proteins is with enzymes that are responsible for maintaining them in the appropriate redox state for their activities, analogous to the DsbB-DsbA interaction. These reactions also appear to involve mixed disulfide intermediates. However the mixed disulfide bonded reaction intermediates in either direction are, in most cases (*28*), difficult to detect probably because such complexes resolve very quickly (*29*). Here, in the case of DsbA, we show that it is possible to detect such intermediates by mutations that alter a *cis* proline residue highly conserved in the structures of proteins with a thioredoxin-like fold. A striking finding is that one of the mutations (P151T) results in accumulation of complexes between DsbA and its substrates, whereas the other (P151S) accumulates complexes between DsbA and the enzyme that oxidizes it, DsbB. Because this proline residue is conserved in most thioredoxin family members, it seems possible that mutational alteration of this residue in other family members will also allow detection of reaction intermediates. However, success in such efforts may depend on the precise geometry and distance of the proline relative to the active site and the nature of the amino acid substituted.

Recently, mutants lacking the C-terminal cysteine of the Cys-X-X-Cys motif of a plant thioredoxin were used to trap putative enzyme-substrate complexes (*30*, *31*). This approach may not be as informative for proteins that act as oxidants such as DsbA, where the formation of such complexes would represent the reverse reaction from that normally carried out by the protein.

Further studies with the DsbA P151T mutant should allow identification of a larger number of DsbA substrates. Analysis of these complexes should also yield a more detailed understanding of the mechanism of DsbA action, including (i) which cysteines in substrate proteins are recognized during its action, (ii) at what point during protein translocation these reactions take place, and (iii) the role of the *cis* proline in this process.

References and Notes

1. J. L. Martin, *Structure* **3**, 245 (1995).
2. C. S. Sevier, C. A. Kaiser, *Nature Rev. Mol. Cell Biol.* **3**, 836 (2002).
3. H. Kadokura, F. Katzen, J. Beckwith, *Annu. Rev. Biochem.* **72**, 111 (2003).
4. J. C. Bardwell, K. McGovern, J. Beckwith, *Cell* **67**, 581 (1991).
5. S. Kamitani, Y. Akiyama, K. Ito, *EMBO J.* **11**, 57 (1992).
6. A. Zapun, J. C. Bardwell, T. E. Creighton, *Biochemistry* **32**, 5083 (1993).
7. U. Grauschopf *et al.*, *Cell* **83**, 947 (1995).
8. S. Kishigami, Y. Akiyama, K. Ito, *FEBS Lett.* **364**, 55 (1995).
9. J. C. Bardwell *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1038 (1993).
10. T. Kobayashi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 11857 (1997).
11. M. Bader, W. Muse, D. P. Ballou, C. Gassner, J. C. Bardwell, *Cell* **98**, 217 (1999).
12. G. Jander, N. L. Martin, J. Beckwith, *EMBO J.* **13**, 5121 (1994).
13. N. J. Darby, T. E. Creighton, *Biochemistry* **34**, 3576 (1995).
14. F. Vinci, J. Couprie, P. Pucci, E. Quéméneur, M. Moutiez, *Protein Sci.* **11**, 1600 (2002).
15. S. Kishigami, E. Kanaya, M. Kikuchi, K. Ito, *J. Biol. Chem.* **270**, 17072 (1995).
16. H. Tian, D. Boyd, J. Beckwith, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4730 (2000).
17. H. Kadokura, M. Bader, H. Tian, J. C. Bardwell, J. Beckwith, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10884 (2000).
18. Materials and methods are available as supporting material on *Science* Online.
19. L. W. Guddat, J. C. Bardwell, T. Zander, J. L. Martin, *Protein Sci.* **6**, 1148 (1997).
20. J. B. Charbonnier, P. Belin, M. Moutiez, E. A. Stura, E. Quéméneur, *Protein Sci.* **8**, 96 (1999).
21. H. Kadokura *et al.*, data not shown.
22. H. Kadokura, J. Beckwith, *EMBO J.* 21, 2354 (2002).
23. U. Grauschopf, A. Fritz, R. Glockshuber, *EMBO J.* **22**, 3503 (2003).
24. R. Ortenberg, F. Katzen, J. Beckwith, unpublished results.
25. G. Richarme, *Biochem. Biophys. Res. Commun.* **252**, 156 (1998).
26. J. Qin, G. M. Clore, W. P. Kennedy, J. Kuszewski, A. M. Gronenborn, *Structure* **4**, 613 (1996).
27. A. Bateman *et al.*, *Nucleic Acids Res.* **30**, 276 (2002).
28. M. Molinari, A. Helenius, *Nature* **402**, 90 (1999).
29. C. Frech, M. Wunderlich, R. Glockshuber, F. X. Schmid, *EMBO J.* **15**, 392 (1996).
30. Y. Balmer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 370 (2003).
31. K. Motohashi, A. Kondoh, M. T. Stumpp, T. Hisabori, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11224 (2001).

