Ex Tetrad arg1- arg2- recessive to WT mutants
a b

- Type 1: 1 Arg+ 3 Arg- (T)
- Type 2: 4 Arg- (PD)

Parental: AB, ab = Tetratype
Ab, aB

ab, ab, AB, AB – Nonparental diptype

tightly linked case

Dist = $\frac{\#T \cdot 100}{2 \#Total}$

arg3- , lost to WT

arg3- x arg1-

Type	2	3
1 Arg+ 3 Arg-	4 Arg-	2 Arg+
T	PD	2 Arg- NPD

1:4:1 ratio, PD:T:NPD => unlinked

weakly linked: Dist = $100 \cdot \frac{T + 6NPD}{2 Total}$

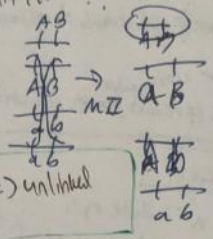
tightly linked – only T, PD

$\chi^2 = \sum_{\text{all phenotypic classes}} \frac{(o-e)^2}{e}$

χ^2	0.016	0.46
p	0.9	0.5

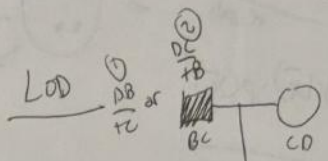
↓ can't reject null hypothesis

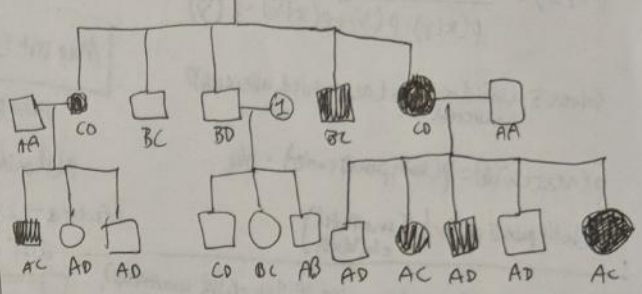
* COMPLEMENTATION TEST USEFUL ONLY W/ RECESSIVE MUTATIONS !!



4 mutants = P
3:1 mutants TT
2:2 mutants NPD

LOD





- autosomal dominant
- 1 must be AC
- cannot be completely linked as in

LOD score – second gen $\theta = 0.08$

$$LOD = \log \left[\frac{\theta^R \cdot (1-\theta)^{NR}}{0.5^{Tot}} \right]$$

$$LOD = \log_{10} \left[\frac{1}{2} \left(\frac{\theta^R \cdot (1-\theta)^{NR}}{0.5^T} \right)_{\text{Phase 1}} + \frac{1}{2} \left(\frac{\theta^R (1-\theta)^{NR}}{0.5^T} \right)_{\text{Phase 2}} \right]$$

R=4 NR=1 R=1 NR=4

LOD = -0.038

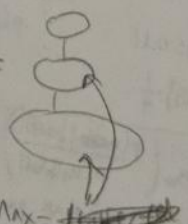
3rd gen LOD – we know phases of parent.

$$LOD = \log_{10} \left[\frac{\theta' \cdot (1-\theta)^2}{0.5^T} \right] = 1.06$$

– middle is uninformative

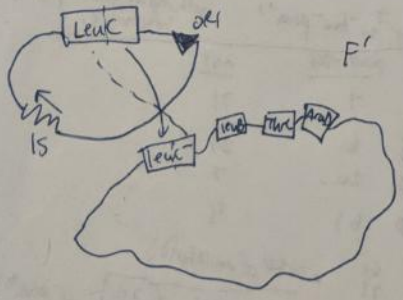
Total = 1.022

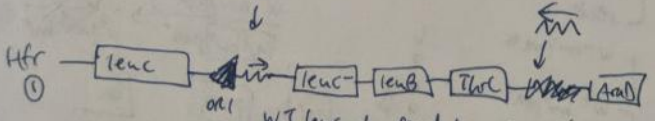
Simple LOD =



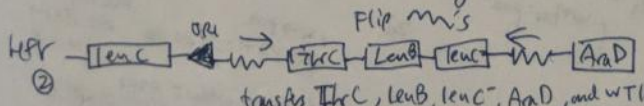
Max = ~~Hoke's~~

offspring · 0.3 – 0.3



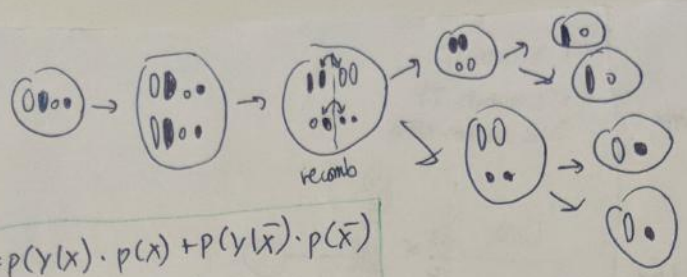


WT leuc transferred long after others.



transfers ThrC, LeuB, leuC⁻, AraD, and WT leuC last.

Use merodiploid (hfr ①) to make w/ double mutants. # of WT phenotypes suggests order of genes.



