Animals and Fungi:

Common Origin, but Independent Approaches to Multicellularity

Main applicants: Gertraud Burger (Montreal), Michael W. Gray*****,† (Halifax), Peter Holland† (Oxford, UK), Nicole King (Berkeley), B. Franz Lang (Montreal), Andrew Roger (Halifax), Inaki Ruiz-Trillo (Halifax)

Collaborators: Mary Berbee (Vancouver), Barry Leadbeater (Birmingham, UK), Hervé Philippe (Montreal), Emma Steenkamp (Pretoria, South Africa)

Other potential collaborators (to be confirmed): Sandra L. Baldauf (York, UK), Scott E. Baker (Pacific Northwest National Laboratory), Edward L. Braun (Florida), Bruce Birren (Fungal Genome Initiative, Berkeley), Laurent Duret (Lyon, France), John Taylor (Fungal Genome Initiative, Berkeley), Ken Wolfe (Dublin, Ireland)

[NOTE: this initial list of investigators will be expanded as appropriate]

†Member, NHGRI Working Group on Comparative Genome Evolution

***** Coordinator; Room 8F-2, Sir Charles Tupper Medical Building, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax NS B3H 1X5; tel. (902)494-2521; FAX (902)494-1355; email: m.w.gray@Dal.ca

Abstract

Motivation: Numerous recent publications have demonstrated convincingly that the most effective approach to understanding the origin of the human genome is comparative animal genomics (e.g., (1-3)). Similarly, to understand the origin of multicellularity and animal-specific gene families that are relevant to human development, health and complex diseases, we need to trace back these features to their ancestral unicellular eukaryotes (protists). It is now well established that animals share a common single-celled ancestor with fungi, and that in the two sister lineages multicellularity arose independently. This insight prompts a host of intriguing, fundamental questions:

- What genome-level changes coincide with the rise of multicellularity in these two kingdoms, and to what extent are these changes different in animals and fungi?
- Which particular genes or genomic features, present in the unicellular ancestors of animals and fungi, were prerequisites for the emergence of multicellularity?
- What is the origin of animal- and fungal-specific gene families?

Clearly, an incisive comparative genomics analysis requires genomic information from very early diverging animals and fungi, as well as their specific unicellular relatives.

Among the unicellular protists most closely related to animals and/or fungi, only a single group (choanoflagellates) has so far been selected for high-coverage genome sequencing. The unicellular bacterivorous choanoflagellate *Monosiga brevicollis* is in the JGI sequencing pipeline. The distantly related *M. ovata* was approved by NHGRI for full genome sequencing, but progress has been slow due to difficulties in supplying sufficiently clean nuclear DNA (note: 'genus *Monosiga*' is a misnomer, as this assemblage is in fact polyphyletic). To address the questions listed above, it will be necessary to sample both within and outside of the choanoflagellates. This sampling will allow general features of each clade to be distinguished from the specializations of particular species, thereby enabling insightful comparison to animals and fungi. Within choanoflagellates, it would be most informative to investigate the genome of a marine loricate species (these have complex bio-architectures and fall basally within the group) and also a colonial choanoflagellate (such as *Salpingoeca* or *Codosiga*). The latter organisms not only represent an important intermediate between uni- and multicellularity, but we also expect that they will be more amenable to genome sequencing than *Monosiga ovata*. In addition to choanoflagellates, our recent molecular explorations have identified several other animal- and fungus-related protists that are attractive for full genome sequencing (**Fig. 1**). Among these candidates are unicellular eukaryotes (i) that diverge prior to Metazoa (*Capsaspora*, *Amoebidium* and *Sphaeroforma*; (4,5)); (ii) that diverge prior to Fungi (amoeboids in the genus *Nuclearia* (6)); and (iii) that are members of the enigmatic apusomonads (e.g., *Amastigomonas*). The latter group has been postulated to diverge prior to the separation of animals and fungi (7). Two other unicellular taxa (*Ministeria* and *Corallochytrium*) are also positioned basally to one or both of the multicellular groups, but are not currently amenable to full genome sequencing.

To investigate commonalities and differences underlying multicellularity in animals and fungi, available fungal genome information (mostly from ascomycetes and basidiomycetes) is insufficient. High-coverage genome data are required for basally diverging fungal groups: (i) chytridiomycetes (chytrids), a highly diverse assemblage of zoosporic fungi (e.g., *Allomyces* and *Spizellomyces*); and (ii) zygomycetes, another large, deep divergence within Fungi. So far, only a single high-coverage zygomycete genome sequence, that of *Rhizopus oryzae*, is available (NHGRI and JGI currently list a single species within chytrids, *Batrachochytrium dendrobatidis*, which is rather distantly related to *Allomyces* and *Spizellomyces*).

Some of the organisms proposed here for genome sequencing have direct relevance to human health. *Capsaspora owczarzaki* is a parasite of a snail that serves as the intermediate host for *Schistosoma*, a digenean platyhelminth that is the causative agent of schistosomiasis, a disabling and often life-threatening disease that afflicts more than 200 million people worldwide. Knowledge of the *Capsaspora* genome sequence would provide insights into host-parasite co-evolution in this system but, more importantly, might also be used to develop strategies in the control of this devastating human parasite. Likewise, the zygomycete *Mortierella* could be developed as a model for zygomycosis, a disease that has been rising steadily in incidence over the past 10 years. *Nuclearia* species are pathogens of commercially exploited fish, as are *Amoebidium* and *Sphaeroforma*. The latter two organisms are the few culturable members of a large group of fish parasites that remain otherwise inaccessible to genomic investigation.

Aims: This proposal aims at filling, in a systematic manner, crucial gaps in our understanding of animal and fungal evolution, thus providing for the first time comprehensive data with which to conduct meaningful comparative genomic analyses of animals, fungi and their unicellular relatives. In particular, we propose that genomes be sequenced to high coverage (6x) from the following species for which preliminary relevant information is available:

- *Capsaspora owczarzaki*, *Amoebidium parasiticum*, *Sphaeroforma arctica* (divergences prior to animals);
- *Nuclearia simplex* (divergence prior to fungi);
- the chytridiomycete fungi *Allomyces macrogynus* and *Spizellomyces punctatus*;
- the zygomycete fungus *Mortierella verticillata*.

We further propose to explore the feasibility of genome projects in

- *Amastigomonas*,
- a colonial choanoflagellate (*Salpingoeca* or *Codosiga*) and
- a loricate choanoflagellate (*Stephanoeca or Acanthocoepis*).

The comprehensive and coherent genomic data that we propose to generate will provide a unique resource for fundamental life sciences research, including parasitology, comparative genomics, pathogenomics, molecular evolution, macromolecular modeling, gene discovery, and development of new experimental model systems.

Complementarity with other genome projects: The Fungal Genome Initiative (FGI) has a strong focus on ascomycetes and basidiomycetes. The FGI currently proposes the sequencing of one chytridiomycete and one zygomycete genome. Within the chytrids and zygomycetes, we have chosen additional species that are relevant to our research questions, but that also complement the FGI's list of organisms. Evidently, the range of fungal species that we selected for sequencing does not cover the chytrids and zygomycetes as a whole. Comprehensive coverage of Fungi will presumably remain the objective of future FGI proposals. We intend to invite FGI members to participate in the current project, in order to benefit from their expertise and to avoid potential overlap.

