

Detection of Ploidy Levels

*Hidden Markov Model Based on
Genotype Likelihoods & Sequencing
Depth*

Samuele Soraggi - KU
Matteo Fumagalli - ICL



MOLECULAR ECOLOGY

[Explore this journal >](#)

ORIGINAL ARTICLE

Genomic evidence for population-specific responses to co-evolving parasites in a New Zealand freshwater snail

Laura Bankers [✉](#), Peter Fields, Kyle F. McElroy, Jeffrey L. Roore, John M. Logsdon Jr., Maurine Neill

First published: 10 May 2017 Full publication available

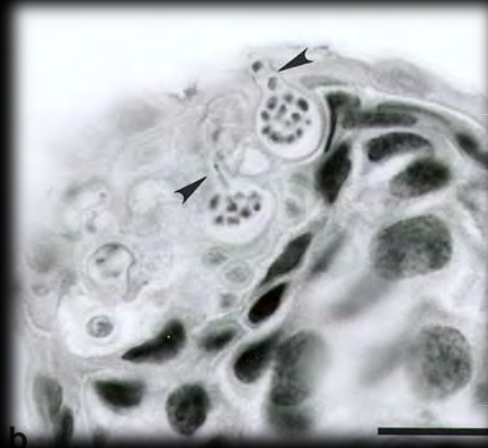
DOI: 10.1111/mec.14146 [View/save citation](#)



Cytometry

PLOIDY

Number of possible alleles at a locus



FROG FUNGI
Diploid assumption

nature
COMMUNICATIONS

Article | OPEN

Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation

Alison E. Wright [✉](#), Iulia Darolti, Natasha I. Bloch, Vicencio Oostra, Ben Sandkam, Severine D. Buechel, Niclas Kolm, Felix Breden, Beatriz Vicoso & Judith E. Mank

