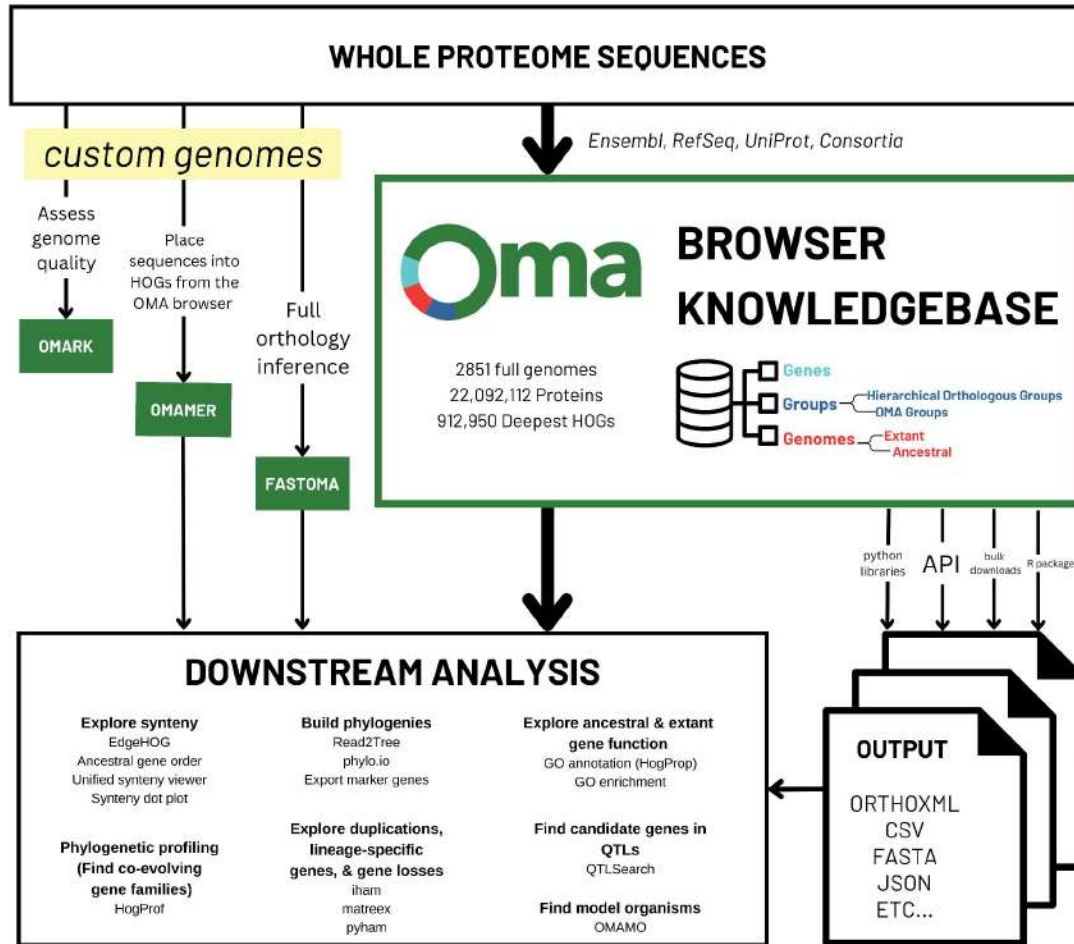


# The OMA Ecosystem

From public and private genomes, orthology inference, a database of over 2800 species, and a variety of downstream analyses for comparative genomics



# The OMA browser

The screenshot shows the OMA browser website. At the top left is the SIB logo and the OMA browser logo. At the top right are navigation links: Explore, Tools, Download, Help, and About. The main header features the OMA logo (a green circle with a red and blue segment) and the text "orthologous matrix browser". Below this is a search bar containing the text: "Blue-light photoreceptor" | proteinid:P53\_RAT | species:"Drosophila melanogaster". Below the search bar is an example text: "Examples: Entry P53\_RAT - 'EWGKQSF' in Tetraodon - Search for 'Blue-light photoreceptor' - 'Drosophila melanogaster' species". Below the example text is a button that says "SCROLL TO DISCOVER MORE". At the bottom of the page, there is a green banner with the following statistics: 2,851 Full genomes, 22,092,112 Proteins, 1,251,567 OMA groups, 912,950 Deepest HOGs, and All.Jul2023 Release.

