



# FunCore: a core gene database for fungi

Wen-Huei Lo<sup>1,2,\*</sup>, Shan-Syue He<sup>1,3,\*</sup>, Yu-Chun Huang<sup>1</sup>, Ming-Ren Yen<sup>1</sup>, and Pao-Yang Chen<sup>1</sup>

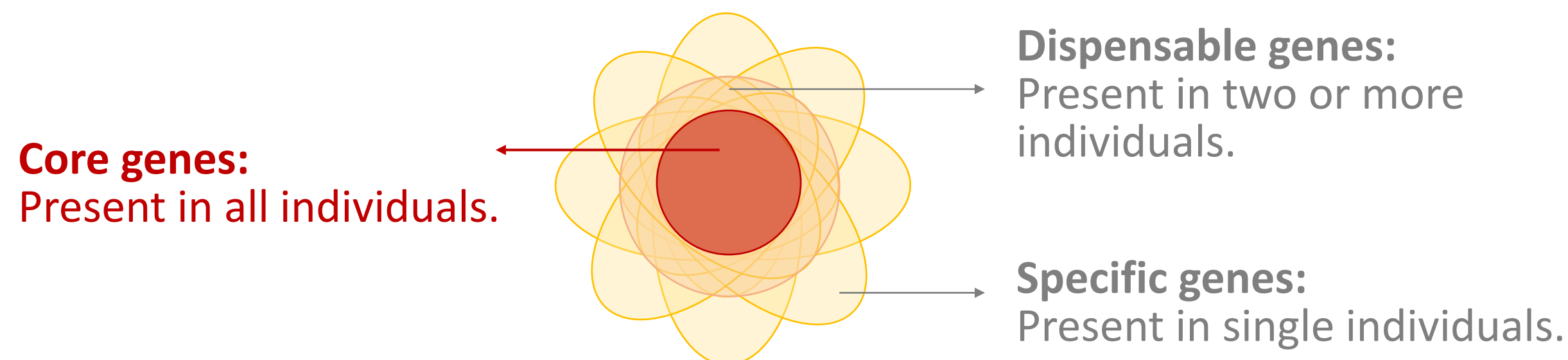
<sup>1</sup>Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan. <sup>2</sup>Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan. <sup>3</sup>Department of Agronomy, National Chung Hsing University, Taichung, Taiwan.

\*Both authors contributed equally to this work.

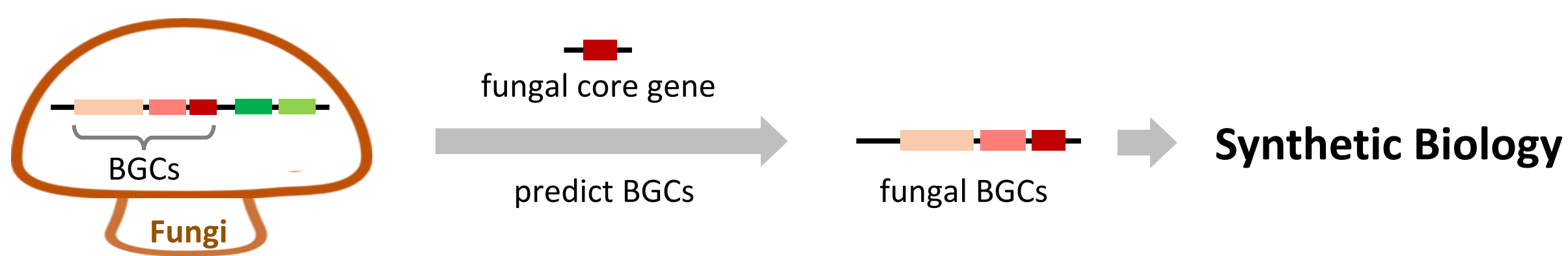


## Background

Core genes exist in all organisms. Some of them play an important role for survival, such as essential genes. Similar gene sequences between organisms can be considered as a core gene and are able to be identified by homology search.



Databases of core gene can be helpful for searching biosynthesis gene clusters (BGCs) which encode the enzymatic pathway for secondary metabolites production. Therefore, the study of fungal core genes opens up a great potential for synthetic biology researches.



