

## An on-line database for zebrafish development and genetics research



Monte Westerfield, Eckehard Doerry\*, Arthur E. Kirkpatrick\*, Wolfgang Driever† and Sarah A. Douglas\*

We have built a relational database of zebrafish developmental and genetic research information accessible via the World Wide Web. Our team of biologists and computer scientists employed a user-centered design process, using input from the research community to tailor the contents and usability of the database. The database supports the broad range of data types generated by zebrafish research including text, images and graphical information about mutations, gene expression patterns and the genetic map. Data are entered both by the database staff and directly by authorized users. The database also maintains links among data, scientists and laboratories, thus facilitating information exchange within the research community.

**Key words:** genes / genetic map / mutations / object-oriented relational database / World Wide Web

©1997 Academic Press Ltd

### Importance of the zebrafish database

THE ZEBRAFISH HAS RECENTLY emerged as a premier organism for studies of vertebrate development and genetics.<sup>1-3</sup> Powerful techniques allow efficient generation and recovery of zebrafish mutations that affect a wide range of genes including those that regulate developmental patterning, organogenesis, physiology and behavior. The functions of many of these genes appear to be conserved among different vertebrate groups.<sup>4-6</sup> Analysis of zebrafish mutations therefore provides insights into gene functions in other vertebrates, including humans.

The remarkable success of research using zebrafish has generated a serious information access problem. Although the use of zebrafish in genetic research is

relatively new, the number of labs and the amount of data generated by these labs are increasing at a phenomenal rate. For example, ongoing genetic screens have already identified over 3000 mutations.<sup>7,8</sup> As mutant lines become available, the information about these lines grows rapidly; studies of each mutant generate a description of its phenotype, genetics, lineage, map location, interactions with other genes, etc. The number of annual zebrafish publications has increased over 400% in the past six years. The information already far exceeds the ability of individual scientists to track and organize it. For these reasons, a concerted effort has begun to establish and maintain a centralized database for the zebrafish community.

The zebrafish database project grew from our earlier World Wide Web (WWW) site, <http://zfish.uoregon.edu>. Due to the dramatic increase in information and the demand for more sophisticated search methods, we have integrated the Web site information and other zebrafish research information into an object-oriented relational database. The resulting database project is unique among biological systems because it incorporates many novel design principles and database features:

- designed with a user-centered collaborative method;
- meets usability and usefulness requirements of biologists;
- developed exclusively for WWW access;
- includes unpublished data;
- depends upon user submission of data;
- integrates genotype and phenotype information;
- supports multi-media data.

### Database design process

From the outset, both biologists and computer scientists participated together to design the zebrafish

From the Institute of Neuroscience and the \*Computer Science Department, University of Oregon, Eugene, OR 97403, USA and †Institut für Biologie, Universität Freiburg D-79104, Freiburg, Germany

©1997 Academic Press Ltd  
1084-9521/97/050477+12 \$25.00/0/sr970173

database. The ultimate usefulness and usability of a database depend upon a careful assessment of the requirements of the users, detailed testing of prototypes by real users, and analysis of the users' interactive behavior while using the database. We<sup>9</sup> followed the basic steps of user-centered (Figure 1; ref 10) and participatory<sup>11</sup> design.

### **Step 1: Develop database and usability requirements**

We began by identifying what types of data are needed in the database and how the database can be used to study biological problems. To identify these requirements, the computer scientists on the project needed to understand the everyday work of zebrafish researchers and its relationship to their use of the database. This was a difficult task because biological research is very specialized, using techniques and vocabulary which are unfamiliar to most computer scientists. Accordingly, the computer scientists interviewed zebrafish researchers, read journal articles, participated in experiments, and attended research talks and lab meetings. We used questionnaires to gather design information from scientists around the world, distributing them at workshops and via the zebrafish WWW site. We also examined existing web-accessible biological databases to evaluate their content and user interfaces and we obtained help and design suggestions from scientists who are developing databases for mouse and *Drosophila*. We used this information to formulate the requirements for the database.

### **Step 2: Iterate detailed design process**

We next began an iterative refinement phase with cycles of design, prototype implementation, and evaluation with real users. In a process known as usability testing, selected pairs of zebrafish scientists evaluated each prototype using typical data submission and retrieval tasks. We videotaped and analysed these sessions to identify problems which we solved with changes implemented in subsequent prototypes

in the iterative design cycle. When we were satisfied with a prototype, we made it available to a small group of zebrafish scientists (acting as beta testers) through the WWW; access to the prototypes was limited to these testers. Each screen contained a comment form allowing our testers to send us feedback. At the end of the beta testing period, we interviewed our testers to assess the prototype design and improve it.

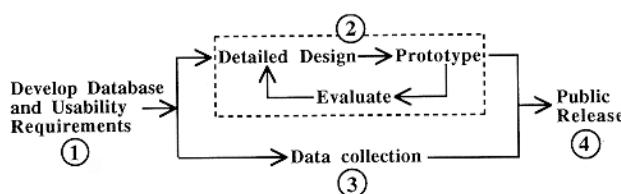
### **Steps 3 & 4: Data collection and public release**

Step 3, data collection, was conducted in parallel with the iterative design cycle. Several laboratories helped by collecting and formatting data for entry into the database. Public release of the database represents an additional step, rather than the end of the user-centered design process; usability analysis continues, allowing the system to evolve to meet the changing needs of the users. The commentary forms for gathering user feedback remain available in the public release. We are also recording (anonymously) the sequence of screens visited by each user and the total number of visits to each screen to identify common usage patterns and to expose areas of confusion. Finally, we are planning to conduct user surveys, periodically interviewing a sample of users to assess usage patterns, good and bad features of the database, and interest in including additional data types.

