

The African coelacanth genome provides insights into tetrapod evolution

bioinformaatika ajakirjaklubi

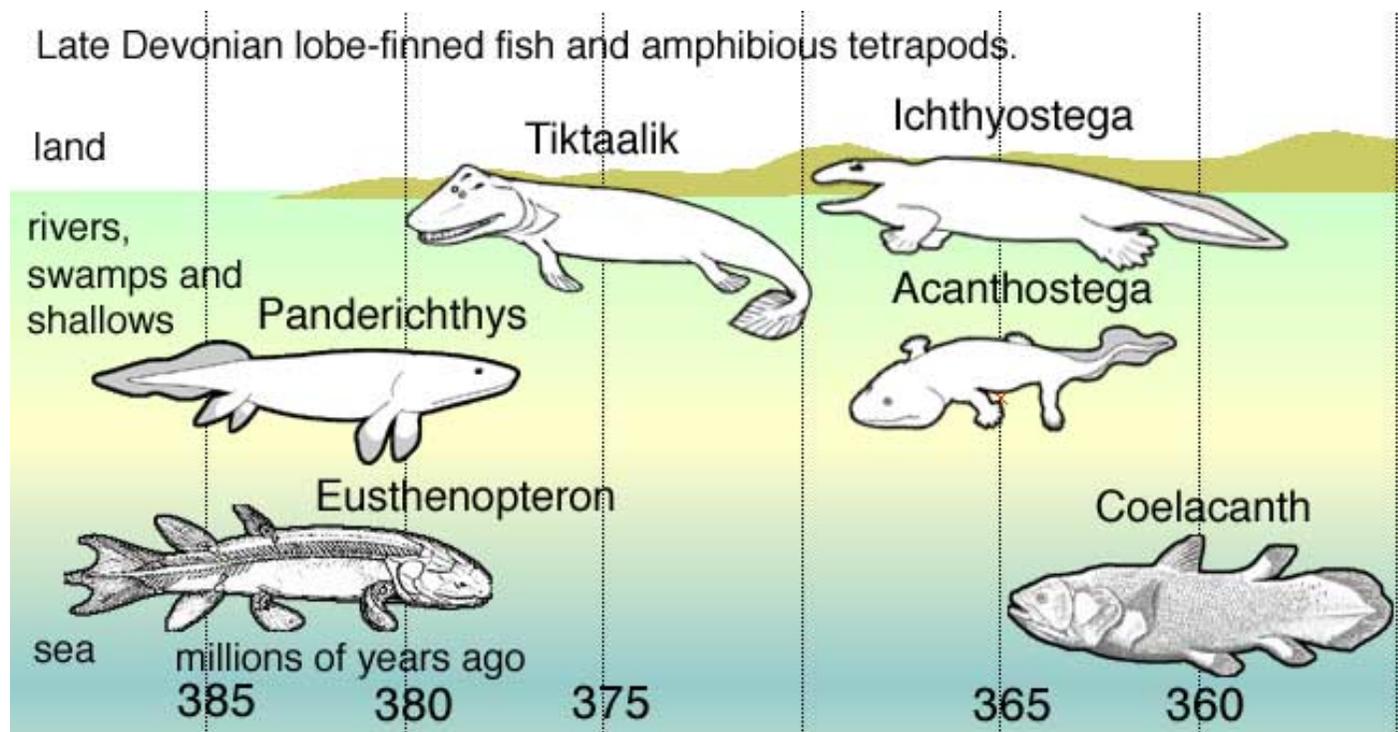
27.05.2013

Ülesehitus

- Täisgenoomi sekveneerimisest
- vankrid mille ette neid andmeid on rakendatud
 - evolutsiooni kiirus – "elus fossiil" - kas tõesti?
 - elu siirdumine maismaale
 - jäsemete areng
 - lämmastikust lahtisaamine
 - loote areng
 - Ig-M/Ig-W

Latimeria chalumnae

- Leiti 1938a India Ookeanist
- Coelacanth /^lsi:ləkæn θ /; Order Coelacanthiformes
- Peeti väljasurnuks – 70 millionit aastat tagasi



millionit aastat

Genome assembly and annotation

- karyotype: 48 Chromosomes
- It was sequenced by Illumina sequencing technology and assembled using the short read genome assembler ALLPATHS-LG
- The draft assembly is 2.86 gigabases (Gb) in size and is composed of 2.18 Gb of sequence plus gaps between contigs.
- Contig N50 size 12.7 kb
- Scaffold N50 size of 924 kb

