

References and Notes

1. G. Simchen, D. Idar, Y. Kassir, *Mol. Gen. Genet.* **144**, 21 (1976).
2. M. E. Budd, K. D. Wittrup, J. E. Bailey, J. L. Campbell, *Mol. Cell. Biol.* **9**, 365 (1989).
3. D. Schild, B. Byers, *Chromosoma* **70**, 109 (1978).
4. D. Stuart, C. Wittenberg, *Genes Dev.* **12**, 2698 (1998).
5. V. Borde, A. S. H. Goldman, M. Lichten, data not shown.
6. K. N. Smith, A. Penkner, K. Ohta, F. Klein, A. Nicolas, personal communication.
7. I. Collins, C. S. Newlon, *Mol. Cell. Biol.* **14**, 3524 (1994).
8. L. H. Johnston, D. H. Williamson, A. L. Johnson, D. J. Fennell, *Exp. Cell Res.* **141**, 53 (1982).
9. F. Baudat, A. Nicolas, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5213 (1997).
10. A. E. Reynolds, R. M. McCarroll, C. S. Newlon, W. L. Fangman, *Mol. Cell. Biol.* **9**, 4488 (1989).
11. R. M. McCarroll, W. L. Fangman, *Cell* **54**, 505 (1988).
12. B. M. Ferguson, W. L. Fangman, *Cell* **68**, 333 (1992).
13. B. J. Brewer, W. L. Fangman, *Cell* **51**, 463 (1987).
14. S. Keeney, C. N. Giroux, N. Kleckner, *Cell* **88**, 375 (1997).
15. H. Tsubouchi, H. Ogawa, *Mol. Cell. Biol.* **18**, 260 (1998).
16. T. Usui *et al.*, *Cell* **95**, 705 (1998).
17. R. S. Cha, B. M. Weiner, S. Keeney, J. Dekker, N. Kleckner, *Genes Dev.* **14**, 493 (2000).
18. R. Rothstein, B. Michel, S. Gangloff, *Genes Dev.* **14**, 1 (2000).
19. F. Uhlmann, K. Nasmyth, *Curr. Biol.* **8**, 1095 (1998).
20. A. M. Miller, K. A. Nasmyth, *Nature* **312**, 247 (1984).
21. K. Shibahara, B. Stillman, *Cell* **96**, 575 (1999).
22. K. Ohta *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 646 (1998).
23. V. Rocco, A. Nicolas, *Genes Cells* **1**, 645 (1996).
24. L. Xu, N. Kleckner, *EMBO J.* **14**, 5115 (1995).
25. Supplemental data are available at www.sciencemag.org/feature/data/1052791.shl.
26. All strains are diploids of the SK1 background, homozygous for the indicated mutations. Complete genotypes, construction details, and protocols for sporulation and conventional agarose gel analyses are available on Science Online (25).
27. The *ars305 ars306 ars307* diploid contains mutations inactivating *ARS305*, *ARS306*, and *ARS307*. The *his4::TEL1* diploid contains a reciprocal translocation with breakpoints at *HIS4* on chromosome III and at *PHO11* on chromosome I. Genotypes, construction details, and protocols for pulsed-field and 2D gel analyses are available on Science Online (25).
28. I. Collins, C. S. Newlon, *Cell* **76**, 65 (1994).
29. We thank C. Newlon and D. K. Bishop for strains; T. Allers for valuable advice; K. Smith, A. Nicolas, and C. Newlon for personal communications; and D. Chat-toraj, J. Haber, G. Hayhurst, C. Newton, R. Shroff, B. Thomas, and M. Yarmolinsky for comments that improved this manuscript.

2 June 2000; accepted 10 August 2000

Genomic Analysis of Gene Expression in *C. elegans*

A. A. Hill,^{1*} C. P. Hunter,^{2*} B. T. Tsung,² G. Tucker-Kellogg,^{1†} E. L. Brown¹

Until now, genome-wide transcriptional profiling has been limited to single-cell organisms. The nematode *Caenorhabditis elegans* is a well-characterized metazoan in which the expression of all genes can be monitored by oligonucleotide arrays. We used such arrays to quantitate the expression of *C. elegans* genes throughout the development of this organism. The results provide an estimate of the number of expressed genes in the nematode, reveal relations between gene function and gene expression that can guide analysis of uncharacterized worm genes, and demonstrate a shift in expression from evolutionarily conserved genes to worm-specific genes over the course of development.

