Title: Evolution and genetic basis of the plant-penetrating ovipositor, a key adaptation in the transition to herbivory within the Drosophilidae

Authors: Julianne N. Pelaez,[1*] Andrew D. Gloss,[2,3*] Julianne F. Ray,[3] Noah K. Whiteman[1§]

Affiliations:

[1]Department of Integrative Biology, University of California, Berkeley

[2]Department of Ecology and Evolution, University of Chicago

[3]Department of Ecology and Evolutionary Biology, University of Arizona

[*]These authors contributed equally

§ Corresponding authors: Whiteman, N.K. (whiteman@berkeley.edu); Pelaez, J.N. (juliannepelaez@gmail.com)

Abstract

Herbivorous insects are extraordinarily diverse, yet are found in only one-third of insect orders. This skew may be driven by barriers to plant colonization coupled with phylogenetic constraint on plant-colonizing adaptations. Physical barriers can be surmounted by key innovations like the plant-penetrating ovipositor. Within Drosophilidae, ovipositor margins densely adorned with hard bristles used to cut into plants evolved repeatedly, but their evolutionary, developmental and genomic basis has only been explored in *Drosophila suzukii*. Here, we addressed this gap using *Scaptomyza*, an herbivorous radiation nested in a detritivorous clade. First, we found that ovipositor bristle number increased markedly as herbivory evolved in *Scaptomyza*. We then dissected the genomic architecture of variation in ovipositor bristle number within *S. flava* using a pooled genome wide association study (pool-GWAS). Variation in ovipositor bristle number in *S. flava* was heritable and associated with single nucleotide polymorphisms (SNPs) within non-coding regions involved in neural development. Genotyping of individual flies replicated the association at a candidate SNP upstream of *Gai*, a neural development gene, and estimated that it contributes to an average gain of ~0.58 bristles/ovipositor in *S. flava*. Neural developmental genes thus underlie variation in this key morphological adaptation, possibly facilitating the evolution of this trait and the colonization of tough tissue of living plants.

Introduction

Herbivorous insects are among the most successful of animal radiations[1,2], representing one quarter of all animal species.[3] Yet, herbivores are found in only one-third of extant insect orders, [1] a macroevolutionary pattern suggesting that colonization of plants is difficult. Indeed, the evolution of herbivory requires multi-faceted adaptations, including: locating appropriate host plants, remaining attached to the host, resisting desiccation, and feeding on nutritionally unbalanced, chemically- and physically-defended plant tissues.[4–7] Despite the limited number of insect orders that have evolved herbivory, *within* some insect orders, it has evolved many times independently, [2] including at least 25 times within Diptera.[8] Identifying whether these clades possess specific precursor traits that underlie this repeated evolution may help resolve this paradox.[9,10]

The plant-penetrating ovipositor may be one such trait, having arisen with major diversification events in the true fruit flies (Tephritidae) and a leaf-mining fly clade (Agromyzidae), together compromising 7,500 species, all leafhoppers (Cicadellidae, ~20,000 species), and several other clades among sawflies (Tenthredinidae), katydids (Tettigoniidae), plant bugs (Miridae), and various others.[11,12] By enabling the insertion of eggs into plant tissue, a penetrating ovipositor overcomes the challenges of attachment to hosts, desiccation, and access to physically-defended plant tissues[6]. It also allows neonate larvae of small-sized leaf-miners to hatch directly into the leaf interior, providing protection from variation in the physical environment (i.e. UV damage, temperature shifts, evaporative potential, and dislodgement by wind and rain), and protection from pathogens, predators, and parasites.[13] Some insects with cutting ovipositors, like agromyzid flies, are able as adults to consume exudates from oviposition wounds, allowing

them to feed on plant tissue even in the absence of piercing/sucking or chewing mouthparts.[14–16]

Convergent morphological evolution within the Drosophilidae offers strong evidence that the plant-penetrating ovipositor is a key adaptation for an herbivorous lifestyle. Although most drosophilid fly species feed on decaying plant tissues, plant-penetrating ovipositors have appeared across at least three drosophilid lineages that evolved herbivory independently: (1) *D. suzukii*, a generalist pest of ripe fruit, in the *D. melanogaster* species group,[14] (2) several leaf-mining species within the subgenus *Scaptomyza* (phylogenetically nested within *Drosophila*), which includes the model herbivore *Scaptomyza flava*, a specialized pest of Brassicaceae crops,[17,18] and (3) *Scaptodrosophila notha*, a fern-mining specialist whose ovipositor cuts into the rachis of living bracken fern (*Pteridium* spp.).[19] All three lineages independently evolved sclerotized ovipositors, studded with sharp, enlarged marginal bristles used by females to pierce or scrape wounds into living plant tissue (Fig. 1).[14,20,21] Drosophilid flies are models for studying the evolution of host plant specialization[22] and herbivory,[14,21,23,24] given the many high-quality reference genome assemblies across the genus and functional data from yeast-feeding *D. melanogaster* that allow for identification of loci involved in adaptations to plant feeding. The ovipositors of herbivorous drosophilids thus offer the opportunity to study the evolution and genetic architecture of this derived morphological trait.

Here, we focus on the evolution of the ovipositor in *S. flava*. In addition to the described morphological changes to the ovipositor, *S. flava* has acquired a unique and stereotyped behavioral repertoire during and after leaf puncturing. Females scoop a hole into the leaf by repeatedly opening the two ovipositor valves laterally (Supplemental Videos S1, S2). Following this, they pause briefly, turn counter clockwise, and use their proboscis to feed on leaf exudates in the wound.

Females create hundreds of feeding punctures per day and oviposit eggs into a small percentage of wounds.[21,25] Hatched larvae immediately begin feeding on mesophyll tissue adjacent to the puncture, mining the leaves through three instar stages.[21,26] The ovipositor of *S. flava* is thus a critical adaptation for egg-laying, and for enabling food acquisition for both the leaf-mining larvae and adult females.

Although the ovipositors of herbivorous drosophilids differ noticeably in many aspects of their shape and size, the ovipositors of almost all herbivorous drosophilids (exceptions include *Scaptomyza flavella*[27] and *Scaptodrosophila megagenys*),[19] share supernumerary bristles, arranged in a single row along the ventral margin of each ovipositor valve (Fig. 1). We therefore focused on bristle number, which, at least in other regions of the fly body, have been well-studied in quantitative genetics and developmental biology.[28–33]

To test if marginal ovipositor bristle number was associated with the evolution of herbivory in *Scaptomyza*, we first investigated whether ancestral increases in the number of ovipositor bristles paralleled the transition to herbivory in *Scaptomyza*. To do this, we traced the evolution of ovipositor bristle number in the paraphyletic subgenus *Drosophila* (in which *Scaptomyza* is nested) using phylogenetic generalized least squares (PGLS) methods and ancestral state reconstruction (ASR). We then reasoned that because gene regulatory evolution is likely the main driver of bristle number variation across species, a quantitative genetic approach within an herbivorous species could be used to identify genetic variants associated with loci controlling ovipositor bristle number. Accordingly, we quantified the additive genetic variance associated with ovipositor bristle number using controlled crosses, and found heritable variation. We employed pool-genome wide association mapping (pool-GWA) to locate genetic polymorphisms associated with variation in ovipositor bristle number. Our GWA was not designed to detect fixed mutations that account for

inter-species differences, but rather to identify genetic variants and candidate developmental processes that may contribute to increased ovipositor bristle number in *S. flava.* Finally, we sought to validate the results of our pooled-GWA study by genotyping individual flies to estimate the effect size of a single nucleotide polymorphism (SNP) that was linked to the gene *Gαi*, a promising candidate involved in bristle development.[34]

Materials and Methods

(a) Phylogeny reconstruction.

We estimated the phylogeny of extant *Scaptomyza* by expanding on previous datasets,[35,36] to include an additional five herbivorous taxa: two with already sequenced genetic markers, *S.* nr. *nigrita* (Nevada) and *S.* nr. *flava* (Arizona),[37] and three others obtained in this study from California, *S.* nr. *nigrita*, *S. flava* ssp. *montana* Hackman 1959, and *S. flaviventris* Hackman 1959. This amounted to a total of 95 taxa, spanning subgenera *Scaptomyza* and *Drosophila*, including the Hawaiian *Drosophila*. The character matrix of DNA sequence data included 11 genes: 16S, 28S, *Adh*, *Cad-r* (rudimentary), *COI*, *COII*, *gstd1*, *gpdh*, *marf*, *ND2*, *n(l)tid)*, *Orco*. To generate DNA sequence data for the additional taxa, DNA extraction and PCR methods (including primers) followed those described previously.[35] PCR amplicons were cleaned and Sanger sequenced in both directions at UC-Berkeley, trimmed in Geneious v.10.0.5, and manually aligned to the other 90 taxa.

A species tree was estimated by both maximum likelihood (ML) in RAxML,[38] and by Bayesian inference using MrBayes v.3.2.4[39] and BEAST v.2.4.6.[40] The Bayesian methods also simultaneously allowed us to infer divergence times to generate a time-calibrated phylogeny.

The concatenated alignment was partitioned by codon and gene, with ribosomal genes given a single partition. The ML analysis used a GTR+gamma model on all partitions. To

determine whether the results were consistent from one run to the next, five independent runs were performed, each with a distinct starting seed. Each run included 1000 bootstrap replicates, and a slow ML search on every 5th tree. The phylogeny with the highest likelihood was used for ancestral character estimation. For the Bayesian analysis, models of sequence evolution were selected for each partition with the Akaike information criterion (AIC), using MrModeltest2 v.2.3[41] and PAUP* v.4.0a.[42] To infer a phylogeny and divergence times, a Markov-Chain Monte-Carlo (MCMC) analysis was performed as previously described.[36] To ensure that the Markov chain adequately converged to a stationary distribution, forty replicate runs were performed and implemented in BEAST v.2.4.6[40] with BEAGLE[43] for multicore processing. The first ten percent of samples were discarded as burn-in. The trees were re-sampled every 250,000 generations, and combined using LogCombiner v.2.4.6[40]. Tracer v.1.6.0[44] was used to confirm that ESS values were sufficient for reliable parameter estimates (ESS >200).

(b) Ovipositor trait evolution across herbivorous and non-herbivorous drosophilids.

