Segregation Ratios

Reference: Mather (1957)

- **Single genes and no linkage.**

SEGREGATION IN DIPLOIDS

Analysis of single gene segregations–determine what type of genetic control exists for a single trait–i.e., single gene dominant, codominant, two genes with partial dominance, etc.

- Need single genes in order to conduct linkage analyses.

- Testcross: Cross heterozygote Aa to homozygous recessive aa
  Intercross: Cross two heterozygotes Aa and Aa to produce an F_2 progeny

- Single genes in diploids: backcross will yield a 1:1; intercross 3:1 if dominant, 1:2:1 if codominant

Deviation from expectation–fitting genetic ratios

- Do the data you observe fit certain expected ratios?

- Use the χ^2 test:

$$\chi^2 = \sum_{j=1}^{n} \left(\frac{(o_j - e_j)^2}{e_j} \right)$$

where the summation is over genotypic classes (*j*), o and e are the observed and expected number of individuals in genotype class *j* , respectively.

Table 1: Antirrhinum majus: self a yellow flowered plant known to be heterozygous. Test for fit to a 3:1 ratio:

Class	Observed number	Expected number	Deviation (o-e)	χ^2
Yellow	208	216.75	-8.75	0.353
Ivory	81	72.25	+8.75	1.056
Total	289	289.00	0.00	1.409

- The χ^2 of 1.409 has 1 degree of freedom (**because we can only fill one class arbitrarily**).

- The probability value of obtaining a deviation as large or larger than 1.409 is between 0.3 and 0.2, which means that we fail to reject the null hypothesis, and therefore, the observed ratio fits a 3:1 ratio.

- **In general, no probability greater than 0.05, i.e. one in 20 should be considered as significant deviation.**

- More complex ratios--e.g. two genes each segregating 3:1--->9:3:3:1--can be analyzed in exactly the same way.

- Special cases of the χ^2 formula for a family segregating into two classes are given by Mather. They are especially easy to use for determining fit to various ratios:

- e.g., for a 1:1 ratio, the χ^2 formula becomes:

$$\chi^2 = \frac{1}{n}(a_1 - a_2)^2$$

where, a_1 and a_2 are the observed numbers in each class, and n is the total number observed. This follows from the general formula for a family segregating into two classes, with an expected ratio of l:1:

$$\chi^2 = \frac{1}{l * n}(a_1 - (l * a_2))^2$$

(Mather, 1957 page 25 for additional ratios)

<u>Testing homogeneity of data from several segregation experiments</u>
(Mendel's 2:1 ratio experiment)

- Because the χ^2 is additive, we can evaluate segregations from several experiments or crosses to determine if they all show the same results.

$$\chi^2_{Total} = \sum_{i=1}^{p}\sum_{j=1}^{n}\left(\frac{\left(o_{ij} - e_{ij}\right)^2}{e_{ij}}\right)$$

where p = number of populations and n = number of genotypic classes

$$\chi^2_{Pool} = \sum_{j=1}^{n} \frac{\left(\sum_{i=1}^{p} o_{ij} - \sum_{i=1}^{p} e_{ij} \right)^2}{\sum_{i=1}^{p} e_{ij}}$$

$$\chi^2_{Heterogeneity} = \chi^2_{Total} - \chi^2_{Pool}$$

degrees of freedom:		
	Total	p(n-1)
	Pool	n - 1
	Heterogeneity	(n-1)(p - 1)

- Thus, the total χ^2 is the sum of the individual χ^2 values; the pooled χ^2 is determined on the sum of the individual observations.

- Mendel's data

- If the data from different studies or crosses is consistent, then the pooled value will be very close to the total value.

- However, if the independent studies produce very different results, the pooled value will be quite different from the total value.

- Thus, in the latter case, the heterogeneity among studies will be high–indicating that something, either due to sampling small numbers in each case, experimental error, or different genetics in different crosses, is causing the studies to vary.

Distinguishing between two possible two-class segregation ratios

- What if we have data such that two different types of segregation are possible.
- How do we decide how many individuals are needed to be able to differentiate between them with some probability?

- Suppose that we suspect our trait to be segregating as either 3:1 or 9:7.

- We have an F_2 family size "n". For a given size, n, the number of recessives, r, that we observe will be equally likely to have arisen from a 3:1 ratio or a 9:7 ratio--this is the *ambiguous segregation ratio*.

- If we observe more than "r" recessives, then the 9:7 ratio is more likely.
- But if we observe fewer than "r" recessives, the 3:1 ratio is more likely.

