A white paper for sequencing the genome of a living fossil: the coelacanth, *Latimeria chalumnae*

Chris T. Amemiya, Benaroya Research Institute at Virginia Mason and University of Washington, Seattle, WA (camemiya@vmresearch.org)

Eric S. Lander, Broad Institute of MIT and Harvard University, Cambridge, MA

Richard M. Myers, Department of Genetics, Stanford University School of Medicine, Stanford, CA

Introduction

The discovery of the living coelacanth, *Latimeria chalumnae*, by Marjorie Courtenay-Latimer off the coast of South Africa in 1938 [1], was one of the most important biological discoveries of the twentieth century. *L. chalumnae* is an extant member of an ancient group of lobe-finned fishes previously known only from fossils and believed to have been extinct since the Late Cretaceous, about 70 million years ago [1,2]. The discovery of a second coelacanth species in Indonesia in 1998, *L. menadoensis* [3-5], was equally surprising, but for other reasons. Both discoveries were met with fanfare, intrigue, disbelief, skepticism, political shenanigans, and intense scrutiny [2,6,7].

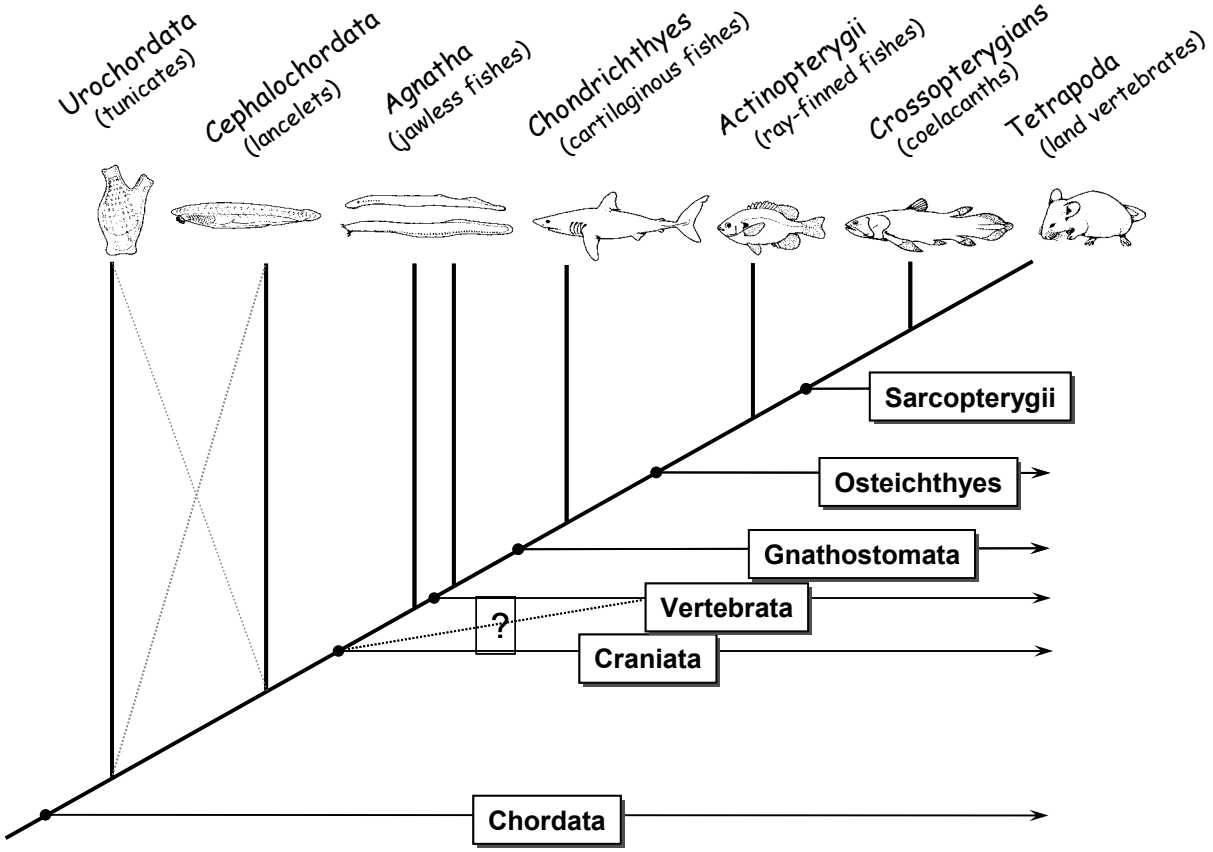

Figure 1. Abbreviated phylogeny of the chordates. The coelacanth lineage is shown in a sister-group relationship with the tetrapods. Note, the lungfish lineage is not included here since its genome is very large and not tractable to a genome sequencing effort. The coelacanth is important because it represents an early divergence of the sarcopterygians and can inform relative to the tetrapods.

Part of the reason why the discoveries drew such fascination is because *Latimeria* is a large, prehistoric-looking creature that has changed very little over evolutionary time, and because of the ramifications involving our own evolutionary history, something in which we are inherently interested [7,8]. Appendix 1 lists several salient facts regarding the general biology of coelacanths. We have also written an up-to-date review of all molecular biological studies concerning *Latimeria* [9]. It is important to note that coelacanths are extremely informative from a phylogenetic standpoint since they represent an ancient lineage (along with the lungfish) that serves as an outgroup to the land vertebrates, and realistically the only such taxon for which a genome sequence can be obtained.

Which species should be sequenced?