- unlinked: #parentals = #recombs
- weakly linked: #parentals ≥ #recombs
- tightly linked: #parentals >>> #recombs

Look at recombination tests with heterozygotes and cross w/ WT or recessive homozygote
- Look at combos of gametes to form diplotypes

$$p(Y) = p(Y|X)\cdot p(X) + p(Y|\bar{X})\cdot p(\bar{X})$$

$$p(Y|X) = \frac{p(X|Y)\cdot p(Y)}{p(X|Y)\cdot p(Y) + p(X|\bar{Y})\cdot p(\bar{Y})}$$

Given 5 children unaffected, p(next child affected)?

p(next child) = p(both parents carrier) · 1/4

p(both parents carrier | 5 unaffected children)

$$\text{Map dist (genetic distance)} = 100 \cdot \frac{\#\text{recombs}}{\#\text{gametes}} \text{ (cM)}$$

A↔B 18 cM = p(recomb) = 0.18

p(double recomb) = p(recomb A) · p(recomb B)

3 factor → 2 factor cross dist = #recombs gametes which is double $\frac{2 \cdot \#\text{double recombs}}{\text{total}} \cdot 100$

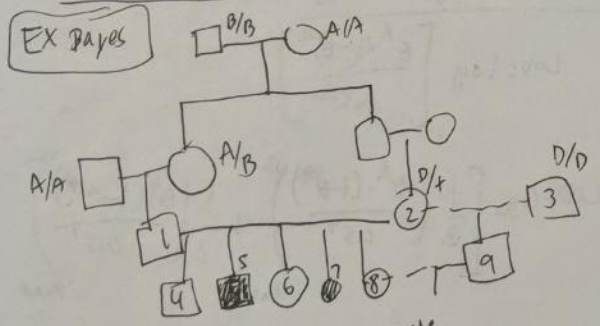
Probability of both parents carriers if 1st child unaffected?

X Y

p(Y|X) = __

So p(X|Y) = Bayes p(X) calc'd p(Y|X̄) = calc'd

EX Bayes



- autosomal recessive

1) Given #3 not carrier → 9 → 1/2 prob carrier
8 → 1 is D/+, 2 is D/+
so 8 can be D/D, D/+ or D/+.

2/3

2) #8↔#9 5 unaffected children, next child affected?

p(next affected) = p(9 carrier) · 1/4 D/+ · D/+ → +/+

$p(Y|X) = (\frac{3}{4})^5$

5 children unaffected; both parents carrier

p(X|Y) = 0.11

p(next) = p(X|Y) · 1/4.

Solve for p(X|Y) p(X) = 1/3 = 1/2 · 2/3

p(Y|X̄) = 1

p(X̄) = 2/3

3 Factor cross

w⁻: white eyes
bw⁻: bent wings
para^ts: paralysis
X-linked recessive

P: ♂ X ♀ $\frac{w^- bw^- para^{ts}}{w^- bw^- para^{ts}}$

↓

- F1: ♂ $w^+ bw^- para^{ts}/Y$ ♀ $\frac{w^+ bw^- para^{ts}}{WT\ X}$
- F1 females will informative meioses b/c het.
- Cross F1 ♀ w/ ♂ $w^- bw^- para^{ts}$ ... both males & female progeny are informative.

	paralyzed	not
Normal eyes, wings	4	71
White eyes, normal wings	6	23
Normal eyes bent wings	20	7
White eyes bent wings	65	4
$w^- bw^- para^{ts}$ recip	65	71
$w^- bw^+ para^{ts}$	6	7
$w^+ bw^- para^{ts}$	4	4
$w^+ bw^+ para^{ts}$	20	23

- Order of mutations:

$\frac{w^- bw^- para^{ts}}{w^+ bw^+ para^+}$ → $\frac{w^- para^{ts} bw^+}{w^+ para^+ bw^-}$

w para bw

- w → para #: 8+43 = 51/200 · 100 cM
- para → bw #: 13+8 → 10.5 cM
- w/o para → w → bw = 43+13/200 = 28 cM

LOD: $LOD = \log_{10}\left(\frac{p(\text{linked})}{p(\text{unlinked})}\right)$ $LOD_\theta = \log_{10}\left(\frac{\theta^R \cdot (1-\theta)^{NR}}{(1/2)^{Total}}\right)$ → Add LODs together for generations.

For F1, where phase parent is unknown → $LOD = \log_{10}\left(\frac{1}{2}(p(\text{phase 1})) + \frac{1}{2}p(\text{phase 2})\right)$

→ In F2 we will know phase of parent

Quick calc = +0.3 for informative, one time penalty −0.3 for unknown phase

need >3 to infer linkage

For a part – random A/B placement

A	B	b	B	b	b	B
A	B	b	b	B	B	b
a	b	B	B	b	B	b
a	b	B	b	B	b	B
	PD	NPD	T			

1:4:1 for unlinked

Chosen for gametes

Mapping

PD, NPD, TT, TT } results probs on second crossover (first is all T)

Total double = 4 # NPD → 4 recomb

T Tetrads from double = 2 # NPD Don't double count tetrads from double!

Total single = T Tetrads − 2#NPD → 2 recomb

$$\text{Distance} = 100 \times \frac{2(T-2NPD)+4(4NPD)}{4\ \#\text{Tetrads}} = 100 \cdot \frac{T+6NPD}{2T}$$

we can also apply this to experiments to count # spores of each type, determine linkage

LOD

Given pedigree and genotypes, consider linkage b/w SSRs and traits

SSR(A, B) w/ D/+

$$P(X|Y) = \frac{P(Y|X)\cdot P(X)}{P(Y)}$$

$$\frac{P(X|Y)}{P(\bar{X}|Y)} = \boxed{\frac{P(Y|X)}{P(Y|\bar{X})}} \cdot \frac{P(X)}{P(\bar{X})}$$

