

SUPPLEMENTARY FIGURES

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

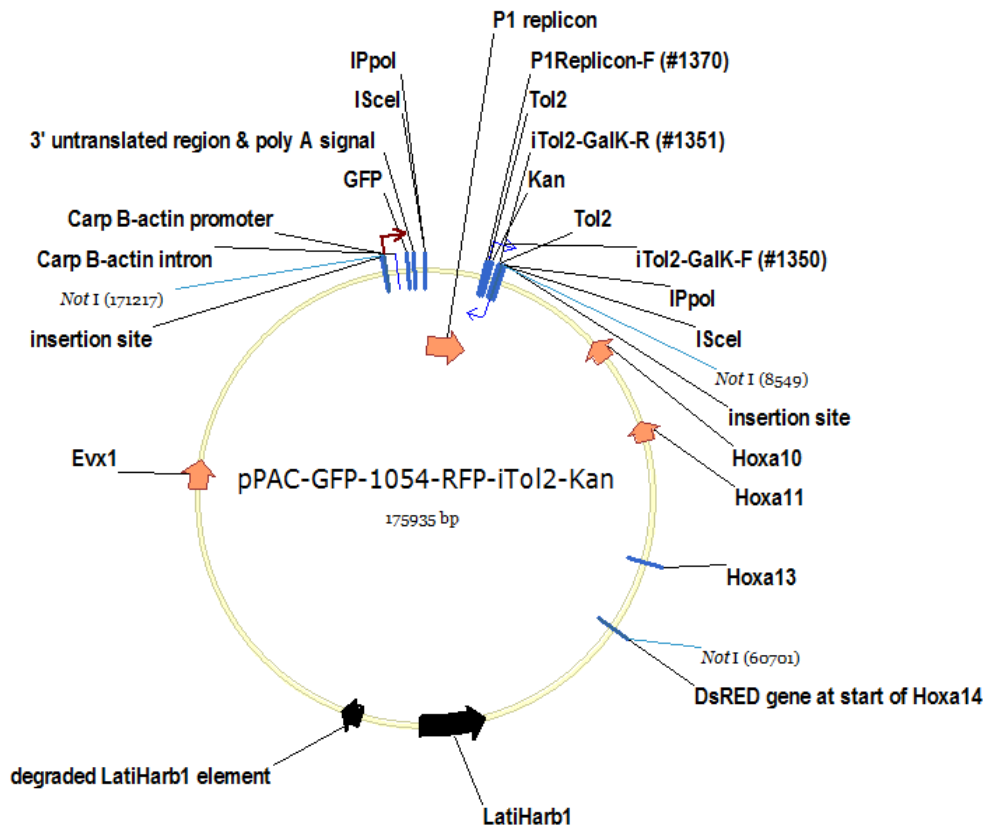
Jeramiah J. Smith, Kenta Sumiyama and Chris T. Amemiya

Supplementary Figure S1 - Amino acid alignment of *lha*-like proteins – The amino acid sequence of *lha* shares sequence similarity with several vertebrate proteins, but is not overtly similar to any particular group of sequences. The alignment includes all vertebrate blast hits to the predicted *lha* gene from *Latimeria* and covers 141 conserved amino acids from *lha*. All sequences aligning to *lha* are labeled by their GenBank identifiers.

