High Resolution Mapping of Candidate Alleles for Desiccation Resistance in *Drosophila melanogaster* under Selection

Marina Telonis-Scott,*[,1] Madeleine Gane,[1] Sarah DeGaris,[1] Carla M. Sgrò,[2] and Ary A. Hoffmann[1]

[1]Department of Genetics, Bio21 Institute, The University of Melbourne, Parkville, Melbourne, Australia

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

***Corresponding author:** E-mail: marinats@unimelb.edu.au.

Associate editor: John H. McDonald

Abstract

The ability to counter periods of low humidity is an important determinant of distribution range in *Drosophila*. Climate specialists with low physiological tolerance to desiccation stress are restricted to the tropics and may lack the ability to further increase resistance through evolution. Although the physiological adaptations to desiccation stress are well studied in *Drosophila* and other ectotherms, factors underlying evolutionary responses remain unknown because of a paucity of genetic data. We address this issue by mapping evolutionary shifts in *D. melanogaster* under selection for desiccation resistance. Genomic DNA from five independent replicate selected, and control lines were hybridized to high density Affymetrix *Drosophila* tiling arrays resulting in the detection of 691 single feature polymorphisms (SFPs) differing between the treatments. While randomly distributed throughout the genome, the SFPs formed specific clusters according to gene ontology. These included genes involved in ion transport and respiratory system development that provide candidates for evolutionary changes involving excretory and respiratory water balance. Changes to genes related to neuronal control of cell signaling, development, and gene regulation provide candidates to explore novel biological processes in stress resistance. Sequencing revealed the nucleotide shifts in a subset of the SFPs and highlighted larger regions of genomic diversity surrounding SFPs. The association between natural desiccation resistance and a 463-bp region of the 5′ promoter region of the *Dys* gene undergoing allele frequency changes in response to selection in the experimental evolution lines was tested in an independent population from Coffs Harbour, Australia. The allele frequencies of 23 SNPs common to the two populations were inferred from the parents of the 10% most and 10% least resistant Coffs Harbour flies. The frequencies of the selected alleles were higher at all sites, with three sites significantly associated with the resistant Coffs Harbour flies. This study illustrates how rapid mapping can be used for discovering natural molecular variants associated with survival to low humidity and provides a wealth of candidate alleles to explore the genetic basis of physiological differences among resistant and susceptible *Drosophila* populations and species.

Key words: *Drosophila*, desiccation, candidate-alleles, natural-adaptation.

Research article

Introduction

Insect distributions are commonly linked to physiological resistance traits including desiccation stress (van Herrewege and David 1997; Le Lagadec et al. 1998; Addo-Bediako et al. 2001). In *Drosophila*, distributional patterns have been connected with genetic variation for physiological tolerances to traits, such as desiccation resistance and cold resistance (Kellermann et al. 2009; Boher et al. 2010; Strachan et al. 2011). The narrow latitudinal range of tropical rainforest species parallels low phenotypic and genetic variances for desiccation resistance and reflects fundamental evolutionary limits that appear independent of neutral variation or phylogeny (Hoffmann et al. 2003; Kellermann et al. 2009). Predicted rainfall patterns leading to drier tropical conditions are anticipated to pose specific challenges to climate specialists that are evolutionarily constrained to extend their range beyond perennially humid habitats (Hoffmann 2010). Variation in distribution-limiting traits such as desiccation resistance is therefore an important indicator of species potential to adapt to changing climatic conditions, particularly for species living close to their physiological limits (Hoffmann and Sgro 2011).

Desiccation resistance is a complex adaptation that can arise through multiple mechanisms including 1) lowered water loss rates, 2) increased bulk water stores, and 3) tolerating more water loss (dehydration tolerance) (Gibbs and Matzkin 2001). The ability to retain more water underlies numerous desert invasions of cactophilic *Drosophila* primarily by reducing respiratory water loss rates to conserve water content (Gibbs and Matzkin 2001; Marron et al. 2003). Low metabolic rates may also permit greater carbohydrate energy utilization during desiccation in desert *Drosophila* despite similar basal reserves as their mesic congeners (Marron et al. 2003). Water retention is also a common mechanism underlying adaptation to simulated laboratory "deserts," although selected populations often follow different evolutionary trajectories depending on their genetic backgrounds and selection regimes (Telonis-Scott et al. 2006). For example, varied patterns of water balance and resource partitioning in selected *D. melanogaster* populations include sequestration of water to the hemolymph (Chippindale et al. 1998; Folk et al. 2001), increased carbohydrate (glycogen) storage (Chippindale et al. 1998), reduced metabolic rate (Hoffmann and Parsons 1989), and greater dehydration

tolerance (Telonis-Scott et al. 2006). Moreover, dehydration tolerance in desert *Drosophila* can be phylogenetically constrained (Marron et al. 2003).

Despite the well-characterized physiology and ecology of desiccation resistance in *Drosophila*, surprisingly little is known about the genetic basis underlying the evolutionary divergence among climatically adapted species. Single gene approaches in *D. melanogaster* have isolated loci with large effects on water balance in adults and wandering larvae, *parched* and *desiccate*, respectively (Kimura et al. 1985; Kawano et al. 2010), and have linked the cold stress candidates *frost* and *smp-30* with the desiccation response (Sinclair et al. 2007). Genome-wide studies suggest that adaptations to low humidity environments may involve more complex genetic architectures. Quantitative trait loci (QTL) mapping identified at least 15 genomic regions contributing to natural variation for desiccation survival in *D. melanogaster* (Foley and Telonis-Scott 2011). At the transcriptome level, desiccation stress elicited differential expression patterns in over 1,000 genes in the desert endemic *D. mojavensis* (Matzkin and Markow 2009) and altered the basal expression of more than 200 genes in experimental evolution populations of *D. melanogaster* (Sorensen et al. 2007). Although genome-wide studies associate desiccation resistance with intervals containing thousands of annotated genes (QTL analysis), or with expression variation in hundreds of genes (microarrays), variation at the allelic level remains poorly defined. Single-gene studies attempt to bridge this gap but are limited in scope given that natural continuous phenotypes are rarely reflected in the discrete form of laboratory mutants and isolates. As such, important questions still remain including why only some species lack the genetic potential to adapt to low-humidity habitats, and specifically, to discover the molecular basis of variation in this trait.

Here, we begin to address these questions by utilizing a rapid and effective genome-wide strategy to systematically map candidate alleles associated with evolutionary shifts in desiccation resistance in *D. melanogaster*. This species is a widespread cosmopolitan climate generalist with naturally high levels of desiccation resistance (David et al. 2004; Kellermann et al. 2009); importantly, it is also a fully sequenced model organism. *Drosophila melanogaster* exhibits substantial evolutionary capacity to adapt to low humidity environments, and studies on this species have provided much insight into the evolutionary physiology of desiccation resistance (Hoffmann and Harshman 1999). We used a short intensive artificial selection treatment to generate highly desiccation divergent lines from a large recently field-derived population. The selected lines survived desiccation on average over 30% longer than the controls, retained 14% more water after 3 h, and stored 6% and 57% more water and glycogen than the controls, respectively (DeGaris S, unpublished data).

The applications of array-based genotyping are multifold and include differentiating between inbred and wild *Arabidopsis* strains (Borevitz et al. 2003, 2007), cryptic mosquito strains (Turner et al. 2005), and highly diverse cline endpoints in *D. melanogaster* (Turner et al. 2008), as well as mapping QTL for longevity between recombinant inbred lines in *D. melanogaster* (Lai et al. 2007). Nevertheless, this approach has been rarely used to map allele frequency changes contributing to the selection response of complex traits. Nuzhdin et al. (2007) inferred QTL contributing to the selection response for starvation resistance in *D. melanogaster* using expression arrays to identify linked markers undergoing allele frequency shifts. Although a more powerful multiallelic approach compared with traditional hitchhiking mapping studies, a more comprehensive mapping platform such as the tiling array would have enabled greater resolution of small selective sweeps (Nuzhdin et al. 2007).

Our design, involving multiple replicate lines combined with comprehensive genome coverage, permitted the detection of at least three regions of chromosome 3R undergoing multiple allele frequency changes in response to selection for desiccation resistance. We used high-density tiling arrays to screen over 2 million loci across the *D. melanogaster* genome, and this process mapped shifts following selection to 691 single-feature polymorphisms (SFPs). We linked SNPs from one of these regions showing evidence of a selective sweep at $<$0.5 kb of the *Dys* gene with natural variation for desiccation resistance and resolved several small haplotype blocks in this region. These data represent the essential next step in connecting laboratory evolution to natural resistance evolution in *Drosophila* and provide new candidate regions for variation in a trait of ecological significance.

Materials and Methods

Drosophila melanogaster Lines and Culture

The artificial selection and control lines were founded from *D. melanogaster* sampled from southern Victoria in 2008. The offspring of 100 field-collected females were pooled and mass bred for two generations in the laboratory prior to the first selection at generation F_3. The experimental flies were reared under controlled density conditions by removing parents after 6 h of oviposition. Virgin F_3 flies were separated by sex under light CO_2 anesthesia at 6 h intervals and allowed at least 48 h recovery prior to desiccation stress. Flies were desiccated in groups of 25 in glass vials topped with gauze in sealed glass tanks containing silica desiccant (RH $<$ 10%) until the approximate LT_{50} of 1,000 flies of each sex. The survivors were randomly allocated into six replicate lines comprised of 70–80 flies of each sex (150–160 flies total per replicate). The controls were established in the same manner as the selected lines barring the substitution of water for desiccant, and an equivalent number of flies were randomly allocated into six replicate lines at the end of the desiccation stress. Selection on subsequent generations was more stringent, with the last 10–20% of surviving males and females used to start the following generation. Although we endeavored to select from 1,000 flies of each sex per replicate line, the number of virgin progeny collected from the preceding generation was sometimes fewer than 2,000 per line. Therefore, the proportion of selected individuals could vary from generation to generation, however no less than

80 flies of each sex were used to start the next generation for selection, (i.e., 160 virgin males and females). The proportion of flies selected, strength of selection, and sex ratios were kept consistent between replicates at each generation. Eight generations of selection were conducted in this manner on the six replicate lines, and the six controls were also maintained at comparable densities. Generation F_8 females from five replicates of each treatment were frozen in 100% ethanol for the DNA hybridizations and sequencing. All flies were maintained at 25 °C under constant light on a dextrose-dead yeast-agar medium in 500-ml bottles.

Desiccation Selection Response Assay

The lines were assessed for desiccation resistance following the final selection round. The desiccation assays were conducted on all lines except Control replicate 4 (which had low progeny numbers at this time for an unknown reason) after two generations following the cessation of selection. Density was controlled in the test flies by minimizing the time for oviposition; 130 females and 150 males were allowed to oviposit in 500-ml bottles for a 2-h interval. The progeny eclosed into mixed sex cohorts and were sorted by sex into groups of 10 under light CO_2 anesthesia and allowed at least 48 h recovery prior to desiccation. Flies were desiccated as described above, and scored at hourly intervals until 100% mortality was reached (LT_{100}).

