

# Session 6: advanced genome profiling



Kamil S. Jaron

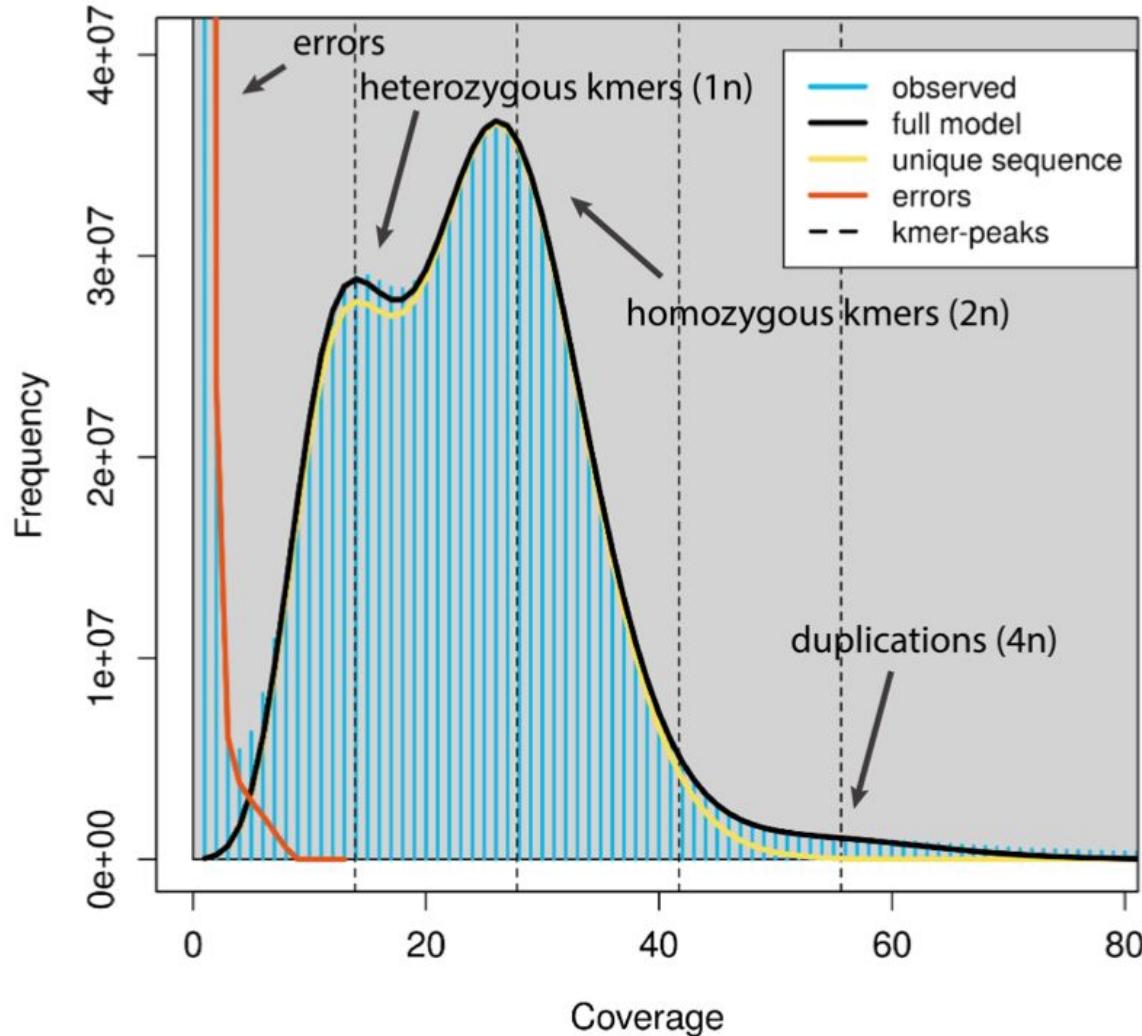
4th June 2025

K-mer workshop for biodiversity genomics						
Time	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday
08:45 - 09:00						
09:00 - 10:00	Registration and Lunch (Measuring)	Lunch	Lunch	Lunch	Lunch	Lunch
10:00 - 10:30						
10:30 - 11:00	Introduction of the K-mer workshop & Overview of the Tree of Life					
11:00 - 11:30						
11:30 - 12:30	Hands-on for Practical Profiling	Lecture on K-mer Profiling and its applications	Hands-on for K-mer Profiling	Hands-on for K-mer Profiling	Hands-on for QC and Data Processing	Work on Headline Genomes
12:30 - 13:30	Registration and Lunch (Measuring)	Lunch	Lunch	Lunch	Lunch	Lunch
13:30 - 14:30						
14:30 - 15:00	Hands-on for Practical Profiling	Lecture on K-mer Profiling and its applications	Hands-on for K-mer Profiling	Hands-on for K-mer Profiling	Hands-on for QC and Data Processing	Work on Headline Genomes
15:00 - 15:30	Break	Break	Break	Break	Break	Break
15:30 - 16:00						
16:00 - 16:30	History and overview of the Tree of Life					
16:30 - 17:00	Break					
17:00 - 18:00	Hands-on for Practical Profiling and its applications of the Tree of Life	Hands-on for Practical Profiling and its applications of the Tree of Life	Hands-on for Practical Profiling and its applications of the Tree of Life	Hands-on for Practical Profiling and its applications of the Tree of Life	Hands-on for Practical Profiling and its applications of the Tree of Life	Hands-on for Practical Profiling and its applications of the Tree of Life
18:00 - 19:00						
19:00 - 20:00	Speed networking					
20:00 - 21:00						
21:00 - 22:00						
22:00 - 23:00						
23:00 - 24:00						

# Reminder

K-mer spectra allow...

- visual inspection of sequencing runs
- Fitting of genome models

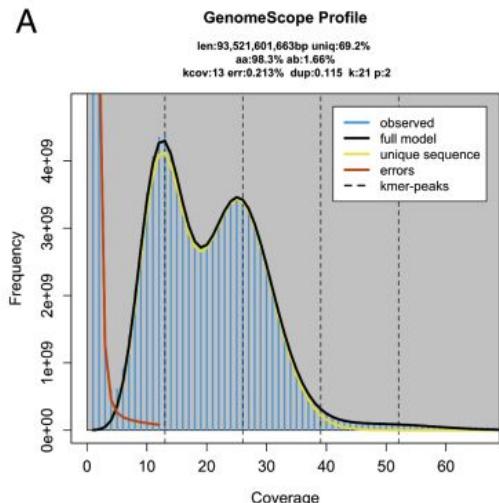


# Reminder

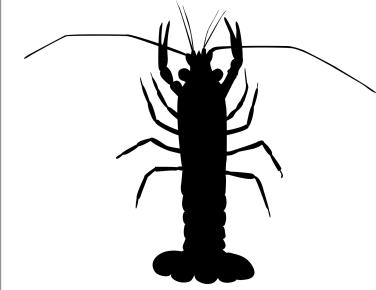
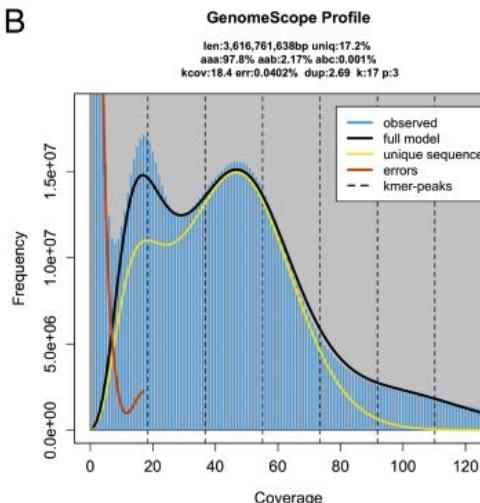
## Coverage patterns



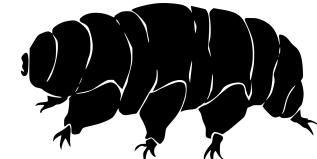
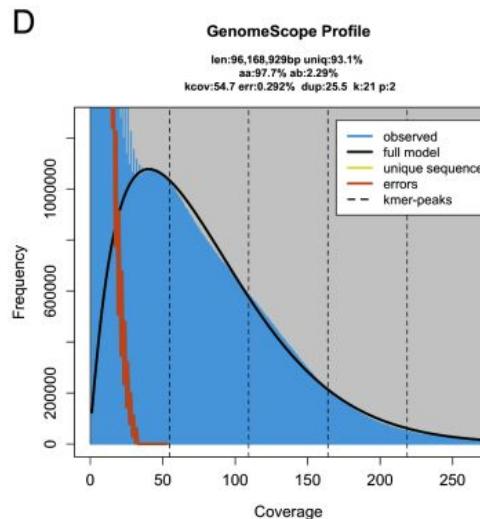
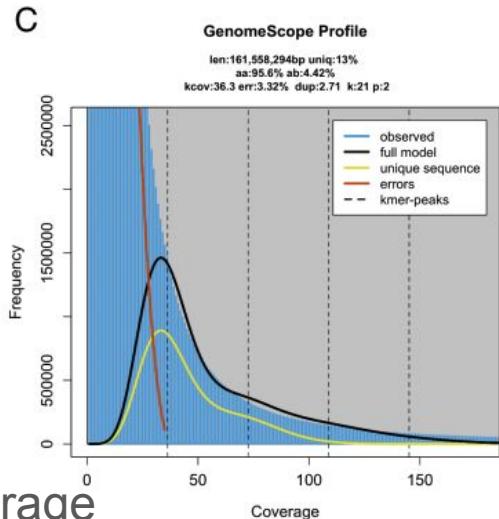
Great



Not enough coverage



Perhaps ok?

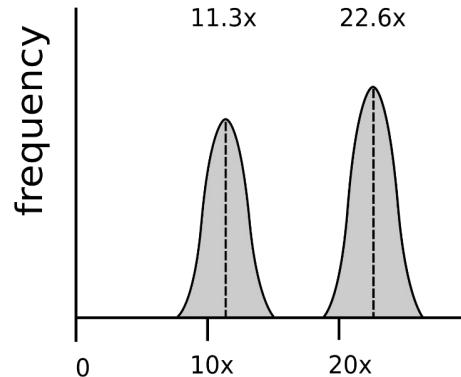


Contamination

# Reminder

The strong expectation of stoichiometry

The  $k$ -mers in a genome show (quite exactly) mean coverages proportional to their copy-number



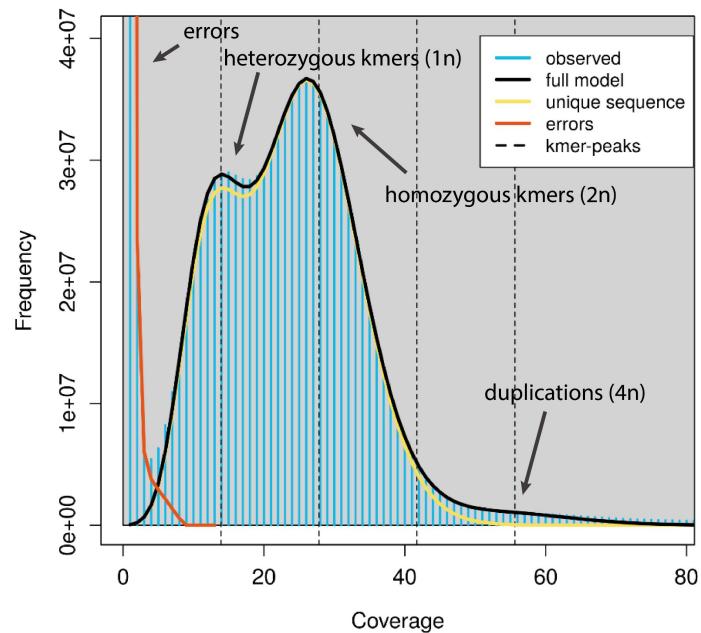
**Deviations are not statistical but technical or biological!!!**

# Part 1: Dissection of GenomeScope model



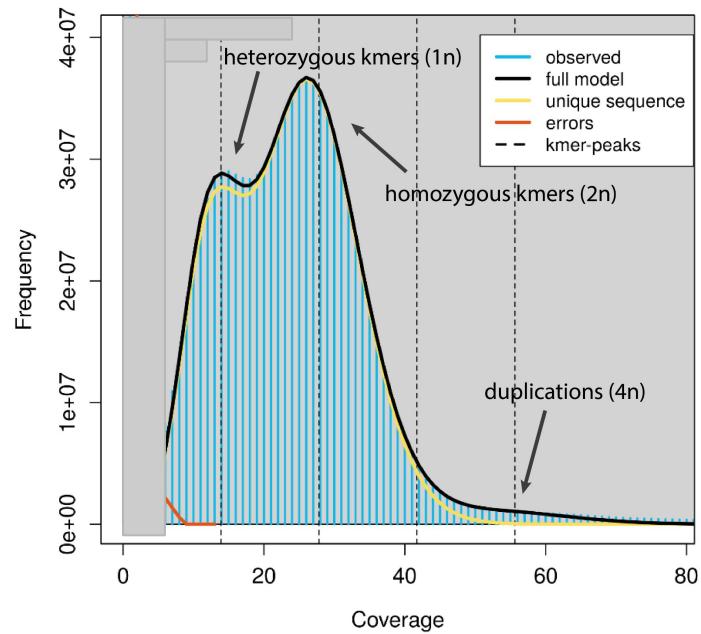
# The idea behind GenomeScope genome model

frequency ~ error peak +  
1n peak +  
2n peak +  
3n peak +  
4n peak



# GenomeScope “error” model

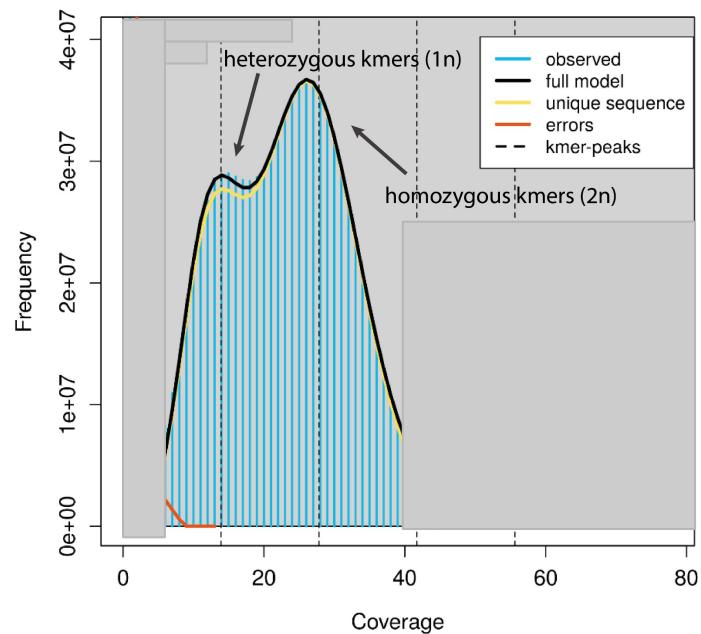
frequency  $\sim 1n$  peak +  
 $2n$  peak +  
 $3n$  peak +  
 $4n$  peak



GenomeScope actually does NOT explicitly model errors  
They are modelled as left-side residuals

