The sea lamprey (*Petromyzon marinus*) is a member of an ancient vertebrate lineage that diverged from our own lineage approximately 500 million years ago, before the evolution of jaws and paired appendages.

The lamprey genome: Illuminating genomic change across eons and embryogenesis

Jeramiah J Smith, Courtney KM Waterbury, Melissa C Keinath, Cody B Saraceno, Vladimir A Timoshevskiy, Nataliya Y Timoshevskaya

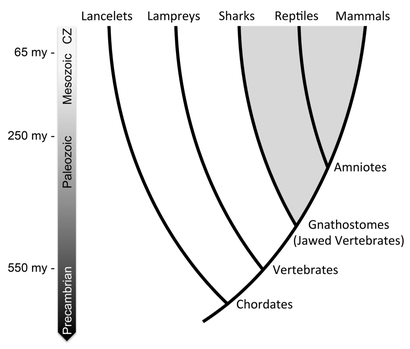
University of Kentucky, Department of Biology

Abstract

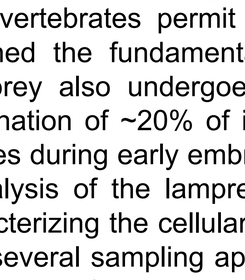
The lamprey genome provides unique insights into both the deep evolutionary history of vertebrate genomes and the maintenance of genome structure/integrity over development. The lamprey lineage diverged from all other vertebrates approximately 500 million years ago. As such, comparisons between lamprey and other vertebrates permit reconstruction of ancient duplication and rearrangement events that defined the fundamental architecture and gene content of all extant vertebrate genomes. Lamprey also undergoes programmatic changes genome structure that result in the physical elimination of ~20% of its genomic DNA (~0.5Gb from a ~2 Gb genome) from all somatic cell lineages during early embryonic development. Here, we outline recent progress in assembly and analysis of the lamprey germline genome, and progress in the development of methods for characterizing the cellular events that mediate DNA elimination. We have integrated information from several sampling approaches and sequencing technologies to achieve a highly contiguous assembly of lamprey genome (including: Illumina fragments/mate pairs, 20X coverage in Pacific Biosciences reads, dense meiotic maps and optical mapping data). This genome assembly has dramatically improved our ability to dissect the molecular basis and genetic outcomes of programmed genome rearrangements (PGRs), and has improved our understanding of the tempo and mode of large-scale duplications and translocations within the ancestral vertebrate lineage. Analysis of the germline genome identifies several genes that are expressed in germline but physically eliminated from all somatic tissues. These eliminated genes correspond to several known oncogenes and appear to identify several other novel oncogene candidates. Complementing this assembly, the development of approaches to in situ analysis of 3D preserved cells has revealed that PGR unfolds through a series of dramatic cellular events that involve the programmatic alteration of several fundamental mechanisms of genome maintenance, including: alignment of chromosomes at metaphase, chromatid cohesion, separation and segregation, and nuclear envelope formation.

