Cross-Study Comparison Reveals Common Genomic, Network, and Functional Signatures of Desiccation Resistance in *Drosophila melanogaster*

Marina Telonis-Scott,*[1] Carla M. Sgrò,[1] Ary A. Hoffmann,[2] and Philippa C. Griffin[2]

[1]School of Biological Sciences, Monash University, Clayton, Melbourne, VIC, Australia

[2]School of BioSciences, Bio21 Institute, University of Melbourne, Parkville, Melbourne, VIC, Australia

*Corresponding author: E-mail: marina.telonisscott@monash.edu.

Associate editor: Ilya Ruvinsky

Abstract

Repeated attempts to map the genomic basis of complex traits often yield different outcomes because of the influence of genetic background, gene-by-environment interactions, and/or statistical limitations. However, where repeatability is low at the level of individual genes, overlap often occurs in gene ontology categories, genetic pathways, and interaction networks. Here we report on the genomic overlap for natural desiccation resistance from a Pool-genome-wide association study experiment and a selection experiment in flies collected from the same region in southeastern Australia in different years. We identified over 600 single nucleotide polymorphisms associated with desiccation resistance in flies derived from almost 1,000 wild-caught genotypes, a similar number of loci to that observed in our previous genomic study of selected lines, demonstrating the genetic complexity of this ecologically important trait. By harnessing the power of cross-study comparison, we narrowed the candidates from almost 400 genes in each study to a core set of 45 genes, enriched for stimulus, stress, and defense responses. In addition to gene-level overlap, there was higher order congruence at the network and functional levels, suggesting genetic redundancy in key stress sensing, stress response, immunity, signaling, and gene expression pathways. We also identified variants linked to different molecular aspects of desiccation physiology previously verified from functional experiments. Our approach provides insight into the genomic basis of a complex and ecologically important trait and predicts candidate genetic pathways to explore in multiple genetic backgrounds and related species within a functional framework.

Key words: desiccation, Drosophila, GWAS, gene overlap.

Introduction

In climate change research there is increasing interest to consider not only the obvious impact of changing temperatures on biodiversity, but also fluctuations in rainfall and humidity (Bonebrake and Mastrandrea 2010; Clusella-Trullas et al. 2011; Chown 2012). Changes in water availability pose specific challenges to terrestrial ectotherms such as insects, impacting activity, range distributions, species richness, and disease vector populations (reviewed in Chown et al. 2011, and references therein). Efforts to understand the responses of insects to water availability are underway, given that physiological responses are an integral component of predicting species responses to climate change (Chown et al. 2011; Hoffmann and Sgrò 2011). Insect water balance physiology is relatively well elucidated (Hadley 1994; Chown and Nicolson 2004; Bradley 2009), and the field is undergoing a crucial shift toward understanding the molecular underpinnings, largely achieved through high-throughput and transgenic technologies in *Drosophila*.

Insects can lose water from the epicuticle, through respiration via the spiracles or across the gut epithelia (Hadley 1994). *Drosophila* balance water via three main mechanisms: Altering water content, slowing water loss rate, and less commonly tolerating water loss (Gibbs and Matzkin 2001; Gibbs et al. 2003). Water retention appears to be the primary mechanism for withstanding desiccation in highly resistant cactophilic *Drosophila* (Gibbs and Matzkin 2001), as well as in *Drosophila melanogaster* (Telonis-Scott et al. 2006). How water is preserved, however, is highly variable depending on the species and/or genetic background. Cactophilic species have repeatedly colonized arid habitats via reduced metabolic rates that stem respiratory and cuticular water loss and extend energy utilization (Gibbs and Matzkin 2001; Marron et al. 2003), while in the laboratory, water retention and sequestration arise from multiple evolutionary pathways less clearly related to metabolic rate (Hoffmann and Parsons 1989a), often acting in concert (Hoffmann and Parsons 1989a; Djawdan et al. 1998; Folk et al. 2001; Telonis-Scott et al. 2006, 2012). Less resolved is the role of respiratory relative to cuticular water loss via discontinuous gas exchange, although this might be an important water budget strategy for insects in general (Chown et al. 2011).

Recent candidate gene-based approaches have seen new developments in understanding specific molecular aspects of this variable ecological trait in *Drosophila*. Diuretic peptide

signaling in ion and water balance regulated by excretion via the Malpighian tubules (MTs) and gut absorption has been shown to impact desiccation resistance in *D. melanogaster* (Kahsai et al. 2010; Terhzaz et al. 2012, 2014, 2015). Specifically, knockdown of the *Capability* (*Capa*) neuropeptide gene enhanced desiccation resistance by reductions in respiratory, cuticular, and excretory water loss, while also cross-conferring cold tolerance (Terhzaz et al. 2015). Cuticular hydrocarbon (CHC) levels are also implicated; a single gene encoding a fatty acid synthase *mFAS* was shown to impact both desiccation resistance and mate choice in Australian rainforest endemics *Drosophila serrata* and *Drosophila birchii* (Chung et al. 2014). In *D. melanogaster*, knockdown of *CYP4G1* also reduced CHC production and impaired survival under low humidity (Qiu et al. 2012). Metabolic signaling has also been shown to impact desiccation resistance, and includes components of the insulin signaling pathways (Söderberg et al. 2011; Liu et al. 2015). Several other single gene studies highlight different mechanisms such as desiccation avoidance in larvae via nociceptors encoded by members of the transient receptor potential (TRP) aquaporin family (Johnson and Carder 2012), cyclic adenosine monophosphate (cAMP)-dependent signaling protein kinase *desi* (Kawano et al. 2010), and potential tissue protection via trehalose sugar accumulation (Thorat et al. 2012).

Collectively, these functional approaches highlight the complexity of insect water balance strategies in controlled backgrounds, but do not explain the variation observed in natural populations and species. Resistance evolves rapidly and is highly heritable (up to 60%) in the cosmopolitan and unusually resistant species *D. melanogaster*, whereas heritability is much lower in less tolerant range-restricted species (Hoffmann and Parsons 1989b; Kellermann et al. 2009). Multiple, genome-wide quantitative trait loci were identified from mapping lines constructed from a natural *D. melanogaster* population, suggesting a polygenetic architecture (Foley and Telonis-Scott 2011). Transcriptomics revealed differential regulation of thousands of genes in response to desiccation in *Drosophila mojavensis* (Matzkin and Markow 2009), while artificial selection for desiccation resistance altered the basal expression of over 200 genes in *D. melanogaster* (Sørensen et al. 2007).

Previously, we used microarray-based genomic hybridization to survey allele frequency shifts in experimental evolution lines of *D. melanogaster* recently derived from the field (Telonis-Scott et al. 2012). We documented shifts in over 600 loci in response to selection for desiccation resistance following a rapid phenotypic response after only 8 generations of selection. This variant identification approach was limited to highlighting candidate genes and regions, and not actual nucleotide polymorphisms apart from those sequenced post hoc. We now expand on this to further explore the unresolved natural genomic complexity of desiccation resistance using a high-throughput sequencing Pool-GWAS (genome-wide association study) approach (Bastide et al. 2013). In a GWAS framework for polygenic traits where many small effect genes contribute to a phenotype, a large sample size is required for adequate power (Mackay et al.

2009). Accordingly, we sampled natural genetic variation from almost 1,000 inseminated wild-caught females, and compared the upper 5% desiccation resistant tail from almost 10,000 F_1 progeny with a control sample chosen randomly from the same progeny set. Harnessing the power of Pool-GWAS, we sequenced a pool of 500 natural "resistant" genomes and compared allele frequencies with a pool of 500 "random control" genomes.

We employed a novel comparative approach with the resulting set of candidate loci and those detected in our previous artificial selection study of flies collected several seasons earlier from the same region of southeastern Australia (Telonis-Scott et al. 2012). Cross-study repeatability of loci associated with complex phenotypes is often low (Sarup et al. 2011), and can be influenced by genetic background, selection intensity, epistatic effects, and gene-by-environment interactions (Mackay et al. 2009; Civelek and Lusis 2014). Multiple genetic solutions appear to underlie the array of desiccation responses, and we further explored this by investigating overlap at the level of individual genes, functional gene ontology (GO) categories, and protein–protein interaction (PPI) networks. Genetic variants contribute to final phenotypes by way of "intermediate phenotypes" (transcript, protein, and metabolite abundances) and correlations should theoretically occur across these biological scales (Civelek and Lusis 2014).

Here we report on a screen of natural nucleotide variants associated with desiccation resistance and, using a powerful analysis approach, demonstrate common cross-study signatures across different hierarchical levels from gene, to network, and function. We found evidence that some functional desiccation candidates may be important in wild populations, and discovered variants linked to multiple physiological responses, consistent with the trait's complex underpinnings at both the physiological and molecular levels.

Results

Genome Wide Differentiation for Natural Desiccation Resistance

We collected over 1,000 *D. melanogaster* isofemale lines from southern Australia, and after one generation of laboratory culture screened over 9,000 female progeny for desiccation resistance. The top ~5% desiccation-resistant flies were selected for Illumina sequencing (~500 flies), together with a random sampling of the same number of flies from the families as the "control" pool. Our design incorporated a subpooling (technical replicate) strategy to both optimize Pool-seq allele frequency estimates and control for technical bias on allele frequency estimates (see Materials and Methods). For the control and desiccation-resistant samples respectively, 162M–259M and 188M–289M raw reads were obtained per technical replicate, and an average of 95.0% of reads mapped to the reference genome after trimming (mean of all 10 replicates; standard deviation [SD] 0.5%).

