Gene mapping and identification

Introduction

Often, when studying a gene, it is critical to determine its position in the genome. In this lesson, we will discuss how tetrad analysis is used to map mutations. The mapping of mutations (and the respective genes) is critical for subsequent analysis of gene function.

Gene mapping using tetrad analysis

After identifying mutations in a screen and putting them into complementation groups, we want to know which genes are affected by the mutations. In the current era of sequencing, one might think it would be easiest to just sequence all mutants to know their mutation. However, sequencing hundreds of mutants resulting from a screen is quite expensive, because as we do not know where the mutation is located, we need to sequence the entire genome for each mutant. Furthermore, to overcome sequencing errors, the nucleotide sequences are usually sequenced several times (on average, more than 10 times) in whole-genome sequencing. This makes whole-genome sequencing a laborious and costly process, which is not feasible for several mutants. Additionally, sequencing the genome of a mutant strain generally reveals many mutations when compared to a standard yeast genome; simply because the strain used in the screen keeps itself to evolve and already had a slightly different genotype compared to the strain that was sequenced for the data base. If more than one mutation is detected when comparing the mutant genome with a reference genome, one still does not know which mutation causes the phenotype.

Therefore, genetic methods are used to narrow down the regions of the genome where the mutation is located (the chromosome or even a specific region on a chromosome). This can be determined by performing a linkage analysis. Here, the distance between two genes in the genome (the unknown gene affected in a mutant found in our screen and a known gene) is determined by measuring the frequency of recombination between the two genes occurring during meiosis. Genes are called linked if they do not segregate independently, indicating that they are located on the same chromosome.

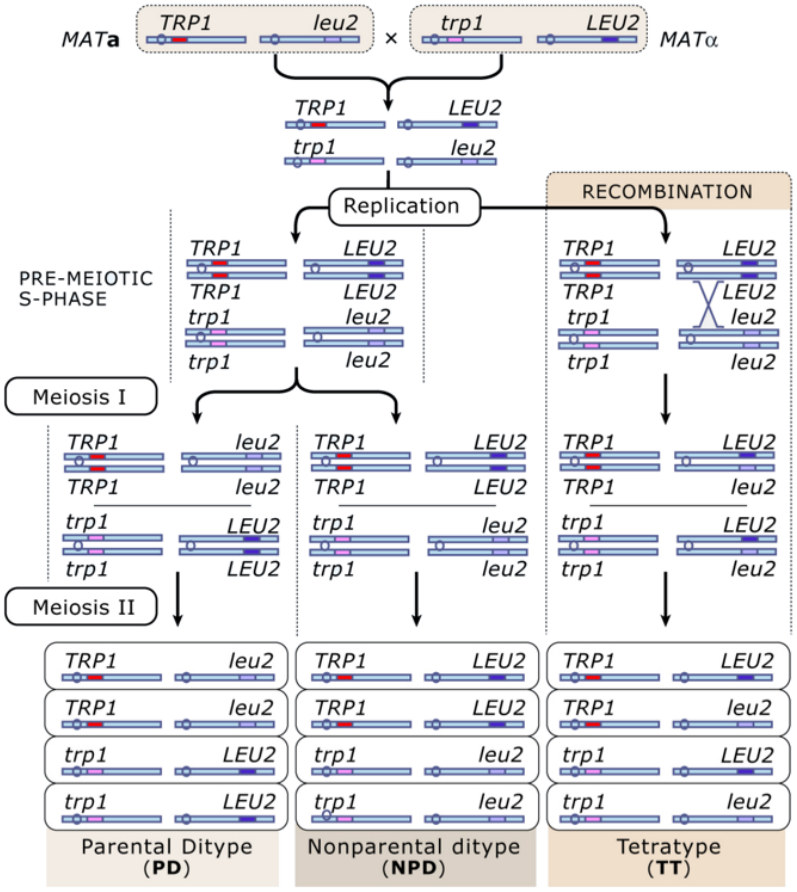
Yeast is especially well suited for these analyses, because the four spores in an ascus are the result of a single meiotic event. This allows formulating and testing specific predictions regarding the segregation patterns that will occur depending on whether two genes are linked or unlinked.

