3

Fungal Genomics

Dr. Ariyah Terasawat

3.1 Introduction

3.1.1 The Fungal Kingdom

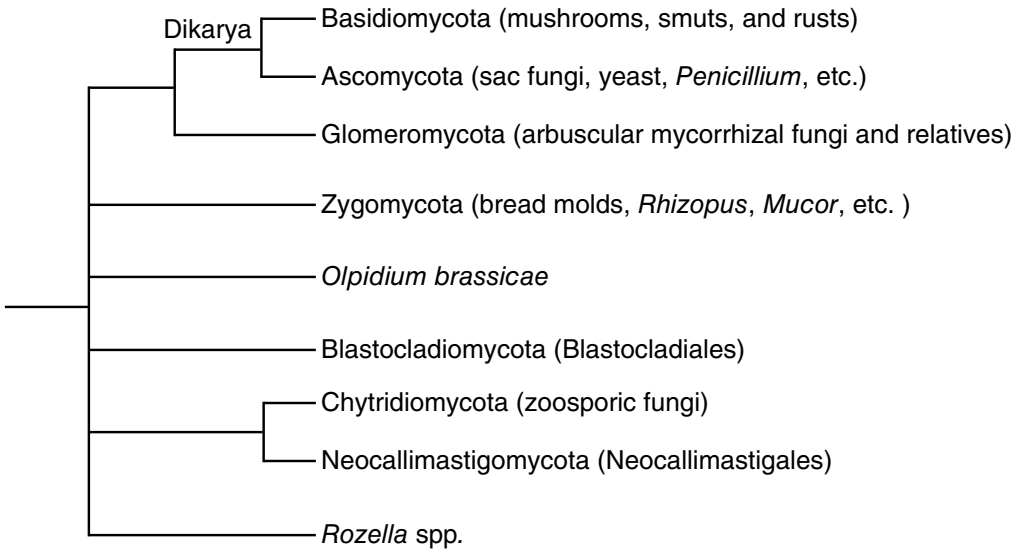


Figure 3.1 Schematic of the current consensus on fungal phylogeny. Fungal phylogenetics is currently in flux and the relationships between particular phyla are still unknown. (Phylogeny redrawn from the Tree of Life web project (Table 1).)

Fungi are eukaryotic organisms (contain a nucleus and membrane-bound organelles) and form one of the kingdoms of life. They lack chlorophyll and are saprobic (live on dead organic matter). Traditionally, fungi were thought to be closely related to plants; however, recent phylogenetic studies have shown that fungi are more closely related to animals than plants. The exact number of fungal species is unknown, but it is estimated to be 1.5 million.

Initial phylogenetic analyses of fungal species had revealed that there were four distinct phyla within the fungal kingdom: the **chytridiomycota, Zygomycota, Ascomycota, and Basidiomycot**a. early diverging species were found in the chytridiomycota and Zygomycota phyla. more recent phylogenetic analyses have suggested that neither the chytridiomycota nor the Zygomycota phyla are monophyletic. monophyletic species are descended from a common evolutionary ancestor and are not shared with any other groups. Recent studies now suggest that there are actually six fungal phyla and four additional unplaced subphyla (Figure 3.1).

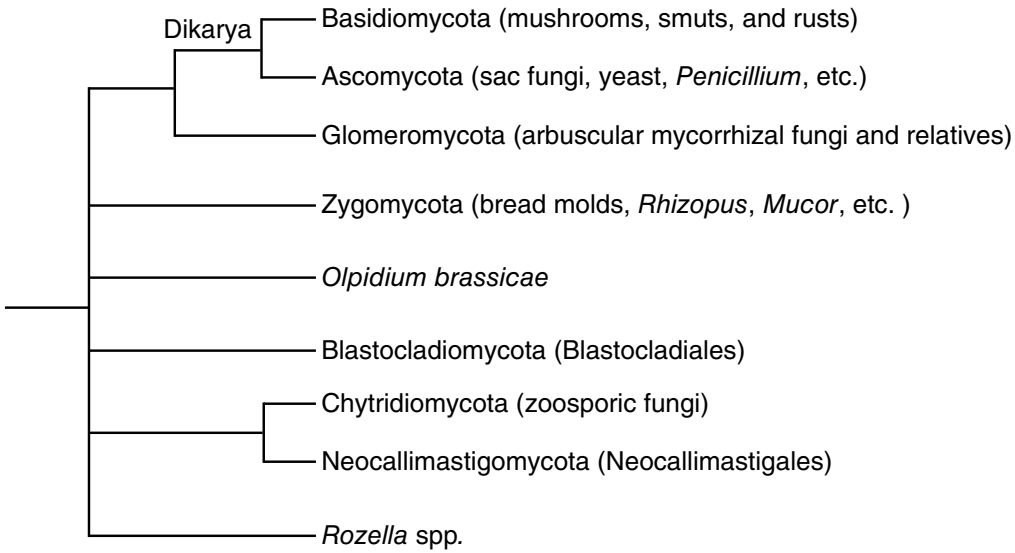
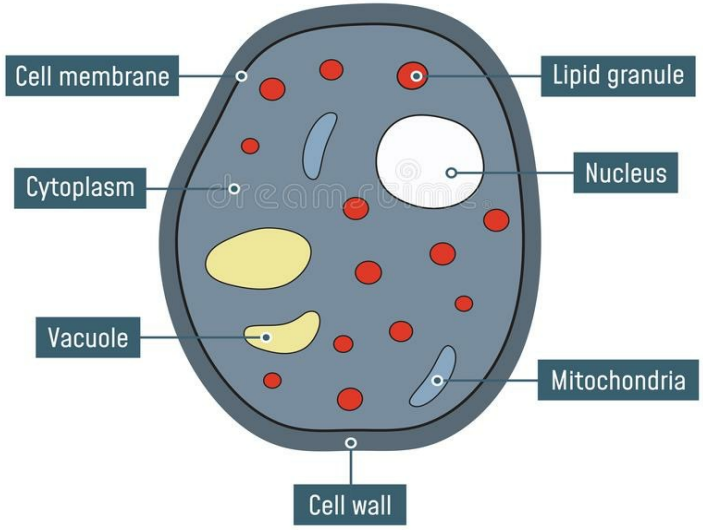


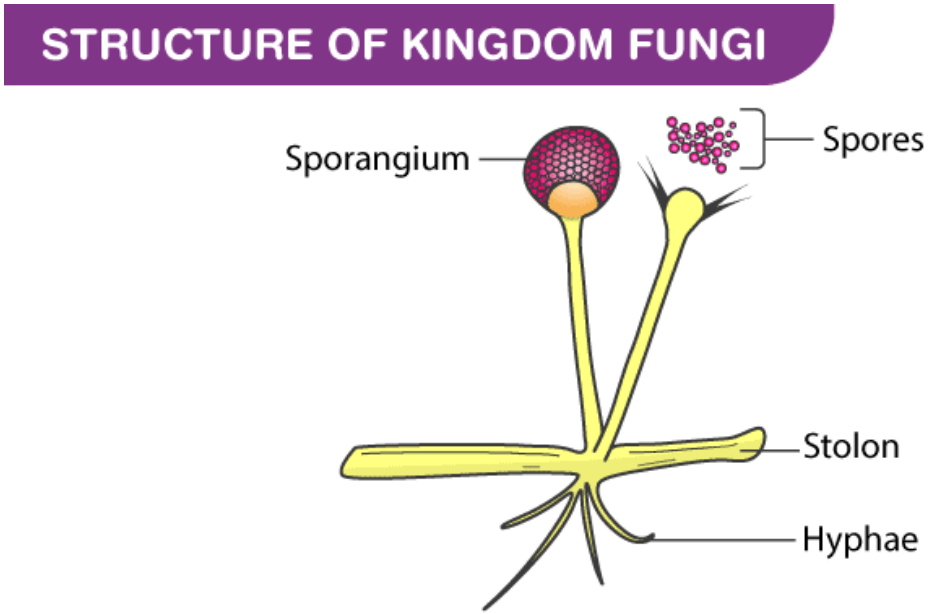
Figure 3.1 Schematic of the current consensus on fungal phylogeny. Fungal phylogenetics is currently in flux and the relationships between particular phyla are still unknown. (Phylogeny redrawn from the Tree of Life web project (Table 1).)

Fungal phylogenetics is far from static or resolved; in fact the placement of specific species within the fungal kingdom has been questioned. For example, gut inhabitants of arthropods (*Trichomycetes* species) that were thought to be members of the Zygomycota are actually protists. Furthermore, other species that were considered fungal as they display heterotrophic, mold-like growth morphology are in fact Stramenopiles (including algae, kelps, and diatoms). The genome sequences of sparsely sampled phyla should help resolve this ambiguity in future years. For clarity, only the four traditional fungal phyla are discussed here.

STRUCTURE OF FUNGI

- Eukaryotic
- Non motile
- Some fungi are single celled, while others are multicellular. Single celled fungi are also called yeasts.
- Fungi cells have a very small nuclei with little repetitive DNA and organelles like plant and animals do.
- The cell walls of fungi contain chitin, which is a hard substance found in the exoskeletons of insects and arthropods





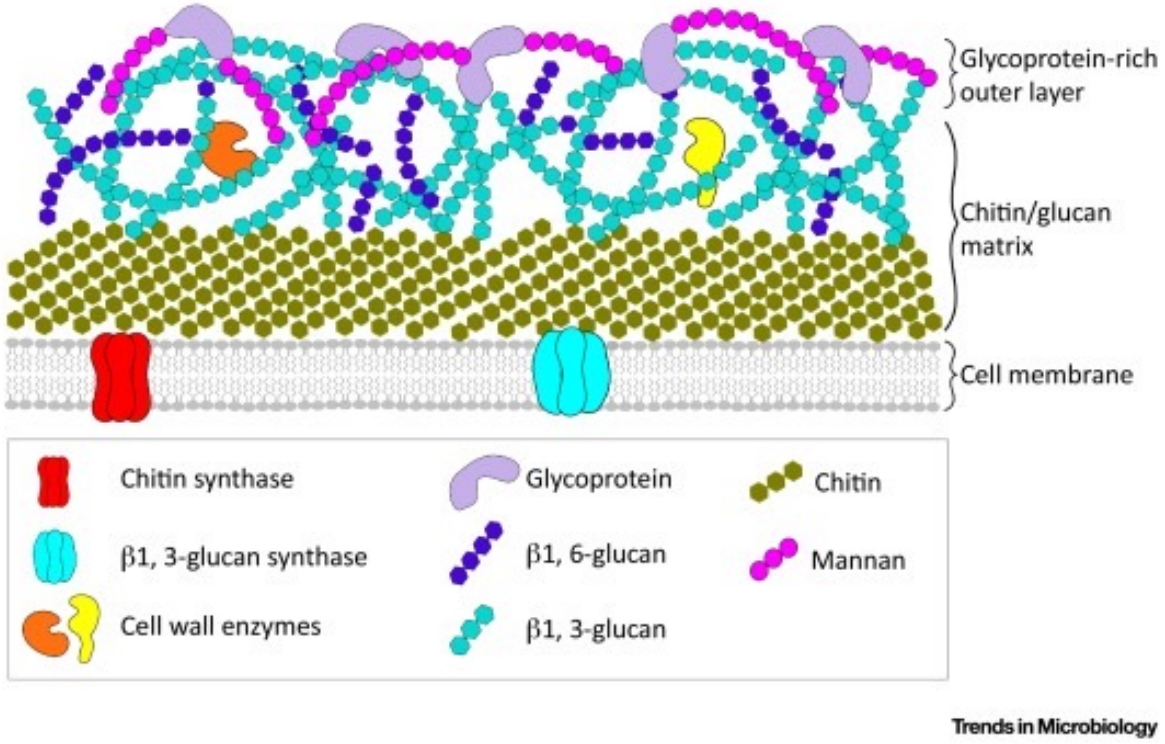
- Multicellular fungi have many branching filaments called hyphae. The are tubular shaped and are split into cell like compartments that are known as septa. The network of hyphae in fungus is called mycelium.

- They are heterotrophs, and can't make their own food so they must obtain nutrients from organic material.
- They can be :
    * Saprobes (absorb nutrients from dead material)
    * Parasites
    * Mutualistic symbionts (association of fungi with algae is termed as lichens, association of fungus and plant root is called mycorrhiza)
- All fungi reproduce through spores. Spores are microscopic cells that disperse from their parental fungus, usually through wind or water. Fungi can produce spores through sexual and asexual reproduction.
- Both asexual and sexual method of reproduction can be opted, asexual reproduction occurs through the release of spores or through mycelial fragmentation, where mycelium separates into multiple pieces that grow separately and in sexual reproduction separate individuals fuse their hyphae together.

CELL WALL

❖ CHARACTARISTICS OF CELL WALL

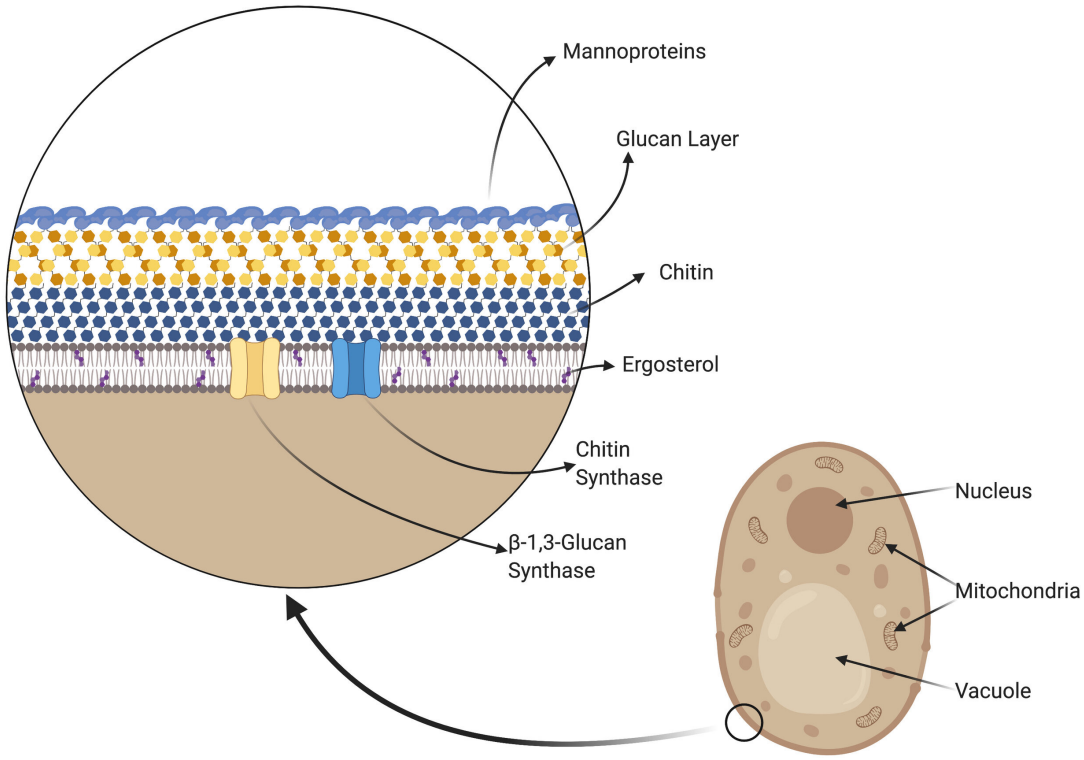
- Tough, flexible and sometimes rigid.
- Gives shape to fungi.
- Gives strength to fungi.
- Provides protection for the protoplasm from ultraviolet rays(presence of melanins).
- Ability to resist lysis by organic solvents such as enzymes, toxins, osmotic integrity.
- Ability to bind to metal ions.
- Secretes enzymes from their cell walls (invertase hydrolyses sucrose to glucose and fructose) and so assisting in nutrition.
- The main identifying characteristic of fungi is the makeup of their cell walls. Many contain a nitrogenous substance known as chitin which is not found in plants, but can be found in the exoskeletons of arthropods and insects.



