

# Identification of 315 genes essential for early zebrafish development

Adam Amsterdam, Robert M. Nissen, Zhaoxia Sun, Eric C. Swindell, Sarah Farrington, and Nancy Hopkins\*

Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

Contributed by Nancy Hopkins, June 8, 2004

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected on April 20, 2004.

**We completed a large insertional mutagenesis screen in zebrafish to identify genes essential for embryonic and early larval development. We isolated 525 mutants, representing lesions in  $\approx 390$  different genes, and we cloned the majority of these. Here we describe 315 mutants and the corresponding genes. Our data suggest that there are roughly 1,400 embryonic-essential genes in the fish. Thus, we have mutations in  $\approx 25\%$  of these genes and have cloned  $\approx 22\%$  of them. Re-screens of our collection to identify mutants with specific developmental defects suggest that  $\approx 50$  genes are essential for the development of some individual organs or cell types. Seventy-two percent of the embryonic-essential fish genes have homologues in yeast, 93% have homologues in invertebrates (fly or worm), and 99% have homologues in human. Yeast and worm orthologues of genes that are essential for early zebrafish development have a strong tendency to be essential for viability in yeast and for embryonic development in the worm. Thus, the trait of being a genetically essential gene is conserved in evolution. This mutant collection should be a valuable resource for diverse studies of cell and developmental biology.**

To identify a significant fraction of the genes essential for early vertebrate development, we developed a method of insertional mutagenesis for the zebrafish using mouse retroviral vectors (1, 2) and applied the method in a large-scale screen. We identified mutants by visual inspection of embryos at 1, 2, and 5 days postfertilization (dpf), by which time they have developed into free-swimming and feeding larvae. Mutants that result in a visible defect by 5 dpf are almost invariably lethal. Here, we describe 315 mutants and their genes and present an analysis of the evolutionary conservation of these genes. The results argue that the mutant collection contains mutations in at least 25% of the genes essential for the development of many different embryonic organs and structures. Re-screens of the collection in our lab to identify mutations in specific developmental processes support this conclusion. This collection should be a valuable resource for diverse studies of cell and developmental biology in this vertebrate.

## Materials and Methods

**Mutagenesis and Gene Cloning.** Retroviral-mediated insertional mutagenesis and the cloning of the mutated genes were carried out as described (1–3).

**Comparative Genomic Analysis.** The amino acid sequence of each fish gene was compared by BLASTP (4) to the reference genomes of *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Homo sapiens*, as well as the nonredundant database of all organisms to find genes absent in the two yeast species but present in other unicellular eukaryotes. Comparative orthologous group (COG) analysis (5) to determine whether homologues were 1:1 orthologues or had other orthology relationships was done by iteratively blasting the top hits from each organism against the genomes of the others. We consider two genes from different species to be 1:1 orthologues if they are not only each other's

reciprocal best BLASTP hits, but also if *no other* gene in either species' genome finds the gene as its top hit. A worm or yeast gene was defined as an "ancestor" of a human/fish gene if several human genes found the worm or yeast gene as their top hit, but no other worm or yeast gene found *any* of those human genes as their top hit.

## Results

**Identification of 25% of the Genes Essential for Early Zebrafish Development.** We isolated 525 insertional mutants that have visible phenotypes by 5 dpf, a time when the embryo has developed from a fertilized egg to a free swimming larva that has begun to feed. Most of these mutations result in lethality, and many mutants die by 5 dpf. Almost all mutants that have not died by 5 days fail to inflate their swim bladder, a phenotype associated with certain death by 2 weeks of age. Thus, we refer to the mutants and mutations as "embryonic lethal(s)" and the mutated genes as "embryonic-essential" genes.

The insertional mutagenesis procedure and the methodology for identifying mutagenic inserts and cloning their flanking DNA have been described (1–3). A summary of the numbers of mutants isolated and flanking sequences and genes cloned is shown in Table 1. The 486 mutants for which we have obtained DNA sequence at the site of the mutagenic insertion probably represent 362 different loci; we have identified the mutated gene for 315 of these, 86 of which we have reported (1, 2). As discussed next, our findings suggest that these 362 loci represent  $\approx 25\%$  of the genes whose mutation leads to an embryonic lethal phenotype. Thus, there are only  $\approx 1,400$  such genes. The following evidence supports this conclusion.

Our collection includes mutations in 5 of 20 (25%) tRNA synthase genes and 26 of the 79 (33%) ribosomal protein genes in the zebrafish genome. We also have mutations in 23 of 97 genes (24%) for which a chemically induced mutant has also been identified and the mutated gene cloned by positional or candidate gene cloning (as of February 2004). Most of the latter genes encode transcription factors, receptors, and ligands. These data argue that we have screened  $\approx 25\%$  of the genes in the fish genome, whether extrapolated from housekeeping-type genes or genes with more specific developmental functions. The data in Table 2 also show that viral integrations do not occur preferentially into housekeeping-type genes because the allele frequency for housekeeping genes is no greater than that for other types of genes.

In theory, a second way to estimate the number of embryonic-essential genes from these data is to apply the Poisson distribu-

Abbreviations: SCE, single-celled eukaryotes; IFT, intraflagellar transport; RNAi, RNA interference; dpf, days postfertilization.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession numbers can be found in Table 3, which is published as supporting information on the PNAS web site).

See accompanying Biography on page 12789.

\*To whom correspondence should be addressed. E-mail: nhopkins@mit.edu.

© 2004 by The National Academy of Sciences of the USA

**Table 1. Insertional mutants recovered**

	Mutants	Loci
Insertional mutants	525	≈390
Mutants with cloned junction DNA	486	362
Mutants with mutated gene identified	438	315

tion to the numbers of single and multiple hits in Table 2 to calculate the number of loci that have not yet been mutated and hence the total number of loci that can give rise to an embryonic-visible phenotype. However, as is almost always seen in mutagenesis screens, the data do not fit the Poisson distribution precisely, making this approach unreliable. Attempting to apply a Poisson distribution under these circumstances usually results in an underestimate of the number of mutable loci (6, 7). For example, by using the number of loci hit twice in our study (62) and keeping either the number of loci hit (362) or the number of mutants obtained (486) fixed, solving the Poisson equation predicts that the total number of mutable loci is either 1,035 or 1,310 mutable loci. This finding is somewhat lower than (but not inconsistent with) the estimate of 1,400 genes obtained above.

