Xenopus Community Recommendations for Future Resources and Goals (Summer 2006)

In March of 2000 and then again in October of 2003, the *Xenopus* Community met to discuss resource goals that would benefit their scientific efforts. Recently, in October of 2005, members of the *Xenopus* community met once more at the NIH to discuss the progress on these important goals and to determine future resources that would further benefit the community. This paper reports on the resources identified by the *Xenopus* community, which are necessary to improve *Xenopus* further as a non-mammalian model system. The report includes the progress that has been made, the consensus recommendations on additional resources needed, and proposals to best achieve those identified needs. The needed resources fall into ten general categories:

1. Genome Resources

Genome Sequencing and Assembly

The Department of Energy's Joint Genome Institute (JGI), acknowledging that *Xenopus* occupies a unique position among vertebrates, expressed interest in sequencing the *Xenopus tropicalis* genome. A white paper was submitted with strong Community support for a sequencing effort, and sequencing began in 2002. The JGI has carried out sequencing of the *X. tropicalis* genome using the DNA of a single female F6 inbred Nigerian frog. The NIH supported BAC end sequencing, and mapping is also used in the genome assembly. The importance of the genomic sequence cannot be overstated. It provides two critical components to *Xenopus* biology: **1)** a complete gene set for genomics and proteomics that will be especially powerful when coupled with the experimental strengths of *Xenopus* and **2)** genomic architecture that is critical for the cloning efforts in *Xenopus* genetics and the analysis of *cis*-regulatory elements using *Xenopus* transgenics. The genome sequence will also provide unique opportunities for biologists to examine genome evolution and function. For example, the evolutionary distance of *Xenopus* from mammals provides an opportunity to identify conserved non-coding elements in vertebrates, which, coupled with powerful transgenic techniques, offers the prospect of a high-throughput system for identifying novel gene *cis*-regulatory elements. Work of this kind will greatly improve annotation of vertebrate genomes by identifying large numbers of conserved regulators. From studying these elements one can anticipate important new insights about gene expression, and the phenotypes associated with mutations in them, which will be an extremely valuable resource for identifying genetic lesions in human disease.

Progress in genomic sequencing has been rapid and fruitful. The JGI sequencing effort has produced an outstanding resource for *Xenopus* researchers and the *Xenopus* Community applauds their crucial contribution to our field. The reports of the JGI Genome Steering committee provide update summaries and plans, and can be found at http://tropicalis.berkeley.edu/home/genomic_resources/genome_seq/genome.html. To summarize, the JGI released the latest version of the genome (version 4.1) in August 2005 that contains a total of 20 million reads compiled from a 3kb insert library, 8kb insert library, 40kb fosmid libraries, and some BAC library end sequencing. The latest assembly is about 8x genomic coverage containing 19,759 scaffolds representing 1.5 Gigabases of sequence. In version 3 (version 4 is slightly improved over this), the N50 is 275 pieces, with 550 scaffolds > 1 Mbase, and 1500 >100kb. (N50 size is based on the total number of nucleotides assembled into contigs; if one considers half the sequence, the scaffold at that position gives the N50 size. The number of scaffolds above that is the N50 number. For comparison this number in the chicken assembly is 37). Efforts are now underway to annotate the genome. There are approximately 28,000 predicted gene models, 55% of which show EST support. Shotgun sequencing is now completed, and efforts are underway to improve genome assembly with a final release planned for the Fall of 2006.

Genome Assembly Improvements

While the available draft genome is an exceptional resource, further improvements to genomic assembly will provide additional dividends. **The *Xenopus* Community has identified further improvements in genome assembly, and finishing of gene-rich regions, as their highest priority.** Given the extent of *X. tropicalis* genome sequencing, it is surprising that there is not an even better assembly. This is in part due to inherent difficulties posed by the genome structure, which may include unclonable regions, but certainly includes large amounts of *Xenopus* specific repeats, many of which are arranged in long stretches that confound assembly. The "easy" and gene-rich regions are well represented in scaffolds, while the sequences that the Stanford finishing group has found to be highly repetitive results in breaks in genomic contiguity. Such sequences include *Xenopus*-specific repeats of 2.5 kb or 5 kb although some repeats are longer than 35 kb fosmids. Frequently, in comparing the BAC sequences and the shotgun assembly, these regions are collapsed in the assembly and result in nearby gaps.

A necessary step in improving the genome will be to generate another assembly that incorporates current BAC end sequences, additional EST sequences, and mapping information. Since Bac End Sequencing had been minimal, efforts were focused on high-throughput Bac end-sequencing sequencing. However, while the JGI is doing another assembly in the summer of 2006, early results show that Bac end-sequencing has not provided much additional contiguity for the amount of sequencing performed partly due to failures of paired-end reads and poor BAC coverage.