Coffs Harbour Association Study

For the association study, the offspring of 30 field-collected females from a banana plantation in Coffs Harbour (30°13′507′′S, 153°05′443′′ E) in 2010 were pooled and mass bred for a generation under controlled density conditions (limiting the time for oviposition as described above). Virgin generation F_2 flies were separated by sex using light CO_2 anesthesia and pair mated at 6–7 days of age. The pairs were transferred to fresh medium after 48 h to create a second block of progeny from each cross. The parents were placed in 100% ethanol and frozen for later genotyping. The offspring were held in mixed sex cohorts for 48 h then separated by sex using light CO_2 anesthesia and transferred to fresh medium in groups of 8–10. Two groups of 8–10 offspring of each sex were assessed from two blocks of 137 crosses (137 crosses × 2 blocks × 2 sexes × 2 groups of 8–10 flies = approximately 10,000 flies). Flies were desiccated as described above and scored for mortality at hourly intervals, with the time for half the flies to die (LT_{50}) determined by linear interpolation. As the Coffs Harbour population was initially cultured in the laboratory at 19 °C, all assays were performed at this temperature for consistency.

Phenotypic Data Analyses

The difference between the selected and control line LT_{100} values were compared with a one-way analysis of variance (ANOVA). While the residual variances departed from normality ($P < 0.0001$, Shapiro–Wilk test, PROC UNIVARIATE, SAS V9.2), inferences were similar using nonparametric or parametric statistics, therefore, ANOVA is shown for simplicity.

For the Coffs Harbour population, the LT_{50} of each sex/replicate combination was adjusted to account for slight differences in mortality across the six tanks used for the desiccation assays. The difference between the overall average LT_{50} across tanks and the average of tank x was subtracted from each vial in tank x. The 10% tails of the F_2 phenotypic distribution were chosen by ranking the female LT_{50s}. This resulted in 14 F_1 parents of field females from each tail of the distribution for genotyping (14/137 "high" parents and 14/137 "low" parent crosses). Phenotypic differences between the two tails were tested with a Student's *t* test (SAS V9.2).

DNA Isolation and Sample Preparation for Array Hybridizations

Groups of 25 females were homogenized in 180 μl buffer ATL (QIAGEN, Valencia, CA) and lysed in 20 μl proteinase K at 56 °C for 3 h, then incubated with 4 μl RNase A (100 mg/ml) for 2 min at room temperature. Genomic DNA was isolated and purified using the DNeasy Blood and Tissue Kit (QIAGEN) according to the manufacturer's instructions. For each replicate line, four separate isolations were combined and concentrated by ethanol precipitation for a total pool of 100 flies per sample, producing a total of 10 samples for the hybridizations. DNA concentration was determined on a Nanodrop Spectrophotometer (Nanodrop Technologies, Wilmington, DE), and quality was assessed using 1% agarose gel electrophoresis.

DNA Fragmentation for Array Hybridizations

DNA fragmentation was carried out according to the following protocol: 9 μg of each sample was digested in a 40 μl reaction containing 0.016 U/μg DNaseI (New England Biolabs Inc, Ipswich, MA), 4 μl One Phor All Buffer (100 mM each of Tris-acetate, magnesium acetate, and potassium acetate in 50 ml H_20), 0.08 μl purified BSA (NEB Inc). Samples were incubated in an Eppendorf PCR machine at 37 °C for 16 min, and the reaction was stopped at 99 °C for 15 min, then cooled at 12 °C for 10 min. Fragments were visualized using 3% agarose gel electrophoresis. This protocol generated fragments consistently ranging approximately 200–15 bp, with an average size of 50 bp.

Sample Labeling, Array Hybridization, Washing, and Scanning

Fragmented DNA (7.5 μg) was labeled and hybridized to GeneChip *Drosophila* Tiling 2.0R Arrays (Affymetrix, Santa Clara, CA) following sections of the Affymetrix Chromatin Immunoprecipitation Assay protocol. Double-stranded fragmented DNA was labeled in 15 μl of double-stranded DNA labeling mix (12 μl 5× TdT buffer, 2 μl TdT, 1 μl 5 mM DNA labeling reagent) by incubating at 37 °C for 60 min, 70 °C for 10 min, and cooling at 4 °C. For each sample, the hybridization cocktail contained 7.5 μg fragmented and labeled DNA target, 3.3 μl control oligonucleotide B2, 100 μl 2× hybridization mix, 14 μl DMSO, and nuclease-free H_20 in a total volume of 200 μl. The hybridization cocktail was heated at 99 °C for 5 min and cooled at 45 °C for 5 min.

Arrays were hybridized at 45 °C at 60 rpm for 16 h, then washed and stained using the fluidics protocol for GeneChip Tiling Arrays. Array hybridization, washing, and scanning were performed at the Ramaciotti Centre (University of New South Wales, NSW, Australia). Raw intensity values for the perfect match (PM) features were exported from CEL files using Affymetrix Tiling Analysis Software (v1.1).

SFP Analysis

The Affymetrix probe annotation (NCBIv36 release) was updated to exclude probes with partial or multiple hits by aligning 2, 877,067 PM probes to the FlyBase release 5.22 excluding chromosome U (Tweedie et al. 2009) using the short read aligner Bowtie (Langmead et al. 2009). Two million seven hundred fifty thousand nine hundred fifty-six probes had unique hits to the 5.22 genome. This included 157 Affymetrix control features, 536 Y chromosome features, and 536 chromosome U-labeled features which were excluded from the final analysis of 2, 750,232 probes.

The data were normalized using a modified rank-based method (Graze et al. 2009). This approach predicted SNPs with 97% accuracy between *D. melanogaster* and *D. simulans* on Drosophila 1.0R tiling arrays (Graze et al. 2009). This normalization method controls for shifts in signal intensities relative to all the features on the array and accounts for artifacts such as background and slide intensity variation among slides. The raw PM intensities of each array were ranked into ten bins as this tended to provide greater dynamic range, sensitivity, and overall stringency compared with ranking data into thirds or quartiles. Allele frequency shifts between the control and selected lines were identified from those features where consistent DNA mismatches between the target and probe sequences sufficiently reduced hybridization efficiency resulting in signal attenuation in one set of treatments compared with the other. This was determined by comparing the normalized hybridization signals between the control and selected lines using Student's *t* tests. Significance was determined across multiple tests using the false discovery rate (FDR) method of Benjamini and Hochberg (1995) at a threshold of 0.2. Analyses were performed using SAS software version 9.2 (SAS Institute, Cary, NC).

Verification of Putative SFPs with Sequencing

Regions of genomic divergence between the control and selected lines were further assessed with single-pass sequencing. A subset of SFPs were chosen for verification based on several criteria including substantial signal attenuation between treatments, clustering of significant SFPs, and mapping to an annotated gene with biological relevance to water balance. Fifteen SFPs targeting 13 genes were sequenced. Primers flanking up to 400 bp either side of the probe sequence were designed using Primer3 (http://frodo.wi.mit.edu/primer3/ [date last accessed 6 December 2011]); sequences and expected product sizes are provided in supplementary table 1 in supplementary file 1, Supplementary Material online. A standard 25 μl polymerase chain reaction (PCR) contained 2.5 μl 1× NEB Buffer, 2.5 μl dNTPs, 1 μl 10 μM forward primer, 1 μl 10 μM reverse primer, 0.21 μl NEB Taq, 16.8 μl H_20, and 1 μl genomic DNA diluted 1:20 in H_20. Products were amplified with a standard PCR profile with annealing temperatures ranging from 55 to 60 °C. The same pooled genomic DNA samples used for the array hybridizations (bulk samples of 100 females) were sequenced from purified PCR products using standard protocols at Macrogen (Korea). Sequencing was performed on both strands only in those cases where one direction yielded poor quality sequence. Multiple alignments were performed using Sequencher V4.7. The chromatograms were visually inspected to identify sites undergoing frequency changes and based on the predominance of two major peaks at the polymorphic sites in the control lines, the secondary peak height calling threshold was set to 40% of the larger peak. This threshold tended to partition major allele changes from the sequencing noise of 100 individuals, albeit at the expense of detecting rare or fixed alleles.

Functional Annotation Enrichment Analysis

SFPs that mapped to a functionally annotated gene were assessed for annotation-term enrichment analysis using DAVID (Dennis et al. 2003; Huang et al. 2009). Three hundred and eighty-eight Flybase IDs were converted to DAVID IDs (*D. melanogaster* background) and were analyzed with the following settings: functional categories = "cog ontology": SP_PIR_keywords, UP_seq_features; "gene ontology": GOTERM_BP_FAT (biological process), GOTERM_MP_FAT (molecular process), GOTERM_CC_FAT (cellular component); "protein domains": INTERPRO, PIR_SUPERFAMILY, SMART; "pathway": KEGG pathways. Gene-term enrichment was first performed to determine which annotation terms were overrepresented in the gene list (custom settings: minimum count = 5; EASE = 0.05). Probability values (EASE scores) were determined using a modified Fisher's exact test (Dennis et al. 2003; Huang et al. 2009) and were corrected for multiple tests with an FDR threshold of 0.1. Functional annotation clustering was then performed to summarize annotation terms into more meaningful biological modules and highlight group-term enrichment of related genes (custom settings: simterm overlap = 3; similarity threshold = 0.75; initial group membership = 5; final group membership = 5; multiple linkage threshold = 0.5). The enrichment score is the log transformed geometric mean of all the EASE scores of each annotation term in the group (Dennis et al. 2003; Huang et al. 2009).

Dys-RF Association Study

The parents of the progeny from the Coffs Harbour 10% phenotypic tails (high and low progeny) were genotyped for the association analysis at the *Dys-RF* locus. DNA was isolated from single flies using 5% w/v Chelex, and PCR was performed using Phusion Flash High Fidelity MasterMix (Finnzymes) following the manufacturer's protocols. This specific DNA polymerase yielded more consistent PCR amplicons from the Chelex-based DNA isolations. Sequencing was performed at Macrogen (Korea).

From the 56 individual dams and sires producing the 10% highest and 10% lowest desiccation tolerant progeny (28

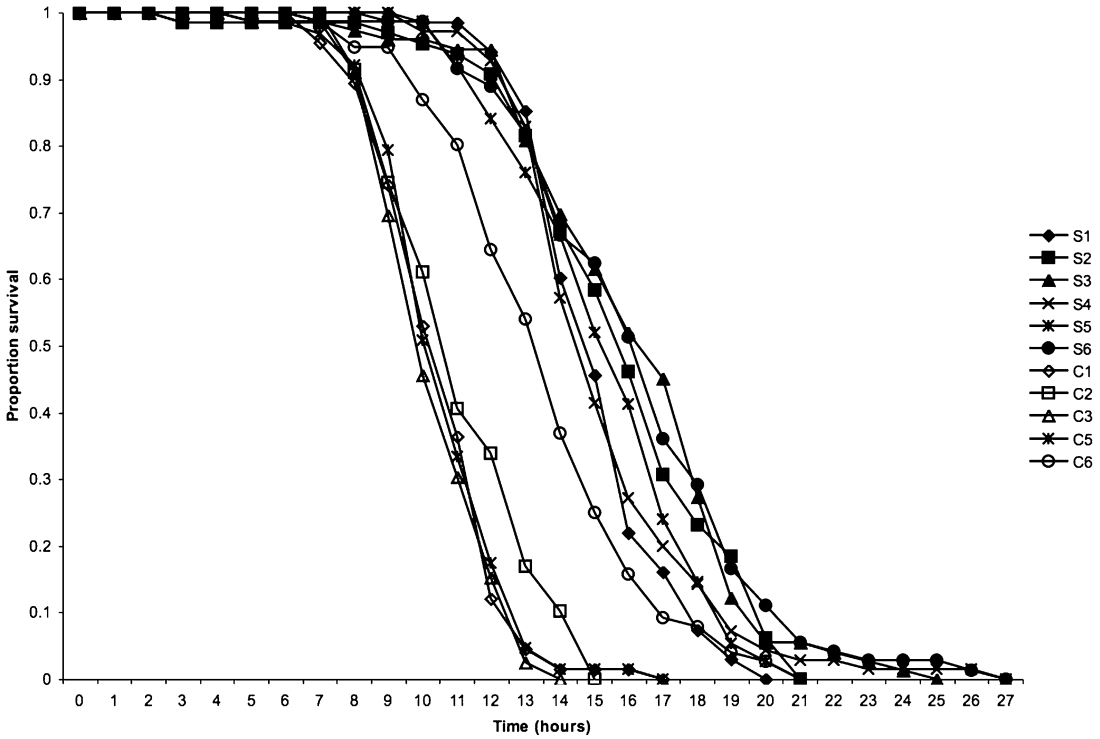

Fig. 1. Average female response to selection for desiccation resistance after nine generations at 25 °C. The five control lines are shown as open markers, and the six selected lines are solid markers.

pairs, 14 high and 14 low), sequences were obtained from 47 parents (24 high and 23 low). Owing to the decrease in sequence quality across some amplicons, 96 alleles were analyzed for markers 3R:15286900-15287069 and 84 alleles for markers 3R:15287112-15287364. The sequences were aligned in Sequencher Version 4.7, and the allele frequencies at 28 polymorphic sites (spanning 463 bp) between the high and low classes were compared using Fisher's exact tests (Proc FREQ, SAS 9.2). Linkage disequilibrium (LD) statistics (r^2) and haplotype patterns across the 463-bp region were generated with Haploview (Barrett et al. 2005).