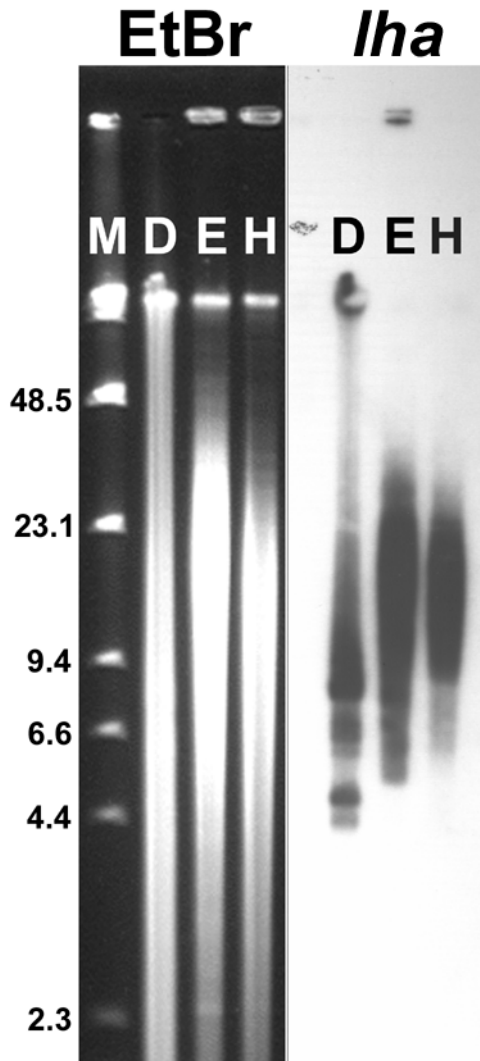
lha QHQGSDPDRIDGAPGPKRKRKRR--FKVAEVDLTHGWDAAADIVLQ--AGSTSLAHEE---CQRITTHQWALGHILGIVVEVHHW
CAH14370 DSDSTKIDMIRIITLPAERKRRPKFKSEKEDLIVTEVITHEAVLFGRETRRSHADADK---DREGIARKITISQPPRSKIDIKHHW
XP_001148200 -----RKPKFKSEKEDLIVTEVITHEAVLFGRETRRSHADADK---DREGIARKITISQPPRSKIDIKHHW
AAI13557 KDSAGPHGRNAGPSTTRAKKRPHTPCQETEVLSKVSHHQLLFGTGLLKAEPTRAYR---VNSRIQAQWALGYCRDWDLDLHHW
AAI26344 KDSAGPHGRNAGPSTTRAKKRPHTPCQETEVLSKVSHHQLLFGTGLLKAEPTRAYR---VNSRIQAQWALGYCRDWDLDLHHW
EAW82324 KDSAGPHGRNAGPSTTRAKKRPHTPCQETEVLSKVSHHQLLFGTGLLKAEPTRAYR---VNSRIQAQWALGYCRDWDLDLHHW
HP_659440 KDSAGPHGRNAGPSTTRAKKRPHTPCQETEVLSKVSHHQLLFGTGLLKAEPTRAYR---VNSRIQAQWALGYCRDWDLDLHHW
XP_001089591 KDLAAGPHGRNAGPSTTRAKKRPHTPCQETEVLSKVSHHQLLFGTGLLKAEPTRAYR---VNSRIQAQWALGYCRDWDLDLHHW
XP_519987 KDSAGPHGRNAGPSTTRAKKRPHTPCQETEVLSKVSHHQLLFGTGLLKAEPTRAYR---VNSRIQAQWALGYCRDWDLDLHHW
XP_539185 KDSAGPHGRNAGPSTTRAKKRPHTPCQETEVLSKVSHHQLLFGSGLRAREPARYR---VNSRIQAQWALGYCRDWDLDLHHW
NP_001075915 KELAARAREHVAS--KTRAKKRPHTPCQETEVLSKVSHHQLLFGSGLRAREPARYR---VNSRIQAQWALGYCRDWDLDLHHW
XP_001381738 KDPLATPSPRHNAGPSTTRAKKRPHTPCQETEVLSKVSHHQLLFGTGLLKAEPTRAYR---VNSRIQAQWALGYCRDWDLDLHHW
XP_001489117 ILPGPGCVPLRPGP--HERITTKRPHFSPQETEVLSKVSHHQLLFGSGLRAREPARYR---VNSRIQAQWALGYCRDWDLDLHHW
XP_001381736 IADPGSTPPOPGH--HERITTKRPHFSPQETEVLSKVSHHQLLFGSGLRAREPARYR---VNSRIQAQWALGYCRDWDLDLHHW
NP_001380063 -----GKRRKAFSHDETELVNIVVHFSAALYGEALRTHSAHQ---LDSQQRNIVFLGYIENSIDDLHHW
XP_424011 -----GKRRKAFSHDETELVNIVVHFSAALYGEALRTHSAHQ---LDSQQRNIVFLGYIENSIDDLHHW
NP_001089601 HQVREIVGASGDPVPTCRKRAAFSHDETELVNIVVHFSAALYGEALRTHSAHQ---LDSQQRNIVFLGYIENSIDDLHHW
XP_520284 KAV-----PAKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLAASA---AWGILRRVWAVATCRRELPEVKKW
XP_001152900 KAV-----PAKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLAASA---AWGILRRVWAVATCRRELPEVKKW
NP_931045 KAV-----PAKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLAASA---AWGILRRVWAVATCRRELPEVKKW
