On the Origin and Distribution of Prokaryotic Cellulases and Metazoan Oxygen Sensing Pathway

Inaugural dissertation

for the attainment of the title of doctor
in the Faculty of Mathematics and Natural Sciences
at the Heinrich Heine University Düsseldorf

presented by

Bing Song

from Xinjiang, China

Düsseldorf, 08 2022

From the Institute of Molecular Evolution
at the Heinrich Heine University Düsseldorf

Published by permission of the
Faculty of Mathematics and Natural Sciences at
Heinrich Heine University Düsseldorf

Supervisor: Prof. Dr. William F. Martin
Co-supervisor: Prof. Dr. Sven Gould

Date of the oral examination: 21/09/2022

Statement of declaration

I hereby declare that the thesis is original and has not been submitted in the same or similar form to other institutions. The work for the thesis was conducted at Heinrich Heine University Düsseldorf. I have not previously failed a doctoral examination procedure.

Düsseldorf, August 1st, 2022

Bing Song

Table of Contents

Summary 1

Introduction 3

1. Cellulases 3
    1.1. Categories and hydrolysis of cellulases 3
    1.2. Ecological and taxonomic distribution of cellulases 5
2. Hypoxia-inducible factor (HIF) 6
    2.1. Finding and subunits of the transcription factor HIF 6
    2.2. Functional enzymes in the HIF pathway 7
    2.3. The evolutionary origins and distribution of the HIF components 9
    2.4. The role of the HIF pathway in metazoan evolution 10

Outline of the dissertation 13

Manuscripts 14

1. Manuscript I: On the distribution of the prokaryotic cellulases in the natural environments 15
2. Manuscript II: The origin and distribution of the main oxygen sensing mechanism across metazoans 52

References 101

Acknowledgements 108

Summary

Cellulase project

Cellulose is the most abundant polysaccharide in plant cell walls and the hardest one to be degraded due to its structural complexity. It is composed of β-D-glucose units linked by (1 -> 4) glycosidic bonds in nature. Cellulases are the key enzymes for the hydrolysis of crystalline cellulosic fibers, and they can synergistically break β-glycosidic linkages in cellulose polymer to release glucose.

Cellulolytic enzymes are secreted by cellulolytic microorganisms and are widely spread in various fiber-based ecosystems. Generally, cellulose decomposition is a community process probably involving a key group of multiple enzymes, as very few degraders can produce the full suite of enzymes. Therefore, the unprecedented increased sequenced metagenome data is a valuable resource to help us investigate the critical biomass-degrading microbial community. Thus, more can be explored about the critical microbial consortium responsible for cellulose degradation across cellulose-rich environments using unprecedented extensive sequencing data.

This study investigated the ecological and taxonomic distribution of three main cellulase types among many metagenome sequences. All the investigated cellulases spread across the cellulose-rich environments with no significant difference. Beta-glucosidases and endoglucanases are more abundant than exoglucanases. Bacterioidetes and Clostridia may harbor ancestral diversity of the cellulase genes, while other groups have gained it recently. Distant bacterial lineages secrete the same cellulases in the same environment, demonstrating the high frequency of LGT of the cellulases and the microbial cooperation among different prokaryotes.

Hypoxia-inducible factor (HIF) project

Oxygen is one of the essential needs for metazoan physiology activities due to today's oxygen content occupying 21% of the atmospheric air. Among numerous survivors under the long earth's oxygen evolution history, metazoans (multicellular animals) have specific interaction mechanisms with earth's oxygen. The hypoxia-inducible factor (HIF) related oxygen metabolic pathway is metazoans' most crucial oxygen-sensing pathway. The discoverers of the mechanism were awarded The Nobel Prize in Physiology or Medicine in 2019.

Thus, the exact origin and distribution of the HIF-related oxygen pathway among all metazoans have recently been among the most popular research interests. While most studies only focused on the phylogenetic evolution of one or some core gene components within specific species or phylum, we investigated the co-evolution of the whole HIF pathway across modern animals along with a broader geological evolutionary perspective.

The findings demonstrate the following. First, the origin of the HIF gene components can trace back to the LMCA almost 800 Ma; second, the ancient widespread HIF oxygen sensing pathway helped metazoans tolerate changing oxygen environments. In addition, the wide oxygen affinity span of the HIF related enzymes reacting with several substrates (HIF, non-HIF, and collagen substrates) supports that these enzymes are mixed-function oxidases instead of only oxygen sensors, and the range of oxygen availability limits their catalytic rate.

Introduction

1. Cellulases

1.1. Categories and hydrolysis of cellulases

Cellulose is the commonest polysaccharide but the only crystalline skeletal form in almost all higher plant cell walls. Native cellulose is a linear polymer of β-D-glucose residues linked by 1 -> 4 links (Haworth 1929; O'sullivan 1997). On average, cellulose crystallites account for about 40% of each plant cell wall weight (Preston 1975). However, it is challenging to destruct straight, inflexible cellulosic chains with a strong capacity for hydrogen bonding. Moreover, the crystalline bundles were embedded in an amorphous hemicellulose-lignin matrix (Fig.1) (Hon 1994; O'sullivan 1997). Therefore, cellulolytic enzymes have been the subject of renewed interest because they catalyze cellulose microfibrils hydrolysis (Jayasekara and Ratnayake 2019).

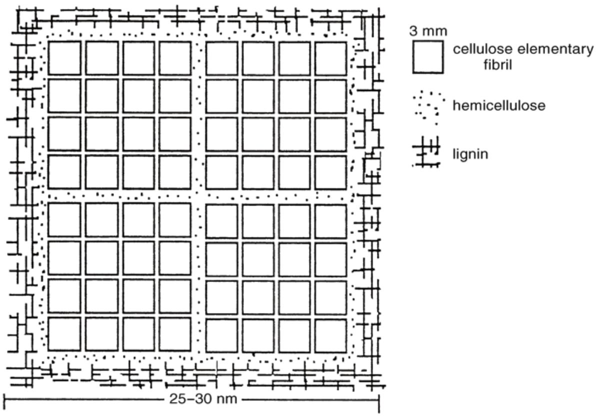


Fig.1 The cross-sectional model of the ultrastructural organization of the wood cell wall (O'sullivan 1997).

Cellulases are composed of multiple groups of enzymes that can hydrolyze the crystalline cellulose into glucose through a cascade of synergistic interactions of these enzymes (Fig.2). Cellulolytic enzymes are mainly classified into three functional categories, exoglucanases or cellobiohydrolases (CBH) (EC 3.2.1.91 (CBH I) and EC 3.2.1.176 (CBH II)), β-glucosidases (βG) (EC 3.2.1.21) and endoglucanases (EG) (EC 3.2.1.4) (Knowles et al. 1987). One notable characteristic of cellulolytic enzymes is the synergistic effect. The power of cooperative action is greater than the sum of the action of the individual enzyme components (Wood 1969). CBH typically hydrolyzes native insoluble cellulose to release the main hydrolysis product, cellobiose (Halliwell and Griffin 1973; Berghem et al. 1976; Henrissat et al. 1985). CBH II synergistically assists with CBH I, as CBH I hydrolyzes from the nonreducing end of the cellulose chain while CBH II acts on the reducing end (Enari and Niku-Paavola 1987). Then EG decomposes large soluble cellodextrins by breaking the interior bonds within a cellulose chain to generate cellobiose after CBH eliminates end-product inhibition exerted by cellobiose (Niku-Paavola et al. 1985). βG is responsible for hydrolyzing extensive shorter degradation product cellobiose into glucose.

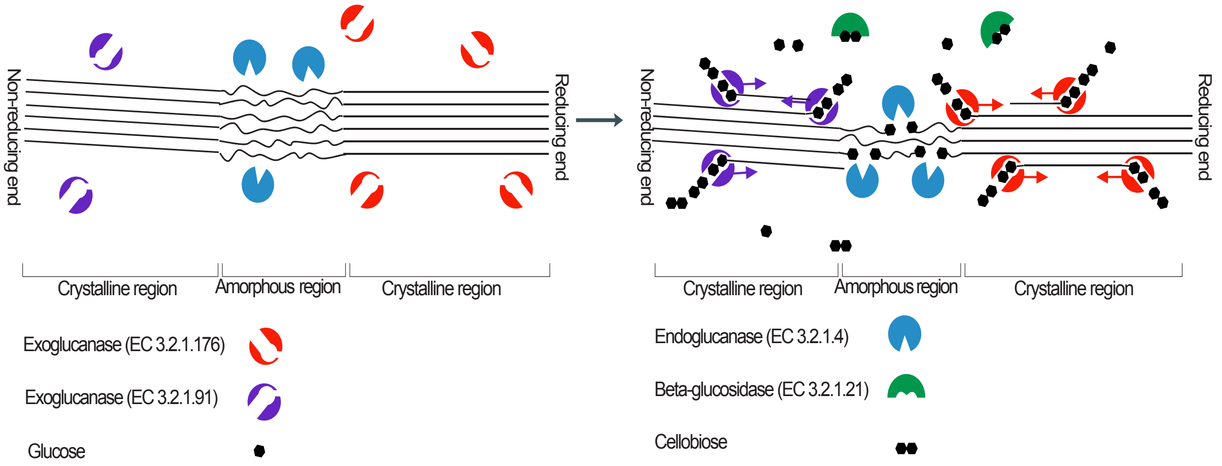


Fig.2 The cellulose degradation process.

1.2. Ecological and taxonomic distribution of cellulases

Cellulases and their producer are involved in cellulose degradation in various ecosystems. Sequences for cellulose utilization are abundant in human samples, intermediate in the phyllosphere and soil, and low in the other ecosystem types like sponge and coral (Berlemont and Martiny 2016). The medium soil macroaggregates show greater cellulose-degrading enzyme abundance (Choi et al. 2018). The relative gene abundances of EG and βG in herbivore gut microbiota are significantly higher than those of carnivore and carnivore-like (Guo et al. 2018). In the best studied cellulolytic environment rumen, the percentage of the major cellulose degrading bacteria is around 10% among 10^{10} bacteria/ml rumen content (Russell et al. 2009; Wilson 2011). In another well studied cellulose degrading environment compost, thermophilic cellulolytic bacteria dominate the community when cellulose degradation occurs in the thermophilic phase (Schloss et al. 2005; Székely et al. 2009). In a bamboo fiber soaking pit of an old paper-making plant, a microbial consortium with six dominant genera including Cloacibacterium, Paludibacter, Exiguobacterium, Acetivibrio, Tolumonas, and Clostridium cooperate through exchanging metabolites (Mai et al. 2019).

In the environment, cellulosic materials are mostly broken down by various microorganisms, including fungi, bacteria, and actinomycetes growing on cellulose. Cellulases are the most frequently identified traits with around 0.15% of genes and the most conserved traits in analyzed fungal genomes (Berlemont 2017). The distribution of functional types of cellulolytic enzymes in bacterial lineages is not random but mostly follows a Brownian motion evolution model (Berlemont and Martiny 2015). 38-80% of potential analyzed bacterial genomes, which include few potential cellulose degraders and more opportunist strains, encode one cellulase or more (Medie et al. 2012; Berlemont and Martiny 2015).

The opportunistic microorganisms do not carry the ability to decompose plant cell wall polymers but benefit from the enzyme activity performed by the cellulose degraders (Allison 2005; Romaní et al. 2006; Goldfarb et al. 2011). 20-62% of sequenced bacteria, which have specific bacterial lifestyles like autotrophy and parasitism, do not contain cellulase genes (Medie et al. 2012; Berlemont and Martiny 2015).

Cellulose degradation is a community process that involves a suite of enzymes, as cellulose degraders usually are not capable of producing the full suite of enzymes. So, a key microbial consortium may exist and play a keystone function in the natural cellulolytic process across different ecosystems. The increasing number of sequenced metagenome data provides a valuable resource for discovering the critical biomass-degrading microbial community (Stewart et al. 2018). Thus, there is still much we can explore about the distribution of the essential suite of cellulose degradation microbiome across major cellulose rich environments using massive metagenome data.

2. Hypoxia-inducible factor (HIF)

2.1. Finding and subunits of the transcription factor HIF

In 1992, Semenza and Wang detected a nuclear factor extracted from human hepatoblastoma cell line Hsep3B that can be induced by hypoxia and bind to the human erythropoietin (EPO) gene enhancer at binding site 1 (Semenza and Wang 1992). This nuclear factor was named hypoxia-inducible factor (HIF) and plays a crucial role in the hypoxic activation of EPO gene expression. However, HIF is later demonstrated to be involved in the transcriptional response to hypoxia in general mammalian cells instead of only binding at a site of human EPO gene enhancer (Wang and Semenza 1993a). It is further identified as a basic-helix-loop-helix-PAS heterodimeric protein complex regulated by cellular O_2 tension (Wang and Semenza 1993a; Wang et al. 1995). It is built of two subunits - 120-kDa HIFA and 91-94-kDa HIFB, both of which are basic-helix-loop-helix-PAS proteins (bHLH) (Wang and Semenza 1995). The RNA and protein levels of HIFA and HIFB can be induced when cells are exposed to 1% O_2 but decay rapidly when exposed to 20% O_2 (Wang et al. 1995).

The HIFA subunit has three closely related isoforms encoded by three distinct gene loci, HIF1A, HIF2A, and HIF3A. HIF1A and HIF2A have similar domain structure and regulation manner, but the functional divergence between the two is also found (Schofield and Ratcliffe 2004; Graham and Presnell 2017). Besides having a similar conserved functional domain to the first two isoforms, HIF3A is also regulated by multiple splicing variants (Maynard et al. 2003). The HIFB subunit has two closely related forms, HIF1B and HIF2B. The HIFA subunit content is highly inducible by deficient oxygen, while the endogenous HIFB subunit is sufficient and present in hypoxic and non-hypoxic cells (Semenza et al. 1996).

2.2. Functional enzymes in the HIF pathway

In the oxygen sensing HIF mechanism, three enzyme groups - HIF prolyl hydroxylases (PHDs), factor inhibiting HIF (FIH), and von-Hippel Lindau tumor suppressor protein (VHL) target the HIFA substrates to function as regulators. When oxygen is sufficient, PHDs and FIH hydroxylate the proline (Pro) and asparagine (Asn) residues on HIFA subunits. The hydroxylated hydroxyl group of HIFA is led to the ubiquitylation (Ub) and immediate degradation targeted by

VHL. PHDs, FIH, and VHL are inhibited when there is no available oxygen to permit their activities.

PHDs are a subfamily of 2-oxoglutarate and dioxygen-dependent enzymes responsible for the hydroxylation of one or two critical prolines in the oxygen-dependent degradation region of HIFA subunits (Schofield and Ratcliffe 2004). There are four PHD isoenzymes, PHD1, PHD2, PHD3, and PHD4 (Bruick and McKnight 2001; Epstein et al. 2001; Ivan et al. 2002; Laitala et al. 2012). PHD1-3 have a high rate of catalysis on the C-terminal degradation domain (CDD) of HIFA subunits, with an order of PHD2 = PHD3 > PHD1 (Tuckerman et al. 2004). PHD1 is mainly present in the nucleus (Metzen et al. 2003). The membrane-anchored PHD2 is degraded by a ubiquitin-independent proteasomal pathway, and stable cytosolic PHD2 functions as an active prolyl hydroxylase (Barth et al. 2009). So, PHD2 is largely located in the cytoplasm. PHD3 is rapidly degraded under normoxia and can regulate a distinct transcriptional response to hypoxia (Jaakkola and Rantanen 2013; Zhang et al. 2014). PHD3 is equally located in the cytoplasm and nucleus (Metzen et al. 2003). Besides the HIF pathway, PHD3 participates in diverse metabolic activities as an ancient signaling protein (Jaakkola and Rantanen 2013; Place and Domann 2013). The endoplasmic reticulum transmembrane prolyl 4-hydroxylase (P4H-TM) is the fourth prolyl 4-hydroxylase (PHD4), which can regulate EPO production and erythropoiesis (Laitala et al. 2012).

FIH is an asparaginyl hydroxylase enzyme that can hydroxylate the asparagine residue (Asn) on the HIFA subunit to block its interaction with the transcriptional co-factors - p300 and the cyclic AMP response element-binding (CREB) protein (CBP) (Lando et al. 2002). It is exclusively present in the cytoplasm (Metzen et al. 2003). FIH prefers to hydroxylate long peptide substrates, and its hydroxylation of HIF1A is more efficient than that of HIF2A (Koivunen et al. 2004; Bracken et al. 2006). Although both are oxygen sensors, FIH is more tolerable to oxygen deprivation than PHDs in cellular activity (Tarhonskaya et al. 2015).

VHL refers to the protein encoded by the von-Hippel Lindau tumor suppressor gene (pVHL). Normally, its function is to bind and transport the hydroxylated proline residue (OH) on the oxygen dependent domain (ODD) of the HIFA subunit for proteasome degradation (Maxwell et al. 1999). The conserved methionine of HIF1A results in it binding more tightly and having a higher affinity on VHL than HIF2A (Tarade et al. 2019).

2.3. The evolutionary origins and distribution of the HIF components

When the HIF-PHD-FIH-VHL system exactly occurred among the metazoans remains unclear. It is possible that this oxygen sensing mechanism occurred in the last metazoan common ancestor (LMCA), as a functional HIF pathway is demonstrated to exist and targets on glycolytic and metabolic enzymes in the simplest animal - the placozoan *Trichoplax adhaerens* (*T. adhaerens*) (Loenarz et al. 2011; Rytkönen and Storz 2011). Remarkable conserved structural features exist in both *T. adhaerens* and human PHDs (Lippl et al. 2018). In addition, the metazoan HIF-mediated oxygen homeostasis characteristics are shared by non-metazoan fungi and amoeba. For example, the fission yeast *Schizosaccharomyces pombe* has the prolyl hydroxylase Ofd1, which can regulate the binding of Nro1 to its dioxygenase domain in an oxygen-dependent manner to control the stability of Sre1N (Lee et al. 2009). During the development of the unicellular soil amoeba *Dictyostelium*, PHD1 functions as an oxygen sensor required for O-glycosylation of Skp1, one subunit of E3SCF-Ub-ligases (West et al. 2007). However, the earliest animal demosponge *Halichondria panicea* can survive under 0.5-4% present atmospheric level (PAL), so it seems likely that LMCA can adapt well under low-oxygen conditions without the HIF pathway (Mills et al. 2014, 2018).

The HIF pathway is thought to be an extreme common trait of all metazoan taxa (Loenarz et al. 2011; Rytkönen and Storz 2011). However, recent studies revealed that this main oxygen sensing pathway might not be that common across metazoans. For instance, despite the loss of major regulatory components of the HIF pathway - HIFA and PHD in marine intertidal crustacean *Tigriopus californicus*, it still can tolerate 24 hours of extreme low oxygen level (Graham and Barreto 2019). During the anoxia process, the pathway related to chitin metabolism and cuticle reorganization has been changed consistently and could play as an alternative solution in animals without complete respiratory structures (Graham and Barreto 2019). Independent losses of the critical regulators of the HIF system have also been found in other Crustacean groups, and alternative mechanisms for metazoan hypoxia response may exist (Graham and Barreto 2020).

2.4. The role of the HIF pathway in metazoan evolution

The last common ancestor of metazoans (multicellular animals) is estimated to arise nearly 800 million years (Ma) ago when the atmospheric oxygen content is 1% of the PAL (Erwin et al. 2011; Lyons et al. 2014). Almost all metazoans use oxygen for their aerobic energy metabolism fuel and physiological enzymatic cofactor. However, the oxygen content in the environment has kept changing during the process of early animal evolution. Too much oxygen supply causes toxic byproducts, and a deficient oxygen supply can impact the rate of aerobic energy production and physiological activities in animals. To sense a broad oxygen spectrum, metazoans use dioxygenases in the HIF mechanism to keep the cellular homeostasis (Kaelin and Ratcliffe 2008).

Under low oxygen conditions, the HIFA subunit is directly transported into the cell nucleus to combine with the HIFB subunit as a heterodimer complex. Then, this heterodimer functions as the transcription factor binding on downstream genes' corresponding hypoxia response elements (HRE) to stimulate or repress their transcriptional activities. While under normal oxygen conditions, HIFA is kept in a state of repression by the activity of PHD. Proline and asparaginyl residues in HIFA subunits are first degraded by PHD and FIH, respectively. Then, the hydroxylated proline residue on the HIFA molecule is ubiquitinated by the VHL and degraded by proteasomes within the cytoplasm (Fig.3).

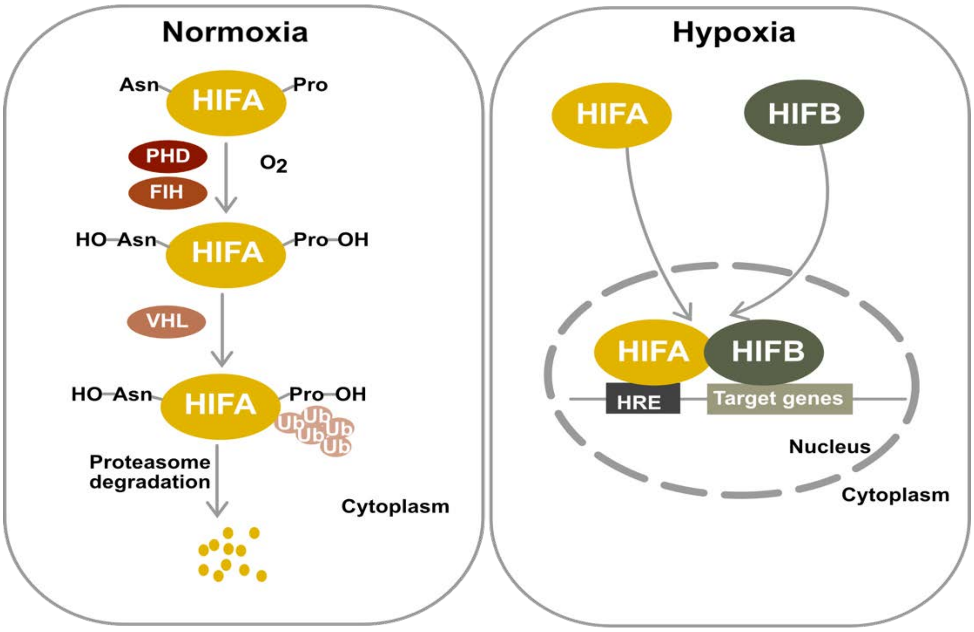


Fig.3 The HIF-PHD-FIH-VHL pathway.

To adapt to similar hypoxic environmental stress, convergent evolution of the HIF related genes PHD2 and HIF2A was found in two high-altitude duck species (Graham and McCracken 2019). In addition, hypoxia-inducible transcription factors provided animals with rudimentary control of stemness and pseudohypoxia driven by HIF1A and HIF2A to allow them to cope with fluctuating oxygen concentrations (Hammarlund et al. 2018). Thus, the animal diversification during the Cambrian explosion (555 - 500 Ma) is not only associated with the increasing oxygenic extent but also contributed by the cellular hypoxia-response machinery (Hammarlund et al. 2018; Hammarlund 2020). The author suggested that oxygen sensing should be more focused as a mechanism that allows multicellularity to cope with and utilize oxygen fluctuations, instead of being considered as a primary hypoxia response for aerobic respiration (Hammarlund et al. 2020).

Outline of the dissertation

In this thesis, there are two projects. The first project focuses on exploring the distribution and the abundance of these main cellulase types among cellulose rich environments (human gut, soil, saline water, and rumen). To see the ecological and phylogenetic distribution of the main cellulose degraders in each environment, all potential cellulose degradation enzyme sequences were searched and identified using the downloaded standard cellulases sequences as queries among large metagenome datasets sourced from mentioned four environments.

The second project is to explore the origin and distribution of the main oxygen sensing mechanism across metazoans. To trace the origin of the HIF-PHD-FIH-VHL system among metazoans, we investigated the phylogenetic distribution and phylogeny of 11 of its critical components across 566 eukaryotic genomes. Furthermore, we compared the earth's oxygen evolution history and the presence-absence pattern (PAP) of HIF components to determine how the HIF pathway evolved along with the environmental oxygen levels.

Manuscripts

1. Bing Song, Fernando D. K. Tria, Josip Skejo, Madeline C. Weiss, William F. Martin. On the distribution of the prokaryotic cellulases in the natural environments. Unpublished work.

2. Bing Song, Luca David Modjewski, Nils Kapust, William F. Martin. The origin and distribution of the main oxygen sensing mechanism across metazoans. In preparation for publication.

1. Manuscript I: On the distribution of the prokaryotic cellulases in the natural environments

Bing Song[1†], Fernando D. K. Tria[1†], Josip Skejo[12†], Madeline C. Weiss[1], William F. Martin[1*]

1 Insitute of Molecular Evolution, Heinrich – Heine University Düsseldorf, Universitätstraße 1, Germany;

2 Evolution Lab, Division of Zoology, Department of Zoology, Faculty of Science, University of Zagreb, Croatia;

† authors of equal contribution;

* corresponding author, e-mail: bill@hhu.de.

Contributed by Bing Song:

I downloaded all datasets used in the study, searched and filtered potential cellulase hits, and reconstructed cellulase clusters. I also wrote and corrected the draft, did all analyses, and made all figures and tables, except the Fig.3.

Abstract

Cellulose is a carbon source widespread in nature. However, it is a difficult task for any organism to get carbon atoms from the cellulose as it has a highly complex structure. Only a few taxonomic groups (bacteria and fungi) are known to decompose cellulose. They do it by producing cellulases, the various enzymes which break beta-glycosidic bonds in the cellulose, getting the simple units from this complex source. The main cellulase types found in nature are endoglucanases, exoglucanases, and beta-glucosidases. In this study, the distribution and the abundance of these main cellulase types are assessed in 27 kinds of the natural environment, for which metagenome data are available. Metagenome sequences were searched for the presence of cellulases by environmental and taxonomic annotations (prokaryotic phyla). All the investigated cellulases seem widespread across environments, and we find no significance among the four most cellulose-rich environments (rumen, saline water, human gut, and soil). Beta-glucosidases and endoglucanases seem to be more common in nature than exoglucanases. Bacterioidetes and Clostridia may harbor ancestral diversity of the cellulase genes, while other groups seem to have gained it recently. The fact that distant bacterial lineages (e.g., Gammaproteobacteria and Actinobacteria) secrete the same cellulases in the same environment indicates not only the high frequency of the lateral gene transfer (LGT) of the cellulases in the cellulose rich environments, but also the extent of the cooperation between different prokaryotes in cellulose degradation.

Introduction

As the most abundant form of photosynthetically fixed carbon in nature, cellulose is produced mainly by plants (Archaeplastida) as the primary structural molecule of their cell walls. In its pure form, cellulose would be crystalline and insoluble (Atalla 1990), but in nature, it is most often found inside a hemicellulose and lignin matrix (Hon 1994). Because of both, its structure and function, cellulose is a challenging molecule for organisms to degrade. The degradation is accomplished through a cascade of reactions, each catalyzed by a different enzyme that typically breaks beta-glycosidic bonds. Those enzymes are referred to as cellulases and are classified into three main functional categories: beta-glucosidases, endoglucanases, and exoglucanases (Knowles et al. 1987). Exoglucanases typically hydrolyze native insoluble cellulose by attacking from chain ends to release cellobiose or cellodextrins (Halliwell and Griffin 1973; Berghem et al. 1976; Henrissat et al. 1985). Endoglucanases penetrate inner linkages within a cellulose chain to catalyze soluble cellodextrins into cellobiose by an endo-type mechanism (Niku-Paavola et al. 1985). Finally, beta - glucosidases hydrolyze cellobiose to glucose, which cells then utilize (Enari and Niku-Paavola 1987). Cellulose reduction into cellobiose is an energy-costly process and primarily involves the action of endoglucanases that make large size cellodextrins available to assimilated fragments. A functionally diverse set of cellulases are secreted by a range of microorganisms in nature, predominantly fungi, and bacteria (Sharma et al. 2016). The abundance of the prokaryotic cells exceeds the quantity of the fungal (eukaryotic) cells in nature (Kikani 2018). Hence, this study focuses on prokaryotic cellulases.

Here we contribute to the knowledge of the distribution and abundance of the cellulases in cellulose-rich environments using metagenomes as a proxy for the underlying microbial community. This approach permitted us to uncover the relative importance of functionally different cellulases and the main prokaryotic lineages responsible for cellulose degradation in nature.

Materials and methods

Compilation of protein sequences for known bacterial cellulases. The protein sequences of 122 bacterial cellulases encompassing three KEGG catalytic functions: 10 exoglucanase sequences (8 for EC 3.2.1.91 and 2 for EC 3.2.1.176), 22 endoglucanase sequences (EC 3.2.1.21) and 90 beta-glucosidase sequences (EC 3.2.1.4) were downloaded from the Uniprot database (Bateman 2019), as of May 2018 (Supplement 1, Supplement 2). EC represents the Enzyme Commission number (https://enzyme.expasy.org).

Identification of cellulases in metagenome nucleic samples. We have obtained 166,896,572 sequences from 2,305 metagenomes belonging to 329 bioprojects (Supplement 3) from NCBI (May 2018). The protein-coding genes from Uniprot were used as queries against the metagenome nucleic sequences. We have used the blastx command of the DIAMOND sequence aligner version v.0.9.24.125 to perform sequence alignments (Buchfink et al. 2014). Hits with e-value > 10^{-5} and query coverage < 50% were discarded. Only the best hit for each query-metagenome sequence was retained, as judged by bitscore, and non-overlapping hits were then selected from each metagenome to avoid oversampling the same gene (Fig.1). The resulting set of metagenome nucleotide hits were translated into protein sequences in all six reading frames with transeq, version 6.6.0.0 (Rice et al. 2000), and the reading frames aligned to the corresponding metagenome nucleotide hits using tblastn, version 2.5.0+ (McGinnis and Madden 2004). The reading frame generating the best alignment, as judged by bitscore was selected and subjected to HMMER version 3.0 HMMscan (Wheeler and Eddy 2013) for the glycoside hydrolase (GH) domains with an e-value cutoff $< 10^{-5}$. Cellulose degradation requires the synergistic action of cellulases from multiple GH families (Gilbert 2010).

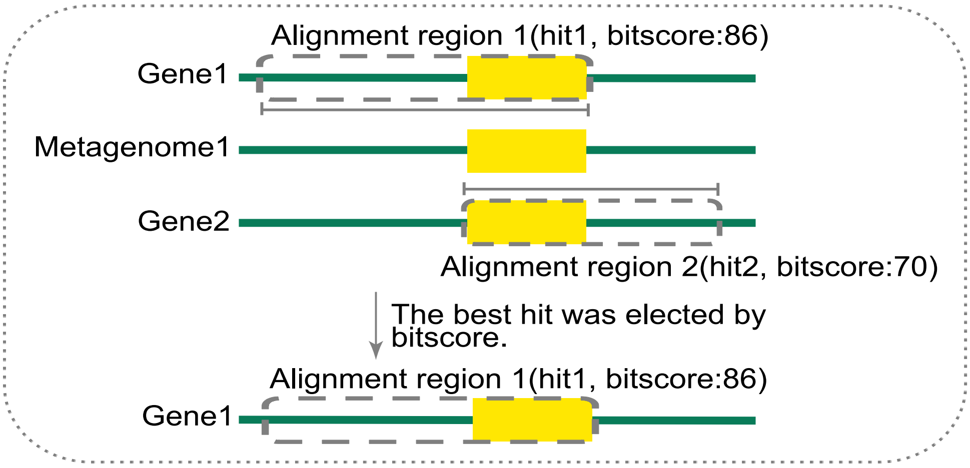


Fig.1 The step of selecting the best hit from overlapping hits within each metagenome. The yellow part means the overlapping region between two hits. The two dotted boxes represent two respective alignment regions within the metagenome1. Finally, hit1 was selected because it has a higher bitscore than hit2.

Construction of metagenome-derived cellulase clusters. The cellulase hits for the different KEGG categories were processed separately to cluster the protein sequences into families. Pairs of the protein sequences were globally aligned using the Needleman–Wunsch algorithm (Needleman and Wunsch 1970) implemented in the EMBOSS package. Protein pairs with global identity values above 25% were retained. The retained identity values of the pairwise alignments were used for clustering using the default setting of the Markov Clustering Algorithm (MCL) (Dongen 2000).

Taxonomical assignment of cellulase genes. The taxonomical rank was assigned to each metagenome-derived cellulase using a database of 5,655 prokaryotic genomes obtained from the NCBI RefSeq database version in June 2012 (Pruitt et al. 2011). We have characterized the cellulases at the phylum level by assigning the taxonomical affiliations based on the best blast hit on the prokaryotic genomes, with an e-value cutoff $< 10^{-5}$.

Correlation among cellulase genes. The Spearman correlation analysis was based on the abundance of exoglucanases, endoglucanases, and beta-glucosidases in each metagenome (Supplement 4). This analysis was done in the Rstudio-Server v1.3.1073 (RStudio Team 2021) and R v4.0.0 (Team 2021) with packages - ggplot2, ggpubr, reshape2, and gridExtra.

Principal component analysis (PCA). The PCA was performed in the Rstudio-Server v1.3.1073 (RStudio Team 2021) and R v4.0.0 (Team 2021) with packages - FactoMineR and factoextra. PCA analysis of the cellulases across the 1,515 metagenomes from the rumen, saline, human gut, and soil is shown in the Supplement 5. The environment is annotated with color. Prediction ellipses are such that with the probability of 0.95.

Results

Beta-glucosidases are the commonest cellulases in nature.

Cellulases were found in 1,735 metagenomes altogether from 225 bioprojects, which includes 35 prokaryotic phyla from 27 distinct environments (Fig. 2). Within the metagenomes as mentioned above, we have quantified overall 12,837 metagenome-derived cellulases. The set of metagenome-derived cellulases encompasses three catalytic functions: exoglucanases (1,042), endoglucanases (5,685), and beta-glucosidases (6,110) (Table. 1). The total number of the recorded beta-glucosidases is a bit (= 1.07 times) higher than endoglucanases and much higher exoglucanases (= 5.86 times), respectively. The ratios are, of course, variable across the prokaryotic phyla. According to the Enzyme Commission number category, we have assigned the metagenome-derived cellulases to the protein clusters and have identified that our 12,837 cellulases belong to 136 groups altogether (Table 1). Endoglucanases exhibit the highest diversity of the clusters, while beta-glucosidases and exoglucanases are much lesser (Table 1).

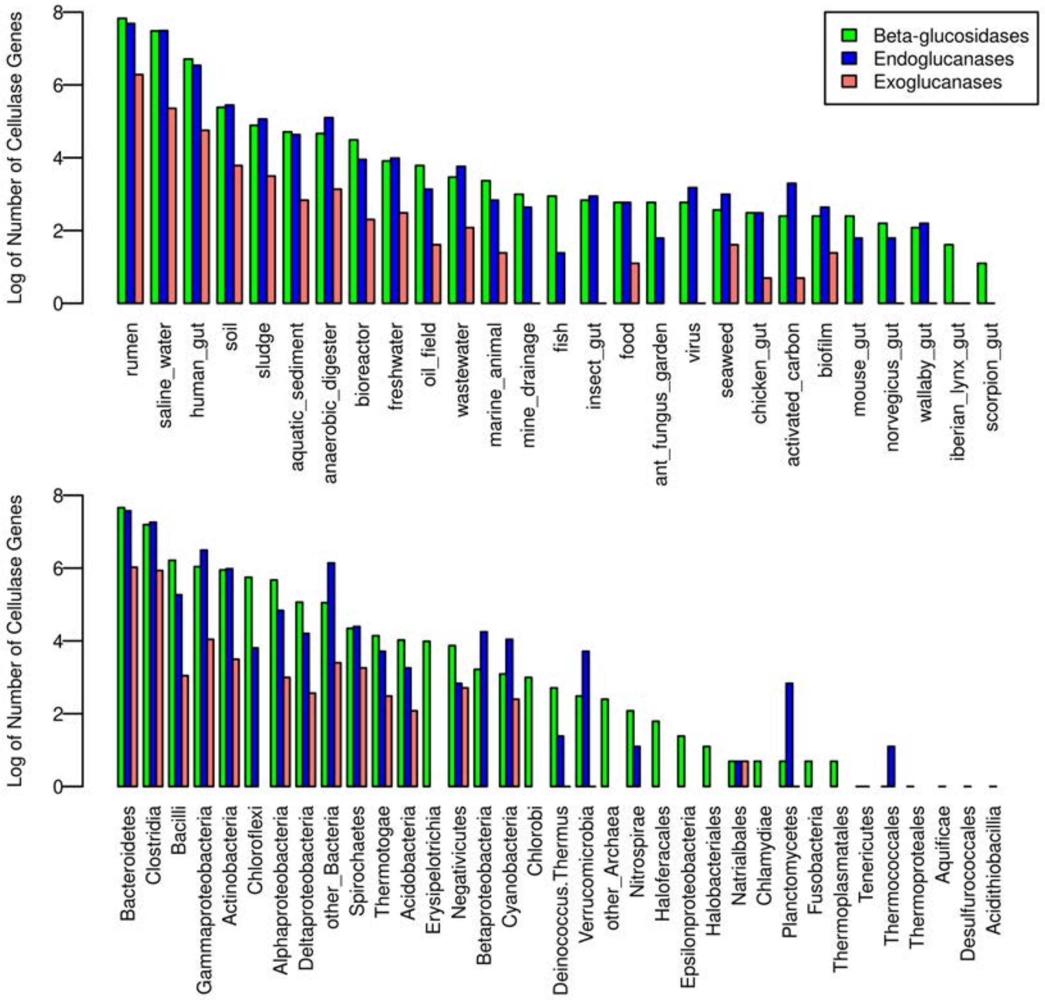


Fig. 2. Ecological and functional overview of the cellulases found within 1,735 metagenomes. The three main types of cellulases (beta-glucosidases, endoglucanases, and exoglucanases) were assigned to the genomes of the 35 prokaryotic groups across 27 environments.

Table 1 Functional overview of the cellulases found within 1,735 metagenomes.

Catalytic function	Number of cellulases	Number of the protein clusters
Beta-glucosidases	6,110	19
Endoglucanases	5,685	97
Exoglucanases	1,042	20
In total	12,837	136

Found protein cluster sizes to which the cellulases belong have a wide range, from only two sequences to up to 2,399 sequences (Supplement 6). All three cellulase types may be found in the same metagenomes, and we have observed roughly one-third of metagenomes (in 672 of them) include three types of cellulases here. In 491 metagenomes, we have observed two cellulase types, usually the combination of endoglucanases and beta-glucosidases. In comparison, in 572 metagenomes, only a single kind of cellulase was found, most commonly beta-glucosidase type (Supplement 7). Beta-glucosidases are found more often together with the endoglucanases than with the exoglucanases (Supplement 7). Exoglucanases are in our results, the least abundant and the least diverse functional type of cellulase (Table 1).

Some prokaryotes break cellulose better than others. Sequence comparison against the proteome of 5,655 prokaryotic genomes was performed to evaluate the taxonomical distribution of the metagenome-derived cellulases and determine the relative contribution of the distinct prokaryotic lineages to cellulose degradation. The highest diversity in the cellulases was found among Bacterioidetes (4,480), Clostridia (3,129), Gammaproteobacteria (1138), Actinobacteria (812), and Bacilli (715) (Supplement 8). Among the tribes mentioned above, we have identified the highest number of sequences and the highest number of clusters of all the types of cellulases in all the environments. The other prokaryotic phyla exhibit much smaller diversity. For example, in Thermococcales, only a few endoglucanases are reported from the saline water. In Chlorobi and Deinococcus-Thermus group, only some taxa exhibit beta-glucosidases, and only in saline water (Fig. 3). In Archaea, only a few phyla show some of the cellulases, which is most probably a result of the recent LGT (Fig. 3).

There is no significant difference between the abundance and composition of the cellulases among the four cellulose-rich environments (soil, human gut, saline water, rumen) (Supplement 5). The four studied cellulose-rich environments are abundant in the representatives of the five phyla mentioned above that exhibit various cellulases.

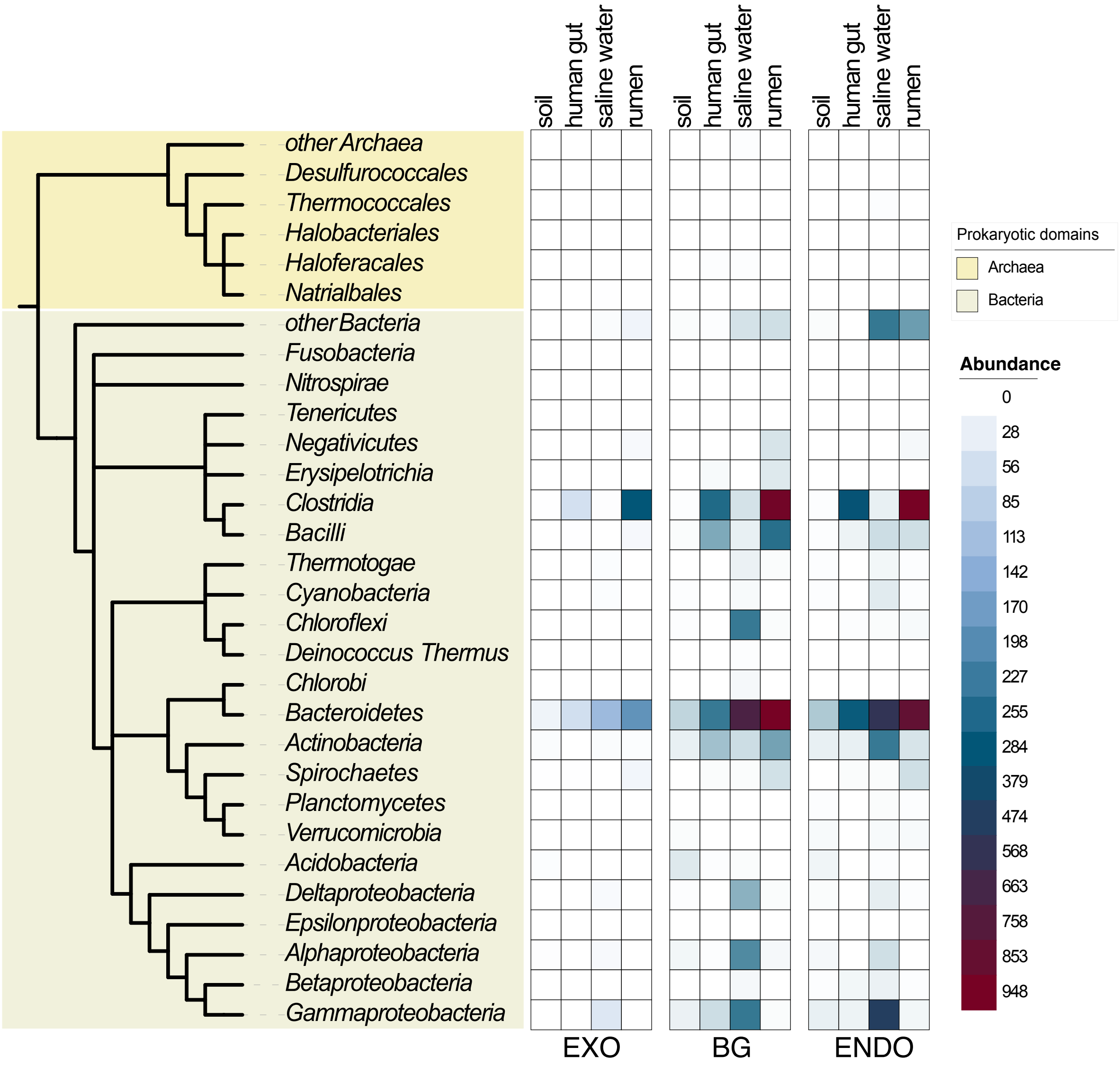


Fig. 3. Environmental and taxonomic distribution of the cellulases. Presence of the three main types of cellulases (EXO – exoglucanases, BG – beta-glucosidases, ENDO – endoglucanases) in the genomes of the main prokaryotic groups, assigned to four environments (soil, human gut, saline water, rumen). The reference tree of the prokaryotes follows the Interactive Tree of Life (Letunic and Bork 2019).

Discussion

Diverse microbial communities have been studied by metagenome sequencing, such as human gut or rumen (Karlsson et al. 2013, Hess et al. 2011), making it possible to evaluate their cellulases' diversity. Metagenomic techniques have certain advantages over the traditional genome sequences (Madhavan et al. 2017), as genome sequencing requires prior cultivation of the axenic (= single species) culture. Existing metagenome data may provide preliminary insights into the unknown diversity of certain enzymes from the uncultivated genomes (Stewart et al. 2018; Nayfach et al. 2019), as we show here for the cellulases. We have recorded endoglucanases, exoglucanases, and β-d-glucosidases mainly in the anaerobic thermophilic cellulolytic environment, such as rumen, saline water, and the human gut. These environments are similar to the previously reported cellulose-rich bioreactors (Sukharnikov et al. 2012). Furthermore, similar cellulases (especially beta-glucosidases) seem present across different environments (Berlemont and Martiny 2016).

No matter which cellulose-rich-environment a metagenome is sampled from, it seems that it will always include Bacterioidetes and Clostridia as the most abundant phyla. These two phyla exhibit the highest genetic diversity from the standpoint of view of the sequences coding for the three main types of cellulases. The two phyla were already reported as one of the most essential cellulose degrading groups (Mai et al. 2019); however, they are not the only groups in which the genes (sequences) were found. Genes for the cellulases, primarily for beta-glucosidases, are widespread between bacterial and archaeal phyla.

From the distribution of the sequences (Fig. 3), it is clear that cellulases may be ancestral to many taxa within Bacterioidetes, Clostridia, and maybe to some groups of Actinobacteria, Gammaproteobacteria, and Bacillia, while all the other groups gained the trait independently (see the scattered distribution of the cellulases in, e.g., Spirochetes and Thermotogae in Fig. 3). Numerous taxa may gain cellulases by living in the cellulose-rich environment together with the prokaryotes which have the cellulase genes. Previous studies on the potential cellulose degraders have yielded similar results on the taxonomic distribution – dominant being Actinobacteria, Bacterioidetes, Firmicutes (including Clostridia and Bacilli), and Proteobacteria (including Deltaproteobacteriaa) (Berlemont and Martiny 2013, Mai et al. 2019). Cooperation in the prokaryotic cellulose-rich communities has already been recognized previously (Choi et al. 2018). Some prokaryotic phyla seem more prone to achieving those genes, for example, Bacilli and Gammaproteobacteria, than others – for instance, Archaea, in which only a few sequences for the cellulose genes were identified.

The abundance of cellulose types and the complexity of the plant cell walls leads to the heterogeneity of cellulases (Wilson et al. 1999), so multiple genes encoding similar function, contributing to high enzymatic concentrations needed for cellulose degradation, may be observed in all the samples of the cellulolytic environments, albeit with variable abundances.

Beta-glucosidases are already, based on whole-genome sequencing, shown to be the most abundant cellulase type in potential opportunistic strains (Berlemont and Martiny 2013). Taking into account that they represent the commonest cellulases (we also report 6,110 out of 12,837 sequences), but at the same time, taking into account how low their diversity is (only 19 out of 136 clusters), we can conclude that beta-glucosidases represent the most transferrable cellulases in nature and that their few types are more often laterally transferred than is the case with endoglucanases, and especially exoglucanases, of which only a few types exist, and those few types are taxonomically concentrated among Bacterioidetes and Clostridia. This is one more reason to suspect that these two bacterial phyla have evolved cellulases independently, while their presence in the other groups represents laterally transferred genes. The cooperation and the lateral gene transfer, necessary for cellulose degradation, may be explained by the black queen hypothesis. Many lineages seem to have gained beta-glucosidases and endoglucanases and are secreting those enzymes in the environment. However, the exoglucanases, the third component, either have a high tendency to be lost (Bai et al. 2017) or are rarely laterally transferred.

Acknowledgments

The authors acknowledge helpful contributions to this manuscript from Falk Nagies, Michael Knopp, Nils Kapust, Sriram G. Garg, Joana C. Xavier, Julia Brückner, and Jessica Wimmer. The stipend of Bing Song was funded by the China Scholarship Council (Bing Song's application number: 201706300100).

References

Atalla RH (1990) The structures of cellulose. Materials Research Society Symposium 197:89–98.

Bai X, Wang X, Wang S, et al (2017) Functional studies of β-glucosidases of Cytophaga hutchinsonii and their effects on cellulose degradation. Front Microbiol 8:140.

Bateman A (2019) UniProt: A worldwide hub of protein knowledge. Nucleic Acids Research 47:506–515.

Berghem LER, Pettersson LG, Axiö-Fredriksson U-B (1976) The mechanism of enzymatic cellulose degradation: purification and some properties of two different 1, 4-β-glucan glucanohydrolases from Trichoderma viride. European Journal of Biochemistry 61:621–630.

Berlemont R, Martiny AC (2016) Glycoside hydrolases across environmental microbial communities. PLoS Computational Biology 12:e1005300.

Berlemont R, Martiny AC (2013) Phylogenetic distribution of potential cellulases in bacteria. Applied and Environmental Microbiology 79:1545–1554.

Buchfink B, Xie C, Huson DH (2014) Fast and sensitive protein alignment using DIAMOND. Nature Methods 12:59–60.

Choi J, Bach E, Lee J, et al (2018) Spatial structuring of cellulase gene abundance and activity in soil. Frontiers in Environmental Science 6:107.

Dongen S Van (2000) A cluster algorithm for graphs. Report Information Systems 10:X.

Enari TM, Niku-Paavola ML (1987) Enzymatic hydrolysis of cellulose: is the current theory of the mechanisms of hydrolysis valid? Critical Reviews in Biotechnology 5:67–87.

Gilbert HJ (2010) The biochemistry and structural biology of plant cell wall deconstruction. Plant Physiology 53:444–455.

Halliwell G, Griffin M (1973) The nature and mode of action of the cellulolytic component C1 of Trichoderma koningii on native cellulose. Biochemical Journal 135:587–594.

Henrissat B, Driguez H, Viet C, Schülein M (1985) Synergism of cellulases from trichoderma reesei in the degradation of cellulose. Bio/Technology 3:722–726.

Hess M, Sczyrba A, Egan R, et al (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science 331:463–467.

Hon DNS (1994) Cellulose: a random walk along its historical path. Cellulose 1:1–25.

Karlsson FH, Tremaroli V, Nookaew I, et al (2013) Gut metagenome in european women with normal, impaired and diabetic glucose control. Nature 498:99–103.

Kikani B (2018) Production and characterization of bacterial thermostable cellulase.

Knowles J, Lehtovaara P, Teeri T (1987) Cellulase families and their genes. Trends in Biotechnology 5:255–261.

Letunic I, Bork P (2019) Interactive tree of life (iTOL) v4: recent updates and new developments. Nucleic Acids Research.

Madhavan A, Sindhu R, Parameswaran B, et al (2017) Metagenome analysis: a powerful tool for enzyme bioprospecting. Applied Biochemistry and Biotechnology 183:636–651.

Mai G, Liu Q, Zhou Y, et al (2019) Metagenomic insights into a cellulose-rich niche reveal microbial cooperation in cellulose degradation. Frontiers in microbiology 10:618.

McGinnis S, Madden TL (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Research 32:20–25.

Metsalu T, Vilo J (2015) ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. Nucleic Acids Research.

Nayfach S, Shi ZJ, Seshadri R, et al (2019) New insights from uncultivated genomes of the global human gut microbiome. Nature 568:505.

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology 48:443–453.

Niku-Paavola ML, Lappalainen A, Enari TM, Nummi M (1985) A new appraisal of the endoglucanases of the fungus Trichoderma reesei. Biochemical Journal 231:75–81.

Pruitt KD, Tatusova T, Brown GR, Maglott DR (2011) NCBI reference sequences (RefSeq): current status, new features and genome annotation policy. Nucleic acids research 40:130–135.

Rice P, Longden L, Bleasby A (2000) EMBOSS: The european molecular biology open software suite. Trends in Genetics 16:276–277.

RStudio Team (2021) RStudio: Integrated Development for R. RStudio, Inc., Boston, MA.

Sharma A, Tewari R, Rana SS, et al (2016) Cellulases: classification, methods of determination and industrial applications. Applied Biochemistry and Biotechnology 179:1346–1380.

Stewart RD, Auffret MD, Warr A, et al (2018) Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. Nature Communications 9:870.

Sukharnikov LO, Alahuhta M, Brunecky R, et al (2012) Sequence, structure, and evolution of cellulases in glycoside hydrolase family 48. Journal of Biological Chemistry 287:41068–41077.

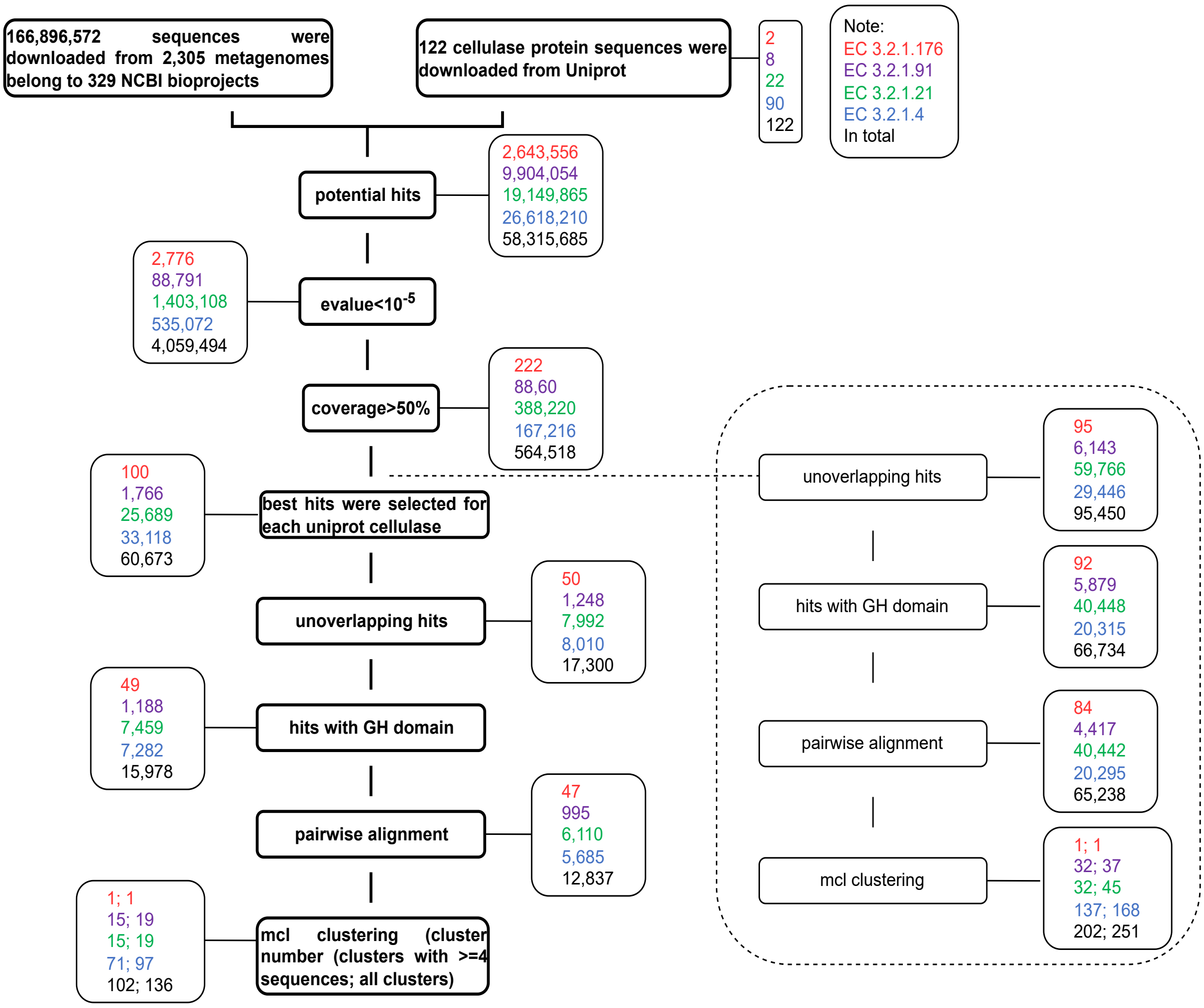
Team RC (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Wheeler TJ, Eddy SR (2013) Nhmmer: dna homology search with profile HMMs. Bioinformatics 29:2487–2489.

