

# Integration of gene maps: mouse chromosome X

L.H. Wang, A. Collins, S. Lawrence and N.E. Morton

## ABSTRACT

A linkage map was constructed from data in GBASE and integrated with physical and other data. The summary map contains 178 loci in 86 cM (164 Mb). Interference is at the level predicted by Carter and Falconer (*J. Genet.* 50: 307-323, 1951). Problems encountered in this exercise include data errors and omissions which an integrable location database could correct if it received some of the effort now devoted to subjective maps. The extent to which homologous sequences have been rearranged in mouse and man cannot be determined until steps are taken to make both maps more reliable. There has been remarkably little attention to problems of map integration that are common to all eucaryotes, and the need for a location database is as acute in the mouse as in man.

## INTRODUCTION

In well studied organisms each locus has an attribute called location, typically a vector of genetic and physical assignments. For example, location in *Drosophila melanogaster* subsumes a genetic distance from a telomere (in cM) and assignment to one or more contiguous salivary bands (Lindsley and Zimm, 1992). A modest beginning has been made in defining a physical location not in base pairs but through a map index that considers tentative order but not distance (Ashburner, 1992), although a physical map that does not estimate location in base pairs is unacceptable.

Locations in mammals are evolving in parallel and the numbers of loci and maps are now so great as to demand integration by objective algorithms operating on a location database to produce a summary map (Morton *et al.*, 1992). A location database includes partial maps that contain a subset of locations for a subset of loci, and also data summarized by lods or locus content on which multiple pairwise mapping

programs can operate (Morton, 1991). The necessity of these developments has been stressed by the NIH/CEPH Collaborative Mapping Group (1992), the Workshop on Statistical Methods in Genetic Mapping (Terwilliger and Ott, 1994) and the first session of the 12th Gene Mapping Workshop.

A location database does not yet exist for most mammalian chromosomes, but linkage data are summarized by LODSOURCE for man (Keats *et al.*, 1991) and by GBASE (1989) for the mouse. We have used the former (Morton and Collins, 1990; Collins *et al.*, 1992; Lawrence *et al.*, 1993), most recently for the human X chromosome (Wang *et al.*, 1994). Here we apply the January, 1994 version of GBASE to construct a map of homologous loci, for which we give the human symbols with mouse synonyms, and for other loci in the mouse with public data. The X chromosome provides a test case to determine whether the need for a location database is as acute in the mouse as in man.

## MATERIAL AND METHODS

Since GBASE has not previously been used to construct maps, we encountered a few difficulties. The format is not friendly to analysis. Double backcrosses

with R recombinants among N progeny were scored as  $\hat{\theta} = R/N$  with  $\text{lod } \hat{Z} = R \log(2R/N) + (N-R) \log[2(N-R)/N]$ . In GBASE some matings were coded as intercrosses (which is of course impossible for loci on the X chromosome). Crossovers in meiotic panels were sometimes not related to the total number of backcross animals. Mapping panels can be used to estimate distance only if recombinants in a defined interval are sampled randomly. Recombinant inbred (RI) lines were evaluated under autosomal inheritance, and these estimates had to be corrected. The correct formula for X-linkage was given by Haldane and Waddington (1931) as  $y = 8\theta/3(1+4\theta)$ , where  $y = R/N$  is the proportion of RI lines that are recombinant and  $\theta$  is the recombination fraction. This inverts to  $\hat{\theta} = 3R/(8N-12R)$  with  $\text{lod } \hat{Z} = R \log(9R/4N) + (N-R) \log[9(N-R)/5N]$ .

A more serious problem is that GBASE is incomplete, especially for strain and species backcrosses. We are grateful to Dilip Tailor for helping to capture data from the Mbx database of the European Backcross Project, but other data are not public (i.e., neither published nor available on Internet). These limitations deserve attention by mouse geneticists, since a summary map can be no better than the location database from which it is derived. Lods were analysed by the MAP program (Morton and Collins, 1990) and integrated with other data by the *ldb* program (Morton *et al.*, 1992). Physical data are extremely sparse, and GBASE does not accommodate them. We therefore used cytogenetic assignments (Evans, 1989) to project the genetic map onto the physical map, (Collins *et al.*, 1992), assuming the same length (164 Mb) as in man (Morton, 1991). Regions were defined as the major conserved sequences (A-H) and the subtelomeric segment (I), within which there may be rearrangements (Brown *et al.*, 1993).

## LINKAGE ANALYSIS

There were 131 loci in the linkage map, one of which (HPRT) contributed nearly 1/3 of the heterogeneity  $\chi^2$ , calculated as  $2 \ln 10(Z_O - Z_E)$ , where  $Z_O$  is the sum of maximum lods and  $Z_E$  is the sum predicted by minimum deviance. The number of degrees of freedom is taken as the number of lods minus the number of informative loci (Rao *et al.*, 1978). Investigation of HPRT led to several anomalous entries in GBASE (Barnard *et al.*, 1989; Borsani *et al.*, 1991; Derry and Barnard, 1992). All contained unpublished map distances on a meiotic panel of recombinants in which choice of the panel was not allowed for (an error with which human geneticists will soon be familiar since there is no theory for estimating map distance from an arbitrary panel). When these references were deleted, the number of degrees of freedom was barely reduced.  $\chi^2$  was halved and became acceptable.

