Zebrafish genomics: from mutants to genes

JOHN H. POSTLETHWAIT (jpostle@oregon.uoregon.edu)

WILLIAM S. TALBOT (talbot@saturn.med.nyu.edu)

Exquisite embryonic lethal mutations have been isolated in hundreds of genes necessary for zebrafish development. Analysis of this resource promises to enhance our understanding of the molecular genetic mechanisms of vertebrate development. This review discusses the state of the zebrafish genome project and the genetic trickery that can expedite molecular isolation of genes disrupted by these mutations.

A mutant phenotype identifies the function of a disrupted gene. Applying this simple concept revolutionized our understanding of genetic control mechanisms in phage, bacteria and yeast, and illuminated cell interactions during development in fruitflies and nematodes. Phage geneticist George Streisinger, searching for a way to apply systematic mutagenesis protocols to vertebrates, introduced the zebrafish *Danio rerio* as a vertebrate that would allow the isolation of mutations in large-scale genetic screens[1]. Such screens have been remarkably successful, having now provided a collection of over 2000 mutations in several hundred genes that are necessary for the normal development of zebrafish embryos[2–17].

Because the fundamental molecular mechanisms of development are shared among vertebrates, the analysis of zebrafish mutations is likely to provide new information generally relevant to other classes of vertebrates. Evidence to support this assertion includes the fact that mutant phenotypes in zebrafish often mimic those in other vertebrates[18–21], an indication of shared gene functions. In addition, expression patterns of orthologous genes are often strikingly similar in zebrafish and tetrapods[22–24]. Finally, the single-cell level phenotypic and epistatic analyses possible in zebrafish[11,18,25–28] are likely to reveal shared gene functions that might be difficult to discover in other organisms.

For the study of zebrafish mutations to reveal new gene functions, investigators must characterize the molecular nature of the disrupted genes. There are three main approaches for the cloning of mutated genes: candidate gene testing, positional cloning and insertional mutagenesis. The candidate gene approach tests cloned genes that have properties expected of the mutated gene to see if any candidate co-segregates with the mutant phenotype and is structurally altered by the mutation. Positional cloning identifies a DNA sequence located near the mutation, and then sequentially isolates overlapping DNA fragments until the gene of interest is reached. Insertional mutagenesis involves the induction of mutations by the integration of a cloned DNA sequence and then identifies DNA flanking the insertion that should include the mutant gene. All three approaches benefit from methods to map loci quickly and to show that disruption of a particular gene causes the mutant phenotype. This review discusses the current state of zebrafish genetic mapping and the prospects for cloning genes identified by mutation.

The genetic map

Until three years ago, no two genes had been shown to be linked in zebrafish. The first genetic map ordered about 400 markers in 29 linkage groups[29]. This work exploited a haploid mapping panel consisting of 94 offspring from a single female who was heterozygous for the laboratory strain AB and the Darjeeling wild-type strain. Most of these markers were anonymous DNA sequences identified as random amplified polymorphic DNAs (RAPDs)[30,31]. Subsequent analysis brought the map to 652 markers assayed by the polymerase chain reaction (PCR) and localized centromeres, thereby consolidating the map to 25 linkage groups, the same as the number of chromosomes[32]. An additional 102 simple sequence length polymorphisms (SSLPs) have recently been scored on an F_2 diploid reference cross[33], and approximately 350 more SSLPs have been scored on a subset of the reference cross [E. Knapik *et al.*, pers. commun.; S. Neuhauss (1996) PhD dissertation, University of Tübingen]. SSLPs are useful because they are highly polymorphic, and co-dominant SSLP alleles simplify the analysis of diploid crosses. Over 100 SSLPs have been located on the RAPD map, thus integrating the two maps (W.S. Talbot *et al.*, unpublished). In addition to these anonymous markers, 11 mutations and about 120 cloned genes have been mapped (Refs 8, 23, 29, 34, 35; J.H. Postlethwait *et al.*, unpublished).

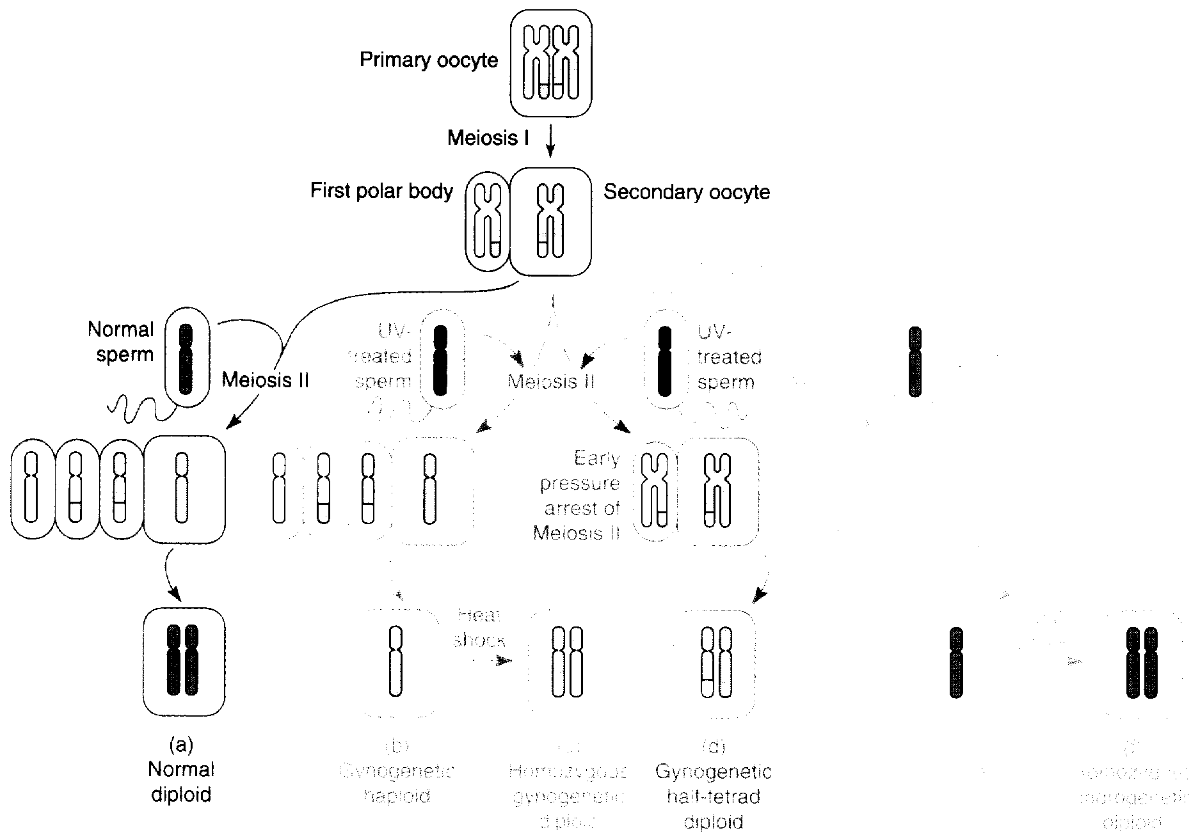
In total, the map positions of about 1200 markers are known (Refs 29, 32, 33; J.H. Postlethwait *et al.*, unpublished; E. Knapik *et al.*, pers. commun.). The linkage map is about 3000 cM long and the genome is about 1.7×10^9 bp, therefore the average intermarker interval is about 2.5 cM or 1500 kb. For comparison, the genetic linkage map of human has 5264 microsatellite markers with an average spacing of 700 kb or 0.7 cM (Ref. 36).

