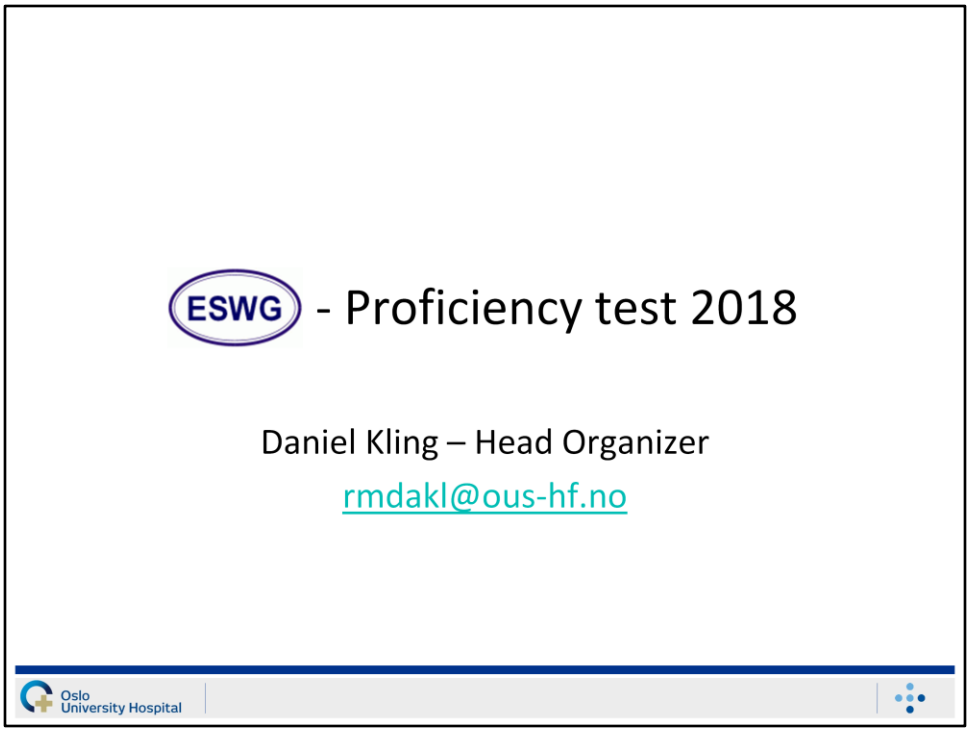
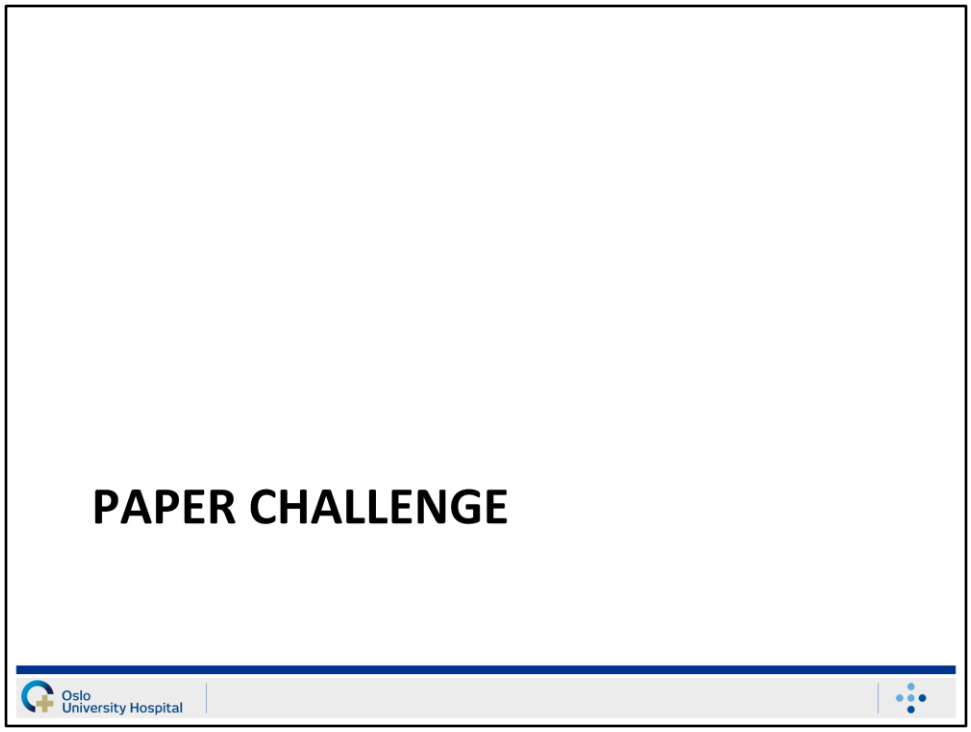


This is a brief summary of the ESWG proficiency test 2018. Complete results are contained in the Excel summary.



We start with the paper challenge.

Disclaimer

All individuals and DNA data are fictional. Any resemblence to known persons is purely coincidental.

Oslo University Hospital

Paper challenge - Background

This year's paper challenge involves an ancient tale from a realm known as Westeros. The history has it that a claimant to the throne presents himself. He proposes that he is the grandchild of the deceased king. At the time, the allegations could not be ascertained and the claimant was regarded as a false pretender.

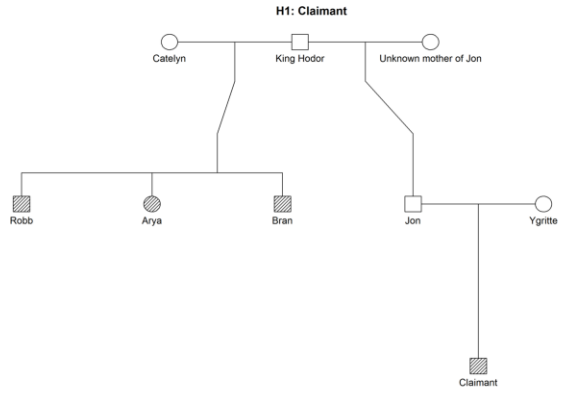
Using modern DNA techniques we are able to obtain samples and results from three known descendants of the King as well as from the Claimant.