posterior odds = odds ratio · prior odds

X = linked
$\bar{X}$ = not linked
Y = data

If p > 0.95, post. odds = 20/1
prior odds = 1/50
odds must = 1000

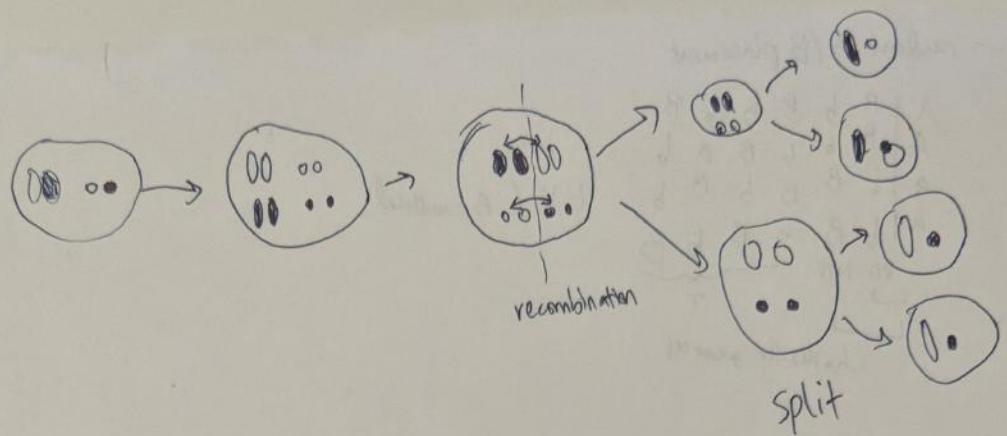
Ex $\frac{DA}{+A}$

LOD = $\log_{10}$ (odds ratio) should be > 3 for significance

Y|X → Assume that $\frac{DA}{+A}$.

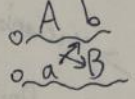
Y|$\bar{X}$ → Assume that $\frac{D}{+}$ $\frac{A}{B}$ Find probs for D/+, A/B separately

$P(\text{data}|X) = \frac{1}{2}(P|\text{phase 1}) + \frac{1}{2}(P|\text{phase 2})$



- unlinked
  - # parentals = # recombs
- weakly linked
  - # parentals > # recombs
- tightly linked
  - # parentals >>> # recombs

To look at recombination, test with heterozygotes, and cross with a wt or recessive homozygote.
Then, look at all combinations of gametes to form diploids.



to make AB
ab
Ab
aB
gametes.

Map Distance (genetic distance) = 100 · $\frac{\#\text{recombs}}{\#\text{gametes}}$ (cM)

Mapping function: relationship b/w physical distance (# crossovers) + genetic distance
cM

→ Identifying SSRs or markers related to mutation trait of interest

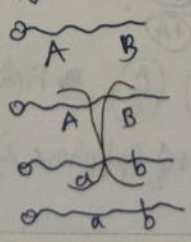
- Get the heterozygote like $\frac{wg^- \text{ (A alleles)}}{wg^+ \text{ (B alleles)}}$
- Cross with recessive homozygote, count the # of each genotype wg^-A, wg^+A, wg^-B, wg^+B received from parent by looking at progeny

- 3 factor crosses can account for double recombs
  - 8 possible genotypes, 4 reciprocal pairs
    - the rare class is product of double crossover, so try diff alignments to see what products result

→ Tetrad Analysis: Distance = f(Tetrad types)

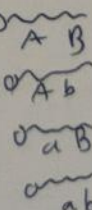
2 haploid → diploid → meiosis → 4 gametes

$\frac{AB}{ab} \times \frac{AB}{ab} \times 2 =$



PD

or w/ recomb



(T)

For tightly linked, this is what we have bc double recombs are rare.

Dist = 100 · $\frac{\sum T}{2\sum}$

Transposon Mutagenesis

λ-phage – bacterial virus → P gene (replication)
→ Int gene – integration gene (into bacterial genome)

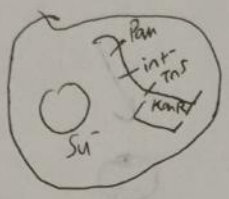
- Infect Su^- bacteria w/ λ Pam int^-::Tn5

Su+ would be able to suppress Pam mutation

Pam → P has nonsense → can't integrate

contains Tn5 transposon w/ KanR gene

– select for Kan R



Situation:
→ can't replicate virus
→ can't integrate into bacteria genome
→ we need Tn5 to jump into genome

Transposons can put KanR or other virus transferred genes into bacterial DNA

Ways to transfer DNA b/w bacteria:

1. Transduction: transfer of DNA via bacteriophage
2. Conjugation: unidirectional transfer via direct contact

$F^+/F'/Hfr \rightarrow F^-$

3. Transformation

---

1. TRANSDUCTION: infects, replicates

P1 phage → infect host, replicate DNA + package, lyses cell

1/300 of these will mistakenly have E. coli genomic DNA

Steps: 1) infect strain that has Tn5, collect lysate (virus)
2) infect WT strain (w/o Tn5)
3) select for Kan R

Donor + recipient genotypes are different to track changes!

↓

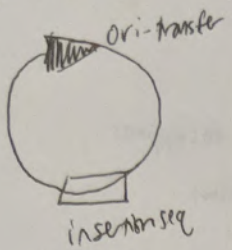
When in recipient, DNA from donor can recombine w/ homologous regions.

