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: transcriptomica *single-cell* de *Arcella intermedia* e caracterização da expressão gênica durante a formação de teca.

São Paulo

2018

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: transcriptomica *single-cell* de *Arcella intermedia* e caracterização da expressão gênica durante a formação de teca.

Versão original

Dissertação apresentada ao Instituto de Biociências da Universidade de São Paulo para obtenção do título de Mestre em Zoologia pelo Programa de Pós-graduação em Zoologia.

Supervisor: Prof. Dr. Daniel J.G. Lahr

São Paulo

2018

Ficha catalográfica

Comissão julgadora:

_____

Prof. Dr. _____

_____

Prof. Dr. _____

_____

Prof. Dr. _____

Orientador

À Tainá e minha família.

"A alegria é a passagem para um estado mais potente do próprio ser."

Baruch Espinoza

Acknowledgements

Agradeço à Tainá Figueiredo, minha companheira, por todo o apoio, companheirismo e existência. À minha família, meus pais (Maria Alice de Souza e Marco Antônio Porfírio de Sousa), minha irmã (Isabelle Porfírio) e meu irmão (Charles Marlon), meus sogros (Isabel Figueiredo e Roberto Silva) e meu cunhado (Ian da Silva), por todo o apoio e envolvimento. Ao Professor Daniel J. G. Lahr, mais que orientador, um exemplo para se espelhar e aprender, não só como cientista mas também como ser humano. Aos membros de laboratório e amigos, Tainá Figueiredo, Giulia Magri, Paulo Hofstatter, Samuel Júnior, Daniel Máximo, Bruna Trevisan, por todo o apoio. Ao Professor Matthew Brown, por ter me acolhido durante quatro meses em seu laboratório e contribuído significativamente com este trabalho e minha formação (Thank you very much Matt!). A Alex Tice, pelo companheirismo e contribuição significativa para este trabalho e minha formação (Thank you very much Alex!), assim como aos outros membros do Laboratório do Professor Matthew (Gabriel Schuler, Christopher Lebeau, Robert Jones e Seungho Kang), com os quais convivi durante minha estadia na Mississippi State University (Thank you guys!). Aos técnicos Manuel Antunes Júnior e Beatriz Vieira Freire, por todo o apoio no Laboratório Molecular. Ao Sr. Wilson, pelos desenhos produzidos para este trabalho. A todos os meus amigos. A todos que contribuíram para a minha formação e o desenvolvimento do presente trabalho. O projeto desta dissertação de mestrado foi financiado pela FAPESP, por meio de dois processos: Mestrado / Fluxo Contínuo (Processo número: 2016/14317-4) e BEPE - Mestrado / Fluxo Contínuo (Processo número: 2017/19388-0).

Contents

List of Figures

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 (Chromoalveolata) (**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*, *Dictyostelium 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

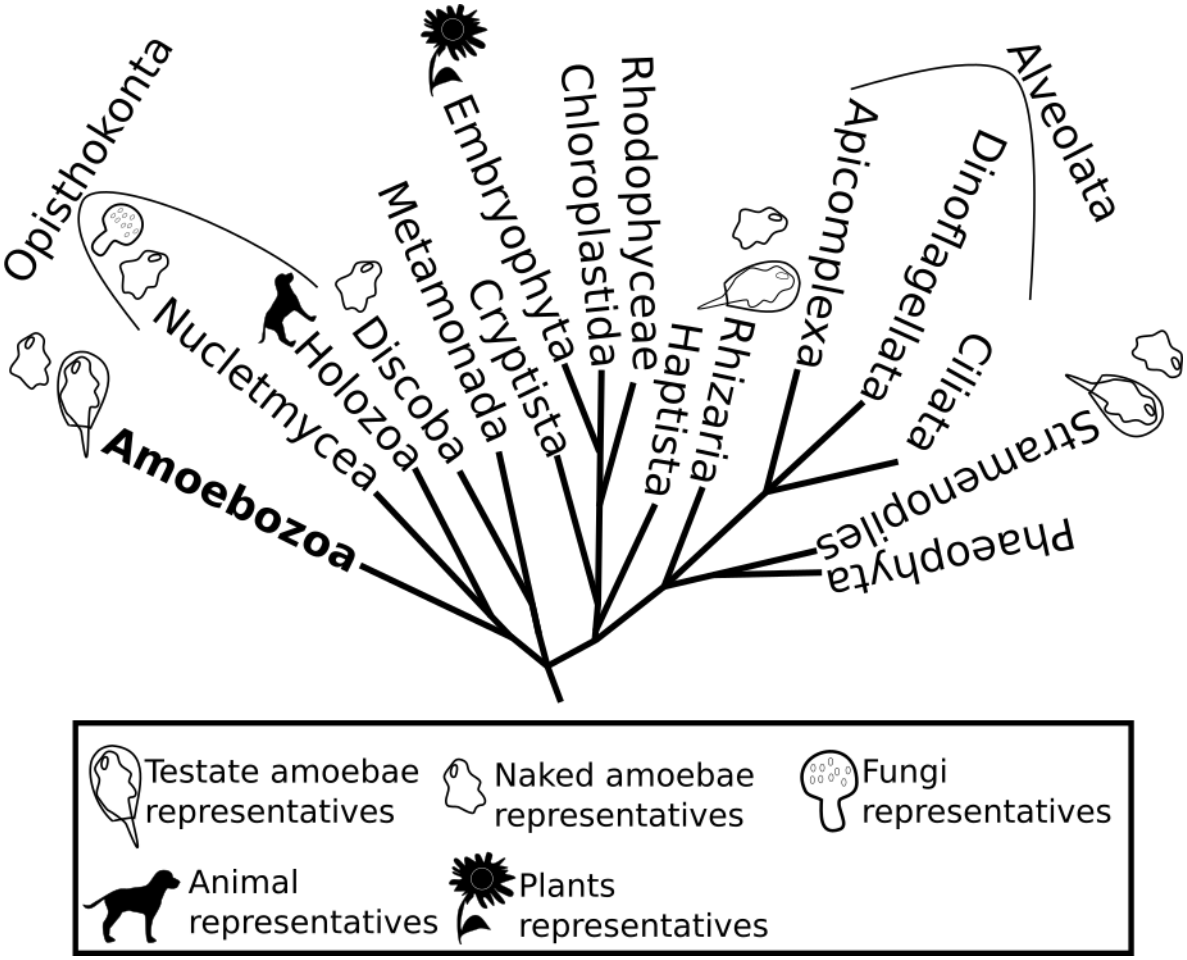

Figure 1.0.1 – Schematic overview of the diversity of eukaryotes based on adl et al. (2018). We highlight the lineages with naked amoebae representatives and testate amoebae representatives, as well as, the groups that comprise Fungi, Animals, and Plants. Amoebozoa, our focus group, is sister group to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

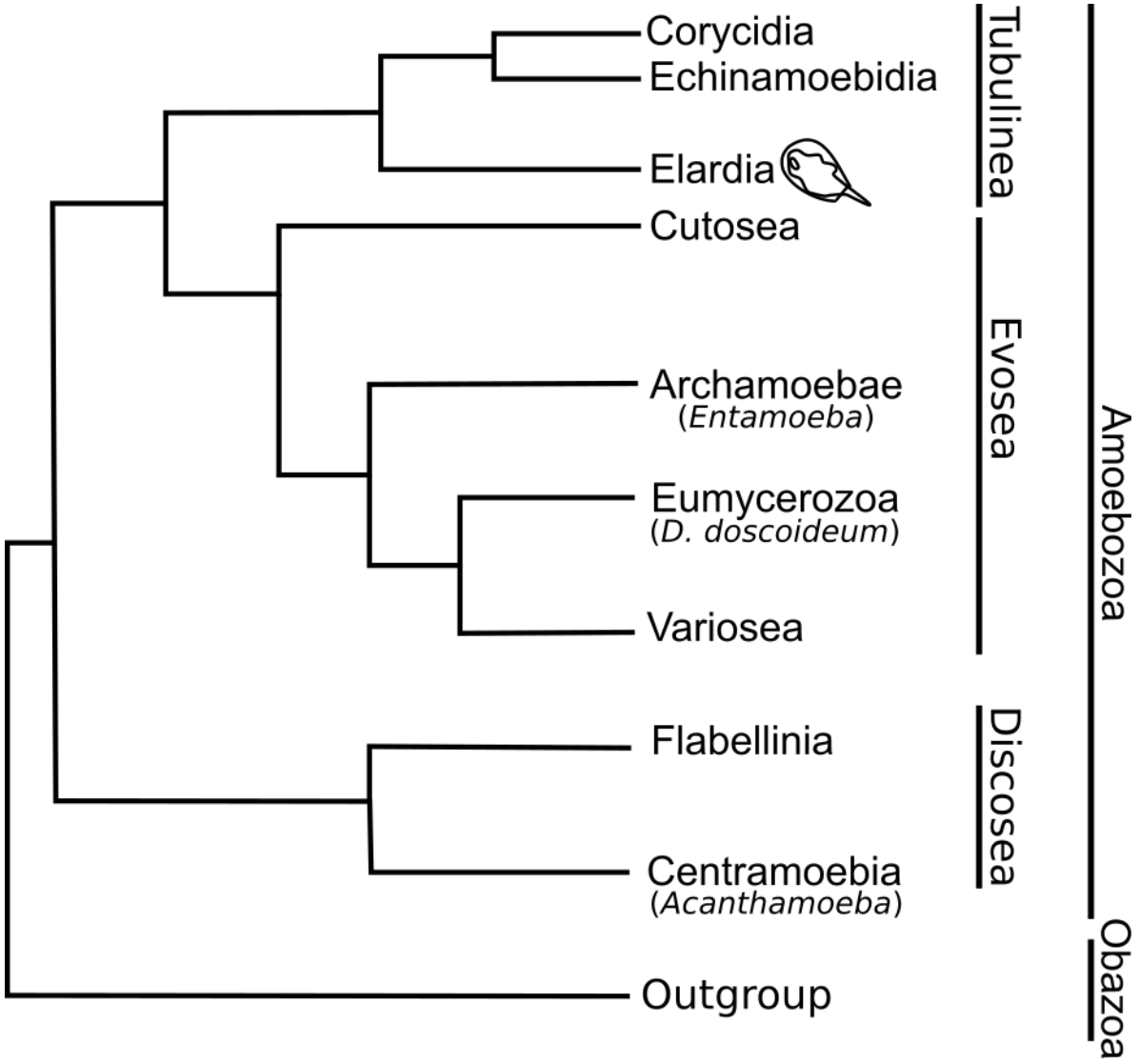

Figure 1.0.2 – Schematic Amoebozoa tree based on the phylogenomic study Kang et al. (2017). Currently, Amoebozoa are classified in tree major lineages (Tubulinea, Evosea and Discosea). Arcellinida are a monophyletic lineage of Elardia.

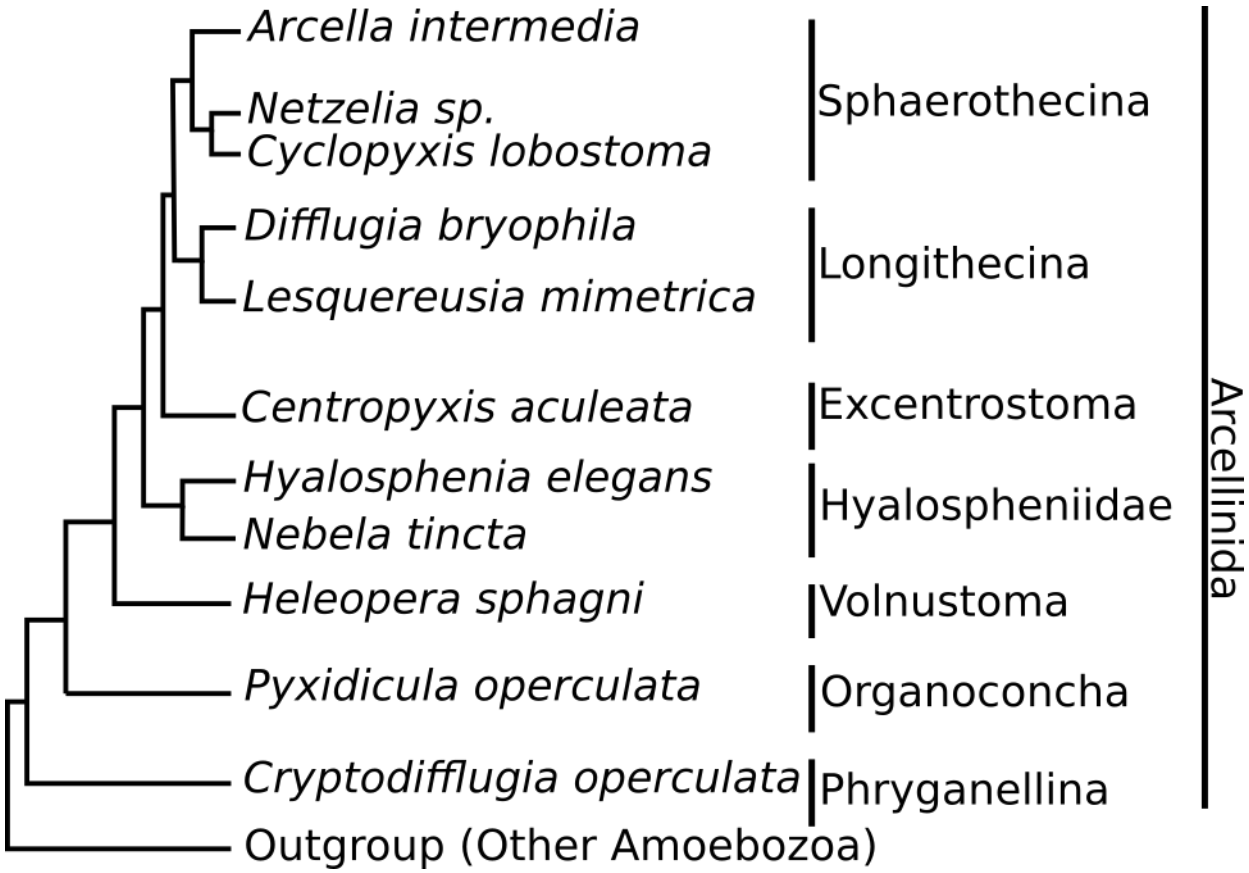

Figure 1.0.3 – Schematic Arcellinida tree based on the phylogenomic Lahr et al. (accepted). Currently, Arcellinida monophyly is recovered and classified in five Infraorders (Sphaerothecina, Longithecina, Excentrostoma, Hyalospheniidae, and Volnustoma) and two Suborders (Organoconcha and Phryganeliina), 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**.

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, Testaeea). 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.

2 Chapter 1: Shell structure and thecagenesis in the genus *Arcella* (Arcellinida:Amoebozoa): A literature overview

2.1 Abstract

Arcellinida is an Amoebozoa lineage characterized by the presence of a shell. A comprehensive cyto-morphological description of the shell structure and formation (thecagenesis) process has been generated in the span of more than a century of literature. Recently, few works have advanced our understanding on the thecagenesis process and none has discussed this rich literature; although, several aspects of the shell formation and structure are still a puzzle, as the relation between genes and shell formation morphogenesis. Here we present an overview of the literature about shell formation and structure, focusing in the genus *Arcella*, the most studied lineage. We show that some issues, raised by this literature, are still open questions. Moreover, based on the current knowledge, we are able to establish new frameworks, aware of the state of the art on this field, to address issues such as the origin and evolution of the shell in *Arcellinida*.

2.2 Introduction

The arcellinid shell (test or carapace) is a hardened outer structure, which almost completely covers these single-cell organisms, with the exception of a single opening called "aperture". The shell is classified in four main types, given its chemical composition (Mitchell et al., 2008): i. shells made of proteinaceous material; ii. shells made of secreted plates (idiosomes) of organic material; iii. shells made of secreted idiosomes of calcium carbonate or biosilica (Yamaoka and Mizuhira, 1987; Aoki et al., 2007); iv. agglutinated shells made of mineral or organic material (xenosomes) recycled from the environment (Châtelet et al., 2015).

Shells are present in diverse lineages, not exclusively in Arcellinida. Together, the shelled organisms comprise the polyphyletic group of "testate amoebae" (Adl et al., 2012, 2002, Meisterfeld, 2002a). Arcellinida is the most diverse group of testate amoebae (currently with around 687 nominal species), and present a rich literature (Kosakyan et al., 2016). The structure and formation process (thecagenesis) of the shell remained a puzzle for many years and intrigued generations of researchers (Moraczewski, 1971; Netzel, 1971; Raikov and Mignot, 1990). Diverse studies present a comprehensive analysis and interpretation of Arcellinida shell, primarily using the genus *Arcella* (Moraczewski, 1971). Possibly, the focus on this genus is due to its abundance in diverse freshwater environments, the capability to easily grow on laboratory cultures, and a literature bias; Together these characteristics define *Arcella* as an interesting model organism of Arcellinida.

This carapace was identified for the first time by Ehrenberg (1832), in the original description of *Arcella vulgaris*. Since this initial observation, and the first discussion of its structure, in 1864 by Wallich (Moraczewski, 1971), the formation process and structure of the shell on *Arcella* received special attention by several authors in the span of more than a century. In the early 1970s, Moraczewski (1971) presented an overview of the literature comprising studies from 1832 to 1967. During this period, a total of 18 studies dealt with this issue, analyzing shell structures and the formation process under light microscope and electron microscopy (Moraczewski, 1971). From 1971 to 1990, nine studies advanced our understanding of shell formation in the genus *Arcella*. Recently, Pchelin (2010) and, Volkova and Alexey (2016) discussed shell formation during the test regeneration on *Arcella vulgaris*. Currently, it is known that *Arcella* has a proteinaceous shell, although the nature of this material is not known.

Here we present a timeline of the literature that discusses shell formation in *Arcella*. We briefly present the studies from 1832 to 1867, discussed by Moraczewski

(see Moraczewski (1971) for the original discussion), and present an overview of the studies between 1971 and 2016. Rather than an exhaustive list and review of articles from the literature regarding shell formation, we present a general overview of the many studies that elucidated the thecagenesis process in the genus *Arcella* as we understand it now. We summarize the current understanding of shell formation and structure on *Arcella*, based on a historical perspective. Finally, we present our perspective of the thecagenesis process and propose a framework that would advance our knowledge about *Arcella*'s shell.

2.3 Literature timeline

To date, 29 studies comprise the key literature that discusses the shell structure and formation in *Arcella* (**Figure 2.3.1**). From 1864 to 1928 researchers focused on the description of the general structure of the shell, enabled by light microscopy; several interpretations of the shell formation process were derived from these descriptions (Moraczewski, 1971). This period was followed by 40 years, between 1928 and 1963, of literature focusing on the ecology of these organisms, but not the nature of their carapace (Moraczewski, 1971). Cambar et al. (1963) renewed the discussion of this issue and introduced the use of the electron microscopy (Moraczewski, 1971), describing the external structure of the *Arcella* shell (Cambar et al., 1963). On the following studies, researchers applied both light and scanning electron microscopy to shed light on the shell formation and structure. Charret and Vivier (1964), and Charret (1967) presented the first micrographs of tangential sections of the shell under transmission electron microscope (Charret and Vivier, 1964; Charret, 1967). Moraczewski (1971) published a study describing direct evidence of the process of shell formation; he presented several possible interpretations of the shell formation process, based on his observations of tan-

gential sections and the previous literature. The period from 1971 to 1990 was a fertile period for the description of the thecagenesis process and shell structure as we understand now. Netzel, a remarkable contributor in Arcellinida shell studies, published seven articles describing the carapace and formation process in different species of the genus *Arcella* (Netzel, 1971;; Netzel, 1972; Netzel,1975a; Netzel, 1975b; Netzel, 1975c; Netzel and Grunewald, 1977; Netzel, 1980). In 1990, Mignot and Raikov contributed to the description established by Netzel and collaborators; they presented additional observations and interpretations based on several tangential sections of *Arcella vulgaris* during shell formation, which together represent the complete process (Mignot and Raikov, 1990). It was only recently, 20 years after Mignot and Raikov's study, that two studies dealt with the shell formation process; Pchelin (2010) and, Volkova and Alexey (2016) studied shell formation during test regeneration on *Arcella vulgaris*. Both studies demonstrated that Arcella is capable of regenerating its carapace, once deprived of this structure by *in vivo* dissection (Pchelin, 2010; Volkova and Alexey, 2016). Currently, no study discusses the possible molecular process and genes underlying shell formation, a topic already pointed out by Netzel and Grunewald (1977), more than 40 years ago.

2.3.1 Description of shell formation and structure in *Arcella*

Arcella presents a general mechanism of thecagenesis: arrangement and secretion (extrusion) of thechagenous granules, derived from the Golgi, cellularly controlled; this process was comprehensively described based on tangential sections of *Arcella*'s cells (Netzel, 1971; Netzel, 1975a; Netzel, 1975b; Netzel and Grunewald, 1977; Mignot and Raikov 1990). The first evidence of shell formation is the appearance and concentration of Golgi derived thecagenous granules in the cytoplasm (**Figure 2.3.2**) (Netzel, 1971; Netzel, 1975a; Netzel, 1975b; Netzel and Grunewald,

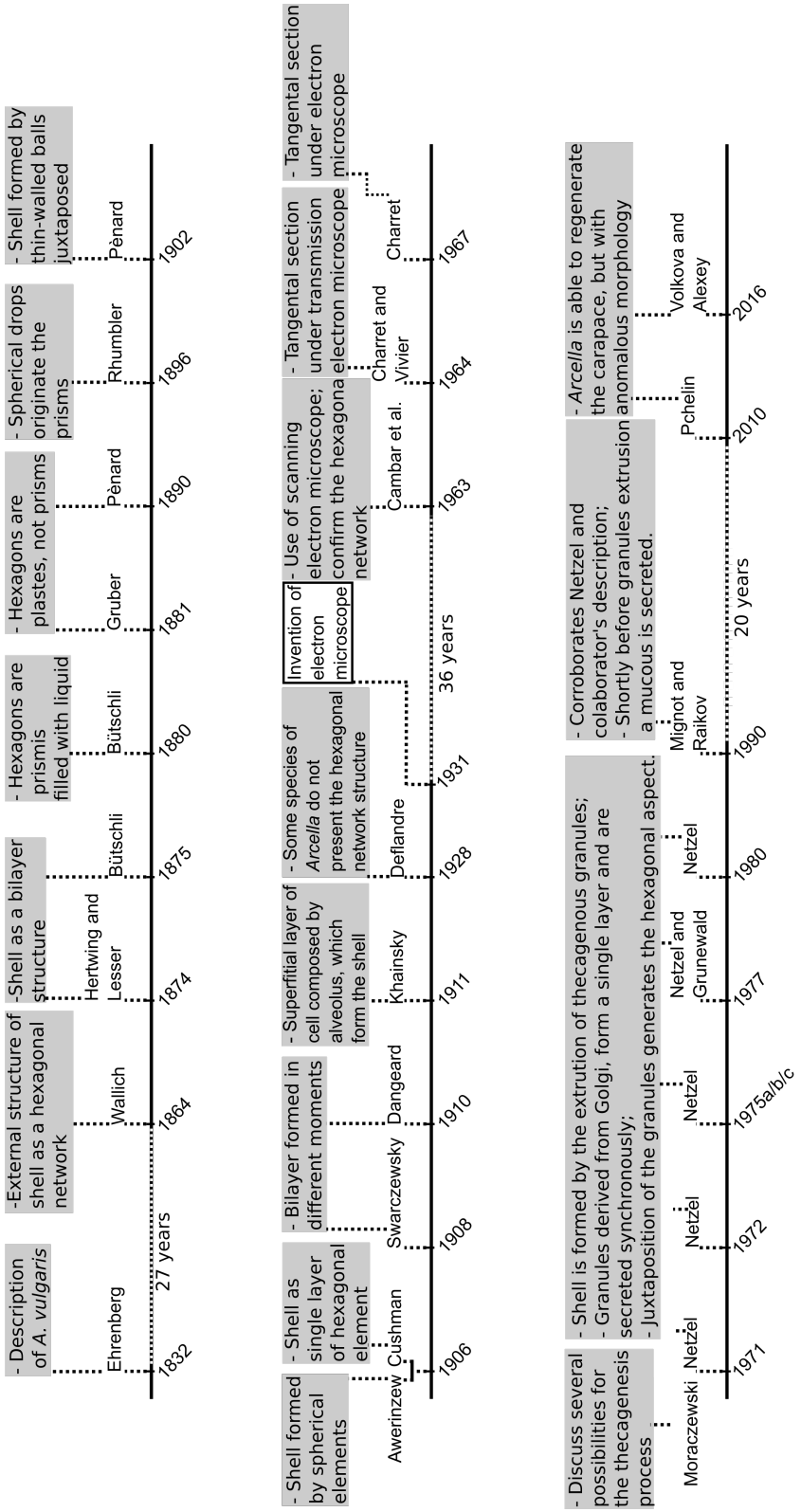

Figure 2.3.1 – Timeline with an overview of the literature that discusses the shell structure and thecagenesis process in the genus *Arcella*. Time scale on years; dashed time scale lines represent long periods lacking studies about shell structure and formation. Gray boxes present some general ideas and contributions presented by each author, which are relevant to the present study.