CELL MEMBRANE

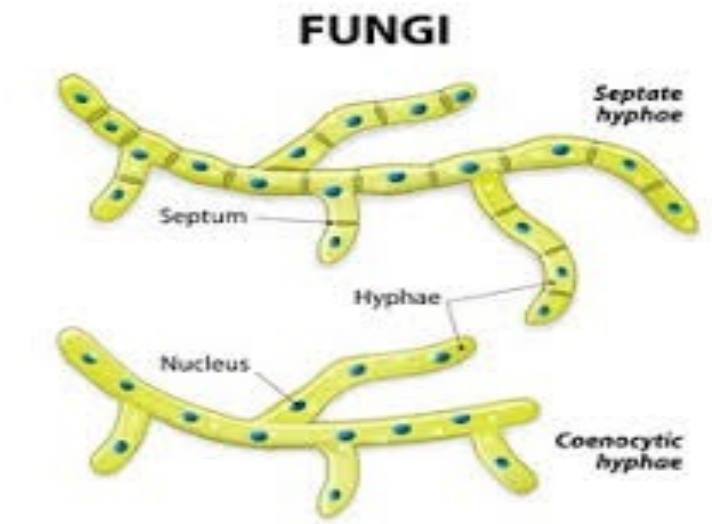
In fungi, the primary membrane sterol is ergosterol, which is unique in contrast to animals. There located various functional proteins, such as enzymes, proton pumps, and ion transporters, on the membrane surface, essential for cell permeability, signal transduction, and cell wall morphogenesis.

- Ergosterol is a sterol that resides on the cell membranes of fungi and acts to maintain cell membrane integrity, similar to mammalian cholesterol.
- Most of the current antifungal agents interfere with ergosterol function in some way, either through inhibition of various steps in ergosterol biosynthesis (allylamines, azoles, morpholines) or by complexing directly with membrane ergosterol (polyenes).
- Polyene antimycotic agents (amphotericin B, nystatin) are a subset of macrolide antibiotics that bind to ergosterol on the cell membranes of fungi. The bound drug molecules form a pore in the ergosterol which allows electrolytes and small molecules to leak out of the cell.
- Azole antifungals (fluconazole, itraconazole, ketoconazole) act to prevent the conversion of lanosterol to ergosterol. Without the protective layer of ergosterol, the cell membrane becomes permeable, leaking intracellular contents. Interestingly, the azoles have an antagonistic effect on the polyene antimycotics—they can only bind to ergosterol.

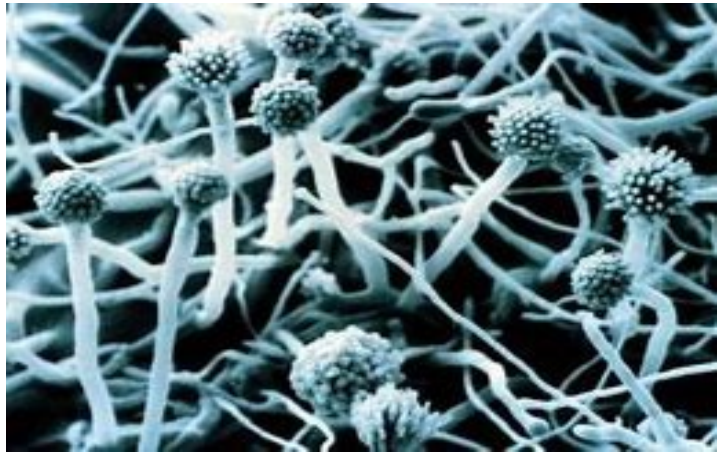


HYPHAE

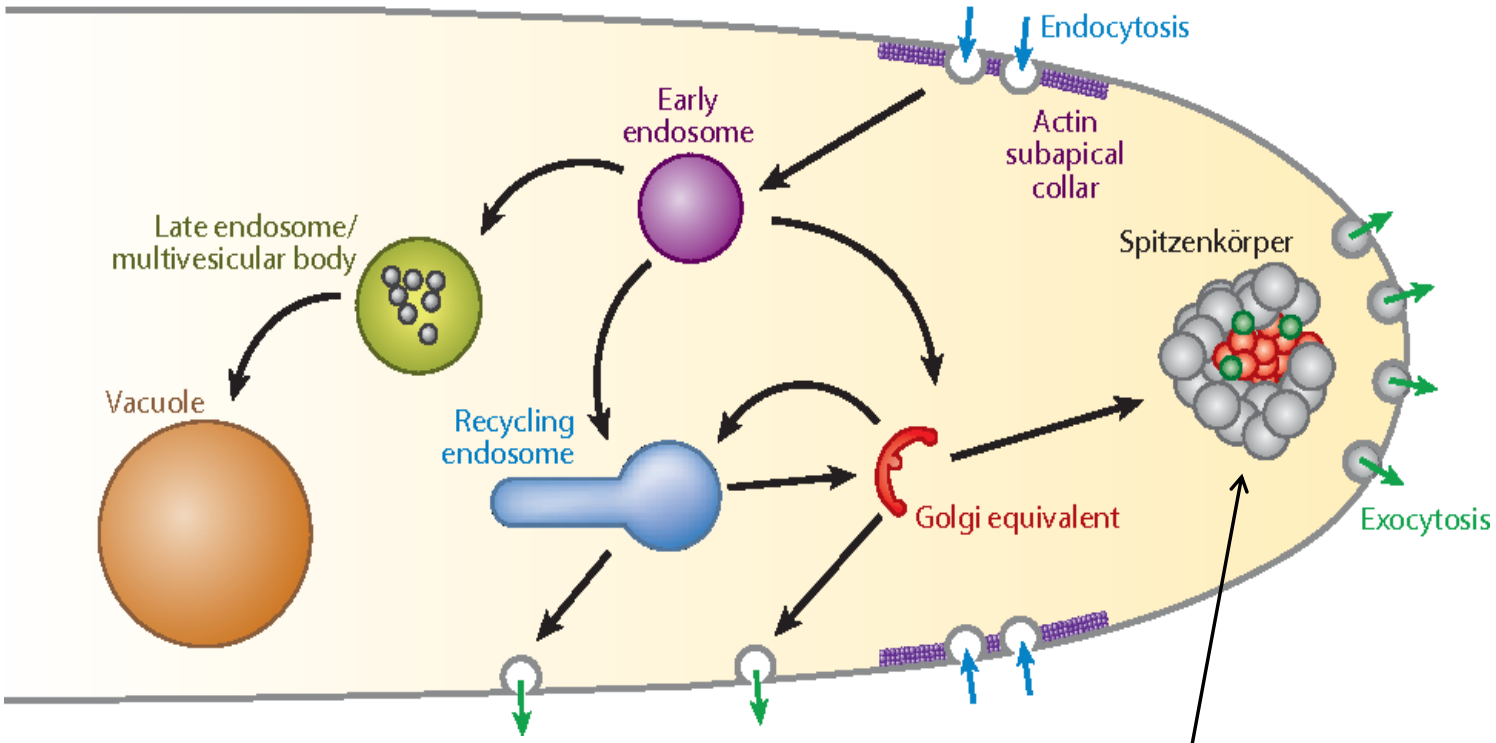
- Hyphae are comprised of hypha, which are the long filamentous branches found in fungi. Hyphae are important structures required for growth in these species, and together, are referred to as mycelium(tangled mass of hypahe)
- Each hypha is comprised of at least one cell encapsulated by a protective cell wall typically made of chitin, and contain internal septa, which serve to divide the cells. Septa are important as they allow cellular organelles (e.g., ribosomes) to pass between cells via large pores. However, not all species of fungi contain septa. The average hyphae are approximately 4 to 6 microns in size.
- Hyphae growth occurs by extending the cell walls and internal components from the tips. During tip growth, a specialized organelle called the spitzenkörper, assists in the formation of new cell wall and membrane structures by harboring vesicles derived from the golgi apparatus and releasing them along the apex of the hypha. As the spitzenkörper moves, the tip of the hypha is extended via the release of the vesicle contents, which form the cell wall, and the vesicle membranes, which create a new cell membrane. As the hypha extends, new septa can be created to internally divide the cells. The characteristic branching of hyphae is the result of the formation of a new tip from a hypha, or the division of a growing tip.



HYPHAE



MYCELIUM



Spitzenkörper

MORPHOLOGICAL CLASSIFICATION

1. Yeast – These occur in the form of round or oval bodies which reproduce by asexual process called budding in which cell develops can't protuberence which enlarges and eventually seperates from parent cell.

   Eg: Saccharomyces cerevesiae(used in production of bread and brewing)

2. Yeast like fungi – These type of fungi grow partly as yeast and partly as elongated cells resembling hyphae. The latter form pseudomycelium.

   Eg: Candida albicans

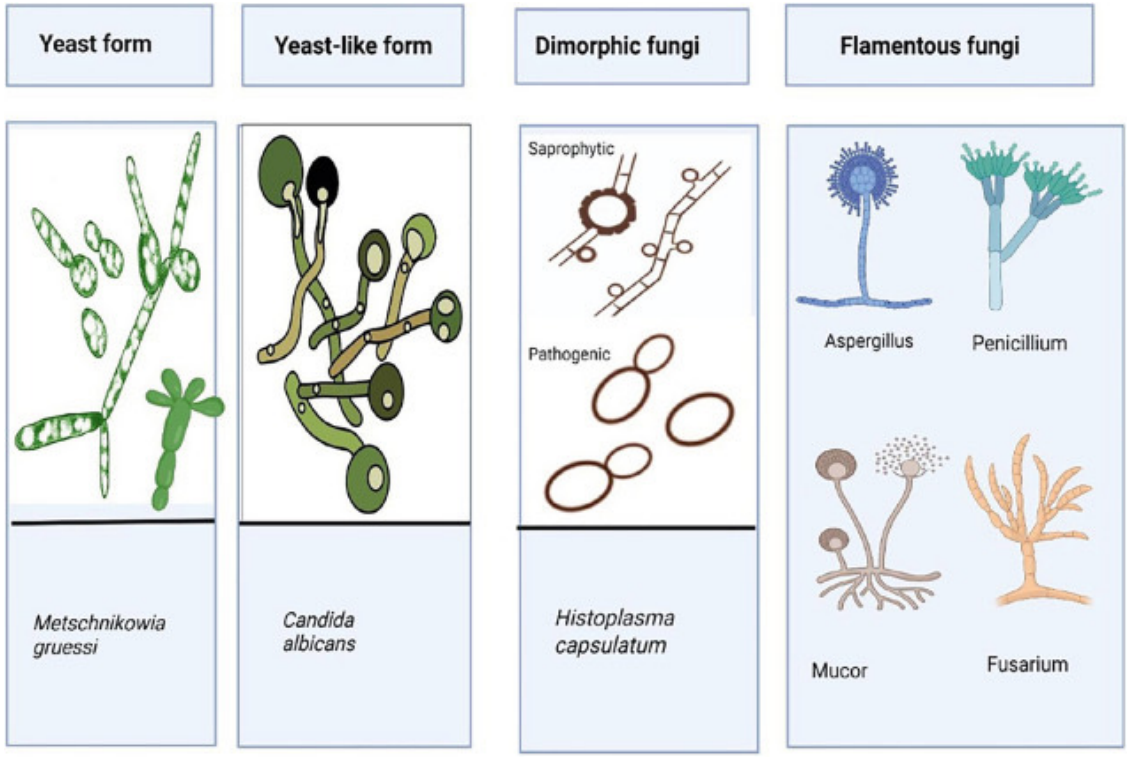
3. Dimorphic fungi – These fungi can exist in the form of both mold and yeast. This is usually brought about by change in temperature and the fungi are also described as thermally dimorphic fungi.

   Eg: Histoplasma capsulatum

4. Molds or filamentous fungi – They have long branching filaments or hyphae which interwine to produce a mass of filaments or mycelium. Their colonies are strongly adherent to the medium and unlike most bacterial colonies it cant be emulsified in water. They reproduce by the formation of different types of spores.

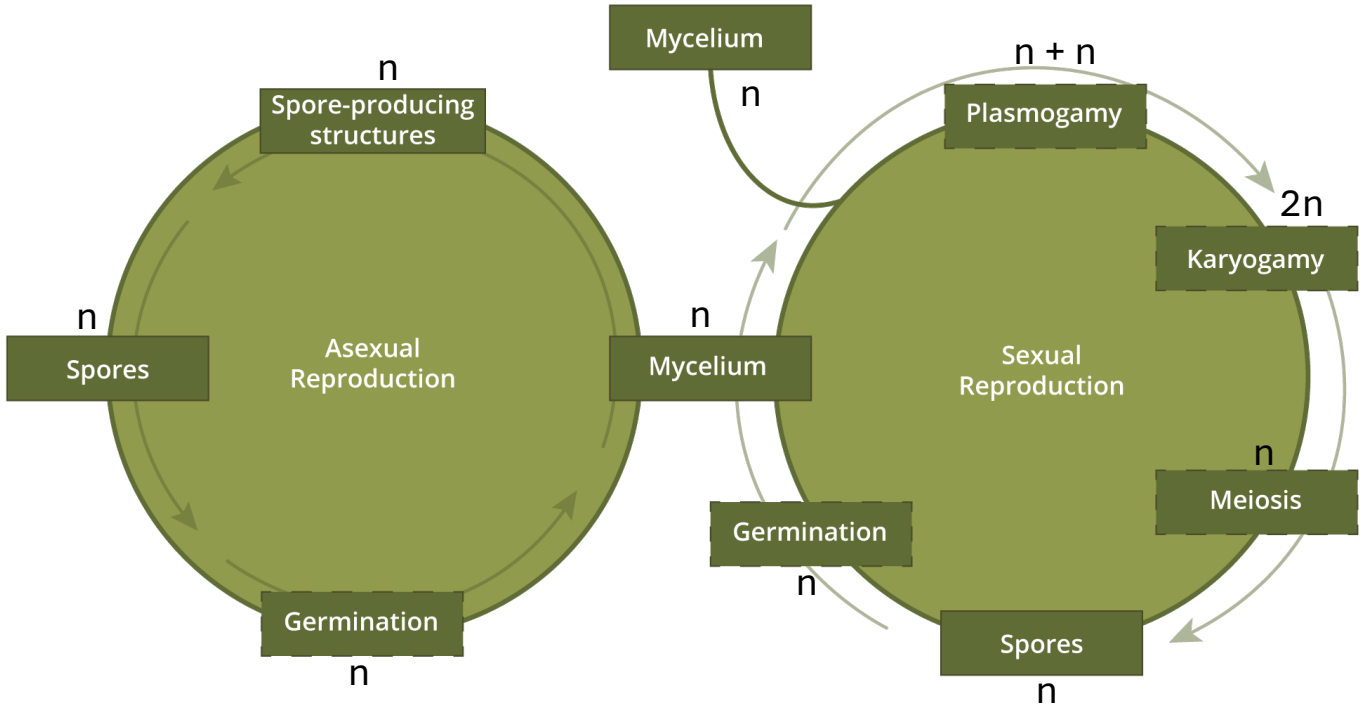
   Eg: Aspergillus(fermentation organism used for production of citric acid)

   Pencillium(source of antibiotic pencillin)



FUNGAL REPRODUCTION

- Fungi reproduce sexually and/or asexually.
- Perfect fungi reproduce both sexually and asexually, while imperfect fungi reproduce only asexually (by mitosis).
- In both sexual and asexual reproduction, fungi produce spores that disperse from the parent organism by either floating on the wind or hitching a ride on an animal. Fungal spores are smaller and lighter than plant seeds. The giant puffball mushroom bursts open and releases trillions of spores. The huge number of spores released increases the likelihood of landing in an environment that will support growth.

