenetic lability and rates of evolution: a comparison of behavioral, morphological and life history traits. In Phylogenies and the Comparative Method in Animal Behavior (Oxford University Press).
36. Blomberg, S.P., Garland, T., Jr., and Ives, A.R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution *57*, 717–745.
37. Lamarck, J.-B. (1809). Zoological Philosophy (University of Chicago Press).
38. Darwin, C. (1859). On the Origin of Species (Harvard University Press).
39. Huxley, J.S. (1942). Evolution, the Modern Synthesis (Allen & Unwin)).
40. Simpson, G.G. (1944). Tempo and Mode in Evolution (Columbia University Press).
41. Mayr, E. (1982). Toward a New Philosophy of Biology: Observations of an Evolutionist (Harvard University Press).
42. Wcislo, W.T. (1989). Behavioral environments and evolutionary change. Annu. Rev. Ecol. Syst. *20*, 137–169.
43. Tinbergen, N. (1951). The Study of Instinct (Clarendon Press).
44. Tinbergen, N. (1954). The hierarchical organization of nervous mechanisms underlying instinctive behaviour. Symp. Soc. Exp. Biol. *4*, 305–312.
45. Dawkins, R. (1976). Hierarchical organization: a candidate principle for ethology. In Growing Points in Ethology, P.P.G. Bateson, and R.A. Hinde, eds. (Cambridge University Press), pp. 7–45.
46. Herrel, A., Meyers, J.J., Nishikawa, K.C., and De Vree, F. (2001). The evolution of feeding motor patterns in lizards: modulatory complexity and possible constraints. Am. Zool. *41*, 1311–1320.
47. Jin, X., and Costa, R.M. (2015). Shaping action sequences in basal ganglia circuits. Curr. Opin. Neurobiol. *33*, 188–196.
48. Wiltschko, A.B., Johnson, M.J., Iurilli, G., Peterson, R.E., Katon, J.M., Pashkovski, S.L., Abraira, V.E., Adams, R.P., and Datta, S.R. (2015). Mapping sub-second structure in mouse behavior. Neuron *88*, 1121–1135.

49. Duistermars, B.J., Pfeiffer, B.D., Hoopfer, E.D., and Anderson, D.J. (2018). A brain module for scalable control of complex, multi-motor threat displays. Neuron *100*, 1474e4–1490.e4.

50. Marques, J.C., Lackner, S., Félix, R., and Orger, M.B. (2018). Structure of the zebrafish locomotor repertoire revealed with unsupervised behavioral clustering. Curr. Biol. *28*, 181e5–195.e5.

51. Kaplan, H.S., Salazar Thula, O., Khoss, N., and Zimmer, M. (2020). Nested neuronal dynamics orchestrate a behavioral hierarchy across timescales. Neuron *105*, 562e9–576.e9.

52. Johnson, R.E., Linderman, S., Panier, T., Wee, C.L., Song, E., Herrera, K.J., Miller, A., and Engert, F. (2020). Probabilistic models of larval zebrafish behavior reveal structure on many scales. Curr. Biol. *30*, 70. e4–82.e4.

53. R Development Core Team (2020). R: A language and environment for statistical computing (R Foundation for Statistical Computing).

54. Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W., and Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat. Neurosci. *21*, 1281–1289.

55. Coughlan, J.M., Dagilis, A.J., Serrato-Capuchina, A., Elias, H., Peede, D., Isbell, K., Castillo, D.M., Cooper, B.S., and Matute, D.R. (2021). Population structure and introgression among recently differentiated Drosophila melanogaster populations. Preprint at bioRxiv. https://doi.org/10.1101/2021.06.25.449842.

56. Shultzaberger, R.K., Johnson, S.J., Wagner, J., Ha, K., Markow, T.A., and Greenspan, R.J. (2019). Conservation of the behavioral and transcriptional response to social experience among drosophilids. Genes Brain Behav. *18*, e12487.

57. Karashchuk, P.R., Rupp, K.L., Dickinson, E.S., Walling-Bell, S., Sanders, E., Azim, E., Brunton, B.W., and Tuthill, J.C. (2021). Anipose: a toolkit for robust markerless 3D pose estimation. Cell Rep. *36*, 109730.