Genome annotation by Ensemble pipeline

- Ensembl gene annotation pipeline (release 66, Feb. 2012)
 - + RNA-seq data generated from muscle
 - 18 Gb of paired-end reads were assembled using Trinity software
 - orthology with other vertebrates
- 19,033 protein-coding genes containing
- 21,817 transcripts

Genome annotation by MAKER

- pipeline used the *L. chalumnae* Ensembl gene set,
- Uniprot protein alignments,
- *L. chalumnae* (muscle) and *L. menadoensis* (liver and testis) 15 RNA-seq data to create gene models

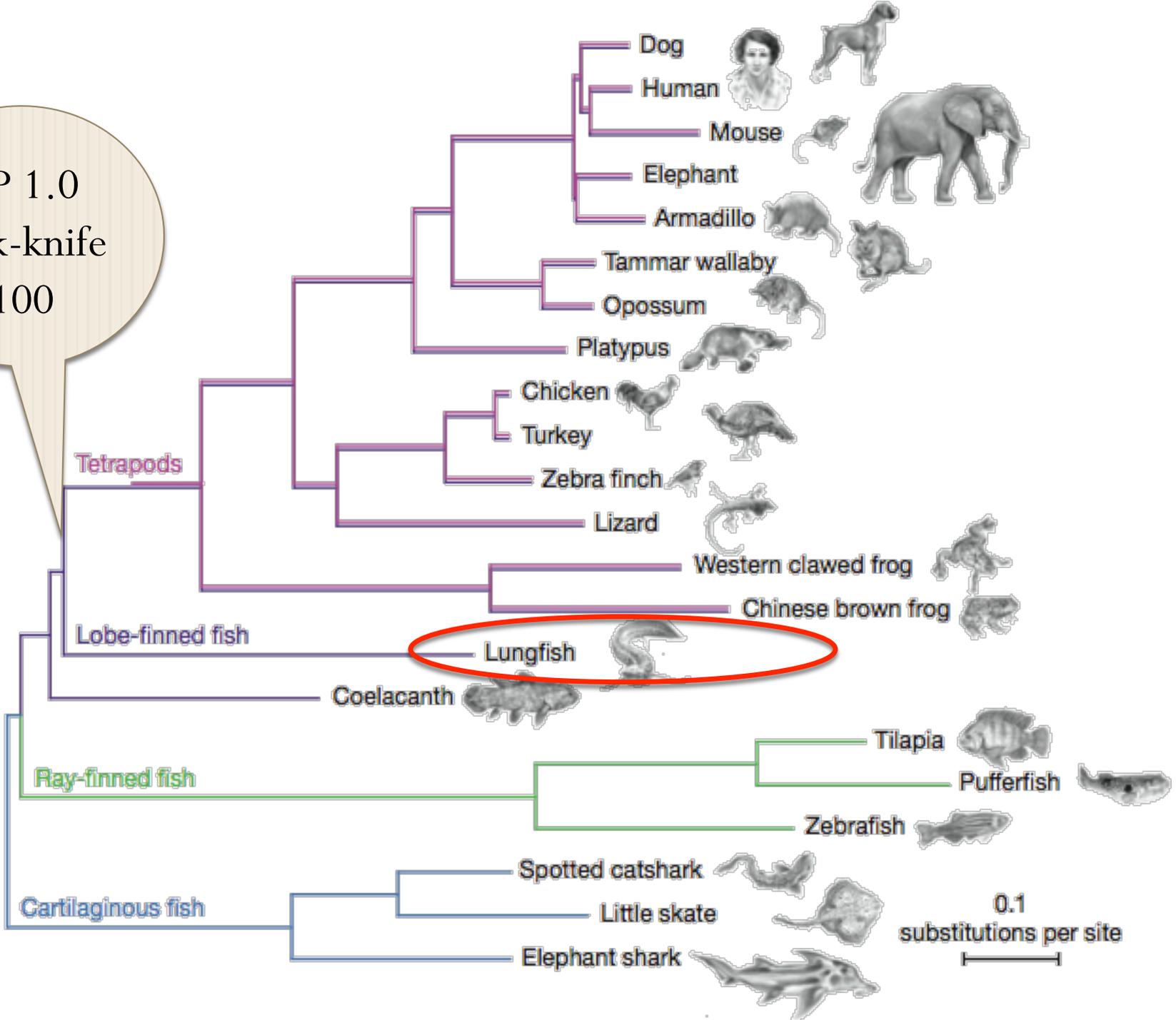
- 29,237 protein-coding gene annotations.
- 2,894 short non-coding RNAs,
- 1,214 long non-coding RNAs,
- > 24,000 conserved RNA secondary structures

