Concepts in Modern Genetics

19.9.2017

(Mon: 12:45 – 14:30; Tue: 7:45 – 9:20 - - - Exam is on 10.1.2018 at 14:30)

Genetic analysis in yeast

The power of yeast

The haploid genome of S. cerevisiae contains 12000 kb of DNA distributed over 16 chromosomes. Its genome shows a high degree of gene density: on average, there is one gene every 2 kb, which is 50 times higher than in humans. The yeast genome contains roughly 6600 genes, most of which do not contain introns. S. cerevisiae also exhibits high conservation of genes and pathways.

Further distinctive features of yeast as a model organism: Yeast uses homologous recombination to repair DNA strand breaks, where a healthy DNA is used as the template or substrate to repair the broken DNA part. This technique is further used in changing the mating type in yeast and to ensure correct segregation of homologous chromosomes during meiosis. DNAs carrying homologies to the yeast genome are effectively integrated into the genome. That way, one can integrate exogenous DNA in a specific location in the yeast genome.

Secondly, yeast contains autonomously replicating sequences (=: ARS) that contain an origin of replication. ARS in a plasmid will multiply in yeast, so external DNA can be expressed in yeast too.

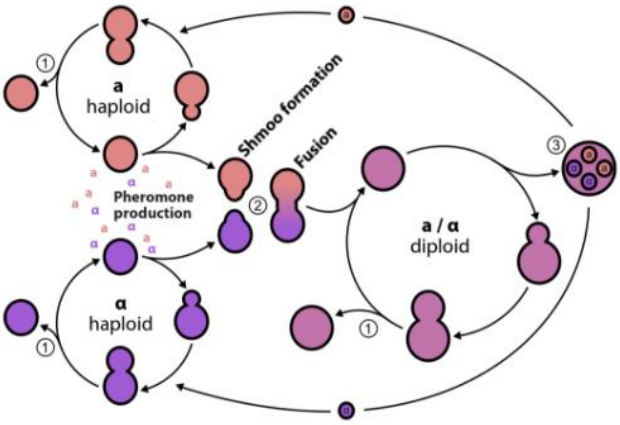
Thirdly, yeast can change its life cycle (haploid or diploid). Haploid life cycle is good to observe gene mutations that are recessive. Diploid life cycle if helpful to maintain such a mutation or to study gene interactions.

Remarkable cell biology of yeast: The thick cell wall of most yeast species also renders them immune to external viruses. Consequently, the yeast genome shows little modification by horizontal gene transfer and is fairly stable. This might also explain why homologous recombination is particularly efficient in yeast, since the risk of inserting foreign DNA is lower compared to other species.

Candida albicans is a pathogen that causes thrush.

Mating type determination and mating in yeast:

MAT-a cells produce the a-factor (pheromone) and they respond to the alpha-factor through the a-receptor that induces a response within the cell. MAT-alpha cells produce the alpha-factor (pheromone) and they respond to the a-factor through a-receptors that induces a response within the cell. Both cell types produce a shmoo that is needed for fusion. Those factors are only produced in haploid stage.



How does this work on a molecular level? The mating type is determined by genes present in the mating type locus (MAT) on chromosome 3. There are two different alleles of the MAT locus: MATa and MATα. The two mating type alleles differ by 700 base pairs of sequences that encode regulators of the two different haploid mating types.

Table 1 Overview over the abbreviations of the yeast sex-determining genes and their functions.

abbreviation	name	function
asg	**a**-specific genes	• α-factor pheromone receptor • a-factor pheromone
αsg	**α**-specific genes	• a-factor pheromone receptor • α-factor pheromone
hsg	haploid-specific genes	• budding-pattern genes • fusion genes • signal-transduction genes • repressor of diploid-specific gene
dsg	diploid-specific genes	• meiotic genes

There are a1 and a2 genes as well as alpha1 and alpha2 genes.

In the haploid stage, alpha1 is a transcriptional activator and alpha2 is a transcriptional repressor. a2 has no function by itself and is thought to be a pseudogene. Also, a1 has no function by itself (has to be bound to alpha2 to be functional). In the diploids, a1/alpha2 dimer is a transcriptional repressor.d

The a sex is the default state, when alpha1/2 is not expressed (only MAT-a is present in the haploid cell). Also, hsg is expressed (obviously). When only alpha2 is present, it will repress the expression of a-specific genes and alpha1 will activate the expression of alpha-specific genes. Therefore, α2 is the ultimate factor that determines the mating type.

In MATa/MATα diploids, the α2 protein has two functions: By itself, it represses asg. Together with a1, it forms a a1/α2-dimer that functions as a repressor for hsg. Since hsg normally represses dsg, the repression of hsg leads to an expression of dsg in diploids. The a1/α2-dimer also represses α1 (thus, αsg are repressed). That means that α2 can extend its function in MATa/MATα diploids due to the ability to form heterodimers with a1.

How does this sex-determining system ensure that MATa/MATα diploids do not mate?

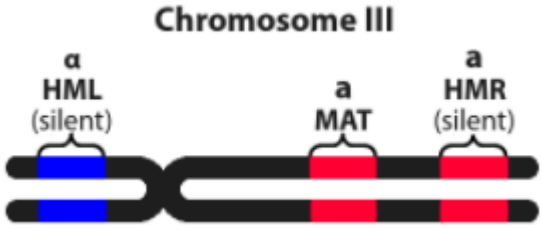
Answer: The a1/α2 complex turns off haploid genes. It inhibits α1; thus, αsg is not expressed and no α-factor is produced (because α1 is required to activate the expression of αsg). On the other hand, the a1/α2 complex allows the expression of α2, the repressor of asg, thus, no asg are expressed. Thus, the repression of both asg and αsg ensures that the diploid does not produce mating pheromones and is non-mating.

Def. Homothallism: This process of converting some cells in a colony from one mating type to the other. An organism is called homothallic if both male and female reproductive structures are present in the same thallus.

Kor.: HO gene := homothallic gene (on chromosome 3).

Yeast cells are heterothallic (like humans) and homothallic.

Haploid yeast switch mating type by replacing the information present at their MAT locus. For example, a cell will switch to an α cell by replacing the MATa allele with the MATα allele. This replacement of one allele of MAT for the other is possible, because yeast cells carry an additional copy of both the MATa and MATα alleles: the HML (Hidden MAT Left) and HMR (Hidden MAT Right) locus (see figure 1-5). These loci typically carry a silenced copy of the MATα or the MATa allele, respectively. The HML and HMR loci are silenced by proteins interacting with sequences flanking the HMR or HML locus, thus forming short regions of heterochromatin that is not transcribed.



The HO gene induces the switching of the mating type. It codes for an endonuclease that cleaves DNA specifically at the MAT locus. Then exonucleases attach to the broken parts and degrade the MAT allele at both ends (it is removed basically). Either the HML (alpha) or HMR (a) will now serve as the repair template and thus, the mating type is changed. A mutation in HO gives an heterothallic organism.

Normally, HMR would want to be used for repair, because it is closer to MAT-a (and MAT-alpha). Behind is a complex search machinery for repair. In reality, it is always the opposite mating type that will be recovered.

Reason: in alpha yeast, HML is farther away from MAT-alpha locus. Therefore, HMR is closer and the new mating type will be MAT-a. In a yeast, the arms of the chromosome are bent, so that HML is physically closer to MAT-a. Also, another loci RE is expressed which favors HML over HMR. In alpha yeast, RE is repressed though (Fkh1 binds RE).

In mitosis of yeast, if the initial mother cell is type a, it will bud and when budding is done, the daughter cell will be type a and mother will become alpha (HO gene activation during budding). This keep continuing for some time like this. They do not form diploids immediately, because the mother cell is not ready to respond to pheromones right away and the daughter cell is still small, so it needs to grow first. Therefore, we have a temporal reproductive barrier.

The relationship between mutation and phenotype

Def. Auxotrophy: A cell that needs besides a C- and N-source at least one additional external nutrient to synthesize all other compounds needed to grow etc.

Such mutations are not limited to amino acid auxotrophy, but exist also for strains that cannot synthesize the bases for DNA and RNA synthesis, or key lipids. For example, a strain that cannot synthesize uracil will be called uracil auxotroph. In the wild, yeast strains are generally complete prototrophs.

Temperature sensitive (=: ts) mutations: Ts mutations are functional at low (permissive) temperatures, yet non-functional at high (restrictive) temperatures, and thus a rise in temperature quickly abolishes protein function.

Dominant-negative (=: dn) mutations: If a dn mutation affects a transcriptional activator, the altered protein retains DNA binding activity, but lacks the ability to transactivate. It can complex with the DNA binding sites and displace the WT protein.

Epistasis

Epistasis occurs in the following scenarios: Whenever two or more loci interact to create new phenotypes, whenever an allele at one locus masks the effects of alleles at one or more other loci, whenever an allele at one locus modifies the effects of alleles at one or more other loci.

Epistasis can be seen on genes that act in the same pathway, but also for genes acting in very different biological processes.

Ex. Same pathway: In yeast, we have the genes ade2 and ade3. A mutation in ade2 will lead to a red color in yeast cells, because there is an accumulation of AIR (mitochondria (rho+) oxidize AIR and it becomes red-AIR). The ade2 gene codes for an enzyme that uses AIR as its substrate. The initial white phenotype can be restored by a second mutation in the ade3 gene. Effectively, ade3 operates more upstream, therefore AIR is never produced and the mutation in ade2 never affects the phenotype in the first place. We say ade3 is epistatic to ade2 (ade3 is epistatic over ade2).

Kor.: ade2 can be mutated in different places, such as ade2-delta (operon mutation) or ade2-108 (mutation at the 108^{th} base – a premature stop codon).

Ex. Different biological processes: In a ts mutation in yeast, we have the sec61-1 allele. This gene codes for ER-embedded translocation channel mediating protein import into the ER. At 37°C, part of the Sec61 protein is misfolded, although the full protein still shows activity. Nevertheless, the misfolded domain of the protein is recognized and leads the degradation of the entire protein by the ERAD machinery.

The phenotype can be restored by a mutation in ubc6 (although at a compromised activity). As a consequence, the sec61-1 protein is no longer degraded and can fulfil its function. Thus, here we can say that the lethality due to the sec61-1 mutation at 37°C is not caused by the fact that the mutation would inactivate the Sec61 protein but to how the cells react to the change in Sec61 organization, a change that they perceive as a stress. In this case, ubc6 is epistatic to sec61-1.

Synthetic lethality: Arginine availability

Arginine is made available to a yeast cell via two ways: There is the CAN1 transporter gene and the arg genes, that make de novo biosynthesis of arginine. CAN1 also imports the toxic canavanine molecule. Therefore, a mutation in CAN1 will render the cell resistant to the toxin and viable as long as the arg genes are functional. A second mutation in *arg* leads to lethality, since arginine is no longer available to the cell, because it is essential.

Isolation of yeast mutants

A typical forward genetic screen has the following structure:

Selection of a biological process (phenotype).
Mutagenesis.
Choosing a screening strategy.
Identification of interesting mutants (screening vs selection).
Identification of the gene resulting in the mutated phenotype (mapping and cloning).

Mutagenesis: EMS produces point mutations by alkylating guanines to produce 6-O-ethylguanine. During replication, the incorporated base is modified. This leads to point mutations.

