Microsporidia is a group of parasites that infect a wide range of species, many of which play important role in the agricultural economics as well as cause many medical troubles. Especially, at least 14 microsporidian species have been confirmed causing dangerous infectious diseases in both immunocompromised and immunocompetent human. Aapproximately 1,400 microsporidia have been reported, which are classified into three groups depending on the host and type of environments, namely the aquasporidia, the terresporidia and the marinosporidia.

Microsporidia was originally classified as fungi by Naegeli (1857). However, the lack of typical eukaryotic components such as mitochondria, Golgi bodies or peroxisomes placed microsporidia together with other amitochondriate protists within the Archezoa kingdom. This "microsporidia-early" hypothesis was further supported by some molecular phylogenies. Despite of those evidences, the placement of microsporidia as an early branching eukaryote was repeatedly debated. Firstly, the phylogeny of microsporidia was prone to suffer from the biases in the reconstruction algorithms. The high evolutionary rate of microsporidian proteins tend to place them together with other fast evolving lineages due to the effect of the long-branch attraction. Furthermore, the presence of the mitochondrial heat shock protein hsp70 proposes the ancestral origin of mitochondrial in microsporidia. As a consequence, it invalidates the main morphological argument for the "microsporidia-early" hypothesis. In 1996, the fungal relationship of microsporidia was at the first time confirmed by molecular phylogenetics studies. Subsequently, several studies based on single or multiple genes placed microsporidia in different positions in the fungal clade, either next to Ascomycota, sister clade of both Ascomycota and Basidiomycota, within the Zygomycota or Crytomycota, or as the sister group of all fungi.

Until now, the exact position of microsporidia in the species tree, especially their relationship to fungi is still unresolved. The difficulties in determining the origin of microsporidia were caused not only by the lack of several cellular components but also by the reduced genomes and metabolism.

Being obligate intracellular parasites, microsporidia successfully reduced their genome and resulted in a genome size in the range of bacterial one. As the smallest eukaryotic genome described so far, the genome of *E.intestinalis* is just 2.3 Mbp, about half the size of the one of *E.coli*. With the low number of protein coding genes (less than 4,000), microsporidia are thought to retain only essential genes required for their survival and development. Furthermore, several of metabolic pathways in microsporidia were missing, such as the citric acid cycle, oxidative phosphorylation, or the *de novo* biosynthesis of nucleotides. This results in an obligatory dependence on many primary metabolites from the hosts. However, the functional uncharacterized hsp70 protein supposes a more complicate genome of the microsporidian ancestor. Consequently, the reduction of microsporidian genomes and metabolic pathways were due to the secondary processes.

Microsporidia become model organisms for studying the minimal eukaryotic genome and the obligate endoparasitic lifestyle in eukaryote. Moreover, a deeper understanding of the physiology of microsporidia is required to better cope with this pathogen. Nevertheless, there is little knowledge regarding the evolutionary and metabolism of microsporidia due to several challenges. Firstly, the obligate endoparasitic lifestyle hinders the study of microsporidia isolated from the host's cells. Yet, no technique for genetic modification is available for microsporidia. Therefore, it raises a need for a comprehensive comparative analysis for the microsporidia.

Hence, we investigated the evolutionary history of the contemporary microsporidia though the reconstruction and investigation of their last common ancestor (LCA). For an intuitive data analysis of the microsporidian phylogenetic profiles, we have developed PhyloProfile, a program that facilitates a dynamic visualization and exploration of the presence-absence patterns of microsporidian genes across several taxa in the tree of life, integrated with other additional information layers. PhyloProfile closes the methodological gap that existed between tools to generate large phylogenetic profiles to delineate the evolutionary history and the contemporary distribution of large – and ultimately complete – gene sets, and the more function-oriented analysis of individual protein. Besides, we propose HamFAS, a new approach for reliable functional annotation of proteins. HamFAS integrates a targeted ortholog search based on the HaMStR algorithm with a weighted assessment of feature architecture similarities (FAS) between orthologs. In brief, we identify for a seed protein orthologs in reference species whose proteins have been functionally annotated based on manually curated assignments to KEGG Ortholog (KO) groups. The FAS scores between the orthologs and seed proteins are calculated. Subsequently, we compute pairwise FAS scores for all reference proteins within a KO group. A group's mean FAS score serves then as cutoff that must be exceeded to warrant transfer of its KO identifier to the seed. We benchmarked the performance of HamFAS using a manually curated KO annotated yeast protein set. HamFAS yielded the best precision (98.5%) when compared with two state-of-the-art annotation tools provided by KEGG, KAAS and BlastKOALA. Furthermore, the ability of inferring orthologs between distantly related species using hidden Markov model profiles allows HamFAS achieving a higher sensitivity than the two other tools, with 47.5% more proteins annotated by HamFAS than by KAAS or BlastKOALA in average.