The identities of the 315 mutated genes we have identified to date are provided in Fig. 1 (additional details about these genes are described in Table 3). A preliminary phenotypic description of the corresponding 315 mutants is provided in Table 3 and images are available at [http://web.mit.edu/ccr/pnas.zebrafish\\_mutant\\_images/index.html](http://web.mit.edu/ccr/pnas.zebrafish_mutant_images/index.html). More detailed phenotypic descriptions of the defects in particular organs and structures will come from careful re-screens of the collection (see below).

If we take 1,400 as the number of zygotic genes whose mutation can lead to an embryonic visible/lethal phenotype in the zebrafish, the 315 genes listed here represent 22% of the total. Because mutated genes were cloned without regard to phenotype, these genes should be representative of the entire set of protein-coding genes that are genetically essential for early zebrafish development.

Fig. 2 provides a summary of the types of proteins encoded by the 315 embryonic essential genes based on their biochemical function. Many of the genes are probably essential for cell viability whereas others are likely to be required for more specific developmental processes. Mutations in cell-essential genes can survive for 1 to at least several days due to maternal supplies in the egg (8). Mutation of some cell-essential genes may not have been detected in our screen because maternal supplies of some genes are sufficient to sustain the embryo beyond 5 dpf, e.g., *Dicer1* (9). About 20% of all of the genes we identified encode proteins of unknown biochemical function.

**Evolutionary Conservation of Embryonic-Essential Zebrafish Genes.**

To better understand the genetic basis of animal development and its evolution, we asked whether the embryonic-essential fish genes have homologues in yeast, other single-celled eukaryotes

(SCEs) (e.g., *Giardia*, *Plasmodium*, *Trypanosoma*, *Chlamydomonas*), worm, fly, and human genomes. Analysis of the completed human and mouse genomes has already indicated that 50% of their genes are present in SCEs and ≈80% in flies and/or worms (10, 11), (with homology defined as a BLASTP E value of <10<sup>-5</sup>), and our analysis of 50 fish genes selected at random from the July 2003 annotation of the zebrafish genome found a similar proportion of conserved genes (data not shown). Applying the same homology definition to our list of 315 embryonic-essential zebrafish genes, however, we found that a higher fraction, 74% of them, have homologues in yeast or other SCEs, and 93% are present in fly and/or worm. By using a more conservative cutoff requiring homology over 80% of the length of the protein and E < 10<sup>-20</sup>, 64% of the fish genes have homologues in yeast or some other SCE whereas 91% are present in flies or worms. Ninety-nine percent of the genes (all but 2 of the 315) have homologues in mammals. Thus, as found in other species (12, 13), the fish genes with essential functions are more likely than randomly selected genes to be conserved across species. The conservation of the 315 embryonic-essential fish genes in yeast and other SCEs, in worm, in fly, and in human is represented in Fig. 3 by using the more liberal definition of homology. The identity of the human, fly, worm, and yeast homologues of each essential fish gene can be found in Table 3.

**Mutant Phenotypes and the Evolutionary Conservation of Genes Required for Specific Developmental Processes.**

About 30% of embryonic zebrafish mutants identified in gross morphological screens have developmentally “specific” and unique phenotypes whereas 70% display relatively “nonspecific” or common syndromes (7, 14). The latter more frequently result from mutations in cell-essential genes. Many mutants can be placed into either of the two broad phenotypic categories by superficial visual inspection. However, to identify mutants with phenotypic defects in specific organs or processes with certainty requires considerable effort. To accomplish this goal, we are re-screening the insertional mutant collection, a process called “shelf-screening.” Summaries of three such screens from our lab are shown in Fig. 3 *Right*.

Black boxes in the three columns in Fig. 3 *Right* identify genes specifically required (i) to prevent cystic kidney, probably because these genes are required for the normal development of kidney epithelial tubes (15), (ii) to form cartilage that appears normal after staining with Alcian blue (R.M.N., A.A., and N.H., unpublished observations) or (iii) for melanocyte pigmentation (E. Maldonado, A.A., and N.H., unpublished observations). Twelve genes when mutated resulted in kidney cysts, 8 in abnormal cartilage condensation as revealed by appearance after Alcian staining, and 11 affect melanocyte pigmentation. Because we have cloned ≈22% of the genes essential for early fish development, these results predict that mutations in ≈55 genes can give rise to cystic kidney, 36 to abnormal cartilage, and 50 to defects in melanocyte pigmentation in the zebrafish. As discussed later, many genes identified in each screen are components of a common pathway, with different pathways or processes emerging for each screen.

Fig. 3 shows the conservation of the genes identified in each of the three screens in the genomes of yeast, other SCEs, invertebrates (fly or worm), and human. Of the 12 genes that can give rise to cystic kidney when mutated, 6 are shared with SCEs but not with yeast. Three of these are homologues of genes identified in *Chlamydomonas* that encode intraflagellar transport (IFT) proteins required for flagellum formation (16) (see *Discussion*).

Formation of cartilaginous structures requires the deposition of proteoglycans in the extracellular matrix (17). Among the eight cartilage mutants identified, four have lesions in genes required for proteoglycan synthesis, predicting that ≈20 such

**Table 2. Allele frequencies**

Loci with one allele	275
Loci with two alleles	62
Loci with three allele	17
Loci with four alleles	6
Loci with five allele	1
Loci with seven alleles	1
Average allele frequency	1.34
Average allele freq, RP and tRS	1.10
Average allele freq, ENU cloned	1.95

RP, ribosomal protein genes; tRS, tRNA synthase genes; ENU cloned, genes identified as the cause of ENU-induced mutants.