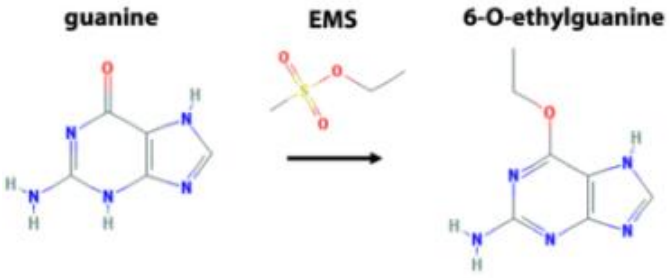


Figure 3-2 Mutagenesis by ethyl methanesulfonate (EMS). EMS alkylates guanines to produce 6-O-ethylguanine, typically causing transition mutations (from G:C to A:T).

For example, exposure to X-rays induces DNA double strand breaks, leading to chromosomal mutations such as deletions.

Def. progeny: daughter cells. Descendants of a mother cell.

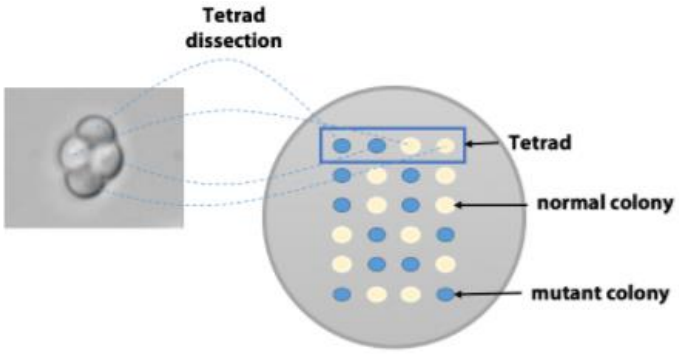


Figure 3-4 Tetrad analysis. When undergoing meiosis, yeast cells produce four haploid spores that are packaged together in an ascus and build a tetrad **(left)**. Tetrads can be dissected, i.e., separated under the microscope and transferred onto agar plates and grown to colonies **(right)**. The spores of each tetrad are placed such that the four colonies grow in one row.

In tetrad analysis, we can check the phenotypes of the discrete spores from the tetrads. Effectively, do they behave like a normal colony or like a mutant colony?

If we know the ratio of muants in a tetrade we can figure out the type of mutation: it can either be due to Mendelian laws or non-Mendelian (several genes affect the mutation, epigenetics).

Here, tetrad analysis can be used to test for single-gene segregation after crossing the mutant cell to a wild type strain. If a single gene is the cause, the mutant allele **should segregate in a 2:2 manner**, such that half of the resulting spores will carry the mutant gene and show the mutant phenotype, and the other half will carry the wild-type gene and show the wild-type phenotype. If the mutant phenotype **does not segregate 2:2, the phenotype is not caused by a single gene**.

Ex. Histidine dependence: A mutation in His1 leads to the inability to de novo biosnythesize histidine, leading to a dependence on histidine provided by the medium (agar plate). It segregates on a 2:2 manner.

To confirm dominance or recessivity, one needs to mate the mutated yeast cells (that cannot grow on histidine lacking mediums) with the wild type. If the yeast cells grows on histidine lacking mediums, the his1-1 mutation is recessive, else it was dominant.

Dominant mutations cannot be complemented. For this, we sheare the DNA of the mutant and insert it in plasmids (2-3 fragments per plasmid) and insert these in wild type yeasts. Those who cannot grow on galactose (assume GALx_D was the dominant mutation for galactose cannot be used as carbon source), possess the mutated genes in the plasmids. Then simply sequence the plasmid and maybe to 1-gene-1-plasmid inserts again. Or you can do signal processing.

Complementation test in yeast:

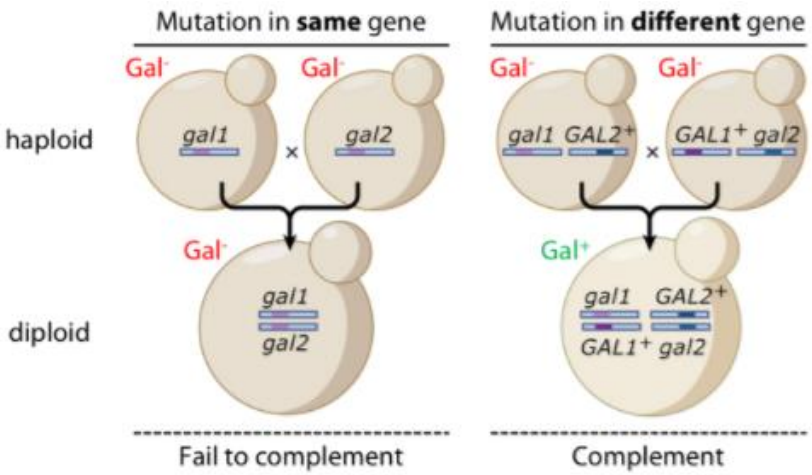


Figure 3-8 Complementation test in yeast. In order to test whether two mutants that were identified in a screen have recessive mutations in the same gene, they are crossed to produce a diploid cell. If the resulting diploid shows the mutant phenotype (Gal⁻, cannot grow without galactose), the mutation is likely in the same gene (left). If the diploid shows the wild-type phenotype (Gal⁺, can grow without galactose), complementation was successful and the mutation is likely in different genes (right). Here, *gal1* and *gal2* are the two identified mutant alleles causing galactose dependence. In the example on the right, *GAL1⁺* and *GAL2⁺* are the wild-type alleles encoding the dominant phenotype Gal⁺. (adapted from ergito.com)

4:0 segregation means that all spores show the mutated phenotype. Thus, the two mutations are tightly linked (on the same gene).

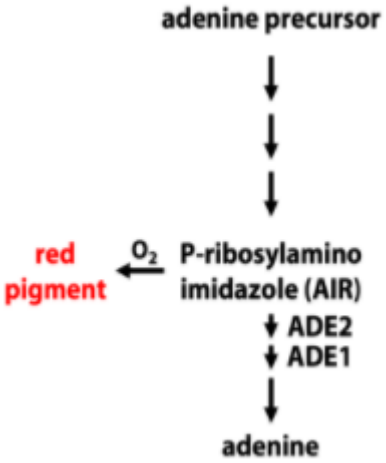
Examples of genetic screens in yeast

Cell cycle: Prepare conditional ts mutants and identify the mutated cell colony (at 23°C they grow, but at 36°C they are stuck in different stages of the cell cycle/budding process – cell cycle and budding occurs parallel).

Def. cell division control genes (=: cdc): Genes associated with the check points of the cell cycle.

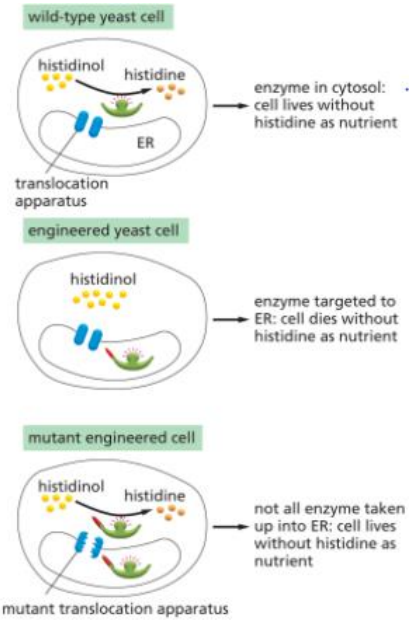
Kor.: cdc28: cells fail to enter the cell cycle and do not replicate their DNA nor does budding occur.
cdc24 and cdc42: mitosis already occurred, but budding was not complete, therefore, cells have big unbudded buds.
cdc3, cdc10, cdc11, cdc12: cytokinesis mutants.

Metabolism: Adenine biosynthesis: Yeast cells will turn red when either ade2 or ade1 is effective due to the accumulation of AIR (and no adenine is synthesized therefore). Also, if the mutants are grown under anaerobic conditions they do not turn red, but are white, so oxidative metabolism is needed for this pathway. This allows for an easy non-selecive method.



Secretory pathway: The following reaction occurs in the cytoplasm in yeast: histidinol + HIS4C → histidine (essential). Cells with some sort

of defect or absence of HIS4C die.



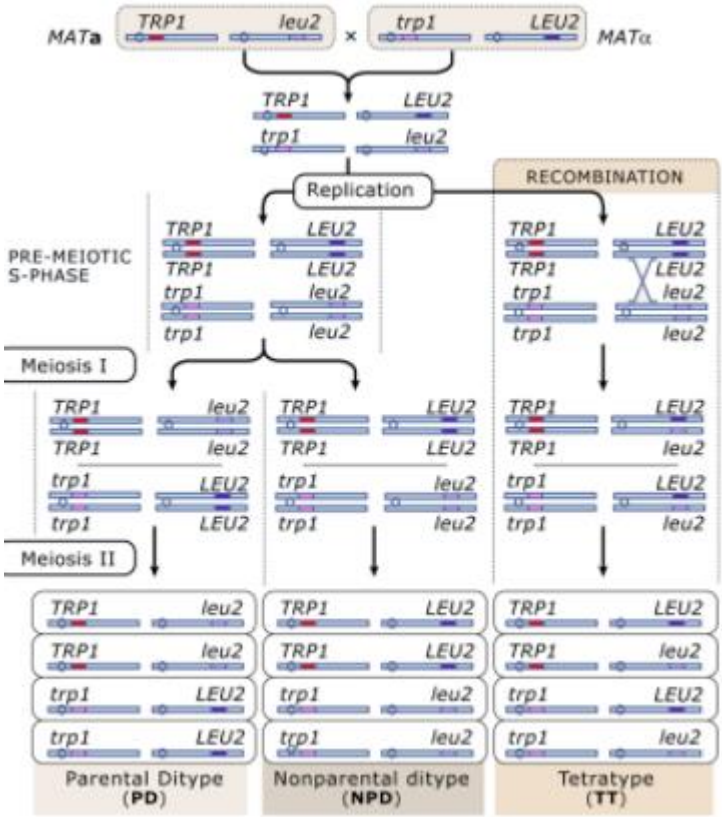
Researchers added a signal sequence to the genes encoding HIS4C such that it will be translocated in the ER. Since the ER is impermeable to histidinol and histidine, those cells will die. Then, genes involved in the translocation process were identified and mutated such that HIS4C partially remained in the cytoplasm leading to viable mutants.

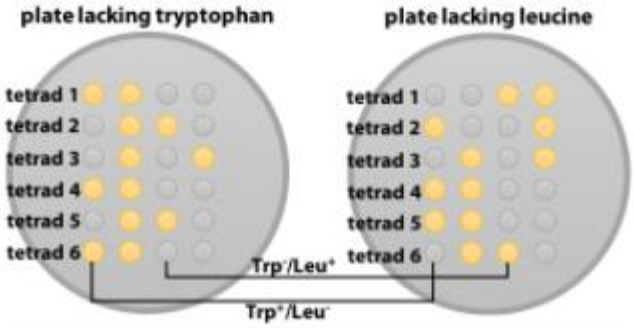
Drug resistance: Cycloheximide binds to ribosomes and renders protein synthesis ineffective. Those cells do not grow. There are mutants in cycloheximide that are also temperature sensitive (crl mutants) where cycloheximide binding to ribosomal structure is affected.
There is one outstanding mutation that fully recovers protein synthesis at high concentrations of cycloheximide under all temperatures: Mutations in the CYH1 gene. The CYH1R allele affects binding of cycloheximide to ribosomes. Thus, the ribosomes can work normally. CYH1S is the dominant allele that does not affect binding positively. Therefore, only CYH1R homozygous cells can grow under these conditions.