## **Overview of the zebrafish database**

### **Relational database searches**

Because of the broad range of zebrafish research, the database system must support many different types of data including images, spatialized graphical data and text information. To make these experimental data available to the research community, we first considered using standard WWW document technology as a standard WWW server could list all the information that the zebrafish community needs. However, a simple WWW site has significant limitations when compared to a database system. For example, all data must be tediously entered and interlinked by hand in the source files, and data access is limited to simple file contents browsing because there is no underlying data model. Zebrafish researchers need to put together complex searches which combine genetic, developmental, image, and other types of data. Only a database system that supports model-based data



**Figure 1.** The user-centered design process.

organization and dynamic relational querying can support such powerful searching capabilities.

For these reasons we chose a commercial object-oriented relational database management system (ORDBMS) which supports:

- user-defined data types;
- a standard language for relational queries;
- image, spatial and temporal data types;
- security, query optimization, and data integrity.

The relational database system supports an abstract data model which separates the logical data model from the low-level data structures and forms the foundation for relational querying. Adding object-oriented modeling<sup>12,13</sup> to a relational database creates important advantages including inheritance and class structures. Its extensible type system allows creation of data types to represent the full range of biological data including images and movies. Moreover, an ORDBMS, like traditional database management systems, provides the support we require for security, query optimization, data integrity and recovery.

#### *Inexpensive, ubiquitous access*

Our goal is to provide easy access to the database for researchers with little or no knowledge of how the database operates. It is also important for researchers world wide to access the database using general purpose, inexpensive hardware and software because most users of the database are biologists with desktop computers and limited database expertise. For these reasons, we have designed and built a WWW-accessible interface which allows users to interact with the database using commonly available WWW browser software like Mosaic or Netscape (Figure 2). A primary challenge in the user interface design was to determine in advance which subset of queries would satisfy the needs of most users but still provide a simple interface. We chose this approach, rather than providing a more powerful, but cumbersome standard query language (SQL) interface. The interface translates browsable lists of search criteria submitted by users (Figure 3A) into SQL queries, submits the queries to the underlying database, and formats the results into standard hypertext markup language (HTML) pages which are displayed by the user's browser software (Figure 3B). Items in the query result are linked (Figure 3B, arrow) to pages of additional more detailed information (Figure 3C). During data submission by users, the interface also

acts as the arbitrator between the research labs and the database, ensuring uniformity of the data types and the assignment of proper descriptors (attributes) to data records, thus enforcing the data model.

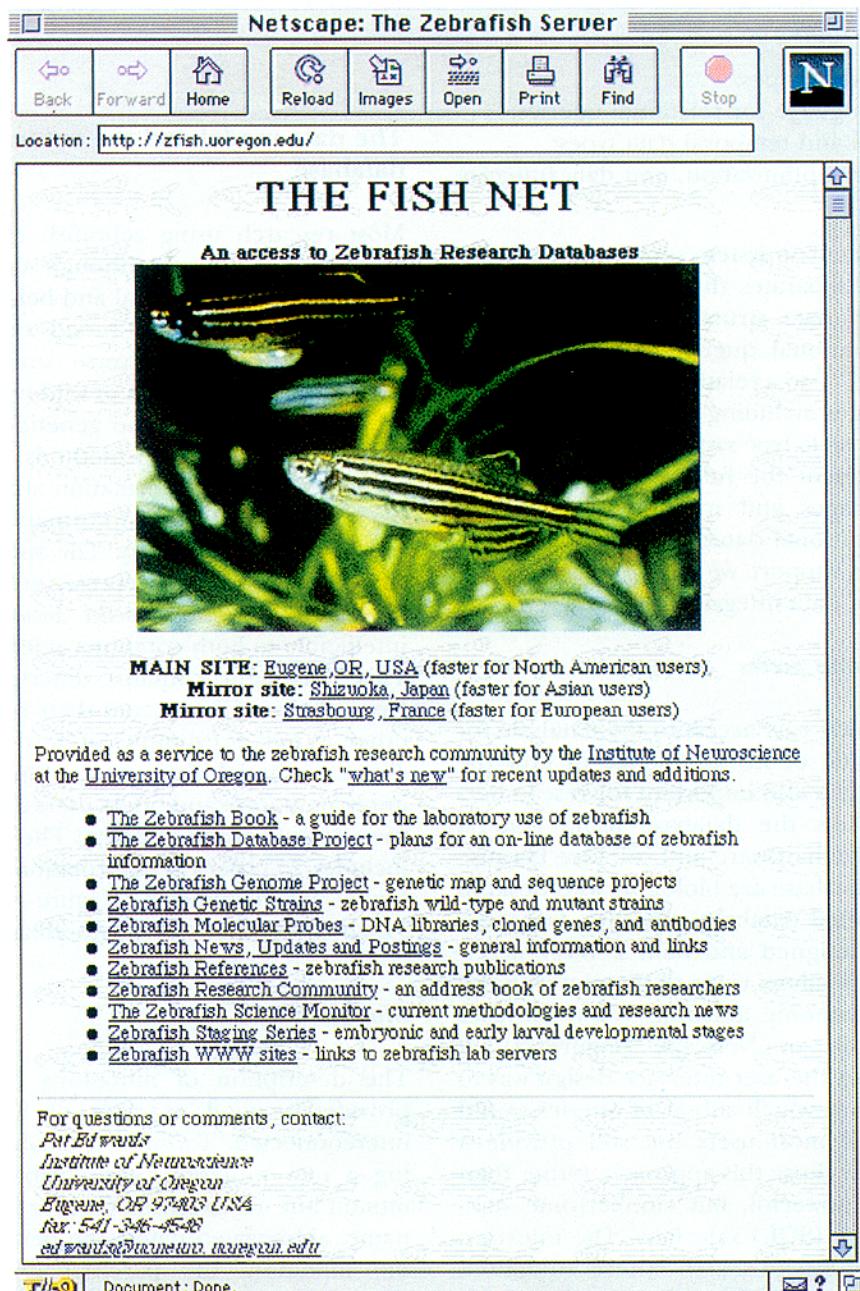
#### **The data model and the contents of the database**