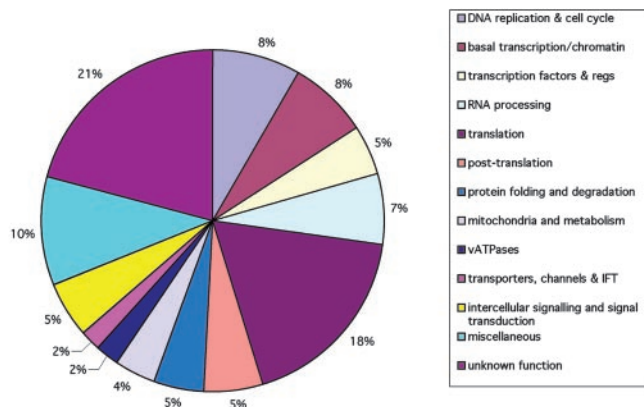
IN YEAST	
<b>Cell cycle</b>	
DNA replication	
2618 Replication Protein A 70kd subunit	
783 DNA pol e subunit B	
1146 DNA polymerase alpha	
1411 MCM7	
471 MCM3-associated protein	
4026 FEN1b (flap structure specific endonuclease 1)	
1244 MCM2	
319 MCM3	
1887 Replication Factor C, subunit RFC5	
2877A Replication Factor, subunit RFC4	
3512C DNA polymerase beta	
<b>mitosis</b>	
2479 wee1a	
2080 NDC80 / HEC (highly expressed in cancer)	
1045 aurora -related ser/thr kinase	
1730 kns1-5 (mitotic kinesin-like 5)	
2769 cell division cycle associated 1-like	
1856 polo-like kinase	
2734 cyclinB1	
486 kinesin-related motor protein EG5	
1113A chromosome adhesion protein SMC1-like	
1143 chromosome adhesion protein SMC4-like 1	
2529 RAD21 homolog	
3235 odc2 homolog	
<b>Transcription</b>	
<b>basal transcription</b>	
2635 RNA polymerase II subunit B	
3639 RNA polymerase I 194 kDa subunit	
3685 RNA polymerase II subunit G	
2718B RNA polymerase II subunit D	
2393 RNA polymerase I polypeptide D	
3635 topoisomerase 2	
1942 RNA polymerase III 39kDa subunit	
2487 TFIIIA	
1124 RNA polymerase I 140 kDa subunit	
998 RNA polymerase II CTD phosphatase	
1621 suppressor of Ty 6 homolog	
<b>chromatin remodeling</b>	
550 SMF2-related	
2786A histone deacetylase 3	
1618 histone deacetylase 1	
<b>transcription cofactor</b>	
296B ski-interacting protein	
889 TAF55	
2143 EXML1	
3079 taube nuss / TAF8	
2491 TAF II 50	
1055B pontin/TIP49	
2394 reclin	
<b>transcription factor</b>	
962 nuclear transcription factor Y, gamma	
<b>RNA processing</b>	
<b>splicing</b>	
1950 spliceosome-associated factor 61	
2791 U4U6-associated splicing factor HPRP3	
3808 small nuclear ribonucleoprotein F	
199 splicing factor UZAF 35kd	
3394A splicing factor 3b, subunit 1, 155kDa	
601 small nuclear ribonucleoprotein D1	
1159 spliceosome-associated protein 49	
1371 U1 small nuclear ribonucleoprotein C	
2271 Splicing factor 3B subunit 10	
3018A U1 small ribonucleoprotein 70kd	
<b>polyadenylation</b>	
2666 Cleavage and polyadenylation specificity factor 3	
2675 Cleavage and polyadenylation specificity factor 1	
2908 Cleavage Stimulation Factor 77kd subunit	
1 Cleavage and polyadenylation specificity factor 4	
<b>RNA export</b>	
3248 nuclear RNA export factor 1	
1464 DEAD/H box 19 RNA helicase	
1951 RAE1 homolog (RNA export)	
4181A QLE1-like, RNA export mediator	
3237 Tho2	
<b>mRNA capping</b>	
3112B mRNA capping enzyme	
<b>Translation</b>	
<b>ribosome biogenesis</b>	
215 DEAD/H box 55 RNA helicase	
1581 nucleolar phosphoprotein Nopp34	
2581 fibrillarin	
3073 DEAD/H box 56 RNA helicase/noh61	
3101 NOP56	
3118 NOPS/NOP58	
2078 nucleolin	
2689 NNP-1/Nop52	
1769 weakly similar to SRF6 (nucleolar protein)	
2 pescidillo	
63 Nop14-like	
1086A DEAD box polypeptide 27	
1019 AD24	
1447 DEAD/H box 97 RNA helicase	
2289 DEAD/H box 51 RNA helicase	
2578 Nop10 (nucleolar)	
3120 YTM1 homolog	

