Towards a Polymer Model of Recombination

Christopher Penfold†,‡,
Alastair Goldman†, Neil Lawrence§, and Guido Sanguinetti‡.

†Department of Molecular Biology and Biotechnology and
‡Department of Computer Science, University of Sheffield;
§School of Computer Science, University of Manchester.
1 Introduction
   - Homologous Recombination
   - Meiosis at a Glance
   - Compartmentalisation

2 Model Formalisms
   - The Ideal Gaussian Chain
   - The Prior Distribution

3 Behaviour of the Prior Distribution
   - Stationary Formations
   - A Little Conjecture
   - Future Work
What is Homologous Recombination?

Shuffling of genetic material across the genome ... motivated by sequence homology.
What is Homologous Recombination?

Shuffling of genetic material across the genome ... motivated by sequence homology.

1. Allelic HR between corresponding locations on homologues.
What is Homologous Recombination?

- Shuffling of genetic material across the genome ... motivated by sequence homology.

1. Allelic HR between corresponding locations on homologues.
2. Ectopic HR due to chance similarities throughout the genome.
What is Homologous Recombination?

Shuffling of genetic material across the genome ... motivated by sequence homology.

1. Allelic HR between corresponding locations on homologues.
2. Ectopic HR due to chance similarities throughout the genome.
What is Homologous Recombination?

Shuffling of genetic material across the genome ... motivated by sequence homology.

1. Allelic HR between corresponding locations on homologues.
2. Ectopic HR due to chance similarities throughout the genome.

- Elevated rates during cellular division, e.g. meiosis.
Chromosome Dynamics

- Telomeres indicated by squares.
- Centromeres indicated as circles.
- Homologous pairs indicated by similar colour.

Figure: The Rabl formation during Leptotene.
Chromosome Dynamics

- Telomeres indicated by squares.
- Centromeres indicated as circles.
- Homologous pairs indicated by similar colour.
- Centromeres confined to the nuclear periphery (Rabl formation).

Figure: The Rabl formation during Leptotene.
Chromosome Dynamics

- Telomeres indicated by squares.
- Centromeres indicated as circles.
- Homologous pairs indicated by similar colour.

1. Centromeres confined to the nuclear periphery (Rabl formation).
2. Centromeres clustered about a common point.

Figure: The Rabl formation during Leptotene.
Chromosome Dynamics

- Telomeres indicated by squares.
- Centromeres indicated as circles.
- Homologous pairs indicated by similar colour.

1. Centromeres confined to the nuclear periphery (Rabl formation).
2. Centromeres clustered about a common point.
3. Paired sister chromatin.

Figure: The Rabl formation during Leptotene.
Figure: The Bouquet formation during Zygotene.
Telomeres attached to the nuclear periphery (Bouquet formation).

Figure: The Bouquet formation during Zygotene.
Figure: The Bouquet formation during Zygotene.

1. Telomeres attached to the nuclear periphery (Bouquet formation).
2. Telomeres clustered.

Christopher Penfold†,‡, Alastair Goldman†, Neil Lawrence§,‡ Towards a Polymer Model of Recombination
Fig. 1

1. Telomeres attached to the nuclear periphery (Bouquet formation).
2. Telomeres clustered.
3. Initiation of Synapsis (four strand bundles).

Figure: The Bouquet formation during Zygotene.
Introduction
Model Formalisms
Behaviour of the Prior Distribution

Homologous Recombination
Meiosis at a Glance
Compartmentalisation

Chromosome Dynamics

Figure: The Bouquet formation during Zygotene.

1. Telomeres attached to the nuclear periphery (Bouquet formation).
2. Telomeres clustered.
3. Initiation of Synapsis (four strand bundles).
4. Distinctly different from Rabl [Zickler and Kleckner, 1998].
Figure: Synapsed Chromosomes during Pachytene.
Chromosomes fully Synapsed.

**Figure:** Synapsed Chromosomes during Pachytene.
Chromosome Dynamics

Figure: Synapsed Chromosomes during Pachytene.

1. Chromosomes fully Synapsed.
2. Telomeres dispersive.
Chromosome Dynamics

Figure: Synapsed Chromosomes during Pachytene.

