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

Vincent Gardeux[1,2], Roel P.J. Bevers[1,‡], Fabrice P.A. David[1,2,3], Emily Rosschaert[1,4], Romain Rochepeau[1], Bart Deplancke[1]

[1]Laboratory of Systems Biology and Genetics, Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.
[2]Swiss Institute of Bioinformatics, CH-1015 Lausanne, Switzerland.
[3]Bioinformatics Competence Center, EPFL, Switzerland.
[4]KU Leuven, Belgium
‡Present address: Genomics England Ltd., E14 5AB London, England.

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 **DGRPool** (dgrpool.epfl.ch), which aggregates phenotyping data of 935 phenotypes across 125 DGRP studies in a common environment. DGRPool enables users to download data and run various tools such as genome-wide association analyses (GWAS) and Phenome-WAS analyses. As a proof-of-concept, DGRPool was used to study the longevity phenotype and uncovered both established and unexpected correlations with other phenotypes such as locomotor activity, sleep duration, and oxidative stress resistance. DGRPool 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.

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/)[1,2]. 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**). 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[3]. 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/)[1,2]. This website hosts the genotyping data, its annotation, and potential covariates, as well as 31 phenotypes from 12 studies (**Table 1**). 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 with known covariates). 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, remains thus 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, DGRPool (dgrpool.epfl.ch), 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 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[4].

In addition, we set out to build important tools for the DGRP community such that DGRPool 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 DGRPool 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)[5] 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.

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[6]. By leveraging the data harmonization and curation efforts in DGRPool, we identified multiple phenotypes that are significantly associated with longevity across 18 different studies, such as oxidative stress resistance[7], sleep duration[8,9], desiccation survival[10,11], and starvation resistance[10,12,13]. Interestingly, we also observed correlations between shorter lifespan and certain phenotypes, such as locomotor activity[14] and food intake[15,16]. 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 DGRPool —which ultimately could become entirely community-driven— are essential not only for catalyzing novel research, but also for leveraging the diversity and richfulness of existing datasets.

Results

A thousand phenotypes across 125 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 DGRPool, 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 DGRPool website, including the modification of any metadata attached to a study (title, authors, description, categories), and the submission or modification of attached phenotypes (see **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 DGRPool repository. If the DOI is not in the database, it triggers the same automated scripts that were originally used to incorporate the 125 studies. The corresponding study is then created on DGRPool, 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, DGRPool hosts 125 studies, including 41 that have already been fully curated, 81 still under curation, and 3 under a submitted status given that the latter were used to test DGRPool's DOI feature. All metadata of these three studies were correctly imported into DGRPool, but not the associated phenotypes, which is also the case for a portion of the other 122 studies. Indeed, in total, 74 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 DGRPool is currently 125, 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 **125** studies (independent of "curation" status), and 2) the **curated dataset**, comprising **41** studies that already underwent tedious curation and contributed about 500 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 125 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 41 curated datasets, the proportion of phenotypes without assigned sex (*NA*) dropped drastically to ~15%. This effect highlights the importance of tedious 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[17–21]. 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[22] 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[23] studied cuticular hydrocarbon composition, considering 66 different cuticular components, while Vonesch et al., 2016[24] studied organismal size traits, regrouping 28 morphological phenotypes such as wing length or intraocular distance. In total, the 41 curated studies aggregate 312 M + 220 F + 132 NA = 664 sex-specific phenotypes, for a total of 500 unique phenotypes (~60%), while the remaining non-curated studies provide another 57 M + 34 F + 267 NA = 358 sex-specific phenotypes, for a total of 329 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[25], 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 DGRPool, 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[26–28], 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 DGRPool 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 enables 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 DGRPool'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 DGRPool (**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**). 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[29]. The longevity phenotype however, assessed in at least six of the studies in DGRPool[27,30–34] 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 (Vonesch et al, 2016)[24] and (Grubbs et al, 2013)[35] 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)[1], (Richardson et al, 2012)[36] and (Huang et al, 2014)[2] are expectedly correlated since all three investigate the influence of the *Wolbachia* endosymbiont. Another interesting correlation is between (Chow et al., 2013)[37] and (Durham et al., 2014)[27] which both studied fecundity and yield a cross-study correlation between remating proportion (Chow et al., 2013)[37] vs. mean fecundity (Durham et al., 2014)[27] (**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 DGRPool phenotype correlation tools, 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 society's strong interest in defining "healthy aging" determinants[38], we continued investigating the "mean longevity" phenotype from (Arya et al, 2010)[30] and we selected 50 phenotypes that were significantly correlated with it at 25% 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[10,12,13] and oxidative stress resistance[7] but some are less intuitive such as desiccation survival[10,11], certain cuticular components of the epicuticle[39], and sleep duration[8,9], whose relationship to longevity is complex and still not fully understood[40]. Although we cannot exclude spurious correlations, some of these more surprising correlations appear biologically highly interesting, illustrating the capacity of DGRPool 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[14], some other cuticular components of the epicuticle[41], and food intake[15,16], suggesting that higher locomotor activity