Loci were entered in the consensus order (Brown *et al.*, 1993), constrained by physical data, and submitted to a bootstrap that increases likelihood by successive rearrangements of unconstrained loci (Morton and Andrews, 1989). We did not constrain by the consensus order, which is based on unspecified data that are not public, integrated in an unspecified way. Changes in location between successive editions are too great to take the consensus map as necessarily correct. The level of interference corresponds closely to the level suggested by Carter and Falconer (1951) for the mouse, with a very low but significant frequency of residual typing errors (Table I). The Haldane model of no interference is excluded by  $\chi^2_1 = 223.52$ . Equally strong evidence was obtained for *Drosophila* by Haldane (1919) and for man by Hulten (1974), Rao *et al.*

**Table I** - Tests of hypotheses on the mouse X (parameters fixed by hypothesis in parentheses).

Interference hypothesis	Mapping parameter p	Error frequency $\epsilon$	Genetic length cM	$\chi^2$
Haldane	(1)	0	120.5	966.29
Kosambi	(.5)	0	95.4	786.42
	(.5)	.0005	94.9	783.77
Carter-Falconer	(.25)	(0)	86.5	742.96
	(.25)	.0005	86.1	734.10
Muller	(0)	0	79.5	825.10
	(0)	.0011	78.8	806.01
Empirical	.262	0	86.8	742.77
	.247	.0005	86.0	733.98

Number of pairs = 868, number of loci = 131, df = 738. All hypotheses rejected by  $\chi^2$  except Carter-Falconer with error

(1977), Morton and Collins (1990), Collins *et al.* (1992), Lawrence *et al.* (1993) and Wang *et al.* (1994). The empirical mapping parameter gives a map length of 86 cM, much less than the human X (211 cM) although the physical maps are similar (164 Mb). This is consistent with the sex-averaged autosomal map which is about 3800 cM in the human (Morton, 1991) and 1600 cM in the mouse (Davisson and Roderick, 1990; Dietrich *et al.*, 1994).

### Map integration

Physical data are sparse, and projection of the genetic map on cytogenetic assignment is required for connectivity (Collins *et al.*, 1992). From this projection 12 widely spaced loci with physical data were entered as a partial map, giving the results in Table II. Table III presents the human loci in the same order as their mouse homologs.

**Table II** - Mouse chromosome X.

Locus	Mouse synonym	Composite Mb	rg*	Cytogenetic band left	Cytogenetic band right	Genetic map cM	Physical Mb	Reference in mouse	Composite Mb (human)
pter		0.00		pter	pter	0.0	0.00	19,99,98	0.00
cen		13.95	A	cen	cen	0.0	13.95	22,98,99	62.00
DXF34	DXHXF34	13.95	A	A1	A2	0.0	13.95	3,99	X
GATA1	Gata1	14.27	B	A1	A2	1.0	14.27	98,99,97	52.39
HcX		17.73					2.9	22,97	
DXWas70		19.37	A	A1	A2	3.9		97,22,3	
CYBB	Cybb	22.10	C	A1	A2	5.5		97,22,98	35.43
WAS	sf	23.19				6.1		97,98,22	54.54
IP1	Td	23.19				6.1		98,97,22	59.20
OTC	Otc(spf)	23.92	C	A1	A2	6.5		97,22,98	36.28
DXS32	DXHXS32	26.10	C	A1	A2	7.7		97,22,3	X
DXS676	DXHXS676	27.19	C	A1	A2	8.4		3,99	X
MAOB	Maob	31.02	C	A2	A2	10.5		97,3,22	44.76
MAOA	Maoa	31.02	C	A2	A2	10.5		97,3,22	44.36
UBE1	Ube1x	32.29	C	A2	A2	11.2		97,22,98	46.80
jk105		32.47				11.3		97	
Xlr1		32.66	C/D	A3	A3	11.4		22,97,3	
PFC	Pfc	33.02	C	A2	A3	11.7	33.02	98,1,22	49.58
DXPas22		33.02	C	A2	A3	11.7		22,97,3	
TIMP1	Timp-1	33.06	C	A2	A3	11.7	33.06	1,97,22	48.65
SYN1	Syn-1	33.06	C	A2	A3	11.8	33.06	1,98,97	48.18
ARAF1	Araf	33.08	C	A2	A3	11.8	33.08	1,97,22	47.74
DXSmh15		33.08	C/D	A3	A3	11.8		22,97,3	
DXSmh141		33.08	C/D	A3	A3	11.8		22,97,3	
DXSmh222		34.16	D	A5	A6	12.2		22,97,3	
DXPas3		34.43	C/D	A3	A3	12.3		22,97,3	
DXWas68		38.50	C/D	A3	A3	13.8		22,97,3	
DXPas4		38.77	C/D	A4	A4	13.9		97,3,99	
DXSmh191		43.37				15.7		22,97	
LAMP2	Lamp2	44.73	D	A5	A6	16.2		3,19,99	131.31
DXSmh172		47.17	D	A5	A6	17.1		97,3,99	
DXSmh67		47.44				17.2		97	
DXSmh66		47.44				17.2		22, 97	
DXSmh10		47.71	D	A5	A6	17.3		22,97,3	
DXNds1		49.33	D	A6	A6	17.9		3,99	

Continue

Table II - Continued.

