

Introduction to QTL (Quantitative Trait Loci) & LOD analysis

Steven M. Carr / Biol 4241 / Winter 2016

- **Quantitative Trait Loci:** contribution of multiple genes to a single trait
- **Linkage** between phenotypic trait & genotype markers
- **X-linked trait**
 - => restricts search to **X chromosome**
 - Cf. **GWAS (Genome-Wide Association Study)**
- **Pedigree analysis** of pairs of brothers
 - **Concordance of X:** do brothers inherit *same X* from mother?
 - **Identity by Descent [D]:** *same X allele*
 - **Discordance [n]:** *different X alleles*
- **Logarithm of Odds (LOD)** calculation
 - Given two *very unlikely* outcomes, which is *less unlikely / more likely*?
 - Is the difference statistically significant?
- **Disclaimer:** The experimental design and numerical data in Hamer et al. (1993) are presented as an example of a simplified QTL / Pedigree / LOD analysis. No opinion is expressed as to any implications of the results. Readers are referred to the MS itself for the author's conclusions.

NEWFOUNDLAND

RESEARCH ARTICLE SCIENCE • VOL. 261 • 16 JULY 1993

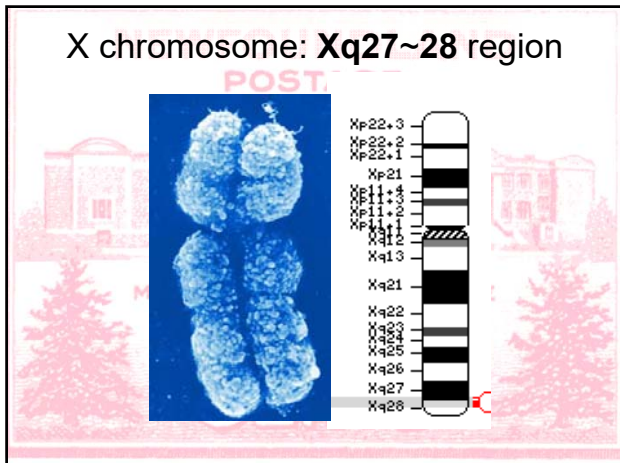
A Linkage Between DNA Markers on the X Chromosome and Male Sexual Orientation

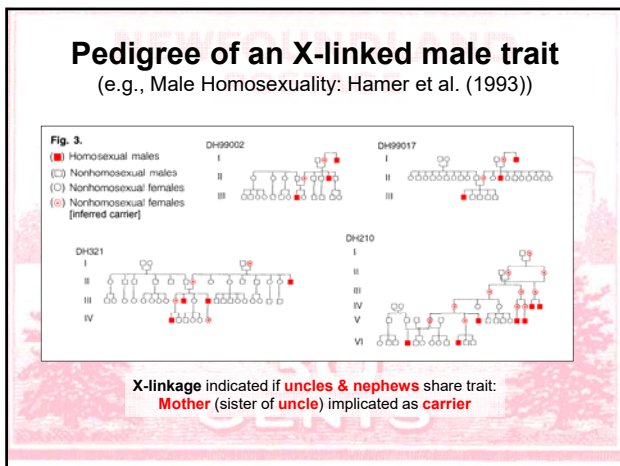
Dean H. Hamer, Stella Hu, Victoria L. Magnuson, Nan Hu, Angela M. L. Pattatucci

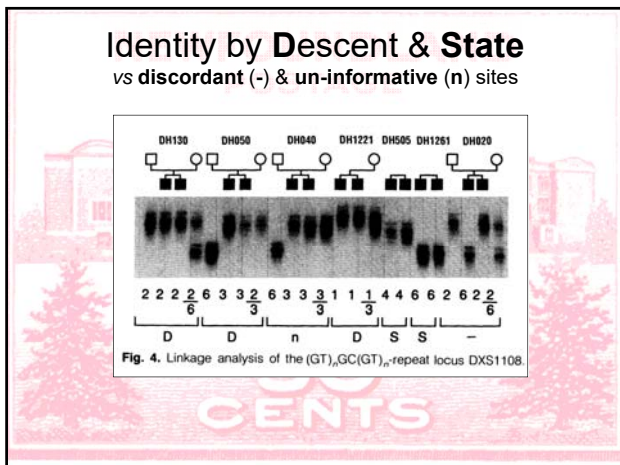
The role of genetics in male sexual orientation was investigated by pedigree and linkage analyses on 114 families of homosexual men. Increased rates of same-sex orientation were found in the maternal uncles and male cousins of these subjects, but not in their fathers or paternal relatives, suggesting the possibility of sex-linked transmission in a portion of the population. DNA linkage analysis of a selected group of 40 families in which there were two gay brothers and no indication of nonmaternal transmission revealed a correlation between homosexual orientation and the inheritance of polymorphic markers on the X chromosome in approximately 64 percent of the sib-pairs tested. The linkage to markers on Xq28, the subtelomeric region of the long arm of the sex chromosome, had a multipoint lod score of 4.0 ($P = 10^{-5}$), indicating a statistical confidence level of more than 99 percent that at least one subtype of male sexual orientation is genetically influenced.

