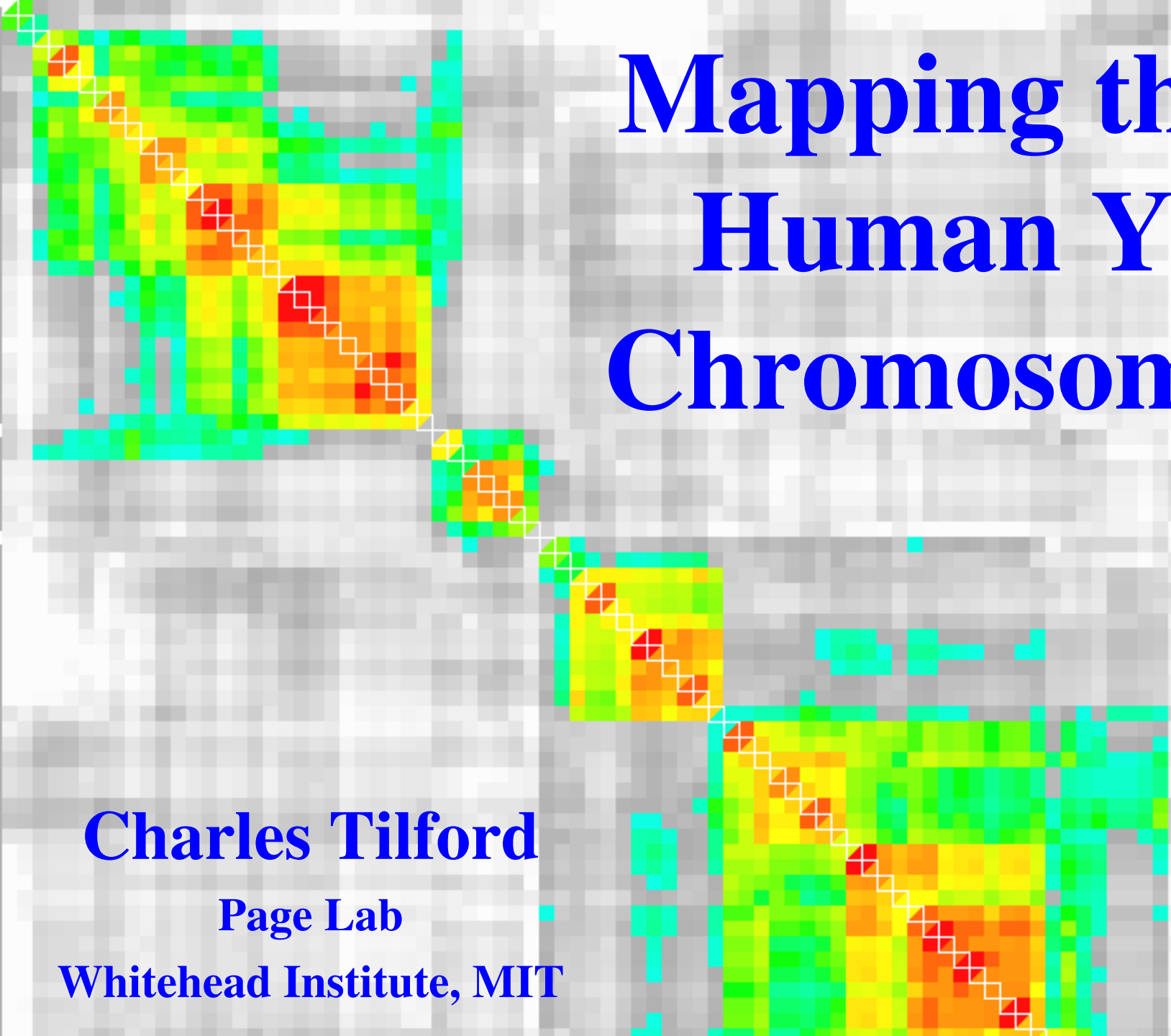


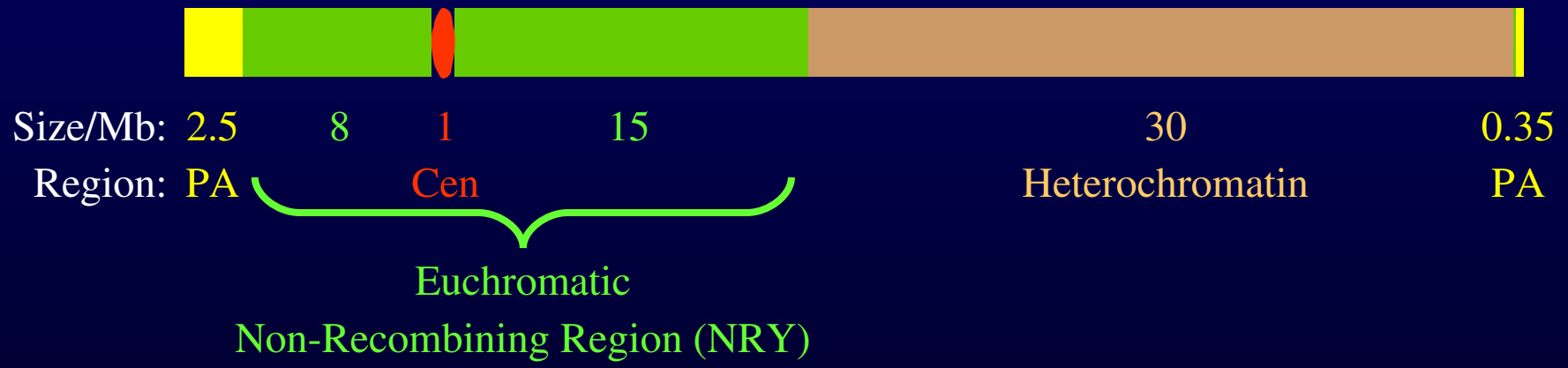
Mapping the Human Y Chromosome

Charles Tilford

Page Lab

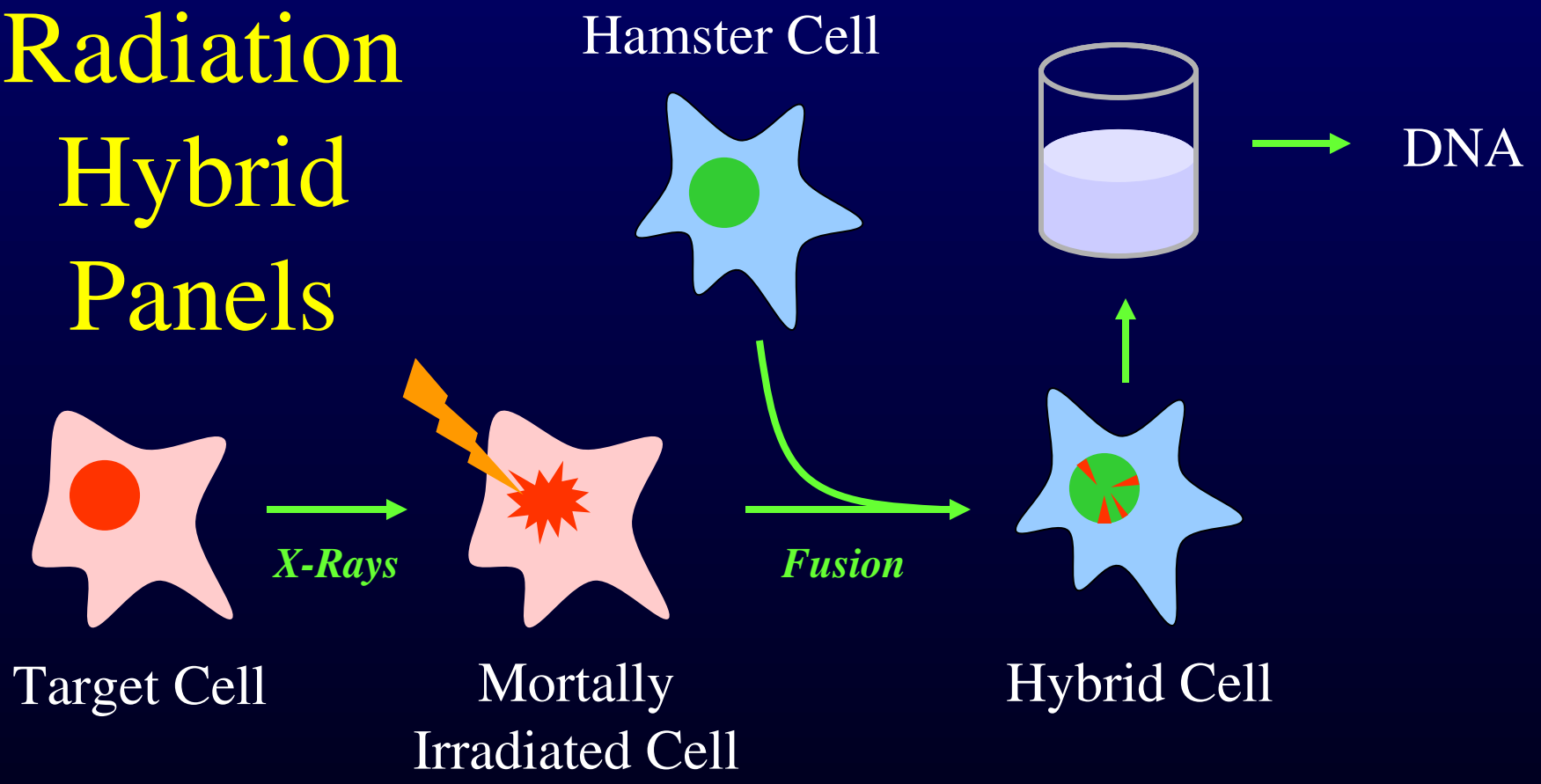
Whitehead Institute, MIT

The Human Y Chromosome



- Male sex determining, passed through male lineage only
- Only the pseudoautosomal regions (PA) recombine with X
- Heterochromatin has 10kb sequence complexity
- NRY contains 20+ genes in two broad categories:
  X-Y homologous, or Y-specific multicopy families.