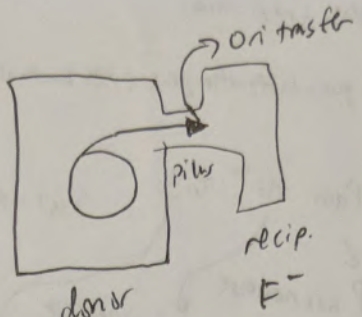
↓

recipient acquires donor sequence

cotrans f = # w/ ~~phenotype~~ / # resistant, ↑ freq = ↑ closer

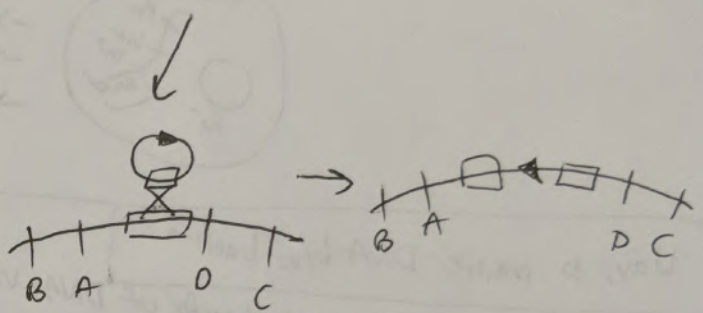
Conjugation – F-plasmid





↙
cell containing free F-plasmid

Hfr – cell w/ F-plasmid integrated into genome

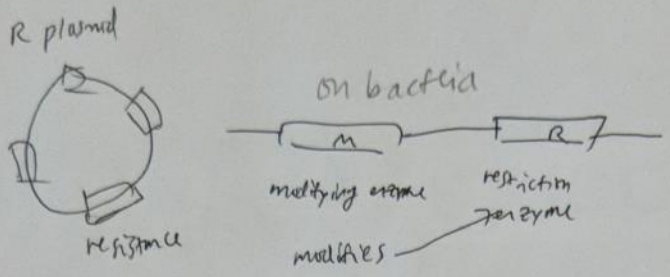


D→C→ B→ A

Transformation

Clone a gene

Find a specific protein



M+ R−, modifies no cut

M− R+ everything cut, bacteria doesn't survive

Generating a library: E.coli, restriction enzyme, mix ligate w/ plasmids, select

1) Library generate from mutant into plasmids
2) Transform a WT strain w/ plasmids, select
3) Lyse open the ones selected for to identify gene

Counting

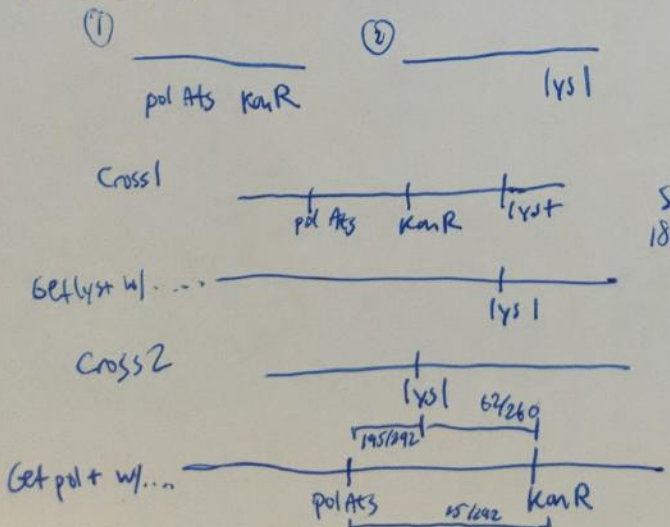
p(random stop codon) = 3/64

codons · p( ) = # random stop codons = # ORFs

$p(\text{ORF long enough}) = \left(\frac{61}{64}\right)^{\#\text{codons}}$

Random for fun: guess → #ORFs · p(ORF long enough) · 6

Cotransduction



Select lys:
183/260 have polAts
62 got KanR

Select pol+
97/292 lys+

277/292 KanR

Freqs
KanR, Lys 62/260
polAts − Lys 183/260

pol+ w/ Lys1: 195/292

pol+ KanS: 15/292
↳ V far

3-factor cross

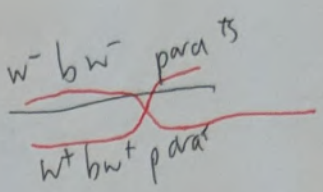
P ♂ × ♀ w⁻bw⁻ parents / w⁻bw⁻ parents

↓

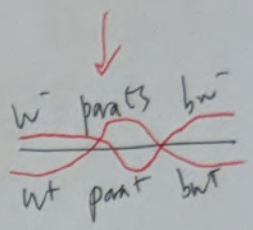
F1 ♂ w⁻bw⁻ parents / ♀ w⁻bw⁻ parents / w⁺bw⁺ parents × ♂ w⁻bw⁻ parents

↓

w⁻bw⁻ parents	65
w⁻bw⁻ parent	4 ← double
w⁺bw⁺ parents	4 ← double
w⁺bw⁺ parent	71
w⁻bw⁺ parents	6
w⁻bw⁺ parent	23
w⁺bw⁻ parents	20
w⁺bw⁻ parent	7
Total	200



vs



Dist w/ parn = $\frac{8+43}{200}$ = 25.5 cM

Dist parn, bw = $\frac{8+13}{200}$ = 10.5 cM

If middle dist considered

Dist w/bw = $\frac{43+13}{200}$ = 28 cM

If transitive 36 cM

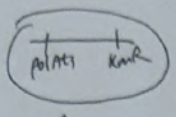
LOD score

$$\frac{P(\text{data} \mid \text{completely linked})}{P(\text{data} \mid \text{loci unlinked})}$$

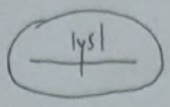
DA

+B

Cotransduction

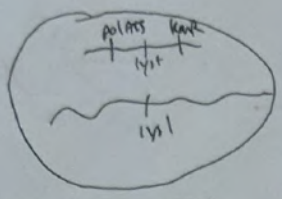


1



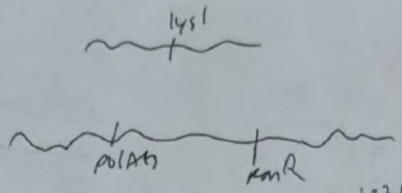
2

① Infect 1 w/ phage.
Infect 2.



pol+ lys+ 97/292 polAts lysI = 195/292
pol+ KanR 277/292 polAts/KanR = 15/292

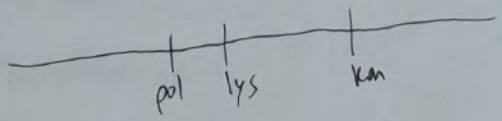
② Infect 2 w/ phage.
Infect 1.



lysS + polAts 183/260
lysR KanS 198/260 lysR KanR = 62/260

Cotrans freq
② pol & lys : 67%.
pol & KanR 5%.

