SUPPLEMENTARY MATHERIAL AND METHODS

A living fossil in the genome of a living fossil: Harbinger transposons in the coelacanth genome

Jeramiah J. Smith, Kenta Sumiyama and Chris T. Amemiya

Prediction of functional elements within *LatiHarb1*

To further characterize this repetitive element, we first asked if its sequence was likely to encode any proteins or other functional elements. *Ab initio* predictions of protein coding elements identified two non-overlapping candidates that were oriented in opposite directions (Figure 1A, Supplementary Tables S2 and S3). A *blastp* search of the GenBank nr database identified several matches for both predicted proteins. No other significant alignments were found when the entire 8.7 kb element was searched (*tblastx*) against this same database ($e<10^{-3}$). One of these predicted proteins aligned to *harbi1* genes from several vertebrate species and to predicted transposases from several metazoan species, all of which contained a conserved *harbinger* transposase (aka Transposase 11) domain (Supplementary Table S4). The presence of a transposase gene within this element, together with flanking inverted 35 bp repeats (Figure 1A), strongly implies that this gene participates the transposition of the element through a cut-and-paste mechanism.

PCR screening for *lha* and *ltase* genes in a *Lm* BAC library – Standard PCR conditions (~1 ng DNA, 50 ng each primer, 1.2 mM $MgCl_2$, 0.3 U Taq polymerase, 1x PCR buffer, 200 μM each of dATP, dCTP, dGTP, dTTP; thermal cycling at 94° C for 4 minutes; 33 cycles of 94° C for 45 s, 60° C for 45 s, 72° C for 30 s; and 72° C for 7 minutes) were used to amplify fragments of the *lha* and *tpase* genes from 96 clones that were drawn from the *Lm* BAC library VMRC4 (Danke et al. 2004). Oligoucleotide primers were designed from *lha* (lha.f - AAGGACACGTGGAGGAGGTA, lha.r - GAGCGTCAAGAGGAAGATGG) and *tpase* genes (Ltpase.f - GCGTGCTGTCGTGAGTATGT, Ltpase.r - GGGGCGATGATACCAATATG**).** Thirty-four of these clones were positive for *lha*. These same 34 clones, and one additional clone, were positive for *tpase* . Overall, the *lha* and *tpase* genes were both very abundant in these clones and were strongly correlated in their distribution among BACs ($G = 116$, $p = 5.5E^{-27}$, $n = 96$, d.f. = 1). Accounting for the possibility of multiple insertions per BAC, we estimate that there were ~52 complete copies of *LatiHarb1* among 96 BAC clones. The distribution of these genes among *Lm* BACs is consistent with our other estimates of *LatiHarb1* abundance, as well as the simultaneous proliferation of *lha* and *tpase* within the *LatiHarb1* element.

Hybridization of a *lha* fragment to *Lm* genomic southern blots and high-density filters – Genomic Southern blots and previously described high-density filters of the *Lm*

BAC library VMRC4 (Danke et al. 2004) were probed with a fragment of *lha*. This 500 bp fragment was amplified using standard PCR conditions (5 ng DNA, 50 ng each primer, 1.2 mM $MgCl_2$, 0.3 U Taq polymerase, 1x PCR buffer, 200 μM each of dATP, dCTP, dGTP, dTTP; thermal cycling at 94° C for 4 minutes; 33 cycles of 94° C for 45 s, 60° C for 45 s, 72° C for 30 s; and 72° C for 7 minutes). Oligonucleotide primers were designed from *lha* (lha.f - AAGGACACGTGGAGGAGGTA, lha.r - GAGCGTCAAGAGGAAGATGG) and the BAC clone VMRC4-45P06 as DNA template. This amplified fragment was purified (Qiagen #28104) and radiolabeled using α^{32}P-dCTP and random priming (New England Biolabs #N1500). Hybridization followed standard conditions (Sambrook and Russell 2001).