Radiation Hybrid Panels



- Each hybrid has a full hamster genome, and some random fraction of the target genome
- X-ray intensity determines fragment size, which determines STS linkage range and ordering resolution

Radiation Hybrid Theory

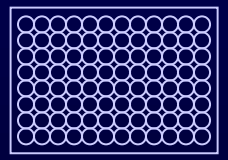
- Fragments are independently retained/rejected by host cell. Each hybrid will have a unique subset of fragments.

- Unlinked markers will associate randomly, since they will always be on different fragments.

- Linked markers with "zero" separation will associate absolutely, and with decreasing frequency as the distance between them (and thus the probability of a radiation-induced break occurring) increases

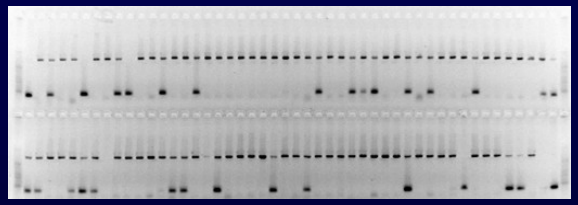
Radiation Hybrids in Action

RH Panel



PCR

Agarose Gel



Scoring

Vector

```
10100100110...
```

Generate Vector for each STS

```
sY14     0000101000011000000...
sY137    1000010010011000010...
sY63     0000000000000000010...
sY140    1000020010001000010...
sY217    0000000000001000010...
SP18-15  0000000000010001000...
SP18-07  0000000000010000000...
SP18-09  0000000000010000000...
...
```

Pairwise Comparison

Theta Matrix *(Large)*

Distance and Linkage Estimates

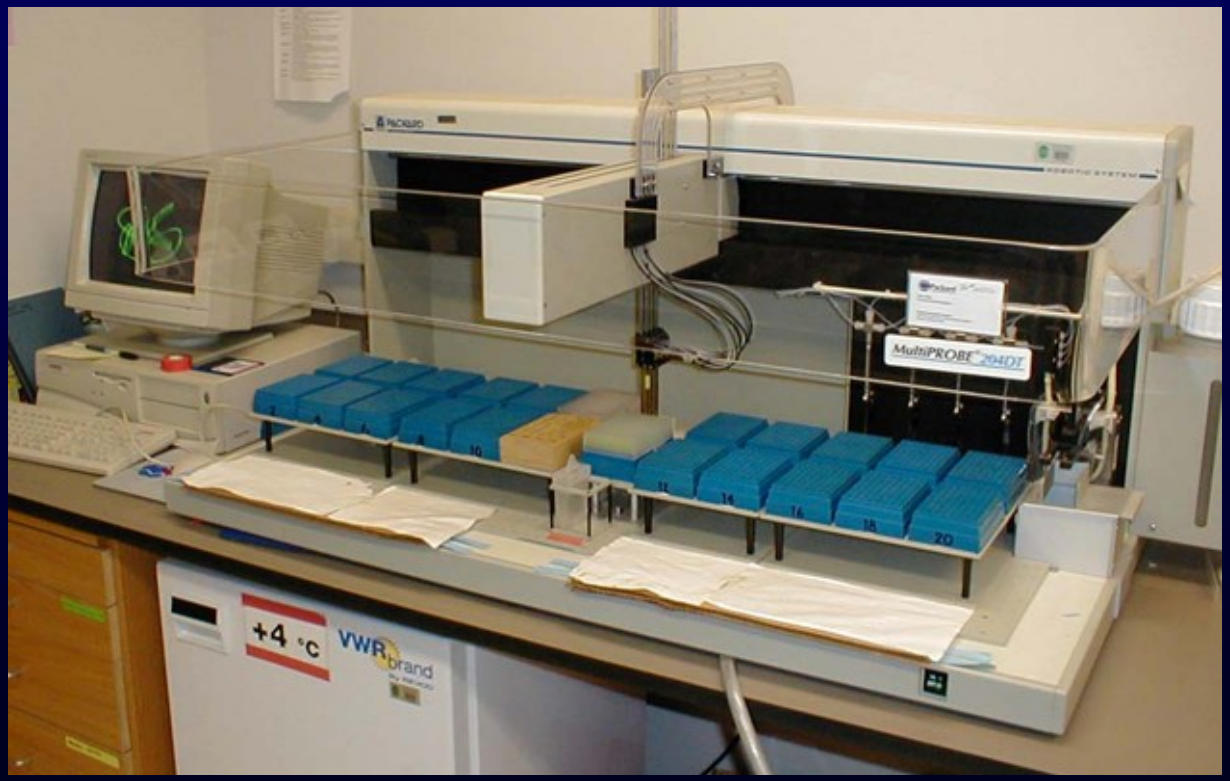
Distance Minimization

STS Order

Problem: Scale of Project

- Over 1000 markers to be mapped
- Each marker requires at least one plate of PCR, two plates for markers that provide useful map information
- PCR reactions are visualized with agarose gels
- Vector data needs to be extracted, stored and analyzed

High Throughput PCR

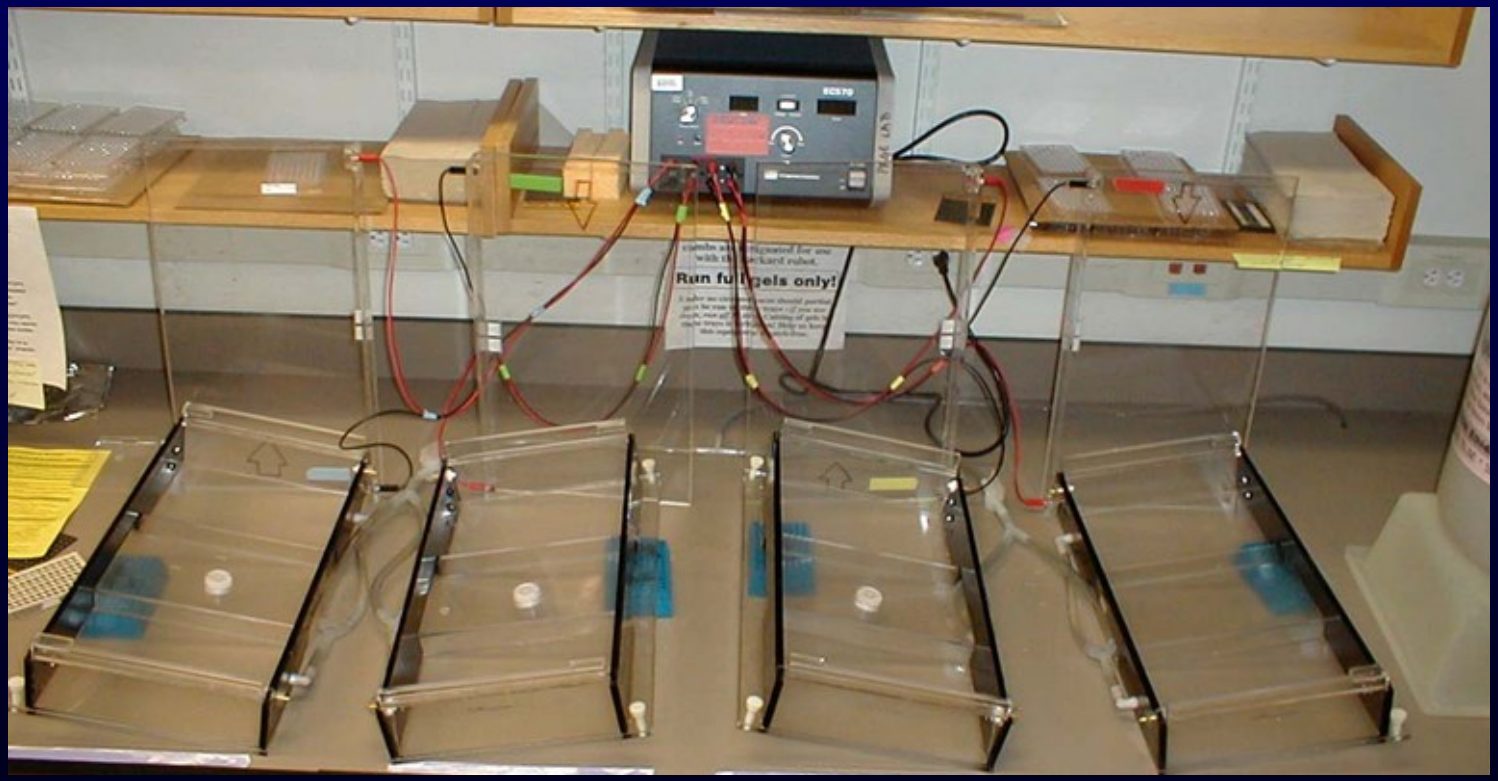


- Packard Multiprobe adapted for 20 x 96-well plates
- Adds template DNA, primers, PCR master mix, mineral oil and loading dye