Thousands of cloned genes and other DNA sequences have been mapped in the human genome by screening somatic cell hybrids and radiation hybrids for the presence of the target sequence[37,38]. The preparation of a panel of somatic cell hybrids[39] and radiation hybrids (C. Kwok, pers. commun.) should provide useful alternatives to meiotic mapping of cloned genes. After the panels are genotyped for several hundred markers, cloned genes will be able to be mapped without identifying polymorphisms. These reagents will not aid the mapping of mutations, however, because these cell lines cannot be used for genetic complementation analysis with animals. In contrast, a panel of deletions and translocations can be used to map mutations to chromosome segments and to localize cloned genes without finding polymorphisms. Directed construction of such a panel is in progress by the screening of γ-ray mutagenized haploid genomes for the absence of DNA markers of known location[8].

The assignment of each linkage group to a cytogenetically identified chromosome is an important goal. The zebrafish karyotype consists of 25 rather similar, roughly metacentric chromosomes[40–42]. Recent results show that replication-banded metaphase spreads from the primary cultures of zebrafish embryos, and caudal fins of adults[43], are of sufficient quality and consistency to

Copyright © 1997 Elsevier Science Ltd. All rights reserved. 0168-9525/97/$17.00
PII: S0168-9525(97)01129-3

Box 1. Manipulating ploidy of the zebrafish genome



In vitro fertilization techniques can vary the ploidy and parental origin of genome sets in zebrafish, leading to six types of animals useful for genetic analysis. In a primary oocyte, crossing-over occurs in Meiosis I, leading to tetrads with recombinant chromatids. (a) Fertilization of secondary oocytes by normal sperm stimulates the completion of Meiosis II and formation of a normal diploid zygote. (b) Fertilizing secondary oocytes with sperm whose genome has been destroyed with ultraviolet (UV) light activates the egg, stimulates the completion of Meiosis II, and results in the development of a gynogenetic haploid zygote, which can develop into a haploid embryo. (c) Gynogenetic haploids are useful for first-generation mutant screens[2,11] and for the genetic mapping of recessive loci[29]. Haploid embryos develop remarkably normally, but have a specific developmental syndrome including a short body axis and a kinked neural tube; haploids die shortly after hatching, but long after many mutant phenotypes are apparent. Subjecting haploid zygotes to a two minute heat shock, 13 minutes after egg activation, suppresses the embryo's first mitotic cleavage division, leading to a gynogenetic diploid embryo. These animals are homozygous for maternal alleles at all loci. (d) Such embryos can give rise to clonal lines of fish. Whereas heat shock makes gynogenetic diploids that are homozygous for all loci, early pressure (EP) treatment produces gynogenetic half-tetrad diploid animals that are homozygous for only some loci. Freshly ovulated oocytes complete crossing-over and Meiosis I and then commence anaphase of Meiosis II about 90 seconds after fertilization. The application of high pressure at this time dissolves the spindle and suppresses Meiosis II, giving rise to animals that are homozygous for loci proximal to the first cross-over on a chromosome arm. A large number of half-tetrad animals develop normally, survive to sexual maturity and reproduce. Geneticists exploit these animals in first-generation mutant screens[10], in the mapping of loci to centromeres[32,34,46], in the analysis of chromosomal interference in recombination[46,47] and in the production of inbred lines[32]. In addition to these animals that have only maternal chromosomes, geneticists can produce animals that have only paternal chromosomes (e and f). γ-Ray irradiation of oviposited secondary oocytes destroys the maternal genome[60]. Fertilization of these γ-irradiated oocytes with normal sperm yields androgenetic haploid embryos (e). These embryos are useful for comparing rates of recombination in male and female meiosis. The application of heat shock to androgenetic haploids at the time of the first mitotic cleavage division results in homozygous androgenetic diploid embryos (f). Some of these animals survive to adulthood, and can be used to study the role of genetic imprinting in zebrafish. *The Zebrafish Book* details these and other methods[61]. (See Box 2 on page 189 for access to this information and other zebrafish resources over the World Wide Web.)

allow the identification of each individual chromosome and the generation of a standard karyotype (A. Amores, pers. commun.). These preparations should enable the cytogenetic mapping of cloned sequences by fluorescent *in situ* hybridization (FISH). An integrated cytogenetic and linkage map will be helpful for the cloning of zebrafish mutations.

Exploiting genetic tricks to map zebrafish mutations

A standard method to map a mutation in mouse is the genome scan, that is an analysis of 40 or more SSLPs spanning the mouse genome in 60–100 animals from an F_1 backcross or F_2 intercross[44]. This effort can require thousands of PCRs. Even more PCRs are required to utilize the genome scan strategy for mutant mapping in

zebrafish because: (1) the zebrafish linkage map is twice the length of the mouse map; (2) zebrafish have 25 chromosomes compared with 20 in mouse; and (3) the ends of zebrafish chromosomes have not yet been placed on the linkage map. Coupling PCR to the genetic tricks Streisinger *et al.*[1] developed for zebrafish (Box 1) has permitted the design of two efficient methods for mapping zebrafish genes with far fewer PCRs, bulked segregant analysis[29,45] and centromere-linkage analysis using half-tetrad animals[32,34].