To test if ovipositor bristle number increased in association with the evolution of herbivory, we first collected ovipositor bristle counts from as many species from our phylogeny for which we could obtain phenotypic data. Because drosophilids exhibit variation in ovipositor bristle morphology, ranging from hardened pegs to flexible structures, distinguishing between bristle types was not straightforward. We therefore counted all bristles, regardless of their morphology and size, on one of the oviscapts (also referred to as ovipositor valves or plates). Bristle counts were obtained either from illustrations or images from the literature, or from wild-caught or lab-reared individuals (Table S1 for source data and measurements). Data from the literature may be undercounted as a result of obscured angles, but, where possible, counts from multiple literature sources were averaged for each species. Whole ovipositors of wild-caught or lab-reared female

flies (n ≤ 10 individuals per species) were dissected and mounted onto slides with Permount mounting media (Fisher Scientific) and coverslips. Ovipositors were photographed under visible light for phenotypic measurements using an EOS Rebel T3i camera (Canon) mounted on a Stemi 508 stereomicroscope (Zeiss) with a 1000 μm scale bar for calibration. Representative illustrations of single oviscapts for 26 species are shown in Fig. S1. Information on larval diet for 50 species was obtained from a previous study[35] (Table S1).

Phylogenetic generalized least-squares regression (PGLS)[45] was used to test for a relationship between the number of ovipositor bristles and larval diet (i.e. oviposition sites), accounting for phylogenetic relatedness on the covariance of these two traits. The PGLS regression between ovipositor bristle number and larval diet was performed using the APE[46] and geiger[47] packages in R, using species for which there was both data on ovipositor bristle number and larval diet (n = 23). Models of trait evolution (Brownian motion (BM) = random drift, Ornstein–Uhlenbeck (OU) = random drift tending towards a single optimum, Early Burst (EB) = rate of evolution exponentially slows through time, and white noise (white) = non-phylogenetic) were separately modeled, and compared using AIC corrected for small sample sizes, as implemented in the R package geiger v. 2.0.6. The Brownian motion model was selected as the best-fit model for ovipositor bristle number. The degree of phylogenetic signal in the residuals was estimated using Pagel's lambda (λ), with the signal ranging from phylogenetic independence (Pagel's λ=0) to phylogenetic dependence (Pagel's λ=1).

To visualize the correlation between the evolution of herbivory and increases in ovipositor bristle number, ancestral state reconstructions (ASR) of ovipositor bristle number and larval diet were conducted and mapped onto each node of the phylogeny. For ovipositor bristle number, we used the anc.ml function in the phytools R package,[48] using ML and the Brownian motion

model. ML ASR on larval diet was performed using the ace function in APE[46] under an equal rates model of evolution (after using AIC to compare the following different models: equal rates, symmetric, all rates different.)

To investigate the possible role ovipositor length or body size may play in ovipositor bristle number, we analyze data on ovipositor bristle number, ovipositor length, and female thorax length, as a proxy for body size, across 67 drosophilid species for which this data was previously reported[49] and 4 species from this study (Table S2). Ovipositor bristle counts were averaged across ovipositor valves and individuals within a species. If only a subset of bristles were counted (i.e. ventral bristles only), these counts were excluded to maintain consistency in counting methods across species.

(c) Mapping population and ovipositor measurements.

To identify specific genetic polymorphisms contributing to variation in bristle number, we used a pool-GWAS approach, which relies on detecting allele frequency differences among pools of individuals with extreme phenotypes from the same population. Two *S. flava* outbred laboratory populations were founded from larvae collected on one or two mustard (Brassicaceae) host plant species near Dover, New Hampshire, USA: 79 larvae from *Turritis glabra* (termed the "NH1" colony) and 58 from mixture of *T. glabra* and *Barbarea vulgaris* (termed "NH2"). To minimize chances of sampling related individuals, only a single larva was typically collected per plant. After pupation in detached leaves, adults were subsequently transferred to one mesh cage per population containing *Arabidopsis thaliana* Col-0 plants, yielding over 200 F1 generation offspring per population. F1 flies were subsequently reared on a mixture of *T. glabra* and *B. vulgaris*, and the adult female F2 offspring were preserved in 95% ethanol and phenotyped for the subsequent GWAS. Ovipositor dissections were performed as described above. Bristles were counted along

the ventral edge of the ovipositor from the smallest bristle at the anterior to the longest bristle at the posterior apex (shown in Fig. 3a), while those along the blunt apical edge of the ovipositor were excluded because their number varied by only one in all *S. flava* specimens studied. We also quantified ovipositor and wing chord length (a proxy for body size) for each specimen using ImageJ. Wing chord was measured from the base to the wing apex following the third longitudinal vein (Fig. 3a). All count and length phenotypes were averaged across two independent measurements per specimen.

Linear regression analyses in a pilot experiment (N = 100, NH1 and NH2 flies) revealed that ovipositor length was positively correlated with ovipositor bristle number ($R2 = 0.13$; $P < 0.001$), but not wing length ($P > 0.1$). We therefore quantified both ovipositor bristle number and ovipositor length for all individuals (NH1, N = 308 flies; NH2, N = 422 flies).

(d) Narrow-sense heritability estimates from mother-daughter regression.

Fifty single male-female pairs (virgin females) from the combined NH1 and NH2 colonies were individually mated on *T. glabra* in Magenta boxes (Sigma-Aldrich) with modified mesh tops. Each box was provisioned with cotton balls soaked in 10% honey solution. For 35 mate pairs that produced daughters, ovipositor length and bristle number were profiled for every mother and at least one of her daughters. Narrow-sense heritability – the proportion of phenotypic variation due to additive genetic effects – of ovipositor bristle number was calculated by regressing the phenotype of each mother against the average phenotype of her daughters and doubling the slope, since ovipositor traits could only be measured in mothers.[50]

(e) Pooled genome sequencing.

Flies in the NH1 and NH2 populations were split into two phenotypically extreme pools for each population, composed of 60-85 females in either the upper or lower 20% tail of the

distribution of residual bristle number (i.e., number of bristles relative to the number expected based on ovipositor length). Residual bristle number was determined by regressing ovipositor bristle number against ovipositor length, and residuals were extracted from the model using the *lm* function in R. Flies were homogenized using a TissueLyser (Qiagen) and a stainless-steel bead, and genomic DNA was extracted from each pool using a DNeasy Blood and Tissue Kit (Qiagen). A single Illumina library per pool was constructed with 100 bp paired-end reads and a 350 bp insert size, and each library was sequenced on one half lane on an Illumina HiSeq 2500 at Arizona State University.

(f) Read mapping and pool-GWA.

Reads were trimmed using Trimmomatic v0.32 with the parameters "TRAILING:3 HEADCROP:2 SLIDINGWINDOW:6:15 MINLEN:50" and mapped to an *S. flava* reference genome assembly (GenBank accession no. GCA_003952975.1) using bwa v0.7.12 [51–53] with the following parameters: for bwa aln, "-o 3 –d 15 -l 100"; for bwa sampe, "-a 1000". PCR and optical duplicate reads were removed using Picard Tools v1.107 (http://picard.sourceforge.net). Unpaired and low-quality reads were removed using the View command in Samtools v1.3.1[54] with parameters "-q 20 –f 0x0002 –F 0x0004 –F 0x0008". Low quality bases were removed using the Samtools mpileup command with parameters "-B –Q 17". Repeat regions and 5 bp windows flanking indels (minimum count > 4) were filtered using Popoolation2[55]. Statistical significance of allele frequency differences per site were estimated using the Cochran-Mantel-Haenszel test using Popoolation2 and custom scripts for sites with a minimum minor allele count of 10, coverage depth of 100, and minor allele frequency of 8% across all pools combined, and a minimum coverage depth of 15 and maximum coverage depth 200 in each pool. To correct *P*-value inflation, we regressed observed against expected –log10(*P*) values (excluding the 1% most significant

SNPs) and divided each –log10(*P*) value by the slope of the regression line. This approach was conservative, yielding a median *P*-value of 0.597. Reads were uploaded to the SRA on NCBI (SRR11252387-SRR11252390).

To enable tests for enrichment of predicted functions for genes harboring the top GWAS associations, we first assigned orthology relationships among protein-coding genes in *S. flava* and other *Drosophila* genomes (FlyBase release FB2013_06) using similarity clustering in orthoMCL, with default parameters and an inflation value of 1.5. Each *S. flava* gene was then annotated with the gene ontology (GO) terms assigned in FlyBase (release 2020_02) to all its predicted orthologs in *D. melanogaster*. Tests for the enrichment of gene ontology (GO) terms among the top 0.05% of SNP associations were implemented using GOWINDA [56], with significance determined from one million permutations. SNPs were assigned to a given gene if they fell within its exons, intronic regions, or were within 2 kb of the outer boundaries of its UTRs. The same methods were used to test if the top 0.05% and 0.001% of SNPs were enriched for orthologs of transcription factors that are differentially regulated among neural cell lineages in *D. melanogaster [57]*. To avoid redundancy in the GO analysis, terms were only considered if they were assigned to at least 25 genes in *S. flava* and were located in levels 4-5 (i.e., positioned 4 to 5 edges from the root of the acyclic graph of GO term relationships) of Developmental Process terms or levels 6-8 of Molecular Function terms.

(g) Validating effect size of a candidate SNP.

GWAS can be confounded by many factors including assortative mating, population structure and sampling artifacts. To validate the *S. flava* pool-GWAS results, we individually genotyped one of the top candidate SNPs associated with ovipositor bristle number and estimated its effect size. For each individual genotyped, ovipositor bristle number and length were measured

as described above. The SNP, located ~48bp upstream of the start codon of the gene *G alpha i subunit* (*Gαi*), was chosen because (1) the SNP was among the top highest ranking SNPs associated with ovipositor bristle number (Fig. 3g), and (2) *Gαi* is involved in neural development, and specifically in asymmetric cell division of neuroblasts, including sensory organ precursor (SOP) cells.[58] Genomic DNA was extracted from 74 individual females from the lab population (derived from NH1 and NH2), following the methods described above. A target region of 500bp around the SNP was Sanger-sequenced in both directions using the primers: gai4F (5'-CATTTCTGTCCATGGCGTCG-3'), and gai9R (5'-GCCGTTAGACAAAGCGCATT-3'), although the reverse strand did not result in usable sequences.

