

Comparing the Circadian Rhythm Cell Signalling Pathways of *Neurospora crassa* and *Aspergillus flavus* Using CMP

Austin Bassett
V00905244
abassett@uvic.ca
University of Victoria

Helen Dun
V00912482
dunhelen@uvic.ca
University of Victoria



Figure 1: *Aspergillus flavus*. Image from Medmyco (https://commons.wikimedia.org/wiki/File:Aspergillus_flavus.jpg, cropped to fit)

ABSTRACT

Cell signalling pathways are an extremely helpful tool for learning and analyzing how proteins interact within a cell and how those small interactions correlate to bigger systems. However, the creation of these pathways are usually done manually and require specific data or preprocessing of the data. In this paper, we use a program we created and a graphing tool to automatically graph the circadian rhythm cell signalling pathways of *Neurospora crassa* and *Aspergillus flavus*. We compare these using the orthology of the proteins in the pathways and discover that the circadian rhythms of *Neurospora crassa* and *Aspergillus flavus* have a lot in common.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, July 2017, Washington, DC, USA
© 2021 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00
<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

CCS CONCEPTS

- Applied computing → Biological networks.

KEYWORDS

fungi, orthologous, circadian rhythm, cellular signalling pathway, protein-protein interactions (ppi)

ACM Reference Format:

Austin Bassett, V00905244 and Helen Dun, V00912482. 2021. Comparing the Circadian Rhythm Cell Signalling Pathways of *Neurospora crassa* and *Aspergillus flavus* Using CMP. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

1 INTRODUCTION

Due to the interestingness of circadian rhythms and complexity of cellular signalling pathways, for our humble project we compare the circadian rhythm pathways of only 2 fungi, *Neurospora crassa* and *Aspergillus flavus*. Using the CMP program, we compared what proteins are and aren't orthologous between the pathways

and graphed this visually so we could see which parts of the circadian rhythm pathways were significantly different or similar between fungi. Finally, we analyzed why sections of the pathway were similar, simpler or more complex and what the benefits of each were.

1.1 Circadian Rhythms

Even with varying bedtimes and wake up times, people typically live on a 24-hour cycle. This process is known as the circadian rhythm. The circadian rhythm is an internal and natural process which repeats roughly every 24 hours and is known for regulating the sleep-wake cycle [1]. Many organisms have some form of circadian rhythm and have been observed in animals, plants and even fungi and cyanobacteria [1]. Circadian rhythms are adjusted by external cues called zeitgebers, and these include light, temperature, and redox cycles [1]. The driving force of the circadian rhythm is the circadian clock. The circadian clock is a biochemical oscillator which cycles using a stable phase while also being synchronized with solar time [2]. Circadian clocks consist of three major components, a central biochemical oscillator, input pathways to the central oscillator, and output pathways tied to distinct phases of the oscillator [2]. These are all put together using cell signalling.

1.2 Cell Signalling Pathways

Cell signalling is how the cells of an organism signal to each other what to do. Signals are usually propagated using different types of proteins, although other molecules are also used. The way the signal is propagated from a protein to another protein to yet another protein can be mapped to a directional pathway. Pathways are not always linear and can be very complex as sometimes multiple different proteins are needed at the same time to propagate signals and signals can be propagated at the same time to different proteins. Visually mapping the pathways is very important for helping scientists and doctors understand how biological systems work at the cellular level and how different medicines affect the pathways at a chemical level.

2 CMP PROGRAM

To achieve our goal, we developed a Python program that reads in and compares the list of proteins of 2 organisms, the orthology between the 2 organisms and other data. There were 3 different databases used for this:

- Circadian Gene Database (CGD)
- Orthologous MAtrix Project (OMA)
- STRING Database

CGD was used to find the lists of circadian rhythm proteins [3]. OMA was used to find orthology between the proteins of organisms [16-33]. STRING was used to graph the output of the program for better visualization and analysis [4-15].

For 2 organisms A and B, the program outputs 8 different sets of proteins as text files: Imperfect Union (A,B), Perfect Union (A,B), Imperfect Except (A,B), Perfect Except (A,B), Imperfect Union (B,A), Perfect Union (B,A), Imperfect Except (B,A) and Perfect Except (B,A). The next few sections will explain, first, the reason why the

Union sets are ordered, then what Union and Except sets are, and, finally, the difference between Imperfect and Perfect sets.