2430 40S ribosomal protein S15	
577B 40S ribosomal protein S5	
10 60S ribosomal protein L36a	
258 60S ribosomal protein L35	
1026 40S ribosomal protein S18	
1061 60S ribosomal protein L7	
1290 40S ribosomal protein S3a	
1444 60S acidic ribosomal protein P1	
1479 40S ribosomal protein SA	
1807 60S ribosomal protein L36	
1974 40S ribosomal protein a8	
1987 60s ribosomal protein L19	
2649 40S ribosomal protein S15A	
2799 40S ribosomal protein s11	
2903 40S ribosomal protein s29	
3076 60S ribosomal protein L12	
1034B 40S ribosomal protein S7	
3893 60S ribosomal protein L28	
2437 60S ribosomal protein L3	
1016 60S ribosomal protein L13	
3820B 60S ribosomal protein L11	
823 60S ribosomal protein L14	
1227 40S ribosomal protein S12	
1284 60S ribosomal protein L24	
3655B 60S ribosomal protein L6	
1422 60S ribosomal protein L9	
<b>ribosomal protein</b>	
<b>RNA synthase</b>	
558 valyl-tRNA synthetase	
2216 glutamyl-tRNA synthetase	
2586 lysyl-tRNA synthetase	
4134 methionine-tRNA synthetase	
3817 seryl-tRNA synthetase	
<b>RNA modification</b>	
343 tRNA nucleotidyl transferase-like	
<b>translation initiation</b>	
3202B poly A binding protein, PABPC1	
929 eukaryotic translation initiation factor 2 gamma	
2230 eukaryotic translation initiation factor 3, subunit 7	
3963 eukaryotic translation initiation factor 3, subunit 4	
2596 eukaryotic translation initiation factor 4A	
2470 eukaryotic translation initiation factor 3, subunit 6	
<b>translation elongation</b>	
1118B eukaryotic translation elongation factor 1 alpha	
4 eukaryotic translation elongation factor 1 gamma	
<b>Post-translation</b>	
<b>protein sorting</b>	
3017 FLJ11749-like (similar to vacuolar protein sorting protein 20)	
5 NAC alpha chain	
96 sec5 homolog	
2462 clathrin, heavy chain	
1058 sec1 homolog	
4153 signal recognition particle, 68kDa	
1520 enthoproin	
1872 cotamer alpha	
526 cotamer zeta1	
472 VIP36-like	
2285 syntaxin binding protein 3	
2499A vacuolar sorting protein 18	
<b>proteolytic synthesis</b>	
3583 Dolichyl-diphosphooligosaccharide-protein glycosyltransferase	
3378 UDP-glucuronic acid/UDP-N-acetylglactosamine dual transporter	
<b>Protein folding</b>	
138 heatshock cognate 70	
383A TCP-1 gamma	
642 TCP-1 beta	
800A TCP-1 eta	
1743 immunoglobulin binding protein	
2473 mortalin-2	
2972B TCP-1 epsilon	
491 TCP-1 alpha	
2062 odc37-like	
<b>Protein degradation</b>	
3593 26S protease regulatory subunit S10B	
3970 SKP1A: S-phase kinase-associated protein 1A	
2939 proteasome beta-subunit C5	
3471 ubiquitin fusion degradation 1 protein	
<b>Mitochondria</b>	
3619 ATP synthase oligomycin sensitivity conferral protein	
887 translocase of inner mitochondrial membrane 23	
526A mitochondrial adp/atap translocator	
<b>Metabolism</b>	
688 ribonucleoside reductase R2	
2694 CAD trifunctional protein	
3510 thymidylate synthase	
1433 adenylosuccinate synthetase	
1903C dihydrofolate reductase 5-succinyltransferase	
559 phosphatidylinositol synthase	
318 ribonucleoside reductase R1	
1841B ornithine decarboxylase	
<b>vATPase</b>	
923 v-ATPase SDF/54kD subunit	
1207 v-ATPase 16 kDa proteolipid subunit	
1988 v-ATPase subunit F	
2188B v-ATPase AC39 subunit	
577a v-ATPase subunit e	
<b>Signal transduction</b>	
3174A pleiotropic regulator 1	
2335 phospholipase C gamma 1	
1002 casein kinase I alpha	

<b>Miscellaneous</b>	
4049 DEAD/H box 16 RNA helicase	
4 dead eye/NIC96-like	
1727 myc-regulated DEAD/H box 18 RNA helicase	
2453 ATP-binding cassette, sub-family E, member 1 (RNase L inhibitor)	
1055A MAK16-like	
1579 nuclear cap binding protein subunit 2	
2452 RNA 3'-terminal phosphate cyclase-like protein	
3282 zinc finger protein 162	
3351 byerin-like	
3013 zinc finger protein 183-like 1	
1552 myb binding protein (P160) 1a-like	
1793 ribosomal protein L7-like	
3625 EBNA1 binding protein 2-like	
340b exosome complex exonuclease RRP41	
1532A WD repeat protein 1	
1858B F-actin capping protein beta subunit	
3721 DNA methyltransferase 1-associated protein 1	
3785 gamma tubulin	
4063B DKFZp596G0123-like (possibly calcium binding transporter)	
<b>Unknown Function</b>	
1034A novel (G10)	
2667 novel (FLJ10290-like)	
2865 novel (RIKEN cDNA 241000417-like)	
447 novel (dentless homolog)	
297 novel (FLJ10498-like)	
932 novel (BAP28-like)	
1370 novel (FLJ12787-like)	
1451 novel (BING4-like)	
2806 novel (FLJ14075-like)	
2404 novel (AKO22581-like)	
821A novel (KIAA0007-like)	
2993 novel (my038-like)	
1241 novel (FLJ13912-like)	
1373 novel (MGC1346-like)	
2345 novel (GTP binding protein NGB)	
2937 novel (CGI-94-like)	
557 novel (C7orf114-like)	
526b novel (FLJ31121-like)	
1437 novel (FLJ10613-like)	
2573A novel (twainb-like)	
2155 novel (FLJ12945-like)	
2246 novel (FLJ11294-like)	
3630A novel (TA-WDRP-like)	
4201B novel (cdc39-like)	
299 novel (nuclcostatin-like)	
239 novel (FLJ14909-like)	
3512B novel (BC-2-like)	
1257 novel (AKO27570-like)	
2329A novel (KIAA0682-like)	
1116A novel (DT1P1A10-like)	
3400 novel (C20orf77-like)	
2735A novel (MGC10433-like)	
2786B novel (AKO25355-like)	
3773 novel (DKFZP564C186-like)	
2914 novel (RIKEN cDNA 943007D24-like)	
1482 novel (pinball wizard, WRB -like)	
3812 novel (AN11 homolog)	
429 novel (DJ434C14.5-like)	
1858A novel (FLJ20291-like)	
3305 novel (wbscr21-like)	
<b>IN NON-FUNGAL SINGLE CELL ORGANISMS</b>	
<b>Intraflagellar transport</b>	
409 IFT81 / camille deficiency-associated, expressed in ventricle 1	
2211 IFT172 / LIM-interacting protein	
3417 IFT57 / HIPPI	
<b>Unknown Function</b>	
1477 novel (FLJ2626-like)	
3764 novel (C21orf59-like)	
1392 MGC35261-like	
3308 novel (leucis specific leucine rich repeat protein-like)	
3959A novel (KIAA0643-like)	
<b>ANIMAL-SPECIFIC (WORM AND FLY)</b>	
<b>Cell cycle</b>	
3820A epithelial cell transforming sequence 2 oncogene	
1326 survivin	
2696 cyclin A2	
<b>Transcription</b>	
<b>chromatin remodeling</b>	
1072 baff3a (swi/snf complex component)	
<b>transcription factor</b>	
349 pou2	
933 pbx4	
1134 sox9a	
2933 eyes absent 1A	
1321 fox1	
2092 caudal	
1059 neurogenin-1	
2109 ttx16/spadetail	
<b>transcription factor regulator</b>	
1715 aryl hydrocarbon receptor nuclear translocator 2A	
<b>RNA processing</b>	
<b>splicing</b>	
1779 proline/glutamine rich splicing factor	
<b>Post-translation</b>	
<b>proteoglycan synthesis</b>	
307 glucuronosyl-transferase	
954 UDP-glucuronic acid decarboxylase	
1516 xylosylprotein beta 1,4-galactosyltransferase 7	