Bristle number and ovipositor length were modeled separately in a generalized linear model, assuming an additive effect of the major allele, using the lm function in R v.3.2.2. The model accounted for host plant species on which flies were reared in the lab, collection locality where flies were originally collected from in the wild, ovipositor length, and whether flies were taken from the high or low ovipositor bristle number pools (described in pool-GWA methods). The estimated effect size was given as β, the shift in bristle number (in standard deviations) expected by a single allelic substitution at the locus. Pairwise measures of linkage disequilibrium (LD) between the candidate SNP and other variant sites (minimum variant frequency >0.05) were calculated using the *LD2* function from the R package pegas, which assumes unphased data.

Results

(a) An increase in ovipositor bristle number coincided with the evolution of herbivory.

PGLS revealed that ovipositor bristle number was affected strongly by both phylogeny (Pagel's $\lambda = 1$) and larval diet (herbivorous or non-herbivorous) ($F_{1,1}=4.33$, $P = 0.05$). The ASRs of ovipositor bristle number and larval diet similarly suggests that there was a dramatic increase

in ovipositor bristle number that coincided with the evolution of herbivory within *Scaptomyza* <13 million years ago (Fig. 2a; Fig. S3 for confidence intervals of bristle number ASR estimates). Although ovipositor bristle number varies within species, this variation is small relative to the large differences observed between species (Fig. 2b). Ovipositor length and body size, which may covary with ovipositor bristle number, could not be taken into account in the phylogenetic analyses because of the lack of data for these species. However, a linear regression of ovipositor bristle number regressed onto ovipositor length, using published data [49] and additional data from this study, indicates that ovipositor bristle number is strongly predicted by ovipositor length in drosophilids ($\beta = 0.53$, $t(53) = 4.3$, $P < 0.001$, adjusted $R^2 = 0.47$; Fig. 2c). Notably, the herbivorous *S. flava* has proportionally more ovipositor bristles relative to ovipositor length than all other species, illustrated by having the highest residual value (Fig. S4).

(b) GWAS on ventral ovipositor bristle number.

Variation in ovipositor bristle number was normally distributed in the NH1 and NH2 outbred laboratory populations of *S. flava* (Fig. 3b), typical of a quantitative trait controlled by multiple loci. Mother-daughter regression revealed that additive genetic differences among individuals accounted for roughly half of this variation ($h^2 = 0.46$; Fig. 3c).

We sought to interrogate the genetic basis for this variation using pool-GWAS. Individuals from two outbred colonies of *S. flava* were separated into four pools with extremely high or low ovipositor bristle numbers relative to the numbers expected given their ovipositor lengths. Thus, although ovipositor length significantly influences ovipositor peg number (Fig. 3d), our pools differed in ovipositor bristle number but not ovipositor length (Fig. 3e).

Whole genome re-sequencing of the four DNA pools (high and low phenotypic extremes from the NH1 and NH2 outbred colonies) resulted in a mean experiment-wide coverage depth per

site of 168X. After mapping sequences to the *S. flava* genome, we quantified the probability that allele frequencies differed between the high and low bristle number pools for 1.7 million SNPs. We found an excess of SNPs with significantly differentiated allele frequencies among pools (Fig. 3f); for example, the false discovery rate when considering the 15 most significant SNPs (i.e., the top 0.001 percentile genome-wide) is 21%, suggesting most are true associations. Because LD decays quickly in *S. flava* [18], the SNPs showing the strongest associations are likely in close proximity to causal polymorphisms or are causal themselves.

(c) Gene ontology enrichment analysis on candidate SNPs.

To gain insight into the developmental basis of variation in ovipositor bristle number, we first tested if SNPs with the strongest GWAS associations (top 0.05% of p-values) were enriched in genes annotated with gene ontology (GO) terms classified as developmental processes. Two terms, regulation of neuron maturation (GO:0014041) and regulation of neuron remodeling (GO:1904799), were significantly enriched (Table 1). Second, to gain insight into the functional basis of this variation, we performed an analogous test for GO terms classified as molecular functions. A single term was significantly enriched: DNA-binding transcription repressor activity, RNA polymerase II-specific (GO:0001227).

Given that the GO enrichment analyses implicated gene regulatory functions involved in neural development, we also tested for enrichment of transcription factors whose orthologs are putatively involved in specifying neural cell fate in *D. melanogaster* [57]. No significant enrichment was detected in the top 0.05% of SNPs in our GWAS. However, two of the top 15 SNPs were located upstream of these transcription factors, *sloppy paired 2 (slp2)* and *Ecdysone-induced protein 75B* (*Eip75B*) (Fig. 3g), reflecting a significant enrichment of neural fate-specifying transcription factors among the strongest GWAS associations (Table 1).

(d) Validation of a top candidate SNP from pool-GWAS.

We genotyped and phenotyped individual flies to validate results from the pool-GWAS by estimating the effect size of the major allele of one SNP identified as a potential causal variant of ovipositor bristle number. This SNP resides in the 5' UTR of *G protein α i subunit* (*Gαi*), a gene that functions in asymmetric cell division that gives rise to sensory organs [58]. Not only did the sign and magnitude of the effect of this allele recapitulate the results from the GWAS, the effect was significant ($\beta = 0.11$ standard deviations, $t(68) = 2.88$, $P < 0.005$), the equivalent of an increase in 0.58 bristles per major allele carried. The proportion of variance explained by this SNP out of the total variance in bristle number was 9.53%, using partial adjusted r^2 values, a sizable proportion compared to that given by the full model, 91.34%. Although ovipositor length is correlated with ovipositor bristle number in natural populations, it did not have an effect in this analysis because of the sampling methods used to correct for ovipositor length within both extreme-phenotype pools. Accordingly, the SNP did not have an effect on ovipositor length ($\beta = 0.02$ standard deviations, $t(69) = 0.177$, $P > 0.05$). Out of 39 variant sites found in the sequenced region, we identified 5 sites with a minimum variant frequency of 0.05, and all had significant *P*-values for LD with the candidate SNP. All were located upstream of the coding sequence of *Gαi* or in an intronic region. Further sequencing will be necessary to identify the causal allele(s) in this region.

Discussion

Ovipositor bristle number presents an excellent lens for studying the evolution and genetic architecture of a complex trait because of its clear adaptive role in feeding and egg-laying among herbivorous drosophilids, and because of the wealth of information regarding the genetics and development of bristles in *D. melanogaster*. Our ancestral state reconstruction of ovipositor bristle number revealed that bristle number increased around the timing of the evolution of herbivory

within *Scaptomyza*, <13 million years ago (Fig. 2a), with extant herbivorous species exhibiting among the highest counts, even after accounting for ovipositor length (Fig. 2c, Fig. S4). Throughout the *Scaptomyza* and *Drosophila* subgenera, ovipositor bristle number is estimated to have increased and decreased repeatedly, suggesting ancestral variation in this trait could have been rapidly selected upon to transform this egg-laying organ into a multipurpose cutting tool and trophic organ.

Although it would be difficult to pinpoint the genetic changes that gave rise to this increase in ovipositor bristle number coinciding with the evolution of herbivory millions of years ago, we sought to leverage GWAS to illuminate genes and gene functions that confer phenotypic variation on which selection for bristle number can act. Transcription factors and other genes that regulate neural development were significantly enriched near the strongest GWAS associations. These SNPs may affect the abundance, location, and/or timing of expression of transcription factors that initiate developmental cascades in neural cell fate specification and in turn influence sensory bristle number. This is consistent with our understanding that insect bristles are developmentally derived from single neural precursor cells (sensory organ precursors or SOP), which differentiate through asymmetric cell divisions to generate multidendritic neurons (to detect pain, temperature and other submodalities), external sense neurons innervating the bristles (for mechanosensation and chemosensation), and cells that produce the shaft, socket and sheath of the bristle [59]. In other *Drosophila* species, genes involved in neural development have similarly been found to underlie interspecific differences in the number of bristles on male genitalia and sexcombs of the forelegs[60], and differences in thoracic bristle number [61,62].

We attempted to limit our investigation to the genetic basis of variation in ovipositor bristle number, rather than ovipositor length. While we accounted for length where possible, we found

that both within *S. flava* and across drosophilid species these two traits are nonetheless positively correlated (Fig. 2c, 3d). This suggests the likely possibility that an increase in ovipositor length and/or body size may have also contributed in part to an increase in bristle number. Recent investigations on the ripe fruit-cutting ovipositor of *D. suzukii* show ovipositor length is driven by increased cell size and cellular reorganization during ovipositor development, which increases length but not width.[63] Indeed, there were many top SNPs in our GWAS located near genes involved in cell elongation, which could have a pleiotropic effect on both increased ovipositor length and bristle number. Future functional work will be needed to address the role of these candidate SNPs in the evolution of multiple ovipositor traits.

Many studies aiming to understand the genetic architecture of adaptive traits have focused on monogenic, Mendelian traits with large effect sizes of candidate loci[64–66] with lower detection thresholds than genetically complex traits. Ovipositor bristle number represents an ideal quantitative trait because it is a discrete, countable trait, highly variable and heritable, and yet heritable variation is underpinned by many loci each of small effect. Despite having a genetic basis similar to many quantitative traits (small effect SNPs underlying variation in the trait) we found a moderately large effect SNP could be detected and validated with individual sequencing. These results suggest that pool-GWAS can be a viable method for pinpointing narrow regions of the genome that underlie variation in quantitative traits. Allele-specific expression analysis, functional experimentation with alternative alleles, and validation of more candidate SNPs are necessary to understand how these polymorphisms, either singularly or in concert with one another, influence development of ovipositor bristle number at the molecular level. The developmental pathways, genes, and regulatory regions identified through our GWAS mapping approach are nonetheless promising candidates for functional and evolutionary studies aiming to identify the genetic origin

of dense ovipositor bristles in *S. flava* and other species in which this trait has evolved in parallel, such as *D. suzukii*, and *S. notha*, given the conservation of these pathways, providing a substrate that could have been repeatedly targeted.

Data accessibility

All data files and R scripts are in the process of being deposited in the Dryad Repository, and are available upon request. Sequence data are available at NCBI sequence read archive: SRR11252387, SRR11252388, SRR11252389, SRR11252390.

Authors' contributions

J.N.P., A.D.G., and N.K.W. conceived the project and designed the research. J.NP., A.D.G., and J.F.R. collected the data. J.N.P and A.D.G. analyzed the data. A.D.G., J.N.P. and N.K.W. contributed to manuscript preparation. J.N.P. and N.K.W. wrote the first version of the paper, and A.D.G. and J.F.R. contributed to the final version.