58. Bowman, D.C., and Lees, J.M. (2013). The Hilbert-Huang transform: a high resolution spectral method for nonlinear and nonstationary time series. Seismol. Res. Lett. *84*, 1074–1080.

59. developers, signal (2014). signal: signal processing. http://r-forge.r-project.org/projects/signal/.

60. Visser, I., and Speekenbrink, M. (2010). depmixS4: an R package for hidden markov model. J. Stat. Software *36*, 1–21.

61. Levine, J.H., Simonds, E.F., Bendall, S.C., Davis, K.L., Amir, El-A.D., Tadmor, M.D., Litvin, O., Fienberg, H.G., Jager, A., Zunder, E.R., et al. (2015). Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. Cell *162*, 184–197.

62. Spedicato, G.A. (2017). Discrete time markov chains with R. The R Journal *9*, 84–104.

63. Huey, R.B., Moreteau, B., Moreteau, J.-C., Gibert, P., Gilchrist, G.W., Ives, A.R., Garland, T., and David, J.R. (2006). Sexual size dimorphism in a *Drosophila* clade, the *D. obscura* group. Zoology (Jena) *109*, 318–330.

64. Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. *3*, 217–223.

65. Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. R News *6*, 7–11.

66. Cooney, C.R., Bright, J.A., Capp, E.J.R., Chira, A.M., Hughes, E.C., Moody, C.J.A., Nouri, L.O., Varley, Z.K., and Thomas, G.H. (2017). Mega-evolutionary dynamics of the adaptive radiation of birds. Nature *542*, 344–347.

67. Ronco, F., Matschiner, M., Böhne, A., Boila, A., Büscher, H.H., El Taher, A., Indermaur, A., Malinsky, M., Ricci, V., Kahmen, A., et al. (2021). Drivers and dynamics of a massive adaptive radiation in cichlid fishes. Nature *589*, 76–81.

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

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Deposited data | | |
| Dryad | This paper | https://doi.org/10.5061/dryad.z8w9ghxfc |
| Experimental models: Organisms/strains | | |
| *D. melanogaster* | Thomas Clandinin Lab | IsoD1 |
| *D. willistoni* | National Drosophila Species Stock Center (NDSSC) | (Heed) H57.30 |
| *D. santomea* | NDSSC | (Gompel) STO.4 |
| *D. persimilis* | NDSSC | 2529.6 |
| *D. mauritiana* | NDSSC | E18912 MS17 |
| *D. virilis* | NDSSC | 3367.1 |
| *D. arizonae* | NDSSC | N/A |
| *D. yakuba* | NDSSC | N/A |
| *D. sechellia* | NDSSC | N/A |
| *D. simulans* | NDSSC | N/A |
| *D. pseudoobscura* | NDSSC | N/A |
| *D. erecta* | NDSSC | N/A |
| *D. teissieri* | NDSSC | N/A |
| *D. melanogaster* | Daniel Matute lab | La69 |
| *D. melanogaster* | Daniel Matute lab | La66 |
| *D. melanogaster* | Daniel Matute lab | Zh29 |
| *D. melanogaster* | Daniel Matute lab | Zh16 |
| *D. melanogaster* | Daniel Matute lab | Zh27 |
| *D. melanogaster* | Daniel Matute lab | Zh42 |
| *D. melanogaster* | Daniel Matute lab | Zh33 |
| *D. melanogaster* | Daniel Matute lab | Zh18 |
| *D. melanogaster* | Daniel Matute lab | Zs30 |
| *D. melanogaster* | Daniel Matute lab | Zh32 |
| *D. melanogaster* | Daniel Matute lab | Zs6 |
| *D. melanogaster* | Daniel Matute lab | Zs8 |
| Software and algorithms | | |
| FlyVR | Thomas Clandinin Lab | github.com/ClandininLab/flyvr |
| TREBLE | York et al.[8] | github.com/ryanayork/TREBLE |
| Evolutionary analyses | This paper | Github.com/ryanayork/fly_locomotor_evolution |
| R Version 3.6.1 | R Core Team[53] | r-project.org/ |
| Gait tracking algorithm | This paper | github.com/Prattbuw/CODE-Evolution-of-Drosophila-Walking |
| DeepLabCut | Mathis et al.[54] | github.com/DeepLabCut |
| Other | | |
| Stepper motors | | SM42HT47-1684B |
| Arduino UNO | | N/A |
| Gshield | | N/A |
| CNC mill | | N/A |
| Camera | Parallax | TSL1401 Linescan Sensor |
| Camera | | N/A |