Study Design of Hamer et al. (1993)

- **QTL (Quantitative Trait Loci)** study:
 - Possible *genetic influence* on (behavioral) *phenotype*
- **Linkage** between phenotype & genotype markers
- Expectations for an **X-linked trait**
 - *Brothers share trait* with *uncle* (or cousin), *but not father*
 - *E.g.*, Sexual orientation of gay male brothers [one subtype]
 - => simplifies search to **X chromosome**
- **Pedigree analysis** of pairs of brothers
 - **Concordance of X:** do brothers inherit *same X* from mother?
 - **Identity by Descent [D]:** *same X allele*
 - **Discordance [n]:** *different X alleles*
 - **Similarity by State [S]:** copies of *similar X allele*
 - (if mother is *homozygous* for marker)
 - **Uninformative [-]:** pedigree data missing
 - **Logarithm of Odds (LOD)** calculation
 - $Z_1 = \text{prob}(D)$, given *observed allele frequencies*
 - $\text{LOD} = \ln[Z_1 / \text{prob}(null)]$, where $p(null) = \text{random expectation}$
 - Cf. Poker Hands







Identity by Descent & State: vs discordant (-)

DH130
 Identical by *Descent*

DH040
 Identical by *State*

DH020
 Discordant (-)

LOD score analysis of Five-Card Draw Poker (Logarithm of Odds)

Given 52 cards, there are $(52 \text{ choose } 5) = 2,598,960 = 2.6 \times 10^6$ possible five-card hands
 The probability of any random hand = 3.85×10^{-7}

$P(\text{flush}) = (13 \text{ choose } 5) = 1,287$ for each of 4 suits = $5,148 / 2,598,960 = 1.98 \times 10^{-3}$

$P(4\text{-of-a-Kind}) = 13 \text{ 4s-of-a-kind, and } 48 \text{ other cards to fill} = (13 \times 48) / 2,598,960 = 2.40 \times 10^{-4}$

$\log_{10}(1.98 \times 10^{-3}) = -2.70$ $\log_{10}((2.40 \times 10^{-4}) / (1.98 \times 10^{-3})) = -0.92$ $\log_{10}(2.40 \times 10^{-4}) = -3.62$

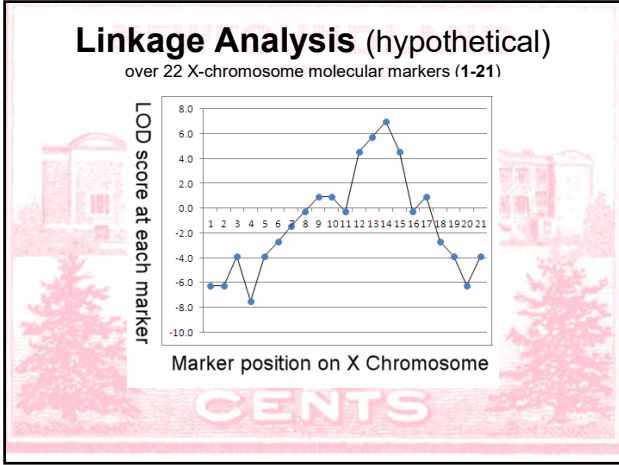
A flush or 4-of-a-kind, though unlikely, are far more likely ($\sim 10^{-3-4}$) than a random hand
 4-of-a-kind is less likely (and therefore beats) a flush

