

ALFREDO LEONARDO PORFIRIO DE SOUSA

Elucidating the molecular machinery of an evolutionary novelty:
Single-cell transcriptomics of *Arcella intermedia* and characterization
of gene expression during shell formation.

Elucidando a maquinaria molecular de uma novidade evolutiva:
transcriptômica *single-cell* de *Arcella intermedia* e caracterização da
expressão gênica durante a formação de teca.

São Paulo

2018

Abstract

The present dissertation aims to shed light on the molecular machinery involved in the process of shell formation (thecagenesis) in *Arcella* (Arcellinida : Amoebozoa). Arcellinida are single-celled testate amoebae organisms, characterized by the presence of an outer shell (test or carapace); it is a monophyletic lineage of Amoebozoa, sister group to a naked amoeboid lineage. No homologous structure to shell is present in the sister group of Arcellinida, thus it is considered an evolutionary novelty. The origin and evolution of the shell in Arcellinida are currently open questions; deciphering its formation process is a key step to address these questions. During each reproductive process by budding division, these organisms build a new shell. In the span of more than a century, several authors have described the thecagenesis process on Arcellinida, primarily focusing on the genus *Arcella*, based on cyto-morphological evidence. Conversely, the absence of molecular data has impaired advances on describing the molecular aspects of shell formation. In this study, we designed and applied a molecular framework to identify candidate genes and develop a molecular model for the shell formation process in *Arcella*; we based this framework on single-cell RNA-sequencing, gene expression profiling, Gene Ontology analysis, and comparative analysis of cyto-morphological with newly generated molecular data. We identify and propose a set of 539 genes as the candidate genes for shell formation, based on expression profiling and biological process assignment. We propose a model for the the shell formation process, which describes the mechanistic aspect of this process, hypothetically based on a molecular machinery conserved in Eukaryotes. Additionally, we identified a massive expansion of the Rab GTPase family, a protein likely to be involved on the process of shell formation. In the lights of the present study, we briefly discuss possible evolutionary scenarios involved on the origin and evolution of the shell and present future perspectives; we propose the shell of Arcellinida as a prosperous model to study the origin and evolution of evolutionary novelties, as well as other evolutionary questions.

Key words: Amoebozoa; Arcellinida; thecagenesis; Molecular model; evolutionary novelty; Rab GTPases.

Abstract

A presente dissertação tem como objetivo lançar luz sobre a maquinaria molecular envolvida no processo de formação de teca (tecagênese) em *Arcella* (Arcellinida: Amoebozoa). Arcellinida são amebas tecadas unicelulares, caracterizadas pela presença de uma teca (carapaça ou concha) externa; é uma linhagem monofilética de Amoebozoa, grupo irmão de alguns organismos amebóides nus. Nenhuma estrutura homóloga à carapaça está presente no grupo irmão de Arcellinida, sendo considerada como uma novidade evolutiva. A origem e evolução da carapaça em Arcellinida são questões em aberto; Decifrar seu processo de formação é um passo fundamental para abordar essas questões. Durante todo processo reprodutivo, por divisão por brotamento, estes organismo constroem uma nova concha. No decorrer de mais de um século, vários autores descreveram o processo de tecagênese nestes organismos, focando principalmente no gênero *Arcella*, baseados em evidências cito-morfológicas. Enquanto isso, a ausência de dados moleculares impede avanços na descrição dos aspectos moleculares da formação de conchas. Neste estudo, projetamos e aplicamos uma *framework* molecular para identificar genes candidatos e desenvolver um modelo molecular para o processo de formação de teca em *Arcella*; Baseamos este *framework* em sequenciamento de RNA *single-cell*, perfil de expressão gênica, análise de *Gene Ontology* e análise comparativa de dados cito-morfológicos e moleculares. Nós identificamos e propomos um conjunto de 539 genes como genes candidatos para a formação de carapaça, com base no perfil de expressão e na atribuição de processos biológica. Propomos um modelo para o processo de formação de carapaça, que descreve o aspecto mecanicista deste processo, hipoteticamente baseado em um mecanismo molecular conservado em Eucariotos. Além disso, identificamos uma expansão maciça da família gênica das Rab GTPase, gene provavelmente envolvida no processo de formação de carapaça. À luz do presente estudo, discutimos brevemente possíveis cenários evolutivos envolvidos na origem e evolução da teca e apresentamos perspectivas futuras; propomos a teca dos Arcellinida como próspero modelo para estudar a origem e evolução das novidades evolutivas, bem como outras questões evolutivas.

