Solving the recombination equation

Ellen Baake

Bielefeld University

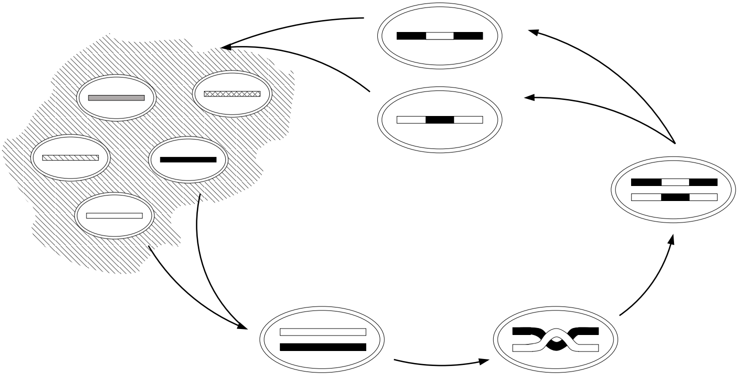
joint work with Michael Baake

1. Recombination
2. Deterministic dynamics forward in time
3. Stochastic partitioning process backward in time

PROBABILISTIC STRUCTURES
IN EVOLUTION
DFG SPP 1590

Universität Bielefeld

Recombination



major source of genetic variability in populations

dynamics of genetic composition of germ cell pool (deterministic limit) ??

Sequences, types, populations

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

letter at site i: $x_i \in X_i$ (finite), $1 \leqslant i \leqslant n$

types: $x := (x_1, \ldots, x_n) \in X_1 \times \cdots \times X_n =: X$

marginal types: $x_I := (x_i)_{i \in I}$, $\varnothing \neq I \subseteq S$

population: $p = \big(p(x)\big)_{x \in X}$ probability measure on X

$p(x) \geqslant 0$ proportion of individuals of type $x \in X$

$$\sum_{x \in X} p(x) = 1$$

Recombining sequences

crossovers between sites i and $i+1$, and between $j > i$ and $j+1$

$$x_1, x_2, \dots, x_n \qquad\qquad y_1, y_2, \dots, y_n$$

$$\downarrow$$

$$x_1, \dots, x_i, y_{i+1}, \dots, y_j, x_{j+1}, \dots, x_n$$

offspring copies from (randomly chosen) ordered pair of parents and replaces a randomly chosen individual in the population

Recombining sequences

crossovers between sites i and $i+1$, and between $j > i$ and $j+1$

$x_1, \dots, x_i, *, \dots, *, x_{j+1}, \dots, x_n$ $\qquad$ $*, \dots, *, x_{i+1}, \dots, x_j, *, \dots, *$

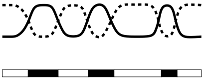
$\downarrow$

$x_1, x_2, \dots, x_n$

offspring copies from (randomly chosen) ordered pair of parents
and replaces a randomly chosen individual in the population
'$*$ at site i' $= X_i$

Partitions

- reco event defines a partition $\mathcal{A}$ of S into at most two parts
  ex.: $\mathcal{A} = \{\{1, \dots, i, j+1, \dots, n\}, \{i+1, \dots, j\}\}$
- $\mathcal{A} = \{S\} = \mathbf{1}$: offspring copies first parent
- $\mathcal{A} = \{A_1, A_2\}$, $A_1, A_2 \neq \varnothing, A_1 \dot\cup A_2 = S$:



- $\varrho(\mathcal{A})$ rate of recombination according to $\mathcal{A}$, $\mathcal{A} \in \mathcal{P}_{\leqslant 2}(S)$

$\rightsquigarrow$ recombination equation:

$$\dot{p}_t(x) = \sum_{\mathcal{A} \in \mathcal{P}_2(S)} \varrho(\mathcal{A}) \big[p_t(x_{A_1}, *)\, p_t(*, x_{A_2}) - p_t(x) \big], \quad x \in X.$$

formidable!