1. Chromosomes fully Synapsed.
2. Telomeres dispersive.
3. Resolution of crossovers.
Towards a Polymer Model of Recombination

[Bass et al.]
Influence of Chromosome Dynamics
Continuous repositioning is expected to influence recombination by “compartmentalising” loci with regard to each other.
Continuous repositioning is expected to influence recombination by “compartmentalising” loci with regard to each other.
Continuous repositioning is expected to influence recombination by “compartmentalising” loci with regard to each other.
Continuous repositioning is expected to influence recombination by “compartmentalising” loci with regard to each other.

Should be especially obvious at ectopic loci.
Continuous repositioning is expected to influence recombination by “compartmentalising” loci with regard to each other.

Should be especially obvious at ectopic loci.

Introduce artificial regions of homology at ectopic loci.
Experimental Motivation
Experimental Motivation

Efficiency of ectopic Recombination, [Goldman and Lichten, 1996].
Experimental Motivation

Efficiency of ectopic Recombination, [Goldman and Lichten, 1996].

\[ \varepsilon|i,j \propto \frac{A_{i,j}f_{i,j}}{f_{i,i} + f_{j,j}} \]  

(1)

where \( f_{i,j} \) is the frequency of recombination between inserts at loci \( i \) and \( j \) and \( A_{i,j} \) a constant accounting for inviable tetrads.
Efficiency of ectopic Recombination, [Goldman and Lichten, 1996].

$$\varepsilon | i, j \propto \frac{A_{i,j} f_{i,j}}{f_{i,i} + f_{j,j}}$$

(1)

where $f_{i,j}$ is the frequency of recombination between inserts at loci $i$ and $j$ and $A_{i,j}$ a constant accounting for inviable tetrads.
Efficiency of ectopic Recombination, [Goldman and Lichten, 1996].

\[ \varepsilon_{i,j} \propto \frac{A_{i,j} f_{i,j}}{f_{i,i} + f_{j,j}} \] (1)

where \( f_{i,j} \) is the frequency of recombination between inserts at loci \( i \) and \( j \) and \( A_{i,j} \) a constant accounting for inviable tetrads.
Model chromosomes \textit{a priori} as an ideal Gaussian chains.
Ideal Gaussian Chain

- Model chromosomes \textit{a priori} as an ideal Gaussian chains.
- More specifically the limiting case of random flight [Yamakawa, 2001, Kleinert, 2006].
Model chromosomes \textit{a priori} as an ideal Gaussian chains.

More specifically the limiting case of random flight [Yamakawa, 2001, Kleinert, 2006].
End-to-end distance distribution:

\[ P(R|N) = \left( \frac{3}{2\pi a^2 N} \right)^{3/2} \exp\left( \frac{|R|^2}{2a^2 N} \right) \]  \quad (2)

where \( N \) represents the number of links and \( a \) the mean monomer length.
Consider an $N$-link chromosome of length $Na$ with ends fixed at positions $x_O$ and $x_N$ respectively.
Consider an $N$-link chromosome of length $Na$ with ends fixed at positions $x_O$ and $x_N$ respectively.

Let the $n$th link pass through $x_n^1$. 
Consider an \( N \)-link chromosome of length \( Na \) with ends fixed at positions \( x_0 \) and \( x_N \) respectively.

Let the \( n \)th link pass through \( x_n^1 \).

We may represent this chromosome as two random flights:
Consider an $N$-link chromosome of length $Na$ with ends fixed at positions $x_O$ and $x_N$ respectively.

Let the $n$th link pass through $x_n^1$.

We may represent this chromosome as two random flights:

1. The first, $w_1$, originating at $x_O$ and terminating at $x_n^1$ after $n$ steps of length $a$. 
Consider an $N$-link chromosome of length $Na$ with ends fixed at positions $x_O$ and $x_N$ respectively.

Let the $n$th link pass through $x_n^1$.

We may represent this chromosome as two random flights:

1. The first, $w_1$, originating at $x_O$ and terminating at $x_n^1$ after $n$ steps of length $a$.
2. The second, $w_2$, beginning at $x_N$ and ending at $x_n^1$ after $(N - n)$-steps of length $a$. 
Consider an $N$-link chromosome of length $Na$ with ends fixed at positions $x_O$ and $x_N$ respectively.

