TURUN YLIOPISTON JULKAISUJA
ANNALES UNIVERSITATIS TURKUENSIS

SARJA - SER. AII OSA - TOM. 253

BIOLOGICA - GEOGRAPHICA - GEOLOGICA

Molecular Evolution of Metazoan Hypoxia-Inducible Factors

Kalle T. Rytkönen

TURUN YLIOPISTO
UNIVERSITY OF TURKU
Turku 2010

From the Division of Genetics and Physiology, Department of Biology, University of Turku, FIN-20014, Finland

Supervised by:

Professor Mikko Nikinmaa
Laboratory of Animal Physiology
Department of Biology
University of Turku, Finland

Professor Craig R. Primmer
Laboratory of Genetics
Department of Biology
University of Turku, Finland

Reviewed by:

Professor Johanna Myllyharju
Oulu Center for Cell-Matrix Research,
Biocenter Oulu and Department of Medical Biochemistry and Molecular Biology,
University of Oulu, FIN-90014, Finland

Professor Jay F. Storz
School of Biological Sciences
University of Nebraska
Lincoln, NE 68588, USA

Examined by:

Associate Professor Jeffrey G. Richards
Department of Zoology,
University of British Columbia,
Vancouver, British Columbia V6T 1Z4, Canada

Cover image: an epaulette shark and coral reef, Heron Island, Australia, Kalle Rytkönen

ISBN 978-951-29-4340-1 (PRINT)
ISBN 978-951-29-4341-8 (PDF)
ISSN 0082-6979
Painosalama Oy – Turku, Finland 2010

To my mother and father

MOLECULAR EVOLUTION OF METAZOAN HYPOXIA-INDUCIBLE FACTORS

Kalle T. Rytkönen

This thesis is based on the following original research contributions, which are referred to in the text by their roman numerals:

I Rytkönen KT, Williams TA, Renshaw GMC, Primmer CR and Nikinmaa M. Molecular evolution of the metazoan oxygen-sensing system: insights from elasmobranchs. *Manuscript*.

II Rytkönen KT, Renshaw GMC, Ashton KJ, Williams-Pritchard G, Leder EH and Nikinmaa M (2010) Elasmobranch qPCR reference genes: a case study of hypoxia preconditioned epaulette sharks. *BMC Molecular Biology* 11:27

III Rytkönen KT, Vuori KAM, Primmer CR and Nikinmaa M (2007) Comparison of hypoxia-inducible factor-1 alpha in hypoxia-sensitive and hypoxia-tolerant fish species. *Comparative Biochemistry and Physiology D Genomics and Proteomics* 2:177-186

IV Rytkönen KT, Ryynänen HJ, Nikinmaa M and Primmer CR. (2008) Variable patterns in the molecular evolution of the hypoxia-inducible factor-1 alpha (HIF-1α) gene in teleost fishes and mammals. *Gene* 420:1-10

ABSTRACT

Most metazoans rely on aerobic energy production, which is dependent on adequate oxygen supply. In the case of reduced oxygen supply (hypoxia), the most profound changes in gene expression are mediated by transcription factors named hypoxia-inducible factors (HIF alpha). These proteins are post-translationally regulated by prolyl-4-hydroxylase (PHD) enzymes that are direct "sensors" of cellular oxygen levels. This thesis examines the molecular evolution of metazoan HIF systems. In early metazoans the HIF system emerged from pre-existing PHD oxygen sensors and early bHLH-PAS transcription factors. In invertebrates our analysis revealed an unexpected diversity of PHD genes and HIF alpha sequence characteristics.

An early branching vertebrate, the epaulette shark (*Hemiscyllium ocellatum*) was chosen for sequencing and hypoxia preconditioning studies of HIF alpha and PHD genes. As no quantitative PCR reference genes were available, this thesis includes the first study of reference genes in cartilaginous fish species. Applying multiple statistical analysis we also discovered that commonly used reference gene software may perform poorly with some data sets. Novel reference genes allowed accurate measurements of the mRNA levels of the studied target genes.

Cartilaginous fishes have three genomic duplicates of both HIF alpha and PHD genes like mammals and teleost fishes. Combining functional divergence and selection analyses it was possible to describe how sequence changes in both HIF alpha and PHD duplicates may have contributed to the differential oxygen sensitivity of HIF alphas. Additionally, novel teleost HIF-1 alpha sequences were produced and used to reveal the molecular evolution of HIF-1 alpha in this lineage rich with hypoxia tolerant species.

CONTENTS

1. INTRODUCTION

1.1. Oxygen and life

1.1.1. Oxygen and energy production in living organisms

This thesis studies the evolution of the oxygen sensing system that takes part in maintaining cellular oxygen homeostasis. For a deeper understanding of oxygen metabolism it is first useful to review the origins of eukaryotic cells. Life on earth originated under anaerobic conditions (Follmann & Brownson, 2009). Oxygen was introduced into our atmosphere as a biological end product from the metabolism of photosynthetic organisms that had evolved the ability to capture solar energy into the chemical energy of reduced carbon bonds. For early anaerobic life forms, and still in modern cells, the toxic byproducts of oxygen were extremely harmful. Many of the cellular systems responding to oxygen may have developed to sequester harmful reactive oxygen species. About 1.5 billion years ago a major evolutionary transition took place in the form of endosymbiosis (Sagan, 1967; Wallin, 1923): an anaerobic pro-eukaryote engulfed an aerobic bacterium to form a eukaryotic cell. The aerobic endosymbiont (former bacteria) that was capable of efficient energy production by oxidative phosphorylation (Mitchell, 1961) ultimately evolved into mitochondria, sub-cellular organelles found in all eukaryotes from yeast to man. The aerobic energy production in mitochondria produces 18 times more energy per mole of glucose compared to anaerobic glycolysis. Subsequently, the evolution and successful radiation of multi-cellular organisms has relied on effective use of oxygen as the final electron acceptor in energy production. Metabolic and gene regulatory pathways, including responses to hypoxia (reduced oxygen supply), have adapted in parallel to integrate and coordinate mitochondrial and glycolytic functions (Webster, 2003). This delicate balance of fine-tuning the mitochondrial energy production without excessive release of reactive oxygen species is still a matter of life and death in all eukaryotes. In a wider sense, this balance can be reflected back to the symbiotic origins of eukaryotic cells in which oxygen byproducts have become part of normal cellular chemistry.

Together with efficient cellular oxygen metabolism metazoans (animals) have evolved a wide array of different physiological and molecular adaptations for oxygen delivery. In invertebrates all the different oxygen transfer possibilities exist from pure diffusion to tracheal or closed circulation. For example, the roundworm *Caenorhabditis elegans* has less than a 1000 cells and is able to meet its aerobic oxygen demand solely by passive diffusion from the atmosphere. Another model species, the fruit fly *Drosophila melanogaster*, has a more complex anatomy with tracheal tubules that transports oxygen to internal tissues and cells. Finally, in vertebrates the heart pumps oxygen to tissues through a circulatory system equipped with red blood cells containing oxygen-binding hemoglobin.

1.1.2. Hypoxia responses

The vast majority of metazoan life forms need continuous and sufficient oxygen supply to support metabolism, but various species have also evolved to cope with hypoxia or anoxia (total lack of oxygen) for limited time periods (Hochachka & Lutz, 2001; Nikinmaa & Rees, 2005). Here hypoxia responses will be considered generally in the context of organisms being exposed to suboptimal oxygen supply. Oxygen availability is more critical in aquatic than terrestrial habitats, since water contains only 1/30th of the

oxygen compared with the same volume of air at the same partial pressure and the rate of diffusion of oxygen in water is only 1/10 000th of that in air (Dejours, 1975). Consequently, the oxygen level in water can change drastically as a result of oxygen consumption by organisms. Due to these environmental pressures, oxygen has been a major force in the evolution of aquatic organisms and various adaptations have evolved in response to reduced oxygen availability, also in vertebrates (Nikinmaa, 2002; Val, 1995). For these reasons particularly water-breathing teleost and cartilaginous fish will be the focus of this thesis.

Next I will briefly describe some of the general characteristics of hypoxia responses, for more details see reviews (Bickler & Buck, 2007; Hochachka & Lutz, 2001; Hoogewijs et al, 2007; Nikinmaa & Rees, 2005; Richards, 2009). When exposed to hypoxia the animal tries to maintain tissue oxygenation. This includes rapid responses increasing the hemoglobin concentration of blood via changes in plasma volume, liberation of red blood cells from storage organs and production of red blood cells (erythropoiesis). Many species also have adaptations in the structure of hemoglobin for better oxygen affinity and enhanced oxygen delivery. Another solution is to produce energy anaerobically by metabolising all sugars in glycolysis. However, glycolysis is less efficient for ATP production and high glucose reserves are necessary for longer survival. Additionally, the accumulation of the end product lactate leads to acidosis.

Animals that are able to tolerate hypoxia (many fish species and some non-mammalian tetrapods) can fast and reversibly reduce energy consumption (metabolic depression or hypometabolism). Efficient energy conservation includes saving ATP in all major cellular processes, for example protein synthesis and degradation, ion pumping (especially Na+/K+ ATPase) and gluconeogenesis. Hypoxia-tolerant species can also sustain aerobic energy production in lower oxygen concentrations than hypoxia-sensitive species. Mitochondrial oxygen consumption is reduced to meet the reduced oxygen supply from the environment (oxyconformance) and aerobic energy production can be sustained for longer time periods. Hypoxia-tolerant animals have also evolved protective mechanisms to avoid oxidative damage in re-oxygenation. The most profound protective responses to hypoxia include drastic changes in gene expression, and many of these are controlled by a transcription factor termed hypoxia inducible factor (HIF), the major regulator of oxygen-dependent gene expression and the subject of this thesis.

1.1.3. Hypoxia-inducible factors

I briefly introduce HIF regulatory systems – more details are available in recent reviews on mammals (Kaelin & Ratcliffe, 2008; Lendahl et al, 2009; Webb et al, 2009), fishes (Nikinmaa & Rees, 2005; Richards, 2009) and invertebrates (Gorr et al, 2006; Hampton-Smith & Peet, 2009). Hypoxia-inducible factors belong to the bHLH-PAS (basic Helix-Loop-Helix – Per-ARNT-Sim) family of transcription factors, which are involved in the regulation of environmentally induced and developmental gene expression (Kewley et al, 2004). In mammals, HIFs exert transcriptional control on gene expression involved in a range of processes including glycolysis, glucose and iron transport, angiogenesis, erythropoiesis, cell-cycle control (Lendahl et al, 2009; Wenger et al, 2005) and oxyconformance (Fukuda et al, 2007); all of these are processes that are crucial to hypoxia responses discussed in the previous section. HIF-1 consists of two subunits (Wang et al, 1995), ARNT

(aryl hydrocarbon nuclear translocator or HIF-beta) and HIF-1alpha, which confers hypoxia sensitivity to HIF. HIF-1 alpha is post-translationally regulated by prolyl-4-hydroxylase (PHD) enzymes that directly use oxygen as a cosubstrate in the hydroxylation reaction and thus are direct "sensors" of cellular oxygen partial pressure (Jaakkola et al, 2001; Myllyharju, 2009). In normoxia PHD enzymes covalently modify two proline residues in the oxygen-dependent degradation (ODD) domain of HIF-1 alpha, and HIF-1 alpha is rapidly broken down. Hydroxylation of the prolines allows binding of von-Hippel-Lindau protein (VHL) and recruits ubiquitin ligase complex that targets of HIF-1 alpha for proteosomal degradation (Jaakkola et al, 2001). In hypoxia hydroxylation does not take place, and HIF is stabilized. Another level of oxygen-dependent regulation by HIF is caused by Factor Inhibiting HIF-1 (FIH), which negatively regulates HIF at the level of transcriptional complexes (Koivunen et al, 2004; Lando et al, 2002).