or food intake is linked to reduced longevity. 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[42], it might seem intuitive that higher activity[43] would lead to greater food intake[44]. However, we did not observe such a correlation. Similarly, higher activity levels may reflect increased mating behaviour[37], 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 DGRPool 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[45]. In response, tools like DGRP2 GWAS (http://dgrp2.gnets.ncsu.edu/)[1,2] 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 DGRPool, aiming to assist researchers with performing GWAS analyses and interpreting the respective output. Specifically, we precalculated GWAS analyses using PLINK2 on every existing phenotype in DGRPool (see **Methods**), thereby considering all ~4M available DGRP variants while correcting for six main covariates (*Wolbachia* status, and five major insertions)[2]. Consequently, users can browse the GWAS results from any phenotype page on DGRPool (**Supp. Figure S7**). 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<0.01 threshold. The tool further runs an ANOVA between the phenotype and the six main 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 covariates. 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 over all existing phenotypes in DGRPool. 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 also provide links to their description in Flybase[4].

As mentioned, these GWAS results are available for each existing phenotype in DGRPool, 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 DGRPool, 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 (**Figure 4A**). This threshold yields on average 382 significant hits per tested phenotype, which is skewed due to some phenotypes leveraging lots of results (median = 38). Conveniently, this threshold seems sufficient for avoiding an over-abundant number of false positives, as is clearly visible from other, less stringent, thresholds (**Supp. Figure S8**). 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 $p \leq 2.64 \times 10^{-8}$ (if removing variants with low MAF or high number of missing values). In our results, the Bonferroni threshold ($p \leq 2.64 \times 10^{-8}$) yielded 73 significant hits on average (median = 0, **Supp. Figure S8**) 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[46]. Thus, while choosing an optimal threshold is in general challenging, our results indicate that any threshold below 1×10^{-5} is reasonable given that at this threshold, the *p*-values appear not over-estimated, as observed on the respective QQplots. We also verified if any variant is over-selected across all phenotypes to uncover a possible bias in our studies (**Figure 4B**), but we did not find such variants, even at different thresholding values (data not shown).

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)[26]. Of note, this study is available in DGRPool under https://dgrpool.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 DGRPool as can be observed on the respective phenotype's page (https://dgrpool.epfl.ch/phenotypes/20**, Figure 4C**). For example, DGRPool's GWAS results are very similar to those of the study (https://dgrpool.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 x 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[47,48]. As described in the author's Ph.D. thesis[49], 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*^{M} 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 DGRPool 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[26], 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 other studies, in particular one study on cuticular hydrocarbon composition[23], and another study investigating *Drosophila* microbiota[22]. 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[50,51], 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

376 falls within the top or bottom 15% performers of the phenotype. We then assessed for each
377 DGRP line how often they are 'extreme' and divided this by the total number of phenotypes
378 in which the DGRP line has been included to obtain a "fraction of extremeness" (FoE).
379 Finally, we filtered for lines which had at least 50 phenotypic measures available to ensure
380 that our values were not driven by a low number of observations (**Figure 5A**). Looking
381 broadly, we observed a mild correlation of fraction of extremeness (FoE) across the sexes
382 (**Figure 5B**, Spearman's $\rho = 0.3514$, $p < 1.57 \times 10^{-5}$). While this may indicate that
383 extremeness is a population-wide feature, it is not sufficiently profound to conclude that
384 DGRP lines are generally extreme in both sexes, which may only be the case for specific
385 DGRP lines.

