# DNA Identification: Inclusion Genotype and LR

**Mark W Perlin, PhD, MD, PhD**
**Cybergenetics, Pittsburgh, PA**

**TrueAllele® Lectures**
**Fall, 2010**

**Cybergenetics**

---

# Mixture Data



Some amount of contributor A genotype

+

Other amount of contributor B genotype

PCR →

Mixture data with genotypes of contributors A & B

---
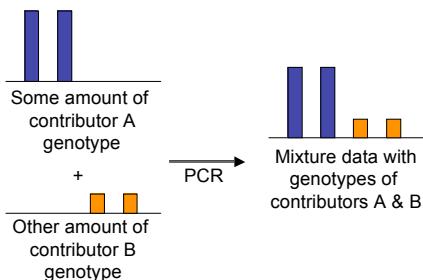
# Quantitative Likelihood



---
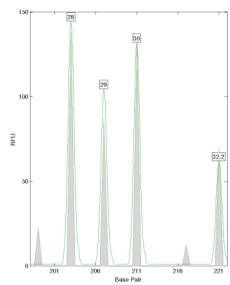
## Qualitative Likelihood



## Punnett Square



|   | 1 | 2 |
|---|---|---|
| 1 | $p_1^2$ | $p_1 p_2$ |
| 2 | $p_1 p_2$ | $p_2^2$ |

## Prior Probability



probability

0.10

$p_1^2$    $2p_1 p_2$    $p_2^2$

1, 1    1, 2    2, 2    …
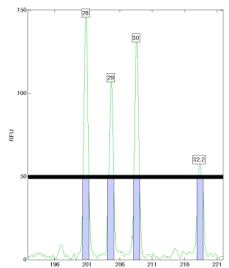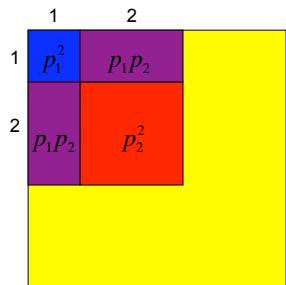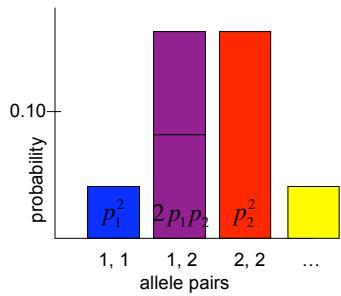
allele pairs

Likelihood Function

---
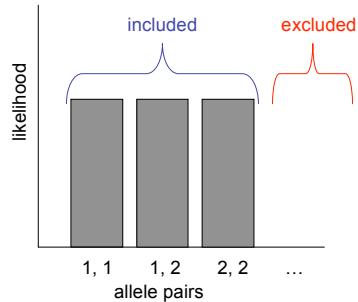
# Genotype Inference

| | allele pair $x$ | prior $\pi_Q(x)$ | likelihood $\lambda_Q(x)$ | product $\lambda_Q(x)\cdot\pi_Q(x)$ | posterior $q(x)$ |
|---|---|---|---|---|---|
| included | 1,1 | $p_1^2$ | 1 | $p_1^2$ | $p_1^2\big/(p_1+p_2+\ldots+p_K)^2$ |
| | 1,2 | $2p_1p_2$ | 1 | $2p_1p_2$ | $2p_1p_2\big/(p_1+p_2+\ldots+p_K)^2$ |
| | ... | $r(x)$ | 1 | $r(x)$ | $r(x)\big/(p_1+p_2+\ldots+p_K)^2$ |
| | $K,K$ | $p_K^2$ | 1 | $p_K^2$ | $p_K^2\big/(p_1+p_2+\ldots+p_K)^2$ |
| excluded | $1,K+1$ | $2p_1p_{K+1}$ | 0 | 0 | 0 |
| | ... | $r(x)$ | 0 | 0 | 0 |
| | $N,N$ | $p_N^2$ | 0 | 0 | 0 |
| | | | | $p_1^2+2p_1p_2+\ldots+p_K^2$ $=(p_1+p_2+\ldots+p_K)^2$ | 1 |