32. We thank members of J.C.A.B. and J.B. laboratories for discussions. J.C.A.B. is a Pew Scholar. J.B. is an American Cancer Society Professor. Supported by NIH grants GM57039 (J.C.A.B.) and GM41883 (J.B.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/303/5657/534/DC1
Materials and Methods
Tables S1 to S3
References

22 September 2003; accepted 26 November 2003

Extensive Gene Traffic on the Mammalian X Chromosome

J. J. Emerson,[1]* Henrik Kaessmann,[1,2]* Esther Betrán,[1,3] Manyuan Long[1]†

Mammalian sex chromosomes have undergone profound changes since evolving from ancestral autosomes. By examining retroposed genes in the human and mouse genomes, we demonstrate that, during evolution, the mammalian X chromosome has generated and recruited a disproportionately high number of functional retroposed genes, whereas the autosomes experienced lower gene turnover. Most autosomal copies originating from X-linked genes exhibited testis-biased expression. Such export is incompatible with mutational bias and is likely driven by natural selection to attain male germline function. However, the excess recruitment is consistent with a combination of both natural selection and mutational bias.

The mammalian X and Y chromosomes changed profoundly in their differentiation from ancestral autosomes (*1*–*3*). Throughout this process, the selective placement of new genes can be driven by gene duplication (*1*, *4*). The Y chromosome has been shown to recruit male-specific genes (*1*, *2*), whereas a few individual X-linked genes have male-specific duplicate counterparts on autosomes (*4*). Furthermore, some male-specific genes appear to be enriched on the X chromosome (*5*, *6*). However, analysis of human genome project data indicated that no pattern exists for gene movements involving the X chromosome in humans (*7*). To elucidate gene movements in the human and mouse genomes, we analyzed duplicate genes produced by retroposition, whereby a mature messenger RNA (mRNA) is reverse-transcribed and reintegrated into the genome (*8*).

Retroposition is an important mechanism of gene copying (*9*) and produces a large number of functional genes in mammalian genomes. It accounts for approximately 10,000 duplication events in the human genome (*10*), of which approximately 10% are functional (*11*). The direction of copying can be inferred from sequence features of each member of the duplicate pair (*12*): the processed retrocopy is intronless, whereas its parental gene usually contains introns (*13*). Retrocopies that recently integrated into the genome may also display a 3′ polyadenylate [poly(A)] tract and flanking direct repeats. We screened annotated genes in the human genome (*7*, *14*, *15*) for functional retropositions (*16*), identifying 655 retroposition events, of which 366 involved interchromosomal movements. Furthermore, with the use of stringent functionality criteria based on both selective constraint, a nonsynonymous to synonymous substitution rate ratio (K_A/K_S) less

[1]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA. [2]Center for Integrative Genomics, BEP, University of Lausanne, CH-1015 Lausanne, Switzerland. [3]Department of Biology, University of Texas at Arlington, Arlington, TX 76019, USA.