Supplementary Note 2, Supplementary Tables 3 and 4, Supplementary Data 1–3 and
Supplementary Fig. 3

The closest living fish relative of tetrapods

- The question of which living fish is the closest relative to **‘the fish that first crawled on to land’** has long captured our imagination: among scientists the odds have been placed on either **the lungfish** or **the coelacanth**.
- Analyses of small to moderate amounts of sequence data (ranging from 1 to 43 genes) has tended to favour the lungfishes.
- The alternative hypothesis that the lungfish and the coelacanth are equally closely related to the tetrapods could not be rejected.
- To perform a reliable analysis we selected 251 genes in which a 1:1 orthology ratio was clear
- The resulting phylogeny, based on 100,583 concatenated amino acid positions

PP 1.0
Jack-knife
100



The slowly evolving coelacanth

- Earlier work showed that Hox and protocadherins have comparatively slower protein-coding evolution in coelacanth than in other vertebrate lineages
- Protein-coding gene evolution was examined using the 251 concatenated proteins.
- Pair-wise distances between taxa were calculated from the branch lengths of the tree using the two-cluster test proposed previously to test for equality of average substitution rates
- The coelacanth proteins that were tested were **significantly more slowly evolving** (0.890 subst.per site) than the **lungfish** (1.05 subst. per site), **chicken** (1.09 subst. per site) and **mammalian** (1.21 subst. per site) orthologues. (P < 10⁻⁶ in all cases)

(Supplementary Data 5)

Transposable elements

- Theoretically, transposable elements may make their greatest contribution to the evolution of a species by generating templates for exaptation to form novel regulatory elements and exons, and by acting as substrates for genomic rearrangement
- The coelacanth genome contains a wide variety of transposable-element superfamilies and has a relatively high transposable-element content (25%).
- Analysis of RNA-seq data and of the divergence of individual transposable-element copies from consensus sequences show that **14 coelacanth transposable-element superfamilies are currently active.**

Supplementary Note 5 and Supplementary Tables 7–10
Supplementary Note 6, Supplementary Table 10 and Supplementary Fig. 5

The adaptation of vertebrates to land

- To understand the aspect of the water-to-land transition and how it has been reflected on genome level when vertebrates have lived on land over the 400 Myr.
- Genes that were present in the last common ancestor of all bony fish (including the coelacanth) but that are missing from tetrapod genomes.

Genes missing from tetrapod genomes

- More than 50 such genes, including components of
 - fibroblast growth factor (FGF) signalling,
 - TGF- β and bone morphogenic protein (BMP) signalling, and
 - WNT signalling pathways, etc...
- Functional annotation:
 - fin development (13 genes);
 - otolith and ear-development(8 genes);
 - kidney development(7 genes);
 - trunk, somite, and tail development(11 genes);
 - eye(13 genes);
 - brain development (23 genes).

(Supplementary Data 7 and Supplementary Fig. 9)

The water-to-land transition

- This implies **that critical characters** in the morphological transition **from water to land** (for example, fin-to-limb transition and remodelling of the ear) **are reflected in the loss of specific genes** along the phylogenetic branch leading to tetrapods.
- Homeobox **genes**, which are responsible for the development of an organism's **basic body plan**, show only **slight differences** between Latimeria, ray-finned fish and tetrapods;

Conserved non-coding elements (CNEs)

- Conserved non-coding elements (CNEs) are strong candidates for **gene regulatory elements**. They can act as promoters, enhancers, repressors and insulators, and have been implicated as major facilitators of evolutionary change.
- To identify CNEs that originated in **the most recent common ancestor** of tetrapods, they predicted CNEs that evolved in various bony vertebrate (that is, ray-finned fish, coela- canth and tetrapod) lineages and assigned them to their likely branch points of origin.