SIB **Oma** browser

Explore ▾ Tools ▾ Download ▾ Help ▾ About ▾

**Oma** | orthologous matrix  
browser

? "Blue-light photoreceptor" | proteinid:P53\_RAT | species:"Drosophila melanogaster" 🔍

**Examples:** Entry P53\_RAT - 'EWGKQSF' in Tetraodon - Search for "Blue-light photoreceptor" - "Drosophila melanogaster" species

SCROLL TO DISCOVER MORE

2,851 Full genomes   22,092,112 Proteins   1,251,567 OMA groups   912,950 Deepest HOGs   All.Jul2023 Release

<https://omabrowser.org/>

<https://oma-stage.vital-it.ch/>

(Please use this one for the exercises)

# The OMA Database: 3 kinds of pages

## genes

### Gene LEIIN03631 (A4I3J2)

Leishmania infantum is on Chromosome 28.

Groups

Genome

## groups

### HOG:0495624.7c with 2 members (DYW\_deaminase domain-containing protein)

[Eukaryota](#) / [Kinetoplastida](#) / [Trypanosomatidae](#) / [Leishmaniinae](#) / [Leishmania](#) / [Leishmania donovani species complex](#)

### OMA GROUP 402875 with 9 members. (Uncharacterized protein)

Fingerprint: MRFARFE

## genomes

### LEIIN - Leishmania infantum

Proteome version: 03-JAN-2001 (Rel. 66, Last updated, Version 1) with 8031 proteins.

### Ancestral genome of Leishmaniinae

with 6 descendant species and 7893 ancestral genes (HOGs).

# Hierarchical Orthologous Groups (HOGs)

**HOG:D0606964 with 42 members** (Ig-like domain-containing protein)

Completeness score: 0.75 ⓘ

Ancestral Genome

Primates / Lower Level ▾

Hierarchical group HOG:0606964 open at level of **Primates**

OPTIONS ▾

## Graphical viewer

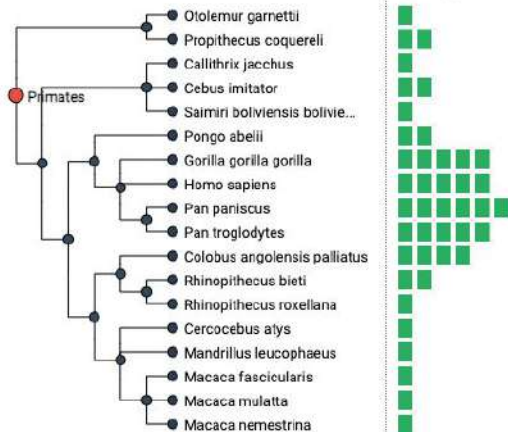
Members

Ancestral GO

Alignment

Ancestral synteny

Similar HOGs >



- A HOG is a gene family
- A collection of orthologs and paralogs which descended from a common ancestral gene

# HOGs

## HOG:D0606964 with 42 members (Ig-like domain-containing protein)

Completeness score: 0.75 ⓘ

Ancestral Genome

Primates / Lower Level ▾

[OrthoXML](#) / [Phyloxml species tree \(Primates\)](#) / [Sequences \(fasta\)](#)