# Simplified GenomeScope genome model

frequency  $\sim 1n$  peak +  
 $2n$  peak



# Modeling coverage as negative binomial

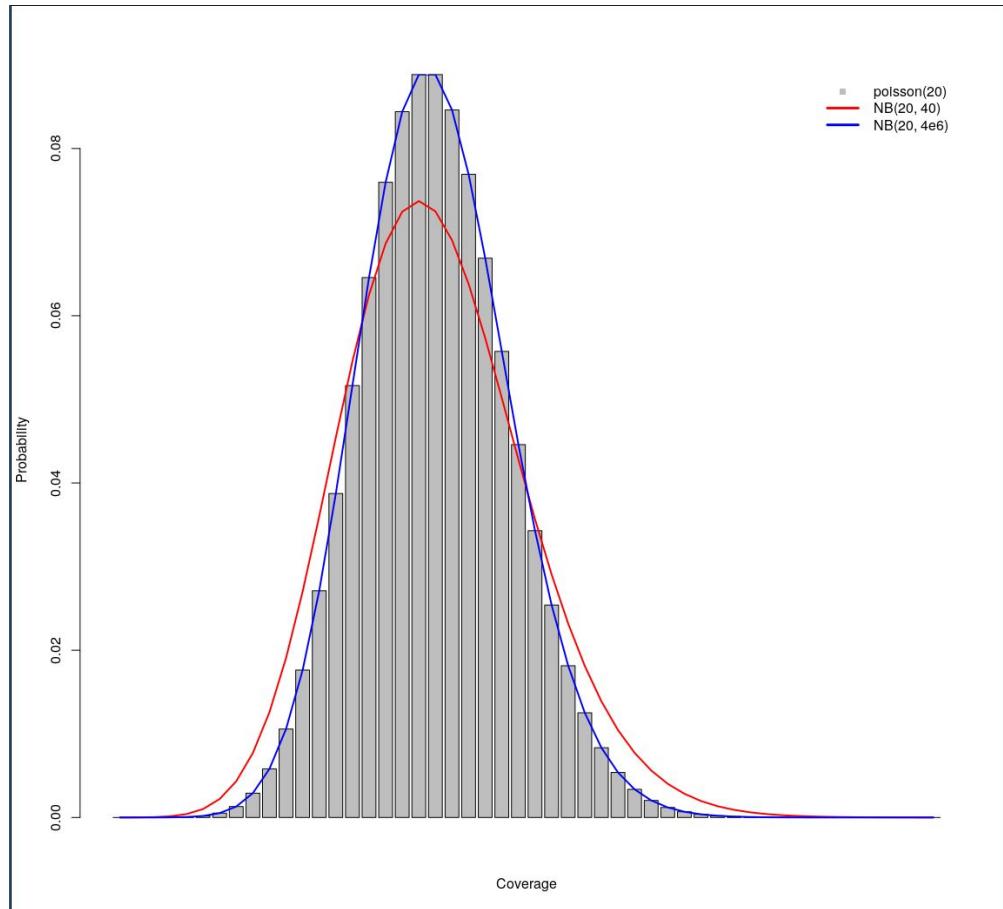
Poisson ( $\lambda$ ) ~

$$\text{NB}(\lambda, \frac{\lambda}{\xi}) \text{ for } \xi > 0$$

$\xi$  - overdispersion parameter

There is a good reason why I express it as this moment as

$$\frac{\lambda}{\xi}$$



# GenomeScope genome model - simplified

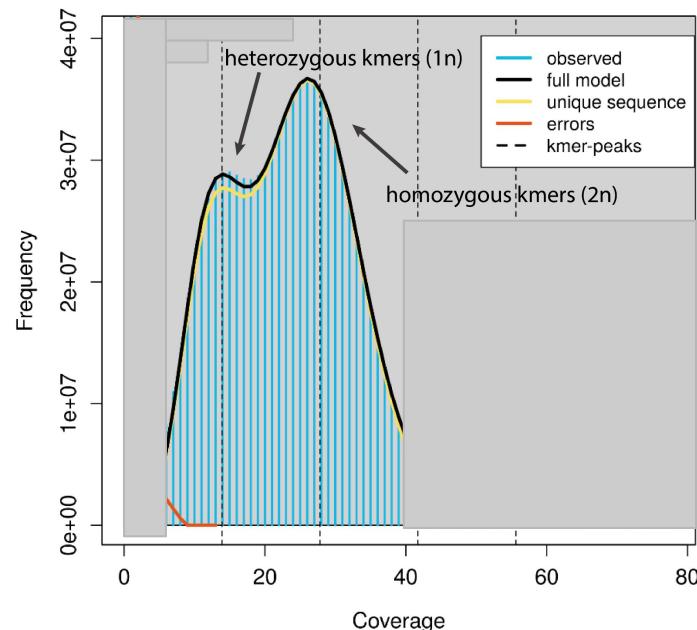
$$\text{Probability} \sim \alpha \text{ NB}(k_{\text{cov}}, k_{\text{cov}} / \xi) + \\ \beta \text{ NB}(2*k_{\text{cov}}, 2*k_{\text{cov}} / \xi)$$

$\alpha$  and  $\beta$  are relative contributions of 1n and 2n peaks

$$\text{So } \alpha + \beta = 1$$

**kcov** is 1n coverage and

$\xi$  is the overdispersion parameter from the last slide



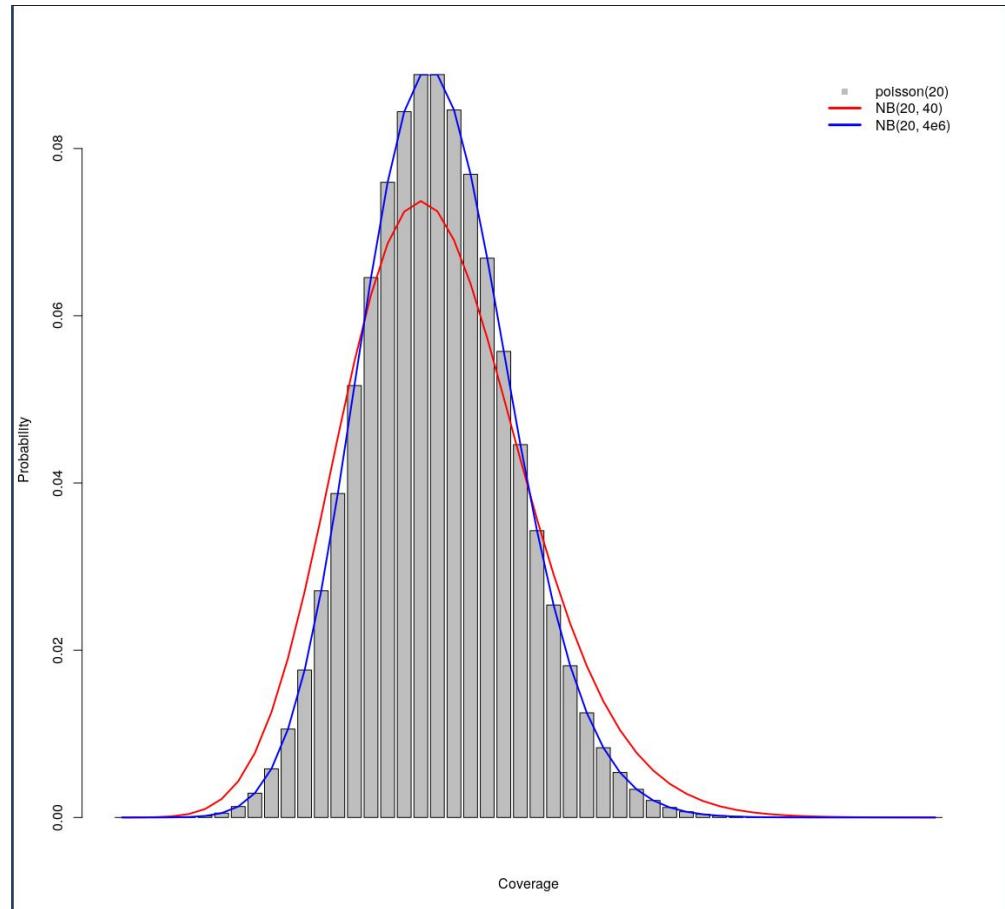
# Modeling coverage as negative binomial

Poisson ( $\lambda$ )  
NB( $\lambda, \frac{\lambda}{\xi}$ ) for  $\xi > 0$

$\xi$  - overdispersion parameter

$$\sum NB(\lambda, \frac{\lambda}{\xi}) = 1$$

Each PROBABILITY distribution sums to 1

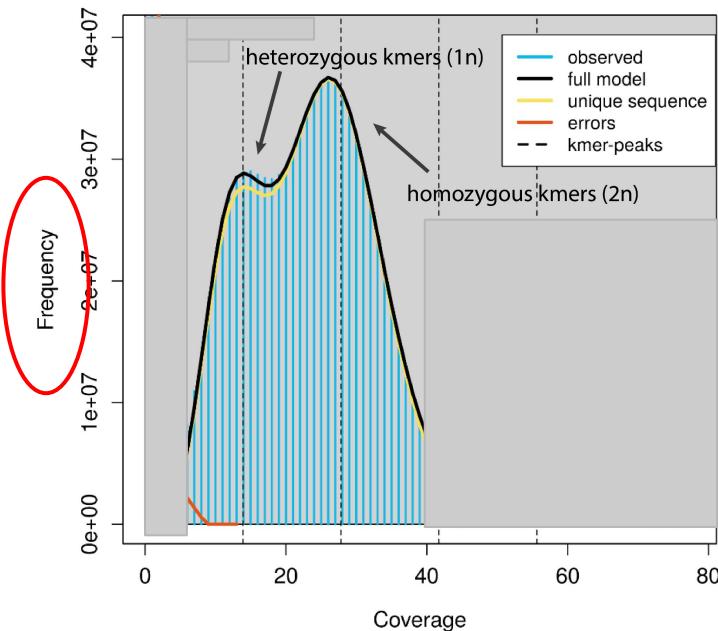


# We need a factor to scale the model to k-mer frequencies

$$\text{Frequency} \sim (\alpha \text{ NB}(\text{kcov}, \text{kcov} / \xi) + \beta \text{ NB}(2*\text{kcov}, 2*\text{kcov} / \xi))$$

Probability  $\neq$  Frequency

-> we need a factor that turns Prob  $\rightarrow$  Freq



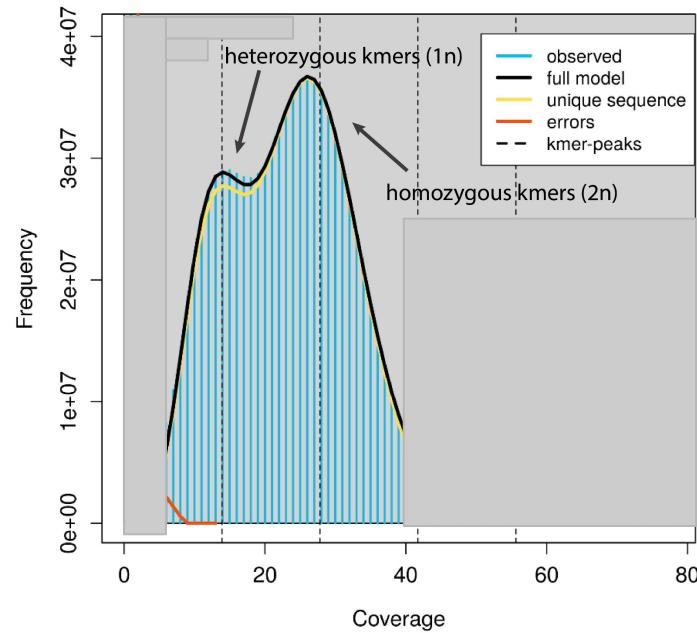
# Working GenomeScope genome model

$$\text{Frequency} \sim (\alpha \text{ NB}(\text{kcov}, \text{kcov} / \xi) + \beta \text{ NB}(2*\text{kcov}, 2*\text{kcov} / \xi)) * \text{length}$$

$\alpha$  and  $\beta$  are relative contributions of 1n and 2n peaks

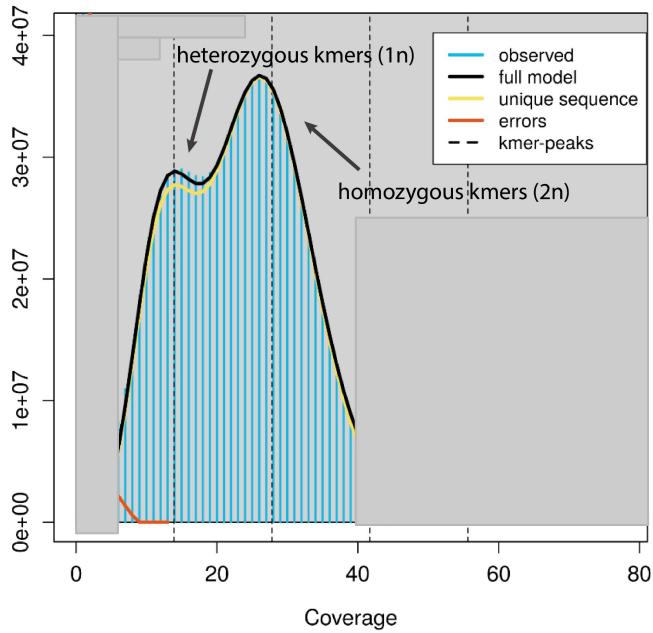
**kcov** is 1n coverage and bias is overdispersion parameter

**length** is a factor that will scale the probability distribution (and moreless corresponds to the haploid genome size)



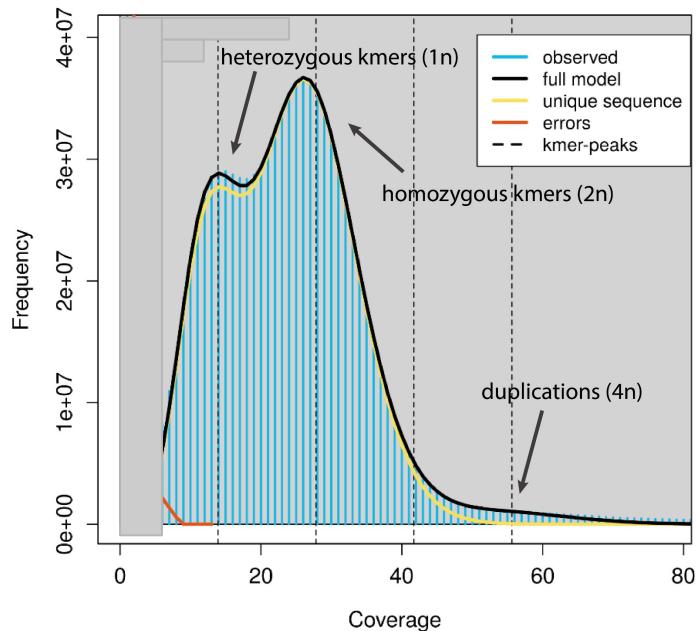
# Why moreless? And how to make it right

**length** is a factor that will scale the probability distribution (and moreless corresponds to the haploid genome size)



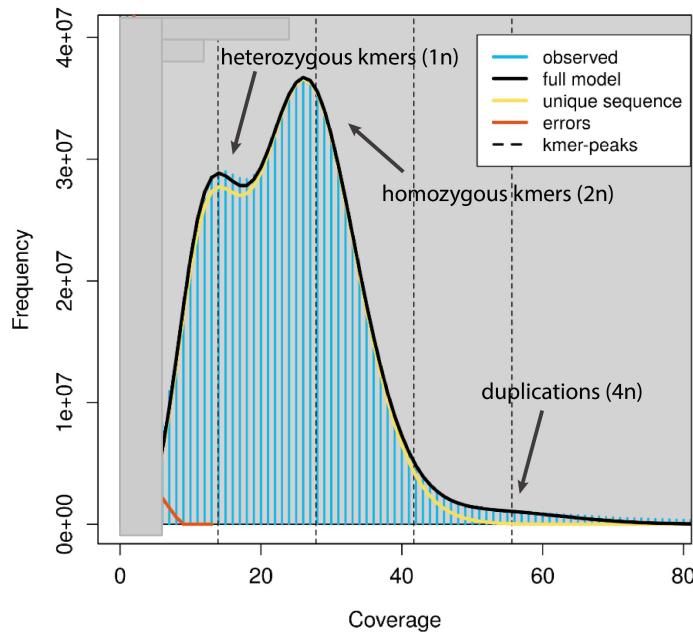
# We ignore every repetition in the genome!

and they represent a large portion of our genomes!