Competing interests

We declare we have no competing interests.

Funding

This work was supported by the National Institute of General Medical Sciences of the National Institute of Health (award number R35GM11981601) to NKW; the National Science Foundation Graduate Research Fellowship (grant number DGE 1752814) to JNP and a UC Berkeley Mentored Research Award to JNP.

Acknowledgements

We thank Kirsten Verster and Esteban Rodas from the Whiteman Lab for their contributions. We also thank Craig Miller, Kristin Scott, and Michael Nachman for comments on previous versions of the manuscript.

References

1. Mitter C, Farrell B, Wiegmann B. 1988 The Phylogenetic Study of Adaptive Zones: Has Phytophagy Promoted Insect Diversification? *Am. Nat.* **132**, 107–128.

2. Wiens JJ, Lapoint RT, Whiteman NK. 2015 Herbivory increases diversification across insect clades. *Nat. Commun.* **6**, 8370.

3. Schoonhoven LM, Van Loon B, van Loon JJA, Dicke M. 2005 *Insect-Plant Biology*. OUP Oxford.

4. Bernays E, Graham M. 1988 On the Evolution of Host Specificity in Phytophagous Arthropods. *Ecology* **69**, 886–892.

5. Bernays EA, Jarzembowski EA, Malcolm SB, Chaloner WG, Harper JL, Lawton JH. 1991 Evolution of insect morphology in relation to plants. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **333**, 257–264.

6. Southwood TRE. 1972 insect/plant relationship--an evolutionary perspective. *Roy Entomol Soc London Symp* **1972, 6**.

7. Strong DR, Lawton JH, Southwood SR, Others. 1984 *Insects on plants. Community patterns and mechanisms*. Blackwell Scientific Publicatons.

8. Wiegmann BM *et al.* 2011 Episodic radiations in the fly tree of life. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 5690–5695.

9. Marazzi B, Ané C, Simon MF, Delgado-Salinas A, Luckow M, Sanderson MJ. 2012 Locating evolutionary precursors on a phylogenetic tree. *Evolution* **66**, 3918–3930.

10. Simpson GG. 1953 The major features of evolution.

11. Dıaz-Fleischer F, Papaj DR, Prokopy RJ, Norrbom AL, Aluja M. 2000 Evolution of fruit ßy oviposition behavior, pp. 811Ð841. *Fruit ßies (Tephritidae): phylogeny and evolution of behavior. CRC, Boca Raton, FL*

12. Dietrich CH, Others. 2013 Overview of the phylogeny, taxonomy and diversity of the leafhopper (Hemiptera: Auchenorrhyncha: Cicadomorpha: Membracoidea: Cicadellidae) vectors of plant pathogens. In *Proceedings of the 2013 International Symposium on Insect Vectors and Insect-Borne Diseases*, pp. 47–69. 台中市: 農業試驗所.

13. Connor EF, Taverner MP. 1997 The Evolution and Adaptive Significance of the Leaf-Mining Habit. *Oikos* **79**, 6–25.

14. Atallah J, Teixeira L, Salazar R, Zaragoza G, Kopp A. 2014 The making of a pest: the evolution of a fruit-penetrating ovipositor in Drosophila suzukii and related species. *Proc. Biol. Sci.* **281**, 20132840.

15. Bendall EE, Vertacnik KL, Linnen CR. 2017 Oviposition traits generate extrinsic postzygotic isolation between two pine sawfly species. *BMC Evol. Biol.* **17**, 26.

16. Ghara M, Kundanati L, Borges RM. 2011 Nature's Swiss Army knives: ovipositor structure mirrors ecology in a multitrophic fig wasp community. *PLoS One* **6**, e23642.

17. Hackman W. 1959 On the genus Scaptomyza Hardy (Dipt., Drosophilidae). *With descriptions of new species from various parts of*

18. Gloss AD *et al.* 2019 Evolution of herbivory remodels a Drosophila genome. *bioRxiv*. , 767160. (doi:10.1101/767160)

19. Thomson JA, Jackson MJ, Bock IR. 1982 CONTRASTING RESOURCE UTILISATION IN TWO AUSTRALIAN SPECIES OF DROSOPHILA FALLEN (DIPTERA) FEEDING ON THE BRACKEN FERN PTERIDIUM SCOPOLI. *Aust. J. Entomol.* **21**, 29–30.

20. Aluja M, Norrbom A. 1999 *Fruit Flies (Tephritidae): Phylogeny and Evolution of Behavior*. CRC Press.

21. Whiteman NK *et al.* 2011 Mining the plant--herbivore interface with a leafmining Drosophila of Arabidopsis. *Mol. Ecol.* **20**, 995–1014.

22. Yassin A, Debat V, Bastide H, Gidaszewski N, David JR, Pool JE. 2016 Recurrent specialization on a toxic fruit in an island Drosophila population. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 4771–4776.

23. Goldman-Huertas B, Mitchell RF, Lapoint RT, Faucher CP, Hildebrand JG, Whiteman NK. 2015 Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3026–3031.

24. Karageorgi M, Bräcker LB, Lebreton S, Minervino C, Cavey M, Siju KP, Grunwald Kadow IC, Gompel N, Prud'homme B. 2017 Evolution of Multiple Sensory Systems Drives Novel Egg-Laying Behavior in the Fruit Pest Drosophila suzukii. *Curr. Biol.* **27**, 847–853.

25. Shakeel M, He XZ, Martin NA, Hanan A, Wang Q. 2009 Diurnal periodicity of adult eclosion mating and oviposition of the european leafminer Scaptomyza flava (Falln)(Diptera Drosophilidae). *N. Z. Plant Prot.* **62**, 80–85.

26. Whiteman NK *et al.* 2012 Genes involved in the evolution of herbivory by a leaf-mining, Drosophilid fly. *Genome Biol. Evol.* **4**, 900–916.

27. Martin N. 2014 Scaptomyza (Bunostoma) flavella (Diptera: Drosophilidae) and the evolution of leaf mining. *1* **47**, 8–11.

28. Dilda CL, Mackay TFC. 2002 The genetic architecture of Drosophila sensory bristle number. *Genetics* **162**, 1655–1674.

29. Gurganus MC, Nuzhdin SV, Leips JW, Mackay TF. 1999 High-resolution mapping of quantitative trait loci for sternopleural bristle number in Drosophila melanogaster. *Genetics* **152**, 1585–1604.

30. Lai C, Lyman RF, Long AD, Langley CH, Mackay TF. 1994 Naturally occurring variation in bristle number and DNA polymorphisms at the scabrous locus of Drosophila melanogaster. *Science* **266**, 1697–1702.

31. Long AD, Lyman RF, Langley CH, Mackay TF. 1998 Two sites in the Delta gene region contribute to naturally occurring variation in bristle number in Drosophila melanogaster. *Genetics* **149**, 999–1017.

32. Long AD, Mullaney SL, Reid LA, Fry JD, Langley CH, Mackay TF. 1995 High resolution mapping of genetic factors affecting abdominal bristle number in Drosophila melanogaster. *Genetics* **139**,

1273–1291.

33. Norga KK *et al.* 2003 Quantitative analysis of bristle number in Drosophila mutants identifies genes involved in neural development. *Curr. Biol.* **13**, 1388–1396.

34. Kopein D, Katanaev VL. 2009 Drosophila GoLoco-Protein Pins Is a Target of Gαo-mediated G Protein–coupled Receptor Signaling. *MBoC* **20**, 3865–3877.

35. Lapoint RT, O'Grady PM, Whiteman NK. 2013 Diversification and dispersal of the Hawaiian Drosophilidae: the evolution of Scaptomyza. *Mol. Phylogenet. Evol.* **69**, 95–108.

36. Katoh T, Izumitani HF, Yamashita S, Watada M. 2017 Multiple origins of Hawaiian drosophilids: phylogeography of Scaptomyza hardy (Diptera: Drosophilidae). *Entomol. Sci.* **20**, 33–44.

37. Gloss AD *et al.* 2014 Evolution in an ancient detoxification pathway is coupled with a transition to herbivory in the drosophilidae. *Mol. Biol. Evol.* **31**, 2441–2456.

38. Stamatakis A. 2006 RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690.

39. Ronquist F, Huelsenbeck JP. 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

40. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014 BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537.

41. Nylander JAA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey JL. 2004 Bayesian phylogenetic analysis of combined data. *Syst. Biol.* **53**, 47–67.

42. Swofford DL, Sullivan J. 2003 Phylogeny inference based on parsimony and other methods using PAUP*. *The phylogenetic handbook: a practical approach to DNA and protein phylogeny* **7**, 160–206.

43. Ayres DL *et al.* 2012 BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. *Syst. Biol.* **61**, 170–173.

44. Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014 Tracer v1. 6. Computer program and documentation distributed by the author.

45. Butler MA, King AA. 2004 Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution. *Am. Nat.* **164**, 683–695.

46. Paradis E, Claude J, Strimmer K. 2004 APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*. **20**, 289–290. (doi:10.1093/bioinformatics/btg412)

47. Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. 2008 GEIGER: investigating evolutionary radiations. *Bioinformatics* **24**, 129–131.

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

49. Craddock EM, Kambysellis MP, Franchi L, Francisco P, Grey M, Hutchinson A, Nanhoo S, Antar S.

2018 Ultrastructural variation and adaptive evolution of the ovipositor in the endemic Hawaiian Drosophilidae. *J. Morphol.* **279**, 1725–1752.

50. Lynch M, Walsh B, Others. 1998 *Genetics and analysis of quantitative traits*. Sinauer Sunderland, MA.

51. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120.

52. Li H, Durbin R. 2009 Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760.

53. Schlötterer C, Tobler R, Kofler R, Nolte V. 2014 Sequencing pools of individuals—mining genome-wide polymorphism data without big funding. *Nat. Rev. Genet.* **15**, 749–763.

54. Li H *et al.* 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.

55. Kofler R, Pandey RV, Schlötterer C. 2011 PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**, 3435–3436.

56. Kofler R, Schlötterer C. 2012 Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics* **28**, 2084–2085.

57. Yang C-P, Fu C-C, Sugino K, Liu Z, Ren Q, Liu L-Y, Yao X, Lee LP, Lee T. 2016 Transcriptomes of lineage-specific Drosophila neuroblasts profiled by genetic targeting and robotic sorting. *Development* **143**, 411–421.

