

ESWG paper challenges – A walkthrough

Athen, May 29, 2014



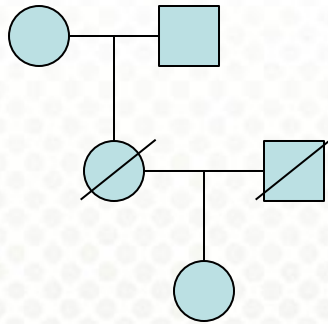
Paper challenges - Outline

- Purpose
- Walkthrough
 - 2010 – Maternal grandparents
 - 2011 – Uncle (Data mother)
 - 2012 – Uncle and aunt
 - 2013 – Duo (Paternity)
 - 2014 – Incest case
- Tips and tricks in Familias
- Recommendations

Paper challenges - Purpose

- Performed to assure the quality of the labs
- Investigate the softwares and approaches
- Illustrate how different labs approach difficult cases

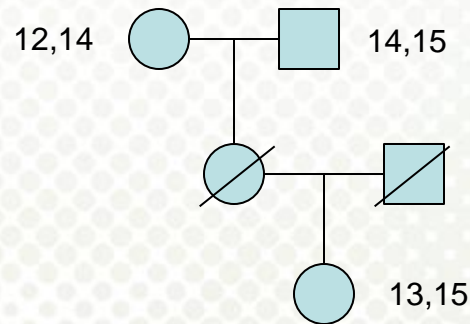
2010 – Maternal grandparents



- X-data and mt-data
- Total LR (Autosomal + mt) ranged from 16.09 to 17628798.70

2010 – Maternal grandparents

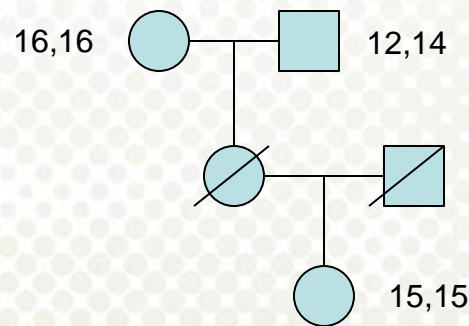
- D8S1179: LR was 1/8d and 1/4d. GM: 12,14 GF: 14,15
C:13,15



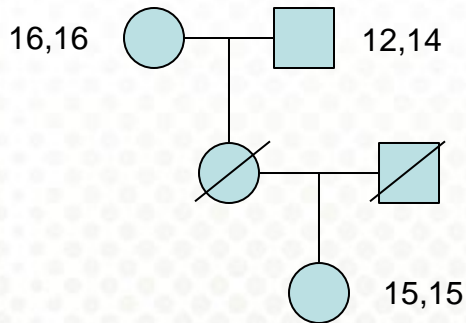
$$\begin{aligned}
 LR &= \frac{P(DNA | H_1)}{P(DNA | H_2)} = \frac{P(GF)P(GM)P(C | GF, GM)}{P(GF)P(GM)P(C)} = \frac{P(12,14)P(14,15)p(13)0.5^2}{P(12,14)P(14,15)P(13,15)} = \\
 &= \frac{0.5^2}{2p(15)} = \frac{1}{8p(15)}
 \end{aligned}$$

2010 – Maternal grandparents (cont.)

- D18S51: LR range from 0.002 to 156.588 (!). GM: 16,16
GF: 12,14 C: 15,15. (Formula extremely complicated)
- Mutation or silent allele
- High discrepancy between presented formulas

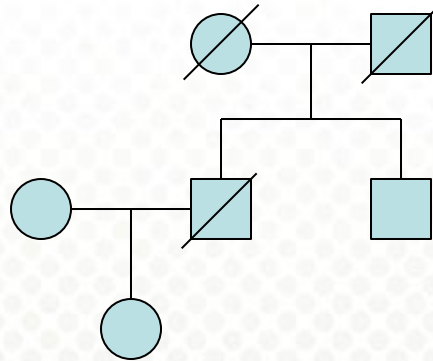


2010 – Maternal grandparents (cont.)