Bulked segregant analysis (BSA) identifies dominant or co-dominant markers linked to mutations by the differential amplification of fragments from two genomic DNA pools, one from wild-type animals and the other from their mutant siblings (Fig. 1). Both alleles of markers unlinked or distantly linked to the mutation are equally represented in both phenotypic pools because of independent assortment and crossing-over. Hence, unlinked marker fragments are amplified from both pools. In contrast, fragments from markers linked to the mutation amplify in only one pool because these alleles co-segregate with the mutation on which the pools are based. Although the pools can consist of either haploid or diploid individuals, the use of haploids simplifies the analysis of dominant markers such as RAPDs because, in haploids, dominant alleles do not obscure their recessive counterparts as they do in diploid heterozygotes.

Centromere-linkage analysis (CLA) relies on diploid half-tetrad animals to assign mutations to linkage groups rapidly (Fig. 2)[32]. Half-tetrad analysis reveals the distance between genes and their centromeres[46,47] because homozygous mutant half-tetrads result almost exclusively from meioses with no cross-overs between the mutant locus and its centromere (Fig. 2, Step 1). Exceptions due to multiple cross-overs are rare because of high recombinational interference in zebrafish[46]. The identification of PCR-based markers near each of the 25 centromeres permits a mutation to be tested for linkage to each centromere[40,42]. This, in turn, allows the mutation to be assigned to one of the 25 linkage groups with a minimum number of PCR assays (Fig. 2, Step 2). CLA allows a mutation to be assigned to a linkage group by scoring 12 homozygous mutant half-tetrads for markers linked to each of the 25 centromeres (300 PCRs), assuming that stocks are inbred and polymorphic with respect to centromeric markers. If one uses polymorphic co-dominant markers and prepares pools of DNA from half-tetrad diploids (i.e. one pool from 12 wild-type animals and the other from 12 mutant animals) then 50 PCRs are sufficient to localize the mutation to a linkage group (25 centromere markers × two phenotypic pools). After the linkage group is determined, additional markers are used to localize it on a chromosome arm within a few centimorgans (Fig. 2, Step 3).

The genetic map and the methodologies of BSA and CLA can expedite the genomic localization of mutations, but how can mutations best be cloned? Three approaches have proven fruitful: insertional mutagenesis, candidate genes and positional cloning.

Insertional mutagenesis

Recent studies show that insertional mutagenesis with retroviral vectors is feasible in zebrafish[12,13]. Screens of F_3 progeny from F_2 animals bearing transgenes identified four lethal mutations linked to, and presumably induced by, retroviral insertions. In three of these cases, sequence analysis of flanking DNA rapidly identified the gene disrupted by the insertion, demonstrating an important advantage of insertional mutagenesis. One application of retroviral vectors is whole-genome mutation screens, but the inefficiency of insertional mutagenesis will make it difficult to isolate mutations in as many genes as identified in the chemical screens. Approximately 35 F_2 families, with two insertions each, must be screened to identify each embryonic lethal mutation because about 1 in 70 retroviral insertions disrupts an essential gene[12]. By contrast, mutagenesis screens employing the chemical mutagen ethylnitrosourea (ENU) identified a lethal mutation in almost all F_2 families[7,9].

Another application of retroviral mutagens is the isolation of tagged alleles of previously identified loci by non-complementation of existing mutations. It is likely that more than 70 000 insertions must be tested to find an insertional allele of a known locus because most loci identified in ENU mutagenesis screens were recovered at a rate of less than one allele in 1000 lethal mutations[7,9] and because about 1 in 70 retroviral insertions hits an essential gene. The germline of an insertionally mutagenized G0 individual transmits, on average, about 11 insertions[12,13]; therefore, on average, about 7000 pair matings between G0 individuals and heterozygous testers need to be screened to find an insertional allele at a previously identified locus. In addition, insertion alleles of some loci might be difficult or impossible to find because the vector might have a target-site bias as with *P* elements in *Drosophila*. In spite of these concerns, improvements in the breeding and screening protocols, coupled with more efficient vectors, perhaps including mobilizable transposons, will make insertional mutagenesis an attractive strategy for the cloning of mutations.

The candidate gene approach: exploiting mouse and human maps

The systematic testing of candidate genes has resulted in the molecular isolation of four zebrafish mutations: *no tail* (Ref. 22), *floating head* (*flh*) (Ref. 23), *no isthmus* (Ref. 24) and *nic1* (Ref. 48). In this approach, the map location of a mutation is compared with the map location of cloned genes that are expressed in or near tissues phenotypically altered by the mutation. This cloning approach benefits from rapid methodologies to map mutations and from a genetic map with many cloned genes to serve as candidates.

The future success of the candidate gene approach will be improved by increasing the density of cloned genes on the zebrafish map. The mapping of cloned genes in zebrafish began in 1994 (Ref. 29) with the localization of *mhcdab*, *snail1* and *msxb*. To map cloned genes, sequences expected to be polymorphic, such as introns or 3′ untranslated regions, are amplified by PCR. Polymorphisms are revealed either by a restriction enzyme, by size differences or by single-strand conformation polymorphism (SSCP) analysis[49]. The map currently contains more than 120 cloned genes, most of which are of interest to developmental biologists and, hence, provide a source of candidate genes for embryonic mutations (J.H. Postlethwait *et al.*, unpublished).