(*Continued on next page*)

*Continued*

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Camera lens | | N/A |
| Solenoid | Mouser | 474-ROB-11015 |
| LEDs | Mouser | 720-SFH4235 |
| Optic Fiber | Edmund optics | N/A |

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Thomas R. Clandinin (trc@stanford.edu).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

- All data reported in this paper are available on Dryad (https://doi.org/10.5061/dryad.z8w9ghxfc).
- All original code has been deposited on github and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Fly strains

All stocks were kept at 25°C on molasses-based food and reared under a light-dark cycle of 12:12 h. The following stocks were obtained from National Drosophila Species Stock Center: *D. willistoni* ((Heed) H57.30), *D. santomea* ((Gompel)STO.4), *D. persimilis* (2529.6), *D. mauritiana* (E18912 MS17), *D. virilis* (3367.1), *D. arizonae*, *D. yakuba*, *D. sechellia*, *D. simulans*, *D. pseudoobscura*, *D. erecta*, and *D. teissieri*. Strains of *D. melanogaster* were collected in Southern Africa, the ancestral range of the species (Coughlan et al.,[55]).

## METHOD DETAILS

### Behavior

#### *Apparatus*

The Coliseum is an enclosed 1m x 1m arena for measuring the unconstrained walking of fruit flies. The arena is sealed from external light via Velcro-attached curtains on the sides and solid walls on the top and bottom. Flies are released into the arena through a hole in the floor by means of an automated dispenser consisting of a vial filled with flies and a servo-gated exit (details in Data S2). Flies enter the arena by crawling up the exit channel. An optical sensor detects that a fly has entered the channel and immediately closes the gate behind the fly to both avoid releasing multiple flies simultaneously and to prevent flies from returning to the vial from the arena. Once a fly is in the arena, it is edge-lit by IR LEDs around the perimeter of the floor and recorded from above by a high-definition camera outfitted with a zoom lens that is sufficiently powerful to capture anatomical details at high resolution. To keep the fly in its field of view the camera is mounted on a 2-axis CNC mill (grbl with gShield + Arduino UNO; stepper motors are SM42HT47-1684B) that updates the camera position as the fly moves.

The position of the CNC mill is controlled by the software flyvr (github.com/ClandininLab/flyvr) by tracking the position of the fly and the camera simultaneously. The fly's position in the camera is first computed by thresholding and extracting the pixels representing the fly, identifying the head-tail axis and orientation, and then calculating the in-frame coordinates of the fly's center. This relative position is then summed with the position of the camera to calculate an absolute position in the Coliseum which is used to update the stepper-motor coordinates of the CNC and thus keep the camera in sync with the fly's movement. The absolute position and heading angle of the fly are recorded for each frame (at 100Hz) and outputted by flyvr at the end of each trial.

#### *Behavior experiments*

Immediately after eclosion, virgin female flies were sorted into vials of 10-20 flies and reared in light-dark chambers (12:12h) at 25°C. 2-4 day old flies were used for tracking in the Coliseum. On the day of the experiment, individual vials would be loaded directly into the automated dispenser and flies would enter the Coliseum one at a time. Flies were allowed to explore the arena freely for up to 20 minutes, after which the animal would be manually removed from the chamber. The floor of the Coliseum was cleaned with

70% ethanol between trials to remove odorants or other stimuli that may affect patterns of locomotor behavior. All experiments were conducted during the same time window to align with the light-dark cycle, from roughly CT0-CT3.