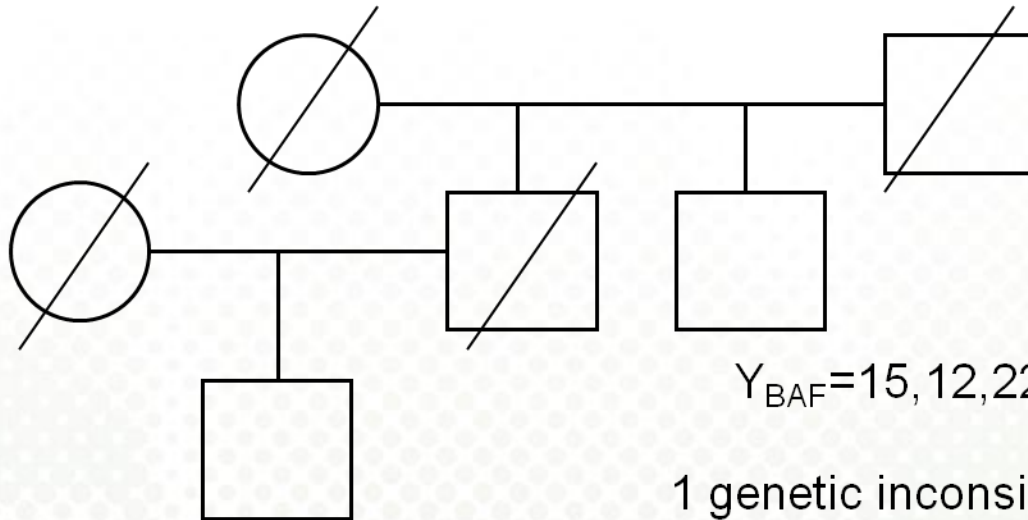
$$\begin{aligned}
 LR &= \frac{P(DNA | H_1)}{P(DNA | H_2)} = \frac{P(GF)P(GM)P(C | GF, GM)}{P(GF)P(GM)P(C)} = \\
 &= \frac{P(12,14)p(15) \left[P(16,16) \left(0.5^2 m(16 \rightarrow 15) * 2 + 0.5(0.5m(16 \rightarrow 15) * 2 + 0.5m(14 \rightarrow 15) * 2) \right) + P(16,s)0.5^2 \right]}{P(16,16)P(12,14)P(15,15)} = \\
 &= \frac{p(16)(m(16 \rightarrow 15) + 0.5m(14 \rightarrow 15)) + 2p(s)0.5^2}{p(15)p(16)} = \langle m(16 \rightarrow 15) = m(14 \rightarrow 15) \rangle = \frac{1.5p(16)m(16 \rightarrow 15) + 0.5p(s)}{p(15)p(16)} = \\
 &= \frac{3m(16 \rightarrow 15)}{2p(15)} + \frac{p(s)}{2p(15)p(16)} = \frac{3\mu}{4p(15)} + \frac{p(s)}{2p(15)p(16)}
 \end{aligned}$$

2011 - Uncle (Data mother)



- Results were pretty much the same for all autosomal markers
- Calculations fairly straightforward
 - No mutation could be observed
- Total LR ranged from 567.42 to 69786.49
- Inconsistency in Y-data

Paper challenge 2011 Y-STRs



$Y_C=15, 12, 22, 28, \dots \mathbf{11} \dots$

$\mu_{Tot}=0.006$

$\Pr(\text{DNA} \mid \text{related}) = p(Y_C) * 1 * \mu_{+1} + p(Y_{BAF}) * 2 * \mu_{-1}$

