Ellen Baake

Bielefeld University

joint work with Frederic Alberti

1. The selection-recombination equation (forward in time)
2. Its solution via genealogical thinking
Recombination

Ancestral lines under selection and recombination
Sequences and selection

individual: sequence of \( n \) sites, \( S = \{1, \ldots, n\} \)

types: \( x := (x_1, \ldots, x_n) \in \prod_{i \in S} X_i =: X, \ X_i = \{0, 1\} \)

marginal types: \( x_U := (x_i)_{i \in U}, \ U \subseteq S \)

reproduction and selection:

- single selected site: \( i_\bullet \in S \) (all others ‘neutral’)
- types \( x \) with \( x_{i_\bullet} = 1 \): reproduce at rate \( 1 \) (‘bad, unfit’)
- types \( x \) with \( x_{i_\bullet} = 0 \): reproduce at rate \( 1 + s \) (‘good, fit’)
- offspring replaces randomly chosen individual

Ellen Baake
Ancestral lines under selection and recombination
Ordering the sites

partial order on $S$:

- $i \preceq j$ if either $i \leq i \leq j$ or $i \geq i \geq j$
- $i \prec j$ if $i \preceq j$ and $i \neq j$
- $D_i := \{j \in S \mid i \preceq j\}$ (i-tail), $C_i := S \setminus D_i$ (i-head), $i \in S \setminus i\bullet$

\[
\begin{array}{cccccccc}
\begin{array}{c}
C_8 \\
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10
\end{array}
\end{array}
\]

\[
\begin{array}{cccccccc}
\begin{array}{c}
D_8
\end{array}
\end{array}
\]

\[
\begin{array}{cccccccc}
\begin{array}{c}
D_2
\end{array}
\end{array}
\]

\[
\begin{array}{cccccccc}
\begin{array}{c}
C_2
\end{array}
\end{array}
\]

- $\rightsquigarrow i\bullet \in C_i$

- (single-crossover) recombination between $C_i$ and $D_i$ (partitions $S$ into $C_i$ and $D_i$) at rate $\varrho_i$
Sequences and recombination

\[ x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n \]

\( Q_i \)

\[ x_1, \ldots, x_i, y_{i+1}, \ldots, y_n \]
Selection-recombination equation

- \( \omega_t = (\omega_t(x))_{x \in X} \) probability measure on \( X \)
  \( (\omega_t(x) \) proportion of individuals of type \( x \) at time \( t \), \( x \in X \) \)
- solves selection-recombination equation (SRE)

\[
\dot{\omega}_t(x) = s \left[ \delta_{x \cdot, 0} - \omega_t(\ast, 0, \ast) \right] \omega_t(x) \\
+ \sum_{i \in S \setminus i \cdot} \rho_i \left[ \omega_t(x_{C_i}, \ast) \omega_t(\ast, x_{D_i}) - \omega_t(x) \right], \quad x \in X
\]
Recombination

\[ x_1, \ldots, x_n \]

\[ y_1, \ldots, y_n \]

\[ Q_i \]

\[ x_1, \ldots, x_i, y_{i+1}, \ldots, y_n \]
Recombination

Ellen Baake
Ancestral lines under selection and recombination
Recombination

Ellen Baake

Ancestral lines under selection and recombination
Selection-recombination equation

- \( \omega_t = (\omega_t(x))_{x \in X} \) probability measure on \( X \)
  (\( \omega_t(x) \) proportion of individuals of type \( x \) at time \( t \), \( x \in X \))
- solves selection-recombination equation (SRE)

\[
\dot{\omega}_t(x) = s \left[ \delta_{x,0} - \omega_t(\ast,0,\ast) \right] \omega_t(x) \\
+ \sum_{i \in S \setminus i_0} \rho_i \left[ \omega_t(x_{C_i}, \ast) \omega_t(\ast, x_{D_i}) - \omega_t(x) \right], \quad x \in X
\]
Research context

- **hitchhiking**: neutral $x_i \ (i \neq i_*)$ selectively favoured if linked with $x_{i_*} = 0$
- **Akin 1979**: “The differential equations which model the action of selection and recombination are nonlinear equations which are impossible to solve explicitly.”
hitchhiking: neutral \( x_i \ (i \neq i^*) \) selectively favoured if linked with \( x_{i^*} = 0 \)

Akin 1979: the “selection-recombination equation is a large, nonlinear system of differential equations that are impossible to solve explicitly”

Stephan, Song, Langley 2006: approximate solution (\( n = 3 \), wild approximations, technical result, numerical evaluation only, not generalisable)
Recombinators

- canonical projection: for $\emptyset \neq U \subseteq S$,
  \[
  \pi_U : X \to \bigtimes_{i \in U} X_i = X_U, \quad \pi_U(x) = (x_i)_{i \in U} = x_U
  \]

- marginal measure wrt sites in $U$: for $\nu \in \mathcal{P}(X)$,
  \[
  \pi_U.\nu = \nu \circ \pi_U^{-1} =: \nu_U
  \]

  type distribution of sites in $U$
  for $x_U \in X_U$:
  \[
  \nu_U(x_U) = \nu(x_U, \ast)
  \]