# Calculating genome size

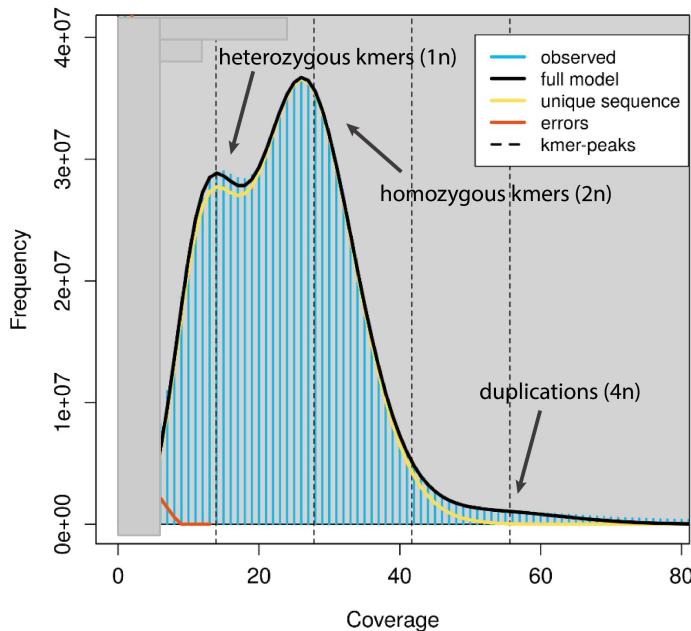
$$\text{coverage} = \frac{\text{sequencing yield}}{\text{genome size}}$$



# Calculating genome size

$$\text{coverage} = \frac{\text{sequencing yield}}{\text{genome size}}$$

$$\text{genome size} = \frac{\text{sequencing yield}}{\text{coverage}}$$

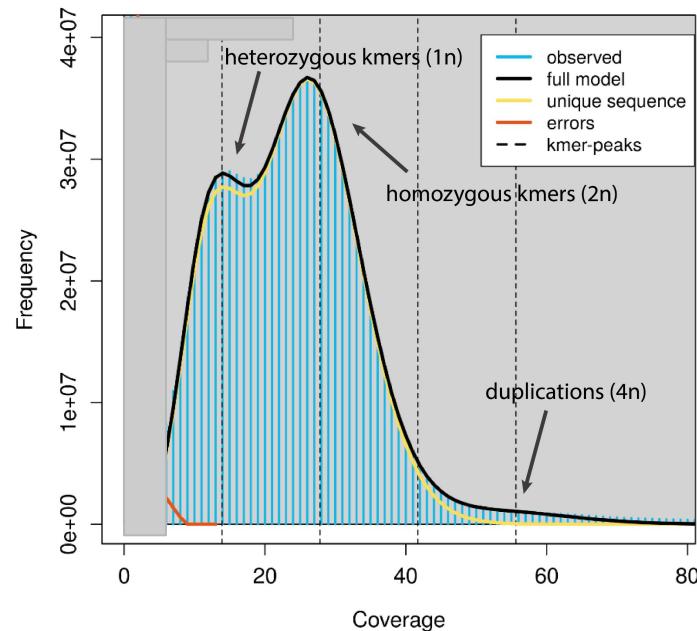


# Calculating genome size

$$\text{coverage} = \frac{\text{sequencing yield}}{\text{genome size}}$$

$$\text{genome size} = \frac{\text{sequencing yield}}{\text{coverage}}$$

$$\text{genome size} = \frac{\text{sequencing "k-mer" yield}}{\text{k-mer coverage} * \text{ploidy}}$$

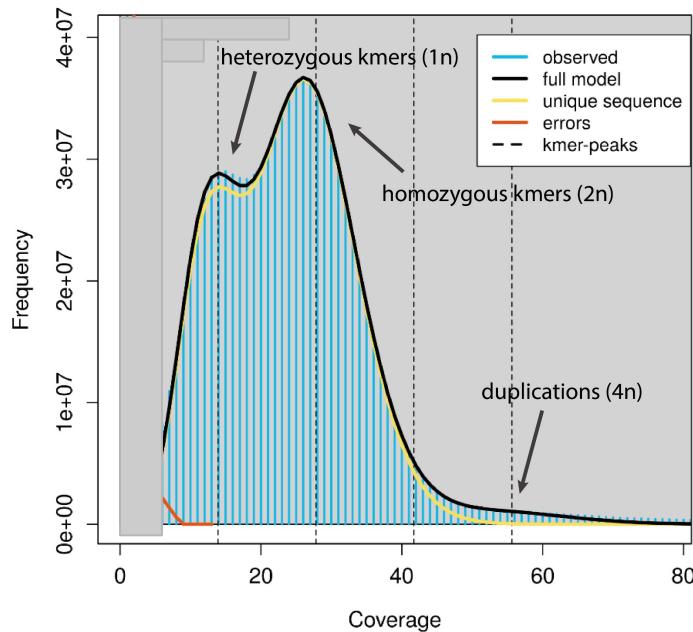


# Calculating genome size

“sum all the non-error k-mer coverages”

in k-mer histogram world:  $\text{sum}(\text{kmer\_cov} * \text{kmer\_freq})$

$$\text{genome size} = \frac{\text{sequencing “k-mer” yield}}{\text{k-mer coverage} * \text{ploidy}}$$



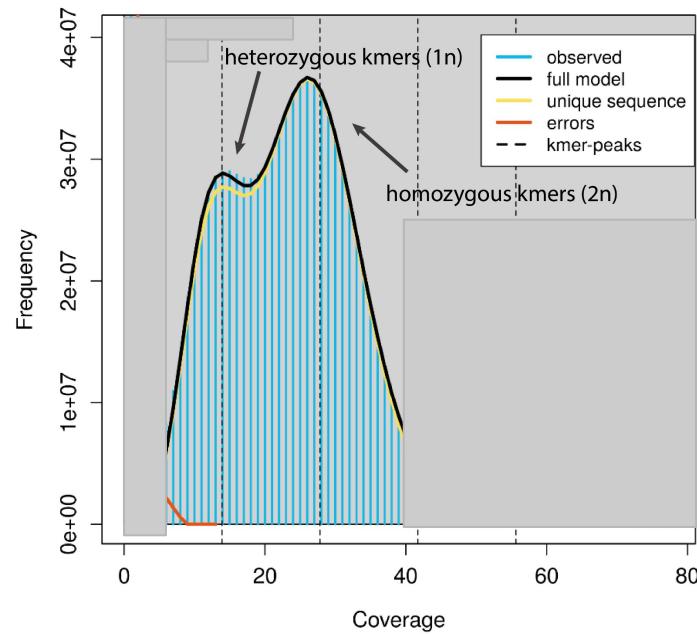
# Working GenomeScope genome model

$$\text{Frequency} \sim (\alpha \text{ NB}(\text{kcov}, \text{kcov} / \xi) + \beta \text{ NB}(2*\text{kcov}, 2*\text{kcov} / \xi)) * \text{length}$$

$\alpha$  and  $\beta$  are relative contributions of 1n and 2n peaks

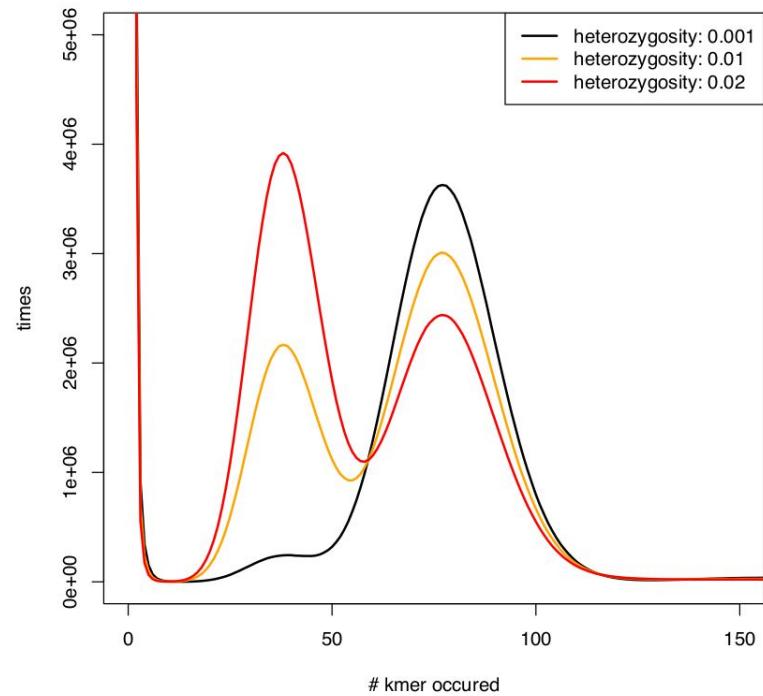
**kcov** is 1n coverage and bias is overdispersion parameter

**length** is a factor that will scale the probability distribution (and moreless corresponds to the haploid genome size)



# $\alpha$ and $\beta$ somehow represent heterozygosity

Frequency  $\sim (\alpha \text{ NB}(\text{kcov}, \text{kcov} / \xi) + \beta \text{ NB}(2 * \text{kcov}, 2 * \text{kcov} / \xi)) * \text{length}$



from Supplementary materials of Vurture et al. 2017

# Expressing $\alpha$ and $\beta$ as function of heterozygosity

**r**  $\Leftrightarrow$  probability of a heterozygous nucleotide (i.e. heterozygosity)

**k**  $\Leftrightarrow$  k-mer size

# Expressing $\alpha$ and $\beta$ as function of heterozygosity

**r**  $\Leftrightarrow$  probability of a heterozygous nucleotide (i.e. heterozygosity)

**k**  $\Leftrightarrow$  k-mer size

$(1 - r)$   $\Leftrightarrow$  prob of a nucleotide being homozygous

$(1 - r)^k$   $\Leftrightarrow$  prob of homozygous ( $2n$ ) k-mer

$1 - (1 - r)^k$   $\Leftrightarrow$  prob of heterozygous ( $1n$ ) k-mer

$\Leftrightarrow$  Each heterozygous site generate 2 het k-mers

# Expressing $\alpha$ and $\beta$ as function of heterozygosity

**r**  $\Leftrightarrow$  probability of a heterozygous nucleotide (i.e. heterozygosity)

**k**  $\Leftrightarrow$  k-mer size

$(1 - r)$   $\Leftrightarrow$  prob of a nucleotide being homozygous

$(1 - r)^k$   $\Leftrightarrow$  prob of homozygous (2n) k-mer

$1 - (1 - r)^k$   $\Leftrightarrow$  prob of heterozygous (1n) k-mer

$\Leftrightarrow$  Each heterozygous site generate 2 het k-mers

$$\alpha = 2 * (1 - (1 - r)^k)$$

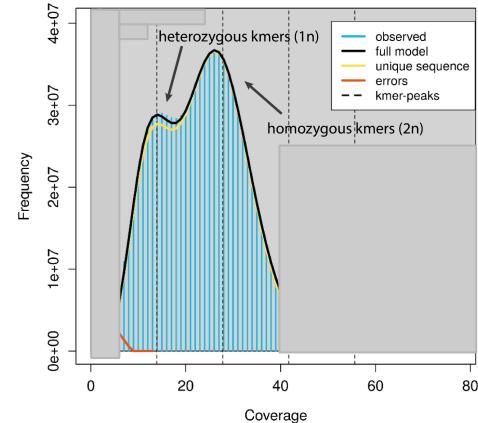
$$\beta = (1 - r)^k$$

# Working GenomeScope genome model in R

```
Frequency ~ (α NB( kcov, kcov / ξ) +  
β NB(2*kcov, 2*kcov / ξ)) * length
```

In R:

```
nls(y ~ ((2*(1-(1-r)^k)) * dnbinom(x, size = kcov / overdispersion, mu = kcov) +  
((1-r)^k) * dnbinom(x, size = kcov^2 / overdispersion, mu = kcov^2)) * length),  
...)
```



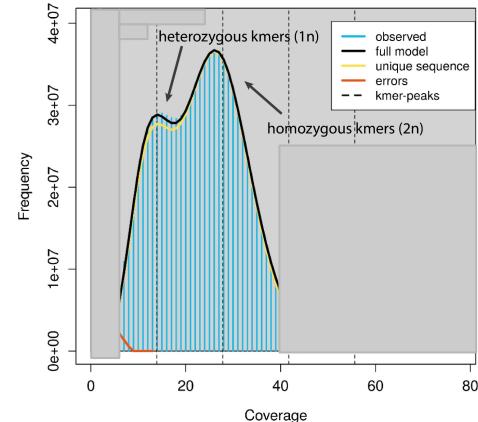
# Working GenomeScope genome model in R

```
Frequency ~ (α NB( kcov, kcov / ξ) +  
β NB(2*kcov, 2*kcov / ξ)) * length
```

In R:

```
nls(y ~ ((2*(1-(1-r)^k)) * dnbinom(x, size = kcov / overdispersion, mu = kcov) +  
((1-r)^k) * dnbinom(x, size = kcov^2 / overdispersion, mu = kcov^2)) * length),  
start = list(r = 0, kcov = 50, overdispersion = 0.5, length = 300e6),  
...)
```

starting conditions



# Working GenomeScope genome model in R

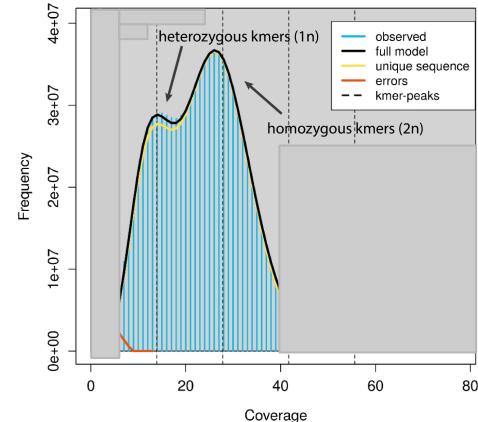
Frequency ~  $(\alpha \text{ NB}(\text{kcov}, \text{kcov} / \xi) + \beta \text{ NB}(2*\text{kcov}, 2*\text{kcov} / \xi)) * \text{length}$

In R:

```
nls(y ~ ((2*(1-(1-r)^k)) * dnbinom(x, size = kcov / overdispersion, mu = kcov) +  
((1-r)^k) * dnbinom(x, size = kcov^2 / overdispersion, mu = kcov^2)) * length),  
start = list(r = 0, kcov = 50, overdispersion = 0.5, length = 300e6),  
control = list(minFactor=1e-12, maxiter=40))
```

starting conditions

setting for the iterative least square search



# Try it out

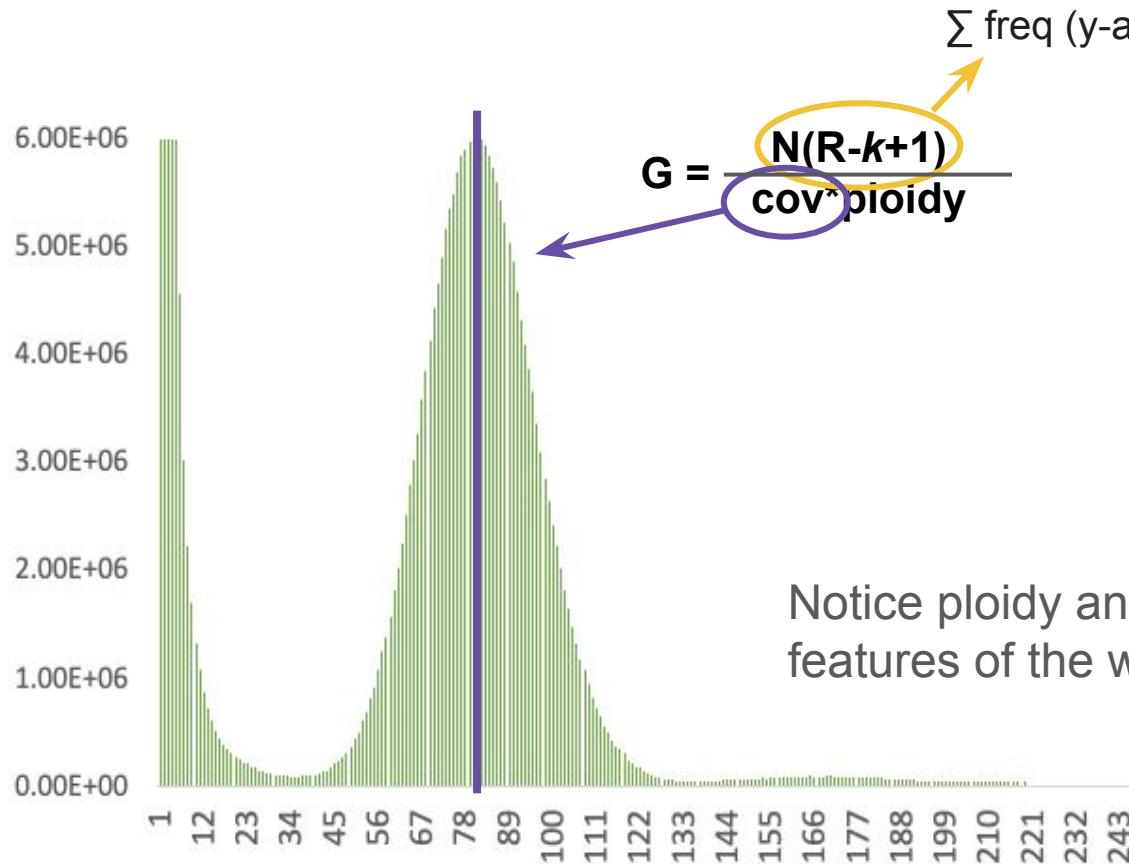
1. Load k-mer histogram to R
2. Write a GenomeScope model yourself
3. Get a bit of intuition about starting conditions and appreciation about how the model is constructed

Tips:

- Singular gradient means that the model could not converge (did not see improvement when changing parameters), usually your starting values are too off
- You can use plotting functions in the GenomeTelescope package to save some time
- Use pen and pencil to think the models through, try things out, have fun

# Reminder

## Genome size estimates



**G** = genome size (bp)

**N** = number of reads

**R** = read length

**k** = k-mer length

**cov** = peak coverage

Notice ploidy and coverage are features of the whole genome

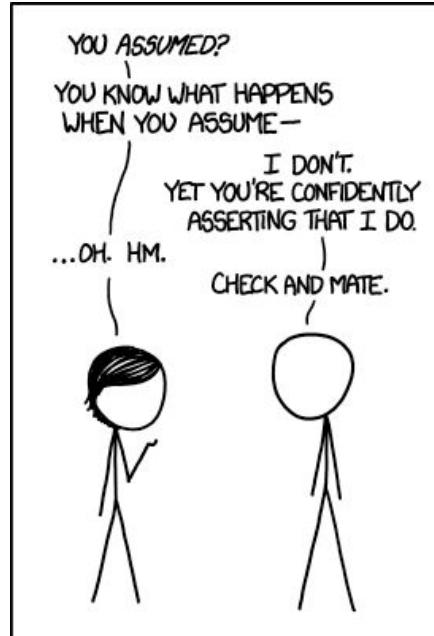
# Smooth segway...