Locus	Mouse synonym	Composite Mb	rg*	Cytogenetic band left	Cytogenetic band right	Genetic map cM	Physical Mb	Reference in mouse	Composite Mb (human)
DXSmh219		49.33	D	A6	A6	17.9		22,97,3	
DXPas5		50.96	D	A5	A6	18.6	50.96	99,22,97	
DXSmh36		51.91				18.9		22,97	
DXSmh59		54.46				19.7		22,97	
HPRT	Hprt	57.00	D	A6	A6	20.5		97,3,22	140.91
Otf3h		57.32	D	A5	A6	20.6		97,3,99	
Hq		57.32	D	A6	A6	20.6		22,97,3	
Xlr2		70.35	D	A6	A6	24.9		22,97,3	
DXS144	DXHXS144	77.98	D	A6	A6	27.3		97,22,98	142.45
DXSmh91		79.89				27.9		22,97	
Tyl		81.80				28.6	22,97		
jk095		82.43				28.8		97	
F9	Cf-9	82.75	D	A6	A6	28.9	82.75	21,17,97	146.81
IDS	Ids	84.02	D	A6	A7	32.9	84.02	17,97,22	155.51
DXS296	DXHS296	84.24	D	A6	A7	32.5		97,22,98	154.42
DXSmh255		84.55	D	A6	A6	31.8		97,3,99	
Bn		84.78				33.1		22,97	
GABRA3	Gabra-3	84.78	D	A7	B	33.1	84.78	13,97,22	158.84
CDR1	Cdr	84.94	D	A6	A7	30.8		97,3,22	147.92
DXS1104		85.08					85.08	13	
DXMit1		85.19	D	A7	B	33.6		3,99	
FMR1	Fmr-1	85.20	D	A6	A7	30.2		97,22,98	153.88
MCF2	Mcf-2	85.60	D	A6	A7	29.3		97,22,21	146.84
F8A	Cf8a	85.64	D	A7	B	29.2	85.64	97,22,13	X
DXS52	DXPas8	85.76	D	A7	B	34.3	85.76	13,3,97	159.99
BGN	Bgn	85.77	D	A7	B	37.3	85.77	17,13,3	160.40
DXBay2		86.14	D	A7	B	35.2	86.14	3,99,13	
DXBay6		86.14	D	A7	B	35.2		3,99	
AVPR2	Avp-2r	86.14	D	A7	B	35.2		3,99	161.11
CDPX2	Bpa	86.17	D	A7	B	35.3		97,22,3	X
L1CAM	L1cam	86.32	D	A7	B	35.9	86.32	98,13,97	160.94
RCP	Rsvp	86.55	D	A7	B	37.3	86.55	18,97,22	161.51
DXS253	DXHXS253	86.80	D	B	B	36.8	86.80	13,97,22	X
G6PD	G6pd	86.87	D	B	B	36.8	86.87	13,97,22	162.68
DXS254	DXHXS254	86.88	D	B	B	37.6	86.88	18,97,22	162.26
DXSel1		86.99	D	B	B	37.1		3,99	
ptd		87.44				39.3		97	
DXSmh64		87.58	E	C	C	39.7		22,97,3	
DXSmh23		87.58	E	C	C	39.7		22,97,3	
XIC	Xce	87.65				39.9		20,98,19	81.51
F8C	Cf-8	87.67	D	B	B	36.8	87.67	13,97,22	162.98
Oat-rs1		87.67				36.8		97	
DMD	Dmd(mdx)	87.69	E	C	C	40.0	87.69	97,22,17	30.02
DXS120	SXHXS120	88.46				40.3		97,99,98	
IP2	Str	89.48	D	A7	B	40.7		3,97,22	162.68

Continue

Table II - Continued.