2.1 Why Order Matters for Union Sets

Some proteins of organism A can be orthologous to many proteins in organism B and vice-versa. Therefore, for the union of orthologous proteins in A and B the program would have to create a node that combines all proteins orthologous with each other. This was infeasible as for the graph part STRING only takes a list of proteins so we would need to create or use some other graphing software for the project. Simplifying the graph of the pathway may also be destructive to information about the proteins. Instead, Union (A,B) is the set of proteins in A that are orthologous to some protein in organism B and Union (B,A) is the set of proteins in B that are orthologous to some protein in organism A.

2.2 What are Union and Except Sets

There are 2 set operations this program performs on the proteins of the 2 selected organisms: Union (U) and Except (-). As mentioned in the previous section, the Union operation is slightly different from the well-known set operation as it does not consolidate the 2 input sets. The Union operation takes the sets of proteins for 2 organisms, A and B, and the orthology between them and returns the subset of proteins of organism A that are orthologous to some protein of organism B. The Except operation takes the same input but does the opposite of the Union function: it returns the subset of proteins of organism A that are not orthologous to any protein of organism B.

2.3 Imperfect vs. Perfect

The difference between Imperfect and Perfect sets is that Imperfect sets have proteins that are orthologous to proteins that may not be part of the circadian rhythm pathway in the other organism. If that did not make sense, here is the explanation again in logical form.

Protein A from Organism X is orthologous to a set of proteins P of Organism Y. For some protein Q in P, Q may or may not be a protein of the cellular signalling pathway being studied in Organism Y. Suppose Q is not. Do we keep that A is orthologous to Q? The Imperfect set keeps the orthology between A and Q and the Perfect set does not.

2.4 Output

Now that that has all been explained, let us go back to the overview of the algorithm. So the program outputs to the ./output/text directory the 8 different sets of proteins as text files. Each protein in the text file is represented by its OMA id and the proteins are separated by newlines. For Union set text files, each protein also lists the set of proteins orthologous to it using OMA ids and with orthology scores. The program also outputs to the ./output/graphs directory the 8 different sets of proteins and the full set of proteins for each organism, but this time it skips the extra Union set stuff and outputs STRING ids instead of OMA ids. These text files can be directly input to the STRING database to get a graph of the set of proteins, which there are some examples of in the next section [4-15]. Along with all of that, the program also outputs how many

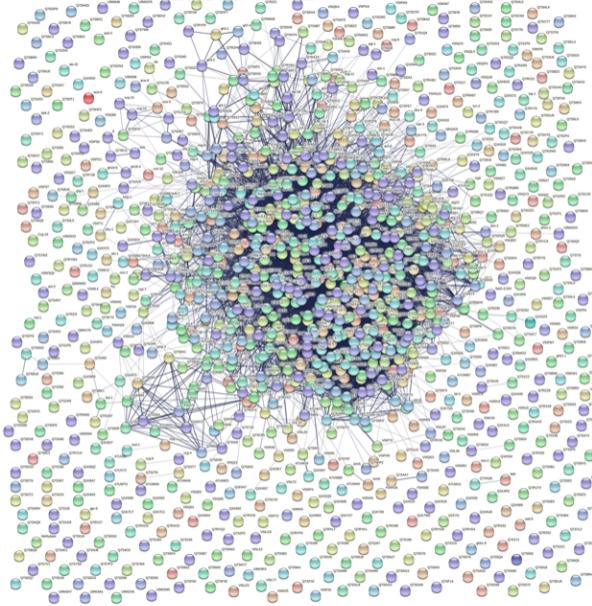


Figure 2: *Neurospora crassa*

proteins are in each set.

We used the program to compare and analyze the phylogeny of the circadian rhythm pathways of the fungi, *Neurospora crassa* and *Aspergillus flavus*.

In this paper we use the program to compare and analyze the phylogeny of the circadian rhythm pathways of the fungi, *Neurospora crassa* and *Aspergillus flavus*.