The nematode *C. elegans* is a genetically accessible model organism that is widely used to study genetics, development, and other biological processes (1, 2). In 1998, the genome of this organism was completely sequenced, and the presence of 19,099 genes, or open reading frames (ORFs), was predicted. This made it possible to use oligonucleotide arrays for genome-wide gene expression monitoring in this metazoan (3, 4). We designed three oligonucleotide arrays (denoted A, B, and C) to monitor the mRNA expression levels of 18,791 (98%) of the predicted worm ORFs; the remaining ORFs were not included because they were almost identical to one or more of the selected ORFs (5). To maximize the number of detected transcripts, we quantitated nematode gene expression in six developmental-

ly staged populations from eggs to adult worms, in isolated oocytes, and in aged worms near the end of their ~2-week lifespan (6). We anticipated that the resulting data set would provide insight into gene expression in the nematode and also serve as a baseline for further experiments in *C. elegans*.

The number of ORFs called “present” by the Affymetrix GeneChip software (7) in any readout across all RNA samples is summarized in Table 1. In total, 10,747 ORFs (56%) were detected in at least one hybridization. This number of detected genes is comparable to the ~10,000 independent genes represented by the current set of *C. elegans* expressed sequence tags (ESTs) (8). We detected most ORFs with ESTs, and most of the ORFs that we did not detect are likely transcripts in very low abundance. For example, we detected 78% of the ORFs on the A array, which had sequence homology to *C. elegans* cDNAs. In contrast, 90% of the ORFs that we failed to detect are represented by at most a single cDNA in the *C. elegans* database ACEDB (version WS6).

Our ability to detect genes as expressed is dependent on many factors, but the most important two are the sensitivity of the oli-

gonucleotide chip and the relative abundance of the transcripts. To estimate the sensitivity of the oligonucleotide arrays, we included in vitro synthesized transcripts in each hybridization (Table 1) (9). As determined by the signal response from these control transcripts, the sensitivity of detection of the arrays ranged between ~1:300,000 and 1:50,000. By way of example, we estimate that embryos contain ~10⁷ transcripts (10); thus, a sensitivity level of 1:300,000 corresponds to detecting a gene expressed at an average of 30 transcripts per embryo. To be reliably detected, transiently expressed genes would need to be expressed at higher levels. Furthermore, as the animal grows, rare transcripts, particularly those expressed in only a few or single cells, would be further diluted by ubiquitous and abundant transcripts. Indeed, preliminary data from isolated gonads indicates the presence of many gonad- and germ line-specific messages that were at low or undetectable levels in whole worms (11). Thus, many of the ~8400 ORFs that were not detected in any experiment are probably genes that are expressed at levels below our limits of detection. Consistent with this, in a separate series of experiments that monitored expression in worms grown under stressful conditions (which should induce gene expression), we detected 611 additional transcripts that were not detected during the unperturbed life cycle (11).

By applying a one-way analysis of variance (ANOVA) to our replicated measurements, we isolated the subset of transcripts that had a significant ($P < 1 \times 10^{-3}$) increase or decrease in frequency (12) at some point during the life cycle and that were called “present” at least once; 4221 (22%) of the ORFs met these criteria. The expression profile of each of these developmentally modulated genes was normalized to have a mean value of zero and a variance of one, and the normalized profiles were clustered by means of a self-organizing map (SOM) (13). Examples of selected clusters are shown in

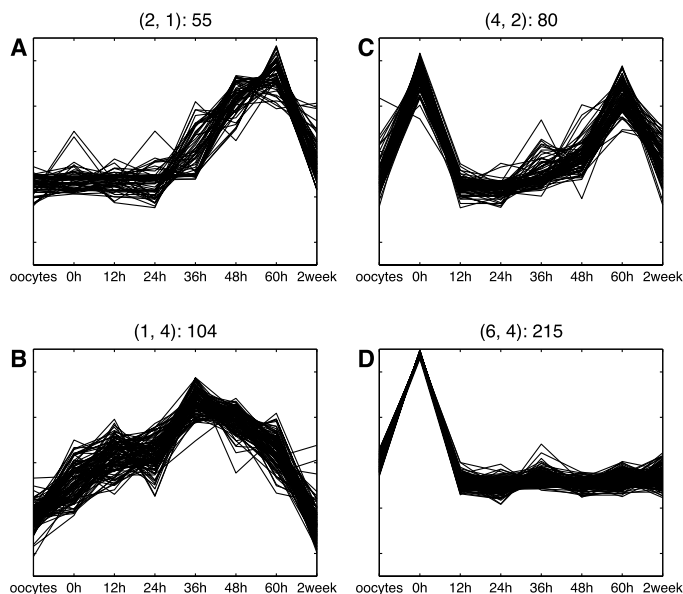
¹Department of Genomics, Genetics Institute, Cambridge, MA 02140, USA. ²Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA.