Wilson DB (2009) Evidence for a novel mechanism of microbial cellulose degradation. Cellulose 16:723–727.

Supplementary_File_1 The study workflow.

The main pipeline is to search and identify the high potential cellulase sequences by e-value, protein coverage, bitscore, alignment region, catalytic domain, and clustering. The partial pipeline within the dotted box is supplementary pipeline without selecting the best hits for each reference Uniprot cellulase gene.



Supplementary_File_2 The 122 reference uniprot cellulase protein sequences In the supplementary table 2, each row contains the uniprot id, protein name, and the lineage name of each downloaded uniprot reference cellulase sequence.

<table>
<tr><th>>uniprot-id|protein names|lineage(Phylum)</th></tr>
<tr><td>>P0C2S5|Cellulose 1,4-beta-cellobiosidase (reducing end) CelS (EC 3.2.1.176)|Firmicutes</td></tr>
<tr><td>>A3DH67|Cellulose 1,4-beta-cellobiosidase (reducing end) CelS (EC 3.2.1.176)|Firmicutes</td></tr>
<tr><td>>P0C2S1|Cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91)|Firmicutes</td></tr>
<tr><td>>P50899|Exoglucanase B (EC 3.2.1.91)|Actinobacteria</td></tr>
<tr><td>>P50900|Exoglucanase-2 (EC 3.2.1.91)|Firmicutes</td></tr>
<tr><td>>P07986|Exoglucanase/xylanase Includes: Exoglucanase (EC 3.2.1.91);(EC 3.2.1.8)|Actinobacteria</td></tr>
<tr><td>>P10474|Endoglucanase/exoglucanase B Includes: (EC 3.2.1.4);(EC 3.2.1.91)|Firmicutes</td></tr>
<tr><td>>P50401|Exoglucanase A (EC 3.2.1.91)|Actinobacteria</td></tr>
<tr><td>>A3DCH1|Cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91)|Firmicutes</td></tr>
<tr><td>>P38535|Exoglucanase XynX (EC 3.2.1.91)|Firmicutes</td></tr>
<tr><td>>P14002|Thermostable beta-glucosidase B (EC 3.2.1.21)|Firmicutes</td></tr>
<tr><td>>A7LXS8|Beta-glucosidase BoGH3A (EC 3.2.1.21)|Bacteroidetes</td></tr>
<tr><td>>P26208|Beta-glucosidase A (EC 3.2.1.21)|Firmicutes</td></tr>
<tr><td>>Q59437|Beta-glucosidase A (EC 3.2.1.21)|Proteobacteria</td></tr>
<tr><td>>P94248|Bifunctional beta-D-glucosidase/beta-D-fucosidase Includes: (EC 3.2.1.21);(EC 3.2.1.38)|Actinobacteria</td></tr>
<tr><td>>P22073|Beta-glucosidase A (BGA) (EC 3.2.1.21)|Firmicutes</td></tr>
<tr><td>>P10482|Beta-glucosidase A (EC 3.2.1.21) |Firmicutes</td></tr>
<tr><td>>P15885|Beta-glucosidase (EC 3.2.1.21)|Firmicutes</td></tr>
<tr><td>>P22505|Beta-glucosidase B (EC 3.2.1.21)|Firmicutes</td></tr>
<tr><td>>P27034|Beta-glucosidase (EC 3.2.1.21)|Proteobacteria</td></tr>
<tr><td>>B9K7M5|1,4-beta-D-glucan glucohydrolase Includes: (EC 3.2.1.74);(EC 3.2.1.21)|Thermotogae</td></tr>
<tr><td>>P16084|Beta-glucosidase A (EC 3.2.1.21)|Firmicutes</td></tr>
<tr><td>>Q46684|Periplasmic beta-glucosidase/beta-xylosidase Includes: (EC 3.2.1.21);(EC 3.2.1.37)|Proteobacteria</td></tr>
<tr><td>>Q7WUL3|Beta-N-acetylglucosaminidase/beta-glucosidase (EC 3.2.1.21);(EC 3.2.1.52)|Actinobacteria</td></tr>
<tr><td>>P0C946|1,4-beta-D-glucan glucohydrolase Includes: (EC 3.2.1.74);(EC 3.2.1.21)|Thermotogae</td></tr>
<tr><td>>Q03506|Beta-glucosidase (EC 3.2.1.21)|Firmicutes</td></tr>
<tr><td>>A7LXU3|Beta-glucosidase BoGH3B (EC 3.2.1.21)|Bacteroidetes</td></tr>
<tr><td>>P12614|Beta-glucosidase (EC 3.2.1.21)|Proteobacteria</td></tr>
<tr><td>>P33363|Periplasmic beta-glucosidase (EC 3.2.1.21)|Proteobacteria</td></tr>
<tr><td>>P38645|Thermostable beta-glucosidase B (EC 3.2.1.21)|Actinobacteria</td></tr>
<tr><td>>Q08638|Beta-glucosidase A (EC 3.2.1.21)|Thermotogae</td></tr>
<tr><td>>Q56078|Periplasmic beta-glucosidase (EC 3.2.1.21)|Proteobacteria</td></tr>
<tr><td>>A0A0U4EBH5|Cellulase CelDZ1 (EC 3.2.1.4)|Firmicutes</td></tr>
<tr><td>>P16216|Endoglucanase 1 (EC 3.2.1.4)|Firmicutes</td></tr>
<tr><td>>P17877|Endoglucanase 1 (EC 3.2.1.4)|Firmicutes</td></tr>
<tr><td>>P16218|Endoglucanase H (EC 3.2.1.4)|Firmicutes</td></tr>
<tr><td>>P26222|Endoglucanase E-2 (EC 3.2.1.4)|Actinobacteria</td></tr>
<tr><td>>A3DC29|Endoglucanase A (EGA) (EC 3.2.1.4)|Firmicutes</td></tr>
<tr><td>>O08342|Endoglucanase A (EC 3.2.1.4)|Firmicutes</td></tr>
<tr><td>>P37651|Endoglucanase (EC 3.2.1.4)|Proteobacteria</td></tr>
<tr><td>>Q5YLG1|Endoglucanase A (EC 3.2.1.4)|Firmicutes</td></tr>
<tr><td>>P19570|Endoglucanase C (EC 3.2.1.4)|Firmicutes</td></tr>
<tr><td>>P28621|Endoglucanase B Includes: (EC 3.2.1.4);(EC 3.2.1.8)|Firmicutes</td></tr>
<tr><td>>P29019|Endoglucanase (EC 3.2.1.4)|Firmicutes</td></tr>
<tr><td>>Q8RSY9|Endoglucanase (EC 3.2.1.4)|Proteobacteria</td></tr>
<tr><td>>P07103|Endoglucanase Z (EC 3.2.1.4)|Proteobacteria</td></tr>
<tr><td>>P54937|Endoglucanase A (EC 3.2.1.4)|Firmicutes</td></tr>
<tr><td>>P04956|Endoglucanase B (EC 3.2.1.4)|Firmicutes</td></tr>
<tr><td>>P0C2S2|Endoglucanase A (EGA) (EC 3.2.1.4)|Firmicutes</td></tr>
<tr><td>>P17974|Endoglucanase (EC 3.2.1.4)|Proteobacteria</td></tr>
</table>

>P54583|Endoglucanase E1 (EC 3.2.1.4)|Actinobacteria
>P18336|Endoglucanase (EC 3.2.1.4)|Actinobacteria
>P23661|Endoglucanase B (EC 3.2.1.4)|Firmicutes
>Q05156|Cellulase 1 (EC 3.2.1.4)|Actinobacteria
>P37700|Endoglucanase G (EC 3.2.1.4)|Firmicutes
>P28623|Endoglucanase D (EC 3.2.1.4)|Firmicutes
>P06564|Endoglucanase (EC 3.2.1.4)|Firmicutes
>P0C2S3|Endoglucanase C (EC 3.2.1.4)|Firmicutes
>P26224|Endoglucanase F (EGF) (EC 3.2.1.4)|Firmicutes
>P27032|Minor endoglucanase Y (EC 3.2.1.4)|Proteobacteria
>P0C2S4|Endoglucanase D (EGD) (EC 3.2.1.4)|Firmicutes
>Q05622|Endoglucanase E (EGE) (EC 3.2.1.4)|Firmicutes
>A3DJ77|Endoglucanase C (EC 3.2.1.4)|Firmicutes
>P58935|Endoglucanase (EC 3.2.1.4)|Proteobacteria
>P20847|Endoglucanase 1 (EC 3.2.1.4)|Firmicutes
>P22533|Beta-mannanase/endoglucanase A Includes: (EC 3.2.1.78);(EC 3.2.1.4)|Firmicutes
>P22541|Endoglucanase A (EC 3.2.1.4)|Firmicutes
>P15329|Putative endoglucanase X (EGX) (EC 3.2.1.4)|Firmicutes
>P13933|Endoglucanase 1 (EC 3.2.1.4)|Actinobacteria
>P06565|Endoglucanase B (EC 3.2.1.4)|Firmicutes
>P27035|Endoglucanase CelA (EC 3.2.1.4)|Actinobacteria
>Q02934|Endoglucanase 1 (EC 3.2.1.4)|Firmicutes
>Q59394|Endoglucanase N (EC 3.2.1.4)|Proteobacteria
>Q59395|Endoglucanase 6 (EC 3.2.1.4)|Proteobacteria
>P15704|Endoglucanase (EC 3.2.1.4)|Firmicutes
>P50400|Endoglucanase D (EC 3.2.1.4)|Actinobacteria
>P25472|Endoglucanase D (EC 3.2.1.4)|Firmicutes
>P23658|Cellodextrinase (EC 3.2.1.4)|Firmicutes
>P37701|Endoglucanase 2 (EC 3.2.1.4)|Firmicutes
>P28622|Endoglucanase 4 (EC 3.2.1.4)|Firmicutes
>P26414|Endoglucanase A (EC 3.2.1.4)|Actinobacteria
>P10474|Endoglucanase/exoglucanase B Includes: (EC 3.2.1.4);(EC 3.2.1.91)|Firmicutes
>P37699|Endoglucanase C (EC 3.2.1.4)|Firmicutes
>A3DDN1|Endoglucanase D (EGD) (EC 3.2.1.4)|Firmicutes
>P23549|Endoglucanase (EC 3.2.1.4)|Firmicutes
>P23659|Endoglucanase Z (EC 3.2.1.4)|Firmicutes
>P58599|Endoglucanase (EC 3.2.1.4)|Proteobacteria
>P17901|Endoglucanase A (EC 3.2.1.4)|Firmicutes
>P26225|Endoglucanase B (EC 3.2.1.4)|Actinobacteria
>P14090|Endoglucanase C (EC 3.2.1.4)|Actinobacteria
>P33682|Endoglucanase 1 (EC 3.2.1.4)|Actinobacteria
>Q01786|Endoglucanase E-5 (EC 3.2.1.4)|Actinobacteria
>P07984|Endoglucanase A (EC 3.2.1.4)|Actinobacteria
>P19424|Endoglucanase (EC 3.2.1.4)|Firmicutes
>P40942|Thermostable celloxylanase Includes: (EC 3.2.1.4);(EC 3.2.1.8)|Firmicutes
>P22534|Endoglucanase A (EC 3.2.1.4)|Firmicutes
>P16630|Endoglucanase S (EC 3.2.1.4)|Proteobacteria
>P19487|Major extracellular endoglucanase (EC 3.2.1.4)|Proteobacteria
>P37696|Probable endoglucanase (EC 3.2.1.4)|Proteobacteria
>P18126|Endoglucanase B (EGB) (EC 3.2.1.4)|Proteobacteria
>Q8ZLB7|Endoglucanase (EC 3.2.1.4)|Proteobacteria
>O85465|Endoglucanase 5A (EC 3.2.1.4)|Firmicutes
>P27033|Endoglucanase C (EC 3.2.1.4)|Proteobacteria
>P14250|Endoglucanase 3 (EC 3.2.1.4)|Fibrobacteres
>P23665|Endoglucanase A (EC 3.2.1.4)|Fibrobacteres
>Q07940|Endoglucanase 4 (EC 3.2.1.4)|Firmicutes
>Q47096|Endoglucanase 5 (EC 3.2.1.4)|Proteobacteria
>Q05332|Endoglucanase G (EC 3.2.1.4)|Firmicutes
>P23340|Endoglucanase C307 (EC 3.2.1.4)|Firmicutes
>P10475|Endoglucanase (EC 3.2.1.4)|Firmicutes

>P37698|Endoglucanase F (EC 3.2.1.4)|Firmicutes
>P06566|Endoglucanase A (EC 3.2.1.4)|Firmicutes
>P07983|Endoglucanase (EC 3.2.1.4)|Firmicutes
>Q8X5L9|Endoglucanase (EC 3.2.1.4)|Proteobacteria
>P10476|Endoglucanase A (EGA) (EC 3.2.1.4)|Proteobacteria
>P23550|Endoglucanase B (EC 3.2.1.4)|Firmicutes
>P29719|Endoglucanase A (EC 3.2.1.4)|Firmicutes
>P23660|Endoglucanase A Includes: (EC 3.2.1.4);(EC 3.2.1.8)|Firmicutes
>P26221|Endoglucanase E-4 (EC 3.2.1.4)|Actinobacteria
>Q8Z289|Endoglucanase (EC 3.2.1.4)|Proteobacteria
>P23548|Endoglucanase (EC 3.2.1.4)|Firmicutes
>P10477|Cellulase/esterase CelE (CtCel5C-CE2) Includes: (EC 3.2.1.4);(EC 3.1.1.72)|Firmicutes

Supplementary_File_3 Downloaded 2305 metagenome ids (GCA_id) and their corresponding project ids (Project_id).

Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id
PRJNA18313	GCA_000208225.1	PRJEB15554	GCA_900118295.1	PRJNA269163	GCA_002011535.1	PRJNA269163	GCA_002010755.1	PRJEB7226	GCA_000944795.1	PRJNA289734	GCA_001625935.1
PRJNA17621	GCA_000208245.1	PRJEB15554	GCA_900118395.1	PRJNA269163	GCA_002011715.1	PRJNA269163	GCA_002010875.1	PRJEB7226	GCA_000937115.1	PRJNA289734	GCA_001625855.1
PRJNA17393	GCA_000205045.1	PRJEB15554	GCA_900118365.1	PRJNA269163	GCA_002010545.1	PRJNA269163	GCA_002011415.1	PRJEB7226	GCA_000944735.1	PRJNA289734	GCA_001625775.1
PRJNA17399	GCA_000205105.1	PRJEB15554	GCA_900118385.1	PRJNA269163	GCA_002010685.1	PRJNA269163	GCA_002010205.1	PRJNA29605	GCA_000206585.1	PRJNA289734	GCA_001629005.1
PRJNA17391	GCA_000205005.1	PRJNA266758	GCA_000986745.1	PRJNA269163	GCA_002010125.1	PRJNA269163	GCA_002011565.1	PRJNA29619	GCA_000206485.1	PRJNA289734	GCA_001629035.1
PRJNA17397	GCA_000205085.1	PRJEB11362	GCA_900066005.1	PRJNA269163	GCA_002010555.1	PRJNA269163	GCA_002010915.1	PRJNA29625	GCA_000206425.1	PRJNA289734	GCA_001628845.1
PRJNA17385	GCA_000205125.1	PRJEB11362	GCA_900065945.1	PRJNA269163	GCA_002011475.1	PRJNA269163	GCA_002010975.1	PRJNA29621	GCA_000206465.1	PRJNA289734	GCA_001625975.1
PRJNA17395	GCA_000205065.1	PRJEB11362	GCA_900065985.1	PRJNA269163	GCA_002010655.1	PRJNA269163	GCA_002011495.1	PRJNA167987	GCA_000342605.1	PRJNA289734	GCA_001629075.1
PRJNA42359	GCA_000205705.1	PRJEB11362	GCA_900065955.1	PRJNA269163	GCA_002011655.1	PRJNA269163	GCA_002010035.1	PRJNA29615	GCA_000206525.1	PRJNA289734	GCA_001626055.1
PRJEB14154	GCA_900088525.1	PRJEB11362	GCA_900065965.1	PRJNA269163	GCA_002010725.1	PRJNA269163	GCA_002010365.1	PRJNA29613	GCA_000206545.1	PRJNA289734	GCA_001628945.1
PRJEB14154	GCA_900088585.1	PRJEB11362	GCA_900065975.1	PRJNA269163	GCA_002010635.1	PRJNA269163	GCA_002010115.1	PRJNA72255	GCA_001188585.1	PRJNA289734	GCA_001626015.1
PRJNA318959	GCA_001753025.1	PRJEB11362	GCA_900065995.1	PRJNA269163	GCA_002010215.1	PRJNA269163	GCA_002010285.1	PRJNA28343	GCA_000207785.1	PRJNA289734	GCA_001628635.1
PRJNA318959	GCA_001752985.1	PRJNA37895	GCA_000206005.1	PRJNA269163	GCA_002011075.1	PRJNA269163	GCA_002010475.1	PRJNA17769	GCA_000207825.1	PRJNA289734	GCA_001628915.1
PRJNA318959	GCA_001753005.1	PRJNA17241	GCA_000206165.1	PRJNA269163	GCA_002010455.1	PRJNA269163	GCA_002010885.1	PRJNA13694	GCA_000204965.1	PRJNA289734	GCA_001625915.1
PRJNA318959	GCA_001753035.1	PRJNA269163	GCA_002011555.1	PRJNA269163	GCA_002011155.1	PRJNA269163	GCA_002011435.1	PRJNA326480	GCA_001757525.1	PRJNA289734	GCA_001625955.1
PRJNA320927	GCA_002163315.1	PRJNA269163	GCA_002010995.1	PRJNA269163	GCA_002011055.1	PRJNA269163	GCA_002010715.1	PRJNA326480	GCA_001757065.1	PRJNA289734	GCA_001625895.1
PRJNA320927	GCA_002163335.1	PRJNA269163	GCA_002011215.1	PRJNA269163	GCA_002010435.1	PRJNA269163	GCA_002011085.1	PRJNA326480	GCA_001757025.1	PRJNA289734	GCA_001625785.1
PRJNA320927	GCA_002163355.1	PRJNA269163	GCA_002010355.1	PRJNA269163	GCA_002011255.1	PRJNA269163	GCA_002010445.1	PRJNA326480	GCA_001757045.1	PRJNA289734	GCA_001628995.1
PRJNA314399	GCA_001623305.1	PRJNA269163	GCA_002011165.1	PRJNA269163	GCA_002011485.1	PRJNA269163	GCA_002011745.1	PRJNA326480	GCA_001757005.1	PRJNA289734	GCA_001628655.1
PRJEB15554	GCA_900118245.1	PRJNA269163	GCA_002011795.1	PRJNA269163	GCA_002009955.1	PRJNA269163	GCA_002010515.1	PRJNA326480	GCA_001756945.1	PRJNA289734	GCA_001625995.1
PRJEB15554	GCA_900118405.1	PRJNA269163	GCA_002010805.1	PRJNA269163	GCA_002011355.1	PRJNA269163	GCA_002010925.1	PRJNA326480	GCA_001757125.1	PRJNA289734	GCA_001629555.1
PRJEB15554	GCA_900118375.1	PRJNA269163	GCA_002011685.1	PRJNA269163	GCA_002010385.1	PRJNA269163	GCA_002011115.1	PRJNA326480	GCA_001757145.1	PRJNA289734	GCA_001626035.1
PRJEB15554	GCA_900118425.1	PRJNA269163	GCA_002011335.1	PRJNA269163	GCA_002010275.1	PRJNA269163	GCA_002011775.1	PRJNA326480	GCA_001756965.1	PRJNA289734	GCA_001629095.1
PRJEB15554	GCA_900118335.1	PRJNA269163	GCA_002011315.1	PRJNA269163	GCA_002011395.1	PRJNA269163	GCA_002011235.1	PRJNA326480	GCA_001756385.1	PRJNA289734	GCA_001629575.1
PRJEB15554	GCA_900118325.1	PRJNA269163	GCA_002011425.1	PRJNA269163	GCA_002010195.1	PRJNA269163	GCA_002011615.1	PRJNA326480	GCA_001756985.1	PRJNA289734	GCA_001628925.1
PRJEB15554	GCA_900118355.1	PRJNA269163	GCA_002010775.1	PRJNA269163	GCA_002011005.1	PRJNA269163	GCA_002010165.1	PRJNA326480	GCA_001757085.1	PRJNA289734	GCA_001629525.1
PRJEB15554	GCA_900118305.1	PRJNA269163	GCA_002011735.1	PRJNA269163	GCA_002010295.1	PRJEB19060	GCA_900157355.1	PRJNA16339	GCA_000208205.1	PRJNA289734	GCA_001628835.1
PRJEB15554	GCA_900118165.1	PRJNA269163	GCA_002010615.1	PRJNA269163	GCA_002010075.1	PRJNA53357	GCA_000347535.1	PRJNA28345	GCA_000207765.1	PRJNA289734	GCA_001629625.1
PRJEB15554	GCA_900118415.1	PRJNA269163	GCA_002010535.1	PRJNA269163	GCA_002009965.1	PRJNA81851	GCA_000829845.1	PRJNA17767	GCA_000207845.1	PRJNA289734	GCA_001629045.1
PRJEB15554	GCA_900118285.1	PRJNA269163	GCA_002011365.1	PRJNA269163	GCA_002010045.1	PRJNA67915	GCA_001684355.1	PRJNA17771	GCA_000207805.1	PRJNA289734	GCA_001625795.1
PRJEB15554	GCA_900118315.1	PRJNA269163	GCA_002010665.1	PRJNA269163	GCA_002011845.1	PRJNA39533	GCA_000205945.1	PRJEB20421	GCA_900186075.1	PRJNA289734	GCA_001628855.1
PRJEB15554	GCA_900118345.1	PRJNA269163	GCA_002010795.1	PRJNA269163	GCA_002011895.1	PRJEB6121	GCA_000821205.1	PRJNA50327	GCA_000205685.2	PRJNA289734	GCA_001628895.1
PRJEB15554	GCA_900118265.1	PRJNA269163	GCA_002010815.1	PRJNA269163	GCA_002010225.1	PRJNA114691	GCA_000299275.1	PRJNA17765	GCA_000207865.1	PRJNA289734	GCA_001629665.1
PRJEB15554	GCA_900117435.1	PRJNA269163	GCA_002009985.1	PRJNA269163	GCA_002011835.1	PRJNA218796	GCA_000698905.1	PRJNA77837	GCA_000283355.1	PRJNA289734	GCA_001629695.1
PRJEB15554	GCA_900118235.1	PRJNA269163	GCA_002010065.1	PRJNA269163	GCA_002010305.1	PRJEB10139	GCA_001458795.1	PRJNA335308	GCA_001875485.1	PRJNA289734	GCA_001629615.1
PRJEB15554	GCA_900118175.1	PRJNA269163	GCA_002011825.1	PRJNA269163	GCA_002011195.1	PRJEB10139	GCA_001458635.1	PRJEB11812	GCA_900067685.1	PRJNA289734	GCA_001629115.1
PRJEB15554	GCA_900118275.1	PRJNA269163	GCA_002011125.1	PRJNA269163	GCA_002010155.1	PRJEB10139	GCA_001458615.1	PRJEB11812	GCA_900067675.1	PRJNA289734	GCA_001628955.1
PRJEB15554	GCA_900118225.1	PRJNA269163	GCA_002011035.1	PRJNA269163	GCA_002011635.1	PRJNA29627	GCA_000206405.1	PRJEB11812	GCA_900067665.1	PRJNA289734	GCA_001628555.1
PRJEB15554	GCA_900118185.1	PRJNA269163	GCA_002010955.1	PRJNA269163	GCA_002010595.1	PRJNA29623	GCA_000206445.1	PRJNA289734	GCA_001629595.1	PRJNA289734	GCA_001629655.1
PRJEB15554	GCA_900118195.1	PRJNA269163	GCA_002011295.1	PRJNA269163	GCA_002010845.1	PRJNA29611	GCA_000206565.1	PRJNA289734	GCA_001625805.1	PRJNA289734	GCA_001626065.1
PRJEB15554	GCA_900118255.1	PRJNA269163	GCA_002011815.1	PRJNA269163	GCA_002009975.1	PRJNA214436	GCA_000495655.1	PRJNA289734	GCA_001625865.1	PRJNA30445	GCA_000206025.1
PRJEB15554	GCA_900118205.1	PRJNA269163	GCA_002011245.1	PRJNA269163	GCA_002011575.1	PRJNA29617	GCA_000206505.1	PRJNA289734	GCA_001629135.1	PRJEB7988	GCA_001040365.1
PRJEB15554	GCA_900118215.1	PRJNA269163	GCA_002010375.1	PRJNA269163	GCA_002011675.1	PRJEB7226	GCA_000945655.1	PRJNA289734	GCA_001628675.1	PRJEB7988	GCA_001369935.1

Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id
PRJEB7988	GCA_001371455.1	PRJEB7988	GCA_001040625.1	PRJEB7988	GCA_001028505.1	PRJEB7988	GCA_001369455.1	PRJEB7988	GCA_001371735.1	PRJEB7988	GCA_001371515.1
PRJEB7988	GCA_001040125.1	PRJEB7988	GCA_001371215.1	PRJEB7988	GCA_001371595.1	PRJEB7988	GCA_001371395.1	PRJEB7988	GCA_001370935.1	PRJEB7988	GCA_001371975.1
PRJEB7988	GCA_001370035.1	PRJEB7988	GCA_001369695.1	PRJEB7988	GCA_001049015.1	PRJEB7988	GCA_001371195.1	PRJEB7988	GCA_001370715.1	PRJEB7988	GCA_001028425.1
PRJEB7988	GCA_001049255.1	PRJEB7988	GCA_001371655.1	PRJEB7988	GCA_001370455.1	PRJEB7988	GCA_001048995.1	PRJEB7988	GCA_001370795.1	PRJEB7988	GCA_001370835.1
PRJEB7988	GCA_001370875.1	PRJEB7988	GCA_001371475.1	PRJEB7988	GCA_001370975.1	PRJEB7988	GCA_001370395.1	PRJEB7988	GCA_001049035.1	PRJNA66351	GCA_002197955.1
PRJEB7988	GCA_001028485.1	PRJEB7988	GCA_001370355.1	PRJEB7988	GCA_001370135.1	PRJEB7988	GCA_001040265.1	PRJEB7988	GCA_001049055.1	PRJNA217777	GCA_001757205.1
PRJEB7988	GCA_001370855.1	PRJEB7988	GCA_001040605.1	PRJEB7988	GCA_001048895.1	PRJEB7988	GCA_001371955.1	PRJEB7988	GCA_001048855.1	PRJNA217777	GCA_001757275.1
PRJEB7988	GCA_001048835.1	PRJEB7988	GCA_001039985.1	PRJEB7988	GCA_001369655.1	PRJEB7988	GCA_001371295.1	PRJEB7988	GCA_001371555.1	PRJNA217777	GCA_001757265.1
PRJEB7988	GCA_001370755.1	PRJEB7988	GCA_001370995.1	PRJEB7988	GCA_001369915.1	PRJEB7988	GCA_001370775.1	PRJEB7988	GCA_001048935.1	PRJNA217777	GCA_001757305.1
PRJEB7988	GCA_001040005.1	PRJEB7988	GCA_001370555.1	PRJEB7988	GCA_001371035.1	PRJEB7988	GCA_001372035.1	PRJEB7988	GCA_001040645.1	PRJNA28347	GCA_000207745.1
PRJEB7988	GCA_001048915.1	PRJEB7988	GCA_001370695.1	PRJEB7988	GCA_001369835.1	PRJEB7988	GCA_001179605.1	PRJEB7988	GCA_001371095.1	PRJNA30009	GCA_000205805.1
PRJEB7988	GCA_001040105.1	PRJEB7988	GCA_001179645.1	PRJEB7988	GCA_001179545.1	PRJEB7988	GCA_001371795.1	PRJEB7988	GCA_001370055.1	PRJEB14308	GCA_900089805.1
PRJEB7988	GCA_001371255.1	PRJEB7988	GCA_001039885.1	PRJEB7988	GCA_001261955.1	PRJEB7988	GCA_001049215.1	PRJEB7988	GCA_001369495.1	PRJNA28363	GCA_000207585.1
PRJEB7988	GCA_001040525.1	PRJEB7988	GCA_001371675.1	PRJEB7988	GCA_001369775.1	PRJEB7988	GCA_001369735.1	PRJEB7988	GCA_001371115.1	PRJNA30011	GCA_000205825.1
PRJEB7988	GCA_001371235.1	PRJEB7988	GCA_001370295.1	PRJEB7988	GCA_001370515.1	PRJEB7988	GCA_001040565.1	PRJEB7988	GCA_000982495.1	PRJNA337945	GCA_001888125.1
PRJEB7988	GCA_001048815.1	PRJEB7988	GCA_001179585.1	PRJEB7988	GCA_001040665.1	PRJEB7988	GCA_001372195.1	PRJEB7988	GCA_001179685.1	PRJNA28367	GCA_000207545.1
PRJEB7988	GCA_001370575.1	PRJEB7988	GCA_001371055.1	PRJEB7988	GCA_001371335.1	PRJEB7988	GCA_001370335.1	PRJEB7988	GCA_001049195.1	PRJNA254808	GCA_001757165.1
PRJEB7988	GCA_001040085.1	PRJEB7988	GCA_001371355.1	PRJEB7988	GCA_001371575.1	PRJEB7988	GCA_001370175.1	PRJEB7988	GCA_001049315.1	PRJNA254808	GCA_001757225.1
PRJEB7988	GCA_001371495.1	PRJEB7988	GCA_001372095.1	PRJEB7988	GCA_001370535.1	PRJEB7988	GCA_001370215.1	PRJEB7988	GCA_001039905.1	PRJNA254808	GCA_001757185.1
PRJEB7988	GCA_001048775.1	PRJEB7988	GCA_001371615.1	PRJEB7988	GCA_001370375.1	PRJEB7988	GCA_001040245.1	PRJEB7988	GCA_001040165.1	PRJNA254808	GCA_001757315.1
PRJEB7988	GCA_001040305.1	PRJEB7988	GCA_001370095.1	PRJEB7988	GCA_001048755.1	PRJEB7988	GCA_001370115.1	PRJEB7988	GCA_001370495.1	PRJNA254808	GCA_001757245.1
PRJEB7988	GCA_001179665.1	PRJEB7988	GCA_001048875.1	PRJEB7988	GCA_001372115.1	PRJEB7988	GCA_001049295.1	PRJEB7988	GCA_001369755.1	PRJEB19326	GCA_900162855.1
PRJEB7988	GCA_001369615.1	PRJEB7988	GCA_001371275.1	PRJEB7988	GCA_001371155.1	PRJEB7988	GCA_001040425.1	PRJEB7988	GCA_001370415.1	PRJNA286221	GCA_001563815.1
PRJEB7988	GCA_001371895.1	PRJEB7988	GCA_001369595.1	PRJEB7988	GCA_001369515.1	PRJEB7988	GCA_001369475.1	PRJEB7988	GCA_001040585.1	PRJNA286221	GCA_001564315.1
PRJEB7988	GCA_001372175.1	PRJEB7988	GCA_001370015.1	PRJEB7988	GCA_001040065.1	PRJEB7988	GCA_001049095.1	PRJEB7988	GCA_000982535.1	PRJNA286221	GCA_001563825.1
PRJEB7988	GCA_001040505.1	PRJEB7988	GCA_001371775.1	PRJEB7988	GCA_001372135.1	PRJEB7988	GCA_000982475.1	PRJEB7988	GCA_001040045.1	PRJNA286221	GCA_001564335.1
PRJEB7988	GCA_001049075.1	PRJEB7988	GCA_001369555.1	PRJEB7988	GCA_001040025.1	PRJEB7988	GCA_001370595.1	PRJEB7988	GCA_001049235.1	PRJNA158313	GCA_000294405.1
PRJEB7988	GCA_001048955.1	PRJEB7988	GCA_001048975.1	PRJEB7988	GCA_001040325.1	PRJEB7988	GCA_001369995.1	PRJEB7988	GCA_001370435.1	PRJNA30129	GCA_000206285.1
PRJEB7988	GCA_001369955.1	PRJEB7988	GCA_001369895.1	PRJEB7988	GCA_001040345.1	PRJEB7988	GCA_001371015.1	PRJEB7988	GCA_001370955.1	PRJNA30131	GCA_000206265.1
PRJEB7988	GCA_001369975.1	PRJEB7988	GCA_001369635.1	PRJEB7988	GCA_001370615.1	PRJEB7988	GCA_001372075.1	PRJEB7988	GCA_001179565.1	PRJNA30127	GCA_000206325.1
PRJEB7988	GCA_001370075.1	PRJEB7988	GCA_001369675.1	PRJEB7988	GCA_001369575.1	PRJEB7988	GCA_001369875.1	PRJEB7988	GCA_001371855.1	PRJNA30135	GCA_000206305.1
PRJEB7988	GCA_001369715.1	PRJEB7988	GCA_001369535.1	PRJEB7988	GCA_001370155.1	PRJEB7988	GCA_001372015.1	PRJEB7988	GCA_001179525.1	PRJNA30133	GCA_000206245.1
PRJEB7988	GCA_001040145.1	PRJEB7988	GCA_001371835.1	PRJEB7988	GCA_001370655.1	PRJEB7988	GCA_001048795.1	PRJEB7988	GCA_001371715.1	PRJNA167727	GCA_000315215.1
PRJEB7988	GCA_001370675.1	PRJEB7988	GCA_001049155.1	PRJEB7988	GCA_001040685.1	PRJEB7988	GCA_001028405.1	PRJEB7988	GCA_001040185.1	PRJNA230141	GCA_000565115.1
PRJEB7988	GCA_001040465.1	PRJEB7988	GCA_001040545.1	PRJEB7988	GCA_001370815.1	PRJEB7988	GCA_001039965.1	PRJEB7988	GCA_001372055.1	PRJNA73973	GCA_000287795.1
PRJEB7988	GCA_001370895.1	PRJEB7988	GCA_001371435.1	PRJEB7988	GCA_001040225.1	PRJEB7988	GCA_001040385.1	PRJEB7988	GCA_001049275.1	PRJNA255918	GCA_000763215.1
PRJEB7988	GCA_001040485.1	PRJEB7988	GCA_001370195.1	PRJEB7988	GCA_001371135.1	PRJEB7988	GCA_001040205.1	PRJEB7988	GCA_001371695.1	PRJNA49615	GCA_000205665.2
PRJEB7988	GCA_001048735.1	PRJEB7988	GCA_001039925.1	PRJEB7988	GCA_001049175.1	PRJEB7988	GCA_001369815.1	PRJEB7988	GCA_001370275.1	PRJNA258478	GCA_001757535.1
PRJEB7988	GCA_001369855.1	PRJEB7988	GCA_001370735.1	PRJEB7988	GCA_001049115.1	PRJEB7988	GCA_001372215.1	PRJEB7988	GCA_001370475.1	PRJNA297058	GCA_001316455.1
PRJEB7988	GCA_001371175.1	PRJEB7988	GCA_001371815.1	PRJEB7988	GCA_001369795.1	PRJEB7988	GCA_001370915.1	PRJEB7988	GCA_001370255.1	PRJNA223357	GCA_001469255.1
PRJEB7988	GCA_001371375.1	PRJEB7988	GCA_001371875.1	PRJEB7988	GCA_001179625.1	PRJEB7988	GCA_001028465.1	PRJEB7988	GCA_001028445.1	PRJNA222667	GCA_001045095.1
PRJEB7988	GCA_001370315.1	PRJEB7988	GCA_001370235.1	PRJEB7988	GCA_001049135.1	PRJEB7988	GCA_001371535.1	PRJEB7988	GCA_001039945.1	PRJDB1119	GCA_001682595.1
PRJEB7988	GCA_000982635.1	PRJEB7988	GCA_001371315.1	PRJEB7988	GCA_001371755.1	PRJEB7988	GCA_001371915.1	PRJEB7988	GCA_001370635.1	PRJNA222659	GCA_001044975.1
PRJEB7988	GCA_001371455.2	PRJEB7988	GCA_001040285.1	PRJEB7988	GCA_001371415.1	PRJEB7988	GCA_001040405.1	PRJEB7988	GCA_001371935.1	PRJNA222664	GCA_001045045.1
PRJEB7988	GCA_001040125.2	PRJEB7988	GCA_001372155.1	PRJEB7988	GCA_001371995.1	PRJEB7988	GCA_001371635.1	PRJEB7988	GCA_000982515.1	PRJDB1115	GCA_001682615.1

Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id
PRJNA260268	GCA_000753565.1	PRJDB4437	GCA_002003205.1	PRJNA234377	GCA_000681455.1	PRJNA261476	GCA_000803525.1	PRJEB9703	GCA_900006455.1	PRJEB9703	GCA_900007015.1
PRJNA260268	GCA_000753555.1	PRJDB4437	GCA_002003245.1	PRJNA28439	GCA_000206945.1	PRJNA261480	GCA_000803465.1	PRJEB9703	GCA_900007225.1	PRJEB9703	GCA_900006065.1
PRJNA222666	GCA_001045155.1	PRJDB4437	GCA_002013255.1	PRJNA28441	GCA_000206925.1	PRJNA271635	GCA_000987625.1	PRJEB9703	GCA_900007455.1	PRJEB9703	GCA_900006545.1
PRJDB1117	GCA_001682555.1	PRJNA28349	GCA_000207725.1	PRJNA213161	GCA_000496385.1	PRJNA271635	GCA_000987635.1	PRJEB9703	GCA_900007305.1	PRJEB9703	GCA_900006365.1
PRJNA222657	GCA_001044995.1	PRJEB19352	GCA_900163725.1	PRJNA306036	GCA_001577415.1	PRJNA271635	GCA_000987685.1	PRJEB9703	GCA_900006175.1	PRJEB9703	GCA_900006105.1
PRJDB1118	GCA_001682535.1	PRJNA28369	GCA_000207525.1	PRJNA306036	GCA_001577405.1	PRJNA271635	GCA_000987615.1	PRJEB9703	GCA_900006895.1	PRJEB9703	GCA_900007155.1
PRJNA222665	GCA_001045125.1	PRJNA28365	GCA_000207565.1	PRJNA306036	GCA_001577435.1	PRJNA271635	GCA_000987605.1	PRJEB9703	GCA_900006925.1	PRJEB9703	GCA_900006355.1
PRJDB1116	GCA_001682575.1	PRJNA28451	GCA_000206825.1	PRJNA62043	GCA_000227035.1	PRJNA261470	GCA_000803365.1	PRJEB9703	GCA_900007355.1	PRJEB9703	GCA_900006495.1
PRJNA290197	GCA_001263475.1	PRJNA28457	GCA_000206765.1	PRJNA62039	GCA_000205505.1	PRJNA261472	GCA_000803385.1	PRJEB9703	GCA_900006075.1	PRJEB9703	GCA_900006755.1
PRJNA290197	GCA_001263485.1	PRJNA28353	GCA_000207685.1	PRJNA62041	GCA_000227015.1	PRJNA260954	GCA_000803605.1	PRJEB9703	GCA_900009855.1	PRJEB9703	GCA_900007365.1
PRJNA290197	GCA_001263465.1	PRJNA28465	GCA_000206685.1	PRJNA39805	GCA_000205645.1	PRJNA261478	GCA_000803445.1	PRJEB9703	GCA_900006515.1	PRJEB9703	GCA_900006915.1
PRJNA290197	GCA_001263535.1	PRJNA28463	GCA_000206705.1	PRJEB14783	GCA_900094055.1	PRJEB22550	GCA_900216985.1	PRJEB9703	GCA_900006905.1	PRJEB9703	GCA_900006775.1
PRJNA290197	GCA_001263455.1	PRJNA28445	GCA_000206885.1	PRJEB14783	GCA_900163505.1	PRJEB22550	GCA_900216975.1	PRJEB9703	GCA_900006885.1	PRJEB9703	GCA_900005995.1
PRJNA222660	GCA_001045025.1	PRJNA28373	GCA_000207485.1	PRJEB14783	GCA_900163525.1	PRJNA17659	GCA_000208305.1	PRJEB9703	GCA_900006145.1	PRJEB9703	GCA_900006615.1
PRJNA222663	GCA_001045145.1	PRJNA28375	GCA_000207465.1	PRJEB14783	GCA_900163515.1	PRJNA17657	GCA_000208285.1	PRJEB9703	GCA_900006785.1	PRJEB9703	GCA_900006115.1
PRJNA259156	GCA_000987305.1	PRJNA28447	GCA_000206865.1	PRJEB14783	GCA_900094065.1	PRJNA171369	GCA_001945615.1	PRJEB9703	GCA_900007065.1	PRJEB9703	GCA_900006825.1
PRJNA222658	GCA_001045015.1	PRJNA28455	GCA_000206785.1	PRJNA32357	GCA_000206225.1	PRJNA244670	GCA_002007545.1	PRJEB9703	GCA_900007215.1	PRJEB9703	GCA_900007205.1
PRJNA222661	GCA_001045005.1	PRJNA28401	GCA_000207205.1	PRJNA32355	GCA_000206205.1	PRJNA244670	GCA_002007525.1	PRJEB9703	GCA_900006425.1	PRJEB9703	GCA_900006095.1
PRJNA287316	GCA_001543945.1	PRJNA28403	GCA_000207185.1	PRJNA32359	GCA_000206185.1	PRJNA360538	GCA_002049745.1	PRJEB9703	GCA_900019255.1	PRJEB9703	GCA_900006535.1
PRJNA287316	GCA_001543535.1	PRJNA28399	GCA_000207225.1	PRJNA17779	GCA_000208265.2	PRJNA283612	GCA_001577815.1	PRJEB9703	GCA_900006685.1	PRJEB9703	GCA_900006025.1
PRJNA28425	GCA_000206045.1	PRJNA28397	GCA_000207245.1	PRJNA374545	GCA_002022945.1	PRJNA281982	GCA_001025305.1	PRJEB9703	GCA_900006645.1	PRJEB9703	GCA_900006415.1
PRJNA28419	GCA_000206105.1	PRJNA28601	GCA_000206625.1	PRJNA374545	GCA_002022935.1	PRJNA281982	GCA_001025115.1	PRJEB9703	GCA_900007295.1	PRJEB9703	GCA_900006715.1
PRJNA20995	GCA_000206365.2	PRJNA28361	GCA_000207605.1	PRJNA374545	GCA_002022905.1	PRJNA360539	GCA_002049765.1	PRJEB9703	GCA_900007125.1	PRJEB9703	GCA_900007075.1
PRJNA28421	GCA_000206085.1	PRJNA28411	GCA_000207105.1	PRJNA374545	GCA_002022885.1	PRJNA283298	GCA_001513565.1	PRJEB9703	GCA_900006055.1	PRJEB9703	GCA_900006965.1
PRJNA28423	GCA_000206065.1	PRJNA49085	GCA_000205725.1	PRJNA374545	GCA_002022925.1	PRJNA360243	GCA_002049705.1	PRJEB9703	GCA_900006955.1	PRJEB9703	GCA_900007255.1
PRJNA28415	GCA_000206145.1	PRJNA49085	GCA_000205745.1	PRJNA193217	GCA_001684395.1	PRJNA360534	GCA_002049725.1	PRJEB9703	GCA_900006465.1	PRJEB9703	GCA_900006435.1
PRJNA28417	GCA_000206125.1	PRJNA28409	GCA_000207125.1	PRJNA28461	GCA_000206725.1	PRJNA239997	GCA_001509935.1	PRJEB9703	GCA_900006085.1	PRJEB9703	GCA_900006235.1
PRJNA29311	GCA_000206345.2	PRJNA245479	GCA_000734835.1	PRJNA28377	GCA_000207445.1	PRJNA239997	GCA_001509945.1	PRJEB9703	GCA_900006695.1	PRJEB9703	GCA_900006225.1
PRJNA28355	GCA_000207665.1	PRJNA273640	GCA_000987175.1	PRJNA28443	GCA_000206905.1	PRJNA239997	GCA_001509955.1	PRJEB9703	GCA_900006625.1	PRJEB9703	GCA_900006125.1
PRJNA28381	GCA_000207405.1	PRJNA273640	GCA_000987145.1	PRJNA28379	GCA_000207425.1	PRJEB10932	GCA_001460915.1	PRJEB9703	GCA_900006555.1	PRJEB9703	GCA_900007285.1
PRJNA28357	GCA_000207645.1	PRJNA273640	GCA_000987365.1	PRJNA28359	GCA_000207625.1	PRJEB10932	GCA_001460895.1	PRJEB9703	GCA_900006525.1	PRJEB9703	GCA_900006485.1
PRJNA293159	GCA_001509455.1	PRJNA273640	GCA_000987125.1	PRJNA28453	GCA_000206805.1	PRJNA19107	GCA_000208185.1	PRJEB9703	GCA_900006975.1	PRJEB9703	GCA_900006815.1
PRJNA293159	GCA_001509435.1	PRJNA273640	GCA_000987135.1	PRJNA28459	GCA_000206745.1	PRJNA78659	GCA_001577735.1	PRJEB9703	GCA_900007115.1	PRJEB9703	GCA_900006505.1
PRJNA28469	GCA_000206645.1	PRJNA273640	GCA_000987265.1	PRJNA28449	GCA_000206845.1	PRJNA168188	GCA_000312065.1	PRJEB9703	GCA_900006635.1	PRJEB9703	GCA_900006445.1
PRJNA28413	GCA_000207085.1	PRJNA273640	GCA_000987465.1	PRJNA261481	GCA_000803545.1	PRJNA163505	GCA_000320525.1	PRJEB9703	GCA_900007465.1	PRJEB9703	GCA_900006805.1
PRJNA28467	GCA_000206665.1	PRJNA273640	GCA_000987475.1	PRJNA261477	GCA_000803505.1	PRJNA169114	GCA_000312085.1	PRJEB9703	GCA_900006565.1	PRJEB9703	GCA_900006765.1
PRJNA343412	GCA_002162635.1	PRJNA273640	GCA_000987285.1	PRJNA322674	GCA_003332265.1	PRJNA301005	GCA_001468985.1	PRJEB9703	GCA_900006475.1	PRJNA27877	GCA_000207945.1
PRJEB9818	GCA_001292545.1	PRJNA273640	GCA_000987235.1	PRJNA322674	GCA_002254365.1	PRJNA28383	GCA_000207385.1	PRJEB9703	GCA_900006995.1	PRJNA88107	GCA_000299075.1
PRJDB4437	GCA_002013235.1	PRJNA273640	GCA_000987225.1	PRJNA322674	GCA_002254325.1	PRJNA28351	GCA_000207705.1	PRJEB9703	GCA_900006725.1	PRJNA43253	GCA_000205765.1
PRJDB4437	GCA_002003125.1	PRJNA273640	GCA_000987255.1	PRJNA261475	GCA_000803425.1	PRJNA28385	GCA_000207365.1	PRJEB9703	GCA_900006585.1	PRJNA43253	GCA_000205785.1
PRJDB4437	GCA_002003145.1	PRJNA64629	GCA_000270545.1	PRJNA61401	GCA_000205625.1	PRJNA296403	GCA_001441265.1	PRJEB9703	GCA_900006795.1	PRJNA27865	GCA_000208065.1
PRJDB4437	GCA_002003185.1	PRJNA283341	GCA_001028805.1	PRJNA260953	GCA_000803585.1	PRJNA200716	GCA_000414385.1	PRJEB9703	GCA_900006195.1	PRJEB18557	GCA_900149225.1
PRJDB4437	GCA_002003225.1	PRJNA283341	GCA_001029185.1	PRJNA260952	GCA_000803565.1	PRJNA200717	GCA_000414405.1	PRJEB9703	GCA_900006215.1	PRJEB18557	GCA_900149235.1
PRJDB4437	GCA_002003165.1	PRJNA283341	GCA_001028855.1	PRJNA261473	GCA_000803405.1	PRJNA289617	GCA_001566565.1	PRJEB9703	GCA_900007235.1	PRJEB18557	GCA_900149205.1
PRJDB4437	GCA_002013275.1	PRJNA234377	GCA_001005475.1	PRJNA261471	GCA_000803485.1	PRJNA50815	GCA_000205545.2	PRJEB9703	GCA_900007045.1	PRJEB18557	GCA_900149215.1

Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id
PRJEB18557	GCA_900149185.1	PRJEB9169	GCA_001458935.1	PRJEB9169	GCA_001459455.1	PRJEB8094	GCA_001408895.1	PRJEB8094	GCA_001403375.1	PRJEB10725	GCA_900069915.1
PRJNA27875	GCA_000207965.1	PRJEB9169	GCA_001460135.1	PRJEB9169	GCA_001460315.1	PRJEB8094	GCA_001408715.1	PRJEB8094	GCA_001403255.1	PRJEB10725	GCA_900000245.1
PRJNA27871	GCA_000208005.1	PRJEB9169	GCA_001460155.1	PRJEB9169	GCA_001459655.1	PRJEB8094	GCA_001408995.1	PRJEB8094	GCA_001408735.1	PRJEB10725	GCA_900000295.1
PRJEB9169	GCA_001459555.1	PRJEB9169	GCA_001460215.1	PRJEB9169	GCA_001460055.1	PRJEB8094	GCA_001403515.1	PRJEB8094	GCA_001403235.1	PRJEB10725	GCA_900000315.1
PRJEB9169	GCA_001460075.1	PRJEB9169	GCA_001460605.1	PRJNA27867	GCA_000208045.1	PRJEB8094	GCA_001408455.1	PRJEB8094	GCA_001403555.1	PRJEB10725	GCA_900069885.1
PRJEB9169	GCA_001459255.1	PRJEB9169	GCA_001458955.1	PRJNA60717	GCA_000205525.2	PRJEB8094	GCA_001282745.1	PRJEB8094	GCA_001408795.1	PRJEB10725	GCA_001517605.1
PRJEB9169	GCA_001459995.1	PRJEB9169	GCA_001459315.1	PRJNA27859	GCA_000208125.1	PRJEB8094	GCA_001403395.1	PRJEB8094	GCA_001403215.1	PRJEB10725	GCA_900069835.1
PRJEB9169	GCA_001458895.1	PRJEB9169	GCA_001460655.1	PRJNA27857	GCA_000208145.1	PRJEB8094	GCA_001403295.1	PRJEB8094	GCA_001282865.1	PRJEB10725	GCA_900069905.1
PRJEB9169	GCA_001460275.1	PRJEB9169	GCA_001459395.1	PRJEB4707	GCA_001511515.1	PRJEB8094	GCA_001403315.1	PRJEB8094	GCA_001403335.1	PRJEB10725	GCA_900069855.1
PRJEB9169	GCA_001460235.1	PRJEB9169	GCA_001459515.1	PRJNA27879	GCA_000207925.1	PRJEB8094	GCA_001408435.1	PRJEB8094	GCA_001408255.1	PRJEB10725	GCA_900000265.1
PRJEB9169	GCA_001459075.1	PRJEB9169	GCA_001459915.1	PRJNA16729	GCA_000208325.1	PRJEB8094	GCA_001403135.1	PRJEB8094	GCA_001408475.1	PRJEB10725	GCA_900069925.1
PRJEB9169	GCA_001459435.1	PRJEB9169	GCA_001459215.1	PRJNA16729	GCA_000208345.1	PRJEB8094	GCA_001409115.1	PRJEB8094	GCA_001408415.1	PRJEB10725	GCA_900000255.1
PRJEB9169	GCA_001460455.1	PRJEB9169	GCA_001459735.1	PRJNA27855	GCA_000208165.1	PRJEB8094	GCA_001403355.1	PRJEB8094	GCA_001408535.1	PRJEB10725	GCA_900069895.1
PRJEB9169	GCA_001458915.1	PRJEB9169	GCA_001459975.1	PRJNA219732	GCA_000508725.1	PRJEB8094	GCA_001408855.1	PRJEB8094	GCA_001403435.1	PRJEB10725	GCA_900069945.1
PRJEB9169	GCA_001459935.1	PRJEB9169	GCA_001460735.1	PRJNA27861	GCA_000208105.1	PRJEB8094	GCA_001408575.1	PRJEB8094	GCA_001408315.1	PRJEB10725	GCA_001544415.1
PRJEB9169	GCA_001459175.1	PRJEB9169	GCA_001460175.1	PRJNA27863	GCA_000208085.1	PRJEB8094	GCA_001403175.1	PRJEB8094	GCA_001282825.1	PRJEB10725	GCA_900000335.1
PRJEB9169	GCA_001460795.1	PRJEB9169	GCA_001460415.1	PRJEB9488	GCA_900005485.1	PRJEB8094	GCA_001403415.1	PRJEB8094	GCA_001409015.1	PRJNA72925	GCA_000245375.1
PRJEB9169	GCA_001460815.1	PRJEB9169	GCA_001460035.1	PRJEB9488	GCA_900005805.1	PRJEB8094	GCA_001403075.1	PRJEB8094	GCA_001403115.1	PRJNA330682	GCA_001714665.1
PRJEB9169	GCA_001459495.1	PRJEB9169	GCA_001459815.1	PRJEB9488	GCA_900004985.1	PRJEB8094	GCA_001408635.1	PRJEB8094	GCA_001403275.1	PRJNA358809	GCA_002049675.1
PRJEB9169	GCA_001459035.1	PRJEB9169	GCA_001460755.1	PRJEB9488	GCA_900005425.1	PRJEB8094	GCA_001403495.1	PRJEB8094	GCA_001408615.1	PRJNA358809	GCA_002049665.1
PRJEB9169	GCA_001460335.1	PRJEB9169	GCA_001459855.1	PRJEB9488	GCA_900005015.1	PRJEB8094	GCA_001408695.1	PRJEB8094	GCA_001408835.1	PRJNA358809	GCA_002049585.1
PRJEB9169	GCA_001459475.1	PRJEB9169	GCA_001460015.1	PRJEB9488	GCA_900005395.1	PRJEB8094	GCA_001408975.1	PRJEB8094	GCA_001408555.1	PRJNA358809	GCA_002049595.1
PRJEB9169	GCA_001459415.1	PRJEB9169	GCA_001459155.1	PRJEB9488	GCA_900005645.1	PRJEB8094	GCA_001408495.1	PRJEB8094	GCA_001408815.1	PRJNA358809	GCA_002049645.1
PRJEB9169	GCA_001460775.1	PRJEB9169	GCA_001459535.1	PRJEB9488	GCA_900005795.1	PRJEB8094	GCA_001408355.1	PRJEB8094	GCA_001409075.1	PRJNA358809	GCA_002049625.1
PRJEB9169	GCA_001459275.1	PRJEB9169	GCA_001460435.1	PRJEB9488	GCA_900004975.1	PRJEB8094	GCA_001408915.1	PRJNA88153	GCA_000299055.1	PRJNA202911	GCA_001482205.1
PRJEB9169	GCA_001460675.1	PRJEB9169	GCA_001460715.1	PRJEB9488	GCA_900005635.1	PRJEB8094	GCA_001403015.1	PRJNA27873	GCA_000207985.1	PRJNA202911	GCA_001482235.1
PRJEB9169	GCA_001460115.1	PRJEB9169	GCA_001459755.1	PRJEB9488	GCA_900005005.1	PRJEB8094	GCA_001403035.1	PRJNA13699	GCA_000208365.1	PRJNA202911	GCA_001482255.1
PRJEB9169	GCA_001460395.1	PRJEB9169	GCA_001459895.1	PRJEB9488	GCA_900005475.1	PRJEB8094	GCA_001408655.1	PRJNA192218	GCA_000496495.1	PRJNA269960	GCA_000972885.1
PRJEB9169	GCA_001459115.1	PRJEB9169	GCA_001460295.1	PRJEB9488	GCA_900005455.1	PRJEB8094	GCA_001408595.1	PRJNA271842	GCA_000960285.1	PRJEB9761	GCA_900078645.1
PRJEB9169	GCA_001459875.1	PRJEB9169	GCA_001459355.1	PRJEB9488	GCA_900005445.1	PRJEB8094	GCA_001408755.1	PRJNA271842	GCA_000960325.1	PRJEB9761	GCA_900078595.1
PRJEB9169	GCA_001459055.1	PRJEB9169	GCA_001459575.1	PRJEB9488	GCA_900004995.1	PRJEB8094	GCA_001403195.1	PRJNA271842	GCA_000960345.1	PRJEB9761	GCA_900078625.1
PRJEB9169	GCA_001459235.1	PRJEB9169	GCA_001458975.1	PRJEB9488	GCA_900005385.1	PRJEB8094	GCA_001409095.1	PRJNA271842	GCA_000960355.1	PRJEB9761	GCA_900078655.1
PRJEB9169	GCA_001459375.1	PRJEB9169	GCA_001459195.1	PRJEB9488	GCA_900005885.1	PRJEB8094	GCA_001408395.1	PRJNA271842	GCA_000960275.1	PRJEB9761	GCA_900078635.1
PRJEB9169	GCA_001459635.1	PRJEB9169	GCA_001459835.1	PRJEB9488	GCA_900005405.1	PRJEB8094	GCA_001282845.1	PRJNA271842	GCA_000960405.1	PRJEB9761	GCA_900078615.1
PRJEB9169	GCA_001460695.1	PRJEB9169	GCA_001458875.1	PRJEB9488	GCA_900005435.1	PRJEB8094	GCA_001408875.1	PRJEB10725	GCA_900000305.1	PRJEB9761	GCA_900078565.1
PRJEB9169	GCA_001460095.1	PRJEB9169	GCA_001458995.1	PRJEB9488	GCA_900005665.1	PRJEB8094	GCA_001409055.1	PRJEB10725	GCA_900000235.1	PRJEB9761	GCA_900078605.1
PRJEB9169	GCA_001460355.1	PRJEB9169	GCA_001458835.1	PRJEB9488	GCA_900005735.1	PRJEB8094	GCA_001408295.1	PRJEB10725	GCA_900069875.1	PRJEB9761	GCA_900078555.1
PRJEB9169	GCA_001459015.1	PRJEB9169	GCA_001460195.1	PRJEB9488	GCA_900005725.1	PRJEB8094	GCA_001408375.1	PRJEB10725	GCA_900069845.1	PRJEB9761	GCA_900078585.1
PRJEB9169	GCA_001459675.1	PRJEB9169	GCA_001459595.1	PRJEB9488	GCA_900005415.1	PRJEB8094	GCA_001408775.1	PRJEB10725	GCA_900069955.1	PRJEB9761	GCA_900078575.1
PRJEB9169	GCA_001459295.1	PRJEB9169	GCA_001459715.1	PRJEB9488	GCA_900005655.1	PRJEB8094	GCA_001409035.1	PRJEB10725	GCA_900000325.1	PRJNA376086	GCA_002058925.1
PRJEB9169	GCA_001458855.1	PRJEB9169	GCA_001460255.1	PRJEB9488	GCA_900005465.1	PRJEB8094	GCA_001408955.1	PRJEB10725	GCA_900069935.1	PRJNA376086	GCA_002059125.1
PRJEB9169	GCA_001460485.1	PRJEB9169	GCA_001459615.1	PRJNA27869	GCA_000208025.1	PRJEB8094	GCA_001408275.1	PRJEB10725	GCA_900000285.1	PRJNA376086	GCA_002058865.1
PRJEB9169	GCA_001459135.1	PRJEB9169	GCA_001459955.1	PRJEB8094	GCA_001408675.1	PRJEB8094	GCA_001403455.1	PRJEB10725	GCA_900069865.1	PRJNA376086	GCA_002058825.1
PRJEB9169	GCA_001460375.1	PRJEB9169	GCA_001459795.1	PRJEB8094	GCA_001403055.1	PRJEB8094	GCA_001403155.1	PRJEB10725	GCA_900000225.1	PRJNA376086	GCA_002058845.1
PRJEB9169	GCA_001459335.1	PRJEB9169	GCA_001459695.1	PRJEB8094	GCA_001403535.1	PRJEB8094	GCA_001403475.1	PRJEB10725	GCA_900000275.1	PRJNA376086	GCA_002058885.1

Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id
PRJNA376086	GCA_002059005.1	PRJNA28603	GCA_000206605.1	PRJEB21624	GCA_900321695.1	PRJEB21624	GCA_900317085.1	PRJEB21624	GCA_900316365.1	PRJEB21624	GCA_900314695.1
PRJNA376086	GCA_002059145.1	PRJNA238866	GCA_000754565.1	PRJEB21624	GCA_900318035.1	PRJEB21624	GCA_900321625.1	PRJEB21624	GCA_900321435.1	PRJEB21624	GCA_900320775.1
PRJNA376086	GCA_002058985.1	PRJNA238866	GCA_000754525.1	PRJEB21624	GCA_900316155.1	PRJEB21624	GCA_900316295.1	PRJEB21624	GCA_900314145.1	PRJEB21624	GCA_900320605.1
PRJNA376086	GCA_002059065.1	PRJNA238866	GCA_000754545.1	PRJEB21624	GCA_900316285.1	PRJEB21624	GCA_900317535.1	PRJEB21624	GCA_900315845.1	PRJEB21624	GCA_900321925.1
PRJNA376086	GCA_002059045.1	PRJNA28407	GCA_000207145.1	PRJEB21624	GCA_900320215.1	PRJEB21624	GCA_900313695.1	PRJEB21624	GCA_900317525.1	PRJEB21624	GCA_900320185.1
PRJNA376086	GCA_002059025.1	PRJNA259931	GCA_001515565.1	PRJEB21624	GCA_900320005.1	PRJEB21624	GCA_900321515.1	PRJEB21624	GCA_900313475.1	PRJEB21624	GCA_900317725.1
PRJNA376086	GCA_002058905.1	PRJEB13176	GCA_900079655.1	PRJEB21624	GCA_900318985.1	PRJEB21624	GCA_900314655.1	PRJEB21624	GCA_900313605.1	PRJEB21624	GCA_900319995.1
PRJNA376086	GCA_002059165.1	PRJEB13176	GCA_900143125.1	PRJEB21624	GCA_900314525.1	PRJEB21624	GCA_900319695.1	PRJEB21624	GCA_900316075.1	PRJEB21624	GCA_900314615.1
PRJNA376086	GCA_002058965.1	PRJNA28387	GCA_000207345.1	PRJEB21624	GCA_900318455.1	PRJEB21624	GCA_900321205.1	PRJEB21624	GCA_900320195.1	PRJEB21624	GCA_900316195.1
PRJNA376086	GCA_002059085.1	PRJNA257561	GCA_000802025.2	PRJEB21624	GCA_900319445.1	PRJEB21624	GCA_900320355.1	PRJEB21624	GCA_900316815.1	PRJEB21624	GCA_900320735.1
PRJNA376086	GCA_002058945.1	PRJNA257561	GCA_000961915.2	PRJEB21624	GCA_900313145.1	PRJEB21624	GCA_900316995.1	PRJEB21624	GCA_900317565.1	PRJEB21624	GCA_900321885.1
PRJNA376086	GCA_002059105.1	PRJNA232936	GCA_000690855.1	PRJEB21624	GCA_900315975.1	PRJEB21624	GCA_900315075.1	PRJEB21624	GCA_900316935.1	PRJEB21624	GCA_900320755.1
PRJNA20811	GCA_000206385.2	PRJNA206543	GCA_000496515.1	PRJEB21624	GCA_900313935.1	PRJEB21624	GCA_900316835.1	PRJEB21624	GCA_900316855.1	PRJEB21624	GCA_900319495.1
PRJEB11544	GCA_900010595.1	PRJNA42235	GCA_000205925.2	PRJEB21624	GCA_900314575.1	PRJEB21624	GCA_900318695.1	PRJEB21624	GCA_900313585.1	PRJEB21624	GCA_900319785.1
PRJNA324509	GCA_001684045.1	PRJNA68063	GCA_000221925.1	PRJEB21624	GCA_900316185.1	PRJEB21624	GCA_900320935.1	PRJEB21624	GCA_900321495.1	PRJEB21624	GCA_900316065.1
PRJNA324506	GCA_001684035.1	PRJNA294671	GCA_001879665.1	PRJEB21624	GCA_900317755.1	PRJEB21624	GCA_900313405.1	PRJEB21624	GCA_900315795.1	PRJEB21624	GCA_900318225.1
PRJNA324507	GCA_001684095.1	PRJNA42239	GCA_000205885.2	PRJEB21624	GCA_900320715.1	PRJEB21624	GCA_900313205.1	PRJEB21624	GCA_900319155.1	PRJEB21624	GCA_900321895.1
PRJNA301899	GCA_001483165.1	PRJEB14843	GCA_900093745.1	PRJEB21624	GCA_900316795.1	PRJEB21624	GCA_900320685.1	PRJEB21624	GCA_900320535.1	PRJEB21624	GCA_900317965.1
PRJNA324366	GCA_001684015.1	PRJEB14843	GCA_900093735.1	PRJEB21624	GCA_900318615.1	PRJEB21624	GCA_900313895.1	PRJEB21624	GCA_900317435.1	PRJEB21624	GCA_900313715.1
PRJNA278302	GCA_001507815.1	PRJEB14843	GCA_900093705.1	PRJEB21624	GCA_900315615.1	PRJEB21624	GCA_900321455.1	PRJEB21624	GCA_900316645.1	PRJEB21624	GCA_900316415.1
PRJNA278302	GCA_001507825.1	PRJEB14843	GCA_900093715.1	PRJEB21624	GCA_900315275.1	PRJEB21624	GCA_900320995.1	PRJEB21624	GCA_900315085.1	PRJEB21624	GCA_900321735.1
PRJNA278302	GCA_001507705.1	PRJEB14843	GCA_900093725.1	PRJEB21624	GCA_900321015.1	PRJEB21624	GCA_900321445.1	PRJEB21624	GCA_900319045.1	PRJEB21624	GCA_900315415.1
PRJNA278302	GCA_001507835.1	PRJNA42245	GCA_000205845.2	PRJEB21624	GCA_900321665.1	PRJEB21624	GCA_900317295.1	PRJEB21624	GCA_900319885.1	PRJEB21624	GCA_900318815.1
PRJNA278302	GCA_001507695.1	PRJNA42237	GCA_000205905.2	PRJEB21624	GCA_900315895.1	PRJEB21624	GCA_900316175.1	PRJEB21624	GCA_900319725.1	PRJEB21624	GCA_900313595.1
PRJNA278302	GCA_001507715.1	PRJNA42241	GCA_000205865.2	PRJEB21624	GCA_900319545.1	PRJEB21624	GCA_900315865.1	PRJEB21624	GCA_900318045.1	PRJEB21624	GCA_900318685.1
PRJNA324521	GCA_001684125.1	PRJNA28393	GCA_000207285.1	PRJEB21624	GCA_900321285.1	PRJEB21624	GCA_900316515.1	PRJEB21624	GCA_900321175.1	PRJEB21624	GCA_900316575.1
PRJNA324510	GCA_001684115.1	PRJNA28391	GCA_000207305.1	PRJEB21624	GCA_900321245.1	PRJEB21624	GCA_900319685.1	PRJEB21624	GCA_900313875.1	PRJEB21624	GCA_900315805.1
PRJNA323583	GCA_002120385.1	PRJNA28395	GCA_000207265.1	PRJEB21624	GCA_900317925.1	PRJEB21624	GCA_900321365.1	PRJEB21624	GCA_900319595.1	PRJEB21624	GCA_900319075.1
PRJNA323583	GCA_002127095.1	PRJNA305465	GCA_001580215.1	PRJEB21624	GCA_900322215.1	PRJEB21624	GCA_900318445.1	PRJEB21624	GCA_900317475.1	PRJEB21624	GCA_900319195.1
PRJNA323583	GCA_002127115.1	PRJNA28389	GCA_000207325.1	PRJEB21624	GCA_900317165.1	PRJEB21624	GCA_900321685.1	PRJEB21624	GCA_900321355.1	PRJEB21624	GCA_900322165.1
PRJNA323583	GCA_002120565.1	PRJNA338239	GCA_002003405.1	PRJEB21624	GCA_900319675.1	PRJEB21624	GCA_900316105.1	PRJEB21624	GCA_900321745.1	PRJEB21624	GCA_900318015.1
PRJNA323583	GCA_002120635.1	PRJNA338239	GCA_001998525.1	PRJEB21624	GCA_900322045.1	PRJEB21624	GCA_900318745.1	PRJEB21624	GCA_900317655.1	PRJEB21624	GCA_900314025.1
PRJNA323583	GCA_002120455.1	PRJNA338239	GCA_001998445.1	PRJEB21624	GCA_900322065.1	PRJEB21624	GCA_900316035.1	PRJEB21624	GCA_900321545.1	PRJEB21624	GCA_900319385.1
PRJNA323583	GCA_002127135.1	PRJNA28437	GCA_000206965.1	PRJEB21624	GCA_900318065.1	PRJEB21624	GCA_900317595.1	PRJEB21624	GCA_900315235.1	PRJEB21624	GCA_900314215.1
PRJNA323583	GCA_002120435.1	PRJNA28431	GCA_000207025.1	PRJEB21624	GCA_900319875.1	PRJEB21624	GCA_900319425.1	PRJEB21624	GCA_900318935.1	PRJEB21624	GCA_900319525.1
PRJNA323583	GCA_002120545.1	PRJNA28435	GCA_000206985.1	PRJEB21624	GCA_900319465.1	PRJEB21624	GCA_900314985.1	PRJEB21624	GCA_900314685.1	PRJEB21624	GCA_900318155.1
PRJNA323583	GCA_002120405.1	PRJNA28371	GCA_000207505.1	PRJEB21624	GCA_900317445.1	PRJEB21624	GCA_900317055.1	PRJEB21624	GCA_900316335.1	PRJEB21624	GCA_900322005.1
PRJNA323583	GCA_002120515.1	PRJNA28429	GCA_000207045.1	PRJEB21624	GCA_900320915.1	PRJEB21624	GCA_900315355.1	PRJEB21624	GCA_900319395.1	PRJEB21624	GCA_900318275.1
PRJNA323583	GCA_002120575.1	PRJNA28427	GCA_000207065.1	PRJEB21624	GCA_900315035.1	PRJEB21624	GCA_900315985.1	PRJEB21624	GCA_900313835.1	PRJEB21624	GCA_900318565.1
PRJNA323583	GCA_002120465.1	PRJNA28433	GCA_000207005.1	PRJEB21624	GCA_900318475.1	PRJEB21624	GCA_900322095.1	PRJEB21624	GCA_900314955.1	PRJEB21624	GCA_900315095.1
PRJNA323583	GCA_002120625.1	PRJEB21624	GCA_900320675.1	PRJEB21624	GCA_900317235.1	PRJEB21624	GCA_900313285.1	PRJEB21624	GCA_900315055.1	PRJEB21624	GCA_900321005.1
PRJNA323583	GCA_002120585.1	PRJEB21624	GCA_900321045.1	PRJEB21624	GCA_900317045.1	PRJEB21624	GCA_900314015.1	PRJEB21624	GCA_900314345.1	PRJEB21624	GCA_900317935.1
PRJNA323583	GCA_002120475.1	PRJEB21624	GCA_900317245.1	PRJEB21624	GCA_900317105.1	PRJEB21624	GCA_900315335.1	PRJEB21624	GCA_900315525.1	PRJEB21624	GCA_900319205.1
PRJNA323583	GCA_002120395.1	PRJEB21624	GCA_900316055.1	PRJEB21624	GCA_900314725.1	PRJEB21624	GCA_900313345.1	PRJEB21624	GCA_900321395.1	PRJEB21624	GCA_900319635.1
PRJNA28405	GCA_000207165.1	PRJEB21624	GCA_900317635.1	PRJEB21624	GCA_900314205.1	PRJEB21624	GCA_900315755.1	PRJEB21624	GCA_900317095.1	PRJEB21624	GCA_900314965.1

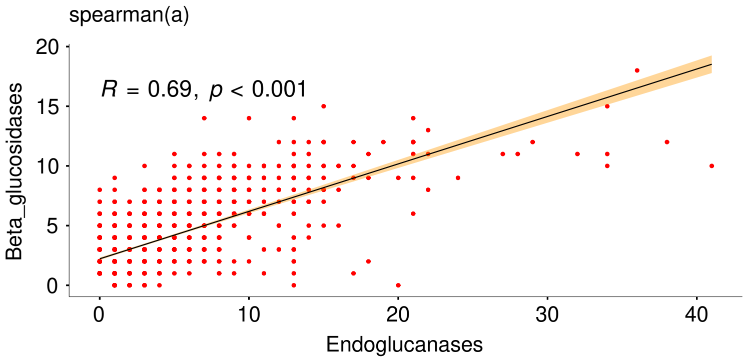
Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id
PRJEB21624	GCA_900315685.1	PRJEB21624	GCA_900315425.1	PRJEB21624	GCA_900320635.1	PRJEB21624	GCA_900314165.1	PRJEB21624	GCA_900315885.1	PRJEB21624	GCA_900317135.1
PRJEB21624	GCA_900313115.1	PRJEB21624	GCA_900318295.1	PRJEB21624	GCA_900319575.1	PRJEB21624	GCA_900320225.1	PRJEB21624	GCA_900315145.1	PRJEB21624	GCA_900319975.1
PRJEB21624	GCA_900318645.1	PRJEB21624	GCA_900314435.1	PRJEB21624	GCA_900319185.1	PRJEB21624	GCA_900319435.1	PRJEB21624	GCA_900313485.1	PRJEB21624	GCA_900313745.1
PRJEB21624	GCA_900314395.1	PRJEB21624	GCA_900316535.1	PRJEB21624	GCA_900314045.1	PRJEB21624	GCA_900314005.1	PRJEB21624	GCA_900314135.1	PRJEB21624	GCA_900319145.1
PRJEB21624	GCA_900313985.1	PRJEB21624	GCA_900313455.1	PRJEB21624	GCA_900315105.1	PRJEB21624	GCA_900315195.1	PRJEB21624	GCA_900318735.1	PRJEB21624	GCA_900316405.1
PRJEB21624	GCA_900313545.1	PRJEB21624	GCA_900315225.1	PRJEB21624	GCA_900320795.1	PRJEB21624	GCA_900315405.1	PRJEB21624	GCA_900316755.1	PRJEB21624	GCA_900314085.1
PRJEB21624	GCA_900315915.1	PRJEB21624	GCA_900322175.1	PRJEB21624	GCA_900318875.1	PRJEB21624	GCA_900318205.1	PRJEB21624	GCA_900317215.1	PRJEB21624	GCA_900317075.1
PRJEB21624	GCA_900320375.1	PRJEB21624	GCA_900317605.1	PRJEB21624	GCA_900314105.1	PRJEB21624	GCA_900320435.1	PRJEB21624	GCA_900320745.1	PRJEB21624	GCA_900319755.1
PRJEB21624	GCA_900314645.1	PRJEB21624	GCA_900314825.1	PRJEB21624	GCA_900320665.1	PRJEB21624	GCA_900318895.1	PRJEB21624	GCA_900317705.1	PRJEB21624	GCA_900318755.1
PRJEB21624	GCA_900318285.1	PRJEB21624	GCA_900321485.1	PRJEB21624	GCA_900321605.1	PRJEB21624	GCA_900314755.1	PRJEB21624	GCA_900320475.1	PRJEB21624	GCA_900318925.1
PRJEB21624	GCA_900315825.1	PRJEB21624	GCA_900318915.1	PRJEB21624	GCA_900316665.1	PRJEB21624	GCA_900314805.1	PRJEB21624	GCA_900313385.1	PRJEB21624	GCA_900319175.1
PRJEB21624	GCA_900318265.1	PRJEB21624	GCA_900316565.1	PRJEB21624	GCA_900318355.1	PRJEB21624	GCA_900318405.1	PRJEB21624	GCA_900320845.1	PRJEB21624	GCA_900313185.1
PRJEB21624	GCA_900313155.1	PRJEB21624	GCA_900321155.1	PRJEB21624	GCA_900317065.1	PRJEB21624	GCA_900319215.1	PRJEB21624	GCA_900317465.1	PRJEB21624	GCA_900319285.1
PRJEB21624	GCA_900321595.1	PRJEB21624	GCA_900314895.1	PRJEB21624	GCA_900314335.1	PRJEB21624	GCA_900320445.1	PRJEB21624	GCA_900317035.1	PRJEB21624	GCA_900320165.1
PRJEB21624	GCA_900316925.1	PRJEB21624	GCA_900314265.1	PRJEB21624	GCA_900317835.1	PRJEB21624	GCA_900317345.1	PRJEB21624	GCA_900313325.1	PRJEB21624	GCA_900316695.1
PRJEB21624	GCA_900321375.1	PRJEB21624	GCA_900321935.1	PRJEB21624	GCA_900320805.1	PRJEB21624	GCA_900317555.1	PRJEB21624	GCA_900317715.1	PRJEB21624	GCA_900318115.1
PRJEB21624	GCA_900319005.1	PRJEB21624	GCA_900313975.1	PRJEB21624	GCA_900316315.1	PRJEB21624	GCA_900315165.1	PRJEB21624	GCA_900319735.1	PRJEB21624	GCA_900315375.1
PRJEB21624	GCA_900318655.1	PRJEB21624	GCA_900320525.1	PRJEB21624	GCA_900317625.1	PRJEB21624	GCA_900318255.1	PRJEB21624	GCA_900318485.1	PRJEB21624	GCA_900319925.1
PRJEB21624	GCA_900314245.1	PRJEB21624	GCA_900318465.1	PRJEB21624	GCA_900317665.1	PRJEB21624	GCA_900319345.1	PRJEB21624	GCA_900317775.1	PRJEB21624	GCA_900315855.1
PRJEB21624	GCA_900314975.1	PRJEB21624	GCA_900315445.1	PRJEB21624	GCA_900320885.1	PRJEB21624	GCA_900320485.1	PRJEB21624	GCA_900320285.1	PRJEB21624	GCA_900320975.1
PRJEB21624	GCA_900317415.1	PRJEB21624	GCA_900319835.1	PRJEB21624	GCA_900318175.1	PRJEB21624	GCA_900316205.1	PRJEB21624	GCA_900315455.1	PRJEB21624	GCA_900313435.1
PRJEB21624	GCA_900322025.1	PRJEB21624	GCA_900313135.1	PRJEB21624	GCA_900315935.1	PRJEB21624	GCA_900319865.1	PRJEB21624	GCA_900313755.1	PRJEB21624	GCA_900317585.1
PRJEB21624	GCA_900319125.1	PRJEB21624	GCA_900320085.1	PRJEB21624	GCA_900317495.1	PRJEB21624	GCA_900320835.1	PRJEB21624	GCA_900316775.1	PRJEB21624	GCA_900314585.1
PRJEB21624	GCA_900314405.1	PRJEB21624	GCA_900320455.1	PRJEB21624	GCA_900319505.1	PRJEB21624	GCA_900314455.1	PRJEB21624	GCA_900317505.1	PRJEB21624	GCA_900321075.1
PRJEB21624	GCA_900319415.1	PRJEB21624	GCA_900314175.1	PRJEB21624	GCA_900319165.1	PRJEB21624	GCA_900313235.1	PRJEB21624	GCA_900313965.1	PRJEB21624	GCA_900314935.1
PRJEB21624	GCA_900317335.1	PRJEB21624	GCA_900317785.1	PRJEB21624	GCA_900321275.1	PRJEB21624	GCA_900319115.1	PRJEB21624	GCA_900318105.1	PRJEB21624	GCA_900320465.1
PRJEB21624	GCA_900318705.1	PRJEB21624	GCA_900315645.1	PRJEB21624	GCA_900315785.1	PRJEB21624	GCA_900314795.1	PRJEB21624	GCA_900315205.1	PRJEB21624	GCA_900319255.1
PRJEB21624	GCA_900319265.1	PRJEB21624	GCA_900315635.1	PRJEB21624	GCA_900320265.1	PRJEB21624	GCA_900321165.1	PRJEB21624	GCA_900315325.1	PRJEB21624	GCA_900318975.1
PRJEB21624	GCA_900316525.1	PRJEB21624	GCA_900313765.1	PRJEB21624	GCA_900322105.1	PRJEB21624	GCA_900317545.1	PRJEB21624	GCA_900315775.1	PRJEB21624	GCA_900319065.1
PRJEB21624	GCA_900316705.1	PRJEB21624	GCA_900321475.1	PRJEB21624	GCA_900319245.1	PRJEB21624	GCA_900318995.1	PRJEB21624	GCA_900318765.1	PRJEB21624	GCA_900321335.1
PRJEB21624	GCA_900320395.1	PRJEB21624	GCA_900314195.1	PRJEB21624	GCA_900317325.1	PRJEB21624	GCA_900319225.1	PRJEB21624	GCA_900313815.1	PRJEB21624	GCA_900318235.1
PRJEB21624	GCA_900322035.1	PRJEB21624	GCA_900317455.1	PRJEB21624	GCA_900314355.1	PRJEB21624	GCA_900316845.1	PRJEB21624	GCA_900316115.1	PRJEB21624	GCA_900314515.1
PRJEB21624	GCA_900319895.1	PRJEB21624	GCA_900320115.1	PRJEB21624	GCA_900318185.1	PRJEB21624	GCA_900316805.1	PRJEB21624	GCA_900314625.1	PRJEB21624	GCA_900320725.1
PRJEB21624	GCA_900318545.1	PRJEB21624	GCA_900315595.1	PRJEB21624	GCA_900318585.1	PRJEB21624	GCA_900317405.1	PRJEB21624	GCA_900321385.1	PRJEB21624	GCA_900313305.1
PRJEB21624	GCA_900313105.1	PRJEB21624	GCA_900320075.1	PRJEB21624	GCA_900315655.1	PRJEB21624	GCA_900313275.1	PRJEB21624	GCA_900321785.1	PRJEB21624	GCA_900316375.1
PRJEB21624	GCA_900314835.1	PRJEB21624	GCA_900318715.1	PRJEB21624	GCA_900319815.1	PRJEB21624	GCA_900319105.1	PRJEB21624	GCA_900315185.1	PRJEB21624	GCA_900315875.1
PRJEB21624	GCA_900318325.1	PRJEB21624	GCA_900315245.1	PRJEB21624	GCA_900320825.1	PRJEB21624	GCA_900315625.1	PRJEB21624	GCA_900317145.1	PRJEB21624	GCA_900313805.1
PRJEB21624	GCA_900313355.1	PRJEB21624	GCA_900314555.1	PRJEB21624	GCA_900319645.1	PRJEB21624	GCA_900314845.1	PRJEB21624	GCA_900316885.1	PRJEB21624	GCA_900321535.1
PRJEB21624	GCA_900322155.1	PRJEB21624	GCA_900313365.1	PRJEB21624	GCA_900317675.1	PRJEB21624	GCA_900318595.1	PRJEB21624	GCA_900313425.1	PRJEB21624	GCA_900321645.1
PRJEB21624	GCA_900318825.1	PRJEB21624	GCA_900320305.1	PRJEB21624	GCA_900314275.1	PRJEB21624	GCA_900318495.1	PRJEB21624	GCA_900319605.1	PRJEB21624	GCA_900314925.1
PRJEB21624	GCA_900317255.1	PRJEB21624	GCA_900316895.1	PRJEB21624	GCA_900313255.1	PRJEB21624	GCA_900316505.1	PRJEB21624	GCA_900320855.1	PRJEB21624	GCA_900313505.1
PRJEB21624	GCA_900314545.1	PRJEB21624	GCA_900317025.1	PRJEB21624	GCA_900317485.1	PRJEB21624	GCA_900316485.1	PRJEB21624	GCA_900320135.1	PRJEB21624	GCA_900315495.1
PRJEB21624	GCA_900316605.1	PRJEB21624	GCA_900319375.1	PRJEB21624	GCA_900315605.1	PRJEB21624	GCA_900320045.1	PRJEB21624	GCA_900318245.1	PRJEB21624	GCA_900319855.1
PRJEB21624	GCA_900314915.1	PRJEB21624	GCA_900314035.1	PRJEB21624	GCA_900316745.1	PRJEB21624	GCA_900319095.1	PRJEB21624	GCA_900318515.1	PRJEB21624	GCA_900318805.1
PRJEB21624	GCA_900317955.1	PRJEB21624	GCA_900315505.1	PRJEB21624	GCA_900316545.1	PRJEB21624	GCA_900313945.1	PRJEB21624	GCA_900320105.1	PRJEB21624	GCA_900315585.1

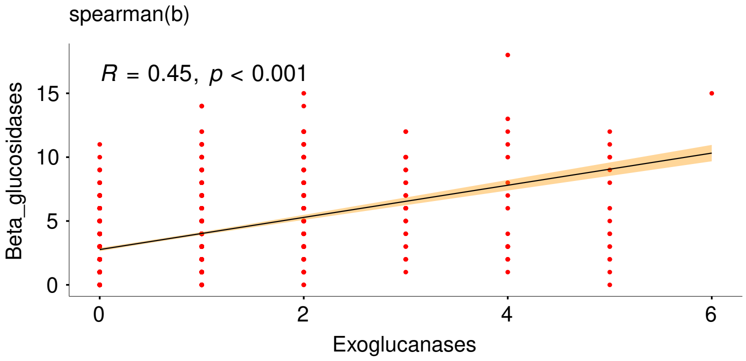
Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id
PRJEB21624	GCA_900318505.1	PRJEB21624	GCA_900317885.1	PRJEB21624	GCA_900316265.1	PRJEB21624	GCA_900315295.1	PRJEB21624	GCA_900314325.1	PRJEB21624	GCA_900317385.1
PRJEB21624	GCA_900316675.1	PRJEB21624	GCA_900318945.1	PRJEB21624	GCA_900313905.1	PRJEB21624	GCA_900317985.1	PRJEB21624	GCA_900321835.1	PRJEB21624	GCA_900318125.1
PRJEB21624	GCA_900318375.1	PRJEB21624	GCA_900315705.1	PRJEB21624	GCA_900313085.1	PRJEB21624	GCA_900317355.1	PRJEB21624	GCA_900317425.1	PRJEB21624	GCA_900315155.1
PRJEB21624	GCA_900317195.1	PRJEB21624	GCA_900314425.1	PRJEB21624	GCA_900318095.1	PRJEB21624	GCA_900318215.1	PRJEB21624	GCA_900319745.1	PRJEB21624	GCA_900314855.1
PRJEB21624	GCA_900319335.1	PRJEB21624	GCA_900313165.1	PRJEB21624	GCA_900316585.1	PRJEB21624	GCA_900317125.1	PRJEB21624	GCA_900321405.1	PRJEB21624	GCA_900319035.1
PRJEB21624	GCA_900320145.1	PRJEB21624	GCA_900321215.1	PRJEB21624	GCA_900320055.1	PRJEB21624	GCA_900315265.1	PRJEB21624	GCA_900316135.1	PRJEB21624	GCA_900319305.1
PRJEB21624	GCA_900320405.1	PRJEB21624	GCA_900314305.1	PRJEB21624	GCA_900320865.1	PRJEB21624	GCA_900316865.1	PRJEB21624	GCA_900321875.1	PRJEB21624	GCA_900315385.1
PRJEB21624	GCA_900320545.1	PRJEB21624	GCA_900316725.1	PRJEB21624	GCA_900314495.1	PRJEB21624	GCA_900313245.1	PRJEB21624	GCA_900319555.1	PRJEB21624	GCA_900319055.1
PRJEB21624	GCA_900321705.1	PRJEB21624	GCA_900316495.1	PRJEB21624	GCA_900321225.1	PRJEB21624	GCA_900313125.1	PRJEB21624	GCA_900320575.1	PRJEB21624	GCA_900320905.1
PRJEB21624	GCA_900317155.1	PRJEB21624	GCA_900313225.1	PRJEB21624	GCA_900321725.1	PRJEB21624	GCA_900321305.1	PRJEB21624	GCA_900314665.1	PRJEB21624	GCA_900320925.1
PRJEB21624	GCA_900316275.1	PRJEB21624	GCA_900320065.1	PRJEB21624	GCA_900320565.1	PRJEB21624	GCA_900316455.1	PRJEB21624	GCA_900317015.1	PRJEB21624	GCA_900315835.1
PRJEB21624	GCA_900321115.1	PRJEB21624	GCA_900316305.1	PRJEB21624	GCA_900317115.1	PRJEB21624	GCA_900321085.1	PRJEB21624	GCA_900319315.1	PRJEB21624	GCA_900319805.1
PRJEB21624	GCA_900315395.1	PRJEB21624	GCA_900317975.1	PRJEB21624	GCA_900320425.1	PRJEB21624	GCA_900320585.1	PRJEB21624	GCA_900316425.1	PRJEB21624	GCA_900313635.1
PRJEB21624	GCA_900318885.1	PRJEB21624	GCA_900320315.1	PRJEB21624	GCA_900320345.1	PRJEB21624	GCA_900322055.1	PRJEB21624	GCA_900316765.1	PRJEB21624	GCA_900313845.1
PRJEB21624	GCA_900320015.1	PRJEB21624	GCA_900318675.1	PRJEB21624	GCA_900317865.1	PRJEB21624	GCA_900317735.1	PRJEB21624	GCA_900314565.1	PRJEB21624	GCA_900319295.1
PRJEB21624	GCA_900319235.1	PRJEB21624	GCA_900318955.1	PRJEB21624	GCA_900319845.1	PRJEB21624	GCA_900318145.1	PRJEB21624	GCA_900318575.1	PRJEB21624	GCA_900314225.1
PRJEB21624	GCA_900319985.1	PRJEB21624	GCA_900314385.1	PRJEB21624	GCA_900317615.1	PRJEB21624	GCA_900320765.1	PRJEB21624	GCA_900316325.1	PRJEB21624	GCA_900315515.1
PRJEB21624	GCA_900313775.1	PRJEB21624	GCA_900320655.1	PRJEB21624	GCA_900318835.1	PRJEB21624	GCA_900317645.1	PRJEB21624	GCA_900317575.1	PRJEB21624	GCA_900321235.1
PRJEB21624	GCA_900318365.1	PRJEB21624	GCA_900314125.1	PRJEB21624	GCA_900321985.1	PRJEB21624	GCA_900313705.1	PRJEB21624	GCA_900316595.1	PRJEB21624	GCA_900319565.1
PRJEB21624	GCA_900318395.1	PRJEB21624	GCA_900321585.1	PRJEB21624	GCA_900318855.1	PRJEB21624	GCA_900321315.1	PRJEB21624	GCA_900318625.1	PRJEB21624	GCA_900316045.1
PRJEB21624	GCA_900314285.1	PRJEB21624	GCA_900314995.1	PRJEB21624	GCA_900313855.1	PRJEB21624	GCA_900315015.1	PRJEB21624	GCA_900319025.1	PRJEB21624	GCA_900317825.1
PRJEB21624	GCA_900321505.1	PRJEB21624	GCA_900319965.1	PRJEB21624	GCA_900315025.1	PRJEB21624	GCA_900318525.1	PRJEB21624	GCA_900313565.1	PRJEB21624	GCA_900313215.1
PRJEB21624	GCA_900316145.1	PRJEB21624	GCA_900316825.1	PRJEB21624	GCA_900318725.1	PRJEB21624	GCA_900318425.1	PRJEB21624	GCA_900315255.1	PRJEB21624	GCA_900321185.1
PRJEB21624	GCA_900320895.1	PRJEB21624	GCA_900321345.1	PRJEB21624	GCA_900314535.1	PRJEB21624	GCA_900318075.1	PRJEB21624	GCA_900315285.1	PRJEB21624	GCA_900317845.1
PRJEB21624	GCA_900321415.1	PRJEB21624	GCA_900315435.1	PRJEB21624	GCA_900317205.1	PRJEB21624	GCA_900320125.1	PRJEB21624	GCA_900318315.1	PRJEB21624	GCA_900318665.1
PRJEB21624	GCA_900314785.1	PRJEB21624	GCA_900322195.1	PRJEB21624	GCA_900314375.1	PRJEB21624	GCA_900322135.1	PRJEB21624	GCA_900320155.1	PRJEB21624	GCA_900314255.1
PRJEB21624	GCA_900313575.1	PRJEB21624	GCA_900316355.1	PRJEB21624	GCA_900320025.1	PRJEB21624	GCA_900315745.1	PRJEB21624	GCA_900315005.1	PRJEB21624	GCA_900313175.1
PRJEB21624	GCA_900320385.1	PRJEB21624	GCA_900319405.1	PRJEB21624	GCA_900318305.1	PRJEB21624	GCA_900320325.1	PRJEB21624	GCA_900318085.1	PRJEB21624	GCA_900315115.1
PRJEB21624	GCA_900321635.1	PRJEB21624	GCA_900314875.1	PRJEB21624	GCA_900320955.1	PRJEB21624	GCA_900313725.1	PRJEB21624	GCA_900314295.1	PRJEB21624	GCA_900316655.1
PRJEB21624	GCA_900313885.1	PRJEB21624	GCA_900313915.1	PRJEB21624	GCA_900317375.1	PRJEB21624	GCA_900316715.1	PRJEB21624	GCA_900319775.1	PRJEB21624	GCA_900316125.1
PRJEB21624	GCA_900319275.1	PRJEB21624	GCA_900318905.1	PRJEB21624	GCA_900319655.1	PRJEB21624	GCA_900320235.1	PRJEB21624	GCA_900320695.1	PRJEB21624	GCA_900316965.1
PRJEB21624	GCA_900313515.1	PRJEB21624	GCA_900313075.1	PRJEB21624	GCA_900316955.1	PRJEB21624	GCA_900314905.1	PRJEB21624	GCA_900315965.1	PRJEB21624	GCA_900320555.1
PRJEB21624	GCA_900321755.1	PRJEB21624	GCA_900319945.1	PRJEB21624	GCA_900315315.1	PRJEB21624	GCA_900320275.1	PRJEB21624	GCA_900320415.1	PRJEB21624	GCA_900313375.1
PRJEB21624	GCA_900316165.1	PRJEB21624	GCA_900314605.1	PRJEB21624	GCA_900314315.1	PRJEB21624	GCA_900313735.1	PRJEB21624	GCA_900318345.1	PRJEB21624	GCA_900318635.1
PRJEB21624	GCA_900321945.1	PRJEB21624	GCA_900314675.1	PRJEB21624	GCA_900313315.1	PRJEB21624	GCA_900318385.1	PRJEB21624	GCA_900319585.1	PRJEB21624	GCA_900315925.1
PRJEB21624	GCA_900318845.1	PRJEB21624	GCA_900321255.1	PRJEB21624	GCA_900317805.1	PRJEB21624	GCA_900317275.1	PRJEB21624	GCA_900321905.1	PRJEB21624	GCA_900318605.1
PRJEB21624	GCA_900315905.1	PRJEB21624	GCA_900319515.1	PRJEB21624	GCA_900319665.1	PRJEB21624	GCA_900320985.1	PRJEB21624	GCA_900316245.1	PRJEB21624	GCA_900314415.1
PRJEB21624	GCA_900314185.1	PRJEB21624	GCA_900319795.1	PRJEB21624	GCA_900317905.1	PRJEB21624	GCA_900321055.1	PRJEB21624	GCA_900319325.1	PRJEB21624	GCA_900320175.1
PRJEB21624	GCA_900316015.1	PRJEB21624	GCA_900321915.1	PRJEB21624	GCA_900321025.1	PRJEB21624	GCA_900317185.1	PRJEB21624	GCA_900321525.1	PRJEB21624	GCA_900314115.1
PRJEB21624	GCA_900313675.1	PRJEB21624	GCA_900313095.1	PRJEB21624	GCA_900316905.1	PRJEB21624	GCA_900320815.1	PRJEB21624	GCA_900313445.1	PRJEB21624	GCA_900315555.1
PRJEB21624	GCA_900313295.1	PRJEB21624	GCA_900319825.1	PRJEB21624	GCA_900320205.1	PRJEB21624	GCA_900318165.1	PRJEB21624	GCA_900315365.1	PRJEB21624	GCA_900313495.1
PRJEB21624	GCA_900321575.1	PRJEB21624	GCA_900315045.1	PRJEB21624	GCA_900314715.1	PRJEB21624	GCA_900314465.1	PRJEB21624	GCA_900321615.1	PRJEB21624	GCA_900320945.1
PRJEB21624	GCA_900321125.1	PRJEB21624	GCA_900321955.1	PRJEB21624	GCA_900321145.1	PRJEB21624	GCA_900317875.1	PRJEB21624	GCA_900320495.1	PRJEB21624	GCA_900313535.1
PRJEB21624	GCA_900318335.1	PRJEB21624	GCA_900321565.1	PRJEB21624	GCA_900315945.1	PRJEB21624	GCA_900316235.1	PRJEB21624	GCA_900317895.1	PRJEB21624	GCA_900316095.1
PRJEB21624	GCA_900314705.1	PRJEB21624	GCA_900320505.1	PRJEB21624	GCA_900319625.1	PRJEB21624	GCA_900318025.1	PRJEB21624	GCA_900319485.1	PRJEB21624	GCA_900316625.1

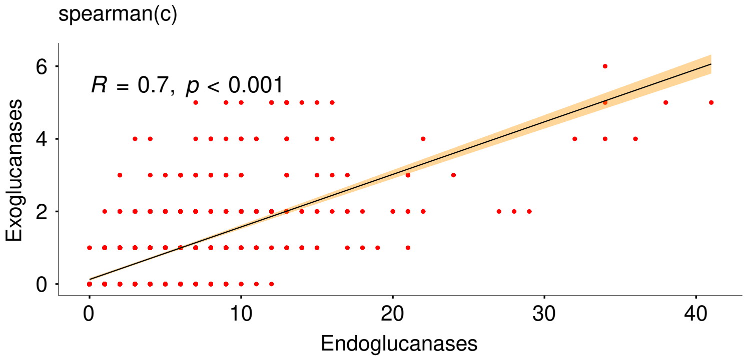
Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id
PRJEB21624	GCA_900322075.1	PRJEB21624	GCA_900319355.1	PRJEB21624	GCA_900315485.1	PRJEB21624	GCA_900316875.1	PRJEB21624	GCA_900321035.1	PRJNA471733	GCA_000622005.1
PRJEB21624	GCA_900313825.1	PRJEB21624	GCA_900321845.1	PRJEB21624	GCA_900321065.1	PRJEB21624	GCA_900316465.1	PRJEB21624	GCA_900320035.1	PRJNA471733	GCA_003149245.1
PRJEB21624	GCA_900313685.1	PRJEB21624	GCA_900315695.1	PRJEB21624	GCA_900316615.1	PRJEB21624	GCA_900313555.1	PRJEB21624	GCA_900319715.1	PRJNA471733	GCA_000712055.1
PRJEB21624	GCA_900317855.1	PRJEB21624	GCA_900316395.1	PRJEB21624	GCA_900315465.1	PRJEB21624	GCA_900318055.1	PRJEB21624	GCA_900315765.1	PRJNA471733	GCA_000964315.1
PRJEB21624	GCA_900321325.1	PRJEB21624	GCA_900316225.1	PRJEB21624	GCA_900320875.1	PRJEB21624	GCA_900313995.1	PRJEB21624	GCA_900316685.1	PRJNA471733	GCA_000703145.1
PRJEB21624	GCA_900313925.1	PRJEB21624	GCA_900313465.1	PRJEB21624	GCA_900318775.1	PRJEB21624	GCA_900315065.1	PRJEB21624	GCA_900319365.1	PRJNA471733	GCA_000702545.1
PRJEB21624	GCA_900313955.1	PRJEB21624	GCA_900314735.1	PRJEB21624	GCA_900318435.1	PRJEB21624	GCA_900321425.1	PRJEB21624	GCA_900313195.1	PRJNA471733	GCA_000702885.1
PRJEB21624	GCA_900314595.1	PRJEB21624	GCA_900319915.1	PRJEB21624	GCA_900315995.1	PRJEB21624	GCA_900321855.1	PRJEB21624	GCA_900315675.1	PRJNA471733	GCA_000421405.1
PRJEB21624	GCA_900317695.1	PRJEB21624	GCA_900314635.1	PRJEB21624	GCA_900321765.1	PRJEB21624	GCA_900317395.1	PRJEB22623	GCA_900230525.1	PRJNA471733	GCA_000622085.1
PRJEB21624	GCA_900314445.1	PRJEB21624	GCA_900321655.1	PRJEB21624	GCA_900317915.1	PRJEB21624	GCA_900314065.1	PRJEB22623	GCA_900230505.1	PRJNA471733	GCA_000424145.1
PRJEB21624	GCA_900318555.1	PRJEB21624	GCA_900316345.1	PRJEB21624	GCA_900317745.1	PRJEB21624	GCA_900318135.1	PRJEB22623	GCA_900230515.1	PRJNA471733	GCA_000621845.1
PRJEB21624	GCA_900317945.1	PRJEB21624	GCA_900321195.1	PRJEB21624	GCA_900322115.1	PRJEB21624	GCA_900318535.1	PRJEB22623	GCA_900230545.1	PRJNA471733	GCA_000701905.1
PRJEB21624	GCA_900315725.1	PRJEB21624	GCA_900316215.1	PRJEB21624	GCA_900315345.1	PRJEB21624	GCA_900319905.1	PRJEB22623	GCA_900230535.1	PRJNA471733	GCA_000420865.1
PRJEB21624	GCA_900316735.1	PRJEB21624	GCA_900314485.1	PRJEB21624	GCA_900316005.1	PRJEB21624	GCA_900321805.1	PRJNA202380	GCA_001684375.1	PRJNA471733	GCA_000424105.1
PRJEB21624	GCA_900316915.1	PRJEB21624	GCA_900321965.1	PRJEB21624	GCA_900322145.1	PRJEB21624	GCA_900321825.1	PRJEB12797	GCA_900089695.1	PRJNA471733	GCA_000702845.1
PRJEB21624	GCA_900320625.1	PRJEB21624	GCA_900318795.1	PRJEB21624	GCA_900320595.1	PRJEB21624	GCA_900321795.1	PRJEB12797	GCA_900089675.1	PRJNA471733	GCA_000420825.1
PRJEB21624	GCA_900313265.1	PRJEB21624	GCA_900319615.1	PRJEB21624	GCA_900319535.1	PRJEB21624	GCA_900314075.1	PRJEB12797	GCA_900089725.1	PRJNA471733	GCA_000424605.1
PRJEB21624	GCA_900320365.1	PRJEB21624	GCA_900313665.1	PRJEB21624	GCA_900319085.1	PRJEB21624	GCA_900315125.1	PRJEB12797	GCA_900089705.1	PRJNA471733	GCA_000621825.1
PRJEB21624	GCA_900319935.1	PRJEB21624	GCA_900315545.1	PRJEB21624	GCA_900313615.1	PRJEB21624	GCA_900316085.1	PRJEB12797	GCA_900089715.1	PRJNA471733	GCA_000701725.1
PRJEB21624	GCA_900317005.1	PRJEB21624	GCA_900314095.1	PRJEB21624	GCA_900316445.1	PRJEB21624	GCA_900317225.1	PRJEB12797	GCA_900089685.1	PRJNA471733	GCA_000526795.1
PRJEB21624	GCA_900320515.1	PRJEB21624	GCA_900317365.1	PRJEB21624	GCA_900320705.1	PRJEB21624	GCA_900318785.1	PRJNA319009	GCA_002631825.1	PRJNA471733	GCA_000424005.1
PRJEB21624	GCA_900314475.1	PRJEB21624	GCA_900316385.1	PRJEB21624	GCA_900314155.1	PRJEB21624	GCA_900318865.1	PRJEB22504	GCA_900215825.1	PRJNA471733	GCA_003149105.1
PRJEB21624	GCA_900321135.1	PRJEB21624	GCA_900317285.1	PRJEB21624	GCA_900320615.1	PRJEB21624	GCA_900315575.1	PRJNA471733	GCA_000702105.1	PRJNA471733	GCA_000686945.1
PRJEB21624	GCA_900322125.1	PRJEB21624	GCA_900314365.1	PRJEB21624	GCA_900321465.1	PRJEB21624	GCA_900316975.1	PRJNA471733	GCA_000621765.1	PRJNA471733	GCA_000420965.1
PRJEB21624	GCA_900317765.1	PRJEB21624	GCA_900319705.1	PRJEB21624	GCA_900321975.1	PRJEB21624	GCA_900316985.1	PRJNA471733	GCA_000701765.1	PRJNA471733	GCA_000701565.1
PRJEB21624	GCA_900314885.1	PRJEB21624	GCA_900321815.1	PRJEB21624	GCA_900315715.1	PRJEB21624	GCA_900317995.1	PRJNA471733	GCA_003149165.1	PRJNA471733	GCA_000702185.1
PRJEB21624	GCA_900314815.1	PRJEB21624	GCA_900315215.1	PRJEB21624	GCA_900321105.1	PRJEB21624	GCA_900316435.1	PRJNA471733	GCA_000621585.1	PRJNA471733	GCA_000420985.1
PRJEB21624	GCA_900319765.1	PRJEB21624	GCA_900321675.1	PRJEB21624	GCA_900320965.1	PRJEB21624	GCA_900321995.1	PRJNA471733	GCA_000423965.1	PRJNA471733	GCA_000701645.1
PRJEB21624	GCA_900319135.1	PRJEB21624	GCA_900319015.1	PRJEB21624	GCA_900319455.1	PRJEB21624	GCA_900316255.1	PRJNA471733	GCA_000424265.1	PRJNA471733	GCA_000619985.1
PRJEB21624	GCA_900314505.1	PRJEB21624	GCA_900313415.1	PRJEB21624	GCA_900315565.1	PRJEB21624	GCA_900320095.1	PRJNA471733	GCA_000702005.1	PRJNA471733	GCA_000621785.1
PRJEB21624	GCA_900321265.1	PRJEB21624	GCA_900318005.1	PRJEB21624	GCA_900313865.1	PRJEB21624	GCA_900314235.1	PRJNA471733	GCA_000526735.1	PRJNA471733	GCA_000424405.1
PRJEB21624	GCA_900314775.1	PRJEB21624	GCA_900316785.1	PRJEB21624	GCA_900321865.1	PRJEB21624	GCA_900321095.1	PRJNA471733	GCA_000421425.1	PRJNA471733	GCA_000702225.1
PRJEB21624	GCA_900315665.1	PRJEB21624	GCA_900313645.1	PRJEB21624	GCA_900319955.1	PRJEB21624	GCA_900318415.1	PRJNA471733	GCA_000421045.1	PRJNA471733	GCA_000621905.1
PRJEB21624	GCA_900320295.1	PRJEB21624	GCA_900315955.1	PRJEB21624	GCA_900320245.1	PRJEB21624	GCA_900315815.1	PRJNA471733	GCA_000424305.1	PRJNA471733	GCA_000424465.1
PRJEB21624	GCA_900313785.1	PRJEB21624	GCA_900317175.1	PRJEB21624	GCA_900322015.1	PRJEB21624	GCA_900313395.1	PRJNA471733	GCA_000424185.1	PRJNA471733	GCA_000688035.1
PRJEB21624	GCA_900315735.1	PRJEB21624	GCA_900316475.1	PRJEB21624	GCA_900320255.1	PRJEB21624	GCA_900315175.1	PRJNA471733	GCA_000702045.1	PRJNA471733	GCA_000701965.1
PRJEB21624	GCA_900317685.1	PRJEB21624	GCA_900314745.1	PRJEB21624	GCA_900315135.1	PRJEB21624	GCA_900313335.1	PRJNA471733	GCA_000702125.1	PRJNA471733	GCA_000687975.1
PRJEB21624	GCA_900317265.1	PRJEB21624	GCA_900313795.1	PRJEB21624	GCA_900321295.1	PRJEB21624	GCA_900316025.1	PRJNA471733	GCA_000688015.1	PRJNA471733	GCA_000424085.1
PRJEB21624	GCA_900321715.1	PRJEB21624	GCA_900317315.1	PRJEB21624	GCA_900316945.1	PRJEB21624	GCA_900320335.1	PRJNA471733	GCA_000526775.1	PRJNA471733	GCA_000703065.1
PRJEB21624	GCA_900319475.1	PRJEB21624	GCA_900320645.1	PRJEB21624	GCA_900314945.1	PRJEB21624	GCA_900320785.1	PRJNA471733	GCA_000687595.1	PRJNA471733	GCA_000687675.1
PRJEB21624	GCA_900318965.1	PRJEB21624	GCA_900317515.1	PRJEB21624	GCA_900314865.1	PRJEB21624	GCA_900317305.1	PRJNA471733	GCA_000702165.1	PRJNA471733	GCA_000421345.1
PRJEB21624	GCA_900318195.1	PRJEB21624	GCA_900321775.1	PRJEB21624	GCA_900322085.1	PRJEB21624	GCA_900313655.1	PRJNA471733	GCA_000621745.1	PRJNA471733	GCA_000421005.1
PRJEB21624	GCA_900315475.1	PRJEB21624	GCA_900313625.1	PRJEB21624	GCA_900314765.1	PRJEB21624	GCA_900314055.1	PRJNA471733	GCA_000701945.1	PRJNA471733	GCA_000702525.1
PRJEB21624	GCA_900317795.1	PRJEB21624	GCA_900313525.1	PRJEB21624	GCA_900315305.1	PRJEB21624	GCA_900315535.1	PRJNA471733	GCA_000424565.1	PRJNA471733	GCA_000424245.1
PRJEB21624	GCA_900316635.1	PRJEB21624	GCA_900321555.1	PRJEB21624	GCA_900317815.1	PRJEB21624	GCA_900316555.1	PRJNA471733	GCA_000701985.1	PRJNA471733	GCA_000703165.1

Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id	Project_id	GCA_id
PRJNA471733	GCA_000701545.1	PRJNA471733	GCA_000424585.1	PRJNA471733	GCA_000622045.1	PRJNA304217	GCA_001699935.1				
PRJNA471733	GCA_000702685.1	PRJNA471733	GCA_000424325.1	PRJNA471733	GCA_003149285.1	PRJNA173069	GCA_000315315.2				
PRJNA471733	GCA_003149065.1	PRJNA471733	GCA_000702965.1	PRJNA471733	GCA_000703025.1	PRJNA41493	GCA_000205965.1				
PRJNA471733	GCA_000420885.1	PRJNA471733	GCA_000424485.1	PRJNA471733	GCA_000711215.1	PRJNA246767	GCA_000691925.1				
PRJNA471733	GCA_000424045.1	PRJNA471733	GCA_000687515.1	PRJNA471733	GCA_000621605.1	PRJNA246767	GCA_000691905.1				
PRJNA471733	GCA_000423925.1	PRJNA471733	GCA_000622125.1	PRJNA471733	GCA_000687575.1	PRJNA257236	GCA_000753695.1				
PRJNA471733	GCA_000424505.1	PRJNA471733	GCA_000526935.1	PRJNA471733	GCA_000424285.1	PRJNA257236	GCA_000753655.1				
PRJNA471733	GCA_000687635.1	PRJNA471733	GCA_000703125.1	PRJNA471733	GCA_000424165.1	PRJNA257236	GCA_000753675.1				
PRJNA471733	GCA_000702905.1	PRJNA471733	GCA_000621565.1	PRJNA471733	GCA_000423945.1	PRJNA167559	GCA_000390105.1				
PRJNA471733	GCA_000687655.1	PRJNA471733	GCA_000420845.1	PRJNA471733	GCA_000621885.1	PRJNA193663	GCA_000448325.1				
PRJNA471733	GCA_000420905.1	PRJNA471733	GCA_000619945.1	PRJNA471733	GCA_000686065.1	PRJNA193664	GCA_000448285.1				
PRJNA471733	GCA_000424065.1	PRJNA471733	GCA_000619965.1	PRJNA471733	GCA_000702285.1	PRJNA193665	GCA_000448305.1				
PRJNA471733	GCA_000424545.1	PRJNA471733	GCA_000702025.1	PRJDB4067	GCA_002003105.1	PRJNA343431	GCA_001875625.1				
PRJNA471733	GCA_000621545.1	PRJNA471733	GCA_000685945.1	PRJDB4067	GCA_002003085.1	PRJNA17635	GCA_000207885.1				
PRJNA471733	GCA_000424205.1	PRJNA471733	GCA_000702085.1	PRJNA73603	GCA_000234475.1	PRJNA184676	GCA_000338435.1				
PRJNA471733	GCA_000702145.1	PRJNA471733	GCA_000424025.1	PRJNA158439	GCA_000401875.1	PRJNA184676	GCA_000338355.1				
PRJNA471733	GCA_000619925.1	PRJNA471733	GCA_000424345.1	PRJNA158439	GCA_000401855.1	PRJNA17633	GCA_000207905.1				
PRJNA471733	GCA_000424525.1	PRJNA471733	GCA_000702065.1	PRJNA158439	GCA_000401775.1	PRJNA38045	GCA_000205605.2				
PRJNA471733	GCA_000712065.1	PRJNA471733	GCA_000702205.1	PRJNA158439	GCA_000401655.1	PRJNA18537	GCA_000205145.2				
PRJNA471733	GCA_000701925.1	PRJNA471733	GCA_000687995.1	PRJNA158439	GCA_000401835.1	PRJNA29437	GCA_001857785.1				
PRJNA471733	GCA_000622105.1	PRJNA471733	GCA_000424385.1	PRJNA158439	GCA_000401815.1	PRJNA29437	GCA_001857765.1				
PRJNA471733	GCA_000424445.1	PRJNA471733	GCA_000526755.1	PRJNA158439	GCA_000401795.1	PRJNA266890	GCA_001029945.1				
PRJNA471733	GCA_000621405.1	PRJNA471733	GCA_000701885.1	PRJNA158439	GCA_000401895.1						
PRJNA471733	GCA_003149325.1	PRJNA471733	GCA_000746155.1	PRJNA226580	GCA_000701245.1						
PRJNA471733	GCA_000686965.1	PRJNA471733	GCA_000701605.1	PRJNA226580	GCA_000701185.1						
PRJNA471733	GCA_000420945.1	PRJNA471733	GCA_000702865.1	PRJNA226580	GCA_000701225.1						
PRJNA471733	GCA_000622065.1	PRJNA471733	GCA_000702265.1	PRJNA226580	GCA_000701205.1						
PRJNA471733	GCA_000424225.1	PRJNA471733	GCA_000746015.1	PRJNA231836	GCA_000732065.1						
PRJNA471733	GCA_000687715.1	PRJNA471733	GCA_000621465.1	PRJNA231836	GCA_000732265.1						
PRJNA471733	GCA_000621965.1	PRJNA471733	GCA_000421365.1	PRJNA231836	GCA_000732285.1						
PRJNA471733	GCA_000621945.1	PRJNA471733	GCA_000424425.1	PRJNA312682	GCA_003010405.1						
PRJNA471733	GCA_000621985.1	PRJNA471733	GCA_000423985.1	PRJNA255238	GCA_000753615.1						
PRJNA471733	GCA_000424365.1	PRJNA471733	GCA_000622025.1	PRJNA255460	GCA_001468665.1						
PRJNA471733	GCA_000703005.1	PRJNA471733	GCA_000687955.1	PRJNA255460	GCA_001468655.1						
PRJNA471733	GCA_000621865.1	PRJNA471733	GCA_003149115.1	PRJNA255460	GCA_001468675.1						
PRJNA471733	GCA_000621925.1	PRJNA471733	GCA_000686125.1	PRJNA255460	GCA_001468605.1						
PRJNA471733	GCA_000687695.1	PRJNA471733	GCA_000686145.1	PRJNA255460	GCA_001468585.1						
PRJNA471733	GCA_000687555.1	PRJNA471733	GCA_000702245.1	PRJNA255460	GCA_001468595.1						
PRJNA471733	GCA_000621805.1	PRJNA471733	GCA_000687615.1	PRJNA255460	GCA_001468645.1						
PRJNA471733	GCA_000621425.1	PRJNA471733	GCA_000703185.1	PRJNA255460	GCA_001468715.1						
PRJNA471733	GCA_000688375.1	PRJNA471733	GCA_000420925.1	PRJNA209200	GCA_000691425.1						
PRJNA471733	GCA_000687495.1	PRJNA471733	GCA_000526815.1	PRJNA281117	GCA_001586955.1						
PRJNA471733	GCA_000421025.1	PRJNA471733	GCA_000619825.1	PRJNA304217	GCA_001699965.1						
PRJNA471733	GCA_000702825.1	PRJNA471733	GCA_000421385.1	PRJNA304217	GCA_001699955.1						
PRJNA471733	GCA_000424125.1	PRJNA471733	GCA_000701625.1	PRJNA304217	GCA_001699945.1						

Supplementary_File_4 Spearman analysis among three cellulase categories. The spearman analysis is to examine the monotone relations between each pair of cellulase categories. In a monotonic relationship, two variables increase or decrease toward the same direction but may not exactly happen at a constant rate. The range of R represents the relationship between two variables. 0.8-1.0, extremely correlated; 0.6-0.8, strongly correlated; 0.4-0.6, middle correlated; 0.2-0.4, weakly correlated; 0.0-0.2, extremely weakly or no correlated.