The most serious deficiency in the genome assembly is the modest coverage in BAC sequencing (currently at 4x) which is very low compared to other assembled vertebrate genomes which often have >15x coverage. BAC genomic coverage is non-uniform; certain regions are better represented than others, but even a cursory inspection of the genome browser shows that much of the genome is inadequately covered by BACs. BAC coverage is an important genomic resource to the community, and libraries adequately covering the genome must be obtained.

BAC Genomic Libraries

Implicit in this proposal is the generation of additional BAC libraries since the current libraries are inadequate. The Community sought the advice of experts in the field to identify a method to produce an unbiased library of adequate depth. A sheared library

would be one possibility, but since the second segment CHORI library also appears unbiased, restriction digestion may also be effective. Therefore, the particular expertise of those generating a new BAC library and an understanding of previous problems will be important.

Two BAC libraries made from Nigerian *X. tropicalis* have been used so far: one made by the Institute for Systems Biology (ISB) and one made by the Children's Hospital Oakland Research Institute (CHORI). The ISB library has an average insert size of 75 kb while the CHORI library has a larger insert size of 175 kb. A third library has been made from the Adiopodoume strain of *X. tropicalis* with an insert size of 120 kb. The CHORI BAC library has two segments. The first segment has larger inserts but is believed to be biased, while the second segment, which has a slightly smaller average insert size of 118kb, appears to have a much better genomic representation. However, this second segment of the library is exhausted and no additional clones can be generated. The Adiopodoume BAC library is currently being evaluated, but at best represents only 6x coverage. **Therefore, BAC library coverage is currently inadequate for the community needs, and an additional BAC library needs to be generated.**

Genome Finishing Resources

Currently, approximately 500 BACs are in different stages of finishing including clones containing genes chosen by the *Xenopus* community, clones with ENCODE sequences, and clones that contain sequences that are at the ends of scaffolds. Stanford is finishing these, but progress has been difficult due to *Xenopus*-specific repeats and concerns with BAC library quality. Since BAC finishing has proven to be difficult, it is felt that high level finishing of the frog genome using current technology will be prohibitively expensive and delay much of the long-range and gap filling data that the Community needs now. Therefore, the Community discussed interim goals that would still satisfy many of the current needs. **Currently, the two most pressing needs for the *Xenopus* Community are 1) a complete gene set for genomic and proteomic analysis, and 2) long-range contiguity for genetics.**

To this end, the Community recommends that genome scaffolds should be mapped and oriented to chromosomes and missing gene-rich regions of the genome should be identified and finished by BAC sequencing. Ongoing annotation efforts will identify genomic regions where gene models are impacted because of assembly gaps. These regions will be identified by community annotators and placed in a BAC finishing queue. Also, BACs at the ends of scaffolds will be identified and need to be finished. Genes not present in the genome but identified by EST sequencing or by users also need to be identified and finished. Therefore since BAC finishing of the *Xenopus* genome is difficult, selective BAC finishing that will provide a high impact to important regions of the genome is a reasonable, obtainable goal and will provide a near-complete gene set for genomics and proteomics as well as long-range contiguity for genetics.

Resources will need to be identified to 1) generate adequate BAC libraries 2) screen BAC libraries for particular clones 3) sequence and finish these identified BAC clones. Other efforts will also improve the genomic assembly and should be pursued: **1)** Happy mapping may be a rapid and powerful method for identifying physical linkage of the genome fragments. However, this technology has not been tried in a vertebrate genome and so funds should be made available for a pilot effort to test feasibility. If successful, this approach will provide long-range contiguity of genomic scaffolds. **2)** FISH mapping could be applied to test specific hypotheses of linkage, particularly because synteny of frog and mammalian genomes has been confirmed for individual scaffolds, and this is likely to extend to longer range synteny. While FISH mapping is useful for limited mapping, it is not adaptable to high throughput analysis. **3)** Chromosome sorting and

sequencing or chromosome-specific HAPPY mapping may be useful and given the differences in chromosomes in *X. tropicalis* should in theory be straightforward. However, reproducible metaphase chromosomes are most easily prepared from cell lines. Although there are anecdotal reports of successes with *X. tropicalis* cell lines, none of these are widely available or well characterized. Resources should be identified to develop these important genomic tools.

Annotation

The Community is very interested in genome annotation, which allows the assembly to be useful in gene identification, and predictions of the proteome and transcriptome. The JGI has provided automated annotation, but optimal annotation requires manual curation. To properly annotate the genome, the *Xenopus* community will need to coordinate amongst its members to annotate regions of the genome in which there is expertise. In particular, a high priority will be to identify those genes that are missing in the genome so that efforts to fill these gaps can proceed without delay. An initial effort has been started at the JGI genome annotation Jamboree in April 2006, and an ideal venue and time for a substantive effort will be at the next *Xenopus* meeting, taking place in September 2006 in Japan. Subsequently, resources need to be identified for continued annotation.