For the purposes of comparative genomics and informing other vertebrate genomes, the coelacanth species chosen is not really relevant and either would suffice. Thus the selection needs to be made on the basis of availability and potential impact. We submit that the species that should be sequenced is the African coelacanth, *Latimeria chalumnae*, as it is more easily accessible and far more data have been procured regarding its biology. While both species are protected under the Convention on International Trade of Endangered Species (CITES), the Indonesian species is more difficult to procure [8]. Insofar as the African coelacanth, an international conservation consortium, the African Coelacanth Ecosystem Programme, has recently been established whose mission includes procurement of specimens/tissues specifically for scientific investigations (see letters by Dorrington and Ribbink and discussion of resources below). Thus, it is possible to obtain materials from the African coelacanth for a genome sequencing effort, including RNAs for cDNA/EST sequencing.

Summary of available resources

Very few genomic resources are currently available for study. This has largely been due to the difficulty in obtaining material that was not substantially degraded since most of the coelacanth specimens from which these tissues were taken have not been alive (or had been severely stressed and very close to death when sampled). The vast majority of the tissue samples collected to date have not been sufficient for isolation of RNA and only a few have been adequate for isolation of library-quality genomic DNA. Consequently, most of the molecular studies to date have utilized PCR-amplified material for analysis [10-16]. This is especially true for all the African coelacanth specimens.

For the Indonesian coelacanth, the specimen was collected alive in a trawl by a local Sulawesi fisherman and brought to the attention of Dr. Mark Erdmann. Erdmann tried to revive the specimen by bringing it to deeper water and having divers swim with it (see cover page), however, it was clear that the specimen was going to expire and he was able to collect a small amount of fresh tissue for preservation in liquid nitrogen, prior to perfusion of the specimen with formalin and deposition in the Indonesian Institute of Sciences. We were able to procure the frozen heart tissue (~4 g) from Erdmann and managed to construct an arrayed BAC library (roughly 7-8X coverage) whose average insert size was 170 kb [17]. This BAC library is valuable for many applications and we have several ongoing projects on various genes and gene families utilizing this resource [18-20]. In addition to the arrayed library, 900,000 unarrayed BAC clones whose average insert size is > 150 kb, were archived at –80 C. We also have DNA from this specimen (100 μg) which could be ostensibly utilized for smaller-insert (shotgun) libraries.

Importantly, a not-for-profit international organization, the African Coelacanth Ecosystem Programme, ACEP (http://acep.co.za/), has recently been established for the purpose of promoting all aspects of the African coelacanth, most notably conservation, outreach/education, and science (see letters by Ribbink and Dorrington). As part of their science mission, ACEP has established a Genome Resources group under the direction of Drs. Rosemary Dorrington and Gregory Blatch (Rhodes University). Dr. Dorrington is establishing an infrastructure for being able to collect tissues from accidental coelacanth catches on the African coast (see letter by Dorrington). This includes distribution of tissue “kits” for long-term preservation of nucleic

acids without the need for refrigeration[1]. Dr. Dorrington has sent over to our (CTA's) lab a sample of blood from a catch off the Comoro Islands that was preserved in this way. We analyzed the blood cells microscopically, determined the genome size using flow cytometry (Fig. 2A), and prepared high molecular weight DNA using agarose embedding [21]. The DNA was subsequently analyzed for quality via pulsed field gel electrophoresis and by partial restriction digestion with *Eco*RI to judge its suitability for BAC cloning. These results are shown in Fig. 2B,C. We subsequently used this DNA to generate a small number of BAC clones to show proof-of-principle that the DNA obtained in this way was of sufficient quality for BAC cloning (not shown) and, therefore, suitable for WGS plasmid libraries. Importantly, we have sufficient amounts (~ 200 μg) of this African coelacanth DNA specimen for a sequencing project. This amount should suffice for all the plasmid and fosmid libraries necessary for the project, although probably not enough if a good-quality BAC library would need to be constructed as well. Alternatively, it is entirely possible that more accidentally-caught specimens will be obtained along the East African coast in the immediate future[2].

Insofar as RNA samples for a parallel EST project, select tissues (liver, gill, blood) have been preserved from two Comoran specimens in "RNA later" and stored at –80 C. Examination of total RNA extracted from these samples suggests that the preservation method was successful in that agarose gel electrophoresis showed RNA smears and not completely degraded samples (Dorrington, unpubl.). Synthesis of cDNA samples from these RNAs has not been attempted as yet. As part of the ACEP mission to develop genome resources, Comoran collaborators are equipped and have the capacity to collect additional tissue samples including gill, blood, nervous and reproductive tissue should any specimens become available through accidental catches by the local fishermen. Kits may also be distributed to collaborators along the East African coast. Lastly, the non-destructive collection of scale (skin) tissue has been done previously [22] (also see letters by Schartl and Dorrington) and is one of the primary objectives of the next ACEP research cruise off the South African coast for the purpose of establishing long-term cell cutures. Cell cultures had previously been attempted with some success, although long-term cultures were not established (Dorrington, unpubl.).

What is the genome size and degree of polymorphism?

The genome size of the African coelacanth has been reported to be 2.8 – 6.6 pg per 1C genome using feulgen densitometry of erythrocytes [23]. In order to get a more accurate estimate we employed flow cytometry of preserved erythrocytes from a Comoran *Latimeria chalumnae* specimen (see above). Visual (microscopic) inspection of blood smears indicated that the erythrocytes were largely intact, with a small (but noticeable) degree of degradation. Flow cytometry analysis was conducted using propidium iodide-stained erythrocytes and chicken nuclei as internal controls. Four separate replicates were conducted, each examining over 25,000 events. A representative output is given in Fig. 2A. By comparison to the chicken standard (2.33 pg/2C) we estimate the *Latimeria chalumnae* genome size to be around 5.5 pg/2C, or 2.75 pg per 1C genome. Due to the superior method of genome size estimation and

[1] It is not really feasible to have a scientist "on call" to be sent to a site where a coelacanth has been collected since these sites are often in extremely remote rural areas that cannot be easily accessed.

