Recombination

Mendel's laws and chromosome theory
Remember **Mendel's law of segregation** (=*Spaltungsregel*). It states that each individual contains two alleles determining the value of a discrete phenotypic trait. During the formation of gametes, alleles are randomly distributed such that each haploid gamete only contains one allelic variant. The offspring obtains one gamete at random from each parent.

Let us consider an example. Crossing true-breeding (homozygous) round (RR) with wrinkled (rr) peas, Mendel first obtained a population of exclusively round (Rr) individuals (**law of dominance**=*Uniformitätsregel*). Di-hybrid crosses (Rr x Rr) yielded a ratio of 3:1 of round (RR, Rr) to wrinkled (rr) peas. This ratio is consistent with the expectation of random segregation of gametes (RR:Rr:rr = 1:2:1) as illustrated by the Punett square below assuming a probability of 0.5 for both the formation of each gamete followed by random assortment:

	paternal gametes		
maternal gametes		**R**	**r**
	R	RR	Rr
	r	Rr	rr

Mendel did not know anything about genes or the biochemical basis of the process in general. Molecular proof of his principle came from direct observations of chromosomes during meiosis in the late 19th century showing that paternal and maternal chromosomes get separated in meiosis and segregate into two different gametes. Each parent contributes a single gamete, and thus a single, randomly chosen allele copy to their offspring and fertilization. This **chromosome theory of inheritance** (Boveri-Sutton theory) was largely due to work by Theodor Boveri and Walter Sutton who recognized that chromosomes may constitute the physical basis of the Mendelian law of heredity.

Mendel's law of independent assortment *(=Unabhängigkeitsregel*) states that even alleles for separate traits are passed on independently from parents to offspring. Let us consider the following **test cross** with two traits:

colour - green (GG, Gg) or yellow (gg)
shape – round (RR, Rr) or wrinkled (rr)

F0: RR GG x rr gg
F1: Rr Gg x rr gg

With independent segregation the heterozygote is expected to produce the following gametes RG:Rg:rG:rg at equal proportion. The homozygote recessive will only produce rg gametes such that we expect the following offspring genotypes at equal frequencies (numerator and denominator indicate paternal and maternal origin, respectively):

$$\frac{RG}{rg}:\frac{rg}{rg}:\frac{Rg}{rg}:\frac{rG}{rg} = 1:1:1:1$$

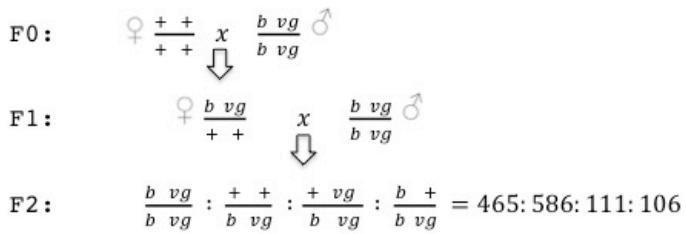
This ratio held pretty much for all crosses among seven traits that Mendel was considering. By the early 1880s the groundwork for the rediscovery of Mendel's principles was laid by the development of cytology. At the turn of the century Theodor Boveri and Walter Sutton independently had formulated the **chromosome theory of inheritance** stating that chromosomes form the physical basis of the Mendelian law of heredity. In case of traits being encoded by genes on different chromosomes, Mendel's ratios were consistent with chromosome theory. In case of **linkage** on the same chromosome, however, no **recombinant genotypes** would be expected:

$$\frac{RG}{rg}:\frac{rg}{rg}:\frac{Rg}{rg}:\frac{rG}{rg} = 1:1:0:0$$

Mendel's results were startling for his coeveals, as the ratio he obtained suggested independent segregation for combinations of any of the seven traits he considered – yet the pea *Pisum sativum* only has seven chromosomes. Wasn't it highly unlikely that the factors Mendel had chosen all resided on different chromosomes? It actually turns out that three of his genes actually were located on chromosome 4, and another 2 on chromosome 1. This should in principle result in deviations from his published ratios. It turns out, however, that the loci on chromosome 1 are so distantly related that they effectively segregate independently. The same is true for 2 out of the three possible combinations of the three loci on chromosome 4 leaving only one combination (loci influencing shape of mature pods + plant height) that would lead to deviation in expected ratios. Mendel, however, does not seem to have published and therefore, presumably, never made - or dismissed - the appropriate cross for this combination of genes tightly linked on one chromosome (Blixt 1975).