High Throughput PCR

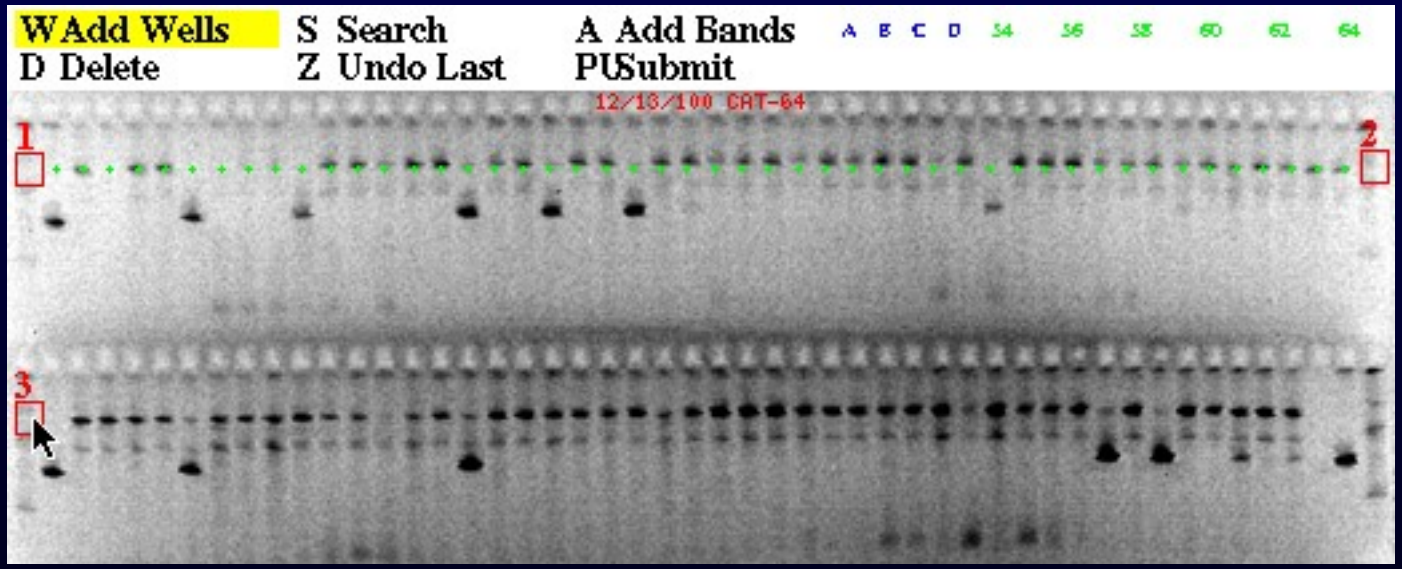
- Small, inexpensive, basic-function thermal cyclers (Hybaid Omn–E)

High Throughput PCR



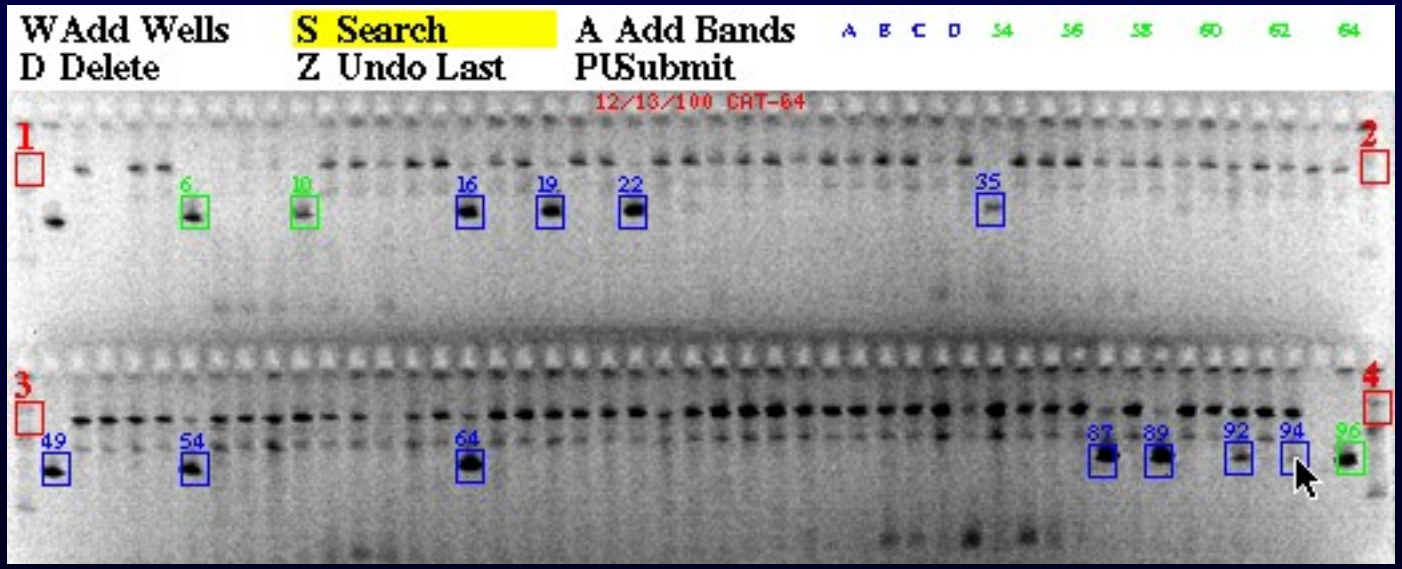
- Four Centipede gel boxes, each with 10 x 50-well combs
- Buffer plumbing system for rapid "dry" gel loading

Data Management - JavaGel



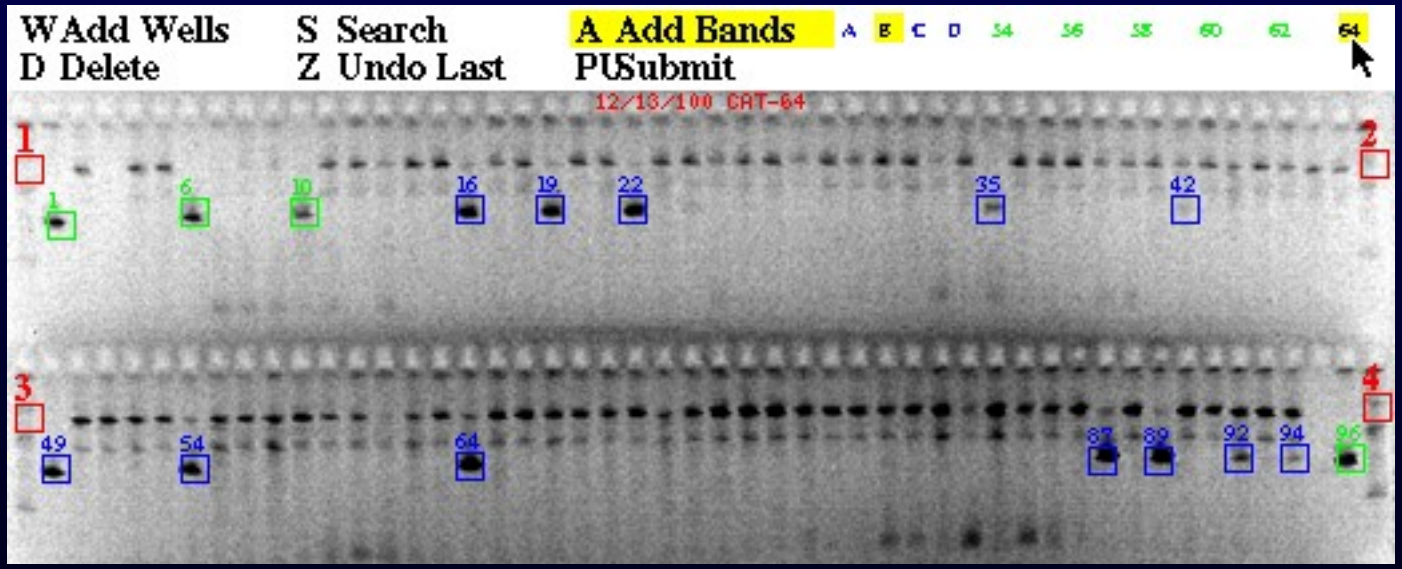
- Java applet for image annotation and data retrieval
- User clicks on the four edge marker lanes

Data Management - JavaGel



- In search mode, clicking on the faintest band finds all other bands of that size with the same or better S/N ratio

Data Management - JavaGel



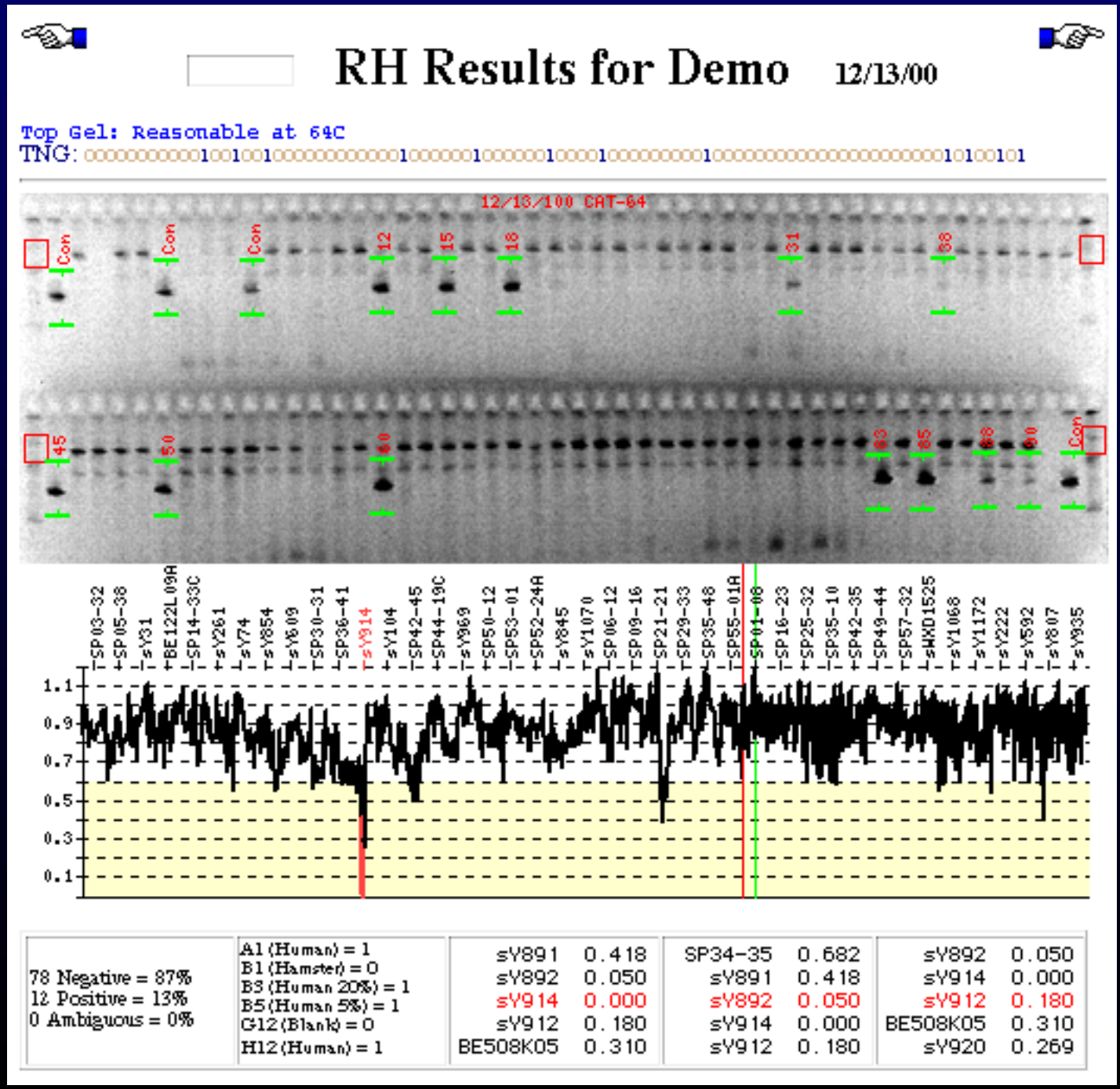
- Quality and annealing temperature information are entered

