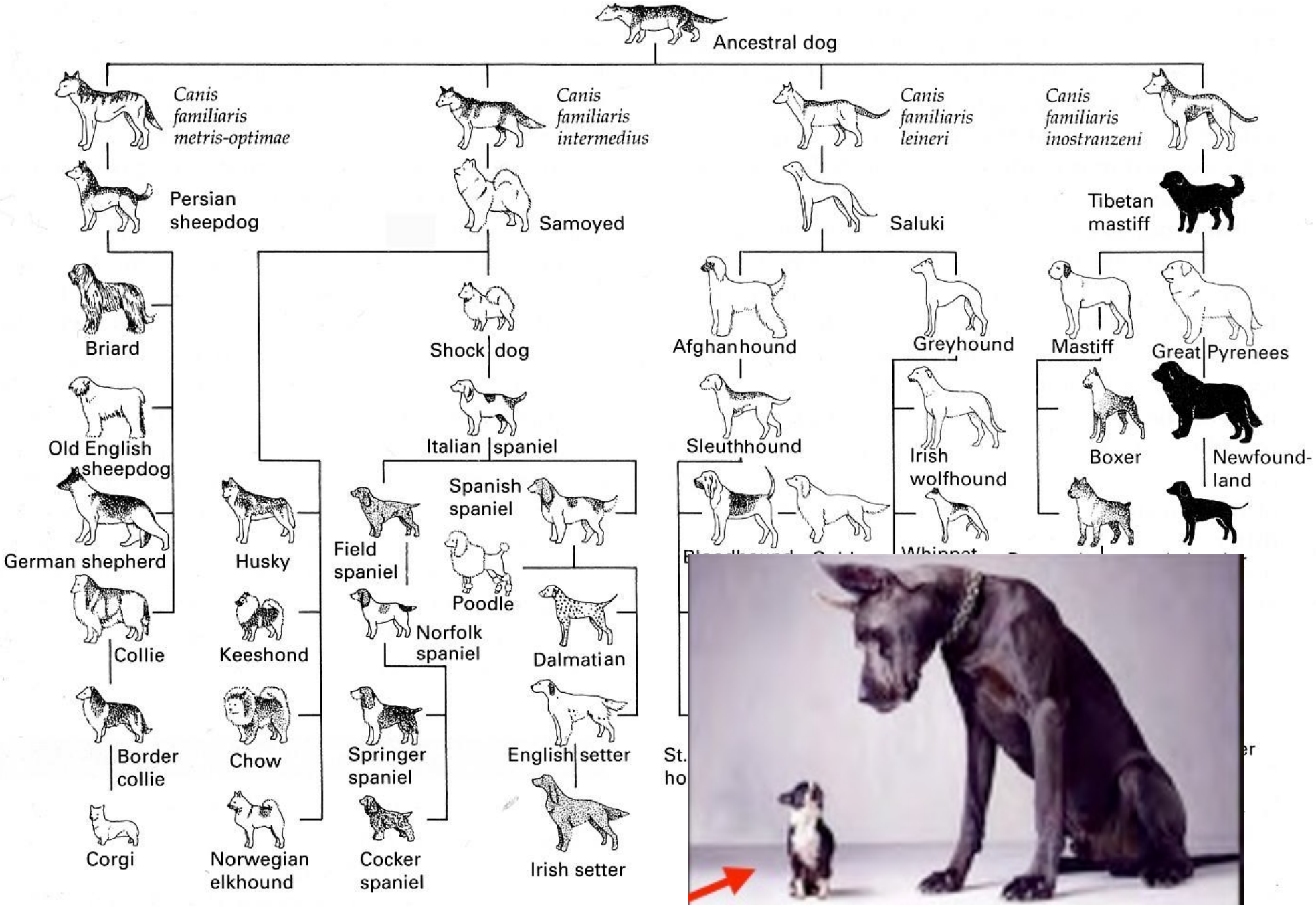
Lecture 4
Quantitative traits and
Mapping, Mapping, Mapping!!!

Quantitative Traits

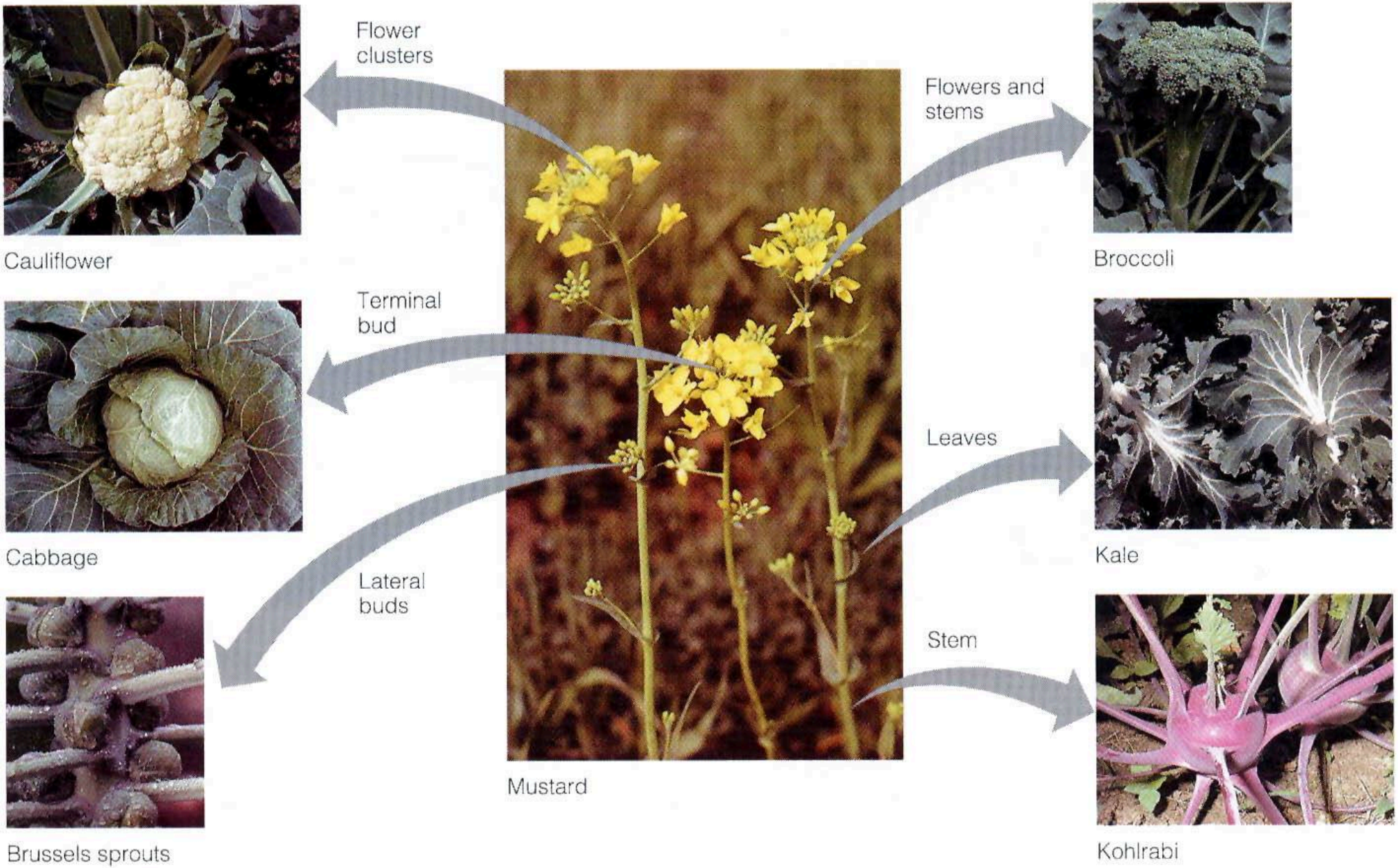
…traits that show a continuous variation in phenotype over a range,

…often result from multiple genes and are further termed polygenic traits.


Ancestral dog
Canis familiaris metris-optimae
Canis familiaris intermedius
Canis familiaris leineri
Canis familiaris inostranzeni
Persian sheepdog
Samoyed
Saluki
Tibetan mastiff
Briard
Shock dog
Afghanhound
Greyhound
Mastiff
Great Pyrenees
Old English sheepdog
Italian spaniel
Sleuthhound
Irish wolfhound
Boxer
Newfound-land
German shepherd
Husky
Spanish spaniel
Field spaniel
Poodle
Collie
Keeshond
Norfolk spaniel
Dalmatian
Border collie
Chow
Springer spaniel
English setter
Corgi
Norwegian elkhound
Cocker spaniel
Irish setter

Brassica oleracea

All these foods are the same species



Discontinuous Traits

- Discontinuous Traits,
  - exhibit only a few distinct phenotypes and can be described in a qualitative manner,

- Mendel and others worked with "inbred" lines,

```
Father:  DD BB cc aa EE FF GG
Mother:  dd BB cc aa EE FF GG
```

A Continuous Trait with a major component. Dog breed size

- "Outbred" transmission,

Father:	dd	BB	CC	AA	EE	FF GG
Mother:	DD	bb	CC	aa	EE	ff gg

- Continuous Traits,
  - Display a spectrum of phenotypes, and must be described in quantitative terms,

https://doi.org/10.1016/j.cub.2021.12.036

https://doi.org/10.1016/j.cub.2021.12.036

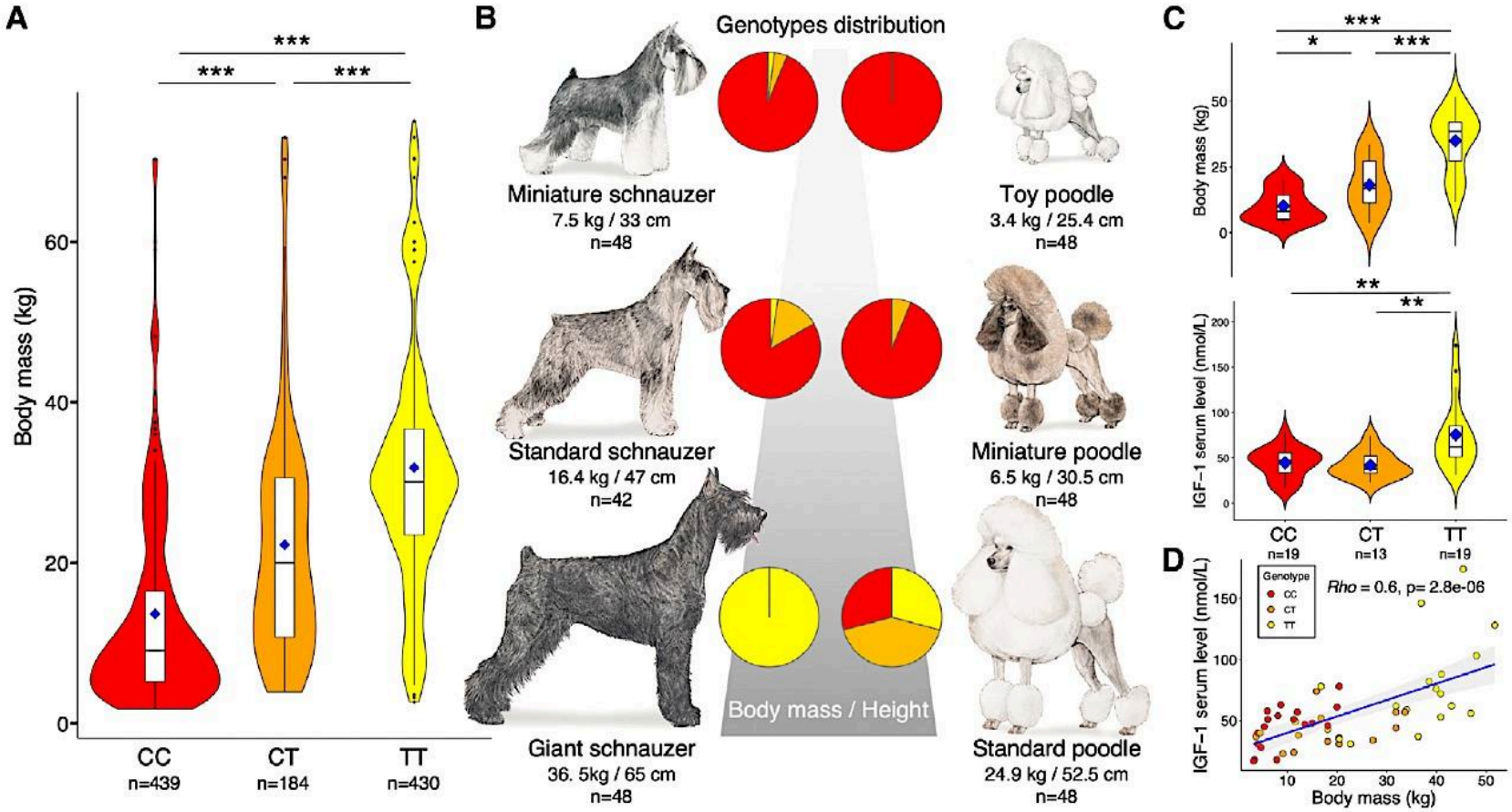


Figure 1. Insulin-like growth factor 1 (*IGF1*) in *Canidae*

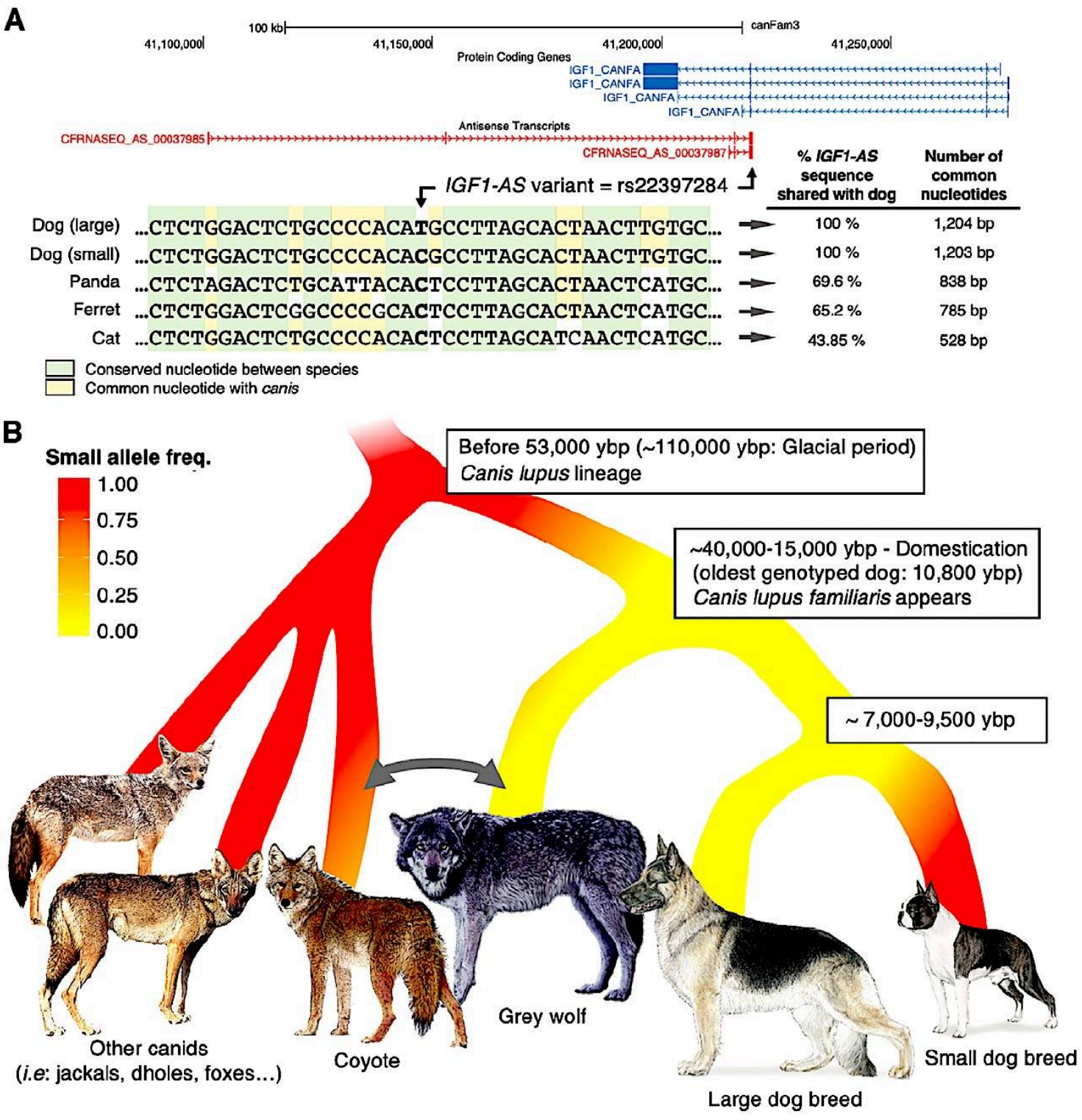
(A) *IGF1-AS* variant genotypes and body mass range collected from 1,162 dogs of 230 breeds. Dots represent outliers. Blue diamonds indicate breed body mass averages, boxplots represent interquartile ranges, and black horizontal bars show median for each (***p < 0.0001, Mann-Whitney-Wilcoxon tests).

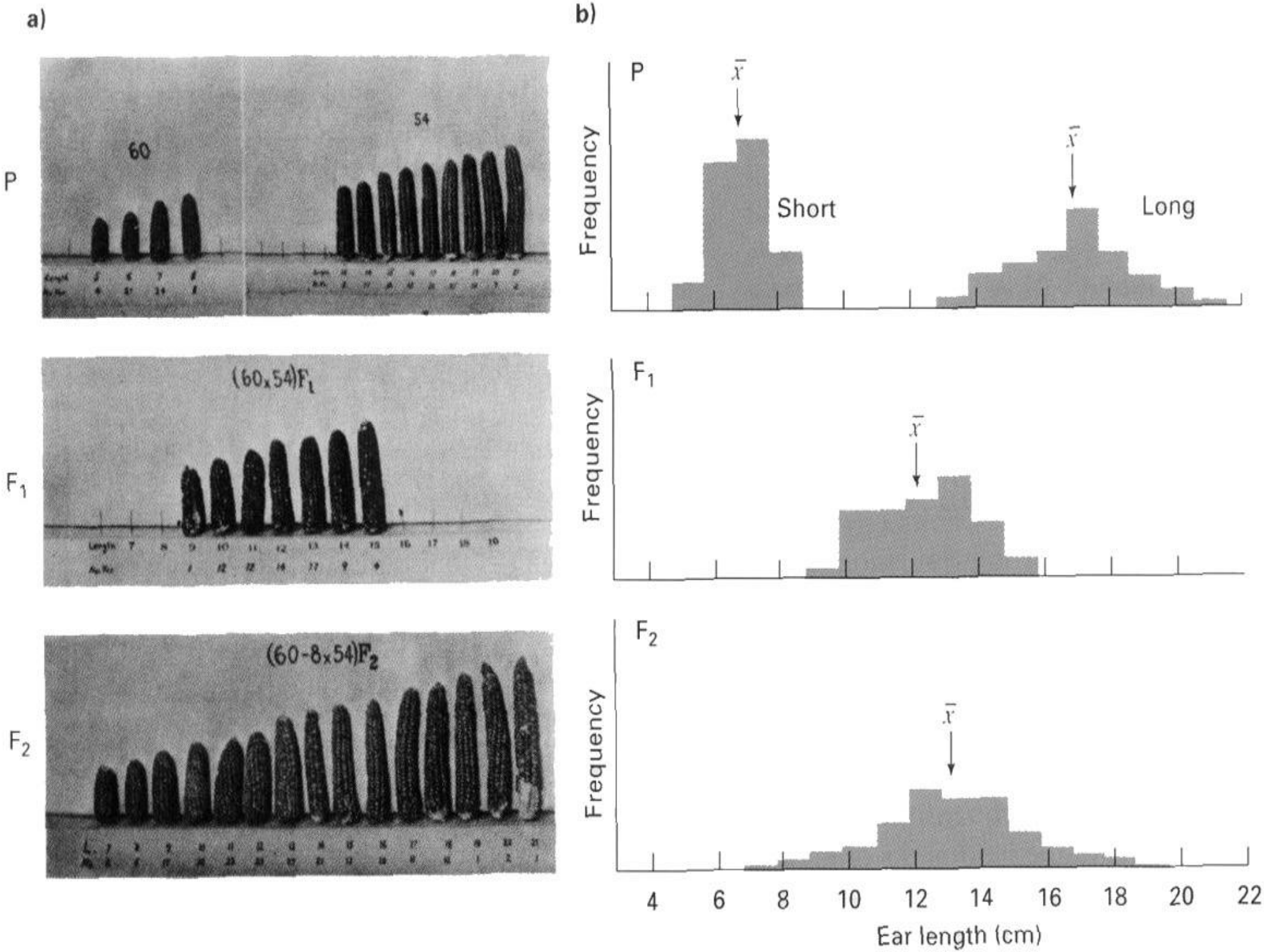
(B) Distribution of *IGF1-AS* alleles in three schnauzer breeds and poodle varieties. Pie chart indicates population proportion based on genotypes. Red, CC; orange, CT; yellow, TT.

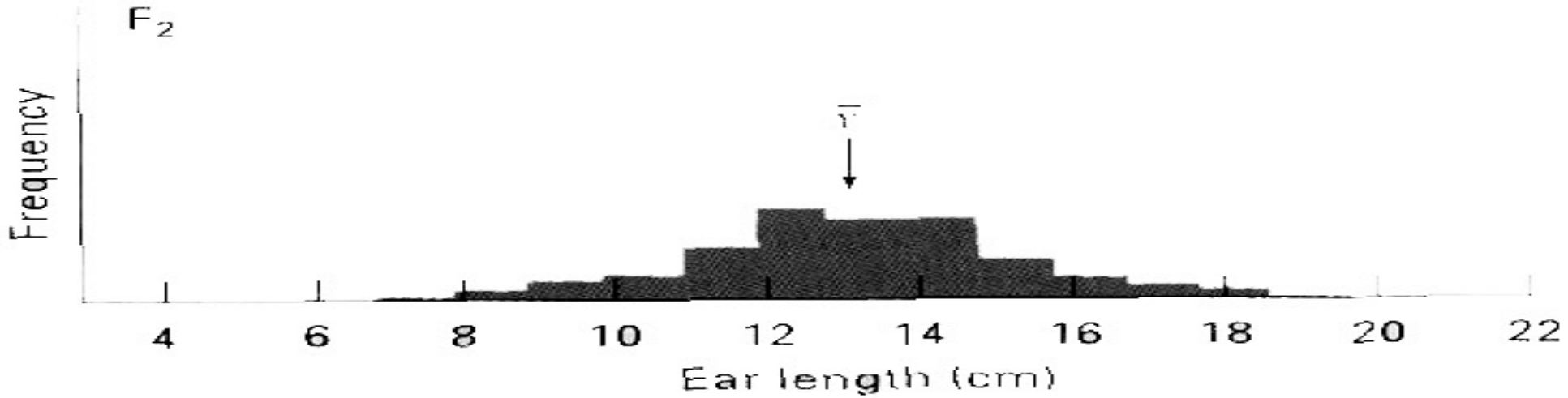
(C) Body mass and serum levels of IGF-1 protein (nmol/L) as functions of *IGF1-AS* genotype. IGF-1 serum protein levels were assayed in 51 dogs, including 13 mixed-breed dogs (*p < 0.01, **p < 0.001, ***p < 0.0001, Mann-Whitney-Wilcoxon tests).

