

---

# Statistical Methods to Detect Sporadic Recombination in DNA Sequence Alignments

Dirk Husmeier

Biomathematics and Statistics Scotland  
at the Scottish Crop Research Institute  
Invergowrie, Dundee DD2 5DA, UK

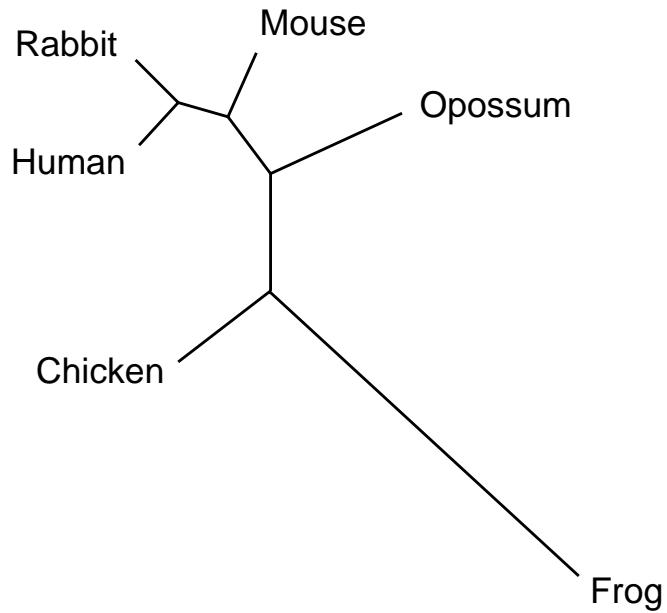
Email: [dirk@bioss.ac.uk](mailto:dirk@bioss.ac.uk)

<http://www.bioss.ac.uk/~dirk>

- Phylogenetics