Let the $n$th link pass through $x_n^1$.

We may represent this chromosome as two random flights:

1. The first, $w_1$, originating at $x_O$ and terminating at $x_n^1$ after $n$ steps of length $a$.
2. The second, $w_2$, beginning at $x_N$ and ending at $x_n^1$ after $(N - n)$-steps of length $a$.

Conditional distribution of the $n$th insert given end positions.
Consider an \( N \)-link chromosome of length \( Na \) with ends fixed at positions \( x_O \) and \( x_N \) respectively.

Let the \( n \)th link pass through \( x_n^1 \).

We may represent this chromosome as two random flights:

1. The first, \( w_1 \), originating at \( x_O \) and terminating at \( x_n^1 \) after \( n \) steps of length \( a \).
2. The second, \( w_2 \), beginning at \( x_N \) and ending at \( x_n^1 \) after \( (N - n) \)-steps of length \( a \).

Conditional distribution of the \( n \)th insert given end positions.

Similarly consider a second chromosome of length \( Ma \) with ends fixed respectively at \( x_P \) and \( x_M \) whose \( m \)th link passes through \( x_n^2 \).
Consider an $N$-link chromosome of length $Na$ with ends fixed at positions $x_O$ and $x_N$ respectively.
Let the $n$th link pass through $x_n^1$.
We may represent this chromosome as two random flights:
1. The first, $w_1$, originating at $x_O$ and terminating at $x_n^1$ after $n$ steps of length $a$.
2. The second, $w_2$, beginning at $x_N$ and ending at $x_n^1$ after $(N - n)$-steps of length $a$.
Conditional distribution of the $n$th insert given end positions.
Similarly consider a second chromosome of length $Ma$ with ends fixed respectively at $x_P$ and $x_M$ whose $m$th link passes through $x_n^2$.
Random walks $w_3$ and $w_4$. 
Consider an $N$-link chromosome of length $Na$ with ends fixed at positions $x_O$ and $x_N$ respectively. Let the $n$th link pass through $x_n^1$. We may represent this chromosome as two random flights:

1. The first, $w_1$, originating at $x_O$ and terminating at $x_n^1$ after $n$ steps of length $a$.
2. The second, $w_2$, beginning at $x_N$ and ending at $x_n^1$ after $(N - n)$-steps of length $a$.

Conditional distribution of the $n$th insert given end positions. Similarly consider a second chromosome of length $Ma$ with ends fixed respectively at $x_P$ and $x_M$ whose $m$th link passes through $x_n^2$. Random walks $w_3$ and $w_4$. Assume physical association is a necessary and sufficient condition for recombination.
$$E[p|X, i, j] = \mathcal{A}\mathcal{B} \int_{x_1^n \in V} \int_{x_2^n \in V} P(w_1 = x_1^n)P(w_2 = x_1^n)$$
$$P(w_3 = x_2^n)P(w_4 = x_2^n)\delta(|x_1^n - x_2^n| < \epsilon)dx_1^ndx_2^n,$$

(3)
The Prior Distribution

\[ \mathbb{E}[p|X, i, j] = AB \int_{x_1^n \in V} \int_{x_2^n \in V} P(w_1 = x_1^n) P(w_2 = x_1^n) P(w_3 = x_2^n) P(w_4 = x_2^n) \delta(|x_1^n - x_2^n| < \epsilon) dx_1^n dx_2^n, \]

(3)

where \( A \) and \( B \) are constants of normalization such that:

\[ A \int_{x_1^n \in V} P(w_1) P(w_2) dx_1^n = 1, \]

(4)

\[ B \int_{x_2^n \in V} P(w_3) P(w_4) dx_2^n = 1. \]

(5)
Chromosome Dynamics Revisited
Chromosome Dynamics Revisited

- Telomere’s confined to the nuclear periphery for majority of Prophase I.
Chromosome Dynamics Revisited

- Telomere’s confined to the nuclear periphery for majority of Prophase I.
- Consider a von Mises-Fisher distribution:

\[
f(X|\mu, \kappa) = C_p(\kappa) \exp(\kappa \mu^T X),
\]

where \(\mu\) is the mean vector, \(\kappa\) the spread parameter and \(C_p(\kappa)\) a normalising constant.
Chromosome Dynamics Revisited