386 Upon considering individual DGRP lines, we can observe to what extent they are extreme for
387 each individual phenotype. In **Figure 5C,** we show the most extreme and "moderate" (i.e.
388 least distinctive) DGRP lines for each sex using an adjusted fraction of extremeness for
389 plotting purposes in which lower scores represent DGRP lines with a high fraction of
390 extremeness. While females of DGRP_879 and males of DGRP_783 tend to be extreme in
391 some cases, for the majority of phenotypes they are considered moderate. Conversely,
392 females of DGRP_757 and males of DGRP_352 are more likely to be labeled as extreme.

393 These examples only represent extremeness for individual DGRP lines of a given sex,
394 however, their counterpart may not be *as* extreme or moderate. We therefore also looked for
395 DGRP lines which can be considered extreme in both females and males, and are
396 potentially more extreme on a population-wide basis. **Figure 5D** describes such populations
397 where the overall fraction of extremeness between males and females differed on average at
398 most 0.05. In these cases, DGRP_852 and DGRP_042 are more likely to be extreme across
399 sexes, which may be attributed to at least two factors. First, this may indicate that the
400 population is generally not healthy if they consistently display a low lifespan, or second, and
401 conversely, well-documented trade-offs of life history traits such as lifespan vs fecundity may
402 be strongly at play here. The former does not however seem to be the case, as shown in
403 **Figure 5E**. Both DGRP_852 and DGRP_042 generally display lifespan values around the
404 mean lifespan of all DGRP lines, suggesting that they are more likely extreme for other
405 phenotypes and are thus not by definition weak lines. However, DGRP_757 and DGRP_765
406 consistently display lower longevity in lifespan studies. These lines may therefore on the one
407 hand be of particular interest for those studying life history traits in an evolutionary context,
408 even though we did not observe strong lifespan and fecundity trade-offs across our
409 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.

Discussion

There are many studies across organisms where collated phenotyping data has led to novel insights[52,53]. 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[5], 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[54]. 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.[55] studied the resistance and tolerance of DGRP flies to the fungal pathogen *Metarhizium anisopliae* (Ma549) and found that the host's defense to Ma549 was correlated with its defense to the bacterium *Pseudomonas aeruginos*a (Pa14). But they also compared this result to several previously published DGRP phenotypes including oxidative stress sensitivity[56], aggression[57], nutritional scores[58], sleep indices[43], and others. Similarly, Zwarts et al.[59] studied the size of the cerebral cortex and the mushroom bodies (MB). They showed that these phenotypes were correlated with phenotypes from other studies like aggression[60] and sleep[43]. Therefore, we believe that DGRPool 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[61] 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 with DGRPool. 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

444 can be used to enhance power and potentially find more small-effect associations. Indeed,
445 researchers are increasingly advocating for collaboration and joining efforts to combine
446 resources[62] to enable more accurate, and reproducible results.

447 Our data collection and harmonization efforts have already enabled us to conduct some
448 interesting cross-study analyses, including an investigation into the presence of biases
449 stemming from outlier DGRP lines. Our "extremeness" analysis revealed that caution is
450 warranted when selecting DGRP lines for specific studies, because, while some DGRP lines
451 may be situated at the outer edge of the phenotypic spectrum by chance, DGRP_757 and
452 DGRP_765 generally display lower lifespans in longevity studies. It is important to note that
453 a shorter lifespan does not necessarily imply lower viability, as populations can still be
454 propagated healthily. However, a shorter lifespan may also result from an impaired
455 development[63] or developmental environment, which may confound the study of healthy
456 aging[64]. Consequently, researchers should consider excluding these extreme lines from their
457 experimental designs to prevent loss of power or potential covariate biases.

458 Furthermore, and beyond our current focus on DGRP lines, we may in the future also
459 consider adding standard *D. melanogaster* lines such as w1118, YWB, YWN or ORB to
460 DGRPool. This is because such lines have often been included as controls in DGRP
461 studies[34], and for most of these, genomic information is also available.