- Sporadic recombination
- Statistical detection methods

# Phylogenetics

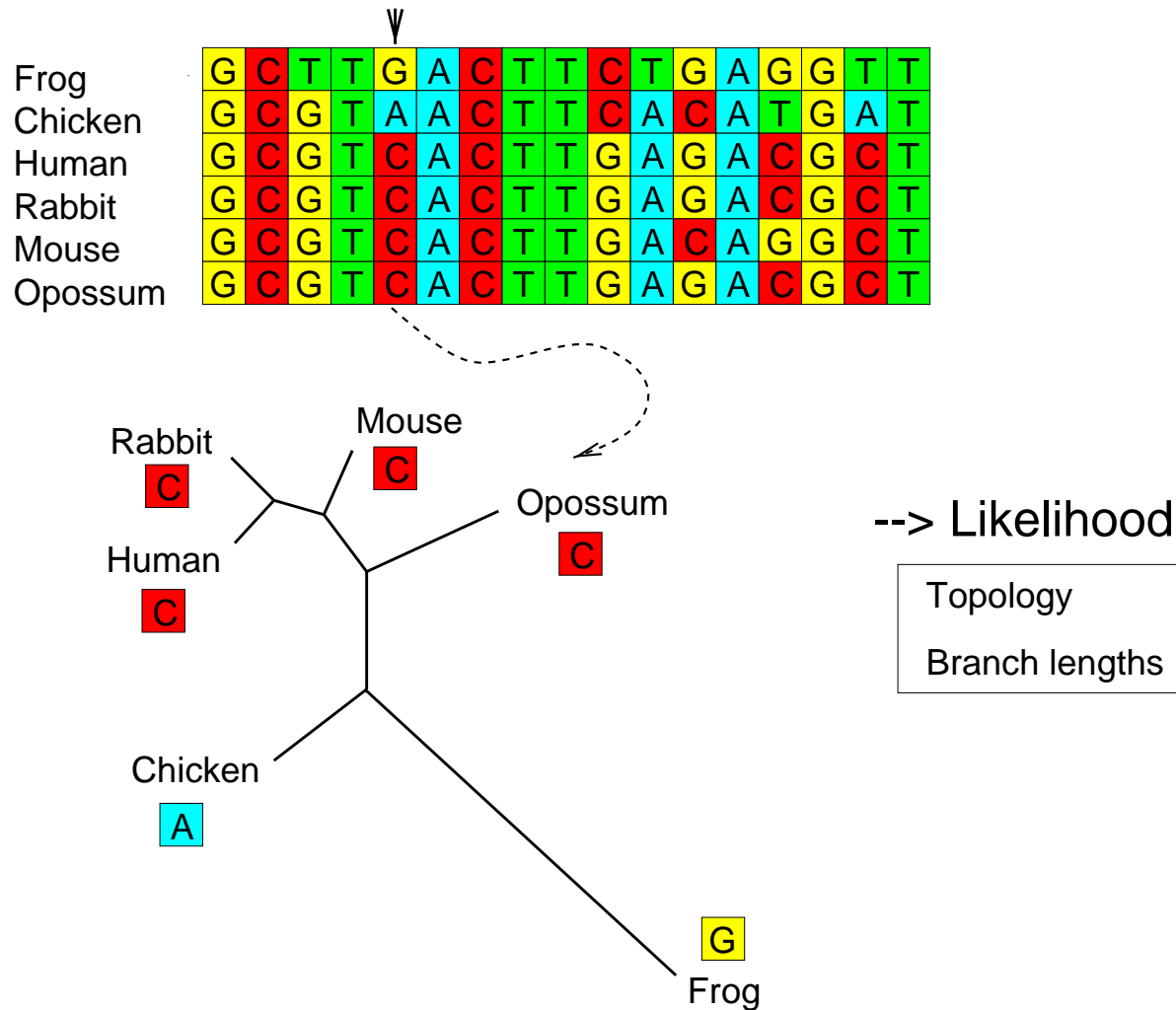
Frog	G	C	T	T	G	A	C	T	T	C	T	G	A	G	G	T	T
Chicken	G	C	G	T	A	A	C	T	T	C	A	C	A	T	G	A	T
Human	G	C	G	T	C	A	C	T	T	G	A	G	A	C	G	C	T
Rabbit	G	C	G	T	C	A	C	T	T	G	A	G	A	C	G	C	T
Mouse	G	C	G	T	C	A	C	T	T	G	A	C	A	G	G	C	T
Opossum	G	C	G	T	C	A	C	T	T	G	A	G	A	C	G	C	T