Vector

Annotated Gel Image

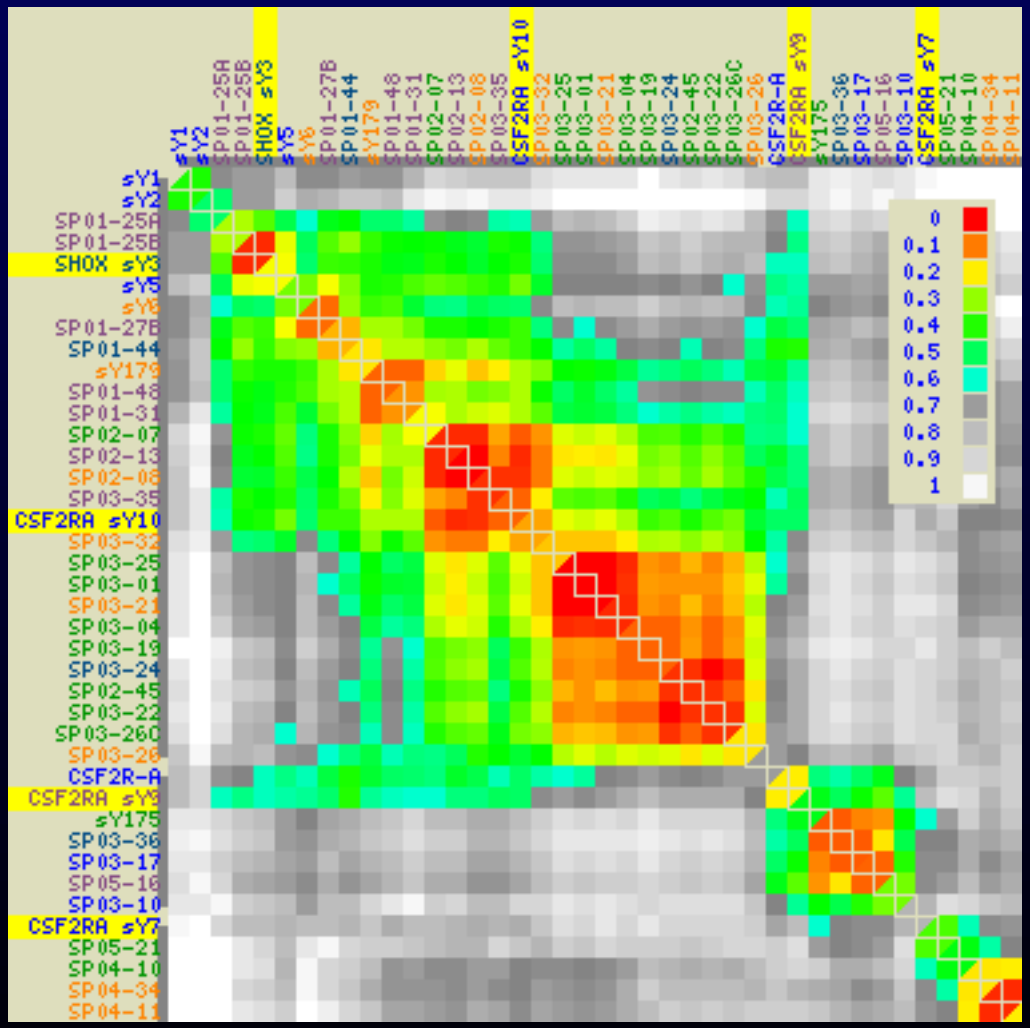
Theta scores for existing markers

Controls, best hits

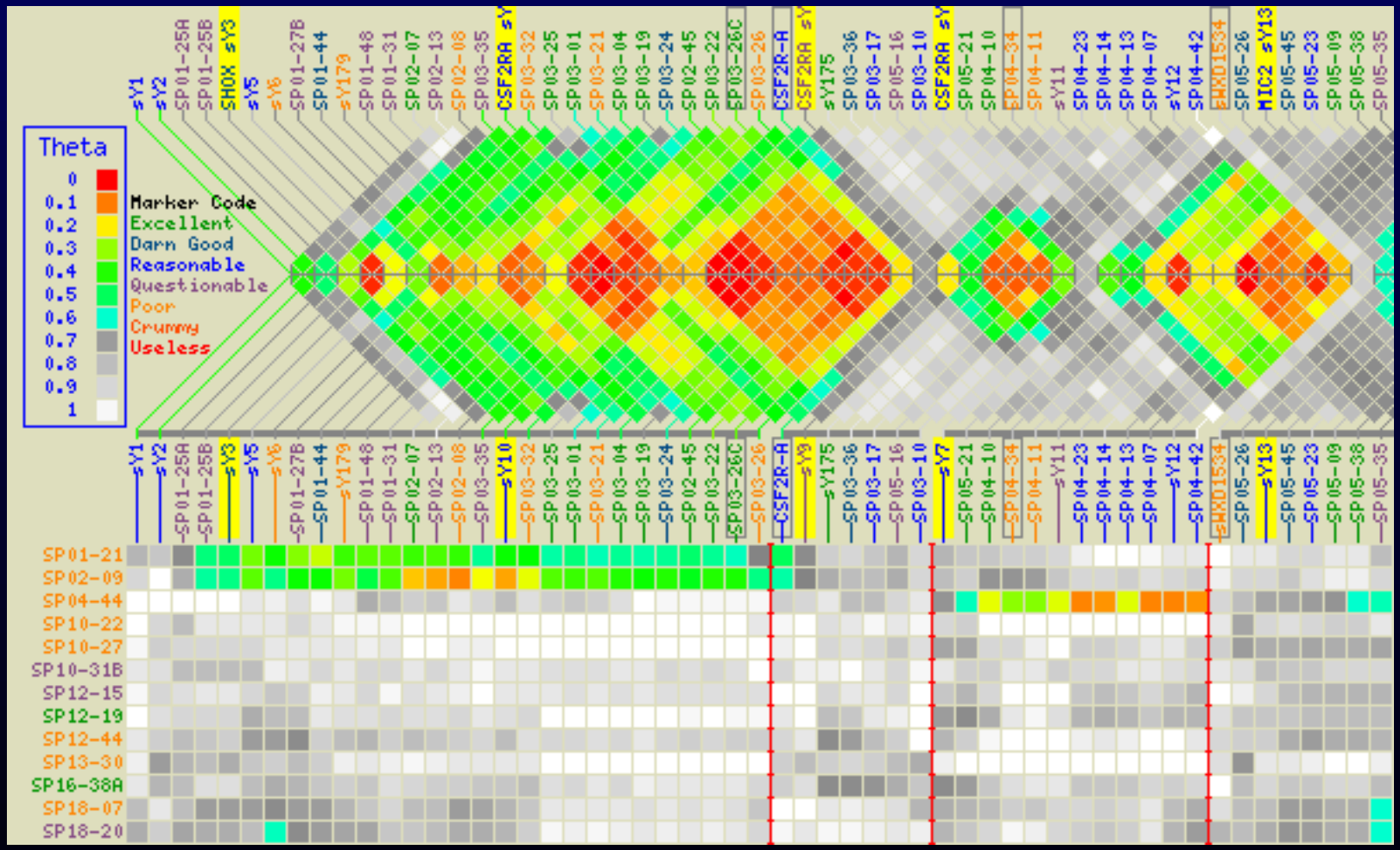


Graphical View of Theta Matrix

- Theta values for likely linkage represented by spectrum
- Likely unlinked values (>0.6) shown as gray scale
- Color of marker name indicates quality of STS
- Diagonal values (technically 0) are split between flanking markers



Diagonal View of Matrix

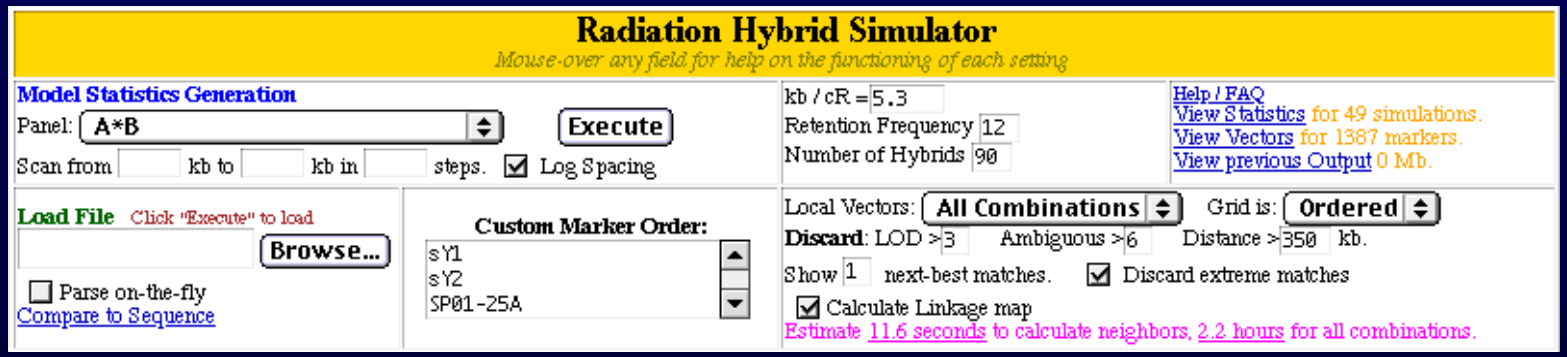


Ordering Markers

- User needs to grossly order markers due to multiple Y–specific repeat families

- Algorithmic ordering minimizes subdiagonal distance:
  - Exhaustive ordering within a small (6-8 marker) sliding window
  - "Random cost" ordering generates excellent order with much larger marker sets
    - Wang et al. A fast random cost algorithm for physical mapping. Proc Natl. Acad. Sci. USA 1994 (91) 11094-11098

