

Reviewed Preprint

v2 • September 30, 2024

Revised by authors

Reviewed Preprint

v1 • August 24, 2023

# DGRPpool: A web tool leveraging harmonized *Drosophila* Genetic Reference Panel phenotyping data for the study of complex traits

Vincent Gardeux, Roel PJ Bevers, Fabrice PA David, Emily Rosschaert, Romain Rochepeau, Bart Deplancke 

Laboratory of Systems Biology and Genetics, Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland • Swiss Institute of Bioinformatics, Lausanne, Switzerland • Bioinformatics Competence Center, EPFL, Switzerland • KU Leuven, Belgium

 [https://en.wikipedia.org/wiki/Open\\_access](https://en.wikipedia.org/wiki/Open_access)

 Copyright information

## Abstract

Genome-wide association studies have advanced our understanding of complex traits, but studying how a GWAS variant can affect a specific trait in the human population remains challenging due to environmental variability. *Drosophila melanogaster* is in this regard an excellent model organism for studying the relationship between genetic and phenotypic variation due to its simple handling, standardized growth conditions, low cost, and short lifespan. The *Drosophila* Genetic Reference Panel (DGRP) in particular has been a valuable tool for studying complex traits, but proper harmonization and indexing of DGRP phenotyping data is necessary to fully capitalize on this resource. To address this, we created a web tool called **DGRPpool** ([dgrpools.epfl.ch](http://dgrpools.epfl.ch)), which aggregates phenotyping data of 1034 phenotypes across 135 DGRP studies in a common environment. DGRPpool enables users to download data and run various tools such as genome-wide (GWAS) and phenome-wide (PheWAS) association studies. As a proof-of-concept, DGRPpool was used to study the longevity phenotype and uncovered both established and unexpected correlations with other phenotypes such as locomotor activity, starvation resistance, desiccation survival, and oxidative stress resistance. DGRPpool has the potential to facilitate new genetic and molecular insights of complex traits in *Drosophila* and serve as a valuable, interactive tool for the scientific community.

### eLife Assessment

Genetic analysis of complex traits in *Drosophila* provides a resource for exploring the relationship between genetic and phenotypic variation. The web tool DGRPpool presented in this paper makes data and results from the *Drosophila* Genetic Reference Panel accessible that will enable downstream analyses of genetic association. The findings of this paper are considered to be **important**, with practical implications beyond a single subfield, supported by **convincing** evidence using appropriate and validated methodology in line with current state of the art.

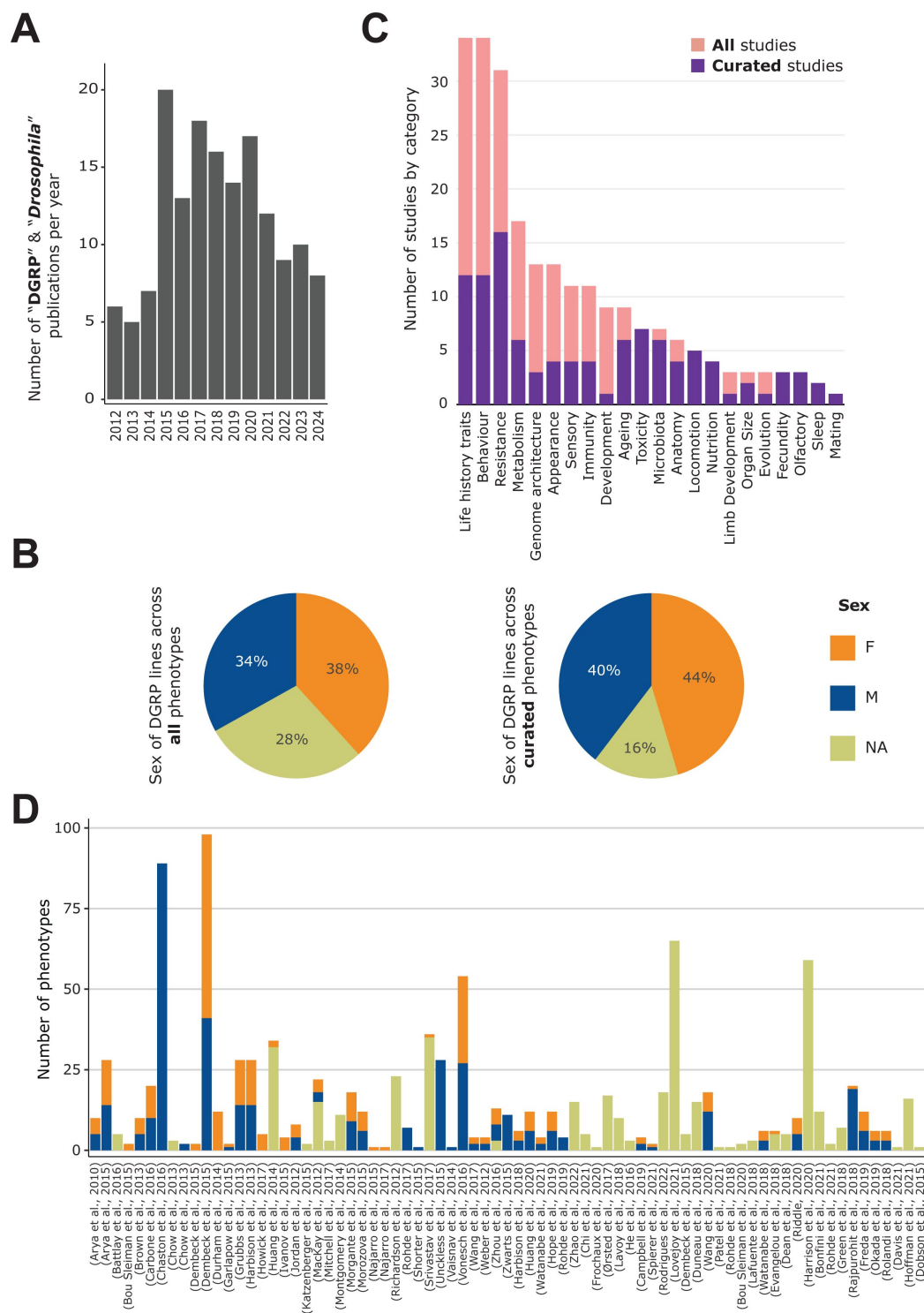
<https://doi.org/10.7554/eLife.88981.2.sa3>

## Introduction

*Drosophila melanogaster* is an excellent model organism for studying genotype-to-phenotype relationships. It is a short-living species and is very easy to maintain in similar laboratory conditions, which limits confounding factors such as the environment. The *Drosophila* Genetic Reference Panel (DGRP) was created in the early 2010s and now consists of 205 inbred lines that are fully sequenced, of which 192 are still available in the Bloomington *Drosophila* Stock Center (<https://bdsc.indiana.edu/><sup>1,2</sup>). The DGRP has proven highly valuable to study the genetic basis of complex traits, as illustrated by the many studies that have used GWAS principles to identify variants that contribute to traits related to morphology, metabolism, behavior, aging, disease susceptibility etc. (Figure 1A<sup>3</sup>). Furthermore, since the DGRP lines were inbred for many generations, they are almost fully homozygous, which simplifies the identification of putatively causal alleles and elucidation of implicated molecular mechanisms<sup>3</sup>. Moreover, the fact that the same lines can be studied by various researchers for diverse traits should leverage these data generation efforts to uncover unexpected correlations between phenotypes or relationships between genetic variants and a wide range of traits.

However, there is currently only one major data resource that aims to compile DGRP information, the DGRP2 website (<http://dgrp2.gnets.ncsu.edu/>)<sup>1,2</sup>. This website hosts the genotyping data, its annotation, and potential known covariates (well-established inversion genotypes and symbiont levels), as well as 31 phenotypes from 12 studies (Table 1<sup>3</sup>). The data is primarily hosted as static files, downloadable from the website, along with limited RNA expression data. In addition, a very important tool, used by the DGRP community, is the possibility for any user to submit their own phenotype files for running a GWAS analysis (corrected for the so-called “known covariates”, i.e. established inversions and *Wolbachia* infection status). This is particularly useful, especially for researchers that do not have the bioinformatics knowledge or capacity to perform these tasks internally. However, the DGRP2 website has not been updated for an extended period as the last referenced paper dates back to 2015, and, except for the GWAS computation, has remained static. This means that any meta-study, which would aim to aggregate datasets across available phenotypes, would require hours (if not days) of work to transform the data into an appropriate and common format. Moreover, the result of such effort would unlikely become available to the rest of the community, and thus any other group would need to redo this work in order to gather similar information, while the data of other phenotyping studies beyond the 12 available would not be easily accessible.

For all these reasons, we decided to create a web application that would both act as a repository of DGRP phenotyping datasets and also as an online tool for assisting researchers with some basic systems genetics-inspired analyses. Our goal was to index all existing literature about DGRP



**Figure 1.**

### General content of the DGRP web tool.

**A.** Pubmed search on "Drosophila DGRP" terms unveiled 155 results from 2012 to 2024 (search made in July 2024). **B.** Sex of the DGRP lines used across all 135 studies (left) and 43 curated studies (right), for each phenotype. **C.** Number of studies per phenotype category. Studies can be assigned to multiple categories. **D.** Number of phenotypes per study and per sex. Studies without attached phenotypes were not plotted. Of note, a given phenotype can be measured for different sexes and thus counted multiple times.

REFERENCE			DGRPpool	DGRP2 (Mackay et al., 2012) (Huang, et al., 2014)
DATA	DGRP lines		342	205
	DGRP studies		135 (43 fully curated)	12
	Phenotypes		838 (1034 sex-specific)	31
	Gene Expression data		External links	✓
TOOLS	GWAS	Calculated on all phenotypes	✓	
		User upload	✓	✓
		Method	Plink2	FastLMM
		Covariates	Wolbachia + 5 Insertions	Wolbachia + 5 Insertions
		Boxplot of REF vs ALT	✓	
		PheWAS of top variants	✓	
	Phenotype correlation	Calculated on all phenotypes	✓	
		User upload	✓	
WEB	URL		<a href="https://dgrpools.epfl.ch/">https://dgrpools.epfl.ch/</a>	<a href="http://dgrp2.gnets.ncsu.edu/">http://dgrp2.gnets.ncsu.edu/</a>
	Backend		Ruby-on-rails + PostgreSQL	NA
	Frontend		Javascript, Plotly	NA
FEATURES	Curation system & tools		✓	
	Publish new studies		✓	
	Interactive plots		✓	

**Table 1.**

### Comparison of the two currently available web portals organizing DGRP phenotyping data.

This table compares different features available in DGRPpool, with DGRP2, the current main resource for DGRP data. It separates the features into 1) **Data**, which summarizes the available phenotyping data, 2) **Tools**, which lists the available tools and options, mainly GWAS, PheWAS and phenotype correlation, 3) **Web**, which describes the website itself, and 4) **Additional features**, that are available in DGRPpool, such as the curation system, the possibility to publish new studies and the interactive plots. Of note, the 838 phenotypes are counted regardless of the associated sex (M, F and/or NA), while the “sex-specific” value is calculated by counting the same phenotype separately for each available sex.

phenotyping data—where possible—in order for users to quickly search through the website using simple keywords. We manually associated each study with broad and tailored categories such as “ageing”, “metabolism”, or “olfactory”. We specifically spent important time curating the datasets to avoid any errors or misrepresentations of datasets. To avoid the “maintenance issue” that is common to online tools, and keep the data up to date, we implemented specific curators’ tools, to help maintain the web application in the future. These tools allow any user to submit a novel dataset, which is then attributed to a curator, in order to manually format and validate all phenotyping data and metadata associated with the study. Importantly, any user can become a curator, as advertised on the main page of the resource, since we strongly believe that such a community-run resource architecture is most optimal to keep a web tool state-of-the-art and allow crowd-based curation work<sup>4</sup>.

In addition, we set out to build important tools for the DGRP community such that DGRPpool would not only be a static repository for downloading phenotyping data but could also be used as an interactive data analysis tool. For example, users can correlate phenotypes together, from the same study or across studies. We also implemented an automated GWAS analysis (using PLINK2, and known covariates) which we pre-calculated on all the phenotyping data that are currently available. Using this data, users can simply browse through their genes or variants of interest and directly find related phenotypes. A PheWAS page also allows exploration of each variant’s impact across multiple phenotypes. Moreover, these tools are applicable to user-submitted phenotypes, so that anyone can upload their own phenotypes to search the DGRPpool database for correlated phenotypes or to run GWAS analyses.