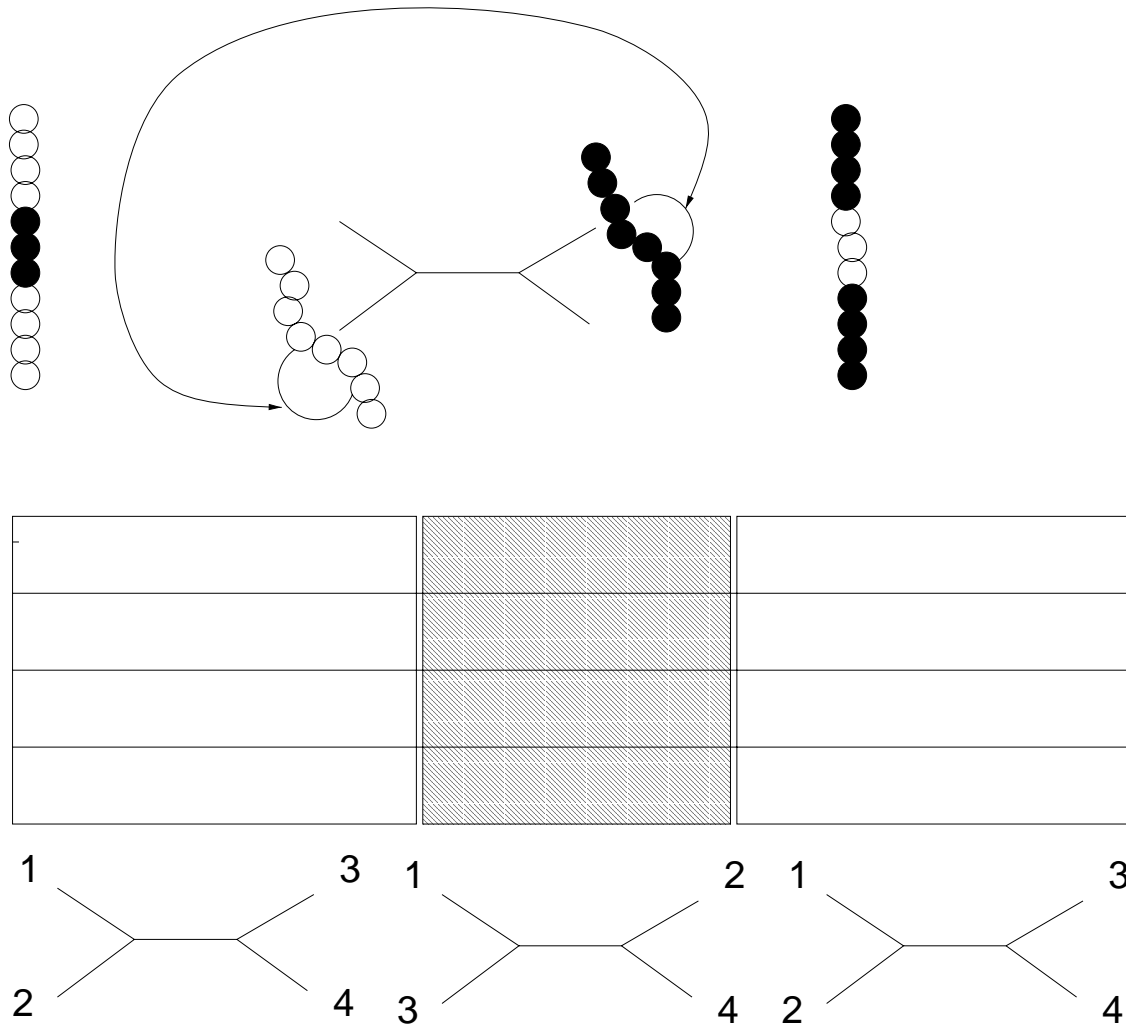
--> Topology

--> Branch lengths

# Statistical Approach to Phylogenetics



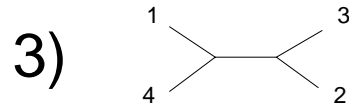
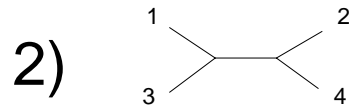
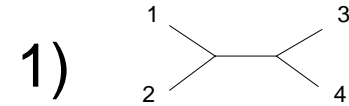
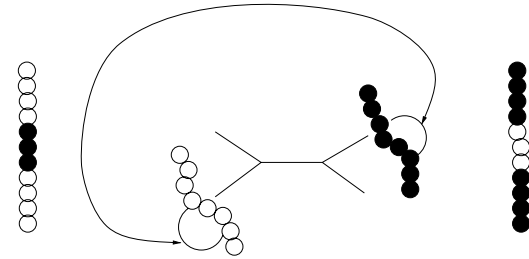
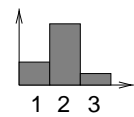
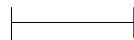
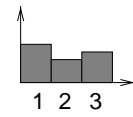
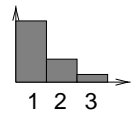
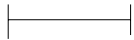
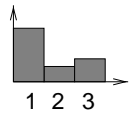
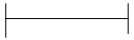
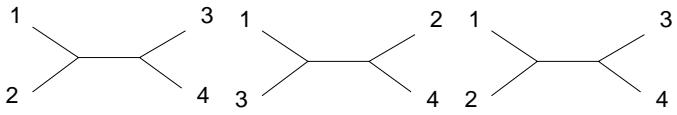
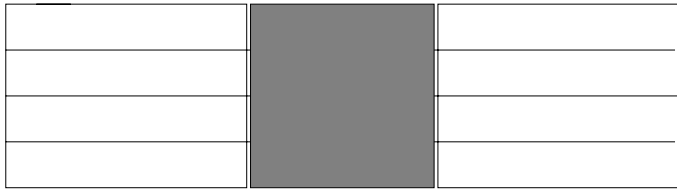
# Recombination