---



Posterior Probability
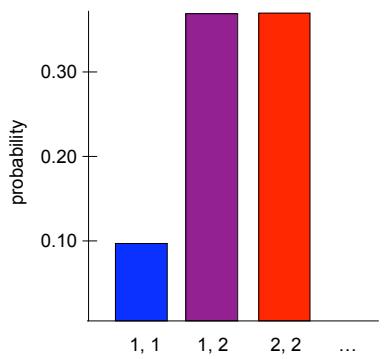
## Likelihood Ratio

$$LR = \frac{q(x)}{r(x)}$$

$$= \frac{r(x)\big/\left(p_1 + p_2 + \ldots + p_K\right)^2}{r(x)}$$

$$= \frac{1}{\left(p_1 + p_2 + \ldots + p_K\right)^2}$$

$$= \frac{1}{inclusion\ probability}$$

---

## Uncertain Genotypes

Bayes theorem updates probability

- prior probability
- likelihood function
- posterior probability

Probability concentration determines likelihood ratio

---

## Uniform population prior
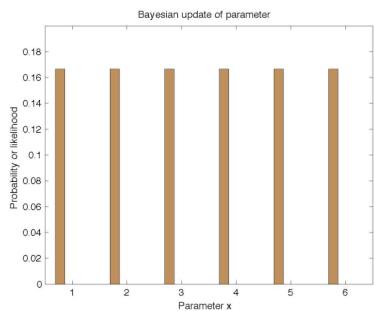
## Realistic population prior
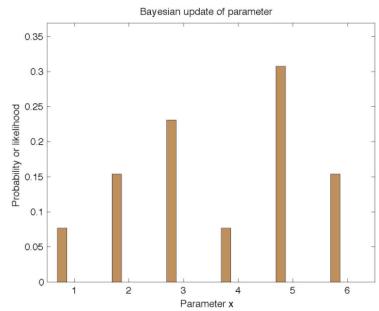


## Likelihood: all pairs are equal
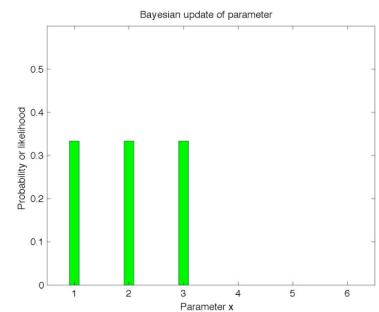


## Filter allele pair list
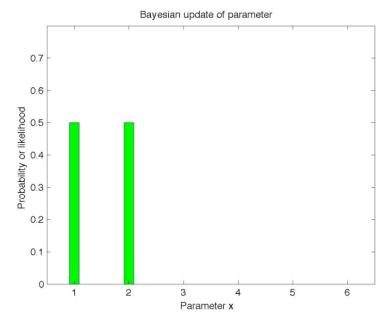
# Unequal: quantitative data

Bayesian update of parameter
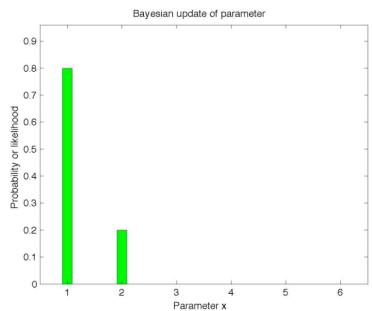
# Population prior probability

Bayesian update of parameter
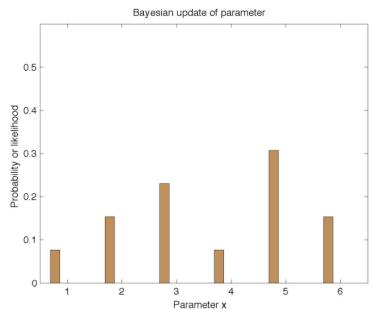
# Combine with equal listing