Our goal is to ensure that DGRP phenotyping data is findable, accessible, interoperable, and reusable (FAIR)<sup>5</sup> to fully leverage the opportunities that stem from this unique genotyping-phenotyping resource. To this end, we made user access our priority, both for removing the bottleneck of data harmonization, and also to allow for better, more reproducible research. This aligns with the harmonization effort undertaken by other human GWAS/PheWAS resources, such as the GWAS Catalog<sup>6</sup>, Open Targets Genetics<sup>7</sup>, MR-base<sup>8</sup>, and the FinnGen portal<sup>9</sup>, which provide extensive examples of effective data use and accessibility. Although the structure of DGRPpool differs from these human databases, we acknowledge the importance of similar meta-data harmonization guidelines. Inspired by the GWAS Catalog’s summary statistics submission guidelines, we propose submission guidelines for DGRP phenotyping data in this paper.

To showcase the potential of our tool in facilitating new biological discoveries, we conducted a proof-of-concept study focusing on the longevity phenotype, a well-studied trait in *Drosophila* research with clear relevance to human longevity<sup>10</sup>. By leveraging the data harmonization and curation efforts in DGRPpool, we identified multiple phenotypes that are significantly associated with longevity across 18 different studies, such as oxidative stress resistance<sup>11</sup>, sleep duration<sup>12,13</sup>, desiccation survival<sup>14,15</sup>, and starvation resistance<sup>14,16,17</sup>. Interestingly, we also observed phenotype-phenotype correlations between “shorter lifespan” and certain phenotypes, such as locomotor activity<sup>18</sup> and food intake<sup>19,20</sup>. These results validate prior knowledge and illustrate how our tool can provide novel biological insights with just a few clicks. Therefore, we firmly believe that tools such as DGRPpool—which ultimately could become entirely community-driven—are essential not only for catalyzing novel research, but also for leveraging the diversity and richness of existing datasets.

## Results

## A thousand phenotypes across 135 studies

To start our data collection, we searched for DGRP studies that reference any phenotyping data and in parallel implemented diverse tools to automatically aggregate these data and their associated metadata from the journals hosting the datasets. However, we quickly realized that it was difficult to automate the entire process. Specifically, the import of phenotyping data proved challenging since i) datasets tended to be hosted in very different formats such as Excel files or PDF, ii) data was stored within the journal's supplementary section, or in external repositories such as Figshare; and iii) the format of the phenotyping data differed from one publication to another. Because of these challenges, we implemented a curation page to manually review, edit, and correct datasets that were automatically aggregated, aiming to prevent errors in the imported datasets. In addition, this allows the curator to add relevant remarks or comments on the study under review, thus providing enhanced context for future analyses of these datasets.

In line with the community-resourcing philosophy of DGRP, we created a specific “curator” role that any logged-in user can claim, again with the underlying rationale of assuring long-term sustainability of our web application. With this role, the user has access to additional functionalities on the DGRP website, including the modification of any metadata attached to a study (title, authors, description, categories), and the submission or modification of attached phenotypes (Supp. Figure S1). Although this may entail a considerable amount of time, we assert that this approach is the most effective means of furnishing high-quality data. Consistent with this philosophy, we have incorporated a functionality on the homepage which empowers any user to submit a DOI as a recommendation for a study that could be absent from the DGRP repository. If the DOI is not in the database, it triggers the same automated scripts that were originally used to incorporate the 135 studies. The corresponding study is then created on DGRP, and its metadata (authors, links, ...) are automatically imported. Once a study has been created, one of three possible labels can be assigned to describe the state of curation of a study: 1) **Submitted** (default), when no curator is yet assigned to the study, 2) **Under curation**, when a curator is assigned, and 3) **Curated** when all phenotyping data and metadata have been curated, and the study received final approval by the curator. At this time, DGRP hosts 135 studies, including 43 that have already been fully curated, 80 still under curation, and 12 under a submitted status (i.e. the DOI was submitted by users as a relevant DGRP study not yet present in the database). In total, 75 studies have attached phenotyping data; 100% of the curated ones, and only 40% of the non-curated ones. Altogether, the total number of studies in DGRP is currently 135, and we expect that this number will continue to grow upon its public release, along with the number of curated studies.

Since the curation process is still ongoing, we will be referring to two different datasets in the manuscript: 1) The **full dataset**, comprising **135** studies (independent of “curation” status), and 2) the **curated dataset**, comprising **43** studies that have already underwent thorough curation and contributed about 505 phenotypes (see below). Of note, for all tools available on the website, it is possible to run these on either all studies or (as is currently the default), only on the curated studies.

For all of the curated studies, we carefully separated the data by sex when information on sex-specific phenotypes was available, or we assigned it as *NA* when flies were sex-mixed, when there was no information on sex, or when the phenotype is inherent to a population (e.g. in the case of non-sexual chromosomal traits, like inversions). We also extracted this information from the phenotyping data itself for the non-curated studies, when available, but when not findable, it was set to *NA*, waiting for a more in-depth curation and careful reading of the paper method's section. Therefore, across all 135 studies, this led to an overall equilibrium between all represented sexes, with slightly more data for females and slightly less unannotated data (**Figure 1B** [↗](#)). However, when focusing only on the 43 curated datasets, the proportion of phenotypes without assigned sex (*NA*) dropped drastically to ~15%. This effect highlights the importance of rigorous curation,



which typically requires the curator to read through the entire manuscript to understand the utilized experimental protocols to select the appropriate sex, even if this information is not explicitly indicated in the phenotyping data itself.

Upon data curation, the assigned curator(s) has to specify a few phenotypic categories for each study, for example, “Metabolism”, “Nutrition”, or “Ageing” (**Figure 1C**). Since these categories are browsable, it facilitates searching for a set of specific studies or linking the studies together. Interestingly, the top annotated categories are either “Behaviour”, “Life History Traits”, or “Resistance”, which is consistent with historical behavioral and immune studies conducted for *Drosophila* as a model organism<sup>21–25</sup>. The number of phenotypes per study ranges from 1 to 89 (**Figure 1D**, Supp Figure S2), with a median of 5, and a mean of 11, revealing that while a low number of phenotypes (usually less than 10) tends to be the norm, some studies aggregate lots of (often similar) phenotypes. An example of the latter is Chaston et al., 2016<sup>26</sup> which investigated the impact of microbiota on nutritional traits. The authors studied 76 different microbial taxa, whose effect was quantified independently, generating a high number of phenotypes. Similarly, Dembeck et al., 2015<sup>27</sup> studied cuticular hydrocarbon composition, considering 66 different cuticular components, while Vonesch et al., 2016<sup>28</sup> studied organismal size traits, regrouping 28 morphological phenotypes such as wing length or intraocular distance. In total, the 43 curated studies aggregate 316 M + 220 F + 133 NA = 669 sex-specific phenotypes, for a total of 505 unique phenotypes (~60%), while the remaining non-curated studies provide another 60 M + 37 F + 268 NA = 365 sex-specific phenotypes, for a total of 333 unique phenotypes (~40%).

## Harmonization and formatting of phenotyping data

DGRP phenotyping data are often available as a supplemental data table, published along with the main paper on the journal’s website. Such data can also be stored on external websites such as Figshare and, as already indicated, the corresponding file can be in varying formats (i.e. Excel, text, or PDF), so it is challenging to entirely automate extraction algorithms. Usually, the data are presented in the form of a matrix, with DGRP lines in rows and phenotypes in columns. But sometimes, they can be in a more “exotic” format<sup>29</sup>, requiring a hands-on approach to format it appropriately. Also, the provided phenotyping data are often not sufficiently self-informative and thus require in-depth reading of the original manuscript to grasp abbreviations or identify the correct measurement units. These are important, in particular, to assure reproducibility, but especially when aggregating multiple studies together such that the scale of the values is similar. In DGRPpool, we therefore created a common matrix format to represent all studies, and we implemented a “Unit” metadata for each phenotype. Then, for each study, we mapped all phenotypes to their appropriate format and units (Supplement Figure S3). This part is fully accessible to the curator, who can update or add any phenotype that would be missing, with their corresponding units and meta-data description.

Another issue that we faced is that phenotypes are often averaged across multiple individual flies and that the authors only provide these “Summary datasets”. This can be problematic in terms of reproducibility, since some figures may show boxplots or distributions of values for each DGRP line, but these plots are not reproducible when only summary data is available (i.e. means or medians). Fortunately, some studies do provide “raw datasets” which contain multiple phenotypic values per DGRP line, often corresponding to replicate flies of the same genotype. These values tend to be of much greater interest since they enable statistical analyses and/or the computation of further summary statistics (not only mean or median, but also the standard error of means or other often non-provided summary values).

Finally, for some studies, phenotyping data were not or no longer available from the journal’s website<sup>30–32</sup>, which is often the journal’s responsibility. However, in all cases, we were able to contact the authors directly to recover the missing datasets.

To avoid such issues in the future, we have formulated a couple of good practice guidelines for authors to facilitate and improve upon our and future datasets with the aim of enabling harmonized and reproducible research. These guidelines are detailed in the Discussion section of this manuscript. All curated datasets in DGRP are formatted following these guidelines (where possible), and phenotypes can now be easily downloaded in a standard TSV format from a particular study, or from a phenotype page.

## How to leverage these datasets by correlating phenotypes

Our formatting and harmonizing of all datasets now enable interesting cross-phenotype analyses to generate new biological insights. One strategy to perform such analyses is to download a summary table that contains all the phenotypes in a common format and that is available from DGRP's front page. However, we deemed this still insufficient as a catalyzing resource, which is why we implemented tools to correlate existing and user-submitted phenotypes with all the other phenotypes in DGRP (Supp. Figure S4).

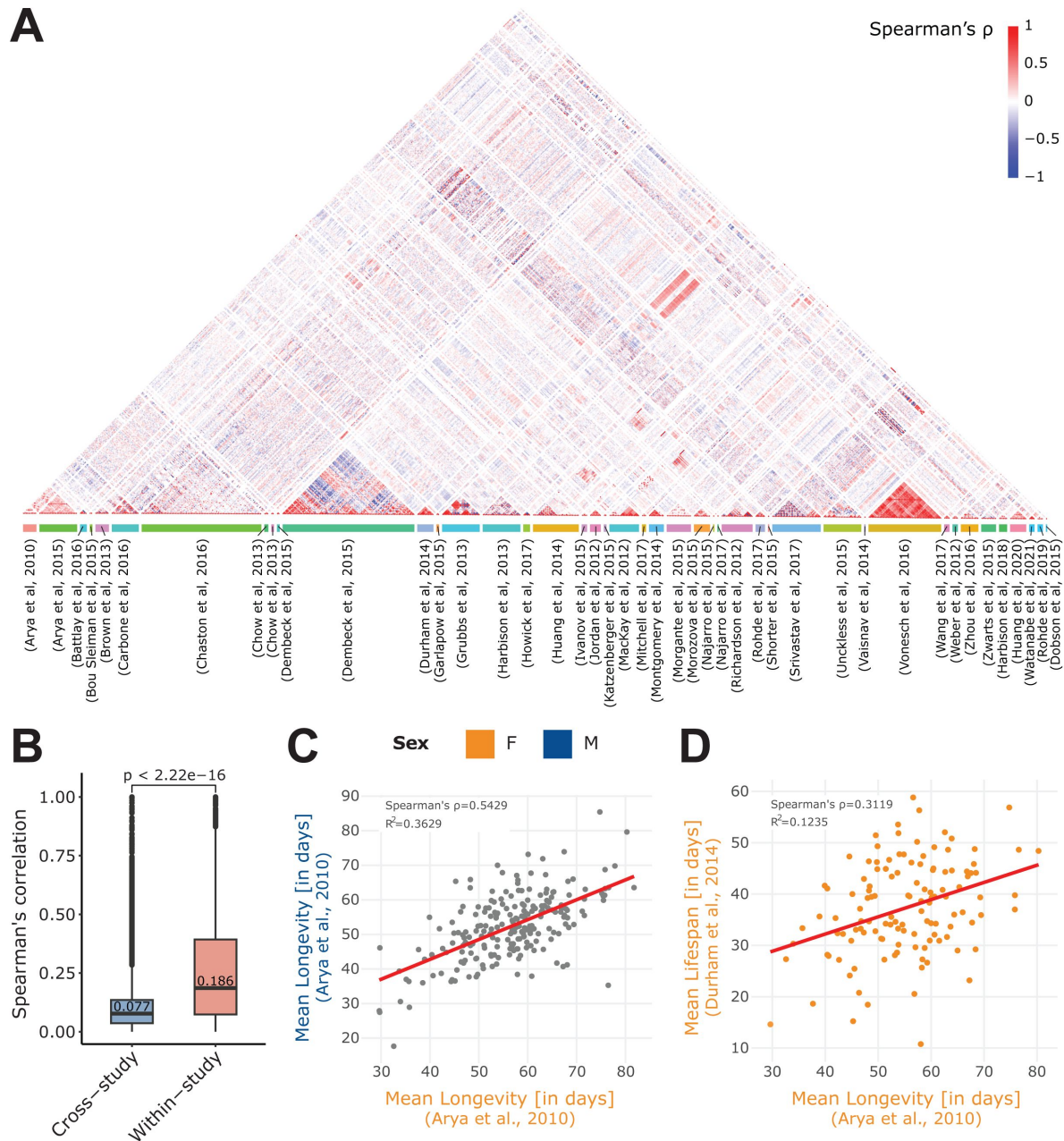
To better understand the structure of these phenotypes, and how they relate together, we also computed a global visualization of the phenotype correlations across all curated studies (**Figure 2A** [↗](#), Supp. Figure S5). Of note, by “phenotype correlations” we mean direct phenotype-phenotype correlations, i.e. a straightforward Spearman's correlation of two phenotypes between common DRGP lines, and we repeated this process for each pair of phenotypes. This revealed a clear trend, with phenotypes belonging to the same study (within-study) correlating in general stronger than those from different studies (**Figure 2B** [↗](#), Supp. Figure S6). This is expected since a given study will typically contain phenotypes that have been acquired for a given research topic, thus they will share similarities. Another potential factor that could explain this similarity is the well-known “batch effect”. Indeed, phenotypes acquired in the same environment (same lab, technician, reagents etc.) may sometimes show greater similarity than those acquired across different labs and conditions<sup>33</sup> [↗](#). The longevity phenotype however, assessed in at least six of the studies in DGRP<sup>31</sup> [↗](#),<sup>34</sup> [↗](#)–<sup>38</sup> [↗](#) across different laboratories, illustrates that phenotype and its measurements not only exhibits strong correlation across sexes (**Figure 2C** [↗](#)), but are also sufficiently robust between laboratories (**Figure 2D** [↗](#)). This example illustrates both the high robustness of results acquired in the context of DGRP studies (stable genotype, stable environment) and the robustness of the phenotype itself, which highlights its potential high heritability.

