Insertional Mutagenesis in Zebrafish Using a Pseudotyped Retroviral Vector

by

Nicholas R. Gaiano

B.S., Biochemistry
Cornell University, 1991

Submitted to the Department of Biology in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

January 1997

© 1997 Massachusetts Institute of Technology
All rights reserved

Signature of Author: ______________________________

Department of Biology
January 14, 1997

Certified by: ______________________________

Nancy Hopkins
Professor of Biology
Thesis Supervisor

Accepted by. ______________________________

Richard A. Young
Professor of Biology
Chairman, Graduate Committee

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

FEB 28 1997

Science

LIBRARIES

Insertional Mutagenesis in Zebrafish Using a Pseudotyped Retroviral Vector

by

Nicholas R. Gaiano

Submitted to the Department of Biology on January 24, 1997
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy in Biology

ABSTRACT

Large-scale chemical mutagenesis screens have recently been performed in zebrafish and have isolated thousands of mutations affecting processes ranging from epiboly and gastrulation to organogenesis and behavior. However, because these lesions are point mutations, the cloning of the mutated genes is likely to be difficult. With this issue in mind, we undertook the development of an insertional mutagenesis methodology for use in zebrafish. Insertional mutagenesis has been shown to be an effective way to mutate and rapidly clone genes in a wide variety of organisms including *D. melanogaster*, *C. elegans*, and mice.

First, a pseudotyped retroviral vector, composed of a nucleocapsid based on the Moloney murine leukemia virus (MoMLV) and an envelope derived from the vesicular stomatitis virus (VSV), was injected into blastula-stage zebrafish embryos and shown to be capable of generating transgenic offspring. In an effort to improve upon the obtained transgenic frequency, two new MLV/VSV vectors, prepared to titers roughly 100-fold higher than that of the previously used virus, were injected into embryos. For one of these, injected fish transmitted, on average, 11 different insertions to 30% of their F1 progeny. At this frequency, it is feasible for a single lab to generate tens to hundreds of thousands of proviral insertions. In a pilot insertional mutagenesis screen of 217 insertions, 3 insertional mutations were isolated, and in two cases the disrupted gene was rapidly cloned. One of these encodes a putative endoribonuclease essential for normal pharyngeal arch development, and the other encodes a novel protein necessary for embryonic maturation. As it is easy to generate many thousands of proviral transgenes in zebrafish, it should now be possible to use this screening method to mutate and then rapidly clone hundreds of gene affecting vertebrate developmental and cellular processes.

Thesis Supervisor: Nancy Hopkins
Title: Professor of Biology

I dedicate this thesis to my
mother and to my grandmother.

ACKNOWLEDGMENTS

I am heavily indebted to the members of the Hopkins lab with whom I have worked over the past four and a half years to develop the method described here. In particular, I thank Shuo Lin, Adam Amsterdam, Miguel Allende, Koichi Kawakami, and Tom Becker. I thank the other members of the Hopkins lab for providing a pleasant environment in which to work and for their frequent words of wisdom. In addition, I thank the fishroom caretakers, especially Caroline McMurtrie and Dean Thompson, who have kept the fish alive and healthy for our use. I thank Richard Mulligan and his lab, especially David Sanders and Michael Sadelain for helpful discussions regarding the manipulation of retroviral vectors, and Pat LaPorte for advice on preparing pseudotyped viral stocks. I would also like to thank Rudolf Jaenisch and Hazel Sive, both of whom have provided valuable suggestions during the course of this work. I thank my parents and Michelle Starz for emotional support. Finally, I would most like to thank my advisor, Nancy Hopkins, who has taught me how to think about science both critically and creatively, and who has provided me with a great deal of insight and guidance throughout the years.

TABLE OF CONTENTS

Abstract .... 2.

Acknowledgments .... 4.

Chapter 1: Introduction: Insertional Mutagenesis in the Study of Animal Developmental Genetics .... 7.
1. The Study of Animal Development .... 8.
2. Vertebrate Model Systems .... 10.
3. Large-Scale Chemical Mutagenesis Screens in Zebrafish .... 12.
4. Insertional Mutagenesis in Invertebrates .... 15.
5. Insertional Mutagenesis in Mice .... 18.
6. Transgenesis in Zebrafish .... 24.
7. MLV/VSV Pseudotyped Retroviral Vectors .... 25.
8. Retroviral Insertional Mutagenesis in Zebrafish .... 29.
9. References .... 29.

Chapter 2: Integration and Germ-Line Transmission of a Pseudotyped Retroviral Vector in Zebrafish .... 39.
1. Abstract .... 40.
2. Introduction .... 41.
3. Results .... 42.
4. Discussion .... 51.
5. Materials and Methods .... 52.
6. References .... 54.

Chapter 3: Highly Efficient Germ-Line Transmission of Proviral Insertions in Zebrafish .... 56.
1. Abstract .... 57.
2. Introduction .... 58.
3. Results .... 59.
   a. High frequency germ-line transmission of proviral integrations .... 59.
   b. Germ-line mosaicism of founders .... 63.
   c. SFG founders transmit an average of 11 proviral insertions to their F1 progeny .... 63.
   d. Analysis of SFG proviral genome structure .... 65.
   e. Analysis of genomic sequences at the sites of SFG insertions .... 68.
4. Discussion .... 71.
5. Material and Methods .... 74.
6. References .... 77.

Chapter 4: Insertional Mutagenesis and Rapid Cloning of Essential Genes in Zebrafish........ 80.
1. Abstract........ 81.
2. Introduction........ 82.
3. Results and Discussion........ 83.
a. Retroviral insertional mutagenesis in zebrafish........ 83.
b. The *no arches* mutant phenotype........ 91.
c. The *no arches* gene product........ 97.
4. Summary........ 104.
5. Material and Methods........ 104.
6. References........ 105.

Chapter 5: Insertional Mutagenesis in Zebrafish Identifies a Novel Gene, *pescadillo*, Which is Essential for Embryonic Development........ 109.
1. Abstract........ 110.
2. Introduction........ 111.
3. Results........ 112.
a. The *pes* gene encodes a highly conserved novel protein........ 113.
b. Expression of the *pes* gene is developmentally regulated........ 116.
c. The *pes* mutation affects the development of a subset of embryonic primordia which correlate with sites of strong *pes* gene expression........ 121.
4. Discussion........ 128.
5. Materials and Methods........ 130.
6. References........ 132.

Chapter 6: Future Prospects for the Use of Retroviral Insertional Mutagenesis in Zebrafish........ 135.
1. Randomness of Retroviral Integration........ 136.
2. The Zebrafish Genome Project........ 141.
3. Insertional Screening Strategies in Zebrafish........ 143.
4. Potential Improvements in Methodology........ 146.
5. Insertional Mutagenesis Beyond Developmental Biology........ 151.
6. Identifying Essential Genes is Only the Beginning........ 153.
7. References........ 155.

CHAPTER 1

INTRODUCTION: INSERTIONAL MUTAGENESIS IN THE STUDY OF ANIMAL DEVELOPMENTAL GENETICS

The Study of Animal Development

Development from a fertilized egg to a mature organism requires the exquisite coordination of many biological processes which can be broadly categorized as pattern formation, cell growth and differentiation, and morphogenesis. These processes include the asymmetric distribution of RNA and protein determinants, signaling between cells to communicate position and specify fate, the migration of cells over large distances, and the regulation of both cell proliferation and cell death. Consequently, the study of development is a highly diverse area of biological research addressing issues ranging from macromolecular structure, to organogenesis, to ever appropriately the evolution of life on earth.

Two primary experimental approaches have been undertaken to examine both vertebrate and invertebrate development: embryological studies and genetic analysis. Embryological studies utilize the manipulation and observation of embryonic cells and tissues to characterize their behavior and potential. Examples of such studies include the characterization of cell lineage during nematode development (1), the transplantation of neural tissue in birds to examine cell fate (2), and the treatment of frog embryos with various factors to examine the nature of cell-fate specification (3). In contrast, genetic studies focus on identifying and understanding the function of genes which control embryological phenomena. Some such studies have included the isolation of mutations which affect pattern formation in the fruit fly (4), the assessment of gene function in mice by targeted disruption (5), and the ectopic expression of genes with putative instructive roles during development (6, 7).

While both embryological and genetic studies have contributed to our understanding of animal development, currently the cellular and genetic mechanisms controlling this process are most thoroughly understood in the invertebrates *Drosophila melanogaster* and *Caenorhabditis elegans*. These

organisms have been popular model systems because they are suited to large-scale genetic screens which permit the mutation and identification of many biologically interesting genes (4, 8). The ability to perform "saturation" mutagenesis screens in these systems has permitted the isolation of mutations in most if not all of the genes necessary during a chosen process of interest. The generation of such mutant collections has been used to thoroughly dissect the genetic pathways in question, and the subsequent cloning of the mutated genes has then allowed the integration of molecular data regarding their gene products with genetic data which characterizes their functions.

Many genes with putative or demonstrated roles during vertebrate development have been identified by virtue of their homology to genes isolated through mutagenesis in *Drosophila* (9-12). Such genes include those in the *hox*, *pax*, and *wnt* families (10, 11), and the neurogenic genes *Notch*, *Delta*, and *numb* (12), among many others. Analyses of the expression patterns of these genes in vertebrates as well as functional studies have suggested many may have conserved functions during both vertebrate and invertebrate development. This conservation along with the versatility of invertebrate experimental systems has provided a great deal of insight into vertebrate development.

While the degree of conservation between vertebrate and invertebrate developmental mechanisms is impressive, there are fundamental differences between these types of organisms. For example, early in fruit fly development the embryo is a multi-nucleated syncytium, while early vertebrate embryos contain predominantly single-nucleated cells (see 13). Another example is that while many of the genes which control neural development exist and may function similarly in both vertebrate and invertebrate, the cellular behavior of neural progenitors is quite different in these organisms (12). Furthermore, the neural crest, which is fundamental to the development of the vertebrate head, among other things, has

no known counterpart in invertebrates. The divergence between vertebrates and invertebrates indicates that a thorough understanding of vertebrate development will require the study of vertebrates directly.

Vertebrate Model Systems

The study of vertebrate development has been undertaken predominantly in four model system: the mouse, the frog, the chicken, and more recently the fish. These systems each have strengths and limitations. The mouse is useful for genetic studies but embryological observation and manipulation are difficult. In contrast, the frog and the chicken are well-suited for embryological studies but not for genetic analysis. The zebrafish is unique among these systems because it is amenable to both embryological studies and genetic analysis.

Embryological studies in amphibia and birds have focused primarily on understanding cell lineage, cell movement, and the nature of the signals which pattern the embryo. For example, early work by Spemann and Mangold showed that transplantation of the dorsal lip of a gastrulating newt embryo to an ectopic location could induce the surrounding tissue to participate in the formation of an embryonic axis (14). This work suggested that signals present in the transplanted tissue could change the developmental fate of adjacent cells, a result which inspired many similar experiments and has subsequently led to the formation of induction-based models of vertebrate development (3, 15).

While embryological studies such as these have been invaluable for characterizing the cellular nature of developmental processes, they have been limited by difficulties in identifying the genes which control those processes. As indicated above, many genes important for vertebrate development have been identified by homology to *Drosophila* genes. In addition, vertebrate genes such as *myoD, Vg1, noggin,* and *cerberus,* have been identified using activity-based selection

assays and subtractive strategies (16-19). The role of such genes has been addressed by studying the effect of their gain-of-function and loss-of-function forms on vertebrate development. Gain-of-function studies have utilized the over-expression of wild-type or constitutively active forms of the genes in question (20, 21), while loss-of-function studies have utilized expression of dominant negative forms of these genes (22, 23) or homologous recombination-mediated gene "knockouts" in mice (5). These efforts have proven informative although they are limited. Gene expression patterns as well as the effects of over-expression can be misleading regarding the endogenous biological function of genes of interest. Furthermore, the targeted disruption of genes in mice is limited to genes that have already been identified, and the costs of such efforts limits the large-scale application of this approach, at least within single laboratories.

An attractive alternative to these approaches would be to perform large-scale 'forward' mutagenesis screens in a vertebrate. During forward mutagenesis screening the organism of choice is treated with a mutagen and then bred to reveal mutations in genes with required functions during a process of interest. Genetic studies can then provide clues regarding the role of these mutated genes, and the ultimate cloning of such genes permits molecular characterization of the processes in question. As indicated above, such screens in invertebrates have proven extremely powerful for identifying biologically-interesting genes. Historically, however, large-scale forward screening in vertebrates has been precluded by the lack of a vertebrate model organism which was well-suited for both mutagenesis and embryonic screening. While mutations in the mouse can be induced using chemical or insertional mutagenesis (see below), the costs of maintaining large numbers of mice, in conjunction with the fact that embryonic lethal mutations cannot be easily scored, makes large-scale screening in mice unfeasible. Such screens are also not feasible in frogs or chickens for reasons including the generation time,

ploidy, and size of these animals. In recent years, however, the zebrafish has become prominent as a vertebrate model organism well-suited for large-scale forward screening, and several recent efforts to perform saturation mutagenesis in zebrafish have underscored the system's utility (for review see 24).

Large-Scale Chemical Mutagenesis Screens in Zebrafish

Large-scale forward mutagenesis has been successfully employed in a variety of organisms including flies, worms, plants, and very recently zebrafish. Such screens are conducted based on the notion that the power of genetic analysis can be maximized by isolating mutations in as many members of a genetic pathway as possible. Furthermore, screens aimed at identifying all members of a given pathway are more likely to identify many previously uncharacterized genes.

The zebrafish is an excellent model system on which to perform large-scale forward mutagenesis for numerous reasons: (a) Zebrafish can be relatively easily maintained in large numbers; (b) they have a 2-3 month generation time; (c) hundreds of embryos can be obtained from single crosses; (d) the embryos are large, transparent, and develop outside of the mother; (f) embryogenesis is rapid, producing swimming larva within 5 days; and (e) haploid and gynogenetic embryos can be generated.

While small-scale screens for mutations affecting zebrafish embryogenesis have yielded several interesting mutations over the years, including *no tail, cyclops,* and *spadetail* (25-27), these screens have not taken full advantage of the potential power of the zebrafish as a genetic model organism. Recently, however, several laboratories have conducted large-scale screens, aimed at saturation level, using the chemical mutagen ethylnitrosourea (ENU) (28-31). The labs of C. Nüsslein-Volhard, M. Fishman, and W. Driever have performed these screens based on breeding strategies designed to uncover mutant phenotypes in the F3 progeny of

mutagenized fish. Crosses were screened for phenotypes affecting pattern formation, organogenesis, and motility, and several thousand mutations were identified. Mutations resulting in non-specific phenotypes were discarded and roughly 1,800 mutations were kept in total which, based upon complementation tests, are thought to represent 400-500 different genes (24, 30, 31).

These screens are potentially of tremendous importance to furthering our understanding of vertebrate development. Mutations have been obtained effecting virtually every observable process during zebrafish embryogenesis including gastrulation, somitogenesis, axon guidance, and the development of the brain, heart, liver, gut, and pigmentation (30, 31). Unfortunately, although these screens were designed to achieve saturation, current estimates based upon allele frequencies suggest that they may have identified mutations in only 40-50% of the genes essential for normal zebrafish embryonic development. Nonetheless, this work is of major importance because it has demonstrated the feasibility of large-scale screening in zebrafish, and more importantly, has provided a vast collection of mutations which create a comprehensive picture of the mutant phenotypes that can occur in this vertebrate.

While chemical mutagenesis screens of this sort are clearly of value, they are hindered by the fact that ENU usually induces point mutations (32), and the identification of genes mutated in such a manner can be quite difficult. Subsequently, although genetic analysis can be performed in the absence of information regarding the molecular nature of the mutated genes, genetic data alone is of limited use. The cloning of the such genes is a necessary step toward fully understanding their role during development.

The identification of genes disrupted by point mutations has proven quite challenging in many organisms including flies, worms, plants, mice, zebrafish, and humans. Such work is often initiated by determining what region of the genome a

mutation is genetically linked to and then isolating genomic DNA in that region and sequencing it and/or using it for activity-based assays and expression analysis to find the gene of interest. This "positional cloning" methodology has been used successfully in numerous systems (33, 33a), although the process is often extremely laborious and usually requires years of intensive work. In particular, in vertebrates which have genomes over 10^9 bp in size, an order of magnitude larger than those of flies and worms for example, the difficulty in cloning point mutations is compounded by the need for very large numbers of genomic markers.

An increasingly popular variation on positional cloning as described above, has been to look for 'candidate' genes which map to the region of the genome containing a mutation of interest, and to attempt to identify one of these as the sought after gene (33). Candidate genes can be divided into three categories: (a) Known genes with biological functions demonstrated by previous studies, (b) previously uncharacterized genes with conserved domains that suggest biological functions, and (c) novel genes with no obvious biological functions. When the apparent biological function of a gene in a given region correlates well with the mutant phenotype in question, that gene can be studied further to determine, for example, if mutant strains possess nonfunctional copies. Several human disease genes have been identified in this way including those responsible for Marfan syndrome (fibrillin) (34), and multiple endocrine neoplasia type 2 (c-RET) (35). If no transcripts are known to map to the region in question, or if those in the region have no predictable function, identifying the mutated gene is more difficult.

The abundance of ENU-induced zebrafish mutations has created the need for a high-resolution genetic map of the zebrafish genome to permit the genomic localization of mutations as well as candidate genes (36). Although relative to other vertebrates, a limited number of zebrafish genes have been cloned to date, efforts are underway to clone and map thousands of expressed sequence tags (ESTs). These will

eventually help expedite the identification of transcription units in a region of interest. Nevertheless, based upon positional cloning efforts in other organisms the identification of zebrafish genes mutated by ENU is likely to remain difficult for at least several years. Among these mutated genes, some are likely to be cloned relatively rapidly by the candidate gene approach outlined above, however as the nature of that approach suggests, efforts to clone novel genes will not benefit.

An alternative to chemical mutagenesis in various animal model systems has been to use insertional mutagenesis methods (see below). Insertional mutagenesis screens, and in particular those using cleanly integrating DNA elements, have been quite effective at mutagenizing and rapid cloning many biologically interesting genes. With these issues in mind and prompted by concerns regarding the clonability of the chemical mutations in zebrafish, we chose to undertake the development of a large-scale insertional mutagenesis methodology in this vertebrate. Based upon the success of such methods for mutating and rapidly cloning genes in flies, worms, and mice, we believed that a similar method in zebrafish would provide a powerful tool for the identification of both known and novel genes with essential roles during vertebrate development.

Insertional Mutagenesis in Invertebrates

Insertional mutagenesis methods take advantage of the fact that the integration of DNA into the genome of most organisms will at some frequency result in the disruption of endogenous gene functions. While insertional mutagenesis is generally less efficient than chemical mutagenesis, its primary advantage is that the mutagenic DNA serves as a molecular 'tag' which can facilitate the cloning of mutated genes (37-39). Molecular methods which take advantage of the nucleotide sequence of a mutagenic insertion can be used to isolate the genomic DNA which flanks that insertion. Genes identified in the flanking genomic DNA

are excellent candidates to be the disrupted gene responsible for the observed mutant phenotype.

Insertional mutagenesis in animal systems has been performed in the invertebrates *D. melanogaster* (37) and *C. elegans* (38) as well as in a vertebrate, the mouse (39). Such efforts in flies and worms have been more effectively applied than those in mice for several reasons, including the fact that these invertebrates are much easier to maintain in large numbers, have a significantly shorter generation times, and have genomes roughly one-twentieth the size of the mouse genome. Furthermore, in flies and worms new insertions are generated simply by crossing different lines and allowing transposons to "hop" to new insertion sites. In mice, however, the generation of insertions usually requires costly and time consuming embryonic manipulations which limit the number of insertions that can be generated.

In flies, the transposition of P elements is used to conduct insertional mutagenesis screens (37, 40, for review see 41). P insertions have been widely used both to generate new mutations, as well to isolate insertional alleles of genes previously identified by chemical lesions. The generation of insertional alleles is a popular alternative to positional cloning efforts because it can greatly simplify the cloning of the mutated gene. P element mutagenesis screens are performed using two types of fly strains, one containing functional transposase but lacking a P element capable of transposition, and the other containing one or several P insertions but no transposase (37). The crossing of these strains mobilizes the P element which then integrates elsewhere and may disrupt the function of an essential gene. Roughly 10-15% of P insertions disrupt essential genes (41).

P element insertional screens have been performed in several different ways. Some screens have used a dominant eye color marker to detect new insertion events (37). Others screens have used enhancer or promoter "traps" to identify

insertions which may have occurred in or near genes (40, 42). Trap constructs contain a reporter which can only be expressed when the insertion occurs near endogenous enhancers or promoters. Consequently, such insertions express their reporter in a pattern similar to that of the endogenous gene and thereby facilitate screening for mutations into genes expressed at a given time or in a given tissue.

Large-scale P element screens have been performed by generating thousands of insertions and screening those insertions for recessive lethal phenotypes. These screens have suggested that P elements are capable of mutating 30-50% of the genes which can be disrupted by chemical mutagenesis, a result which indicates that the insertion of P elements is not entirely random (41).

In addition to being useful for forward mutagenesis in flies, P elements can be used to isolate mutations in genes of interest which have already been characterized either genetically or molecularly. For such efforts, strains containing numerous P elements and/or P elements close to the locus of interest are often used (43). The simultaneous transposition of multiple P insertions increases the likelihood of hitting the gene in question. Furthermore, the fact that P insertions often 'hop' to new insertion sites within 100 Kb of the site of origin makes inserts near the gene of interest useful for generating insertional alleles (44). If insertional inactivation is not achieved even by integration events very close to the sequence in question, an alternative is to screen for imprecise excision of such elements to identify events which have deleted regions of the adjacent genomic DNA (41). In general, P element-mediated insertional mutagenesis has facilitated the identification of many genes with important role during fly development, including *prospero, torso, easter,* and *Toll* (45-48) among others.

In the nematode *C. elegans,* the transposon Tc1 has been used to induce insertional mutations (38, 43). This methodology is similar to that of P element mutagenesis in flies, but is limited in comparison for several reasons. First, because

insertion-free strains of C. *elegans* have not been isolated, and the 'low-copy' number strains harbor about 30 insertions, the identification of mutagenic integration events is significantly complicated. In addition, enhancer and promoter traps, of the sort used in flies, have not been developed for use in C. *elegans*. Nevertheless, the isolation of Tc1 insertions into loci of interest has facilitated the cloning of many C. *elegans* genes, including *unc-52*, *daf-1*, and *fem-3* (49-51) among others. 'Mutator' strains are used to isolate insertions into a given gene or region and these can be screened for gene disruption (38). The imprecise excision of Tc1 can also be used to disrupt local genes in a manner similar to that used in flies (43). Furthermore, the generation of worm collections comprising Tc1-insertion "libraries," in conjunction with PCR-based screening strategies has permitted the identification of mutagenic insertions into genes of interest (52, 53).

Insertional Mutagenesis in Mice

In the past, insertional mutagenesis in vertebrates has been limited to mice and has used retroviruses and plasmid DNA as mutagens. The first demonstration that retroviruses could disrupt genes was obtained in tissue culture during studies which infected RSV-transformed rat B31 cells with Moloney murine leukemia virus (MoMLV) (54). This work found that two 'revertant' B31 clones (no longer transformed) had MoMLV proviral insertions into the transforming RSV provirus which disrupted the expression of the *src* oncogene. In support of the notion that the MoMLV insertions had caused these disruptions, this work also showed that deletion of the MoMLV genome, leaving a single long terminal repeat, could 'back revert' these clones to regain transformed properties.

The first indication that proviral insertion could disrupt mouse genes *in vivo* was the observation that an allele of the *dilute* mutation in mice was associated with the integration of an endogenous retrovirus (55). Similar to the *in vitro* study

described above, this work also showed that deletion of the proviral integration was associated with loss or 'reversion' of the mutant phenotype. Contemporary with these observations was the development of transgenic technologies in mice, using either retroviral infection (56, 57), or the injection of plasmid DNA into embryos (58, for review see 59). These methods permitted the introduction of exogenous DNA into the mouse genome during embryogenesis and made the generation of transgenic mouse lines possible. Subsequently, it was found that the integration of retroviral proviruses as well as of injected plasmid DNA could be used as tools to disrupt essential genes in mice.

The mechanism of insertion as well as the structure of integrated transgenes made using either retroviral infection or plasmid microinjection are significantly different. Retroviruses are surrounded by a bi-lipid membrane containing envelope proteins which mediate penetration into the host cell (60). Once inside a cell the RNA genome of the virus is reverse transcribed to produce a double-stranded DNA copy called the 'provirus', which can then be stably integrated into a host chromosome through the activity of the integrase protein. Such proviral insertions are precise, maintain the linear organization of the retroviral genome, and result in a short direct repeat of 4-6 nucleotides in the genomic DNA at the site of integration.

The mechanism which mediates the integration of microinjected plasmid DNA is not well understood. To generate transgenes in this manner, several hundred copies of plasmid are injected directly into the pronuclei of fertilized mouse embryos (59). This DNA often appears to undergo concatamerization and then integrates seemingly haphazardly into the mouse genomic DNA. Plasmid transgenes can be single-copy or multi-copy ranging from several to hundreds of head-to-tail or tail-to-tail repeats. Often large deletions (10-100 Kb) in the genomic DNA are created by the integration event and other rearrangements such as duplications and translocations may occur as well (59, 61, 62).

It was first observed that plasmid transgenes could induce insertional mutations in mice during a study which inbred six mouse lines harboring human growth hormone transgenes generated by plasmid microinjection (63). Two of these lines had recessive lethal mutations apparently caused by the transgenic insertions (61, 62). Subsequently, as many transgenic mouse lines have been produced during the course of a wide range of studies, it has been found that roughly 8% of these lines have mutant phenotypes associated with and apparently caused by the insertions (39). While numerous of these mutations have been studied further, and include *limb deformity, inversion of embryonic turning,* and *germ-cell deficient* (64-66), often the identification of the mutated gene has been severely hindered by the imprecise nature of plasmid transgene insertion. This is because deletions and other rearrangements at the insertion site can theoretically cause the simultaneous disruption of several genes. Consequently, the primary theoretical advantage of insertional mutagenesis, the ability to rapidly identify the mutated genes, has not been widely realized using this methodology.

In contrast, the cloning of mouse mutations induced by the proviral insertions of retroviruses has proceeded relatively quickly (for review see 67). Proviral insertions integrate cleanly into the genome and alter the genomic DNA in a simple predictable manner (60). As a result, the disruption caused by a given proviral insertion can be easily localized to within a few base pairs, and attention can be focused directly on the lesion responsible for the mutant phenotype.

The use of retroviruses to induce insertional mutations in mice has taken several paths including the use of both exogenous and endogenous viruses (68-71). Early efforts using exogenous viruses generated proviral transgenes with either wild-type MoMLV or replication-defective MoMLV-based retroviral vectors. Mouse embryos were infected before or after implantation and transgenic lines were established which harbored single proviral insertions (68, 69). Those insertions

were then inbred and mutant phenotypes were identified. This work demonstrated that roughly 5% of proviral insertions disrupt essential genes in mice (72).

Proviral insertions have been shown to mutate endogenous mouse genes in several different ways. The insertion at the *dilute* locus is in the third intron of a myosin heavy chain (73, 74), and disrupts expression by causing aberrant splicing from the splice donor of exon 3 to the splice acceptor of the retroviral envelope gene (74). This transcript then polyadenylated prematurely in the 3′ proviral LTR. In the case of the Mov13 mutation, which was induced by a proviral insertion into the first intron of the a1(I) collagen gene, transcription of that gene is at 1-5% of wild-type levels (75). This is thought to be the result of the disruption of a *cis*-acting regulatory element within the first intron (76). Curiously, in the cases of both *dilute* and Mov13, the proviral disruptions are tissue-specific with these genes functional in certain tissues and not in others (74, 77). Other examples of the mechanisms by which proviral insertions can disrupt gene expression come from *in vitro* studies which have shown that proviral insertions can lead to premature polyadenylation (78), and that transcription from the viral LTR can antagonize endogenous genes transcription, or can create hybrid transcripts which initiate in the virus (79).

In addition to the infection of pre- or post-implantation of mouse embryos with MoMLV-derived viruses mentioned above, efforts to use exogenous retroviruses as insertional mutagens in mice have also used embryonic stem (ES) cells to introduce insertions into the mouse genome. These studies were designed to increase the efficiency of insertional mutagenesis screening either by introducing many insertions into the germ line simultaneously (80), or by allowing pre-selection of insertions into genes (81, 82). Studies designed to increase insertion number have infected ES cells repeatedly in culture to generate clones with over 10 insertions which could give rise to mice with multiple insertions (80). This work has

generated several insertional mutations including one in the gene *nodal* which is essential for anterior/posterior axis formation in mice (83, 84).

To preselect for insertions into genes, several studies have used retroviral 'gene traps' (81, 82). These constructs are similar in principle to the P element promoter and enhancer traps described earlier. The use of retroviral gene traps in mice has included vectors containing genes such as *neo*, *β-gal*, or *β-geo* as reporters which can only be expressed when inserted downstream of an endogenous promoter. Retroviral gene traps have been designed with or without splice acceptor (SA) sequences. Those without an SA tend to have lower 'activation' frequencies, or the frequencies with which the reporter genes are expressed (85), than those traps with an SA which allows the reporter to splice into endogenous transcripts (81). In either case, ES cells have been infected with gene trap vectors and activation events have been detected *in vitro*. ES cell clones containing activated traps have then been used to generate germ-line chimeric mice, and these activated insertions then inbred to screen for mutant phenotypes. Roughly 40% of activated gene trap insertions disrupt essential genes in mice (82) which is nearly a tenfold increase in frequency from that observed with unselected insertions (5%) (72). While this method has proven fruitful, once again the expensive nature of mouse work and the problems inherent to screening embyros which develop *in utero* have limited the application of this approach.

As mentioned above, endogenous retroviruses have also been used to generate insertional mutations in mice (71). This work has focused on the use of two different mouse strains, SWR/J and RF/J, which are crossed to produce viremic hybrid offspring. When such female hybrids are then crossed to SWR/J males the progeny often harbor new insertions which can subsequently be inbred to screen for mutant phenotypes (86). This method is similar in strategy to those used in flies and worms which take advantage of genetic crosses to mobilize insertional

elements. While it has been demonstrated that new proviral insertions generated in this way can be mutagenic (71) this method has not been utilized to a significant extent.

In general, while numerous insertional mutagenesis methods in mice have been developed over the years, none of these methods have been widely used. Consequently, the contribution of insertional mutagenesis to the study of vertebrate development has been limited to the occasional fortuitous isolation of either interesting novel mutations or of insertional alleles of previously identified loci (69, 70, 87, 88). The most aggressive efforts to conduct forward insertional screens in mice have focused primarily on gene trap screens, and while these efforts have been productive they have been limited in scale.

There are two likely reasons why random insertional mutagenesis in mice has not been widely performed. The first of which, as already indicated, is the cost of maintaining large numbers of mice and of conducting the embryonic manipulations required to carry out such screens. Central to this issue are the *in utero* development of the mouse embryo and the relatively small number of progeny produced in a given cross. Furthermore, because plausible screens can only isolate a handful of mutations, investigators with focused objectives are unlikely to attain mutations of interest to them.

Another reason why forward insertional screens in mice may have received limited attention was the advent in the late 1980's of targeted gene disruption (5). The ability to create mutations in any cloned mouse gene has provided an attractive tool whereby the function of genes implicated as important for a certain process could be tested by 'reverse' genetics. Consequently, efforts to address the genetic control of mouse development has shifted toward this method. Recently, however, it has grown increasingly apparent that the predicted role of genes, identified either biochemically or through the use of activity-based or subtractive strategies, can

sometimes be difficult to reconcile with the mutant phenotype. This realization has re-emphasized the value of forward genetics screens in vertebrates and has turned attention to the zebrafish as a system in which such screens are feasible.

Our efforts have centered on the development of a large-scale insertional mutagenesis strategy in zebrafish. Fundamental to any insertional mutagenesis method is the efficient introduction of foreign DNA into the genome of the organism being mutagenized. Therefore, the development and improvement of methods for generating transgenic zebrafish was the first step toward an insertional mutagenesis protocol.

Transgenesis in Zebrafish

The first transgenic zebrafish were generated by injecting plasmid DNA into the cytoplasm of one-cell stage embryos (89-91). In these studies, injected embryos were raised to adulthood and the germ-line transmission of plasmid sequences was detected either by gene expression or by PCR. These studies found that 5-20% of injected embryos, as adults, transmitted the foreign DNA to their progeny and that this DNA was inherited in a Mendelian fashion suggesting that it was stably integrated into the fish genome. Subsequent work has cloned the junction fragments between plasmid sequences and zebrafish genomic DNA and has confirmed that the plasmid DNA is integrated (P. Culp and N. Hopkins, unpublished data). Southern blot analysis of numerous transgenes of this sort in zebrafish indicates that they are similar in structure to those in mice and can be single- or multi-copy and often appear to be in tandem arrays (89, 91).

Recently, the generation of insertional mutations using plasmid transgenes in zebrafish has been shown to be possible. Of the more than 60 such transgenes which have been bred to homozygosity by several different labs, two are tightly linked to mutant phenotypes (D. Grunwald and J. Campos-Ortega, personal

communications). In neither case has the disrupted transcript been identified. In the past, large-scale insertional mutagenesis using this method has been considered, but the variable and usually low efficiency with which transgenes can be generated by plasmid microinjection has dampened enthusiasm for such a method. In addition, concerns regarding the extent to which complex transgene structures might complicate the identification of the mutated genes have further reduced interest in such efforts.

Various attempts have been made to devise alternative methods for introducing DNA into the zebrafish genome including the electroporation of embryos (92), the use of high-velocity microprojectiles (93), and the injection of protein-mediated integration systems such as those of P elements (94) and Tc1-like transposons (Z. Ivics and P. Hackett, personal communication). To date, none of these methods has been shown to be capable of generating transgenic zebrafish lines. However, in recent years a novel class of retroviral vector has been developed which is capable of stably integrating proviral DNA into zebrafish cells and of generating transgenic zebrafish (95, 96). These vectors are pseudotyped retroviruses which contains the envelope of the vesicular stomatitis virus and a nucleocapsid based on the Moloney murine leukemia virus. The work presented in this thesis is founded on the use of such vectors to generate transgenic zebrafish and insertional mutations.

MLV/VSV Pseudotyped Retroviral Vectors

Murine retroviral vectors are replication-defective viruses which are capable of infecting cells, but not of generating infectious particles from those cells (97). These viruses are produced in packaging cell lines which express the viral nucleocapsid and envelope proteins but are unable to package RNA encoding those proteins into viral particles. Retroviral-based constructs containing a packaging

sequence and other viral sequences necessary for provirus production, in addition to the gene sequences of interest to the investigator, are introduced into such packaging cell lines and infectious particles containing the engineered RNA genomes are produced. These particles can then be used to infect target cells and to stably introduce the DNA of choice into their genome.