Additional Amphibian Genomes

The value of the *X, tropicalis* genome has been substantial to the *Xenopus* community, and given the power of comparative genomics, sequencing additional amphibian genomes is likely to have as substantial a benefit as it has in other species. Given *X. tropicalis* as a reference genome, sequencing the allotetraploid *X. laevis* genome appears feasible and would greatly benefit many *Xenopus* researchers, especially proteomics work for *Xenopus* cell biologists. Results from Uffe Hellsten and Dan Rokhsar indicate that the genomic sequence from *X. laevis* can be deconvoluted into two genomes that are evolutionarily as distant as rat and mouse (Hellsten et al. submitted). These genomes can be aligned with *X. tropicalis,* thus simplifying the otherwise complex task of assembly of polymorphic sequence reads. Another interesting candidate would be the spade-foot toad, which has an extremely small genome for amphibians (0.9 GB). A toad genome would provide interesting evolutionary information, and the spade-foot toad develops a dimorphism during development as either a herbivore or carnivore which is of particular interest to evolutionary biologists.

2. Generation of ESTs and Full length cDNA collections

Progress on ESTs

The generation of both *Xenopus laevis* and *Xenopus tropicalis* ESTs has been extraordinary. With the support of the NIH and DOE/JGI in the U.S., the Sanger Center in the U.K., the NIBB in Japan, and the CNRS in France, *Xenopus tropicalis* is currently fifth in the table of ESTs at NCBI, with 1,044,182 ESTs listed on June 9, 2006 and *Xenopus laevis* is 14th with 525,247 ESTs listed. Many tissues have been sampled including embryonic and tadpole stages as well as adult tissues. These are listed at the IMAGE site.

The emphasis initially was on obtaining ESTs in *X. laevis*, as laid out in the March 2000 plan. Since that time, however, large efforts, first at the Sanger Centre and then by JGI, have led to a rapid increase in *X. tropicalis* ESTs. Due to the tremendous depth of sequencing, many putative full-length clones have been identified and sequencing has confirmed that they are full length. **This has prompted the Community to work aggressively toward a new goal, to generate what will become an extraordinary and unique resource for *Xenopus*: a genome-wide unique, full-length cDNA library (Unigene set) in expression ready vectors.** Given the power of functional assays and

expression cloning in *Xenopus*, the community felt strongly that this new goal is one of the highest priorities.

Issues of full-length

The ready availability of the sequenced clones through the IMAGE consortium and the HGMP (now Geneservice) has provided new molecular markers, and full length clones for functional analyses. Full length cDNAs are particularly important to the Community because they facilitate the generation of Unigene sets which can be powerful for functional assays, one of the particular strengths of *Xenopus*, and as a collection, make for highly efficient expression cloning assays. In providing a tool for biochemical experiments, the predicted full-length proteome is an essential prerequisite for analysis of mass spectroscopy data. At the time of the original proposal (2000), technologies for generating full-length cDNA libraries appeared promising. However, many of these technologies have not proven to be advantageous. Based on Bruce Blumberg's work, technologies to enrich for full length clones only marginally improve the fraction of clones that are full length and appear to lead to a loss of diversity in the library.

Therefore, informatics approaches have been designed and implemented to identify EST clones that have a high likelihood of being full length. Cluster analysis has been performed on all of the current EST sequences, and Mike Gilchrist at the Sanger/Wellcome Trust/Cancer Research UK Gurdon Institute and Lukas Wagner at the *Xenopus* Gene Collection (XGC) at the NCBI have identified a number of clones that are putatively full-length. Notably, a concerted effort has been made at the Gurdon Institute to assemble a collection of full-length cDNAs from *X. tropicalis* in expression vectors, suitable for functional assays of individual or pooled clones. In addition, a Unigene set of over 12,000 cDNAs and a smaller, full-length cDNA set for *X. laevis* are available from RZPD.

The priority for future EST and cDNA resources is to identify, in an expression-ready vector, a full length clone set for as large a fraction of the genes in the *X. tropicalis* genome as is feasible. Plans are also being made for how to distribute this Unigene library to researchers as sets, or subsets, of clones that can be used for expression studies. This goal is motivated by the powerful embryonic and cell biological assays that are available to exploit this resource. With this resource, *Xenopus* will be in a unique position for genomic and proteomic level analysis.

Vector Choice

Many of the ESTs have been sequenced from the CS expression vectors, and it makes good sense to maintain the practice of generating and sequencing cDNAs in this CMV and SP6 promoter-containing plasmid since these provide the most useful current platform for expression studies. pSport vectors have also been used for making cDNA libraries. These contain convenient Gateway technology, but compared to the CS vectors, they are less suitable for generating mRNA for use in expression screens and functional assays. Bioinformatic approaches have identified full-length clones in both pSport6 and in CS vectors. In many cases, putative full-length clones for a particular gene can be identified in the CS vectors. However, currently in approximately 1600 cases, a full-length candidate could only be identified in the pSport6 vector.