<b>Metabolism</b>	
2688 multifunctional protein ADE2	
<b>vATPase-associated</b>	
112 v-ATPase Acl5	
3681 v-ATPase membrane sector associated protein M8-9	
<b>Transporters and Channels</b>	
3475 NAK ATPase subunit 1 (ATP1A1A.1)	
801 spirater homolog	
4166 PKD2	
<b>Cell-cell communication</b>	
1389 N-cadherin	
1688 glypican-6	
1780B wnt5	
975 nodal-related 1/squnt	
2718A nodal-related 2/cyclops	
2557 laminin C1	
1113B laminin B1	
3263 notch1a	
<b>Signal transduction</b>	
2039A synembryn-like	
2226 nagie oko (MAGUK family)	
3208 protein kinase C lambda	
<b>Miscellaneous</b>	
2820 clasp1	
286 MBD2-interacting zinc finger	
3332 NF2a	
2653 myosin-binding subunit of myosin phosphatase	
1463 CD36-like 2	
1230 actopaxin	
<b>Unknown Function</b>	
591 novel (ars2-like)	
1594 novel (DKFZP434B168-like)	
1518 LCMR1-like	
2729B novel (hepatocellular carcinoma-associated antigen 127-like)	
1073 novel (KIAA0877-like)	
1532B novel (FLJ12716-like)	
<b>ANIMAL-SPECIFIC (FLY, NOT WORM)</b>	
<b>Transcription</b>	
<b>chromatin remodeling</b>	
2639B Sds3 (Suppressor of Defective Silencing 3)	
<b>transcription cofactor</b>	
2639A TAF1B/RNA polymerase I, B	
<b>transcription factor</b>	
2561 B-myb	
399 NRF-1	
<b>transcription factor regulator</b>	
1487 site-1 protease of SREBP	
<b>Protein degradation</b>	
904 mindbomb (E3 ubiquitin ligase)	
<b>Cell-cell communication</b>	
2293 chordino	
229 smoothend	
781 deltaA	
<b>Miscellaneous</b>	
2026A inhibitor of apoptosis protein with a caspase recruitment domain	
2737 selenoprotein H	
<b>Unknown Function</b>	
428 novel (AK000411)	
<b>VERTEBRATE-SPECIFIC</b>	
<b>Transcription</b>	
<b>transcription factor</b>	
272 nuclear protein NP95	
548 vHNF-1	
<b>Protein degradation</b>	
2648 Fbox protein 5	
<b>Miscellaneous</b>	
1182 telomere binding protein 2	
1011 Opa-interacting protein 2 (3' exoribonuclease family)	
1722 PCD7/ES18-like	
2091 RING-HBOXX zinc finger protein	
340A claudin j	
2217 serine protease inhibitor HGFAI	
<b>Unknown Function</b>	
3662 novel (D13S106E-like)	
3505 GJ383J.3-like	
1573 novel (RIKEN cDNA 5730590G19-like)	
3634 BM039-like	
1262 SIL (TAL1-interrupting locus-like)	
81 PRDM5/PFM2	
572 novel (FLJ20508-like)	
994 novel (FLJ22611-like)	
2151 novel (pan-epithelial glycoprotein)	
296a novel (MGC4607-like)	
459 novel (arf2-like)	
<b>FISH-SPECIFIC</b>	
<b>Unknown Function</b>	
37 novel, fish specific	
43 novel, fish specific	

**Fig. 1. Genes essential for zebrafish embryonic development identified by insertional mutagenesis. Genes are listed by mutant number and sorted by evolutionary conservation and gene function. Phenotypic descriptions are available in Table 3, and images are available at [http://web.mit.edu/ccr/pnas.zebrafish\\_mutant\\_images/index.html](http://web.mit.edu/ccr/pnas.zebrafish_mutant_images/index.html).**



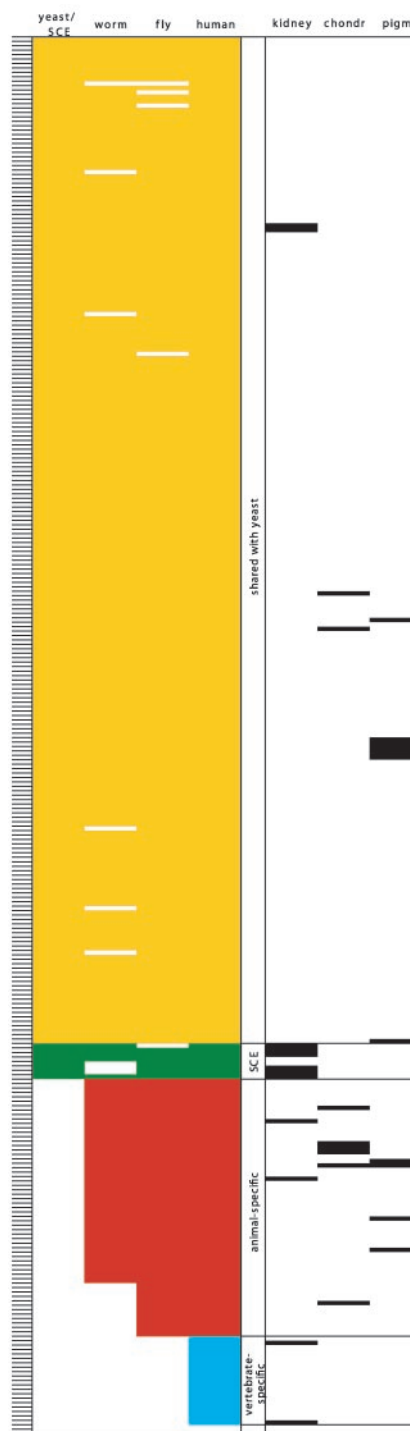
**Fig. 2.** Types of genes whose mutation in zebrafish leads to an embryonic visible phenotype. The genes are assigned to the same categories as in Fig. 1 although some categories have been combined.

genes would be found in a saturation screen in the fish. Three of these four genes are animal specific. In *C. elegans*, mutations in homologues of these four genes led to a “squashed vulva” phenotype (18, 19). Thus, whereas the genes are presumably required for proteoglycan synthesis in both worm and fish, they are involved in the formation and structural integrity of very different body parts in the two organisms.

