

# Habitat Variability Correlates with Duplicate Content of *Drosophila* Genomes

Takashi Makino,<sup>\*,1</sup> and Masakado Kawata<sup>\*,1</sup>

<sup>1</sup>Department of Ecology and Evolutionary Biology, Graduate School of Life Sciences, Tohoku University, Sendai, Japan

\*Corresponding author: E-mail: tamakino@m.tohoku.ac.jp; kawata@m.tohoku.ac.jp.

Associate editor: Manolo Gouy

## Abstract

The factors limiting the habitat range of species are crucial in understanding their biodiversity and response to environmental change. Yet the genetic and genomic architectures that produce genetic variation to enable environmental adaptation have remained poorly understood. Here we show that the proportion of duplicated genes ( $P_D$ ) in the whole genomes of fully sequenced *Drosophila* species is significantly correlated with environmental variability within the habitats measured by the climatic envelope and habitat diversity. Furthermore, species with a low  $P_D$  tend to lose the duplicated genes owing to their faster evolution. These results indicate that the rapid relaxation of functional constraints on duplicated genes resulted in a low  $P_D$  for species with lower habitat diversity, and suggest that the maintenance of duplicated genes gives organisms an ecological advantage during evolution. We therefore propose that the  $P_D$  in a genome is related to adaptation to environmental variation.

**Key words:** habitat diversity, habitat distributions, evolvability, adaptation, duplicated genes.

## Introduction

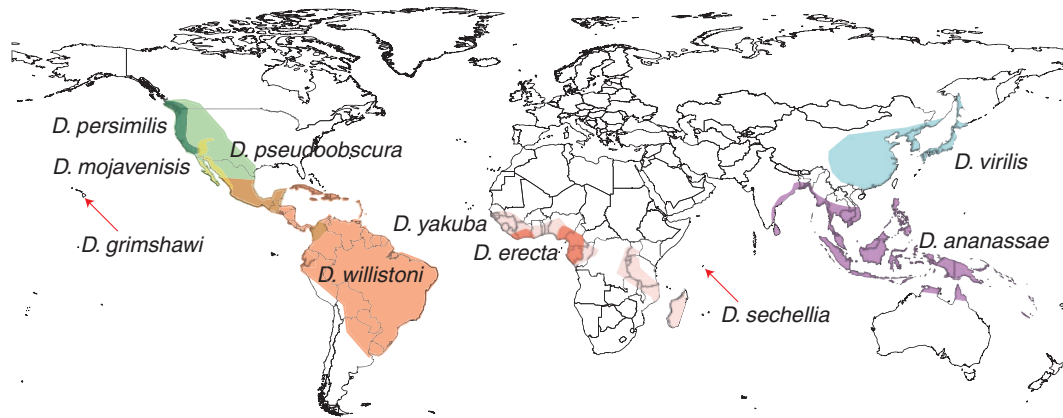
The factors that constrain the evolution of habitat range is of critical importance for understanding the evolution of biodiversity and conservation because these factors are closely related to historical processes creating the current biodiversity and to adaptation to current and future global climate changes (Root et al. 2003; Bridle and Vines 2007; Roy et al. 2009). Even within closely related groups like *Drosophila*, some species have narrow restricted ranges and inhabit one or a few habitat types, whereas others have wider ranges and live in diverse environments (Kirkpatrick and Barton 1997). Kellermann et al. (2009) showed that low genetic variation in cold and desiccation tolerance limits the distributions of species. This indicates that a lack of genetic variation in key traits within a species is related to failure to expand in range and adapt to environmental change. However, what determines the ability of species to generate genetic variation (i.e., their evolvability) remains unknown. Restricted-range species might have genetic and genomic architectures that do not allow high variation.

We focused on gene duplication as a source of genome-wide genetic variation. One system that produces and maintains a large amount of genetic variation is buffering. By buffering the deleterious consequences of mutations, genetic variation can accumulate in a genome. An obvious mechanism for buffering genetic variation is redundancy, and one of the main factors that generates redundancy is gene duplication (Wilkins 1997; Hartman et al. 2001). This type of mutation is particularly common in eukaryotes, and in yeast, duplication rates are reportedly faster than point mutation rates (Lynch et al. 2008). After gene duplication, one of the pair is redundant, and as such, the functional constraints are

relaxed, and one or both copies can differentiate as long as their original function is maintained (Ohno 1970). Therefore, under these relaxed functional constraints, mutations are likely to accumulate in duplicated genes. Furthermore, the functional redundancy of duplicated genes can be maintained for extensive periods of time (Dean et al. 2008). Genetic variations in the duplicated genes within a population are likely to be maintained by their buffering effect (Wilkins 1997; Hartman et al. 2001). Therefore, gene duplication is a major source of genetic variation. In fact, Kliebenstein showed that not only younger tandem duplicated genes but also older duplicated genes elevated intraspecific gene expression variation in a population (Kliebenstein 2008).

Previous studies reported that duplicated olfactory receptor (*Or*) and gustatory receptor (*Gr*) genes were likely to be lost in specialist *Drosophila* species with host specificity; however, these studies focused only on the particular gene families associated with odor response, and did not consider the relationship between the duplicated gene content and habitat (McBride 2007; McBride et al. 2007). The habitats of specialists are necessarily restricted by those of their hosts; hence, host specificity would be expected to be related to habitat (Markow and O'Grady 2007). If species with a larger proportion of duplicated genes ( $P_D$ ) have a greater potential to generate genetic variation for more traits, they might show increased environmental adaptability. Namely, duplicated genes would have contributed to adaptation to diverse environments within the ranges of species.

We propose that species with a higher duplicated gene content would have distribution ranges with higher environmental variability. We tested this hypothesis using *Drosophila* species that had been fully sequenced (Clark et al. 2007) and



**Fig. 1.** Habitat distributions of *Drosophila* species. Habitat distributions of *D. yakuba* (pink), *D. erecta* (red), *D. ananassae* (purple), *D. pseudoobscura* (light green), *D. persimilis* (green), *D. willistoni* (orange), *D. mojavensis* (yellow), and *D. virilis* (blue) are shown. Red arrows indicate islands inhabited by island endemic species (*D. sechellia* and *D. grimshawi*). The habitat distribution of the cosmopolitan species *D. melanogaster* is shown in figure S1, [Supplementary Material](#) online.

had documented habitat ranges (figs. 1 and S1, [Supplementary Material](#) online) (Markow and O'Grady 2006). Environmental variability within their habitat range (here referred to as habitat variability) was measured using two different indices, one relating to climatic envelope and the other to habitat diversity. We estimated their climatic envelopes using bioclimatic variables from WORLDCLIM (Hijmans et al. 2005). Variability of Köppen climate classification (Kottek et al. 2006) within their range was estimated as habitat diversity using the Brillouin's index as a measure of species diversity (Margalef 1958; Legendre and Legendre 1998). These indices were used as measurements to indicate the adaptability of species to environmental variability within their habitats. Examining the effect of duplicated genes on the expansion, contraction, and/or conservation of the habitat ranges of these species during their evolution allowed us to explore the importance of basal genetic diversity for adaptation.