It was determined by the Community that new cDNA libraries (see below) would be made using the pCS expression vectors. For any current putative full-length cDNAs that exist in both pSport6 and pCS, the pCS clone will be chosen for full-length sequencing. For the 1600 clones that are only available in pSport6, full-length sequencing will be completed for these clones. If they are indeed full length, then they will be converted so that they are ready for expression analysis. Resources to convert these clones need to be identified.

Additional cDNA Libraries and EST Sequencing

Enormous progress has been made in EST and full length clone identification and sequencing, but additional cDNA library construction and continued sequencing is justified. Most of the current existing cDNA libraries focus on egg and early embryonic stages as well as some adult tissues. In order to maximize new cDNA discovery, additional cDNA libraries will be made from tadpole stages and adult tissues that have not been represented. In addition, libraries appear not to have been mined evenly to generate the maximum EST representation possible, and further sequencing of some libraries is needed to provide important depth as well.

An in depth bioinformatics analysis has been presented by Mike Gilchrist (Wellcome Trust Gurdon Institute) and Lukas Wagner (NCBI) which shows that oocyte, egg, and early embryonic stages have been intensely sequenced and further sequencing of these tissue libraries is unlikely to obtain adequate novel clones to justify costs. However, two existing *X. tropicalis* libraries (brain and testis) appear to continue to provide novel clones and deserve additional depth of sequencing.

Based on these analyses, the Community identified 11 tadpole stages and adult tissues that were not adequately represented and whose cDNA library construction and sequencing would likely identify additional novel clones. Several groups have already agreed to provide the necessary tissues. The JGI and NCI/NIH groups have agreed to make the libraries.

The 11 identified tissues for new cDNA library construction are:

1) St 25 embryos (*X. tropicalis*)
2,3,4) Metamorphic tissues (*X. laevis* and *X. tropicalis*):
- **1) Limb bud**
- **2) Tail**
- **3) Gut**

5) Bone (*X. laevis* and *X. tropicalis*)
6) Thymus (*X. laevis* and *X. tropicalis*)
7) Spleen (*X. laevis* and *X. tropicalis*)
8) Eye (*X. tropicalis*)
9) Olfactory bulb (*X. laevis*)
10) Pancreas (*X. laevis* and *X. tropicalis*)
11) Testis (*X. tropicalis*)

In addition, it was determined that Bruce Blumberg's *X. tropicalis* brain library should be sequenced further.

The exact division of labor for library making and EST sequencing is in advanced stages of planning. Discussions are now underway to organize the creation and sequencing of these libraries. It is anticipated that much of the additional sequencing will be completed by the fall of 2006.

The proposed plan is to sequence 20,000 ESTs from each library. At that point, they will be evaluated for diversity, novelty, and full-length clone representation and then decisions will be made about which libraries will be sequenced to greater depth.

Full-length Sequencing

A definitive plan for full-length clone identification and sequencing has been formulated, and this plan has been endorsed by the Genome sequencing steering committee. In summary, the NCBI and the Wellcome Trust Gurdon Institute have been collaborating on the identification of full-length clones from the EST databases. To maximize the usefulness of the full-length candidates, the picked and re-arrayed putative

full-length clones will be subjected to full-length sequencing. This is necessary to verify the quality of the cDNA, but also to characterize its splicing pattern. Currently there is no plan to collect all possible splice variants of genes, but it is important to know what splice pattern is represented in the full-length collection, and it is likely to be a valuable future goal to have more than one splice form represented in the Unigene collection.

Currently, full-length sequencing is most advanced for *X. laevis* with nearly 8,000 non-redundant full length sequences with plans to increase this to 10,000. An aggressive effort is now underway for *Xenopus tropicalis* which should be completed with a high priority, since there are likely to be more full-length clones already available in the large *X. tropicalis* EST set and because these can be aligned with the genome sequence to determine completeness.

The JGI, NCBI, Sanger Center, and Genoscope have all agreed to continue full-length cDNA clone sequencing efforts acknowledging the importance of this goal. A coordinated plan to use these available sequencing resources effectively is well underway.

Distribution of Clones and Libraries

In order to facilitate the distribution of the large number of cDNA clones generated, most clones (including all of the JGI clones) will be sent to the IMAGE distributors. These clones will be duplicated and a set sent to the RZPD in Europe, a set sent to a US distributor, and one set maintained at IMAGE. IMAGE agreed that they would provide clones to PIs who provided tissues, mRNA, or libraries at no cost.

On a related note, members of the Wellcome Trust Gurdon Institute have arranged for Geneservice to change their relatively restrictive clone re-distribution policy on *X. tropicalis* clones to bring it in line with the IMAGE policy.

The full length set and its subsequent additions and revisions would be distributed widely to many distributors since there will be considerable demand and interest in keeping costs to a minimum. It is expected that this set will change with time as additional full length clones are identified or generated and as errors are detected. Currently, these additions and corrections will simply be added to the library rather than by replacing clones. Users would purchase these “updates” at a cost considerably less than the full library.