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

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

In conclusion, we propose that DGRPool 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[54]. 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 DGRPool, 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. DGRPool 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 DGRPool 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.

All codes used to produce the figures of this manuscript are also available on our GitHub: https://github.com/DeplanckeLab/DGRPool

Web application

The DGRPool 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 v.2.16.1[29] 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[65] API. 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[66], ArrayExpress[67], or the Sequence Read Archive (SRA)[68] can be associated manually with any study, for example for referencing additional gene expression data that would be published along with the phenotyping data.

Phenotypes correlated with longevity

We computed the correlation of the "mean longevity" phenotype from (Arya et al, 2010)[30] and selected 50 phenotypes that were significantly correlated with it using a 25% FDR threshold. For this, we used the phenotype correlation tools available in DGRPool (result list available at https://dgrpool.epfl.ch/phenotypes/1315/compute_correlation) which makes our results reproducible and freely accessible, following the FAIR principles.

GWAS

GWAS analyses (whether pre-calculated, or using the web tool) use Plink2 v2.00a3LM (1 Jul 2021). 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) with options "--glm --geno 0.2 --maf 0.05". We corrected the model for six main covariates (*Wolbachia* status, and 5 major insertions) that were described in [2] and also used on the DGRP2 website. Of note, these covariates are phenotypes, and thus are also available as a separate, browsable study on DGRPool (https://dgrpool.epfl.ch/studies/17).

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 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.

Acknowledgments

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.* The Drosophila melanogaster Genetic Reference Panel. *Nature* **482**, 173–178 (2012).
2. Huang, W. *et al.* Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Res.* **24**, 1193–1208 (2014).
3. Bou Sleiman, M. S. *et al.* Genetic, molecular and physiological basis of variation in Drosophila gut immunocompetence. *Nat. Commun.* **6**, 7829 (2015).
4. Gramates, L. S. *et al.* FlyBase: a guided tour of highlighted features. *Genetics* **220**, iyac035 (2022).
5. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
6. Piper, M. D. W. & Partridge, L. Drosophila as a model for ageing. *Biochim. Biophys. Acta BBA - Mol. Basis Dis.* **1864**, 2707–2717 (2018).
7. Finkel, T. & Holbrook, N. J. Oxidants, oxidative stress and the biology of ageing. *Nature* **408**, 239–247 (2000).
8. Bushey, D., Hughes, K. A., Tononi, G. & Cirelli, C. Sleep, aging, and lifespan in Drosophila. *BMC Neurosci.* **11**, 56 (2010).
9. Thompson, J. B., Su, O. O., Yang, N. & Bauer, J. H. Sleep-length differences are associated with altered longevity in the fruit fly *Drosophila melanogaster*. *Biol. Open* **9**, bio054361 (2020).
10. Rion, S. & Kawecki, T. J. Evolutionary biology of starvation resistance: what we have learned from *Drosophila*: Starvation resistance in *Drosophila*. *J. Evol. Biol.* **20**, 1655–1664 (2007).
11. Hoffmann, A. A. & Harshman, L. G. Desiccation and starvation resistance in Drosophila: patterns of variation at the species, population and intrapopulation levels. *Heredity* **83**, 637–643 (1999).
12. Chippindale, A. K., Chu, T. J. F. & Rose, M. R. Complex Trade-Offs and the Evolution of Starvation Resistance in Drosophila melanogaster. *Evolution* **50**, 753 (1996).
13. Jang, T. & Lee, K. P. The genetic basis for mating-induced sex differences in starvation resistance in Drosophila melanogaster. *J. Insect Physiol.* **82**, 56–65 (2015).
14. Magwere, T. *et al.* Flight Activity, Mortality Rates, and Lipoxidative Damage in Drosophila. *J. Gerontol. Ser. A* **61**, 136–145 (2006).
15. Lee, K. P. *et al.* Lifespan and reproduction in *Drosophila*□: New insights from nutritional geometry. *Proc. Natl. Acad. Sci.* **105**, 2498–2503 (2008).
16. Piper, M. D. W. & Partridge, L. Dietary Restriction in Drosophila: Delayed Aging or Experimental Artefact? *PLoS Genet.* **3**, e57 (2007).