The Lamprey Genome

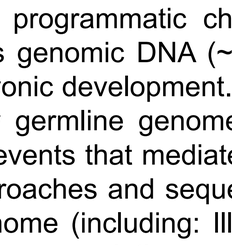
Deep Evolutionary Lineage



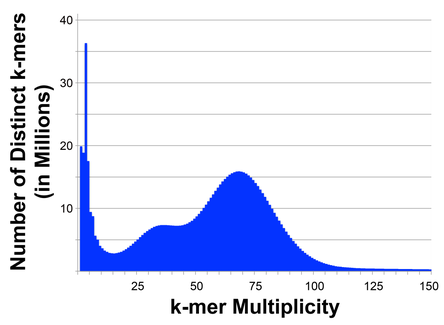
1N = 99 Chromosomes



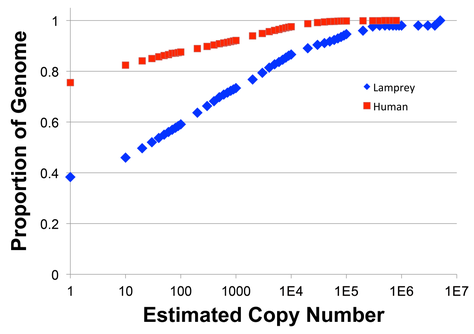
1N = 84



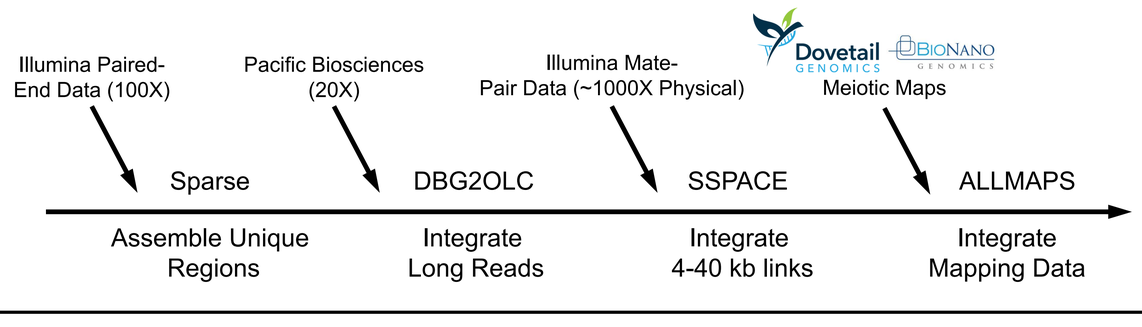
Highly Polymorphic



Highly Repetitive

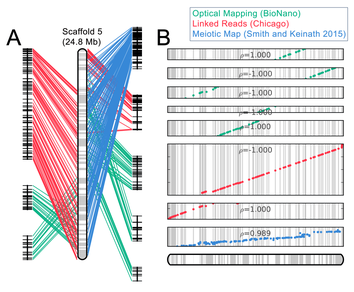


Hierarchical Assembly Pipeline

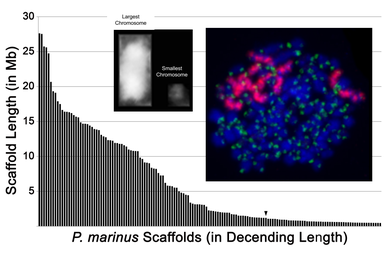


Large-Scale Contiguity of the Lamprey Assembly

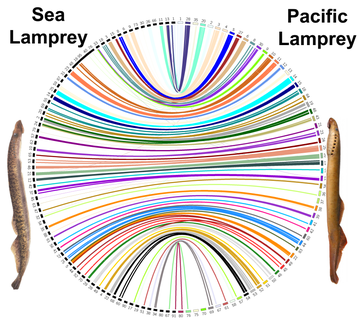
Consensus Scaffolding



Scaffolds Approximate Chromosomes



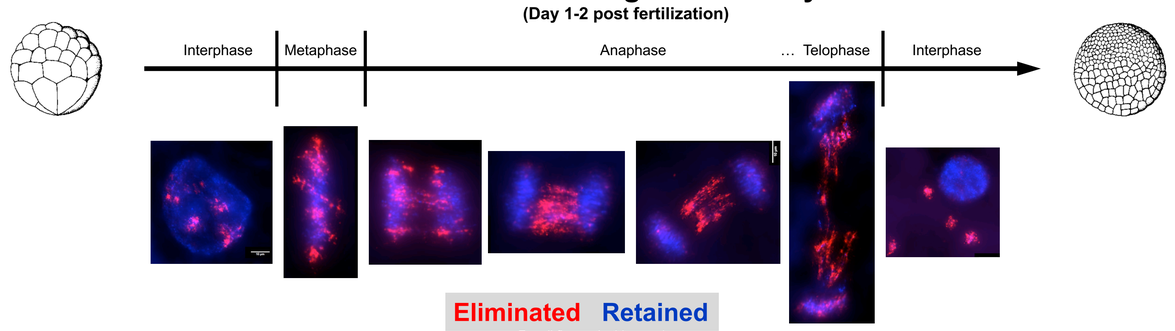
Chromosome-Scale Conservation



Programmed Genome Rearrangement

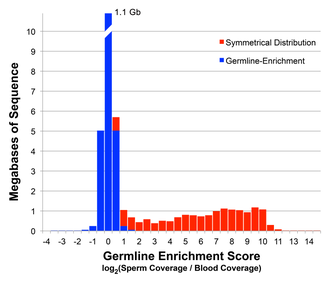
DNA Loss Through the Cell Cycle

(Day 1-2 post fertilization)