Completing the Full Length Set

Continued sequencing of novel cDNA libraries will facilitate the generation of full length Unigene sets. However, if one looks at the experience from human and mouse, even with more than four times as many ESTs as *Xenopus*, some full-length clones have not been identified. In particular, these are transcripts expressed at very low levels or are long transcripts that are not easily isolated as full-length clones. Alternative methods will need to be identified and resources allocated for more directed methods to generate a more complete full length Unigene set. An obvious possibility is to exploit the recent genome assembly to predict full length clones that are not yet available, and to amplify these in a directed way from appropriate tissues or libraries, and to use genomic resources from organisms (human and mouse) where sequence information is more complete.

The Community strongly recommends that the ultimate goal of the *Xenopus* Unigene set is to have a full length clone representative of every identifiable gene in the *X. tropicalis* genome. This Genomic Unigene set will establish the core of the *Xenopus* transcriptome and proteome, and in conjunction with a genome-wide microarray (section 5), will allow researchers to exploit fully the tremendous advantages of *Xenopus* as a model organism.

3. *X. tropicalis* Genetics

In March of 2000, initial data on the feasibility of genetic manipulations in *X. tropicalis* prompted the Community to recommend further development of these powerful approaches for biological analysis. Therefore they recommended that pilot projects for *X. tropicalis* mutagenesis be funded that would define the conditions for generating mutants. These "trailblazer" projects would be charged with making their protocols quickly available to the Community. It was also anticipated that interesting mutants would be generated but not pursued by the respective PI. These mutants would be freely available to the Community and understood to be collaborations. Three RFA's were awarded to PIs: Richard Harland, Derek Stemple/Lyle Zimmerman, and Paul Mead. Enrique Amaya, Rob Grainger, and Frank Conlon who have also been generating *X. tropicalis* mutants, agreed to join this group and make protocols and reagents available.

During the recent October 2005 meeting, the mutagenesis and phenotyping group reported outstanding progress. Rob Grainger's group showed that ENU mutagenesis has been successful in identifying specific phenotypes in the eye and a number of other tissues during development, and that meganuclease offers a simple and easy method for transgenesis. Richard Harland's group showed that interesting mutants can be identified, propagated through multiple generations, and mechanisms such as changes in proliferation can be identified. In addition, the Mead, Amaya and Grainger labs also suggest insertional mutagenesis strategies may be effective and lead to rapid cloning of mutations. Results from Rebecca Heald's lab indicate that *X. tropicalis* egg extracts are effective for cell biology, biochemistry, and proteomic studies. Derek Stemple showed that sequencing PCR amplicons from animals generated from mutagenized sperm could be used to identify mutants in a reverse genetic "TILLING" approach. Sequencing is also being used to determine absolute mutagenesis rates, in order to determine optimal doses of mutagen. Lyle Zimmerman showed that early gynogenesis can be effective for screening and identifying mutants and that these mutants can be mapped using AFLP mapping and genomic sequencing.

The mutagenesis group also made the following recommendations that will help the larger community become effective in mutagenesis and phenotyping. First, **optimal mutagen doses must be identified**. As of 2005, an optimal dose for gamma-ray mutagenesis with approximately 1% hit rates for three different loci has been identified. Current ENU mutagenic rates are similar or better than mouse or zebrafish but a consensus about the optimal dose, and protocol for mutagenesis, is not yet established. Procedures for both sperm mutagenesis and spermatogonial mutagenesis are effective, but their relative advantages need to be investigated further. Methods must be standardized and mutation hit rates determined for varying doses of ENU. EMS may be better than ENU for mutagenesis of sperm suspensions and should be tested. Other agents including the alkylating agent chlorambucil and ICR191 (which causes frameshifts) are being used by Frank Conlon and should be evaluated.

Second, transgenesis has been a powerful strategy to understand *cis*-regulation in the frog. With the short generation time of *X. tropicalis*, transgenic animals can be propagated and used in genetic screens. Although the REMI method of Amaya and Kroll works well, simpler techniques to create transgenic animals have been tested and meganuclease appears to be a very simple and effective strategy as shown by the Pieler, Grainger and Amaya groups. It opens the possibility of insertional mutagenesis strategies, which as documented by Amaya, Grainger and Mead, can be very rapid ways to go from mutant to mutagenized gene. **The community felt that new transgenic methods and the establishment of transgenic lines are of significant benefit to mutagenesis efforts and phenotyping and should be aggressively pursued.**