Likelihood Ratio Test



PLOIDY

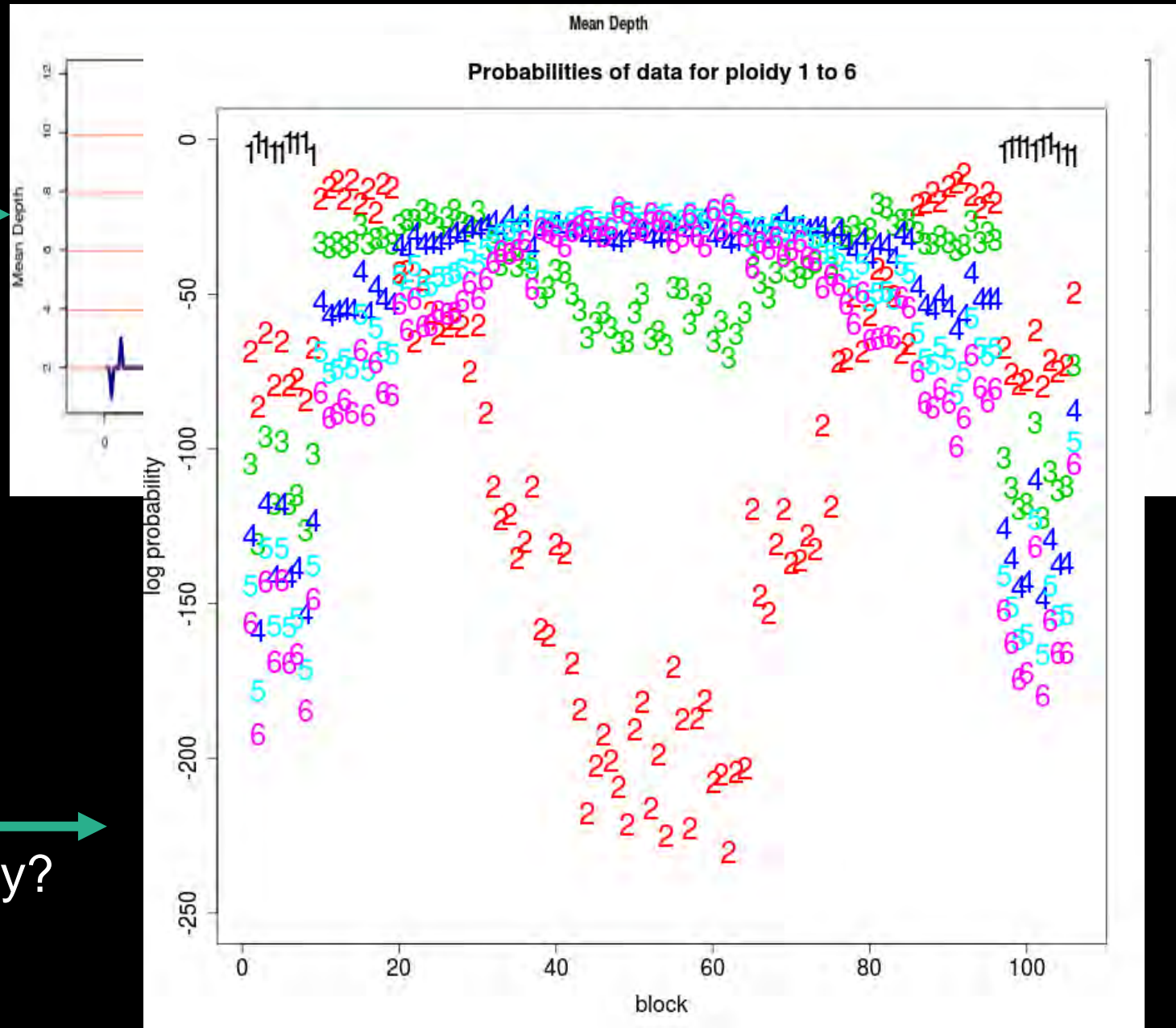
Number of possible alleles at a locus

Probabilities of sequenced data

involve

Genotype likelihoods

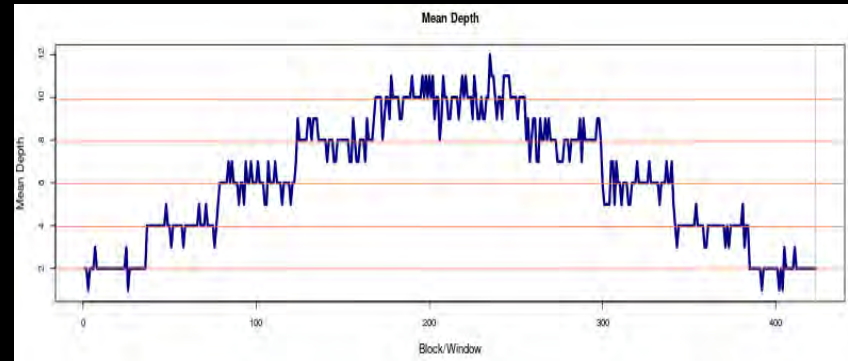
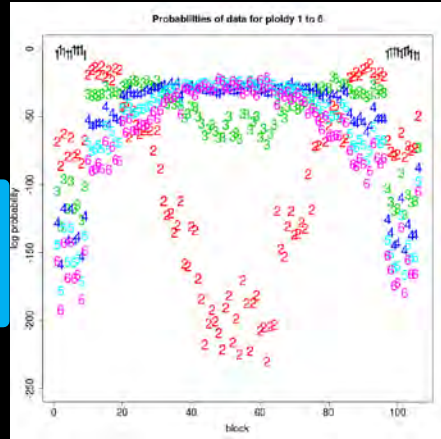
why?



