Table of Contents

I- Introduction to the course

II- Yeast as a model system

  1- Different types of yeast
  2- Sexual reproduction
  3- Life within a cell wall
  4- Metabolism
  5- Prototrophy/auxotrophy
  6- Sex determination

III- Forwards genetics approach for gene identification in yeast

  1- Point mutations, knockout and conditional mutations
  2- Metabolic screens (auxotrophy mutants)
  3- Morphological screens (cdc mutants)
  4- Physics-based screens (late sec mutants)
  5- Screening with molecular reporters (early sec mutants)
  6- Mutant selection (karyogamy screen)

IV- Epistasis analyses

  1- Suppressor mutations
  2- Synthetic lethality
  3- Non-allelic non-complementation
  4- Allelic complementation
  5- Multicopy suppressors
  6- Pathway ordering

V- Gene cloning

  1- Gene mapping
  2- Cloning by complementation
  3- Cloning of multicopy suppressors
  4- ORF identification
  5- Identity tests

VI- Non-mendelian genetics

  1- Mitochondrial genetics
  2- Prions
  3- Selfish DNA
I- Introduction to the course

Classical vs. modern genetics

Classical genetics, as founded by Gregor Mendel, thought of a gene as a free-floating entity that accounted for a trait. Through the work of Morgan, it has then become clear that these genes are not free floating but located on chromosomes, where a gene corresponds to a chromosomal locus. It has therefore become possible to map these loci relative to each other. Therefore classic genetics is based on the analysis of phenotypes arising from mutations, the phenotypes arising from the combination of mutations, and on mapping these mutations relative to each other on the genome. With progresses in the 1960’s, it has become evident that chromosomes are made of DNA, and that genes are encoded in the DNA by the sequence of the nucleotides. Discoveries in biochemistry and molecular biology led to the insight that most processes in cells are conducted by proteins and that the sequence of the genes encode for the sequence of the proteins. Therefore, in modern genetics a gene is rather thought as a sequence of nucleotides than only as a locus.

Topics of modern genetics

With advancing knowledge and methods genetics is becoming more important in many fields of biology. With modern methods geneticists try to answer the following key questions:

- How many genes are in a genome, how many of them are involved in each essential functions supported by the genome?
- Which genes are essential under experimental conditions?
- How do genes function together in various conditions?
- How does a cell develop its phenotype from its genome?
- How is it possible that different type of cells can be generated out of the same genome?
- How does the genome regulate its own expression? (Epigenetics; Chromosome/Nucleus organization)

At the beginning of modern genetics it was assumed that the majority of genes are essential for the survival of the cell, just as parts of a perfect machine would be. Nowadays it is known that only about 15% of the genes are essential under any specific experimental conditions. This is due to the fact that important functions are covered by overlapping genes and are encoded more than once in the genome. It is also due to the fact that some genes become necessary only under other experimental/growth/survival conditions.

Redundancy ensures that a single defect can’t kill the cell. However, one must be careful to not assume that the genome is structured that way as a manner to prevent individual mutations to be lethal: the genomes that we are working with do not have these mutations (yet), and therefore could not have survived such mutations that have not appeared in their history. Thus, the
redundancy must be there for other reasons.

Three reasons have been hypothesized:

1- it is possible that the genes present in the genome are targeted by drugs in nature or are sensitive to specific environmental conditions, such that there is selection for the emergence of rescue pathway even in absence of mutation.

2- Some processes might function in a highly stochastic manner, such that high fidelity can be achieved only if redundant mechanisms coexist (for example, if the DNA polymerase makes mistakes at a rate of $10^{-6}$, DNA repair recognize mistakes at a frequency of $10^{-3}$, a combination of several mechanisms will be needed to reach a fidelity of less than $10^{-10}$ mistake per cycle.

3- Redundancy might provide other advantages, such as the ability to become “evolvable”, i.e., to simplify the emergence of advantageous mutations. Still there are a few essential genes which all have a broad variety of functions.

Tools of modern genetics

There are two major approaches to genetic problems: Forward genetics and reverse genetics. 

**Forward genetics** starts with an interesting function and then tries to isolate mutants where this function is affected. These mutants will be used in a second step to identify the genes that are affected. Therefore the approach is: function to mutant to gene to protein.

**Reverse genetics** on the other hand first identifies a protein, and then scans for the underlying gene, after that the gene is mutated and the phenotype is determined. Here the approach is: protein to gene to mutant to phenotype. These techniques allow determining easily the positive or negative function of a protein.
II- Yeast as a model system

1- Different types of yeast

Fungi are excellent model systems due to their close relatedness to animalia. Indeed, fungi and animalia are the two kingdoms forming the class of the opistokonts (See Figure 1 Tree of Life). Among fungi, the most used model systems are the unicellular fungi, known as yeasts. There are two main types of yeasts. Representative species of types are used as model systems in research.

The most frequent yeasts are budding yeasts (most of which, but not all, are ascomycetes). All these different organisms form round cells that divide by forming de novo a daughter, which emerge in the form of a bud from the surface of its mother. They represent a large number of species, some of which are only distantly related to each other (at the range of hundreds of millions of years). The species most typically used in research is the bakers yeast, *Saccharomyces cerevisiae*, originally known for its usage in bread, wine and beer making. Related species are also used by researchers, such as:

- *Saccharomyces carlbergensis*.
- *Kluyveromyces lactis*, a related species, is studied in the context of its usage in industry and cheese making.
- *Pichia pastoris*, a more distant species, is used for the expression of recombinant proteins and as a model system to study secretion.
Finally, some budding yeast species are known and studied for their roles as pathogens. These include *Candida albicans*, which infects animals, including humans, and *Ustilago maydis* (a basidiomycete), which infects corn.

The second forms of yeasts are the species generically called as fission yeasts (which represent a sub-branch of the ascomycetes), or *Schizosaccharomyces*. These yeast species form rod-shaped cells that divide in their middle by fission. They are very distantly related to *Saccharomyces cerevisiae* and most budding ascomycetes, from which they have diverged between 300 Mio and 1 Billion years ago, and hence from which their genomic sequence and genome organization is highly divergent. There are only four species of fission yeast known so far, and two, *Schizosaccharomyces pombe* and *Schizosaccharomyces japonicus*, are used as model systems. *Schizosaccharomyces pombe* the more commonly used one of the two. Both species were identified in breweries, where they participate in fermentation.

In these lectures, we will focus our attention to *Saccharomyces cerevisiae*, which is not only the most widely used model system, but also the species of highest biotechnological relevance.

### 2- Why *S. cerevisiae*?

*S. cerevisiae* is an ideal model organism to answer questions in genetics for multiple reasons:

- It grows and divides relatively rapidly on standard media (generation time is about 90 minutes at 30°C on rich medium).
- It has a simple life-cycle, switching between haploid (1n) and diploid (2n) states.
- During the haploid state, recessive mutations cause a detectable change of the phenotype, since the cell has only the mutated allele. Thus, recessive mutations can be observed directly.
- The conjugation process, whereby two haploid cells form one diploid cell, allows combining mutated alleles of a same gene and determining which are dominant and which are recessive.
- Upon meiosis i.e. the formation of 4 haploid cells out of one diploid, mutations can be redistributed - a process called recombination - such as to generate novel genotypes. This allows generating and studying the phenotype of double and multiple mutants.
- Single cells can be separated and incubated to give raise to isogenic strains via asexual reproduction (budding), i.e. colonies of yeast cells with identical genomes, also called clones. Isogenic strains are important because they permit studying the effect of different mutations in the context of the same genetic background.
- Most yeast species are immune to external viruses due to their thick cell wall. Consequently the genome of the cells cannot be modified by horizontal gene transfer. This has for consequence that the genome is fairly stable, and that homologous recombination is particularly efficient.
- Since *S. cerevisiae* cells also undergo mating, a form of sexuality, it is a handy model organism to study meiosis.

### 3- Sexual reproduction

Haploid yeast cells exist in sexual types, called mating types (sexes): MATa and MATα. Depending on how sexuality is regulated, the cells can be homo- or heterothallic. Heterothallic strains can stably exist in either one of the mating types, i.e., if a colony is grown from a single
haploid MATa cell, all cells in the colony will be MATa (haploid). Similarly, it is also possible to
grow colonies of MATα (haploid) or MATα/MATα (diploid) cells. Homothalic strains exist only as
MATa/MATα. This is due to the fact that the haploid homothallic cells change their mating type (!) during mitosis. Therefore a culture starting i.e. from a MATa cell will rapidly contain both MATa and
MATα cells, which will therefore soon mate with each other and form MATa/MATα diploid cells.
The cells of homothallic strains are therefore rarely observed in the haploid state. The reason why
cells have developed this mode of growth is that the diploid state has the advantage that each cell
always possesses two copies of each gene, even in the G1 phase of the cell cycle. This allows the
repair of damaged genes (for example due to a double strand break) using the homologous copy
as a template. It also reduces the noise of transcription: If a gene exists as a single copy, then
stochastic start of transcription will cause the amount of the encoded protein to vary over time. If
there are two copies of the same gene in the cell, it is unlikely that stochastic fluctuations are
synchronous on both copies. Therefore, the amount of the protein produced by these genes is
more stable.

4- Life within a cell wall

The cell wall of yeast cells provides mechanical stability and protection. However, it makes
migration of the cell more difficult. For this reason many species of fungi are unicellular. Single
cells can easily be moved by a stream of water or a gust of wind. Another possibility is to migrate
via directed cell division, where newly formed cells always grow in the direction of i.e. the highest
nutrient concentration. Thus many yeast populations form large networks of chains (the cells
remain attached to each other), whereby the population progresses forward from generation to
generation.

