RESEARCH ARTICLE SCIENCE • VOL. 261 • 16 JULY 1993

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

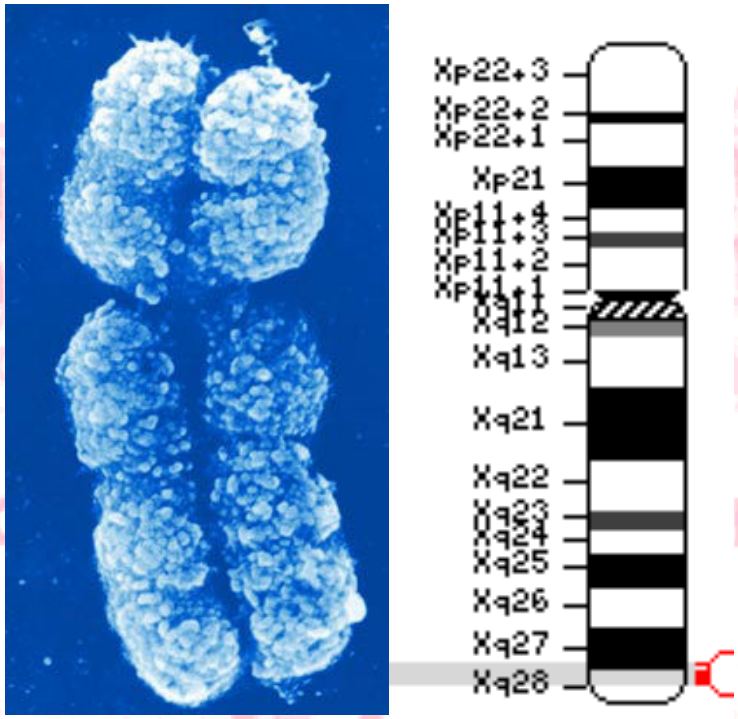
Dean H. Hamer, Stella Hu, Victoria L. Magnuson, Nan Hu, Angela M. L. Pattatucci

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.

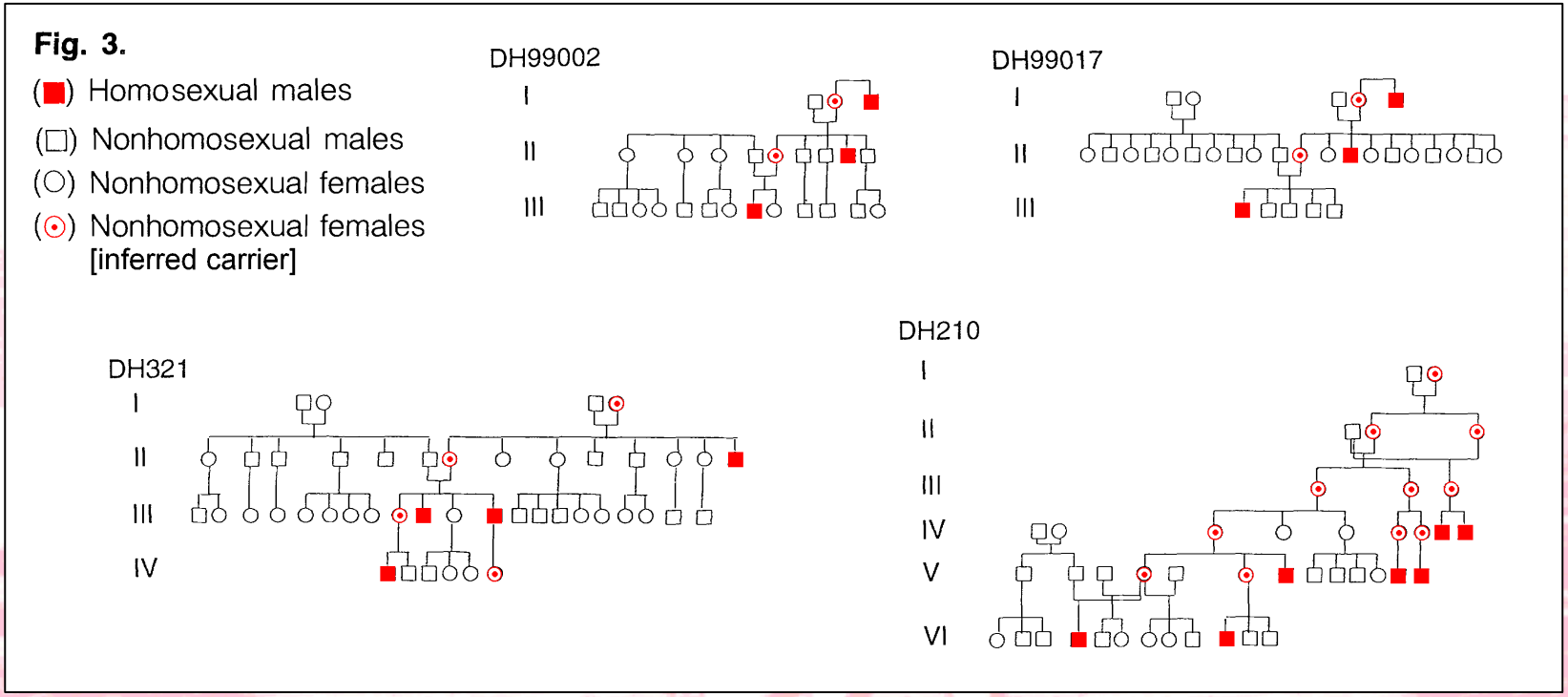
Study Design of Hamer *et al.* (1993)