A similar regulatory system, having HIF as a central regulator of the hypoxia response, is found in fishes (Nikinmaa & Rees, 2005), the fruit fly (*D. melanogaster*) (Nambu et al, 1996) and the roundworm (*C. elegans*) (Jiang et al, 2001). Table 1 shows some examples of HIF targets that are conserved across metazoans. Studies in *C. elegans* have revealed additional possible widespread roles including heat acclimation, behavioral responses to oxygen and carbon dioxide, neural development and aging (Bretscher et al, 2008; Chang & Bargmann, 2008; Mehta et al, 2009; Pocock & Hobert, 2008; Treinin et al, 2003). Involvement of HIF in temperature acclimation has also been demonstrated for a poikilothermic teleost fish (*Carassius carassius*) (Rissanen et al, 2006). However, most of our knowledge on the HIF system originates from mammalian tissues, tumors and cell lines (Lendahl et al, 2009), which may bias our understanding when metazoans are considered as a whole. For example, in mammals HIF targets more than 100 hypoxia responsive genes at hypoxia response elements (HRE) that have a consensus HRE motif (NRCGTG) (Lendahl et al, 2009; Wenger et al, 2005). In teleost fishes, only three HIF target genes have been analyzed in detail: Insulin-like growth factor binding protein 1 (IGFBP-1) (Kajimura et al, 2006), lactate dehydrogenase-B gene (*Ldh-B*) (Rees et al, 2009), and Cbp/p300-interacting transactivator CITED3 (Ng et al, 2009). Strikingly, only IGFBP-1 was induced via the canonical HRE, whereas *Ldh-B* had non-canonical HRE (GATGTG)

Table 1. Examples of HIF target genes conserved across metazoans (modified from Hampton-Smith and Peet 2009).

Function	Gene Product	C. elegans	D. melanogaster	Mammals
Negative feedback regulation of HIF	PHD3, PHD2, Fga, Egl-9 [a]	(Bishop et al, 2004)	(Lavista-Llanos et al, 2002)	(Metzen et al, 2005)
Lactate metabolism	LDH, dmLDH [b]		(Gorr et al, 2004)	(Firth et al, 1995)
pH regulation	CA, CAH-4 [c]	(Bishop et al, 2004)		(Ivanov et al, 1998)
Collagen synthesis	P4H, P4Hα2 [d]	(Shen et al, 2005)		(Takahashi et al, 2000)
TOR signaling, Inhibition of growth	REDD1, Scylla [e]		(Reiling & Hafen, 2004)	(Brugarolas et al, 2004)

[a] HIF prolyl-4-hydroxylases
[b] lactate dehydrogenases
[c] carbonic anhydrases
[d] collagen prolyl-4-hydroxylases
[e] TOR = target of rapamycin

and CITED3 activation may involve yet another core sequence.

In invertebrates only one HIF alpha transcription factor can be detected, whereas in vertebrates three functional duplicates of HIF (HIF-1 - HIF-3 alpha) are presently known (Heidbreder et al, 2003; Law et al, 2006; Ratcliffe, 2007). In mammals HIF-1 alpha and HIF-2 alpha, which is also called endothelial PAS domain protein 1 (EPAS1), both function as transcriptional activators. HIF-1 alpha is ubiquitously expressed, whereas the expression pattern of HIF-2 alpha is more restricted to certain cell types and conditions (Wiesener et al, 2002). In addition to their role in the hypoxic response, HIFs appear to be crucial for the basal transcription of their target genes (Mason et al, 2004; Stroka et al, 2001) and are important in developmental processes (Dunwoodie, 2009). Splice variant of HIF-3 alpha, the mouse IPAS negatively regulates HIF-1 alpha (Makino et al, 2002), and human HIF-3 alpha4 variant negatively regulates HIF-1 alpha (Jang et al, 2005; Makino et al, 2002) and HIF-2 alpha (Maynard et al, 2007).

Functional divergence of HIF-1 alpha and HIF-2 alpha is relevant both from the evolutionary and medical perspectives. Generally HIF-1 alpha is more involved in acute short-term hypoxia responses whereas HIF-2 alpha associates more with long term hypoxia responses (Holmquist-Mengelbier et al, 2006; Lendahl et al, 2009; Rahman & Thomas, 2007). HIF-2 alpha is less efficiently hydroxylated by PHDs (becomes stabilized in higher oxygen levels) (Appelhoff et al, 2004) and FIH (Bracken et al, 2005; Koivunen et al, 2004). The ability to activate different sets of target genes is not based on the unique DNA-binding to HREs of hypoxia responsive genes, but rather protein-protein interactions at C-terminal transactivation domains (Hu et al, 2007; Lau et al, 2007). It is possible that HIF-1 alpha and HIF-2 alpha may partly target the same genes but under different conditions or cell types (Holmquist-Mengelbier et al, 2006), but ortholog specific DNA binding has been described (Mole et al, 2009). HIF-1 alpha activates glycolytic enzymes and is more closely involved in the regulation of metabolism than HIF-2 alpha. HIF-2 alpha in turn is more involved in stem cell control and differentiation, which makes it particularly interesting in terms of cancer research (Lendahl et al, 2009).

This thesis also analyzes the evolution of PHDs that are direct oxygen sensing enzymes regulating HIF alphas. PHD enzymes belong to the family of iron (II) and 2-oxoglutarate -dependent dioxygenases that covalently modify two proline residues in the ODD domain of HIF alpha (Myllyharju, 2009; Wenger et al, 2009). The two target prolines in HIF-1 alpha are Pro-402 (N-terminal ODD domain, NODD domain) and Pro-564 (C-terminal ODD domain, CODD domain), and the core hydroxylation motifs (LXXLAP) are widely conserved in vertebrate HIFs. In vertebrates three PHD enzymes are present: PHD1 (*EGLN2*, HIF-P4H-1), PHD2 (*EGLN1*, HIF-P4H-2) and PHD3 (*EGLN3*, HIF-P4H-3) (Myllyharju, 2009). PHDs share very little sequence identity with the other types of vertebrate prolyl hydroxylases, collagen P4Hs and a trasmembrane P4H (P4H-TM, also known as PHD4) that is likely to be involved in the regulation of HIF (Koivunen et al, 2007). There is only one PHD homolog in *D. melanogaster* (*fatiga*) and *C. elegans* (*EGLN*).

Recent advances in comparative genomics and molecular evolution offer a possibility for comparative studies that can create novel insights into the diversity of HIF signaling in non-model organisms. As the conceptual and methodological frame of this thesis is from the field of molecular evolution, before moving into

the specific research contributions, next section introduces molecular evolution as a discipline and discusses the outcomes of gene duplications.

1.2. Molecular evolution

1.2.1. Principles of molecular evolution

Evolution is a process in which organisms adapt genetically to varying environments (Darwin, 1859). Changes in genotype, or DNA sequences, are the raw material for adaptations that are exposed to natural selection on the level of phenotypes of individuals. Molecular evolution is the process of evolution at the scale of nucleotide sequences (DNA or RNA) or protein sequences. In the 1940-1950s deoxyribonucleic acid (DNA) was discovered as the genetic material of living organisms (Avery et al, 1944; Hershey & Chase, 1952) and the structure of DNA was solved (Watson & Crick, 1953). Before these findings, mechanisms of evolution were studied in the fashion of Mendelian crossing experiments (Mendel, 1866) limited to within-species studies. The knowledge of the molecular genetic entity revolutionized the study of evolution. By extracting the actual genetic entities, genes (DNA) or their protein products, the evolutionary change in genes could be studied between any pair of species as long as homology could be identified between the gene pair. The early studies of molecular evolution were conducted on protein sequences with some fundamental outcomes. First, it was observed that amino acid changes take place very rarely in functionally important proteins or protein regions. Second, the number of amino acid substitutions between two species is often nearly proportional to the time since divergence of the species (Margoliash & Smith, 1965; Zuckerland & Pauling, 1965). Further, substitution rates of important genes are low, but the number of substitutions that have beneficial effects on the fitness of individual organisms is much lower still. This was formalized as the neutral theory of molecular evolution (Kimura, 1968; King & Jukes, 1969), which states that by far most nucleotide and amino acid mutations are selectively neutral. The neutral theory does not exclude the possibility of beneficial or adaptive mutations, but lays emphasis on random genetic drift driving the mutations to extinction or fixation. Ohta (Ohta, 1973) extended this framework in "the nearly neutral theory", and increasing amounts of data have suggested that adaptive substitutions may not be as rare as originally proposed (the selectionist view) (Nei, 2005), up to a recent report claiming adaptive substitutions as "pervasive" in vertebrate evolution (Studer et al, 2008). Together, both "neutralist" and "selectionist" views have been useful in the process of understanding the molecular mechanisms of evolution (Nei et al, 2008; Wagner, 2008). Additionally, evolutionary importance of non-protein coding DNA that has regulatory roles in gene expression has become evident (Chan et al, 2010; Visel et al, 2009).

1.2.2. Evolution by gene duplication

Gene duplication is the process by which a chromosome or a portion of DNA is duplicated, resulting in an additional copy of a gene or genes. Whole genome duplication is a duplication event where all the genetic material of an organism is doubled. Prior to the discovery of DNA as the genetic material, cytologists studying *Drosophila* produced substantial evidence for the evolutionary importance of duplications by chromosomal studies (reviewed by Taylor & Raes, 2004). In 1970 Susumo Ohno published *Evolution by Gene Duplication,* where he

conceptualized the possible evolutionary outcomes of gene duplications that have later been only slightly refined (Force et al, 1999; Hahn, 2009; Zhang, 2003). Briefly, the 4 major outcomes are:

1) Pseudogenization, or loss of the other duplicate, for example, the loss of chemosensory receptor genes (Nei et al, 2008).

2) Conservation of both copies in rare cases of high demand for the gene product (dosage), for example, multidrug resistant gene 1 in malaria parasite (Price et al, 2004).

3) Subfunctionalization, which is the division of ancestral functions among duplicates, for example, specialization of digestive enzyme functions in leaf-eating monkey (Zhang et al, 2002).

4) Neofunctionalization, which is the evolution of a new function in one of the duplicates, for example, glutamate dehydrogenase 2 in neurons (Burki & Kaessmann, 2004).

In addition to changes in gene coding sequences the changes in regulatory sequences of the genes can be considered in the above frameworks. That is, even without purely novel function (enzymatic etc.) subtle changes of the ancestral gene expression patterns spatially and temporally may contribute to novel phenotypic adaptations. Subfunctionalization has also been described by the duplication, degeneration, complementation (DDC) model (Force et al, 1999). In the DDC model the mutations causing subfunctionalization are neutral and both gene copies are preserved as a result of mutations that have removed different subsets of the original functions from each gene copy. The mutations are not deleterious as the function lost is still performed by the other copy of the gene. In many cases the "new" function of one copy may be an ancient secondary property that became its main function after the duplication, that is "turning hobby into a job" (Conant & Wolfe, 2008).

Whole genome duplications likely contributed significantly to the major evolutionary transitions of life. It is now well-established that two rounds of genome duplication took place during the early evolution of vertebrates (Kuraku et al, 2009; Ravi et al, 2009). Also in the early metazoan evolution gene or genome duplications were important (Lundin, 1999), and duplication events leading to the expansion of transcription factor families may have contributed to the evolution of multicellularity (Degnan et al, 2009). Generally, changes in transcription factor proteins have played central roles in evolution (Lynch & Wagner, 2008) and the most widely studied ones are developmental transcription factors including hox genes (Gehring et al, 2009; Ravi et al, 2009). Changes in environmentally regulated transcription factors, like HIF, have been significant in evolution (Gu et al, 2000), even though their phenotypic effects are not as readily quantifiable compared to the morphological effects of purely developmental transcription factors. Of interest is also the recent debate on the relative contribution of changes in regulatory DNA versus protein coding DNA to phenotypic evolution. Even though much evidence has emerged for the importance of regulatory changes in non-coding DNA, substitutions in protein coding DNA, especially in transcription factors, still remain of crucial research interest (Lynch & Wagner, 2008). In this thesis I concentrate on the changes in protein coding sequences and study the molecular evolution of a physiologically relevant

transcription factor responsible for oxygen regulation. The divergence following the early vertebrate genome duplications of HIF alpha and its regulatory enzymes may serve as viable examples of physiologically important sub- and neofunctionalizations.