## Materials and Methods

### Fully Sequenced *Drosophila* Species

The genomes of 12 *Drosophila* species (*Drosophila melanogaster*, *D. sechellia*, *D. simulans*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. mojavensis*, *D. virilis*, and *D. grimshawi*) have been fully sequenced (Clark et al. 2007). However, the coverage of the genome assemblies for *D. simulans* is comparatively poor. As a result, the number of identified orthologs of *D. melanogaster* in the *D. simulans* genome is relatively low (Heger and Ponting 2007), even though these are among the closest related *Drosophila* species. We therefore excluded *D. simulans* from our analyses.

### *Drosophila* Gene Sequences

The protein sequences corresponding to protein-encoding genes from the 11 *Drosophila* species were downloaded from the EnsemblMetazoa database, release 4 (<http://metazoa.ensembl.org>). In some cases, a non-*melanogaster* gene was split into two genes as a result of sequence or assembly errors

(fig. S2A, [Supplementary Material](#) online). To minimize these errors, we conducted a homology search using the Basic Local Alignment Search Tool (BLAST; *D. melanogaster* protein sequences vs. non-*melanogaster* protein sequences), combined the sequences of physically neighboring genes in a non-*melanogaster* genome into one sequence when the neighboring genes did not show homology to each other ( $E$  value  $< 10^{-5}$ ), and identified the best hit for the same gene among those in *D. melanogaster* (table S1, [Supplementary Material](#) online). The merged genes were treated as a single gene in this study. For non-*melanogaster* species, we also combined the nucleotide sequences of separate genes by the same process.

### Duplicated Genes

To identify duplicated genes in the *Drosophila* genomes, we conducted an all-to-all BLAST search for all protein sequences used in this study. Genes with a homologue ( $E$  value  $< 10^{-5}$  and query coverage  $> 30\%$ ) in the same species were identified as candidate duplicated genes. Importantly, we found that our results were robust at different  $E$  value cut-offs ( $10^{-5}$ ,  $10^{-10}$ , and  $10^{-20}$ ), and the trend in our results did not change as a result of different cut-off values.

### Synonymous Substitution Rate between Duplicated Genes

To examine the distribution of gene duplication timing for the duplicated genes, we aligned the sequences of the duplicated gene pairs derived from the EnsemblMetazoa database by using the T-COFFEE multiple sequence alignment program (Notredame et al. 2000) and estimated the synonymous substitution rate ( $K_S$ ) between a duplicated gene and its closest paralogue by the Yang and Nielsen method (Yang and Nielsen 2000), which was implemented in the Phylogenetic Analysis by Maximum Likelihood (PAML) program package (Yang 1997). Note that the closest paralogues were determined by identifying the best BLAST hit from the duplicated gene candidates. The distributions of  $K_S$  are shown in figure S3, [Supplementary Material](#) online.

### Gene Collapse Based on Sequence Similarity

There are many non-divergent duplicated gene pairs created as a result of recent duplication events or assembly errors in a genome. Homologous gene pairs with  $K_S < 0.1$  were collapsed into a single gene (a2–a3–a4 and b1–b2 in fig. S2B, [Supplementary Material](#) online). The collapsed genes were classified as duplicated genes on the basis of a BLAST hit in comparison with genes not recently duplicated. If at least one gene in the homologous gene cluster (a2, a3, and a4 in fig. S2B, [Supplementary Material](#) online) had a duplicated gene partner (a1 in fig. S2B, [Supplementary Material](#) online;  $K_S \geq 0.1$ ;  $E$  value  $< 10^{-5}$  and query coverage  $> 30\%$ ), the collapsed gene was defined as a duplicated gene. If not, the collapsed gene (b1–b2 in fig. S2B, [Supplementary Material](#) online) was defined as a singleton. The duplicated genes in *Drosophila* are summarized in table S2, [Supplementary Material](#) online.

### Lineage-Specific Gene Losses

Orthologs were defined by reciprocal best hits between different species by using the results of the all-to-all BLAST. If the orthologous relationship was obtained by one-to-one best hits, we defined the orthologs as one-to-one. Such a relationship indicated that there has been no lineage-specific gene duplication and loss after speciation. We identified orthologous gene clusters using one-to-one orthologous relationships for closely related species and their outgroups to investigate gene-loss events during evolution (fig. S4A, [Supplementary Material](#) online). We did not use genes without orthologs in the outgroups, because it was not easy to predict their ancestral state. If there was no gene-loss event in either of the closely related species, we obtained orthologous trios where possible (e.g., species 1B–species 2B–outgroup B in fig. S4B, [Supplementary Material](#) online). When orthologous trios were not available, we inferred that gene-loss events occurred in either lineage of the closely related species (fig. S4A, [Supplementary Material](#) online). For the comparison, we typically used *D. melanogaster* as the outgroup. When we investigated gene-loss events for species that were in the clade including *D. melanogaster*, we used other closely related species as the outgroups (tables 1 and 2). Note that there were other possible outgroups for the comparisons of some species; however, even when we used other outgroups for estimating the proportion of lost duplicated genes, our result did not change. To identify a species' lost duplicated genes generated before speciation from another species, we focused on the gene similarity between a species and the outgroup species (fig. S4B, [Supplementary Material](#) online). We defined a species (e.g., species 1 in fig. S4B, [Supplementary Material](#) online) as having a lost duplicated gene (species 1A in fig. S4B, [Supplementary Material](#) online) when the following were observed: an inferred ortholog (species 2A in fig. S4B, [Supplementary Material](#) online) in the compared species (species 2 in fig. S4B, [Supplementary Material](#) online) was a duplicated gene, and the similarity between the duplicated gene and its duplicated gene partner (similarity between species 2A and species 2B in fig. S4B, [Supplementary Material](#) online) was lower than that between either of the duplicated

**Table 1.** Differences in Evolutionary Events of Genes between Closely Related Species.

Species	Outgroup	Habitat Variability		$P_D$	Gene Losses		Fast Evolving Genes		Duplicated Gene Pairs with Fast Evolving Genes	
		Climatic Tolerance	Habitat Diversity		Number of Lineage-Specific Gene Losses	$P_D$ for Lineage-Specific Lost Genes	Number of One-to-One Orthologous Trios	Number of Significant Fast-Evolving Genes	Number of Duplicated Gene Pairs in Orthologous Trios	Number of Duplicated Gene Pairs Having Significant Fast Evolving Gene
<i>Drosophila melanogaster</i>	<i>D. yakuba</i>	1,119	3.11	0.53	497	133/497	11,952	482/11,952	3,193	104/3,193
<i>D. sechellia</i>	1	1	0	0.49	430	270/430		1,441/11,952	433/3,193	
<i>P</i> value						$< 2.2 \times 10^{-16}$		$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	
<i>D. pseudoobscura</i>	<i>D. melanogaster</i>	599	2.71	0.53	204	98/204	10,995	203/10,995	2,599	53/2,599
<i>D. persimilis</i>		287	1.18	0.5	401	229/401		1,502/10,995	507/2,599	
<i>P</i> value						0.042		$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	
<i>D. yakuba</i>	<i>D. melanogaster</i>	240	1.87	0.5	304	149/304	12,156	712/12,156	3,472	352/3,472
<i>D. erecta</i>		65	1.05	0.5	259	145/259		1,082/12,156	608/3,472	
<i>P</i> value						0.12		$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$	