(D) Positive correlation observed between body mass and IGF-1 serum level (*Rho* Spearman test). Blue line shows the regression line; gray area represents confidence interval.

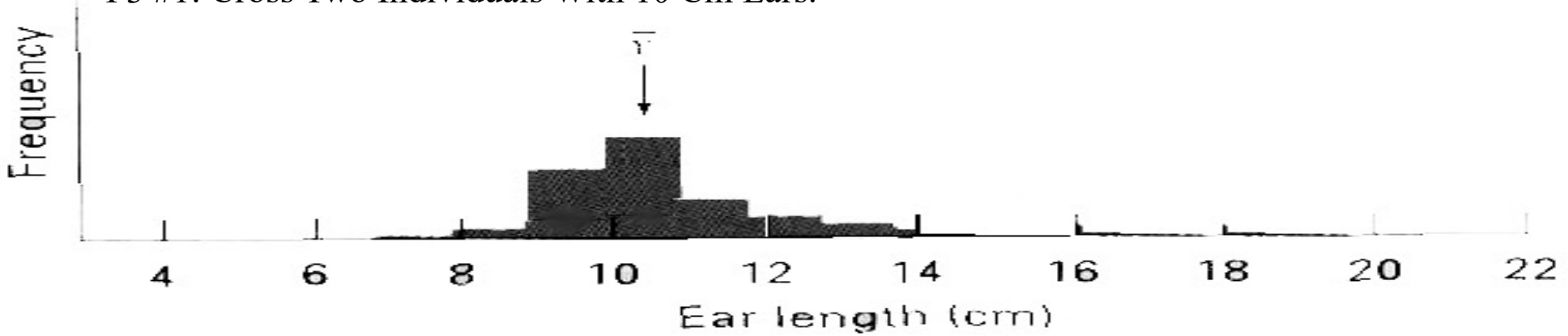
See also Table S1 and Data S1.


A
100 kb
canFam3
41,100,000
41,150,000
41,200,000
41,250,000
Protein Coding Genes
IGF1_CANFA
IGF1_CANFA
IGF1_CANFA
IGF1_CANFA
Antisense Transcripts
CFRNASEQ_AS_00037985
CFRNASEQ_AS_00037987
IGF1-AS variant = rs22397284
% IGF1-AS sequence shared with dog
Number of common nucleotides
Dog (large) ...CTCTGGACTCTGCCCCACATGCCTTAGCACTAACTTGTGC... 100 % 1,204 bp
Dog (small) ...CTCTGGACTCTGCCCCACACGCCTTAGCACTAACTTGTGC... 100 % 1,203 bp
Panda ...CTCTAGACTCTGCATTACACTCCTTAGCACTAACTCATGC... 69.6 % 838 bp
Ferret ...CTCTGGACTCGGCCCCGCACTCCTTAGCACTAACTCATGC... 65.2 % 785 bp
Cat ...CTCTGGACTCTGCCCCACACTCCTTAGCATCAACTCATGC... 43.85 %
528 bp
Conserved nucleotide between species
Common nucleotide with canis
B
Small allele freq.
1.00
0.75
0.50
0.25
0.00
Before 53,000 ybp (~110,000 ybp: Glacial period)
Canis lupus lineage
~40,000-15,000 ybp - Domestication
(oldest genotyped dog: 10,800 ybp)
Canis lupus familiaris appears
~ 7,000-9,500 ybp
Other canids
(i.e: jackals, dholes, foxes...)
Coyote
Grey wolf
Large dog breed
Small dog breed

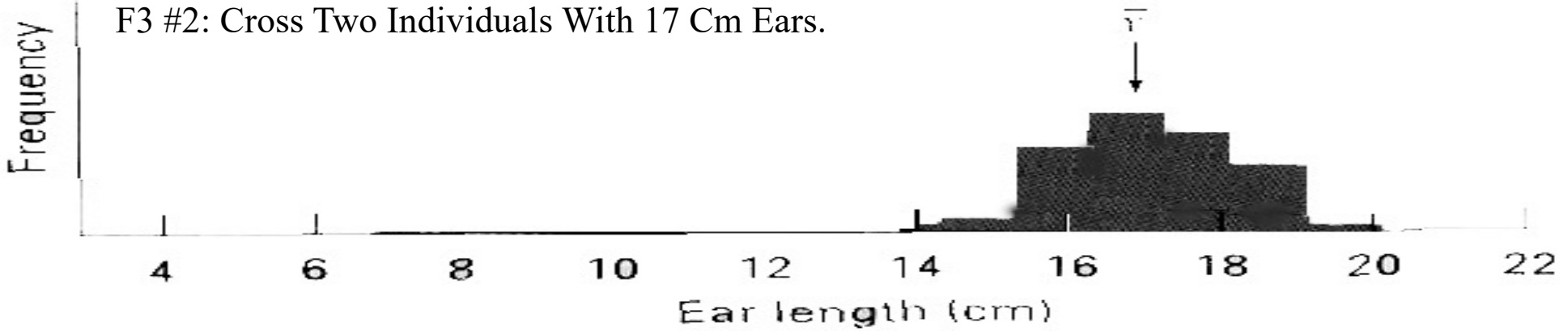

a)
b)
P
F1
F2
60
54
(60x54)F1
(60-8x54)F2
Frequency
Short
Long
Ear length (cm)
4
6
8
10
12
14
16
18
20
22



F3 #1: Cross Two Individuals With 10 Cm Ears.

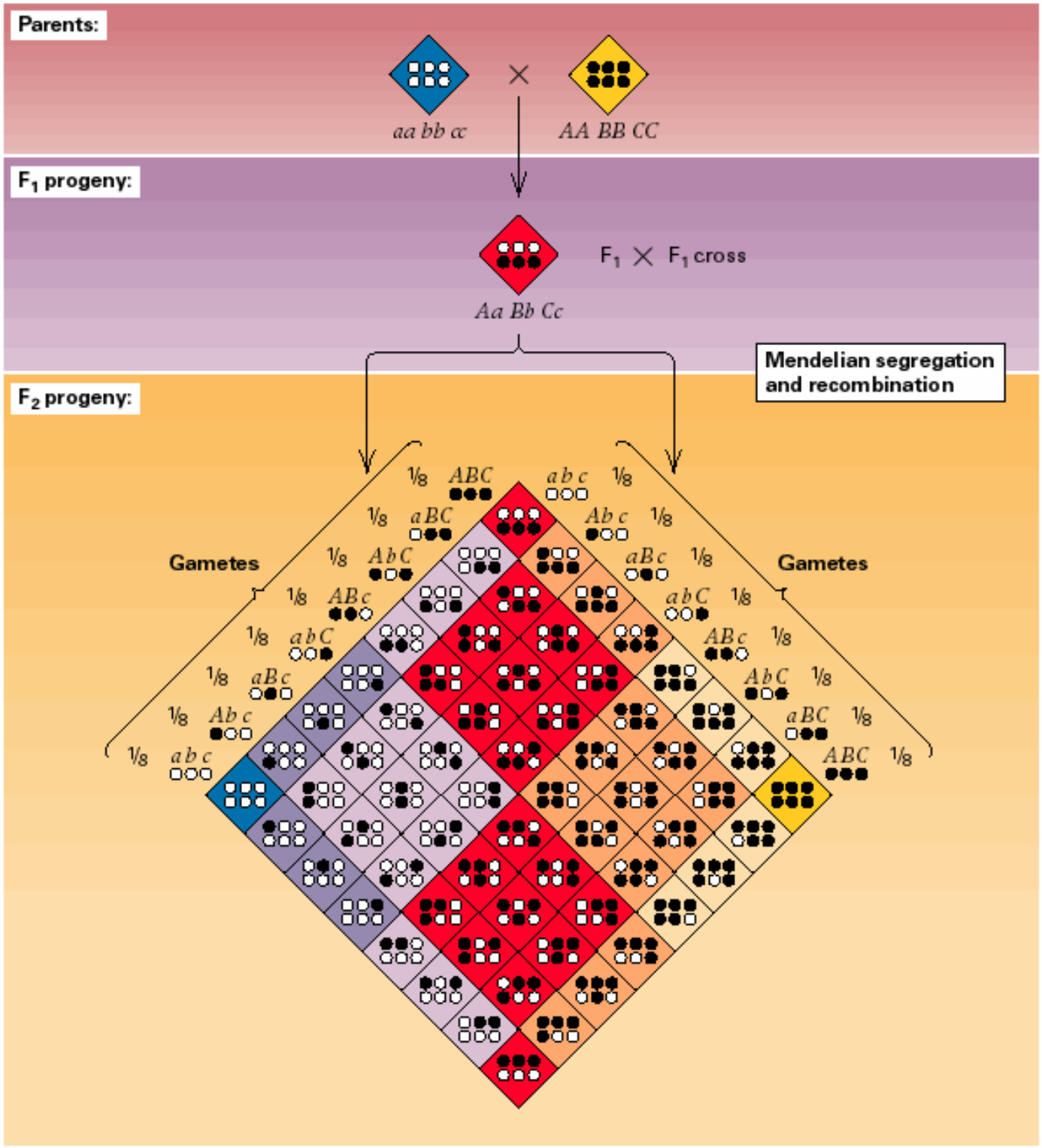


F3 #2: Cross Two Individuals With 17 Cm Ears.

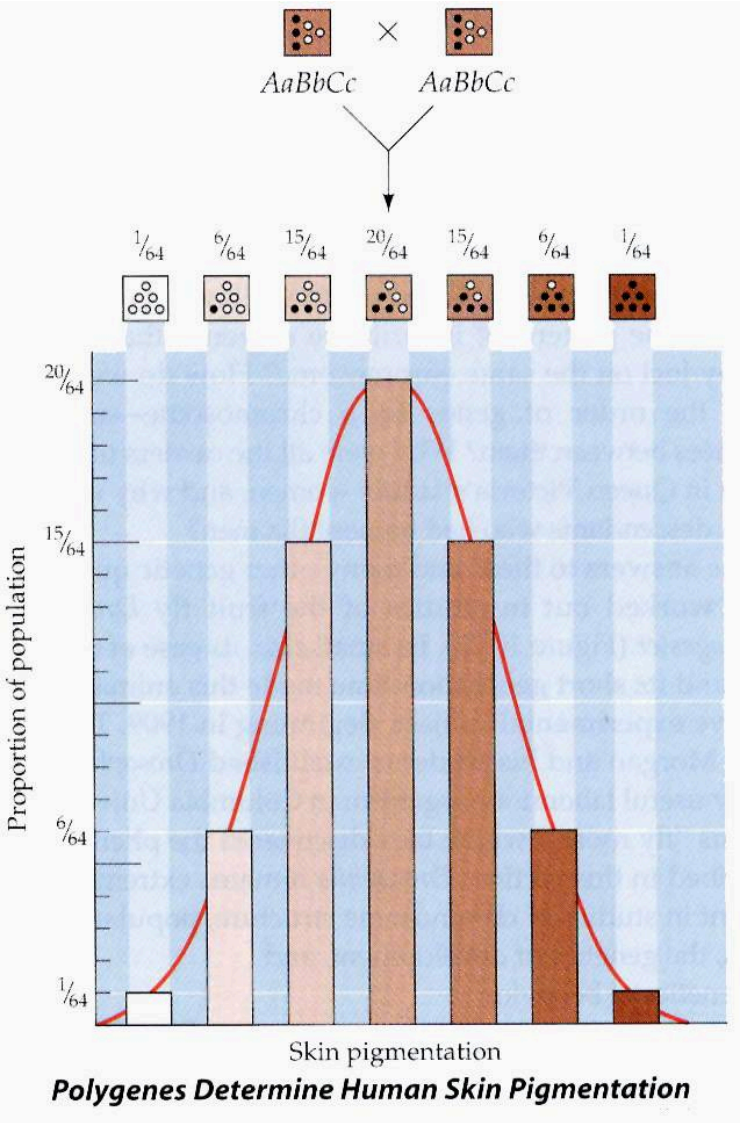


Polygenic Traits and Mendel

- It is possible to provide a Mendelian explanation for continuous variation by considering numbers of genes contributing to a phenotype,
  - the more genes, the more phenotypic categories,
  - the more categories, the more the variation seems continuous.


Parents:
aa bb cc
AA BB CC
F1 progeny:
F1 × F1 cross
Aa Bb Cc
Mendelian segregation and recombination
F2 progeny:
Gametes
Gametes
1/8 ABC
1/8 aBC
1/8 AbC
1/8 ABc
1/8 abC
1/8 aBc
1/8 Abc
1/8 abc
1/8 abc
1/8 Abc
1/8 aBc
1/8 abC
1/8 ABc
1/8 AbC
1/8 aBC
1/8 ABC

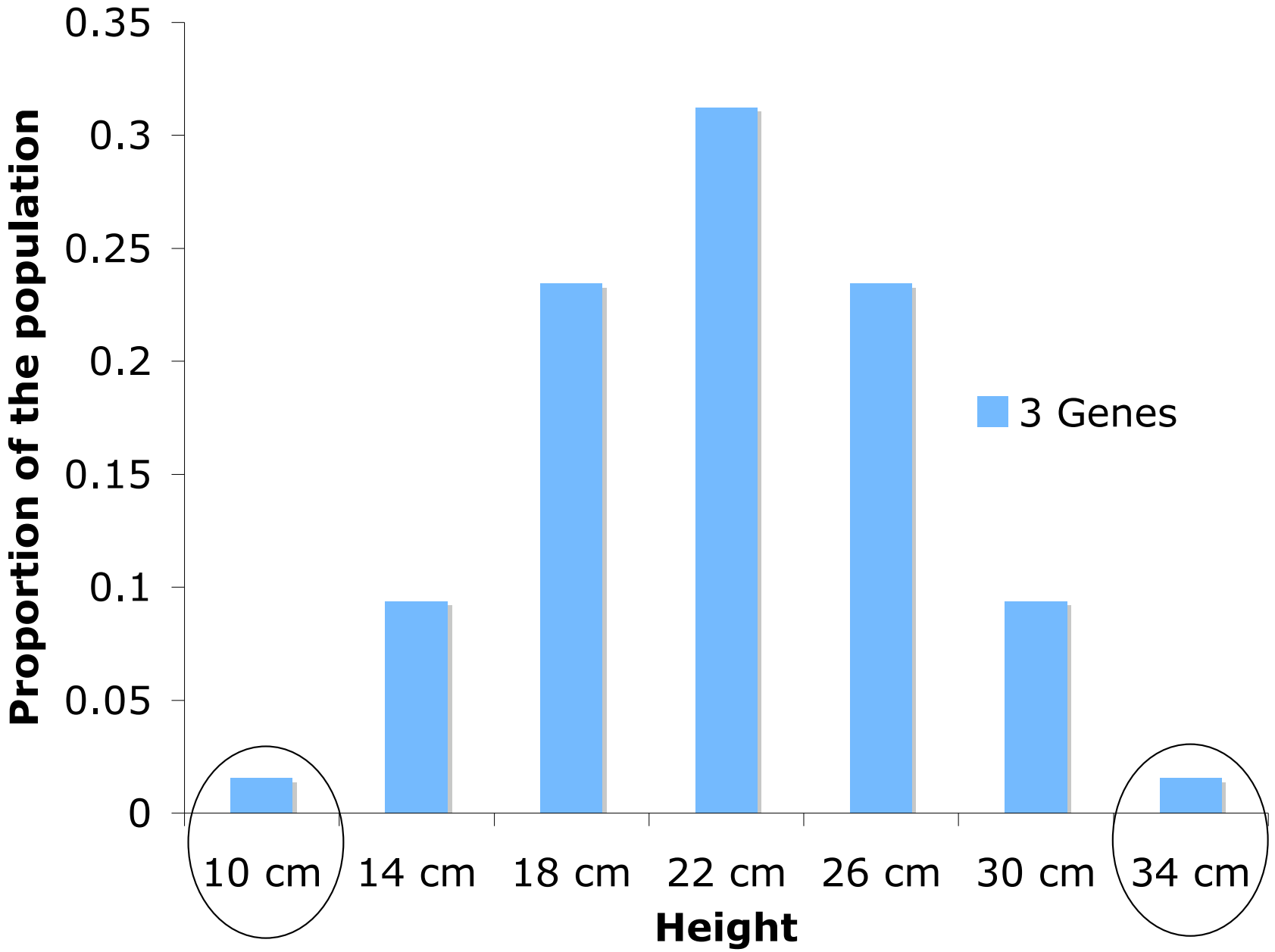
Melanin Pigmentation in twins

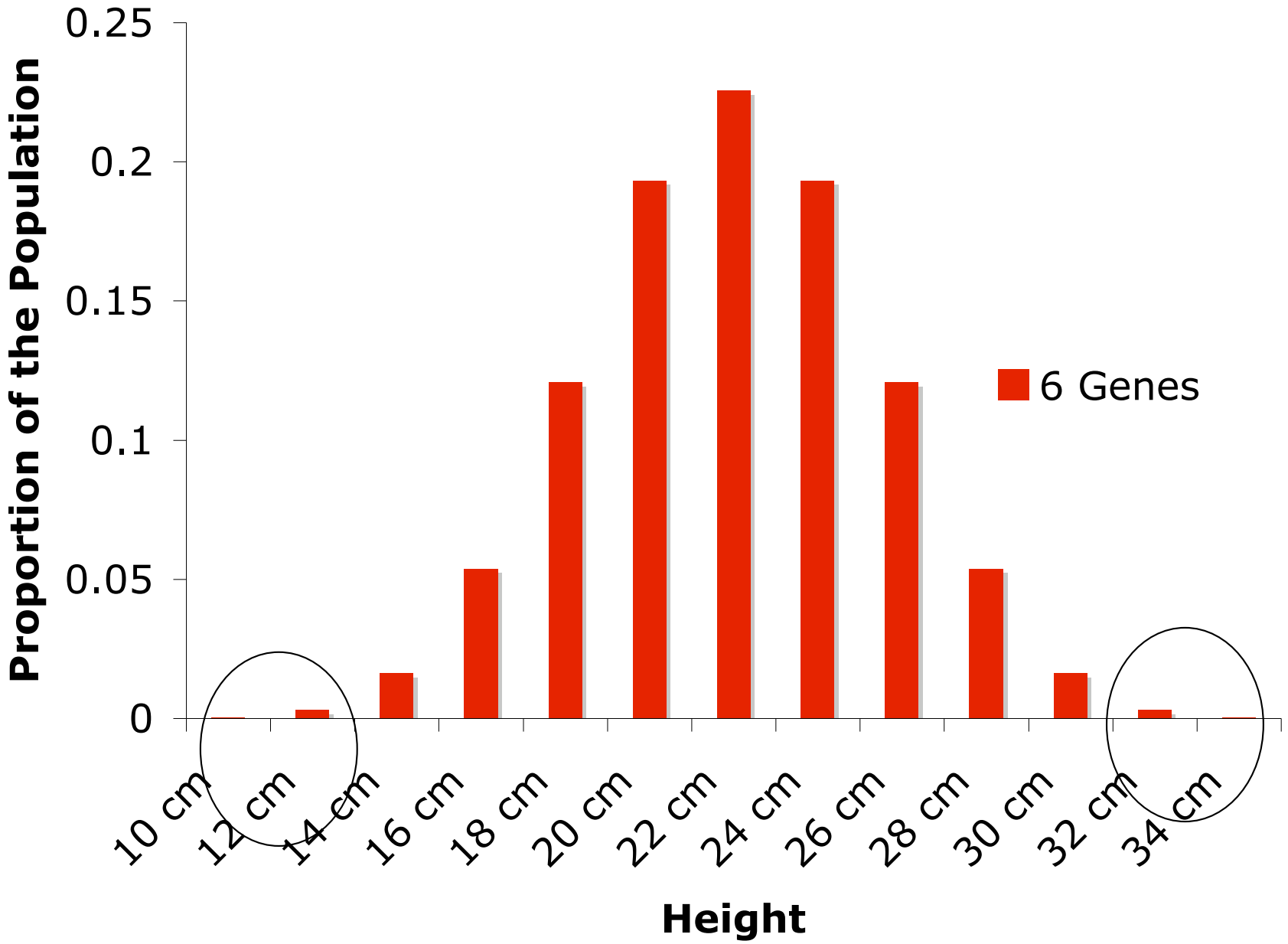


Polygenes Determine Human Skin Pigmentation

Twins, One White and One Black, Get Ready to Start Middle School: 'I Notice People Doing Double-Takes'

Inside Edition 9 hours ago


0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Proportion of the population
3 Genes
10 cm
14 cm
18 cm
22 cm
26 cm
30 cm
34 cm
Height


Proportion of the Population
0.25
0.2
0.15
0.1
0.05
0
6 Genes
10 cm
12 cm
14 cm
16 cm
18 cm
20 cm
22 cm
24 cm
26 cm
28 cm
30 cm
32 cm
34 cm
Height

Continuous Variation

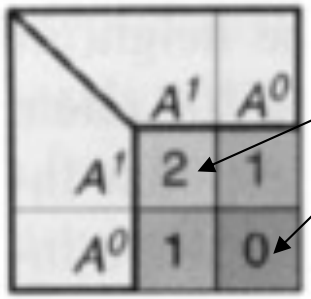
- Two or more genes that influence the same phenotype, often in an additive way,
  - **additive allele:** contributes a set amount to the phenotype,
  - **non-additive allele:** does not contribute to the phenotype.
- The affect of each allele on the phenotype is relatively small, and roughly equivalent,
- Substantial variation is observed when multiple genes control a single trait,
- Must be studied in large populations.