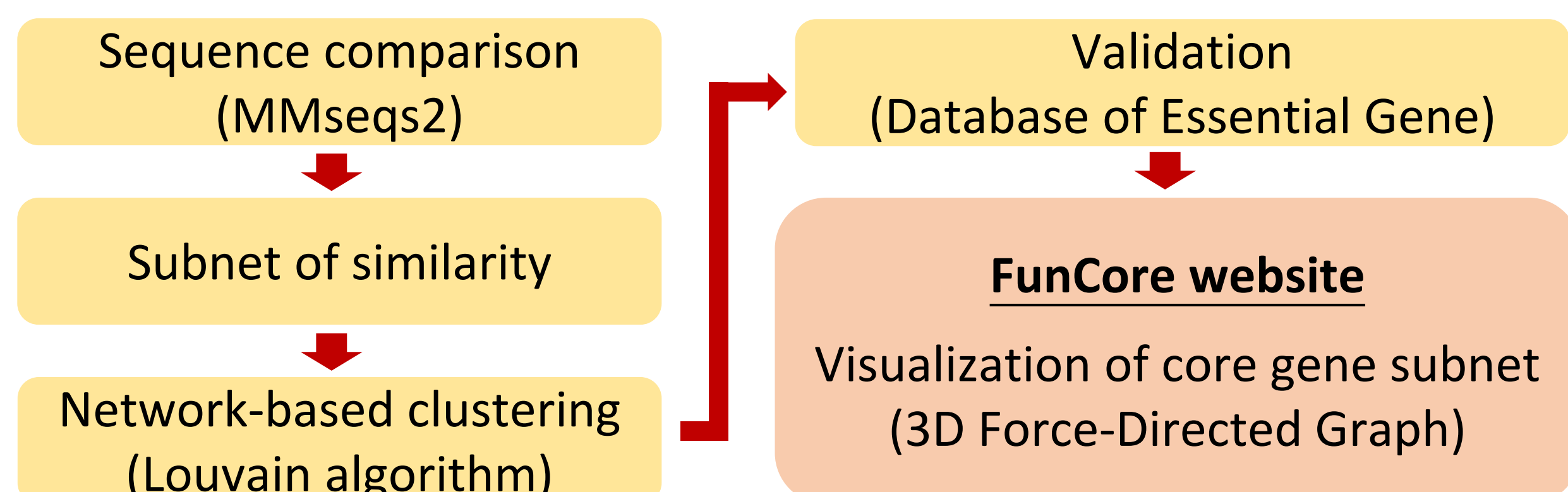
**Aim:** In fungi, very little is known about their core genes. Hence we are going to build a fungal core gene database (FunCore).

## Materials & Methods

### Materials

- One fungal genome from each genus.
- 4,272,737 protein sequences of 347 fungus from Joint Genome Institute.

### Pipeline of building fungal core gene database.



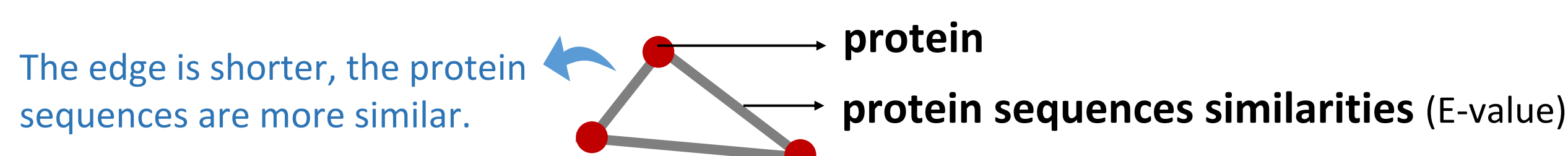
### Homology search for duplicate genes.

Compare 2 protein sequences and use E-value to describe their similarity.