**Table 2.** Comparison of Lost Duplicated Genes.

Species	Species 1 with Higher Climatic Tolerance				Species 2 with Lower Climatic Tolerance				Outgroup	P Values ( $\chi^2$ Test of the Significant Difference in the Lost Duplicated Genes)	
	Climatic Tolerance	Number of Lost Duplicated Genes	Total Number of Lost Genes	Proportion of Lost Duplicated Genes	Species	Climatic Tolerance	Number of Lost Duplicated Genes	Total Number of Lost Genes			Proportion of Lost Duplicated Genes
<i>Drosophila ananassae</i>	385	114	269	0.42	<i>D. sechellia</i>	1	281	458	0.61	<i>D. pseudoobscura</i>	$1.1 \times 10^{-6}$
<i>D. ananassae</i>	385	148	310	0.48	<i>D. yakuba</i>	240	198	390	0.51	<i>D. pseudoobscura</i>	
<i>D. ananassae</i>	385	128	292	0.44	<i>D. erecta</i>	65	147	266	0.55	<i>D. pseudoobscura</i>	0.0090
<i>D. ananassae</i>	385	137	277	0.49	<i>D. persimilis</i>	287	362	635	0.57	<i>D. willistoni</i>	0.042
<i>D. erecta</i>	65	84	171	0.49	<i>D. sechellia</i>	1	207	355	0.58	<i>D. ananassae</i>	
<i>D. melanogaster</i>	1,119	133	497	0.27	<i>D. sechellia</i>	1	270	430	0.63	<i>D. yakuba</i>	$<2.2 \times 10^{-16}$
<i>D. melanogaster</i>	1,119	94	287	0.33	<i>D. yakuba</i>	240	142	283	0.50	<i>D. ananassae</i>	$3.5 \times 10^{-5}$
<i>D. melanogaster</i>	1,119	88	285	0.31	<i>D. erecta</i>	65	105	174	0.60	<i>D. ananassae</i>	$1.1 \times 10^{-9}$
<i>D. melanogaster</i>	1,119	152	333	0.46	<i>D. ananassae</i>	385	133	278	0.48	<i>D. pseudoobscura</i>	
<i>D. melanogaster</i>	1,119	135	294	0.46	<i>D. persimilis</i>	287	361	599	0.60	<i>D. willistoni</i>	$6.8 \times 10^{-5}$
<i>D. melanogaster</i>	1,119	164	324	0.51	<i>D. pseudoobscura</i>	599	223	394	0.57	<i>D. willistoni</i>	
<i>D. melanogaster</i>	1,119	187	366	0.51	<i>D. willistoni</i>	462	218	448	0.49	<i>D. virilis</i>	
<i>D. mojavensis</i>	133	109	233	0.47	<i>D. grimshawi</i>	28	221	410	0.54	<i>D. melanogaster</i>	
<i>D. persimilis</i>	287	326	574	0.57	<i>D. sechellia</i>	1	254	414	0.61	<i>D. willistoni</i>	
<i>D. persimilis</i>	287	349	594	0.59	<i>D. yakuba</i>	240	185	346	0.53	<i>D. willistoni</i>	
<i>D. persimilis</i>	287	362	628	0.58	<i>D. erecta</i>	65	135	231	0.58	<i>D. willistoni</i>	0.011
<i>D. pseudoobscura</i>	599	200	373	0.54	<i>D. sechellia</i>	1	285	455	0.63	<i>D. willistoni</i>	
<i>D. pseudoobscura</i>	599	207	382	0.54	<i>D. yakuba</i>	240	202	376	0.54	<i>D. willistoni</i>	
<i>D. pseudoobscura</i>	599	222	394	0.56	<i>D. erecta</i>	65	148	248	0.60	<i>D. willistoni</i>	
<i>D. pseudoobscura</i>	599	205	382	0.54	<i>D. ananassae</i>	385	156	293	0.53	<i>D. willistoni</i>	
<i>D. pseudoobscura</i>	599	98	204	0.48	<i>D. persimilis</i>	287	229	401	0.57	<i>D. melanogaster</i>	0.042
<i>D. pseudoobscura</i>	599	223	404	0.55	<i>D. willistoni</i>	462	229	479	0.48	<i>D. virilis</i>	0.034
<i>D. virilis</i>	362	81	178	0.46	<i>D. mojavensis</i>	133	144	262	0.55	<i>D. melanogaster</i>	
<i>D. virilis</i>	362	78	167	0.47	<i>D. grimshawi</i>	28	246	446	0.55	<i>D. melanogaster</i>	$6.0 \times 10^{-8}$
<i>D. willistoni</i>	462	191	438	0.44	<i>D. sechellia</i>	1	303	492	0.62	<i>D. virilis</i>	0.037
<i>D. willistoni</i>	462	213	458	0.47	<i>D. yakuba</i>	240	228	424	0.54	<i>D. virilis</i>	
<i>D. willistoni</i>	462	204	453	0.45	<i>D. erecta</i>	65	170	304	0.56	<i>D. virilis</i>	0.0042
<i>D. willistoni</i>	462	220	473	0.47	<i>D. ananassae</i>	385	180	330	0.55	<i>D. virilis</i>	0.030
<i>D. willistoni</i>	462	208	449	0.46	<i>D. persimilis</i>	287	379	638	0.59	<i>D. virilis</i>	$2.7 \times 10^{-5}$
<i>D. yakuba</i>	240	112	267	0.42	<i>D. sechellia</i>	1	207	354	0.58	<i>D. ananassae</i>	$6.4 \times 10^{-5}$
<i>D. yakuba</i>	240	149	304	0.49	<i>D. erecta</i>	65	145	259	0.56	<i>D. melanogaster</i>	



copies and its best hit homolog in the outgroup (similarity between species 2A and outgroup A or that between species 2B and outgroup B in fig. S4B, [Supplementary Material](#) online), as determined by a BLAST search.

### Relative Rate Test for Orthologous Gene Pairs

To detect fast-evolving genes after the speciation of closely related species (fig. S4C, [Supplementary Material](#) online), we conducted a relative rate test using protein sequences aligned by T-COFFEE for the following closely related species and their outgroups: *D. melanogaster*, *D. sechellia*, *D. yakuba* (outgroup); *D. pseudoobscura*, *D. persimilis*, *D. melanogaster* (outgroup); and *D. yakuba*, *D. erecta*, *D. melanogaster* (outgroup) (Tajima 1993). We used the orthologous trios of closely related species and their outgroups obtained earlier, and counted the number of significant fast-evolving genes for each species ([table 1](#)).