Palavras-chave: Amoebozoa; Arcellinida; tecagênese; modelo molecular; novidade evolutiva; Rab GTPases.

1 General Introduction

Amoebae are organisms characterized by amoeboid movement, in at least one stage during their life cycle. These organisms constitute a paraphyletic group traditionally recognized as Sarcodina (Page, 1976). Currently, amoebae are classified in diverse eukaryotic groups such as, Rhizaria, Amoebozoa, Heterolobosea (Excavata), Stramenopila, Nucleariida (Opisthokonta), and Actinophryidae (Chromalveolata) (**Figure 1.0.1**) (Adl et al. 2005; Pawlowski 2008; Brown et al. 2013; Adl et al., 2018); the majority of amoebae diversity is shared between Rhizaria and Amoebozoa. Amoebozoa is a monophyletic super-group of Eukaryotes (**Figures 1.0.1 - 1.0.2**), comprising a large diversity of organisms presenting vastly diverging morphology, life cycles, and cellular structures (Kang et al., 2017; adl., 2018). The last decade saw advances in describing the diversity and evolution of this group. Amoebozoa include some of the well known amoebae, such as *Amoeba proteus*, *Dicystelium discoideum* and the pathogens *Entamoeba histolytica* and *Acanthamoeba castellanii* (**Figure 1.0.2**). Among them, the monophyletic Arcellinida, a group of organisms that present an external structure, the shell (test), covering these single-cell organisms (**Figures 1.0.2 - 1.0.3**). The Arcellinida have been classified as a member of Amoebozoa only in the past decade (Nikolaev, 2005); in recent years morphological and molecular studies, including phylogenomics, have improved our understanding of the diversity and evolution of this group (Kozakyan et al., 2016; Lahr et al., accepted). The presence of a proteinaceous shell, a hardened outer structure with a single aperture, is the key characteristic of Arcellinida. Several authors studied the morpho-cytological process of shell formation in Arcellinida (Netzel, 1971; Netzel, 1972; Netzel, 1975a; Netzel, 1975b; Netzel, 1975c; Netzel and Grunewald, 1977; Netzel, 1980; Mignot and Raikov, 1990). On the other hand, the origin and evolution of the arcellinid shell remain open questions; and no