[2] Reports of accidental catches are apparently on the rise. In Tanzania, for example, 29 coelacanths have been captured since September 2003 (Ribbink, unpubl.):
http://www.planetark.com/dailynewsstory.cfm/newsid/36941/newsDate/22-Jun-2006/story.htm
http://www.planetark.com/dailynewsstory.cfm/newsid/28150/newsDate/15-Nov-2004/story.htm
http://www.alertnet.org/thenews/newsdesk/L21826592.htm

considerably larger sample size we trust these estimates much more than the previous estimates.

The degree of polymorphism is difficult to establish due to limited population sampling. The one study where this issue has been addressed [22] examined 47 specimens from eight populations from South Africa, Madagascar, Mozambique, Comoros, Tanzania and Kenya (range > 3500 km). Studies of both mitochondrial DNA (SNPs) and nuclear DNA (microsatellites) indicated that population subdivision was low or nonexistent and that the level of individual heterozygosity was very low across the range. This may be a function of the life history of coelacanth (low fecundity, long gestation, slow moving, etc.) as well as the geographic events that have shaped the African coastline in the western Indian Ocean. Interestingly, the Indonesian coelacanth, while considered a different species, is not that different genetically and all of the microsatellites used for the above-cited population study (of the African coelacanth) were obtained from BAC sequences from the Indonesian coelacanth (also see letter by Schartl). The low level of polymorphism seen in the African coelacanth is in keeping with what we have seen in our very limited genome sequencing of the Indonesian coelacanth. Of the > 2 MB of finished sequence that we've generated, 93 KB represent overlaps of BAC ends (from presumably the other allele) and the number of SNPs detected between alleles was 5 (0.005%)[3]; we did not observe any indels. Taken together, the results suggest that, for whatever reason, coelacanths are quite genetically homogeneous and the polymorphism problems encountered with such organisms as *Ciona*, *Fugu*, amphioxus and zebrafish, will likely not be as major an issue[4]. Should it be mandated, a more thorough assessment of the level of polymorphism will be undertaken.

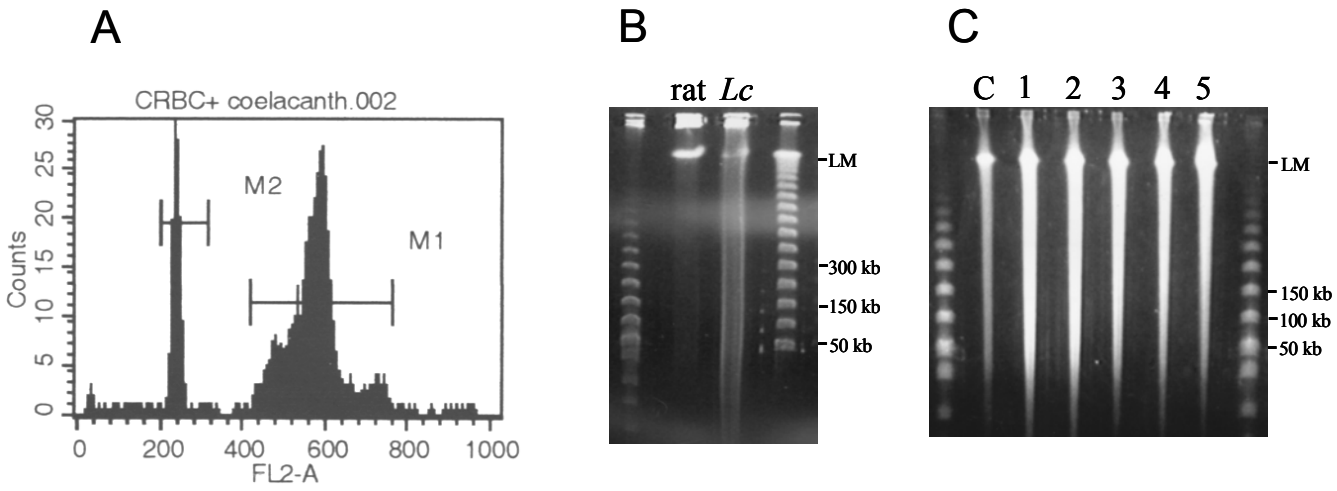

Figure 2. Assessment of the quality of preserved coelacanth blood and its genomic DNA. A blood sample that had been preserved from an accidental catch in the Comoros was analyzed by flow cytometry (A) and embedded in agarose for preparation of high molecular weight DNA (B,C). (A) Flow cytometric analysis of *L. chalumnae* blood sample. Coelacanth cells were washed in PBS (with 50 mM EDTA), combined with chicken red blood cell nuclei, stained in propidium iodide, and analyzed by flow cytometry. Several thousand cells were analyzed in four separate experimental runs. The left peak represents the chicken red blood cells (2.33 pg per 2C nucleus) and the right peak represents the coelacanth sample (primarily erythrocytes). Number of events counted is given on the left (x 1000). Based on these results we can conclude that the cells that we received from the coelacanth were of good quality

[3] This includes a number of gene families, including *Hox* and immunoglobulins, and may not be completely representative of the genome.
[4] The low level of polymorphism should result in longer contigs and would be advantageous from a fiscal standpoint as well.

(i.e., not overly hemolysed) and that the estimated genome size is ~ 5.5 pg/2C. (B,C) Analysis of coelacanth genomic DNA prepared from the same specimen as in (A). (B) Agarose-embedded *Latimeria* genomic DNA was run on a pulsed field gel along with a similarly prepared sample from brown Norway rat. The rat DNA is of good quality and had been previously used to generate a BAC library as part of the NHGRI BAC Resources program (VMRC11, http://bacpac.chori.org/library.php?id=270). As can be seen, some degradation was evident in the coelacanth (*Lc*) sample, however, the majority of the DNA was still in the well (i.e., was of very high molecular weight). (C) Agarose-embedded DNA was subjected to an *Eco*RI-*Eco*RI methylase competition reaction prior to electrophoresing. In this experiment, DNA was partially digested with a standard amount of *Eco*RI and increasing amounts of methylase (tracks 1-5). Track C represents an untreated control sample. This experiment showed that the DNA was sensitive to competition by *Eco*RI methylase, which blocks available *Eco*RI sites (note the increased amount of DNA in the limiting mobility (LM) band in track 5). We conclude from these experiments that the DNA is of good integrity for cloning and that the method of blood preservation used by Dr. Dorrington is adequate.