Conduct a kinship analysis to uncover whether the Claimant is related to the three descendants of the king the way he claims. The pedigree corresponding to the claimant's allegations is depicted below. In the alternative hypothesis, the Claimant is unrelated to the descendants who are themselves still full siblings.

Oslo University Hospital

Brief description of this year's paper challenge. It involves an extended kinship case with three full siblings and an alleged paternal half nephew of those.

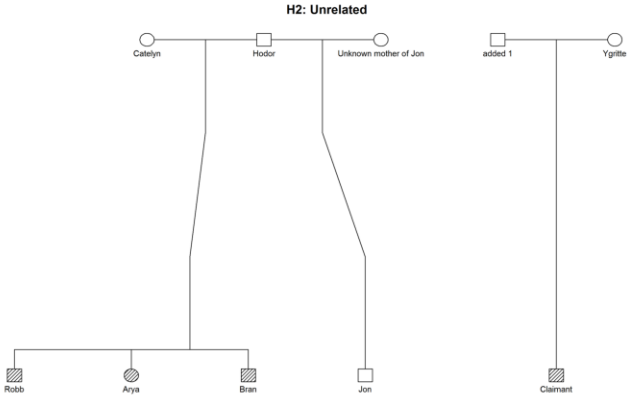
Paper challenge – Claimed relationship



Oslo University Hospital

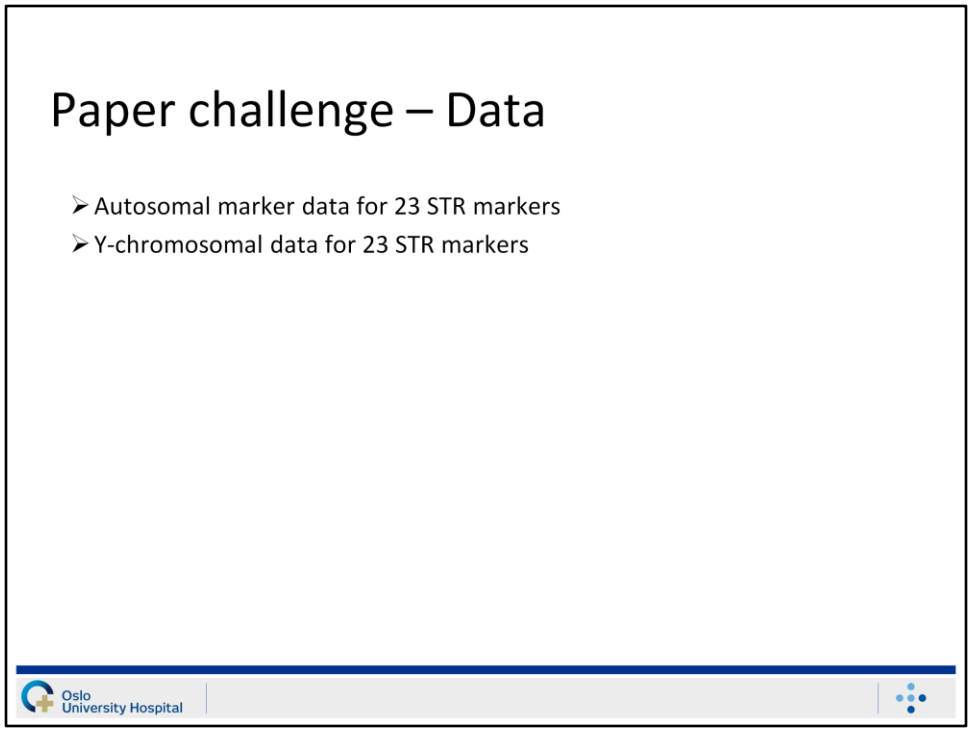
The pedigree corresponding to H1. We note that Y-chromosomal markers will be relevant. X-chromosomal markers are not relevant.

Paper challenge – Alternative



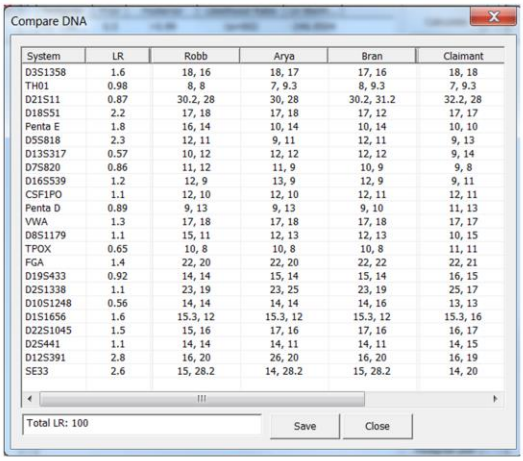
Oslo University Hospital

The pedigree corresponding to H2



Brief summary of the genetic marker data.

Paper challenge – Autosomal results



System	LR	Robb	Arya	Bran	Claimant
D3S1358	1.6	18, 16	18, 17	17, 16	18, 18
TH01	0.98	8, 8	7, 9.3	8, 9.3	7, 9.3
D21S11	0.87	30.2, 28	30, 28	30.2, 31.2	32.2, 28
D18S51	2.2	17, 18	17, 18	17, 12	17, 17
Penta E	1.8	16, 14	10, 14	10, 14	10, 10
D5S818	2.3	12, 11	9, 11	12, 11	9, 13
D13S317	0.57	10, 12	12, 12	12, 12	9, 14
D7S820	0.86	11, 12	11, 9	10, 9	9, 8
D16S539	1.2	12, 9	13, 9	12, 9	9, 11
CSF1PO	1.1	12, 10	12, 10	12, 11	12, 11
Penta D	0.89	9, 13	9, 13	9, 10	11, 13
VWA	1.3	17, 18	17, 18	17, 18	17, 17
D8S1179	1.1	15, 11	12, 13	12, 13	10, 15
TPOX	0.65	10, 8	10, 8	10, 8	11, 11
FGA	1.4	22, 20	22, 20	22, 22	22, 21
D19S433	0.92	14, 14	15, 14	15, 14	16, 15
D2S1338	1.1	23, 19	23, 25	23, 19	25, 17
D10S1248	0.56	14, 14	14, 14	14, 16	13, 13
D1S1656	1.6	15.3, 12	15.3, 12	15.3, 12	15.3, 16
D22S1045	1.5	15, 16	17, 16	17, 16	16, 17
D2S441	1.1	14, 14	14, 11	14, 11	14, 15
D12S391	2.8	16, 20	26, 20	16, 20	16, 19
SE33	2.6	15, 28.2	14, 28.2	15, 28.2	14, 20