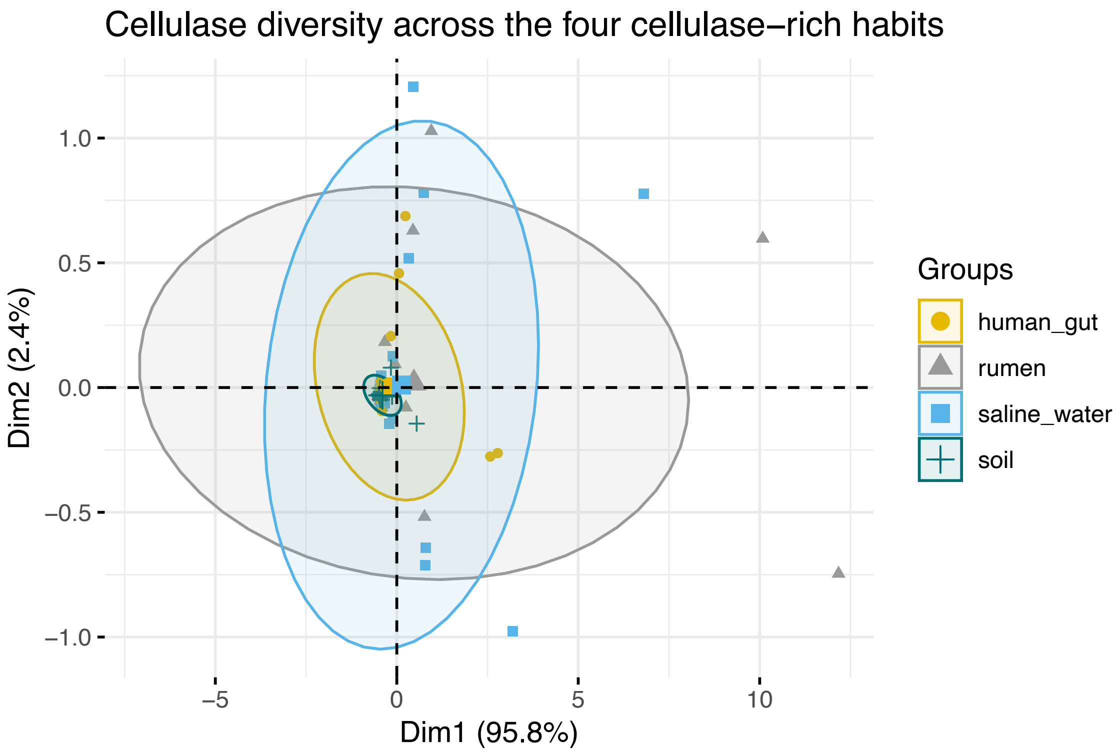






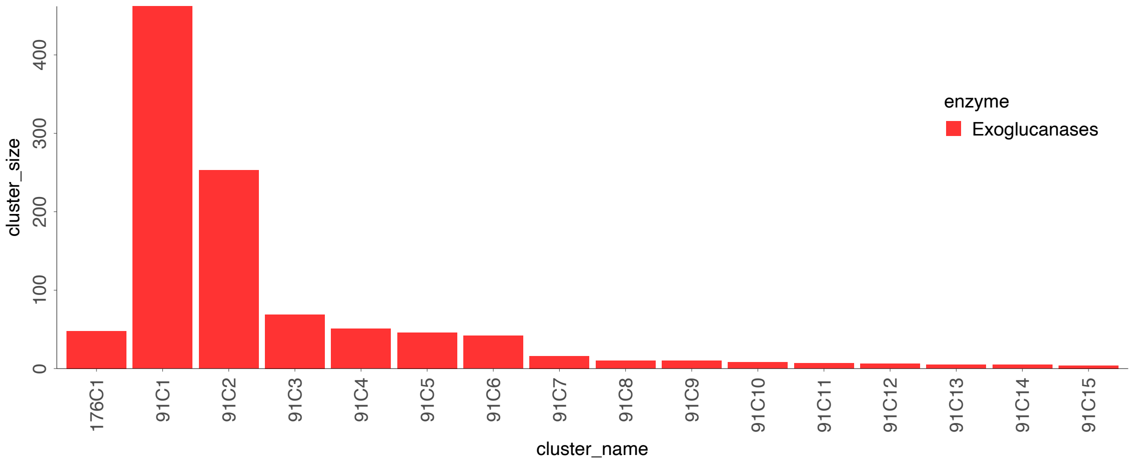
Supplementary_File_5 The principal component analysis (PCA) analysis of cellulases among rumen, saline water, human gut, and soil.

The PCA of the cellulases across the 1,515 metagenomes among rumen, saline water, human gut, and soil. X and Y-axis show principal component 1 and principal component 2, which explain 95.8% and 2.4% of the total variance, respectively. Prediction ellipses are such that with the probability of 0.95, a new observation from the same group will fall inside the ellipse.

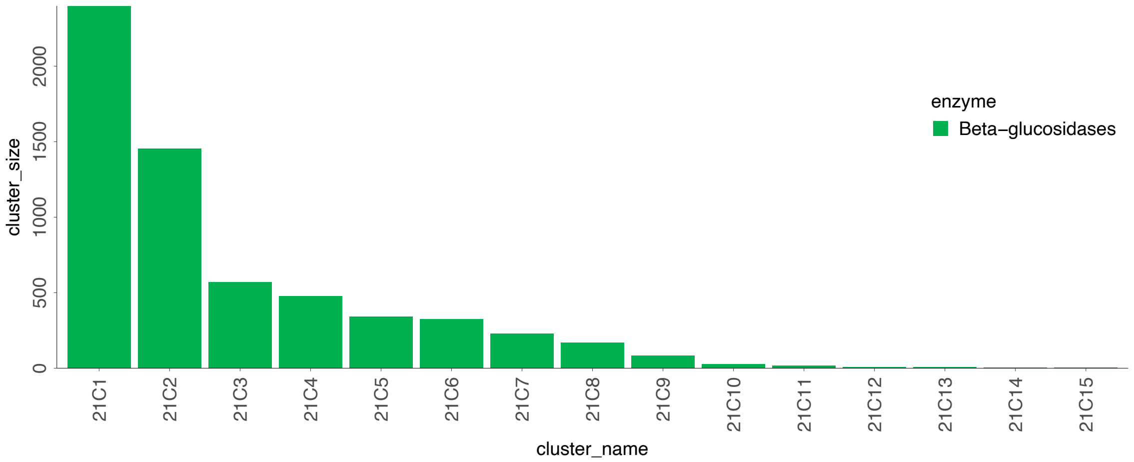


Supplementary_File_6 The size of each cellulase protein cluster

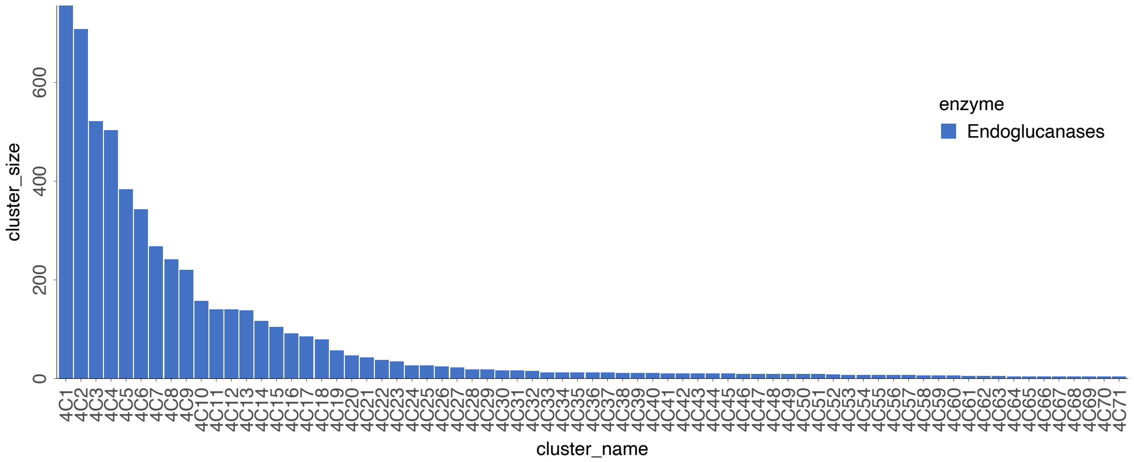
Each cluster is the smallest unit classified by its sequence similarity. The cluster size is the sequence amount clustered together with the global identity > 30%. Clusters with more than or equal to 4 sequences are shown here.



The cluster size of the 16 exoglucanases protein clusters (n>=4)



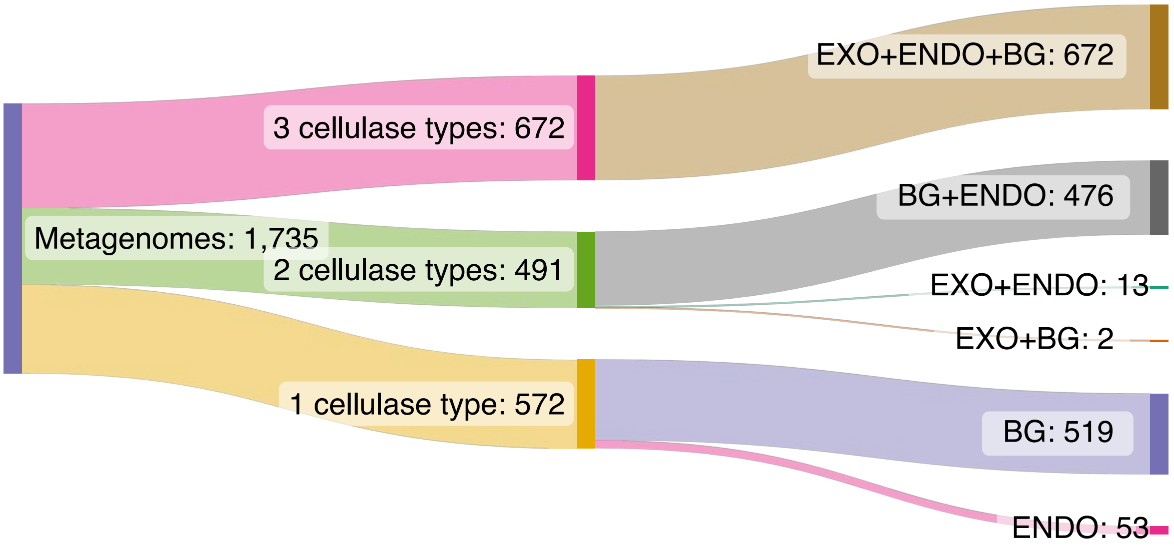
The cluster size of the 15 beta–glucosidases protein clusters (n>=4)



The cluster size of the 71 endoglucanases protein clusters (n>=4)

Supplementary_File_7 Distribution of the cellulase types across 1,735 metagenomes.

Note: 3 cellulase types – metagenomes having three cellulase types, two cellulase types – metagenomes having two cellulase types, one cellulase type – metagenomes having only one cellulase type. EXO – exoglucanases, BG – beta-glucosidases, ENDO – endoglucanases.



Supplementary_File_8 35 Prokaryotic phyla identified with cellulase gene

Phyla	Beta-glucosidases	Endoglucanases	Exoglucanases	Row_sum
Bacteroidetes	2121	1946	413	4480
Clostridia	1331	1420	378	3129
Gammaproteobacteria	420	661	57	1138
Actinobacteria	383	396	33	812
Bacilli	500	194	21	715
other_Bacteria	156	464	30	650
Alphaproteobacteria	291	126	20	437
Chloroflexi	313	45	0	358
Deltaproteobacteria	158	67	13	238
Spirochaetes	77	81	26	184
Thermotogae	63	41	12	116
Betaproteobacteria	25	70	0	95
Acidobacteria	56	26	8	90
Cyanobacteria	22	57	11	90
Negativicutes	48	17	15	80
Verrucomicrobia	12	41	1	54
Erysipelotrichia	54	0	0	54
Planctomycetes	2	17	1	20
Deinococcus-Thermus	15	4	1	20
Chlorobi	20	0	0	20
other_Archaea	11	0	0	11
Nitrospirae	8	3	0	11
Haloferacales	6	0	0	6
Natrialbales	2	2	2	6
Epsilonproteobacteria	4	0	0	4
Thermococcales	1	3	0	4
Halobacteriales	3	0	0	3
Chlamydiae	2	0	0	2
Fusobacteria	2	0	0	2
Thermoplasmatales	2	0	0	2
Tenericutes	1	1	0	2
Thermoproteales	1	0	0	1
Aquificae	0	1	0	1
Desulfurococcales	0	1	0	1
Acidithiobacillia	0	1	0	1

2. Manuscript II: The origin and distribution of the main oxygen sensing mechanism across metazoans

Bing Song[1†], Luca David Modjewski[1†], Nils Kapust[1], William F. Martin[1*]

1 Department of Biology, Institute for Molecular Evolution, Heinrich-Heine-Universität Düsseldorf, D-40225, Düsseldorf, Germany

† These authors contributed equally to this work.

* Corresponding author E-mail: bill@hhu.de

Written by Bing Song:

I downloaded all reference protein sequences from NCBI, KEGG, EMBL, and UniprotKB using python script and searched and filtered all potential HIF-related hits. I also wrote and corrected the draft, did all analyses, and made all figures and tables, except the supplementary Table 2.

Abstract

Oxygen sensing mechanisms are essential for metazoans, their origin and evolution in the context of oxygen in earth history are of interest. To trace the evolution of a main oxygen sensing mechanism among metazoans, the hypoxia-inducible factor, HIF, we investigated the phylogenetic distribution and phylogeny of 11 of its components across 566 eukaryote genomes. The HIF based oxygen sensing machinery in eukaryotes can be traced as far back as 800 million years (Ma) ago, likely to the last metazoan common ancestor (LMCA), and arose at a time when the atmospheric oxygen content corresponded roughly to the Pasteur point, or roughly 1% of present atmospheric level (PAL). By the time of the Cambrian explosion (541 to 485 Ma) as oxygen levels approached those of the modern atmosphere, the HIF system with its key components HIF1A, HIF1B, PHD1, PHD4, FIH and VHL was well established across metazoan lineages. HIF1A is more widely distributed and therefore may have evolved earlier than HIF2A and HIF3A, and HIF1B and is more widely distributed than HIF2B in invertebrates. PHD1, PHD4, FIH, and VHL appear in all 13 metazoan phyla. The O_2 consuming enzymes of the pathway, PHDs and FIH, have a lower substrate affinity, Km, for O_2 than terminal oxidases in the mitochondrial respiratory chain, in line with their function as an environmental signal to switch to anaerobic energy metabolic pathways. The ancient HIF system has been conserved and widespread during the period when metazoans evolved and diversified together with O_2 during earth history.

Introduction

Oxygen on earth stems from cyanobacterial photosynthesis. During earth history, there were two main phases of atmospheric oxygen content change. The first was the great oxidation event, GOE, around 2.4 billion years (Ga) ago (Bekker et al. 2004), followed by almost 2 Ga of low oxygen levels and a second, late rise of oxygen corresponding to the origin of land plants and animals about 500 million years (Ma) ago (Lyons et al. 2014). The metazoan lineage arose roughly 700-1000 Ma ago, long before the origin of land plants, during a phase of earth history when oxygen levels were much lower than today's. The first metazoans were thus well adapted to low oxygen environments from the outset of their evolution, but adapted to rising oxygen levels during evolution, particularly with the transition to life on land in a highly oxygenic atmosphere starting about 450 Ma ago (Martin et al. 2021). Oxygen sensing pathways in animals are integral to their evolution during changing oxygen concentrations (Hammarlund et al. 2020) in earth history. The main oxygen sensing pathway in animals is mediated by hypoxia-inducible factors, HIFs (Semenza 2001; Kaelin and Ratcliffe 2008; Zhang et al. 2014; Ivan and Kaelin 2017). The alpha subunit of HIF1, HIF1A is posttranslational modified by prolyl hydroxylases (PHD1–4), which catalyze the O_2 dependent hydroxylation of prolyl residues, inducing ubiquitinylation of HIF by the von Hippel Lindau tumor suppressor (VHL) (Brihimi-Horn and Pouysségur 2009), leading to HIF1A degradation by proteasomes (Fig. 1). The factor inhibiting HIF, FIH, hydroxylates asparaginyl residues in the alpha subunit of HIFs (HIF1 – 3A), in an O_2 dependent reaction, which inhibits proline hydroxylation (Fig. 1). HIF1A and PHDs are constitutively expressed. When O_2 is lacking, the alpha subunit of HIFs is not hydroxylated by PHDs and therefore not degraded, but accumulates and is transported into the cell nucleus to create a heterodimer with the beta subunit of HIFs (HIF1B and 2B). The HIFA HIFB heterodimer acts as the transcription factor, which binds to corresponding hypoxia response elements (HRE) of the promoters of the HIF-target genes to stimulate or repress the downstream gene transcription activities which can elicit a series of biological responses (Brihimi-Horn and Pouysségur 2009) (Fig. 1). As it relates to energy metabolism, the main biological response governed by the HIF pathway is a shift from aerobic energy metabolism to anaerobic energy metabolism, which in land mammals, diverts pyruvate flux away from mitochondria and O_2 dependent terminal oxidases towards cytosolic fermentations (Samanta and Semenza 2018).

The mechanism of the main oxygen sensing pathway is well known and its origin and distribution have also been widely studied from the phylogenetic perspective (Rytkönen and Storz 2011; Rytkönen et al. 2011, 2013; Mills et al. 2018; Rytkönen 2018; Graham and Barreto 2019; Graham and Barreto 2020) but less so from the geological perspective (Taylor and McElwain 2010; Hammarlund 2020). One hypothesis is that the main oxygen sensing pathway may have already occurred in the common ancestor of metazoans (Hammarlund et al. 2018; Hammarlund 2020). PHD like prolyl hydroxylases have been reported in unicellular eukaryotes such as Dictyostelium (Van der Wel et al. 2005) and Schizosaccharomyces pombe (Lee et al. 2009) and pathogenic proteobacteria like Vibrio cholerae (Aravind and Koonin 2001; McDonough et al. 2006) and Pseudomonas aeruginosa (Aravind and Koonin 2001; Scotti et al. 2014). The HIFA/PHD/VHL pathway is conserved in all animals but is not found in choanoflagellates (Monosiga brevicollis) or other protists (Loenarz et al. 2011), it is present in all eumetazoans, except Ctenophera (Mills et al. 2018). A recent study showed that choanoflagellates have PHD but not HIF and VHL (Rytkönen 2018). HIF1A is conserved among most metazoans while HIF2A appeared later

(Graham and Presnell 2017). Some studies have suggested that the HIF pathway was lacking in the last common ancestor of animals and is not ubiquitous across metazoans (Mills et al. 2018; Graham and Barreto 2020). From the geological perspective, the HIF system requires the presence of O_2, with HIF1A, HIF1B, PHD2, and VHL being the oldest components, HIF2A and PHD3 were suggested to have appeared around 460 to 421 Ma ago when the atmospheric oxygen level was lower than the present, whereby PHD1 and HIF3A are thought to have arisen most recently (about 312 Ma ago) with contemporary atmospheric oxygen levels (Taylor and McElwain 2010). FIH1 is missing in fruit flies and nematodes but not in intermediate beetles (Taylor and McElwain 2010).

Here, we focus on the origin and distribution of the HIF/PHD/FIH/VHL oxygen sensing pathway within metazoans from the perspective of oxygen in earth history, addressing the two phases of the appearance of its essential components: First, the presence of HIF/PHD/FIH/VHL pathway related genes — HIFAs, HIFBs, PHDs, FIH, and VHL — across metazoan and non-metazoan phyla, and the oxygen affinity of oxygen sensing enzymes (PHDs and FIH) were compared to terminal oxidases in mitochondria by the measure of substrate affinity for O_2, $K_m^{app}(O_2)$.

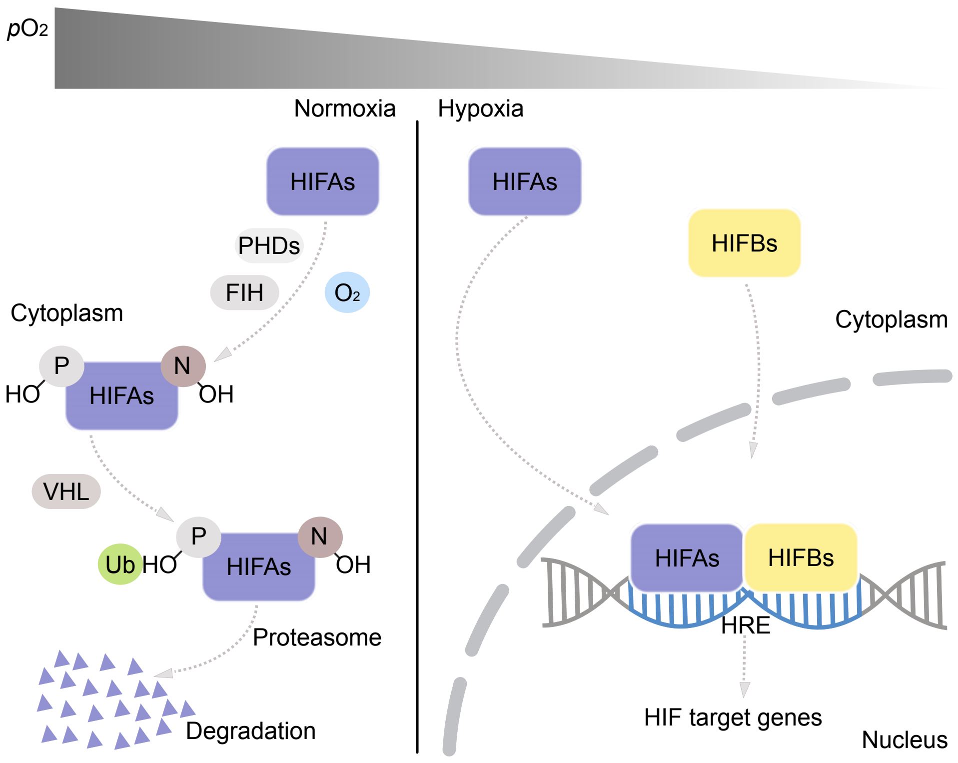


Fig. 1 The HIF oxygen sensing pathway.

Under normoxic conditions, proline (P) and asparaginyl (N) residues in HIFAs are first recognized and degraded by activated prolyl hydroxylases (PHDs) and asparaginyl hydroxylase (FIH). The hydroxylated proline residues (P-OH) bind on HIFAs and are ubiquitinated by the VHL and degraded by proteasomes within the cytoplasm. Under hypoxia, PHDs and FIH are inactivated and HIFAs are transported into the nucleus and form heterodimers with HIFBs. This heterodimer binds on the hypoxia responsive element (HRE) in the DNA structure and functions as a transcription factor activating or repressing the transcription activities of the downstream genes.

Methods

Identification and presence of HIF related genes. From NCBI (Agarwala et al. 2018), KEGG (Kanehisa et al. 2016), EMBL (Nightingale et al. 2017), and UniprotKB (Bateman 2019) (September 2020), 422 eukaryotic HIF related sequences from 11 gene categories were downloaded (Supplementary Table 1). Protein sequences from 566 complete eukaryotic genomes were downloaded from the NCBI RefSeq (O'Leary et al. 2016) database (July and October 2020) (Supplementary Table 2). The 566 genomes were classified into 15 phyla according to NCBI taxonomy. The proteins of the 566 eukaryotic genomes were blasted against the eukaryotic HIF related sequence database using diamond v2.0.1.139 (Buchfink et al. 2015), to identify potential HIF related genes. To determine the best corresponding HIF related gene in each genome, the best hit from each gene category for each genome was selected with an identity threshold of 25% and an e-value cut-off of 10-7. Duplicated sequences for the same identified HIF related gene were removed by only taking one of the best hits.

Presence and absence patterns of the selected hits were used to demonstrate the evolution of HIF related genes within metazoans over the last 3.0 Ga of earth's history (Fig. 2). The reference phylogenetic tree in Fig. 2 was taken from Telford et al. 2015 and the divergence time of the last common ancestor of each representative phylum on the phylogenetic tree was estimated from the free public database (Kumar et al. 2017) used for the estimation of time-scaled phylogenies (Dos Reis et al. 2015; Gold et al. 2015; Delsuc et al. 2018; Tedersoo et al. 2018). The presence of the genes in each phylum is based on the respective gene being present in at least one species within the indicated phylum and represented by filled colored circles, if the gene was not present in the respective phylum, the space was left blank. The time of the earliest eukaryote fossil record (Eme et al. 2014) is displayed in the time scale at the bottom in light brown and the earliest animal fossil record (Hoyal Cuthill and Han 2018) is indicated by a dark brown bar in the time scale.

Oxygen affinity of oxygen sensing enzymes and terminal oxidases. $K_m^{app}(O_2)$ values of oxygen sensing enzymes (PHD and FIH) and terminal oxidases were collected from literature (Supplementary Tables 3 and 4). Km is the Michaelis constant, which represents the concentration of substrate needed to achieve half of the maximum reaction speed (Vmax) of the enzyme (Northrop 1998). $K_m^{app}(O_2)$ represents the concentration of oxygen required to achieve half of the Vmax of the examined enzyme; the lower the Km the higher the affinity of the enzyme for the substrate.

Result

The HIF oxygen sensing pathway traces to the last metazoan common ancestor. The two non-metazoan phyla in our dataset (Choanoflagellata and Ascomycota) possess the oxygen sensors PHD, FIH, and VHL. This result is based on the genome of Salpingoeca rosetta, a choanoflagellate (Supplementary Table 2). The choanoflagellates are a group of simple unicellular eukaryotes considered to be the closest relatives of metazoans (Carr et al. 2008). It has been previously reported that choanoflagellates have PHD, but not HIF and VHL (Rytkönen 2018). Ascomyceta fungi also possess HIF homologues, the result of Ascomycota is based on nine fungi genomes (Supplementary Table 2). Ascomycota are the largest phylum of fungi predating metazoans, within which Ofd1 — the prolyl 4-hydroxylase-like 2-OG-Fe(II) dioxygenase — mediates the degradation and accumulation of the N-terminal transcription factor (Sre1N) (Hughes and Espenshade 2008). Porifera is the sister group of all other animals

(Simion et al. 2017) and has been shown to contain HIF1A, HIF1B, PHD1, PHD2, PHD4, FIH, and VHL based on the sponge species Amphimedon queenslandica (Supplementary Table 2). Trichoplax adhaerens which belongs to the phylum Placozoa is discussed as the last common ancestor of all animals (Schierwater and DeSalle 2018) and was previously shown to possess the key components of the oxygen sensing mechanism — HIFA, PHD and VHL (Loenarz et al. 2011). Our finding shows a more complete result as T. adhaerens not only contains HIFA and PHD homologs but also HIFB, FIH, and VHL homologues. The main oxygen sensing pathway can thus be traced back to at least the last common ancestor of metazoans, which lived roughly 800 Ma ago, when the atmospheric oxygen content was around 1% PAL (Erwin et al. 2011) (Fig. 2).

The HIF pathway is conserved and widely distributed across metazoans. The 13 metazoan phyla examined all possess HIF1A, HIF1B, PHD1, PHD4, FIH, and VHL. This shows that the core components of the main oxygen sensing pathway are relatively conserved and widespread across metazoans. Moreover, HIF1A is more widely distributed across all phyla and therefore probably evolved earlier than HIF2A and HIF3A (Fig. 2). HIF1B is more widely distributed than HIF2B, whereas PHD1 and PHD4 may have arisen earlier than PHD2 and PHD3 as they are found in all phyla examined as shown in Fig. 2. It was previously shown that HIF1A is more conserved among metazoans and probably arose before HIF2A (Graham and Presnell 2017), consistent with our findings. However, previous studies have shown that HIF homologs are ubiquitous across metazoans, with the exception of Porifera and Ctenophora, and HIF2A only appears in vertebrates (Loenarz et al. 2011; Rytkönen et al. 2011; Graham and Presnell 2017; Hammarlund et al. 2018; Rytkönen 2018). Accordingly, it has been suggested that the HIF/PHD/FIH/VHL oxygen sensing pathway might represent a recent lineage specific invention among recent animals (Hashimoto et al. 2016; Graham and Barreto 2019; Graham and Barreto 2020). The copepod Tigriopus californicus seems to have lost the HIF pathway but still tolerates nearly anoxic conditions for at least 24 hours, the genes involved in cuticle reorganization and ion transport may act as the potential solution to low oxygen availability as a replacement to the HIF pathway (Graham and Barreto 2019). Ramazzottius varieornatus, one member of the most stress-tolerant tardigrade species, has lost HIF1A, PHD, and VHL selectively (Hashimoto et al. 2016). Three of the four orders of Copepoda were shown to have lost the use of the HIF pathway, but both barnacles and copepods have not fully lost VHL which suggests HIF-independent functions of VHL (Graham and Barreto 2020).

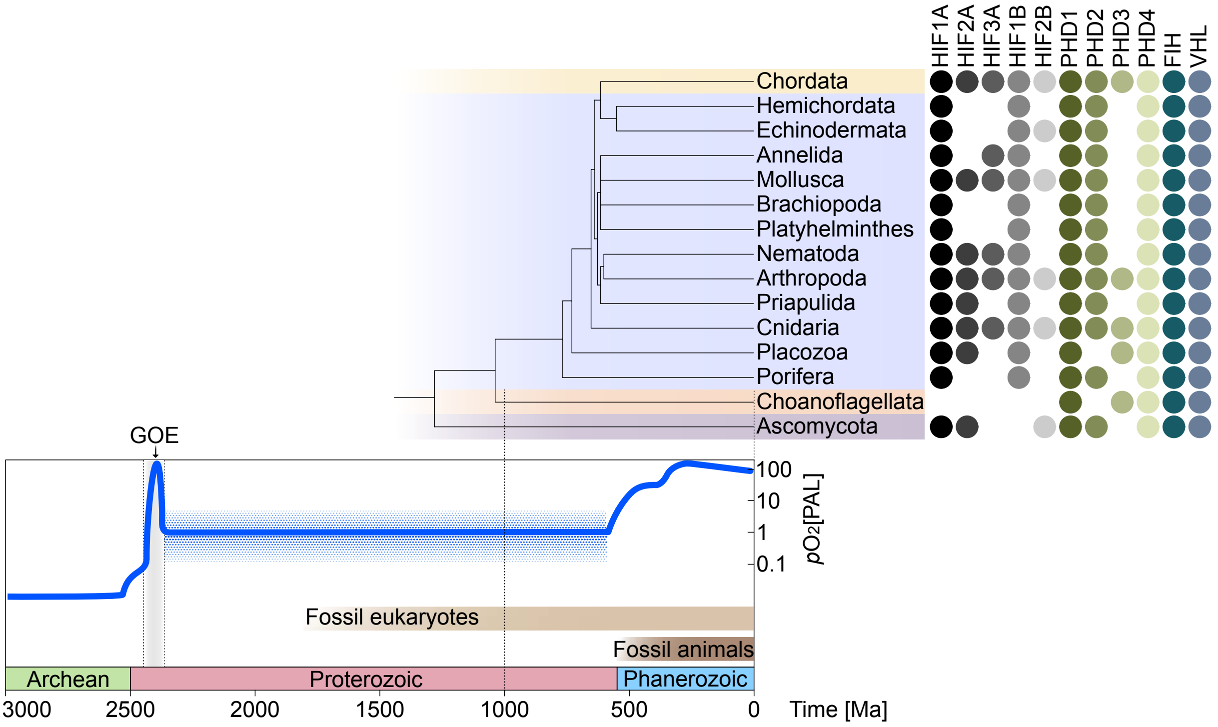


Fig. 2 The evolution of the main oxygen sensing pathway across metazoans over the earth's oxygen history.

The oxygen curve describes the earth's oxygen content during each geological time scale (Walker et al. 2018), the detailed information has been obtained from recent publications (Bekker et al. 2004; Lyons et al. 2014). More than 2500 million years (Ma) ago, the earth oxygen content was almost at an anaerobic state (Catling and Zahnle 2020); during the Proterozoic period (2500 to 542 Ma), despite the great oxidation event (GOE) that happened at the beginning of this period, the oxygen content was still very low with around 1% present atmospheric levels (PAL) (Allen et al. 2019); during the Phanerozoic period (541 to 0 Ma), the atmospheric oxygen content increased constantly to the modern level (Lenton et al. 2016; Krause et al. 2018). The phylogenetic tree shown in the figure is scaled to the timeframe at the bottom of the figure. It is connected to the presence-absence pattern (PAP) of oxygen sensing genes. The PAP shows the distribution of the 11 gene categories across 13 metazoan phyla (Chordata, Hemichordata, Echinodermata, Annelida, Mollusca, Brachiopoda, Platyhelminthes, Nematoda, Arthropoda, Priapulida, Cnidaria, Placozoa, Porifera) and two non-metazoan phyla (Choanoflagellata and Ascomycota) that include 566 species (Supplementary Table 2). The oxygen sensing regulators (HIF1A, HIF2A, HIF3A, HIF1B, and HIF2B) at the left of the PAP are the substrates of the oxygen sensors (PHD1, PHD2, PHD3, PHD4 and FIH) in oxygen-dependent reactions before the ubiquitination (Ub) of the tumor suppressor gene (VHL). (HIF – hypoxia-inducible factor; PHD – HIF prolyl hydroxylases; FIH – factor inhibiting HIF; VHL – von Hippel Lindau tumor suppressor).

The oxygen affinity of terminal oxidases is far higher than that of oxygen sensing enzymes. The $K_m^{app}(O_2)$ of terminal oxidases were obtained from both prokaryotes and eukaryotes, the $K_m^{app}(O_2)$ values of oxygen sensing enzymes (PHDs and FIH) were obtained from the literature for eukaryotes (Supplementary

Tables 3 and 4). Fig. 3 shows the distribution of the $K_m^{app}(O_2)$ values of terminal oxidases, PHDs and FIH, which are 0.0034 to 33 µM, 11.54 to 530 µM and 12 to 270 µM, respectively. Low $K_m^{app}(O_2)$ values indicate higher oxygen affinity, this shows that the oxygen affinity of terminal oxidases is far higher than that of PHD and FIH.

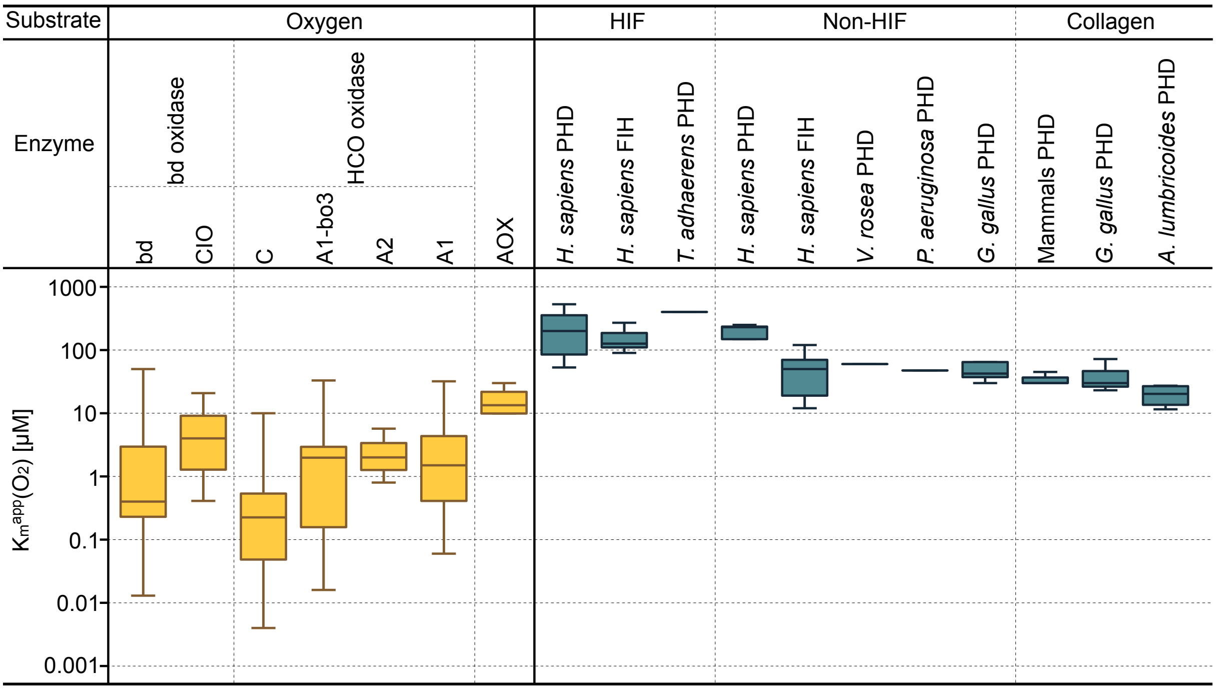


Fig. 3 The $K_m^{app}(O_2)$ values of terminal oxidases and oxygen sensing enzymes. On the left (yellow) $K_m^{app}(O_2)$ values of the three main families of terminal oxidases are shown and indicated on the right (dark cyan) are $K_m^{app}(O_2)$ values of oxygen sensing enzymes (PHDs and FIH). For cytochrome bd oxidases, two subtypes (bd and CIO) are shown; for HCOs, four subtypes (C, A1-bo3, A2, and A1) are shown. On the top the substrates of the enzymes are indicated; oxygen as substrate for terminal oxidases, as their main function is to reduce oxygen within organism organelles. The $K_m^{app}(O_2)$ values of oxygen sensors (PHDs and FIH) are combined based on their substrates, which are sorted into three groups — HIF

protein peptides, non-HIF protein peptides, and collagen. All underlying $K_m^{app}(O_2)$ values are detailed in Supplementary Tables 3 and 4. (AOX: Alternative Oxidase, CIO: Cyanide Insensitive Oxidase, HCO: Heme Copper Oxidase.)

Discussion

In previous studies, the characteristic domains of HIF, PHD, FIH, and VHL were used to identify HIF related homologs as the characteristic domains are usually the most conserved and representative part of one protein family. Here, complete HIF related protein sequences were used to find possible HIF pathway homologues from sequenced genomes. This identified HIF pathway homologues that were not detected in previous studies. We used Diamond (Buchfink et al. 2014) for searching, which is generally faster and often more accurate than the traditional Basic Local Alignment Search Tool (BLAST; Altschul et al. 1990), which helped to identify previously missing HIF pathway genes.

The HIF/PHD/FIH/VHL pathway was present during the Cambrian explosion. Genes for enzymes of the HIF oxygen sensing pathway — PHD1, PHD4, FIH, VHL — are distributed across all examined phyla (Fig. 2), tracing their first appearance to lineages during the course of metazoan and non-metazoan phylum diversification. HIF1A was the progenitor from which HIF2A and HIF3A arose, and HIF1B may have emerged before HIF2B as it is more widely distributed, giving rise to the hypoxia-inducible transcription factor (HIF1), which is composed of HIF1A and HIF1B, and is the major regulator of oxygen homeostasis (Wang et al. 1995; Semenza 2007). PHD3 only appears in five of the 13 phyla analyzed but was present in Choanoflagellata. Previous work reported that PHD3 may function as an ancient signaling protein due to its involvement in several cell signaling mechanisms (Place and Domann 2013), which enables large animals to sense and deliver oxygen for development and metabolism. Therefore, the retention of PHD3 in Choanoflagellata might not be due to the oxygen sensing pathway but might relate to other functions.

It has been suggested that rudimentary HIF1A-dominant stemness control can function in such a way as to generate oxygen gradients within a tissue in the presence of 1% oxygen and thus maintain the inner tissue's hypoxic cell stemness, with HIF2A-driven pseudohypoxia maintaining stemness of cells in nearby well-vascularized and oxygenated tissue with more than 1 to 3% oxygen (Hammarlund et al. 2018). During early animal evolution, refined HIF stemness control might have played an essential role in promoting the larger size of animals and in energy metabolism. Before the Cambrian, low oxygen levels were sufficient to meet the physiological requirements of the simple invertebrates, but after the Cambrian, higher oxygen would have interfered with animal the stemness, which would implicate a role of HIF in the evolution of animal size at the Cambrian explosion. At the same time, collagen hydroxylation at proline residues would have led to more rigid invertebrate body structures and better fossilization properties (Towe 1970).

The functional ranges of both HIF1A and HIF2A (5 to 24% PAL or 1 to 5% O_2) hydroxylation (Holmquist-Mengelbier et al. 2006) suggest that they are a relic from the time of O_2 sensing during low O_2 concentrations in the Cambrian (15 to 20% PAL or 3 to 4% O_2), marking a role for the HIF pathway during the adaptation of larger animals to the rising atmospheric oxygen concentrations. Under low oxygen concentrations of around 1%, both HIF1A and HIF2A are stabilized with HIF1A driving the hypoxia induced reactions which can create oxygen gradients in tissue and thus result in hypoxia and cell stemness within invertebrate tissue. HIF2A may have gained a role for creating the pseudohypoxic phenotype in order to facilitate the activation of genes related with promoting stemness within vertebrate tissue, while HIF2A maintained HIF2A-driven pseudohypoxia, promoting stemness even at oxygen levels of roughly 5% (Hammarlund et al. 2018). Clearly HIF dependent oxygen sensing was integral to the Cambrian explosion (541 to 485 Ma) and likely facilitated an increase in animal diversity (Knoll and Carroll 1999).

Low O_2 affinity of HIF and high O2 affinity of terminal oxidases. The $K_m^{app}(O_2)$ values in Fig. 3 show that the oxygen affinity of O_2 sensing enzymes is lower than that of terminal oxidases by one to two orders of magnitude. The O_2 affinity of A1 type terminal oxidases in mitochondria is on the order of 10 µM, corresponding to 1% [v/v] O_2 in air reflecting the environmental concentration of O_2 at the time of mitochondrial origin (Zimorski et al. 2019) which is, in turn, very close to the O_2 concentration in functioning mitochondria of human tissues because of the oxygen cascade from air to blood to capillaries (Martin et al. 2021). This affinity seems to have changed little during evolution. The situation is different with HIF however, as different animals respond to hypoxia at different levels. For example, Caenorhabditis elegans prefers O_2 levels around 7%, and responds to hypoxia only at about 1% O_2 (Branicky and Schafer 2008), reflecting diversification of oxygen sensing physiology during metazoan evolution (Kaelin and Ratcliffe 2008; Hampton-Smith and Peet 2009).

PHDs and FIH can target alternative substrates other than HIF, due to their involvement in both HIF and non-HIF related pathways. Their protein hydroxylation is neither unique nor ubiquitous. The substrates of PHDs and FIH can be sorted into three groups (Strowitzki et al. 2019) (Fig. 3). Within the oxygen metabolizing enzymes, FIH has a relatively high oxygen affinity compared to the PHDs suggesting that FIH is more hypoxia-tolerant than the PHDs (Tarhonskaya et al. 2015). The $K_m^{app}(O_2)$ values of FIH are normally between 90 up to 270 µM (Koivunen et al. 2004; Hangasky et al. 2014; Wilson et al. 2020) while that of PHDs are 30 to 250 µM (Hirsilä et al. 2003). One study also reported an outlier $K_m^{app}(O_2)$ value of PHD around 1700 µM, which is not considered here (Dao et al. 2009). PHD2 was identified as the main regulator of the normal development of growth plate chondrocytes in the avascular environment, as it can inactivate HIF1A to avoid prolonged HIF1A-induced skeletal dysplasia (Stegen et al. 2019). Overexpressed HIF1A can reprogram cellular metabolism from respiration to fermentation (Stegen et al. 2019). Appropriate control of PHD2 on HIF1A activation is also important for collagen synthesis. For the comparison of the same oxygen sensing enzyme type from the same substrate group, it can be seen that larger sized vertebrates have higher oxygen affinity than smaller sized vertebrates and invertebrates on HIF-based substrates and collagen but not non-HIF substrates. This might reflect selection for a greater O_2 demand for structural rigidity among larger vertebrates that are adapted to life above the soil line.

Within the terminal oxidases, the $K_m^{app}(O_2)$ values of other bacteria except for Alphaproteobacteria, archaea, and eukaryotes have also been added here. The mean $K_m^{app}(O_2)$ values of these seven of terminal oxidase subtypes, from high to low, are AOX (14.91 µM) > CIO (8.40 µM) > A1-bo3 (4.38 µM) > A1 (4.32 µM) > bd (4.18 µM) > A2 (2.83 µM) > C (1.62 µM) respectively (Degli Esposti et al. 2019). Cytochrome c oxidase was reported to exhibit a much smaller $K_m^{app}(O_2)$ than the mean physiological oxygen concentration compared to other oxygen utilizing enzymes in mammals, while the reaction rates of other oxygen-consuming enzymes could be limited by the physiological oxygen tensions commonly present in organs (Vanderkooi et al. 1991). The lower affinity for O_2 of HIF dependent O_2 sensing oxygenases relative to terminal oxidases suggests that the former have undergone evolutionary adaptations in response to changing O_2 environments that reflect the physiological needs of the whole animal as opposed to its mitochondria, which are supplied via circulatory and respiratory systems with roughly the same concentration of O_2 as the mitochondria of the first unicellular eukaryotes more than 1.5 billion years ago.

Acknowledgements

This work received financial support from the Deutsch-Israelische Projektkooperation (DIP) der Deutschen Forschungsgemeinschaft (MA 1426/23-1). The stipend of Bing Song was funded by the China Scholarship Council (Bing Song's application number: 201706300100).

Author Contributions

William F. Martin & Nils Kapust conceived and designed the study, Bing Song & Luca David Modjewski collected and analyzed the data, Bing Song & Nils Kapust drafted the initial version of the manuscript and all authors contributed to final version of the manuscript. And all authors agree to the submission and declare no conflicts of interest. We thank Julia Bruekner, who corrected the grammar for our first version manuscript.

Data Accessibility

All data mentioned in the text are available on request.

References

Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bollin, C., Bolton, E. et al. (2018). Database resources of the national center for biotechnology information. Nucleic Acids Res. 46:8–13.

Allen, J. F., Thake, B. & Martin, W. F. (2019). Nitrogenase inhibition limited oxygenation of earth’s proterozoic atmosphere. Trends Plant Sci. 24:1022–1031.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Aravind, L. & Koonin, E. V. (2001). The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate-andiron-dependent dioxygenases. Genome Biol. 2:1–8.

Bateman, A. (2019). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 47:506–515.

Bekker, A., Holland, H. D., Wang, P. L., Rumble, D. I. I. I., Stein, H. J., Hannah, J. L. et al. (2004). Dating the rise of atmospheric oxygen. Nature 427:117–120.

Branicky, R. S. & Schafer, W. R. (2008). Oxygen homeostasis: how the worm adapts to variable oxygen levels. Curr. Biol. 18:559–560.

Brahimi-Horn, M. C. & Pouysségur, J. (2009). HIF at a glance. J. Cell Sci. 122:1055–1057.

Buchfink, B., Xie, C. & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12:59–60.

Carr, M., Leadbeater, B. S., Hassan, R., Nelson, M. & Baldauf, S. L. (2008). Molecular phylogeny of choanoflagellates, the sister group to Metazoa. PNAS 105:16641–16646.

Catling, D. C. & Zahnle, K. J. (2020). The archean atmosphere. Sci. Adv. 6:1420.

Dao, J. H., Kurzeja, R. J., Morachis, J. M., Veith, H., Lewis, J., Yu, V. et al. (2009). Kinetic characterization and identification of a novel inhibitor of hypoxia-inducible factor prolyl hydroxylase 2 using a time-resolved fluorescence resonance energy transfer-based assay technology. Anal. Biochem. 384:213–223.

Degli Esposti, M., Mentel, M., Martin, W. & Sousa, F. L. (2019). Oxygen reductases in alphaproteobacterial genomes: physiological evolution from low to high oxygen environments. Front. Microbiol. 10:499.

Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M. K., Turon, X. et al. (2018). A phylogenomic framework and timescale for comparative studies of tunicates. BMC Biol. 16:1–14.

Dos Reis, M., Thawornwattana, Y., Angelis, K., Telford, M. J., Donoghue, P. C. & Yang, Z. (2015). Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Curr. Biol. 25:2939–2950.

Eme, L., Sharpe, S. C., Brown, M. W. & Roger, A. J. (2014). On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. Cold Spring Harb. Perspect. Biol. 6:a016139.

Erwin, D. H., Laflamme, M., Tweedt, S. M., Sperling, E. A., Pisani, D. & Peterson, K. J. (2011). The Cambrian conundrum: early divergence and later ecological success in the early history of animals. Science 334:1091–1097.

Gold, D. A., Runnegar, B., Gehling, J. G. & Jacobs, D. K. (2015). Ancestral state reconstruction of ontogeny supports a bilaterian affinity for Dickinsonia. Evol. Dev. 17:315–324.