1.3. Outline of the thesis

In this thesis the methods of molecular evolution and genomics were employed to study the evolution of the HIF transcription factor system at various taxonomic levels. The first chapter gives a comprehensive overview of the molecular evolution of the metazoan oxygen sensing system. After utilizing the newly available genome data to evaluate the presence of HIF alpha and PHD genes in invertebrate metazoans (I), an early diverging cartilaginous fish, the epaulette shark (*Hemiscyllium ocellatum*), was chosen for studies of HIF alpha and PHD genes (I, II). Novel epaulette shark sequences served as the out-group for detailed protein level functional divergence analysis of vertebrate gene duplicates (I). The responses of hypoxia-tolerant epaulette sharks were functionally studied in hypoxia and after hypoxia preconditioning. As to date no comparative studies have established reference genes for quantitative PCR (qPCR) in cartilaginous fishes, the second chapter examines the suitability of 9 reference candidates for hypoxia studies in the epaulette shark (II). The remainder of the thesis concentrates on HIF-1 alpha, which is particularly central in the acute hypoxia responses. The HIF-1 alpha gene was sequenced from an extensive collection of hypoxia-sensitive and hypoxia-tolerant fish species and the presence of any protein signatures associated with oxygen dependency was investigated (III). The novel teleost sequences enabled a more detailed comparison of molecular evolution of HIF-1 alpha in the water-breathing vertebrates vs. air-breathing vertebrates and particularly the evidence for positive selection was evaluated with maximum likelihood models of codon substitutions (IV).

2. MATERIAL AND METHODS

2.1. Animal experiments, sampling, novel sequences and quantitative PCR

For papers I and II epaulette sharks were collected from the wild and a hypoxia preconditioning protocol was used to elicit the shark's responses to hypoxia in the laboratory. Hypoxia insult was two hours of hypoxia at 0.34 mg O_2/l (5% of air saturation). Cerebellum, heart, gill and eye tissue samples were collected from 10 controls, 10 individuals with a single hypoxic insult and 10 individuals with hypoxia preconditioning (8 hypoxic insults, 12 hours apart). RNA was extracted, cDNA produced, and a collection of degenerate (universal) primers was used to obtain primary sequence fragments. Degenerate primers based on alignments of tetrapod (human, mouse, chicken, frog) and teleost (zebrafish, medaka, stickleback, fugu) sequences were designed using netprimer (http://www.premierbiosoft.com/netprimer). "Touch-down" PCRs (see II) were conducted, the identities of the sequences were verified using BLAST searches and with alignments by hand. To obtain the complete coding sequences primers for rapid amplification of cDNA ends (RACE) were designed and PCRs were conducted according to manufacturer's instruction with the SMART kit (BD Biosciences, San Jose, CA). For chapter III, the teleost fish samples were collected from the wild and processed as above. Here the degenerate primers were designed based on the sequences of the teleost species available at the time.

Based on the novel sequences obtained, qPCR primers were designed for each gene and tested (I and II). For each gene a specific and efficient pair was chosen and qPCRs were run on a 7900HT Fast Real-Time PCR System (Applied Biosystems) with Maxima SYBR Green qPCR Master Mix (2x) (Fermentas, St.Leon-Rot, Germany) using a 2-step protocol [initial denaturation at 95 °C for 15 min, 40 cycles of denaturation at 95 °C for 15 s and annealing/extension at 60 °C for 1 min] with dissociation curve analysis and all primers at 75 nM. Samples were run in triplicates, with no-template controls, and standard curves were embedded in the experimental block design. Standard curves were constructed from a five point dilution series (1/10 to 1/910) from pooled cDNA. In the reference gene study (II) relative quantity values were examined by analysis of variance (ANOVA) among the control and treatment groups, and then with pairwise testing. This was followed with NormFinder (Andersen et al, 2004) and geNorm (Vandesompele J et al, 2002) analysis. The ANOVA analysis was supplemented with graphical analysis of the residuals, normality tests and homogeneity tests implemented in SPSS 11 (SPSS Inc.). For the target genes (I) we normalized the mRNA expression levels to the geometric mean of EF and UBQ and normalized relative quantity values were examined by analysis of variance (ANOVA) (I).

2.2. Database searches, sequence alignments, phylogenies and shared synteny

In order to investigate the presence of HIF and PHD genes we conducted comprehensive TBLASTN searches in the available invertebrate and early vertebrate genome sequences (see I) and NCBI non-redundant protein database. If expected/hypothesized hits were not found we collected the genomic sequences surrounding the primary BLAST hits in the genome contigs and re-predicted the gene structure using three independent pieces of software: GENESCAN (http://genes.mit.edu/GENSCAN.html), Augustus (http://augustus.gobics.de/) and FGENESH (http://linux1.softberry.com).

For vertebrates, nucleotide sequences were collected from Ensembl selecting the longest transcripts. This Ensembl collection was supplemented with Genbank sequences using Blast searches (I). In chapters III and IV sequences were collected from Genbank and Ensembl.

Multiple sequence alignments were built with CLUSTAL W (Thompson et al. 1994) (III, IV) or MUSCLE (Edgar, 2004) using the default parameters (I). The alignments were manually curated to identify and remove poorly aligned regions. ProtTest 1.4 (Abascal et al, 2005) was used to obtain a substitution model that best fit the data (JJT+G) and this was used in PhyML (Guindon & Gascuel, 2003) to obtain the maximum likelihood (ML) phylogeny with bootstrapping (100 replicates). The ML phylogenies were used in subsequent analyses (I). In chapter IV ML (DNAml) and parsimony (DNAPars) methods (PHYLIP) were used.

Shared synteny was evaluated in human, zebrafish and tetraodon by collecting the genes flanking HIF and PHD at their genomic loci in Ensembl (I). Protein hydrophilicity plots were made using the Hopp-Woods scale (http://arbl.cvmbs.colostate.edu/molkit/hydropathy/) (III).

2.3. Amino acid substitution rates

The average mammalian and teleost amino acid substitution rates were calculated using two different out-groups (I). Sequences from the invertebrate *Amphioxus* were used to calculate evolutionary rates between HIF and PHD paralogs; we then evaluated rate shifts between amniotes and teleosts within each vertebrate duplication using the epaulette shark sequences as out-groups. First, pairwise amino acid distances from the out-group were calculated in MEGA 3.1 (Kumar et al, 2004), and then the average of mammals and teleost was used for substitution rate calculations. The average pairwise amino acid sequence divergence estimates for mammals, teleost, cypriniformes, perciformes and rodentia were calculated in MEGA 3.1, using the pairwise deletion option for gaps (II). Evolutionary rates (substitutions/aa site/year $x10^9$) in these groups were estimated from the average of pairwise amino acid divergences within a given group.

2.4. Model testing of selection pressures

Comparison of relative fixation rates of nonsynonymous (dN, amino acid changing) and synonymous (dS, silent) substitutions offer an insight to the evolution that has been acting on protein coding genes. The dN/dS rate ratio (ω = dN/dS) can be used to measure the selective pressure at the protein level. In genes under no selective pressure an equal number of synonymous substitutions per synonymous site and nonsynonymous substitutions per nonsynonymous site are expected (neutral evolution, $\omega = 1$). In the case that there are significant constraints to retain amino acid identity relatively fewer nonsynonymous substitutions are expected (negative selection, $\omega < 1$). In contrast, a higher number of nonsynonymous than synonymous substitutions provides evidence for positive Darwinian selection (positive selection $\omega > 1$).

To avoid synonymous substitution saturation the duplicate-specific gene phylogeny was not used, but a species phylogeny and concatenated sequence data from each set of paralogs (for example, HIF-1 alpha, HIF-2 alpha and HIF-3 alpha concatenated) (I). The partition dataset option (G) (Yang & Swanson, 2002) was used in codeml, which is part of the PAML package (Yang, 2007), to test if the gene paralogs have experienced statistically significant differences in selection pressures. Nested

likelihood ratio tests were performed for the following series of model comparisons: First, mode of selection was equal (w1 = w2 = w3); Second, selection pressures were significantly different in one of the three paralogs but equal in two others (w1 = w2 ≠ w3 or w1 ≠ w2 = w3); Third, selective pressures were different for each (w1 ≠ w2 ≠ w3). Inside a model, different combinations of free and fixed parameters were tested (Mgene=0,1,3,4 and Mgene=3 with fixed κ), including substitution rate (s), transition/transversion ratio (κ), codon frequencies (πs) and dN/dS ratio (ω). HIF-1 alpha and HIF-2 alpha NODD and CODD domains (see above) were analyzed together in a dataset of 4 partitions for both mammals and teleosts. Here, likelihood ratio tests (LRT) were done for each model with only substitution rate parameters, and for the best fitting model more parameters were added and tested.

Evidence of variation in selection pressures on branches of the HIF-1 alpha phylogeny and along HIF-1 alpha codons was tested using the maximum-likelihood methods implemented in PAML (Yang, 2007) (II). A collection of branch-specific, site-specific and branch-site specific models was employed to study the selective pressures and evidence for positive selection on protein coding sequences. Negative and positive selection was also tested using single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL) and random effects likelihood (REL) methods of the HypHy package (Pond & Frost, 2005) via the public web implementation www.datamonkey.org. Additionally, estimates of dN and dS substitution rate variation for individual codons as well as across the entire gene across each of the aligned datasets were made in SNAP (http://www.hiv.lanl.gov/content/sequence/SNAP/README.html).

2.5. Functional divergence analyses

Analyses of functional divergence (FD) were performed on two periods during the evolutionary history of the HIF and PHD genes: firstly, on the divergence of the vertebrate paralogs following their duplication early in vertebrate evolution (using Amphioxus sequences as out-groups); and secondly, on divergence within each paralogous group after the split between teleosts and amniotes. Two computational methods for identifying residues under functional divergence were used. The Type I method of Gu (Gu, 1999), implemented in DIVERGE (Gu & Vander Velden, 2002), uses a maximum likelihood procedure to compare evolutionary rates in two pre-defined clades of sequences. DIVERGE also implements a second (Type II) method (Gu, 2006), which identifies "conserved-but-different" residues between two clades of sequences. The third method, which is conceptually similar to the Type II analysis, uses a simple distance-based approach in which BLOSUM substitution scores are used to quantify the radical or conservative nature of substitutions between two clades of sequences, with the score corrected for alignment column conservation (Toft et al, 2009; Williams et al, 2010).

3. MAIN RESULTS AND DISCUSSION

The specific study questions and main results of the four chapters in this thesis are summarized in Table 2 and discussed below. In the following overview I will concentrate on the main findings – for specific details see original research contributions.

3.1. Emergence of the HIF oxygen-response system in invertebrates

The HIF-PHD oxygen-sensing system provides an excellent example of how pre-existing components can be recruited into a new system during evolution. The PAS domain responsible for HIF alpha dimerization has ancient origins, with prokaryote homologs involved in environmental sensing (Taylor & Zhulin, 1999). Oxygen-sensing homologs of PHDs were already present in non-metazoan eukaryotes (Hughes & Espenshade, 2008; West et al, 2007) and during metazoan evolution were recruited into a bHLH-PAS transcription factor circuitry. Early metazoan gene duplications of bHLH-PAS transcription factors resulted in the appearance of HIF alpha -like transcription factors before the divergence of Cnidarian and Bilaterian lineages. In bilaterians, the HIF-PHD interaction was further refined by the emergence of two separate proline motifs that permit more precise coordination of oxygen-dependent HIF regulation (Figure 1). The pre-Chordate origins of this innovation could not be deduced from studies on *Drosophila* or *Caenorhabditis*, as these model organisms have only one predicted proline in the HIF gene product. This result suggests that work on the HIF oxygen-sensing systems of these species should be only very carefully projected to other bilaterian species (I).

Table 2. The specific study questions and main results of the four chapters included in this thesis.

