Jawless Fishes of the World

Jawless Fishes of the World:

Volume 1

Edited by

Alexei Orlov and Richard Beamish

Cambridge
Scholars
Publishing

Jawless Fishes of the World: Volume 1

Edited by Alexei Orlov and Richard Beamish

This book first published 2016

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-4438-8582-7
ISBN (13): 978-1-4438-8582-9

TABLE OF CONTENTS

Volume 1

Part 1: Evolution, Phylogeny, Diversity, and Taxonomy

Part 2: Ecology and Life History

Volume 2

Part 3: Demography, Stock Assessment, Fisheries and Conservation

Part 4: Species Interactions

PREFACE

Hagfishes and lampreys, the extant jawless fishes of the world, are rare gems. With approximately 110–120 described species (roughly 70–80 hagfish and at least 40 lamprey species), they represent less than 0.2% of all extant vertebrates. However, as sole survivors at the base of the vertebrate family tree (surviving at least four of the five mass extinction events since the Cambrian explosion), these two lineages—dating back some 400–500 million years—can tell us much about the evolution of early vertebrates. The origin of vertebrates represents one of the major jumps in animal evolution, with the active, sentient vertebrates being distinguished from the largely sessile non-vertebrate chordates (the lancelets and tunicates) by a suite of major innovations (e.g., a cranium and pronounced cephalization; paired sense organs; an axial skeleton and muscle segmentation; more complex circulatory, respiratory, digestive, and endocrine systems; and a glomerular kidney). Given the poor fossil record, hagfishes and lampreys—so-called living fossils—are helping us understand the events that occurred at the dawn of vertebrate evolution.

Whether the extant jawless fishes are monophyletic (with hagfishes and lampreys as each other's closest living relative) or paraphyletic (with hagfishes representing an earlier offshoot from the vertebrate family tree and lampreys sharing an ancestor more recently with the jawed vertebrates) has been the subject of much debate. Although recent molecular evidence strongly supports hagfishes and lampreys as a monophyletic group, it is nevertheless important to recognize the long independent evolutionary histories of each lineage and their significant morphological, physiological, and ecological differences. Hagfishes are exclusively marine (and, unique among vertebrates, are isomotic with the environment), lack image-forming eyes and even remnant vertebral arches, and show direct development. Lampreys, in contrast, were the first extant vertebrates to invade fresh water (and the anadromous species are able to osmoregulate in both fresh and salt water), have well-developed eyes as adults and rudimentary vertebral arches, and show indirect development (undergoing a dramatic metamorphosis after the prolonged larval stage).

Hagfishes and lampreys are also being used increasingly in biomedical and biomimetic research. For example, the slime of hagfishes—one of the traits that make them so reviled by fishermen—is a source of inspiration for the production of extra strong fabrics and petroleum-free plastics. The protein threads of hagfish slime are finer than spider silk (and 100 times thinner than human hair) but just as strong. Hagfish slime (or a synthetic gel containing its fibers) is also being explored for treating accident victims or use during surgery; since it expands when it contacts blood (i.e., saline water), it could potentially be used to stop blood flow. In New Zealand, hagfish slime is used by the Maori as a cleaning agent. The lamprey central nervous system has long been used as a model in neurological studies, and lampreys—since they are unique among known vertebrates in their ability to recover nearly full function after complete spinal cord transection—are being used extensively in research into spinal cord regeneration. The relatively simple and well-studied neural networks of lampreys have also led to the development of lamprey-inspired biorobotic systems for studying high level motor tasks. In addition, the buccal gland secretions of parasitic lampreys are being investigated as a potential source of novel anti-coagulants (comparable to hirudin from medicinal leeches), local anaesthetics, and immunosuppressants, and the ability of lampreys to survive the programmed loss of the gall bladder and biliary tree during metamorphosis and to store and tolerate high concentrations of iron in various body tissues make them excellent model organisms for research into treatment for cholestasis and hemochromatosis in humans.

Despite their incredible scientific value, however, hagfishes and lampreys are generally underappreciated (or actively loathed) by the general public. Sometimes referred to as “scavengers of the deep” and “vampires of the deep,” respectively, hagfishes and lampreys are often considered pests of commercial food fisheries—hagfishes given their tendency to feed on commercially-valuable fish caught on lines and in pots; lampreys due largely to the decimation of the commercial and recreational fisheries in the Great Lakes following the invasion and spread of the sea lamprey. More respect is warranted, of course, particularly given the important ecosystem services that hagfishes and lampreys perform. The burrowing and scavenging behavior of hagfishes helps in the turnover of substrate and recycling of organic material in deep-sea environments (so that “earthworms of the deep” might be a better name for them), and they are important food items for pinnipeds and other marine mammals. Similarly, larval lampreys are important in nutrient cycling in rivers and streams, the carcasses of spent anadromous lampreys are a significant source of marine-derived nutrients in freshwater systems, and lampreys—during all stages—are important food sources for other animals. Lampreys are also ecosystem engineers; the burrowing and feeding activities of larval lampreys significantly increase substrate oxygen levels and the nest-building activity of spawning lampreys increases streambed complexity in ways that appear to benefit other organisms.

When hagfishes and lampreys are appreciated, they are often overexploited. Traditionally, hagfishes were fished in Asia for their meat—for consumption by humans and domestic animals and for use as bait. These fisheries expanded in Japan in the 1940s, when other fish stocks were being depleted, and in Korea in the 1980s—particularly following increased demand for hagfish skin, used for "eelskin" leather products. After local Asian stocks were exhausted, hagfish fisheries expanded into North America and New Zealand, followed, unfortunately, by serial depletion of most of these stocks. Lampreys (despite the bad reputation now given to them by sea lamprey in the Great Lakes) have also been long appreciated as food by many cultures (e.g., the Maori, Native Americans in the Columbia River Basin, and throughout much of Europe and Asia). Traditional and commercial fisheries continue throughout many of these areas, although lamprey numbers have declined in recent decades, due largely to the effects of industrialization and urbanization, and some fisheries (e.g., for Caspian lamprey in Russia and Azerbaijan) are no longer viable.