Measuring recombination: genetic maps

In the decades after the rediscovery of Mendel's laws it soon became apparent that reality differed from both Mendel's ratios and the prediction of chromosome theory. Genes did not always segregate independently (1:1:1:1 ratio in test crosses) nor did they show full linkage (1:1:0:0). At this time **Thomas Hunt Morgan** introduced the fruit fly as a model for genetics to test for patterns of inheritance using a variety of easy scorable phenotypes such as eye colour, wing shape, etc. His crosses confirmed the general prediction of chromosome theory by documenting a close correlation between chromosomal variants and the patterns inheritance they determined. However, he also found a substantial number of recombinant gene pairs that did substantially deviate from the 1:1:1:1 ratio. For example, in a testcross between a female wild type fly homozygous for normal colouration (+) and long wings (+) with a homozygous male mutant of black body colouration (b) and vestigial (rudimentary) wings (vg) he obtained ratios that were neither predicted by Mendelian laws nor by chromosome theory:



This corresponds to $\frac{111+106}{465+586+111+106} = \frac{217}{1268} = 0.171$, hence 17.1% recombinant genotypes (Morgan 1914). Repeating this kind of experiment with many traits yielded recombinant frequencies anywhere between 0 (no recombination) and 1 (full recombination) providing clear evidence for **crossing-over** in the female germline. Morgan reasoned that the frequency of recombination and hence the strength of linkage may be related to the distance on a chromosome with more recombination occurring between more distantly related loci. Inspired by this hypothesis his then 19-year old undergraduate student **Alfred Henry Surtevant** (who as a child had constructed pedigrees of his father's horses) went through the data and constructed the first **genetic linkage map** – working out the order and linear distances between genes using "three-factor crosses". Distance was arbitrarily expressed in units of recombinant frequencies as multiple of 0.01 (or 1%), a unit later denoted as one **centimorgan (cM** = 1 recombinant in 100 offspring).

Today genetic maps do no longer require phenotypic assays, but can directly be constructed from genotypic markers using informative crosses (as shown above), **pedigree data** from natural populations. Recently, it has also become possible to construct genetic maps from **single-sperm** sequencing. In any case, it is important that markers are to be densely spaced to avoid missing out on **double-crossovers** leading to an underestimation of map lengths.

Recombination frequencies often differ between sexes leading to differences in the total map length. In humans, for example, the total female, male, and sex-averaged lengths of the final maps have been estimated to be 44, 27, and 35 Morgans (1M=100cM), respectively **Sex-specific genetic maps** thus differ by almost a factor of two. Sex-linkage is clade specific. In the small tailed-opposum *Monodelphis domestica* e.g. the pattern is reverted and in *Drosophila,* sex-specific differences are at a maximum with recombination being entirely restricted to the female germline (male map length = 0).

Recombination frequencies do not only differ between species and sexes, they also differ significantly among chromosomes within species. As any successful meiosis requires one **obligate cross-over** per chromosome arm, recombination rate tends to be elevated in small chromosomes. This effect, for instance, is strongly pronounced in the avian karyotype with few large chromosomes and many small micro-chromosomes. Moreover, recombination also varies along the genome, often with strongly reduced recombination in regions of dense heterochromatin such as around the centromere.

Testing for recombination between two markers – the four gamete test
Consider two bi-allelic loci of ancestral allelic types *A* and *B*. In the ancestor, only *AB* haplotypes will exist. Once a mutation occurs on locus 1 we obtain an additional haplotype *aB*. Once locus 2 also receives a mutation we will end up with three haplotypes. Depending on where the mutation occurs this will yield *ab* or *Ab*. Let us here assume the mutation occurs on the ancestral haplotype.

AB → mutation: AB, aB → mutation: AB, aB, Ab → recombination: AB, aB, Ab, ab (cross-over between aB x Ab)

Under the infinite-sites assumption (i.e. repeat mutations have zero probability), the probability of a repeat mutation is zero, the maximum number of haplotypes is three accordingly. Any haplotype configuration with four haplotypes will be due to recombination. The four gamete test can estimate a minimum number of recombination events, but has low power. Morever, it is only suitable if the mutation rate is significantly smaller than the recombination rate and the infinite site assumption holds.

Evolutionary consequences of recombination
Recombination is a central parameter for many aspects of evolution.