Variance in allele frequency among the five technical replicates for each pool was low (supplementary fig. S1,
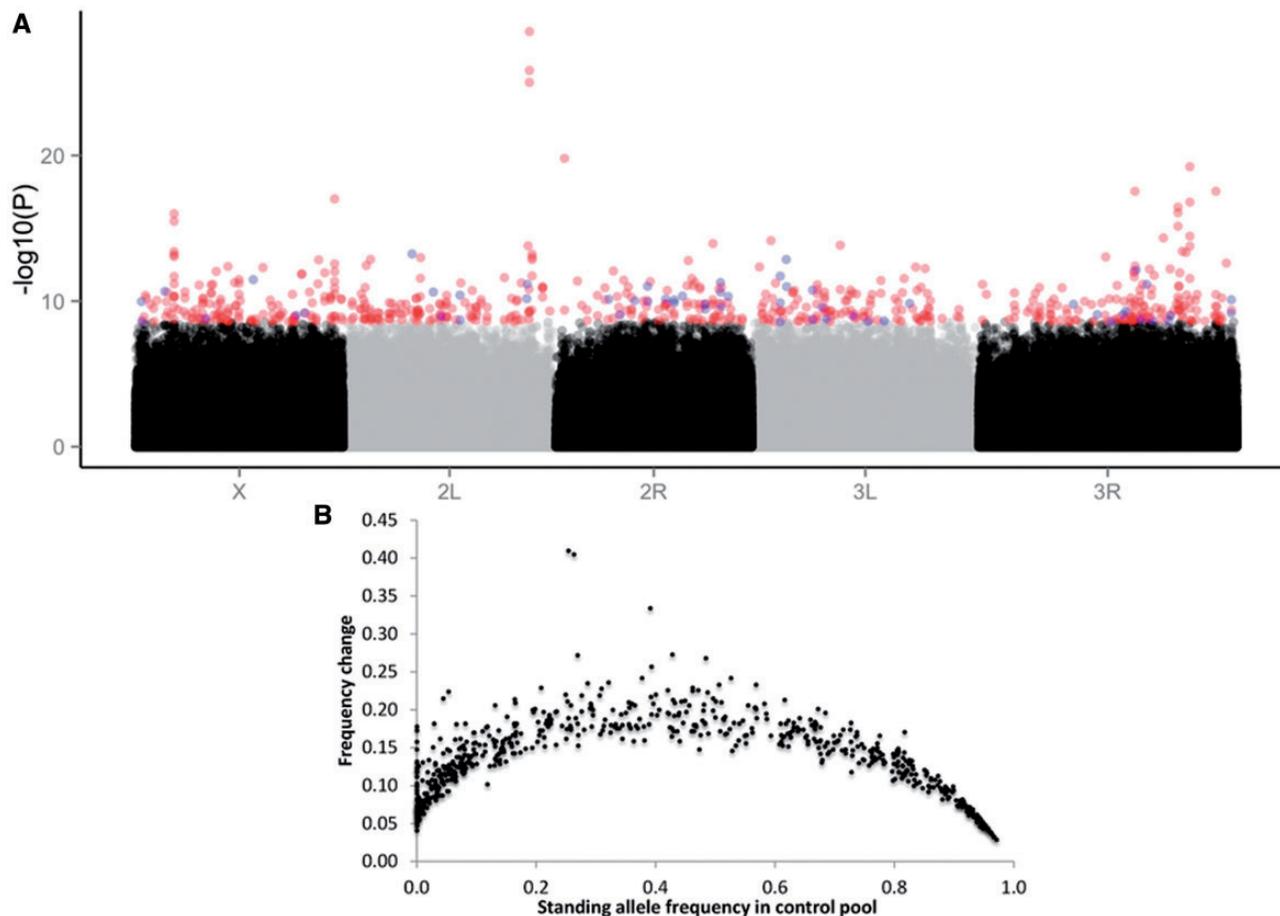
Fig. 1. (A) Manhattan plot of genome-wide *P* values for differentiation between desiccation-resistant and random control pools, shown for the entire genome where each point represents one SNP. SNPs highlighted in red showed differentiation above the 0.05% threshold level determined from the null *P* value distribution (see text for details); SNPs highlighted in blue were above the 0.05% threshold and also detected in Telonis-Scott et al. (2012). (B) Standing allele frequency in the control pool plotted against the frequency change in the selected pool for the 648 differentiated SNPs (FDR 0.05%).

Supplementary Material online; mean across control replicates: 0.0028, median: 0.0015; mean across resistant replicates: 0.0027, median: 0.0016), corresponding to a mean SD of ~4.5% around the mean allele frequency. The mean pairwise difference in allele frequency estimates was 5.5% and 5.4% for control and desiccation-resistant replicates respectively, consistent with previously published estimates (4–6%; Kofler et al. 2016).

For a test subset of 1,803 randomly sampled single nucleotide polymorphisms (SNPs) that showed no significant differentiation between control and desiccation-resistant pools, the replicates accounted for the majority of total variance in allele frequency. Replicates within pool category (control vs. desiccation resistant; supplementary fig. S1, Supplementary Material online) explained between 15% and 100% (mean: 87%; median: 93%) of what was a relatively low total variance (mean: 0.0029, median: 0.0019). The mean concordance correlation coefficient was 0.98 for each pool category, higher than a comparable *Drosophila* study that reported a correlation of 0.898 between 2 replicates with 20 individuals (Zhu et al. 2012). We therefore combined the technical replicates into a single control and resistant pool respectively for further analyses. After processing, the mean coverage depth

across the genome was 487× and 568× per sample, respectively (equivalent to 0.97× and 1.1× per individual in the pools of 500 flies).

Allele Frequency Changes

The allele frequencies of 3,528,158 SNPs were compared between the 2 pools using Fisher's exact tests with a permutation-based false discovery rate (FDR) correction chosen based on the top 0.05% of the simulated null *P* values. Six hundred forty-eight SNPs were differentiated between desiccation resistant and control pools (fig. 1 and supplementary table S1, Supplementary Material online). The SNPs were nonrandomly distributed across chromosomes ($X_4^2 = 43.12$, $P < 0.0001$), highly abundant on the X chromosome ($X_1^2 = 36.98$, $P < 0.0001$), while SNPs were underrepresented on both 2L and 3L (2L: $X_1^2 = 6.41$, $P < 0.01$; 3L: $X_1^2 = 3.97$, $P < 0.05$). In the desiccation pool, the allelic distribution of the 648 SNPs ranged from low to fixed (4–100%; fig. 1B and supplementary table S1, Supplementary Material online). The increase in frequency of the favored "desiccation" alleles compared with the controls ranged from 3% to 41%, (median increase 14%; supplementary table S1, Supplementary Material online), where

the largest shifts tended to occur in common intermediate frequencies (fig. 1B).

Inversion Frequencies

To test whether inversions were associated with desiccation resistance, allele frequencies were checked at SNPs diagnostic for the following inversions: In(3R)P (Anderson et al. 2005; Kapun et al. 2014); In(2L)t (Andolfatto et al. 1999; Kapun et al. 2014); In(2R)NS, In(3L)P, In(3R)C, In(3R)K, and In(3R)Mo (Kapun et al. 2014). In(3L)P, In(3R)K, and In(3R)Mo were not detected in this population, and for all other inversions investigated, the control and desiccation-resistant pool frequencies did not differ more than 6%.

Genomic Distribution

We examined these candidates using a comprehensive approach encompassing SNP, gene, functional ontology, and network levels. First, we examined SNP locations with respect to gene structure and putative effects of SNPs on genome features. The 648 SNPs mapped to 382 genes and were largely noncoding (79%; supplementary table S2, Supplementary Material online). Coding variants comprised nearly 15% of candidate SNPs; largely synonymous substitutions and the remainder nonsynonymous missense variants predicted to result in amino acid substitution with length preservation (9% and 1.4%; supplementary table S2, Supplementary Material online). Based on proportions of features from all annotated SNPs compared with the candidates, intron variants were significantly underrepresented ($X_1^2 =7.06, P < 0.01$; supplementary table S2, Supplementary Material online), while SNPs were enriched in 5′UTRs, 5′UTR premature start codon sites, and splice regions ($X_1^2 = 8.20, P < 0.01$; $X_1^2 = 31.24, P < 0.0001$; $X_1^2 = 6.40, P < 0.05$, respectively; supplementary table S2, Supplementary Material online).

GO Functional Enrichment

For the GO analyses, we found no overrepresented terms using the stringent Gowinda gene mode, but identified several categories using SNP mode with a 0.05% FDR threshold. GO terms were enriched for the broad categories of "chromatin," "chromosome," and more specifically "double-stranded RNA binding," although significance of this category was due solely to nine differentiated SNPs in the DIP1 gene. The term "gamma-secretase complex" was also enriched due to three SNPs in the pen-2 gene.

Desiccation Candidate Genes

Chown et al. (2011) comprehensively summarized the primary mechanisms underpinning desiccation resistance, from the critical first phase of stress sensing and behavioral avoidance, to the manifold physiological trajectories available to counter water stress. These include water content variation, water loss rates (desiccation resistance), and tolerance of water loss (dehydration tolerance). We expanded on this review to incorporate recent molecular research and summarized the primary mechanisms and known candidate genes relevant to D. melanogaster stress sensing and water balance

(table 1 and references therein). Informed by this body of work, we applied a "top-down" mechanistic approach to screening desiccation candidate genes in our data (e.g., from initial hygrosensing to downstream key physiological pathways; table 1). We did not observe differentiation in SNPs mapping to the TRP ion channels, the key hygrosensing receptor genes. However, we did map SNPs to genes involved in stress sensing specific to the MT fluid secretion signaling pathways, including two of the six cAMP and cyclic guanosine monophosphate (cGMP) signaling phosphodiesterases dnc and pde9, as well as klu, indicated in NF-kB ortholog signaling in the renal tubules (table 1). A number of the stress-sensing SAPK pathway genes were differentiated (table 1). Key genes involved in the insulin signaling pathway implicated in metabolic homeostasis and water balance were also differentiated, including the insulin receptor InR and insulin receptor binding Dok (table 1).

Network Analysis: Modeling Genomic Differentiation for Desiccation in a PPI Context

Given the genomic complexity underpinning desiccation resistance, we also implemented a higher-order analysis to provide an overview of potential architectures beyond the SNP and gene levels using PPI networks. The SNPs mapped to 382 genes which were annotated to 307 seed proteins occurring in the full background D. melanogaster PPI network. The seed proteins produced a large first-order network with 3,324 nodes and 51,299 edges.

To determine if broader functional signatures could be ascertained from the complex network, we employed functional enrichment analyses on the 3,324 nodes (table 2 and supplementary table S3, Supplementary Material online). Strikingly, the network was enriched for pathways involved in gene expression, from transcription through RNA processing to RNA transport, metabolism, and decay (table 2). The network analysis also revealed further complexity of the stress response with enrichment of proteins involved in regulating innate insect immunity (table 2), where multiple pathways were overrepresented including cytokine, interleukin, and toll-like receptor signaling cascades (table 2). Developmental/gene regulation signaling pathways were significant in the pathway analysis; the KEGG database showed enrichment for Notch and Mitogen-activated protein kinase (MAPK) signaling, while insulin and NGF pathways were highlighted from the Reactome database (table 2).