Results

Desiccation Selection Response

The effect of selection was significant in a one-way ANOVA ($F_{1,81} = 119.2$, $P < 0.0001$); moreover, while the selection response was similar between the replicate selected lines ($F_{5,40} = 1.89$, $P = 0.118$), there was significant variation among the replicate control lines ($F_{4,32} = 5.66$, $P < 0.002$) due to lower mortality in one of the control replicates (fig. 1). Replicate control lines no longer differed when this line was excluded from the analysis ($F_{3,25} = 2.21$, $P = 0.112$). We suspect that the unusual behavior of this line reflected a methodological issue because it did not behave aberrantly when retested at a later stage (De Garis S, unpublished data). There was a significant shift in the survival curve of the selected lines with 100% mortality occurring on average 6.7 h after the controls, representing an average survival increase of 29% (fig. 1) or 7.75 h and 33% survival increase when excluding the outlying control line.

SFP Discovery

Regions of genomic divergence between the desiccation resistant and control lines were mapped using Affymetrix GeneChip *Drosophila* Tiling 2.0R Arrays with a modified DNA hybridization protocol and SFP analysis. Our goal was to identify a general molecular signal of small strongly differentiated regions following artificial selection rather than to identify replicate-specific responses across large regions of the genome. Five selected and five control lines were chosen for genomic analysis, and our methodology maximized biological replication at the treatment level with relatively few (10) arrays. The normalized signals of 2,750,232 unique 25 bp PM features were compared between the selected and control treatments using Student's *t* tests. Six hundred and seventy-one SFPs were identified at a FDR of 0.05, with a further 16 and 4 SFPs at FDRs of 0.1 and 0.2, respectively (supplementary table 2, cited as supplementary file 2.1, Supplementary Material online). To balance Type1 and Type II error rates, we performed further analyses on the SFPs significant at FDR 0.05, although SFPs were later confirmed at all thresholds. The SFPs were randomly distributed across the genome with no significant departure from random expectations in the comparison of features across the genome ($\chi^2 = 2.86$, degrees of freedom [df] = 4, $P = 0.58$) or when the analysis was carried out on the basis of the individual chromosome arms (cited as supplementary table 2 in supplementary file 1, Supplementary Material online). The majority of significant features mapped to annotated genes including introns, exons, and untranslated regions, but at least a third of all features spanned intergenic regions flanking gene clusters (table 1).

Table 1. Summary of DNA Regions of the 691 Significant SFPs (FDR 20%).

DNA Sequence Region	SFPs
Exon	116
Exon/intron	12
Exon/UTR	6
Intron	262
Intron/UTR	5
UTR	52
UTR/IGR	2
IGR	236

NOTE.—Dual categories indicate a probe that spans two regions either in the same transcript or in alternative transcripts. UTR, untranslated region; IGR, intergenic region.

Sequence Confirmation

We chose a subset of 15 SFPs targeting 13 genes to identify sequence divergence between the control and selected lines at the nucleotide level. Sequencing up to 400 bp either side of the SFPs confirmed polymorphisms in 12 of the 13 genes investigated (table 2 and fig. 2). The same genomic DNA pools used for the DNA–DNA array duplexes were sequenced. Rather than sequencing many individuals from each population to infer exact allele frequencies, sequencing on mass (100 females per sample) quickly confirmed sites undergoing allele frequency changes. Secondary peak calling was utilized to determine population polymorphisms similarly to calling heterozygous sites in individuals. Lower peaks were only considered if their heights were at least 40% of the upper peak to reduce the extensive ambiguous base calling resulting from the "noisy" signals of 100 individuals per amplicon. This threshold predominantly detected shifts in alleles from low/intermediate to high frequencies following selection (table 2). Note however, that the alleles shown are an estimate of the total allelic composition at each site and do not necessarily reflect rare or fixed alleles (table 2).

Eleven SFPs were verified for ten genes, whereas SNPs were detected outside the Affymetrix probe for *Dys* (region 2, 18 bp) and *SNF4Agamma* (150 bp and 178 bp either side of the probe) (table 2). The multiple alignments for each region confirmed are given in supplementary file 4, Supplementary Material online. While relatively rarer compared with true SFP detection, sequence polymorphisms can be detected near a monomorphic feature due to labeling polymorphisms and different hybridization kinetics (Rostoks et al. 2005; Borevitz et al. 2007; Turner et al. 2008).

At least two segregating sites were detected for 10 of the 12 genes, including two highly differentiated regions of the *Dys* and *cdi* genes with 29 and 34 sites out of 678 bp and 375 bp, respectively (fig. 2 and table 2). Of the three *Dys* SFPs, the most differentiated region spanned the 5′

Table 2. Sequencing Summary of the <1 kb Regions Surrounding 15 SFPs from 13 Genes.

Gene	Genome Coordinate	DNA Region	Size (bp)	Sites	Allele Change, Control-Selected	SFP Verified
Dys (region 2)	3R:15301867-15302296	Intron	430	1	CT-C	18 bp from probe
cdi	3R:14914539–14914913	Intron	375	34	AG-G; AC-C; AT-T; CG-G; GC-C; CA-A; GT-T; GC-C; GT-T; GT-T; GC-C; GC-C; GC-C; CA-A; CA-C; GC-C; GT-T; GT-T; AT-A; GC-G; GT-T; GC-G; GCT-T; GC-G; GC-G; GT-T; GC-G; GC-G; GT-G; CT-T; A-AG; AG-G; GC-C; AT-T	3R:14914754, 3R:14914762, 3R:14914771
Abd-B	3R:12757222–12757624	Exon/intron	420	10	GT-TG; AC-C; GCT-T; AGT-A; AG-G; AG-A; AT-A; AT-A; GT-G; AT-A	3R:12757492, 3R:12757503, 3R:12757504
CG31431	3R:17698959–17699378	Exon/UTR	420	3	CT-T; TA-A; CA-A	3R:17699153
beatVII	3R:22495768–22496287	Intron	520	3	CT-TC; GT-G; GC-C	3R22495936
CG7638	3L:1103922–1104241	Exon/intron	320	3	CT-T; AG-A; CT-C	3L:11104116, 3L:11104128
Mol	2L:14981445–14980926	Intron	520	3	GC-C; AG-G; CT-C	2L:14981743, 2L:14981747
CG11069	3R20722832–20722434	Exon[a]/intron	399	1	AG-A	3R:20722538
SNF4Aγ	3R:17022665–17022368	Intron	298	4	AT-T; AC-C; C-AC; C-CT	3R:17022393
CG14304	3R:14426924–14427243	Intron	320	2	GT-T; AT-T	80; 120 bp away from probe
CG7084	3R:18129401–undetermined	Intron/exon/3′ UTR	—	3	1-bp insertion-del 27 bp insertion-del 187–394 bp Insertion-del	3R:18129078
ome	3L:14740423-undetermined	3′ UTR	—	—	Undetermined	Indel at probe
dpr2	2L:10945222–10945411	Intron	190	—	—	—

NOTE.—Sequences were obtained from DNA pools of 100 females, and allelic differentiation between the control and selected lines resulting from positive selection for desiccation resistance are represented as "allele changes." The bulk sequencing data were scored based on visual inspection of chromatograms and conservative base peak height analysis at each site. As this scoring method may exclude the detection of low-frequency alleles in a line, the shifts from segregating control states to a single high-frequency allele in the selected lines are intended as polymorphism estimations and do not represent the fixation of alleles following selection. Genome coordinates correspond to the *D. melanogaster* genome build from FlyBase V2011_01. The size (bp) represents the aligned sequence where polymorphisms could be scored with confidence.

[a] Nonsynonymous base substitution.

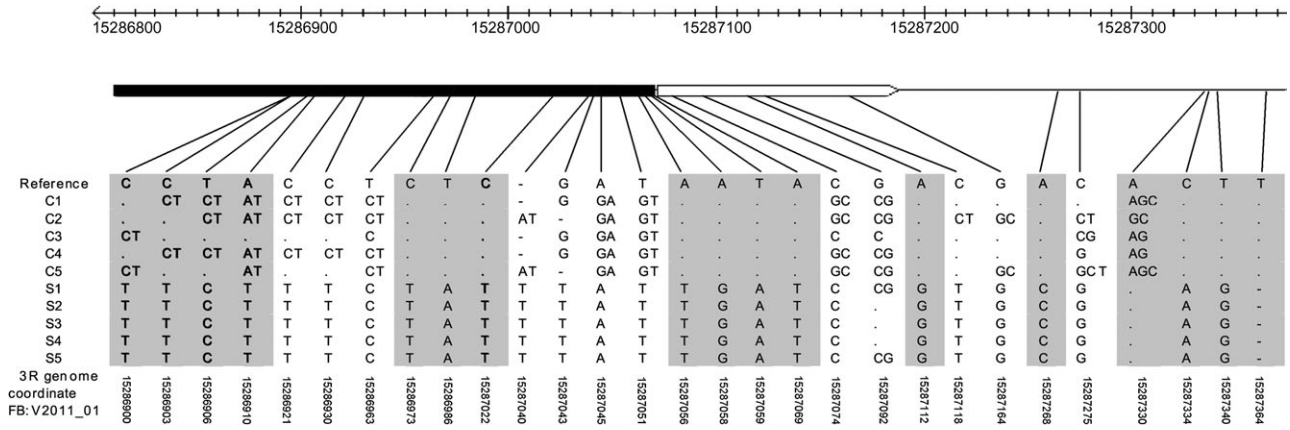

Fig. 2. Allele frequency changes at the 5′ promoter-proximal region of the *Dys-RF* isoform following selection for desiccation resistance. Alleles are scored from population sequencing of 100 females per line may not reflect rare alleles. The Affymetrix feature sequence at each variable site is shown as the probe sequence, C1–C5 = control lines, S1–S5 = selected lines. The gray boxes highlight sequence tiled on the array and bolded sites = significant SFPs. The schematic shows the 5′UTR as the black box, the coding region = white arrow, and the intron = the black line. The chromosome 3R genome coordinates correspond to the *D. melanogaster* genome build from FlyBase V2011_01.

promoter-proximal region of the *Dys-RF* isoform with 29 sites in 463 bp across the first exon including the 5′ UTR, coding region, and intron (fig. 2). The two SFPs in this region were tiled within a 100-bp window, whereas the third SFP indicated a single SNP in an intron over 100-kb downstream (regions 1 and 2, respectively, fig. 2 and table 2). At the 3R:15286900 SFP (region 1), the low signal in the selected lines resulted from C-T and T-C transitions at positions 6, 9, 12, and 16 of the 25mer, whereas a C-T transition at position 12 affected hybridization intensities of the feature-targeting chromosome coordinate 3R:15287022 (fig. 2). This region included two putative alternative promoter sequences (Neural Network Promoter Prediction, scores > 0.9, http://www.fruitfly.org/seq_tools/promoter.html accessed 6 December 2011). The first UTR-predicted promoter sequence contained five SNPS (including SFP 3R:15286900) and four SNPS occurred at the second site proximal to the start codon of the coding region. Sequencing 0.5 kb downstream of this polymorphic region did not reveal any further variation for another 0.5 kb of intronic sequence (data not shown).