*These authors contributed equally to this work
†To whom correspondence should be addressed. E-mail: mlong@uchicago.edu

than 0.5, and expression evidence (*16*), we identified 94 functional retroposed pairs where the retrogenes do not share chromosome linkage with their parental copy (tables S1 and S2). We also applied our screening process to the mouse genome (*15*–*17*) and identified 105 pairs. This screen provides a conservative sample of functional duplications for a rigorous analysis of gene movements among chromosomes. To study neutral patterns of gene movements, we also generated a data set of 1859 human nonfunctional retropseudogenes and their corresponding parental genes in a separate genome-wide screen (*16*).

Analysis of chromosomal locations of parental genes reveals that the human X chromosome, in comparison to the autosomes, harbors a substantial excess of genes that generate new retrocopies. The X chromosome is the only outlier in the genome (Fig. 1A; table S3), a pattern that is also reflected in mouse (Fig. 1C; table S4). Previous work in humans (*18*) reported a random insertion pattern of processed pseudogenes generated from retroposition. We thus assume that (i) chromosomes generate a number of retrogenes proportional to the number of genes on that chromosome and (ii) chromosomes accept retroposed copies in proportion to the size of the chromosome and, thereby, we were able to calculate the expected number of retrocopies that each chromosome generates (*16*). On the basis of these expectations, we found that the human X chromosome has generated a 299% excess of retrogenes, and we conducted a Monte Carlo sampling test, showing that this excess is significant (P = 0.00012, Table 1). Similarly, in the mouse genome, the X chromosome has generated a 309% excess (P = 0.00011) relative to the autosomes.

This nonrandom excess of parental genes on the X chromosome may be explained in two ways: (i) a mechanistic bias, such as either an insertion/deletion bias or an excess expression of X-linked genes in gonads, or (ii) natural selection, which selects for preservation of genes copied to the autosomes. A genome-wide analysis of both expressed sequence tag (EST) tissue expression (15,666 Unigene clusters) as well as gonad microarray data (10,857 Affymetrix probes) provides evidence against the first hypothesis. For both data sets, neither testis nor ovary shows a level of X expression that significantly exceeds that of the autosomes (fig. S1). If a mechanistic bias explains the pattern, then

Table 1. Multinomial Monte Carlo resampling test of the difference between observation and the expected frequencies. X→ and →X refer to the predictions and observations for genes leaving and entering the X, respectively; A→ and →A predictions and observations for genes leaving and entering an autosome, respectively. The *P* values were calculated from Monte Carlo simulations (*16*) by analyzing the movements involving both autosomes and the X chromosome. Excess = (Observed – Expected)/Expected.

Direction of retroposition	Expected	Observed	Excess (%)	*P*
Human				
X→	3.76	15	299	0.00012
A→	90.24	79	−12	
→X	3.61	13	260	0.00244
→A	90.39	81	−10	
Mouse				
X→	4.16	17	309	0.00011
A→	100.84	88	−13	
→X	4.62	16	246	0.00015
→A	100.38	89	−11	