Conserved non-coding elements (CNEs)

- 44,200 ancestral tetrapod CNEs that originated after the divergence of the coelacanth lineage was identified.
- They represent 6% of the 739,597 CNEs that are under constraint in the bony vertebrate lineage.
- They compared **the ancestral tetrapod CNEs** to **mouse embryo** ChIP-seq (chromatin immunoprecipitation followed by sequencing). This resulted in **a sevenfold enrichment** in the p300 binding sites for our candidate CNEs and **confirmed that these CNEs are indeed enriched** for gene regulatory elements.

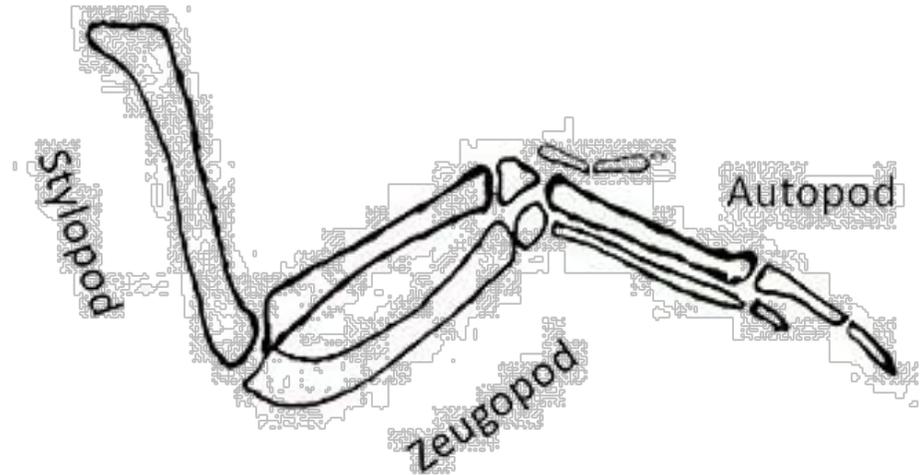
Ancestral tetrapod CNEs and GO annotation of closest genes

- Each tetrapod CNE was assigned to the gene whose transcription start site was closest, and gene-ontology (GO) category enrichment was calculated for those genes.
- **The most enriched categories** were involved with **smell perception** (for example, sensory perception of smell, detection of chemical stimulus and olfactory receptor activity).
- morphogenesis (radial pattern formation, hind limb morphogenesis, kidney morphogenesis)
- cell differentiation (endothelial cell fate commitment, epithelial cell fate commitment)
- immunoglobulin VDJ recombination

also,

Limbs development

- A major innovation of tetrapods is the evolution of limbs characterized by digits. The limb skeleton consists of:
 - a stylopod
 - the zeugopod
 - an autopod