Many hagfish and lamprey species are now of conservation concern. On the IUCN (International Union for the Conservation of Nature) Red List of Threatened Species (as of July 2015), 24 hagfish and 32 lamprey species have been evaluated (approximately 32 and 78% of described hagfish and lamprey species, respectively). Of the 24 hagfish species evaluated, three are considered at some risk of extinction (Critically Endangered 1, Vulnerable 1, Near Threatened 1), 15 are listed as Least Concern, and six are considered Data Deficient (i.e., with insufficient information to assess risk of extinction). Of the 32 generally recognized lamprey species evaluated, eight are considered at some risk of extinction (Critically Endangered 2, Endangered 1, Vulnerable 3, Near Threatened 2), 20 are listed as Least Concern, and four are considered Data Deficient. There is a rapidly growing body of research, therefore, being conducted in support of conservation and management. As outlined below, many of the chapters in this book relate to the conservation needs of hagfishes and lampreys.

Beyond their value in evolutionary and biomedical research, their ecosystem services, and their economic value, hagfishes and lampreys have always garnered respect and study from at least a small dedicated group of scientists interested in these fascinating creatures for their own sake. Many of the biomedical applications, for example, are offshoots of such basic research, and hagfishes and lampreys continue to astonish and amaze us. Studies using baited video cameras, for example, have captured images showing hagfishes actively hunting and preying on other fishes (suggesting that their ecological role in deep-sea ecosystems is more diverse than previously thought) and others showing the instantaneous effectiveness of hagfish slime at deterring predators; in less than half a second of being grasped by a shark or other fish, jets of slime filled the mouth and gill chamber of the predator, causing it to visibly choke and move away. A study looking at coiling behavior in Pacific hagfish demonstrated that individuals showed significant “handedness,” that is, individuals preferentially coiled in a clockwise or counterclockwise direction. Recent discoveries in lampreys include finding a chemical alarm cue in sea lamprey (reminiscent of the well-studied "Schreckstoff" found in cyprinids) that is released into the water when a conspecific is injured, eliciting avoidance behavior. Also interesting is the discovery of adipose tissue in the dorsal rope of sexually-mature male sea lamprey that produces heat when the male is in the presence of an ovulated female; this is the only known example of a thermogenic fat in a non-mammal. Interestingly, handedness in lampreys (in terms of both attachment to a host during parasitic feeding and mating behavior) has long been observed. Hagfishes and lampreys are also physiological wonders: Pacific hagfish are able to tolerate and recover from exposure to high concentrations of ammonia while burrowed inside their meals of decomposing carcasses, and anadromous sea lamprey have a high capacity for urea excretion that enables them to feed on basking sharks (the world’s second largest fish) and other ureosmotic elasmobranchs.

The growing appreciation for hagfishes and lampreys is indicated by the number of special symposia and publications dedicated to these jawless fishes in recent years, including this book, which is the main output from a one-day symposium at the 2014 American Fisheries Society annual meeting. *Jawless Fishes of the World* is a truly international effort, including contributions from authors from four continents (Asia, Europe, North America, South America) and 13 countries (Brazil, Canada, Chile, Costa Rica, Czech Republic, Germany, Greece, Japan, Mexico, Portugal, Russia, Spain, and the United States of America). Consisting of 25 chapters organized into four sections, *Jawless Fishes of the World* provides a wealth of information on a broad range of topics, and includes chapters on some of the better-studied wide-ranging species (e.g., Pacific hagfish and anadromous sea, Arctic, and Pacific lampreys) and many of the less well-known species (e.g., lamprey species with more limited freshwater distributions). It is exciting to see such diversity included in a single volume and such attention paid to traditionally “non-charismatic” species.

Part 1 of *Jawless Fishes of the World*—which focuses on the evolution, phylogeny, diversity, and taxonomy of hagfishes and lampreys—shows that there has been no waning of interest in the evolution of these ancestral vertebrates. The first chapter, for example, takes advantage of the now-available sea lamprey and Arctic lamprey genomes to explore the timing of two large-scale genome duplication events that occurred sometime during the evolution of early vertebrates—and that likely permitted the large number of vertebrate-specific innovations mentioned above. Several other chapters in this section tackle hagfish and lamprey taxonomy; taxonomy is challenging because, compared to bony fishes (with their vertebrae, scales, fin rays, and many other meristic and morphometric characters), there are relatively few taxon-distinctive morphological characters in hagfishes and lampreys. Furthermore, given the morphological conservation seen in these lineages over long periods of evolutionary time (modern hagfishes and lampreys look eerily similar to those captured in the fossil record more than 300 million years ago), it is not surprising that extant species—that diverged from each other so comparatively recently—are difficult to distinguish. However, the chapter describing the use of gill pore papilla number and structure for resolving lamprey phylogeny sounds like the sort of thing that J. D. McPhail and C. C. Lindsey had in mind in 1970 when they suggested in their *Freshwater Fishes of Northwestern Canada and Alaska* that “an imaginative search for new taxonomic characters in lampreys might be fruitful.” Molecular characters, particularly mitochondrial DNA sequence, have also helped to resolve lamprey phylogenies, showing, for example, that some species (e.g., *Lethenteron* sp. S)—despite being morphologically indistinguishable from other species—represent divergent evolutionary lineages. Other chapters in this section cover the taxonomy and diversity of several less well-known species or populations, including: hagfishes from Mexico, Central America, and Chile; the pouched and short-headed lampreys in Chile (representatives of the two Southern Hemisphere families which receive more attention in Australia and New Zealand); and the West Transcaucasian and Alaskan brook lampreys (which include the first descriptions of the adult and larval stages, respectively). There has also been no diminution of interest in lamprey “paired species,” a term first coined in 1959 by G. Zanandrea, although J. C. C. Loman made the observation more than 100 years ago that European river and brook lampreys are morphologically similar (particularly as larvae) but the non-parasitic brook lamprey delays metamorphosis and accelerates sexual maturation relative to the river lamprey (which “inserts” a parasitic feeding phase between metamorphosis and sexual maturation). Investigation of paired species has been reinvigorated in recent years with the advent of molecular DNA techniques, which have shown that many paired species are “barcode indistinguishable” and that they sometimes—but not always—show evidence of contemporary gene flow when sympatric. As explored in two chapters in this section, this has renewed the debate regarding whether paired species are “real” (i.e., distinct) species. The population of western brook lamprey in Morrison Creek on Vancouver Island that is able to produce some individuals capable of feeding parasitically adds to this fascinating debate.