RH "Simulator"

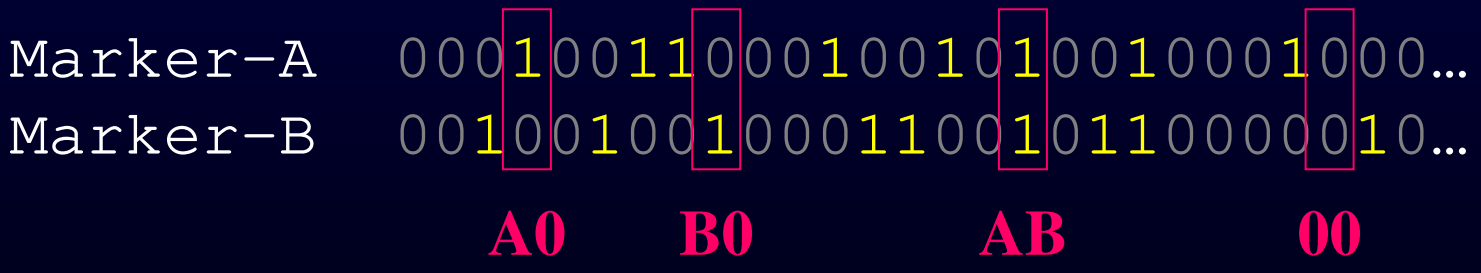


- Novel method for determining linkage likelihood
- Developed due to concern that existing algorithms would be less likely to find linkage in marker pairs with differing copy number :

```
Triploid 0101000101001100100110001001 01
Haploid  0100000100000000001100000000 0
```

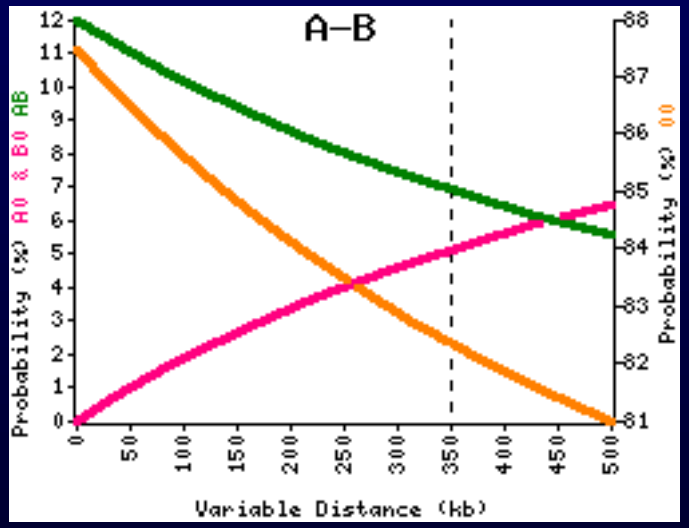
RH Simulator Theory

- System is provided with "models" of marker arrangement:
  - A-B A-B A-B A-B A-B B B *etc.*
  - Each model provides linkage information only, not order
- For each model, the probability is calculated for a hybrid being one of four classes in a pairwise comparison:

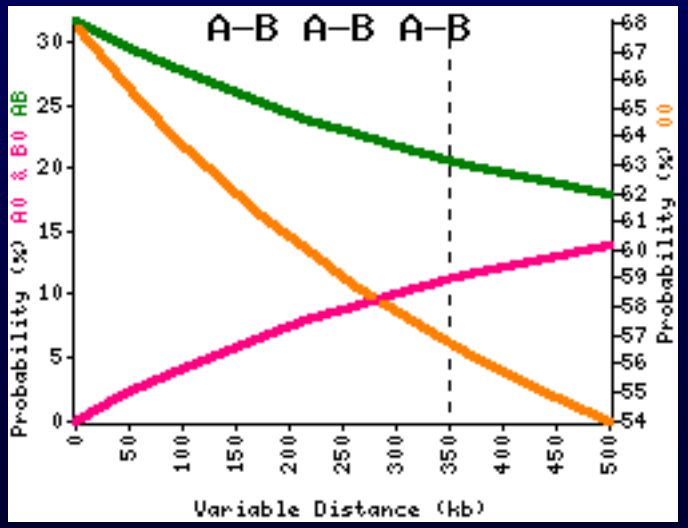


- These probabilities depend on:
  - The model, the distance between linked markers, the retention frequency (RF) and the average kilobases per centiRay. The last two values need to be observed from a subset of actual data.

Model Probability Curves



Linked Haploid



Linked Triploid

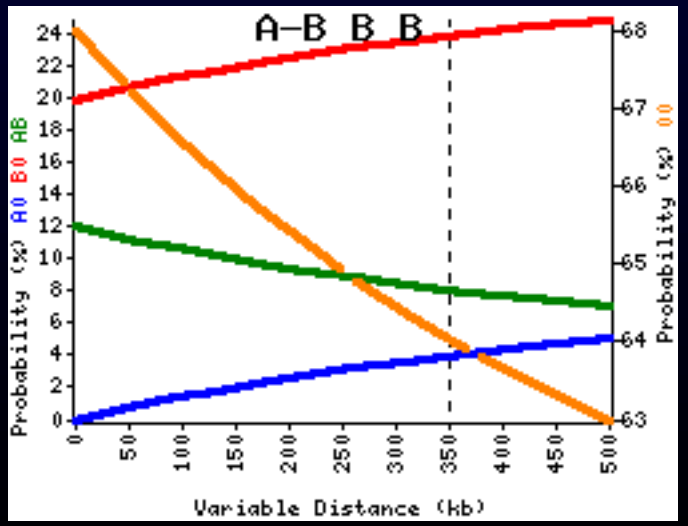
A0 = $A^+ B^-$

B0 = $A^- B^+$

Shown in pink when probabilities are equal

AB = $A^+ B^+$

00 = $A^- B^-$ *shown on right axis*



Haploid - Triploid

Overall Probability

- Given a vector pair, the number of hybrid pairs that are 00, AB, A0 & B0 are summed (for example, 79, 8, 1 & 2)

- Using the probability curves, the overall probability that a model would generate that set of 4 sums can be calculated for each distance. The distance with the best (greatest) probability is then reported for the model.

- These "model probabilities" can then be compared to each other to determine the likelihood of one model over the others

Ambiguous Data

- Markers are typically tested twice, on different days
- On average, about 1-3% of hybrids show conflicts between duplicate trials (i.e. one trial is positive, the other negative)
- Conflicts result from error (often gel loading, but also PCR failure) or marginal assays with "consistent inconstancy"
- Such ambiguous data are represented as "2" in the vector
- Program will analyze all possible vector pairs when ambiguous data are encountered, and report the average results with standard deviations