*To whom correspondence should be addressed. E-mail: ahill@genetics.com (A.A.H.) and hunter@mcb.harvard.edu (C.P.H.)

†Present address: Millennium Predictive Medicine, 700 One Kendall Square, Third Floor, Cambridge, MA 02139, USA.

REPORTS

Fig. 1. Selected SOM clusters of expression profiles for developmentally modulated genes. A SOM was used to partition a total of 4221 genes that were significantly modulated (ANOVA, $P < 1 \times 10^{-3}$) over the worm life cycle into 36 clusters with similar normalized expression profiles. Four representative clusters are shown. The heading of each panel indicates the index of the cluster in the full SOM (e.g., 2, 1), followed by the number of genes in the cluster. Because all profiles have been normalized to the same amplitude, there are no units indicated on the y axes.



Complete gene listings for each cluster are available as supplemental information (14). (A) This cluster of 55 genes, with peak expression in the egg-laying young adult, included the vitellogenin *vit-6*. (B) This cluster of 104 genes included the cuticular collagen *dpy-13*. The cluster was significantly enriched in several WormPD classes, as described in the text. (C) This cluster of 80 genes was enriched in genes related to embryogenesis, germ line development or function, and cell-cycle progression, as described in the text. (D) This cluster of 215 genes was enriched in known or putative transcription factors, as described in the text.

Table 1. Summary of the number of ORFs detected by array experiments. An ORF was scored as detected if it was called "present" (7) one or more times across all samples. The number of distinct ORFs detected in each worm sample for each array design (A, B, or C) is tabulated. A total of 10,747 distinct genes was detected. The sensitivity of arrays is indicated in ppm (e.g., 10 ppm = 1:100,000).

Developmental time	Array			Total genes ever called "present"
	A	B	C	
Oocytes	3153	1110	1726	5980
0 hours	4345	1275	1525	7133
12 hours	2552	699	1773	5016
24 hours	1956	786	1878	4612
36 hours	3547	896	2355	6787
48 hours	3475	768	2138	6373
60 hours	3964	373	2403	6732
2 weeks	579	713	719	2006
Sensitivity (9) average* (ppm)	4.1 (2 to 18)	6.3 (2 to 18)	4.4 (2 to 18)	
Nonredundant genes called "present" by array	5189	2123	3447	10,747
Genes called "present" (% of genes on array)	78	37	52	

*Numbers in parentheses indicate the range.

Fig. 1 [see supplemental information for a complete listing of all cluster and frequency data (14)]. A number of previously characterized genes had life-cycle expression profiles that matched our expectations. For example, K07H8.6 (*vit-6*) (cluster A) is a vitellogenin, known to be expressed abundantly in the intestine of the late larval/adult hermaphrodite *C. elegans* (2). F30B5.1 (*dpy-13*) (cluster B) is a cuticular collagen, expressed in waves during the four larval stages as cuticle is synthesized and in lower levels before and after the larval stages (15).

To interpret the expression profile clusters, we matched the gene classifications in the Proteome WormPD database (16) to the genes in each cluster. Examination of the gene clusters indicated several expression profiles that were readily interpretable in terms of worm biology. Cluster B contained 104 genes whose expression increased through development to a peak at 60 hours, then declined in 2-week-old worms. This cluster was enriched in markers of metabolic activity, including oxidoreductases (7-fold enrichment as compared to the genome as a

whole), amino acid metabolism genes (17-fold enrichment), carbohydrate metabolism genes (15-fold enrichment), and protein synthesis genes (7-fold enrichment). Cluster C contained 80 genes that were up-regulated in eggs and then again later in the reproductively active, egg-laying adult worm. This cluster is enriched in a small set of genes that we classified as being linked to embryogenesis, germ line development or function, and cell-cycle progression. Examples of these genes include F38E1.7 [*mom-2*, required for polarization of the EMS cell (17)], C08B11.1 [*zyg-11*, required for zygote formation (18)], and Y39A1A.12 (putative origin recognition complex subunit).