Table 1: Features of the species proposed for genome sequencing

Species [Taxonomy]	**Nuclear genome size[1] (Mb)**	**ESTs (# of available readings)**	**% GC content[2]**	**Mitochon-drial DNA sequence**	**Culture**	**Other features and relevance**	**Provision of nuclear DNA**	**Priority: Full (6x)/ Draft (2x); Pilot data**
Capsaspora owczarzaki [choanozoan relative]	20-24[3]	8,870	50	complete	axenic	Medical relevance (schistosomiasis)	easy	**(1) 6x coverage**
Allomyces macrogynus [Fungi, Chytridio-mycota]	~ 30[5]	5,078	62	complete	axenic	Best-developed chytrid model, ideal for teaching purposes[6]. Typical flagellar apparatus; photo- and chemotaxis	easy; nuclear DNA available	**(2) 6x coverage**
Sphaero-forma arctica [Ichthyo-sporea]	n.d.	8,006	49	n.d.	axenic	'Multicellular-like' features	easy	**(3) 6x coverage;** requires pilot sequencing
Nucleraria simplex [Nucleariida]	n.d.	3,313	36	complete	monoxenic (can be grown on various bacterial strains)	Pathogen of fish; unique phylogenetic position, prior to the divergence of Fungi and close to fungi/animals	difficult; ~ 30 µg of clean nuclear DNA available	**(4) 6x coverage;** requires pilot sequencing
Amoebidium parasiticum [Ichthyo-sporea]	n.d.	3,632	54	almost complete	axenic	'Multicellular-like' features	easy	**(5) 2x coverage;** requires pilot sequencing
Spizello-myces punctatus [Fungi; Chytridio-mycota]	n.d.	5,365	50	complete	axenic	Molecular studies of tRNA editing[7]; typical flagellar apparatus	easy	**(6) 2x coverage;** requires pilot sequencing
Mortierella verticulata [Fungi, Zygomycota]	n.d.	5,724	53	complete	axenic	Medical relevance (zygomycosis in animals)	easy	**(7) 2x coverage;** requires pilot sequencing
***Salpingoeca* or *Codosiga* sp.** [Choano-flagellata]	n.d.	-	n.d.	n.d.	monoxenic culture are easily established	Colonial choanoflagellate; typical flagellar apparatus	difficult, but much easier than *Monosiga*[4]	**(8) 6x coverage;** requires pilot sequencing
***Amastigo-monas* sp.** [Apusozoa]	n.d.	-	n.d.	n.d.	two monoxenic species growing well in culture are available	Minimally derived apusomonad; likely diverging prior to opisthokonts	difficult, but much easier than *Monosiga*[4]	**(9) 6x coverage;** requires pilot sequencing
Stephanoeca* or *Acantho-coepis sp. [loricate choa-noflagellate]	n.d.	-	n.d.	n.d.	ATCC (growing on mixed food bacteria)	Basally diverging choanoflagellate with complex exoskeleton	difficult	**(10) 2x coverage;** requires pilot sequencing

[1] In cases lacking precise estimates , we would propose to determine both the genome size and the feasibility of the genome project through pilot sequencing.

[2] Based upon EST sequences.

[3] Estimated by pulsed-field gel electrophoresis; see **Fig. 2**.

[4] Colonial choanoflagellates attach to surfaces, behavior that facilitates elimination of food bacteria; the same consideration applies to apusomonads, which slide on surfaces in order to catch bacteria. A condition for efficient removal of food bacteria is that the latter do not attach to surfaces , and that the eukaryote in question will ingest the food bacteria completely.

[5] (8)

[6] e.g., (9)

[7] (10-12)

[8] (13)

Introduction

Phylogenetic analyses have shown convincingly that the eukaryotic clades Metazoa (animals) and Fungi derive from a common ancestor that existed ~1 billion years ago. The taxonomic supergroup that contains animals and fungi, Opisthokonta, also includes various unicellular taxa such as Choanoflagellata, Ichthyosporea, *Capsaspora*, nucleariids and *Ministeria* (**Fig. 1**). Within Fungi, two major but little explored groups stand out: (i) 'zoosporic' fungi, which others have classified as protists related to fungi (due to their developed flagellar apparatus, a 'primitive' character shared with Metazoa and many of their unicellular relatives; (14,15)), and (ii) zygomycetes. The well-studied fungi such as Basidiomycota ('mushrooms'), Euascomycota (e.g., *Neurospora* and *Aspergillus*), budding yeasts and fission yeasts are sister groups of zygomycetes.

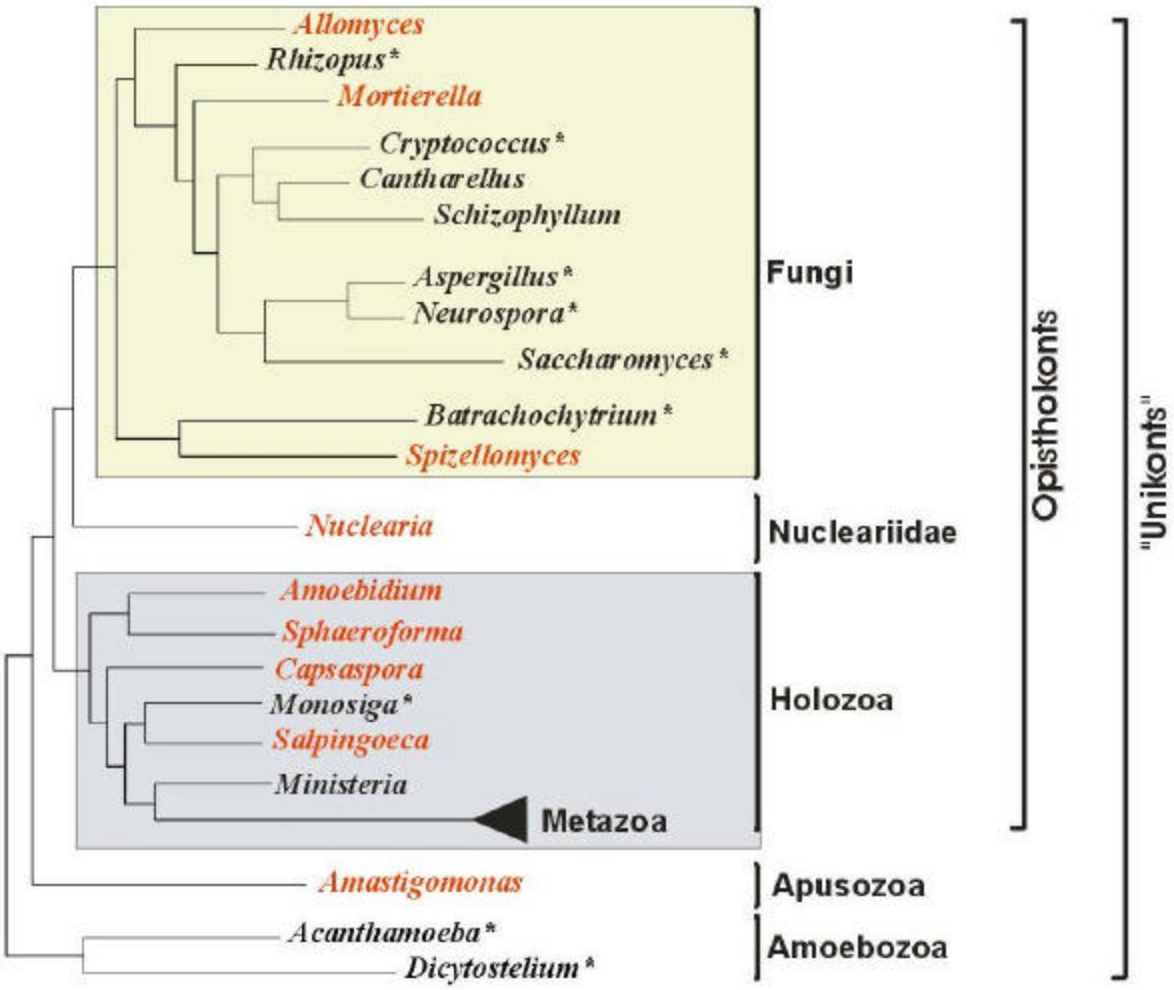

Figure 1: **Phylogenetic tree of opisthokonts based on molecular data.**
The tree is based on published (4-7) and unpublished information available to the authors of this proposal (Lang BF, Burger G, Ruiz-Trillo I, Steenkamp E). Species for which we propose genome sequencing are labeled in **red**; asterisks (*) indicate species for which genome sequences are either available or that are slated for completion.

This proposal focuses on comprehensive comparative genomics of opisthokonts. The data generated in this project will provide insights into fundamental biological questions, from a genomics and comparative-molecular perspective:

- What was the genomic make-up of the opisthokont ancestor that gave rise to animals and fungi?
- Which are the specific gene sets that correlate with multicellularity in animals and fungi, and to what extent do the sets overlap?
- Is there a genetic predisposition in opisthokonts to 'invent' multicellularity?
- What are the trends and strategies in the evolution of genome complexity within the two kingdoms, and is there a correlation with the evolution of the body plan?
- What is the origin of animal- and fungal-specific gene families?

The following section provides a brief introduction to these issues.

Origin of multicellular animals. Few evolutionary transitions have garnered so much attention from the scientific community. Despite ongoing controversies about the nature of the earliest putative soft-bodied macro-organisms, dated around 570-550 million years ago (Ediacaran fauna), paleobiological evidence suggests that at least some of these fossils are derived from multicellular animals (16,17). It is now widely accepted that most of the metazoan phyla appeared in the fossil record during a brief period of time known as the Cambrian explosion, approximately 520 million years ago (18). The potential selective advantages that fostered the emergence of multicellular animals have been widely discussed, but the genetic basis for this key evolutionary transition remains elusive.