### Divergence of Duplicated Gene Pairs in Orthologous Trios

We examined whether an acceleration of the evolutionary rates of duplicated genes occurred in species with low habitat variability and low  $P_D$ . We focused on all the duplicated gene pairs in the aforementioned dataset of orthologous trios, to minimize any effect of lineage-specific extra gene duplications on evolutionary rates (fig. S4D, [Supplementary Material](#) online). Note that as long as a relationship among the three species is observed, no gene loss events have occurred in any of the lineages of the orthologous trios. An extra gene copy generated by lineage-specific gene duplication might cause a relaxation of the functional constraints on the gene copies in the lineage; therefore, we used duplicated gene pairs derived from a duplication event before the speciation of the closely related species and their outgroups. Note that no recent gene duplication events occurred after speciation in the datasets from these trios. We counted the number of duplicated gene pairs in which at least one partner was a significantly fast-evolving gene for each species ([table 1](#)).

### Habitat Area and Habitat Variability

The habitat areas for the *Drosophila* species were obtained from the literature (Ashburner et al. 1982; Piano et al. 1997; Reed and Markow 2004; Markow and O'Grady 2006) (online: <http://scitechlab.wordpress.com/2008/11/02/the-humble-fruit-fly-drosophila-melanogaster>). Habitat variability was estimated from climatic envelope and habitat diversity using the Köppen climate classification. Climatic envelope is the range of temperatures, rainfall, and other climate-related parameters in which a species currently exists. We estimated climatic envelope using principal component analysis (PCA) with WORLDCLIM (Hijmans et al. 2005). We obtained world spacial data and the WORLDCLIM climatic dataset (10 minutes latitude/longitude) from DIVA-GIS (<http://www.diva-gis.org>). The habitat area was measured as the number of grid squares on the climate map. We then extracted the climatic values from 19 bioclimatic variables used for BIOCLIM (Hijmans et al. 2005) in the habitat area of each *Drosophila*

species. We performed PCA using the bioclimatic variables for all of the species, and found that the first 2 principal components (PCs) explained 93.4% of the total variance ([table S3](#), [Supplementary Material](#) online). The contribution of PC1 and PC2 is 79.9% and 13.5%, respectively. PCA plots (x-axis: PC1 and y-axis: PC2) and the correlation circle are shown in figure S5 and S6, [Supplementary Material](#) online, respectively. On the basis of PCA results, we also plotted values of PC1 and PC2 for each species (fig. S5, [Supplementary Material](#) online). We used 107,865 cells (PC1: 799 × PC2: 135) by weighting the relative contribution to PC1 and PC2 for estimating climatic envelope, and defined the number of cell grids overlapping points in the 107,865 cell grids as the climatic envelope of *Drosophila* species.

The Köppen climate classification map was used for estimating the *Drosophila* species' habitat diversity (Kottek et al. 2006). This climate map consists of a grid of squares (0.5° latitude/longitude) in which a certain climate is classified by temperature, precipitation, and vegetation (fig. S1, [Supplementary Material](#) online). The habitat area was measured as the number of grid squares on the climate map. The number of grids varied among the *Drosophila* species, and therefore habitat diversity was calculated using varieties (through logarithmic transformation) in the climatic environment among grid squares for each species using the Brillouin's index, which is robust to sample size (Margalef 1958).

### Model Selection

All of the following statistical analyses were executed in R (<http://www.r-project.org>). We applied model selection using regression to examine which genomic factors affect habitat features (bioclimatic variables and habitat area). We explored the set of predictors of the explanatory variables using the stepwise Akaike's Information Criterion procedure, and determined the set of variables that yielded the lowest score. In addition, we conducted a multivariate analysis of variance (MANOVA) in which two variables (climatic envelope and habitat area) were used as response variables, and genome size, number of genes and  $P_D$  were used as explanatory variables ([tables S4 and S5](#), [Supplementary Material](#) online).

To remove any phylogenetic constraints on the relationship between genetic architecture and habitat, we used a robust phylogeny derived from *Drosophila* 12 Genomes Consortium (Clark et al. 2007). Using this phylogenetic tree, we selected a model by applying the generalized least squares model with the Brownian model, as described earlier, and measured the phylogenetically independent contrasts (PICs) (Felsenstein 1985). We performed linear regression analyses for selected explanatory variables using the estimated PICs.

### Gene Ontology

To investigate whether the lineage-specific lost or fast-evolving duplicated genes of species with low  $P_D$  were enriched in some particular functional categories, we examined the Gene Ontology (GO) database entries for the duplicated genes between species with different  $P_D$

(*D. melanogaster*–*D. sechellia* and *D. pseudoobscura*–*D. persimilis*) The GO identifiers (ids) and GO “slim” annotations for the biological processes of *D. melanogaster* were downloaded from <ftp://ftp.geneontology.org/pub/go/gene-associations/> and [ftp://ftp.geneontology.org/pub/go/GO\\_slims](ftp://ftp.geneontology.org/pub/go/GO_slims), respectively. We excluded those classified as GO:0008150 (biological process unknown). The frequency of each GO id was counted for the *D. melanogaster* genes. For the other species, we used the GO ids of the most similar homolog in *D. melanogaster*. To analyze the GO data for genes that had been lost in the *D. melanogaster* lineage (mel1 in fig. S2C, [Supplementary Material](#) online), we used the GO id of the most similar homolog retained in *D. melanogaster* (mel2 in fig. S2C, [Supplementary Material](#) online) for the orthologs retained in the *D. sechellia* genome (sec1 in fig. S2C, [Supplementary Material](#) online) of the lost genes. The enrichment of GO ids for the genes in species having a low  $P_D$  was compared with that in species having a high  $P_D$ . We calculated the  $P$  value for each GO id by comparing two different gene sets. The estimated  $P$  values were adjusted using Bonferroni correction.

## Results and Discussion

### Recently Duplicated Genes

We identified duplicated genes by similarity search (blastp) for each species and estimated the synonymous substitution rate ( $K_S$ ) between a duplicated gene and its closest paralogue of 11 fully sequenced *Drosophila* species. We observed that the duplicated gene pairs tended to have  $K_S < 0.1$  (fig. S3, [Supplementary Material](#) online). This observation is consistent with high rates of gene duplications and losses (Lynch and Conery 2000). There was apparent bias in the number of recent duplication events in particular species. Although the recent burst of gene duplications observed in some particular lineages is biologically feasible, we question the reliability of the enrichment of recently duplicated gene pairs. Indeed, it is difficult to distinguish recently duplicated genes from artifacts of genome assembly. On the other hand, diverged duplicated genes maintained in a genome are obviously derived from ancient gene duplication events (not artifacts), and most of the substitutions can be attributed to diverged duplicated genes (not recently duplicated genes). Therefore, we focused on diverged duplicated genes in the following analyses, and homologous gene pairs with a  $K_S < 0.1$  were collapsed into a single gene (see Materials and Methods). We found that there was a significant positive correlation between the proportion of total duplicated genes including recently duplicated genes and climatic envelope estimated by bioclimatic variables ( $R^2 = 0.62$ ,  $P = 0.0066$ ; see next section in more detail), but there was no significant correlation between the proportion of only recently duplicated genes and climatic envelope ( $R^2 = 0.26$ ,  $P = 0.13$ ). This observation indicates that in comparison with evolutionarily maintained duplicates, recent duplicates are unlikely to contribute to the climatic envelope. This could be attributed to the lower divergence of recently duplicated genes.