① lys kan 24%.
lys polAts 67%.



3.4

HW

$f(a/a) = 600 \rightarrow 0.12$

$f(A/a) = 2000 \rightarrow 0.4$

$f(A/A) = 2400 \Rightarrow 0.48$

$p = f(A/A) + \frac{1}{2} f(A/a) = 0.48 + 0.2 =$ **0.68**

$q = f(a) = 0.32$

$H_0 \Rightarrow$ in HWE

$H_1 \rightarrow$ not in HWE

Exp
$f(a/a) = q^2 = 0.1024 \Rightarrow 512$
$f(A/a) = 2pq = 0.4352 \Rightarrow 2176$
$f(A/A) = p^2 = 0.4624 \Rightarrow 2312$

$\chi^2 = \sum \frac{(o-e)^2}{e} = \frac{(2312-2400)^2}{2312} + \frac{(2176-2000)^2}{2176} + \frac{(600-512)^2}{512}$

$= 32.71$. $p << 0.005$. Reject H_0.

Selection

Genotype	Freq	Freq after	Δ
A/A	p^2	p^2	0
A/a	$2pq$	$2p(1-s')$	$-2ps'$
a/a	q^2	$q^2(1-s)$	$-q^2s$

$\Delta q_{wt} + \Delta q_{mut} = 0$

$\frac{1}{2}(-2ps') - q^2s\nu = 0$

$\nu = ps' + q^2s$

Linkage

Controls: $A1 = 0.9$, $A2 = 0.1$, $B1 = 0.5$, $B2 = 0.5$

Cases:

	B1	B2
A1	30	90
A2	12	1

$D = P_{AB}P_{ab} - P_{Ab}P_{aB} = -0.059$

$D_{min} = \max\{-P_AP_B, -q_aq_b\} = \max\left\{-\frac{120}{133}\cdot\frac{42}{133}, -\frac{13\cdot 91}{133^2}\right\} = -0.0668$

$P(A1/B1) = 30/133$ $\quad P(A1) = 120/133$

$P(A2/B2) = 1/133$ $\quad P(B1) = 42/133$

$P(A1/B2) = 90/133$ $\quad P(A2) = 13/133$

$P(A2/B1) = 12/133$ $\quad P(B2) = 91/133$

$D_{max} = \min\left\{\frac{120}{133}\cdot\frac{91}{133}, \frac{42\cdot 13}{133^2}\right\} = 0.0309$

$D' = \frac{-0.059}{-0.0668} = 0.883$

allele correlation: $r^2 = \frac{D^2}{\quad} = 0.1828$

If r (recomb rate) $= 0.3$

$D_5 = (1-r)^5 \cdot (-0.059) = -0.0099$

Odds ratio

	SPD	Healthy
C	6213	14002
T	120	1054

$\text{Odds ratio}_{C,SPD} = \frac{(\text{ppl w/ C and SPD})(\text{ppl w/ T \& no})}{(\text{ppl w/ T and SPD})(\text{ppl w/ C \& no})}$

$= 3.897$.

more likely

SNP & SPD are...

H_0 - not asso

H_1 - assoc

$\chi^2 = \frac{(ad-bc)^2(\text{ratio total})}{(\text{row})(\text{col})(\text{total})(\text{col})} > 224.02$ $\quad df = 1$

$\rightarrow p < 0.005$

$\rightarrow$ reject H_0. They are assoc

(b) $\rightarrow$ measured 10^5 SNPs

$p' = \frac{0.05}{10^5} = 5\times10^{-7}$

prev p-value is still under threshold.

EXAM 3

- $p = f(A)$, $q = f(a)$ (freq of allele) $p+q=1$
- $f(A/A) + f(A/a) + f(a/a) = 1$

(A) • $p = f(A/A) + \frac{1}{2} f(A/a)$ (freq of genotype)

(A) • $q = \frac{1}{2} f(A/a) + f(a/a)$

By random mating:
$f(A/A) = p^2$
$f(A/a) = 2pq$
$f(a/a) = q^2$
$p_1 = p_0$

χ^2 to obtain p_{exp}, q_{exp} from experimental data (df = 1)
- calc random HW expected $p^2, 2pq, q^2$

$\chi^2 = \sum \frac{(o-e)^2}{e}$, df = # classes − # est params − 1

Mutation & Selection

$A \xrightarrow{\mu} a$

$\Delta q_{mut} = \mu \cdot p = \mu$ if $p = 1$

S = selective disadv ($0 \to 1$: no effect → all die)

$f = 1 - S$

(A) $\Delta f_{sel} + \Delta f_{mut} = 0$

(A) $\Delta q_{sel} + \Delta q_{mut} = 0$

Alternatives:
- $\Delta q_{inbreeding} + \Delta q_{sel} + \Delta q_{mut} = 0$
- $\Delta q_{sel} + \Delta q_{mut} = 0$

Δq_{sel} against homozygous (from random mating) + Δq_{sel} against homo (from inbreeding) + Δq_{mut}

Ex	Genotype	f	f_{after}	Δf
Het balance	A/A	p^2	p^2	
	A/a	$2pq$	$2pq(1+h)$	$2hq$
	a/a	q^2	$q^2(1-S)$	$-Sq^2$