Other patterns of gene distribution among the clusters were observed, at least one of which emphasizes the influence of assay limitations on the clusters and the need to examine the raw data (19) when interpreting clustered profiles. For example, cluster D is enriched three to four times in known or putative transcription factors and in rare messages [212 out of 263 transcripts in this cluster had frequencies that never exceeded 30 parts per million (ppm); $P < 0.001$ by hypergeometric statistics]. This cluster exhibited the highest expression in the egg and exhibited low or undetectable expression at later times. However, many of these genes were expressed in the egg just slightly above our limit of detection and sank below our sensitivity at later developmental stages. Thus, the prevalence of transcription factors in this cluster is to some extent driven by the limits of our ability to detect rare messages in older animals, as opposed to the biology of specific transcription factors. This finding appears to have a parallel in reports that many messages that are found in embryonic sea urchin cDNA libraries are no longer detectable in adults (20).

Directed searches for genes that were significantly down-regulated in 2-week-old worms also revealed functionally related transcripts. These include the muscle-related genes T22E5.5 (gene, *mup-2*; protein, troponin-T; 15-fold down-regulated at 2 weeks as compared to its mean level between 0 and 60 hours), M03F4.2 (*act-4*; actin; 17-fold down-regulated), and F07A5.7 (*unc-15*; paramyosin; 13-fold down-regulated), as well as Y57G11C.12 (similar to ubiquinone oxidoreductase subunit; fivefold down-regulated) and C44B12.2 (*ost-1*; osteonectin; 22-fold down-regulated). Together, these results suggest impaired muscle function, reduced metabolic activity, and extracellular matrix defects in aged worms, consistent with the aged worm phenotype.

Recent full-genome sequence comparisons (21, 22) have revealed widespread similarities and important differences between the yeast, fly, and worm genomes. Such se-

REPORTS

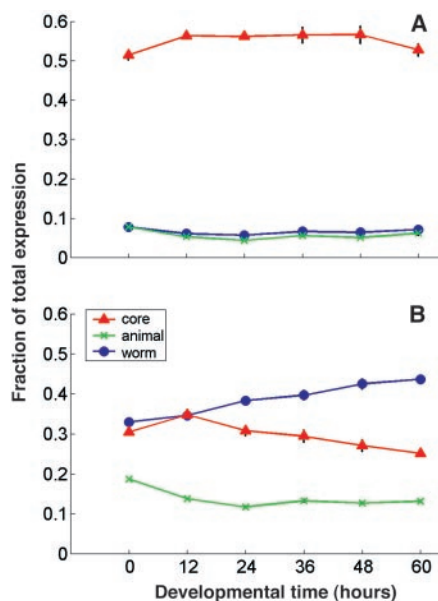


Fig. 2. (A) The fraction of total gene expression among 728 highly expressed (>30 ppm) non-modulated genes related to the inferred ancestries of the genes. These genes were defined as those that were called “present” 12 or more times and did not change significantly across the worm life cycle (ANOVA, $P > 1 \times 10^{-2}$). Lines indicate the fraction of total gene expression due to core, animal, and worm genes. Error bars are ± 1 SD, based on the variation of subsets of replicated data. Fractions do not add to one because not all genes were classified as core, animal, or worm. (B) Same plot for 4221 modulated genes, defined as those that were detected and changed significantly during the worm life cycle (ANOVA, $P < 1 \times 10^{-3}$).

quence comparisons provide us with a glimpse of the evolutionary process and are a useful source of information about the function of uncharacterized genes. However, these comparisons do not address the issue of how yeast, fly, and worm compare at the level of expressed transcripts. We formulated a simple “phylogenetic” model for gene expression in the developing worm. In this model, we partitioned nematode genes into three classes on the basis of sequence similarity: “core” genes (shared among yeast, worm, and fly), “animal” genes (shared between worm and fly), and “worm” genes (unique to the worm) (23). We hypothesized that, during development, the expression of core genes would remain relatively high and constant, reflecting a primary role of these genes in common cellular processes. In contrast, we theorized that animal and worm genes would make up a smaller fraction of the transcriptome and be highly developmentally modulated, reflecting their probable role in defining multicellular processes and worm-specific development.