1977; Mignot and Raikov 1990). The granules concentrate near the cell membrane in a cytoplasmic region close to the shell aperture; this region is the site formation for the cell division bud (Netzel and Grunewald, 1977), named as thecagenous bud by Mignot and Raikov (1990), which grows through the shell aperture (**Figure 2.3.2 and Figure 2.3.3**); according to Netzel (1971, 1975, and 1977), the granules become arranged in a single layer and are synchronously secreted during the budding division, forming an amorphous structure in the edge of the cell membrane (**Figure 2.3.4 A**); shortly a defined shape of the granules structures is observed (**Figure 2.3.4 B**). The granules extrusion is followed by an abrupt cell division bud growth and cytoplasmic flux from the old shell towards the new (daughter) shell, the shell is externalized and involved by a special type of pseudopod, the pseudopodial dome (**Figure 2.3.5**) (Netzel, 1971; Mignot and Raikov, 1990). This cytoplasmic flux determines the general size and morphology of the daughter shell (**Figure 2.3.5**). Once this shape is determined, cytoplasmic flux between both shells does not modify the shell morphology; this cytoplasmic flux is followed by cytokinesis, the final stage of *Arcella* reproduction (**Figure 2.3.6**).

In vivo, under light microscope, the first evidence of shell formation is the appearance of the cell division bud (**Figure 2.3.7 A-B**). This bud grows abruptly and the shell becomes visible (**Figure 2.3.7 C**); a cytoplasmic flux towards the daughter shell determines its shape (**Figure 2.3.7 D**). Briefly, it is possible to identify an abrupt retraction of the pseudopodial dome (**Figure 2.3.7 D**). An intense cytoplasmic flux is observed between both cells; no modification of the daughter shell morphology is observed (**Figure 2.3.7 E**), and finally, the cells divide by cytokinesis (**Figure 2.3.7 F-G**). Two separated individuals are observed (**Figure 2.3.7 H-I**).

The structure of the shell is determined by the intrinsic characteristics of the thecagenous granules and the shell formation process. Externally, the shell structure

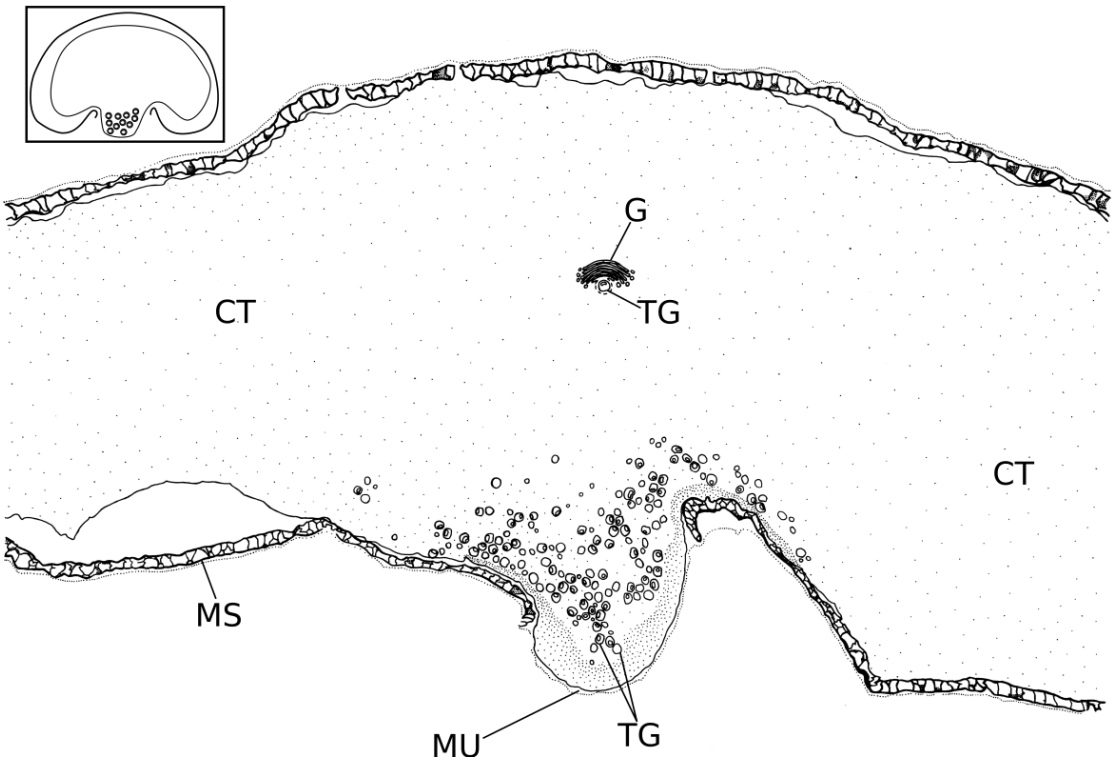

Figure 2.3.2 – Reproduction of a tangential section of *Arcella vulgaris* during the formation of the thecagenous bud (TB), based on the Figure 4 from Mignot and Raikov (1990); Golgi representation was based on Netzel and Grunewald (1977). The scheme shows the concentration of the thecagenous granules (TG) on the growing bud, near the shell (SH) aperture; these granules are derive from Golgi (G). Box present a schematic representation of the tangential section orientation; organism seen in side view. CT- Cytoplasm; MS - Mother shell; MU- Mucus-like product, secreted shortly before TG secretion.

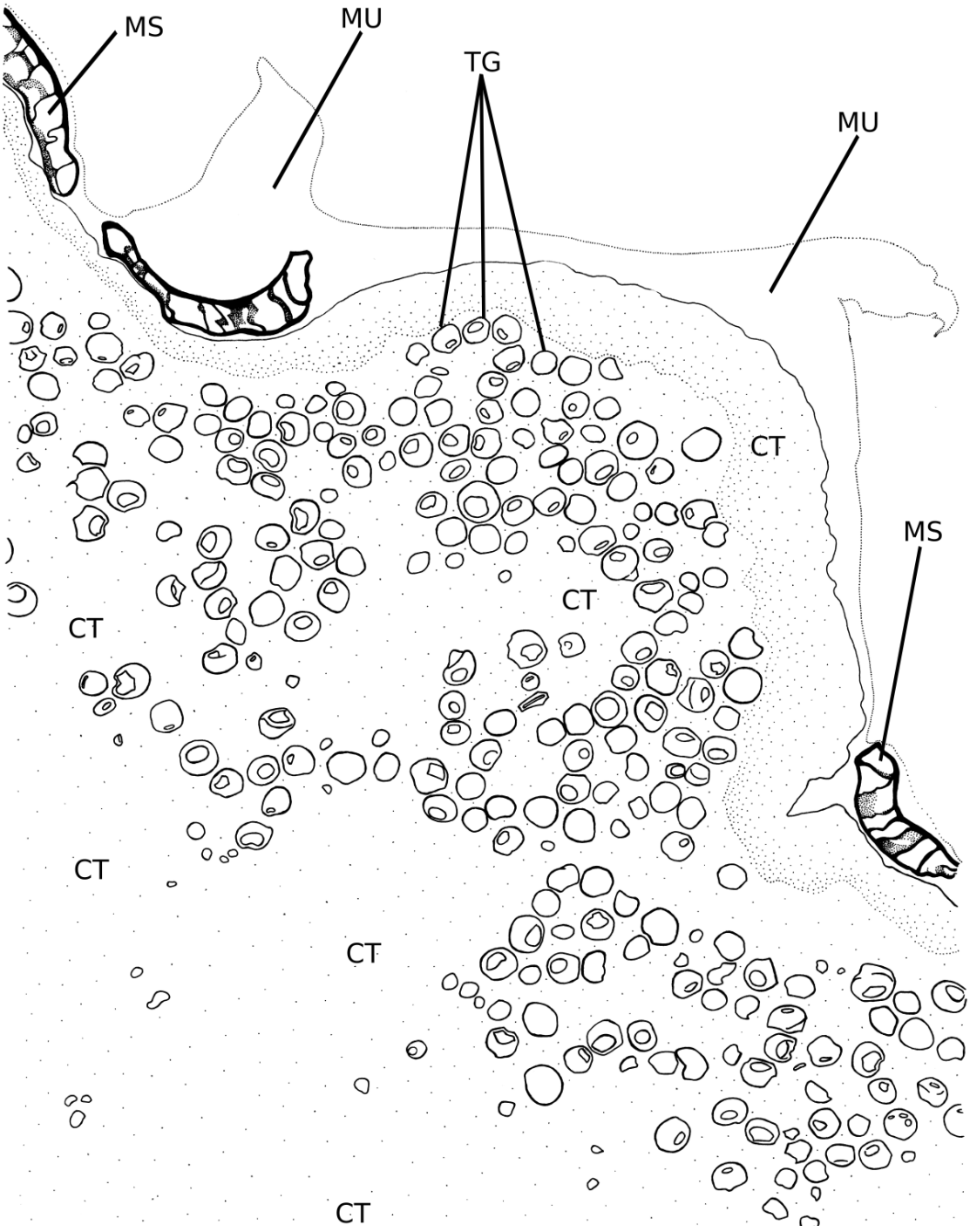

Figure 2.3.3 – Reproduction of a tangential section of *Arcella vulgaris* during the formation of the thecagenous bud (TB), based on Figure 5 from Mignot and Raikov (1990). The representation shows the detail of the bud formation and the mucus-like product (MU), secreted shortly before the extrusion of thecagenous granules (TG); the mucus composition is not known. MS - Mother shell.

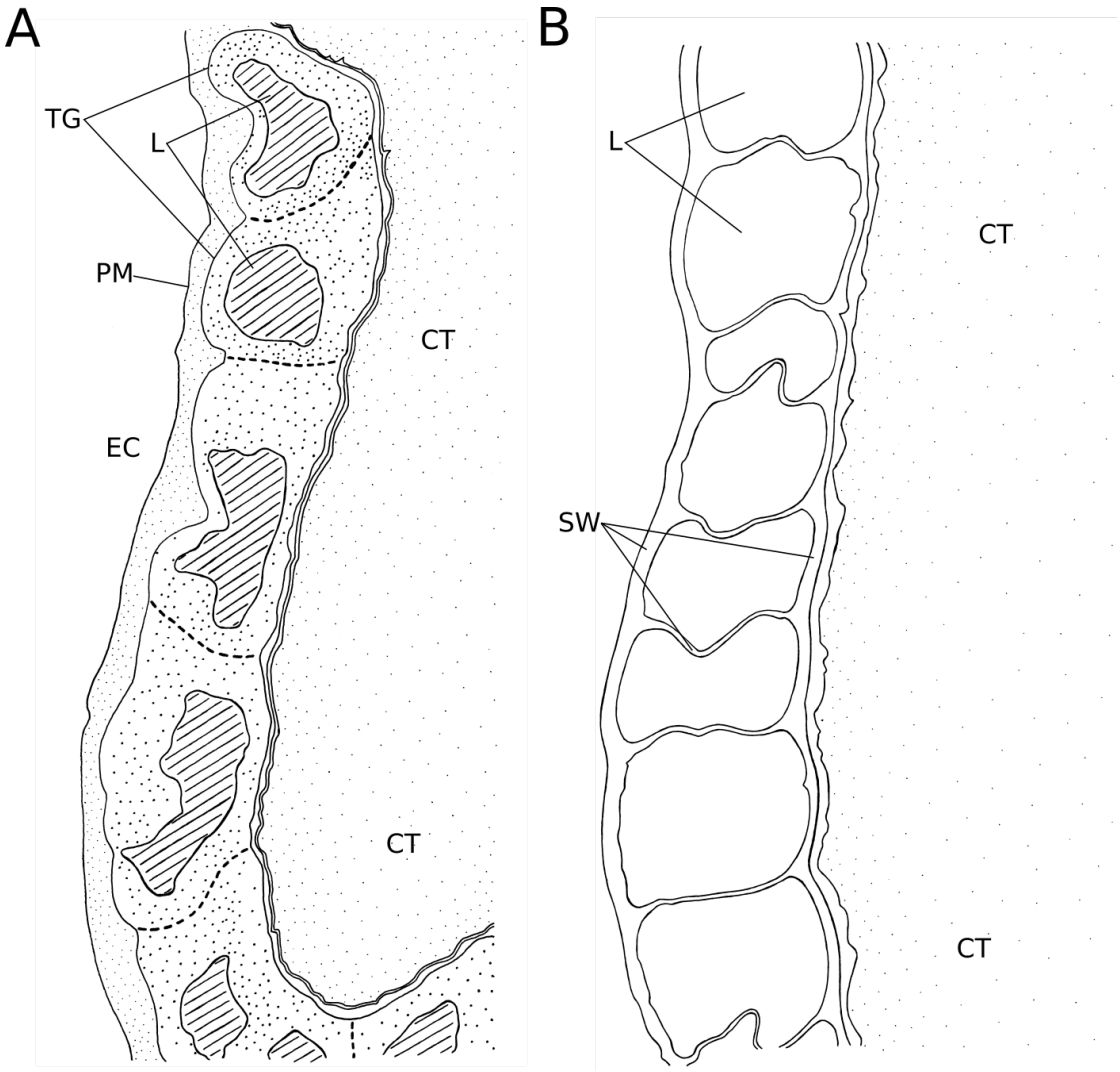

Figure 2.3.4 – Reproduction of a tangential section of *Arcella dentata*. **A**. Shows the thecagenous granules (TG) organized on a single layer in the edge of the Plasma Membrane (PM), in a late stage of thecagenesis, based on Figure 25B from Netzel and Grunewald (1977). The thecagenous granules (TG) present a prominent alveolar lumen (L). **B**. Shows the thecagenous material externalized and with defined structure. The alveolar lumen is maintained after the extrusion of the thecagenous material. CT - Cytoplasm; SW - Shell walls; EC - Extracellular region.

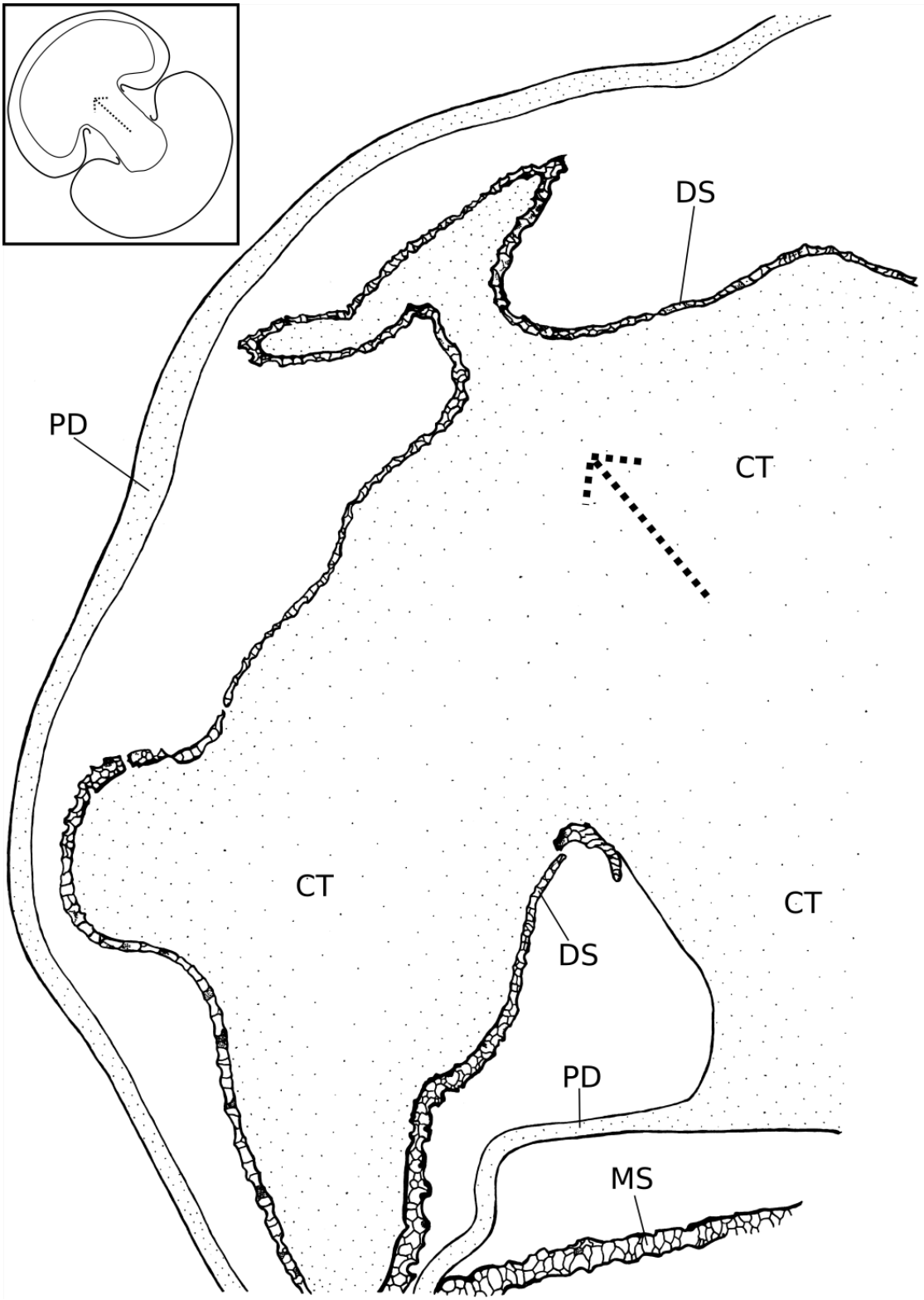

Figure 2.3.5 – Schematic reproduction of a tangential section of *Arcella vulgaris* during cytoplasmic flux that shapes the forming shell (Daughter shell-DS). The scheme shows the pseudopodial dome (PD), which covers the forming shell (DS). Box present a schematic representation of the tangential section orientation. MS - Mother shell; CT - Cytoplasm.

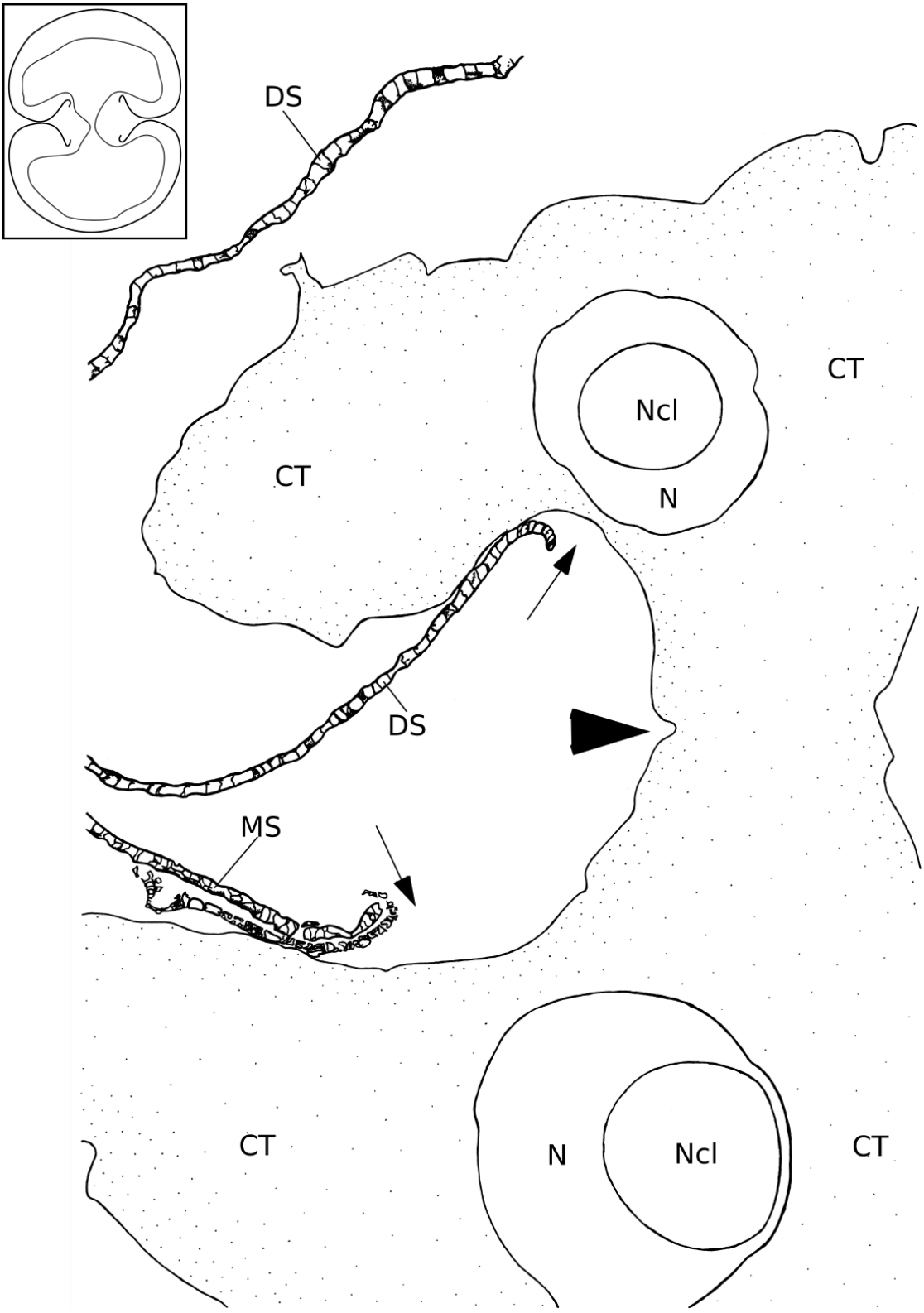

Figure 2.3.6 – Schematic reproduction of a tangential section of *Arcella vulgaris* during cytokinesis, the last stage of reproduction. The figure shows the two shells, the mother and daughter shells (MS and DS, respectively), and the cytokinesis site (arrow head). Box present a schematic representation of the tangential section orientation. N - Nucleus; NCL - Nucleolus; CT - Cytoplasm. The arrows shows the apperture region.

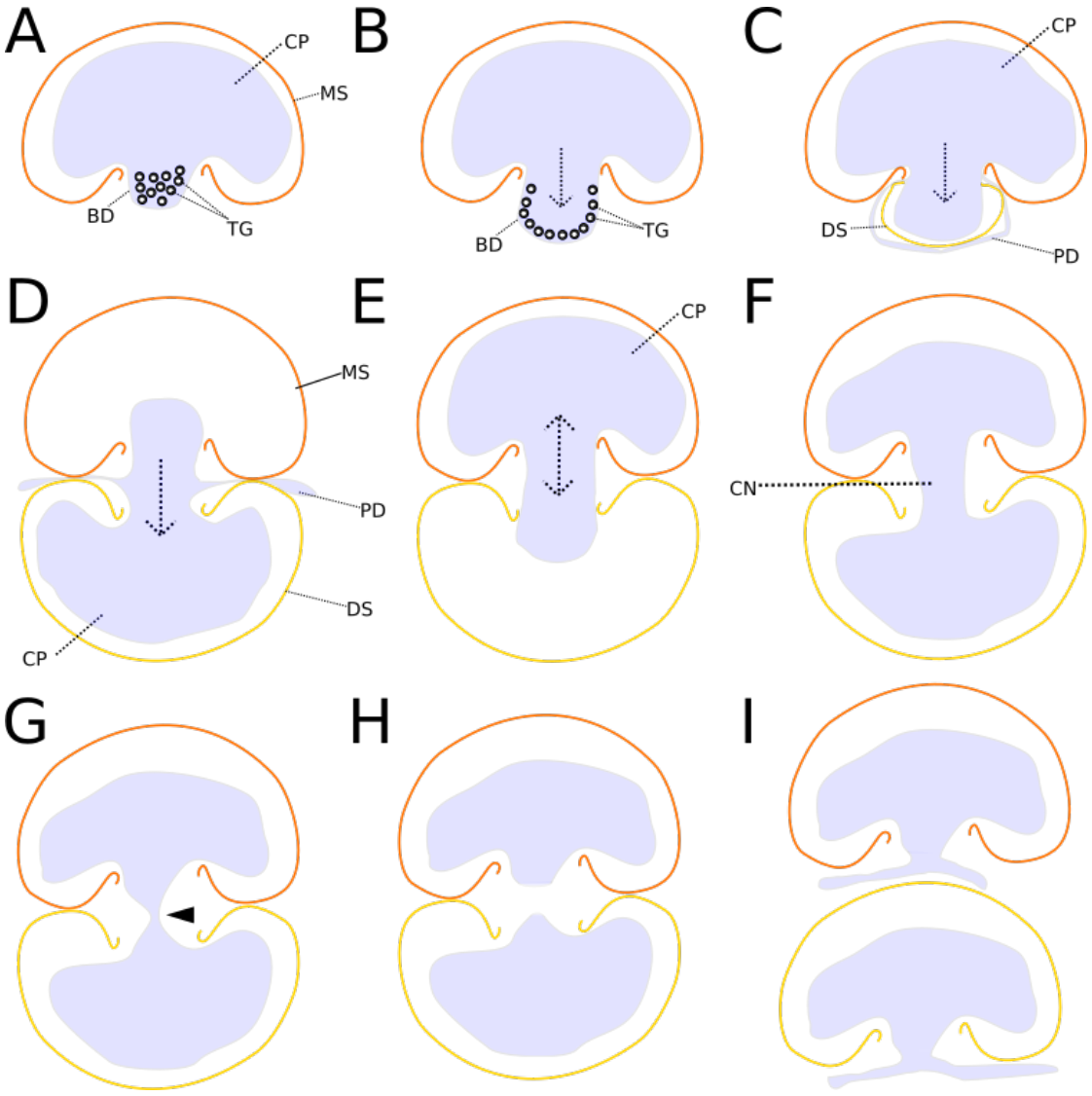

Figure 2.3.7 – Representation of shell formation and reproduction by budding cell division in *Arcella*, observed under the light microscope. **A.** Bud division (BD) formation and concentration of thecagenous granules in the edge of the plasma membrane (PM), near to the regions of the mother shell (MS - orangeish line) aperture. Cytoplasm (light blue) and thecagenous granules (TG) shown. TG are not observable under light microscope, but are shown based on the interpretations of tangential sections. **B.** Division bud growth and thecagenous granules arranged on a single layer near the PM. Dashed arrow represents the direction of bud growth. **C.** The daughter shell (DS - yellowish line) is observed externally. Dashed arrow represents the direction of bud growth and cytoplasmic flux. **D.** An abrupt growth of the division bud and cytoplasmic flux determines the general shape of the DS. Dashed arrow represent the direction of bud growth and cytoplasmic flux. **E.** Intense cytoplasmic flux between both shells is observed. Dashed arrow represent the direction of cytoplasmic flux. **F.** Cytoplasm equally distributed between the shells, with a cytokinesis neck (CN) present. **G.** Cell cytokinesis (arrow head). **H - I.** Two individuals completely formed.