Partitions

- $\mathcal{P}(S)$: set of all partitions of S
- $\preccurlyeq$ partial order (refinement)

Example: $n = 3$

$$\begin{array}{ccccc} & & \{\{1\,2\,3\}\} & & \\ & \prec & \curlyvee & \succ & \\ \{\{1\,2\}\{3\}\} & & \{\{1\,3\}\{2\}\} & & \{\{1\}\{2\,3\}\} \\ & \succ & \curlyvee & \prec & \\ & & \{\{1\}\{2\}\{3\}\} & & \end{array}$$

History

- first study of recombination dynamics:
  Jennings 1917, Robbins 1918
- iterative procedure to determine solution:
  Geiringer 1944, Bennett 1954
- genetic algebras: ....., Lyubich 1992
- Haldane linearisation: McHale and Ringwood, 1983 (transforms nonlinear dynamics into linear one in higher-dimensional space by adding multilinear transforms of p_t)

Aim: Do it right....

Recombinators

- canonical projection: for $\varnothing \neq I \subseteq S$,

$$\pi_I : X \to \times_{i \in I} X_i = X_I, \quad \pi_I(x) = (x_i)_{i \in I} = x_I$$

- marginal measure wrt sites in I: for $\nu \in \boldsymbol{P}(X)$,

$$\pi_I . \nu = \nu \circ \pi_I^{-1} =: \nu^I$$

  type distribution of sites in I
  for $x_I \in X_I$: $\quad \nu^I(x_I) = \nu(x_I, *)$

- recombinator: for $\mathcal{A} = \{A_1, \ldots, A_m\} \in \mathcal{P}(S)$,

$$\boldsymbol{P}(X) \longrightarrow \boldsymbol{P}(X)$$

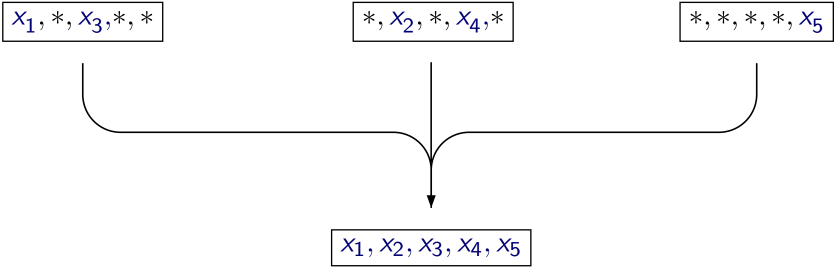
$$R_{\mathcal{A}}(\nu) := \nu^{A_1} \otimes \ldots \otimes \nu^{A_n}$$

$$\big(R_{\mathcal{A}}(\nu)\big)(x) = \nu(x_{A_1}, *) \cdot \ldots \cdot \nu(*, x_{A_n})$$

  distribution of sequences randomly pieced together according to $\mathcal{A}$

Recombinators

$R_{\mathcal{A}}(p) := p^{A_1} \otimes \ldots \otimes p^{A_n}$ for $\mathcal{A} = \{\{1,3\},\{2,4\},\{5\}\}$:



(generalised) recombination equation:

$$\dot{p}_t = \sum_{\mathcal{A} \in \mathcal{P}_{\geqslant 2}(S)} \varrho(\mathcal{A}) \big(R_{\mathcal{A}} - \mathbb{1}\big)(p_t),$$

existence and uniqueness of solution, and forward invariance of $\boldsymbol{P}(X)$, via standard methods

Factorisation property of recombinators

for $\varnothing \neq U \subseteq S,\ \mathcal{A} \in \mathcal{P}(U),\ \nu^U \in \boldsymbol{P}(X_U)$:

marginal recombinator: $R^U_{\mathcal{A}}(\nu^U)$ defined as before, with S replaced by U

Lemma

$S = U \,\dot\cup\, V$, $\mathcal{A} \in \mathcal{P}(U)$, $\mathcal{B} \in \mathcal{P}(V)$, $\nu \in \boldsymbol{P}(X) \rightsquigarrow$