Most research using zebrafish centers on development and genetics,<sup>2</sup> although there are increasing numbers of physiological and behavioral studies. The database supports the broad range of data types generated in these diverse studies including text descriptions and images of wild-type and mutant fish, graphical displays of the genetic map, physiological records, and laboratory methods.<sup>14</sup> In addition, the database contains information about researchers and laboratories facilitating information exchange within the research community. The specifications for how these types of data are represented in the database are described in a *data model document* written to be intelligible to both computer scientists and biologists. This document contains descriptions of the data types, their attributes, and their relationships to each other, as well as examples of how to use the database. The data model serves as the blueprint for database implementation and offers users a concise overview of the contents of the database. The current data model includes 25 classes of information, most of which are highly interconnected (Figure 4; see also [http://zfish.uoregon.edu/zf\\_Sinfo/dbase/db.html](http://zfish.uoregon.edu/zf_Sinfo/dbase/db.html)).

#### *Mutations*

The description of mutations in the data model provides a good example of how information is interconnected. To enter a database record describing a new mutation, the attributes describing the mutant line are specified (Figure 4). These include its name, abbreviated name, the person who discovered the mutation, the parental lineage, the mutant phenotypes, and the lab that currently has it. The genetic composition of this stock is specified by the chromosomes it contains with each chromosome being specified by the alterations it contains. The complete description of the mutant is thus provided by the combination of information describing the alterations contained in each chromosome and the full set of altered chromosomes.

The lineage of a stock is specified by the mother and father used to generate that stock. This information is especially useful for entering information about stocks containing more than one mutation



**Figure 2**

**Figure 2.** The graphical interface. The interface allows users to communicate with the database using standard WWW browsers that run on desk top computers. No other specialized user-side software is required.

created by crossing parents with different single mutations (Figure 5). Each parental stock is specified by its chromosomes which are each specified by the alterations they contain. The lineage of a parental line is specified by its parents, and so on. The data model associates a fish with its parents, so, for example, when information about the parent's genotype is updated, the updated information automatically appears in the description of the offspring. Thus, data entered at different times and in different contexts are interconnected by the relationships defined by the data model. This relational property of the database also allows scientists to retrieve information in many different ways. For example, one can search for and display mutations by phenotype, genetic map location, mutagen, type of chromosomal rearrangement, lab of origin, double mutants derived from particular single mutations, or any combinations of these or other attributes (Figure 3).

### **Images**

Many studies using zebrafish rely on anatomical or morphological analyses. Much of the information obtained from these studies is recorded in photomicrographs of anatomical structures, gene expression patterns, or labeled cells in wild-type and mutant embryos. In some studies, photomicrographs recorded during development of live embryos are made into time-lapse movies. Thus, images are included in many of the data records in the database and we have defined a data type which represents images (Figure 4). The attributes of images include the stock and developmental stage of the fish, the type and orientation of the image, the magnification, and the anatomical structures shown. Anatomical structures are defined, as a function of developmental age, by the anatomical atlas, a dictionary of hierarchically arranged terms. The developmental staging atlas includes images that illustrate the morphological features that define each stage as well as a text description of the stage. Similarly, the attributes of cloned genes include images of their expression patterns linked to text descriptions which, in turn, are linked to developmental stages. This organization of the data model allows the scientist to combine various attributes into complex searches, such as finding all images of gene expression patterns in a particular region of the CNS at a particular developmental stage in a particular mutant.

## **Data sources, entry and integrity**

### *Data entered by database staff*

Some types of data generated from studies of zebrafish are currently available in other public databases. For example, most zebrafish research publications are listed in Medline or other literature reference resources. The database staff searches these public sources on a regular basis, and downloads information into the zebrafish database. Similarly, DNA, gene and protein sequences are available through GENBANK and other on-line sources. The zebrafish database provides links from zebrafish gene or protein names to the raw data in these other sources.