The first step during successful retroviral infection is penetration of the viral particle into the cell being infected (60). This step typically involves an interaction between viral coat or envelope proteins and components, usually proteins, on the surface of the host cell. For example, the envelope protein of ecotropic MoMLV interacts with a cationic amino acid transporter on the surface of murine cells and permits the virus entry into such cells (98). Cells without this 'receptor' cannot be penetrated by ecotropic MoMLV and are therefore resistant to infection. In general, the interaction between envelope proteins and cell surface receptors plays a large role in determining viral host range.

In the past, the host range of the standard retroviral vectors was not thought to include zebrafish cells. For example, efforts to infect a zebrafish-derived cell line with an amphotropic murine retrovirus have proven unsuccessful (95). Such limitations are likely to result from an absence of the appropriate receptors on the surface of zebrafish cells. However, a novel class of retrovirus, first identified over twenty years ago, has been engineered in recent years for use as a retroviral vector which can be concentrated to very high titers, and can infect cells from a wide variety of species including zebrafish.

These novel vectors are hybrid or 'pseudotyped' viruses which contain the envelope glycoprotein (G-protein) of vesicular stomatitis virus (VSV) and an MoMLV-type nucleocapsid (95, 99). VSV is a rhabdovirus with a broad natural host range which includes mammals and insects (100). This host range is conferred by the envelope G-protein which may use the membrane phospholipid

phosphitadylserine (PTS) as its receptor. Evidence in support of this notion has come from studies showing that addition of excess PTS can specifically block binding of VSV to the cell surface (101). Additional work has shown that vesicles containing the VSV G-protein fuse preferentially with other vesicles composed of the acidic phospholipids PTS or phosphatidic acid (102). In light of this evidence, and since PTS is a common membrane component in a wide variety of organisms, it is a good candidate to mediate penetration of viral particles with the VSV-G envelope into the cells of many different organisms.

The production of pseudotyped viruses containing components of VSV and retroviruses was first observed when cells infected with either MLV or AMV (avian myeloblastosis virus) were superinfected with VSV (103). A small fraction of the viral particles produced were resistant to neutralization by anti-VSV antisera and had host ranges characteristic of either MLV or AMV. This work suggested that the VSV core could be surrounded by a retroviral envelope and attain the host range of that retrovirus. It was later shown that the inverse could also be true and that the VSV envelope could confer broadened host range upon RSV (Rous sarcoma virus) core particles (104).

While this work demonstrated that VSV and various retroviruses could form hybrid particles, the requirements for the formation of these particles was not clear. An understanding of this process was of interest because the production of retroviral vectors with VSV host range represented a potentially powerful tool for use in human gene therapy. To address this issue, T. Friedmann and colleagues investigated the ability of MLV-based vectors to form VSV pseudotyped particles with the VSV G-protein as the only VSV protein contribution. This work found that the VSV G-protein was sufficient to confer the VSV host range to MLV vectors in the absence of any other VSV-derived proteins, and that the simultaneous presence of the MLV envelope protein on the surface of such particles was

unnecessary (105). These results indicated that the development of MLV/VSV pseudotyped retroviral vectors in the absence of live VSV would be possible.

Subsequent efforts to produce MLV/VSV pseudotypes have transiently transfected a construct expressing the VSV G-protein into a 293 cell-derived packaging cell line which expresses the gag-pol protein of MLV as well as a retroviral construct for packaging (95, 99). The virus produced by these cells can then be concentrated to titers of over 10^9 cfu/ml by ultracentrifugation and can infect a wide range of cell types otherwise refractory to MLV infection (95, 106, 107). Upon cell penetration, mediated by the VSV envelope, the MLV core particle can produce a proviral DNA and integrate that DNA into the genome of the host cell.

The observation that MLV/VSV vectors could infect zebrafish cells was first made by Burns *et al.* using the embryonic zebrafish cell line ZF4 (95). Infection of this cell line with a *neo* expressing virus demonstrated that *neo* resistance could be conferred to these cells and suggested that the virus had stably integrated into them. Subsequent efforts with a second embryonic zebrafish cell line PAC2 showed both by *neo* resistance and by Southern blot analysis that the proviral DNA was likely to be integrated (see Chapter 1) (96).

This work demonstrated that MLV/VSV vectors would permit the use of retroviral vector technology in zebrafish. Our interest in these vectors stemmed from an interest in developing a large-scale insertional mutagenesis strategy in zebrafish. While the infection of cultured cells lines was encouraging, three fundamental questions remained regarding the utility of such vectors as insertional mutagens: (a) Could these vectors be used to infect the zebrafish germ line? (b) If so, would the production of proviral transgenes occur efficiently enough to make the generation of many thousands of insertions feasible? (c) Would proviral insertions into the zebrafish genome disrupt essential genes?

Retroviral Insertional Mutagenesis in Zebrafish

The work presented in the subsequent chapters demonstrates that MLV/VSV vectors can be used to generate transgenic zebrafish at very high efficiency and that proviral insertions can disrupt essential genes. This work has significant implications for both the future exploration of zebrafish development as well as the study of vertebrate biology in general (see Chapter 6). By facilitating the cloning of mutated genes this method will permit the assignment of unique and essential biological functions to hundreds of known or previously uncharacterized vertebrate genes.

The following chapter describes the first demonstration that the injection of an MLV/VSV virus into blastula-stage zebrafish embryos can infect the cells destined to become the germ line. The work presented in Chapter 3 then details a 100-fold improvement in germ-line infection efficiency which made the generation of thousands of insertions by small labs feasible. Next, Chapter 4 describes a pilot insertional mutagenesis screen during which three insertional mutations were isolated, and two of the mutated genes were identified. Preliminary analyses of these genes and their corresponding mutant phenotypes are presented in both Chapters 4 and 5. Finally, Chapter 6 addresses the potential value of this methodology to the study of zebrafish developmental genetics in particular, and to that of vertebrate biology in general. Both limitations and possible improvements are discussed.

REFERENCES

1. Sulston JE; Horvitz HR (1977) Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev Biol* **56**: 110-56.

2. Le Douarin NM (1986) Cell line segregation during peripheral nervous system ontogeny. *Science* **231**: 1515-22.

3. Sive HL (1993) The frog prince-ss: a molecular formula for dorsoventral patterning in *Xenopus*. *Genes Dev* **7**: 1-12.

4. Nusslein-Volhard C; Wieschaus E (1980) Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**: 795-801.

5. Capecchi MR (1989) The new mouse genetics: altering the genome by gene targeting. *Trends Genet* **5**: 70-6.

6. Zhang M; Kim HJ; Marshall H; Gendron-Maguire M; Lucas DA; Baron A; Gudas LJ; Gridley T; Krumlauf R; Grippo JF (1994) Ectopic Hoxa-1 induces rhombomere transformation in mouse hindbrain. *Development* **120**: 2431-42.

7. Mlodzik M; Gibson G; Gehring WJ (1990) Effects of ectopic expression of caudal during *Drosophila* development. *Development* **109**: 271-7.

8. Ferguson EL; Horvitz HR (1985) Identification and characterization of 22 genes that affect the vulval cell lineages of the nematode Caenorhabditis elegans. *Genetics* **110**: 17-72.

9. Lobe CG; Gruss P (1989) Mouse versions of fly developmental control genes: legitimate or illegitimate relatives? *New Biol* **1**: 9-18.

10. Krumlauf R (1992) Evolution of the vertebrate Hox homeobox genes. *Bioessays* **14**: 245-52.

11. Puelles L; Rubenstein JL (1993) Expression patterns of homeobox and other putative regulatory genes in the embryonic mouse forebrain suggest a neuromeric organization. *Trends Neurosci* **16**: 472-9.

12. Lewis J (1996) Neurogenic genes and vertebrate neurogenesis. *Curr Opin Neurobiol* **6**: 3-10.

13. Gilbert SF (1994) in *Developmental Biology* (Sinauer Associates, Inc, Sunderland, Massachusetts).

14. Spemann H; Mangold H (1924) Induction of embryonic primordia by implantation of organizers from different species. *Wilhelm Roux' Arch Entwicklungsmech Organ* **100**: 599-638.

15. Kimelman D; Christian JL; Moon RT (1992) Synergistic principles of development: overlapping patterning systems in *Xenopus* mesoderm induction. *Development* **116**: 1-9.

16. Davis RL; Weintraub H; Lassar AB (1987) Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**: 987-1000.

17. Rebagliati MR; Weeks DL; Harvey RP; Melton DA (1985) Identification and cloning of localized maternal RNAs from *Xenopus* eggs. *Cell* **42**: 769-77.

18. Smith WC; Harland RM (1992) Expression cloning of noggin, a new dorsalizing factor localized to the Spemann organizer in *Xenopus* embryos. *Cell* **70**: 829-40.

19. Bouwmeester T; Kim S-H; Sasai Y; Lu B; DeRobertis EM (1996) Cerberus is a head-inducing secreted factor expressed in the anterior endoderm of Spemann's organizer. *Nature* **382**: 595-601.

20. Nye JS; Kopan R; Axel R (1994) An activated Notch suppresses neurogenesis and myogenesis but not gliogenesis in mammalian cells. *Development* **120**: 2421-30.

21. Umbhauer M; Marshall CJ; Mason CS; Old RW; Smith JC (1995) Mesoderm induction in *Xenopus* caused by activation of MAP kinase. *Nature* **376**: 58-62.

22. Amaya E; Musci TJ; Kirschner MW (1991) Expression of a dominant negative mutant of the FGF receptor disrupts mesoderm formation in *Xenopus* embryos. *Cell* **66**: 257-70.

23. Damm K; Heyman RA; Umesono K; Evans RM (1993) Functional inhibition of retinoic acid response by dominant negative retinoic acid receptor mutants. *Proc Natl Acad Sci USA* **90**: 2989-93.

24. Eisen JS (1996) Zebrafish make a big splash. *Cell* **87**: 969-77.

25. Halpern ME; Ho RK; Walker C; Kimmel CB (1993) Induction of muscle pioneers and floor plate is distinguished by the zebrafish no tail mutation. *Cell* **75**: 99-111.

26. Hatta K; Kimmel CB; Ho RK; Walker C (1991) The cyclops mutation blocks specification of the floor plate of the zebrafish central nervous system. *Nature* **350**: 339-41.

27. Kimmel CB; Kane DA; Walker C; Warga RM; Rothman MB (1989) A mutation that changes cell movement and cell fate in the zebrafish embryo. *Nature* **337**: 358-62.

28. Mullins M; Hammerschmidt M; Haffter P; Nüsslein-Volhard C (1994) Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. *Curr Biol* **4**: 189-202.

29. Solnica-Krezel L; Schier AF; Driever W (1994) Efficient recovery of ENU-induced mutations from the zebrafish germline. *Genetics* **136**: 1401-20.

30. Haffter P; Granato M; Brand M; Mullins M; Hammerschmidt M; Kane D; Odenthal J; van Eeden F; Jiang Y; Heisenberg C-P; Kelsh R; Furutani-Seiki M; Vogelsang L; Beuchle D; Schach U; Fabian C; Nüsslein-Volhard C (1996) The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* **123**: 1-36.

31. Driever W; Solnica-Krezel L; Schier A; Neuhauss S; Maliki J; Stemple D; Stainier D; Zwartkruis F; Abdelilah S; Rangini Z; Belak J; Boggs C (1996) A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123**: 37-46.

32. Shibuya T; Morimoto K (1993) A reveiw of the genotoxicity of 1-ethyl-1-nitrosourea. *Mut Res* **297**: 3-38.

33. Collins FS (1995) Positional cloning moves from perditional to traditional. *Nat Genet* **9**: 347-50.

33a. Dietrich WF; Copeland NG; Gilbert DJ; Miller JC; Jenkins NA; Lander ES (1995) Mapping the mouse genome: current status and future prospects. *Proc Natl Acad Sci USA* **92**: 10849-53.

34. Dietz HC; Cutting GR; Pyeritz RE; Maslen CL; Sakai LY; Corson GM; Puffenberger EG; Hamosh A; Nanthakumar EJ; Curristin SM; Stetten G; Meyers DA; Francomano CA (1991) Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* **352**: 337-9.

35. Mulligan LM; Kwok JB; Healey CS; Elsdon MJ; Eng C; Gardner E; Love DR; Mole SE; Moore JK; Papi L; Ponder MA; Telenius H; Tunnacliffe A; Ponder BAJ (1993) Germ-line mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. *Nature* **363**: 458-60.

36. Postlethwait JH; Johnson SL; Midson CN; Talbot WS; Gates M; Ballinger EW; Africa D; Andrews R; Carl T; Eisen JS; Horne S; Kimmel CB; Hutchinson M; Johnson M; Rodriguez A (1994) A genetic linkage map for the zebrafish. *Science* **264**: 699-703.

37. Cooley L; Kelley R; Spradling A (1988) Insertional mutagenesis of the *Drosophila* genome with single P elements. *Science* **239**: 1121-8.

38. Herman RK; Shaw JE (1987) The transposable genetic element Tc1 in the nematode *Caenorhabditis elegans*. *Trends Genet* **3**: 222-5.

39. Gridley T; Soriano R; Jaenisch R (1987) Insertional mutagenesis in mice. *Trends Genet* **3**: 162-6.

40. Bier E; Vaessin H; Shepherd S; Lee K; McCall K; Barbel S; Ackerman L; Carretto R; Uemura T; Grell E; Jan LY; Jan YN (1989) Searching for pattern and mutation in the *Drosophila* genome with a P-lacZ vector. *Genes Dev* **3**: 1273-87.

41. Spradling AC; Stern DM; Kiss I; Roote J; Laverty T; Rubin GM (1995) Gene disruptions using P transposable elements: an integral component of the *Drosophila* genome project. *Proc Natl Acad Sci USA* **92**: 10824-30.

42. O'Kane CJ; Gehring WJ (1987) Detection in situ of genomic regulatory elements in *Drosophila*. *Proc Natl Acad Sci USA* **84**: 9123-7.

43. Kaiser K; Sentry JW; Finnegan DJ (1995) Eukaryotic transposable elements as tools to study gene structure and function, in *Mobile Genetic Elements*, ed, Sherratt DJ (Oxford University Press, Oxford) pp 69-100.

44. Tower J; Karpen GH; Craig N; Spradling AC (1993) Preferential transposition of *Drosophila* P elements to nearby chromosomal sites. *Genetics* **133**: 347-59.

45. Doe CQ; Chu-LaGraff Q; Wright DM; Scott MP (1991) The prospero gene specifies cell fates in the *Drosophila* central nervous system. *Cell* **65**: 451-64.

46. Sprenger F; Stevens LM; Nusslein-Volhard C (1989) The *Drosophila* gene torso encodes a putative receptor tyrosine kinase. *Nature* **338**: 478-83.

47. Chasan R; Anderson KV (1989) The role of easter, an apparent serine protease in organizing the dorsal-ventral pattern of the *Drosophila* embryo. *Cell* **56**: 391-400.

48. Hashimoto C; Hudson KL; Anderson KV (1988) The Toll gene of *Drosophila*, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. *Cell* **52**: 269-79.

49. Rogalski TM; Williams BD; Mullen GP; Moerman DG (1993) Products of the unc-52 gene in *Caenorhabditis elegans* are homologous to the core protein of the mammalian basement membrane heparan sulfate proteogycan. *Genes Dev* **7**: 1471-84.

50. Georgi LL; Albert PS; Riddle DL (1990) daf-1, a C. *elegans* gene controlling dauer larva development, encodes a novel receptor protein kinase. *Cell* **61**: 635-45.

51. Rosenquist TA; Kimble J (1988) Molecular cloning and transcript analysis of fem-3, a sex-determination gene in *Caenorhabditis elegans*. *Genes Dev* **2**: 606-16.

52. Rushforth AM; Saari B; Anderson P (1993) Site-selected insertion of the transposon Tc1 into a *Caenorhabditis elegans* myosin light chain gene. *Mol Cell Biol* **13**: 902-10.

53. Zwaal RR; Broeks A; van Meurs J; Groenen JT; Plasterk RH (1993) Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon insertion mutant bank. *Proc Natl Acad Sci USA* **90**: 7431-5.

54. Varmus HE; Quintrell N; Ortiz S (1981) Retroviruses as mutagens: insertion and excision of a nontransforming provirus alter expression of a resident transforming provirus. *Cell* **25**: 23-36.

55. Jenkins NA; Copeland NG; Taylor BA; Lee BK (1981) Dilute (d) coat colour mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome. *Nature* **293**: 370-4.

56. Jaenisch R (1976) Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. *Proc Natl Acad Sci USA* **73**: 1260-4.

57. Jaenisch R; Jahner D; Nobis P; Simon I; Lohler J; Harbers K; Grotkopp D (1981) Chromosomal position and activation of retroviral genomes inserted into the germ line of mice. *Cell* **24**: 519-29.

58. Gordon JW; Ruddle FH (1981) Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science* **214**: 1244-6.

59. Palmiter RD; Brinster RD (1986) Germ-line transformation of mice. *Ann. Rev. Genet* **20**:465-99.

60. Coffin JM (1991) Retroviridae and their replication, *in Fundamental Virology*, eds Fields BN; Knipe DM (Raven Press, Ltd., New york) pp 645-708.

61. Covarrubias L; Nishida Y; Mintz B (1986) Early postimplantation embryo lethality due to DNA rearrangements in a transgenic mouse strain. *Proc Natl Acad Sci USA* **83**: 6020-4.

62. Covarrubias L; Nishida Y; Terao M; D'Eustachio P; Mintz B (1987) Cellular DNA rearrangements and early developmental arrest caused by DNA insertion in transgenic mouse embryos. *Mol Cell Biol* **7**: 2243-7.

63. Wagner EF; Covarrubias L; Stewart TA; Mintz B (1983) Prenatal lethalities in mice homozygous for human growth hormone gene sequences integrated in the germ line. *Cell* **35**: 647-55.

64. Woychik RP; Maas RL; Zeller R; Vogt TF; Leder P (1990) 'Formins': proteins deduced from the alternative transcripts of the limb deformity gene. *Nature* **346**: 850-3.

65. Yokoyama T; Copeland NG; Jenkins NA; Montgomery CA; Elder FFB; Overbeek PA (1993) Reversal of left-right asymmetry: a situs invertus mutation. *Science* **260**: 679-82.

66. Pellas TC; Ramachandran B; Duncan M; Pan SS; Marone M; Chada K (1991) Germ-cell deficient (gcd), an insertional mutation manifested as infertility in transgenic mice. *Proc Natl Acad Sci USA* **88**: 8787-91.

67. Meisler MH (1992) Insertional mutagenesis of 'classical' and novel genes in transgenic mice. *Trends Genet* **8**: 341-4.

68. Jaenisch R; Harbers K; Schnieke A; Lohler J; Chumakov I; Jahner D; Grotkopp D; Hoffman E (1983) Germline integration of Moloney murine leukemia virus at the Mov13 locus leads to recessive lethal mutation and early embryonic death. *Cell* **32**: 209-16.

69. Soriano P; Gridley T; Jaenisch R (1987) Retroviruses and insertional mutagenesis in mice: proviral integration at the Mov 34 locus leads to early embryonic death. *Genes Dev* **1**: 366-75.

70. Conlon FL; Barth KS; Robertson EJ (1991) A novel retrovirally induced embryonic lethal mutation in the mouse: assessment of the developmental fate of embryonic stem cells homozygous for the 413.d proviral integration. *Development* **111**: 969-81.

71. Spence SE; Gilbert DJ; Swing DA; Copeland NG; Jenkins NA (1989) Spontaneous germ line virus infection and retroviral insertional mutagenesis in eighteen transgenic Srev lines of mice. *Mol Cell Biol* **9**: 177-84.

72. Jaenisch R (1988) Transgenic animals. *Science* **240**: 1468-75.

73. Mercer JA; Seperack PK; Strobel MC; Copeland NG; Jenkins NA (1991) Novel myosin heavy chain encoded by murine dilute coat colour locus. *Nature* **349**: 709-13.

74. Seperack PK; Mercer JA; Strobel MC; Copeland NG; Jenkins NA (1995) Retroviral sequences located within an intron of the dilute gene alter dilute expression in a tissue-specific manner. *EMBO J* **14**: 2326-32.

75. Hartung S; Jaenisch R; Breindl M (1986) Retrovirus insertion inactivates mouse alpha 1(I) collagen gene by blocking initiation of transcription. *Nature* **320**: 365-7.

76. Barker DD; Wu H; Hartung S; Breindl M; Jaenisch R (1991) Retrovirus-induced insertional mutagenesis: mechanism of collagen mutation in Mov13 mice. *Mol Cell Biol* **11**: 5154-63.

77. Kratochwil K; von der Mark K; Kollar E; Jaenisch R; Mooslehner K; Schwarz M; Haase K; Gmachi I; Harbers K (1989) Retrovirus-induced insertional mutation in Mov13 mice affects collagen I expression in a tissue specific manner. *Cell* **57**: 807-16.

78. Marchetti A; Buttitta F; Miyazaki S; Gallahan D; Smith GH; Callahan R (1995) Int-6, a highly conserved, widely expressed gene, is mutated by mouse mammary tumor virus in mammary preneoplasia. *J Virol* **69**: 1932-8.

79. Frankel WN; Potter TA; Rajan TV (1989) Effect of proviral insertion of transcription of the murine B2m^b gene. *J Virol* **63**: 2623-28.

80. Robertson E; Bradley A; Kuehn M; Evans M (1986) Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. *Nature* **323**: 445-8.

81. Friedrich G; Soriano P (1991) Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev* **5**: 1513-23.

82. Chen J; DeGregori J; Hicks G; Roshon M; Scherer C; Shi E; Ruley HE (1994) Gene trap retroviruses *Meth Mol Genet* **4**: 123-140.

83. Conlon FL; Lyons KM; Takaesu N; Barth KS; Kispert A; Herrmann B; Robertson EJ (1994) A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development* **120**: 1919-28.

84. Collignon J; Varlet I; Robertson EJ (1996) Relationship between asymmetric nodal expression and the direction of embryonic turning. *Nature* **381**: 155-8.

85. Chang W; Hubbard C; Friedel C; Ruley HE (1993) Enrichment of insertional mutants following retrovirus gene trap selection. *Virology* **193**: 737-47.

86. Jenkins NA; Copeland NG (1985) High frequency germline acquisition of ecotropic MuLV proviruses in SWR/J-RF/J hybrid mice. *Cell* **43**: 811-9.

87. Miao GG; Smeyne RJ; D'Arcangelo G; Copeland NG; Jenkins NA; Morgan JI; Curran T (1994) Isolation of an allele of reeler by insertional mutagenesis. *Proc Natl Acad Sci USA* **91**: 11050-4.

88. Keller SA; Liptay S; Hajra A; Meisler MH (1990) Transgene-induced mutation of the murine steel locus. *Proc Natl Acad Sci USA* **87**: 10019-22.

89. Stuart GW; McMurray JV; Westerfield M (1988) Replication, integration and stable germ-line transmission of foreign sequences injected into early zebrafish embryos. *Development* **103**: 403-12.

90. Stuart GW; Vielkind JR; McMurray JV; Westerfield M (1990) Stable lines of transgenic zebrafish exhibit reproducible patterns of transgene expression. *Development* **109**: 577-84.

91. Culp P; Nusslein-Volhard C; Hopkins N (1991) High-frequency germ-line transmission of plasmid DNA sequences injected into fertilized zebrafish eggs. *Proc Natl Acad Sci USA* **88**: 7953-7.

92. Buono RJ; Linser PJ (1992) Transient expression of RSVCAT in transgenic zebrafish made by electroporation. *Mol Mar Biol Biotechnol* **1**: 271-5.

93. Zelenin AV; Alimov AA; Barmintzev VA; Beniumov AO; Zelenina IA; Krasnov AM; Kolesnikov VA (1991) The delivery of foreign genes into fertilized fish eggs using high-velocity microprojectiles. *FEBS lett* **287**: 118-20.

94. Gibbs PDL; Gray A; Thorgaard G (1994) Inheritance of P element and reporter gene sequences in zebrafish. *Mol Mar Biol and Biotech* **3**: 317-26.

95. Burns JC; Friedmann T; Driever W; Burrascano M; Yee JK (1993) Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. *Proc Natl Acad Sci USA* **90**: 8033-7.

96. Lin S; Gaiano N; Culp P; Burns JC; Friedmann T; Yee JK; Hopkins N (1994) Integration and germ-line transmission of a pseudotyped retroviral vector in zebrafish. *Science* **265**: 666-9.

97. Mulligan RC (1993) The basic science of gene therapy. *Science* **260**: 926-32.

98. Kim JW; Closs EI; Albritton LM; Cunningham JM (1991) Transport of cationic amino acids by the mouse ecotropic retrovirus receptor. *Nature* **352**: 725-8.

99. Yee J-K; Miyanohara A; LaPorte P; Bouic K; Burns JC; Friedmann T (1994) A general method for the generation of high-titer, pantropic retroviral vectors: Highly efficient infection of primary hepatocytes. *Proc Natl Acad Sci USA* **91**: 9564-8.

100. Wagner RR (1991) Rhabdoviridae and their replication, *in Fundamental Virology*, eds Fields BN; Knipe DM (Raven Press, Ltd., New york) pp 489-503.

101. Schlegel R; Willingham C; Pastan I (1982) Saturable binding sites for vesicular stomatitis virus on the surface of Vero cells. *J Virol* **43**: 871-5.

102. Eidelman O; Schlegel R; Tralka T; Blumenthal R (1984) pH-dependent fusion induced by vesicular stomatitis virus glycoprotein reconstituted into phospholipid vesicles. *J Biol Chem* **259**: 4622-8.

103. Zavada J (1972) Pseudotypes of vesicular stomatitis virus with the coat of murine leukemia and of avian myeloblastosis viruses. J Gen Virol 15: 183-91.

104. Weiss RA; Boettiger D; Love DN (1974) Phenotypic mixing between vesicular stomatitis virus and avian RNA tumor viruses. *Cold Spring Harbor Symp Quant Biol* **39**: 913-8.

105. Emi N; Friedmann T; Yee J-K (1991) Pseudotype formation of murine leukemia virus with the G protein of vesicular stomatitis virus. *J Virol* **65**: 1202-7.

106. Lu J-K; Chen TT; Allen SK; Matsubara T; Burns JC (1996) Production of transgenic dwarf surfclams, *Mulinia lateralis*, with pantropic retroviral vectors. *Proc Natl Acad Sci USA* **93**: 3482-6.

107. Matsubara T; Beeman RW; Shike H; Besansky NJ; Mukabayire O; Higgs S; James AS; Burns JC (1996) Pantropic retroviral vectors integrate and express in cells of the malaria mosquito, *Anopheles gambiae*. *Proc Natl Acad Sci USA* **93**: 6181-5.

CHAPTER 2

INTEGRATION AND GERM-LINE TRANSMISSION OF A PSEUDOTYPED RETROVIRAL VECTOR IN ZEBRAFISH

ABSTRACT

The zebrafish is rapidly becoming a popular model system for the study of vertebrate development because it is ideal for both embryological studies and genetic analysis. To determine if a retroviral vector pseudotyped with the envelope glycoprotein of the vesicular stomatitis virus could infect zebrafish embryos, and in particular, the cells destined to become the germ line, a pseudotyped virus was injected into blastula-stage zebrafish embryos. Fifty-one embryos were allowed to develop and eight transmitted proviral DNA to their progeny. Founders were mosaic, but as expected, transgenic F1's transmitted proviral DNA in a Mendelian fashion to the F2 progeny. Transgenic F1 fish inherited a single integrated provirus and a single founder could transmit more than one viral integration to its progeny. These results demonstrate that this pantropic pseudotyped vector, originally developed for human gene therapy, will make the use of retroviral vectors in zebrafish possible.

INTRODUCTION

That retroviruses could be used to deliver foreign DNA into the genome of an animal was first demonstrated by infecting preimplantation stage mouse embryos with the Moloney murine leukemia virus (MoMLV) and obtaining germ line transmission of an integrated provirus (1). Subsequently, the ability of retroviruses to integrate exogenous DNA into the genome of infected cells has been exploited for gene therapy (2), for cell lineage studies (3-5), and for studies of insertional mutagenesis (6, 7). In addition, the use of retroviral gene traps in conjunction with mouse embryonic stem cells has proven quite effective in the search for and mutagenesis of genes expressed during mouse development (7-9).

In recent years the zebrafish, *Danio rerio*, has become a popular model system for vertebrate developmental studies because it offers the opportunity to combine classical genetic analysis, including large scale mutagenesis, with an easily accessible and manipulatable embryo. Genetic studies of the zebrafish benefit from the 2-3 month generation time, the ability of females to lay hundreds of eggs routinely , and the small size of the adults, while embryological studies benefit from the large, transparent embryos, detailed fate maps, and the fact that single identified cells can be studied in living embryos (10). In the past, the application of retroviral vector technology to the zebrafish system was not feasible due to the limited host range of the standard vectors. However, a recent report has demonstrated that a pseudotyped retroviral vector, which can be concentrated to very high titers, can infect cultured fish cells, including those derived from zebrafish embryos (11). This virus contains an MoMLV-based genome surrounded by an envelope containing the glycoprotein (G-protein) of the vesicular stomatitis virus (VSV), completely replacing the retroviral *env* glycoprotein. As a result of the presence of the VSV G-protein, this

pseudotyped virus has the broad host range characteristic of VSV, and upon entry into a permissive cell will integrate retroviral sequences into the host genome.

The ability of the MoMLV(VSV) pseudotyped virus to infect cultured fish cell lines suggested that it might be possible to infect zebrafish embryos and to obtain germ line transmission of integrated proviral DNA. However, because zebrafish embryos develop very rapidly and at 28°C while murine retroviruses generally require over 6 hours at 37°C to synthesize and integrate proviral DNA (12), it was unclear whether or not germ-line transmission could be obtained efficiently or at all in zebrafish. In these studies we describe the infection of the zebrafish germ line with this pseudotyped virus and discuss its potential use in the study of zebrafish development.

RESULTS

A concentrated stock of the pseudotyped virus LZRNL(G) (Fig. 1) was generated essentially as described (11). This virus was titered on cultured zebrafish cells by infecting an established zebrafish cell line, PAC2 cells, and selecting for clones in media containing G418. LZRNL contains the neomycin phospho-transferase gene (*neo*) and thus can confer G418 resistance to infected cells. Control mouse 3T3 cells were infected under the same conditions. The titer of the virus was 6.7×10^6 cfu/ml on zebrafish PAC2 cells, and 2.5×10^7 cfu/ml on mouse 3T3 cells.

To generate transgenic zebrafish, we injected LZRNL(G) virus into the blastoderm, among the cells of blastula stage embryos, at approximately the 2000-4000 cell stage. On the basis of the virus titer on PAC2 cells and the volume injected, we estimate that 50-100 infectious units were injected into each embryo. To test germ-line transmission of proviral DNA, injected embryos were raised to sexual maturity, mated, and DNA from 24-hour-old pools of their F1 progeny were tested

Figure 1

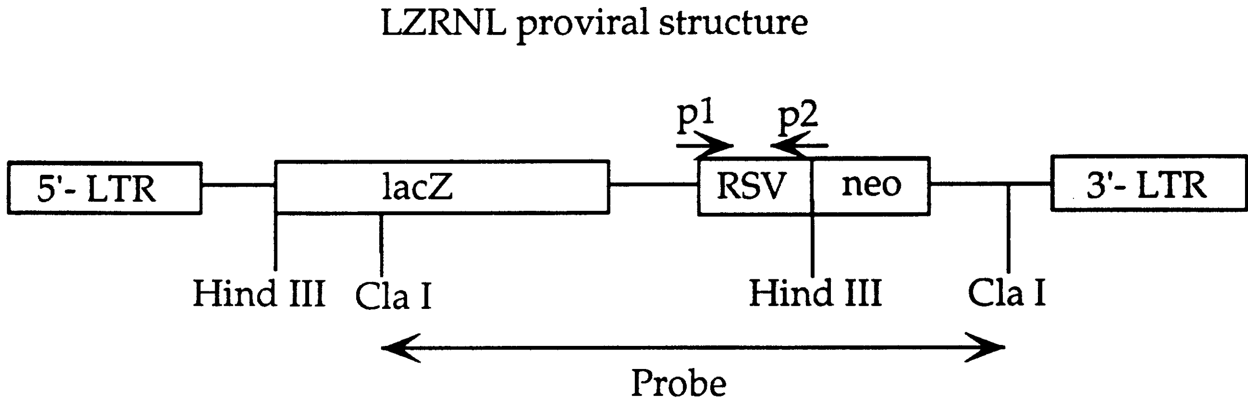


FIG. 1 Map of LZRNL. The locations of the PCR primers used to identify transgenic fish are indicated with arrows. The *Cla I* fragment was used as the probe for Southern blot analysis. The construct is not drawn to scale.

for the presence of LZRNL sequences by PCR. In total, 8 out of 51 fish examined showed germ line transmission of the retroviral sequences (Fig. 2A). Because only 50-100 F1 embryos were collected to test for germ line transmission, founders which transmitted proviral sequences to less than 1% of their offspring may have been overlooked.

Because the virus was injected into blastula-stage embryos containing a large number of potential target cells, the embryos were very likely to be mosaic for the presence of integrated viral sequences. To determine if the founder fish had mosaic germ lines, individual F1 progeny from each founder were analyzed by PCR for the presence of proviral DNA (Fig. 2B). As shown in Table 1, all eight founders did indeed have mosaic germ lines and transmitted proviral DNA to less than 5% of their F1 progeny.

Live transgenic F1 fish were identified by isolating genomic DNA from caudal fin clips and using PCR to test for the presence of the viral transgene. Two transgenic F1's identified in this way were then mated to non-transgenic fish and individual F2 embryos were screened by PCR. If a transgenic F1 fish contained an integrated provirus, that provirus should have been transmitted to 50% of the F2 progeny. The first F1 tested transmitted the transgene to 11 out of 25 of the F2 progeny (44%) (Fig. 2C) and the second F1 tested transmitted the transgene to 8 out of 17 of the F2 progeny (47%). These frequencies are consistent with Mendelian transmission and support the notion that the proviral DNA is integrated into the zebrafish genome.