Common Signatures of Desiccation Resistance across Studies

The key finding of this study is the degree of genomic overlap with our previous mapping of alleles associated with artificially selected desiccation phenotypes (Telonis-Scott et al. 2012). Although the flies were sampled from comparable southeastern Australian locations, the study designs were different. Specifically, the earlier study focused on resistance generated by artificial selection, whereas this study focused on natural variation in resistance. Despite the differences in study design, we identified 45 genes common to both studies

Table 1. Summary of Primary Molecular Responses and Candidate Genes for Desiccation Response Mechanisms Identified in the Literature.

Mechanism	Molecular Function	Genes	Current Study (F_1 wild-caught)	Telonis-Scott et al. (2012) (F_8 artificial selection)
Hygrosensing: Sensory moisture receptors (antennae, legs, gut, mouthparts)				
Temperature-activated transient receptor potential ion channels[b,c,d]	TRP ion channels	*trpl, trp, Trpη, TrpA1, Trpm, Trpml, pain, pyx, wtrw, nompC, iav, nan, Amo*	—	*trpl*
Stress sensing: Malpighian tubules (principal cells)				
Tubule fluid secretion signaling pathways				
cAMP signaling[e,f,g]	cAMP activation	*Dh44-R2*	—	—
cGMP signaling[e,f,g]	cGMP activation	*Capa, CapaR, Nplp1-4*	—	—
	Receptors	*Gyc76c, CG33958, CG34357*	—	*CG34357*
	Kinases	*dg1-2*	—	—
cAMP + cGMP signaling[e,f,g,h]	Hydrolyzing phosphodiesterases	*dnc, pde1c, pde6, pde9, pde11*	*dnc, pde9*	*dnc, pde1c, pde9, pde11*
Calcium signaling[e,g]	Calcium-sensitive nitric oxide synthase	*Nos*	—	*Nos*
Desiccation specific	Diuretic neuropeptide, NFkB signaling[i,j,k]	*Capa, CapaR, trpl, Relish, klu*	*klu*	*klu, trpl*
	Anti-diuretic/neuropeptide[l]	*ITP, sNPF, Tk*	—	*ITP, sNPF*
Stress sensing: Stress responsive pathway				
Stress activated protein kinase pathway[m,n]	MAPK Jun kinase	*Aop, Aplip1, Atf-2, bsk, Btk29A, cbt, Cdc 42, Cka, cno, CYLD, Dok, egr, Gadd45, hep, hppy, Jra, kay, medo, Mkk4, msn, Mtl, Pak, puc, pyd, Rac 1-2, Rho1, shark, slpr, Src42A, Src64B, Tak1, Traf6*	*Aop, Dok, hep, kay, msn, Mtl, puc, Rac1, Src64B*	*Mtl, slpr*
Metabolic homeostasis and water balance				
Components of insulin signaling pathway	Receptor[o,p]	*InR*	*InR*	—
	Receptor binding	*chico, Ilp1-8, dock, Dok, poly*	*Dok*	*Poly*
Resistance mechanisms: Water loss barriers				
Cuticular hydrocarbons	Fatty acid synthases[q], desaturases[r], transporter[s], oxidative decarbonylase[t]	*CG3523, CG3524 desat1-2, FatP, Cyp4g1*	*CG3523*	—
Primary hemolymph sugar/tissue-protectant				
Trehalose metabolism	Trehalose-phosphatase[b,u], trehlase activity[b,u], phosphoglycerate mutase	*Tps1, Treh, Pgm, CG5171, CG5177, CG6262, crc*	*Treh, Pgm*	—

NOTE.—Candidate genes differentiated in the current study and in Telonis-Scott et al. (2012) are shown and are underlined where overlapping.

[a]For brevity several comprehensive reviews are cited and readers are referred to references therein. Note that water budgeting via discontinuous gas exchange was not included due to a lack of "bona fide" molecular candidates, and desiccation tolerance due to aquaporins was omitted as genes were not differentiated in either of our studies.

[b]Chown et al. (2011).
[c]Johnson and Carder (2012).
[d]Fowler and Montell (2013).
[e]Day et al. (2005).
[f]Davies et al. (2012).
[g]Davies et al. (2013).
[h]Davies et al. (2014).
[i]Terhzaz et al. (2012).
[j]Terhzaz et al. (2014).
[k]Terhzaz et al. (2015).
[l]Kahsai et al. (2010).
[m]Huang and Tunnacliffe (2004).
[n]Huang and Tunnacliffe (2005).
[o]Söderberg et al. (2011).
[p]Liu et al. (2015).
[q]Chung et al. (2014).
[r]Sinclair et al. (2007).
[r]Foley and Telonis-Scott (2011).
[t]Qiu et al. (2012).
[u]Thorat et al. (2012).

Table 2. Pathway Enrichment of the GWAS and Telonis-Scott et al. First-Order PPI Networks Investigated Using FlyMine (FDR $P < 0.05$).

	Current Study: GWAS	Telonis-Scott et al. (2012)
Database	**Pathway**	**Pathway**
KEGG	Ribosome	Ribosome
	Notch signaling pathway	Notch signaling pathway
	MAPK signaling pathway	
	Dorso-ventral axis formation	
	Progesterone-mediated oocyte maturation	
		Wnt signaling pathway
		Endocytosis
		Proteasome
Reactome	Immune system	Immune system
	Adaptive immune system	*Adaptive immune system*
	Innate immune system	*Innate immune system*
	Signaling by interleukins	*Signaling by interleukins*
	(Toll-like receptor cascades)	*Toll-like receptor cascades*
		TCR signaling
		(B cell receptor signaling)
	Gene expression	**Gene expression**
	Metabolism of RNA	*Metabolism of RNA*
	Nonsense-mediated decay	*Nonsense-mediated decay*
	Transcription	
	mRNA processing	
	(RNA transport)	
	Metabolism of proteins	*(Metabolism of proteins)*
	Translation	*(Translation)*
	Cell cycle	**Cell cycle**
		Cell cycle checkpoints
		Regulation of mitotic cell cycle
		Mitotic metaphase and anaphase
	G0 and early G1	*G0 and early G1*
		G1 phase
		(G1/S transition)
		G2/M transition
		(DNA replication/S phase)
		DNA repair
		Double-strand break repair
	Signaling pathways	**Signaling pathways**
	Insulin signaling pathway	
	Signaling by NGF	*Signaling by NGF*
		Signaling by FGFR
		Signaling by NOTCH
		Signaling by Wnt
		Signaling by EGFR
		Development
		Axon guidance
		Membrane trafficking
		Gap junction trafficking and regulation
		Programed cell death
		Apoptosis
		Hemostasis
		Hemostasis

NOTE.—KEGG and Reactome pathway databases were queried. All significant terms are reported, with Reactome terms grouped together under parent terms in italic text (full list of terms available in supplementary table S3, Supplementary Material online). Parent terms without brackets are the names of pathways that were significantly enriched; parent terms in parentheses were not listed as significantly enriched themselves, but are reported as a guide to the category contents.

out of the 382 and 416 genes for the current and 2012 study, respectively (Fisher's exact test for overlap, $P < 2 \times 10^{-16}$, odds ratio = 5.90; supplementary table S4, Supplementary Material online). Bearing gene length bias in mind, we examined the average length of genes in both studies as well as the gene overlap to the average gene length in the *Drosophila* genome (supplementary fig. S3, Supplementary Material online). Although there was bias toward longer genes in the overlap list, this was also apparent in the full gene lists from both studies.

The overlap genes mapped to all of the major chromosomes and their differentiated variants were mostly noncoding SNPs located in introns (supplementary table S4, Supplementary Material online). Two SNPs were synonymous (*Strn-Mlick* coding for a protein kinase and *Cht7* chitin binding; supplementary table S4, Supplementary Material online) while a 3′UTR variant was predicted to putatively alter protein structure in a nondisruptive manner, most likely impacting expression and/or protein translational efficiency (*jbug*, response to stimulus; supplementary table S4, Supplementary Material online). One nonsynonymous SNP was predicted to alter protein function as a missense variant (*fab1*, signaling; supplementary table S4, Supplementary Material online). The increase in frequency of the favored desiccation alleles compared with the controls ranged from 5% to 27% with a median increase of 16.4%.

Despite the strong overlap, there were important differences between the two studies. For example, our previous work revealed evidence of a selective sweep in the 5′ promoter region of the *Dys* gene in the artificially selected lines, and we confirmed a higher frequency of the selected SNPs in an independent natural association study in resistant flies from Coffs Harbour (Telonis-Scott et al. 2012). In this study, we found no differentiation at the *Dys* locus between our resistant and random controls. However, both resistant and control pools had a high frequency of the selected allele detected in Telonis-Scott et al. (2012) (supplementary table S5, Supplementary Material online).

The 45 genes common to both studies were analyzed for biological function; while many genes are obviously pleiotropic (i.e., assigned multiple GO terms), almost half were involved in response to stimuli (cellular, behavioral, and regulatory), including signaling such as Rho protein signal transduction and *cAMP*-mediated signaling (table 3). Other categories of interest include spiracle morphogenesis, sensory organ development including development of the compound eye and stem cell, and neuroblast differentiation (table 3).

With regard to specific stress response/desiccation candidates (table 1), 4 of the 45 genes overlapped, including the cAMP/cGMP signaling phosphodiesterases *pde9* and *dnc*, MAPK pathway genes *Mtl* and *klu*, indicated in *NF-kB* orthologue signaling in the renal tubules (table 1). Furthermore, there was overlap among mechanistic categories where different genes in the same pathways were detected. For example, additional key cAMP/cGMP hydrolyzing phosphodiesterases *pde1c* and *pde11* were significant in Telonis-Scott et al. (2012), as well as *slpr* (MAPK pathway) and *poly* (insulin signaling) (table 1).

Table 3. Overrepresentation of GO Categories for the 45 Core Candidate Genes Overlapping between the Current and Our Previous (Telonis-Scott et al. 2012) Genomic Study.