What is reasonable coverage?

We propose a genome coverage of 7X for the African coelacanth. We feel that the requirement for full coverage is justified in that this is the only genome representing the basal sarcopterygian node within the tree of life (see Fig. 1). WGS sequencing will be accomplished in the usual manner using a cadre of different-sized templates as has been done for other large genome sequencing projects conducted at the Broad Institute (e.g., mouse, short-tailed opossum, dog, stickleback). The resulting genome sequence will be of sufficient contiguity and quality for preliminary analyses and gene discovery by comparative methods. To improve on the utility of the sequence for these and other types of analyses, we advocate some level of automated finishing ("genome improvement") on selected regions of the genome. Fosmid (and or BAC) clones would serve as the templates for the necessary sequencing reactions. The methods, computational tools, and laboratory pipelines for automated genome improvement are already in place at Stanford Human Genome Center.

EST sequencing

In order to aid in gene discovery and annotation we recommend sequencing ESTs. At present, the appropriate cDNA libraries are not available, however, efforts are in place to procure fresh tissues for mRNA isolation and cDNA library construction (see letter by Dorrington). Some tissues are already archived.

Size of research community

The size of the current community is small and difficult to estimate, but probably consists of less than 20 active laboratories. Moreover, most of these labs have only published sparingly on the coelacanth and the majority of the papers in the last fifteen years have concerned molecular systematic investigations based on small or incomplete datasets. This level of activity is in contrast to that of 1950s and '60s when there were labs fully dedicated to the study of coelacanth anatomy and zoology [2,6,7]. This drop-off in coelacanth research activity has primarily been due to the lack of study materials and resources rather than lack of interest in this living fossil [8]. Indeed, the recent generation of one resource, a BAC library from the Indonesian coelacanth [17], has led directly to investigations in several areas: protocadherins [18], *Hoxc8* early enhancer regulation [19], discovery of novel Hox-cluster genes [24-26], characterization of the *Hsp70* gene family (G. Blach, unpubl.), characterization of the

mesotocin-vasotocin locus (B. Venkatesh, unpubl.), comparative genomics of the *Hox* and immunoglobulin heavy-chain loci [20] (C. Amemiya, unpubl.), and characterization of *Otx2* and its *cis*-regulatory elements [S. Aizawa, ms. accepted in PNAS). In addition, the sequence data generated by some of the above projects contributed to a population genetic study of African coelacanths [22] (see letter by Schartl) and to the identification of ancient SINE elements in the coelacanth genome that had been retained in high copy in the coelacanth genome and became largely lost but coopted (exapted) for novel functions in tetrapod genomes [27,28] (see letter by Bejerano). The recent paleontological discovery of a new transitional species, *Tiktaalik roseae* [29-31], will surely pique interest in the coelacanth even more and it is expected that researchers studying the evo-devo of aquatic-terrestrial adaptations will utilize the *Latimeria* genome sequence for making biological inferences (see letters by Tabin and Shubin).

What are unique aspects of the coelacanth that justify a whole genome sequence?

Phylogenetic position

The lobe-finned fishes are widely regarded as the group that gave rise to the tetrapods [6,15,32]. There are three issues directly relevant to its phylogenetic position for which the genome sequence will be important: (1) annotation and informing other genomes; (2) higher-level systematics; and (3) evo-devo and better understanding of the adaptations involved in aquatic to terrestrial habitats (e.g., fin-limb transition).

1. Annotating and informing other genomes

The coelacanth genome sequence would greatly enhance efforts to annotate the human genome. Currently, the genomes of teleost species are used as the most distant references to identify human functional elements [33,34]. However, teleost genomes are considerably derived relative to the common ancestor of teleosts and tetrapods due to the teleost whole-genome duplication [35-41] (see letters by Noonan and Postlethwait). There is substantial evidence that coelacanth genomes have not experienced such an event [17,18,42] [Amemiya, unpubl.]. In addition, comparison of coelacanth, teleost and tetrapod genomes will identify functional sequences unique to each lineage. Specifically, deep analysis of tetrapod genomes that use coelacanth genome sequence as an outgroup will reveal tetrapod-specific features, as these will be absent in coelacanth. Teleost genomes are too divergent for this purpose, due to whole-genome duplication and the genomic diversity of teleost species.

2. Higher-level systematics

The question as to whether lungfish or coelacanth is more closely related to tetrapods has been addressed by several groups using small or incomplete molecular datasets, the majority of which concluded that the lungfish was more closely related to the tetrapods [10,14,15]. Analysis of whole mitochondrial sequences, however, have not been able to satisfactorily resolve the trichotomy [43] and Takezaki *et al.* [12], using the largest nuclear dataset available, were likewise unable to resolve the interrelationships (see letter by Takezaki). The narrow window of divergence of the three higher taxa and their long branch lengths confound current molecular datasets. Their solution, which has been subjected to computational modeling and statistical rigor, is to examine a larger number of gene sequences from the respective taxa. The same approach has been used recently to examine the higher level interrelationships of deuterostomes [44] with the conclusion that urochordates (sea squirts) and not the traditionally held amphioxus (lancelets), are more closely related to vertebrates.

