The Ecoresponsive Genome of Daphnia pulex

John K. Colbourne,1,2 Michael E. Pfrender,2† Donald Gilbert,1,2 W. Kelley Thomas,4 Abraham Tucker,4,14 Todd H. Oakley,5 Shinichi Tokishita,6 Andrea Aerts,7 Georg J. Arnold,8 Malay Kumar Basu,2† Darren J. Bauer,4 Carla E. Cáceres,10 Liran Carmel,§ Claudia Casola,3 Jeong-Hyeon Choi,1 John C. Detter,2 Qunfeng Dong,3∗ Sergei Dushkevich,2 Brian D. Eads,1,3 Thomas Fröhlich,9 Kerry A. Geiler-Samorì,3 Daniel Gerlach,11,12# Phil Hatcher,4 Sanjuro Jogdeo,16 Jeroen Krijgsman,12,1† Evgenia V. Krivtsova,11 Dietmar Kültz,13 Christian Laforsch,14 Erika Lindquist,7 Jacqueline Lopez,7 J. Robert Manak,15,16† Jean Muller,16,16§§ Jasmyntang,7 Rupali P. Patwardhan,15,15 Samuel Pitluck,2 Ellen J. Pritham,17 Andreas Rechtsteiner,2,18 Mina Rho,18 Igor B. Rogozin,9 Onur Sakarya,5,19 Afsal Salamov,7 Sarah Schack,3,17 Harris Shapiro,7 Yasufumi Shiga,10 Courtney Skalitzky,15 Zachary Smith,7 Alexander Souvorov,9 Way Sun,5 Zuoqiang Tan,5,6†† Daisy Tsuchiya,1 Hank Tu,9†† Harmjan Vos,12,12‡‡ Mei Wang,5 Yuri I. Wolf,7 Hideo Yamagata,6 Takaju Yamada,16 Yuzhen Ye,28 Joseph R. Shaw,5,19 Justen Andrews,1,3 Teresa J. Crease,20 Haixu Tang,1,18 Susan M. Lucas,5 Hugh M. Robertson,21 Peer Bork,26 Eugene V. Koonin,5 Evgeny M. Zdobnov,13,22 Igor V. Grigoriev,7 Michael Lynch,3 Jeffrey L. Boore,23,24

We describe the draft genome of the microcrustacean Daphnia pulex, which is only 200 megabases and contains at least 30,907 genes. The high gene count is a consequence of an elevated rate of gene duplication resulting in tandem gene clusters. More than a third of Daphnia’s genes have no detectable homologs in any other available proteome, and the most amplified gene families are specific to the Daphnia lineage. The coexpansion of gene families interacting within metabolic pathways suggests that the maintenance of duplicated genes is not random, and the analysis of gene expression under different environmental conditions reveals that numerous paralogs acquire divergent expression patterns soon after duplication. Daphnia-specific genes, including many additional loci within sequenced regions that are otherwise devoid of annotations, are the most responsive genes to ecological challenges.

Daphnia pulex, or the water flea, is a keystone species of freshwater ecosystems: a principal grazer of algae, a primary forage for fish (1), and a sentinel of lentic (still water) inland ecosystems. Their populations are defined by the boundaries of ponds and lakes, are sensitive to modern toxicants in the environment, and thus are used to assess the ecological impact of environmental change (2, 3). Daphnia exhibit a range of context-dependent development of specialized phenotypes, such as switching between clonal and sexual reproduction in response to diverse habitats (2, 3), and as a model for ecological and evolutionary research (4). Daphnia pulex is a crustacean arthropod, the group most closely allied with the insects (9), and thus allows the cataloging of genes that likely evolved in the pancrustacean ancestor of at least some lineages of insects and Crustacea (fig. S1). Although the branchiopod D. pulex represents only a single crustacean lineage—which contains more than 40,000 known species with striking levels of phenotypic diversity—the genus and its order (the Cladocera) date to the Permian (4). Because Daphnia’s ecology is superbly understood, access to its genome sequence (fig. S2 and table S1) allows studying environmental influences on gene functions in ways that are difficult in even the best-developed genomic model species. Traits observed in laboratories are likely a small subset of the phenotypic variation that is expressed in natural ecosystems, and a focus on laboratory studies may partly explain why over 50% of many eukaryotic genomes are without experimentally determined functional annotations (11).