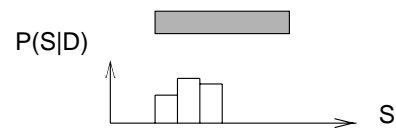
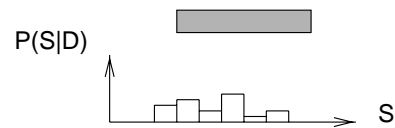
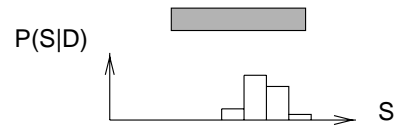
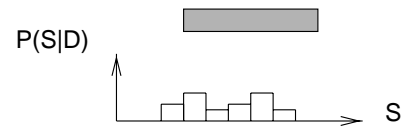
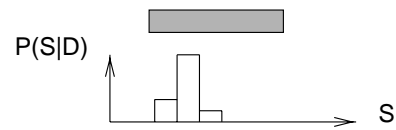
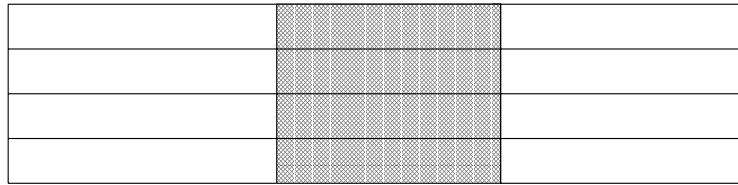
---

Detecting Recombination with a  
Sliding Window MCMC Method

Dirk Husmeier & Frank Wright  
*Bioinformatics*  
In Print

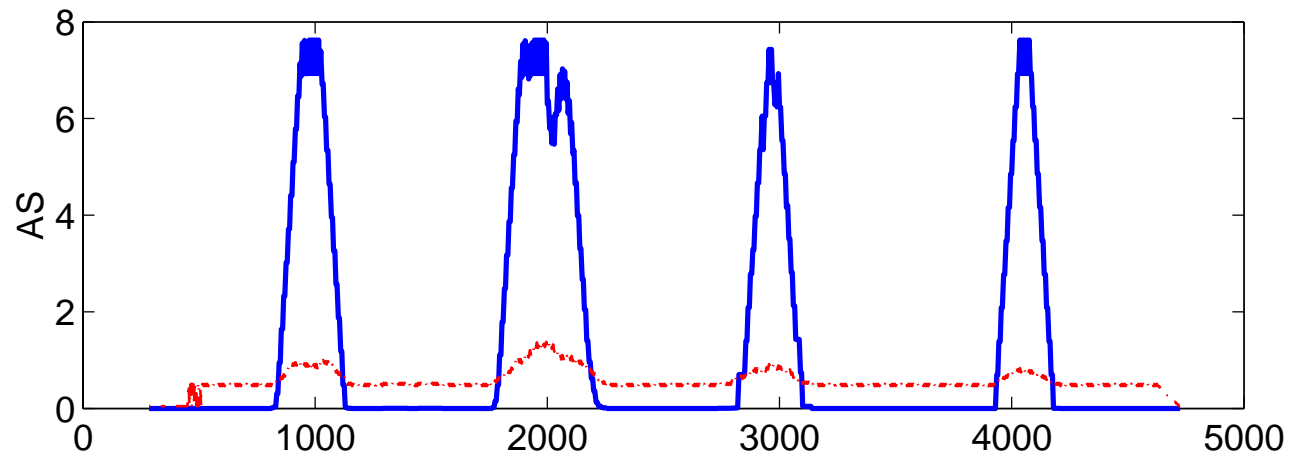
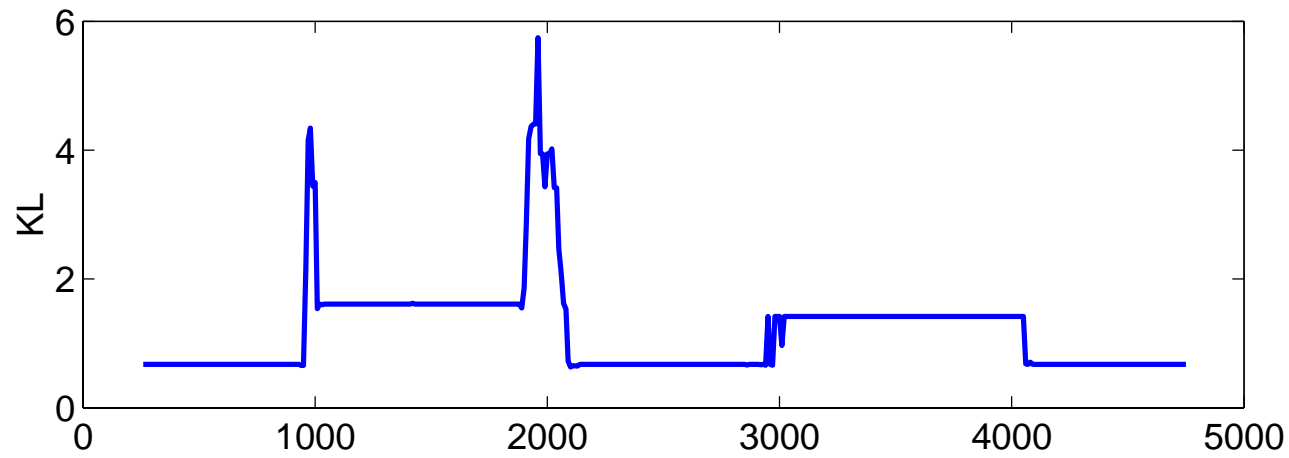