Locus	Mouse synonym	Composite Mb	rg*	Cytogenetic band left	Cytogenetic band right	Genetic map cM	Physical Mb	Reference in mouse	Composite Mb (human)
Bmp2b-ps1		91.78	E	C	C	41.7		97,3,99	
ZFX	Zfx	93.06	E	D	D	42.2		97,22,98	25.01
DXSmh120		93.06	E	C	C	42.2		97,3,99	
93/YB/002		93.32				42.3		97	
POLA	Pola	93.44	E	D	D	42.8	93.44	99,14,3	25.02
DXCrc140		93.68	E	D	D	42.5	93.68	97,14,22	
ASB		93.70				42.6		97,99,98	
DXCrc190		93.83	E	D	D	42.5		97,3,99	
DXCrc28		93.83	E	D	D	42.5	93.83	14,97,3	
pDP1068		94.13					94.13	14	
BSXB2		94.13					94.13	14,13,20	
PT26		94.18					94.18	99,14	
Bhd		96.23				44.4		22,97	
AR	Tfm(Ar)	96.75	F	D	D	44.8		97,3,22	68.43
DXCrc131		96.75	F	D	D	44.8		97,3,99	
jk090		98.32				45.7		97	
DXCrc169		99.19	F	D	D	46.2		22,97,3	
Gs		99.89				46.6		22,97	
DXCrc94		101.28				47.4		22,97	
EDA	Ta	101.63	F	D	D	47.6		97,22,98	72.04
GJB1	Gjb-1	101.74	F	D	D	48.4	101.74	10,3,97	73.89
AGMX1		101.87				48.7		97,22,3	
Ym		101.92				48.8		97	
CCG1	Ccg-1	101.98	F	D	D	47.8	101.98	10,97,22	74.87
Mintd		102.00	F	D	D		48.9	22,97,3	
DXCrc98		102.09	F	D	D	49.2	102.09	10,97,22	
RPS4	Rps4	102.42	F	D	D	50.2	102.42	10,11,97	78.22
PHKA1	Phka	102.84	F	D	D	50.6	102.84	10,11,98	79.40
DXCrc177		103.02	F	D	D	50.7	103.02	10,97,22	
DXS393	DXHXS393	103.52	F	D	D	49.8		3,99	X
DXPas31		103.52	F	D	D	49.8	103.52	3,99,20	
DXCrc318		103.64	F	D	D	50.7	103.64	10,97,22	
XIST	Xist	103.84	F	D	D	50.6	103.84	10,98,11	81.16
DXCrc187		103.85	F	D	D	50.8		22,97,3	
DXPas29		103.91	F	D	D	50.4	103.91	3,99,20	
DXPas19		104.02	F	D	D	50.4	104.02	20,16,3	
Slf		104.07				50.9		97	
DXSmh44		104.07	F	D	D	50.9		97,3,99	
MNK	Mo	104.07	C		F4	50.9		97,22,98	85.22
DXPas2		104.07	F	D	E	50.9		22,97,3	
DXPas28		104.11	F	D	D	50.4	104.11	3,99,20	
DXCrc13		104.28	F	D	D	51.0		22,97,3	
DXCrc47		104.28	F	D	D	51.0		97,3,99	
DXPas27		104.32	F	D	D	50.4	104.32	16,20,3	
Odc-ps13		104.70	F	D	E	51.2		22,97,3	

Continue

Table II - Continued.

Locus	Mouse synonym	Composite Mb	rg*	Cytogenetic band left	Cytogenetic band right	Genetic map cM	Physical Mb	Reference in mouse	Composite Mb (human)
PGK1	Pgk-1	104.92	F	D	D	51.4		98,97,22	85.83
DXPas23		106.20	F	D	D	51.9		3,99	
DXCrc323		106.41	F	D	D	52.1		97,3,99	
DXCrc112		106.41	F	D	D	52.1		97,3,99	
DXPas24		108.96	F	D	E	53.3		3,99	
DXCrc180		110.45	F	F1	F1	54.0		3,99	
DXPas25		110.45	F	F1	F1	54.0		3,99	
DXCrc23		110.45	F	F1	F1	54.0		3,99	
GLA	Ags	115.35	F	F1	F1	56.4		3,97,22	104.56
DXMit3		116.20	F	F1	F1	56.8		3,99	
DXMit4		121.53	F	F1	F1	59.4		3,99	
DXPas10		121.53	F	F1	F1	59.4		3,99	
DXWas17		127.49	F	F1	F1	62.3		22,97,3	
PLP	Plp(jp)	129.83	F	F1	F1	63.4	129.83	99,3,97	106.94
Scar		130.50	F	F1	F1	63.8		3,99	
DXS178	DXHXS178	131.51	F	F1	F1	64.4		3,97,22	104.68
DXS101	DXHXS101	132.86	F	F1	F1	65.3		97,22,98	105.20
DXCrc48		136.23	F	F1	F1	67.3		3,99	
DXSmh43		137.24	F	F1	F1	67.9		97,3,99	
NHS		138.09				68.5		98,22,97	
COL4A5	Col4A5	141.96	F	F1	F1	70.8		3,19,99	117.60
DXS674	DXHXS674	144.83	G	F1	F1	72.6		3,19,99	56.78
DXS679	DXHXS679	144.83	G	F1	F1	72.6		3,19,99	56.78
ALAS2	Alas-2	144.83	G	F1	F1	72.6		3,19,99	56.78
Syx2		146.85	G/H	F1	F1	73.8		3,99	
Xp1		146.85				73.8		22,97	
DXSmh225		147.02				73.9		97	
PDHA1	Pdha-1	147.18	H	F3	F4	74.0		3,19,99	23.50
HYP	Hyp	147.86	H	F3	F3	74.5		3,97,22	23.50
Hst3		149.54				75.5		22,97	
Li		149.54				75.5		22,97	
DXPas11		149.54				75.5		97	
DXSel501		149.54	H	F3	F3	75.5		3,99	
DXWas31		150.55	H	F3	F3	76.1		97,3,99	
DXPas18		150.55	H	F3	F3	76.1		22,97,3	
Cbx-rs1		150.72	H	F3	F3	76.2		22,97,3	
Gja6		150.89	H	F3	F3	76.3		22,97,3	
DXPas1		150.89	H	F3	F3	76.3		22,97,3	
Crm		152.41				77.2		22,97	
DXMit12		152.41	H	F3	F3	77.2		3,99	
PHKA2	Phka2	152.41	H	F3	F3	77.2		3,19,99	19.90
PRPS2	Prps2	153.42	H	F3	F3	77.9		3,19,99	15.58
DXCrc157		153.42	H	F3	F3	77.9		3,99	
GLRA2	Glra2	155.10	H	F3	F3	78.9		97,22,3	19.98
GRPR	Grpr	155.78	H	F3	F3	79.3		97,22,98	19.90

Continue

Table II - Continued.