Genome sequence, assembly, and mapping to chromosomes. The D. pulex genome was assembled using JAZZ (12) from 1,554,564 quality-filtered nuclear sequence reads (8.7-fold coverage) from a naturally inbred isoclonal daphnid dubbed “The Chosen One” (TCO) (supporting online material [SOM] 1.1). The version 1.1 draft genome assembly comprises 19,008 contigs arranged within 5191 scaffolds that sum to a genome size of ~200 Mb (table S2). Two-hundred-eighty scaffolds link to construct 118 super-scaffolds (tables S3 and S4). Microsatellite markers (13) place 73 large scaffolds (73.9 Mb total) on the 12 chromosomes (table S5). We estimate that the draft assembly is high quality and includes ~80% of Daphnia’s nuclear genome (SOM 1.2, tables S6 and S7, and figs. S3 to S5). We determine that 3598 missing regions (59%) contain duplicated genes, whereas others are heterochromatic regions, including the centromeres and telomeres. We estimate that 25% of the genome may be heterochromatic (table S8 and fig. S6). The ends of D. pulex chromosomes appear to consist of long stretches of TTAGG repeats with flanking regions (30 to 40 Kb) internal to these repeats consisting of repetitive sequences, including at least two kinds of satellite sequences (SOM 1).

Gene inventory. A minimum set of 30,907 protein-encoding genes was predicted for D. pulex, with 26,867 gene models having the following support (tables S9 to S14 and fig. S7): (i) 145,578 expressed sequence tags (ESTs) from 37 separate conditions validating 10,578 genes; (ii) whole-genome tiling microarrays examining gene expression under six different conditions that detect 186,269 transcriptionally active regions (TARS) validating 57,294 exons from 14,135 genes (additional TARS suggest gene models not yet included within the minimum set); (iii) similarity to proteins from other (non-daphnid) genomes that detect 19,641 D. pulex genes (blast e < 10−5); (iv) 18,765 genes identified in protein similarity searches against a preliminary draft genome sequence for D. magna (SOM 2), which belongs to a separate subgenus (7); (v) more than 11,000 D. pulex peptide sequences detected by tandem mass spectrometry, of which 93% map to 1273 gene models in the minimum set; (vi) 716 conserved single-copy eukaryotic genes, of which D. pulex is missing only two (table S15), confirming that expected genes are included in the assembly; and (vii) 13,105 loci identified as paralogs by nucleotide sequence similarity searches for each predicted gene against the complete gene list (e < 10−20). Measures of the relative rate of nonsynonymous nucleotide substitutions to the substitution rate at synonymous sites (Ks/Ka) indicate that the paralogs within our gene set generally show evidence of purifying selection (fig. S8).

To ensure that the gene count was not inflated by the erroneous assembly of alleles of the same locus as unique gene copies, we conducted comparative genomic hybridizations of labeled TCO DNA on microarrays. We detected no correlation between the read coverage and the mean fluorescence units of probes representing genes (fig. S9). Counts can also be inflated by inclusion of pseudogenes. However, manual annotations suggest that pseudogenes account for only 4 to 6% of large paralogous family members in Daphnia (see companion studies (14)).

Many non–protein-encoding genes were also identified in the D. pulex genome (SOM 3). Fifty microRNA genes are annotated, and 27 are validated using tiling microarrays (table S16 and fig. S10).
We estimate that 78% of these intron gains are unique to this lineage and that 22% occurred in parallel with gains in other lineages (fig. S16).