Previous ideas of a stepwise acquisition of structural and genetic complexity within Metazoa have been revised in the light of new genetic and genomic data from basal metazoan lineages. The accumulating data reveal that most of the key regulatory genes and signaling molecules deployed during development of multicellular animals (and affected in many human genetic disorders) are widely conserved across Metazoa, even to the most basally branching animal taxa. Yet many of these ancient gene families have not been detected so far outside the metazoans, raising crucial and unsolved questions about their origins. For example, within the homeobox superfamily, the entire ANTP and PRD classes have thus far only been detected in multicellular animal genomes (these include Hox, Dlx, En, Emx, Msx, Otx and others). The same is true for many other transcription factors (e.g. T-box, Pax families) and intercellular signalling molecules (Wnt, FGFs, ephrins, TGFb, Hh genes, delta). For example, even the cnidarians (simple diploblastic and deep-branching metazoans) have been found to possess a small set of Hox genes that control early axial patterning in animals (19) and a large diversity of Wnt genes involved in cell-cell signaling (20,21). Similarly, many structural features of cell and tissue organization (such as desmosomes, cell junctions and extracellular matrix proteins) are ubiquitously distributed throughout Metazoa, e.g., (22)).These findings raise the question of whether homologs, or perhaps progenitors, of these regulatory and structural proteins could be present in the unicellular relatives of animals. It is certainly conceivable that some of the key gene families used in animal development had already appeared/diversified in their unicellular ancestors (reviewed in (23)). Alternatively, it is likely that some of these proteins are truly metazoan-specific, and that they were assembled by re-arrangement of smaller protein modules in the ancestors of animals. One such example is the *hedgehog* gene family (encoding key signaling proteins), for which a homolog to one part of the encoded protein has been found in a choanoflagellate (24). The current proposal aims at testing these hypotheses, for all metazoan gene families, through genome comparisons among diverse metazoans and their protistan relatives.

Origin of fungi. Fungi comprise organisms that are exceptionally diverse both morphologically and biochemically, and include some of the best studied of eukaryotic model systems, including the brewer's/baker's yeast *Saccharomyces cerevisiae*, the fission yeast *Schizosaccharomyces pombe*, and the common bread molds *Neurospora crassa* and *Aspergillus nidulans*. Nevertheless, there is no generally accepted and precise definition of the taxon Fungi (e.g., (25-27). This situation is mostly attributable to the great variation of fungal morphology and lifestyle, which has led to diverse classification schemes. Only molecular phylogenies have shown beyond doubt that organisms previously termed fungi (e.g., oomycetes) are actually members of unrelated eukaryotic lineages. In turn, numerous species that were traditionally associated with protists ('protoctists') are now recognized as fungi. Examples include chytridiomycetes (e.g., (14,28)), *Microsporidia* (29), *Pneumocystis carinii* (30) and *Hyaloraphidium curvatum* (31,32). Finally, molecular phylogenies have recently placed previously unclassified organisms such as *Nuclearia simplex* basally to Fungi ((6); our unpublished data).

Taxa that are now known to be affiliated with fungi include unicellular organisms with yeast-like, rhizoid-carrying and biphasic-zoosporic properties, as well as amoeboid cells, filamentous molds and multicellular mushrooms. This observation predicts additional fungal relatives among poorly characterized amoeboid and flagellated opisthokonts.

Inspection of fungal phylogenetic trees does not indicate an overall evolutionary trend towards multicellularity; nor is it clear whether multicellularity evolved only once and was lost several times independently, or emerged several times independently within Fungi.

Origin of opisthokonts. The origin of opisthokonts is a highly controversial topic. According to phylogenetic analysis based on rRNA sequences, the closest opisthokont neighbors are the apusomonads ((7); **Fig. 1**). However, these conclusions are tenuous at best. A number of molecular phylogenies infer, with significant support, that the closest sister-group of opisthokonts is Amoebozoa (including slime molds such as *Dictyostelium*, a variety of lobose amoebae such as *Acanthamoeba*, and the human parasite *Entamoeba*). This inference has prompted a classification of opisthokonts plus Amoebozoa as the superkingdom 'unikonts', to the exclusion of all other eukaryotes (which are termed 'bikonts'). This distinction, if substantiated, may be key to tracing the evolutionary root of eukaryotes (33).

The nuclear genomes of two amoebozoans (*Entamoeba histolytica* and *Dictyostelium discoideum*) have been fully sequenced and that of another (*Acanthamoeba*) is underway. These genomes, together with the data to be generated in this project, will provide a unique resource for investigating the origin of opisthokonts and assessing the validity of the potentially fundamental 'unikont'/'bikont' concept.

Conclusion: Numerous comparative genomics studies have been applied in yeasts and other eukaryotes (e.g., (34-36)) and have proven to constitute a most powerful approach to understanding genome evolution. The studies that we propose on a taxonomically broad lscale are expected to provide insights at an even deeper level. This proposal is intended to provide comprehensive data that will allow:

- comparative sequence-based analyses of gene content;
- exploration of factors affecting and mechanisms underlying gene family expansion and contraction;
- consolidation of the phylogenetic tree of fungi, animals and opisthokonts as a whole, and placement of species of controversial affiliation;
- subsequent mapping of genomic features to the species tree; and
- correlation with body plan, biochemical capacity, ecological niche(s) and other biological features.

Choice of Organisms

The choice of organisms is critical for this large-scale genomics project. It will be important to choose species that are minimally derived (i.e., it will be of little value to characterize species having genomes with highly reduced gene contents resulting in major loss of function). Characterization of diverse, independent unicellular lineages closely related to animals and fungi is further crucial to enable robust inferences to be made about the genetic innovations that appeared when animals and fungi separated.

Logical candidates for better sampling at the base of the metazoan branch are *Capsaspora owczarzaki*; a colonial choanoflagellate such as *Salpingoeca* sp.; and the ichthyosporeans *Sphaeroforma arctica and Amoebidium parasiticum*. A more balanced coverage of Fungi will be provided by the two chytrids, *Spizellomyces punctatus* and *Allomyces macrogynus*, plus the zygomycete *Mortierella verticillata*. To characterize the branch diverging prior to Fungi, we propose *Nuclearia simplex*. Finally, to investigate protists diverging immediately prior to the opisthokonts, we propose to explore members of the apusomonads (*Amastigomonas* sp. or a related species).

The species proposed here are well described in the literature (for a summary of features, see Table 1). We have live cultures, cDNA libraries and the EST data generated as part of the Genome Canada-supported Protist EST Program (PEP; http://megasun.bch.umontreal.ca/pepdb/pep_main.html), as well as complete or almost complete mitochondrial genome sequences for most species. In addition, we have conducted phylogenomic analyses for these organisms using all currently available data, in order to determine the evolutionary affiliations of these species. In a few instances, we have genome size estimates, as well as other relevant information that strengthens the case for genome projects for these particular organisms.

For the less well-characterized species we intend to assess the feasibility of large-scale genome sequencing by limited pilot sequencing. Such data not only provide estimates of genome size, intron density, intergenic regions and content of mobile and repetitive elements, but are also a rigorous indicator of potential contaminations with food bacteria or hidden endosymbionts. In addition, we propose to explore related taxa if required. In the case of *Nuclearia simplex*, e.g., two different isolates are recognized and available from public strain collections. Isolate #4 turned out to be heavily contaminated with various food bacteria, which makes purification of nuclear DNA a cumbersome task. Isolate #2, in contrast, yielded (a small amount of) clean DNA, sufficient for pilot sequencing, and also allowed us to isolate and completely sequence the mitochondrial genome of this isolate.

Priority and Completeness of Sequencing Projects

We have prioritized the proposed genome sequencing projects by considering both their scientific importance and the ease of providing clean nuclear DNA (see **Table 1**). The most logical approach would be to start with one full genome sequence each in the animal and fungal lineages, followed by those projects that require pilot sequencing for feasibility studies. We have assigned lowest priority to *Salpingoeca/Codosiga* sp., *Amastigomonas* sp. and the loricates. Despite their high intrinsic level of scientific interest, the logistics of DNA purification from these organisms will require more time and effort than for axenic taxa.

Although one might argue that high-coverage genome sequencing is desirable in all instances, we recognize that the finishing of sequencing projects is particularly costly. We have therefore identified three out of the nine species for which high coverage draft sequences might be sufficient. Note, however, that the genomes of all proposed species are expected to be small (for nine species, an estimate of about 200-300 Mb in total) in comparison to animals and plants.