3. Evo-devo

Much of the interest on *Latimeria* has focused on evolution and development of its unusual morphology, which includes fleshy fins (that somewhat resemble primordial limbs), a hollow nerve cord, poor ossification of skeleton yet presence of a rigid notochord that persists throughout its lifetime, lack of defined ribs, and a unique bi-lobate caudal fin region, the structure of which has been maintained in coelacanths since the middle Devonian [2,6,45,46]. The genome sequence will allow characterization of developmentally important genes that could be involved in evo-devo of these structures.

Interrogation of these genes can be done in a surrogate biological system in order to deduce function. For example, coelacanth enhancer sequences for *Otx2* (transcription factor necessary for brain development) and *Hoxc8* (transcription factor necessary for axial development) have been studied in a mouse transient transgenic system [19] [Aizawa, in press] (see letter by Aizawa). Moreover, the opsin system in coelacanths has been interrogated using an *in vitro* assay in order to make inferences with regard to the molecular coevolution of visual perception with its restricted visual environment [13,47] (see letter by Yokoyama).

Lastly, many paleontologists and neontologists alike are very interested in the coelacanth's possession of lobe-fins and the genetic regulatory mechanisms controlling their development [31,45,48]. Fish and tetrapods differ in their proterygium vs. metapterygium component and consequent dermal versus endochondral bone component of appendages [49]. The coelacanth is the most ancient extant lineage containing the metapterygium as primary component of the fin-limb (as all future tetrapods would have). The genomic sequence will certainly contribute to creative investigations on this front (see letters by Shubin and Tabin), as well as to many other interesting questions regarding early adaptations in the transition to life on land.

Decreased molecular evolutionary rate

The complete genome sequence (609 KB) of the coelacanth protocadherin cluster has recently been determined and compared to the homologous human and zebrafish protocadherin clusters [18]. In gene number and organization, human and coelacanth protocadherin clusters are similar, while zebrafish has two highly divergent clusters, arising from the teleost whole-genome duplication. The protocadherin cluster is a tandem gene array prone to gene duplication, loss, and gene conversion. These processes are much less pronounced in coelacanth protocadherins than in protocadherins of mammals and especially zebrafish, indicating the coelacanth genome may not have undergone major rearrangement events since the divergence of coelacanths from other vertebrates.

This idea is further supported by analysis of 33 coelacanth *Hox* genes from the Indonesian coelacanth, that indicated that 32 of these genes have orthologs in the four mammalian *HOX* clusters [42]. This suggests that the organization of coelacanth *Hox* genes is similar to that in mammals, and unlike that in zebrafish and other teleosts, which have six or seven HOX clusters as a consequence of the teleost whole-genome duplication [35-37,39,41,50,51]. In order to corroborate these inferences we have cloned and sequenced all four of the coelacanth HOX clusters [17], and shown that these clusters are very similar in content and organization to that of mammals [Amemiya, unpubl.].

It is notable that relative rate tests on the protocadherins [18], *Hox* genes [Amemiya, unpubl.] and a few other nuclear genes [10], confirm that the overall rate of molecular evolution in *Latimeria* is considerably slower than in tetrapods and teleost fishes. In the case of the HOX

clusters this goes for both coding and conserved noncoding sequences [Prohaska, Amemiya, unpubl.].

Retention of conserved noncoding elements

The coelacanth should be useful for identifying conserved noncoding elements (CNEs) in other vertebrates by virtue of its phylogenetic position, the nonduplicated state of its genome (relative to teleost fishes), and the fact that its genome may be evolving slower. For example, we have used global alignments to identify a cadre of CNEs in the HOX clusters [19] [Amemiya, unpubl.] and in the *Otx2* region [Aizawa, in press]. Shown in Appendix 2 is a VISTA plot of the entire HOX-A cluster of the coelacanth with that of the human. This plot shows that CNEs are conserved throughout the cluster, particularly at the 3' end (which is known to contain many regulatory regions involved in axial patterning). It also shows the presence of a specific CNE at the 5' end of *Hoxa14* of *Latimeria*. This CNE is probably a proximal promoter for *Latimeria Hoxa14*, however, it is retained in all tetrapods even though the gene has been lost. This element is extremely conserved amongst tetrapods and would be considered an "ultraconserved" element [52]. Preliminary functional analysis of this sequence suggests that it has retained biological activity in the tetrapods and that it may be a master control element for the HOX-A cluster [Amemiya, unpubl.].

Retention of ancestral molecular characters

1. Genes

The slow rate of evolution of the coelacanth and of its genome may have contributed to the retention of ancestral genes that have been lost in teleost and tetrapod lineages. For example we have found that the coelacanth has retained a *Hox14* paralog which is only found in cartilaginous fishes and amphioxus [24-26,53]. The function of *Hox14* genes is, as yet, unknown; however, based on preliminary data from the coelacanth and skate, it is likely that this gene serves an important role in axial and appendicular patterning. It is tempting to speculate on the role that these genes may have had in evolution and development of the tetrapod limb.

Similar to the *Hox14* story, the coelacanth has retained features of the cartilaginous fish immunoglobulin heavy chain locus [54] (see also letter by Litman). This not only holds for the organization of variable region genes but for the possession of a heavy chain constant region isotype, IgW, found in cartilaginous fishes [20,55,56]. The finding of IgW in the coelacanth and lungfish has far-reaching implications for the evolution of heavy chain isotypes and of isotype switching [57,58][5].

2. Retrotransposons

As with the *Hox14* gene and the immunoglobulin heavy chain locus, the slower rate of molecular evolution in the coelacanth may also have influenced the turnover rate of retrotransposon (SINE) elements. Most retrotransposon families undergo expansion and rapid turnover during evolution [59]. In the case of *Latimeria*, two such SINE families have been shown to predate the

[5] The isotype switching problem is of considerable interest to comparative immunologists since the molecular machinery involved seems to have roles in somatic hypermutation, another conserved process that is necessary for *de novo* antibody diversification.

coelacanth-tetrapod divergence [27,28]. These families are propagated and maintained in the coelacanth genome as typical SINE-like families, but have undergone significant turnover in the tetrapod genomes, even adopting new functions (exaptation). How did these families maintain their integrity in the coelacanth genome over 400 millions years of evolution? If these two examples are any indication perhaps we can look to the coelacanth genome as a repository of such ancient elements in order to study how the tetrapod lineage has coopted these elements for function and to understand the regulatory logic of these primordial parasitic elements (see letters by Bejerano, Schartl and Haussler).