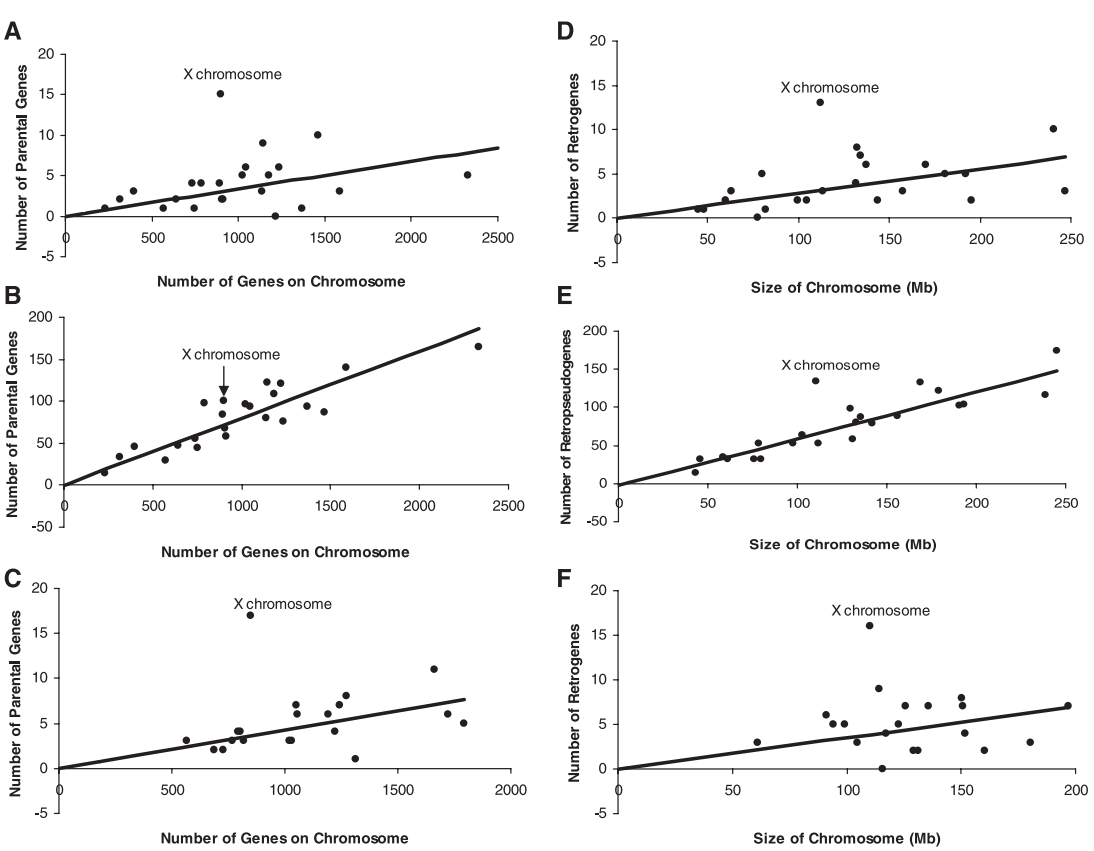

Fig. 1. Regressions for the parental genes of retrogenes in (**A**) human and (**C**) mouse and for the parental genes of retropseudogenes (**B**) in human. Regressions for the size of a chromosome in (**D**) human and (**F**) mouse and for the retropseudogenes (**E**) in human. In the plots, X is shown as 75% of its size as predicted by the model (*15*), although allowing X to assume 100% of its size does not change the results. Probabilities for the hypothesis that the chromosome with the highest observed/expected ratio [where the expected number is calculated as in (*16*)] is an outlier are calculated using Grubbs and Dixon outlier tests (*16*). For every distribution [except (B)], the X has the largest ratio and is an outlier with $P < 0.005$ and $P < 0.01$ for the Grubbs and Dixon tests, respectively; (B) shows no such outliers.

retropseudogenes should show the same pattern as functional genes. When we examine the distribution of the 1859 retropseudogenes, we see no excess of X-linked parental genes (Fig. 1B; $P > 0.1$, table S3), suggesting that some force other than a mutational bias determines the fate of functional retrogenes.

