Genetic recombination, reaction systems, partitioning, and the solution of the differential equation

Ellen Baake, Bielefeld University

joint work with Michael Baake

1. Recombination
2. Deterministic dynamics forward in time: reaction system
3. Solution via stochastic partitioning process backward in time

Universität Bielefeld

SFB 1283

Sequences, types, populations

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

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

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

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

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

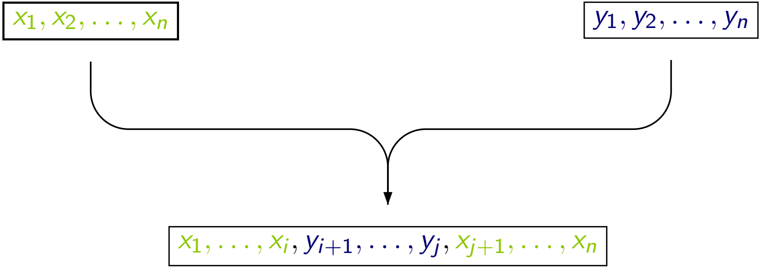
$p(x) \geqslant 0$ proportion of individuals of type $x \in X$

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

Recombining sequences

on the occasion of sexual reproduction:

offspring pieced together from (randomly chosen) pair (x, y) of parents

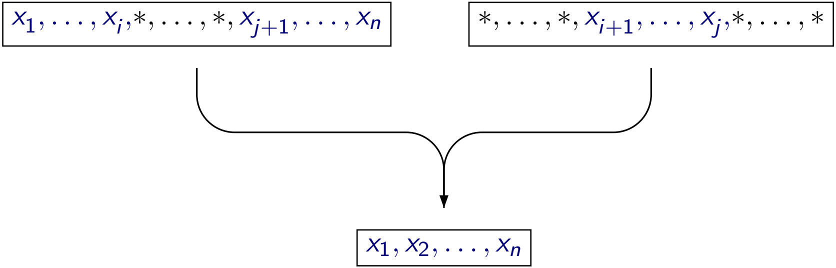


replaces a randomly chosen individual in the population

Recombining sequences

on the occasion of sexual reproduction:

offspring pieced together from (randomly chosen) pair (x, y) of parents



replaces a randomly chosen individual in the population
'$*$ at site i' $= X_i$

Recombination equation

- reco event defines a partition $\mathcal{A}$ of S into at most two parts
  ex.: $\mathcal{A} = \big\{\{1, \dots, i, j+1, \dots, n\}, \{i+1, \dots, j\}\big\}$
- $\mathcal{A} = \{S\} =: \mathbf{1}$: offspring copies first parent
- $\mathcal{A} = \{A_1, A_2\}$, $A_1, A_2 \neq \varnothing, A_1 \dot\cup A_2 = S$: two parents
- dynamics:
  $\varrho_{\mathcal{A}}$ rate of recombination according to $\mathcal{A}$, $\mathcal{A} \in \mathcal{P}_{\leqslant 2}(S)$

$\rightsquigarrow$ recombination equation:

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

equivalent to a reaction system (Müller & Hofbauer 2016)

large, nonlinear ODE system

Recombinators

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

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

- 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 $\mathcal{A} = \{A_1, \dots, A_m\} \in \mathcal{P}(S)$,

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

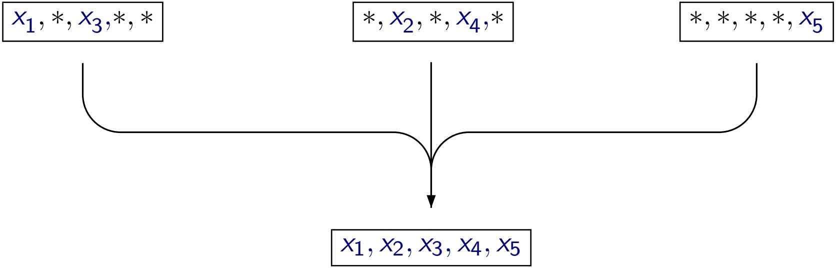
$$R_{\mathcal{A}}(\nu) := \nu^{A_1} \otimes \dots \otimes \nu^{A_m}$$

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

  distribution of sequences sampled from ν and randomly pieced together according to $\mathcal{A}$