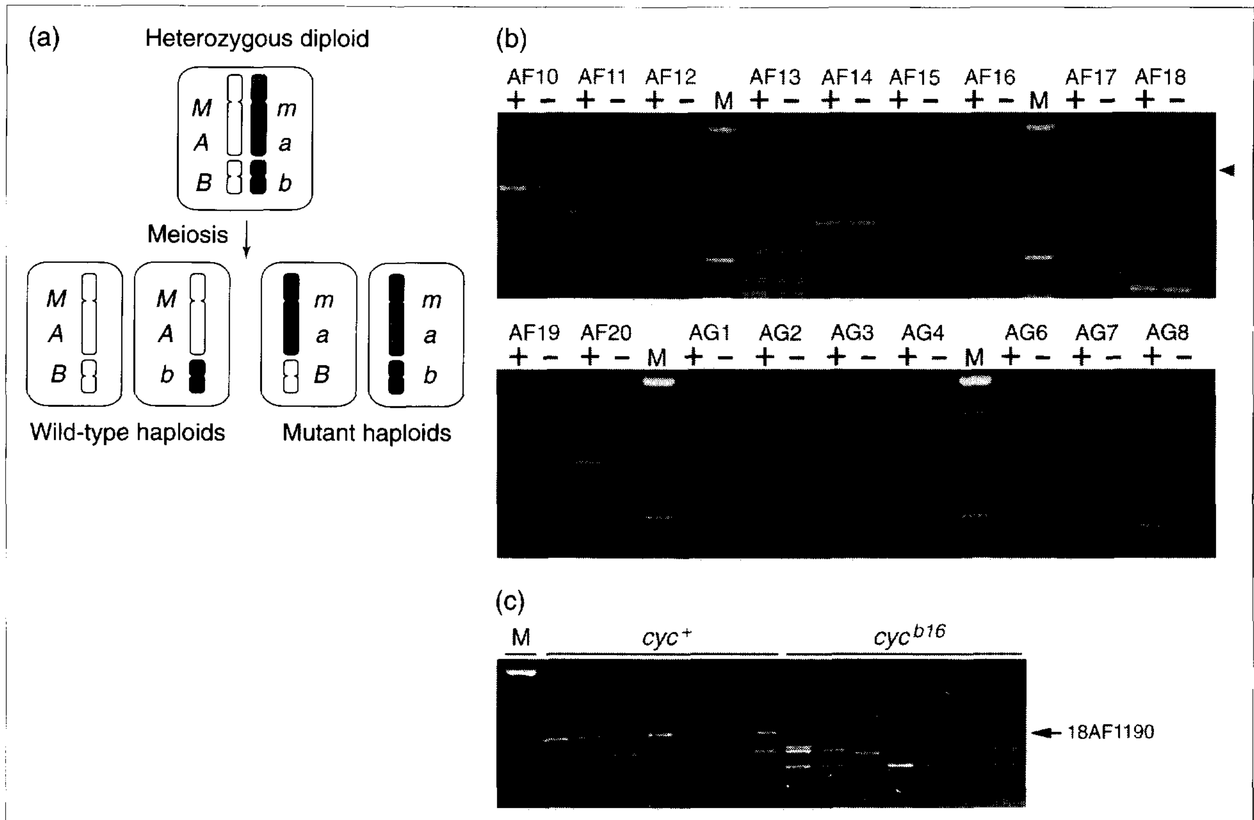


FIGURE 1. Bulked segregant analysis. (a) Animals carrying the mutation (*m*) in a laboratory strain background (shaded chromosomes) are crossed to a genetically divergent wild-type mapping strain (open chromosomes) to produce F_1 females heterozygous for the mutation and for many other genetic markers (top). Genomic DNA is isolated from individual haploid embryos produced from these females and two DNA pools are prepared (bottom), one phenotypically wild type and the other mutant. Linked markers are identified by the differential amplification of fragments from the two pools. In the example shown, the wild-type pool contains allele *A*, but not *a*, and the mutant pool contains allele *a*, but not *A*. If locus *A* is a dominant random-amplified polymorphic DNA (RAPD) marker, the marker fragment will amplify from the wild-type pool, but not from the mutant pool. If the locus is a co-dominant simple-sequence length polymorphism (SSLP), then fragments of different sizes will amplify from the two pools. Both alleles of unlinked loci (e.g. *B* and *b*) are present in the wild-type and mutant pools due to independent assortment, so fragments from these loci will not amplify differentially. (b) Identification of RAPD markers linked to *cyc* by bulked segregant analysis. Wild-type (+) and mutant (−) DNA pools were scored by PCR with 18 different 10 nucleotide oligomer primers. As expected, most marker fragments were not differentially amplified from the two pools. Two of the fragments that were differentially amplified [one generated by primer AF18 (indicated by arrowhead) and another generated by primer AG3] were shown to be linked to *cyc* by analyzing their segregation in individual haploid embryos. (c) Analysis of the 1190 bp AF18 RAPD marker, called 18AF1190 (arrow), in 14 haploid individuals from a Darjeeling (DAR) × cyc^{b16} mapping family. All of the wild-type animals inherited the dominant RAPD allele (i.e. the allele that amplifies the fragment) and all of the mutants inherited the recessive (non-amplifying) allele. Other marker fragments segregating independently of *cyc* are also visible. Abbreviation: M, marker.

If the location of a cloned gene on the zebrafish map could be predicted from the position of its presumed ortholog on human and mouse maps, then information from these dense mammalian maps could be used to help formulate candidate gene hypotheses for zebrafish mutations. The human and mouse genomes, for example, differ by about 150 chromosome rearrangements and have conserved chromosome segments with an average size of about 9 cM (Refs 50, 51). When a gene is mapped in mouse, one can predict with about 95% confidence that its ortholog resides in a specific, identified and conserved chromosome segment in humans. Can this approach extend to zebrafish?

Given the large number of chromosome rearrangements that have occurred between human and mouse genomes in the 100 million years since the radiation of placental mammals[52], it came as a surprise to discover that many large chromosome segments are conserved in the genomes of humans and zebrafish, which have been evolving separately for about 420 million years[53]. For example, apparent orthologs (or in one case, a homolog) of ten genes occupying the long arm of human chromosome 2 (about 5% of the length of the human genome) have been mapped in zebrafish (J.H. Postlethwait *et al.*, unpublished) and in mouse. Seven of these ten genes reside on LG9 and three on LG6 in zebrafish (Fig. 3). These genes are also on two different chromosomes in mouse, although the inferred chromosome breakpoint is in a different location. The results suggest that the common ancestor of mammals and zebrafish might have had these ten genes on one chromosome, followed by two independent chromosome fission events in the zebrafish and mouse lineages, although other models are possible.