For *cdi*, three SNPs were detected at positions 4, 12, and 21 (sense strand) of the feature at 3R:14914751 (A-G and C-T transitions and an A-T transversion, respectively, table 2). For 33/34 *cdi* SNPs, one of the two most common control alleles shifted to a high frequency following selection, in contrast to *Dys* where 8/29 SNPs where represented by a single alternate high-frequency allele each in the control and selected lines. This suggests that for *Dys*, directional selection for desiccation resistance was strong enough to induce rapid allele shifts from low starting frequencies in the founding populations, whereas shifts at *cdi* appear to be from more intermediate starting frequencies. In addition, the *Dys-RF* substitutions were the most consistent across the 5 lines, whereas *cdi* ranged from shifts in 3–5 of the lines. SNPs ranging from 1 to 10 sites were detected in a further 8 genes, whereas two SFPs indicated the presence of indels for two genes *CG7084* and *ome* (table 2). Population sequencing showed clear shifts in the abundance of indels between the control and selected lines, but apart from one site at *CG7084*, the multiple alleles could not be separated and accurately scored. Nonetheless, indels can be detected with this rapid mapping approach, and the alleles can be later determined with individual sequencing.

Overall, the sequence analysis revealed a substantial amount of genomic differentiation that was disproportionate to SFP detection. Given that features are tiled on average every 38 bp from the center of the oligo, we chose the *Dys-RF* region to investigate why only 2 of the 17 features in this region were detected in the SFP analysis. Close inspection of the tiling scheme excluded 12 SNPs from the tiling layout, and while five SNPs were detected on the arrays, an additional 12 variable sites remained unaccounted for. Plotting the average normalized signal of each probe by treatment showed large differences in probe hybridization intensities across this region (fig. 3*a*). The highest signal intensities were observed for features targeting undifferentiated sequence, although signal attenuation was greatest at sites harboring SNPs. Five features were at the extreme low end of the signal distribution in all lines, whereas the two SFPs were due to extreme signal attenuation in only one treatment (fig. 3*a*). Mapping DNA variants with array features as markers is dependent on factors such as the hybridization affinity of the oligo sequence and SNP positions in the 25mer such that SNPs central to the feature (between positions 6 and 13) are more likely to be detected (Rostoks et al. 2005). Accordingly, we plotted the position of the SNPs with positions 1 and 25 the origin and end of the 25mer (sense strand) and found that only target sequence harboring SNPs between positions 6 and 16 hybridized in at least one treatment (fig. 3*b*).

Biological Annotation Enrichment Analysis

The 671 SFPs mapped to 388 annotated genes (DAVID *D. melanogaster* background) and were enriched for 47 annotation terms (FDR 0.1, supplementary table 4, cited as supplementary file 2.2, Supplementary Material online) and one Kegg pathway. The Kegg pathway analysis highlighted four SNPs that target four genes at several junctions

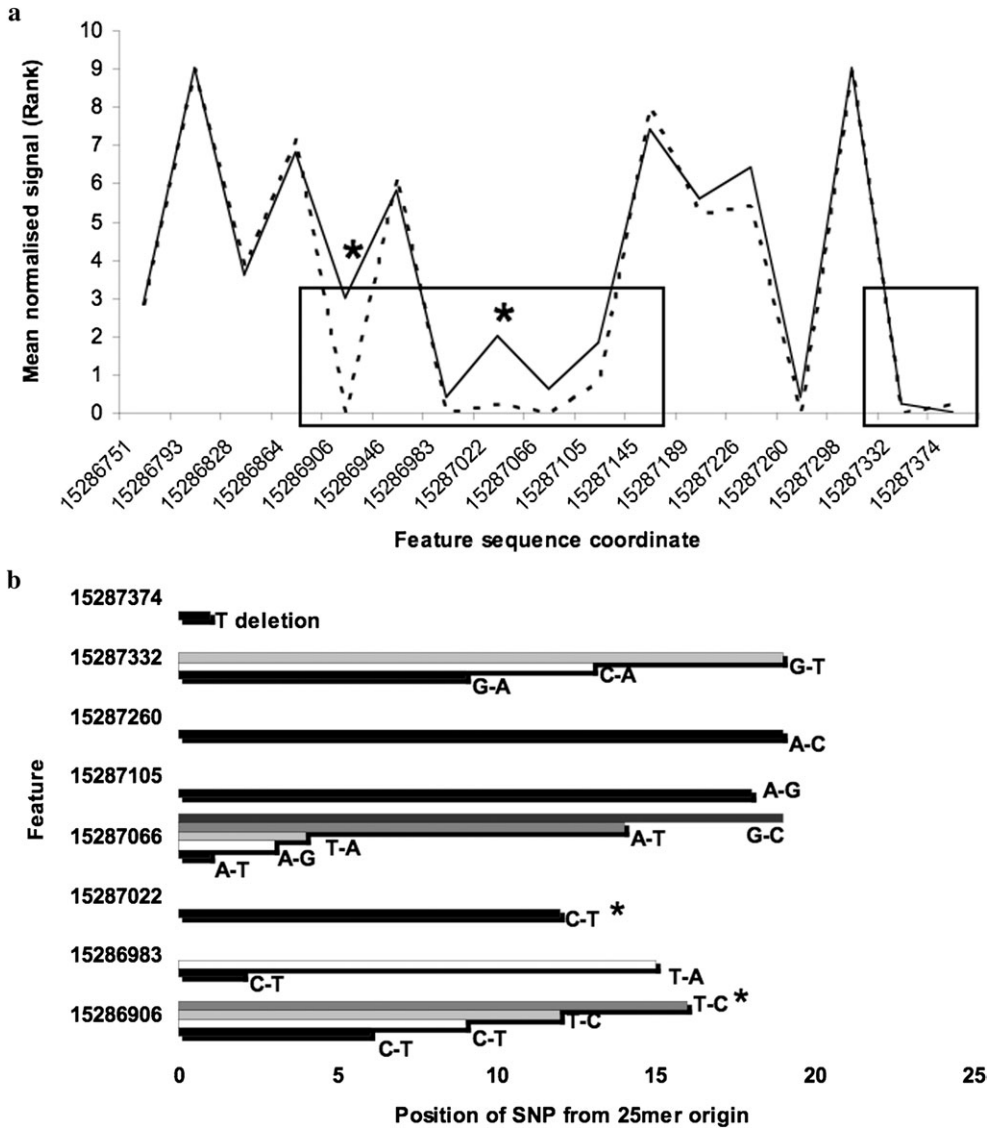

FIG. 3. (*a*) Average normalized signal of the probes tiled across the differentiated 463 bp 5′ promoter-proximal region of the *Dys-RF* isoform. Solid line = control lines, dashed line = selected lines. The boxes indicate probes targeting sites with SNPs and highlight the reduced hybridization intensity compared with undifferentiated sites, * = significant SFPs. (*b*) Position of the SNP from the Affymetrix 25mer probe origin (sense strand) for those features targeting differentiated regions. The feature position genome coordinate, FBV2011_01, is shown on the *y* axis, and allele frequency shifts are shown from the control to the selected lines. * = significant SFPs.

in the phosphatidylinositol signaling system (supplementary figure 1 in supplementary file 1, Supplementary Material online). Functional annotation clustering reduced term redundancy by organizing the annotation output into five smaller modules (nominal FDR-corrected EASE of 0.1 for each term, table 3). Although the results between the two enrichment analyses were concordant, clustering can result in the omission of enriched terms without closely related neighbors (Huang et al. 2009). In this case, 18 genes associated with respiratory system development related to open tracheal development were overrepresented with gene-term enrichment analysis but not with functional annotation clustering (supplementary table 5 cited as supplementary file 2.3, Supplementary Material online). In the cluster analysis, the highest and lowest enrichment scores of 7.59 and 2.71 (table 3) are equivalent to <0.0001 and 0.001, respectively, on the nonlog scale (Dennis et al. 2003). Modules 1–4 were connected to varying degrees due to overlapping gene annotations and biological themes, whereas module 5 clustered independently. The two most enriched modules were functionally related to nervous system development at the level of the neuron, including cell–cell communication (1) and neuronal differentiation (2) (table 3). Module 1 comprised of highly overrepresented members of the immunoglobulin protein superfamily (IgSF). Sequencing of the IGSF genes *beatVII* and *dpr* confirmed the former SFP, whereas the *dpr2* SFP appeared to be a false positive (table 2). False positives may be biological (i.e., such as repetitive sequence binding) or technical (Turner et al. 2008). It is possible that the large poly C, T, and dinucleotide

Table 3. Functional Annotation Clustering Analysis of 388 Annotated Genes from 671 SFPs (FDR 0.05) to Highlight Significantly Enriched Annotation Modules.

Category	Term	FDR-Corrected EASE	Gene Number	Fold Enrichment
Module 1: enrichment score = 7.59				
IPR007110	Immunoglobulin-like	3.02×10^{-07}	23	5.36
IPR003599	Immunoglobulin subtype	1.65×10^{-05}	18	5.71
IPR013151	Immunoglobulin	6.72×10^{-05}	16	6.01
IPR013783	Immunoglobulin-like fold	2.90×10^{-04}	18	4.73
SM00409	Immunoglobulin	6.32×10^{-04}	18	4.30
Module 2: enrichment score = 6.61				
GO:0030182	Neuron differentiation	2.30×10^{-05}	36	2.92
GO:0000904	Cell morphogenesis involved in differentiation	1.49×10^{-04}	29	3.17
GO:0048812	Neuron projection morphogenesis	1.66×10^{-04}	28	3.25
GO:0031175	Neuron projection development	1.78×10^{-04}	28	3.23
GO:0048667	Cell morphogenesis involved in neuron morphogenesis	1.92×10^{-04}	28	3.22
GO:0048666	Neuron development	2.21×10^{-04}	31	2.96
GO:0048858	Cell projection morphogenesis	4.69×10^{-04}	29	3.00
GO:0000902	Cell morphogenesis	5.07×10^{-04}	35	2.62
GO:0032990	Cell part morphogenesis	9.39×10^{-04}	29	2.90
GO:0030030	Cell projection organization	0.00213	30	2.72
GO:0032989	Cellular component morphogenesis	0.00257	37	2.37
GO:0007409	Axonogenesis	0.00291	21	3.52
Module 3: enrichment score = 5.68				
GO:0007444	Imaginal disc development	2.76×10^{-05}	36	2.90
GO:0048569	Postembryonic organ development	9.73×10^{-05}	29	3.24
GO:0048563	Postembryonic organ morphogenesis	1.44×10^{-04}	28	3.27
GO:0007560	Imaginal disc morphogenesis	1.44×10^{-04}	28	3.27
GO:0009791	Postembryonic development	3.29×10^{-04}	35	2.6
GO:0002165	Instar larval or pupal development	4.32×10^{-04}	34	2.70
GO:0007552	Metamorphosis	5.63×10^{-04}	31	2.84
GO:0048707	Instar larval or pupal morphogenesis	7.49×10^{-04}	30	2.87
GO:0009886	Postembryonic morphogenesis	0.00107	30	2.82
GO:0035114	Imaginal disc-derived appendage morphogenesis	0.01478	22	3.05
GO:0035107	Appendage morphogenesis	0.01793	22	3.01
GO:0048737	Imaginal disc-derived appendage development	0.01911	22	3.00
GO:0048736	Appendage development	0.02304	22	2.96
GO:0035120	Postembryonic appendage development	0.02679	21	3.04
Module 4: enrichment score = 3.85				
GO:0045449	Regulation of transcription	0.03318	46	1.91
GO:0003700	Transcription factor activity	0.02574	30	2.45
GO:0006355	Regulation of transcription, DNA-dependent	0.03907	37	2.33
SP_PIR_KW	DNA-binding	0.02574	29	2.45
Module 5: enrichment score = 2.71				
SP_PIR_KW	Ion transport	0.07437	17	3.29

NOTE.—Categories include: IPR, INTERPRO Protein domains; SM, SMART Protein domains; GO, gene ontology; SP_PIR_KW:, keywords.

repeats flanking the *dpr2* SFP affected probe binding, in any case we did not detect polymorphisms in the 190-bp amplicon surrounding the probe.