58. Schaefer M, Petronczki M, Dorner D, Forte M, Knoblich JA. 2001 Heterotrimeric G proteins direct two modes of asymmetric cell division in the Drosophila nervous system. *Cell* **107**, 183–194.

59. Orgogozo V, Schweisguth F, Bellaïche Y. 2001 Lineage, cell polarity and inscuteable function in the peripheral nervous system of the Drosophila embryo. *Development* **128**, 631–643.

60. Nagy O *et al.* 2018 Correlated Evolution of Two Copulatory Organs via a Single cis-Regulatory Nucleotide Change. *Curr. Biol.* **28**, 3450–3457.e13.

61. Skaer N, Simpson P. 2000 Genetic Analysis of Bristle Loss in Hybrids between Drosophila melanogaster and D. simulans Provides Evidence for Divergence of cis-Regulatory Sequences in the achaete–scute Gene Complex. *Dev. Biol.* **221**, 148–167.

62. Marcellini S, Simpson P. 2006 Two or four bristles: functional evolution of an enhancer of scute in Drosophilidae. *PLoS Biol.* **4**, e386.

63. Green JE, Cavey M, Médina Caturegli E, Aigouy B, Gompel N, Prud'homme B. 2019 Evolution of Ovipositor Length in Drosophila suzukii Is Driven by Enhanced Cell Size Expansion and Anisotropic Tissue Reorganization. *Curr. Biol.* **29**, 2075–2082.e6.

64. Lamichhaney S, Han F, Berglund J, Wang C, Almén MS, Webster MT, Grant BR, Grant PR, Andersson L. 2016 A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* **352**, 470–474.

65. Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE. 2009 On the origin and spread of an adaptive

allele in deer mice. *Science* **325**, 1095–1098.

66. Rost S *et al.* 2004 Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. *Nature* **427**, 537–541.

67. Dmitriev DA, Rakitov RA. 2008 Decoding of superimposed traces produced by direct sequencing of heterozygous indels. *PLoS Comput. Biol.* **4**, e1000113.

Figures & Tables:

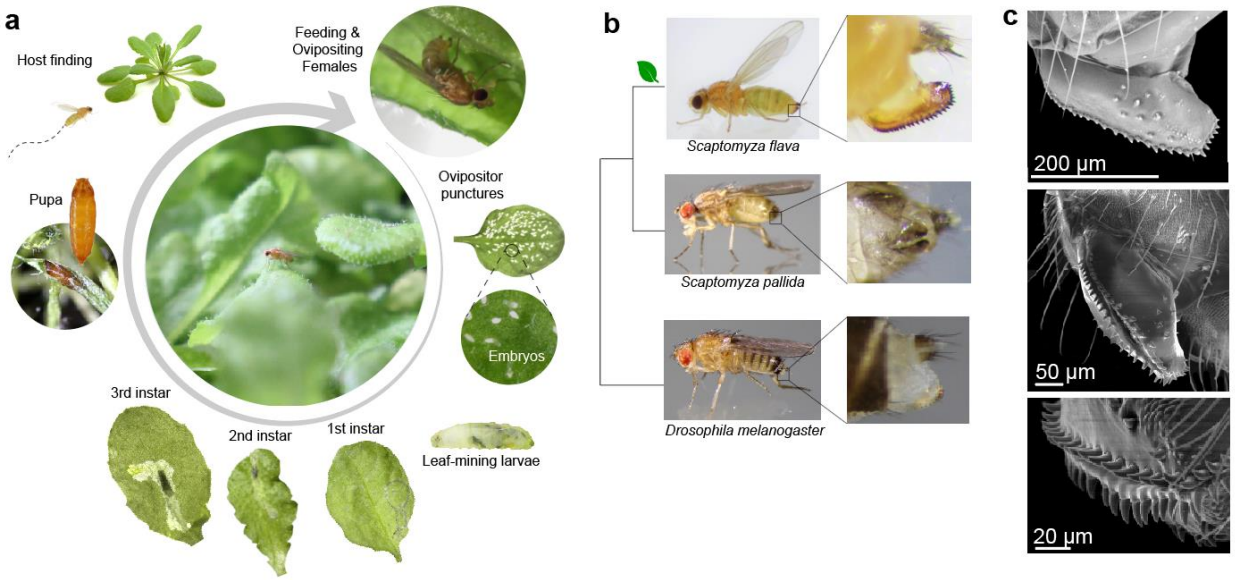


Figure 1.

Overview of the function and morphology of the female ovipositor in the herbivorous drosophilid *Scaptomyza flava.* (a) Life cycle of *S. flava* is strongly dependent on host plants for female nutrition, larval development, and mating sites. Female on the underside of *Arabidopsis thaliana* leaf uses her dentate ovipositor to scoop out numerous leaf punctures for feeding and egg-laying. (b) Comparison of the ovipositors (insets) of herbivorous and non-herbivorous *Scaptomyza* species and *D. melanogaster*. (c) Scanning electron micrographs of the ovipositor of *S. flava*, showing from left to right lateral, ventrolateral, and close-up ventrolateral view. Hardened, curved bristles line the ventral and apical sides of the ovipositor, with shorter, more stout bristles on lateral sides.

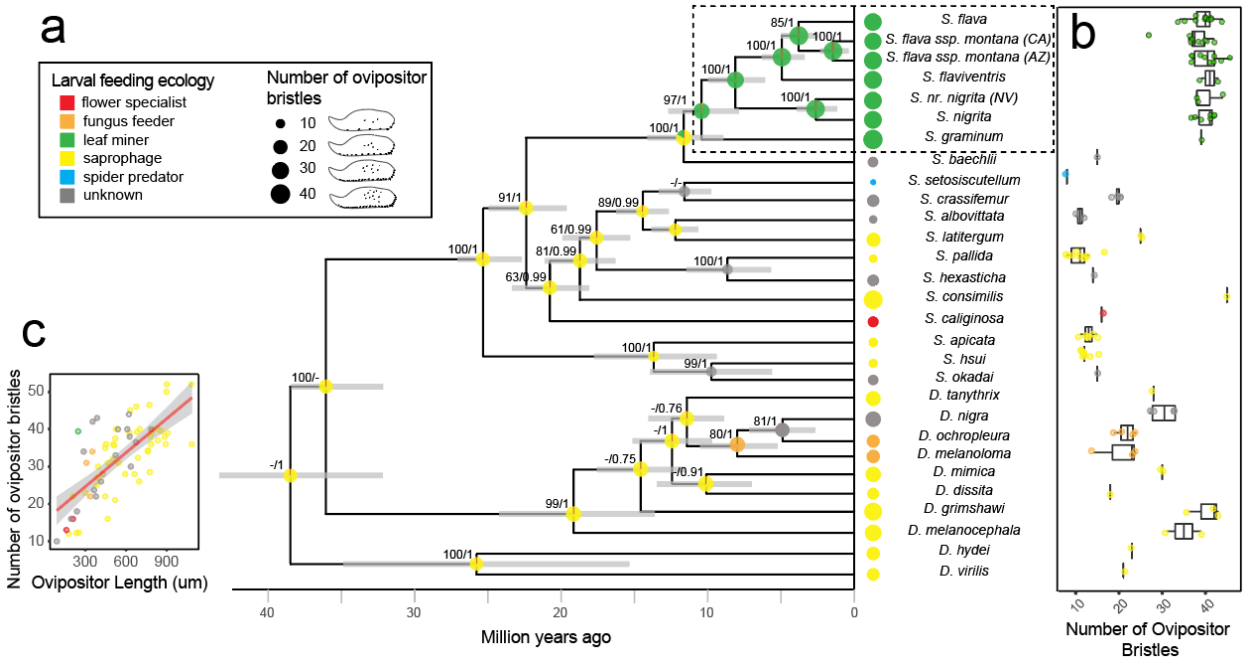


Figure 2.

Evolution of ovipositor bristle number in drosophilids. (a) Time-calibrated phylogeny of herbivorous *Scaptomyza* and their non-herbivorous drosophilid relatives, based on Maximum Likelihood (ML) and Bayesian analyses using 11 genes (16S, 28S, *Adh*, *Cad-r* (rudimentary), *COI*, *COII*, *gstd1*, *gpdh*, *marf*, *ND2*, *n(l)tid)*, *Orco*) and fossil and biogeographic time calibrations. Branch support is indicated by the bootstrap value (≥50%) from ML analysis and the Bayesian posterior probability (≥0.9). Bars at the nodes indicate the 95% highest posterior density interval around the mean node age. Pie graphs at the nodes show probabilities of ancestral character states for larval diets from ML ASR, with yellow = detritivorous, green = leaf-mining, orange = fungus feeding, blue = spider predation, gray = unknown, and the size of each pie-graph represents the ancestral character estimation of the number of bristles on each ovipositor valve again from ML ASR. Circles at the tips represent the number of ovipositor bristles within each extant species, averaged over the sampled individuals shown in (b). (c) Scatterplot of the number of ovipositor bristles per valve for *Drosophila* and *Scaptomyza* species, as a function of ovipositor length and

color coded according as non-herbivorous *Drosophila* = blue, non-herbivorous *Scaptomyza* = purple, herbivorous *Scaptomyza* = green. Phenotypic measurements for 67 species were taken from Craddock et al. and 4 species from this study (see Supplemental Methods and Table S1 for details).