Search



Graphical viewer

Members

Ancestral GO

Alignment

Ancestral synteny

Similar HOGs >

Protein ID	Cross reference	Domain Architectures	Taxon
CERAT42572	★ <a href="#">A0A2K5KVK2</a>		Cercocebus atys
MACFA00770	📄 <a href="#">ENSMFAG00000035597.1</a>		Macaca fascicularis
MACMU01493	★ <a href="#">A0A5F8APY1</a>		Macaca mulatta
MACNE44496	★ <a href="#">A0A2K6B118</a>		Macaca nemestrina
MANLE27898	★ <a href="#">A0A2K6A627</a>		Mandrillus leucophaeus
COLAP38015	★ <a href="#">A0A2K5HL19</a>		Colobus angolensis palliatus
COLAP38016	★ <a href="#">A0A2K5INJ2</a>		Colobus angolensis palliatus
COLAP38023	★ <a href="#">A0A2K5HLG0</a>		Colobus angolensis palliatus
COLAP38026	★ <a href="#">A0A2K5JAM3</a>		Colobus angolensis palliatus
RHIBE28045	★ <a href="#">A0A2K6M7G8</a>		Rhinopithecus bieti

OMA ID

# Ancestral genomes

The collection of HOGs at a given taxonomic level

## Ancestral genome of Primates

with 24 descendant species and 38457 ancestral genes (HOGs).

Remove HOGs with completeness score below

Search



Genome information

Ancestral genes

PHYLOGENETIC FILTER:

Choose a target genome

Ancestor genome

Descendant genome

Ancestral Gene Order

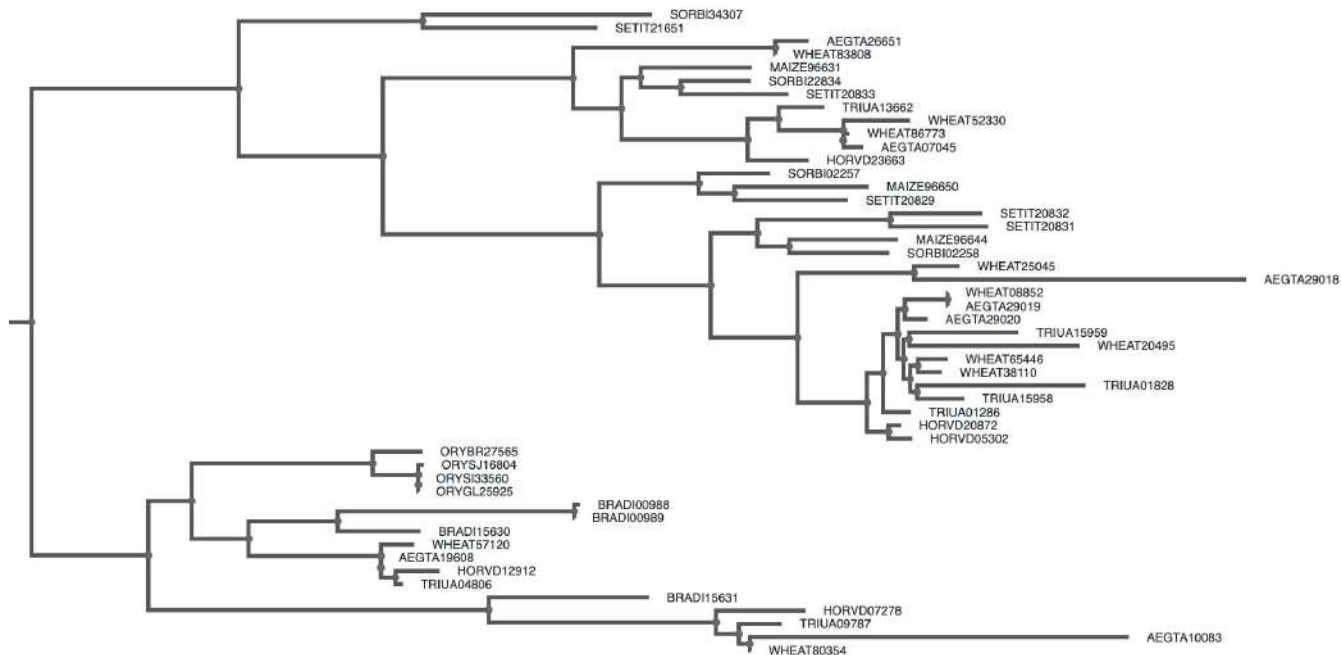
HOG ID	Root HOG ID	Completeness	Nr genes in HOG	Description
<a href="#">HOG:D0912633.1a</a>	<a href="#">HOG:D0912633</a>	1.00	24	autophagy related 16 like
<a href="#">HOG:D0912535.3b.8a.7b.4b</a>	<a href="#">HOG:D0912535</a>	1.00	24	leukocyte cell derived chemotaxin
<a href="#">HOG:D0911480.5b.3b</a>	<a href="#">HOG:D0911480</a>	1.00	24	prostaglandin G/H synthase
<a href="#">HOG:D0911480.5a.2b</a>	<a href="#">HOG:D0911480</a>	1.00	24	prostaglandin G/H synthase
<a href="#">HOG:D0911074.2b.12b</a>	<a href="#">HOG:D0911074</a>	1.00	25	thioredoxin domain containing
<a href="#">HOG:D0911067.13d.9b</a>	<a href="#">HOG:D0911067</a>	1.00	24	paraoxonase
<a href="#">HOG:D0909668</a>	<a href="#">HOG:D0909668</a>	1.00	25	alkB homolog
<a href="#">HOG:D0909574.1b.7a.4b</a>	<a href="#">HOG:D0909574</a>	1.00	24	kinase regulatory subunit
<a href="#">HOG:D0909570.1a.6g</a>	<a href="#">HOG:D0909570</a>	1.00	24	rna helicase
<a href="#">HOG:D0909570.1a.6d.20a.23a.11b</a>	<a href="#">HOG:D0909570</a>	1.00	24	rna helicase
<a href="#">HOG:D0908691.1b.1b.2a.1b</a>	<a href="#">HOG:D0908691</a>	1.00	24	5'-nucleotidase