Recombinators

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

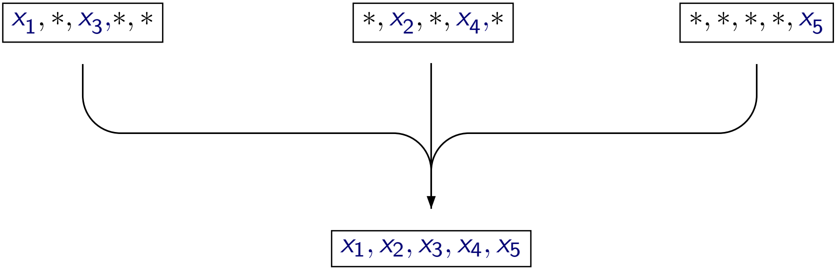


(generalised) recombination equation:

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

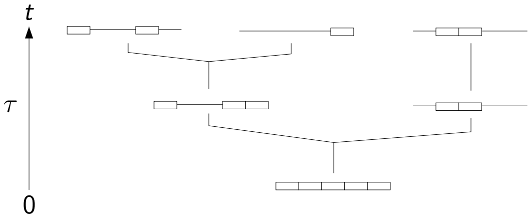
still equivalent to a reaction system (Alberti 2021)

Recombination and partitioning



recombination forward in time = splitting up backward in time

Partitioning process



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

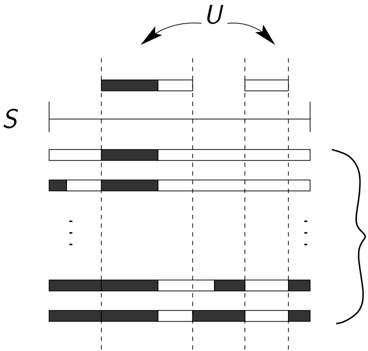
$\Sigma_0 = \{\{1,2,3,4,5\}\} = \mathbf{1}$

- $(\Sigma_\tau)_{\tau \geqslant 0}$ partitioning of genetic material of an individual backward in time
- ancestral recombination graph in law-of-large-numbers regime

Partitioning process

marginal recombination rates:

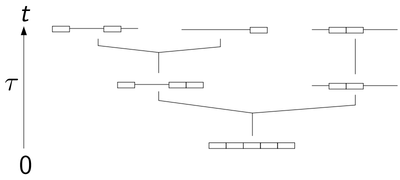
$$\varrho_{\mathcal{B}}^{U} := \sum_{\substack{\mathcal{A}\in\mathcal{P}(S) \\ \mathcal{A}|_U=\mathcal{B}}} \varrho_{\mathcal{A}}, \qquad \varnothing \neq U \subseteq S,\ \mathcal{B} \in \mathcal{P}(U)$$



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

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

Partitioning process



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

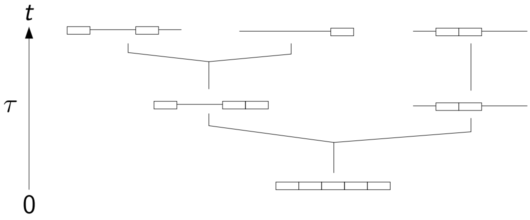
$\Sigma_0 = \{\{1,2,3,4,5\}\}$

$(\Sigma_\tau)_{\tau \geqslant 0}$ Markov chain in continuous time with state space $\mathcal{P}(S)$

- present individual: $\Sigma_0 = \{S\} = \mathbf{1}$
- at time τ before the present: $\Sigma_\tau = \mathcal{B} = \{B_1, \dots B_m\}$
  each B_i corresponds to ancestor 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 $(\mathcal{B} \setminus B_i) \cup \mathfrak{b}_i \prec \mathcal{B}$
- $\rightsquigarrow$ Markov generator $Q = (Q_{\mathcal{B}\mathcal{C}})_{\mathcal{B},\mathcal{C} \in \mathcal{P}(S)}$

Connecting the present and the past

Construction of type in present population (at time t):



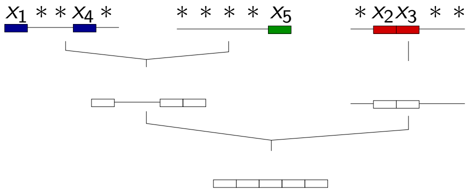
$\Sigma_t = \{\{1,4\},\{2,3\},\{5\}\}$

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

1. run $(\Sigma_\tau)_{\tau \geqslant 0}$ (untyped, backward)

Connecting the present and the past

Construction of type in present population (at time t):