Typically retroviral DNA integrates into a host chromosome as a single copy in a manner that maintains the linear organization of its viral genome (13). To confirm that the proviral sequences were integrated in the expected arrangement in the infected fish cells and transgenic animals we performed Southern blot analysis. Genomic DNA from two clones of cultured PAC2 cells infected with LZRNL(G), as

Figure 2

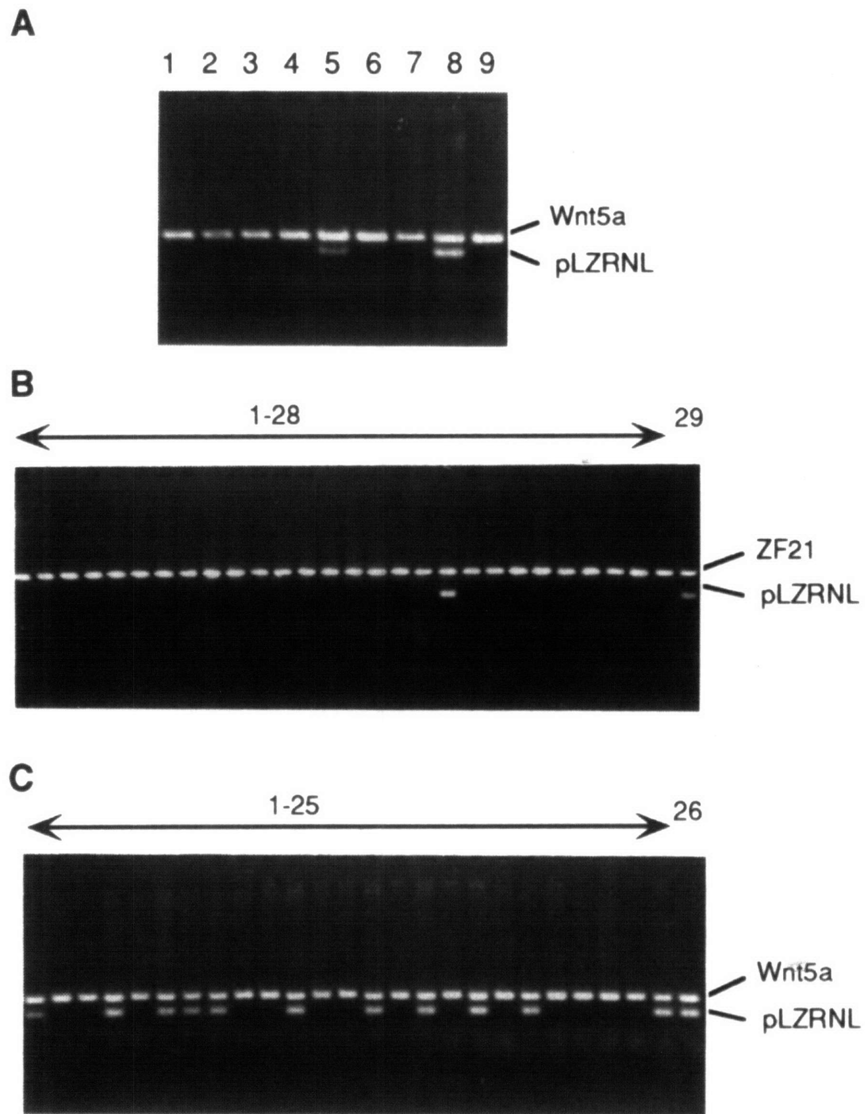


FIG. 2 Polymerase chain reaction (PCR) analysis to detect germ-line transmission of proviral DNA. (A) PCR analysis using DNA extracted from pools of the F1 progeny of individual injected potential founder fish. Two pairs of primers were used for PCR analysis of zebrafish genomic DNA. The first pair of primers is specific to LZRNL proviral DNA (see Fig. 1) and generate a 290 bp PCR product. The second pair of primers is specific to the zebrafish Wnt5A gene and generates a 387 bp PCR product. Lanes 1-7 show the PCR products generated using DNA from pools of embryonic F1 progeny. Note that lane 5 is positive for LZRNL sequence suggesting that the fish which gave rise to those embryos is a transgenic founder. Lane 8 shows the PCR products generated using DNA from a zebrafish cell line that contains a proviral insertion. Lane 9 shows the PCR product generated using DNA from an uninjected fish. (B) PCR analysis performed using DNA extracted from individual F1 progeny of a positive founder fish. In this case, the second pair of primers is specific to the zebrafish homeobox gene, ZF21, generating a 475 bp PCR product. Lanes 1-28 show the PCR products generated using DNA from individual F1 fish. Lane 29 shows the PCR products generated using DNA from a zebrafish cell line that contains a proviral insertion. Note that the F1 represented by lane 19 is transgenic. (C) DNA was extracted from the individual F2 embryos of a transgenic F1 fish and was analyzed by PCR. Lanes 1-25 are DNA from individual F2 fish. Lane 26 shows the PCR products generated using DNA from a zebrafish cell line that contains a proviral insertion.

Table 1. Mosaicism of germ-line transmission of proviral DNA from founders to the F1 generation

Founder	Transgenic F1's	Mosaicism
m4	6/306	2.0%
f6	3/128	2.3%
f12	2/86	2.3%
f13	1/20	5.0%
f36	2/110	1.8%
f44	1/60	1.7%
f45	2/50	4.0%
m49	1/115	0.9%

well as from the transgenic F1 progeny of two different founders (m4 and f13, see Table 1) was digested with the restriction enzyme *Dra I*. In addition, genomic DNA from the two infected PAC2 clones and from the transgenic F1 progeny of all eight founders was digested with *Hind III*. *Dra I* does not cleave within the LZRNL sequence and should yield one or more fragments, depending upon the number of integrations, with sizes larger than the proviral genome (6.6 Kb). *Hind III* cleaves twice within the LZRNL sequence (see Fig. 1) and is expected to yield a 3.7 Kb internal fragment for all insertions, and two junction fragments with sizes dependent upon the presence of *Hind III* sites in the surrounding genomic sequences. The *Cla I* fragment of pLZRNL containing *lacZ*, Rous sarcoma virus (RSV), and *neo* sequences, was used as the hybridization probe (Fig. 1), and was expected to hybridize to both the internal 3.7 Kb *Hind III* fragment and to the 3' junction fragment.

As shown in Fig. 3A, samples digested with *Dra I* revealed only single bands of variable size, indicating the presence of single copies of integrated provirus (Fig 3A, lanes 1-4). In the case of *Hind III* digests a 3.7 Kb fragment of LZRNL was detected in the genomic DNA of the transgenic fish and the PAC2 clones (Fig 3A, lanes 5-8, Fig 3B, lanes 1-6). In addition, each lane has a second band of variable size, presumably representing the 3' junction fragment. The variation in size between these junction fragments suggests that the sites of proviral integration were distinct in each case.

The fact that approximately 50-100 infectious units were injected into each embryo suggested that multiple integration events could readily have occurred in each embryo. To determine whether or not a founder fish transmitted more than one proviral integration through its germ line the insertions of three different transgenic F1's from a single founder (m4) were compared by Southern blot. Genomic DNA was digested with *Hind III* and probed as described earlier. As

Figure 3

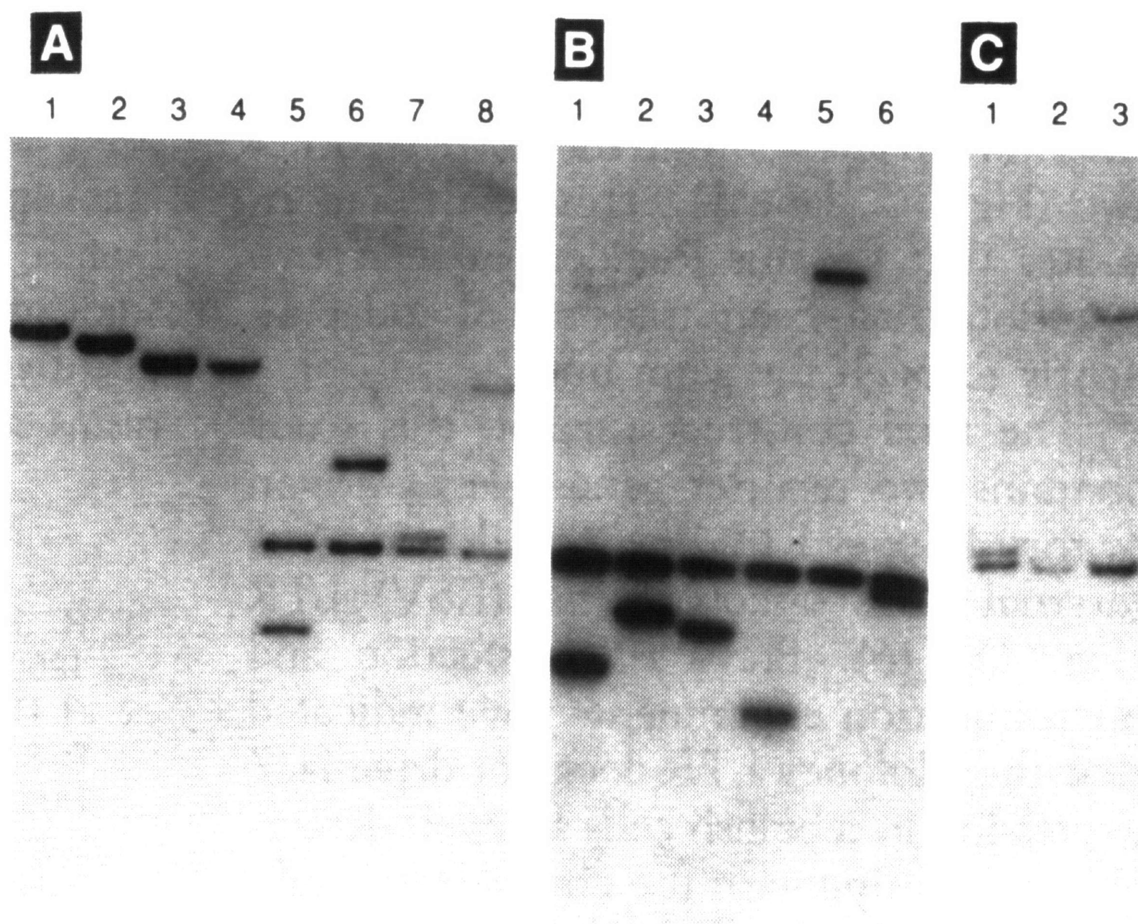


FIG. 3 Confirmation of proviral DNA integration by Southern blot analysis.
(A) Southern blot analysis of genomic DNA from two clones of zebrafish PAC2 cells generated by virus infection (lanes 1, 2, 5, and 6) and from two transgenic F1 fish from founders m4 (lanes 3 and 7), and f13 (lanes 4 and 8). Genomic DNA samples were digested with *Dra I* (lanes 1-4) and *Hind III* (lanes 5-8), and were probed with the *Cla I* fragment of pLZRNL (shown in Fig. 1). (B) Southern blot analysis of *Hind III* digests of genomic DNA from the transgenic F1 progeny of founders f6, f12, f36, f44, f45, m49 (lanes 1-6, respectively). (C) Genomic DNA from three transgenic F1 progeny of a single founder fish was digested with *Hind III* and probed with the *Cla I* fragment of pLZRNL. The hybridization pattern in lane 1 is different from those in lanes 2 and 3 indicating that the genomic sites of integration are different.

expected, all three F1's had the internal 3.7 Kb band as well as a 3' junction fragment. A comparison of the junction fragments, however, indicated that two of the fish had the same insertion (Fig 3C, lanes 2 and 3), but that the third fish had an insertion different from that of the other two (Fig 3C, lane 1). This result demonstrates that for this founder, at least two insertions had been transmitted independently through the germ line.

The viral construct used in this study contains two reporter genes, *lacZ*, driven by the Moloney LTR, and *neo*, driven by an internal RSV LTR (Fig. 1) (14). Previous transfection and microinjection experiments have indicated that the Moloney LTR does not drive *lacZ* expression in zebrafish cells at levels detectable by staining with the chromogenic substrate, X-gal (N. Gaiano, S. Lin and N. Hopkins, unpublished data). Therefore, we were not surprised to find that neither PAC2 clones infected with LZRNL(G) nor embryos containing the LZRNL transgene stained blue with X-gal in our present studies.

DISCUSSION

The results presented here demonstrate that the LZRNL(G) pseudotyped virus can be used to infect zebrafish embryos and to generate transgenic zebrafish. Additional technological advances promise to make the infection of zebrafish embryos with this class of viral vectors an extremely powerful tool for the study of vertebrate development. Such advances should include the identification of suitable promoter and promoter-enhancer combinations for gene expression in zebrafish cells.

We are interested in using this class of pseudotyped vector to generate insertional mutants in zebrafish. Mutagenesis in zebrafish has typically been performed using gamma rays or chemical mutagens such as *N*-ethyl-*N*-nitrosourea

(ENU) (15-17). However, a drawback of these mutagenesis methods is that the genes identified cannot be cloned readily because of the current absence of high-resolution genetic and physical maps of the zebrafish genome. Genes mutated by insertional mutagenesis would be more amenable to cloning since the integrated exogenous DNA would provide a 'tag' for the insertion site.

One possible strategy for studies of insertional mutagenesis would be to generate as many transgenic lines as possible, breed them to homozygosity, and screen them for mutant phenotypes. Although this approach is currently possible, it would be made more feasible by an increase in transgenic frequencies. Such an increase might be achievable by injecting a higher titer virus stock earlier in embryonic development. The stock of virus used in this study was more than ten-fold lower in titer than stocks which have been obtained with other vectors (11).

These results suggest that the advantages of retroviral vector technology, which have been well-documented in the mouse and chicken, will now be applicable to the zebrafish. This advance should accelerate the study of zebrafish development, and consequently help to elucidate the general mechanisms controlling vertebrate development.

MATERIALS AND METHODS

Preparation of viral stocks: A 293-derived cell line stably expressing the MoMLV *gag* and *pol* genes as well as the LZRNL genome was transiently transfected with a plasmid encoding the VSV G-protein driven by the human cytomegalovirus promoter. Virus containing supernatant was collected 48-72 hours later and concentrated as previously described (11). Concentrated pseudotyped virus was resuspended in TNE (50 mM Tris-HCl, pH 7.8 / 130 mM NaCl / 1 mM EDTA) containing 8μg/ml polybrene (Sigma).

Titering of viral stocks: Concentrated stocks of LZRNL(G) were diluted and used to infect NIH 3T3's, and zebrafish PAC2 cells for three hours in the presence of 8 μg/ml of polybrene. Approximately 4 hours after the completion of the infection the infected cells were trypsinized, serially diluted into both selective media and nonselective media, and plated. The plates were crystal violet stained 10-14 days later and the number of colonies was counted. The PAC2 cell line was derived from 24 hr. old embryos and is maintained in Lebowitz-15 media supplemented with 15% FBS and 5% zebrafish embryo extract (P. Culp and N. Hopkins, unpublished data).

Embyro injections: For injection, dechorionated eggs were incubated in Holtfreter's solution (0.9 mM $CaCl_2$ / 60 mM NaCl / 0.65 mM KCl / 2.4 mM $NaHCO_3$ / pH to 7.4 with HCl) at 27°C for about 4-5 hours and the late blastula-stage embryos were injected with a total of 10-20 nl of virus into multiple locations in each embryo using a glass needle and a dissecting microscope.

PCR detection of founder fish: DNA was extracted from pools of 50-100 F1 embryos at 24 hours of development, or from individual fish by incubation for 4-12 hours at 55°C in a lysis buffer (10 mM Tris-HCl pH 8.0 / 10 mM EDTA / 100 mM NaCl / 0.4% SDS / 0.2 mg/ml proteinase K). DNA was precipitated by ethanol and dissolved in TE (pH 8.0). Approximately 10 ng of DNA was used for PCR using AmpliTaq Polymerase (Perkin Elmer, Cetus). The reaction was carried out at 94°C/30", 60°C/45", 72°C/60" for 32 cycles with an initial 2 minute denaturation step at 94°C. The two primers used to detect the presence of LZRNL DNA sequence yield a 290 bp PCR product. The 5' primer (P1) is 5'GGGAATGTAGTCTTATGCAATAC3'. The 3' primer (P3) is 5'GCACACCAATGTGGTGAATGGTC3'. A pair of internal control primers homologous to the zebrafish Wnt5A gene (5'CAGTTCTCACGTCTGCTAC-

TTGCA3' and 5'ACTTCCGGCGTGTTGGAGAATTC3') or to the ZF21 gene (5'GAA-CTAGCAGCAGCGCTATGAAC3' and 5'ATGTAGTTTCCTCATCCAAGGG3') was included in each reaction.

Southern blot analysis: Genomic DNA from PAC2 clones or individual fish was digested with the appropriate restriction enzymes, was run through a 0.8% agarose gel using electrophoresis, and was then transferred to Hybond N^+ membrane (Amersham). Hybridizations were carried out as suggested by the vender in the presence of a probe labeled with ^{32}P-α-ATP using a random primed labeling kit (Boehringer Mannheim). The probe was derived from pLZRNL by *Cla I* digestion (see Fig. 1).

REFERENCES

1. Jaenisch R (1976) Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. *Proc Natl Acad Sci USA* **73**: 1260-4.

2. Mulligan RC (1993) The basic science of gene therapy. *Science* **260**: 926-32.

3. Soriano P; Jaenisch R (1986) Retroviruses as probes for mammalian development: allocation of cells to the somatic and germ cell lineages. *Cell* **46**: 19-29.

4. Huszar D; Sharpe A; Hashmi S; Bouchard B; Houghton A; Jaenisch R (1991) Generation of pigmented stripes in albino mice by retroviral marking of neural crest melanoblasts. *Development* **113**: 653-60.

5. Cepko CL; Ryder EF; Austin CP; Walsh C; Fekete DM (1993) Lineage analysis using retrovirus vectors. *Methods Enzymol* **225**: 933-60.

6. Gridley T; Soriano P; Jaenisch R (1987) Insertional mutagenesis in mice *Trends Genet* **3**: 162-6.

7. Friedrich G; Soriano P (1991) Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev* **5**: 1513-23.

8. von Melchner H; DeGregori JV; Rayburn H; Reddy S; Friedel C; Ruley HE (1992) Selective disruption of genes expressed in totipotent embryonal stem cells. *Genes Dev* **6**: 919-27.

9. Chen J; DeGregori J; Hicks G; Roshon M; Scherer C; Shi E; Ruley HE (1994) Gene trap retroviruses *Meth Mol Genet* **4**: 123-140.

10. Kimmel CB (1989) Genetics and early development of zebrafish. *Trends Genet* **5**: 283-8.

11. Burns JC; Friedmann T; Driever W; Burrascano M; Yee JK (1993) Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. *Proc Natl Acad Sci USA* **90**: 8033-7.

12. Roe T; Reynolds TC; Yu G; Brown PO (1993) Integration of murine leukemia virus DNA depends on mitosis. *EMBO J* **12**: 2099-108.

13. Varmus HE (1982) Form and function of retroviral proviruses. *Science* **216**: 812-20.

14. Xu L; Yee JK; Wolff JA; Friedmann T (1989) Factors affecting long-term stability of Moloney murine leukemia virus-based vectors. *Virology* **171**: 331-41.

15. Grunwald D; Striesinger G (1992) Induction of recessive lethal and specific locus mutations in the zebrafish with ethyl nitrosourea. *Genet Res* **59**: 103-16.

16. Solnica-Krezel L; Schier AF; Driever W (1994) Efficient recovery of ENU-induced mutations from the zebrafish germline. *Genetics* **136**: 1401-20.

17. Mullins M; Hammerschmidt M; Haffter P; Nüsslein-Volhard C (1994) Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. *Curr Biol* **4**: 189-202.

CHAPTER 3

HIGHLY EFFICIENT GERM-LINE TRANSMISSION OF PROVIRAL INSERTIONS IN ZEBRAFISH

ABSTRACT

An important technology in model organisms is the ability to make transgenic animals. In the past, transgenic technology in zebrafish has been limited by the relatively low efficiency with which transgenes could be generated using either DNA microinjection or retroviral infection. Previous efforts to generate transgenic zebrafish using retroviral vectors utilized a pseudotyped virus with a genome based on the Moloney murine leukemia virus and the envelope glycoprotein of the vesicular stomatitis virus. This virus was injected into blastula-stage zebrafish and 16% of the injected embryos transmitted proviral insertions to their offspring, with most founders transmitting a single insertion to about 2% of their progeny. In an effort to improve this transgenic frequency we have generated pseudotyped viral stocks of two new Moloney-based genomes. These viral stocks have titers up to two orders of magnitude higher than that used previously. Injection of these viruses resulted in a dramatic increase in transgenic efficiency: over three different experiments 83% (110/133) of the injected embryos transmitted proviral insertions to 24% of their offspring. Furthermore, founders for one of the viruses transmitted an average of 11 different insertions through their germ line. These results represent a 50- to 100-fold improvement in the efficiency of generating transgenic zebrafish, making it now feasible for a single lab to rapidly generate tens to hundreds of thousands of transgenes. Consequently, large-scale insertional mutagenesis strategies, previously limited to invertebrates, may now be possible in a vertebrate.

INTRODUCTION

Traditionally, the generation of transgenic zebrafish has been achieved by the microinjection of plasmid DNA into the cytoplasm of the one-cell stage embryo (1-3). Although this method is useful, efficiency is variable, and transgenes are frequently present in tandem arrays and can have complex unpredictable structures (1, 3). More recently retroviral infection has emerged as a method for generating transgenic zebrafish (4). In initial studies, Burns *et al.* demonstrated that a pseudotyped retroviral vector, containing a genome based on the Moloney murine leukemia virus (MLV) and the envelope glycoprotein (G-protein) of the vesicular stomatitis virus (VSV), was able to infect a cultured zebrafish cell line (5). This result was important because previously the host range of the standard retroviral vectors did not permit infection of fish cells (5), and as a result the zebrafish was inaccessible to retroviral vector technology.

Subsequently, our laboratory showed that retroviruses pseudotyped with the VSV G-protein are able to infect the zebrafish germ line following injection of a concentrated stock of an MLV/VSV pseudotyped virus into blastula-stage zebrafish embryos (4) (see Chapter 2). In these studies, 16% (8/51) of the potential founders tested transmitted proviral insertions to 2-3% of their F1 progeny, with founders transmitting 1-2 different insertions. These results suggested that pseudotyped retroviral vectors could be useful tools for generating transgenic zebrafish, and that if the transgenic frequency could be increased substantially that they might also prove to be effective insertional mutagens.

To determine if we could improve the efficiency of generating transgenic fish using retroviral vectors, we constructed two new MLV-based genomes and generated viral stocks from these constructs with titers up to two orders of magnitude higher that the previously used viral stock (4). Injection of these new

viral stocks into blastula-stage embryos resulted in as much as a 50- to 100-fold increase in the efficiency of generating transgenic insertions as compared to previously obtained results using either plasmid microinjection (3) or retroviral infection (4). These results suggest that the efficiency of generating transgenic zebrafish using pseudotyped retroviral vectors is correlated with the titer of the viral stock *in vitro*, and that at the highest titer tested to date, the germ line of every injected fish can harbor many different proviral integrations. Consequently it is now feasible for a small lab to generate tens to hundreds of thousands of proviral transgenes in zebrafish. This work represents a major advance in transgenic technology in zebrafish, and may make large-scale insertional mutagenesis and the rapid identification of phenotypically interesting genes possible in this vertebrate system.

RESULTS

High Frequency Germ-Line Transmission of Proviral Integrations

Two pseudotyped viruses, SFG(G) and NK(G), were constructed for these studies and have MLV-based genomes (see Fig. 1A) and an envelope containing the VSV G-protein. These viruses are similar to the MLV/VSV pseudotyped virus previously shown to be capable of stably integrating proviral DNA into the zebrafish genome (4, 5). Concentrated stocks of SFG(G) and NK(G) were prepared from stable producer cell lines, and were titered on both mouse 3T3 cells and zebrafish PAC2 cells. Because of complications in determining the titer of these viruses on PAC2 cells (see Materials and Methods) only the titers on 3T3 cells will be given hereafter.

The SFG(G) virus stock (2×10^9 cfu/ml on 3T3 cells) was microinjected into zebrafish embryos at about the 1000-cell stage. Following microinjection the embryos were incubated at either 26°C or 28°C. Although many embryos (50-80%)

Figure 1

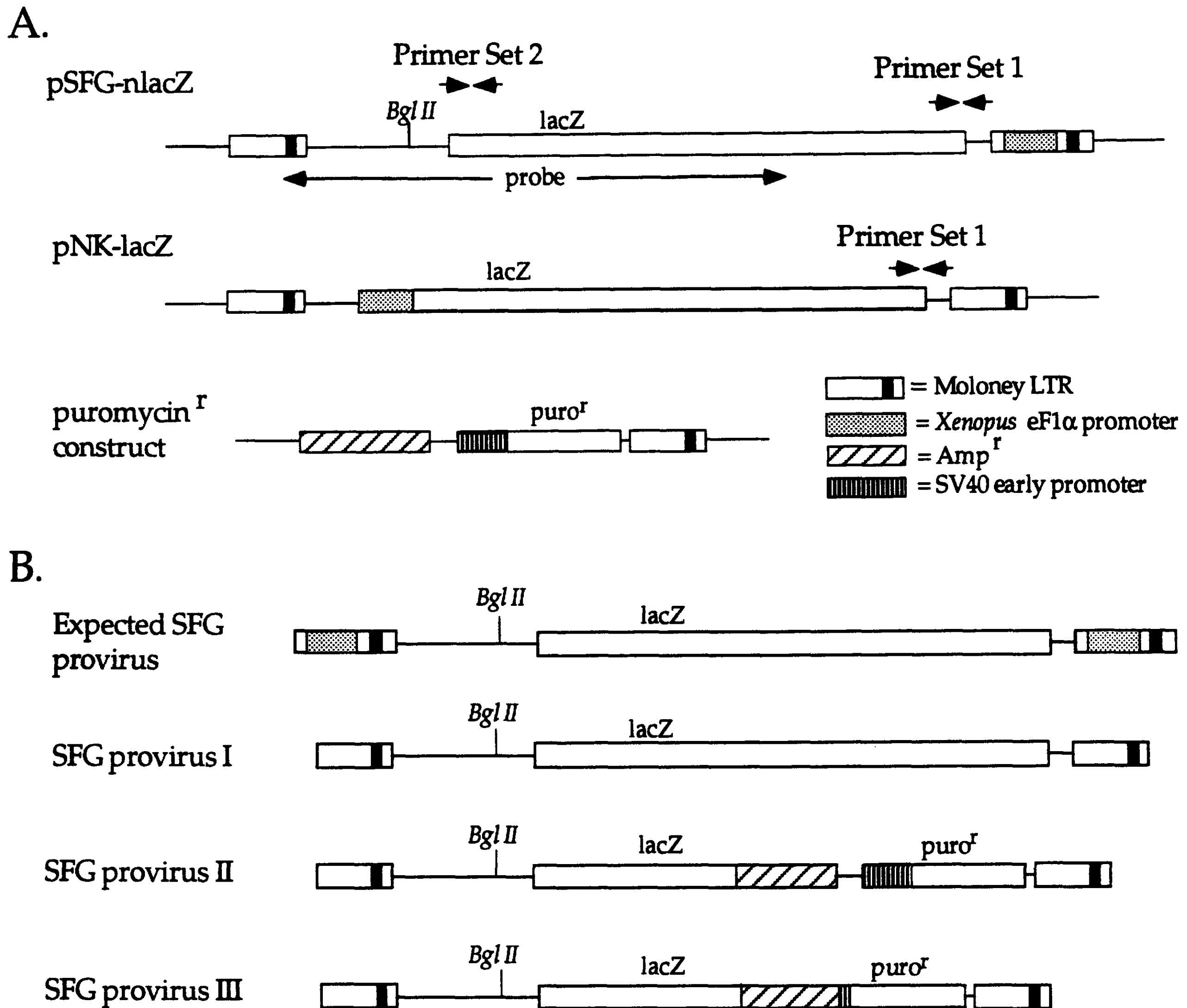


FIG. 1 Schematic representation of the plasmids used, and of the predicted and observed proviral structures. (A) Maps of pSFG-nlacZ (with 3'-LTR modified to contain the *Xenopus* eF1α promoter), pNK-lacZ, and of the construct used to confer resistance to puromycin. The locations of primer sets 1 and 2 are indicated, as is the region of pSFG-nlacZ used as a probe for Southern blots. (B) The expected SFG proviral genome and the actual SFG proviral genomic structures.

did not survive to the next day, or were malformed, the majority of those that appeared normal at 24 hours grew to adulthood. To detect germ-line transmission of proviral DNA, the injected embryos were raised to adulthood and mated, and genomic DNA from pools of 24-hour-old F1 embryos was tested for the presence of proviral sequences by PCR. As shown in Table 1, 90/106 (85%) of the potential SFG founders tested (experiments 1 and 2) were found to transmit proviral DNA to their F1 progeny.

Although aliquots of the same virus stock were used in both SFG(G) experiments listed in Table 1, the frequency of germ-line transmission in the first experiment, 71% (40/56), is significantly lower than the 100% (50/50) obtained in the second experiment. The primary difference between these two experiments was the temperature the embryos were incubated at after injection, suggesting that viral infection occurs more efficiently at 28°C than at 26°C. We have obtained similar results which support this conclusion using two other MLV/VSV viruses (N. Gaiano, M. Allende, and N. Hopkins, unpublished data).

Initial injections with an undiluted stock of the virus NK(G) (2×10^9 cfu/ml on 3T3 cells), resulted in all injected embryos being dead or severely malformed by the next day. Injection of four-fold dilutions of the concentrated NK(G) stock resulted in survival rates similar to those observed using SFG(G) and were used to generate the potential NK founders. Of 27 potential NK founders tested, 20 (74%) were found to transmit proviral DNA to their F1 progeny (Table 1).

The toxicity observed upon injection of some viral stocks into blastula-stage embryos could be a function of the fusogenic nature of the VSV G-protein (6). Due to the inherent variability in the efficiency of transient transfection of the VSV-G expressing plasmid, a step required to make MLV/VSV pseudotyped viruses (see ref. 7), the amount of VSV G-protein may vary between virus stocks. Such variability could account for the difference in toxicity seen between the SFG(G) and NK(G)

Table 1. High-frequency germ-line transmission of proviral insertions.

Expt.	Virus Injected	Temp†	Titer on 3T3's (cfu/ml)††	Potential Founders Tested	Founders Identified
1	SFG(G)	26 °C	2×10^9	56	40 (71%)
2	SFG(G)	28 °C	2×10^9	50*	50 (100%)
3	NK(G)	28 °C	5×10^8	27	20 (74%)

Injected embryos were raised, genomic DNA was isolated from pools of their F1 progeny and was tested by PCR for the presence of proviral DNA. In some cases, F1 pools were not tested in this manner but instead F1 embryos were raised and tested by isolating DNA from fin clips.

† The temperature at which embryos were incubated for 24 hours after injection.

†† The titer on 3T3 cells is shown for comparison because an accurate determination of SFG(G) titer on PAC2 cells was complicated by limited or no *lacZ* expression (see Materials and Methods).

* In experiment 2 a total of five potential founders with between 8 and 21 F1 progeny tested were found to be negative but were considered inconclusive, due to the small number of progeny tested, and are not listed.

viruses used here, although contaminants in the viral stocks, derived from the producer cells used to make the virus, could also be responsible.

Germ-Line Mosaicism of Founders

To identify individual transgenic F1 fish without sacrificing them, genomic DNA was isolated from caudal fin clips of 6-8 week-old F1 progeny of identified founders, and tested by PCR for the presence of proviral DNA. As shown in Fig. 1B and as discussed below, the SFG(G) virus stock contains a mixture of three viral genomes. Primer set 2 (see Fig. 1A), which detects all three SFG proviruses, was used to detect F1's transgenic for SFG proviruses. Primer set 1 (see Fig. 1A) was used to detect F1's transgenic for NK proviruses.

The percentage of transgenic offspring from SFG founders ranged from 12-70% with an average of 29% (Table 2). The percentage of transgenic offspring from NK founders was somewhat lower and ranged from 3-28% with an average of 13%. A plausible explanation for the difference in both the frequency of germ-line transmitting founders (see above) and the frequency of transgenic F1's from these founders between the SFG(G) and NK(G) injections would be that the SFG(G) stocks used for injection were 4-fold higher in titer than the NK(G) stock used (see Table 1).

SFG Founders Transmit an Average of 11 Proviral Insertions to Their F1 Progeny

In order to examine the number of different insertions being transmitted through the germ line of the founder fish, Southern blot analysis was performed on genomic DNA from fin clips of individual transgenic F1's from SFG founders. The DNA was digested with *Bgl* II, which cuts once within the all three SFG proviral genome types (see Fig. 1B, and below), and analyzed by Southern blot. Depending upon the location of *Bgl* II sites in the genomic DNA adjacent to the insertion, each different insertion was expected to yield two junction fragments of diagnostic sizes.

Table 2. Analysis of germ-line transmission of proviral insertions from individual SFG and NK founders

Founder	Transgenic F1 progeny (%)	# of insertions transmitted to F1's
SFG1	12%	—
SFG8	39%	—
SFG17	23%	—
SFG26	19%	—
SFG35	70%	11
SFG48	23%	5
SFG49	28%	7
SFG51	39%	7
SFG52	36%	12
SFG54	21%	6
SFG57	20%	14
SFG59	56%	14
SFG62	24%	12
SFG64	12%	12
SFG66	33%	10
SFG67	30%	10
SFG69	25%	19
SFG73	17%	9
SFG77	57%	22
SFG80	21%	6
SFG81	13%	6
SFG89	28%	12
Average:	**29%**	**11**
NK3	24%	—
NK4	24%	—
NK5	3%	—
NK6	8%	—
NK7	7%	—
NK10	4%	—
NK11	28%	—
NK12	12%	—
NK16	15%	—
NK20	9%	—
Average:	**13%**	—

Transgenic F1's were identified by isolating genomic DNA from caudal fin clips and testing by PCR for proviral sequences. Genomic DNA from identified transgenic F1 fish was then digested with *Bgl* II which cuts once in the proviral sequence, and Southern blot analysis was performed. Junction fragment sizes were compared between fish and those with identical patterns were classified as having the same insertion.

An example of a Southern blot used to compare insertions is shown in Fig. 2. The proviral insertions in 19 transgenic offspring from founder SFG77 were compared. Among these fish, 14 have one insertion, 4 have two insertions, and 1 has three insertions (lane 6) with some insertions being present in more than one fish. In total these fish harbor 16 different insertions among them.

Progeny from 18 outcrossed founders were analyzed by Southern blot and were found to contain 194 different insertions, indicating that on average each founder transmits 11 different insertions to its F1 progeny (see Table 2). Although the majority of transgenic F1 fish (65%) were found to have 1 proviral insertion each, individual F1 fish were frequently found with 2, 3 or 4 different proviral insertions (26%, 7%, and 2%, respectively). The mosaicism of individual insertions in the germ line of founders varied, with some insertions being present in <1% of the F1 progeny from a given founder, and others being present in as much as 14% of the F1 progeny from a given founder. Of 187 insertions analyzed roughly 30% were found to be transmitted to at least 3% of the founder's progeny.

