

DNA Identification: Stochastic Effects

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

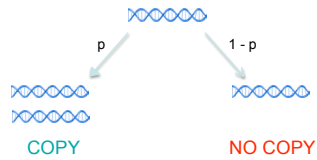
TrueAllele® Lectures
Fall, 2010



Cybergenetics

Cybergenetics © 2003-2010

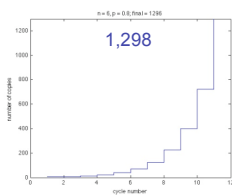
PCR is a random process



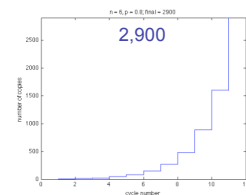
PCR efficiency is not 100% efficient. A strand copies with probability p , and doesn't copy with probability $1-p$.

STR peak is a random variable

$p = 80\%$, $n = 6$

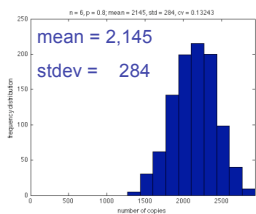


One amplification

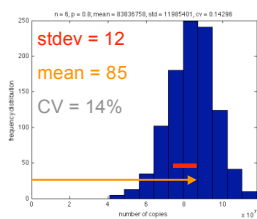


Another amplification

STR peak height measurement reflects probability distribution

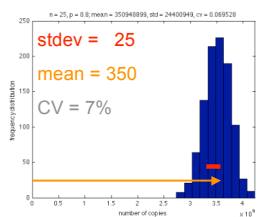


Relative peak certainty: coefficient of variation

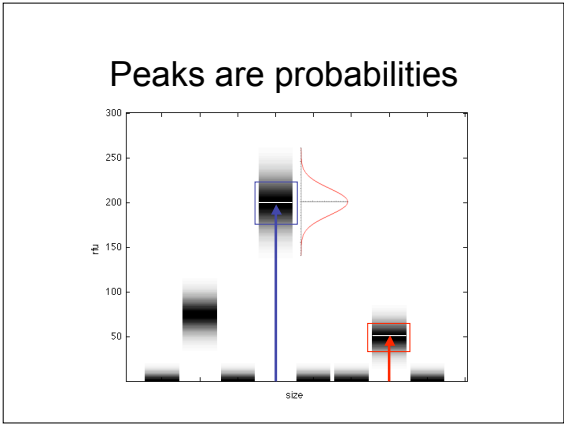


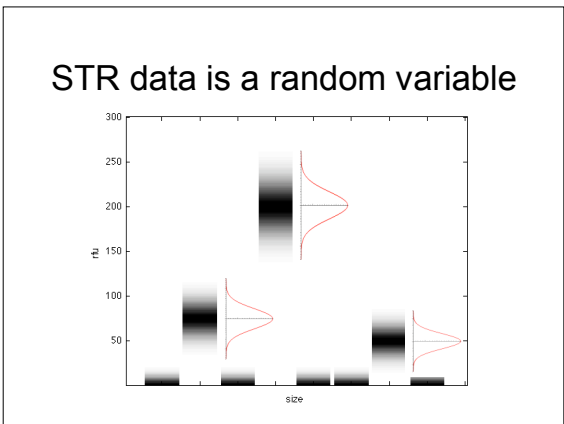
$$CV = \frac{\text{standard deviation}}{\text{mean value}}$$

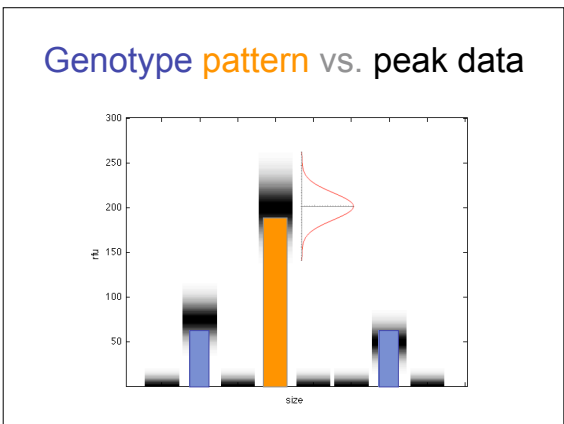
Four times the peak height, gives twice the peak certainty



$$CV = \frac{\text{standard deviation}}{\text{mean value}}$$

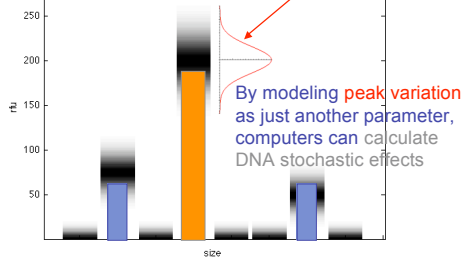




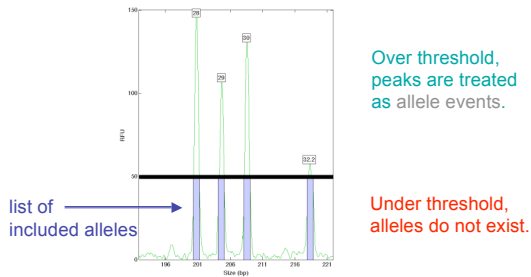


Calculate stochastic effects

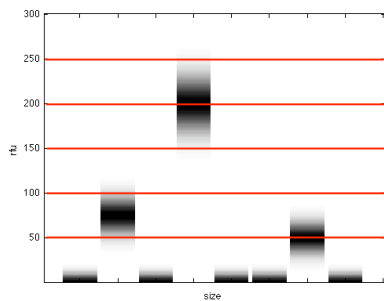
Computers can solve for genotype probabilities and other random variables, like peak variation



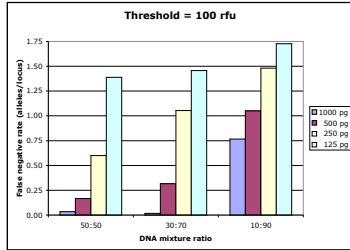
Some apply a threshold



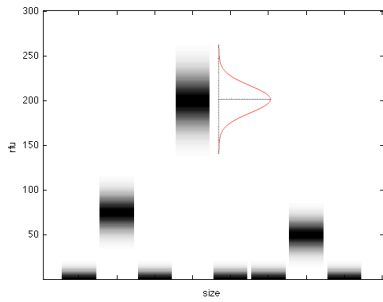
Since peaks are probabilities, thresholds introduce error



False allele exclusion rate



Probability preserves information



Data probability

- arises from PCR randomness
- models stochastic effects
- helps explain allele drop out
- compares with genotype patterns
- preserves identification information
- established normative science