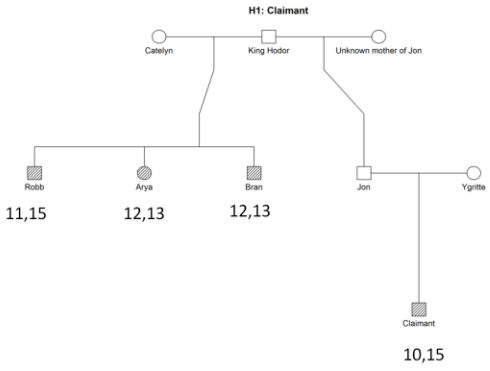
LR: 100
LR: 20-30 if you remove linked
LR: 130 if you account for linked

Oslo University Hospital

Result from calculations of the likelihod ratios (LRs). Most labs got similar results with a median around 100. Some labs reported LRs in the range 20-30, where one of the markers in the pair vWA-D12S391 had been removed due to linkage. Some labs reported LRs of about 130 where linkage had been accounted for in the indicated pair of markers. Overall, very consistent results.

Paper challenge – Autosomal results

Manual calculations for a single marker, D8S1179



Founder genotypes

Hodor	Catelyn
11,12	13,15
12,15	11,13
11,13	12,15
13,15	11,12

Oslo University Hospital

The following slides present manual calculations for one of the markers, D8S1179. We start by enumerating all possible genotypes for the relevant founders, in this case Catelyn and Hodor. Given the data, we see that there are four different combinations we need to consider.

Paper challenge – Autosomal results

Manual calculations for a single marker, D8S1179

Hodor	Catelyn
11,12	13,15
12,15	11,13
11,13	12,15
13,15	11,12

Founder probabilities

IBD probability between grandfather and grandson

$$Likelihod(data|H1) = 2p_{11}p_{12} \cdot 2p_{13}p_{15} \cdot 0.5^6 \cdot (0.5 \cdot 2p_{10}p_{15}) + \cdots$$

Transmission probabilities to descendants

$$Likelihod(data|H2) = 2p_{10}p_{15} \cdot (2p_{11}p_{12} \cdot 2p_{13}p_{15} \cdot 0.5^6 + \cdots + 2p_{13}p_{15} \cdot 2p_{11}p_{12} \cdot 0.5^6)$$

Oslo University Hospital

The likelihood for H1 will be a summation over the possible founder genotypes (i.e. Hodor's and Catelyn's) and the transmission probabilities to their known children, which in this case is 0.5^6. Finally we compute the so called identical by descent (IBD) probabilities between Hodor and the Claimant. For all genotypes where Hodor and the Claimant do not share any alleles the probability simply becomes 0.5*p(genotype of claimant), because there is a 0.5 probability that grandparent and grandchild share zero alleles. For genotypes where Hodor and the Claimant share at one allele the probability becomes 0.5*p(genotype of claimant)+0.5*p(allele not shared), where that first part is explained as above and the second relates to the fact that grandparent and grandchild shares one allele IBD with probability 0.5 and the allele that is not shared must be from the population.

The likelihood for H2 contains the same founder genotypes

but the Claimant’s genotype is now unrelated to those.

Paper challenge – Autosomal results

Manual calculations for a single marker, D8S1179

Hodor	Catelyn
11,12	13,15
12,15	11,13
11,13	12,15
13,15	11,12

$$LR = \frac{0.5^6 \cdot (2p_{11}p_{12} \cdot 2p_{13}p_{15} \cdot 0.5 \cdot 2p_{10}p_{15} + \cdots + 2p_{13}p_{15} \cdot 2p_{11}p_{12} \cdot (0.5 \cdot 2p_{10}p_{15} + 0.5 \cdot 0.5p_{10}))}{{p_8}^2 \cdot 0.5^6 \cdot (2 \cdot 2p_{17}p_{18} \cdot 2p_{16}p_{17} + 2 \cdot 2p_{17}p_{18} \cdot 2p_{16}p_{18})}$$

$$= \frac{0.5 \cdot 2p_{10}p_{15} + 0.5 \cdot 2p_{10}p_{15} + (0.5 \cdot 2p_{10}p_{15} + 0.5 \cdot 0.5p_{10}) + (0.5 \cdot 2p_{10}p_{15} + 0.5 \cdot 0.5p_{10})}{4 \cdot 2p_{10}p_{15}}$$

$$= \frac{2p_{15} + 2p_{15} + (2p_{15} + 0.5) + (2p_{15} + 0.5)}{16p_{15}} = \frac{8p_{15} + 1}{16p_{15}} = \frac{8 \cdot 0.1 + 1}{16 \cdot 0.1} = 1.125$$

Oslo University Hospital

Summarizing and simplifying.