## Cross-study correlations highlight phenotype relationships

**Figure 2A** [↗](#) also highlights interesting cross-study correlations. For example, we can see a strong correlation between (Vonesh et al, 2016)<sup>28</sup> [↗](#) and (Grubbs et al, 2013)<sup>39</sup> [↗](#) which is perhaps expected since both studies examine fly morphology traits. The first one measures different organismal size traits such as eye interocular distance, or wing length, while the second studies leg and antenna development from imaginal discs, resulting in measuring phenotypes such as leg and bone length (**Figure 3A** [↗](#)). Similarly, three studies: (MacKay et al, 2012)<sup>1</sup> [↗](#), (Richardson et al, 2012)<sup>40</sup> [↗](#) and (Huang et al, 2014)<sup>2</sup> [↗](#) are expectedly correlated since all three investigate the influence of the *Wolbachia* endosymbiont. Another interesting correlation is between (Chow et al., 2013)<sup>41</sup> [↗](#) and (Durham et al., 2014)<sup>31</sup> [↗](#) which both studied fecundity and yield a cross-study correlation between remating proportion (Chow et al., 2013)<sup>41</sup> [↗](#) vs. mean fecundity (Durham et al., 2014)<sup>31</sup> [↗](#) (**Figure 3B** [↗](#)). While potentially conceptually obvious, this correlation suggests that females that are more likely to mate with multiple males tend to also produce a greater number of eggs.

These examples were all generated using the DGRP phenotype correlation tool, supporting our notion that it can leverage cross-study comparisons of multiple phenotypes to unveil potentially new interesting phenotype interaction/associations. As a further proof of concept and given

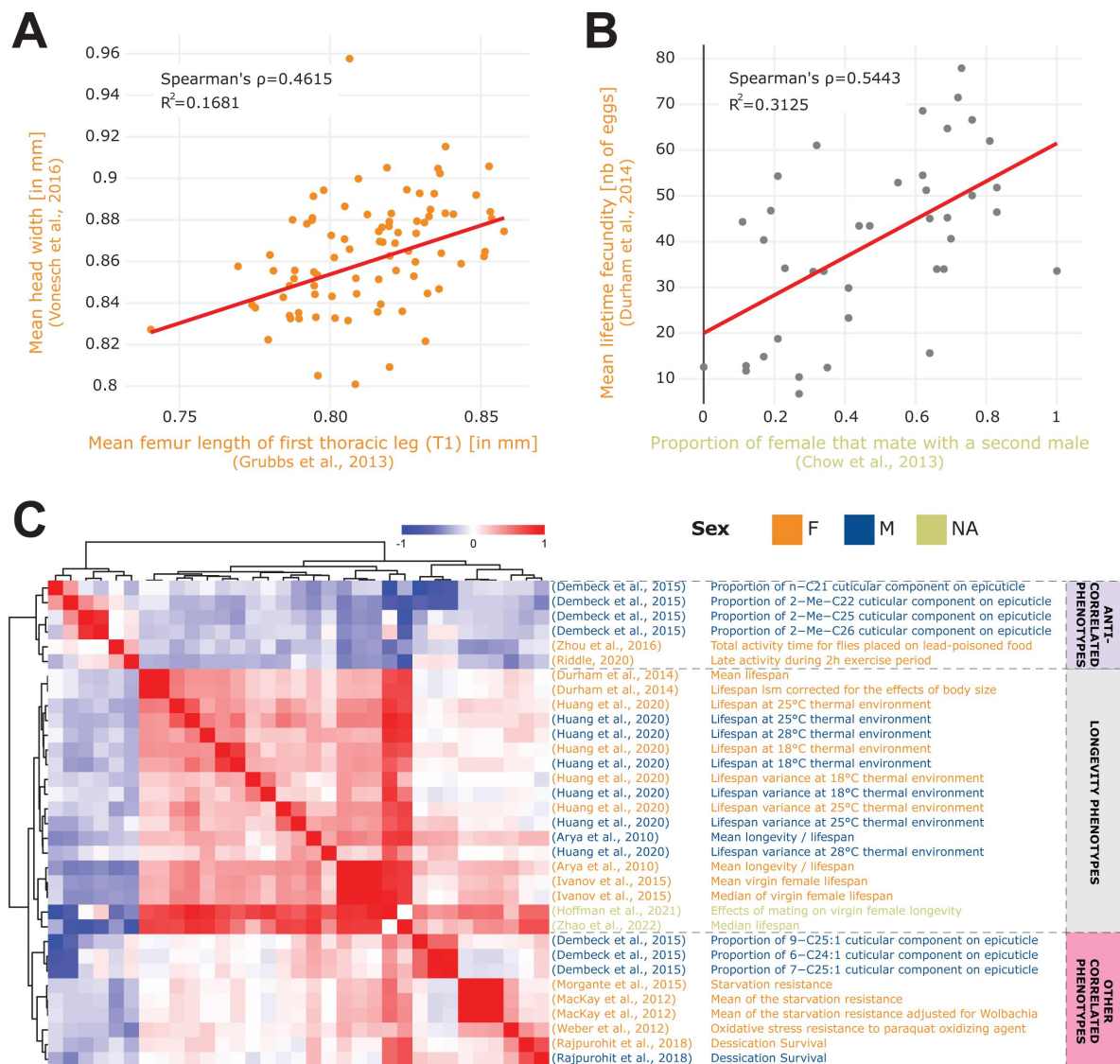




**Figure 2.**

### Within- and cross-study phenotype correlations.

A. Spearman's correlation of all phenotypes available in the 43 curated studies. Of note, we separately computed the phenotype correlations when data per sex were available (M, F or NA), and we restricted the computation to quantitative (non-categorical) phenotypes. Phenotypes are grouped by study (colored box at the bottom of the plot). B. Absolute value of the Spearman's correlation of pairs of phenotypes that originated from the same study (within-study) and those that originated from two different studies (cross-study). Of note, displayed values are median. Mean values are 0.099 for cross-study, and 0.259 for within-study. Again, we restricted the calculation to the 43 curated studies. C. Correlation of two longevity phenotypes from the same study (Arya, et al, 2010)<sup>34</sup>, revealing a strong correlation between Female (F) and Male (M) longevity. D. Correlation of two phenotypes from different studies: mean lifespan (Durham et al, 2014)<sup>31,32,33</sup> and mean longevity (Arya et al, 2010)<sup>34</sup>. Of note, both the C and D plots were generated using the "phenotype correlation" tool in DGRPpool.



**Figure 3.**

### Phenotype correlations contribute new biological insights.

**A.** Correlation of mean femur length (Grubbs et al., 2013)<sup>39</sup> vs. mean head width (Vonesch et al., 2016)<sup>28</sup> showing the significant cross-study association of organismal size traits. **B.** Correlation of remating proportion (Chow et al., 2013)<sup>41</sup> vs. mean fecundity (Durham et al., 2014)<sup>31</sup>. **C.** The 33 phenotypes correlated with longevity (Arya et al., 2010)<sup>34</sup> at a 5% FDR threshold (Spearman's correlation), revealing three main groups of phenotypes: lifespan phenotypes (middle rows), other correlated phenotypes (bottom rows) and anti-correlated phenotypes (top rows). Of note, both the A and B plots were generated using the "phenotype correlation" tool in DGRPpool.

society's strong interest in defining "healthy aging" determinants<sup>42</sup>, we continued investigating the "mean longevity" phenotype from (Arya et al, 2010)<sup>34</sup> and we studied the 33 phenotypes that were significantly correlated with it at 5% FDR threshold (**Figure 3C**). The hierarchical clustering clearly separated the phenotypes into three clusters: longevity-like phenotypes (strongly correlated together), other longevity-associated phenotypes (correlated with longevity), and phenotypes that seem antagonistic to longevity (anti-correlated phenotypes). Among the phenotypes that positively correlated with longevity, some may be expected such as starvation resistance<sup>14,16,17</sup> and oxidative stress resistance<sup>11</sup> but some are less intuitive such as desiccation survival<sup>14,15</sup>, and certain cuticular components of the epicuticle<sup>43</sup>. We further investigated the results for an alternative Pearson's correlation test, at 25% FDR threshold (Supp. Figure S7), and recapitulated most of these findings, with the addition of sleep duration<sup>12,13</sup>, whose relationship to longevity is complex and still not fully understood<sup>44</sup>. Although we cannot exclude spurious correlations, some of these more surprising correlations appear biologically highly interesting, illustrating the capacity of DGRPpool to unveil new research avenues that seem worth exploring in greater molecular detail. Also of interest is the group of often unexpected phenotypes that significantly anti-correlates with longevity. These include locomotor activity<sup>18</sup>, and some other cuticular components of the epicuticle<sup>45</sup>, suggesting that higher locomotor activity is linked to reduced longevity. At 25% FDR (Supp. Figure S7), we can also see an anti-correlation with food intake<sup>19,20</sup>, which potentially recapitulates previous results stating that the limitation of food intake increases an organismal lifespan<sup>46</sup>. Whether these are direct or indirect links remains unanswered, but appears worthy for a more in-depth scrutiny that is beyond the scope of this paper.

Inversely, our analyses also revealed that some expected phenotype correlations could not be detected. For example, in the context of metabolic energy expenditure<sup>47</sup>, it might seem intuitive that higher activity<sup>48</sup> would lead to greater food intake<sup>49</sup>. However, we did not observe such a correlation. Similarly, higher activity levels may reflect increased mating behaviour<sup>41</sup>, but this was also not observed. These are just a few examples of several cases where expected correlations did not materialize, collectively signifying that the genetic architecture underlying such traits appears inherently complex.

These proof-of-concept examples demonstrate in our opinion the utility of the DGRP lines and by extension DGRPpool to serve as powerful tools that will facilitate the identification of non-intuitive phenotype correlations and their underlying molecular basis as well as the discovery of putative genotype to phenotype relationships, as detailed below.

## From phenotypes to associated genotypes

The goal of most DGRP phenotyping studies is to eventually be able to link the phenotypes to potentially causal variants or sets of variants<sup>50</sup>. In response, tools like DGRP2 GWAS (<http://dgrp2.gnets.ncsu.edu/>)<sup>1,2</sup> have been put in place to accommodate geno-phenotype relationship analyses.

With the goal of providing an integrative analytical environment, we therefore also implemented GWAS tools within DGRPpool ([https://dgrpools.epfl.ch/check\\_pheno](https://dgrpools.epfl.ch/check_pheno)), aiming to assist researchers with performing GWAS analyses and interpreting the respective output. Specifically, we precalculated GWAS analyzes using PLINK2 on every existing phenotype in DGRPpool (see **Methods**), thereby considering all ~4M available DGRP variants while correcting for six known covariates (*Wolbachia* status, and five major insertions)<sup>2</sup>. Consequently, users can browse the GWAS results from any phenotype page on DGRPpool (Supp. Figure S8). These comprise a QQplot, for assessing the validity of the results, or potentially over-estimated p-values, and a Manhattan plot, for visualizing the significant loci across the *D. melanogaster* genome. It also displays a table with the top 1000 associated variants and allows the user to download the table of all significant hits, at a p-value<0.001 threshold. The tool further runs a gene set enrichment analysis of the results filtered at p<0.001 to enrich the associated genes to gene ontology terms, and Flybase

phenotypes. We also provide an ANOVA and a Kruskal-Wallis test between the phenotype and the six known covariates to uncover potential confounder effects (prior correction), which is displayed as a “warning” table to inform the user about potential associations of the phenotype and any of the six known covariates. It is important to note that these ANOVA and Kruskal tests are conducted for informational purposes only, to assess potential associations between well-established inversions or symbiont infection status and the phenotype of interest. However, all known covariates are included in the model regardless, and PLINK2 will automatically correct for them, irrespective of the results from the ANOVA or Kruskal tests.