5- Metabolism

Saccharomyces cerevisiae can perform fermentation without respiration, by suppressing the
respiratory activity of mitochondria as long as there is glucose in the medium. This enables the
study of anaerobic as well as aerobic metabolic pathways. It also allows the study of the otherwise
lethal mutations of the mitochondria. The respiratory function of mitochondria is not essential in S.

![Figure 2: Fermentation vs. Respiration](image)
Fermentation: glucose -> ethanol + carbon dioxide + 2 ATP
Respiration: glucose -> carbon dioxide + 36 ATP
Fermentation has a lower output of energy molecules (ATP), however it enables a faster consumption of
the valuable carbon source (glucose).
cerevisiae, hence the mutations $\gamma^0 = \text{no mitochondrial DNA}; \gamma^- = \text{mutated mitochondrial DNA}; \gamma^+ = \text{mitochondrial DNA can all survive. Glucose is an optimal carbon source and yeast compete with bacteria over glucose. In terms of ethanol, yeast are resistant up to 20% EtOH, but ethanol can kill bacteria.}

Growth: yeast grows in 2 phases:
   i. log phase: in this phase cells grow exponentially and use up all the energy source. If two carbon sources are present in the medium such as glucose and ethanol then they first use glucose and then after a diauxic shift continue using ethanol for respiration.
   ii. stationary phase: after all the ethanol (or the carbon sources) are used up, the cell enters a state of quiescence called the stationary phase. The $\rho^0$ and $\rho^-$ cells, which cannot respire, do not do the diauxic shift and go directly to the stationary phase when glucose is used up.

Diauxic shift:

Microorganisms grown in the presence of 2 different carbohydrates in the nutrient solution (in other words in the log phase) show two phases of growth. In the first phase the carbohydrate, which induces a quicker cell growth or proliferation, is metabolised. After the first carbohydrate source has been exhausted the cells start metabolising the second carbohydrate, hence the term diauxic shift. During the diauxic shift a lag phase can be observed due to the fact that the cells need to produce new enzymes to metabolise the second carbohydrate. For yeast, this second carbon source is generally that ethanol generated during fermentation of the first carbon source.

6- Prototrophy/auxotrophy

A prototroph micro-organism is an organism able to synthesize all essential compounds out of inorganic substances, and thus can grow on minimal media. All yeast are prototroph with the exception that they need a carbon source of organic nature, which also serves as source of energy. They also need a few vitamins that they cannot synthesize themselves. For all other organic compounds needed for anabolism, such as lipids, amino acids and nucleotides, they can synthesize them themselves. However, they are also able to pick up most of these compounds from their environment if available, such as to save on synthesis.

Some mutants have lost one or several enzymes required for the synthesis of specific compounds and therefore require that this compound is being provided from outside. Such a mutant is called auxotroph for the compound in question.
7- **Sex determination**

The sex genes in Yeast are located on chromosome III, which is a rather small chromosome (300 kb):

The MAT region consists of two open reading frames (arrows in the scheme).

- a2 is a pseudogene without a function
- a1 ~ α1, they only differ in the N-terminus
- α2 homeobox-type-protein

**How do the transcription factors (TF) regulate the sex determination?**

<table>
<thead>
<tr>
<th>abbreviation</th>
<th>name (codes for, for example)</th>
<th>MATα</th>
<th>MATα/MTα</th>
<th>MATα/MTα</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASG</td>
<td>a specific genes</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>αSG</td>
<td>alpha specific genes</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>HSG</td>
<td>haploid specific genes</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>DSG</td>
<td>diploid specific genes</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

**Table 1:** Overview over the abbreviations of the yeast – sex – determining genes and their functions together with the expression patterns of the sex determining genes in the different sexes of yeast: + indicates expression of certain genes and - that the genes are not expressed, i.e. in MATα, the ASG and HSG are expressed and αSG and DSG are repressed.

**How does this work in more detail?**

MATα is the default state of a yeast, and in MATα types genes ASG and HSG are always expressed, α-factor and α-factor-receptor leading to the sex a. The MATα cells expresses the MAT genes α1 and α2, which both encode transcription factors. The α1 protein activates the expression of the αSG, and particularly the genes coding for the α-factor-receptor and α-factor. The α2 protein represses the ASG. Thus the cells containing the MATα version of the MAT locus express the
mating identity $\alpha$. In MATa/MAT$\alpha$, the $a1$ protein becomes functional and the $\alpha2$ protein has two functions:

i. alone, it represses ASG

ii. with $a1$, $\alpha2$ forms the $a1\alpha2$-dimer, which is a HSG-repressor (since HSG represses DSG, DSG is now expressed) and also represses $a1$ (thus $\alpha$SG are repressed)

That means that $\alpha2$ can extend its function in MATa/MAT$\alpha$. As a default state, the ASG have a strong promoter which is regulated by repression, while the $\alpha$SG have a lousy promoter which is regulated by activation. The mating types are defined using two tricks: First, MATa is the default mating type and MAT$\alpha$ represses the MATa related genes. Second, for the diploid type the both genes are needed (like complementation).

Hidden Mating-Type Loci

Homothallic haploid yeast mother cells switch mating type after every cell division. This ensures that haploid cells of opposite mating types are next to each other and can mate, resulting in the preferred diploid form. This is the reason why these strains are called homothallic: They never stay haploid MATa or MAT$\alpha$, but rapidly convert into diploids, whether they come from a single MATa or a single MAT$\alpha$ cell. Thus, they form only colonies of diploid cells. Heterothallic cells are unable to switch mating type. Therefore, three types of colonies exist: colonies of haploid MATa cells, colonies of haploid MAT$\alpha$ cells, and colonies of diploid cells.

To switch their mating type, the mother cell’s genome has to contain the genes for both mating types, $a$ and $\alpha$. The alleles MATa and MAT$\alpha$ are present as silenced copies at the loci HML ($\alpha$) and HMR ($\alpha$). During G1, the HO endonuclease is expressed in the cells that were a mother already in the cycle before. This endonuclease cleaves the genome on the MAT cassette and the only way to repair the locus is to use the homologue copies in HML$\alpha$ and HMR$\alpha$. In a MATa cell, if HMR$\alpha$ locus is used for repair the mating type switches from MATa to MAT$\alpha$. If HML$\alpha$ locus is used, the mating type still remains MATa. In a MAT$\alpha$ cell, if HML$\alpha$ is used the cell switches to MATa, if HMR$\alpha$ is used it remains as MAT$\alpha$. In 70-80% of the cases the repair mechanism will lead to a mating type change: Yeasts don’t want to stay haploid for a long time, because if there are double strand breaks, they can’t repair them. But if they are diploid they can even repair it in G1. They can use homologues chromosomes.

The fact that only the mother cell switches its mating type is driven by Ash1 (Asymmetric expression of HO). Ash1 is transcriptional repressor that prevents expression of the HO endonuclease in the nucleus inherited by the daughter (the previous bud). Ash1 represses HO only in the bud, because its mRNA is generated during the late stages of budding (during mitosis), and transported along actin into the bud. Therefore, the protein is only formed in the bud and represses the HO gene only in the daughter cell. Therefore the
daughter cell keeps the original mating type of its mother. In the mother cell HO is not repressed and initializes the process by cutting the DNA.

HML/R are repressed by Sir1,2,3,4 (Silencing Receptor). Sir2 is a histone deacetylase, i.e. removes acetylgroups from a histone, and promotes thereby the condensation of chromatin at the HM loci, making them inaccessible for transcription.

✓ Why are sir- mutant cells resistant to pheromone? The Sir proteins function to repress both HMLα (Hiden Mating type locus Left) and HMRα (Hiden Mating type locus Right). Thus a sir- mutant cell behaves as a pseudo-diploid. Thus sir- mutants express both HMLα and HMRα, and are genetically similar to the MATα/MATα cells and repress the hsg, asg and the asg genes. For this reason, they are resistant to pheromone.

✓ What is the difference between haploid cells that have a mutation inactivating the MAP kinase pathway involved in pheromone signaling compared to cells lacking the SIR genes?

In MAP kinase pathway mutants, the cells still express the hsg genes, which repress the dsg genes, and hence they cannot undergo meiosis.

In sir- mutants, the cells behave as diploids. Under starvation they will start meiosis and manage to do two viable spores.
III- Forwards genetics approach for gene identification in yeast

1- Point mutations, knockout and conditional mutations

- temperature sensitive (ts) mutants: missense mutations leading to protein folding problems

  - heat-sensitive: wt at permissive temperature (often room temp.) and mutant at higher restrictive temperature (~36°C). Missense mutation in the core of the protein leads to a small change in the folding at higher temperature. The protein is degraded before the chaperones are able to fold the protein correctly.

  - cold-sensitive: cells grow at high and die at cold temperature

- regulatory promoters: i.e. glucose that represses promoter

In general it is easy to create different types of conditional mutations if you know the genome, if not then mostly temperature-sensitive mutations are used.

How are mutations in essential genes identified?

→ screen for temperature sensitive genes

Because the temperature sensitive mutations cause misfolding of the mutant protein and its degradation, some additional mutations that affect the pathway of degradation of misfolded protein can cause a reversion of the phenotype. This is true when the mutated protein can maintain enough function despite being misfolded.