Paper challenge – Y results

Y STRs	Robb	Arya	Bran	Claimant
DYS576	18		18	18
DYS389I	13		13	13
DYS448	19		19	19
DYS389II	29		29	29
DYS19	14		14	14
DYS391	10		10	10
DYS481	21		21	21
DYS549	12		12	12
DYS533	13		13	13
DYS438	12		12	12
DYS437	14		14	14
DYS570	17		17	18
DYS635	23		23	23
DYS390	24		24	24
DYS439	11		11	11
DYS392	13		13	13
DYS643	10		10	10
DYS393	13		13	13
DYS458	18		18	18
DYS385	11,15		11,15	11,15
DYS456	16		16	16
YGATAH4	12		12	12

Inconcistency

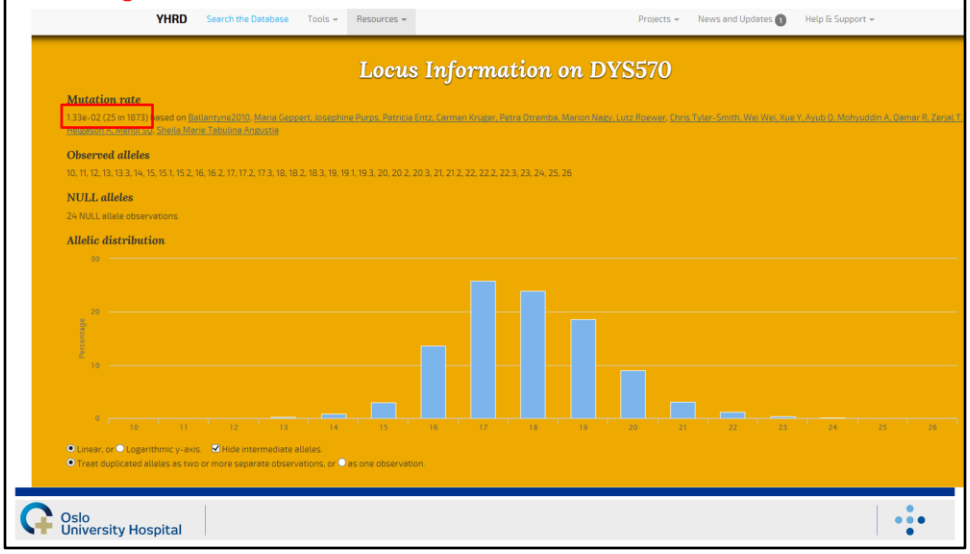
$H_{R,B}$ H_C

Oslo University Hospital

The Y-results contained an inconsistency at DYS570. Not recognized by all labs. Essentially we observe two haplotypes, denoted HR,B and HC.

Paper challenge – Y results

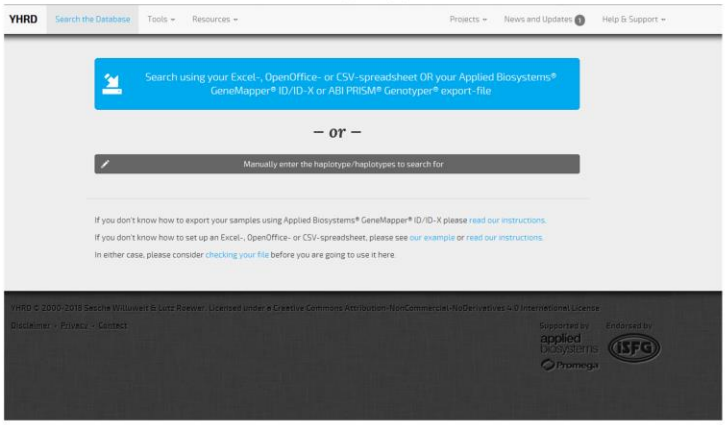
High mutation rate!



We can find the estimated mutation rate for our locus at YHRD. The overall mutation rate is >1% (0.01).

Paper challenge – Y results

Let's search YHRD to find if haplotypes have been observed!

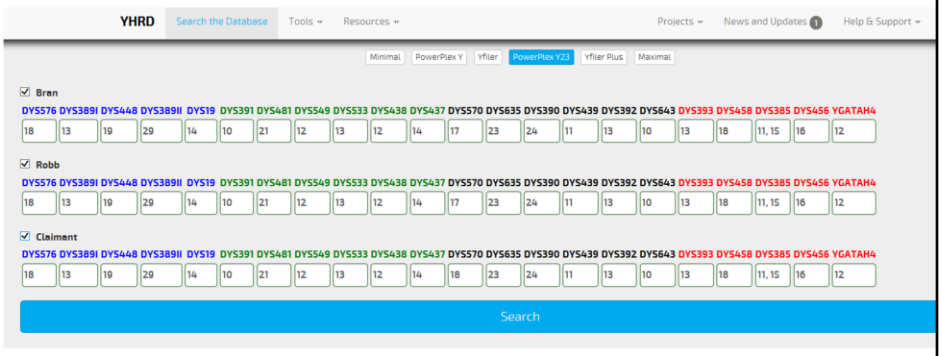


Oslo University Hospital

The data was prepared for upload to YHRD.

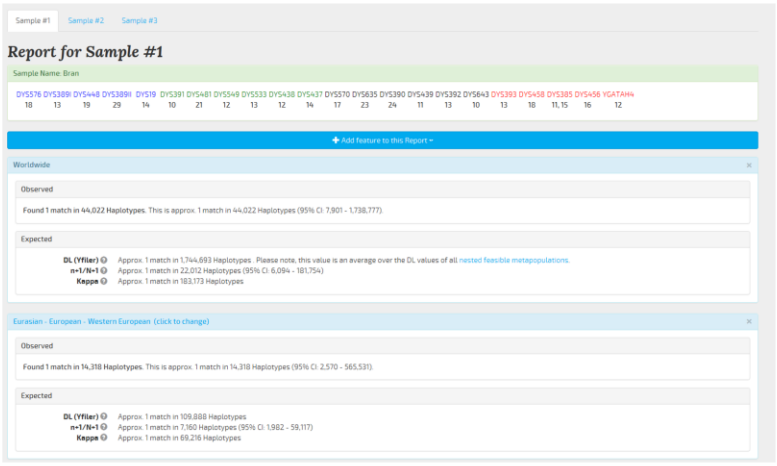
Paper challenge – Y results

Let's search YHRD to find if haplotypes have been observed!



We find the three haplotypes and note that Bran and Robb have identical while the Claimant has another one.