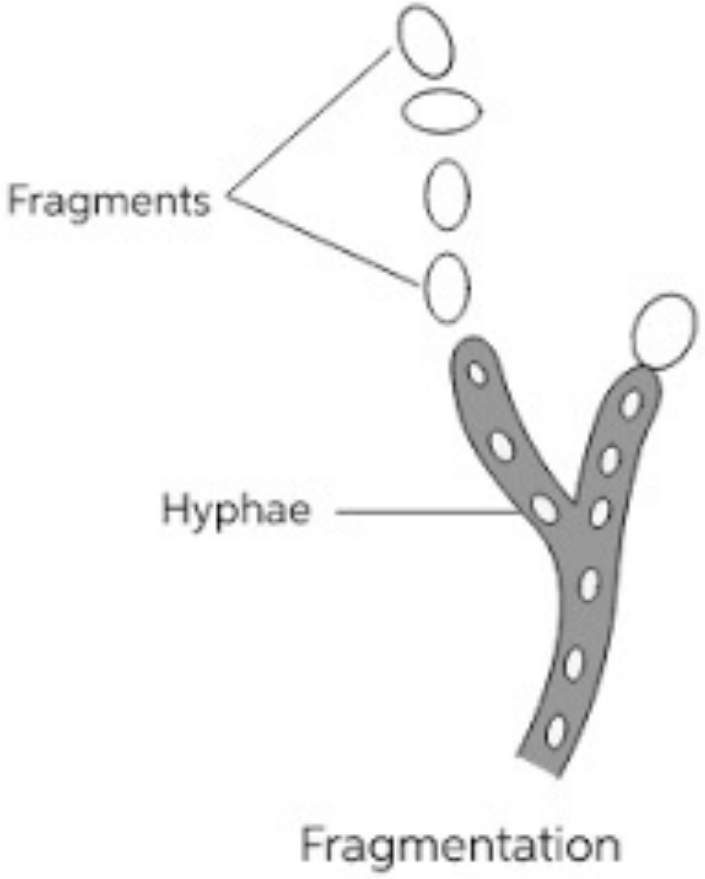


ASEXUAL REPRODUCTION

- Fungi reproduce asexually by:

1. Fragmentation
2. Budding
3. Producing spores

❖ Fragmentation : Fragmentation is seen in various types of fungi such as molds, yeasts, mushrooms. They do reproduction by fragmentation using a specific type of structure called hyphae. Hyphae can be defined as each of the branching filaments that make up the mycelium of a fungus. It is branched portion of parent fungi body and they can easily get rid of it. During the life cycle of hyphae, the obtain food and other nutrients from the parent fungal body. By doing this, hyphae eventually grow and become mature, and ultimately, they become ready for fertilization. Now a piece of hyphae breaks off from the parent body and enters into a growth phase as an individual body. Eventually they also mature and grow hyphae and this way the cycle continues.



❖ Budding : Budding, which is another method of asexual reproduction, occurs in most yeasts and in some filamentous fungi.

- In this process, a bud develops on the surface of either the yeast cell or the hypha, with the cytoplasm of the bud being continuous with that of the parent cell.
- The nucleus of the parent cell then divides; one of the daughter nuclei migrates into the bud, and the other remains in the parent cell.
- The parent cell is capable of producing many buds over its surface by continuous synthesis of cytoplasm and repeated nuclear divisions. After a bud develops to a certain point and even before it is severed from the parent cell, it is itself capable of budding by the same process.
- In this way, a chain of cells may be produced. Eventually, the individual buds pinch off the parent cell and become individual yeast cells.
- Buds that are pinched off a hypha of a filamentous fungus behave as spores; that is, they germinate, each giving rise to a structure called a germ tube, which develops into a new hypha

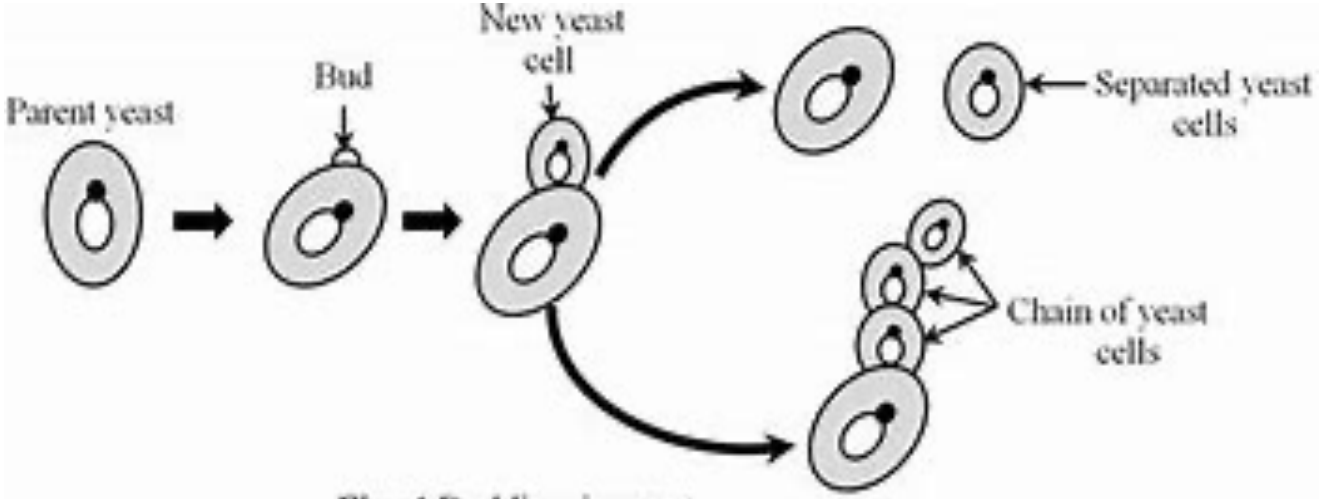


Fig. 4 Budding in yeast

❖ Spore formation : Spore formation is the characteristic feature of fungi.

- Different fungi forms different types of spore.
- Asexual spores in fungi can be of:

1. Sporangiospore:

- This is a type of aerial spore.
- These asexual spore are produced in a sac like structure called sporangia (singular;saprangium).

- Sporangium are produced at the end of special aerial hyphae called sporangiophore.
- Sporangium contains large numbers of haploid spores, which are released by rapture of

  sporangial wall

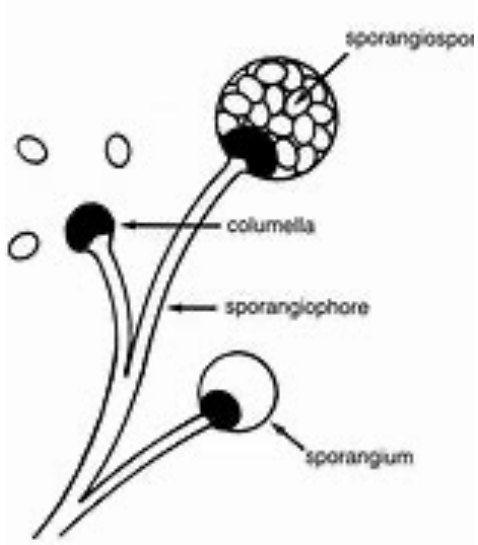
  Examples: *Rhizopus*

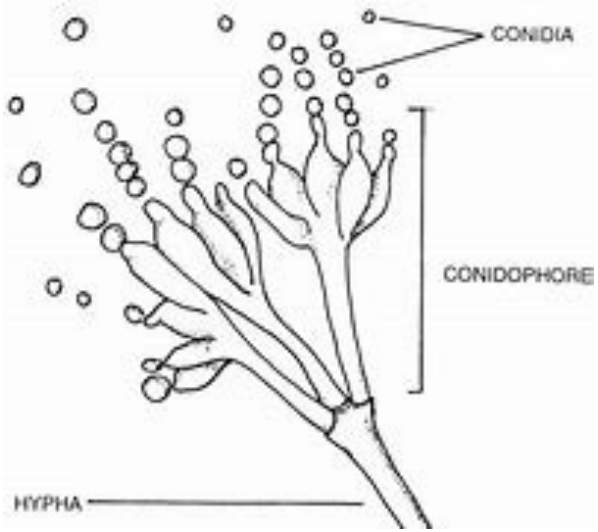
2. Conidiospore:

- Type of aerial spore found in fungi.

- Conidiospore or conidia are single celled, bicelled or multicelled structure born on the tip or side of aerial hyphal structure called conidiophore
- Conidia are different from sporangiospore as these are not produced inside sporangium or any sac like structure.
- Conidia are born singly or in chain

  Examples: *Penicillium, Apergillus*

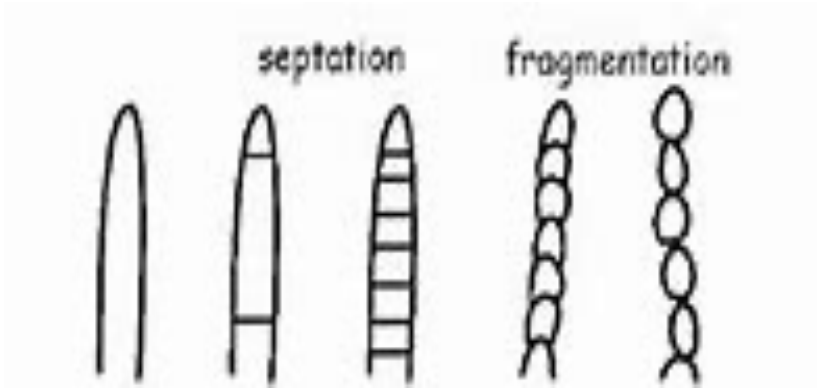




3. Arthrospore:

- They are a type of asexual vegetative spore in fungi.

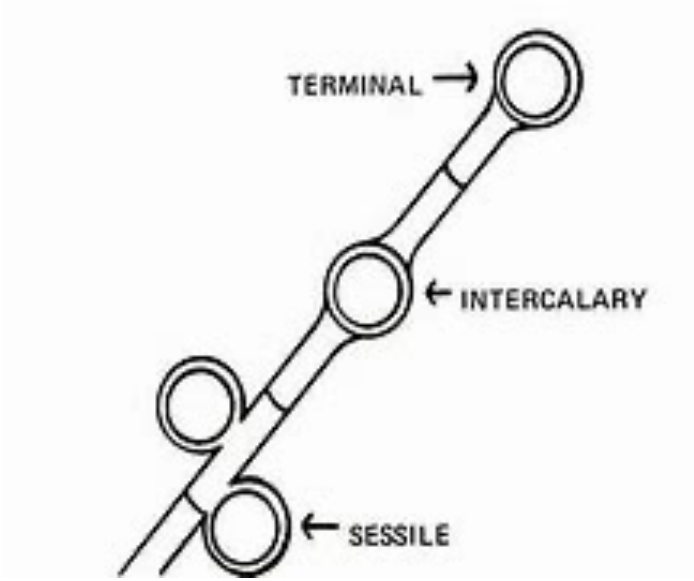
- Arthrospore are very primitive type of spore formed by the breaking up of fungal mycelium
- A spore is formed by separation followed by fragmentation of hyphae
- Examples: *Trichosporium, Geotrichum, Coccididious imitis*



4. Chlamydospore:

- A type of vegetative spore.

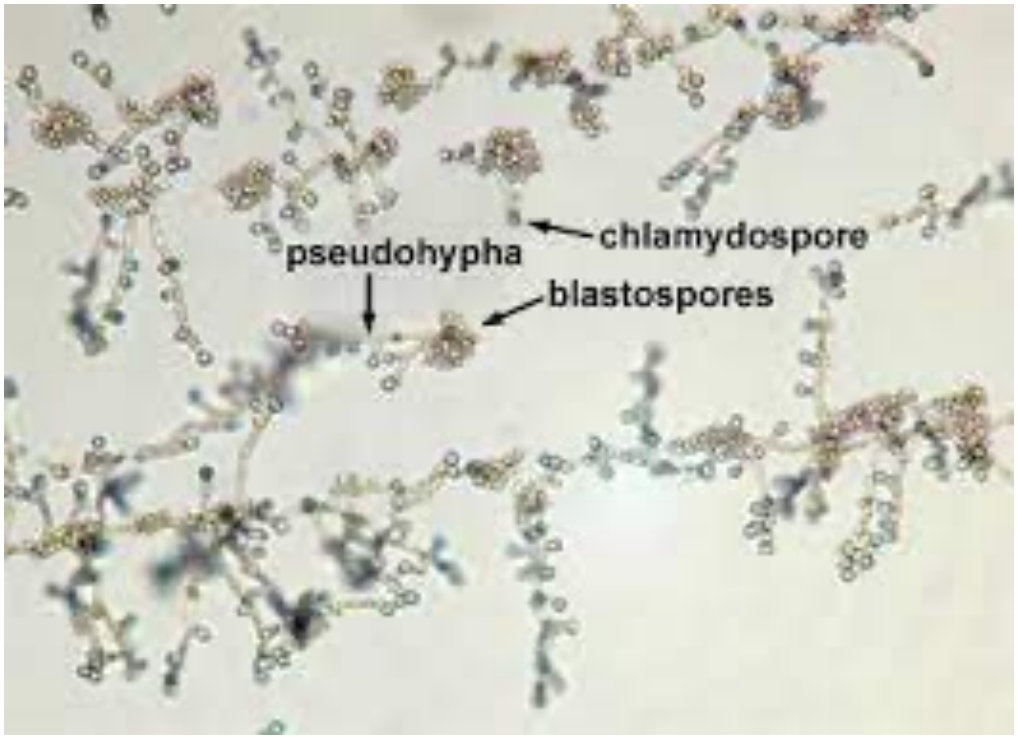
- These are usually formed during unfavorable condition and are thick walled single celled spore, which are highly resistant to adverse condition.
- Hyphal cell or portion of hyphae contracts, loose water, round up and develops into thick walled chalmydospore.
- When favorable condition returns, each chlamydospore give rise to a new individual fungi.
- Examples: ascomycetes, basidiomycetes, zygomycetes,
- *Histoplasma capsulatum, Candida albicans*



5. Blastospore:

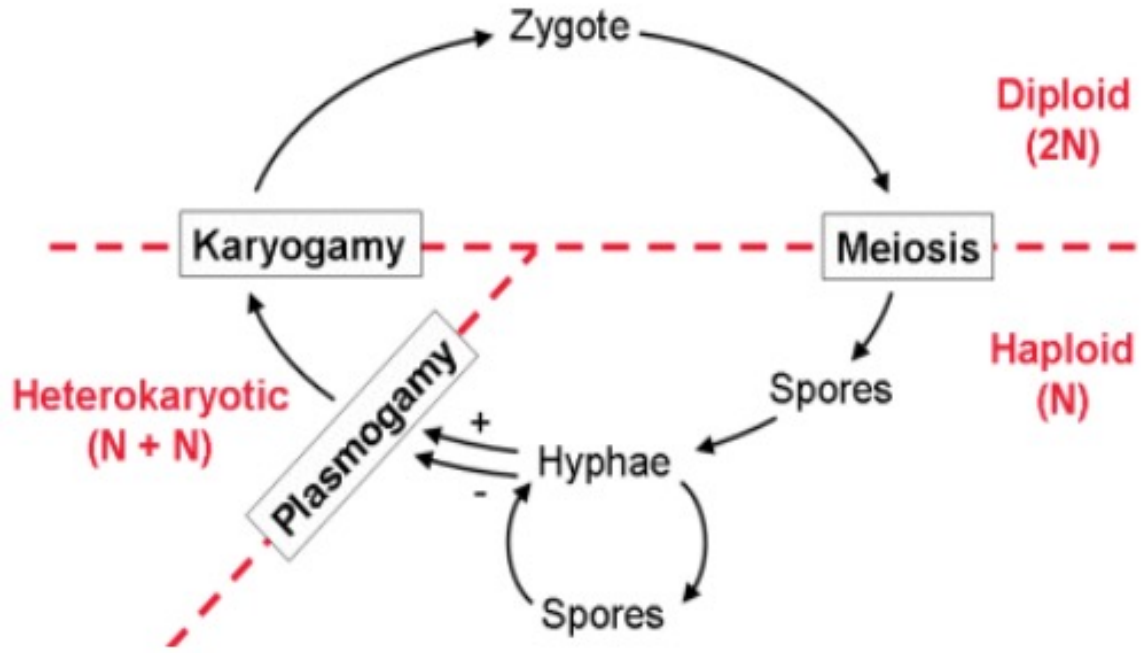
- A type of vegetative spore.