Gene mapping and identification in yeast

Tetrad analysis: Let A, B be the dominant alleles of a gene and a,b the recessive alleles. One is interested in the frequency. One crosses to haploid yeast cells, resulting in one diploid yeast cell that can undergo sporulation (tetrad formation) via the removal of an N-source.

Definitions: PD := parental ditype, NPD := non-parental ditype, TT := tetratype.





tetrad	Trp⁺/Leu⁺	Trp⁺/Leu⁻	Trp⁻/Leu⁺	Trp⁻/Leu⁻	type
1	0	2	2	0	PD
2	0	2	2	0	PD
3	2	0	0	2	NPD
4	2	0	0	2	NPD
5	1	1	1	1	TT
6	1	1	1	1	TT

Figure 5-1 The three classes of tetrads produced by random segregation of two genes (here: *TRP1/trp1*, and *LEU2/leu2*) are: (1) parental ditypes (PD); (2) nonparental ditypes (NPD); and (3) tetratypes (TT). Six exemplary tetrads resulting from the shown cross were dissected and grown on agar plates. The colonies were replica-plated onto plates lacking (i) tryptophan and (ii) leucine. Colonies that grew without tryptophan and leucine are abbreviated in the table below with Trp⁻ and Leu⁻, respectively. The table shows how the tetrad type can be determined by comparison of colony growth on the two plates. (adapted from ergito.com)

Unlinked gene pairs, and no crossing over between either gene and its centromere, yields PD = NPD and TT = 0.

Unlinked gene pairs, with a crossing over between either gene and its centromere yields PD = NPD and TT > 0.

Now, A and B, are located on the same chromosome and positioned relatively close to each other such that there is less than 25% chance of a crossover between them. A is closer to the centromere than B. Thus, if we cross a wild type of genotype Ab with another of genotype aB, in >50% of the tetrads, there was no crossover at all between the two loci, such that after meiosis I, one nucleus contains the sister chromatids Ab/Ab and the other one contains the sister chromatids aB/aB. After meiosis II, we will therefore have two spores of genotype Ab and two spores of genotype aB. Hence, >50% of the tetrads resulting from the cross will be PD.

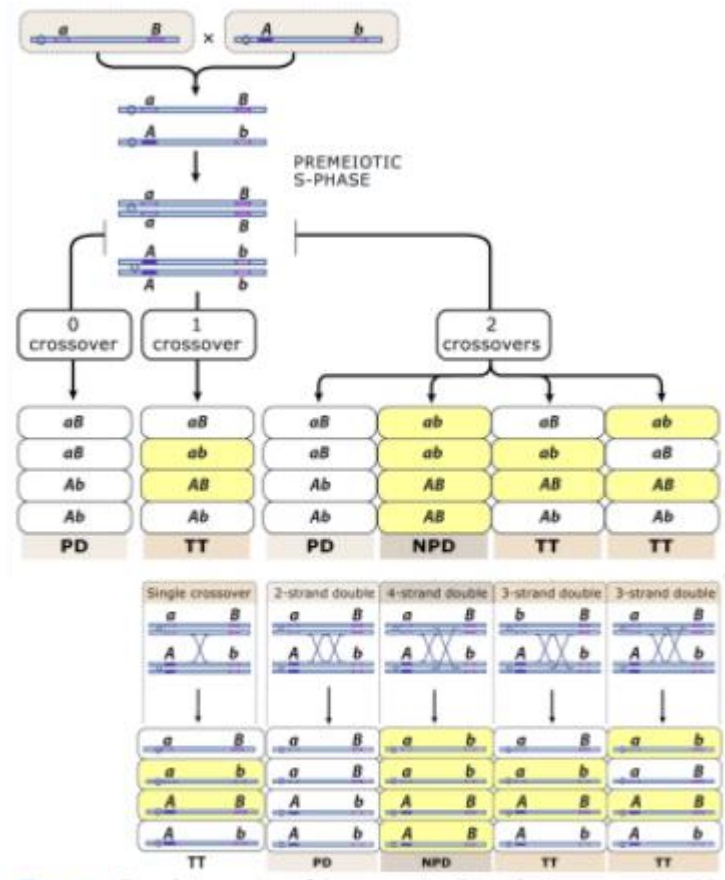


Figure 5-2 Tetrad types arising from a cross involving the two genes A and B that are closely linked, with A being closer to the centromere than B. Possible tetrad types are indicated in case of 0 crossover (more than 50% of the tetrads), 1 crossover (rare) or 2 crossovers (very rare). Recombinant spores are marked in yellow. (adapted from ergito.com)

As a consequence, for linked genes (less than 25% chance of crossover) PD>TT>NPD.

Finally, let's consider the case where the two loci A and B are completely unlinked to each other and far from centromeres. These loci can rearrange in all possible manners with the same probabilities. This can lead to the following 6 tetrads:

AB AB ab ab
Ab Ab aB aB
Ab AB aB ab
AB Ab ab aB
AB Ab aB ab
Ab AB ab aB

These represent one PD, one NPD and four TT. Since they all come with the same probability, this means that we if two genes are completely unlinked the proportions of the different tetrads will be 1:1:4 (PD:NPD:TT).

Consider a cross between trp1 and a hypothetical mutation in gene B. We obtain the following numbers of tetrads:

40 PD; 40 NPD; and 20 TT.

What does this mean for the possible centromere linkage of B?

Because PD = NPD, we know that trp1 and B are unlinked to each other.

What about the linkage of B to the centromere? Remember that linkage is determined by the number of recombinants divided by the number of total progeny. In our example, there are 20 TT tetrads and a total of 100 tetrads. Because each tetrad in this case has 4 spores, two of which are recombinant, there are a total of 40 recombinant spores. The total progeny equals 400 spores (from the 100 total tetrads). Therefore, the linkage is calculated by 40 recombinants/400 total progeny = 10 map units. Thus, B is 10 centimorgan away from its centromere.

For yeast, this mapping function is usually expressed by an equivalent function in terms of tetrads: linkage = (½ · TT)/number of total tetrads. In our example, (½ · 20)/100 = 10%, which corresponds to 10 centimorgan (cM).

In the above example, the genes are unlinked to each other but they are centromere-linked (close to the centromere).

How does one , in practice, determine whether a tetrad is PD/NPD/TT?

Answer: In order to determine, whether a tetrad is PD, NPD, or TT, one isolates the spores by dissecting each ascus, aligns the four spores of each ascus in one row on an agar plate and grows them to colonies. Depending on the gene of interest, selection criteria are introduced by replica-plating in order to determine the phenotype of each spore. By comparison with the parent generation, the type of the tetrad can be determined.

Understanding frequencies: In genetic linkage, assume <25cM and if all tetrads show PD, then the distance between these two loci is non-measurably small.

In centromere linkage, if NPD=0 **and** they are closely linked, then use formula: linkage = (0.5 * TT)/(number of total tetrads) if and only if yeast is the model organism.
If NPD>0, then linkage = 0.5 * (TT + 6*NPD)/(number of total tetrads).

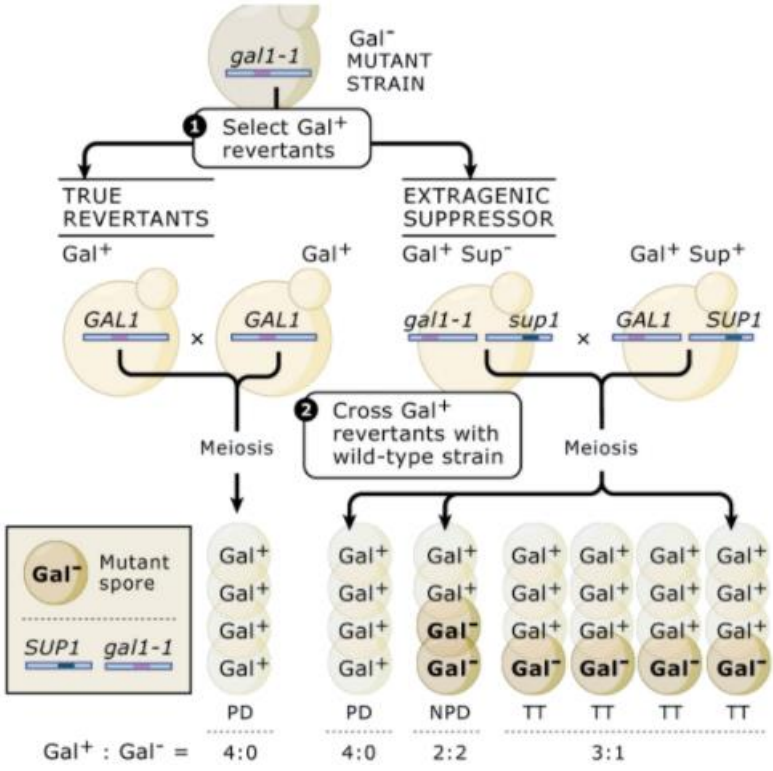
Gene interactions

Complementing a recessive phenotype: Example of galactose: Let there be a mutation in the GAL gene, such that yeast cells cannot grow on a medium with galactose. Which gene is affected? Make use of library DNA (contains plasmids with up to 3 genes per plasmid): Let a bacteria amplify the plasmid first and use a selectable marker such as ura3 mutation (cells will be auxotrophic to uracil). Also, yeast cells on average take up one plasmid per transformation. Selection is done by mediums lacking uracil, so those cells with the URA3 will biosynthesize uracil.

In our example, yeast cells cannot grow on galactose lacking mediums, so we make use of plating and screen for those that were able to grow on the medium nonetheless (since they acquired a functional copy from one of the plasmids). Then, sequence the plasmid. As a last step to confirm our findings, make new plasmids with the single genes of the original plasmid on it (normally, a plasmid has around 2-3 genes on it, don't forget that S.cerevisiae has around 6600 genes), search the yeast genome in a database to identify the exact genomic fragment and select again to identify the single gene plasmid that was viable.

Some auxotrophy markers: HIS3 (histidine prototrophy), LEU2, URA3, TRP1 – all these genes allow to perform biosynthesis of the respective molecule.

Epistasis and suppression analysis: A second mutation can be induced to restore the wild type phenotype. Did the second mutation occur in the same gene or in another gene that was epistatic to the first gene? In yeast, we can answer this via tetrad analysis:
True revertants ⇔ second mutation occurred in the same gene ⇔ every progeny is PD (4:0)
Extragenic suppressor ⇔ second mutation occurred in another gene ⇔ progeny is PD, NPD and TT



In rare cases, double mutants behave like wild types, although the two mutations are in the same gene. Thus, the two mutations complement each other. This situation is called intragenic complementation or allelic complementation. When observed, this situation 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.

Difference in up- and downstream in biosynthetic and regulatory pathways:

For a biosynthetic/metabolic pathway, the "upstream" gene will be epistatic to the downstream gene. This is because the intermediate upstream of the first gene/enzyme accumulates, but the blockade at this point will eliminate everything downstream.

For a regulatory pathway the order is reversed: a downstream mutant usually determines the phenotype, because any components upstream of it become irrelevant. Thus, the downstream gene will be epistatic over the upstream gene.

Non-Mendelian inheritance

Non-mendelian inheritance occurs in epigenetics and in cytosplasmic structures, such as mitochondria, plastides of plants and in plasmids for example.

If we have a gene inheritance pattern that is 4:0 or 0:4, then this is a strong indication for genes outside of the nucleus (for example, mtDNA inheritance).

Kor.: Yeast can grow on fermentable carbon sources such as glucose. Thus, it does need a functional mitochondrion, since it can do fermentation to satisfy its ATP needs.