Part 2 deals with the ecology and life history of lampreys, the study of which has generally been more feasible than that of deep-sea hagfishes. In lampreys, habitat requirements of the prolonged freshwater larval stage have been relatively well-characterized, although the secretive nature of the larvae makes accurate population assessment difficult. Far less is known about the parasitic feeding phase, particularly in anadromous species; this is a theme also echoed throughout Parts 3 and 4. The complex and finely-orchestrated life history of migratory lampreys, which experience different habitat requirements at each stage and well-timed movements between them, makes them particularly vulnerable to environmental perturbation (e.g., climate change, dams that act as barriers to migration and alter natural stream flow regimes). Chapters in this section describe advances in electrofishing methods for assessing and monitoring larval lamprey populations, a mark-recapture study that helps fill gaps in our knowledge about the oceanic phase of the anadromous sea lamprey, and investigations into the potential effects of climate and anthropogenic changes on life history of the anadromous Pacific lamprey. Also of great interest is the report of landlocked sea lamprey in Spain, where some post-metamorphic feeding was observed in the Portodemouros Reservoir but no individuals were found to complete their life cycle in fresh water—presumably due to a limited prey base. Given the apparent ease with which sea lamprey colonized the Great Lakes, it is important to understand that it may not be osmoregulatory constraints that generally restrict sea lamprey populations to feeding at sea; predicted climate-driven increases in the productivity of fresh water relative to marine systems (as well as stocking of lakes and reservoirs with host fishes) could potentially permit anadromous lampreys to become invasive in fresh water.

Part 3 focuses on research in support of hagfish and lamprey conservation and fisheries management. As mentioned above, knowledge of the basic biology and demography of deep-sea hagfishes and the oceanic feeding phase of anadromous lampreys is generally lacking—dangerously so in some species. Chapters in this section review the conservation status of lampreys in central Europe; present key baseline information gathered from an exploratory Pacific hagfish fishery off the coast of Baja California in Mexico; and shed more light on the enigmatic marine phase of Arctic, Pacific, and western river lampreys, that is, on their distribution (both geographically and in the water column), population trends, and their potential effect on (and dependence on) host fish populations. Two other chapters outline recent advances in the artificial propagation and conservation genetics and genomics of Pacific lamprey in the Columbia River basin. As is being done with European river lamprey (e.g., in Finland, Estonia, and Latvia), methods are being developed to rear Pacific lamprey larvae in hatcheries as part of a supplementation program to reintroduce or augment locally or functionally extinct populations. Rearing strategies to help get more early-stage larvae through the "survival bottleneck" would also be highly valuable to researchers interested in heritability of feeding type in paired species and, among other things, the relative contribution of genetic versus environmental factors in lamprey sex determination. The recent technological advancements in conservation genomics is helping to identify management units (given that lampreys do not home to their natal streams and thus do not form populations that can be easily delineated based on geographic location), and predict and monitor the success of adult translocations.

Part 4 explores the interactions between parasitic lampreys and their hosts. This section includes fascinating insights (from the late Philip Cochran—a rare gem himself—and his long-time collaborator John Lyons) into the ancient predator-prey relationship between silver lamprey and American paddlefish in the Mississippi River drainage. The silver lamprey is not a recent invader to fresh water, and it and the paddlefish appear to have shared a long evolutionary history living in "peaceful co-existence." The paddlefish is unlikely to be much affected by one or a few parasitic silver lamprey (given its large body size and its breaching behavior); the silver lamprey, in turn, has a few tricks of its own "up its sleeve" (e.g., a relatively larger oral disc where its co-occurs with paddlefish and attachment sites within the paddlefish's branchial cavity that presumably provide some protection from dislodgement when the paddlefish breaches). The silver lamprey-paddlefish "story" serves as a good counterpoint to the negative image of blood-sucking sea lamprey in the Great Lakes.

Hagfishes and lampreys have survived—as important components of their respective (and dynamic) ecosystems—for hundreds of millions of years. Over evolutionary time, hagfishes—living in predominantly deep-sea habitats—were likely protected from the periodic upheaval experienced in shallow sea and terrestrial ecosystems. Lampreys, as generalists, also managed to persist when other—perhaps more specialized—lineages perished. For example, since lampreys appear not to be philopatric (locating spawning tributaries by attraction to larval pheromones, a good indicator of contemporary rather than historical larval rearing habitat quality), they are likely less finely tuned to local environmental conditions than species that exhibit strong site fidelity. That parasitic species lack strong host specificity (more than 50 host species, for example, have been reported for anadromous sea lamprey) means that they are less likely to be impacted from changes in the distribution and abundance of individual host species; "facultative ureotelism" even allows them to feed on basking and other large sharks! "Facultative parasitism" or at least the apparent ease with which lamprey feeding and migratory type can evolve (e.g., allowing populations to "abandon" the parasitic feeding phase and associated migrations when the substantial mortality incurred during this stage is no longer compensated for by the opportunity to attain a larger size and higher fecundity) may also allow lampreys "to keep their options open" under changing conditions. The "staying power" exhibited by hagfishes and lampreys to date, however, should not allow us to be complacent; although these two lineages have survived through four mass extinction events, individual species did not. Research in support of the management and conservation of the extant jawless fishes of the world—those species that are commercially valuable and those that are not—therefore needs to continue.