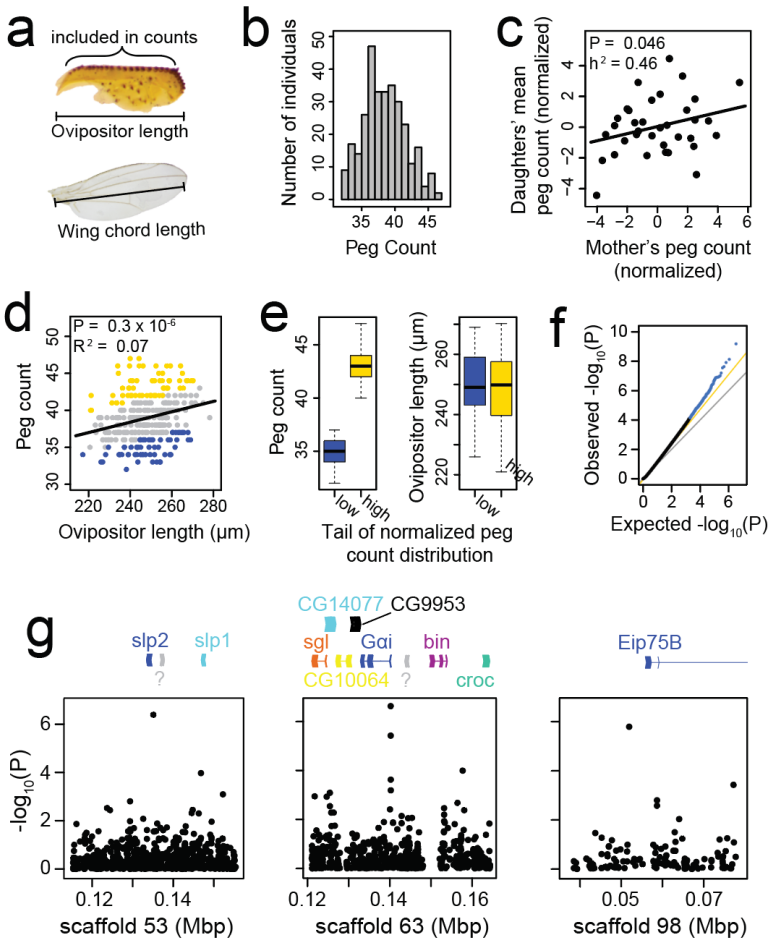


Figure 3.

Genome-wide association mapping implicates conserved neural developmental pathways in development of the densely serrated, plant-cutting ovipositor of *S. flava*. Panels b, d, and e show the NH1 pool-GWAS mapping population only. **(a)** Ovipositor bristle counts included those lining the cutting-margin of the ovipositor. Wing chord length (a proxy for body size) was taken from the base of the wing to the apex along the third longitudinal vein. **(b)** Ovipositor bristle number variation is normally distributed in *S. flava*. **(c)** Ovipositor bristle count, expressed as the

residuals from a lineage regression of bristle count against ovipositor length, is heritable in the narrow sense (h^2 = 0.46) in mother-daughter regression analysis (N = 35). **(d)** After regressing out the effect of ovipositor length on bristle count, pools of phenotypically extreme individuals for genome sequencing were constructed by combining individuals in the upper (yellow) or lower (blue) 20% tail of the ovipositor bristle count distribution. **(e)** Individuals in the low ovipositor bristle count pool had ~20% fewer bristles, but not statistically different ovipositor lengths, than individuals in the upper 20% tail of the phenotype distribution. **(f)** The genome-wide distribution of pool-GWAS *P*-values, shown for common SNPs (minor allele frequency > 8%) interrogated by pooled genome sequencing, was inflated relative to the expected distribution (gray line) if most SNPs were not associated with ovipositor variation. After adjusting *P*-values to eliminate inflation, based on the expected distribution shown by the yellow line, an excess of strong SNP-phenotype associations persisted, consistent with an enrichment of true positive associations. The top 0.05% associations are plotted in blue; all other SNPs are plotted in black. **(g)** Genes encoding transcription factors that differ in expression level among neural cell types are enriched near SNPs most strongly associated with ovipositor bristle count variation. SNPs upstream of *slp2* and *Eip75B* are the 3rd and 10th highest ranking SNPs genome-wide (FDR = 21%). **(j)** Effect size validation of a candidate SNP identified from the pooled GWAS on ovipositor bristle number. The candidate SNP was located upstream of the gene *G-alpha i subunit (Gαi),* which encodes a G Protein involved in asymmetric mitosis of sensory organ precursor cells. *S. flava Gαi* gene region and the location of the focal candidate SNP identified from the GWAS, 48 base pairs upstream of the start codon. Effect of the candidate SNP on ovipositor bristle number and ovipositor length in *Scaptomyza flava*.

Table 1.

Enriched gene sets co-localized with SNPs exhibiting the strongest GWAS associations.

Candidate SNPs	Enriched Gene Set	Intersections with SNPs			Fold Enrichment	P	P (Bonferroni)	Genes Intersecting Candidate SNPs (named by orthology to D. melanogaster)*
		Observed	Expected	Possible				
Top 0.05%	Regulation of neuron maturation (GO:0014041); Regulation of neuron remodeling (GO:1904799, redundant to GO:0014041)	8	2.362	28	3.4	0.000675	0.068175	Dhc64C (x5), Fas2, hdc, Mical
Top 0.05%	DNA-binding transcription repressor activity, RNA polymerase II-specific (GO:0001227)	10	2.899	64	3.5	0.000465	0.073935	CG1233, CG7987, chn, dpn, E(spl)mbeta-HLH, E5, h, HHEX, lms, Mad
Top 0.001%	Transcription factors differentially expressed during neural cell differentiation (Yang et al. 2016)	2	0.053	40	37.8	0.00122	NA	Eip75B, slp2

* In one case, multiple *S. flava* genes belonged to orthology clusters that contained more than one *D. melanogaster* gene. "Dhc64C (x5)" indicates that five *S. flava* genes belonging to the same orthology cluster as *D. melanogaster* Dhc64C intersected with the candidate SNP set, but only one representative gene name is listed for simplicity.

Supplemental Figures

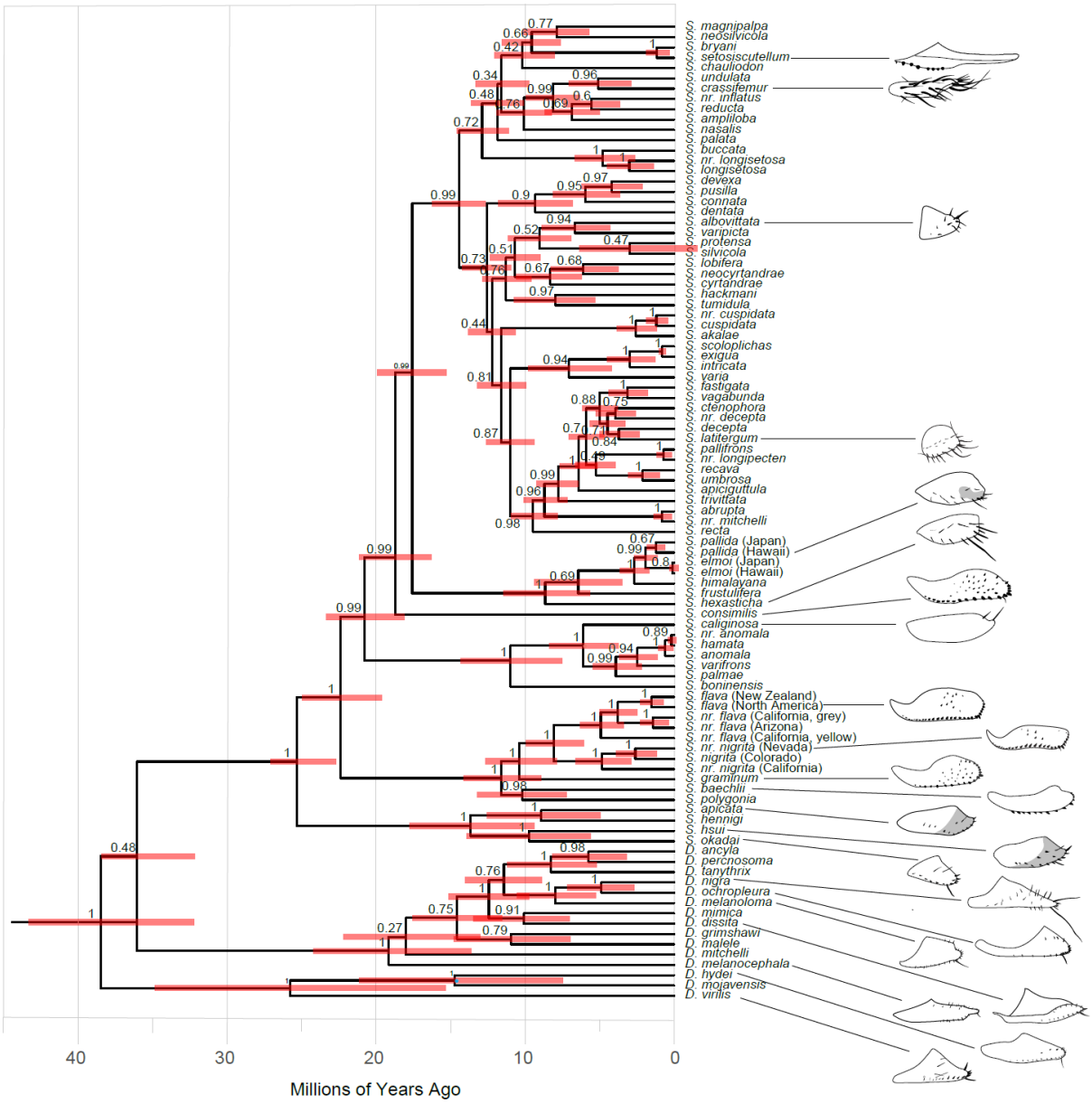


Figure S1 – Bayesian tree.

Time calibrated phylogeny of subgenus *Drosophila* and *Scaptomyza* inferred from Bayesian analysis, using 11 genes (16S, 28S, *Adh*, *Cad-r* (rudimentary), *COI*, *COII*, *gstd1*, *gpdh*, *marf*, *ND2*, *n(l)tid)*, *Orco*) and fossil and biogeographic time calibrations. Nodes indicate posterior probabilities, red bars represent 95% highest posterior density interval around the mean node age.

Illustrations of female ovipositors, indicating bristle number, were drawn from sources indicated in Table S1.