XP_001092549 KAV-----PAKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLAASA---AWGILRRVWAVATCRRELPEVKKW
XP_001092434 KAV-----PAKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLAASA---AWGILRRVWAVATCRRELPEVKKW
AAH21580 KAV-----PAKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLAASA---AWGILRRVWAVATCRRELPEVKKW
XP_001501275 KAV-----PAKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLAASA---AWGILRRVWAVATCRRELPEVKKW
NP_001094544 KAV-----PAKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLAASA---AWGILRRVWAVATCRRELPEVKKW
XP_548443 KAV-----PAKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLAASA---AWGILRRVWAVATCRRELPEVKKW
XP_001079981 KAV-----PAKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLAASA---AWGILRRVWAVATCRRELPEVKKW
NP_919316 KAV-----PAKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLAASA---AWGILRRVWAVATCRRELPEVKKW
NP_001016353 NAT-----PAKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLVTSI---AWHDLKRWAVSTCHRELAEVKKW
AAI28615 NAT-----PAKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLVTSI---AWHDLKRWAVSTCHRELAEVKKW
NP_001090371 NAT-----PAKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLVTSI---AWHDLKRWAVSTCHRELAEVKKW
NP_001025863 NASEP-----KPAKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLAASA---AWHDLKRWAVSTCHRELAEVKKW
XP_691225 NASE-----CAKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLAANS---AWHDLKRWAVSTCHRELAEVKKW
XP_001520341 NTA-----PVKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLAASA---AWGILRRVWAVATCRRELPEVKKW
CAF89557 NASE-----AAKRRKUIFSEREVEIIVVEELKHLLWVHFNAGVPLAANS---AWGILRRVWAVATCRRELPEVKKW
NP_001013508 NSEFDPVYRPPGGLHRRPKRKRSPFQEVVELLSEVQKRHLVGVKTRGSKDLNRR---VLSVTERINERSECHKREVEIIVKRW
NP_001070739 NVSSSLALQ---DINIVRYRKRARFVSEVHLLDENPDRHRLVGVKTRGSKDLNRR---KNAETEARVETSEGDREINELIKRW
NP_001081092 AHEDLFDGVCPCITLPLRTRKRFVDEELNVLQEVHDMVSQLFGEISVKTPTVVK---DREGIARWALGVITVISEIKRW
NP_001089108 AHEDLFDGVCPCITLPLRTRKRFVDEELNVLQEVHDMVSQLFGEISVKTPTVVK---DREGIARWALGVITVISEIKRW