Paper challenge – Y results

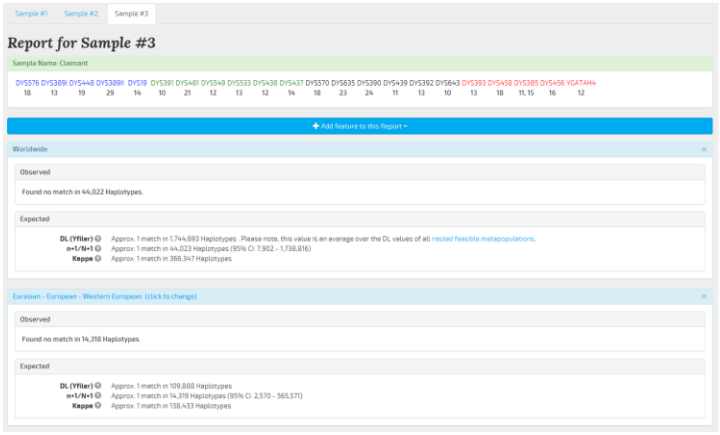


1 match in Western European metapopulation for $H_{R,B}$

Oslo University Hospital

The first haplotype as one match in the metapopulation we are interested in. Different frequency models are given. There is no right or wrong frequency, but we note that the resulting frequency estimates will be very different.

Paper challenge – Y results

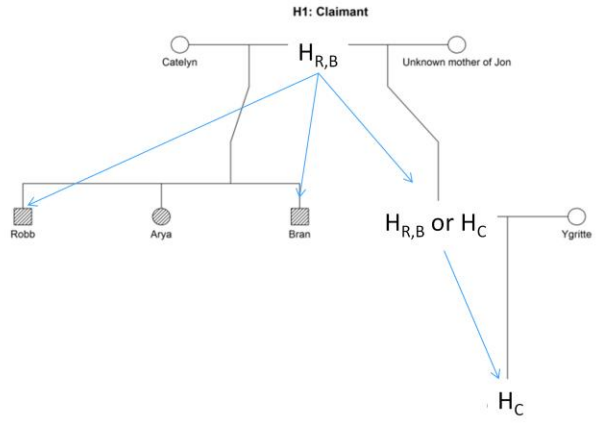


0 matches in Western European metapopulation for H_C

Oslo University Hospital

The first haplotype as no matches in the metapopulation we are interested in. Different frequency estimates are given.

Paper challenge – Y results



Oslo University Hospital

Illustrating one possible scenario given H1. The grandpaternal haplotype is HR,B and has been transmitted to Robb and Bran without mutations, while it has been subject to a mutation in the transmission to the Claimant.

Paper challenge – Y results

Let's get down to formulas!

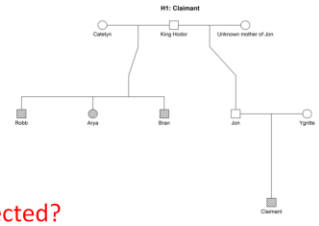
No mutation to Robb and Bran

Probability of $H_{B,R}$

$$Likelihood(Data|H_1) = [(1-\mu)^2 \cdot (2 \cdot \mu \cdot 0.5(1-\mu)) \cdot p_{H_{B,R}}] + [(1-\mu)^2 \cdot (\mu \cdot 0.5)^2 \cdot p_{H_C}]$$

Single mutation to Claimant

$$Likelihood(Data|H_2) = (1-\mu)^2 p_{H_{B,R}} \times p_{H_C}$$



Can be neglected?

$$LR = \frac{[(1-\mu)^2 \cdot (\mu(1-\mu)) \cdot p_{H_{B,R}}] + [(1-\mu)^2 \cdot (\mu \cdot 0.5)^2 \cdot p_{H_C}]}{(1-\mu)^2 p_{H_{B,R}} \times p_{H_C}} \approx \frac{\mu(1-\mu)}{p_{H_C}}$$

Oslo University Hospital

This is a suggested way to compute the likelihood ratio in this case. The method described here has been called the Grandparent-centric approach and is the most logic one. YHRD allows you to compute the kinship. We then have to assume one of the haplotypes to be «upstream» or ancestral and the second one to be «downstream» or offspring. The YHRD method does not (as far as we are concerned) allow for more than two profiles, but will work with the previously stated adaptions.

Explanation of the formulas: We will only consider mutations for the marker where we have actually seen a mutation. The likelihood for H1 has two parts, first we consider the grandpaternal haplotype to be HR,B. There must then have been either a mutation from Hodor to Jon or from Jon to the Claimant. And since we consider the simplication that mutation only occurs as single-steps we get 2 * (mu*0.5*(1-mu)). The second part relates to HC as the grandpaternal

haplotype. This, however, requires two mutation to the brothers and is therefore unlikely. The likelihood for H2 is simple. The HB,R haplotype is shared by the two known brothers and no mutation has occurred in the transmission from Hodor. Finally the LR may be simplified if we remove the second part in the numerator relating to the double mutation probability.

Paper challenge – Y results

We get LR=180 when using p_{H_C}=1/14,000. The frequency of $H_{R,B}$ does not affect results.

If we combine the automsomal and the Y LR:s we get

$LR_{combined}=100 \cdot 180=18{,}000$

A range of different LRs depending on frequency of haplotype!

Oslo University Hospital

Using the results we get approximately LR=180 (or LR=163 on YHRD) if we use the (n+1)/(N+1) method to estimate the haplotype frequency. If we combine the LR from the autosomal markers with the Y-results we obtain a combined LR of about 18,000.

The final results from the labs depended, chiefly, on the frequency estimate for the Y-chromosomal haplotype.

Paper challenge – Simulations

Following the results, your lab decides to perform a simulation study to evaluate whether including further relatives could improve the results. We may assume we have already analyzed all available genetic markers. Evaluate the following scenarios,

a) Exhumation of the mother of the descendants (Catelyn).
b) Exhumation of the mother of the Claimant (Ygritte).
c) Exhumation of the King (Hodor).
d) Exhumation of all the above mentioned individuals.

Report the median of your simulations as well as the probability that the LR will exceed 1000 for each given scenario (assuming the Claimant is related to the descendants as proposed). Perform at least 1000 simulations.

What is your conclusion regarding the four possible scenarios? For instance, can we reach a more certain conclusion by only exhuming one extra individual (a-c) or do we need to exhume all three of them (d).

Oslo University Hospital

The second part of the paper challenge included simulations as a mean to find out if exhumation should be conducted.