### *Submissions by authorized users*

A goal of the zebrafish database project is to archive as much information as possible. Although only a fraction of the data obtained in scientific studies is available to the research community in peer reviewed journals, much of the unpublished information may be useful and important for current and future studies. The data obtained from screens for new mutations provide a good example. Recent screens have identified over 3000 mutations affecting zebrafish morphology and developmental patterning. Only a few hundred of these have been described in any detail in peer reviewed publications and these descriptions are necessarily incomplete due to space constraints imposed by the journals. A complete set of photographs and phenotypic descriptions of these mutants will never appear in journals even though the information has been obtained and may be useful. If these data are not preserved in a database, much of the information will eventually be lost. For these reasons, we have designed the zebrafish database to accept submissions directly from authorized scientists, without review. This means that some records contain preliminary or incomplete data whereas others, based on published information, are more complete and accurate. To distinguish among these sources, each record is marked with a tag which identifies whether a researcher or database staff member submitted it and whether it has been published. During a database search, the user can therefore determine a level of confidence for each data item.

**A**

Current TASK: <input type="button" value="ZFIN HOME"/>	<input type="button" value="Find STOCKS"/>	QUESTIONS COMMENTS
Search for <input type="text" value="mutant"/>	<input type="button" value="RESET"/> <input type="button" value="SEARCH"/>	
NAME contains: <input type="text"/>	<input text"="" type="button" value="Lab: &lt;input type="/>	
Mutated CHROMOSOME(1-25): <input type="text"/>	carries ALLELE: <input type="text"/>	
MUTATION TYPE: <input type="text"/>	phenotype <input type="text"/>	contains fused eye

**B**

Search Results (searching mutant lines)				
MUTANT	Allele(s)	Mut. Type	Chrom#	Lab
<b>cyclops(cyc)</b>	b15	UNKNOWN	12	Eugene
<b>cyclops(cyc)</b>	b213	translocation	12	Eugene
<b>cyclops(cyc)</b>	b229	UNKNOWN	12	Eugene
<b>cyclops(cyc)</b>	b238	translocation	12	Eugene
<b>cyclops(cyc)</b>	b352	UNKNOWN	12	Eugene
<b>cyclops(cyc)</b>	b361	point mutation	12	Eugene
<b>cyclops(cyc)</b>	b325	UNKNOWN	12	Eugene
<b>cyclops(cyc)</b>	b323	UNKNOWN	12	Eugene
<b>cyclops floating head(cx Tb)</b>	b325	UNKNOWN	12	Eugene
<b>cyclops floating head(cx Tb)</b>	b326	UNKNOWN	13	Eugene
<b>cyclops floating head(cx Tb)</b>	b323	UNKNOWN	12	Eugene
<b>cyclops floating head(cx Tb)</b>	b327	UNKNOWN	13	Eugene
<b>cyclops(cx)</b>	b323	UNKNOWN	12	Eugene
<b>cyclops(cx)</b>	b325	UNKNOWN	12	Eugene

**C**

QUESTION ANSWERS

Current TASK:

Find STOCKS

Record ID: ZDB-240824-147 Last updated: Never modified Update this Record Mutation History

**MUTANT line: cyclops (cyc)**



**Brief Instructions**

- Use the "Criteria" pane to describe the data for which you are looking.
- Press the "SEARCH" button. All matching lines will appear in this pane.
- From the individual data items in more detail by clicking on them. Use the "BACK" button in the browser to return to this search after examining a data item. Alternatively, you may simply click to the search by clicking on the "Task..." link located in the "Task Path" near the top of this screen.

In general, you may click on any link shown in the "Task Path" appearing at the top of this screen to return to a corresponding previous screen. For instance, clicking on the "ZFIN HOME" will take you back to the starting (first) screen to start a new search.

**ORIGIN/AVAILABILITY:**

DISCOVERER: Walker, Charlotte	LAB: Eugene
ORIGINAL STOCK NUMBER:	CURRENT SOURCE: Teratology, BII

**MUTANT ALLELES:**

ALLELE	CHROM#	MUTANT CHROMOSOME
b16	12	<input type="text"/>

**BREEDING:**

Maternal Strain: G20K026	Segregation: Mendelian
Paternal Strain: UNKNOWN	

**PHENOTYPE INFORMATION:**

DESCRIPTION: Fused eye. Planar body with tail curved down. Fin plate missing or patchy. Hatching and larval development normal.

AFFECTED PARTS: None described.

IMAGES of this line:  ZDB-IMAGE-970312-16 (new);  ZDB-IMAGE-970312-16 (new);  ZDB-IMAGE-960821-7 (UNKNOWN);  UNKNOWN (seen during UNKNOWN stage).

**PUBLICATIONS:**

Specify Primary Publication: The cyclops mutation blocks specification of the floor plate of the zebrafish central nervous system.

Figure 3 (see caption on facing page).