Can you think when this is not true?

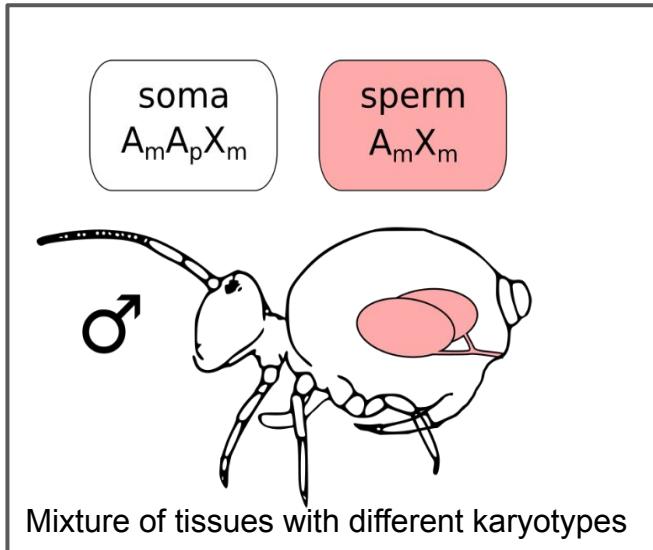


Notice ploidy and coverage are features of the whole genome

# Part 2: Violating GenomeScope assumptions



# Conditions naturally violating GenomeScope assumptions



# Estimating coverage and genome size of cobionts

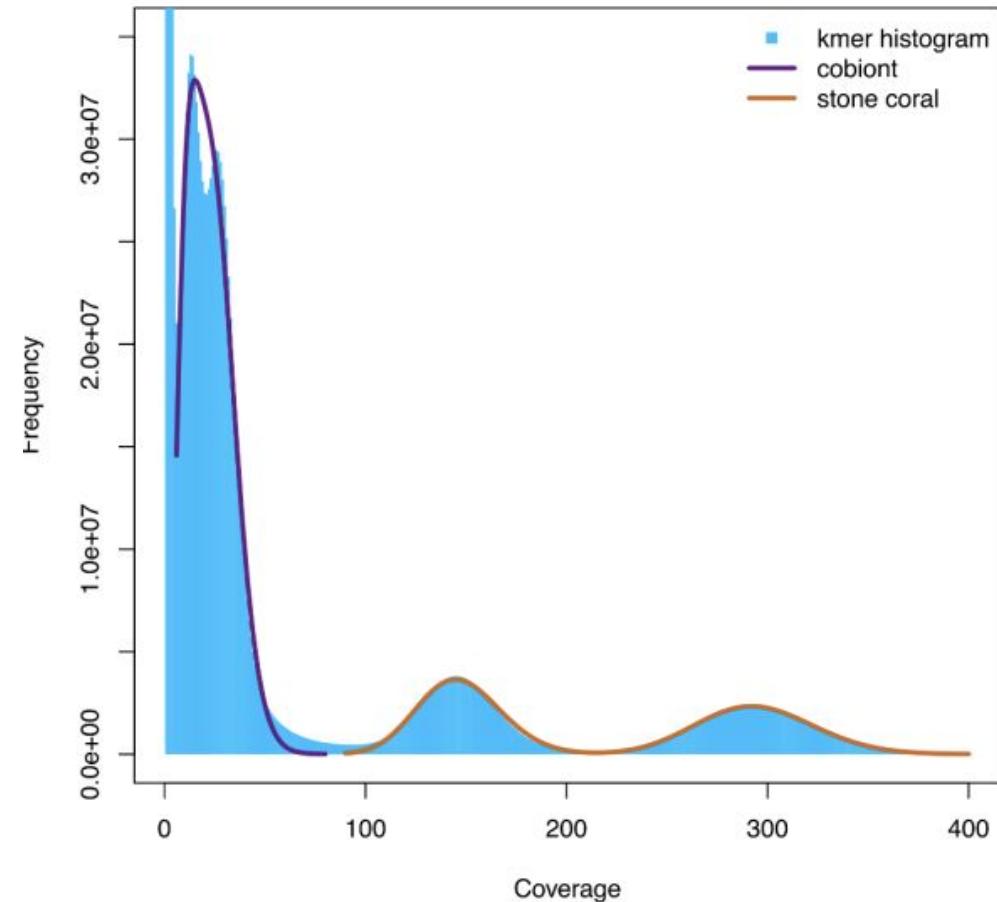


# Obligate cobionts



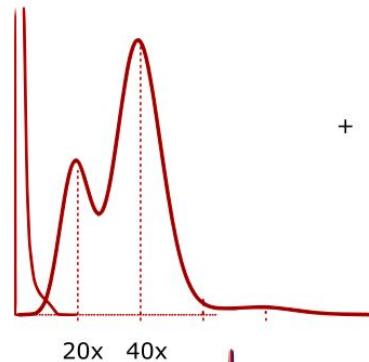
Spectrum of two diploid genomes with very different coverages and genome sizes

## B. Stone coral and its cobiont

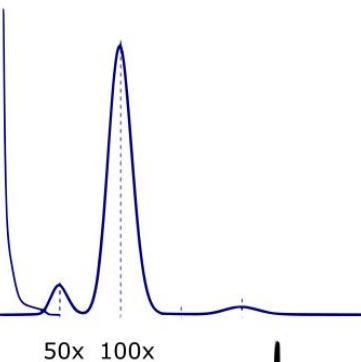


# Expectations from the co-biont spectra

Cobiont 1

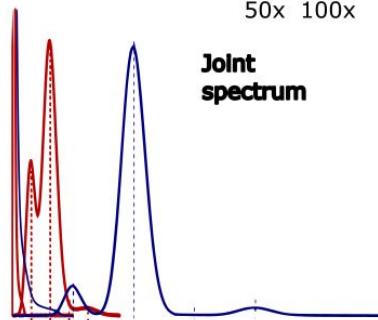


Cobiont 2



+

Joint  
spectrum

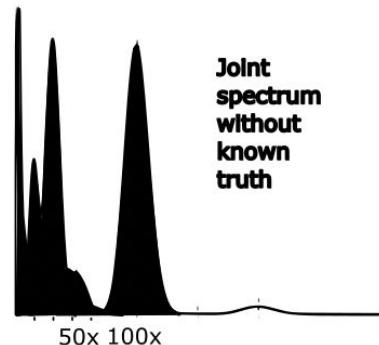


=

Assuming very distinct contributors!

- Stochiometry is retained within genomes
- Stochiometry is violated between genomes
- Joint spectrum can be a bit messy

Joint  
spectrum  
without  
known  
truth



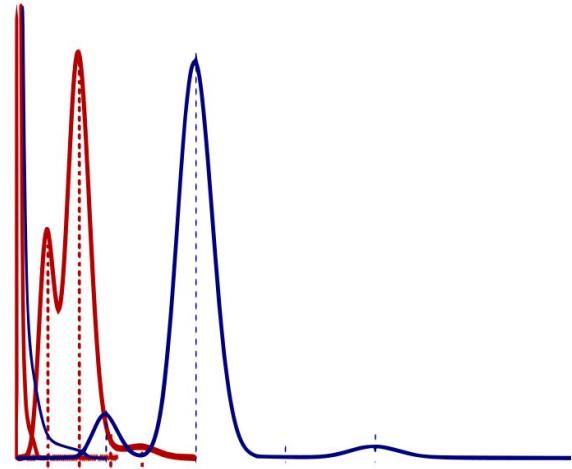
# Idea behind cobiont model

Originally

$$\text{Frequency} \sim (\alpha \text{ NB}(kcov, kcov / \xi) + \beta \text{ NB}(2*kcov, 2*kcov / \xi)) * \text{length}$$

The adjustment conceptually

$$\begin{aligned} \text{Frequency} \sim & (\alpha \text{ NB}(kcov, kcov / \xi) + \\ & \beta \text{ NB}(2*kcov, 2*kcov / \xi)) * \text{length} \\ & + \\ & (\alpha \text{ NB}(kcov, kcov / \xi) + \\ & \beta \text{ NB}(2*kcov, 2*kcov / \xi)) * \text{length} \end{aligned}$$



# Separate parameters of the two models

The adjustment

$$\begin{aligned} \text{Frequency} \sim & (\alpha \text{ NB}(k\text{cov1}, k\text{cov1} / \xi) + \\ & \beta \text{ NB}(2*k\text{cov1}, 2*k\text{cov1} / \xi)) * \text{length1} \\ & + \\ & (\gamma \text{ NB}(k\text{cov2}, k\text{cov2} / \xi) + \\ & \delta \text{ NB}(2*k\text{cov2}, 2*k\text{cov2} / \xi)) * \text{length2} \end{aligned}$$

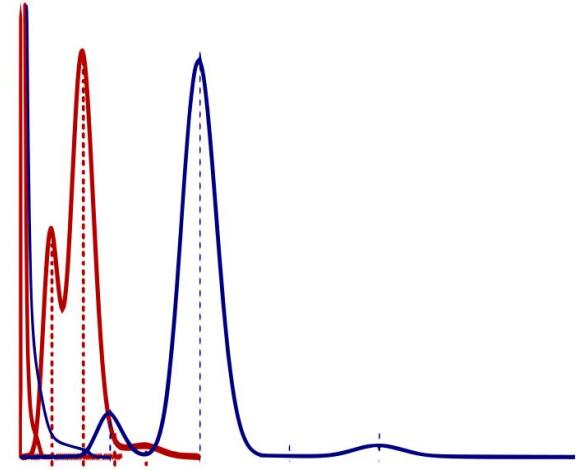
**kcov1, kcov2** - respective coverages

**length1, length2** - respective genome lengths

$$\alpha + \beta = 1$$

$$\gamma + \delta = 1$$

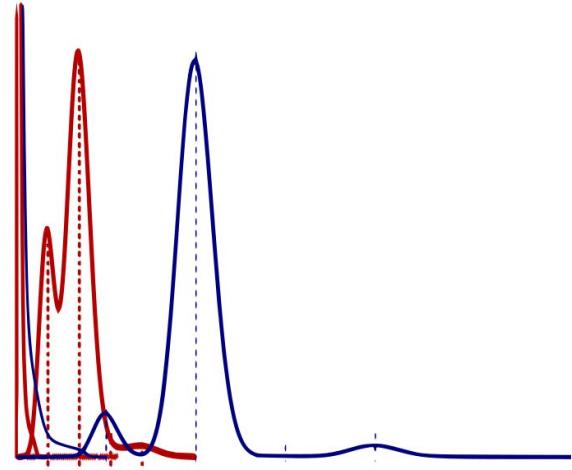
**$\xi$**  - overdispersal; single parameter (scales with coverage)



# Co-biont model in R

In R (if both haploid):

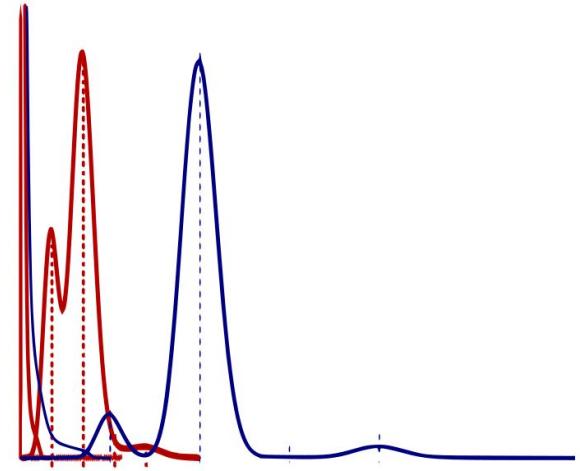
```
lichen_model <- nlsLM(y ~ (dnbinom(x, mu = kcov, size = kcov/overdispersion) * length1) +  
                         (dnbinom(x, mu = kcov2, size = kcov2/overdispersion) * length2),  
                         start = list(kcov = kmerEst, kcov2 = kmerEst2,  
                                      overdispersion = 0.5, length1 = lengthEst, length2 = lengthEst),  
                         control = list(minFactor = 1e-12, maxiter = 40))
```



# Co-biont model in R

In R (if one is diploid):

```
lichen_model <- nlsLM(y ~ (a           * dnbinom(x, mu = kcov,      size =      kcov/overdispersion) +  
                           (1 - a) * dnbinom(x, mu = 2*kcov, size = 2*kcov/overdispersion)) *  
                           length1 +  
                           dnbinom(x, mu = kcov2, size = kcov2/overdispersion) * length2,  
                           start = list(kcov = kmerEst, kcov2 = kmerEst2, a = 0.3,  
                           overdispersion = 0.5, length1 = lengthEst, length2 = lengthEst),  
                           control = list(minFactor = 1e-12, maxiter = 40))
```



# Important bottleneck - very sensible starting values

- Explain singular gradient here

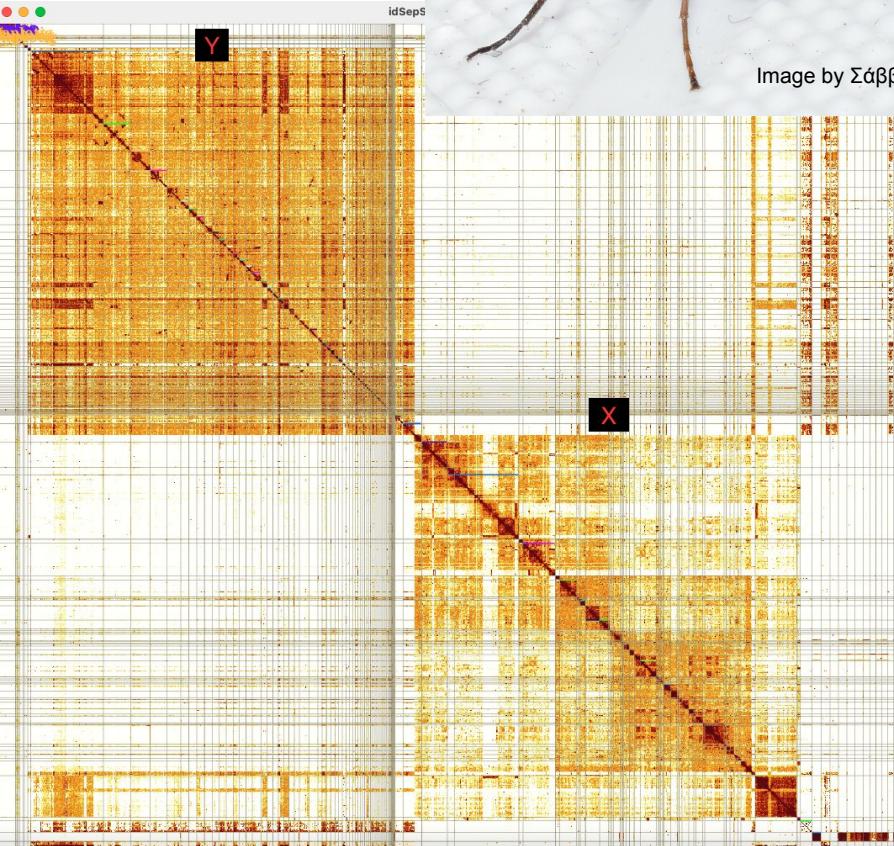
# How the non-linear least square method works?