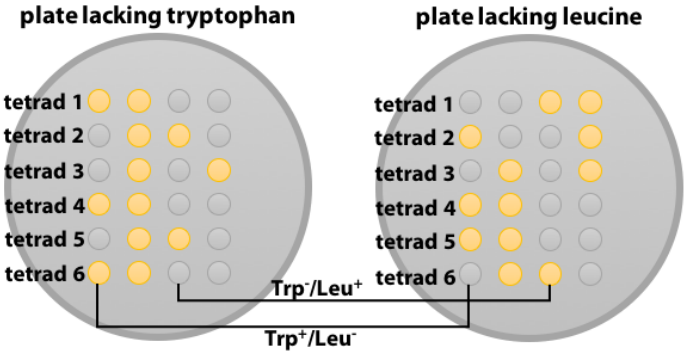
To understand how a linkage analysis is performed, we must first look at the three types of tetrads that can occur when two genes are segregating during meiosis. First, let us consider a genetic cross in which we will follow mutations in two genes that are on different chromosomes, i.e., the two genes are not linked to each other. The two mutations we will follow are called *a*, on chromosome III, and *b*, on chromosome IV.

In yeast nomenclature, capital letters refer to the allele conferring a dominant phenotype (*A*) and lower-case letters refer to the allele conferring a recessive phenotype (*a*). The wild-type allele is usually written as *A* or *A+*, whereas the mutant allele can contain a description of the allele (e.g. *ade2-203*). The phenotypes are usually written as Trp^- (a tryptophan auxotroph that requires tryptophan for growth) or Trp^+ (a wild-type cell that can grow without tryptophan).

In order to perform a dihybrid cross, the two parents are mixed together such that the two haploid parents can mate, and *A/a B/b* heterozygous diploids form. The diploids are then isolated and sporulation is induced by nitrogen starvation. The tetrads are then dissected and analyzed to determine the segregation of the *a* and *b* mutations.

Three types of tetrads arise as a result of the normal segregation of the markers: the parental ditype, the nonparental ditype, and the tetratype (see figure 5-1).





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)

In the absence of recombination, there are two possible outcomes for the tetrads. First, the two parental sets of markers segregate away from each other at meiosis I. The resulting tetrads have two spores bearing the genotype of one parent (*Ab*, in figure 5-1: *TRP1-leu2* as the parent with mating type MATa) and the other two spores bearing the genotype of the other parent (*aB,* in figure 5-1: *trp1-LEU2* as the parent with mating type MATα). This class of tetrads, resembling the two kinds of parental spores, is called a parental ditype (PD).

Second, the next class of tetrad is that in which segregation at meiosis I occurred in the opposite fashion. This results again in a tetrad with two different classes of spores, *AB* (*TRP1-LEU2*) and *ab* (*trp1-leu2*). This class of tetrad that has two kinds of nonparental spores is called a nonparental ditype (NPD). Because segregation at meiosis I is random and each configuration of homolog pairs occurs at equal frequency (according to Mendel's First Law), for two unlinked genes, the number of occurring PD tetrads and NPD tetrads is equal. This prediction is one of the fundamental tests of genetic linkage in yeast. Therefore, when testing the possible linkage of two mutations with unknown genetic positions, the first question is, "'Is the number of PDs equal to the number of NPDs?"' If the answer is yes, the mutations (and therefore the genes) are unlinked. If the answer is no, there is linkage. We will see below how this linkage can be further analyzed.

The third type of tetrad arises as the result of crossing-over events or recombination between one of the genes and the centromere of the chromosome on which it is located. Crossing over occurs after replication, when there are four chromatids. It results in a tetrad that contains four spores, each with a distinct genotype. In our example of the unlinked *A* and *B* genes, the four kinds of spores will be AB (*TRP1-LEU2*), ab (*trp1-leu2*), Ab (*TRP1-leu2*) and aB (*trp1-LEU2*).

Such a tetrad that has four genotypically different spores is called a tetratype (TT).

Therefore, we can summarize the outcome of linkage analyses as follows:

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, let's imagine that two genes, *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. Of the rest (<50%), the majority had only one crossover between *A* and *B*, because both genes are close to each other such that the likelihood of a crossover between both genes is small. After meiosis I, one nucleus will contain the *A* sisters with one having recombined, the other will have the *a* sisters with one having recombined. Therefore, the *A* nucleus will contain one chromosome *Ab* and the other *AB*, while the *a* mother will have one chromosome *aB* and the other being *ab*. After meiosis II, this combination produces a tetrad of type TT (*Ab*, *AB*, *aB*, *ab*).