Calculating the Number of Genes

- If you know the frequency of *either* extreme trait, then the number of genes (*n*) can be calculated using the formulae…

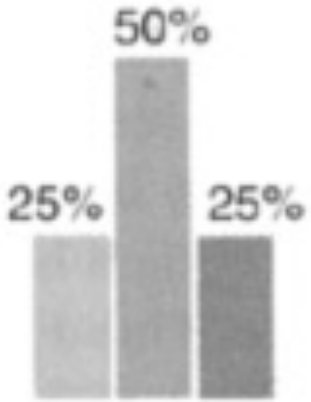
$$\frac{1}{4^n} = \text{ratio of F2 individuals expressing } \underline{\text{either}} \text{ extreme phenotype.}$$

The number of genes: n = 1



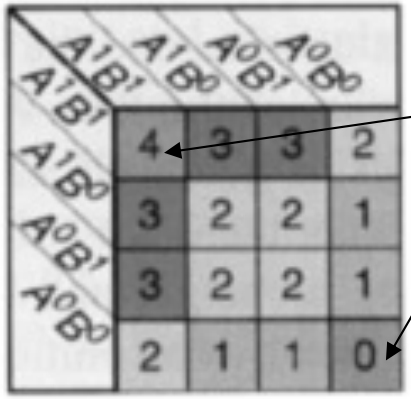
1/4 show either extreme trait.

$$\frac{1}{4^n} = \frac{1}{4^1}$$



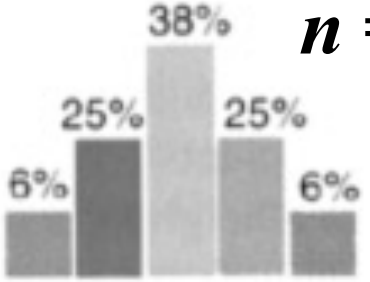
$n = 1$

The number of genes: n = 2



1/16 show either Extreme trait.

$$\frac{1}{4^n} = \frac{1}{16} = \frac{1}{4^2}$$



$n = 2$

- **Inbred strain #1, mean height = 24 cm,**
- **Inbred strain #2, mean height = 24 cm,**
- **F1 mean height = 24 cm,**
- **F2, 12 - 36 cm range, mean = 24 cm,**
  - **4/1000 are 12 cm, smallest class**
  - **4/1000 are 36 cm, largest class**

a. What mode of inheritance? **Quantitative.**

b. How many gene pairs? $1/4^n = 1/250$, **n = 4**

4/1000 = 1/250 are the extreme class,

solve $1/4^n$ **= ratio of extreme class**

$4^2 = 16$, $4^3 = 64$, $4^4 = 256$, **etc.**

- **Inbred strain #1, mean height = 24 cm,**
- **Inbred strain #2, mean height = 24 cm,**
- **F1 mean height = 24 cm,**
- **F2, 12 - 36 cm range, mean = 24 cm,**
  - **4/1000 are 12 cm, smallest class**
  - **4/1000 are 36 cm, largest class**

c. How much does each allele contribute?

range: largest (36 cm) - smallest (12 cm) = 24 cm

alleles: 4 genes, 8 potential additive alleles

24 cm / 8 additive alleles = 3 cm / additive allelle

- **Inbred strain #1, mean height = 24 cm,**
- **Inbred strain #2, mean height = 24 cm,**
- **F1 mean height = 24 cm,**
- **F2, 12 - 36 cm range, mean = 24 cm,**
  - **4/1000 are 12 cm, smallest class**
  - **4/1000 are 36 cm, largest class**

d. Indicate one possible set of P1, F1.

Base Height is 12 cm, so each P1, F1 has 4 additive alleles.

P1 AABBccdd x aabbCCDD

F1 AaBbCcDd x AaBbCcDd

Research Article

Common DNA Variants Accurately Rank an Individual's Extreme Height

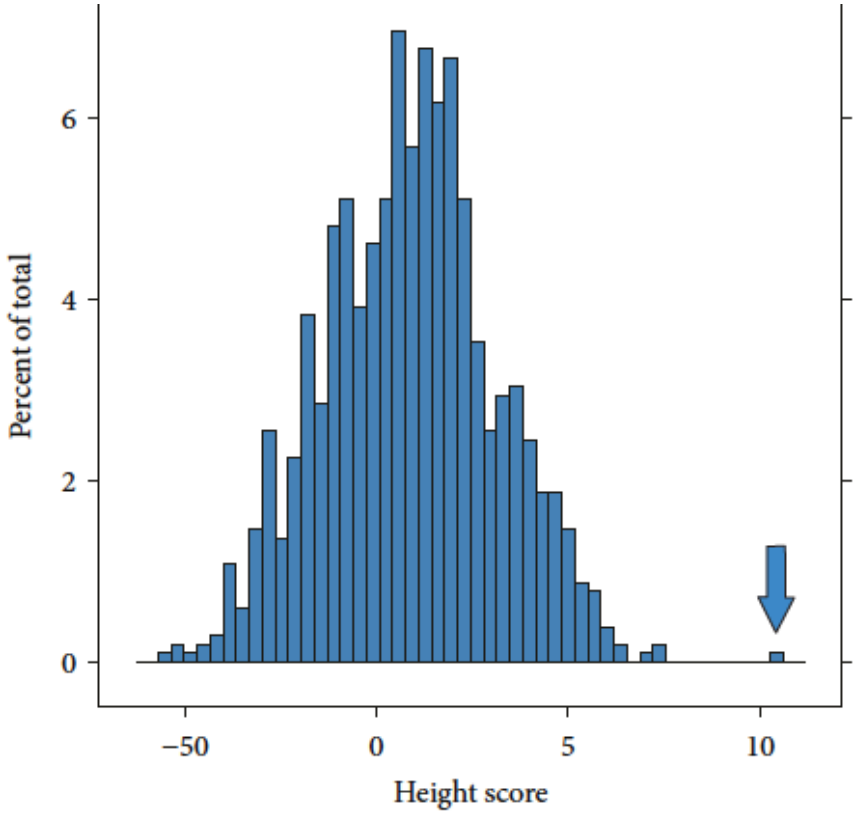


FIGURE 2: Height score distribution calculated using the 2910 SNPs. Mr. Bradley's height score (10.32, indicated by the arrow) ranked highest when compared to the 1020 individuals from ADNI and Cache County, while the next highest was 7.43. The mean height score within the ADNI and Cache County data was 0.98 with a standard deviation of 2.22, making Mr. Bradley's height score 4.2 standard deviations above the mean.

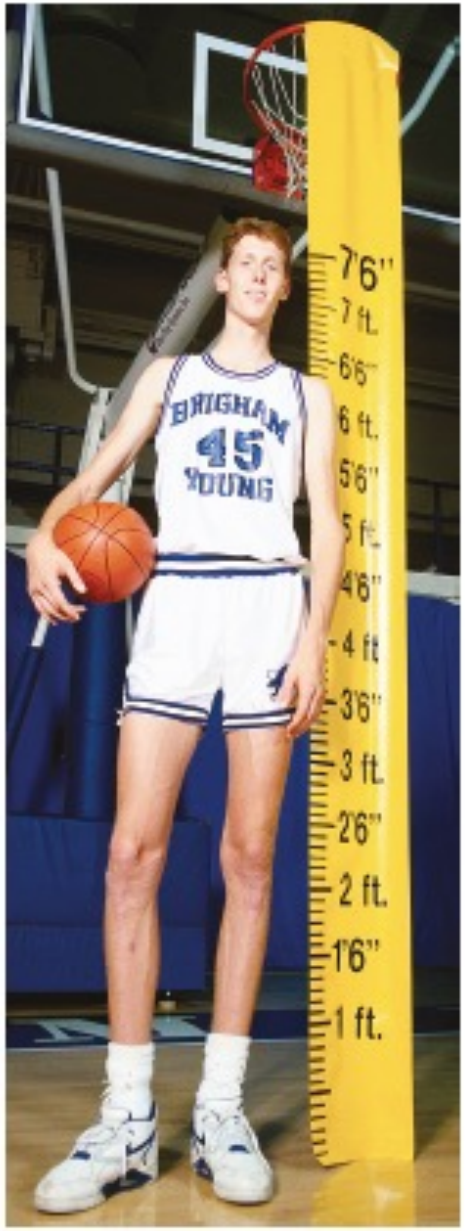
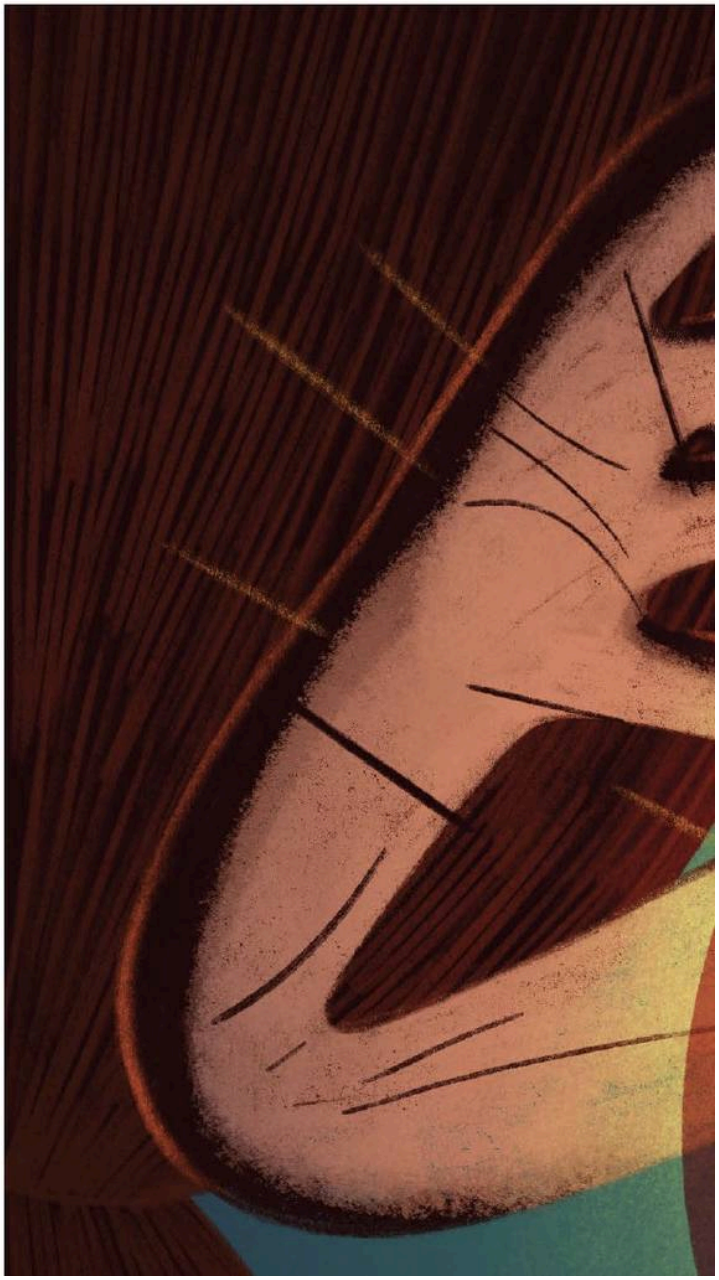


FIGURE 1: Shawn Bradley is 2.29 m (7' 6") tall with no known medical conditions. Mr. Bradley played basketball for Brigham Young University from 1990 to 1991. He played in the National Basketball Association from 1993–2005. Photo courtesy of BYU photography.

THE CONTROVERSIAL EMBRYO TESTS THAT PROMISE A HEALTHIER BABY

Some companies offer tests that rank embryos based on their risk of developing complex diseases such as schizophrenia or heart disease. Are the tests accurate — or ethical? **By Max Kozlov**

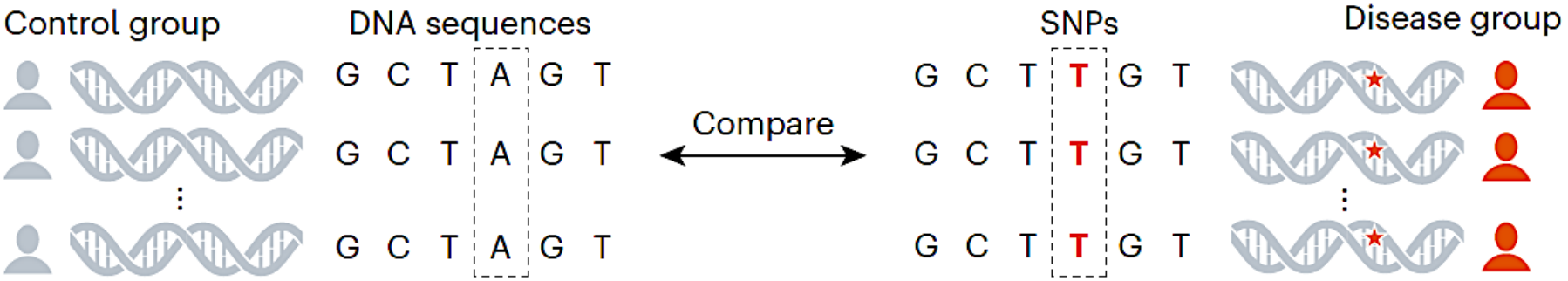


SCORING EMBRYOS

Some companies are selling genetic tests for embryos generated by *in vitro* fertilization that they say can identify an embryo with the lowest risk of developing common diseases, such as diabetes and some cancers. The method relies on sequencing the embryo's genome and comparing it with existing data on genetic risk factors in adult populations.

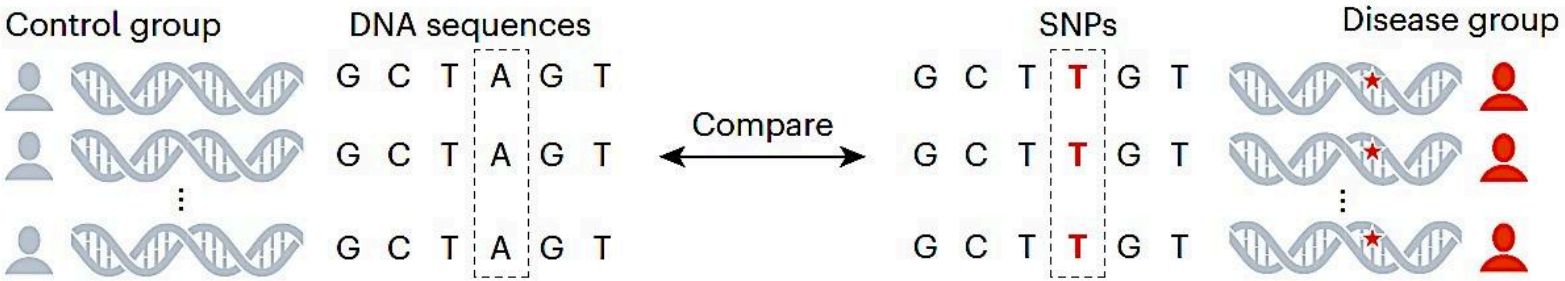
Genetics and risk

Researchers have discovered links between individual DNA letter changes, called single nucleotide polymorphisms (SNPs), and the risk of having a disease. To do this, they compare genome and health data from thousands of people with or without various diseases.

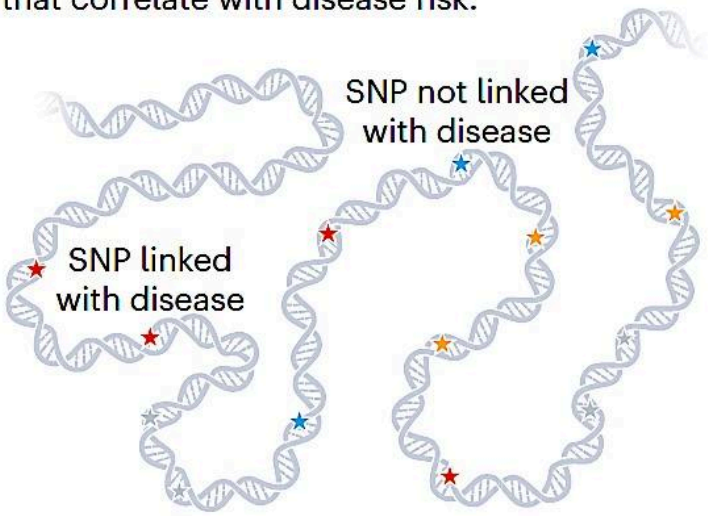


They look across the genome for SNPs that correlate with disease risk.

Combining information on all SNPs yields a prediction for how likely someone is to develop a disease.

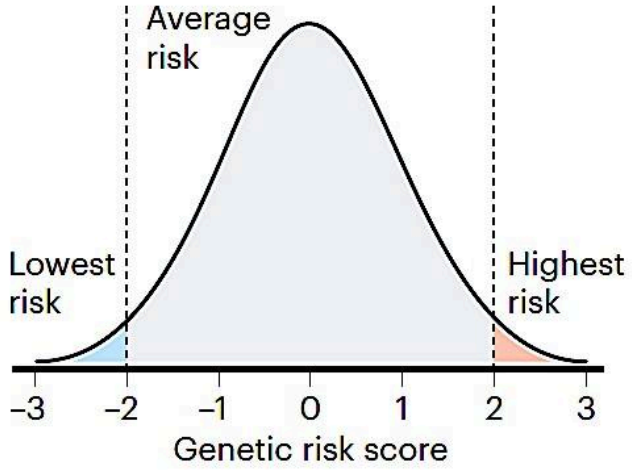


They look across the genome for SNPs that correlate with disease risk.



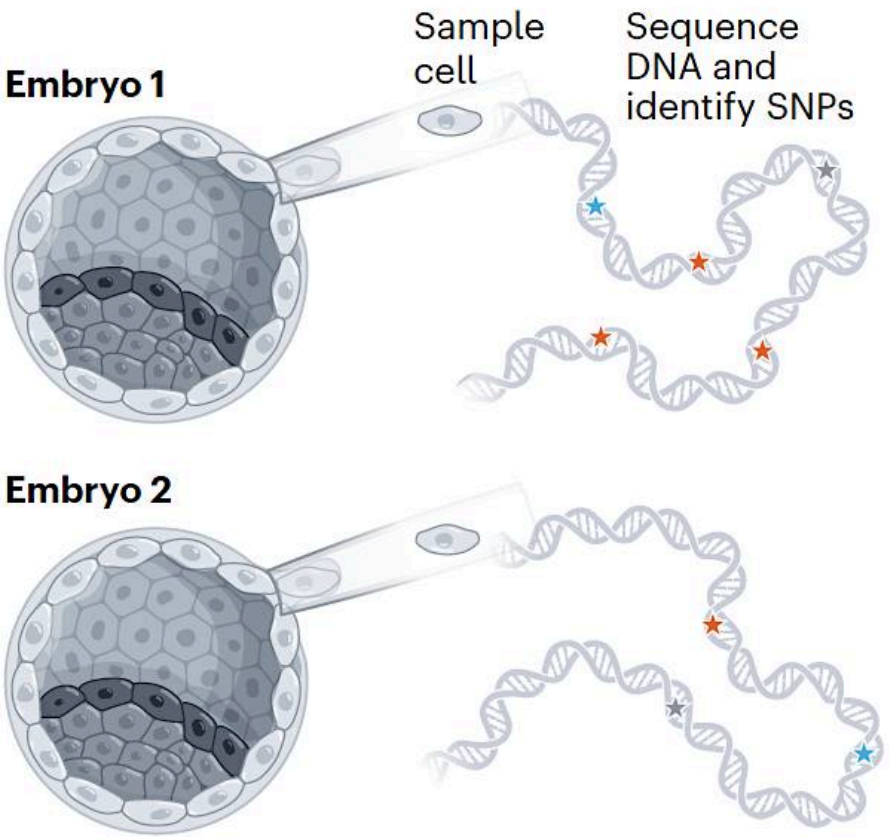
Control for age, sex and ancestry

Combining information on all SNPs yields a prediction for how likely someone is to develop a disease.



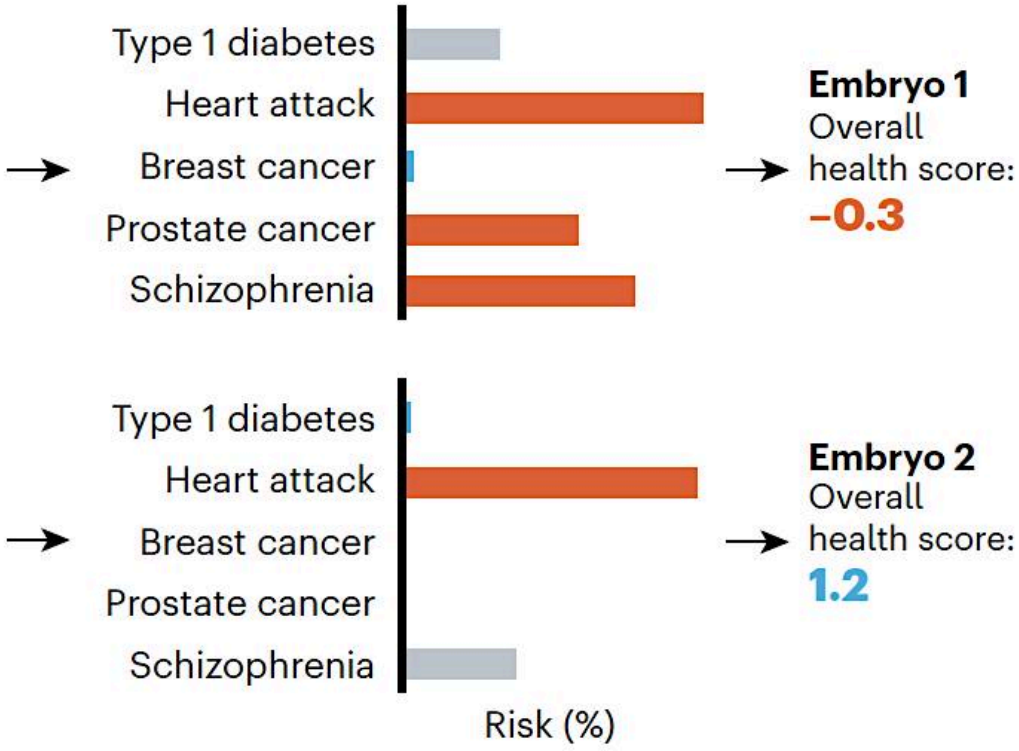
Embryo sampling

To apply this information to an embryo, clinicians take a few cells from embryos that are about 5 days old. They extract and sequence the DNA and look for SNPs.