Module 2 was enriched for genes involved in aspects of neuronal development, cell morphogenesis, and axonogenesis (table 4, supplementary table 4 cited as supplementary table 2.2, Supplementary Material online). Genes clustered to module 3 were enriched for anatomic development, particularly organ and appendage formation, as well as functioning in neuronal development (module 2), including *Dys* (table 4, supplementary table 5 in supplementary file 2.3, Supplementary Material online). The developmental modules included a substantial number of transcription regulators (module 2 = 13/36, module 3 = 18/36) overlapping with the 46 DNA binding/regulatory genes clustered in module 4 (table 4, supplementary table 4, cited as supplementary file 2.2, Supplementary Material online). Module 5 comprised a functionally distinct and unique cluster of genes related to ion transport (table 4, supplementary table 4, cited as supplementary file 2.2, Supplementary Material online). Although functional annotation clustering analysis excluded the open tracheal development genes, a comparison of this gene enrichment group to the clustered data showed that 15/18 genes were represented in modules 2, 3, and 4 (supplementary table 5 cited as supplementary file 2.3, Supplementary Material online). Data mining of the literature and protein–protein interaction databases showed many instances of genetic interactions between genes/gene products within and between the modules; these are summarized in supplementary table 5 cited as supplementary file 2.3, Supplementary Material online.

Validation of the Mapping Approach: Dys and Desiccation Tolerance Association Study

Finally, we utilized an indirect association approach to test the degree of overlap between candidate alleles for

Table 4. Frequencies of the Selected Allele in the Coffs Harbour Parents of the 10% Most (High) and Least (Low) Desiccation Resistant Progeny for 23 Shared Sites of the 463 bp Region of *Dys-RF*.

3R Genome Coordinate, FB:V2011_01	Selected Allele	High Lines, Allele Frequency (*n*), 95% CI	Low Lines, Allele Frequency (*n*), 95% CI	Fisher's Exact Test, Two-Tailed *P*
15286900	T	0.270 (48), *0.142–0.399*	0.108 (46), *0.018–0.198*	0.065
15286903	T	0.270 (48), *0.142–0.399*	0.108 (46), *0.018–0.198*	0.065
15286906	C	0.270 (48), *0.142–0.399*	0.108 (46), *0.018–0.198*	0.065
15286910	T	0.270 (48), *0.142–0.399*	0.108 (46), *0.018–0.198*	0.065
15286921	T	0.270 (48), *0.142–0.399*	0.108 (46), *0.018–0.198*	0.065
15286930	T	0.270 (48), *0.142–0.399*	0.108 (46), *0.018–0.198*	0.065
15286963	C	0.270 (48), *0.142–0.399*	0.108 (46), *0.018–0.198*	0.065
15286973	T	0.270 (48), *0.142–0.399*	0.108 (46), *0.018–0.198*	0.065
15286986	A	0.270 (48), *0.142–0.399*	0.108 (46), *0.018–0.198*	0.065
15287022	T	0.270 (48), *0.142–0.399*	0.108 (46), *0.018–0.198*	0.065
15287040	T	0.333 (48), *0.197–0.469*	0.130 (46), *0.033–0.227*	0.027
15287043	T	0.208 (48), *0.090–0.325*	0.086 (46), *0.005–0.168*	0.147
15287056	T	0.208 (48), *0.090–0.325*	0.086 (46), 0.005–0.168	0.147
15287058	G	0.187 (48), *0.074–0.300*	0.086 (46), *0.005–0.168*	0.233
15287059	A	0.208 (48), *0.090–0.325*	0.086 (46), *0.005–0.168*	0.147
15287069	T	0.208 (48), *0.090–0.325*	0.086 (46), *0.005–0.168*	0.147
15287112	G	0.642 (42), *0.501–0.784*	0.340 (44), *0.200–0.480*	0.009
15287118	T	0.642 (42), *0.501–0.784*	0.363 (44), *0.221–0.505*	0.017
15287268	C	0.238 (42), *0.112–0.636*	0.090 (44), *0.005–0.175*	0.083
15287275	F	0.333 (42), *0.194–0.472*	0.295 (44), *0.160–0.430*	0.817
15287334	A	0.238 (42), *0.112–0.363*	0.090 (44), *0.005–0.175*	0.083
15287340	G	0.238 (42), *0.112–0.363*	0.090 (44), *0.005–0.175*	0.083
15287364	—	0.190 (42), *0.074–0.306*	0.090 (44), *0.005–0.175*	0.223

NOTE.—Associations of the selected alleles with desiccation resistant Coffs harbour families were tested with Fisher's exact tests and sites significant at $\alpha < 0.05$ are underlined. The 95% binomial confidence intervals for the population allele frequencies are shown in italics.

desiccation resistance from artificial selection and alleles occurring in naturally tolerant *D. melanogaster* (Weeks et al. 2002; Chenoweth and Visscher 2009). We compared the frequencies of SNPs in the 3R:15286900–15287364 (FlyBase V2011_01) region of the *Dys RF* isoform between the parents of the most desiccation resistant and susceptible progeny from a wild-derived population from Coffs Harbour (CH) Australia. To avoid environmental and maternal effects on phenotypic measurements of flies directly from the wild, the F_1 progeny of 30 wild caught isofemales were mass mated, and their progeny were used to establish single-pair matings to generate F_3 field derived flies for phenotyping. Based on the 10% tails of the F_3 phenotypic distribution from 137 crosses, 14 F_2 pairs representing the high progeny and 14 F_2 pairs representing the low progeny were selected for sequencing. Desiccation resistance substantially differed between the high and low female progeny from these crosses, forming two discrete phenotypic classes (paired *t*-test, high vs. low; $t = 23.46$; df $= 24$; $P < 0.0001$), with high females surviving an average 8.9 h longer than low females (fig. 4). Based on the sequencing outcome, 24 of the possible 28 high dams and sires and 23 of the 28 low dams and sires were included in the SNP analysis (the phenotypic data above reflects these crosses only). Twenty-three SNPs were shared between the CH and the selected populations (table 4), and five SNPs were unique to CH (supplementary table 6 cited as supplementary file 3, Supplementary Material online). The selected allelic state segregated at a higher frequency among the parents of the high CH progeny for all 23 common SNPs, and this was significant at $\alpha < 0.05$ for sites 3R:15287040, 15287112, 15287118, and marginal at $\alpha = 0.07$ for an additional ten SNPs (table 4). Allele frequencies significantly differed between the high and low CH families for an additional SNP at 3R:15287017 that did not change in response to selection (supplementary table 6 cited as supplementary file 3, Supplementary Material online, $P = 0.005$).

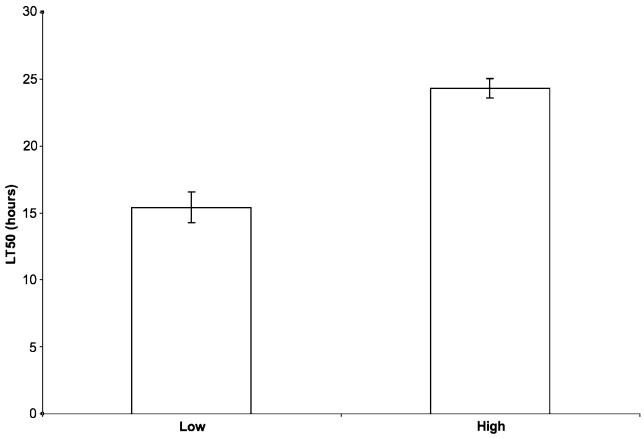

FIG. 4. Average survival to desiccation (LT_{50}) at 19 °C in the 10% most and 10% least resistant families, three generations from collection in Coffs Harbour, Australia. Error bars are standard deviations of the means.

Patterns of LD were extensive across the 463-bp region tested and could be resolved to several small haplotype blocks (fig. 5). Haplotypes in complete LD comprise six SNPs at positions 3R:15286900-15286930, three SNPs at 3R:1528963-15286986, four SNPs at 3R:15287043-15287069, and two SNPs at positions 3R:152871112–15287118. When