Broader impacts

From the moment of its discovery in 1938, *Latimeria* captured the hearts of scientists and the public at large all over the world. Politics, exploitation, greed, intrigue, fraud, and rivalry have been as much a part of its history as is the pure quest for knowledge [2,7,8]. Coelacanth conservation became an issue when a monetary value was attached to this fish, a value that has increased dramatically when museums started looking for live specimens for display and the rumored value of coelacanth notochord fluid is being sold for $1000 a drop in China [7]. The existence of a black market for the coelacanth resulted in *L. chalumnae* being placed in Appendix I of CITES in 1989 [8]. Thus, genomic resources are necessary, valuable and inexhaustible tools for continued study of a protected species such as *Latimeria*. More importantly, the genome sequencing of *Latimeria* will do much to foster both goodwill and the promotion of science and technology among the consortium of countries in the African Coelacanth Ecosystem Programme (see letters by Dorrington and Ribbink).

SUMMARY

We advocate 7X WGS sequencing of the African coelacanth, *Latimeria chalumnae.* The coelacanth is a key taxon that fills a huge void in genomic comparisons within and among vertebrates. There are many interesting biological questions that are unique to the coelacanth, or which the coelacanth (by virtue of its phylogenetic position) can help to address. Preliminary BAC-based sequencing data have been procured from the closely-related *Latimeria menadoensis* that suggest that the coelacanth genome will be very useful as a comparative genomics tool. Efforts have been initiated to procure high-quality tissue samples for this project.

Reference List

1. Smith J.L.B. (1940). A living coelacanth fish from South Africa. Trans. R. Soc. S. Afr. *28*: 1-106.

2. Smith J.L.B. (1956). Old Fourlegs. The Story of the Coelacanth. (London: Longman Green).

3. Erdmann M.V., Caldwell R.L., and Moran J.V. (1998). Indonesian 'king of the sea' discovered. Nature *395*: 335.

4. Pouyaud L., Wirjoatmodjo S., Rachmatika I., Tjakrawidjaja A., Hadiaty R., and Hadie W. (1999). A new species of coelacanth. Genetic and morphologic proof. C. R. Acad. Sci. III *322*: 261-267.

5. Holder M.T., Erdmann M.V., Wilcox T.P., Caldwell R.L., and Hillis D.M. (1999). Two living species of coelacanths? Proc Natl Acad Sci U S A *96*: 12616-20.

6. Thompson K.S. (1991). Living fossil: the story of the coelacanth. (New York: W. W. Norton).

7. Weinberg S. (2000). A fish caught in time: the search for the coelacanth. (New York: HarperCollins).

8. Aquamarine Fukushima. (2002). Aquamarine Symposium: The Coelacanth, Fathom the mystery. (Iwaki City: Aquamarine Fukushima).

9. Modisakeng K.W., Amemiya C.T., Dorrington R.A., and Blatch G.L. (2006). Molecular biology studies on the coelacanth: A review. South African Journal of Science *in press*.

10. Brinkmann H., Venkatesh B., Brenner S., and Meyer A. (2004). Nuclear protein-coding genes support lungfish and not the coelacanth as the closest living relatives of land vertebrates. Proc. Natl. Acad. Sci. U. S. A. *101*: 4900-4905.

11. Inoue J.G., Miya M., Venkatesh B., and Nishida M. (2005). The mitochondrial genome of Indonesian coelacanth Latimeria menadoensis (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths. Gene *349*: 227-235.

12. Takezaki N., Figueroa F., Zaleska-Rutczynska Z., Takahata N., and Klein J. (2004). The phylogenetic relationship of tetrapod, coelacanth, and lungfish revealed by the sequences of forty-four nuclear genes. Mol. Biol. Evol. *21*: 1512-1524.

13. Yokoyama S. and Tada T. (2000). Adaptive evolution of the African and Indonesian coelacanths to deep-sea environments. Gene *261*: 35-42.

14. Zardoya R. and Meyer A. (1997). The complete DNA sequence of the mitochondrial genome of a "living fossil," the coelacanth (Latimeria chalumnae). Genetics *146*: 995-1010.

15. Zardoya R. and Meyer A. (1996). Evolutionary relationships of the coelacanth, lungfishes, and tetrapods based on the 28S ribosomal RNA gene. Proc. Natl. Acad. Sci. U. S. A *93*: 5449-5454.

16. Chiu C.H., Nonaka D., Xue L., Amemiya C.T., and Wagner G.P. (2000). Evolution of Hoxa-11 in Lineages Phylogenetically Positioned along the Fin-Limb Transition. Mol Phylogenet Evol *17*: 305-316.

17. Danke J., Miyake T., Powers T., Schein J., Shin H., Bosdet I., Erdmann M., Caldwell R., and Amemiya C.T. (2004). Genome resource for the Indonesian coelacanth, Latimeria menadoensis. J. Exp. Zoolog. A Comp Exp. Biol. *301*: 228-234.

18. Noonan J.P., Grimwood J., Danke J., Schmutz J., Dickson M., Amemiya C.T., and Myers R.M. (2004). Coelacanth genome sequence reveals the evolutionary history of vertebrate genes. Genome Res. *14*: 2397-2405.

19. Shashikant C., Bolanowski S.A., Danke J., and Amemiya C.T. (2004). Hoxc8 early enhancer of the Indonesian coelacanth, Latimeria menadoensis. J. Exp. Zoolog. B Mol. Dev. Evol. *302*: 557-563.