Age, diet, circadian rhythm, and environment were controlled to facilitate interspecific comparisons, potentially influencing individual species' expression of locomotor behavior. However, a standardized diet has minimal effects on the behavior of diverse species, even on specialist species such as *D. sechellia*, *D. arizonae*, and *D. erecta*.[56] Future work comparing the locomotion of wild-caught flies in natural conditions would be useful for corroborating the controlled laboratory experiment presented here.

## Generating a universal behavior space with TREBLE

### *Determining window size*

Following previous work on the kinematics of walking in *D. melanogaster*[8,15] we calculated the following kinematic parameters for each trial: translational velocity (cm/second), angular velocity (degree/second), and sideslip (cm/second). Per-frame position and heading angles from the Coliseum were used for these calculations. As in York et al.,[8] angular and sideways velocity values were normalized to the first frame in the window and proceeding velocity values were adjusted to ensure that the second frame was always positive. Since we weren't concerned with turning direction, the absolute value of angular velocity was used. To control for rare occasions of jitter in the camera movement we smoothed each velocity parameter using a Nadaraya-Watson kernel estimate (ksmooth function in R; bandwidth = 0.25) and down sampled the data to 30Hz (which removed noise without affecting the structure of the velocity parameters).

A core component of the TREBLE framework is identifying the temporal structure of a behavioral dataset by iteratively sampling data in increasing window sizes and empirically analyzing variation in the downstream behavior spaces.[8] Here, we performed this iterative window search by randomly sampling 50 trials and testing the effect of window sizes (from 10ms to 1 second, 20ms intervals) on the structure and temporal properties of behavior space (Figures S1E–S1I). The goal of this procedure is to identify a window size that minimizes variation across replicates in both the structure and temporal sequencing of movement through behavior space. Variation in behavior space structure was assessed by measuring the Procrustes and mean inter-point Euclidean distances between all replicates (Figures S1F and S1G). Here, Procrustes distance is a measure of the global differences between two spaces.[8] First, the sets are scaled to match each other in size, shifted to match positions in space, rotated until the distances between the points are minimized, and then compared by calculated the square root of the sum of squared differences between the two configurations of points.[23] Mean inter-point Euclidean distance complements this global assessment by providing a local measurement (i.e. between adjacent points) of structural variation. In both cases the mean stabilized, and variance decreased, at windows around 300ms in size.

Temporal variation was measured by analyzing the timing of recurrence in behavior space, based on the concept that these are essentially state spaces reflecting the function of a continuous dynamic system.[8] Recurrence was measured by calculating the duration of time it took to return to the local neighborhood of each point in behavior space. The proportions of points that were recurrent as a function of 10ms temporal bins, ranging from 0 to 2 seconds, were then measured (Figure S1E). From this we obtained the mean recurrence time across all temporal bins (Figure S1H) and the overall proportion of recurrent points across all bins (Figure S1I) for each window size. As has been seen previously in studies of fruit fly walking,[8] there was a narrow range of window sizes (~140-360ms) that displayed a peak of recurrence at around 250ms (Figure S1E) and dissipated with increasing window size. Taken together, these analyses indicated that a window size of 333ms optimally reduced variation in the structure of behavior space while capturing the temporal patterns present.

### *Embedding temporal windows into behavior space*

We next produced windows for all trials using the size identified above. For each frame of a given trial (denoted here as frame i), and for a window size w, we individually extracted the kinematic parameters (x) for time i:i+w and then concatenated them linearly into a single vector of length 3x. These window vectors were then collected into a single matrix spanning all flies and containing 3,890,645 windows. Further details of this procedure can be found in York et al.[8] We then used the R implementation of the UMAP algorithm[23] to non-linearly embed the windows into a 2-dimensional behavior space containing all flies from all species and strains. To facilitate downstream analyses, we also produced version of the space with simplified 2-d coordinates by decomposing each point onto a grid of size *n* bins x *n* bins using the TREBLE function bin_space. Most analyses described below use a 64x64 grid size, containing 4,096 unique bins, unless otherwise noted.

### *Characterizing behavior space*

Stereotypy in movement through behavior space was assessed by creating a vector field representation. Using the 64x64 grid representation of behavior space, we identified all points associated with a given bin and then calculated the mean direction for all instances of a trajectory leaving that bin (in xy coordinates). Mean direction was represented by plotting the mean trajectory out of each bin via an arrow, the direction and magnitude of corresponded to the xy coordinates calculate above (as in Figure 2D). Arrow color reflected the polar coordinates of the vector for each bin.