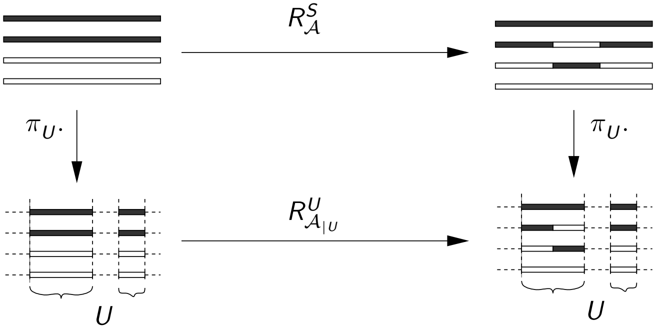
$$R_{\mathcal{A} \cup \mathcal{B}}(\nu) \;=\; R^U_{\mathcal{A}}(\nu^U) \otimes R^V_{\mathcal{B}}(\nu^V).$$

Marginalisation property of recombinators

Lemma

$\mathcal{A} \in \mathcal{P}(S)$, $\varnothing \neq U \subseteq S$, $\nu \in \boldsymbol{P}(X) \rightsquigarrow$

$$\pi_U.(R^S_{\mathcal{A}}(\nu)) \;=\; R^U_{\mathcal{A}|_U}(\pi_U.\nu).$$



$\mathcal{A}|_U$ restriction of $\mathcal{A}$ to U

Marginalisation consistency of recombination equation

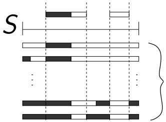
Proposition

$\varnothing \neq U \subseteq S$, p_t solution of recombination equation with $p_0 \in \boldsymbol{P}(X) \rightsquigarrow (p_t^U)_{t \geqslant 0}$ on X_U solves

$$\frac{\mathrm{d}}{\mathrm{d}t} p_t^U = \sum_{\mathcal{A} \in \mathcal{P}_{\geqslant 2}(U)} \varrho^U(\mathcal{A}) \left(R_{\mathcal{A}}^U - \mathbb{1}\right)(p_t^U)$$

with $p_0^U = \pi_U . p_0$ and marginalised rates

$$\varrho^U(\mathcal{A}) := \sum_{\substack{\mathcal{B} \in \mathcal{P}(S) \\ \mathcal{B}|_U = \mathcal{A}}} \varrho^S(\mathcal{B})$$



$\mathcal{A} \in \mathcal{P}(U)$

$\{\mathcal{B} \in \mathcal{P}(S) \colon \mathcal{B}|_U = \mathcal{A}\}$

Solving the recombination equation

$\mathcal{B} = \{B_1, \dots, B_m\}$, p_t solution of recombination equation $\rightsquigarrow$

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} R_{\mathcal{B}}(p_t) &= \sum_{i=1}^{m} \Big(\frac{\mathrm{d}}{\mathrm{d}t} p_t^{B_i} \Big) \otimes \bigotimes_{j:j \neq i} p_t^{B_j} \\
&= \sum_{i=1}^{m} \sum_{\mathfrak{b}_i \in \mathcal{P}_{\geqslant 2}(B_i)} \varrho^{B_i}(\mathfrak{b}_i) \, \big(R_{\mathfrak{b}_i}^{B_i} - \mathbb{1}\big)(p_t^{B_i}) \otimes \bigotimes_{j:j \neq i} p_t^{B_j} \\
&= \sum_{i=1}^{m} \sum_{\mathfrak{b}_i \in \mathcal{P}_{\geqslant 2}(B_i)} \varrho^{B_i}(\mathfrak{b}_i) \big(R_{(\mathcal{B} \setminus B_i) \cup \mathfrak{b}_i} - R_{\mathcal{B}}\big)(p_t) \\
&= \sum_{\mathcal{C} \prec \mathcal{B}} Q_{\mathcal{B}\mathcal{C}} \big(R_{\mathcal{C}} - R_{\mathcal{B}}\big)(p_t) = \sum_{\mathcal{C} \preccurlyeq \mathcal{B}} Q_{\mathcal{B}\mathcal{C}} R_{\mathcal{C}}(p_t) \\
&= \sum_{\mathcal{C} \in \mathcal{P}(S)} Q_{\mathcal{B}\mathcal{C}} R_{\mathcal{C}}(p_t)
\end{aligned}
$$