Bayesian update of parameter
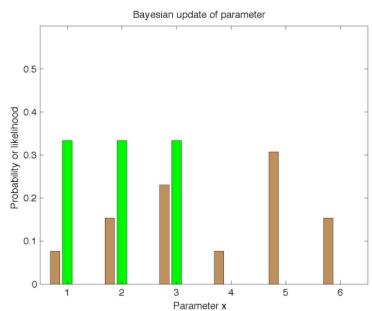
Posterior proportional to prior
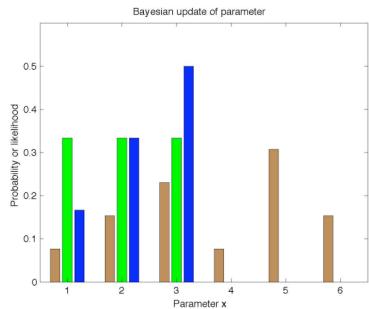


Population genotype prior
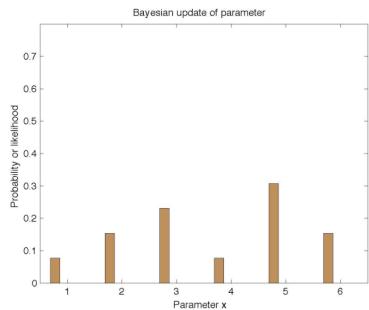


Likelihood: smaller allele pair list
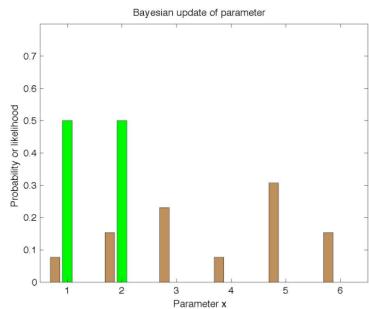
## Posterior focuses probability

Bayesian update of parameter

## Population genotype prior
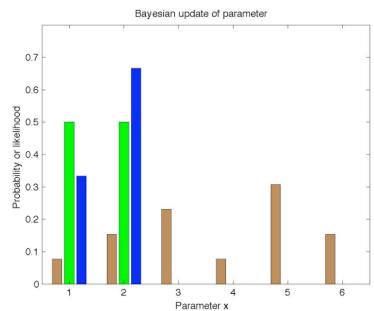
Bayesian update of parameter

## Quantitative likelihood function

Bayesian update of parameter
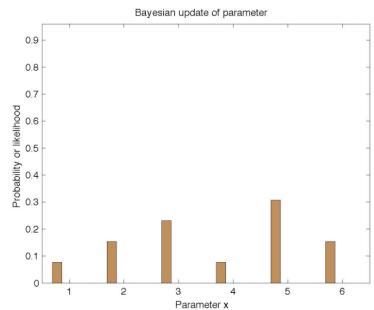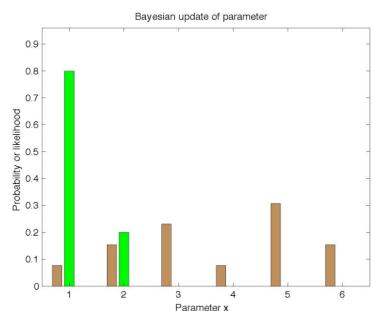
## Evidence genotype posterior

Bayesian update of parameter
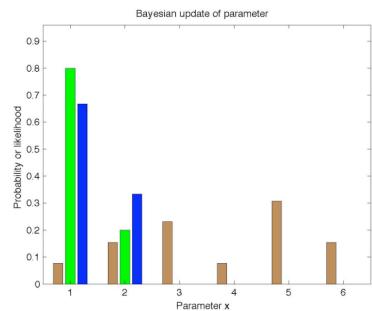


## Conclusions

- No real inclusion vs. LR controversy: same construction
- All uncertain genotypes are probability distributions
- Efficacy measured by preserving identification information
  apply log(LR) information measure to inclusion method
- Inclusion has justification for use in court, since is a LR
- PI statistic understandable via inclusion likelihood function
- Relevance of inclusion?  Depends on the data.

**Cybergenetics**