Analysis of SFG Proviral Genome Structures

As mentioned above, the SFG(G) virus stock contains three different viral genomes. To examine the structures of these proviruses in transgenic fish, genomic DNA from individual fish harboring the different proviruses was digested with various combinations of the following restriction enzymes: *Xba I, Pvu II, Eco RV, Eco RI, Bam HI, Nhe I, Hinc II, Sac I, Hind III, Nco I, Sph I,* and *Dra I.* Based on Southern blot analysis of these digests, restriction maps were constructed which were consistent with every digest examined. Structural maps of the three SFG proviral genomes (SFG provirus types I, II, and III) are shown in Fig. 1B. All three SFG proviral genomes were found to have wild-type MLV LTR's. This result was surprising because based on the plasmid used to generate the SFG producer cell line

Figure 2

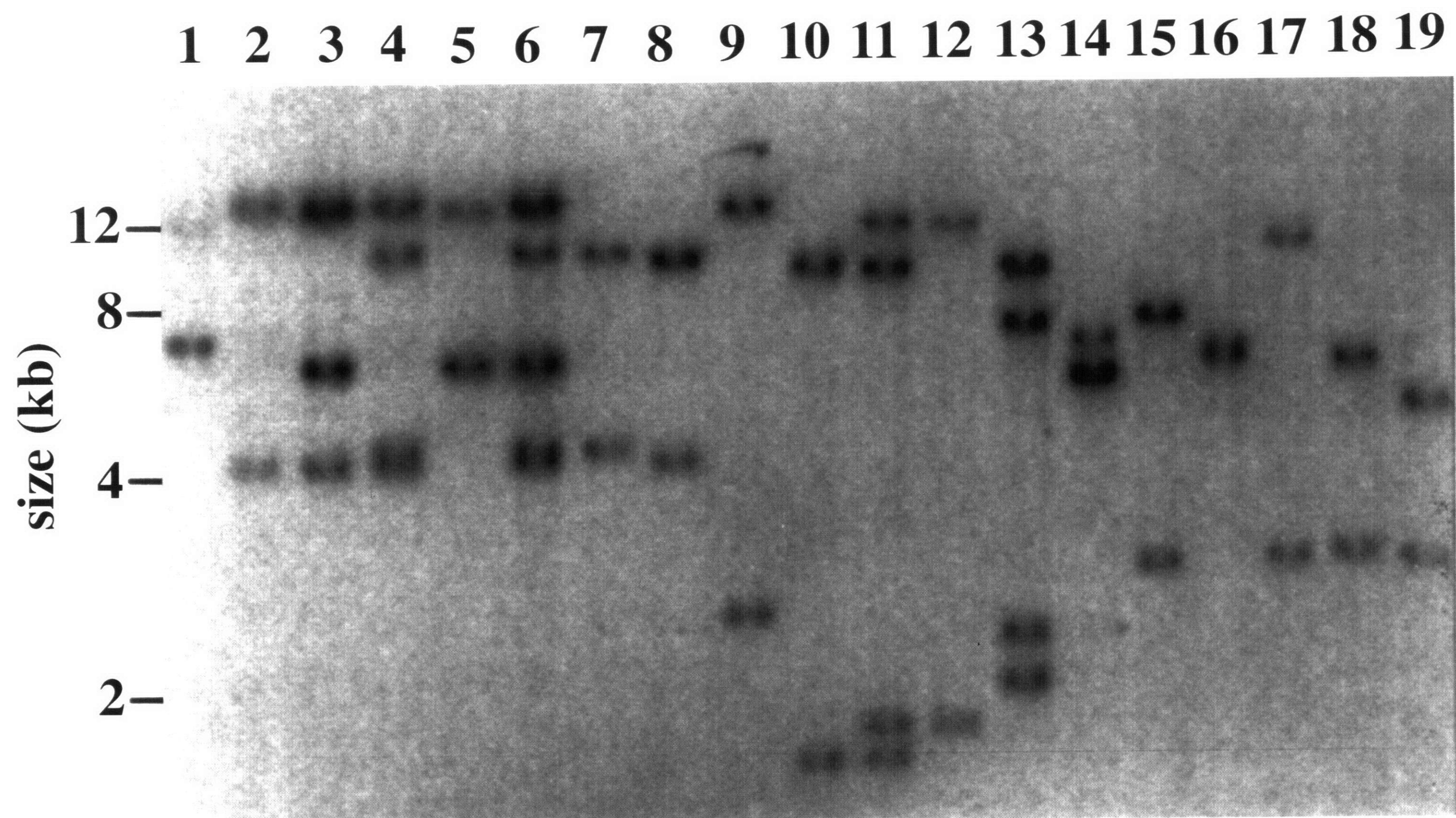


FIG. 2. Southern blot analysis of genomic DNA from transgenic F1 progeny of founder SFG77. DNA was digested with *Bgl* II and probed with the sequence indicated in FIG. 1A. Each insertion is expected to produced two junction fragments with sizes characteristic of the site of integration. The result of segregation during meiosis is apparent in lanes 2-7, and 10-12. For example, the F1 represented in lane 11 has four bands representing 2 insertions. These two insertions can be seen independently in the F1's represented in lanes 10 and 12.

(pSFG-nlacZ, see Fig. 1A), SFG proviruses were expected to possess modified LTR's with the *Xenopus* eF1α promoter in the U3 region (see Fig. 1B). In addition to the unexpected LTR structure found, two of the three SFG proviral genomes (type II and III) were found to have large deletions in the *lacZ* gene and to contain sequences from the puromycin resistance construct which had been used to allow drug selection of stable producer clones (see Materials and Methods, and Fig. 1A). The absence of the desired viral genome in the SFG(G) stocks, and the presence of the three different viral genomes in these stocks is likely to be the result of DNA rearrangements which occurred during the transfection of pSFG-nlacZ and the puromycin resistance construct into the 293-derived packaging cell line.

Analysis of genomic sequences at the sites of SFG proviruses

A nonrandom bias in the selection of proviral integration sites could significantly impact the utility of retrovirus-mediated insertional mutagenesis. Studies in both the mouse and chick have suggested that retroviruses integrate either randomly or with a preference for genes sequences, either of which would be acceptable for insertional screens (8-10). However, because the randomness of murine proviral insertions into the zebrafish genome has not been directly addressed it remains formally possible that such insertions might have a gross bias toward certain sequences, such as repetitive elements for example, and may not integrate into gene sequences.

To address this issue genomic DNA flanking 19 randomly selected SFG proviral transgenes was examined for evidence of insertion site bias. Using inverse PCR (11) genomic fragments adjacent to the 5'-LTR were amplified and subsequently subcloned and sequenced. The 19 fragments ranged from 172 bp to 1332 bp and their sequences (14,386 bp in total) were compared to each other, and used to search the sequence databases. No significant sequence similarity or obvious consensus

sequence was observed among the 19 fragments within 100 bp of the integration sites. Three fragments were found to contain dinucleotide repeats (2 CA, 1 TA), and two contained DANA elements, members of a family of short interspersed elements estimated to comprise 10% of the zebrafish genome (12). Based upon the published frequencies of such repetitive sequences in the zebrafish genome, the frequency with which these elements were found was consistent with what would be expected for random insertion site selection (12, 13). These results, together with an analysis of 23 additional junction fragments from LZRNL proviruses (K. Kawakami and N. Hopkins, unpublished data), suggest that proviral insertion site selection in zebrafish is not subject to gross bias. However, these studies are insufficient to rule out more subtle biases which might impact the extent to which all genes in the genome are targets for insertional mutagenesis using proviral insertions.

During the course of this analysis, four of the SFG flanking sequences were found to have homology to putative gene sequences in the public databases. While in three cases the homologies were weak and are of questionable significance, the fourth was a strong homology between the translated sequence adjacent to insertion SFG54A and the pig 3-oxoacid CoA transferase gene (31/35 amino acid identity) (14). Subsequent studies have indicated that this insertion lies two base pairs downstream of an exon/intron boundary in this putative fish gene (K. Kawakami and N. Hopkins, unpublished data) (Fig 3). Efforts to determine the effect of insertion 54A on the expression of this putative gene were complicated by the fact that the gene appears to be duplicated in the genome. Nevertheless this integration event supports the notion that proviral insertions into the zebrafish genome can land in gene sequences. Mutagenesis screens are the most direct way to determine if such insertions can disrupt gene function (see Chapters 4 and 5).

Figure 3

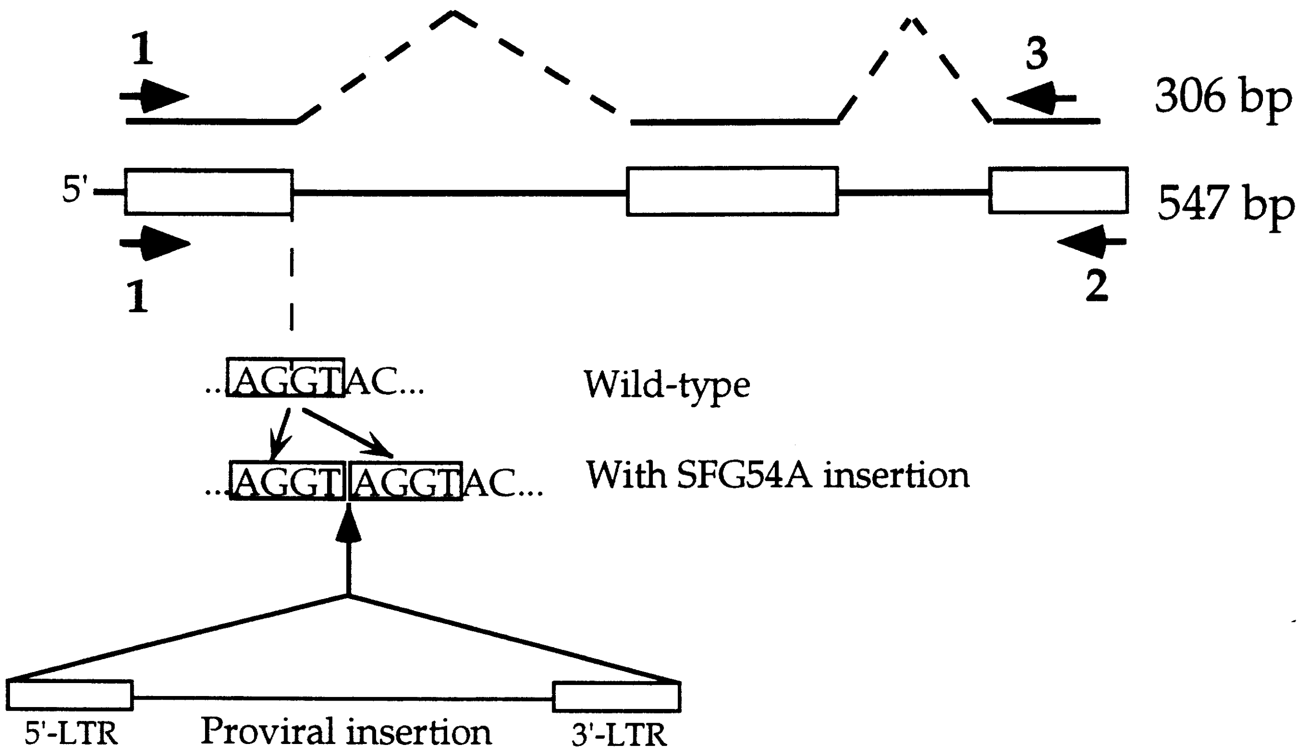


FIG. 3. Insertion SFG54A has integrated into a zebrafish copy of the 3-oxoacid CoA-transferase gene. Primers 1 and 2 were used to amplify a 547 bp genomic DNA fragments, and primers 1 and 3 were used to amplify a 306 bp RT-PCR product. Sequencing of these fragments confirmed the splicing events predicted based upon amino acid homology to the pig protein. The 3' junction of insertions 54A was also isolated and a 4 bp duplication was identified on either side of the provirus. Such a duplication is typical for MLV intergation events (15).

DISCUSSION

The results presented here demonstrate that MLV/VSV pseudotyped retroviral vectors can be used to generate transgenic zebrafish with extremely high efficiency. At previous transgenic frequencies, using either DNA microinjection or retroviral infection, the generation of 100 potential transgenic founders would typically result in the germ-line transmission of 10-20 transgenes (3, 4). The current work represents a dramatic improvement in transgenic frequency and shows that it is now possible, using retroviral infection, for 100 potential founders to result in the germ-line transmission of over 1,000 transgenes. One possible explanation for the increased transgenic frequency seen here as compared to previous results using the pseudotyped retrovirus LZRNL(G) is that the titers of both the SFG(G) and NK(G) stocks used here were roughly 100-fold higher than that of the LZRNL(G) stock used previously (4).

Based upon the transgenic frequencies reported here, large-scale insertional mutagenesis, a technique which has not been practical in vertebrates model systems, may now be possible in zebrafish. Although large-scale chemical mutagenesis can be performed in zebrafish (16, 17) and has yielded many interesting mutants, an insertional mutagenesis strategy could be a powerful alternative. This is because mutagenic insertions provide a molecular tag to facilitate the cloning of mutated genes, circumventing the laborious positional cloning methods often required to clone chemically mutated genes.

In fruit flies, large-scale insertional mutagenesis is possible because P-elements can be used to generate many thousands of insertions which can be screened for integration events of interest (18). While previously the generation of many thousands of transgenic insertions was theoretically possible in vertebrate systems such as the mouse and the zebrafish, the resources and time required to do

so were prohibitive. The present work, however, indicates that it is now feasible for a single lab to rapidly produce as many as 100,000-200,000 transgenes in zebrafish. This could be achieved by generating 10,000-20,000 founders, each of which would transmit 10 insertions to its F1 progeny as shown here (Table 2). We estimate that it would take 4-6 people roughly three months to generate the founders.

Based upon the size of the zebrafish genome (1.6×10^9 bp), a screen involving 200,000 insertions would have, on average, one insertion every 8 Kb. If the average gene spans roughly 10 Kb, then a screen of this size would be expected to have potentially mutagenic insertions into most of the genes in the genome. The actual efficiency of such a screen would depend upon whether or not proviruses integrate randomly into the zebrafish genome and upon the mutagenicity of proviral DNA integrated into zebrafish genes. If, for example, proviral insertions into the zebrafish genome preferentially occur into intergenic sequences, then the likelihood of integrating into and mutating genes would be greatly reduced. However, studies of retroviral integration in the mouse and chicken suggest that proviral insertions occur either at random (8) or possibly with a preference for transcribed regions of the genome (9, 10). Furthermore, a comparison of a total of 42 proviral insertion sites in zebrafish has indicated that there is no gross integration site bias and that insertions may integrate into transcribed regions of the genome (see above).

There are several possible ways a large-scale insertional mutagenesis screen in zebrafish might be conducted. One strategy would simply be to generate insertions, breed them to homozygosity, and screen for mutant phenotypes. Although such a screen is labor intensive and limited by the time and space required, it should be possible to screen several thousand insertions in this manner. It is unknown how many mutants would be generated in a screen of this size, because the number of proviral insertions into the zebrafish genome required to produce a mutant phenotype remains to be determined. We have conducted a pilot screen to

determine this number (see Chapter 4). In mice 5% of proviral insertions disrupt essential genes (19).

The inbreeding strategy mentioned above might be an effective way to isolate insertional mutants, although it is limited by the need to maintain very large numbers of individual lines. As a result, using such a strategy a lab could only screen a small fraction of the hundreds of thousands of transgenes that could be generated. An alternative strategy, which would permit the screening of a much larger number of insertions, would be to screen haploid embryos. Haploid zebrafish embryos are easy to generate, and undergo relatively normal early development (20). Transgenic F1 fish, hemizygous for proviral insertions, could be used to generate haploid F2 embryos to be screened for mutant phenotypes. A phenotype observed in 50% of the haploid embryos would indicate the presence of a mutagenic insertion.

Another possible approach to insertional mutagenesis in zebrafish could utilize retroviral gene traps. Gene traps are constructs containing a reporter gene which can only be expressed after integration into a transcribed endogenous gene (21). Such 'activated' integrations are likely to disrupt the function of the gene into which they have integrated and typically express the trap reporter in a temporal and spatial pattern similar to that of the endogenous gene (22, 23). The generation of 20,000 founders, which could be maintained in 100 15-gallon fish tanks, would permit the screening of 200,000 gene trap insertions. F1 embryos from founder matings could be screened for trap activations and those containing expression patterns of interest could be isolated for further study. If gene traps are activated in fish cells at efficiencies similar to those in mice (21), then 200,000 gene trap insertions in zebrafish could contain as many as 20,000 activated trap patterns, a significant percent of which would be expected to have a mutant phenotype when bred to homozygosity (22).

MATERIAL AND METHODS

Plasmids Used to Make Retroviral Constructs: pSFG(ECT-) (obtained from R. Mulligan, MIT) contains deletions in the U3 region of the 3'-LTR which remove the MLV transcriptional regulatory elements. The *Xenopus* eF1α promoter (24) was placed into the 3'-LTR and the resulting construct, pSFG-nlacZ (Fig. 1A), was expected to generate proviruses with this promoter driving a nuclear localized *E. coli β-galactosidase* (*lac*Z) gene (Fig. 1B). pNK-lacZ (Fig. 1A) was derived from pLZRNL (25). The eF1α promoter was placed upstream of the *lacZ* gene, and the RSV LTR and neomycin phosphotransferase gene were removed.

Generation of Stable Retroviral Producer Clones: pSFG-nlacZ (SFG), and pNK-lacZ (NK) were each transfected into a 293 gag-pol packaging cell line (293GP; obtained from Viagene, Inc.) with a construct containing the puromycin acetyltransferase gene driven by the SV40 early promoter and with an MLV LTR providing the polyadenylation signal (Fig. 1A). The packaging cell line used does not express any envelope protein but does express the gag-pol protein required to make infectious retroviral core particles (5). Puromycin resistant cell clones were screened for virus production by transient transfection of a construct expressing the VSV G-protein from the human cytomegalovirus promoter and subsequent titering on mouse 3T3 cells. This construct, pHCMV-G (7), provides the envelope protein necessary to produce infectious pseudotyped virus, and such virus is indicated with the designation (G).

Titering was performed by infecting mouse 3T3 cells and zebrafish PAC2 cells (4, P. Culp and N. Hopkins, unpublished data) with serial dilutions of virus and then staining 48 hours later with the chromogenic substrate X-gal to detect β-gal

activity in infected cells. An SFG(G) producing clone was identified which produced titers of 5-10 x 10^6 cfu/ml on 3T3 cells, and 1 x 10^4 cfu/ml on zebrafish PAC2 cells. The SFG(G) virus was unexpectedly found by Southern blot to be a mixture of three different viral genomes, representing recombinants of the plasmids used to construct the virus-producing cell line (see Results). An NK producer clone was identified which produced *lacZ* titers of 5-10 x 10^6 cfu/ml on 3T3 cells, and 1 x 10^5 cfu/ml on PAC2 cells. Southern blot analysis indicated that virus from this clone produces the expected proviral genome (data not shown).

Virus-containing supernatant from the selected SFG and NK clones was concentrated as previously described (5, 7) to *lacZ* titers of 1-2 x 10^9 cfu/ml on 3T3 cells. The *lacZ* titers of the concentrated SFG(G), and NK(G) stocks on PAC2 cells were 5-10 x 10^6 cfu/ml, and 1-2 x 10^8 cfu/ml, respectively. Previous studies in our lab have indicated that *lacZ* may be a substantially less effective reporter in zebrafish PAC2 cells than in mouse 3T3 cells (P. Culp and N. Hopkins, unpublished data). Therefore the relative titer of these stocks on 3T3 and PAC2 cells was estimated by comparing the amount of integrated proviral DNA in both cell types after infection with the same dilution of a virus stock. For both SFG(G) and NK(G) the amount of integrated proviral DNA in PAC2 cells was found to be roughly 2-fold less than that in 3T3 cells (data not shown). The *lacZ* titers of the concentrated SFG(G) and NK(G) stocks used for injection into embryos were 2 x 10^9 cfu/ml on 3T3 cells and were therefore estimated to be roughly 1 x 10^9 cfu/ml on PAC2 cells.

Generation and Identification of Transgenic Founder Fish: Ten to twenty nanoliters of the concentrated SFG(G) and NK(G) viral stocks, containing 8 μg/ml polybrene, were injected into 4-5 locations among the blastomeres of blastula-stage zebrafish embryos (roughly 512-2,000 cell stage). Injected embryos were raised to sexual maturity and mated either to each other or to wild-type fish. Genomic DNA was

prepared from pools of the F1 progeny as previously described (4), and was tested by the polymerase chain reaction (PCR) for the presence of proviral DNA. The nucleotide sequence of the primers used to detect both SFG and NK founders (primer set 1, see Fig. 1A) is as follows: The upstream primer sequence is 5'-ATATCGACGGTTTCCATATGGG-3' and is within the coding sequence of the *lacZ* gene. The downstream primer sequence is 5'-GTACTCTATAGGCTTCAGCTGG-3' and is within the MLV-derived sequences downstream of the *lacZ* gene. This set of primers amplifies a 232 bp sequence in SFG and about a 200 bp sequence in NK. Primers designed to detect sequences within the zebrafish Wnt5a gene were used as internal controls in each reaction and are the same as those previously described (4). The PCR program used is as follows: 1 minute at 94°C, 1 minute at 60°C, 1 minute at 72°C for 32 cycles, with an initial denaturation step at 94°C for 2 minutes, and a final elongation step at 72°C for five minutes.

<u>Identification of Transgenic F1 Fish</u>: The F1 progeny of founders were raised and individual genomic DNA samples were prepared from caudal fin clips by incubation in 10 mM Tris-HCl (8.0), 10 mM EDTA, 0.2 mg/ml proteinase K for at least 3 hours at 55°C. PCR was then used to detect the presence of proviral sequences. Primer set 1 (Fig. 1A) was used to identify transgenic progeny from NK founders, while a second set of primers, primer set 2 (see Fig. 1A), was used to identify transgenic progeny from SFG founders. Primer set 2 was designed when it was found that the SFG(G) stock generated three different proviruses, two of which were not detectable by primer set 1 (see Results). The nucleotide sequence of primer set 2 is as follows: The upstream primer sequence is 5'-ATCCTCTAGACTGCC-ATGG-3' and includes the start codon of the *lacZ* gene. The downstream primer sequence is 5'-ATCGTAACCGTGCATCTG-3' and is within the coding sequence of the *lacZ* gene. This set of primers amplifies about a 340 bp sequence. All identified

transgenic offspring from a single founder were kept together, and subsequently tail DNA was re-isolated from these fish for Southern Blot analysis.

Southern Blot Analysis: Genomic DNA was digested with the indicated restriction enzymes, electrophoresed through a 0.8% agarose gel and blotted to Hybond N^+ nylon membranes (Amersham). Radiolabeled probes were made using the Random Primed DNA Labeling Kit (Boehringer Mannheim). Hybridizations were carried out at 65°C in a Robbins Scientific Model 2000 Hybridization Incubator in a solution containing of 0.25 M Na_2HPO_4 (pH 7.4), 1 mM EDTA, 10 mg/ml BSA, and 7% SDS. Filters were washed 3 x 20 minutes with 0.1x SSC and 0.1% SDS at 65°C.

Inverse PCR: Genomic DNA from fin clips was isolated as described above. 5 µg of DNA was triple-digested overnight at 37°C with *Bgl II/Bam HI/Bcl I* all of which generate compatible overhangs. Digested samples were then phenol extracted and ethanol precipitated in the presence of 0.2 M NaCl. Ligations were performed with a DNA concentration of 2 µg/ml at 14°C overnight using New England Biolabs T4 DNA ligase and the recommended buffer. Ligations were ethanol precipitated in the presence of 2 M ammonium acetate and PCR was performed with 0.2-0.5 µg of ligated DNA.

REFERENCES

1. Stuart GW; McMurray JV; Westerfield M (1988) Replication, integration and stable germ-line transmission of foreign sequences injected into early zebrafish embryos. *Development* **103**: 403-12.

2. Stuart GW; Vielkind JR; McMurray JV; Westerfield M (1990) Stable lines of transgenic zebrafish exhibit reproducible patterns of transgene expression. *Development* **109**: 577-84.

3. Culp P; Nüsslein-Volhard C; Hopkins N (1991) High-frequency germ-line transmission of plasmid DNA sequences injected into fertilized zebrafish eggs. *Proc Natl Acad Sci USA* **88**: 7953-7.

4. Lin S; Gaiano N; Culp P; Burns J; Friedman T; Yee J-K; Hopkins N (1994) Integration and germ-line transmission of a pseudotyped retroviral vector in zebrafish. *Science* **265**: 666-9.

5. Burns JC; Friedmann T; Driever W; Burrascano M; Yee JK (1993) Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. *Proc Natl Acad Sci USA* **90**: 8033-7.

6. Wagner RR (1991) Rhabdoviridae and their replication, in *Fundamental Virology*, eds Fields BN; Knipe DM (Raven Press, Ltd., New york) pp 489-503.

7. Yee JK; Miyanohara A; LaPorte P; Bouic K; Burns JC; Friedmann T (1994) A general method for the generation of high-titer, pantropic retroviral vectors: highly efficient infection of primary hepatocytes. *Proc Natl Acad Sci USA* **91**: 9564-8.

8. Withers-Ward ES; Kitamura Y; Barnes JP; Coffin JM (1994) Distribution of targets for avian retrovirus DNA integration in vivo. *Genes Dev* **8**: 1473-87.

9. Rohdewohld H; Weiher H; Reik W; Jaenisch R; Breindl M (1987) Retrovirus integration and chromatin structure: Moloney Murine Leukemia proviral integration sites map near DNase I-hypersensitive sites. *J Virol* **61**: 336-43.

10. Scherdin U; Rhodes K; Breindl M (1990) Transcriptionally active genome regions are preferred targets for retrovirus integration. *J Virol* **64**: 907-12.

11. Ochman H; Medhora MM; Garza D; Hartl DL (1990) Amplification of flanking sequences by inverse PCR. In *PCR Protocols: A guide to Methods and Applications*, eds, Innis MA; Gefland DH, Sninsky JJ; White TJ (Academic Press, Inc, San Diego) pp 219-27.

12. Izsvak Z; Ivics Z; Garcia-Estefania D; Fahrenkrug SC; Hackett PB (1996) DANA elements: A family of composite, tRNA-derived short interspersed DNA elements associated with mutational activities in zebrafish (*Danio rerio*). *Proc Natl Acad Sci USA* **93**: 1077-81.

13. Goff DJ; Galvin K; Katz H; Westerfield M; Lander ES; Tabin CJ (1992) Identification of polymorphic simple sequence repeats in the genome of the zebrafish. *Genomics* **14**: 200-2.

14. Lin TW; Bridger WA (1992) Sequence of a cDNA clone encoding pig heart mitochondrial CoA transferase. *J Biol Chem* **267**: 975-8.

15. Coffin JM (1991) Retroviridae and their replication, in *Fundamental Virology*, eds Fields BN; Knipe DM (Raven Press, Ltd., New york) pp 645-708.

16. Mullins M; Hammerschmidt M; Haffter P; Nüsslein-Volhard C (1994) Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. *Curr Biol* **4**: 189-202.

17. Solnica-Krezel L; Schier AF; Driever W (1994) Efficient recovery of ENU-induced mutations from the zebrafish germline. *Genetics* **136**: 1401-20.

18. Cooley L; Kelley R; Spradling A (1988) Insertional mutagenesis of the *Drosophila* genome with single P elements *Science* **239**: 1121-8.

19. Jaenisch R (1988) Transgenic animals. *Science* **240**: 1468-74.

20. Westerfield M (1995) ed*The Zebrafish Book* (Univ. of Oregon Press, Eugene).

21. Chen J; DeGregori J; Hicks G; Roshon M; Scherer C; Shi E; Ruley HE (1994) Gene trap retroviruses *Meth Mol Genet* **4**: 123-140.

22. Friedrich G; Soriano P (1995) Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev* **5**: 1513-23.

23. Wurst W; Rossant J; Prideaux V; Kownacka M; Joyner A; Hill DP; Guillemot F; Gasca S; Cado D; Auerbach A; Ang SL (1995) A large-scale gene-trap screen for insertional mutations in developmentally regulated genes in mice. *Genetics* **139**: 889-99.

24. Johnson AD; Krieg PA (1994) pXeX, a vector for efficient expression of cloned sequences in Xenopus embryos. *Gene* **147**: 223-6.

25. Xu L; Yee JK; Wolff JA; Friedmann T (1989) Factors affecting long-term stability of Moloney murine leukemia virus-based vectors. *Virology* **171**: 331-41.

CHAPTER 4

INSERTIONAL MUTAGENESIS AND RAPID CLONING OF ESSENTIAL GENES IN ZEBRAFISH

ABSTRACT

Large-scale chemical mutagenesis screens in zebrafish have led to the isolation of thousands of lethal mutations in genes essential for embryonic development. However, the cloning of these mutated genes is difficult at present as it requires positional cloning methods. In *Drosophila*, chemical mutagenesis screens were complemented with P-element insertional mutagenesis which facilitated the cloning of many genes that had been identified by chemical lesions. To facilitate the cloning of vertebrate genes that are important during embryogenesis, we have developed an insertional mutagenesis strategy in zebrafish using a retroviral vector. Here, in a pilot screen of 217 proviral insertions, we obtained three insertional mutants with embryonic lethal phenotypes, and identified two of the disrupted genes. One of these, *no arches*, is essential for normal pharyngeal arch development, and is homologous to the recently characterized *Drosophila* zinc-finger gene *clipper*, which encodes a novel type of ribonuclease. As it is easy to generate tens to hundreds of thousands of proviral transgenes in zebrafish, it should now be possible to use this screening method to mutate and then rapidly clone a large number of genes affecting vertebrate developmental and cellular processes.

INTRODUCTION

In invertebrates, such as fruit flies and nematodes, forward genetic screens have been quite successful at disrupting and subsequently identifying novel biologically interesting genes (1-4). Recently, in an effort to do the same in a vertebrate model system several large-scale forward screens have been performed in zebrafish using the chemical mutagen ethylnitrosourea (ENU) (5, 6). These screens have generated thousands of mutations in processes ranging from early embryonic cell movements, such as epiboly and gastrulation, to those important for organogenesis and behavior. While these efforts represent a major advance towards an understanding of vertebrate development, they suffer from the fact that ENU induces point mutations, and that positional cloning methods, which are often laborious and slow, will need to be employed to identify the disrupted genes. Although the mutant phenotypes are informative, the molecular characterization of the perturbed gene products is essential to a complete understanding of their role in development.

As an alternative, or possibly an adjunct, to ENU mutagenesis, we have developed an insertional mutagenesis strategy in zebrafish. The primary advantage of insertional mutagenesis is that the mutagenic DNA is likely to lie in the disrupted gene thereby facilitating its identification. In other animal models the use of cleanly integrating DNA elements such as P-elements in fruit flies (7, 8), Tc1 in nematodes (9, 10), and retroviruses in mice (11, 12) has demonstrated that the identification of genes disrupted by insertional mutagenesis can proceed rapidly. The development of an insertional mutagenesis strategy in zebrafish is of particular value because the *ex utero* development, as well as the transparency of the zebrafish embryo make it the vertebrate model best suited for forward genetic screening.

To induce insertional mutations in zebrafish we infected blastula-stage embryos with a retroviral vector and generated proviral insertions which were transmitted to subsequent generations (13, 14). These insertions were then bred to homozygosity and the resulting embryos screened for mutant phenotypes. The results of this pilot screen, which identified three retrovirally-induced insertional mutations and two of the mutated genes, are described below. In addition, for one of these mutations, an initial characterization of the mutant phenotype and disrupted gene are presented.

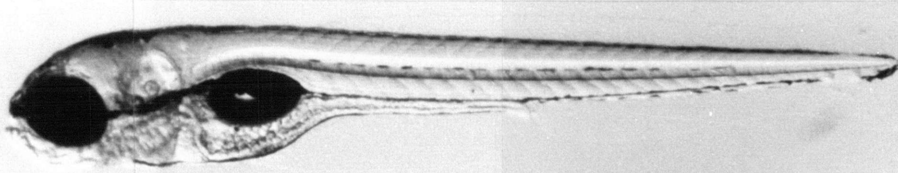
RESULTS AND DISCUSSION

Retroviral Insertional Mutagenesis in Zebrafish

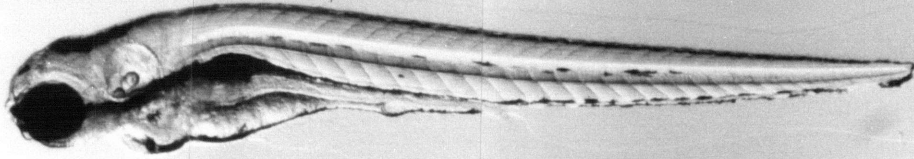
The infection of blastula-stage zebrafish embryos with MLV/VSV (Moloney murine leukemia virus/vesicular stomatitis virus) pseudotyped retroviral vectors (15, 16) results in the germ-line transmission of proviral insertions (13, 14). Integration of proviral DNA into the mouse genome has been shown to be capable of disrupting essential genes (11, 12, 17). In an effort to isolate insertional mutations in zebrafish we generated and inbred 217 proviral transgenes and screened for mutant phenotypes that were linked to those transgenes. The virus used to generate the insertions SFG(G) has been described elsewhere (14) (see Chapter 3). Three proviral transgenes, referred to as 38M, 67D, and 80A, were found to be linked to mutant phenotypes as all mutant embryos were transgenic (data not shown). These three phenotypes are lethal, including a variety of defects in the head, visceral organs, and fins, and in all three cases mutant embryos develop edema by day 5 of development (Fig 1).