Graham, A. M. & Presnell, J. S. (2017). Hypoxia Inducible Factor (HIF) transcription factor family expansion, diversification, divergence and selection in eukaryotes. PLoS One 12:e0179545.

Graham, A. M. & Barreto, F. S. (2019). Loss of the HIF pathway in a widely distributed intertidal crustacean, the copepod Tigriopus californicus. PNAS 116:12913–12918.

Graham, A. M. & Barreto, F. S. (2020). Independent losses of the hypoxia-inducible factor (HIF) pathway within Crustacea. Mol Biol Evol 37:1342–1349.

Hammarlund, E. U., Von Stedingk, K. & Påhlman, S. (2018). Refined control of cell stemness allowed animal evolution in the oxic realm. Nat. Ecol. Evol. 2:220–228.

Hammarlund, E. U. (2020). Harnessing hypoxia as an evolutionary driver of complex multicellularity. Interface Focus 10:20190101.

Hammarlund, E. U., Flashman, E., Mohlin, S. & Licausi, F. (2020). Oxygen-sensing mechanisms across eukaryotic kingdoms and their roles in complex multicellularity. Science 370.

Hampton-Smith, R. J. & Peet, D. J. (2009). From polyps to people: a highly familiar response to hypoxia. Ann. N. Y. Acad. Sci. 1177:19–29.

Hangasky, J. A., Gandhi, H., Valliere, M. A., Ostrom, N. E. & Knapp, M. J. (2014). The rate-limiting step of O2 activation in the α-ketoglutarate oxygenase factor inhibiting hypoxia inducible factor. Biochemistry 53:8077–8084.

Hashimoto, T., Horikawa, D. D., Saito, Y., Kuwahara, H., Kozuka-Hata, H., Shin, T. et al. (2016). Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. Nat. Commun. 7:1–14.

Hirsilä, M., Koivunen, P., Günzler, V., Kivirikko, K. I. & Myllyharju, J. (2003). Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. J. Biol. Chem. 278:30772–30780.

Holmquist-Mengelbier, L., Fredlund, E., Löfstedt, T., Noguera, R., Navarro, S., Nilsson, H. et al. (2006). Recruitment of HIF-1α and HIF-2α to common target genes is differentially regulated in neuroblastoma: HIF-2α promotes an aggressive phenotype. Cancer Cell 10:413–423.

Hoyal Cuthill, J. F. & Han, J. (2018). Cambrian petalonamid Stromatoveris phylogenetically links Ediacaran biota to later animals. Palaeontology 61:813–823.

Hughes, B. T. & Espenshade, P. J. (2008). Oxygen-regulated degradation of fission yeast SREBP by Ofd1, a prolyl hydroxylase family member. EMBO J. 27:1491–1501.

Ivan, M. & Kaelin Jr, W. G. (2017). The EGLN-HIF O2-sensing system: multiple inputs and feedbacks. Mol. Cell 66:772–779.

Kaelin Jr, W. G. & Ratcliffe, P. J. (2008). Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. Mol. Cell 30:393–402.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44:457–462.

Knoll, A. H. & Carroll, S. B. (1999). Early animal evolution: emerging views from comparative biology and geology. Science 284:2129–2137.

Koivunen, P., Hirsilä, M., Günzler, V., Kivirikko, K. I. & Myllyharju, J. (2004). Catalytic properties of the asparaginyl hydroxylase (FIH) in the oxygen sensing pathway are distinct from those of its prolyl 4-hydroxylases. J. Biol. Chem. 279:9899–9904.

Krause, A. J., Mills, B. J., Zhang, S., Planavsky, N. J., Lenton, T. M. & Poulton, S. W. (2018). Stepwise oxygenation of the Paleozoic atmosphere. Nat. Commun. 9:1–10.

Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. (2017). TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol 34:1812–1819.

Lee, C. Y. S., Stewart, E. V., Hughes, B. T. & Espenshade, P. J. (2009). Oxygen-dependent binding of Nro1 to the prolyl hydroxylase Ofd1 regulates SREBP degradation in yeast. EMBO J. 28:135–143.

Lenton, T. M., Dahl, T. W., Daines, S. J., Mills, B. J., Ozaki, K., Saltzman, M. R. et al. (2016). Earliest land plants created modern levels of atmospheric oxygen. PNAS 113:9704–9709.

Loenarz, C., Coleman, M. L., Boleininger, A., Schierwater, B., Holland, P. W., Ratcliffe, P. J. et al. (2011). The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, Trichoplax adhaerens. EMBO Rep. 12:63–70.

Lyons, T. W., Reinhard, C. T. & Planavsky, N. J. (2014). The rise of oxygen in earth's early ocean and atmosphere. Nature 506:307–315.

Martin, W. F., Tielens, A. G. & Mentel, M. (2021). Mitochondria and anaerobic energy metabolism in eukaryotes. Düsseldorf University Press, Düsseldorf.

McDonough, M. A., Li, V., Flashman, E., Chowdhury, R., Mohr, C., Liénard, B. M. et al. (2006). Cellular oxygen sensing: crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). PNAS 103:9814–9819.

Mills, D. B., Francis, W. R., Vargas, S., Larsen, M., Elemans, C. P., Canfield, D. E. et al. (2018). The last common ancestor of animals lacked the HIF pathway and respired in low-oxygen environments. Elife. 7:e31176.

Modjewski LD (2021) Relationship between Earth's atmospheric oxygen level and the emergence of hypoxia-inducible factors and related enzymes in eukaryotes.

Nightingale, A., Antunes, R., Alpi, E., Bursteinas, B., Gonzales, L., Liu, W. et al. (2017). The proteins API: accessing key integrated protein and genome information. Nucleic Acids Res. 45:539–544.

Northrop, D. B. (1998). On the meaning of Km and V/K in enzyme kinetics. J. Chem. Educ. 75:1153.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R. et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44:733–745.

Place, T. L. & Domann, F. E. (2013). Prolyl-hydroxylase 3: evolving roles for an ancient signaling protein. Hypoxia 1:13.

Rytkönen, K. T. & Storz, J. F. (2011). Evolutionary origins of oxygen sensing in animals. EMBO Rep. 12:3–4.

Rytkönen, K. T., Williams, T. A., Renshaw, G. M., Primmer, C. R. & Nikinmaa, M. (2011). Molecular evolution of the metazoan PHD-HIF oxygen-sensing system. Mol Biol Evol 28:1913–1926.

Rytkönen, K. T. (2018). Evolution: oxygen and early animals. Elife 7:e34756.

Samanta, D. & Semenza, G. L. (2018). Metabolic adaptation of cancer and immune cells mediated by hypoxia-inducible factors. Biochim Biophys Acta Rev Cancer 1870:15–22.

Schierwater, B. & DeSalle, R. (2018). Placozoa. Curr. Biol. 28:97–98.

Scotti, J. S., Leung, I. K., Ge, W., Bentley, M. A., Paps, J., Kramer, H. B. et al. (2014). Human oxygen sensing may have origins in prokaryotic elongation factor Tu prolyl-hydroxylation. PNAS 111:13331–13336.

Semenza, G. L. (2001). HIF-1, O2, and the 3 PHDs: how animal cells signal hypoxia to the nucleus. Cell 107:1–3.

Semenza, G. L. (2007). Hypoxia-inducible factor 1 (HIF-1) pathway. Science's STKE 407:8.

Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D. J., Di Franco, A. et al. (2017). A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. Curr. Biol. 27:958–967.

Stegen, S., Laperre, K., Eelen, G., Rinaldi, G., Fraisl, P., Torrekens, S. et al. (2019). HIF-1α metabolically controls collagen synthesis and modification in chondrocytes. Nature 565:511–515.

Strowitzki, M. J., Cummins, E. P. & Taylor, C. T. (2019). Protein hydroxylation by hypoxia-inducible factor (HIF) hydroxylases: unique or ubiquitous? Cells 8:384.

Tarhonskaya, H., Hardy, A. P., Howe, E. A., Loik, N. D., Kramer, H. B., McCullagh, J. S. et al. (2015). Kinetic investigations of the role of factor inhibiting hypoxia-inducible factor (FIH) as an oxygen sensor. J. Biol. Chem.290:19726–19742.

Taylor, C. T. & McElwain, J. C. (2010). Ancient atmospheres and the evolution of oxygen sensing via the hypoxia-inducible factor in metazoans. Physiology 25:272–279.

Tedersoo, L., Sánchez-Ramírez, S., Koljalg, U., Bahram, M., Döring, M., Schigel, D. et al. (2018). High-level classification of the fungi and a tool for evolutionary ecological analyses. Fungal Divers. 90:135–159.

Telford, M. J., Budd, G. E. & Philippe, H. (2015). Phylogenomic insights into animal evolution. Curr. Biol. 25:876–887.

Towe, K. M. (1970). Oxygen-collagen priority and the early metazoan fossil record. PNAS 65:781–788.

Van der Wel, H., Ercan, A. & West, C. M. (2005). The Skp1 prolyl hydroxylase from Dictyostelium is related to the hypoxia-inducible factor-α class of animal prolyl 4-hydroxylases. J. Biol. Chem. 280:14645–14655.

Vanderkooi, J. M., Erecinska, M. & Silver, I. A. (1991). Oxygen in mammalian tissue: methods of measurement and affinities of various reactions. Am. J. Physiol. Cell Physiol. 260:1131–1150.

Walker, J. D., Geissman, J. W., Bowring, S. A., & Babcock, L. E. (2018). GSA geological time scale v. 5.0. Geol. Soc. Am.

Wang, G. L., Jiang, B. H., Rue, E. A. & Semenza, G. L. (1995). Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. PNAS 92:5510–5514.

Wilson, J. W., Shakir, D., Batie, M., Frost, M. & Rocha, S. (2020). Oxygen-sensing mechanisms in cells. FEBS J. 287:3888–3906.

Zhang, P., Yao, Q., Lu, L., Li, Y., Chen, P. J. & Duan, C. (2014). Hypoxia-inducible factor 3 is an oxygen-dependent transcription activator and regulates a distinct transcriptional response to hypoxia. Cell Rep. 6:1110–1121.

Zimorski, V., Mentel, M., Tielens, A. G. & Martin, W. F. (2019). Energy metabolism in anaerobic eukaryotes and earth's late oxygenation. Free Radic. Biol. Med. 140:279–294

Supplementary_Table_1 422 eukaryotic HIF-related sequences of 11 gene categories

Organism_name	Organism_id	Protein_name	New_protein_name	Protein_id	New_protein_id	Sequence	Url	Database	Download_time
Mus musculus (Mouse)	10090	HIF1A_MOUSE	HIF1A	Q61221	U10090_1_HIF1A_1	TIISLDFGSDDTETEDQQLE	mat=tab&columns=linea	UniProtKB	8/17/20 9:57
Mus musculus	10090	factor 1-alpha isoform	HIF1A	NP_001300849.1	N10090_2_HIF1A_2	DTIISLDFGSDDTETEDQQL	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Mus musculus	10090	factor 1-alpha isoform	HIF1A	NP_001300848.1	N10090_3_HIF1A_3	DTIISLDFGSDDTETEDQQL	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Mus musculus	10090	actor 1-alpha isoform	HIF1A	XP_017170450.1	N10090_4_HIF1A_4	LAPAAGDTIISLDFGSDDTE	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Homo sapiens (Human)	9606	HIF1A_HUMAN	HIF1A	Q16665	U9606_1_HIF1A_5	AAGDTIISLDFGSNDTETDD	mat=tab&columns=linea	UniProtKB	8/17/20 9:57
Homo sapiens	9606	factor 1-alpha isoform	HIF1A	NP_851397.1	N9606_2_HIF1A_6	QTECVLKPVESSDMKMTQL	gov/entrez/eutils/efetch	NCBI	8/17/20 9:57
Homo sapiens	9606	factor 1-alpha isoform	HIF1A	NP_001230013.1	N9606_3_HIF1A_7	KKEPDALTLLAPAAGDTIISL	gov/entrez/eutils/efetch	NCBI	8/17/20 9:57
Xenopus laevis (African clawed frog)	8355	HIF1A_XENLA	HIF1A	Q9I8A9	U8355_1_HIF1A_8	LKQEPESLTVLAPDAGDEIIP	mat=tab&columns=linea	UniProtKB	8/17/20 9:57
Xenopus laevis (African clawed frog)	8355	le factor 1, alpha sub	HIF1A	xla:380141	K8355_2_HIF1A_9	LKQEPESLTVLAPDAGDEIIP	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Xenopus laevis (African clawed frog)	8355	a; hypoxia-inducible fac	HIF1A	xla:445838	K8355_3_HIF1A_10	KLKQEPESLTALAPDAGDDII	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Salmo salar (Atlantic salmon)	8030	hypoxia-inducible factor	HIF1A	sasa:106598919	K8030_1_HIF1A_11	TEDMRPVKKELEESSEAEV	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Salmo salar (Atlantic salmon)	8030	poxia-inducible factor 1	HIF1A	sasa:106610969	K8030_2_HIF1A_12	VLSGIEEKMVLSLEQTEDMR	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Salmo salar (Atlantic salmon)	8030	ypoxia-inducible factor	HIF1A	sasa:100194993	K8030_3_HIF1A_13	YRFLAKSGGFVWAETQATV	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Cricetulus griseus (Chinese hamster)	10029	a-inducible factor 1-alph	HIF1A	cge:100751293	K10029_1_HIF1A_14	AGDTIISLDFGSDDTETEDQ	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Cricetulus griseus	10029	ctor 1-alpha isoform X2	HIF1A	XP_007647201.1	N10029_2_HIF1A_15	LDFGSDDTETEDQQLEDVP	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Cricetulus griseus	10029	ctor 1-alpha isoform X1	HIF1A	XP_007647200.1	N10029_3_HIF1A_16	DFGSDDTETEDQQLEDVPL	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Bos taurus (Bovine)	9913	HIF1A_BOVIN	HIF1A	Q9XTA5	U9913_1_HIF1A_17	APAAGDTIISLDFGSNDTETD	mat=tab&columns=linea	UniProtKB	8/17/20 9:57
Bos taurus (cow)	9913	hypoxia-inducible factor	HIF1A	bta:281814	K9913_2_HIF1A_18	APAAGDTIISLDFGSNDTETD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Ictalurus punctatus (channel catfish)	7998	ia inducible factor 1 al	HIF1A	ipu:108263197	K7998_1_HIF1A_19	YRMLVKGGGFVWVETQAT	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Ictalurus punctatus (channel catfish)	7998	hypoxia induced factor	HIF1A	ipu:100305068	K7998_2_HIF1A_20	DKSVMDMMKLFKKDPLTC	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Danio rerio (zebrafish)	7955	hypoxia-inducible factor	HIF1A	dre:797150	K7955_1_HIF1A_21	QILSLQQTNSTKIKQEEHQE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Danio rerio (zebrafish)	7955	ia inducible factor 1 su	HIF1A	dre:393202	K7955_2_HIF1A_22	ELDILKLFKPESLNCSLESS	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Rattus norvegicus (Rat)	10116	HIF1A_RAT	HIF1A	O35800	U10116_1_HIF1A_23	PAAGDTIISLDFGSDDTETE	mat=tab&columns=linea	UniProtKB	8/17/20 9:57
Rattus norvegicus (Norway rat)	10116	ypoxia inducible factor	HIF1A	rno:29560	K10116_2_HIF1A_24	LAPAAGDTIISLDFGSDDTE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Oryctolagus cuniculus (rabbit)	9986	xia inducible factor 1 a	HIF1A	ocu:100009579	K9986_1_HIF1A_25	LLAPAAGDTIISLDFGSNDTE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Manis javanica (Malayan pangolin)	9974	ia-inducible factor 1-alp	HIF1A	mjv:108404522	K9974_1_HIF1A_26	LLAPAAGDTIISLDFGGSDTE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Ovis aries (sheep)	9940	ia-inducible factor 1-alp	HIF1A	oas:443519	K9940_1_HIF1A_27	APAAGDTIISLDFGSNDTET	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Capra hircus (goat)	9925	xia inducible factor 1 a	HIF1A	chx:100861391	K9925_1_HIF1A_28	APAAGDTIISLDFGSNDTET	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Bos indicus (zebu cattle)	9915	hypoxia-inducible factor	HIF1A	biu:109565231	K9915_1_HIF1A_29	APAAGDTIISLDFGSNDTETD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Camelus dromedarius (Arabian camel)	9838	ia-inducible factor 1-alp	HIF1A	cdk:105086214	K9838_1_HIF1A_30	SSLFDKLKKEPDALTLLAPA	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Sus scrofa (pig)	9823	hypoxia-inducible factor	HIF1A	ssc:396696	K9823_1_HIF1A_31	APAAGDTIISLDFGSNDTET	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Equus przewalskii (Przewalski's horse)	9798	xia inducible factor 1 a	HIF1A	epz:103562790	K9798_1_HIF1A_32	TLLAPAAGDTIISLDFGSNDT	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Equus caballus (horse)	9796	hypoxia-inducible factor	HIF1A	ecb:100061166	K9796_1_HIF1A_33	APAAGDTIISLDFGSNDTDT	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Equus asinus (ass)	9793	xia inducible factor 1 a	HIF1A	eai:106848533	K9793_1_HIF1A_34	LAPAAGDTIISLDFGSNDTDT	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Loxodonta africana (African savanna elephant)	9785	hypoxia-inducible factor	HIF1A	lav:100659617	K9785_1_HIF1A_35	LAPAAGDTIISLDFGSNDTES	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Physeter catodon (sperm whale)	9755	ia-inducible factor 1-alp	HIF1A	pcad:102978574	K9755_1_HIF1A_36	LAPAAGDTIISLDFGSNDTET	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Delphinapterus leucas (beluga whale)	9749	ia-inducible factor 1-alp	HIF1A	dle:111173087	K9749_1_HIF1A_37	APAAGDTIISLDFGSNDTET	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Orcinus orca (killer whale)	9733	hypoxia-inducible factor	HIF1A	oor:101271390	K9733_1_HIF1A_38	APAAGDTIISLDFGSNDTET	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Odobenus rosmarus divergens (Pacific walrus)	9708	ia-inducible factor 1-alp	HIF1A	oro:101374637	K9708_1_HIF1A_39	LLAPAAGDTIISLDFGSNDTE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Panthera pardus (leopard)	9691	hypoxia-inducible factor	HIF1A	ppad:109268360	K9691_1_HIF1A_40	APAAGDTIISLDFGSNDTET	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Felis catus (domestic cat)	9685	hypoxia-inducible factor	HIF1A	fca:101083686	K9685_1_HIF1A_41	APAAGDTIISLDFGSNDTET	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Ailuropoda melanoleuca (giant panda)	9646	hypoxia-inducible factor	HIF1A	aml:100474777	K9646_1_HIF1A_42	LLAPAAGDTIISLDFGSNDTE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Vulpes vulpes TameXAggressive cross (red fox)	9627	hypoxia-inducible factor	HIF1A	vvp:112921294	K9627_1_HIF1A_43	DKLKKEPDALTLLAPAAGD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Canis lupus familiaris (dog)	9615	hypoxia-inducible factor	HIF1A	cfa:480348	K9615_1_HIF1A_44	LAPAAGDTIISLDFGSNDTE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Pongo abelii (Sumatran orangutan)	9601	hypoxia-inducible factor	HIF1A	pon:100173994	K9601_1_HIF1A_45	PAAGDTIISLDFGSNDTETD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Pan troglodytes (chimpanzee)	9598	ia-inducible factor 1-alp	HIF1A	ptr:746195	K9598_1_HIF1A_46	QTECVLKPVESSDMKMTQL	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Pan paniscus (bonobo)	9597	xia inducible factor 1 a	HIF1A	pps:100993223	K9597_1_HIF1A_47	AAGDTIISLDFGSNDTETDD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Gorilla gorilla gorilla (western lowland gorilla)	9595	ia-inducible factor 1-alp	HIF1A	ggo:101136350	K9595_1_HIF1A_48	PAAGDTIISLDFGSNDTETD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Macaca mulatta (rhesus monkey)	9544	ia-inducible factor 1-alp	HIF1A	mcc:704678	K9544_1_HIF1A_49	QTECVLKPVESSDMKMTQL	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Macaca fascicularis (crab-eating macaque)	9541	xia inducible factor 1 a	HIF1A	mcf:101866092	K9541_1_HIF1A_50	AAGDTIISLDFGSNDTETDD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Callithrix jacchus (white-tufted-ear marmoset)	9483	xia inducible factor 1 a	HIF1A	cjc:100411678	K9483_1_HIF1A_51	VESEDTSSLFDKLKKEPDAL	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Desmodus rotundus (common vampire bat)	9430	hypoxia-inducible factor	HIF1A	dro:112311145	K9430_1_HIF1A_52	LAPAAGDTIISLDFGSNDTE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Rousettus aegyptiacus (Egyptian rousette)	9407	hypoxia-inducible factor	HIF1A	ray:107509336	K9407_1_HIF1A_53	LAPAAGDTIISLDFGSNDTE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Pteropus alecto (black flying fox)	9402	hypoxia-inducible factor	HIF1A	pale:102897284	K9402_1_HIF1A_54	APAAGDTIISLDFGSNDTET	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Coturnix japonica (Japanese quail)	93934	ia-inducible factor 1-alp	HIF1A	cjo:107315399	K93934_1_HIF1A_55	DTNSLFEKLKQEPDALTVLA	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Corvus cornix cornix (hooded crow)	932674	hypoxia-inducible factor	HIF1A	ccw:104687294	K932674_1_HIF1A_56	PAAGDTIISLDFSSNESDEQ	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57

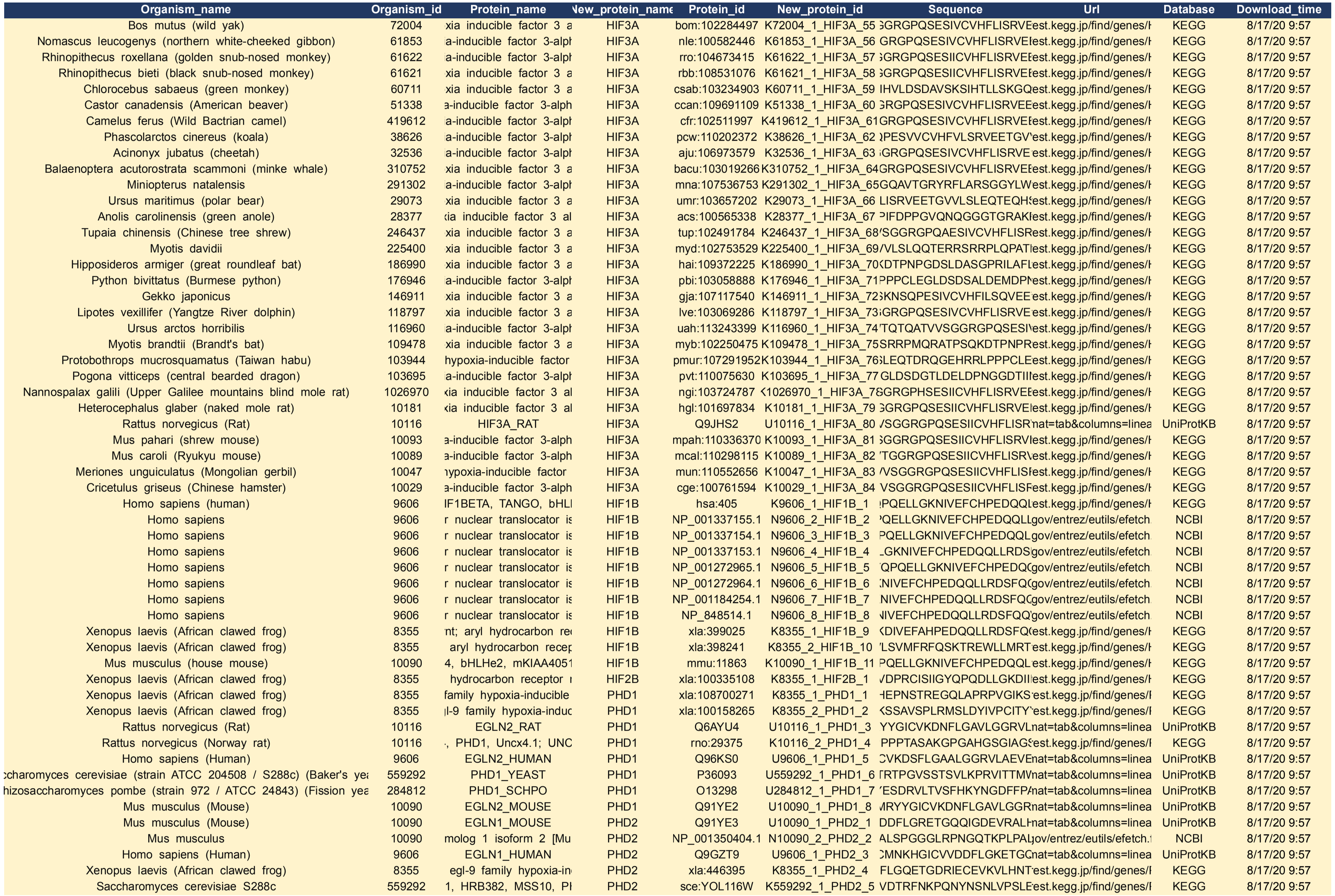
1

Organism_name	Organism_id	Protein_name	ew_protein_name	Protein_id	New_protein_id	Sequence	Url	Database	Download_time
Sarcophilus harrisii (Tasmanian devil)	9305	hypoxia-inducible factor	HIF1A	shr:100931887	K9305_1_HIF1A_57	ALTLLAPAAGDTIISLDFDN	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Ornithorhynchus anatinus (platypus)	9258	ia-inducible factor 1-alp	HIF1A	oaa:100080039	K9258_1_HIF1A_58	GDTIIYLDFGNNETEEQQL	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Pygoscelis adeliae (Adelie penguin)	9238	hypoxia-inducible factor	HIF1A	padl:103924259	K9238_1_HIF1A_59	LFEKLKQEPDALTVLAPAAG	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Parus major (Great Tit)	9157	hypoxia-inducible factor	HIF1A	pmaj:107205385	K9157_1_HIF1A_60	EKLKQEPDALTVLAPAAGDT	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Serinus canaria (common canary)	9135	ia-inducible factor 1-alp	HIF1A	scan:103826360	K9135_1_HIF1A_61	GDTIISLDFSSNESDEQQC	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Meleagris gallopavo (turkey)	9103	hypoxia-inducible factor	HIF1A	mgp:100545267	K9103_1_HIF1A_62	APAAGDTIISLDFSSNESDE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Gallus gallus (Chicken)	9031	HIF1A_CHICK	HIF1A	Q9YIB9	U9031_1_HIF1A_63	DALTVLAPAAGDTIISLDFSS	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Numida meleagris (helmeted guineafowl)	8996	ia-inducible factor 1-alp	HIF1A	nmel:110402046	K8996_1_HIF1A_64	FSKDDLDDTNSLFEKLKQE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Falco peregrinus (peregrine falcon)	8954	xia inducible factor 1 a	HIF1A	fpg:101923837	K8954_1_HIF1A_65	VPLYNDVMLPSSSEKLQNIN	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Bubalus bubalis (water buffalo)	89462	ia-inducible factor 1-alp	HIF1A	bbub:102392717	K89462_1_HIF1A_66	DKLKKEPDALTLLAPAAGDT	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Columba livia (rock pigeon)	8932	hypoxia-inducible factor	HIF1A	clv:102090001	K8932_1_HIF1A_67	EVPLYNDVMLPSSSEKLQN	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Anas platyrhynchos (mallard)	8839	ia-inducible factor 1-alp	HIF1A	apla:101793247	K8839_1_HIF1A_68	DTNNLFEKLKQEPDALTVL	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Alligator mississippiensis (American alligator)	8496	xia inducible factor 1 a	HIF1A	amj:106736850	K8496_1_HIF1A_69	PVERPDTKTTQIFKTDVEDT	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Chrysemys picta bellii (western painted turtle)	8478	xia inducible factor 1 a	HIF1A	cpic:101937400	K8478_1_HIF1A_70	LAPAPGDTIISLDFSSNESD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Chelonia mydas (green sea turtle)	8469	xia inducible factor 1 a	HIF1A	cmy:102942025	K8469_1_HIF1A_71	VPLYNDVMLPSSSEKLQNIN	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Xenopus (Silurana) tropicalis (western clawed frog)	8364	ypoxia-inducible factor	HIF1A	xtr:496583	K8364_1_HIF1A_72	EKLKQEPESLTVLAPDAGD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Paralichthys olivaceus (Japanese flounder)	8255	ypoxia-inducible factor	HIF1A	pov:109628061	K8255_1_HIF1A_73	VVESSSPDMSPTLLKEEEE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Lates calcarifer (barramundi perch)	8187	ypoxia-inducible factor	HIF1A	lcf:108892479	K8187_1_HIF1A_74	QEEEKVVVESSSPDMSPAL	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Oreochromis niloticus (Nile tilapia)	8128	a-inducible factor 1-alph	HIF1A	onl:100703737	K8128_1_HIF1A_75	DQLVVEISQVELSPVQPKK	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Amphiprion ocellaris (clown anemonefish)	80972	ypoxia-inducible factor	HIF1A	aoce:111578922	K80972_1_HIF1A_76	PVKEEQQQEDETIVLSSEPD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Oryzias latipes (Japanese medaka)	8090	a-inducible factor 1-alph	HIF1A	ola:101158737	K8090_1_HIF1A_77	SVKEEPQEAGEVGAESDKT	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Xiphophorus maculatus (southern platyfish)	8083	ypoxia-inducible factor	HIF1A	xma:102217674	K8083_1_HIF1A_78	QQDEVKAEVKKSEAVQAD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Poecilia reticulata (guppy)	8081	ypoxia-inducible factor	HIF1A	pret:103458897	K8081_1_HIF1A_79	QQQQDEVQAEVKKSEAVQ	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri)	8022	HIF1A_ONCMY	HIF1A	Q98SW2	U8022_1_HIF1A_80	EEEESSEPEVSPVLLKEEKS	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Esox lucius (northern pike)	8010	ypoxia-inducible factor	HIF1A	els:105015558	K8010_1_HIF1A_81	KKERIEEEEVEEESSEAEM	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Electrophorus electricus (electric eel)	8005	ypoxia-inducible factor	HIF1A	eee:113576281	K8005_1_HIF1A_82	EEEEEEQQEEEERGEGEN	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Astyanax mexicanus (Mexican tetra)	7994	ypoxia-inducible factor	HIF1A	amex:103022448	K7994_1_HIF1A_83	ELVLSLHQSGSMVVKQEKK	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Latimeria chalumnae (coelacanth)	7897	xia inducible factor 1 a	HIF1A	lcm:102348279	K7897_1_HIF1A_84	PEALTLLAPAAGDTIISLDFS	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Callorhinchus milii (elephant shark)	7868	xia inducible factor 1 al	HIF1A	cmk:103185383	K7868_1_HIF1A_85	EPEALTQLAPAAGDTIISLDF	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Oncorhynchus tshawytscha (Chinook salmon)	74940	ypoxia-inducible factor	HIF1A	otw:112261708	K74940_1_HIF1A_86	VKKELEEEESSEPEVSPVLL	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Panthera tigris altaica (Amur tiger)	74533	xia inducible factor 1 a	HIF1A	ptg:102951241	K74533_1_HIF1A_87	SETPKPLRSSADPALNQEV	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Bos mutus (wild yak)	72004	xia inducible factor 1 a	HIF1A	bom:102288255	K72004_1_HIF1A_88	PAAGDTIISLDFGSNDTETD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Tribolium castaneum (red flour beetle)	7070	milar; hypoxia-inducible	HIF1A	tca:655772	K7070_1_HIF1A_89	SGVECKDEIYSSSQLASVKT	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Podarcis muralis (common wall lizard)	64176	hypoxia-inducible factor	HIF1A	pmua:114602212	K64176_1_HIF1A_90	KLKKEPDALTYLAPAAGDTII	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Nomascus leucogenys (northern white-cheeked gibbon)	61853	ia-inducible factor 1-alp	HIF1A	nle:100588621	K61853_1_HIF1A_91	AAGDTIISLDFGSNDTETDD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Rhinopithecus roxellana (golden snub-nosed monkey)	61622	ia-inducible factor 1-alp	HIF1A	rro:104655242	K61622_1_HIF1A_92	AAGDTIISLDFGSNDTETDD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Rhinopithecus bieti (black snub-nosed monkey)	61621	xia inducible factor 1 a	HIF1A	rbb:108544440	K61621_1_HIF1A_93	AGDTIISLDFGSNDTETDD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Chlorocebus sabaeus (green monkey)	60711	xia inducible factor 1 a	HIF1A	csab:103229134	K60711_1_HIF1A_94	PAAGDTIISLDFGSNDTETD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Ficedula albicollis (collared flycatcher)	59894	xia inducible factor 1 a	HIF1A	fab:101809981	K59894_1_HIF1A_95	KLKQEPDALTVLAPAAGDT	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Taeniopygia guttata (zebra finch)	59729	hypoxia-inducible factor	HIF1A	tgu:100226791	K59729_1_HIF1A_96	LTVLAPAAGDTIISLDFSSN	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Austrofundulus limnaeus (annual killifish)	52670	ypoxia-inducible factor	HIF1A	alim:106521368	K52670_1_HIF1A_97	EEKDQEEEEVVVESSISVLV	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Castor canadensis (American beaver)	51338	a-inducible factor 1-alph	HIF1A	ccan:109682911	K51338_1_HIF1A_98	LFDKLKKEPDALTLLAPAAG	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Geospiza fortis (medium ground-finch)	48883	xia inducible factor 1 a	HIF1A	gfr:102040649	K48883_1_HIF1A_99	ALTVLAPAAGDTIISLDFSS	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Camelus ferus (Wild Bactrian camel)	419612	ia-inducible factor 1-alp	HIF1A	cfr:102514077	419612_1_HIF1A_10	SSLFDKLKKEPDALTLLAPA	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Saimiri boliviensis boliviensis (Bolivian squirrel monkey)	39432	ia-inducible factor 1-alp	HIF1A	sbq:101044689	K39432_1_HIF1A_101	PAAGDTIISLDFGSNDTETD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Alligator sinensis (Chinese alligator)	38654	hypoxia-inducible factor	HIF1A	asn:102386418	K38654_1_HIF1A_102	DVEDTNSLFEKLKQEPDAL	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Phascolarctos cinereus (koala)	38626	hypoxia-inducible factor	HIF1A	pcw:110208444	K38626_1_HIF1A_103	DALTLLAPAAGDTIISLDFDN	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Anser cygnoides domesticus (swan goose)	381198	ia-inducible factor 1-alp	HIF1A	acyg:106037204	381198_1_HIF1A_10	TNNLFEKLKQEPDALTVLAP	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Kryptolebias marmoratus (mangrove rivulus)	37003	ypoxia-inducible factor	HIF1A	kmr:108235270	K37003_1_HIF1A_105	KEEQEEEQVEIESSITIRVG	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Thamnophis sirtalis	35019	hypoxia-inducible factor	HIF1A	tsr:106551842	K35019_1_HIF1A_10	SPEMGMRTELNKSDLEDSS	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Falco cherrug (Saker falcon)	345164	ia-inducible factor 1-alp	HIF1A	fch:102058543	345164_1_HIF1A_10	VPLYNDVMLPSSSEKLQNIN	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Acinonyx jubatus (cheetah)	32536	hypoxia-inducible factor	HIF1A	aju:106971299	K32536_1_HIF1A_108	APAAGDTIISLDFGSNDTET	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Xiphophorus couchianus (Monterrey platyfish)	32473	ypoxia-inducible factor	HIF1A	xco:114134418	K32473_1_HIF1A_10	QQDEVKAEVKKSEAVRADN	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Balaenoptera acutorostrata scammoni (minke whale)	310752	xia inducible factor 1 a	HIF1A	bacu:103008124	310752_1_HIF1A_11	APAAGDTIISLDFGSNDTET	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Takifugu rubripes (torafugu)	31033	ypoxia-inducible factor	HIF1A	tru:101071027	K31033_1_HIF1A_111	ELLQEEEEEKFIIESSSPDIS	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57

Organism_name	Organism_id	Protein_name	New_protein_name	Protein_id	New_protein_id	Sequence	Url	Database	Download_time
Bos mutus grunniens (Wild yak) (Bos grunniens)	30521	HIF1A_BOSMU	HIF1A	Q0PGG7	U30521_1_HIF1A_112	PAAGDTIISLDFGSNDTETD	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Lonchura striata domestica (Bengalese finch)	299123	hypoxia-inducible factor	HIF1A	lsr:110484111	K299123_1_HIF1A_113	LTVLAPAAGDTIISLDFSSNE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Miniopterus natalensis	291302	hypoxia-inducible factor	HIF1A	mna:107525927	K291302_1_HIF1A_114	YNDVMLPSSNEKLQSINLA	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Ursus maritimus (polar bear)	29073	xia inducible factor 1 a	HIF1A	umr:103672464	K29073_1_HIF1A_115	LFDKLKKEPDALTLLAPAAG	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Cyprinodon variegatus (sheepshead minnow)	28743	ypoxia-inducible factor	HIF1A	cvg:107087575	K28743_1_HIF1A_116	QQGEEESAAEANEALKVKE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Canis lupus dingo	286419	ble factor 1-alpha [Cani	HIF1A	XP_025298220.1	N286419_1_HIF1A_117	LAPAAGDTIISLDFGSNDTE	ov/entrez/eutils/efetch.f	NCBI	8/17/20 9:57
Anolis carolinensis (green anole)	28377	xia inducible factor 1 al	HIF1A	acs:100555313	K28377_1_HIF1A_118	EPDALTYLAPAAGDTIIALDF	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Rhincodon typus (whale shark)	259920	a-inducible factor 1-alph	HIF1A	rtp:109915465	K259920_1_HIF1A_119	PAAGDTVVSLDFSPSEMQQ	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Tupaia chinensis (Chinese tree shrew)	246437	xia inducible factor 1 a	HIF1A	tup:102486959	K246437_1_HIF1A_120	LFDKLKKEPDALTLLAPAAG	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Cynoglossus semilaevis (tongue sole)	244447	ypoxia-inducible factor	HIF1A	csem:103376909	K244447_1_HIF1A_121	TEDVKPVKEEQEEEKAEVL	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Myotis davidii	225400	xia inducible factor 1 a	HIF1A	myd:102751343	K225400_1_HIF1A_122	YNDVMLPSSNEKLQSINLAM	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Larimichthys crocea (large yellow croaker)	215358	ypoxia-inducible factor	HIF1A	lco:104921576	K215358_1_HIF1A_123	KEELQLEEEEAVVESSQPE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Apteryx mantelli mantelli (brown kiwi)	202946	xia inducible factor 1 a	HIF1A	aam:106483102	K202946_1_HIF1A_124	APAAGDTIISLDFSSNESDE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Athene cunicularia (burrowing owl)	194338	hypoxia-inducible factor	HIF1A	acun:113480536	K194338_1_HIF1A_125	VPLYNDVMLPSSSEKLQNIN	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Hipposideros armiger (great roundleaf bat)	186990	xia inducible factor 1 a	HIF1A	hai:109383747	K186990_1_HIF1A_126	LAPAAGDTIISLDFGSNDTE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Seriola lalandi dorsalis (Yellowtail amberjack)	1841481	ypoxia-inducible factor	HIF1A	slal:111651106	K1841481_1_HIF1A_12	EKEEEEEEEEEEEEEKAAV	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Pseudopodoces humilis (Tibetan ground-tit)	181119	xia inducible factor 1 a	HIF1A	phi:102099564	K181119_1_HIF1A_128	DALTVLAPAAGDTIISLDFSS	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Python bivittatus (Burmese python)	176946	hypoxia-inducible factor	HIF1A	pbi:103059081	K176946_1_HIF1A_129	DALTYLAPAAGDTIIALDFGN	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Paramormyrops kingsleyae	1676925	a-inducible factor 1-alph	HIF1A	pki:111851026	K1676925_1_HIF1A_13	EEVEEDTHEEPKEDVEGPE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Empidonax traillii (willow flycatcher)	164674	hypoxia-inducible factor	HIF1A	etl:114065853	K164674_1_HIF1A_131	DALTVLAPAAGDTIISLDFSS	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Sinocyclocheilus anshuiensis	1608454	ypoxia-inducible factor	HIF1A	sanh:107697305	K1608454_1_HIF1A_13	EVDILKLFKPESLKCPMESP	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Cyanistes caeruleus (blue tit)	156563	hypoxia-inducible factor	HIF1A	ccae:111929532	K156563_1_HIF1A_133	LKQEPDALTVLAPAAGDTII	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Boleophthalmus pectinirostris (great blue-spotted mudskipper)	150288	ypoxia-inducible factor	HIF1A	bpec:110155253	K150288_1_HIF1A_134	PVKEEQLEAVEEPVEESSP	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Gekko japonicus	146911	xia inducible factor 1 a	HIF1A	gja:107119536	K146911_1_HIF1A_135	TLLAPAAGDTIIALDFGTDDS	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Eospalax fontanierii baileyi (Plateau zokor) (Eospalax baileyi)	146132	HIF1A_EOSFB	HIF1A	Q309Z6	U146132_1_HIF1A_136	LTLLAPAAGDTIISLDFGSDD	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Pelodiscus sinensis (Chinese soft-shelled turtle)	13735	ia-inducible factor 1-alph	HIF1A	pss:102457736	K13735_1_HIF1A_137	TLLAPAPGDTIVSLDFDSNE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Monodelphis domestica (opossum)	13616	xia inducible factor 1 a	HIF1A	mdo:100027212	K13616_1_HIF1A_138	DALTLLAPAAGDTIISLDFDN	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Nipponia nippon (crested ibis)	128390	hypoxia-inducible factor	HIF1A	nni:104010140	K128390_1_HIF1A_139	MANEFKLELVEKLFAIDTEAK	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Nanorana parkeri	125878	ia-inducible factor 1-alph	HIF1A	npr:108791966	K125878_1_HIF1A_140	EPESLTVLAPDAGDEIIALDF	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Lipotes vexillifer (Yangtze River dolphin)	118797	xia inducible factor 1 a	HIF1A	lve:103073387	K118797_1_HIF1A_141	NDVMLPSSNEKLQNINLAM	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Ursus arctos horribilis	116960	ia-inducible factor 1-alph	HIF1A	uah:113242413	K116960_1_HIF1A_142	LFDKLKKEPDALTLLAPAAG	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Scleropages formosus (Asian bonytongue)	113540	ypoxia-inducible factor	HIF1A	sfm:108919562	K113540_1_HIF1A_143	VSEDTEELKFSEPTGLEAD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Myotis brandtii (Brandt's bat)	109478	xia inducible factor 1 a	HIF1A	myb:102245337	K109478_1_HIF1A_144	DTIISLDFGSNDTETDDQQL	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Hippocampus comes (tiger tail seahorse)	109280	ypoxia-inducible factor	HIF1A	hcq:109522901	K109280_1_HIF1A_145	QKDQEKVVVLSCSLEKSPTL	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Maylandia zebra (zebra mbuna)	106582	a-inducible factor 1-alph	HIF1A	mze:101483044	K106582_1_HIF1A_146	EDQLVVEISQVELSPVQPKK	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Nothobranchius furzeri (turquoise killifish)	105023	ypoxia-inducible factor	HIF1A	nfu:107392302	K105023_1_HIF1A_147	EEQEEEELVVETTQLIPVKA	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Protobothrops mucrosquamatus (Taiwan habu)	103944	hypoxia-inducible factor	HIF1A	pmur:107294915	K103944_1_HIF1A_148	PLYNDVMLPSSNEKLQNINL	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Pogona vitticeps (central bearded dragon)	103695	ia-inducible factor 1-alph	HIF1A	pvt:110090015	K103695_1_HIF1A_149	KKEPDALTYLAPAAGDTIIAL	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Nannospalax galili (Upper Galilee mountains blind mole rat)	1026970	xia inducible factor 1 al	HIF1A	ngi:103747967	K1026970_1_HIF1A_15	LLAPAAGDTIISLDFGSDDTE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Heterocephalus glaber (naked mole rat)	10181	xia inducible factor 1 al	HIF1A	hgl:101715845	K10181_1_HIF1A_151	AGDIIISLDFSSNDTETDDQ	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Mus pahari (shrew mouse)	10093	a-inducible factor 1-alph	HIF1A	mpah:110323979	K10093_1_HIF1A_152	PAAGDTIISLDFGSDDTETE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Mus caroli (Ryukyu mouse)	10089	a-inducible factor 1-alph	HIF1A	mcal:110306746	K10089_1_HIF1A_153	AAGDTIISLDFGSDGQCLDT	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Meriones unguiculatus (Mongolian gerbil)	10047	a-inducible factor 1-alph	HIF1A	mun:110556968	K10047_1_HIF1A_154	AAGDTIISLDFGSDDTETED	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Gallus gallus	9031	omain-containing protei	HIF2A	NP_990138.1	N9031_1_HIF2A_1	YTSAVLTPNKPWPVEVKSH	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Gallus gallus	9031	-containing protein 1 is	HIF2A	XP_015139105.1	N9031_2_HIF2A_2	YTSAVLTPNKPWPVEVKSH	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Gallus gallus	9031	-containing protein 1 is	HIF2A	XP_015139104.1	N9031_3_HIF2A_3	APAYTSAVLTPNKPWPVEVK	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Gallus gallus	9031	domain protein 1, parti	HIF2A	ASF89071.1	N9031_4_HIF2A_4	VLTPNKPWPVEVKSHAAQG	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Homo sapiens (Human)	9606	EPAS1_HUMAN	HIF2A	Q99814	U9606_1_HIF2A_5	GKAILPPSQPWATELRSHST	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Homo sapiens	9606	containing protein 1 iso	HIF2A	XP_011531000.1	N9606_2_HIF2A_6	EESSAYGKAILPPSQPWATE	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Xenopus laevis (African clawed frog)	8355	a, hlf, mop2, pasd2; er	HIF2A	xla:443990	K8355_1_HIF2A_7	SFNSVQTSPSKQWPLEVKN	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Xenopus laevis (African clawed frog)	8355	alpha, hlf, mop2, pasd2	HIF2A	xla:444145	K8355_2_HIF2A_8	ISSVQTSPSKPWPLEVKNLN	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Xenopus tropicalis	8364	ain-containing protein 1	HIF2A	NP_001005647.1	N8364_1_HIF2A_9	SVQTSPSKQWPLEVKNLNS	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Phrynocephalus erythrurus	500582	e factor 2a [Phrynoceph	HIF2A	ALS35218.1	N500582_1_HIF2A_10	PNKPWPPEVKNQTPQPENL	ov/entrez/eutils/efetch.f	NCBI	8/17/20 9:57
Phrynocephalus przewalskii	171649	e factor 2a [Phrynoceph	HIF2A	ALS35219.1	N171649_1_HIF2A_11	PNKPWPPEVKNQTPQLEKL	ov/entrez/eutils/efetch.f	NCBI	8/17/20 9:57
Rattus norvegicus (Rat)	10116	EPAS1_RAT	HIF2A	Q9JHS1	U10116_1_HIF2A_12	LPPGQPWATELRSHSAQSE	nat=tab&columns=linea	UniProtKB	8/17/20 9:57

3

Organism_name	Organism_id	Protein_name	New_protein_name	Protein_id	New_protein_id	Sequence	Url	Database	Download_time
Mus musculus (Mouse)	10090	EPAS1_MOUSE	HIF2A	P97481	U10090_1_HIF2A_13	PPGQPWVSGLRSHSAQSES	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Homo sapiens (Human)	9606	HIF3A_HUMAN	HIF3A	Q9Y2N7	U9606_1_HIF3A_1	GRGPQSESIVCVHFLISQVE	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Homo sapiens	9606	factor 3-alpha isoform	HIF3A	NP_690007.1	N9606_2_HIF3A_2	RGPQSESIVCVHFLISQVEE	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Homo sapiens	9606	factor 3-alpha isoform	HIF3A	NP_071907.4	N9606_3_HIF3A_3	RRPIQRGAPSQKDTPNPGD	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Homo sapiens	9606	factor 3-alpha isoform	HIF3A	NP_690009.1	N9606_4_HIF3A_4	AVSKSIHTLLSKGQAVTGQY	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Homo sapiens	9606	actor 3-alpha isoform X	HIF3A	XP_024307417.1	N9606_5_HIF3A_5	KDTPNPGDSLDTPGPRILAF	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Homo sapiens	9606	actor 3-alpha isoform X	HIF3A	XP_016882631.1	N9606_6_HIF3A_6	QAVTGQYRFLARSGGYLWT	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Homo sapiens	9606	actor 3-alpha isoform X	HIF3A	XP_016882630.1	N9606_7_HIF3A_7	LDTPGPRILAFLHPPSLSEA	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Homo sapiens	9606	actor 3-alpha isoform X	HIF3A	XP_016882627.1	N9606_8_HIF3A_8	VSGGRGPQSESIVCVHFLIS	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Homo sapiens	9606	actor 3-alpha isoform X	HIF3A	XP_016882626.1	N9606_9_HIF3A_9	EETGVVLSLEQTEQHSRRP	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Homo sapiens	9606	HIF3A-6_HUMAN	HIF3A	Q9Y2N7-6	E9606_10_HIF3A_10	APSQKDTPNPGDSLDTPGP	ins?offset=0&size=-1&r	EMBL	8/17/20 9:57
Homo sapiens	9606	HIF3A-3_HUMAN	HIF3A	Q9Y2N7-3	E9606_11_HIF3A_11	ARSGGYLWTQTQATVVSGG	ins?offset=0&size=-1&r	EMBL	8/17/20 9:57
Homo sapiens	9606	HIF3A-5_HUMAN	HIF3A	Q9Y2N7-5	E9606_12_HIF3A_12	QTEQHSRRPIQRGAPSQKD	ins?offset=0&size=-1&r	EMBL	8/17/20 9:57
Homo sapiens	9606	HIF3A-4_HUMAN	HIF3A	Q9Y2N7-4	E9606_13_HIF3A_13	TLTSRGRTLNLKAATWKVLN	ins?offset=0&size=-1&r	EMBL	8/17/20 9:57
Mus musculus (Mouse)	10090	HIF3A_MOUSE	HIF3A	Q0VBL6	U10090_1_HIF3A_14	SGGRGPQSESIICVHFLISRV	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Mus musculus	10090	actor 3-alpha isoform X	HIF3A	XP_011248930.1	N10090_2_HIF3A_15	SLEQTEQHTRRPPRLSASSQ	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Mus musculus	10090	factor 3-alpha isoform	HIF3A	NP_001156422.1	N10090_3_HIF3A_16	VSGGRGPQSESIICVHFLISR	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Mus musculus	10090	HIF3A-2_MOUSE	HIF3A	Q0VBL6-2	E10090_4_HIF3A_17	IFDFIHPCDQEELQDALTPRP	ins?offset=0&size=-1&r	EMBL	8/17/20 9:57
Callithrix jacchus (white-tufted-ear marmoset)	9483	xia inducible factor 3 a	HIF3A	cjc:100413902	K9483_1_HIF3A_18	QTEQHSRRPIQRGSPSQKD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Callithrix jacchus	9483	ctor 3-alpha isoform X4	HIF3A	XP_008986513.1	N9483_2_HIF3A_19	RRPIQRGSPSQKDASNSGG	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Manis javanica (Malayan pangolin)	9974	ia-inducible factor 3-alp	HIF3A	mjv:108398976	K9974_1_HIF3A_20	TQTQATVVSGGRGPQSESI	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Ovis aries (sheep)	9940	ia-inducible factor 3-alp	HIF3A	oas:100135687	K9940_1_HIF3A_21	VTQTQATVVSGGRGPQSESI	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Capra hircus (goat)	9925	xia inducible factor 3 a	HIF3A	chx:102168854	K9925_1_HIF3A_22	TQTQATVVSGGRGPQSESI	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Bos indicus (zebu cattle)	9915	ia-inducible factor 3-alp	HIF3A	biu:109572549	K9915_1_HIF3A_23	VTQTQATVVSGGRGPQSESI	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Bos taurus (cow)	9913	hypoxia-inducible factor	HIF3A	bta:507699	K9913_1_HIF3A_24	GGRGPQSESIVCVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Camelus dromedarius (Arabian camel)	9838	ia-inducible factor 3-alp	HIF3A	cdk:105104022	K9838_1_HIF3A_25	TQTQATVVSGGRGPQSESIV	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Sus scrofa (pig)	9823	ia-inducible factor 3-alp	HIF3A	ssc:100621574	K9823_1_HIF3A_26	ALDSDAVGQSIHALLSKGQA	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Equus przewalskii (Przewalski's horse)	9798	xia inducible factor 3 a	HIF3A	epz:103566291	K9798_1_HIF3A_27	GRGPQSESIVCVHFLISRVEE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Equus caballus (horse)	9796	ia-inducible factor 3-alp	HIF3A	ecb:100071095	K9796_1_HIF3A_28	TQTQATVVSGGRGPQSESI	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Equus asinus (ass)	9793	xia inducible factor 3 a	HIF3A	eai:106845173	K9793_1_HIF3A_29	LWTQTQATVVSGGRGPQS	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Physeter catodon (sperm whale)	9755	hypoxia-inducible factor	HIF3A	pcad:102987456	K9755_1_HIF3A_30	GRGPQSESIVCVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Delphinapterus leucas (beluga whale)	9749	ia-inducible factor 3-alp	HIF3A	dle:111180713	K9749_1_HIF3A_31	GRGPQSESIVCVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Orcinus orca (killer whale)	9733	hypoxia-inducible factor	HIF3A	oor:101281082	K9733_1_HIF3A_32	GRGPQSESIVCVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Odobenus rosmarus divergens (Pacific walrus)	9708	ia-inducible factor 3-alp	HIF3A	oro:101372866	K9708_1_HIF3A_33	GRGPQSESIVCVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Panthera pardus (leopard)	9691	ia-inducible factor 3-alp	HIF3A	ppad:109253198	K9691_1_HIF3A_34	AGYSPDDLIGCSAYEYIHAL	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Felis catus (domestic cat)	9685	ia-inducible factor 3-alp	HIF3A	fca:101091575	K9685_1_HIF3A_35	SRRPAQRGTPSQKDAPDPG	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Ailuropoda melanoleuca (giant panda)	9646	ia-inducible factor 3-alp	HIF3A	aml:100473029	K9646_1_HIF3A_36	GRAPQSESIVCVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Vulpes vulpes TameXAggressive cross (red fox)	9627	ia-inducible factor 3-alp	HIF3A	vvp:112931230	K9627_1_HIF3A_37	VTQTQATVVSGGRGPHSESI	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Canis lupus familiaris (dog)	9615	ia-inducible factor 3-alp	HIF3A	cfa:476429	K9615_1_HIF3A_38	HFLISRVEETGVVLSLEQTER	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Pongo abelii (Sumatran orangutan)	9601	ia-inducible factor 3-alp	HIF3A	pon:100442498	K9601_1_HIF3A_39	RRPIQQGAPSQKDTPNPGD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Pan troglodytes (chimpanzee)	9598	a-inducible factor 3-alph	HIF3A	ptr:456149	K9598_1_HIF3A_40	SRRPIQRGAPSQKDTPNPAD	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Pan paniscus (bonobo)	9597	xia inducible factor 3 a	HIF3A	pps:100979253	K9597_1_HIF3A_41	GRGPQSESIVCVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Gorilla gorilla gorilla (western lowland gorilla)	9595	ia-inducible factor 3-alp	HIF3A	ggo:101146641	K9595_1_HIF3A_42	GRGPQSESIVCVHFLISRVEE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Macaca mulatta (rhesus monkey)	9544	ia-inducible factor 3-alp	HIF3A	mcc:717188	K9544_1_HIF3A_43	GRGPQSESIVCVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Macaca fascicularis (crab-eating macaque)	9541	xia inducible factor 3 a	HIF3A	mcf:102138103	K9541_1_HIF3A_44	GRGPQSESIVCVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Desmodus rotundus (common vampire bat)	9430	hypoxia-inducible factor	HIF3A	dro:112312920	K9430_1_HIF3A_45	GRGPQSESIICVHFLISRVEE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Rousettus aegyptiacus (Egyptian rousette)	9407	ia-inducible factor 3-alp	HIF3A	ray:107515999	K9407_1_HIF3A_46	TQTQATVVSGGRGPQSESI	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Pteropus alecto (black flying fox)	9402	hypoxia-inducible factor	HIF3A	pale:102896353	K9402_1_HIF3A_47	GGRGPQSESIVCVHFLISLVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Sarcophilus harrisii (Tasmanian devil)	9305	ia-inducible factor 3-alp	HIF3A	shr:100927264	K9305_1_HIF3A_48	AVSKSIHTLLSKGQAVTGQYR	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Gallus gallus (chicken)	9031	AS domain-containing p	HIF3A	gga:107049602	K9031_1_HIF3A_49	SAGGGAALRCLVLICEPIAHP	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Bubalus bubalis (water buffalo)	89462	ia-inducible factor 3-alp	HIF3A	bbub:102408468	K89462_1_HIF3A_50	GRGPQSESIVCVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Alligator mississippiensis (American alligator)	8496	xia inducible factor 3 a	HIF3A	amj:102558818	K8496_1_HIF3A_51	CDDRMVEVAGYTPAELLGCS	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Chrysemys picta bellii (western painted turtle)	8478	xia inducible factor 3 a	HIF3A	cpic:101943880	K8478_1_HIF3A_52	PEPDPARQANGSPEELEPDT	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Chelonia mydas (green sea turtle)	8469	xia inducible factor 3 a	HIF3A	cmy:102946043	K8469_1_HIF3A_53	PDPAHQANGSPEESEPDTG	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Panthera tigris altaica (Amur tiger)	74533	xia inducible factor 3 a	HIF3A	ptg:102960124	K74533_1_HIF3A_54	HFLISRVEETGVVLSLEQTER	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57