In agreement with this model, we found that core genes were more likely to be detect-

Table 2. The number of core, animal, and worm genes among detected genes and genes with either high or low expression levels at all developmental time points. Core genes were more likely than animal or worm genes to be detected and were also more likely to be expressed at levels above 30 ppm. Percentages in parentheses are fractions of the total number of genes in each class (core, animal, or worm). *f*, frequency.

Genes	Detected	Never detected	Minimum expression ($f > 30$ ppm)	Maximum expression ($f < 30$ ppm)
Core	2329 (80%)	580 (20%)	116 (4%)	2077 (71%)
Animal	2114 (67%)	1041 (33%)	18 (0.6%)	2552 (81%)
Worm	4099 (47%)	4686 (54%)	28 (0.3%)	7662 (87%)

ed and more likely to be highly expressed than animal or worm genes (Table 2). Approximately 80% of core genes were detected, as opposed to only 67% of animal genes and 47% of worm genes. Similar observations have been made from an analysis of EST databases (5). However, the quantitative nature of our data coupled with the temporal information allows a deeper analysis. For example, 4% of core genes were detected at frequencies greater than 30 ppm at each developmental time point, but less than 1% of animal or worm genes were detected ($P < 1 \times 10^{-3}$, χ^2 test). Similarly, in terms of the fraction of total transcripts, core genes accounted for most of the transcripts among nondevelopmentally modulated, highly expressed genes (Fig. 2A). In contrast, worm genes accounted for a larger proportion of transcripts among the developmentally modulated genes, and this proportion rose during the course of development (0 to 60 hours), whereas the fraction of expression due to core genes concurrently dropped (Fig. 2B). To determine if the increase in worm-specific expression was driven by a small number of abundant, structural transcripts, we examined the likelihood of worm, animal, and core genes being called “present” at each developmental time and found the same trend, with slightly reduced amplitude. This suggests that, although abundant genes contribute importantly to the trends in Fig. 2B, lower abundance messages also contribute substantially to the increase in worm-specific expression. Thus, the trends in expression during nematode development are consistent with a simple model that envisages the multicellular organism in terms of an ancient cellular core that is organized and regulated by newer genes that evolved from that core. Although this model is too simplistic to be predictive of individual gene functions, it is striking that groups of genes with similar histories are expressed in accord with such a framework.

References and Notes

- W. B. Wood, Ed., *The Nematode Caenorhabditis elegans* (Cold Spring Harbor Laboratory Press, Plainview, NY, 1988).
- D. L. Riddle, T. Blumenthal, B. J. Meyer, J. R. Priess, Eds., *C. elegans II* (Cold Spring Harbor Laboratory Press, Plainview, NY, 1997).
- D. J. Lockhart et al., *Nature Biotechnol.* **14**, 1675 (1996).
- L. Wodicka, H. Dong, M. Mittmann, M. H. Ho, D. J. Lockhart, *Nature Biotechnol.* **15**, 1359 (1997).
- The *C. elegans* Sequencing Consortium, *Science* **282**, 2012 (1998).
- Nematodes were grown at 25°C with concentrated OP50 as a food source. For each developmental stage, embryos representing all 12 hours of embryonic development were isolated by hypochlorite treatment of embryos and adults collected from growing mixed-stage populations. For the 0-hour sample, a 1-ml aliquot of embryos was flash-frozen in liquid nitrogen and stored at -80°C . For the larval preparations, the embryos were plated at low density on seeded plates and allowed to develop for the specified time. Additional details on nematode growth conditions, preparation of purified oocytes, and RNA isolation and labeling procedures are available as supplemental information (14).
- The Affymetrix GeneChip software provided for each transcript an “absolute decision,” which predicted whether the gene was “present” or “absent.” The algorithm used to calculate the absolute decision is described in the Affymetrix GeneChip Analysis Suite User Guide (Affymetrix, Santa Clara, CA). Essentially, a gene was called “present” if its specific hybridization intensity was significantly above array background and noise levels. Additional details are available as supplemental information (14).
- M. Mochii, S. Yoshida, K. Morita, Y. Kohara, N. Ueno, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 15020 (1999).
- Eleven control transcripts were spiked into each hybridization. The abundances of these transcripts (stated in terms of control transcripts per total transcripts), ranged from 1:300,000 (or 3 ppm) to 1:1000 (or 1000 ppm). These abundances were calculated by assuming an average worm RNA size of 1000 ribonucleotides (3). We plotted the absolute decision (7) as a function of abundance for each of these transcripts, scoring “present calls” as one, and “absent calls” as zero. By logistic regression of these 11 data points, we estimated the likelihood of receiving a present call as a function of transcript abundance. In general, the likelihood of receiving a present call was unity for the highest abundance control transcripts and close to zero for the least abundant control transcript. Using the regression, we defined the sensitivity of detection for each hybridization as the transcript abundance at which the likelihood of receiving a present call fell below 70%. This sensitivity estimated the absolute transcript level at which the hybridization signal became indistinguishable from noise. Because of several sources of variation that affect the quantitation of the control transcripts, we think that our sensitivity estimates and the frequencies (12) derived from them are accurate to within a factor of ~ 3 . Details about the false-positive and false-negative rates of the present call, their impact on our detection limit, and further comments on the uncertainty of our quantitation are available as supplemental information (14).
- We isolated 1.2 mg of total RNA from 1 ml of embryos (~ 7 million embryos). With an average of 3% polyadenylated RNA and an average transcript length of 1 kb, this corresponds to 10 million transcripts per embryo.
- A. A. Hill, C. P. Hunter, B. T. Tsung, G. Tucker-Kellogg, E. L. Brown, data not shown; L. R. Baugh, A. A. Hill, C. P. Hunter, D. Mootz, data not shown.
- The GeneChip software provided a specific hybridization intensity value or “average difference” (AD) for each transcript that was proportional to transcript