17. Arch, M., Vidal, M., Koiffman, R., Melkie, S. T. & Cardona, P.-J. Drosophila melanogaster as a model to study innate immune memory. *Front. Microbiol.* **13**, 991678 (2022).
18. Dissel, S. Drosophila as a Model to Study the Relationship Between Sleep, Plasticity, and Memory. *Front. Physiol.* **11**, 533 (2020).
19. Flatt, T. Life-History Evolution and the Genetics of Fitness Components in *Drosophila melanogaster*. *Genetics* **214**, 3–48 (2020).
20. Harnish, J. M., Link, N. & Yamamoto, S. Drosophila as a Model for Infectious Diseases. *Int. J. Mol. Sci.* **22**, 2724 (2021).
21. O'Kane, C. J. Drosophila as a Model Organism for the Study of Neuropsychiatric Disorders. in *Molecular and Functional Models in Neuropsychiatry* (ed. Hagan, J. J.) vol. 7 37–60 (Springer Berlin Heidelberg, 2011).
22. Chaston, J. M., Dobson, A. J., Newell, P. D. & Douglas, A. E. Host Genetic Control of the Microbiota Mediates the Drosophila Nutritional Phenotype. *Appl. Environ. Microbiol.* **82**, 671–679 (2016).
23. Dembeck, L. M. *et al.* Genetic architecture of natural variation in cuticular hydrocarbon composition in Drosophila melanogaster. *eLife* **4**, e09861 (2015).
24. Vonesch, S. C., Lamparter, D., Mackay, T. F. C., Bergmann, S. & Hafen, E. Genome-Wide Analysis Reveals Novel Regulators of Growth in Drosophila melanogaster. *PLOS Genet.* **12**, e1005616 (2016).
25. Hope, K. A. *et al.* The Drosophila Gene Sulfateless Modulates Autism-Like Behaviors. *Front. Genet.* **10**, 574 (2019).
26. Battlay, P., Schmidt, J. M., Fournier-Level, A. & Robin, C. Genomic and Transcriptomic Associations Identify a New Insecticide Resistance Phenotype for the Selective Sweep at the *Cyp6g1* Locus of *Drosophila melanogaster*. *G3 GenesGenomesGenetics* **6**, 2573–2581 (2016).
27. Durham, M. F., Magwire, M. M., Stone, E. A. & Leips, J. Genome-wide analysis in Drosophila reveals age-specific effects of SNPs on fitness traits. *Nat. Commun.* **5**, 4338 (2014).
28. Najarro, M. A., Hackett, J. L. & Macdonald, S. J. Loci Contributing to Boric Acid Toxicity in Two Reference Populations of *Drosophila melanogaster*. *G3 GenesGenomesGenetics* **7**, 1631–1641 (2017).
29. Ackermann, M. *et al.* Effects of assay conditions in life history experiments with Drosophila melanogaster: Assay environment in life history experiments. *J. Evol. Biol.* **14**, 199–209 (2001).
30. Arya, G. H. *et al.* Natural Variation, Functional Pleiotropy and Transcriptional Contexts of *Odorant Binding Protein* Genes in *Drosophila melanogaster*. *Genetics* **186**, 1475–1485 (2010).
31. Huang, W. *et al.* Context-dependent genetic architecture of Drosophila life span. *PLOS Biol.* **18**, e3000645 (2020).
32. Ivanov, D. K. *et al.* Longevity GWAS Using the *Drosophila* Genetic Reference Panel. *J. Gerontol. A. Biol. Sci. Med. Sci.* **70**, 1470–1478 (2015).
33. Zhao, X. *et al.* The metabolome as a biomarker of aging in *Drosophila melanogaster*. *Aging Cell* **21**, (2022).
34. Hoffman, J. M., Dudeck, S. K., Patterson, H. K. & Austad, S. N. Sex, mating and repeatability of *Drosophila melanogaster* longevity. *R. Soc. Open Sci.* **8**, 210273 (2021).
35. Grubbs, N. *et al.* New Components of Drosophila Leg Development Identified through Genome Wide Association Studies. *PLoS ONE* **8**, e60261 (2013).
36. Richardson, M. F. *et al.* Population Genomics of the Wolbachia Endosymbiont in Drosophila melanogaster. *PLoS Genet.* **8**, e1003129 (2012).
37. Chow, C. Y., Wolfner, M. F. & Clark, A. G. Large Neurological Component to Genetic Differences Underlying Biased Sperm Use in *Drosophila*. *Genetics* **193**, 177–185 (2013).
38. Friedman, S. M. Lifestyle (Medicine) and Healthy Aging. *Clin. Geriatr. Med.* **36**, 645–653 (2020).