- It is a budding spores usually formed at the terminal end of hyphae.
- These spore may remains attached to hyphae and bud further to gibe branching chain of blastospores.
- Examples: ascomycetes, basidiomycetes, zygomycetes
- *Candida albicans*



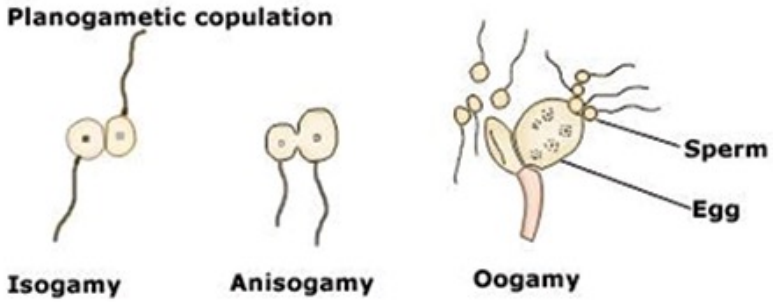
SEXUAL REPRODUCTION

- Sexual reproduction is carried out by diffusion of compatable nuclei from two parent at a definite state in the life cycle of fungi.
- The process of sexual reproduction involves three phases:
  - Plasmogamy: fusion of protoplasm
  - Karyogamy: fusion of nucleus
  - Meiosis: reductional nuclear division
- Various methods by which compatible nuclei are brought together in plasmogamy. Some are:
  1. Gametic copulation
  2. Gamete- gametangial copulation
  3. Gametangial copulation
  4. Somatic copulation
  5. Spermatization



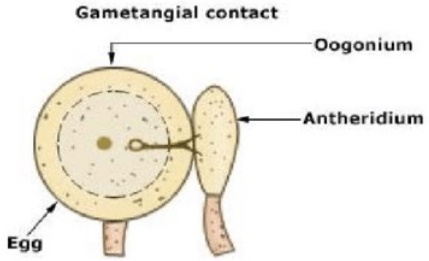
1. Planogametic copulation:

- Fusion of two naked gametes, one or both of them are motile
  - Isogamous – gamates of similar morphology.
  - Anisogamous – Fusion of two games which differ in size or form.
  - Oogamous – Small motile male gamate and large immobile female gamate.



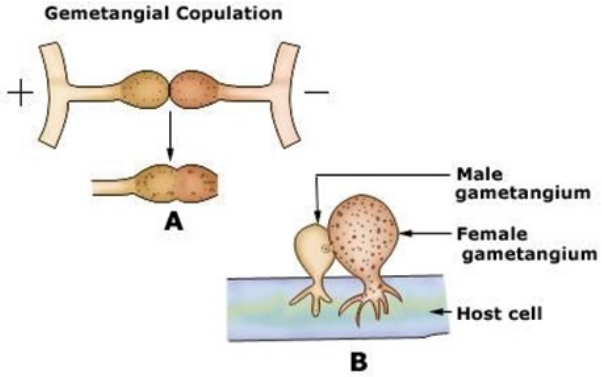
2. Gamete-gametangial copulation:

- Male and female gametangia (organ or cell in which gamates are formed) comes into contact but do not fuse.
- A fertilization tube formed from where male gametangium enters the female gametangium and male gamate passes through this tube.



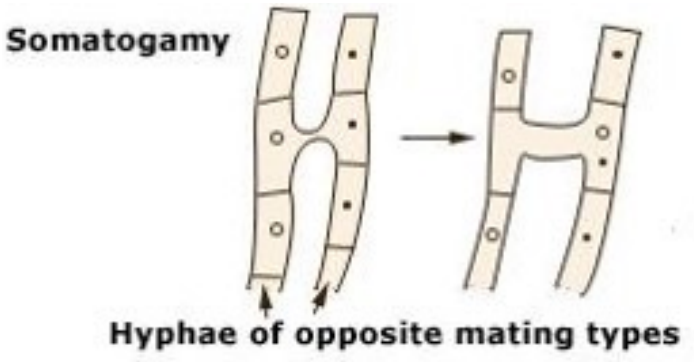
3. Gametangial copulation;

- Two gametangia or their protoplast fuse and give rise to zygospore



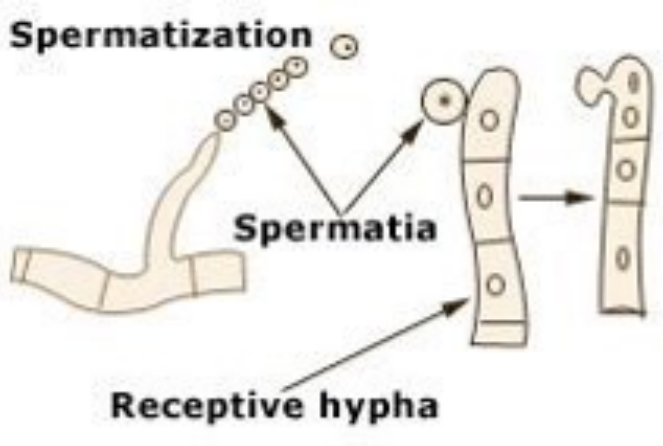
4. Somatic copulation:

- Also known as somatogamy.
- In this process fusion of somatic cell (cells except sperm and egg cells) occurs
- This sexual fusion of undifferentiated vegetative cell results in dikaryotic hyphae, so the process is also called dikarotization



5. Spermetization:

- It is an union of special male structure called spertatium with a female receptive structure.
- Spermatium empties its content into receptive hyphae during plasmogamy.

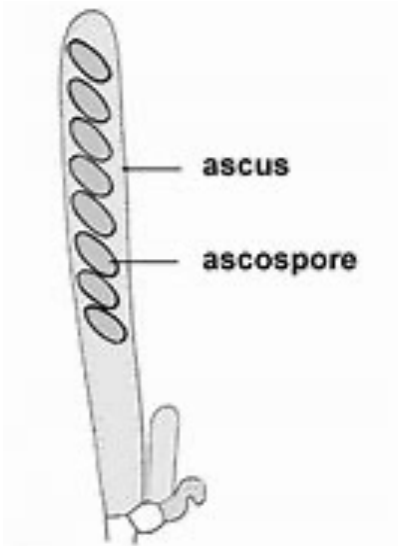


SEXUAL SPORES

- As a result of sexual reproduction sexual sores are produced.
- Sexual spores are fewer in number than asexual spores.
- Sexual spores in fungi can be of :

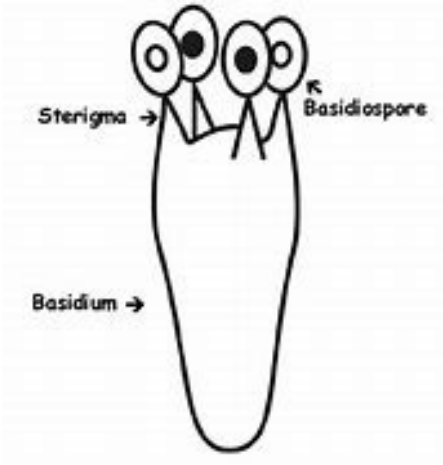
1. Ascospore:

- It is usually single celled produced in a sac called ascus (plural;asci) and usually there are 4-8 ascospore in an ascus but the number may vary from species to species
- The ascospore are usually arranged in a linear order. In some case ascospores are long, narrow and are arranged in parallel order.



2. Basidiospore:

- It is a reproductive spore produced by basidiomycetes.
- This single celled spores are born in a club shaped structure called basidium
- These basidiospore aerves as main air dispersal unit for the fungi.

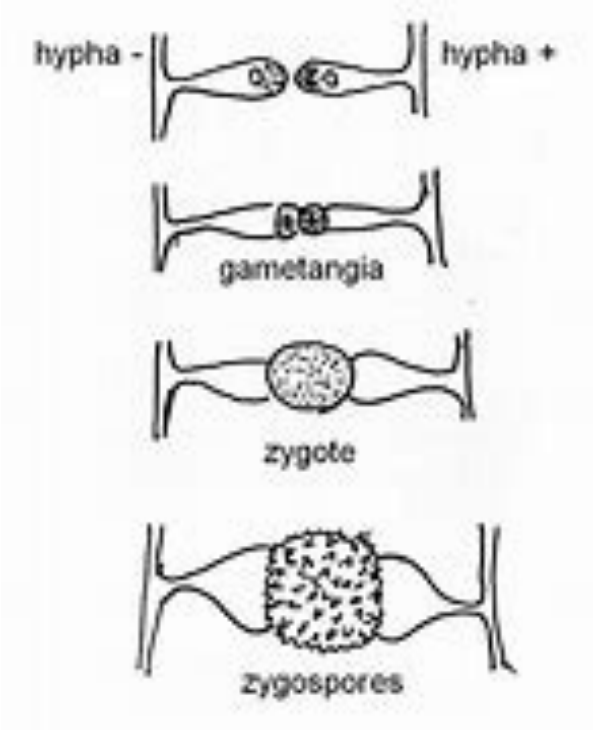


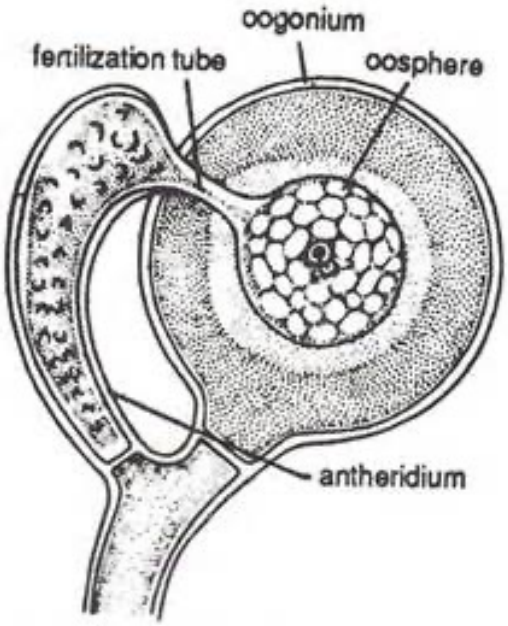
3. Zygospore:

- Zygospores are thick walled spores formed when two sexually compatible hyphae or gametangia of certain fungi fuse together.
- In suitable condition, zygospore germinates to produce a single vertical hyphae which forms a aporangium and releases its spores.

4. Oospore:

- These are formed within a special female structure called Oogonium.
- Fertilization of egg by male gamete in female sex organ give rise to oospoes.
- There are one or more oospores in each oogonium.





Fungal states

- **Teleomorph**: the sexual reproductive stage (morph),***Emericella nidulans***.

- **Anamorph**: an asexual reproductive stage (morph), often mold-like (e.g. *Aspergillus nidulans* .

- **Holomorph**: the whole fungus, including all anamorphs and the teleomorph.

Homothallic fungi

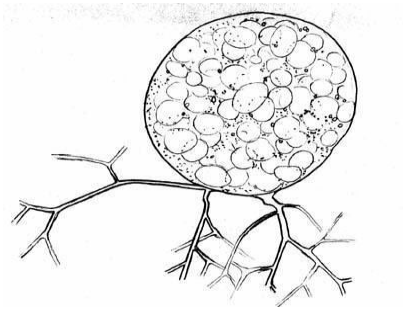
- possess both male and female nuclei derived from the same thallus for sexual reproduction. They do not need a partner for sexual reproduction. This is a form of self-fertilization or selling. Homothallism is a common condition in fungi . it causes reduced genetic variability.
- Self – fertile fungi

Heterothallic fungi

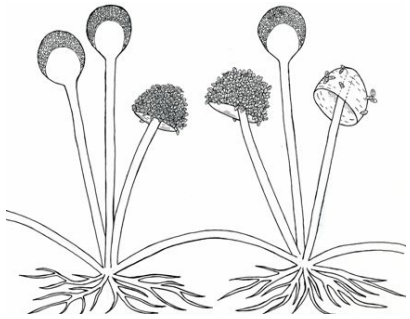
- are the fungal strains which bear one type of mating type. They are unisexual in nature. Sexual reproduction of heterothallic fungi occurs between two different compatible mycelia. the genetic variation within the populations is high.
- Self – sterile fungi.

TYPES OF FUNGI

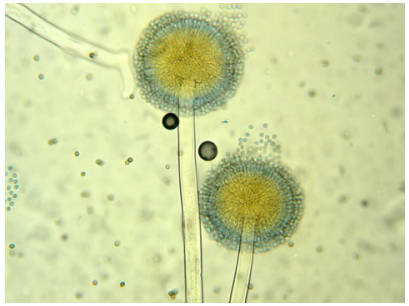
- There are mainly five types of fungi
  * Chytridiomycota
  * Zygomycota
  * Ascomycota
  * Basidiomycota
  * Deuteromycota



Chytridiomycota



Zygomycota



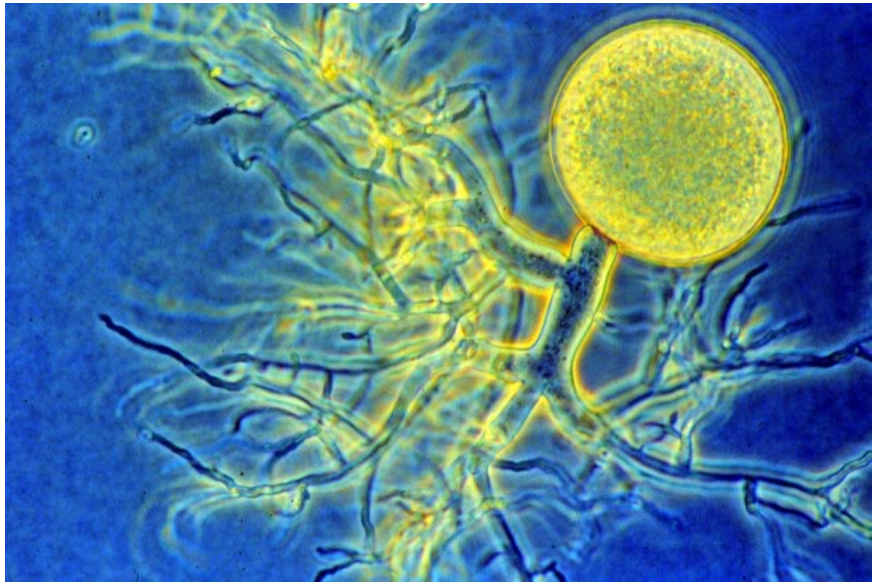
Deuteromycota

Ascomycota

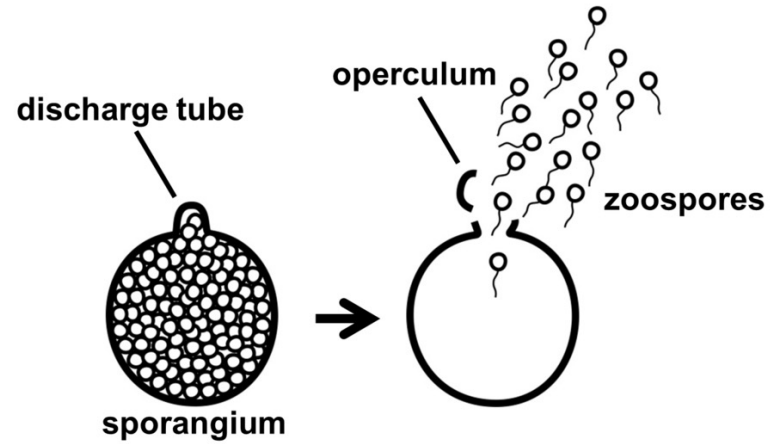
Basidiomycota

Phylum Chytridiomycota

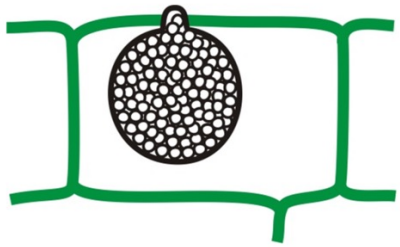
- Earliest fungal phlyum to diverge
- Relatively simple; most unicellular
- Coenocytic hypha
- ONLY FUNGI with flagellated cells
- Most have no sexual reproduction
- Most decomposers; few cause disease
- Some species are saprotrophic; others are parasites of plants, animals, algae and other fungi