# Window Method: Illustration



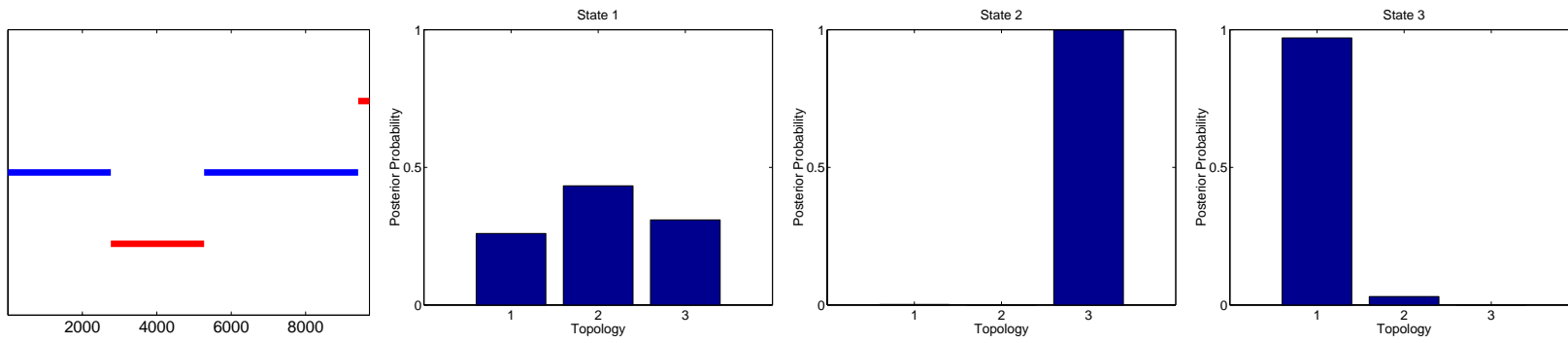
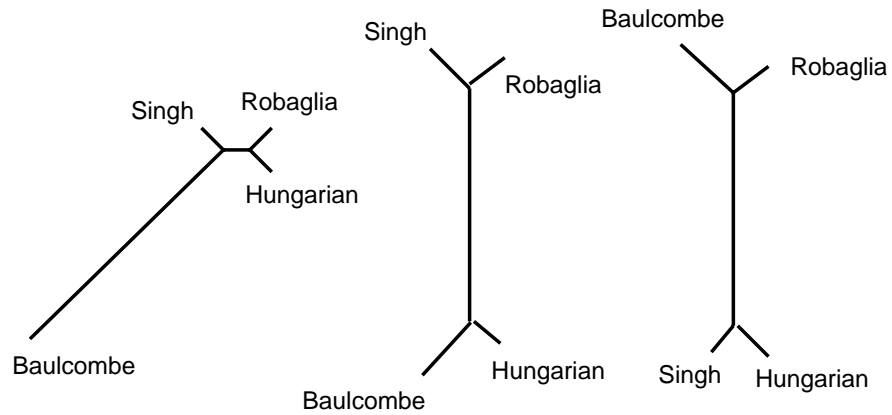
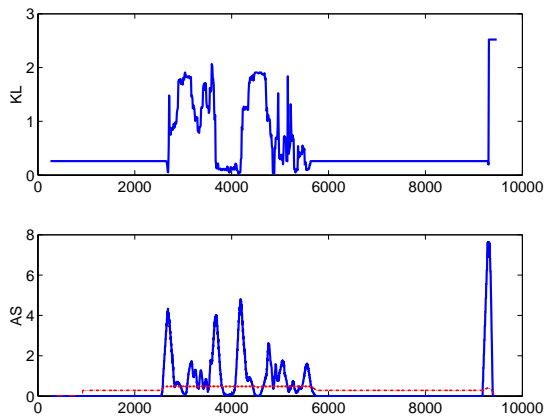
## Window Method: Illustration