- We have to study the segregation in a large family to ensure that "r" recessives will show a deviation from expectation on either hypothesis of a size that could occur only with that probability chosen as the maximum for classification.

- For the general case, the two expected segregations are the following:

  - $l_1 : 1$ and $l_2 : 1$.
  - Observed segregation that results equal χ^2 for both hypotheses is:-
  - $\sqrt{l_1 l_2} : 1$

- In our case ratios are 3:1 and 9/7:1

- The segregation pattern giving equal χ^2 for both hypotheses is $\sqrt{\frac{27}{7}} : 1$ which equals 1.964:1.
- The number of recessives, "r", equals: $\frac{1}{2.964} n$
- Dominants = $\frac{1.964}{2.964} n$

- The ambiguous segregation in terms of *n* is $\frac{1.964n}{2.964} : \frac{n}{2.964}$

- To understand how we derived these numbers, compare a 3:1 ratio, in which

  ¾ n are dominant and ¼ n are recessive

- We can think of this ratio as the borderline "observed" ratio that separates 3:1 from 9:7. We use this ratio as the "observations" in the formula from the previous section.

$$\chi^2 = \frac{1}{3n}(a_1 - 3a_2)^2$$

  which is Mather's short formula for 2-class segregation of 3:1.

- The χ^2 for this segregation can be obtained from the tabled value, at 1 df and the chosen probability. In this case, we look for = 0.05, but realize that the actual value is 0.025 since we are only interested in deviation in one direction from 3:1 (i.e. toward more recessives).

- We get the following:

$$3.841 = \frac{\left(\frac{1.964n}{2.964} - \frac{3n}{2.964}\right)^2}{3n} = \frac{\left(\frac{-1.036n}{2.964}\right)^2}{3n} = \frac{0.122n^2}{3n}$$

$$\boldsymbol{n = 94.31}$$

- Thus, we need to test 95 plants to be 97.5% sure that the segregation we see is either 3:1 or 9:7; that is, to distinguish between the two hypotheses.

- So if we see more than 32 recessives out of 95 plants, we assume 9:7 is correct;

$$32 \left(= \frac{1}{2.964} * 95\right)$$

- If we observe fewer than 32 recessives, then 3:1 is the right segregation.

- Bailey (1961) presents a generalized equation for determining number of individuals to distinguish two segretations: (this doesn't work if one ratio is 1:1)

$$n \geq \left(\frac{1+\sqrt{xy}}{\sqrt{x}-\sqrt{y}}\right)^2 \chi^2_{\alpha,1}$$,where x:1 and y:1 are the two segregations to be distinguished.

EUPLOIDY (REVIEW):

Chromosome numbers are usually whole multiples of a basic number

- **Haploid:** Also monoploid, a plant having the "n" number of chromosomes or basic chromosomes.
  -Having one set of chromosomes--e.g. arising from a gamete from a diploid species
- **Polyhaploidy**: A plant developed from gamete of species with n>x (e.g., alfalfa, n=2x), still can be called a haploid.

- **<u>Doubled haploids</u>**
  -Haploidy can be induced in some plants: e.g. small grains through anther/pollen culture or interspecific hybridization (in some cases).
  -Why would you do this?
  -The haploid will only have one allele/locus
  -Double chromosomes with colchicine or other method-->homozygocity for every loci attained instantly.

- **Polyploid**
  - Two or more basic genomes and/or more than two chromosome sets of a given genome.

  Many species are polyploids (70% grasses, 23% legumes, for example)
  Others are "**paleopolyploid**" or diploidized polyploids, e.g. maize and soybean
  It is proposed that any species with >10 chromosomes could be a paleopolyploid

  - Gene mapping experiments have shown duplicated regions within a genome quite well;
    e.g. soybean (Shoemaker et al., 1996).
  - Every soybean linkage group has homoeology to other groups--i.e. ***synteny*** of loci-
    large scale duplication is evidence of ancient polyploidy.
  - Small scale duplications can be due to other factors.

<u>Types of polyploids</u>:

- Genomes are labeled by capital letters: e.g., wheat has three homoeologous genomes: A, B, D
- Autopolyploidy--many sets of same genome: AA---> AAAA
- Allopolyploidy--two or more distinct genomes: AA + BB----> AABB or AB---> AABB
- Segmental allopolyploidy--genomes are somewhat differentiated, but can still pair
  $A_1A_1 + A_2A_2$---->$A_1A_1\ A_2A_2$
- Autoallopolyploids--mixture of common and different genomes

Chromosomal Pairing in polyploids

- Allopolyploids will not show any pairing higher than bivalents 7II=7 bivalents.
- Autopolyploids may show higher-order pairing. 7IV=7 quadrivalents etc.