### Updating data

A database record can be updated at any time by the original submitter, but, if a change is requested by a user who did not submit the original data, the update must be performed by the database staff. The database records the previous and the updated value, retaining a history of the changes. A record may need to be updated for several different reasons. First, simple errors occur during data entry: these errors can be corrected directly by the submitter during the data entry process. Second, some data records are incomplete because associated information was unknown when the data were first submitted. For example, a mutation may be entered before its map location is known. The database allows submission of partial records because information should be available in a timely manner, even if it is preliminary. The original submitter of a data record retains the privilege to update the record, so new information can be added as it becomes available. The database staff makes all other updates. Third, database records become outdated: due to the nature of scientific discovery, information describing a biological system or process changes as new studies are completed. Database records describing previous studies may thus, become incorrect or obsolete. For example, a gene may be cloned by more than one laboratory and entered into the database under different names. When the genes are shown to be identical, one of the records will need to be deleted or changed. This creates a more difficult problem because outdated records need to be identified and decisions made about how to deal with them. To solve such problems, the database staff communicates with the original data submitters. An oversight committee from the zebrafish research community helps make especially difficult decisions.

**Figure 3.** Searching for a mutation by phenotype. (A) The user sets up the search for mutations that cause fused eyes, by specifying the phenotype. (B) The results of the search include a list of mutant alleles with a fused eye phenotype. Each allele name is a 'hot-link' to more detailed information. (C) Detailed information about the *cyclops<sup>b16</sup>* mutation. Information specified for each attribute of the mutant data type is returned to the user. Many of these attributes are 'hot-links' to other data records.

### How the database deals with biological problems

#### Specifying the resolution or precision of the data

A general problem we have encountered during design of the database is how to specify the precision with which data were recorded. Developmental staging is a good example. Zebrafish develop very rapidly, growing from the fertilized egg to a free swimming larva in three days.<sup>15</sup> Developmental processes can change significantly over a very short period of time. To describe these processes adequately, precise developmental staging is required. However, the precision of staging varies widely among different studies: in some cases, investigators accurately stage development within a few minutes, whereas in other studies, the stage of the embryo from which data were obtained is only roughly specified, for example 1 day, 2 days, etc. These differences in precision can create ambiguities during data entry and searches. When for example, an investigator submits data that were obtained from embryos at a range of times during the second day of development, they need to be identified differently from data obtained at a particular developmental stage during that day. Similarly, when an investigator requests records from 1-day embryos, should these include all data from any stage between 24 and 48 hours, or only those data which were obtained precisely at 24 hours?

Our solution to this problem is to map the developmental stage to a time frame in hours. The higher the resolution specified, the narrower the time frame. For example, an embryo coarsely staged as 1 day is mapped to a broad time frame of 24–48 hours; a record from an embryo staged precisely with the standard staging series, like Blastula: 30%-epiboly, is mapped to a narrower time frame, in this case, 4.66–5.24 hours. Searches for records are specified in a similar manner. The user can request a range of developmental stages at whatever resolution is required. For example, searching for embryos between 1 and 3 days would be a broad search whereas embryos between Gastrula Dome and Gastrula Bud stages would produce a narrow, high-precision search. To carry out the search, the search criteria are, again, mapped to a time frame in hours. The database then retrieves records from developmental stages (time frames) that intersect the requested time frame. Matching records are displayed in priority order, based on how closely they match the requested time frame. Records from developmental

stages that fall completely within the time frame (i.e. the actual age of the embryo is positively known to have been within the requested period) are displayed first, records whose specified developmental stage partially overlaps or completely includes the requested time frame (i.e., the actual age of the embryo might be within the target period) are displayed last. In this manner, the investigator specifies the level of precision both during submission and retrieval of data records.

Spatial resolution is also a problem for specifying anatomical structures. Some data records contain very precise anatomical descriptions, whereas other contain more general terms. For example, the expression pattern of a gene may include the ventral thalamus. Some researchers would use the precise term, ventral thalamus, whereas others would use the progressively less specific terms diencephalon, forebrain, anterior central nervous system or head. Thus, if an investigator requests records of genes expressed in the ventral thalamus, should these include gene expression data loosely specified as from the head or only those data obtained specifically from the ventral thalamus? The hierarchical organization of the Anatomical Atlas specifies the relationships among anatomical structures, allowing users to search broadly for data records from the head or specifically for records from the ventral thalamus. The database also enforces a standardized anatomical nomenclature; during data submission, users must describe anatomical structures using standard terms selected from those listed in the Anatomical Atlas.

#### *Ambiguity in experimental data*

In some cases, data from different studies may differ or contradict each other. This means that searches may return ambiguous results. For example, the genetic map is constructed from data describing the segregation of markers among the offspring of a reference cross between parents that carry polymorphic forms of the markers. The location of a marker relative to other markers on the genetic map is calculated by counting the frequency of appearance of each polymorphic form among the offspring. The precision with which the marker is located thus depends upon the number of offspring examined. Moreover, the calculated positions of markers may vary among different reference crosses and some markers are not detectable in all crosses. To accommodate these variations, the database requires that submitters specify the reference cross and the preci-

sion used to calculate the position of each marker as it is entered. During searches, the database allows the user to select among available reference cross data and to view the precision of measurement of each marker. Serious ambiguities due to variations among reference crosses are referred to a chromosome committee of zebrafish scientists who make decisions about how the standardized genetic map should appear.