$\Delta q_{sel} = \Delta f(a/a) + \frac{1}{2}\Delta f(A/a) = -Sq^2 + \frac{1}{2}(2hq)(+q)$

$q = h/s$

Inbreeding:
- increasing # homozygotes
- affects auto recessive #

F = inbreeding coeff = $\sum_{A_1, A_2 \ldots} p(\text{homozygous})$

$\Delta q_{inbreeding} = -SFq$

prob of someone affected by a disease who is inbred $\Rightarrow Fq$

Ex. $f(a/a)$ = % inbred (Fq) + % not inbred (q^2)

% of ppl as a result of inbreeding: $Fq \cdot q_{inbred}$

$Fqf + q^2$

Ex: mutation rate $10^{-6} = \mu$
$S = 0.9$, $h = 0.03$
steady state
$\Delta q_{sel\ homoz} + \Delta q_{sel\ hetero} + \Delta q_{mut} = 0$
$-Sq^2 + hq + \mu = 0$
$\Rightarrow q$ = steady state q

$F = 1/8$, % inbred = 15%

$0 = \mu + (-Sq^2)(0.85) + 0.15(-SFq)$ → steady state q

Recessive
$-q^2 S + \mu = 0$ $\quad q = \sqrt{\frac{\mu}{S}}$

Dom: $-\frac{1}{2}(2sq) + \mu = 0$
$\mu = Sq$

Haplotype: A b

LE expected $\Rightarrow p(AB) = p(A)p(B)$ etc

~~LE~~ estimated $\Rightarrow p_A = p_{AB} + p_{Ab}$ etc

LE → LD
- pop admixture
- ~~founder~~ effect
- selection

$D = p_{AB} \cdot p_{ab} - p_{Ab} \cdot p_{aB}$

$r^2 = \frac{D^2}{p_A q_a p_B q_b}$

$D_{min} = \max\{-p_A p_B, -q_a q_b\}$

$D_{max} = \min\{p_A q_b, q_a p_B\}$

$D' = \begin{cases} D/D_{min} & D < 0 \\ D/D_{max} & D > 0 \end{cases}$

$p_{AB}^{(n+1)} = (1-r)p_{AB}^{(n)} + r\, p_A p_B$, r = recomb rate

$D_n = (1-r)^n D_0$

$r = 0.5 \Rightarrow$ unlinked

Brother-sister Mating: $F = 1/4$

1st Cousins: $F = 1/16$

Ex | Consider autosomal recessive trait present in 1/250000 newborns. 6/13.

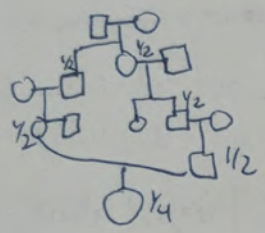
$f(b) \Rightarrow, b^2 = \frac{1}{250000}$, $b = \frac{1}{500}$

- homozygotes have 90% as many offspring as others.
what is μ for SS?

recessive → $-Sq^2 + \mu = 0$, $S = 0.1$
$q = b$

- consider effect of a change in mating patterns ST every person mates w/ cousin of one of their parents.
Frequency of recessive trait in 1st generation after?

$S = 0 = h = 0$ -



$4\left(\frac{1}{2}\right)^7 = F$ $f(bb) = Fq$

$f(N) = 0.0505$

↓

put into situation where $S = 0.9$, for dominant disease

$\Delta q = -qS = -0.9 \cdot 0.0505$

↓

$\frac{1}{2}(-2qS)$

$f(N)_{final} - f(N)_{init} = -0.9 \cdot 0.0505$

GWAS

- Case/control
  - cross-sectional (once in time)
  - cohort longitudinal - collected over time
  - case-cohort - look at cases/controls in specific group/pop

Observed:

	Cases	Controls	Total
A0	a	b	a+b
A1	c	d	c+d
Total	a+c	b+d	a+b+c+d

$$OR(A0) = \frac{(w/\ disease, A0)(w/o\ disease, A1)}{(w/o\ disease, A0)(w/\ disease, A1)}$$

↑ OR, more conc.

H_0 - not assoc

H_1 - assoc

Est

	Cases	Controls
A0	$\frac{(a+b)(a+c)}{N}$	..
A1	..	..

↓ Significance of OR ↓

$$\chi^2 = \frac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(b+d)(a+c)}, \quad df = 1 \quad (\#rows-1)(\#cols-1)$$

- 1 est case freq
- 1 allele freq

$$\chi^2 = \sum \frac{(o_i - E_i)^2}{E_i}$$

If $p <$ threshold:
- reject null hypothesis
- also χ^2 is high

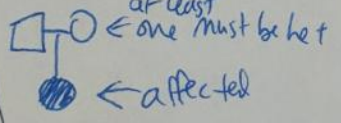
$p >$ threshold
- accept (don't reject) null hypothesis
- χ^2 is low → matches up w/ expected

TRIOS

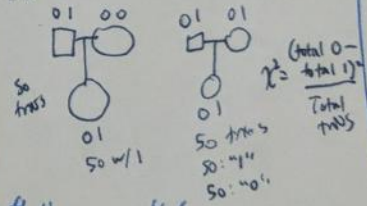
$$\chi^2_{TDT} = \frac{(n_1 - n_2)^2}{(n_1 + n_2)}$$

H_0 - alleles transmitted 1:1

H_A - preferential transmission



Ex 250 trios



*Count for each trio # times an allele was passed down from a het parent

$$\bar{X} = \frac{\sum f_i x_i}{N}$$ (f_i: # in class, x_i: val of class)

$$Var\ s^2 = \frac{\sum f_i (x_i - \bar{X})^2}{N-1}$$

s, standard dev $= \sqrt{s^2}$

$$\sigma_p^2 = \sigma_g^2 + \sigma_e^2$$

Broad sense heritability

$$H^2 = \frac{\sigma_g^2}{\sigma_p^2} = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