Locus	Mouse synonym	Composite Mb	rg*	Cytogenetic band left	Cytogenetic band right	Genetic map cM	Physical Mb	Reference in mouse	Composite Mb (human)
Paf		156.45				79.7		22,97	
DXCrc181		157.29	H	F3	F3	80.2		3,99	
AMELX	Amel	157.80	H	F3	F3	80.5	157.80	97,22,99	11.87
DXYMov15		159.50	I	F3	F3	81.8		3,99	
Syx3		159.50	I	F3	F3	81.8		3,99	
Sxa		163.88	I	F3	F3	84.9	163.88	97,3,99	
Telxy		163.93	I	F3	F3	85.4		97,22,3	
Telo2		163.99	I	F3	F3	85.9		3,99	
STS	Sts	163.99				85.9		22,97	6.55
YA2L		164.00				86.0		22,97	
qter		164.00	I	qter	qter	86.0	164.00	22,99,98	164.00

\*Region of synteny in mouse.

**References to Table II.**

01. Derry, J.M.J. and Barnard, P.J. (1992). Physical linkage of the A-raf-1, properdin, synapsin 1, and TIMP genes on the human and mouse X chromosomes. *Genomics* 12: 632-638.
03. Brown, S.D.M., Avner, P., Boyd, Y., Chapman, V., Rastan, S., Sefton, L., Thomas, J.D. and Herman, G.E. (1993). *Mammalian Genome* 4: S269-S281.
10. Cooper, P., Keer, J.T., McCabe, V.M., Hamvas, R.M.J., Brown, S.D.M., Rastan, S. and Brockdorff, N. (1993). Physical mapping of 200 kb of the mouse X chromosome in the vicinity of the Xist locus. *Genomics* 15: 570-575.
11. Hamvas, R.M.J., Zinn, A., Keer, J.T., Fischer, E.M.C., Beer-Romero, P., Brown, S.D.M. and Page, D.C. (1992). *Genomics* 12: 363-367.
13. Brown, S.D.M., Avner, P., Boyd, Y., Chapman, V., Rastan, S., Sefton, L., Thomas, J.D. and Herman, G.E. (1993). *Mammalian Genome* 4: S281(A).
14. Brown, S.D.M., Avner, P., Boyd, Y., Chapman, V., Rastan, S., Sefton, L., Thomas, J.D. and Herman, G.E. (1993). Mouse X chromosome. *Mammalian Genome* 4, S281.  
  
B-Hamvas, R.M.J., Larin, Z., Brockdorff, N., Rastan, S., Lehrach, H., Chartier, F.L., Brown, S.D.M. (1993). YAC clone contigs surrounding the ZFX and POLA loci in the mouse X chromosome. *Genomics* 17: 52-58.
16. Brown, S.D.M., Avner, P., Boyd, Y., Chapman, V., Rastan, S., Sefton, L., Thomas, J.D. and Herman, G.E. (1993). Mouse X chromosome. *Mammalian Genome* 4: S281.  
  
D-Heard, E., Simmler, M.C., Larin, Z., Rougeulle, C., Courtier, B., Lehrach, H. and Avner, P. (1993). Physical mapping and YAC contig analysis of the region surrounding XIST on the mouse X-chromosome. *Genomics* 15: 559-569.
17. Chatterjee, A., Faust, C.J. and Herman, G.E. (1993). Genetic and physical mapping of the biglycan gene on the mouse X chromosome. *Mammalian Genome* 4: 33-26.
18. Faust, C.J., Levinson, B., Gitschier, J. and Herman, G.E. (1992). Extension of physical map in the region of the mouse X chromosome homologous to human Xq28 and identification of an exception of conserved linkage. *Genomics* 13: 1289-1295.
19. Wang, L.H., Collins, A., Lawrence, S., Keats, B.J. and Morton, N.E. (1994). Integration of gene maps: chromosome X. *Genomics* 22 (in press).
20. Simmler, M.C., Cattanach, B.M., Rasberry, C., Rougeulle, C. and Avner, P. (1993). Mapping the murine Xce locus with (CA)<sub>n</sub> repeats. *Mammalian Genome* 4: 523-530.
21. Faust, C.J., Verkerk, A.J.M.H., Wilson, P.J., Morris, C.P., Hopwood, J.J., Oostra, B.A. and Herman, G.E. (1993). Genetic mapping on the mouse X chromosome of human cDNA clones for the Fragile X and Hunter syndromes. *Genomics* 12: 814-817.
- 22, 97, 98. Mouse genetic map (this study).
99. Projected genetic map (this study).