	Specific study questions	Main results
I	When did HIF and PHD mediated gene regulation arise and diversify?	The presence of PHDs in metazoan genomes predates the emergence of HIFs, which are confidently detectable in Cnidarians. Cartilaginous fish and other vertebrates have 3 duplicates of both HIF and PHD.
	Did functional divergence take place during HIF alpha evolution?	Functional divergence was detected in both HIF alpha paralogs and orthologs in both the PAS dimerization domains and oxygen dependent degradation (ODD) domains.
	How HIF-2 alpha evolved to respond to oxygen differently than HIF-1 alpha?	Elevated substitution rates in the HIF-2 alpha NODD domain compared to the other ODD domains, together with the known differences in PHD2 vs. PHD1/3 substrate specificity, may contribute to differential oxygen sensitivity between HIF-1 alpha and HIF-2 alpha paralogs.
II	What are the most suitable reference genes for acute hypoxia studies in cartilaginous fishes?	In the studied conditions eef1 and ubq. The sequences we provided are useful reference gene candidates for other studies in cartilaginous fishes.
	Do the results of common reference gene software agree with basic statistical analysis (ANOVA)?	Not necessarily, the genes selected for testing can bias the results of the software.
III	Can we observe protein signatures in HIF-1 alpha protein associated with hypoxia-sensitivity versus hypoxia-tolerance in teleost fish species?	No clear amino acid signatures which could be associated with the oxygen requirements of the species were found. HIF-1 alpha proteins are shorter than their mammalian counterparts due to deletions around CODD domain. However, in the absence of structural data the significance of the deletions remains to be elucidated.
IV	Have HIF-1 alpha proteins evolved differently in water-breathing vertebrates compared to air-breathing vertebrates?	The overall evolutionary rate in teleost HIF-1 alpha was twice as fast as in mammalian HIF-1 alpha, however, the crucial interaction domains were under stringent negative selection in all vertebrates.
	Is there evidence for positive selection in teleost HIF-1 alpha?	Some evidence was found in the HIF-1 alpha bHLH-PAS domain.

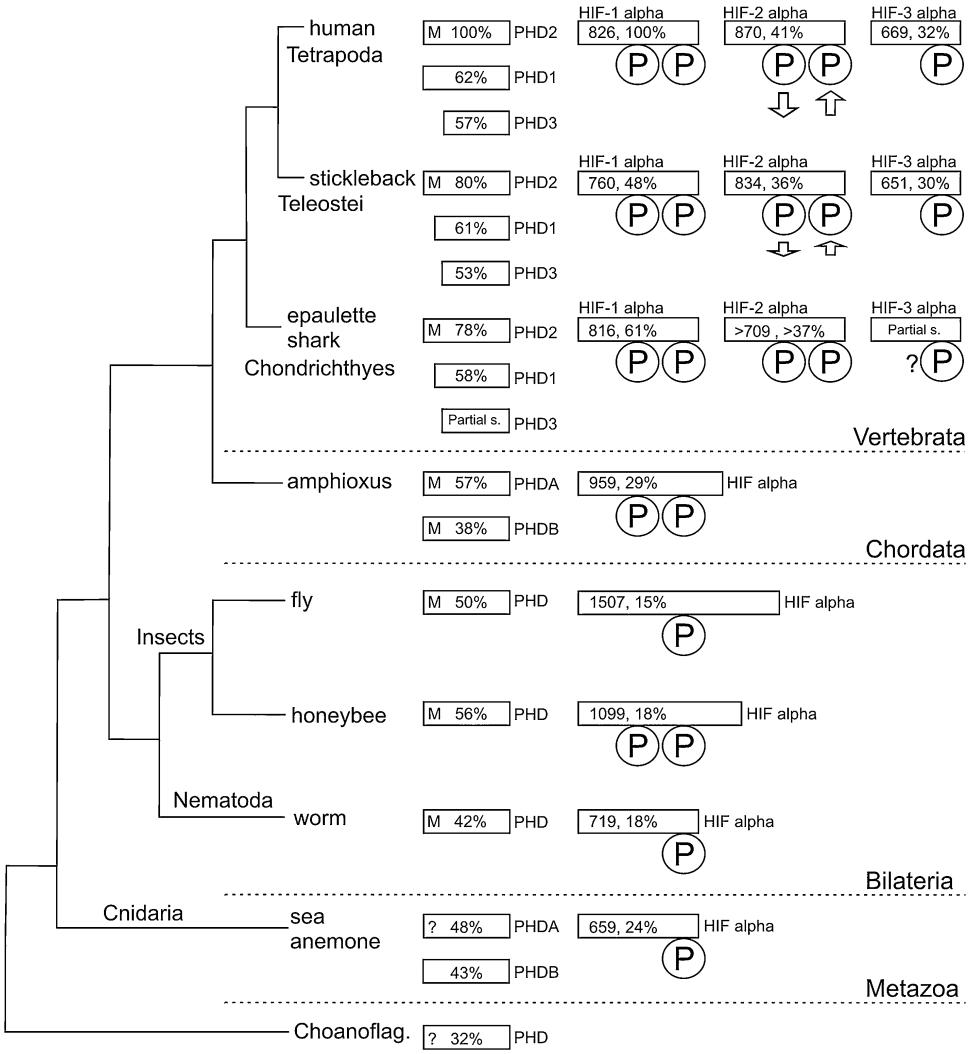

Figure 1. Overview of the Hypoxia inducible factor alpha (HIF alpha) and HIF Prolyl-4-hydroxylase (PHD) gene products in selected metazoan lineages. The left column of blocks shows the PHD proteins and the columns to the right show HIF alpha proteins. In PHD blocks the symbol M indicates the presence of a MYND (myeloid, Nervy, and DEAF-1)-type Zn2+ finger domain. In HIF alpha blocks the number indicates the length of the protein in amino acids. For both proteins the most ubiquitously expressed human paralogs (hsPHD2 and hsHIF-1 alpha) were chosen as a reference for amino acid identity comparisons that are indicated in percentage values in the blocks (for PHDs the catalytic domain and for HIF alphas whole CDS). In HIF the P symbol indicates proline hydroxylation motifs, in case where there is only one symbol it aligns with the CODD domain motif and when there are two symbols these indicate presence of N-terminal and C-terminal ODD domain motifs. The arrows below the P symbols indicate a change in relative importance of the ODD in the regulation. Cnidarians have only one prolyl hydroxylation motif whereas insect and chordate HIF alphas have two proline hydroxylation sites, corresponding to the vertebrate CODD core and NODD core. *C. elegans* has only one proline site that is very divergent from both C-terminal and N-terminal motifs. In the Arthropoda lineage the honeybee *(Apis mellifera) HIF homolog has both NODD and CODD motifs present (as does also the grass shrimp (Palaemonetes pugio), and the red flour beetle (Tribolium castaneum), but the model organism fruit fly (Drosophila melanogaster)* is an exception and has only CODD. Inspection of other Drosophila species suggests that NODD is present in other insects (bees and beetles), but was specifically lost in the Drosophila genus.

3.2. Molecular evolution of the HIF oxygen-response system in vertebrates

3.2.1. Hypoxia studies in epaulette sharks

Cartilaginous fishes, which include the elasmobranchs, are an early-branching gnathostome lineage that diverged from the lineage leading to tetrapods and teleosts approximately 450 million years ago (MYA) (Sansom et al, 1996). We chose an elasmobranch, the epaulette shark, for our studies of early vertebrate oxygen-sensing system for the following reasons. First, elasmobranch genes serve as an good (that is, relatively close) out-group for the molecular analysis of mammalian and teleost gene duplicates. Secondly, the epaulette shark is interesting from a physiological perspective. Epaulette sharks have adapted to tolerate hypoxia at a relatively high temperature, often being exposed to intermittent hypoxia during nocturnal low tides on shallow reef platforms (Renshaw et al, 2002). Studies of the HIF system in these fishes may prove useful in elucidating the evolution of hypoxia tolerance in early ancestral vertebrates in general (I, II).

The protective responses of epaulette sharks were studied in hypoxia and after hypoxia preconditioning. As to date no comparative studies have established suitable reference genes for quantitative PCR (qPCR) in cartilaginous fishes for any physiological conditions, the second chapter of this thesis provided the first one to do this. As hypoxia is a very strong stress factor, genes belonging to a number of functional categories are expected to be transcriptionally regulated in the course of hypoxic insult. Based on literature it is challenging to shortlist recommendable reference gene candidates for hypoxia studies. 9 reference candidates from various functional categories were sequenced and mRNA expression was monitored in four tissues: cerebellum, heart, gill and eye. The best ranking genes in our study were *eukaryotic translation elongation factor 1 beta* (*eef1b*), *ubiquitin* (*ubq*) and *polymerase (RNA) II (DNA directed) polypeptide F* (*polr2f*). The performance of the *ribosomal protein L6* (*rpl6*) was tissue-dependent (II).

The most remarkable finding of this study was an observation of clear discrepancy in the results of the very commonly used reference gene software (geNorm and NormFinder) with the ANOVA results. In our cerebellum data set ANOVA indicated statistically significant differences between treatments for genes that were ranked as the most stable candidates by both reference gene programs. Previously, NormFinder has been recommended over other methods, such as geNorm or Bestkeeper (Pfaffl et al, 2004), because it takes account of both the intra-group and the inter-group variation, whereas the latter methods do not have this ability (Hibbeler et al, 2008). Our results were not due to great intra-group individual variation due to the ecological sampling or laboratory procedures and we carefully validated the ANOVA results. In the cerebellum transcription of most genes, even in functionally different categories, was up-regulated in the single hypoxic insult. This may have influenced NormFinder ranking as it may conform to the mRNA expression pattern that the whole data set shares together (Andersen et al, 2004). These observations indicate that it is always necessary to include basic statistical tests in the analysis of reference genes for qPCR (II).

The actual target genes in our study, HIF-1 alpha, HIF-2 alpha, PHD2 and PHD1, are ubiquitously transcribed in the studied tissues. In epaulette shark hypoxia or hypoxia preconditioning did not alter the mRNA expression levels of HIF alphas and PHDs suggesting mainly post-translational regulation like in most studied animals. In response to hypoxia the mRNA expression of

HIF alphas is up-regulated in some hypoxia tolerant species, both in mammals (Shams et al, 2004) and teleosts (Law et al, 2006; Rahman & Thomas, 2007; Rissanen et al, 2006), but this was not the case for epaulette shark. Based on our results we predict that epaulette shark's adaptations to intermittent hypoxia do not require changes in the basal mRNA expression levels of activatory HIF alphas and that in this vertebrate species HIF activity is mainly post-translationally regulated (I).

3.2.2. Molecular evolution of vertebrate HIF alpha and PHD duplicates

The genome duplications that took place after the stem lineage of vertebrates was separated from invertebrates led to the refinement of oxygen sensing; here we have shown that cartilaginous fishes, but not lamprey, contain three duplicates of both HIFs and PHDs in their genomes (Figure 1). Unexpectedly, a novel cartilaginous HIF-3 alpha homolog grouped phylogenetically with teleost HIF-3 alphas. This may suggest differential loss of ancestral HIF-3 alpha duplicates in cartilaginous fishes/teleosts on the one hand, and tetrapods on the other. The acquisition of the inhibitory activity of HIF-3 alpha may have proceeded via multiple evolutionary mechanisms including faster rate of amino acid substitutions (relaxed selective constraint), exon loss and gain, insertions/deletions and alternative splicing (I).

Both vertebrate HIF-1 alpha and HIF-2 alpha are activatory transcription factors and have functionally diverged from ancestral HIF alpha forms. Taking into account the substitution rates and selection pressures we observed during vertebrate evolution, our results are more consistant with some predictions of the subfuctionalization model of both HIF-1 alpha and HIF-2 alpha rather than neofunctionalization of HIF-2 alpha or HIF-1 alpha. We detected functionally divergent sites both in the conserved PAS domains and in the HIF-PHD interaction domains. In functional divergence analysis we found clusters of significant sites in a particular location in teleost HIF-1 alpha PAS domains (see Figure 2, details in next section). These results together with the recent suggestion that HIF-2 alpha PAS B would require ligands for dimerization whereas HIF-1 alpha PAS B would not (Scheuermann et al, 2009) emphasize that important functional evolution may have occurred in the PAS dimerization domains. This is interesting as PAS domains are generally very conserved in all bHLH-PAS proteins (I).