Fragments of the genomic DNA flanking the three potentially mutagenic insertions were isolated, then single-copy sequences were identified and these

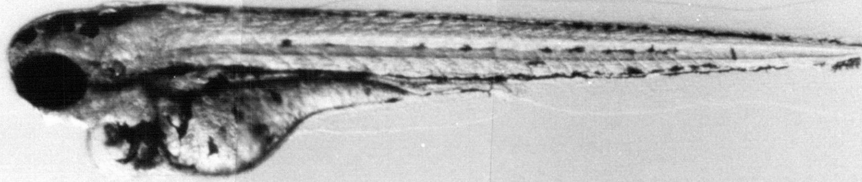
a



b



c



d

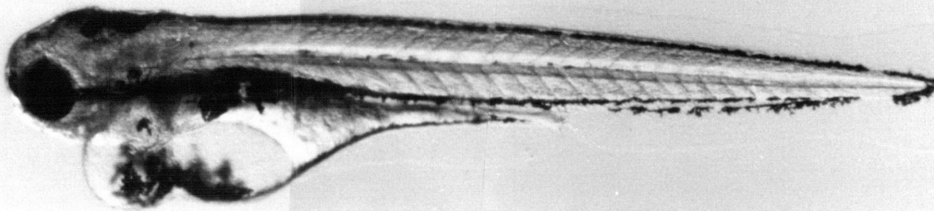


FIG. 1 Comparison of wild-type and mutant embryos on day 5 of development. (A) Wild-type embryo. (B, C, and D) Mutant embryos obtained by inbreeding insertions 38M, 67D, and 80A, respectively. Inbreeding crosses were screened as described previously (5).

sequences used as probes on Southern blots to determine the genotype of embryos generated by inbreeding the insertions (Fig. 2). For insertion 38M, we analyzed 110 mutants and 272 wild-type siblings; for insertion 67D, 133 mutants and 252 wild-type siblings; and for insertion 80A, 109 mutants and 241 wild-type siblings. In all three cases, every mutant embryo examined was found to be homozygous for the insertion in question, and every phenotypically wild-type sibling was found to be either heterozygous or non-transgenic. With the number of embryos examined in each case, if a mutation and the associated insertion were 1 cM apart, we would have observed 3-4 recombinants. No recombinants were observed, indicating that in all three cases the insertions and mutations are tightly linked.

These linkage data suggested that proviral insertions, 38M, 67D, and 80A, were likely to have induced the mutant phenotypes observed upon inbreeding. In an effort to identify the genes mutated by these insertions, 2-5 Kb fragments of genomic DNA flanking the three proviruses were sequenced. These sequences were used to search the Genbank sequence database using the BLAST algorithm (18). For insertions 38M and 67D, putative gene sequences with significant homology to genes in the database were identified. The only significant homology found in the sequence adjacent to insertion 80A was a recently characterized zebrafish short interspersed repeat called a DANA element (19).

DNA sequence adjacent to the site of insertion 38M was found to be highly homologous to a *Drosophila* zinc-finger gene, as well as to putative gene sequences identified by the *C. elegans* and yeast genome projects. The *Drosophila* gene, *clipper* (*clp*) was identified in an expression screen for proteins that bind single-stranded nucleic acids (20), and has recently been shown to encode a novel type of ribonuclease (21). Based on the homology between the fish and fly sequences, the intron and exon structure of the zebrafish gene was predicted (Fig. 3A). Reverse transcriptase (RT) PCR demonstrated that this zebrafish sequence is expressed in day

Figure 2

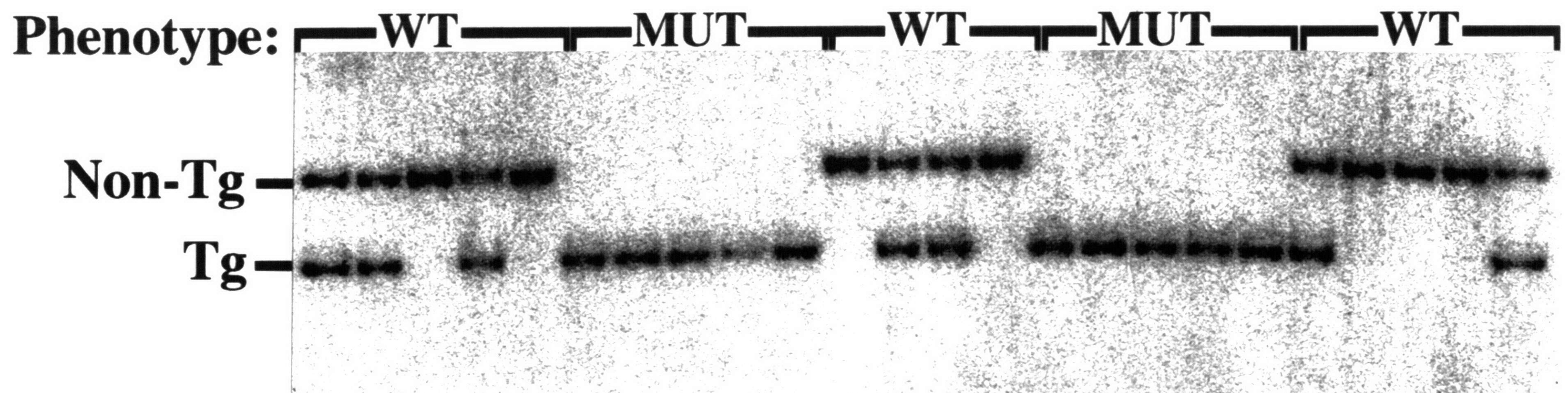


Figure 3

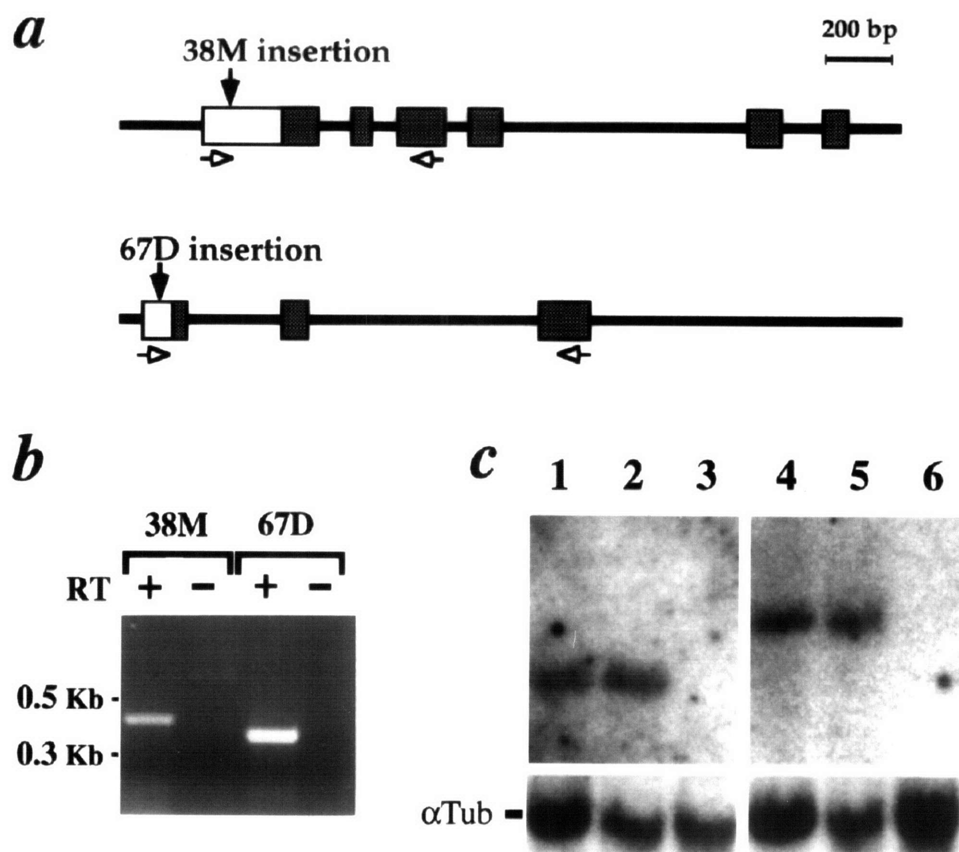


FIG. 2 Linkage analysis between insertion 38M and the mutant phenotype resulting from its inbreeding. Phenotypically wild-type and mutant embryos from a cross of two fish heterozygous for insertion 38M were sorted on day 5, and individual embryos were genotyped by Southern blot. Tg, transgenic

FIG. 3 Proviral insertions 38M and 67D have integrated into and disrupted expressed zebrafish genes. (A) The genomic site of insertion 38M contains amino acid homology to the *Drosophila* gene *clp* (shaded) (22). The genomic site of insertion 67D contains amino acid homology to a human expressed sequence tag (shaded). In both cases the exons depicted encode only the 5′ portion of the fish genes. (B) RT-PCR showing that insertions 38M and 67D have both landed in exon sequence. The primers used are shown in (a). The identity of these PCR products and the predicted splicing events were confirmed after subcloning and sequencing. In both cases, the amplified sequences extend upstream (white boxes) of the exon sequences identified by homology to other species (shaded). The absence of open reading frames which extend past the homologous regions suggests that the insertions have occurred into the 5'-UTR of these genes. (C) Northern blot analysis of embryos on day 5 showing the disruption of zebrafish genes by insertions 38M and 67D. See text for details. An α-tubulin probe was used to confirm the presence of RNA in all lanes.

5 embryos and that insertion 38M has integrated within exon sequence likely to be the 5'-UTR of the gene (Fig. 3B). A more extensive description of this gene and the corresponding mutant phenotype will be presented below.

DNA sequence at the site of insertion 67D were found to be highly homologous to a human expressed sequence tag (EST) of unknown function, as well as to putative gene sequence identified in the C. *elegans* and yeast genome projects. Intron and exon structure was predicted based on homology between the human and fish sequences (Fig. 3A). RT-PCR was used to demonstrate that the zebrafish gene is in fact expressed in day 5 embryos, and that insertion 67D has integrated into exon sequence likely to be the 5'-UTR of the gene (Fig. 3B). A more extensive description of this gene and the corresponding mutant phenotype will be presented in Chapter 5.

To determine whether the expression of the genes adjacent to insertions 38M and 67D was disrupted by the proviral insertions, northern blot analysis was used (Fig. 3C). RNA isolated from wild-type embryos on day 5 of development was found to contain a 1.9 Kb transcript when probed with a cDNA fragment from the gene at the site of insertion 38M, and a 2.3 Kb transcript when probed with a cDNA fragment from the gene at the site of insertion 67D (Fig 3C, lanes 1 and 4, respectively). Insertions 38M and 67D were inbred and RNA was made on day 5 of development from sorted pools of phenotypically wild-type and mutant embryos. For both insertions, the phenotypically wild-type embryos expressed transcripts of the expected size (Fig. 3C, lanes 2 and 5), whereas no transcripts could be detected in the mutant embryos (lanes 3 and 6).

These results demonstrate that the proviral insertions 38M and 67D have either disrupted the expression of the genes into which they have integrated, or in homozygous embryos have resulted in a reduction or deletion of the tissues in which those genes are normally expressed. Considering also the tight linkage between these two insertions and the associated mutations, we believe that the

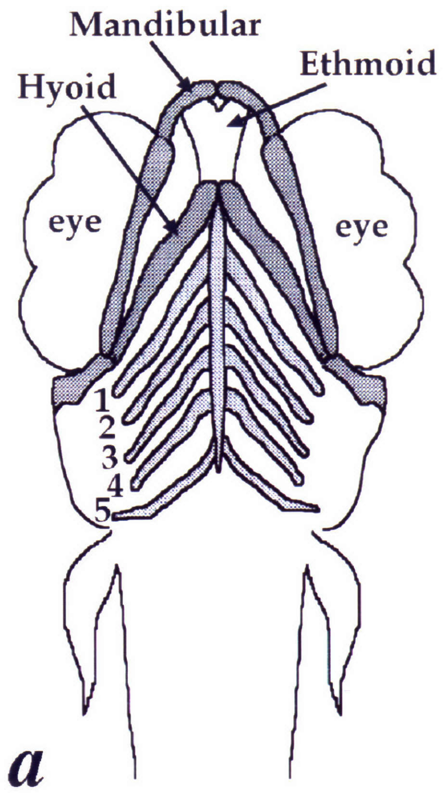
insertional disruption of these genes is responsible for the observed phenotypes, although definitive proof will require rescue of the mutant phenotypes by addition of functional copies of the disrupted genes.

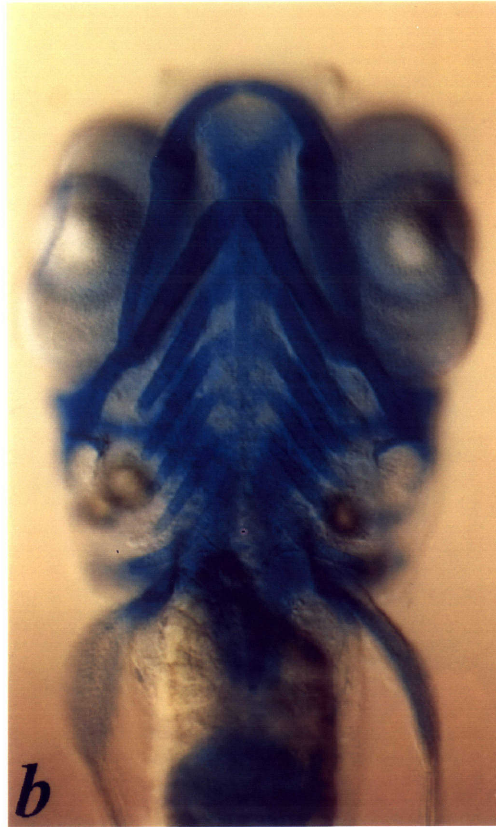
The *no arches* mutant phenotype

The gene mutated by insertion 38M is of interest because its disruption results in defects in the head, and in particular the pharyngeal arches, a set of structures unique to vertebrates and derived in large part from the cells of the neural crest (22). We have called this gene *no arches* (*nar*), and have initiated studies to address its role in vertebrate development. Initial observations of *nar* mutant embryos reveal that while they appear normal during the first 2 days of development, by day 3 they have smaller heads and eyes than their wild-type siblings, and the growth of their pectoral fins is stunted. On days 4 and 5 the difference in head size between mutant and wild-type embryos becomes more extreme, and the pharyngeal arches are largely absent in the mutant. No obvious defects are present in the trunk and the tail of mutant embryos, although the development of the liver and gut is slightly retarded.

To examine the structure of the head skeleton, day 5 mutant and wild type embryos were treated with alcian blue which stains cartilaginous structures (23). Staining of wild-type embryos shows the seven pharyngeal arches which include two jaw arches, the mandibular and the hyoid, and five branchial arches (Fig. 4A, B). Also evident is the cartilage of the ethmoid plate, the most anterior portion of the neurocranium. In mutant embryos the cartilage of the branchial arches is largely or entirely absent, and that of the jaw arches is absent or malformed (Fig. 4C). The cartilage of the neurocranium is present but reduced in size and the ethmoid plate extends far less anteriorly than in wild-type embryos. Interestingly, the cartilage of both the pharyngeal arches and of the ethmoid plate is thought to be neural crest derived (24, 25).

Figure 4





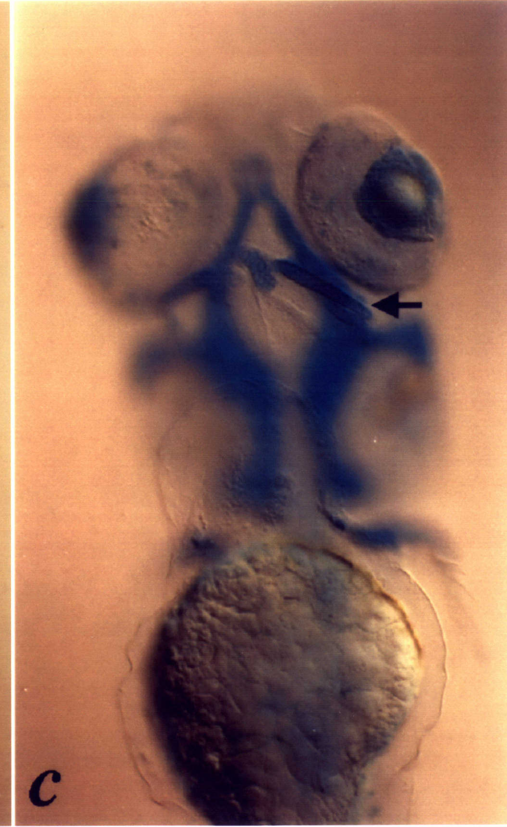


FIG. 4 Alcian blue staining of wild-type and *nar* mutant embryos on day 5 of development. (A) Schematic representation of the ventral view of a wild-type embryo. (B, C) Ventral view of stained wild-type and mutant embryos, respectively. The absence of the cartilage of the branchial arches in the mutant is apparent and malformed pieces of the jaw arches are present (arrow).

The reduction in pharyngeal arch structures in *nar* mutant embryos suggests that either the progenitor cells don't proliferate sufficiently in this region, or that cell death reduces cell numbers, or both. Evidence from a comparison of sections of *nar* mutants and wild-type siblings on day 3 of development suggested that there is significantly more cell death present in the head region of mutants than that of wild-type embryos (data not shown). A similar comparison between sections of day 5 embryos showed no increased cell death in the mutant as compared to the wild-type suggesting that this phenomenon occurs during a limited time window.

To address the role of cell death in *nar* mutants, embryos obtained by inbreeding 38M heterozygous fish were live stained with acridine orange. Acridine orange has been shown to stain dying cells in *Drosophila* (26), and more recently in zebrafish (27). Embryos were stained on days 1 through 4. On days 1 and 2, when *nar* mutants cannot be morphologically distinguished from wild-type siblings, no obvious differences in staining was detected among the pool of injected embryos (not shown). However, significant differences were apparent by day 3 when mutant embryos showed large amounts of cells death in the pharyngeal arch region, the head, and the pectoral fins, while wild-type siblings showed little cell death in these areas (Fig. 5). By day 4, cell death in mutant embryos decreased to levels comparable to those found in wild-type siblings (not shown).

These results demonstrate that cell death plays a role in the acquisition of the *nar* mutant phenotype. The fact that acridine orange is thought to stain apoptotic cells in particular (26, 27) and that the cell death observed in *nar* mutants occurs only during a limited developmental time period suggests that the cell death seen in these mutants is likely to be apoptotic and not the result of cellular degeneration caused by hypoxia or external factors (i.e. bacterial infection).

Figure 5

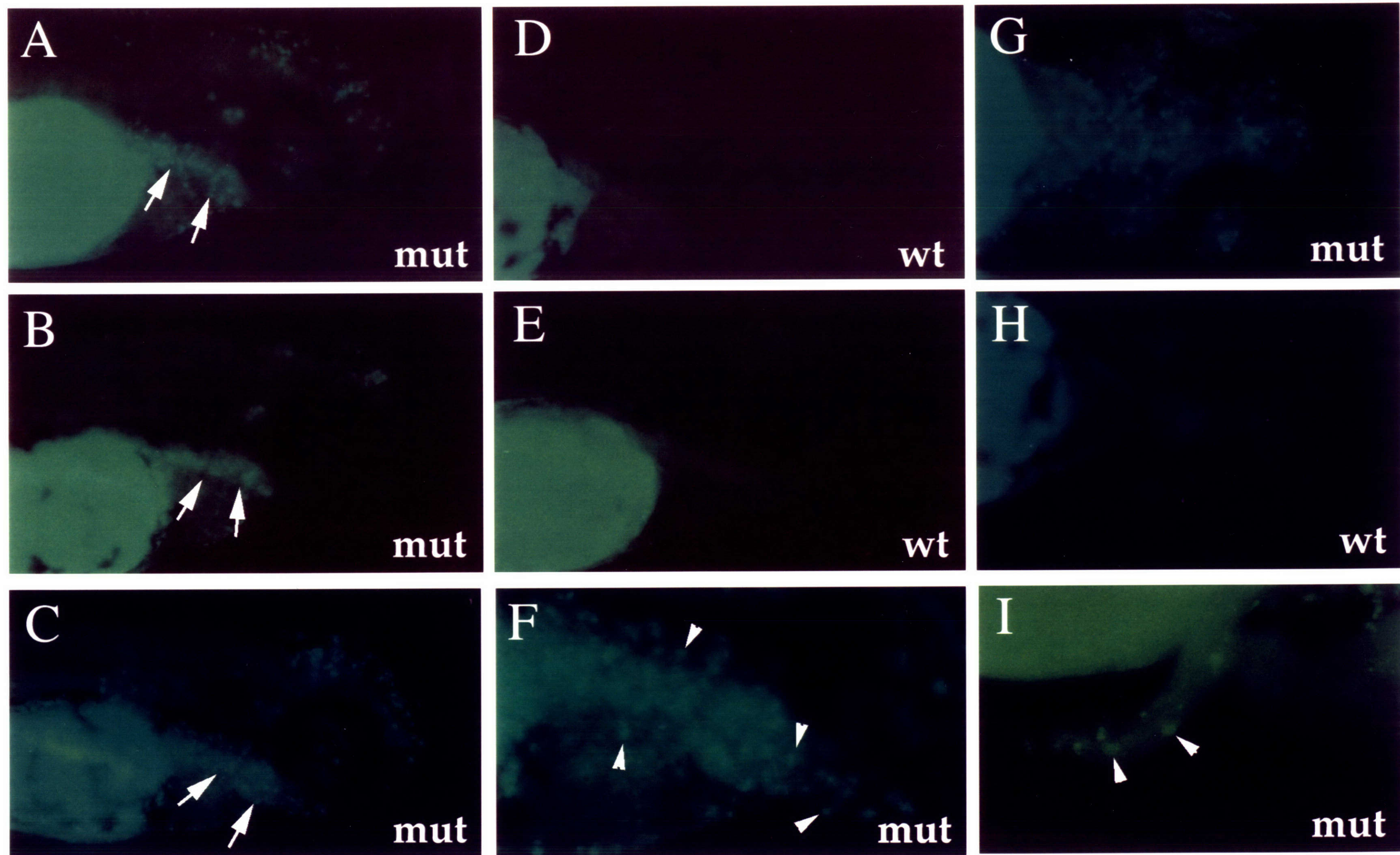


FIG. 5 Acridine orange staining of wild-type and *nar* mutant embryos on day 3 of development. (A-E) Lateral view of *nar* mutants (a-c), and wild type siblings (d, e), showing that mutants have many stained cells in the developing pharyngeal arches (arrows) and in the head, while wild-type embyros have few to none. (F) Close-up of the pharyngeal arch region of the mutant embryo in (c). Individual stained cells are apparent (arrowheads). (G, H) Ventral views of a mutant (g) and wild-type sibling (h) show the dramatic increase in cell death in the mutant as compared to wild-type. (I) Ventral view of the pectoral fin of a mutant embryo shows dying cells (arrowheads). Anterior is to the right in all cases. Yolk auto-fluorescence can be seen in both mutant and wild-type embryos.

The *nar* gene product

Three *nar* cDNA clones were isolated from an adult zebrafish cDNA library (obtained from D. Grunwald, Univ. of Utah). These clones were sequenced and were found to include two differentially-polyadenylated forms of the *nar* message suggesting that there are at least two different transcripts present in the adult. One of these is similar in size and likely to be the 1.9 Kb message detected by northern blot in day 5 embryos (Fig. 3C, lanes 1 and 2), and the other transcript is 712 bp shorter, lacking nearly all of the 3'-UTR present in the 1.9 Kb message. Each of these messages is polyadenylated at a distinct AATAAA sequence present downstream of the protein coding sequence (Fig. 6).

Based upon the *nar* cDNA sequence a full-length *nar* coding region was predicted (Fig. 6) and used to search the translated public EST database, and a human EST was identified with homology to the C-terminus of the fish protein. This EST clone was truncated at the 5' end and did not include homology to the N-terminal third of the Nar protein. Northern analysis using a blot of RNA from multiple human tissues, and the human *nar* (*hnar*) EST as a probe, suggested that, in humans, this gene may be ubiquitously expressed (not shown). Subsequently, 5'-RACE was used to isolate the 5' end of the *hnar* cDNA using RNA from 293 cells (kidney derived) as the template, and the full-length protein coding region was determined.

The sequence of the zebrafish and human Nar proteins, as well as the that of homologous Drosophila protein Clipper (Clp), were aligned, and the degree of evolutionary conservation assessed (Fig. 7). Clp contains seven zinc-finger domains, and has been shown biochemically to be a novel type of endoribonuclease which specifically cleaves RNA hairpins (21). Six of the seven zinc-fingers present in the Clp protein are conserved in both the fish and human Nar proteins including the five at the N-terminus which possess the nuclease activity in Clp (Fig. 7). The conserved C-terminal zinc finger is similar in structure to those of retroviral

Figure 6: *nar* cDNA sequence

```
CTCGAGCAGCGCTGTGGAAGTGGACGTTTCTGAAACCAACTGTTTATTTACAATGTTTTTCTGCTTTACCGGTTCACCCTTGCGCTTTAT    90

TAAAAAACGTTTACATAGATAATATATCTGCGTGCTTTTTCAGCCATTCTCTCTTTTTTACGTCTGATATTAAACCACAACAAAGCCAAA   180

CTGATAACTTGTTAGCCATTAAGACTTGCTTTACTTTTCTGATCAAGTGTTGTAACGAGAAGAGGTTCAGCTGCTGGTTAAAATATTAAT   270

AATATACTCAATATAATTCCAAAATGCAGGAATTAATCGCCACTGTCGATCACATCAAGTTTGATCTTGAAATCGCTGTTGAACAACAAT   360
                            M  Q  E  L  I  A  T  V  D  H  I  K  F  D  L  E  I  A  V  E  Q  Q  L

TAGGCGCTCAGCCGTTGCCATTTCCAGGAATGGATAAATCTGGCGCTGCTGTATGCGAGTACTTCATGAGAGCTGCTTGTATGAAAGGTG   450
 G  A  Q  P  L  P  F  P  G  M  D  K  S  G  A  A  V  C  E  Y  F  M  R  A  A  C  M  K  G  G

GAATGTGTCCCTTCAGACACATCAGTGGAGAAAAGACTGTGGTGTGCAAGCACTGGCTTAGAGGCTTGTGCAAGAAAGGAGATCAATGTG   540
 M  C  P  F  R  H  I  S  G  E  K  T  V  V  C  K  H  W  L  R  G  L  C  K  K  G  D  Q  C  E

AATTTTTACACGAGTACGACATGACTAAGATGCCCGAGTGTTATTTCTACTCTAAATTTGGGGAGTGCAGTAATAAAGAGTGTCCATTCT   630
 F  L  H  E  Y  D  M  T  K  M  P  E  C  Y  F  Y  S  K  F  G  E  C  S  N  K  E  C  P  F  L

TGCACATTGACCCAGAATCTAAGATTAAAGATTGCCCGTGGTATGATAGGGGTTTTTGTAAACATGGTCCGGATTGCAGGCACAGACACA   720
 H  I  D  P  E  S  K  I  K  D  C  P  W  Y  D  R  G  F  C  K  H  G  P  D  C  R  H  R  H  T

CTAGAAGAGTCATTTGTGTAAATTACCTGGTTGGCTTTTGTCCAGAGGGCAAGTCATGCAAATTTATGCACCCAAGATTTGAACTTCCGA   810
 R  R  V  I  C  V  N  Y  L  V  G  F  C  P  E  G  K  S  C  K  F  M  H  P  R  F  E  L  P  M

TGGGTGCGACTGAACAGCCACCACTGCCACAACAAGTACAGACTCAGCAAAAGCAGCAAAACATGCAGCCCATAAACAGGTCATCACAGT   900
 G  A  T  E  Q  P  P  L  P  Q  Q  V  Q  T  Q  Q  K  Q  Q  N  M  Q  P  I  N  R  S  S  Q  S

CACTCATCCAGTTAACCAACCCCAACATCAGCAACAACAACCACCAGAGAATTCCAAACGCTGTTGGGATAGTGCACTCTAACAGCCACA   990
 L  I  Q  L  T  N  P  N  I  S  N  N  N  H  Q  R  I  P  N  A  V  G  I  V  H  S  N  S  H  M

TGGGTGGAGCTCGTGGCCCTCGCCCGCTGGACCAAGTTACATGTTACAAGTGTGGTGAAAAGGGCCATTACGCAAACAAATGCACAAAAG  1080
 G  G  A  R  G  P  R  P  L  D  Q  V  T  C  Y  K  C  G  E  K  G  H  Y  A  N  K  C  T  K  G

GACACCTGGCGTTTTTGAGTGGACAGTGAAGCTTTCTGTAAAAAGGATCTCCAATAAATGTGGTCATCACGTGAGGTATTCGATTCCTGG  1170
 H  L  A  F  L  S  G  Q  *

TACATTACCCTCTCACTGAGAAGGGTGTAGTTGAGTGACATAAATGTGGCTCTACCTCACCACCATCACGTATGCCAAGCGCATCACAGT  1260

CACATAACACTTTCAACAGGTAGAACACATCAGTAAAAGATTTTTATATTGAGTTTTGTTTTTCTTTCACATGAAACCTTACATACCCGA  1350

TTTCATGCTTTGAAACGTGAATCTCTGCTATTTTATTTAGAGCTGGAAGATTTTTTGAATGCTTAAAGGGATAGTTCACCCAAAAATGAC  1440

AATAACCTGCATGTTTTTTTGTTTATTTCAAATGTTGAACACAAAAGGAGATATTTTAGAGAATTTTGGTTGCTGTCACCCATTTCCTTC  1530

CATAGTAGGGAAAAAATTACTATGGAAGAGGATGTCAACAACTAACATTTCTCAGAATATCTTATTTTGTGTTTAACAGAAGAAAGAAAT  1620

TGAAATAGATTTTGAACAAGGGAAAGTAAATGATGGCAGAATGTTCTTCTTTGGTGAACTATCCCTTTACATTTTAAAGTGATAAAGCAA  1710

AATATATGAAAGTCTTTTGGGGAACTTTTGGCAATAGCAGCCAGTGTGATCTGAAATTGAAGAGCACCGTTTTGCAGCATCAAAATGGCT  1800

TGTTTATATAAATTACATTTAAAACAGATGTTTGGGTTGTACTTACAATAAAGATTATCTGGATTTAAAAAAAAAAAAA             1879
```

Figure 7: Alignment of zebrafish Nar, human Nar, and Drosophila Clp proteins.

```
1   MQELIATVDHIKFDLEIAVEQQLGAQPLPPPGMDKSGAAV Fish
1   MQEIIASVDHIKFDLEIAVEQQLGAQPLPPPGMDKSGAAV Human
1   MDILLANVSGLQFKAERDLIEQVGAIPLPFYGMDKSIAAV Fly

41  CEYFMRAA--CMKGGMCPPRHISGEKTVVCKHWLRGLCKK Fish
41  CEFFLKAA--CGKGGMCPFRHISGEKTVVCKHWLRGLCKK Human
41  CNFITRNGQECDKGSACPFRHIRGDRTIVCKHWLRGLCKK Fly

79  GDQCEFLHEYDMTKMPECYFYSKFGECSNKECPFLHIDPE Fish
79  GDQCEFLHEYDMTKMPECYFYSKFGECSNKECPFLHIDPE Human
81  GDQCEFLHEYDMTKMPECYFYSRFNACHNKECPFLHIDPQ Fly

119 SKIKDCPWYDRGFCKHGPDCRHRHTRRVICVNYLVGFCPE Fish
119 SKIKDCPWYDRGFCKHGPLCRHRHTRRVICVNYLVGFCPE Human
121 SKVKDCPWYKRGPCRHGPHCRHQHLRRVLCMDYLAGPCPE Fly

159 GKSCKFMHPRFELP-MGATEQPPLPQQVQT-------QQK Fish
159 GPSCKFMHPRFELP-MGTTEQPPLPQQTQP-------PAK Human
161 APSCKHMHPHFELPPLAELGKDQLHKKLPTCHYCGELGHK Fly

191 QQN----MQPINRSSQSLIQLTNPNISNNN----HQRIPN Fish
191 QSN----NPPLQRSS-SLIQLTSQN-SSPN----QQRTPQ Human
201 ANSCKQYVGSLEHRN-NINAMDHSG-GHSGGYSGHSGHIE Fly

223 AVGIVHSNSHMGGA-----RGPRPLDQVTCYKCGEKGHYA Fish
221 VIGVMQSQNSSAGN-----RGPRPLEQVTCYKCGEKGHYA Human
239 GADDMQSNHHSQPHGPGFVKVPTPLEEITCYKCGNKGHYA Fly

258 NKCTKGHLAFLSGQ                           Fish
256 NRCTKGHLAFLSGQ                           Human
279 NKCPKGHLAFLSNQHSHK                       Fly
```

FIG. 6 *nar* cDNA sequence. The putative coding region encodes a protein product of 271 amino acids. There are two in-frame stop codons 24 bp upstream of the predicted initiation methionine. Two different polyadenylation signals present in the 3′-UTR are indicated (underlined), as is the insertion site of provirus 38M in the 5′-UTR (ATAT, raised). This sequence is not likely to contain the complete 5′ end of transcript based on the fact that it ends with an *Xho I* site which was an enzyme used to construct the cDNA library.

FIG. 7 Alignment of the putative zebrafish and human Nar proteins with the *Drosophila* homologue Clp. The amino acid identity (shaded) and the zinc-finger motifs (boxes) are shown. The fish and human proteins are 86% identical, while the fish and fly proteins are roughly 50% identical. This alignment was obtained using the Lasergene software (DNAStar, Inc.).

nucleocapsid proteins which bind to the single-stranded RNA genome of retroviruses (28). The presence of an RNA binding motif at the C-terminus of Nar, in conjunction with a putative ribonuclease at the N-terminus, suggest that the Nar protein may function during zebrafish embryogenesis by binding to and cleaving RNA.

To examine the temporal pattern of expression of *nar* during zebrafish development, as well as expression in adult fish, northern blot analysis was performed. Two different *nar* transcripts of roughly 1.9 Kb and 1.2 Kb were detected in unfertilized eggs indicating that these messages are maternally supplied (Fig. 8A). Consistent with this was the detection of both messages in adult females, and in particular the ovaries (Fig. 8B). The size of these two messages suggests that they correspond to the two cloned cDNA sequences. Very low levels of the 1.9 Kb transcript are present in adult males and in adult females with their ovaries removed.

During embryogenesis only the 1.9 Kb transcript was detected and was found to be present during gastrulation, neurulation, and somitogenesis (6-24 hours) (Fig. 8A). The transcript level increases significantly during this period and remains high through 48 hours but then decreases by 96 hours. The timing of *nar* expression during embryogenesis is consistent with a role in the formation of the pharyngeal skeleton, which normally takes place during days 2 and 3 of development (25).