3 RESULTS

The STRING database can provide a graph of proteins which shows the connection between proteins and the confidence of which the nodes are connected [4-15]. For the edge confidence the thickness of the line determines the confidence of the connection, STRING uses the following four values for confidence, 0.150 (low), 0.400 (medium), 0.700 (high), and 0.900 (highest) [4-15]. The value of 0.150 represents the thinnest line, while 0.900 represents the thickest line. STRING provided an average clustering coefficient, which indicates how connected the graph is, with higher values meaning higher connectivity [4-15]. Finally, STRING provided an average degree node value, which indicates how many interactions a protein has on average in the network [4-15]. The full graphs for *Neurospora crassa* and *Aspergillus flavus* are drastically different.

The *Neurospora crassa* graph contains 1150 protein nodes with not all nodes being connected. The majority of these nodes are densely packed into the middle. The reason these nodes are arranged this way is due to the fact that each of them have an extremely high number of connections, and the average node degree was 20.7 which is clearly located in the middle. The average clustering coefficient was 0.276 which isn't very high meaning a lower connectivity.

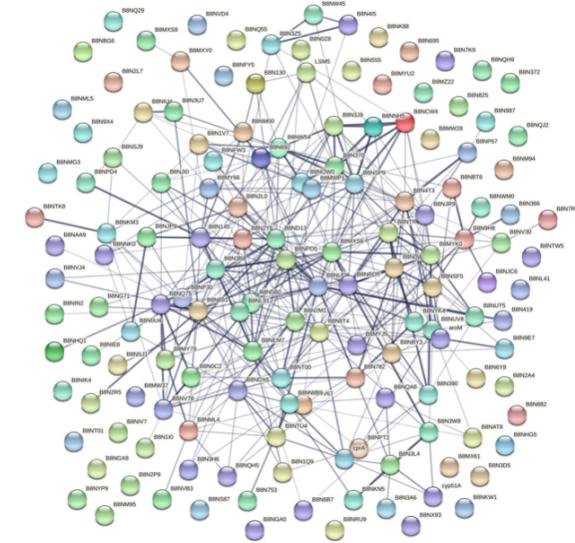


Figure 3: *Aspergillus flavus*

The *Aspergillus flavus* graph only contains 148 protein nodes with not all nodes being connected. *Aspergillus*'s nodes were also located more in the middle, but had more of a spread to them instead of being densely packed. The average node degree was 5.8 which is significantly less than *Neurospora*. The average clustering coefficient was 0.335 which isn't very high meaning a lower connectivity, but more connected than *Neurospora crassa*.

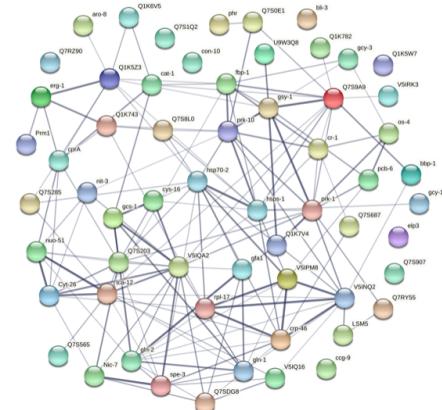
The nodes which were not connected in both graphs were due to CGD saying they were, while STRING did not think they were.

3.1 Order: (*Neurospora crassa*, *Aspergillus flavus*)

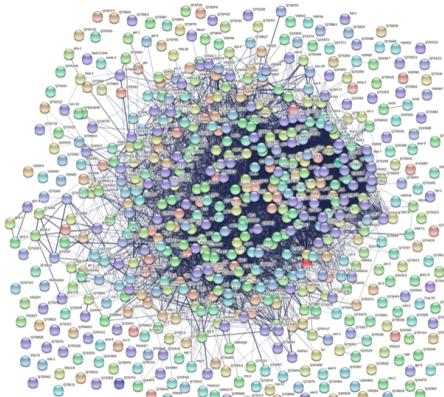
While many of the proteins of the *Neurospora crassa* are orthologous to proteins in *Aspergillus flavus*, not many are orthologous to proteins in *Aspergillus flavus*'s circadian rhythm (52.5% of proteins in Imperfect Union → 5.8% of proteins in Perfect Union). This is either because the proteins of the imperfect set but not of the perfect set are loosely connected to the circadian rhythm and propagate input and output signals, and would evolve to or from proteins in other pathways. Or this is because the list of circadian rhythm proteins used is unfortunately out-of-date. If the first is true, then it means more pathways or evolved, more complex pathways use the circadian rhythm in *Neurospora Crassa* than in *Aspergillus flavus*.