### $P_D$ Associated with Habitat Diversity

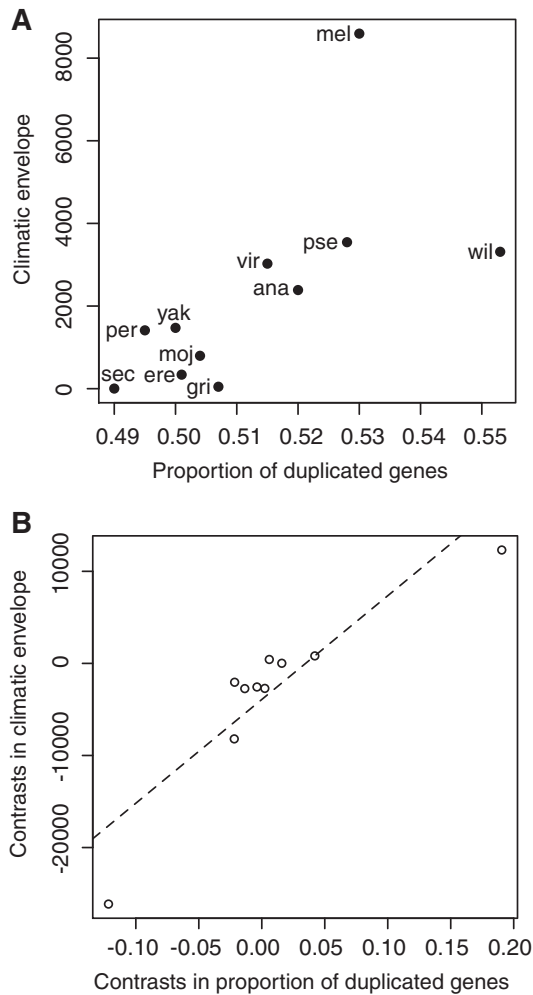
To investigate the relationship between genomic architecture and habitat, we employed a linear model in which genome size (Nardon et al. 2005) from (Bosco et al. 2007), the  $P_D$  and number of genes were used as explanatory variables, and climatic envelope and habitat area were used as response variables, removing the phylogenetic constraints.  $P_D$  was selected as the sole explanatory variable for climatic envelope ( $R^2 = 0.82$ ,  $P = 0.00032$ ; fig. 2).  $P_D$  and number of genes were selected as explanatory variables for habitat area, but the regression coefficient was statistically significant only for the former ( $R^2 = 0.45$ ,  $P = 0.024$ ); these results were not changed by using a MANOVA for the two response variables (table S4, [Supplementary Material](#) online). We then examined the effects of climatic envelope and habitat area on  $P_D$ , and only climatic envelope was selected as an explanatory variable. These results indicate that  $P_D$  is strongly correlated with climatic envelope.

We were concerned that two extreme contrasts might be driving the relationship in figure 2B. The two extreme contrasts were generated by the large climatic envelope values for *D. melanogaster* and *D. pseudoobscura*, which have closely related species with opposite features, that is, *D. sechellia* and *D. persimilis*, respectively. However, even when we removed *D. melanogaster* and *D. pseudoobscura* from our analyses, we observed the same trends ( $R^2 = 0.58$ ,  $P = 0.029$ ; fig. S7A, [Supplementary Material](#) online).

Some species' habitats are known to have been expanded by human activity, particularly in the case of *D. melanogaster*, which has been spread around the world. We therefore repeated the analysis without *D. melanogaster*, and confirmed that the results were not affected (table S4, [Supplementary Material](#) online). It has been reported that *D. virilis* is a holarctic species (Ashburner et al. 1982; Miroi et al. 2008). When *D. virilis* and/or *D. melanogaster* were removed from the analysis, the results did not change (data not shown).

We suspected that the differences in genome coverage among *Drosophila* species might correlate with  $P_D$ , and therefore examined this relationship. We obtained data on the genome coverage of all *Drosophila* species except for *D. melanogaster* from EnsemblMetazoa, and found no correlation between the genome coverage and the  $P_D$  ( $R^2 = 0.014$ ,  $P = 0.76$ ).

To reinforce our results, we investigated the relationship between  $P_D$  and the habitat diversity of *Drosophila* within their range based on Köppen climate classification. The classification considers not only temperature and precipitation but also vegetation (Kotteck et al. 2006). Environmental diversity within the habitat was estimated using the Brillouin's index (Margalef 1958). Similar results were obtained even when we used a different measure of environmental variability with a different climatic dataset ( $R^2 = 0.93$ ,  $P = 7.7 \times 10^{-6}$ ; fig. S8 and table S5, [Supplementary Material](#) online). These results strongly support the contention that in *Drosophila* species,  $P_D$  is correlated with habitat variability.



**Fig. 2.** Correlation between climatic envelope and  $P_D$ . (A) Relationship between climatic envelope and  $P_D$ . The x-axis indicates  $P_D$  for 11 *Drosophila* species (mel, *D. melanogaster*; sec, *D. sechellia*; yak, *D. yakuba*; ere, *D. erecta*; ana, *D. ananassae*; pse, *D. pseudoobscura*; per, *D. persimilis*; wil, *D. willistoni*; moj, *D. mojavensis*; vir, *D. virilis*; and gri, *D. grimshawi*). The y-axis indicates climatic envelope estimated by WORLDCLIM datasets. (B) Relationship between contrasts in climatic envelope and  $P_D$ . The x-axis indicates PICs in  $P_D$ , and the y-axis indicates PICs in the climatic envelope. The dashed line represents the regression line.