Recombination generates phenotypic variation
It has long been a question how continuous phenotypic traits can be encoded by discrete factors (genes). Recombination is key to an answer. Imagine a polygenic trait encoded by three Mendelian factors, say locus A, B and C. Crossing pure bred A_1A_1 B_1B_1 C_1C_1 (white) and A_2A_2 B_2B_2 C_2C_2 (dark red) will result in virtually continuous variation in colouration already in the F2 generation from white to dark red assuming co-dominance and free recombination between loci. In case of linkage, one would have to wait longer, but eventually recombination will unlink loci and introduce novel combinations. The Swedish plant breeder **Herman Nielsson-Ehle** conducted a series of experiments with purebred winter wheat having either white or brown seed. In the F2 cross he obtained continuous variation in colour.

Can we predict the number of genotypes generated by recombination? First, consider two loci in an F1 cross between homozygous parents AA BB and aa bb. Each F1 individual in heterozygous state Aa Bb will produce four gametes assuming free recombination ab, Ab, aB, AB. From those 10 different diploid offspring (e.g. aabb, aaBb, aaBB, etc.) can be formed. In general, 2^ngamete combinations (haplotypes) can be formed from n loci assuming free recombination. From these

$$2^{n-1}(2^n+1)$$

distinct diploid genotypes can be formed (counting AB/ab and Ab/aB as distinct) (**Figure 1**). The number of possible diploid genotypes increases rapidly: from 10 loci more than half a million genotypes can be formed. This insight settled the bitter debate between biometricians and Mendelians reconciling Darwin's concept of natural selection on continuous variation and Mendelian hereditary principles by discrete genetic factors.

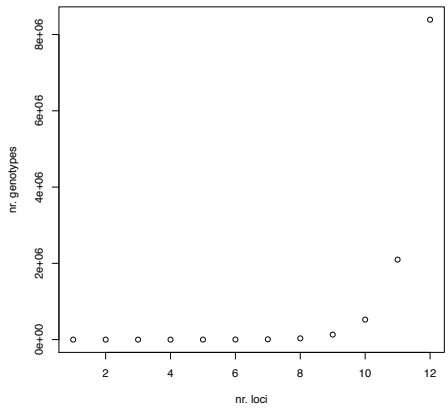


Figure 1: Relationship between the number of Mendelian factors (freely recombining loci) and the number of recombinant, diploid genotypes that can result after free recombination.

Recombination speeds up adaptation

Without recombination, beneficial mutations must be established sequentially; with recombination favourable mutations at different loci can be combined. Consider two loci in a large, haploid, asexual population (no recombination) with alleles a and b. A favourable mutation A will lead to an increase of the Ab haplotype. Another favourable mutation B occurring shortly after on the ancestral haplotype will result in a novel haplotype aB. Now haplotypes Ab and aB compete and the haplotype with the mutations conferring less of a selective advantage will eventually be lost from the population. In a sexual population with recombination, the haplotype AB can be formed which will quickly spread to fixation. In this case, both advantageous mutations survive.

Similarly, if an advantageous mutation happens to end up in a haplotype with deleterious mutations, it will not reach fixation (Hill-Robertson interference). In the end it is the fitness of the haplotype that matters. Recombination can break the linkage between deleterious mutations and the advantageous mutation randomizing the haplotypic background. With recombination selection will thus act on the locus, without recombination it will act on the entire haplotype.

Recombination counters the accumulation of deleterious mutations

Deleterious mutations accumulate in asexual populations via Muller's ratchet; we speak of the **mutation load** of a population. Under the justified assumption that most mutations are deleterious, we expect a certain number of mutations U per generation reducing the fitness of an individual (and the population). In a balance between deleterious mutation U=0.1 and selection s=0.02, and equilibrium frequency is reached at U/s =5 deleterious mutations per genome. In a population of 1000 individuals there are on average only 1000 $e^{-U/s}$ = 6.7 individuals free of mutation. These individuals with the fittest genotype will eventually be lost by genetic drift and can never be recovered – decreasing the fitness of the entire population. Unless there is backmutation or recombination recreating mutation-free haplotypes the expectation in the long run is that the population will go extinct.

Literature: (Barton et al. 2007; Futuyma 2013; Nielsen and Slatkin 2013)

Barton NH, Briggs DEG, Eisen JA, Goldstein DB, Patel NH. 2007. Evolution. 1st edition. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press

Blixt S. 1975. Why didn't Gregor Mendel find linkage? Nature 256:206–206.

Futuyma DJ. 2013. Evolution. 3rd ed. Sinauer Associates

Morgan TH. 1914. No crossing over in the male of drosophila of genes in the second and third pairs of chromosomes. Biol. Bull. 26:195–204.

Nielsen R, Slatkin M. 2013. An Introduction to Population Genetics: Theory and Applications. Sunderland, Mass: Macmillan Education