Solving the recombination equation

Theorem

p_t solution of recombination equation with $p_0 \in \boldsymbol{P}(X)$, and $\mathcal{B} \in \mathcal{P}(S) \rightsquigarrow$

$$\frac{\mathrm{d}}{\mathrm{d}t} R_{\mathcal{B}}(p_t) = \sum_{\mathcal{C} \in \mathcal{P}(S)} Q_{\mathcal{B}\mathcal{C}} R_{\mathcal{C}}(p_t),$$

with Markov generator $Q = (Q_{\mathcal{B}\mathcal{C}})_{\mathcal{C} \in \mathcal{P}(S)}$

$$Q_{\mathcal{B}\mathcal{C}} = \begin{cases} \varrho^{B_i}(\mathfrak{b}_i), & \text{if } \mathcal{C} = (\mathcal{B} \setminus B_i) \cup \mathfrak{b}_i \text{ for some} \\ & \mathfrak{b}_i \in \mathcal{P}_{\geqslant 2}(B_i) \text{ and exactly one } i, \\ -\sum_{i=1}^{|\mathcal{B}|} \sum_{\mathfrak{b}_i \in \mathcal{P}_{\geqslant 2}(B_i)} \varrho^{B_i}(\mathfrak{b}_i), & \text{if } \mathcal{C} = \mathcal{B}, \\ 0, & \text{otherwise.} \end{cases}$$

Solution of the recombination equation

column vector $\varphi_t := \big(\varphi_t(\mathcal{B})\big)_{\mathcal{B}\in\mathcal{P}(S)}$ with $\varphi_t(\mathcal{B}) := R_{\mathcal{B}}(p_t)$

$\rightsquigarrow$ linear system: $\dot{\varphi}_t = Q\varphi_t$

solution: $\varphi_t = \mathrm{e}^{tQ}\varphi_0$

first component: $p_t = \varphi_t(\mathbf{1}) = \sum_{\mathcal{A}\in\mathcal{P}(S)} \big(\mathrm{e}^{tQ}\big)_{\mathbf{1}\mathcal{A}} R_{\mathcal{A}}(p_0)$

Theorem

The recombination equation has the solution

$$p_t = \sum_{\mathcal{A}\in\mathcal{P}(S)} a_t(\mathcal{A})\, R_{\mathcal{A}}(p_0)$$

with $a_t(\mathcal{A}) := \big(\mathrm{e}^{tQ}\big)_{\mathbf{1}\mathcal{A}}$.

Interpretation??

Partitioning process

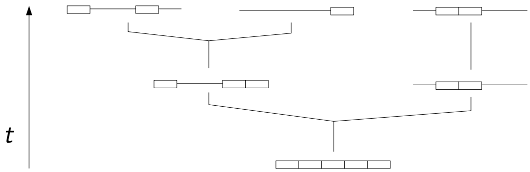
$$Q_{\mathcal{B}\mathcal{C}} = \begin{cases} \varrho^{B_i}(\mathfrak{b}_i), & \text{if } \mathcal{C} = (\mathcal{B} \setminus B_i) \cup \mathfrak{b}_i \text{ for some} \\ & \mathfrak{b}_i \in \mathcal{P}_{\geqslant 2}(B_i) \text{ and exactly one } i, \\ -\sum_{i=1}^{|\mathcal{B}|} \sum_{\mathfrak{b}_i \in \mathcal{P}_{\geqslant 2}(B_i)} \varrho^{B_i}(\mathfrak{b}_i), & \text{if } \mathcal{C} = \mathcal{B}, \\ 0, & \text{otherwise.} \end{cases}$$

Q generates partitioning process $\{\Sigma_t\}_{t \geqslant 0}$:

- Markov chain in continuous time with state space $\mathcal{P}(S)$
- progressive refinement: if $\Sigma_t = \mathcal{B}$, replace part B_i of $\mathcal{B}$ by $\mathfrak{b}_i \in \mathcal{P}_{\geqslant 2}(B_i)$ at rate $\varrho^{B_i}(\mathfrak{b}_i)$;
  for $1 \leqslant i \leqslant |\mathcal{B}|$, independently of all other parts

Partitioning process

Q generates $\{\Sigma_t\}_{t\geqslant 0}$:



$\Sigma_t = \{\{1,4\},\{2,3\},\{5\}\}$

$\Sigma_0 = \{\{1,2,3,4,5\}\}$

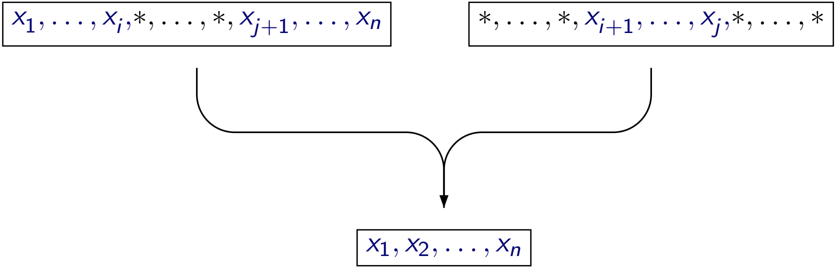
semigroup:

$$\left(\mathrm{e}^{tQ}\right)_{\mathcal{B}\mathcal{C}} = \mathbb{P}\left(\Sigma_t = \mathcal{C} \mid \Sigma_0 = \mathcal{B}\right)$$

in particular:

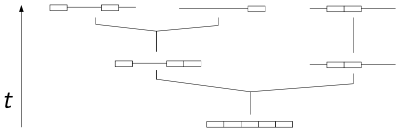
$$a_t(\mathcal{A}) = \left(\mathrm{e}^{tQ}\right)_{\mathbf{1}\mathcal{A}} = \mathbb{P}\left(\Sigma_t = \mathcal{A} \mid \Sigma_0 = \mathbf{1}\right)$$

Partitioning and recombination



recombination forward in time = splitting up backward in time

Partitioning and recombination



$\Sigma_t = \{\{1,4\},\{2,3\},\{5\}\}$

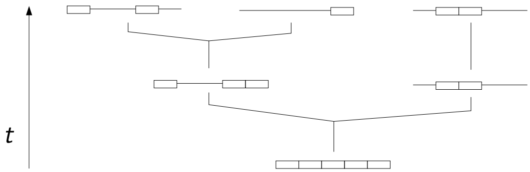
$\Sigma_0 = \{\{1,2,3,4,5\}\}$

$\{\Sigma_t\}_{t \geqslant 0}$ describes partitioning of genetic material of an individual backward in time

- present individual: $\Sigma_0 = \{S\} = \mathbf{1}$
- at time t before the present: $\Sigma_t = \mathcal{B} = \{B_1, \ldots B_m\}$
  each B_i corresponds to parent that contributed sites in B_i
- B_i-individual splits up into $\mathfrak{b}_i$ at rate $\varrho^{B_i}(\mathfrak{b}_i)$, $\mathfrak{b}_i \in \mathcal{P}_{\geqslant 2}(B_i)$, independently for all i
- $\rightsquigarrow$ transition from $\mathcal{B}$ to $\mathcal{C} \prec \mathcal{B}$ at rate $Q_{\mathcal{BC}}$!