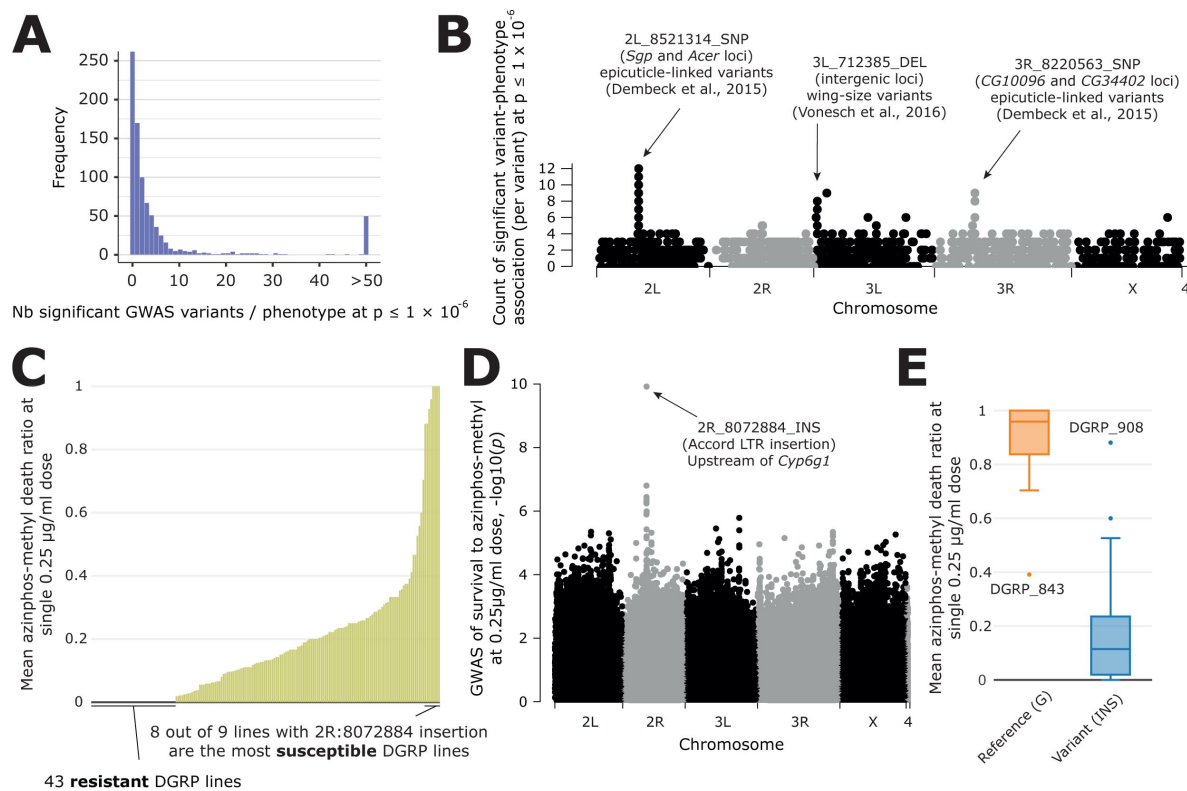
The interface also allows plotting an independent boxplot for each variant to visualize the effect of each allele on the phenotype. Importantly, for each variant, we also implemented a PheWAS button to visualize the effect of a particular variant across all existing phenotypes in DGRP. We also annotated all the variants for impact (non-synonymous effects, stop-codon gain, etc.) and for potential regulatory effect (transcription factor binding motif disruption), which should assist researchers with prioritizing the variants in terms of potential consequences. For all of these variants, we provide links to their description in Flybase<sup>4</sup>.

As mentioned, these GWAS results are available for each existing phenotype in DGRP, directly from the phenotype’s page. But users can also submit their own phenotype files (through the ‘Tool’ menu in the header) and visualize the same information for their own phenotypes. The GWAS analysis runs in the backend and takes about 1-2 minutes before displaying the results. This is implemented using a queuing system which prevents overloading the server in case of a peak of users or requests.

After having run GWAS on all phenotypes in DGRP, we observed the distribution of the number of significant variants per phenotype at  $p \leq 1 \times 10^{-5}$  threshold, which is an often-used arbitrary threshold for GWAS analyses in DGRP studies (Supp. Figure S9). This threshold yields on average 118 significant hits per tested phenotype, which is skewed due to some phenotypes leveraging lots of results (median = 20). Interestingly, from the shape of the distribution, it seems that this threshold may not be sufficiently stringent for avoiding some false positives in the results. This may be specific to the GWAS parameters we used (*--quantile-normalize --variance-standardize*), which already provide a more stringent list of hits than without normalizing the phenotypes. However, in this configuration, the  $p \leq 1 \times 10^{-6}$  threshold appears to be a good alternative for avoiding an excessive number of false positives (Figure 4A), as evident from the distribution of GWAS results across different thresholds (Supp. Figure S9). Another very often used threshold, is the Bonferroni one, which is much more stringent and varies from  $p \leq 1.126 \times 10^{-8}$  (if considering all 4M variants) to  $p \leq 2.67 \times 10^{-8}$  (if removing variants with low MAF or high number of missing values). In our results, the Bonferroni threshold ( $p \leq 2.67 \times 10^{-8}$ ) yielded 20 significant hits on average (median = 0, Supp. Figure S9) which could be limiting for many studies as it may mask potentially interesting variants that, while minimally contributing on an individual basis, may collectively point to implicated pathways or biological processes<sup>51</sup>. Thus, while choosing an optimal threshold is in general challenging, our results indicate that any threshold below or equal to  $1 \times 10^{-6}$  is reasonable given that at this threshold, the *p*-values appear not overestimated. We also verified if any variant is over-selected across all phenotypes to uncover a possible bias in our studies (Figure 4B). We found that only three loci are clearly overrepresented, corresponding to two studies with many correlated phenotypes (Dembeck et al, 2015 and Vonesch et al, 2016)<sup>27,28</sup>. Both studies were already prominent in Figure 2A, both in terms of their number of phenotypes and high correlation. One study comprises 66 cuticular hydrocarbon composition traits<sup>27</sup> and the other involves 27 organismal size traits<sup>28</sup>. These results are thus expected and do not indicate an overestimation of certain variants.

As a proof-of-concept and a validation of our approach, we compared our results with a randomly selected study that identified several variants associated with survival to azinphos-methyl at different doses (0.25, 0.5, 1, and 2 µg/ml)<sup>30</sup>. Of note, this study is available in DGRP under





**Figure 4.**

### Overview of GWAS results across phenotypes and one case study.

**A.** Distribution of the number of significant variants after a GWAS, for each phenotype available in DGRP. Of note, all values >50 have been set to 50, for easier visualization. **B.** For each variant, we plotted the number of times it was significantly associated with a phenotype (y-axis = number of occurrences). It is worth noting that we chose a Manhattan plot for representing this information, but this is not a “real” GWAS Manhattan plot. **C.** Case study on survival to azinphos-methyl exposure (Battlay et al., 2016)<sup>30</sup>, here to a 0.25 µg/ml dose. This plot was extracted from the phenotype’s page on DGRP at <https://dgrp.epfl.ch/phenotypes/20>. **D.** Manhattan plot (taken from DGRP’s result page [https://dgrp.epfl.ch/phenotypes/20/gwas\\_analysis](https://dgrp.epfl.ch/phenotypes/20/gwas_analysis)) showing the association of variants to “survival at 0.25 µg/ml dose” phenotype. **E.** Boxplot (taken from DGRP’s result page [https://dgrp.epfl.ch/phenotypes/20/gwas\\_analysis](https://dgrp.epfl.ch/phenotypes/20/gwas_analysis)), showing the effect of the top variant, 2R:8072884, which is a long insertion.

<https://dgrpools.epfl.ch/studies/3>. In particular, this study showed that survival to azinphos-methyl is highly variable among DGRP lines, even at a “low” 0.25 µg/ml dose. Importantly, the results of this study are reproduced in DGRPools as can be observed on the respective phenotype’s page (<https://dgrpools.epfl.ch/phenotypes/20>, **Figure 4C**). For example, DGRPools’s GWAS results are very similar to those of the study ([https://dgrpools.epfl.ch/phenotypes/20/gwas\\_analysis](https://dgrpools.epfl.ch/phenotypes/20/gwas_analysis), **Figure 4D**) and show a strong association at a 2R locus. Interestingly, the top variant we found, 2R:8072884 ( $p = 1.966 \times 10^{-26}$ ), a 509bp insertion polymorphism, is the *Accord* LTR insertion. It is annotated as located upstream of the *Cyp6g1* gene and has a high likelihood to be the main causal gene<sup>52,53</sup>. As described in the author’s Ph.D. thesis<sup>54</sup>, the minor allele at this variant — which corresponds to NOT having the insertion — correctly genotypes eight out of nine susceptible DGRP lines that are homozygous for the ancestral *Cyp6g1*<sup>M</sup> arrangement at this locus (DGRP lines 091, 486, 642, 776, 802, 821, **843**, 852, and 857). The presence of the *Accord* LTR insertion is associated with increased resistance to organophosphates, suggesting that derived alleles of *Cyp6g1* confer organophosphate resistance in the DGRP (**Figure 4E**).

These results show that DGRPools is able to accurately reproduce results from existing studies, and that new biological findings can be leveraged from its interactive results and plots. Revisiting the same organophosphate study<sup>30</sup>, the PheWAS page present in the GWAS results shows that this top variant is not only significant at other doses, but that it is also significant in the context of another study investigating *Drosophila* microbiota<sup>26</sup>. This could help with fine-tuning putative causal variants, but also with uncovering potential associations between certain phenotypes that in turn could enable studies aimed at providing underlying genetic and molecular mechanisms.

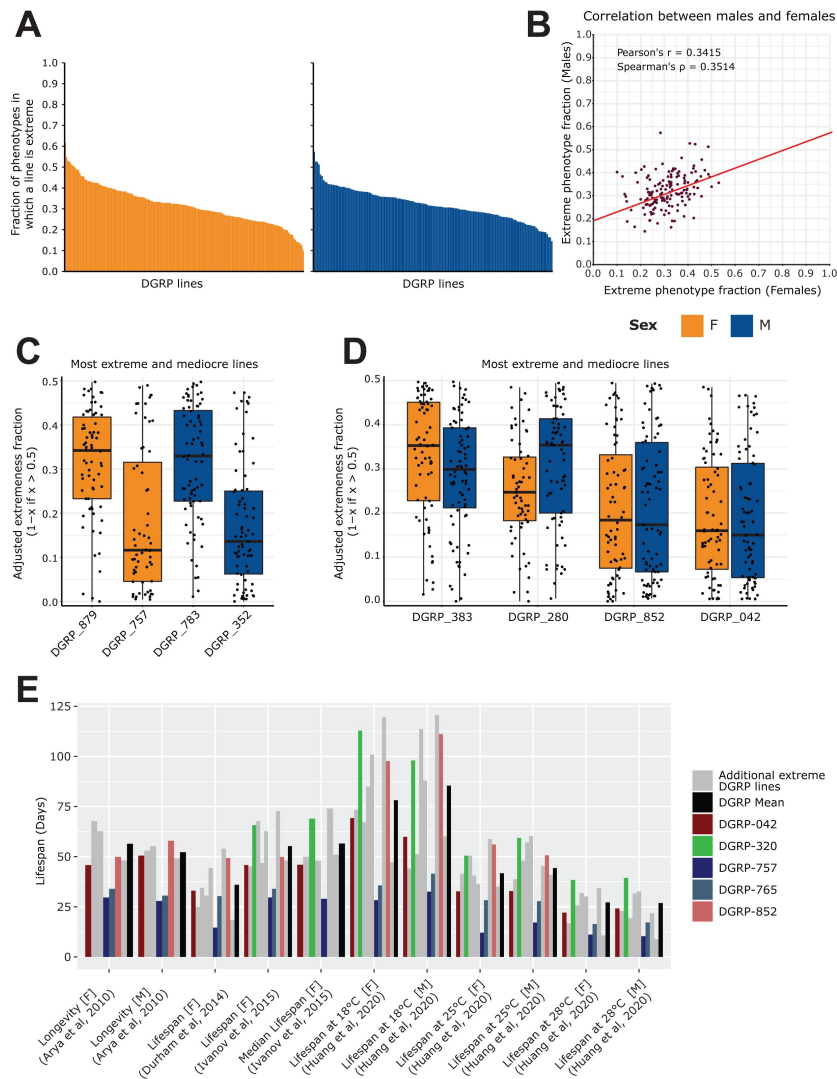
## Extreme phenotypes

After having collected and harmonized thousands of DGRP phenotypes, we investigated if we could identify outliers amongst DGRP lines that would potentially bias phenotypic associations. Indeed, if a particular DGRP line is repeatedly ranked in the extreme of all phenotypes, it could be that there are unknown cofactors that make the line “weaker” in general, or inversely. Although it is difficult to judge what phenotype is particularly advantageous or disadvantageous due to the presence of potential trade-offs<sup>55,56</sup>, we can determine how often a DGRP line is in the top or bottom 15% of a given phenotype. By focusing on phenotypes that are likely impacting overall viability, we ranked DGRP lines for each associated phenotype. Upon ranking the DGRP lines, we calculated whether the rank falls within the top or bottom 15% performers of the phenotype. We then assessed for each DGRP line how often they are ‘extreme’ and divided this by the total number of phenotypes in which the DGRP line has been included to obtain a “fraction of extremeness” (FoE). Finally, we filtered for lines with at least 50 available phenotypic measures to ensure that our values were not driven by a low number of observations (**Figure 5A**). Overall, we observed a modest correlation in the FoE across the sexes (**Figure 5B**, Spearman’s  $\rho = 0.3514$ ,  $p < 1.57 \times 10^{-5}$ ). While this suggests that extremeness is a population-wide feature, the correlation is not sufficiently strong to conclude that DGRP lines are generally extreme in both sexes, which may only apply to specific lines.