Narrow Sense: ratio of additive genetic var to total phenotypic diff

$$h^2 = \frac{M' - M}{M^* - M}$$

M = mean, parent gen

M^* = mean parents selected

M' = mean progeny

Mult Hypothesis Correction

decision / truth	H_0 not rej.	H_0 rej.
H_0 true	truly null	false pos (I)
H_0 false	false negative (II)	truly alt

*p-value = probability (Type I error) (α)

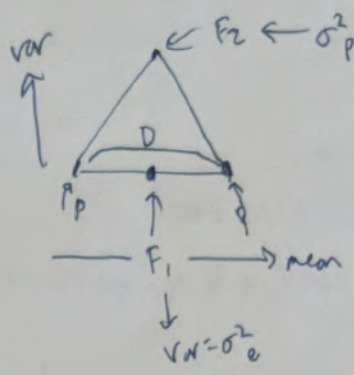
*Power = 1 − prob we made Type II error
= $1 - \beta$

*Control: prob of making at least 1 Type 1 error in a family of tests

New p-value for N tests → $p(A_i) \leq \alpha/N$ each test

Covariance b/w pairs of relatives

$$Cov(x,y) = \frac{\Sigma f_i (x_i - \bar{x})(y_i - \bar{y})}{N-1}$$



$$h = \frac{D^2}{8\sigma_g^2}$$

r = lin correlation coeff = $\frac{Cov(x,y)}{S_x S_y}$

Ex - monozygotic twins → $r = H^2$, $\sigma^2_p \simeq \sigma^2_e$

↑ Corr. coeff.

Ex. Full siblings $r = \frac{H^2}{2}$

EXAM #2

Strain 1 Strain 2

F1 X or

F2 x other F2

converges to homozygous recomb

$(1-\frac{1}{8})^n$ genome heterozygous

n=0 @ F1 → random F2 x F2

QTL – minor contributors over major one

Strong phenotype x strain w/o phenotype

↓

P1 backcross 2 gen

more and more like strain 1

↓

7 rounds

→ if you want something specific, select for the trait at every gen

Transposon – can confer resistance into bacteria like Tn5 – KanR

For E coli, we use λ phage

λ P_{am} int^- Tn5

Su+ – suppress P_{am}

Su^- – only way to get KanR is to jump in replication

- P replication P_{am} = amber mutation in replication
- Int integration int^- – can't integrate into bacterial genome

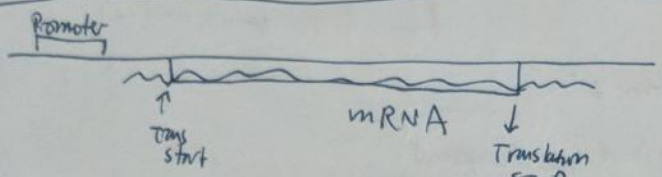
Strategy: Infect Ecoli w/ λ P_{am} int^- Tn5

select for kanR

These have Tn5 put in randomly somewhere

*Identifying where gene A is

This can knock out A if Tn5 jumps into middle of it.



Gene Seq – p(stop codon) = 3/64

#ORFs (length ≥ k) = (p(not stop))^k • #total ORFs • b, k = #codons

p(ORF ≥ k) = (p(not stop))^k

total ORFs = # codons in genome • p(stop) = # random stops

Mutations

Missense: base change

Nonsense: codon → stop (via base change)

Frameshift: indel

Screen – identifying an altered prop

Selection – parent strain can't grow, desired mutant can

Base Analog – something inserted, interpreted as one thing, can bond to another

Base-Modifying – chemical damage – repairs hard errors

DNA – intercalate – □ b/w bases – miscopying by polymerase

Suppressors: reversants – mutants that reverse effect of primary mutation

1) Back mutation

2) Intragenic – mutation in same gene

3) Extragenic – compensating mutation in diff gene

↳ Nonsense suppressors (like Am suppressor Su+)

↓

works via mutations to tRNA (codon is right, but attached AA isn't)

PCR

TRANSDUCTION

1) infect strain that has Tn5, collect virus

2) infect WT strain (or other strain)

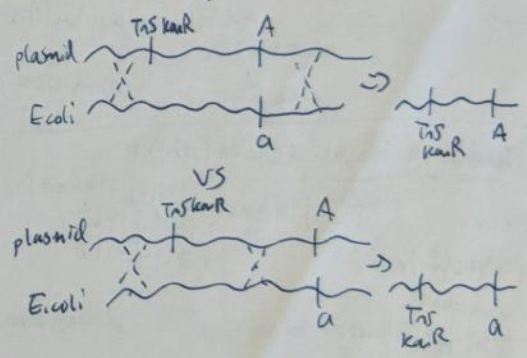
~~observe recombinations~~

3) select for trait

4) screen for phenotype

Cotransf b/w trait 1, trait 2 = # w/ trait 2 / # selected by trait 1

Cotransf ↑ dist ↓



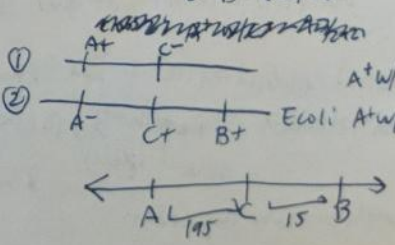
Ex: ① A⁻ B+ C+

② A+ B⁻ C⁻ Ecoli

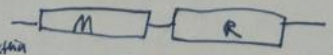
Cross 1: C+ A⁻ = 183/260

C+ B+ = 62/260

→ stuff from plasmid!