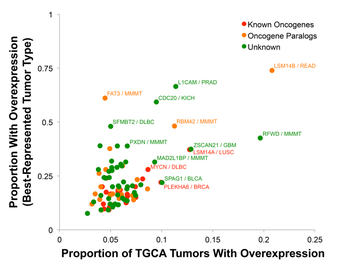


For all figures in this panel

Eliminated Single-Copy Regions

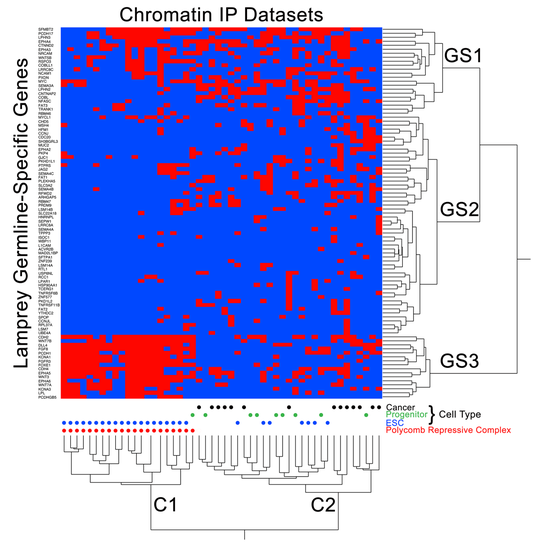


Expression of Eliminated Homologs in Human Cancer



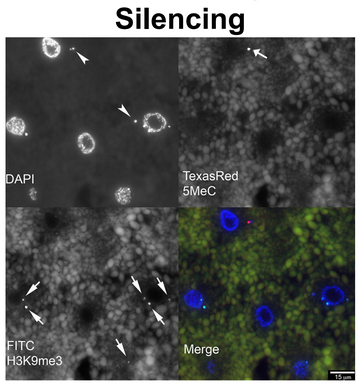
Functions of Eliminated Genes

Chromatin IP Datasets

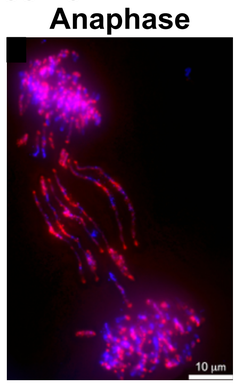


PGR and Epigenetic Silencing

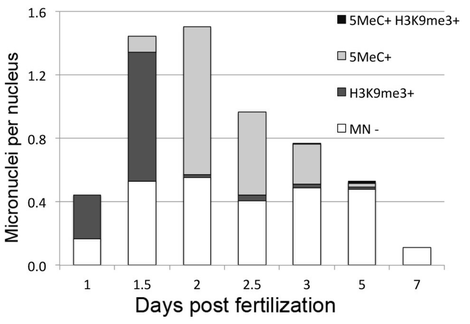
Overlap With Epigenetic Silencing



Lagging Chromatin in Anaphase

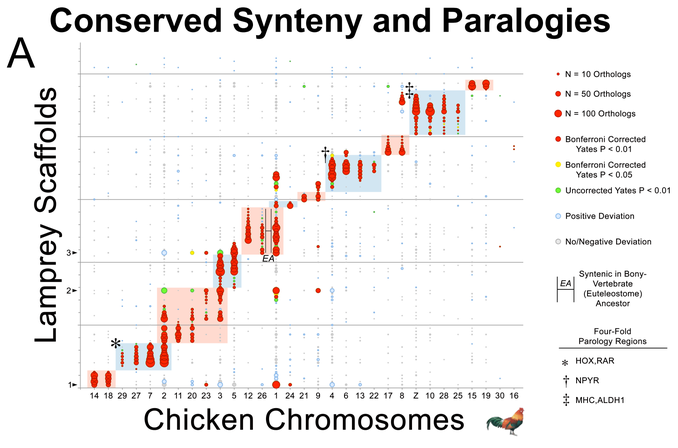


Dynamic Changes During Development

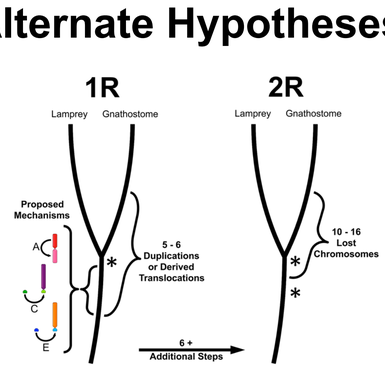


Genome Evolution, Duplication, and Comparative Studies

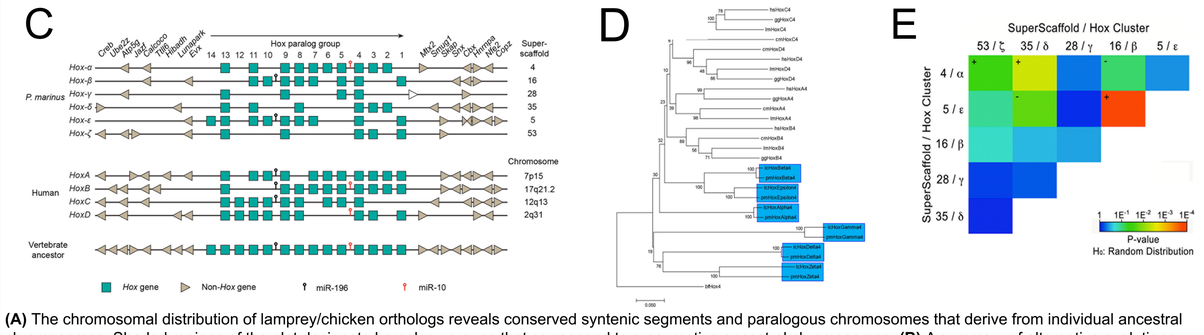
Conserved Synteny and Paralogies



B Alternate Hypotheses



Evolution of Hox clusters



(A) The chromosomal distribution of lamprey/chicken orthologs reveals conserved syntenic segments and paralogous chromosomes that derive from individual ancestral chromosomes. Shaded regions of the plot designate homology groups that correspond to presumptive ancestral chromosomes. **(B)** A summary of alternative evolutionary models explaining the distribution of paralogous segments in gnathostome genomes. Asterisks denote whole genome duplication events that have been proposed under two alternate evolutionary models. Mechanisms underlying three pre-R1 duplications are depicted under the classical "1R" model. **(C)** Six hox clusters can be identified within the sea lamprey genome assembly. Cluster designations α through ζ follow the convention of Mehta et al. **(D)** The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 3.88610197 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches. **(E)** Tests for enrichment of 2-copy duplicates among all pairs of HOX-bearing chromosomes (SuperScaffolds). Colors correspond to the degree to which the counts of shared duplicates on each pair of chromosomes deviates from the expected value given an identical number of chromosomes and paralogs retained on each chromosome. Plus and minus symbols indicate the direction of deviation from expected for chromosome pairs with P<0.1.