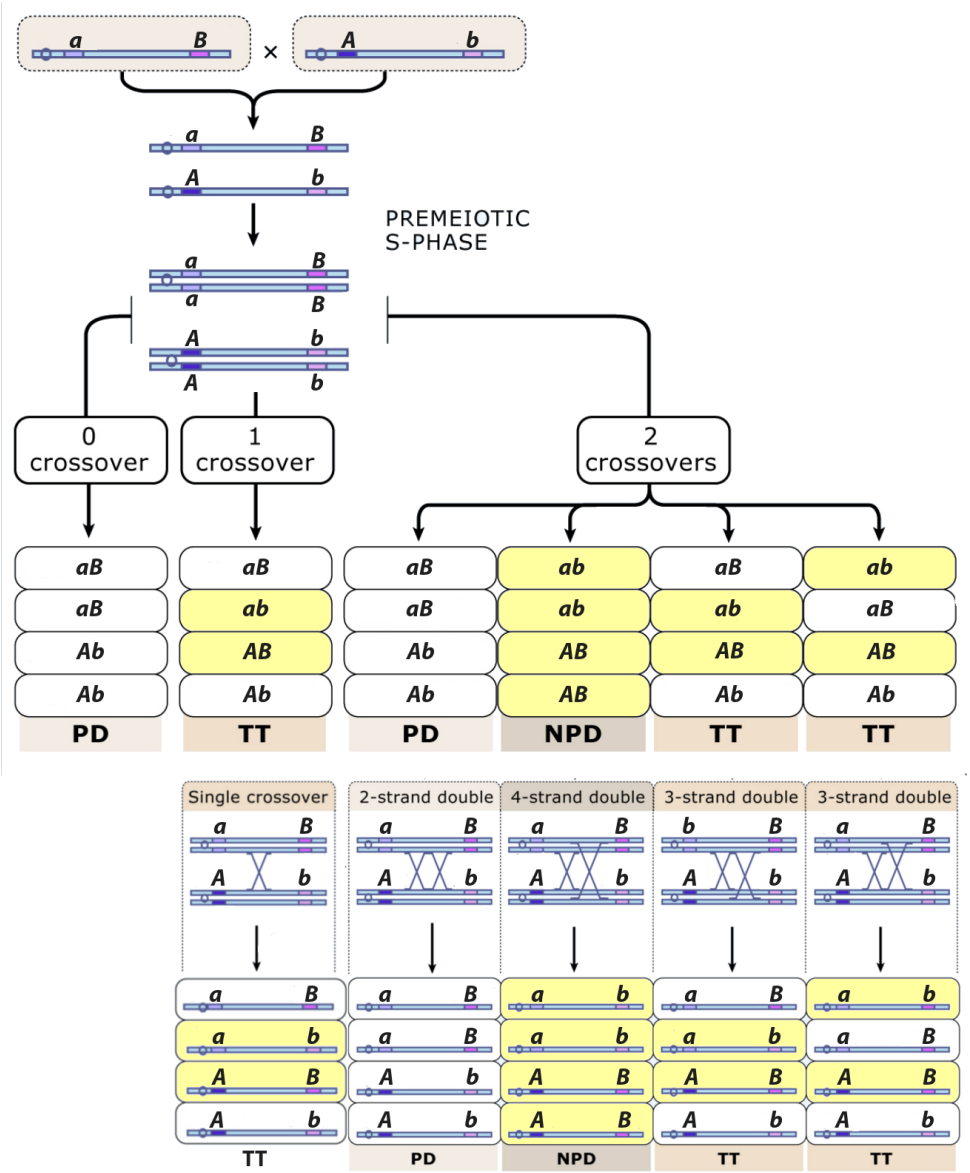


Figure 5-2 Tetrad types arising from a cross involving the two genes A and B that are closely linked, with A being closer to the centromere than B. Possible tetrad types are indicated in case of 0 crossover (more than 50% of the tetrads), 1 crossover (rare) or 2 crossovers (very rare). Recombinant spores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

Among the few cells that make two crossovers between *A* and *B*, some of them will make these crossovers across independent sister chromatids, such that after meiosis I, one nucleus will have two *AB* sisters, and the other will have two *ab* sisters. This results in the formation of a tetrad of the type NPD. Note that only few of the double crossovers lead to this situation (1/4 to be precise), such that NPD tetrads are very rare.

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).

In summary, by examining the tetrad patterns we observe across multiple asci, we can determine i) whether two genes in the yeast genome are unlinked, tightly linked, or loosely linked to each other, ii) whether both genes are tightly or loosely linked to the centromere, or iii) whether one or both genes are not linked to the centromere.

Crossover frequency and genetic distance

Now that we have examined the mechanisms behind tetrad formation, we want to look at how the frequency of tetrad types can be used to physically map the position of a gene by linkage analysis. Before we do so, we need to review how the genetic distance of genes can be determined using recombination frequencies.