2- Metabolic screens (auxotrophy mutants)

Example: Trp biogenesis

TRP1 yeast cells (wild type) can synthesize their own tryptophan by themselves. The trp1 mutant cells grows only on media, which contain tryptophan (TRP1 is centromere linked):

remove Trp of medium, replace it by the last Trp-precursor molecule

if they still don’t grow, the gene required for the transformation of last precursor to Trp is missing

mapping of metabolic pathways: can find out which genes are involved in which steps of the i.e. Trp biogenesis

Example: Adenine biosynthesis

- ADE genes \(\rightarrow\) essential to synthesize adenine out of inorganic substances
- ade \(\rightarrow\) need an adenine source and some of them produce a red coloured metabolite
- ade2 \(\rightarrow\) red (only after diauxic shift, only in \(\gamma\) cells)
- ade2-103 \(\rightarrow\) transfers an early codon into an ochre stop codon (\(\rightarrow\) no functional Trp1 protein synthesized, the cells are red when they reach stationary phase)

ade2 mutants growing on EtOH plates become red rapidly
If they turn white, there’s something that can suppress the ochre mutation, or they have lost the mitochondrial genome → useful to identify a variety of stop tRNA mutations, as well as respiratory mutations.

Auxotrophy genes are **conditionally essential** i.e. mutations in LYS gene → lost ability to produce lysine out of inorganic medium. Therefore needs lysine in medium to survive. The mutation has a phenotype only on media lacking lysine. The same is true for ade and trp mutations that are conditionally deleterious, i.e., show a growth phenotype only on media lacking adenine or tryptophan, respectively.

### 3- Morphological screens (example: the cdc mutants)

The first such screen was carried out in yeast by Lee Hartwell (who got the Nobel Prize for it in 2001) to identify cell division genes (CDC = cell division control).

The cdc mutants stop cell division and enter cell division arrest, but they keep growing and increase their volume. Depending on where the cells are arrested in their cycle (depending on which function is affected), the cdcx mutant cells arrest uniformly at the corresponding arrest point in the cell cycle, all showing the same phenotype (for example as large, round, non-budded cells when arrested in G1).

Cdc null mutations are lethal in all mating types, haploid and diploids (if homozygous). Therefore, only conditional alleles can be isolated, and analyzed.

Because they wanted to identify conditional mutations in CDC genes, the first step that Hartwell and his team took was to generate random conditional mutations. They chose to go for temperature sensitive mutations. A library of about 2000 independent clones each containing one distinct temperature sensitive (ts) mutation in one unidentified gene was generated. This collection was subsequently screened, visually, for mutations that affected cell division.

In order to generate random mutations, there are in principle two options:

**Chemically mutagenesis**

This approach generally lead to the generation of point mutations and is an excellent strategy to create temperature sensitive mutations. Ethylmethylsulfonate (EMS) is a methylaing agent and is often used as chemical mutagene. For example, the methylation of DNA bases leads to a misincorporation during DNA replication. The amount of EMS determines the frequency of point mutations incorporated into the genome. A yeast genome contains about 6000 genes where from 15 % are essential. If you have a lethality frequency of about 15 % after mutagenesis, you can assume that you have inactivated about one gene per cell in average, because in 15% of the cells you have inactivated an essential gene.

**Physically mutagenesis**

Irradiation (X-ray) inserts doublestrand breaks and therefore creates disruptions, which are much bigger mutations.

Once the mutagenesis has been carried out, the cells are spread on plates and grown to individual colonies. Thereby, each colony represents a pure clone where all cells contain the mutation originally incorporated during mutagenesis.

**Identification of temperature-sensitive mutation**
We are looking for mutations that do not affect much the growth of the cells at 24°C (the permissive temperature), but become lethal, or at least stop proliferation of the cells at 37°C (the restrictive temperature). The general problem is that you have to kill the cells to identify temperature-sensitive mutations. How can you handle it?

Solution of Gutenberg

1. Petri dish with your colonies
2. Attach velvet on the top of a stamp
3. Put the petridish with the colonies upside down onto the velvet, some cells attach to the little pili of the velvet
4. Transfer them by putting the stamp onto new Petri dishes, so you get an exact copy of your original petridish
5. The stamp/ replica plates can now be grown at different temperature
6. Compare the Petri dishes grown at permissive (24°C) and at restrictive (37°C) temperature. Colonies that died at restrictive and grew at permissive temperature are temperature-sensitive. You can now isolate the ts strain.

Screening for cdc mutations

An aliquot for each of the ca 2000 strains of the ts collection was inoculated into as many individual tubes filled with growth medium and these cultures were first grown at permissive temperature until they had reached log phase. Subsequently, the tubes were placed at 37°C for 4 more hours. Samples were taken, fixed and stained to visualize the nuclei and amount of DNA. The cdc mutants were those that caused all the cells in the sample to arrest at the same point in the cell cycle. For example, cdc28-1 cells had no bud and a single nucleus with 1N DNA content (failed to enter the cell cycle). The cdc24 and cdc42 mutant cells formed large, un-budded cells with several nuclei (failed to bud but entered mitosis). The cdc3, cdc10, cdc11 and cdc12 mutant cells formed chains of cells that failed to separate from each other (cytokinesis mutants). DNA replication mutants arrested with a large bud, a single nucleus with 1N DNA content. The cells that were unable to enter mitosis arrested with the same phenotype, with the only difference that the nuclei contained 2N DNA, etc…

4- Physics-based screens (late sec mutants)

identify genes that are involved in membrane biosynthesis: late secretory protein (Sec)

- late secretory pathway: translocation from ER to Golgi, from Golgi to plasma membrane or to vacuole
- mutants are not able to secrete proteins and cell wall can’t assemble → accumulation of proteins in the cytoplasm → cells get denser compared with normal cells

Identification of late secretory protein (Sec)

Procedure: Temperature sensitive mutants were grown at high temperature → separated by density gradient centrifugation (in high density sucrose solution) → mutants which have a higher density are located in pellet → mutants were allowed to restart growth at permissive temperature → mutants identified affected the late secretory pathway

5- Screening with molecular reporters (example: early sec mutants)
The goal here is to gain information about the functionality of a particular protein via another one.

The mutations identified cover the late steps in secretion such as translocation from ER to Golgi and from Golgi to plasma membrane or to the vacuole. Early steps such as translocation of proteins into the ER and ER homeostasis are not affected by these mutations.

To identify the mutations in the early steps of the secretory pathway, another screen was done using \textit{SUC2-HIS4} fusion gene. His4 is an enzyme functioning in the conversion of Histidinol to Histidine in the cytoplasm. Suc2 functions in digestion of saccharose to glucose and is expressed on ER surface and cotranslationally imported into the ER. When SUC2 is fused to HIS4, His4 is imported into ER and cannot have access to Histidinol which is in the cytoplasm. Thus, cells that have only the Suc2-His4 fusion protein as source of His4 activity cannot synthesize Histidine and die. But the early sec mutants show defects in translocating the protein into ER. Therefore, some of the Suc2-His4 fusion protein is left in the cytoplasm, where His4 converts Histidinol to Histidine. These cells become viable on \textit{–HIS} medium at semi-restrictive temperature. The mutations cause defects in the maintenance of ER and are very toxic. If cells are grown at restrictive temperature, ER contents leak into cytoplasm and the cells die. Still, this approach allowed identifying quite a few mutations and all genes involved in the formation of the translocon, for example.

\textbf{6- Mutant selection (example: karyogamy screen)}

Karyogamy= “fusion of the nuclei”

During the mating process of \textit{S. cerevisiae}, mating types \textit{a} and \textit{α} cells attract each other by releasing \textit{a}- and \textit{α}-factor, respectively. The result is the stop of cell division in G1 (in order for both cells to have the same DNA content), polarization and the formation of a so-called shmoo, which brings the two cells close to each other (no migration possible!). Then, the cytoplasm fuse and in a second step the nuclei as well, which is referred to as karyogamy. The first bud forms right in the middle of the connecting bridge, between the two original mother cells, and the first division transforms this bud into the first diploid daughter. I.e., this daughter contains the genomic information from both parents.

In karyogamy-mutant cells, only the cytoplasms fuse, while the nuclei stay separate. Daughter cells bud only from one mother cell and inherit the genome only of that mother cell.

In order to identify genes involved in karyogamy, we need to be able to detect karyogamy-mutants, and therefore daughter cells that have the cytoplasm from both mother cells, yet the genetic information only from one of them. Ideally, after mutagenization, one could use selection criteria where only daughters of a karyogamy-defective mating survive.

Since mitochondria have their own genome, and are inherited by daughters independently of the nucleus, they are a suitable marker for the cytoplasm. \(\rho^0\) mutant cells have no mitochondrial DNA (but they still do have mitochondrial!) and therefore are not able to do respiration. Such cells stop dividing (but do not die – they go into stationary phase) when plated on ethanol or glycerol plates.

On the other hand, to follow the status of the nucleus, we use \textit{CAN1}, which encodes for an amino acid transporter for arginine. This transporter also imports into the cell a toxic compound, canavanine, which is an arginine-analogue and can be integrated into proteins instead of arginine. Cells that lack this transporter, through a recessive mutation, \textit{can1Δ} are therefore resistant to canavanine, provided that they are \textit{ARG}⁺ and can synthesize arginine by themselves.
We now combine the two mutations ($\rho^o$ and $can1\Delta$) on one mating type, let's say MATa. And let the other mating type (now: $\alpha$) to have both functional, wt genotypes. We now mutagenize only MATa, because that's the nucleus we can select for (MATa nuclei eventually ‘die’ with their cells). After combining the two strains, we will then be able to select the karyogamy-mutants on canavanine-containing plates with glycerol as sole carbon source. Since the daughters of karyogamy defective crosses contain a cytoplasm mixing that of both parents, they inherit intact mitochondria from the MATa parent, can respire and use glycerol as sole carbon source. Because all of these daughters are still haploid, half of them, those that are containing the MATa nucleus, are $can1^-$ and resistant to canavanine (therefore only ‘nuclear’ progenitors of MATa would survive in our example). Since these cells are those containing the mutageneized nucleus, they are most likely to contain the mutation that disrupted karyogamy.