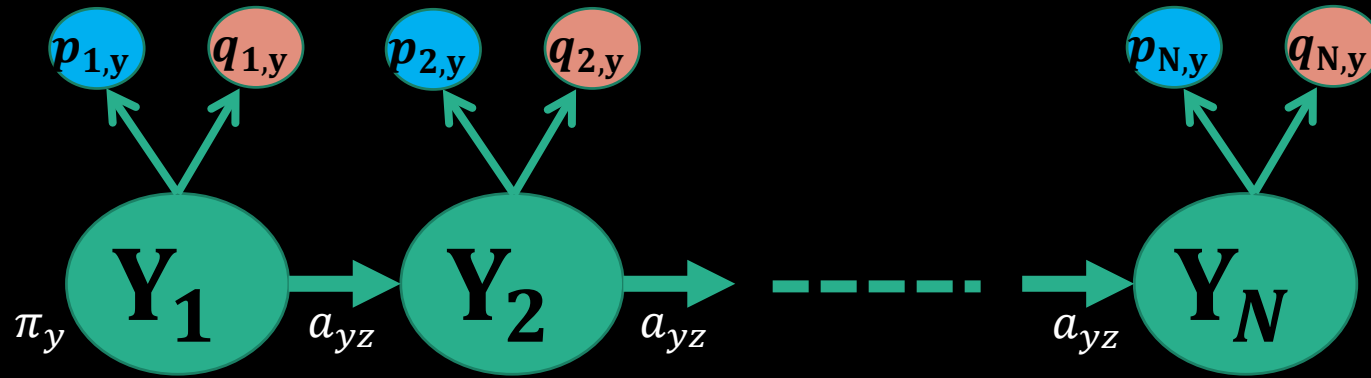
MODEL

Discrete Hidden Markov Model

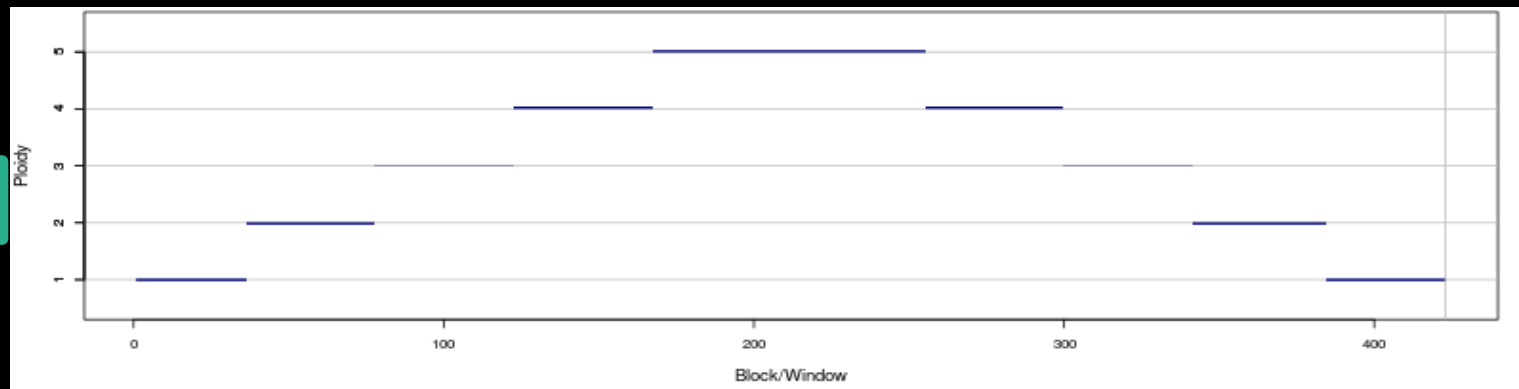
Prob of sequenced data



Distribution of depth



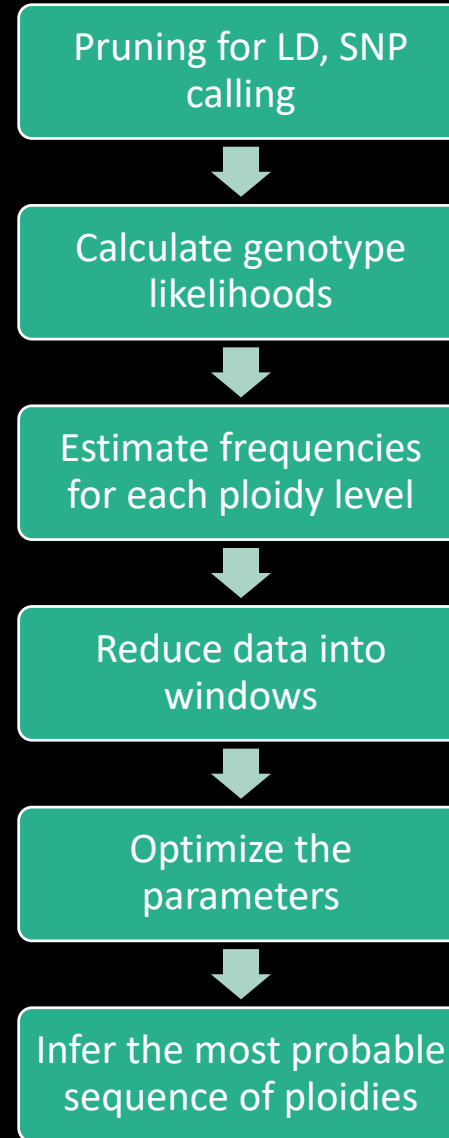
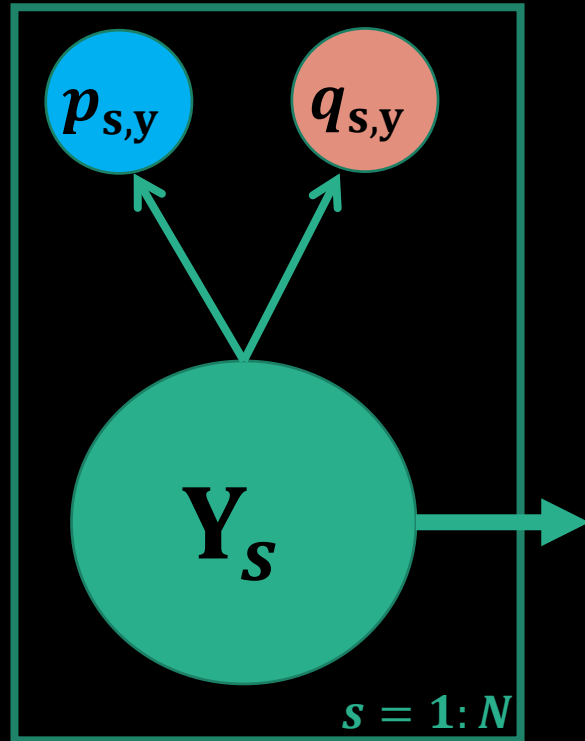
Unknown Ploidy



Other methods

- QuantiSNP PennCNV GenoCN etc. use a **continuous** HMM
- More complex models – work well in detecting **CNVs** but...
- ... it is hard to deal with **LD**
- Do not use info on **genotypes** they detect CNVs