We pursued a phylogenetic approach to reconstruct a robust phylogeny of microsporidia and further identified the placement of microsporidia in the eukaryotic tree of life. To approach that, we used the data available from eleven microsporidia sequencing projects at the Broad Institute and the Joint Genome Institute. We successfully identified a microsporidia core set comprising of 80 one-to-one orthologous groups, which was used as a basic for the phylogeny reconstruction. Our reconstructed phylogenetic tree from a diverse set of taxa, including 48 fungi representing the full fungal diversity as well as other representative species for both unikonts and bikonts, strongly supported the hypothesis that microsporidia form the early branch of fungal clade. The analyzed data explained this microsporidia-fungi relationship significantly better than any other hypotheses.

Our phylogenetic profile study revealed the dynamic evolutionary history of microsporidian genomes. From 2% the proteins in the compact microsporidia *E.intestinalis* to 49% the proteins of *E.eadis* were not found in other microsporidian species. Those proteins were thought to be useful for investigate the host range of each microsporidium. On a contrary, the other retained proteins were expected to be essential and evolutionary old. We estimated 94% of the microsporidian LCA proteins that could be tracked back to the last eukaryotic common ancestor and 3% proteins were shared the common ancestor with fungi. Only 3% LCA proteins are microsporidia specific, which supposed to be important for the obligate parasitic lifestyle of microsporidia.

The functional annotation and metabolic pathway analysis of the microsporidian LCA gave more insight about the adaptation of microsporidia into their parasitic lifestyle, likewise the origin of the reduction in microsporidia. The presence of E1, E3 components of the pyruvate dehydrogenase complex and the mitochondrial hsp70 protein once again confirmed the hypothesis about the mitochondrial origin of the ancestral microsporidia. In additional, several novel proteins have been found in the microsporidian LCA that could complement some gapped metabolic pathways. They suggested a more complex genome and metabolism in the LCA. However, the microsporidian LCA still lacked many main pathways such as the effective energy production TCA cycle, or key enzymes that required for *in vivo* synthesis of critical metabolites like purines and pyrimidines. In summary, we suppose that the parasitic lifestyle already occurred in the microsporidian LCA and the reduction of microsporidia, therefore, were indeed an ancestral state following by further losses and gains during the evolution of each individual microsporidian lineage.

Die Mikrosporidien sind sporenbildende Parasiten, die verschiedenen Organismen infizieren. Die Mikrosporidiose beeinträchtigen die Agrarökonomie als auch viele medizinische Probleme. Insbesondere wurden mindestens 14 Mikrosporidien bestätigt, die gefährliche infektiöse Krankheiten sowohl bei immunkompromittierten als auch bei immunkompetenten Menschen verursachen. Etwa 1.400 Mikrosporidien wurden beschrieben, die abhängig vom Wirtsarten und der Art der Umgebung in drei Gruppen zugeordnet sind, nämlich den Aquasporidien, den Terresporidien und den Marinosporidien.

Mikrosporidien wurden ursprünglich von Naegeli (1857) als Pilz klassifiziert. Wegen des Mangels vieler typischen eukaryotischen Komponenten wie Mitochondrien, Golgi-Apparat oder Peroxisomen wurden Mikrosporidien nachher jedoch zusammen mit anderen amitochondrischen Protisten innerhalb des Archezoa-Reiches gruppieren. Diese "Microsporidia-early" Hypothese wurde durch einige molekulare Phylogenien unterstützt. Trotz dieser Beweise wurde die Platzierung von Mikrosporidien als früh verzweigende Eukaryoten erörtert. Zum einem könnten die Phylogenie von Mikrosporidien durch die Rekonstruktionsalgorithmen verzerren. Die hohe evolutionäre Rate der mikrosporidischen Proteinen neigt dazu, sie zusammen mit anderen schnell evolvierten Abstammungen aufgrund der Wirkung des Baumbildungsartefakts (long-branch-attraction) zu platzieren. Darüber hinaus schlägst das mitochondriale Hitzeschockprotein hsp70 den Ursprung der Mitochondrien vor. Als Konsequenz invalidierte es das morphologische Hauptargument für die "Microsporidia-early" Hypothese. Im Jahr 1996 wurde die Verwandtschaft zwischen Pilzen und Mikrosporidien erstmals durch molekulare phylogenetische Studien bestätigt. Anschließend wurden verschiedenen Untersuchungen, die auf einzelnen oder mehreren Genen basierten, Mikrosporidien in verschiedenen Positionen in der Pilzklade platziert, entweder neben Ascomycota, als Schwestergruppe von beiden Ascomycota und Basidiomycota, innerhalb der Zygomycota oder Crytomycota, oder als Schwestergruppe aller Pilze.