Indeed, both empirical and statistical lines of evidence support the hypothesis of natural selection on gene location. What selective pressures explain this export of functional retrogenes? One model of sexual antagonism (*19*) predicts that gene copies that benefit males at a cost to females would be more likely found on the autosomes than on the X chromosome. A favorable gene that has an effect on the fitness of heterozygous carriers and benefits males at the expense of reducing female fitness is more likely to spread if it is autosomal rather than X-linked, because an X-linked gene spends two-thirds of its time in females compared with one-half for an autosomal gene and thus the X chromosome becomes "demasculinized." Another explanation is that the condensation of the X chromosome during male meiosis silences many X-linked genes (*20*). Such precocious inactivation of the X chromosome would create two different genomic environments in which autosomal genes may carry out male functions more effectively than if they were X-linked. This inactivation process may drive the X chromosome to export functional retroposed gene copies. Both selective models predict that the X-derived autosomal retrogenes that develop beneficial male expression to enhance functions during male meiosis would be more likely to survive and spread on autosomes. Thus, selection will cause biased fixation or preservation of retrogenes created by X-to-autosome retroposition over evolutionary time scales, although retroposition events happen randomly, as we have shown (Fig. 1B).

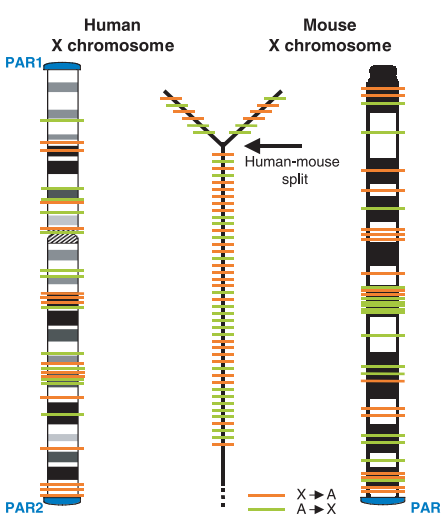

Fig. 2. Cytologic location of the parental genes (orange) and retrogenes (green) in the X chromosome of human and mouse. Pseudoautosomal regions (PAR1 and PAR2 in human and PAR in mouse) are indicated (*2*). None of the genes maps to these small regions known to have been acquired from autosomes by translocation in the recent past. The approximate age of the duplication events involving autosome-to-X retroposition (A→X, green) and X-to-autosome retroposition (X→A, orange) is indicated in a schematic phylogeny.

To analyze expression patterns of human retrogenes, we performed similarity searches of retrogene sequences against all available ESTs from the Unigene database (*16*). Of the 15 previously identified X-derived retrogenes in humans, 3 are known genes that developed male germline function after leaving the X: *Pgk-2*, *Pdha2*, and *hnRNP* (*4*, *21*, *22*). We identified four more genes with male-specific function: a transketolase, a subunit of a transcription factor, a ras-related protein, and a ribose phosphate pyrophosphokinase (*23*). Two of the mouse retrogenes have also been characterized as male-germline genes that originate on the X chromosome, including *Pdha2* and *Cst-64*-like genes (*24*). Other retrogenes that fit this pattern from the literature are reported in supporting online material (table S5). Furthermore, in genes with expression data available, ~77% (10/13) of the autosomal retrogenes originating from the X chromosome have testis expression, whereas only ~44% (20/45) of the other autosomal retrogenes have testis expression, a significant difference ($P < 0.05$). Thus, the number of retrogenes exported from the X chromosome is incompatible with mutational bias, revealing the force of selection for these male-germline genes. Autosomes are also the preferred location for many *Drosophila* and *C. elegans* male-specific genes (*9*, *25*–*27*).

A previous study noted an abundance of X-linked genes expressed in spermatogonia in mouse (10/25 genes specific to mitosis map to the X) (*6*). However, a contrasting pattern is found for later stages of spermatogenesis. For meiotic and postmeiotic expression (*28*), there are few X-linked genes (1/27 and 2/78, respectively) (table S6). Thus, for genes expressed during X inactivation and in haploid cells, there is a dearth of X-linked genes.

As with retrogene origination, the X chromosome is the only outlier (tables S3 and S4) in the genome with respect to recruitment, accepting an excess of functional copies from autosomes (Table 1, $P = 0.00244$, Fig. 1D). This result is in contrast to data from the *Drosophila* genome (*9*), but data from mouse corroborate the human pattern (Table 1, $P = 0.00015$, Fig. 1F). We applied the model of random retrogene insertion to predict how many genes a chromosome is expected to accept, confirming that the mammalian X chromosome recruits a significant excess of functional retrogenes (Table 1). Two hypotheses may explain the biased X recruitment: (i) a mechanistic bias, a disparity in the molecular mechanisms of either maintaining or accepting retrogenes, or (ii) natural selection favoring the fixation and maintenance of retrogenes.