$p(Y_C) \approx (0+1)/(N+1) = 1/238$

$\Pr(\text{DNA} \mid \text{unrelated}) = p(Y_C) * p(Y_{BAF})$

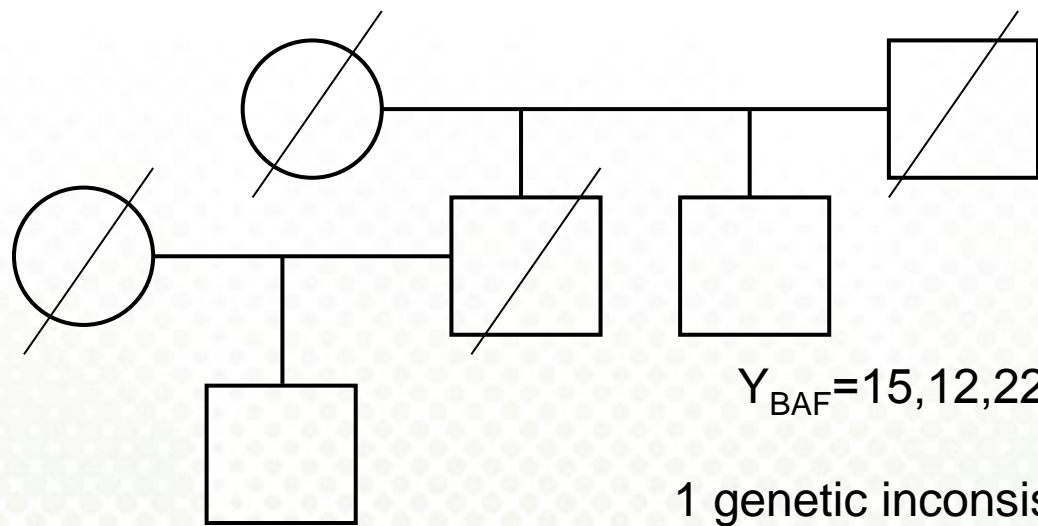
$p(Y_{BAF}) \approx (1+1)/(N+1) = 2/238$

$LR = [p(Y_C) * 1 * \mu_{+1} + p(Y_{BAF}) * 2 * \mu_{-1}] / [p(Y_C) * p(Y_{BAF})] \approx$

$[p(Y_C) * 1 * \mu_{tot} * 0.5 + p(Y_{BAF}) * 2 * \mu_{tot} * 0.5] / [p(Y_C) * p(Y_{BAF})] \approx 1.785$

$[3 * \mu_{tot} * 0.5] / [p(Y_C)] \approx 2.142$

Paper challenge 2011 Y-STRs



$Y_{BAF}=15,12,22,28,\dots 12\dots$

1 genetic inconsistency in marker DYS439

$Y_c=15,12,22,28,\dots 11\dots$

$$\mu_{Tot}=0.006$$

$$\Pr(\text{DNA} \mid \text{related})=p(Y_c)*1*\mu_{+1}+p(Y_{BAF})*2*\mu_{-1}$$

$$p(Y_c)\approx(0+1)/(N+1)=1/238$$

$$\Pr(\text{DNA} \mid \text{unrelated})=p(Y_c)*p(Y_{BAF})$$

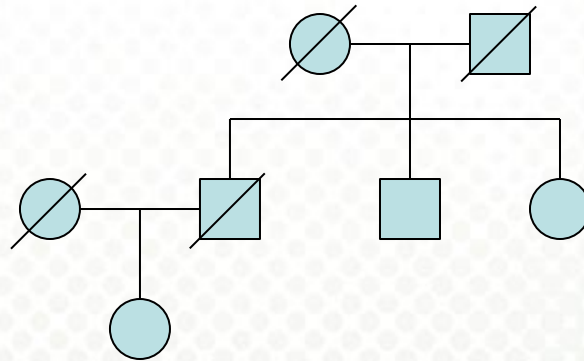
$$p(Y_{BAF})\approx(1+1)/(N+1)=2/238$$

$$LR=[p(Y_c)*1*\mu_{+1}+p(Y_{BAF})*2*\mu_{-1}]/[p(Y_c)*p(Y_{BAF})] \approx$$

$$[p(Y_c)*1*\mu_{tot}^{*0.5}+p(Y_{BAF})*2*\mu_{tot}^{*0.5}]/[p(Y_c)*p(Y_{BAF})] \approx 1.785$$