**Origin and preservation of Daphnia pulex genes.** Daphnia’s gene catalog shows more universal bilaterian genes than other arthropods (8096; black in Fig. 1A) and thus shares the highest number of genes with human (table S25). Only 1383 genes (4.5%) appear pancrustacean (green in Fig. 1A). Remarkably, over 36% of the minimal set of *D. pulex* genes have no detectable homology to those in the other species (Fig. 1A), which can partly be explained by the disproportionate expansion of gene families distinctive to this crustacean lineage ($\chi^2 = 450.55, P < 0.0001$) (table S26 and Fig. 1B) and fast divergence for some genes (enlarged beige fraction in Fig. 1A). A phylogenetic accounting of the expansions and contractions of all gene families within pancrustacean and representative deuterostome genomes (tables S27 and S28) suggests a net increase in the number of paralogs within the lineage leading to *Daphnia* (Fig. 1C). By reconstructing gene family histories across a phylogeny (SOM IV2), we count 17,424 new and 1079 lost genes in the branch leading to *Daphnia*. By contrast, the sum of inferred gains and losses along the longest series of branches in the insect phylogeny—originating from the shared pancrustacean ancestor with *Daphnia*—only reaches 8981 gained loci with 3040 gene losses. Therefore, the overall elevated *Daphnia* gene count appears to result from both gaining and retaining more genes.

To better understand gene duplication in the *Daphnia* genome, we examined the age distribution of gene duplicates by estimating $K_s$ for 66,502 pair-wise combinations of paralogs showing >40% sequence similarity and by comparing this distribution to that of 12,570 nematode and 64,783 human gene pairs (Fig. 1D). The single-pair duplicates within the youngest cohort ($K_s < 0.01$) suggest that *D. pulex* genes duplicate at a rate three times as high as those measured for fly and nematode and 30% greater than human, even when we exclude nearly identical gene copies that may be biased by gene conversion (table S29 and figs. S17 and S18).

In the genomes of many species, new duplicate genes are found in clusters (fig. S19) (15). The *D. pulex* genome shows ~20% (table S30) of all genes tightly arranged in clusters of 3 to 80 paralogs and with elevated numbers of tandemly duplicated genes at intervening intervals of 1000 to 2000 bp (fig. S20). The age distribution and positioning of gene duplicates indicate that *Daphnia* has not experienced whole-genome duplication, but the genome is instead characterized by a high and historically steady rate of tandem duplication (Fig. 1D).

Nine gene families have expanded independently in *Daphnia* and other aquatic lineages, including vertebrates (tables S26 and S31). These include photoreactive or photore sponsive gene families (cryptochromes, opsins, and G proteins). The *D. pulex* genome shows 46 opsins (table S32 and figs. S21 and S22), of which 42 derive from two rhodopsin subfamilies, one ciliar ops in subfamily, and a newly discovered lineage that we have named arthropsins (SOM 4). Arthropsins are ancestral to the chordate melanosin lineage and thus appear to have been retained in *Daphnia*, despite their loss from all other available bilateral animal genomes. The expansion of these gene families suggests that adaptations to a more compact genome.

1Department of Biology II and GeoBio Center Munich, Ludwig-Maximilians-University Munich, Großenbergerstrasse 2, 82152 Planegg-Martinsried, Germany. 2Gene Expression, Roche NimbleGen Inc., 504 South Rosa Rd, Madison WI 53719, USA. 3Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL) Heidelberg, Meyerhoffstrasse 1, D-69117 Heidelberg, Germany. 4Department of Biology, University of Texas at Austin, 1 University Station, Austin, TX 78712, USA. 5Laboratory for Functional Genome Analysis (LAFUGA), University of California, Santa Cruz, CA 95064, USA. 6School of Informatics and Computer Science, New York University Langone Medical Center, New York, NY 10016, USA. 7Department of Molecular Cancer Research, University Medical Center Utrecht, Utrecht University, Utrecht, 3584 CG Utrecht, Netherlands.
plex light regime in aquatic environments (16, 17) can be influential in shaping the gene content of these organisms.