Upon considering individual DGRP lines, we can observe to what extent they are extreme for each individual phenotype. In **Figure 5C**, we show the most extreme and “moderate” (i.e. least distinctive) DGRP lines for each sex using an adjusted FoE for plotting purposes in which lower scores represent DGRP lines with a high FoE. While females of DGRP\_879 and males of DGRP\_783 tend to be extreme in some cases, for the majority of phenotypes they are considered moderate. Conversely, females of DGRP\_757 and males of DGRP\_352 are more likely to be labeled as extreme.

These examples only represent extremeness for individual DGRP lines of a given sex, however, their counterpart may not be as extreme or moderate. We therefore also looked for DGRP lines which can be considered extreme in both females and males and are potentially more extreme on a population-wide basis. **Figure 5D** describes such populations where the overall FoE between males and females differed on average at most 0.05. In these cases, DGRP\_852 and DGRP\_042 are





**Figure 5.**

### Analysis of extremeness among DGRP lines across 40 phenotypes.

**A.** Fraction of extremeness of a given DGRP line. DGRP lines are assigned as 'extreme' in a phenotype when they are in the top or bottom 15% of the phenotypic spectrum. Phenotypes were selected based on the curated studies which had the following categories assigned to them: Life history traits, Immunity, Toxicity, Resistance, Fecundity, Aging. DGRP lines were included if they had at least 50 phenotypic measures. **B.** Scatter plot for the fraction of extremeness of DGRP lines. On the x-axis, the fraction of extremeness is plotted for females, whereas males are plotted on the y-axis. **C.** Most extreme and moderate DGRP lines per sex. On the x-axis, the adjusted fraction of extremeness is provided. Individual fractions of extremeness per phenotype were retrieved for each DGRP line. The fraction was adjusted by 1 minus the fraction of extremeness if the fraction of extremeness was above 0.5. Because extremeness can range from 0 to 0.15 or 0.85 to 1, we adjusted the fraction of extremeness for plotting purposes. DGRP lines with a low adjusted fraction of extremeness are therefore more extreme, whereas a high adjusted fraction of extremeness is representative of more moderate DGRP lines. **D.** Extreme and moderate DGRP line pairings. On the x-axis, the adjusted fraction of extremeness is provided. Extreme and moderate line pairings were retrieved by searching for DGRP lines for which the fraction of extremeness between females and males was not greater than 0.05 while still having the highest and lowest average fraction of extremeness (across sex). **E.** Looking at phenotypes from [Figure 2D](#) marked as longevity/lifespan, for DGRP lines which are in the top 5 of fraction of extremeness for each respective sex, including DGRP\_852 and DGRP\_042 (red shades) from 5D. We specifically highlight DGRP\_757, DGRP\_765 in blue shades to show that they are across multiple studies in the lower end of the lifespan as is expected given that the lifespan trait is robust across studies. Similarly, DGRP\_320 shows a trend in which it displays above average lifespan. Other extreme DGRP lines which were in each respective top 5 are displayed in gray.

more likely to be extreme across sexes, which may be attributed to at least two factors. First, this may indicate that the population is generally not healthy if they consistently display a low lifespan, or second, and conversely, well-documented trade-offs of life history traits such as lifespan vs fecundity may be strongly at play here. The former does not however seem to be the case, as shown in **Figure 5E**. Both DGRP\_852 and DGRP\_042 generally display lifespan values around the mean lifespan of all DGRP lines, suggesting that they are more likely extreme for other phenotypes and are thus not by definition weak lines. However, DGRP\_757 and DGRP\_765 consistently display lower longevity in lifespan studies. These lines may therefore on the one hand be of particular interest for those studying life history traits in an evolutionary context, even though we did not observe strong lifespan and fecundity trade-offs across our phenotype dataset. On the other hand though, it may be advisable not to include DGRP\_757 and DGRP\_765 when studying the genetic basis of these complex traits as their outlier status may not reflect common genetic principles.

In conclusion, this analysis showed that while certain lines exhibit lower longevity or outlier behavior for specific traits, we found no evidence of a general pattern of extremeness across all traits. Therefore, the data do not support the idea of 'super normals' or any other inherently biased lines that could significantly affect genetic studies.

## Discussion

There are many studies across organisms where collated phenotyping data has led to novel insights<sup>57,58</sup>. Even though the *Drosophila* Genetic Reference Panel was formally released more than ten years ago, the resulting phenotype data of over 100 studies has so far not been combined into a single accessible resource. We anticipate that providing wider access to this data, as driven by FAIR principles<sup>5</sup>, will therefore facilitate our general understanding of the relationship between genotypes and phenotypes.

We have previously shown that using a subset of this resource effectively enabled us to establish a relationship between mitochondrial haplotypes and feeding behavior which we experimentally validated<sup>59</sup>. Next to our own study, other studies have used a similar approach and compared their results to already published phenotypes. For example, Wang et al.<sup>60</sup> studied the resistance and tolerance of DGRP flies to the fungal pathogen *Metarhizium anisopliae* (Ma549) and found that the host's defense to Ma549 is correlated with its defense to the bacterium *Pseudomonas aeruginosa* (Pa14). But they also compared this result to several previously published DGRP phenotypes including oxidative stress sensitivity<sup>61</sup>, aggression<sup>62</sup>, nutritional scores<sup>63</sup>, sleep indices<sup>48</sup>, and others. Similarly, Zwarts et al.<sup>64</sup> studied the size of the cerebral cortex and the mushroom bodies (MB). They showed that these phenotypes are correlated with phenotypes from other studies like aggression<sup>65</sup> and sleep<sup>48</sup>. Therefore, we believe that DGRPpool will either aid with validating the findings of a given study (i.e. higher bacterial resistance linked to overall resistance phenotypes) or by placing a study's phenotype data into a wider context (for example, linking brain size to behavioral phenotypes). Moreover, having access to multiple studies studying similar phenotypes can also be of help for meta-analyses and increased statistical power. In the case of longevity for example, there are multiple studies that aggregated this phenotype, across similar or complementary DGRP lines. Therefore, one could conduct a meta-GWAS analysis<sup>66</sup> by leveraging the replicates or combining the different lines into a single dataset. This tends to be a challenging process given the need for data harmonization and curation, which is exactly what we aimed to address by establishing DGRPpool. Of course, since similar DGRP lines across laboratories still have the same genotype, they should not be treated as biological replicates, but phenotypes could be averaged across similar lines, which would reduce hidden covariates such as laboratory adaptation or batch effects. Moreover, complementary lines can be used to enhance power and potentially find more small-effect associations. Indeed, researchers are increasingly advocating for collaboration and joining efforts to combine resources<sup>67</sup> to enable more accurate, and reproducible results.

We recognize certain limitations of the current web tool, particularly the lack of eQTL or gene expression data integration. Properly integrating DGRP GWAS results with gene expression data in a fair and robust manner would require uniform processing of multiple public datasets, necessitating the cataloging and standardization of all available datasets through a consistent pipeline. Moreover, incorporating a “cell type” or “tissue” layer would be essential, as gene expression data from whole flies is not directly comparable to data from specific tissues or even specific conditions. Since phenotypes are often tissue-dependent, this information is vital. However, implementing these layers presented too big of a challenge and was beyond the scope of this paper.

Our data collection and harmonization efforts have already enabled us to conduct some interesting cross-study analyses, including an investigation into the presence of biases stemming from outlier DGRP lines. Our “extremeness” analysis revealed that caution is warranted when selecting DGRP lines for specific studies, because, while some DGRP lines may be situated at the outer edge of the phenotypic spectrum by chance, DGRP\_757 and DGRP\_765 generally display lower lifespans in longevity studies. It is important to note that a shorter lifespan does not necessarily imply lower viability, as populations can still be propagated healthily. However, a shorter lifespan may also result from an impaired development<sup>58</sup> or developmental environment, which may confound the study of healthy aging<sup>59</sup>. Consequently, researchers should consider excluding these extreme lines from their experimental designs to prevent loss of power or potential covariate biases. Furthermore, and beyond our current focus on DGRP lines, we may in the future also consider adding standard *D. melanogaster* lines such as w1118, YWB, YWN or ORB to DGRPpool. This is because such lines have often been included as controls in DGRP studies<sup>38</sup>, and for most of these, genomic information is also available.

Finally, in order to sustain the value of the DGRP as a resource and to promote more findings, we provide the following guidelines for future DGRP phenotyping studies:

- • When available, report the raw datasets with values per fly. Optionally, but only in addition, the summary datasets can be provided, with values averaged across flies.
- • Provide the data as a separate Excel or text file (TSV/CSV) in the form of a matrix, with DGRP lines in rows and phenotypes in columns. Avoid reporting the values in the form of a PDF or an image, because it complicates data extraction afterward.
- • Clearly define the abbreviations in the tables and the units used for all phenotypes, so that the phenotyping dataset is self-explanatory and does not require an extended search in the main manuscript.
- • Report all DGRP lines in the first column of the phenotyping file, and the corresponding sex in the second column (M, F, or NA), before all phenotypes. Be careful to use the same format for all DGRP lines (e.g. DGRP\_XXX).
- • Pick a common format for all NA values. Whether reporting NA, or as an empty cell. But avoid mixing different formats.

In conclusion, we propose that DGRPpool has two primary purposes within the *Drosophila* community and beyond. First, it can be used to evaluate potential associations between phenotypes and contribute to understanding the genetic architecture underlying complex traits. Second, it can serve as a catalyst for further research and inform broader validation experiments, as exemplified in our previous work<sup>59</sup>. In the latter study, the validation of our hypothesis would not have been feasible without a harmonized dataset of phenotype data, as the connection between mitochondrial haplotypes and food intake would have remained theoretical. To maximize the benefits of DGRPpool, it should therefore remain subject to all FAIR principles, which unfortunately are still too often only implemented in terms of “open” and “sharing.” In other words, when large amounts of data are made publicly available without systematic curation or

homogenization, data interoperability and reproducibility can be highly problematic. DGRP Pool is in this regard a crucial initial step towards making DGRP phenotyping data widely accessible and usable for the entire *Drosophila* research community.

## Methods

### Data availability

All phenotyping data aggregated in DGRP Pool can be downloaded in a common format on each phenotype page. In the “Download” section on the front page, we also provide four .tsv files containing 1) All studies and their metadata (authors, citation, ...), 2) All phenotypes and their metadata (name, description, unit, ...), 3) All DGRP lines and their metadata (name, bloomington accession, ...), and 4) a global file with all numerical phenotypes across all studies, formatted following our recommendations.

We also provide an API for programmatic access to the data hosted on our website. The API is described at this page: <https://dgrpools.epfl.ch/home/api>.

All codes used to produce the figures of this manuscript are available at this GitHub repository: <https://github.com/DeplanckeLab/DGRPools>. The website code is available as a Docker container at this GitHub repository: [https://github.com/DeplanckeLab/DGRPools\\_web](https://github.com/DeplanckeLab/DGRPools_web).