Dead frog with chytridiomycosis

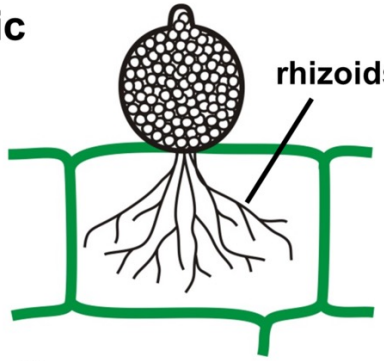


holocarpic

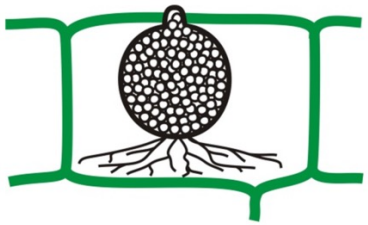


eucarpic monocentric

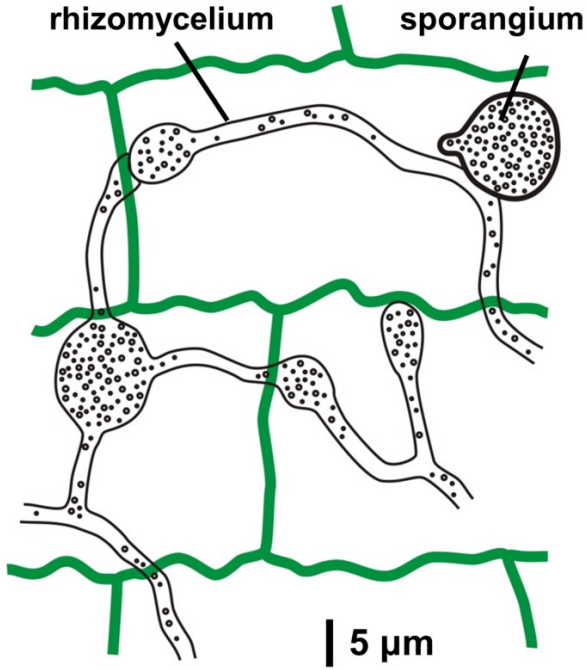
epibiotic



endobiotic

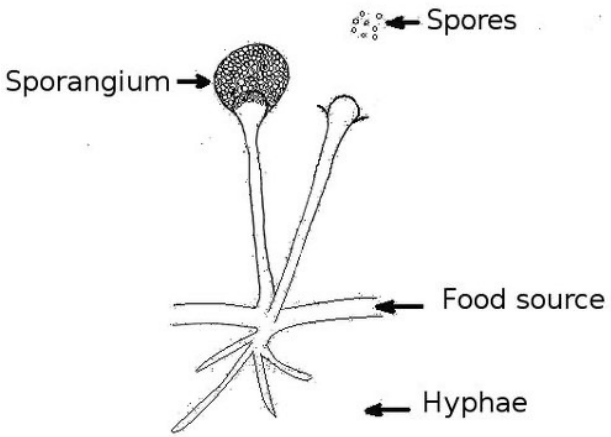


eucarpic polycentric



Phylum Zygomycota: The Conjugation Fungi

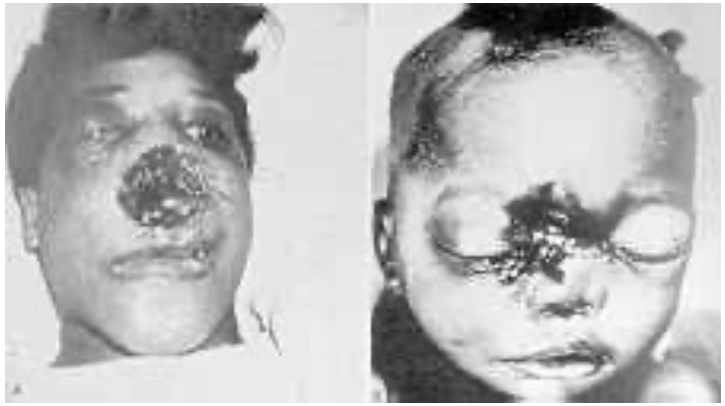
The Zygomycota, or conjugation fungi, include molds, such as those that invade breads and other food products. The identifying characteristics of the Zygomycota are the formation of a zygospore during sexual reproduction and the lack of hyphal cell walls (non-septate or coenocytic hypha) except in reproductive structures. Many (~100 species) are known plant root symbionts



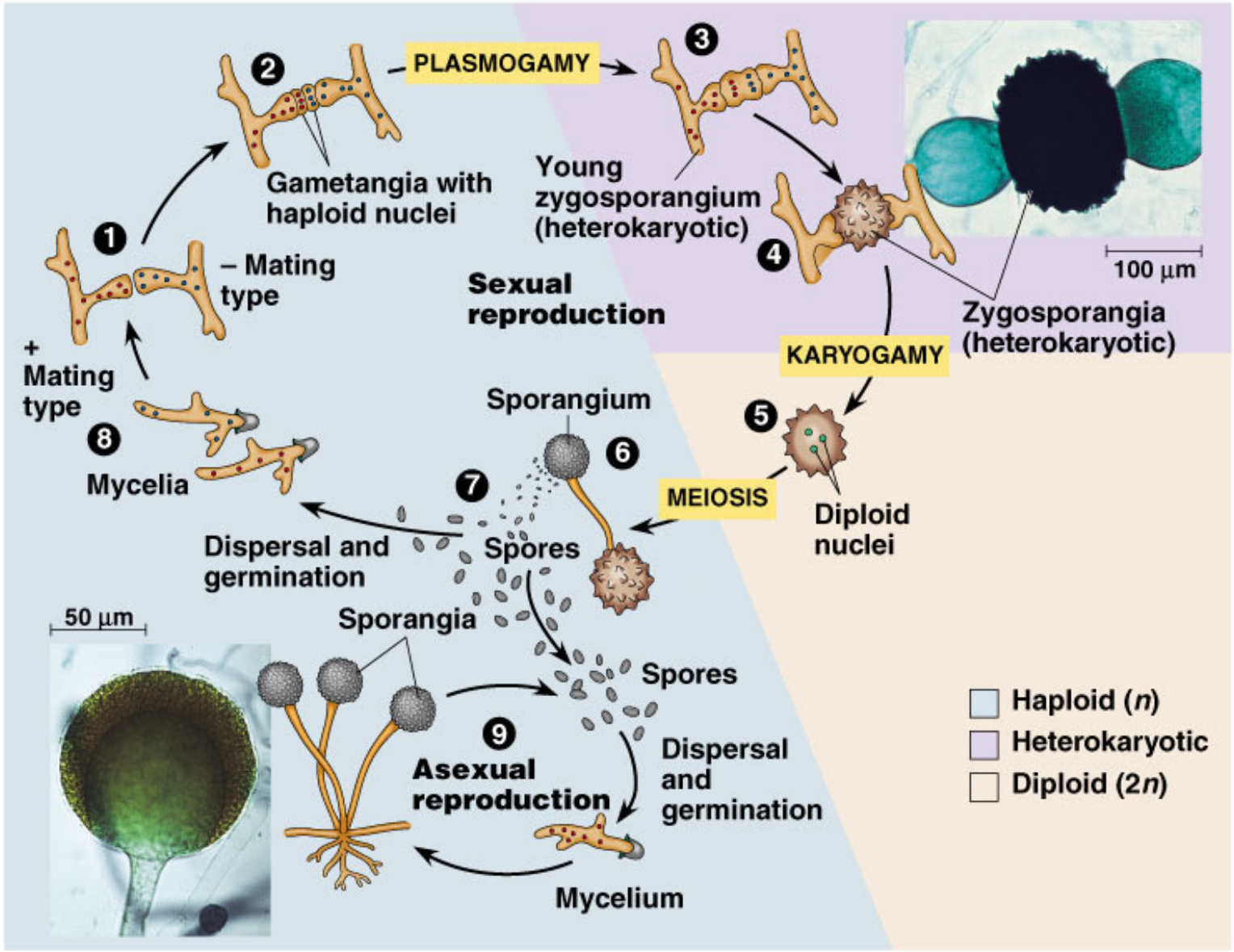
- several species of zygomycota cause serious human infections and animals
- They are being increasingly used in the biological control of insect pests of crops.

Rhinocerebral zygomycosis (disease)

- Sexual Reproduction - zygosporangia
- Asexual reproduction– by common (sporangia – bags of asexual spores)



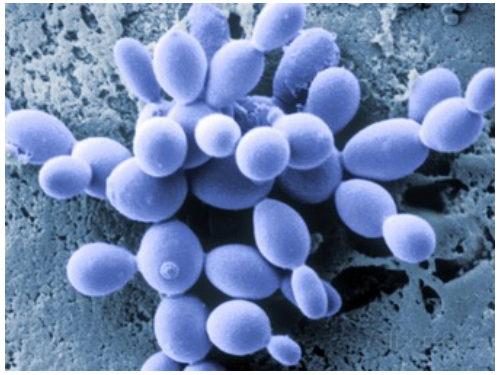
Zygomycosis

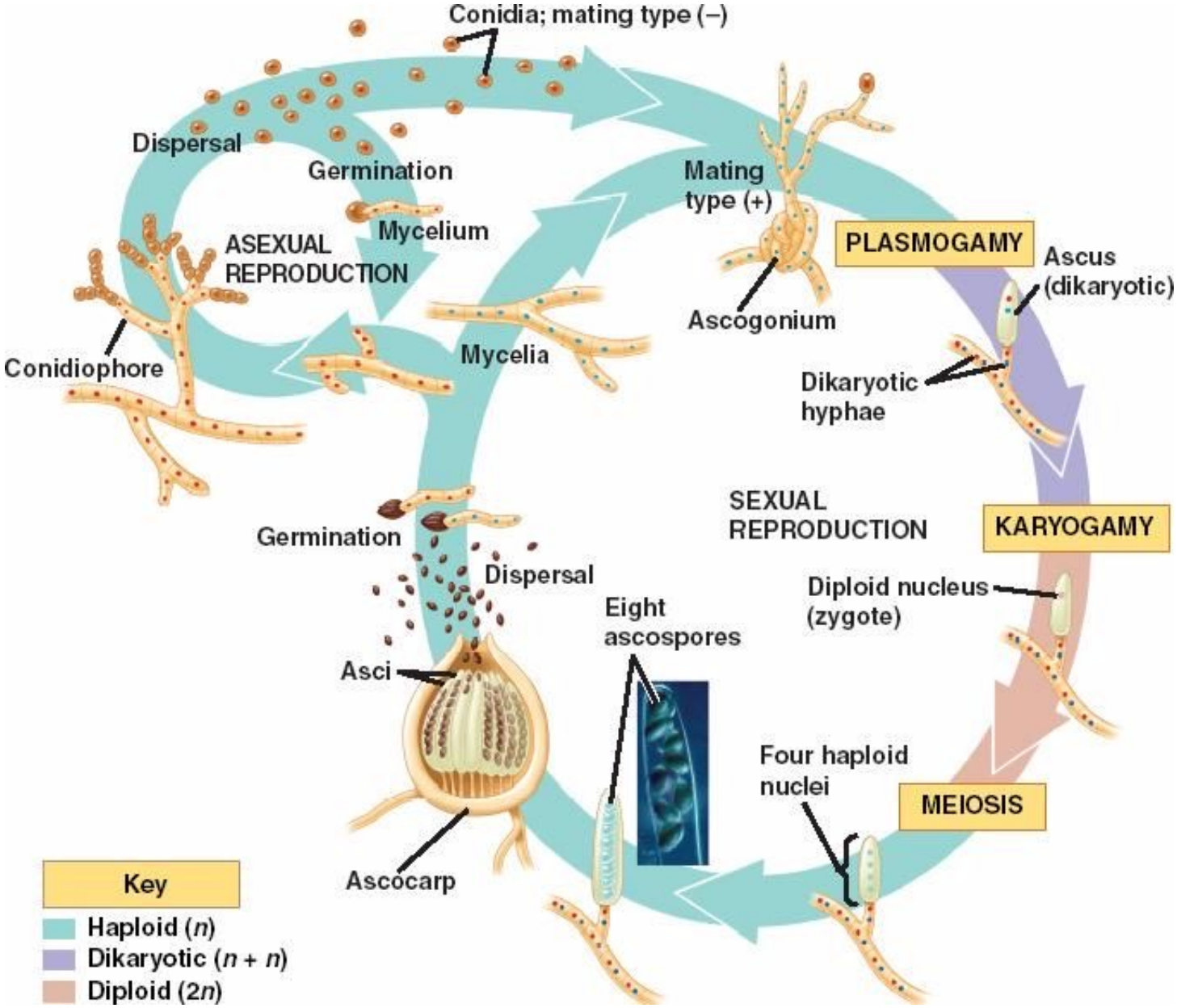


Phylum Ascomycota

"sac fungi" or "cup fungi"

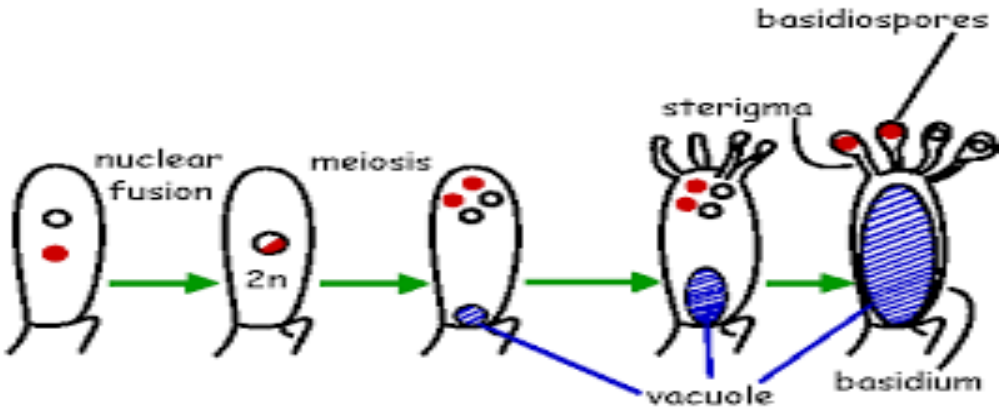
- Includes yeasts, powdery mildews, molds
- Hyphae with perforated septa
- Asexual reproduction by conidiophores
- The Ascomycota are morphologically diverse. The group includes organisms from unicellular yeasts to complex cup fungi.
- Characterized by:
- First, they can produce conidiophores for asexual reproduction.
- Secondly, ascomycota produce structures for sexual reproduction called Asci.
- There are many famous and infamous organisms: *Saccharomyces cervisiae* (baker's yeast), *Penicillium chrysogenum* (penicillin), *Morchella esculenta* (morels), *Neurospora crassa*,
- Used in genetic studies and molecular studies *Aspergillus flavus* (aflatoxin and ochratoxin). *Claviceps purpurea* (Ergot)