**Table III** - Loci with homologies in man (Wang et al., 1994).

Locus	Rank*	Composite Mb (human)	rg** (human)	Cytogenetic band (human)		Genetic map (human) cM	Composite Mb (mouse)
				left	right		
pter	1	0.00		pter	pter	0.0	00.00
cen	0	62.00		cen	cen	100.0	13.95
GATA1	1	52.39	46	p11.23	p11.23		14.27
CYBB	1	35.43	42	p21.1	p21.1	75.0	22.10
IP1	0	59.20	5M	p11.21	cen		23.19
WAS	0	54.54	4Q	p11.3	p11.22	96.8	23.19
OTC	1	36.28	43A	p21.1	p21.1	76.2	23.92
MAOA	2	44.36	44N	p11.4	p11.3	83.4	31.02
MAOB	1	44.76	44N	p11.4	p11.3		31.02
UBE1	1	46.80	45	p11.23	p11.23		32.29
PFC	2	49.58	45	p11.3	p11.23	95.3	33.02
SYN1	1	48.18	45	p11.23	p11.23	91.3	33.06
TIMP1	1	48.65	45	p11.3	p11.23	94.8	33.06
ARAF1	2	47.74	45	p11.3	p11.23		33.08
LAMP2	1	131.31	79	q24	q24		44.73
HPRT	2	140.91	80	q26.1	q26.1	160.4	57.00
DXS144	1	142.45	80	q26.2	q26.2	161.8	77.98
F9	1	146.81	81	q26.3	q27.1	166.7	82.75
IDS	1	155.51	86	q27.3	q28	192.4	84.02
DXS296	1	154.42	85	q27.3	q28	193.5	84.24
GABRA3	2	158.84	86	q28	q28		84.78
CDR1	1	147.92	81	q27.1	q27.2		84.94
FMR1	1	153.88	84	q27.3	q27.3	191.9	85.20
DXS52	2	159.99	87	q28	q28	205.3	85.76
BGN	1	160.40	87	q28	q28		85.77
AVPR2	1	161.11	88	q28	q28	205.6	86.14
L1CAM	1	160.94	88	q28	q28	205.6	86.32
RCP	1	161.51	88	q28	q28	205.6	86.55
G6PD	1	162.68	89	q28	q28	205.6	86.87
DXS254	1	162.26	88	q28	q28		86.88
XIC	0	81.51	65	q13.2	q13.2		87.65
F8C	2	162.98	89	q28	q28	206.6	87.67
DMD	2	30.02	3T	p21.3	p21.2	63.6	87.69
IP2	0	162.68	8Q	q27	q28	205.6	89.48
ZFX	1	25.01	30A	p22.1	p22.1	49.9	93.06
POLA	1	25.02	30N	p22.1	p21.3		93.44
AR	2	68.43	56	q11.2	q12	101.6	96.75
EDA	0	72.04	59A	q12	q13.1	103.9	101.63
GJB1	1	73.89	60	q13.1	q13.1		101.74
CCG1	1	74.87	61	q13.1	q13.1	108.3	101.98
RPS4	1	78.22	63	q13.1	q13.1		102.42
PHKA1	1	79.40	63	q13.1	q13.1		102.84
XIST	1	81.16	65	q13.2	q13.2		103.84
MNK	1	85.22	67	q13.2	q13.3	108.3	104.07
PGK1	1	85.83	68	q13.3	q13.3	107.6	104.92

Continue

Table III - Continued

Locus	Rank*	Composite Mb (human)	rg** (human)	Cytogenetic band (human)		Genetic map (human) cM	Composite Mb (mouse)
				left	right		
GLA1	1	104.56	78A	q21.3	q22	125.7	115.35
PLP	1	106.94	78A	q21.33	q22	127.7	129.83
DXS178	1	104.68	78A	q21.33	q22	126.4	131.51
DXS101	0	105.20		q22	q22	127.7	132.86
COL4A5	1	117.60	78N	q22	q22	127.7	141.96
ALAS2	1	56.78	51	p11.21	p11.21		144.83
DXS679	0	56.78	51	p11.22	p11.21		144.83
DXS674	0	56.78	51	p11.22	p11.21		144.83
MCF2	1	146.84	81	q26.3	q27.1		146.56
PDHA1	0	23.50	29	p22.1	p22.1		147.18
HYP	0	23.50	29	p22.2	p22.1	45.2	147.86
PHKA2	r	19.90	2Q	p22.2	p22.1	38.2	152.41
PRPS2	0	15.58	26	p22.3	p22.2		153.42
GLRA2	1	19.98	28	p22.1	p21.3	38.2	155.10
GRPR	r	19.90	2Q	p11.22	cen		155.78
AMELX	1	11.87	23	p22.31	p22.1	29.6	157.80
STS	1	6.55		p22.32	p22.32	18.6	163.99
qter	1	164.00		qter	qter	210.8	164.00

\*rank in man (2 = reference locus, 1 = other framework locus, 0 = locally unordered, r = cytogenetically assigned).

\*\*region (human).