In this assay, upon mating wild-type cells would die on glycerol+canavanine medium, because the fused nucleus would be heterozygote for $CAN1$, and therefore the daughters are not resistant to canavanine, because they express the arginine transporter ($can1^-$ is obviously a recessive mutation). At the same time, the parents of the cross also do not survive on this medium: the MATa parent because it is $\rho^o$ and cannot grow with glycerol as sole carbon source, and the MATa parent because it is $CAN1$, and is killed by canavanine.

Cycloheximid can be used as an alternative to canavanine. It inhibits protein synthesis by binding to ribosomes and freezing them to the mRNA, resulting in arresting the whole machinery. Cells that are resistant have a mutation in one of the ribosome subunits ($cyh1$) that prevents cycloheximid binding. The crucial point is that, as in the previous example, the heterozygotes are sensitive to the drug and die in its presence.

**IV- Epistasis analyses**

Epistasis

Definition: the interaction between two or more genes that control a single phenotype

- i.e. expression of a gene at one locus can mask or suppress the expression of a second gene at another locus
- $ade2$ $ade1 \rightarrow$ red $\rightarrow$ $ade2$ upstream of $ade1$; $ade2$ is epistasic to $ade1$
Concepts of Modern Genetics, Part Y. Barral

16

- \( ade2 \ ade3 \rightarrow \text{white} \rightarrow \text{ade3 upstream of ade2; ade3 is epistatic to ade2} \)

1- Complementation analysis

**Classical**: cross of \( TRP1 \times trp1-1 \rightarrow [TRP1^+] \) cells, so the WT is dominant and \( TRP1 \) complement the \( trp1-1 \) mutation.

2- Suppressor mutations

Example as above:

- \( ade2 \rightarrow \text{red} \)
- \( ade3 \rightarrow \text{white} \)
- \( ade2 \ ade3 \rightarrow \text{white} \)

Thus, the \( ade3 \) mutation suppresses the red colour phenotype of the \( ade2 \) mutant cells. Note that this suppression is phenotype specific. Indeed, the \( ade3 \) mutation does not suppress the adenine auxotrophy phenotype of the \( ade2 \) mutant cells.

**Secondary mutations** restoring growth at 37°C of the sec61-1 mutant cells were isolated by mutagenesis and selection. In this case, one can expect to identify revertants, i.e., mutations in the sec61-1 locus that restore the function of the protein. None of those were identified in this case, but instead, real secondary suppressor mutations, i.e., mutations in other genes, such that the double mutants no-longer showed the growth phenotype of the sec61-1 cells at 37°C, without correcting the Sec61-1 protein itself. These mutations all affected genes of a single pathway now known as the **ER-associated degradation pathway (ERAD)**. This pathway is involved in the degradation of misfolded proteins in the ER.

Explanation: If you prevent degradation of a misfolded protein, like Sec61-1, it is let in place. In this case, this helps the cells survive, because all Sec61-1 proteins are misfolded at 37°C, and ERAD degrades it all. Since secretion, and hence Sec61, is essential for survival, degradation of the entire pool of Sec61-1 is actually what kills the cell. In fact, if not degraded, Sec61-1 is still able to function sufficiently at 37°C, such as to allow enough translocation into the ER and the cell to survive. Screening for secondary suppressor mutations for sec61-1 allowed the identification of several components of the ERAD, such as UBC7 (an ubiquitin-conjugating enzyme).

It is a typical example of **epistasic interaction**, where the effects of one gene are modified by one or several other genes.

3- Synthetic lethality

The opposite of suppression is synthetic lethality. This situation corresponds to the case where two mutations have little effect on their own, at least at the permissive temperature, but have a strong effect when combined, such that the combination is lethal, at least in some conditions.

The concept of synthetic lethality can be extended in the concept of negative genetic interaction, where the phenotype considered is not necessarily growth, but any phenotype, and where each mutation alone have little effect on that phenotype, but the combination of both leads to a strong phenotype.

Because the read out is life or death, synthetic lethality is the easiest negative genetic interaction to score and has therefore been extensively used in the last decade. One need however to distinguish between two types of cases: synthetic lethality of two null mutations and synthetic...
lethality between two point mutations (such as for example between two temperature sensitive alleles).

**Synthetic lethality of two null mutations**

Three types of situations can lead to synthetic lethality.

1. **Inactivation of the two copies of a duplicated gene**

   In many genomes essential and highly expressed genes happen to be duplicated. This is also the case in yeast. A clear example is the case of the histone genes, which are all duplicated in yeast. As a consequence, there are for example two identical genes coding for histone H3. Mutating any of the two does not lead to cell death; the cells are still growing nearly as fast as wild type cells (but not completely, otherwise there would be no selection pressure for keeping the two genes – note that a few percent difference are enough to give a strong selective advantage to the fittest strain). However, inactivation of both H3 genes is obviously lethal.

2. **Inactivation of two parallel mechanisms for achieving the same essential cellular function**

   In most cases, however, synthetic lethality is not due to mutations in duplicated genes. The large majority of the cases corresponds to mutations in genes that have overlapping roles in the cell. For example when two pathways can both lead to the same outcome, mutation of any gene in one pathway does not kill the cells because the other pathway still keep the cells running. However, mutation of both pathways kills the cells.

   For example, cells need arginine for protein synthesis. Arginine can be made available either through synthesis inside the cell, or through import from the environment (if there is any arginine in the medium). Accordingly, the cells express both a set of enzymes catalyzing the biosynthesis of arginine (the ARG genes), and the CAN1 gene, which encodes an arginine transporter. On arginine containing medium, arg- mutations are not lethal. Likewise, in ARG+ strains the can1-mutations have little phenotype. However, combination of the can1- mutation with any mutation in the ARG pathway is obviously lethal.

3. **Inactivation of quality control mechanisms in the context of a first mutation affecting the fidelity of a process**

   Many cellular processes are monitored by quality control pathways that ensure that the cell detects defects and rapidly reacts to them. For example DNA damage rapidly leads to activation of the so-called DNA damage checkpoint, which in turn orchestrates an ordered response involving repair pathways. The gene RAD9 is involved in checkpoint response. As long as the cells are not stressed, the RAD9 gene is not essential. Likewise, many genes involved in the precise regulation of DNA replication and chromatin organization are not essential as long as the repair pathways are present. However, combining a rad9- mutation with such DNA and chromatin organization pathways becomes immediately lethal, because the cell does not repair the damage anymore.

Synthetic lethal interaction studies have been carried out in a genome-wide basis, testing every pair of null mutations possible, using a collection containing each strain where one of the about 5000 non-essential gene in yeast is disrupted. These large-scale studies have also allowed classifying (also called clustering) genes in groups of genes that show the same set of genetic interactions. In general, these genes are then predicted to work together in a common pathway or even in a common complex. Importantly, the genes in such a cluster generally show no genetic interaction with each other. Indeed, if knocking out any of the genes in a pathway inactivates the
pathway, then knocking out a second gene in the same pathway will not have any additional effect. If the pathway is inactivated, it is not possible to inactivate it further.

Therefore, if two non-essential genes are involved in the same process and the double mutant combining the two null mutations is not sicker than the single mutants, we would conclude that the two gene products act together in the same pathway, or in other words, depend on each other for function. If the double mutant is sicker than the single mutants, we must conclude that the two genes function in parallel, separate pathways, or in other words, are at least partially independent of each other for function.

*Synthetic lethal interactions between two point mutations in two distinct genes*

In this case the argument above does not work anymore. Indeed, the mutations in consideration no-longer inactivate necessarily the pathway under consideration. For example, a ts mutation does not inactivate the function of the gene at permissive temperature, but it may simply make the protein less efficient at permissive temperature. A second mutation might kill the cell now because it leads to the full inactivation of the ts protein. Accordingly, combining ts mutations in genes coding for two different subunits of the same complex or two components of the same essential pathway frequently lead to a synthetic lethal phenotype. For example, the products of the genes cdc3, cdc10, cdc11 and cdc12 are so-called septin proteins that are all homologous to each other and assemble into a linear octamer in the following arrangement: cdc11-cdc10-cdc12-cdc3-cdc3-cdc12-cdc10-cdc11. This octamer can assemble tail to tail into filaments that are involved in the control of cytokinesis in fungi and animalia. Temperature sensitive mutations have been identified in each of these genes. The combination of any pair of these mutations is lethal, with only very few exceptions. In all cases, the double mutants are much sicker than the single mutants.

Thus, in the opposite to when combining null mutations, when combining reduction of function, point mutations the observation of a synthetic phenotype generally indicate that the product of the two genes function together, or depend on each other for full function.

### 4- Non-allelic non-complementation

Non-allelic non-complementation describes the case where two mutants that have a similar phenotype are crossed and the diploid behaves as a mutant, despite these two mutations not being in the same gene. The fact that the heterozygous diploid is sick indicates that there is no complementation and in general would tell us that the two mutations are in the same gene. But linkage analysis (see below section on gene mapping) might still indicate that the mutations are unlinked, and therefore not in the same gene. This situation is rare but when observed it is indicative that two subunits of a same complex are mutated with point mutations such as to express inactive form of each protein. If the complex assembled with these proteins is active only if it contains the wild type copies of both proteins, the diploid might be very sick.

For example, let’s take the two hypothetical genes A and B coding for the proteins Ap and Bp. The $a^*$ and $b^*$ alleles of A and B respectively code for inactive proteins $a^*p$ and $b^*p$. Let’s hypothesize that Ap and Bp are functional only when in a Ap-Bp dimer. The diploid $A/a^* B/b^*$ expresses the proteins Ap, $a^*p$, Bp and $b^*p$ and can therefore form the complexes Ap-Bp, $a^*p$-Bp, Ap-$b^*p$ and $a^*p$-$b^*p$. Only one of these four complexes is active, leading to a strong reduction of function already observed in the heterozygous diploid.