The discovery of extensive sharing of chromosome segments between zebrafish and humans enables the use of comparative gene mapping to suggest candidates for zebrafish mutations. If a zebrafish mutation lies in a

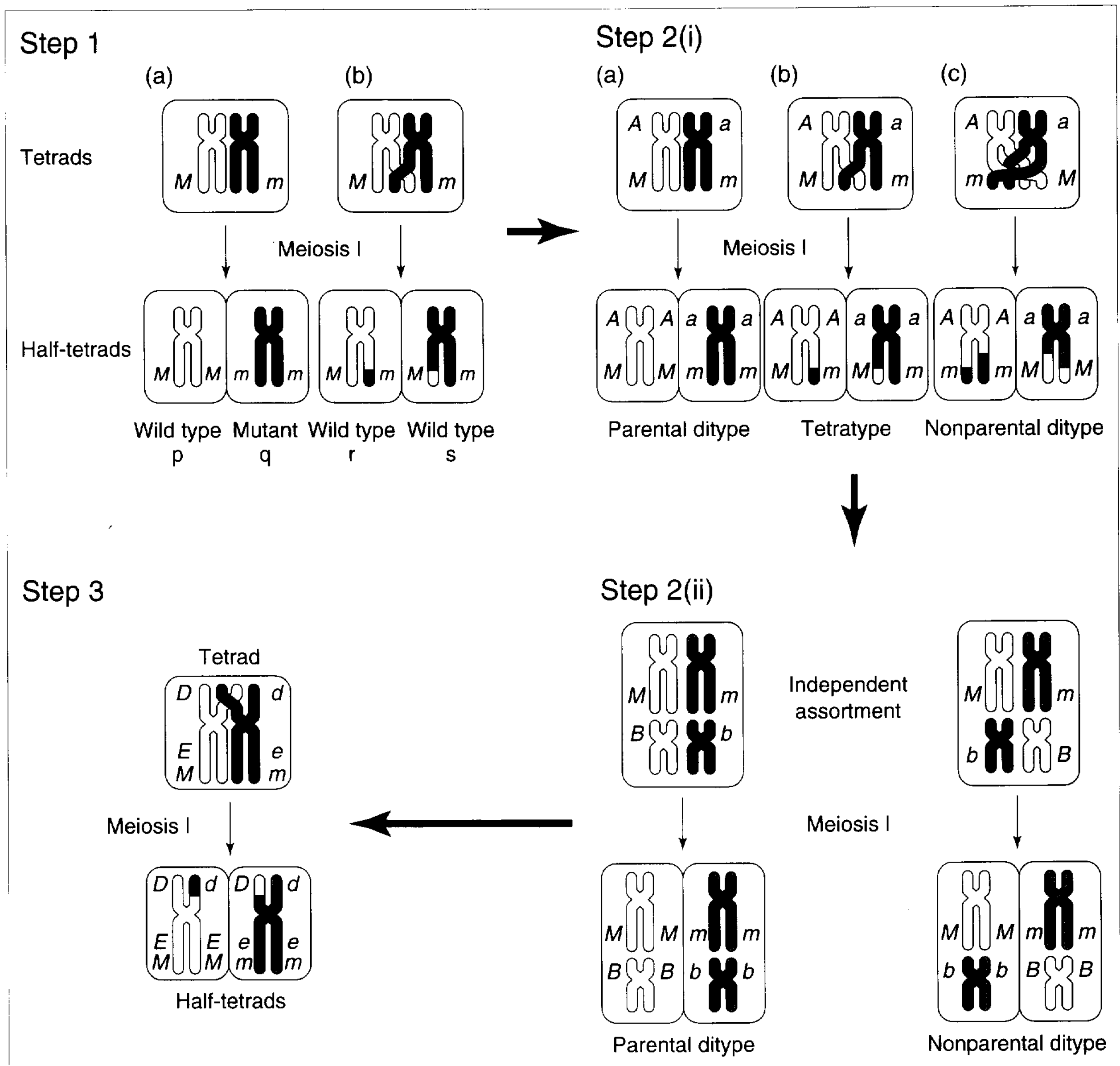


FIGURE 2. Centromere-linkage analysis. Step 1. The distance between a recessive mutation *m* and the centromere is calculated from the frequency of homozygous mutants in a family of half-tetrad diploids[46]. (a) A meiosis with no crossing-over between the centromere and the mutation yields 50% homozygous wild-type half-tetrads (p) and 50% homozygous mutant half-tetrads (q). (b) A meiosis with a single cross-over between the mutation and the centromere yields all heterozygous phenotypically wild-type half-tetrads (r and s). For a recessive mutation, the recombinant chromatid-bearing heterozygotes r and s cannot be distinguished from the nonrecombinant homozygous wild-type p. The frequency of recombinant half-tetrads (r+s) is 1–2q because p = q and p+q+r+s = 1. In a meiosis with a single cross-over between the mutation and the centromere, only one of the two chromatids in each half-tetrad is a recombinant chromatid; thus, the frequency of recombinant chromatids is half the frequency of recombinant half-tetrads, or (1–2q)/2. The map distance between the mutation and its centromere is defined as 100 times the frequency of recombinant chromatids, or 100(1–2q)/2. Genetic distances are underestimated for loci distal to the centromere because of chiasma interference. A correction assuming complete interference is given by the equation: map distance (cM) = 100(−ln 2q)/2 (Ref. 46). Step 2. The linkage group bearing the mutation is found by comparing the frequencies of half-tetrads arising from parental and nonparental ditype tetrads for the mutation and each of 25 centromere markers. A ditype tetrad has chromatids of two different genotypes relative to the scored markers; a tetratype tetrad has chromatids of four different types. Parental ditype tetrads have allele arrangements like the parents; nonparental ditypes have allele arrangements different from the parents. As can be seen from the figure, the type of tetrad can be inferred from the genotype of a half-tetrad. Step 2(i). (a, b) If the mutation *m* is linked to a specific centromere marker *A*, then homozygous mutant nonparental ditypes (*A m/A m* half-tetrads) will be rare compared with homozygous mutant parental ditypes (*a m/a m* half-tetrads). This is because nonparental ditypes appear only after (c) four-strand double cross-overs, which are rare due to interference. Step 2(ii). By contrast, if the mutation *m* is not linked to a given centromere marker *B*, then homozygous mutant nonparental ditypes (*m B/m B*) and homozygous mutant parental ditypes (*m b/m b*) will be found equally often due to independent assortment. Step 3. The chromosome arm on which the mutation resides is determined by examining homozygous mutant half-tetrads for their genotypes at two readily distinguished loci *D* and *E* located on opposite chromosome arms at the distance from the centromere calculated in Step 1. Homozygous mutant half-tetrads will almost always be homozygous for a locus *e* closely linked in coupling on the same arm, but will often or occasionally be heterozygous for a locus *d* located in coupling the same distance from the centromere on the opposite chromosome arm.