In the functional divergence analysis of PHD paralogs, four of the detected functional divergence positions in PHD3 were close to the channel that binds HIF-1 alpha CODD peptide (Figure 3, E260G, D278R, R281L, G294E/A). We measured the minimum Euclidean distances between these positions and the CODD residues. All four residues are within 6 Angstroms of the HIF CODD (with the closest site being 294 at < 3 Angstroms), well within the usual range over which protein-protein interactions can occur (Gloor et al, 2005), and thus may be involved in the catalytic properties of the enzyme. Elsewhere, some of the FD positions with the greatest statistical significance (PHD1/3vs2 D246V/I, PHD1vs3 S247P and PHD3vs1 S248K) are concentrated in a beta2beta3 loop (Figure 3) that has been experimentally characterized. In PHD2, this loop displays considerable conformational changes upon ligand binding via Arg-252 and Asp-254 (Chowdhury et al, 2009), and is reported to determine the substrate specificity of the enzyme towards HIF CODD or NODD (Flashman et al, 2008; Villar et al, 2007).

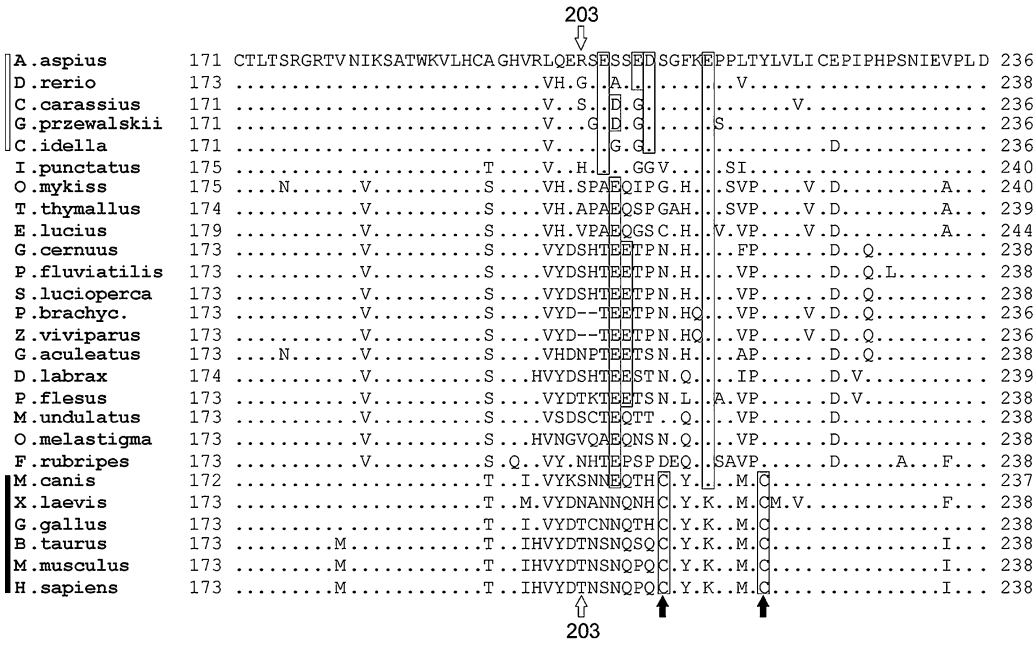

Figure 2. Alignment of the predicted HIF-1α protein for 20 teleost species and six other vertebrates (black bar) in the linking region between PAS A and PAS B domains. The mammalian HIF-1α is highly conserved in this region and only three species are shown (the last three sequences). Our analysis of functional divergence (I) identified in teleost two acidic residues (N205T/E/A, H209P/D, in boxes) and two cysteines (C210N/S and C219Y, in boxes with black arrows). Putative positively selected codon 203 (IV) is marked with a white arrow After codon 203 teleosts have an overrepresentation of acidic residues (D, aspartic acid; E, glutamic acid, in boxes) compared to mammals: most teleosts have three and most Cypriniformes (white bar) four acidic amino acids. A dot indicates the same amino acid in that position as in the first line.

In addition to analysis of HIF alpha and PHD paralogs, we studied HIF alpha and PHD gene orthologs inside major air-breathing (mammals) and water-breathing (teleosts) lineages. In our study orthologous genes mostly shared similar sequence characteristics suggesting that the functions of their products are more similar to each other

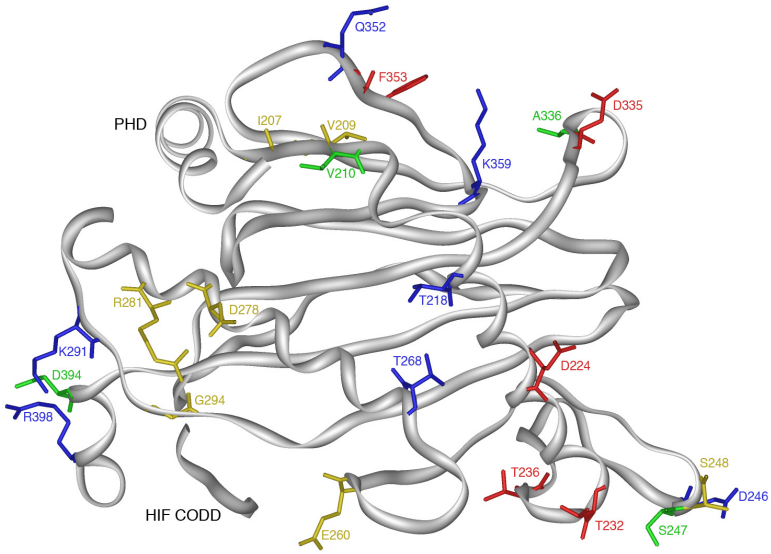

Figure 3. Residues under functional divergence in vertebrate HIF Prolyl-4-hydroxylases (PHDs). Colors denote the part of the phylogeny in which those sites experienced functional divergence: PHD1/3 after split from PHD2, blue; PHD2 lineage, red; PHD1 lineage, green; PHD3 lineage, yellow. A HIF fragment corresponding to the C-terminal oxygen-dependent degradation domain (CODD) is visible in the bottom-left of the image. The *Homo sapiens* PHD2 structure (Chowdury et al., 2009) was visualized in iMol (http://www.pirx.com/iMol/) and numbering is according to hsPHD2.

than those of paralogous genes. However, our results supported the notion that functionally relevant divergence is common between orthologs and not only between paralogs (Studer & Robinson-Rechavi, 2009). For example, when considering the functionally divergent sites preceding HIF-1 alpha PAS B and the teleost specific insertions preceding HIF-1 alpha NODD, mammalian HIF-1 alpha resembles more closely vertebrate HIF-2 alpha proteins than teleost HIF-1 alpha. (I)

Physiologically one of the most relevant questions is how HIF-2 alpha evolved to respond to oxygen differently than HIF-1 alpha. The oxygen sensitivity of post-translational HIF degradation is governed by the interactions of the HIF NODD and CODD with PHDs. First, considering PHDs, in PHD3 we detected FD positions close to the channel that binds HIF-1 alpha CODD peptide and in the PHD1/3 in the PHD beta2beta3 loop, which mediates PHD substrate specificity on HIF ODD peptides (Chowdhury et al, 2009; Flashman et al, 2008; Villar et al, 2007). We found that specifically PHD1, but not PHD2, has experienced some rate shifts in its early vertebrate evolution. Experimental studies have shown that PHD2 and PHD1/3 have different selectivity to the HIF NODD and CODD, so that all PHDs hydroxylate the principal proline site at CODD efficiently. However, only PHD2 efficiently hydroxylates the NODD proline site (Appelhoff et al, 2004; Hirsila et al, 2003; Koivunen et al, 2006). Secondly, we noticed that HIF alpha NODD and CODD had evolved differently, with especially HIF-2 alpha NODD exhibiting a faster rate of evolution and less stringent negative selection. Together this suggests that the evolution in both HIF-2 alpha NODD and PHD1/3 have resulted in less intensive oxygen dependent surveillance of HIF-2 alpha than HIF-1 alpha. The desensitization of HIF-2 alpha to PHDs could have created more opportunities for HIF-2 alpha to be particularly involved in long-term hypoxia adaptations, especially in the teleost lineage where more functionally divergent sites were detected in HIF-2 alpha than in HIF-1 alpha ODD. On the other hand, relaxation of oxygen-dependent regulation in HIF-2 alpha may have indirectly created a window for HIF-2 alpha to evolve specific regulatory functions (Covello et al, 2006; Mastrogiannaki et al, 2009) (I).

When considering structural evolution of proteins, conformationally diverse and dynamic proteins are expected to exhibit high evolvability (Tokuriki & Tawfik, 2009). HIF alpha is a transcription factor protein with less constrained structure in its C-terminal part of the protein, whereas PHDs are enzymes and their evolution is highly constrained by the structure of the catalytic domain. However, in the case of the differential oxygen sensitivities of HIF-1 alpha and HIF-2 alpha, we predict that the functional changes are due to sequence changes in both the enzyme and the substrate (transcription factor) duplicates and it is hardly feasible to assign either component with higher evolvability. For PHD, the available structure enabled more detailed analysis of the functional divergence sites, whereas for complete HIF alphas there was no structure information. Overall, we predict that HIF-1 alpha and PHD2 have retained relatively more of the ancient interaction characteristics that were present in the invertebrate HIF-PHD regulatory system, whereas HIF-2 alpha and PHD1/3 have experienced more functional divergence from the ancient state (I).

3.2.3. Molecular evolution of HIF-1 alpha in hypoxia-tolerant vs. hypoxia-sensitive vertebrates

For comparative evolutionary studies of oxygen-dependent systems teleost fishes

exhibiting a wide array of hypoxia-tolerant and hypoxia-sensitive species are a good choice. Since HIF-1 alpha is the most oxygen sensitive of the HIF alpha paralogs, in the third chapter we asked if we can observe any protein signatures in HIF-1 alpha associated with hypoxia-sensitivity vs. hypoxia-tolerance in teleost fish species. HIF-1 alpha was sequenced from a collection of hypoxia-tolerant and hypoxia-sensitive fish species sampled in Europe. Sequence analysis revealed that all teleost HIF-1 alpha proteins are shorter than their mammalian counterparts due to teleost specific deletions adjacent to the CODD. Without structural information on HIF alphas it is difficult to predict the functional significance of these deletions, but their presence could affect the HIF-PHD interactions (III). After careful examination of the novel protein sequences we could not find clear amino acid signatures that could be associated with the oxygen requirements of the fish species. However, in the C-terminal transactivation domain (C-TAD) of HIF-1 alpha two amino acid positions were specific to ostariophysi/cypriniformes-lineage, which generally includes hypoxia tolerant species. These were an aspartic acid and a histidine at positions +5 and +7 after Asn-803, which confers hypoxia dependent regulation via Factor inhibiting HIF (FIH). These positions are tentative candidates for *in vitro* studies using FIH enzyme and cyprinid like HIF peptides (III).

In the last chapter of this thesis (IV) the novel teleost gene sequences were used to compare the rate and mode of HIF-1 alpha molecular evolution between water-breathing teleost fishes and air-breathing mammals. In addition, the evolution of this gene within these groups was investigated in more detail. To this end amino acid substitution rate estimations were combined with various likelihood methods for estimating codon substitutions to detect different modes of selection. The predicted average evolutionary rate in teleost HIF-1 alpha proteins was approximately twice as fast as in mammalian HIF-1 alpha, but the predicted crucial interaction domains were found to be under stringent negative selection in all vertebrates (IV). We also evaluated evidence for positive selection in mammalian and teleost lineages with ML models of codon substitutions. No evidence for positive selection that was supported with all available methods was found in any of the studied lineages. The most convincing support for positive selection was not found in the oxygen-dependent degradation domain that confers HIF-1 alpha oxygen sensitivity, but in the bHLH-PAS domain that is responsible for DNA binding and dimerization (Figure 2.). This site 203 in human HIF-1 alpha is in the region linking the PAS A (85-158) and PAS B (228-298) domains (Card et al, 2005; Wang et al, 1995). In protein level analysis (I) we found significant functional divergence at positions N205T/E/A, H209P/D, C210N/S and C219Y in the same region of teleost HIF-1 alpha. Here, the functionally divergent sites in teleost HIF-1 alpha constitute a short region rich in acidic amino acids which is lacking from tetrapods, whereas teleosts lack two cysteine residues which are conserved in tetrapods and shark. Thus, some extra acidic residues have arisen and (redox-sensitive) cysteines have been specifically lost in teleost HIF-1 alpha. Together the analysis predicts that the above mentioned teleost specific features in HIF-1 alpha may underlie functionally significant differences between these two vertebrate lineages. These results (I, IV) also emphasize the advantage of combining both protein and DNA level models in the analysis of molecular evolution to achieve higher predictive power.