### 5- Allelic complementation
On the opposite, the cross above might lead to a diploid that behaves like wild type, despite of the mutations being both in the same gene, as indicated later by linkage analysis. This situation again is rare, and when observed indicates that the gene might code for a protein with two independent domains that can act independently of each other, such that an allele where the first domain is inactivated but the second functions normally may be complemented by another allele where the first domain functions normally but the second does not.

6- Multicopy suppressors

In some cases, overexpression of a protein, for example by putting the corresponding gene on a multicopy plasmid, can suppress the phenotype of one mutant of interest. The overexpressed protein is called a multicopy or dosage suppressor. The interpretation of this situation is again dependent on the nature of the original mutation.

If the suppressed mutant is a null allele, such as a disruption, we will consider that the dosage suppressor acts either instead or downstream of the missing protein. Indeed if the protein Ap functions by activation of the Bp protein, overexpressing Bp might generate enough function even in the absence of Ap. Alternatively, if Ap and Bp are very similar proteins, Bp might become able to substitute for Ap.

If the suppressed mutation is a point mutation leading to the destabilization of the protein product, the suppressor is likely to be an interactor protein. By mass action, overexpression of the suppressor protein may drive complex formation, and the complex may help the affected protein to stay stable. Alternatively, the suppressor protein might titrate an upstream inhibitor of the mutated protein or a downstream effector. Therefore, dosage suppressors of point mutations can identify more components of a given pathway, but the interpretation of the results require more attention and subsequent tests.

7- Pathway ordering

Once genes functioning in a common pathway are identified, it is generally important to identify which of the genes act first (upstream) and which act later (downstream) relative to each other. If all mutants have the same phenotype, such an analysis is very difficult. However, the use of overexpression studies can help, as outlined above.

The situation is much easier if dominant mutations have been identified, or if some mutations have opposite phenotype. Let’s take the example of mitosis entry. Deleting the kinase gene SWE1 advances the timing of mitosis, while deleting the gene MIH1 delays the entry into mitosis. Thus, Swe1 is an inhibitor of mitosis and MIH1 is an activator. The double mutant swe1- mih1- shows an advanced timing of mitosis, similar to the swe1- single mutant. This means that Mih1 cannot accelerate mitosis when Swe1 is absent, or in other words, that the mih1 mutation does not affect the timing of mitosis when Swe1 is absent. Thus, Mih1 would be predicted to act as a repressor of Swe1. In fact, Mih1 is the phosphatase that removes the phosphates deposited by Swe1 on its substrates, and particularly on the master regulator of mitosis, the cyclin-dependent kinase Cdk1. Thus, indeed, Mih1 inhibits, or here reverts, the function of Swe1.
V- Gene cloning

Once mutants have been isolated, the next important step is to identify the genes that have been mutated. Three approaches can be followed. Historically, the first one was to map the position of the mutations on the genetic map of the chromosomes and to use this information to narrow down the position of the mutation to a single coding gene on the physical map of the same chromosome.

When possible, the second approach was to clone the genes by complementation, using genomic libraries cloned into shuttle vectors. Shuttle vectors are plasmids that can be maintained both in E. coli (for DNA amplification, purification and molecular biology) and in yeast (for complementation and functional studies). Genomic libraries are mixtures of about $10^8$ different plasmids each containing a different fragment of the yeast genome. These fragments are originally isolated to be around 5-10 Kb long in average (an average yeast gene is about 2 Kb). Complementation cloning works best for conditional mutants. In this case, the mutant cells are grown at permissive temperature and subsequently transformed with a DNA sample of the library. The cells take up in average one plasmid during transformation. The cells are subsequently plated on rich medium and the plates are incubated at the restrictive temperature. The cells that have received a plasmid containing the wild type copy of the mutated gene can now resume growth at the restrictive temperature and hence form a colony on the plate grown under this condition. All the other cells die or fail to proliferate, because of the mutation. The colonies formed at restrictive temperature can be collected, the plasmid rescued and recovered in E. coli, and the fragment contained in the plasmid can be subsequently sequenced. This allows the identification of the few genes contained in the complementing fragments. One of them is the gene of interest. If several fragments are recovered, the gene of interest is in the overlapping region between these different fragments.

Obviously, if the mutation is dominant this approach does not work. It is also very difficult to use this approach if there is no growth phenotype associated with the mutation of interest.

The most recent approach consists in sequencing the entire genome of the mutant strain and to compare it with the sequence of a related wild type strain. The difficulty here is that even within a same colony there is already some genetic variability simply due to the fact that replication, even if very faithful, introduces mistakes at a frequency between $10^{-8}$ – $10^{-9}$.

1- Gene mapping

The classical approach to map two genes in a genome is to measure the distance between these genes in centiMorgans (centiMorgans reflect the frequency of recombination between two genes and is hence a mean to estimate their distance on the genome). Therefore the amount of recombinants over the total amount of offspring is calculated.

A progeny is called recombinant if its genotype is different at a specific locus compared to its parent’s genotype. (example: Ab x aB are parental genotypes, therefore AB and ab are recombinant offspring)

If two loci are located on the same chromosome, the frequency of recombination reflects the probability of a crossing over event between these loci. The frequency of recombination is taken as an estimate for the physical distance between these loci: the longer the sequence in between, the more likely is a crossing-over to happen in this sequence. (If two loci are far away from each other, the probability of recombination is maximal, that means 50%.)

If the distance between two genes is long enough, several crossing overs may occur. This can lead to non-recombinant offspring. In the classical approach, these are not taken into account,
(because you're only counting the recombinant ones as an estimate for the crossing overs). Therefore, the amount of crossing overs is underestimated and the genes will be located too close to each other.

The advantage of ascomycetes is, that all spores (offspring) from one diploid cell are located in one ascus (in yeast 4 spores). From the different distribution of recombinant and non-recombinant spores in one ascus, you can conclude to their genesis (crossing over-dependent or not). Three types of tetrads can occur in asci, therefore you only have to count the different tetrads instead all spores.

When we cross AB x ab we obtain three types of possible tetrad:

<table>
<thead>
<tr>
<th># of spore/ type of tetrad</th>
<th>Parental ditype (PD, like one of the parent and 2 types of spores)</th>
<th>Non-Parental ditype (NPD, not like the parents and 2 types of spores)</th>
<th>Tetratype (TT, all spores are different)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spore 1 : ab (non-recombinant)</td>
<td>aB (recombinant)</td>
<td>aB (recombinant)</td>
<td></td>
</tr>
<tr>
<td>Spore 2 : ab (non-recombinant)</td>
<td>aB (recombinant)</td>
<td>ab (non-recombinant)</td>
<td></td>
</tr>
<tr>
<td>Spore 3 : AB (non-recombinant)</td>
<td>Ab (recombinant)</td>
<td>AB (non-recombinant)</td>
<td></td>
</tr>
<tr>
<td>Spore 4 : AB (non-recombinant)</td>
<td>Ab (recombinant)</td>
<td>Ab (recombinant)</td>
<td></td>
</tr>
</tbody>
</table>

If the two genes are not linked (frequency of recombinant = 50%) we have exactly ¼ of each possibility to catch a spore ab, aB, Ab or AB.

Every tetrad hasn’t the same probability to occur, though. It can also vary in function of the frequency of recombinant, i.e. if the two genes are close or if one or both are near to a centromere. Now we will study example of different repartition of probability of each tetrad type depending of frequency of recombinant.

**Example 1: Frequency of recombinant= 50% (genes are completely unlinked)**

To calculate the probability of each tetrad, we draw a tree to explain the proposition.

Assignments: We choose to do this sequence of gene A in the tree: a, a, A and A. This has not effect on probability because we are interested on couple and not the arrangement of couple, in others words: ab/ab/AB/AB and ab/AB/ab/AB are equivalent.

By this tree (see next page) we see the proportion 1/6 PP, 1/6 NPD and 4/6 TT = 1:1:4 when the two genes aren’t linked. In other words this tree represents the random distribution possibilities of two non-linked alleles.
Example 2: on different chromosome, close to centromere

Consider the following case:

Two alleles – i.e. A and B – lie on different chromosomes (i.e. I and II), but each close to the centromere. Two parents with recessive mutations in A and B (a and b) respectively, are crossed. During meiosis, the chromosomes are distributed to the spores such that either the parental or a „new”, non-parental combination of chromosomes is observed (see Fig. 9). Note that crossovers will not affect the location of A and B (and thus the observed tetrad genotypes/phenotypes concerning A and B) because A and B lie close to the centromere. As a consequence of this, no tetratypes are observed in the tetrad analysis. The mixing of homolog parental chromosomes during meiosis allows the formation of parental and non-parental ditypes with the same frequency, i.e. PD:NPD:TT = 1:1:0, PD=NPD.

Example 3: Different chromosome, one close to the centromere

If however, one of the alleles is slightly away from the centromere of its chromosome the observed frequencies of tetrads change, due to crossover. The fact that allele B is now away from the centromere of its chromosome leads to a changed segregation behavior in the case of crossovers;

If no crossover occurs: the same tetrads as in the case of two alleles close to the centromere are obtained (see before).

If one crossover occurs: additionally to PD and NPD, tetratypes are observed: NPD=PD, TT<(PD+NPD), see Fig. 10

The frequency of observed tetratypes can now be used to estimate the distance d between the centromere and the B-locus (T. H. Morgan and A. H. Sturtevant):

\[ d = \frac{1}{2} \frac{TT}{(TT+PD+NPD)} = \frac{1}{2} \frac{TT}{total} < 20\% \text{ if } d>20\% : \text{more than one crossover probable} \]
The factor \( \frac{1}{2} \) in front of the ratio between tetratypes and the sum of all tetrad frequencies is explained by the fact that only one of the chromosomes had a crossover (at least as far as can be determined by this assay). Thus, the analysis and quantification of tetrads can be used for distance measurements.