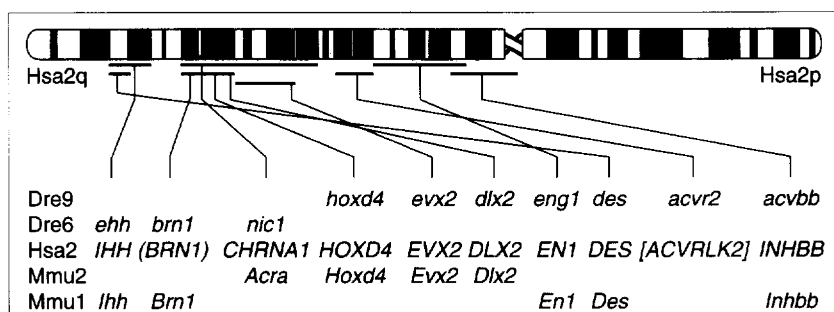


FIGURE 3. Conservation of syntenies: the zebrafish map and human chromosome 2. The orders of cloned genes mapped in zebrafish (Dre) linkage groups 6 and 9 compared with presumably orthologous genes located on human (Hsa2) and mouse (Mmu) chromosomes. Gene order is ignored in the representations of mammalian chromosomes. Genes: *ehh, IHH* and *Ihh, echidna* and *Indian hedgehog*; *brn1, BRN1* and *Brn1, brain1* (Refs 62, 63) (the location of *BRN1* on Hsa2q is inferred from comparative mapping with mouse); *nic1, CHRNA1* and *Acra*, nicotinic acetylcholine receptor α subunit; *hoxd4, HOXD4* and *Hoxd4, homeobox d*; *evx1, EVX1* and *Evx1, even-skipped homologue-1*; *dlx2, DLX2* and *Dlx2, distal-less homeobox-2*; *eng1, EN1* and *En1, engrailed 1*; *acvr2* and its homolog *ACVRLK2*, activin receptor; *acvbb, INHβB* and *Inhbb, activin β b* and *inhibin β B*.

conserved chromosome segment, then human genes located in this segment become candidate orthologs for the zebrafish mutation. Subsequent molecular genetic studies can test these candidates. Reciprocally, and importantly, the phenotype of a zebrafish mutation, which can be understood with great cellular precision, can suggest functions for the human gene. Indeed, molecular and phenotypic characterization of zebrafish mutations that disrupt genes with human counterparts might provide an efficient way to assign functions to human genes known only by sequences that are identified by the human genome project. Thus, zebrafish might help move the human genome project from a descriptive phase toward functional genomics. To make this approach more useful, further research must define the boundaries of conserved chromosome segments.

Positional cloning in the zebrafish

The candidate gene approach for cloning mutations is limited by the number of genes for which sequences are available. Until most vertebrate genes are sequenced, researchers will often have to employ positional cloning methods to identify the molecular nature of genes disrupted by mutation. Recent work suggests that the positional cloning of zebrafish genes is feasible and that several advantages of the zebrafish genetic system facilitate this approach.

Identifying tightly linked markers

The first step in a positional cloning project is the isolation of a DNA probe near the mutation of interest. This usually occurs by finding a DNA polymorphism that segregates in close association with the mutation in genetic mapping crosses. In organisms with dense genetic maps, such markers can be found by localizing the mutation on the genetic map and determining which of the neighboring markers is linked most tightly.

The current average intermarker interval in the zebrafish genetic map is about where the mouse map was in the early 1990s. This density is such that most genes will not be close enough to a mapped marker to justify starting a chromosomal walk. This will become less of an impediment as more markers are added to the map and the average intermarker interval decreases.

To identify markers tightly linked to mutations in the absence of an extremely dense genetic map, we have employed BSA with RAPD markers[29,45]. SSLPs can also be used, but RAPD markers have an important advantage that makes them especially useful. RAPD markers require little initial effort to develop because 10 nucleotide oligomers (10mer) of arbitrary sequence serve as primers in RAPD PCR assays, and each RAPD primer can amplify about five bands that are polymorphic among laboratory strains[29,31]. Thus, BSA with RAPD markers allows an essentially unlimited number of genetic markers (there are 4^{10} possible 10mer sequences) to be scored for linkage to a particular locus without having to derive individual primer sets for each marker tested, most of which will, of course, be unlinked to the mutation of immediate interest. Amplified-fragment length polymorphism[54] markers are similarly advantageous and they have also proved useful for BSA.

High resolution mapping

Because large numbers of zebrafish embryos can be produced and genotyped with RAPDs and other PCR-based genetic markers, it is possible to map with high resolution the positions of markers identified by BSA. For example, a locus initially identified as a RAPD marker was found not to recombine with *flh* among 1332 haploid offspring[23]. An adult female zebrafish can repeatedly produce clutches of about 100 or more embryos. Each clutch can be grown in a single petri dish for several days, at which time each haploid embryo yields enough DNA for several thousand PCR assays. Thus, a large number of embryos can be obtained in a short time from a few females that are heterozygous for the markers to be assayed. A mapping cross containing 1000 individual haploid embryos is sufficient to resolve genetic markers separated by as little as 0.1 cM, which corresponds to a physical distance of about 60 kb on average[29]. Mapping at this resolution is useful for determining which of a set of markers is closest to a mutation and, therefore, which should be used to isolate genomic clones in the region of the mutation.