The mechanistic hypothesis may account for part of the pattern. For example, if recombination occurs less frequently on the X chromosome, as is observed in humans (*29*, *30*), deleterious chromosomal rearrangements caused by ectopic exchanges between duplicate genes would be less common than on autosomes (*31*). Another possibility is that an insertional bias exists on the X chromosome.

To test the insertion/deletion bias hypothesis of X-retrogene recruitment, we examined human retropseudogene distribution. Notably, the X chromosome is an outlier for the regression in this direction of movement (Fig. 1E; table S3). The number of retropseudogenes entering the X chromosome is nearly twice that predicted by both the random model and the regression results. A similar bias has been observed in repetitive retroelements (*32*, *33*), as predicted if these elements exploit the same machinery as duplication via retroposition. Consistently, an upper limit for this bias is observed with LINE1 elements, which show a ~100% excess of total insertion lengths on the X chromosome in humans (*32*, *33*). This excess suggests that a bias in extra insertions or fewer deletions of both functional and inactivated retroposed genes exists for the X chromosome and accounts for part of the observed pattern. However, a Fisher's exact test of the 2×2 contingency table comparing the number of X-linked versus autosomal retropseudogenes to the X-linked versus autosomal functional retrogenes shows that the functional genes disproportionately enter the X chromosome compared with nonfunctional genes ($P < 0.05$, table S9), suggesting that the mechanistic bias is not only reason for the recruitment excess.

Such a bias may also result from selection on recessive advantageous retropositions to the X chromosome that are beneficial when hemizygous in males (*34*); i.e., a distinct selective sieve will operate for the genes generated in this direction. In contrast, the selective sieve would favor the insertion to autosomes with at least partially dominant effects, consistent with the aforementioned observation of X-to-autosome retroposition. Of these X-linked retrogenes, we detect female expression for only ~14% (1/7) for which expression data are available, whereas in the autosomal genes ~71% (32/45) are expressed in female tissues ($P < 0.01$). This paucity of female expression suggests that retrogenes entering the X chromosome tend to avoid female expression, possibly decreasing disadvantageous effects on females and thus facilitating the spread of such insertions on the chromosome.

The presence of these two patterns in both humans and mouse suggests their importance in the evolution of mammalian X chromosomes. Our sample of functional retroposed genes in the mammalian genomes is likely at least an order of magnitude smaller than the actual number (*10*, *11*). Notably, our analyses exclude retrocopies maintaining introns, such as partially processed retrogenes (*35*) or chimeric genes (*36*), which would implicate even more genes. Finally, other mechanisms of interchromosomal gene movement are also likely influenced by the aforementioned selective forces. Thus, we expect many more genes to be subject to the gene traffic described herein.

To elucidate the age of retrogene movements, we dated the human duplications involving X-linked parents or retrogenes both by comparison to the mouse genome sequence and by sequence divergence analysis (*16*). Most copies that escape X linkage (12/15) as well as most copies that obtain X linkage (10/13) originated before the human-mouse split (Fig. 2, tables S7 and S8). Duplicates in the mouse genome show the same pattern, consistent with this notion. Thus, both patterns result from ancient evolutionary forces common to eutherian mammals. However, this process appears to be an ongoing characteristic of eutherian X evolution, because 6/28 events have occurred subsequent to the human-mouse split in the human lineage, 6/33 retropositions have occurred within the past ~80 million years in the mouse lineage, and some of these retroduplicate pairs have high sequence similarity (>95%) at synonymous sites. This chromosome-biased gene origination appears to be an important process actively driving the differentiation of the X chromosome in mammals and suggests that this differentiation is still in progress.