Calculate scores

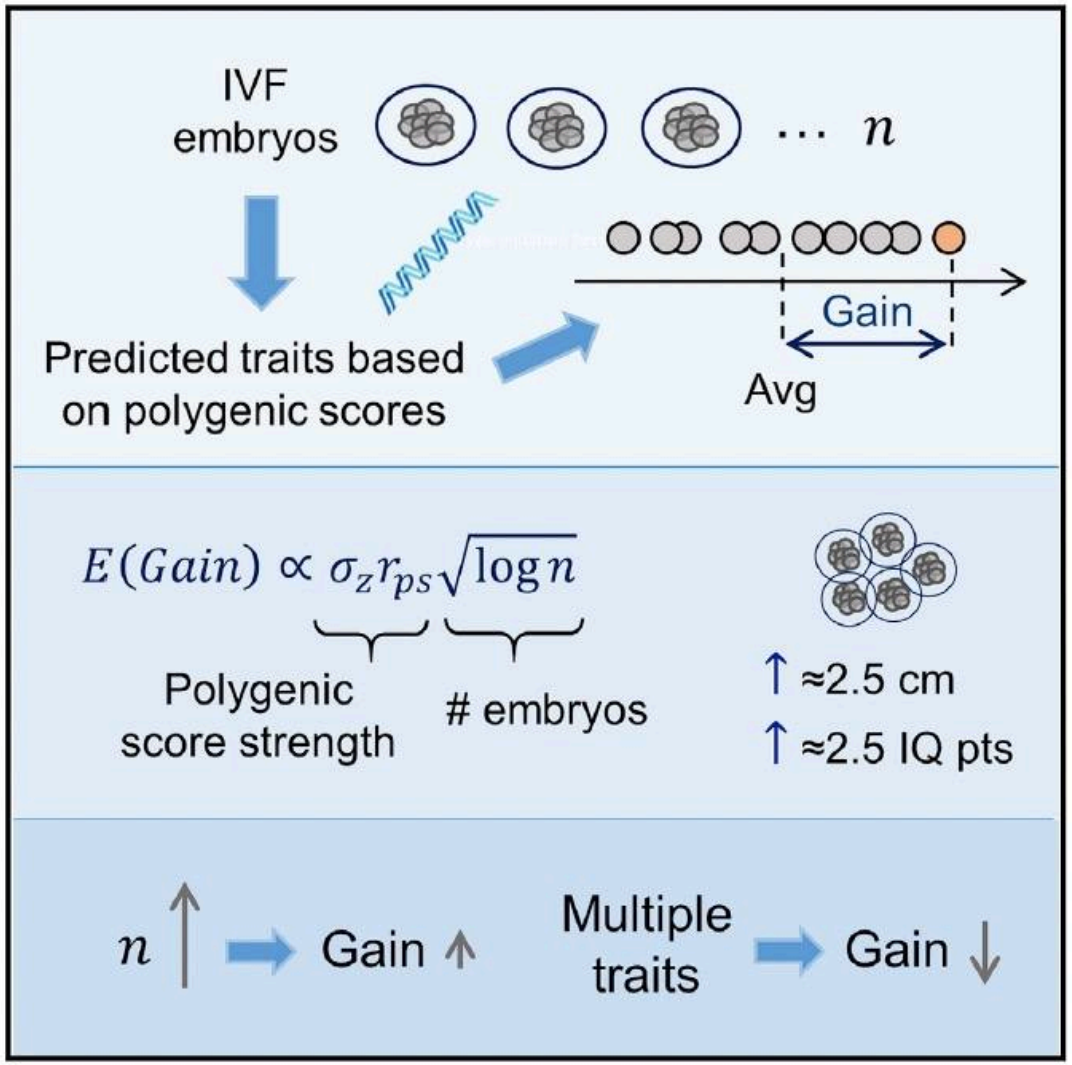
One company, Genomic Prediction, uses this method to help clients to identify embryos with a low risk of developing 12 disorders, including diabetes and some cancers. Each embryo's overall score is calculated by combining the risk of each disease and weighting them by their effect on life expectancy. A higher health score suggests a lower overall risk.



Cell https://doi.org/10.1016/j.cell.2019.10.033

Screening Human Embryos for Polygenic Traits Has Limited Utility

Graphical Abstract



Authors

Ehud Karavani, Or Zuk, Danny Zeevi, ..., Max Lam, Todd Lencz, Shai Carmi

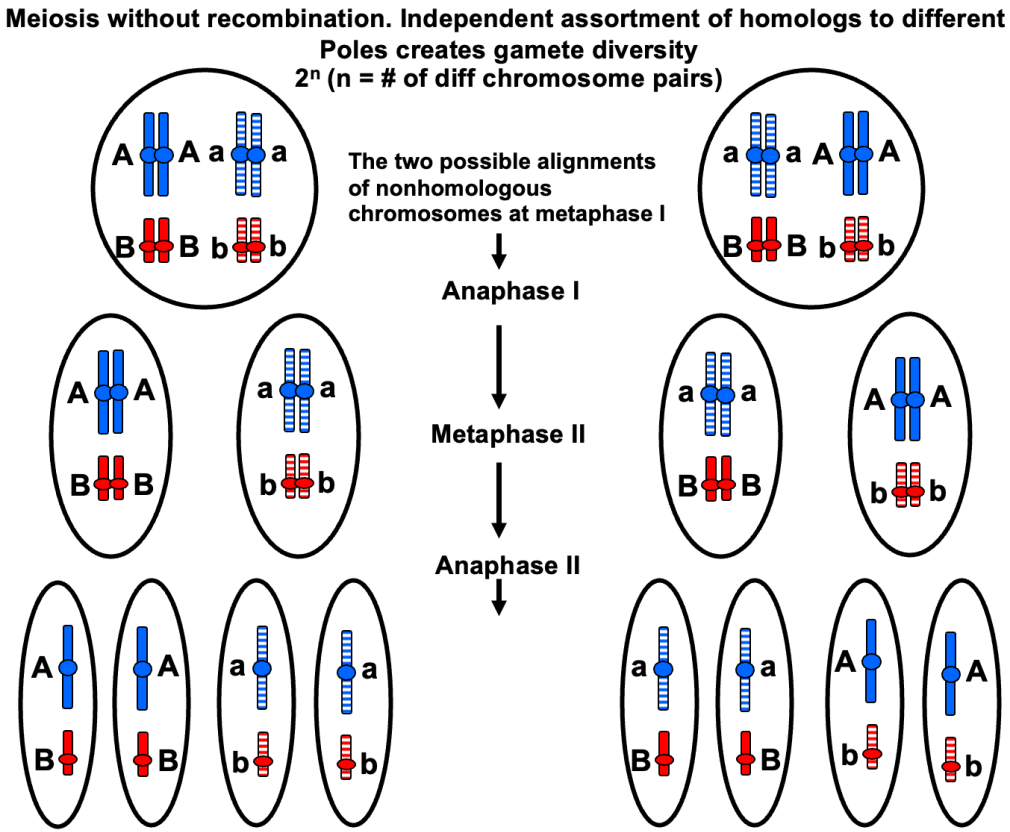
Correspondence

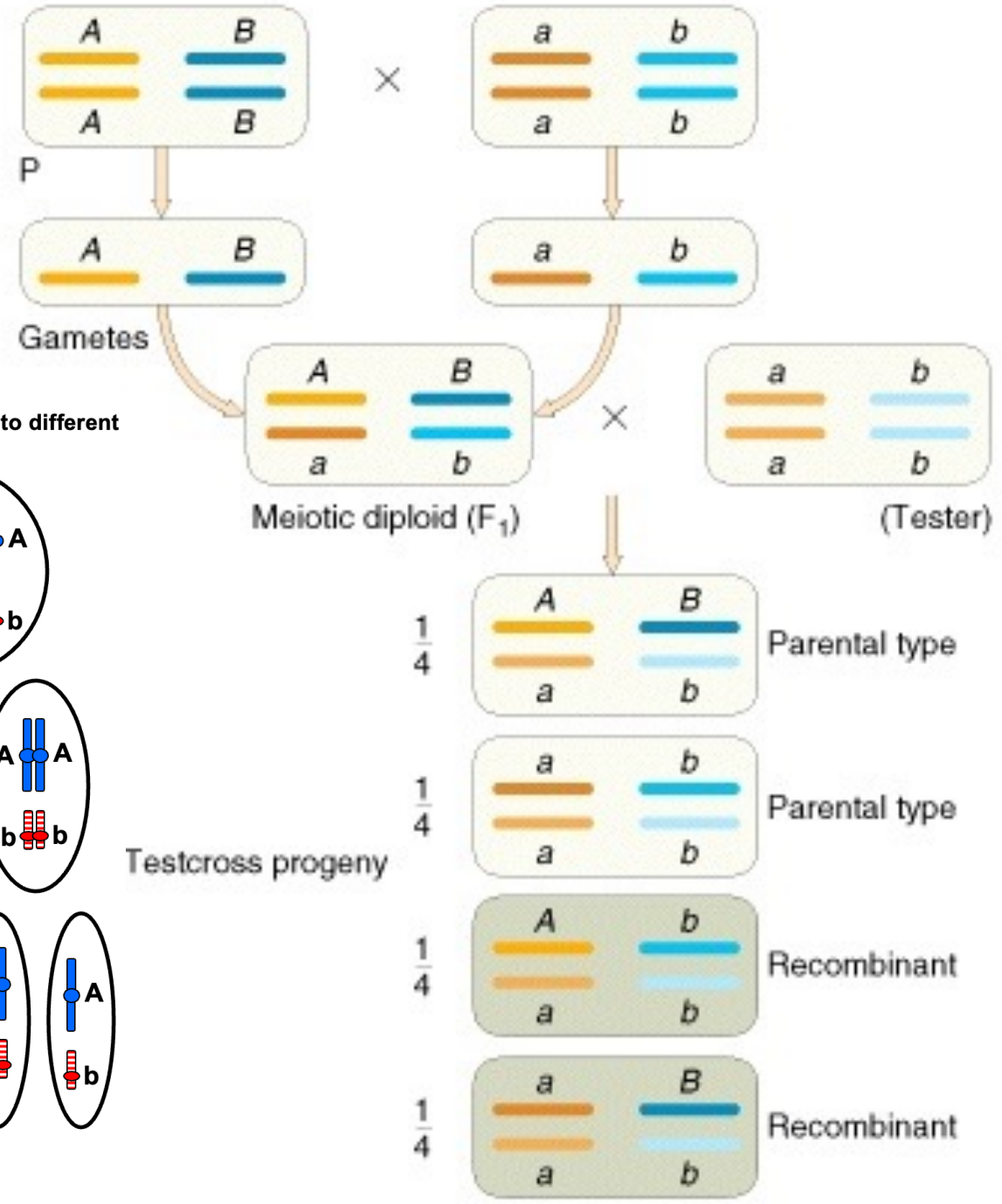
tlencz@northwell.edu (T.L.), shai.carmi@huji.ac.il (S.C.)

In Brief

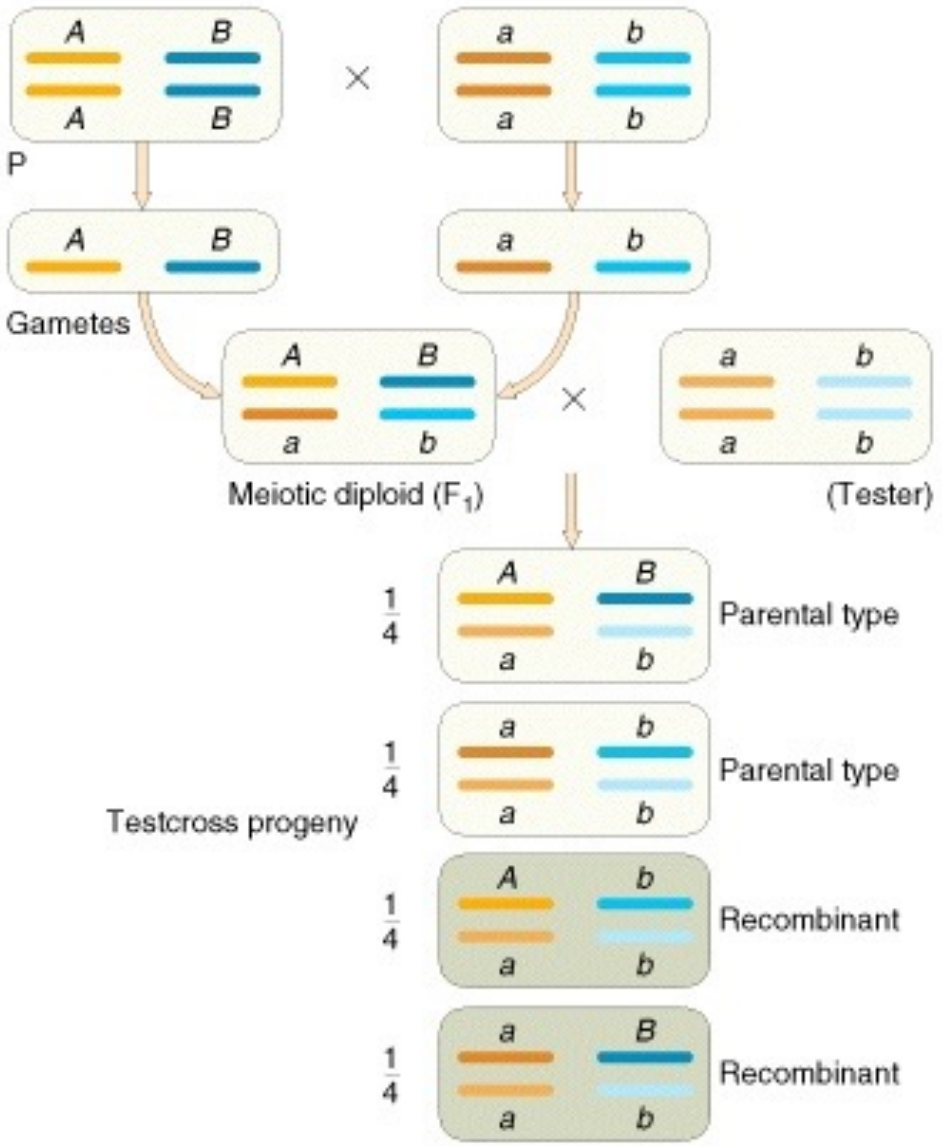
Recent progress in genetic testing of embryos has made it technically feasible to profile IVF embryos for polygenic traits such as height or IQ, but simulations, models, and empirical data show that the gain in trait value when selecting the top-scoring embryo is currently limited and uncertain.

Independent assortment

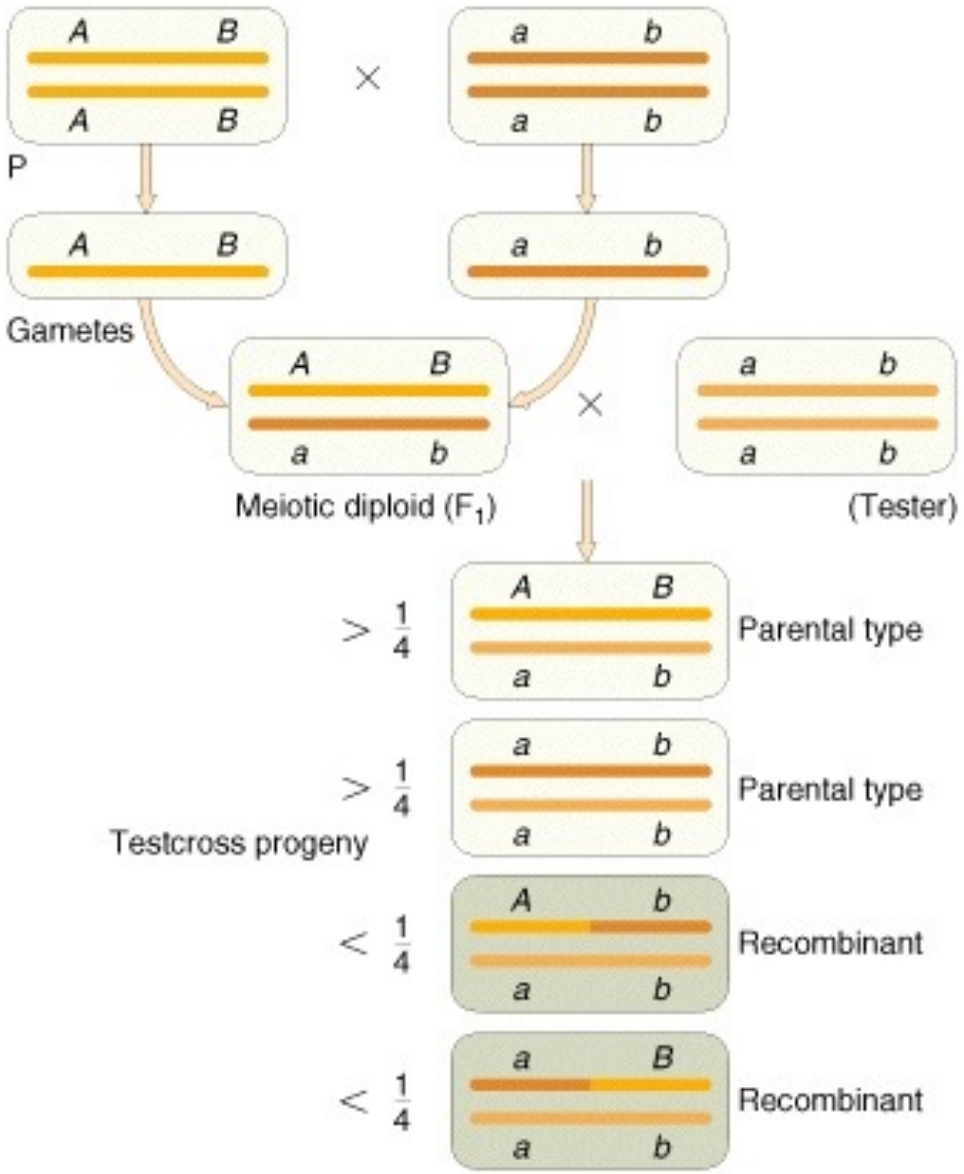




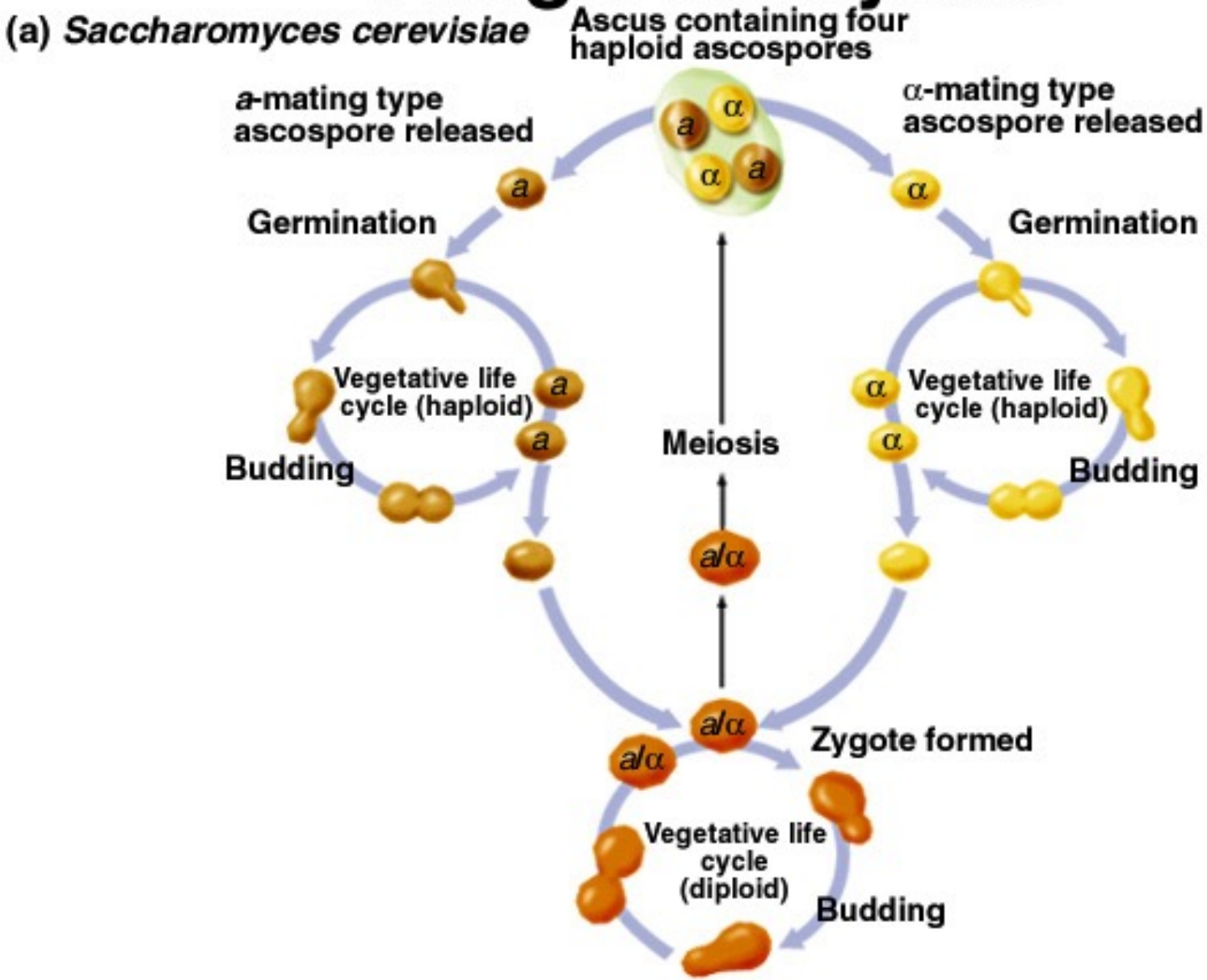
Independent assortment



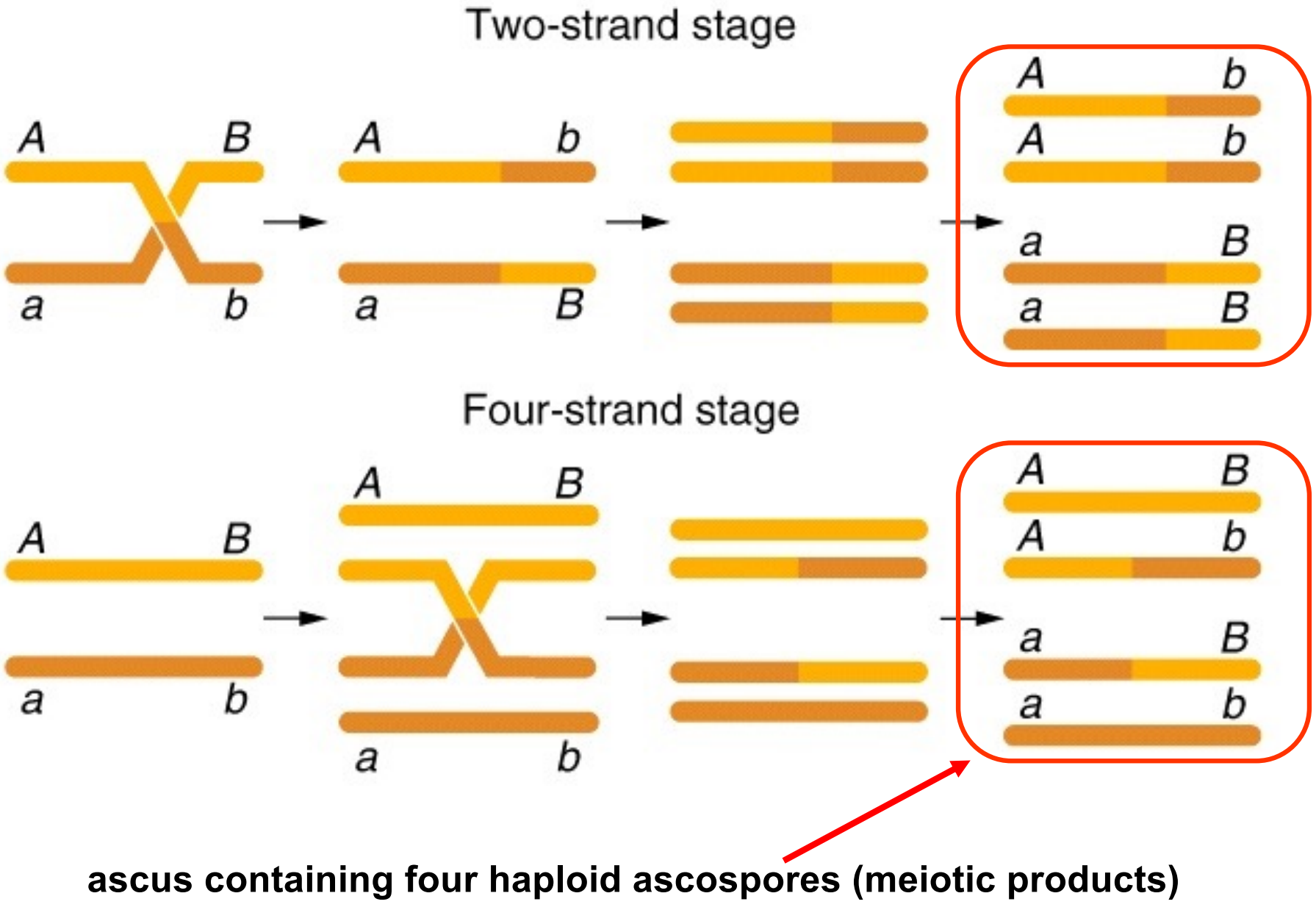
Linkage



Fungal life cycles



Recombination happens at the 4-strand stage



ascus containing four haploid ascospores (meiotic products)

Measuring distance between genes.