Finally, **more genetic screens must be pursued and a frog genetics community should be encouraged**. ENU mutagenesis, gamma-ray mutagenesis, and inbreeding of wild-caught animals have all produced mutants. Mutants have been identified by haploid analysis, gynogenetic diploid screening, and standard F3 generational screens. The possibility for using TILLING for reverse genetics looks particularly promising as a tool for many researchers. The ability to generate mutations in known genes would rapidly facilitate the ability of many investigators to study cellular and developmental phenomena, and underlying regulatory pathways, with far greater rigor than is now possible (e.g. by morpholino treatments). In addition, the possibility of generating null mutations in this manner offers other powerful prospects. Because BAC transgenesis has now been shown to work effectively in *Xenopus*, it should be possible to rescue such nulls with BAC clones and study the regulation of a given gene by mutating specific sequences (both regulatory and coding) during such BAC "rescue" experiments. Therefore, the continued development of frog genetics is a high priority and additional resources are necessary to continue projects initiated by the RFA. The recent Program Announcement over the next three years (2006, 2007, and 2008) is well suited for mutagenesis and phenotyping applications. This should encourage new labs to initiate projects. The current mutagenesis group has regular meetings where progress and needs are discussed and ideas shared for improving the system. It remains essential that the mutagenesis groups meet biannually for discussion.

The short-term goal of these preliminary mutagenesis efforts will be to identify the conditions that are necessary to carry out efficient mutagenesis screens for specific phenotypes; success in such screens may justify larger scale screens that approach saturation. Thus, optimal mutagen and dosing must be identified, mutant founders generated, screening strategies and phenotyping carefully considered, and resources made available to perform a large scale screen. Such efforts are likely to produce many more mutants than the investigators who identified them can reasonably study and will surely lead to an expansion of the *Xenopus* community. It is likely that such screens would be done most efficiently in the context of a stock center, which can effectively distribute mutants (see below). Genetics in *Xenopus* will foster new models for human diseases, with the attendant insights and possibilities for direct benefits to human health.

4. Genetic Mapping

In the March 2000 meeting, the *Xenopus* community recommended that pilot projects be funded to test the feasibility of *Xenopus* genetics. Three proposals were funded and in the October 2003 meeting, forward genetic screens were shown to be feasible and effective in identifying mutants, and these mutants will improve our understanding of vertebrate biology. To facilitate cloning of mutants, the *Xenopus* community strongly recommended that a meiotic map be made (which would be greatly facilitated with the then available genomic sequence from the JGI).

The University of Houston and Baylor College of Medicine obtained funding from the NIH and are developing a microsatellite map with a goal of 5000 mapped markers. Their preliminary map (as of Jun. 2006) has 1050 markers in 10 large linkage groups and 7 small cluster groups at an LOD score of 10.0. Their preliminary efforts for meiotic mapping have focused primarily on tri- and tetra-nucleotide repeats because these microsatellites are less likely to produce artifactual polymorphisms due to polymerase stutter. However, these markers are now exhausted and additional mapping will rely on di-nucleotide repeats, which are plentiful throughout the genome. These efforts are now in production phase and additional markers are expected regularly. Validation of markers will be done by FISH, which will begin once 500 markers are mapped. Preliminary assessments of map size (approximately 1440 cM) show that it appears to

be of adequate size to allow for mapping of mutants. **The Community strongly encourages continuation of these efforts that will provide immediate dividends for mapping mutants and long-range genome assembly and validation. Their continued support is needed.**

Looking forward, the highest density maps will be single nucleotide polymorphisms (SNPs), which will provide the highest resolution for mapping purposes. Therefore, identifying resources to sequence DNA from separate strains of frogs is needed to identify these SNPs. The JGI Genome Project used DNA from a Nigerian strain individual. N. Pollet's group has created a BAC library from an Ivory Coast strain and the JGI has agreed to BAC end sequence five plates of the library to assess its quality and use for SNP identification. In addition, non-Nigerian strains will be used to generate new cDNA libraries for the ongoing full-length cDNA identification efforts (**section 2**). This may also provide valuable SNP data and should provide sufficient information to identify which strain of *X. tropicalis* has the highest density of SNPs relative to Nigerian and therefore should be sequenced with low coverage for genome-wide identification. With this SNP data in hand, resources will need to be identified to create a SNP-chip and a snip-SNP map. SNP-chips are microarrays that can provide thousands of genotypes concurrently and are extremely powerful and rapid for genetic mapping. Snip-SNPs are SNPs that introduce a restriction site and therefore simply the polymorphism assay. These tools would greatly complement the microsatellite map and provide important tools for genetic mapping as well as provide validation for the genomic assembly. A JGI Community Sequencing Project proposal has been submitted requesting low-level sequencing of known polymorphic strains (Population A and TGA). While the proposal received a high score for scientific merit, the Department of Energy has decided to focus exclusively on model systems directly relevant to the DOE mission. Therefore, resources to identify SNPs in polymorphic strains must still be identified. This is a critical resource for future genetic mapping efforts.

5. Physical Map

In October 2003, the Genome Sequencing Center (GSC) at the Washington University School of Medicine initiated a BAC end sequencing project as well as BAC fingerprinting to generate a physical map of the *X. tropicalis* genome (see sequencing reports and http://www.genome.wustl.edu/genome_group.cgi?GROUP=2). Physical maps have played critical roles in the validation and assembly of other genomes, and are especially important for long-range contiguity. Such a map is anticipated to be equally important for *Xenopus*.