---



# Potato Virus Y

(1) Hungarian, (2) Singh, (3) Robaglia, (4) Baulcombe

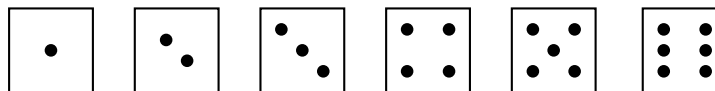
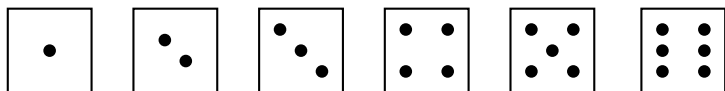
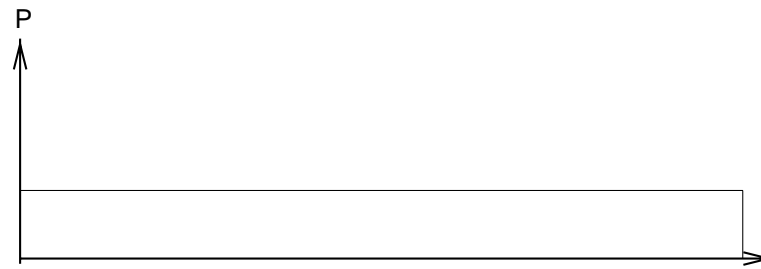
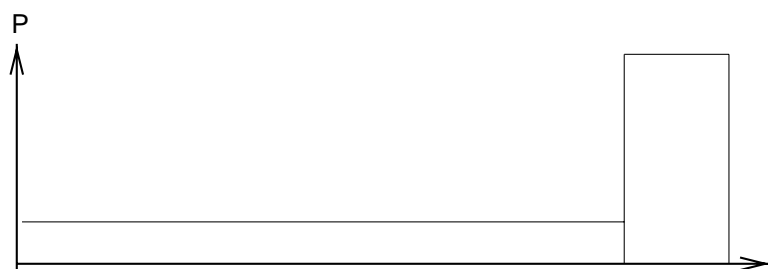
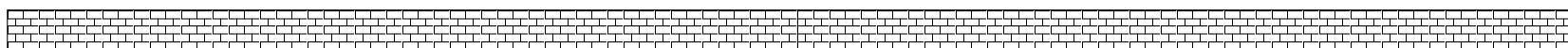
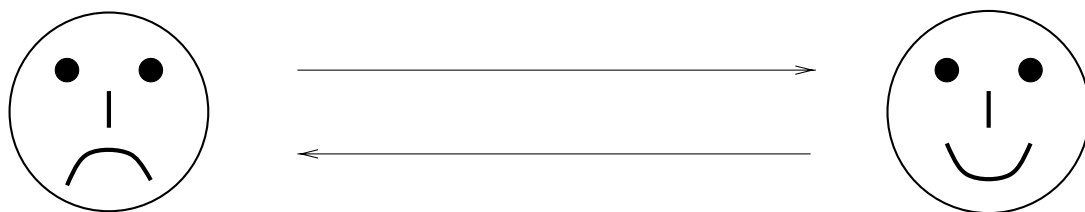