Several discrepancies suggest rearrangements within the major syntenic blocks, but these may in part be due to errors in order in either organism. Block A is defined on DXF34, which is assigned cytogenetically to the pericentromeric region in man but not better localized. Block B is defined on GATA1 which is in region 46 of man. Boyd *et al.* (1994) have shown that TFE3 in region 47 is within 500 kb of GATA1. Although public linkage data support a location proximal to DXWAs70, this interrupts block A and may be in error. Block C covers at least regions 42-45 in man. Within it IP1 is insecurely placed. Block D covers at least regions 79-89. The mouse locus Xce for the X inactivation centre is conventionally taken as the homolog of XIC in Man (O'Brien *et al.*, 1994). If so, the location of Xce appears to be grossly in error, presumably because of phenotype misclassification. In man it maps close to XIST in region 65. F8A, a transcribed sequence in human F8 intron 22, maps 2 Mb proximally in the mouse suggesting a transposition in one or the other line of descent from a common ancestor. Block E covers region 30 in man. Inclusion of IP2 in the mouse as a homolog of Str may be in error, since the assignment distal to region 80 in man is supported by linkage and deletions. The discrepancy is not great, since the estimated

distance to the border of block D is only 2 cM. In block F there is a gap between PGK1 in human region 68 and GLA in region 78A, corresponding to band q21 in man. MCF2 is assigned to the distal long arm in man but maps between blocks G and H in the mouse. No homologous loci have been definitely assigned to the subtelomeric block I, corresponding to regions 1-25 in man, although STS may belong to it. Public linkage data on STS in the mouse are sparse and the location distal to DXYMov15 is insecure. It is not clear whether the latter belongs to the major pseudoautosomal region adjacent to STS in man, or to the minor pseudoautosomal region that in man is adjacent to qter. Clearly much remains to be learned about chromosomal location in mouse and man, and to this end public location databases would be invaluable.

## DISCUSSION

There are catalogues of genetic variants for *Drosophila*, mouse and man, giving for each expressed locus its genetic and physical location, phenotypes of alleles, and references (Lyon and Searle, 1990; Lindsley

and Zimm, 1992; McKusick, 1992). None of these catalogues yet attempts to cover loci that are not expressed, to give gene frequencies or heterozygosity, to provide the data on which locations are based, to specify location with any precision, or to list vectors or sequences. For this information one must go to one or more species-specific databases, the content and logic of which are *in status nascendi*. The researcher often wants to ask a single question as he would of a catalogue: "tell me everything you know about this locus", but instead he must search for countless iotas of information that are often too incomplete, out of date, or erroneous to be useful. Surprisingly little attention has been devoted to this problem, especially by the biologists who are most affected but defer to informaticians with less interest in the genome. In desperation biologists turn to single chromosome workshops that meet regularly to vote on consensus maps of a subset of loci. The position of each locus in a consensus map is supported by at least one member of the workshop, but the evidence (if any) on which this location is based is not accessible to other scientists. Since workshops are not supported by a location database, the evidence must be reiterated in sessions that could more profitably be spent updating a public and integrable database. This paradigm is a serious weakness in the Genome Project. The tenacity of single chromosome committees is admirable, but there is a more reliable and cost-effective way to construct maps.

A successor to GBASE would be an invaluable source of information for the mouse if the linkage data were complete, if physical data were included with some precision, and if algorithms were adopted to integrate genetic and physical maps. Better databases would stimulate better algorithms and the use that justifies existence of a database. Without these developments any map must be flawed, and the value of consensus mouse maps to guide human genetics will be diminished by the obscurity of the data and methods on which they are based.

We have no doubt that a single chromosome committee working by suitable algorithms with a comprehensive database could create a better map than ours, but as the volume of data increases exponentially it is inconceivable that consensus maps without a location database can have acceptable accuracy. This paper will have served its purpose if it provokes serious consideration of the means, objectives and control of error in mapping the genomes of higher organisms and especially the authentication of summary maps by a location database that is public, reliable, complete and amenable to alternative integration algorithms.

## RESUMO

Um mapa da ligação foi construído de dados em GBASE e integrado com dados físicos e outros dados. O mapa resumido contém 178 loci em 86 cM (164 Mb). Interferência está no nível predito por Carter e Falconer (*J. Genet.* 50: 307-323, 1951). Problemas encontrados neste exercício incluem erros de dados e omissões que uma base de dados de locações integráveis poderiam corrigir se recebessem algum esforço direcionado a mapas subjetíveis. O grau pelo qual seqüências homólogas têm sido organizados no camundongo e homem não podem ser determinados até passos serem tomados para fazer ambos mapas mais confiáveis. Marcadamente pouca atenção tem sido direcionada para problemas de integração de mapas que são comuns para todos os eucariontos e a necessidade de uma base de dados de locação é tão importante no camundongo como é no homem.