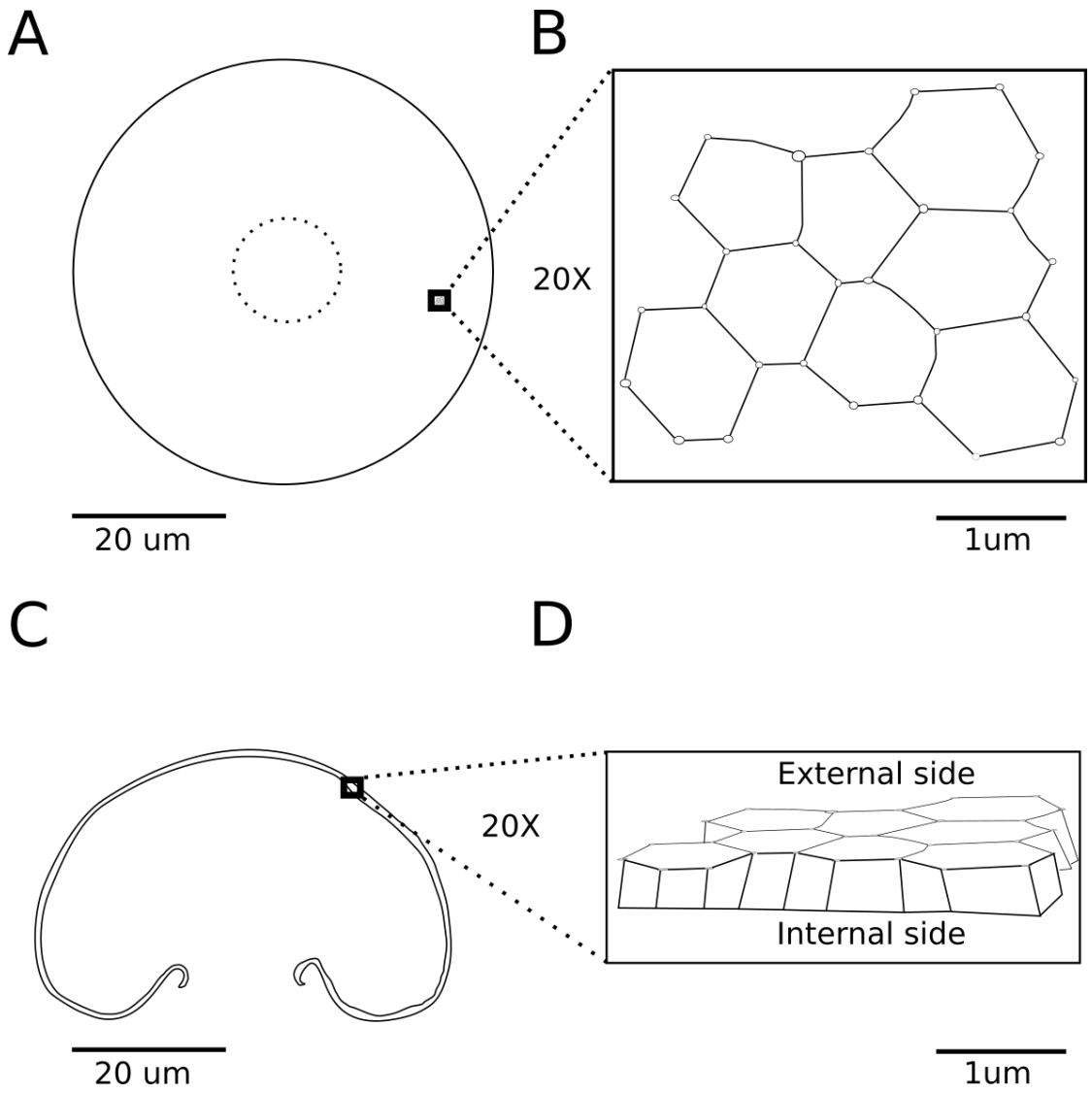

Figure 2.3.8 – Schematic representation of general *Arcella* shell structure, based on scanning electron microscope micographs of *Arcella hemisphaerica* from Lahr and Lopes (2009). **A.** Representation of intact shell in the dorsal view. **B.** Representation of the external hexagonal (honeycomb-like) shell structure of *Arcella*. **C.** Representation of a tangential section of *Arcella* shell in a lateral view. **D.** Representation of the prismatic structure of the shell, based on tangential sections of the shell.

is observed as a regular hexagonal network (or honeycomb-like structure) (**Figure 2.3.8 A**), first noted by Wallich (1864); this observation was consistently corroborated by other authors (Pènard, 1902; Awerinzew, 1906; Cushman and Henderson, 1906; Deflandre, 1928; Cambar, 1963; Charret and Vivier, 1964; Moraczewski, 1971; Netzel, 1971; Netzel, 1975a; Netzel, 1975b; Netzel and Grunewald, 1977; Mignot and Raikov 1990). According to Rhumbler (1896), the external structure observed of the shell was derived from the nature of its building blocks, spherical drops (Phâasomen); similarly, other authors proposed that the building blocks should be thin-walled balls (Pènard, 1902) or spherical elements (Awerinzew, 1906), which juxtaposed would generate a hexagonal shape (Pènard, 1902; Awerinzew, 1906). Moraczewski (1971), proposed that circular alveolus (vesicles) derived from the ER were the units involved in the origin of a honeycomb structure. Netzel (1975a and 1975b), identified and proposed the participation of vesicles (thecagenous granules) as the building blocks of the shell; these thecagenous granules are derived from Golgi and were identified to be directly involved on shell formation (Netzel, 1975a; Netzel, 1975b; Mignot and Raikov, 1990). Finally, it is the extrusion and juxtaposition of these granules, which contain the shell material, that generate the hexagonal network aspect of the shell (Netzel, 1975a; Netzel, 1975b; Netzel, 1977; Mignot and Raikov, 1990). The composition of the proteinaceous material that compose the shell is not known. Tangential section demonstrates that the shell is formed as a bi-layer structure (**Figure 2.3.8 B**). Several authors described this structure (Hertwing and Lesser, 1874; Bütschli, 1875; Pènard, 1902; Awerinzew, 1906; Swarczewsky, 1908; Dangeard, 1910); initially, was thought that each layer was deposited in different moments (Swarczewsky, 1908; Dangeard, 1910). Netzel (1975), and after corroborated by Mignot and Raikov (1990), described that the thecagenous granules present an alveolar inner region; Netzel (1975) demonstrated

that this alveolar inner region is maintained through the thecagenesis process and is present in the mature shell (as shown on **Figure 2.3.4**).

2.4 Final considerations

The literature discussing the thecagenesis process and structure of the shell in *Arcella* developed in the span of more than a century, with some fertile periods being followed by less productive ones. Currently, we have a comprehensive description on the literature; this description is based on several interpretations and hypothesis proposed by generations of researchers, most of them not corroborated nowadays, but which led to new studies and findings. The use of new technologies, as electron microscope, and different techniques, as tangential sections, had special impact on our current understanding. Netzel, from 1971 to 1980, and, Mignot and Raikov (1990), present a complete description of the thecagenesis process and shell structure in the genus *Arcella*; these authors established such complete description by considering the previous literature, generating new data and, deeply describing the cyto-morphology of this process.

Finally, Netzel and Grunewald stated, in 1977, something still valid now a days, 41 years later: "... one must admit, that in thecamoebae these relations [gene products with morphogenesis] are far from clear". Basically, the thecagenesis process is based on: i. secretory granules, the thecagenous granules, which contain the material that forms the shell; ii. morphogenesis of the shell by cell division bud formation and growth, and cytoplasmic flux, once the granules are secreted (extruded). Thus, we propose that the thecagenesis process in *Arcella* involves two basic cell processes present in other eukaryotes, biogenesis and secretion of secretory granules, and cell budding division. The first, involves different sequential steps, protein synthesis, vesicle transport from Endoplasmic Reticulum to Golgi protein

transport, protein sorting and packing from Golgi to cell membrane through secretory vesicle (Tooze and Stinchcombe, 1992; Vázquez-Martínez, 2011); the second, while being involved on cell division (reproduction) of this organisms, a typical budding cell division (Netzel and Grunewald, 1977), also functions as the driving force of morphogenesis of the shell in *Arcella* (Netzel and Grunewald, 1977). Coupling the available cyto-morphological knowledge and newly generated molecular data of *Arcella* during shell formation, represents a promising framework to be established and shed light on what Deflandre already mentioned, in 1953, as "les facultés morphogénétiques de la cellule ... exteriorisèes" during shell formation. We present and develop these ideas in the following chapters.

2.5 ***Reference***

Adl, S. M., Simpson, A. G. B., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S. et al (2012). The revised classification of eukaryotes. Journal of Eukaryotic Microbiology, 59: 429–514, 2012.

Aoki, Y., Hoshino, M., & Matsubara, T. (2007). Silica and testate amoebae in a soil under pine–oak forest. Geoderma, 142(1-2), 29-35.

Awerinzew, S. (1907). Die Struktur und die chemische Zusammensetzung der Gehäuse bei den Süßwasserrhizopoden. Arch. Protistenk, 8, 91-111.

Bütschli, O. (1875). Zur Kenntniss der Fortpflanzung bei Arcella vulgaris Ehrb. Archiv für mikroskopische Anatomie, 11(1), 459-467.

Bütschli, O. (1880). Protozoa, Bronns Klassen und Ordnungen im Tierreich. CF Winter, Leipzig.

Cambar R., Thomas R. et Le Blanc M. (1963). Recherches sur la constitution de la thèque des Arcelles (Genre *Arcella*, Rhizopode testacé), observations au microscope électronique. C. R. Acad. Sci., 256., 1364-1367.

Charret, R. (1967). Caractères cytologiques du Thécamoebien *Arcella polypora*. Protistologica, 3(1), 73-78.

Charret, R. & Vivier, E. (1964)Ultrastructure de *Arcella polypora* Rhizopoda Testacea. III Europ. reg. conf. El. Micr. vol. B, 187-188..

Cushman, J. A., & Henderson, W. P. (1906). A Preliminary Study of the Finer Structure of Arcella. The American Naturalist, 40(479), 797-802.

Dangeard, P. A. (1910). Etudes sur le developpement et la structure des organismes inferieurs. I. Les Amibes. Le Botaniste, 11, 4-57.

Deflandre, G. (1928). Le genre Arcella Ehrenberg. Archiv für Protistenkunde, 64(1), 152-287.

Deflandre, G. (1953). Ordres des Testaceolobosa (de Saedeleer, 1934), Testaceafilosa (de Saedeleer, 1934), Thalamia (Haeckel, 1862) ou Thecamoebiens (Auct.)(Rhizopoda Testacea). Traité de Zoologie, 1, 97-148.

Du Chatelet, E. A., Bernard, N., Delaine, M., Potdevin, J. L., & Gilbert, D. (2015). The mineral composition of the tests of 'testate amoebae'(Amoebozoa, Arcellinida): The relative importance of grain availability and grain selection. Revue de micropaleontologie, 58(3), 141-154.

Gruber, A. (1881). Die Theilung der monothalamen Rhizopoden.

Hertwig R. & Lesser E. (1874). Ober Rhizopoden und denselben nahstehenden Organismen, Arch. mikrosk, Anat Entw. Mech., 10 supl. 110-125.

Khainsky, A. (1911). Untersuchungen über Arcellen. Arch. Protistenk, 21, 165-185.

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., & Lopes, S. G. B. C. (2009). Evaluating the taxonomic identity in four species of the lobose testate amoebae genus Arcella Ehrenberg, 1832. Acta Protozoologica, 48(2), 127-142.

Meisterfeld, R. (2002). Order arcellinida kent, 1880. The illustrated guide to the Protozoa, 2, 827-860.

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.

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.

Mitchell, E. A., Payne, R. J., & Lamentowicz, M. (2008). Potential implications of differential preservation of testate amoeba shells for paleoenvironmental reconstruction in peatlands. Journal of Paleolimnology, 40(2), 603-618.

Moraczewski, J. (1971). La composition chimique de la coque d'Arcella discoides Ehrbg. Acta Protozoologica, 8, 423-437.

Moraczewski, J. (1971). Structure et formation de la coque d'Arcella. Acta protozool, 8(27-33), 423-438.

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

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, Testaeea). 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.

Pchelin, I. M. (2010). Testate amoeba Arcella vulgaris (Amoebozoa, Arcellidae) is able to survive without the shell and construct a new one. Protistology, 6(4), 251-257.

Pchelin, I. M. (2010). Testate amoeba Arcella vulgaris (Amoebozoa, Arcellidae) is able to survive without the shell and construct a new one. Protistology, 6(4), 251-257.

Penard, E. (1902). Faune rhizopodique du bassin du Léman.

Penard, M. E. (1890). Études sur les Rhizopodes d'eau douce. Imprimerie Aubert-Schuchardt.

Rhumbler L. (1896). Beitrfige zur Kenntnis der Rhizopoden. Z. wiss. Zool., 61, 118-126.

Swarczewsky, B. (1908). Über die Fortpflanzungserscheinungen bei Arcella vulgaris. Ehrbg. Arch. Protistenk, 12, 173-212.

Tooze, S. A., & Stinchcombe, J. C. (1992, October). Biogenesis of secretory granules. In Seminars in cell biology (Vol. 3, No. 5, pp. 357-366). Academic Press.

Volkova, E., & Smirnov, A. (2016). Regeneration of test in testate amoebae of the genus Arcella (Tubulinea, Arcellinida). European journal of protistology, 55, 128-140.

Wallich G. (1864). On structural variation among the Difflugians Rhizopods, Ann. Mag. nat. Hist., 13, 53-57.

Yamaoka, I., & Mizuhira, V. (1987). X-ray microanalysis of the mineral components in the scales of an amoeba, Cochliopodium sp.(Testacea). Cell and tissue research, 247(3), 633-637.

3 Chapter 2: Gene expression profiling of *A.intermedia* (Arcellinida : Amoebozoa) during shell formation and after 12h of shell formation: Shedding light on the molecular machinery involved in shell formation

3.1 Abstract

Arcellinida are characterized by the presence of a shell. The shell formation (thecagenesis) process has been well characterized based on cyto-morphological studies. A more complete understanding is impaired by the lack of molecular data to decipher the molecular machinery underlying this process. Aiming to shed light on the molecular aspect of the thecagenesis process, we designed a single-cell RNA-sequencing experiment to generate gene expression profiling of *Arcella intermedia* during two different life stages, shell formation and after twelve hours of shell formation. Transcriptome annotation, differential gene expression and Gene Ontology analyzes enabled us to molecularly describe these two life stages. We propose an annotated list of genes expressed by *Arcella intermedia* as candidates genes possibly involved in shell formation. Understanding the relation between gene products and shell morphogenesis is necessary to shed light on the origin and evolution of shell on *Arcellinida*.

3.2 Introduction

The presence of an external shell (test) is the key characteristic of Arcellinida, a monophyletic amoeboid lineage in the supergroup Amoebozoa (Nikolaev, 2005). The shell is a hardened outer structure, with a single aperture, that covers the cell of the organism. A new shell is formed before the reproduction process (budding division): after the cellular division, one daughter-cell keeps the old shell while the other daughter-cell keeps the newly formed one (Mignot and Raikov, 1990). Despite

its diversity, Arcellinida presents a general mechanism of thecagenesis: extrusion and arrangement of thechagenous granules cellularly controlled (Netzel, 1971; Netzel, 1975a; Netzel, 1975b; Netzel, 1975c; Netzel and Grunewald, 1977; Mignot and Raikov 1990). The most detailed shell formation descriptions are derived from studies of the genus Arcella (Moraczewski, 1971; Netzel, 1971; Netzel and Heunert, 1971; Netzel, 1975a; Netzel, 1975b; Netzel and Grunewald, 1977), however other diverse genera of Arcellinida have been studied (Netzel, 1972a; Netzel, 1975c; Netzel, 1976; Netzel, 1983; Harrison, 1976; Ogden, 1989; Ogden and Meisterfeld, 1989; Ogden and Pitta; 1989). Unlike the comprehensive morpho-cytological description available in the literature, we lack information about the molecular machinery involved in shell formation.

Aiming to shed light on the genes expressed during shell formation, we designed single-cell (whole-organism) RNA-sequencing experiments to generate gene expression profiling. We used *Arcella intermedia* (**Figure 3.4.1**) (Sphaerothecina : Arcellinida) as our model organism. We generated RNA-seq from single-cells in three different life cycle stages: i. Shell forming cells (SF); ii. Trophic cells (TC) after 4 hours of shell formation; iii. Trophic cells (TC) after 12 hours of shell formation. We annotated and performed Gene Ontology analysis to characterize the genes expressed both in shell forming and trophic cells.

Here we briefly describe the shell formation process observed under light microscope on *Arcella intermedia* strain used on this study. We present and discuss the newly generated single-cell transcriptomes; through Gene Ontology enrichment (GO) we identify the Biological Function, Cellular Component and Molecular Function from the genes expressed by *A. intermedia*, during shell formation and during trophic stage. Based on differential gene expression we discuss the biological processes that characterize the up-regulated genes on each life-cycle stage. The combined approach of RNA-seq, Gene Ontology analysis and differential gene

expression analysis enabled us to identify 539 genes assigned to biological processes described on the morpho-cytological descriptions of shell formation.

3.3 Materials and methods

3.3.1 Lineages strains and culture

We used *Arcella intermedia* isolate 6 as our model organism (**Figure 3.4.1**). We establish this strain in 2013 by isolating one individual of *A. intermedia* from an artificial water tank within Sao Paulo city, Brazil (University of Sao Paulo campus, coordinates lat - 23.565720, long -46.730512). *A. intermedia* strain is a clonal and mono-eukaryotic culture maintained as stock cultures on standard liquid media (20 ml of sterilized spring water + 250ul of "Cereal Grass Media"-FisherScientific, cat - Nº BP9727 -500), with periodic culture replication, in the Laboratory of Evolutionary Protistology, Institute of Biosciences, University of São Paulo, Brazil. For RNA extraction, we established a culture with 50 individuals from a stock culture and performed the single-cell RNA extractions after one week of culture growth. This same isolate was previously used on different studies, our laboratory established its shell morphology and morphometry variability (Porfírio-Sousa, 2017), described the growth curve and generation time (Ribeiro et al., submitted), generated a whole culture reference transcriptome and included it on a phylogenomic study of Amoebozoa (Kang et al., 2017). The *Arcella intermedia* isolate 6 reference transcriptome (hereafter *A. intermedia* Reference Transcriptome) is the most complete and comprehensive *A. intermedia* transcriptome available for Arcellinida; it was previously generated by whole culture RNA extraction from 40 dense cultures (in total around 40,000 individuals) and has been deposited with National Center for Biotechnology Information under the BioProject PRJNA380424.

We used this *A. intermedia* Reference Transcriptome as a reference in different steps of the present study.

3.3.2 Culture observation

We observed the development of different individuals in cultures under light microscopy. We observed and generated time-lapses of individuals during shell formation, and determined the generation time from *A. intermedia* isolate 6. We obtained the time-lapses taking pictures under light microscope, we assembled the pictures using a time-lapse assembler.

3.3.3 Gene Ontology Enrichment Analysis and Gene expression analysis

Library preparation and RNA-seq

We generated full-length RNA-seq from single cells of *A. intermedia* using Smart-seq2 protocol (Picelli et al., 2014). We extracted RNA from single-cell in three different life cycle stage: i. shell formation stage, 12 single-cell replicates; ii. trophic individual (control-1) after 4 hours of shell formation, 6 single-cell replicates; iii. trophic individuals (control-2) after 12 hours of shell formation, 6 single-cell replicates. During the experiment we kept the cells on the standard liquid media under the same conditions. We generated Illumina libraries following Nextera XT protocol. For sequencing, we made a single library pool with the 24 samples of *Arcella intermedia* and sequenced on a single lane of an Illumina HiSeq4000 machine.

Transcriptome processing, assembly and quality control

We trimmed low quality bases and index primer sequences from raw sequence reads using fastp software (Chen, 2018). We used the cleaned raw sequence reads files from all samples as assembly input and for downstream analysis. We performed RNA-Seq De novo Assembly Using Trinity for each of the 24 *Arcella intermedia* transcriptomes. We translated the assembled transcriptomes to protein using TransDecoder (https://github.com/TransDecoder/TransDecoder/wiki). We cleaned the transcriptomes and translated transcriptomes for bacterial sequences through personalized scripts, we implemented DIAMOND algorithm similarity search (Buchfink et al., 2014) on a personalized comprehensive OrthoMCL DB and cleaned transcripts and proteins addressed to bacterial sequence. To assess the robustness of our data, we compared the number of contigs present in each single-cell assembled transcriptome with the *A. intermedia* Reference Transcriptome (Kang et al., 2017). Additionally, we performed a quantitative measurement from the completeness of the transcriptomes using BUSCO v3 (Simão et al., 2015); Based on Eukaryotic OrthoDB BUSCO performed a search on each assembly for 303 genes that are expected to be present in all eukaryotes. For differential gene expression, we performed a quality check of each single-cell samples (shell forming and trophics) and its biological replicates comparing the whole transcriptome gene expression correlation using Trinity.

Gene expression profiling

We performed gene expression analysis using Trinity-supported companion utilities. We used RSEM (Li et al., 2011) to identify and quantify transcript expression by mapping each of the 24 single-cell transcriptomes (12 shell forming and

12 trophic individuals samples) to the *A. intermedia* Reference Transcriptome and assigning the samples reads to Reference transcripts. We assembled a transcriptome including the Reference Transcriptome and the single-cell transcriptomes generated in this study, since no change on mapping rate was identified we opted to keep only the currently available A. intermedia Reference Transcriptome as our reference. We performed differential gene expression analysis between shell formation samples and trophic individuals using edgeR.

Gene Ontology

We annotated the *A. intermedia* Reference Transcriptome using Eukaryotic Non-Model Transcriptome Annotation Pipeline (EnTAP) (Hart, 2017). This pipeline comprises: i. Similarity Search: optimized search against Uniprot, UniProtKB/TrEMBL and RefSeq release 87, implemented through DIAMOND program; ii. Contaminant Filtering and Best Hit Selection: selects final annotation and identifies potential contaminants; iii. Orthologous Group Assignment: independent assignment of translated protein sequences to gene families (eggNOG). Includes protein domains (SMART/Pfam), Gene Ontology (GO) terms, and KEGG pathway assignment; iv. InterProScan: sequence search against the families of InterPro databases to assign protein domains, Gene Ontology terms, and pathway information (Hart, 2017). Once we annotated the *A. intermedia* Reference Transcriptome, we have a comprehensive annotated transcriptome of *Arcella intermedia* to base the Gene Ontology analysis and gene expression profiling.

We performed Gene Ontology (GO) analysis for the genes expressed by shell forming and trophic individuals using Ontologizer. We present the GO analysis in two different perspectives: i. considering all genes expressed by shell forming and trophic individuals, aiming to describe in a general way both stages; ii. considering

the differentially expressed genes between shell forming and trophic individuals, aiming to characterize and differentiate each stage. We used Revigo for GO analysis visualization.

3.4 Results and Discussion

3.4.1 Shell formation of *Arcella intermedia* under light microscope

The shell formation and cell division process in *Arcella intermedia* takes around 30 minutes (**Figure 3.4.1 A – P**), from the first evidence of bud formation (**Figure 3.4.1 - A**) until a completely formed shell, followed by cell cytokinesis and individual dislocation (**Figure 3.4.1 – P**). The appearance of a division bud is the first evidence observed and is characterized by a darker cytoplasm in the region of the aperture. (**Figure 3.4.1 A - B**); the bud appearance is followed by a sudden growth of the bud and the appearance of a new shell, the darker region disappears (**Figure 3.4.1 C - D**). The newly formed shell already presents the characteristic morphology and morphometry of *A. intermedia* and the cytoplasm movements between both shells (**Figure 3.4.1 E - J**). The cytoplasmic flux is followed by the cytokinesis (**Figure 3.4.1 E - J**). Finally, the new individual and shell are completely formed (**Figure 3.4.1 M - P**). Thus, the strain used for this study presents a typical thecagenesis process, as described in chapter 1; the bud appearance is a remarkable characteristic, which enable to identify the thecagenesis process.

