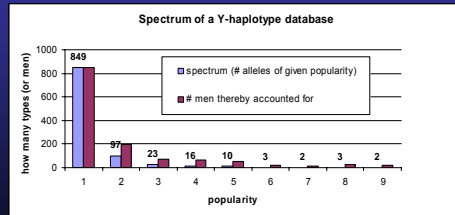


## Y database $p$ -values Slatkin's neutrality test

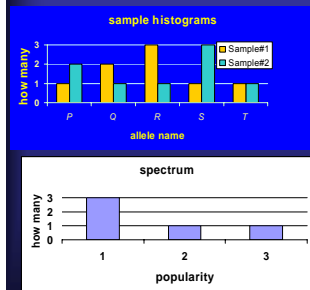
Charles H. Brenner, Ph.D.  
DNA-VIEW™  
Visiting Scholar, School of Public Health, UC Berkeley  
Consultant in forensic DNA mathematics  
<http://dna-view.com>

## Spectrum

- $N=1300$  (from Reliagene US Black, 11 loci)
- # different alleles =  $1005 = 849 + 97 + 23 + \dots$
- # men =  $N = 1300 = 849 + 2 \cdot 97 + 3 \cdot 23 + \dots$



## Spectrum vs. Histogram



Two different samples:

PQQRRTST  
PPQRSSST

Same spectrum for both:

3 1 1  
( PST QQ RRR  
QRT PP SSS )

## Why spectra?

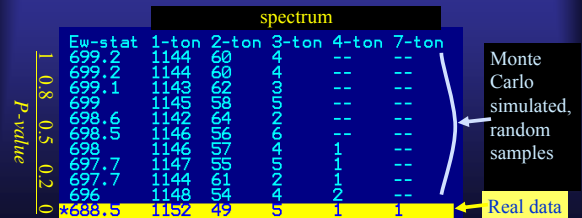
- Evolution favors some spectra over others.
- For a given spectrum, evolution does not favor one histogram over others
  - (assuming that the names are arbitrary)
- Therefore, the sample spectrum tells you something about evolution. I.e. is spectrum consistent with the model of "infinitely many selectively neutral alleles"?
  - All alleles (=haplotypes) are selectively neutral
  - Mutation always creates a new allele
  - Each generation is randomly selected from previous (with mutation)

## Possible spectra

- Suppose  $N=8$  men and  $k=5$  distinct alleles
  - For given  $N$  & assuming " $\infty$  alleles" model,  $k$  depends on mutation rate & random chance
    - If very high mutation rate,  $k=N$
    - If zero mutation rate,  $k=1$
- Possible spectra?
  - 3 1 1 (3 singletons, 1 pair, 1 trio)
  - 2 3 (2 singletons, 3 pairs)
  - 4 0 0 1 (4 singletons, 1 quad)
- Some spectra are more (311) or less (4001) likely under the model, hence tend to support or refute the model.
  - That's the principle behind Slatkin's neutrality test.

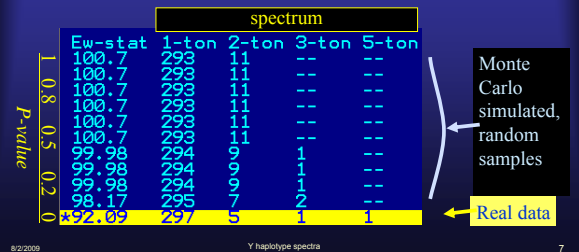
## Inconsistent with neutral model

- Caucasian ABI (old) 1276 haplotypes
- $p$ -value=0.00 (Slatkin exact test by Monte-Carlo) for conformity to model of  $\infty$ ly many selectively neutral types



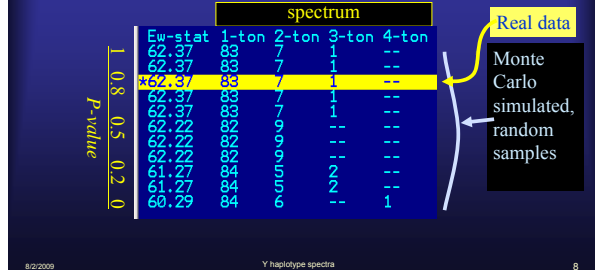
## Inconsistent with neutral model

- Portuguese data; 315 haplotypes
- $p\text{-value}=0.00$  (Slatkin exact test by Monte-Carlo) for conformity to model of  $\infty$ ly many selectively neutral types



## Consistent with neutral model

- Macedonian data; 100 haplotypes
- $p\text{-value}=0.8$  (Slatkin exact test by Monte-Carlo) for conformity to model of  $\infty$ ly many selectively neutral types



## p-values for Slatkin neutrality test

name	n	nloci	p-val	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	a15	a16
Krakow IES	435	17	0.80	403	16														
Macedonian	99	17	0.79	82	7	1													
Indian KL	304	17	0.65	286	9														
Thai ABI 2009	241	17	0.59	229	6														
Krakow(Sanak)	189	17	0.57	179	5														
Chinese KL	326	17	0.55	290	18														
Vietnamese ABI	103	17	0.54	97	3														
Filipino ABI	106	17	0.51	101	2														
Native Amr ABI	106	17	0.50	106															
African ABI	59	17	0.25	39	6	1		1											
Malaysian KL	338	17	0.09	315	5	1													
SE Hispanic ABI	597	17	0.07	543	22	2	1												
Black (US) ABI	977	17	0.03	917	28	1													
Asian ABI	327	17	0.01	309	7	1													
Native Am Reliagene	104	11	0.01	77	8	1			1										
Black(US) Reliagene	1300	11	0	849	97	23	16	10	3	2	3	2							
Caucasian Reliagene	1242	11	0	691	81	24	12	5		1	4	2							
Hispanic Reliagene	452	11	0	348	16	6	2	2	2		1			1					
Caucasian ABI 2009	1300	17	0	1174	38	8	1	2											
Black (US) ABI 2009	1300	17	0	1215	34	4	1												
Portuguese	313	17	0	295	5	1	1												
Jew Ashk. Hammer	496	10	0	176	28	19	7	5	1	1				1	1	2	1	1	
Caucasian ABI	1272	17	0	1148	49	5	1												

## Comments about non-neutrality

- Big excess of low and zero  $p$ -values.
  - And several medium  $p$ -values are with small samples for which little spectral variety is possible.
- Partly due to very common types, explainable as very successful haplotypes.
- But typically there is a singleton excess and a deficiency of doubletons. Why? What mechanism?

The end