Margaret F. Docker
Department of Biological Sciences
University of Manitoba
Winnipeg, Manitoba, Canada

PART 1:

EVOLUTION, PHYLOGENY, DIVERSITY, AND TAXONOMY

CHAPTER ONE

MOLECULAR EVOLUTION IN THE LAMPREY GENOMES AND ITS RELEVANCE TO THE TIMING OF WHOLE GENOME DUPLICATIONS

TEREZA MANOUSAKI, HUAN QIU,
MIYUKI NORO, FALK HILDEBRAND,
AXEL MEYER AND SHIGEHIRO KURAKU

Background

The genomes of two lamprey species have been sequenced, and this has provided the basis for genome-wide comparison of molecular evolution between jawless fishes and the rest of vertebrates. Molecular phylogenetic analyses of jawless fish genes increased our knowledge of the evolutionary time scale of diversification of hagfishes and lampreys, as well as of gene redundancy in their genomes. It was shown that the ancestor of jawed vertebrates experienced two rounds of whole genome duplications (Dehal & Boore, 2005). However, it has been controversial whether this event occurred before or after the ancestors of extant jawless fishes diverged from the lineage which gave rise to jawed vertebrates (e.g., Escriva et al., 2002; reviewed in Kuraku, 2013). Recent molecular phylogenetic studies showed that the whole genome duplications occurred before the radiation of all extant vertebrates including hagfishes and lampreys (Kuraku et al., 2009a), and this scenario has been confirmed by later studies including the genomic analysis of *Petromyzon marinus* (sea lamprey) (Hoffmann et al., 2010; Smith et al., 2013). In this chapter, we analyze peculiar characteristics of the lamprey genomes, focusing mainly on protein-coding regions, to propose potential factors that act as barriers to the understanding of the timing of whole genome duplications.

Jawless fish in molecular phylogenetics

Extant jawless fishes, also called cyclostomes, are comprised of hagfishes and lampreys and diverged from the stem lineage that gave rise to jawed vertebrates (gnathostomes) about 600-500 million years ago (Kuraku et al., 2009b; Figure 1-1). They have been proven to be a monophyletic group, based on molecular phylogenies of both mitochondrial and nuclear genes (reviewed in Kuraku, 2008; Figure 1-1). So far, the so-called ‘phylogenomics’ approach, involving more than hundreds of genes on nuclear genomes (Kumar et al., 2012), has not been applied to elucidate the relationships between hagfish, lamprey and jawed vertebrates in such a high resolution as demonstrated for other long-standing questions in animal phylogeny (e.g. Misof et al., 2014). This is mainly due to the lack of large-scale sequence information for hagfish (see Delsuc et al., 2006).

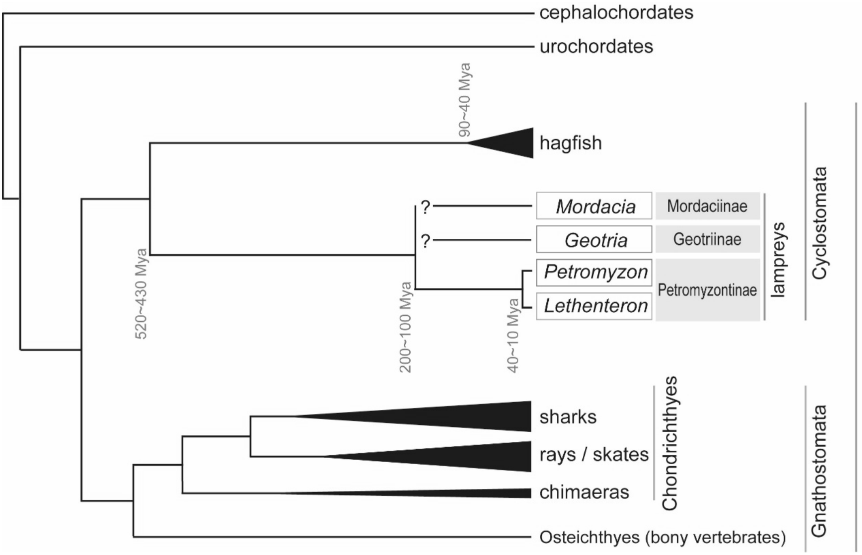

Figure 1-1. Overview of the lamprey phylogeny based on molecular data. Evolutionary time scale within the cyclostome lineage is based on previous literature (Kuraku & Kuratani, 2006; Kuraku et al., 2009b). Branch lengths in the other lineages roughly correspond to evolutionary times inferred from molecular data (Hedges et al., 2006). The phylogenetic relationships among the Mordaciinae, Geotriinae and Petromyzontinae remain to be carefully analyzed with multiple genes (reviewed in Kuraku, 2008).