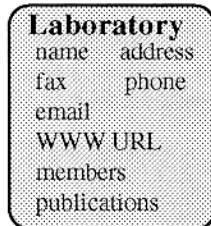
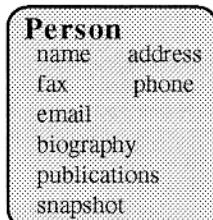
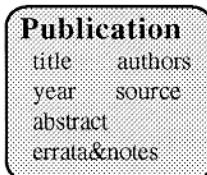
#### **How the database deals with WWW-based interface design problems**

Interface design for any Web-accessible database is severely constrained by basic limitations in the technology of the Web itself, primarily because the WWW environment was designed to support distribution of static documents, rather than the dynamic connections required to interact efficiently with a database. The limitations of HTML and Hypertext Transfer Protocol (HTTP), the communication protocol that controls interactions between the user and database server, impose several constraints.

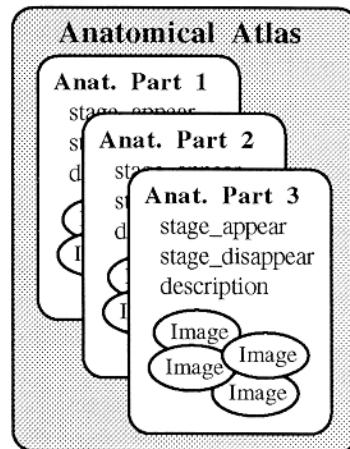
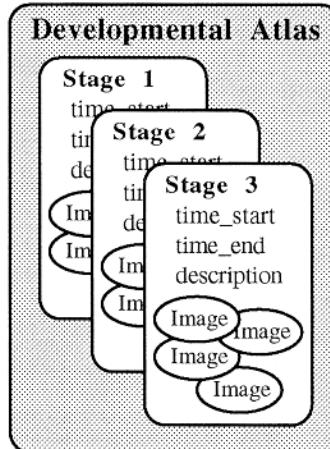
- *Anonymous Transactions.* The HTTP protocol does not distinguish among transactions at the user level, meaning that transactions are essentially anonymous.
- *Discrete, Stateless Transactions.* The HTTP protocol supports only discrete transactions between user and server. Each transaction retrieves a single HTML document. Both the HTTP server and the user treat each transaction independently of preceding transactions. Although both typically maintain a record of document access, this record plays no role in shaping the retrieval or appearance of future documents.

These constraints interfere with some of the basic principles and techniques of modern interface design. For example, anonymous transactions do not support user authentication and discrete, stateless transactions do not support persistent information (e.g. navigation, context-sensitive help), most forms of event-based feedback (e.g. immediate error-checking, graying of inappropriate menu choices) or interaction techniques like direct manipulation and animation. To ameliorate these limitations of the WWW environment, we use the database itself to store individual user and interface state information. To provide enhanced user interaction and to make the interface more responsive, the database generates the

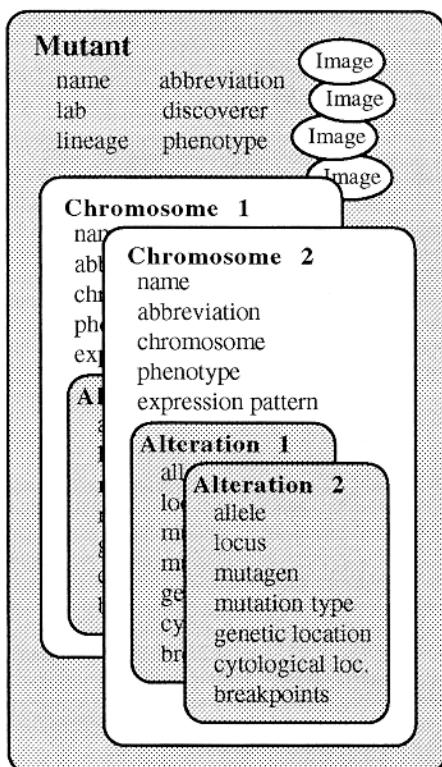
## Community Resources



## Reference Resources

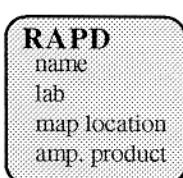
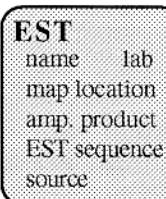
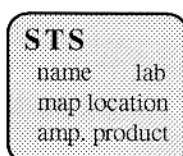
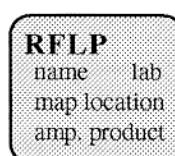
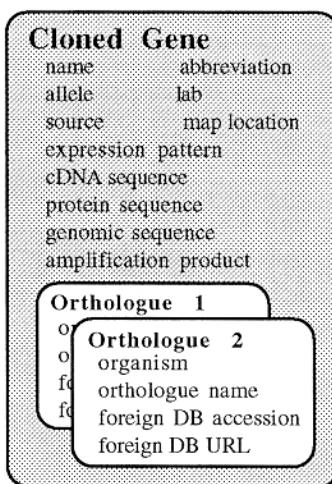


## Mutations

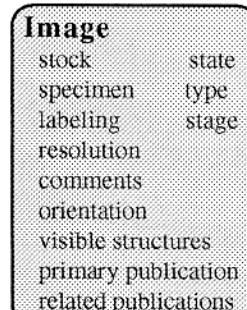


## Experimental Data

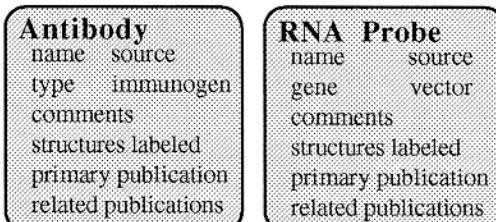
### Map Markers



### Images



### Labels



**Figure 4.** Graphical overview of the data model. The data model specifies the attributes of each type of data. Data types are listed in bold lettering, the attributes that describe each data type are listed below.