3.10.2017

Question: Screening for karyogamy (yeast) mutants: Yeasts do not fuse their nuclei, therefore they have to distinct nuclei in one cell body.
Prerequisites: have a recessive selectable mutation (for example mutate CAN1 gene (such that the mutant has no functional CAN1) – makes them resistant to canavanine, but cannot take up arganine anymore either (still has the biosynthesis genes for arganine)).
Moreover, we need a cytoplasmic marker: make use of mitochonodrial genetics, designated rho (rho+ - can do respiration; rho- - cannot do respiration but can live; rho0 – lost their genome and cannot survive). Rho- and rho0 cannot use ethanol and glycerol as their only carbon source therefore, but rho+ can use those moleclues as a carbon source.

Setup: yeast a is CAN1-sensitive (dies) and rho+ (grows on glycerol). Alpha-yeast is CAN1-resistant and rho- (cannot grow on glycerol). Crossing them leads to a strain with the following features: it is a/alpha, CAN1-sensitive/CAN1-resistant and rho+ (shares same phenotype as yeast a).

The karyogamy mutant will generate new buds (alpha, CAN1-resistant and it will have the rho+ mtDNA, because cytoplasm is mixed nonetheless). The buds (haploid) will be canavanine resistant and will be able to grow on glycerol unlike the alpha yeast.

Mutagenesis: Mutate the alpha yeast, because the alpha yeast bud in the karyogamy mutant grows and it needs the CAN1-resistance to grow in the first place.

High-level lecture

On ade2 mutants (mutation in ade2-108): How can we retain the white phenotype when it is an ade2 mutant (turns red due to oxidation of AIR). Two possibilities: rho0 double mutants or translational machinery double mutant.

Objective: We try to identify a double mutant that can read through the premature stop codon and still synthesize adenine even with ade2-108 mutation (this is a suppression analysis).

Proposition for the approach: Cross ade2-108 mutX with WT rho+ ➔ it's offspring will then be ade2-108 AND rho+. Initially, also feed the ade2-108 mutants with EMS to get a second mutation.
Another way: Put the cells on ethanol/glycerol plate and we will immediately see which one are the mutants we want (ade2-108), since it could possibly be that there is a rho0 that gives the same phenotype.

Next: delete ade2 in the mutants altogether such that they are ade2-delta AND ura3-delta. Also, put those deleted genes on a plasmid. (5FOA can be taken up by ura3 wild types, which kills them (similar idea as in CAN1 mutants), so ura3 mutants do not die on 5FOA, but are auxotrophic for uracil).

Also, make use of adenine antibodies so as to identify biosynthesized adenine despite ade2-108 mutation.

Our mutation is SUP35 (termination factor). This can be used manipulate termination in yeast (and also in other organisms that have SUP35 or homologs, probably).

Ade2-108 is recessive so a diploid with a wild type will be white simply (to 99%). After the formation of tetrads we expect 2 reds and 2 whites, but we get 4 whites, so therefore it is non-mendelian. From above, we know it cannot be a mitochondrial mutation (all are rho+).

We know mutations occur with a rate of about 10**-7. But there is another factor that occurs with a rate of 10**-5 (100 times more probable). This unknown factor is dominant and inherited by all spores, therefore non-mendelian and not on a chromosome.

Identification: Make use of the previously discussed karyogamy mutants – cross it with the mutant (:= PSI+): PSI+ with CAN1-resistant rho- and karyogamy ade2-108 mutant. The combined cell has rho+ from PSI+ and is white, but it is not diploiad (only n+n). The budding offspring will be alpha rho+ and PSI+ (PSI+ is definitely in the cytoplasm).

Then, take a PSI- cell to perform **cytoplasmic transfection (electroporation)**: Drop some PSI+ in the PSI- cell so that we can **identify where it is DNA, RNA or protein**. Then, use DNAases, RNAases and proteinases to identify the object. In the actual experiment, none of them actually worked.
Assumption: A huge protein that proteinases have troubles with.
Approach: Turn the protein complex in its primary structure and use proteinases. The whole cell was rendered PSI-, therefore it was a protein affected phenotype.

We deduce: the PSI+ has **prions**: They affect active proteins and turn them into prion conformation. Only very little of the prion form is needed so that all the active proteins change into the prion form. Prions also form those amyloid fibers which are resistant proteases.

Structure: PSI- ➔ overexpress SUP35 ➔ PSI+. SUP35 structure: it has a termination and a pretermination region. The pretermination has a QN rich domain, that is charged right before the termination region starts.
Overexpression: Either the SUP35-N or SUP35-C: The N case leads to the PSI+ phenotype and the C case leads to PSI-, so we make a SUP35-deltaN and cross it with PSI+. Tetrad analysis: 2 spores are PSI- ➔SUP35-deltaN and 2 spores are PSI+ ➔ sup35-C.

We can thereofre convert SUP35 to SUP35-PSI+ (N terminal) and this is a prion (keep in mind that prions are autocatalytic – only need to "touch" the normal protein).

This is why translation termination is not so effective anymore.

Short overview of mutants in yeast

Auxotrophies in yeast: ura3, his1, ade3 or ade2-108, trp1, leu2, cys3-1, pro3-delta2, gal1. Yeast rely on an external source. Some amino acid imports also import toxins like: Canavanine by CAN1 (and arginine), 5FOA by ura3 wild types (no death if uracil cannot be imported, therefore they are auxotrophic to uracil).

Further structural mutations in yeast: rho0 and rho-. In rho0, there is no mitochondrion anymore, since all mtDNA has been lost. No glycolysis possible, but fermentation still works. In rho-, glycolysis of mitochondrium is disturbed, but yeast still lives. Both cannot grow on neither ethanol nor glycerol as carbon source which can be used as screening criterium. Note that mtDNA is cytoplasmic, so segregation pattern will always be 4:0 or 0:4 (non-Mendelian). Useful for karyogamy mutants, since cytoplasmic genetic information and structural objects are inherited, but genomic information remains distinctly separated in two nuclei (yeast is n+n).

Karyogamy related mutants (cell cycle control):

Cdc28: no cell cycle entry and no bud.

Cdc24 and cdc42: Mitosis occurred, but a large bud on the mother yeast remains (n+n).

Cdc3/10/11/12: cytokinesis mutants.

HO gene mutant: yeast is heterothallic and can no longer change its mating type by DSB and HDR of the opposite mating type locus.

Protein expression mutants: Cycloheximide binds to ribosomes and inhibits protein synthesis. CYH1R homozygous (since gene recessive) recover wt phenotype (CYH1S is dominant and will therefore not affect ribosomal translation inhibition by cycloheximide).

Calculation tricks: if PD = NPD, then unlinked genes. For tetrad analysis if TT != 0: 0.5*TT/total number of tetrads = frequency = map distance in cM. Same formula can be used in centromere linkage if and only if NPD = 0. Else NPD > 0: linkage = 0.5 * (TT + 6*NPD)/total number of tetrads.

Genes close to centromeres will rather rarely undergo recombination (which means that they increase the likelihood of PD in tetrad analysis). Moreover, if PD = NPD then two genes are unlinked (unlinked implies the genes are on different chromosomes. Example: trp1 is near centromere (1cM) on Chr. 4 and

met14 is near centromere on Chr.9 – both are unlinked). This principle is used with genetic markers when one tries to identify a mutation (mutation always inherited with marker: closely linked, but PD=NPD then both are close to centromere). Using markers on different chromosomes we can also identify the chromosome by crossing the mutant with a wild type from a different genetic background (in yeast: USA and CH, in C. elegans: Bristol and Hawaii) (=> SNPs, maybe CNV etc.). Mutation will most likely be on the chromosome which has the least foreign DNA material in its genome (count number of local strains like Bristol where the mutation occurred).

Forward Genetics

On C. Elegans

It has 5 autosomes and one sex chromosome, around 20k genes distributed in 100m bp (high gene density, since its genome is 20 times smaller than that of humans – has 35% homologs to human genes).

Unusual features of C. Elegans: Males have only one sex chromosome (X0), while hermaphrodites have XX. Hermaphrodites are self-fertile females that can do internal fertilization (basically cloning). Males are generated during non-disjunction of chromosome X during meiosis 1. They are needed to ensure genetic variability.

Def. eutely: An adult of a species has a genetically determined number of cells. In the case of C. Elegans, males have 1031 somatic cells and hermaphrodites have 959 somatic cells. In their case, the developmental fate of every single cell has been mapped, so that one can predict the differentiation pattern of every single cell during embryogenesis.

On Drosophila

Drosophila has only four pairs chromosomes, three autosomes (2, 3, and 4) and one pair of sex chromosomes (XX in females, XY in males). Chromosome number 4 is very small and contains only a few genes. The Drosophila genome is about 140 million base pairs long. It is estimated that 50% of the 15’000 genes have human homologs.

Balancer chromosomes: Special kind of chromosomes used for screening homozygous recessive lethal mutations (often during embryo development) in Drosophila. They contain a dominant marker and a recessive lethal mutation. Also, some DNA regions are inverted, others are duplicated and the genes can be differently arranged. They are crippled such so that no recombination can occur with the “normal” chromosome – the crippled nature of the balancer chromosome makes it no longer really homologous to the wild type chromosome.

The Heidelberg screen cannot identify mutations in maternal genes, genes affecting the patterning of internal structures such as the nervous system, tissue-specific genes and mutations in small genes such as miRNAs (the larger the gene the higher the probability of inducing a mutation in the gene).

Drawbacks of F2 screens: First, they are time-consuming and labor-intensive, since several generations are required for recessive mutations to show their effect. Second, they do not allow to identify the functions of essential genes during later stages of development because of lethality.

Forward genetic methods to screen F1 populations (clonal screens using Flp/FRT recombinase)

Questions for moodle: “How are the Flp sites flanked? Where is the second flank? One is in a promoter region, but where will the second one be located? Since we are often interested in transversion and no deletion or inversion, the second Flp site should be on another chromosome arm, but can it be “just somewhere, where the “target” gene is?”

What is the prerequisite for homologous recombination to occur in mitotic cells?

Answer: The chromosomes must be paired during mitosis. This often occurs in Drosophila and far less so in other animal models such as the mouse.

The Flp/FRT system allows to make homozygous mutants tissue-specific. Effectively, the Flp site will be placed in a regulatory region of the genome, for example in the eyeless-promoter region where Flp is expressed if and only if the eyeless promoter is bound to.

From yeast genetics, we know that mutations in the Minute gene lead to recessive lethality and in the heterozygous case to dominant growth inhibition. In Drosophila, we can thus use the Flp/FRT recombination system to kill descrete cells (do not forget that there is only one copy of Flp/FRT and after recombination and mitosis, there can be a normal (or sometimes called “twin”) daughter and one homozygous Flp/FRT daughter cell).

Modifiers: Enhancers and Suppressors

Those are mutations in a second gene that can either worsen or milden the phenotype of a single mutation. Enhancers generally make the phenotype more extreme with lethality being the most extreme case.

Def. amorph: The strongest type of mutation that can be generated such as a nonsense mutation (deletion). Usually recessive, but not a rule.

Def. hypomorph: Not a complete abolishment of the gene activity, the phenotype shows a reduced appearance. Usually recessive, but not a rule.

Def. hypermorph: Overexpression of a gene, more extreme phenotype.

Def. antimorph: Dominant negative mutation on the wild type allele.

Def. neomorph: A mutation that leads to a new function of the gene.

Def. penetrance: Per centage of a population that expresses a certain phenotype, given the genotype (especially in dominant alleles). It implies that a phenotype does not necessarily have to expressed even if the genotype is present. This is **non-Mendelian** even if the genotype inheritance follows a Mendelian pattern.