Variation in the percent of behavior space covered by individuals and species/strains (Figure 3E) was measured by calculating the number of unique bins in behavior space were visited as a function of the total number of bins. The significance of per-bin variance as function of species (Figure 3F) was measured using a Kruskal-Wallis test. First, we calculated the total proportion of time spent in each bin (as a fraction of overall trial time) for all flies. We then ran a Kruskal-Wallis test (kruskal.test function in the R package stats) for each bin, comparing species and using the bin-wise proportions from each fly of a given species to account for intraspecific variation.

Patterns of temporal sequencing through behavior space were inferred by calculating the autocorrelation of position in behavior space. To do so, we assigned each bin in behavior space a unique 1-d identifier, creating a single numeric representation of behavior space position (a metric outputted by bin_space). This vector was then used to calculate autocorrelation using the acf function in R (lag time = 3 seconds), for all species (Figure 2G). Analyses of the return time distribution to specific portions of behavior space were calculated in the same fashion as the recurrence tests run during the iterative window search procedure (outlined above). Here, we measured recurrence timing across all bins for each species and then measured the mean distribution of return times over a span of 3 seconds (as seen in Figure 2H).

#### *Behavior space power tests, intra- inter-species variance, and inter-trial variance*

Animal behaviors can be influenced by a variety of factors - such as context, physiological state, or personality - in addition to genotype.[42] If the variance of a given behavior is strongly driven by contextual factors, then identifying consistent species differences may be difficult or impossible. With this in mind, we sought to calculate the intra-individual variability in behavior space statistics and to compare this to the variance observed with and between strains and species.

To address the first goal, we performed a power test exploring the number of individuals needed to capture the overall structural and frequency statistics within behavior space for each species. First, we calculated probability density functions (pdfs) of xy coordinates in behavior space for each species using all individuals (grid points = 100, bandwidth = 1). We then compared pdfs calculated using a range of included trials (1 to 10 trials) to the full pdf for each species using correlation. To factor in intra-trial variability, we bootstrapped each trial sample size 10 times (i.e., 2 trials were randomly chosen 10 times and correlated to the full pdf; then 3 trials were randomly chosen 10 times…), randomly permuting which trials were used per species. Comparing mean correlations across permutations, we found that each species and strain had a high level of autocorrelation with a small number of trials included, for most reaching a correlation >0.9 using just 6 trials and converging >0.95 with 10 trials used (Figures S2A and S2B). These results indicated that, for the given behavior and experimental set up, intra-species variation in walking could be captured well within the sample sizes collected (Figure S2A).

We next assessed the relationships of intra-individual, intra-species, and inter-species variation by bootstrapping behavior and comparing distributions in behavior space using cosine similarity. We first selected trials longer than 5,000 frames (166.66 seconds), resulting in 373 trials. Intra-individual variance was measured by randomly selecting 2,000 time points 10 times per trial (without replacement), resulting in a total 3,730 permutations. A pdf (grid points = 100, bandwidth = 1) was calculated each permutation. The intra-trial similarity was calculated by computing the cosine similarity of the pdfs from the 10 permutations per trial (Figure S2B). Intra-strain variance was computed in the same fashion except for 100 bootstraps we performed per strain (Figure S2B). Interspecific variance was measured by bootstrapping xy coordinates (2,000 per shuffled) from the full data set 1,000 times (Figure S2B). The distributions of all three measures were compared using a Kruskal-Wallis test followed by a post-hoc Dunn's test (dunn.test function in the R package dunn.test). We found that all three significantly differed. Intra-individual similarity was greater than intra-strain while intra-strain similarity was greater than inter-strain (Figure S2B).

Stationarity was assessed by splitting each trial into evenly sized halves, comparing the halves' distributions within behavior space via correlation, and then analyzing the mean intra-trial correlations for each species. We reasoned that, for stationary behavior, the two trial portions should be highly correlated. On the other hand, if the behavior of the flies was governed by multiple states within a single trial, then the portions would be uncorrelated (and would be reflected by different distributions within behavior space). We found that each species displayed a correlation of at least 0.8 (Figure S2C), representing strong intra-trial correlations and indicating that the our data appear to overwhelmingly represent stationary behavior.