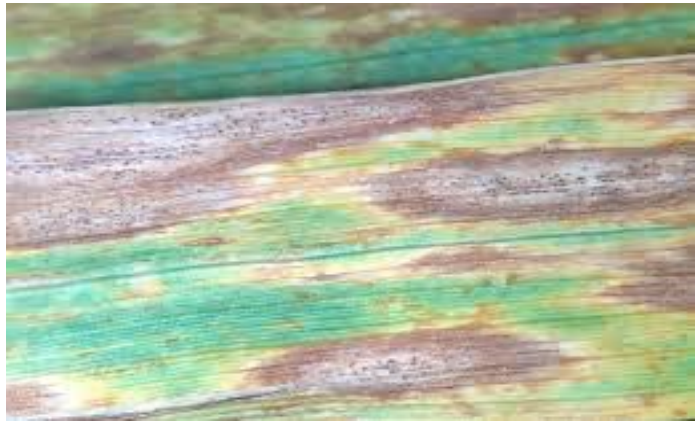


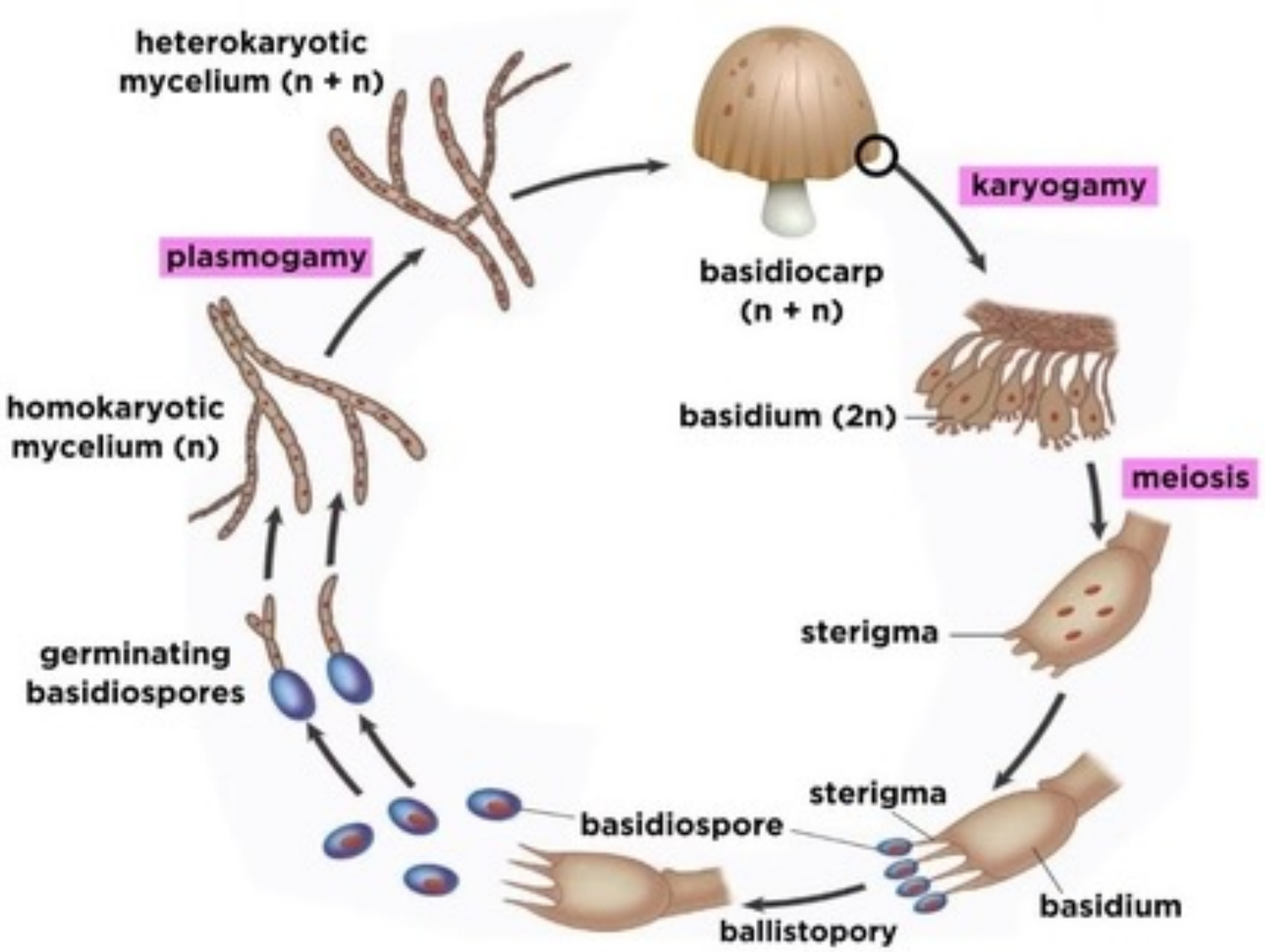

Conidia; mating type (–)
Dispersal
Germination
Mycelium
ASEXUAL REPRODUCTION
Conidiophore
Mycelia
Mating type (+)
Ascogonium
PLASMOGAMY
Ascus (dikaryotic)
Dikaryotic hyphae
SEXUAL REPRODUCTION
KARYOGAMY
Diploid nucleus (zygote)
MEIOSIS
Four haploid nuclei
Eight ascospores
Asci
Ascocarp
Germination
Dispersal
Key
Haploid (n)
Dikaryotic (n + n)
Diploid (2n)

Phylum Basidiomycota

- Mushrooms, bracket fungi, puffballs
- The members include rusts, smuts, mushrooms, puff balls, toad stools, bracket fungi etc.
- Basidiospores are developed exogenously on the horn-shaped structure, called sterigmata (generally 4) on the Basidium






heterokaryotic
mycelium (n + n)
karyogamy
basidiocarp
(n + n)
basidium (2n)
meiosis
sterigma
sterigma
basidiospore
basidium
ballistopory
germinating
basidiospores
homokaryotic
mycelium (n)
plasmogamy

Examples of Phylum Basidiomycota

Orange Jelly

Pigskin Poison Puffball

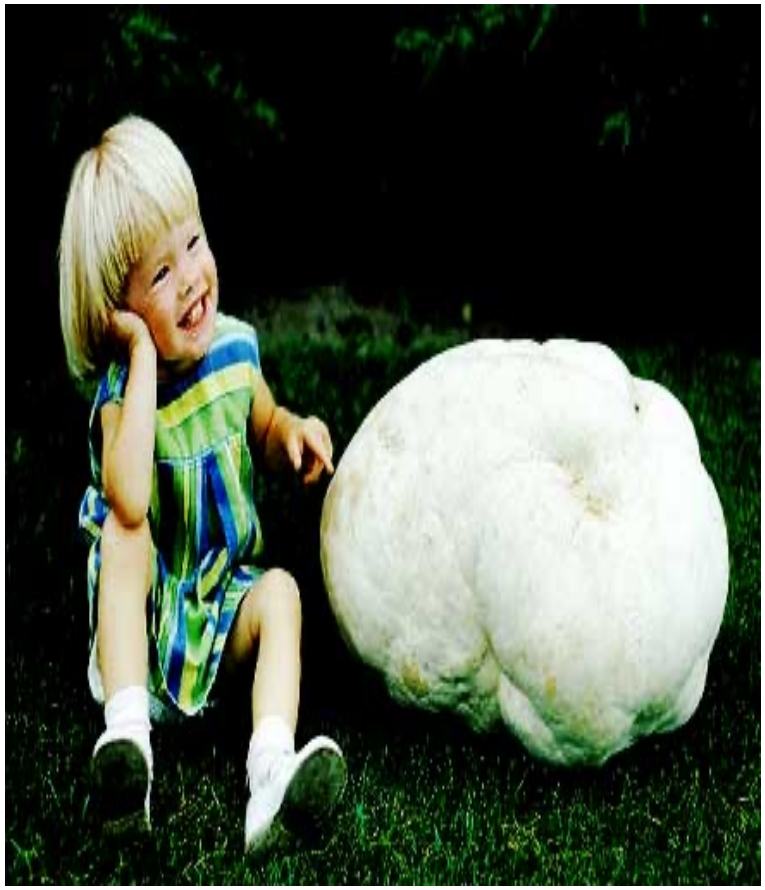
Fly Agaric

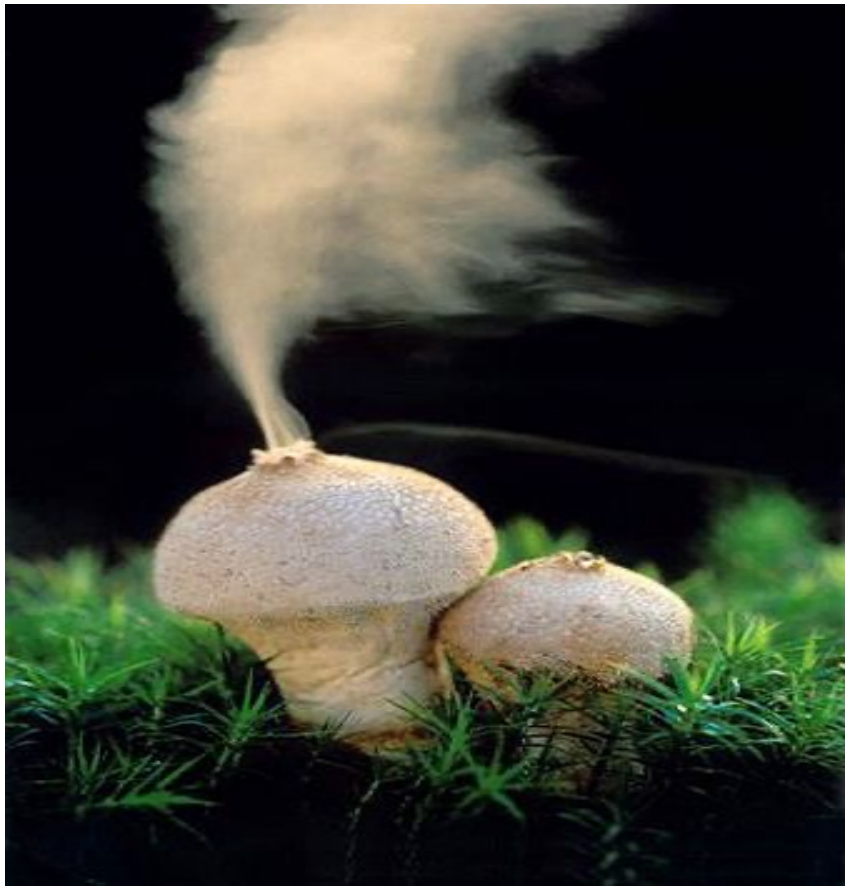
Bird's Nest Fungus

Star Stinkhorn

Shelf Fungus

Fairy Ring





- Basidiomycota play a significant role in the carbon cycle. Unfortunately, Other Basidiomycota cause diseases in animals, including humans. Basidiomycota frequently attack the wood in buildings and other structures, which has negative economic consequences for humans.

PHYLUM DEUTEROMYCOTA

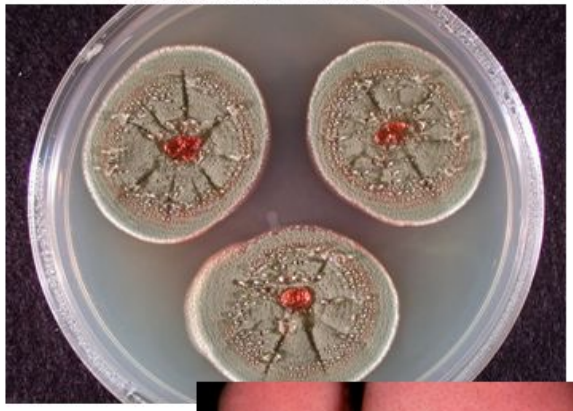
No Longer Exist!!

- 22,000 species.
- No known sexual stage.
- Saprophytic, parasitic and predatory.
- Many produce conidia.
- Most classified as Ascomycota.
- *Fusarium* wilt of tomato, potato and cotton.
- Athletes foot, ring worm
- Imperfect fungi

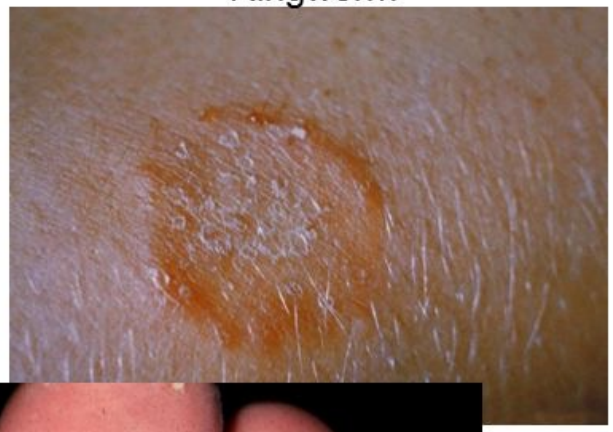
(*Penicillium* sp. *and Aspergillus* sp.)