4. CONCLUDING REMARKS

Oxygen is crucial for the energy production of most metazoans. Regulation of gene functions by transcription factor proteins plays a central role in evolution. This thesis provides an evolutionary overview of the oxygen-dependent HIF transcription factor system. In early metazoan evolution the HIF system emerged from pre-existing PHD oxygen sensors and early bHLH-PAS transcription factors. The presence of PHDs in metazoan genomes predates the emergence of HIFs, which can be confidently detected in Cnidarians. Our analysis revealed an unexpected diversity of PHD genes and HIF sequence characteristics suggesting that the simple oxygen sensing systems of *Caenorhabditis* and *Drosophila* may not be typical of other bilaterian invertebrates.

For the experimental component of this thesis we chose to use a cartilaginous fish, the epaulette shark, as our main study species. Cartilaginous fish represent a basal lineage of vertebrates that diverged from the ancestor tetrapods and teleosts approximately 450 million years ago. For this reason, comparison of HIF system in cartilaginous fish with that of other vertebrates should allow inferences regarding the ancestral state of the system. The work included hypoxia preconditioning studies of the shark, and provided the first study of quantitative PCR reference genes for any cartilaginous fish species. Generally, the sequences studied provide a better starting point for future qPCR studies in this early branching vertebrate lineage. A crucial observation was that results from reference gene software may be biased by the collection of tested reference genes. This emphasizes the importance of basic statistical analysis of variance in reference gene evaluation. After validation of the proper reference genes, it was possible to monitor mRNA levels of target genes, *HIF alpha* and *PHD*.

Cartilaginous fishes have three duplicate copies of both HIF alpha and PHD, like tetrapods and teleosts. Based on our analysis and earlier mammalian results we predict that HIF-3 alpha genes have neofunctionalized to gain inhibitory functions in all vertebrates. Functional divergence was detected in both the HIF alpha PAS dimerization domains and HIF alpha ODD domains that interact with PHDs. Our analysis of functional divergence identified sites that may underlie the differential preference of PHDs for specific HIF duplicates. For PHD3 we found four FD sites in the vicinity of HIF binding channel and for PHD1/3 additional sites in the beta2beta3 loop, which mediates PHD substrate specificity (Chowdhury et al, 2009; Flashman et al, 2008; Villar et al, 2007). Our evolutionary analysis also revealed that the HIF-2 alpha NODD has experienced a faster rate of evolution and less stringent negative selection than other ODD cores. Taken together, these results suggest that the functional divergence in both HIF-2 alpha ODD and PHD1/3 may have resulted in less intensive oxygen-dependent regulation of HIF-2 than of HIF-1 alpha. Additionally, novel teleost HIF-1 alpha sequences were produced and it was found that teleost HIF-1 alpha proteins are shorter than their mammalian counterparts due to teleost specific deletions adjacent to the CODD core. However, the strongest evidence for positive selection was found in the PAS dimerization domain, not the ODD domain responsible for oxygen dependent regulation. Generally, our results supported the notion that functionally relevant divergence may take place between orthologs as well as paralogs (Studer & Robinson-Rechavi, 2009).

In the future, interesting avenues of research include the study of teleost specific deletions in HIF-1 alpha, the

role of the additional HIF duplicates in hypoxia-tolerant cyprinids and significance of cyprinid specific changes in HIF alpha hydroxylation motifs. With the increasing medical interest in the regulation of the HIF system by specific inhibitors of PHD activity (Harten et al, 2010; Myllyharju, 2009), evolutionary insights into PHD-HIF interactions will be useful from both the ecological and medical perspectives.

ACKNOWLEDGEMENTS

First and most importantly I would like to thank my supervisors Mikko Nikinmaa and Craig Primmer. Mikko, I thank you for all the material and immaterial support and guidance you provided. Mikko, thank you also for the confidence you had on me in the sense that I could exploit my own research interests inside the general framework of our projects. Craig, thank you for moving with your group to Turku at the right time, this made the research directions of my thesis possible. Craig, I thank you for the guidance and all your positive input. I sincerely thank the reviewers of the thesis Jay Storz and Johanna Myllyharju for their work and excellent comments on the thesis manuscript.

I have been part of two research groups during my PhD: Nikinmaa's group at the animal physiology and Primmer's group, the PnP, at the genetics. I want to express my greatest gratitude for members of both of these groups, support from both of these groups has been essential for the completion of this thesis. Many excellent scientists were transferring their knowledge to me in the laboratory and theory. When I very first started at animal physiology Kristiina "Krisu" Vuori taught RNA protocols and gave me invaluable support in getting my PhD started. Then, during the years everybody at animal physiology has been very kind and helpful. I thank Lotta Leveelahti, Wolfgang Wasser Nina Vuori, Eeva Rissanen, Piia Leskinen, Mirella Kanerva, Anna Lindross, Minna Vainio, Tiina Henttinen, Olli Arjamaa, Hanna Tranberg, Tomi Streng, Virpi Salonen and others.

In 2005, after working few months at animal physiology, all the PCR machines were moved to the laboratory of genetics and for following four years I worked at genetics. In the beginning Laura Buggiotti and JP Vähä showed me how to do things in the lab better and faster, thank you for this. I particularly want to thank Laura for being the sunshine of Italy in the darkness of Finland and organizing many activities in and out of the lab. The members of PnP group and other people at genetics were essential for my well-being and inspiration. I Thank Pop (Akarapong Swatdipong) for sharing the same room for four full years: you were the best possible officemate: you were always positive, helpful and friendly. I thank Reza Zahiri for cheerful presence during my final period in the office. I thank Anti Vasemägi for always having time for comments and extremely positive attitude. I thank Erica Leder for all the theoretical and practical supervision and advise - you were a great support not only in the article you co-authored but all the time. I thank Paula Lehtonen and Anni Tonteri for always being very positive and generating nice atmosphere. I thank Ville Aukee and Meri Lindqvist for always having time for me and solving many dilemmas. I sincerely thank Irma Saloniemi, Pirjo Lehtola, Raija Rouhiainen, Satu Koivumäki, Tatjaana Saarinen, Mikhael Ozerov, Niklas Wahlberg, Julien Leneveu, Mikko Nieminen, Harri Savilahti, Elsi Pulkkinen, Roghelio Fernandez Diaz, Seppo Nokkala, Christina Nokkala, Sanna Huttunen, Heidi Viitaniemi, Siim Kahar, Veronika Laine and others. I also thank Juha Merilä and all the scientists involved in the Centre of Excellence in Evolutionary Genetics and Physiology. I am grateful to Academy of Finland and Mikko Nikinmaa for funding. I thank Pekka Pamilo for organizing excellent work-shops.

In my second paper I got invaluable support on how to formalize statistical analyses to a scientific paper from Heikki Ryynänen - thank you Heikki. In the final period of my PhD the research on sharks was made possible by collaboration with Gillian Renshaw. Gillian, I thank you for all the scientific help, all the unforgettable

moments at Heron and taking care of me as you own son. Thanks for Grant Pritchard, Kevin Ashton, Jiri Neuzil and others at Griffith University. My last but not least scientific collaborator, Tom Williams, I met in Iowa 2009. Tom, I thank you for sharing the dorm room in Iowa and all the scientific help and great insights after that. During these years I was also was fortunate to supervise hardworking undergraduate students: big thanks for Kirsi Mikkola and Petra Vainio for their invaluable help in laboratory work. I also thank visiting student Arash Akbarzadeh.

In Turku I was lucky to have friends that also shared the shine and misery of this field called science. I want to thank Teijo Pellinen, Anssi Malinen, Arsi Rosengren, Tuomas Huovinen, Matti Salo and Jukka Rissanen for bringing light to unresolved questions and many discussion that went on and off around science along other more cheerful activities. I thank Anna Toivomäki-Pellinen and Katri Huovinen for tolerating molecular insights at odd hours. I thank the bio-orientated company of Sampo Lahtinen, Antti Valanne, Matti Lahti, Urpo Lamminmäki and Joonas Jämsen.

Originally I am not from Turku and to keep my roots and mind clear it has been crucial to meet up with other aboriginals from my hometown Jämsänkoski. Thank you Antti Blom, Matti Pulli, Mikko Kaamanen (Raisio), Sami Manninen, Timo Aaltonen, Mika Pylvänen, Lauri Toivonen, Juhani Koskela and other cheerful mates. The importance of Topi Valtakoski and Kalle Kukkamäki for this thesis should not be underestimated – if that 7 kg pike would not have escaped when you took me fishing that summer evening 1991, I might have lost my interest in fishes a long time ago.

Thank you mother, father and Roosa for always supporting me and giving me all that you have. During these years with you I have always had a peaceful place to load up my batteries and return to work full of energy. I thank Eeva-Liisa Virnes and other relatives for support. Finally, I want to thank Katja for being there for me every day and showing me that the world is not just about science.

REFERENCES

Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21:** 2104-2105

Andersen CL, Jensen JL, Orntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Research* **64:** 5245-5250

Appelhoff RJ, Tian YM, Raval RR, Turley H, Harris AL, Pugh CW, Ratcliffe PJ, Gleadle JM (2004) Differential function of the prolyl hydroxylases PHD1, PHD2, and PHD3 in the regulation of hypoxia-inducible factor. *Journal of Biological Chemistry* **279:** 38458-38465

Avery OT, MacLeod CM, McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *Journal of Experimental Medicine* **79:** 137-158

Bickler PE, Buck LT (2007) Hypoxia tolerance in reptiles, amphibians, and fishes: Life with variable oxygen availability. *Annual Review of Physiology* **69:** 145-170

Bishop T, Lau KW, Epstein ACR, Kim SK, Min J, O'Rourke D, Pugh CW, Gleadle JM, Taylor MS, Hodgkin J, Ratcliffe PJ (2004) Genetic analysis of pathways regulated by the von Hippel-Lindau tumor suppressor in Caenorhabditis elegans. *Plos Biology* **2:** 1549-1560

Bracken CP, Whitelaw ML, Peet DJ (2005) Activity of hypoxia-inducible factor 2 alpha is regulated by association with the NF-kappa B essential modulator. *Journal of Biological Chemistry* **280:** 14240-14251

Bretscher AJ, Busch KE, de Bono M (2008) A carbon dioxide avoidance behavior is integrated with responses to ambient oxygen and food in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 8044-8049

Brugarolas J, Lei K, Hurley RL, Manning BD, Reiling JH, Hafen E, Witter LA, Ellisen LW, Kaelin WG (2004) Regulation of mTOR function in response to hypoxia by REDD1 and the TSC1/TSC2 tumor suppressor complex. *Genes & Development* **18:** 2893-2904

Burki F, Kaessmann H (2004) Birth and adaptive evolution of a hominoid gene that supports high neurotransmitter flux. *Nature Genetics* **36:** 1061-1063

Card PB, Erbel PJA, Gardner KH (2005) Structural basis of ARNT PAS-B dimerization: Use of a common beta-sheet interface for hetero- and homodimerization. *Journal of Molecular Biology* **353:** 664-677

Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, Myers RM, Petrov D, Jonsson B, Schluter D, Bell MA, Kingsley DM (2010) Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. *Science* **327:** 302-305

Chang AJ, Bargmann CI (2008) Hypoxia and the HIF-1 transcriptional pathway reorganize a neuronal circuit for oxygen-dependent behavior in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 7321-7326

Chowdhury R, McDonough MA, Mecinovic J, Loenarz C, Flashman E, Hewitson KS, Domene C, Schofield CJ (2009) Structural Basis for Binding of Hypoxia-Inducible Factor to the Oxygen-Sensing Prolyl Hydroxylases. *Structure* **17:** 981-989

Conant GC, Wolfe KH (2008) Turning a hobby into a job: How duplicated genes find new functions. *Nature Reviews Genetics* **9:** 938-950

Covello KL, Kehler J, Yu HW, Gordan JD, Arsham AM, Hu CJ, Labosky PA, Simon MC, Keith B (2006) HIF-2 alpha regulates Oct-4: effects of hypoxia on stem cell function, embryonic development, and tumor growth. *Genes & Development* **20:** 557-570

Darwin C (1859) On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. London: John Murray.