---

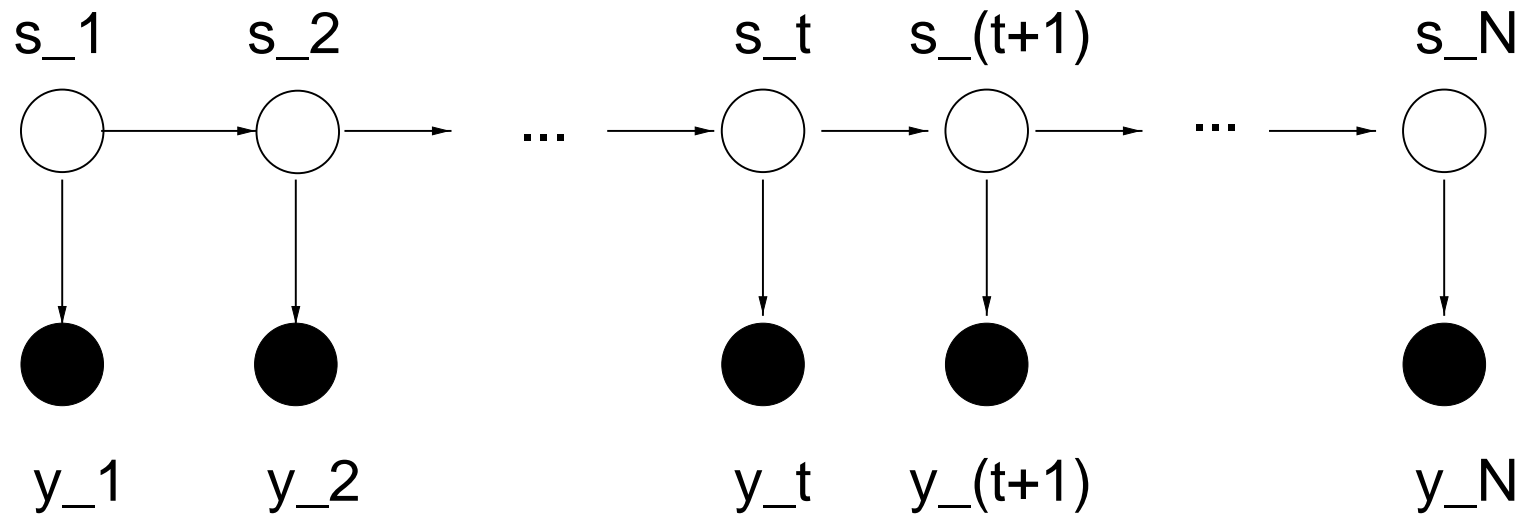
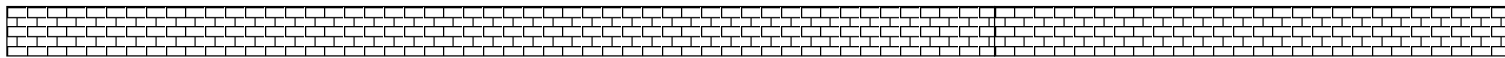
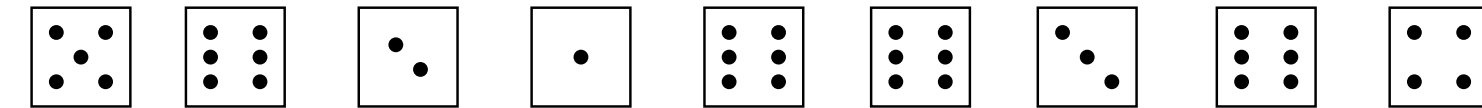
# Detecting Recombination with Hidden Markov Models