Examples of Phylum Deuteromycota

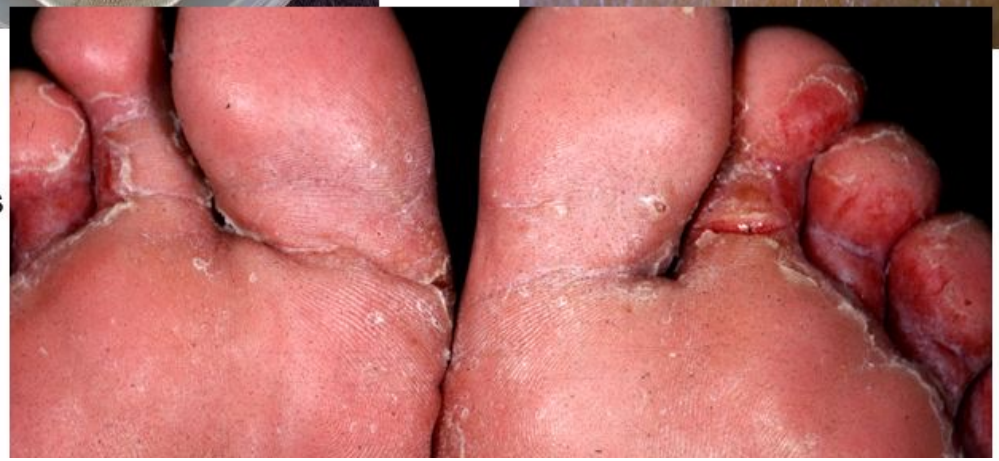
Penicillium notatum



Ringworm

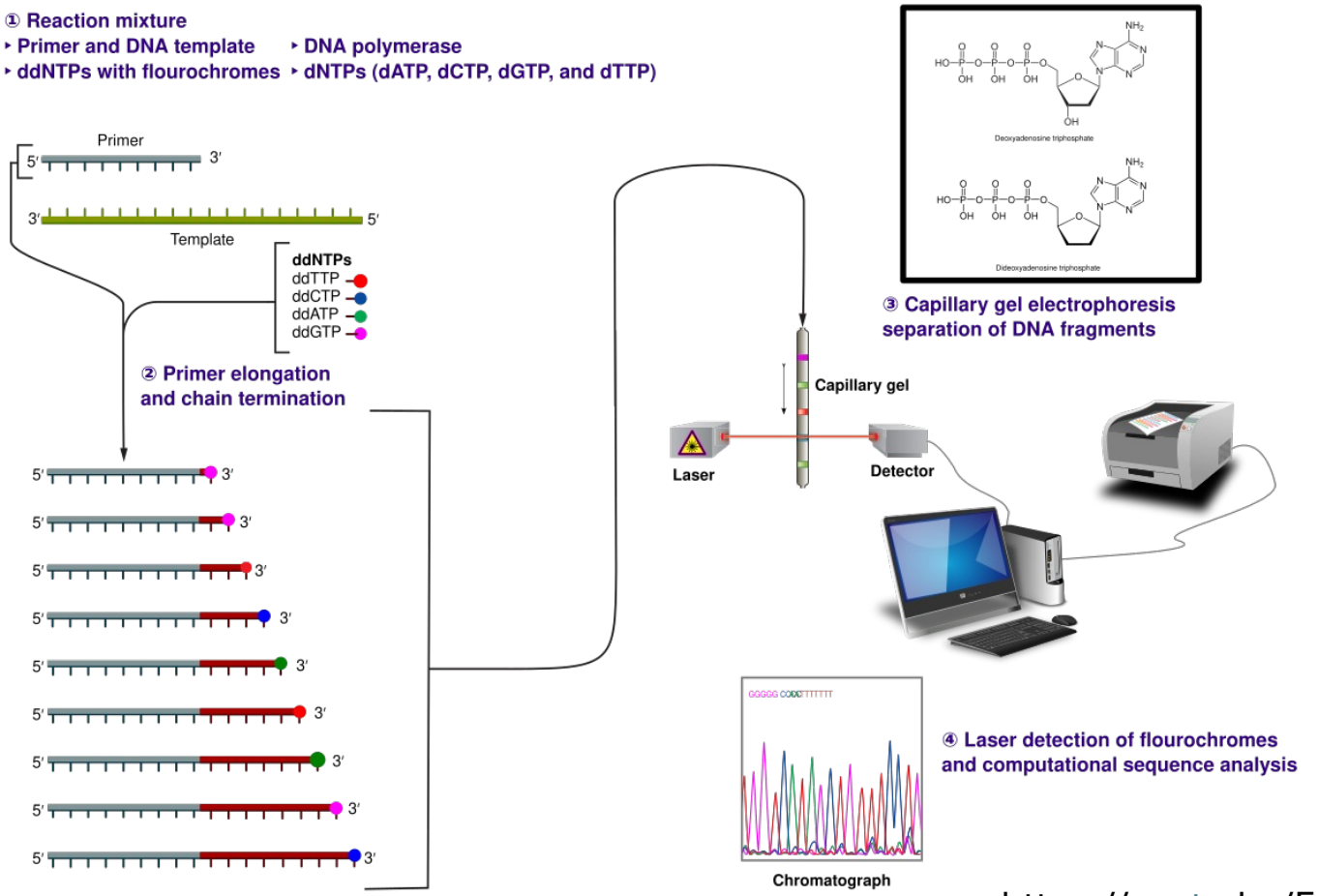


Athlete's Foot



3.2 Genome Sequencing

3.2.1 Sanger Sequencing



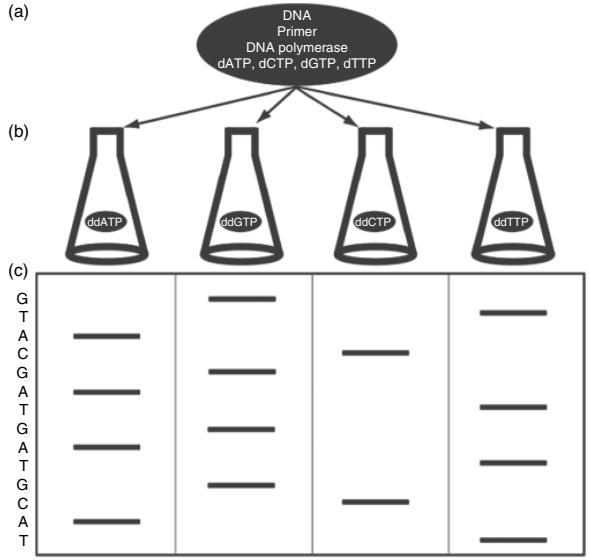


Figure 3.2 Schematic of the Sanger sequencing method. (a) Four separate DNA extension reactions are carried out. Materials required include single-stranded DNA, DNA polymerase, DNA primers, and all four dNTPs. One of the dNTPs is radioactively labeled to enable visualization in part c). (b) Each of the four reactions contains a different dideoxynucleoside triphosphate (ddNTP). Synthesis continues until a ddNTP is incorporated, terminating extension reaction. (c) Products are separated based on size on a polyacrylamide gel and the sequence can be determined.

https://youtu.be/FvHRio1yyhQ?si=k3lqmSsmra6M628L

3.2.2 Next-Generation Sequencing

3.2.2.1 Roche/454 GS FLX Pyrosequencer https://youtu.be/bNKEhOGvcaI?si=UffBOneqlPvaBrx9

3.2.2.2 Illumina Genome Analyzer https://youtu.be/CZeN-IgjYCo?si=soDI5ClEkiwtwK8A

3.2.2.3 PacBio SMRT Sequencing https://youtu.be/_lD8JyAbwEo?si=NeWPCD1tvlnedMBe

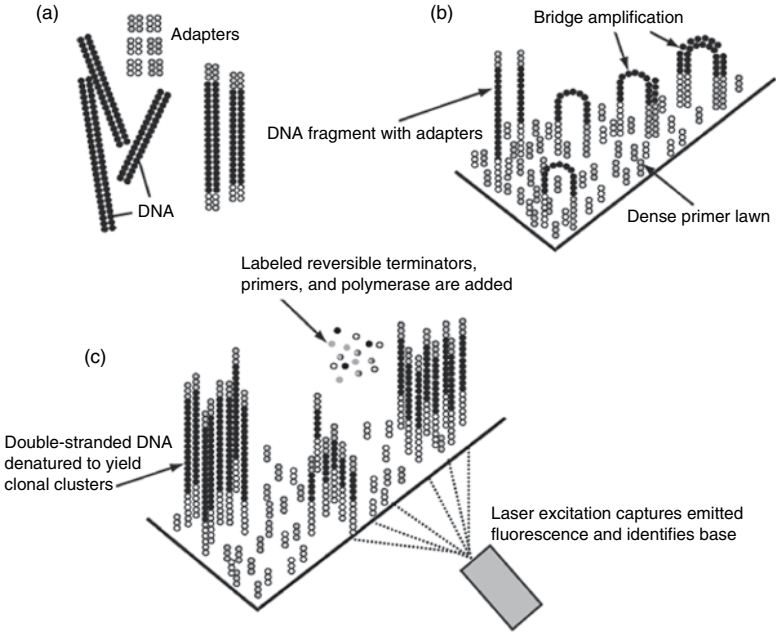


Figure 3.3 Schematic of the Illumina genome analyzer sequencing technology. (a) DNA is fragmented and adapters are ligated to both ends of the fragments. (b) Single-stranded fragments bind randomly to the surface of the flow cell (see main text). (c) Sequencing by synthesis (see main text).

3.3 Bioinformatics Tools

3.3.1 Locating Homologs

Sequence similarity searches are an essential component of genomic studies. They allow researchers to identify homologs and conserved structural motifs, and help assign putative functions to unannotated genes in *de novo* genomes. Since 2002, there has been an exponential increase in the quantity of genetic data available in public databases such as NCBI (Table 3.1). To utilize this deluge of genetic data it is imperative we have efficient similarity search techniques.

Table 3.1 Useful online resources.

Database	URL address
SGD	www.yeastgenome.org
CGD	www.candidagenome.org
AspGD	www.aspgd.org
CADRE	www.cadre-genomes.org.uk
Aspergillus fumigatus database	www.aspergillusgenome.org
CandidaDB	www.candidagenome.org
NCBI	www.ncbi.nlm.nih.gov
Sanger Institute	www.sanger.ac.uk
EMBL	www.embl.de
DDBJ	www.ddbj.nig.ac.jp
Swiss-Prot	http://expasy.org/sprot/
Fungal Tree of Life	https://aftol.umn.edu/
CGOB	http://cgob.ucd.ie/
Genomes OnLine Database (GOLD)	https://gold.jgi.doe.gov/

3.3.1.1 Global and Local Alignments

https://www.youtube.com/watch?v=dYuktSSPfYQ

3.3.1.3 FASTA

Like BLAST, FASTA is a program for rapid alignment of pairs of protein or DNA sequences and was the first widely used algorithm utilized for database similarity searching. FASTA begins by locating subsequences above a particular word length from the database sequence to subsequences of the query sequence. in FASTA the word length parameter is termed *ktup* and it is equivalent to W in BLAST searches. FASTA generates diagonal lines on a dotplot where residues match up (Figure 3.5).

The FASTA algorithm next locates diagonal regions in the alignment matrix that contain as many *ktup* matches as possible with short distances separating them (Figure 3.5). The top 10 highest scoring diagonal regions are retained and correspond to high-scoring local alignments that do not contain gaps.

FASTA then determines which of the adjacent diagonals can be joined together, thereby increasing the overall length of the alignment. For each diagonal that is connected, a joining penalty is invoked, and the overall score is determined by addition of the net scores of individual diagonals minus the joining penalties. The score of the enlarged diagonals is referred to as *initn*. All enlarged diagonals are ranked based on score and the highest scoring ones are aligned optimally using a local alignment strategy. Finally, FASTA assesses the significance of the alignment by randomly generating sequences of similar length and composition as the query sequences and calculates the probability that an alignment would be seen by random chance.

https://youtu.be/WlfF1IzHHxc?si=IQb-LPNwIk8H_HNe

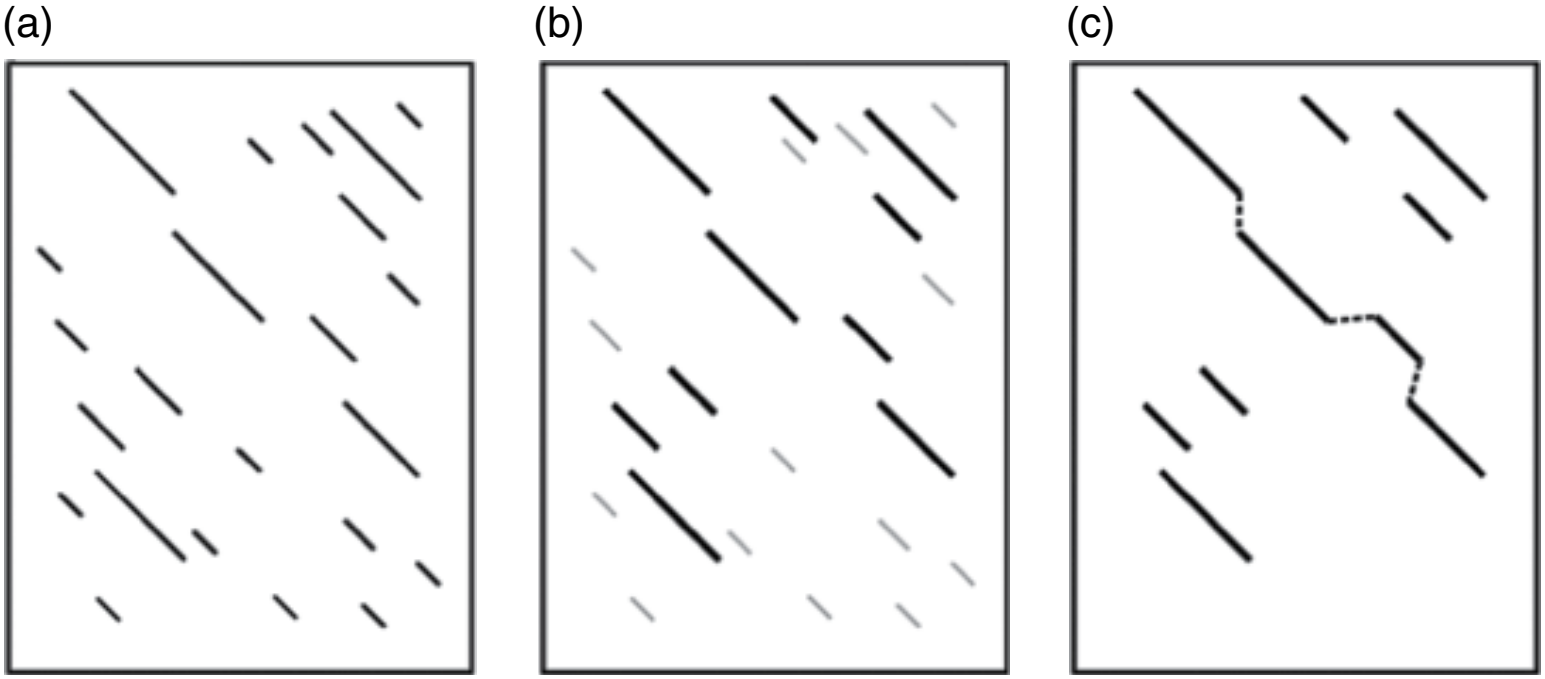


Figure 3.5 Steps taken by the FASTA algorithm when searching a database. (a) Words common to the query and target sequence are located. FASTA connects words close to one another and these are represented by diagonal lines. (b) The top 10 diagonals are selected for further analysis. (c) Diagonals are aligned optimally using a local alignment strategy.

3.3.1.2 BLAST

The **Basic local Alignment Search Tool (BLAST**) is the most commonly used method for locating homologs in a sequence database. BLAST is both sensitive and efficient at locating regions of sequence similarity between nucleotide or protein sequences.

The BLAST algorithm begins by “seeding” the search with a small subset of letters (query word) from the query sequence (Figure 3.4).

The query word as well as related words (where conservative substitutions have been introduced) are located. All words are scored by a scoring matrix and this yields the “neighborhood” (Figure 3.4).

BLAST uses a neighborhood threshold (T) to determine which words are closely related to the original query word. increasing the value of T implies that only closely related sequences are considered, while decreasing it allows for distantly related sequences to be considered.

The original query word is aligned to a word above the neighborhood threshold (Figure 3.4).

The BLAST algorithm then proceeds to extend the alignment in both directions, tracking the alignment score by addition of matches, mismatches, and gaps. The maximal length of the alignment is determined by the number of positions aligned versus the cumulative score of the alignment. The alignment extension continues until the number of mismatches starts to decrease the cumulative score of the alignment; if this decrease is large enough (above a predefined value X, Figure 3.4), the alignment procedure ceases and the resultant alignment is called the high scoring segment (HSP). A score threshold is defined by the algorithm, and if the HSP clears this score the alignment is reported in the BLAST result file.

Finally, the biological significance of an HSP is determined. BLAST uses the e-value to calculate the number of HSPs that would have a score greater than S by chance alone. lower values of e imply greater biological significance; in essence, e can infer whether the HSP is a false positive.