Degnan BM, Vervoort M, Larroux C, Richards GS (2009) Early evolution of metazoan transcription factors. *Current Opinion in Genetics & Development* **19:** 591-599

Dejours P (1975) Principles of Comparative Respiratory Physiology. North-Holland: Amsterdam, pp. 1–253.

Dunwoodie SL (2009) The Role of Hypoxia in Development of the Mammalian Embryo. *Developmental Cell* **17:** 755-773

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *Bmc Bioinformatics* **5:** 1-19

Firth JD, Ebert BL, Ratcliffe PJ (1995) Hypoxic regulation of lactate-dehydrogenase-A - interaction between hypoxia-inducible factor-1 and CAMP response elements. *Journal of Biological Chemistry* **270:** 21021-21027

Flashman E, Bagg EAL, Chowdhury R, Mecinovic J, Loenarz C, McDonough MA, Hewitson KS, Schofield CJ (2008) Kinetic rationale for selectivity toward N- and C-terminal oxygen-dependent degradation domain substrates mediated by a loop region of hypoxia-inducible factor prolyl hydroxylases. *Journal of Biological Chemistry* **283:** 3808-3815

Follmann H, Brownson C (2009) Darwin's warm little pond revisited: from molecules to the origin of life. *Naturwissenschaften* **96:** 1265-1292

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151:** 1531-1545

Fukuda R, Zhang HF, Kim JW, Shimoda L, Dang CV, Semenza GL (2007) HIF-1 regulates cytochrome oxidase subunits to optimize efficiency of respiration in hypoxic cells. *Cell* **129:** 111-122

Gehring WJ, Kloter U, Suga H (2009) Evolution of the hox gene complex from an evolutionary ground state. In *Hox Genes* Vol. 88, pp 35-61.

Gloor GB, Martin LC, Wahl LM, Dunn SD (2005) Mutual information in protein multiple sequence alignments reveals two classes of coevolving positions. *Biochemistry* **44:** 7156-7165

Gorr TA, Gassmann M, Wappner P (2006) Sensing and responding to hypoxia via HIF in model invertebrates. *Journal of Insect Physiology* **52:** 349-364

Gorr TA, Tomita T, Wappner P, Bunn HF (2004) Regulation of Drosophila hypoxia-inducible factor (HIF) activity in SL2 cells - Identification of a hypoxia-induced variant isoform of the HIF alpha homolog gene similar. *Journal of Biological Chemistry* **279:** 36048-36058

Gu X (1999) Statistical methods for testing functional divergence after gene duplication. *Molecular Biology and Evolution* **16:** 1664-1674

Gu X (2006) A simple statistical method for estimating Type-II (Cluster-Specific) functional divergence of protein sequences. *Molecular Biology and Evolution* **23:** 1937-1945

Gu X, Vander Velden K (2002) DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. *Bioinformatics* **18:** 500-501

Gu YZ, Hogenesch JB, Bradfield CA (2000) The PAS superfamily: Sensors of environmental and developmental signals. Annual Review of Pharmacology and Toxicology **40:** 519-561

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52:** 696-704

Hahn MW (2009) Distinguishing Among Evolutionary Models for the Maintenance of Gene Duplicates. *Journal of Heredity* **100:** 605-617

Hampton-Smith RJ, Peet DJ (2009) From Polyps to People A HIghly Familiar Response to Hypoxia. In *Hypoxia and Consequences from Molecule to Malady*, Peers C, Haddad GG, Chandel NS (eds), Vol. 1177, pp 19-29.

Harten SK, Ashcroft M, Maxwell PH (2010) Prolyl Hydroxylase Domain Inhibitors: A Route to HIF Activation and Neuroprotection. *Antioxidants & Redox Signaling* **12:** 459-480

Heidbreder M, Frohlich F, Johren O, Dendorfer A, Qadri F, Dominiak P (2003) Hypoxia activates HIF-3 alpha at the transcriptional level. *Naunyn-Schmiedebergs Archives of Pharmacology* **367:** 51

Hershey AD, Chase M (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. *Journal of General Physiology* **36:** 39-56

Hibbeler S, Scharsack JP, Becker S (2008) Housekeeping genes for quantitative expression

studies in the three-spined stickleback Gasterosteus aculeatus. *Bmc Molecular Biology* **9:** 18

Hirsila M, Koivunen P, Gunzler V, Kivirikko KI, Myllyharju J (2003) Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. *Journal of Biological Chemistry* **278:** 30772-30780

Hochachka PW, Lutz PL (2001) Mechanism, origin, and evolution of anoxia tolerance in animals. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **130:** 435-459

Holmquist-Mengelbier L, Fredlund E, Lofstedt T, Noguera R, Navarro S, Nilsson H, Pietras A, Vallon-Christersson J, Borg A, Gradin K, Poellinger L, Pahlman S (2006) Recruitment of HIF-1 alpha and HIF-2 alpha to common target genes is differentially regulated in neuroblastoma: HIF-2 alpha promotes an aggressive phenotype. *Cancer Cell* **10:** 413-423

Hoogewijs D, Terwilliger NB, Webster KA, Powell-Coffman JA, Tokishita S, Yamagata H, Hankeln T, Burmester T, Rytkonen KT, Nikinmaa M, Abele D, Heise K, Lucassen M, Fandrey J, Maxwell RH, Pahlman S, Gorr TA (2007) From critters to cancers: bridging comparative and clinical research on oxygen sensing, HIF signaling, and adaptations towards hypoxia. *Integrative and Comparative Biology* **47:** 552-577

Hu CJ, Sataur A, Wang LY, Chen HQ, Simon MC (2007) The N-terminal Transactivation domain confers target gene specificity of hypoxia-inducible factors HIF-1 alpha and HIF-2 alpha. *Molecular Biology of the Cell* **18:** 4528-4542

Hughes BT, Espenshade PJ (2008) Oxygen-regulated degradation of fission yeast SREBP by Ofd1, a prolyl hydroxylase family member. *Embo Journal* **27:** 1491-1501

Ivanov SV, Kuzmin I, Wei MH, Pack S, Geil L, Johnson BE, Stanbridge EJ, Lerman MI (1998) Down-regulation of transmembrane carbonic anhydrases in renal cell carcinoma cell lines by wild-type von Hippel-Lindau transgenes. *Proceedings of the National Academy of Sciences of the United States of America* **95:** 12596-12601

Jaakkola P, Mole DR, Tian YM, Wilson MI, Gielbert J, Gaskell SJ, von Kriegsheim A, Hebestreit HF, Mukherji M, Schofield CJ, Maxwell PH, Pugh CW, Ratcliffe PJ (2001) Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O-2-regulated prolyl hydroxylation. *Science* **292:** 468-472

Jang MS, Park JE, Lee JA, Park SG, Myung PK, Lee DH, Park BC, Cho S (2005) Binding and regulation of hypoxia-inducible factor-1 by the inhibitory PAS proteins. *Biochemical and Biophysical Research Communications* **337:** 209-215

Jiang H, Guo R, Powell-Coffman JA (2001) A homolog of mammalian hypoxia-inducible factor-1 alpha is required for adaptation to low oxygen in Caenorhabditis elegans. *Developmental Biology* **235:** 411

Kaelin WG, Ratcliffe PJ (2008) Oxygen sensing by metazoans: The central role of the HIF hydroxylase pathway. *Molecular Cell* **30:** 393-402

Kajimura S, Aida K, Duan CM (2006) Understanding hypoxia-induced gene expression in early development: In vitro and in vivo analysis of hypoxia-inducible factor 1-regulated zebra fish insulin-like growth factor binding protein 1 gene expression. *Molecular and Cellular Biology* **26:** 1142-1155

Kewley RJ, Whitelaw ML, Chapman-Smith A (2004) The mammalian basic helix-loop-helix/PAS family of transcriptional regulators. *International Journal of Biochemistry & Cell Biology* **36:** 189-204

Kimura M (1968) Evolutionary rate at molecular level. *Nature* **217:** 624

King JL, Jukes TH (1969) Non-Darwinian evolution. *Science* **164:** 788

Koivunen P, Hirsila M, Gunzler V, Kivirikko KI, Myllyharju J (2004) Catalytic properties of the asparaginyl hydroxylase (FIH) in the oxygen sensing pathway are distinct from those of its prolyl 4-hydroxylases. *Journal of Biological Chemistry* **279:** 9899-9904

Koivunen P, Hirsila M, Kivirikko KI, Myllyharju J (2006) The length of peptide substrates has a marked effect on hydroxylation by the hypoxia-inducible factor prolyl 4-hydroxylases. *Journal of Biological Chemistry* **281:** 28712-28720

Koivunen P, Tiainen P, Hyvarinen J, Williams KE, Sormunen R, Klaus SJ, Kivirikko KI, Myllyharju J (2007) An endoplasmic reticulum transmembrane

prolyl 4-hydroxylase is induced by hypoxia and acts on hypoxia-inducible factor alpha. *Journal of Biological Chemistry* **282:** 30544-30552

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* **5:** 150-163

Kuraku S, Meyer A, Kuratani S (2009) Timing of Genome Duplications Relative to the Origin of the Vertebrates: Did Cyclostomes Diverge before or after? *Molecular Biology and Evolution* **26:** 47-59

Lando D, Peet DJ, Gorman JJ, Whelan DA, Whitelaw ML, Bruick RK (2002) FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Genes & Development* **16:** 1466-1471

Lau KW, Tian YM, Raval RR, Ratcliffe PJ, Pugh CW (2007) Target gene selectivity of hypoxia-inducible factor-alpha in renal cancer cells is conveyed by post-DNA-binding mechanisms. *British Journal of Cancer* **96:** 1284-1292

Lavista-Llanos S, Centanin L, Irisarri M, Russo DM, Gleadle JM, Bocca SN, Muzzopappa M, Ratcliffe PJ, Wappner P (2002) Control of the hypoxic response in Drosophila melanogaster by the basic helix-loop-helix PAS protein similar. *Molecular and Cellular Biology* **22:** 6842-6853

Law SHW, Wu RSS, Ng PKS, Yu RMK, Kong RYC (2006) Cloning and expression analysis of two distinct HIF-alpha isoforms - gcHIF-1alpha and gcHIF-4alpha - from the hypoxia-tolerant grass carp, Ctenopharyngodon idellus. *Bmc Molecular Biology* **7:** 15

Lendahl U, Lee KL, Yang H, Poellinger L (2009) Generating specificity and diversity in the transcriptional response to hypoxia. *Nature Reviews Genetics* **10:** 821-832

Lundin LG (1999) Gene duplications in early metazoan evolution. *Seminars in Cell & Developmental Biology* **10:** 523-530

Lynch VJ, Wagner GP (2008) Resurrecting the role of transcription factor change in developmental evolution. *Evolution* **62:** 2131-2154

Makino Y, Kanopka A, Wilson WJ, Tanaka H, Poellinger L (2002) Inhibitory PAS domain protein (IPAS) is a hypoxia-inducible splicing variant of the hypoxia-inducible factor-3 alpha locus. *Journal of Biological Chemistry* **277:** 32405-32408

Margoliash E, Smith EL (1965) Structural and functional aspects of cytochrome c in relation to evolution. Pp. 221–242 in V. Bryson and H. J. Bogel, eds. Evolving genes and proteins. New York: Academic Press.