Provision of pure nuclear DNAs

The purification of large amounts of high-molecular-weight nuclear DNA from axenic eukaryotes (**Table 1**, projects 1,2,3,5,6,7) is straightforward. This is not the case for bacterivorous protists (projects 4,8,9,10) that have to be grown on live bacteria. In these cases, it is highly advantageous to replace the mix of food bacteria with a single bacterial species (monoxenic culture), whose genomic nucleotide composition differs significantly from that of the protist. Typically, several food bacteria are tested according to this criterion. Since protist nuclear DNAs are heavily contaminated with bacterial DNA even when isolated from stationary phase cultures, nuclear DNA is further purified by CsCl gradient centrifugation.

The following applicants, who have experience with the particular species and the outlined procedures, will ensure the provision of pure nuclear DNAs:

- ***Capsaspora* and *Sphaeroforma*:** I. Ruiz-Trillo and A. Roger
- ***Allomyces, Spizellomyces* and *Mortierella*:** B.F. Lang
- ***Amoebidium*:** G. Burger
- **Bacterivorous protists (*Nuclearia*, various choanoflagellates and *Amastigomonas*):** B. F. Lang, in collaboration with E. Steenkamp and N. King

Detailed Descriptions of Organisms

Capsaspora owczarzaki. Recent molecular phylogenetic analyses have shown that *Capsaspora owczarzaki,* a filose amoeboid symbiont of the pulmonate snail *Biomphalaria glabrata* (and initially believed to be a nucleariid), is one of the unicellular lineages branching close to choanoflagellates ((5); BFL, unpublished data using the complete mtDNA sequence). Besides its key phylogenetic position, *Capsaspora owczarzaki* has relevance to human health because its host, *B. glabrata,* is also the intermediate host of the digenean flatworm *Schistosoma mansoni*, the cause of widespread schistosomiasis in humans. This chronic disease affects millions of people: according to the World Health Organization, more than 600 million people are currently at risk for infection with a schistosoma species, especially in developing countries (WHO Expert Committee, 1993). Furthermore, military and diplomatic personnel, as well as travelers to developed countries (some 5 million individuals; CDC 1984, 1990, 1993) are at risk each year.

Interestingly, *Capsaspora owczarzaki* not only parasitizes the intermediate host of *S. mansoni* but has also been found to attack and kill the sporocysts of this worm living inside the snail (37). Thus, genome sequences of all the organisms pertinent to the transmission of schistosomiasis would be valuable. So far, the genome sequences of one host, humans, and the parasite, *S. mansoni*, have been completed, and the *Biomphalaria* genome is currently being sequenced (http://www.genome.gov/11007951). Along with the foregoing sequences, that of *Capsaspor*a should offer significant insights into genes implicated in the host-parasite relationship, and also aid in identifying, at the molecular level, the possible role of *Capsaspora* in influencing susceptibility of the snail to flatworm infection.

In addition, two important technical issues make *Capsaspora owczarzaki* an ideal candidate for this project: ease of culture and modest genome size. First, *Capsaspora* grows axenically, providing an ideal source of high quality, bacteria-free DNA. Furthermore, we have determined the size of the *Capsaspora owczarzaki* genome to be ~ 20-24 Mb (12 chromosomes ranging from 0.60 Mb to 4 Mb; **Fig. 2**, IR-T and AJR, unpublished). This value is less than half the size of the 50-Mb nuclear genome of *Trichoplax*, the metazoan with the simplest body plan. The G+C content for the coding regions is ~50%, based on the sequence data obtained from the 8,100 EST sequences generated through PEP. It should be noted that the *Capsaspora* mtDNA, which we have sequenced completely, exhibits conspicuous similarities with that of its closest relative, the choanoflagellate *Monosiga brevicollis*, with respect to genome size, genome organization, and gene and intron content.

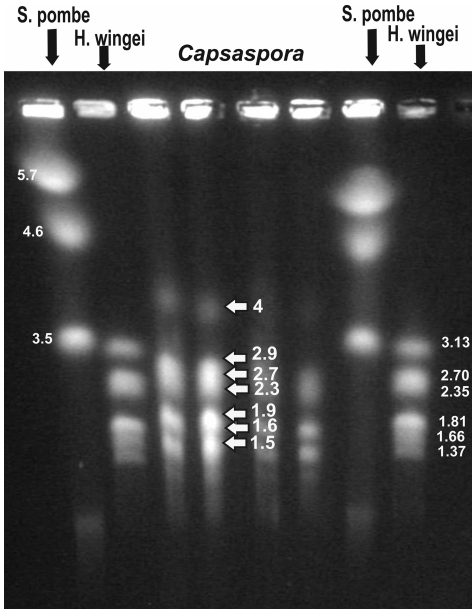

Figure 2. PFGE of *C. owczarzaki* chromosomes. Conditions have been optimized to resolve the largest chromosomes (in the range of 3-5 Mb; separations under other conditions not shown). The markers used were chromosomes from *S. pombe* and *H. wingei.* Marker size bands and the putative sizes of *C. owczarzaki* chromosomes are both shown.

Amoebidium parasiticum* and *Sphaeroforma arctica. Both of these unicellular organisms are recognized members of Ichthyosporea (38). Evolutionarily, they are the earliest diverging members of Holozoa (**Fig. 1**), which justifies sequencing the genomes of both species (one full and one draft). Like *Capsaspora*, ichthyosporeans offer the advantage of axenic culture and provide an ideal source of high quality DNA free of bacterial contamination. As we have only a rough estimate of their genome sizes, pilot sequencing should explore the feasibility of large-scale genome sequencing with respect to size, density of repeat elements, etc. EST sequences are available for both species through PEP, and a large part of the *Amoebidium* mitochondrial genome sequence has been published, revealing a unique multi-chromosome organization (39). This information allows us to position this species as the deepest evolutionary divergence at the base of the metazoan clade (4) – prior to choanoflagellates, and also prior to *C. owczarzaki* (BFL, unpublished).

Sphaeroforma arctica may provide special insights into multicellularity origin(s), because despite being unicellular, it has unique 'multicellular-like' features also found in colonial choanoflagellates. The species forms spherical structures composed of several cells, which at maturity release smaller cells. We believe some cell-to-cell communication may be needed to co-ordinate growth and maturation of these spherical bodies. Again, the current proposal could be crucial to determining whether *Sphaeroforma* might have different signaling genes than other unicellular opisthokonts, allowing a comparison to those present in metazoans. Interestingly, similar changes of morphology and cellular behavior are known for *Amoebidium parasiticum* (40,41).

***Nuclearia simplex*.** Nucleariids are naked amoebae that are characterized by radiating filopodia, which they use to gather food bacteria or algae. According to the most recent published phylogenies, nucleariids branch prior to Fungi (6). However, the statistical support for this position remains unsupported by likelihood ratio tests and a divergence prior to opisthokonts or prior to Ichthyosporea cannot be excluded. Several thousand ESTs each from two *Nuclearia simplex* strains have been sequenced (BFL; collaboration E. Steenkamp), and with a database comprising ~100 deduced protein sequences selected from this dataset, significant support is obtained for the divergence of *Nuclearia* prior to Fungi (unpublished results), confirming the potentially outstanding value of this group for comparisons with Fungi, and opisthokonts as a whole.

To date, no axenic *Nuclearia* strains have been described; in fact, culturing nucleariids such as *N. simplex* is cumbersome under standard conditions (on Petri dishes with solid agar and live *E. coli*), which has motivated attempts to optimize growth conditions. Replacement of *E. coli* with alternative food bacteria (isolated from protist cultures) has not only permitted us to establish for the first time vigorously growing liquid cultures for two *N. simplex strains*, it also facilitated the isolation of mitochondrial and nuclear DNA that is sufficiently pure for random genome sequencing.

Chytridiomycete and zygomycete fungi. The two latter groups have long been considered 'odd balls' within Fungi. Chytrids in particular are unusual, as they are the only fungi to produce flagellated spores (zoospores) during their life cycle. In fact, chytrids have only recently been recognized as a fourth fungal phylum, based on ultrastructural features (14,15) and molecular phylogenetic analyses (e.g., (28,42,43)). This affiliation suggests that the fungal ancestor had a flagellar apparatus, like many of its closest opisthokont relatives, and that flagella were lost for good in the traditional fungal phyla (ascomycetes, basidiomycetes and zygomycetes). Thus, chytrids could become unique fungal model systems for the analysis of genes encoding the flagellar apparatus.