We were faced with four different alternatives, exhuming Catelyn, Ygritte, Hodor or all of them. Labs were supposed to report the median from the simulations as well the probability that we could obtain a LR>1000 in each alternative.

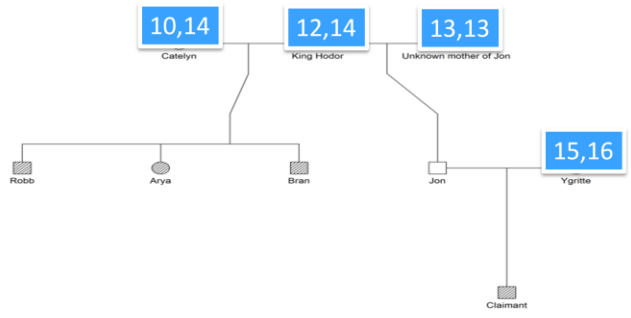
Paper challenge – Simulations

Simulations work by considering the pedigrees (H1 and H2); It starts by randomly drawing genotypes for the founders of the pedigree and subsequently randomly «creating» the children. In our case we were only interested in H1, i.e. in essence the true positive rate. We would like to simulate 1000 (or more) of H1 and count the number of times the LR exceeds 1000.

The next slides try to illustrate how a single simulation is conducted.

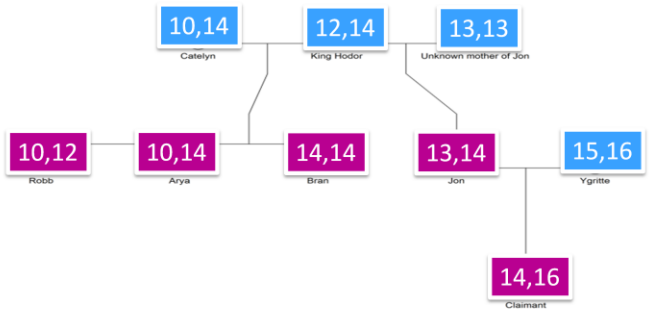
Paper challenge – Simulations

1. Simulate founders (randomly draw genotypes from the population frequencies). Here illustrated for a single marker.



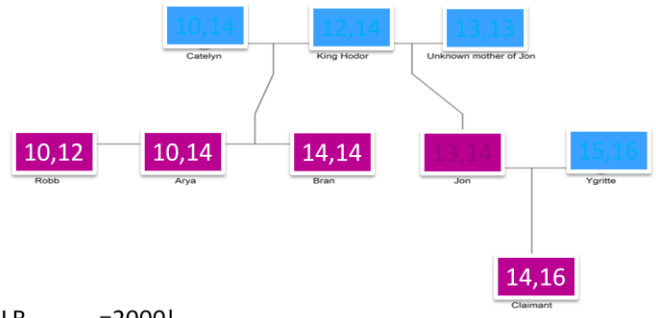
Paper challenge – Simulations

2. Simulate non-founders. Using laws of inheritance we know that there is a 50/50 chance for parents to transmit either of their alleles. This is basically just as flipping a coin and deciding what allele to pick for each child.



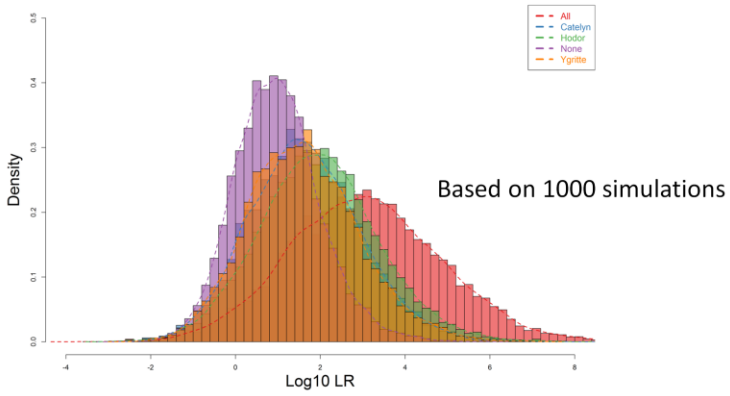
Paper challenge – Simulations

3. Finally, we compute the LR based on a subset of persons, starting with the once we have, i.e. Robb, Arya, Bran and the Claimant. We then add the exhumed persons one by one. We repeat steps 1-3 for all genetic markers and get a combined LR.



$LR_{combined}$=2000!

Paper challenge – Simulations



Interpretation: The distributions overlap, there is most information in exhuming all and, as expected, least information if none is exhumed.

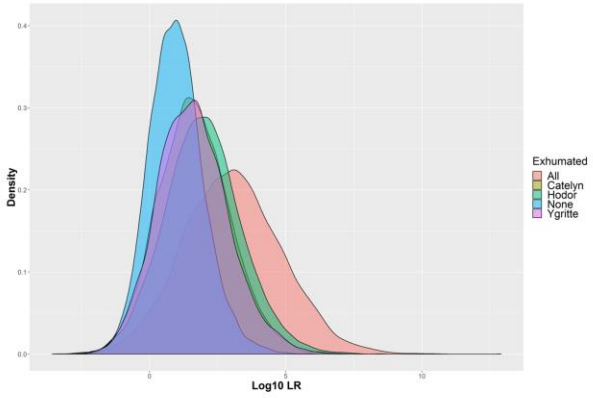
Oslo University Hospital

The results may be visualized in what is called a kernel density plot combined with histograms. Histograms provide an easy way to illustrate how many LRs falls within a certain range.

A simulation is where we randomly draw the genotypes for the founders (Hodor, Catelyn and Ygritte) and then randomly draw alleles from their genotypes and transmit them to their children (gene dropping). At the end we only retain the genotypes from those genotyped (different depending on the alternatives we consider) and compute the LR. Finally we can summarize the 1,000 simulations (or more) and create plots. The distribution plot above represent one way where we see that adding all individuals will shift the distribution to the right thus indicating an increase in information. Next slide present an easier way to understand/interpret these distributions.