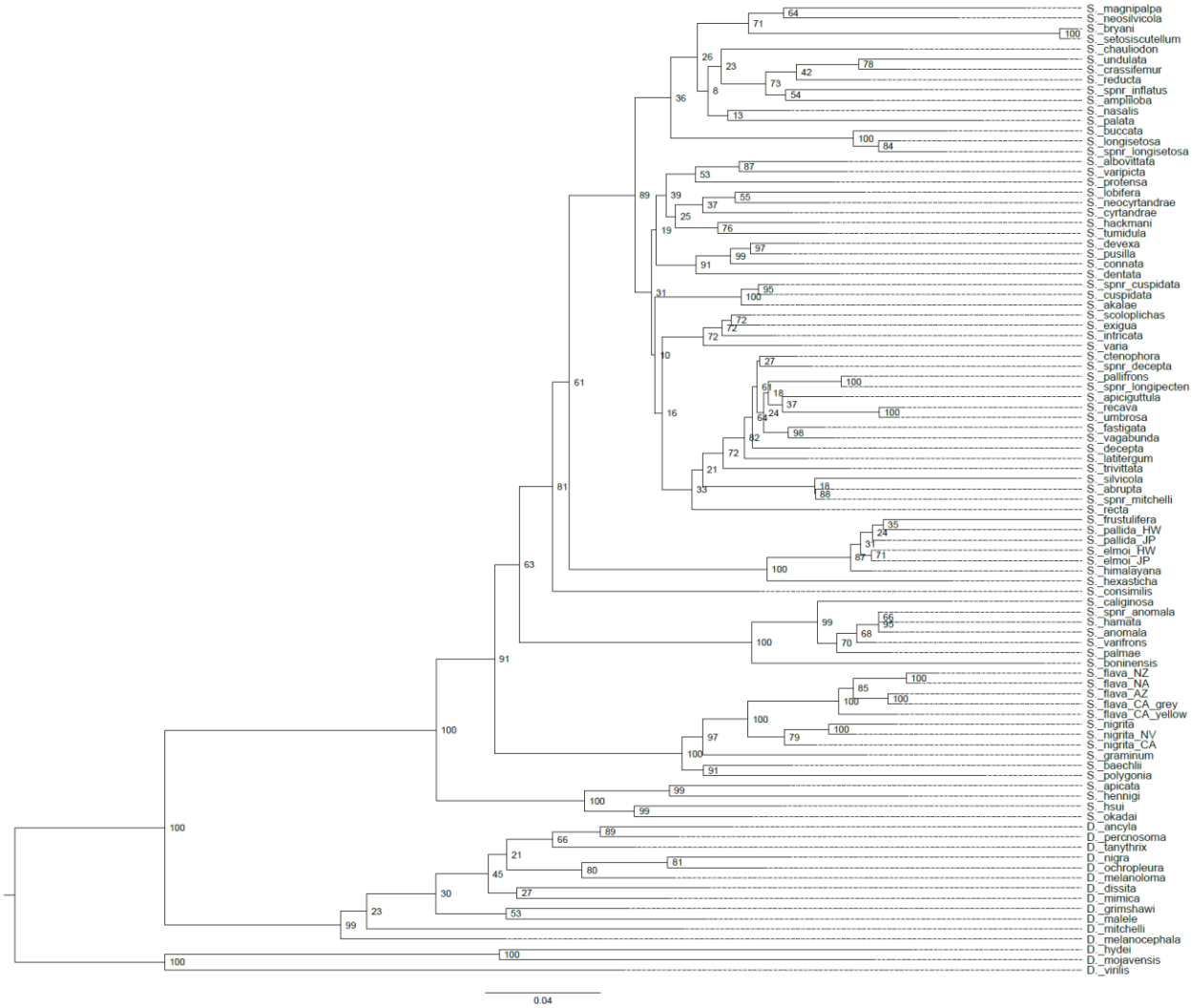
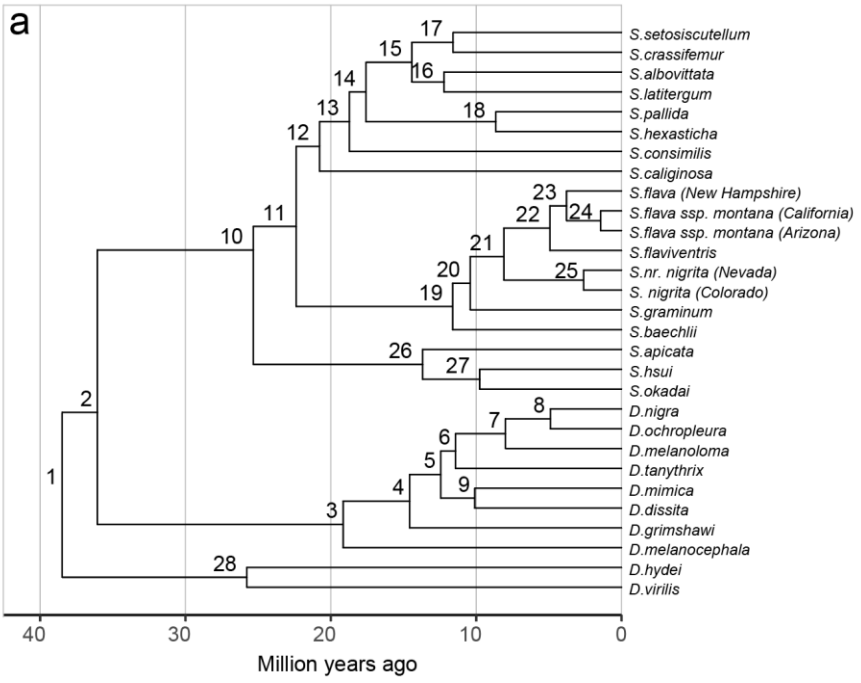


Figure S2 – ML tree.

Phylogeny of subgenus *Drosophila* and *Scaptomyza* inferred from maximum likelihood (ML), using 11 genes (16S, 28S, *Adh*, *Cad-r* (rudimentary), *COI*, *COII*, *gstd1*, *gpdh*, *marf*, *ND2*, *n(l)tid)*, *Orco*) and fossil and biogeographic time calibrations. Nodes represent bootstrap value (≥50%) from ML analysis.



b

Node	Lower bound	Ovisenislla number ASR	Upper bound
1	15.32	17.59	32.91
2	16.11	16.41	32.52
3	22.44	13.58	36.02
4	23.74	10.88	34.63
5	22.80	9.45	32.25
6	22.49	9.46	31.95
7	21.34	9.25	30.59
8	21.83	8.31	30.14
9	21.33	10.15	31.49
10	15.71	12.74	28.46
11	17.34	11.28	28.63
12	16.92	10.92	27.84
13	17.00	10.49	27.50
14	15.62	10.34	25.95
15	13.60	10.22	23.82
16	13.07	10.89	23.96
17	11.60	10.96	22.56
18	9.69	11.10	20.79
19	25.16	10.23	35.39
20	27.90	9.58	37.49
21	31.51	8.79	40.30
22	34.70	7.36	42.06
23	35.30	6.74	42.04
24	36.42	4.86	41.27
25	35.93	6.58	42.50
26	9.85	12.53	22.38
27	9.26	11.47	20.73
28	14.07	18.00	32.06

Figure S3 – CI for ASR on ovipositor bristle number.

Confidence intervals for ancestral state estimations of ovipositor bristle number. Values in the table correspond to respective nodes given in the phylogeny.

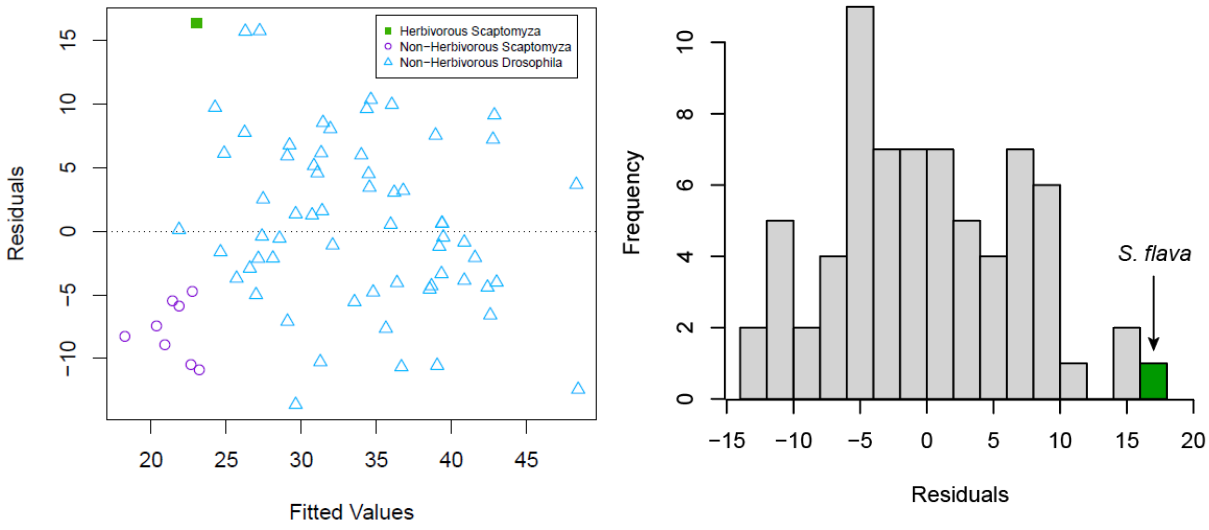


Figure S4 – Residual Analysis.

Residual analysis on linear regression model on ovipositor bristle number. Ovipositor bristle number, ovipositor length, and larval feeding ecology were obtained from Craddock et al. 2018, with additional data from this study (Suppl. Dataset S3). (a) Scatter plot of residuals on the y-axis and fitted values (estimated responses) on the x axis. (b) Histogram of residuals from (a).

Table S1.

Ovipositor bristle number and larval feeding ecology data for 95 species included in phylogenetic analyses to estimate the ancestral character states and perform phylogenetic generalized least squares. * Data from [35].