3.2 Order: (*Aspergillus flavus*, *Neurospora crassa*)

In comparison to the huge disparity between the Imperfect and Perfect Union sets of *Neurospora crassa*, *Aspergillus flavus*'s Imperfect and Perfect Union sets are not very different at all (44.6% of proteins in Imperfect Union → 39.2% of proteins in Perfect Union). This disparity seems to be evidence of a lack of up-to-date knowledge about circadian rhythm pathways by the CGD, or at least



(a) Perfect Union



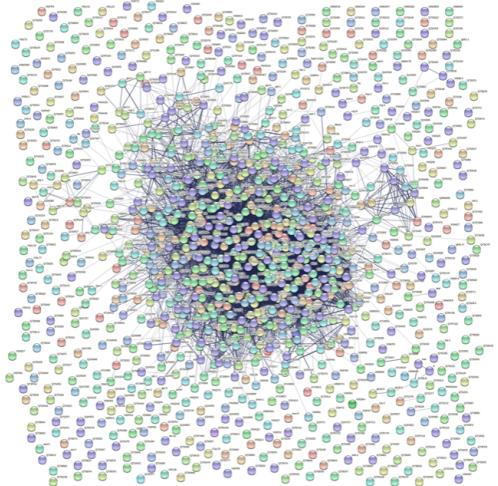
(b) Imperfect Union

Figure 4: Union of *Neurospora crassa*, *Apergillus flavus*

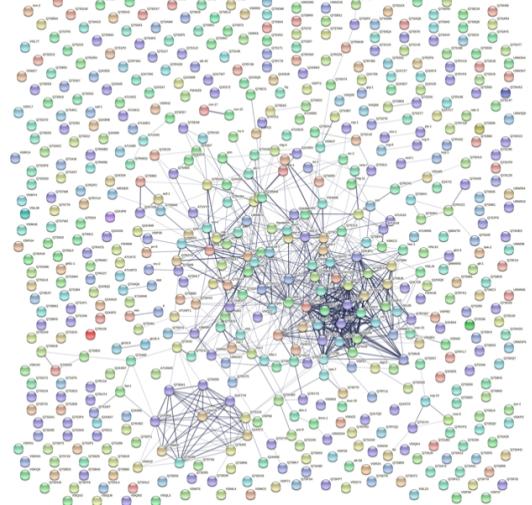
an inconsistency in the way the CGD considers some proteins of the circadian rhythm pathway and others not. Thus the previous idea that other pathways have evolved from a few output-signal-propagating proteins is inconclusive for these graphs, but it may still pertain to future, more accurate graphs.

4 DISCUSSION

The project had multiple issues as a result of using 3 databases together. Other than using each protein's sequence, there was no way to map which protein was which in each database and each database didn't always have every protein. On occasion, one database's organism's list of proteins would include proteins from different variants of that organism. Some proteins were labelled as interacting with each other as part of the circadian rhythm pathway



(a) Perfect Except

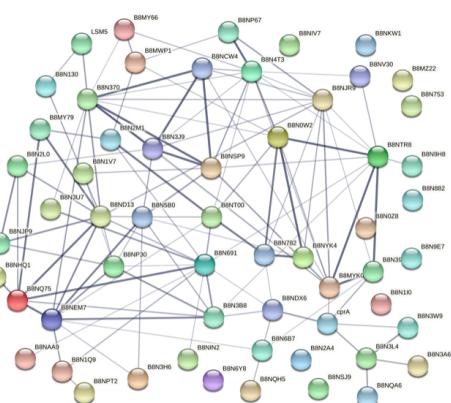


(b) Imperfect Except

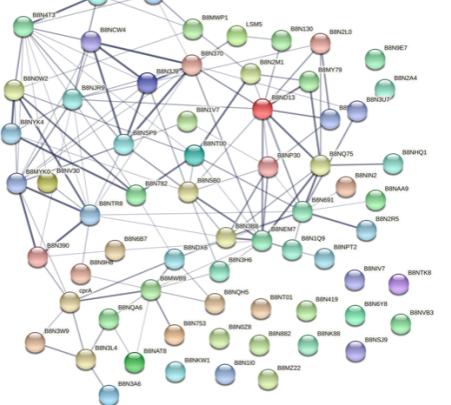
Figure 5: Union of *Neurospora crassa*, *Apergillus flavus*

according to CGD but had no interactions with other proteins according to STRING. Due to these issues, the process of extracting, comparing and graphing the data was not ideal.