Our knowledge of the biochemical and genetic properties of chytrids as well as zygomycetes is marginal, compared to what is known about ascomycetes and basidiomycetes. In this sense, chytrids and zygomycetes are 'forgotten treasures' that await exploration by genomics, an undertaking that could change our view of Fungi altogether.

Zygomycota are interesting from an additional perspective. They are a central group in the sense that they branch at a position intermediate between chytrids and ascomycetes/basidiomycetes, thus occupying an important position in comparative analyses. Moreover, they include numerous groups of pathogens. For instance, *Rhizopus* species cause mucormycosis, a type of disease that is characterized by vascular invasion with hyphae and necrosis of tissue. In its acute form, it is followed by death within a few days.

Traditional taxonomy has subdivided Zygomycota into Zygomycetes and Trichomycetes; however, these taxonomic groupings do not hold up in molecular phylogenies where even Zygomycetes appears to be paraphyletic (e.g., (44)). Given the low resolution of zygomycete phylogenies with single genes, phylogenomics is indispensable for a robust re-classification.

The fungal strains proposed for genome sequencing are all axenic and easily cultivated. Their genome sizes are expected to be in the range of 20 to at most 50 Mb, as in a variety of other fungi, including the chytrid *Batrachochytrium* (see http://www.genome.gov/10002154 and hyperlinks from this page) and the zygomycete *Rhizopus oryzae* (http://www.broad.mit.edu/annotation/fungi/rhizopus_oryzae/).

Exploration projects: apusomonads and choanoflagellates. These projects have been labeled 'exploration projects' because there is currently no information available to assure us that the production of clean DNA will be possible – a limitation shared with most other non-axenic protists. Nonetheless, the phylogenetic positions of these groups (diverging prior to opisthokonts or Metazoa, respectively), as well as their organismal characteristics (flagellar apparatus; colonial growth possibly representing a transition towards a multicellular organization), are strong incentives for genome projects. Identification of the most amenable species will follow procedures previously developed with success for *Nuclearia*. From a set of alternative species, the isolation of pure RNA and DNA fractions will be explored and, if necessary, culture conditions will be optimized (liquid cultures; establishment of monoxenic strains: i.e., strains able to grow on a single food bacterium). As in other cases, we propose that the feasibility of genome projects be determined by the analysis of a limited number of random genome sequences.

Broader Scientific Interest

The proposed comparative genomics projects will provide significant insights into hotly debated and controversial topics on multicellularity, as well as animal, fungal and opisthokont origins. A broad range of researchers will undoubtedly benefit from the genomic data, which will nurture a wide range of studies, as exemplified below.

Comparative genomics and bioinformatics: Species have been chosen for their pivotal phylogenetic positions: *Capsaspora*, the two ichthyosporeans and a colonial choanoflagellate as lineages closely related to Metazoa; *Nuclearia simplex* as a sister group to Fungi; chytrids and zygomycetes as sister groups to the ascomycetes and basidiomycetes; and an apusomonad species because it appears to diverge prior to opisthokonts. These phylogenetic characteristics make these taxa key for comparison of genomic features, and the investigation of traits related to the evolution of a multicellular lifestyle and developmental programs. This comparative genome proposal will further be of key importance for the identification of animal-specific and fungal-specific gene families, and elucidation of their origins in the respective unicellular ancestors.

Molecular evolution: The genomes of these organisms will also provide insights into long-standing evolutionary questions, such as the expansion and diversification of gene families and the evolution of regulatory elements. With the data generated we will be able to show which of the animal- and fungal-specific gene classes are more ancient and were already present in the common ancestors, and which ones appeared at later transition points from protists to Metazoa and Fungi.

Functional genomics and 'evo-devo': Once the genome sequences become available, an experimental functional genomics platform could be launched in order to analyze the ancestral function of animal-specific and fungal-specific gene families. Such a development would provide insights into the evolution and acquisition of gene function, especially relevant in the case of gene families involved in development, cell adhesion, and cell-to-cell communication. *Capsaspora owczarzaki, Sphaeroforma arctica* and *Amoebidium parasiticum*, as well as the three selected fungal species, all have the advantage of growing readily and axenically, in contrast to related species such as choanoflagellates. Thus, post-sequencing functional analyses, such as genome-wide microarray expression studies and proteomics, will be much more tractable in comparison with choanoflagellates. In addition, traditional biochemical studies that have so far been restricted to mitochondrial components (e.g., (11,12)) may be more easily extended to include the analysis of nucleus-encoded components.

Metazoan systematics: Multi-gene and phylogenomic approaches are gaining momentum, and are the preferred methods for obtaining more robust topologies. Thus, due to their pivotal phylogenetic positions, sequences from early diverging opisthokonts will not only be key to determining which genes are animal-specific, but also to providing a critical resource (as a close outgroup to animals) for improving phylum-level phylogenomic studies of metazoans.

Fungal systematics, cell biology and inter-species relationships: Fungal systematics has been largely dominated by investigations of ascomycetes and basidiomycetes, and within these subgroups, is again dominated by yeast projects. With these initiatives rapidly coming to fruition, interest in extending research to other (even previously unknown) fungal groups is growing. This expansion is even more important considering that fungal genomics has been a motivator for understanding a multitude of eukaryotic (and specifically human) functions: an understanding limited, however, by the highly reduced gene complement of the widely used budding yeast and fission yeast models. For instance, model investigations on the eukaryotic flagellar apparatus might benefit considerably from establishing a chytridiomycete model system.

Eukaryotic microbiology and protistology: Eukaryotic microbes (protists) represent an extremely diverse group of organisms in terms of morphology, body plan, ecology, biochemistry and molecular processes. To date, the genomic information gathered from protists derives mainly from human parasites (e.g. *Plasmodium*, *Trypanosoma,* etc.). Obtaining the full genome sequences of additional unicellular lineages will contribute to our understanding of the diversity, systematics and evolution of eukaryotic microbes in general.

Parasitology:The *Capsaspor*a genome sequence could offer significant insights into the human disease schistosomiasis, with potential implications for the development of therapies for the more than 200 million people affected annually. In addition, a large number of fungi are known to be opportunistic pathogens that are difficult to treat, given the variety of fungal life styles. New pathogens are constantly being added to the long list of mycoses (caused, e.g., by certain *Pneumocystis*, *Rhizopus* and *Aspergillus* species) that are particularly threatening to immuno-compromised patients. A balanced comparative genome analysis of selected members of the fungal kingdom would be a key first step to understanding the diversity of their genetic and functional properties.