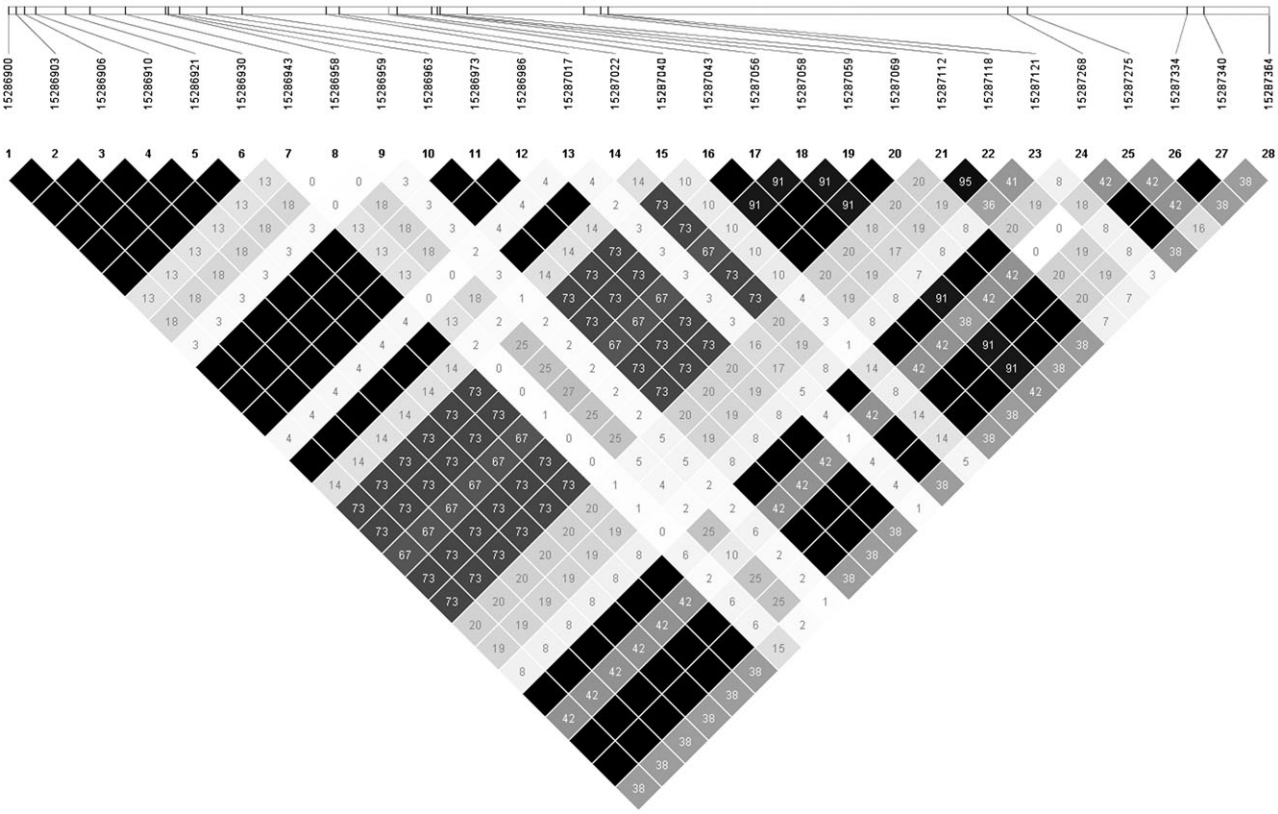

FIG. 5. Haploview LD analysis of 28 segregating sites in 463 bp of the *Dys-RF* isoform in Coffs Harbour. The r^2 for all pairwise comparisons between sites are shown, where $r^2 = 0$ = white boxes, $0 < r^2 < 1$ = shades of gray, $r^2 = 1$ = black boxes.

the allele frequencies were compared between the high and low groups by haplotype, the marginal differences in allele frequencies from the single-marker analysis were significant for SNPs at 3R:15286900–15286930 ($\chi^2 = 7.90$; df = 2; $P < 0.05$) and remained significant at 3R:15287112-15287118 ($\chi^2 = 14.52$; df = 3; $P < 0.01$). The other significant sites from the single-marker analysis (3R:15287017 and 15287040) were not linked to any other sites (fig. 5).

Discussion

SFP Analysis

Drosophila distributional patterns appear to be closely associated with evolutionary potential for physiological tolerances to ecological traits such as cold and desiccation tolerance (Hoffmann et al. 2003; Kellermann et al. 2009). Although the genetics of many stress tolerance traits are unknown, their elucidation could considerably expand our understanding of the differences in physical tolerances between species, improve predictions of climate change effects on populations, and provide potential genetic indicators of environmental stress. Specifically, allele frequency shifts at key loci undergoing evolutionary change could be used as genetic markers to monitor effects of environmental perturbation on populations well before the occurrence of range shifts or extinction (Hoffmann and Daborn 2007). Our data provides new candidate alleles associated with adaptive evolution for desiccation resistance in a broadly distributed tolerant *Drosophila* species. Model organisms such as *D. melanogaster* can provide valuable inroads to understanding the molecular ecology of climatic traits in nonmodel *Drosophila*, such as Australian rainforest endemics. We used highly replicated divergent lines and relatively few high-density tiling arrays to discover sites undergoing allele frequency changes in *D. melanogaster* under selection for desiccation resistance. Although the efficacy of array-based genotyping is well established (Borevitz et al. 2003, 2007; Rostoks et al. 2005; Turner et al. 2005, 2008; Nuzhdin et al. 2007), our study utilized a unique combination of array genotyping, artificial selection, and association analysis to simply and effectively link candidate alleles from the laboratory to resistance in the field. This approach demonstrates the usefulness of short-term laboratory evolution to understand the evolution of complex traits, despite the potential for confounding experimental artifacts from an artificial design (see Harshman and Hoffmann 2000).

We performed our analysis at the probe level as pooling probe statistics with a sliding-windows approach can result less accurate FDR estimates on this platform (Turner et al. 2008). Furthermore, more complex "regional" analyses are problematic owing to the partial genome coverage due to the tiling strategy itself as well as ineffective hybridization resulting from limits in tiling probe specificity/sensitivity or SNP position (see Zhang et al. 2003). The *Dys-RF* feature analysis illustrates some of these technical issues. In this light, the 0.02% of probes with differentiated signals is

not an explicit measurement of the genomic selection response. Rather, we aimed to map short highly replicated molecular signals in non-laboratory adapted flies to ultimately test candidate associations with natural tolerances among populations and species. Next-generation sequencing technologies provide the coverage and sensitivity required to fully characterize allele frequency changes in traits undergoing selection, particularly for smaller genomes such as budding yeast (Parts et al. 2011). Resequencing at the level of replication required to map convergent evolutionary responses is still in the early phases for larger genomes such as *Drosophila* and is so far limited to selection for body size in long-term laboratory strains (Turner et al. 2011). This will change in the immediate future as technological advances continue to improve both sequencing throughput and cost. In the interim, our method presents a cost-effective way to map replicated adaptive responses to poorly defined complex traits in the absence of a full population analysis preselection and postselection.

Sequence Analysis

Although the high sequencing confirmation rate is promising for a larger scale investigation of evolutionary responses to desiccation resistance, the array analysis tended to underestimate the diversity at many loci as discussed above. Despite this limitation, our data provide a valuable snapshot of alleles undergoing consistent frequency changes, and we surmise that at least a proportion of the significant SFPs represent markers to larger regions undergoing allele changes in response to selection (i.e., *Dys-RF*, *cdi*, *Abd-B*). In the absence of known candidates for desiccation resistance, we examined genes with large signal differences between treatments and with a variety of biological functions that may be relevant to stress resistance. We confirmed segregating sites at two genes that may function in the fly sensory organs *CG7638* and *mol* (Dubnau et al. 2003; Buffin and Gho 2010). The detection of stress is vital to mounting an adequate stress response, and evidence from gene expression studies suggests that genes encoding phototransducers are enriched in both laboratory induced and naturally desiccation resistant *Drosophila* (Sorensen et al. 2007; Matzkin and Markow 2009). Other signaling genes include protein kinases; *SNF4Aγ*, CG14305 and *cdi*, and immunoglobulins *Beat-VII* and CG31431 (Tweedie et al. 2009). Shifts in the frequency of indels at *ome* and *CG7084* highlight more specific desiccation candidates; *ome* codes for an enzyme that modifies a protein in the developing epicuticle, and *CG7084* is one of the most highly expressed genes in the Malpighian tubules (MTs) which together with the hindgut form the fly renal system and site of excretory water balance (discussed below) (Wang et al. 2004; Chihara et al. 2005). Other genes of interest include *CG11069* which encodes a molecule involved in the transmembrane movement of substances, and the transcription factor *Abd-B* which is expressed at high levels in the larval hind gut and trachea (Tweedie et al. 2009).

The regions sequenced around the *cdi* and *Dys* SFPs were highly differentiated, with 34 and 29 sites, respectively, in less than 500 bp. *Dys* is one of the largest genes in the *Drosophila* genome (Neuman et al. 2001, 2005), and although we might expect to detect more SFP by chance, sequencing 1-k upstream of this region showed that the differentiation clustered specifically to the 5′ end of RF transcript. Given that the mapping resolution is unknown and may be low at the chromosome level due to incomplete probe coverage, further genotyping is required to determine if this is indeed a localized signature of selection or the mark of widespread differentiation extending within or beyond the *Dys* locus.

Validating the Mapping Approach

Artificial selection studies of stress resistance in *Drosophila* have so far mostly failed to link nucleotide-level candidates with natural variation in stress traits. Our data contribute to making this essential leap by connecting allelic differentiation at the *Dys-RF* locus in experimental evolution populations with desiccation tolerant flies only a few generations from the field. Quite strikingly, 23 of the selected SNPs segregated at a higher frequency among the desiccation tolerant Coffs Harbour (CH) families. The significant sites reflect strong differentiation between the desiccation tolerant and susceptible CH families. Although there was a high degree of SNP overlap between the laboratory lines and the CH population, there was some divergence between the populations, including an additional SNP at position 3R:15287017 that had significantly different allele frequencies between the high and low CH families. This SNP clusters with a number of SNPs common to the selected and CH populations in the alternative promoter region of the RF transcript. This presents an intriguing region to examine *cis*-regulatory alleles that may affect alternative transcription start sites and/or transcript-specific expression, in addition to further exploring the status of the CH-specific SNP as an alternate resistance allele or silent polymorphism. LD analysis revealed high levels of LD across this small region and the CH SNPs resolved to several haplotype blocks that are considerably narrower than predicted natural LD blocks (Haddrill et al. 2005). Strong LD in naturally tolerant flies combined with the rapid shift from very low to very high frequency alleles in the selected lines are suggestive of a selective sweep in this region in response to positive selection.

It is interesting to note that the three CH SNPS with significant frequencies of the selected allele at 3R:15287040, 15287112, and 15287118 appear not to be in LD with the surrounding alleles, apart from the complete linkage observed between the latter SNPs. This is also the case for the CH-specific SNP at position 3R:15287017. These data highlight this region as a potentially localized region affecting variation for desiccation resistance and provides positional candidates to explore causal allele variants. The *Dys* gene is highly pleiotropic with varied roles in the establishment of cell polarity and wing vein morphogenesis, muscle cell development and homeostasis, regulation of neurotransmitters, and neuronal synaptic plasticity (Tweedie et al. 2009). Mutations in the human homolog of *Dys* (*DMD*) are known to cause the muscle wasting associated with Duchenne muscular dystrophy. In *D. melanogaster*, two

isoforms are required for the integrity of the musculature (Neuman et al. 2005; van der Plas et al. 2006). Targeted knockdowns of all of the *Dys* transcripts results in progressive muscle degeneration of larvae and adult flies (van der Plas et al. 2006) but so far the individual function of the *Dys-RF* isoform remains unclear.

Gene Annotation Enrichment Analysis

Clustering the entire set of 388 annotated genes highlighted biologically meaningful associations among the SFPs, with a particular emphasis on genetic regions harboring genes important for insect water balance.

Candidate Mechanisms at the Tissue Level for Changes in Water Balance

The cluster analysis revealed specific insights into potential changes to excretory water balance in the insect renal system, through a distinct cluster of genes associated with ion transport. In insects, an open circulatory system results in the tight coupling of ion homeostasis with water balance (reviewed in Coast et al. 2002; Beyenbach 2003; Gade 2004). Primary osmoregulation occurs in the MTs, which in conjunction with the hindgut form the insect excretory system (reviewed in Dow and Davies 2001). Although the MT secrete KCl- or NaCl- rich urine that is isosmotic to the hemolymph, water, and other solutes transfer by diffusion and drain to the hindgut that selectively reabsorbs water, ions, and other necessary metabolites. MT and hindgut function are independently regulated by neuroendocrine factors, including diuretic and antidiuretic peptides that either increase or decrease water loss by regulating tubule secretion and/or ion and water retake in the hindgut. Differential regulation of these hormonal peptides provides an obvious way to budget water in *Drosophila*. We mapped SFPs that target the candidate anti-diuretic neuropeptide *ITP* and the neuropeptide *sNPF*. *ITP* products colocalize with the neuropeptides *sNPF* and *DTK* in the fly brain, and loss of function produces a dramatic decrease in both desiccation and starvation survival (Kahsai et al. 2010). A likely scenario is that *sNPF* and *DTK* are accessories to *ITP* and are released during desiccation stress (Kahsai et al. 2010). Other likely candidates for MT water balance include K+ channel genes *KCNQ*, *slo*, sodium channel, and symporter genes *para*, *NaCP60E*, *CG7720*, *Picot*, chloride channel genes *pHCl*, *GluCalpha* among others related to calcium, chloride, and cation channel activity such as *Itp-r83A* and *trpl*, *pHCl*, and *nAcRα-30D*, *nAcRα-34E* (supplementary file 2.1, 2.3, Supplementary Material online). The genes *KCNQ*, *slo*, *Itp-r83A*, *NcKx30C*, and *trpl* are all expressed in the MT, which are the likely sites of sodium balance during desiccation in selected populations of *D. melanogaster* (Folk and Bradley 2005) and water-homeostasis channels in *Anopheles gambiae* (Liu et al. 2011). We mapped several other genes that are known to be expressed in the MT including transcription factors *cad*, *Pxt1*, and *ct*, major signal transducers *Nos*, *dunce*, *sl*, and *Itp-r83A*.