Term ID	Term/Subterms	FDR P value[a]	Genes
Cellular/behavioral response to stimulus, defense, and signaling			
GO:0050896	Response to stimulus (cellular response to stimulus, regulation of response to stimulus)	0.007	*Abd-B, dnc, fz, mam, nkd, slo, ct, fru, klu, Rbf, klg, santa-maria, jbug, fab1, CG33275, Mtl, Trim9, cv-c, Gprk2, Ect4, CG43658*
GO:0048585	Negative regulation of response to stimulus (Rho protein signal transduction, signal transduction)	0.042	*dnc, mam, slo, fru, Trim9, cv-c*
GO:0023052	Signaling (cAMP-mediated signaling, cell communication)	0.031	*Amph, Pde9, dnc, fz, mam, nkd, ct, klu, Rbf, santa-maria, fab1, CG33275, Mtl, Trim9, cv-c, Gprk2, Ect4, CG43658*
Development			
GO:0048863	Stem cell differentiation	0.014	*Abd-D, mam, nkd, klu, nub*
GO:0014016	Neuroblast differentiation	0.048	
GO:0045165	Cell fate commitment	0.010	*Abd-D, dnc, fz, mam, nkd, ct, klu, klg, nub*
GO:0035277	Spiracle morphogenesis, open tracheal system	0.010	*Abd-D, ct, cv-c*
GO:0007423	Sensory organ development (eye, wing disc, imaginal disc)	0.008	*fz, mam, ct, klu, klg, sdk, Amph, Nckx30c, Mtl*
GO:0048736	Appendage development	0.020	*fz, mam, ct, fru, CG33275, Mtl, Fili, nub, cv-c, Gprk2, CG43658*
GO:0007552	Metamorphosis	0.001	*fz, mam, ct, fru, CG33275, Mtl, Fili, cv-c, Gprk2, CG43658*
Cellular component organization			
GO:0030030	Cell projection organization	0.020	*dnc, fz, ct, fru, GluClalpha, Nckx30c, Mtl, Trim9, nub, cv-c*

NOTE.—Note that most genes are assigned to multiple annotations and therefore are shown purely as a guide to putative function.
[a]Bonferroni–Hochberg FDR and gene length corrected P values using the Flymine v40.0 database Gene Ontology Enrichment. The terms were condensed and trimmed using REVIGO.

In addition to the significant gene-level overlap between the two studies, we observed significant overlap between the PPI networks constructed from each set of candidate genes (summarized in fig. 2). We examined the overlap by comparing the first-order networks 1) at the level of node overlap and 2) at the level of edge (interaction) overlap. Three hundred seven GWAS seeds formed a network of 3,324 nodes, while our previous set of candidates mapped to 337 seeds forming a network that comprised 3,215 nodes. One thousand eight hundred twenty-six, or ~55%, of the nodes overlapped between networks, which was significantly higher than the null distribution estimated from 1,000 simulations (mean simulated overlap ± SD: 1,304 ± 143 nodes; observed overlap at the 99.9th percentile; fig. 2A and C). The overlap at the interaction level was not significantly higher than expected (mean simulated overlap ± SD: 25,704 ± 4,024 edges; observed overlap: 30,686 edges; 89th percentile; fig. 2B and D). Hive plots were used to visualize the separate networks (supplementary fig. S4A and B, Supplementary Material online) as well as the overlap between the networks (supplementary fig. S4C, Supplementary Material online).

Given the high degree of shared nodes from distinct sets of protein seeds, we compared the networks for biological function (fig. 2E). Both networks produced long lists of significantly enriched GO terms (supplementary table S3, Supplementary Material online) and assessing overlap was difficult due to the nonindependent nature of GO hierarchies. However, a comparison of functional pathway enrichment revealed some differences. Although both networks showed enrichment in the immune system pathways, the ribosome, and Notch and NGF signaling pathways, as well as some involvement of gene expression and translation pathways, the network constructed from the Telonis-Scott et al. (2012) gene list showed extensive involvement of cell cycle pathway proteins, as well as significant enrichment for Wnt signaling, EGRF signaling, DNA double-strand break repair, axon guidance, gap junction trafficking, and apoptosis pathways, which were not enriched in the GWAS network. The GWAS network was specifically enriched for dorso-ventral axis formation, oocyte maturation, and many more gene expression pathways than the Telonis-Scott et al. (2012) gene network (table 2 and supplementary table S3, Supplementary Material online).

Discussion

Genomic Signature of Desiccation Resistance from Natural Standing Genetic Variation

We report the first large-scale genome-wide screen for natural alleles associated with survival of low humidity conditions in *D. melanogaster*. By combining a powerful natural association study with high-throughput Pool-seq, we
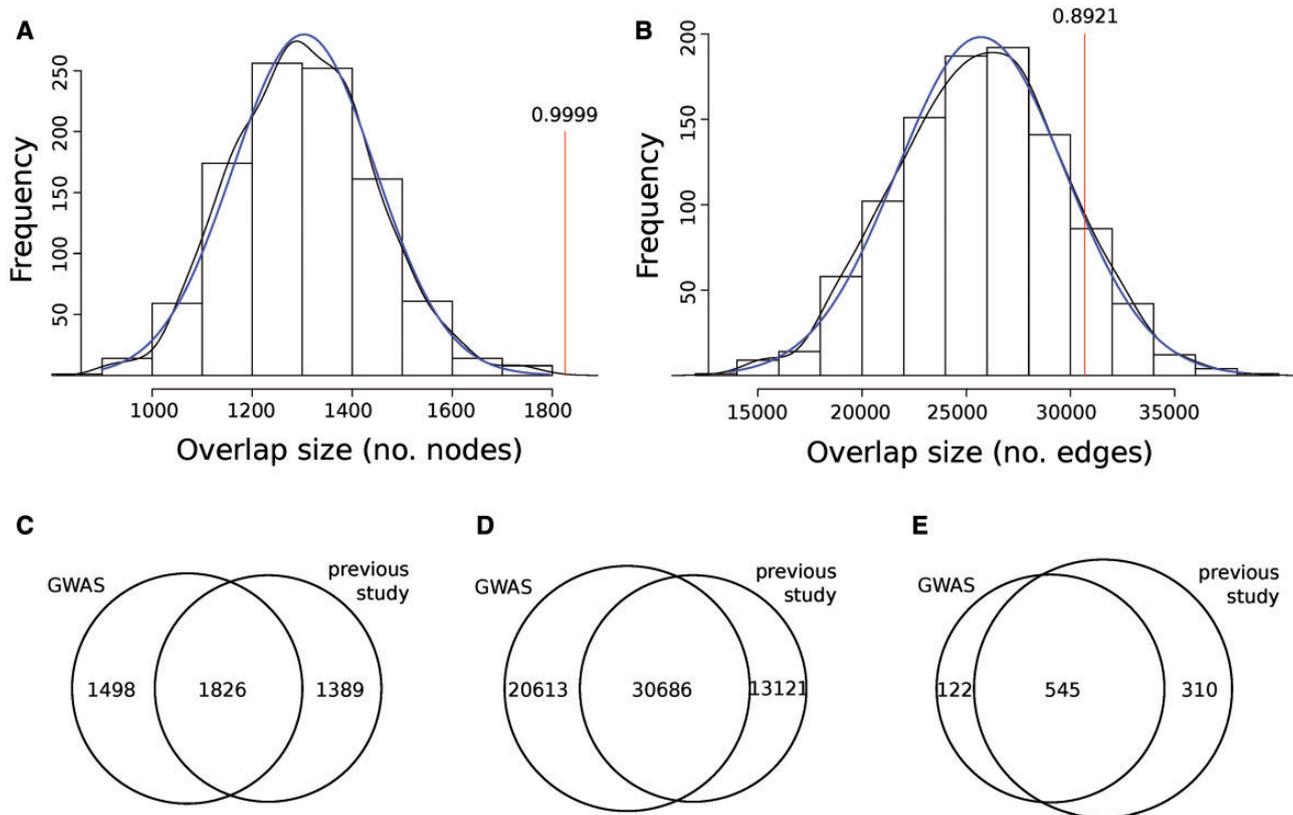
Fig. 2. Observed versus simulated network overlap between the GWAS first-order PPI network and the first-order PPI network from the Telonis-Scott et al. (2012) gene list. Overlap is presented in terms of (*A*) node number and (*B*) edge number. Networks were simulated by resampling from the entire *Drosophila melanogaster* gene list (r5.53) as described in the text. The histogram shows the distribution of overlap measures for 1,000 simulations. The black line represents the histogram as density and the blue line shows the corresponding normal distribution. The red vertical line shows the observed overlap from the real networks, labeled as a percentile of the normal distribution. Area-proportional Venn diagrams summarized the extent of overlap for PPI networks constructed separately from the GWAS candidates and Telonis-Scott et al. (2012) candidates (labeled "previous study"). (*C*) Overlap expressed as the number of nodes. Approximately 55% of the nodes overlapped between the two studies which was significantly higher than simulated expectations (99.9% percentile; 2B). (*D*) Overlap expressed in the number of interactions. This was not significantly higher than expected by simulation (89th percentile; 2B). (*E*) Overlap at the level of GO biological process terms (directly compared GO terms by name; this was not tested statistically because of the complex hierarchical nature of GO terms and is presented for interest).

mapped SNPs associated with desiccation resistance in 500 naturally derived genotypes. Although Pool-seq has limited power to reveal very low frequency alleles, to reconstruct phase or adequately estimate allele frequencies in small pools (Lynch et al. 2014), the Pool-GWAS approach has been validated by retrieval of known candidate genes involved in body pigmentation, using similar numbers of genotypes as assessed here (Bastide et al. 2013).

Although studies in controlled genetic backgrounds report large effects of single genes on desiccation resistance (Table 1), our current and previous screens revealed similarly large numbers of variants (over 600) mapping to over 300 genes. A polygenic basis of desiccation resistance is consistent with quantitative genetic data for the trait (Foley and Telonis-Scott 2011), and we also anticipate more complexity from natural *D. melanogaster* populations with greater standing genetic variation than inbred laboratory strains. However, some false positive associations are likely in the GWAS due to the limitations of quantitative trait mapping approaches. Pooled evolve-and-resequence experiments routinely yield vast numbers of candidate SNPs, partly due to high levels of standing linkage disequilibrium (LD) and hitchhiking neutral alleles

(Schlötterer et al. 2015). Large population sizes from species with low levels of LD such as *D. melanogaster* somewhat circumvent this issue, and a notable feature of our experimental design is our sampling. Here, we selected the most desiccation-resistant phenotypes directly from substantial standing genetic variation where one generation of recombination should largely reflect natural haplotypes.

We observed the largest candidate allele frequency differences in loci that had intermediate allele frequencies in the control pool. This is in line with population genetics models which predict that evolution from standing genetic variation will initially be strongest for intermediate frequency alleles, which can facilitate rapid change (Long et al. 2015). However, we also observed a number of low frequency alleles which if beneficial can inflate false positives due to long-range LD (Orozco-terWengel et al. 2012; Nuzhdin and Turner 2013; Tobler et al. 2014). Nonetheless, spurious LD associations cannot fully account for our candidate list as evidenced by the desiccation functional candidates and cross-study overlap. Furthermore, chromosome inversion dynamics can also generate both short and long-range LD and contribute numerous false positive SNP phenotype associations (Hoffmann

and Weeks 2007; Tobler et al. 2014), but we found little association between the desiccation pool and inversion frequency changes, consistent with the lack of latitudinal patterns in desiccation resistance/and trait/inversion frequency associations in east Australian *D. melanogaster* (Hoffmann et al. 2001; Weeks et al. 2002).