Lamprey genome sequencing

The first jawless fish species for which whole genome sequencing was started was the sea lamprey *Petromyzon marinus*. Its genome was sequenced with the so-called Sanger method using DNA extracted from the adult liver. An early version of the genome assembly was made public in 2007 at the UCSC Genome Browser (http://genome.ucsc.edu/), which is still available there as version 3 (petMar1; http://hgdownload.soe.ucsc.edu/goldenPath/petMar1/bigZips/). Later, an improved genome assembly, designated version 7 (petMar2), was generated and used as the final product in the genome-wide analysis by the genome consortium (Smith et al., 2013). In the meantime, it was reported that the sea lamprey experiences programmed genomic rearrangement (PGR) in somatic cell lineages (Smith et al., 2009; Smith et al., 2012). It is thus likely that DNA extracted from the source material for the whole genome sequencing was incomplete and heterogeneous, derived from a mixture of somatic cells with differentially rearranged genomes. In 2013, the genome assembly of another northern hemisphere species, the Arctic lamprey *Lethenteron camtschaticum* (formerly known as *L. japonicum*), based on Roche 454 sequencing platform, was also released (Mehta et al., 2013; http://jlampreygenome.imcb.a-star.edu.sg/). This *L. camtschaticum* project employed genomic DNA extracted from the mature testis, which could have contributed to a larger size and higher continuity of the genome assembly because germline cells possess the intact genome (Table 1-1). This project focused on the evolutionary history of *Hox* gene clusters – a long-standing theme regarding cyclostome gene phylogeny (reviewed in Kuraku, 2011; Kuraku and Meyer, 2009). To provide a comparison of assembly statistics between these two genomes, we recomputed basic metrics (see Bradnam et al., 2013) using the latest genome scaffold sequences downloaded from Ensembl (for *P. marinus*) and NCBI (for *L. camtschaticum*) (Table 1-1).

Table 1-1. Assembly statistics of the two lamprey genomes.

Species	**Assembly ID**	**Total # of bases (Gbp)**	**# of scaffolds**	**Scaffold N50 (Kbp)**	**% N**	**Max scaffold length (Kbp)**	**Min scaffold length (Kbp)**
Petromyzon marinus	petMar2	0.886	25,006	79.7	26.8	3,631	0.201
Lethenteron camtschaticum	LetJap1	1.031	86,125	923.6	17.2	11,640	0.867

The statistics in this table are based on computations using all publicly available scaffolds and are partly different from those in the respective publications reporting the genome analyses.

To analyze completeness of protein-coding landscape of the two lamprey genomes, we used a program pipeline CEGMA which reports the number of detected genes among 248 conserved genes (CEG, core eukaryotic genes) (Parra et al., 2009). As a result, the *L. camtschaticum* genome assembly was shown to cover 80% (199/248) of the CEGs, while 69% was detected in the *P. marinus* genome (172/248) (Table 1-2). This suggests that the *L. camtschaticum* genome assembly covers more protein-coding genes. The genome consortia of *P. marinus* and *L. camtschaticum* reported 26,046 and 17,829 protein-coding genes, respectively (Smith et al., 2013; http://jlampreygenome.imcb.a-star.edu.sg/). These two species are thought to have diverged relatively recently, i.e., 40-10 million years ago (Kuraku & Kuratani, 2006) and possess very similar karyotypic features (reviewed in Caputo Barucchi et al., 2013). It is unlikely that the genomic contents and protein-coding landscape largely differs between the two genomes. Thus, the difference in the number of predicted genes is likely caused by the difference in the completeness of genome assemblies or the difference in gene prediction methods.

Table 1-2. Protein-coding landscape in the two lamprey genome assemblies.

Species	**# CEGs detected by CEGMA**			**# of predicted genes**
	Complete	**Partial**	**All**	
Petromyzon marinus	140	32	172	26.046
Lethenteron camtschaticum	141	58	199	17.829

See Parra et al. (2009) for details of the criteria for categorizing genes detected in genome assembly into ‘complete’ and ‘partial’.

GC-content

Peculiarity of lamprey genes in terms of GC-content was already reported before the whole genome sequence of lampreys became available (Kuraku & Kuratani, 2006; reviewed in Kuraku, 2008). In the comprehensive analysis of the *P. marinus* genome consortium, we performed an intensive investigation of its base composition (Smith et al., 2013). The *P. marinus* genome exhibited relatively high overall GC-content (45.9%), and protein-coding regions, particularly synonymous nucleotide sites, especially had high GC-content (Supplementary Figure 6 in Smith et al., 2013; also see Figure 1-2).

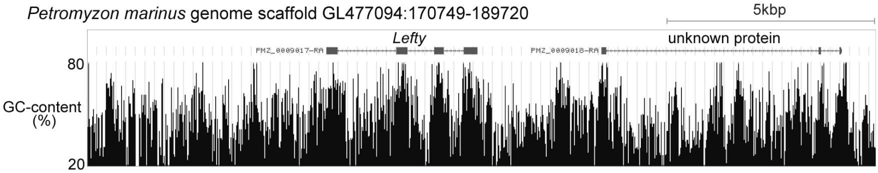

Figure 1-2. Browser view of GC-content in a selected region in the *Petromyzon marinus* genome. The graph of GC-content was obtained as GC-percent track at the UCSC Genome Browser, for the *P. marinus* genomic scaffold GL477094 (base position 170749-189720) containing a homolog of the *Lefty* gene (PMZ_0009017-RA). Note that the exons of this gene tend to have high GC-content (70-80%). The other 'unknown' gene in this view (PMZ_0009018-RA on the right) does not have any obvious homolog in other species and might be a lamprey-specific gene. In such a case, GC-content might serve as an indicator of protein-coding nature of genomic sequences.

Here we have analyzed the *Lethenteron camtschaticum* genome and compared some characteristics about GC-content with other vertebrates including *P. marinus* (Table 1-3). The *L. camtschaticum* genome exhibited markedly higher overall GC-content (48.0%) than the *P. marinus* genome (45.9%) (Figure 1-3, centerfold page i). Similarly, overall GC-content of protein-coding regions showed a comparable difference between the two species (Table 1-3).

The difference of global GC-content in the whole genome sequences of the two lampreys might be caused by either the respective choices of DNA source tissue (liver versus testis, in light of programmed genomic rearrangement), sequencing platform (Sanger versus Roche 454) or assembly methods (Arachne versus Newbler), rather than reflecting the genuine genome compositions. This might also hold for the difference in GC-content of protein-coding regions described above. The lamprey genomes would provide an interesting system to study how epigenetic information is organized in the genome with exceptional GC compartmentalization between coding (GC-rich) and non-coding (GC-poor) regions, as little is known about epigenetic regulation of this group of animals (see Tweedie et al., 1997; Covelo-Soto et al., 2014).