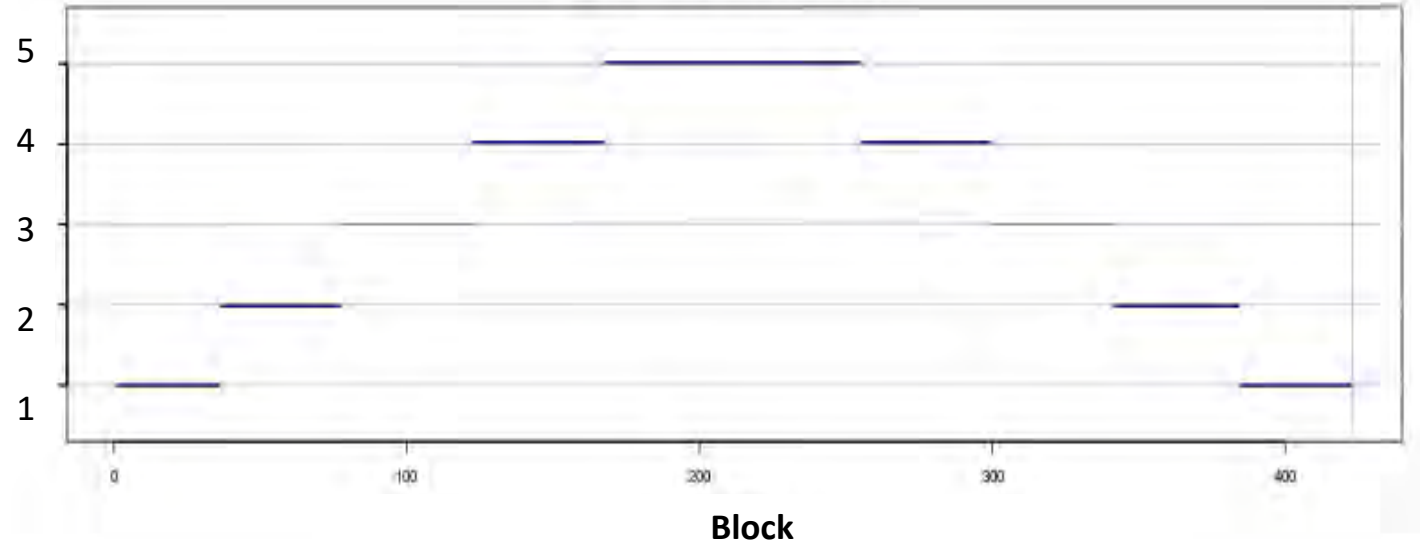
WORKFLOW



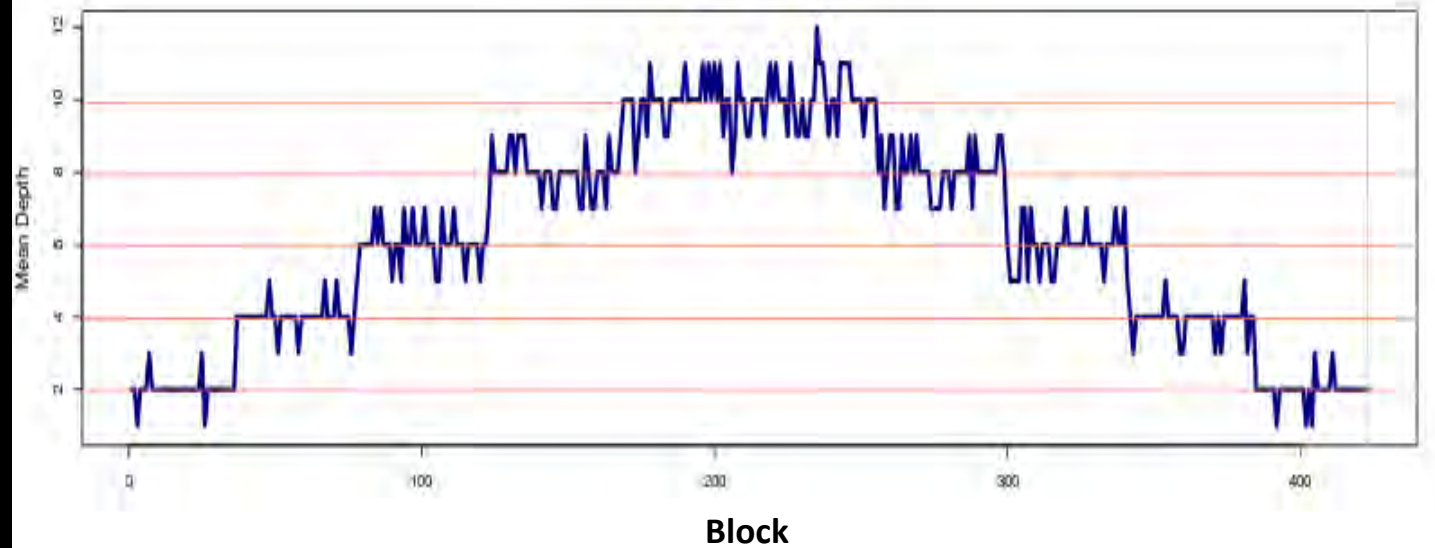
Results

- 20 Individuals
- 2X Depth for ploidy 1
- Small window (5 loci)