Transformation



M+R⁻ – modifies, doesn't cut any phage

M^-R^+ – everything it cut inc. bacteria

Generate a library – Ecoli, restriction enzyme, mix/ligate w/ plasmids, select

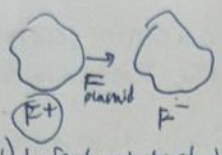
1) Library generate from mutant into plasmids

2) transform a WT strain w/ plasmid, select

3) Lyse open one that you want

* Clone a gene? * Find a gene?

CONJUGATION

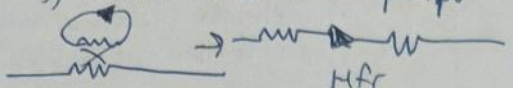


Hfr = cell w/ integrated F plasmid into genome

*FUNCTIONAL COMPLEMENT

1) Infect mutant strain w/ P1 phage
2) Use lysate to infect WT
3) Select for KanR and desired phenotype

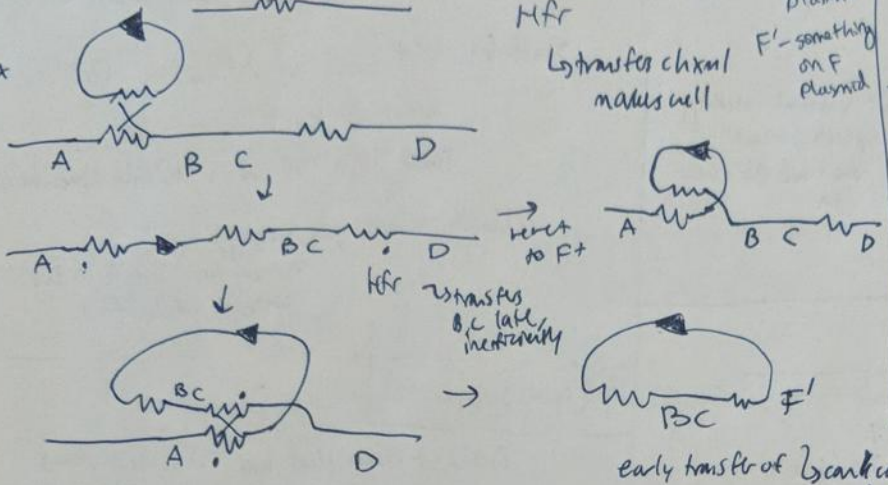
F+ plasmid (empty)
F+ → F- does not transfer chromosomal markers well!



F' - something on F plasmid

↳ transfer chromosomal markers well

Ex

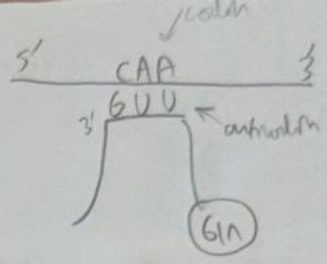


↳ revert to F+

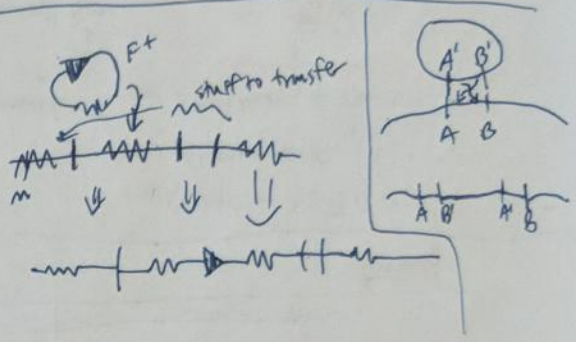
↳ transfers B, C late, inefficiently

early transfer of ↳ can't complement B, C — dom/rec gene?

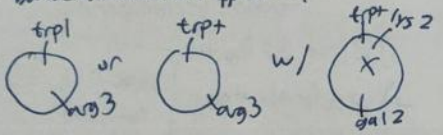
tRNA



= 5' TTG 3'
3' AAC

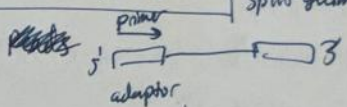


PSET 3.4: bacterial mutation vs suppression?

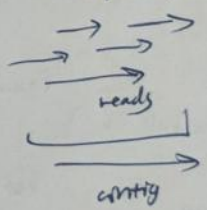


Transduct Q into X.
Select w/ lysine.
Are all resulting bacteria WT w/ respect to trp? or no?
↳ bacteria ↳ extragenic suppressor

split genome into set size pieces, attach adaptors



duplicate, cut into pieces



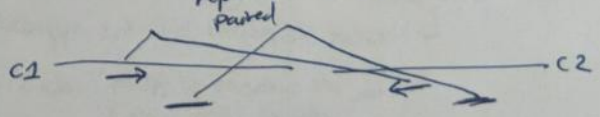
Ex. C G A T C T T A A G C C T T A T A

These 2 are a read-pair

paired end reads help us determine b/w

→ ←

repetitive elements



Needleman Wunsch: Global align

$F(i,j) = \max \begin{cases} F(i-1,j-1) + score(i,j) \\ F(i-1,j) + d \\ F(i,j-1) + d \end{cases}$

$P(0,0) = 0$
$F(i,0) = d \cdot i$
$F(0,j) = d \cdot j$

space = mn
complexity = $O(n^2 - 1)mn$

n = # sequences
2^n - 1 refers to boxes checked per step

d = gap pen.

Smith Waterman: Local

add 0 to F(i,j) eq → conserved regions via evolution

↳ SIMILARITY
↳ DISSIMILARITY

Dot plots to identify diff. genes → simply fxn

*In exons: conserved regions + low base changes (nonsyn)

* Introns/elsewhere: ∃ frameshifts, gaps (indels), etc among related species

*indels are clean when ∃ AA

*watch for early stop codons