Extending our earlier observations (2), 9 of 11 genes we identified as required for normal melanocyte pigmentation in fish encode v-ATPase subunits or associated proteins or proteins otherwise involved in intracellular vesicles. Whereas the main v-ATPase subunits are found in yeast, zebrafish required several animal-specific v-ATPase-associated proteins in addition.

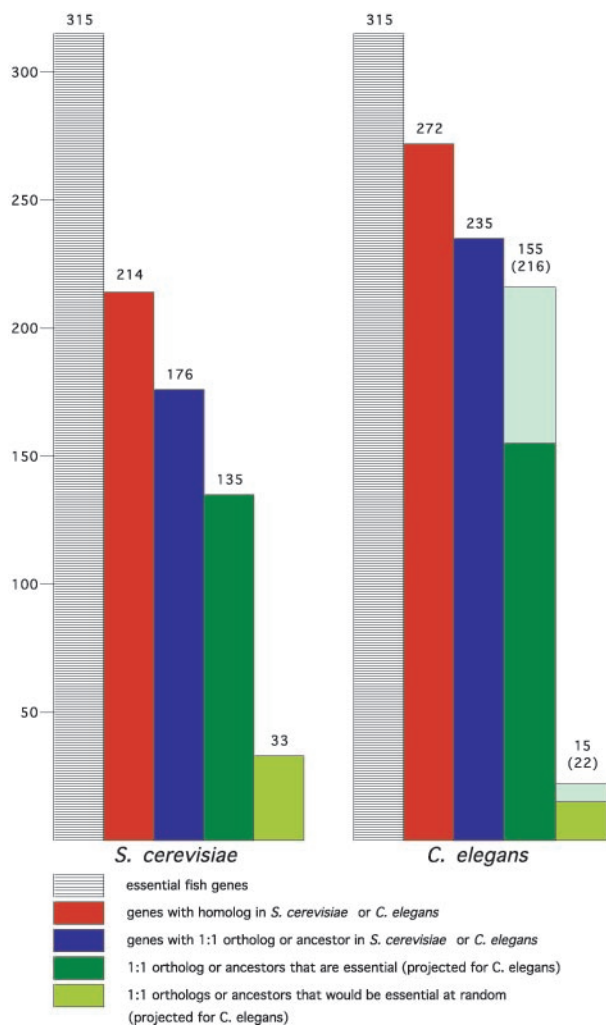
**The Trait of Being a Genetically Essential Gene Is Conserved in Evolution.** As Fig. 3 above reveals, and has been observed in other species (12, 13), genes that are essential in zebrafish are highly conserved in evolution. We next asked whether genes that are genetically essential in zebrafish are also genetically essential in other species. It is convenient to ask this question of genes that have counterparts in yeast and worm because a large fraction of the annotated genes in these species have been deleted [yeast (12, 20)] or knocked down by using RNA interference (RNAi) [worm (13, 21)]. To make functional comparisons of genes across these species, however, it is necessary first to identify the yeast and worm genes that are not merely homologues of the essential fish genes, but that have orthologous relationships to them, because orthologues are more likely to perform the same function in their respective species. We determined which zebrafish genes have clear orthologous relationships with their homologues in *S. cerevisiae* and *C. elegans* by comparative orthologous group (COG) analysis (5) (see *Materials and Methods*). Because the zebrafish genome is not yet fully sequenced and annotated, it cannot be used for this purpose, so we used the human genome as a “surrogate vertebrate” genome. We considered both genes that were 1:1 orthologues and genes for which a single yeast or worm gene is an “ancestor” of two or more human paralogues. The results of the analysis are shown in Fig. 4 for yeast (*Left*) and for worm (*Right*).

The left-most bars in Fig. 4 *Left* and *Right* represent the list of 315 essential fish genes, the red columns next to them show the number of these genes that have homologues in yeast or worm, and the blue columns show the fraction of homologues that have an orthologous relationship to the vertebrate genes. Having identified the yeast and worm orthologues, we then asked which of these genes are essential in their respective species. The results are shown by the dark green columns in Fig. 4.



**Fig. 3.** Evolutionary conservation of essential zebrafish genes. Horizontal lines on the *Left* represent 315 different genes. The genes are listed in the same order as in Fig. 1. In the first four colored columns, the presence of a colored box indicates the presence of one or more homologous genes (BLASTP E value of  $<10^{-5}$ ) in yeast (either *S. cerevisiae* or *S. pombe*), *C. elegans*, *D. melanogaster*, or *H. sapiens*. Green boxes indicate genes without homologues in yeast but with homologues in other SCEs such as *Giardia*, *Plasmodium*, *Trypanosoma*, and/or *Chlamydomonas*. Thus, yellow boxes represent genes conserved through yeast, green are those found in SCEs other than yeast, red are those found in invertebrates but not SCEs, and blue are vertebrate-specific. The last two genes (with no colored boxes) seem to be fish-specific. The black boxes in the last three columns indicate genes whose mutation leads to one of three phenotypes: cystic kidney, chondrogenesis defects, or reduction or lack of melanocyte pigmentation.