Bisher ist die exakte Position der Mikrosporidien im Spezienbaum, insbesondere deren Pilzverwandtschaft immer noch ungelöst. Die Schwierigkeiten bei der Bestimmung des Ursprungs von Mikrosporidien wurden nicht nur durch mehrere mangelnden zellulären Komponenten verursacht, sondern auch durch ihre reduzierten Genome und den Metabolismus.

Als obligatorische intrazelluläre Parasiten verminderten Mikrosporidien ihren Genome erfolgreich und führten zu einer Größe im Bereich der bakteriellen Genome. Das kleinste beschriebene eukaryotische Genom von E.intestinalis ist nur 2,3 Mbp, etwa halb so groß wie das von E.coli. Der geringen Anzahl von protein-kodierenden Genen (weniger als 4.000) annimmt, dass Mikrosporidien nur essentielle Gene enthalten, die für ihr Überleben und ihre Entwicklung benötigt werden. Darüber hinaus fehlten mehrere Stoffwechselwege in Mikrosporidien wie der Zitronensäurezyklus, die oxidative Phosphorylierung oder die de novo Biosynthese von Nukleotiden. Dies führt zu einer obligaten Abhängigkeit von vielen primären Metaboliten von den Wirten. Das uncharakterisierte hsp70-Protein setzt jedoch ein komplizierteres Genom des mikrosporidischen Vorfahren voraus. Folglich waren die Reduktion von mikrosporidischen Genomen und Stoffwechselwegen auf die Sekundärprozesse zurückzuführen.

Mikrosporidien werden zu Modellorganismen für die Untersuchung des minimalen eukaryotischen Genoms und der obligaten endoparasitischen Lebensweise in Eukaryoten. Darüber hinaus ist ein tieferes Verständnis der mikrosporidischen Physiologie erforderlich, um mit diesem Pathogen besser zu bewältigen. Dennoch gibt es aufgrund verschiedener Herausforderungen geringe Kenntnisse über die mikrosporidische Evolution und ihren Metabolismus. Zum einen behindert die obligate endoparasitische Lebensweise die Untersuchung von Mikrosporidien, die aus den Wirtszellen isoliert wurden. Für Mikrosporidien steht jedoch keine Technik zur genetischen Modifikation zur Verfügung. Daher bedarf es einer umfassenden vergleichenden Analyse der Mikrosporidien.

Daher untersuchten wir die Entwicklungsgeschichte der heutigen Mikrosporidien durch die Rekonstruktion und Untersuchung ihrer letzten gemeinsamen Vorfahre (LCA). Für eine intuitive Datenanalyse der phylogenetischen Profile der Mikrosporidien entwickelten wir PhyloProfile, ein Programm für dynamische Visualisierung und Untersuchung der mit anderen zusätzlichen Informationsebenen integrierten An- und Abwesenheitsmuster der mikrosporidischen Genen über mehrere Taxa im Lebensbaum ermöglicht. PhyloProfile schließt die methodologische Lücke, die zwischen Softwares bestand, um große phylogenetische Profile zu erzeugen, um die evolutionäre Geschichte und die gegenwärtige Verteilung großer - und schließlich vollständiger - Genmengen und die funktionalere Analyse einzelner Proteine zu beschreiben.