Inferred/Simulated Ploidy levels – Window size = 5

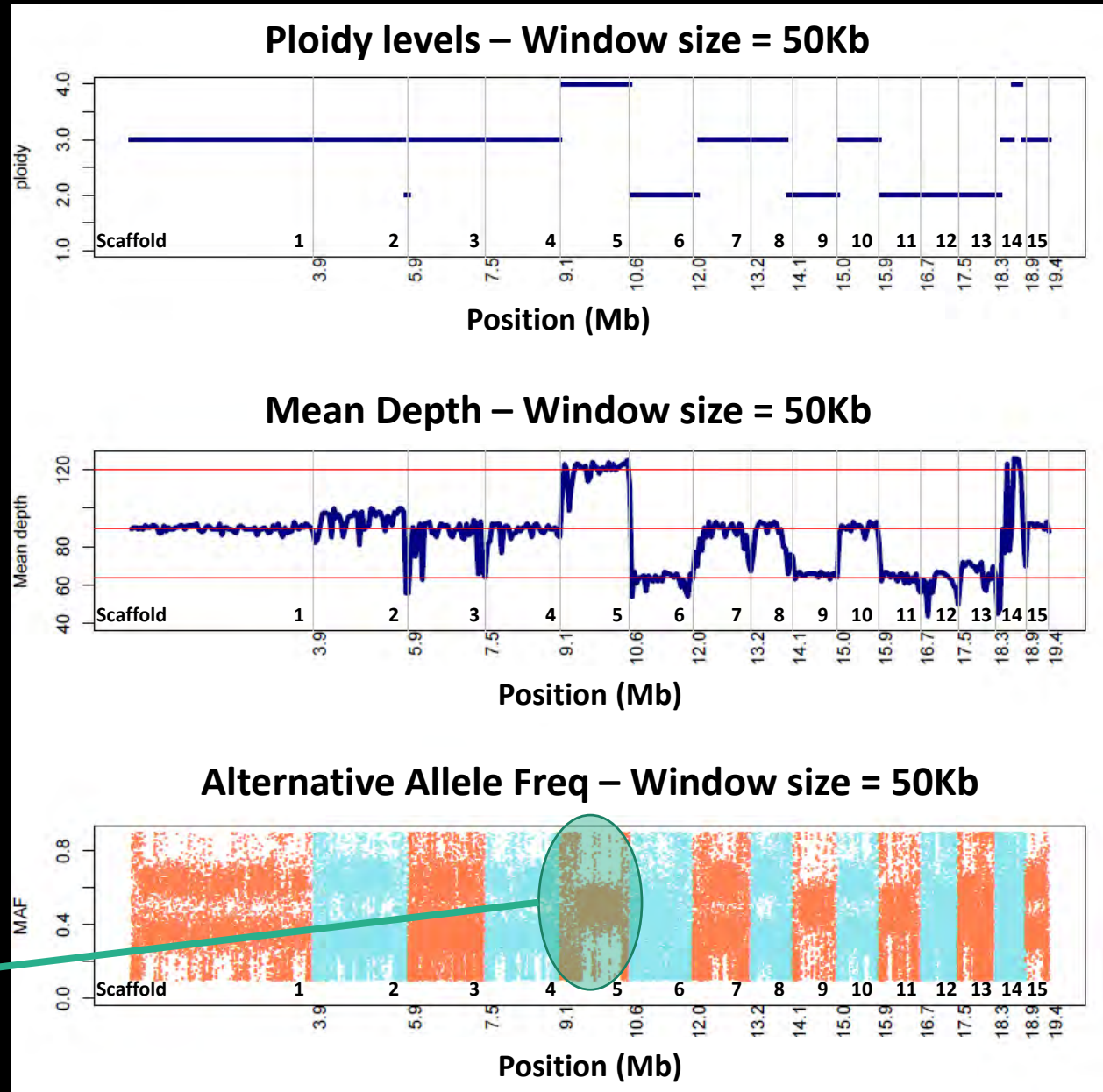


Mean Depth – Window size = 5



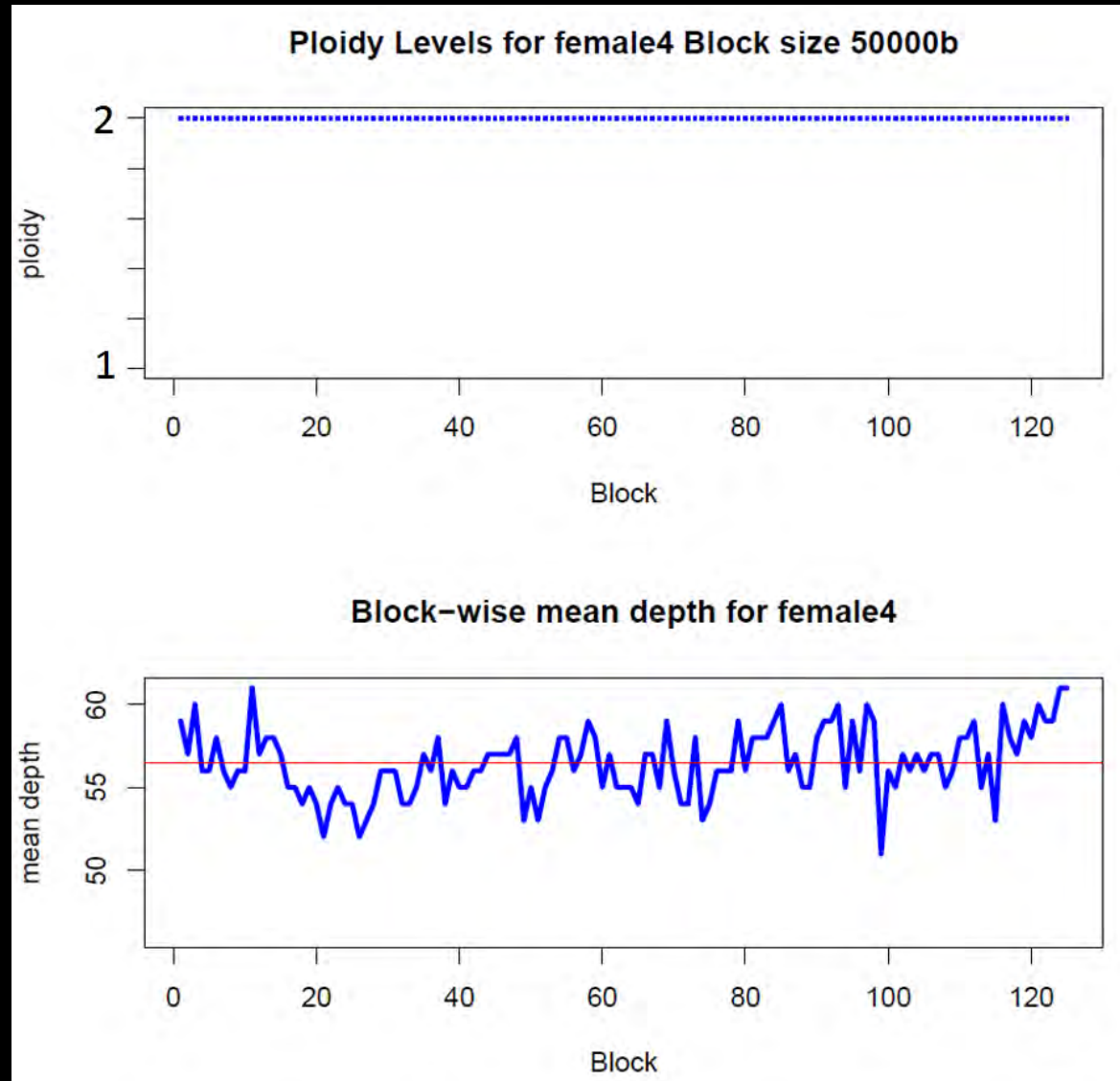
Results

- Real data (Fungi)
- ~20 Mb
- 50Kb window



Results

- Real data (Guppy fishes, 2 males, 1 female)
- Other statistical analysis: diploid males, (maybe) diploid female
- ~6 Mb
- 50Kb window



What this HMM does:

- Reduces the data into windows
- Considers genotype info to assign ploidy levels
- Starting from a maximum number of ploidy levels, filters out “not likely” states during the optimization process.

What it does NOT (yet?):

- Distinguish CNV and actual ploidy level



?

Thank you