Kor.: Apart from the penetrance, the expressivity of a phenotype can vary heavily across different individuals even if their genotypes are all the same. It implies that a phenotype is continuous or discrete on a scale and non-binary. This trait is often referred to as the intensity.

Reconsidering gene identification using novel mapping strategies

When genes are located on different chromsomes, their recombination frequency in the offspring will typically be 50:50. The same holds true for genes on the same chromosome that are far apart from one another, because crossover events are frequent enough to occur.

When recombination frequency is below 50%, then linkage is indicated (approaching 0% when genes are immeasurably close to each other meaning that they are always intherited together as a haplotype).

In what cases is the probability of a crossover between two linked genes NOT proportional to the distance between them?

Answer: An exception is if two genes are at or close to the centromere. Also, if the distance between two genes is more than 30 m.u., the distance will be underestimated because the frequency of double crossovers is increased.

Genetic distances: Recombination frequency = Nr. of recombinants/total number of flies x 100 = Map distance. The unit of distance are called map units (m.u.) or centimorgans (cM).

Using genetic markers to pinpoint the location of a mutation: Let C be a chromosome with known locations O1-5 on C (wild type) and M1-5 (mutant), where M and O are different alleles. Cross a wild type with a mutant and analyze their progeny to pinpoint location of the mutation: C will undergo meiosis and the chromosomes of the progeny will have M and O alleles. Sequence C of the progeny (especially those of the mutants) and identify the common mutant allele: The mutation can either lie in the allele itself or be in its close neighbourhood (it can be inherited as a haplotype).

Def. integrated genome map: Physical genome map + recombination map. The physical map shows a gene's possible action on the cellular level while the recombination map reveals the effect of the gene on the phenotypic level. Physical map resolution is on a discrete DNA base (several bases or in kilobases) level and output is the genotype. Recombination maps are measured in map units or cM and are continous values based on a recombination frequency.

In which regions of the genome are the gene distances shown on physical and on recombination maps out of proportion?

Answer: In regions where recombination occurs more frequently (recombination hot spots) and less frequently (cold spots) than on average. Examples are centromeric regions: here, recombination is lower. Hot spots are for example CNVs, e.g., trinucleotid repeats cause "fragile sites" within the DNA sequence that are more prone to recombination. Chromosome ends are also hotspots for recombination (it is unclear why).

Def. maternal effect gene: Genes expressed in the mother, but not in the oocyte. Those proteins are then imported into the zygote (basically additional genes for the offspring coming from the mother). There is also the paternal effect which is extremely rare though.

23.10.2017

High-level lecture

Def. incomplete penetrance: Individual has mutant alleles, but exhibits wild type phenotype.

Quantitive trait loci (=: QTL) mapping

Due to noise, a binary genotype can have a continous distribution as its phenotype (organisms are no perfect computer machines).

Bem.: genotype + environmental factors ➔ phenotype.

Def. Quantitative trait locus: A region of the genome containing one or more genes that affect variation in a quantitative trait, which is identified by its linkage to polymorphic marker loci.

Recombinant inbred lines (=: RILs) can be used to study quantitive phenotypes (basically, translocate two individuals from nature to the lab and breed them and then keep interbreeding and anaylze the phenotype at all stages).

A QTL never determines the entire degree of variation. Some QTLs may even show epistatic interactions between each other.

QTL mapping: do the values show a normal distribution? Do statstical tests and find p (e.g. use beta and r values from a linear regression analysis). Calculate LOD score (likelihood of disequilibrium):
LOD = -log(p). Repeat this for many markers distributed across the genomes. Plot LOD score of the markers along chromosome positions.

Reverse Genetics

If Sanger sequencing yields “5'-TAGATCCGA-3'”, then the template sequence would be just the complementary sequence: 3'-ATCTAGGCT-5'.

The basic biochemical process of SBS is relatively simple (figure 1-6a). Template molecules are attached to a solid surface (called a flow cell), a primer is annealed to the template strand and a polymerase is bound. The individual cycle of the sequencing by synthesis reaction then proceeds as follows.

1. The flow cell is flooded with a solution containing the four different types of fluorescently labeled, chain-terminating nucleotides.
2. The polymerase uses the template strand to extend the primer with the complementary nucleotide. The terminating group on the newly incorporated nucleotide prevents the incorporation of additional nucleotides.
3. Any non-incorporated nucleotides are washed away.
4. The fluorescence signal of the sample is measured and reveals, which of the four nucleotides has just been incorporated.
5. Both the fluorophore and the terminating group are removed by chemical cleavage and washed away. This prepares the reaction for the next cycle.

Without this synchrony, the fluorescent signal from the different molecules in a cluster would be out of register and become uninterpretable. The SBS reaction therefore had to be designed such that each reaction is prevented from progressing until all molecules in the flow cell have completed that reaction step. This requires i) a substantial waiting time at each reaction step to ensure all molecules have completed the reaction and ii) the time-consuming flushing in and washing out of the reagents that is needed for each individual step in the reaction cycle.

Another problem is that there is a finite chance of a molecule skipping the reaction: Initially this is not a problem, because the signal from the other molecules in the cluster will overpower the signal from those few molecules that have fallen out of lockstep. But, since there is no way for the molecules that have fallen out of lockstep to get back into lockstep, the fraction of out-of-lockstep molecules increases with every cycle. Eventually, after 200-300 reaction cycles so many molecules of the cluster will be out of lockstep that their random fluorescence signal drowns out the signal from those molecules that have remained in lockstep.

3rd generation sequencing:

PacBio: SMRT

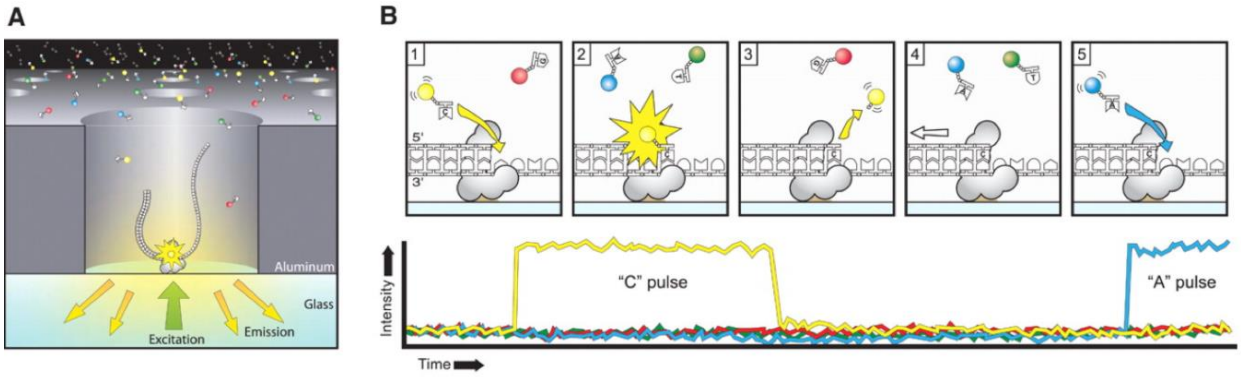


Figure 1-7 Principle of the PacBio single-molecule real-time sequencing technology. (A) Individual DNA polymerase molecules are immobilized at the bottom of 100 nm diameter nanowells etched into the bottom of the reaction chamber. **(B)** Individual fluorescently labeled nucleotides diffuse in and out of these nanowells and may be incorporated into the growing strand when they are complementary to the next unpaired nucleotide on the template strand. During the time between successful binding and the formation of the phosphodiester bond, the gamma phosphate-linked fluorophore is excited and emits an optical signal that is picked up by the instrument's optics. The color of the fluorescence (trace bottom right) reveals which base is incorporated. (from Eid *et al., Science,* 2016)

The challenge for such an approach is that the signal that can be obtained from a single fluorophore molecule is very limited. Also the fluorescence noise in the sample is substantial, in particular given that the reaction mix contains a large amount of unincorporated fluorescently labeled nucleotides.

Sequence assembly: Map-based assembly and shotgun sequencing.

Transgenic animals

How can you test whether your engineered fusion-protein faithfully reflects the localization and action of the wild-type protein?

Answer: You can cross animals expressing the fusion protein with animals that lack the function of the protein of interest (a mutant). Assuming that your mutant has a specific phenotype, if the fusion protein can rescue the mutant phenotype, you can be pretty sure that the fusion protein localizes and functions like the wild-type protein.

Electroporation (brief electric shock on cell membrane of bacteria, yeast or tissue culture cells) is used to add transgenes, since the membrane becomes permeable for a short duration. Plants are **bombarded** with the transgene that go through the cell wall with a special modified gun. **Injection** is done for mammal cells. Both in plant cells and in mammal cells, we can add transgenes via viral vectors.

The **gene gun method** is used as an alternative to electroporation in non plant cells and as default in plants (and all other organisms with a cell wall). One coats a heavy metal such as gold or tungsten with DNA and then blasts them onto a cell. The nucleus can be targeted but also organelles and even plastids to inject transgenic DNA. Cells can be labelled like this as well.

For what reason may integrated arrays be preferred to extra-chromosomal arrays?

Answer: The integrated array allows for stable inheritance of the transgenic construct; this is important to establish a transgenic line that can be used for further experiments.

As an example, let's look at a cross using a balancer for the second chromosome that you have already encountered before. The balancer carries the mutation Cy, which confers bent up wings, and the X-chromosomal cis-homozygous mutation w-, which confers white eyes. For the flies in G1, we can't yet tell the genotype. From the phenotype we know that they contain the balancer and the transgene (TG); however, these flies could have different genotypes:

If the TG is on the second chromosome, these flies would have the genotype TG / Cy.

If the TG is NOT on the second chromosome, these flies would have the genotype + / Cy; TG (the TG is now on another chromosome, with the ; separating the different chromosomes)

Therefore, we need another cross to determine where the TG is inserted. We therefore cross the red-eyed Cy flies with balancer flies (white eyes, Cy). For the progeny of this cross, there are two possibilities:

Three phenotypes occur if the TG inserted on the 2nd chromosome; four phenotypes occur if the TG inserted on the 3rd chromosome. Thus, if white-eyed, non-curly flies occur in the progeny, we know that the TG is not on the 2nd chromosome, because in all other flies, either the TG or the balancer are present. If the Cy marker and the transgene segregate independently, they are on different chromosomes.

Reverse genetics in invertebrates

Def. retrotransposon: Basically, a part of the non-coding region in the genome. A retrotransposon is first transcribed into mRNA and then reverse transcribed back to DNA and inserted in the genome again. Thus, the retrotransposon keeps its original position and it has copied itself into a new location (cf. "normal" transposons).

Kor.: There are two types of retrotransposons: Long terminal repeat transposons (=: LTR transposons) and non-LTR transposons. LTR transposons possess two long terminal sequences that encode the transposase and integrase. Non-LTR transposons lack the integrase, but they encode and endonuclease.

Kor.: In drosophila, we call them P elements. The autonomous version of the P element encodes the so called P transposase itself, while in the non-autonomous case, the transposase has been deleted from the P element, but it can still relocate if P transposase is supplied externally.

Kor.: EP elements are special P elements for overexpression in drosophila. They contain an UAS sequence that can be activated by the transcription factor Gal4.

Def. enhancer trap: transposable element + reporter gene + original of replication + screening gene (e.g. ampicillin resistance).