Figure 8

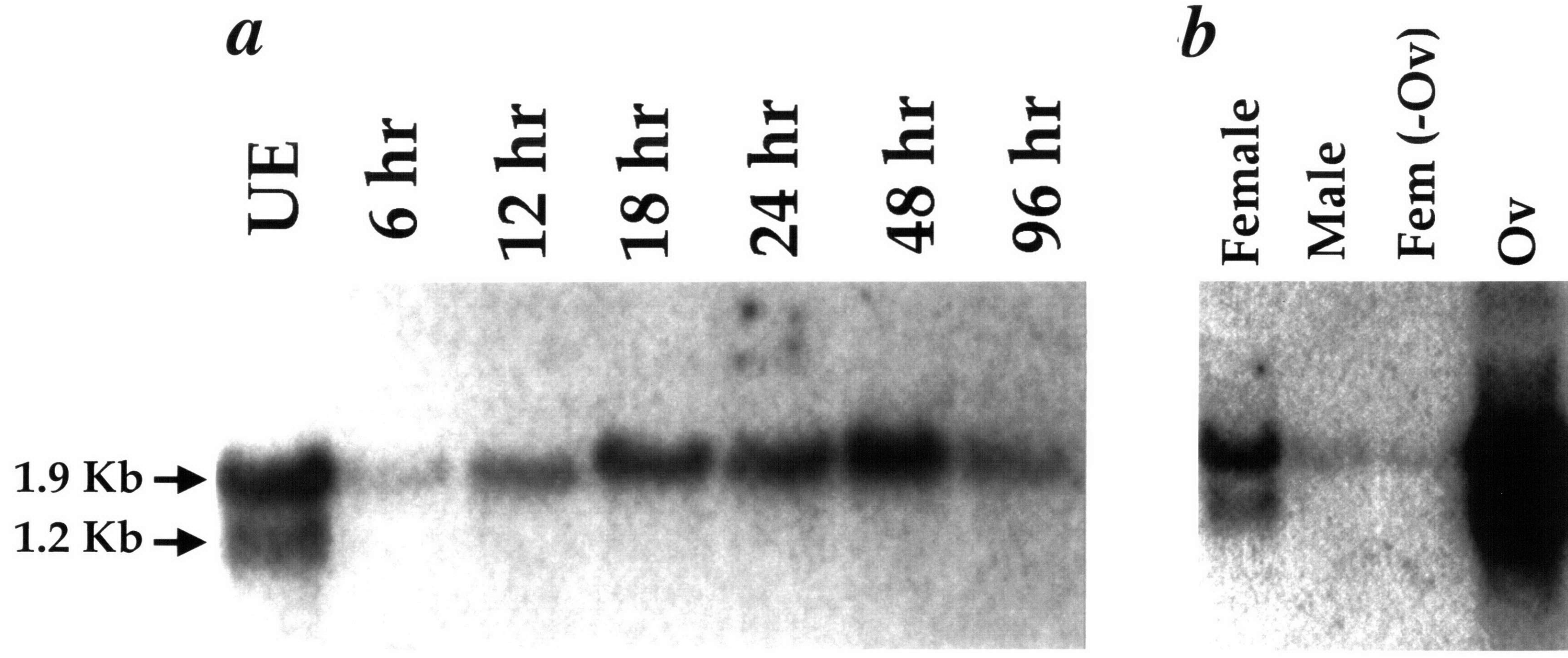


FIG. 8 Northern blot analysis of *nar* expression during embryogenesis and in adult fish. Comparable amounts of total RNA are present as judged by the amounts of 18S rRNA present (not shown).

SUMMARY

Our pilot insertional mutagenesis screen in zebrafish using a retroviral vector has identified three insertional mutants. In two cases we have found that the proviruses have inserted into and disrupted highly-conserved genes essential for zebrafish embryogenesis. Preliminary analysis of one of these genes indicates that it encodes a zinc-finger endoribonuclease which has temporally regulated expression during development and is essential for normal pharyngeal arch development.

Using this methodology on a larger scale, it should now be possible to isolate hundreds of insertional mutants in zebrafish and to easily clone the mutated genes. The use of alternative strategies, such as the screening of haploids or gynogenotes (29), or the development of a retroviral gene trap (30, 31) for use in zebrafish, could make the isolation of thousands of insertional mutants possible (see Chapter 6 for further discussion).

MATERIAL AND METHODS

Southern and Northern blots: Genomic DNA was isolated after lysis at 55°C in 100 mM Tris (8.0), 5 mM EDTA, 200 mM NaCl, 0.4% SDS, and 100 μg/ml Proteinase K. Total RNA was prepared using the RNAzol B method (Tel-Test, Friendswood, TX). All gels were transferred to Hybond N+ membranes (Amersham), and filters were hybridized and washed as previously described (14).

Reverse transcriptase PCR and 5'-RACE: For RT-PCR, first strand cDNA synthesis was performed at 42°C for 1 hr using 1 μg of total RNA. PCR was performed in 67 mM Tris (8.8), 16.6 mM $(NH_4)_2SO_4$, 10 mM βME, 170 μg/ml BSA, 2 mM $MgCl_2$, with 0.2 μM of each dNTP, and 0.4 μM of each primer. The PCR program used was as

follows: 30 cycles of 15 seconds at 94°C, 30 seconds at 55°C, and 45 seconds at 72°C with an initial 2 minute denaturation step and a final 5 minute elongation step. 5'-RACE (rapid amplification of cDNA ends) (32) was performed using a kit from Gibco BRL/Life Technologies (Version 2.0) according to the provided instruction manual.

Alcian blue staining: Embryos were fixed in 4% paraformaldehyde for 2 hours at room temperature, and then stained overnight with 0.1% alcian blue in 80% ethanol/20% glacial acetic acid. Embryos were rehydrated into PBS, treated for 1-2 hours in a 1% trypsin solution saturated with sodium borate, and were bleached in 1% KOH (w/v), 3% H_2O_2, and 100 mM NaCl. Stained embryos were then washed with PBS and cleared with glycerol.

Acridine orange staining: Acridine orange (1 mg/ml in PBS) was injected into the yolk of embryos from crosses inbreeding insertion 38M. Injected embryos were incubated for 2-3 hours at 28°C to allow sufficient diffusion throughout the embryos which was facilitated by the embryonic circulatory system. Individual embryos were then examined using fluorescence microscopy and the presence of stained cells noted.

REFERENCES

1. Nüsslein-Volhard C; Wieschaus E (1980) Mutations affecting segments number and polarity in *Drosophila*. *Nature* **287**: 795-801.

2. Schüpbach T (1987) Germ line and soma cooperate during oogenesis to establish the dorsoventral pattern of egg shell and embryo in *Drosophila melanogaster*. *Cell* **49**: 699-707.

3. Seydoux G; Savage C; Greenwald I (1993) Isolation and characterization of mutations causing abnormal eversion of the vulva in *Caenorhabditis elegans*. *Dev Biol* **157**: 423-36.

4. Hekimi S; Boutis P; Lakowski B (1995) Viable maternal-effect mutations that affect the development of the nematode *Caenorhabditis elegans*. *Genetics* **141**: 1351-64.

5. Haffter P; Granato M; Brand M; Mullins M; Hammerschmidt M; Kane D; Odenthal J; van Eeden F; Jiang Y; Heisenberg C-P; Kelsh R; Furutani-Seiki M; Vogelsang L; Beuchle D; Schach U; Fabian C; Nüsslein-Volhard C (1996) The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* **123**: 1-36.

6. Driever W; Solnica-Krezel L; Schier A; Neuhauss S; Maliki J; Stemple D; Stainier D; Zwartkruis F; Abdelilah S; Rangini Z; Belak J; Boggs C (1996) A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123**: 37-46.

7. Cooley L; Kelley R; Spradling A (1988) Insertional mutagenesis of the *Drosophila* genome with single P elements *Science* **239**: 1121-8.

8. Bier E; Vaessin H; Shepherd S; Lee K; McCall K; Barbel S; Ackerman L; Carretto R; Uemura T; Grell E; Jan LY; Jan YN (1989) Searching for pattern and mutation in the *Drosophila* genome with a P-lacZ vector. *Genes Dev* **3**: 1273-87.

9. Tabuse Y; Nishiwaki K; Miwa J (1989) Mutations in a protein kinase C homolog confer phorbol ester resistance on *Caenorhabditis elegans*. *Science* **243**: 1713-6.

10. Rogalski TM; Williams BD; Mullen GP; Moerman DG (1993) Products of the unc-52 gene in Caenorhabditis elegans are homologous to the core protein of the mammalian basement membrane heparan sulfate proteoglycan. *Genes Dev* **7**: 1471-84.

11. Schnieke A; Harbers K; Jaenisch R (1983) Embryonic lethal mutation in mice induced by retroviral insertion into the a1(I) collagen gene. *Nature* **304**: 315-20.

12. Gridley T; Gray DA; Orr-Weaver T; Soriano P; Barton DE; Francke U; Jaenisch R (1990) Molecular analysis of the Mov 34 mutation: transcript disrupted by proviral integration in mice is conserved in *Drosophila*. *Development* **109**: 235-42.

13. Lin S; Gaiano N; Culp P; Burns JC; Friedmann T; Yee JK; Hopkins N (1994) Integration and germ-line transmission of a pseudotyped retroviral vector in zebrafish. *Science* **265**: 666-9.

14. Gaiano N; Allende M; Amsterdam A; Kawakami K; Hopkins N (1996) Highly efficient germ-line transmission of proviral insertions in zebrafish. *Proc Natl Acad Sci USA* **93**: 7777-82.

15. Burns JC; Friedmann T; Driever W; Burrascano M; Yee JK (1993) Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to very high titer and efficient gene transfer into mammalian and nonmammalian cells. *Proc Natl Acad Sci USA* **90**: 8033-7.

16. Hopkins N (1993) High titers of retrovirus (vesicular stomatitis virus) pseudotypes, at last. *Proc Natl Acad Sci USA* **90**: 8759-60.

17. Meisler MH (1992) Insertional mutation of 'classical' and novel genes in transgenic mice. *Trends Genet* **8**: 341-4.

18. Altschul S; Gish W; Miller W; Myers E; Lipman D (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-10.

19. Izsvak Z; Ivics Z; Garcia-Estefania D; Fahrenkrug SC; Hackett PB (1996) DANA elements: A family of composite, tRNA-derived short interspersed DNA elements associated with mutational activities in zebrafish (*Danio rerio*). *Proc Natl Acad Sci USA* **93**: 1077-81.

20. Stroumbakis ND; Li Zhihua; Tolias PP (1994) RNA- and single-stranded DNA-binding (SSB) proteins expressed during *Drosophila melanogaster* oogenesis: a homolog of bacterial and eukaryotic mitochondrial SSBs. *Gene* **143**: 171-7.

21. Bai C; Tolias P (1996) Cleavage of RNA hairpins mediated by a developmentally regulated CCCH zinc-finger protein. *Mol Cell Biol* (in the press).

22. Gilbert SF (1994) Early vertebrate development: neurulation and ectoderm. In *Development Biology* (Sinauer Associates, Inc), pp 244-94.

23. Dingerkus G; Uhler DL (1977) Enzyme clearing of alcian blue stained whole small vertebrates for demonstration of cartilage. *Stain Tech* **32**: 229-31.

24. Langille RM; Hall BK (1988) Role of neural crest in development of the trabeculae and branchial arches in embryonic sea lamprey, *Pertomyzon marinus* (L) *Development* **102**: 301-10.

25. Schilling TF; Kimmel CB (1994) Segment and cell type lineage restrictions during pharyngeal arch development in the zebrafish embryo. *Development* **120**: 483-94.

26. Abrams JM; White K; Fessler LI; Steller H (1993) Programmed cell death during *Drosophila* embryogenesis. *Development* **117**: 29-43.

27. Furutani-Seiki M; Jiang Y-J; Brand M; Heisenberg C-P; Houart C; Beuchle D; van Eeden FJM; Granato M; Haffter P; Hammerschmidt M; Kane D; Kelsh R;

Mullins M; Odenthal J; Nüsslein-Volhard C (1996) Neural degeneration mutants in the zebrafish, *Danio rerio*. *Development* **123**: 229-39.

28. Summers MF (1991) Zinc finger motif for single-stranded nucleic acids? Investigations by nuclear magnetic resonance. *J Cell Biochem* **45**: 41-8.

29. Westerfield M (1995) ed*The Zebrafish Book* (Univ. of Oregon Press, Eugene).

30. Friedrich G; Soriano P (1995) Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev* **5**: 1513-23.

31. Hicks GG; Shi EG; Chen J; Roshon M; Williamson D; Scherer C; Ruley HE (1995) Retrovirus gene traps. *Methods Enzymol* **254**: 263-75.

32. Frohman MA; Dush MK; Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci USA* **85**: 8998-9002.

CHAPTER 5

INSERTIONAL MUTAGENESIS IN ZEBRAFISH IDENTIFIES A NOVEL GENE, *PESCADILLO*, WHICH IS ESSENTIAL FOR EMBRYONIC DEVELOPMENT

ABSTRACT

Recently our laboratory described an efficient method for generating retroviral provirus insertions in the zebrafish germ line and we showed that provirus insertions induce embryonic mutations at a frequency of roughly 1 mutant per 70 insertions. To date we have isolated four insertional mutants, and using the proviruses as a molecular tag, have cloned the genes disrupted in three of them. The proviruses in all three mutants lie within or just 5' of the first coding exon, point in the opposite transcriptional orientation from the gene, and disrupt transcription. Here we present a molecular characterization of a gene identified by this method and describe the associated mutant phenotype. The *pescadillo (pes)* gene is predicted to encode a protein of 582 amino acids with no recognizable functional motifs, which is highly conserved from yeast to humans. *pes* mRNA is expressed widely and dynamically during the first three days of embryogenesis. Prominent sites of expression are the eyes and optic tectum on day 1, the fin buds, liver primordium and gut on day 2, and the branchial arches on day 3. Beginning at day 3 of embryogenesis, *pes* mutant embryos exhibit small eyes, a reduced brain and visceral skeleton, shortened fins, and a lack of expansion of the liver and gut, and then die on the sixth day of development. These results demonstrate the power of insertional mutagenesis in zebrafish for rapidly finding and characterizing novel genes essential for embryonic development. Depending upon the extent to which all genes are accessible to proviral insertion, the wider application of this approach could lead to the rapid identification of the majority of genes that are required for embryonic development of this vertebrate.

INTRODUCTION

The zebrafish is a superb model organism for identifying genes essential in vertebrate development (1-4). The ability to breed and maintain large numbers of adult animals in the laboratory makes classical genetics feasible, while the optical transparency of the zebrafish embryo makes it ideal for visualizing early developmental processes. This year two labs completed systematic large-scale mutant screens for embryonic lethal and visible mutations in the zebrafish (5, 6). Using ENU as the mutagen, the labs of Nüsslein-Volhard and Driever recovered several thousand chemically-induced point mutations which affect diverse aspects of early development in the zebrafish. About 70% of the mutants were considered to be nonspecific, while about 30% are associated with more specific, usually lethal, defects in patterning and morphogenesis. The latter include mutations affecting gastrulation, pattern formation, organogenesis, structural organization of the CNS, and basic behaviors. Altogether 400-500 genes with relatively specific developmental defects have been identified by complementation tests based on phenotypes encountered in these screens. It is estimated, although very roughly, that there are about 2,500 genes with essential or visible functions in the fish embryo and that about half were identified in the chemical mutagenesis screens (6).

Despite the wealth of new genetic information emerging from chemical mutagenesis screens in the zebrafish, knowledge about the molecular nature of the affected genes and their products will not be immediately forthcoming. Cloning the mutated genes will depend upon the development of reagents for positional cloning in the zebrafish. For now this technology remains laborious and expensive due to the large size of the zebrafish genome (1.6×10^{9}bp).

As an alternative to chemical mutagenesis, we recently developed a method for generating insertional mutants in zebrafish utilizing integration of retroviral

proviruses into the genome (7-9, see also Chapters 2-4). Although the frequency of mutagenesis is considerably lower than that of chemical mutagenesis, the molecular tag provided by the inserted retroviral provirus allows the immediate isolation of flanking genomic fragments, which are likely to include the disrupted gene. Of the four zebrafish insertional mutants isolated thus far, we quickly cloned genes disrupted in three of them. We believe these disrupted genes are likely to be responsible for the phenotypes of the respective mutants. Cloning was extremely rapid because in all three cases the provirus that is genetically linked to the mutant phenotype integrated close to coding sequences of the gene it disrupted, because so many gene sequences are now available in the database, and because the disrupted genes are highly conserved (9).

Here we describe the molecular characterization of one of the genes mutated by proviral insertion and present a preliminary characterization of the mutant phenotype. The gene *pescadillo* (*pes*) encodes a novel protein of unknown function that is very highly conserved across species: homologous sequences are present in human, mouse, *C. elegans* and yeast. These results presented here provide strong support for the prediction that genetic analysis in zebrafish will identify novel genes essential for vertebrate development. If the retroviral mutagen we have used integrates at random into the fish genome, given the efficiency of mutagenesis we have observed to date, it should be possible for a number of fish labs together to identify and clone the majority of the genes essential for the early development of this vertebrate species within several years.

RESULTS

In an ongoing insertional mutagenesis screen, visible or lethal phenotypes were sought among the progeny of crosses between pairs of F2 fish heterozygous for

a single identical proviral insertion (9). We screened by observing at least 25 live embryos under a dissecting microscope on days one, two, and five post-fertilization, and scored for consistent abnormalities visible in 25% of the embryos as described by Haffter *et al.* (6). Four recessive lethal mutations tightly linked to proviral insertions have been identified. The *nar* mutation and gene are described elsewhere (9, see Chapter 4). The 80A mutation has not been studied further since a disrupted gene has not yet been identified for this mutant. Here we describe the gene and phenotype associated with proviral insertion 67D which we have named *pescadillo* (*pes*). The fourth mutation, *dead eye,* will not be discussed.

The *pes* gene encodes a highly conserved novel protein

Previously we reported that the proviral insertion designated 67D is genetically linked to, and presumably caused, the zebrafish *pescadillo* mutation (9). This insertion lies within the first coding exon of a gene which was discovered because of high homology between genomic sequences flanking the 67D insertion and a human Expressed Sequence Tag (EST) (Genbank accession number R13806). The regions of homology were presumed to be exon sequences and were used to design primers for RT-PCR. Analysis of the amplified RT-PCR products confirmed the predicted intron-exon structure in the region and showed that the provirus lies 80 base pairs upstream of the putative methionine initiation codon (9, and data not shown).

To learn more about the *pes* gene and its encoded protein, we screened a zebrafish cDNA library prepared from day 3 embryonic mRNA using a 300 bp cDNA fragment amplified by RT-PCR as a probe. Two clones with cDNA inserts of apparently equal size were obtained and one was sequenced. This clone includes a 2214 base pair insert. A putative protein coding region was identified in this sequence based on homology with the human EST sequence. This coding region

corresponds to the longest ORF found and encodes a protein of 582 amino acids (Fig. 1). The region surrounding the presumed methionine ATG initiation codon conforms to the consensus translation start site (10) and is preceded by an in-frame stop codon located 21 nucleotides upstream. The 5' and 3' untranslated regions are 99 and 368 bp, respectively. The entire cDNA sequence upstream of the proviral insertion point is contiguous with the genomic DNA, suggesting that the insertion is within the first exon. It cannot be ruled out, however, that the cDNAs obtained are incomplete and that there are additional exons further upstream. No poly A sequence was found at the 3' terminus of this cDNA.

Three human cDNA clones containing the EST identified by computer search were obtained from ATCC and the longest was sequenced in its entirety. Comparison of the putative proteins encoded by the human and zebrafish transcripts show 74% identity (Fig. 1). Further database searches using the predicted *pes* amino acid sequence were carried out against the translated Genbank database (TBLASTN) (11). Significantly homologous sequences were identified in cosmid clones of genomic DNA isolated in the *C. elegans* and yeast genome projects (Genbank accession number D75131 and yeast ORF name YGR103W, respectively), and in mouse EST sequences (Genbank accession number AA003101). Amino acid identity between the zebrafish *pes* protein and the *S. cerevisiae* homologue is about 54% in the amino-terminal third of the protein, 39% overall (Fig. 1).

Analysis of the predicted *pes* protein sequence did not reveal any known motifs which would suggest a possible cellular function. A highly acidic region at the carboxy terminus of the protein is conserved among the zebrafish, human and yeast genes (underlined in Fig. 1). Most of the *pes* protein is predicted to have an alpha-helical structure as determined by analysis with the Robson/Garnier secondary structure algorithm (M. Robinson, Amgen, personal communication).

Figure 1

```
1   MGGLQKKKYESGSATNYITRNKARKKLSLSLADFRRLCILKGIYPHEPKHKKKVNKGSTAARTYYLLKDIRFLLHEPIVGKFREYKKIFVRKLRKAYGKAEWSAVERLKENKPGYKLDHII Zf
1   MGGLEKKKYERGSATNYITRNKARKKLQLSLADFRRLCILKGIYPHEPKHKKKVNKGSTAARTFYLIKDIRFLLHEPIVNKFREYKKVFVRKLRKAYGKSEWNTVERLKDNKPNYKLDHII Hu
1   MR--IKKKNTRGNARNFITRSQAVRKLQVSLADFRRLCIFKGIYPREPRNKKKANKGSTAPTTFYYAKDIQYLMHEPVLAKFREHKTFARKLTRALGRGEVSSAKRLEENRDSYTLDHII Yst

121 KERYPTFIDALRDVDDALSMCFLFSTFARTGKCHVQTIQLCRRLSVEWMNYIISSRSLRKVFLSIKGIYYQAEVLGQTITWIVPYQFAHNHPTDVDYRVMATFTELYTTLLGFVNFRLYQ Zf
121 KERYPTFIDALRDLDDALSMCFLFSTFPRTGKCHVQTIQLCRRLTVEFMHYIIAARALRKVFLSIKGIYYQAEVLGQPIVWITPYAFSHDHPTDVDYRVMATFTEFYTTLLGFVNFRLYQ Hu
119 KERYPSFPDAIRDIDDALNMLFLFSNLPSTNQVSSKIINDAQKICNQWLAYVAKERLVRKVFVSIKGVYYQANIKGEEVRNLVPFKFPENIPSDVDFRIMLTFLEFYSTLLHFVLYKLYT Yst

241 TLNLVYPPKLDGQGEISLKAEFEEDYALES--ESYTEKL----------------SALSASLARMVASVEEEEAELDHFPTEGEDQEKMEVREKMEQQQSKQKK---------------- Zf
241 LLNLHYPPKLEGQAQAEAKAG-EGTYALDS--ESCMEKL----------------AALSASLARVVVPATEEEAEVDEFPTDGEMSAQEEDRRKELEAQEKHKK---------------- Hu
239 DSGLIYPPKLDLKKDKIISG--LSSYILESRQEDSLLKLDPTEIEEDVKVESLDASTLKSALNADEANTDETEKEEEQEKKQEKEQEKEQNEETELDTFEDNNKNKGDILIQPSKYDSPV Yst

327 --LFEGLKFFLNREVPRESLAFVIRCFGGEVSWDKSLC-IGSTYEATDETITHIVDRPSMDKQYINRYYIQPQWVYDSVNAKIQLPVEEYFLGVTLPPHLSPFVEETEGDYVPPEKLKL Zf
326 --LFEGLKFFLNREVPREALAFIIRSFGGEVSWDKSLC-IGATYDVTDSRITHQIVDRPGQQTSVIGRCYVQPQWVFDSVNARLLLPVAEYFSGVQLPPHLSPFVTEKEGDYVPPEKLKL Hu
357 ASLFSAFVFYVSREVPIDILEFLILSCGGNVISEAAMDQIENKKDIDMSKVTHQIVDRPVLKNKVAGRTYIQPQWIFDCINKGELVPANKYLPGEALPPHLSPWGDAI--GYDPTA---- Yst

444 MALQRGEKP----QAEEDEEEEGEEEEDDEEDEEDDEQSEDEE---EAEEEANLAEMEEKRSQGKSLSVKVTPGKAKAENRARAAEEEKAEEKRLAIMMMKKKEKYLYDKIMFGKKRKVR Zf
443 LALQRGEDPGNLNESEEEEEEEDDNNEGDGDEEGENEEEEEDAEAGSEKEEEARLAALEEQRMEGK--KPRVMAGTLKLEDKQRLAQEEESEAKRLAIMMMKKREKYLYQKIMFGKRRKIR Hu
471 -PVEEGEEEESESESESEDQVE-EEDQEVVAGEEDDDDDEELQAQKELELEAQGIKYSETSEADKDVN--------KSKNKKRKVDEEE-EEKKLKMIMMSNKQKKLYKKMKYSNAKKEE Yst

557 EANKLAAKRKAHDDASKADKKKKK---KC     582   Zf
561 EANKLAEKRKAHDEAVRSEKKAKKARPE      588   Hu
580 QAENL--KKKKKQIAKQKAKLNKLDSKK      605   Yst
```

FIG. 1 Zebrafish *pes* predicted protein sequence and amino acid alignment with *pes* homologs. Amino acid sequence alignment of the predicted zebrafish *pes* protein with the human and yeast homologs as compiled by the Lasergene alignment tool. Shaded regions correspond to identities among the sequences. Underlined amino acids correspond to a highly acidic region conserved in *pes* homologs. The zebrafish and human cDNA sequences have been deposited in the Genbank database (accession numbers U77627 and U78310, respectively).

Expression of the *pes* gene is developmentally regulated

If mutation of the *pes* gene is responsible for the mutant phenotype in *pes/pes* embryos, gene expression would be expected to occur at or before the mutant phenotype becomes visible at day 3 of embryogenesis. Thus, we analyzed the timing and tissue distribution of expression of the *pes* gene during embryogenesis by northern blot and *in situ* hybridization.

A 300 bp *pes* cDNA fragment was radiolabeled and was hybridized to RNA prepared from several embryonic stages and adult fish in a northern blot (Fig. 2). Two transcripts, 2.2 Kb and 1.9 Kb, were detected, with the larger being more prevalent during embryogenesis. Unfertilized eggs and gastrulating embryos (6 hours post-fertilization) have low levels of the transcripts, while strong zygotic expression is seen by 12 hours. Transcript levels decrease after about 24 hours. In adult fish, only females have detectable *pes* RNA and dissection of the ovaries shows that it is restricted to this organ (Fig. 2).

A 2 Kb cDNA fragment of the *pes* gene was used to synthesize digoxigenin-labeled RNA probes for whole mount *in situ* hybridization. Sense-strand-specific probe did not produce signal at the stages examined. Using antisense-strand probe, *pes* transcripts could not be detected in 3 or 6 hour embryos, but were detected beginning at 12 hours in the eye and brain primordia (not shown). Between 18 hours and three days of development, the expression of *pes* message is highly dynamic. At 18 to 24 hours, strong expression is detected in the eye, forebrain, tectum, and somites, while lower levels of transcript are seen in the hindbrain and in cells flanking the hindbrain (Fig. 3A, and data not shown). During the second day of development (28 to 36 hours) expression levels begin to decrease and by 48 hours the distribution of transcripts is restricted to the ganglion cell layer of the eye, the ventral forebrain, cells in the posterior tectum at the midbrain-hindbrain boundary, rows of cells in the pharyngeal arches, the pectoral fin buds, the liver and pancreatic

Figure 2

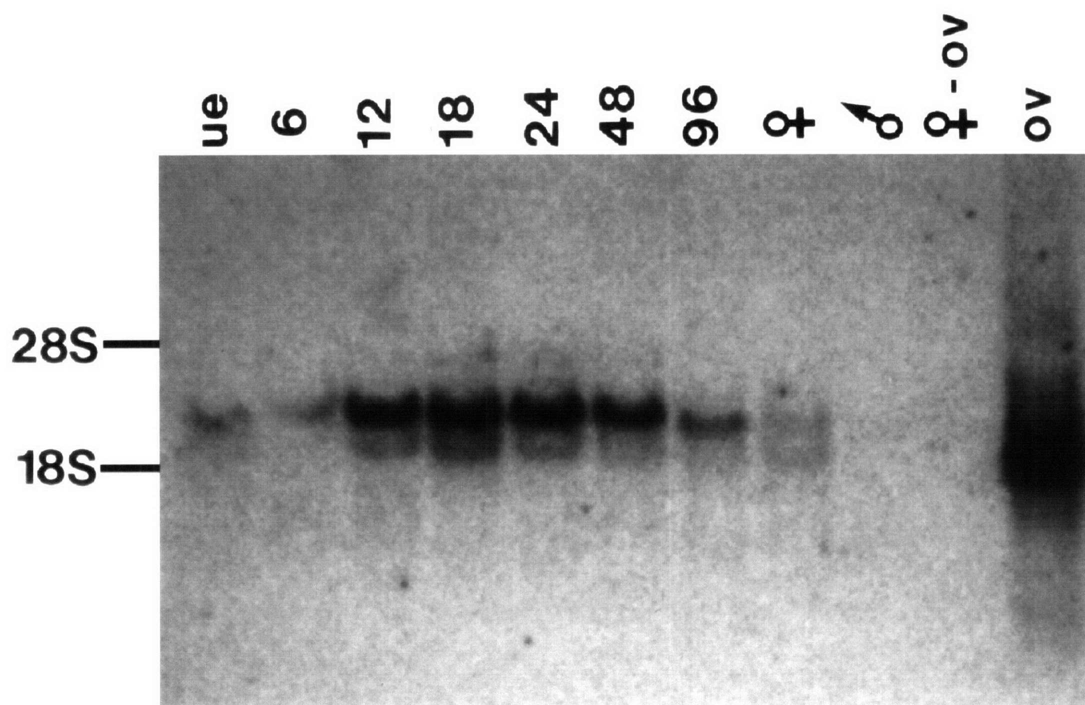


Figure 3

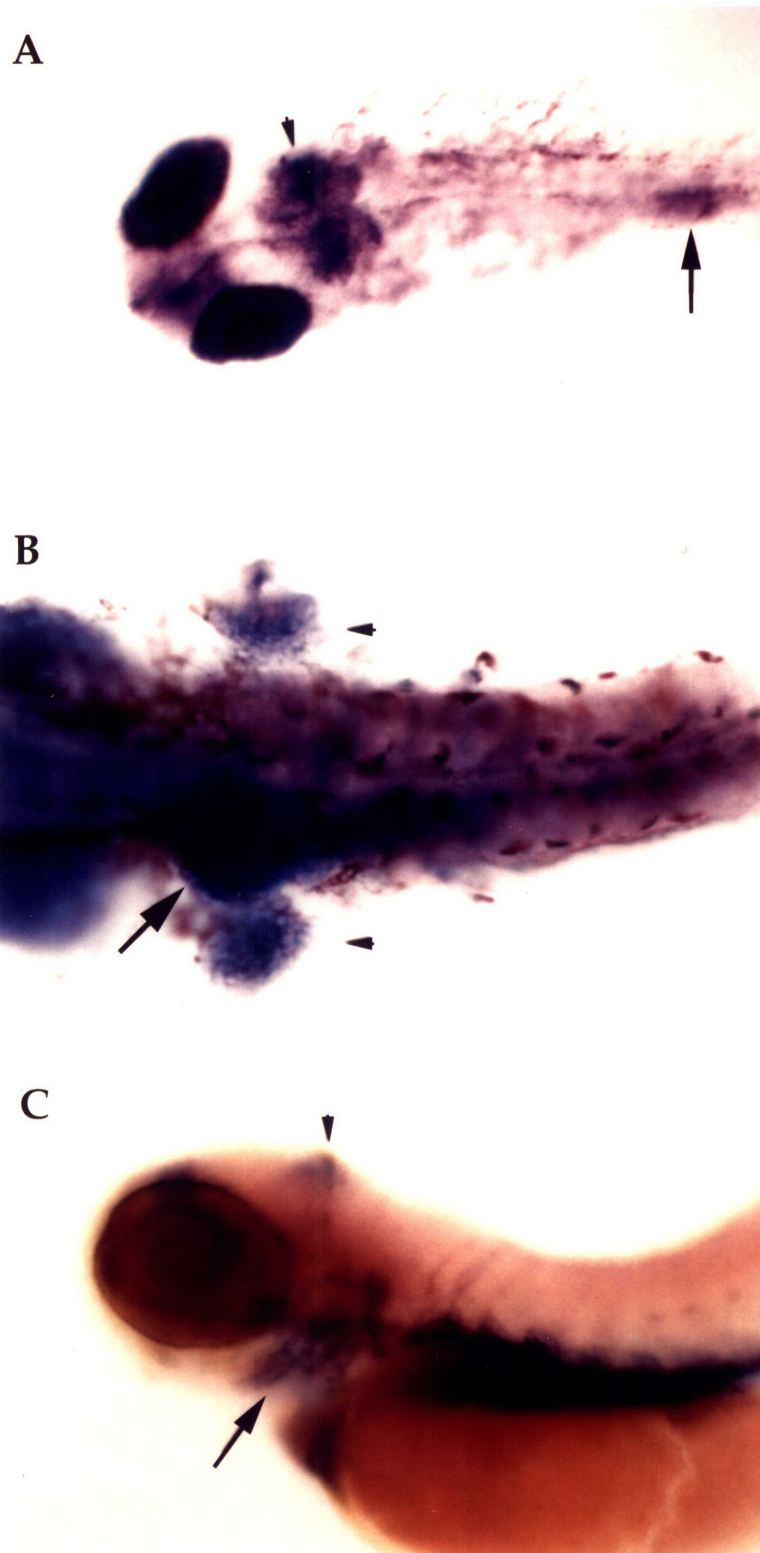


FIG. 2 Northern analysis of *pes* expression. A *pes* cDNA fragment was hybridized to total RNA isolated from the indicated embryonic stages and adult zebrafish by northern analysis. The numbers shown correspond to hours post-fertilization; ue, unfertilized egg RNA; ov, ovary. The last four lanes on the right correspond to RNA from adult female, adult male, adult female without ovary, and ovary. A 2.2 Kb product is prevalent during embryogenesis while a 1.9 Kb band is expressed at low levels in the embryo but is the more abundant product in adult ovary. RNA from adult male and from females whose ovaries have been removed have no detectable *pes* product.

FIG. 3 Restricted expression of *pes* between days one and three post-fertilization. (A) 24 hour old embryo whose yolk has been removed, observed dorsally, anterior is to the left. Note heavy expression in the eye primordia, in the forebrain, the tectum (arrowhead), and in the asymmetrically localized liver primordium (arrow). (B) 48 hours. The embryo was dissected as in (A) and shows heavy hybridization in the developing liver (arrow) and in the fin buds (arrowheads). Anterior is to the left. (C) 72 hours. Expression of *pes* in the tectum is reduced to a row of cells (arrowhead), compare to stain at 28 hours (A). Expression in the pharyngeal arches can be seen at this stage (arrow). The pancreas and gut express heavily (dark stain to the right of arch staining). Anterior is to the left and dorsal is up. Bars A, C=100μm; B=50μm.

primordia, and the presumptive gut (Fig. 3B, and data not shown). At 72 hours, the tectal row of cells continues to express *pes* transcripts. RNA levels in the liver are lower than at 48 hours while the developing gut continues to express high levels of *pes* mRNA. Striking expression is seen in a series of stripes coincident with the expanding branchial arches (Fig. 3C). The distribution of *pes* transcripts in embryos older than 72 hours was not analyzed in whole mount material due to the difficulty in obtaining efficient penetration of probe to all tissues.