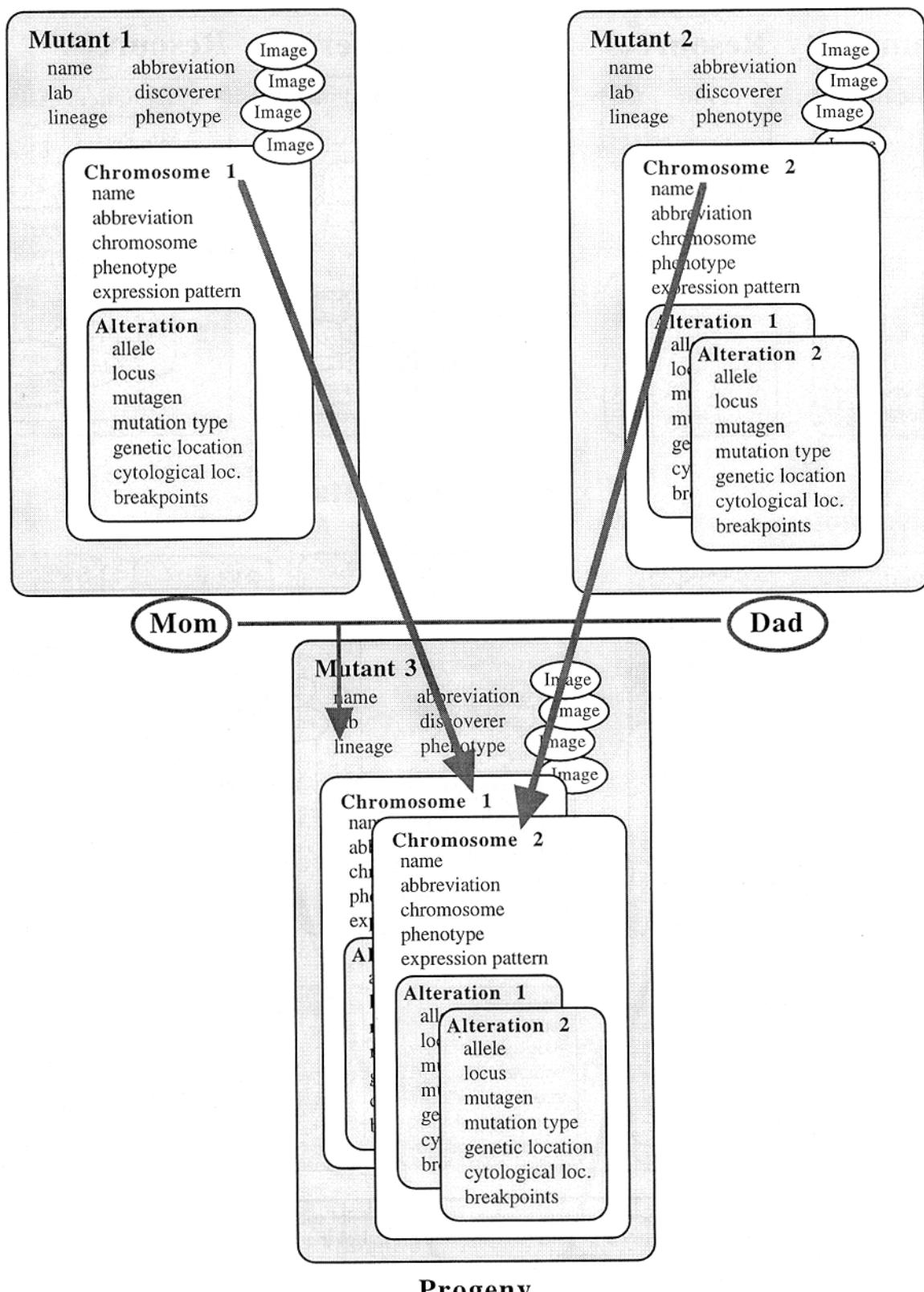


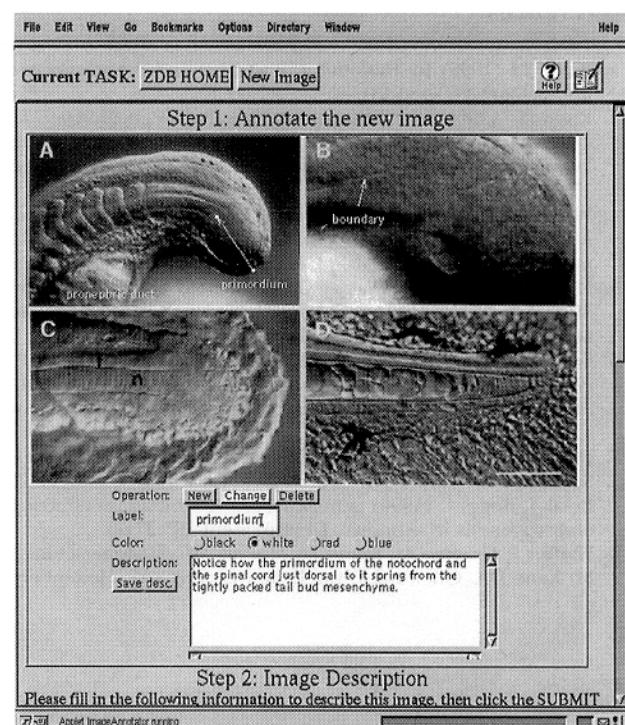
Figure 5 (see caption on facing page).