### Web application

The DGRP Pool web application is hosted on a virtual machine at EPFL. All compute-intensive calculations (i.e. GWAS) are performed on an HPC within EPFL and results are then moved to the virtual machine’s local storage. The back-end is implemented with Ruby-on-Rails (RoR) 7 and all data is stored in a PostgreSQL relational database. The front-end uses different JavaScript libraries and is set to enable interactive usage. For instance, the application implements bootstrap tooltips to display HTML texts within tooltips, plotly.js v2.16.1 to generate the scatter plots, bar plots and box plots, using *scattergl*, *bar* and *box* modes respectively, or JQuery autocomplete for phenotype search combined with a SOLR search engine running on the server side (used for the phenotype comparison tool).

### Semi-automated referencing of studies and/or phenotypes

To submit a new study, any user can submit a DOI from the front page. Then, all metadata associated with this study (authors, journal, date, ...) are automatically imported from the Crossref API<sup>70</sup>. When the study is created, it acquires the “Submitted” state, and administrators are notified. Then, a curator is assigned to the study and needs to manually verify all information. A specific curator page allows him/her to 1) edit the metadata, 2) edit the categories associated with the study, or 3) add/remove/modify the phenotyping data and edit their names/types/units.

Identifiers from GEO<sup>71</sup>, ArrayExpress<sup>72</sup>, or the Sequence Read Archive (SRA)<sup>73</sup> can be associated manually with any study, for example for referencing additional gene expression data that would be published along with the phenotyping data.

### GWAS

GWAS analyses (whether pre-calculated, or using the web tool) were computed using Plink2 v2.00a3LM (1 Jul 2021) with these parameters: “--glm hide-covar --quantile-normalize --variance-standardize --geno 0.2 --maf 0.01”. It runs on all available variants in the DGRP database which is using the dm3 assembly (4’438’427 variants: 3’963’420 SNPs, 293’363 deletions, 169’053 insertions and 12’591 MNPs). We corrected the model for six known covariates (*Wolbachia* status, and 5

major insertions) that were described in (Huang et al, 2014)<sup>2</sup>[2](#) and also used on the DGRP2 website. Of note, these known covariates can also be considered as phenotypes, and thus are also available as a separate, browsable study on DGRPpool (<https://dgrpools.epfl.ch/studies/17>). For each phenotype, we also provide (similar to the DGRP2 website), both a Kruskal-Wallis test and an ANOVA test to inform the user about the association between the phenotype and the six known covariates. We also provide the results of a Shapiro-Wilk normality test of the phenotype distribution. Of note, the Kruskal-Wallis test is used for a single factor (independent variables) at a time, unlike the ANOVA test which is handling multiple factors simultaneously (as it is performed in a multifactorial design).

## Extremeness

Fraction of extremeness was calculated for each phenotypic spectrum separately by ranking the values with ties being assigned the minimum rank. We then calculated a cut-off to assign ranks in the bottom or upper 15% of a phenotypic range. This rank cut-off was further rounded up to be more inclusive on either end (i.e. if the cut-off was 1.2 or 1.8, the cut-off would become 2). Phenotypes equal or lower than the cut-off were assigned -1, whereas phenotypes equal to the max rank minus the cutoff or higher were assigned 1. Remaining phenotypic values were assigned 0. DGRP lines with phenotypic values of either -1 or 1 were then considered extreme for a given phenotype.

To calculate the overall fraction of extremeness for each DGRP line, we counted the number of times a DGRP line was assigned -1 or 1 and divided this by the total number of phenotypes available for that particular DGRP line. For most of our analyses, we only included DGRP lines for which at least 50 phenotypes were available unless stated otherwise.

The adjusted fraction of extremeness was calculated by dividing the phenotypic ranking by the max rank of a given phenotype. Values were adjusted with 1 minus the value if the value was above 0.5 (e.g. if  $x = 0.91$ , the adjusted value is  $1 - 0.91 = 0.09$ ). Only adjusted fraction of extremeness values below 0.15 are therefore considered extreme. As no rounding was performed in this case, it is possible for DGRPs to be assigned -1 and labeled as extreme, even though the DGRP line may have a value of 0.167. Further analysis shows that this 'violation' only takes place for 1.1% (417 out of 36,753) of the observations. At a *per* DGRP view, this would amount to less than 1 per 50 phenotypes, the cut-off for the number of phenotypes which a line needs to adhere to in order to be included in our analysis.

## Acknowledgements

The authors gratefully acknowledge the help and suggestions from Nathan M. Fiorellino and Jasper Deplancke in the early stages of the development of this web tool. This work was funded by the Ecole Polytechnique Fédérale de Lausanne (EPFL) and SNSF Project Grant (#310030\_197082) to BD.

## Author contributions

RB and BD initiated the project. VG, RB and BD wrote the article. RR implemented the automatic pipeline to retrieve phenotypic data from articles. VG and ER curated the studies. FPAD designed and implemented the web application and its database. FPAD designed and set up the unified format to represent phenotype data. FPAD and VG implemented the different tools (GWAS, PheWAS, Correlation). VG tested the web application extensively. VG and RB performed supporting analyses (e.g. GWAS, extremeness analysis).

## Competing interests

The authors declare that they have no conflict of interest.



## References

1. Mackay T. F. C., et al. (2012) **The *Drosophila melanogaster* Genetic Reference Panel** *Nature* **482**:173–178
2. Huang W., et al. (2014) **Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines** *Genome Res* **24**:1193–1208
3. Bou Sleiman M. S., et al. (2015) **Genetic, molecular and physiological basis of variation in *Drosophila* gut immunocompetence** *Nat Commun* **6**
4. Gramates L. S., et al. (2022) **FlyBase: a guided tour of highlighted features** *Genetics* **220**
5. Wilkinson M. D., et al. (2016) **The FAIR Guiding Principles for scientific data management and stewardship** *Sci Data* **3**
6. Sollis E., et al. (2023) **The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource** *Nucleic Acids Research* **51**:D977–D985
7. Ghousaini M., et al. (2021) **Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics** *Nucleic Acids Research* **49**:D1311–D1320
8. Hemani G., et al. (2018) **The MR-Base platform supports systematic causal inference across the human phenome** *eLife* **7**
9. Kurki M. I., et al. (2023) **FinnGen provides genetic insights from a well-phenotyped isolated population** *Nature* **613**:508–518
10. Piper M. D. W., Partridge L (2018) ***Drosophila* as a model for ageing** *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1864**:2707–2717
11. Finkel T., Holbrook N. J (2000) **Oxidants, oxidative stress and the biology of ageing** *Nature* **408**:239–247
12. Bushey D., Hughes K. A., Tononi G., Cirelli C (2010) **Sleep, aging, and lifespan in *Drosophila*** *BMC Neurosci* **11**
13. Thompson J. B., Su O. O., Yang N., Bauer J. H (2020) **Sleep-length differences are associated with altered longevity in the fruit fly *Drosophila melanogaster*** *Biology Open* **9**
14. Rion S., Kawecki T. J (2007) **Evolutionary biology of starvation resistance: what we have learned from *Drosophila*: Starvation resistance in *Drosophila*** *Journal of Evolutionary Biology* **20**:1655–1664
15. Hoffmann A. A., Harshman L. G (1999) **Desiccation and starvation resistance in *Drosophila*: patterns of variation at the species, population and intrapopulation levels** *Heredity* **83**:637–643

16. Chippindale A. K., Chu T. J. F., Rose M. R (1996) **Complex Trade-Offs and the Evolution of Starvation Resistance in *Drosophila melanogaster*** *Evolution* **50**
17. Jang T., Lee K. P (2015) **The genetic basis for mating-induced sex differences in starvation resistance in *Drosophila melanogaster*** *Journal of Insect Physiology* **82**:56–65
18. Magwere T., et al. (2006) **Flight Activity, Mortality Rates, and Lipoxidative Damage in *Drosophila*** *The Journals of Gerontology: Series A* **61**:136–145
19. Lee K. P., et al. (2008) **Lifespan and reproduction in *Drosophila* : New insights from nutritional geometry** *Proc. Natl. Acad. Sci. U.S.A* **105**:2498–2503
20. Piper M. D. W., Partridge L (2007) **Dietary Restriction in *Drosophila*: Delayed Aging or Experimental Artefact?** *PLoS Genet* **3**
21. Arch M., Vidal M., Koiffman R., Melkie S. T., Cardona P.-J (2022) ***Drosophila melanogaster* as a model to study innate immune memory** *Front. Microbiol* **13**
22. Dissel S (2020) ***Drosophila* as a Model to Study the Relationship Between Sleep, Plasticity, and Memory** *Front. Physiol* **11**
23. Flatt T (2020) **Life-History Evolution and the Genetics of Fitness Components in *Drosophila melanogaster*** *Genetics* **214**:3–48
24. Harnish J. M., Link N., Yamamoto S (2021) ***Drosophila* as a Model for Infectious Diseases** *IJMS* **22**
25. O’Kane C. J., Hagan J. J. (2011) ***Drosophila* as a Model Organism for the Study of Neuropsychiatric Disorders** *Molecular and Functional Models in Neuropsychiatry* :37–60
26. Chaston J. M., Dobson A. J., Newell P. D., Douglas A. E (2016) **Host Genetic Control of the Microbiota Mediates the *Drosophila* Nutritional Phenotype** *Appl Environ Microbiol* **82**:671–679
27. Dembeck L. M., et al. (2015) **Genetic architecture of natural variation in cuticular hydrocarbon composition in *Drosophila melanogaster*** *eLife* **4**
28. Vonesch S. C., Lamparter D., Mackay T. F. C., Bergmann S., Hafen E (2016) **Genome- Wide Analysis Reveals Novel Regulators of Growth in *Drosophila melanogaster*** *PLoS Genet* **12**
29. Hope K. A., et al. (2019) **The *Drosophila* Gene *Sulfateless* Modulates Autism-Like Behaviors** *Front. Genet* **10**
30. Battlay P., Schmidt J. M., Fournier-Level A., Robin C (2016) **Genomic and Transcriptomic Associations Identify a New Insecticide Resistance Phenotype for the Selective Sweep at the *Cyp6g1* Locus of *Drosophila melanogaster*** *G3 Genes | Genomes | Genetics* **6**:2573–2581
31. Durham M. F., Magwire M. M., Stone E. A., Leips J (2014) **Genome-wide analysis in *Drosophila* reveals age-specific effects of SNPs on fitness traits** *Nat Commun* **5**

32. Najarro M. A., Hackett J. L., Macdonald S. J (2017) **Loci Contributing to Boric Acid Toxicity in Two Reference Populations of *Drosophila melanogaster*** *G3 Genes | Genomes | Genetics* **7**:1631–1641
33. Ackermann M., et al. (2001) **Effects of assay conditions in life history experiments with *Drosophila melanogaster*: Assay environment in life history experiments** *Journal of Evolutionary Biology* **14**:199–209
34. Arya G. H., et al. (2010) **Natural Variation, Functional Pleiotropy and Transcriptional Contexts of Odorant Binding Protein Genes in *Drosophila melanogaster*** *Genetics* **186**:1475–1485
35. Huang W., et al. (2020) **Context-dependent genetic architecture of *Drosophila* life span** *PLoS Biol* **18**
36. Ivanov D. K., et al. (2015) **Longevity GWAS Using the *Drosophila* Genetic Reference Panel** *GERONA* **70**:1470–1478
37. Zhao X., et al. (2022) **The metabolome as a biomarker of aging in *Drosophila melanogaster*** *Aging Cell* **21**
38. Hoffman J. M., Dudeck S. K., Patterson H. K., Austad S. N (2021) **Sex, mating and repeatability of *Drosophila melanogaster* longevity** *R. Soc. open sci* **8**
39. Grubbs N., et al. (2013) **New Components of *Drosophila* Leg Development Identified through Genome Wide Association Studies** *PLoS ONE* **8**
40. Richardson M. F., et al. (2012) **Population Genomics of the Wolbachia Endosymbiont in *Drosophila melanogaster*** *PLoS Genet* **8**
41. Chow C. Y., Wolfner M. F., Clark A. G (2013) **Large Neurological Component to Genetic Differences Underlying Biased Sperm Use in *Drosophila*** *Genetics* **193**:177–185
42. Friedman S. M (2020) **Lifestyle (Medicine) and Healthy Aging** *Clinics in Geriatric Medicine* **36**:645–653
43. Wang Z., et al. (2022) **Desiccation resistance differences in *Drosophila* species can be largely explained by variations in cuticular hydrocarbons** *eLife* **11**
44. Panel Consensus Conference, et al. (2015) **Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society on the Recommended Amount of Sleep for a Healthy Adult: Methodology and Discussion** *Sleep* **38**:1161–1183
45. Nghiem D., Gibbs A. G., Rose M. R., Bradley T. J (2000) **Postponed aging and desiccation resistance in *Drosophila melanogaster*** *Experimental Gerontology* **35**:957–969
46. McCracken A. W., Buckle E., Simons M. J. P (2020) **The relationship between longevity and diet is genotype dependent and sensitive to desiccation in *Drosophila melanogaster*** *Journal of Experimental Biology jeb* **230**185 <https://doi.org/10.1242/jeb.230185>
47. Chatterjee N., Perrimon N (2021) **What fuels the fly: Energy metabolism in *Drosophila* and its application to the study of obesity and diabetes** *Sci. Adv* **7**