- Telomere’s confined to the nuclear periphery for majority of Prophase I.
- Consider a von Mises-Fisher distribution:

  \[ f(X|\mu, \kappa) = C_p(\kappa) \exp(\kappa \mu^T X), \]  

  where \( \mu \) is the mean vector, \( \kappa \) the spread parameter and \( C_p(\kappa) \) a normalising constant.
- Time-dependant spread parameter, \( \kappa = \kappa(t) \).
Chromosome Dynamics Revisited

- Telomere’s confined to the nuclear periphery for majority of Prophase I.
- Consider a von Mises-Fisher distribution:

\[ f(X|\mu, \kappa) = C_p(\kappa) \exp(\kappa \mu^T X), \]  

where \( \mu \) is the mean vector, \( \kappa \) the spread parameter and \( C_p(\kappa) \) a normalising constant.
- Time-dependant spread parameter, \( \kappa = \kappa(t) \).
- Given *in vivo* data could estimate maximal Likelihood solution, e.g. [Banerjee et al., 2004].

Christopher Penfold†,‡, Alastair Goldman†, Neil Lawrence§, Towards a Polymer Model of Recombination
Chromosome Dynamics Revisited

- Telomere’s confined to the nuclear periphery for majority of Prophase I.
- Consider a von Mises-Fisher distribution:

\[ f(\mathbf{X}|\mu, \kappa) = C_p(\kappa) \exp(\kappa \mu^T \mathbf{X}), \]  

where \( \mu \) is the mean vector, \( \kappa \) the spread parameter and \( C_p(\kappa) \) a normalising constant.
- Time-dependant spread parameter, \( \kappa = \kappa(t) \).
- Given \textit{in vivo} data could estimate maximal Likelihood solution, e.g. [Banerjee et al., 2004].
- Analytically intractable even for stationary boundary conditions.
Chromosome Dynamics Revisited

- Telomere’s confined to the nuclear periphery for majority of Prophase I.
- Consider a von Mises-Fisher distribution:

\[ f(X|\mu, \kappa) = C_p(\kappa) \exp(\kappa \mu^T X), \tag{6} \]

where \( \mu \) is the mean vector, \( \kappa \) the spread parameter and \( C_p(\kappa) \) a normalising constant.
- Time-dependant spread parameter, \( \kappa = \kappa(t) \).
- Given \textit{in vivo} data could estimate maximal Likelihood solution, e.g. [Banerjee et al., 2004].
- Analytically intractable even for stationary boundary conditions.
- Optimal histogram, smoothing etc. [Shinomoto et al., 2006].
Is the Prior Sensible?

- Clustered Centromeres:

Figure: Recombination frequencies for Rabl formation. (Left) Data from Goldman et al. (1996) and (Right) model data ($\kappa = 100$).
Is the Prior Sensible?

- Clustered Centromeres:

Figure: Recombination frequencies for bouquet formation. (Left) Data from Goldman et al. (1996) and (Right) model data ($\kappa = 100$).
Is the Prior Sensible?

- Clustered Telomeres:

Figure: Recombination frequencies for bouquet formation. (Left) Data from Goldman et al. (1996) and (Right) model data ($\kappa = 100$).
Is the Prior Sensible?

Clustered Telomeres:

Figure: Recombination frequencies for bouquet formation. (Left) Data from Goldman et al. (1996) and (Right) model data ($\kappa = 10$).
Is the Prior Sensible?

- Clustered Telomeres:

**Figure:** Recombination frequencies for bouquet formation. (Left) Data from Goldman et al. (1996) and (Right) model data ($\kappa = 100$).
Preliminary Conclusions

- Rabl promotes allelic interaction near centromeres.
Preliminary Conclusions

- Rabl promotes allelic interaction near centromeres.
- Bouquet promotes allelic interaction near telomeres.
Rabl promotes allelic interaction near centromeres.
Bouquet promotes allelic interaction near telomeres.
Away from “promoted” regions ectopic efficiencies approach allelic ...
Preliminary Conclusions

- Rabl promotes allelic interaction near centromeres.
- Bouquet promotes allelic interaction near telomeres.
- Away from “promoted” regions ectopic efficiencies approach allelic ... though frequencies fall exponentially.
Preliminary Conclusions

- Rabl promotes allelic interaction near centromeres.
- Bouquet promotes allelic interaction near telomeres.
- Away from “promoted” regions ectopic efficiencies approach allelic ... though frequencies fall exponentially.
- Clustering heavily promotes interaction.
Preliminary Conclusions

- Rabl promotes allelic interaction near centromeres.
- Bouquet promotes allelic interaction near telomeres.
- Away from “promoted” regions ectopic efficiencies approach allelic ... though frequencies fall exponentially.
- Clustering heavily promotes interaction.