Alfred Henry Sturtevant

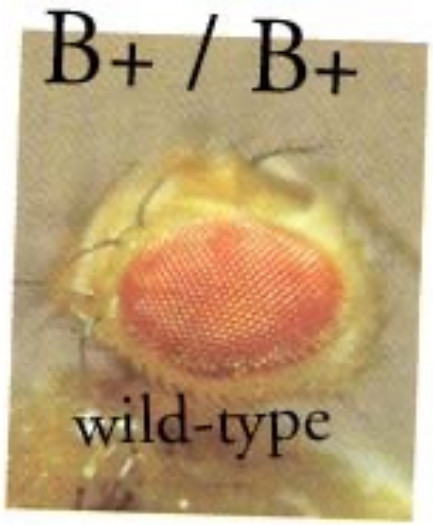
Photo courtesy of Cold Spring Harbor Laboratory Archives.

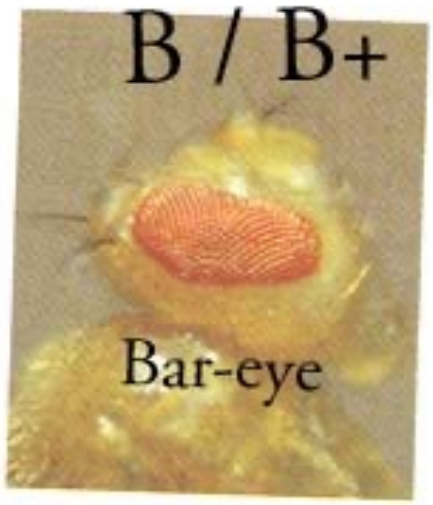
- Alfred Sturtevant, an undergraduate in Morgan's lab proposed that the % recombinants could be used to measure distance between genes.
- Using this approach he constructed first genetic map.

Wild type
white^{+}

white^{-}

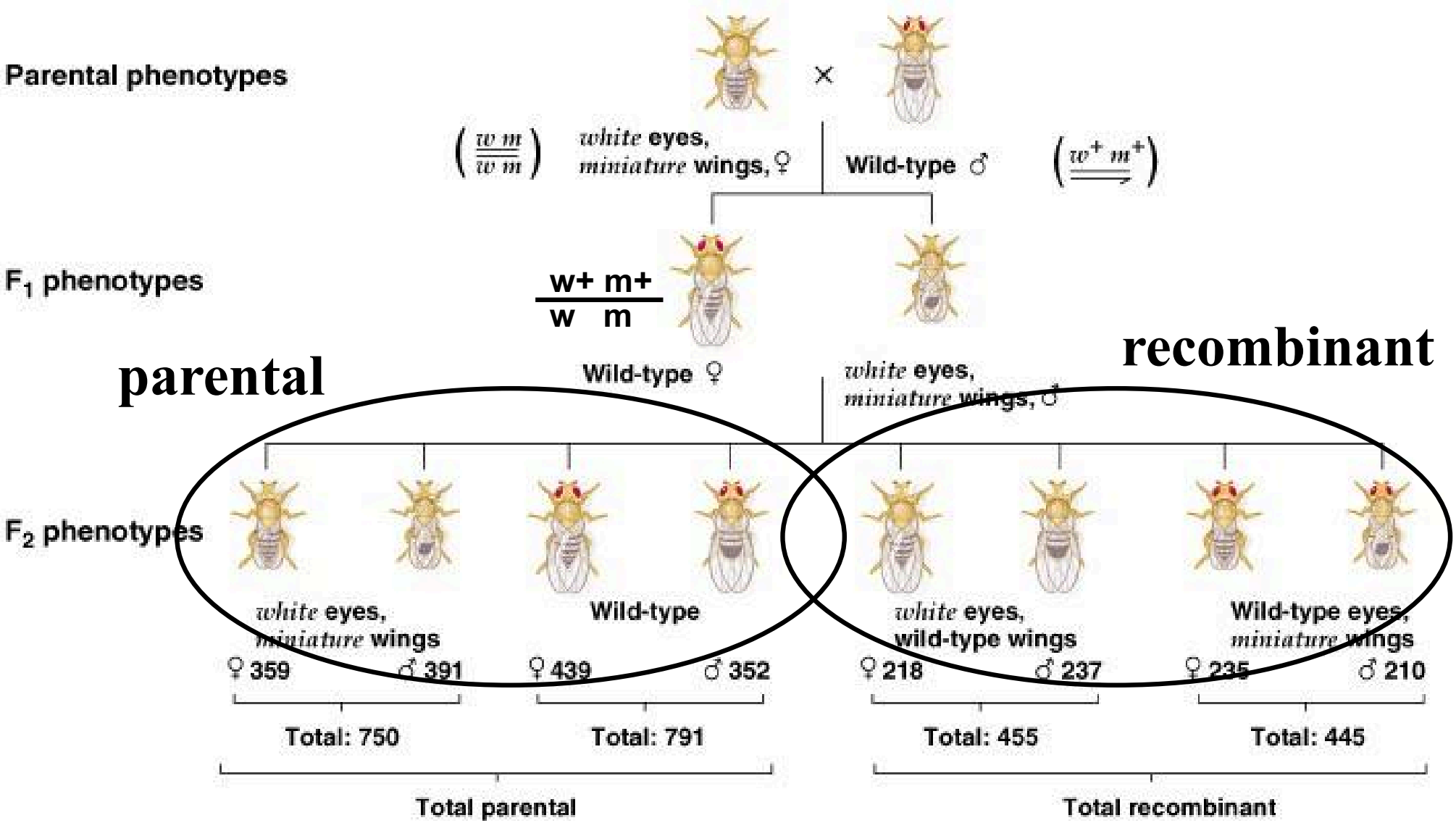
The gene white makes a red pigment.





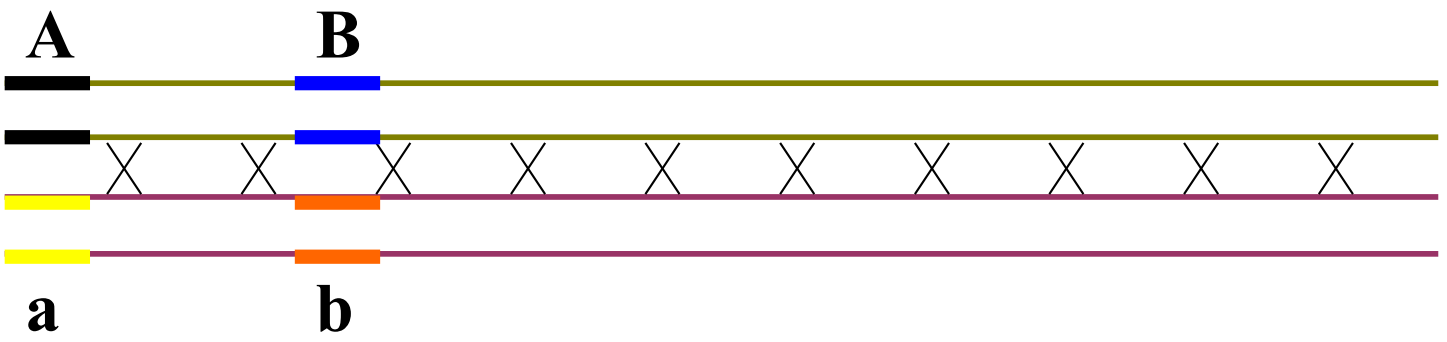
m (Miniature wings)

BUT, the ratio is not 1:1:1:1

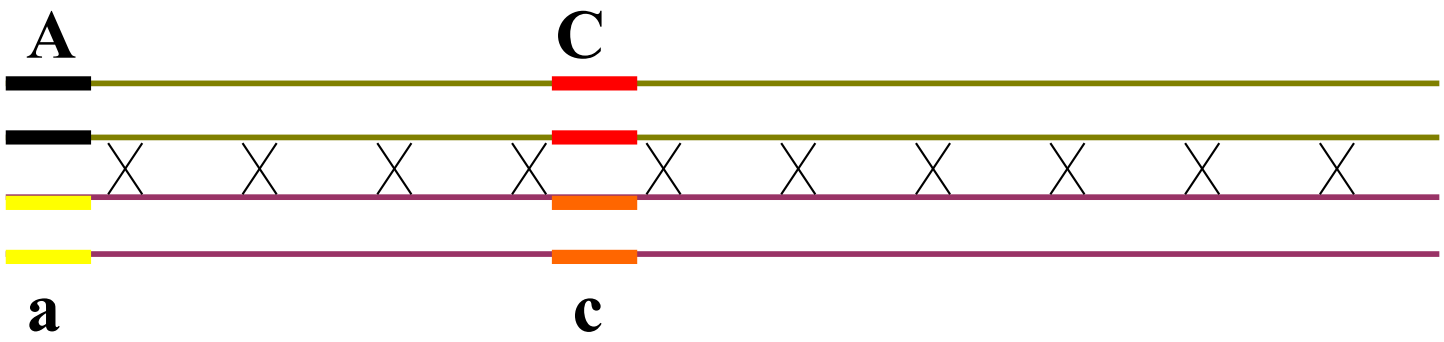


Sturtevant proposed:

- **Alleles that lie on the same chromosome tend to remain together during meiosis.**
- **Chiasmata (sites of attachment between homologs) were sites of reciprocal exchange between homologs, a process that he referred to as crossing over.**
- **Crossing over produced recombinant gametes.**
- **The closer two genes are, the less likely crossing over will occur between then and fewer recombinants will be produced.**

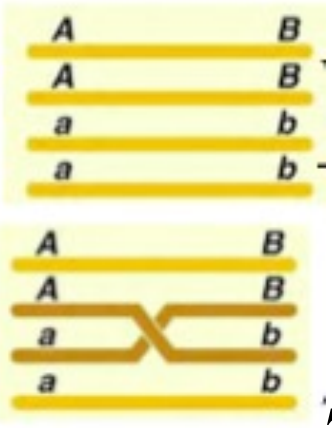


In this example, there is a 2/10 chance of recombination.



In this example, there is a 4/10 chance of recombination.

white (w) and *miniature (m)* mapping



wm

++

***w m* /+ + X *w m* / Y**

↓

770 red eyes, normal wings (++) } **P**

716 white eyes, miniature wings (*wm*) }

401 red eyes, miniature wings (+ *m*) } **R**

318 white eyes, normal wings (*w* +) }

Rf = 401 + 318/ 2205 = 0.326; 32.6% recombination; 32.6 map units (mu); 32.6 centiMorgans (cM)

white (w) and *yellow (y)* mapping

***w y* /+ + X *w y* / Y**

1159	**red eyes, brown body (++)**	**} P**
1017	**white eyes, yellow body (*wy*)**	
12	**red eyes, yellow body (+ *y*)**	**} R**
17	**white eyes, brown body (*w* +)**	

Rf = 12 + 17 / 2205 = 0.013 or 1.3 mu

yellow (y) and *miniature (m)* mapping

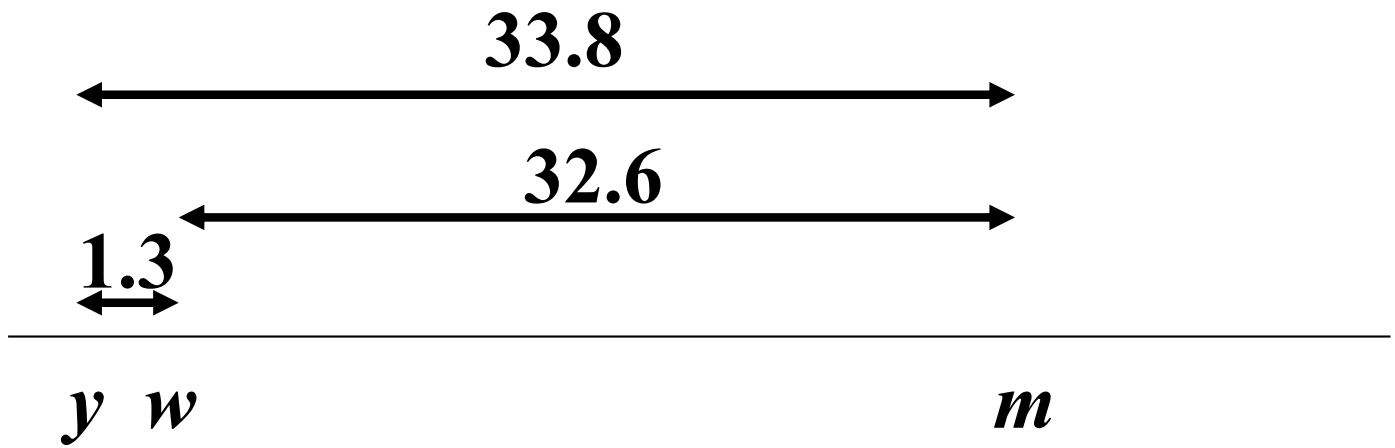
y m **/+ + X *y m* / Y**

759	**brown body, normal wings (++)**	**P**
700	**yellow body, miniature wings (*ym)***	
417	**brown body, miniature wings (+ *m)***	**R**
329	**yellow body, normal wings (*y* +)**	

Rf = 417 + 329 / 2205 = 0.338 or 33.8 mu

Genetic map

***y* - *w* 1.3 mu**
***w* - *m* 32.6 mu**
***y* - *m* 33.8 mu**



Results are consistent with the gene order: *y w m*

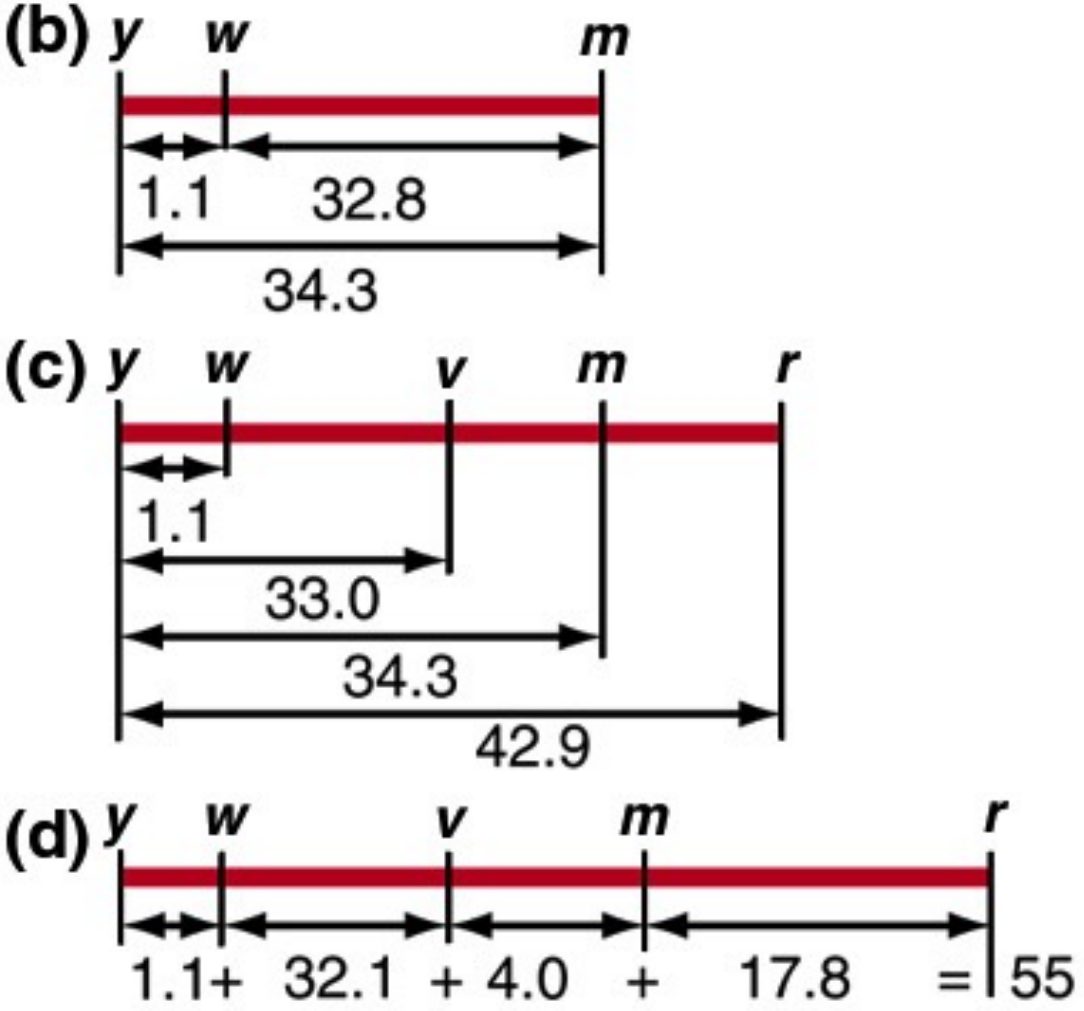
$$\text{Genetic map distance (recombination frequency)} = \frac{\text{\# of recombinant gametes}}{\text{Total gametes}} \times 100$$

Genetic map distance measured in centiMorgans (cM)

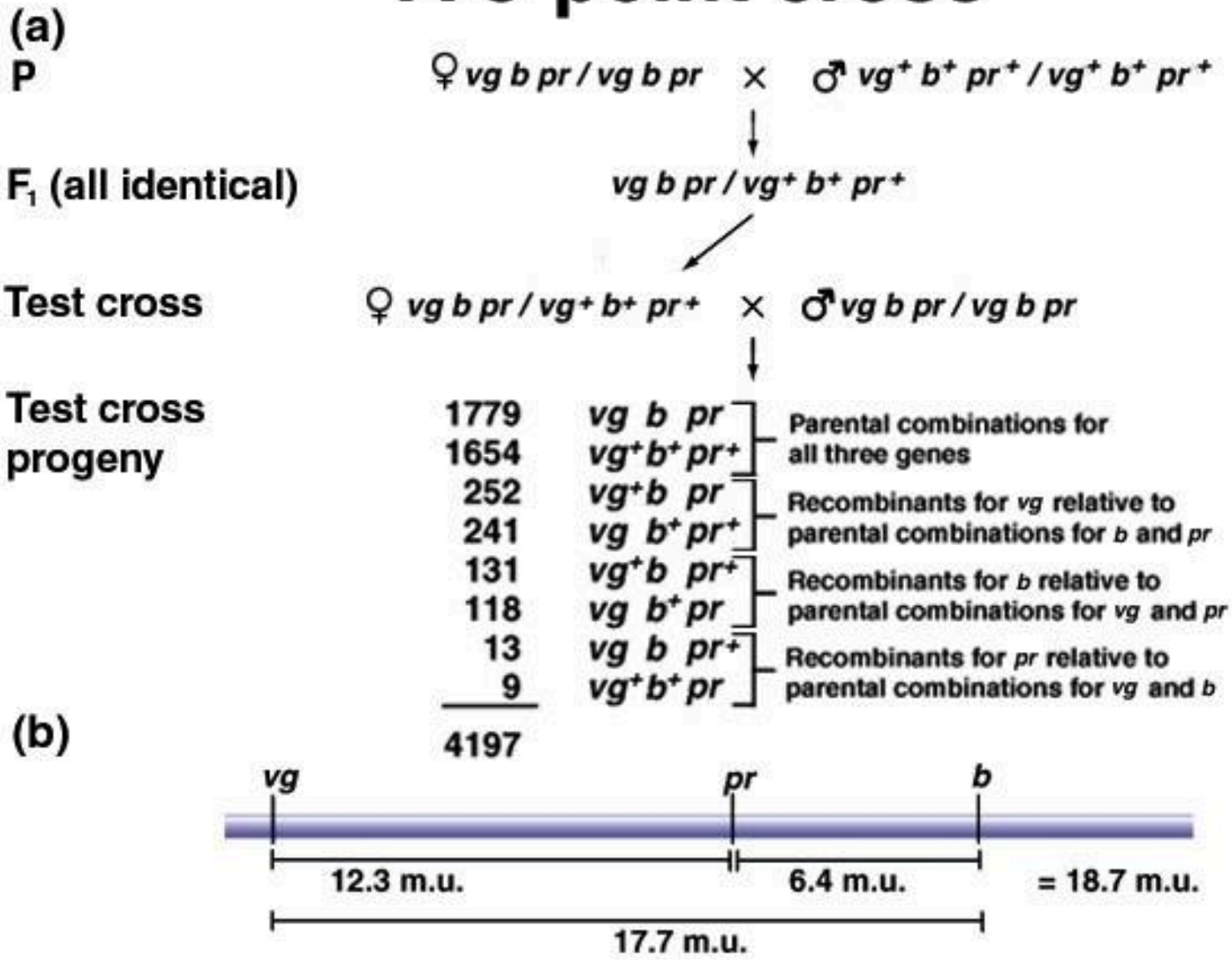
The first genetic map

(a)

y-w	1.1
y-v	33.0
y-m	34.3
y-r	42.9
w-v	32.1
w-m	32.8
w-r	42.1
v-m	4.0
v-r	24.1
m-r	17.8