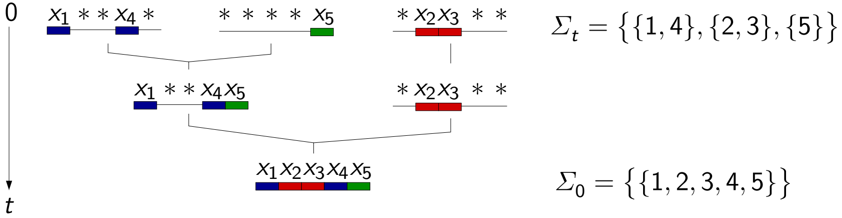
$\Sigma_t = \{\{1,4\},\{2,3\},\{5\}\}$

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

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

Connecting the present and the past

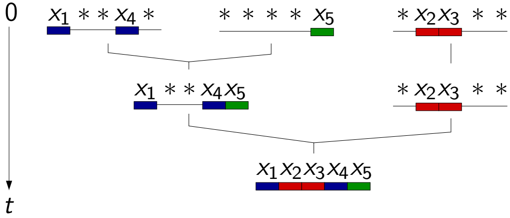
Construction of type in present population (at time t):



1. run $(\Sigma_\tau)_{\tau \geqslant 0}$ (untyped, backward)
2. assign colours (parents) and letters (types)
   If $\Sigma_t = \mathcal{A} = \{A_1, \dots, A_m\}$: 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 $R_{\mathcal{A}}(p_0)$

Connecting the present and the past

Construction of type in present population (at time t):



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

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

Theorem

$$p_t = \sum_{\mathcal{A} \in \mathcal{P}(S)} \mathbb{P}\left(\Sigma_t = \mathcal{A} \mid \Sigma_0 = \mathbf{1}\right) R_{\mathcal{A}}(p_0)$$

- stochastic representation of a deterministic solution
- nonlinear $\longrightarrow$ linear !

Solution of the recombination equation

semigroup:

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

in particular:

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

$\rightsquigarrow$ solution:

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

$\rightsquigarrow$ $a_t(\mathcal{A})$ for given Q ?

Example: two parents, single breakpoint

$\varrho_{\mathcal{A}} > 0 \Rightarrow \mathcal{A} = \mathcal{A}_k = \big\{\{1, 2, \ldots, k\}, \{k+1, \ldots, n\}\big\}$
for some $1 \leq k < n$

$\rightsquigarrow \varSigma_\tau$ interval partition for all τ
(contiguous blocks, e.g. $\mathcal{C} = \big\{\{1, 2\}, \{3, 4, 5\}, \{6, 7\}\big\}$)

Theorem

$a_t(\mathcal{C}) = 0$ if $\mathcal{C}$ is not an interval partition; otherwise,

$$a_t(\mathcal{C}) = \prod_{k \in G(\mathcal{C})} \big(1 - \exp(-t\varrho_{\mathcal{A}_k})\big) \prod_{\ell \in S \setminus (G(\mathcal{C}) \cup \{n\})} \exp(-t\varrho_{\mathcal{A}_\ell}),$$

where $G(\mathcal{C})$ is the set of breakpoints defining $\mathcal{C}$.

(e.g., $G\Big(\big\{\{1, 2\}, \{3, 4, 5\}, \{6, 7\}\big\}\Big) = \{2, 5\}$)

General partitions

$a_t(\mathcal{A})$

- via recursions (Q triangular!); done
- explicitly via combinatorial tools:
  inclusion-exclusion, Möbius inversion; nearly done

  $a_t(\mathcal{A})$ in terms of

$$r_{\mathbf{1}}^{U} := \exp\Big(- \sum_{\mathbf{1} \neq \mathcal{B} \in \mathcal{P}(U)} t \varrho_{\mathcal{B}}^{U} \Big), \quad U \subseteq S$$

ex.: $S = \{1,2,3\} \rightsquigarrow$

$$a_t(\{S\}) = r_{\mathbf{1}}^{S},$$

$$a_t\big(\{\{i,j\},\{k\}\}\big) = r_{\mathbf{1}}^{\{i,j\}} - r_{\mathbf{1}}^{\{i,j,k\}},$$

$$a_t\big(\{\{1\},\{2\},\{3\}\}\big) = 1 - r_{\mathbf{1}}^{\{1,2\}} - r_{\mathbf{1}}^{\{1,3\}} - r_{\mathbf{1}}^{\{2,3\}} + 2 r_{\mathbf{1}}^{\{1,2,3\}}.$$

Conclusion

- nonlinear reco equation (=reaction system) forward in time solved via linear Markov chain backward in time
- further reaction systems where this may work?

References

review article:

E. Baake and M. Baake, Ancestral lines under recombination,
in: Probabilistic Structures in Evolution
(E. Baake and A. Wakolbinger, eds.), EMS Press, Berlin, 2021

and references therein