Implications of *Daphnia’s* genome structure. Tandemly duplicated gene clusters are predisposed to homogenization by gene conversion and unequal crossing-over (18). If common, concerted evolution can maintain sequence and functional similarities among paralogs. We examined copied DNA segments among all paralogs in the genome (SOM V.1) and observed that 47% of the genes show tracts of nonallelic gene conversion compared with 12 to 18% of genes in five *Drosophila* species (tables S33 to S38 and figs. S23 and S24).Thus, concerted evolution is affecting more than 1 Mb (8%) of all protein-coding sequences in *Daphnia*, especially when duplicates are oriented on the same strand, with a similar conversion rate (converted pairs of paralogs/total pairs of paralogs analyzed) and number of events per pair as *Drosophila*. The greater proportion of converted genes in *D. pulex* is mainly attributed to the greater number of targets for gene conversion within the genome, including tandemly duplicated gene clusters with intervening genes. Conversion events in *Daphnia* are less common among the youngest duplicates and within gene families containing only two paralogs.

One example of widespread gene conversion is found in the di-domain hemoglobin genes. Hemoglobin levels in the hemolymph of daphniids can rise by more than one order of magnitude in response to reduced oxygen availability in aquatic habitats, which fluctuates in diurnal and seasonal cycles (Fig. 2A). In *Daphnia*, a tandemly duplicated gene cluster of hemoglobin (Hb) genes contributes to the protein’s varying composition (19). We sequenced and assembled the full *D. magna* cluster to compare with the arrangement of eight clustered *D. pulex* hemoglobin genes (figs. S25 and S27). (*D. pulex* also has three nonclustered Hb genes.) Notably, the two species show almost identical gene arrangements within an interval of ~23.5 Kb (table S39) except for the obvious absence of Hb6 from the *D. magna* cluster (Fig. 2B). In both species, a noncoding RNA gene interrupts the cluster between Hb4 and Hb5, and hypoxia response elements plus ancillary sequences are preserved within the regulatory regions of each gene. Thus, the duplication and subsequent divergence of hemoglobins must have occurred before the divergence time of *D. pulex* and *D. magna*.

However, a phylogenetic analysis of protein-coding sequences (SOM V.2) suggests that most hemoglobin genes have duplicated independently within each species (Fig. 2C). A separate phylogenetic reconstruction using sequences from intergenic regions recovers a tree that is consistent with duplication before speciation (Fig. 2D). Because the support values at nodes for both trees are equally strong, we conclude that gene conversion tracts are homogenizing the protein coding regions. The hemoglobin gene clusters in both species are homologs because of ancestral gene duplications, yet the duplication history of

![Fig. 1. The *Daphnia pulex* gene repertoire. (A) Comparison of genes among *D. pulex*, *Drosophila melanogaster*, *Pediculus humanus*, *Tribolium castaneum*, and *Strongylocentrotus purpuratus* (urchin), and *Galbus gallus*, *Xenopus tropicalis*, and *Homo sapiens*, showing the core bilaterian genes (black), vertebrate (blue), insect (aqua), and pancrustacean (green) specific genes, patchy or ancient orthologs present in at least one arthropod and one deuterostome genome but lost in other lineages (pink), multiple copy homologs (yellow and beige), and species-specific genes (white). (B) Distribution of *D. pulex* gene family sizes comparing genes with and without detectable homology to other genomes. (C) History of gene family expansions and losses among pancrustacean plus representative deuterostome genomes with the outgroup *Nematostella vectensis*. Tree topology is fixed from the assumed species phylogeny and used to map gene family histories by a combination of gene similarity and character-state optimization with Dollo parsimony (SOM IV.2). Branch lengths scaled to differences between inferred gene gains and losses. Scale bar corresponds to 1000 genes gained. Gene gains along each branch of the tree are scaled by the maximum value along the branch leading to *D. pulex* (blue); gene losses along each branch are scaled by the maximum loss along the branch leading to *Caenorhabditis elegans* (yellow). (D) Frequency of pair-wise genetic divergence at silent sites (Ks) among all gene duplicates in the *D. pulex*, *C. elegans*, and *H. sapiens* genomes, for genes with >100 aligned amino acids and percent identity >40% (66,502, 12,570, and 64,783 pair-wise comparisons for the three genomes, respectively). The vertical axis differs for each tree.

www.sciencemag.org  SCIENCE VOL 331  4 FEBRUARY 2011  557
genes is obfuscated by independent gene conversions facilitated by their ordered arrangement in the genomes.