- recombinator: for $i \in S \setminus i_*$,
  \[
  \mathcal{P}(X) \to \mathcal{P}(X)
  \]
  \[
  R_i(\nu) := \nu_{C_i} \otimes \nu_{D_i}
  \]
  \[
  (R_i(\nu))(x) = \nu(x_{C_i}, \ast) \cdot \nu(\ast, x_{D_i})
  \]

  distribution of sequences pieced together from random $i$-heads and $i$-tails
Selection-recombination equation

\[ \dot{\omega}_t(x) = s \left[ \delta_{x_{i\bullet},0} - \omega_t(\ast, 0, \ast) \right] \omega_t(x) \]

\[ + \sum_{i \in S \setminus i_\bullet} \rho_i \left[ \omega_t(x_{C_i}, \ast) \omega_t(\ast, x_{D_i}) - \omega_t(x) \right], \quad x \in X \]

or

\[ \dot{\omega}_t = s(F - f(\omega_t))\omega_t + \sum_{i \in S \setminus i_\bullet} \rho_i (R_i - \mathbb{1})\omega_t = \Psi_{\text{sel}}(\omega_t) + \Psi_{\text{rec}}(\omega_t) \]

with

\[ F\omega(x) = \delta_{x_{i\bullet},0} \omega(x) \]

and

\[ f(\omega) := \omega_{\{i\bullet\}}(0) \]

(proportion of fit individuals = proportion of individuals with 0 at selected site)
Recursive solution

- $(i_k)_{0 \leq k < n}$ non-decreasing permutation of $S$ (sensu $\preceq$)
- $i_0 = i_*$
- $C^{(k)} = C_{i_k}, D^{(k)} = D_{i_k}, q^{(k)} := q_{i_k}$ ($k > 0$)

1 2 3 4 5 6 7 8 9 10

\[ i_9 \quad i_5 \quad i_1 \quad i_0 \quad i_2 \quad i_3 \quad i_4 \quad i_6 \quad i_7 \quad i_8 \]

hierarchy of solutions:

- $(\omega^{(0)}_t)_{t \geq 0}$ solution of SRE with $q^{(\ell)} = 0$ for all $\ell$
  (pure selection equation)
- $(\omega^{(k)}_t)_{t \geq 0}$ solution of SRE truncated at $k$
  (with $q^{(\ell)} = 0$ for all $\ell > k$), \( 1 \leq k < n \)
Recursive solution

Theorem

The solution of the pure selection equation is given by

\[ \omega_t^{(0)} = \frac{e^{stF} \omega_0}{e^{st f(\omega_0)} + 1 - f(\omega_0)} =: \varphi_t(\omega_0), \]

and the solutions \( \omega_t^{(k)} \) can be computed via the recursion

\[ \omega_t^{(k)} = e^{-\varrho^{(k)}_t} \omega_t^{(k-1)} + \omega_{C(k),t}^{(k-1)} \otimes \int_0^t \varrho^{(k)} e^{-\varrho^{(k)}_\tau} \omega_{D(k),\tau}^{(k-1)} d\tau. \]

analytical proof ✓

(genealogical) meaning?
$i$-splitting events on every line at rate $q_i$, $i \in S \setminus i_\bullet$ (backward!)

$S = \{1, 2, 3\}$, $i_\bullet = 1$:
$i$-splitting events on every line at rate $\varrho_i$, $i \in S \setminus i_\bullet$ (backward!)

$S = \{1, 2, 3\}$, $i_\bullet = 1$

$s = 0$: partitioning of sequence across ancestors
sampling from $\omega_0$
Ancestral graph, $s > 0$

$$x \sim \varphi_t(\omega_0) \quad \text{and} \quad x \sim \omega_t^{(0)}$$

$$\omega_t^{(0)} = \varphi_t(\omega_0)$$
Ancestral graph, $s > 0$

\[
\begin{align*}
    x &\sim \varphi_t(\omega_0) \\
    y &\sim \varphi_\tau(\omega_0) \\
    (x_1, y_2, y_3) &\sim \omega_t^{(1)}
\end{align*}
\]

decomposition according to leftmost 2-event on ancestral line of tail (since $\Psi_{\text{sel}}(\nu_{\{i,\bullet\}} \otimes \nu_{\{S \setminus i,\bullet\}}) = \Psi_{\text{sel}}(\nu_{\{i,\bullet\}}) \otimes \nu_{\{S \setminus i,\bullet\}}$)

\[
\begin{align*}
    \sim \omega_t^{(1)} &= e^{-\varrho_2 t} \varphi_t(\omega_0) + \varphi_{\{1\},t}(\omega_0) \otimes \int_0^t \varrho_2 e^{-\varrho_2 \tau} \varphi_{\{2,3\},\tau}(\omega_0) \, d\tau \\
    &= e^{-\varrho^{(1)} t} \omega_t^{(0)} + \omega_{C(1),t}^{(0)} \otimes \int_0^t \varrho^{(1)} e^{-\varrho^{(1)} \tau} \omega_{D(1),\tau}^{(0)} \, d\tau
\end{align*}
\]
Ancestral graph, $s > 0$

\[ x \sim \varphi_t(\omega_0) \]

\[ y \sim \varphi_\tau(\omega_0) \]

\[ z \sim \varphi_\sigma(\omega_0) \]

\[ (x_1, y_2, z_3) \sim \omega_t^{(2)} \]

decomposition according to \textbf{leftmost} $k$-event on ancestral line of $D^{(k)}$

\[ \omega_t^{(k)} \rightarrow \omega_t = e^{-g^{(k)}t} \omega_t^{(k-1)} + \omega^{(k-1)}_{C(k-1), t} \otimes \int_0^t g^{(k)} e^{-g^{(k)}\sigma} \omega^{(k-1)}_{D(k-1), \sigma} d\sigma \]