... a little conjecture

“Rabl is responsible for the early association between homologous pairs and encourages interaction between centromeric regions. Such early association may reduce the homology search space for later stages of meiosis.

The Bouquet promotes interaction between homologues at telomerically located regions. Such interaction suggest telomeres are a likely candidate locus for the initiation of Synapsis.”
... and now for the Bayesian part!

- We have data!
We have data!

Likelihood: \( L(p|\mathcal{D}) = \binom{N}{n} p^n (1 - p)^{N-n} \).
... and now for the Bayesian part!

- We have data!
- Likelihood: \( L(p|D) = \binom{N}{n} p^n (1 - p)^{N-n} \).
- Evaluate Posterior distribution.
... and now for the Bayesian part!

- We have data!
- Likelihood: \( L(p|\mathcal{D}) = \binom{N}{n} p^n (1-p)^{N-n} \).
- Evaluate Posterior distribution.
- Allow distribution to vary by multiplicative constant, \( a \).
- Maximal likelihood solution to \( a \).
We have data!

Likelihood: \( L(p|\mathcal{D}) = \binom{N}{n} p^n (1 - p)^{N-n} \).

Evaluate Posterior distribution.

Allow distribution to vary by multiplicative constant, \( a \).

Maximal likelihood solution to \( a \).

Make inferences about the Frequency and Efficiency of ectopic Recombination.
... and now for the Bayesian part!

- **We have data!**
- Likelihood: \( L(p|D) = \binom{N}{n} p^n (1 - p)^{N-n} \).
- Evaluate Posterior distribution.
- Allow distribution to vary by multiplicative constant, \( a \).
- Maximal likelihood solution to \( a \).
- Make inferences about the Frequency and Efficiency of ectopic Recombination.
- Find maximal likelihood solution with regards to a number of intermediate chromosome formations.
... and now for the Bayesian part!

- We have data!
- Likelihood: \( L(p|D) = \binom{N}{n} p^n (1 - p)^{N-n} \).
- Evaluate Posterior distribution.
- Allow distribution to vary by multiplicative constant, \( a \).
- Maximal likelihood solution to \( a \).
- Make inferences about the Frequency and Efficiency of ectopic Recombination.
- Find maximal likelihood solution with regards to a number of intermediate chromosome formations.
- Mutant systems: less data, prior distribution more important.
... and now for the Bayesian part!

- We have data!
- Likelihood: \( L(p|\mathcal{D}) = \binom{N}{n} p^n (1 - p)^{N-n} \).
- Evaluate Posterior distribution.
- Allow distribution to vary by multiplicative constant, \( a \).
- Maximal likelihood solution to \( a \).
- Make inferences about the Frequency and Efficiency of ectopic Recombination.
- Find maximal likelihood solution with regards to a number of intermediate chromosome formations.
- Mutant systems: less data, prior distribution more important.
This is a BBSRC funded studentship.
Acknowledgments

- This is a BBSRC funded studentship.
- ML group at Uni. Sheffield.
Acknowledgments

This is a BBSRC funded studentship.
ML group at Uni. Sheffield.
Neil Lawrence, Alastair Goldman and Guido Sanguinetti.
Acknowledgments

- This is a BBSRC funded studentship.
- ML group at Uni. Sheffield.
- Neil Lawrence, Alastair Goldman and Guido Sanguinetti.
- PIMMS organisers.
Acknowledgments

- This is a BBSRC funded studentship.
- ML group at Uni. Sheffield.
- Neil Lawrence, Alastair Goldman and Guido Sanguinetti.
- PIMMS organisers.
- Thank you for listening ... any questions?