### The Influence of Effective Population Size on Genomic Architecture

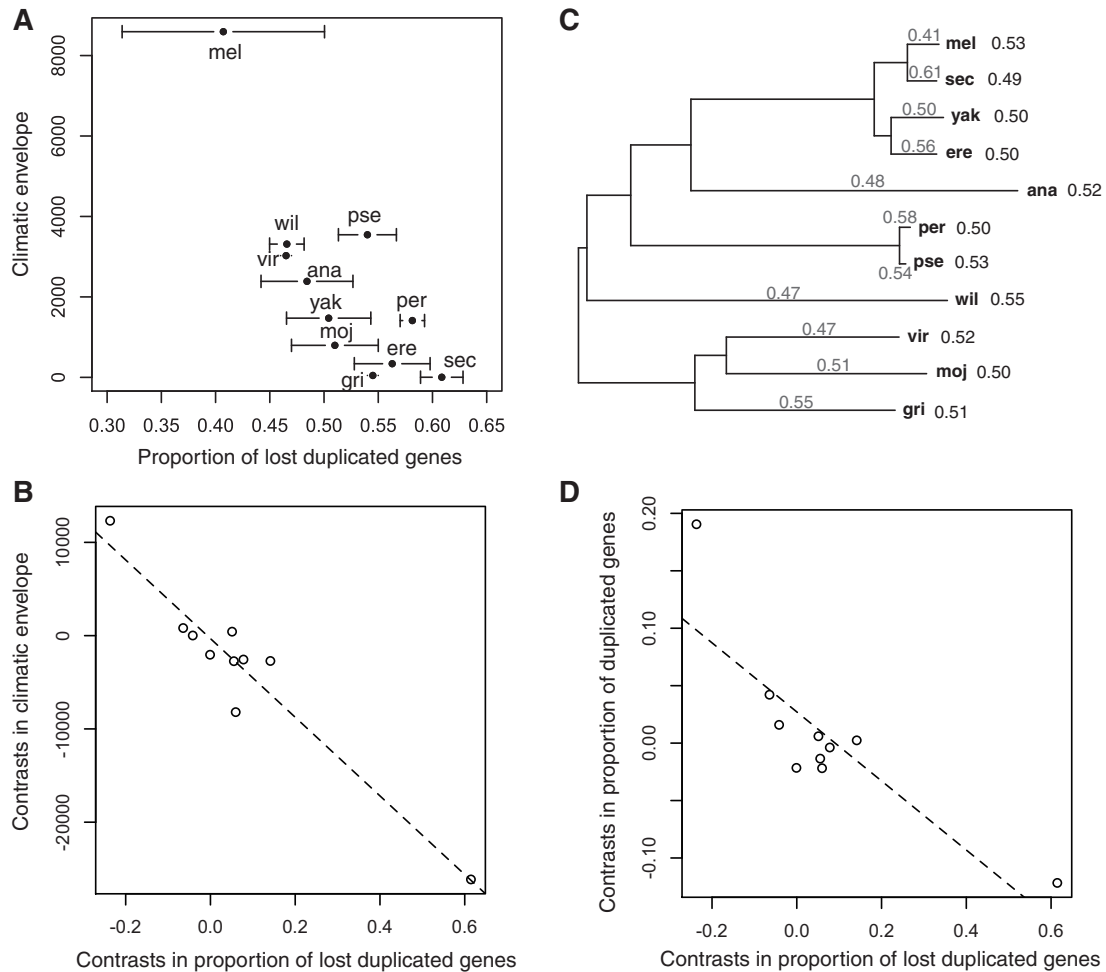
It has been proposed that there are correlations between effective population size and genomic contents (Lynch and Conery 2003). In the genus *Drosophila*, several studies have shown that the evolutionary rates of genes are faster in the host-specific species *D. sechellia*, which has a small effective population size, than in the cosmopolitan species *D. simulans* (Kliman et al. 2000; McBride 2007). Similarly, Singh et al. (2009) observed that the evolutionary rates of genes are likely to be accelerated in the host-specific species *D. sechellia* and *D. erecta*, which have smaller effective population sizes, compared with *D. melanogaster* and *D. yakuba*. These studies tend to suggest that the genes of species with small population sizes evolve fast, possibly due to less effective natural

selection. Petit and Barbadilla (2009) examined the effective population sizes of many of the *Drosophila* species used in the present study and reported that selection efficiency is correlated with effective population size, which, in turn, is correlated with levels of genomic codon bias, proportion of adaptive substitutions, and repetitive sequences. We therefore examined the relationship between effective population size and climatic envelope using synonymous polymorphism in the genes of the seven *Drosophila* species (*D. melanogaster*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. mojavensis*, and *D. virilis*) reported in Petit and Barbadilla (2009). Our results showed that there was no significant correlation between effective population size and climatic envelope ( $R^2 = 0.061$ ,  $P = 0.63$ ; fig. S9A, [Supplementary Material](#) online). In addition, the results of a linear model, in which effective population size and climatic envelope were used as explanatory variables, showed that  $P_D$  was explained by climatic envelope ( $R^2 = 0.85$ ,  $P = 0.0089$ ) but not by the effective population size. Further, we examined the relationship between  $P_D$  and climatic envelope after removing species with a small effective population size (*D. sechellia* and *D. erecta*), because the correlation between selection efficiency and population size was strong when the host-specific species were compared with generalist species (Petit and Barbadilla 2009; Singh et al. 2009). However, we still found a significant correlation between  $P_D$  and climatic envelope ( $R^2 = 0.86$ ,  $P = 0.00086$ ; fig. S9B, [Supplementary Material](#) online), suggesting that differences in  $P_D$  among species are not explained by the effective population sizes.

### Evolutionary Processes in Divergence of Duplicate Content

We next investigated the evolutionary processes responsible for differences in the  $P_D$  between closely related species with different habitat variability. First, we examined the conservation of  $P_D$  by fitting it to the phylogenetic tree using a Brownian motion model and calculating Pagel's lambda (Pagel 1999). We found that lambda was  $1.2 \times 10^{-7}$ , and that the value differed significantly from 1 under Brownian motion evolution by comparison of likelihoods ( $P = 3.3 \times 10^{-6}$ ). This indicates that the phylogeny does not explain the distribution of  $P_D$  among the *Drosophila* species. We therefore examined whether the loss of duplicated genes occurred more frequently in species with low  $P_D$ . We focused on two species pairs, *D. melanogaster*–*D. sechellia* and *D. pseudoobscura*–*D. persimilis*, in which the habitat variability and  $P_D$  differed even though they were closely related phylogenetically (figs. 2A and S8A, [Supplementary Material](#) online) and found that species with low habitat variability and low  $P_D$  (*D. sechellia* and *D. persimilis*) tended to lose duplicated genes (figs. S4A and S4B, [Supplementary Material](#) online, and [table 1](#)). *D. pseudoobscura* and *D. persimilis* diverged more recently than did *D. melanogaster* and *D. sechellia*, and, in fact, the former species pair can easily interbreed. Even though divergence times were different between species pairs, we observed consistent results in which species with low habitat variability tended to lose duplicated genes, when





**Fig. 3.** Correlation between climatic envelope (or  $P_D$ ) and the  $P_D$  of lost genes. (A) Relationship between climatic envelope and the  $P_D$  of lost genes. The x-axis indicates the proportion of lost duplicated genes for 11 *Drosophila* species (mel, *D. melanogaster*; sec, *D. sechellia*; yak, *D. yakuba*; ere, *D. erecta*; ana, *D. ananassae*; pse, *D. pseudoobscura*; per, *D. persimilis*; wil, *D. willistoni*; moj, *D. mojavensis*; vir, *D. virilis*; and gri, *D. grimshawi*). The y-axis indicates climatic envelope estimated by WORLDCLIM datasets. Error bars indicate standard error for the  $P_D$  of lost genes derived from table 2. (B) Relationship between contrasts in climatic envelope and the average proportion of lost duplicated genes for each species. The x-axis indicates PICs in the  $P_D$  of lost genes, and the y-axis indicates PICs in the climatic envelope. The dashed line represents the regression line. (C) The  $P_D$  of total genes for each species (black) and the average proportion of lost duplicated genes for each branch (gray) on phylogenetic tree. The phylogenetic tree is from *Drosophila* 12 Genomes Consortium (2007). (D) Relationship between contrasts in the  $P_D$  of total genes and the average proportion of lost duplicated genes for each species. The x-axis indicates PICs in the  $P_D$  of lost genes, and the y-axis indicates PICs in the  $P_D$  of total genes. The dashed line represents the regression line.