# OMA's Downstream analyses



# Make gene trees\*

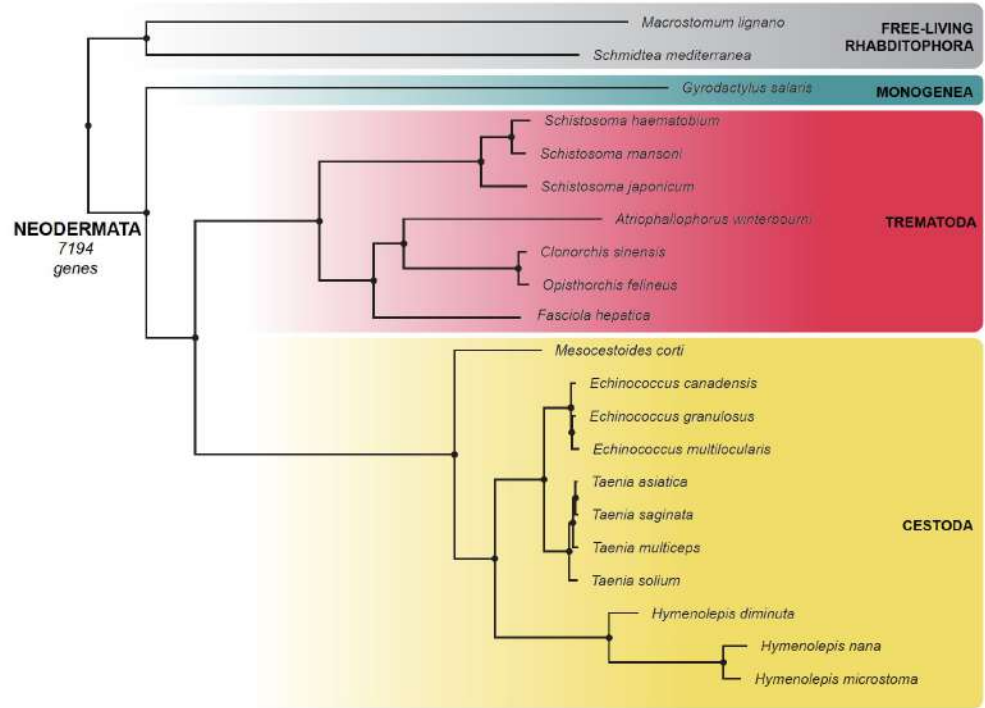
Create **gene** trees for a clade of interest using the genes in the **HOG** at that taxonomic level