Sample Model Prediction Output

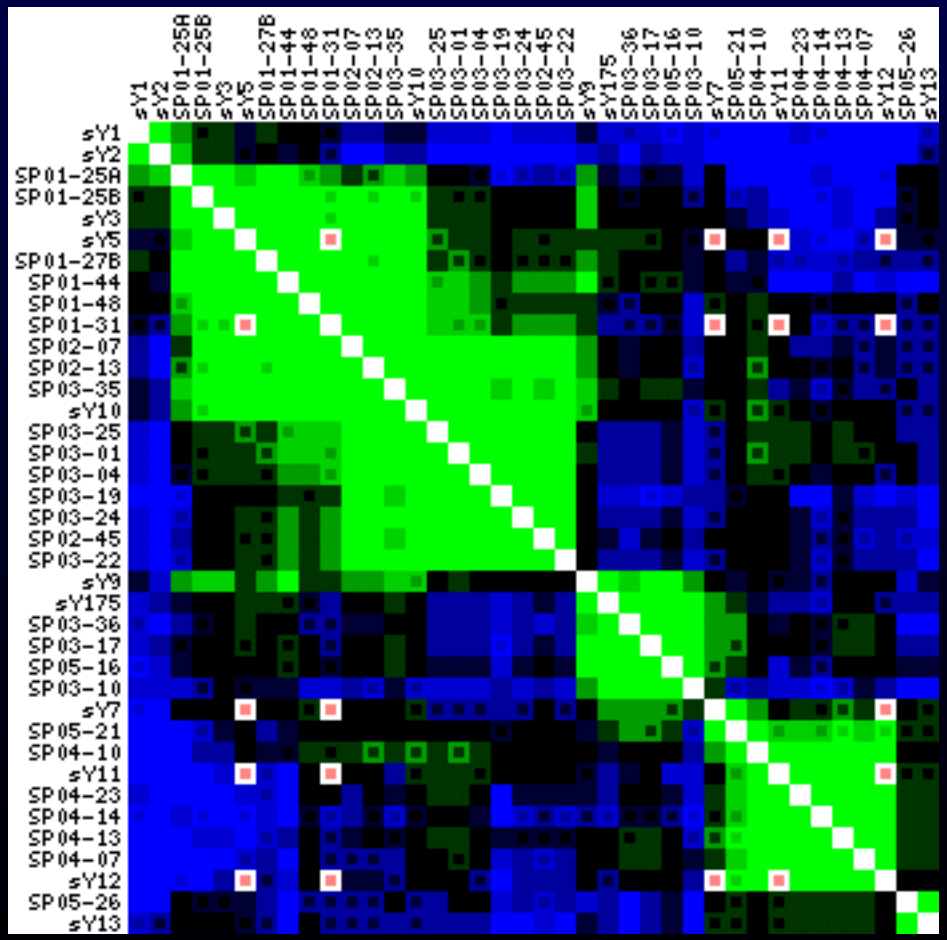
A = sY1
B = sY2

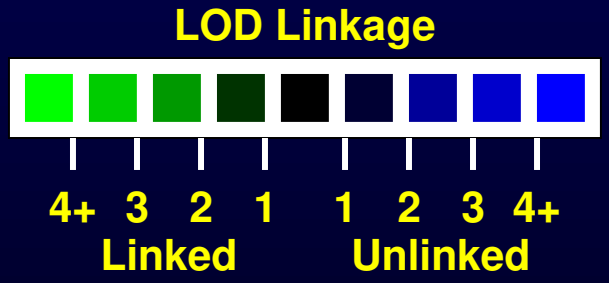
Model	$-\log_{10}(P)$		Distance
A–B B	1.95 ± 0.12	at	0 ± 0 kb
A–B B B	3.04 ± 0.16	at	0 ± 0 kb
A–B B B B	5.44 ± 0.30	at	0 ± 0 kb
A–B A–B	5.85 ± 0.34	at	256 ± 29 kb
A A–B B	6.16 ± 0.19	at	0 ± 0 kb
A A–B B B	6.97 ± 0.14	at	0 ± 0 kb
A B B	8.22 ± 0.55		Unlinked Model
A–B B B B B	8.37 ± 0.41	at	0 ± 0 kb
A–B A–B A–B	8.62 ± 0.47	at	194 ± 22 kb
A A–B B B B	9.13 ± 0.25	at	0 ± 0 kb
A B	9.32 ± 0.53		Unlinked Model
A B B B	9.41 ± 0.59		Unlinked Model
B B–A A A	10.43 ± 0.26	at	0 ± 0 kb
			etc...

- Observation - likelihoods are insufficient to choose a single model with high certainty.

Linkage Likelihood with Models

- By subtracting the best linked and unlinked models, a likelihood of linkage for STS pairs can be generated.





Inner color represents minimum LOD between standard deviations

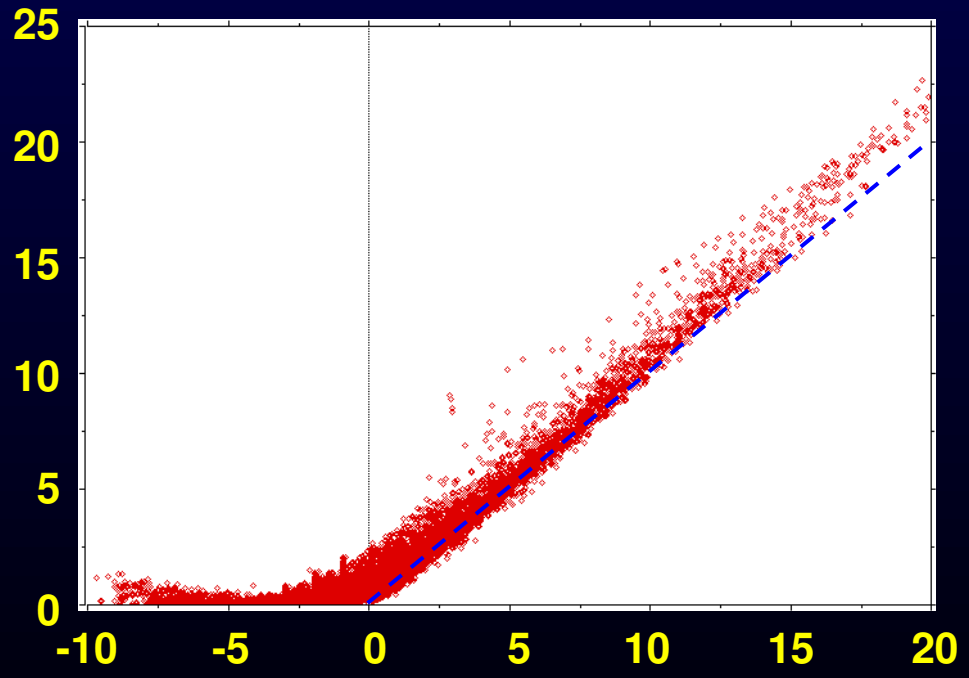
Marker pairs exceeding user-set maximum for ambiguous data

Comparison to Standard Methods

- LOD linkage results are similar to those obtained by standard algorithms, but tend to be lower.

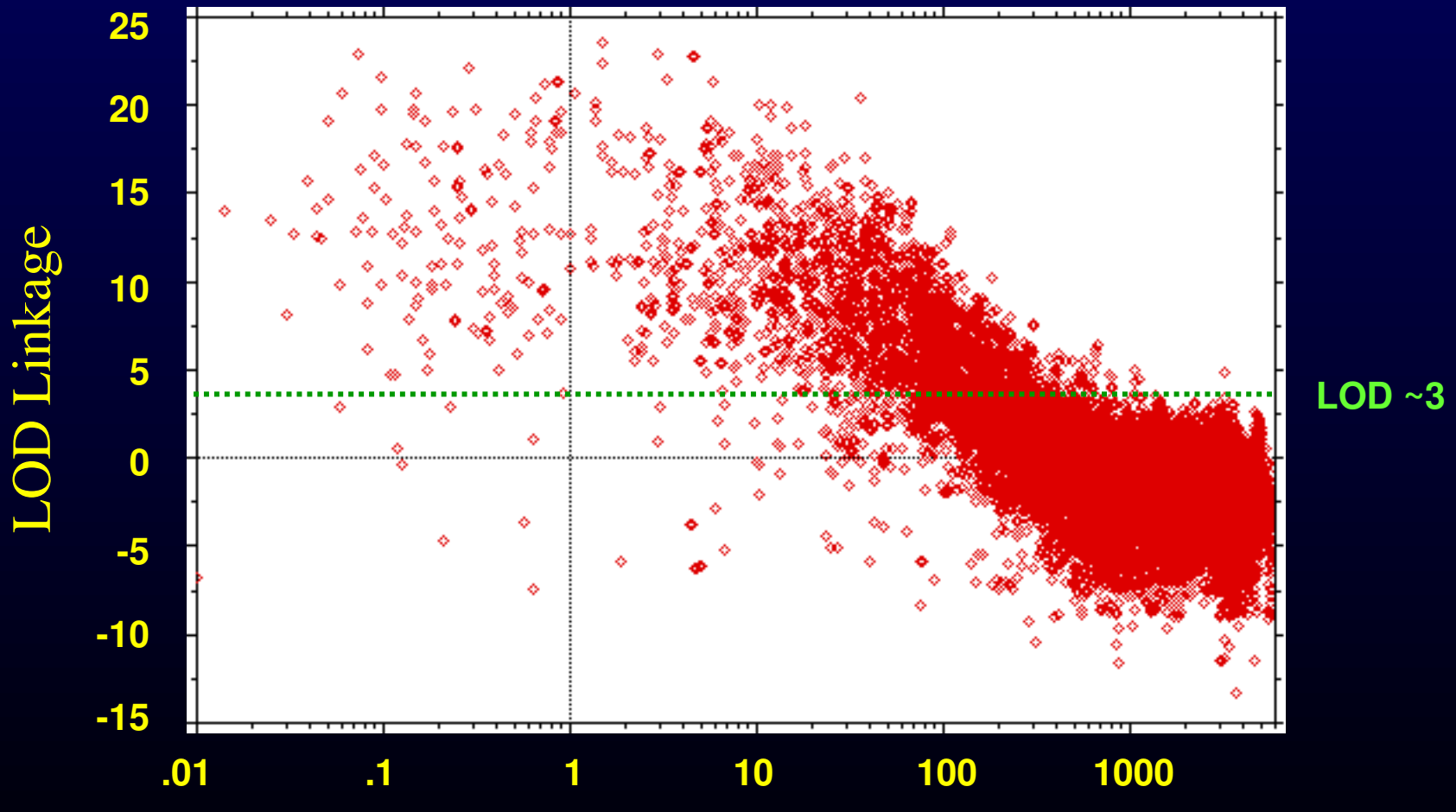
LOD Linkage as per:

Jones HB. Pairwise analysis of radiation hybrid mapping data. *Ann. Hum. Genet.* 60, 351-357 (1996)