- **QTL (Quantitative Trait Loci)** study:
  Possible *genetic influence* on (behavioral) *phenotype*
- **Linkage** between phenotype & genotype markers
- Same-Sex orientation of gay males [one subtype]
  - Behavior shared with gay cousin / uncle *but not* father
  - => **X-linked inheritance**
  - => simplifies search to **X chromosome**
- **Pedigree analysis** of pairs of brothers
  - Genetic **concordance** of **X** inherited from mother
    - Similarity by **S**tate [**S**]
    - Identity by **D**escent [**D**]: copies of *same allele*
    - Discordant [**n**]: brothers get *different* alleles
    - Uninformative [-]
- **Logarithm of Odds** (**LOD**) calculation
  - Z_1 = prob(**D**), given *observed* allele frequencies
  - **LOD** = *ln*[Z_1 / prob(*null*)], where p(*null*) = *random* expectation
  - Cf. Five-Card Stud

X chromosome: **Xq27~28** region



X-linked male homosexuality?



X-linkage is indicated where **uncles & nephews** share trait:
Mother (sister of **uncle**) is implicated as **carrier**

Identity by **(S)tate** & **(D)escent**

vs **discordant** (**-**) & **un-informative** (**n**) sites

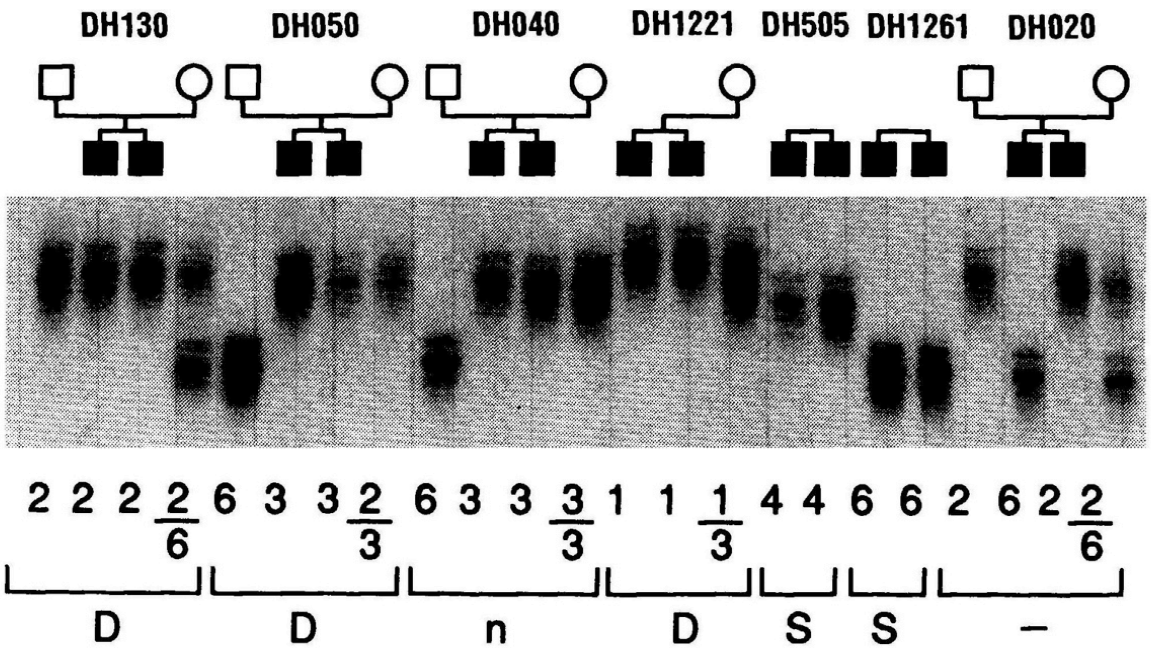
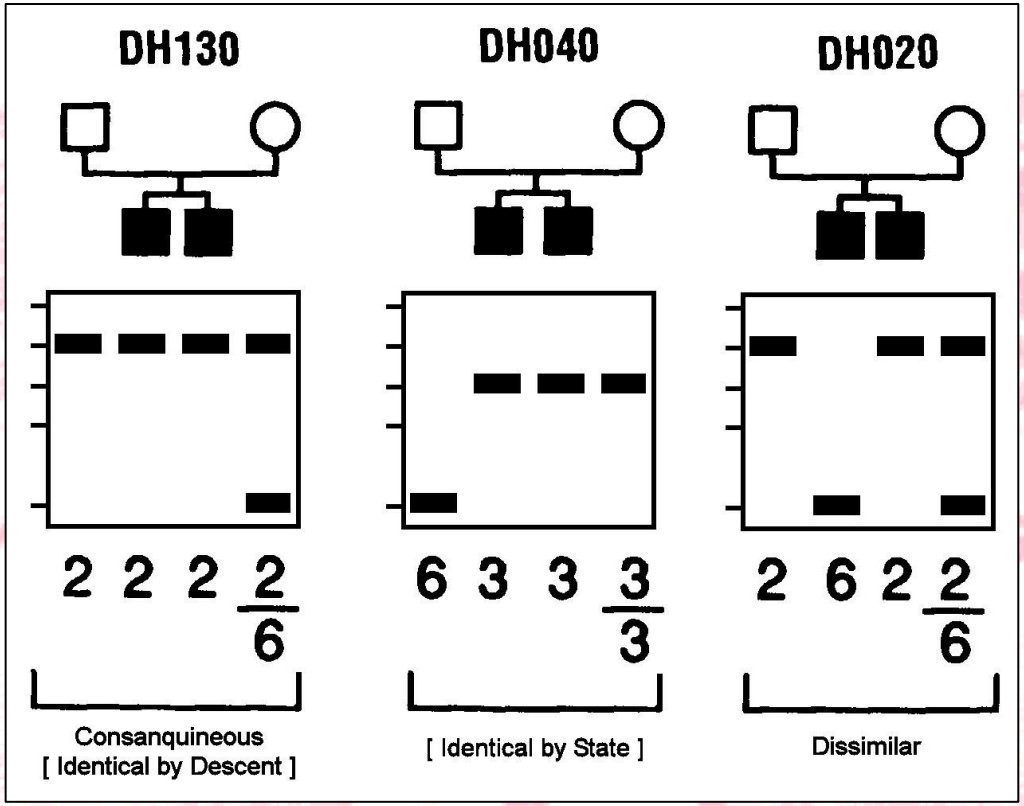


Fig. 4. Linkage analysis of the $(GT)_nGC(GT)_n$-repeat locus DXS1108.