However, the latest physical mapping efforts have not generated the long-range contiguity that had been hoped due to the limitations in the BAC libraries available. As a consequence, the current physical map has ~8000 contigs and is of limited value. Since the primary limitation appears to be BAC library quality, additional BAC libraries must be generated that provide adequate coverage of the genome. Physical mapping along with genetic mapping will provide important long range contiguity that will greatly improve the quality of the genomic sequence.

The Community strongly recommends that another BAC library be made and additional physical mapping be performed. Because the GSC has transitioned to a capillary based method for fingerprinting which is incompatible with agarose based methods, they will not be able to generate additional fingerprints. However, physical mapping will provide additional contiguity and an alternative center must be identified to generate a better physical map once a new high-quality BAC library is validated.

The Community sought advice from experts who unanimously suggested that 15x coverage would be a minimum goal. Since 1/3 of BAC end sequences are likely to fall

into repeats and therefore not be informative, **the Community requests that a new BAC library be made, end sequenced, and fingerprinted with a target of 25x coverage.** The long-term goals are to 1) validate the genomic assembly and provide long-range contiguity, and 2) determine and make available a BAC-tiling path along the genome.

6. Microarrays

Microarray technology offers the possibility to analyze a very large set of genes in a single experiment. It has proven to be extremely effective in identifying changes in gene expression and gene regulatory pathways in human, mouse, and in *Xenopus*. The Community recognizes this potential, and a few efforts have been made to create *Xenopus* chips. However, the generation of effective microarrays requires significant investments in personnel and hardware. **The goal of the Community is to have widely available and reasonably affordable microarrays containing all available *X. laevis* genes and the entire *X. tropicalis* gene set.**

The Community in March 2000 recommended that a publicly funded project to create relatively inexpensive microarrays should be an important priority, and reiterated that support in 2003. A community based and subsidized operation at the academic level has not come to fruition. A few academic labs, most prominently Ken Cho's group, Christof Niehrs' group and Naoto Ueno's group have chips that are available on a collaborative basis. The Amaya group has also developed a collection of longer oligonucleotides paralleling the effort to identify full length cDNAs by the Cambridge group. Also, several commercially available chips have been and are being made. The status of these is outlined here.

A good quality chip for *X. laevis* was created by Affymetrix with NIH funding. This has been quite successful for the Community and recently Affymetrix has agreed to generate, at their own cost, two microarrays, one for *X. laevis* and one for *X. tropicalis*. There are currently about 28,000 putative *X. tropicalis* genes (based on the latest genome draft) and about 15,000 unique *X. laevis* genes (based on EST sequences). Steve Klein (NIH) has assembled a group of volunteers from the *Xenopus* community to serve as advisors in the design of this latest version. It is hoped that the new chips will be available by the end of 2006. It is worth stating again that the Community recognizes and commends the commitment of both the NIH and Affymetrix to this endeavor.

In addition, the JGI has contracted with Nimblegen to create a genome-wide *X. tropicalis* gene array. This "gene chip" is being used to determine the gene expression profile at various stages of development and will validate gene models for which there is currently no EST evidence. Once completed, this array design will be available to the community in addition to the data generated from these pilot experiments. In addition, with the availability of the *X. tropicalis* genome sequence, it would be possible for Nimblegen to create a *X. tropicalis* genome tiling array. Genomic information will also be useful for other kinds of microarrays, such as splice site microarrays or promoter microarrays, and these should be further considered now that a high quality draft genome is available.

Therefore, the community will have choices on available microarrays although the technology remains expensive. The importance of this technology is not questioned, but the best way to design and make a microarray widely available to the community should be readdressed. Recently the NIH produced a Program Announcement soliciting applications for genomic and genetic analysis in *Xenopus* for each of the next three years (2006, 2007, and 2008). NIH staff considered this PA to be an ideal opportunity for interested investigators to develop microarray technologies and applications involving microarrays.

7. Database

An internet database has been identified by the Community as a pressing need. It should provide a central repository of information that will greatly facilitate information exchange. A great deal of information is being generated daily by the *Xenopus* community but lacks an efficient portal for effective transmission.

The Xenbase website, maintained by Peter Vize, has received NIH funds to expand its role as the central internet source for information on *Xenopus* research. Xenbase will be reorganized using the GMOD (Generic Model Organism Database) modular schema called Chado. Chado is a set of modules for building a relational database (http://www.gmod.org/schema/index.shtml).