### Gait analysis

#### *Limb tracking*

Behavior videos were first contrast enhanced using custom Matlab scripts to optimize the view of the animal's legs. The head, thorax, abdomen, and leg tarsi were automatically tracked from the top-down videos using DeepLabCut and Anipose.[54,57] Tracked data points were used to analyze walking kinematics using a custom Python script (https://github.com/Prattbuw/CODE-Evolution-of-Drosophila-Walking). Position time series data associated with the tarsi was smoothed using a moving average with a time window of ~80ms. Head, abdomen, and leg tarsi positions were normalized to the thorax to calculate positions relative to the body. Heading angle was calculated based on head and thorax positions relative to an allocentric reference (i.e. experimental chamber). Tarsi positions were rotated and translated based on deviations from a common heading angle (allocentric angle of 0°). This transformation was necessary for extracting stance and swing onsets as this maximizes the position signal of each leg tarsi signal along the body axis. Steps were extracted by identifying the peaks and troughs of the tarsi position signals. Steps consist of two phases: swing (trough to peak) and stance (peak to trough). Number of legs in stance was calculated by summing the number of legs touching the ground at each sample point. Leg phase (as in Figure 3A) was calculated using a Hilbert transformation of the position of each leg at all frames (HilbertTransform function in the R package hht[58]). The output values were then smoothed using a Savitzky-Golay filter (order = 3, filter length = 15) using the function sgolayfilt in the R package signal.[59]

#### *Gait analysis*

Flies were considered walking when body velocity was greater than 5 mm/s, and the maximum likelihood of each of the six tracked points was greater than 0.99. We identified a small number of rare cases in which flies were estimated to have 0 or 1 legs in stance

while moving at low velocities. Given this, all instances in which these stances were detected at body velocities less than 10mm/s were replaced with 6 leg stances. We then identified all continuous walking bouts longer than 333ms, yielding 1,123 bouts representing 216,729 time points.

The walking bout data set was then used to train Hidden Markov Models (HMMs) modelling the states underlying the number of legs in stance over time using the R package depmixS4.[60] We used the function depmix to train three models (2 hidden states, 3 hidden states, 4 hidden states) using the formula stance ~ 1. We then optimized the model parameters using expectation maximization via the fit function in depmixS4. The fits of the three models were compared using log likelihood AIC, and BIC (Table S1), revealing that a model with 3 hidden states fit the data best and paralleling previous observations in fruit flies.[24,26,27] Per-frame state designations were estimated from the posterior probabilities of the models. The distributions of HMM states in behavior space were calculated by identifying the xy coordinates in behavior space for all time points in which a given state occurred. These were then used to calculate probability density functions as function of behavior space position for each state (as in Figure 3D) using the function kde2d in the R package MASS (grid points = 200, bandwidth = 2).

## Phylogenetic analyses

### *Calculating structure, frequency, and transitions*

*Structure* was measured by identifying the unique bins that each species or individual (depending on the analysis) visited. A 64 × 64 binary matrix was generated and bins visited were filled with 1 while bins not visited were filled with 0.

*Frequency* was inferred by calculating probability density functions (pdfs) from xy position in behavior space. Here, greater density in a specific region would reflect a higher frequency of occurrence of the movement represented by that portion of behavior space. We calculated pdfs for all individuals using the function kde2d in the R package MASS (grid points = 100, bandwidth = 1). To enable comparison each pdf was normalized by dividing all values by the maximum. From these we calculated a mean pdf for each species by averaging each point across all individuals of a species, in addition to the standard error for each bin (used for downstream phylogenetic analyses).