48. Harbison S. T., McCoy L. J., Mackay T. F. (2013) **Genome-wide association study of sleep in *Drosophila melanogaster*** *BMC Genomics* **14**
49. Garlapow M. E., Huang W., Yarboro M. T., Peterson K. R., Mackay T. F. C. (2015) **Quantitative Genetics of Food Intake in *Drosophila melanogaster*** *PLoS ONE* **10**
50. Mackay T. F. C., Huang W. (2018) **Charting the genotype–phenotype map: lessons from the *Drosophila melanogaster* Genetic Reference Panel** *WIREs Dev Biol* **7**
51. Uffelmann E., et al. (2021) **Genome-wide association studies** *Nat Rev Methods Primers* **1**
52. Daborn P. J., et al. (2002) **A single P450 allele associated with insecticide resistance in *Drosophila*** *Science* **297**:2253–2256
53. Schmidt J. M., et al. (2010) **Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 locus** *PLoS genetics* **6**
54. Battlay P. (2019) **The quantitative genetics of insecticide resistance in *Drosophila melanogaster***
55. Zwaan B., Bijlsma R., Hoekstra R. F. (1995) **Direct selection on life span in *Drosophila Melanogaster*** *Evolution* **49**:649–659
56. Rose M., Charlesworth B. (1980) **A test of evolutionary theories of senescence** *Nature* **287**:141–142
57. Greene D., et al. (2023) **Genetic association analysis of 77,539 genomes reveals rare disease etiologies** *Nat Med* **29**:679–688
58. Doust C., et al. (2022) **Discovery of 42 genome-wide significant loci associated with dyslexia** *Nat Genet* **54**:1621–1629
59. Bevers R. P. J., et al. (2019) **Mitochondrial haplotypes affect metabolic phenotypes in the *Drosophila* Genetic Reference Panel** *Nat Metab* **1**:1226–1242
60. Wang J. B., Lu H.-L., St. Leger R. J. (2017) **The genetic basis for variation in resistance to infection in the *Drosophila melanogaster* genetic reference panel** *PLoS Pathog* **13**
61. Jordan K. W., et al. (2012) **Genome-Wide Association for Sensitivity to Chronic Oxidative Stress in *Drosophila melanogaster*** *PLoS ONE* **7**
62. Shorter J., et al. (2015) **Genetic architecture of natural variation in *Drosophila melanogaster* aggressive behavior** *Proc. Natl. Acad. Sci. U.S.A* **112**
63. Unckless R. L., Rottschaefer S. M., Lazzaro B. P. (2015) **A Genome-Wide Association Study for Nutritional Indices in *Drosophila*** *G3 Genes | Genomes | Genetics* **5**:417–425
64. Zwarts L., et al. (2015) **The genetic basis of natural variation in mushroom body size in *Drosophila melanogaster*** *Nat Commun* **6**
65. Zwarts L., et al. (2011) **Complex genetic architecture of *Drosophila* aggressive behavior** *Proc. Natl. Acad. Sci. U.S.A* **108**:17070–17075

66. Zeggini E., Ioannidis J. P (2009) **Meta-analysis in genome-wide association studies** *Pharmacogenomics* **10**:191–201
67. McCarthy M. I., et al. (2008) **Genome-wide association studies for complex traits: consensus, uncertainty and challenges** *Nat Rev Genet* **9**:356–369
68. May C. M., Doroszuk A., Zwaan B. J (2015) **The effect of developmental nutrition on life span and fecundity depends on the adult reproductive environment in *Drosophila melanogaster*** *Ecol Evol* **5**:1156–1168
69. Iliadi K. G., Knight D., Boulianne G. L (2012) **Healthy Aging – Insights from *Drosophila*** *Front. Physio* **3**
70. Hendricks G., Tkaczyk D., Lin J., Feeney P (2020) **Crossref: The sustainable source of community-owned scholarly metadata** *Quantitative Science Studies* **1**:414–427
71. Barrett T., et al. (2009) **NCBI GEO: archive for high-throughput functional genomic data** *Nucleic Acids Research* **37**:D885–D890
72. Brazma A (2003) **ArrayExpress--a public repository for microarray gene expression data at the EBI** *Nucleic Acids Research* **31**:68–71
73. Kodama Y., Shumway M., Leinonen R. (2012) **on behalf of the International Nucleotide Sequence Database Collaboration. The sequence read archive: explosive growth of sequencing data** *Nucleic Acids Research* **40**:D54–D56

## Editors

Reviewing Editor

**Jonathan Flint**

University of California, Los Angeles, Los Angeles, United States of America

Senior Editor

**David James**

University of Sydney, Sydney, Australia

### Reviewer #1 (Public Review):

This is a technically sound paper focused on a useful resource around the DRGP phenotypes which the authors have curated, pooled, and provided a user-friendly website. This is aimed to be a crowd-sourced resource for this in the future. The authors should make sure they coordinate as well as possible with the NC datasets and community and broader fly community.

<https://doi.org/10.7554/eLife.88981.2.sa2>

### Reviewer #2 (Public Review):

In the present study, Gardeux et al provide a web-based tool for curated association mapping results from DRP studies. The tool lets users view association results for phenotypes and compare mean phenotype ~ phenotype correlations between studies. In the manuscript, the authors provide several example utilities associated with this new resource, including pan-

study summary statistics for sex, traits, and loci. They highlight cross-trait correlations by comparing studies focused on longevity with phenotypes such as oxphos and activity. Strengths: -Considerable efforts were dedicated toward curating the many DRG studies provided. -Available tools to query large DRP studies are sparse and so new tools present appeal Weaknesses: The creation of a tool to query these studies for a more detailed understanding of physiologic outcomes seems underdeveloped. These could be improved by enabling usages such as more comprehensive queries of meta-analyses, molecular information to investigate given genes or pathways, and links to other information such as in mouse rat or human associations.

<https://doi.org/10.7554/eLife.88981.2.sa1>

### Author response:

The following is the authors' response to the original reviews.

We would like to thank the reviewers for their positive and constructive comments on the manuscript.

We committed in our original rebuttal letter to implement the following revisions to both DGRPpool and the corresponding manuscript to address the reviewers' comments:

(1) We agree with reviewer #1 that normalizing the data could potentially improve the GWAS results. Thus, for computing the GWAS results, we are now using these two additional options in PLINK2: "--quantile-normalize --variance-standardize". We assessed the impact of these options on the overall results, which revealed only minor improvements of the results, globally being a bit more stringent. In this direction, we also now filter the top results with a nominal p-value of 0.001 instead of 0.01, also because it provided better results for the new gene set enrichment step.

(2) We added a KRUSKAL test next to the ANOVA test for assessing the links between the phenotypes and the 6 known covariates, as well as a Shapiro-Wilk test of normality.

(3) We agree with both reviewers that gene expression information is of interest. As mentioned before, adding gene expression data to the portal would have required extensive work, beyond the current scope of this paper, which primarily focuses on phenotypes and genotype-phenotype associations. Nonetheless, we included more gene-level outlinks to Flybase. Additionally, we now link variants and genes to Flybase's online genome browser, JBrowse. By following the reviewers' suggestions, we aim to guide DGRPpool users to potentially informative genes.

(4) Consistent with the latter point, and in agreement with reviewer #2, we acknowledge that additional tools could enhance DGRPpool's functionality and facilitate meta-analyses for users. Therefore, we developed a gene-centric tool that now allows users to query the database based on gene names. Moreover, we integrated ortholog databases into the GWAS results. This feature will enable users to extend *Drosophila* gene associations to other species if necessary.

(5) We amended the manuscript to describe all the new tools and features that were developed and implemented. In short, the new features include a new gene-centric page with diverse links (Phenotypes, Genome Browser JBrowse, Orthologs ...), a variant-centric page (variant details, and PheWAS), an API for programmatic access to the database, and other statistical outputs and filtering options.

We will detail these advances in the point-by-point response below and in the revised manuscript.



**Reviewer #1 (Public Review):**

*This is a technically sound paper focused on a useful resource around the DRGP phenotypes which the authors have curated, pooled, and provided a user-friendly website. This is aimed to be a crowd-sourced resource for this in the future.*

*The authors should make sure they coordinate as well as possible with the NC datasets and community and broader fly community. It looks reasonable to me but I am not from that community.*

We thank the reviewer for the positive comments. We will leverage our connections to the fly and DGRP communities to make the resource as valuable as possible. DGRPpool in fact already reflects the input of many potential users and was also inspired by key tools on the DGRP2 website. Furthermore, it also rationalizes why we are bridging our results with other resources, such as linking out to Flybase, which is the main resource for the Drosophila community at large.

*I have only one major concern which in a more traditional review setting I would be flagging to the editor to insist the authors did on resubmission. I also have some scene setting and coordination suggestions and some minor textual / analysis considerations.*

*The major concern is that the authors do not comment on the distribution of the phenotypes; it is assumed it is a continuous metric and well-behaved - broad gaussian. This is likely to be more true of means and medians per line than individual measurements, but not guaranteed, and there could easily be categorical data in the future. The application of ANOVA tests (of the "covariates") is for example fragile for this.*

*The simplest recommendation is in the interface to ensure there is an inverse normalisation (rank and then project on a gaussian) function, and also to comment on this for the existing phenotypes in the analysis (presumably the authors are happy). An alternative is to offer a kruskal test (almost the same thing) on covariates, but note PLINK will also work most robustly on a normalised dataset.*

We thank the reviewer for raising this interesting point. Indeed, we did not comment on the distribution of individual phenotypes due to the underlying variability from one phenotype to another, as suggested by the reviewer. Some distributions appear normal, while others are clearly not normally distributed. This information is 'visible' to users by clicking on any phenotype; DGRPpool automatically displays its global distribution if the values are continuous/quantitative. Now, we also provide a Shapiro-Wilk test to assess the normality of the distribution.

We acknowledge the reviewer's concerns regarding the use of ANOVA tests. However, we want to point out that the ANOVA test is solely conducted to assess whether any of the well-established inversions or symbiont infection status (that, for simplification, we call "covariates" or "known covariates") are associated with the phenotype of interest. This is merely informational, to help the user understand if their phenotype of interest is associated with a known covariate. But all of these known covariates are put in the model in any case, so PLINK2 will automatically correct for them, whatever is the output of the ANOVA test.

Still, we amended the manuscript to better explain this, and we added a Kruskal-Wallis test (in addition to the ANOVA test) in the results, so the users can have a better overview of potentially associated known covariates. We added this text on p. 10 of the revised manuscript:

“The tool further runs a gene set enrichment analysis of the results filtered at  $p < 0.001$  to enrich the associated genes to gene ontology terms, and Flybase phenotypes. We also provide an ANOVA and a Kruskal-Wallis test between the phenotype and the six known covariates to uncover potential confounder effects (prior correction), which is displayed as a “warning” table to inform the user about potential associations of the phenotype and any of the six known covariates. It is important to note that these ANOVA and Kruskal tests are conducted for informational purposes only, to assess potential associations between well-established inversions or symbiont infection status and the phenotype of interest. However, all known covariates are included in the model regardless, and PLINK2 will automatically correct for them, irrespective of the results from the ANOVA or Kruskal tests. “

We also acknowledge in the manuscript (Methods section) that the Kruskal-Wallis test is used for a single factor (independent variables) at a time. This is unlike the ANOVA test that we initially performed, which was handling multiple factors simultaneously (given that it was performed in a multifactorial design). For a more direct comparison with our ANOVA model, we ran separate Kruskal-Wallis tests for each factor, but then we acknowledged its potential limitations compared to our multifactorial ANOVA, since each of these tests treats the factor in question as the only source of variation, not considering other factors. But since the test is not intended for interactions or combined effects of these factors, we deem it to be sufficient.

Nevertheless, we concur with the reviewer that normalizing the data could potentially enhance GWAS results. Consequently, we have rerun the GWAS analyses using the PLINK2 --quantile-normalize and --variance-standardize options. We have updated all results on the website and also updated the plots in the manuscript, accordingly.

*Minor points:*

*On the introduction, I think the authors would find the extensive set of human GWAS/PheWAS resources useful; widespread examples include the GWAS Catalog, Open Targets PheWAS, MR-base, and the FinnGen portal. The GWAS Catalog also has summary statistics submission guidelines, and I think where possible meta-data harmonisation should be similar (not a big thing). Of course, DRGP has a very different structure (line and individuals) and of course, raw data can be freely shown, so this is not a one-to-one mapping.*

Thank you for the suggestion. We cited these resources in the Introduction.