Organism_name	Organism_id	Protein_name	ew_protein_name	Protein_id	New_protein_id	Sequence	Url	Database	Download_time
Bos mutus (wild yak)	72004	xia inducible factor 3 a	HIF3A	bom:102284497	K72004_1_HIF3A_55	GRGPQSESIVCVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Nomascus leucogenys (northern white-cheeked gibbon)	61853	a-inducible factor 3-alph	HIF3A	nle:100582446	K61853_1_HIF3A_56	GRGPQSESIVCVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Rhinopithecus roxellana (golden snub-nosed monkey)	61622	a-inducible factor 3-alph	HIF3A	rro:104673415	K61622_1_HIF3A_57	GRGPQSESIICVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Rhinopithecus bieti (black snub-nosed monkey)	61621	xia inducible factor 3 a	HIF3A	rbb:108531076	K61621_1_HIF3A_58	GRGPQSESIICVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Chlorocebus sabaeus (green monkey)	60711	xia inducible factor 3 a	HIF3A	csab:103234903	K60711_1_HIF3A_59	IHVLDSDAVSKSIHTLLSKGQ	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Castor canadensis (American beaver)	51338	a-inducible factor 3-alph	HIF3A	ccan:109691109	K51338_1_HIF3A_60	RGPQSESIVCVHFLISRVEE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Camelus ferus (Wild Bactrian camel)	419612	a-inducible factor 3-alph	HIF3A	cfr:102511997	K419612_1_HIF3A_61	GRGPQSESIVCVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Phascolarctos cinereus (koala)	38626	a-inducible factor 3-alph	HIF3A	pcw:110202372	K38626_1_HIF3A_62	PESVVCVHFVLSRVEETGV	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Acinonyx jubatus (cheetah)	32536	a-inducible factor 3-alph	HIF3A	aju:106973579	K32536_1_HIF3A_63	GRGPQSESIVCVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Balaenoptera acutorostrata scammoni (minke whale)	310752	xia inducible factor 3 a	HIF3A	bacu:103019266	K310752_1_HIF3A_64	GRGPQSESIVCVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Miniopterus natalensis	291302	a-inducible factor 3-alph	HIF3A	mna:107536753	K291302_1_HIF3A_65	GQAVTGRYRFLARSGGYLW	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Ursus maritimus (polar bear)	29073	xia inducible factor 3 a	HIF3A	umr:103657202	K29073_1_HIF3A_66	LISRVEETGVVLSLEQTEQH	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Anolis carolinensis (green anole)	28377	ia inducible factor 3 al	HIF3A	acs:100565338	K28377_1_HIF3A_67	IFDPPGVQNQGGGTGRAK	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Tupaia chinensis (Chinese tree shrew)	246437	xia inducible factor 3 a	HIF3A	tup:102491784	K246437_1_HIF3A_68	SGGRGPQAESIVCVHFLISR	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Myotis davidii	225400	xia inducible factor 3 a	HIF3A	myd:102753529	K225400_1_HIF3A_69	VLSLQQTERRSRRPLQPAT	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Hipposideros armiger (great roundleaf bat)	186990	xia inducible factor 3 a	HIF3A	hai:109372225	K186990_1_HIF3A_70	DTPNPGDSLDASGPRILAF	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Python bivittatus (Burmese python)	176946	a-inducible factor 3-alph	HIF3A	pbi:103058888	K176946_1_HIF3A_71	PPPCLEGLDSDSALDEMDP	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Gekko japonicus	146911	xia inducible factor 3 a	HIF3A	gja:107117540	K146911_1_HIF3A_72	KNSQPESIVCVHFILSQVEE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Lipotes vexillifer (Yangtze River dolphin)	118797	xia inducible factor 3 a	HIF3A	lve:103069286	K118797_1_HIF3A_73	GRGPQSESIVCVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Ursus arctos horribilis	116960	a-inducible factor 3-alph	HIF3A	uah:113243399	K116960_1_HIF3A_74	TQTQATVVSGGRGPQSESI	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Myotis brandtii (Brandt's bat)	109478	xia inducible factor 3 a	HIF3A	myb:102250475	K109478_1_HIF3A_75	SRRPMQRATPSQKDTPNPR	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Protobothrops mucrosquamatus (Taiwan habu)	103944	hypoxia-inducible factor	HIF3A	pmur:107291952	K103944_1_HIF3A_76	LEQTDRQGEHRRLPPPCLE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Pogona vitticeps (central bearded dragon)	103695	a-inducible factor 3-alph	HIF3A	pvt:110075630	K103695_1_HIF3A_77	GLDSDGTLDELDPNGGDTII	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Nannospalax galili (Upper Galilee mountains blind mole rat)	1026970	xia inducible factor 3 al	HIF3A	ngi:103724787	1026970_1_HIF3A_7	GRGPHSESIICVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Heterocephalus glaber (naked mole rat)	10181	xia inducible factor 3 al	HIF3A	hgl:101697834	K10181_1_HIF3A_79	GRGPQSESIICVHFLISRVE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Rattus norvegicus (Rat)	10116	HIF3A_RAT	HIF3A	Q9JHS2	U10116_1_HIF3A_80	VSGGRGPQSESIICVHFLISR	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Mus pahari (shrew mouse)	10093	a-inducible factor 3-alph	HIF3A	mpah:110336370	K10093_1_HIF3A_81	GGRGPQSESIICVHFLISRV	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Mus caroli (Ryukyu mouse)	10089	a-inducible factor 3-alph	HIF3A	mcal:110298115	K10089_1_HIF3A_82	TGGRGPQSESIICVHFLISR	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Meriones unguiculatus (Mongolian gerbil)	10047	ypoxia-inducible factor	HIF3A	mun:110552656	K10047_1_HIF3A_83	VSGGRGPQSESIICVHFLIS	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Cricetulus griseus (Chinese hamster)	10029	a-inducible factor 3-alph	HIF3A	cge:100761594	K10029_1_HIF3A_84	VSGGRGPQSESIICVHFLIS	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Homo sapiens (human)	9606	IF1BETA, TANGO, bHL	HIF1B	hsa:405	K9606_1_HIF1B_1	PQELLGKNIVEFCHPEDQQL	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Homo sapiens	9606	r nuclear translocator i	HIF1B	NP_001337155.1	N9606_2_HIF1B_2	QELLGKNIVEFCHPEDQQLL	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Homo sapiens	9606	r nuclear translocator i	HIF1B	NP_001337154.1	N9606_3_HIF1B_3	PQELLGKNIVEFCHPEDQQL	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Homo sapiens	9606	r nuclear translocator i	HIF1B	NP_001337153.1	N9606_4_HIF1B_4	GKNIVEFCHPEDQQLLRDS	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Homo sapiens	9606	r nuclear translocator i	HIF1B	NP_001272965.1	N9606_5_HIF1B_5	QPQELLGKNIVEFCHPEDQ	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Homo sapiens	9606	r nuclear translocator i	HIF1B	NP_001272964.1	N9606_6_HIF1B_6	NIVEFCHPEDQQLLRDSFQ	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Homo sapiens	9606	r nuclear translocator i	HIF1B	NP_001184254.1	N9606_7_HIF1B_7	NIVEFCHPEDQQLLRDSFQ	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Homo sapiens	9606	r nuclear translocator i	HIF1B	NP_848514.1	N9606_8_HIF1B_8	IVEFCHPEDQQLLRDSFQQ	gov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Xenopus laevis (African clawed frog)	8355	nt; aryl hydrocarbon re	HIF1B	xla:399025	K8355_1_HIF1B_9	DIVEFAHPEDQQLLRDSFQ	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Xenopus laevis (African clawed frog)	8355	aryl hydrocarbon recep	HIF1B	xla:398241	K8355_2_HIF1B_10	LSVMFRFQSKTREWLLMRT	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Mus musculus (house mouse)	10090	4, bHLHe2, mKIAA4051	HIF1B	mmu:11863	K10090_1_HIF1B_11	PQELLGKNIVEFCHPEDQQL	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Xenopus laevis (African clawed frog)	8355	hydrocarbon receptor	HIF2B	xla:100335108	K8355_1_HIF2B_1	DPRCISIIGYQPQDLLGKDII	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Xenopus laevis (African clawed frog)	8355	family hypoxia-inducible	PHD1	xla:108700271	K8355_1_PHD1_1	EPNSTREGQLAPRPVGIKS	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Xenopus laevis (African clawed frog)	8355	l-9 family hypoxia-induc	PHD1	xla:100158265	K8355_2_PHD1_2	SSAVSPLRMSLDYIVPCITY	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Rattus norvegicus (Rat)	10116	EGLN2_RAT	PHD1	Q6AYU4	U10116_1_PHD1_3	YYGICVKDNFLGAVLGGRVL	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Rattus norvegicus (Norway rat)	10116	, PHD1, Uncx4.1; UNC	PHD1	rno:29375	K10116_2_PHD1_4	PPPTASAKGPGAHGSGIAG	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Homo sapiens (Human)	9606	EGLN2_HUMAN	PHD1	Q96KS0	U9606_1_PHD1_5	VKDSFLGAALGGRVLAEVE	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
ccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's ye	559292	PHD1_YEAST	PHD1	P36093	U559292_1_PHD1_6	RTPGVSSTSVLKPRVITTM	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
hizosaccharomyces pombe (strain 972 / ATCC 24843) (Fission ye	284812	PHD1_SCHPO	PHD1	O13298	U284812_1_PHD1_7	ESDRVLTVSFHKYNGDFFP	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Mus musculus (Mouse)	10090	EGLN2_MOUSE	PHD1	Q91YE2	U10090_1_PHD1_8	MRYYGICVKDNFLGAVLGGR	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Mus musculus (Mouse)	10090	EGLN1_MOUSE	PHD2	Q91YE3	U10090_1_PHD2_1	DDFLGRETGQQIGDEVRALH	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Mus musculus	10090	molog 1 isoform 2 [Mu	PHD2	NP_001350404.1	N10090_2_PHD2_2	ALSPGGGLRPNGQTKPLPAL	ov/entrez/eutils/efetch.	NCBI	8/17/20 9:57
Homo sapiens (Human)	9606	EGLN1_HUMAN	PHD2	Q9GZT9	U9606_1_PHD2_3	MNKHGICVVDDFLGKETG	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Xenopus laevis (African clawed frog)	8355	egl-9 family hypoxia-in	PHD2	xla:446395	K8355_1_PHD2_4	FLGQETGDRIECEVKVLHNT	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Saccharomyces cerevisiae S288c	559292	1, HRB382, MSS10, P	PHD2	sce:YOL116W	K559292_1_PHD2_5	VDTRFNKPQNYNSNLVPSLE	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57

Organism_name	Organism_id	Protein_name	ew_protein_nam	Protein_id	New_protein_id	Sequence	Url	Database	Download_time
Rattus norvegicus (Rat)	10116	EGLN1_RAT	PHD2	P59722	U10116_1_PHD2_6	KSDSSKDIRGDKITWIEGKE	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Xenopus laevis (African clawed frog)	8355	family hypoxia-inducible	PHD3	xla:100036879	K8355_1_PHD3_1	GNRLGQYYVKERSKAMVA	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Xenopus laevis (African clawed frog)	8355	-9 family hypoxia-induci	PHD3	xla:100127282	K8355_2_PHD3_2	GNRLGQYYVKERSKAMVA	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Rattus norvegicus (Rat)	10116	EGLN3_RAT	PHD3	Q62630	U10116_1_PHD3_3	YYVKERSKAMVACYPGNGT	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Rattus norvegicus (Norway rat)	10116	SM-20; egl-9 family hy	PHD3	rno:54702	K10116_2_PHD3_4	NGALRDGQLAGPRAGVSKF	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Mus musculus (Mouse)	10090	EGLN3_MOUSE	PHD3	Q91UZ4	U10090_1_PHD3_5	YYVKERSKAMVACYPGNGT	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Mus musculus (house mouse)	10090	2, Hif-p4h-3, Phd3, SM	PHD3	mmu:112407	K10090_2_PHD3_6	YYVKERSKAMVACYPGNGT	est.kegg.jp/find/genes/	KEGG	8/17/20 9:57
Homo sapiens (Human)	9606	EGLN3_HUMAN	PHD3	Q9H6Z9	U9606_1_PHD3_7	YYVKERSKAMVACYPGNG	nat=tab&columns=linea	UniProtKB	8/17/20 9:57
Homo sapiens (human)	9606	-TM, PH-4, PH4, PHD	PHD4	hsa:54681	K9606_1_PHD4_1	LSLQEFSNMDLRDFHKYMR	est.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Homo sapiens	9606	yl 4-hydroxylase isofor	PHD4	NP_808807.2	N9606_2_PHD4_2	SHHTWLYQGEGAHHIMRAIF	gov/entrez/eutils/efetch.	NCBI	9/15/20 17:25
Homo sapiens (Human)	9606	HIF1N_HUMAN	FIH	Q9NWT6	U9606_1_FIH_1	NWNWINKQQGKRGWGQLT	at=tab&columns=lineag	UniProtKB	9/15/20 17:25
Homo sapiens (human)	9606	, HHC1, HYPOC1, NS	FIH	hsa:846	K9606_2_FIH_2	LWSGFSREVPFSNCSRDCL	/rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Xenopus laevis (African clawed frog)	8355	cible factor 1, alpha s	FIH	xla:108696241	K8355_1_FIH_3	FNWNWINKQQAKHGWGQL	/rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Xenopus laevis (African clawed frog)	8355	inducible factor 1, alph	FIH	xla:432095	K8355_2_FIH_4	FNWNWINKQQAKHGWGQL	/rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Danio rerio (Zebrafish) (Brachydanio rerio)	7955	HIF1N_DANRE	FIH	P59723	U7955_1_FIH_5	QQAKRNWGPLTSNLLLIGM	at=tab&columns=lineag	UniProtKB	9/15/20 17:25
Strongylocentrotus purpuratus	7668	g HIF [Strongylocentrot	FIH	ASL69981.1	N7668_1_FIH_6	YQKKNSWGPLTSNLLLVSME	gov/entrez/eutils/efetch.	NCBI	9/15/20 17:25
Siniperca chuatsi	119488	yl hydroxylase [Siniperc	FIH	QDA76730.1	N119488_1_FIH_7	NWINKQQAKRNWGQLTSNL	ov/entrez/eutils/efetch.f	NCBI	9/15/20 17:25
Mus musculus (Mouse)	10090	HIF1N_MOUSE	FIH	Q8BLR9	U10090_1_FIH_8	WNWINKQQGKRGWGQLT	at=tab&columns=lineag	UniProtKB	9/15/20 17:25
Homo sapiens (Human)	9606	VHL_HUMAN	VHL	P40337	U9606_1_VHL_1	QPYPTLPPGTGRRIHSYRG	at=tab&columns=linea	UniProtKB	9/15/20 17:25
Homo sapiens (Human)	9606	VHLL_HUMAN	VHL	Q6RSH7	U9606_2_VHL_2	NHSPRIVLPVWLNYYGKLL	at=tab&columns=linea	UniProtKB	9/15/20 17:25
Homo sapiens	9606	se tumor suppressor is	VHL	NP_001341652.1	N9606_3_VHL_3	WLNFDGEPQPYPTLPPGT	gov/entrez/eutils/efetch.	NCBI	9/15/20 17:25
Homo sapiens	9606	se tumor suppressor is	VHL	NP_937799.1	N9606_4_VHL_4	FCNRSPRVVLPVWLNFDGE	gov/entrez/eutils/efetch.	NCBI	9/15/20 17:25
Homo sapiens	9606	r supressor isoform 1,	VHL	AFS51418.1	N9606_5_VHL_5	EELGAEEEMEAGRPRPVLR	gov/entrez/eutils/efetch.	NCBI	9/15/20 17:25
Homo sapiens	9606	L, partial [Homo sapie	VHL	AAA98614.1	N9606_6_VHL_6	YRRLDIVRSLYEDLEDHPNV	gov/entrez/eutils/efetch.	NCBI	9/15/20 17:25
Homo sapiens	9606	VHL-3_HUMAN	VHL	P40337-3	E9606_7_VHL_7	GTHDGLLVNQTELFVPSLN	eins?offset=0&size=-1&	EMBL	9/15/20 17:25
Canis lupus familiaris (Dog) (Canis familiaris)	9615	VHL_CANLF	VHL	Q5Q9Z2	U9615_1_VHL_8	PTLPPGTGRRIHSYRGHLW	at=tab&columns=linea	UniProtKB	9/15/20 17:25
Canis lupus familiaris	9615	umor suppressor isofor	VHL	XP_013977009.1	N9615_2_VHL_9	FDGEPQPYSTLPPGTGRRI	gov/entrez/eutils/efetch.	NCBI	9/15/20 17:25
Pan paniscus (bonobo)	9597	Hippel-Lindau tumor su	VHL	pps:100980663	K9597_1_VHL_10	NHSPRIVLPVWLNYYGELL	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Pan paniscus (bonobo)	9597	Hippel-Lindau tumor s	VHL	pps:100986949	K9597_2_VHL_11	QPYPTLPPGTGRRIHSYRG	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Macaca fascicularis (crab-eating macaque)	9541	Hippel-Lindau tumor su	VHL	mcf:102137687	K9541_1_VHL_12	CYEELLPGRDFCIHNFRSH	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Macaca fascicularis (crab-eating macaque)	9541	Hippel-Lindau tumor s	VHL	mcf:107128974	K9541_2_VHL_13	SLNVDGQPIFANITLPVYTL	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Danio rerio (zebrafish)	7955	el-Lindau disease tum	VHL	dre:791202	K7955_1_VHL_14	PMVVNNKEMYLPASLENG	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Danio rerio (zebrafish)	7955	von Hippel-Lindau-like p	VHL	dre:795635	K7955_2_VHL_15	SDGAKLLVSFSEVYFPAPAQ	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Rhinopithecus bieti (black snub-nosed monkey)	61621	Hippel-Lindau tumor s	VHL	rbb:108512296	K61621_1_VHL_16	QPYPTLPPGTGRRIHSYRG	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Rhinopithecus bieti (black snub-nosed monkey)	61621	Hippel-Lindau tumor su	VHL	rbb:108515508	K61621_2_VHL_17	RDFCIHNFRSHPWLFRDAR	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Chlorocebus sabaeus (green monkey)	60711	Hippel-Lindau tumor su	VHL	csab:103223849	K60711_1_VHL_18	PVWLNCYEELLPGRDFCIH	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Chlorocebus sabaeus (green monkey)	60711	Hippel-Lindau tumor s	VHL	csab:103227999	K60711_2_VHL_19	QPYPTLPPGTGRRIHSYRG	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Oryctolagus cuniculus (rabbit)	9986	Hippel-Lindau tumor s	VHL	ocu:100350053	K9986_1_VHL_20	PVWLNFHGEPQPYPTLPP	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Manis javanica (Malayan pangolin)	9974	dau disease tumor su	VHL	mjv:108408138	K9974_1_VHL_21	SGPGTTEKFFQAEVQICRR	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Ovis aries (sheep)	9940	pel-Lindau disease tum	VHL	oas:101122897	K9940_1_VHL_22	SYRGHLWLFRDAGTSDGLL	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Capra hircus (goat)	9925	Hippel-Lindau tumor s	VHL	chx:102182921	K9925_1_VHL_23	SYRGHLWLFRDAGTSDGLL	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Bos indicus (zebu cattle)	9915	pel-Lindau disease tum	VHL	biu:109576248	K9915_1_VHL_24	SYRGHLWLFRDAGTSDGLL	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Bos taurus (cow)	9913	pel-Lindau disease tum	VHL	bta:540957	K9913_1_VHL_25	SYRGHLWLFRDAGTSDGLL	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Camelus dromedarius (Arabian camel)	9838	pel-Lindau disease tum	VHL	cdk:105093118	K9838_1_VHL_26	DGLLVNQTELFVPSLNVDG	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Sus scrofa (pig)	9823	pel-Lindau disease tum	VHL	ssc:100739306	K9823_1_VHL_27	GRRIHSYRGHLWLFRDAGT	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Equus przewalskii (Przewalski's horse)	9798	Hippel-Lindau tumor s	VHL	epz:103561463	K9798_1_VHL_28	PGTGRRIHSYRGHLWLFRD	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Equus caballus (horse)	9796	pel-Lindau disease tum	VHL	ecb:100058139	K9796_1_VHL_29	PGTGRRIHSYRGHLWLFRD	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Equus asinus (ass)	9793	Hippel-Lindau tumor s	VHL	eai:106835626	K9793_1_VHL_30	PGTGRRIHSYRGHLWLFRD	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Physeter catodon (sperm whale)	9755	dau disease tumor su	VHL	pcad:102997043	K9755_1_VHL_31	GRRIHSYRGHLWLFRDAGT	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Orcinus orca (killer whale)	9733	pel-Lindau disease tum	VHL	oor:101272871	K9733_1_VHL_32	YRGHLWLFRDAGTSDGLLV	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Odobenus rosmarus divergens (Pacific walrus)	9708	dau disease tumor su	VHL	oro:101370163	K9708_1_VHL_33	YPTLPPGTGRRIHSYRGHLW	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Panthera pardus (leopard)	9691	dau disease tumor su	VHL	ppad:109254368	K9691_1_VHL_34	PTLPPGTGRRIHSYRGHLW	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Felis catus (domestic cat)	9685	pel-Lindau disease tum	VHL	fca:101092590	K9685_1_VHL_35	PTLPPGTGRRIHSYRGHLW	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Ailuropoda melanoleuca (giant panda)	9646	pel-Lindau disease tum	VHL	aml:100473201	K9646_1_VHL_36	YPTLPPGTGRRIHSYRGHLW	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Vulpes vulpes TameXAggressive cross (red fox)	9627	pel-Lindau disease tum	VHL	vvp:112933697	K9627_1_VHL_37	YPTLPPGTGRRIHSYRGHLW	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25

Organism_name	Organism_id	Protein_name	New_protein_name	Protein_id	New_protein_id	Sequence	Url	Database	Download_time
Pongo abelii (Sumatran orangutan)	9601	pel-Lindau disease tum	VHL	pon:100173372	K9601_1_VHL_38	PQPYPTLPPGTGRRIHSYRG	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Pan troglodytes (chimpanzee)	9598	ndau disease tumor sup	VHL	ptr:749769	K9598_1_VHL_39	PQPYPTLPPGTGRRIHSYRG	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Gorilla gorilla gorilla (western lowland gorilla)	9595	ndau disease tumor sup	VHL	ggo:101149351	K9595_1_VHL_40	EPQPYPTLPPGTGRRIHSYR	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Macaca mulatta (rhesus monkey)	9544	ndau disease tumor sup	VHL	mcc:106996285	K9544_1_VHL_41	PQPYPTLPPGTGRRIHSYRG	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Callithrix jacchus (white-tufted-ear marmoset)	9483	Hippel-Lindau tumor s	VHL	cjc:103788323	K9483_1_VHL_42	GTGRRIHSYRGHLWLFRDAG	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Desmodus rotundus (common vampire bat)	9430	pel-Lindau disease tum	VHL	dro:112301235	K9430_1_VHL_43	GLLVNQTELFVPSLNVDGQF	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Rousettus aegyptiacus (Egyptian rousette)	9407	ndau disease tumor sup	VHL	ray:107498366	K9407_1_VHL_44	YPTLPPGTGRRIHSYRGHLW	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Sarcophilus harrisii (Tasmanian devil)	9305	pel-Lindau disease tum	VHL	shr:111719447	K9305_1_VHL_45	RDANTGDGLLVNQTELFVPS	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Ornithorhynchus anatinus (platypus)	9258	ndau disease tumor sup	VHL	oaa:114806529	K9258_1_VHL_46	WMFRDEETNDRLLVNQKELF	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Pygoscelis adeliae (Adelie penguin)	9238	pel-Lindau disease tum	VHL	padl:103913610	K9238_1_VHL_47	ITLPVFTLKERCLQVVRSLVK	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Parus major (Great Tit)	9157	pel-Lindau disease tum	VHL	pmaj:107210253	K9157_1_VHL_48	LWLFRDAGTNDGLLVNQQEL	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Meleagris gallopavo (turkey)	9103	pel-Lindau disease tum	VHL	mgp:109369884	K9103_1_VHL_49	VVRSLVRPGDYRKLDIVRSLY	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Gallus gallus (chicken)	9031	ndau disease tumor sup	VHL	gga:416117	K9031_1_VHL_50	DFEGRPRYYPVLRPRTGRI	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Numida meleagris (helmeted guineafowl)	8996	pel-Lindau disease tum	VHL	nmel:110405041	K8996_1_VHL_51	MHSYRGHLWLFRDAGTHDG	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Falco peregrinus (peregrine falcon)	8954	Hippel-Lindau tumor s	VHL	fpg:101921920	K8954_1_VHL_52	TNDGLLVNQQELFVAAPNV	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Bubalus bubalis (water buffalo)	89462	pel-Lindau disease tum	VHL	bbub:102411667	K89462_1_VHL_53	HSYRGHLWLFRDAGTSDGLL	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Columba livia (rock pigeon)	8932	pel-Lindau disease tum	VHL	clv:102086490	K8932_1_VHL_54	PNVSTADITLPVFTLKERCLQ	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Alligator mississippiensis (American alligator)	8496	Hippel-Lindau tumor s	VHL	amj:102557707	K8496_1_VHL_55	LWLFRDAGTDDRLLVNKMEL	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Chrysemys picta bellii (western painted turtle)	8478	Hippel-Lindau tumor s	VHL	cpic:101937051	K8478_1_VHL_56	WLFRDAETDDGLLVNQTELF	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Chelonia mydas (green sea turtle)	8469	Hippel-Lindau tumor s	VHL	cmy:102935375	K8469_1_VHL_57	WLFRDAETDDGLLVNQTELF	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Xenopus (Silurana) tropicalis (western clawed frog)	8364	pel-Lindau disease tumo	VHL	xtr:549121	K8364_1_VHL_58	AETDVGLMVNKKEVYVPNP	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Xenopus laevis (African clawed frog)	8355	mor suppressor, E3 ubi	VHL	xla:108714636	K8355_1_VHL_59	REAETDVGLVVNKKEIYVPSF	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Paralichthys olivaceus (Japanese flounder)	8255	pel-Lindau disease tumo	VHL	pov:109628898	K8255_1_VHL_60	TGDPLRVNCKELFLPKPAEG	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Lates calcarifer (barramundi perch)	8187	pel-Lindau disease tumo	VHL	lcf:108881483	K8187_1_VHL_61	DEPLKVNCKELFLPKPAEGG	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Oreochromis niloticus (Nile tilapia)	8128	pel-Lindau disease tumo	VHL	onl:100699102	K8128_1_VHL_62	TDDPLKVNSKELFLPKPERG	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Amphiprion ocellaris (clown anemonefish)	80972	pel-Lindau disease tumo	VHL	aoce:111566652	K80972_1_VHL_63	TDEPLRGNCKELLVPKPEEA	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Oryzias latipes (Japanese medaka)	8090	pel-Lindau disease tumo	VHL	ola:101157853	K8090_1_VHL_64	TDEPLKVNGKEFFLPKPAEN	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Xiphophorus maculatus (southern platyfish)	8083	pel-Lindau disease tumo	VHL	xma:102232438	K8083_1_VHL_65	ASDEPLKVNCKELFVPKPSG	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Salvelinus alpinus (Arctic char)	8036	pel-Lindau disease tumo	VHL	salp:111966645	K8036_1_VHL_66	TDEALKANSKEMFLPKPMEN	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Salmo salar (Atlantic salmon)	8030	Hippel-Lindau tumor su	VHL	sasa:100380306	K8030_1_VHL_67	EALKANSKEMFLPKPMENG	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Esox lucius (northern pike)	8010	dau disease tumor sup	VHL	els:105017201	K8010_1_VHL_68	EPLKANSKDLFLPKPMENGP	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Ictalurus punctatus (channel catfish)	7998	Hippel-Lindau tumor su	VHL	ipu:108271809	K7998_1_VHL_69	DAETDDPMFVNNKEMYLPN	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Oncorhynchus tshawytscha (Chinook salmon)	74940	pel-Lindau disease tumo	VHL	otw:112217436	K74940_1_VHL_70	TDDALKANSKEMFLPKPMEN	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Panthera tigris altaica (Amur tiger)	74533	Hippel-Lindau tumor s	VHL	ptg:102955568	K74533_1_VHL_71	GHLWLFRDAGTYDGLLVNQT	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Drosophila melanogaster (Fruit fly)	7227	VHL_DROME	VHL	Q9V3C1	U7227_1_VHL_72	GERMHVRSQRIFQPIRVRVP	iat=tab&columns=lineag	UniProtKB	9/15/20 17:25
Bos mutus (wild yak)	72004	Hippel-Lindau tumor s	VHL	bom:102274048	K72004_1_VHL_73	ERCLQVVRSLVKPEDYRRL	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Tribolium castaneum (red flour beetle)	7070	pel-Lindau disease tumo	VHL	tca:103313576	K7070_1_VHL_74	QIDKEFIYFPKTTKEYLQNML	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Podarcis muralis (common wall lizard)	64176	pel-Lindau disease tum	VHL	pmua:114591793	K64176_1_VHL_75	AGTDDSLLVNQAELFVATRN	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Caenorhabditis elegans	6239	VHL_CAEEL	VHL	Q19213	U6239_1_VHL_76	FDGCKVLVNEKEVFWPEPA	iat=tab&columns=lineag	UniProtKB	9/15/20 17:25
Nomascus leucogenys (northern white-cheeked gibbon)	61853	ndau disease tumor sup	VHL	nle:100600696	K61853_1_VHL_77	FCNRSPRVVLPVWLNFDGE	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Rhinopithecus roxellana (golden snub-nosed monkey)	61622	pel-Lindau disease tum	VHL	rro:104667143	K61622_1_VHL_78	PQPYPTLPPGTGRRIHSYRG	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Ficedula albicollis (collared flycatcher)	59894	Hippel-Lindau tumor s	VHL	fab:101814622	K59894_1_VHL_79	NDGLLVNQQELFVAAPNVTK	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Taeniopygia guttata (zebra finch)	59729	pel-Lindau disease tum	VHL	tgu:100231028	K59729_1_VHL_80	WLFRDAGTNDGLLVNQQEL	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Austrofundulus limnaeus (annual killifish)	52670	pel-Lindau disease tumo	VHL	alim:106518020	K52670_1_VHL_81	ATDEPLRVNNKELFVPKPTDT	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Geospiza fortis (medium ground-finch)	48883	pel-Lindau disease tum	VHL	gfr:102035931	K48883_1_VHL_82	PVFTLKERCLQVVRSLVSPM	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Monopterus albus (swamp eel)	43700	pel-Lindau disease tumo	VHL	malb:109964425	K43700_1_VHL_83	TDEPLRVNGKELFLPKPEDF	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Camelus ferus (Wild Bactrian camel)	419612	pel-Lindau disease tum	VHL	cfr:116657219	K419612_1_VHL_84	SDGLLVNQTELFVPSLNVDG	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Seriola dumerili (greater amberjack)	41447	pel-Lindau disease tumo	VHL	sdu:111221786	K41447_1_VHL_85	TDEPLKVNSKELFLPKPTEG	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Alligator sinensis (Chinese alligator)	38654	pel-Lindau disease tum	VHL	asn:102378530	K38654_1_VHL_86	RLVLPVWLDFEGRPQSYPVL	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Phascolarctos cinereus (koala)	38626	pel-Lindau disease tum	VHL	pcw:110204865	K38626_1_VHL_87	NTGDALLVNQTELFVPSLNE	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Kryptolebias marmoratus (mangrove rivulus)	37003	pel-Lindau disease tumo	VHL	kmr:108235772	K37003_1_VHL_88	DEPLRGNTKELFVPKPADAR	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Acinonyx jubatus (cheetah)	32536	ndau disease tumor sup	VHL	aju:106972341	K32536_1_VHL_89	YPTLPPGTGRRIHSYRGHLW	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Xiphophorus couchianus (Monterrey platyfish)	32473	pel-Lindau disease tumo	VHL	xco:114135786	K32473_1_VHL_90	ASDEPLKVNCKELFVPKPSG	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Pangasianodon hypophthalmus (striped catfish)	310915	pel-Lindau disease tumo	VHL	phyp:113545808	K310915_1_VHL_91	FRDAETDDPMLVNNKEMYL	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25
Balaenoptera acutorostrata scammoni (minke whale)	310752	Hippel-Lindau tumor s	VHL	bacu:103007825	K310752_1_VHL_92	TGRRIHSYRGHLWLFRDAGT	/rest.kegg.jp/find/genes/	KEGG	9/15/20 17:25

Organism_name	Organism_id	Protein_name	New_protein_name	Protein_id	New_protein_id	Sequence	Url	Database	Download_time
Lonchura striata domestica (Bengalese finch)	299123	pel-Lindau disease tum	VHL	lsr:110477103	K299123_1_VHL_93	LWLFRDAGTNDGLLVNQQE	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Thalassiosira pseudonana CCMP1335	296543	vhl1; prefoldin protein	VHL	HAPSDRAFT_3	K296543_1_VHL_94	GDTVITRYNLADTIYAKAEVD	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Miniopterus natalensis	291302	pel-Lindau disease tum	VHL	mna:107534686	K291302_1_VHL_95	YRGHLWLFRDAGTYDGLLV	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Ursus maritimus (polar bear)	29073	Hippel-Lindau tumor s	VHL	umr:103669642	K29073_1_VHL_96	FVPSLNVDGQPIFANITLPVY	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Cyprinodon variegatus (sheepshead minnow)	28743	pel-Lindau disease tum	VHL	cvg:107094015	K28743_1_VHL_97	LKVNHKELFVPKPSDEDELY	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Anolis carolinensis (green anole)	28377	Hippel-Lindau tumor su	VHL	acs:100560920	K28377_1_VHL_98	TDDSLLVNQAELFVASHNT	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Tupaia chinensis (Chinese tree shrew)	246437	Hippel-Lindau tumor s	VHL	tup:102486323	K246437_1_VHL_99	DGLLVNQTELFVPSLNVDG	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Larimichthys crocea (large yellow croaker)	215358	pel-Lindau disease tum	VHL	lco:104933452	K215358_1_VHL_100	NEPLRANCKEMFLPKPVEG	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Apteryx mantelli mantelli (brown kiwi)	202946	Hippel-Lindau tumor s	VHL	aam:106494625	K202946_1_VHL_101	ADITLPVFTLKERCLQVVRS	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Egretta garzetta (little egret)	188379	pel-Lindau disease tum	VHL	egz:104121781	K188379_1_VHL_102	LFRDAGTNDGLLVNQQELF	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Hipposideros armiger (great roundleaf bat)	186990	Hippel-Lindau tumor s	VHL	hai:109387293	K186990_1_VHL_103	TDNFNQTHKISEGTFADIYR	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Seriola lalandi dorsalis (Yellowtail amberjack)	1841481	pel-Lindau disease tum	VHL	slal:111667871	K1841481_1_VHL_104	TDEPLKVNSKELFLPKPTEG	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Pseudopodoces humilis (Tibetan ground-tit)	181119	Hippel-Lindau tumor s	VHL	phi:102106299	K181119_1_VHL_105	GTNDGLLVNQQELFVAAPNV	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Viola hederacea (Australian violet)	180952	VHL1_VIOHE	VHL	P84522	U180952_1_VHL_106	SCAMISFCFTEVIGCSCKN	at=tab&columns=lineag	UniProtKB	9/15/20 17:25
Empidonax traillii (willow flycatcher)	164674	pel-Lindau disease tum	VHL	etl:114067683	K164674_1_VHL_107	LWLFRDAGTNDGLLVNQQE	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Boleophthalmus pectinirostris (great blue-spotted mudskipper)	150288	pel-Lindau disease tum	VHL	bpec:110168396	K150288_1_VHL_108	DDPLKVNSKELFLPTPATSG	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Gekko japonicus	146911	Hippel-Lindau tumor s	VHL	gja:107122229	K146911_1_VHL_109	SLPPGTGRRMHSYLEHLWL	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Pelodiscus sinensis (Chinese soft-shelled turtle)	13735	pel-Lindau disease tum	VHL	pss:102455661	K13735_1_VHL_110	DGLLVNQTEMFVPTHNING	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Monodelphis domestica (opossum)	13616	Hippel-Lindau tumor s	VHL	mdo:100024305	K13616_1_VHL_111	TGDGLLVNQTELFVPTLNE	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Nipponia nippon (crested ibis)	128390	pel-Lindau disease tum	VHL	nni:104017014	K128390_1_VHL_112	DITLPVFTLKERCLQVVRSL	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Nanorana parkeri	125878	pel-Lindau disease tum	VHL	npr:108786037	K125878_1_VHL_113	ELDTDVPLLVNKKEIYVPSP	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Lipotes vexillifer (Yangtze River dolphin)	118797	Hippel-Lindau tumor s	VHL	lve:103080565	K118797_1_VHL_114	YPTLPPGTGRRIHSYRGDLW	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Ursus arctos horribilis	116960	pel-Lindau disease tum	VHL	uah:113257162	K116960_1_VHL_115	YPTLPPGTGRRIHSYRGHLW	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Myotis brandtii (Brandt's bat)	109478	Hippel-Lindau tumor s	VHL	myb:102255400	K109478_1_VHL_116	RFEKSLTSALLVCSVVICFS	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Hippocampus comes (tiger tail seahorse)	109280	pel-Lindau disease tum	VHL	hcq:109522503	K109280_1_VHL_117	SDDLLRVNGKELFLPALAD	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Maylandia zebra (zebra mbuna)	106582	pel-Lindau disease tum	VHL	mze:101468402	K106582_1_VHL_118	DDPLKVNSKELFLPKPAGG	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Protobothrops mucrosquamatus (Taiwan habu)	103944	pel-Lindau disease tum	VHL	pmur:107298606	K103944_1_VHL_119	RDARSGDSLLVNQTELFVA	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Pogona vitticeps (central bearded dragon)	103695	pel-Lindau disease tum	VHL	pvt:110077876	K103695_1_VHL_120	WLFRDAGTDDSLLVNQAELF	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Nannospalax galili (Upper Galilee mountains blind mole rat)	1026970	Hippel-Lindau tumor s	VHL	ngi:103751433	K1026970_1_VHL_121	WLFRDAGTHDGLLVNQTELF	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Heterocephalus glaber (naked mole rat)	10181	Hippel-Lindau tumor s	VHL	hgl:101697654	K10181_1_VHL_122	DGLLVNQTELFVPSLNVDG	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Rattus norvegicus (Rat)	10116	VHL_RAT	VHL	Q64259	U10116_1_VHL_123	WLFRDAGTHDGLLVNQTEL	at=tab&columns=lineag	UniProtKB	9/15/20 17:25
Mus pahari (shrew mouse)	10093	pel-Lindau disease tum	VHL	mpah:110316548	K10093_1_VHL_124	DGLLVNQTELFVPSLNVDG	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Mus musculus (Mouse)	10090	VHL_MOUSE	VHL	P40338	U10090_1_VHL_125	HLWLFRDAGTHDGLLVNQ	at=tab&columns=lineag	UniProtKB	9/15/20 17:25
Mus caroli (Ryukyu mouse)	10089	pel-Lindau disease tum	VHL	mcal:110296136	K10089_1_VHL_126	HLWLFRDAGTHDGLLVNQ	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Meriones unguiculatus (Mongolian gerbil)	10047	pel-Lindau disease tum	VHL	mun:110548945	K10047_1_VHL_127	WLFRDAGTHDGLLVNQTEL	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25
Cricetulus griseus (Chinese hamster)	10029	dau disease tumor sup	VHL	cge:100751916	K10029_1_VHL_128	DGLLVNQTELFVPSLNVDG	rest.kegg.jp/find/genes	KEGG	9/15/20 17:25

Supplementary_Table_2 566 eukaryotic RefSeq genomes

REFSEQ ASSEMBLY	SPECIES	SUPERKINGDOM	KINGDOM	PHYLUM	CLASS	ORDER	FAMILY	GENUS	ASSEMBLY LEVEL	CLASSIFICATION
GCF_000165445.2	Microcebus murinus	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cheirogaleidae	Microcebus	chromosome	vertebrate
GCF_902635505.1	Sarcophilus harrisii	Eukaryota	Metazoa	Chordata	Mammalia	Dasyuromorphia	Dasyuridae	Sarcophilus	chromosome	vertebrate
GCF_002837175.2	Physeter catodon	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Physeteridae	Physeter	chromosome	vertebrate
GCF_000002295.2	Monodelphis domestica	Eukaryota	Metazoa	Chordata	Mammalia	Didelphimorphia	Didelphidae	Monodelphis	chromosome	vertebrate
GCF_011064425.1	Rattus rattus	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Muridae	Rattus	chromosome	vertebrate
GCF_009834535.1	Camelus ferus	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Camelidae	Camelus	chromosome	vertebrate
GCF_009663435.1	Callithrix jacchus	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cebidae	Callithrix	chromosome	vertebrate
GCF_008632895.1	Mastomys coucha	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Muridae	Mastomys	chromosome	vertebrate
GCF_008692025.1	Phocoena sinus	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Phocoenidae	Phocoena	chromosome	vertebrate
GCF_006542625.1	Nomascus leucogenys	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Hylobatidae	Nomascus	chromosome	vertebrate
GCF_000409795.2	Chlorocebus sabaeus	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cercopithecidae	Chlorocebus	chromosome	vertebrate
GCF_000247795.1	Bos indicus	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Bovidae	Bos	chromosome	vertebrate
GCF_011762505.1	Arvicanthis niloticus	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Muridae	Arvicanthis	chromosome	vertebrate
GCF_002776525.3	Piliocolobus tephrosceles	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cercopithecidae	Piliocolobus	chromosome	vertebrate
GCF_001704415.1	Capra hircus	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Bovidae	Capra	chromosome	vertebrate
GCF_003369695.1	Bos indicus x Bos taurus	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Bovidae	Bos	chromosome	vertebrate
GCF_011762595.1	Tursiops truncatus	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Delphinidae	Tursiops	chromosome	vertebrate
GCF_004126475.1	Phyllostomus discolor	Eukaryota	Metazoa	Chordata	Mammalia	Chiroptera	Phyllostomidae	Phyllostomus	chromosome	vertebrate
GCF_000001635.26	Mus musculus	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Muridae	Mus	chromosome	vertebrate
GCF_008122165.1	Gorilla gorilla gorilla	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Hominidae	Gorilla	chromosome	vertebrate
GCF_003339765.1	Macaca mulatta	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cercopithecidae	Macaca	chromosome	vertebrate
GCF_004115265.1	Rhinolophus ferrumequinum	Eukaryota	Metazoa	Chordata	Mammalia	Chiroptera	Rhinolophidae	Rhinolophus	chromosome	vertebrate
GCF_002863925.1	Equus caballus	Eukaryota	Metazoa	Chordata	Mammalia	Perissodactyla	Equidae	Equus	chromosome	vertebrate
GCF_002263795.1	Bos taurus	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Bovidae	Bos	chromosome	vertebrate
GCF_004115215.1	Ornithorhynchus anatinus	Eukaryota	Metazoa	Chordata	Mammalia	Monotremata	Ornithorhynchidae	Ornithorhynchus	chromosome	vertebrate
GCF_000002285.3	Canis lupus familiaris	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Canidae	Canis	chromosome	vertebrate
GCF_002007445.1	Ailuropoda melanoleuca	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Ursidae	Ailuropoda	chromosome	vertebrate
GCF_002880755.1	Pan troglodytes	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Hominidae	Pan	chromosome	vertebrate
GCF_002742125.1	Ovis aries	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Bovidae	Ovis	chromosome	vertebrate
GCF_007565055.1	Rhinopithecus roxellana	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cercopithecidae	Rhinopithecus	chromosome	vertebrate
GCF_002880775.1	Pongo abelii	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Hominidae	Pongo	chromosome	vertebrate
GCF_000003625.3	Oryctolagus cuniculus	Eukaryota	Metazoa	Chordata	Mammalia	Lagomorpha	Leporidae	Oryctolagus	chromosome	vertebrate
GCF_900094665.1	Mus caroli	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Muridae	Mus	chromosome	vertebrate
GCF_000803125.2	Camelus dromedarius	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Camelidae	Camelus	chromosome	vertebrate
GCF_013052645.1	Pan paniscus	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Hominidae	Pan	chromosome	vertebrate
GCF_007474595.1	Lynx canadensis	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Felidae	Lynx	chromosome	vertebrate
GCF_006229205.1	Suricata suricatta	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Herpestidae	Suricata	chromosome	vertebrate
GCF_000181335.3	Felis catus	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Felidae	Felis	chromosome	vertebrate
GCF_000003025.6	Sus scrofa	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Suidae	Sus	chromosome	vertebrate
GCF_000001405.39	Homo sapiens	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Hominidae	Homo	chromosome	vertebrate
GCF_008728515.1	Papio anubis	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cercopithecidae	Papio	chromosome	vertebrate
GCF_000317375.1	Microtus ochrogaster	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Cricetidae	Microtus	chromosome	vertebrate
GCF_009829155.1	Mustela erminea	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Mustelidae	Mustela	chromosome	vertebrate
GCF_003255815.1	Theropithecus gelada	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cercopithecidae	Theropithecus	chromosome	vertebrate
GCF_000364345.1	Macaca fascicularis	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cercopithecidae	Macaca	chromosome	vertebrate
GCF_000001895.5	Rattus norvegicus	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Muridae	Rattus	chromosome	vertebrate
GCF_900095145.1	Mus pahari	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Muridae	Mus	chromosome	vertebrate
GCF_006547405.1	Globicephala melas	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Delphinidae	Globicephala	scaffold	vertebrate
GCF_000956105.1	Propithecus coquereli	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Indriidae	Propithecus	scaffold	vertebrate
GCF_000164845.3	Vicugna pacos	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Camelidae	Vicugna	scaffold	vertebrate
GCF_000299155.1	Elephantulus edwardii	Eukaryota	Metazoa	Chordata	Mammalia	Macroscelidea	Macroscelididae	Elephantulus	scaffold	vertebrate
GCF_009764315.1	Trachypithecus francoisi	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cercopithecidae	Trachypithecus	scaffold	vertebrate
GCF_000687225.1	Ursus maritimus	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Ursidae	Ursus	scaffold	vertebrate
GCF_000349705.1	Leptonychotes weddellii	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Phocidae	Leptonychotes	scaffold	vertebrate

REFSEQ ASSEMBLY	SPECIES	SUPERKINGDOM	KINGDOM	PHYLUM	CLASS	ORDER	FAMILY	GENUS	ASSEMBLY LEVEL	CLASSIFICATION
GCF_000151885.1	Dipodomys ordii	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Heteromyidae	Dipodomys	scaffold	vertebrate
GCF_000327345.1	Myotis davidii	Eukaryota	Metazoa	Chordata	Mammalia	Chiroptera	Vespertilionidae	Myotis	scaffold	vertebrate
GCF_000412655.1	Myotis brandtii	Eukaryota	Metazoa	Chordata	Mammalia	Chiroptera	Vespertilionidae	Myotis	scaffold	vertebrate
GCF_000952055.2	Aotus nancymaae	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Aotidae	Aotus	scaffold	vertebrate
GCF_000493695.1	aenoptera acutorostrata scamn	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Balaenopteridae	Balaenoptera	scaffold	vertebrate
GCF_004348235.1	Phoca vitulina	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Phocidae	Phoca	scaffold	vertebrate
GCF_010015895.1	Lontra canadensis	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Mustelidae	Lontra	scaffold	vertebrate
GCF_001685135.1	Manis javanica	Eukaryota	Metazoa	Chordata	Mammalia	Pholidota	Manidae	Manis	scaffold	vertebrate
GCF_002201575.1	Neomonachus schauinslandi	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Phocidae	Neomonachus	scaffold	vertebrate
GCF_000298275.1	Orycteropus afer afer	Eukaryota	Metazoa	Chordata	Mammalia	Tubulidentata	Orycteropodidae	Orycteropus	scaffold	vertebrate
GCF_003265705.1	Callorhinus ursinus	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Otariidae	Callorhinus	scaffold	vertebrate
GCF_000331955.2	Orcinus orca	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Delphinidae	Orcinus	scaffold	vertebrate
GCF_001466805.2	Rousettus aegyptiacus	Eukaryota	Metazoa	Chordata	Mammalia	Chiroptera	Pteropodidae	Rousettus	scaffold	vertebrate
GCF_000280705.1	Jaculus jaculus	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Dipodidae	Jaculus	scaffold	vertebrate
GCF_011800145.1	Mirounga leonina	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Phocidae	Mirounga	scaffold	vertebrate
GCF_000349665.1	Mesocricetus auratus	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Cricetidae	Mesocricetus	scaffold	vertebrate
GCF_000208655.1	Dasypus novemcinctus	Eukaryota	Metazoa	Chordata	Mammalia	Cingulata	Dasypodidae	Dasypus	scaffold	vertebrate
GCF_009828535.2	Hylobates moloch	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Hylobatidae	Hylobates	scaffold	vertebrate
GCF_003327715.1	Puma concolor	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Felidae	Puma	scaffold	vertebrate
GCF_000151735.1	Cavia porcellus	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Caviidae	Cavia	scaffold	vertebrate
GCF_000215625.1	Mustela putorius furo	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Mustelidae	Mustela	scaffold	vertebrate
GCF_001595765.1	Miniopterus natalensis	Eukaryota	Metazoa	Chordata	Mammalia	Chiroptera	Vespertilionidae	Miniopterus	scaffold	vertebrate
GCF_003584765.1	Ursus arctos horribilis	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Ursidae	Ursus	scaffold	vertebrate
GCF_004664715.1	Peromyscus leucopus	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Cricetidae	Peromyscus	scaffold	vertebrate
GCF_000321225.1	Odobenus rosmarus divergens	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Odobenidae	Odobenus	scaffold	vertebrate
GCF_000298355.1	Bos mutus	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Bovidae	Bos	scaffold	vertebrate
GCF_000296755.1	Erinaceus europaeus	Eukaryota	Metazoa	Chordata	Mammalia	Eulipotyphla	Erinaceidae	Erinaceus	scaffold	vertebrate
GCF_004028035.1	Eumetopias jubatus	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Otariidae	Eumetopias	scaffold	vertebrate
GCF_000325575.1	Pteropus alecto	Eukaryota	Metazoa	Chordata	Mammalia	Chiroptera	Pteropodidae	Pteropus	scaffold	vertebrate
GCF_002204375.1	Meriones unguiculatus	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Muridae	Meriones	scaffold	vertebrate
GCF_002940915.1	Desmodus rotundus	Eukaryota	Metazoa	Chordata	Mammalia	Chiroptera	Phyllostomidae	Desmodus	scaffold	vertebrate
GCF_000260255.1	Octodon degus	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Octodontidae	Octodon	scaffold	vertebrate
GCF_000296735.1	Chrysochloris asiatica	Eukaryota	Metazoa	Chordata	Mammalia	frotheria (SUPERORDE	Chrysochloridae	Chrysocloris	scaffold	vertebrate
GCF_001857705.1	Panthera pardus	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Felidae	Panthera	scaffold	vertebrate
GCF_000276665.1	Chinchilla lanigera	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Chinchillidae	Chinchilla	scaffold	vertebrate
GCF_002288905.1	Enhydra lutris kenyoni	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Mustelidae	Enhydra	scaffold	vertebrate
GCF_012274545.1	Fukomys damarensis	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Bathyergidae	Fukomys	scaffold	vertebrate
GCF_000696425.1	Galeopterus variegatus	Eukaryota	Metazoa	Chordata	Mammalia	Dermoptera	Cynocephalidae	Galeopterus	scaffold	vertebrate
GCF_003709585.1	Acinonyx jubatus	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Felidae	Acinonyx	scaffold	vertebrate
GCF_000164805.1	Carlito syrichta	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Tarsiidae	Carlito	scaffold	vertebrate
GCF_000247695.1	Heterocephalus glaber	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Bathyergidae	Heterocephalus	scaffold	vertebrate
GCF_000308155.1	Eptesicus fuscus	Eukaryota	Metazoa	Chordata	Mammalia	Chiroptera	Vespertilionidae	Eptesicus	scaffold	vertebrate
GCF_000334495.1	Tupaia chinensis	Eukaryota	Metazoa	Chordata	Mammalia	Scandentia	Tupaiidae	Tupaia	scaffold	vertebrate
GCF_000181275.1	Sorex araneus	Eukaryota	Metazoa	Chordata	Mammalia	Eulipotyphla	Soricidae	Sorex	scaffold	vertebrate
GCF_000283155.1	Ceratotherium simum simum	Eukaryota	Metazoa	Chordata	Mammalia	Perissodactyla	Rhinocerotidae	Ceratotherium	scaffold	vertebrate
GCF_000147115.1	Myotis lucifugus	Eukaryota	Metazoa	Chordata	Mammalia	Chiroptera	Vespertilionidae	Myotis	scaffold	vertebrate
GCF_000001905.1	Loxodonta africana	Eukaryota	Metazoa	Chordata	Mammalia	Proboscidea	Elephantidae	Loxodonta	scaffold	vertebrate
GCF_000464555.1	Panthera tigris altaica	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Felidae	Panthera	scaffold	vertebrate
GCF_009761245.1	Sapajus apella	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cebidae	Sapajus	scaffold	vertebrate
GCF_001890085.1	Hipposideros armiger	Eukaryota	Metazoa	Chordata	Mammalia	Chiroptera	Hipposideridae	Hipposideros	scaffold	vertebrate
GCF_001698545.1	Rhinopithecus bieti	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cercopithecidae	Rhinopithecus	scaffold	vertebrate
GCF_000243295.1	Trichechus manatus latirostris	Eukaryota	Metazoa	Chordata	Mammalia	Sirenia	Trichechidae	Trichechus	scaffold	vertebrate
GCF_001305755.1	Equus asinus	Eukaryota	Metazoa	Chordata	Mammalia	Perissodactyla	Equidae	Equus	scaffold	vertebrate
GCF_003160815.1	Vulpes vulpes	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Canidae	Vulpes	scaffold	vertebrate