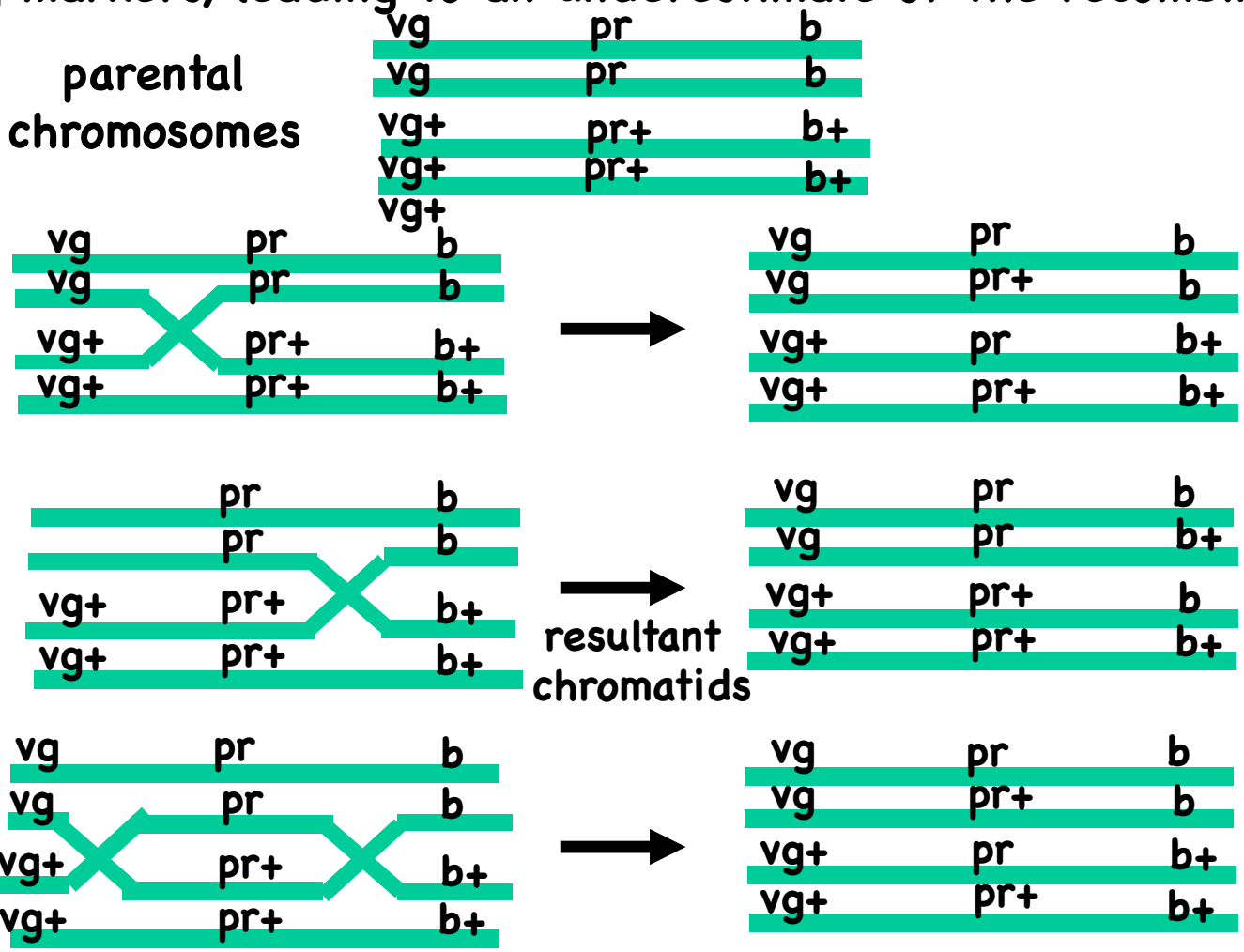


A 3-point cross



$$\frac{252 + 241 + 131 + 118}{4197} \times 100 = vg\text{-}b = 17.7 \text{ m. u.}$$

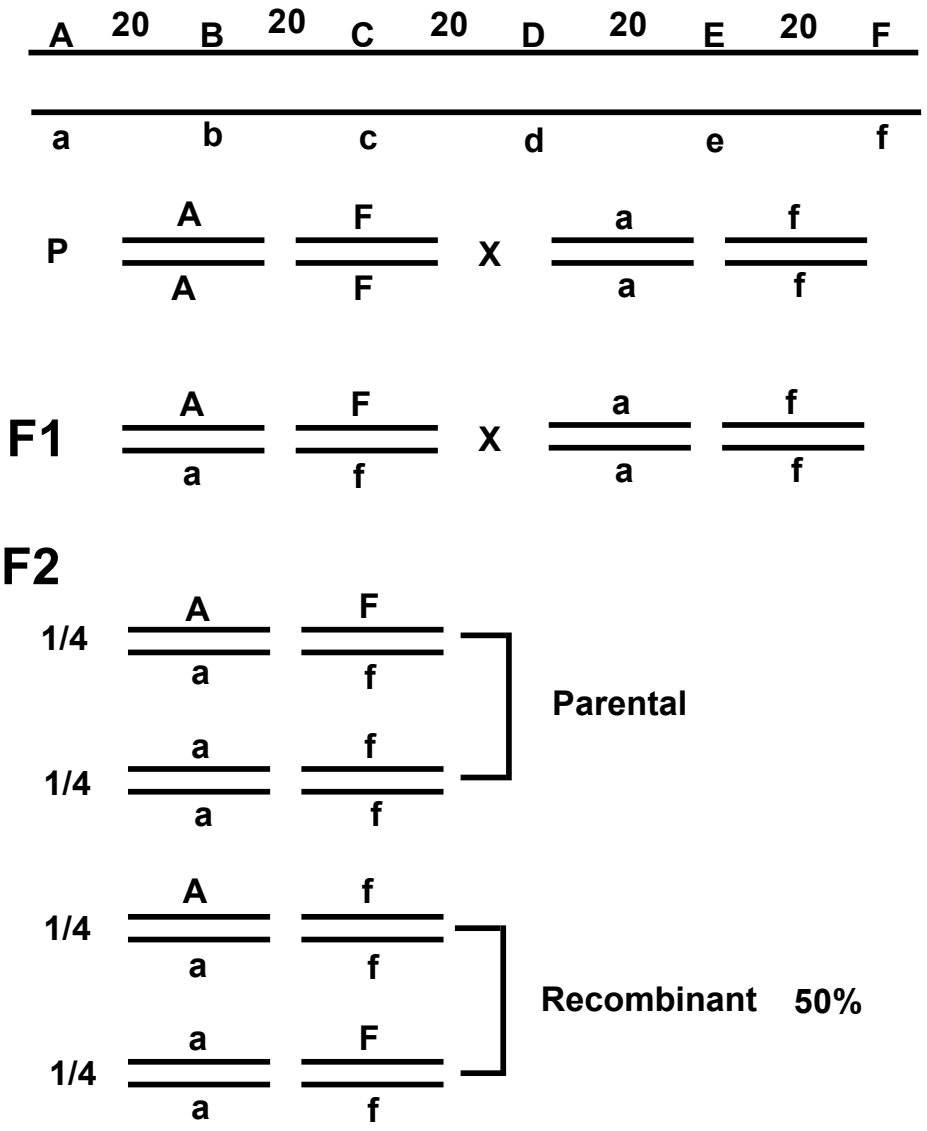
Double crossovers between two markers can result in parental genotypes for flanking markers, leading to an underestimate of the recombination distance.



These double crossovers can be identified if there are markers (alleles of genes) in the middle (in this case pr, pr+) whose recombination can be followed.

$$\frac{252 + 241 + 131 + 118 + \boxed{13 + 13 + 9 + 9}}{4197}$$

The largest map distance that can be measured between 2 genes is 50 map units



Genes far away on a chromosome act as if they are unlinked

No crossovers

RF = $\frac{0}{4}$ = 0%

One crossover

(Can be between any nonsister pair.)

RF = $\frac{2}{4}$ = 50%

Two crossovers

(Holding one crossover constant and varying the position of the second produces four equally frequent two-crossover meioses.)

RF = $\frac{0}{4}$ = 0% — Two-strand double crossover

RF = $\frac{2}{4}$ = 50% — Three-strand double crossover

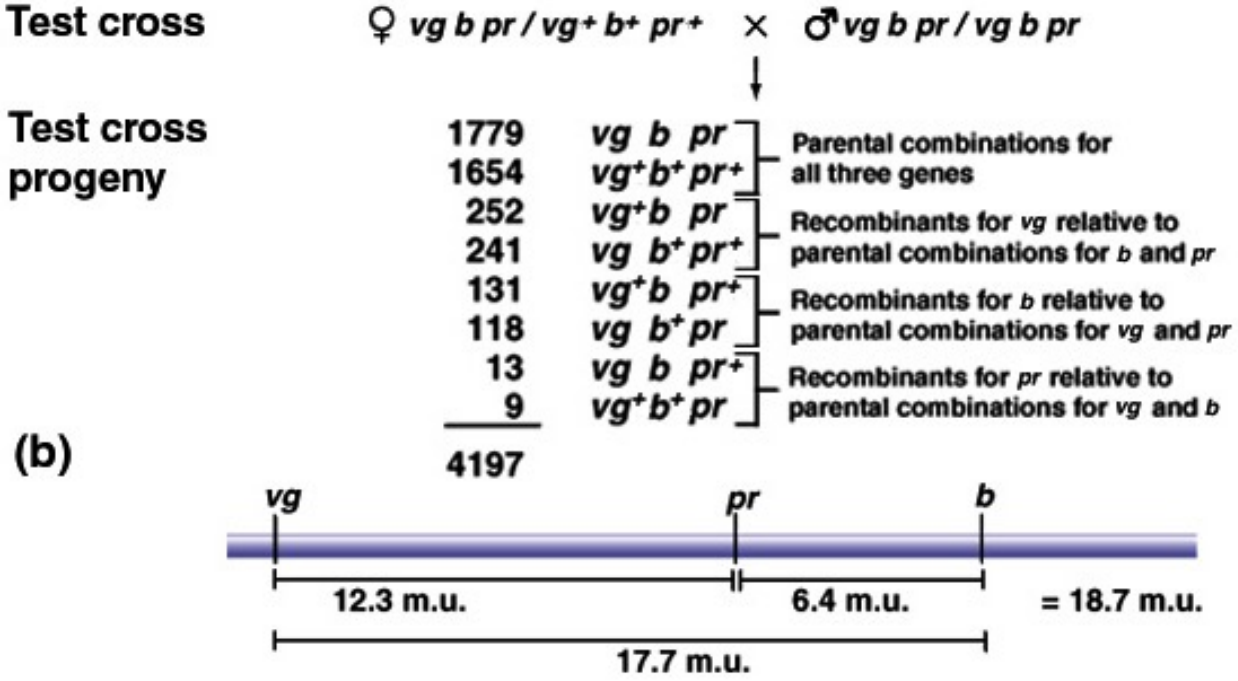
RF = $\frac{2}{4}$ = 50% — Three-strand double crossover

RF = $\frac{4}{4}$ = 100% — Four-strand double crossover

Average two-crossover RF = $\frac{8}{16}$ = 50%

One or more crossovers looks, on average, like 50% recombination.

Interference



If the probability of one recombinant in a given interval (vg--b) is p then the probability of two such events is p X p. But this is not always the case.

b--pr = 0.064
vg--pr = 0.123

0.123 X .064 = .0079 = .79% Expected

$$\frac{13 + 9}{4197} = 0.52\% \text{ Observed}$$

The occurrence of one crossover reduces the likelihood of a second crossover = interference

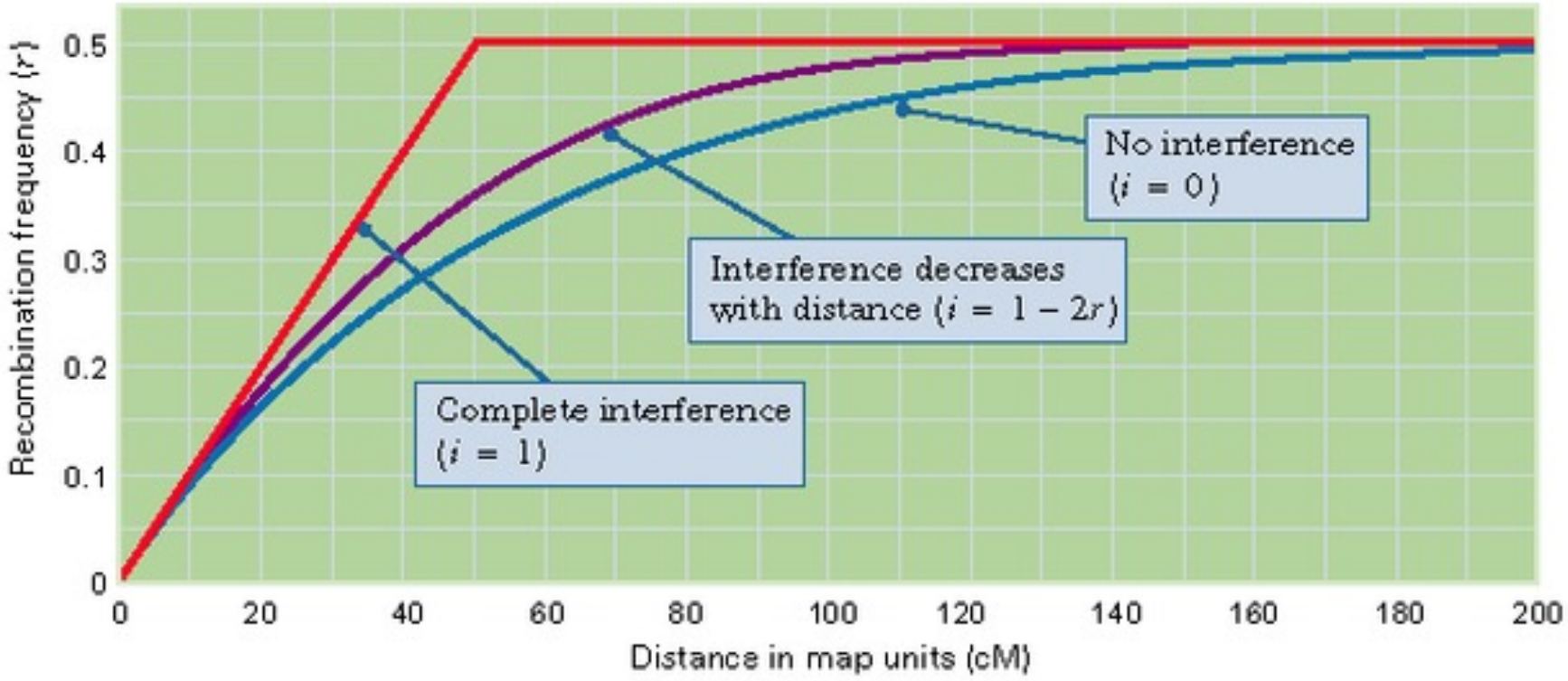
$$\text{coefficient of coincidence} = \frac{\text{Frequency of observed doubles}}{\text{Frequency of expected doubles}} = \frac{0.52}{0.79} = 0.66$$

Interference = 1 - coefficient of coinc.

I = 0 = crossovers do not interfere with each other

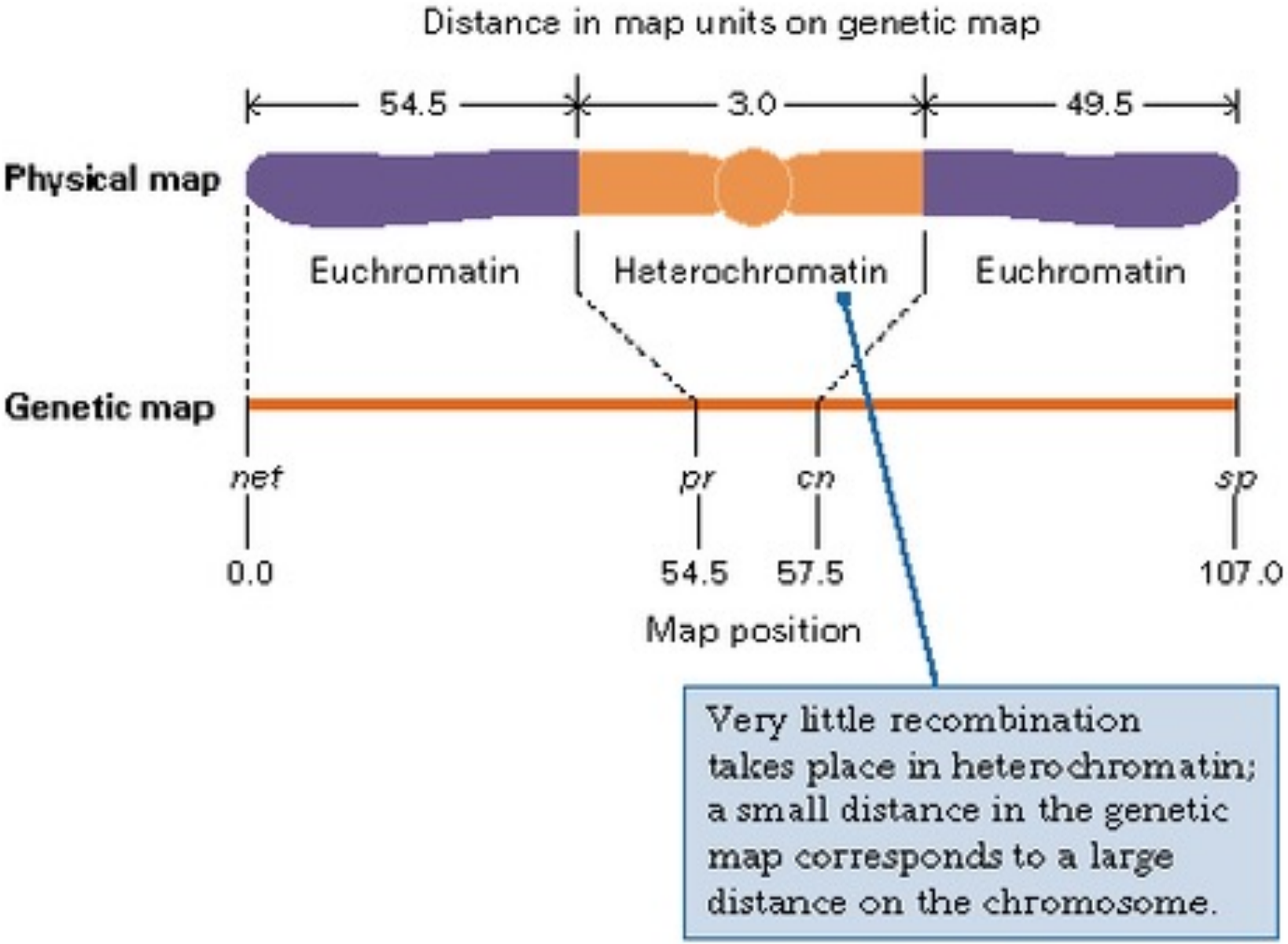
I = 1 = There are no double crossovers

When interference is low we need some method for correcting for the presence of multiple crossovers in order to generate an accurate genetic map



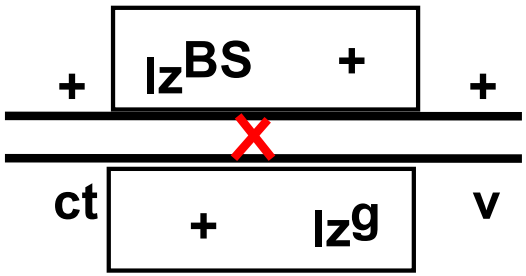
Bottom line: best to measure distances over a series of small distances rather than over one large distance

Map units/physical distance depends on the nature of the chromatin involved

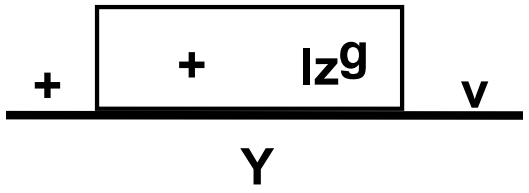


The unit of recombination is the basepair

Therefore, recombination can also occur within a gene



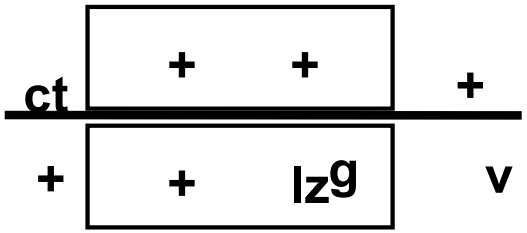
X

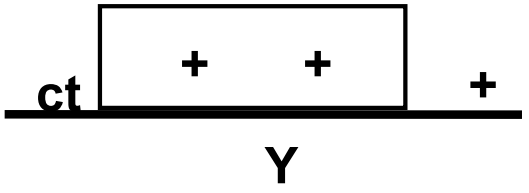


→ 16 thousand progeny $\frac{lz}{lz}$

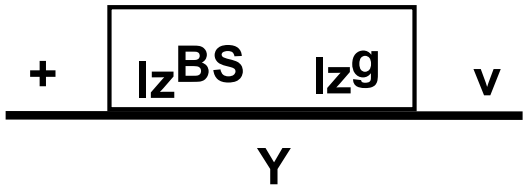
134 males and females lz+ = 8 X 10^{-3}

Reversion rate = 10^{-6}





Only a few of the reciprocal recombinants (5)
They are semilethal.