Identity by **Descent** & **State:**

consanguineous (**IbD** or **IbS**) *vs* dissimilar **X** chromosomes



Trait mapping to **RS** region of **Xq28**

Proband	Marker Locus																				
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	n	D	D	D	D	D	n	D	n	D	D	n	n	D	n	D	n	n	n	D	D	D
2			D		n		n	n	n	D	n	D	n	n	n	n	n	D	n	n	D	D
3	n			n	D	D	D	n	D	n	D	n	D	D	D	D	n	D	D	n	n	D
4	D	D			n		D	D	n	D	D	D	n	n	n	n	n	D	n	n	n	D
5						D			D	D	n	D	D	D	D	D	D	n	D	D	D	D
6	n	D	n							n		D	D	D	n	D	n	D	D	D	D	n
7	D	n	D	n	D	D	n						D	n	D	D	n	D	n	n	D	D
8	n					n	D	D	D	n	D	D	n			n	n	D	n	D	D	n
9	D	D				n			n	n	D		n			n	n	D	D	n	n	n
10	D	D	D	S	S	D	S	D	S	D	S						S	S	S	S	D	S
11	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
12	S				S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
13					S	S	S	S	S	S		S	S	S	S	S	S	S	S	S	S	S
14					S	S	S	S		S			S	S	S	S	S	S	S	S	S	S
15	S	S		S		S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
16						S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
17	S	S	S			S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
18	S	S			S		S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
19		S	S	S	S		S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
20	S	S	S		S	S		S				S	S	S	S	S	S	S	S	S	S	S
21			S	S	S		S			S		S	S	S	S	S	S	S	S	S	S	S
22										S		S	S	S	S	S	S	S	S	S	S	S
23			S	S					S			S	S	S	S	S	S	S	S	S	S	S
24	S	S	S					S	S				S	S	S	S	S	S	S	S	S	S
25	S	S				S	S		S	S		S				S	S	S	S	S	S	S
26	S	S					S		S		S				S		S	S	S	S	S	S
27				S		S	S	S	S	S	S	S					S	S	S	S	S	S
28	S	S	S		S	S	S	S	S	S							S	S	S	S	S	S
29	S		S	S	S	S	S	S	S	S	S	S	S					S	S	S	S	S
30	n		n	n	n	n		n	n	D	D	D			n		n		n	n	n	n
31						n	D	D					n			n	n		n			n
32			D	D	n	D	D				D	D				n	n	n		n		
33	S	D	D	n	n	D	D	n	n	n	n		S			S			n	S	n	
34	S	S	S	S	S	S	S	S	S		S	S	S	S		S	S		S	S	S	
35			S	S	S	S	S	S	S	S	S	S	S	S		S	S					
36		S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S		S			

Linkage Analysis

over 22 X-chromosome molecular markers (**A-V**)

Table 2. Summary of linkage results. Linkage analysis was performed on 40 male homosexual sib-pairs; 22 X chromosome markers were used (*30*). The five marker loci on distal Xq28 are in boldface.

Locus	Location	AL*	HET†	Sib-pairs‡			z_1§	$2\ln L(z_1)$‖	P¶
				[D]	[S]	[–]			
A. .KAL	p22	6	0.77	5	16	14	0.51	0.01	ns
B. .DXS996	p22	11	0.84	7	14	18	≤.5	≤0	ns
C. .DXS992	p	8	0.87	6	13	19	≤.5	≤0	ns
D. .DMD1	p21	9	0.78	3	10	23	≤.5	≤0	ns
E. .DXS993	p11	6	0.80	3	14	17	≤.5	≤0	ns
F. .DXS991	p	8	0.77	8	14	14	0.57	0.61	ns
G. .DXS986	q	10	0.71	7	20	10	0.65	2.11	ns
H. .DXS990	q	7	0.76	4	19	13	0.55	0.25	ns
I. .DXS1105	q	5	0.48	3	20	9	≤.5	≤0	ns
J. .DXS456	q21	10	0.85	8	20	8	0.75	7.95	0.00241
K. .DXS1001	q26	10	0.82	8	16	13	0.60	1.09	ns
L. .DXS994	q26	5	0.75	7	17	13	0.55	0.26	ns
M. .DXS297	q27	5	0.70	5	21	8	0.71	4.25	0.01963
N. .FMR	q27	17	0.79	6	17	14	0.56	0.45	ns
O. .FRAXA	q27	8	0.72	4	17	13	0.56	0.38	ns
P. .DXS548	q27	6	0.67	7	20	7	0.73	5.21	0.01123
Q. .GABRA3	q28	4	0.35	2	23	3	0.74	2.39	ns
R. .DXS52	**q28**	**12**	**0.79**	**9**	**22**	**6**	**0.81**	**11.83**	**0.00029**
S. .G6PD	**q28**	**2**	**0.36**	**4**	**24**	**2**	**0.85**	**6.38**	**0.00577**
T. .F8C	**q28**	**2**	**0.41**	**5**	**24**	**3**	**0.82**	**6.56**	**0.00522**
U. .DXS1108	**q28**	**6**	**0.71**	**8**	**22**	**4**	**0.85**	**12.87**	**0.00017**
V. .DXYS154#	**q28**	**10**	**0.71**	**8**	**22**	**5**	**0.83**	**12.84**	**0.00017**
R/S/T/U/V	**q28**		**0.99**	**12**	**21**	**7**	**0.82**	**18.14**	**0.00001**