We can use the phenomenon of crossovers and the resulting frequency of tetrad types to determine the order of genes along a chromosome and the approximate distances between those genes. This depends on the following basic principles:

- Crossovers are equally likely to occur at any point along the length of a chromosome.
- The probability of a crossover between two genes is proportional to the distance between the two genes. That is, the greater the distance between the two genes, the greater the probability that a crossover will occur between them during meiosis.

It logically follows that the greater the frequency of crossovers between two genes, the greater the distance between them.

The number of recombinants acts as a measure for the distance between the two genes that took part in the crossover event. This distance is expressed in terms of a genetic map unit (m.u.) or a centimorgan (cM) and is defined as the distance between genes for which one product of meiosis in 100 is recombinant. A recombination frequency (RF) of 1% is equivalent to 1 centimorgan.

Linkage is determined by the number of recombinants divided by the number of total progeny. In our example of peas, the frequency of crossovers is 10 recombinants per 100 total progeny, thus 10%, which is equivalent to a distance between the genes of 10 centimorgans.

Calculating gene-centromere linkage from tetrad frequencies

Beyond looking at the distance between genes, it is also possible to map the position of a gene relative to its centromere, provided that at least one known centromere-linked gene is present in the cross. This is because the centromeres of non-sister chromatids (from homologous chromosomes) always segregate away from each other in meiosis I, and the centromeres of sister chromatids always segregate away from each other in meiosis II.

When tetratype (TT) tetrads are produced, you have seen above this indicates that a crossover has occurred between a gene and its centromere. Now, imagine two genes are on different chromosomes and therefore unlinked. The only case in which one does not observe tetratypes is when both genes are very tightly linked (in very close proximity) to the centromere. Thus, for two unlinked genes, the frequency of TT tetrads depends on the linkage of each gene to its centromere. This property can be used to determine whether a gene is linked to its centromere or not. In practice, studies to determine gene-centromere linkage are done with a gene of interest and a mutant in a known centromere-linked

gene. The gene *trp1*, which causes an auxotrophic phenotype requiring tryptophan for survival, is very close to the centromere of chromosome IV and is often used in such studies.

To start, let us look at the case where two genes are very tightly linked to their centromeres. As mentioned before, *trp1* is tightly centromere-linked and is on chromosome IV. A second centromere-linked gene is *met14*, which lies on chromosome XI. Since these two genes are on different chromosomes, they are not linked to each other. Therefore, in a cross of *trp1* and *met14*, tetrad analysis will show that PD = NPD. However, because there will be very little recombination between either gene and its centromere, the frequency of TT tetrads will be extremely low.

By determining the frequency of TT tetrads, the possible linkage of any marker to its centromere can be determined by a genetic cross using the gene of interest and a known marker that is not linked to the marker of interest (is present on a different chromosome), but known to be tightly centromere-linked. For example, the *trp1* gene is approximately one map unit from its centromere, meaning that only one percent of the tetrads will show a recombination event between *trp1* and its centromere. Therefore, the TT tetrads in this cross will arise almost entirely by crossovers between the unknown gene and its centromere, as shown for *leu2* in figure 5-1.

To understand this principle, let us 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*?

1. Because PD = NPD, we know that *trp1* and *B* are unlinked to each other.
2. What about the linkage of *B* to thecentromere? 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 = $(\frac{1}{2} \cdot \text{TT})$/number of total tetrads. In our example, $(\frac{1}{2} \cdot 20)/100 = 10\%$, which corresponds to 10 centimorgan (cM).

Calculating gene-gene linkage from tetrad frequencies

Now, let us consider the scenario where two genes are linked to each other. By tetrad analysis we can determine the degree of genetic linkage between such markers. We will restrict ourselves to the case above where the distance is less than 25% crossover (hence less than 25 centimorgan). In the extreme cases when two genes are completely linked, with no detectable recombination between them, we expect that tetrad analysis will reveal that all tetrads are PD tetrads. The distance between these two loci is then non-measurably small.

As for centromere linkage, in the case of closely linked genes, we can calculate the degree of linkage based on the frequencies of the different types of tetrads produced. If there are no NPD, than the distance is as above for the distance to the centromere: linkage = $(\frac{1}{2} \cdot \text{TT})$/number of total tetrads, expressed in centimorgan. In case there are a few NPDs, then the calculation needs to take them in consideration and this leads to the Perkins formula, which we will not demonstrate here: linkage = $\frac{1}{2} \cdot (\text{TT} + 6 \cdot \text{NPD})$/number of total tetrads.