

1. **Research Unit:** Laboratory of Environmental and Evolutionary Systems Biology
2. **Supervisor:** Dr hab. Anna Muszewska
3. **Supervisor (email):** musze@ibb.waw.pl
4. **Project title (English):** Nondomain proteins encoded by fungal genomes
5. **Project title (Polish):** Białka niedomenowe kodowane w genomach grzybów
6. **Description of the project (up to 500 words):**

Fungi are important for the health and human economy as pathogens of crops, fermenting organisms, with additional use in bioreactors. In our previous projects we have linked the distribution of diverse gene classes to the ecology of fungi [1,2]. Currently, only homologs of defined and relatively conserved protein families are widely used for genomic comparative studies. However, (fungal) genomes harbour a fraction of genes encoding protein without detectable protein domains. The main idea of the project is to explore the properties of these uncharacterized proteins. We will use sequences of fungal genomes and use diverse tools to predict functional categories and characterize these proteins. One of the aims of this project is to maximize the usage of available protein sequences to comprehend the underpinnings of fungal ecology and adaptability. The project will integrate information on the evolution of particular fungal groups, their characteristic structures, ecological properties with the identified in the course of this project protein properties. We will try to answer questions whether some lineages/**taxa are more** susceptible to having non-domain proteins? Which categories of non-domain proteins are found in fungi and do they change across the fungal tree of life? Does the size of the **genome** influence the share of non-domain proteins? Are non-domain protein-coding genes colocalized? Are they randomly distributed in the genome? Does the **lifestyle** of the fungus and its ecological strategy impact on the number of non-domain proteins? How about the properties of the **proteins** without domains, do they differ from the observed feature distribution for the proteome of origin (amino-acid composition, length, predicted cellular localization)?

7. **References related to conducted /planned research (maximum 3):**

- Muszewska, Anna, Marta M. Stepniewska-Dziubinska, Kamil Steczkiewicz, Julia Pawlowska, Agata Dziedzic, and Krzysztof Ginalski. 2017. "Fungal Lifestyle Reflected in Serine Protease Repertoire." *Scientific Reports* 7 (1): 9147. <https://doi.org/10.1038/s41598-017-09644-w>
- Muszewska, Anna, Kamil Steczkiewicz, Marta Stepniewska-Dziubinska, and Krzysztof Ginalski. 2017. "Cut-and-Paste Transposons in Fungi with Diverse Lifestyles." *Genome Biology and Evolution* 9 (12): 3463–77. <https://doi:10.1093/gbe/evx261>
- Mistry, Jaina, Sara Chuguransky, Lowri Williams, Matloob Qureshi, Gustavo A. Salazar, Erik L. L. Sonnhammer, Silvio C. E. Tosatto, et al. 2020. "Pfam: The Protein Families Database in 2021." *Nucleic Acids Research*, October. <https://doi.org/10.1093/nar/gkaa913>.

8. **Scholarship amount (net):** 3000 PLN for mid-term evaluation, after mid-term evaluation, change to 57% professor's remuneration (currently it would be 3242 PLN net).