The data were enriched for gene terms relating to open tracheal system development, and while this was omitted from the cluster analysis (see Results), we discuss this group based on biological relevance to insect water balance. The tracheal system is a branched, tubular, and structure that delivers oxygen to the tissues through the external tracheal apertures called spiracles (Samakovlis et al. 1996; Affolter and Caussinus 2008). In *Drosophila*, epidermal water loss can occur through the cuticle or across the spiracles during the open phase in respiration, and this occurs predominantly across the spiracles in xeric and mesic *Drosophila* including *D. melanogaster* (reviewed in Lehmann and Schutzner 2010). Since the tracheal system is assumed to be water saturated, discontinuous gas exchange is thought to keep the spiracles closed for longer to increase CO_2 loss relative to respiratory water loss (Hoffmann and Harshman 1999; Lehmann and Schutzner 2010). This link has proven inconclusive in desiccation-resistant selected lines of *D. melanogaster* (Hoffmann and Harshman 1999), but is thought to save more water in *Drosophila* than other insects, and has been shown to occur more frequently in xeric species that lose water less rapidly than in their mesic counterparts, such as *D. melanogaster* (Marron et al. 2003). Here, SFPs targeted 18 genes associated with tracheal system development, including tube development, primary branching, branch fusion, and chitin biosynthesis (*DAAM*, *Hs6st*, *dys*, and *dp*, supplementary file 2.3, Supplementary Material online). Many genes overlap with several modules significant in the cluster analysis, including the genes encoding cell adhesion molecules *if*, *mew*, *Fas2*, and transcription factors *Ubx* and *ct* (discussed below). Our selected lines show distinct changes in water balance including increased water storage and slower water loss rates which are likely functionally related to changes in the excretory and respiratory systems either independently or in concert. Analyses of water balance mechanisms between desert and mesic *Drosophila* showed that excretory water loss explained a significantly smaller proportion of water loss rates than respiratory water loss (Gibbs et al. 2003), and it would be fascinating to further connect the genetic candidates to the physiological adaptations here and, ultimately, to divergent species.

Candidate Stress Signaling Networks

The cluster ontology analysis also resulted in some less obvious but nonetheless compelling candidates. The most highly represented group comprised a suite of GO terms associated with the immunoglobulin protein superfamily (IgSF) representing up to 17% of the *Drosophila* repertoire (Hynes and Zhao 2000; Vogel et al. 2003). IgSF proteins play a major role in cell–cell communication, signal transduction, and neuronal development and provide a structural platform for essential protein interactions during the life of a neuron including migration, axon pathfinding, synapse formation as well as the maintenance of adult neuronal networks (reviewed in Rougon and Hobert 2003; Maness and Schachner 2007). SFPs clustered to multiple members of the *Beat*, *dpr*, and *sidestep* families (supplementary file 2.3, Supplementary Material online), and while interesting, this group warrants

further investigation given that DAVID does not account for the selection bias for GO categories enriched for longer genes and, particularly, those populated by gene families.

In spite of the potential for overrepresentation bias in the GO analysis, several IgSF are known to genetically interact with genes encoding molecules known to modulate the neuronal response to changing environments (module 2) (Rougon and Hobert 2003), suggesting a biological role in neuronal signaling (the full list is given in supplementary table 5 cited as supplementary file 2.3, Supplementary Material online). Two well-studied interactions include *Fas2* (modules 1 and 2, supplementary table 5 cited as supplementary file 2.3, Supplementary Material online) that encodes a cell surface adhesion molecule (reviewed in Rougon and Hobert 2003). *Fas2* interacts with *Amph* (module 2), which plays a novel role in *Drosophila* postsynaptic vesicle exocytosis to regulate *Fas2* protein cycling (Mathew et al. 2003). *Fas2* also interacts with *beat* proteins (reviewed in Rougon and Hobert 2003), particularly *beat1a* and its complement *beat1c* (Pipes et al. 2001). A final example of molecular "cross-talk" is for SFPs targeting genes encoding adhesive molecules and their receptors which guide neuronal development, including two of the five interacting α subunits of the *Drosophila* integrin αβ heterodimeric receptors *mew* and *if* (Hynes and Zhao 2000).

Recent adaptation resulting in genomic differentiation among related and/or interacting genes presents intriguing possibilities to study evolution from a network perspective (Turner et al. 2008). Our data reveal several attractive candidates to explore in a neural-network context. In order adapt to fluctuating environments, organisms must rapidly detect, then respond to environmental stressors. Stress responses across a range of taxa are mediated through signal transduction either specifically or via molecular cross-talk between pathways (Ruis and Schüller 1995; Chinnusamy et al. 2004). In *Drosophila*, distinct subsets of neurons are recruited into the stress response circuitry to modulate behavioral responses during starvation and oxidative stress, and this is context dependent based on the hormonal environment of the brain (Neckameyer and Matsuo 2008). Furthermore, starvation stress elicits different factors, such as transcription factors *foxo* and transducers of regulated CREB activity (TORCs) in response to insulin-regulated signaling pathways to function in the maintenance of energy balance in the fly brain (Kramer et al. 2008). Although gene expression studies suggest that phototransduction might be important for stress detection during desiccation (Sorensen et al. 2007; Matzkin and Markow 2009), knowledge of stress signaling during water deprivation in flies is scant. However, desiccation resistance is known to be affected by variable neuropeptide control of the renal system (Kahsai et al. 2010), and it is likely that other neuronal signaling components play an important role in water balance but remain to be identified.

Candidate Sites of Regulatory Divergence

Finally, cluster analysis revealed an overrepresentation of DNA transcription regulators involved in complex developmental phenotypes, such as organ and muscle development, pigment, and cuticle synthesis suggest that regulation of gene expression is modified by selection for desiccation resistance. Artificial selection for different stress resistances in *D. melanogaster* including heat knockdown, desiccation, starvation, and chill coma resistance are known to alter basal gene expression in adults (Sorensen et al. 2007; Telonis-Scott et al. 2009). Of the five coding regions of the 13 genes sequenced, only one base substitution resulted in an amino acid change from an isoleucine to a threonine at position 3R:20722513 of the CG11069 gene. For this subset of genes, the level of synonymous substitutions and differentiation around UTRs and predicted promoter regions suggest that gene regulation is affected by selection for desiccation resistance.

Overlap with Previous Studies

Although our large number of SFPs did not overlap with the few known desiccation candidate genes (i.e., *desiccate*, *TotA*, *Smp-30*, and *Frost*) (Sinclair et al. 2007; Kawano et al. 2010), we did observe overlap with QTL and genome-wide expression studies. Foley and Telonis-Scott (2011) reported at least 15 QTL affecting female survival to desiccation and strong associations with CHC composition suggested that the epicuticle barrier forms an important component of the survival strategy. Here, SFPs occur within the most significant peak of seven of the desiccation QTL including 1 at 7F (X chromosome), 9 spanning QTL 25B, 28D, 37C, 45C, and 47D-E (2L, 2R), and 3 at QTL 84D-E (3R). This expands considerably to almost half the SFP when the QTL confidence intervals are considered and includes genes such as *Dys*, *mol*, *SN4Agamma*, *KCNQ*, *ct*, *ed*, *CadN*, *trpl*, and members of the immunoglobulin *beat* family. Nonetheless, some overlap is expected by chance given the large genomic intervals encompassed by the QTL (particularly chromosome two, see Foley and Telonis-Scott 2011) but is still suggestive in a comparative sense to previously described desiccation "hotspots." A small fraction of genes (14/262) with altered basal expression patterns in *D. melanogaster* following selection for desiccation resistance including immunuglobulins, tracheal system, metabolic, and sodium symporter functions (Sorensen et al. 2007), whereas functional clusters, such as the immunoglobulin and cell–cell adhesion genes overlapped with the transcriptional response to desiccation in *D. mojavensis* (Matzkin and Markow 2009). In *D. melanogaster*, desiccation resistance can evolve in diverse ways depending on the genetic and environmental backgrounds and given that our mapping is to a much higher resolution than previous studies, it is not surprising that the overlap between studies seems small. However, the genetic changes common to different populations and species provide strong functional candidates for further analysis.

Conclusions

We have found that tiling arrays provide an effective way of rapidly mapping genomic regions differentiated between

selected and control lines. By having a large number of replicate lines, regions that have repeatedly and independently differentiated can be readily identified with some confidence and used to suggest pathways and mechanisms underlying evolved shifts in stress resistance. We independently corroborated the association between the desiccation resistant phenotype and a potential *cis*-regulatory region of the *Dys* locus in a natural population. This study also provides numerous candidates for further genetic analysis of diverged populations and species, with the eventual goal of understanding genetic opportunities and constraints for resistance evolution and shifts in the climatic niche of species.

Data Accessibility

Sequence data from individual flies (Coffs Harbour flies) are deposited with the EMBL/GenBank data libraries under accession numbers JN886039–JN886084.

Supplementary Material

Supplementary files 1, 2, 3, and 4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We are grateful to Lauren McIntyre for valuable statistical advice. We also thank Robert Good for bioinformatic support and Charles Robin as well as Sui-Fai (Ronald) Lee for helpful discussions. Finally, we thank Jennifer Shirriffs, Lea Rako, Allen Rako, Vanessa White for technical assistance, and Tom Turner and an anonymous reviewer for comments that improved the manuscript. This work was supported by Discovery and Laureate Fellowship grants from the Australian Research Council.

References

Addo-Bediako A, Chown SL, Gaston KJ. 2001. Revisiting water loss in insects: a large scale view. *J Insect Physiol.* 47:1377–1388.

Affolter M, Caussinus E. 2008. Tracheal branching morphogenesis in *Drosophila*: new insights into cell behaviour and organ architecture. *Development* 135(12):2055–2064.

Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2):263–265.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Series B-Methodol.* 57(1):289–300.

Beyenbach KW. 2003. Transport mechanisms of diuresis in Malpighian tubules of insects. *J Exp Biol.* 206(21):3845–3856.

Boher FBF, Godoy-Herrera R, Bozinovic F. 2010. The interplay between thermal tolerance and life history is associated with the biogeography of Drosophila species. *Evol Ecol Res.* 12(8):973–986.

Borevitz JO, Hazen SP, Michael TP, et al. (15 co-authors). 2007. Genome-wide patterns of single-feature polymorphism in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A.* 104(29):12057–12062.

Borevitz JO, Liang D, Plouffe D, Chang HS, Zhu T, Weigel D, Berry CC, Winzeler E, Chory J. 2003. Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Res.* 13(3):513–523.

Buffin E, Gho M. 2010. Laser microdissection of sensory organ precursor cells of *Drosophila* microchaetes. *Plos One* 5(2):e9285.