**Example 4: Genes are on the same chromosome**

Now let’s assume the two loci are on the same chromosome and that we have A & B on one chromosome and a & b on its homologous one.

1. If they are very close to each other, there won’t be a crossover. Therefore we have: PD>>NPD
2. If there is one crossover between them, the spores are always aB, ab, AB & Ab, i.e. the tetrad is TT. Thus: 1 crossover -> always TT.
3. If the genes lie way apart from each other then the ratio for PD:NPD:TT is 1:1:4.
4. If we assume two crossovers, there are four possibilities for the spores:

![Figure 7: Possible scenarios in crossover events and its results on two genes (A and B) that lie far enough from the centromere and closer than 50cM. (Taken from Introduction to Genetic Analysis, Griffiths et al.)](http://www.swan.ac.uk/genetics/big322/Yeast%20Book/7.htm)

![Figure 8: A comparison of recombination of two genes on homologous chromosomes or on different chromosomes. Including no crossover scenarios.](http://www.swan.ac.uk/genetics/big322/Yeast%20Book/7.htm)
The important thing is that the appearance of NPD is specific for a double crossover. So $4 \times \text{NPD}$ gives us the number of double crossovers. TT - $2 \times \text{NPD}$ is equal to the number of single crossovers, since $2 \times \text{NPD}$ is the number of TT due to double crossovers. Combining this we get Perkins's formula (1949, David D. Perkins):

$$cM = \frac{1}{2} \left[ \frac{\text{TT} + 6 \times \text{NPD}}{\text{TT} + \text{NPD} + \text{TT}} \right]$$

2- Cloning by complementation

After randomly introducing mutations in the genome, one screens for the desired mutations. However, after identifying the mutants and describing their phenotype one has to find out which gene was mutated.

This is done by a process called complementational cloning. In this process the wild type gene is reintroduced into the mutant and reverts the phenotype. It is necessary to extract the gene to identify it, recombination or integration into a yeast chromosome is not doable. Thus one uses shuttle plasmids. Shuttle plasmids are plasmids which can propagate in two different organisms.

Thus, another approach was developed, in order to be able to recover the complementing DNA: Random insertion of pieces of DNA from the yeast genome into the 2µm plasmid of yeast -> creation of a plasmid library. One of these plasmids should contain TRP1.

How are these plasmids isolated and amplified? -History of plasmid vectors in S.cerevisiae-

Electron microscopy of budding yeast showed that each cell contained about fifty plasmids. As those plasmids were 2 micrometers in circumference, they were called 2micron plasmids. Unfortunately, it was originally not possible to isolate these plasmids and use them for complementation. However, this discovery showed that yeast can contain plasmids, therefore it was tried to create artificial plasmids. In order to do this the yeast genome was digested with endonucleases and subsequently cyclised with ligase, this treatment resulted in plasmids. To check if the plasmids were taken in by the yeast and replicated, it was attempted to use this plasmids to complement mutations in the tryptophan biosynthesis. After transformation, the trp-yeast cells were put in medium that did not contain any tryptophan. Surprisingly the mutation trp1 was easily complemented, but not any other mutation. As it was unlikely that a mutation in biosynthesis would be dominant negative, it was suspected that the other plasmids were simply not replicated and thus lost. Analysis of the plasmid complementing trp1 showed that it indeed contained a sequence which allowed its replication. This sequence was called ARS (autonomous replication sequence). By combining ARS with other plasmids it was possible to also complement the other trp mutations. Such plasmids were called Yeast replicating plasmid or YRp.

However, these plasmids were still very unstable as they lacked a kinetochore to ensure proper segregation and were rapidly lost if there was not a strong selection pressure. Thus it was tried to make them more stable by including a centromere. This was done by giving the endonucleases
less time to digest the DNA, creating bigger fragments. As the centromere of chromosome four is relatively close to the TRP1 gene, this actually worked, and it was possible to create a plasmid which was much more stable – the loss rate was around 1-5%. This plasmids were called Yeast centromereic plamids or YCp.

**Shuttle Plasmids**

In order to be stable in two organisms – in this case S.cerevisiae and E.coli – the shuttle Plasmid contains a bacterial component (green) and a yeast component (orange). The bacterial component includes an ORI (Origin of replication), so the plasmid can propagate in E.coli and a selectable marker, i.e., an antibiotic resistance gene.

The yeast component includes an ARS, which allows the plasmid to propagate in yeast, centromereic DNA which assembles a kinetochore so the plasmid is correctly segregated, a selectable marker. In such plasmids approximately 10 kbp long fragments of the yeast genome can be randomly cloned to form a genomic library. These libraries can then be screened for complementing genomic fragments as a method to clone the wild type copy of the gene mutated in a mutant of interest.

To determine which gene is responsible for the complementation it is necessary to have overlapping fragments. Therefore a genome library usually contains $10^7$ different plamids.

In yeast, the shuttle plasmid can propagate itself as well without integrating into the genome, thanks to the origin of replication contained in the 2µm sequence, as well as segregation sequences involved in the proper partition of the plasmid between mother and bud at mitosis.

Under what conditions does complementation work?

- Mutation has to be recessive.
- A screen for the mutation has to be doable.
- The gene in question must not be bactericide.
- The gene in question should not be too big.

**Combinations allowing the formation of functional plasmids**

- **YRp (Yeast Replicating plasmid)**: ARS1, TRP1, pBR322 (AmpR)

  =$=$ unstable (no centromere), asymmetrical segregation, plasmids go to the bud very inefficiently. This causes a large increase in the number of plasmids present in the mother cell, while only a fraction of the buds become one or two copies of the plasmid. The rest of the buds do not inherit any.

  Variants of this plasmids were produced were TRP1 was replaced with other selection markers such as LEU2, URA3, HIS3… (Genes involved in the biosynthesis of leucine, uracil and histidine, respectively)

- **YEp (Yeast Episomal plasmid)**: 2 µm plasmid, pBR322 (AmpR), various selection markers available in different versions of the plasmid (TRP1, LEU2, URA3, HIS3…)

  =$=$ very stable, about 20 copies/cell,

- **YCp (Yeast centromereic plasmid)**: ARS1, TRP1, CEN4, pBR322 (AmpR)
Again, variants of this plasmids contain other selection markers in place of TRP1.

CEN4 was identified thanks to the fact that it is close to TRP1 on the genome.

Take-home message:
- CEN4 provides a sequence that makes the plasmid highly stable, in the sense that it is well propagated in the population and rarely lost by any of the daughters at mitosis.
- A replication origin, such as ARS1, is needed for the plasmid to replicate
- 2µm-based plasmids are stable and present in multiple copies in the cell, due to the fact that the 2µm sequence contains both a replication origin and a segregation sequence.

CEN4 and ARS1 were found thanks to the analysis of the TRP1 locus. However, now TRP1 can be replaced by other selection methods based on the complementation of any other auxothrophic mutation by the corresponding wild-type gene, such as LEU2, ADE2, HIS3, HIS4, URA3.

Therefore, currently yeast strains used in the laboratory as “wild-type” strain contain therefore the following auxotrophy markers to allow plasmid selection:

trp1-1 leu2 ade2 his3 his4 lys2 ura3

3- Cloning of multicopy suppressors

In the SUC2-HIS4 sec61-1 double mutant cells grown at semi-restrictive temperature, some of the Suc2-His4 protein is mislocalized to the cytoplasm. Therefore, these cells can grow on plates containing histidinol instead of histidine. This is how the sec61-1 mutant was actually isolated, by screening for mutations that allow the SUC2-HIS4 cells to grow on histidinol+ histidine- plates. At 37°C and above, the sec61-1 mutant cells are unable to translocate any of the ER resident proteins and secreted proteins into the ER, and therefore they stop growing and die.

What is the sequence of Sec61-1? What does SEC61 encode for?

Plasmids of a genomic library based on the 2µm (YEp) were transformed into the sec61-1 mutant cells and plasmids that promoted growth of these cells at 37°C were isolated. The first assumption was that these plasmids would contain the SEC61 gene, which would then complement the sec61-1 mutation. Indeed, the majority of the plasmids identified contained all one and the same gene, which could be shown to be the SEC61 gene. However, analysis also indicated that some other plasmids contained another gene, which when overexpressed made the sec61-1 cells able to grow again at 37°C. In this case, one does not speak anymore of complementation, but of suppression: The gene on these plasmids does not complement the sec61-1 mutation because it does not provide the information to make again a normal Sec61 protein. This gene helps the sec61-1 cells by coding for a protein that stabilizes or helps the sec61-1 protein to better function despite of its
mutation. We say that it suppresses the phenotype of the sec61-1 mutation. In this case, the gene was called SSS1, for Suppressor of Sec Sixty-one, number one.

Thus, SSS1 is an example of a **multi-copy suppressor gene** of a mutation, here sec61-1. It should not be confused with a suppressor mutation, which is a mutation on a gene of the genome that, when combine with another mutation, suppresses the phenotype of that second mutation. A multi-copy suppressor suppresses the primary mutation when the suppressor gene is overactivated (here by surexpression), while the suppressor mutation generally causes suppression by inactivation of a secondary gene. Thus, genes identified by multi-copy suppressor screens encode generally subunits of the same complex as the primary mutant protein, or activators of the primary mutant protein. They suppress either by stabilization, by mass-action law, of the well-assembled complex, or by boosting the activity of the mutated protein. This is called **dosage-dependent suppression**. In contrast, suppressor genes identified by suppressor mutations generally encode for proteins involved in the inhibition of the primary mutant protein, or in their degradation. This is called **secondary mutation suppression**.

In the case of SSS1, it turned out that Sss1 is part, together with Sec61, of the translocon. Thus, Sss1 overexpression stabilizes the Sec61-1 containing complexes, and hence, the Sec61-1 protein.