Cross-Study Comparison Reveals a Core Set of Candidates for a Complex Trait

Our data contain many candidate genes, functions, and pathways for insect desiccation resistance, but partitioning genuine adaptive loci from experimental noise remains a significant challenge. We therefore focus our efforts on our novel comparative analysis, where we identified 45 robust candidates for further exploration. Common genetic signatures are rarely documented in cross-study comparisons of complex traits (Sarup et al. 2011; Huang et al. 2012; but c.f. Vermeulen et al. 2013), and never before for desiccation resistance. Some degree of overlap between our studies likely results from a shared genetic background from similar sampling sites, highlighting the replicability of these responses. One limitation of this approach is the overpopulation of the "core" list by longer genes. This result reflects the gene length bias inherent to GWAS approaches, an issue that is not straightforward to correct in a GWAS let alone in the comparative framework applied here. A focus on the common genes functionally linked to the desiccation phenotype can provide a meaningful framework for future research, and also suggests that some large-effect candidate genes exhibit natural variation contributing to the phenotype (table 1).

Other feasible candidates include genes expressed in the MTs (Wang et al. 2004) or essential for MT development (St Pierre et al. 2014), providing further research opportunity focused on the primary stress sensing and fluid secretion tissues. Genes include one of the most highly expressed MT transcripts CG7084, as well as *dnc*, *Nckx30C*, *slo*, *cv-c*, and *nkd*. Other aspects of stress resistance are also implicated: *CrebB*, for example, is involved in regulating insulin-regulated stress resistance and thermosensory behavior (Wang et al. 2008). *nkd* also impacts larval cuticle development, and shares a PDZ signaling domain also present in the desiccation candidate *desi* (Kawano et al. 2010; St Pierre et al. 2014). Finally, several core candidates were consistently reported in recent studies suggesting generalized roles in stress responses and fitness including transcriptome studies on MT function (Wang et al. 2004), inbreeding and cold resistance (Vermeulen et al. 2013), oxidative stress (Weber et al. 2012), and genomic studies exploring age-specific SNPs on fitness traits (Durham et al. 2014), and genes likely under spatial selection in flies from climatically divergent habitats (Fabian et al. 2012).

Several of the enriched GO categories from the core list also make biological sense, including stimulus response (almost half the suite), defense, and signaling, particularly in pathways involved in MT stress sensing (Davies et al. 2012, 2013, 2014). Functional categories enriched for sensory organ development were highlighted, consistent with environmental sensing categories from desert *Drosophila* transcriptomics (Matzkin and Markow 2009; Rajpurohit et al. 2013) and stress detection via phototransduction pathways in *D. melanogaster* under a range of stressors including desiccation (Sørensen et al. 2007).

Evidence for Conserved Signatures at Higher Level Organization

Beyond the gene level, we observed a striking degree of similarity between this study and that of Telonis-Scott et al. (2012) at the network level, where significantly overlapping protein networks were constructed from largely different seed protein sets. This supports a scenario where the "biological information flow from DNA to phenotype" (Civelek and Lusis 2014) contains inherent redundancy, where alternative genetic solutions underlie phenotypes and functions. Resistance to perturbation by genetic or environmental variation appears to increase with increasing hierarchical complexity, that is, phenotype "buffering" (Fu et al. 2009). The same functional network in different populations/genetic backgrounds can be impacted, but the exact genes and SNPs responding to perturbation will not necessarily overlap. In *Drosophila*, different sets of loci from the same founders mapped to the same PPI network in the case of two complex traits, chill coma recovery and startle response (Huang et al. 2012). Furthermore, interspecific shared networks/pathways from different genes have been documented for numerous complex traits (Emilsson et al. 2008; Aytes et al. 2014; Ichihashi et al. 2014), providing opportunities for investigating taxa where individual gene function is poorly understood.

Although the null distribution is not always obvious when comparing networks and their functions partly because of reduced sensitivity from missing interactions (Leiserson et al. 2013), here, the shared network signatures portray a plausible multifaceted stress response involving stress sensing, rapid gene accessibility, and expression. The predominance of immune system and stress response pathways are of interest with reference immunity/stress pathway "cross-talk," particularly in the MTs, where abiotic stressors elicit strong immunity transcriptional responses (Davies et al. 2012). The higher order architectures reflect crucial aspects of rapid gene expression control in response to stress, from chromatin organization (Alexander and Lomvardas 2014) to transcription, RNA decay, and translation (reviwed in de Nadal et al. 2011). In *Drosophila*, variation in transcriptional regulation of individual genes can underpin divergent desiccation phenotypes (Chung et al. 2014), or be functionally inferred from patterns of many genes elicited during desiccation stress (Rajpurohit et al. 2013). Variants with potential to modulate expression at different stages of gene expression control were evident in our GWAS data alone, for example, enrichment for 5′UTR and splice variants at the SNP level, and for chromatin and chromosome structure terms at the functional level.

Genetic Overlap Versus Genetic differences—Causes and Implications

Despite the common signatures at the gene and higher order levels between our studies, pervasive differences were observed. These can arise from differences in standing genetic variation in the natural populations where different allelic compositions impact gene-by-environment interactions and epistasis. Our network overlap analyses highlight the potential for different molecular trajectories to result in similar functions (Leiserson et al. 2013), consistent with the multiple physiological pathways potentially available to *D. melanogaster* under water stress (Chown et al. 2011). Experimental design also presumably contributed to the study differences. Strong directional selection as applied in the 2012 study can increase the frequency of rare beneficial alleles and likely reduced overall diversity than did the single-generation GWAS design. Furthermore, selection regimes often impose additional selective pressures (e.g., food deprivation during desiccation) resulting in correlated responses such as starvation resistance with desiccation resistance and these factors differ from the laboratory to the field. Finally, evidence suggests in natural *D. melanogaster* that standing genetic variation can vary extensively with sampling season (Itoh et al. 2010; Bergland et al. 2014). Although we cannot compare levels of standing variation between our two studies, evidence suggests that this may have affected the *Dys* gene, where our GWAS approach revealed no differentiation between the control and resistant pools at this locus, but rather detected the alleles associated with the selective sweep in Telonis-Scott et al. (2012) in both pools at high frequency. Whether this resulted from sampling the later collection after a long drought (2004–2008; www.bom.gov.au/climate, last accessed January 13, 2016) is unclear, but this observation highlights the relevance of temporal sampling to standing genetic variation when attempting to link complex phenotypes to genotypes.

Materials and Methods

Natural Population Sampling

Drosophila melanogaster was collected from vineyard waste at Kilchorn Winery, Romsey, Australia, in April 2010. Over 1,200 isofemale lines were established in the laboratory from wild-caught females. Species identification was conducted on male F_1 to remove *Drosophila simulans* contamination, and over 900 *D. melanogaster* isofemale lines were retained. All flies were maintained in vials of dextrose, cornmeal media at 25 °C, and constant light.

Desiccation Assay

From each isofemale line, ten inseminated F_1 females were sorted by aspiration without CO_2 and held in vials until 4–5 days old. For the desiccation assay, 10 females per line (over 9,000 flies in total) were screened by placing groups of females into empty vials covered with gauze and randomly assigning lines into 5 large sealed glass tanks containing trays of silica desiccant (relative humidity [RH] <10%). The tanks were scored for mortality hourly and the final 558 surviving
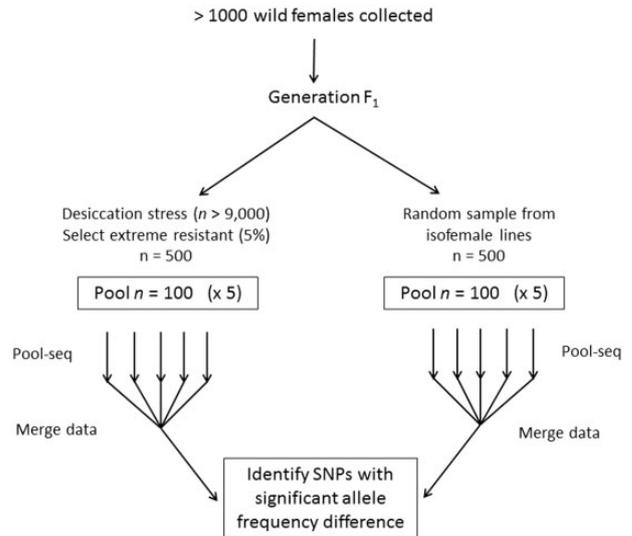

Fig. 3. Experimental design. See text for further explanation.

individuals were selected for the desiccation-resistant pool (558 flies, <6% of the total). The control pool constituted the same number of females sampled from over 200 randomly chosen isofemale lines that were frozen prior to the desiccation assay. We chose this random control approach instead of comparing the phenotypic extremes (i.e., the late mortality tail vs. the "early mortality" tail) because unfavorable environmental conditions experienced by the wild-caught dam may inflate environmental variance due to carryover effects in the F_1 offspring (Schiffer et al. 2013).

DNA Extraction and Sequencing

We used pooled whole-genome sequencing (Pool-seq) (Futschik and Schlötterer 2010) to estimate and compare allele frequency differences between the most desiccation-resistant and randomly sampled flies from a wild population to associate naturally segregating candidate SNPs with the response to low humidity. For each treatment (resistant and control), genomic DNA was extracted from 50 female heads using the Qiagen DNeasy Blood and Tissue Kit according to the manufacturer's instructions (Qiagen, Hilden, Germany). Two extractions were combined to create five pools per treatment, which were barcoded separately to create technical replicates (total: 10 × pools of 100 flies, $n = 1{,}000$). The DNA was fragmented using a CovarisS2 machine (Covaris, Inc., Woburn, MA), and libraries were prepared from 1 μg genomic DNA (per pool of 100 flies) using the Illumina TruSeq Prep module (Illumina, San Diego, CA). To minimize the effect of variation across lanes, the ten pools were combined in equal concentration (fig. 3) and run multiplexed in each lane of a full flow cell of an Illumina HiSeq 2000 sequencer. Clusters were generated using the TruSeq PE Cluster Kit v5 on a cBot, and sequenced using Illumina TruSeq SBS v3 chemistry (Illumina). Fragmentation, library construction, and sequencing were performed at the Micromon Next-Gen Sequencing facility (Monash University, Clayton, Australia).