References

1. Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C. *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science*, **309,** 1559-1563.
2. Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420,** 520-562.
3. Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A. *et al.* (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*, **297,** 1301-1310.
4. Lang, B.F., O'Kelly, C., Nerad, T., Gray, M.W. and Burger, G. (2002) The closest unicellular relatives of animals. *Curr Biol*, **12,** 1773-1778.
5. Ruiz-Trillo, I., Inagaki, Y., Davis, L.A., Sperstad, S., Landfald, B. and Roger, A.J. (2004) *Capsaspora owczarzaki* is an independent opisthokont lineage. *Curr Biol*, **14,** R946-947.
6. Steenkamp, E.T., Wright, J. and Baldauf, S.L. (2006) The protistan origins of animals and fungi. *Mol Biol Evol*, **23,** 93-106.
7. Cavalier-Smith, T. and Chao, E.E. (2003) Phylogeny of Choanozoa, Apusozoa, and other protozoa and early eukaryote megaevolution. *J Mol Evol*, **56,** 540-563.
8. Ohja, M., Turler, H. and Turian, G. (1977) Characterization of the *Allomyces* genome. *Biochem Biophys Acta*, **478,** 392-406.
9. Olson, L.W. (1984) *Allomyces* - a different fungus. *Opera Botanica*, **73,** 1-96.
10. Laforest, M.J., Roewer, I. and Lang, B.F. (1997) Mitochondrial tRNAs in the lower fungus *Spizellomyces punctatus*: tRNA editing and UAG 'stop' codons recognized as leucine. *Nucleic Acids Res*, **25,** 626-632.
11. Bullerwell, C.E. and Gray, M.W. (2005) *In vitro* characterization of a tRNA editing activity in the mitochondria of *Spizellomyces punctatus*, a chytridiomycete fungus. *J Biol Chem*, **280,** 2463-2470.
12. Laforest, M.J., Bullerwell, C.E., Forget, L. and Lang, B.F. (2004) Origin, evolution, and mechanism of 5' tRNA editing in chytridiomycete fungi. *RNA*, **10,** 1191-1199.
13. Dykova, I. and Lom, J. (2004) Advances in the knowledge of amphizoic amoebae infecting fish. *Folia Parasitol (Praha)*, **51,** 81-97.
14. Barr, D.S. (2001) Chytridiomycota. In McLaughlin, D. J., McLaughlin, E. G. and Lemke, P. A. (eds.), *The Mycota VIIA*. Springer Verlag, Berlin, Heidelberg, pp. 93-112.
15. Barr, D.J.S. (1990) Phylum Chytridiomycota. In Margulis, L., Corliss, J. O., Melkonian, M. and Chapman, D. J. (eds.), *Handbook of Protoctista*. Jones & Bartlett, Boston, pp. 454-466.
16. Clapham, M.E., Narbonne, G.M. and Gehling, J.G. (2003) Paleoecology of the oldest known animal communities: Ediacaran assemblages at Mistaken Point, Newfoundland. *Paleobiol*, **29,** 527-544.
17. Conway-Morris, S. (2005) Ediacarans. *Curr Biol*, **15,** R8.
18. Valentine, J.W. and Jablonski, D. (2003) Morphological and developmental macroevolution: a paleontological perspective. *Int J Dev Biol*, **47,** 517-522.
19. Finnerty, J.R. and Martindale, M.Q. (1999) Ancient origins of axial patterning genes: Hox genes and ParaHox genes in the Cnidaria. *Evol Dev*, **1,** 16-23.
20. Miller, D.J., Ball, E.E. and Technau, U. (2005) Cnidarians and ancestral genetic complexity in the animal kingdom. *Trends Genet*, **21,** 536-539.
21. Kusserow, A., Pang, K., Sturm, C., Hrouda, M., Lentfer, J., Schmidt, H.A., Technau, U., von Haeseler, A., Hobmayer, B., Martindale, M.Q. *et al.* (2005) Unexpected complexity of the Wnt gene family in a sea anemone. *Nature*, **433,** 156-160.
22. Technau, U., Rudd, S., Maxwell, P., Gordon, P.M., Saina, M., Grasso, L.C., Hayward, D.C., Sensen, C.W., Saint, R., Holstein, T.W. *et al.* (2005) Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. *Trends Genet*, **21,** 633-639.
23. King, N. (2004) The unicellular ancestry of animal development. *Dev Cell*, **7,** 313-325.
24. Snell, E.A., Brooke, N.M., Taylor, W.R., Casane, D., Philippe, H. and Holland, P.W.H. (2006) An unusual choanoflagellate protein released by Hedgehog autocatalytic processing. *Proc R Soc B*, **in press**.

25. Alexopolous CJ, M.C., and Blackwell M. (1996) *Introductory Mycology*. John Wiley and Sons, New York.
26. Margulis, L., Olendzenski, L., Dolan, M. and MacIntyre, F. (1996) Diversity of eukaryotic microorganisms: computer-based resources, "The Handbook of Protoctista" and its "Glossary". *Microbiologia*, **12,** 29-42.
27. Margulis, L. (1992) Biodiversity: molecular biological domains, symbiosis and kingdom origins. *Biosystems*, **27,** 39-51.
28. James, T.Y., Porter, D., Leander, C.A., Vilgalys, R. and Longcore, J.E. (2000) Molecular phylogenetics of the Chytridiomycota supports the utility of ultrastructural data in chytrid systematics. *Can J Bot*, **78,** 226-350.
29. Keeling, P.J., Luker, M.A. and Palmer, J.D. (2000) Evidence from beta-tubulin phylogeny that microsporidia evolved from within the fungi. *Mol Biol Evol*, **17,** 23-31.
30. Edman, J.C., Kovacs, J.A., Masur, H., Santi, D.V., Elwood, H.J. and Sogin, M.L. (1988) Ribosomal RNA sequence shows *Pneumocystis carinii* to be a member of the fungi. *Nature*, **334,** 519-522.
31. Ustinova, I., Krienitz, L. and Huss, V.A. (2000) *Hyaloraphidium curvatum* is not a green alga, but a lower fungus; *Amoebidium parasiticum* is not a fungus, but a member of the DRIPs. *Protist*, **151,** 253-262.
32. Forget, L., Ustinova, J., Wang, Z., Huss, V.A. and Lang, B.F. (2002) *Hyaloraphidium curvatum*: a linear mitochondrial genome, tRNA editing, and an evolutionary link to lower fungi. *Mol Biol Evol*, **19,** 310-319.
33. Richards, T.A. and Cavalier-Smith, T. (2005) Myosin domain evolution and the primary divergence of eukaryotes. *Nature*, **436,** 1113-1118.
34. Rubin, G.M., Yandell, M.D., Wortman, J.R., Gabor Miklos, G.L., Nelson, C.R., Hariharan, I.K., Fortini, M.E., Li, P.W., Apweiler, R., Fleischmann, W. *et al.* (2000) Comparative genomics of the eukaryotes. *Science*, **287,** 2204-2215.
35. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. and Lander, E.S. (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature*, **423,** 241-254.
36. Gaillardin, C., Duchateau-Nguyen, G., Tekaia, F., Llorente, B., Casaregola, S., Toffano-Nioche, C., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M. *et al.* (2000) Genomic exploration of the hemiascomycetous yeasts: 21. Comparative functional classification of genes. *FEBS Lett*, **487,** 134-149.
37. Owczarzak, A., Stibbs, H.H. and Bayne, C.J. (1980) The destruction of *Schistosoma mansoni* mother sporocysts in vitro by amoebae isolated from *Biomphalaria glabrata*: an ultrastructural study. *J Invertebr Pathol*, **35,** 26-33.
38. Mendoza, L., Taylor, J.W. and Ajello, L. (2002) The class Mesomycetozoea: a heterogeneous group of microorganisms at the animal-fungal boundary. *Annu Rev Microbiol*, **56,** 315-344.
39. Burger, G., Forget, L., Zhu, Y., Gray, M.W. and Lang, B.F. (2003) Unique mitochondrial genome architecture in unicellular relatives of animals. *Proc Natl Acad Sci U S A*, **100,** 892-897.
40. Whisler, H.C. (1968) Developmental control of *Amoebidium parasiticum*. *Dev Biol*, **17,** 562-570.
41. Whisler, H.C. and Fuller, M.S. (1968) Preliminary observations on the holdfast of *Amoebidium parasiticum*. *Mycologia*, **60,** 1068-1079.
42. Paquin, B., Forget, L., Roewer, I. and Lang, B.F. (1995) Molecular phylogeny of *Allomyces macrogynus*: congruency between nuclear ribosomal RNA- and mitochondrial protein-based trees. *J Mol Evol*, **41,** 657-665.
43. Gunderson, J.H., Elwood, H., Ingold, A., Kindle, K. and Sogin, M.L. (1987) Phylogenetic relationships between chlorophytes, chrysophytes, and oomycetes. *Proc Natl Acad Sci U S A*, **84,** 5823-5827.
44. Seif, E., Leigh, J., Liu, Y., Roewer, I., Forget, L. and Lang, B.F. (2005) Comparative mitochondrial genomics in zygomycetes: bacteria-like RNase P RNAs, mobile elements and a close source of the group I intron invasion in angiosperms. *Nucleic Acids Res*, **33,** 734-744.

APPENDIX

In the initial evaluation of this proposal, one of the concerns raised by the Coordinating Committee was that some of the organisms proposed here might be too divergent for meaningful comparative analyses of genes involved in multicellularity. In fact, recent publications (many of them based on newly available genome sequences) have identified a growing number of developmental gene homologs across eukaryotes as different as humans, slime molds, fungi and plants. These new data are best explained by the existence of a common set of ancient developmental genes that are the basis for multicellularity, with the particular patterns of development in animals being achieved through the action of supplementary, animal-specific genes.

This Appendix is aimed at addressing issues of species selection for the investigation of the origin(s) of multicellularity. We review recent literature on this topic and show that available new data derived from certain unicellular organisms are indeed highly informative, constituting a unique means to identify both fungal- and animal-specific gene families responsible for a multi-celled body plan.