we expanded the estimation for all species used in our study (table 2). In addition, we found that there was a strong negative correlation between the loss rates of duplicated genes and climatic envelope among the species, even after phylogenetic constraints were removed ( $R^2 = 0.92$ ,  $P = 1.1 \times 10^{-5}$ ; fig. 3). Furthermore, we found a significant negative correlation between  $P_D$  and the loss rates of duplicated genes ( $R^2 = 0.73$ ,  $P = 0.0017$ ; fig. 3C and D). Note that the extreme contrasts in figure 3B and D did not drive the relationships (fig. S7B, Supplementary Material online;  $R^2 = 0.78$ ,  $P = 0.0038$  and Fig. S7C;  $R^2 = 0.45$ ,  $P = 0.068$ ). The negative correlation in figure S7C, Supplementary Material online, after removal of the extreme contrasts was not statistically significant, but it is possible due to the low statistical power of the small dataset. This indicates that the functional constraints on duplicated genes of species with low habitat variability are more relaxed

than those of species with high habitat variability. *D. sechellia* is thought to have lost *Or* and *Gr* genes associated with odor response in compensation for specializing on *Morinda citrifolia*, which is toxic to other *Drosophila* (McBride 2007; McBride et al. 2007). This is likely to be a case of antagonistic pleiotropy (trade-offs) (Hoffmann 2010). McBride reported that not only *Or/Gr* genes but also randomly chosen genes in *D. sechellia* were fast-evolving compared with those in the closely related cosmopolitan species *D. simulans* (McBride 2007). Our findings derived from genome-wide analyses suggest that DNA decay occurred in climatic specialists rather than generalists (Hoffmann and Willi 2008; Hoffmann 2010), although it is difficult to distinguish the hypothesis from that of antagonistic pleiotropy (Hoffmann 2010). However, we suggest that species with low habitat variability might have lost the functional constraints on genes in general. We



conducted a relative rate test (Tajima 1993) to detect lineage-specific fast-evolving genes to understand the general trends in differences in the functional constraints on genes between closely related species with different habitat ranges. The number of fast-evolving genes in species with high habitat variability and high  $P_D$  (*D. melanogaster* and *D. pseudoobscura*) was significantly smaller than that in species with low habitat variability and low  $P_D$  (*D. sechellia*,  $P < 2.2 \times 10^{-16}$ ,  $\chi^2$  test; and *D. persimilis*,  $P < 2.2 \times 10^{-16}$ ,  $\chi^2$  test; fig. S4C, **Supplementary Material** online, and **table 1**). Furthermore, duplicated gene pairs including significantly fast-evolving genes were enriched in species with low habitat variability and low  $P_D$  (*D. sechellia*,  $P < 2.2 \times 10^{-16}$ ,  $\chi^2$  test; and *D. persimilis*:  $P < 2.2 \times 10^{-16}$ ,  $\chi^2$  test; fig. S4D, **Supplementary Material** online, and **table 1**). These results imply that the low  $P_D$  in species with low habitat variability was caused by losses both of duplicated genes and of sequence similarity between duplicated gene pairs.

We examined whether lineage-specific lost and fast-evolving genes in species with low  $P_D$  were enriched in particular functional categories using gene ontology (<http://www.geneontology.org>). As a result, we detected little enrichment of functional categories for the genes in species with low  $P_D$ ; the lost genes in *D. sechellia* and *D. persimilis* were enriched only in metabolic process ( $P = 1.5 \times 10^{-7}$ ) and response to stimulus ( $P = 9.9 \times 10^{-4}$ ), respectively. Note that the enrichment of lost genes related to metabolic process in *D. sechellia* could be caused by a trade-off associated with specializing on the fruits of *M. citrifolia*, which contain substrates toxic to other *Drosophila* species (Markow and O'Grady 2007). This indicates that both the loss and relaxation of functional constraints are common for genes in species with low habitat variability, rather than being specific to particular genes. Species would need not only cold and desiccation tolerances but also physiological, morphological, behavioral, and certain other adaptations to live in heterogeneous environments.

We also conducted the analyses using the closely related species *D. yakuba* and *D. erecta*; although *D. yakuba* has a wider distribution in Africa than the specialist *D. erecta* (Markow and O'Grady 2006), both species inhabit tropical regions and have similar  $P_D$ . We found no significant difference in the  $P_D$  of lineage-specific lost genes between these species, which was consistent with our hypothesis (**table 1**). However, the duplicated gene pairs containing significantly fast-evolving genes and the fast-evolving genes themselves were both enriched in *D. erecta* ( $P < 2.2 \times 10^{-16}$ ,  $\chi^2$  test; **table 1**). Notably, this difference was smaller than that for species pairs with different habitat variability and  $P_D$  (*D. melanogaster*–*D. sechellia* and *D. pseudoobscura*–*D. persimilis*; **table 1**). Although both *D. yakuba* and *D. erecta* have low habitat variability, these results might also be affected by the host specificity, narrow habitat area and/or small population size of *D. erecta*. Overall, our results suggest that in species with low habitat variability, duplicated genes have been lost from the genome, whereas in species with high habitat variability, high  $P_D$  has been maintained in the genome.

## Cause and Effect of Habitat Variability

Adaptation to homogeneous environments (e.g., host specialization) is probably the main cause of habitat range restriction, because our results show that species with low habitat variability and low  $P_D$  tend to lose duplicated genes (**fig. 3**, **tables 1** and **2**). In addition, we also showed that there is no evidence to suggest that species with high habitat variability have gained a greater number of duplicated genes than those with low habitat variability. This indicates that habitat variability cannot be the cause of increasing  $P_D$ . We propose that selection for retaining genetic diversity operated efficiently in species with high habitat variability. Under this selection, duplicated genes in species with high habitat variability were maintained, in contrast to those species with low habitat variability. Therefore, the loss of duplicated genes could be a reason for restricting habitat expansion to habitats with lower variability after species have adapted to homogeneous environments and lost the genes. Compared with more generalist species, host-specific species (*D. sechellia*, *D. erecta*, and *D. mojavensis*) and island endemic species (*D. sechellia* and *D. grimshawi*) are unable to expand their distributions to heterogeneous environments due to a lack of genetic variation conferred by retention of duplicated genes. Therefore,  $P_D$  can be both a cause and an effect of habitat variability in *Drosophila* species (**fig. 3**, **tables 1** and **2**).

## Conclusion

Our findings show that the  $P_D$  in a genome strongly correlates with the habitat variability of a species. Variable environments within a species' range must promote the maintenance of duplicated genes. A recent study predicted that duplicated genes could be maintained in gene regulatory networks in randomly fluctuating environments (Tsuda and Kawata 2010). The expression of duplicated genes was more diverse than that of singletons (Kliebenstein 2008; Ha et al. 2009; Dong et al. 2011), and therefore, individuals with more duplicated genes have advantages in diverse environments because they produce more genetically variable offspring. Kellermann et al. (2009) showed that specialist species lacked genetic variation in key traits, thereby limiting their ability to adapt to changed conditions. Our results indicate that genetic and genomic architecture, such as the  $P_D$  in a genome, are fundamental constraints on the production of genetic variation for adaptation to new and varied environments.