Data Processing and Mapping

Processing was performed on a high-performance computing cluster using the Rubra pipeline system that makes use of the Ruffus Python library (Goodstadt 2010). The final variant-calling pipeline was based on a pipeline developed by Clare Sloggett (Sloggett et al. 2014) and is publicly available at https://github.com/griffinp/GWAS_pipeline (last accessed January 13, 2016). Raw sequence reads were processed per sample per lane, with P1 and P2 read files processed separately initially. Trimming was performed with trimmomatic v 0.30 (Lohse et al. 2012). Adapter sequences were also removed. Leading and trailing bases with a quality score below 30 were trimmed, and a sliding window (width = 10, quality threshold = 25) was used. Reads shorter than 40 bp after trimming were discarded. Quality before and after trimming was examined using FastQC v 0.10.1 (Andrews 2014).

Posttrimming reads that remained paired were used as input into the alignment step. Alignment to the *D. melanogaster* reference genome version r5.53 was performed with bwa v 0.6.2 (Li and Durbin 2009) using the bwa aln command and the following options: No seeding (-l 150); 1% rate of missing alignments, given 2% uniform base error rate (-n 0.01); a maximum of 2 gap opens per read (-o 2); a maximum of 12 gap extensions per read (-e 12); and long deletions within 12 bp of 3′ end disallowed (-d 12). The default values were used for all other options. Output files were then converted to sorted bam files using the bwa sampe command and the "SortSam" option in Picard v 1.96.

At this point, the bam files were merged into one file per sample using PicardMerge. Duplicates were identified with Picard MarkDuplicates. The tool RealignerTargetCreator in GATK v 2.6-5 (DePristo et al. 2011) was applied to the ten bam files simultaneously, producing a list of intervals containing probable indels around which local realignment would be performed. This was used as input for the IndelRealigner tool, which was also applied to all ten bam files simultaneously.

To investigate the variation among technical replicates, allele frequency was estimated based on read counts for each technical replicate separately, for 100,000 randomly chosen SNPs across the genome. After excluding candidate desiccation-resistance SNPs and SNPs that may have been false positives due to sequencing error (those with mean frequency <0.04 across replicates), 1,803 SNPs remained. For each locus, we calculated the variance in allele frequency estimate among the five control replicates, and the variance among the five desiccation-resistance replicates. We compared this with the variance due to pool category (control vs. desiccation resistant) using analysis of variance (ANOVA). We also calculated the mean pairwise difference in allele frequency estimate among control and among desiccation-resistance replicates, and the mean concordance correlation coefficient was calculated over all pairwise comparisons within pool category (using the epi.ccc function in the epiR package; Stevenson et al. 2015).

The following approach was then taken to identify SNPs differing between the desiccation-resistant and control groups. Based on the technical replicate analysis, we merged the five "desiccation-resistant" and the five random control samples into two bam files. A pileup file was created with samtools v. 0.1.19 (Li et al. 2009) and converted to a "sync" file using the mpileup2sync script in Popoolation2 v. 1.201 (Kofler et al. 2011), retaining reads with a mapping quality of at least 20 and bases with a minimum quality score of 20. Reads with unmapped mates were also retained (option −A in samtools mpileup). For this step, we used a repeat-masked version of the reference genome created with RepeatMasker 4.0.3 (Smit et al. 2013-2015) and a repeat file containing all annotated transposons from *D. melanogaster* including shared ancestral sequences (Flybase release r5.57) plus all repetitive elements from RepBase release v19.04 for *D. melanogaster*. Simple repeats, small RNA genes, and bacterial insertions were not masked (options −nolow −norna −no_is). The rmblast search engine was used.

Association Analysis

We then performed Fisher's exact test with the Fisher test script in Popoolation2 (Kofler et al. 2011). We set the minimum count (of the minor allele) = 10, minimum coverage (in both populations) = 20, and maximum coverage = 10,000. The test was performed for each SNP (sliding window off). Although there are numerous approaches to calculating genome-wide significance thresholds in a Poolseq framework, currently there is no consensus on an appropriate, standardized statistical method. Methods include simulation approaches (Orozco-terWengel et al. 2012; Bastide et al. 2013; Burke et al. 2014), correction using a null distribution based on estimating genetic drift from replicated artificial selection line allele frequency variance (Turner and Miller 2012), use of a simple threshold based on a minimum allele frequency difference (Turner et al. 2010), and the FDR correction suggested by Storey and Tibshirani (2003) (used by Fabian et al. 2012).

We determined based on our design to calculate the FDR threshold by comparison with a simulated distribution of null *P* values following the approach of Bastide et al. (2013). For each SNP locus, a *P* value was calculated by simulating the distribution of read counts between the major and minor allele according to a beta-binomial distribution with mean $\alpha/(\alpha+\beta)$ and variance $(\alpha\beta)/((\alpha+\beta)^2(\alpha+\beta+1))$. This model allowed for two types of sampling variation: Stochastic variation in allele sampling across the phenotypes and the background variation in the representation of alleles in each pool caused by library preparation artifacts or stochastic variation driven by unequal coverage of the pools. These sources of variation have also been recognized elsewhere as being important in interpreting pooled sequence allele calls and identifying significant differentiation between pooled samples (Lynch et al. 2014). The value of $\alpha = 37$ was chosen by chi-square test as the best match to the observed *P* value distribution ($\alpha = 10$ to $\alpha = 40$ were tested), with a corresponding $\beta = 43.15$ to reflect the coverage depth difference between the random control ($487\times$) and desiccation-resistant ($568\times$) pools. These values were then used to simulate a null data set $10\times$ the size of the observed data

set. Because the null P value distribution still did not fit the observed P value distribution particularly well (supplementary fig. S2, Supplementary Material online), these values were not used for a standard FDR correction. Instead, a P value threshold was calculated based on the lowest 0.05% of the null P values similar in approach to Orozco-terWengel et al. (2012).

Inversion Analysis

In *D. melanogaster*, several cosmopolitan inversion polymorphisms show cross-continent latitudinal patterns that explain a substantial proportion of clinal variation for many traits including thermal tolerance (reviewed in Hoffmann and Weeks 2007). Patterns of disequilibria generated between loci in the vicinity of the rearrangement can obscure signals of selection (Hoffmann and Weeks 2007), and inversion frequencies are routinely considered in genomic studies of natural populations sampled from known latitudinal clines (Fabian et al. 2012; Kapun et al. 2014). We assessed associations between inversion frequencies and our control and resistant pools at SNPs diagnostic for the following inversions: In3R(Payne) (Anderson et al. 2005; Kapun et al. 2014); In(2L)t (Andolfatto et al. 1999; Kapun et al. 2014); In(2R)Ns, In(3L)P, In(3R)C, In(3R)K, and In(3R)Mo (Kapun et al. 2014). Between one and five diagnostic SNPs were investigated for each inversion (depending on availability). We considered an inversion to be associated with desiccation resistance if the frequency of its diagnostic SNP differed significantly between the control and desiccation-resistant pools by a Fisher's exact test.

Genome Feature Annotation

The list of candidate SNPs was converted to vcf format using a custom R script, setting the alternate allele as the allele that increased in frequency from the control to the desiccation-resistant pool, and setting the reference allele as the other allele present in cases where both the major and minor alleles differed from the reference genome. The genomic features in which the candidate SNPs occurred were annotated using snpEff v4.0, first creating a database from the *D. melanogaster* v5.53 genome annotation, with the default approach (Cingolani et al. 2012). The default annotation settings were used, except for setting the parameter "-ud 0" to annotate SNPs to genes only if they were within the gene body. Using χ^2 tests on SNP counts, we tested for over- or under-enrichment of feature types by calculating the proportion of 'all' features from all SNPs detected in the study and the proportion of "candidate" SNPs identified as differentiated in desiccation-resistant flies (Fabian et al. 2012).

GO Analysis

GO analyses was used to associate the candidates with their biological functions (Ashburner et al. 2000). Previously developed for biological pathway analysis of global gene expression studies (Wang et al. 2010), typical GO analysis assumes that all genes are sampled independently with equal probability. GWAS violate these criteria where SNPs are more likely to

occur in longer genes than in shorter genes, resulting in higher type I error rates in GO categories overpopulated with long genes. Gene length as well as overlapping gene biases (Wang et al. 2011) were accounted for using Gowinda (Kofler and Schlötterer 2012), which implements a permutation strategy to reduce false positives. Gowinda does not reconstruct LD between SNPs but operates in two modes underpinned by two extreme assumptions: 1) Gene mode (assumes all SNPs in a gene are in LD) and 2) SNP mode (assumes all SNPs in a gene are completely independent).

The candidates were examined in both modes, but given the stringency of gene mode, high-resolution SNP mode output was utilized for the final analysis with the following parameters: 10^6 simulations, minimum significance $=1$, and minimum genes per GO category $= 3$ (to exclude gene-poor categories). P values were corrected using the Gowinda FDR algorithm. GO annotations from Flybase r5.57 were applied.

Candidate Gene Analysis

In conjunction with the GO analysis, we also examined known candidate genes associated with survival under low RH to better characterize possible biological functions of our GWAS candidates. We took a conservative approach to candidate gene assignment, where the SNPs were mapped within the entire gene region without up- and downstream extensions. Genes were curated from FlyBase and the literature (table 1), predominantly using detailed functional experimental evidence as criteria for functional desiccation candidates.

Network Analysis

We next obtained a higher level summary of the candidate gene list using PPI network analysis. The full candidate gene list was used to construct a first-order interaction network using all *Drosophila* PPIs listed in the *Drosophila* Interaction Database v 2014_10 (avoiding interologs) (Murali et al. 2011). This "full" PPI network includes manually curated data from published literature and experimental data for 9,633 proteins and 93,799 interactions. For the network construction, the igraph R package was used to build a subgraph from the full PPI network containing the candidate proteins and their first-order interacting proteins (Csardi and Nepusz 2006). Self-connections were removed. Functional enrichment analysis was conducted on the list of network genes using FlyMine v40.0 (www.flymine.org, last accessed January 13, 2016) with the following parameters for GO enrichment and KEGG/Reactome pathway enrichment: Benjamini–Hochberg FDR correction $P < 0.05$, background $=$ all *D. melanogaster* genes in our full PPI network, and Ontology (for GO enrichment only) $=$ biological process or molecular function.