*Transitions* were identified by first clustering points in behavior space based on their graph properties. Many clustering methods identify structures based on the density of points across some number of dimensions. However, the TREBLE behavior space contains information about both point density *and* the temporal sequencing between points. We therefore sought a clustering method that could capture both important aspects. Furthermore, to facilitate rapid comparisons across individuals and species, we prioritized methods that minimized assumptions and extensive model fitting. We opted to use Louvain clustering, a graph-based approach common in other nonlinear dimensionality reduction application such as scRNA-seq.[61] First, we created an undirected graph with two columns using the function graph_from_data_frame in the R package igraph in which the first column represented the xy coordinate *from* which a trajectory was leaving while the second was the xy coordinate the trajectory was going *to*. This resulted in a graph with 3,890,645 rows corresponding to a full set of feature windows and their corresponding 2-d movement patterns in behavior space. We then used the igraph function cluster_louvain to do the clustering. The procedure yielded 7 clusters that corresponded to recognizable features based on point density and movement through behavior space (Figures S2D–S2F). Furthermore, each cluster was associated with a characteristic joint distribution of kinematic parameters used (Figure S2E), reflecting distinct components of movement. Transition probabilities between Louvain clusters were inferred using Markov models. We created models for all flies using the function markovchainfit in the R package markovchain[62] using the Louvain cluster designations from cluster_louvain as input. The transition matrices for each fly were extracted and averaged per-species to create a mean transition matrix for each. The standard error of each transition was also calculated per-species for downstream phylogenetic analyses.

We calculated morphospaces for *structure*, *frequency*, and *transitions* using PCA (Figures 5A, 5D, and 5G). The input matrices for PCA were produced by linearizing and combining the species mean values for each trait. For example, the 64 × 64 *frequency* matrices were linearized into 4,096 element vectors and horizontally combined to create a 4,096 × 24 matrix that was then used to run a PCA.

### *Body size comparisons*

Body size variation across pure species (Figure S4A) was measured from the videos recorded during trials in the Coliseum. An ellipse was fit to best match landmarks on the fly corresponding to anterior, posterior, and lateral boundaries. Body size was inferred by calculating the area of the fitted ellipse for each trial (Area = $\pi$*minor axis*major axis). Though this method allowed for an assessment of the general distribution of body sizes, and their variance, it is not a well standardized measure. To facilitate comparisons of body size with behavioral traits we instead opted to use the standard measurements of thorax length.[63] Relationships between thorax length and % of behavior space covered, *frequency* PC1, and *transitions* PC1 were first measured using Pearson correlation (Figures S4B–S4D).

We then compared the explanatory power of size and species of origin using a series of generalized linear models for three traits related to the frequency and sequencing of behavior: time spent moving, *frequency* PC1, and *transitions* PC1. For each trait, we constructed four models. The first predicted the trait using just size (trait ~ size), the second using just species of origin (trait ~ species), the third using both size and species of origin (trait ~ size+species), and the fourth allowing for an interaction between size and species of origin (trait ~ size*species). Model fits were compared using Akaike information criterion (AIC) (Figures S4E–S4G).

To assess the effect of body size on the statistics of behavior, we normalized translational velocity by dividing it by thorax length ('body normalized translational velocity'). We then re-generated parameter windows using this measure and embedded the resulting windows into the TREBLE behavior space using the predict function in the R UMAP implementation. Two probability density

functions were generated for each species, one from windows created with translational velocity and the other for those incorporating body normalized translational velocity. These were then compared using a cophyloplot, as seen in Figure S4H.

### Mapping trait evolution

Phylogenetic signal was measured via Blomberg's *K* using the R package phytools.[64] For the traits *structure* and *frequency*, we calculated phylogenetic signal for all bins in behavior space, treating each as a unique phenotypic measurement, and using the species means and standard errors calculated above. We assessed the extent to which the number of bins chosen for these traits may affect the calculation of phylogenetic signal by calculating the phylogenetic signal distributions of a range of resolutions (2×2 to 100×100) for the trait *frequency*. Comparing mean phylogenetic signal revealed that the measure began to level-off around a resolution of 20×20 bins and become stable at a resolution of 30×30 bins (Figures S2G and S2H), demonstrating that a relatively broad range of possible resolutions could be used in comparing this trait across species. For *transitions* we calculated phylogenetic signal of each transition probability contained in the mean species-level transition matrices, also factoring in the standard error. This resulted in distributions of phylogenetic signal across all dimensions of the three traits (Figure 6A). Variation in these distributions was tested using a Kruskal-Wallis test followed by a post-hoc Dunn's test.