Duality

Construction of type in present population:



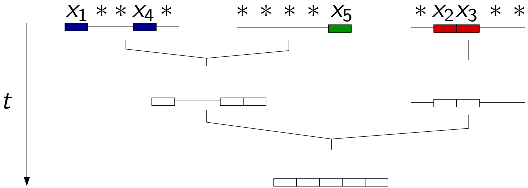
$$\Sigma_t = \{\{1,4\},\{2,3\},\{5\}\}$$

$$\Sigma_0 = \{\{1,2,3,4,5\}\}$$

1. run $\{\Sigma_t\}_{t \geqslant 0}$ (untyped, backward)

Duality

Construction of type in present population:



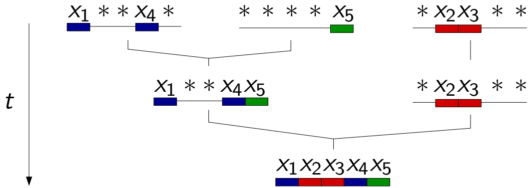
$\Sigma_t = \{\{1,4\},\{2,3\},\{5\}\}$

$\Sigma_0 = \{\{1,2,3,4,5\}\}$

1. run $\{\Sigma_t\}_{t \geqslant 0}$ (untyped, backward)
2. assign colours (parents) and letters (types)
   If $\Sigma_t = \mathcal{A} = \{A_1, \ldots, A_m\}$: draw letters at sites in A_i from $p_0^{A_i}$, independently for $1 \leqslant i \leqslant m \rightsquigarrow$ type distribution $R_{\mathcal{A}}(p_0)$

Duality

Construction of type in present population:



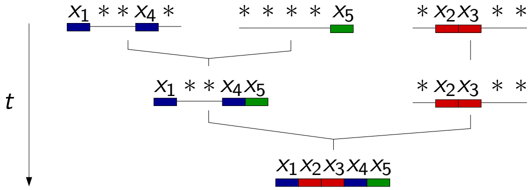
$\Sigma_t = \{\{1,4\},\{2,3\},\{5\}\}$

$\Sigma_0 = \{\{1,2,3,4,5\}\}$

1. run $\{\Sigma_t\}_{t \geqslant 0}$ (untyped, backward)
2. assign colours (parents) and letters (types)
   If $\Sigma_t = \mathcal{A} = \{A_1, \ldots, A_m\}$: draw letters at sites in A_i from $p_0^{A_i}$, independently for $1 \leqslant i \leqslant m \rightsquigarrow$ type distribution $R_{\mathcal{A}}(p_0)$
3. propagate colours and letters forward in time
   $\rightsquigarrow$ type distribution $R_{\mathcal{A}}(p_0)$

Duality

Construction of type in present population:



$$\Sigma_t = \{\{1,4\},\{2,3\},\{5\}\}$$

$$\Sigma_0 = \{\{1,2,3,4,5\}\}$$

$$p_t = \sum_{\mathcal{A} \in \mathcal{P}(S)} \underbrace{\mathbb{P}\big(\Sigma_t = \mathcal{A} \mid \Sigma_0 = \mathbf{1}\big)}_{a_t(\mathcal{A})} R_{\mathcal{A}}(p_0) = \mathbb{E}\big(R_{\Pi}(p_0)\big)$$

stochastic representation of deterministic solution

Applications and connections

- recursive evaluation of e^{tQ} (Q triangular !) (with M. Salamat, E. Shamsara, 2016)
- closed formula for single crossovers (E.&M.B. 2003)
- analogous solution for discrete-time system

$$p_{t+1} = \sum_{\mathcal{A} \in \mathcal{P}(S)} r(\mathcal{A}) \, R_{\mathcal{A}}(p_t) \rightsquigarrow p_t = \sum_{\mathcal{A} \in \mathcal{P}(S)} a_t(\mathcal{A}) \, R_{\mathcal{A}}(p_0)$$

  (E.&M.B. 2016, S. Martínez 2016a, 2016b)
- tree representation for solution for single crossovers in discrete time (with U. von Wangenheim, 2014, and M. Esser, work in progress)
- a different tree representation for general partitions in discrete time, and quasistationary distribution (S. Martínez 2016a, 2016b)

Job ad

1 postdoc and 1 PhD position available at Bielefeld early next year

- to work on recombination and/or mutation & selection
- with Fernando Cordero and/or Ellen & Michael Baake

interested? contact us !