3.4.2 Dataset

We generated 24 single-cell transcriptomes of *Arcella intermedia*: 12 shell forming individual replicates (SF1-SF12); 6 trophic individual replicates after 4

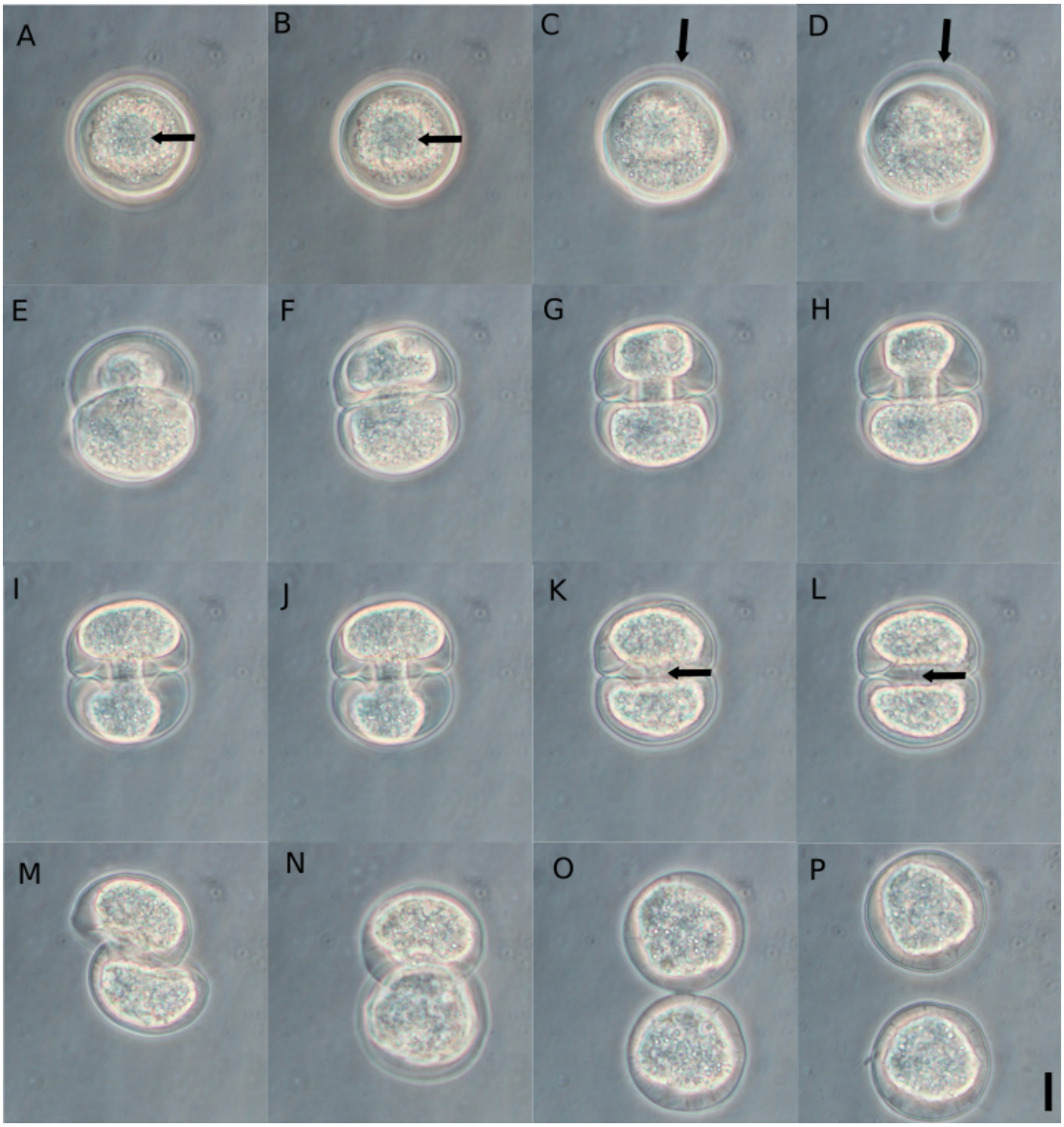

Figure 3.4.1 – Representation of shell formation and reproduction by budding division in *Arcella intermedia* observed under the light microscope. A – B. Identification of shell formation and cell division starting-point based on the appearance of a darker cytoplasm in the aperture region (arrow), due to the bud growth; C - D. First evidence of a new shell (Arrow). E. Newly formed shell with defined morphology and morphometry; F – J. Cytoplasmic flux between both shells; K - L. Cytoplasm equally distributed between the shells and cell cytokinesis (arrow). M – P. Two individuals completely formed and moving a part from each other.

hours of shell formation (TC1-TC6); 6 trophic individual replicates after 12 hours of shell formation (TC7-TC12). The basic statistics from the transcriptomes are present on (**Figure 3.4.2**). We obtained sequence depth between 8 – 13 millions of reads per transcriptome (**Figure 3.4.2 A**). The sequence depth comprises reads that represent both *Arcella intermedia* and contaminant sequences (mainly bacterial carry-over), as well as low quality reads.

The mapping rate from each transcriptome to the assembled *A. intermedia* Reference Transcriptome ranges between 1.5 – 6.5 millions reads for 23 of the 24 samples. One trophic cell replicate (TC10) presents a low mapping rate of around 270 thousand reads (**Figure 3.4.2 A**). Regarding sequence depth and biological replicates 23 single-cell transcriptomes, from the 24 generated, are suitable for gene discovery and differential gene expression analysis; Ziegenhain et al. (2017) demonstrated that transcriptomes generated through smart-seq2 protocol reaches gene discovery saturation at 1 million reads and Sims et al. (2014) demonstrated that biological replicates improves Differential Gene Expression detection power significantly regardless of sequencing depth. Part of the sequenced reads are not mapped to *Arcella intermedia* (**Figure 3.4.2 A**); since *Arcella intermedia* is maintained on bacteria rich media, diverse bacteria sequences are expected to be deeply sequenced, as well as virus and *Arcella* low quality sequences that are not possible to map.

The transcriptomes' BUSCO score ranges between 150 - 230 genes, from 303 genes searched, for 21 of the 24 samples (**Figure 3.4.2 B**). Based on this, each of the single-cell transcriptome present, in a relative way, around 50% to 75% of all genes expected to be expressed by an eukaryote. Replicates SF7, SF12 and TC10 present BUSCO score comparatively lower (112, 42 and 3 genes, respectively); Both SF12 and TC10 also present the lowest mapping rates (**Figure 3.4.2 A - B**). The *A. intermedia* Reference Transcriptome presents a BUSCO score of 273 genes.

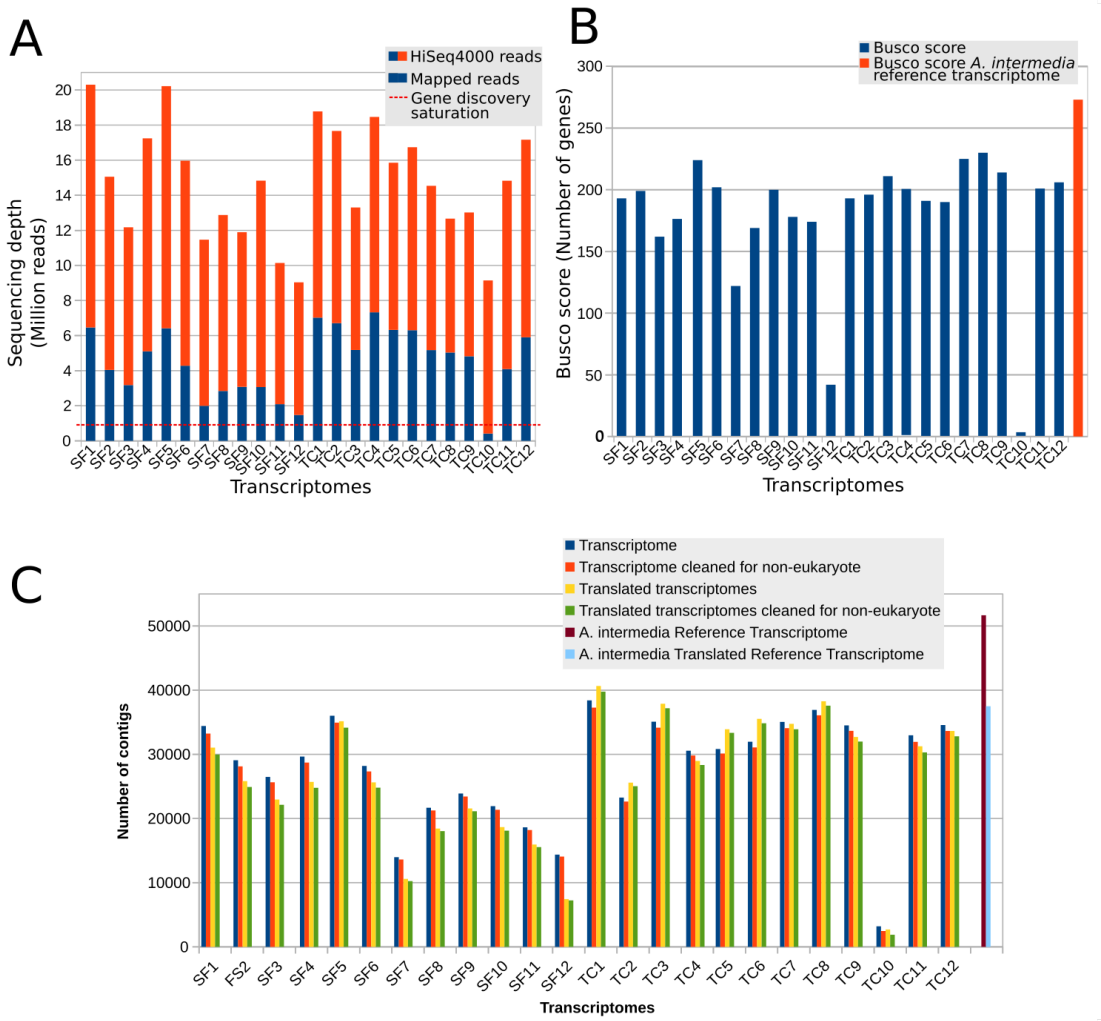

Figure 3.4.2 – Transcriptomes basic statistics, see legends for plot colors explanation. **A**. Sequence depth obtained by HiSeq4000 sequencing machine and the mapping rate from each single-cell transcriptome to the *Arcella intermedia* Reference Transcriptome, both measured on million reads. Red dashed line represents the number of reads necessary to reach gene discovery saturation with Smart-seq2 protocol (Ziegenhain et al., 2017); **B**. BUSCO score measured on number of reads identified from an evolutionarily-informed expectations of 303 gene in any eukaryotic complete transcriptome. BUSCO score from the *A. intermedia* Reference Transcriptome shown as a reference; **C**. Number of contigs recovered on assembled transcriptome and translated transcriptomes. Number of contigs recovered from the *A. intermedia* Reference Transcriptome shown as a reference.

The assembly from each transcriptome recovers between 19 - 40 thousand contigs, except samples SF7, SF12 and TC10 assemblies; consistently these low complexity samples (based on mapped reads and BUSCO score) recover less than 15 thousand contigs (**Figure 3.4.2 C**). Transcriptome translation maintains consistent the number of identified contigs (**Figure 3.4.2 C**). Non-eukaryote clean-up slightly decreases the number of contigs assembled (**Figure 3.4.2 C**). The *A. intermedia* Reference Transcriptome presents around 50 thousand transcripts and 38 thousand translated transcripts (**Figure 3.4.2 C**). Comparatively to the reference transcriptome and between replicates, most of the single-cell transcriptomes are rich representation from *Arcella intermedia* regarding to expected contig numbers. Together these contigs are the genetic units assembled from the transcriptomic data and represent complete genes, fragmented genes, gene isoforms and assembly artifacts.

Shell forming individual replicates are highly correlated based on gene expression and constitute a distinct SF cluster from the trophic individuals replicates (**Figure 3.4.3**); trophic individuals after 12 hours of shell formation replicates are also highly correlated and comprise a distinct cluster (**Figure 3.4.3**). Conversely, trophic individuals after 4 hours of shell formation replicates (TC) are poorly correlated with all replicates and constitute two different replicate clusters (**Figure 3.4.3**); low correlated and discrepant biological replicates are not suitable for Differential Gene Expression (DGE) analysis due to artifactual effects (REF). Based on that, TC1-TC6 replicates will not be considered on DGE analysis.

Based on the transcriptomes statistics and correlation, as well as preliminary DGE analysis (not shown), we consider 5 shell forming individual replicates (SF1, SF2, SF4, SF5 and SF6) and 5 trophic individual replicates (TC7, TC8, TC9, TC11 and TC12) as the final dataset for the following analysis.

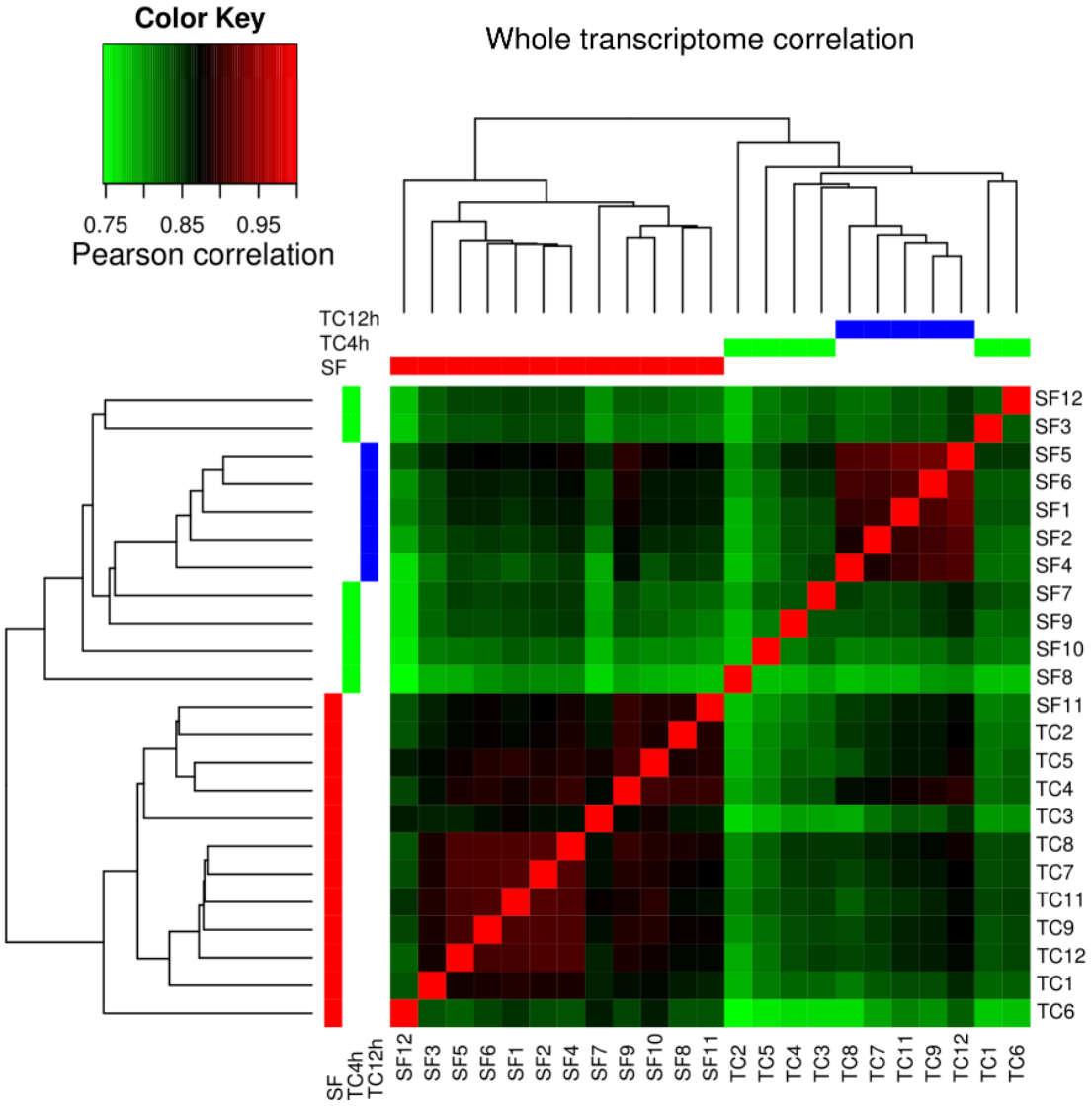

Figure 3.4.3 – Whole transcriptome gene expression correlation plot. Left and up trees represent the relative similarity between samples based on gene expression correlation. Red bars represent shell forming cells (SF), green bars represent trophic cells after 4 hours of shell formation (TC1-TC6) and blue bars represent trophic cells after 12 hours of shell formation (T7-TC12).

3.4.3 Expression Profiling

Transcriptome annotation and GO analysis

We successfully annotated 19,935 from 37,493 contigs of the *Arcella intermedia* transcribed Reference Transcriptome (**Figure 3.4.4**). We mapped the 19,935 annotated *A. intermedia* contigs to OrthoMCL groups database through Map your proteins to OrthoMCL groups workflow; these contigs represent 10,398 different genes (**Figure 3.4.4**). Similarity search shows that most of the annotated contigs are mainly similar to eukaryotic genes and primarily hit well-studied amoebozoan lineages (**Figure 3.4.5 A**), independently of database considered (**Figure 3.4.5 B - D**); Opisthokonta (with animals and fungi) and Viridiplantae (as *Arapdopsis thaliana*) are the second most represented lineages (Figure **3.4.5 A - D**), as expected due to Opistokonta and Amoebozoa close phylogenetic relationship and, Opistokonta and Viridiplantae available genomic data. A set of 17558 contigs of *A. intermedia* Reference Transcriptome were not annotated and may represent not only assembly artifacts (e.g. unreal genetic unit), but also genes of *A. intermedia* not present in other eukaryotes, thus demanding a bigger effort to be identified and annotated.

We performed Gene Ontology (GO) analysis on the 10,398 genes through Ontologizer tool; Ontologizer resolved a total of 5,751 different genes, assigning Biological Process, Cellular Component and Molecular Function GO terms. The biological meaningful genetic units (i.e. real genes) not resolved by Gene Ontology analysis demonstrate the need of more effort for GO annotation. Even though, the relative number of genes annotated through OrthoMCL and GO annotated from

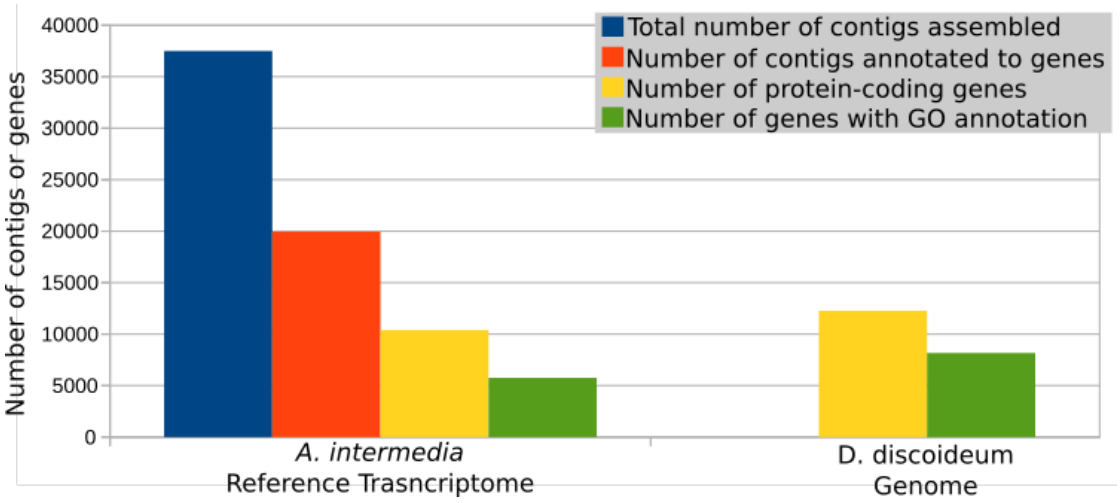

Figure 3.4.4 – Comparative plot between *Arcella intermedia* and *Dictyostelium discoideum* of number of genes annotated and resolved by GO analysis. The *Arcella intermedia* dataset consists on the *Arcella intermedia* transcribed Reference Transcriptome and the D. discoideum dataset consists on its Genome annotation.

our dataset is similar to the identified for the genome of *Dictyostelium discoideum*, a traditional model organisms of Amoebozoa (**Figure 3.4.4**). Based on this, we have a comprehensive annotated dataset of *A. intermedia*; we used this annotated reference transcriptome to base our Gene Ontology enrichment analysis, describing the general gene profiling of Arcella intermedia during shell formation and trophic stage (after 12 hours of shell formation), as well as the differential gene expression profiling between both stages.

Genes expressed by shell forming and trophic cells

We generated a shell forming (SF) reference transcriptome and a trophic cell (TC) reference transcriptome concatenating and assembling their respective sample replicates; the five SF replicates considered assembled a reference with 32,291 contigs and the five TC replicates assembled a reference with 36,451 contigs.

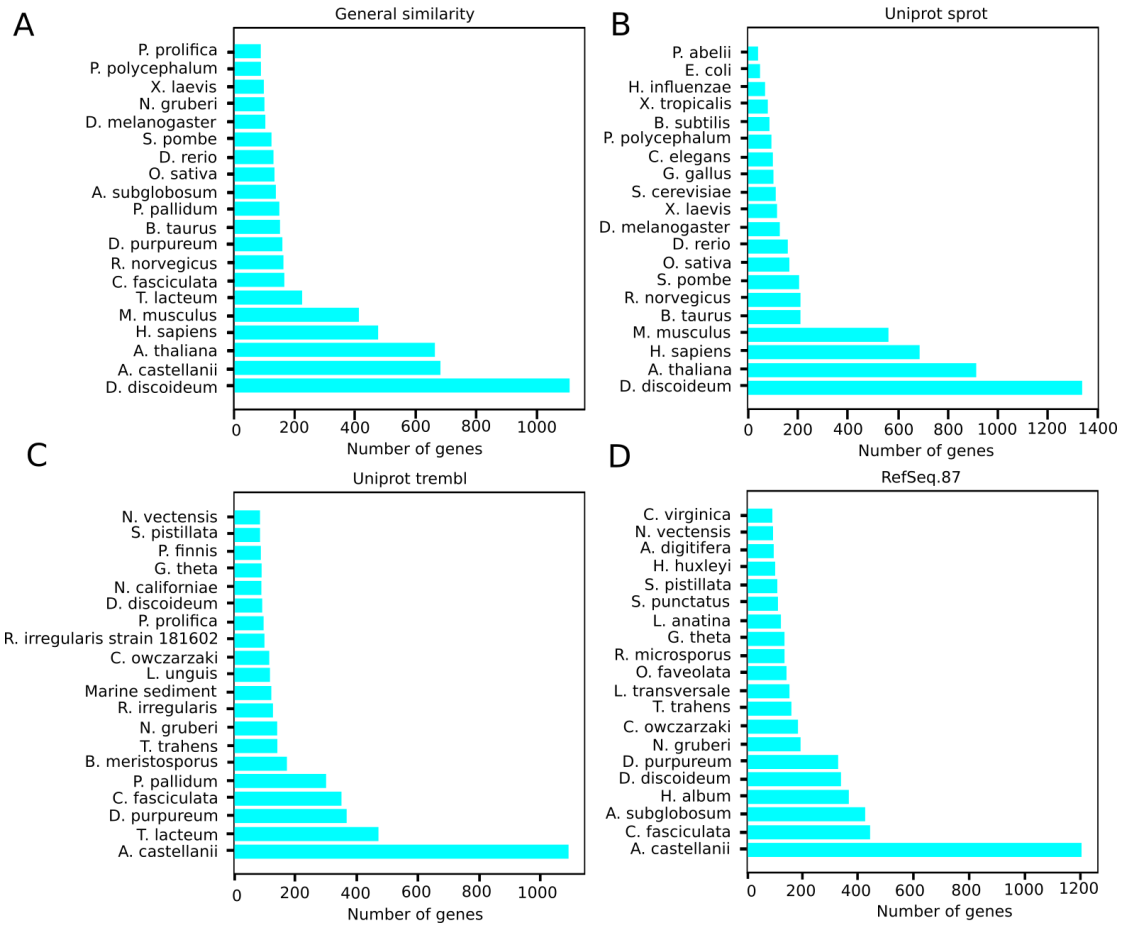

Figure 3.4.5 – Similarity search hit distribution showing the top 10 species. **A**. Ten most hit species considering together the tree databases used. **B**. Uniprot Database. **C**. Uniprot and Trembl database; **D**. RefSeq87 database.

We mapped SF and TC assemblies to the *A. intermedia* Reference Transcriptome and both SF and TC present all 10,398 genes annotated for *Arcella intermedia*.

In total 351 Biological Process GO terms (BP), 104 Cellular Component GO terms (CC) and 110 Molecular Function GO terms (MF) describe shell forming and trophic individuals transcriptomes (for complete table access: **Supplementary Table 3.1** at ⟨https://www.dropbox.com/sh/hqmqm5d284cfaen/AACshsScIT0jK3n5zBQOfPg2a?dl=0⟩). The Biological Function terms function-

ally characterize the gene expression on each stage. Both SF and TC present genes involved on diverse processes, ranging from mitosis (i.e. SF: GO:0000087; TC: GO:0000236), cytokinesis (i.e. SF: GO:0007105); protein process (i.e. SF: GO:0000209; TC: GO:0006624), response to environment (i.e. SF: GO:0032104; TC: GO:0009646), growth regulation (i.e. SF: GO:0040010), vesicle transportation control (i.e. SF: GO:0006888; TC: GO:0098927) to cell-wall biogenesis (SF: GO:0071852) and cuticle formation (TC: GO:0035017). Since the 5,751 successfully resolved genes are present on SF and TC replicates, not taking to account the differential gene expression, both replicates present the same set of genes assigned to GO terms.