20. Saha N.R., Ota T., and Cheng J.-F.A.C.T. (2006). Genomics and the evolution of the immune system: rearranging genes in the coelacanth.), pp. 215.

21. Amemiya C.T., Ota T., and Litman G.W. (1996). Construction of P1 artificial chromosome (PAC) libraries from lower vertebrates. In Analysis of Nonmammalian Genomes, Lai E, Birren B, eds. (San Diego: Academic Press), pp. 223-256.

22. Schartl M., Hornung U., Hissmann K., Schauer J., and Fricke H. (2005). Genetics: relatedness among east African coelacanths. Nature *435*: 901.

23. Cimino M.C. and Bahr G.F. (1974). The nuclear DNA content and chromatin ultrastructure of the coelacanth *Latimeria chalumnae*. Exptl. Cell Res. *88*: 263-272.

24. Powers T.P. and Amemiya C.T. (2004). Evidence for a Hox14 paralog group in vertebrates. Curr. Biol. *14*: R183-R184.

25. Powers T.P. and Amemiya C.T. (2004). Evolutionary plasticity of vertebrate Hox genes. Current Genomics *5*: 459-472.

26. Hoegg S. and Meyer A. (2005). Hox clusters as models for vertebrate genome evolution. Trends Genet. *21*: 421-424.

27. Bejerano G., Lowe C.B., Ahituv N., King B., Siepel A., Salama S.R., Rubin E.M., Kent W.J., and Haussler D. (2006). A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature *441*: 87-90.

28. Nishihara H., Smit A.F., and Okada N. (2006). Functional noncoding sequences derived from SINEs in the mammalian genome. Genome Res. ..

29. Shubin N.H., Daeschler E.B., and Jenkins F.A., Jr. (2006). The pectoral fin of Tiktaalik roseae and the origin of the tetrapod limb. Nature *440*: 764-771.

30. Daeschler E.B., Shubin N.H., and Jenkins F.A., Jr. (2006). A Devonian tetrapod-like fish and the evolution of the tetrapod body plan. Nature *440*: 757-763.

31. Ahlberg P.E. and Clack J.A. (2006). Palaeontology: a firm step from water to land. Nature *440*: 747-749.

32. Forey P.L. (1988). Golden jubilee for the coelacanth *Latimeria chalumnae*. Nature *336*: 727-732.

33. Woolfe A., Goodson M., Goode D.K., Snell P., McEwen G.K., Vavouri T., Smith S.F., North P., Callaway H., Kelly K. et al. (2005). Highly conserved non-coding sequences are associated with vertebrate development. PLoS. Biol. *3*: e7.

34. Thomas J.W., Touchman J.W., Blakesley R.W., Bouffard G.G., Beckstrom-Sternberg S.M., Margulies E.H., Blanchette M., Siepel A.C., Thomas P.J., McDowell J.C. et al. (2003). Comparative analyses of multi-species sequences from targeted genomic regions. Nature *424*: 788-793.

35. Amores A., Force A., Yan Y.L., Joly L., Amemiya C., Fritz A., Ho R.K., Langeland J., Prince V., Wang Y.L. et al. (1998). Zebrafish hox clusters and vertebrate genome evolution. Science *282*: 1711-1714.

36. Amores A., Suzuki T., Yan Y.L., Pomeroy J., Singer A., Amemiya C., and Postlethwait J.H. (2004). Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish. Genome Res. *14*: 1-10.

37. Cresko W.A., Yan Y.L., Baltrus D.A., Amores A., Singer A., Rodriguez-Mari A., and Postlethwait J.H. (2003). Genome duplication, subfunction partitioning, and lineage divergence: Sox9 in stickleback and zebrafish. Dev. Dyn. *228*: 480-489.

38. Taylor J.S., Van de P.Y., and Meyer A. (2001). Genome duplication, divergent resolution and speciation. Trends Genet. *17*: 299-301.

39. Taylor J.S., Braasch I., Frickey T., Meyer A., and Van de P.Y. (2003). Genome duplication, a trait shared by 22000 species of ray-finned fish. Genome Res. *13*: 382-390.

40. Van de P.Y., Taylor J.S., Braasch I., and Meyer A. (2001). The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. J. Mol. Evol. *53*: 436-446.

41. Vandepoele K., De Vos W., Taylor J.S., Meyer A., and Van de P.Y. (2004). Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proc. Natl. Acad. Sci. U. S. A *101*: 1638-1643.

42. Koh E.G., Lam K., Christoffels A., Erdmann M.V., Brenner S., and Venkatesh B. (2003). Hox gene clusters in the Indonesian coelacanth, Latimeria menadoensis. Proc Natl Acad Sci U S A *100*: 1084-8.

43. Brinkmann H., Denk A., Zitzler J., Joss J.J., and Meyer A. (2004). Complete mitochondrial genome sequences of the South american and the Australian lungfish: testing of the phylogenetic performance of mitochondrial data sets for phylogenetic problems in tetrapod relationships**.** J. Mol. Evol. *59***:** 834-848.

44. Delsuc F., Brinkmann H., Chourrout D., and Philippe H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates**.** Nature *439***:** 965-968.

45. Ahlberg P.E. (1992). Coelacanth fins and evolution**.** Nature *358***:** 459.

46. Andrews S.M. (1977). The axial skeleton of the coelacanth, *Latimeria***.** In Problems in vertebrate evolution, Andrews SM, Miles RS, Walker AD, eds. (London: Linnean Society of London), pp. 271-288.

47. Yokoyama S., Zhang H., Radlwimmer F.B., and Blow N.S. (1999). Adaptive evolution of color vision of the Comoran coelacanth (Latimeria chalumnae)**.** Proc. Natl. Acad. Sci. U. S. A *96***:** 6279-6284.