lha RDRVSIKRNIAITARR-----ATGGGPPASLTPLEEYARHISTLEVEGIPGGHIDLRK-----
CAH14370 DDKRPTDMLAPRQGLSG-----PGIGCHTPIHLAHEPATEAALITRAGHFPRAEIDGDSFS----
XP_001148200 DDKRPTDMLAPRQGLSG-----PGAGGRAPTVLTAHERATESALITRAGRFPRAEIDGDSFS----
AAI13557 RDLAVVRRKIGDLKKAH-----GSPGSGKQALAITPVEVYAKTSCQALPEGFSLEPPRATQVD---
AAI26344 RDLAVVRRKIGDLKKAH-----GSPGSGKQALAITPVEVYAKTSCQALPEGFSLEPPRATQVD---
EAW82324 RDLAVVRRKIGDLKKAH-----GSPGSGKQALAITPVEVYAKTSCQALPEGFSLEPPRATQVD---
HP_659440 RDLAVVRRKIGDLKKAH-----GSPGSGKQALAITPVEVYAKTSCQALPEGFSLEPPRATQVD---
XP_001089591 RDLAVVRRKIGDLKKAH-----GSPGSGKQALAITPVEVYAKTSCQALPEGFSLEPPRATQVD---
XP_519987 RDLAVVRRKIGDLKKAH-----GSPGSGKQALAITPVEVYAKTSCQALPEGFSLEPPRATQVD---
XP_539185 RDLAVVRRKIGDLKCG-----PGPGAPALAITPVEVYAKTSCRAPPEGVDEPPRATQAD---
NP_001075915 RDLAVVRRKIGDL-----PRAQALAITPLEDAVAKTSCQALPEGLPPELRRATQID---
XP_001381738 RDLAVVRRKIGDLKKAH-----GSPGSGKQALAITPVEVYAKTSCQAGPVGYELRALRSQLD---
XP_001489117 RDLGVVRRKLAECPRAP-----AESVAPGLITPVERHVAETYSAPAPPGESQATEPISPRILN---
XP_001381736 RDLGVVRRKIGDLKKAH-----TEGQCALVITPVERHVAETYSANHCQESGAEHSP-----
XP_001380063 RDLALDVKKITSKHP-----GSSSRT--HLTPLEKIVASTV--TQASDDEEPE--ALDPGE-----
XP_424011 RDLALDVKKITSKHP-----PHIRAGGLPKRPLITPLEKIVASTV--LQASHDEPE--ITLDPQGSW---
NP_001089601 RDLALDVKKITHRTSGKA-----PLPGVAPLLEKLSPLEDLVASTIGHHCITLDBQEBHPMEGVER---
XP_520284 SDLTTEVRRKIAQVRAAVEGEG--APGPTTEEDGAGGPGTCGGSGGGGPAVAVLLITPQCRICHLAGEVTIISLPSITTEHPVALG-----
XP_001152900 SDLTTEVRRKIAQVRAAVEGEG--APGPTTEEDGAGGPGTCGGSGGGGPAVAVLLITPQCRICHLAGEVTIISLPSITTEHPVALG-----
NP_931045 SDLTTEVRRKIAQVRAAVEGEG--APGPTTEEDGAGGPGTCGGSGGGGPAVAVLLITPQCRICHLAGEVTIISLPSITTEHPVALG-----
XP_001092549 SDLTTEVRRKIAQVRAAVEGEG--APGPTTEEDGAGGPGTCGGSGGGGPAVAVLLITPQCRICHLAGEVTIISLPSITTEHPVALG-----
XP_001092434 SDLTTEVRRKIAQVRAAVEGEG--APGPTTEEDGAGGPGTCGGSGGGGPAVAVLLITPQCRICHLAGEVTIISLPSITTEHPVALG-----
AAH21580 SDLTTEVRRKIAQVRAAVEGEG--APGPTTEEDGAGGPGTCGGSGGGGPAVAVLLITPQCRICHLAGEVTIISLPSITTEHPVALG-----
NP_001501275 SDLTTEVRRKIAQVRAAVEGEG--APGPTTEEDGAGGPGTCGGSGGGGPAVAVLLITPQCRICHLAGEVTIISLPSITTEHPVALG-----
NP_001094544 SDLTTEVRRKIAQVRAAVEGEG--APGPTTEEDGAGGPGTCGGSGGGGPAVAVLLITPQCRICHLAGEVTIISLPSITTEHPVALG-----
XP_548443 SDLTTEVRRKIAQVRAAVEGEG--APGPTTEEDGAGGPGTCGGSGGGGPAVAVLLITPQCRICHLAGEVTIISLPSITTEHPVALG-----
XP_001079981 SDLTTEVRRKIAQVRAAVEGEG--APGPTTEEDGAGGPGTCGGSGGGGPAVAVLLITPQCRICHLAGEVTIISLPSITTEHPVALG-----
NP_919316 SDLTTEVRRKIAQVRAAVEGEG--APGPTTEEDGAGGPGTCGGSGGGGPAVAVLLITPQCRICHLAGEVTIISLPSITTEHPVALG-----
NP_001016353 SDLTTEVRRKISAKAAVEG-----EAD-----VETPPVHLITPQCRICHLAGEVTIISLPSITTEHPVALG-----
AAI28615 SDLTTEVRRKISAKAAVEG-----EAD-----VETPPVHLITPQCRICHLAGEVTIISLPSITTEHPVALG-----
NP_001090371 SDLTTEVRRKISAKAAVEG-----EAD-----VETPPVHLITPQCRICHLAGEVTIISLPSITTEHPVALG-----
NP_001025863 SDLTTEVRRKIAQVRAAVEG-----GGESQISGGAGTEEDVETATAAPVLLITPQCRICHLAGEVTIISLPSITTEHPVALG-----
XP_691225 SDLTTEVRRKIAQVRAAVEG-----EAD-----CTTPVILITPQCRICHLAGEVTIISLPSITTEHPVALG-----
XP_001520341 SDLTTEVRRKIAQVRAAVEGEG--PGEGHAAAG-----DADGDLGAPVLLITPQCRICHLAGEVTIISLPSITTEHPVALG-----
CAF89557 SDLTTEVRRKIAQVRAAVEG-----SAD-----SAAVPLISAKQRICHLAGEVTIISLPSITTEHPVALG-----
NP_001013508 ADLACDTRKRVAAKAGATSQAQ--LAQ---HLPIETVHQLIQLSHSRINSSFALDDQDDDDDLAHLSEISQSSSWAHGRPHS--
NP_001070739 SDCDITVRLAAQVTCIVVSR--LARNHILSLPHEVIVESILELDDKPIEATVQSRSSRSHDDQEVCE--DEEEDGCVVGH-----
NP_001081092 HDLRRKRNKLSVTEPPGHLEA--SVPVFPVPHQR---LIQETIHPHQQGISLDTSSAIFSHAAITVQDREYLDQTVVQ---