Based on the morpho-cytological description of shell formation revised on the first chapter, we identified five GO terms (from the total of 351) that describe processes similar to the shell formation process: 1. GO:0071554, described as "cell wall organization or biogenesis" and represented by 155 and *A. intermedia* genes (**Supplementary Tables 3.2**); 2. GO:0043062, described as "extracellular structure organization" and represented by 48 genes and 144 contigs (**Supplementary Tables 3.3**); 3. GO:0045229, described as "external encapsulating structure organization" and represented by 113 genes and 215 contigs (**Supplementary Tables 3.4**); 4. GO:0046903, described as "secretion" and represented by 390 genes and 911 contigs (**Supplementary Tables 3.5**); 5. GO:0032940, described as "secretion by cell" and represented by 353 genes and 827 contigs (**Supplementary Tables 3.6**). In total these five GO terms are represented by 539 different genes, with several genes being assigned to more than one of these GO terms. Each gene can be represented by more than one assembled contig; this can represent not only assembly artifact or alternative RNA splicing, but also the presence of multiple copies of the same gene, originated by gene duplication, on *Arcella* (for the complete list of the genes, and the contig's ID and length (len), of *A. intermedia*

Reference Transcriptome, assigned to each of the five GO terms access: **Supplementary Tables 3.2 - 3.6** at ⟨https://www.dropbox.com/sh/hqmqm5d284cfaen/AACshsScIT0jK3n5zBQOfPg2a?dl=0⟩).

Differential gene expression profiling

Differential gene expression analysis identify 727 contigs that differentiates shell forming individuals from trophic individuals after 12h of shell formation (**Figure 3.4.6 A**); From these contigs 260 are up regulated during shell formation and down regulated on trophic individuals, while 467 are up regulated on trophic individuals and down regulated on shell forming individuals. The replicates are highly correlated within samples and lowly correlated between samples regarding the differentially expressed genes (**Figure 3.4.6 B**).

The 260 contigs up-regulated during shell formation represent 92 different genes annotated and GO resolved, and the 467 contigs up-regulated in trophic individuals represent 131 different genes annotated and GO resolved. In total 328 Biological Process (BP), 94 Cellular Component (CC) and 91 Molecular Function (MF) GO terms represent the 52 up-regulated genes during shell formation (for complete table access: **Supplementary Table 3.7** at ⟨https://www.dropbox.com/sh/hqmqm5d284cfaen/AACshsScIT0jK3n5zBQOfPg2a?dl=0⟩) - 367 BP, 130 CC and 200 MF GO terms represent the 131 up-regulated genes on trophic individuals (for complete table access: **Supplementary Table 3.8** - at ⟨https://www.dropbox.com/sh/hqmqm5d284cfaen/AACshsScIT0jK3n5zBQOfPg2a?dl=0⟩). The higher number of GO terms than genes identified is due to the same gene being assigned to more than one GO term with similar descriptions.

Shell forming individuals are characterized by 39 different Biological Process (BP) related to M phase of cell-cycle (mitosis + cytokinesis); these GO terms

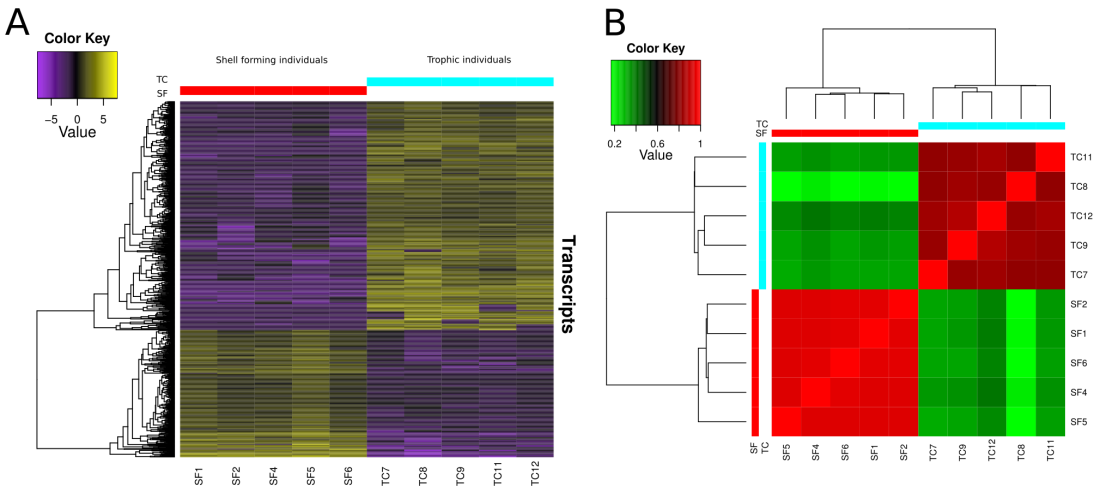

Figure 3.4.6 – Differential gene expression plots, SF1-6 are the shell forming individuals replicates and TC7-12 are the trophic individuals after 12 hours of shell formation. **A**. differentially expressed transcripts between shell forming individuals (red bar) and trophic individuals (blue bar). Color key represent arbitrary values to represent gene expression levels; purple represent down-regulated transcripts and yellow represents up-regulated genes. **B**. Correlation plot between samples based on differentially expressed genes expression. Left and up branch schemes represent the relative correlation between replicates.

describe BP ranging from bud site selection (i.e. GO:0031106 and GO:0032185), mitosis phase transitions (i.e. G0:0051781 and GO:0010389), DNA dynamics (i.e. GO:0006265 and GO:006338), nuclear division (i.e. GO:0051783 and GO:0007088) to cytokinesis (i.e G0:2000689 and GO:0000915) (for complete table access: **Supplementary Table 3.7** - at ⟨https://www.dropbox.com/sh/hqmqm5d284cfaen/AACshsScIT0jK3n5zBQOfPg2a?dl=0⟩). The expression of diverse genes involved on M phase of cell cycle is expected, as described on the first chapter, shell formation and cell division occurs in the same moment.

Trophic individuals up-regulate genes assigned to 67 BP GO terms that describe mitosis processes (for complete table access: **Supplementary Table 3.8** - at ⟨https://www.dropbox.com/sh/hqmqm5d284cfaen/AACshsScIT0jK3n5zBQOfPg2a?

dl=0⟩); these GO terms describe BP ranging from mitotic prometaphase (GO:0000236), anaphase (GO:0051322), mitotic spindle checkpoint (GO:0071174) to regulative processes of mitosis (i.e. GO:0045841 and GO:0051785). It is clear that trophic individuals after 12h of shell formation are already preparing to the next cell division. These genes are involved on mitosis and are up-regulated by cells after 12 hours of the last mitosis and shell formation; the genes must be functional translated proteins by the time the cell is forming the new shell and dividing, around 12 hours later.

GO analysis from the differentially expressed genes does not recover the five GO terms that describe processes similar to the shell formation process; the 539 genes assigned to these GO terms are not differentially expressed between SF and TC.

3.4.4 Conclusions

We generated a comprehensive single-cell RNA-seq dataset for *Arcella intermedia* during two different life stage, shell forming/reproduction and trophic cells after 12 hours of shell formation. Aiming to shed light on shell formation we applied a combined approach of gene annotation, Gene ontology analysis and differential gene expression; we identified and assigned GO terms to genes expressed on each stage and to genes differentially expressed between them. We focused on the Biological Function GO terms related to processes described for shell formation on *Arcella* and mitosis, since shell formation and cell division are simultaneous processes.

We identified that: i. both shell forming and trophic cell express genes involved on mitosis; ii. some mitotic genes are up-regulated on SF and down-regulated on TC, while others are down-regulated on SF and up-regulated on TC;

iii. TC after 12 hours of shell formation present diverse mitotic genes up-regulated; iv. 539 genes assigned to 5 different GO terms described to biological processes involved on shell formation are expressed on SF and TC, and present no differential expression.

This is the first study in *Arcellinida* to apply single-cell RNA-seq aiming to shed light on the organism development and shell formation. We propose the annotated list of 539 genes expressed by *Arcella intermedia* (for the complete list of the genes, and the contig's ID and length (len), of *A. intermedia* Reference Transcriptome, assigned to each of the five GO terms access: **Supplementary Tables 3.2 - 3.6** at ⟨https://www.dropbox.com/sh/hqmqm5d284cfaen/AACshsScIT0jK3n5zBQOfPg2a?dl=0⟩) as candidates genes possibly involved in shell formation. A complete list of the 10,398 annotated genes will be available on the publication of this chapter. Despite our focus aiming to shed light on candidate genes involved on shell formation, the dataset generated and available by this study enable to investigate other aspects of these organisms, as metabolic pathways, stimuli responses and cell signaling, on a single-cell perspective. Additionally, there is a significant fraction of contigs that were not annotated and, annotated genes with no GO annotation. This set of contigs possibly presents some biological meaningful genetic units (i.e. real genes), representing a dataset for gene discovery of *A. intermedia*.

3.5 References

Buchfink, B., Xie, C., Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. Nature methods, 12(1), 59.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. bioRxiv, 274100. Harrison, F. W., Dunkelberger, D., Watabe, N., Stump, A. B. (1976). The cytology of the testaceous rhizopod Lesquereusia spiralis (Ehrenberg) Penard. I. Ultrastructure and shell formation. Journal of Morphology, 150(2), 343-357.

Hart, A. (2017). EnTAP: Software to Improve the Quality and Functional Annotation of De Novo Assembled Non Model Eukaryotic Transcriptomes.

Hedley, R. H., Ogden, C. G., Mordan, N. J. (1976). Manganese in the shell of Centropyxis (Rhizopodea: Protozoa). Cell and tissue research, 171(4), 543-549.

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.

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics, 12(1), 323.

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.

Moraczewski, J. (1971). La composition chimique de la coque d'Arcella discoides Ehrbg. Acta Protozoologica, 8, 423-437.

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

Netzel, H. (1972b). Die Schalenbildung bei Difflugia oviformis (Rhizopoda, Testacea). Zeitschrift für Zellforschung und Mikroskopische Anatomie, 135(1), 55-61.

Netzel, H. (1975). Morphologie und Ultrastruktur von Centropyxis discoides (Rhizopoda, Testacea). Arch. Protistenkd, 117, 369-392.

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, Testaeea). 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. (1976). Die Abscheidung der Gehäusewand bei Centropyxis discoides (Rhizopoda, Testacea). Arch. Protistenk., 118, 53-91.

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

Netzel, H. (1983). Gehäsewandbildung durch mehrphasige Sekretion bei der Thekamöbe Netzelia oviformis (Rhizopoda, Testacea). Archiv für Protistenkunde, 127(4), 351-381.

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.

Picelli, S., Faridani, O. R., Björklund, Å. K., Winberg, G., Sagasser, S., & Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nature protocols, 9(1), 171.

Porfírio-Sousa, A. L., Ribeiro, G. M., Lahr, D. J. (2017). Morphometric and genetic analysis of Arcella intermedia and Arcella intermedia laevis (Amoebozoa, Arcellinida) illuminate phenotypic plasticity in microbial eukaryotes. European journal of protistology, 58, 187-194.

Ribeiro, G. M.; Prado, P. I., Coutinho, R. M., Rillo, M. C., Junior, S. P., Porfírio-Sousa, A. L.and Lahr, D. J. G. Growth rate modulation enables coexistence in a competitive exclusion scenario between microbial eukaryotes. Submitted to PeerJ.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, 31(19), 3210-3212. Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. Nature Reviews Genetics, 15(2), 121.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M.,& Enard, W. (2017). Comparative analysis of single-cell RNA sequencing methods. Molecular cell, 65(4), 631-643.

4 Chapter 3: "Translating" morpho-cytological evidences on a molecular interpretation of shell formation in *Arcella* (Arcellinida:Amoebozoa).

4.1 Abstract

Cyto-morphological studies have generated a comprehensive description of the shell formation in Arcellinida. These single-celled organisms build a new shell during each reproduction process, based on two main subcellular processes, granular secretion and budding division. To date, no molecular study discussing the thecagenesis process in *Arcellinida* has been published. The gap on our knowledge about the molecular machinery involved on shell formation impairs the raise of hypothesis to the origin and evolution of such a characteristic structure. However, the cyto-morphological descriptions available enable useful insights to derive a molecular model for the thecagenesis process.

4.2 Introduction

As reviewed in chapter 1, several authors described the shell formation (thecagenesis) process in *Arcella* (Arcellinida:Amoebozoa); together these authors established the cyto-morphological description of this process as we understand now (Netzel, 1975a; Netzel, 1975b; Netzel and Grunewald, 1977; Mignot and Raikov, 1990). The two key cellular processes underlying thecagenesis are the thecagenous granules exocytosis and morphogenesis of the shell by cell division bud formation and growth (Netzel, 1975a; Netzel, 1975b; Netzel and Grunewald, 1977; Mignot and Raikov, 1990). The first, is the characteristic process of regulated exocytosis of secretory granules (Peterson et al., 1987; McCaffery and Gillin, 1994; Burgoyne and Morgan, 2003; Tran and Hagen, 2017), a well-described process in well-known

Opisthokonta (animal and fungi) (Burgoyne and Morgan, 2003; Tran and Hagen, 2017); the second, is the reproduction by budding, a process observed in a variety of diverse eukaryotes (Schaudinn, 1903; Meyer and Abel, 1975) and well-described in *Saccharomyces cerevisiae* (Bi and Park, 2012).

Regulated exocytosis of secretory granules is a characteristic type of traffic of secretory vesicles (Burgess and Kelly, 1987; Burgoyne and Morgan, 2003; Tran and Hagen, 2017). It differs from the constitutive secretion by involving the accumulation of storage vesicles in the cytoplasmic region near the plasma membrane (PM), and the necessity of a stimulus for secretion to take place (Burgoyne and Morgan, 2003; Tran and Hagen, 2017); the constitutive vesicular secretion is a continuous process, it neither involves storage vesicles nor secretion stimulation (Burgoyne and Morgan, 2002; Tran and Hagen, 2017). The dense granular aspect of storage vesicles (secretory granules) is due to the accumulation of secreted material (Burgoyne and Morgan, 2003); this material is kept on these granules and are released once the vesicles are stimulated to synchronously fuse to the PM (Burgoyne and Morgan, 2003; Tran and Hagen, 2017). The process of granule secretion has been comprehensively described in model Opisthokonta (e.g. *Saccharomyces*, *Drosophila*, *C. elegans*, mice, and humans) and diverse cell types (Burgoyne and Morgan, 2003; Bembenek et al., 2007; Tran and Hagen, 2017). In mammals alone, tree different classes of cells, the endocrine, exocrine, and neuronal cells, represented by more then 42 cell types, present regulated exocytosis of secretory granules (for a non-exhaustive list of cells with secretory granules check table 1 from Burgoyne and Morgan, 2003). Also, the material secreted by these granules are diverse, ranging from peptide, secretion proteins, hormones, and neurotransmitters (Burgoyne and Morgan, 2003), to cite some. Different then initially expected (Burgoyne and Morgan, 2003), a similar molecular machinery controls secretion of granules, independent of organism or cell type, including the participation of conserved homologous proteins.

Budding division is a reproductive process well described in yeast *Saccharomyces*. The molecular machinery underlying this process involves several proteins homologous in diverse eukaryotes, as well as *Saccharomyces*' exclusive proteins (Bi and Park, 2012). Generally, this machinery controls cellular process characteristic of eukaryotic cells, as cell polarity, cytoplasmic flux, cell morphogenesis, and cytokinesis (Sanders and Field, 1995; Guertin et al., 2002; Bi and Park, 2012); together these processes enable the origin of a new organism from the parental cell, in a morphogenetically controlled way (Sanders and Field, 1995; Bi and Park, 2012).

Two different evolutionary scenarios may explain the cyto-morphological similarities of a subcellular process between diverse organisms: conservation of homologous features or convergent evolution (Andrews, 2000; Perry et al., 2006; Elde et al., 2007). In the first scenario, the similarities are due to conservation of a molecular machinery, keeping the homologous proteins, from a common ancestor; in the second, similar molecular machinery originates in two divergent lineages by convergent origin and evolution of proteins with analogous functions (Elde et al., 2007). Accumulating evidence shows that cyto-morphological similarity of the granule exocytosis process shared by *Saccharomyces*, *Drosophila*, *C. elegans*, mice and humans, lineages that comprise the monophyletic group of Opisthokonta, is due to conservation of the same homologous protein toolkit for granule secretion (Burgoyne and Morgan, 2003); on the other hand, studies of the molecular machinery involved on granule exocytosis in Ciliata (*Paramecium* and *Tetrahymena*) demonstrate that different set of proteins underlie the granule secretion machinery, when compared to Opisthokonta's machinery (Peterson et al., 1987; Madeddu et al., 1995; Chilcoat et al., 1996; Verbsky et al., 1998; Cowan et al., 2005). Although being different, the proteins of Opisthokonta and Ciliata present functional convergence (Almers, 1990; Erxleben and Plattner, 1994; Erxleben et al., 1997; Vayssie et al., 2000; Froissard et al., 2002; Turkewitz, 2004; Elde et al., 2007). Whether by conservation (based

on homologous proteins) or convergence (based on analogous proteins), similar subcellular processes, even between distantly related lineages, share comparable principles of cellular operation and present protein with similar functions (Andrews, 2000; Perry et al., 2006; Elde et al., 2007).

Aiming to shed light on the molecular machinery involved in shell formation in *Arcella*, we reviewed the literature about the molecular machineries that underlie regulated granule secretion and division by budding, focusing on the Opisthokonta lineage. Based on the molecular machinery conserved in Opisthokonta, and reviewed here, we hypothesize a molecular model of shell formation in *Arcella* derived from the cyto-morphological descriptions of thecagenesis. We include tree main aspects of thecagenesis: i. origin of thecagenous material from ER; ii. synchronous secretion of thecagenous granules; iii. shell morphogenesis by bud formation and growth during cell reproduction. We propose this model in an exploratory perspective, we describe the thecagenesis process assigning possible molecular machineries to each step, given what we know of similar subcellular processes from other organisms (Animal and Fungi). We hypothesize candidate proteins to be involved in this processes in *Arcella*, based on the possible scenario of homology between *Arcella* (Arcellinida : Amoebozoa), Animals and Fungi, due to the close relation between Amoebozoa and Opisthokonta; we use the transcriptomic data generated on the previous chapter to fundament our hypothesis, we identify whether these proteins are expressed in *Arcella intermedia* or not, and confirm their presence in other lineages of Amoebozoa. We argue that this model is based on an evolutionarily-informed and testable interpretation of the two key cellular processes underlying thecagenesis, regulated secretion of thecagenous granules and division bud formation; it can be the starting point for further advances to our understanding of the molecular aspect of shell formation process and the origin and evolution of *Arcella*'s shell.

4.3 Material and Methods

4.3.1 Literature review

We reviewed the literature on regulated secretion of secretory granules and cell division by budding; we focused on seminal studies and reviews that summarize the current knowledge of each process. Based on these works we describe the molecular machineries and proteins involved on these subcellular processes (here after referred as proteins of interest), focusing on Opisthokonta, a closely related lineage to Amoebozoa.

4.4 Dataset

We assembled a dataset with the proteins of interest from *Dictyostelium discoideum*, *Saccharomyces cerevisiae* and *Homo sapiens* collected from uniprot knowledgebase (UniProt Consortium). Based on this dataset we performed BLASTP searches to identify similar proteins from the translated *A. intermedia* Reference Transcriptome (reference used and described in chapter 2).

We checked the proteins' identity by multiple sequence alignment analysis. We assembled a dataset of each protein of interest from OrthoMCL DB and aligned them with the identified protein from *A. intermedia*. Based on the BLASTP searches and alignments we are able to check the protein identity. For clarity, we just present the alignment of *S. cerevisiae*, *D. discoideum* and *A. intermedia*, representing the alignment of each protein.

Our laboratory (Laboratory of Evolutionary Protistology - Biosciences Institute, University of Sao Paulo) and collaborating laboratory (Social and Evolutionary Protistology - Biological Sciences, Mississippi State University) maintain a comprehensive Amoebozoa dataset that was used as reference on the present study;

this dataset represents the compilation and translation of data available by two different studies, Kang et al., 2017 (BioProject PRJNA380424) and Lahr et al. submitted (BioProject accession number currently not available). We assessed the robustness of the identification of the proteins of interest on *Arcella intermedia* based on an evolutionary perspective. Based on the proteins identified on *Arcella intermedia* we performed BLASTP search to identify similar proteins on other Arcellinida lineages present on our personalized Amoebozoa transcriptomic dataset; we considered seven Arcellinida lineages, *Arcella intermedia*, *Cyclopyxis lobostoma*, *Difflugia sp.*, *Centropyxis sp.*, *Nebela tincta*, *Heleopera sphagni*, *Cryptodifflugia operculata*, that together represent all Arcellinida infraorder recovered by a phylogenomic study (Lahr et al., submitted). We generated multiple sequence alignment for each protein including lineages from Arcellinida, other Amoebozoa (*Dictyostelium discoideum* and *Entamoeba dispar*), Metazoa (*Caenorhabditis elegans*, *Drosophila melanogaster*and *Homo sapiens*), Fungi (*Saccharomyces cerevisiae*, *Neurospora crassa* and *Aspergillus fumigatus*), Archaeplastida (*Arabidopsis thaliana* and *Volvox carteri*) and Excavata (*Giardia lamblia*); based on the alignments we built phylogenetic trees for each protein. We performed multiple sequence alignment using GUIDANCE tool (Sela et al., 2015). We inferred maximum-likelihood trees using ModelFinder (Kalyaanamoorthy, 2017) and obtained branch supports with the ultrafast bootstrap (Hoang et al.,2017), both implemented in the IQ-TREE software (Nguyen et al., 2014). Alignment analysis and phylogenetic trees enable us to check the accuracy of gene identification and confirm its presence on Arcellinida lineage, excluding the possibility of identification on *A. intermedia* due transcriptome assembly artifact.

4.5 Results and discussion

4.5.1 Secretion of secretory granules

Regulated exocytosis of protein through secretory granules involves interactions between the Endoplasmic Reticulum (ER), Golgi and the cell plasma membrane (PM) (Burgoyne and Morgan, 2002). These interactions are based on vesicle traffic between these cellular compartments and between the Golgi and the PM, and are controlled by coordinated molecular machineries (Burgoyne and Morgan, 2002). The ER is involved in the synthesis of proteins that will be targeted to the cis-Golgi stack. Inside Golgi, these proteins are processed and modified, before leaving to the cytoplasm through granules derived from the trans-Golgi stack (Burgoyne and Morgan, 2002). Finally, the exocytic granules remain in the cytoplasm and synchronously fuse to the PM when stimulated (Burgoyne and Morgan, 2002). Here, we focus on the mechanisms involved in the traffic of vesicles from ER to Golgi and the regulation of the synchronous exocytosis of the granules on Opistokota.

Vesicle formation and transport from the Endoplasmic Reticulum to Golgi on Opisthokonta

The vesicle formation from Endoplasmic Reticulum is a COPII-mediated process (Barlowe et al., 1994; Jensen and Schekman, 2011); it involves a set of proteins conserved on eukaryotes (Barlowe et al., 1994; Cao and Barlowe, 2000; Jensen and Schekman, 2011). As shown in **Figure 4.5.1**, the dynamic between the proteins Sar1, Sec12, Sec23, Sec24, Sec13, Sec31, Bet3, Bet5, Trs20, Trs23, Trs31, Trs33, Uso1, Hrr25 and, Rab1 controls the COPII-mediated vesicle formation and traffic from ER to Golgi (Cao and Barlowe, 2000; Jensen and Schekman, 2011). The

activated protein Sar1 is the responsible for starting the ER membrane deformation and recruitment of the other proteins necessary for the vesicle formation; Sec12 catalyzes the activation of Sar1 from an inactive form (**Figure 4.5.1 A**). Once activated on the ER membrane, Sar1 recruits a complex of the proteins Sec23 and Sec24, the inner coat proteins; while Sec23 subunit binds to Sar1, Sec24 binds to a cargo protein that is going to be transported from ER (**Figure 4.5.1 B**). The formation of more pre-budding complex (Sar1, Sec23, Sec24 and cargo) continues (**Figure 4.5.1 C - F**) in parallel to the recruitment of the outer coat proteins (**Figure 4.5.1 E - F**). The outer complex is formed by Sec13 and Sec31; these proteins binds to the pre-budding complex, constituting the full COPII coat (**Figure 4.5.1 F**). The COPII coat assembly is followed by vesicle fission from the ER membrane; the vesicle containing the cargo is ready to be delivered to Golgi. On a site of the Sec23 protein, a complex of proteins (TRAPPI) binds to the COPII complex (**Figure 4.5.1 D**); TRAPII recruits Rab1, a member of the Rab GTPase family responsible on the regulation of vesicle traffic between ER to Golgi (**Figure 4.5.1 E - F**). Rab1 recruits the protein Uso1, involved on the anchoring of the vesicle to Golgi, through snares. Hrr25 protein controls the release of TRAPPI from the COPII complex, and by phosphorylating Sec23, Hrr25 regulates COPII disassembly(**Figure 4.5.1 G**); the vesicle containing the cargo is now ready to fuse with Golgi. The movement of the vesicles from the ER to Golgi is based on the interaction of myosin and actin filaments (Cao and Barlowe, 2000; Jensen and Schekman, 2011); this same complex (vesicle+myosin+actin filament) is the motor force of granules transport from Golgi to the cytoplasmic region near to the plasma membrane (Tran and Hagen, 2017).