Table 1-3. Global and protein-coding GC-content in the two lamprey genomes.

Species	**Genome**		**Overall GC % of coding regions**
	Overall GC %	**GC % of 10Kbp non-overlapping windows**	
Petromyzon marinus	45.9	45 ± 3	56.3
Lethenteron camtschaticum	48.0	47 ± 4	59.6

Gene model

As the lamprey genomes have peculiar features in their protein-coding sequences (see below), gene prediction based on training with those features is expected to enhance its sensitivity and precision. The *Petromyzon marinus* genome consortium employed the program package MAKER (Cantarel et al., 2008) for genome-wide gene prediction, and it produced a gene typical of vertebrate genomes (Table 1-2) (Smith et al., 2013). In order to predict lamprey genes more precisely, we independently sought to implement lamprey-specific features in gene prediction. First, we built transcriptome assembly using all Sanger sequence reads of *P. marinus* available in NCBI dbEST (as of March 2008). In the assembled transcript contigs, we inferred open reading frames (ORFs) with identical lengths and high sequence conservation (≥70% positive match at the amino acid level, with a methionine corresponding to the putative start codon) in comparison with their jawed vertebrate homologs. Among 828 putative ORFs selected as above, we identified 132 ORF sequences that were contained in the *P. marinus* genome assembly petMar1 (version 3) with presumably full intronic and 2Kbp-long flanking sequences. Using them, we executed the training module of AUGUSTUS version 2.0.3 (Stanke & Waack, 2003) as instructed in its manual. The resulting parameter files for *P. marinus* gene model were passed to the developer of AUGUSTUS and are now available in the default species list (with the species identifier 'lamprey') of the installable program package (http://bioinf.uni-greifswald.de/augustus/ binaries/) and web server (http://bioinf.uni-greifswald.de/webaugustus/ prediction/create). This alternative gene prediction platform provides a complementary approach to exploit genomic resources of lampreys, although it remains to be carefully assessed whether the species parameters for *P. marinus* performs well for other lamprey species.

Codon usage bias and amino acid composition

Before whole genome sequences of lampreys became available, we performed analyses on codon usage bias and amino acid composition with 173 protein-coding genes of *Petromyzon marinus* that were available in GenBank (Qiu et al., 2011). In this study based on the relatively small data set, we suggested that lampreys have peculiar patterns of codon usage bias and amino acid composition. More recently, with the whole genome sequences of *P. marinus*, we performed more comprehensive analyses on those characteristics and confirmed that the peculiarity in the sequences of lamprey genes and peptides is genome-wide (Smith et al., 2013; Figure 1-4a and 1-4b, centerfold page ii). In addition, we revealed that GC-content in protein-coding regions is the major factor contributing to the peculiarity of codon usage bias and amino acid composition (Figure 1-4c and 1-4d, centerfold page ii). Our analyses did not support the relevance of codon usage bias to gene expression levels (Supplementary Figure 10 of Smith et al., 2013). It is of particular interest whether this lamprey-specific pattern is shared with other jawless fish genomes.

Homopolymeric amino acid (HPAA) tracts

More recently, we focused on homopolymeric amino acid (HPAA) tracts in peptide sequences (or single amino acid repeats, such as ‘QQQQQQQQ’; see Mularoni et al., 2010) and carried out a cross-species comparison of their frequencies (Noro et al., 2015). Our interest originated from a particular case of lamprey *Emx* genes (reviewed in Kuraku, 2010). Lampreys possess at least two *Emx* genes (*EmxA* and *EmxB*; Tank et al., 2009), and their gene products have a Q-tract and an A-tract at equivalent locations in the sequences of the two *Emx* gene products (Figure 1-5a, centrefold page iii-iv; Noro et al., 2015). A comparison with their hagfish orthologs without conspicuous HPAA tracts indicates that the insertions of the HPAA tracts occurred in the lamprey lineage after the split of the hagfish lineage (Figure 1-5a, centerfold page iii-iv). Our reanalysis confirmed the result by Tank et al. (2009) supporting lamprey lineage-specific *Emx* gene duplication, whereas the support was significantly weakened when the HPAA tracts were deleted from the multiple sequence alignment (Noro et al., 2015).

Inspired by the example of *Emx* genes, we performed a genome-wide survey of HPAA tract insertion. Our survey revealed a significant abundance of HPAA tracts in the overall protein-coding landscape of the sea lamprey genome, compared to that in the human and zebrafish (Noro et al., 2015). It also detected significant enrichment of G-tracts and Q-tracts unique to the sea lamprey (Noro et al., 2015; Figure 1-5b, centerfold page iii-iv). It is unknown what biochemical reasons underlie this species difference in HPAA tract insertion. If the trend of HPAA tract insertion reflects similarly on multiple sequences with similar property, namely paralogs, phylogenetic signals in those sequences might be weakened or erased by the secondary effects. This can result in erroneous alignments and molecular phylogeny inferences.

Perspectives

Several studies have reported gene duplications in the cyclostome lineage (Fried et al., 2003; Stadler et al., 2004; Tank et al., 2009). More recently, a genome-wide analysis suggested that a duplication event at the genome scale introduced lineage-specific duplicates (Mehta et al., 2013). Our analyses have highlighted several unique aspects of molecular evolution that seem to be characteristic of the cyclostome lineage. Above all, the unique sequence property of lamprey protein-coding genes is remarkable. It possibly drove convergent sequence evolution among ancient paralogs towards unexpected similarity, resulting in erroneous proximity between the paralogs in inferred molecular phylogeny. Molecular phylogeny inference based on amino acid sequences, as often practiced, is apparently supposed to circumvent the effect of GC-content and codon usage bias but is still prone to unfavorable effect of amino acid composition. Some of the gene duplications attributed to the cyclostome lineage so far could be explained by this possible artefact, and it can mislead our interpretation of the timing of whole genome duplications. Thus, extra cautions should be exercised in analyzing gene family trees involving lamprey genes (Figure 1-6).