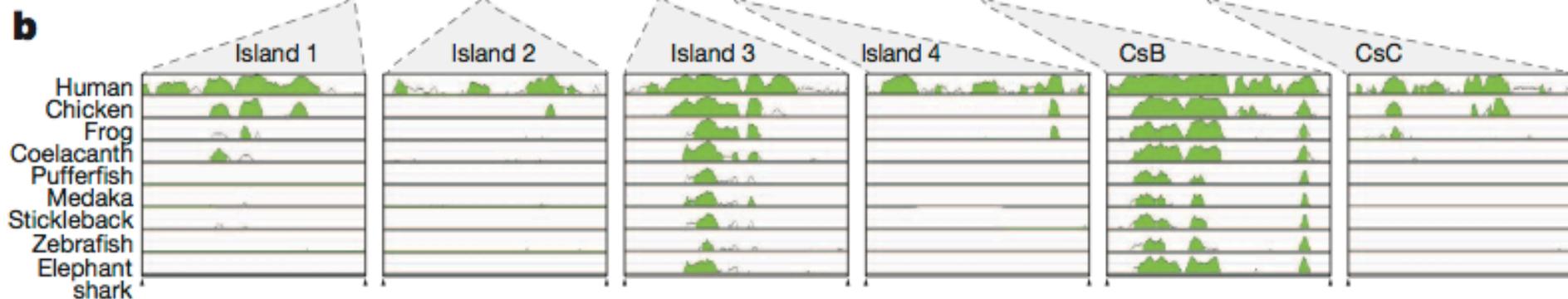
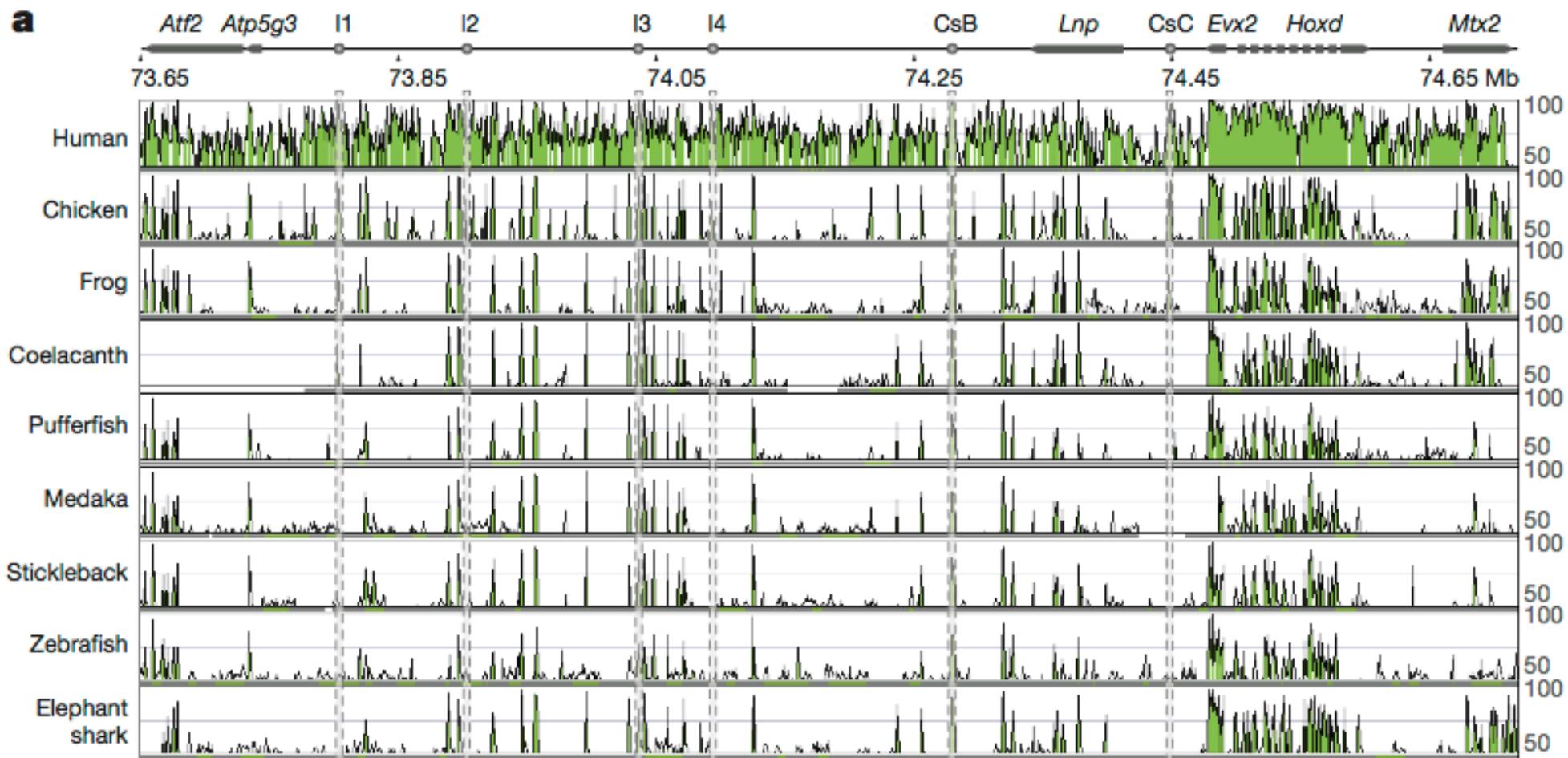