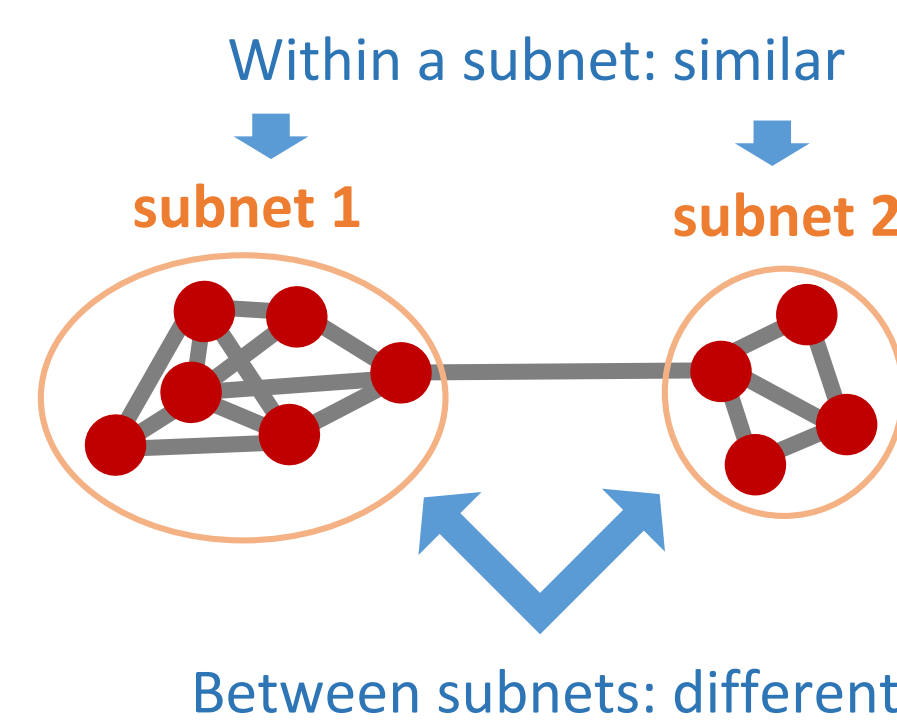


### Network-based clustering to identify densely connected subnets.

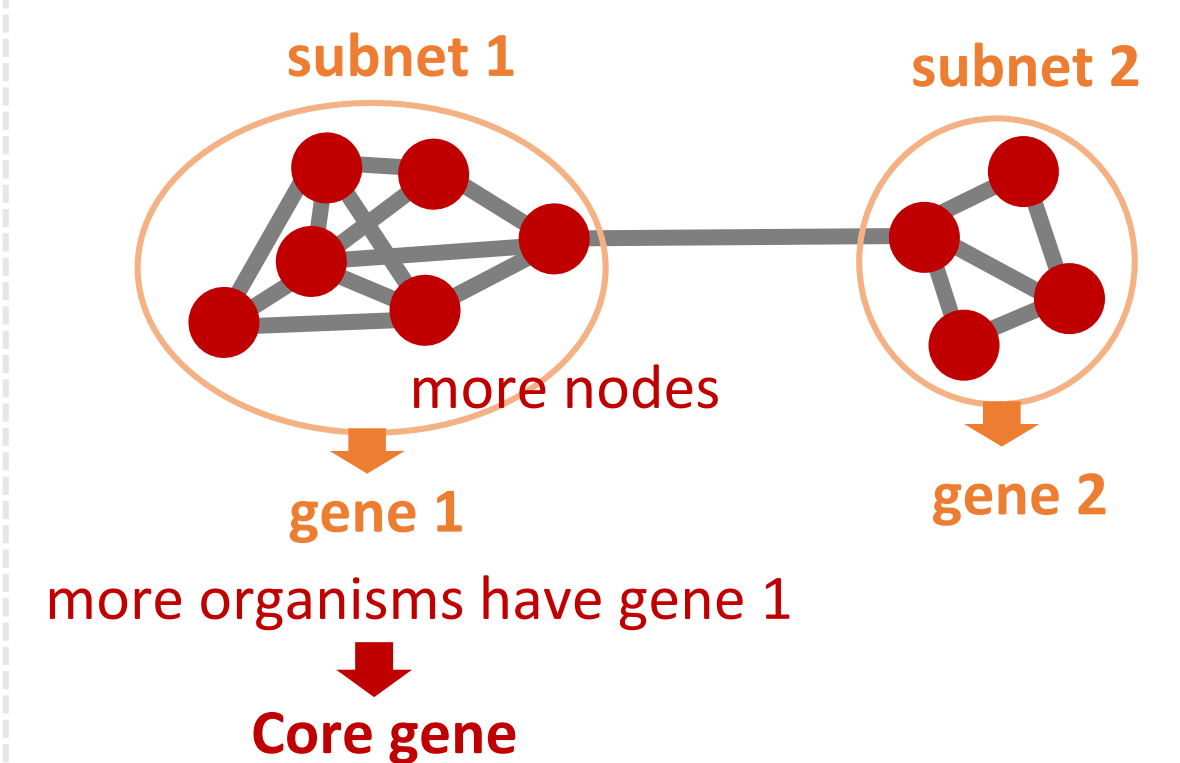
Use network to show the similarities of protein sequences.