Chi-Square Test

- **A fundamental problem is genetics is determining whether the experimentally determined data fits the results expected from theory (i.e. Mendel's laws as expressed in the Punnett square).**
- **How can you tell if an observed set of offspring counts is legitimately the result of a given underlying simple ratio? For example, you do a cross and see 290 purple flowers and 110 white flowers in the offspring. This is pretty close to a 3/4 : 1/4 ratio, but how do you formally define "pretty close"? What about 250:150?**

Goodness of Fit

- **Mendel had no way of solving this problem. Shortly after the rediscovery of his work in 1900, Karl Pearson and R.A. Fisher developed the "chi-square" test for this purpose.**
- **The chi-square test is a "goodness of fit" test: it answers the question of how well do experimental data fit expectations.**
- **We start with a theory for how the offspring will be distributed: the "null hypothesis". We will discuss the offspring of a self-pollination of a heterozygote. The null hypothesis is that the offspring will appear in a ratio of 3/4 dominant to 1/4 recessive.**

Formula

- **First determine the number of each phenotype that have been observed and how many would be expected given basic genetic theory.**
- **Then calculate the chi-square statistic using this formula.**
- **The "X" is the Greek letter chi; the "∑" is a sigma; it means to sum the following terms for all phenotypes. "obs" is the number of individuals of the given phenotype observed; "exp" is the number of that phenotype expected from the null hypothesis.**
- **Note that you must use the <u>number</u> of individuals, the counts, and NOT proportions, ratios, or frequencies.**

$$X^2 = \sum \frac{(obs - \exp)^2}{\exp}$$

Example

- **As an example, you count F2 offspring, and get 290 purple and 110 white flowers. This is a total of 400 (290 + 110) offspring.**

- **We expect a 3/4 : 1/4 ratio. We need to calculate the expected numbers (you MUST use the numbers of offspring, NOT the proportion!!!); this is done by multiplying the total offspring by the expected proportions. This we expect 400 * 3/4 = 300 purple, and 400 * 1/4 = 100 white.**

- **Thus, for purple, obs = 290 and exp = 300. For white, obs = 110 and exp = 100.**

- **Now it's just a matter of plugging into the formula:**

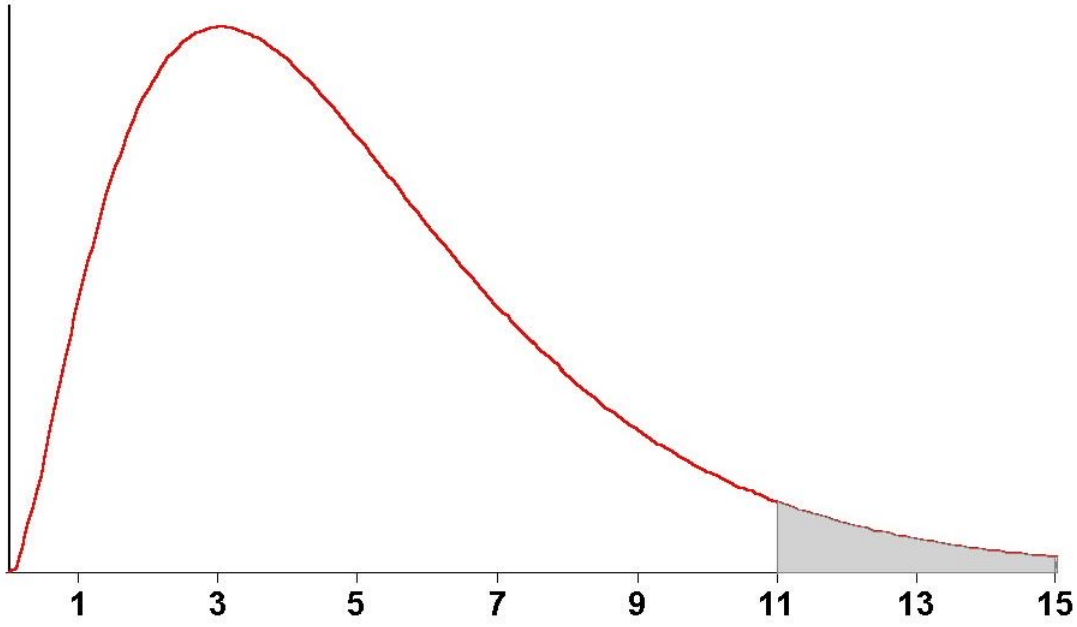
$$X^2 = (290 - 300)^2 / 300 + (110 - 100)^2 / 100$$
$$= (-10)^2 / 300 + (10)^2 / 100$$
$$= 100 / 300 + 100 / 100$$
$$= 0.333 + 1.000$$
$$= 1.333.$$

- **This is our chi-square value: now we need to see what it means and how to use it.**

Chi-Square Distribution

- **Although the chi-square distribution can be derived through math theory, we can also get it experimentally:**

- **Let's say we do the same experiment 1000 times, do the same self-pollination of a Pp heterozygote, which should give the 3/4 : 1/4 ratio. For each experiment we calculate the chi-square value, them plot them all on a graph.**



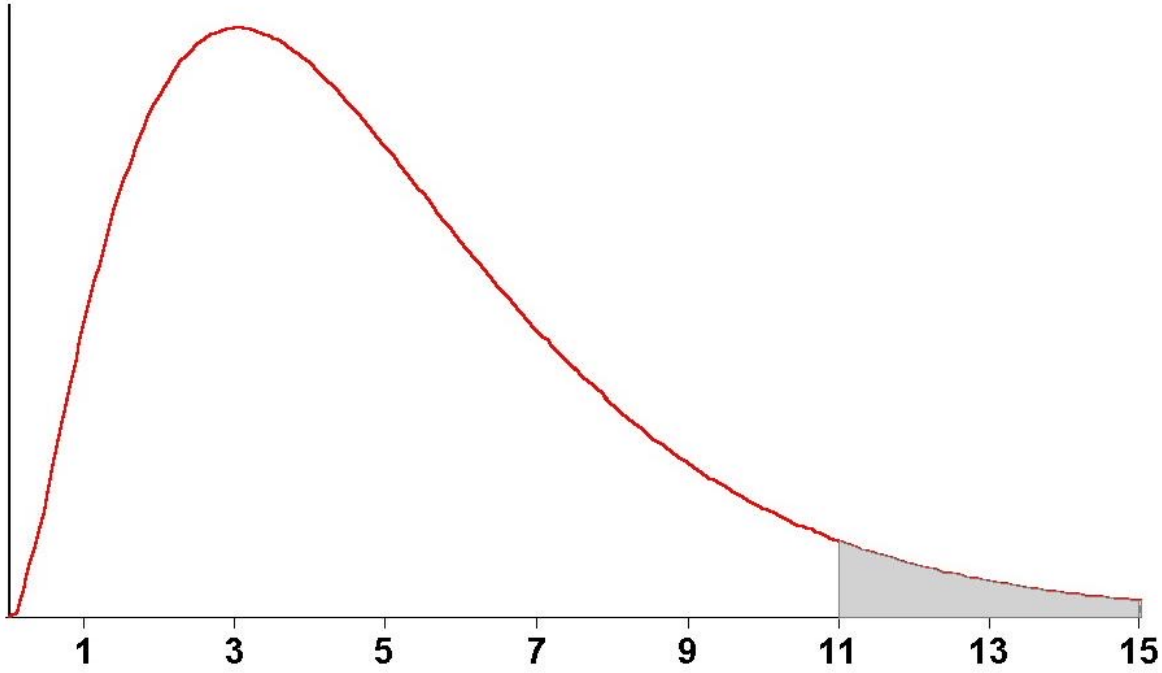
- **The x-axis is the chi-square value calculated from the formula. The y-axis is the number of individual experiments that got that chi-square value.**

Chi-Square Distribution, p. 2

- **You see that there is a range here: if the results were perfect you get a chi-square value of 0 (because obs = exp). This rarely happens: most experiments give a small chi-square value (the hump in the graph).**

- **Note that all the values are greater than 0: that's because we squared the (obs - exp) term: squaring always gives a non-negative number.**

- **Sometimes you get really wild results, with obs very different from exp: the long tail on the graph. Really odd things occasionally do happen by chance alone (for instance, you might win the lottery).**



The Critical Question

- **how do you tell a really odd but correct result from a WRONG result? The graph is what happens with real experiments: most of the time the results fit expectations pretty well, but occasionally very skewed distributions of data occur even though you performed the experiment correctly, based on the correct theory.**

- **The simple answer is: you can never tell for certain that a given result is "wrong", that the result you got was completely impossible based on the theory you used. All we can do is determine whether a given result is likely or unlikely.**

- **Key point: There are 2 ways of getting a high chi-square value: an unusual result from the correct theory, or a result from the wrong theory. These are indistinguishable; because of this fact, statistics is never able to discriminate between true and false with 100% certainty.**

- **Using the example here, how can you tell if your 290: 110 offspring ratio really fits a 3/4 : 1/4 ratio (as expected from selfing a heterozygote) or whether it was the result of a mistake or accident-- a 1/2 : 1/2 ratio from a backcross for example? You can't be certain, but you can at least determine whether your result is reasonable.**

Reasonable

- **What is a "reasonable" result is subjective and arbitrary.**
- **For most work (and for the purposes of this class), a result is said to not differ significantly from expectations if it could happen at least 1 time in 20. That is, if the difference between the observed results and the expected results is small enough that it would be seen at least 1 time in 20 over thousands of experiments, we "fail to reject" the null hypothesis.**
- **For technical reasons, we use "fail to reject" instead of "accept".**
- **"1 time in 20" can be written as a probability value p = 0.05, because 1/20 = 0.05.**
- **Another way of putting this. If your experimental results are worse than 95% of all similar results, they get rejected because you may have used an incorrect null hypothesis.**

Degrees of Freedom

- A critical factor in using the chi-square test is the "degrees of freedom", which is essentially the number of independent random variables involved.

- Degrees of freedom is simply the number of classes of offspring minus 1.

- For our example, there are 2 classes of offspring: purple and white. Thus, degrees of freedom (d.f.) = 2 -1 = 1.

Critical Chi-Square

- Critical values for chi-square are found on tables, sorted by degrees of freedom and probability levels. Be sure to use $p = 0.05$.

- If your calculated chi-square value is greater than the critical value from the table, you "reject the null hypothesis".

- If your chi-square value is less than the critical value, you "fail to reject" the null hypothesis (that is, you accept that your genetic theory about the expected ratio is correct).

Chi-Square Table

Table 5-2

Critical Values of the χ^2 Distribution

df \ p	0.995	0.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005	df
1	.000	.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879	1
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597	2
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838	3
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860	4
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750	5
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548	6
7	0.989	1.690	2.833	6.346	12.017	14.067	16.013	18.475	20.278	7
8	1.344	2.180	3.490	7.344	13.362	15.507	17.535	20.090	21.955	8
9	1.735	2.700	4.168	8.343	14.684	16.919	19.023	21.666	23.589	9
10	2.156	3.247	4.865	9.342	15.987	18.307	20.483	23.209	25.188	10
11	2.603	3.816	5.578	10.341	17.275	19.675	21.920	24.725	26.757	11
12	3.074	4.404	6.304	11.340	18.549	21.026	23.337	26.217	28.300	12
13	3.565	5.009	7.042	12.340	19.812	22.362	24.736	27.688	29.819	13
14	4.075	5.629	7.790	13.339	21.064	23.685	26.119	29.141	31.319	14
15	4.601	6.262	8.547	14.339	22.307	24.996	27.488	30.578	32.801	15

Linkage?

- T. H. Morgan crossed *Drosophila melanogaster* wildtype flies with pure breeding purple-eyed, vestigial winged flies
- He then crossed F_1 females heterozygous for purple-eyes (pr) and vestigial wings (vg) with tester males homozygous recessive for pr and vg
- i.e. $pr^+/pr \cdot vg^+/vg$ X $pr/pr \cdot vg/vg$

F1 of testcross

Class	Phenotype	O	Type
pr^+ vg^+	Wildtype eyes and wings	1339	Parental
pr vg	Purple eyes, vestigial wings	1195	Parental
pr^+ vg	Wildtype eyes, vestigial wings	151	Recombinant
pr vg^+	Purple eyes, wildtype wings	154	Recombinant
Total		2839	

Linked

- Parental vs recombinant
- 1339+1195 vs 151+154
- Recombinant frequency = (151+154)/2839
- 305/2839 = 10.7% recombinant frequency
- Linked in cis conformation
  - pr+ vg+ / pr vg

Null hypothesis: the genes are unlinked

- H_0: based on Mendel's first law of equal segregation the observed progeny fits a 1:1:1:1 ratio of phenotypes

  pr+· vg+: pr ·vg+ : pr+ · vg : pr · vg

Class	O	E	$(O-E)^2$	$(O-E)^2/E$
$pr^+ vg^+$	1339	709.75	$(629.25)^2$	557.88
$pr\ vg^+$	1195	709.75	$(485.25)^2$	331.76
$pr^+ vg$	151	709.75	$(-558.75)^2$	439.87
$pr\ vg^+$	154	709.75	$(-555.75)^2$	435.16
Total	2839		Chisquare=	1764.82
For 4 phenotypic classes df= 4-1= 3				

Linked?

- So df =3, chi-square =1764.82
- Rejection level p=0.05 the X^2 =7.815. In fact on the chi-square table for p=0.005 the X^2 = 12.838
- Conclusion: given the chi-square =1764.82 and df=3 a deviation from the expected 1:1:1:1 ratio at least this large would occur by chance alone much less than 5% of the time so we reject the null hypothesis that the observed ratio of pr+· vg+: pr ·vg+ : pr+ · vg : pr · vg is 1:1:1:1.

Another Example: from Mendel

phenotype	observed	expected proportion	expected number
round yellow	315	9/16	312.75
round green	101	3/16	104.25
wrinkled yellow	108	3/16	104.25
wrinkled green	32	1/16	34.75
total	556	1	556

Finding the Expected Numbers

- You are given the observed numbers, and you determine the expected proportions from a Punnett square.
- To get the expected numbers of offspring, first add up the observed offspring to get the total number of offspring. In this case, 315 + 101 + 108 + 32 = 556.
- Then multiply total offspring by the expected proportion:
  --expected round yellow = 9/16 * 556 = 312.75
  --expected round green = 3/16 * 556 = 104.25
  --expected wrinkled yellow = 3/16 * 556 = 104.25
  --expected wrinkled green = 1/16 * 556 = 34.75
- Note that these add up to 556, the observed total offspring.

Calculating the Chi-Square Value

- Use the formula.
- $X^2 = (315 - 312.75)2 / 312.75$
  $+ \ (101 - 104.25)2 / 104.25$
  $+ \ (108 - 104.25)2 / 104.25$
  $+ (32 - 34.75)2 / 34.75$

$= \ 0.016 + 0.101 + 0.135 + 0.218$
$= 0.470.$

$$X^2 = \sum \frac{(obs - \exp)^2}{\exp}$$

D.F. and Critical Value

- Degrees of freedom is 1 less than the number of classes of offspring. Here, 4 - 1 = 3 d.f.
- For 3 d.f. and p = 0.05, the critical chi-square value is 7.815.
- Since the observed chi-square (0.470) is less than the critical value, we fail to reject the null hypothesis. We accept Mendel's conclusion that the observed results for a 9/16 : 3/16 : 3/16 : 1/16 ratio.
- It should be mentioned that all of Mendel's numbers are unreasonably accurate.

Table 5-2
Critical Values of the χ^2 Distribution

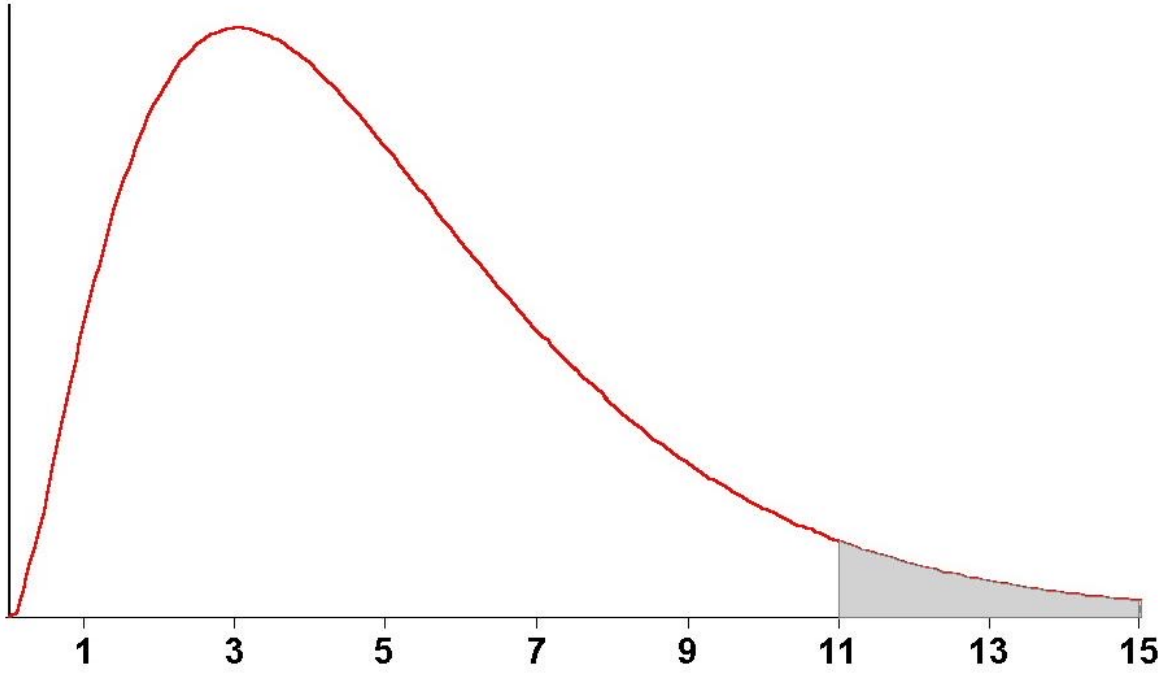
df \ p	0.995	0.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005	df
1	.000	.000	0.016	0.455	2.706	3.841	5.024	6.635	7.879	1
2	0.010	0.051	0.211	1.386	4.605	5.991	7.378	9.210	10.597	2
3	0.072	0.216	0.584	2.366	6.251	7.815	9.348	11.345	12.838	3
4	0.207	0.484	1.064	3.357	7.779	9.488	11.143	13.277	14.860	4
5	0.412	0.831	1.610	4.351	9.236	11.070	12.832	15.086	16.750	5
6	0.676	1.237	2.204	5.348	10.645	12.592	14.449	16.812	18.548	6
7	0.989	1.690	2.833	6.346	12.017	14.067	16.013	18.475	20.278	7
8	1.344	2.180	3.490	7.344	13.362	15.507	17.535	20.090	21.955	8
9	1.735	2.700	4.168	8.343	14.684	16.919	19.023	21.666	23.589	9
10	2.156	3.247	4.865	9.342	15.987	18.307	20.483	23.209	25.188	10
11	2.603	3.816	5.578	10.341	17.275	19.675	21.920	24.725	26.757	11
12	3.074	4.404	6.304	11.340	18.549	21.026	23.337	26.217	28.300	12
13	3.565	5.009	7.042	12.340	19.812	22.362	24.736	27.688	29.819	13
14	4.075	5.629	7.790	13.339	21.064	23.685	26.119	29.141	31.319	14
15	4.601	6.262	8.547	14.339	22.307	24.996	27.488	30.578	32.801	15

Chi-Square Distribution, p. 2

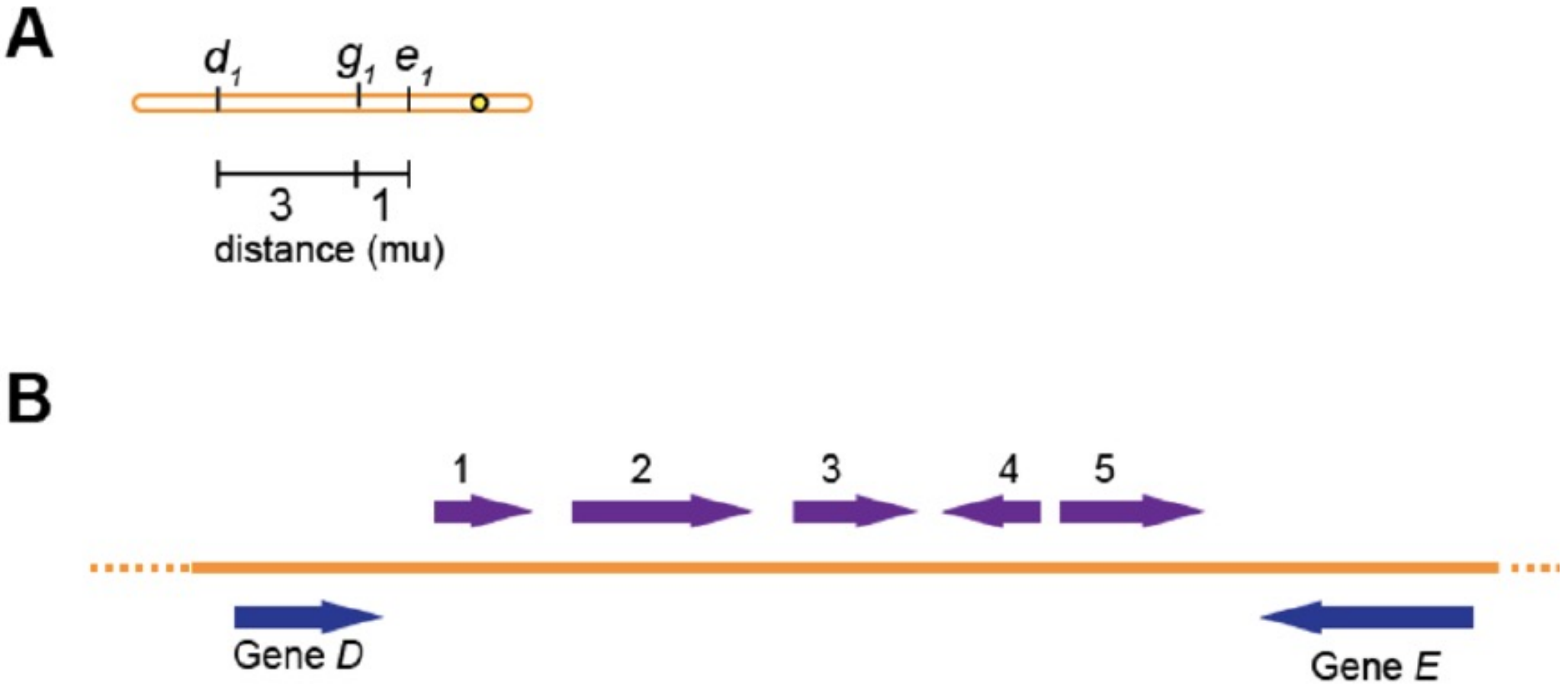
- **You see that there is a range here: if the results were perfect you get a chi-square value of 0 (because obs = exp). This rarely happens: most experiments give a small chi-square value (the hump in the graph).**

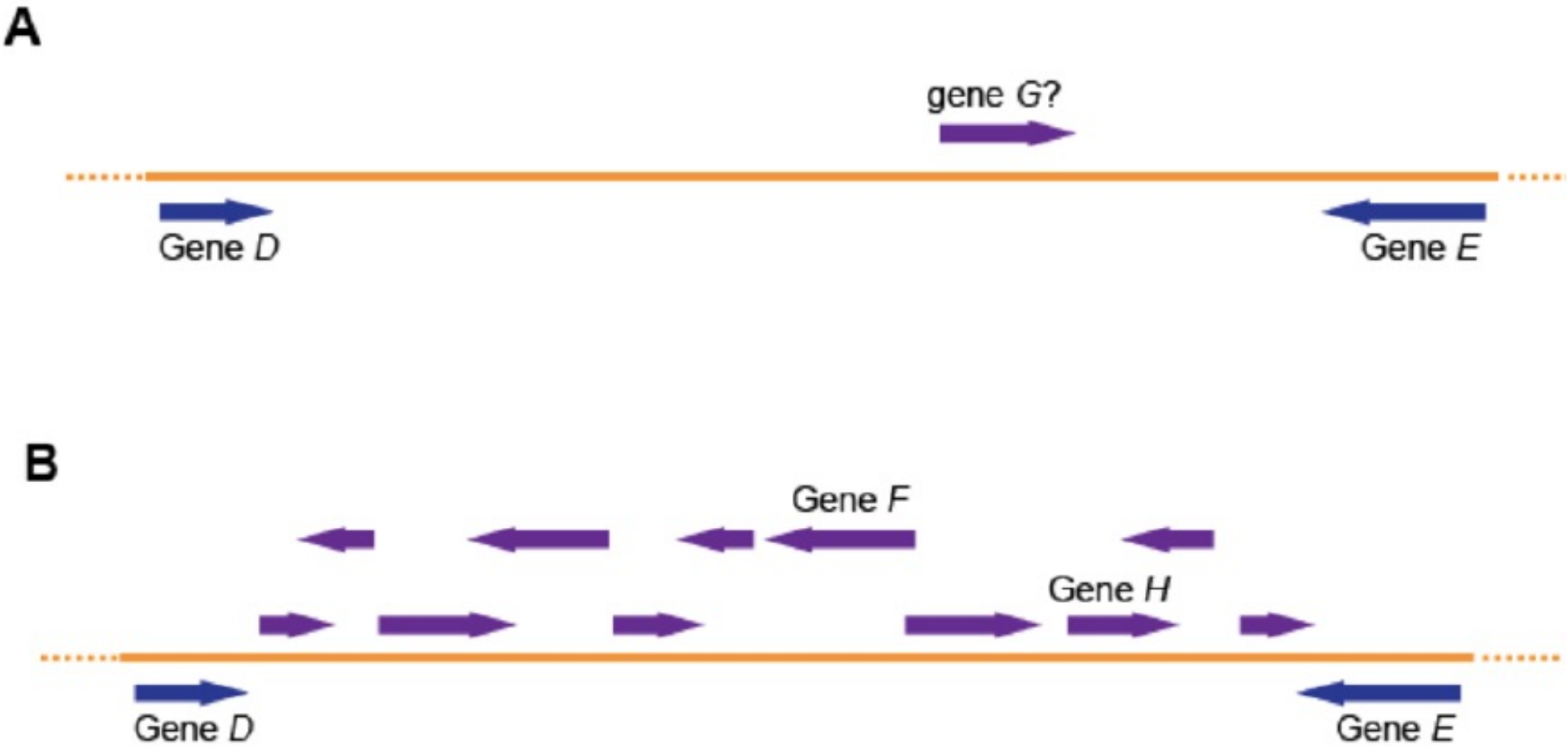
- **Note that all the values are greater than 0: that's because we squared the (obs - exp) term: squaring always gives a non-negative number.**