For the future, it would be best to use better curated or not out-of-date databases. The next step in the CMP program that was infeasible for this project would be to develop a graphing application which would display the pathway similar to STRING. However, instead of assigning the colours to each node as STRING does, we would assign nodes 1 of 2 colours, red and green. Red would indicate proteins are not orthologous, or that the proteins are from the Except set. Green would indicate proteins are orthologous, or that the proteins are from the Union set. This would allow a user to easily see which parts of the pathway are orthologous and which are



(a) Perfect Union



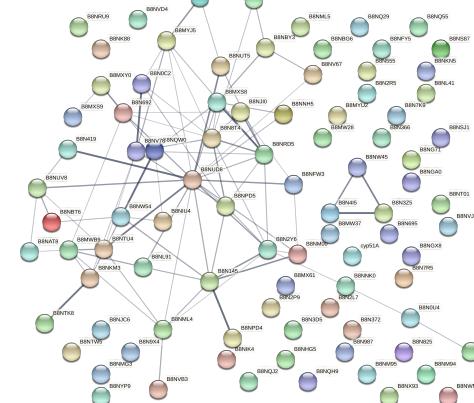
(b) Imperfect Union

Figure 6: Union of *Aspergillus flavus*, *Neurospora crassa*,

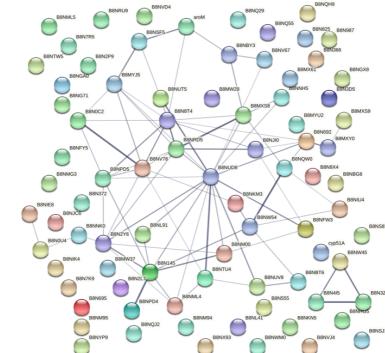
not. It would also be good if the graph was zoomable and the nodes of the graph were clickable so users can find more information about the protein.

5 CONCLUSION

In conclusion, the CMP Program outputs and the STRING graphs showed that there is indeed quite a bit of overlap of circadian rhythm pathways of *Neurospora crassa* and *Aspergillus flavus*. Graphs of pathway comparisons, along with the numbers, are great tools for analysis and will help scientists and doctors better understand the difference and similarities between and evolution of cellular signalling pathways. By comparing the pathways between organisms, it may also become easier to predict the effects



(a) Perfect Except



(b) Imperfect Except

Figure 7: Except of *Aspergillus flavus*, *Neurospora crassa*,

of medicine or the chances of a disease that affects one organism affecting another. As a difficult topic currently being researched, furthering our understanding of cell signalling pathways in all their complexity is one of the next steps in the field of Bioinformatics.

REFERENCES

- [1] "Circadian rhythm," Wikipedia.[Online]. Available: https://en.wikipedia.org/wiki/Circadian_rhythm. [Accessed: 26-Nov-2021.]
- [2] "Circadian clock," Wikipedia.[Online]. Available: https://en.wikipedia.org/wiki/Circadian_clock. [Accessed: 26-Nov-2021.]
- [3] Shujing Li, Ke Shui, Ying Zhang, Yongqiang Lv, Wankun Deng, Shahid Ullah, Luoying Zhang, Yu Xue, CGDB: a database of circadian genes in eukaryotes, Nucleic Acids Research, Volume 45, Issue D1, January 2017, Pages D397–D403,<https://doi.org/10.1093/nar/gkw1028>
- [4] Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ, von Mering C. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 2021 Jan 8;49(D1):D605-D612. doi: 10.1093/nar/gkaa1074. Erratum in: Nucleic Acids Res. 2021 Oct 11;49(18):10800. PMID: 33237311; PMCID: PMC7779004.
- [5] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019 Jan 8;47(D1):D607-D613. doi: 10.1093/nar/gky1131. PMID: 30476243; PMCID: PMC6323986.
- [6] Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C. The STRING database in 2017: quality-controlled protein-protein association networks,