- Minimising residual squares (similarly to any other linear model you know)
- No exact solution, so you guess some values, calculate gradients for changinge those values and iteratively change them to minimise residual squares
- Gauss Newton method

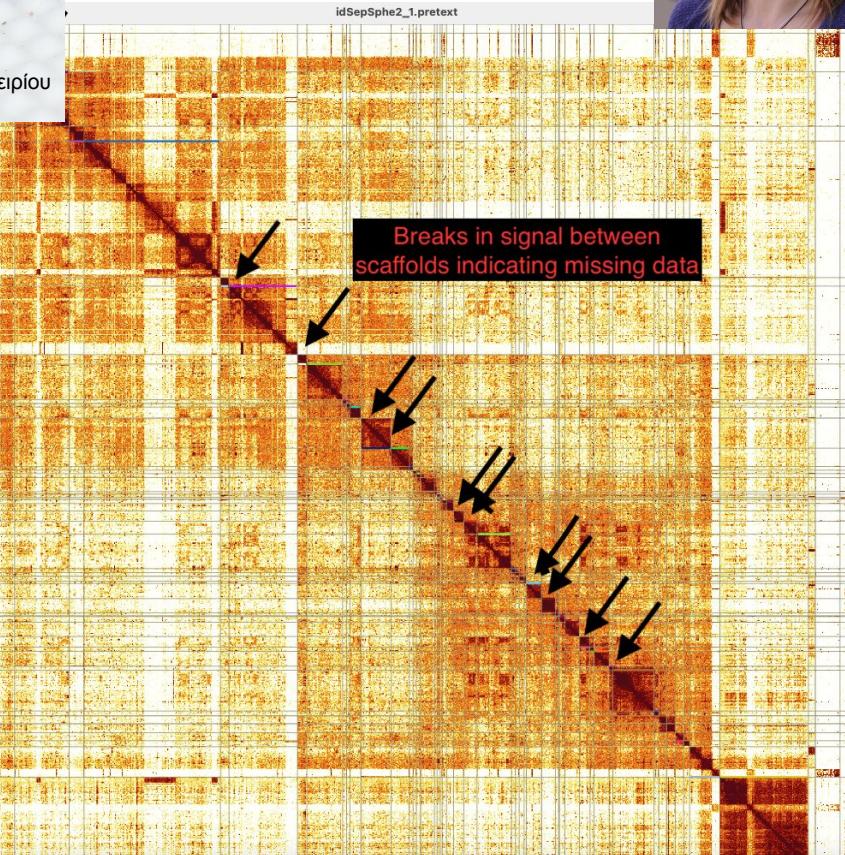
# Estimating the size of heterogametic sex chromosomes



# A march fly



Curated by Joanna Collins



# XY sex chromosome model

Workflow for XY species:

1. Assume heterozygosity / heterogametic size (possible to est from homogametic sex)
2. Fit to heterogametic sex k-mer spectra with a fixed parameter & infer the other

$$\text{Frequency} \sim (\alpha \text{ NB}(k_{cov}, k_{cov} / \text{overdispersion}) + \beta \text{ NB}(2*k_{cov}, 2*k_{cov} / \text{overdispersion}) * (\text{length} - XY\_len) + \text{NB}(k_{cov}, k_{cov} / \text{overdispersion}) * XY\_len)$$

autosomes

X and Y chromosomes

Bold parameters are fixed constants (fixed parameters estimated in female)

# Sex chromosome model - XY 4peaks

Frequency ~  $(\alpha \text{ NB}(\text{kcov}, \text{kcov} / \text{bias}) +$  **autosomes**

$\beta \text{ NB}(2*\text{kcov}, 2*\text{kcov} / \text{bias}) +$

$\gamma \text{ NB}(3*\text{kcov}, 3*\text{kcov} / \text{bias}) +$

$\delta \text{ NB}(4*\text{kcov}, 4*\text{kcov} / \text{bias}) ) * \text{length\_diplo} +$

$(\epsilon \text{ NB}(\text{kcov}, \text{kcov} / \text{bias}) +$

$\zeta \text{ NB}(2*\text{kcov}, 2*\text{kcov} / \text{bias}) ) * \text{length\_mono}$  **X and Y chromosomes**

If we would like to make it fully “GenomeScope”-like

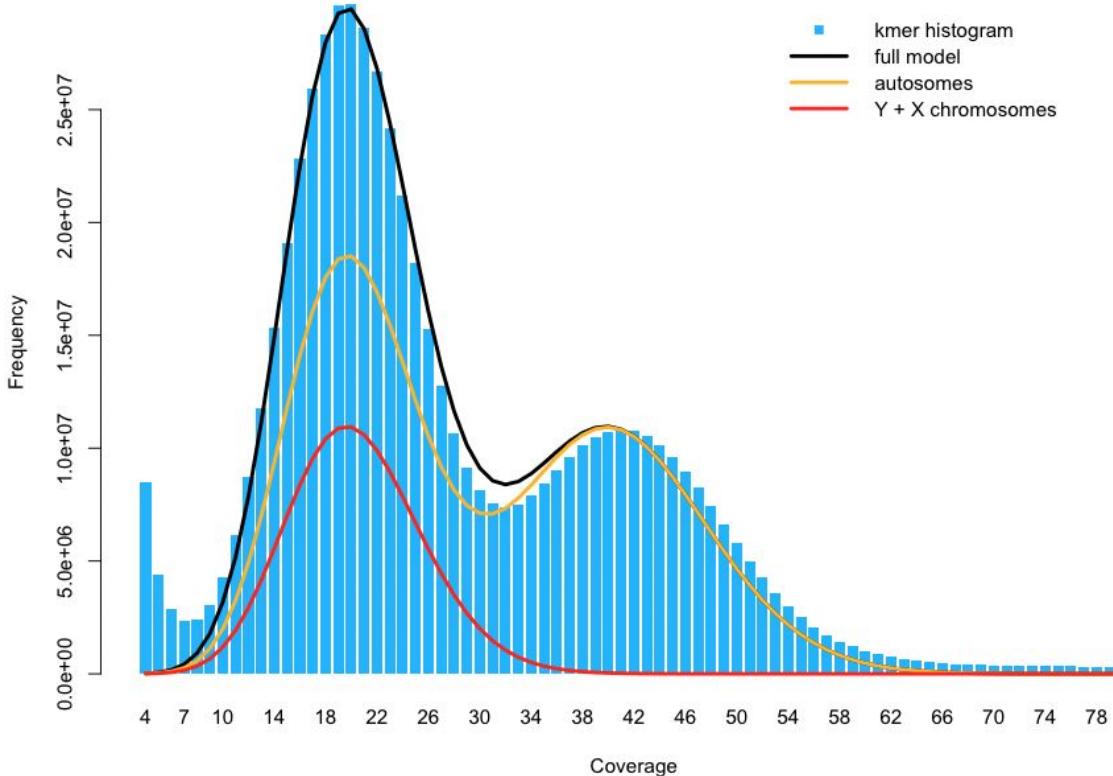
# Let's stick with XY sex chromosome model

Frequency ~ 
$$\begin{aligned} & (\alpha \text{ NB}(kcov, kcov / \text{overdispersion}) + \text{autosomes} \\ & \beta \text{ NB}(2*kcov, 2*kcov / \text{overdispersion}) * (\text{length} - XY\_len) + \\ & \text{NB}(kcov, kcov / \text{overdispersion}) * XY\_len \end{aligned}$$

**X and Y chromosomes**

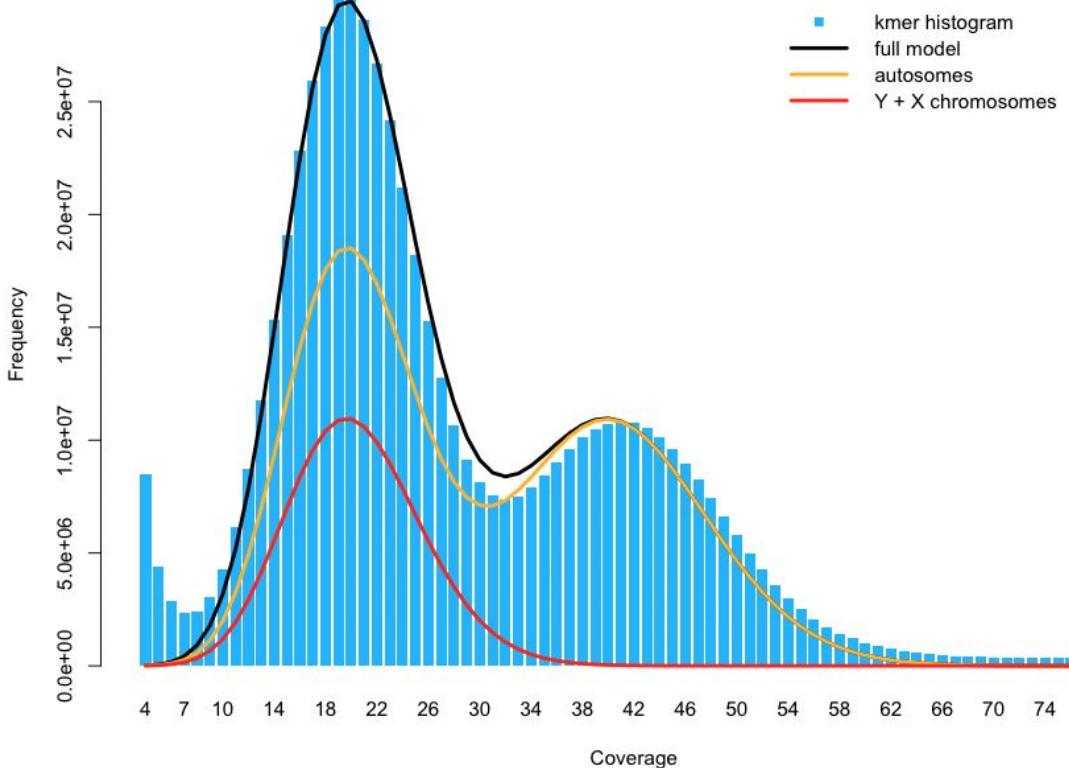
# Estimating spectra composition by fixing heterogametic size

Heterozygosity: 1.5% assumed X+Y size: 141 Mbp out of total 458.08 Mbp



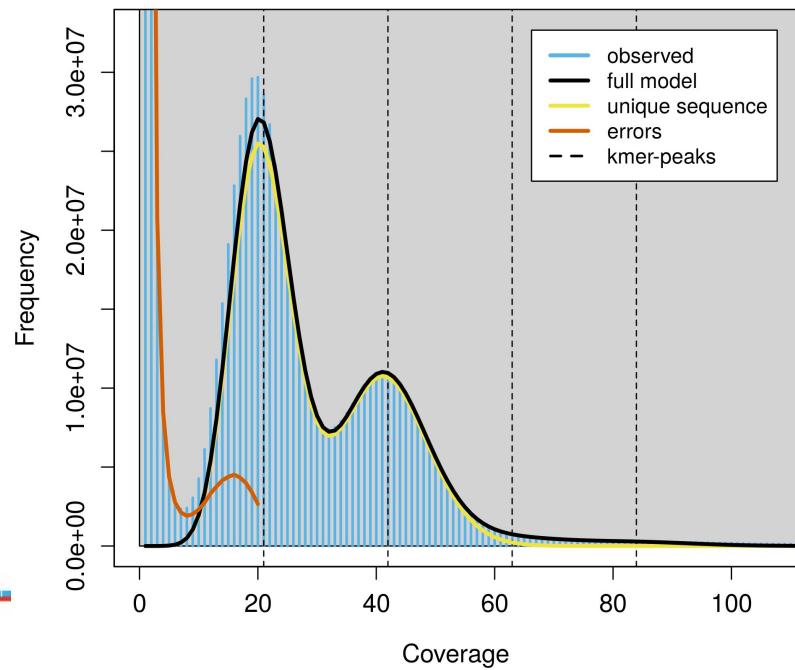
# Estimating spectra composition by fixing heterogametic size

Heterozygosity: 1.5% assumed X+Y size: 141 Mbp out of total 458.08 Mbp



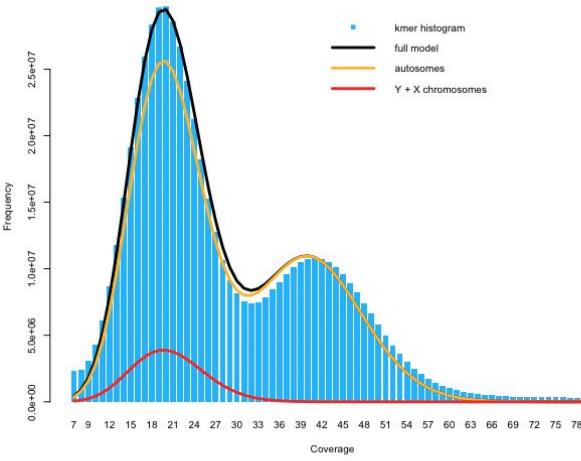
GenomeScope Profile

len:645,592,265bp uniq:54.7%  
aa:98.1% ab:1.95%  
kcov:21 err:0.193% dup:0.222 k:31 p:2

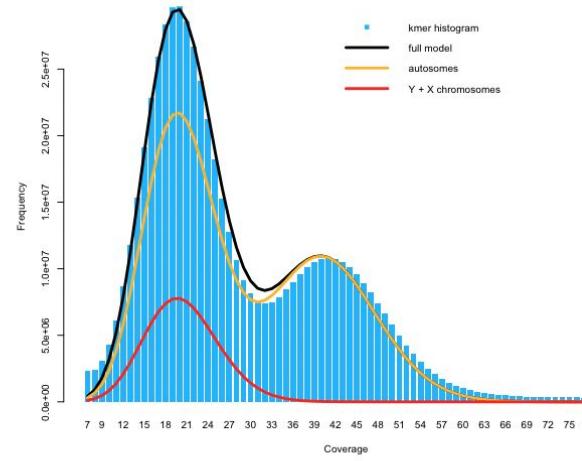


# How heterozygous given heterogametic size?

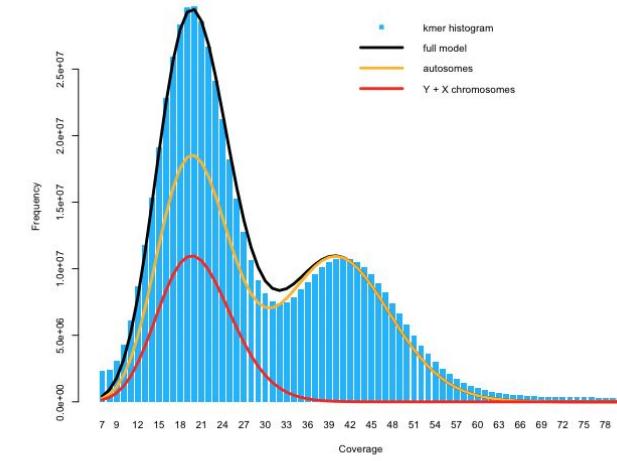
Heterozygosity: 1.93% assumed X+Y size: 50 Mbp out of total 412.4 Mbp



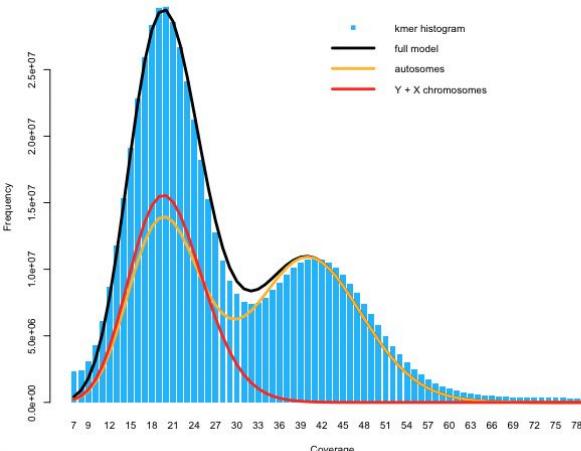
Heterozygosity: 1.7% assumed X+Y size: 100 Mbp out of total 437.4 Mbp



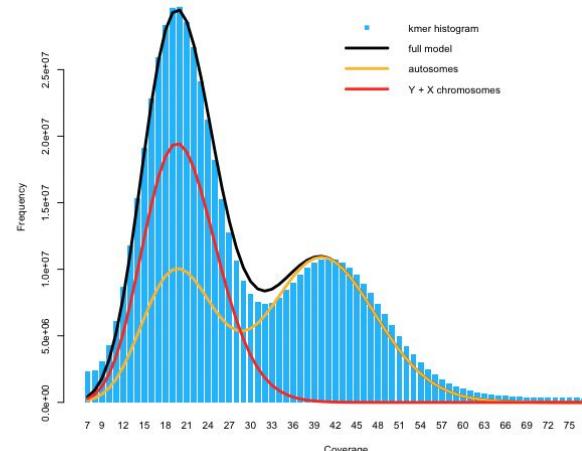
Heterozygosity: 1.5% assumed X+Y size: 141 Mbp out of total 457.9 Mbp



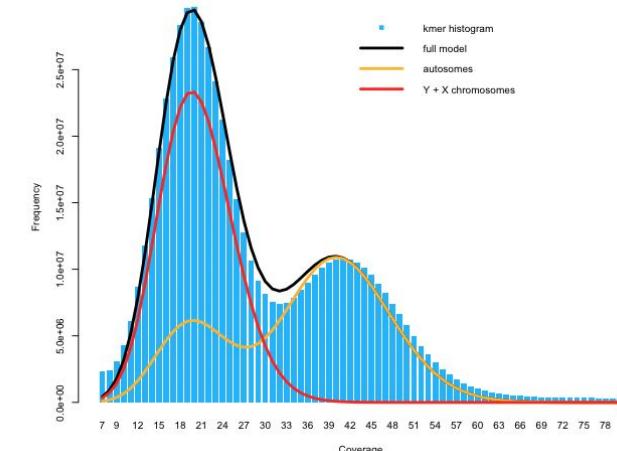
Heterozygosity: 1.19% assumed X+Y size: 200 Mbp out of total 487.4 Mbp