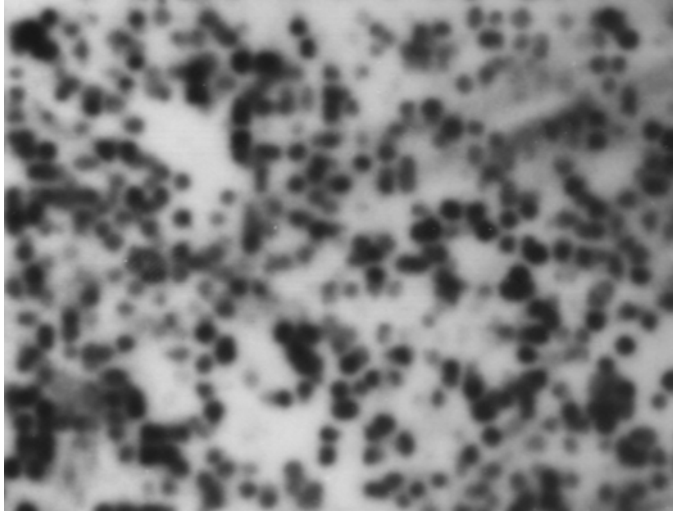
Supplementary Figure S2 - Structure of the pPAC-GFP-1054-RFP-iTol2-Kan PAC clone – This clone was used for generating the *Tg* (*pPAC-GFP-1054-RFP-iTol2-Kan*) *B6C3* mouse transgenic line.



Supplementary Figure S3 - Southern hybridizations of *lha* gene. – M = Low Range PFG Marker (NEB #N0350S), marker labels are in kilobases. D, E, and H represent *Lm* genomic heart DNA digested with *Dra*I, *Eco*RI and *Hind*III respectively. Left panel is the ethidium bromide-stained gel and the right panel is the subsequent Southern blot-hybridization with the *lha* probe.



Supplementary Figure S4 - Hybridization of *lha* gene to a high-density filter of the *Lm* BAC library. – The *Lm* BAC library was spotted onto 22 cm² nylon filters in a 4 x 4 grid pattern (18,432 individual clones per filter) for hybridization screening (Danke et al. 2004). This region of a high density filter represents ~400 clones from the library spotted in duplicate.



Literature Cited

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 Zool A Comp Exp Biol* 301:228-234.

Higgins DG, Sharp PM. 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* 73:237-244.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876-4882.