$$[3*\mu_{tot}^{*0.5}]/[p(Y_c)] \approx 2.142$$

2012 – Uncle and aunt

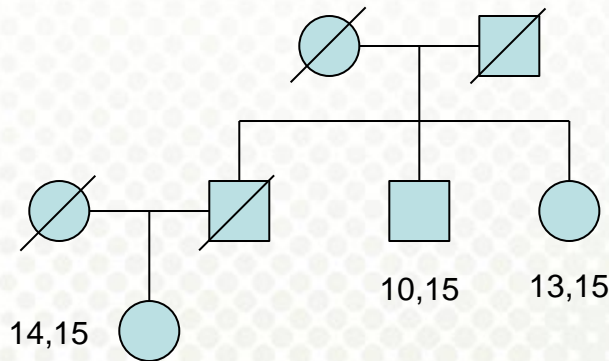


- Total LR ranged from 58.33 to 9153976.52!
- Some markers show similar LR (between labs) for all submitted formulas
 - D8S1179, D21S11 (mutation) and D18S51 (rare allele) deviates the most
 - Formulas quite large, easy to do mistakes (Confirm results with software)

2012 – Uncle and aunt

– D8S1179, C: 14,15; BAF: 10,15; SAF: 13,15

- LR ranged from 0.125 to 23.830



- We sum over possible genotypes for the grandparents

Grandmother	Grandfather
10,13	15,15
10,13	14,15
10,13	15,x
10,15	13,15

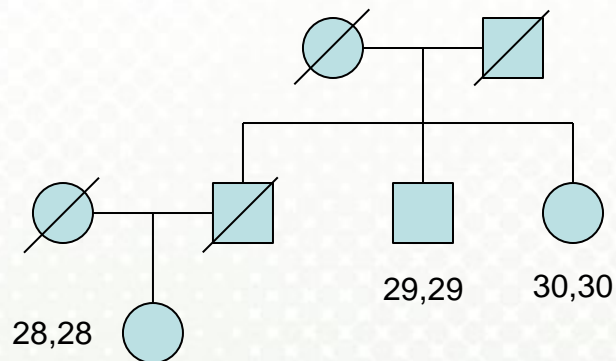
$$\begin{aligned}
 LR &= \frac{P(DNA | H_1)}{P(DNA | H_2)} = \frac{P(C | GM, GF)P(BAF | GM, GF)P(SAF | GM, GF)}{P(C)P(BAF | GM, GF)P(SAF | GM, GF)} = \\
 &= \frac{2 \left[P(10,13) \left(P(15,15)p(14)0.5^3 + P(14,15)0.5^4(0.5p(14) + 0.5p(15)) + P(15,x)p(14)0.5^5 \right) + P(10,15)P(13,15)p(14)0.5^4(0.5^2 + 0.5^2) \right]}{P(14,15)2 \left[P(10,13) \left(P(15,15)0.5^2 + P(15,x)0.5^4 \right) + P(10,15)P(13,15)0.5^4 \right]} = \\
 &= \frac{\left[\left(p(15)0.5^4 + 0.5^4(0.5p(14) + 0.5p(15)) \right) + (1 - p(14) - p(15))0.5^5 \right] + p(15)0.5^5}{2p(14)p(15) \left[\left(p(15)0.5^3 + (1 - p(15))0.5^4 \right) + p(15)0.5^4 \right]} = \frac{6p(15) + 1}{8p(15)[1 + 2p(15)]}
 \end{aligned}$$

2012 – Uncle and aunt

- D21S11, C: 28,28; BAF: 29,29; SAF: 30,30
 - LR ranged from 0.004 (u/8b) to 0.5
 - Silent allele or mutation