Our knowledge of cyclostome genome compositions is still limited to Northern Hemisphere lampreys. Evolving DNA sequencing technology has enabled economical genome sequencing, and whole genome sequencing of hagfish and Southern Hemisphere lampreys are anticipated. With those

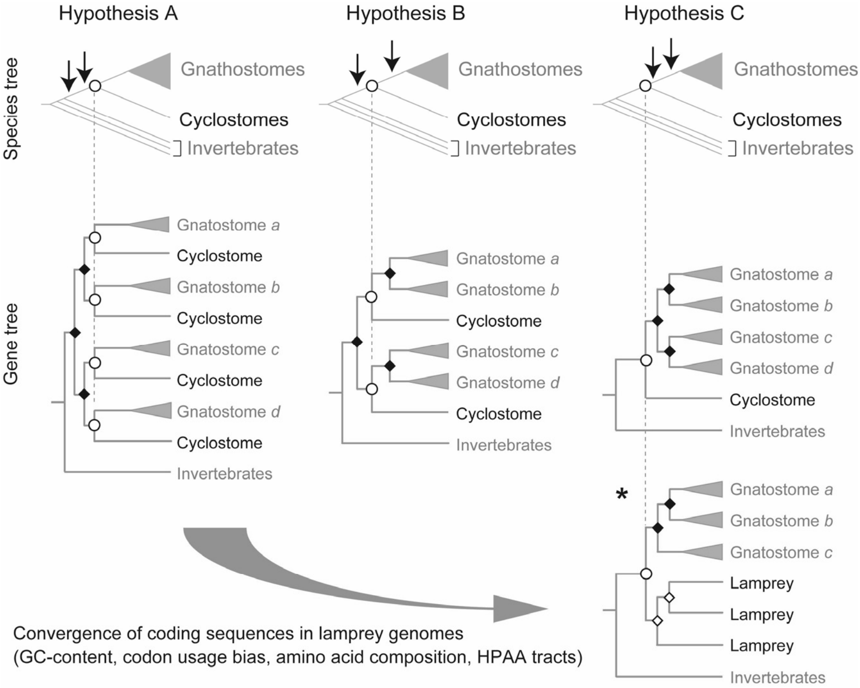

Figure 1-6. Possible causes of misinterpretation on molecular phylogeny involving lamprey genes. Alternative scenarios regarding the timing of two whole genome duplications (Hypothesis A-C) are shown with species trees and hypothetical gene trees. In the species trees, open circles indicate gnathostome-cyclostome split, and black arrows indicate the timing of whole genome duplication. In the hypothetical gene trees, a black diamond indicates gene duplication giving rise to multiple gnathostome paralogs, while a white diamond represents gene duplication giving rise to multiple cyclostome paralogs. Hypothesis A, with both rounds of whole genome duplications before the split between cyclostomes and gnathostomes, has been supported by a series of recent studies (Hoffmann et al., 2010; Kuraku et al., 2009a; Smith et al., 2013). In reality, some gene families exhibit molecular phylogeny depicted in the right bottom corner (*), with multiple cyclostome genes exclusively clustering with each other. This phylogenetic pattern, with gene duplications after the split between cyclostomes and gnathostomes, is incongruent with Hypothesis A and is rather compatible with Hypothesis C. We propose that the incongruence is, at least partly, caused by convergence of lamprey sequences discussed in this chapter.

resources, more comprehensive comparison of genomic features is expected to provide an increased understanding of what in the genome makes the phenotypic differences between jawless fishes and other chordates.

Acknowledgements

The authors acknowledge Maria Anisimova and Christophe Dessimoz for insightful discussion and Yuichiro Hara and Mitsutaka Kadota for critical reading of the manuscript.

References

Bradnam K.R., Fass J.N., Alexandrov A., Baranay P., Bechner M., Birol I., Boisvert S., Chapman J.A., Chapuis G., Chikhi R., Chitsaz H., Chou W.C., Corbeil J., Del Fabbro C., Docking T.R., Durbin R., Earl D., Emrich S., Fedotov P., Fonseca N.A., Ganapathy G., Gibbs R.A., Gnerre S., Godzaridis E., Goldstein S., Haimel M., Hall G., Haussler D., Hiatt J.B., Ho I.Y., Howard J., Hunt M., Jackman S.D., Jaffe D.B., Jarvis E.D., Jiang H., Kazakov S., Kersey P.J., Kitzman J.O., Knight J.R., Koren S., Lam T.W., Lavenier D., Laviolette F., Li Y., Li Z., Liu B., Liu Y., Luo R., Maccallum I., Macmanes M.D., Maillet N., Melnikov S., Naquin D., Ning Z., Otto T.D., Paten B., Paulo O.S., Phillippy A.M., Pina-Martins F., Place M., Przybylski D., Qin X., Qu C., Ribeiro F.J., Richards S., Rokhsar D.S., Ruby J.G., Scalabrin S., Schatz M.C., Schwartz D.C., Sergushichev A., Sharpe T., Shaw T.I., Shendure J., Shi Y., Simpson J.T., Song H., Tsarev F., Vezzi F., Vicedomini R., Vieira B.M., Wang J., Worley K.C., Yin S., Yiu S.M., Yuan J., Zhang G., Zhang H., Zhou S. & Korf I.F. 2013. Assemblathon 2: evaluating *de novo* methods of genome assembly in three vertebrate species. *GigaScience* 2, 10. doi 10.1186/2047-217X-2-10.