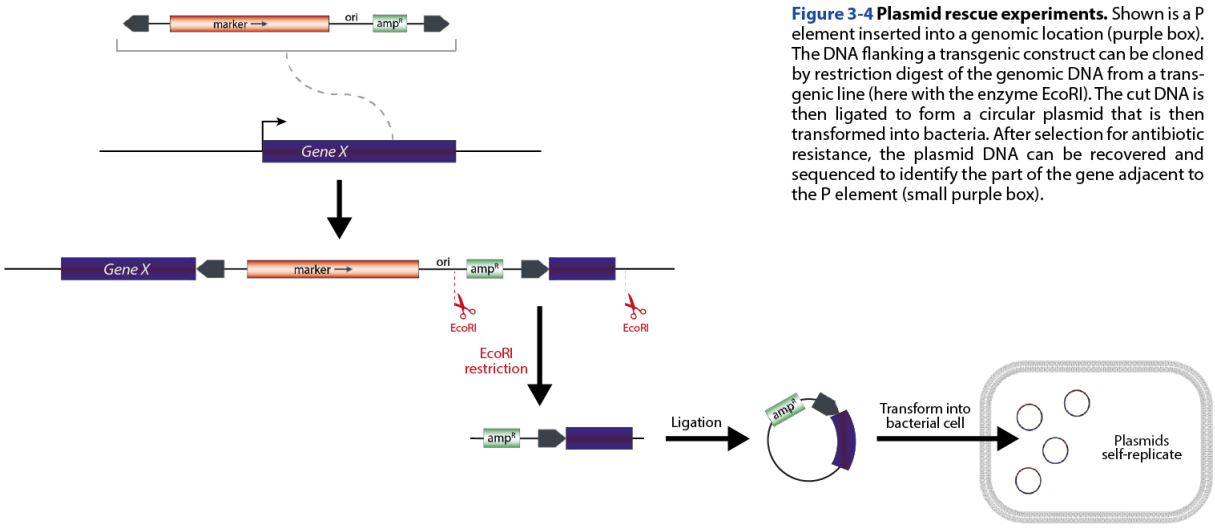


Figure 3-4 Plasmid rescue experiments. Shown is a P element inserted into a genomic location (purple box). The DNA flanking a transgenic construct can be cloned by restriction digest of the genomic DNA from a transgenic line (here with the enzyme EcoRI). The cut DNA is then ligated to form a circular plasmid that is then transformed into bacteria. After selection for antibiotic resistance, the plasmid DNA can be recovered and sequenced to identify the part of the gene adjacent to the P element (small purple box).

What experiment could you perform to detect on a molecular level whether a P element has been excised or not, and what kind of deletion has been induced by the excision of the P element?

Answer: Since you know the original position where the P element was inserted, you can design primers that flank the P element: the forward primer would bind in the genomic region that is 5' from the P element insertion and the reverse primer would bind 3' of the P element insertion, in the gene. Then, a PCR is performed on DNA extracted from flies where the transposon has been mobilized. If the P element has not been excised, the PCR product will include the P element sequence and thus be big (let's say 3 kb). If the P element has been excised, but the DNA has not been deleted, this will give a fragment that is smaller and includes only the genomic sequence of this region. If a deletion has occurred and this deletion has removed part of the gene, then the PCR product would be even smaller, or- in the extreme case if the deletion has removed the part where the reverse primer would bind- no PCR product would be obtained.

What problems can you think of to be associated with gene silencing using RNAi?

Answer: The biggest problem are off-target effects, because RNAs are small (20-25 nucleotides) and can bind also other genes than the desired one, leading to the downregulation of a non-related gene. Another problem is that some genes are not effectively downregulated by RNAi. Also, in some tissues, like the brain, RNAi does not work efficiently.

Def. Ectopic insertion of a transgene: Microinject the transgene into oocytes and add these oocytes into a foster mother. Hope for the transgene to be taken up by the oocytes in order to produce transgenic mice.

A much more controlled way is the use of transposons. In vertebrates, transposons are normally dormant. The Sleeping Beauty (=: SB) transposon is a system rescued from fish that can be used in vertebrates. SB can incise itself in any TA site in a chromosome (the human genome has 200m TA sites).

It is realized by manipulating ES cells (mouse) and then add them in a blastocyst (chimeric mouse produced). The chimeric mouse will be heterozygous for the transgene, therefore cross it with another heterozygous mouse to produce mice homozygous for the transgene.

The positive-negative method for selection

How can we ensure, transgenic DNA has been successfully implanted into a chromosome of a mammal such as the mouse? In between the homologous sequences of the gene to be deleted, we need a positive marker, such as resistance to a certain antibiotic. Outside of the homologous sequence, there will be a negative marker that makes the organism sensitive to a substance, such that it dies in its presence. Therefore, when random insertion occurs, but no homologous recombination, then the negative marker remains in the genome and cells in a suitable medium will die (e.g. thymidine kinase gene (=: tk gene) and the guanine analog ganciclovir – ganciclovir is phosphorylated by the tk gene, which causes premature DNA chain termination and apoptosis). When homologous recombination takes place, the negative marker is lost and only the positive marker will remain, because it has been integrated permamently into the genome.

Genome editing

Method 1: Zinc-finger nucleases
Method 2: Transcription activator-like effector nucleases
Method 3: CRISPR/Cas9

Comparison: CRISPR/Cas9 vs. RNAi

Common points:

- Both are mediated by small noncoding RNAs with a target specificity of roughly 20 nucleotides
- Both work in combinations with a ribo-nucleoprotein complex to target specific nucleic acids sequences
- Both originate from defense mechanisms against foreign DNA

Differences:

- Technological aspects:
    - RNAi: fast, easy to use (transfection, feeding, etc.), suitable for large-scale genetic screens
    - CRISPR: labor intensive, requires transgenic animals
- Molecular consequences:
    - RNAi: Transient gene downregulation, does not change genome sequence
    - CRISPR: permanent gene knockout, changes genome sequence
- Mode and place of action:
    - RNAi: in the cytoplasms, target nucleic acid is mRNA
    - CRISPR: in the nucleus, target dsDNA
- Genome functions that can be assessed:
    - RNAi: Protein coding sequences
    - CRISPR: Protein coding, Introns, Promoter

- Origin:
    - RNAi: is an endogenous eukaryotic pathway that uses cellular machinery (RISC)
    - CRISPR: derives from a bacterial defense mechanism

6.11.2017

In a transcrptional reporter gene, we not only know whether a gene is transcribed, but also where in the cell. A translational reporter gene is fused with the target gene and might not be visible if the protein is degraded.

Small non-coding RNAs

Difference between miRNA and siRNA:

miRNAs are excised by Dicer from their short (roughly 70 nucleotide long in mammals) and imperfect stem loop precursor as a single small RNA species, which may accumulate to tremendous levels in the cell (up to 50,000 molecules/cell).

siRNAs, on the other hand are produced by consecutive cuts by Dicer along a perfect, long dsRNA molecule. Therefore, siRNAs, unlike miRNAs, are always part of a population that is distributed along their long dsRNA precursor.

miRNAs in C. Elegans

lin-4 inhibits lin-14. Loss-of-function in lin-4: Needed for appropriate transition in larval stages (L1 to L2). When ko'ed there are inappropriate cell-division pattern recapitulations at later stages. A mutation in the protein coding region of lin-14 can compensate the lin-4 phenotype; a new phenotype occurs, where the L1 stage is skipped.

A mutation in let-7 causes the reappearance of larval cell fates during adult development. A mutation in lin-41 can partially suppress the let-7 phenotype.

Let-7 and lin-4 are both miRNAs to their RNA targets. They typically bind in the 3' UTR of their targets.

Translational inhibition by miRNA is favored over degradation for a quick and flexible response to stress by organisms. When the stress is over, translation can be resumed again without producing mRNA again, which takes some time to recover the mRNA pool.

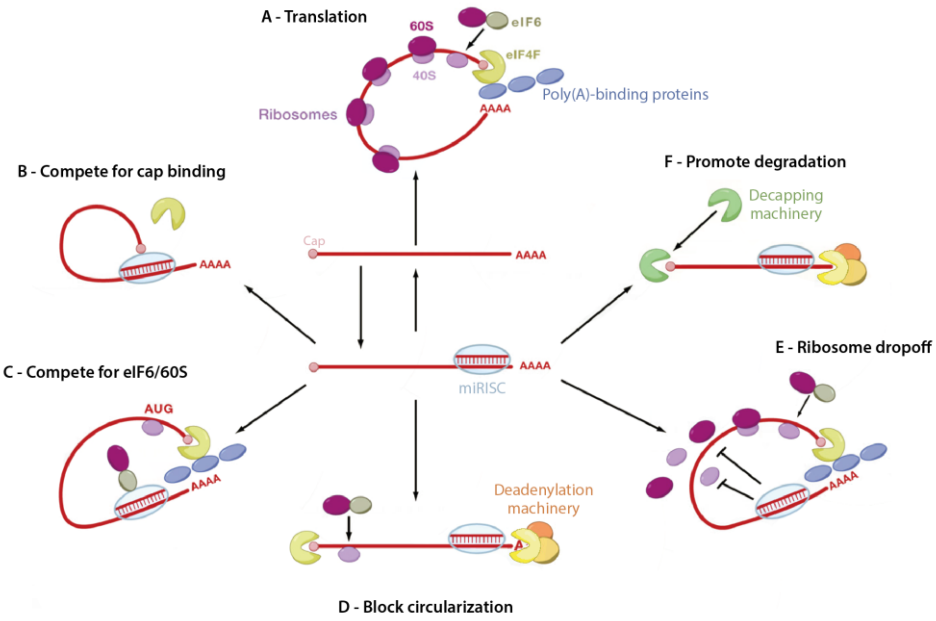


Figure 1-10 Possible mechanisms of translational suppression by miRNA-containing RISCs (miRISCs). (A) Non-repressed mRNAs recruit initiation factors and ribosomal subunits and form circularized structures that enhance translation. **(B)** When miRISCs bind to mRNAs, they can repress initiation at the cap-recognition stage or **(C)** the 60S-recruitment stage. **(D)** miRISCs can induce deadenylation of the mRNA and thereby inhibit circularization of the mRNA. **(E)** miRISCs can also repress a post-initiation stage of translation by inducing ribosomes to drop off prematurely. **(F)** Finally, miRISCs can promote mRNA degradation by inducing deadenylation followed by decapping. (Adapted from R.W. Carthew and E.J. Sontheimer, *Cell*, 2009)

Comparison: animal miRNA vs. plant miRNA production

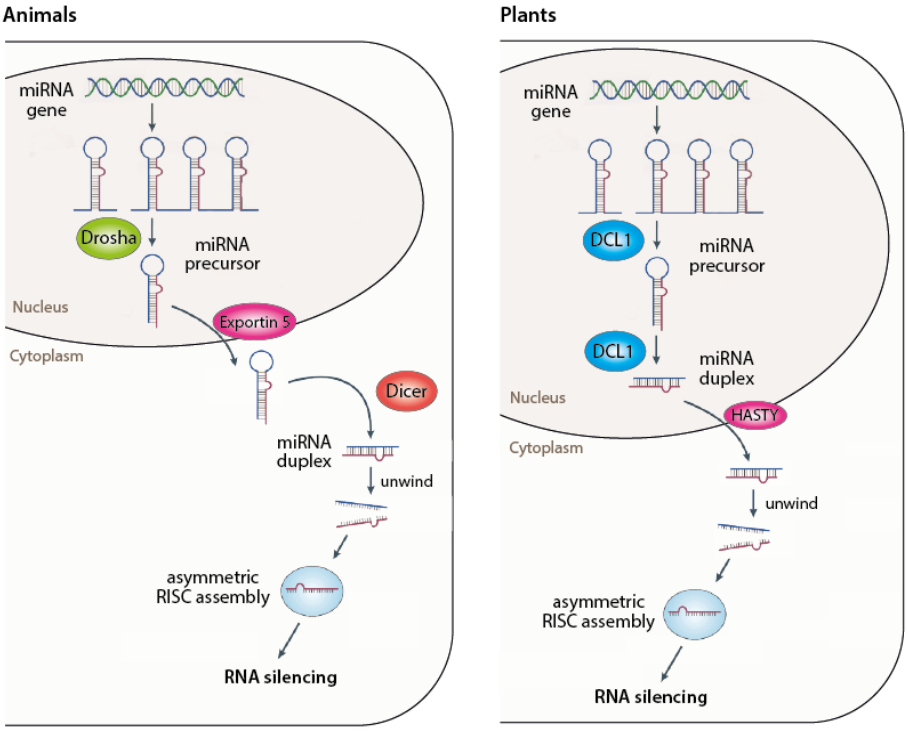


Figure 1-7 Biogenesis and RISC assembly in animals (left) and plants (right). (Adapted from L. He and G.J. Hannon, *Nature*, 2004 and E.J. Chapman and J.C. Carrington, *Nat. Rev. Genet.*, 2007)

Questions

If 99% of animal miRNAs display imperfect mismatches to their target in order to avoid slicing, why have the catalytic residues of their corresponding AGO proteins been conserved in evolution?