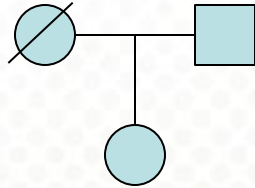
- We sum over possible genotypes for the grandparents

Grandmother	Grandfather
29,30	29,30
28,29	29,30
28,30	29,30
28,s	29,30
29,s	29,30
30,s	29,30
x,s	29,30



$$\begin{aligned}
 LR &= \frac{P(DNA | H_1)}{P(DNA | H_2)} = \frac{P(C | GM, GF)P(BAF | GM, GF)P(SAF | GM, GF)}{P(C)P(BAF | GM, GF)P(SAF | GM, GF)} = \\
 &= \frac{p(28)2 \left[P(29,30)0.5^2 \left(P(x,s)0.5^4 + P(28,s)0.5^3 + P(29,s)0.5^3 + P(30,s)0.5^3 + P(29,30)0.5^3 m(29 \rightarrow 28)[0.5^2 + 0.5^2] + P(28,29)0.5^3 m(29 \rightarrow 30) + P(28,30)0.5^3 m(30 \rightarrow 29) \right) \right]}{P(28,28) \left[P(29,30)P(29,30)0.5^4 \right]} \\
 &= \frac{0.5^4 \left[(2p(s)(1-p(30)-p(29)-p(28))+4p(28)p(s)+4p(29)p(s)+4p(30)p(s)+p(29)p(30)\mu+2p(28)p(29)\mu+2p(28)p(30)\mu) \right]}{0.5^2 p(28)p(29)p(30)} \\
 &= \frac{2p(s)+2p(28)p(s)+2p(29)p(s)+2p(30)p(s)+2p(28)p(29)\mu+2p(28)p(30)\mu+p(29)p(30)\mu}{4p(28)p(29)p(30)}
 \end{aligned}$$

2013 – Duo (Paternity)

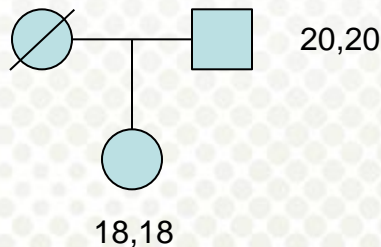


- All markers except D2S1338 shows similar LR values
- X-chromosomal data
 - With haplotypes
 - Recombination rates
 - Large deviation in the results
 - FamLinkX exercise

2013 – Duo (Paternity)

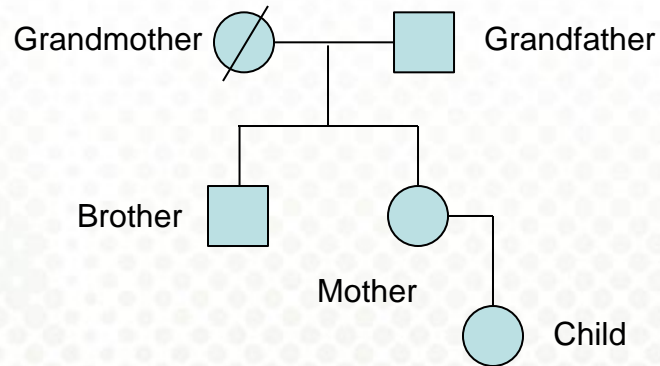
– D2S1338, AF: 20,20 C: 18,18

- Two step mutation or silent allele
- LR range from 0.000009 (μ^2) to 3.697(!)
- 8 labs submitted formulas from the same software



$$\begin{aligned}
 LR &= \frac{P(DNA | Paternity)}{P(DNA | NonPaternity)} = \frac{P(AF)P(C | AF)}{P(AF)P(C)} = \frac{P(20,20)P(18)m(20 \rightarrow 18)}{P(18,18)P(20,20)} = \\
 &= \frac{m(20 \rightarrow 18)}{p(18)} = \left\langle \begin{array}{l} \text{Next step depends} \\ \text{on the mutation model} \end{array} \right\rangle = \frac{\mu / 2 * r}{a} = \langle \text{Given } r=0.1 \rangle = \frac{\mu}{20a}
 \end{aligned}$$