In situ hybridized embryos were sectioned to confirm and further specify the sites of expression inferred from the whole mount preparations (data not shown). We ascertained that high levels of *pes* message are found in all pharyngeal arch primordia and in the nascent anterior neurocranium (the ethmoid plate) at day 2 post fertilization. We also observed heavy expression in the gut epithelium and in the pancreas at day 3.

Previous northern blot analysis had shown that *pes* transcripts could not be detected in five day old homozygous mutant embryos (9). To rule out the possibility that the absence of *pes* RNA at day 5 is due to loss of expressing tissues, *in situ* hybridization was carried out on progeny obtained from crosses of heterozygous *pes*/+ parent fish prior to the appearance of the mutant phenotype. When one or two day old embryos were tested, approximately 25% of the animals showed no detectable staining (38/163 over 6 experiments) whereas 100% of embryos obtained from wild-type crosses (177/177 over 7 experiments) were stained (data not shown). This furthers the notion that in homozygous mutant embryos *pes* transcripts are either not synthesized at all or are present in undetectable amounts. A similar analysis done on three day old embryos, when the *pes* phenotype becomes apparent, confirmed that it is the mutant animals in which hybridization signal is not observed.

The *pes* mutation affects the development of a subset of embryonic primordia which correlate with sites of strong *pes* gene expression

The *pes* mutant phenotype is first evident under a dissecting microscope on the third day of development as a reduction in the size of the head and eyes and incomplete extension of the jaw when compared to wild-type. To further analyze the deficiency within the cranial skeleton, we stained differentiated cartilage in mutant and wild-type embryos with alcian blue (12). The early pharyngeal skeleton normally consists of a series of seven distinct arches: the mandibular (P1) and hyoid (P2), both of which will form the jaw, and five branchial arches (P3 to P7), which eventually will support the gills (13) (Fig. 4). In addition, alcian blue labels the developing neurocranium, which underlies the brain, and cartilage in the fin primordia.

A striking aspect of the *pes* phenotype is the absence of stained cartilage in the five branchial arches (P3 to P7) and a severe reduction of the jaw arches (P1 and P2) relative to wild-type. This is apparent by 80 hours post-fertilization, shortly after cartilage begins to develop in the arches (Fig. 4A top). In day 5 wild-type embryos, the skeletal architecture is more complex than at day 3, whereas in mutants, cartilage in the anterior jaw arches appears unchanged (Fig. 4A, center). Although cartilage fails to differentiate in the branchial arches of mutant embryos, mesenchymal tissue is organized segmentally in this region as can be seen in longitudinal sections through the head region of *pes/pes* embryos (Fig. 5B). Occasionally, a few cells within the first (P3), and less often the second (P4), branchial arch stain lightly by day 5, and in some mutant embryos, muscle fibers can be seen juxtaposed to the undifferentiated branchial arch primordia (not shown). In addition to the arch defects, the ethmoid plate, the anterior portion of the neurocranium, appears shortened in mutant embryos and cartilage in the pectoral fins is reduced, resulting in short fins (Fig. 4A, top, and not shown).

Figure 4

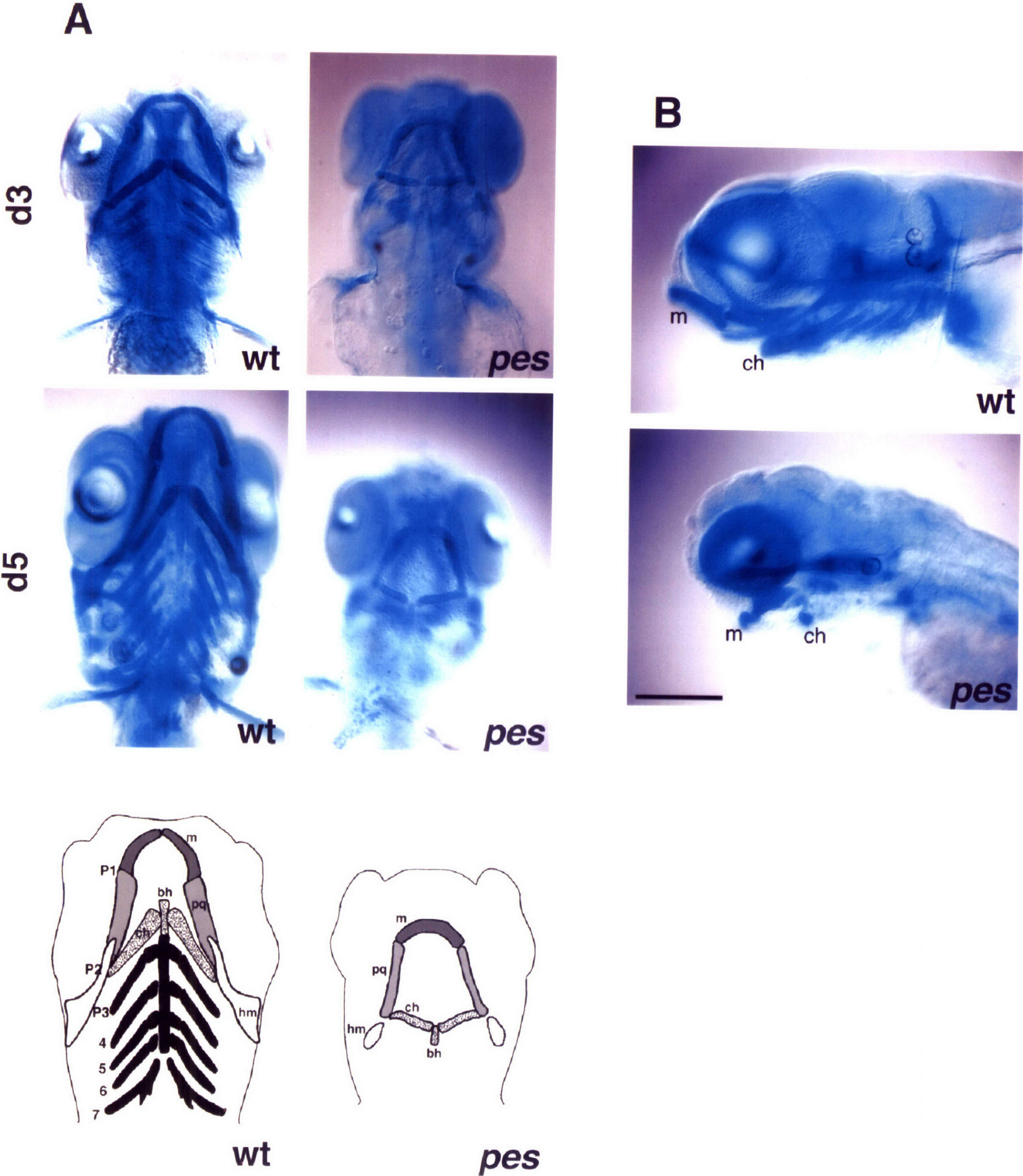
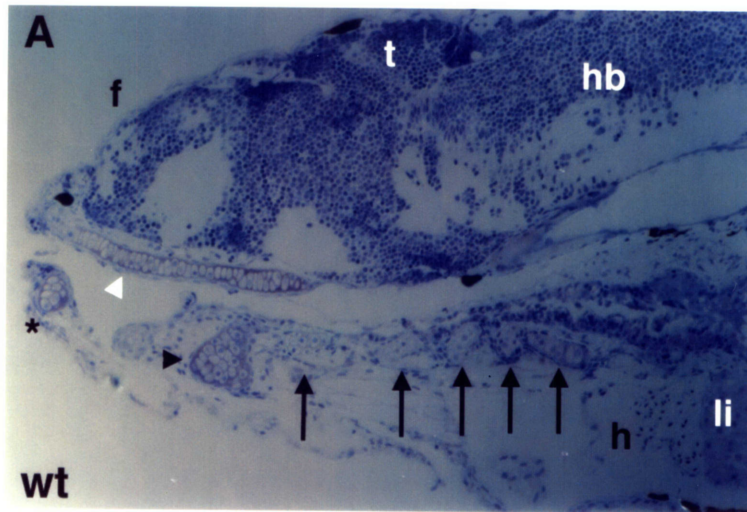


Figure 5



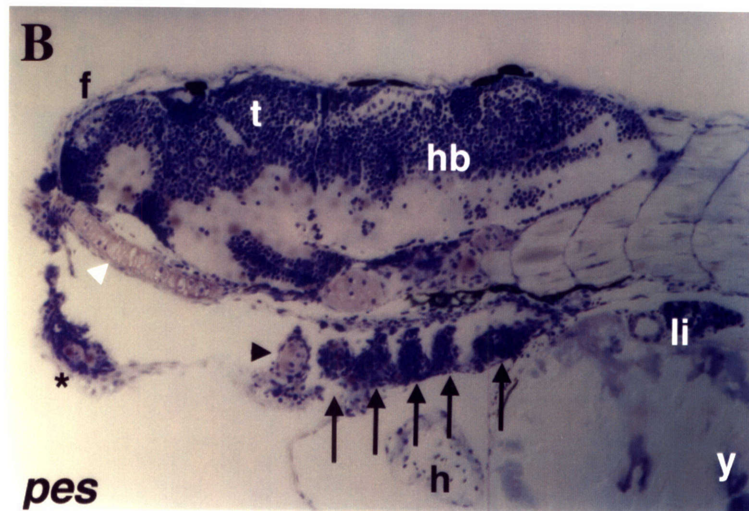


FIG. 4 Pharyngeal cartilage formation is defective in *pes* mutants. (A) Ventral views of three (top) and five (center) day old wild-type (wt) and mutant (*pes*) embryos stained with alcian blue. Note the differing head sizes between wild-type and mutant at day 3 and the lack of growth of cartilage in the mutant by day 5. The principal cartilaginous elements of the wild-type and mutant pharyngeal skeleton are shown schematically below the stained embryos: the first arch (P1), including Meckel's cartilage (m) and palatoquadrate (pq); the second arch (P2), including the hyomandibular (hm), the paired ceratohyals (ch) and the medial basihyal (bh), and branchial arches, P3 to P7. In the mutant, both elements of Meckel's cartilage have fused at the midline, the hyomandibular is reduced, the ceratohyals are pointing ventro-caudally, and the cartilage of the branchial arches absent. (B) Lateral views of day 5 wild-type (wt) and mutant (*pes*) embryos stained with alcian blue. The short, ventrally protruding, Meckel's cartilage (m) and ceratohyal (ch) are the only stained structures seen in the pharyngeal region of the mutant. Bar=100μm.

FIG. 5 Sagittal sections of five day old wild-type (A), and *pescadillo* (B) mutant embryos. Anterior is to the left, dorsal is up, and the plane of sections is medial. In *pes* mutant animals, the brain, particularly the tectum (t), is smaller than in wild-type animals, the neurocranium is shorter and thicker, and the posterior pharyngeal arches (arrows) lack differentiated cartilage (A, B). At the position of the five branchial arches however (arrows in B), mesenchymal tissue is organized into segmental bundles. The liver (li) is very reduced and the yolk (y) has not been consumed. e, eye; f, forebrain; h, heart; hb, hindbrain; li, liver; t, tectum (dorsal midbrain); y, yolk; asterisk, Meckel's cartilage (first arch, P1); black arrowhead, hyoid (second arch, P2); black arrows, branchial arches (P3 through P7); white arrowhead, anterior neurocranium. Bar=100μm.

Transverse sections through the trunk region reveal that the internal organs of mutant and wild-type embryos are indistinguishable in size at day 3 of embryogenesis (Fig. 6A vs. 6B). However, between days three and five of development, striking differences appear: In wild-type embryos the liver grows and extends over the yolk surface and, concomitantly, the yolk is rapidly consumed (Fig. 6C); In mutants the liver does not grow substantially after day 3 and the yolk is not consumed (Fig. 6D). Moreover, the gut is markedly reduced in mutants and the anterior expansion of the intestine that forms the stomach in cyprinid fishes (14) does not develop. In addition to the striking differences in the expansion of the liver and gut, the pancreas, which appears darkly labeled by the counterstain in sections of five day-old wild-type embryos, was not discernible in mutants. Furthermore, in mutants the axial musculature shows signs of degeneration and the body wall muscle, which surrounds the abdominal cavity in wild-type embryos, is absent (Fig. 6C vs. 6D).

At the level of resolution presented here, there appears to be a striking correlation between defects in *pes* mutant embryos and regions where *pes* is normally expressed at high levels earlier in development: the branchial arches, liver and gut, as well as brain, eyes, fin buds, and ethmoid plate. In most cases the primordia develop, but between days three and five they fail to expand. Importantly, other structures appear normal in mutants at day 5, even though many other regions of the animal are severely affected. Tissues that appear normal include the notochord, the pronephros and renal tubules (Fig. 6C vs. 6D, and not shown). Expression of *pes* message was not detected in these tissues at any stage in wild-type animals. To firmly establish that the *pes* phenotype is entirely restricted to cells that normally express *pes* product will require further experiments. Likewise, we cannot yet rule out that *pes* mRNA expression occurs at some point during embryogenesis in tissues that appear normal in mutant animals.

Figure 6

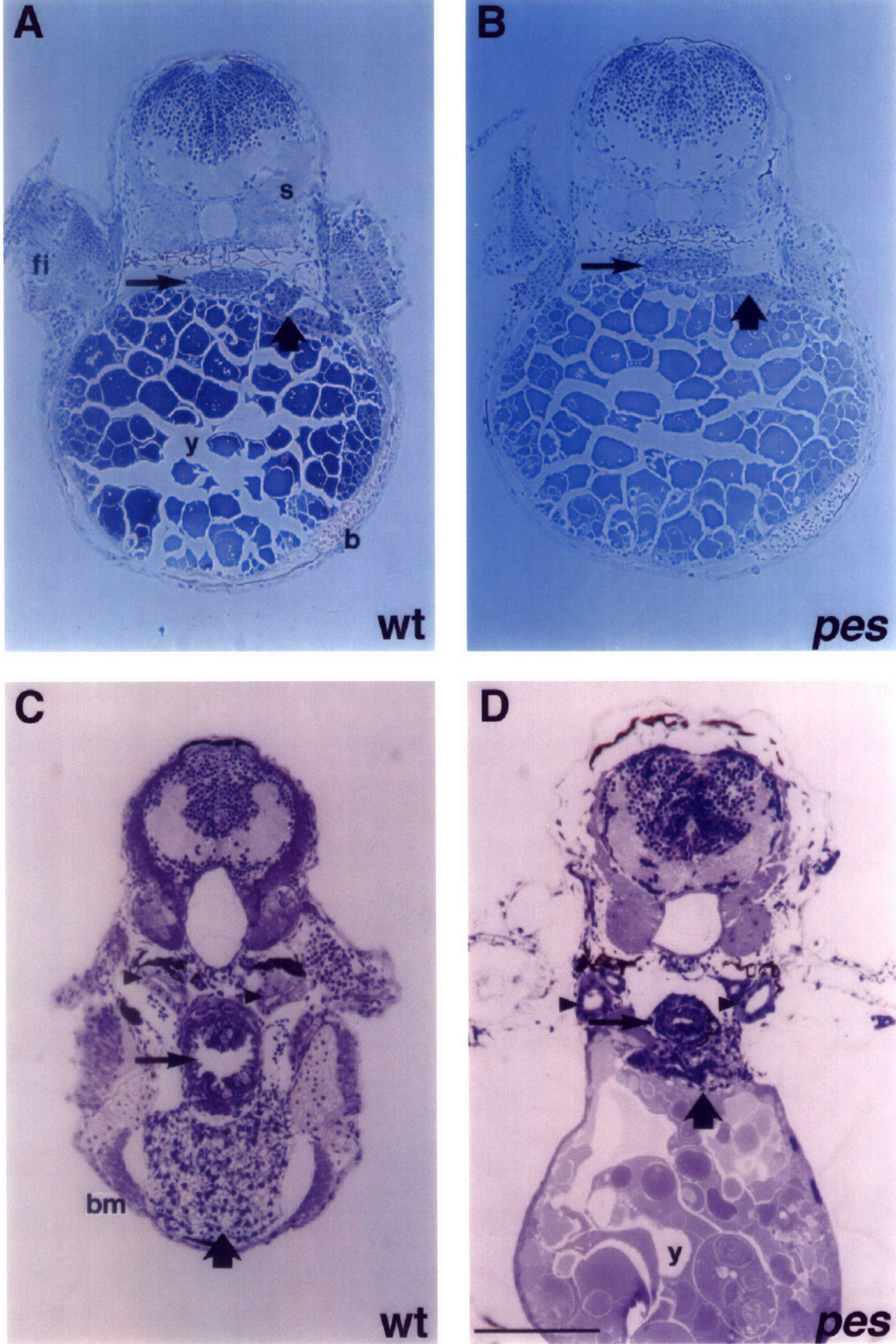


FIG. 6 Comparison of day 3 and day 5 cross sections through the anterior trunk region of wild-type and *pes* mutant embryos. (A, B) At day 3, relative sizes of liver (thick arrow) and gut (thin arrow) in the mutant (B) are virtually indistinguishable from wild-type (A), as are trunk muscles (s) and hindbrain. Both animals were sectioned at the level of the pectoral fins (fi). Circulating blood is visible (b). (C, D) At day 5, the liver (thick arrows) has greatly expanded in wild-type (C), but not in the mutant (D). Also in the mutant the yolk (y) is unconsumed, the gut (thin arrows) has failed to expand, the body wall muscles (bm), which surround the abdomen, are absent, the axial musculature has started to degenerate and all body cavities have expanded and filled with fluid. The renal tubules (arrowheads), however, appear to be of similar size and position in mutant and wild-type. Bar A, B=50μm; C, D=100μm.

DISCUSSION

We have presented evidence that *pescadillo* is an insertional mutant and have described the gene whose disruption is likely to be responsible for the mutant phenotype. We had previously shown (a) genetic linkage of provirus 67D to a mutant phenotype (b) the provirus lies in the 5' exon of a gene that is expressed in embryos, and (c) that the provirus abolishes detectable gene expression. Here we have provided additional strong support that *pes* is an insertional mutant in the gene we identified by showing that at least many sites of *pescadillo* expression in wild-type embryos correspond to regions of the embryo that fail to reach normal size in *pes/pes* mutants. The evidence presented suggests that it is highly likely that a mutation in the *pes* gene is responsible for the mutant phenotype observed. Definitive proof will require either rescue of the mutant phenotype by introducing the gene or its product into mutant animals, or possibly targeted mutation of these genes. Neither technology has yet been reported in zebrafish. Targeted disruption could be performed in mice (15-18) since the *pes* gene is so highly conserved among vertebrates, but only a positive result would be informative since mice and fish may differ in their genetic redundancy and since homologous genes could serve different functions even among vertebrate species.

We rapidly identified genes disrupted in three of the four insertional mutants isolated to date. The most recent case required only 3 weeks of work by a single individual (the *dead eye* gene, which is not discussed here). The reasons for this speed were (a) the proximity of the mutagenic proviral insertions to coding sequences, and (b) the fact that the coding sequence was homologous to sequences present in the database. We have not yet located a gene near the previously described 80A insertion (9). We may not have enough sequence data yet (1 Kb on one side, 4 Kb on the other). Alternatively, this insertional mutant may involve a

gene sequence that is not in the database or one that is not evolutionarily conserved, or the 80A insertion may lie in a large intron or in a regulatory sequence distant from the coding region of the disrupted gene.

It is interesting to note that the proviruses in three of the four mutants we have isolated are less than 500 bp upstream of the translation initiation codon of the mutated genes, and that in all three cases the proviruses point in the opposite transcriptional orientation from the gene they disrupt (unpublished data). While these data are in small quantity, they raise important questions regarding the randomness of proviral integrations into zebrafish genes. Despite much effort, whether mouse and chicken C-type retroviruses integrate randomly into the host cell genome remains controversial. Some evidence suggests that proviruses integrate preferentially into the 5' end of genes (19-21), or into hot spots throughout the genome (22), while other evidence indicates that integration is essentially random and that all regions of the genome are comparable targets (23).

Because the mutagenic insertions we have isolated are a biased pool they are of limited use for evaluating the randomness of proviral integration into the zebrafish genome. However, the fact that all three appear to have integrated similarly into the corresponding disrupted genes suggests several possibilities: (a) Proviral insertions in zebrafish are biased toward the 5' end of genes, and/or (b) integrations into the 5' end of genes are much more likely to disrupt gene function. The fact that only about 1 in 70 insertions appears to cause an embryonic mutation in zebrafish makes a bias toward the 5' end of genes seem unlikely, unless insertions have very limited mutagenicity in such regions. Alternatively, proviral insertions may occur randomly throughout the zebrafish genome, but only those into the 5' end of genes may be disruptive. If, for example, insertions into intron sequences are spliced out they may have little or no effect on the gene into which they have integrated. The analysis of many more zebrafish insertional mutations will be

required to thoroughly address this issue, and such an analysis may be of value for the future design of more mutagenic viruses.

The further implications of the results presented here to the future potential of insertional mutagenesis in zebrafish will be discussed in Chapter 6

MATERIALS AND METHODS

Animals: Zebrafish (*Danio rerio*) were kept and raised essentially according to standard conditions (24) and using practices established in our laboratory (25). The aquarium systems used were designed specifically for housing large numbers of animals in small containers (26) and were purchased from K.-J. Schwarz Glas Aquarienbau (Göttingen, Germany). Fertilization was achieved by natural spawning and embryos were raised at 28°C and staged according to Kimmel *et al.* (27). The insertional mutant pilot screen was carried out by inbreeding fish harboring identical proviral insertions (9) and scoring their progeny for several morphological criteria under low magnification as described (6).

Northern blot and cDNA Isolation: For the northern analysis, total RNA (15μg) from each sample was fractionated on a 2M formaldehyde agarose gel and transferred to a nylon filter (Hybond N+, Amersham). The blot was probed with a radiolabeled 292 bp RT-PCR cDNA fragment (corresponding to nucleotides 61 to 353 of the *pes* cDNA). Exposure was performed on Kodak BioMax MS film for 6 days. The same cDNA fragment was used as a probe to screen a three day embryonic cDNA library (obtained from Kai Zinn, CalTech). 5×10^5 plaques were screened and two positive clones were identified and isolated. Both clones contained inserts of identical length by PCR and one of them was sequenced in its entirety on both strands. The human cDNA clones were purchased from American Type Culture

Collection (Rockville, MD), and were sequenced on both strands. Sequence alignment was accomplished by using the Lasergene software (DNAStar, Inc.).

Histology: Alcian blue staining was done as described (9). The nomenclature for skeletal elements is that described by Schilling *et al.* (13). For tissue sectioning, embryos were fixed in 4% paraformaldehyde/PBS, dehydrated and embedded in Polybed 812 epoxy resin (Polysciences). Specimens were cut into 1-2µm sections which were subsequently counterstained with a solution of 0.05% crystal violet, 0.01% methylene blue, and 0.05% borax at 95°C.

In situ hybridization: UTP-11-digoxigenin labeled RNA probes were prepared as suggested by the manufacturer (Boehringer Mannheim Biochemicals). The probe used was an *in vitro* transcription product of a 2 Kb fragment of the *pes* cDNA (corresponding to nucleotides 214 to 2214 of the *pes* cDNA). Anesthetized embryos were fixed in 4% paraformaldehyde in PBS at 4°C for 12-16 hours and were dehydrated in methanol at -20°C for at least 1 hour. *In situ* hybridization was carried out essentially after Jowett and Lettice (28). Proteinase K treatment was for 10 minutes at 10µg/ml for embryos up to 24 hours old and at 25µg/ml for older embryos. Prehybridization and hybridization temperature was 65°C with a probe concentration of 1µg/ml. Hybridized embryos were cleared in glycerol or in methyl salicylate and were photographed under a Nikon Microphot SA microscope.

REFERENCES

1. Streisinger G; Walker C; Dower N; Knauber D; Singer F (1981) Production of clones of homozygous diploid zebrafish (*Brachydanio rerio*). *Nature* **291**: 293-6.

2. Kimmel CB (1989) Genetics and early development of zebrafish. *Trends Genet* **5**: 283-8.

3. Rossant J; Hopkins N (1992) Of fin and fur: mutational analysis of vertebrate embryonic development. *Genes Dev* **6**: 1-13.

4. Nüsslein-Volhard C (1994) Of flies and fishes. *Science* **266**: 572-4.

5. Driever W; Solnica-Krezel L; Schier A; Neuhauss S; Maliki J; Stemple D; Stainier D; Zwartkruis F; Abdelilah S; Rangini Z; Belak J; Boggs C (1996) A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123**: 37-46.

6. Haffter P; Granato M; Brand M; Mullins M; Hammerschmidt M; Kane D; Odenthal J; van Eeden F; Jiang Y; Heisenberg C-P; Kelsh R; Furutani-Seiki M; Vogelsang L; Beuchle D; Schach U; Fabian C; Nüsslein-Volhard C (1996) The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* **123**: 1-36.

7. Lin S; Gaiano N; Culp P; Burns J; Friedman T; Yee J-K; Hopkins N (1994) Integration and germ-line transmission of a pseudotyped retroviral vector in zebrafish. *Science* **265**: 666-9.

8. Gaiano N; Allende M; Amsterdam A; Kawakami K; Hopkins N (1996) Highly efficient germ-line transmission of proviral insertions in zebrafish. *Proc. Natl. Acad. Sci. USA* **93**: 7777-82.

9. Gaiano N; Amsterdam A; Kawakami K; Allende M; Becker T; Hopkins N (1996) Insertional mutagenesis and rapid cloning of essential genes in zebrafish. *Nature* **383**: 829-32.

10. Kozak M (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* **15**: 8125-48.

11. Altschul S; Gish W; Miller W; Myers E; Lipman D (1990) Basic Local Alignment Search Tool. *J Mol Biol* **215**: 403-10.

12. Dingerkus G; Uhler D (1977) Enzyme clearing of alcian blue stained whole small vertebrates for demonstration of cartilage. *Stain Tech* **32**: 229-31.

13. Schilling T; Piotrowski T; Grandel H; Brand M; Heisenberg C-P; Jiang Y-J; Beuchle D; Hammerschmidt M; Kane D; Mullins M; van Eeden F; Kelsh R; Furutani-Seiki M; Granato M; Haffter P; Odenthal J; Warga R; Trowe T; Nüsslein-Volhard C (1996) Jaw and branchial arch mutants in zebrafish I: branchial arches. *Development* **123**: 329-44.

14. Takashima F; Hibiya T (1995) An atlas of fish histology 2nd ed (Tokyo: Kodansha Ltd)

15. Mansuor SL; Thomas KR; Capecchi MR (1988) Dispruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**: 348-52.

16. Joyner AL; Skarnes WL; Rossant J (1989) Production of a mutation in mouse *En-2* gene by homologous recombination in embryonic stem cells. *Nature* **338**: 153-6.

17. Schwartzberg PO; Goff SP; Robertson EJ (1989) Germ-line transmission of a c-abl mutation produced by targeted gene disruption of ES cells. *Science* **246**: 799-803.

18. Zijlstra M; Li E; Sajjadi F; Subramani S; Jaenisch R (1989) Germ-line transmission of a disrupted β_2 microglobulin gene produced by homologous recombination in embryonic stem cells. *Nature* **342**: 435-8.

19. Vijaya S; Steffen D; Robinson H (1986) Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. *J Virol* **60**: 683-92.

20. Rohdewohld H; Weiher H; Reik W; Jaenisch R; Breindl M (1987) Retrovirus integration and chromatin structure: Moloney Murine Leukemia proviral integration sites map near DNase I-hypersensitive sites. *J Virol* **61**: 336-43.

21. Scherdin U; Rhodes K; Breindl M (1990) Transcriptionally active genome regions are preferred targets for retrovirus integration. *J Virol* **64**: 907-12.

22. Shih C-C; Stoye JP; Coffin JM (1988) Highly preferred targets for retrovirus integration. *Cell* **53**: 531-7.

23. Withers-Ward ES; Kitamura Y; Barnes JP; Coffin JM (1994) Distribution of targets for avian retrovirus DNA integration in vivo. *Genes Dev* **8**: 1473-87.

24. Westerfield M (1995)*The Zebrafish Book*. Eugene: University of Oregon Press.

25. Culp P; Nüsslein-Volhard C; Hopkins N (1991) High-frequency germ-line transmission of plasmid DNA sequences injected into fertilized zebrafish eggs. *Proc Natl Acad Sci USA* **88**: 7953-7.

26. Mullins M; Hammerschmidt M; Haffter P; Nüsslein-Volhard C (1994) Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. *Curr Biol* **4**: 189-202.

27. Kimmel CB; Ballard W; Kimmel S; Ullmann B; Schilling T (1995) Stages of embryonic development of the zebrafish. *Dev Dyn* **203**: 253-310.

28. Jowett T; Lettice L (1994) Whole mount *in situ* hybridization on zebrafish embryos using a mixture of digoxigenin- and fluorescein-labelled probes. *Trends Genet* **10**: 73-4.

CHAPTER 6

FUTURE PROSPECTS FOR RETROVIRAL INSERTIONAL MUTAGENESIS IN ZEBRAFISH

The results presented in Chapters 2-5 demonstrate that an MLV/VSV pseudotyped retroviral vector can be used to generate large numbers of transgenic zebrafish and to induce insertional mutations in this vertebrate. However, the ultimate value of this nascent methodology has yet to be determined as several issues remain unresolved. Perhaps most important among these issues is the randomness of proviral insertion site selection in the zebrafish genome. If only a small fraction of essential zebrafish genes can be mutated by proviral integration then this method will be of limited use. Also relevant to the value of this method is the extent to which the zebrafish genome project will facilitate the cloning of genes mutated by chemical lesions. If the cloning of chemical alleles by positional or candidate approaches could be performed rapidly and routinely then the primary advantage of insertional mutagenesis over chemical mutagenesis would no longer exist. Finally, while the current method is adequate to perform screens designed to isolate hundreds of insertional mutations, further technological advances could significantly improve the efficiency of this method and allow the isolation of thousands of insertional mutations.

Randomness of Retroviral Integration

The utility of a large-scale mutagenesis method is determined in large part by its ability to mutate most if not all of the genes necessary during a process of interest. Mutagen biases could result in some genes being frequent targets while other may rarely if ever be hit, a possibility which would limit the number of genes which could be mutated by a mutagenesis strategy. Evidence for the nonrandom integration of insertional elements has come from studies in numerous organisms including yeast, flies, mice and birds. For example, as mentioned in Chapter 1, large-scale insertional mutagenesis screens in fruit flies have suggested that only 30-50% of the genes which can be mutagenized by chemical mutagens can also be disrupted by P

element insertion (1). The reason for this bias is unknown. It is possible that only certain chromatin conformations are accessible to P insertion, or that these elements are directed to certain sites by the integrase protein or other cellular factors. Integration site bias has also been seen in yeast, where the Ty retrotransposon family integrates preferentially either upstream of tRNA genes or into heterochromatic regions of the genome (2, 3). Furthermore, in vertebrates it has been suggested that retroviral integrations may occur preferentially into regions of the genome with open chromatin conformation and/or undergoing active transcription (4-6).

Evidence that chromatin conformation may influence proviral DNA insertion has come from studies of insertions in murine and avian cells (4, 5). These studies have examined the proximity of apparently unselected insertions to DNase I hypersensitive sites, which are believed to correspond to genomic region that are more accessible to exogenous factors than tightly-packed chromatin. Similar studies have also suggested that retroviruses may integrate preferentially into transcribed and CpG-rich regions of the genome which are thought to correspond to the 5′ ends of genes (6). While these studies are suggestive they are not compelling. They have each considered a very small number of integration sites and some have acknowledged the possibility that the insertions may not have been entirely randomly selected (4, 5).

Perhaps the most convincing study to date indicating that retroviruses may have preferred integration targets has been that of Shih *et al.* examining the integration of avian leukosis virus (ALV) (7). This work suggested that about 20% of proviral insertions occur into 500-1,000 "hot spots" throughout the genome. However, later work by the same group has produced results which do not corroborate the notion that such hot spots exist (8). While this group has chosen to favor the more recent work a satisfactory explanation for the discrepancy has not yet been provided and the issue remains unresolved. Somewhat consistent with the

work in mice mentioned above, however, this work has suggested that while all tested regions of the genome were accessible targets for proviral integration, within those regions insertion site selection might be determined by local structural features of the target DNA.

The location of mutagenic proviruses in mouse insertional mutants suggests that some aspect of this process may be biased. Mutagenic insertions have often been found inserted into the 5' end of the disrupted genes (9-11). This observation is consistent with the notion that insertions occur preferentially into the CpG- and DNase I hypersensitive site-rich 5′ end of genes. However, it is also possible that insertions into these regions are more likely to perturb gene function and therefore might be over-represented among insertions selected as mutagenic.

Theoretically, it should possible to assess the randomness of proviral insertions by considering (a) the size of the mouse genome, (b) the mutagenic frequency of proviral insertions, (c) the number of essential genes in mice, and (d) the average insertion-sensitive target size for an essential gene. However, because estimates of the number of essential genes in mice vary from 5,000 to 25,000 (12) and because the average insertion-sensitive target size can only be guessed, such a calculation is not a worthwhile endeavor.

The relevance of the above mentioned studies in murine and avian cells to the randomness of retroviral insertional mutagenesis in zebrafish is not clear. Those studies do not sufficiently address the extent to which all genes in a target genome may be disrupted by proviral insertion. It remains possible that most or all genes are in fact targets, but that within a given gene, insertions are likely to occur into the 5′ end. Furthermore, these studies have considered the integration of retroviruses in the context of their natural hosts. In contrast, the method presented here uses components of a mouse retrovirus to generate proviral insertions in fish cells. It is possible that proviral integration biases which may exist in mice and birds may be

the result of recognition events which are specific to these organisms and their natural retroviruses, but which are not conserved in zebrafish. For example, pre-integration core particles might interact with cellular proteins which direct them to certain genomic sites. Recent evidence which supports this notion has identified a putative interaction between the integrase of the human immuno-deficiency virus (HIV) and the protein *ini-1* which is related to the yeast transcription factor SNF5 (13). In addition, studies creating fusions of the HIV integrase to the λ repressor protein have shown that this protein can direct integration events to the λ operator DNA sequence (14, 15).

Ultimately, the ability of murine-type proviral insertions to mutate most or all zebrafish genes will need to be determined by isolating large numbers of insertional mutations. The statistically-premature isolation of multiple alleles of given genes would begin to suggest that integration biases or hot spots exist. Even if such biases do exist, as is the case for P elements in flies, unless these biases are extreme this method will still facilitate the mutagenesis and cloning of a significant fraction of the essential genes in zebrafish.