Answer: Perhaps the 1% sliced miRNA targets is very important for growth and development, or, alternatively, maybe the slicer residues of AGOs are used for other processes. Indeed, the slicer activity of AGO was shown to process precurser miRNAs without the requirement of Dicer. This is a rare event, however, it was shown to be important during early embryogenesis in zebrafish.

Can you think of a specific example for an alteration that a) changes miRNA function on a global level or b) affects a small number of miRNAs?

Answer:

Global changes: loss-of function mutations in Pol-II, Drosha, Dicer, Exportin-5, AGO; duplication of a dicer gene, gain-of-function mutation in Drosha or Dicer that enhances enzymatic activity.

Specific changes: duplication of a miRNA gene, mutation of a miRNA gene that disturbs hairpin structure formation, mutation in seed sequence disturbs interaction with target, mutation of miRNA target sites.

4.12.2017

Human diseases

Introduction and approaches to study human genetic diseases

Genetics allows us to understand diseases on a molecular basis.

Symptoms are observed by the patient, are subjective and cannot be measured directly. Signs are also subjective but they can be observed by a third party. If a phenotype is observed by the patient itself, it is called a symptom.

Microsatellites are repeats of di-, tri- or tetranucleotides. There are 50k such sequences in the genome and can be detected by the PCR. Also, they are nearly trivial to amplfily. Minisatellites are repeats of 10 to 60 bp. They are not normally amplified by the PCR.

Chromosomal abnormalities are linked to spontaneous fetal abortions.

Identifying the genetic basis of the disease

Recessive-autosomal: 3:1 (both parents in the beginning were heterozygous).

X-linked domaninance: 1:2 (mutated allele on mother). Ex.: hypophosphatemic rickets (low phosphate concentrations in blood (mutation in PHEX gene). PHEX also regulates FGF23. Mutations lead to short stature and bone deformities).

The genetic disease onset can occur later on in life, which is why in a statistical analysis of diseases mapping (LOD score specifically), there can be an individual with a disease genotype, but the phenotype is not present (for example when the indivudla is still very young. Or in adults, other genes can compensate for the risk allele or extreme environmental compensation can occur (which means that the disease penetrance is not 100%)).

Showing that disease gene has been found indeed

When a candidate has been identified in the human, we can look at the mutation in a model organism and see if similar phenotypes occur. Also the mutation must not occur in a random control population (could be just a SNP).
Biochemically, compare mutated protein to normal protein and compare activity to derive new conclusions.

Forward genetics: clinical symptom – biochemical stem identification: show protein involved in affected subjects. Find gene of protein.
Reverse genetics: clinical symptom – identification of allelic variant: identify gene, show that variants (mutations) renders protein function and abnormal pathway affected in subjects. Find protein from gene and do positional cloning: strategy to identify gene of unknown fuction based on genomic mapping.

05.12.2017

Monogenic vs polygenic diseases

3 types of diseases (traditional): genetically determined (high penetrance), environmentally determined (infections), genetically and environmentally determined (metabolic diseases, psychological diseases, developmental form of congenital heart diseases, cancer (breast, ovarian, colon)).

Key concepts of complex diseases: familial concentration of disease without specific pattern of inheritance; absence of clear biochemical defect resulting from single abnormal gene; considerable variation in severity and expression of phenotype (between and within families); most affected individuals have unaffected parents; sex differences and ethnic/racial differences.

Linkage studies

Large multigenerational single family (or multiple smaller families in clear homogeneous disease); defined mode of inheritance; single locus responsible; known penetrance; genetic homogeneity.

These things are obviously not the case in complex diseases. Available methods: case control studies and TDT.

Transmission disequilibrium testing: tests for distortion in transmission of alleles from a heterozygous parent to an affected offspring. Under no association with the diseasem alleles A and B have an equal chance of being transmitted from a heterozygous parent. If however allele B increases the rik of disease, this allele will be preferntially transmitted to the affected offspring.

Addendum: relevant GGB chapters

Bacterial genetics

A special type or recombination in DNA is the jumping of genes to other parts of the DNA. This process is called transposition, its enzymes are transposases which are responsible for the movement of the transposons and they work independently of the rest of the DNA. The genes can jump to non-homologous sequences. This process has been especially well examined in bacteria, where the transposon typically looks like a phage or plasmid.

Transposons are in the host cell's genome edited. The transposons DNA possesses the DNA sequence for the necessary tools to relocate in the genome (cutting and editing). One can take advantage of this function to add genes, so that they are carried over as well.

Since too many transpositions would eventually destroy the host cell completely, the transposons have developed very subtle mechanisms so that a transposition occurs every 10**3 to 10**8 depending on the type of the transposition. (A transposon might jump into a gene and inactivate it.)

Bacteria as model systems:

Bacteria are haploid, they have a short generation duration, they reproduce asexually, huge populations can be easily cultivated.

Transposon mutagenesis:

Transposons jump randomly into a gene of a bacterium often leading to its inactivation. That way, one can observe its effects on survival and reproducibility. Since the transposons is known, one can easily figure out the inactivated gene with PCR or NGS-methods. Normally, a transposon contains further marker genes, e.g. the immunity to an antibiotic so that the affected bacteria can be identified more easily.

The following criteria define a good transposon for mutagenesis:

1) Transposition occurs with a high frequency for more mutations.
2) The transposon has a low selectivity, such that all genes are affected about equally.
3) Transposon contains a selectable gene, so that the affected bacteria can be selected more easily.
4) The transposon can induce transposition in many different bacteria kinds, so it can be used and observed in different bacteria.

Methods on how to insert a transposon in a bacterium's genome and how to identify it:
Add the desired gene into the donor bacterium via a plasmid and let it multiply there (often in E.coli). Then let the recipient bacterium grow in the same medium as the donor in order to ensure cell-to-cell contact. Through conjugation, the plasmid will get to the recipient. The recipient should be unable to express the genetic information on the plasmid, so that it can only express the desired gene when through transposons it jumps over to the recipient's genome and is permanently integrated there. In case of a resistance gene, one can add an antibiotic to select for the desired bacteria. Also, remove the donor by changing growth conditions.

ADDENDUM – Quiz: Many molecular processes in bacteria are very similar in eukaryotes which is why bacterial lifeforms can be good model organisms to observe molecular processes. Also, they are haploid, reproduce very quickly, they form large cell populations on limited space etc.

E. Coli's genome is 4.6 million bp long.

Base pair substitutions occur with a frequency of 2*10**-10 per cell division and per base pair. Base pair substitions are 10 times more frequent than indels.

Def. indel: **in**sertion and/or **del**etion.

17.3.2017

Yeast genetics:

Advantages of yeast: short generation duration (only 90 minutes, that is almost as fast as bacteria; human cells take 24h to duplicate under optimal conditions);
experiments can be done in either haploid or diploid stage and growth occurs in both stages;
plasmids can be very easily added to the yeast and replicated as episomes there (since the origin of replication in yeast are known);
two haploid yeast cells can merge together into a diploid cell and one can observe dominant or recessive alleles;
a diploid cell can give rise to 4 haploid cells through meiosis which allows the observation of different phenotypes of multiple mutants (those 4 cells are called ascospores and they are held together in a "sack");
a single cell can be isolated and incubated and through asexual reproduction (budding) one can observe colonies of cells with identical genetic background (the "offspring" is considered to be a clone from the original single cell) (important to identify mutants and the loci of the mutation in the genome);
small compact genome

Def. episome: A piece of DNA that can exist and replicate autonomously in the cytoplasm or on a chromosome (mainly found in bacteria, but also in yeast).

Facts about yeast:

G1-phase: no budding, chromosomes are diffuse
S-phase: budding, sister chromatids are created (duplication of chromosomes)
G2-phase: budding, cell nucleus is next to the bud, sister chromatids still diffuse
Mitosis: chromosomes are segregated between mother and sister cell. In G1-phase, the cell nuclei are distributed through cytokinesis

Yeast has 16 chromosomes, 6600 genes and about 12 million base pairs. It does not have many introns, genes are rather short in length and therefore gene density is quite high. It also possesses a mitochondrial genome and the 2 mu plasmid. Genetic redundancy is low in yeast.

Def. Tetrade: Sack of 4 haploid cells (spores). This form occurs when environmental conditions are not in favour of the yeast cell. A tetrade consumes less energy and is more resistant to unfavourable environmental conditions. The cell wall is very thick of the sack (ascus) and protects the tetrade.

Mutagenesis methods:

Def. homologous recombination: A type of genetic recombination in which nucleotide sequences are exchanged between two similar or identical molecules of DNA. It is most widely used by cells to accurately repair harmful breaks that occur on both strands of DNA, known as double-strand breaks.

A chemically-induced mutation can be achieved with ethylmethansulfonate (EMS). EMS is used in temperature-sensitive mutations, since it methylates the DNA bases which results into the DNA polymerase adding the wrong base. It is often used for point mutations.
Intense radiation of UV-light leads to transitions and transversions.

We call such cells, conditional mutants that can only live under certain circumstances. It is especially helpful in diploid organisms in order to observe their mutations and phenotypes.
Heat sensitive mutations for example influence the stability of proteins. Cold sensitive mutations disturb protein interaction so that their respective reactions do not occur or are hindered.

Def. suppression analysis: Organism has a mutation. A second mutation is introduced or the gene activity of another gene is amplified in order to recover the initial phenotype. The mutated phenotype is no longer observed.

Def. synthetic lethality: Organism has a mutation. A second mutation is introduced or the gene activity of another gene is amplified so that the organism can no longer live under the current conditions. Organisms therefore die and this method is applied in order to observe protein interactions and signalling pathways.

Kor.: Synthetic lethality can identify redundancies in cell. It can also identify partial loss of functions, when it is a linear pathway.

ADDENDUM – Quiz:

Yeast also have very similar molecular and cellular processes like eukaryotes. Also, they are haploid, but they can also exist in a diploid stage. In order to observe lethal mutations, one mutates yeast in such a way, that their lethal mutation is only lethal under certain circumstances, such as temperature for example.

21.3.2017

In yeast, there are about 6000 genes of which 15% of all expressed genes have an unknown function and cellular purpose in yeast. The removal of those genes does not yield to any definite conclusion of their function, because no clear phenotype can be observed in yeast.