39. Wang, Z. *et al.* Desiccation resistance differences in Drosophila species can be largely explained by variations in cuticular hydrocarbons. *eLife* **11**, e80859 (2022).
40. Consensus Conference Panel *et al.* 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 (2015).
41. Nghiem, D., Gibbs, A. G., Rose, M. R. & Bradley, T. J. Postponed aging and desiccation resistance in Drosophila melanogaster. *Exp. Gerontol.* **35**, 957–969 (2000).
42. Chatterjee, N. & Perrimon, N. What fuels the fly: Energy metabolism in *Drosophila* and its application to the study of obesity and diabetes. *Sci. Adv.* **7**, eabg4336 (2021).
43. Harbison, S. T., McCoy, L. J. & Mackay, T. F. Genome-wide association study of sleep in Drosophila melanogaster. *BMC Genomics* **14**, 281 (2013).
44. Garlapow, M. E., Huang, W., Yarboro, M. T., Peterson, K. R. & Mackay, T. F. C. Quantitative Genetics of Food Intake in Drosophila melanogaster. *PLOS ONE* **10**, e0138129 (2015).
45. Mackay, T. F. C. & Huang, W. Charting the genotype–phenotype map: lessons from the *Drosophila melanogaster* Genetic Reference Panel. *WIREs Dev. Biol.* **7**, (2018).
46. Uffelmann, E. *et al.* Genome-wide association studies. *Nat. Rev. Methods Primer* **1**, 59 (2021).
47. Daborn, P. J. *et al.* A single P450 allele associated with insecticide resistance in Drosophila. *Science* **297**, 2253–2256 (2002).
48. Schmidt, J. M. *et al.* Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 locus. *PLoS Genet.* **6**, e1000998 (2010).
49. Battlay, P. The quantitative genetics of insecticide resistance in Drosophila melanogaster. (University of Melbourne, 2019).
50. Zwaan, B., Bijlsma, R. & Hoekstra, R. F. Direct selection on life span in *Drosophila Melanogaster*. *Evolution* **49**, 649–659 (1995).
51. Rose, M. & Charlesworth, B. A test of evolutionary theories of senescence. *Nature* **287**, 141–142 (1980).
52. Greene, D. *et al.* Genetic association analysis of 77,539 genomes reveals rare disease etiologies. *Nat. Med.* **29**, 679–688 (2023).
53. Doust, C. *et al.* Discovery of 42 genome-wide significant loci associated with dyslexia. *Nat. Genet.* **54**, 1621–1629 (2022).
54. Bevers, R. P. J. *et al.* Mitochondrial haplotypes affect metabolic phenotypes in the Drosophila Genetic Reference Panel. *Nat. Metab.* **1**, 1226–1242 (2019).
55. Wang, J. B., Lu, H.-L. & St. Leger, R. J. The genetic basis for variation in resistance to infection in the Drosophila melanogaster genetic reference panel. *PLOS Pathog.* **13**, e1006260 (2017).
56. Jordan, K. W. *et al.* Genome-Wide Association for Sensitivity to Chronic Oxidative Stress in Drosophila melanogaster. *PLoS ONE* **7**, e38722 (2012).
57. Shorter, J. *et al.* Genetic architecture of natural variation in *Drosophila melanogaster* aggressive behavior. *Proc. Natl. Acad. Sci.* **112**, (2015).
58. Unckless, R. L., Rottschaefer, S. M. & Lazzaro, B. P. A Genome-Wide Association Study for Nutritional Indices in *Drosophila*. *G3 GenesGenomesGenetics* **5**, 417–425 (2015).
59. Zwarts, L. *et al.* The genetic basis of natural variation in mushroom body size in Drosophila melanogaster. *Nat. Commun.* **6**, 10115 (2015).
60. Zwarts, L. *et al.* Complex genetic architecture of *Drosophila* aggressive behavior. *Proc. Natl. Acad. Sci.* **108**, 17070–17075 (2011).
61. Zeggini, E. & Ioannidis, J. P. Meta-analysis in genome-wide association studies. *Pharmacogenomics* **10**, 191–201 (2009).
62. McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* **9**, 356–369 (2008).
63. May, C. M., Doroszuk, A. & Zwaan, B. J. The effect of developmental nutrition on life span and fecundity depends on the adult reproductive environment in *D rosophila*