### Clustering to detect densely connected subnets



### Core gene prediction



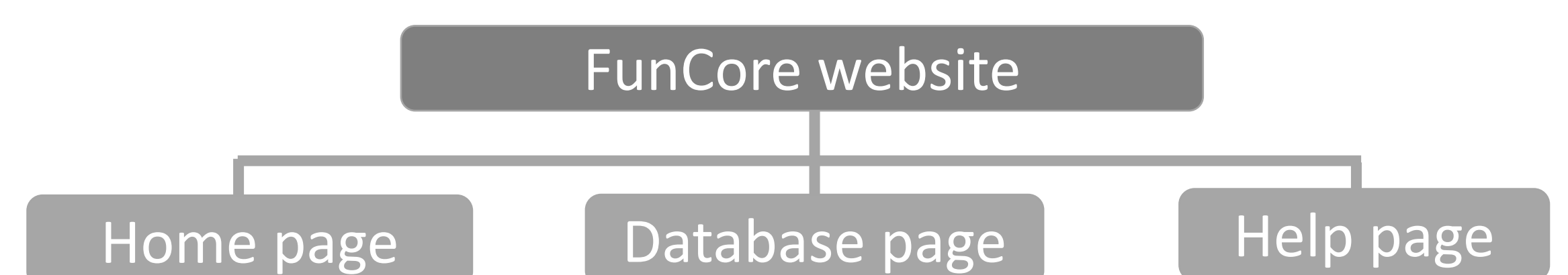
### Validation

Compare our predicted core genes with known essential gene of fungi from Database of Essential Gene.