There are two major hypotheses about the origins of the autopod;

- that it was a novel feature of tetrapods,
- that it has antecedents in the fins of fish

The HOX-D cluster

- They examined the Hox regulation of limb development in ray-finned fish, coelacanth and tetrapods to address these hypotheses.
- Among the **six cis-regulatory sequences** previously identified in this gene desert, **three** sequences show sequence **conservation restricted to tetrapods**.
- One regulatory sequence (**island 1**) is **shared by tetrapods and coelacanth**, but not by ray-finned fish.



Experiment

- When tested in a transient transgenic assay in mouse, the coelacanth sequence of **island 1** was able to drive reporter expression in a limb-specific pattern

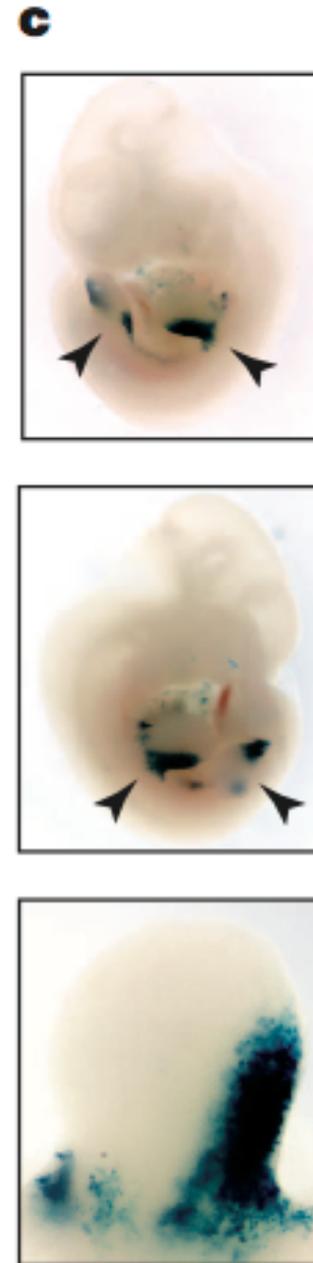


Figure 2C. An upstream gene desert identifies conserved limb enhancers.

Amemiya C.T. et al. Nature 2013

- This suggests that island 1 was a lobe-fin developmental enhancer in the fish ancestor of tetrapods that was then coopted into the autopod enhancer of modern tetrapods.
- In this case, the autopod developmental regulation was derived from an ancestral lobe-finned fish regulatory element.