LOD Linkage as per Simulation

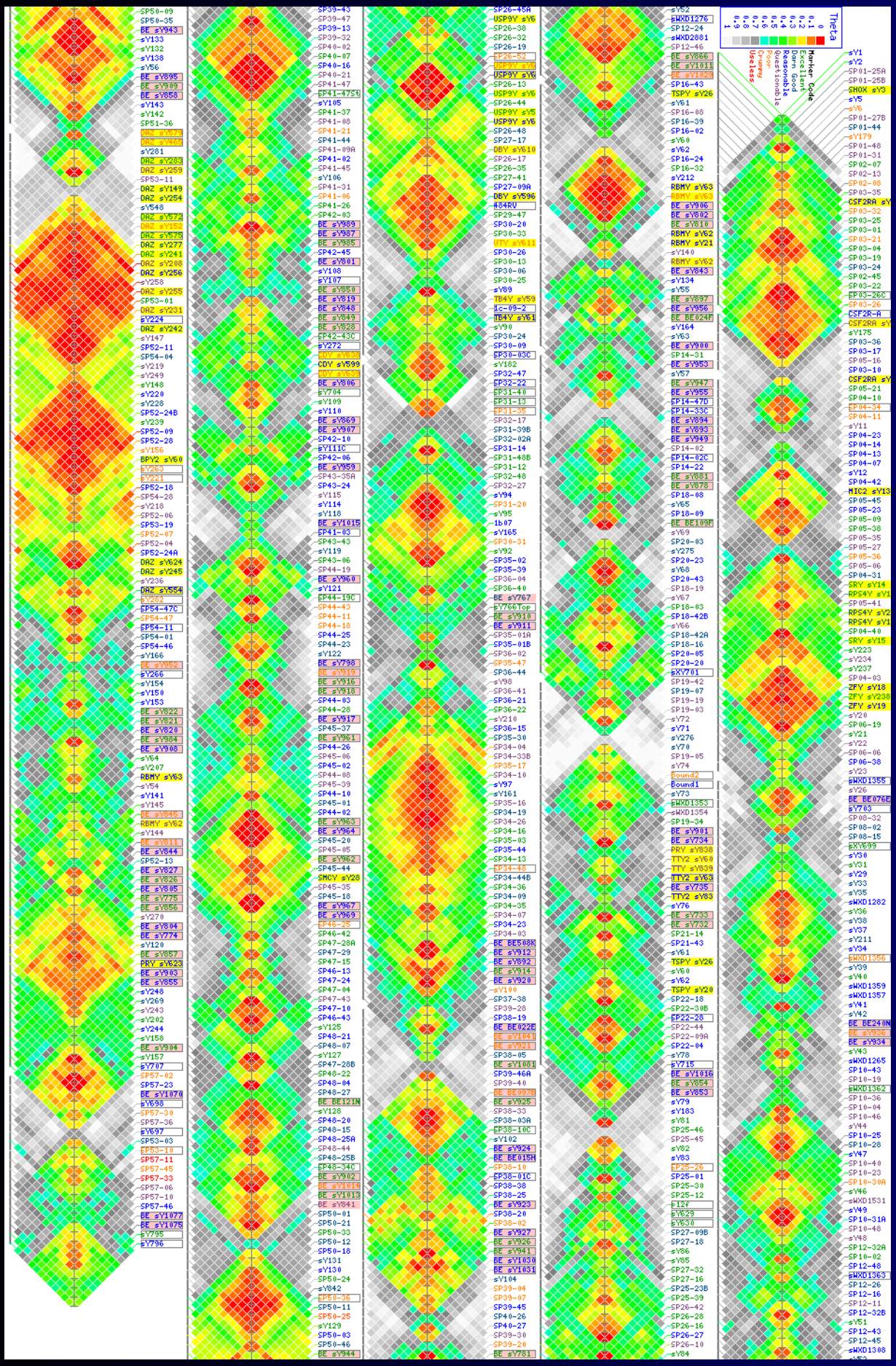
LOD score vs. Distance



Determined by ePCR

Conclusions

- Model simulation provides a stringent assay for determining linkage between RH vectors
- Likelihoods of copy number based on retention frequency are too similar to allow reliable choice of one model over all others when comparing markers.
- Final map has 700+ markers, with 11 gaps. Sequence comparison shows reasonable order within contiguous segments.


ThetaEa
Marker Code
Excellent
Damn Good
Reasonable
Questionable
Useless
SP50-09
SP50-35
BE sY943
sY133
sY132
sY138
sY56
BE sY895
BE sY909
BE sY858
sY143
sY142
SP51-36
DAZ sY574
DAZ sY462
sY281
DAZ sY283
DAZ sY259
SP53-11
DAZ sY149
DAZ sY254
sY548
DAZ sY572
DAZ sY152
DAZ sY575
DAZ sY277
DAZ sY241
DAZ sY208
DAZ sY256
sY258
DAZ sY255
SP53-01
DAZ sY231
sY224
DAZ sY242
sY147
SP52-11
SP54-04
sY219
sY249
sY148
sY220
sY228
SP52-24B
sY239
SP52-09
SP52-28
sY156
BPY2 sY60
sY263
sY221
SP52-18
SP54-28
sY218
SP52-06
SP53-19
SP52-07
SP52-04
SP52-24A
DAZ sY624
DAZ sY245
sY236
DAZ sY554
sY282
SP54-47C
SP54-47
SP54-11
SP54-01
SP54-46
sY166
BE sW952
sY266
sY154
sY150
sY153
BE sY822
BE sY821
BE sY820
BE sY884
BE sY908
sY64
sY207
RBMY sY63
sY54
sY141
sY145
BE sY845
RBMY sY62
sY144
BE sY811
BE sY844
SP52-13
BE sY827
BE sY826
BE sY805
BE sY775
BE sY856
sY270
BE sY804
BE sY774
sY120
BE sY857
CRY sY629
BE sY803
BE sY855
sY248
sY269
sY243
sY202
sY244
sY158
BE sY904
sY157
sY707
SP57-02
SP57-23
BE sY1070
sY698
SP57-30
SP57-36
sY697
SP53-03
SP53-10
SP57-11
SP57-45
SP57-33
SP57-06
SP57-10
SP57-46
BE sY1077
BE sY1075
sY795
sY796
SP39-43
SP39-47
SP39-13
SP39-32
SP40-02
SP40-07
SP40-16
SP40-21
SP41-47
SP41-47SY
sY105
SP41-37
SP41-08
SP41-21
SP41-44
SP41-09A
SP41-02
SP41-45
sY106
SP41-31
SP41-06
SP41-26
SP42-03
BE sY989
BE sY987
BE sY985
SP42-45
BE sY801
sY108
sY107
BE sY850
BE sY819
BE sY848
BE sY849
BE sY828
SP42-43C
sY272
CDY sY636
CDY sY599
CDY sY634
BE sY806
sY704
sY109
sY110
BE sY869
BE sY907
SP42-10
sY111C
SP42-06
BE sY959
SP43-35A
SP43-24
sY115
sY114
sY118
BE sY1015
SP41-03
SP43-43
sY119
SP43-06
SP44-19
BE sY960
sY121
SP44-19C
SP44-43
SP44-11
SP44-18
SP44-25
SP44-23
sY122
BE sY798
BE sY919
BE sY916
BE sY918
SP44-03
SP44-28
BE sY917
SP45-37
BE sY961
SP44-26
SP45-06
SP45-02
SP44-08
SP45-39
SP44-10
SP45-01
SP44-02
BE sY963
BE sY964
SP45-20
SP45-05
BE sY962
SP45-44
SMCY sY28
SP45-35
SP45-18
BE sY967
BE sY969
SP46-25
SP46-42
SP47-28A
SP47-29
SP47-15
SP46-13
SP47-24
SP47-04
SP47-43
SP47-10
SP46-43
sY125
SP48-21
SP48-07
sY127
SP47-28B
SP48-22
SP48-04
SP48-27
BE BE121N
sY128
SP48-20
SP48-15
SP48-25A
SP48-44
SP48-25B
SP48-34C
BE sY902
BE sY1014
BE sY1013
BE sY841
SP50-01
SP50-21
SP50-33
SP50-12
SP50-18
sY131
sY130
SP50-24
sY842
SP50-36
SP50-11
SP50-25
sY129
SP50-03
SP50-46
BE sY944
SP26-45A
USP9Y sY6
SP26-36
SP26-32
SP26-19
SP26-52
USP9Y sY6
USP9Y sY6
SP26-13
USP9Y sY6
SP26-44
USP9Y sY5
USP9Y sY6
SP26-48
SP27-17
DBY sY610
SP26-17
SP26-35
SP27-41
SP27-09A
DBY sY596
484RV
SP29-47
SP30-20
SP30-33
UTY sY611
SP30-26
SP30-13
SP30-06
SP30-25
sY89
TB4Y sY59
4c-09-2
TB4Y sY61
sY90
SP30-24
SP30-09
SP30-03C
sY182
SP32-47
SP32-22
SP31-40
SP31-15
SP31-35
SP32-17
SP31-39B
SP32-02A
SP31-14
SP31-48B
SP31-12
SP32-48
SP32-27
sY94
SP31-20
sY95
1b07
sY165
SP30-31
sY92
SP35-02
SP35-39
SP36-04
SP36-40
BE sY767
sY766Top
BE sY910
BE sY911
SP35-01A
SP35-01B
SP36-02
SP35-47
SP36-44
sY98
SP36-41
SP36-21
SP36-22
sY210
SP36-15
SP35-30
SP34-04
SP34-33B
SP35-17
SP34-10
sY97
sY161
SP35-16
SP34-19
SP34-26
SP34-16
SP35-03
SP35-44
SP34-13
SP34-48
SP34-44B
SP34-36
SP34-09
SP34-35
SP34-07
SP34-23
SP34-03
BE BE568K
BE sY912
BE sY892
BE sY914
BE sY920
sY100
SP37-38
SP39-28
SP38-19
BE BE022E
BE sY1041
BE sY921
SP38-05
BE sY1081
SP39-46A
SP39-40
BE BE097B
BE sY925
SP38-33
SP38-03A
SP38-10C
sY102
BE sY924
BE BE015H
SP38-11
SP38-01C
SP38-38
SP38-25
BE sY923
SP38-20
SP38-02
BE sY927
BE sY926
BE sY941
BE sY1030
BE sY1031
sY14
SP39-04
SP39-07
SP39-45
SP40-26
SP40-27
SP39-30
SP39-20
BE sY781
sY52
sWXD1276
SP12-24
sWXD2881
SP12-46
BE sY866
BE sY1011
BE sY1029
SP16-43
TSPY sY26
sY61
SP16-08
SP16-39
SP16-02
sY60
sY62
SP16-24
SP16-32
sY212
RBMY sY63
RBMY sY63
BE sY906
BE sY802
BE sY810
RBMY sY62
RBMY sY21
sY140
RBMY sY62
BE sY843
sY134
sY55
BE sY897
BE sY956
BE BE024F
sY164
sY63
BE sY900
SP14-31
BE sY953
sY57
BE sY947
BE sY955
SP14-47D
SP14-33C
BE sY894
BE sY893
BE sY949
SP14-02
SP14-02C
SP14-22
BE sY881
BE sY878
SP18-08
sY85
SP18-09
BE BE109F
sY69
SP20-03
sY275
SP20-23
sY68
SP20-43
SP18-19
sY67
SP18-03
SP18-42B
sY66
SP18-42A
SP18-16
SP20-05
SP20-20
sXY701
SP19-42
SP19-07
SP19-19
SP19-03
sY72
sY71
sY276
sY70
SP19-05
sY74
Bound2
Bound1
sY73
sWXD1353
sWXD1354
SP19-34
BE sY901
BE sY734
PRY sY838
TTY2 sY60
TTY sY839
TTY2 sY63
BE sY735
TTY2 sY83
sY76
BE sY733
BE sY732
SP21-14
SP21-43
sY61
TSPY sY26
sY60
sY62
TSPY sY20
SP22-18
SP22-30B
SP22-28
SP22-44
SP22-09A
SP22-04
sY78
sY715
BE sY1016
BE sY854
BE sY853
sY79
sY183
sY81
SP25-46
SP25-45
sY82
sY83
SP25-26
SP25-01
SP25-30
SP25-12
P12F
sY629
sY630
SP27-09B
SP27-18
sY86
sY85
SP27-32
SP27-16
SP25-23B
SP25-39
SP26-42
SP26-28
SP26-16
SP26-27
SP26-10
sY84
sY1
sY2
SP01-25A
SP01-25B
SHOX sY3
sY5
sY6
SP01-27B
SP01-44
sY179
SP01-48
SP01-31
SP02-07
SP02-13
SP02-08
SP03-35
CSF2RA sY
SP03-32
SP03-25
SP03-01
SP03-21
SP03-04
SP03-19
SP03-24
SP02-45
SP03-22
SP03-26C
SP03-26
CSF2R-A
CSF2RA sY
sY175
SP03-36
SP03-17
SP05-16
SP03-10
CSF2RA sY
SP05-21
SP04-10
SP04-34
SP04-11
sY11
SP04-23
SP04-14
SP04-13
SP04-07
sY12
SP04-42
MIC2 sY13
SP05-45
SP05-23
SP05-09
SP05-38
SP05-35
SP05-27
SP05-36
SP05-06
SP04-31
SRY sY14
RPS4Y sY1
SP05-41
RPS4Y sY2
RPS4Y sY1
SP04-40
SRY sY15
sY223
sY234
sY237
SP04-03
ZFY sY18
ZFY sY238
ZFY sY19
sY20
SP06-19
sY21
sY22
SP06-06
SP06-38
sY23
sWXD1355
sY26
BE BE076E
sY703
SP08-32
SP08-02
SP08-15
sXY699
sY30
sY31
sY29
sY33
sY35
sWXD1282
sY36
sY38
sY37
sY211
sY34
sWXD1358
sY39
sY40
sWXD1359
sWXD1357
sY41
sY42
BE BE240N
BE sY936
BE sY934
sY43
sWXD1265
SP10-43
SP10-19
sWXD1362
SP10-36
SP10-04
SP10-46
sY44
SP10-25
SP10-28
sY47
SP10-40
SP10-23
SP10-30A
sY46
sWXD1531
sY49
SP10-31A
SP10-48
sY48
SP12-32A
SP10-02
SP12-48
sWXD1363
SP12-26
SP12-16
SP12-11
SP12-32B
sY51
SP12-43
SP12-45
sWXD1308