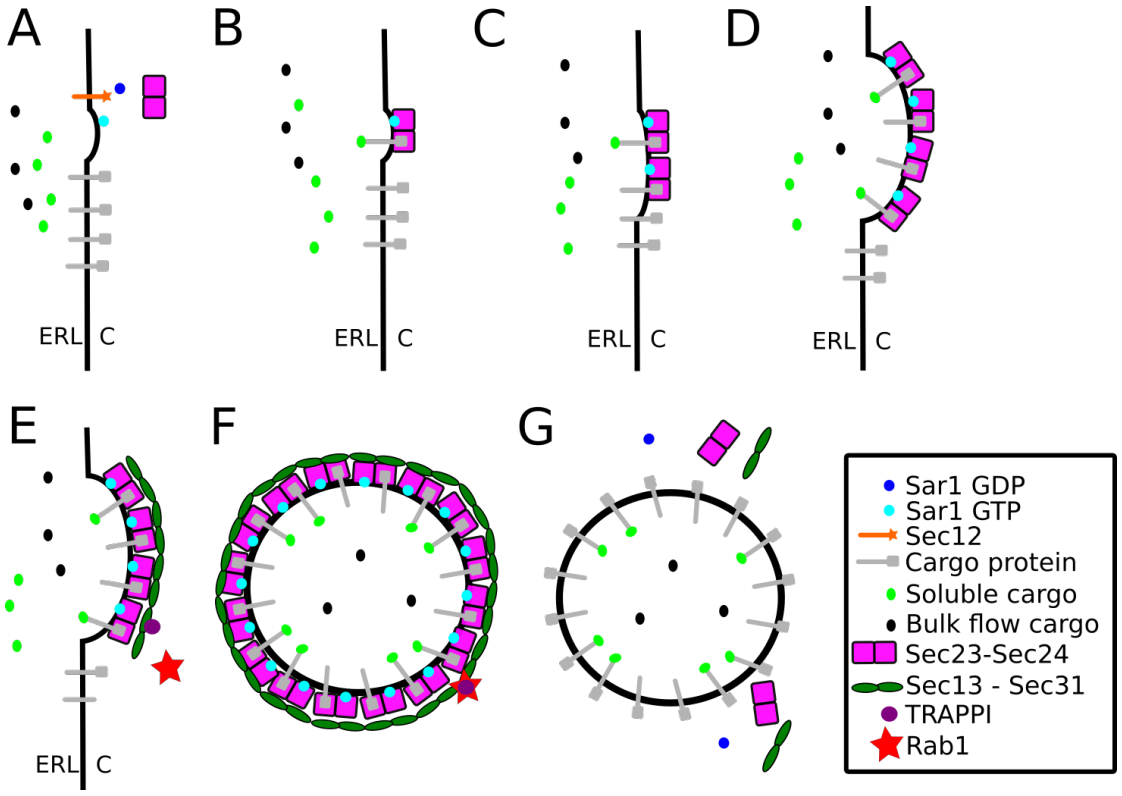

Figure 4.5.1 – Schematic representation of the COPII-mediated vesicle budding and protein sorting at the Endoplasmic Reticulum. **A**. ER's membrane deformation due to the activation and recruitment of Sar1; Sec12 activates Sar1 from the GDP form to the GTP (activated) form. Cargo inside ER's lumen (ERL) shown. **B**. Recruitment of Sec23-Sec24 complex; this complex interacts with the activated form of Sar1 and cargo proteins present on ER's membrane; the cargo proteins are responsible to sort soluble cargo from ER's lumen (ERL). **C - D**. The recruitment of Sec23-Sec24 complex, and its intereaction with Sar1 and cargo, continues and forms the pre-budding complex. **E - F**. In parallel to the formation of the pre-budding complex, a outer protein coat (Sec13 - Sec31) is recruited to the budding vesicle. TRAPPI and Rab1 attaches to the COPII coat and has important role on the transport and fusion of the vesicle to the Golgi membrane (not shown). **G**. COPII coat disassembles shortly before the vesicle fuse to the Golgi membrane. Check box for legend.

Regulated secretion of exocytic granules derived from Golgi on Opisthokonta

The regulated secretion of exocytic granules is characterized by a synchronized fusion of the granules to the plasma membrane. This synchronization is possible due to a molecular machinery composed by: Actin, myosin, Rho-GTPases, SNAREs, Arp2, Arp3, WASp, and RoK (**Figure 4.5.2**) (Rousso et al., 2016; Tran and Hagen, 2017). Actin plays diverse roles in regulated exocytosis. F-actin, together with myosin, form a physical barrier that prevents the premature fusion of the granules; once a positive stimulus for the regulated secretion is present, a clearance of this F-actin based barrier is observed, enabling the granules to fuse with the PM (Rousso et al., 2016; Tran and Hagen, 2017). F-actin and myosin (actomyosin machinery) are also involved on the stability of the granule fusion; these two proteins are recruited to the vesicle membrane, briefly after the granules passes the F-actin based barrier. A Rho-GTPase is responsible to regulate the F-actin recruitment (Rousso et al., 2016; Tran and Hagen, 2017). F-actin and myosin modulate the formation and dynamic of the fusion pore; consequently, the granule's membrane and the PM's membrane fuse forming a pore, through which the material will be released (Rousso et al., 2016; Tran and Hagen, 2017). The F-actin and myosin present in the granule membrane are responsible to generate a compression force that expels the granule cargo. Arp2/Arp3 complex, WASp and RoK are involved on the organization of the actomyosin machinery on the granules membrane; while Arp2/Arp3 are actin nucleators, RoK regulates myosin recruitment (Rousso et al., 2016; Tran and Hagen, 2017).

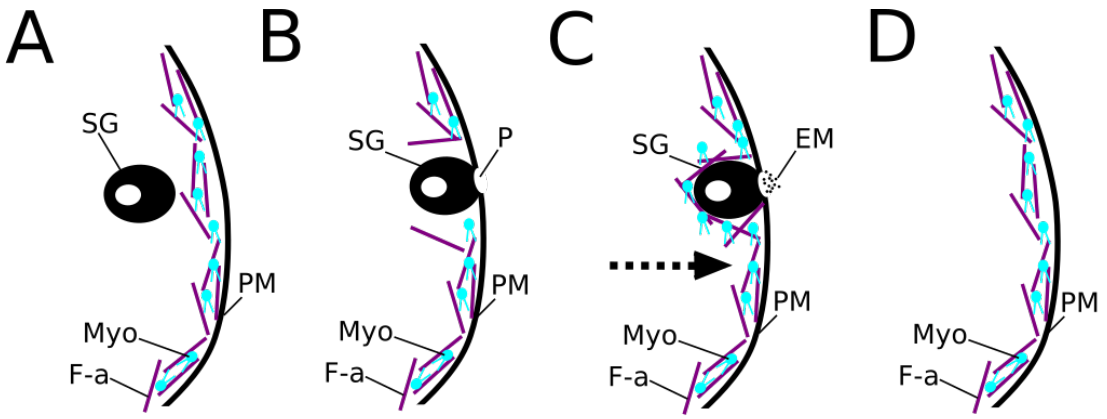

Figure 4.5.2 – Schematic representation of secretory regulated (SG) exocytosis of granules controlled by actinomyosin complex, single granule shown. **A**. A physical barrier of F-actin (F-a) and myosin (Myo) prevents the premature fusion of the secretory granule to the plasma membrane (PM). **B**. Upon the presence of a stimulus to the secret the granules, a clearance of the actinomyosin between the granule and the PM is observed. Consequently, the membranes fuse and form a exocytic pore (P). **C**. Shortly after the membrane fusion, F-actin and myosin are recruited to the granules' membrane; this actinomyosin complex generates an expel force and the exocytic material (EM) is completely secreted. Dotted arrow represents the direction of the expel force. **D**. The secretory granules are synchronously secreted and after secretion the F-actin and myosin complex reorganizes the physical barrier.

4.5.2 Cell division by budding

Cell division by budding is a characteristic reproduction process, differs from the cell division by fission by involving a bud growth (Schaudinn, 1903) and development before cytokinesis. Diverse eukaryotic lineages present budding division, as plants (e.g. tobacco mesophyll cells), diverse testate amoebae (e.g. Arcellinida and Rhizaria lineages) and fungi (e.g. *Saccharomyces cerevisiae*). Since *S. cerevisiae* is a traditional model organism, a comprehensive description of the budding division process is available for this organism, including the description of the molecular machinery.

Cell division by budding on *Saccharomyces cerevisiae*

Bud formation and cell division on *S. cerevisiae* is a process based on cell polarization controlled by actin, septins, and cytoplasmic microtubules (Kaksonen et al., 2006; Moseley and Goode, 2006; Moore et al., 2009; Oh and Bi, 2011). The interaction between actin, Arp2/3 complex, formins (Bnr1 and Bni1) and Cdc42 is the driving force for the polarized formation of the division bud. The polymerization of F-actin filaments is controlled by the Arp2/3 complex (**Figure 4.5.3. A**), while that, the formins (BnR1 and Bni1) are responsible to nucleate these growing filaments of actin (**Figure 4.5.3. B**). The protein Cdc42 is the responsible for the polarization of this process. The polymerization of F-actin filaments serves as the driving force for division bud formation and growth; he cytokinesis neck is signalized by the formin Bni1 (**Figure 4.5.3. C - E**). During each cell division, the next bud division site is determined by the recruitment of landmark proteins; these that signalize at the Plasma Membrane the next presumptive bud site (**Figure 4.5.3. E**).

4.5.3 Insights on the molecular machinery of thecagenesis process

The detailed studies about regulated exocytosis of granules and division by budding, based on lineages of Opisthokonta, give us useful insights to interpret the thecagenesis process in *Arcella*, enabling us to propose a model of the molecular machinery underlying this process. Based on the cyto-morphological description of the thecagenesis process, as reviewed in chapter 1, it is clear that *Arcella*'s cell synthesize intracellularly the proteinaceous material that constitute the shell; this material is stored inside thecagenous granules (exocytic granules) derived from Golgi, and is released once the granules synchronously fuse to the plasma

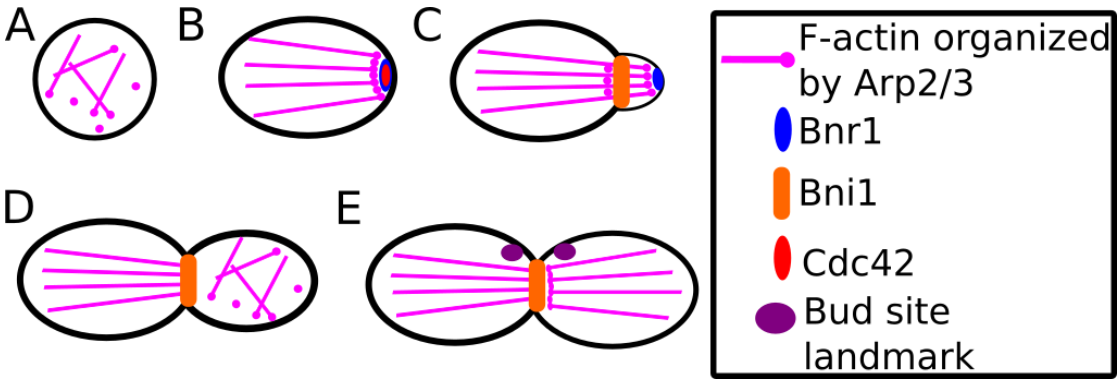

Figure 4.5.3 – Schematic representation the the division bud formation on *Saccharomyces cerevisiae*. **A** Growing F-actin filaments organized by the Arp2/3 complex. **B** The growing F-actin filaments are nucleated by a formin (Bnr1); the cell polarization and nucleation direction is determined by Cdc42. **C**. Formation of division neck and bud growth. Polarized polymerization of F-actin is the driving force for the bud growth. **D - E** F-actin filaments are nucleated by a forming (Bni1) toward the division neck, this region is the cytokinesis site. Check box for legend.

membrane. Following the release of the material, a growing cell division bud, based on cytoplasmic flux, determines the general shape of the new shell (daughter shell). Thus, granules exocytosis and cell division bud growth are the main cellular process underlying the morphogenesis of the shell (Netzel, 1975a; Netzel, 1975b; Netzel and Grunewald, 1977; Mignot and Raikov, 1990). Deriving the current knowledge about this processes on closely related organisms (Opisthokonta) to *Arcella*, on an evolutionarily-informed interpretation, we propose a model of the molecular machinery necessary to control key steps of shell formation on *Arcella*.

Model for the molecular machinery involved in the thecagenesis process of *Arcella*

Our current model for the molecular machinery of the thecagenesis process is presented on **Figures 4.5.4 - 4.5.6**. For clarity, we divided our model in

three distinct stages and 12 steps. Stage one (**Figures 4.5.4**) represents the thecagenesis process from the synthesis of the shell material to the thecagenous granules organization on a single layer in the edge of the Plasma Membrane, steps one to three (**Figures 4.5.4 A - E**). The step one of our model comprises the the synthesis of the proteinaceous shell material at the Endoplamic Reticulum (ER) and the transport of this material from ER to Golgi's cis-stack (GCS) (**Figures 4.5.4 A**). Although the absence of direct evidence of ER participation on the thecagenesis process, it is known that ER is a key eukaryotic organelle to participate on the exocytic pathway of proteins. Thus, the proteins that constitute the shell are probably synthesized by ribosomes (not discussed), stored inside ER's lumen, and subsequently sorted into vesicles (ERV); these vesicles must travels toward Golgi and fuse with Golgi's membrane, at its cis-stack (GCS) (**Figures 4.5.4 A**). As discussed, a COPII-mediated vesicular traffic from ER to Golgi (presented on **Figure 4.5.1**) is a conserved molecular machinery on eukaryotes, from Opisthikonta to plants; *Arcella intermedia* presents most of the proteins of COPII-mediated machinery, only Sec12 is absent on its transcriptome (**Figure 4.5.7**); Sec12 is also absent on the transcriptome of other Arcellinida, as well as on *Dictyostelium discoideum* and *Entamoeba dispar* genomes (**Figure 4.5.7**). The active Sar1 initiates the vesicle budding and is the precursor for the assembly of the remaining components of COPII; Sar1 activation is controlled by Sec12. It is clear that Arcellinida, and also *D. discoideum* and *E. dispar*, may present a protein with similar function of Sec12, since the whole process of vesicle budding and traffic from the ER depends on Sar1 activation; additionally, the complete machinery for COPII-mediated traffic, but Sec12, is expressed by *A. intermedia* and corroborates with the hypothesis that this mechanism of vesicle traffic is present on this lineage. F-actin filaments and myosin are responsible for the transport of the vesicles from ER to Golgi. Both F-actina and myosin present several paralogous on *A. intermedia*, not discussed

here, since the evolution of F-actin and myosin gene family is not the scope from the present; Lahr et al. (2010) investigated the expansions on the actin gene family in Arcellinida, a strain of *A. vulgaris* alone presented around 40-50 actin paralogous (Lahr et al., 2010).

The step two of our model comprises the traffic of exocytic granules from Golgi towards the plasma membrane. Inside Golgi, the proteins move from Golgi's cis-stack to the trans-stack and suffer modifications (not discussed); at the trans-stack, these modified proteins represent the proteinaceous material (thecagenous material) that is going to be released and form the shell. Dense granules (thecagenous granules), derived from Golgis's trans-stack membrane and containing the thecagenous material, travel toward the plasma membrane near to the aperture of the shell (**Figure 4.5.4 B**). The traffic of these granules is clearly a directional process, probably controlled by polarized polymerization of cytoskeletal apparatus apparatus, the F-actin and myosin complex (**Figure 4.5.4 B**).

The step three of our model comprises the formation of the division bud (DB) and the concentration of thecagenous granules in this region (**Figure 4.5.4 C - D**). As in division bud formation in *Saccharomyces cerevisiae*, cytoskeletal proteins are expected to control polarized cytoplasmic movements. Probably, on *Arcella*, actin filaments are responsible to generate the division bud, which grows through the aperture (**Figure 4.5.4 C**). Generally, on eukaryotic cells, proteins as Cdc42 (a small GTPase) and formins are responsible to coordinate the directional polymerization of the actin filaments, controlling the driving force of the actin polymerization; possibly, just as in *S. cerevisiae*, these proteins are involved on the polymerization of actin, necessary to the bud growth o *Arcella*. Consequently, cdc42 and formins would control the actin polymerization toward the shell aperture. In parallel, actin and myosin complexes may be also responsible to concentrate the thecagenous granules in this region (**Figure 4.5.4 D**). We propose that the

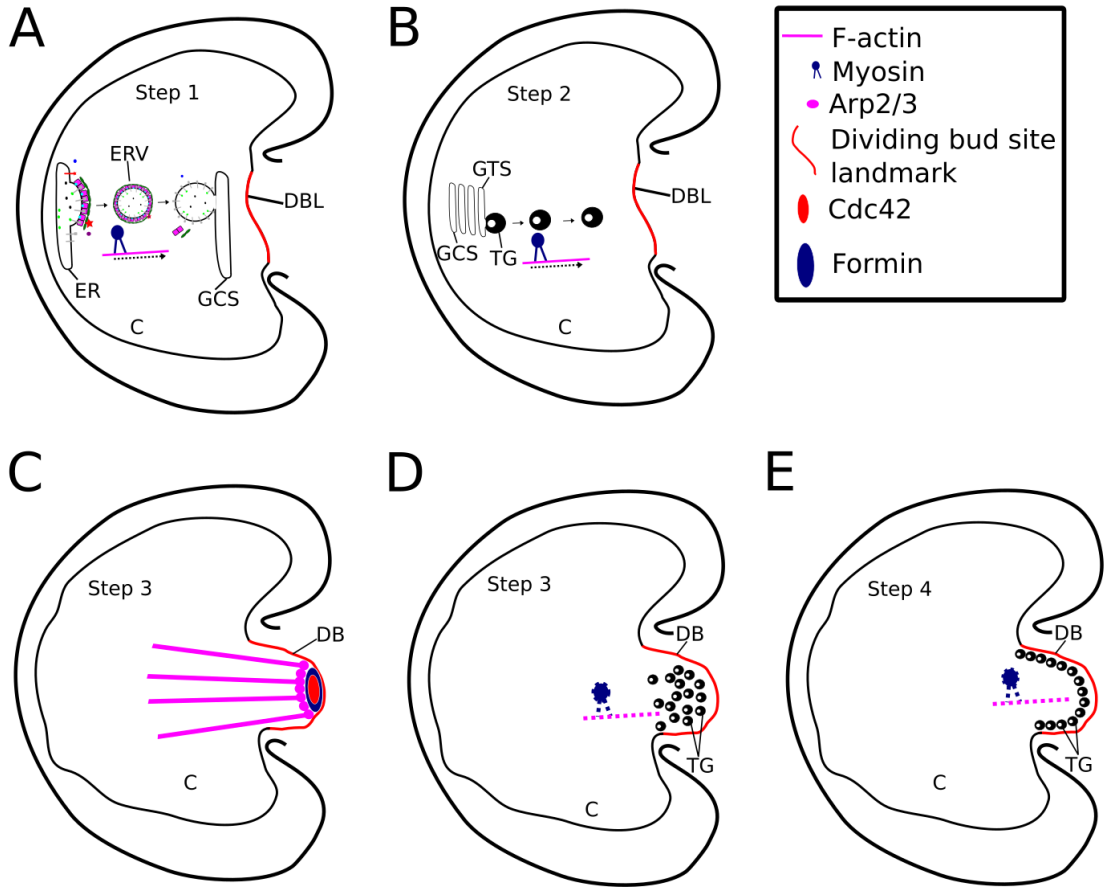

Figure 4.5.4 – Schematic representation of our model for the molecular machinery of the thecagenesis process on *Arcella* - Stage one. The relative position on the cytoplasm (C) and sizes of the structures is not realistic. **A**. Representation of protein sorting and transport from the Endoplasmic Reticulum (ER) to the Golgi's cis-stack (GCS), trhough ER's vesicle (ERV). Division bud landmark (DBL) is the presumptive site where the division bud is going to grow, see discussion in the text. **B**. Representation of thecagenous granules (TG) budding and transport from Golgi's trans-stack (GTS) to the Plasma Membrane. **C**. Representation of the bud growth through the shell aperture, as an actin-based process. **D**. In parallel to the division bud (DB) formation, thecagenous granules (TG), derived from Golgi, concentrate in the region of the division bud. **E** The thecagenous granules organized in a single layer adjacent to the plasma membrane. Detailed description of this stage is presented on the text. Check box for legend.

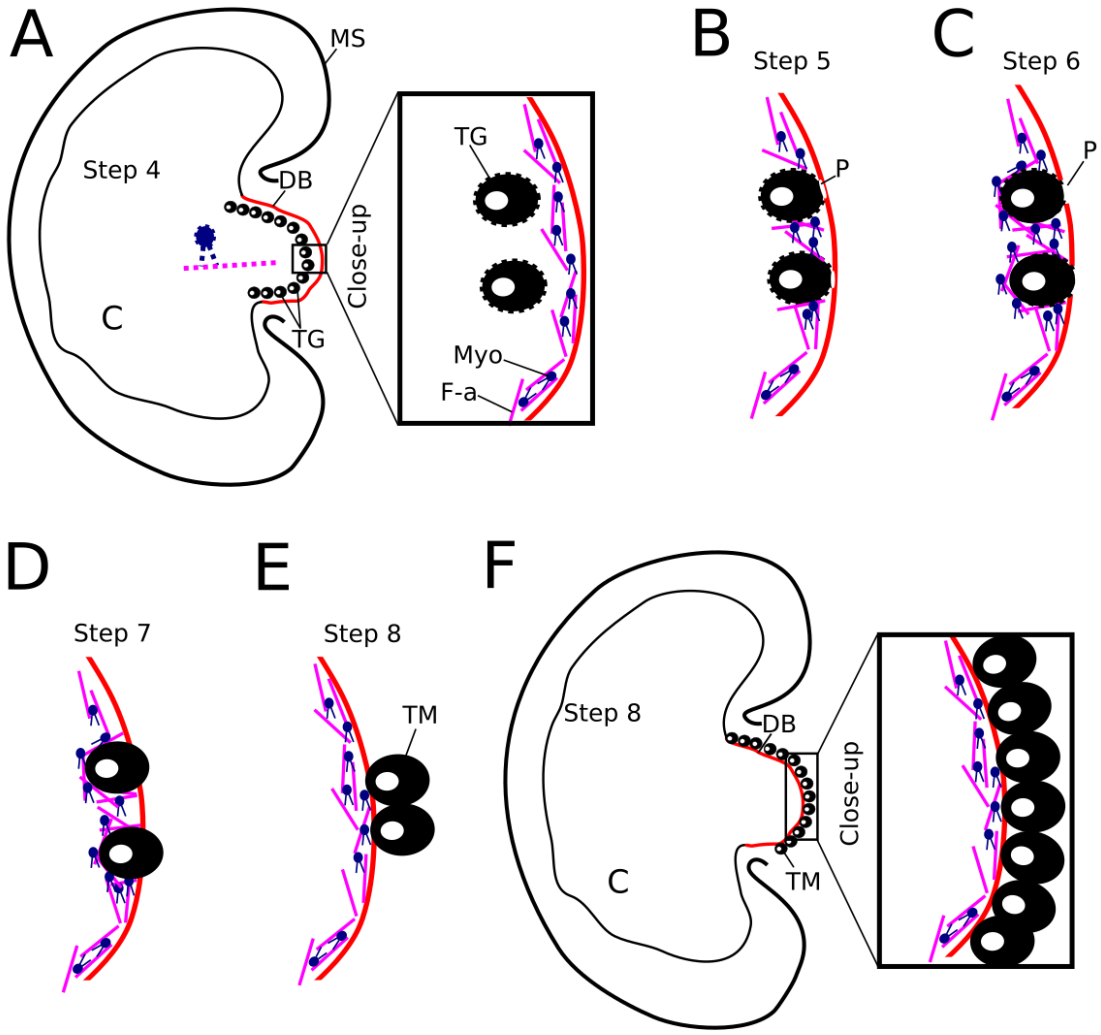

Figure 4.5.5 – Schematic representation of our model for the molecular machinery of the thecagenesis process on *Arcella* - Stage two. The relative position on the cytoplasm (C) and sizes of the structures is not realistic. **A**. Representation of the last step of stage one showing the detail of two granules in a single layer and the physical barrier organized by F-actin and myosin. Actinomyosin complexes participate in the organization of the granules in a single layer. **B**. Representation of the clearance of the F-actin and myosin physical barrier, and the fusion of the granules to the Plasma Membrane (PM). A exocytic pore (P) is shown. **C**. Representation of the recruitment of F-actin and myosin to the granules' membrane. **D - E**. Representation of the expel force generated by the F-actin and myosin complex; this force extrudes the thecagenous material (TM) extracellularly. **F**. Representation of the thecagenous material extracellularly and the F-actin and myosin barrier reorganized. Detailed description of this stage is presented on the text.