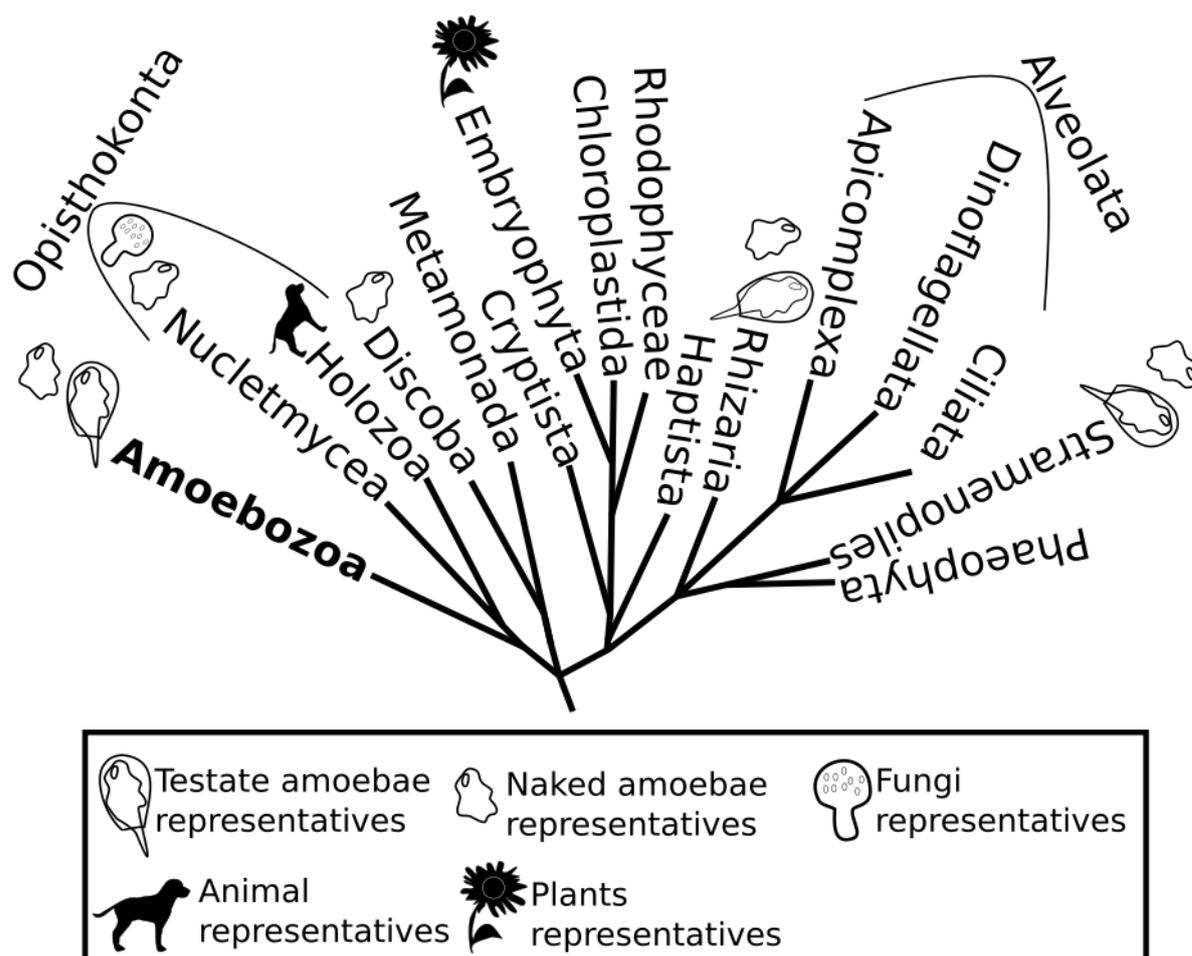


Figure 1.0.1 – Schematic overview of the diversity of eukaryotes based on adl et al. (2018). We highlight the lineages with naked amoebae representatives and testate amoebae representatives, as well as, the groups that comprise Fungi, Animals, and Plants. Amoebozoa, our focus group, is sister group to the eukaryotic lineages comprising Animals and Fungi. Testate amoebae lineages evolved in three different lineages of eukaryotes.

molecular data regarding shell formation is available, impairing the elucidation of its evolutionary story.

The main goal of the present study was to identify candidate genes, and its encoded proteins, possibly involved in shell formation of *Arcella intermedia* (Arcellinida:Amoebozoa), shedding light on the molecular process of shell formation.