Chenoweth SF, Visscher PM. 2009. Association Mmpping in outbred populations: power and efficiency when genotyping parents and phenotyping progeny. *Genetics* 181(2):755–765.

Chihara CJ, Song C, LaMonte G, Fetalvero K, Hinchman K, Phan H, Pineda M, Robinson K, Schneider GP. 2005. Identification and partial characterization of the enzyme of omega: one of five putative DPP IV genes in *Drosophila melanogaster*. *J Insect Sci.* 5:26.

Chinnusamy V, Schumaker K, Zhu JK. 2004. Molecular genetic perspectives on cross-talk and specificity in abiotic stress signalling in plants. *J Exp Bot.* 55(395):225–236.

Chippindale AK, Gibbs AG, Sheik M, Yee KJ, Djawdan M, Bradley TJ, Rose MR. 1998. Resource acquisition and the evolution of stress resistance in *Drosophila melanogaster*. *Evolution* 52(5):1342–1352.

Coast GM, Orchard I, Phillips JE, Schooley DA. 2002. Insect diuretic and antidiuretic hormones. In: Evans PD, editor. Advances in insect physiology. Vol. 29. London: Academic Press Ltd. p. 279–409.

David JR, Allemand R, Capy P, Chakir M, Gibert P, Petavy G, Moreteau B. 2004. Comparative life histories and ecophysiology of Drosophila melanogaster and D. simulans. *Genetica* 120(1–3):151–163.

Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. 2003. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4:R60.

Dow JAT, Davies SA. 2001. The *Drosophila melanogaster* malpighian tubule. In: Advances in insect physiology. Salt Lake City (UT): Academic Press. p. 1–83.

Dubnau J, Chiang AS, Grady L, et al. (12 co-authors). 2003. The staufen/pumilio pathway is involved in *Drosophila* long-term memory. *Curr Biol* 13(4):286–296.

Foley BR, Telonis-Scott M. 2011. Quantitative genetic analysis suggests causal association between cuticular hydrocarbon composition and desiccation survival in *Drosophila melanogaster*. *Heredity* 106(1):68–77.

Folk DG, Bradley TJ. 2005. Adaptive evolution in the lab: unique phenotypes in fruit flies comprise a fertile field of study. *Integr Comp Biol.* 45(3):492–499.

Folk DG, Han C, Bradley TJ. 2001. Water acquisition and partitioning in *Drosophila melanogaster*: effects of selection for desiccation-resistance. *J Exp Biol.* 204(19):3323–3331.

Gade G. 2004. Regulation of intermediary metabolism and water balance of insects by neuropeptides. *Annu Rev Entomol.* 49:93–113.

Gibbs AG, Fukuzato F, Matzkin LM. 2003. Evolution of water conservation mechanisms in *Drosophila*. *J Exp Biol.* 206(7):1183–1192.

Gibbs AG, Matzkin LM. 2001. Evolution of water balance in the genus *Drosophila*. *J Exp Biol.* 204(13):2331–2338.

Graze RM, McIntyre LM, Main BJ, Wayne ML, Nuzhdin SV. 2009. Regulatory divergence in *Drosophila melanogaster* and *D. simulans*, a genomewide analysis of allele-specific expression. *Genetics* 183(2):547–561.

Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P. 2005. Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res.* 15(6):790–799.

Harshman LG, Hoffmann AA. 2000. Laboratory selection experiments using *Drosophila*: what do they really tell us? *Trends Ecol Evol.* 15(1):32–36.

Hoffmann AA. 2010. Physiological climatic limits in *Drosophila*: patterns and implications. *J Exp Biol.* 213(6):870–880.

Hoffmann AA, Daborn PJ. 2007. Towards genetic markers in animal populations as biomonitors for human-induced environmental change. *Ecol Lett.* 10(1):63–76.

Hoffmann AA, Hallas RJ, Dean JA, Schiffer M. 2003. Low potential for climatic stress adaptation in a rainforest *Drosophila* species. *Science* 301(5629):100–102.

Hoffmann AA, Harshman LG. 1999. Desiccation and starvation resistance in *Drosophila*: patterns of variation at the species, population and intrapopulation levels. *Heredity* 83:637–643.

Hoffmann AA, Parsons PA. 1989. An integrated approach to environmental stress tolerance and life-history variation—desiccation tolerance in *Drosophila*. *Biol J Linn Soc.* 37(1–2): 117–136.

Hoffmann AA, Sgro CM. 2011. Climate change and evolutionary adaptation. *Nature* 470(7335):479–485.

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 4(1):44–57.

Hynes RO, Zhao Q. 2000. The evolution of cell adhesion. *J Cell Biol.* 150(2):F89–F95.

Kahsai L, Kapan N, Dircksen H, Winther AME, Nassel DR. 2010. Metabolic stress responses in *Drosophila* are modulated by brain neurosecretory cells that produce multiple neuropeptides. *Plos One* 5(7):e11480.

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(50):38889–38897.

Kellermann V, van Heerwaarden B, Sgro CM, Hoffmann AA. 2009. Fundamental evolutionary limits in ecological traits drive *Drosophila* species distributions. *Science* 325(5945):1244–1246.

Kimura K, Shimozawa T, Tanimura T. 1985. Water-loss through the integument in the desiccation-sensitive mutant *parched*, of *Drosophila melanogaster*. *J Insect Physiol* 31(7):573–580.

Kramer JM, Slade JD, Staveley BE. 2008. *foxo* is required for resistance to amino acid starvation in *Drosophila*. *Genome* 51(8):668–672.

Lai CQ, Leips J, Zou W, Roberts JF, Wollenberg KR, Parnell LD, Zeng ZB, Ordovas JM, Mackay TFC. 2007. Speed-mapping quantitative trait loci using microarrays. *Nat Methods.* 4(10):839–841.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10(3):10.

Le Lagadec MD, Chown SL, Scholtz CH. 1998. Desiccation resistance and water balance in southern African keratin beetles (Coleoptera: Trogidae): the influence of body size and habitat. *J Comp Physiol B-Biochem Syst Environ Physiol.* 168:112–122.

Lehmann FO, Schutzner P. 2010. The respiratory basis of locomotion in *Drosophila*. *J Insect Physiol.* 56(5):543–550.

Liu K, Tsujimoto H, Cha SJ, Agre P, Rasgon JL. 2011. Aquaporin water channel *AgAQP1* in the malaria vector mosquito *Anopheles gambiae* during blood feeding and humidity adaptation. *Proc Natl Acad Sci U S A.* 108(15):6062–6066.

Maness PF, Schachner M. 2007. Neural recognition molecules of the immunoglobulin superfamily: signaling transducers of axon guidance and neuronal migration. *Nat Neurosci.* 10(1):19–26.

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(3):261–270.

Mathew D, Popescu A, Budnik V. 2003. *Drosophila* amphiphysin functions during synaptic Fasciclin II membrane cycling. *J Neurosci.* 23(33):10710–10716.

Matzkin LM, Markow TA. 2009. Transcriptional regulation of metabolism associated with the increased desiccation resistance of the cactophilic *Drosophila mojavensis*. *Genetics* 182(4):1279–1288.

Neckameyer WS, Matsuo H. 2008. Distinct neural circuits reflect sex, sexual maturity, and reproductive status in response to stress in *Drosophila melanogaster*. *Neurosci* 156(4):841–856.

Neuman S, Kaban A, Volk T, Yaffe D, Nudel U. 2001. The dystrophin/utrophin homologues in *Drosophila* and in sea urchin. *Gene* 263(1–2):17–29.

Neuman S, Kovalio M, Yaffe D, Nudel U. 2005. The *Drosophila* homologue of the *dystrophin* gene—introns containing promoters are the major contributors to the large size of the gene. *FEBS Lett.* 579(24):5365–5371.

Nuzhdin SV, Harshman LG, Zhou M, Harmon K. 2007. Genome-enabled hitchhiking mapping identifies QTLs for stress resistance in natural *Drosophila*. *Heredity* 99(3):313–321.

Parts L, Cubillos FA, Warringer J, et al. (14 co-authors). 2011. Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Res.* 21(7):1131–1138.

Pipes GC, Lin Q, Riley SE, Goodman CS. 2001. The *Beat* generation: a multigene family encoding IgSF proteins related to the Beat axon guidance molecule in *Drosophila*. *Development* 128(22):4545–4552.

Rostoks N, Borevitz JO, Hedley PE, Russell J, Mudie S, Morris J, Cardle L, Marshall DF, Waugh R. 2005. Single-feature polymorphism discovery in the barley transcriptome. *Genome Biol.* 6(6):R54.

Rougon G, Hobert O. 2003. New insights into the diversity and function of neuronal immunoglobulin superfamily molecules. *Annu Rev Neurosci.* 26:207–238.

Ruis H, Schüller C. 1995. Stress signaling in yeast. *BioEssays* 17(11):959–965.

Samakovlis C, Hacohen N, Manning G, Sutherland DC, Guillemin K, Krasnow MA. 1996. Development of the *Drosophila* tracheal system occurs by a series of morphologically distinct but genetically coupled branching events. *Development* 122(5):1395–1407.

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(4):435–443.

Sorensen JG, Nielsen MM, Loeschcke V. 2007. Gene expression profile analysis of *Drosophila melanogaster* selected for resistance to environmental stressors. *J Evol Biol.* 20(4): 1624–1636.

Strachan LA, Tarnowski-Garner HE, Marshall KE, Sinclair BJ. 2011. The evolution of cold tolerance in Drosophila larvae. *Physiol Biochem Zool.* 84(1):43–53.

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(10):1837–1847.

Telonis-Scott M, Hallas R, McKechnie SW, Wee CW, Hoffmann AA. 2009. Selection for cold resistance alters gene transcript levels in *Drosophila melanogaster*. *J Insect Physiol.* 55(6):549–555.

Turner TL, Hahn MW, Nuzhdin SV. 2005. Genomic islands of speciation in *Anopheles gambiae*. *Plos Biol.* 3(9):1572–1578.

Turner TL, Levine MT, Eckert ML, Begun DJ. 2008. Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. *Genetics* 179(1):455–473.

Turner TL, Stewart AD, Fields AT, Rice WR, Tarone AM. 2011. Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in Drosophila melanogaster. *Plos Genet.* 7(3):e1001336.

Tweedie S, Ashburner M, Falls K, et al. (11 co-authors). 2009. FlyBase: enhancing *Drosophila* gene ontology annotations. *Nucleic Acids Res.* 37:D555–D559.

van der Plas MC, Pilgram GS, Plomp JJ, de Jong A, Fradkin LG, Noordermeer JN. 2006. *Dystrophin* is required for appropriate retrograde control of neurotransmitter release at the *Drosophila* neuromuscular junction. *J Neurosci.* 26(1):333–344.

van Herrewege J, David JR. 1997. Starvation and desiccation tolerances in *Drosophila*: comparison of species from different climatic origins. *Ecoscience* 4:151–157.

Vogel C, Teichmann SA, Chothia C. 2003. The immunoglobulin superfamily in *Drosophila melanogaster* and *Caenorhabditis elegans* and the evolution of complexity. *Development* 130(25):6317–6328.

Wang J, Kean L, Yang JL, Allan AK, Davies SA, Herzyk P, Dow JAT. 2004. Function-informed transcriptome analysis of *Drosophila* renal tubule. *Genome Biol.* 5(9):R69.

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(6):756–763.

Zhang L, Miles MF, Aldape KD. 2003. A model of molecular interactions on short oligonucleotide microarrays. *Nat Biotechnol.* 21(7):818–821.