- made broadly accessible. *Nucleic Acids Res.* 2017 Jan 4;45(D1):D362-D368. doi: 10.1093/nar/gkw937. Epub 2016 Oct 18. PMID: 27924014; PMCID: PMC5210637.
- [7] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015 Jan;43(Database issue):D447-52. doi: 10.1093/nar/gku1003. Epub 2014 Oct 28. PMID: 25352553; PMCID: PMC4383874.
- [8] Franceschini A, Lin J, von Mering C, Jensen LJ. SVD-phy: improved prediction of protein functional associations through singular value decomposition of phylogenetic profiles. *Bioinformatics*. 2016 Apr 1;32(7):1085-7. doi: 10.1093/bioinformatics/btv696. Epub 2015 Nov 26. PMID: 26614125; PMCID: PMC4896368.
- [9] Franceschini A, Lin J, von Mering C, Jensen LJ. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 2013 Jan;41(Database issue):D808-15. doi: 10.1093/nar/gks1094. Epub 2012 Nov 29. PMID: 23203871; PMCID: PMC3531103.
- [10] Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 2011 Jan;39(Database issue):D561-8. doi: 10.1093/nar/gkq973. Epub 2010 Nov 2. PMID: 21045058; PMCID: PMC3013807.
- [11] Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julian P, Roth A, Simonovic M, Bork P, von Mering C. STRING 8-a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* 2009 Jan;37(Database issue):D412-6. doi: 10.1093/nar/gkn760. Epub 2008 Oct 21. PMID: 18940858; PMCID: PMC2686466.
- [12] von Mering C, Jensen LJ, Kuhn M, Chaffron S, Doerks T, Krüger B, Snel B, Bork P. STRING 7-recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res.* 2007 Jan;35(Database issue):D358-62. doi: 10.1093/nar/gkl825. Epub 2006 Nov 10. PMID: 17098935; PMCID: PMC1669762.
- [13] von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, Jouffre N, Huynen MA, Bork P. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* 2005 Jan 1;33(Database issue):D433-7. doi: 10.1093/nar/gki005. PMID: 15608232; PMCID: PMC539959.
- [14] von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* 2003 Jan 1;31(1):258-61. doi: 10.1093/nar/gkg034. PMID: 12519996; PMCID: PMC165481.
- [15] Snel B, Lehmann G, Bork P, Huynen MA. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* 2000 Sep 15;28(18):3442-4. doi: 10.1093/nar/28.18.3442. PMID: 10982861; PMCID: PMC110752.
- [16] Adrian M Altenhoff, Clément-Marie Train, Kimberly J Gilbert, Ishita Mediratta, Tarcisio Mendes de Farias, David Moi, Yannis Nevers, Hale-Seda Radoykova, Victor Rossié, Alex Warwick Vesztrocy, Natasha M Glover, Christophe Dessimoz, OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more, *Nucleic Acids Research*, Volume 49, Issue D1, 8 January 2021, Pages D373-D379, <https://doi.org/10.1093/nar/gkaa1007>
- [17] Adrian M Altenhoff, Natasha M Glover, Clément-Marie Train, Klara Kaleb, Alex Warwick Vesztrocy, David Dylus, Tarcisio M de Farias, Karina Zile, Charles Stevenson, Jiao Long, Henning Redestig, Gaston H Gonnet, Christophe Dessimoz, The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces, *Nucleic Acids Research*, Volume 46, Issue D1, 4 January 2018, Pages D477-D485, <https://doi.org/10.1093/nar/gkx1019>
- [18] Adrian M. Altenhoff, Nives Skunca, Natasha Glover, Clément-Marie Train, Anna Sueki, Ivana Piližota, Kevin Gori, Bartłomiej Tomiczek, Steven Müller, Henning Redestig, Gaston H. Gonnet, Christophe Dessimoz, The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements, *Nucleic Acids Research*, Volume 43, Issue D1, 28 January 2015, Pages D240–D249, <https://doi.org/10.1093/nar/gku1158>
- [19] Adrian M. Altenhoff, Adrian Schneider, Gaston H. Gonnet, Christophe Dessimoz, OMA 2011: orthology inference among 1000 complete genomes, *Nucleic Acids Research*, Volume 39, Issue suppl_1, 1 January 2011, Pages D289–D294, <https://doi.org/10.1093/nar/gkq1238>