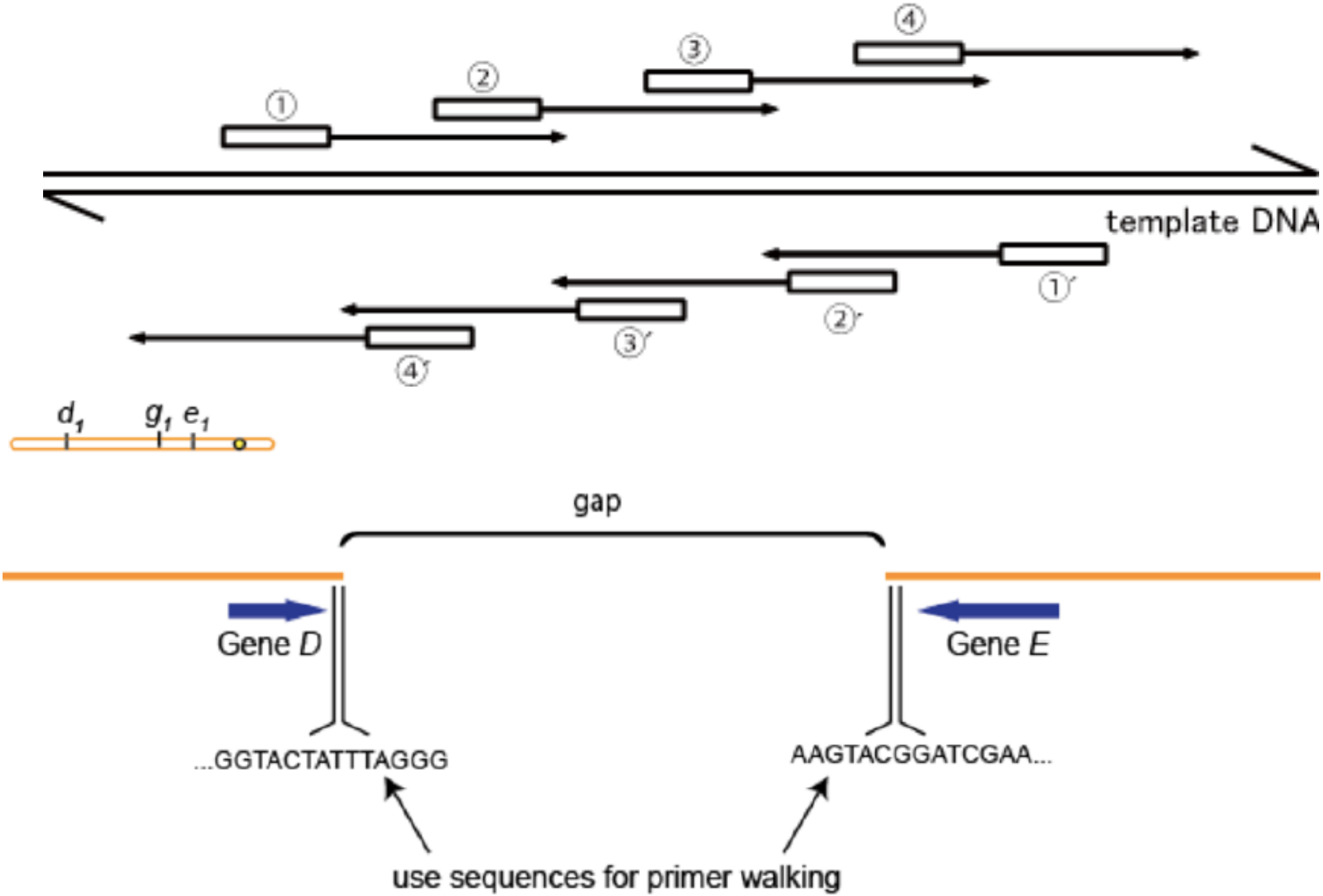
- **Sometimes you get really wild results, with obs very different from exp: the long tail on the graph. Really odd things occasionally do happen by chance alone (for instance, you might win the lottery).**

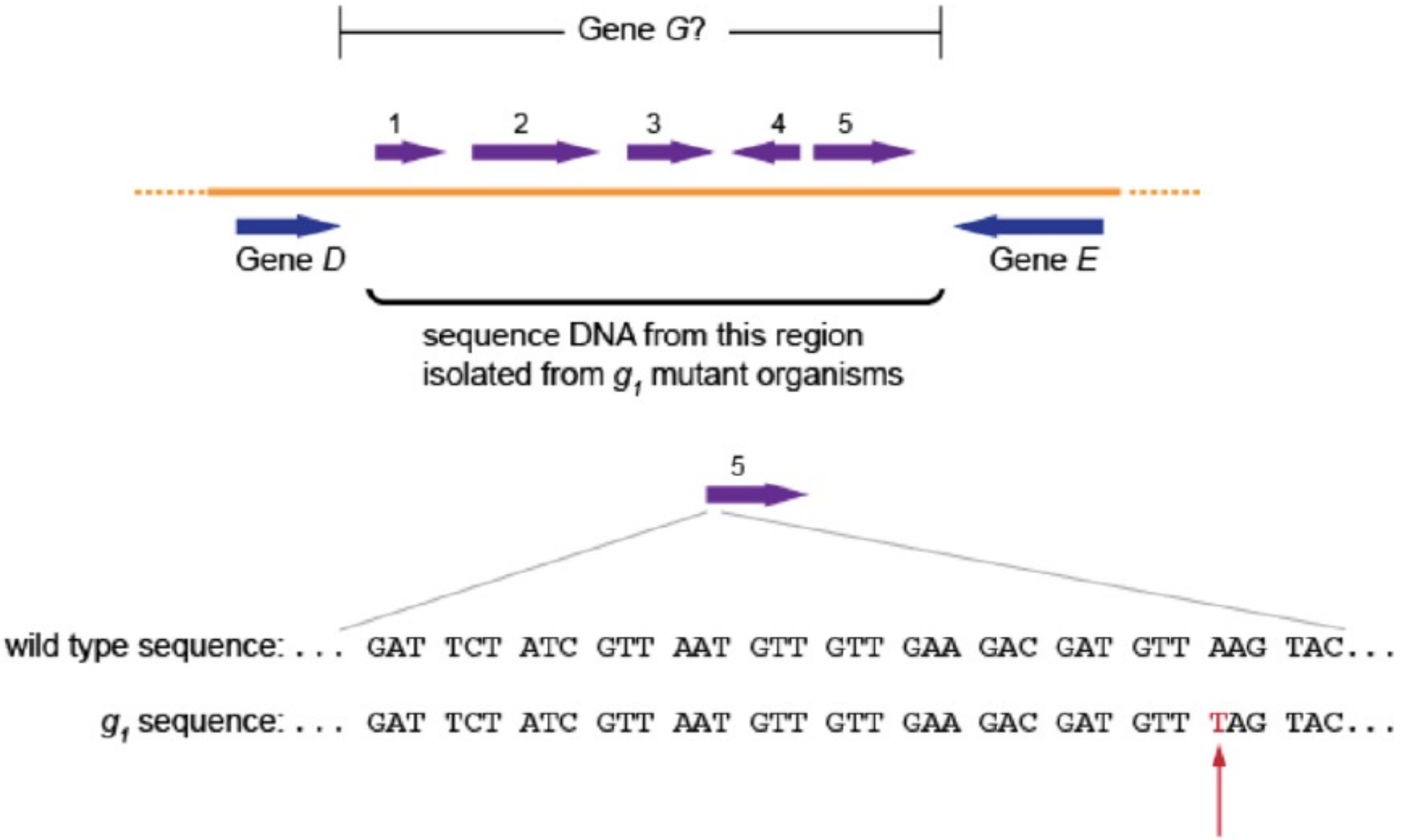


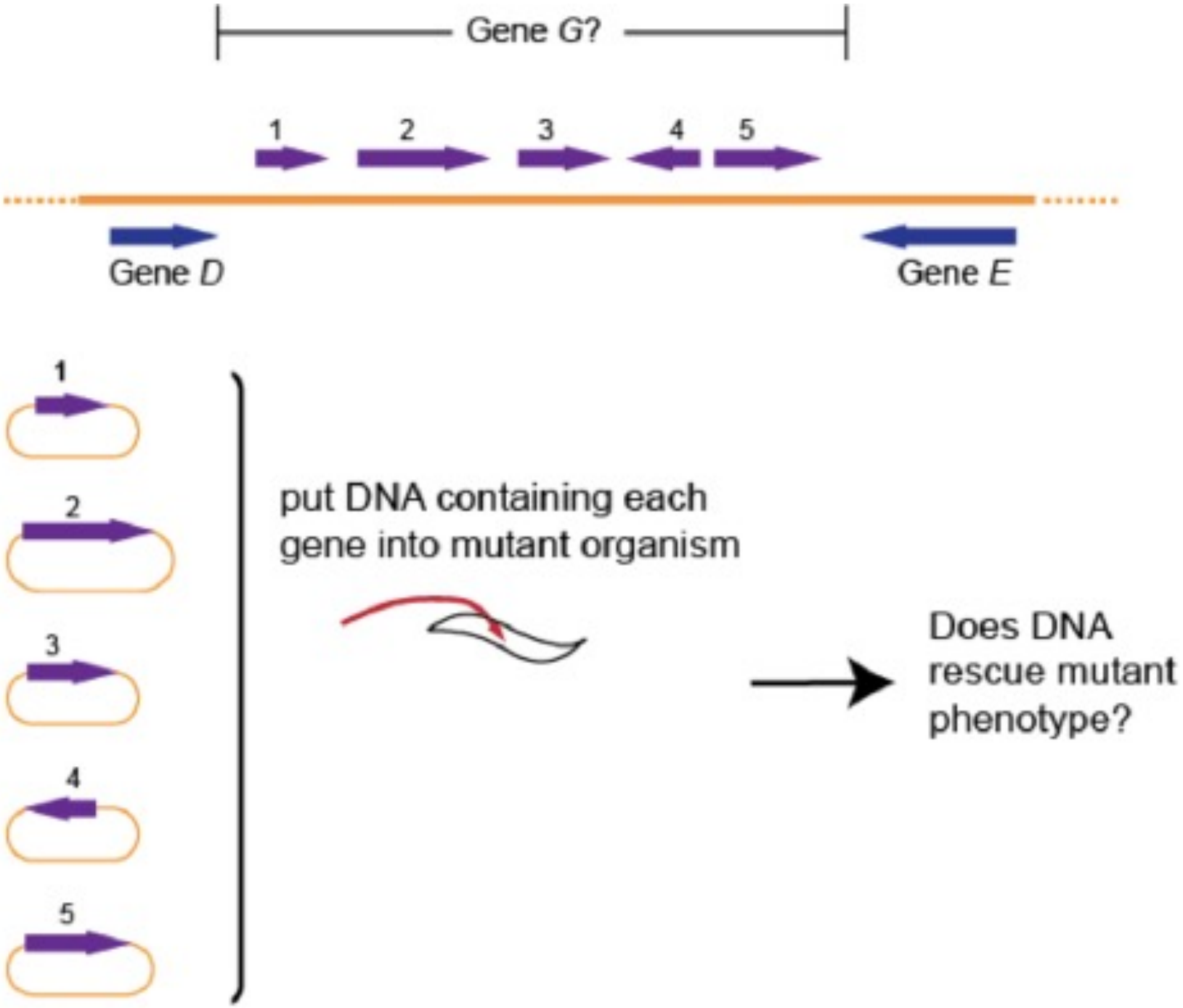
Which is the right gene??



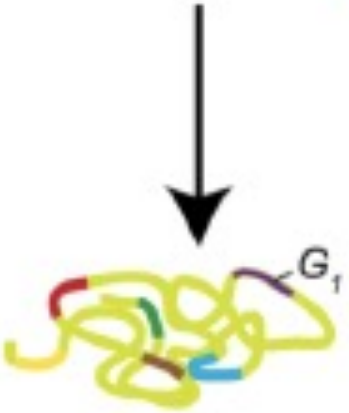

A
gene G?
Gene D
Gene E
B
Gene F
Gene H
Gene D
Gene E


①
②
③
④
template DNA
①′
②′
③′
④′
d_1
g_1
e_1
gap
Gene *D*
Gene *E*
...GGTACTATTTAGGG
AAGTACGGATCGAA...
use sequences for primer walking

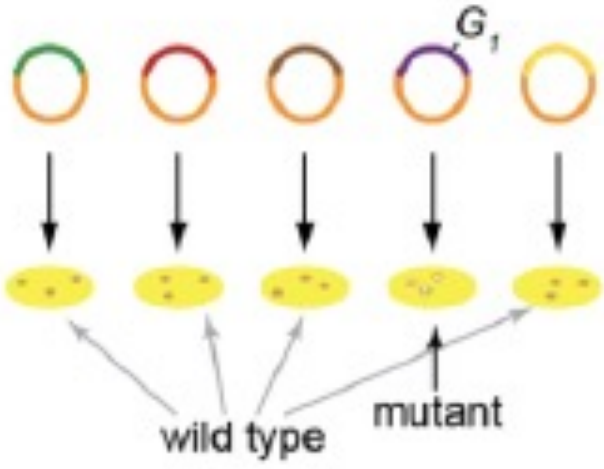

Gene G?
1
2
3
4
5
Gene D
Gene E
sequence DNA from this region
isolated from g_1 mutant organisms
5
wild type sequence: . . . GAT TCT ATC GTT AAT GTT GTT GAA GAC GAT GTT AAG TAC. . .
g_1 sequence: . . . GAT TCT ATC GTT AAT GTT GTT GAA GAC GAT GTT TAG TAC. . .


Gene G?
1
2
3
4
5
Gene D
Gene E
1
2
3
4
5
put DNA containing each gene into mutant organism
Does DNA rescue mutant phenotype?

dominant clones



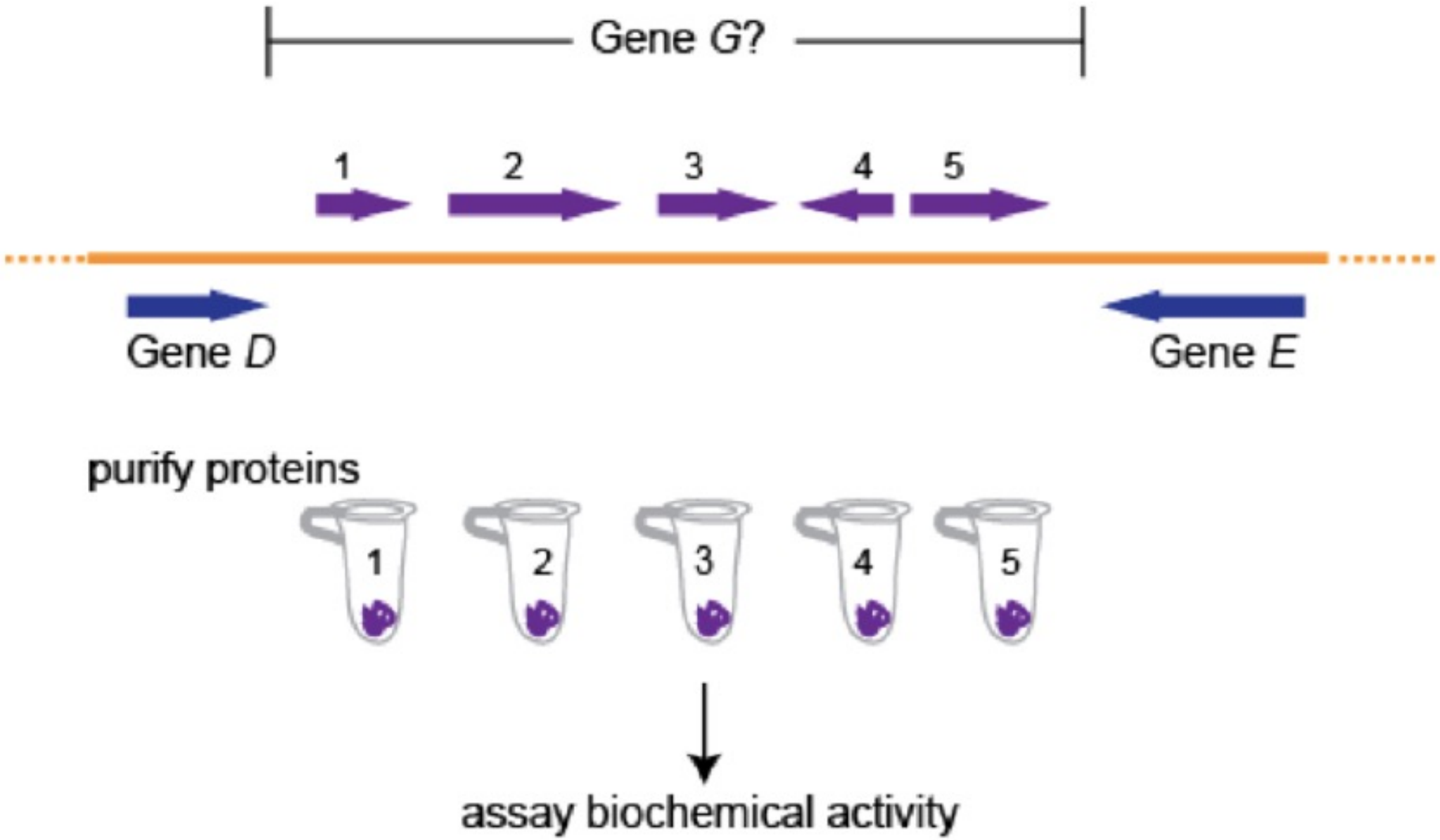
Cut genomic DNA



Create library of genomic DNA
Insertions in plasmid vector

Transform into wild type yeast

Which plasmid results in a mutant phenotype


Gene G?
1
2
3
4
5
Gene D
Gene E
purify proteins
1
2
3
4
5
assay biochemical activity

"Today, if you're a criminal and you've left your DNA at the crime scene, you may as well turn yourself in now. We will catch you," Mark Lindquist said at a June 22 news conference. The Pierce County, Wash., prosecutor was announcing the arrest of Gary Hartman for the 1986 killing of 12-year-old Michella Welch.

Crime solvers embraced genetic genealogy

The Golden State killer case was just the beginning

CAUGHT BY DNA Sacramento County Sheriff Scott Jones (left) discusses the April arrest of Joseph James DeAngelo

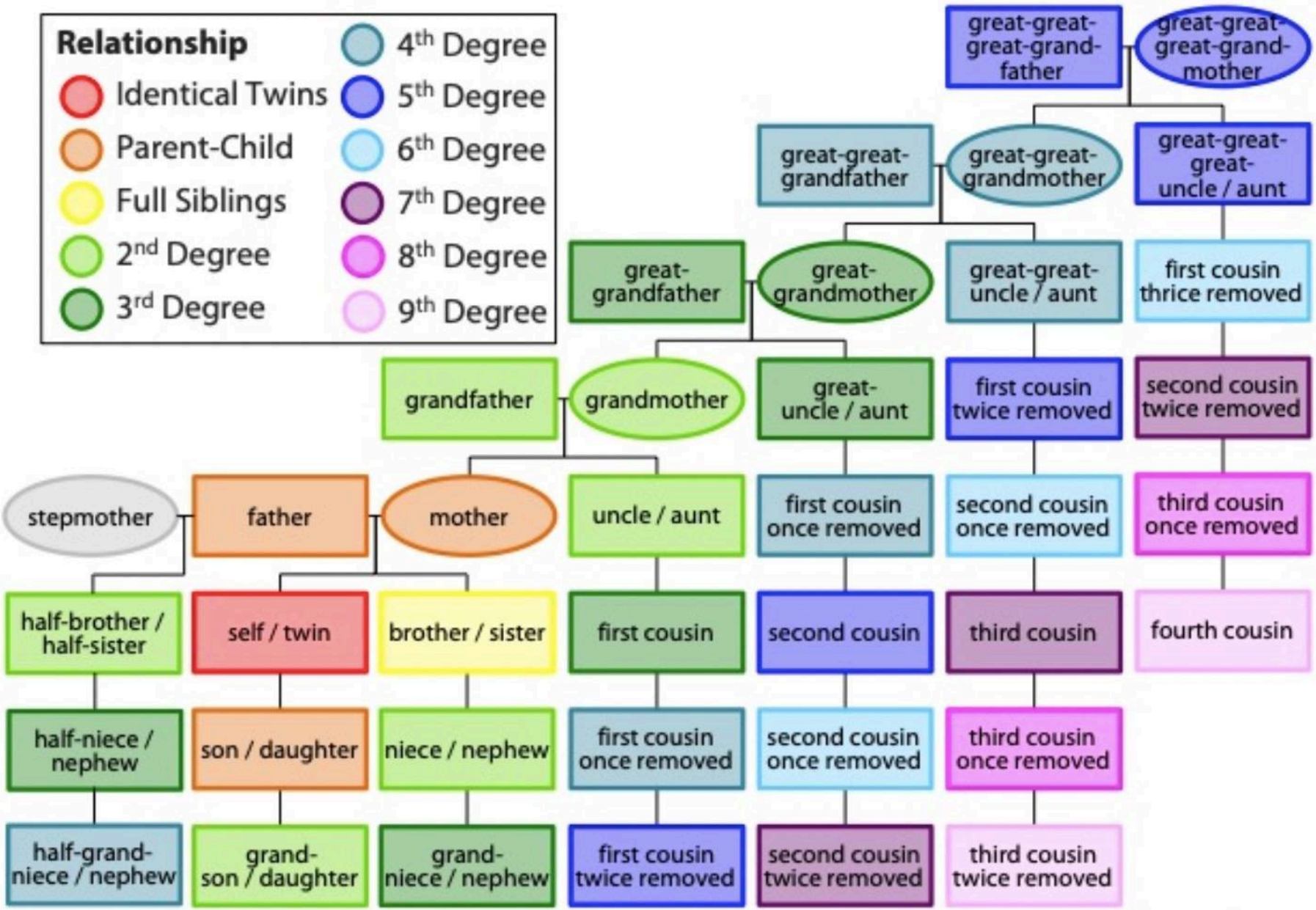


Figure 1: Pedigree showing the degrees of relatedness, as defined by the expected amount of shared DNA. Each relationship is defined with respect to the red "self / twin" box.

Chromosomes are mosaics of ancestral chromosomes