- abundance (3). We normalized the AD values to "frequency" values by referring the AD values on each chip to a calibration curve constructed from the AD values for the 11 control transcripts with known abundances that were spiked into each hybridization (9). This "frequency normalization" allowed comparison of transcript measurements across multiple array experiments. Frequency values for each gene were expressed in number concentrations (transcripts per million, or ppm), under the assumptions described (9).
13. P. Tamayo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 2907 (1999).
 14. Supplemental information is available at www.sciencemag.org/feature/data/1053496.shl.
 15. I. L. Johnstone, J. D. Barry, *EMBO J.* **15**, 3633 (1996).
 16. M. C. Costanzo *et al.*, *Nucleic Acids Res.* **28**, 73 (2000). (The Proteome WormPD database is available at <http://www.proteome.com/databases/index.html>).
 17. C. E. Rocheleau *et al.*, *Cell* **90**, 707 (1997).
 18. P. W. Carter, J. M. Roos, K. J. Kemphues, *Mol. Gen. Genet.* **221**, 72 (1990).
 19. Additional details are available as supplemental information (14).
 20. J. H. Xin, B. P. Brandhorst, R. J. Britten, E. H. Davidson, *Dev. Biol.* **89**, 527 (1982).
 21. S. A. Chervitz *et al.*, *Science* **282**, 2022 (1998).
 22. G. M. Rubin *et al.*, *Science* **287**, 2204 (2000).
 23. To classify worm genes, we used the sequence comparison results of Rubin *et al.* (22). In the period between the date of our chip design (early 1999) and the date of this more recent sequence comparison (early 2000), some of the worm ORFs that had been included in the chip designs had been deleted or altered by worm sequence curators (for details on the curation of the Wormpep database, see http://www.sanger.ac.uk/Projects/C_elegans/wormpep/). To reconcile worm ORFs in the sequence comparison with worm ORFs on the arrays, we limited our analysis to worm ORFs on the arrays that were unchanged between the two data sets, using curation lists maintained at the Sanger Centre (the Wormpep.history file). Thus, we identified on the arrays 2905 worm genes shared between yeast, worm, and fly (core genes); 3150 worm genes shared between worm and fly, but not yeast (animal genes); and 8741 worm genes that were unique to the worm (worm genes).