References and Notes

1. B. T. Lahn, N. M. Pearson, K. Jegalian, *Nature Rev. Genet.* **2**, 207 (2001).
2. H. Skaletsky *et al.*, *Nature* **423**, 825 (2003).
3. J. A. Marshall Graves *et al.*, *Cytogenet. Genome Res.* **96**, 161 (2002).
4. J. R. McCarrey, *Bioscience* **44**, 20 (1994).
5. M. J. Lercher, A. O. Urrutia, L. D. Hurst, *Mol. Biol. Evol.* **20**, 1113 (2003).
6. P. J. Wang, J. R. McCarrey, F. Yang, D. C. Page, *Nature Genet.* **27**, 422 (2001).
7. J. C. Venter *et al.*, *Science* **291**, 1304 (2001).
8. B. Lewin, *Genes VII* (Oxford University Press, New York, 2000).
9. E. Betran, K. Thornton, M. Long, *Genome Res.* **12**, 1854 (2002).
10. P. M. Harrison *et al.*, *Genome Res.* **12**, 272 (2002).
11. L. Z. Strichman-Almashanu, M. Bustin, D. Landsman, *Genome Res.* **13**, 800 (2003).
12. E. Betrán, W. Wang, L. Jin, M. Long, *Mol. Biol. Evol.* **19**, 654 (2002).
13. J. Brosius, *Science* **251**, 753 (1991).
14. E. S. Lander *et al.*, *Nature* **409**, 860 (2001).
15. T. Hubbard *et al.*, *Nucleic. Acids. Res.* **30**, 38 (2002).
16. Materials and methods are available as supporting material on *Science* Online.
17. R. H. Waterston *et al.*, *Nature* **420**, 520 (2002).
18. Z. Zhang, P. Harrison, M. Gerstein, *Genome Res.* **12**, 1466 (2002).
19. C.-I. Wu, E. Y. Xu, *Trends Genet.* **19**, 243 (2003).
20. C. Richler, H. Soreq, J. Wahrman, *Nature Genet.* **2**, 192 (1992).
21. H. H. Dahl *et al.*, *Genomics* **8**, 225 (1990).
22. D. J. Elliott *et al.*, *Hum. Mol. Genet.* **9**, 2117 (2000).
23. M. Taira *et al.*, *J. Biol. Chem.* **265**, 16491 (1990).
24. B. Dass *et al.*, *J. Biol. Chem.* **276**, 8044 (2001).
25. M. Parisi *et al.*, *Science* **299**, 697 (2003).
26. J. M. Ranz, C. I. Castillo-Davis, C. D. Meiklejohn, D. L. Hartl, *Science* **300**, 1742 (2003).
27. V. Reinke *et al.*, *Mol. Cell* **6**, 605 (2000).
28. T. Fujii *et al.*, *EMBO Rep.* **3**, 367 (2002).
29. A. Kong *et al.*, *Nature Genet.* **31**, 241 (2002).
30. H. A. Wichman, *Genetica* **86**, 287 (1992).
31. C. H. Langley *et al.*, *Genet. Res.* **52**, 223 (1988).
32. J. A. Bailey, L. Carrel, A. Chakravarti, E. E. Eichler, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6634 (2000).
33. Z. Gu, H. Wang, A. Nekrutenko, W. H. Li, *Gene* **259**, 81 (2000).
34. B. Charlesworth, J. A. Coyne, N. H. Barton, *Am. Nat.* **130**, 113 (1987).
35. M. B. Soares *et al.*, *Mol. Cell. Biol.* **5**, 2090 (1985).
36. A. Courseaux, J.-L. Nahon, *Science* **291**, 1293 (2001).