2014 – Incest case



– Three hypotheses

- H1: The Grandfather is the father of the child
- H2: The Brother of the mother is the father of the child
- H3: Another man, unrelated to the other persons is the father of the child

2014 – Incest case

- Complex case though the calculations are generally easy, besides D12S391
- We compute

$$LR_1 = \frac{P(DNA | H_1)}{P(DNA | H_3)} \quad \text{and} \quad LR_2 = \frac{P(DNA | H_2)}{P(DNA | H_3)}$$

2014 – Incest case (cont.)

- Starting with LR1, we see that the LR for the individual markers are computed as if the case were a regular paternity case

$$\begin{aligned} LR_1 &= \frac{P(DNA | H_1)}{P(DNA | H_3)} = \langle \text{Simplifies to} \rangle = \frac{P(M | GF)P(GF)P(C | M, GF)}{P(M | GF)P(GF)P(C | M)} = \\ &= \frac{P(C | M, GF)}{P(C | M)} \end{aligned}$$

- The genotype of the brother is irrelevant to this LR for all the given markers
- LR ranged from $2.4E+8$ – $1E+10$
 - Cluster of 14(!) labs located at $2.74E+9$

2014 – Incest case (cont.)

- Continue with LR2, we see that the LR for the some markers are computed as if the case were a regular paternity case (Similar to LR1)

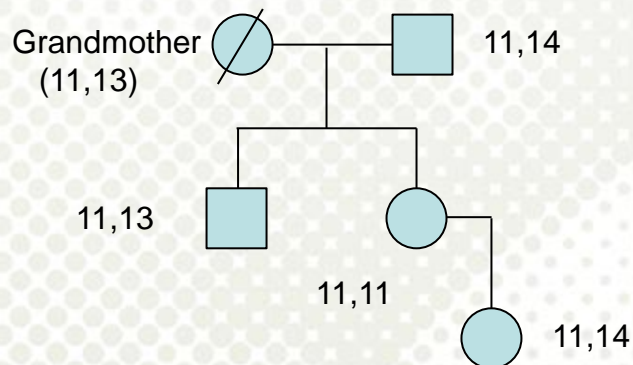
$$\begin{aligned}
 LR_2 &= \frac{P(DNA | H_2)}{P(DNA | H_3)} = \left\langle \begin{array}{l} \text{Simplifies to} \\ \text{for some markers} \end{array} \right\rangle = \\
 &= \frac{P(M | GF, GM)P(B | GF, GM)P(GF)P(C | M, B)}{P(M | GF, GM)P(B | GF, GM)P(GF)P(C | M)} = \frac{P(C | M, B)}{P(C | M)}
 \end{aligned}$$

- The genotype of the grandfather is irrelevant for these markers!
- Total LR ranged from 0.0 (?) to 4.11E+6

2014 – Incest case (cont.)

- D2S441, C: 11,14 M: 11,11 B: 11,13 and GF: 11,14
 - Mutation is needed (Silent allele not possible)
 - 6 labs got $\mu/4c$ and 6 labs got $\mu/2c$

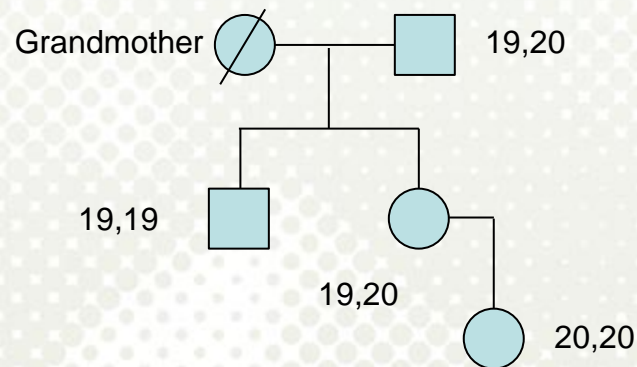
$$LR_2 = \frac{P(DNA | H_2)}{P(DNA | H_3)} = \frac{P(11,14)P(11,13)0.5^4 m(13 \rightarrow 14)0.5}{P(11,14)P(11,13)0.5^4 p(14)} = \frac{\mu}{4p(14)}$$