Cantarel B.L., Korf I., Robb S.M.C., Parra G., Ross E., Moore B., Holt C., Sánchez Alvarado A. & Yandell M. 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* 18, 188-196. doi 10.1101/gr.6743907.

Caputo Barucchi V., Giovannotti M., Nisi Cerioni P. & Splendiani A. 2013. Genome duplication in early vertebrates: insights from agnathan cytogenetics. *Cytogenetetics & Genome Research* 141, 80-89. doi 10.1159/000354098.

Covelo-Soto L., Morán P., Pasantes J.J. & Pérez-García C. 2014. Cytogenetic evidences of genome rearrangement and differential epigenetic chromatin modification in the sea lamprey (*Petromyzon marinus*). *Genetica* 142, 545-554. doi 10.1007/s10709-014-9802-5.

Dehal P. & Boore J.L. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biology* 3, e314. doi 10.1371/journal.pbio.0030314.

Delsuc F., Brinkmann H., Chourrout D. & Philippe H. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 439, 965-968. doi 10.1038/nature04336.

Escriva H., Manzon L., Youson J. & Laudet V. 2002. Analysis of lamprey and hagfish genes reveals a complex history of gene duplications during early vertebrate evolution. *Molecular Biology and Evolution* 19, 1440-1450.

Fried C., Prohaska S.J. & Stadler P.F. 2003. Independent Hox-cluster duplications in lampreys. *Journal of Experimental Zoology Part B: Molecular Development and Evolution* 299, 18-25. doi 10.1002/jez.b.37.

Hedges S.B., Dudley J. & Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22, 2971-2972. doi 10.1093/bioinformatics/btl505.

Hoffmann F.G., Opazo J.C. & Storz J.F. 2010. Gene cooption and convergent evolution of oxygen transport hemoglobins in jawed and jawless vertebrates. *Proceedings of the National Academy of Sciences of the United States of America* 107, 14274-14279. doi 10.1073/pnas.1006756107.

Kumar S., Filipski A.J., Battistuzzi F.U., Kosakovsky Pond S.L. & Tamura K. 2012. Statistics and truth in phylogenomics. *Molecular Biology and Evolution* 29, 457-472. doi 10.1093/molbev/msr202.

Kuraku S. 2008. Insights into cyclostome phylogenomics: pre-2R or post-2R. *Zoological Science* 25, 960-968. doi: 10.2108/zsj.25.960.

Kuraku S. 2010. Palaeophylogenomics of the vertebrate ancestor—impact of hidden paralogy on hagfish and lamprey gene phylogeny. *Integrative and Comparative Biology* 50, 124-129. doi 10.1093/icb/icq044.

Kuraku S. 2011. *Hox* gene clusters of early vertebrates: do they serve as reliable markers for genome evolution? *Genomics, Proteomics & Bioinformatics* 9, 97-103. doi 10.1016/S1672-0229(11)60012-0.

Kuraku S. 2013. Impact of asymmetric gene repertoire between cyclostomes and gnathostomes. 2013. *Seminars in Cell and Developmental Biology* 24, 119-127. doi
10.1016/j.semcdb.2012.12.009.

Kuraku S. & Kuratani S. 2006. Time scale for cyclostome evolution inferred with a phylogenetic diagnosis of hagfish and lamprey cDNA sequences. *Zoological Science* 23, 1053-1064. doi
http://dx.doi.org/10.2108/zsj.23.1053.

Kuraku S. & Meyer A. 2009a. The evolution and maintenance of *Hox* gene clusters in vertebrates and the teleost-specific genome duplication. *The International Journal of Developmental Biology* 53, 765-773. doi 10.1387/ijdb.072533km.

Kuraku S., Meyer A. & Kuratani S. 2009b. Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? *Molecular Biology and Evolution* 26, 47-59. doi 10.1093/molbev/msn222.

Kuraku S., Ota K.G. & Kuratani S. 2009c. Jawless fishes (Cyclostomata). In S.B. Hedges & S. Kumar (eds.): *The Timetree of Life*. Pp. 317-319. New York: Oxford University Press.

Mehta T.K., Ravi V., Yamasaki S., Lee A.P., Lian M.M., Tay B.H., Tohari S., Yanai S., Tay A., Brenner S. & Venkatesh B. 2013. Evidence for at least six Hox clusters in the Japanese lamprey (*Lethenteron japonicum*). *Proceedings of the National Academy of Sciences of the United States of America* 110, 16044-16049. doi 10.1073/pnas.1315760110.

Misof B., Liu S., Meusemann K., Peters R.S., Donath A., Mayer C., Frandsen P.B., Ware J., Flouri T., Beutel R.G., Niehuis O., Petersen M., Izquierdo-Carrasco F., Wappler T., Rust J., Aberer A.J., Aspock U., Aspock H., Bartel D., Blanke A., Berger S., Bohm A., Buckley T.R., Calcott B., Chen J., Friedrich F., Fukui M., Fujita M., Greve C., Grobe P., Gu S., Huang Y., Jermiin L.S., Kawahara A.Y., Krogmann L., Kubiak M., Lanfear R., Letsch H., Li Y., Li Z., Li J., Lu H., Machida R., Mashimo Y., Kapli P., McKenna D.D., Meng G., Nakagaki Y., Navarrete-Heredia J.L., Ott M., Ou Y., Pass G., Podsiadlowski L., Pohl H., von Reumont B.M., Schutte K., Sekiya K., Shimizu S., Slipinski A., Stamatakis A., Song W., Su X., Szucsich N.U., Tan M., Tan X., Tang M., Tang J., Timelthaler G., Tomizuka S., Trautwein M., Tong X., Uchifune T., Walzl M.G., Wiegmann B.M., Wilbrandt J., Wipfler B., Wong T.K., Wu Q., Wu G., Xie Y., Yang S., Yang Q., Yeates D.K., Yoshizawa K., Zhang Q., Zhang R., Zhang W.,