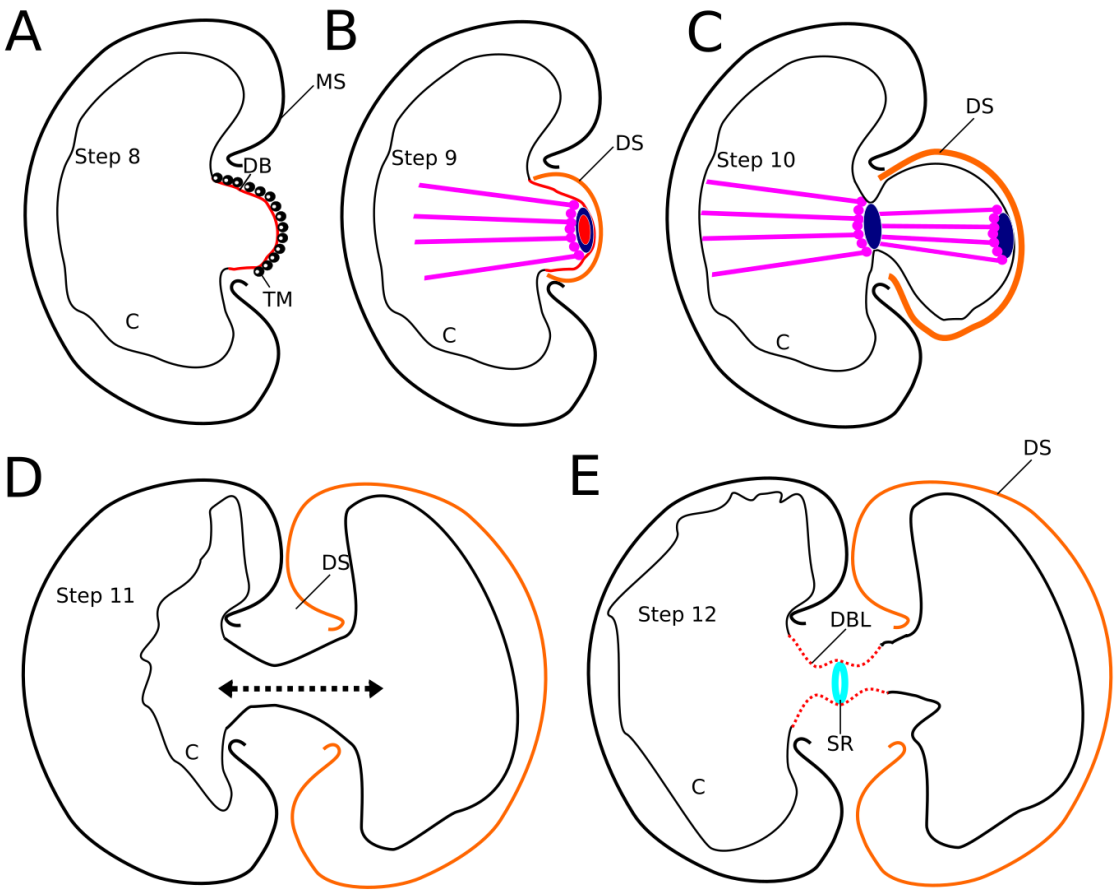

Figure 4.5.6 – Schematic representation of our model for the molecular machinery of the thecagenesis process on *Arcella* - Stage three. The relative position on the cytoplasm (C) and sizes of the structures is not realistic. **A**. Representation of the last step of stage two showing the thecagenous material (TM) extracellularly. **B**. Shortly after the granules exocytosis, the daughter shell (DS) is observed. We represent formin (blue) and Cdc42 (red) to be responsible for the polarization of F-actin growth. **C**. Representation of F-actin polymerization as the driving force of cytoplasmic flux and division bud growth. The bud growth models the newly formed shell (DS). **D**. Representation of the daughter shell with the defined shape and cytoplasmic flux (dotted arrow) involved to cytokinesis. **E**. Representation of the cytokinesis process controlled by a septin (SP) and the possible signalization of the next bud division site by a division bud landmark (DBL - dotted red lines).

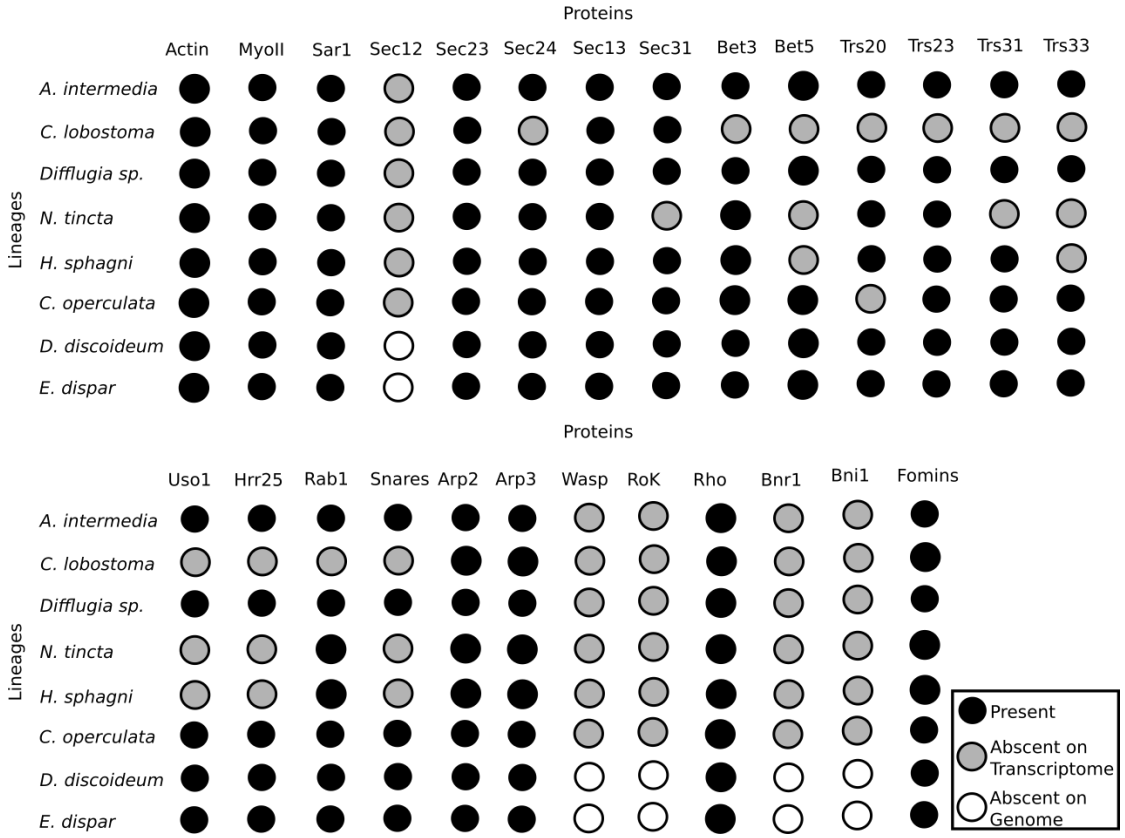

Figure 4.5.7 – Plot of the presence and absence of the proteins of interest on *Arcella*, other Amoebozoa and some selected eukaryotic lineages. Most of the proteins involved on regulated granule secretion and budding division are present on *A. intermedia* and other Arcellinida; some regulators of actin and myosin are absent. The phylogenetic trees and alignments representing the presence of the identified proteins are present in **Supplementary Figures 4.1 - 4.19**, access: ⟨https://www.dropbox.com/sh/hqmqm5d284cfaen/AACshsScIT0jK3n5zBQOfPg2a?dl=0⟩

division bud site is a region predetermined by a division bud site landmark (DBL - **Figures 4.5.4 A - E**), present even before the beginning of the thecagenous process, similarly to what was discussed for bud site on *S. cerevisiae*; this landmark can be signalized by a protein that is responsible to recruit Cdc42 and formins, the organizers of the actin filaments. Both, Cdc42 and formins are present on *A. intermedia* transcriptome, on other Arcellinida and, *D. discoideum* and *E. dispar* genomes (**Figure 4.5.7** and **Figure 4.7.16**).

The step 4 of our model comprises the organization of the thecagenous granules on a single layer in the division bud region. While the division bud develops and more granules derive from Golgi, member of the cytoskeletal apparatus, probably F-actin and myosin, organizes the granules on a single layer (**Figure 4.5.4 E**). This organization is followed by the synchronous extrusion of these granules, which is described as the second stage of our model (**Figure 4.5.4**).

The second stage of our model represents the process of regulated exocytosis of the thecagenous granules, described in four steps (steps five to eight). On Opisthokonta, the main factor that prevents premature fusion of granule to the plasma membrane, thus enabling a synchronous regulated secretion, is the presence of a physical barrier constituted by F-actin filaments and myosin; it is a key step for shell formation the synchronous extrusion of the thecagenous material from the granules, thus a similar physical barrier, if not homologous, is expected to prevent premature granule fusion (**Figure 4.5.5 A**). When stimulated to release the granules, step 5, this barrier must clear between the thecagenous granule and plasma membrane, thus the granules would be able to fuse to the membrane forming a pore (P) (**Figure 4.5.5 B**). On several eukaryotes, actin and myosin (Actinomyosin complex) participate on this process and are quickly recruited to the membrane of the fused granule (**Figure 4.5.5 C**); as in Opisthokonta, the actinomyosin complex may generate stability to the thecagenous granule fusion and exocytic pore dynamic;

the proteins Arp2, Arp3, WASp and RoK are involved in the recruitment of the actinomyosin complex to the granules' membrane. The complex of actinomyosin and its organizers (Arp2, Arp3, WASp and RoK) are possibly responsible to generate the expel force that extrudes synchronously the proteinaceous material that constitutes the shell (Step 7 - **Figure 4.5.5 D**). Once extruded the granules, the physical barrier of actinomyosin remodels (Step 8 - **Figure 4.5.5 E - F**). The secreted material will be seen as a premature externalized daughter shell (DS), which will be molded by division bud growth (Stage 3 - **Figure 4.5.6**). Arp2 and Arp3 are expressed by *Arcella intermedia* and is present in other Arcellinida (**Figure 4.5.7**); we did not identify the presence of WASp and RoK on *Arcellinida* (**Figure 4.5.7**).

The third stage of our model comprises the morphogenesis of the newly formed daughter shell (DS) and the cytokinesis process (steps nine to twelve - **Figure 4.5.6 B - E**). The cyto-morphological evidences of the thecagenesis process clearly demonstrates that, once the shell is externally visible (step nine - **Figure 4.5.6 B**) its shape is determined by the division bud growth (Netzel, 1975a; Netzel, 1975b; Netzel and Grunewald, 1977); as in *S. cerevisiae*, possibly the polymerization of actin filaments is the responsible for the division bud growth, and the indirect driving force for molding the new shell on *Arcella* (steps ten and eleven - **Figure 4.5.6 C - D**). The shell shape is determined by a single unidirectional cytoplasmic movement; subsequent cytoplasmic movement between the old shell and the daughter shell, which is involved to the cytokinesis process, does not modify the daughter shell shape (Step twelve - **Figure 4.5.6 D**). Finally, the cytokinesis process is observed, probably a proteins from the group of septins controls this process. The division bud and thecagenous granule secretions is always polarized and through the shell aperture; we argue that, as in the process of bud site selection on *S. cerevisiae*, is possible that a division bud landmark determines the proper region of the Plasma Membrane where the next division bud must grow and the

thecagenous granules secreted (i.e. in the region of the shell aperture) (Step twelve - **Figure 4.5.6 D**).

4.6 Conclusion

Here we proposed the first model for the molecular machinery of the thecagenesis process on *Arcella*. This process is based on regulated secretion of granules and division by budding, two subcellular processes present in closely related model organisms (Animals and Fungi). Based on the current knowledge about the molecular machineries that control the regulated secretion and bud growth on Opisthokonta, we derived an evolutionarily-informed exploratory interpretation for the molecular mechanism underlying the thecagenesis process on *Arcella*. We proposed some candidate proteins that may compose this machinery, and identified their presence on *Arcella intermedia* and other Arcellinida transcriptomes.

Actin and myosin, and other eukaryotic conserved proteins (small GTPase, Rab GTPase and formins), are the key components of the molecular machinery underlying the regulated granule secretion and division bud development on Opisthokonta. Whether an homologous machinery is present on *Arcella* and involved on the thecagenesis process, as we propose, can be tested. Our current model for the molecular machinery of the thecagenesis process enables us to shed light on some key aspects, as: i. concentration of thecagenous granules always in the edge of the Plasma Membrane, near to the shell aperture; ii. synchronized extrusion of the thecagenous granules containing the thecagenous material; iii. relation between cell division by budding and shell formation, two synchronous processes. Based on the current knowledge of eukaryotic subcellular processes and molecular machineries, a molecular machinery involving cytoskeletal molecules and cell polarization would address these key aspects.

Despite our focus on *Arcella*, the cyto-morphological description of shell formation of *Arcella* is similar to the shell formation process in other *Arcellinida*, despite the diversity of shell morphology on this lineage. Our model can be tested and serves a background for future advances on our understanding about shell origin and evolution on *Arcellinida*. Currently, based on our model, we can hypothesize that shell formation is a process that borrows key subcellular processes, present and conserved on diverse eukaryotes, to build a hardened outer structure that covers the cell organism. Moreover, cell division by budding is not exclusive to Arcellinida, it is a characteristic of other testate amoeba lineages (traditionally classified as Sarcodina) (Netzel and Grumewald, 1774); the shell formation process based on eukaryotic key subcellular processes and molecules (e.i. cell polarization through actin polymerization), coupled with the convergent origin of division by budding on diverse testate amoeba lineages, would shed light on how such complex structure, the shell, appeared and evolved independently at least eight times on diverse Eukaryotic lineages.

4.7 ***References***

Almers, W. (1990) Exocytosis. Annu. Rev. Physiol. 52, 607–624. amoebae demonstrate high diversity of microbial eukaryotes in the Neoproterozoic. (Accepted in Current Biology).

Andrews, N.W. (2000) Regulated secretion of conventional lysosomes. Trends Cell Biol. 10, 316–321

Barlowe, C., Orci, L., Yeung, T., Hosobuchi, M., Hamamoto, S., Salama, N., Rexach, M. F., Ravazzola M., Amherdt, M., & Schekman, R. (1994). COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. Cell, 77(6), 895-907.

Bembenek, J. N., Richie, C. T., Squirrell, J. M., Campbell, J. M., Eliceiri, K. W., Poteryaev, D., Spang, A., Golden, A. & White, J. G. (2007). Cortical granule exocytosis in C. elegans is regulated by cell cycle components including separase. Development, 134(21), 3837-3848.

Bi, E., & Park, H. O. (2012). Cell polarization and cytokinesis in budding yeast. Genetics, 191(2), 347-387.

Brown, M.W. Phylogenomics and ancestral morphological reconstruction of Arcellinida testate

Burgess, T. L., & Kelly, R. B. (1987). Constitutive and regulated secretion of proteins. Annual review of cell biology, 3(1), 243-293.

Burgoyne, R. D., & Morgan, A. (2003). Secretory granule exocytosis. Physiological reviews, 83(2), 581-632.

Cao, X., & Barlowe, C. (2000). Asymmetric requirements for a Rab GTPase and SNARE proteins in fusion of COPII vesicles with acceptor membranes. The Journal of cell biology, 149(1), 55-66.

Chilcoat, N.D. et al. (1996) Granule lattice protein 1 (Grl1p), an acidic, calcium-binding protein in Tetrahymena thermophila dense-core secretory granules, influences granule size, shape, content organization, and release but not protein sorting or condensation. J. Cell Biol. 135, 1775–1787.

Cowan, A.T. et al. (2005) Genetic, genomic, and functional analysis of the granule lattice proteins in Tetrahymena secretory granules. Mol. Biol. Cell 16, 4046–4060

Elde, N. C., Long, M., Turkewitz, A. P. (2007). A role for convergent evolution in the secretory life of cells. Trends in cell biology, 17(4), 157-164.

Erxleben, C. and Plattner, H. (1994) Ca 2+ release from subplasmalemmal stores as a primary event during exocytosis in Paramecium cells. J. Cell Biol. 127, 935–945.

Erxleben, C. et al. (1997) Microdomain Ca 2+ activation during exocytosis in Paramecium cells. Superposition of local subplasmalemmal calcium store activation by local Ca 2+ influx. J. Cell Biol. 136, 597–607.

Froissard, M. et al. (2002) N-ethylmaleimide-sensitive factor is required to organize functional exocytotic microdomains in Paramecium. Genetics 161, 643–650.

Guertin, D. A., Trautmann, S., McCollum, D. (2002). Cytokinesis in eukaryotes. Microbiology and Molecular Biology Reviews, 66(2), 155-178.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., Vinh, L. S. (2017). UFBoot2: improving the ultrafast bootstrap approximation. Molecular biology and evolution, 35(2), 518-522.

Jensen, D., & Schekman, R. (2011). COPII-mediated vesicle formation at a glance. J Cell Sci, 124(1), 1-4.

Kaksonen, M., Toret, C. P., Drubin, D. G. (2006). Harnessing actin dynamics for clathrin-mediated endocytosis. Nature reviews Molecular cell biology, 7(6), 404.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nature methods, 14(6), 587.

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.

Lahr, D. J., Nguyen, T. B., Barbero, E., & Katz, L. A. (2010). Evolution of the actin gene family in testate lobose amoebae (Arcellinida) is characterized by two distinct clades of paralogs and recent independent expansions. Molecular biology and evolution, 28(1), 223-236.

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.& Madeddu, L. et al. (1995) A large multigenic family codes for the polypeptides of the crystalline trichocyst matrix in Paramecium. Mol. Biol. Cell 6, 649–659.

Mccaffery, J. M., & Gillin, F. D. (1994). Giardia lamblia: ultrastructural basis of protein transport during growth and encystation. Experimental parasitology, 79(3), 220-235.

Meyer, Y., & Abel, W. O. (1975). Budding and cleavage division of tobacco mesophyll protoplasts in relation to pseudo-wall and wall formation. Planta, 125(1), 1-13.

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.

Moore, J. K., Stuchell-Brereton, M. D., Cooper, J. A. (2009). Function of dynein in budding yeast: mitotic spindle positioning in a polarized cell. Cell motility and the cytoskeleton, 66(8), 546-555.

Moseley, J. B., Goode, B. L. (2006). The yeast actin cytoskeleton: from cellular function to biochemical mechanism. Microbiology and Molecular Biology Reviews, 70(3), 605-645.

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

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

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

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., Minh, B. Q. (2014). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular biology and evolution, 32(1), 268-274.

Cao, X., Barlowe, C. (2000). Asymmetric requirements for a Rab GTPase and SNARE proteins in fusion of COPII vesicles with acceptor membranes. The Journal of cell biology, 149(1), 55-66.

Oh, Y., & Bi, E. (2011). Septin structure and function in yeast and beyond. Trends in cell biology, 21(3), 141-148.

Perry, A.J. et al. (2006) Convergent evolution of receptors for protein import into mitochondria. Curr. Biol. 16, 221–229.

Peterson, J. B., Nelson, D. L., Ling, E., & Angeletti, R. H. (1987). Chromogranin A-like proteins in the secretory granules of a protozoan, Paramecium tetraurelia. Journal of Biological Chemistry, 262(36), 17264-17267.

Rousso, T., Schejter, E. D., & Shilo, B. Z. (2016). Orchestrated content release from Drosophila glue-protein vesicles by a contractile actomyosin network. Nature cell biology, 18(2), 181.

Sanders, S. L., & Field, C. M. (1995). Cell Division: Bud-site selection is only skin deep. Current Biology, 5(11), 1213-1215.

Schaudinn, F. (1903). Untersuchungenuer die Fortpflanzung einiger Rhizopoden.(Vorlafige Mittheilung). Arbert Kaiserlichen Gesundheitsamte, 19, 547-576.

Sela, I., Ashkenazy, H., Katoh, K., & Pupko, T. (2015). GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Research, 43(W1), W7-W14.

5 Chapter 4: Phylogenetic study reveals massive RabGTPase family expansion in Amoebozoa.

5.1 Abstract

Rab GTPase is a paralogue-rich protein family involved on intracellular vesicular traffic control. It is characteristic to Eukaryotes and has expanded in diverse lineages. Given its key role on cellular processes, Rab GTPase functions have been well studied on different model organisms, and the evolutionary story has been described on phylogenetic studies. To date, Rabs have been studied only on a small part of the Amoebozoa diversity. This group is proposed to have a massive expansion of the Rab GTPase family. Here we present a phylogenetic study of the Rab GTPase, focused on Amoebozoa diversity, based on recently available transcriptomic data. We discuss the general pattern of Rab expansion on three major lineages of Amoebozoa (Tubulinea, Evosea and Discosea).

5.2 Introduction

Vesicular movement is a key characteristic of eukaryotes (Stainer, 1970; Hutagalung and Novick, 2011; Elias et al., 2012) mainly controlled by the Rab GTPase proteins family (Elias et al., 2012). Rab GTPases comprise a paralogue-rich protein family that regulates all stages of membrane trafficking via vesicles (Zerial and McBride, 2001; Stenmark and Olkkonen, 2001; Hutagalung and Novick, 2011). These proteins control from vesicle budding, cargo sorting and transportation to vesicle tethering and fusion (Segev, 1988; Bacon et al., 1989; Plutner, 1991; Eathiraj et al., 2005; Pfeffer, 2005; Hutagalung and Novick, 2011; Elias et al., 2012).

The Rab GTPase family has a lineage-specific composition. While human genome presents at least 60 Rab paralogues (Bock, 2001), Yeast present 11 Rab

(Lipatova, 2015) and Arabidopsis 57 Rab (Rutherford and Moore, 2002). Along with conserved orthologues present in these three lineages, several rabs are lineage-specific. This composition specificity is related to multiple Rab GTPAse family radiations (Rutherford and Moore, 2002). Such radiations may give rise to proteins that perform lineage-specific roles (Rutherford and Moore, 2002).

The Rab GTPase family has been studied in two of the three major lineages of Amoebozoans: Evosea (represented by *Dictyostelium discoideum*, *Entamoeba catellanii* and *Mastigamoeba balamuthi*) and Discosea (represented by Acanthamoeba castellanii) (Saito-Nakano et al., 2005; Diekmann et al., 2011; Elias et al., 2012). Saito-Nakano et al. (2005), reported the presence of around 90 putative Rab genes on *Entamoeba histolytica*, representing a massive repertoire of Rab and difficulty to assign Rab subfamilies (Diekmann et al., 2011). These lineages represent only a small fraction of the known Amoebozoa diversity. Additionally, Rab GTPases has not been studied in Tubulinea, the third major Amoebozoan lineage (Kang et al., 2017).

Here we present a broad phylogenetic study of Rab GTPase family in Amoebozoa. This is the first study of the Rab GTPase family that considers the diversity of an whole eukaryotic supergroup. We present the Rab GTPases from the three major lineage that compose Amoebozoa and present newly identified Rab paralogues. We demonstrate the Rab GTPase family expansion on Amoebozoa and that each Amoebozoa major group are characterized by exclusive Rab's paralogues. Additionally, we present the motifs that characterize Rab GTPase and enable to identify the family members.

5.3 Material and Method

5.3.1 Rab GTPase database

Rab GTPase is an well annotated gene family for diverse Eukaryotes; Elias et al. (2012) assembled a comprehensive protein database of manually curated Rab sequences from Genomes and Expressed Sequence Tag (EST) survey, comprising 1453 sequences from 55 eukaryotic lineages (Elias et al., 2012). We used this database (here after Elias et al. database) to base our search for Rab sequences on our database of Amoebozoa. As described on **Chapter 3**, our laboratory (Laboratory of Evolutionary Protistology - Biosciences Institute, University of Sao Paulo) and collaborating laboratory (Social and Evolutionary Protistology - Biological Sciences, Mississippi State University) maintain a comprehensive Amoebozoa dataset, it is the compilation and translation of data available by two different studies, Kang et al., 2017 (BioProject PRJNA380424) and Lahr et al. accepted (BioProject accession number currently not available), as well as available genomes of Amoebozoa. From the dataset of Amoebozoa we selected the transcriptomic or genomic data from: Tubulinea, Evosea and Discosea representatives. These lineages are representatives of Amoebozoa diversity.

We used the Rab sequences from Elias et al. database as query on a blastp search to identify Rab sequences from selected lineages of Amoebozoa. The identified sequences composed our preliminary Amoebozoa Rab GTPase dataset. We aligned this Amoebozoa dataset and built a phylogenetic tree, using Ran sequences as our outgroup. Rab GTPase family and Ran are members of the Ras superfamily (Wennerberg, 2005), and Ran is the closest member to Rab GTPase and has been proposed as the outgroup to study Rab GTPase evolution (). This approach enabled us to select the sequences that represent Rab GTPase members from the sequences identified by the blastp search, since this search identified not only

Rab sequences, but also sequences that represent members of the other families the Ras superfamily (see Wennerberg et al., 2005 for a description of the Ras superfamily and its members relationship). From this same phylogenetic tree we identified artifactual duplication patterns of the same protein in each lineage; the artifactual pattern consists on a single lineage presenting several copies, slightly different (probably due to assembly artifacts or alternative RNA splicing), of the same protein sequence. We manually curated our Rab GTPase Amoebozoa dataset excluding these artifactual duplication. By doing these we assembled our final Rab GTPase Amoebozoa dataset, cleaned for non-Rab sequences and artifactual duplications; we used this final dataset for the present study of the evolution of Rab GTPAse family on Amoebozoa.