Heterozygosity: 0.9% assumed X+Y size: 250 Mbp out of total 512.4 Mbp

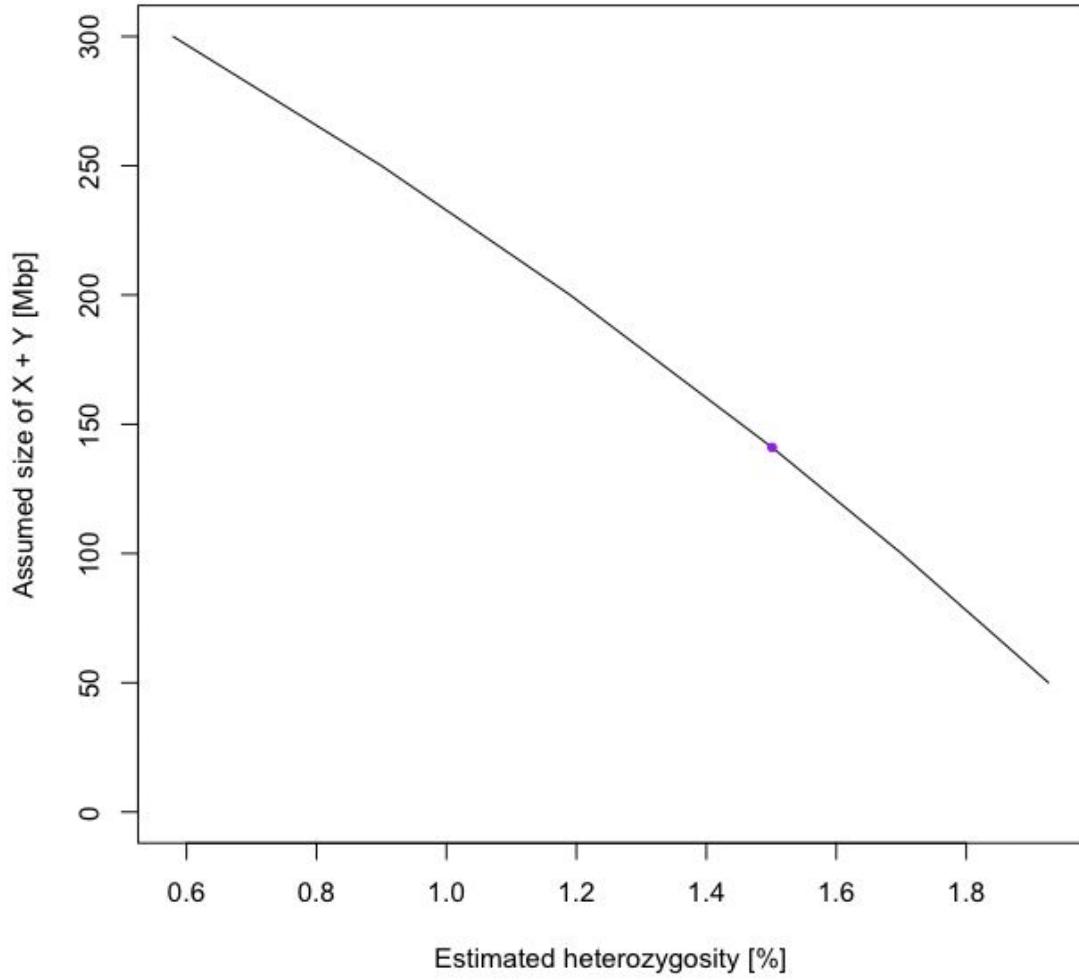


Heterozygosity: 0.58% assumed X+Y size: 300 Mbp out of total 537.4 Mbp

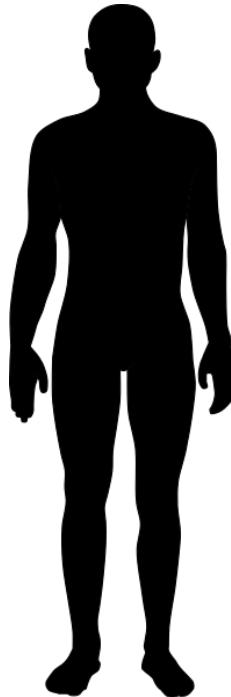


# Heterogametic size ~ heterozygosity

- Gives meaningful range of heterogametic sizes / heterozygosities



# Testing the same principle on the HG002 human



T2T based

Total

3.11 Gbp

Disomic

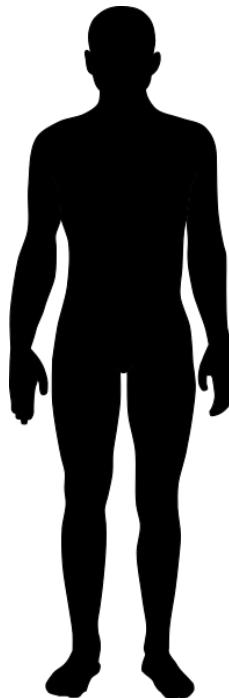
2.9 Gbp

Monosomic

0.21 Gbp



# Testing the same principle on the HG002 human



T2T based

Total

3.11 Gbp

Disomic

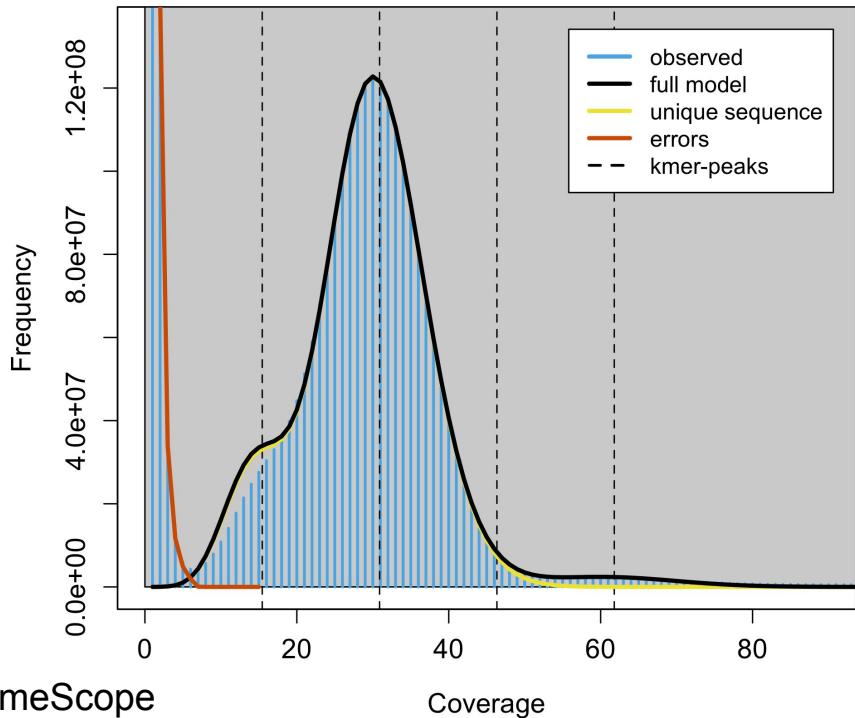
2.9 Gbp

Monosomic

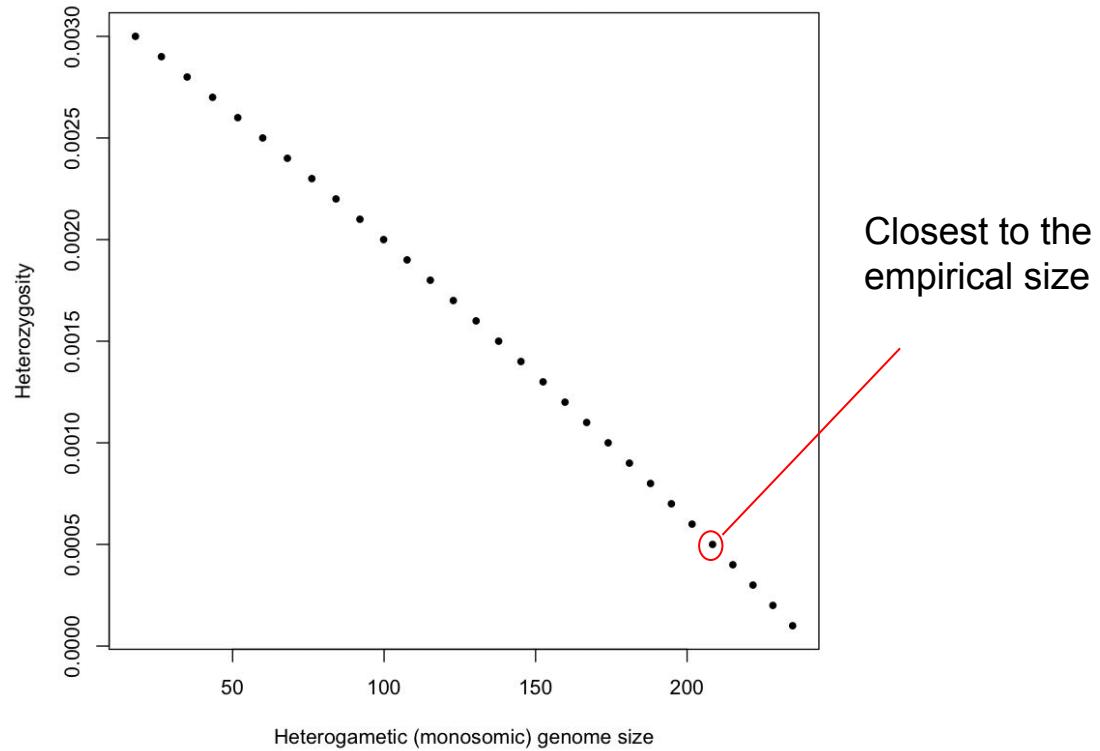
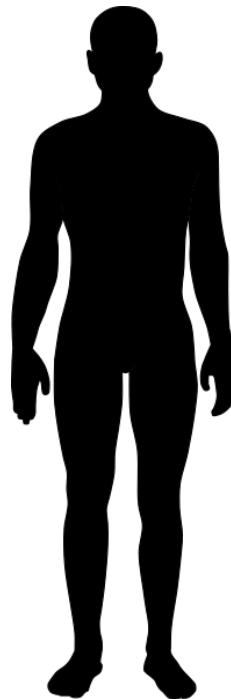
0.21 Gbp

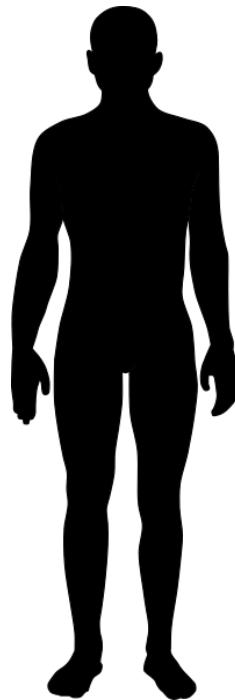
GenomeScope Profile

len:2,873,667,792bp uniq:72.3%  
aa:99.6% ab:0.392%  
kcov:15.4 err:0.112% dup:0.271 k:21 p:2