719 *melanogaster* . *Ecol. Evol.* **5**, 1156–1168 (2015).
720 64. Iliadi, K. G., Knight, D. & Boulianne, G. L. Healthy Aging – Insights from Drosophila.
721 *Front. Physiol.* **3**, (2012).
722 65. Hendricks, G., Tkaczyk, D., Lin, J. & Feeney, P. Crossref: The sustainable source of
723 community-owned scholarly metadata. *Quant. Sci. Stud.* **1**, 414–427 (2020).
724 66. Barrett, T. *et al.* NCBI GEO: archive for high-throughput functional genomic data.
725 *Nucleic Acids Res.* **37**, D885–D890 (2009).
726 67. Brazma, A. ArrayExpress--a public repository for microarray gene expression data at
727 the EBI. *Nucleic Acids Res.* **31**, 68–71 (2003).
728 68. Kodama, Y., Shumway, M., Leinonen, R., & on behalf of the International Nucleotide
729 Sequence Database Collaboration. The sequence read archive: explosive growth of
730 sequencing data. *Nucleic Acids Res.* **40**, D54–D56 (2012).
731

Figures

Figure 1. General content of the DGRPool web tool. A. Pubmed search on "*Drosophila* DGRP" terms unveiled 131 results from 2012 to 2023 (search made on March 2023). **B.** Sex of the DGRP lines used across all 125 studies (left) and 41 curated studies (right), for each phenotype. Studies have only been curated up to study 41 at the time of writing. **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.

Figure 2. Within- and cross-study phenotype correlations. A. Spearman's correlation of all phenotypes available in the 41 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 0.260 for within-study. Again, we restricted the calculation to the 41 curated studies. **C.** Correlation of two longevity phenotypes from the same study (Arya et al, 2010)[30], 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)[27] and mean longevity (Arya et al, 2010)[30]. Of note, both the C and D plots were generated using the "phenotype correlation" tool in DGRPool.

Figure 3. Phenotype correlations contribute new biological insights. A. Correlation of mean femur length (Grubbs et al., 2013)[35] vs. mean head width (Vonesch et al., 2016)[24] showing the significant cross-study association of organismal size traits. **B.** Correlation of remating proportion (Chow et al., 2013)[37] vs. mean fecundity (Durham et al., 2014)[27]. **C.** 50 phenotypes correlated with longevity (Arya et al, 2010)[30] at a 25% FDR threshold, 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 DGRPool.

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 DGRPool. Of note, all values >1000 have been set to 1000, 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)[26], here to a 0.25 µg/ml dose. This plot was extracted from the phenotype's page on DGRPool at https://dgrpool.epfl.ch/phenotypes/20. **D.** Manhattan plot (taken from DGRPool's result page https://dgrpool.epfl.ch/phenotypes/20/gwas_analysis) showing the association of variants to "survival at 0.25 µg/ml dose" phenotype. **E.** Boxplot (taken from DGRPool's result page https://dgrpool.epfl.ch/phenotypes/20/gwas_analysis), showing the effect of the top variant, 2R:8072884, which is a long insertion.

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.