majority of interface screens dynamically. Raw application pages, containing static elements interspersed with scripted processing instructions, are stored as text in a database table. We use two types of scripted instructions. Database scripts, executed by the server, control the content and appearance of the returned page. JavaScript, executed by the client, controls user-side activities like checking submitted information for errors, enforcing correct nomenclature, and updating the navigation and help information. Although this combination of HTML scripting enhancements and storing user interface information in the database itself solved many of the problems associated with plain HTML, we also needed to use Java for more sophisticated graphical user interfaces with displayed animation and interaction techniques like zooming and panning. Our Java applets provide interactive graphical displays for annotating images (Figure 6) and for viewing the genetic map.

## Future prospects

The zebrafish database is still only partially complete. Our immediate goals are to finish designing interfaces to handle all the major data types represented in the data model. We will also need to populate these classes with data. Our longer-term goals include developing methods to access and integrate information from other databases. For example, cross-species comparisons of gene expression patterns, mutant phenotypes, and genetic linkage maps will provide important insights into vertebrate development and evolution. Such studies will be greatly facilitated by methods that allow cross-database searches. We also hope to develop methods for choosing search criteria graphically based upon anatomical and developmental parameters. Graphical searches will provide a much more natural means for biologists to query the database. Additional data types will also be added as

**Figure 5.** Specification of a double mutant made by crossing two single mutants. The parental genotypes (top) are associated with the new double mutant record (bottom) by the lineage attribute. Updates to chromosomal information for either parental stock automatically appear in the description in the double mutant.



**Figure 6.** A Java applet to annotate images. To annotate a data record on an image, the database sends the user an applet which runs in the user's browser. The user then draws and edits arrows and labels on the image and writes a text description. The annotated image is then submitted to the database.

future new directions develop within the zebrafish research community.

## Accessing the zebrafish database

The database can be accessed at: <http://zfish.uoregon.edu>. To gain authorization for data submission or to obtain additional information, contact:

[zfinadm@cs.uoregon.edu](mailto:zfinadm@cs.uoregon.edu).

## Acknowledgements

We thank Paul Bloch, Lauradel Collins, Don Pate and Mike McHorse for expert technical help, Pat Edwards for help with data entry, and our colleagues in the zebrafish research community who have provided suggestions and help with the beta testing. Supported by the W.M. Keck Foundation and the NSF (BIR-9507401).

## References

1. Balter M (1995) In Toulouse, the weather-and the science-are hot. *Science* 269:480-481
2. Eisen JS (1996) Zebrafish make a big splash. *Cell* 87:969-977
3. Travis J (1996) Gone Fishing! Scientists use mutant zebra fish to learn how vertebrate embryos develop. *Science News* 150:360-361
4. Westerfield M, Wegner J, Jegalian BG, DeRobertis EM, Püschel AW (1992) Specific activation of mammalian Hox promoters in mosaic transgenic zebrafish. *Genes Dev* 6:591-598
5. 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
6. Schulte-Merker S, Van Eeden FJM, Halpern ME, Kimmel CB, Nüsslein-Volhard C (1994) *no tail (ntl)* is the zebrafish homologue of the mouse *T* (*Brachyury*) gene. *Development* 120:109-1015
7. Driever W, Solnica-Krezel L, Schier AF, Neuhauss SC, Malicki J, Stemple DL, Stainier DY, Zwartkruis F, Abdelilah S, Rangini Z, Belak J, Boggs C (1996) genetic screen for mutations affecting embryogenesis in zebrafish. *Development* 123:37-46
8. Haffter P, Granato M, Brand M, Mullins MC, Hammerschmidt M, Kane DA, Odenthal J, van-Eeden FJ, Jiang YJ, Heisenberg CP, Kelsh RN, Furutani-Seiki M, Vogelsang E, 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. *Development* 123:1-36
9. Doerry E, Douglas S, Kirkpatrick T, Westerfield M (1997) Moving beyond html to create a multimedia database with user-centered design: A case study of a biological database. Tech. Report CIS-TR-97-02 Computer and Information. Computer and Information Science Dept. University of Oregon
10. Newman WM, Lamming MG (1995) Interactive System Design. Addison-Wesley, Reading, MA
11. Schuler D, Namioka A (1993) Participatory Design: Principles and Practices. Lawrence Erlbaum Association, Hillsdale, NJ
12. Gray P, Krishnarao G, Kulkarni G, Paton NW (1992) Object Oriented Databases: A Semantic Data Model Approach. Prentice-Hall, New York
13. Stonebraker M (1996) Object-Relational DBMSs The Next Great Wave. Morgan Kaufmann Publishers, San Francisco
14. Westerfield M (1995) The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (*Danio rerio*) 3rd. Univ. of Oregon Press, Eugene
15. Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF (1995) Stages of embryonic development of the zebrafish. *Dev Dyn* 203:253-310