4- **ORF identification**

A 10 kbp long fragment contains in average approximately 4-5 different genes. Thus, it is helpful if there are several plasmids which complement the Mutation, as one only has to consider the genes which are present in all plasmids.

The gene must be within the overlapping region.

**Propportion of a start and stop codons:**

1 start codon for every 64 codons

3 stop codons for every 64 codons

This means that every single sequence has 6 frames that can be read.

**What should be the size to have a real gene?**

By Poisson distribution you can determine the frequency of randomly open reading frames. In yeast the size of the genome is $12 \times 10^6$ nucleotides, therefore the probability to get open reading frames of 100 codons by random is $12 \times 10^{-6}$. 

**How can you find genes less then 100 codons?**

The best solution is by comparing genomes that are close enough in order to have fixed amino acids sequence but also far away enough so that the rest of the genome is not that conserved. For instance, these comparative studies in yeast are done with different yeast genomes.

- The average size of the gene is around 600-700 aminoacids (2Kb).
- The smallest gene is around 10 amino acids.
- The biggest gene is more than 12kb long.
5- **Identity tests**

A plasmid reverting the phenotype does not necessarily contain the wild type copy of the mutated gene. It could also carry another gene, which corrects the phenotype when over expressed (in the case of a multi-copy suppressor, see above).

If one wants to determine which gene was mutated, it is necessary to distinguish between complementation and suppression.

This is done by looking whether the gene identified by cloning is genetically linked to the original mutation. In order to do this, a marker can be inserted at or right next to the locus of the cloned gene on the chromosome of an otherwise wild type strain. This strain can now be crossed with the original mutant, and segregation analysis will then establish whether the inserted marker is genetically linked to the original mutation (then the original mutation is most likely in the cloned gene), or not (then the cloned gene is most likely a multicopy suppressor).

**Disrupt a gene or insert a marker by homologous recombination**

![Diagram of plasmid insertion into the mutant locus in yeasts](http://bioweb.wku.edu/courses/biol350/CloningVectEuk9/Review.html)
VI- Non-mendelian genetics

1- Mitochondrial genetics

Yeast is a wonderful organism to study mitochondria because, when grown on glucose, i.e. Saccharomyces cerevisiae has the ability to live by glycolysis alone and can survive mutations in mitochondrial genome that arrest oxidative phosphorylation. Such mutations are lethal in many other eukaryotes.

When a mitochondrial genome mutant ρ− mates with wild type cell ρ+, the daughter cells after sporulation can inherit either more copies of the mutant DNA or more copies of the wild-type DNA. This is called non-Mendelian or cytoplasmic inheritance. Successive mitotic divisions further enrich for either type of DNA, so after some generations the resulting cells will contain either ρ+ or ρ− phenotype.

Figure 12: Daughter cells after meiosis can inherit the mitochondrial genome of only one of the parent cells. The mutation ρ− in a mitochondria encoded protein is denoted by blue dots. When parent cells fuse, the mitochondrial genomes also form a continuous reticulum which contains mutant and wild type copies of the gene (blue and red dots). During rounds of vegetative growth, either type of genome is enriched due to the nature of the mutation i.e. most often, a wild type genome is inherited, unless the ρ− contains highly repeated origins of replication. The respirating cells can be selected by growth on a medium containing a carbon source that cannot be used for glycolysis, i.e. glycerol.

Mitochondrial Cytoplasmic Transmission

To prove that mitochondrial genome is shared through cytoplasm, a kar\(^{-}\) mutation, which prevents the fusion of nuclei after mating, was used.

Figure 13: Steps in the yeast mating pathway. Kar5p is localized to the vicinity of the spindle pole body (SPB), the initial site of fusion between haploid nuclei during karyogamy. Kar5p is thought to be required for the completion of nuclear membrane fusion.

To show cytoplasmic transmission, the following crossing was made:

\[
\text{can}^1\rho^1\text{kar}^1\text{ade}^2 \times \text{CAN}^1\rho^+\text{kar}^1\text{leu}^2
\]

- Parent cells do not grow on medium without adenine and leucine.
- The zygote is killed by canavanine (there is one parent harboring intact arginine transporter allele CAN1\(^+\)).
- Only the \text{can}^1 daughter cells will grow on canavanine.
- The \text{can}^1 daughter cells usually have \rho^+ phenotype. That confirms the cytoplasmic inheritance.
**2- Prions**

A remarkable case of non-mendelian inheritance is also characterized by the inheritance of a number of dominant traits propagated by Prions. Prions are proteins that form aggregates in a self-templated manner. These aggregates are generally in the form of amyloid fibers. The process is self-templated in the sense that when a protein containing a prion domain switches to the aggregating conformation it starts to catalyze the same conformation switch on the other copies of the same protein, in a sequence specific manner. Prion-like domains (PrD) are frequently (but not always) characterized by the presence of asparagine and glutamine repeat sequences. These poly-N and poly-Q sequences are prone to switch from an α-Helix (soluble) to a β-Sheet (aggregating) conformation. The β-sheet conformation forms a hydrophobic surface that promotes both the aggregation and the autocatalytic conversion of the other copies of the same protein from the alpha-helix into the β-sheet conformation. In general, but not always, the aggregated form of the protein is then inactive. As a consequence, prion domains auto-amplify the inactivation of the protein in which they are located. Because the aggregates are generally located in the cytoplasm, they are segregated symmetrically at mitosis and during meiosis. Thereby, they act as “infectious” entities that spread into the entire colony derived from a single “infected” cell and become inherited in a non-Mendelian manner during mitosis.

A clear case of the prion mechanism is the case of the PSI+ phenotype in yeast (noted [PSI+]). The hallmark of [PSI+] cells is that they suppress stop-mutations at high rates. For example, the ade2-101 and the ade1-14 mutations prevent adenine biosynthesis. The mutant cells cannot grow on medium lacking adenine and form red colonies on medium containing adenine (once the cells enter the respiration phase). In both cases the underlying mutation (ade2-101 and ade1-14) is the exchange of a single nucleotide, causing the introduction of a premature stop codon in the open-reading-frame of the gene. As a consequence, translation leads to the formation of a truncated and inactive product. However, [PSI+]ade2-101 and [PSI+] ade1-14 cells do grow on medium...
lacking adenine and do not form red colonies. They produce at low level the full-length Ade2 and Ade1 proteins, respectively, due to substantial level of read-through of the STOP codon. Remarkably, [psi-] ade2-101 and [psi-] ade1-14 cells are able to form colonies on adenine-free medium, at a low frequency ($10^{-3}$-$10^{-6}$), and these colonies are white instead of red. The cells of these colonies have all the hallmark of the [PSI+] prion: upon crosses, the diploid shows the [PSI+] phenotype, and this phenotype segregates 4+:0- at meiosis. A mendelian segregation would yield a 2:2 ratio. The frequency of appearance is much higher than what would be expected if it were due to a mutation (reversion frequency: $10^{-9}$). Remarkably, when cells are grown in presence of a protein-denaturing agent such as guanidinium chloride, a high frequency of reversion ([PSI+] to [psi-]) is observed, indicating that the vector of inheritance is protein based.

The [PSI+] phenotype disappears in case of mutations in the SUP35 gene:

The Sup35 protein contains a prion-domain and is at the same time a translation factor involved in STOP codon recognition and therefore it is involved in translation termination. If the amount of Sup35 protein is reduced, it leads to ade2-101 mutation suppression, due to reading through the premature stop and synthesis of the full-length protein. When Sup35 switches to its prion form and aggregates, this reduces the amount of active Sup35 present in the cell. As a consequence, ade2-101 cells containing the aggregated form of Sup35 are white and can grow normally on adenine-free medium.

Cells in which the prion-domain of the SUP35 gene is deleted cannot switch to the [PSI+] phenotype, and do not inherit this phenotype once crossed with [PSI+] cells. Thus, the PrD of Sup35 is required for the generation of the phenotype. It is also required for the formation of Sup35 aggregates. These aggregates are observed only in the [PSI+] cells. These aggregates disappear in presence of guanidinium, and this correlates to the conversion of the [PSI+] phenotype of these cells to [psi-]

Therefore, the model for prion behavior of the [PSI+] trait is the following: Due to the fact that Sup35 aggregates are floating in the cytoplasm, they are inherited by all daughter cells. Likewise, due to their autocatalytic activity, the aggregates turn the newly synthetized molecules of Sup35 into their prion form.

This is the reason why doing a cross between [PSI+] and [psi-] cells leads to the formation of only [PSI+] spores after meiosis.

A switch between the [PSI+] (β-sheet) and the [psi-] (α-Helix) conformation can happen spontaneous when the cell is affected by stress.

Stress will induce the cell to a change in gene expression and switching to a SUP35+ conformation will permit to control this changes.

(In some species the sexual state is dependent on prions activity and also in humans they play an important role, for example in memory storing or like in the Creutzfeldt–Jakob disease)

Cross: “if it were a dominant allele”

$ade2-101[PSI^+] \times ade2-101[psi^-] \Rightarrow$ diploid of [PSI+] $\Rightarrow$ 2 spores of [PSI+] phenotype and 2 spores [psi-].
However, the reality looks like this:

\[ \text{ade2-101[PSI']} \times \text{ade2-101[psi]} \Rightarrow \text{diploid of [PSI']} \Rightarrow 4 \text{ spores of [PSI']} \text{ phenotype. (non-mendelian)} \]

In addition, there are always [psi'] phenotype (red cells) occurring at a frequency of \(10^{-3}\) both in diploid cells and in spores. It was confirmed by sequencing analysis, that the \text{ade2-101} locus did not mutate, it was still the same allele in both [psi'] and [PSI'] phenotype after crossing. These results indicated that there is something in cytoplasm allowing translation to go through STOP codon in \text{ade2-101} mutant and restore normal adenine synthesis. The [PSI'] phenotype is inherited in non-Mendelian manner.