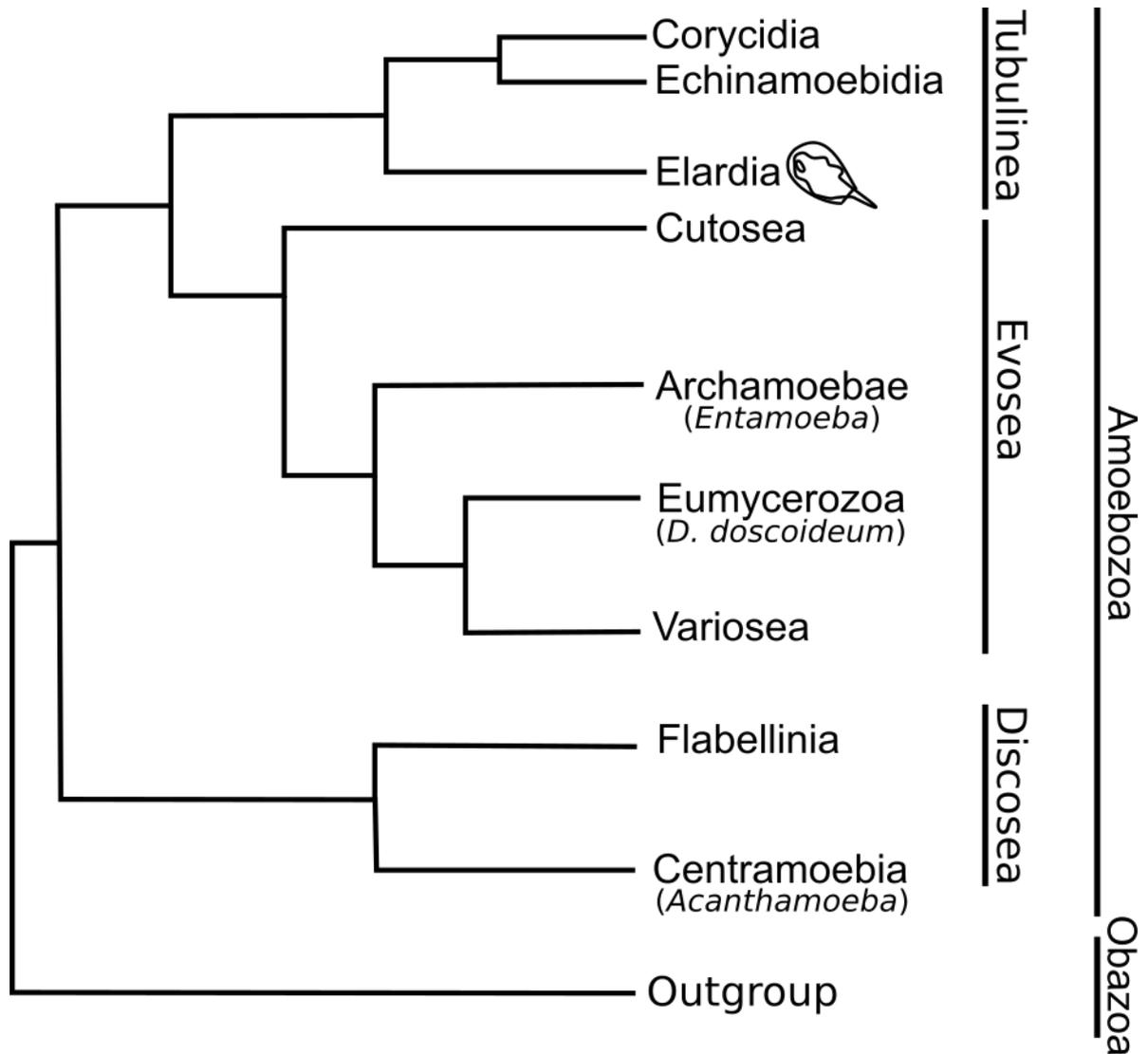


Figure 1.0.2 – Schematic Amoebzoa tree based on the phylogenomic study Kang et al. (2017). Currently, Amoebzoa are classified in tree major lineages (Tubulinea, Evosea and Discosea). Arcellinida are a monophyletic lineage of Elardia.