Relevance of Unicellular Organisms and Fungi for Investigating the Origins of Multicellularity

A paradox in the transition to multicellularity is that its underlying molecular mechanisms must have evolved in unicellular protists. In fact, numerous genes that are clearly related to multicellularity in animals are present in strictly unicellular choanoflagellates such as *Monosiga* (for details, see below), where they are likely involved in two processes: sex and predation, both of which require cell-cell recognition, adhesion and endocytosis or fusion (King *et al.*, 2003; King, 2004). Accordingly, a comprehensive identification of genes involved in multicellularity requires an understanding of the rudimentary steps underlying this organizational pattern in the unicellular ancestors of multicellular lineages. Our intention has been to choose a set of species that is not too divergent but that nevertheless represents the spectrum of extant descendants of animal and fungal ancestors.

The emerging picture of how animals and fungi originated – based on a plethora of recent molecular data – is that their ancestors already possessed a substantial part of the genetic toolkit needed to construct a multicellular body plan. Whereas most of these genes are ubiquitous in animals, they are notably absent in certain fungal lineages such as budding and fission yeasts, which have a dramatically reduced number of nuclear genes and a unicellular lifestyle for all of their members (for a comparison of yeast, flies and worms, see also Hazkani-Covo *et al.*, 2004). In yeasts, the loss of genome complexity and multicellular structures is secondary, not to be confused with the situation in primitively unicellular eukaryotes (e.g., *Monosiga*) that exhibit high genome complexities. In contrast, some of their ascomycete relatives such as *Neurospora crassa*, *Aspergillus nidulans* and *Podospora anserina* have more than twice the number of genes, including numerous (estimated at ~ 100; Poggeler & Kuck, 2004) developmental genes that are involved in building the multicellular structures (fruiting bodies) typical of these fungi (e.g., Borkovich *et al.*, 2004; Malagnac *et al.*, 2004; Poggeler & Kuck, 2004; Huang *et al.*, 2005; Moore & Meskauskas, 2006). The list of these fungal genes includes ones encoding a family of WD40 repeat proteins (striatin, zinedin and SG2NA-related proteins; Poggeler & Kuck, 2004), NADPH oxidase (Lalucque & Silar, 2003; Malagnac *et al.*, 2004) and a host of signal transduction pathways that are absent or reduced in number in yeast (e.g., histidine kinases with various response regulators, heterotrimeric G proteins, cAMP receptors, serine/threonine kinases, and calcium signaling pathways; for details see Borkovich *et al.*, 2004).

Evidently, this set of fungal developmental genes is much smaller than in animals (for missing signaling pathways see Borkovich, 2004). On the other hand, a comparative analysis of the *Dictyostelium* genome not only reveals a large number of animal-like developmental genes, but also identifies some genes found in vertebrates that were lost in invertebrates such as *C. elegans* and *Drosophila* (Kay, 2002; Williams *et al.*, 2005).

It follows from the examples given above that secondary loss of developmental genes is quite common in unicellular fungi, but also in highly complex multicellular animals. In order to identify a comprehensive set of genes involved in multicellular life styles, a genomic survey of organisms close to the animal/fungal origin is in order. In this context, there is obviously little appeal in surveying further extremely reduced genomes such as those of budding and fission yeasts. Rather, we propose to focus on protist and fungal (in particular chytridiomycete and zygomycete) lineages for which no genome sequence is currently published, including eukaryotic microbes that diverged prior to opisthokonts but later than slime molds (e.g., *Anastigomonas*; see **Fig. 1**).

Published Data on Descendents of the Unicellular Ancestor of Animals

Choanoflagellates are the closest unicellular relatives of animals. In recognition of this relationship, Choanoflagellata and Metazoa, together with Ichthyosporea and *Capsaspora*, are now grouped together in the Holozoa ((Lang *et al.*, 2002); Lang *et al.*, unpublished; see **Fig. 1**). Choanoflagellates represent the only clade for which data relevant to this proposal have been published. One available data set is the collection of 5000 ESTs from two choanoflagellate species, *Monosiga brevicollis* and a *Proterospongia*-like species, reported by (King *et al.*, 2003): a data set sufficient to demonstrate that choanoflagellates express a wide variety of protein families involved in animal cell-cell interactions, such as cadherins, C-type lectins and tyrosine kinases (TK). The TK family, for example, represents a well-known group of signal transduction proteins involved in cell proliferation and regulation in animals. Aimed at understanding the role such a gene family was playing in a unicellular organism, King *et al.* (2003) inhibited TK activity and were able to show that TKs are indeed required during choanoflagellate proliferation. Moreover, by analysis of the response to nutrient availability, it was also shown that choanoflagellates interpret extracellular signals through a TK signaling pathway. These data show for the first time that a signaling pathway is present in a unicellular organism and, more importantly, what role it plays in such an organism.

The *hedgehog* gene family provides another example of how data from unicellular members of Holozoa can be used to understand the origin of features typical for animals. Hedgehog (Hh) proteins are important cell-cell signaling entities involved in animal development. Interestingly, Hh proteins have also recently been implicated in human obesity, diabetes, and osteoporosis (Suh *et al.*, 2006). Through analysis of an EST library from the choanoflagellate *Monosiga ovata*, a homolog of the autocatalytic domain of the Hh protein has recently been found in this unicellular organism (Snell *et al.*, 2006). Moreover, it has been shown that this Hh protein (the only one found so far in a non-metazoan organism) has a unique modular structure. By comparative analysis of animal and choanoflagellate Hh proteins, it was deduced that the animal *hedgehog* genes most probably evolved by a fusion of two distinct genes (Snell *et al.*, 2006).

Unpublished Data on Unicellular Members of Holozoa

Multicellular-like features of the unicellular holozoan Capsaspora owczarzaki*.*

Under the auspices of the Genome Canada Protist EST Program (http://megasun.bch.umontreal.ca/pepdb/pep_main.html), we have constructed cDNA libraries from *Capsaspora owczarzaki*, from which we have so far sequenced 8870 ESTs, assembled into 2303 unique gene clusters. As expected from the phylogenetic position of *Capsaspora* (**Fig. 1**), most of these clusters have their highest BLAST scores to animal homologs. Among the genes found to date, several are involved in signaling pathways, such as protein kinases, a Bruton's tyrosine kinase and a membrane-associated guanylate kinase (MAGI) protein. Interestingly, MAGI proteins have so far been described only in animals (see (Adell et al., 2004)) and they are known to specifically regulate adhesion and plasticity in animal tissue barriers, known as tight-junctions (Funke et al., 2005). Since tight-junctions connect different tissues and regulate the passage of molecules among them, they are of direct relevance to human health. Phylogenetic analysis of the complete *Capsaspora* MAGI sequence shows that this protein is clearly a member of the MAGI family (see **Fig. A1**). Furthermore, the domain architecture is unique because although it displays the same overall structure as its animal homologs, it is missing the last five or (in the case of *Danio rerio*) two PDZ domains (see **Fig. A2**). How this unique modular architecture affects the biological role and the specific function of this protein in a single-celled, non-tissued organism such as *Capsaspora owczarzaki* is an intriguing question. Functional analysis of this gene is likely to provide insight into how this crucial animal protein evolved.

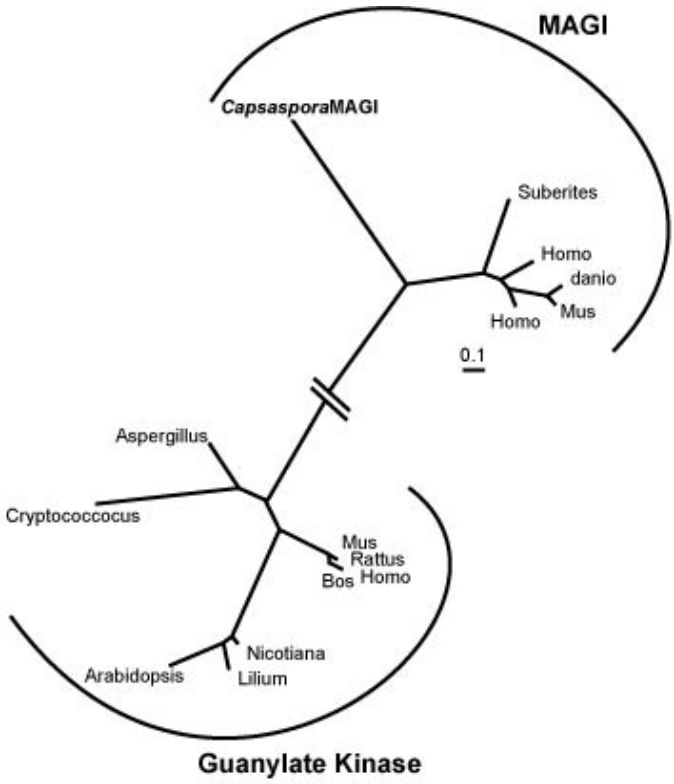

Figure A1: Phylogenetic tree of guanylate kinases and MAGI family proteins.