*AL is the number of different alleles observed in 62 to 150 independent chromosomes. †HET is the calculated heterozygosity; HET = $1 - \Sigma f_i^2$, where f_i = frequency of the ith allele. ‡[D] is the observed number of concordant-by-descent pairs; [S] is the observed number of concordant-by-state pairs; [–] is the observed number of discordant pairs; noninformative pairs are not included in this analysis. §z_1 is the estimated probability that two homosexual brothers share the marker locus by-descent (*31*). ‖$L(z_1)$ is the ratio of the likelihoods of the observed data at z_1 versus the null hypothesis of $z_1 = 1/2$ (*31*). ¶P (one-sided) was calculated by taking $2\ln L(z_1)$ to be distributed as a chi-squared statistic at one degree of freedom; ns· $P > 0.05$. # Only the maternal, X-linked contribution was considered for this sex-linked locus (*23*).

LOD score analysis (simplified)

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

Concordant	non-Concordant	ratio	odds ratio (0.8)	LOD	Chi Sq
0	40	0.00	0.00	-15.9	40.00
2	38	0.05	0.00	-14.7	32.40
4	36	0.11	0.00	-13.5	25.60
6	34	0.18	0.00	-12.3	19.60
8	32	0.25	0.00	-11.1	14.40
10	30	0.33	0.00	-9.9	10.00
12	28	0.43	0.00	-8.7	6.40
14	26	0.54	0.00	-7.5	3.60
16	24	0.67	0.00	-6.3	1.60
18	22	0.82	0.00	-5.1	0.40
20	20	1.00	0.00	-3.9	0.00
22	18	1.22	0.00	-2.7	0.40
24	16	1.50	0.03	-1.5	1.60
26	14	1.86	0.54	-0.3	3.60
28	12	2.33	8.71	0.9	6.40
30	10	3.00	139.38	2.1	10.00
32	8	4.00	2230.07	**3.3**	14.40
34	6	5.67	35681.19	**4.6**	19.60
36	4	9.00	570899.08	**5.8**	25.60
38	2	19.00	9134385.23	**7.0**	32.40
40	0	—	146150163.73	**8.2**	40.00

Marker	Concordant	non-Concordant	odds ratio (0.8)	LOD
A	16	24	0.00	-6.3
B	16	24	0.00	-6.3
C	20	20	0.00	-3.9
D	14	26	0.00	-7.5
E	20	20	0.00	-3.9
F	22	18	0.00	-2.7
G	24	16	0.03	-1.5
H	26	14	0.54	-0.3
I	28	12	8.71	0.9
J	30	10	139.38	2.1
K	32	8	2230.07	3.3
L	34	6	35681.19	4.6
M	36	4	570899.08	5.8
N	38	2	9134385.23	**7.0**
O	34	6	35681.19	4.6
P	30	10	139.38	2.1
Q	26	14	0.54	-0.3
R	22	18	0.00	-2.7
S	20	20	0.00	-3.9
T	16	24	0.00	-6.3
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 $p = z_1 = ½$)

$$2\sum_{J=1}^{N} \ln[L_J(z_1)] =$$

$$2\left(\sum^{G} \ln[(1 - z_1)/(1/2)] + \sum^{D} \ln[(z_1)/(1/2)] + \sum_{J=1}^{S} \ln\{f_J(z_1 + f_J - f_J z_1)/[f_J(1 + f_J)/2]\}\right) =$$

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

LOD score assignments over the X:

high scores (linkage) in **RS region** of **Xq28**