# Given heterozygosity fit a model



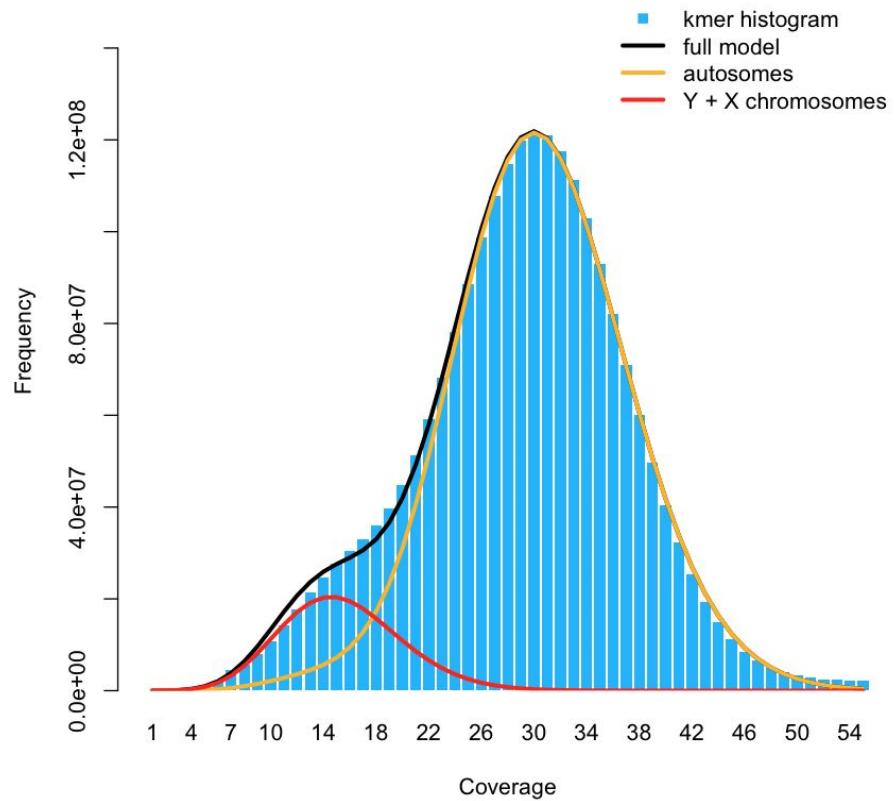


Fitted / Real values

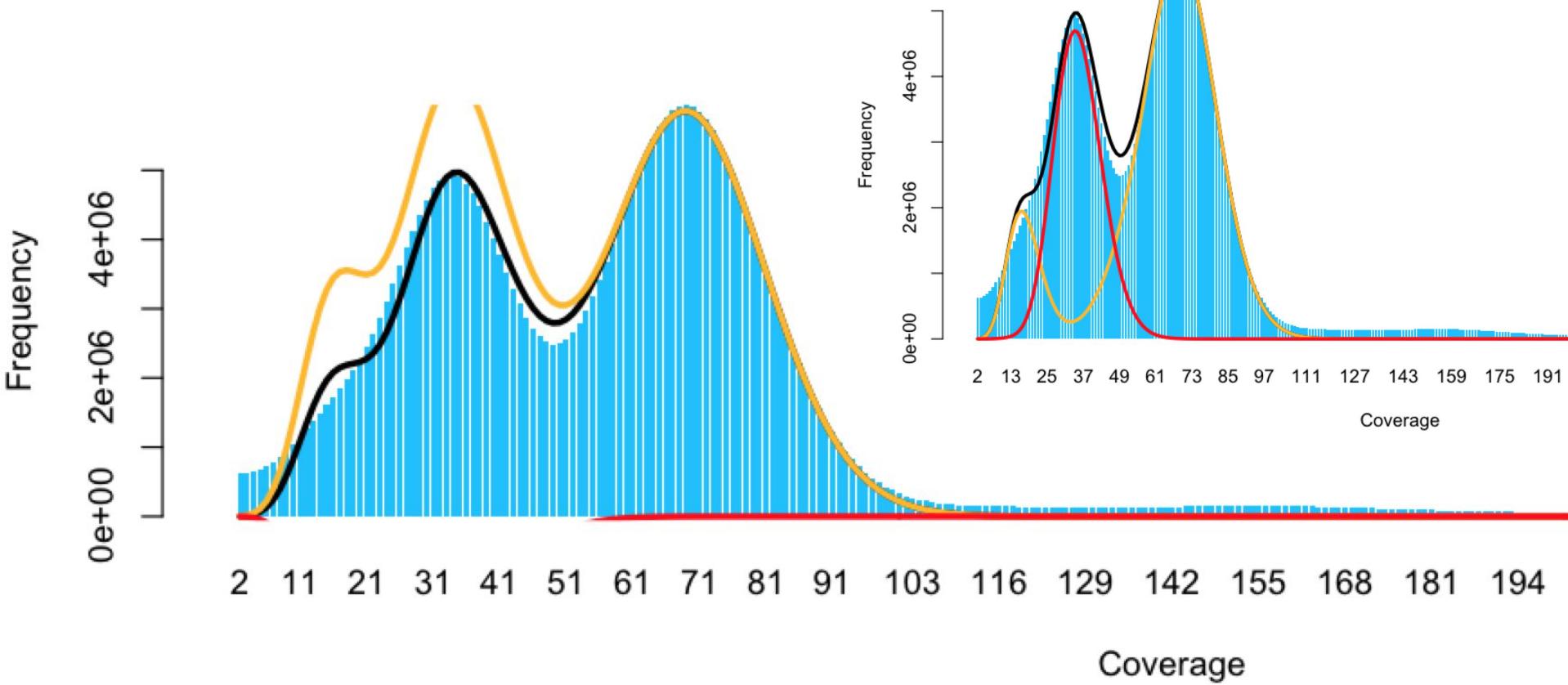
Total  
3.03 / 3.11 Gbp

Disomic  
2.75 / 2.9 Gbp

Monosomic  
0.288 / 0.21 Gbp

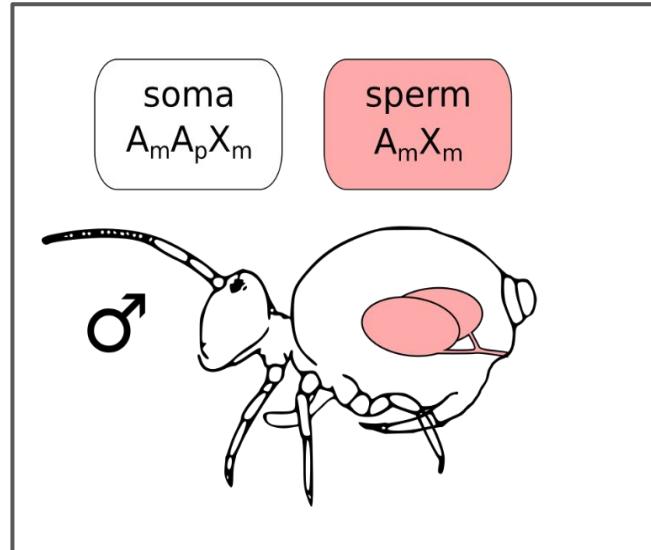


# Convergence problems of more complex models



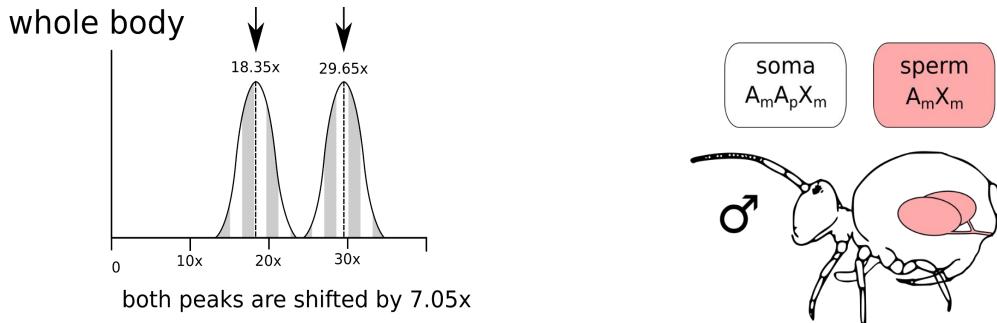
# Bonus slides

## Mixed karyotypes



# Two tissue model

PGE Male



Need to model the 1n and 2n coverage as independent variables

$$\text{Frequency} \sim (\alpha * \text{NB}(\text{kcov}, \text{kcov} / \text{bias}) + \beta * \text{NB}(\text{kcov2}, \text{kcov2} / \text{bias})) * \text{length}$$

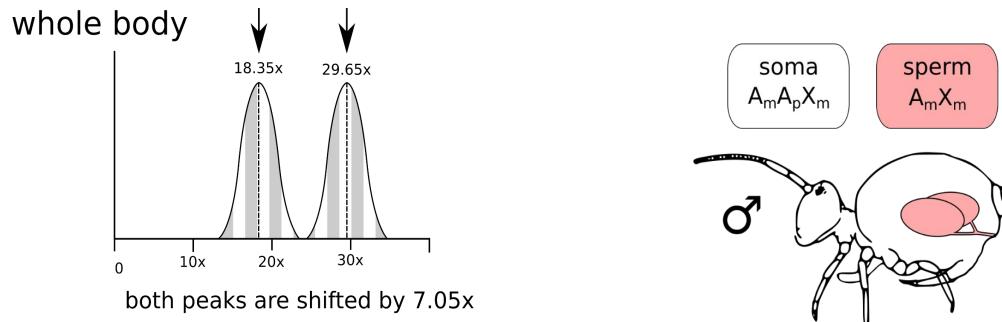
$$f_h = 1 - \frac{(c_A - c_X)}{c_X}$$

$c_A$  - autosomal coverage peak  
(2n peak, kcov2)

$c_X$  - X-chromosome coverage peak  
(1n peak, kcov)

# Two tissue model

PGE Male



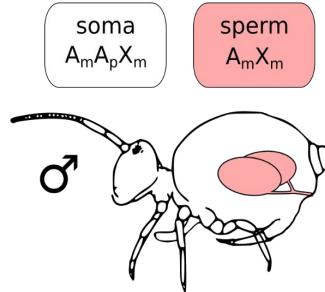
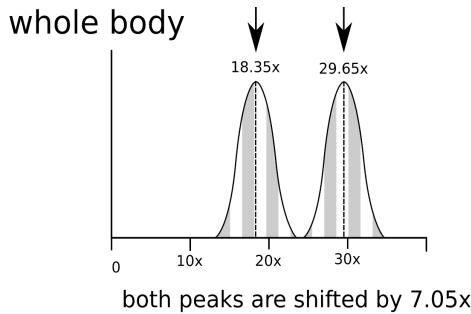
Need to model the 1n and 2n coverage as independent variables

$$\text{Frequency} \sim (\alpha * \text{NB}(\text{kcov}, \text{kcov} / \text{bias}) + (1 - \alpha) * \text{NB}(\text{kcov2}, \text{kcov2} / \text{bias})) * \text{length}$$

Modeled as  $\alpha$   
as we don't care  
about the actual  
heterozygosity

# Two tissue model

PGE Male



Need to model the 1n and 2n coverage as independent variables

$$\text{Frequency} \sim (\alpha * \text{NB}(\text{kcov}, \text{kcov} / \text{bias})) + (1 - \alpha) * \text{NB}(\text{kcov2}, \text{kcov2} / \text{bias}) * \text{length}$$

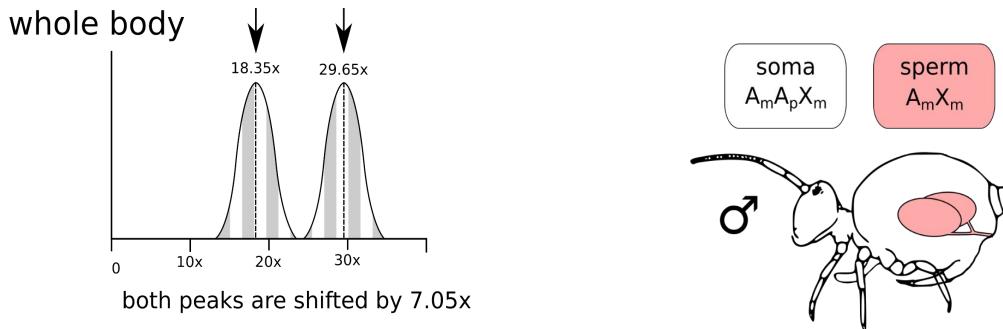
Modeled as  $\alpha$   
as we don't care  
about the actual  
heterozygosity

**kcov** and **kcov2** are the  
important modeled parameters

**bias** is still a single parameter in  
the model :-)

# Two tissue model

PGE Male



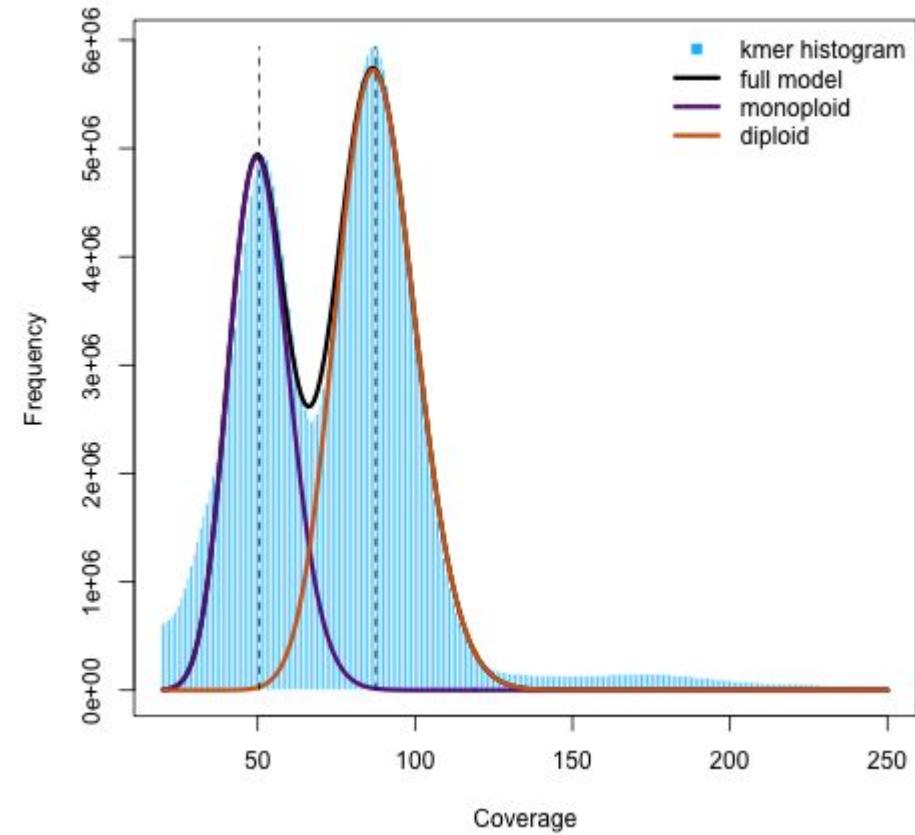
Need to model the 1n and 2n coverage as independent variables

```
Frequency ~ ( α * NB(kcov, kcov / bias) +  
              (1 - α) * NB(kcov2, kcov2 / bias)) * length
```

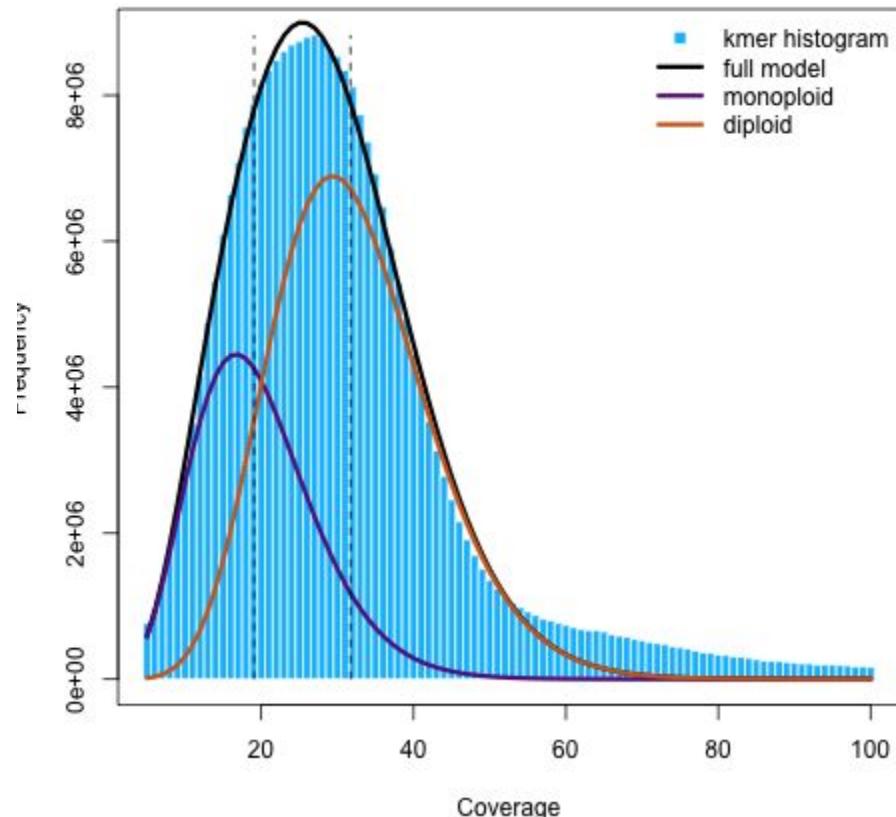
```
kmerEst <- 50  
PGE_model <- nlsLM(y ~ ((het * dnbinom(x, size = kmercov / bias, mu = kmercov)) +  
                           ((1 - het) * dnbinom(x, size = kmercov2 / bias, mu = kmercov2))) * length,  
                     start = list(kmercov = kmerEst, kmercov2 = (2*kmerEst), bias = 0.5, length = 280e6, het = 0.6),  
                     control = list(minFactor=1e-12, maxiter=40))
```

Simple PGE model fit the k-mer spectra really well

Afus1

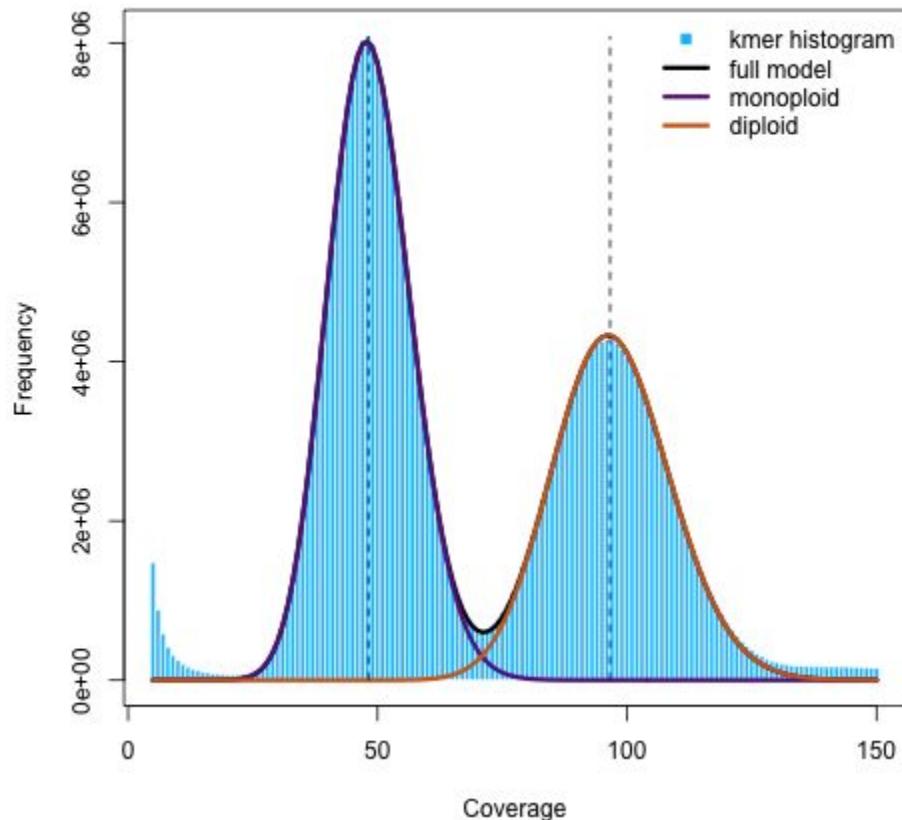


BH3-2



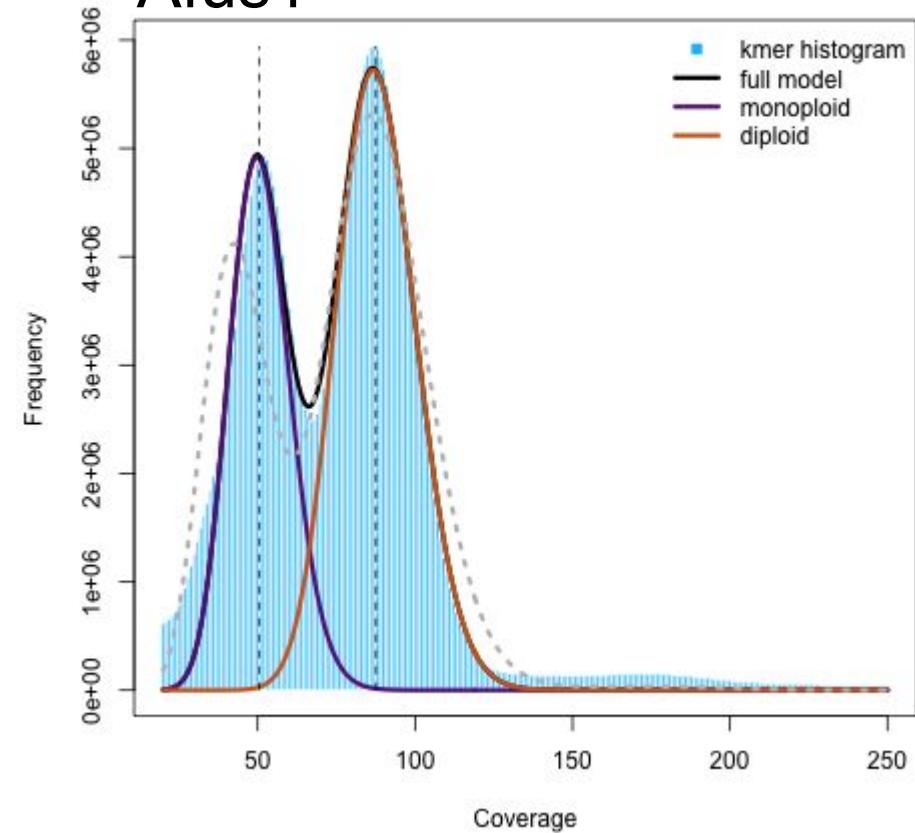
Simple PGE model fit the k-mer spectra of non-PGE species really well too