Bulked segregant screens for tightly linked markers have been successful for most mutants so far attempted, including *flh* (Ref. 23), *cyclops* (*cyc*; Ref. 29; W.S. Talbot, M. Gates, C. Kimmel and J.H. Postlethwait, unpublished) and *one-eyed pinhead* (Ref. 35), suggesting that this approach is a generally useful way to initiate positional cloning projects. The recent construction of large-insert zebrafish genomic libraries in YAC, BAC and PAC vectors should facilitate the isolation of contigs surrounding these linked markers (C. Amemiya, L. Zon and G. Silverman, pers. commun.; D. Smoller, pers. commun.). Given that the zebrafish genome is expected to show a

high level of syntenic and sequence conservation with the highly compact genome of pufferfish[55], it might be possible to expedite positional cloning by 'comparative walking'. Chromosomal walks could be initiated in zebrafish and continued in pufferfish, checking clones along the way back against the zebrafish genome to measure progress toward the mutant locus. This could accelerate a walk by a factor of nearly four, the size differential between the two teleost genomes.

Identifying the mutated gene in a contig

The final phase of a positional cloning project is the identification of the mutated gene within the genomic DNA isolated from the linked marker, a region that can encompass hundreds of kilobases and contain many transcription units. In mouse and human, techniques for isolating exons from genomic clones, including exon trapping and cDNA selection, have proved useful in gene identification (reviewed in Ref. 56). We expect this to be the case in zebrafish as well, particularly when a mutant phenotype provides information about the probable expression pattern of the mutated gene, as will be the case for many interesting developmental mutations with tissue-specific effects.

Molecular characterization of chromosomal rearrangements can also aid identification of mutated genes. This is especially true if inversions and translocations with breakpoints in the gene of interest can be identified, because these lesions can disrupt single genes and they can be detected in simple genomic DNA blotting experiments. γ-Rays are highly mutagenic in zebrafish[57] and genetic characterization of several γ-ray-induced *cyc* mutations suggests that some of these are discrete lesions useful in the identification of genes in the *cyc* region, while others are clearly translocations of large chromosomal segments (W.S. Talbot, E. Egan, C. Kimmel and J.H. Postlethwait, unpublished). A physical characterization of these mutations is required to determine the structures of these rearrangements and their usefulness in positional cloning.

Transgenic rescue of a mutant phenotype is perhaps the most efficient way to identify a clone containing a wild-type allele of a mutated gene. Cloned genes microinjected into early zebrafish embryos are inherited by a subset of the progeny of the injected cell and are expressed at later stages in those cells[58,59]. Microinjection of genomic clones containing the wild-type *flh* gene into *flh*$^-$ mutant embryos can partially rescue the phenotype of some injected animals, suggesting that mutated genes that act early in development can be identified by the ability of wild-type alleles to confer partial or mosaic rescue of mutant embryos (W.S. Talbot, Y-L. Yan and J.H. Postlethwait, unpublished).

Finally, clear evidence that attributes a mutant phenotype to a lesion in a specific gene comes from the demonstration that several mutant alleles all have significant sequence alterations in that gene. Once the mutation is cloned, the fun can begin, and the information and materials gained from the cloning can be used to help learn how vertebrate embryos develop.

Conclusions

Evidence summarized here suggests that the zebrafish genetic system has matured to the point that tools are now available for the molecular identification of genes disrupted by mutations. Insertional mutagenesis will provide immediately clonable mutations; the use of centromere-linkage analysis will expedite the mapping of mutations; syntenies shared with mammalian genomes will accelerate the identification of candidate genes for mapped mutations; bulked segregant analysis will facilitate identification of markers closely linked to mutations; and large insert genomic libraries will provide material for the initiation of positional cloning projects. As more cloned genes are added to the map, candidate gene cloning will become more frequent and, until then, positional cloning will catch mutations of high interest. Molecular analysis of zebrafish mutations, coupled with the exquisite phenotypic characterization possible in this species, will complement developmental genetic information obtained in other vertebrates and hence enrich our understanding of the development of all vertebrates.

Acknowledgements

We thank S. Amacher, J. Eisen, A. Fritz, D. Kane, C. Kimmel, A. Schier and M. Westerfield for helpful comments on the manuscript. We also thank E. Egan, M. Gates, S. Horne and Y-L. Yan for allowing us to cite unpublished results. The Mouse Genome Database [http://www.informatics.jax.org/ (data retrieved 05/12/96)], Mouse Genome Informatics and The Jackson Laboratory, Bar Harbor, ME, USA are acknowledged, along with Online Mendelian Inheritance in Man (http://www3.ncbi.nlm.nih.gov/omim/). This work was supported by grants P01HD22486 and R01RR10715.

Box 2. Internet resources for zebrafish development and genetics

Lists of zebrafish mutations:
http://zebrafish.mgh.harvard.edu/database.html
http://zfish.uoregon.edu/zf_info/zfbook/zfstm.html

Genetic nomenclature:
http://zfish.uoregon.edu/zf_info/zfbook/lab_desig.html
http://zfish.uoregon.edu/zf_info/zfbook/chapt7/7.1.html

Atlas of zebrafish anatomy:
http://zebrafish.mgh.harvard.edu/anatomy.html

Embryonic staging series:
http://zfish.uoregon.edu/zf_info/zfbook/stages/stages.html

Movies of zebrafish development:
http://fishnet.bio.temple.edu/movies/cleave.mov
http://weber.u.washington.edu/~fishscop/

Reference list to zebrafish literature:
http://zfish.uoregon.edu/zf_info/zfbook/zfrefs.html

Zebrafish sequence analysis project:
http://lenti.med.umn.edu/zebrafish/zfish_top_page.html

Zebrafish general information:
http://zebra.sc.edu/
http://zfish.uoregon.edu/
http://www-igbmc.u-strasbg.fr/index.html
http://wwweb.mpib-tuebingen.mpg.de/abt.3/