Supplement Figures

Supplemental Figure S1. **Screenshot from the curator's view for a given study - Metadata section.** This screenshot shows the metadata section of the editing page for a study, where the curator can edit any of the fields. We expect the curator to set a description (short abstract) for the study, and associate some categories. The curator can also deactivate a phenotype if he/she considers that it is not a proper phenotype (like the number of replicates). Once the curation is done, the "Status" field can be changed to "Validated", which signifies that the curation process is finished, allowing the study to be widely visible to the users.

Supplemental Figure S2. **Number of phenotypes per study.** Studies have only been curated up to study 41 at the time of writing. Studies without attached phenotypes were not plotted. We here disregard the sex and thus count the unique phenotypes irrespective of the available sex associated with them. The 41 curated studies have 500 different phenotypes (~60%), while the remaining (S42-S125) studies provide another 329 phenotypes (~40%).

Supplemental Figure S3. **Screenshot from the curator's view for a given study - Phenotype section.** This screenshot shows the phenotype section of the editing page for a study, where the curator can create or update the phenotyping data associated with the study. Here, the data is from (Huang et al, 2020)[31], taken as an example study. It is divided into 4 columns (from left to right): 1) dataset type (raw or summary), 2) phenotypes, 3) DGRP lines, and 4) actions. If the curator submits or updates a phenotype, a parsing script is then run to check the data format, and then the data is updated in the DGRPool database. For each study, the curator can submit, update or delete a unique summary dataset, containing summary data for each DGRP line (for e.g. mean or median values). The curator can also submit multiple raw datasets, if the raw data is available for this study. Raw data means that the phenotyping data is not summarized, i.e. there are multiple values for the same DGRP line (e.g. because of replicate flies). **Note:** Gray phenotypes are deactivated phenotypes, i.e. not treated as real phenotypes (here, it is a block number for each fly).

Supplemental Figure S4. **Screenshot from the phenotype correlation tool result page.** This screenshot shows the results obtained after running the phenotype vs phenotype correlation tool, available directly from a phenotype page, by clicking the "Compute Correlation" button. Of note, there is also the possibility to run this tool from the "Tool" section displayed on the banner of the DGRPool website on any user-submitted phenotype file.

Supplemental Figure S5. Spearman's correlation of all phenotypes available in the 41 curated studies. Here, we applied a binary coloring using a fixed threshold to better visualize the correlations. All correlations above abs(Spearman's ρ) > 0.3 are shown in black (therefore anti-correlated phenotypes are also in black), the others are in white.

Supplemental Figure S6. Comparison of correlation within and cross-study. We calculated the 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 0.170 for cross-study, and 0.287 for within-study. These values are calculated across all phenotypes (125 studies).

Supplemental Figure S7. Screenshot from the GWAS result page. This screenshot shows the results obtained after running the GWAS analysis, available directly from a phenotype page, by clicking the "GWAS" button. Of note, there is also the possibility to run this tool from the "Tool" section displayed on the banner of the DGRPool website on any user-submitted phenotype file.

845 **Supplemental Figure S8. Distribution of the number of GWAS hits per phenotype depending**
846 **on the significance threshold.**

847 ## Tables

			DGRPool	DGRP2
REFERENCE			This study	(Mackay et al., 2012) (Huang, Massouras, et al., 2014)
DATA	*DGRP lines*		341	205
	DGRP studies		125 (41 fully curated)	12
	Phenotypes		935	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*		https://dgrpool.epfl.ch/	http://dgrp2.gnets.ncsu.edu/
	Backend		Ruby-on-rails + PostgreSQL	NA
	Frontend		Javascript, Plotly	NA
FEAT.	Curation system & tools		✓	
	Publish new studies		✓	
	Interactive plots		✓	

848

849 **Table 1**. **Comparison of the two currently available web portals organizing DGRP phenotyping**
850 **data.** This table compares different features available in DGRPool, with DGRP2, the current main
851 resource for DGRP data. It separates the features into 1) **Data**, which summarizes the available
852 phenotyping data, 2) **Tools**, which lists the available tools and options, mainly GWAS, PheWAS and
853 phenotype correlation, 3) **Web**, which describes the website itself, and 4) **Additional features**, that
854 are available in DGRPool, such as the curation system, the possibility to publish new studies and the
855 interactive plots.