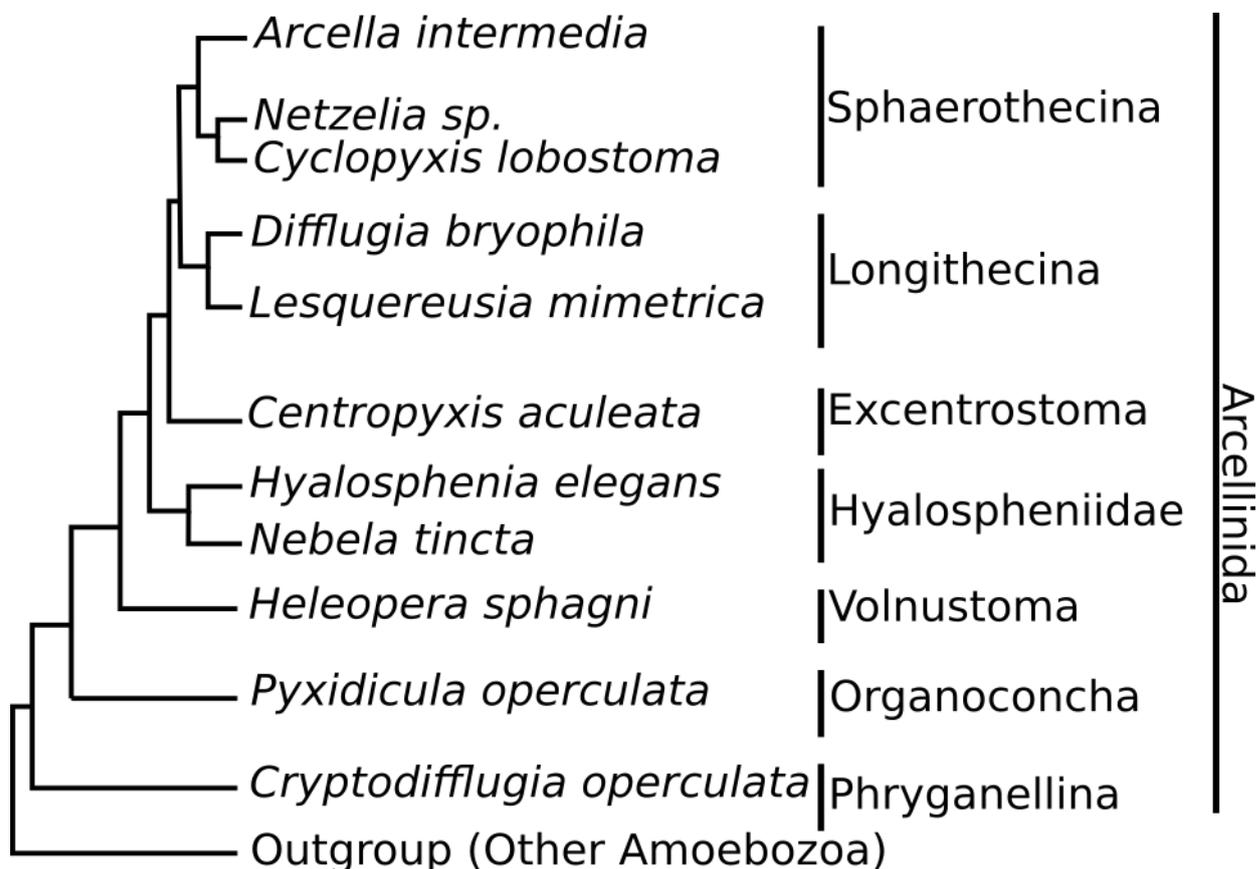


Figure 1.0.3 – Schematic Arcellinida tree based on the phylogenomic Lahr et al. (accepted). Currently, Arcellinida monophyly is recovered and classified in five Infraorders (Sphaerothecina, Longithecina, Excentrostoma, Hyalospheniidae, and Volnustoma) and two Suborders (Organoconcha and Phryganelliina), estimated to comprise around 800-2,000 morphospecies.