## REFERENCES

- Ashburner, M.** (1992). Cytogenetic map. In: *The Genome of Drosophila melanogaster*. Lindsley DL, Zimm GG. Academic Press, San Diego, pp. 1117-1133.
- Barnard, P.J., Derry, J.M.J., Ryder-Cook, A.S. and Barnard, E.A.** (1989). Localization of the GABA-A receptor alpha-3 subunit gene on the mouse X chromosome. *Cytogenet. Cell Genet.* 51: 958 (Abstract).
- Borsani, G., Toniorenzi, R., Simmler, M.C., Dandolo, L., Amand, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C., Wilard, H.F., Awner, P. and Ballabio, A.** (1991). Characterization of a murine gene expressed from the inactive X chromosome. *Nature* 351: 325-328.
- Boyd, Y., Blair, H., Uwechue, I.C., Read, V., Maslen, G.L. and Laval, S.H.** Detailed comparative mapping of the conserved segments which compose Xp11.2 in man and mouse. Abstract, *5th X Chromosome Workshop* (in press).
- Brown, S.D.M., Avner, P., Boyd, Y., Chapman, V., Rastan, S., Sefton, L., Thomas, J.D. and Herman, G.E.** (1993). Mouse X chromosome. *Mammalian Genome* 4: S269-S281.
- Carter, T.C. and Falconer, D.S.** (1951). Stocks for detecting linkage in the mouse and the theory of their design. *J. Genet.* 50: 307-323.
- Collins, A., Keats, B.J., Dracopoli, N., Shields, D.C. and Morton, N.E.** (1992). Integration of gene maps: chromosome 1. *Proc. Natl. Acad. Sci. USA.* 89: 4598-4602.
- Davisson, M.T. and Roderick, T.H.** (1990). Linkage map. In: *Genetic Variants and Strains of the Laboratory Mouse* (Lyon, M.F. and Searle, A.G., eds.). 2nd edn. Oxford Press, Oxford, pp. 416-427.
- Derry, J.M.J. and Barnard, P.J.** (1992). Physical linkage of the A-raf-1, properdin, synapsin I, and TIMP genes on the human and mouse X chromosomes. *Genomics* 12: 632-638.
- Dietrich, W.F., Miller, J.C., Steen, R.G., Merchant, M., Damron, D., Nahf, R., Gross, A., Joyce, D.C., Wessel, M., Dridge, R.D., Marquis, A., Stein, L.D., Goodman, N., Page, D.C. and Lander, E.S.** (1994). A genetic map of the mouse with 4,006 simple sequence length polymorphisms. *Nat. Genet.* 7: 220-339.

- Evans, E.P.** (1989). Standard normal chromosomes. In: *Genetic Variants and Strains of the Laboratory Mouse* (Lyon, M.F. and Searle, A.G., eds.). 2nd edn. Oxford Press, Oxford, pp. 576-581.
- GBASE** (1989). *The Genetic Database of the Mouse*. User Manual, Version 1.1. The Jackson Laboratory, Bar Harbor, Maine.
- Haldane, J.B.S.** (1919). The combination of linkage values, and the calculation of distances between the loci of linked factors. *J. Genet.* 8: 299-309.
- Haldane, J.B.S.** and **Waddington, C.H.** (1931). Inbreeding and linkage. *Genetics* 16: 357-374.
- Hulten, M.** (1974). Chiasma distribution at diakinesis in the normal human male. *Hereditas* 76: 55-78.
- Keats, B.J.B., Sherman, S.L., Morton, N.E., Robson, E.B., Buetow, K.H., Cartwright, P.E., Chakravarti, A., Francke, U., Green, P.P. and Ott, J.** (1991). Guidelines for human linkage maps. (ISLM, 1990). *Ann. Hum. Genet.* 55: 1-6.
- Lawrence, S., Collins, A., Keats, B.J., Hulten, M. and Morton, N.E.** (1993). Integration of gene maps: chromosome 21. *Proc. Natl. Acad. Sci. USA.* 90: 7210-7214.
- Lindsley, D.L. and Zimm, G.G.** (1992). The genome of *Drosophila melanogaster*. Academic Press, San Diego, pp. 1132.
- Lyon, M.F. and Searle, A.G.** (1990). *Genetic Variants and Strains of the Laboratory Mouse*. 2nd edn. Oxford Press, Oxford, pp. 876.
- McKusick, V.A.** (1992). *Mendelian Inheritance in Man*. 10th edn. Johns Hopkins University press, Baltimore, pp. 2320.
- Morton, N.E.** (1991). Gene maps and location databases. *Ann. Hum. Genet.* 55: 235-241.
- Morton, N.E.** (1991). Parameters of the human genome. *Proc. Natl. Acad. Sci. USA* 88: 7474-7476.
- Morton, N.E. and Andrews, V.** (1989). MAP, an expert system for multiple pairwise linkage analysis. *Ann. Hum. Genet.* 53: 263-269.
- Morton, N.E. and Collins, A.** (1990). Standard maps of chromosome 10. *Ann. Hum. Genet.* 54: 235-251.
- Morton, N.E., Collins, A., Lawrence, S. and Shields, D.C.** (1992). Algorithms for a location database. *Ann. Hum. Genet.* 56: 223-232.
- NIH/CEPH Collaborative Mapping Group** (1992). A comprehensive genetic linkage map of the human genome. *Science* 258: 67-86.
- O'Brien, S.J., Peters, J., Searle, A.G., Womack, J.E., Johnson, P.A. and Graves, J.A.M.** (1994). Report of the committee on comparative gene mapping. In: *Human Gene Mapping*, 1993 (Cuticchia, A.J. and Pearson, P.L., eds.). John Hopkins University Press, Baltimore, pp. 846-892.
- Rao, D.C., Morton, N.E., Lindsten, J., Hulten, M. and Yee, S.** (1977). A mapping function for man. *Hum. Hered.* 27: 99-104.
- Rao, D.C., Keats, B.J.B., Morton, N.E., Yee, S. and Lew, R.** (1978). Variability of human linkage data. *Am. J. Hum. Genet.* 30: 516-529.
- Terwilliger, J. and Ott, J.** (1994). Workshop on Statistical Methods in Genetic Mapping. *Genomics* 22: 496-497.
- Wang, L.H., Collins, A., Lawrence, S., Keats, B.J. and Morton, N.E.** (1994). Integration of gene maps: chromosome X. *Genomics* 22: 590-604.

(Received December 19, 1994)