“This aligns with the harmonization effort undertaken by other human GWAS/PheWAS resources, such as the GWAS Catalog, Open Targets PheWAS, MR-base, and the FinnGen portal, which provide extensive examples of effective data use and accessibility. Although the structure of DGRP differs from these human databases, we acknowledge the importance of similar meta-data harmonization guidelines. Inspired by the GWAS Catalog's summary statistics submission guidelines, we propose submission guidelines for DGRP phenotyping data in this paper. “

*For some authors coming from a human genetics background, they will be interpreting correlations of phenotypes more in the genetic variant space (eg LD score regression), rather than a more straightforward correlation between DRGP lines of different individuals. I would encourage explaining this difference somewhere.*

We understand that this is a potential issue and we made the distinction clearer in the manuscript to avoid any confusion. We added this text on p.7, at the beginning of the correlation results section:

“Of note, by “phenotype correlations”, we mean direct phenotype-phenotype correlations, i.e. a straightforward Spearman’s correlation of two phenotypes between common DRGP lines, and we repeated this process for each pair of phenotypes. “

*This leads to an interesting point that the inbred nature of the DRGP allows for both traditional genetic approaches and leveraging the inbred replication; there is something about looking at phenotype correlations through both these lenses, but this is for another paper I suspect that this harmonised pool of data can help.*

We agree with the reviewer and hope that more meta-analyses will be made possible by leveraging the harmonized data that are made available through DGRPpool.

*I was surprised the authors did not crunch the number of transcript/gene expression phenotypes and have them in. Is this because this was better done in other datasets? Or too big and annoying on normalisation? I'd explain the rationale to leave these out.*

This is a very good point and is in fact something that we initially wanted to do. However, to render the analysis fair and robust, it would require processing all datasets in the same way. This implies cataloging all existing datasets and processing them through the same pipeline. In addition, it would require adding a “cell type” or “tissue” layer, because gene expression data from whole flies is obviously not directly comparable to gene expression data from specific tissues or even specific conditions. This would be key information as phenotypes are often tissue-dependent. Consequently, and as implied by the reviewer, we deemed this too big of a challenge beyond the scope of the current paper. Nevertheless, we plan to continue investigating this avenue in a potential follow-up paper.

We still added a gene-centric tool to be able to query the GWAS results by gene. We also added orthologs and Flybase gene-phenotype information, both in this new gene-centric tool and also in all GWAS results.

*I think 25% FDR is dangerously close to "random chance of being wrong". I'd just redo this section at a higher FDR, even if it makes the results less 'exciting'. This is not the point of the paper anyway.*

We agree with the reviewer that this threshold implies a higher risk of false positive results. However, this is not an uncommonly used threshold (Li et al., PLoS biology, 2008; Bevers et al., Nature Metabolism, 2019; Hwangbo et al, Elife, 2023), and one that seems robust enough in our analysis since similar phenotypes are significant in different studies at different FDR thresholds.

Nevertheless, we revisited these results with a stronger threshold of 5% FDR in the main Figure 3C. Most of the conclusions were maintained, except for the relation between longevity and “food intake”, as well as “sleep duration”. We modified the manuscript accordingly, notably removing these points from the abstract, and tuning down the results section. We kept the 25% FDR results as supplemental information.

*I didn't buy the extreme line piece as being informative. Something has to be on the top and bottom of the ranks; the phenotypes are an opportunity for collection and probably have known (as you show) and cryptic correlations. I think you don't need this section at all for the paper and worry it gives an idea of "super normals" or "true wild types" which ... I just don't think is helpful.*

We appreciate the reviewer’s feedback on the section regarding extreme DGRP lines and understand the concern about potential implications of “super normals” or “true wild types.”

This section aimed to explore whether specific DGRP lines consistently rank in the extremes of phenotypic measures, particularly those tied to viability-related traits. Our hypothesis was that if particular lines consistently appear at the top or bottom, this might suggest some inherent bias or inbreeding-related weakness that could influence genetic association studies.

However, as per the analyses presented, we did not discover support for this phenomenon. Importantly, the observed mild correlation in extremeness across sexes, while not profound, further suggested that this phenomenon is not a consistent population-wide feature.

Nevertheless, we consider that this message is still important to convey. In response to the reviewer's feedback, we have provided a clearer conclusion of this paper section by adding the following paragraph:

“In conclusion, this analysis showed that while certain lines exhibit lower longevity or outlier behavior for specific traits, we found no evidence of a general pattern of extremeness across all traits. Therefore, the data do not support the idea of 'super normals' or any other inherently biased lines that could significantly affect genetic studies. “

*I'd say "well-established inversion genotypes and symbiot levels" rather than generic covariates. Covariates could mean anything. You have specific "covariates" which might actually be the causal thing.*

We thank the author for the suggestion. We agree and modified the manuscript accordingly.

*I wouldn't use the adjective tedious about curation. It's a bit of a value judgement and probably places the role of curation in the wrong way. Time-consuming due to lack of standards and best practice?*

We thank the author for the suggestion. We agree and modified the manuscript accordingly, replacing the occurrences by “thorough” and “rigorous” which correspond better to the initial intended meaning.

#### **Reviewer #2 (Public Review):**

##### *Summary:*

*In the present study, Gardeux et al provide a web-based tool for curated association mapping results from DRP studies. The tool lets users view association results for phenotypes and compare mean phenotype ~ phenotype correlations between studies. In the manuscript, the authors provide several example utilities associated with this new resource, including pan-study summary statistics for sex, traits, and loci. They highlight cross-trait correlations by comparing studies focused on longevity with phenotypes such as oophos and activity.*

##### *Strengths:*

*-Considerable efforts were dedicated toward curating the many DRG studies provided.*

*-Available tools to query large DRP studies are sparse and so new tools present appeal*

##### *Weaknesses:*

*The creation of a tool to query these studies for a more detailed understanding of physiologic outcomes seems underdeveloped. These could be improved by enabling usages such as more comprehensive queries of meta-analyses, molecular information to investigate given genes or pathways, and links to other information such as in mouse rat or human associations.*

We appreciate the reviewer's kind comments.

Regarding the tools, we concur with the reviewer that incorporating additional tools could enhance DGRPtool and facilitate users in conducting meta-analyses. Therefore, we developed two new tools: a gene-centric tool that enables users to query the database based on gene names, and a variant-centric tool mostly for studying the impact of specific genomic loci on phenotypes. Additionally, in all GWAS results, we added links to ortholog databases, thereby allowing users to extend fly gene associations to other species, if required.

Furthermore, we added links to the Flybase database, for variants, phenotypes, and genes that are already present in Flybase. We also link out to a 'genome browser-like' view (Flybase's JBrowse tool) of the GWAS results centered around the affected variants/genes.

Finally, we now also perform a gene-set enrichment analysis for each GWAS result, both in the Flybase gene-phenotype database and the Gene Ontology (GO) database.

#### **Reviewer #2 (Recommendations For The Authors):**

*(1) The authors discuss how current available DRG databases are basically data-dump sites and there is a need for integrative queries. Clearly, they spent (and are spending) considerable efforts into curating associations from available studies so the current resource seems to contain several areas of missed opportunities. The most clear addition would be to integrate gene-level queries. For example which genes underlie associations to given traits, what other traits map to a specific gene, or multiple genes which map to traits. This absence of integration is somewhat surprising given the lab's previous analyses of eQTL data in DRPs (<https://doi.org/10.1371/journal.pgen.1003055>) and readily available additional data (ex. 10.1101/gr.257592.119 ,flybase) simple intersections between these at the locus level would provide much deeper molecular support for searching this database.*

The point raised by the reviewer concerning eQTL / transcriptomic data is in fact similar to the one raised by reviewer #1. We strongly agree with both reviewers that incorporating eQTL results in the tool would be very valuable, and this is in fact something that we initially wanted to do. However, to render the analysis fair and robust, it would require re-processing multiple public datasets in the same way. This would imply cataloging all existing datasets and processing them through the same pipeline. In addition, it would require adding a "cell type" or "tissue" layer, because gene expression data from whole flies is obviously not directly comparable to gene expression data from specific tissues or even specific conditions. This would be key information as phenotypes are often tissue-dependent. Consequently, we deemed implementing all these layers too big of a challenge beyond the scope of the current paper, but we plan to continue investigating this avenue in a potential follow-up paper.

As mentioned before, we still integrated gene-level queries in a new tool, querying genes in the context of GWAS results. We acknowledge that this is not directly related to gene expression, and thus not implicating eQTL datasets (at least for now), but we think that it is for now a good alternative, reinforcing the interpretation of the GWAS results.

Since this point was raised by both reviewers, we added a discussion about this in the manuscript.

"We recognize certain limitations of the current web tool, particularly the lack of eQTL or gene expression data integration. Properly integrating DGRP GWAS results with gene expression data in a fair and robust manner would require uniform processing of multiple public datasets, necessitating the cataloging and standardization of all available datasets through a consistent pipeline. Moreover, incorporating a "cell type" or "tissue" layer would be essential, as gene expression data from whole flies is not directly comparable to data from

specific tissues or even specific conditions. Since phenotypes are often tissue-dependent, this information is vital. However, implementing these layers presented too big of a challenge and was beyond the scope of this paper. “

*(2) Another area that would help to improve is to provide either a subset or the ability to perform a meta-analysis of the studies proposed to see where phenotype intersections occur, as opposed to examining their correlation structure. For any given trait the PLINK data or association results seem already generated so running together and making them available seems fairly straightforward. This can be done in several ways to highlight the utility (for example w/wo specific covariates from Huang et al., 2014 and/or comparing associations that occur similarly or differently between sexes).*

We are not 100% sure what the reviewer refers to when mentioning “phenotype intersection”, but we interpreted it as a “PheWAS capability”. Currently, in DGRPool, for every variant, there is a PheWAS option, which scans all phenotypes across all studies to see if several phenotypes are impacted by this same variant.

We tried to make this tool more visible, both in the GWAS section of the website, but also in the “Check your phenotype” tool, when users are uploading their own data to perform a GWAS. We have also created a “Variants” page, accessible from the top menu, where users can view particular variants and explore the list of phenotypes they are significantly associated with.

From both result pages, users can download the data table as .tsv files.

*(3) As pointed out by the authors, an advantage of DRGs is the ease of testing on homozygous backgrounds. For each phenotype queried (or groups of related phenotypes would be of interest too), I imagine that subsetting strains by the response would help to prioritize lines used for follow-up studies. For example, resistant or sensitive lines to a given trait. This is already done in Fig 4C and 4E but should be an available analysis for all traits.*

For all quantitative phenotypes, we show the global distribution by sex, followed by the sorted distribution by DGRP line. Since the data can be directly downloaded from the corresponding plots, resistant and sensitive lines can then be readily identified for all phenotypes.

*(4) To researchers beyond the DRP community, one feature to consider would be seeing which other associations are conserved across species. While doing this at the phenotype level might be tricky to rename, assigning gene-level associations would make this streamlined. For example, a user could query longevity, subset by candidate gene associations then examine outputs for what is associated with orthologue genes in humans (ex. <https://www.ebi.ac.uk/gwas/docs/file-downloads>) or other reference panels such as mice and rats.*

In all GWAS results, and in the gene-centric tool, we have added links to ortholog databases. In short, when clicking on a variant, users can see which gene is potentially impacted by this variant (gene-level variant annotation). When clicking on these genes, the user can then open the corresponding, detailed gene page.

To address the reviewer’s comment, in the gene page, we have added two orthologous databases (Flybase and OrthoDB), which enables cross-species association analyses.

*(5) Related to enabling a meta-data analysis, it would be helpful to let users download all PLINK or DGRP tables in one query. This would help others to query all data*



| *simultaneously.*

We would like to kindly point out that all phenotyping data can already be downloaded from the front page, which includes the phenotypes, the DGRP lines and the studies' data and metadata. However, we did not provide the global GWAS results through a single file, because the data is too large. Instead, we provide each GWAS dataset via a unique file, available per phenotype, on the corresponding GWAS result page of this phenotype. This file is filtered for  $p < 0.001$ , and contains GWAS results (PLINK beta, p and FDR) as well as gene and regulatory annotations.

| *(6) Following analysis of association data an interesting feature would be to enable users to subset strains for putative LOF variants at a given significant locus. This is commonly done for mouse strains (ex. via MGI).*

The GWAS result table available for each phenotype can be filtered for any variant of interest. We added the capability to filter by variant impact; LOF variants being usually referred to as HIGH impact variants.

| *(7) Viewing the locus underlying annotation can also provide helpful information. For example, several nice fly track views are shown in 10.1534/g3.115.018929, which would help users to interpret molecular mechanisms.*

We now link the GWAS results out to Flybase's JBrowse genome browser.

<https://doi.org/10.7554/eLife.88981.2.sa0>