5.3.2 Rab GTPase family phylogenetic study

We combined our final Rab GTPase Amoebozoa dataset with the Elias et al. database. We aligned this combined Rab GTPase database and inferred the maximum-likelihood trees. We performed all alignments on this study with Guidance algorithm (Landan, 2008; Sela, 2015), using MAFFT as the alignment program. We inferred all the maximum-likelihood trees using ModelFider (Kalyaanamoorthy, 2017) and obtained branch supports with the ultrafast bootstrap (Hoang et al.,2017), both implemented in the IQ-TREE software (Nguyen et al., 2015). The best-fitting model for ML analyses was LG+G4.

5.3.3 Rab GTPase motif search

Rab GTPase protein can be unequivocally identified based on twelve highly conserved motifs (Moore, 1995; Stenmark and Olkkonen, 2001). These motifs

are known to functional regions of Rabs, both due to be binding sites or enable proper protein fold. We identified Rab GTPase motifs using the MEME suite motif discovery (*Bailey et al., 2009*). We performed motif analysis considering all sequences from our final Amoebozoa Rab GTPase dataset together and considering separately each Rab GTPase identified member.

5.4 Results and Discussion

Our phylogenetic study identified 70 paralogues of Rab GTPase on Amoebozoa (**Figure 5.4.1 and Supplementary Figure 5.1**), unequivocally identified based on the twelve highly conserved motifs (Moore, 1995; Stenmark and Olkkonen, 2001) (**Figure 5.4.2**). From these paralogues, 39 are present on at least one major lineage of Amoebozoa (Tubulinea, Evosea or Discosea), comprising 17 paralogues that are present on other Eukaryotes and 22 paralogues possibly exclusive to Amoebozoa (**Figure 5.4.1 and Figure 5.4.3**). The remaining 29 paralogues are only present on the genomes of *Dictyostelium discoideum* (19 paralogue) and *Acanthamoeba castellanii* (10 paralogues).

5.4.1 Conserved Rab GTPases

Rabs 1A, 1B, 1C, 1D, 8, 2, 4, 14, 18, 11, 6, 5, 24, 21, 7A, 32A, and 32B are present in at least one of the major Amoebozoa lineages (Tubulinea, Evosea and Discosea) and constitute well supported clades (**Figure 5.4.1 and Supplementary Figure 5.1**). These Rabs are conserved members from the Rab GTPase family (**Figure 5.4.4**). They are present in animal, fungi, plants, ciliates and other diverse Eukaryotes (Gabernet-Castello et al., 2011; Elias et al., 2012) and present conserved localization and functions in the cell (Hutagalung and Novick,

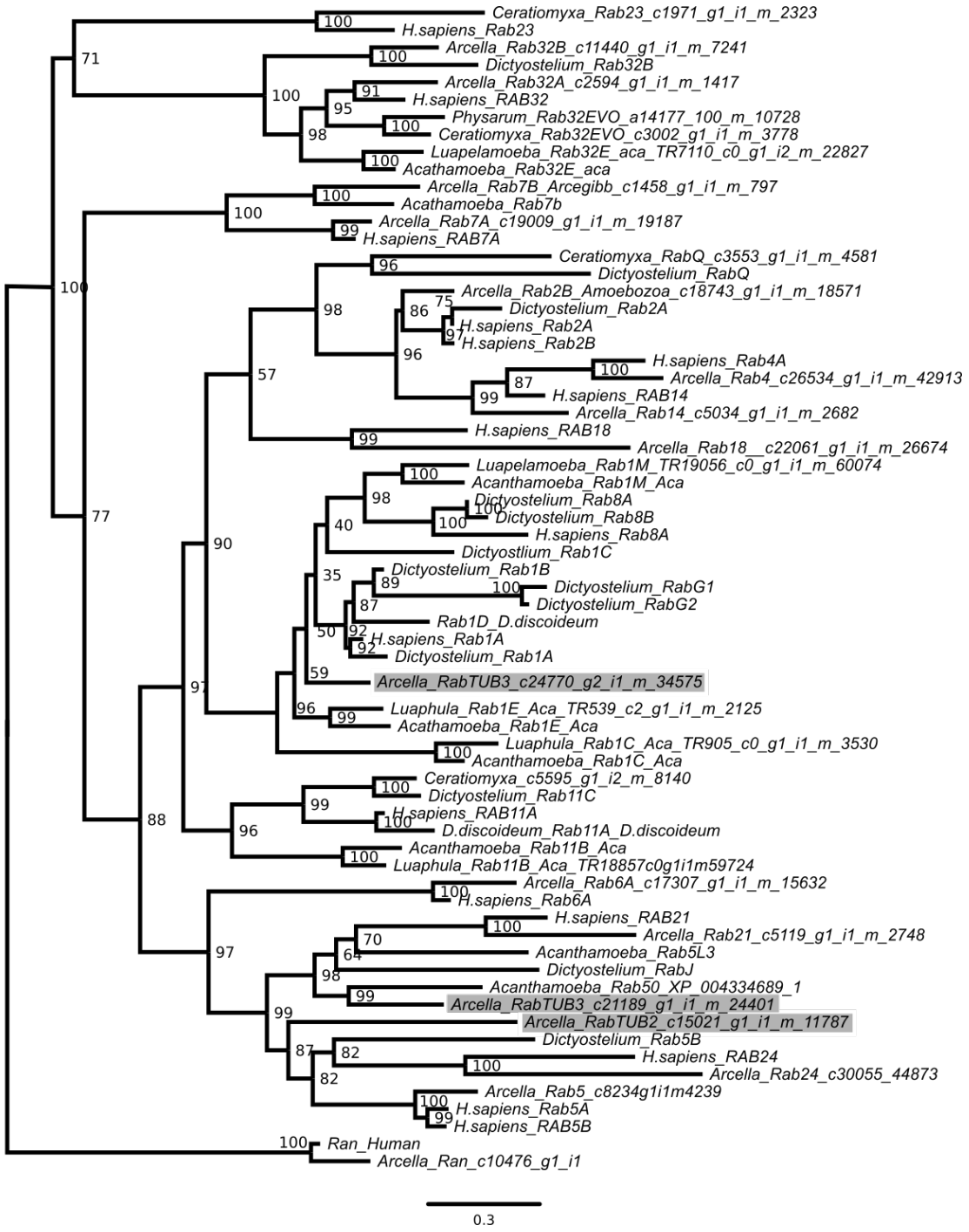

Figure 5.4.1 – Evolutionary relationships of core Rab clades present in Amoebozoa. Each member of Rab GTPase is represented by one Amoebozoa lineage (*Arcella*, *Dictyostelium* or *Achantamoeba*) and *H. sapiens*. The tree is rooted considering Ran as the outgroup. We highlight the Rab members that we identified to be exclusive to Tubulinea. Bootstrap values for branch support are sown. The complete phylogenetic tree, including the Rabs exclusively identified on the genomes of *Dictyostelium* and *Acanthamoeba*, as well as the Elias et al. database, is presented on **Supplementary Figure 5.1**.

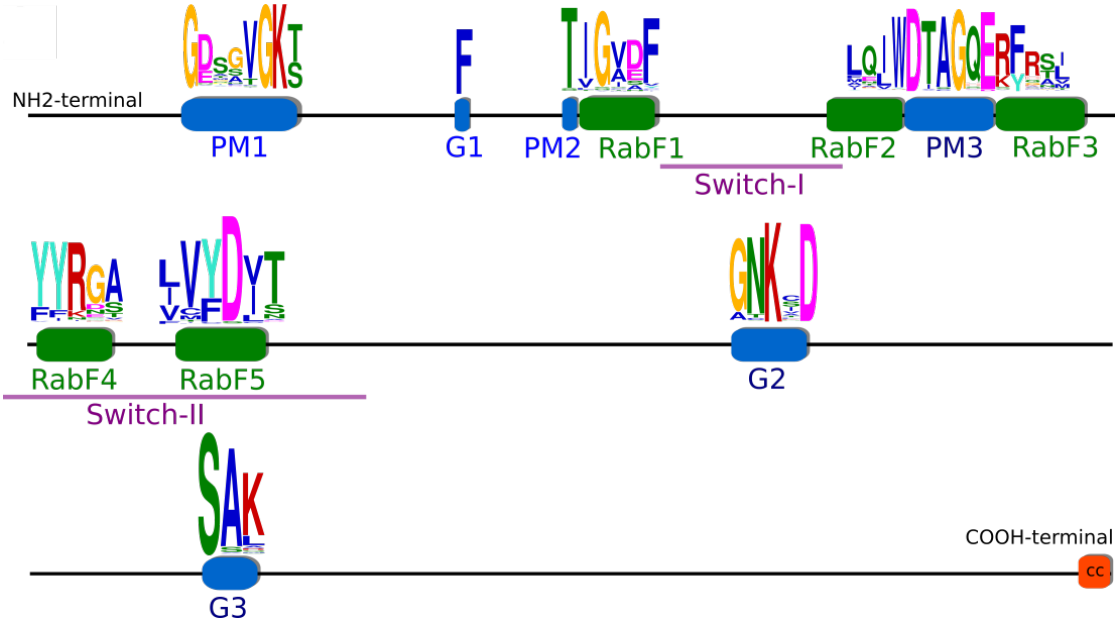

Figure 5.4.2 – Schematic representation of the Rab GTPase seqeunce and its twelve highly conserved motifs, based on Moore (1995) and Stenmark and Olkkonen (2001). The 70 paralogues of Rab GTPase identified on Amoebozoa present the twelve highly motifs, the amino acid sequence of these regions on all Amoeobozoa database are shown. In green, the Rab-specific residues (RabF1-5); In blue the binding motifs, G (guanine-base-binding motif) and PM (phosphate/magnesium-binding motif); In orange, CC terminal motif.

2011). Rabs 5, 21, 24 are involved endocytic pathways (i.e. regulation of early endocytic pathway by Rab5), while Rabs 18, 8, 1, 14, 4, 2, 11 are involved in exocytic pathways (i.e. transport from endoplasmic reticulum (ER) to the Golgi apparatus regulated by Rab1). Rabs 2, 4 and 14 compose a well supported clade but their monophyly is only obtained by alignment and phylogenetic analysis considering only Rabs 2, 4 and 14, including Rab 11 as a out group (not shown).

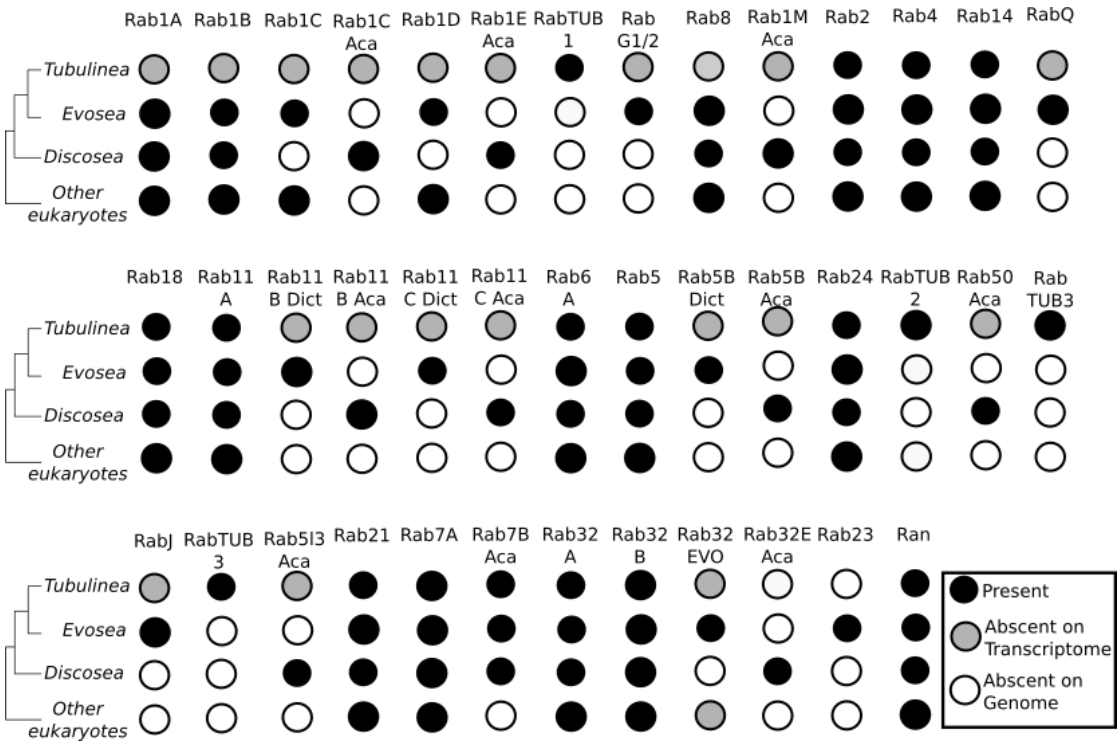

Figure 5.4.3 – Presence and absence plot of the core Rab clades on the major lineages of Amoebozoa. Transcriptomes are informative about the presence of the Rab paralogue, but is an incomplete representation of the organism's genome, thus the absence of a paralogue on a trasncriptome (gray circles) does not necessarily mean that the organism lacks it. Genomes are informative about the absence of a given paralogue (white circle).

5.4.2 Rab GTPases exclusive to the major lineages of Amoebozoa

Rab1C-Aca, Rab1E-Aca, RabTub1, RabG1/G2, Rab1M-Aca, RabQ, Rab11B-Dict, Rab11B-Aca, Rab11C, Rab5B-Dict, Rab5B-Aca, RabTub2, Rab50-Aca, Rab-TUB3, RabJ, Rabl3 -Aca, Rab7B-Aca, Rab32E-Dict, and Rab32E-Aca are present in at least one major lineage of Amoebozoa (Tubulinea, Evosea and Discosea). Three Amoebozoa Rab GTPase members are exclusive to Tubulinea, RabTUB1, RabTUB2, and RabTUB3 (**Figure 5.4.1 and Supplementary Figure 5.1**).

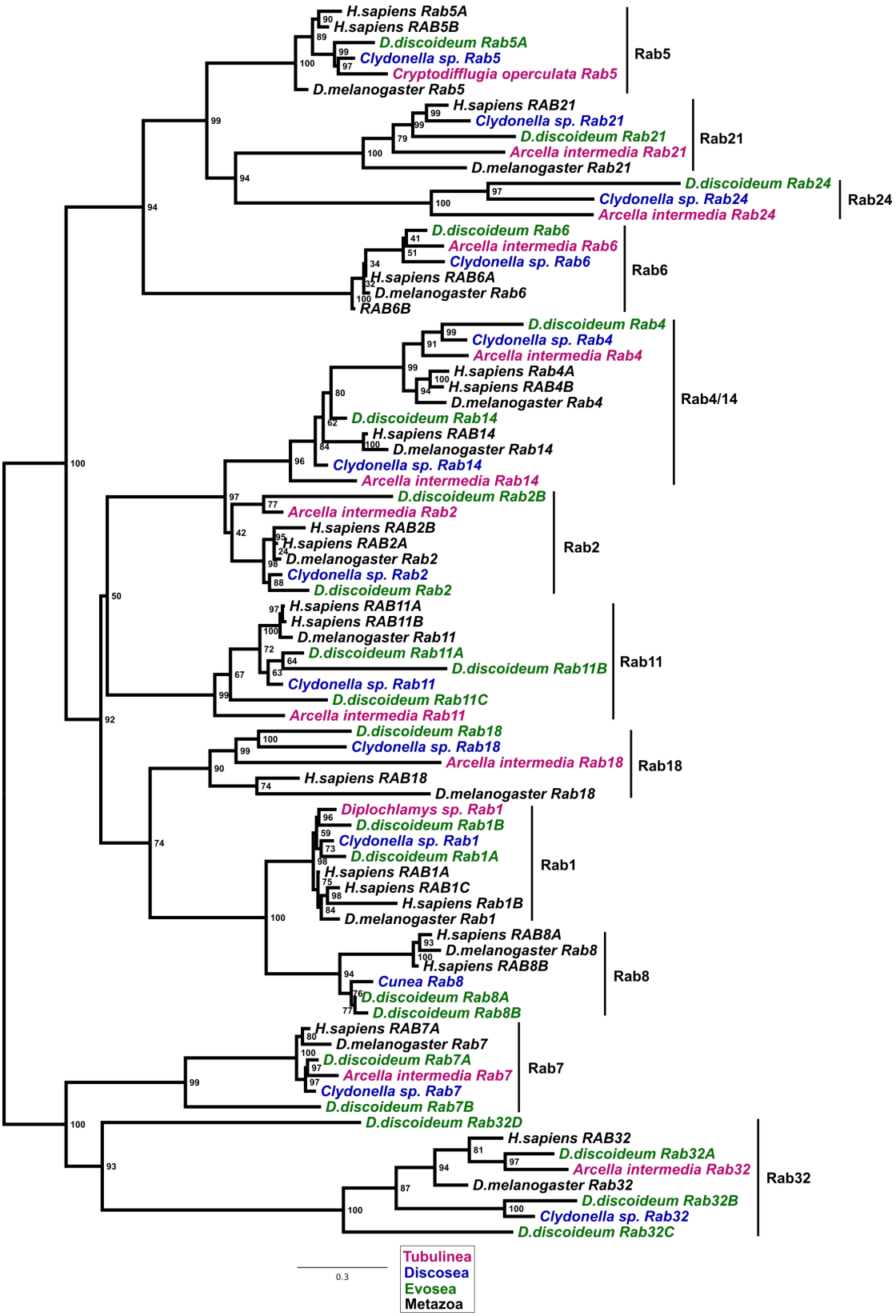

Figure 5.4.4 – Evolutionary relationships of Rab clades that are present in Amoebozoa and diverse eukaryotic group. Each member of Rab GTPase is represented by one member of each major lineage of Amoebozoa and Metazoa (*H. sapiens* and *D. melanogaster*). Bootstrap values for branch support are sown.

RabTUB1 is present in several lineages of Tubulinea (**Supplementary Figure 5.1**) and is closely related to Rab1 (**Figure 5.4.1 and Supplementary Figure 5.1**). RabTUB2 is present in several lineages of Tubulinea (**Supplementary Figure 5.1**) and is closely related to Rab5 and Rab24 (**Figure 5.4.1 and Supplementary Figure 5.1**). RabTUB3 is present in several lineages of Tubulinea (**Supplementary Figure 5.1**) and is closely related to Rab50-Aca, RabJ, Rab5L3, and Rab21 (**Figure 5.4.1 and Supplementary Figure 5.1**). The Rabs assigned as "Aca" in this present study are Rab members annotated on *Acanthamoeba castellanii*. We identified Rab1C-Aca, Rab1E-Aca, Rab1M-Aca, Rab11b-Aca, Rab5B-Aca, Rab50-Aca, Rab5L3-Aca, and Rab32E-Aca exclusively on Discosea lineages, represented on the trees by *Achantamoeba castellanii* and *Luapelamoeba* (**Figure 5.4.1 and Supplementary Figure 5.1**). The Rab7B-Aca is present in the three major lineages of Amoebozoa (**Figure 5.4.2 and Supplementary Figure 5.1**). RabG1/2, RabQ, Rab11B, Rab11C, Rab5B, RabJ, and Rab32E are Rab members annotated on *Dictyostelium discoidem*; these Rabs are present in other lineages of Evosea (**Supplementary Figure 5.1**) and are exclusive to this major group (**Figure 5.4.2 and Supplementary Figure 5.1**).

5.5 Conclusions

Here we demonstrated that a massive expansion of the Rab GTPase family, previously described for *Entamoeba histolytica* (Saito-Nakano et al., 2005), is observed on Amoebozoa as a whole. This expansion has generated a diverse Rab repertoire on Amoebozoa, which makes Amoebozoa a challenging group to comprehensively study the Rab GTPase family. Here we focused on a broad perspective of the Amoebozoa group diversity, identifying the general pattern of expansion on its three major lineages, instead of an effort to identify as much

Rab paralog as possible. We chose a conservative approach to check the identity of the paralogues that we report unequivocally as Rabs, based on the highly conserved motifs. With phylogenetic analysis, we demonstrated the presence of several paralogues that are conserved on diverse eukaryotic lineages and has their cellular function described. Additionally, we identified 22 paralogues exclusive to Amoebozoa, including three exclusive to Tubulinea. The lineages with genomes present several exclusive Rabs, absent even on the transcriptomic data of closely related lineages; three main reasons may explain this observation: i. these paralogues are really specif of these lineages (e.g. *Dictyostelium*), thus are not present on the genomes of other lineages; ii. these paralogues are not expressed by the cell, thus are not identified by transcriptomics; iii. the expression of these paralogues is strictly controlled, thus are rare on transcriptomes. Given the diversity of Rabs present exclusive on Genomes, possibly the three reasons are true, depending on the considered paralogue. It is clear that the 70 paralogues presented here do not fully represent the diversity of Rabs on Amoebozoa. More studies are necessary to address this issue and shed light on the evolution of the Rab GTPase family on Amoebozoa. Functional studies are necessary to understand the role of the newly identified Rab GTPase family member. Our study indicates that the Rab GTPase family is more diverse in Amoebozoa than in Humans, its composition varies greatly even inside the same eukaryotic supergroup and diverse new members must be present in non-model organisms.

5.6 ***References***

Bacon, R. A., Salminen, A., Ruohola, H., Novick, P., Ferro-Novick, S. (1989). The GTP-binding protein Ypt1 is required for transport in vitro: the Golgi apparatus is defective in ypt1 mutants. The Journal of Cell Biology, 109(3), 1015-1022.

Bailey, T. L., Bodén, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Noble, W. S. (2009) "MEME SUITE: tools for motif discovery and searching", Nucleic Acids Research, 37:W202-W208.

Bock, J. B., Matern, H. T., Peden, A. A., & Scheller, R. H. (2001). A genomic perspective on membrane compartment organization. Nature, 409(6822), 839-841.

Diekmann, Y., Seixas, E., Gouw, M., Tavares-Cadete, F., Seabra, M. C., & Pereira-Leal, J. B. (2011). Thousands of rab GTPases for the cell biologist. PLoS computational biology, 7(10), e1002217.

Eathiraj, S., Pan, X., Ritacco, C., Lambright, D. G. (2005). Structural basis of family-wide Rab GTPase recognition by rabenosyn-5. Nature, 436(7049), 415.

Elias, M., Brighouse, A., Castello, C. G., Field, M. C., Dacks, J. B. (2012). Sculpting the endomembrane system in deep time: high resolution phylogenetics of Rab GTPases. J Cell Sci, jcs-101378.

Gabernet-Castello, C., DuBois, K. N., Nimmo, C., Field, M. C. (2011). Rab11 function in Trypanosoma brucei; identification of conserved and novel interaction partners. Eukaryotic cell, EC-05098.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., Vinh, L. S. (2017). UFBoot2: improving the ultrafast bootstrap approximation. Molecular biology and evolution, 35(2), 518-522.

Hutagalung, A. H., & Novick, P. J. (2011). Role of Rab GTPases in membrane traffic and cell physiology. Physiological reviews, 91(1), 119-149.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A., Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nature methods, 14(6), 587.

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.

Lipatova, Z., Hain, A. U., Nazarko, V. Y., & Segev, N. (2015). Ypt/Rab GTPases: principles learned from yeast. Critical reviews in biochemistry and molecular biology, 50(3), 203-211.

Nguyen, J. V. (2015). U.S. Patent No. 9,087,319. Washington, DC: U.S. Patent and Trademark Office.

Pfeffer, S. R. (2005). Structural clues to Rab GTPase functional diversity. Journal of Biological Chemistry, 280(16), 15485-15488.

Plutner, H., Cox, A. D., Pind, S., Khosravi-Far, R., Bourne, J. R., Schwaninger, R., ... Balch, W. E. (1991). Rab1b regulates vesicular transport between the endoplasmic reticulum and successive Golgi compartments. The Journal of cell biology, 115(1), 31-43.

Rutherford, S., & Moore, I. (2002). The Arabidopsis Rab GTPase family: another enigma variation. Current opinion in plant biology, 5(6), 518-528.

Saito-Nakano, Y., Loftus, B. J., Hall, N., & Nozaki, T. (2005). The diversity of Rab GTPases in Entamoeba histolytica. Experimental parasitology, 110(3), 244-252.

Segev, N., Mulholland, J., Botstein, D. (1988). The yeast GTP-binding YPT1 protein and a mammalian counterpart are associated with the secretion machinery. Cell, 52(6), 915-924.

Sela, I., Ashkenazy, H., Katoh, K., Pupko, T. (2015). GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Research, 43(W1), W7-W14.

Stenmark, H., & Olkkonen, V. M. (2001). The rab gtpase family. Genome biology, 2(5), reviews3007-1.

Stenmark, H., Parton, R. G., Steele-Mortimer, O., Lütcke, A., Gruenberg, J., & Zerial, M. (1994). Inhibition of rab5 GTPase activity stimulates membrane fusion in endocytosis. The EMBO journal, 13(6), 1287.

Wennerberg, K., Rossman, K. L., & Der, C. J. (2005). The Ras superfamily at a glance. J cell Sci, 118(5), 843-846.

Zerial, M., & McBride, H. (2001). Rab proteins as membrane organizers. Nature reviews Molecular cell biology, 2(2), 107-117.

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 (Pchelin, 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 hour 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 thecagenesis 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.