Dirk Husmeier & Frank Wright  
*Journal of Computational Biology*  
In Print

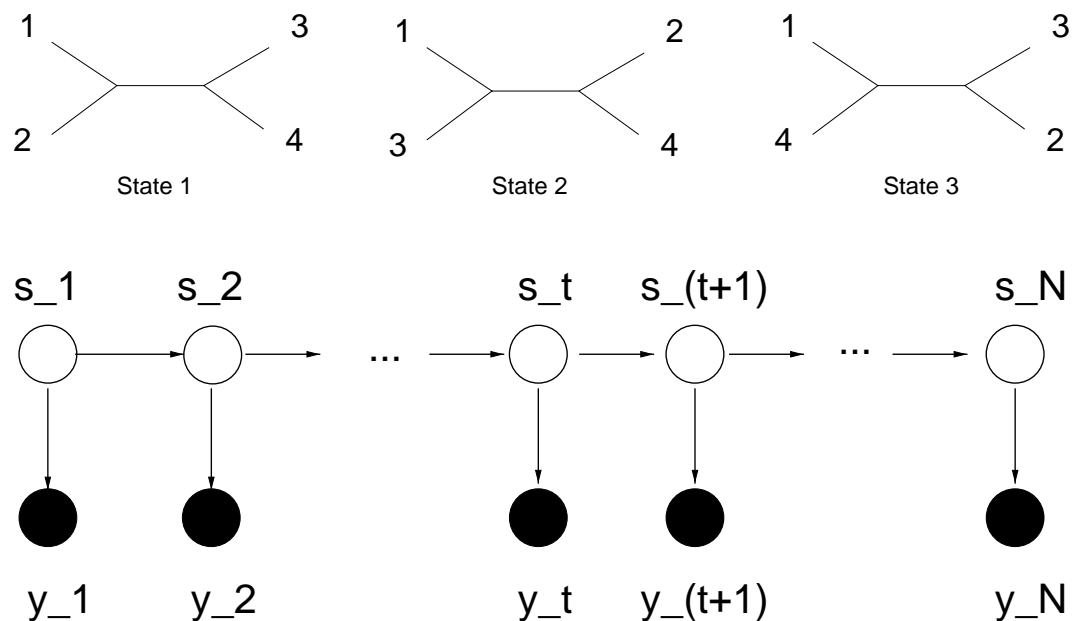
# Introduction to HMMs 1



## Introduction to HMMs 2



## Modelling Recombination with HMMs



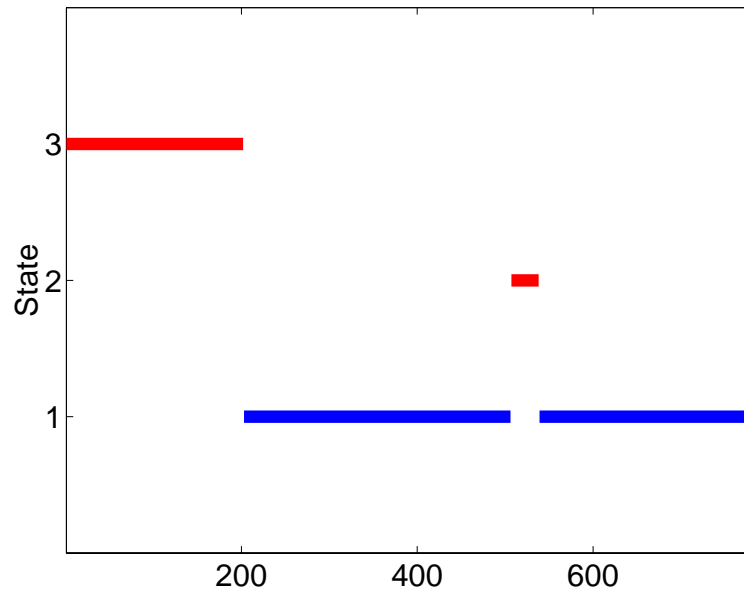
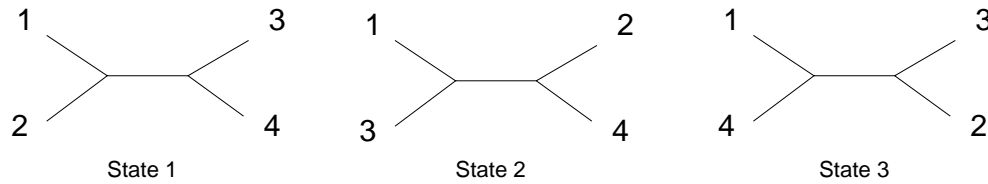
AG C ATCGTTCTATTTTACCGGCTCCCG  
 TG T GTCGCTCAAGATTGCCATCGCGCG  
 TG T CGTGGTCTAGATTGCCATCGCGCG  
 TG T ATCGCTCTAGTTTGCCAGCTCCCG

Find optimal sequence  $s_1, s_2, \dots, s_N \longrightarrow$  Maximise  $P(s_1, s_2, \dots, s_N | \mathbf{D})$

# Neisseria

DNA alignment, 787 nucleotides (argF gene)

- 1) Neisseria gonorrhoeae
- 2) Neisseria meningitidis
- 3) Neisseria cinerea
- 4) Neisseria mucosa



# Potato Virus Y

(1) Hungarian, (2) Singh, (3) Robaglia, (4) Baulcombe