**Evolutionary diversification of duplicated genes.** Gene duplication is an important source of evolutionary novelty. After duplication, one copy is commonly disabled by mutation and becomes a pseudogene. This fate may be avoided if selection maintains both copies through gene dosage, novel function, or by subdividing the gene’s original function into multiple components (20). We conducted microarray experiments to determine the magnitude of functional divergence among paralogs, then traced (21) and tested (22) whether their patterns of gene transcription differ in 1 to 12 ecologically relevant conditions as a function of $K_s$ (table S40 and SOM VI.1). As expected, many recent duplicates ($K_s < 0.05$) have indistinguishable gene expression patterns for the tested conditions (47%) (Fig. 3A). Within many gene families, divergence in expression patterns correlates with age (figs. S28 and S29). We found that long-wavelength opsins most similar in sequence have the same expression patterns (correlation > 0.9) but then diverge in their response to shared conditions as they age, at an estimated rate of 0.6% per 10% synonymous nucleotide substitutions. A similar pattern is observed for the di-domain hemoglobins, albeit with more rapid divergence in expression.

---

**Fig. 2.** Evolution of *Daphnia* di-domain hemoglobin (Hb) genes. (A) When deprived of oxygen, many species (here *D. magna*) increase hemoglobin concentration in the hemolymph by a factor of 15 to 20 within a single molting, coloring the body red. (B) Organization of the Hb gene cluster in the *D. magna* and the *D. pulex* genomes. Black boxes are exons. Gray boxes represent an RNA gene. Vertical bars are hypoxia response elements (HREs), and asterisks show ancillary elements. Conserved HREs are linked by hatches. Open boxes represent highly similar sequences. The lengths of intergenic regions are shown in parentheses. *Daphnia pulex* genes Hb9 to Hb11 are located on separate sequence scaffolds. (C) Phylogenetic tree (SOM V.2) from nucleotide sequences of Hb genes in *D. pulex* (red) and in *D. magna* (black). Outgroup Hb cDNA sequences are from *Ascaris suum* and *Pseudoterranova decipiens*. Scale bar shows mean number of differences (0.1) per nucleotide along each branch. Posterior probability node support <100% are shown. (D) Phylogenetic tree based on nucleotide sequences of intergenic regions between the stop codons and the downstream TATA of the neighboring gene. Posterior probabilities <100% are shown.
In contrast to the steady expression divergence of many duplicates, we observed an equally large fraction of recently arisen paralogs—with nearly identical sequences—that differ in their expression in at least one condition (Fig. 3A). Although we could confidently detect locus-specific expression for only a fraction of the youngest duplicates represented on the microarray (table S40), a plot of the maximum difference in the expression response of paralogs to an identical condition suggests that, on average, newly duplicated genes may differ in expression by as much as a factor of 1.9 (Fig. 3B). These may be cases in which new regulatory programs were created by the gene duplication itself through a failure to copy regulatory elements or when a duplicate is integrated within a new genomic location (23).

Gene conversion, homogenizing nonregulatory nucleotide sequences, can contribute to this class of highly differentially expressed (DE) paralogs at low sequence divergence ($K_s$). We tested whether gene conversion accounts for the differences in the evolutionary rates of expression divergence by comparing duplicates ($K_s < 2$) on the basis of their structural arrangements in the genome (SOM V12). Neighboring paralogs within tandem gene clusters were just as likely to diverge in expression as dispersed duplicates outside of clusters ($\chi^2 = 0.027, P = 0.87$). Globally, gene conversion reduces the expression-level divergence of paralogs ($\chi^2 = 11.9, P = 0.0005$) (table S41), yet we detected no significant impact on the observed fractions of divergently expressed paralogs when we removed duplicated genes with signatures of gene conversion (table S42). Although adjacent genes are often coexpressed (24), the local placement of genes within tandem gene clusters has no clear effect on gene expression divergence in *D. pulex.*

We thus conclude that paralogs, even in tandem, frequently acquire divergent expression patterns at, or soon after, the time of duplication.