48. Tabin C. and Laufer E. (1993). Hox genes and serial homology**.** Nature *361***:** 692-693.

49. Wagner G.P. and Chiu C.H. (2001). The tetrapod limb: a hypothesis on its origin**.** J Exp Zool *291***:** 226-40.

50. Taylor J.S. and Brinkmann H. (2001). 2R or not 2R? Trends Genet *17***:** 488-9.

51. Taylor J.S., Van de Peer Y., Braasch I., and Meyer A. (2001). Comparative genomics provides evidence for an ancient genome duplication event in fish**.** Philos Trans R Soc Lond B Biol Sci *356***:** 1661-79.

52. Bejerano G., Pheasant M., Makunin I., Stephen S., Kent W.J., Mattick J.S., and Haussler D. (2004). Ultraconserved elements in the human genome**.** Science *304***:** 1321-1325.

53. Ferrier D.E. (2004). Hox genes: Did the vertebrate ancestor have a Hox14? Curr. Biol. *14***:** R210-R211.

54. Amemiya C.T., Ohta Y., Litman R.T., Rast J.P., Haire R.N., and Litman G.W. (1993). V_H gene organization in a relict species, the coelacanth *Latimeria chalumnae*: evolutionary implications**.** Proc. Natl. Acad. Sci. USA *90***:** 6661-6665.

55. Rumfelt L.L., Lohr R.L., Dooley H., and Flajnik M.F. (2004). Diversity and repertoire of IgW and IgM VH families in the newborn nurse shark**.** BMC. Immunol. *5***:** 8.

56. Anderson M.K., Strong S.J., Litman R.T., Luer C.A., Amemiya C.T., Rast J.P., and Litman G.W. (1999). A long form of the skate IgX gene exhibits a striking resemblance to the new shark IgW and IgNARC genes**.** Immunogenetics *49***:** 56-67.

57. Ota T., Rast J.P., Litman G.W., and Amemiya C.T. (2003). Lineage-restricted retention of a primitive immunoglobulin heavy chain isotype within the Dipnoi reveals an evolutionary paradox**.** Proc. Natl. Acad. Sci. U. S. A *100***:** 2501-2506.

58. Stavnezer J. and Amemiya C.T. (2004). Evolution of isotype switching**.** Semin. Immunol. *16***:** 257-275.

59. Weiner A.M. (2002). SINEs and LINEs: the art of biting the hand that feeds you**.** Curr. Opin. Cell Biol. *14***:** 343-350.

60. Carroll R.L. (1988). Vertebrate paleontology and evolution. (New York: W.H. Freeman and Co.).

61. Clack J.A. (2002). Gaining Ground: the origin and evolution of tetrapods. (Bloomington: Indiana University Press).

Appendix 1. Fishy facts about coelacanths.

Discovery of living coelacanths	1938 -- Marjorie Courtenay-Latimer; off coast of South Africa; *Latimeria chalumnae* [1] 1998 -- Mark Erdmann; Sulawesi, Indonesia; *Latimeria menadoensis* [3-5].The two species are quite similar; estimates of mtDNA divergence suggest a divergence time of roughly 30-40 MYA [11].
Paleontology and evolution	Existed primarily around 360-80 MYA, with peak at 230 MYA; diverged from tetrapods an estimated 350-400 MYA; there are around 120 or so extinct species [6].
Classification	They are placed in the Sarcopterygii (lobe-fin vertebrates). The Sarcopterygii includes coelacanths, lungfishes and tetrapods (see Fig. 1), and many extinct lobe-finned fishes.
Phylogeny and systematics	Despite numerous attempts to clarify sarcopterygian relationships, the coelacanth-lungfish-tetrapod relationships based on both morphological and molecular criteria are still controversial [12] (see letter by Takezaki). It is also important to recognize that many extinct lineages coincided (coexisted) with early sarcopterygians [6,60].
Designation as a “living fossil”?	Highly valid since close anatomical comparisons of *Latimeria* with fossil coelacanths show that its basic bauplan and skeleton have changed very little throughout history.
Designation as “old four legs”?	This moniker by J. L. B. Smith [2] is not really valid. While the coelacanth clearly has appendicular skeletons that resemble limbs, its fins do not have an underlying muscular structure to support much weight. Other fossil species such as *Acanthostega* and *Tiktaalik* are much more *bona fide* transitional species [29,30,60,61].
Etymology	*Coelacanthus* means “hollow spine” and refers to the hollow neural and haemal spines of the vertebrae that connect to the tubular bones supporting the upper and lower caudal-fin rays. Notably all coelacanths, extinct and extant, possess these hollow spines.
Unique skeletal features	Skeleton is cartilaginous; notochord replaces typical bony vertebral column, contains an oily substance and persists throughout adulthood. Coelacanths possess fleshy paired fins and fleshy dorsal and anal fins; they have an extra dorsal fin and unusual tail fin.
Reproduction and life history	They are ovoviparous, have huge eggs, and give birth to a small number (<100) of live young. Gestation time is estimated to be 14 months. The lifespan of the coelacanth is estimated to be 80 years [6].

Appendix 2. VISTA plot of HOX-A cluster from coelacanth vs. HOX-A cluster from human. The reference sequence is coelacanth. The blue boxes represent exons. The blue peaks are regions of high nucleotide homology in the exons whereas the pink peaks are regions of high nucleotide homology within noncoding regions (CNEs). There are several CNEs throughout the HOX-A cluster, particularly at the 3' end, which contains more retinoic acid receptor elements. The CNE just 5' of the *Hoxa14* gene is interesting in that it is likely the proximal promoter that regulates *Hoxa14* in the coelacanth. Paralogous group-14 genes have been lost in tetrapods [24,25], however, all tetrapods (including chicken and frog) have retained this regulatory element, which may be a master regulatory element for the HOX-A cluster [Amemiya, unpubl.].