Cross-Study Comparison: Overlap with Our Previous Work

Following on from the comparison between our gene list and the candidate genes suggested from the literature, we quantitatively evaluated the degree of overlap between the candidate genes identified in this study with alleles mapped in our previous study of flies collected from the same geographic

region (Telonis-Scott et al. 2012). The "overlap" analysis was performed similarly for the GWAS candidate gene and network-level approaches. First, the gene list overlap was tested between the two studies using Fisher's exact test, where the 382 genes identified from the candidate SNPs (this study) were compared with 416 genes identified from candidate single feature polymorphisms (Telonis-Scott et al. 2012). The contingency table was constructed using the number of annotated genes in the *D. melanogaster* R5.53 genome built as an approximation for the total number of genes tested in each study, and took the following form:

	Not in GWAS	In GWAS
Not in Telonis-Scott et al. (2012)	N – intersect-A–B	A-intersect
In Telonis-Scott et al. (2012)	B-intersect	Intersect(A,B)

where A = gene list from Telonis-Scott et al. (2012), $n = 416$; B = gene list from this study, $n = 382$; N = total number of genes in *D. melanogaster* reference genome r5.53, $n = 17,106$.

To annotate genome features in the overlap candidate gene list, we used the SNP data (this study), which better resolves alleles than array-based genotyping. Due to this limitation, we did not compare exact genomic coordinates between the studies, but considered overlap to occur when differentiated alleles were detected in the same gene in both studies. We did however examine the allele frequencies around the *Dys-RF* promoter, which was sequenced separately revealing a selective sweep in the experimental evolution lines (Telonis-Scott et al. 2012).

The overlapping candidates were tested for functional enrichment using a basic gene list approach. As genes rather than SNPs were analyzed here, Fly Mine v40.0 (www.flymine.org) was applied with the following parameters: Gene length normalization, Benjamini–Hochberg FDR correction $P < 0.05$, Ontology = biological process, and default background (all GO-annotated *D. melanogaster* genes).

Finally, we constructed a second PPI network from the candidate genes mapped in Telonis-Scott et al. (2012) to examine system-level overlap between the two studies. The 416 genes mapped to 337 protein seeds and the network was constructed as described above. The degree of network overlap between the two studies was tested using simulation. In each iteration, a gene set of the same length as the list from this study and a gene set of the same length as the list from Telonis-Scott et al. (2012) were resampled randomly from the entire *D. melanogaster* mapped gene annotation (r5.53). Each resampled set was then used to build a first-order network as previously described. The overlap network between the two resampled sets was calculated using the graph.intersection command from the igraph package, and zero-degree nodes (proteins with no interactions) were removed. Overlap was quantified by counting the number of nodes and the number of edges in this overlap network. The observed overlap was compared with a simulated null distribution for both node and edge number, where $n = 1,000$ simulation iterations.

Supplementary Material

Acknowledgments

References

Alexander JM, Lomvardas S. 2014. Nuclear architecture as an epigenetic regulator of neural development and function. *Neuroscience* 264:39–50.

Anderson AR, Hoffmann AA, McKechnie SW, Umina PA, Weeks AR. 2005. The latitudinal cline in the In(3R)Payne inversion polymorphism has shifted in the last 20 years in Australian *Drosophila melanogaster* populations. *Mol Ecol.* 14:851–858.

Andolfatto P, Wall JD, Kreitman M. 1999. Unusual haplotype structure at the proximal breakpoint of In(2L)t in a natural population of *Drosophila melanogaster*. Genetics 153:1297–1311.

Andrews S. 2014. FastQC. Available from: http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 25:25–29.

Aytes A, Mitrofanova A, Lefebvre C, Alvarez Mariano J, Castillo-Martin M, Zheng T, Eastham James A, Gopalan A, Pienta Kenneth J, Shen MM, et al. 2014. Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. *Cancer Cell* 25:638–651.

Bastide H, Betancourt A, Nolte V, Tobler R, Stöbe P, Futschik A, Schlötterer C. 2013. A genome-wide, fine-scale map of natural pigmentation variation in *Drosophila melanogaster*. *PLoS Genet.* 9:e1003534

Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. 2014. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. *PLoS Genet.* 10:e1004775

Bonebrake TC, Mastrandrea MD. 2010. Tolerance adaptation and precipitation changes complicate latitudinal patterns of climate change impacts. *Proc Natl Acad Sci U S A.* 107:12581–12586.

Bradley TJ. 2009. Animal osmoregulation. Oxford: Oxford University Press.

Burke MK, Liti G, Long AD. 2014. Standing genetic variation drives repeatable experimental evolution in outcrossing populations of *Saccharomyces cerevisiae*. *Mol Biol Evol.* 31:3228–3239.

Chown SL. 2012. Trait-based approaches to conservation physiology: forecasting environmental change risks from the bottom up. *Philos Trans R Soc B.* 367:1615–1627.

Chown SL, Nicolson SW. 2004. Insect physiological ecology: mechanisms and patterns. Oxford: Oxford University Press.

Chown SL, Sørensen JG, Terblanche JS. 2011. Water loss in insects: an environmental change perspective. *J Insect Physiol.* 57:1070–1084.

Chung H, Loehlin DW, Dufour HD, Vaccarro K, Millar JG, Carroll SB. 2014. A single gene affects both ecological divergence and mate choice in *Drosophila*. *Science* 343:1148–1151.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6:80–92.

Civelek M, Lusis AJ. 2014. Systems genetics approaches to understand complex traits. *Nat Rev Genet.* 15:34–48.

Clusella-Trullas S, Blackburn TM, Chown SL. 2011. Climatic predictors of temperature performance curve parameters in ectotherms imply complex responses to climate change. *Am Nat.* 177:738–751.

Csardi G, Nepusz T. 2006. The igraph software package for complex network research. *InterJournal, Complex Systems* 1695. Available from: http://igraph.org.

Davies SA, Cabrero P, Overend G, Aitchison L, Sebastian S, Terhzaz S, Dow JA. 2014. Cell signalling mechanisms for insect stress tolerance. *J Exp Biol.* 217:119–128.

Davies SA, Cabrero P, Povsic M, Johnston NR, Terhzaz S, Dow JA. 2013. Signaling by *Drosophila* capa neuropeptides. *Gen Comp Endocrinol.* 188:60–66.

Davies SA, Overend G, Sebastian S, Cundall M, Cabrero P, Dow JA, Terhzaz S. 2012. Immune and stress response 'cross-talk' in the *Drosophila* Malpighian tubule. *J Insect Physiol.* 58:488–497.

Day JP, Dow JA, Houslay MD, Davies SA. 2005. Cyclic nucleotide phosphodiesterases in *Drosophila melanogaster*. *Biochem J.* 388:333–342.

de Nadal E, Ammerer G, Posas F. 2011. Controlling gene expression in response to stress. *Nat Rev Genet.* 12:833–845.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 43:491–498.

Djawdan M, Chippindale AK, Rose MR Bradley TJ. 1998. Metabolic reserves and evolved stress resistance in *Drosophila melanogaster*. *Physiol Zool.* 71:584–594.

Durham MF, Magwire MM, Stone EA, Leips J. 2014. Genome-wide analysis in *Drosophila* reveals age-specific effects of SNPs on fitness traits. *Nat Commun.* 5:4338

Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, Carlson S, Helgason A, Walters GB, Gunnarsdottir S, et al. 2008. Genetics of gene expression and its effect on disease. *Nature* 452:423–428.

Fabian DK, Kapun M, Nolte V, Kofler R, Schmidt PS, Schlötterer C, Flatt T. 2012. Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. *Mol Ecol.* 21:4748–4769.

Foley BR, Telonis-Scott M. 2011. Quantitative genetic analysis suggests causal association between cuticular hydrocarbon composition and desiccation survival in *Drosophila melanogaster*. *Heredity* 106:68–77.

Folk DG, Han C, Bradley TJ. 2001. Water acquisition and partitioning in *Drosophila melanogaster*: effects of selection for desiccation-resistance. *J Exp Biol.* 204:3323–3331.

Fowler MA, Montell C. 2013. *Drosophila* TRP channels and animal behavior. *Life Sci.* 92:394–403.

Fu J, Keurentjes JJB, Bouwmeester H, America T, Verstappen FWA, Ward JL, Beale MH, de Vos RCH, Dijkstra M, Scheltema RA, et al. 2009. System-wide molecular evidence for phenotypic buffering in *Arabidopsis*. *Nat Genet.* 41:166–167.

Futschik A, Schlötterer C. 2010. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics* 186:207–218.

Gibbs AG, Fukuzato F, Matzkin LM. 2003. Evolution of water conservation mechanisms in *Drosophila*. *J Exp Biol.* 206:1183–1192.

Gibbs AG, Matzkin LM. 2001. Evolution of water balance in the genus *Drosophila*. *J Exp Biol.* 204:2331–2338.

Goodstadt L. 2010. Ruffus: a lightweight Python library for computational pipelines. *Bioinformatics* 26:2778–2779.

Hadley NF. 1994. Water relations of terrestrial arthropods. San Diego (CA): Academic Press.

Hoffmann AA, Hallas R, Sinclair C, Mitrovski P. 2001. Levels of variation in stress resistance in *Drosophila* among strains, local populations, and geographic regions: patterns for desiccation, starvation, cold resistance, and associated traits. *Evolution* 55:1621–1630.

Hoffmann AA, Parsons PA. 1989a. An integrated approach to environmental stress tolerance and life-history variation: desiccation tolerance in *Drosophila*. *Biol J Linn Soc.* 37:117–136.

Hoffmann AA, Parsons PA. 1989b. Selection for increased desiccation resistance in *Drosophila melanogaster*: additive genetic control and correlated responses for other stresses. *Genetics* 122:837–845.

Hoffmann AA, Sgrò CM. 2011. Climate change and evolutionary adaptation. *Nature* 470:479–485.

Hoffmann AA, Weeks AR. 2007. Climatic selection on genes and traits after a 100 year-old invasion: a critical look at the temperate-tropical clines in *Drosophila melanogaster* from eastern Australia. *Genetica* 129:133–147.

Huang W, Richards S, Carbone MA, Zhu D, Anholt RR, Ayroles JF, Duncan L, Jordan KW, Lawrence F, Magwire MM, et al. 2012. Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proc Natl Acad Sci U S A.* 109:15553–15559.

Huang Z, Tunnacliffe A. 2004. Response of human cells to desiccation: comparison with hyperosmotic stress response. *J Physiol.* 558:181–191.