**Functional importance of expanded gene families.** To investigate the functional role of paralogs and their preservation, we examined interacting genes with known function. A total of 1908 genes representing 563 enzymes were charted onto the global metabolic pathway for *D. pulex* by referencing the metabolic enzyme networks of three insects and four vertebrates (Fig. 4) (SOM VII.1). Of these, 38 gene families were amplified in Pancrustacea, and 32 are expanded in the lineage leading to *Daphnia* (figs. S30 and S31 and tables S43 and S44). Half (19 of 38) of the amplified genes are nonrandomly clustered within seven distinct pathways ($P < 0.03$ by exact binomial test and $P < 0.03$ by network permutation analysis) (Fig. 4, A to G, and fig. S32). These data, showing coexpression of genes within pathways, suggest that duplicated genes can be interdependent.

A study of the expression patterns of duplicated genes from this metabolic network (SOM VII.2) reveals greater average similarity between genes from coexpressing and interacting families (same Kyoto Encyclopedia of Genes and Genomes map ID in table S43) than between genes from nonassociating families ($r = 3.30, P = 0.0025$). This pattern suggests nonindependent functional divergence of expanding genes within pathways (e.g., tables S45 to S48 and figs. S33 and S34). One example involves nine phylogroups of fucosyltransferase paralogs that share 95% amino acid similarity (colored lines in fig. S35) and have independently diversified to express seven transcriptional profiles shared with interacting glycosyltransferase paralogs. Such a pattern of codivergence suggests a decoupling of duplication history and functional association. To test this prediction, we estimated the ratio of among-group variance to total variance in differential expression ($D_{ae}$) for phylogroups of fucosyltransferase paralogs and for expression profile clusters (SOM VII.2). We detect no significant subdivision of expression patterns for fucosyltransferase paralogs based on phylogeny (blue nodes in fig. S35; $D_{ae} = 0.0042, P = 0.89$). By contrast, clusters based on transcriptional profiles, and including distinctly related paralogs and interacting glycosyl transferase paralogs, show significant subdivision ($D_{ae} = 0.0836, P = 0.002$).

**Ecoresponsive genes.** The *D. pulex* genome contains many duplicated genes with unknown homology. Although this may diminish with the availability of more crustacean genomes, these unknown genes appear to play important roles in the animal’s ecology. ESTs from 37 cDNA libraries representing transcriptomes of daphnids exposed to biotic ecological challenges, abiotic ecological stressors, and different life-history stages in laboratory environments (table S10) show that genes unique to the *Daphnia* lineage, and genes that reside within tandemly duplicated gene clusters, are significantly over-represented among transcriptomes under ecological conditions (Fig. 5A and table S49). $\chi^2 = 265.1, P = 2.66 \times 10^{-58}$ and $\chi^2 = 41.0, P = 1.23 \times 10^{-09}$, respectively. Whole-genome-tiling-expression microarray experiments show differential expression to be twice as frequent in genomic regions devoid of gene models (intergenic) when *D. pulex* are exposed to environmental challenges compared with conditions of life history (table S50 and fig. S36).

We count 34,844 transcriptionally active regions (TARs) within unannotated regions of the genome, showing predictable exon-intron intervals supporting additional gene models not yet included within the minimum set (TAR-genes) (table S12) and that are condition-dependent in their regulation. By partitioning the differentially expressed genome by experimental conditions, between 72% and 85% of the transcriptome uniquely responded to one of the three conditions (Fig. 5B). In all, 73% of differential regulation under biotic or abiotic stressors requires additional gene models or extensions.

**Evolutionary perspectives.** *Daphnia pulex* paralogs follow different evolutionary trajectories that are determined, in part, by their initial transcriptional expression patterns. At least half appear to acquire divergent expression patterns at or near the time of origin. Interacting and coexpressing genes can also appear to be codiverging in their responses to environmental conditions. These observations suggest that the persistence of this distinctive class of functionally divergent gene duplicates is due to preservation by entrainment (PBE). Entrainment is defined as the process of increasing the initial probability of preserving a duplicated gene through its functional interaction with existing or newly interacting genes sharing regulatory programs. Because biological processes can be governed by interdependent regulation of interacting genes, there are three likely evolutionary