Acknowledgments

Steve Rozen	*Programming Consultation*
Helen Skaletsky	*Statistics / Programming Consultation*
Loreall Pooler	*Marker Typing*
Chad Nussbaum	*Mapping Consultation / Reagents*
Tom Hudson	*Mapping Consultation*
Eric Lander	*Idea Guru*
David Page	*Patron and Mentor*

RH Simulator Theory I

- System is provided with "models" of marker arrangement:
  - A-B A-B A-B A-B A-B B B *etc.*
  - Each model provides linkage information only, not order

- Two basic levels of probability calculations:
  - Probability of a linked pair (A-B) becoming separated is function of distance between them (d) and "intensity of radiation" ($\lambda = 0.01/kb_{per}cR$): $\theta = 1 - e^{-\lambda * d}$
  - Probability of retaining a fragment is retention frequency (RF)
  - Both RF and λ are experimentally determined
  - RF and λ may vary across chromosome, but are treated as constant

RH Simulator Theory II

- We can simplify the system by considering only the four possible outcomes for a pairwise comparison of a hybrid:
  - both positive (AB), both negative (00), or A0 or B0
- Models are composed of independent components, calculate probabilities for each of three component types:

	A	B	A-B
pAB	0	0	$(1-\theta)\ RF + \theta\ RF^2$
p00	1-RF	1-RF	$(1-\theta)\ (1-RF) + \theta\ (1-RF)^2$
pA0	RF	0	$\theta\ RF\ (1-RF)$
pB0	0	RF	$\theta\ RF\ (1-RF)$

RF = retention frequency, θ = break probability (function of distance)

RH Simulator Theory III

- Find total probabilities for the model as:

$p00_{Tot}$ Product (p00) for all components

$pA0_{Tot}$ (Product (pA|0) for all components) - $p00_{Tot}$
where pA|0 = pA0 + p00

$pB0_{Tot}$ (Product (pB|0) for all components) - $p00_{Tot}$
where pB|0 = pB0 + p00

pAB_{Tot} $1 - (p00_{Tot} + pA0_{Tot} + pB0_{Tot})$

- Calculate these total probabilities for each model, and for a finite set of distances (generally 12)

RH Simulator Theory IV

- We can now apply these four probabilities to determine the probability that a given model would generate an observed vector pair at a specified distance:

$$P = \frac{p00^{w}\ pA0^{x}\ pB0^{y}\ pAB^{z}\ N!}{w!\ x!\ y!\ z!}$$

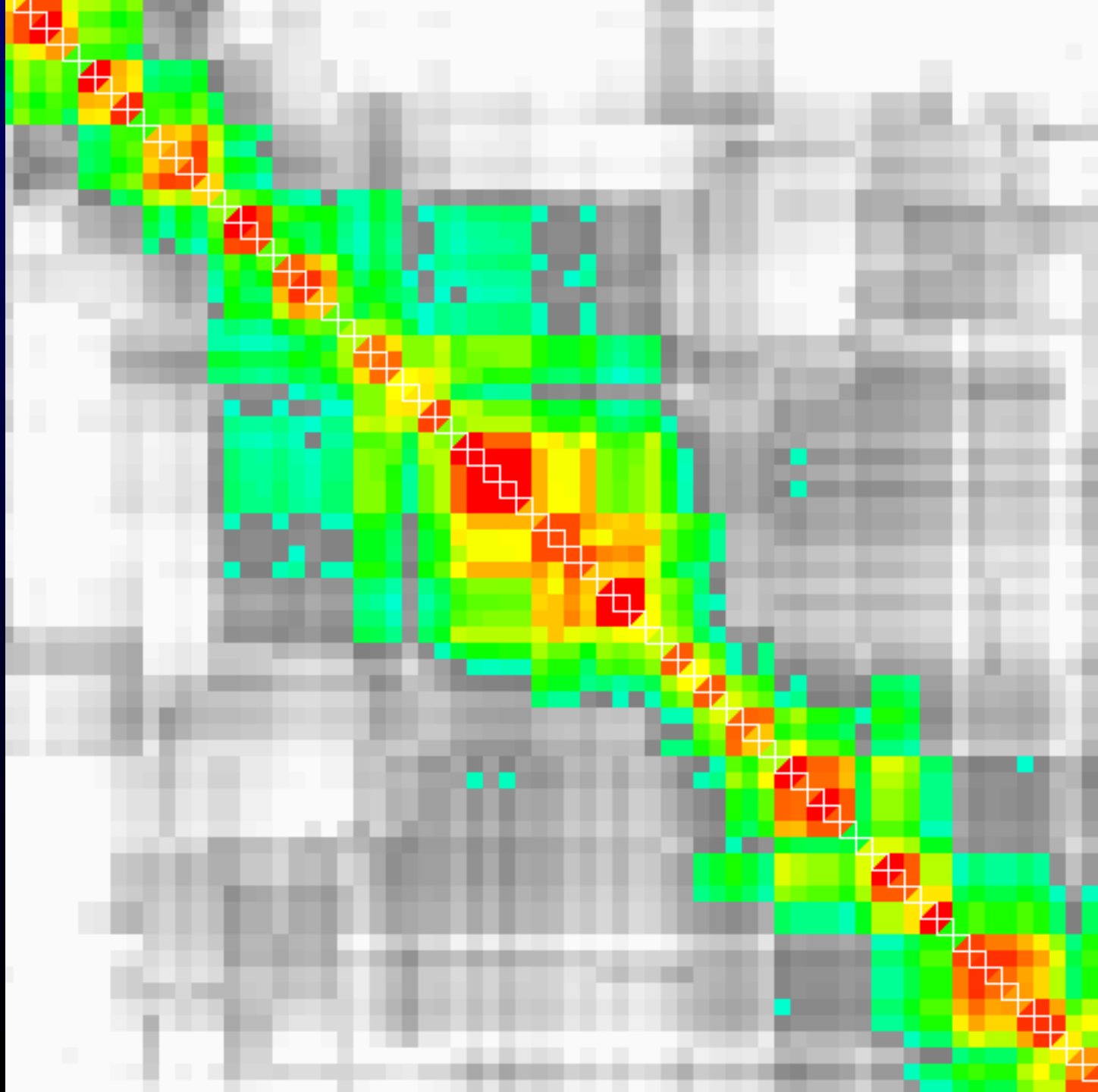
Where:

`w` = number of hybrids 00
`x` = number of hybrids A0
`y` = number of hybrids B0
`z` = number of hybrids AB
`N` = total number of hybrids

Gap Analysis

Gap	Left STS	Link LOD	LOD=3 Copy# Lft	Rgt	Distance / Comment
	SP03-26	0.10	1-2	1-2	No Seq Data (not real gap)
1	SP03-10	2.52	1-2	2-5	No Seq Data
	SP04-42	2.23	1-3	1-3	No Seq Data (not real gap)
2	sY23	2.41	1-4	2-5	439 kb
3	sY1029	-0.41	1-2	2-5	125 kb + Near site of inversion...
4	sY900	-2.32	1-3	1-4	160 kb +
5	SP14-22	-1.36	1	1-2	540 kb
6	sXY701	-2.00	1-2	1-3	270 kb
7	sY715	1.20	6+	1-2	Centromere
8	SP30-03	-0.92	1-2	1-2	100 kb
9	SP37-38	0.69	1-2	1-3	147 kb
10	SP51-36	-0.91	1-2	1-4	??
11	sY707	1.93	1-4	1-2	??

"Well Behaved" Markers



"Poorly Behaved" Markers