Construction of a large-insert construct containing *LatiHarb1* for mouse transgenic analysis – A BAC clone (GenBank AC147788) containing the *Latimeria menadoensis* posterior *HoxA* genes (Powers and Amemiya 2004) was modified by recombineering (Copeland et al. 2001) to supplant *Hoxa14* with a red fluorescence protein (RFP) gene (from pDsRed-Express-1, Clontech). To facilitate recombineering, the intact BAC insert (168 kb) was first transferred to a P1 artificial chromosome (PAC) vector via conventional *Not*I cloning. The PAC vector, in turn, had been previously modified by insertion of a green fluorescence protein (GFP) gene under the control of a carp actin promoter (from pFRMwg, (Gibbs and Schmale 2000)). This construct, pPAC-GFP-1054-RFP, was originally generated so as to be able to assess the expression pattern of the coelacanth *Hoxa14* gene in a transgenic mouse system. Serendipitous discovery of two copies of *LatiHarb1* in its genomic sequence, however, allowed us to subsequently use the transgenic mouse to assess the *in vivo* transcriptional potential of *LatiHarb1*. The incorporation of the first-generation GFP marker in the vector sequence was done in order to detect transgenic founder animals by virtue of constitutive, albeit weak, GFP expression. Finally, addition of an *iTol2* cassette into the PAC vector backbone via recombineering was carried out to facilitate transgenesis in mouse embryos using the *Tol2* transposase system (Suster et al. 2009). The final large-insert targeting construct for mouse transgenesis was denoted pPAC-GFP-1054-RFP-iTol2-Kan. Its complete nucleotide sequence and full details of its construction are available upon request.

Construction of the Tg (pPAC-GFP-1054-RFP-iTol2-Kan) B6C3 transgenic mouse line – Approximately 5ng/μl of pPAC-GFP-1054-RFP-iTol2-Kan was mixed with 25ng/μl Tol2 transposase mRNA in injection buffer (10mM Tris-HCl and 0.1mM EDTA in DEPC treated water, pH 7.4) and injected into both pronuclei and cytoplasm of fertilized eggs of B6C3F1 mice following previously reported protocols (Suster et al. 2009; Sumiyama et al. 2010). In total, 243 fertilized eggs were injected and transferred to the oviducts of pseudopregnant MCH female mice, yielding 37 live births (16 males and 21 females). One transgenic female was identifies among these as confirmed by PCR screening for the integrated dsRED gene using the following primers: dsRED.lo1 – CGCCGTCCTCGAAGTTCATCA (internal primer for dsRED gene) and LmHox14Promup1 – ATATGCATTTTGCTGTGCCACTG (5' upstream region of the LmHoxa14 gene). The transgenic line was designated *Tg (pPAC-GFP-1054-RFP-iTol2-*

Kan) B6C3. Adapter-ligation PCR was performed as previously described, in order to determine whether integration was mediated by *Tol2* transposase (Sumiyama et al. 2010). Our results showed that this particular integration was not transposase mediated but was the product of a random integration, as produced by conventional pronuclear injection. We confirmed PAC integration in F1s and F2s of the Tg (pPAC-GFP-1054-RFP-iTol2-Kan) B6C3 mice by both PCR amplification of dsRED and GFP expression.

Literature Cited

Copeland NG, Jenkins NA, Court DL. 2001. Recombineering: a powerful new tool for mouse functional genomics. Nat Rev Genet 2:769-779.

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

Gibbs PD, Schmale MC. 2000. GFP as a Genetic Marker Scorable Throughout the Life Cycle of Transgenic Zebra Fish. Mar Biotechnol (NY) 2:107-125.

Powers TP, Amemiya CT. 2004. Evidence for a Hox14 paralog group in vertebrates. Curr Biol 14:R183-R184.

Sambrook J, Russell DW. 2001. Molecular Cloning: A laboratory Manual 3rd Eds. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. 2344 p.

Sumiyama K, Kawakami K, Yagita K. 2010. A simple and highly efficient transgenesis method in mice with the Tol2 transposon system and cytoplasmic microinjection. Genomics 95:306-311.

Suster ML, Sumiyama K, Kawakami K. 2009. Transposon-mediated BAC transgenesis in zebrafish and mice. BMC Genomics 10:477.