![Fig. 3. Functional diversification of duplicated genes from 12 microarray experiments. (A) The fraction of duplicated genes with similar versus divergent DE patterns as a function of their pair-wise divergence at silent sites ($K_s$). (B) Regression ($r = 0.29$) of the maximum observed difference (treatment versus control) between duplicated genes among the 12 conditions as a function of the age of duplicated genes inferred from $K_s$. Red points are significant values ($P < 0.05$, analysis of variance). The regression line $y$-axis intercept (ln 0.642 ± 0.009) suggests that, on average, newly duplicated genes may differ in expression by as much as a factor of 1.9 at particular conditions, which is significantly different from zero ($r = 68.7, P = 2 \times e^{-16}$) and validated by tiling microarray data ($r = 0.16; t = 75.3, P < 2 \times e^{-16}$).](www.sciencemag.org)
outcomes for these interacting duplicated genes (Fig. 6). Genes with expression patterns unchanged at the time of duplication may continue to share the condition-specific regulation of existing interacting genes (Fig. 6A). In this scenario, selection for gene dosage may increase the probability that gene duplicates are preserved (25). Alternatively, duplicates may initially have divergent expression patterns but have inappropriate transcriptional responses to environmental conditions or lack appropriately coregulated interacting genes (Fig. 6B). Duplicates within this category are most likely lost. In contrast, genes with divergent expression patterns at the time of duplication, yet with regulation sufficiently similar to the expression patterns of a different interacting gene, may have combined products that are beneficial under a distinct environmental condition (Fig. 6C). In this scenario, the likelihood for preservation of these new gene duplicates is increased. Thus, when genes are advantageous at the time of duplication, their coding regions are subject to purifying selection from the start and are entrained to a distinct regulatory pattern dictated by condition-specific gene-gene interactions. Although the likelihood of converging on a beneficial gene expression profile near the time of duplication is very small, in the case of Daphnia, PBE is facilitated by the high rate of gene duplication, resulting in co-regulated interacting genes that can potentially define environment-specific transcriptomes, which may increase with the complexity of interactions between organisms and their environments.

In conclusion, by examining genome structure and the functional responses of genes to environmental conditions within species with tractable ecologies, we further our understanding of gene-environment interactions in an evolutionary context. Many responsive genes to ecological conditions have unknown function, and information from laboratory model species may be insuf-
efficient because of a lack of homology or experimentally demonstrated functions in response to the environment. Thus, ecological genomics requires empirical annotations of new genome sequences from a broader diversity of species, tested under a variety of natural conditions.

References and Notes
15. Q. Zhou et al., Genomes Res. 18, 146 (2008).
26. We thank M. Frazer (JGP, P. Cherbas (CGB), R. Green, and T. Takova (Roche NimbleGen, Inc.). The work conducted by the U.S. Department of Energy Joint Genome Institute (JGI) was supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-05CH11231 and in collaboration with the Daphnia Genomics Consortium (DGC). This project was also supported by NSF grants 0221837 and 0328516 and NIH grant R24GM07827401. Coordination infrastructure for the DGC is provided by the Center for Genomics and Bioinformatics (CGB) at Indiana University, which is supported in part by the METAcyt Initiative of Indiana University, funded in part through a major grant from the Lilly Endowment, Inc. Additional contributions and acknowledgments are provided in the SOM. Our work benefits from and contributes to the Daphnia Genomics Consortium. Daphnia pulex genome assembly version 1.1 and annotations are deposited at DNA Data Bank of Japan, European Molecular Biology Laboratory, and GenBank databases under accession ACJG00000000. ESTs (FE274839 to FE425949) are in GeneBank. Microarray platforms GPL11200 to GPL11201 and data GSE25823 are deposited at National Center for Biotechnology Information Gene Expression Omnibus database.

Supporting Online Material www.sciencemag.org/cgi/content/full/331/6017/555/DC1
Materials and Methods
SOM Text Figs. S1 to S36 Tables S1 to S50 References 14 September 2010; accepted 16 December 2010 10.1126/science.1197761

www.sciencemag.org

SCIENCE VOL 331 4 FEBRUARY 2011

561

Downloaded from www.sciencemag.org on 3 February 2011