## Ocin2

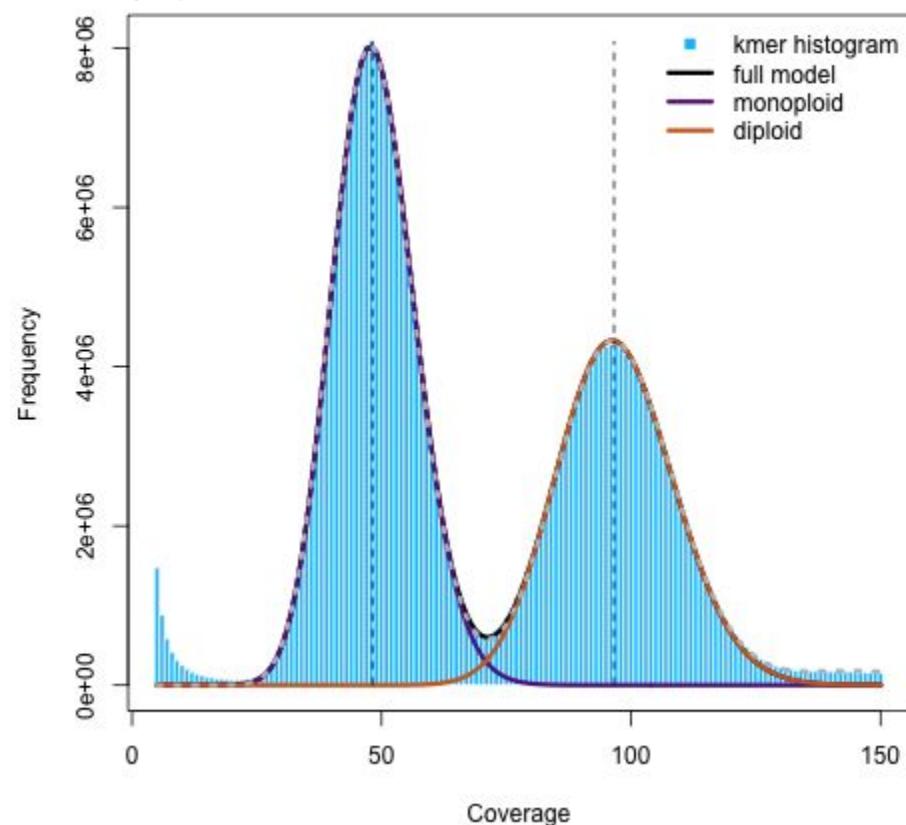


GenomeScope (gray dashed line) vs Simple PGE fits

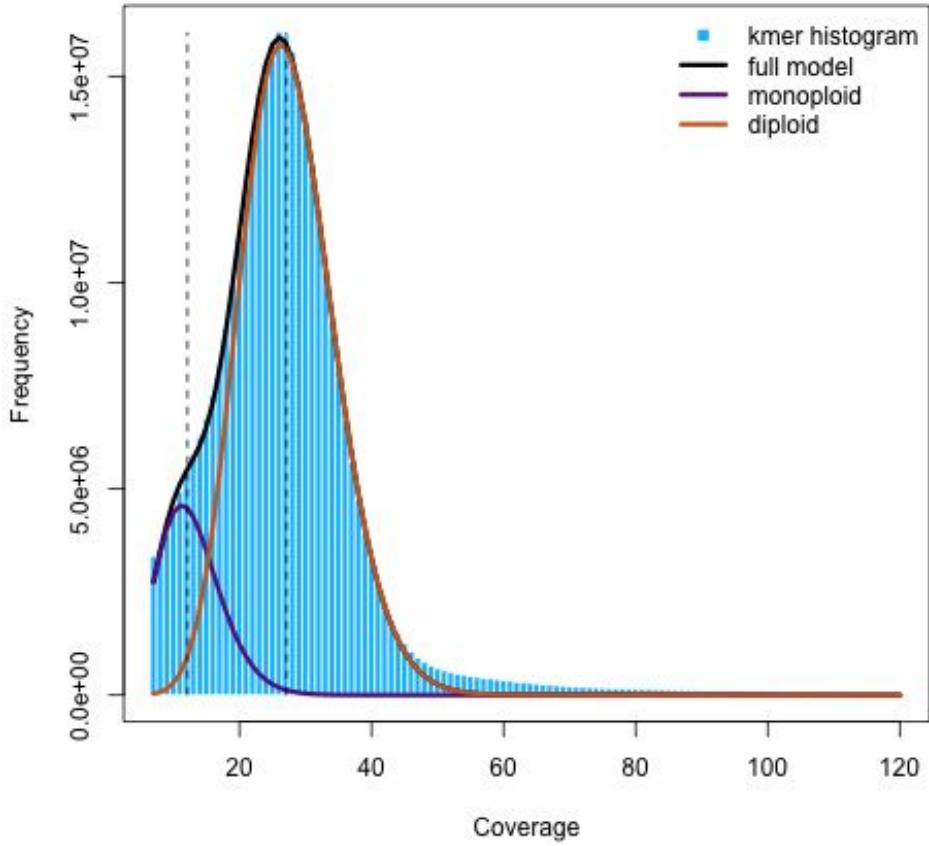
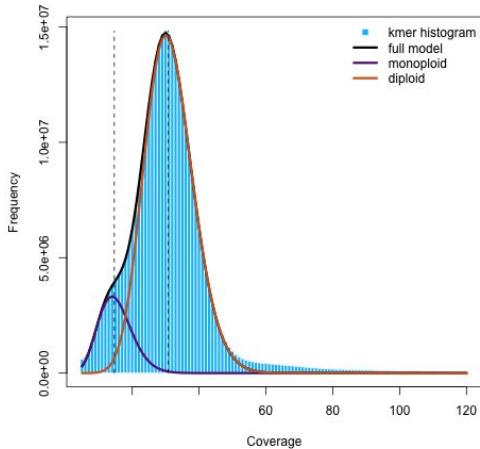
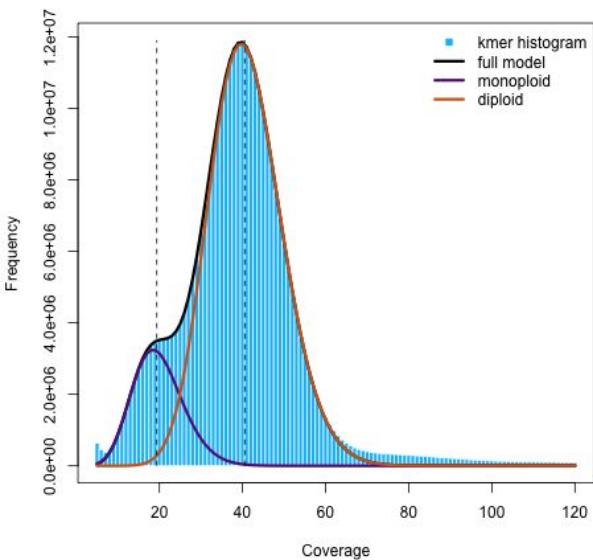
Afus1



Ocin2



# Females

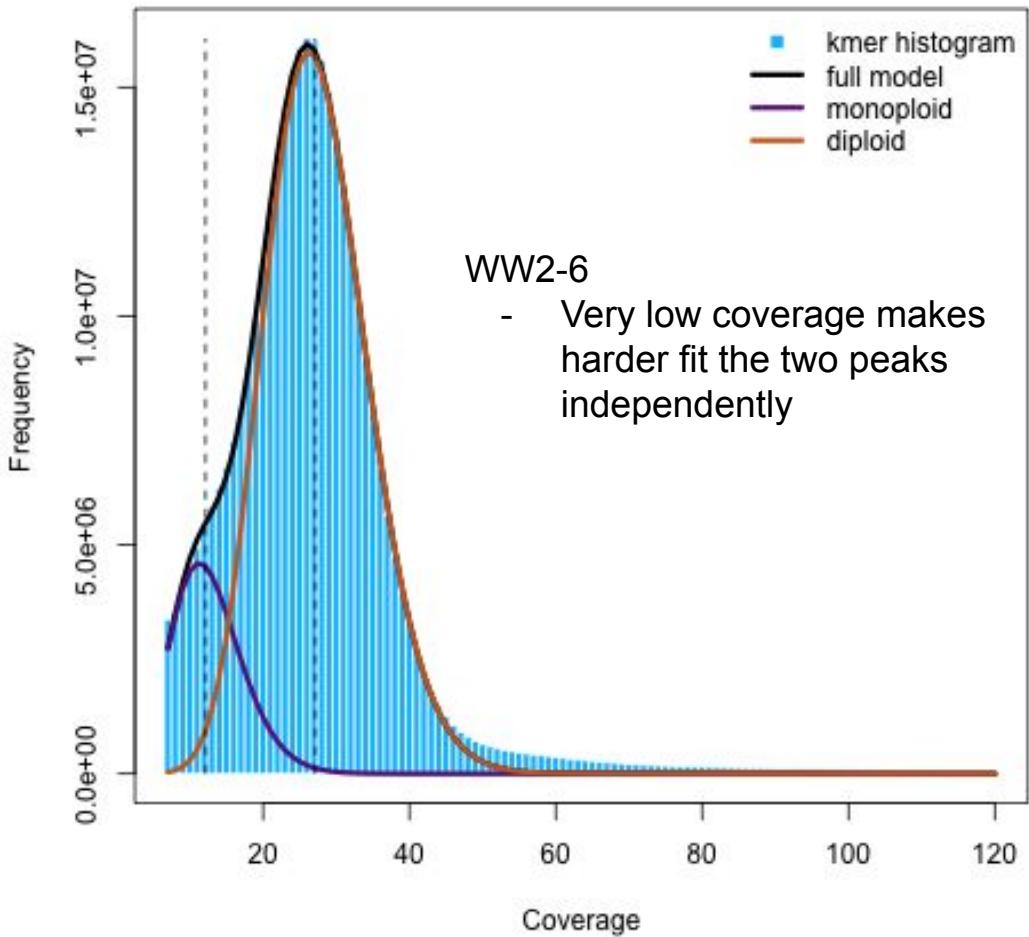


## Table of 1n and 2n est

ind	cov_1n	cov_2n
Afus1	51.05	88.05
BH3-2	19.55	32.21
Ocin2	48.68	97.09
WW5-3	15.11	31.23
WW5-5	19.85	41.05
WW2-6	12.61	27.59

## Table of 1n and 2n est

ind	cov_1n	cov_2n	Estimated sperm
Afus1	51.05	88.05	0.275
BH3-2	19.55	32.21	0.352
Ocin2	48.68	97.09	0.005
WW5-3	15.11	31.23	-0.066
WW5-5	19.85	41.05	-0.068
WW2-6	12.61	27.59	-0.232



Estimated sperm
0.275
0.352
0.005
-0.066
-0.068
-0.232

# Explicit PGE model

```
nls(y ~ ((1-(1-r)^k) * dnbinom(x, size = kmercov_p / bias, mu = kmercov_p) + heterozygous
```

```
(1-(1-r)^k) * dnbinom(x, size = kmercov_m / bias, mu = kmercov_m) + heterozygous
```

```
((1-r)^k) * dnbinom(x, size = (kmercov_p + kmercov_m) / bias, mu = kmercov_p + kmercov_m) * homozygous
```

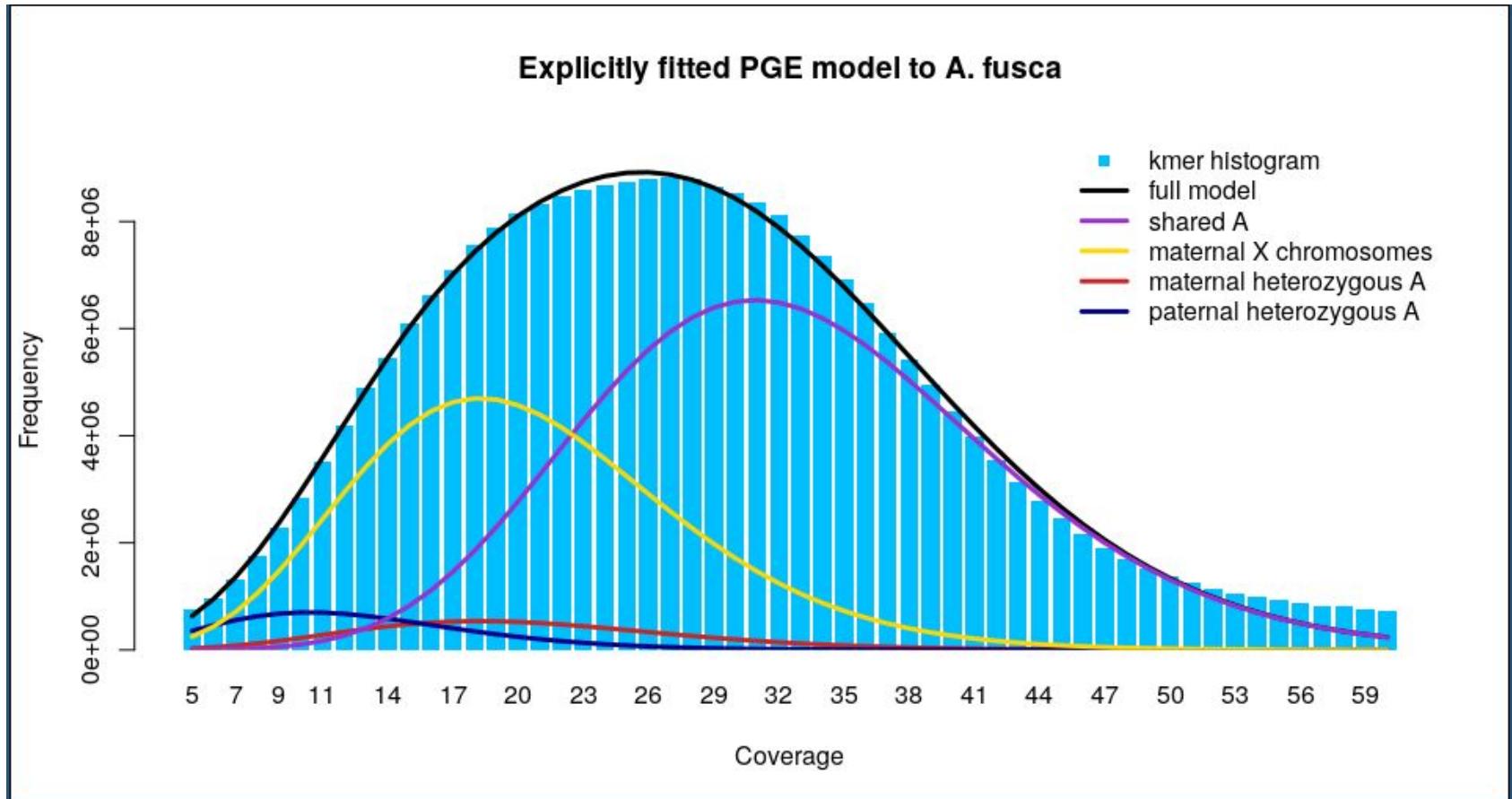
```
length * fraction_diploid + Autosomal
```

```
dnbinom(x, size = kmercov_m / bias, mu = kmercov_m) * X chromosome(s)  
length * (1 - fraction_diploid),
```

```
start = list(kmercov_p=estKcov, kmercov_m=estKcov, bias=bEst, length=estLen, fraction_diploid = estF),
```

```
control = list(minFactor=1e-12, maxiter=max_iterations))
```

# Explicit PGE model fit to BH3-2



# Summary

- K-mer spectra is an extremely efficient and powerful way to look at genomic data
- We can construct alternative genome models. Some pointers:
  - keep the parametric space sane
  - co-estimate parameters whenever we can, fix parameters when not
  - model fit is extremely dependent on initial values
- We are developing **GenomeTelescope** to facilitate the alternative genome models
  - <https://github.com/KamilSJaron/GenomeTelescope>
  - Two Tissue Model could potentially tell us about GRCs already, but likely we will need a better model for that