Species	**Larval Diet***	**Source for Bristle Counts**	**Number of Ovipositor Bristles**
D. ancyla	saprophagy		
D. dissita	saprophagy	Throckmorton 1966	18
D. grimshawi	saprophagy	Craddock et al. 2018	36
			42
			43
D. hydei	saprophagy	Baechli et al. 2004	23
D. malele	unknown		
D. melanocephala	saprophagy	Craddock et al. 2018	39
		Hardy 1967	31
D. melanoloma	fungus feeding	Throckmorton 1966	23
D. mimica	saprophagy	Craddock et al. 2018	30
			30
D. mitchelli	unknown		
D. mojavensis	saprophagy		
D. nigra	unknown	Hardy et al. 2001	28
		Throckmorton 1966	27
		Craddock et al. 2018	33
			33
D. ochropleura	fungus feeding	Hardy et al. 2001	21
		Craddock et al. 2018	19
		Craddock et al. 2018	23
		Craddock et al. 2018	24
D. percnosoma	saprophagy		

D. tanythrix	saprophagy	Craddock et al. 2018	28
D. virilis	saprophagy	Sturtevant 1921	20
S. abrupta	unknown		
S. akalae	unknown		
S. albovittata	unknown	Craddock & Kambysellis 1997	12
		Craddock et al. 2018	10
S. ampliloba	unknown		
S. anomala	unknown		
S. nr. anomala	unknown		
S. apicata	saprophagy	Mount Hood National Forest, Camp Creek	11
		Mount Hood National Forest, Trout Creek	13
			13
		Lab colony reared from wild caught individuals from Strawberry Creek, Berkeley, California	12
			14
			15
S. apiciguttula	saprophagy		
S. baechlii	unknown	Sidorenko 1993	15
S. boninensis	unknown	Okada 1973	
S. bryani	spider predation		
S. buccata	unknown		
S. caliginosa	flower specialist	Craddock et al. 2018	16
S. chauliodon	spider predation		
S. connata	unknown		
S. consimilis	saprophagy	Baechli et al. 2004	45
S. crassifemur	unknown	Throckmorton 1966	20
		Craddock et al. 2018	18
		Grimaldi	20
		Hardy 1965	20
S. ctenophora	unknown		
S. cuspidata	saprophagy		
S. nr. cuspidata	saprophagy		
S. cyrtandrae	saprophagy		
S. decepta	unknown		
S. nr. decepta	unknown		
S. dentata	unknown		
S. devexa	unknown		
S. elmoi (Hawaii)	saprophagy		
S. elmoi (Japan)	saprophagy		
S. exigua	saprophagy		
S. fastigata	unknown		
S. flava (North America)	leaf-mining		34
			35

		Lab colony, reared from wild caught individuals from New Hampshire	38
			38
			40
			40
			40
			40
			41
			41
			41
			44
S. flava (New Zealand)	leaf-mining		
S. flava ssp. montana (California)	leaf-mining	Strawberry Creek, Berkeley, California	39
			37
		Lab colony, reared from wild caught individuals from Strawberry Creek, Berkeley, CA	27
			36
			37
			37
			38
			40
			40
			42
S. flava ssp. montana (Arizona)	leaf-mining	Lab colony, reared from wild caught individuals from Arizona	37
			37
			37
			39
			40
			41
			42
			42
			44
			46
S. flaviventris	leaf-mining	Santa Cruz, California	39
			43
S. frustulifera	unknown		
S. graminum	leaf-mining	Baechli et al. 2004	39
S. hackmani	saprophagy		
S. hamata	unknown		
S. hennigi	unknown		
S. hexasticha	unknown	Okada 1973	14
S. himalayana	unknown		
S. hsui	saprophagy	Thomas Creek, Nevada	11
			11

			12
			12
			12
			12
			13
		Mount Hood National Forest, Still Creek, Oregon	12
		Rocky Mountain Biological Lab, near Gothic, Colorado	12
			15
S. nr. inflatus	unknown		
S. intricata	saprophagy		
S. latitergum	saprophagy	Throckmorton 1966	25
S. lobifera	unknown		
S. nr. longipecten	saprophagy		
S. longisetosa	unknown		
S. nr. longisetosa	unknown		
S. magnipalpa	spider predation		
S. nr. mitchelli	unknown		
S. nasalis	unknown		
S. neocyrtandrae	saprophagy		
S. neosilvicola	spider predation		
S. nigrita (Colorado)	leaf-mining	Rocky Mountain Biological Lab, near Gothic, Colorado	37
			38
			39
			41
			41
			42
			42
S. nr. nigrita (California)	leaf-mining		
S. nr. nigrita (Nevada)	leaf-mining	Lab colony, reared from wild caught individuals from Lake Tahoe, Nevada	38
			38
			44
S. okadai	unknown	Hackman 1959	15
S. palata	unknown		
S. pallida (Japan)	saprophagy		
S. pallida (North American)	saprophagy	Mount Hood National Forest, Still Creek, Oregon	8
			8
			9
			10
			11
			12
			12
		Mount Hood National Forest, Camp Creek, Oregon	13

		Baechli et al. 2004	17
S. pallifrons	unknown		
S. palmae	flower specialist		
S. polygonia	saprophagy		
S. protensa	unknown		
S. pusilla	unknown		
S. recava	unknown		
S. recta	unknown		
S. reducta	unknown		
S. scoloplichas	saprophagy		
S. setosiscutellum	spider predation	Hardy 1965	8
S. silvicola	unknown		
S. trivittata	unknown		
S. tumidula	saprophagy		
S. umbrosa	unknown		
S. undulata	unknown		
S. vagabunda	unknown		
S. varia	saprophagy		
S. varipicta	saprophagy		
S. varifrons	unknown		

Table S2.

Ovipositor length, ovipositor bristle number, and larval feeding ecology data for 67 species from Craddock et al. 2018 and 4 species from this study, used to evaluate the relationship between ovipositor length and bristle number. Ovipositor bristle counts were averaged across individuals and ovipositor valves, when both were present.

Species	**Larval Diet**	**Thorax length (mm)**	**Ovipositor length (um)**	**Ovipositor bristle number**
D. ambochila	decaying bark	1.57	858.5	39.5
D. bostrycha	decaying bark	NA	630	45
D. craddockae	decaying bark	1.92	763	34.375
D. cyrtoloma	decaying bark	3.37	906	39
D. engyochracea	decaying bark	2.47	901	52
D. hemipeza	decaying bark	2.46	701	40

D. melanocephala	decaying bark	3.07	789	39
D. nigribasis	decaying bark	2.84	898	50
D. oahuensis	decaying bark	2.95	835	40
D. orphnopeza	decaying bark	2.11	1085	36
D. peniculipedis	decaying bark	NA	465	31
D. primaeva	decaying bark	3	446.5	35
D. pullipes	decaying bark	2.42	835	37
D. silvestris	decaying bark	2.59	772	46.5
D. sproati	decaying bark	NA	892	36
D. mimica	decaying fruit	1.77	393.5	30
D. adunca	decaying leaves	2.66	775.5	28.5
D. antecedans	decaying leaves	1.73	300	23
D. conjectura	decaying leaves	NA	209	22
D. diamphidiopoda	decaying leaves	2.17	519	21
D. kambysellisi	decaying leaves	1.42	383	25
D. tanythrix	decaying leaves	2.39	663	28
D. waddingtoni	decaying leaves	1.31	465	16
*S.apicata**	decaying leaves	NA	253.32	12.33
*S.hsui**	decaying leaves	NA	235.35	12.2
*S.pallida**	decaying leaves	NA	177.75	12
S. caliginosa	flower specialist	0.9	209	16
S. sp. 2	flower specialist	NA	160	13
D. fungiperda	fungus	1.6	308	31
D. iki	fungus	1.74	429.5	28
D. nigella	fungus	1.77	353	34
D. ochropleura	fungus	1.52	335.3333333	22

*S. flava**	leaf-mining	NA	247.52	39.42
D. longiseta	decaying leaves	2.92	697	26
D. plantibia	multiple	2.71	786.5	40
D. imparisetae	multiple	1.19	391	27
D. disjuncta	multiple	2.2	521	37.5
D. grimshawi	multiple	2.13	681.25	39.25
D. eximia	multiple	1.27	525	40
D. hirititibia	multiple	1.06	785	36
D. hawaiiensis	sap flux	2	546	31
D. musaphilia	sap flux	NA	1081	52
D. picticornis	sap flux	1.77	512.66	35.66
D. recticilia	sap flux	NA	687.33	32.33
D. silvarentis	sap flux	NA	594	28
D. heedi	sap flux	1.53	447	22
D. cilifera	decaying stem	2.17	501	32
D. adiastola	decaying stem	2.14	451.5	36
D. assita	decaying stem	1.48	780	38
D. clavisetae	decaying stem	2.71	625	39
D. differens	decaying stem	3.09	886	38
D. limitata	decaying stem	NA	759	34
D. ornata	decaying stem	2.48	676	46
D. setosimentum	decaying stem	2.13	505	36
D. cilifemorata	unknown	1.71	378	22
D. comatifemora	unknown	1.96	415	26
D. eurypeza	unknown	1.48	354.5	42
D. fasciculisetae	unknown	2.49	673.5	36.5

D. formella	unknown	NA	621	44
D. haleakalae	unknown	1.9	288	34
D. hamifera	unknown	2.71	541	40
D. hirtipalpus	unknown	NA	627	38
D. longiperda	unknown	2	386	43
D. mulli	unknown	1.77	786	40
D. nigra	unknown	2.18	523	33
D. soonae	unknown	1.645	364.33	23.66
D. stigma	unknown	2.32	635	30
D. truncipenna	unknown	3.22	609	40
S. albovittata	unknown	0.79	90	10
S. crassifemur	unknown	1.98	238	18
S. sp. 1	unknown	1	195	16

Supplemental Methods:

Candidate SNP (upstream of *gai*) validation: DNA was extracted from whole adult flies with the Qiagen DNeasy blood and tissue kit, following provided protocol in kit. PCR primers for *Gαi* were designed in Geneious, using default settings and a target region size of 500bp. The primer sequences are as follows: gai4F: CATTTCTGTCCATGGCGTCG; gai9R: GCCGTTAGACAAAGCGCATT. PCR methods were the same as those in Gloss et al. (2013). PCR clean up and sequencing were performed at the UC Berkeley DNA Sequencing Facility. PCR products were sequenced in both directions.

For trimming, alignment, and base calling, we used Geneious v.10.0.5 (Biomatters Ltd.) Sequences were trimmed with the Trimming Tool in Geneious, and regions with more than a 5% chance of an error per base were removed. Sequences were aligned with default "Geneious Alignment" setting (Cost matrix: 65% similarity, Gap open penalty: 12, Gap extension penalty: 3, Refinement Iterations: 2, Alignment type: Global alignment with free end gaps). The final alignment was 331 base pairs in length. A significant fraction of sequences had convoluted regions, upstream of the SNP. Manual base-calling was therefore performed on all sequences. Double peaks were called as heterozygotes. Sequences with convoluted regions were analyzed with Indelligent v.1.2, which also identified indels.[67] Variant sites, including the candidate SNP, were identified in Geneious, using a minimum variant frequency cutoff of 0.05.

We performed a test for Hardy-Weinberg equilibrium (HWE) for the focal SNP among all individuals using the *hw.test* function from the pegas package. The SNP was found to be in HWE ($\chi 2 = 3.15$, $P > 0.05$).