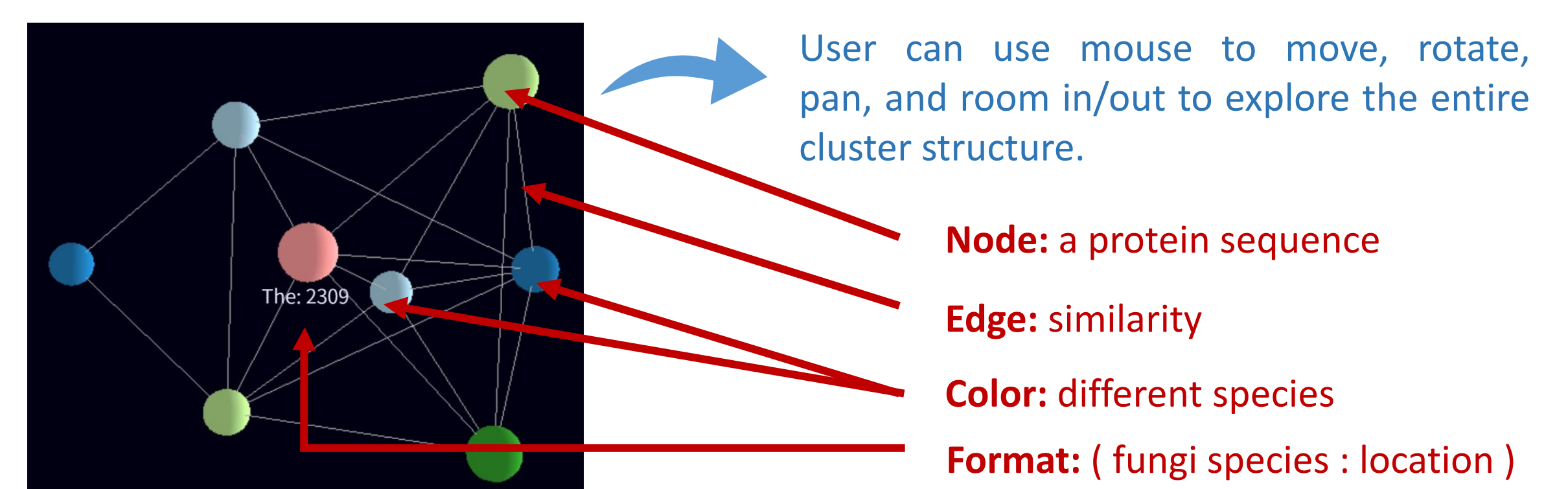
## Results

### FunCore Website

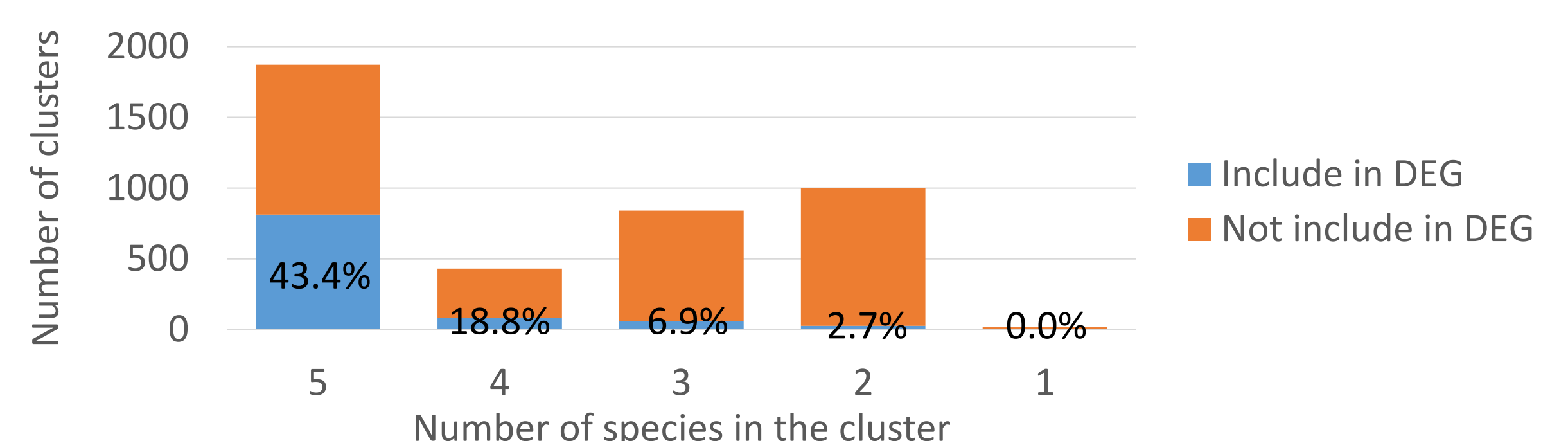
#### 1. Website architecture



#### 2. Display the core gene subnets in Three-Dimensional.



### Preliminary result from 5 fungus analysis



## Summary

- Construct a pipeline for identifying core genes and build the first fungal core gene database.
- Show the protein sequences similarities by 3D graph.
- Use a network-based approach for clustering.

## Reference

- Steinegger, M., and Söding, J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature biotechnology* 35(11): 1026-1028.
- Blondel, V. D. et al. (2008) Fast unfolding of communities in large networks. *J. Stat. Mech.* 2008(10): P10008.
- Luo, H. et al. (2013) DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nucleic acids research* 42(D1): D574-D580.