One indication of the randomness of retrovirally-induced mutations in zebrafish will be the extent to which the distribution of mutant phenotypes obtained using this method will parallel those obtained by chemical mutagenesis. Although only six insertional mutations have been identified to date (Table 1), the phenotypes of these mutations correspond to the broad phenotypic classes obtained during chemical screens. For example, while about 15% of the chemical mutations have CNS degeneration phenotypes (16), 1 of the 6 insertional mutations exhibited CNS degeneration (*dead eye*). Similarly, while roughly 30% of the chemical mutations have "specific" phenotypes (16), 2 of the 6 insertional mutations exhibit specific defects (those caused by insertions 399 and D1). While these numbers are too low to

Table 1: Summary of zebrafish insertional mutants isolated as of 12/96.

Insertion ID	Mutant name	Disrupted gene	Mutant Phenotype	Ref.
38M	*no arches* (*nar*)	Homology to *Drosophila clipper*, a zinc finger ribonuclease, and to human, *C. elegans* and yeast genes.	Recessive embryonic lethal. Pharyngeal arch and fin defects. Arches largely or completely absent.	A
67D	*pescadillo* (*pes*)	Homology to uncharacterized genes in human, mouse and yeast.	Recessive embryonic lethal. Numerous organs fail to expand (i.e. liver, gut). Retarded development.	A, B
80A	-	unidentified	Recessive embryonic lethal. Small eyes and head. Severe edema.	A
404	*dead eye* (*dye*)	Homology to human and frog genes of unknown function, and similarity to yeast NIC96 protein.	Recessive embryonic lethal. Extensive cell death in the tectum and neural tube. Small head and eyes.	B
399	-	unidentified	Recessive larval lethal. Small eyes. Photoreceptor cell layer missing or reduced.	C
D1	-	unidentified	Dominant viable. Heterozygotes have breaks in adult body stripe pattern.	C

A, Gaiano N *et al.* (27)
B, Allende M *et al.* (28)
C, Hopkins N *et al.*, unpublished results.

draw significant conclusions from, they provide a promising start which suggests that a wide range of phenotypes will be attainable by insertional mutagenesis.

The Zebrafish Genome Project

As discussed in earlier chapters, the primary benefit of performing insertional mutagenesis instead of, or in addition to, chemical mutagenesis is that the mutated genes can be rapidly cloned. However, if it were possible to rapidly clone chemically-mutated alleles through the use of positional cloning, the candidate gene approach, and/or using rescue assays, the value of insertional mutagenesis would be in question. In recent years a zebrafish genome project has created genetic linkage maps including hundreds of markers as well as cloned zebrafish genes and expressed sequence tags (ESTs) (17, 18). However, several issues still represent significant obstacles to the cloning of chemical alleles including (a) the size of the zebrafish genome (1.6×10^9 bp), (b) the fact that large regions of the genome still have very few markers, and (c) the lack of a physical map.

In spite of these potential limitations, several aspects of the zebrafish system have generated optimism that the cloning of chemically-mutated genes will proceed rapidly. Among these are the relative ease with which large numbers of embryos can be genotyped to genetically map a mutation of interest in fine detail. The more recombination events examined, the more precisely the genetic location of a mutation can be determined. In addition, the isolation of expressed sequences close to a mutation, in conjunction with the ability to perform large numbers of simultaneous *in situ* hybridizations in zebrafish, may facilitate the identification of mutated genes by identifying transcripts which are expressed in the affected tissues. However, this approach requires the efficient isolation of most if not all expressed sequences in the region, which is a formidable and uncertain task. Furthermore, if

the mutated gene is not expressed in a pattern that obviously correlates with the timing and location of the mutant phenotype such a gene could be easily overlooked.

Another method which is being examined to clone chemically-mutated zebrafish genes is the use of microinjection rescue assays. In *C. elegans*, for example, the injection into mutant animals of a functional copy of the mutated gene can often rescue the mutant phenotype (19, 20). In fact the screening of yeast artificial chromosomes (YACs) or cosmids for rescue activity is a widely used method for narrowing the region of the *C. elegans* genome which may contain a mutated gene of interest.

The utility of mutant rescue methods in zebrafish remains to be determined. Current limitations include the fact that unlike in *C. elegans*, where much of the physical map is complete allowing the immediate isolation of YAC and cosmid clones spanning the region in question, in zebrafish time-consuming chromosomal walks are required to isolate such sequences. Furthermore, DNA microinjected into zebrafish embryos is often distributed unevenly in the embryo (21), and the rescue of only limited numbers of cells may be difficult to detect. Recent efforts to improve the distribution of microinjected DNA may help to alleviate this problem, and include the injection into embryos prior to activation (S. Burgess, personal communication), and the co-injection of a nuclear localization signal (NLS) peptide with the DNA to be tested (P. Alestrom, personal communication). However, the general utility of these methods is uncertain.

The speed with which mutated genes can be identified by positional cloning methods in zebrafish remains to be determined. Most likely, some genes will be cloned within the next few years using either the candidate gene approach or the identification of novel genes in the region of interest. In fact, the disrupted genes responsible for the zebrafish *no tail* and *floating head* mutant phenotypes have already been cloned using these methods (22, 23). However, based upon the difficulty

often encountered during the cloning of chemically-mutated genes in animals with well-characterized genomes, such as those of flies and mice, it seems improbable that the cloning of a given gene of interest will reliably proceed in a short and predictable time frame.

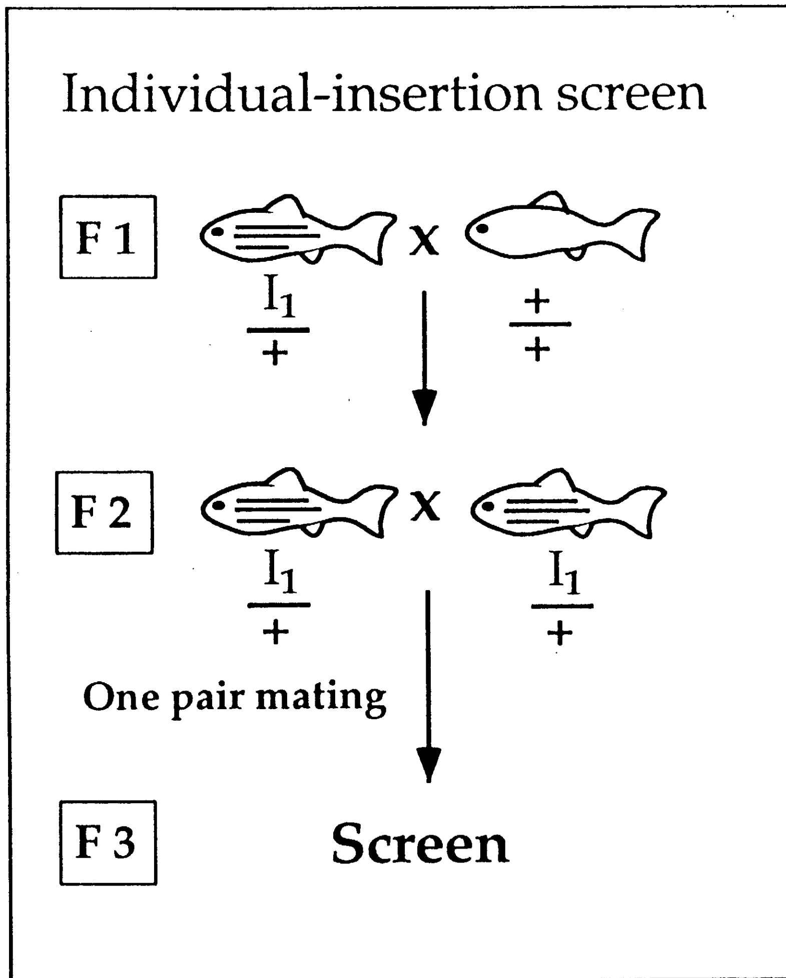
Based upon these considerations, insertional mutagenesis in zebrafish is likely to provide a valuable source of mutations in the near future for which the disrupted gene is easily clonable. As it is thought that roughly 40% of the genes essential for embryogenesis in zebrafish were mutated during the chemical screens (24), about half of the insertional mutations isolated are likely to be in genes which were not mutated in these screens. Furthermore, it is likely that a significant number of the insertional mutations will be allelic to already identified chemical mutations. Assuming that retrovirally-induced mutant phenotypes occur in a distribution similar to those induced by ENU, for every 100 insertional mutations isolated, 30 will have specific defects and 10-15 of those will be allelic to chemical mutations, thereby facilitating their molecular cloning.

Insertional Screening Strategies in Zebrafish

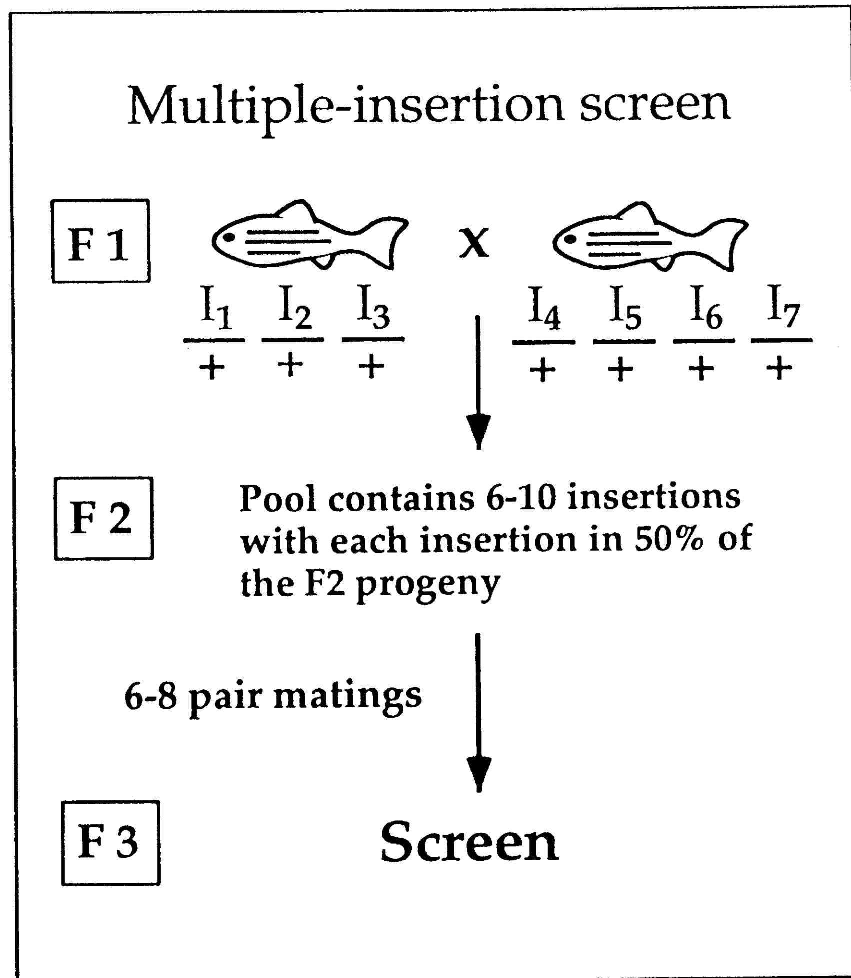
The ability to efficiently generate many thousands of proviral insertions in zebrafish has made large-scale insertional mutagenesis feasible in this system for the first time (25). While previously the ability to generate insertions limited the potential use of such a method, currently the ability to screen insertions for mutant phenotypes is limiting. This is because the identification and maintenance of thousands of insertions consumes a great deal of both time and space. Furthermore, the subsequent tracking of individual insertions by Southern blot and PCR can be quite laborious. Based upon the single-insertion inbreeding strategy used in the pilot screen described here (Fig. 1A), we estimate that a lab of 10-12 people could screen about 10,000 insertions in 2-3 years, isolating roughly 150 mutations. Such a screen

Figure 1: Strategies for conducting insertional mutagenesis screens in zebrafish

A



B



would only include 5-10% of the insertions that could be easily generated by a lab of that size, and while 150 mutations would certainly be of value, screening strategies which could isolate mutations more efficiently, and thereby in larger numbers, would be preferred.

An alternative strategy, based upon the use of transgenic F1 fish harboring 3-5 insertions, has recently been devised (Fig. 1B). Although the identification of large number of such F1's is labor intensive, the subsequent maintenance and inbreeding of insertions by this method would be greatly simplified. F1 fish with 3-5 insertions can be crossed to generate F2 pools which contain 6-8 insertions, each of which is present in 50% of the F2 fish. Eight successful crosses within each F2 pool would be statistically likely to inbreed most (90%) of the 6-10 insertions allowing the identification of mutant phenotypes. This breeding scheme is similar in nature to that employed by the chemical mutagenesis screens. However, while those screens identified, on average, 1 mutation in each F2 pool, the mutagenic frequency we have observed to date (1 mutagenic insertion in 70) suggests that an insertional screen of this sort would identify 1 mutation among about 15 F2 pools (if only 5 insertions are actually homozygosed in each pool). Using this multiple-insertion screening method we estimate that a lab of 10-12 people could screen roughly 25,000 insertions in 2-3 years, isolating about 350 insertional mutations. While this represents a 2- to 3-fold increase in the number of mutations isolated, as compared to the single-insertion method, there is one noteworthy drawback: the presence of 6-8 insertions among the F2 pool will complicate the cloning of the DNA flanking the mutagenic insertion and will delay both linkage analysis and the cloning of mutated gene.

The accuracy of the estimates we have projected regarding the utility of insertional screening can only be determined by undertaking such screens. As several labs in the zebrafish community are currently planning to do so, both the actual feasibility and productivity of these methods should be forthcoming. It

remains possible that these efforts will demonstrate that insertional screening of this sort is not effective enough to warrant continuation. Whether or not this is found to be the case, other methods which might facilitate the screening of large numbers of insertions include the screening of haploid or gynogenetic embryos, both of which can be generated using commonly practiced methods (26), or the use of retroviral gene traps (see below), a method which awaits the development of such vectors for use in zebrafish.

Potential Improvements in Methodology

There are numerous technical improvements which could dramatically enhance the utility of retroviral insertional mutagenesis screening in zebrafish. These include increasing the mutagenic frequency, increasing the germ-line transmission frequency of insertions to the F1 generation, the development of a retrovirus which can express a dominant visible reporter, and/or the development of a retroviral gene trap for use in zebrafish.

The mutagenic frequency observed with the SFG virus described in Chapters 3-5 is roughly 1 mutagenic insertion in 70 (27, 28), which is 3.5-fold lower than that seen in mice. This reduced efficiency has several plausible explanations. First, as suggested above, it is possible that proviral insertions in mice integrate preferentially into genes, a circumstance which would increase the mutagenic frequency in this organism. Such a bias may not exist when using these vectors in zebrafish. It is also possible that signals present on MLV-based genomes, such as the promoter-enhancer, the splice acceptor, and the polyadenylation signal may be more likely to perturb gene function in mouse cells than in fish cells.

Another explanation is that mice and zebrafish require different numbers of genes for embryonic development. Based upon the recent chemical screens in zebrafish it is estimated that there are roughly 2,500 genes in zebrafish which are

essential for embryogenesis (24). In mice, chemical mutagenesis of the T locus it has been used to estimate that there are 5,000-10,000 genes essential for embryogenesis (29). While such a difference could account for the difference in mutagenic frequencies in these organisms, the fact that the zebrafish genome is roughly half the size of the mouse genome suggests that unless zebrafish genes are half the size of mouse genes, in theory, only half as many essential fish genes would be required to obtain a mutagenic frequency similar to that seen in the mouse. It is in fact possible that the genomic size of the average zebrafish gene is smaller than that of the average mouse gene, and therefore might constitute a smaller target, although this has yet to be demonstrated. It is also possible that the genetic control of zebrafish development is more redundant than that of mice such that fewer essential functions can be disrupted by single gene mutations. Furthermore, the estimated number of essential genes in both the mouse and the zebrafish may be inaccurate.

Of these issues raised above the most easily improved upon is the possibility that the SFG proviruses are not as mutagenic as other viruses might be. The inclusion of both splice acceptors and polyadenylation signals which have been shown to be strong signals in zebrafish cells could increase the likelihood that proviral sequences either splice into endogenous transcripts or terminate transcription prematurely. Further study will be required to assess the extent to which a more mutagenic virus can be constructed and produced at the titers required for generating large numbers of insertions. The other issues raised above address characteristics inherent to the use of murine vectors in the zebrafish and would be difficult to improve upon.

In addition to increasing the mutagenic frequency, an increase in the germ-line transmission frequency of insertions to the F1 generation would facilitate the multiple-insertion screen strategy outlined in Fig. 1B. To date, increases in this frequency have coincided with an increase in the number of F1 fish with multiple

insertions. Based upon the most recent germ-line transmission data (roughly 50% better than that described in Chapter 3, A. Amsterdam and N. Hopkins, unpublished data) 4 founders are needed to generate 3 F1 fish harboring 10 insertions. This is because at the current frequency, this strategy calls for the identification of founders which transmit insertions to more than half of their offspring (which about 50% of the founders do) and then those founders are mated to each other to enhance the number of F1's harboring multiple insertions. If instead, all founders produced several F1 fish harboring 3-4 distinct insertions, then only 25% as many founders would need to be generated and they would not need to be tested for those with the highest germ-line transmission frequencies. The injection of higher titer virus stocks earlier in development might make such an increase possible.

Another technical improvement which could significantly increase the efficiency of a retroviral insertional mutagenesis screen would be a retroviral vector which could express a dominant visible reporter after germ-line transmission. The ideal reporter would be the green fluorescent protein (GFP) which could allow the identification of live transgenic embryos simply by examining embryo pools under a fluorescence microscope (30). This would circumvent the need to identify transgenic fish by tail-cutting and PCR and would save considerable time and space. *LacZ* might also be a useful reporter and could be detected in live embryos, as has been shown with plasmid transgenes in zebrafish (31), using the substrate FDG which fluoresces upon cleavage by *β-gal*.

Efforts to express reporter genes from retroviral vectors after passage through the zebrafish germ line have not been successful although such studies have been very limited to date. For example, of about 15 proviral transgenes generated using the NK(G) virus (see Chapter 3) which have been tested, none have been found which express *lacZ* at levels detectable by the chromogenic substrate X-gal (M. Allende and N. Hopkins, unpublished data). Curiously, at least some NK(G)

proviral insertions can express *lacZ* at levels detectable by X-gal in injected embryos. Further analysis of these transgenes after germ-line transmission is necessary to determine whether there is no expression at all or if expression is at levels too low to be detected by X-gal staining.

The development of a retroviral 'expression vector' for use in zebrafish would be valuable not only for insertional mutagenesis but also for studies designed to examine the role of bioactive genes in this vertebrate. However, considerable efforts to develop such vectors in mice have proven disappointing and raise concerns about the likelihood of developing expression vectors for use in zebrafish in the near future. Nonetheless, based upon the limited amount of relevant data currently available in zebrafish and the fact that retroviral vectors are expressed after passage through the germ line in chickens (32), it is premature to assume that the development of this technology in zebrafish will pose a significant problem.

Finally, the most exciting potential technical improvement which would enhance the utility of insertional mutagenesis in zebrafish would be the development of a retroviral gene trap. As described in Chapter 1, most gene traps are constructs which contain a reporter gene without a promoter (33, 34). These constructs are expressed only when they integrate in the correct orientation downstream of the promoter of an endogenous gene. Because the reporter gene is then driven by that cellular promoter, it is usually expressed in a spatial and temporal pattern similar to that of the endogenous gene. Furthermore, gene trap insertions which have their reporter activated generally perturb the expression of the endogenous gene into which they have integrated. When such a gene is essential for a given process its disruption may result in a scorable mutant phenotype.

Gene trap screening would provide two important advantages to retroviral insertional mutagenesis in zebrafish. First, since activated gene traps have generally inserted into genes, the identification of such integration events prior to screening

could dramatically reduce the number of insertions which would need to be screened to isolate a mutant phenotype. In mice, for example, activated gene trap insertions are 5- to 10-fold more likely to have disrupted an essential gene than unselected proviral insertions (cf. 33 & 35). Second, the pattern of expression of activated gene trap reporters would allow insertions to be screened in the heterozygous state for those which have integrated into genes expressed at a given time in a given tissue of interest.

Fundamental to the design of a retroviral gene trap for use in zebrafish is the inclusion of a reporter gene which can be easily detected even when expressed at relatively low levels. This is important to assure that cellular promoters which drive both strong and weak gene expression can produce detectable reporter levels. In addition, as is the case with all new vectors worthy of consideration for use in this context, to be useful for large-scale screening a gene trap retrovirus which can be prepared to high titers is desirable.

Efforts to create gene trap retroviruses usable for insertional mutagenesis screening in zebrafish have been unsuccessful so far. Numerous traps which can be activated in tissue culture have been developed, including some identical or similar to those used in mice (N. Gaiano and N. Hopkins, unpublished data). However, of the numerous traps with *β-gal* or *β-geo* as their reporters, none have produced titers high enough for use in embryos. Several GFP gene traps have also been created, although to date none of these have produced detectable GFP levels when tested in tissue culture. As 'brighter' GFP mutants are generated they are being tested for use in gene traps. Alternatives to these reporters include screening for the expression of an activated RNA molecule by *in situ* hybridization, or of a protein by antibody staining. These methods are worthy of consideration although they include more costly and time-consuming detection methods.

If and when a workable retroviral gene trap is developed for use in zebrafish it is likely to revolutionize forward mutagenesis screening in vertebrates. This is because, while tens of thousands of insertions can be screened by the methods described above, hundreds of thousands of insertions could be screened relatively simply using a gene trap. As mentioned in earlier chapters, 10,000-20,000 founder fish would transmit as many as 200,000 insertions to their F1 progeny (25). The founders could be maintained in a relatively small space and the F1 progeny screened directly for activated trap patterns. In theory, about 10,000 crosses would be necessary to carry out such a screen and these crosses could be performed relatively easily within a few months. Because only fish harboring insertions with patterns of interest to those screening would be raised to adulthood, the raising and maintaining of large numbers of fish would not be required. Such a method would be accessible to both large and small labs, and could be used to identify genes with very specific roles during a given aspect of development even if such genes are few in number.

Insertional Mutagenesis Beyond Developmental Biology

While this thesis has focused on the application of insertional mutagenesis for the study of vertebrate developmental genetics, the use of this method has broader potential value. Currently, the genome projects in numerous organisms, including mice and humans, are attempting not only to map and sequence these genomes but to identify and map all expressed sequences. In animals, these latter efforts have focused on the isolation of expressed sequence tags (ESTs) which can then be placed on the genomic map in question. The genes represented by these ESTs can then be studied if the sequence in question appears to encode a protein of interest, or can simply await the identification of genetic functions which map near them and which they may encode.

The abundance of gene sequences which are being generated, not only through EST isolation but by other methods such as sequencing of the yeast and C. *elegans* genomes, for example, has created a need for methods to assign functions to these genes. In yeast, a community-wide effort is currently engaged in the systematic disruption of the roughly 6,000 open reading frames identified in this eukaryote (36). While such an effort may seem extreme, upon its completion all genes with a function which is unique and essential to the growth of this eukaryote will be identified. These genes will represent a valuable resource in the pursuit of a comprehensive understanding of eukaryotic cell biology. While it is true that such efforts will not readily identify functions which are encoded by more than one gene, this limitation is inherent to most genetic screens and does not detract from the value of those functions which are uncovered.

The ease with which targeted gene-disruption can be achieved in yeast has made this approach possible. While, in theory, the same approach could be undertaken in mice using homologous recombination-mediated gene targeting, the cost of such an effort is prohibitive. Forward mutagenesis screens in animal systems including zebrafish, have disrupted many essential gene functions, although for various reasons these efforts have focused on the isolation of a subset of the induced mutations. These reasons include the particular interests of the investigators, and the fact that the maintenance of large numbers of mutations can be cumbersome. Furthermore, during the course of chemical mutagenesis screens, such as those performed in fruit flies and in zebrafish for example, many mutant phenotypes fall into a few broad classes which exhibit seemingly nonspecific general defects. Since such phenotypes are likely to be caused by a wide range of cellular defects, many of which may be of little interest to those screening, and because the cloning of chemical mutated genes is a large undertaking worthy of only the most 'interesting' mutants, nonspecific mutations are usually discarded. Consequently, genes with

essential roles in animal development and survival are regularly being identified by mutation and then discarded without the acquisition of any information regarding their molecular identity or function.

The identification of all essential genes in a vertebrate animal could dramatically advance our understanding of the basic principles governing vertebrate biology. The so called 'nonspecific' developmental phenotypes referred to above may be caused by mutations in genes central to basic cellular processes which are either currently poorly understood, or which may have yet to be discovered. The widespread application of retroviral insertional mutagenesis in zebrafish, perhaps as a community-wide collaborative effort, could lead to the mutagenesis and rapid cloning of most if not all essential genes in this vertebrate. While the relationship of essential genes in zebrafish to those in other vertebrates, including humans, has yet to be determined, the high degree of conservation in the biological processes of animals suggests that these genes would be of potential relevance to an understanding of vertebrate biology in general and perhaps human disease as well.

Identifying Essential Genes is Only the Beginning

The method described in this thesis is designed to permit the rapid mutagenesis and cloning of hundreds of genes with essential roles in vertebrate developmental and cell biological processes. However, the identification of these genes is only the first step toward an understanding of the processes in which they play a part. A more thorough understanding will require the subsequent analysis of the RNA and protein products encoded by such genes and the activity of these molecules during the processes in question. Furthermore, the identification of regulatory and/or functional interactions among genes is essential to a complete understanding of their functions.

Several lines of research could be useful to characterize the role of newly identified gene products. The identification of genes which are homologous to those characterized in other systems may immediately suggest functions for the 'new' gene products. Similarly, the presence of protein domains with homology to previously characterized motifs would suggest certain functions. For example, as indicated above, the fact that the Nar protein is highly homologous to the fly protein Clp suggests that Nar is an endoribonuclease. Further studies of this protein should verify that the fish protein has such an activity and then attempt to identify the endogenous targets of this activity.

An early step toward the characterization of the function of either novel genes or genes with homology to those previously identified would be to determine where in the organism and in the cell such proteins are localized. Analysis of the protein sequence might provide clues as to whether a protein is likely to be, for example, transmembrane, nuclear, or secreted. Experimentally this issue would be addressed by *in situ* hybridization and/or northern blot analysis to determine where RNAs are localized, and antibody stainings to determine where the proteins are located both on the organismal and subcellular levels. It will also be important to determine if such protein products are modified post-translationally by the addition of phosphate or carbohydrate groups, or by proteolytic cleavages which might be required for the transition from an inactive to an active form or vice versa.

The characterization of interactions between genes affecting a process of interest would be of great value for constructing a comprehensive picture of that process. Such interactions might be suggested by genetic studies such as epistasis analyses among mutations isolated either by insertional, chemical or radiation mutagenesis. Biochemical studies as well as studies utilizing the yeast two-hybrid system could also be performed to identify protein-protein or protein-nucleic acid interactions. Furthermore, for putative transcriptional regulators, downstream

targets might be identified by ectopic expression studies which either look for the upregulation of postulated target genes, or use the overexpression of the potential regulators to enrich for expression from unknown targets which might then be identified by differential screening methods.

Ultimately, the central goal of the method developed in this thesis, as well as the subsequent studies analyzing the function of the identified genes, is to contribute to a comprehensive understanding of vertebrate biology on the molecular, cellular, and organismal levels. This knowledge in conjunction with that obtained from studies performed in a wide variety of other organisms will shed light on the remarkable nature of life on earth and may provide valuable insight into its origins and evolution. In addition, practically speaking such knowledge should continue to facilitate an understanding of the causes of human disease and should permit the design of both preventative and therapeutic treatments for such conditions.

REFERENCES

1. Spradling AC; Stern DM; Kiss I; Roote J; Laverty T; Rubin GM (1995) Gene disruptions using P transposable elements: an integral component of the *Drosophila* genome project. *Proc Natl Acad Sci USA* **92**: 10824-30.

2. Devine SE; Boeke JD (1996) Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. *Genes Dev* **10**: 620-33.

3. Zou S; Ke N; Kim JM; Voytas DF (1996) The Saccharomyces retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci. *Genes Dev* **10**: 634-45.

4. Vijaya S; Steffen D; Robinson H (1986) Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. *J Virol* **60**: 683-92.

5. Rohdewohld H; Weiher H; Reik W; Jaenisch R; Breindl M (1987) Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J Virol* **61**: 336-43.

6. Scherdin U; Rhodes K; Breindl M (1990) Transcriptionally active genome regions are preferred targets for retrovirus integration. *J Virol* **64**: 907-12.

7. Shih C-C; Stoye JP; Coffin JM (1988) Highly preferred targets for retrovirus integration. *Cell* **53**: 531-7.

8. Withers-Ward ES; Kitamura Y; Barnes JP; Coffin JM (1994) Distribution of targets for avian retrovirus DNA integration in vivo. *Genes Dev* **8**: 1473-87.

9. Harbers K; Kuehn M; Delius H; Jaenisch R (1984) Insertion of retrovirus into the first intron of a1(I) collagen gene leads to embryonic lethal mutation in mice. *Proc Natl Acas Sci USA* **81**: 1504-8.

10. Soriano P; Gridley T; Jaenisch R (1987) Retroviruses and insertional mutagenesis in mice: proviral integration at the Mov 34 locus leads to early embryonic death. *Genes Dev* **1**: 366-75.

11. Harbers K; Muller U; Grams A; Li E; Jaenisch R; Franz T (1996) Provirus integration into a gene encoding a ubiquitin-conjugating enzyme results in a placental defect and embryonic lethality. *Proc Natl Acad Sci USA* **93**: 12412-7.

12. Milkos GLG; Rubin GM (1996) The role of the genome project in determining gene function: insights from model organisms. *Cell* **86**: 521-9.

13. Kalpana GV; Marmon S; Wang W; Crabtree GR; Goff SP (1994) Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science* **266**: 2002-6.

14. Bushman FD (1994) Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. *Proc Natl Acad Sci USA* **91**: 9233-7.

15. Bushman FD (1995) Targeting retroviral integration. *Science* **267**: 1443-4.

16. Mullins M; Hammerschmidt M; Haffter P; Nüsslein-Volhard C (1994) Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. *Curr Biol* **4**: 189-202.

17. Postlethwait JH; Johnson SL; Midson CN; Talbot WS; Gates M; Ballinger EW; Africa D; Andrews R; Carl T; Eisen JS; Horne S; Kimmel CB; Hutchinson M; Johnson M; Rodriguez A (1994) A genetic linkage map for the zebrafish. *Science* **264**: 699-703.

18. Knapik EW; Goodman A; Atkinson OS; Roberts CT; Shiozawa M; Sim CU; Weksler-Zangen S; Trolliet MR; Futrell C; Innes BA; Koike G; McLaughlin MG; Pierre L; Simon JS; Vilallonga E; Roy M; Chiang P-W; Fishman MC; Driever W; Jacob HJ (1996) A reference cross DNA panel for zebrafish (*Danio rerio*) anchored with simple sequence length polymorphisms. *Development* **123**: 451-460.

19. Clark DV; Baillie DL (1992) Genetic analysis and complementation by germ-line transformation of lethal mutations in the unc-22 IV region of Caenorhabditis elegans. *Mol Gen Genet* **232**: 97-105.

20. Lee RC; Feinbaum RL; Ambros V (1993) The C. *elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**: 843-54.

21. Westerfield M; Wegner J; Jegalian BG; DeRobertis EM; Puschel AW (1992) Specific activation of mammalian Hox promoters in mosaic transgenic zebrafish. *Genes Dev* **6**: 591-8.

22. Schulte-Merker S; van Eeden FJ; Halpern ME; Kimmel CB; Nusslein-Volhard C (1994) no tail (ntl) is the zebrafish homologue of the mouse T (Brachyury) gene. *Development* **120**: 1009-15.

23. Talbot WS; Trevarrow B; Halpern ME; Melby AE; Farr G; Postlethwait JH; Jowett T; Kimmel CB; Kimelman D (1995) A homeobox gene essential for zebrafish notochord development. *Nature* **378**: 150-7.

24. Haffter P; Granato M; Brand M; Mullins M; Hammerschmidt M; Kane D; Odenthal J; van Eeden F; Jiang Y; Heisenberg C-P; Kelsh R; Furutani-Seiki M; Vogelsang L; Beuchle D; Schach U; Fabian C; Nüsslein-Volhard C (1996) The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* **123**: 1-36.

25. Gaiano N; Allende M; Amsterdam A; Kawakami K; Hopkins N (1996) Highly efficient germ-line transmission of proviral insertions in zebrafish. *Proc Natl Acad Sci USA* **93**: 7777-82.

26. Westerfield M (1995) ed*The Zebrafish Book* (Univ. of Oregon Press, Eugene).

27. Gaiano N; Amsterdam A; Kawakami K; Allende M; Becker T; Hopkins N (1996) Insertional mutagenesis and rapid cloning of essential genes in zebrafish. *Nature* **383**: 829-32.

28. Allende ML; Amsterdam A; Becker T; Kawakami K; Gaiano N; Hopkins N (1996) Insertional mutagenesis in zebrafish identifies two gene, pescadillo and dead eye, essential for embryonic development. *Genes Dev* **10**: 3141-55.

29. Shedlovsky A; Guenet J-L; Johnson LL; Dove WF (1986) Induction of recessive lethal mutations in the T/t-H-2 region of the mouse genome by a point mutagen. *Genet Res* **47**: 135-42.

30. Amsterdam A; Lin S; Hopkins N (1995) The Aequorea victoria green fluorescent protein can be used as a reporter in live zebrafish embryos. *Dev Biol* **171**: 123-9.

31. Lin S; Yang S; Hopkins N (1994) lacZ expression in germline transgenic zebrafish can be detected in living embryos. *Dev Biol* **161**: 77-83.

32. Briskin MJ; Hsu RY; Boggs T; Schultz JA; Rishell W; Bosselman RA (1991) Heritable retroviral transgenes are highly expressed in chickens. *Proc Natl Acad Sci USA* **88**: 1736-40.

33. Friedrich G; Soriano P (1991) Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev* **5**: 1513-23.

34. von Melchner H; DeGregori JV; Rayburn H; Reddy S; Friedel C; Ruley HE (1992) Selective disruption of genes expressed in totipotent stem cells. *Genes Dev* **6**: 919-27.

35. Jaenisch R (1988) Transgenic animals. *Science* **240**: 1468-75.

36. Goffeau A; Barrell BG; Bussey H; Davis RW; Dujon B; Feldmann H; Galibert F; Hoheisel JD; Jacq C; Johnston M; Louis EJ; Mewes HW; Murakami Y; Philippsen P; Tettelin H; Oliver SG (1996) Life with 6000 genes. *Science* **274**: 546, 563-7.