Mason SD, Howlett RA, Kim MJ, Olfert IM, Hogan MC, McNulty W, Hickey RP, Wagner PD, Kahn CR, Giordano FJ, Johnson RS (2004) Loss of skeletal muscle HIF-1 alpha results in altered exercise endurance. *Plos Biology* **2:** 1540-1548

Mastrogiannaki M, Matak P, Keith B, Simon MC, Vaulont S, Peyssonnaux C (2009) HIF-2 alpha, but not HIF-1 alpha, promotes iron absorption in mice. *Journal of Clinical Investigation* **119:** 1159-1166

Maynard MA, Evans AJ, Shi W, Kim WY, Liu FF, Ohh M (2007) Dominant-negative HIF-3 alpha 4 suppresses VHL-null renal cell carcinoma progression. *Cell Cycle* **6:** 2810-2816

Mehta R, Steinkraus KA, Sutphin GL, Ramos FJ, Shamieh LS, Huh A, Davis C, Chandler-Brown D, Kaeberlein M (2009) Proteasomal Regulation of the Hypoxic Response Modulates Aging in C-elegans. *Science* **324:** 1196-1198

Mendel G (1866) Versuche über Pflanzen-Hybriden. *Verh. Naturforsch. Ver. Brünn* 4: 3–47 (in English in 1901, J. R. Hortic. Soc. 26: 1–32)

Metzen E, Stiehl DP, Doege K, Marxsen JH, Hellwig-Burgel T, Jelkmann W (2005) Regulation of the prolyl hydroxylase domain protein 2 (phd2/egln-1) gene: identification of a functional hypoxia-responsive element. *Biochemical Journal* **387:** 711-717

Mitchell P (1961) Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* **191:** 144

Mole DR, Blancher C, Copley RR, Pollard PJ, Gleadle JM, Ragoussis J, Ratcliffe PJ (2009) Genome-wide Association of Hypoxia-inducible Factor (HIF)-1 alpha and HIF-2 alpha DNA Binding with Expression Profiling of Hypoxia-inducible Transcripts. *Journal of Biological Chemistry* **284:** 16767-16775

Myllyharju J (2009) HIF Prolyl 4-Hydroxylases and their Potential as Drug Targets. *Current Pharmaceutical Design* **15:** 3878-3885

Nambu JR, Chen W, Hu S, Crews ST (1996) The Drosophila melanogaster similar bHLH-PAS gene encodes a protein related to human hypoxia-inducible factor 1 alpha and Drosophila single-minded. *Gene* **172:** 249-254

Nei M (2005) Selectionism and neutralism in molecular evolution. *Molecular Biology and Evolution* **22:** 2318-2342

Nei M, Niimura Y, Nozawa M (2008) The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nature Reviews Genetics* **9:** 951-963

Ng PKS, Chiu SK, Kwong TFN, Yu RMK, Wong MML, Kong RYC (2009) Functional characterization of two CITED3 homologs (gcCITED3a and gcCITED3b) in the hypoxia-tolerant grass carp, Ctenopharyngodon idellus. *Bmc Molecular Biology* **10:** 101

Nikinmaa M (2002) Oxygen-dependent cellular functions - why fishes and their aquatic environment are a prime choice of study. *Comparative Biochemistry and Physiology a-Molecular and Integrative Physiology* **133:** 1-16

Nikinmaa M, Rees BB (2005) Oxygen-dependent gene expression in fishes. *American Journal of Physiology-Regulatory Integrative and Comparative Physiology* **288:** R1079-R1090

Nikinmaa M, Tervonen V (2004) Regulation of blood haemoglobin concentration in hypoxic fish. In: Proceedings of the 7th International Symposium on Fish Physiology, Toxicology, and Water Quality, Tallinn, Estonia, May 12-15, 2003, G. L. Rupp and M. D. White (eds). U.S. Environmental Protection Agency, Ecosystems Research Division, Athens, Georgia, USA. EPA 600/R-04/049, pp 243-252

Ohno S (1970) Evolution by gene duplication. New York: Springer-Verlag.

Ohta T (1973) Slightly deleterious mutant substitutions in evolution. *Nature* **246:** 96-98

Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. *Biotechnology Letters* **26:** 509-515

Pocock R, Hobert O (2008) Oxygen levels affect axon guidance and neuronal migration in Caenorhabditis elegans. *Nature Neuroscience* **11:** 894-900

Pond SLK, Frost SDW (2005) A genetic algorithm approach to detecting lineage-specific variation in selection pressure (vol 22, pg 478, 2005). *Molecular Biology and Evolution* **22:** 1157-1157

Price RN, Uhlemann AC, Brockman A, McGready R, Ashley E, Phaipun L, Patel R, Laing K, Looareesuwan S, White NJ, Nosten F, Krishna S (2004) Mefloquine resistance in Plasmodium falciparum and increased pfmdr1 gene copy number. *Lancet* **364:** 438-447

Rahman MS, Thomas P (2007) Molecular cloning, characterization and expression of two hypoxia-inducible factor alpha subunits, HIF-1 alpha and HIF-2 alpha, in a hypoxia-tolerant marine teleost, Atlantic croaker (Micropogonias undulatus). *Gene* **396:** 273-282

Ratcliffe PJ (2007) HIF-1 and HIF-2: working alone or together in hypoxia? *Journal of Clinical Investigation* **117:** 862-865

Ravi V, Lam K, Tay BH, Tay A, Brenner S, Venkatesh B (2009) Elephant shark (Callorhinchus milii) provides insights into the evolution of Hox gene clusters in gnathostomes. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 16327-16332

Rees BB, Figueroa YG, Wiese TE, Beckman BS, Schulte PM (2009) A novel hypoxia-response element in the lactate dehydrogenase-B gene of the killifish Fundulus heteroclitus. *Comparative Biochemistry and Physiology a-Molecular & Integrative Physiology* **154:** 70-77

Reiling JH, Hafen E (2004) The hypoxia-induced paralogs scylla and charybdis inhibit growth by down-regulating S6K activity upstream of TSC in Drosophila. *Genes & Development* **18:** 2879-2892

Renshaw GMC, Kerrisk CB, Nilsson GE (2002) The role of adenosine in the anoxic survival of the epaulette shark, Hemiscyllium ocellatum. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **131:** 133-141

Richards JG (2009) Metabolic and molecular responses of fish to hypoxia. In *Hypoxia in Fishes*, Richards JG, Farrell, A.P., Brauner, C.J. (ed), pp 443–485. San Diego: Elsevier

Rissanen E, Tranberg HK, Sollid J, Nilsson GE, Nikinmaa M (2006) Temperature regulates hypoxia-inducible factor-1 (HIF-1) in a poikilothermic vertebrate, crucian carp (Carassius carassius). *Journal of Experimental Biology* **209:** 994-1003

Sagan L (1967) On origin of mitosing cells. *Journal of Theoretical Biology* **14:** 225

Sansom IJ, Smith MM, Smith MP (1996) Scales of thelodont and shark-like fishes from the Ordovician of Colorado. *Nature* **379:** 628-630

Scheuermann TH, Tomchick DR, Machius M, Guo Y, Bruick RK, Gardner KH (2009) Artificial ligand binding within the HIF2 alpha PAS-B domain of the HIF2 transcription factor. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 450-455

Shams I, Avivi A, Nevo E (2004) Hypoxic stress tolerance of the blind subterranean mole rat: Expression of erythropoietin and hypoxia-inducible factor 1 alpha. *Proceedings of the National Academy of Sciences of the United States of America* **101:** 9698-9703

Shen C, Nettleton D, Jiang M, Kim SK, Powell-Coffman JA (2005) Roles of the HIF-1 hypoxia-inducible factor during hypoxia response in Caenorhabditis elegans. *Journal of Biological Chemistry* **280:** 20580-20588

Stroka DM, Burkhardt T, Desbaillets I, Wenger RH, Neil DAH, Bauer C, Gassmann M, Candinas D (2001) HIF-1 is expressed in normoxic tissue and displays an organ-specific regulation under systemic hypoxia. *Faseb Journal* **15:** 2445-2453

Studer RA, Penel S, Duret L, Robinson-Rechavi M (2008) Pervasive positive selection on duplicated and nonduplicated vertebrate protein coding genes. *Genome Research* **18:** 1393-1402

Studer RA, Robinson-Rechavi M (2009) How confident can we be that orthologs are similar, but paralogs differ? *Trends in Genetics* **25:** 210-216

Takahashi Y, Takahashi S, Shiga Y, Yoshimi T, Miura T (2000) Hypoxic induction of prolyl 4-hydroxylase alpha(I) in cultured cells. *Journal of Biological Chemistry* **275:** 14139-14146

Taylor BL, Zhulin IB (1999) PAS domains: Internal sensors of oxygen, redox potential, and light. *Microbiology and Molecular Biology Reviews* **63:** 479

Taylor JS, Raes J (2004) Duplication and divergence: The evolution of new genes and old ideas. *Annual Review of Genetics* **38:** 615-643

Toft C, Williams TA, Fares MA (2009) Genome-Wide Functional Divergence after the Symbiosis of Proteobacteria with Insects Unraveled through a Novel Computational Approach. *Plos Computational Biology* **5:** e1000344

Tokuriki N, Tawfik DS (2009) Protein Dynamism and Evolvability. *Science* **324:** 203-207

Treinin M, Shliar J, Jiang HQ, Powell-Coffman JA, Bromberg Z, Horowitz M (2003) HIF-1 is required for heat acclimation in the nematode Caenorhabditis elegans. *Physiological Genomics* **14:** 17-24

Val AL (1995) Oxygen-transfer in fish - morphological and molecular adjustments. *Brazilian Journal of Medical and Biological Research* **28:** 1119-1127

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, *et al* (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3:** research0034.0031-0034.0011

Villar D, Vara-Vega A, Landazuri MO, Del Peso L (2007) Identification of a region on hypoxia-inducible-factor prolyl 4-hydroxylases that determines their specificity for the oxygen degradation domains. *Biochemical Journal* **408:** 231-240

Visel A, Rubin EM, Pennacchio LA (2009) Genomic views of distant-acting enhancers. *Nature* **461:** 199-205

Wagner A (2008) OPINION Neutralism and selectionism: a network-based reconciliation. *Nature Reviews Genetics* **9:** 965-974

Wallin IE (1923) The mitochondria problem. *American Naturalist* **57:** 255-261

Wang GL, Jiang BH, Rue EA, Semenza GL (1995) Hypoxia-inducible factor-1 is a basic-helix-loop-PAS heterodimer regulated by cellular oxygen tension. *Proceedings of the National Academy of Sciences of the United States of America* **92:** 5510-5514

Watson JD, Crick FHC (1953) Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid. *Nature* **171:** 737-738

Webb JD, Coleman ML, Pugh CW (2009) Hypoxia, hypoxia-inducible factors (HIF), HIF hydroxylases and oxygen sensing. *Cellular and Molecular Life Sciences* **66:** 3539-3554

Webster KA (2003) Evolution of the coordinate regulation of glycolytic enzyme genes by hypoxia. *Journal of Experimental Biology* **206:** 2911-2922

Wenger RH, Camenisch G, Stiehl DP, Katschinski DM (2009) HIF Prolyl-4-hydroxylase Interacting Proteins: Consequences for Drug Targeting. *Current Pharmaceutical Design* **15:** 3886-3894

Wenger RH, Stiehl DP, Camenisch G (2005) Integration of Oxygen Signaling at the Consensus HRE. *Science STKE* **r12**

West CM, van der Wel H, Wang ZA (2007) Prolyl 4-hydroxylase-1 mediates O-2 signaling during development of Dictyostelium10.1242/dev.000893. *Development* **134:** 3349-3358

Wiesener MS, Jurgensen JS, Rosenberger C, Scholze C, Horstrup JH, Warnecke C, Mandriota S, Bechmann I, Frei UA, Pugh CW, Ratcliffe PJ, Bachmann S, Maxwell PH, Eckardt KU (2002) Widespread, hypoxia-inducible expression of HIF-2 alpha in distinct cell populations of different organs. *Faseb Journal* **16:** 271

Williams TA, Codoner FM, Toft C, Fares MA (2010) Two chaperonin systems in bacterial genomes with distinct ecological roles. *Trends in Genetics* **26:** 47-51

Yang ZH (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24:** 1586-1591

Yang ZH, Swanson WJ (2002) Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Molecular Biology and Evolution* **19:** 49-57

Zhang JZ (2003) Evolution by gene duplication: an update. *Trends in Ecology & Evolution* **18:** 292-298

Zhang JZ, Zhang YP, Rosenberg HF (2002) Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nature Genetics* **30:** 411-415

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. Pp. 97–166 in V. Bryson and H. J. Vogel, eds. Evolving genes and proteins. New York: Academic Press.