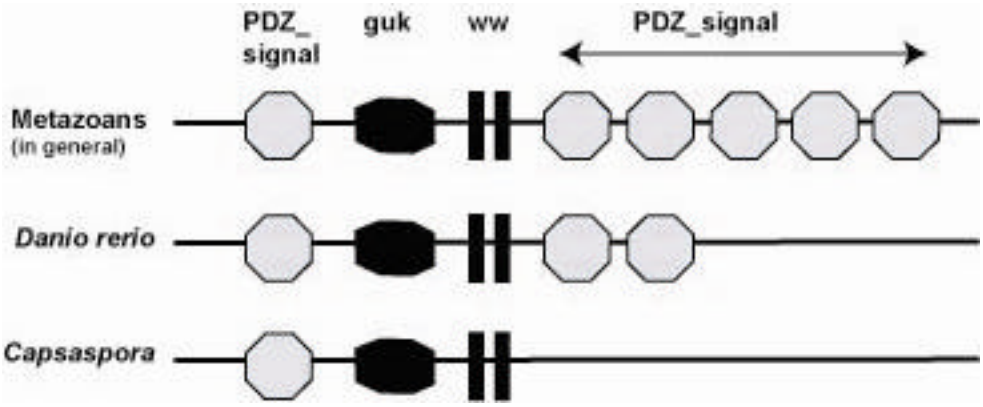

Figure A2: Modular construction of Hh protein homologs.

Multicellular-like features of the unicellular holozoan Sphaeroforma arctica.

An image of the ichthyosporean *Sphaeroforma arctica*,exhibiting "multi-celled" features provides a visual reinforcement of what we have explained in the proposal (**Fig. A3**). We expect that the genome of this protist, which according to currently available data branches prior to the opisthokonts (**Fig. 1**), encodes cell-cell communication genes. Comparison of such genes with those found in the other "completely" unicellular lineages and metazoans will likely provide additional insights into the origins of multicellularity.

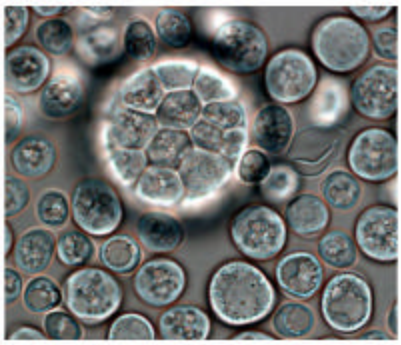

Figure A3: Multi-celled cluster of *Sphaeroforma arctica*.

Conclusions

The examples listed above (TK of *M. brevicollis*, Hh of *M. ovata* and MAGI of *Capsaspora owczarzaki*), all obtained from modest EST projects, provide us with a glimpse of the "*genetic starter kit*" of the putative ancestral metazoan. Notably, these examples substantiate the main argument of this proposal, which posits that clues about the transition to animal and fungal multicellularity are contained in the genomes of their single-celled relatives. Finally, it should be noted that most of these developmental ("multicellular") protein families are relevant to human health; knowledge of how these genes evolved could eventually have an impact on the development of efficient therapeutic strategies.

In closing, we would like to emphasize that we have listed in this proposal the only organisms at the base of animals and fungi that in our experience are actually tractable for genome sequencing. We are aware of no other organisms in this part of the tree that would be "less diverged". Our assessment is that the best hope we have of understanding the common ancestral animal and fungal genomes is through comparison of the genomes of the taxa we mention with those of "higher" animals and fungi. The species we have selected are an unbiased set of representatives of these taxa: there are no major "groups" in this part of the tree that we have ignored. The "long branches" of these species in phylogenies merely indicate that the genes from these organisms that we use for phylogenetic analysis are evolving somewhat faster than their homologs in other taxa in the tree. This phenomenon is completely separate from the issue of whether or not these organisms contain homologs of metazoan marker proteins. In other words, the primary sequences of the genes of these species may be "fast-evolving", but they may still contain metazoan marker gene homologs. We agree with the Coordinating Committee that it is possible that some of these species may not contain the metazoan marker proteins we will be seeking; however, this is something we cannot know in advance. We submit that if we are going to answer the questions we pose here, these are the only taxa we currently know whose genomes may help us. As with any comparative genomics project, this proposal carries with it an element of risk; however, we feel strongly that the questions are "big" enough to warrant accepting such a risk.

Literature Cited

Adell T, Gamulin V, Perovic-Ottstadt S, Wiens M, Korzhev M, Muller IM, Muller WE. 2004. Evolution of metazoan cell junction proteins: the scaffold protein MAGI and the transmembrane receptor tetraspanin in the demosponge *Suberites domuncula*. *J Mol Evol 59*:41-50.

Borkovich KA, Alex LA, Yarden O, Freitag M, Turner GE, Read ND, Seiler S, Bell-Pedersen D, Paietta J, Plesofsky N, Plamann M, Goodrich-Tanrikulu M, Schulte U, Mannhaupt G, Nargang FE, Radford A, Selitrennikoff C, Galagan JE, Dunlap JC, Loros JJ, Catcheside D, Inoue H, Aramayo R, Polymenis M, Selker EU, Sachs MS, Marzluf GA, Paulsen I, Davis R, Ebbole DJ, Zelter A, Kalkman ER, O'Rourke R, Bowring F, Yeadon J, Ishii C, Suzuki K, Sakai W, Pratt R. 2004. Lessons from the genome sequence of *Neurospora crassa*: tracing the path from genomic blueprint to multicellular organism. *Microbiol Mol Biol Rev 68*:1-108.

Funke L, Dakoji S, Bredt DS. 2005. Membrane-associated guanylate kinases regulate adhesion and plasticity at cell junctions. *Annu Rev Biochem 74*:219-245.

Hazkani-Covo E, Levanon EY, Rotman G, Graur D, Novik A. 2004. Evolution of multicellularity in Metazoa: comparative analysis of the subcellular localization of proteins in *Saccharomyces*, *Drosophila* and *Caenorhabditis*. *Cell Biol Int 28*:171-178.

Huang S, Yuan S, Dong M, Su J, Yu C, Shen Y, Xie X, Yu Y, Yu X, Chen S, Zhang S, Pontarotti P, Xu A. 2005. The phylogenetic analysis of tetraspanins projects the evolution of cell-cell interactions from unicellular to multicellular organisms. *Genomics 86*:674-684.

Kay RR. 2002. Chemotaxis and cell differentiation in *Dictyostelium*. *Curr Opin Microbiol 5*:575-579.

King N. 2004. The unicellular ancestry of animal development. *Dev Cell 7*:313-325.

King N, Hittinger CT, Carroll SB. 2003. Evolution of key cell signaling and adhesion protein families predates animal origins. *Science 301*:361-363.

Lalucque H, Silar P. 2003. NADPH oxidase: an enzyme for multicellularity? *Trends Microbiol 11*:9-12.

Lang BF, O'Kelly C, Nerad T, Gray MW, Burger G. 2002. The closest unicellular relatives of animals. *Curr Biol 12*:1773-1778.

Malagnac F, Lalucque H, Lepere G, Silar P. 2004. Two NADPH oxidase isoforms are required for sexual reproduction and ascospore germination in the filamentous fungus *Podospora anserina*. *Fungal Genet Biol 41*:982-997.

Moore D, Meskauskas A. 2006. A comprehensive comparative analysis of the occurrence of developmental sequences in fungal, plant and animal genomes. *Mycol Res 110*:251-256.

Poggeler S, Kuck U. 2004. A WD40 repeat protein regulates fungal cell differentiation and can be replaced functionally by the mammalian homologue striatin. *Eukaryot Cell 3*:232-240.

Snell EA, Brooke NM, Taylor WR, Casane D, Philippe H, Holland PWH. 2006. An unusual choanoflagellate protein released by Hedgehog autocatalytic processing. *Proc R Soc B in press*.

Suh JM, Gao X, McKay J, McKay R, Salo Z, Graff JM. 2006. Hedgehog signaling plays a conserved role in inhibiting fat formation. *Cell Metab 3*:25-34.

Williams JG, Noegel AA, Eichinger L. 2005. Manifestations of multicellularity: *Dictyostelium* reports in. *Trends Genet 21*:392-398.