Außerdem schlagen wir HamFAS vor, einen neuen Ansatz zur zuverlässigen funktionellen Annotation von Proteinen. HamFAS integriert eine gezielte Ortholog-Suche basierend auf dem HaMStR-Algorithmus mit einer gewichteten Bewertung von Feature Architektur Ähnlichkeiten (FAS) zwischen Orthologen. In Kurze identifizieren wir für ein Seedprotein Orthologe in Referenzspezies, deren Proteine anhand von manuellen Zuordnungen zu KEGG-Ortholog (KO) -Gruppen annotiert wurden. Die FAS-Werte zwischen den Orthologen und den Seedproteinen werden berechnet. Anschließend berechnen wir paarweise FAS-Werte für alle Referenzproteine innerhalb einer KO-Gruppe. Der mittlere FAS-Wert einer Gruppe dient dann als einen Cutoff, der überschritten werden muss, um die Übertragung seines KO-Identifikation an den Seed zu rechtfertigen. Wir benchmarkten die Performance von HamFAS mit einem manuell kuratierten KO-annotierten Hefeprotein-Set. HamFAS ergab die beste Genauigkeit (98,5%) im Vergleich zu zwei State-of-the-Art Annotationstools von KEGG, KAAS und BlastKOALA. Darüber hinaus konnte HamFAS mit der Fähigkeit der Profilen des Hidden Markov Modelles, Orthologe zwischen entfernt verwandten Arten zu bestimmen, eine höhere Empfindlichkeit als die beiden anderen Tools erreichen, wobei 47,5% mehr Proteine von HamFAS annotiert sind als durchschnittlich von KAAS oder BlastKOALA.

Wir verfolgten einen phylogenetischen Ansatz, um eine robuste Phylogenie von Mikrosporidien zu rekonstruieren und weiter identifizierten die Platzierung von Mikrosporidien in der eukaryotischer Baum. Um dies zu erreichen, verwendeten wir die Daten von elf Microsporidia-Sequenzierungsprojekten am Broad Institute und am Joint Genome Institute. Wir erfolgreich identifizierten einen mikrosporidischen Geneset, der aus 80 eins-zu-eins-orthologen Gruppen besteht, die als Grundlage für die Phylogenie-Rekonstruktion verwendet wurden. Unser rekonstruierter phylogenetischer Baum aus einer Vielzahl von Taxa, darunter 48 Pilze, die die gesamte Pilzdiversität sowie andere repräsentative Arten sowohl für Unikonten als auch für Bikonten repräsentieren, unterstützte kräftig die Hypothese, dass Mikrosporidien den frühen Zweig der Pilzgruppe bilden. Die analysierten Daten erklärten diese Mikrosporidien-Pilz-Verwandtschaft signifikant besser als alle anderen Hypothesen.

Unsere Studie der phylogenetischen Profilen zeigte die dynamische Evolutionsgeschichte der mikrosporidischen Genomen. Von 2% der Proteine in der kompakten Mikrosporidie E.intestinalis bis zu 49% die Proteine von E.adis wurden in anderen Mikrosporidien nicht gefunden. Diesen Proteinen wurde angenommen, dass sie dienlich sind, um den Wirtsbereich jeder Mikrosporidium zu untersuchen. Im Gegensatz dazu wurde anderen Proteinen essentiell und evolutionär alt erwartet. Wir schätzten 94% der Proteinen des mikrosporidischen LCAs, die auf den letzten gemeinsamen Vorfahren der Eukaryoten zurückverfolgt werden konnten, und 3% Proteine, die den gemeinsamen Vorfahren mit Pilzen teilten. Nur 3% LCA-Proteine sind Mikrosporidien-spezifisch, was für die obligatorische parasitäre Lebensweise von Mikrosporidien wichtig sein sollte.

Die funktionelle Annotation und die Analyse der Stoffwechselwege des mikrosporidischen LCAs gaben mehr Verständnis über die Anpassung von Mikrosporidien an ihre parasitäre Lebensweise, ebenso der Ursprung der Reduktion von Mikrosporidien. Die Anwesenheit von E1, E3-Komponenten des Pyruvat-Dehydrogenase-Komplexes und des mitochondrialen hsp70-Proteins bestätigte erneut die Hypothese über den mitochondrialen Ursprung der anzestralen Mikrosporidien. Zusätzlich wurden mehrere neue Proteine im mikrosporidischen LCA gefunden, die einige Lücken Stoffwechselwege ergänzen könnten. Sie wiesen ein kompliziertes Genom und aufwändigen Metabolismus im LCA hin. Der mikrosporidischen LCA fehlten jedoch immer noch viele primären Stoffwechselwege, wie der effektiven Energieproduktion Citratzyklus, oder Schlüsselenzyme, die für die in vivo Synthese von kritischen Metaboliten wie Purinen und Pyrimidinen benötigt werden. Zusammenfassend nehmen wir an, dass die parasitische Lebensweise bereits in der mikrosporidischen LCA vorkam. Und die Reduktion der Mikrosporidien war ein angestammter Zustand, der durch weitere Ausfälle und Zuwachs der Genen während der Entwicklung jeder einzelnen mikrosporidischen Linie folgte.