Better terms reflect the type of segregation that each shows:

- ***Disomic*** **polyploids** = allopolyploids
- ***Polysomic*** **polyploids** = autopolyploids
- **Amphidiploid:** Induced by doubling of 2x to 4x after wide crosses to restore fertility.

- Man-made polyploids.
  - Triticale (wheat x rye)
  - Raphanobrassica (radish x cabbage)

Endopolyploidy (De Rocher et al., 1990)(Galbraith et al., 1991)

- Amplification of the nuclear genome in somatic cells.
- Easily test with flow cytometry (measures the size of nuclei).
- Systemic in *Arabidopsis* and ice plant-up to 64C in some cells of all tissues except inflorescences. Increased as tissues matured**.** Importance not clear–possibly related to gene regulation.

Nomenclature On Chromosome Numbers And DNAContent

n=gametophytic chromosome complement
2n=sporophytic chromosome complement

x=basic chromosome number, in one complete chromosomal set (e.g. a diploid gamete 1x=1n)
2x=number of chromosomes in a diploid sporophyte.
4x=number of chromosomes in a tetraploid sporophyte (i.e. four sets).

Examples

Diploid	Corn	2n=2x=20
Triploid	Banana	2n=3x=33
Tetraploid	Alfalfa	2n=4x=32
Hexaploid	Wheat	2n=6x=42
Octaploid	Sugarcane	2n=8x=80

Relative amount of DNA in a cell:

C="constant"–relative amount of DNA in a single chromosome set (a gamete in a diploid organism)
2C=DNA in a normal cell of a diploid organism
4C=amount of DNA in a diploid cell after "S" but before mitotic telophase

Absolute amount of DNA:

Value given in picograms (10^{-12} g)
1 pg = 0.965 x 10^{9} nt pairs.

SEGREGATION IN POLYPLOIDS

General reference: Burnham (1962)

Pairing and segregation

1. Cytology: Counting the number of bivalents and multivalents in meiosis will give some idea on the relatedness of the genomes. However, small chromosomes may form multivalents infrequently, possibly leading to the incorrect conclusion that the two genomes are different. Likewise, genes for preferential pairing may give a similar impression.

2. Segregation patterns: Tetraploid segregation.
   Disomic inheritance: The case for disomic polyploids (i.e., "allopolyploids") strict bivalent pairing:

   AAaa selfed –> two possibilities

   1. Homologues are homozygous; homoeologues are heterozygous:
      AA and aa : All gametes are Aa (one from each set of paired homologues).
      Progeny are all AAaa.

   2. Homologues are heterozygous:
      Aa and Aa : Thus, gametes form in the ratio 1AA:2Aa:1aa
      progeny are 15 A-:1aaaa
      (1AAAA:4AAAa:6AAaa:4Aaaa:1aaaa)

- AAaa testcross:
  - 3A-:1aaaa for case 2
  - All A- if case 1 above).

- **Tetrasomic inheritance**: The case for polysomic polyploids (i.e., "autopolyploids")- any chromosome can pair with up to 3 homologues multiple homologues, Therefore, can have higher order pairing, e.g., quadrivalents can occur.

  AAaa selfed –> gametes 1AA:4Aa:1aa –> progeny 35A-:1aaaa

  (1AAAA:8AAAa:18AAaa:8Aaaa:1aaaa)

  AAaa testcrosses (x aaaa) –> progeny 5A-:1aaaa

With tetraploids, five different genotypes and multiple alleles per locus are possible:

AAAA	quadriplex	$A_1A_2A_3A_4$, abcd	tetragenic
AAAa	triplex	$A_1A_1A_2A_3$, aabc	trigenic
AAaa	duplex	$A_1A_1A_2A_2$, aabb	diallelic duplex
Aaaa	simplex	$A_1A_1A_1A_2$, aaab	simplex
aaaa	nulliplex	$A_1A_1A_1A_1$, aaaa	nulliplex

Generally, the $A_1A_2A_3A_4$ notation is preferred because it can accommodate multiple alleles, and it avoids the problem of having several letters (abcd) that are usually associated with multiple loci.