Data and files to reproduce the plots are available at
http://familias.name/eswg2018/paper_challenge.zip

Paper challenge – Simulations



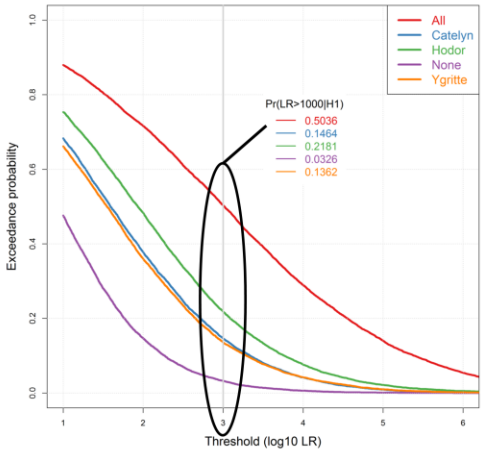
Alternative representation (NB! Different colors)

Oslo University Hospital

The results may be visualized in what is called a kernel density plot. Data and files to reproduce the plots are available at http://familias.name/eswg2018/paper_challenge.zip

A simulation is where we randomly draw the genotypes for the founders (Hodor, Catelyn and Ygritte) and then randomly draw alleles from their genotypes and transmit them to their children (gene dropping). At the end we only retain the genotypes from those genotyped (different depending on the alternatives we consider) and compute the LR. Finally we can summarize the 1,000 simulations (or more) and create plots. The distribution plot above represent one way where we see that adding all individuals will shift the distribution to the right thus indicating an increase in information. Next slide present an easier way to understand/interpret these distributions.

Paper challenge – Simulations



16 labs did the second part!
10 labs got the results presented here
6 labs got the same deviating results

Interpretation: There is 50% chance to obtain LR>1000 if we exhume all

Oslo University Hospital

The results from the simulation are illustrated in an exceedance plot. Briefly, for each given scenario/alternative we have computed 1,000 simulation. For each set of 1,000 simulations we compute the probability that the LR will exceed a threshold (sliding on the x-axis). If we specifically look at 3 on the x-axis (which is on a logarithmic scale and thus indicates LR=1000) we can see where the different lines intersect the grey vertical line. The y-values indicate the probability that we will get a LR>1000 if we exhume the different relatives. Essentially this tells us that there is a 50% chance of getting a LR>1000 if we exhume all while only about 13-20% if exhume only one of them and about 3% if none are exhumed.

Paper challenge – Summary

- Complex kinship case.
- Combining autosomal and Y marker results.
- Accessing YHRD.
- Constructing pedigrees.
- Consistent results.
- Simulations.

Oslo University Hospital

Summary of the paper challenge

PROFICIENCY TEST

Oslo University Hospital

Summary

- 47 participants
  - 36 did paper challenge
  - 43 did wet exercise
- 13 labs do sequencing
- 33 labs use Familias, 16 use Excel, 4 DNA-view
- Consistent results on the paper challenge (Autosomal part)
- More diverging results on the Y markers
  - Some labs did not realize the inconsistency
- Wet exercise results – Some typing differences have been addressed. Some labs report data also for SNP markers

Oslo University Hospital

A short summary of some interesting points from the questionnaire and the exercise.

Wet exercise

Your local police are investigating a case in which a young woman, Laura Dean, was found dead in her apartment. There are no signs of violence. Later, and in the same area, a lifeless baby is found in a container wrapped in a towel identical to towels in Laura Dean's apartment.

The police suspect that there is a connection between the two cases and ask for a DNA-test of the two bodies in order to decide whether Laura Dean is the mother of the child.

a) Perform a maternity test to clarify the maternal relationship between Laura Dean and the child (the samples are labelled Woman and Child).

Oslo University Hospital

The first part of the wet exercise was a simple maternity case. All labs reported LRs in favor of maternity. Labs using only 15-16 markers reported LRs about 10^4 to 10^6. Labs typing more than 20 markers reported LRs above 10^6 except one lab.

Wet exercise

Given the DNA results from the DNA- tests, the police conclude that Laura Dean is the biological mother of the child. During the investigation, the police are approached by a woman claiming that her son, vanished since the case became public, very likely is the father of the baby. The police decide to test the claimed relation between this woman and the child.

b) Perform a test to clarify the claimed biological relationship between this woman (sample labelled Grandmother) and the child. Consider the maternity as confirmed and include the DNA-data from mother and child in task a).

Report the likelihood ratios (LR) for the individual genetic markers included in the tests as well as the combined LR. State which frequency database you have used for the calculations. Part b) is optional while a) is mandatory to obtain the certificate.

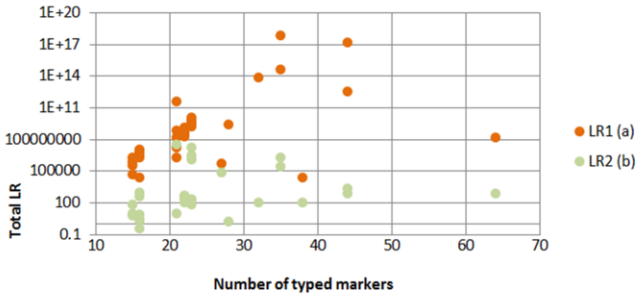
Oslo University Hospital

The second part involved a deficient paternity case where the alleged paternal grandmother was presented. Most labs reported low LRs (10^0-10^3) while a few reported higher (10^7-10^8). See next slide for correlation between the number of markers used and the LRs in both parts of the wet exercise.