**Fig. 4.** Essentialness of genes is evolutionarily conserved. Shown is an analysis of *S. cerevisiae* (Left) and *C. elegans* (Right) genes that are homologous to the essential fish genes. In each case, the leftmost columns represent 315 essential fish genes, the red columns show how many of these have homologues in yeast (214) or worm (272), and the blue columns show which have a 1:1 orthologue or “ancestor” gene in yeast (176) or worm (235). The dark green columns represent the number of these yeast or worm orthologues that are essential in their respective species, 135 of the 176 yeast genes, and 155 of the 235 worm genes. The pea green columns show the number that would be predicted to be essential at random, 33 of the 176 yeast genes, and 15 of the 235 worm genes. Thus, the difference between the dark green and pea green columns is the enrichment. In the case of the worm, the pale green extensions to the two green columns represent projections of how many of the orthologues would be found to be essential if 100% of the worm genes had been successfully knocked down by RNAi; this estimate prorates both for the reported failure rate of RNAi and the number of genes for which no RNAi data have been reported. This calculation estimates that 216 of the 235 worm genes are likely to be essential whereas only 22 would be expected to be at random. Note that the percentage of worm orthologues that are essential that is stated in the text does not include the genes for which no RNAi data have been reported.

In yeast, only 19% of protein-coding genes are essential for cell viability under optimal growth conditions, despite the fact that most “nonessential” yeast genes seem to be single-copy (12). In contrast, among the yeast genes that are 1:1 orthologues or ancestors of the vertebrate genes in our study, 77% are essential in yeast. This is a 4-fold enrichment relative to the yeast genome at large (compare the number observed in dark green with the number predicted if random in pea green).

In *C. elegans*, RNAi analysis of protein-coding genes reveals that only 7% are required for embryo viability, an additional 1.5% for other developmental and/or physiological processes, and another 1.5% for wild-type growth (21). Yet among the worm genes that are orthologous to the vertebrate genes we identified and for which RNAi data have been reported, 72% are required for embryonic viability and a further 6% are associated with postembryonic developmental phenotypes, a total of 78%. This is an enrichment of nearly 10-fold over the worm genome at large. In the large RNAi screens in *C. elegans* that have been published to date, it was estimated that only 78% of embryonic-essential genes were detected, largely due to inefficiencies in the RNAi technology (21). Correcting for this failure rate, we conclude that nearly all of the worm orthologues of essential fish genes might in fact be essential for embryonic development (see lighter colored extensions on the green columns in Fig. 4 Right). In summary, this analysis reveals that genetically essential genes have a strong tendency to retain this special status through evolution from yeast to vertebrates.

## Discussion

We have described the completion of a large genetic screen in zebrafish, the isolation of insertional mutations in  $\approx 25\%$  of the embryonic-essential genes of the fish, and the molecular cloning of  $\approx 22\%$  of all such genes. This collection of mutants will be a valuable resource for the study of many cellular and developmental processes in a vertebrate. Many of the genes we identified are probably required for cell viability, others for more specific developmental processes including patterning, differentiation or physiology. Twenty percent of the genes encode proteins that have no known or sufficiently clearly identifiable biochemical function. About three quarters of the genes we cloned have homologues in yeast or other single-celled organisms, and  $\approx 25\%$  are animal-specific with 7% overall being vertebrate-specific.

Our results imply that there are only  $\approx 1,400$  genes that when mutated result in a visible, usually lethal phenotype in the zebrafish embryo and 5-day-old larva. This number is fewer than the 2,400 such genes proposed by Haffter *et al.* (7) from data obtained in a large chemical mutagenesis screen by using ENU as the mutagen. We do not think the discrepancy is due to the inability of retroviral vectors to target genes than can be mutated by ENU because genes mutated by ENU and cloned by positional or candidate gene cloning were mutated at the same efficiency as genes encoding ribosomal proteins or tRNA synthetases in our screen. Rather, it seems likely that the discrepancy reflects inaccuracies in both calculations. The failure to achieve saturation in either screen and the fact that the data for single and multiple hits do not fit the Poisson distribution precisely in either screen make highly accurate calculations impossible.

The phenotypic descriptions of most of our mutants remain preliminary, and re-screens of the collection are needed to identify the specific defects in most mutants. The first three such re-screens to be completed, which we summarized here, are revealing in this respect. The screen for cystic kidney in particular supports our conclusion that our mutant collection contains 25% of the genes essential for diverse developmental processes in the embryo. Of 12 genes identified in the kidney cyst screen, 7 seemed to be novel when first cloned, whereas another is PKD2, a gene known to be mutated in human polycystic kidney disease (22). In humans, cystic kidney disease results from a failure of epithelial cells in kidney tubes and ducts to differentiate properly and to cease dividing. This defect can result from defects in primary cilia located on the epithelial cells of kidney tubules and ducts. When we gained access to the mostly unpublished sequences of 13 genes that encode IFT proteins in *Chlamydomonas*, and which are required for flagellum formation or function in that species, we found that 3 of our 7 “novel”

genes were in fact fish homologues of 3 of the 13 IFT genes. It is possible that others of the “novel” genes identified in the fish kidney cyst screen will prove to be yet unidentified IFT genes, or other genes involved in a pathway linking cilia to cell differentiation and cell division in vertebrates. The finding that we have mutations in 3 of 13 IFT genes whose sequences are available is consistent with the notion that our mutant collection includes 25% of the genes essential for processes involved in embryonic development. The results of the kidney cyst screen also suggest that genetic screens in vertebrate animals can reveal many of the genes in a pathway and argue that redundancy may be no more of a problem in genetic screens in vertebrates than it has been in genetic screens in invertebrate animals.

The data from the three shelf screens described here, along with the data showing the conservation of the fish genes in the genomes of single-celled eukaryotes and invertebrate animals, show how many of the same genes used in SCEs and invertebrates are used for different biological tasks in vertebrate animals although their cellular functions are similar in the different species. Thus, IFT genes used by *Chlamydomonas* to make motile flagella are used to make cilia with different sensory roles in different organisms, for example, on chemosensory neurons sensing salt concentrations in the worm (23) and on kidney epithelial cells in vertebrates (24). Similarly, the same proteoglycan synthesis enzymes that the worm uses to make a properly shaped vulva (18, 19) are used in the fish to make a vertebrate-specific structure, cartilage. And the production of melanin pigment in the fish, which takes place in acidic subcellular compartments (25), requires the same genes to acidify these compartments that yeast cells use to acidify vacuoles (26).