To compare the temporal patterns of trait evolution among *structure*, *frequency*, and *transitions* we computed variable rates models using the R implementation of the software BayesTraits (evolution.rdg.ac.uk, v.3), and its wrapper btw (github.com/rgriff23/btw). To facilitate multivariate comparisons, we used independent contrasts models for all three traits and applied Markov-Chain Monte Carlo chains run with 10,010,000 iterations and thinned every 1,000. For *frequency* and *transitions* we modelled all PCs that accounted for >90% of variance in the data (13 and 4, respectively). The first two PCs were used for *transitions*. Three models were created for each trait. We tested for convergence of all models using the Gelman-Rubin diagnostic in the R package coda (function gelman.diag; 3 chains),[65] requiring a value of 1 to proceed with analysis. We required models to have effective sizes of at least 200 and selected the best model for each trait via Bayes factor using btw. To compare the distributions of relative evolutionary rates, we extracted the rates reported at each node in the tree by BayesTraits and normalized them by the maximum value for each trait (as seen in Figure 6B). Variation in rate distributions was tested using a Kruskal-Wallis test followed by a post-hoc Dunn's test.

We performed an adapted version of the analyses in Cooney et al.[66] and Ronco et al.[67] to estimate morphospace filling over time. First, we estimated ancestral states for each trait using the fastAnc function in phytools using the mean rate-transformed tree from BayesTraits. We then calculated the values of each trait along the species tree in time intervals of 0.1 million years. At each interval, trait values for all extant branches were identified and then linearly predicted between nodes. To evaluate morphospace filling we performed this procedure for the first two PCs of each trait, as seen in Figure S5A. To estimate the percent of morphospace filled as a function of time (Figure S5B), we calculated the cumulative coverage of the xy coordinates of the 2-d morphospaces in 0.1 million year intervals.

The evolutionary rates of specific components of *frequency* and *transitions* were inferred by computing phylogenetically independent contrasts (PIC) with the R package ape (pic function).[68] For *frequency*, PIC was calculated for every bin in behavior space. For *transitions*, PIC was calculated for each transition probability between Louvain clusters.

### Trait meta-analysis

To compare the patterns observed here to other *Drosophila* traits, we conducted a review of the fruit fly evolutionary literature and identified studies that measured traits in at least 7 species present in our phylogenetic tree. This yielded measurements for 56 traits across 18 individual studies (see Table S2 for citations). In addition, we incorporated *frequency* (PC1, PC2), *transitions* (PC1, PC2), *structure* (% of behavior space covered), and the frequency of occurrence in the 7 Louvain clusters. Overall, the final data set contained 78 traits with 105 species represented at least once (Table S2; Data S1).

To facilitate comparison across traits, each measure was converted to a 0-1 scale by first adding the absolute value of the minimum and then dividing by the maximum value. We then inferred evolutionary rate ($\sigma^2$) using a single-rate Brownian motion model in phytools. Rates were then compared across trait types using a Kruskal-Wallis test followed by a post-hoc Dunn's test.

Estimates of evolutionary rate can be biased by sample size and the phylogenetic distance of the clade being compared. To test if such biases were present in our data set, we created a linear model predicting $\sigma^2$ from sample size and distance (inferred by the maximum branch length of the phylogenetic tree for a given subset of species; lm function in stats). While sample size and distance accounted for very little of the variation in $\sigma^2$ ($R^2 = 0.07$), and sample size did not significantly predict the outcome variable ($P = 0.82$), we did find that distance was marginally significantly predictive ($P = 0.02$). If this association between phylogenetic distance and evolutionary rate were to be unevenly distributed across trait categories, then the observed rate differences might have arisen from artifacts rather than real signal. To control for this potential limitation, we compared the residuals of the linear model (i.e. $\sigma^2$ with the effects of sample size and distance regressed out) using a Kruskal-Wallis test. We found that the five trait types still differed significantly ($P = 8.7 \times 10^{-5}$) (Figure S6A).