\*Uses external software outside of the OMA Ecosystem

# Making species trees\*

Create **species** trees for a clade of interest using the genes in the **OMA Group (strict Orthologous Group)** at that taxonomic level




# Infer gene function and perform GO enrichment analysis

Gene Ontology Function Projection

Email   
We will send an email to this address once the predictions are ready.

Name of Dataset

Sequence File (fasta format)\*  No file chosen

Captcha\*  I'm not a robot 

Transfer Gene Ontology annotations from genes in the same HOG from well-studied model species to other species

## GO ENRICHMENT

You can download the detailed result for this GO enrichment analysis by clicking [here](#)

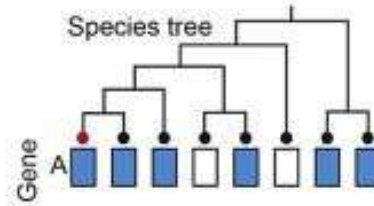
GO_ID	GO_Name	study_count	pop_count	Study_Entries	study_n	p_uncorrected	p_bonferroni	p_1
GO:0033041	sweet taste receptor activity	6	32	DROME01345, DROME03317, DROME04384, DROME04694, DROME05340, DROME05633	6	1.428750553e-16	0	0
GO:0001582	detection of chemical stimulus involved in sensory perception of sweet taste	6	34	DROME01345, DROME03317, DROME04384, DROME04694, DROME05340, DROME05633	6	2.120447249e-16	0	0

Perform a GO enrichment analysis to find overrepresented functions in a set of extant or ancestral genes

# Phylogenetic profiling

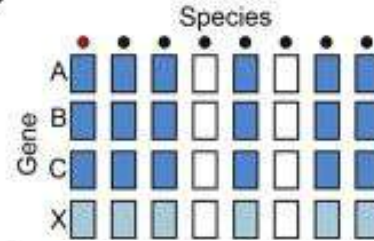
Phylogenetic profiles for human genes and orthogroups

- Ortholog present
- Ortholog absent
- Homo sapiens*



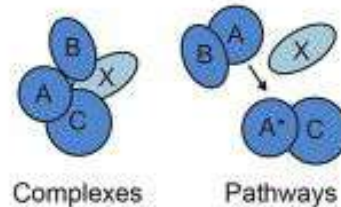
Genome-wide unbiased identification of co-evolving modules

- Uncharacterized
- Homo sapiens*



Prediction of human gene function

- Uncharacterized



# Phylogenetic profiling

**HOG:D0679560 with 323 members** (intraflagellar transport protein 57 homolog)

Completeness score: 0.44 ⓘ

Ancestral Genome

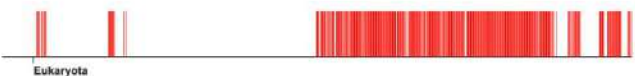
Eukaryota / Lower Level ▾

This HOG has 49 similar HOGs:

## Co-evolving HOGs

Search



HOG ID	Jaccard similarity ⓘ		Description
<a href="#">HOG:D0680756</a>	0.8515625		intraflagellar transport
<a href="#">HOG:D0680535</a>	0.8515625		intraflagellar transport
<a href="#">HOG:D0680689</a>	0.84375		intraflagellar transport

Related functions

Graphical viewer

Members

Ancestral GO

Alignment

Ancestral synteny

Similar HOGs ▾

BASED ON:

Similar domains

# Gene content of ancestral genomes

## Ancestral genome of Viridiplantae

with 85 descendant species and 42339 ancestral genes (HOGs).