The fact that there is such a small number of embryonic-essential genes and that they include genes that comprise coherent genetic pathways of development suggests that the

genetically essential genes have a unique status in biological processes. Consistent with this possibility, we found that the yeast or worm orthologues of genes that are essential in fish have a high probability of also being essential in these species. Thus genes that can be detected in genetic screens, and in particular those that are essential for early viability, seem to have retained this special status through evolution. The implications of this observation are not known, but we suggest that these genes may be all or most of the genes that are absolutely required for many biological processes whereas most other genes may serve to assist these critical genes in making biological processes more robust. Evolution may have required that the number of genetically essential genes be small and that they remain the same genes.

By identifying about a quarter of the embryonic lethal genes in the zebrafish, our studies provide strong evidence that genetic screens in vertebrate animals, as in invertebrates, can reveal molecular pathways of development. Whereas redundancy may hide some genes from this analysis, the fact that the genes that we do find for particular phenotypes can define a biochemical complex or pathway suggests that forward genetic screens can indeed identify many if not most of the genes that are key players in specifying complex developmental processes in vertebrates.

We thank E. Maldonado and J. Davenport for assistance in screening for mutants, G. Golling and K. Anderson for help with gene identification, Maryann Haldi for phenotypic documentation, M. Cunningham for maintenance of the mutant lines of fish, and T. Angelini, C. Doller, and S. Farrington for maintenance of our zebrafish colony. We thank Bob Bosselman for his interest and support. This work was supported by a National Institutes of Health grant from the National Center for Research Resources (to N.H. and A.A.) and a grant from Amgen (to N.H.). Additional funding came from the Ford Foundation (to N.H.). R.M.N. was supported by a postdoctoral fellowship from the National Institutes of Health.

1. Amsterdam, A., Burgess, S., Golling, G., Chen, W., Sun, Z., Townsend, K., Farrington, S., Haldi, M. & Hopkins, N. (1999) *Genes Dev.* **13**, 2713–2724.
2. Golling, G., Amsterdam, A., Sun, Z., Antonelli, M., Maldonado, E., Chen, W., Burgess, S., Haldi, M., Artzt, K., Farrington, S., *et al.* (2002) *Nat. Genet.* **31**, 135–140.
3. Amsterdam, A. (2003) *Dev. Dyn.* **228**, 523–534.
4. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402.
5. Tatusov, R. L., Koonin, E. V. & Lipman D. J. (1997) *Science* **278**, 631–637.
6. Meneely, P. M. & Herman R. K. (1979) *Genetics* **92**, 99–115.
7. Haffter, P., Granato, M., Brand, M., Mullins, M. C., Hammerschmidt, M., Kane, D. A., Odenthal, J., van Eeden, F. J. M., Jiang, Y.-J., Heisenberg, C.-P., *et al.* (1996) *Development (Cambridge, U.K.)* **123**, 1–36.
8. Kane, D. A., Hammerschmidt, M., Mullins, M. C., Maischein, H.-M., Brand, M., van Eeden, F. J. M., Furutani-Seiki, M., Granato, M., Haffter, P., Heisenberg, C.-P., *et al.* (1996) *Development (Cambridge, U.K.)* **123**, 57–66.
9. Wienholds, E., Koudijs, M. J., van Eeden, F. J., Cuppen, E. & Plasterk R. H. (2003) *Nat. Genet.* **35**, 217–218.
10. International Human Genome Sequencing Consortium (2001) *Nature* **409**, 860–921.
11. Mouse Genome Sequencing Consortium (2002) *Nature* **420**, 520–562.
12. Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucan-Danila, A., Anderson, K., Andre, B., *et al.* (2002) *Nature* **418**, 387–391.
13. Fraser, A. G., Kamath, R. S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M. & Ahringer, J. (2000) *Nature* **408**, 325–330.
14. Driever, W., Solnica-Krezel, L., Schier, A. F., Neuhauss, S. C. F., Malicki, J., Stemple, D. L., Stainier, D. Y. R., Zwartkruis, F., Abdelilah, S., Rangini, Z., *et al.* (1996) *Development (Cambridge, U.K.)* **123**, 37–46.
15. Sun, Z., Amsterdam, A., Pazour, G. J., Cole, D. G., Miller, M. S. & Hopkins, N. (2004) *Development (Cambridge, U.K.)*, in press.
16. Rosenbaum, J. L. & Witman, G. B. (2002) *Nat. Rev. Mol. Cell Biol.* **3**, 815–825.
17. Schwartz, N. B. & Domsicz, M. (2002) *Glycobiology* **12**, 57R–68R.
18. Herman, T. & Horvitz, H. R. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 974–979.
19. Hwang, H.-Y., Olson, S. K., Esko, J. D. & Horvitz H. R. (2003) *Nature* **423**, 439–443.
20. Winzler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J. D., Bussey, H., *et al.* (1999) *Science* **285**, 901–906.
21. Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., *et al.* (2003) *Nature* **421**, 231–237.
22. Mochizuki, T., Wu, G., Hayashi, T., Xenophontos, S. L., Veldhuisen, B., Saris, J. J., Reynolds, D. M., Cai, Y., Gabow P. A., Pierides, A., *et al.* (1996) *Science* **272**, 1339–1342.
23. Cole, D. G., Diener, D. R., Himelblau, A. L., Beech, P. L., Fuster, J. C. & Rosenbaum, J. L. (1998) *J. Cell Biol.* **141**, 993–1008.
24. Nauli, S. M., Alenghat, F. J., Luo, Y., Williams, E., Vassilev, P., Li, X., Elia, A. E., Lu, W., Brown, E. M., Quinn, S. J., *et al.* (2003) *Nat. Genet.* **33**, 129–137.
25. Bhatnagar, V., Anjiah, S., Puri, N., Darshanam, B. N. & Ramaiah, A. (1993) *Arch. Biochem. Biophys.* **307**, 183–192.
26. Stevens, T. H. & Forgac M. (1997) *Annu. Rev. Cell Dev. Biol.* **13**, 779–808.