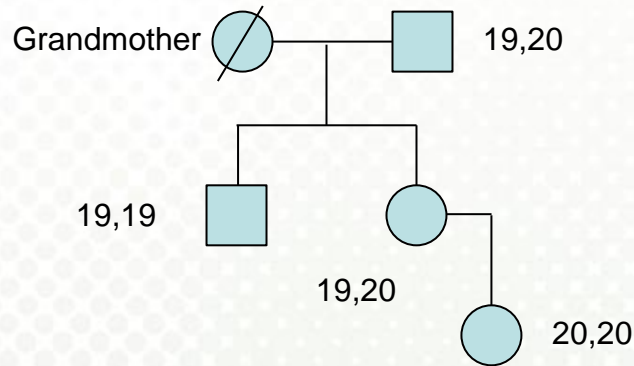
2014 – Incest case (cont.)

- D12S391, C: 20,20 M: 19,20 B: 19,19 and GF: 19,20
 - Mutation or silent allele
 - We sum over possible genotypes for the GM
 - GM: 19,20; 19,s; 20,s; 19,19 or 19,x
 - Again, extremely complicated, formula below is “simple”

Grandmother
19,20
19,s
20,s
19,19
19,x



2014 – Incest case (cont.)



$$LR_2 = \frac{P(DNA | H_2)}{P(DNA | H_3)} =$$

$$= \frac{P(19,20) \left(P(19,20)0.5^4 m(19 \rightarrow 20) + P(19,19)0.5^3 m(19 \rightarrow 20) + P(19,x)0.5^5 m(19 \rightarrow 20) + P(19,s)0.5^6 + P(20,s)0.5^6 \right)}{P(19,20) \left(p(20) \left[P(19,20)0.5^4 + P(19,19)0.5^3 + P(19,x)0.5^5 \right] \right)} =$$

$$= \frac{\left(P(19,20)0.5^5 \mu + P(19,19)0.5^4 \mu + P(19,x)0.5^6 \mu + P(19,s)0.5^6 + P(20,s)0.5^6 \right)}{\left(2p(19)p(20)0.5^5 \left[2p(20) + 2p(19) + p(x) \right] \right)} =$$

$$= \frac{0.5^6 \left[2P(19,20)\mu + 4P(19,19)\mu + P(19,x)\mu + P(19,s) + P(20,s) \right]}{0.5^5 \left(2p(19)p(20) \left[1 + p(20) + p(19) \right] \right)} =$$

$$\frac{\left[\mu p(19) \left(1 + p(20) + p(19) \right) + p(s) \left(p(19) + p(20) \right) \right]}{\left(2p(19)p(20) \left[1 + p(20) + p(19) \right] \right)} = \frac{\mu}{2p(20)} + \frac{p(s) \left(p(19) + p(20) \right)}{2 \left(p(19)p(20) \left[1 + p(20) + p(19) \right] \right)}$$

Tips and tricks in Familias

- Mutation model
 - Use Extended stepwise model in Familias 3
- Mutation range
 - Set to 0.00001 for one-step mutation only
- Fill database with alleles
 - A problem with mutation models
 - For example, observed mutation from 14 to 15, we need alleles 13, 14 and 15 (at least!).

Recommendations

- Never calculate the LR based on manual formulas alone
- Use validated softwares, such as Familias or similar softwares
- Simplifications on formulas

ESWG paper challenges – A walkthrough

Athen, May 29, 2014