Plasmids can be very easily added and used in yeast.

Methods: complementation and sequencing

In sequencing, one simply sequences the mutant's genome and finds the mutations and compares it to the wild type. easy and cheap.

In complementation, one transforms the plasmids. yeast can very easily take up extra-chromosomal DNA. One checks whether a plasmid can save the yeast cell in its mutation. This complements the mutated phenotype and restores the phenotype. Then, simply sequence the plasmid which is even easier.

27.3.2017

On drosophila:

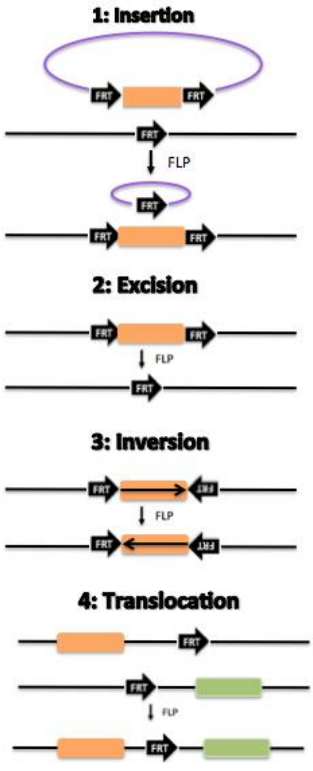
Facts: 4 chromosomes (3 autosomes and 1 sex chromosome) and about 180 million bp. Females have a segmented coloring of their abdomen, while males have their abdomen completely black. 15000 genes, 50% of them are homologous to humans. It is often used as a model animal for higher order animals such as humans. Also, it is often applied in embryonal research and its phenotype can often be observed by the naked eye or with a microscope.

Def. balancer chromosome: A chromosome with 3 criteria: It contains a lethal recessive mutation when it is homozygous, they possess one or more inverted DNA parts which lead to deletion or duplication during meiosis if crossing-over occurs, they possess a dominant phenotype marker.
Balancer chromosomes are used in model organisms in order to select the wanted organisms more easily and in order to preserve an otherwise lethal mutation (which is on the mutated chromosome) in a species. Ex: Drosophila.

EMS is an extremely effective mutagen that typically leads to transitions (point mutations).

Def. Flp-FRT recombination system Analogous to the Cre-loxP system. FRT (flippase recognition targets) sites and the flippase transgene have to be introduced artificially into most organisms. It is applied during mitosis to produce specific mutations.

Clonal screens vs. F2-screens: Advantage clonal screens: No need for a 2nd generation. Recessive phenotypes can be observed in F1 generation (homozygous mutants are produced in the F1 generation through homologues recombination in mitosis). Also, one can induce tissue-specific mutations or control mutations at a certain stage in development with Flp recombinase (Flp-FRT recombination system or Cre-loxP system or CRISPR/Cas9). Since a homozygous embryo is produced at a later stage in its development, one can omit in specific situations a lethal homozygous mutation.



8.4.2017

RNAi and CRISPR/Cas9

Def. Watson-Crick base pair rule: Ratio of A-T and G-C is 1:1. A pairs with T (pair with U in tRNA) and G pairs with C. There are alternative pairings that might occur in tRNA called reverse pairing, Hoegsteen pairing and Wobble pairing.

There are different types of RNA-interference, such as miRNA, piwiRNA and siRNA.

On micro RNA (miRNA): Its information is encoded in the introns of DNA and the genetic information is sometimes found in close neighborhood of other genes. miRNA does not code for a protein, it is used to suppress or to negatively regulate the expression of one or more mRNA. miRNA has complementary sequences according to the Watson-Crick

base pair rule and can bind to the mRNA. Since it is not 100% specific, it might also bind to other mRNAs and negatively regulate their protein expression.

How miRNA is made: The sequences is first transcribed into a primary transcript, the pri-miRNA. Then, splicing removes the introns, a polyadenyl chain is added and a 5' cap is added to the other end.
The pri-miRNA sequence contains inverted repeats that can hybridize and form hair pin structures. A protein complex consisting of Drosha and DCGR8, called a microprocessor, recognizes the pri-mRNA. Drosha cuts off the miRNA-stem-loop out of the pri-miRNA and isolates thus the pre-miRNA (60-80 bp long).
Pre-miRNA is exported into the cytoplasm with exportin 5. There the dicer protein recognizes double-stranded pre-miRNA and cuts it in such a way that we finally get miRNA-duplex (double stranded miRNA). Lastly, unneeded sequences are removed such that the mature single stranded miRNA is in the cell. The miRNA can now bind to the RISC-complex (RNA induced silencing complex).
The RISC-miRNA-complex can regulate protein expression in two ways now. If the miRNA is perfectly complementary to the mRNA, the complex will recognize the mRNA, bind to it and cut it in half. The cell will recognize the destroyed mRNA as non-functional and degrade it.
If the miRNA is not perfectly complementary to the mRNA, the RISC-miRNA-complex will still bind to the mRNA and negatively regulate its protein expression (translational block). The efficiency of translation is reduced. In such a case, a miRNA can influence several hundred genes.

On small interfering RNA (siRNA): siRNA is always perfectly complementary to the targeted mRNA. It is always used to destroy the mRNA (plants make use of siRNA to protect themselves from foreign mRNA). siRNA is commonly used for gene-knockdown experiments, where the RNA is rendered inactive instead of the DNA (such is the case in a gene-knockout experiment). There are two common ways: One can either synthesize the siRNA chemically and add it to the cell or write the corresponding DNA and add it to a vector. This vector will be introduced into a cell that can translate the genetic information on the vector. The DNA for the siRNA will contain inverted repeats such that it can form short hairpins. We call the transcribed DNA shRNA. The dicer enzyme will cut the double-stranded shRNA into siRNA.
The chemical approach will only knockdown the target RNA for some time, since it is degraded by the cell. On the other hand, the vector approach will always generate new siRNA, since the genetic information is transcribed more than enough.

How can one design such a siRNA sequence: As a general rule, it should be specific to the first 50-100 bp of the targeted mRNA, have a CG-percentage of 50%, start with two A's and internal hybridizations should be impossible.
Since a siRNA might also partially bind to another mRNA and negatively regulate the protein expression similar to miRNAs, off-targets effects (unwanted phenotypes) can occur.
Firstly, one can make use of bioinformatical methods to choose a siRNA specific enough. Secondly, 3 experiments with independently different siRNAs will be conducted and they all have the exact same target mRNA. If all 3 experiments show the same phenotype, it is unlikely that an off-target effect has been observed. In a follow-up control experiment, one uses one of the siRNAs to knockdown another gene in order to confirm that the phenotype is not the result of an injection of an siRNA (Ex.: through a non-specific immune reaction).

On CRISPR/Cas9: The system has its origin in bacteria where it functions as a bacterial immune system. The bacterial genome possesses clustered regularly interspace short palindromic repeats (CRISPR) parts. They are made of short repeat sequences (20-40 bp in length) with spacer sequences in between. The

spacer sequences contain the genetic information of a viral genome or other harmful DNA for example. After translation, the pre-crRNA is cut into smaller pieces called crRNA. Each of them possess one segment of the repeat sequence and one segment of the spacer sequence.
The crRNA binds to a tracrRNA (trans-crRNA). The result is a heterodimeric double stranded RNA.
This heterodimeric RNA binds to a multi-domain protein called Cas9. This huge complex is now able to recognize invasive harmful DNA and destroy it.

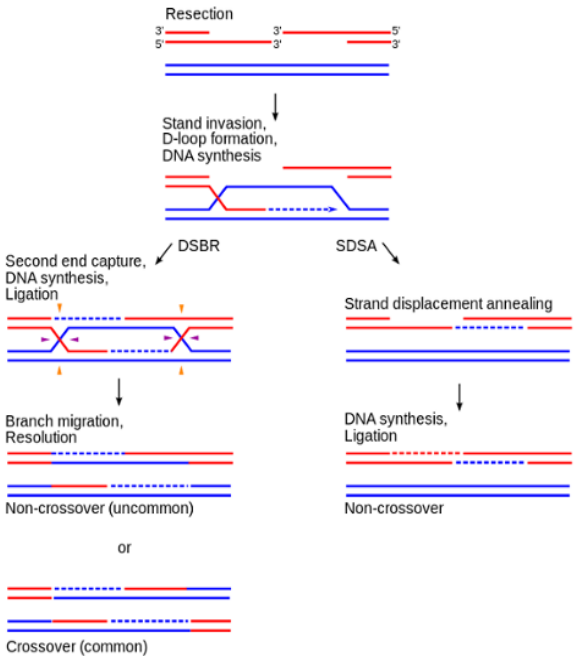
The Cas9 protein complex unwinds the target DNA and the spacer crRNA part can hybridize to the complementary DNA strand, if there is next to the foreign DNA a PAM sequence (protospacer adjacent motif, typically 2-6 bp long, in the case of Cas9 it is NGG). This DNA, registered by the spacer-sequence as harmful, is destroyed.

Def. single guide RNA (sgRNA): A long continuous RNA consisting of crRNA and tracrRNA.

How to do it in the lab: It is incredibly straightforward if the genome is known. First, one searches a PAM sequence (in S. pyogenes it is NGG, which occurs relatively often). The first 20 bp (from the 5' end) are then cloned in single guide RNA, which is in an expression vector with the right promoters already. This defines our spacer sequence. Another vector responsible for the expression of the Cas9 protein is inserted into a cell together with the first expression vector.

Def. non-homologous end joining (NHEJ): After a double strand break, the broken ends are simply joined together without a template. If it occurs in the protein coding region of the genome it will lead to the destruction of the protein (deletion/loss-of-function).

Def. homology directed pair (HDR): Uses the homologous template DNA for repair (has to be intact).



Def. oncogene: A gene that is capable of causing cancer. Such genes are expressed in high rates in mutated tumor cells.

QUESTIONS:

1. **Using CRISPR/Cas9 for germ line gene editing. What are the benefits and what are the risks?**

Benefits: Curing genetic diseases, but only easy for monogenetic diseases, since diseases involving several genes, it will be very hard to understand all these networks as of now.
Easier to make KO mice etc. with CRISPR/Cas9 to study them (already carried out in research)

Risks: Random damages to the genome can occur; unknown consequences might lead to new problems. It is much safer to choose an embryo that is heterozygous to a genetic disease, only makes sense for homozygous embryos that are really rare.
Is it ethical to produce "superhumans" while some people cannot have access to improve their genome -> could lead to an unfair system.
Introducing new genes might cause problems for coming generations and offspring might be unhappy about the parent's decision.

2. **Using CRISPR/Cas9 to generate gene drives: What are the benefits and what are the risks?**

   Benefits: Curing illnesses (e.g. Malaria), but maybe the illness itself might evolve and become resistant.

   Risks: It is not known, what gene drive might do to ecosystems.

3. **Using CRISPR/Cas9 to generate GMOs for agriculture. What are the benefits and what are the risks?**

   Using CRISPR/Cas9 it will be impossible to distinguish between wild type organisms and CRISPR/Cas9 organisms.

ADDENDUM – Quiz:

The CRISPR/Cas9 system can be used to insert a sequence in a genome through HDR (homology directed repair). It is also used to create targeted mutations in the cellular genome. Since a double strand break is often not perfectly repaired, but contains small insertions, the result is often an insertion mutation.

The natural function of CRISPR/Cas9 is to save detrimental DNA in the bacteria's genome and to destroy such DNA when it is found in the cell (such DNA originates from viruses).

The RNAi system can be used to permanently KO gene activity or for a certain period of time.