Molecular markers and polysomic polyploids Wu et al. (1992)

- If only 1 of 4 homologues has a particular band (i.e. simplex for that band size) = SDRF or single dose restriction fragment or more generally (including PCR-based markers) = SDA, single dose allele.
  - An SDRF will segregate 1:1 if the parent carrying the band is crossed to a plant without the band.
  - Selfing of the parent (P_1) carrying the SDRF will segregate to a 3:1 ratio.
  - Double haploids with absolute homozygosity could be used as the other parent.
  - P_1 parent should be highly heterozygous.

Example: Handout

Tetrasomic Inheritance

In diploids:

- *Reductional separation*: sister chromatids separate to **same pole** attached to the **same centromere**
- *Equational separation*: sister chromatids to separate to **different** poles

- Reductional or equational separation depends on the location of the locus in elation to the entromere.
  - e.g. close to centromere–little chance of crossing over and therefore reductional at Anaaphase I.
  - Far away from the centromere - more likely a crossover will result.

Tetraploids:

- Consider equational chromosomes to mean that sister chromatids are attached to different centromeres by a crossing over.
  - Equational chromosomes may now separate to the same pole because two homologues go to each pole.

Therefore, the gametic series is dependant on the amount of equational separation at a given locus.

Three models of segregation: see handout showing four homologues

1. *Random chromosome segregation* (Muller, 1914).
   - If no crossing-over takes place between gene and the centromere the two chromatid carrying the dominant allele will g to the same pole and separe in the second division. They never appear in the same gamete. The gametes that we then observe from random chromosome pairing of a tetraploid are:
     - simplex: 1Aa : 1aa
     - duplex: 1AA : 4Aa : 1aa
     - triplex : 1AA : 1Aa
   - Commonly discussed ratios–e.g., 35:1 for a duplex.
   - Assumes no recombination and is therefore only relevant to the centromere and loci closely linked to it. (100% reductional separations at Anaphase I).

2. *Random chromatid segregation* (Haldane, 1930)
   - Each chromosome could have multiple pairing partners at any time at the most extreme, a chromosome could cross over with all three homologues.
   - 6/7 equational, 1/7 reductional separations at Anaphase I (for loci beyond the cross-overs); i.e., in only 1/7 of the possible separations at a particular locus will the sister chromatids be joined, since the chromatids segregate randomly due to an infinite number of cross overs.
- Therefore, any chromatid can end up in a gamete with any other chromatid with equal frequency, leading to 28 gamete types, 4 of which contain sister alleles; i.e., there are 8 chromatids, and one can end up with any of the other seven, one of which is its sister.
- The genes will segregate at random with respect to the centromeres and gametic series will become:
  - simplex: 1Aa : 12Aaa : 15aa
  - duplex: 3AA : 8Aa : 3aa
  - triplex : 15AA : 12Aa : 1aa

3. *Maximum equational segregation* (Mather, 1935, pp. 70-77)
   - Each chromosome forms a cross over with only one other chromosome at a time, allowing all distal loci to separate equationally at meiosis I; thus, 100% equational separations or *maximum* equational separations.
   - The two dominant allele carrying chromatid will be joined to two differente centromeres, each of which also has a recessive allele carrying chromatids in a simplex organism.
   - Following random disjunction of the quadrivalent into two and two, the two dominant carrying chromatids will appear in the same interphase nucleus in 1/3 of the cases and the same gamete 1/6 of the tetrads.
   - With complete equational separation gametic series will be:
     - simplex: 1AA : 10Aa : 13 aa
     - duplex: 2AA : 5Aa : 2aa
     - triplex : 13AA : 10Aa : 1aa

Double Reduction

- *Double reduction*: In autopolyploids, if separation for any locus is equational the two chromatids from one chromosome may be present together in one interphase nucleus but joined to separate centromeres allowing them to enter the same gamete. Sister chromatids in the same gamete, reducing the genetic content of a gamete twice, instead of once.
- Normally, two of the four chromosomes end up together in a gamete, reducing the genetic content in half. With double reduction gametes, the two chromosomes in the gamete are the same, at least at some loci; i.e., they are sister chromatids, and genetic content is reduced to ¼ when compared to the parental plant.

Requirements for double reduction:

1. **Formation of quadrivalents** (q): Allows any two (of the four) homologous chromosomes to migrate to the same pole. With bivalent pairing, the two chromosomes will migrate to the opposite poles and does not allow equational chromosomes to end in same gamete.