References

1 Streisinger, G. *et al.* (1981) *Nature* 291, 293–296
2 Kimmel, C.B. (1989) *Trends Genet.* 5, 283–288
3 Driever, W., Stemple, D., Schier, A. and Solnica-Krezel, L. (1994) *Trends Genet.* 10, 152–159
4 Mullins, M.C., Hammerschmidt, M., Haffter, P. and Nüsslein-Volhard, C. (1994) *Curr. Biol.* 4, 189–202
5 Solnica-Krezel, L., Schier, A.F. and Driever, W. (1994) *Genetics* 136, 1401–1420
6 Riley, B.B. and Grunwald, D.J. (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 5997–6001
7 Driever, W. *et al.* (1996) *Development* 123, 37–46
8 Fritz, A., Rozowski, M., Walker, C. and Westerfield, M. (1996) *Genetics* 144, 1735–1745
9 Haffter, P. *et al.* (1996) *Development* 123, 1–36
10 Henion, P.D. *et al.* (1996) *Dev. Genet.* 18, 11–17
11 Moens, C.B. *et al.* (1996) *Development* 122, 3981–3990
12 Gaiano, N. *et al.* (1996) *Nature* 383, 829–832
13 Allende, M.L. *et al.* (1996) *Genes Dev.* 10, 3141–3155
14 Felsenfeld, A.L. (1996) *Nat. Genet.* 14, 258–263
15 Eisen, J.S. (1996) *Cell* 87, 969–977
16 Grunwald, D.J. (1996) *Science* 274, 1634–1635
17 Holder, N. and McMahon, A. (1996) *Nature* 384, 515–516
18 Halpern, M.E., Ho, R.K., Walker, C. and Kimmel, C.B. (1993) *Cell* 75, 99–111
19 Weinstein, B.M., Stemple, D.L., Driever, W. and Fishman, M.C. (1995) *Nat. Med.* 1, 1143–1147
20 Ransom, D.G. *et al.* (1996) *Development* 123, 311–319
21 Weinstein, B.M. *et al.* (1996) *Development* 123, 303–309
22 Schulte-Merker, S. *et al.* (1994) *Development* 120, 1009–1015
23 Talbot, W.S. *et al.* (1995) *Nature* 378, 150–157
24 Brand, M. *et al.* (1996) *Development* 123, 179–190
25 Kimmel, C.B., Warga, R.M. and Schilling, T.F. (1990) *Development* 108, 581–594
26 Ho, R.K. and Kane, D.A. (1990) *Nature* 348, 728–730
27 Halpern, M.E. *et al.* (1995) *Development* 121, 4257–4264
28 Melby, A.E., Warga, R. M. and Kimmel, C.B. (1996) *Development* 122, 2225–2237
29 Postlethwait, J.H. *et al.* (1994) *Science* 264, 699–703
30 Williams, J.G.K. *et al.* (1990) *Nucleic Acids Res.* 18, 6531–6535
31 Johnson, S.L., Midson, C.N., Ballinger, E.W. and Postlethwait, J.H. (1994) *Genomics* 19, 152–156
32 Johnson, S.L. *et al.* (1996) *Genetics* 142, 1277–1288
33 Knapik, E.W. *et al.* (1996) *Development* 123, 451–460
34 Johnson, S.L., Africa, D., Horne, S. and Postlethwait, J.H. (1995) *Genetics* 139, 1727–1735
35 Schier, A.F. *et al.* (1996) *Development* 124, 327–342
36 Dib, C. *et al.* (1996) *Nature* 380, 152–154
37 Hudson, T.J. *et al.* (1995) *Science* 270, 1945–1954
38 Schuler, G.D. *et al.* (1996) *Science* 274, 540–546
39 Ekker, M. *et al.* (1996) *Genomics* 33, 57–64
40 Endo, A. and Ingalls, T.H. (1968) *J. Hered.* 59, 382–384
41 Schreeb, K.H., Groth, G., Sachsse, W. and Freundt, K.J. (1993) *J. Exp. Anim. Sci.* 36, 27–31
42 Pijnacker, L.P. and Ferwerda, M.A. (1995) *Genome* 38, 1052–1055
43 Daga, R.R., Thode, G. and Amores, A. (1996) *Chromosome Res.* 4, 29–32
44 Dietrich, W. *et al.* (1992) *Genetics* 131, 423–447
45 Michelmore, R.W., Paran, I. and Kesseli, R.V. (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 9828–9832
46 Streisinger G. *et al.* (1986) *Genetics* 112, 311–319
47 Kauffman, E.L. *et al.* (1995) *Genomics* 30, 337–341
48 Sepich, D. and Westerfield, M. (1993) *Soc. Neurosci.* 19, 1294
49 Beier, D.R., Dushkin, H. and Sussman, D.J. (1992) *Proc. Natl. Acad. Sci. U. S. A.* 89, 9102–9106
50 Nadeau, J.H. and Taylor, B.A. (1984) *Proc. Natl. Acad. Sci. U. S. A.* 81, 814–818
51 Copeland, N.G. *et al.* (1993) *Science* 262, 57–66
52 Novacek, M.J. (1992) *Nature* 356, 121–125
53 Ahlberg, P. and Milner, A. (1994) *Nature* 368, 507–514
54 Vos, P. *et al.* (1995) *Nucleic Acids Res.* 23, 4407–4414
55 Brenner, S. *et al.* (1993) *Nature* 366, 265–268
56 Monaco, A.P. (1994) *Curr. Opin. Genet. Dev.* 4, 360–365
57 Chakrabarti, S., Streisinger, G., Singer, F. and Walker, C. (1983) *Genetics* 103, 109–123
58 Westerfield, M. *et al.* (1992) *Genes Dev.* 6, 591–598
59 Reinhard, E. *et al.* (1994) *Development* 120, 1767–1775
60 Corley-Smith, G.E., Lim, C.J. and Brandhorst, B.P. (1996) *Genetics* 142, 1265–1276
61 Westerfield, M. (1995) *The Zebrafish Book*, University of Oregon Press
62 Sampath, K. and Stuart, G.W. (1996) *Biochem. Biophys. Res. Commun.* 219, 565–571
63 Avraham, K.B. *et al.* (1993) *Genomics* 18, 131–133

J.H. Postlethwait *is in the Institute of Neuroscience, University of Oregon, Eugene, OR 97403, USA.*
W.S. Talbot *is in the Skirball Institute of Biomolecular Medicine and the Department of Cell Biology, New York University Medical Center, 540 First Avenue, New York, NY 10016, USA.*

It's easy to subscribe to *TIG*

Just email **journals@elsevier.co.uk**

or call **+44 1865 843300** or **+1 914 524 9200**

or fax **+44 1865 843940** or **+1 914 333 2444**

All you need to provide is your name, address, the month from which you would like your subscription to start, your credit card number and its expiry date. (Please do not send credit card details by email.)

50% off all new student subscriptions!

Did you know that as a student you are entitled to a **special discount** on a personal subscription to *Trends in Genetics?*

See the subscription order form for details.