When cells of [PSI'] phenotype are mutagenised, it does not affect the frequency of reversion to the [psi'] phenotype, it was still \(10^{-3}\). But when [PSI'] cells were grown in higher temperature (affects protein folding), frequency of reverse phenotype increased up to \(10^{-2}\) or even \(10^{-1}\). These results also support the hypothesis that [PSI'] phenotype is based on protein structure.

[PSI'] phenotype disappears in case of mutations in two genes:

Purified from [psi'] cells, the Sup35 protein is soluble, but purified from [PSI'], it is insoluble and forms amyloid fibers, that are resistant to proteases. If one incubates [psi'] cells with amyloid Sup35 protein, these cells develop the [PSI'] phenotype at high frequency.

Conclusion: Sup35 protein is a prion, which changes the conformation of other proteins.

Distribution of prions via Hsp104: This is a chaperone, which recognizes amyloid aggregates and cuts the amyloid fibers into small pieces, which become seeds, necessary for the inheritance of the [PSI'] phenotype. Hsp104 is responsible for prion spread. If Hsp104 is not working, inheritance of prions decreases.

The prion phenotype is an infectious phenotype.

\[ \text{Figure 13: Four known types of prion: toxic amyloid, inactive amyloid, active amyloid, and self-activating enzyme. All are infectious and self-generating (red arrows). From Wickner et al. 2004} \]
Evolutionary reasons for and consequences of PrD: For a long time, prion phenotype were thought to be due to a proteostasis disease, the aggregated conformation being pathologic. However, the fact that PrD domains are highly conserved in evolution and widespread in many different types of proteins indicates that there is selection pressure in favor of their formation and maintenance. In the case of the SUP35 gene, the hypothesis is the following:

When cells grow for extended periods of time in the same environmental conditions, there is no selection pressure on the genes that are not contributing substantially to fitness under those growth conditions. As a consequence, these genes are likely to accumulate mutations. Among mutations that inactivate these genes, the most frequent ones are the mutations introducing premature stop codons. Changing the growth conditions changes the requirement on the genome, and many genes that were not required before may start to be so under the new growth regime. Allowing to read through stop codons, at least at low level, permits to survive even if the genes that are now required have acquired one or few stop codons when they were not needed. Since changing growth conditions is perceived by the cell first as a stress, the observation that switch to the [PSI+] phenotype is highly enhanced by stress is in good agreement with the model above. The classical objection is that reading through stops would affect all proteins that would all become too long. These extra amino acids may also affect growth. However, inspection of the yeast genome reveals that the vast majority of active open-reading-frames (real genes) have several stop codons following each other at their end. This is probably the signature of the fact that indeed yeast cells have spent a substantial fraction of their evolutionary history in the [PSI+] state, and that during these periods evolution has selected the appearance of stop mutations helping the cell to cope with a low efficiency of translation termination.

3- Selfish DNA

The cell wall prevents foreign viral DNA or RNA from entering the cell (vide supra). However, yeast cells carry internal viruses, also called selfish genetic material.

Three types of selfish DNA are known:

2 micron circular plasmids replicate 20-200 times/cell cycle by hijacking the cell's replication machinery. These plasmids do not correspond to a virus.

Ty (transposons of yeast) are similar to retroviruses. They encode a reverse transcriptase and enveloping proteins. They do form viral particle in the cell, but never escape the cell and therefore never infect other cells. They propagate into new cells only during mating.

Killer viruses: The viral RNA encodes its own independent replication system and toxins, which selectively kill neighboring cells that do not carry the virus. The virus is actually advantageous for the carrier because it eliminates competitors.
CONGRATULATIONS!! IT'S A BOUNCING BABY BIOCHEMIST!!!

Yes!

How we get scientists
Annex: Sequencing methods

1-The first sequencing method: Sanger Chain Termination

In short the first method was based on terminating the DNA polymerization at one particular nucleotide and then comparing the piece by visualizing them with gel-electrophoresis. This termination achieved by substituting one particular dNTP with ddNTP.

Figure 14: The different NTPs. The hydroxyl group at the 3’ position is necessary to form phosphodiester bonds (the backbone of DNA). If this is missing as in ddNTPs the polymerization ends at that nucleotide.

Figure 15: An example of how the results of a Sanger-Termination Method is interpreted. In this example the sequence of the analyzed DNA fragment is TGACCAGACT.
2- Shotgun sequencing: A combination of biochemistry and bioinformatics

In this method DNA of interest is chopped into small pieces and then sequenced. This method is usually used if the gene of interest or the DNA fragment is too long. After the small pieces are sequenced they are stitched back together by a computer. The fragmentation occurs randomly thus every piece contains some complimentary information to one other fragment. With the help of certain softwares the pieces can be added together to finally retrieve the sequence of the gene of interest.

Hierarchical Shotgun Sequencing Method

- Genomic DNA
- BAC Library
- Create Contig Map
- Sequence Each Contig with Shotgun Approach
- Align Contiguous Sequences
- Generate Finished Sequence

Figure 16: A schematic illustration of shotgun sequencing

3- Next generation sequencing

A cheap and a high throughput technology.

Pyrosequencing

This is a fast method that again combines biochemistry/biophysics with bioinformatics.

- Nucleotides are sequentially released the complementary nucleotide binds and releases diphosphate (PP)
- Diphosphate binds sulfarylase with AMP yielding ATP
- ATP activates Luciferase and light is released
- The released light is detected by a photomultiplier and according to its intensity the number of nucleotides is assigned to that position.
Figure 17: A schematic illustration of pyrosequencing

Illumina Sequencing

Figure 18: Steps of illumina sequencing.
Illumina sequencing uses a similar technique as in pyrosequencing. However instead of sequentially releasing color tagged dNTPs, the machine releases all four types of dNTPs at the same time and the matching one naturally binds the DNA. After it binds, a laser is activated and the laser cuts of the tag. And the next color tagged dNTP binds. and so on. For earlier steps see figure 18.

**Single Molecule Real Time Testing (SMRT) Technology**

The essential trick is light detection at extremely low levels of molecules $20 \times 10^{-21}$. Once more the dNTPs are tagged with fluorophore molecules. After each addition a light reaction occurs after the cleavage of the fluorophore tag.

*Figure 19: Illustration of SMRT technology*
**Ion-semiconductor sequencing**

The detection of the polymerase reaction is based on $\text{H}^+$ release during the dNTP hydrolysis reaction. The nucleotides are added sequentially and the increase in $\text{H}^+$ are matched with the dNTP release.

![Diagram](image)

**Figure 20:** A simplified drawing of a well, a bead containing DNA template, and the underlying sensor and electronics. Protons ($\text{H}^+$) are released when nucleotides (dNTP) are incorporated on the growing DNA strands, changing the pH of the well ($\Delta\text{pH}$). This induces a change in surface potential of the metal-oxide-sensing layer, and a change in potential ($\Delta V$) of the source terminal of the underlying field-effect transistor.
Which method is most suited to my sample?

<table>
<thead>
<tr>
<th>Method</th>
<th>Single-molecule real-time sequencing (Pacific Bio)</th>
<th>Ion semiconductor (Ion Torrent sequencing)</th>
<th>Pyrosequencing (454)</th>
<th>Sequencing by synthesis (Illumina)</th>
<th>Sequencing by ligation (SOLiD sequencing)</th>
<th>Chain termination (Sanger sequencing)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read length</td>
<td>2900 bp average[38]</td>
<td>200 bp</td>
<td>700 bp</td>
<td>50 to 250 bp</td>
<td>50+35 or 50+50 bp</td>
<td>400 to 900 bp</td>
</tr>
<tr>
<td>Accuracy</td>
<td>87% (read length mode), 99% (accuracy mode)</td>
<td>96%</td>
<td>99.9%</td>
<td>98%</td>
<td>99.9%</td>
<td>99.9%</td>
</tr>
<tr>
<td>Reads per run</td>
<td>35–75 thousand [39]</td>
<td>up to 5 million</td>
<td>1 million</td>
<td>up to 3 billion</td>
<td>1.2 to 1.4 billion</td>
<td>N/A</td>
</tr>
<tr>
<td>Time per run</td>
<td>30 minutes to 2 hours [40]</td>
<td>2 hours</td>
<td>24 hours</td>
<td>1 to 10 days, depending upon sequencer and specified read length[41]</td>
<td>1 to 2 weeks</td>
<td>20 minutes to 3 hours</td>
</tr>
<tr>
<td>Cost per 1 million bases (in US$)</td>
<td>$2</td>
<td>$1</td>
<td>$10</td>
<td>$0.05 to $0.15</td>
<td>$0.13</td>
<td>$2400</td>
</tr>
<tr>
<td>Advantages</td>
<td>Longest read length.</td>
<td>Less expensive equipment.</td>
<td>Long read size.</td>
<td>Potential for high sequence yield, depending upon sequencer model and desired application.</td>
<td>Low cost per base.</td>
<td>Long individual reads. Useful for many applications.</td>
</tr>
<tr>
<td>Disadvantages</td>
<td>Fast. Detects 4mC, 5mC, 6mA.</td>
<td>Fast.</td>
<td>Fast.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low yield at high accuracy. Equipment can be very expensive.</td>
<td>Runs are expensive. Homopolymer errors.</td>
<td>Equipment can be very expensive.</td>
<td>Slower than other methods.</td>
<td>More expensive and impractical for larger sequencing projects.</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: This comparison gives an overview of the next generation sequencing techniques. Most importantly the sample size is the first determining factor in which method to use. (From www.wikipedia.org)