We reviewed the literature that presents seminal studies, discussions, and reviews about shell formation in *Arcella*; we present this review on **Chapter 1** as an overview of the literature regarding the cyto-morphological description of shell formation in the genus *Arcella*. We performed a single-cell transcriptomic experiment of *Arcella intermedia* aiming to shed light on the genes involved in the shell formation process; we identified candidate genes that may be involved in this process and propose an annotated gene list that can be further studied and tested; we present this experiment, its results, and discussion in **Chapter 2** as a gene-expression profiling of *Arcella intermedia* during shell formation. We combined the morpho-cytological knowledge present in the literature, reviewed in **Chapter 1**, and the newly generated transcriptomic data, presented in **Chapter 2**, to propose the first molecular interpretation of the shell formation process; we propose our molecular interpretation of shell formation in **Chapter 3**, "translating" the morpho-cytological evidence on a molecular interpretation of the mechanisms of shell formation in *Arcella*. We identified a massive expansion of the Rab GTPase family in Amoebozoa, we describe this observation in **Chapter 4** as a phylogenetic study revealing a massive RabGTPase family expansion in Amoebozoa. Finally, we summarize our findings and propositions, and discuss the perspectives of the present work on the dissertation's **Final considerations**.

6 Final considerations

In this thesis, we discussed the cyto-morphological and molecular aspects of the thecagenesis process on Arcellinida (Amoebozoa), focusing on the genus *Arcella*. We reviewed the literature that comprehensively describes shell structure and formation process on *Arcella*, generated transcriptomic data for *Arcella intermedia*, and combined both for the proposition of a model for the molecular machinery involved in shell formation on these organisms. We presented the literature review in **Chapter 1**, describing how a detailed description, based on a cyto-morphological framework, has been developed in the span of more than a century, with the contribution of several authors; in three different moments this literature saw long periods of slower advances, the first between 1838 and 1864, the second between 1928 and 1963, and the third in our contemporary time, from 1990 to date, which is represented by only two studies (Pchelina, 2010; Volkova and Alexey, 2016). Currently, the involved molecular machinery, shell origin, and shell evolution, to cite some, are still puzzles. The lack of additional data and a new framework, impairs further advances. In **Chapter 2**, we presented the newly generated single-cell transcriptomic data, which describes the gene expression of *A. intermedia* during shell formation and after twelve hours of shell formation. We demonstrated that each single-cell transcriptome is an accurate representation of *A. intermedia*'s transcriptome. From the set of expressed genes successfully annotated by GO analysis, 539 genes are assigned to biological processes described on the cyto-morphological studies of thecagenesis on *Arcella*, thus we propose them as candidate genes to be involved on shell formation. Moreover, a significant part of the transcriptome was not successfully GO resolved, and is available for gene discovery on *Arcella*. Rather than a sense of substitution, the generation of molecular data aims to enhance the interpretative power of the cyto-morphological

data, and vice versa. In **Chapter 3**, we presented the model developed for the molecular machinery involved on shell formation. This model takes in account the cyto-morphological description of shell formation on *Arcella*, the newly generated molecular data, and the general description of subcellular molecular machineries on closely related eukaryotic lineages. Our model addresses that the mechanistic part of the thecogenesis process is likely to involve proteins that are conserved on Eukaryotes. In **Chapter 4**, we presented a phylogenetic study of the Rab GTPase family; our analysis identified that the pattern of massive expansion of Rabs is characteristic of Amoebozoa.

The present state of the literature about shell formation implies two different perspectives; one of a slower advance, the other of an open niche for research and rapid development. The cyto-morphological studies, using different techniques (e.g. *in vivo* observations and tangential sections) and technologies (e.g. light microscopy, scanning electron microscopy, and transmission electron microscopy), has shown us one more example of how an established framework, cultivated by generations of researchers, impacts our understanding about the natural world. Currently, we have available a diverse bulk of advanced technology and technique for molecular biology (e.g. transcriptomics, proteomics, and genomics) and cytological studies (e.g. immunoelectron microscopy and *in vivo* Real-time 3D time-lapse imaging), to cite some. Our transcriptomic experiment has shown how power-full and useful such technique can be, even on a non-traditional model organism, including a single-cell accuracy. Combined, the gradually corroborated cyto-morphological evidence and newly generated data will benefit our knowledge about the shell formation process, as well as the origin and evolution of this structure.