Completeness Score =  
Number of species in the HOG/  
number species in the clade

Remove HOGs with completeness score below

Search



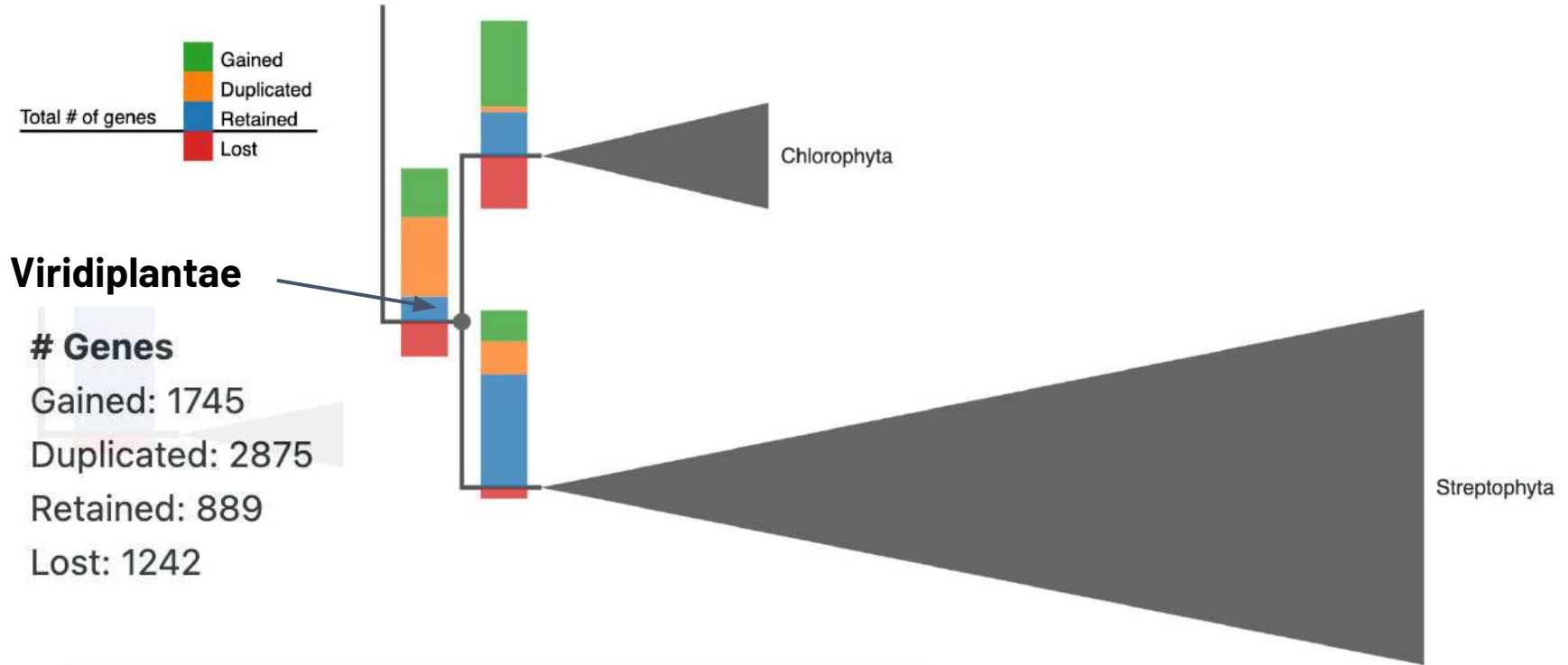
Genome information

Ancestral genes >

Ancestral Gene Order <

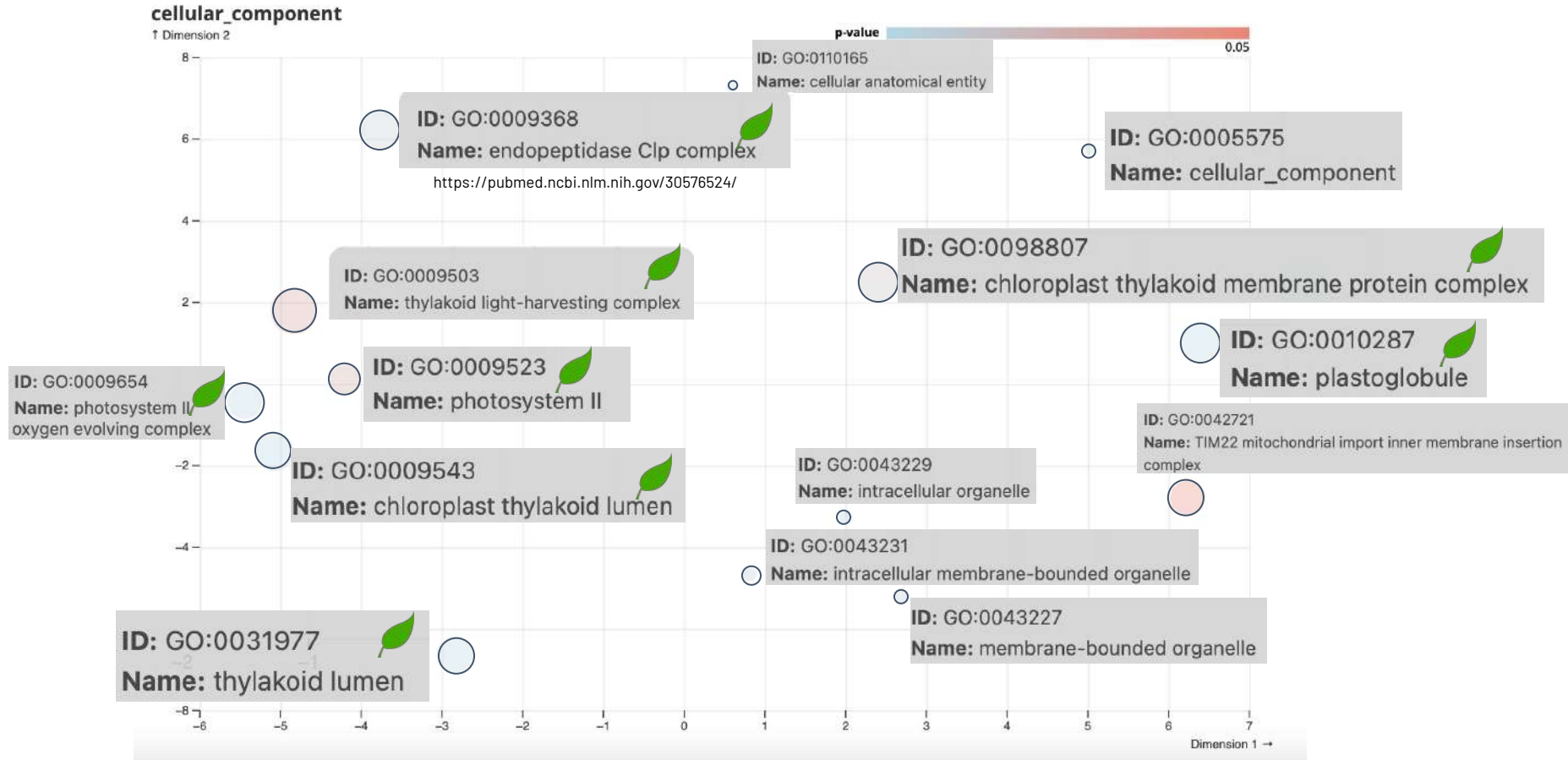
HOG ID	Root HOG ID	Evolutionary event	Completeness	Nr genes in HOG	Description
<a href="#">HOG:D0271083</a>	<a href="#">HOG:D0271083</a>	gained	1.00	454	EGF-like calcium-binding domain-containing protein
<a href="#">HOG:D0271236</a>	<a href="#">HOG:D0271236</a>	gained	0.988	302	Hyaluronan/mRNA-binding protein domain-containing protein
<a href="#">HOG:D0271378</a>	<a href="#">HOG:D0271378</a>	gained	0.976	235	t-SNARE coiled-coil homology domain-containing protein
<a href="#">HOG:D0271254</a>	<a href="#">HOG:D0271254</a>	gained	0.976	104	ATP-dependent Clp protease proteolytic subunit
<a href="#">HOG:D0271213</a>	<a href="#">HOG:D0271213</a>	gained	0.976	145	Chlorophyll a-b binding protein chloroplastic
<a href="#">HOG:D0271177</a>	<a href="#">HOG:D0271177</a>	gained	0.976	276	phosphoglycerate mutase-like protein