REFSEQ ASSEMBLY	SPECIES	SUPERKINGDOM	KINGDOM	PHYLUM	CLASS	ORDER	FAMILY	GENUS	ASSEMBLY LEVEL	CLASSIFICATION
GCF_000292845.1	Ochotona princeps	Eukaryota	Metazoa	Chordata	Mammalia	Lagomorpha	Ochotonidae	Ochotona	scaffold	vertebrate
GCF_003031525.1	ocaena asiaeorientalis asiaeor	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Phocoenidae	Neophocaena	scaffold	vertebrate
GCF_001604975.1	Cebus capucinus imitator	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cebidae	Cebus	scaffold	vertebrate
GCF_000696695.1	Equus przewalskii	Eukaryota	Metazoa	Chordata	Mammalia	Perissodactyla	Equidae	Equus	scaffold	vertebrate
GCF_000235385.1	Saimiri boliviensis boliviensis	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cebidae	Saimiri	scaffold	vertebrate
GCF_001984765.1	Castor canadensis	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Castoridae	Castor	scaffold	vertebrate
GCF_000442215.1	Lipotes vexillifer	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Lipotidae	Lipotes	scaffold	vertebrate
GCF_000151845.1	Pteropus vampyrus	Eukaryota	Metazoa	Chordata	Mammalia	Chiroptera	Pteropodidae	Pteropus	scaffold	vertebrate
GCF_000767855.1	Camelus bactrianus	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Camelidae	Camelus	scaffold	vertebrate
GCF_000236235.1	Ictidomys tridecemlineatus	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Sciuridae	Ictidomys	scaffold	vertebrate
GCF_003426925.1	Urocitellus parryii	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Sciuridae	Urocitellus	scaffold	vertebrate
GCF_000955945.1	Cercocebus atys	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cercopithecidae	Cercocebus	scaffold	vertebrate
GCF_900631625.1	Zalophus californianus	Eukaryota	Metazoa	Chordata	Mammalia	Carnivora	Otariidae	Zalophus	scaffold	vertebrate
GCF_000181295.1	Otolemur garnettii	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Galagidae	Otolemur	scaffold	vertebrate
GCF_900497805.2	Vombatus ursinus	Eukaryota	Metazoa	Chordata	Mammalia	Diprotodontia	Vombatidae	Vombatus	scaffold	vertebrate
GCF_003676395.1	Lagenorhynchus obliquidens	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Delphinidae	Lagenorhynchus	scaffold	vertebrate
GCF_000622305.1	Nannospalax galili	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Spalacidae	Nannospalax	scaffold	vertebrate
GCF_000951045.1	Mandrillus leucophaeus	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cercopithecidae	Mandrillus	scaffold	vertebrate
GCF_000500345.1	Peromyscus maniculatus bairdi	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Cricetidae	Peromyscus	scaffold	vertebrate
GCF_000313985.2	Echinops telfairi	Eukaryota	Metazoa	Chordata	Mammalia	frotheria (SUPERORDER	Tenrecidae	Echinops	scaffold	vertebrate
GCF_000956065.1	Macaca nemestrina	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cercopithecidae	Macaca	scaffold	vertebrate
GCF_004785775.1	Grammomys surdaster	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Muridae	Grammomys	scaffold	vertebrate
GCF_000223135.1	Cricetulus griseus	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Cricetidae	Cricetulus	scaffold	vertebrate
GCF_001458135.1	Marmota marmota marmota	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Sciuridae	Marmota	scaffold	vertebrate
GCF_003676075.2	Marmota flaviventris	Eukaryota	Metazoa	Chordata	Mammalia	Rodentia	Sciuridae	Marmota	scaffold	vertebrate
GCF_002288925.2	Delphinapterus leucas	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Monodontidae	Delphinapterus	scaffold	vertebrate
GCF_002102435.1	Odocoileus virginianus texanus	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Cervidae	Odocoileus	scaffold	vertebrate
GCF_000951035.1	Colobus angolensis palliatus	Eukaryota	Metazoa	Chordata	Mammalia	Primates	Cercopithecidae	Colobus	scaffold	vertebrate
GCF_005190385.1	Monodon monoceros	Eukaryota	Metazoa	Chordata	Mammalia	Artiodactyla	Monodontidae	Monodon	scaffold	vertebrate
GCF_000260355.1	Condylura cristata	Eukaryota	Metazoa	Chordata	Mammalia	Eulipotyphla	Talpidae	Condylura	scaffold	vertebrate
GCF_002099425.1	Phascolarctos cinereus	Eukaryota	Metazoa	Chordata	Mammalia	Diprotodontia	Phascolarctidae	Phascolarctos	contig	vertebrate
GCF_013103735.1	Etheostoma cragini	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Percidae	Etheostoma	chromosome	vertebrate
GCF_001858045.2	Oreochromis niloticus	Eukaryota	Metazoa	Chordata	Actinopteri	Cichliformes	Cichlidae	Oreochromis	chromosome	vertebrate
GCF_002234675.1	Oryzias latipes	Eukaryota	Metazoa	Chordata	Actinopteri	Beloniformes	Adrianichthyidae	Oryzias	chromosome	vertebrate
GCF_013347855.1	Anguilla anguilla	Eukaryota	Metazoa	Chordata	Actinopteri	Anguilliformes	Anguillidae	Anguilla	chromosome	vertebrate
GCF_009819885.1	Catharus ustulatus	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Turdidae	Catharus	chromosome	vertebrate
GCF_000238955.4	Maylandia zebra	Eukaryota	Metazoa	Chordata	Actinopteri	Cichliformes	Cichlidae	Maylandia	chromosome	vertebrate
GCF_900246225.1	Astatotilapia calliptera	Eukaryota	Metazoa	Chordata	Actinopteri	Cichliformes	Cichlidae	Astatotilapia	chromosome	vertebrate
GCF_000372685.2	Astyanax mexicanus	Eukaryota	Metazoa	Chordata	Actinopteri	Characiformes	Characidae	Astyanax	chromosome	vertebrate
GCF_009769535.1	Thamnophis elegans	Eukaryota	Metazoa	Chordata	Lepidosauria	Squamata	Colubridae	Thamnophis	chromosome	vertebrate
GCF_900880675.1	Sparus aurata	Eukaryota	Metazoa	Chordata	Actinopteri	Spariformes	Sparidae	Sparus	chromosome	vertebrate
GCF_009829145.1	Chiroxiphia lanceolata	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Pipridae	Chiroxiphia	chromosome	vertebrate
GCF_000242695.1	Lepisosteus oculatus	Eukaryota	Metazoa	Chordata	Actinopteri	Semionotiformes	Lepisosteidae	Lepisosteus	chromosome	vertebrate
GCF_008692095.1	Etheostoma spectabile	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Percidae	Etheostoma	chromosome	vertebrate
GCF_004354835.1	Perca flavescens	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Percidae	Perca	chromosome	vertebrate
GCF_000002035.6	Danio rerio	Eukaryota	Metazoa	Chordata	Actinopteri	Cypriniformes	Danionidae	Danio	chromosome	vertebrate
GCF_002910315.2	Salvelinus alpinus	Eukaryota	Metazoa	Chordata	Actinopteri	Salmoniformes	Salmonidae	Salvelinus	chromosome	vertebrate
GCF_001465895.1	Nothobranchius furzeri	Eukaryota	Metazoa	Chordata	Actinopteri	Cyprinodontiformes	Nothobranchiidae	Nothobranchius	chromosome	vertebrate
GCF_000241765.3	Chrysemys picta bellii	Eukaryota	Metazoa	Chordata	arcopterygii (SUPERCLAS	Testudines	Emydidae	Chrysemys	chromosome	vertebrate
GCF_000523025.1	Cynoglossus semilaevis	Eukaryota	Metazoa	Chordata	Actinopteri	Pleuronectiformes	Cynoglossidae	Cynoglossus	chromosome	vertebrate
GCF_007364275.1	Archocentrus centrarchus	Eukaryota	Metazoa	Chordata	Actinopteri	Cichliformes	Cichlidae	Archocentrus	chromosome	vertebrate
GCF_900964775.1	Scleropages formosus	Eukaryota	Metazoa	Chordata	Actinopteri	Osteoglossiformes	Osteoglossidae	Scleropages	chromosome	vertebrate
GCF_009819535.1	Lacerta agilis	Eukaryota	Metazoa	Chordata	Lepidosauria	Squamata	Lacertidae	Lacerta	chromosome	vertebrate
GCF_003850225.1	Anas platyrhynchos	Eukaryota	Metazoa	Chordata	Aves	Anseriformes	Anatidae	Anas	chromosome	vertebrate

REFSEQ ASSEMBLY	SPECIES	SUPERKINGDOM	KINGDOM	PHYLUM	CLASS	ORDER	FAMILY	GENUS	ASSEMBLY LEVEL	CLASSIFICATION
GCF_902150065.1	Myripristis murdjan	Eukaryota	Metazoa	Chordata	Actinopteri	Holocentriformes	Holocentridae	Myripristis	chromosome	vertebrate
GCF_002775205.1	Xiphophorus maculatus	Eukaryota	Metazoa	Chordata	Actinopteri	Cyprinodontiformes	Poeciliidae	Xiphophorus	chromosome	vertebrate
GCF_902167405.1	Gadus morhua	Eukaryota	Metazoa	Chordata	Actinopteri	Gadiformes	Gadidae	Gadus	chromosome	vertebrate
GCF_000951615.1	Cyprinus carpio	Eukaryota	Metazoa	Chordata	Actinopteri	Cypriniformes	Cyprinidae	Cyprinus	chromosome	vertebrate
GCF_007399415.2	Gopherus evgoodei	Eukaryota	Metazoa	Chordata	rcopterygii (SUPERCLAS	Testudines	Testudinidae	Gopherus	chromosome	vertebrate
GCF_900324465.2	Anabas testudineus	Eukaryota	Metazoa	Chordata	Actinopteri	Anabantiformes	Anabantidae	Anabas	chromosome	vertebrate
GCF_902362185.1	Chanos chanos	Eukaryota	Metazoa	Chordata	Actinopteri	Gonorynchiformes	Chanidae	Chanos	chromosome	vertebrate
GCF_011077185.1	Oxyura jamaicensis	Eukaryota	Metazoa	Chordata	Aves	Anseriformes	Anatidae	Oxyura	chromosome	vertebrate
GCF_011800845.1	Zootoca vivipara	Eukaryota	Metazoa	Chordata	Lepidosauria	Squamata	Lacertidae	Zootoca	chromosome	vertebrate
GCF_001660625.1	Ictalurus punctatus	Eukaryota	Metazoa	Chordata	Actinopteri	Siluriformes	Ictaluridae	Ictalurus	chromosome	vertebrate
GCF_005281545.1	Epinephelus lanceolatus	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Serranidae	Epinephelus	chromosome	vertebrate
GCF_002163495.1	Oncorhynchus mykiss	Eukaryota	Metazoa	Chordata	Actinopteri	Salmoniformes	Salmonidae	Oncorhynchus	chromosome	vertebrate
GCF_901001165.1	Salmo trutta	Eukaryota	Metazoa	Chordata	Actinopteri	Salmoniformes	Salmonidae	Salmo	chromosome	vertebrate
GCF_900963305.1	Echeneis naucrates	Eukaryota	Metazoa	Chordata	Actinopteri	Carangiformes	Echeneidae	Echeneis	chromosome	vertebrate
GCF_004027225.2	Strigops habroptila	Eukaryota	Metazoa	Chordata	Aves	Psittaciformes	Psittacidae	Strigops	chromosome	vertebrate
GCF_902148855.1	Sphaeramia orbicularis	Eukaryota	Metazoa	Chordata	Actinopteri	Kurtiformes	Apogonidae	Sphaeramia	chromosome	vertebrate
GCF_006149115.1	Oncorhynchus nerka	Eukaryota	Metazoa	Chordata	Actinopteri	Salmoniformes	Salmonidae	Oncorhynchus	chromosome	vertebrate
GCF_002872995.1	Oncorhynchus tshawytscha	Eukaryota	Metazoa	Chordata	Actinopteri	Salmoniformes	Salmonidae	Oncorhynchus	chromosome	vertebrate
GCF_012275295.1	Melopsittacus undulatus	Eukaryota	Metazoa	Chordata	Aves	Psittaciformes	Psittaculidae	Melopsittacus	chromosome	vertebrate
GCF_900496995.1	Aquila chrysaetos chrysaetos	Eukaryota	Metazoa	Chordata	Aves	Accipitriformes	Accipitridae	Aquila	chromosome	vertebrate
GCF_010993605.1	Petromyzon marinus	Eukaryota	Metazoa	Chordata	Hyperoartia	Petromyzontiformes	Petromyzontidae	Petromyzon	chromosome	vertebrate
GCF_902148845.1	Salarias fasciatus	Eukaryota	Metazoa	Chordata	Actinopteri	Blenniiformes	Blenniidae	Salarias	chromosome	vertebrate
GCF_900634415.1	Cottoperca gobio	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Bovichtidae	Cottoperca	chromosome	vertebrate
GCF_003368295.1	Carassius auratus	Eukaryota	Metazoa	Chordata	Actinopteri	Cypriniformes	Cyprinidae	Carassius	chromosome	vertebrate
GCF_901000725.2	Takifugu rubripes	Eukaryota	Metazoa	Chordata	Actinopteri	Tetraodontiformes	Tetraodontidae	Takifugu	chromosome	vertebrate
GCF_001663975.1	Xenopus laevis	Eukaryota	Metazoa	Chordata	Amphibia	Anura	Pipidae	Xenopus	chromosome	vertebrate
GCF_010909765.1	Amblyraja radiata	Eukaryota	Metazoa	Chordata	Chondrichthyes	Rajiformes	Rajidae	Amblyraja	chromosome	vertebrate
GCF_000247815.1	Ficedula albicollis	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Muscicapidae	Ficedula	chromosome	vertebrate
GCF_900700375.1	Denticeps clupeoides	Eukaryota	Metazoa	Chordata	Actinopteri	Clupeiformes	Denticipitidae	Denticeps	chromosome	vertebrate
GCF_009819705.1	Hippoglossus hippoglossus	Eukaryota	Metazoa	Chordata	Actinopteri	Pleuronectiformes	Pleuronectidae	Hippoglossus	chromosome	vertebrate
GCF_901001135.1	Rhinatrema bivittatum	Eukaryota	Metazoa	Chordata	Amphibia	Gymnophiona	Rhinatrematidae	Rhinatrema	chromosome	vertebrate
GCF_902500255.1	Thalassophryne amazonica	Eukaryota	Metazoa	Chordata	Actinopteri	Batrachoidiformes	Batrachoididae	Thalassophryne	chromosome	vertebrate
GCF_901933205.1	Camarhynchus parvulus	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Thraupidae	Camarhynchus	chromosome	vertebrate
GCF_009762535.1	Notolabrus celidotus	Eukaryota	Metazoa	Chordata	Actinopteri	Labriformes	Labridae	Notolabrus	chromosome	vertebrate
GCF_001577835.2	Coturnix japonica	Eukaryota	Metazoa	Chordata	Aves	Galliformes	Phasianidae	Coturnix	chromosome	vertebrate
GCF_009829125.1	eriophthalmus magnuspinnatu	Eukaryota	Metazoa	Chordata	Actinopteri	Gobiiformes	Gobiidae	Periophthalmus	chromosome	vertebrate
GCF_011004845.1	Esox lucius	Eukaryota	Metazoa	Chordata	Actinopteri	Esociformes	Esocidae	Esox	chromosome	vertebrate
GCF_001522545.3	Parus major	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Paridae	Parus	chromosome	vertebrate
GCF_013100865.1	Trachemys scripta elegans	Eukaryota	Metazoa	Chordata	rcopterygii (SUPERCLAS	Testudines	Emydidae	Trachemys	chromosome	vertebrate
GCF_008822105.2	Taeniopygia guttata	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Estrildidae	Taeniopygia	chromosome	vertebrate
GCF_013339905.1	Hippoglossus stenolepis	Eukaryota	Metazoa	Chordata	Actinopteri	Pleuronectiformes	Pleuronectidae	Hippoglossus	chromosome	vertebrate
GCF_000002315.6	Gallus gallus	Eukaryota	Metazoa	Chordata	Aves	Galliformes	Phasianidae	Gallus	chromosome	vertebrate
GCF_000146605.3	Meleagris gallopavo	Eukaryota	Metazoa	Chordata	Aves	Galliformes	Phasianidae	Meleagris	chromosome	vertebrate
GCF_009078355.1	angasianodon hypophthalmu	Eukaryota	Metazoa	Chordata	Actinopteri	Siluriformes	Pangasiidae	Pangasianodon	chromosome	vertebrate
GCF_000233375.1	Salmo salar	Eukaryota	Metazoa	Chordata	Actinopteri	Salmoniformes	Salmonidae	Salmo	chromosome	vertebrate
GCF_000090745.1	Anolis carolinensis	Eukaryota	Metazoa	Chordata	Lepidosauria	Squamata	Dactyloidae	Anolis	chromosome	vertebrate
GCF_001444195.1	Xiphophorus couchianus	Eukaryota	Metazoa	Chordata	Actinopteri	Cyprinodontiformes	Poeciliidae	Xiphophorus	chromosome	vertebrate
GCF_002078875.1	Numida meleagris	Eukaryota	Metazoa	Chordata	Aves	Galliformes	Numididae	Numida	chromosome	vertebrate
GCF_900634795.2	Betta splendens	Eukaryota	Metazoa	Chordata	Actinopteri	Anabantiformes	Osphronemidae	Betta	chromosome	vertebrate
GCF_004329235.1	Podarcis muralis	Eukaryota	Metazoa	Chordata	Lepidosauria	Squamata	Lacertidae	Podarcis	chromosome	vertebrate
GCF_009650955.1	Corvus moneduloides	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Corvidae	Corvus	chromosome	vertebrate
GCF_000972845.2	Larimichthys crocea	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Sciaenidae	Larimichthys	chromosome	vertebrate
GCF_009819795.1	Aythya fuligula	Eukaryota	Metazoa	Chordata	Aves	Anseriformes	Anatidae	Aythya	chromosome	vertebrate
GCF_010645085.1	Acipenser ruthenus	Eukaryota	Metazoa	Chordata	Actinopteri	Acipenseriformes	Acipenseridae	Acipenser	chromosome	vertebrate

4

REFSEQ ASSEMBLY	SPECIES	SUPERKINGDOM	KINGDOM	PHYLUM	CLASS	ORDER	FAMILY	GENUS	ASSEMBLY LEVEL	CLASSIFICATION
GCF_902827115.1	'seudochaenichthys georgianu	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Channichthyidae	Pseudochaenichthys	chromosome	vertebrate
GCF_900634775.1	Gouania willdenowi	Eukaryota	Metazoa	Chordata	Actinopteri	Blenniiformes	Gobiesocidae	Gouania	chromosome	vertebrate
GCF_003957555.1	Calypte anna	Eukaryota	Metazoa	Chordata	Aves	Apodiformes	Trochilidae	Calypte	chromosome	vertebrate
GCF_009769545.1	Cyclopterus lumpus	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Cyclopteridae	Cyclopterus	chromosome	vertebrate
GCF_900747795.1	Erpetoichthys calabaricus	Eukaryota	Metazoa	Chordata	Cladistia	Polypteriformes	Polypteridae	Erpetoichthys	chromosome	vertebrate
GCF_003331165.1	Xiphophorus hellerii	Eukaryota	Metazoa	Chordata	Actinopteri	Cyprinodontiformes	Poeciliidae	Xiphophorus	chromosome	vertebrate
GCF_901765095.1	Microcaecilia unicolor	Eukaryota	Metazoa	Chordata	Amphibia	Gymnophiona	Siphonopidae	Microcaecilia	chromosome	vertebrate
GCF_000004195.4	Xenopus tropicalis	Eukaryota	Metazoa	Chordata	Amphibia	Anura	Pipidae	Xenopus	chromosome	vertebrate
GCF_900634625.1	Parambassis ranga	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Ambassidae	Parambassis	chromosome	vertebrate
GCF_005870125.1	Lonchura striata domestica	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Estrildidae	Lonchura	chromosome	vertebrate
GCF_002021735.2	Oncorhynchus kisutch	Eukaryota	Metazoa	Chordata	Actinopteri	Salmoniformes	Salmonidae	Oncorhynchus	chromosome	vertebrate
GCF_900324485.2	Mastacembelus armatus	Eukaryota	Metazoa	Chordata	Actinopteri	Synbranchiformes	Mastacembelidae	Mastacembelus	chromosome	vertebrate
GCF_902459505.1	Geotrypetes seraphini	Eukaryota	Metazoa	Chordata	Amphibia	Gymnophiona	Dermophiidae	Geotrypetes	chromosome	vertebrate
GCF_900700415.1	Clupea harengus	Eukaryota	Metazoa	Chordata	Actinopteri	Clupeiformes	Clupeidae	Clupea	chromosome	vertebrate
GCF_000633615.1	Poecilia reticulata	Eukaryota	Metazoa	Chordata	Actinopteri	Cyprinodontiformes	Poeciliidae	Poecilia	chromosome	vertebrate
GCF_000230535.1	Pelodiscus sinensis	Eukaryota	Metazoa	Chordata	rcopterygii (SUPERCLAS	Testudines	Trionychidae	Pelodiscus	scaffold	vertebrate
GCF_001891065.1	Hippocampus comes	Eukaryota	Metazoa	Chordata	Actinopteri	Syngnathiformes	Syngnathidae	Hippocampus	scaffold	vertebrate
GCF_000687285.1	Phaethon lepturus	Eukaryota	Metazoa	Chordata	Aves	Pelecaniformes	Phaethontidae	Phaethon	scaffold	vertebrate
GCF_003945725.1	Corapipo altera	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Pipridae	Corapipo	scaffold	vertebrate
GCF_002901205.1	Cyanistes caeruleus	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Paridae	Cyanistes	scaffold	vertebrate
GCF_001515625.1	Sinocyclocheilus rhinocerous	Eukaryota	Metazoa	Chordata	Actinopteri	Cypriniformes	Cyprinidae	Sinocyclocheilus	scaffold	vertebrate
GCF_000687185.1	Egretta garzetta	Eukaryota	Metazoa	Chordata	Aves	Pelecaniformes	Ardeidae	Egretta	scaffold	vertebrate
GCF_000331425.1	Pseudopodoces humilis	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Paridae	Pseudopodoces	scaffold	vertebrate
GCF_000277835.1	Geospiza fortis	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Thraupidae	Geospiza	scaffold	vertebrate
GCF_001715985.3	Manacus vitellinus	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Pipridae	Manacus	scaffold	vertebrate
GCF_000690725.1	Stegastes partitus	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Pomacentridae	Stegastes	scaffold	vertebrate
GCF_003342845.1	Nothoprocta perdicaria	Eukaryota	Metazoa	Chordata	Aves	Tinamiformes	Tinamidae	Nothoprocta	scaffold	vertebrate
GCF_001640805.1	Lates calcarifer	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Centropomidae	Lates	scaffold	vertebrate
GCF_001443285.1	Poecilia latipinna	Eukaryota	Metazoa	Chordata	Actinopteri	Cyprinodontiformes	Poeciliidae	Poecilia	scaffold	vertebrate
GCF_000690775.1	Eurypyga helias	Eukaryota	Metazoa	Chordata	Aves	Gruiformes	Eurypygidae	Eurypyga	scaffold	vertebrate
GCF_000700745.1	Antrostomus carolinensis	Eukaryota	Metazoa	Chordata	Aves	Caprimulgiformes	Caprimulgidae	Antrostomus	scaffold	vertebrate
GCF_001443325.1	Poecilia mexicana	Eukaryota	Metazoa	Chordata	Actinopteri	Cyprinodontiformes	Poeciliidae	Poecilia	scaffold	vertebrate
GCF_003259725.1	Athene cunicularia	Eukaryota	Metazoa	Chordata	Aves	Strigiformes	Strigidae	Athene	scaffold	vertebrate
GCF_000747805.1	Chaetura pelagica	Eukaryota	Metazoa	Chordata	Aves	Apodiformes	Apodidae	Chaetura	scaffold	vertebrate
GCF_002776465.1	Amphiprion ocellaris	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Pomacentridae	Amphiprion	scaffold	vertebrate
GCF_000225785.1	Latimeria chalumnae	Eukaryota	Metazoa	Chordata	Coelacanthimorpha	Coelacanthiformes	Coelacanthidae	Latimeria	scaffold	vertebrate
GCF_000690835.1	Fulmarus glacialis	Eukaryota	Metazoa	Chordata	Aves	Procellariiformes	Procellariidae	Fulmarus	scaffold	vertebrate
GCF_000709895.1	alearica regulorum gibbericep	Eukaryota	Metazoa	Chordata	Aves	Gruiformes	Gruidae	Balearica	scaffold	vertebrate
GCF_000709325.1	Cuculus canorus	Eukaryota	Metazoa	Chordata	Aves	Cuculiformes	Cuculidae	Cuculus	scaffold	vertebrate
GCF_000699005.1	Picoides pubescens	Eukaryota	Metazoa	Chordata	Aves	Piciformes	Picidae	Picoides	scaffold	vertebrate
GCF_000691975.1	Corvus brachyrhynchos	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Corvidae	Corvus	scaffold	vertebrate
GCF_000708225.1	Nipponia nippon	Eukaryota	Metazoa	Chordata	Aves	Pelecaniformes	Threskiornithidae	Nipponia	scaffold	vertebrate
GCF_001604755.1	Lepidothrix coronata	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Pipridae	Lepidothrix	scaffold	vertebrate
GCF_902827165.1	Trematomus bernacchii	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Nototheniidae	Trematomus	scaffold	vertebrate
GCF_003724035.1	Tachysurus fulvidraco	Eukaryota	Metazoa	Chordata	Actinopteri	Siluriformes	Bagridae	Tachysurus	scaffold	vertebrate
GCF_000690535.1	Cariama cristata	Eukaryota	Metazoa	Chordata	Aves	Cariamiformes	Cariamidae	Cariama	scaffold	vertebrate
GCF_000692075.1	Opisthocomus hoazin	Eukaryota	Metazoa	Chordata	Aves	Opisthocomiformes	Opisthocomidae	Opisthocomus	scaffold	vertebrate
GCF_000165045.1	Callorhinchus milii	Eukaryota	Metazoa	Chordata	Chondrichthyes	Chimaeriformes	Callorhinchidae	Callorhinchus	scaffold	vertebrate
GCF_902150015.1	Tyto alba alba	Eukaryota	Metazoa	Chordata	Aves	Strigiformes	Tytonidae	Tyto	scaffold	vertebrate
GCF_000691785.1	Leptosomus discolor	Eukaryota	Metazoa	Chordata	Aves	Coraciiformes	Leptosomidae	Leptosomus	scaffold	vertebrate
GCF_002872115.1	Paramormyrops kingsleyae	Eukaryota	Metazoa	Chordata	Actinopteri	Osteoglossiformes	Mormyridae	Paramormyrops	scaffold	vertebrate
GCF_000691405.1	Haliaeetus albicilla	Eukaryota	Metazoa	Chordata	Aves	Accipitriformes	Accipitridae	Haliaeetus	scaffold	vertebrate
GCF_000691845.1	Merops nubicus	Eukaryota	Metazoa	Chordata	Aves	Coraciiformes	Meropidae	Merops	scaffold	vertebrate
GCF_001515645.1	Sinocyclocheilus grahami	Eukaryota	Metazoa	Chordata	Actinopteri	Cypriniformes	Cyprinidae	Sinocyclocheilus	scaffold	vertebrate

REFSEQ ASSEMBLY	SPECIES	SUPERKINGDOM	KINGDOM	PHYLUM	CLASS	ORDER	FAMILY	GENUS	ASSEMBLY LEVEL	CLASSIFICATION
GCF_000695815.1	Acanthisitta chloris	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Acanthisittidae	Acanthisitta	scaffold	vertebrate
GCF_001077635.1	Thamnophis sirtalis	Eukaryota	Metazoa	Chordata	Lepidosauria	Squamata	Colubridae	Thamnophis	scaffold	vertebrate
GCF_001266775.1	Austrofundulus limnaeus	Eukaryota	Metazoa	Chordata	Actinopteri	Cyprinodontiformes	Rivulidae	Austrofundulus	scaffold	vertebrate
GCF_001642345.1	Rhincodon typus	Eukaryota	Metazoa	Chordata	Chondrichthyes	Orectolobiformes	Rhincodontidae	Rhincodon	scaffold	vertebrate
GCF_003031625.1	Empidonax traillii	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Tyrannidae	Empidonax	scaffold	vertebrate
GCF_000455745.1	Alligator sinensis	Eukaryota	Metazoa	Chordata	rcopterygii (SUPERCLAS	Crocodylia	Alligatoridae	Alligator	scaffold	vertebrate
GCF_001952655.1	Monopterus albus	Eukaryota	Metazoa	Chordata	Actinopteri	Synbranchiformes	Synbranchidae	Monopterus	scaffold	vertebrate
GCF_000344595.1	Chelonia mydas	Eukaryota	Metazoa	Chordata	rcopterygii (SUPERCLAS	Testudines	Chelonioidea	Chelonia	scaffold	vertebrate
GCF_900518725.1	Notechis scutatus	Eukaryota	Metazoa	Chordata	Lepidosauria	Squamata	Elapidae	Notechis	scaffold	vertebrate
GCF_001682695.1	Pygocentrus nattereri	Eukaryota	Metazoa	Chordata	Actinopteri	Characiformes	Serrasalmidae	Pygocentrus	scaffold	vertebrate
GCF_000971095.1	Anser cygnoides domesticus	Eukaryota	Metazoa	Chordata	Aves	Anseriformes	Anatidae	Anser	scaffold	vertebrate
GCF_900518735.1	Pseudonaja textilis	Eukaryota	Metazoa	Chordata	Lepidosauria	Squamata	Elapidae	Pseudonaja	scaffold	vertebrate
GCF_000703405.1	Apaloderma vittatum	Eukaryota	Metazoa	Chordata	Aves	Trogoniformes	Trogonidae	Apaloderma	scaffold	vertebrate
GCF_003343035.1	Apteryx rowi	Eukaryota	Metazoa	Chordata	Aves	Apterygiformes	Apterygidae	Apteryx	scaffold	vertebrate
GCF_001649575.1	Kryptolebias marmoratus	Eukaryota	Metazoa	Chordata	Actinopteri	Cyprinodontiformes	Rivulidae	Kryptolebias	scaffold	vertebrate
GCF_000239415.1	Haplochromis burtoni	Eukaryota	Metazoa	Chordata	Actinopteri	Cichliformes	Cichlidae	Haplochromis	scaffold	vertebrate
GCF_000337955.1	Falco peregrinus	Eukaryota	Metazoa	Chordata	Aves	Falconiformes	Falconidae	Falco	scaffold	vertebrate
GCF_000708025.1	Charadrius vociferus	Eukaryota	Metazoa	Chordata	Aves	Charadriiformes	Charadriidae	Charadrius	scaffold	vertebrate
GCF_000337975.1	Falco cherrug	Eukaryota	Metazoa	Chordata	Aves	Falconiformes	Falconidae	Falco	scaffold	vertebrate
GCF_000186305.1	Python bivittatus	Eukaryota	Metazoa	Chordata	Lepidosauria	Squamata	Pythonidae	Python	scaffold	vertebrate
GCF_004143745.1	Phasianus colchicus	Eukaryota	Metazoa	Chordata	Aves	Galliformes	Phasianidae	Phasianus	scaffold	vertebrate
GCF_001431845.1	Calidris pugnax	Eukaryota	Metazoa	Chordata	Aves	Charadriiformes	Scolopacidae	Calidris	scaffold	vertebrate
GCF_000708925.1	Phalacrocorax carbo	Eukaryota	Metazoa	Chordata	Aves	Pelecaniformes	Phalacrocoracidae	Phalacrocorax	scaffold	vertebrate
GCF_001970005.1	Paralichthys olivaceus	Eukaryota	Metazoa	Chordata	Actinopteri	Pleuronectiformes	Paralichthyidae	Paralichthys	scaffold	vertebrate
GCF_900067755.1	Pogona vitticeps	Eukaryota	Metazoa	Chordata	Lepidosauria	Squamata	Agamidae	Pogona	scaffold	vertebrate
GCF_000695195.1	Chlamydotis macqueenii	Eukaryota	Metazoa	Chordata	Aves	Gruiformes	Otididae	Chlamydotis	scaffold	vertebrate
GCF_000337935.1	Columba livia	Eukaryota	Metazoa	Chordata	Aves	Columbiformes	Columbidae	Columba	scaffold	vertebrate
GCF_001185365.1	Pantherophis guttatus	Eukaryota	Metazoa	Chordata	Lepidosauria	Squamata	Colubridae	Pantherophis	scaffold	vertebrate
GCF_000699145.1	Aptenodytes forsteri	Eukaryota	Metazoa	Chordata	Aves	Sphenisciformes	Spheniscidae	Aptenodytes	scaffold	vertebrate
GCF_000239395.1	Neolamprologus brichardi	Eukaryota	Metazoa	Chordata	Actinopteri	Cichliformes	Cichlidae	Neolamprologus	scaffold	vertebrate
GCF_000788275.1	Boleophthalmus pectinirostris	Eukaryota	Metazoa	Chordata	Actinopteri	Gobiiformes	Gobiidae	Boleophthalmus	scaffold	vertebrate
GCF_000239375.1	Pundamilia nyererei	Eukaryota	Metazoa	Chordata	Actinopteri	Cichliformes	Cichlidae	Pundamilia	scaffold	vertebrate
GCF_000738735.2	Corvus cornix cornix	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Corvidae	Corvus	scaffold	vertebrate
GCF_000699105.1	Pygoscelis adeliae	Eukaryota	Metazoa	Chordata	Aves	Sphenisciformes	Spheniscidae	Pygoscelis	scaffold	vertebrate
GCF_000935625.1	Nanorana parkeri	Eukaryota	Metazoa	Chordata	Amphibia	Anura	Dicroglossidae	Nanorana	scaffold	vertebrate
GCF_900080235.1	Labrus bergylta	Eukaryota	Metazoa	Chordata	Actinopteri	Labriformes	Labridae	Labrus	scaffold	vertebrate
GCF_000732505.1	Cyprinodon variegatus	Eukaryota	Metazoa	Chordata	Actinopteri	Cyprinodontiformes	Cyprinodontidae	Cyprinodon	scaffold	vertebrate
GCF_001723915.1	Gavialis gangeticus	Eukaryota	Metazoa	Chordata	rcopterygii (SUPERCLAS	Crocodylia	Gavialidae	Gavialis	scaffold	vertebrate
GCF_003984885.1	Neopelma chrysocephalum	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Pipridae	Neopelma	scaffold	vertebrate
GCF_007115625.1	Serinus canaria	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Fringillidae	Serinus	scaffold	vertebrate
GCF_000737465.1	Haliaeetus leucocephalus	Eukaryota	Metazoa	Chordata	Aves	Accipitriformes	Accipitridae	Haliaeetus	scaffold	vertebrate
GCF_000705375.1	Tinamus guttatus	Eukaryota	Metazoa	Chordata	Aves	Tinamiformes	Tinamidae	Tinamus	scaffold	vertebrate
GCF_000281125.3	Alligator mississippiensis	Eukaryota	Metazoa	Chordata	rcopterygii (SUPERCLAS	Crocodylia	Alligatoridae	Alligator	scaffold	vertebrate
GCF_000385455.1	Zonotrichia albicollis	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Passerellidae	Zonotrichia	scaffold	vertebrate
GCF_000826765.1	Fundulus heteroclitus	Eukaryota	Metazoa	Chordata	Actinopteri	Cyprinodontiformes	Fundulidae	Fundulus	scaffold	vertebrate
GCF_000687375.1	Pelecanus crispus	Eukaryota	Metazoa	Chordata	Aves	Pelecaniformes	Pelecanidae	Pelecanus	scaffold	vertebrate
GCF_000690715.1	Colius striatus	Eukaryota	Metazoa	Chordata	Aves	Coliiformes	Coliidae	Colius	scaffold	vertebrate
GCF_000709365.1	Tauraco erythrolophus	Eukaryota	Metazoa	Chordata	Aves	Musophagiformes	Musophagidae	Tauraco	scaffold	vertebrate
GCF_001447265.1	Sturnus vulgaris	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Sturnidae	Sturnus	scaffold	vertebrate
GCF_003597395.1	Chelonoidis abingdonii	Eukaryota	Metazoa	Chordata	rcopterygii (SUPERCLAS	Testudines	Testudinidae	Chelonoidis	scaffold	vertebrate
GCF_002109545.1	Acanthochromis polyacanthus	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Pomacentridae	Acanthochromis	scaffold	vertebrate
GCF_001527695.2	rotobothrops mucrosquamatu	Eukaryota	Metazoa	Chordata	Lepidosauria	Squamata	Viperidae	Protobothrops	scaffold	vertebrate
GCF_002814215.1	Seriola lalandi dorsalis	Eukaryota	Metazoa	Chordata	Actinopteri	Carangiformes	Carangidae	Seriola	scaffold	vertebrate
GCF_000690875.1	Gavia stellata	Eukaryota	Metazoa	Chordata	Aves	Gaviiformes	Gaviidae	Gavia	scaffold	vertebrate

REFSEQ ASSEMBLY	SPECIES	SUPERKINGDOM	KINGDOM	PHYLUM	CLASS	ORDER	FAMILY	GENUS	ASSEMBLY LEVEL	CLASSIFICATION
GCF_005870065.1	Oreochromis aureus	Eukaryota	Metazoa	Chordata	Actinopteri	Cichliformes	Cichlidae	Oreochromis	scaffold	vertebrate
GCF_000485575.1	Poecilia formosa	Eukaryota	Metazoa	Chordata	Actinopteri	Cyprinodontiformes	Poeciliidae	Poecilia	scaffold	vertebrate
GCF_000698965.1	Struthio camelus australis	Eukaryota	Metazoa	Chordata	Aves	Struthioniformes	Struthionidae	Struthio	scaffold	vertebrate
GCF_004355925.1	Anarrhichthys ocellatus	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Anarhichadidae	Anarrhichthys	scaffold	vertebrate
GCF_002260705.1	Seriola dumerili	Eukaryota	Metazoa	Chordata	Actinopteri	Carangiformes	Carangidae	Seriola	scaffold	vertebrate
GCF_000735185.1	Notothenia coriiceps	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Nototheniidae	Notothenia	scaffold	vertebrate
GCF_002925995.2	Terrapene carolina triunguis	Eukaryota	Metazoa	Chordata	rcopterygii (SUPERCLAS	Testudines	Em	Terrapene	scaffold	vertebrate
GCF_000695765.1	Mesitornis unicolor	Eukaryota	Metazoa	Chordata	Aves	Gruiformes	Mesitornithidae	Mesitornis	scaffold	vertebrate
GCF_000699245.1	Pterocles gutturalis	Eukaryota	Metazoa	Chordata	Aves	Ciconiiformes	Pteroclidae	Pterocles	scaffold	vertebrate
GCF_001447785.1	Gekko japonicus	Eukaryota	Metazoa	Chordata	Lepidosauria	Squamata	Gekkonidae	Gekko	scaffold	vertebrate
GCF_001723895.1	Crocodylus porosus	Eukaryota	Metazoa	Chordata	rcopterygii (SUPERCLAS	Crocodylia	Crocodylidae	Crocodylus	scaffold	vertebrate
GCF_003342905.1	Dromaius novaehollandiae	Eukaryota	Metazoa	Chordata	Aves	Casuariiformes	Dromaiidae	Dromaius	scaffold	vertebrate
GCF_000696875.1	Nestor notabilis	Eukaryota	Metazoa	Chordata	Aves	Psittaciformes	Psittacidae	Nestor	scaffold	vertebrate
GCF_000710305.1	Buceros rhinoceros silvestris	Eukaryota	Metazoa	Chordata	Aves	Bucerotiformes	Bucerotidae	Buceros	scaffold	vertebrate
GCF_001515605.1	Sinocyclocheilus anshuiensis	Eukaryota	Metazoa	Chordata	Actinopteri	Cypriniformes	Cyprinidae	Sinocyclocheilus	scaffold	vertebrate
GCF_003945595.1	Pipra filicauda	Eukaryota	Metazoa	Chordata	Aves	Passeriformes	Pipridae	Pipra	scaffold	vertebrate
GCF_902827175.1	Gymnodraco acuticeps	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Bathydraconidae	Gymnodraco	scaffold	vertebrate
GCF_008315115.1	Sander lucioperca	Eukaryota	Metazoa	Chordata	Actinopteri	Perciformes	Percidae	Sander	scaffold	vertebrate
GCF_002922805.1	Oryzias melastigma	Eukaryota	Metazoa	Chordata	Actinopteri	Beloniformes	Adrianichthyidae	Oryzias	scaffold	vertebrate
GCF_001039765.1	Apteryx mantelli mantelli	Eukaryota	Metazoa	Chordata	Aves	Apterygiformes	Apterygidae	Apteryx	contig	vertebrate
GCF_000002985.6	Caenorhabditis elegans	Eukaryota	Metazoa	Nematoda	Chromadorea	Rhabditida	Rhabditidae	Caenorhabditis	complete genome	invertebrate
GCF_004354385.1	Drosophila innubila	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	chromosome	invertebrate
GCF_009650485.1	Drosophila albomicans	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	chromosome	invertebrate
GCF_003672135.1	Ooceraea biroi	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Ooceraea	chromosome	invertebrate
GCF_902459465.1	Asterias rubens	Eukaryota	Metazoa	Echinodermata	Asteroidea	Forcipulatida	Asteriidae	Asterias	chromosome	invertebrate
GCF_002204515.2	Aedes aegypti	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Culicidae	Aedes	chromosome	invertebrate
GCF_006345805.1	Octopus vulgaris	Eukaryota	Metazoa	Mollusca	Cephalopoda	Octopoda	Octopodidae	Octopus	chromosome	invertebrate
GCF_000002335.3	Tribolium castaneum	Eukaryota	Metazoa	Arthropoda	Insecta	Coleoptera	Tenebrionidae	Tribolium	chromosome	invertebrate
GCF_005508785.1	Acyrthosiphon pisum	Eukaryota	Metazoa	Arthropoda	Insecta	Hemiptera	Aphididae	Acyrthosiphon	chromosome	invertebrate
GCF_000237925.1	Schistosoma mansoni	Eukaryota	Metazoa	Platyhelminthes	Trematoda	Strigeidida	Schistosomatidae	Schistosoma	chromosome	invertebrate
GCF_003254395.2	Apis mellifera	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Apidae	Apis	chromosome	invertebrate
GCF_009870125.1	Drosophila pseudoobscura	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	chromosome	invertebrate
GCF_002022765.2	Crassostrea virginica	Eukaryota	Metazoa	Mollusca	Bivalvia	Ostreoida	Ostreidae	Crassostrea	chromosome	invertebrate
GCF_000001215.4	Drosophila melanogaster	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	chromosome	invertebrate
GCF_003369915.1	Drosophila miranda	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	chromosome	invertebrate
GCF_000224145.3	Ciona intestinalis	Eukaryota	Metazoa	Chordata	Ascidiacea	Enterogona	Cionidae	Ciona	chromosome	invertebrate
GCF_003590095.1	Trichoplusia ni	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Noctuidae	Trichoplusia	chromosome	invertebrate
GCF_002706865.1	Spodoptera litura	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Noctuidae	Spodoptera	chromosome	invertebrate
GCF_902652985.1	Pecten maximus	Eukaryota	Metazoa	Mollusca	Bivalvia	Pectinoida	Pectinidae	Pecten	chromosome	invertebrate
GCF_902806685.1	Aphantopus hyperantus	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Nymphalidae	Aphantopus	chromosome	invertebrate
GCF_000754195.2	Drosophila simulans	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	chromosome	invertebrate
GCF_000004555.1	Caenorhabditis briggsae	Eukaryota	Metazoa	Nematoda	Chromadorea	Rhabditida	Rhabditidae	Caenorhabditis	chromosome	invertebrate
GCF_004382195.1	Drosophila sechellia	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	chromosome	invertebrate
GCF_011750605.1	Drosophila busckii	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	chromosome	invertebrate
GCF_004382145.1	Drosophila mauritiana	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	chromosome	invertebrate
GCF_009193385.2	Nasonia vitripennis	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Pteromalidae	Nasonia	chromosome	invertebrate
GCF_003990815.1	Daphnia magna	Eukaryota	Metazoa	Arthropoda	Branchiopoda	Diplostraca	Daphniidae	Daphnia	chromosome	invertebrate
GCF_000214255.1	Bombus terrestris	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Apidae	Bombus	chromosome	invertebrate
GCF_009731565.1	Danaus plexippus plexippus	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Nymphalidae	Danaus	chromosome	invertebrate
GCF_003676215.2	Rhopalosiphum maidis	Eukaryota	Metazoa	Arthropoda	Insecta	Hemiptera	Aphididae	Rhopalosiphum	chromosome	invertebrate
GCF_010883055.1	Belonocnema treatae	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Cynipidae	Belonocnema	chromosome	invertebrate
GCF_003073045.1	Pomacea canaliculata	Eukaryota	Metazoa	Mollusca	Gastropoda	Architaenioglossa	Ampullariidae	Pomacea	chromosome	invertebrate
GCF_000005575.2	Anopheles gambiae str. PEST	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Culicidae	Anopheles	chromosome	invertebrate
GCF_008121235.1	Drosophila subobscura	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	chromosome	invertebrate

REFSEQ ASSEMBLY	SPECIES	SUPERKINGDOM	KINGDOM	PHYLUM	CLASS	ORDER	FAMILY	GENUS	ASSEMBLY LEVEL	CLASSIFICATION
GCF_902806645.1	Crassostrea gigas	Eukaryota	Metazoa	Mollusca	Bivalvia	Ostreoida	Ostreidae	Crassostrea	chromosome	invertebrate
GCF_001040885.1	Strongyloides ratti	Eukaryota	Metazoa	Nematoda	Chromadorea	Rhabditida	Strongyloididae	Strongyloides	chromosome	invertebrate
GCF_000005975.2	Drosophila yakuba	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	chromosome	invertebrate
GCF_003227725.1	Camponotus floridanus	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Camponotus	scaffold	invertebrate
GCF_001194135.1	Octopus bimaculoides	Eukaryota	Metazoa	Mollusca	Cephalopoda	Octopoda	Octopodidae	Octopus	scaffold	invertebrate
GCF_001901225.1	ermatophagoides pteronyssinu	Eukaryota	Metazoa	Arthropoda	Arachnida	Sarcoptiformes	Pyroglyphidae	Dermatophagoides	scaffold	invertebrate
GCF_001594045.1	Atta colombica	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Atta	scaffold	invertebrate
GCF_009176525.2	Contarinia nasturtii	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Cecidomyiidae	Contarinia	scaffold	invertebrate
GCF_001298355.1	Papilio machaon	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Papilionidae	Papilio	scaffold	invertebrate
GCF_001465965.1	Polistes dominula	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Vespidae	Polistes	scaffold	invertebrate
GCF_000236325.1	Drosophila eugracilis	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	scaffold	invertebrate
GCF_003227715.1	Harpegnathos saltator	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Harpegnathos	scaffold	invertebrate
GCF_001594055.1	Trachymyrmex zeteki	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Trachymyrmex	scaffold	invertebrate
GCF_000956235.1	Wasmannia auropunctata	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Wasmannia	scaffold	invertebrate
GCF_000002075.1	Aplysia californica	Eukaryota	Metazoa	Mollusca	Gastropoda	Aplysiida	Aplysiidae	Aplysia	scaffold	invertebrate
GCF_000209185.1	Culex quinquefasciatus	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Culicidae	Culex	scaffold	invertebrate
GCF_002443255.1	Varroa destructor	Eukaryota	Metazoa	Arthropoda	Arachnida	Mesostigmata	Varroidae	Varroa	scaffold	invertebrate
GCF_000671375.1	Centruroides sculpturatus	Eukaryota	Metazoa	Arthropoda	Arachnida	Scorpiones	Buthidae	Centruroides	scaffold	invertebrate
GCF_001263275.1	Habropoda laboriosa	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Apidae	Habropoda	scaffold	invertebrate
GCF_003640425.2	Galleria mellonella	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Pyralidae	Galleria	scaffold	invertebrate
GCF_000002995.3	Brugia malayi	Eukaryota	Metazoa	Nematoda	Chromadorea	Rhabditida	Onchocercidae	Brugia	scaffold	invertebrate
GCF_000326865.1	Helobdella robusta	Eukaryota	Metazoa	Annelida	Clitellata	Hirudinida	Glossiphoniidae	Helobdella	scaffold	invertebrate
GCF_000836215.1	Papilio polytes	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Papilionidae	Papilio	scaffold	invertebrate
GCF_000209225.1	Nematostella vectensis	Eukaryota	Metazoa	Cnidaria	Anthozoa	Actiniaria	Edwardsiidae	Nematostella	scaffold	invertebrate
GCF_011952275.1	mbus vancouverensis nearctic	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Apidae	Bombus	scaffold	invertebrate
GCF_000217595.1	Linepithema humile	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Linepithema	scaffold	invertebrate
GCF_000789215.1	Bactrocera dorsalis	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Tephritidae	Bactrocera	scaffold	invertebrate
GCF_000005175.2	Drosophila mojavensis	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	scaffold	invertebrate
GCF_000524195.1	Echinococcus granulosus	Eukaryota	Metazoa	Platyhelminthes	Cestoda	Cyclophyllidea	Taeniidae	Echinococcus	scaffold	invertebrate
GCF_000327385.1	Lottia gigantea	Eukaryota	Metazoa	Mollusca	Gastropoda	Patellogastropoda	Lottidae	Lottia	scaffold	invertebrate
GCF_000236305.1	Drosophila rhopaloa	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	scaffold	invertebrate
GCF_000224215.1	Drosophila kikkawai	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	scaffold	invertebrate
GCF_003704095.1	Pocillopora damicornis	Eukaryota	Metazoa	Cnidaria	Anthozoa	Scleractinia	Pocilloporidae	Pocillopora	scaffold	invertebrate
GCF_003651465.1	Formica exsecta	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Formica	scaffold	invertebrate
GCF_010614865.1	Stegodyphus dumicola	Eukaryota	Metazoa	Arthropoda	Arachnida	Araneae	Eresidae	Stegodyphus	scaffold	invertebrate
GCF_000648655.2	Copidosoma floridanum	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Encyrtidae	Copidosoma	scaffold	invertebrate
GCF_000612105.2	Orussus abietinus	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Orussidae	Orussus	scaffold	invertebrate
GCF_000757685.1	Nilaparvata lugens	Eukaryota	Metazoa	Arthropoda	Insecta	Hemiptera	Delphacidae	Nilaparvata	scaffold	invertebrate
GCF_001186385.1	Diuraphis noxia	Eukaryota	Metazoa	Arthropoda	Insecta	Hemiptera	Aphididae	Diuraphis	scaffold	invertebrate
GCF_000184785.3	Apis florea	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Apidae	Apis	scaffold	invertebrate
GCF_001854935.1	Bemisia tabaci	Eukaryota	Metazoa	Arthropoda	Insecta	Hemiptera	Aleyrodidae	Bemisia	scaffold	invertebrate
GCF_001483705.1	Eufriesea mexicana	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Apidae	Eufriesea	scaffold	invertebrate
GCF_900239965.1	Bicyclus anynana	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Nymphalidae	Bicyclus	scaffold	invertebrate
GCF_000149515.1	Caenorhabditis remanei	Eukaryota	Metazoa	Nematoda	Chromadorea	Rhabditida	Rhabditidae	Caenorhabditis	scaffold	invertebrate
GCF_000715545.1	Opisthorchis viverrini	Eukaryota	Metazoa	Platyhelminthes	Trematoda	Opisthorchiida	Opisthorchiidae	Opisthorchis	scaffold	invertebrate
GCF_000006295.1	Pediculus humanus corporis	Eukaryota	Metazoa	Arthropoda	Insecta	Phthiraptera	Pediculidae	Pediculus	scaffold	invertebrate
GCF_000500325.1	Leptinotarsa decemlineata	Eukaryota	Metazoa	Arthropoda	Insecta	Coleoptera	Chrysomelidae	Leptinotarsa	scaffold	invertebrate
GCF_001594065.1	Cyphomyrmex costatus	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Cyphomyrmex	scaffold	invertebrate
GCF_000949405.1	Vollenhovia emeryi	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Vollenhovia	scaffold	invertebrate
GCF_000181795.1	Trichinella spiralis	Eukaryota	Metazoa	Nematoda	Enoplea	Trichinellida	Trichinellidae	Trichinella	scaffold	invertebrate
GCF_000003605.2	Saccoglossus kowalevskii	Eukaryota	Metazoa	Hemichordata	Enteropneusta	Enteropneusta	Harrimaniidae	Saccoglossus	scaffold	invertebrate
GCF_000188095.3	Bombus impatiens	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Apidae	Bombus	scaffold	invertebrate
GCF_001442555.1	Apis cerana	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Apidae	Apis	scaffold	invertebrate
GCF_003589595.1	Hyposmocoma kahamanoa	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Cosmopterigidae	Hyposmocoma	scaffold	invertebrate