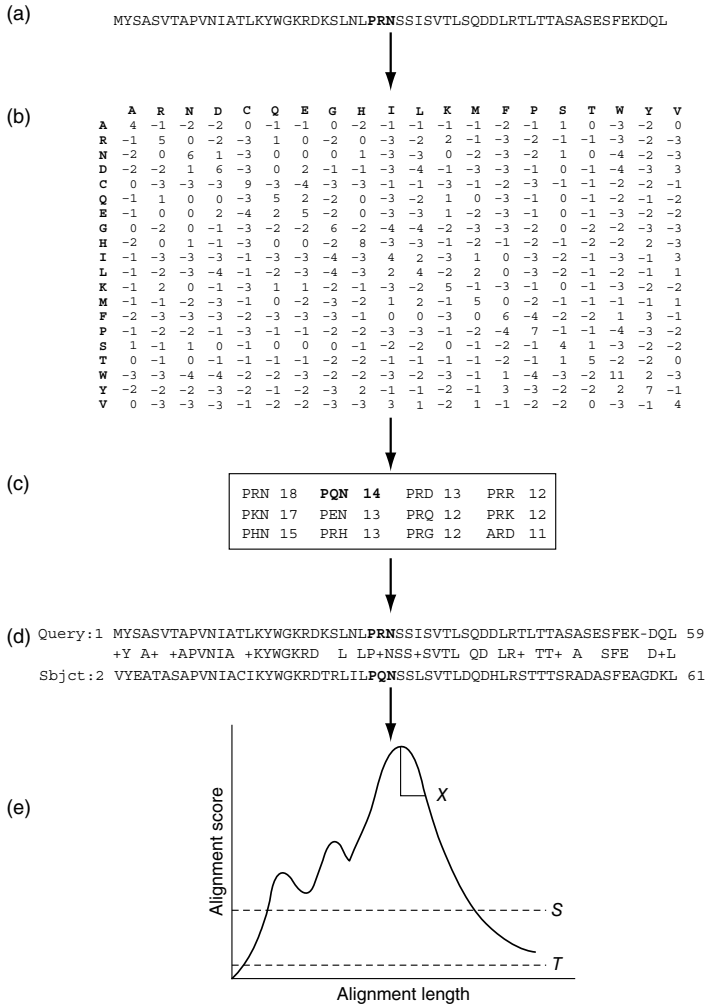


Figure 3.4 Steps taken by the BLAST algorithm when searching a database. (a) The query sequence is compared to a scoring matrix (b), and scores for query words of a given length (three in this case) are calculated (c); query words greater than a certain threshold (T) are used to search the database. (d) The algorithm attempts to extend the alignment either side of the query word that has a hit in the target database. (e) Extension continues until the alignment score falls off more than the allowable significance decay, X.

https://youtu.be/WRKQGwh_Mw0?si=jJG1lQMfjnQdbclB

3.3.2 Multiple Sequence Alignment

Multiple e sequence alignment (MSA) is a method that allows us to infer the interrelationships between DNA or protein families. While pairwise alignments are useful for locating homologs in databases and illustrating conservation between two sequences, they are not as informative as MSA. MSA has the ability to locate conserved residues/domains among thousands of sequences, which can provide insights into important evolutionary and physicochemical processes. MSA is the first step in phylogenetic analysis and is commonly used when designing primers for DNA amplification.

MSA is much more computationally intensive and difficult when compared to the pairwise strategy employed by BLAST and FASTA. One of the most commonly used MSA algorithms is CLUSTAL and it utilizes progressive alignment to efficiently align all sequences of interest. CLUSTAL follows three steps:

1 An initial assessment of how closely related different sequences are to one another by performing pairwise alignments.

2 A guide tree is generated based on the pairwise alignment scores.

3 Sequences are aligned progressively, guided by the phylogenetic tree. closely related sequences are aligned first, and then additional sequences and groups are aligned.

CLUSTAL refines its progressive alignments by implementing a number of alignment penalties. For example, gap insertion and extension penalties exist to reflect that the chances of a gap within a hydrophilic region is more likely, as these are generally loops or random coil regions where gaps are more common. Similarly, residue-specific penalties are enforced so that domains that are rich in glycine are more likely to have an adjacent gap than positions that are rich in valine, for example.

3.3.2 Multiple Sequence Alignment

https://youtu.be/TZaA_-4j19w?si=u40B5jdfuizlYOwe

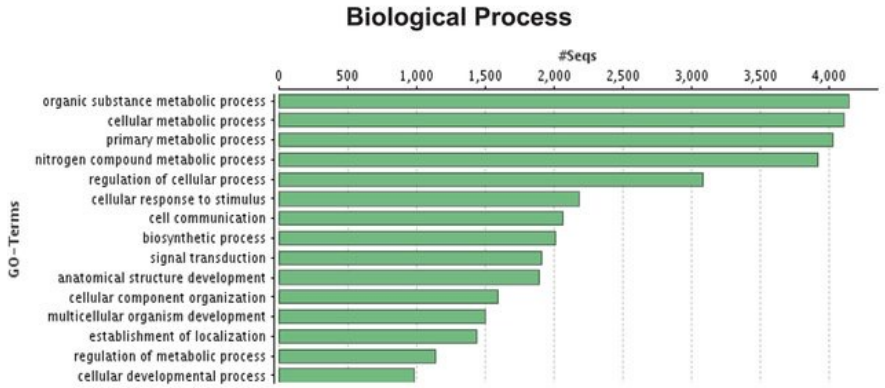
3.3.3 Gene Ontology

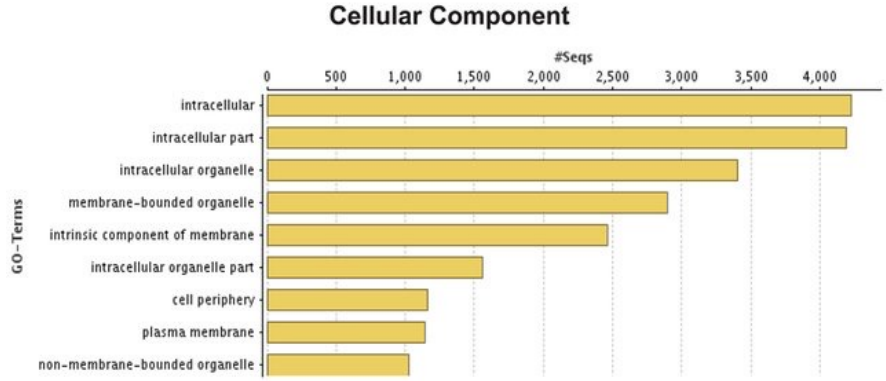
When the first comparison between two complete eukaryotic genomes (yeast and nematode worm) was performed, researchers were surprised to discover that a high proportion of genes displayed orthology between these two dis- tantly related organisms (which diverged ~ 1.6 billion years ago). Orthologs are genes that are derived from a common ancestor and commonly have the same function. Following from this, knowledge of the biological role of an ortholog in one species can be used to illuminate the putative function of the other ortholog. However, organizing biological data from multiple species databases is a major challenge and is made harder when different databases use different terminol- ogies to describe the same process.

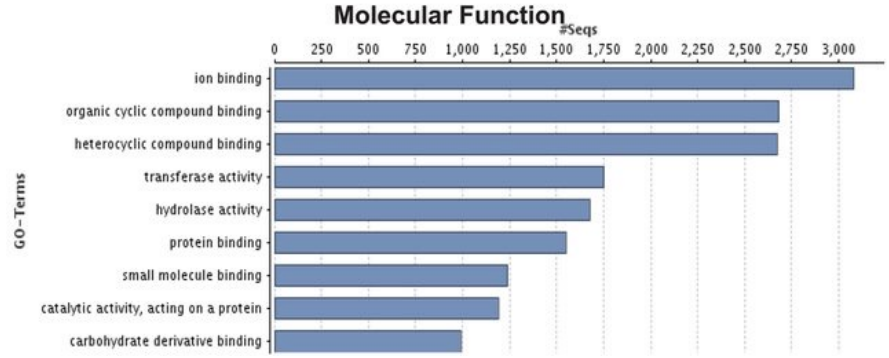
To overcome these difficulties, the Gene Ontology (GO) consortium was set up in 2000 with the goal of producing a structured, precisely defined, common, controlled vocabulary for describing the roles of genes and gene products in any organism. Ontology terms provide a framework for storing and querying different databases using the same search terms. The GO consortium provides detailed annotations for twelve important model organisms (*Arabidopsis thaliana*, *Caenorhabditis elegans*, *Danio rerio*, *Dictyostelium discoideum*, *Drosophila melanogaster*, *Escherichia coli*, *Gallus gallus*, *Mus musculus*, *Rattus norvegicus*, *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe*). collectively, those 12 species are referred to as the GO reference genomes.

The GO consists of over 26,000 terms arranged in three branches:
1 *Cellular Component*: an individual component of a cell, but part of some larger object, such as an anatomical structure (nuclear membrane, for example).
2 *Biological Process*: describes broad biological goals, such as mitosis or purine metabolism.
3 *Molecular Function*: describes the roles carried out by individual gene products; examples include transcription factors and DNA binding.

The annotation of newly sequenced fungi can be greatly accelerated by comparisons to the GO reference genomes. *De novo* genes can be assigned putative functions based on sequence similarity to existing genes in one of the model organisms. The fact that two of the model organisms are fungi (*S. cerevisiae* and *Schiz. pombe*) makes the GO resource highly applicable to genome annotation in newly sequenced fungal genomes.







Genomics and the Fungal Tree of Life

Phylogenetics

The goal of phylogenetics is to arrange a set of populations, species, individuals, or genes into a logical arrangement that infers the evolutionary relationships among them. Evolutionary relationships infer the historical development of species and are usually presented as an evolutionary tree (Figure 3.1). Traditional methods of fungal systematics such as vegetative cell morphology, sexual states, physiological responses to fermentation, and growth tests can assign fungal species to particular genera and families. The fungal fossil record is poor, however, and fungi exhibit few morphological characters; therefore an alternative approach is desirable. Fungal sequence data (RNA, DNA, and protein) have been used successfully to infer evolutionary relationships among species. In many cases, aligned sequences (Section 3.3.2) are processed as a distance matrix. Species that are most closely related will have a small distance, while distantly related species will have a larger distance measure. Phylogenetic algorithms such as UPGMA (Unweighted Pair Group Method with Arithmetic Mean), minimum evolution, and neighbor joining are used to represent distance matrices as phylogenetic trees. The choice of phylogenetic markers for inferring the fungal tree of life is a contentious issue. Ideally, a phylogenetic marker should be ubiquitous throughout the species under consideration, present in single copy, have slowly evolving sites, and be unlikely to undergo horizontal gene transfer. For this reason, a significant majority of accepted relationships between fungal organisms are determined using 18S ribosomal DNA. However, single-gene analyses are dependent on the phylogenetic markers having an evolutionary history that reflects that of the entire organism, an assumption that is frequently violated. Also, individual genes contain a limited number of sites and, in turn, limited resolution. An alternative approach to single gene phylogenies is multigene phylogenies. These attempt to combine all available phylogenetic markers. There are two commonly used methods to do this: concatenated multigene phylogeny reconstruction and supertree analysis.

Concatenated Multigene Phylogenies

Multigene concatenation essentially appends many aligned genes together to give a large super alignment. Combining the data increases their informativeness, helps resolve nodes and basal branching, and improves phylogenetic accuracy. Numerous species phylogenies have been derived by concatenation of universally distributed genes. Recently, the Fungal Tree of Life consortium (Table 3.1) used six housekeeping genes (18S rRNA, 28S rRNA, 5.8S rRNA, elongation factor 1-alpha, and two RNA polymerase ii subunits (RPB1 and RPB2)) from 199 fungal species to reconstruct the evolutionary history of the fungal kingdom. As well as showing the evolutionary history of all fungal phyla, this analysis showed that the loss of spore flagella from early diverging fungi (similar to extant chytrids) coincided with the development of novel spore dispersal mechanisms leading to the diversification of terrestrial fungi.

Supertrees

Supertree methods take all input trees and generate a single representative species phylogeny (Figure 3.7). Individual input trees are derived from single genes. Comparative fungal genomic analyses have shown that less than 1% of all fungal genes are universally distributed. This situation implies that when we reconstruct multigene phylogenies we are ignoring 99% of the genes found in fungi. Ideally, we would use 100% of the gene data. Supertree methods enable us to do this.

Supertree methods generate a phylogeny from a set of input trees that possess fully or partially overlapping sets of taxa (Figure 3.7). Therefore, supertree methods take as input a set of phylogenetic trees and return a phylogenetic tree that represents the input trees. This type of analysis yields a phylogeny that maximizes the number of genes used and therefore is truly representative of the entire genome. A supertree analysis of 103 complete fungal genomes identified 4,753 individual gene families. Individual phylogenies for each gene family were reconstructed and the complete set was summarized by supertree techniques. This analysis showed that within the Saccharomycotina, a monophyletic (single) clade containing C. albicans and close relatives is evident. Species within this clade translate the codon CTG as serine rather than leucine. A second monophyletic clade containing genomes that have undergone a whole-genome duplication (S. cerevisiae and close relatives) is also evident. Supertree techniques are becoming more popular in phylogenetic analysis and will be useful in reconstructing the Fungal Tree of Life as additional fungal genomes become available.

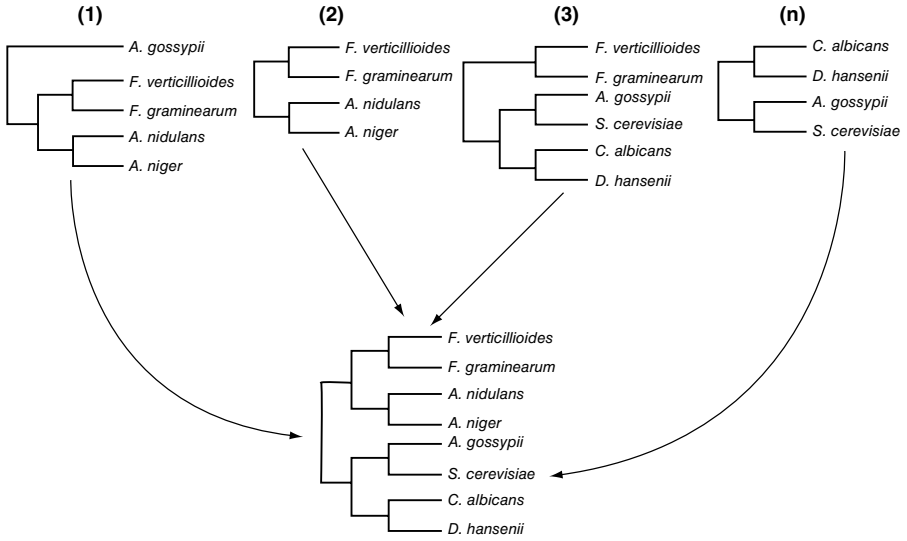


Figure 3.7 Representation of supertree reconstruction. Supertree methods take individual gene trees and express them as a single representative phylogeny. Thousands of trees (expressed as n) can be used as input for supertree techniques.

https://youtu.be/WRKQGwh_Mw0?si=D6T4XjAUBidmhwLY

3.6 Online Fungal Genomic Resources

3.6.1 The Joint Genome Institute Fungi Portal

3.6.2 Saccharomyces, Candida, and Aspergillus Genome Databases

Table 3.1 Useful online resources.

Database	URL address
SGD	www.yeastgenome.org
CGD	www.candidagenome.org
AspGD	www.aspgd.org
CADRE	www.cadre-genomes.org.uk
Aspergillus fumigatus database	www.aspergillusgenome.org
CandidaDB	www.candidagenome.org
NCBI	www.ncbi.nlm.nih.gov
Sanger Institute	www.sanger.ac.uk
EMBL	www.embl.de
DDBJ	www.ddbj.nig.ac.jp
Swiss-Prot	http://expasy.org/sprot/
Fungal Tree of Life	https://aftol.umn.edu/
CGOB	http://cgob.ucd.ie/
Genomes OnLine Database (GOLD)	https://gold.jgi.doe.gov/

Conclusion