# Tracking the evolutionary history of a gene family



**Filtered HOGs.** We only show HOGs with a completeness score above 30%.

# Ancestral GO enrichment of genes gained at the **Viridiplantae** level





# Resources and References

- Orthology: definitions, inferences and impact on species phylogeny inference (Fernández et al., 2019) <https://arxiv.org/abs/1903.04530>
- Inferring orthology and paralogy (Altenhoff and Dessimoz, 2012) <https://people.inf.ethz.ch/adriaal/orthology-bookchapter.pdf>
- Quest for Orthologs (consortium): <https://questfororthologs.org/>

# Before the break

- Do Module 1 of the OMA Academy
  - <https://oma-stage.vital-it.ch/oma/academy/>

## Module 1: Finding orthology with the OMA Browser

The OMA browser serves as an access point for the OMA database, which contains precomputed homology data for over extant and ancestral genomes for over 2800 species (see the [latest list of species](#)).

The OMA browser focuses on three main data types: genes, groups, and genomes. Gene-centric pages provide detailed information about a specific gene, including its sequence, cross-references, functional annotations, and evolutionary data. Group-centric pages classify genes into OMA Groups (Orthologous Groups; OGs) and Hierarchical Orthologous Groups (HOGs) to define families and subfamilies, and Hierarchical Orthologous Groups (HOGs) to define families and subfamilies. Genome-centric pages offer information about extant or ancestral species, associated genes, related genomes, and a synteny viewer.

[Back to home](#) / [Reset](#)

1.1. Browsing the gene page	▼
1.2. Exploring Hierarchical Orthologous Groups	▼
1.3. Browsing the Genomes page	▼

- Break at 10:45



Gitpod

CDEs

Resources ▾

Solutions ▾

Platform ▾

Cust

# Always ready-to-code.

The developer platform for on-demand cloud development environments. Create software faster



## Software

We will be mainly working on an [GitPod](#), an online integrated development environment that allows users to write, edit, and run code directly in a web browser. GitPod is a cloud-based environment where all software, code, and files needed for the course are stored and processed, so you do not need to install or configure anything locally.

You can access the GitPod here:

<https://gitpod.io/#https://github.com/DessimozLab/SIBBiodiversityBioinformatics2023>

Participants need to sign up for a GitPod account via Github and/or LinkedIn to access 50 hours per month for free, which is ample time to complete the exercises. After logging in, create a new workspace by choosing SIBBiodiversityGenomics2023, Browser Editor, and Large configuration (8 cores, 16 GB RAM, 50 GB storage).

**New Workspace**  
Create a new workspace in the **Yannis' Org** organization.

`<>` **github.com/DessimozLab/SIBBiodiversi...**  
Context URL

`>` **VS Code · 1.81.1**  
Editor · Browser

**Large**  
Class · Up to 8 cores, 16GB RAM, 50GB storage

**Continue**

**Autostart with these options for this repository.**  
Don't worry, you can reset this anytime in your [preferences](#).

# New Workspace

Create a new workspace in the **Yannis' Org** organization.



**github.com/DessimozLab/SIBBiodiversi...**

Context URL



**VS Code** · 1.81.1

Editor · Browser



**Large**

Class · Up to 8 cores, 16GB RAM, 50GB storage



**Continue**

**Autostart with these options for this repository.**

Don't worry, you can reset this anytime in your [preferences](#).