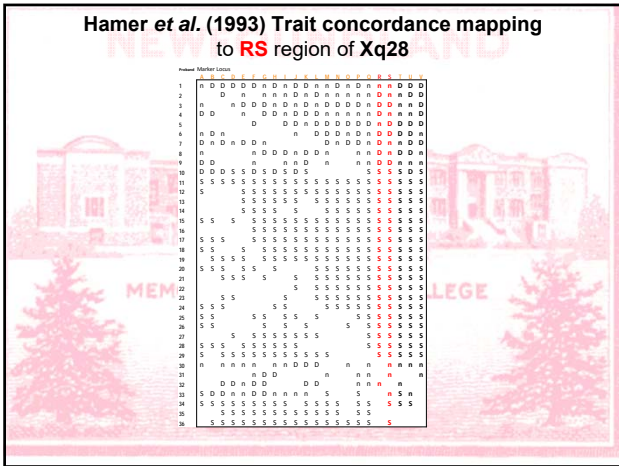
Note: evaluate $(n \text{ choose } k) = n! / k!(n-k)!$

LOD score analysis (simplified)

Consider 21 possible concordance ratios: *What odds?*
 Consider 21 successive markers with observed ratios: *Is there Linkage?*

Concordance	non-concordance	ratio	odds ratio (O.R.)	LOD	Chi Sq	Marker	Concordance	non-concordance	odds ratio (O.R.)	LOD
0	40	0.00	0.00	-15.9	40.00	A	16	24	0.00	-6.3
2	38	0.05	0.00	-14.7	32.40	B	16	24	0.00	-6.3
4	36	0.11	0.00	-13.5	25.60	C	20	20	0.00	-3.9
6	34	0.18	0.00	-12.3	19.60	D	14	26	0.00	-7.5
8	32	0.25	0.00	-11.1	14.40	E	20	20	0.00	-3.9
10	30	0.33	0.00	-9.9	10.00	F	22	18	0.00	-2.7
12	28	0.43	0.00	-8.7	6.40	G	24	16	0.03	-1.5
14	26	0.54	0.00	-7.5	3.60	H	26	14	0.54	-0.3
16	24	0.67	0.00	-6.3	1.60	I	28	12	8.71	0.9
18	22	0.82	0.00	-5.1	0.40	J	30	10	139.38	2.1
20	20	1.00	0.00	-3.9	0.00	K	32	8	2230.07	3.3
22	18	1.22	0.00	-2.7	0.40	L	34	6	35681.19	4.6
24	16	1.50	0.03	-1.5	1.60	M	36	4	570899.08	5.8
26	14	1.86	0.54	-0.3	3.60	N	38	2	9134385.23	7.0
28	12	2.33	8.71	0.9	6.40	O	34	6	35681.19	4.6
30	10	3.00	139.38	2.1	10.00	P	30	10	139.38	2.1
32	8	4.00	2230.07	3.3	14.40	Q	26	14	0.54	-0.3
34	6	5.67	35681.19	4.6	19.60	R	22	18	0.00	-2.7
36	4	9.00	570899.08	5.8	25.60	S	20	20	0.00	-3.9
38	2	19.00	9134385.23	7.0	32.40	T	16	24	0.00	-6.3
40	0	—	146150163.73	8.2	40.00	U	20	20	0.00	-3.9





LOD scores (long calculation)

Let z_1 = the probability that a pair of brothers share a marker allele by-descent
 (z_1 is an unknown parameter that is estimated from the data).
 $\Gamma(z_1)$ = the likelihood of the observed data at z_1
 $\Gamma(1/2)$ = the likelihood of the observed data under the null hypothesis of $z_1 = 1/2$
 $L(z_1) = \Gamma(z_1) / \Gamma(1/2)$ = The ratio of the unlikely observed event to the unlikely random event.
 (where brothers share any allele with $p = z_1 = 1/2$)

$$2 \sum_{j=1}^{N_j} \ln[L_j(z_1)] =$$

$$2 \left(\sum_{i=1}^D \ln[(1 - z_1)(1/2)] + \sum_{i=1}^D \ln[z_1(1/2)] + \sum_{j=1}^S \ln[(f_j z_1 + f_j - f_j z_1)(f_j(1 + f_j)/2)] \right) =$$

$$2 \left(G \ln(2 - 2z_1) + D \ln(2z_1) + \sum_{j=1}^S \ln[2z_1 + 2f_j - 2f_j z_1](1 + f_j) \right)$$