8

REFSEQ ASSEMBLY	SPECIES	SUPERKINGDOM	KINGDOM	PHYLUM	CLASS	ORDER	FAMILY	GENUS	ASSEMBLY LEVEL	CLASSIFICATION
GCF_000806365.1	Fopius arisanus	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Braconidae	Fopius	scaffold	invertebrate
GCF_001853355.1	Bactrocera latifrons	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Tephritidae	Bactrocera	scaffold	invertebrate
GCF_001313835.1	Polistes canadensis	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Vespidae	Polistes	scaffold	invertebrate
GCF_000365465.2	Parasteatoda tepidariorum	Eukaryota	Metazoa	Arthropoda	Arachnida	Araneae	Theridiidae	Parasteatoda	scaffold	invertebrate
GCF_000371365.1	Musca domestica	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Muscidae	Musca	scaffold	invertebrate
GCF_013357705.1	Chelonus insularis	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Braconidae	Chelonus	scaffold	invertebrate
GCF_000648675.2	Cimex lectularius	Eukaryota	Metazoa	Arthropoda	Insecta	Hemiptera	Cimicidae	Cimex	scaffold	invertebrate
GCF_000390285.2	Anoplophora glabripennis	Eukaryota	Metazoa	Arthropoda	Insecta	Coleoptera	Cerambycidae	Anoplophora	scaffold	invertebrate
GCF_003260585.2	Monomorium pharaonis	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Monomorium	scaffold	invertebrate
GCF_003987935.1	Bombyx mandarina	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Bombycidae	Bombyx	scaffold	invertebrate
GCF_000485595.1	Priapulus caudatus	Eukaryota	Metazoa	Priapulida	Priapulimorpha	Priapulimorphida	Priapulidae	Priapulus	scaffold	invertebrate
GCF_000090795.1	Amphimedon queenslandica	Eukaryota	Metazoa	Porifera	Demospongiae	Haplosclerida	Niphatidae	Amphimedon	scaffold	invertebrate
GCF_001654015.2	Drosophila navojoa	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	scaffold	invertebrate
GCF_000341935.1	Cephus cinctus	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Cephidae	Cephus	scaffold	invertebrate
GCF_000255335.1	Galendromus occidentalis	Eukaryota	Metazoa	Arthropoda	Arachnida	Mesostigmata	Phytoseiidae	Galendromus	scaffold	invertebrate
GCF_000503995.1	Ceratosolen solmsi marchali	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Agaonidae	Ceratosolen	scaffold	invertebrate
GCF_000507365.1	Necator americanus	Eukaryota	Metazoa	Nematoda	Chromadorea	Strongylida	Ancylostomatidae	Necator	scaffold	invertebrate
GCF_002042975.1	Orbicella faveolata	Eukaryota	Metazoa	Cnidaria	Anthozoa	Scleractinia	Merulinidae	Orbicella	scaffold	invertebrate
GCF_000696155.1	Zootermopsis nevadensis	Eukaryota	Metazoa	Arthropoda	Insecta	Blattodea	Termopsidae	Zootermopsis	scaffold	invertebrate
GCF_001313825.1	Dinoponera quadriceps	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Dinoponera	scaffold	invertebrate
GCF_004153925.1	Osmia bicornis bicornis	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Megachilidae	Osmia	scaffold	invertebrate
GCF_000599845.2	Trichogramma pretiosum	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Trichogrammatidae	Trichogramma	scaffold	invertebrate
GCF_000648695.1	Onthophagus taurus	Eukaryota	Metazoa	Arthropoda	Insecta	Coleoptera	Scarabaeidae	Onthophagus	scaffold	invertebrate
GCF_008802855.1	Photinus pyralis	Eukaryota	Metazoa	Arthropoda	Insecta	Coleoptera	Lampyridae	Photinus	scaffold	invertebrate
GCF_000236285.1	Drosophila bipectinata	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	scaffold	invertebrate
GCF_001687245.1	Rhagoletis zephyria	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Tephritidae	Rhagoletis	scaffold	invertebrate
GCF_002217835.1	Drosophila obscura	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	scaffold	invertebrate
GCF_000696795.2	Halyomorpha halys	Eukaryota	Metazoa	Arthropoda	Insecta	Hemiptera	Pentatomidae	Halyomorpha	scaffold	invertebrate
GCF_000472105.1	Drosophila suzukii	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	scaffold	invertebrate
GCF_010583005.1	Odontomachus brunneus	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Odontomachus	scaffold	invertebrate
GCF_000224195.1	Drosophila elegans	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	scaffold	invertebrate
GCF_000005925.1	Drosophila willistoni	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	scaffold	invertebrate
GCF_000239435.1	Tetranychus urticae	Eukaryota	Metazoa	Arthropoda	Arachnida	Trombidiformes	Tetranychidae	Tetranychus	scaffold	invertebrate
GCF_001015335.1	Stomoxys calcitrans	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Muscidae	Stomoxys	scaffold	invertebrate
GCF_004143615.1	Acropora millepora	Eukaryota	Metazoa	Cnidaria	Anthozoa	Scleractinia	Acroporidae	Acropora	scaffold	invertebrate
GCF_000355655.1	Dendroctonus ponderosae	Eukaryota	Metazoa	Arthropoda	Insecta	Coleoptera	Curculionidae	Dendroctonus	scaffold	invertebrate
GCF_003070985.1	Temnothorax curvispinosus	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Temnothorax	scaffold	invertebrate
GCF_003710045.1	Nomia melanderi	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Halictidae	Nomia	scaffold	invertebrate
GCF_000220905.1	Megachile rotundata	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Megachilidae	Megachile	scaffold	invertebrate
GCF_000572035.2	Microplitis demolitor	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Braconidae	Microplitis	scaffold	invertebrate
GCF_003789085.1	Penaeus vannamei	Eukaryota	Metazoa	Arthropoda	Malacostraca	Decapoda	Penaeidae	Penaeus	scaffold	invertebrate
GCF_002803265.2	Melanaphis sacchari	Eukaryota	Metazoa	Arthropoda	Insecta	Hemiptera	Aphididae	Melanaphis	scaffold	invertebrate
GCF_003013835.1	Diabrotica virgifera virgifera	Eukaryota	Metazoa	Arthropoda	Insecta	Coleoptera	Chrysomelidae	Diabrotica	scaffold	invertebrate
GCF_000475195.1	Diaphorina citri	Eukaryota	Metazoa	Arthropoda	Insecta	Hemiptera	Liviidae	Diaphorina	scaffold	invertebrate
GCF_000699065.1	Lucilia cuprina	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Calliphoridae	Lucilia	scaffold	invertebrate
GCF_000836235.1	Papilio xuthus	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Papilionidae	Papilio	scaffold	invertebrate
GCF_000183805.2	Loa loa	Eukaryota	Metazoa	Nematoda	Chromadorea	Rhabditida	Onchocercidae	Loa	scaffold	invertebrate
GCF_000188075.2	Solenopsis invicta	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Solenopsis	scaffold	invertebrate
GCF_000143395.1	Atta cephalotes	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Atta	scaffold	invertebrate
GCF_009602425.1	Actinia tenebrosa	Eukaryota	Metazoa	Cnidaria	Anthozoa	Actiniaria	Actiniidae	Actinia	scaffold	invertebrate
GCF_000004095.1	Hydra vulgaris	Eukaryota	Metazoa	Cnidaria	Hydrozoa	Anthoathecata	Hydridae	Hydra	scaffold	invertebrate
GCF_000469605.1	Apis dorsata	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Apidae	Apis	scaffold	invertebrate
GCF_000517525.1	Limulus polyphemus	Eukaryota	Metazoa	Arthropoda	Merostomata	Xiphosura	Limulidae	Limulus	scaffold	invertebrate
GCF_000699045.2	Agrilus planipennis	Eukaryota	Metazoa	Arthropoda	Insecta	Coleoptera	Buprestidae	Agrilus	scaffold	invertebrate

REFSEQ ASSEMBLY	SPECIES	SUPERKINGDOM	KINGDOM	PHYLUM	CLASS	ORDER	FAMILY	GENUS	ASSEMBLY LEVEL	CLASSIFICATION
GCF_000222465.1	Acropora digitifera	Eukaryota	Metazoa	Cnidaria	Anthozoa	Scleractinia	Acroporidae	Acropora	scaffold	invertebrate
GCF_002532875.1	Varroa jacobsoni	Eukaryota	Metazoa	Arthropoda	Arachnida	Mesostigmata	Varroidae	Varroa	scaffold	invertebrate
GCF_000591075.1	Eurytemora affinis	Eukaryota	Metazoa	Arthropoda	Hexanauplia	Calanoida	Temoridae	Eurytemora	scaffold	invertebrate
GCF_000151625.1	Bombyx mori	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Bombycidae	Bombyx	scaffold	invertebrate
GCF_000204515.1	Acromyrmex echinatior	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Acromyrmex	scaffold	invertebrate
GCF_011952255.1	Bombus vosnesenskii	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Apidae	Bombus	scaffold	invertebrate
GCF_012932325.1	Thrips palmi	Eukaryota	Metazoa	Arthropoda	Insecta	Thysanoptera	Thripidae	Thrips	scaffold	invertebrate
GCF_006496715.1	Aedes albopictus	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Culicidae	Aedes	scaffold	invertebrate
GCF_011952205.1	Bombus bifarius	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Apidae	Bombus	scaffold	invertebrate
GCF_005281655.1	Nylanderia fulva	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Nylanderia	scaffold	invertebrate
GCF_001594115.1	Trachymyrmex septentrionalis	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Trachymyrmex	scaffold	invertebrate
GCF_000806345.1	Zeugodacus cucurbitae	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Tephritidae	Zeugodacus	scaffold	invertebrate
GCF_001186105.1	Amyelois transitella	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Pyralidae	Amyelois	scaffold	invertebrate
GCF_002571385.1	Stylophora pistillata	Eukaryota	Metazoa	Cnidaria	Anthozoa	Scleractinia	Pocilloporidae	Stylophora	scaffold	invertebrate
GCF_002217175.1	Folsomia candida	Eukaryota	Metazoa	Arthropoda	Collembola	Entomobryomorpha	Isotomidae	Folsomia	scaffold	invertebrate
GCF_003426905.1	Ctenocephalides felis	Eukaryota	Metazoa	Arthropoda	Insecta	Siphonaptera	Pulicidae	Ctenocephalides	scaffold	invertebrate
GCF_001654025.1	Drosophila arizonae	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	scaffold	invertebrate
GCF_002891405.2	Cryptotermes secundus	Eukaryota	Metazoa	Arthropoda	Insecta	Blattodea	Kalotermitidae	Cryptotermes	scaffold	invertebrate
GCF_001263575.1	Neodiprion lecontei	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Diprionidae	Neodiprion	scaffold	invertebrate
GCF_000233415.1	Drosophila biarmipes	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	scaffold	invertebrate
GCF_011865705.1	Megalopta genalis	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Halictidae	Megalopta	scaffold	invertebrate
GCF_000330985.1	Plutella xylostella	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Plutellidae	Plutella	scaffold	invertebrate
GCF_000764305.1	Hyalella azteca	Eukaryota	Metazoa	Arthropoda	Malacostraca	Amphipoda	Hyalellidae	Hyalella	scaffold	invertebrate
GCF_900245975.1	Drosophila guanche	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	scaffold	invertebrate
GCF_000187915.1	Pogonomyrmex barbatus	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Pogonomyrmex	scaffold	invertebrate
GCF_001625305.1	Branchiostoma belcheri	Eukaryota	Metazoa	Chordata	Leptocardii	Amphioxiformes	Branchiostomidae	Branchiostoma	scaffold	invertebrate
GCF_004193835.1	Ostrinia furnacalis	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Crambidae	Ostrinia	scaffold	invertebrate
GCF_001652005.1	Ceratina calcarata	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Apidae	Ceratina	scaffold	invertebrate
GCF_001412515.2	Diachasma alloeum	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Braconidae	Diachasma	scaffold	invertebrate
GCF_011630105.1	Anneissia japonica	Eukaryota	Metazoa	Echinodermata	Crinoidea	Comatulida	Comatulidae	Anneissia	scaffold	invertebrate
GCF_001949145.1	Acanthaster planci	Eukaryota	Metazoa	Echinodermata	Asteroidea	Valvatida	Acanthasteridae	Acanthaster	scaffold	invertebrate
GCF_000220665.1	Drosophila ficusphila	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	scaffold	invertebrate
GCF_002113885.1	Mizuhopecten yessoensis	Eukaryota	Metazoa	Mollusca	Bivalvia	Pectinoida	Pectinidae	Mizuhopecten	scaffold	invertebrate
GCF_000262585.1	Manduca sexta	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Sphingidae	Manduca	scaffold	invertebrate
GCF_000457365.1	Biomphalaria glabrata	Eukaryota	Metazoa	Mollusca	Gastropoda	Basommatophora	Planorbidae	Biomphalaria	scaffold	invertebrate
GCF_002938995.1	Vanessa tameamea	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Nymphalidae	Vanessa	scaffold	invertebrate
GCF_000002235.5	Strongylocentrotus purpuratus	Eukaryota	Metazoa	Echinodermata	Echinoidea	Echinoida	Strongylocentrotidae	Strongylocentrotus	scaffold	invertebrate
GCF_002938485.1	Sitophilus oryzae	Eukaryota	Metazoa	Arthropoda	Insecta	Coleoptera	Curculionidae	Sitophilus	scaffold	invertebrate
GCF_001412225.1	Nicrophorus vespilloides	Eukaryota	Metazoa	Arthropoda	Insecta	Coleoptera	Silphidae	Nicrophorus	scaffold	invertebrate
GCF_000697945.2	Frankliniella occidentalis	Eukaryota	Metazoa	Arthropoda	Insecta	Thysanoptera	Thripidae	Frankliniella	scaffold	invertebrate
GCF_000344095.2	Athalia rosae	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Tenthredinidae	Athalia	scaffold	invertebrate
GCF_001856805.1	Pieris rapae	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Pieridae	Pieris	scaffold	invertebrate
GCF_003268045.1	Sipha flava	Eukaryota	Metazoa	Arthropoda	Insecta	Hemiptera	Aphididae	Sipha	scaffold	invertebrate
GCF_000699445.1	Schistosoma haematobium	Eukaryota	Metazoa	Platyhelminthes	Trematoda	Strigeidida	Schistosomatidae	Schistosoma	scaffold	invertebrate
GCF_002006095.1	Pseudomyrmex gracilis	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Pseudomyrmex	scaffold	invertebrate
GCF_000224235.1	Drosophila takahashii	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	scaffold	invertebrate
GCF_000347755.3	Ceratitis capitata	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Tephritidae	Ceratitis	scaffold	invertebrate
GCF_000150275.1	Trichoplax adhaerens	Eukaryota	Metazoa	Placozoa	Trichoplacoidea	Trichoplacoidea	Trichoplacidae	Trichoplax	scaffold	invertebrate
GCF_000005155.2	Drosophila grimshawi	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	scaffold	invertebrate
GCF_001594075.1	Trachymyrmex cornetzi	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Formicidae	Trachymyrmex	scaffold	invertebrate
GCF_001039355.2	Lingula anatina	Eukaryota	Metazoa	Brachiopoda	Lingulata	Lingulida	Lingulidae	Lingula	scaffold	invertebrate
GCF_002156985.1	Helicoverpa armigera	Eukaryota	Metazoa	Arthropoda	Insecta	Lepidoptera	Noctuidae	Helicoverpa	scaffold	invertebrate
GCF_004010815.1	Aphis gossypii	Eukaryota	Metazoa	Arthropoda	Insecta	Hemiptera	Aphididae	Aphis	scaffold	invertebrate
GCF_000003815.1	Branchiostoma floridae	Eukaryota	Metazoa	Chordata	Leptocardii	Amphioxiformes	Branchiostomidae	Branchiostoma	scaffold	invertebrate

REFSEQ ASSEMBLY	SPECIES	SUPERKINGDOM	KINGDOM	PHYLUM	CLASS	ORDER	FAMILY	GENUS	ASSEMBLY LEVEL	CLASSIFICATION
GCF_001188975.1	Bactrocera oleae	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Tephritidae	Bactrocera	scaffold	invertebrate
GCF_001417965.1	Exaiptasia diaphana	Eukaryota	Metazoa	Cnidaria	Anthozoa	Actiniaria	Aiptasiidae	Exaiptasia	scaffold	invertebrate
GCF_000188695.1	Salpingoeca rosetta	Eukaryota	Fungi/Metazoa	Choanoflagellata	Choanoflagellata	Craspedida	Salpingoecidae	Salpingoeca	scaffold	invertebrate
GCF_001856785.1	Myzus persicae	Eukaryota	Metazoa	Arthropoda	Insecta	Hemiptera	Aphididae	Myzus	scaffold	invertebrate
GCF_001272555.1	Dufourea novaeangliae	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Halictidae	Dufourea	scaffold	invertebrate
GCF_002892825.2	Ixodes scapularis	Eukaryota	Metazoa	Arthropoda	Arachnida	Ixodida	Ixodidae	Ixodes	contig	invertebrate
GCF_003285975.2	Drosophila ananassae	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	contig	invertebrate
GCF_003285875.2	Drosophila novamexicana	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	contig	invertebrate
GCF_001937115.1	Aethina tumida	Eukaryota	Metazoa	Arthropoda	Insecta	Coleoptera	Nitidulidae	Aethina	contig	invertebrate
GCF_003285735.1	Drosophila virilis	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	contig	invertebrate
GCF_003286155.1	Drosophila erecta	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	contig	invertebrate
GCF_012274295.1	Osmia lignaria	Eukaryota	Metazoa	Arthropoda	Insecta	Hymenoptera	Megachilidae	Osmia	contig	invertebrate
GCF_003285725.1	Scaptodrosophila lebanonensis	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Scaptodrosophila	contig	invertebrate
GCF_004324835.1	Dendronephthya gigantea	Eukaryota	Metazoa	Cnidaria	Anthozoa	Alcyonacea	Nephtheidae	Dendronephthya	contig	invertebrate
GCF_003286085.1	Drosophila persimilis	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	contig	invertebrate
GCF_002093755.1	Drosophila serrata	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	contig	invertebrate
GCF_003285905.1	Drosophila hydei	Eukaryota	Metazoa	Arthropoda	Insecta	Diptera	Drosophilidae	Drosophila	contig	invertebrate
GCF_000146045.2	Saccharomyces cerevisiae	Eukaryota	Fungi	Ascomycota	Saccharomycetes	Saccharomycetales	Saccharomycetaceae	Saccharomyces	complete genome	invertebrate
GCF_000226095.1	hermothelomyces thermophilu	Eukaryota	Fungi	Ascomycota	Sordariomycetes	Sordariales	Chaetomiaceae	Thermothelomyces	complete genome	invertebrate
GCF_000182925.2	Neurospora classa	Eukaryota	Fungi	Ascomycota	Sordariomycetes	Sordariales	Sordariaceae	Neurospora	chromosome	invertebrate
GCF_000002655.1	Aspergillus fumigatus	Eukaryota	Fungi	Ascomycota	Eurotiomycetes	Eurotiales	Aspergillaceae	Aspergillus	chromosome	invertebrate
GCF_000002495.2	Pyricularia oryzae	Eukaryota	Fungi	Ascomycota	Sordariomycetes	Magnaporthales	Pyriculariaceae	Pyricularia	chromosome	invertebrate
GCF_000219625.1	Zymoseptoria tritici	Eukaryota	Fungi	Ascomycota	Dothideomycetes	Mycosphaerellales	Mycosphaerellaceae	Zymoseptoria	chromosome	invertebrate
GCF_000240135.3	Fusarium graminearum	Eukaryota	Fungi	Ascomycota	Sordariomycetes	Hypocreales	Nectriaceae	Fusarium	chromosome	invertebrate
GCF_001653235.2	Pochonia chlamydosporia	Eukaryota	Fungi	Ascomycota	Sordariomycetes	Hypocreales	Clavicipitaceae	Pochonia	chromosome	invertebrate
GCF_000002945.1	Schizosaccharomyces pombe	Eukaryota	Fungi	Ascomycota	Schizosaccharomycetes	Schizosaccharomycetales	Schizosaccharomycetaceae	Schizosaccharomyces	chromosome	invertebrate

Supplementary_Table_3 Oxygen Km values of PHD and FIH

Category	Enzyme	Organism	$K_m^{app}(O_2)$ [uM]	Protein substrate	Reference
PHD	*Pseudomonas* Prolyl Hydroxylase(PPHD)	*Pseudomonas aeruginosa*	47.5±9.2	pH 7.5, 37°C, 30min, 20-residue elongation factor Tu fragment	Scotti et al. 2014
PHD	Prolyl L-hydroxylase(PLH)	*Vinca rosea*	60	poly-L-proline	Tanaka et al. 1980
PHD	Prolyl L-hydroxylase(PLH)	*Vinca rosea*	60	poly(L-proline)	Chrispeels 1984
PHD	*Trichoplax adhaerens* PHD(TaPHD$^{64\text{-}300}$)	*Trichoplax adhaerens*	>400	TaHIFa ODD 25mer peptide	Lippl et al. 2018
PHD	Prolyl-hydroxylase (PH)	*Gallus gallus*	40-45	(Pro-Pro-Gly)5 · $4H_2O$ or (Pro-Pro-Gly) · $9H_2O$	Myllylä et al. 1977
PHD	Prolyl-hydroxylase (PH)	*Gallus gallus*	72	collagen	Vanderkooi et al. 1991
PHD	Prolyl-hydroxylase (PH)	*Homo sapiens*	45	collagen	Cianci 2004
PHD	Prolyl-3-hydroxylase (P3H)	*Gallus gallus*	30	a [2,3-^{3}H]proline-labelled biologically prepared polypeptide	Tryggvason et al. 1979
PHD	Prolyl-4-hydroxylase (P4H)	*Gallus gallus*	190	(Pro-Pro-Gly)$_5$	De Jong and Kemp 1984
PHD	collagen prolyl-4-hydroxylase (CPH)	*Cavia porcellus*	30	collagen	Hutton et al. 1967
PHD	collagen prolyl-4-hydroxylase (CPH)	*Gallus gallus*	30	collagen	Hutton et al. 1967
PHD	collagen prolyl-4-hydroxylase (CPH)	*Rattus norvegicus*	30	collagen	Hutton et al. 1967
PHD	Protocollagen Proline Hydroxylase	*Ascaris lumbricoides*	11.54	proline-labeled protocollagen from the cuticle of Ascaris	Fujimoto and Prockop 1969
PHD	Protocollagen Proline Hydroxylase	*Gallus gallus*	23.08	proline-labeled protocollagen from the cuticle of Ascaris	Fujimoto and Prockop 1969
PHD	Protocollagen hydroxylase (PCH)	*Ascaris lumbricoides*	12.82	^{14}C-proline-labeled protocollagen, Ascoris muscle	Chvapil et al. 1970
PHD	Protocollagen hydroxylase (PCH)	*Ascaris lumbricoides*	27.27	^{14}C-proline-labeled protocollagen, Ascoris cuticle(fluid)	Chvapil et al. 1970
PHD	Protocollagen hydroxylase (PCH)	*Ascaris lumbricoides*	230.7	^{14}C-proline-labeled protocollagen, Rot skin	Chvapil et al. 1970
PHD	Protocollagenproline hydroxylase(PPH)	*Ascaris lumbricoides*	16.13	adult *A. lumbricoides* muscle	Cain and Fairbairn 1971
PHD	Protocollagenproline hydroxylase(PPH)	*Ascaris lumbricoides*	25.42	adult *A. lumbricoides* eggs	Cain and Fairbairn 1971
PHD	human type I collagen prolyl-4-hydroxylase (CP4H-I)	*Homo sapiens*	40	(Pro- Pro-Gly)10	Hirsilä et al. 2003
PHD	HIF prolyl 4-hydroxylases isoenzyme-1(HIF- P4H-1)	*Homo sapiens*	230	(Pro- Pro-Gly)10	Hirsilä et al. 2003
PHD	HIF prolyl 4-hydroxylases isoenzyme-2 (HIF-P4H-2)	*Homo sapiens*	250	(Pro- Pro-Gly)10	Hirsilä et al. 2003
PHD	HIF prolyl 4-hydroxylases isoenzyme-3 (HIF-P4H-3)	*Homo sapiens*	230	(Pro- Pro-Gly)10	Hirsilä et al. 2003
PHD	HIF prolyl 4-hydroxylases isoenzyme-1 (HIF-P4H-1)	*Homo sapiens*	200	N- terminal biotinylated 19'mer peptide(HIF-1αC19$^{556–574}$)	Tuckerman et al. 2004
PHD	HIF prolyl 4-hydroxylases isoenzyme-2 (HIF-P4H-2)	*Homo sapiens*	200	N- terminal biotinylated 19'mer peptide(HIF-1αC19$^{556–574}$)	Tuckerman et al. 2004
PHD	HIF prolyl 4-hydroxylases isoenzyme-3 (HIF-P4H-3)	*Homo sapiens*	200	N- terminal biotinylated 19'mer peptide(HIF-1αC19$^{556–574}$)	Tuckerman et al. 2004
PHD	HIF prolyl 4-hydroxylases isoenzyme-2 (HIF-P4H-2)	*Homo sapiens*	100±20	HIF-1aODDD	Koivunen et al. 2006
PHD	HIF prolyl 4-hydroxylases isoenzyme-2 (HIF-P4H-2)	*Homo sapiens*	53±13	HIF-1a(556–574) peptide	Berchner-Pfannschmidt et al. 2008
PHD	HIF prolyl 4-hydroxylases isoenzyme-2 (HIF-P4H-2)	*Homo sapiens*	54±10	HIF-1a(556–574) peptide	Berchner-Pfannschmidt et al. 2008
PHD	PHD2$^{181\text{-}426}$	*Homo sapiens*	229±60	HIF-1α$^{556–574}$ CODD	Ehrismann et al. 2007
PHD	PHD2$^{1\text{-}426}$	*Homo sapiens*	250	HIF-1α$^{556–574}$ CODD	Ehrismann et al. 2007
PHD	His$_6$ –PHD2$^{181–426}$	*Homo sapiens*	76±11	His$_6$ –HIF-1α$^{530–652}$ CODD	Ehrismann et al. 2007
PHD	His$_6$ –PHD2$^{181–426}$	*Homo sapiens*	81±28	His$_6$ –HIF-1α$^{530–698}$ CODD	Ehrismann et al. 2007
PHD	His$_6$ –PHD2$^{181–426}$	*Homo sapiens*	85±17	His$_6$ –HIF-1α$^{344–503}$ NODD	Ehrismann et al. 2007
PHD	His$_6$ –PHD2$^{181–426}$	*Homo sapiens*	67±10	His$_6$ –HIF-1α$^{502–697}$ CODD	Ehrismann et al. 2007
PHD	HIF prolyl 4-hydroxylases isoenzyme-2(HIF-P4H-2)	*Homo sapiens*	250	a 19-residue HIF1a peptide	Myllyharju 2008
PHD	HIF prolyl 4-hydroxylases isoenzyme-2(HIF-P4H-2)	*Homo sapiens*	100	a 248-residue HIF1a ODDD fragment	Myllyharju 2008
PHD	HIF prolyl 4-hydroxylases isoenzyme-2(HIF-P4H-2)	*Homo sapiens*	65-85	123-195-residue HIF1a and HIF2a ODDD fragments	Myllyharju 2008
PHD	PHD2$^{181\text{-}426}$	*Homo sapiens*	>450	HIF-1α CODD 19-mer peptide	Tarhonskaya et al. 2014
PHD	Human prolyl-hydroxylase 2 (HsPHD2)	*Homo sapiens*	150±50	HIF-2αODDD	Lorenzo et al. 2014
PHD	Human prolyl-hydroxylase 2 (HsPHD2)	*Homo sapiens*	460±30	HIF-1a C-terminal oxygen-dependent degradation domain(CODD)	Tarhonskaya et al. 2015
PHD	Human prolyl-hydroxylase 2 (HsPHD2)	*Homo sapiens*	>450	HIF-1a N-terminal oxygen-dependent degradation domain(NODD)	Tarhonskaya et al. 2015

Category	Enzyme	Organism	$K_m^{app}(O_2)$ [uM]	Protein substrate	Reference
PHD	Human prolyl-hydroxylase 2 (HsPHD2)	*Homo sapiens*	>450	HIF-2a C-terminal oxygen-dependent degradation domain(CODD)	Tarhonskaya et al. 2015
PHD	Human prolyl-hydroxylase 2 (HsPHD2)	*Homo sapiens*	410±80	HIF-1a N-terminal oxygen-dependent degradation domain(NODD)	Tarhonskaya et al. 2015
PHD	$PHD2^{181\text{-}426}$	*Homo sapiens*	216 ± 30	HIF-1α CODD 19-mer peptide in the presence of a cyclic-14-mer peptide	Kearney 2015
PHD	$PHD2^{181\text{-}426}$	*Homo sapiens*	355±80	HIF-1α CODD 19-mer peptide in the absence of a cyclic-14-mer peptide	Kearney 2015
PHD	HIF prolyl 4-hydroxylases isoenzyme-2(HIF-P4H-2)	*Homo sapiens*	530±90	CODD	Pektas et al. 2015
FIH	factor inhibiting HIF	*Homo sapiens*	90±20	$HIF\text{-}1\alpha^{788\text{–}822}$ CAD(C-terminal transactivation domain)	Koivunen et al. 2004
FIH	factor inhibiting HIF	*Homo sapiens*	145±33	$HIF\text{-}1\alpha^{788\text{–}806}$ CAD(C-terminal transactivation domain)	Ehrismann et al. 2007
FIH	factor inhibiting HIF	*Homo sapiens*	150±30	$HIF\text{-}1\alpha^{788\text{–}822}$ CAD(C-terminal transactivation domain)	Ehrismann et al. 2007
FIH	factor inhibiting HIF	*Homo sapiens*	237±28	His_6 –$HIF\text{-}1\alpha^{653\text{–}826}$ CAD(C-terminal transactivation domain)	Ehrismann et al. 2007
FIH	factor inhibiting HIF	*Homo sapiens*	110±73	$HIF\text{-}2\alpha^{832\text{–}866}$ CAD(C-terminal transactivation domain)	Ehrismann et al. 2007
FIH	factor inhibiting HIF(FIH-1)	*Homo sapiens*	12±3	plasmids coding for thioredoxin (Trx)-6His-tagged mNotch1	Wilkins et al. 2009
FIH	factor inhibiting HIF(FIH-1)	*Homo sapiens*	19±4	plasmids coding for thioredoxin (Trx)-6His-tagged mNotch2	Wilkins et al. 2009
FIH	factor inhibiting HIF(FIH-1)	*Homo sapiens*	70±20	plasmids coding for thioredoxin (Trx)-6His-tagged mNotch2	Wilkins et al. 2009
FIH	factor inhibiting HIF(FIH-1)	*Homo sapiens*	12±4	plasmids coding for thioredoxin (Trx)-6His-tagged mNotch3	Wilkins et al. 2009
FIH	factor inhibiting HIF(FIH-1)	*Homo sapiens*	110±20	plasmids coding for thioredoxin (Trx)-6His-tagged mNotch3	Wilkins et al. 2009
FIH	factor inhibiting HIF(FIH-1)	*Homo sapiens*	90±25	hHIF-1a	Wilkins et al. 2009
FIH	factor inhibiting HIF(FIH-1)	*Homo sapiens*	270±50	$HIF\text{-}1\alpha^{788\text{–}826}$ CTAD(C-terminal transactivation domain)(150uM)	Hangasky et al. 2014
FIH	factor inhibiting HIF	*Homo sapiens*	200±40	$HIF\text{-}1\alpha^{788\text{–}826}$ CTAD(C-terminal transactivation domain)(80uM)	Hangasky et al. 2014
FIH	factor inhibiting HIF	*Homo sapiens*	110±30	HIF-1A C-terminal transactivation Domain(CTAD)	Tarhonskaya et al. 2015
FIH	factor inhibiting HIF	*Homo sapiens*	110±10	HIF-2A C-terminal transactivation Domain(CTAD)	Tarhonskaya et al. 2015
FIH	factor inhibiting HIF	*Homo sapiens*	40±10	$1CA_{20\text{-}mer}$(consensus ankyrin)	Tarhonskaya et al. 2015
FIH	factor inhibiting HIF	*Homo sapiens*	50±10	tnkrs-135-mertnkrs-1(a 135-mer fragment of tankyrase-1)	Tarhonskaya et al. 2015
FIH	factor inhibiting HIF	*Homo sapiens*	56±15	2CA(consensus ankyrin)	Tarhonskaya et al. 2015
FIH	factor inhibiting HIF	*Homo sapiens*	120±20	3CA(consensus ankyrin)	Tarhonskaya et al. 2015

Supplementary_Table_4 Oxygen Km values of oxygen reductases

Category	Enzyme name	Abbrevation	Organism	$K_m^{app}(O_2)$ [uM]	Reference
bd oxidase	cytochrome bd type quinol oxidase	bd	*Azotobacter vinelandii*	18	Hoffman et al. 1979
bd oxidase	cytochrome bd type quinol oxidase	bd	*Azotobacter vinelandii*	0.48±0.16	Juenemann et al. 1995
bd oxidase	cytochrome bd type quinol oxidase	bd	*Azotobacter vinelandii*	20-50	Juenemann et al. 1995
bd oxidase	cytochrome bd type quinol oxidase	bd	*Escherichia coli*	0.4	D'Mello et al. 1994
bd oxidase	cytochrome bd type quinol oxidase	bd	*Escherichia coli*	0.26	Juenemann et al. 1995
bd oxidase	cytochrome bd type quinol oxidase	bd	*Escherichia coli*	0.38	Kita et al. 1984
bd oxidase	cytochrome bd type quinol oxidase	bd	*Klebsiella pneumoniae*	0.02	Smith et al. 1990
bd oxidase	cytochrome bd type quinol oxidase	bd	*Escherichia coli*	2	Kolonay et al. 1994
bd oxidase	cytochrome bd type quinol oxidase	bd	*Axotobacter vinelandii*	5	Juenemann et al. 1995
bd oxidase	cytochrome bd type quinol oxidase	bd	*Axotobacter vinelandii*	5.7	Kolonay et al. 1994
bd oxidase	cytochrome bd type quinol oxidase	bd	*Axotobacter vinelandii* UW136 Cyd$^+$	0.052±0.0016	D'Mello et al. 1994
bd oxidase	cytochrome bd type quinol oxidase	bd	*Axotobacter vinelandii* UW136 Cyd$^+$	0.013±0.001	D'Mello et al. 1994
bd oxidase	cytochrome bd type quinol oxidase	bd	*Axotobacter vinelandii* UW136 Cyd$^+$	0.25±0.066	D'Mello et al. 1994
bd oxidase	cytochrome bd type quinol oxidase	bd	*Axotobacter vinelandii* UW136 Cyd$^+$	0.27±0.16	D'Mello et al. 1994
bd oxidase	cytochrome bd type quinol oxidase	bd	*Axotobacter vinelandii* UW136 Cyd$^+$	2.96±0.96	D'Mello et al. 1994
bd oxidase	cytochrome bd type quinol oxidase	bd	*Axotobacter vinelandii* UW136 Cyd$^+$	4.93±2.05	D'Mello et al. 1994
bd oxidase	cytochrome bd type quinol oxidase	bd	*Escherichia coli*	1.75 ±0.13	Juenemann et al. 1995
bd oxidase	cytochrome bd type quinol oxidase	bd	*Azotobacter vinelandii*	4.10 ±0.34	Juenemann et al. 1995
bd oxidase	cytochrome bd type quinol oxidase	bd	*Axotobacter vinelandii* UW136 Cyd$^+$	0.0034±0.0014	D'Mello et al. 1996
bd oxidase	cytochrome bd type quinol oxidase	bd	*Synechocystis sp. strain* PCC 6803	0.35	Pils and Schmetterer 2001
bd oxidase	cytochrome bd (CydAB) type quinol oxidase	bd	*Escherichia coli*	0.024	Rice and Hempfling 1978
bd oxidase	cytochrome bd (CydAB) type quinol oxidase	bd	*Campylobacter jejuni*	0.041±0.0039	Jackson et al. 2007
bd oxidase	cytochrome bd (CydAB) type quinol oxidase	bd	*Campylobacter jejuni*	0.040	Jackson et al. 2007
bd oxidase	cytochrome bd (CydAB) type quinol oxidase	bd	*Campylobacter jejuni*	0.81±0.015	Jackson et al. 2007
bd oxidase	cytochrome bd (CydAB) type quinol oxidase	bd	*Campylobacter jejuni*	0.85±0.056	Jackson et al. 2007
bd oxidase	cytochrome bd (CydAB) type quinol oxidase	bd	*Gluconobacter oxydans*	2.03±0.15	Miura et al. 2013
bd oxidase	cyanide-insensitive terminal quinol oxidase (CIO)	CIO	*Gluconobacter oxydans*	20.8±3.6	Miura et al. 2013
bd oxidase	cyanide-insensitive terminal quinol oxidase (CIO)	CIO	*Pseudomonas aeruginosa*	4.0±2.1	Arai et al. 2014
bd oxidase	cyanide-insensitive terminal quinol oxidase (CIO)	CIO	*Pseudomonas aeruginosa*	0.41±0.1	Arai et al. 2014
HCO oxidase	cytochrome c oxidase(COX)	C	*Campylobacter jejuni*	0.10	Gupta et al. 2009
HCO oxidase	cytochrome c oxidase(COX)	C	Hamster	0.8	Wilson et al. 1977

Category	Enzyme name	Abbrevation	Organism	$K_m^{app}(O_2)$ [uM]	Reference
HCO oxidase	Pseudomonas cytochrome c-551	C	*Pseudomonas aeruginosa*	0.028	Yamanaka et al. 1961
HCO oxidase	Pseudomonas cytochrome c-551	C	*Vigna unguiculata*	0.1	Gupta et al. 2009
HCO oxidase	cytochrome c oxidase(COX)	C	*Homo sapiens*	0.1-10	Zimorski et al. 2019
HCO oxidase	cytochrome c oxidase(COX)	C	Rat	0.04	Degn and Wohlrab 1971
HCO oxidase	cytochrome c oxidase(COX)	C	Rat	0.05-0.5	Vanderkooi et al. 1991
HCO oxidase	cytochrome c oxidase(COX)	C	Beef	0.95	Petersen et al. 1976
HCO oxidase	cytochrome c oxidase(COX)	C	Soybean	0.147	Gupta et al. 2009
HCO oxidase	cytochrome c oxidase(COX)	C	None	1	Massari et al. 1996
HCO oxidase	cytochrome cao oxidase	C	Thermophilic bacterium PS3	5.5	Sone et al. 1990
HCO oxidase	cytochrome c oxidase(COX)	C	None	0.5	Cooper and Brown 2008
HCO oxidase	cytochrome c oxidase(COX)	C	None	30	Cooper and Brown 2008
HCO oxidase	cytochrome c oxidase(COX)	C	a-proteobacteria	0.08	Degli Esposti et al. 2019
HCO oxidase	cytochrome c oxidase(COX)	C	pigeon	<0.27	Vanderkooi et al. 1991
HCO oxidase	cytochrome c oxidase(COX)	C	Rat	0.5	Vanderkooi et al. 1991
HCO oxidase	cytochrome c oxidase(COX)	C	Synechocystis sp. strain PCC 6803	1	Pils and Schmetterer 2001
HCO oxidase	cbb3-type oxidase isoenzyme 1	C	*Pseudomonas aeruginosa*	0.25±0.02	Arai et al. 2014
HCO oxidase	cbb3-type oxidase isoenzyme 1	C	*Pseudomonas aeruginosa*	0.044±0.022	Arai et al. 2014
HCO oxidase	cbb3-type oxidase isoenzyme 1	C	*Pseudomonas aeruginosa*	0.0066	Arai et al. 2014
HCO oxidase	cbb3-type oxidase isoenzyme 2	C	*Pseudomonas aeruginosa*	0.23±0.08	Arai et al. 2014
HCO oxidase	cbb3-type oxidase isoenzyme 2	C	*Pseudomonas aeruginosa*	0.032±0.021	Arai et al. 2014
HCO oxidase	cbb3-type oxidase isoenzyme 2	C	*Pseudomonas aeruginosa*	0.0065	Arai et al. 2014
HCO oxidase	cytochrome cbb3 oxidases	C	*Bradyrhizobium japonicum*	0.0557±0.0242	Preisig et al. 1996
HCO oxidase	cytochrome cbb3 oxidases	C	*Bradyrhizobium japonicum*	0.0193±0.007	Preisig et al. 1996
HCO oxidase	cytochrome cbb3 oxidases	C	*Bradyrhizobium japonicum*	0.004±0.0021	Preisig et al. 1996
HCO oxidase	cytochrome bo3(Cyo)	A1-bo3	*Gluconobacter oxydans*	3.07±0.51	Miura et al. 2013
HCO oxidase	cytochrome bo3(Cyo)	A1-bo3	*Gluconobacter oxydans*	2.80±0.31	Miura et al. 2013
HCO oxidase	cytochrome bo3(Cyo)	A1-bo3	*Pseudomonas aeruginosa*	3.2±1.3	Arai et al. 2014
HCO oxidase	cytochrome bo3(Cyo)	A1-bo3	*Pseudomonas aeruginosa*	0.25±0.04	Arai et al. 2014
HCO oxidase	cytochrome bo'($b_{562}o$, bo_3, and oo_3) oxidase(CyoABCD)	A1-bo3	*Escherichia coli* AN2342	0.16±0.51	D'Mello et al. 1995
HCO oxidase	cytochrome bo'($b_{562}o$, bo_3, and oo_3) oxidase(CyoABCD)	A1-bo3	*Escherichia coli* AN2342	0.15±0.055	D'Mello et al. 1995
HCO oxidase	cytochrome bo'($b_{562}o$, bo_3, and oo_3) oxidase(CyoABCD)	A1-bo3	*Escherichia coli* AN2342	0.085±0.0029	D'Mello et al. 1995

Category	Enzyme name	Abbrevation	Organism	$K_m^{app}(O_2)$ [uM]	Reference
HCO oxidase	cytochrome bo'(b_{562}o, bo_3, and oo_3) oxidase(CyoABCD)	A1-bo3	*Escherichia coli* AN2342	0.016±0.0036	D'Mello et al. 1995
HCO oxidase	cytochrome o type quinol oxidase (cyoABCDE)	A1-bo3	*Escherichia coli*	0.2	Rice and Hempfling 1978
HCO oxidase	cytochrome o type quinol oxidase (cyoABCDE)	A1-bo3	*Escherichia coli*	2.9	Kita et al. 1984
HCO oxidase	cytochromeo (558) oxidase	A1-bo3	*Acetobacter suboxydans* ATCC 621	33	Daniel 1970
HCO oxidase	cytochromeo (565) oxidase	A1-bo3	*Acetobacter suboxydans* ATCC 621	2.9	Daniel 1970
HCO oxidase	cytochrome o oxidase	A1-bo3	*Acetobacter aceti*	16.9	Matsushita et al. 1992
HCO oxidase	cytochrome o type quinol oxidase (cyoABCDE)	A1-bo3	Thermophilic bacterium PS3	0.09	Sone and Fujiwara 1991
HCO oxidase	aa3-type cytochrome c oxidases (aa3)	A1	*Bacillus cereus*	5	Garcia-Horsman et al. 1991
HCO oxidase	aa3-type cytochrome c oxidases (aa3)	A1	*Aeropyrum pernix* K1	32	Ishikawa et al. 2002
HCO oxidase	cytochrome a1 oxidase	A1	*Acetobacter aceti*	4.4	Matsushita et al. 1992
HCO oxidase	A1 type cytochrome c oxidases (aa3)	A1	*Pseudomonas aeruginosa*	4.3±1.0	Arai et al. 2014
HCO oxidase	A1 type cytochrome c oxidases	A1	a-proteobacteria	0.2-10	Degli Esposti et al. 2019
HCO oxidase	cytochrome caa3 oxidase	A2	Thermophilic bacterium PS3	5.7	Sone et al. 1990
HCO oxidase	cytochrome caa3 oxidase	A2	*Bacillus cereus*	2	Garcia-Horsman et al. 1991
HCO oxidase	A2 type cytochrome c oxidases	A2	a-proteobacteria	0.8	Degli Esposti et al. 2019
HCO oxidase	A1 type cytochrome c oxidases (aa3)	A1	None	2	Kita et al. 1984
HCO oxidase	AOX type quinonol:O_2 oxidoreductase	AOX	Melon	30	Gupta et al. 2009
HCO oxidase	AOX type quinonol:O_2 oxidoreductase	AOX	*Glycine max* L. cv. Ransom	13.40	Ribas-Carbo et al. 1994
HCO oxidase	AOX type quinonol:O_2 oxidoreductase	AOX	*Glycine max* L. cv. Ransom	21.80	Ribas-Carbo et al. 1994
HCO oxidase	AOX type quinonol:O_2 oxidoreductase	AOX	*Glycine max* L. cv. Ransom	12.3	Ribas-Carbo et al. 1994
HCO oxidase	AOX type quinonol:O_2 oxidoreductase	AOX	*Glycine max* L. cv. Ransom	18.2	Ribas-Carbo et al. 1994
HCO oxidase	AOX type quinonol:O_2 oxidoreductase	AOX	*Vigna radiata* L.	9.9	Ribas-Carbo et al. 1994
HCO oxidase	AOX type quinonol:O_2 oxidoreductase	AOX	*Vigna radiata* L.	25.2	Ribas-Carbo et al. 1994

References

Barth S, Edlich F, Berchner-Pfannschmidt U, et al (2009) Hypoxia-inducible factor prolyl-4-hydroxylase PHD2 protein abundance depends on integral membrane anchoring of FKBP38. Journal of Biological Chemistry 284:23046–23058.

Bruick RK, McKnight SL (2001) A conserved family of prolyl-4-hydroxylases that modify HIF. Science 294:1337–1340.

Epstein ACR, Gleadle JM, McNeill LA, et al (2001) C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. Cell 107:43–54.

Erwin DH, Laflamme M, Tweedt SM, et al (2011) The Cambrian conundrum: Early divergence and later ecological success in the early history of animals. Science 334:1091.

Graham AM, Barreto FS (2019) Loss of the HIF pathway in a widely distributed intertidal crustacean, the copepod Tigriopus californicus. Proceedings of the National Academy of Sciences of the United States of America 16:12913–12918.

Graham AM, Barreto FS (2020) Independent Losses of the Hypoxia-Inducible Factor (HIF) Pathway within Crustacea. Molecular Biology and Evolution 37:1342–1349.

Graham AM, McCracken KG (2019) Convergent evolution on the hypoxia-inducible factor (HIF) pathway genes EGLN1 and EPAS1 in high-altitude ducks. Heredity 122:819–832.

Hammarlund EU (2020) Harnessing hypoxia as an evolutionary driver of complex multicellularity. Interface Focus 10:20190101.

Hammarlund EU, Flashman E, Mohlin S, Licausi F (2020) Oxygen-sensing mechanisms across eukaryotic kingdoms and their roles in complex multicellularity. Science 370:421.

Hammarlund EU, Von Stedingk K, Påhlman S (2018) Refined control of cell stemness allowed animal evolution in the oxic realm. Nature Ecology and Evolution 2:220–228.

Ivan M, Haberberger T, Gervasi DC, et al (2002) Biochemical purification and pharmacological inhibition of a mammalian prolyl hydroxylase acting on hypoxia-inducible factor. Proceedings of the National Academy of Sciences of the United States of America 99:13459–13464.

Jaakkola PM, Rantanen K (2013) The regulation, localization, and functions of oxygen-sensing prolyl hydroxylase PHD3. Biological Chemistry 394:449–457.

Kaelin WG, Ratcliffe PJ (2008) Oxygen Sensing by Metazoans: The Central Role of the HIF Hydroxylase Pathway. Molecular Cell 30:393–402.

Koivunen P, Hirsilä M, Günzler V, et al (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.

Laitala A, Aro E, Walkinshaw G, et al (2012) Transmembrane prolyl 4-hydroxylase is a fourth prolyl 4-hydroxylase regulating EPO production and erythropoiesis. Blood 120:3336–3344.

Lando D, Peet DJ, Gorman JJ, et al (2002) FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. Genes and Development 16:1466–1471.

Lee CYS, Stewart E V., Hughes BT, Espenshade PJ (2009) Oxygen-dependent binding of Nro1 to the prolyl hydroxylase Ofd1 regulates SREBP degradation in yeast. EMBO Journal 28:135–143.

Lippl K, Boleininger A, McDonough M, et al (2018) Born to sense: biophysical analyses of the oxygen sensing prolyl hydroxylase from the simplest animal Trichoplax adhaerens. Hypoxia 6:57–71.

Loenarz C, Coleman ML, Boleininger A, et al (2011) The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, Trichoplax adhaerens. EMBO Reports 12:63–70.

Lyons TW, Reinhard CT, Planavsky NJ (2014) The rise of oxygen in Earth's early ocean and atmosphere. Nature 506:307–315.

Maxwell PH, Wlesener MS, Chang GW, et al (1999) The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. Nature 399:271–275.

Metzen E, Berchner-Pfannschmidt U, Stengel P, et al (2003) Intracellular localisation of human HIF-1α hydroxylases: Implications for oxygen sensing. Journal of Cell Science 116:1319–1326.

Mills DB, Francis WR, Vargas S, et al (2018) The last common ancestor of animals lacked the HIF pathway and respired in low-oxygen environments. eLife 7:e31176.

Mills DB, Ward LM, Jones CA, et al (2014) Oxygen requirements of the earliest animals. Proceedings of the National Academy of Sciences of the United States of America 111:4168–4172.

Place TL, Domann FE (2013) Prolyl-hydroxylase 3: evolving roles for an ancient signaling protein. Hypoxia 1:13–27.

Rytkönen KT, Storz JF (2011) Evolutionary origins of oxygen sensing in animals. EMBO Reports 12:3–4.

Schofield CJ, Ratcliffe PJ (2004) Oxygen sensing by HIF hydroxylases. Nature Reviews Molecular Cell Biology 5:343–354.

Semenza GL, Jiang BH, Leung SW, et al (1996) Hypoxia response elements in the aldolase A, enolase 1, and lactate dehydrogenase A gene promoters contain essential binding sites for hypoxia-inducible factor 1. Journal of Biological Chemistry 271:32529–32537.

Semenza GL, Wang GL (1992) A nuclear factor induced by hypoxia via de novo protein synthesis binds to the human erythropoietin gene enhancer at a site required for transcriptional activation. Molecular and Cellular Biology 12:5447–5454.

Tarade D, Lee JE, Ohh M (2019) Evolution of metazoan oxygen-sensing involved a conserved divergence of VHL affinity for HIF1α and HIF2α. Nature Communications 10:1–12.

Tarhonskaya H, Hardy AP, Howe EA, et al (2015) Kinetic investigations of the role of factor inhibiting hypoxia-inducible factor (FIH) as an oxygen sensor. Journal of Biological Chemistry 290:19726–19742.

Tuckerman JR, Zhao Y, Hewitson KS, et al (2004) Determination and comparison of specific activity of the HIF-prolyl hydroxylases. FEBS Letters 576:145–150.

Wang G., Semenza G (1993a) Characterization of hypoxia-inducible factor 1 and regulation of DNA binding activity by hypoxia. Journal of Biological Chemistry 268:21513–21518.

Wang GL, Jiang BH, Rue EA, Semenza GL (1995) Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. Proceedings of the National Academy of Sciences of the United States of America 92:5510–5514.

Wang GL, Semenza GL (1993b) General involvement of hypoxia-inducible factor 1 in transcriptional response to hypoxia. Proceedings of the National Academy of Sciences of the United States of America 90:4304–4308.

Wang GL, Semenza GL (1995) Purification and characterization of hypoxia-inducible factor. Journal of Biological Chemistry 270:1230–1237.

West CM, van der Wel H, Wang ZA (2007) Prolyl 4-hydroxylase-1 mediates O2 signaling during development of Dictyostelium. Development 134:3349–3358.

Zhang P, Yao Q, Lu L, et al (2014) Hypoxia-inducible factor 3 is an oxygen-dependent transcription activator and regulates a distinct transcriptional response to hypoxia. Cell Reports 6:1110–1121.

Allison SD (2005) Cheaters, diffusion and nutrients constrain decomposition by microbial enzymes in spatially structured environments. Ecology Letters 8:626–635.

Berghem LER, Pettersson LG, Axiö-Fredriksson U-B (1976) The mechanism of enzymatic cellulose degradation. European Journal of Biochemistry 61:621–630

Berlemont R (2017) Distribution and diversity of enzymes for polysaccharide degradation in fungi. Scientific Reports 7:1–11.

Berlemont R, Martiny AC (2016) Glycoside hydrolases across environmental microbial communities. PLoS Computational Biology 12:e1005300.

Berlemont R, Martiny AC (2015) Genomic potential for polysaccharide deconstruction in bacteria. Applied and Environmental Microbiology 81:1513–1519.

Choi J, Bach E, Lee J, et al (2018) Spatial structuring of cellulase gene abundance and activity in soil. Frontiers in Environmental Science 6:107.

Enari TM, Niku-Paavola ML (1987) Enzymatic hydrolysis of cellulose: is the current theory of the mechanisms of hydrolysis valid? Critical Reviews in Biotechnology 5:67–87.

Goldfarb KC, Karaoz U, Hanson CA, et al (2011) Differential growth responses of soil bacterial taxa to carbon substrates of varying chemical recalcitrance. Frontiers in Microbiology 2:94.

Guo W, Mishra S, Zhao J, et al (2018) Metagenomic study suggests that the gut microbiota of the giant panda (Ailuropoda melanoleuca) may not be specialized for fiber fermentation. Frontiers in Microbiology 9:229.

Halliwell G, Griffin M (1973) The nature and mode of action of the cellulolytic component C1 of Trichoderma koningii on native cellulose. Biochemical Journal 135:587–594.

Haworth WN (1929) The constitution of sugars. Journal of Chemical Education 7:545.

Henrissat B, Driguez H, Viet C, Schülein M (1985) Synergism of cellulases from trichoderma reesei in the degradation of cellulose. Bio/Technology 3:722–726.

Hon DNS (1994) Cellulose: a random walk along its historical path. Cellulose 1:1–25.

Knowles J, Lehtovaara P, Teeri T (1987) Cellulase families and their genes. Trends in Biotechnology 5:255–261.

Mai G, Liu Q, Zhou Y, et al (2019) Metagenomic insights into a cellulose-rich niche reveal microbial cooperation in cellulose degradation. Frontiers in microbiology 10:618.

Medie FM, Davies GJ, Drancourt M, Henrissat B (2012) Genome analyses highlight the different biological roles of cellulases. Nature Reviews Microbiology 10:227.

Niku-Paavola ML, Lappalainen A, Enari TM, Nummi M (1985) A new appraisal of the endoglucanases of the fungus Trichoderma reesei. Biochemical Journal 231:75–81.

O'sullivan AC (1997) Cellulose: The structure slowly unravels. Cellulose 4:173–207.

Preston RD (1975) X-ray analysis and the structure of the components of plant cell walls. Physics Reports 21:183–226.

Romaní AM, Fischer H, Mille-Lindblom C, Tranvik LJ (2006) Interactions of bacteria and fungi on decomposing litter: Differential extracellular enzyme activities. Ecology 87:2559–2569.

Russell JB, Muck RE, Weimer PJ (2009) Quantitative analysis of cellulose degradation and growth of cellulolytic bacteria in the rumen. FEMS Microbiology Ecology 67:183–197.

Schloss PD, Hay AG, Wilson DB, et al (2005) Quantifying bacterial population dynamics in compost using 16S rRNA gene probes. Applied Microbiology and Biotechnology 66:457–463.

Stewart RD, Auffret MD, Warr A, et al (2018) Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. Nature Communications 9:870.

Székely AJ, Sipos R, Berta B, et al (2009) DGGE and T-RFLP analysis of bacterial succession during mushroom compost production and sequence-aided T-RFLP profile of mature compost. Microbial Ecology 57:522–533.

Wilson DB (2011) Microbial diversity of cellulose hydrolysis. Current Opinion in Microbiology 14:1–5.

Wood TM (1969) The cellulase of Fusarium solani. Resolution of the enzyme complex. The Biochemical journal 115:457–464.

Acknowledgements

Firstly, I want to thank Prof. Martin for giving me this precious chance to study abroad, which will also be one of the essential experiences in my life.

I want to thank Prof. Sven for being my second supervisor and always sharing his ideas and interesting papers in the group email.

I want to thank Verena and Renate, who helped me a lot with many detailed but essential things, e.g., registration, residence permit, and graduation.

A special thanks to Nicole, Fernando, Madeline, and Josip, who helped me with my cellulase project.

I thank Nils and Luca sincerely for collaborating with me on the HIF project during 2020.07-2021.05.

Thanks to Luca, Fernando, Madeline, Andrey, and Nils for reading and commenting on my thesis.

I want to thank all other colleagues in the institute for their help every time.

I thank the China Scholarship Council (CSC) for providing the scholarship.

Finally, I want to thank my boyfriend Yiyuan, my friends (Guohua Meng, Yaxi, Xu, Yichu Zhou, Wei Liu, Lin Wang), and my family (mother, father, and brother), who are the reason I live and insist.