Many of the whole genome sequences in the database were determined from inbred individuals. Therefore, these sequences do not provide information about the genetic variation of the population. Although species have gene copy number variations in their genomes, it is highly unlikely that inbreeding or the founder effect immediately reduces their  $P_D$ . Therefore,  $P_D$  can be estimated even from the genomic sequences of inbred lines as a representative value of an individual of a population. We suggest that  $P_D$  is an excellent genetic indicator for adaptation to habitat diversity. Whole genomes can now be sequenced comparatively easily, and techniques continue to rapidly improve (Metzker 2010).

Further analyses of duplicated genes in additional species will clarify the relationship between genetic factors and habitat distributions that depend on habitat variability. If the relationship between  $P_D$  and habitat variability applies to other organisms, it allows us to predict which species are unlikely to survive to environmental change, which could aid future biodiversity conservation efforts. This study shows the first evidence that genome-wide duplicated gene content determines ecological traits. Our results provide new insight into the evolution of duplicated genes, that their maintenance might confer an ecological advantage to an organism during evolution.

## Supplementary Material

Supplementary figures S1–S9 and tables S1–S5 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

## Acknowledgments

We thank A.A. Hoffman, J. Kitano, and J. Bridle for comments on the manuscript. The study was supported by the Global Centres of Excellence Program “Centre for ecosystem management adapting to global change” (J03) of the Ministry of Education, Culture, Sports, Science, and Technology of Japan to T.M. and M.K.

## References

- Ashburner M, Thompson J, Carson HL. 1982. The genetics and biology of *Drosophila*, Vol. 3b. San Diego (CA): Academic Press.
- Bosco G, Campbell P, Leiva-Neto JT, Markow TA. 2007. Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. *Genetics* 177:1277–1290.
- Bridle JR, Vines TH. 2007. Limits to evolution at range margins: when and why does adaptation fail? *Trends Ecol Evol*. 22:140–147.
- Clark AG, Eisen MB, Smith DR, et al. (417 co-authors). 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450: 203–218.
- Dean EJ, Davis JC, Davis RW, Petrov DA. 2008. Pervasive and persistent redundancy among duplicated genes in yeast. *PLoS Genet*. 4: e1000113.
- Dong D, Yuan Z, Zhang Z. 2011. Evidences for increased expression variation of duplicate genes in budding yeast: from cis- to trans-regulation effects. *Nucleic Acids Res*. 39:837–847.
- Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat*. 125:1–15.
- Ha M, Kim ED, Chen ZJ. 2009. Duplicate genes increase expression diversity in closely related species and allopolyploids. *Proc Natl Acad Sci U S A*. 106:2295–2300.
- Hartman JL 4th, Garvik B, Hartwell L. 2001. Principles for the buffering of genetic variation. *Science* 291:1001–1004.
- Heger A, Ponting CP. 2007. Evolutionary rate analyses of orthologs and paralogs from 12 *Drosophila* genomes. *Genome Res*. 17: 1837–1849.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol*. 25:1965–1978.
- Hoffmann AA. 2010. A genetic perspective on insect climate specialists. *Austr J Entomol*. 49:93–103.
- Hoffmann AA, Willi Y. 2008. Detecting genetic responses to environmental change. *Nat Rev Genet*. 9:421–432.
- Kellermann V, van Heerwaarden B, Sgro CM, Hoffmann AA. 2009. Fundamental evolutionary limits in ecological traits drive *Drosophila* species distributions. *Science* 325:1244–1246.
- Kirkpatrick M, Barton NH. 1997. Evolution of a species' range. *Am Nat*. 150:1–23.
- Kliebenstein DJ. 2008. A role for gene duplication and natural variation of gene expression in the evolution of metabolism. *PLoS One*. 3: e1838.
- Kliman RM, Andolfatto P, Coyne JA, Depaulis F, Kreitman M, Berry AJ, McCarter J, Wakeley J, Hey J. 2000. The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* 156:1913–1931.
- Kottek M, Grieser J, Beck C, Rudolf B, Rubel F. 2006. World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift* 15:259–263.
- Legendre P, Legendre L. 1998. Numerical ecology, 2nd English ed. Amsterdam: Elsevier.
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
- Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* 302:1401–1404.
- Lynch M, Sung W, Morris K, et al. (11 co-authors). 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci U S A*. 105:9272–9277.
- Margalef DR. 1958. Information theory in ecology. *Gen Syst*. 3:36–71.
- Markow T, O'Grady P. 2006. *Drosophila*: A guide to species identification and use. San Diego (CA): Academic Press.
- Markow TA, O'Grady PM. 2007. *Drosophila* biology in the genomic age. *Genetics* 177:1269–1276.
- McBride CS. 2007. Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. *Proc Natl Acad Sci U S A*. 104:4996–5001.
- McBride CS, Arguello JR, O'Meara BC. 2007. Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* 177: 1395–1416.
- Metzker ML. 2010. Sequencing technologies—the next generation. *Nat Rev Genet*. 11:31–46.
- Mirol PM, Routtu J, Hoikkala A, Butlin RK. 2008. Signals of demographic expansion in *Drosophila virilis*. *BMC Evol Biol*. 8:59.
- Nardon C, Decelie G, Loevenbruck C, Weiss M, Vieira C, Biemont C. 2005. Is genome size influenced by colonization of new environments in dipteran species? *Mol Ecol*. 14:869–878.
- Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol*. 302: 205–217.
- Ohno S. 1970. Evolution by gene duplication. New York: Springer.
- Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877–884.
- Petit N, Barbadilla A. 2009. Selection efficiency and effective population size in *Drosophila* species. *J Evol Biol*. 22:515–526.
- Piano F, Craddock EM, Kambysellis MP. 1997. Phylogeny of the island populations of the Hawaiian *Drosophila grimshawi* complex: evidence from combined data. *Mol Phylogenet Evol*. 7: 173–184.

- Reed LK, Markow TA. 2004. Early events in speciation: polymorphism for hybrid male sterility in *Drosophila*. *Proc Natl Acad Sci U S A*. 101: 9009–9012.
- Root TL, Price JT, Hall KR, Schneider SH, Rosenzweig C, Pounds JA. 2003. Fingerprints of global warming on wild animals and plants. *Nature* 421:57–60.
- Roy K, Hunt G, Jablonski D, Krug AZ, Valentine JW. 2009. A macroevolutionary perspective on species range limits. *Proc Biol Sci*. 276: 1485–1493.
- Singh ND, Larracuenta AM, Sackton TB, Clark AG. 2009. Comparative genomics on the *Drosophila* phylogenetic tree. *Annu Rev Ecol Evol Systemat*. 40:459–480.
- Tajima F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* 135:599–607.
- Tsuda ME, Kawata M. 2010. Evolution of gene regulatory networks by fluctuating selection and intrinsic constraints. *PLoS Comput Biol*. 6: e1000873.
- Wilkins AS. 1997. Canalization: a molecular genetic perspective. *Bioessays* 19:257–262.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 13:555–556.
- Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol*. 17:32–43.