Shared worm genes were defined as those that had significant homology to one or more genes in yeast or fly (BLASTP expectation values $E < 1 \times 10^{-10}$, as described by Rubin *et al.*) Discrepancies between these numbers of ORFs and comparable counts in Rubin *et al.* are accounted for by the facts (i) that 2% of worm ORFs were not monitored directly by the arrays and (ii) that, to be conservative, we excluded from consideration ORFs that had either been deleted or altered in the Wormpep database in the interval between the chip design and the sequence comparison of Rubin *et al.*

24. We thank M. Whitley for advice on array experiments; K. Griffiths for array bioinformatics support; E. Wilson, A. Velasco, and H. Horton for technical assistance; J. Freeman for bioinformatics support during the chip design process; G. Sherlock for sharing yeast and worm sequence comparison data; M. Yandell for sharing yeast, fly, and worm sequence comparison data; and D. Mootz for providing purified oocytes.

26 June 2000; accepted 29 September 2000

Song Replay During Sleep and Computational Rules for Sensorimotor Vocal Learning

Amish S. Dave and Daniel Margoliash*

Songbirds learn a correspondence between vocal-motor output and auditory feedback during development. For neurons in a motor cortex analog of adult zebra finches, we show that the timing and structure of activity elicited by the playback of song during sleep matches activity during daytime singing. The motor activity leads syllables, and the matching sensory response depends on a sequence of typically up to three of the preceding syllables. Thus, sensorimotor correspondence is reflected in temporally precise activity patterns of single neurons that use long sensory memories to predict syllable sequences. Additionally, "spontaneous" activity of these neurons during sleep matches their sensorimotor activity, a form of song "replay." These data suggest a model whereby sensorimotor correspondences are stored during singing but do not modify behavior, and off-line comparison (e.g., during sleep) of rehearsed motor output and predicted sensory feedback is used to adaptively shape motor output.

In reinforcement learning, systems learn through interaction with the environment by trying to optimize some measure of performance. Biological systems may experience a substantial delay between premotor activity and assessment of performance through sensory feedback (1). This delay poses the problem of how to reward or punish a premotor circuit when that circuit is participating in a different task by the time the reward or punishment is computed. Reinforcement learning is further complicated in systems such as vocal learning, where the mapping of sensory feedback (fundamentally represented as frequency versus time) onto motor output (mus-

cle dynamics) is of high dimensionality (a many-to-many dynamic mapping). Methods developed in the field of machine learning solve the problem of reinforcement learning with delayed reward (2), and a variety of biological solutions have been proposed to the problem of learning sequences of actions (3). Here, we report on neuronal data that represent a solution to the problem of sensorimotor mapping in the bird vocal-motor ("song") system. The physiological properties observed during sleep also suggest an algorithmic implementation for reinforcement learning of song.

Zebra finch songs are organized hierarchically, with one or more notes composing a syllable, and sequences of syllables forming a motif, which are repeated to form song. We investigated neurons in the forebrain nucleus robustus archistriatalis (RA), whose descending projections represent the output of the

forebrain song system. During singing, RA neurons exhibit short bursts of activity, whose identity varies with the note that immediately follows the burst (4). In awake birds, outside the context of vocalizations, RA neurons are regularly firing. RA neurons also prominently burst "spontaneously" and respond to sounds, but only during sleep (5). With the goal of comparing motor, auditory, and ongoing bursting activity, we recorded single neurons in the RA of singing male zebra finches, permitted the animals to fall asleep by turning off the lights, and then tested the same neurons' sensory and ongoing discharge properties (6, 7).

The spiking patterns of RA neurons in singing birds consisted of phasic patterns of premotor excitation superimposed over a background of profound inhibition (4) (Fig. 1, B and C). This premotor activity was virtually invariant for multiple occurrences of the same sound. After the lights were turned off, RA auditory responses were initially weak but gained strength with time, reflecting the gradual transition into sleep (5). Responses to playback of the bird's own song (BOS) also consisted of phasic patterns of excitation separated by inhibition that were similar for multiple occurrences of the same sound, differing mainly in the strength of response rather than pattern (8).

The timing of auditory responses to the BOS was very well aligned to the timing of premotor activity (Fig. 1F). The only exceptions were instances of silence following the end of a motif or the end of song, where the auditory response could include an additional burst that corresponded with the syllable that would have followed if the song had continued without pause. To compare motor and auditory activity, we analyzed the singing-related activity surrounding each syllable of song (4, 9). The spike patterns from the response to the BOS playback were then compared with the spike patterns from premotor activity derived from

Department of Organismal Biology and Anatomy, University of Chicago, 1027 East 57 Street, Chicago, IL 60637, USA.

*To whom correspondence should be addressed. E-mail: dan@bigbird.uchicago.edu