Huang Z, Tunnacliffe A. 2005. Gene induction by desiccation stress in human cell cultures. *FEBS Lett.* 579:4973–4977.

Ichihashi Y, Aguilar-Martínez JA, Farhi M, Chitwood DH, Kumar R, Millon LV, Peng J, Maloof JN, Sinha NR. 2014. Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. *Proc Natl Acad Sci U S A.* 111:E2616–E2621.

Itoh M, Nanba N, Hasegawa M, Inomata N, Kondo R, Oshima M, Takano-Shimizu T. 2010. Seasonal changes in the long-distance linkage disequilibrium in *Drosophila melanogaster*. *J Hered.* 101:26–32.

Johnson WA, Carder JW. 2012. *Drosophila* nociceptors mediate larval aversion to dry surface environments utilizing both the *painless* TRP channel and the DEG/ENaC subunit, *PPK1*. *PLoS One* 7:e32878

Kahsai L, Kapan N, Dircksen H, Winther AM, Nassel DR. 2010. Metabolic stress responses in *Drosophila* are modulated by brain neurosecretory cells that produce multiple neuropeptides. *PLoS One* 5:e11480

Kapun M, van Schalkwyk H, McAllister B, Flatt T, Schlötterer C. 2014. Inference of chromosomal inversion dynamics from Pool-Seq data in natural and laboratory populations of *Drosophila melanogaster*. *Mol Ecol.* 23:1813–1827.

Kawano T, Shimoda M, Matsumoto H, Ryuda M, Tsuzuki S, Hayakawa Y. 2010. Identification of a gene, *Desiccate*, contributing to desiccation resistance in *Drosophila melanogaster*. *J Biol Chem.* 285:38889–38897.

Kellermann V, van Heerwaarden B, Sgrò CM, Hoffmann AA. 2009. Fundamental evolutionary limits in ecological traits drive *Drosophila* species distributions. *Science* 325:1244–1246.

Kofler R, Nolte V, Schlötterer C. 2016. The impact of library preparation protocols on the consistency of allele frequency estimates in Pool-Seq data. *Mol Ecol Resour.* 16:118–122.

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

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

Leiserson MDM, Eldridge JV, Ramachandran S, Raphael BJ. 2013. Network analysis of GWAS data. *Curr Opin Genet Dev.* 23:602–610.

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

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.

Liu Y, Luo J, Carlsson MA, Nässel DR. 2015. Serotonin and insulin-like peptides modulate leucokinin-producing neurons that affect feeding and water homeostasis in *Drosophila*. *J Comp Neurol*. 523:1840–1863.

Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B. 2012. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res*. 40:W622–W627.

Long A, Liti G, Luptak A, Tenaillon O. 2015. Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nat Rev Genet*. 16:567–582.

Lynch M, Bost D, Wilson S, Maruki T, Harrison S. 2014. Population-genetic inference from pooled-sequencing data. *Genome Biol Evol*. 6:1210–1218.

Mackay TFC, Stone EA, Ayroles JF. 2009. The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet*. 10:565–577.

Marron MT, Markow TA, Kain KJ, Gibbs AG. 2003. Effects of starvation and desiccation on energy metabolism in desert and mesic *Drosophila*. *J Insect Physiol*. 49:261–270.

Matzkin LM, Markow TA. 2009. Transcriptional regulation of metabolism associated with the increased desiccation resistance of the cactophilic *Drosophila mojavensis*. *Genetics* 182:1279–1288.

Murali T, Pacifico S, Yu J, Guest S, Roberts GG, Finley RL Jr. 2011. DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for *Drosophila*. *Nucleic Acids Res*. 39:D736–D743.

Nuzhdin SV, Turner TL. 2013. Promises and limitations of hitchhiking mapping. *Curr Opin Genet Dev*. 23:694–699.

Orozco-terWengel P, Kapun M, Nolte V, Kofler R, Flatt T, Schlötterer C. 2012. Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. *Mol Ecol*. 21:4931–4941.

Picard. Picard A set of Java command line tools for manipulating high-throughput seqeuncing (HTS) data and formats. Available from: http://broadinstitute.github.io/picard.

Qiu Y, Tittiger C, Wicker-Thomas C, Le Goff G, Young S, Wajnberg E, Fricaux T, Taquet N, Blomquist GJ, Feyereisen R. 2012. An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. *Proc Natl Acad Sci U S A* 109:14858–14863.

Rajpurohit S, Oliveira CC, Etges WJ, Gibbs AG. 2013. Functional genomic and phenotypic responses to desiccation in natural populations of a desert Drosophilid. *Mol Ecol*. 22:2698–2715.

Sarup P, Sørensen JG, Kristensen TN, Hoffmann AA, Loeschcke V, Paige KN, Sørensen P. 2011. Candidate genes detected in transcriptome studies are strongly dependent on genetic background. *PLoS One* 6:e15644

Schiffer M, Hangartner S, Hoffmann AA. 2013. Assessing the relative importance of environmental effects, carry-over effects and species differences in thermal stress resistance: a comparison of Drosophilids across field and laboratory generations. *J Exp Biol*. 216:3790–3798.

Schlötterer C, Kofler R, Versace E, Tobler R, Franssen SU. 2015. Combining experimental evolution with next-generation sequencing: A powerful tool to study adaptation from standing genetic variation. *Heredity* 114:431–440.

Sinclair BJ, Gibbs AG, Roberts SP. 2007. Gene transcription during exposure to, and recovery from, cold and desiccation stress in *Drosophila melanogaster*. *Insect Mol Biol*. 16:435–443.

Sloggett C, Wakefield M, Philip G, Pope B. 2014. Rubra—flexible distributed pipelines. figshare. Available from: http://dx.doi.org/10.6084/m9.figshare.895626.

Smit AFA, Hubley R, Green P. 2013-2015. RepeatMasker Open-4.0.3. Available from: http://www.repeatmasker.org.

Söderberg JA, Birse RT, Nässel DR. 2011. Insulin production and signaling in renal tubules of *Drosophila* is under control of tachykinin-related peptide and regulates stress resistance. *PLoS One* 6:e19866

Sørensen JG, Nielsen MM, Loeschcke V. 2007. Gene expression profile analysis of *Drosophila melanogaster* selected for resistance to environmental stressors. *J Evol Biol*. 20:1624–1636.

St Pierre SE, Ponting L, Stefancsik R, McQuilton P. 2014. FlyBase 102—advanced approaches to interrogating FlyBase. *Nucleic Acids Res*. 42:D780–D788.

Stevenson M, Nunes T, Heuser C, Marshall J, Sanchez J, Thornton R, Reiczigel J, Robison-Cox J, Sebastiani P, Solymos P, et al. 2015. epiR: Tools for the analysis of epidemiological data. R package version 0.9-62. Available from: http://CRAN.R-project.org/package=epiR.

Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 100:9440–9445.

Telonis-Scott M, Gane M, DeGaris S, Sgrò CM, Hoffmann AA. 2012. High resolution mapping of candidate alleles for desiccation resistance in *Drosophila melanogaster* under selection. *Mol Biol Evol*. 29:1335–1351.

Telonis-Scott M, Guthridge KM, Hoffmann AA. 2006. A new set of laboratory-selected *Drosophila melanogaster* lines for the analysis of desiccation resistance: response to selection, physiology and correlated responses. *J Exp. Biol*. 209:1837–1847.

Terhzaz S, Cabrero P, Robben JH, Radford JC, Hudson BD, Milligan G, Dow JA, Davies SA. 2012. Mechanism and function of *Drosophila* capa GPCR: a desiccation stress-responsive receptor with functional homology to human neuromedinU receptor. *PLoS One* 7:e29897

Terhzaz S, Overend G, Sebastian S, Dow JA, Davies SA. 2014. The *D. melanogaster* capa-1 neuropeptide activates renal NF-kB signaling. *Peptides* 53:218–224.

Terhzaz S, Teets NM, Cabrero P, Henderson L, Ritchie MG, Nachman RJ, Dow JA, Denlinger DL, Davies SA. 2015. Insect capa neuropeptides impact desiccation and cold tolerance. *Proc Natl Acad Sci U S A*. 112:2882–2887.

Thorat LJ, Gaikwad SM, Nath BB. 2012. Trehalose as an indicator of desiccation stress in *Drosophila melanogaster* larvae: a potential marker of anhydrobiosis. *Biochem Biophys Res Commun*. 419:638–642.

Tobler R, Franssen SU, Kofler R, Orozco-terWengel P, Nolte V, Hermisson J, Schlötterer C. 2014. Massive habitat-specific genomic response in *D. melanogaster* populations during experimental evolution in hot and cold environments. *Mol Biol Evol*. 31:364–375.

Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV. 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat Genet* 42:260–263.

Turner TL, Miller PM. 2012. Investigating natural variation in *Drosophila* courtship song by the evolve and resequence approach. *Genetics* 191:633–642.

Vermeulen CJ, Sørensen P, Kirilova Gagalova K, Loeschcke V. 2013. Transcriptomic analysis of inbreeding depression in cold-sensitive *Drosophila melanogaster* shows upregulation of the immune response. *J Evol Biol*. 26:1890–1902.

Wang B, Goode J, Best J, Meltzer J, Schilman PE, Chen J, Garza D, Thomas JB, Montminy M. 2008. The insulin-regulated CREB coactivator TORC promotes stress resistance in *Drosophila*. *Cell Metab*. 7:434–444.

Wang J, Kean L, Yang J, Allan AK, Davies SA, Herzyk P, Dow JA. 2004. Function-informed transcriptome analysis of *Drosophila* renal tubule. *Genome Biol*. 5:R69

Wang K, Li M, Hakonarson H. 2010. Analysing biological pathways in genome-wide association studies. *Nat Rev Genet*. 11:843–854.

Wang L, Jia P, Wolfinger RD, Chen X, Zhao Z. 2011. Gene set analysis of genome-wide association studies: methodological issues and perspectives. *Genomics* 98:1–8.

Weber AL, Khan GF, Magwire MM, Tabor CL, Mackay TF, Anholt RR. 2012. Genome-wide association analysis of oxidative stress resistance in *Drosophila melanogaster*. *PLoS One* 7:e34745

Weeks AR, McKechnie SW, Hoffmann AA. 2002. Dissecting adaptive clinal variation: markers, inversions and size/stress associations in *Drosophila melanogaster* from a central field population. *Ecol Lett*. 5:756–763.

Zhu Y, Bergland AO, González J, Petrov DA. 2012. Empirical validation of pooled whole genome population re-sequencing in *Drosophila melanogaster*. *PLoS One* 7:e41901.