The existing Chado modules are:

- sequence - for sequences/features
- cv - for controlled-vocabs/ontologies
- general - dbxrefs
- organism - taxonomic data
- pub - publication and references
- companalysis - augments sequence module with computational analysis data
- map - non-sequence maps
- genetic - genetic and phenotypic data
- expression - gene expression

For Xenbase, an expression data set will be compiled that can be queried. In order to effectively use this queried database, Peter Vize and Nicolas Pollet's labs are developing an ontology adapted from zfin and Hausen's anatomy book to identify expression patterns effectively. Microarray data will be available. Functional data will be available for a wide variety of genes. The Community had several requests that should be easily implemented including a database of *Xenopus* protocols, integration of all of the EST databases, fate maps, control elements for transcription, and data regarding commercial antibodies that do or do not work in *Xenopus*. Finally the database will allow for users to login and provide data and annotation.

While the database will likely expand with time, initially it will not be able to encompass genome sequencing and annotation. However, the JGI is committed to providing a well-annotated genome assembly that can be handed off to other genome curation sites.

Currently the funding for Xenbase is not sufficient to support a staff of curators that would be needed to keep the database frequently updated. However, once the current database is implemented, and given the quantity and quality of genomic and proteomic data that is being generated, the database should be in a position to expand and become the crucial central repository of *Xenopus* data.

8. Stock Center

The Community identified a US stock center as a very high priority goal. Investigators have generated a large number of extremely useful transgenic and inbred polymorphic lines. These lines are difficult to distribute and maintain on a large scale. Additionally the transfer of animals across national lines is also particularly difficult. A stock center can expedite all of these issues and greatly increase the availability of these lines to investigators. A stock center would also be able to optimize husbandry conditions, which will prove to be invaluable in increasing scientific productivity. Importantly, it could provide facilities for coordinated and collaborative genetic screens: providing a place to

house animals for screening and for investigators to visit to learn methodologies and perform screens.

A European stock center has just recently been approved for funding by the Welcome Trust, and will be managed by Matt Guille and Liz Jones at Portsmouth, England. The UK will define its stock center broadly to include animal stocks as well as molecular reagents. The animal stock center will have the capacity to care for both *X. laevis* as well as *X. tropicalis*. 4000 animals can be held which translates to approximately 200 lines. They will develop the facilities to store sperm so a much larger number of lines could potentially be held as frozen stocks. Finally the stock center will also carefully investigate optimizing husbandry conditions for *Xenopus* species.

Some progress in the United States has been made towards creating a *Xenopus* stock center, with one candidate location being the Marine Biological Laboratory (MBL) at Woods Hole, site of the 2004 *Xenopus* meeting, and a location that is frequented by many *Xenopus* researchers. A second possibility is a Stock Center at Cincinnati, under the direction of Chris Wylie. The community in general felt that the stock center at a place like Woods Hole would have distinct advantages since it would be easy to make it the *Xenopus* training center (because of lodging, meeting facilities, etc.) and because of connections to existing courses (e.g. the Embryology Course).

9. Training Center

Training centers to develop skilled *Xenopus* researchers remains a priority for the *Xenopus* Community. The Cold Spring Harbor course regularly gets 30-50 applicants per year for 16 places. The UK stock center intends to incorporate a course as well into the stock center efforts. Also there was general consensus that the stock centers should make provisions to teach husbandry skills to investigators. The Community felt that it would be ideal to create a *Xenopus* Stock Center (**section 8**) that could also serve as a training center and that continued training of *Xenopus* researchers is essential.

10. Transitional Funding

The March 2000 committee recognized that *X. laevis* investigators might wish to pursue *X. tropicalis* for particular studies. Since *X. tropicalis* does require some differences in its care and use, resources were requested to facilitate this transition. The NIH then offered a package of $25,000 for this transitional funding. **The Community strongly recommends that this offer of funds continues in order to foster the growth of *X. tropicalis* genetics and genomics.** Pilot projects have just recently shown that genetic screens in *X. tropicalis* genome are feasible. Also many investigators have reported that polymorphisms are likely responsible for variable effects in the results of morpholino oligonucleotide experiments. *X. tropicalis* has been inbred so as to minimize polymorphisms. Therefore the Community feels that more and more investigators will find these funds extremely helpful to take advantage of *Xenopus tropicalis*. Finally, for investigators to take full advantage of the *Xenopus* genomics and genetics PA they will require preliminary data to validate their approaches. Without a mechanism for entry of new investigators into these areas, the full impact of the PA may not be achieved.

Acknowledgments from the Community

The *Xenopus* community greatly appreciates and acknowledges the efforts of the NIH, XGC@NCI, Sanger/Wellcome Trust, Genescope, and the JGI/DOE for its efforts to provide resources for investigation of *Xenopus* biology. In addition, Steve Klein has been instrumental in bringing the Community together to

coordinate planning. There is wide consensus that these efforts have revolutionized the study of *Xenopus* biology.

Robert Grainger, University of Virginia, rmg9p@virginia.edu
Richard Harland, University of California, Berkeley, Harland@berkeley.edu
Tim Grammer, University of California, Berkeley
Mustafa Khokha, Yale University and University of California