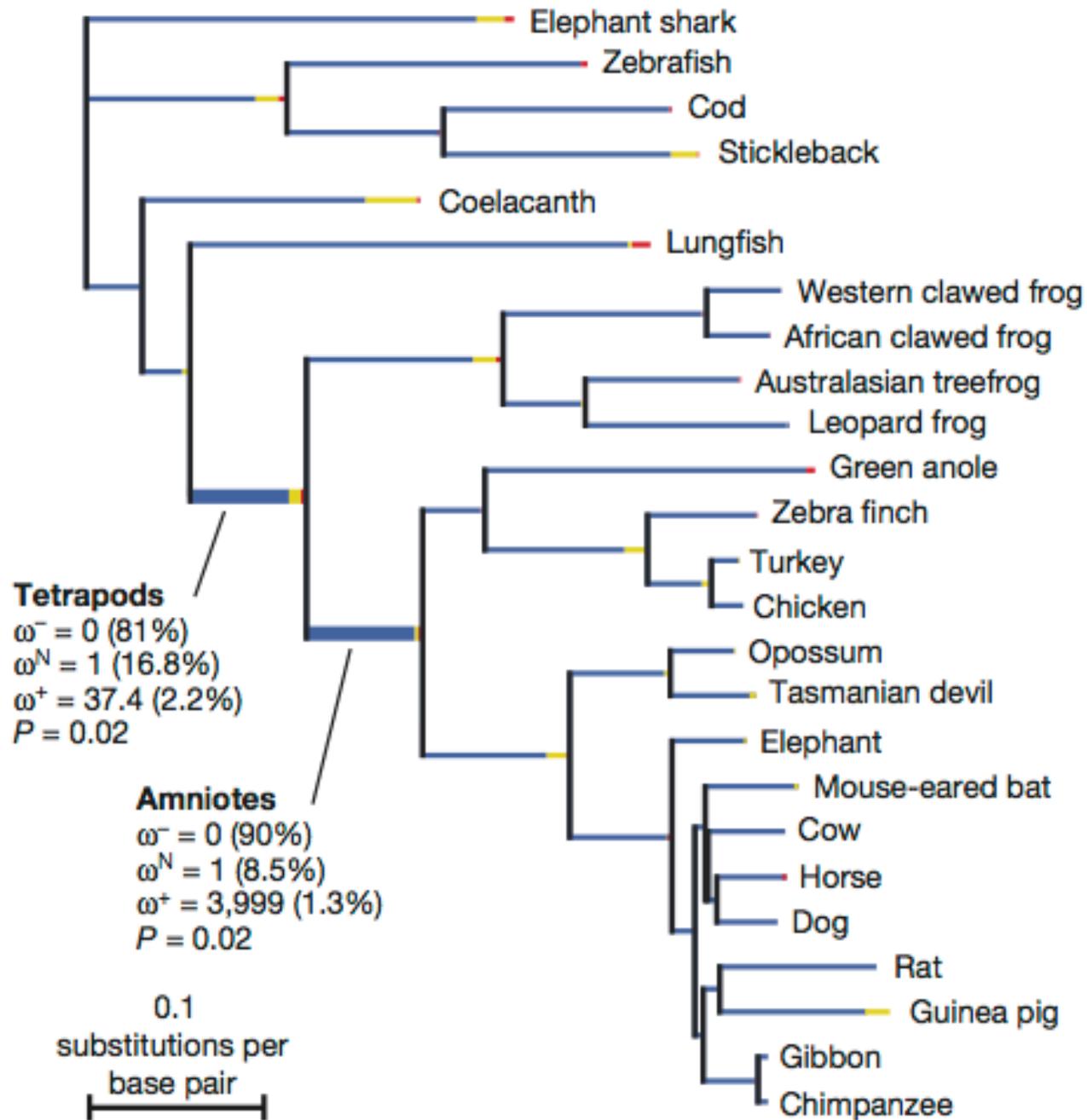
Changes in the urea cycle

Liigest lammastikust vabanemine:

- In aquatic environments >> ammonia
- on land >> urea or uric acid
- **The widespread and almost exclusive occurrence of urea excretion in amphibians, some turtles and mammals has led to the hypothesis that the use of urea as the main nitrogenous waste product was a key innovation in the vertebrate transition from water to land.**

Changes in the urea cycle

- HYPHY package, which estimates the ratio of synonymous (dS) to non-synonymous (dN) substitutions (ω values) among different branches and among different sites (codons) across a multiple-species sequence alignment.
- The **rate-limiting enzyme** of the hepatic urea cycle, **carbamoyl phosphate synthase I (CPS1)**, **only one branch** of the tree shows a strong signature of selection ($P = 0.02$), namely **the branch leading to tetrapods and the branch leading to amniotes**.



Conclusions

- Genome is assembled with a reasonable quality and annotated well for whole genome comparisons with other genomes
- Slow evolution speed is observable in the case of selected (251) coding genes. However, analyzing other regions do not support the "living fossil" at genome level.
- The adaptation of vertebrates to land
 - 50 genes were lost in tetrapode lineages
 - Many genes were classified in important developmental categories
 - Changes in the urea cycle
 - Changes in the physiological environment of the developing embryo and fetus
- The coelacanth lacks immunoglobulin-M (they have Ig-W)

The adaptation of vertebrates to land

- More than 50 genes were lost in tetrapode lineages
 - fin development (13 genes); otolith and ear development (8 genes); kidney development (7 genes); trunk, somite and tail development (11 genes); eye (13 genes); and brain development (23 genes).
- Conserved non-coding elements (CNEs)
 - identified 44,200 ancestral tetrapod CNEs that originated after the divergence of the coelacanth lineage.
- Changes in the urea cycle
- changes in the physiological environment of the developing embryo and fetus
- The coelacanth lacks immunoglobulin-M