- [20] Adrian Schneider, Christophe Dessimoz, Gaston H. Gonnet, OMA Browser—Exploring orthologous relations across 352 complete genomes, *Bioinformatics*, Volume 23, Issue 16, 15 August 2007, Pages 2180–2182, <https://doi.org/10.1093/bioinformatics/btm295>
- [21] Glover NM, Altenhoff A, Dessimoz C. 2019. Assigning confidence scores to homoeologs using fuzzy logic. *PeerJ* 6:e6231 <https://doi.org/10.7717/peerj.6231> [22] Clément-Marie Train, Natasha M Glover, Gaston H Gonnet, Adrian M Altenhoff, Christophe Dessimoz, Orthologous Matrix (OMA) algorithm 2.0: more robust to asymmetric evolutionary rates and more scalable hierarchical orthologous group inference, *Bioinformatics*, Volume 33, Issue 14, 15 July 2017, Pages i75–i82, <https://doi.org/10.1093/bioinformatics/btx229>
- [22] Altenhoff AM, Gil M, Gonnet GH, Dessimoz C (2013) Inferring Hierarchical Orthologous Groups from Orthologous Gene Pairs. *PLoS ONE* 8(1): e53786. <https://doi.org/10.1371/journal.pone.0053786>
- [23] Roth, A.C., Gonnet, G.H., Dessimoz, C. Algorithm of OMA for large-scale orthology inference. *BMC Bioinformatics* 9, 518 (2008). <https://doi.org/10.1186/1471-2105-9-518>
- [24] Christophe Dessimoz, Brigitte Boeckmann, Alexander C. J. Roth, Gaston H. Gonnet, Detecting non-orthology in the COGs database and other approaches grouping orthologs using genome-specific best hits, *Nucleic Acids Research*, Volume 34, Issue 11, 1 June 2006, Pages 3309–3316, <https://doi.org/10.1093/nar/gkl433>
- [25] Dessimoz C. et al. (2005) OMA, A Comprehensive, Automated Project for the Identification of Orthologs from Complete Genome Data: Introduction and First Achievements. In: McLysaght A., Huson D.H. (eds) Comparative Genomics. RCG 2005. Lecture Notes in Computer Science, vol 3678. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11554714_6
- [26] Altenhoff, A., Boeckmann, B., Capella-Gutierrez, S. et al. Standardized benchmarking in the quest for orthologs. *Nat Methods* 13, 425–430 (2016). <https://doi.org/10.1038/nmeth.3830>
- [27] Dalquen DA, Altenhoff AM, Gonnet GH, Dessimoz C (2013) The Impact of Gene Duplication, Insertion, Deletion, Lateral Gene Transfer and Sequencing Error on Orthology Inference: A Simulation Study. *PLoS ONE* 8(2): e56925. <https://doi.org/10.1371/journal.pone.0056925>
- [28] Brigitte Boeckmann, Marc Robinson-Rechavi, Ioannis Xenarios, Christophe Dessimoz, Conceptual framework and pilot study to benchmark phylogenomic databases based on reference gene trees, *Briefings in Bioinformatics*, Volume 12, Issue 5, September 2011, Pages 423–435, <https://doi.org/10.1093/bib/bbr034>
- [29] Altenhoff AM, Dessimoz C (2009) Phylogenetic and Functional Assessment of Orthologs Inference Projects and Methods. *PLoS Comput Biol* 5(1): e1000262. <https://doi.org/10.1371/journal.pcbi.1000262>
- [30] Brigitte Boeckmann, Marc Robinson-Rechavi, Ioannis Xenarios, Christophe Dessimoz, Conceptual framework and pilot study to benchmark phylogenomic databases based on reference gene trees, *Briefings in Bioinformatics*, Volume 12, Issue 5, September 2011, Pages 423–435, <https://doi.org/10.1093/bib/bbr034>
- [31] Altenhoff AM, Dessimoz C (2009) Phylogenetic and Functional Assessment of Orthologs Inference Projects and Methods. *PLoS Comput Biol* 5(1): e1000262. <https://doi.org/10.1371/journal.pcbi.1000262>
- [32] Altenhoff A.M., Dessimoz C. (2012) Inferring Orthology and Paralogy. In: Anisimova M. (eds) Evolutionary Genomics. Methods in Molecular Biology (Methods and Protocols), vol 855. Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-61779-582-4_9