Based on the diversity and evolution of Amoebozoa as we understand now, the shell of Arcellinida is an evolutionary novelty (i.e. there is no homologous, or intermediate, structure on the sister group of Arcellinida). Some evolutionary

scenarios can be applied to hypothesize the origin and evolution of shell as, concerted evolution, gene duplication, and Lateral Gene Transfer (LGT), of the proteins involved on the process of shell formation. Based on the first scenario, the concerted evolution of proteins present on the hypothetical naked ancestor of Arcellinida, with no additional copy of preexisting proteins, would lead to the origin of the molecular machinery involved on shell formation. Differently, gene duplication implies the emergence of new copies from preexisting genes on the hypothetical naked ancestor of Arcellinida, followed by subfunctionalization of each copy, with the consequent emergence of novel morphogenetic traits. On the Lateral Gene Transfer scenario, it is expected that gain of new genes from other lineages by the hypothetical naked ancestor of Arcellinida, would be the driving force for the origin of new molecular machinery and cellular processes, necessary to build a shell. Based on the molecular model proposed on **Chapter 3**, at least the mechanistic aspect of the shell formation process may be explained by the involvement of molecular processes present in other eukaryotes, possibly with the participation of gene duplication and subfunctionalization; however, the synthesis of the thecagenous material, so far of unknown nature, may be explained by the gain of new genes. **Chapter 4** demonstrates a gene family expanding through gene duplication, leading to the presence of diverse paralogues of slightly different functional proteins. Although, being in an speculative stage of discussion about the origin and evolution of shell on *Arcella*, it is clear how crucial is to understand the molecular basis of the shell formation process; we will be able to address different aspects about the biology of *Arcellinida*, as well as general aspects of biology.

In the lights of the present work and discussion, we propose the establishment of a new framework to study the shell formation process on *Arcellinida*; a molecular based framework, coupling both molecules sequencing and cytological study, evolutionarily-informed and, historically aware of the state-of-the-art. Moreover,

we propose the shell of *Arcellinida* as a model to study the origin and evolution of evolutionary novelties, as well as other evolutionary questions. It is a current discussion what does or does not define and characterize an evolutionary novelty, or even if such novelties exist. Similarly, discoveries on Arcellinida may shed light on issues as deep homology or even the predictable aspect of evolution. *Arcellinida*, as well as Amoebozoa, have been consistently studied in the past decade and currently present a comprehensive phylogenomic paradigm, based on morphology, ecology and phylogenomics incredible advance on near well studied traditional model organisms, but as well a group that has seen rapid advance, with a consistent Phylogenomics. Finally, with the present dissertation we aim to promote new future advances on the studies of the thecagenesis process in Arcellinida, with a molecular framework likely to be applied to other testate amoebae lineages.

1.1 *References*

Adl, S. M., Bass, D., Lane, C. E., Lukeš, J., Schoch, C. L., Smirnov, A., Agatha S., Berney C., Brown M.W., Burki F., Cárdenas P., Čepička I., Chistyakova L., Del Campo J., Dunthorn M., Edvardsen B., Eglit Y., Guillou L., Hampl V., Heiss A.A., Hoppenrath M., James T.Y., Karpov S., Kim E., Kolisko M., Kudryavtsev A., Lahr D.J.G., Lara E., Le Gall L., Lynn D.H., Mann D.G., Massana I Molera R., Mitchell E.A.D., Morrow C., Park J.S., Pawlowski J.W., Powell M.J., Richter D.J., Rueckert S., Shadwick L., Shimano S., Spiegel F.W., Torruella I., Cortes G., Youssef N., Zlatogursky ., & Zhang Q. (2018). Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *Journal of Eukaryotic Microbiology*.