37. We thank K. Thornton, Y. Chen, T. Nagylaki, C.-I. Wu, J. Spofford, B. L. Sidlauskas, T. M. Martin, and two anonymous reviewers for helpful comments. Supported by grants NSF Career MCB-0238168 and NIH GM-065429-01A1; a Packard Fellowship (to M.L.), an NSF Predoctoral Fellowship (to J.J.E.), an Emmy Noether fellowship from the Deutsche Forschungsgemeinschaft (to H.K.); and the startup funds (to E.B.) from University of Texas at Arlington.

Supporting Online Material

www.sciencemag.org/cgi/content/full/303/5657/537/DC1
Materials and Methods
Tables S1 to S9
Fig. S1
References

4 August 2003; accepted 26 November 2003

A Map of the Interactome Network of the Metazoan *C. elegans*

Siming Li,[1]* Christopher M. Armstrong,[1]* Nicolas Bertin,[1]* Hui Ge,[1]* Stuart Milstein,[1]* Mike Boxem,[1]* Pierre-Olivier Vidalain,[1]* Jing-Dong J. Han,[1]* Alban Chesneau,[1,2]* Tong Hao,[1] Debra S. Goldberg,[3] Ning Li,[1] Monica Martinez,[1] Jean-François Rual,[1,4] Philippe Lamesch,[1,4] Lai Xu,[5]† Muneesh Tewari,[1] Sharyl L. Wong,[3] Lan V. Zhang,[3] Gabriel F. Berriz,[3] Laurent Jacotot,[1]‡ Philippe Vaglio,[1]‡ Jérôme Reboul,[1]§ Tomoko Hirozane-Kishikawa,[1] Qianru Li,[1] Harrison W. Gabel,[1] Ahmed Elewa,[1]‖ Bridget Baumgartner,[5] Debra J. Rose,[6] Haiyuan Yu,[7] Stephanie Bosak,[8] Reynaldo Sequerra,[8] Andrew Fraser,[9] Susan E. Mango,[10] William M. Saxton,[6] Susan Strome,[6] Sander van den Heuvel,[11] Fabio Piano,[12] Jean Vandenhaute,[4] Claude Sardet,[2] Mark Gerstein,[7] Lynn Doucette-Stamm,[8] Kristin C. Gunsalus,[12] J. Wade Harper,[5]† Michael E. Cusick,[1] Frederick P. Roth,[3] David E. Hill,[1]¶ Marc Vidal[1]¶#

To initiate studies on how protein-protein interaction (or "interactome") networks relate to multicellular functions, we have mapped a large fraction of the *Caenorhabditis elegans* interactome network. Starting with a subset of metazoan-specific proteins, more than 4000 interactions were identified from high-throughput, yeast two-hybrid (HT=Y2H) screens. Independent coaffinity purification assays experimentally validated the overall quality of this Y2H data set. Together with already described Y2H interactions and interologs predicted *in silico*, the current version of the Worm Interactome (WI5) map contains ~5500 interactions. Topological and biological features of this interactome network, as well as its integration with phenome and transcriptome data sets, lead to numerous biological hypotheses.

To further understand biological processes, it is important to consider protein functions in the context of complex molecular networks. The study of such networks requires the availability of proteome-wide protein-protein interaction, or "interactome," maps. The yeast *Saccharomyces cerevisiae* has been used to develop a eukaryotic unicellular interactome map (*1*–*6*). *Caenorhabditis elegans* is an ideal model for studying how protein networks relate to multicellularity. Here we investigate its interactome network with HT-Y2H.

As Y2H baits, we selected a set of 3024 worm predicted proteins that relate directly or indirectly to multicellular functions (*7*). Gateway-cloned open reading frames (ORFs) were available in the *C. elegans* ORFeome 1.1 (*8*) for 1978 of these selected proteins. Of these, 81 autoactivated the Y2H *GAL1::HIS3* reporter gene as Gal4 DNA binding domain fusions (DB-X), and 24 others conferred toxicity to yeast cells. The remaining 1873 baits were screened against two different Gal4 activation domain libraries (AD-wrmcDNA and