Wet exercise

Samples consisted of diluted blood (EDTA) on FTA cards. Two labs reported problems with amplications and received undiluted blood instead.

a) A plain maternity test (43 labs completed)
b) A deficient paternity case (alleged grandmother) (39 labs completed)

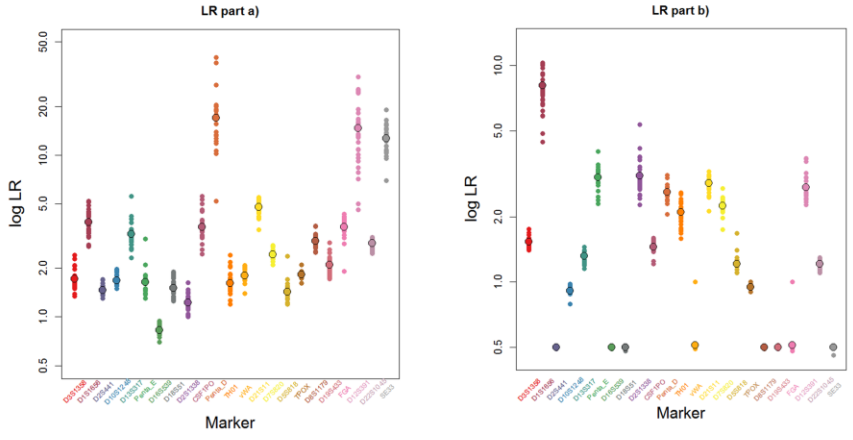


Oslo University Hospital

Figure visualizing the spread in LRs for the two different parts of the wet exercise. One lab used >150 markers and was not included in this graph, see Excel summary for complete information.

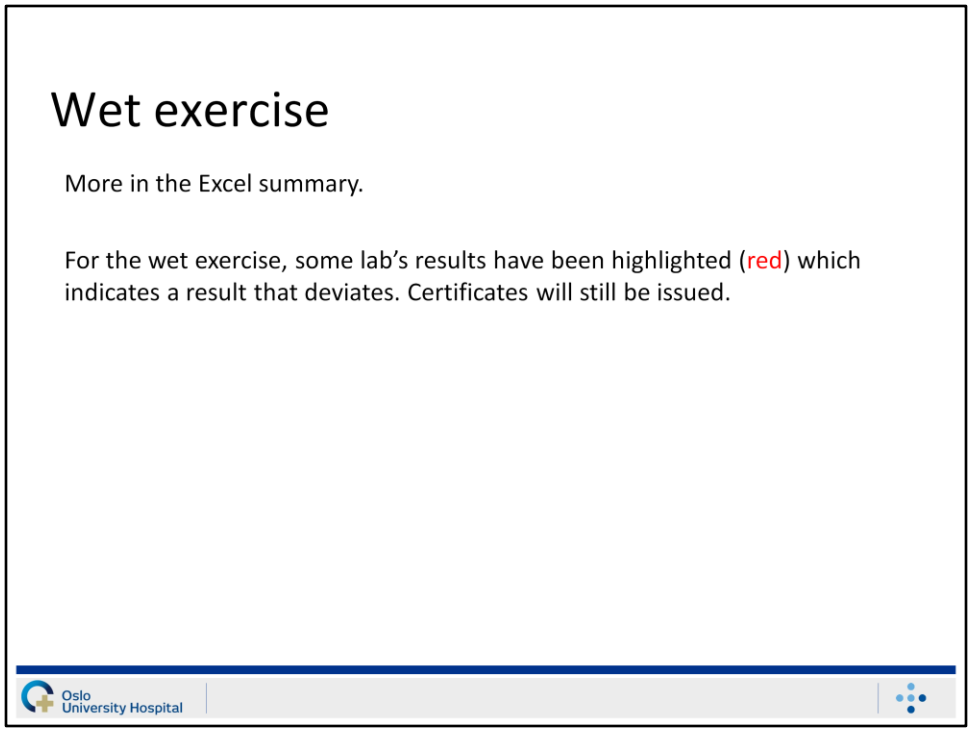
Wet exercise

Per marker results

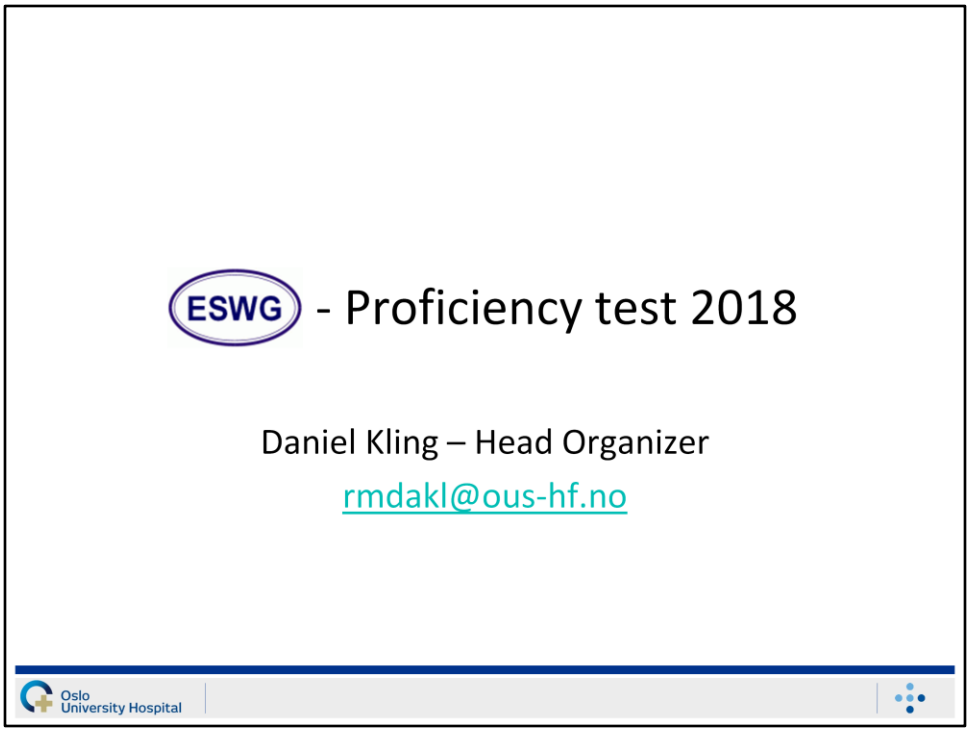


Oslo University Hospital

Plot illustrating results from the wet exercise for a few of the commonly typed STR markers. The plot illustrates the individual results from each lab (logarithmic scale on the y-axis) and the variation as well as the mean.



The final Excel report summarizes the results. The individual lab may find the information and summarize it as they like.



Feel free to ask any questions!