2. **Formation of equational chromosomes** (e): Crossing-over occurs to separate sister chromatids to different centromeres; thus, double reduction only refers to loci distal to the cross-over.

3. **Equational chromosomes migrate to same anaphase I pole** (a): Sister chromatids go to the same pole, but attached to different centromeres. If the equational chromosomes do not go to the same pole, there is no chance for the sister chromatids to end up in the same gamete.

4. **Correct orientation at Metaphase II:** so that the two sister chromatids (now attached to different centromeres) will go to the same gamete cell. Probability = ½

Calculating double reduction:

Pr{double reduction, α} = q * e * a * ½

q = probability often assumed = 1, although this is usually not the case in actuality

e = depends on the crossover frequency

a = a if random

Therefore, Maximum equational segregation $\alpha = 1 * 1 * a * ½ = 1/6$ $a = 1/3$
Random chromatid segregation $\alpha = 1 * 6/7 * a * ½ = 1/7$ $a = 1/3$

Biological significance of double reduction:

- Increases the number of recessive phenotypes recovered.
- Increases the inbreeding coefficient.

A triplex locus will not produce any homozygous recessive offspring upon selfing without double reduction.

e.g.,	Disomic	AAaa		1/16 aaaa
	Tetrasomic	AAaa	Random chromosome segregation (CS)	1/36 aaaa
			Maximum equational segregation (MES)	1/20 aaaa
			Random chromatid segregation	1/22 aaaa

Empirical data: Burnham (1962)

Datura stramonium (Blakeslee et al., 1923)

2 unlinked factors, A and P

Genotype	Obs Dom	Obs Rec	% Rec	Exp Rec CS	Exp Rec MES
AAaa x aaaa	518	137	20.9**	16.7	22.2
AAaa x AAaa	3383	118	3.4**	2.8	4.9
PPpp x pppp	905	179	16.5	16.7	22.2
PPpp x PPpp	9199	225	2.4	2.8	4.9

** Sig. diff. from expected number based on RCs and toward MES

This and other data indicate that A is further from its centromere than P is from its centromere.

Double reduction ratios:

For example: Duplex AAaa:

- Random chromosome:

  aa gametes are present in frequence 1/6

  aaaa genotypes in progeny in frequency 1/36 = 0.028

- Maximum equational segregation:

  Double reduction occurs 1/6 of time; ½ of these gametes will be aa and ½ AA.

  Of the remaining 5/6 when no double reduction occurs, 1/6 will be aa.

  Thus, the number of aa gametes:

  aa from non-double reduction = 1/6* 5/6 = 0.138

  aa from double reduction = 1/12 = 0.083

  total aa = 0.221

  The number of aaaa progeny = 0.049 = 4.9%

New method for estimating segregation: (Jackson and Jackson, 1996)

Meiotic configuration method:

- Assumes a maximum of 2 chiasmata/bivalent and 4/quadrivalent.

- Determines the theoretically expected numbers of II, chain IV, and circle IV using available empirical data.
- Determines the contribution of each configuration to the gamete genotypes.
- This model fits empirical data better than other methods

DIHYBRID F_2 RATIOS IN AUTOTETRAPLOIDS

Table 93 - Burnham (1962)

- AAaa BBbb selfed (2 genes, unlinked):
  1225 AB : 35Ab : 35aB :1ab with random chromosome segregation.
- Frequency of recessive class is very low in all cases–need large populations needed to uncover them
- Also affected by double reduction.
- Compare diploid AaBb F_2 ratio: (9:3:3:1)!

FAMILY SIZE TO ENSURE ONE OF DESIRED GENOTYPE

Mather (1957), Table III, p. 143; Sedcole (1977); Hanson [, 1960? #621]

Ex. 1. Self Aa, how many progeny (n) to grow to be 99% sure that one is 'aa', or that all aren't A-?

- We know that $(^3/_4)^n$ will have at least one "A" allele.
- We don't want all "n" individuals to be A- more than 1% of the time.
- One plant–75% chance that it will be A-; 2 plants (.75)(.75) = 56.25% that both are A-

thus:

P{not finding the desired type}^{n} = Error rate
$(3/4)^n = 0.01$
n log (3/4) = log (0.01)
n = 16.0

Ex. 2. F_2 seg for two genes. Want double homozygous dominant, AABB

The A-B- class will comprise 9/16 of F_2 progeny, 1 of the 9 will be AABB
Therefore, $(8/9)^n = 0.01$
n = 39.1 = 40 A-B- F_2 plants to be 99% sure of including at least one AABB.