Adl, S. M., Simpson, A. G., Farmer, M. A., Andersen, R. A., Anderson, O. R., Barta, J. R., Bowser S.S., Brugerolle G., Fensome R.A., Fredericq S., James T.Y., Karpov S., Kugrens P., Krug J., Lane C.E., Lewis L.A., Lodge J., Lynn D.H., Mann D.G., McCourt R.M., Mendoza L., Moestrup O., Mozley-Standridge S.E., Nerad T.A., Shearer C.A., Smirnov A.V., Spiegel F.W., & James, T. Y. (2005). The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *Journal of Eukaryotic Microbiology*, 52(5), 399-451.

Brown, M. W., Sharpe, S. C., Silberman, J. D., Heiss, A. A., Lang, B. F., Simpson, A. G., & Roger, A. J. (2013). Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *Proceedings of the Royal Society of London B: Biological Sciences*, 280(1769), 20131755.

Jan, P. (2008). The twilight of Sarcodina: a molecular perspective on the polyphyletic origin of amoeboid protists. *Protistology*, 5(4).

Kang, S., Tice, A. K., Spiegel, F. W., Silberman, J. D., Pánek, T., Čepička, I., Kostka. M., Kosakyan, A., Alcântara, D. M. C., Roger A. J., Shadwick, L.L., Smirnov, A., Kudryavtsev, A., Lahr, D.J.G., & Brown, M.W. (2017). Between

a pod and a hard test: the deep evolution of amoebae. *Molecular biology and evolution*, 34(9), 2258-2270.

Kosakyan, A., Gomaa, F., Lara, E., & Lahr, D. J. (2016). Current and future perspectives on the systematics, taxonomy and nomenclature of testate amoebae. *European journal of protistology*, 55, 105-117.

Lahr, D.J.G., Kosakyan, A., Lara, E., Mitchell, E. A. D., Morais, M., Porfirio-Sousa, A. L., Ribeiro, G. M., Tice, A. K., Pánek, T., Kang, S., Matthew W. Phylogenomics and ancestral morphological reconstruction of Arcellinida testate amoebae demonstrate high diversity of microbial eukaryotes in the Neoproterozoic. (Accepted in *Current Biology*).

Mignot, J. P., Raikov, I. B. (1990). New ultrastructural data on the morphogenesis of the test in the testacean *Arcella vulgaris*. *European journal of protistology*, 26(2), 132-141.

Netzel, H. (1971). Die Schalenbildung bei der Thekamöben-Gattung *Arcella* (Rhizopoda, Testacea). *Cytobiologie*, 3, 89-92.

Netzel, H. (1972). Die bildung der Gehäusewand bei der Thekamöbe *Centropyxis discoides* (Rhizopoda, Testacea). *Zeitschrift für Zellforschung und Mikroskopische Anatomie*, 135(1), 45-54.

Netzel, H. (1975a). Die Entstehung der hexagonalen Schalenstruktur bei der Thekamöbe *Arcella vulgaris* var. *multinucleata* (Rhizopoda, Testacea). *Arch. Protistenk*, 117, 321-357.

Netzel, H. (1975b). The origin of the hexagonal shell structure in the testate amoeba *Arcella vulgaris* var. *multinucleata* (Rhizopoda, Testacea). *Arch Protistenk Bd*, 1975b, 117, 321-357.

Netzel, H. (1975c). Struktur und Ultrastruktur von *Arcella vulgaris* var. *multinucleata* (Rhizopoda, Testacea). *Arch. Protistenkd*, 117, 219-245.

Netzel, H. (1980). Amöben als Baumeister. *Biologie in unserer Zeit*, 10(6), 183-190.

Netzel, H., & Grunewald, B. (1977). Morphogenesis in shelled rhizopod *Arcella dentata*. *Protistologica*, 13(2), 299-319.

Nikolaev, S. I., Mitchell, E. A., Petrov, N. B., Berney, C., Fahrni, J., & Pawlowski, J. (2005). The testate lobose amoebae (order Arcellinida Kent, 1880) finally find their home within Amoebozoa. *Protist*, 156(2), 191-202.

Page, F. C. (1976). A revised classification of the Gymnamoebia (Protozoa: Sarcodina). *Zoological Journal of the Linnean Society*, 58(1), 61-77.