Optimal Adaptation via Controlled Mutation and Recombination in Evolutionary Algorithms and in Nature

Roman V. Belavkin

Faculty of Science and Technology
Middlesex University, London NW4 4BT, UK

December 11, 2017
Bioinformatics Educational Seminar
Tokyo University of Science
Adaptation rate and AMR

Evolutionary Dynamics

Mutation

Recombination
Adaptation rate and AMR

Evolutionary Dynamics

Mutation

Recombination
Anti-Microbial Resistance (AMR) Challenge

- In autumn 2016, all members of the United Nations agreed to sign a declaration to fight antimicrobial resistance, the secretary general declaring it a ‘fundamental threat’ to global health and safety.
- New antibiotics can be developed, but they are more expensive than their predecessors (Gelband et al., 2015).
- Bacteria can adapt in just a few hours after introduction of antibiotics (Zhang et al., 2011).

Objective

Understanding how can microbes achieve such an astonishing rate of adaptation.
Mechanisms of adaptation

Mutation $a \rightarrow b$

\[
\begin{align*}
    a &= \text{ATAGGACTCA} \\
    b &= \text{ATGGGATTCA}
\end{align*}
\]

- Mutation rate $\mu$
- Fitness

Recombination $a \rightarrow ab$

\[
\begin{align*}
    a &= \text{ATAGGACTCA} \\
    b &= \text{GTAGGCTAGT} \\
    ab &= \text{ATAGGCTAGA} \\
    ba &= \text{GTAGGACTCT}
\end{align*}
\]

- Recombination rate $\mu$
- Fitness of parents
- Similarity of parents

Hypothesis

Organisms use optimal control of parameters of these mechanisms to maximize their adaptation rates.
Adaptation rate and AMR

Evolutionary Dynamics

Mutation

Recombination
Natural Selection and Adaptation

Definition (Fitness)

A real-valued function $f: \Omega \rightarrow \mathbb{R}$ such that

- $f(a) \leq f(b)$ means $b$ has a better reproductive success than $a$.
- $f(a) \leq f(b)$ means $b$ is better adapted to the environment than $a$.
- Fitness can be absolute or relative (e.g. may depend on other individuals or environment).

Beneficial if $f(a) < f(b)$
Neutral if $f(a) = f(b)$
Deleterious if $f(a) > f(b)$
Considered Euclidean space $\mathbb{R}^l$ of $l$ traits

There exists an optimal individual $T \in \mathbb{R}^l$

Fitness = negative Euclidean distance to $T$:

$$f(a) = -d(T, a)$$

Probability of adaptation:

$$P_+ := P(m < n \mid n, r) = \sum_{m=0}^{n-1} P(m \mid n, r)$$

(Fisher, 1930):

- How does $P_+$ depend on mutation radius $r$?
- $P_+$ decreases exponentially with $r$ for all $n = d(T, a) \in [0, \infty)$
- Evolution is more likely to occur via small mutations
Geometry of a Hamming space $\mathcal{H}_\alpha^l$

- Instead of Euclidean space consider Hamming space of DNA strings:
  
  $$\mathcal{H}_\alpha^l := \{0, \ldots, \alpha - 1\}^l$$

- $\{0, \ldots, \alpha - 1\}$ finite alphabet of size $\alpha$:
  
  $$\{0, 1, 2, 3\} \quad \{A, C, T, G\} \quad \alpha = 4$$

- $\{0, \ldots, \alpha - 1\}^l$ set of all strings of length $l$

- Hamming distance between $a, b \in \mathcal{H}_\alpha^l$:
  
  $$d(a, b) = l - \sum_{i=1}^{l} \delta_{a_i b_i}, \quad \delta_{a_i b_i} = \begin{cases} 
  1 & \text{if } a_i = b_i \\
  0 & \text{otherwise} 
  \end{cases}$$

- Hamming space $\mathcal{H}_\alpha^l$ has finite diameter $l$. 
Evolutionary Dynamics

- Probability space \((\Omega \equiv \mathcal{H}_s^l, 2^\Omega, P_t)\) at time \(t \geq 0\)
- \(P_t(\omega)\) is a probability distribution of genotypes at time \(t \geq 0\)
- Distribution of fitness levels

\[
P_t(n) := P_t\{\omega : f(\omega) = n\}
\]

- Markov evolutionary dynamics:

\[
P_{t+s}(m) = \left[ \sum_{n=0}^{l} P_{\mu}(m \mid n)(\cdot) \right]^s P_t(n)
\]

Optimal evolutionary dynamics

Find optimal \(T(\cdot) = \sum_{n=0}^{l} P(m \mid n)(\cdot)\)
Adaptation rate and AMR

Evolutionary Dynamics

Mutation

Recombination
Simple Point Mutation

Assumptions

Independence: mutations at different sites are independent:

\[ P(b \mid a) = P(b_1, \ldots, b_l \mid a_1, \ldots, a_l) = \prod_{i=1}^{l} P(b_i \mid a_i) \]

Neutrality: selection does not act on sites \( P(b_i \mid a_i) = P(b_j \mid a_j) \).

One parameter: \( P(b_i \mid a_i) = \mu \) for all \( i \in \{0, \ldots, \alpha - 1\} \), called the mutation rate (Jukes & Cantor, 1969).

Random mutation radius

In this case mutation radius \( r \) is binomial random variable:

\[ P_\mu(r) = \binom{l}{r} \mu^r (1 - \mu)^{l-r} \]
Problem

*Find conditional probability* $P(d(\top, b) = m \mid d(\top, a) = n, d(a, b) = r)$

\[
P_{t+1}(m) = \sum_{n=0}^{l} \left( \sum_{r=0}^{l} P(m \mid n, r) P(r \mid n) \right) P_t(n)
\]
Mutation

Theorem (Belavkin (2011))

\[
P(m \mid n, r) = \sum_{r_+ = 0}^{r} (\alpha - 2)r_0^\left(\frac{n-r_+}{r_0}\right)(\alpha - 1)^r \left(\frac{l-n}{r_-}\right) \left(\frac{n}{r_+}\right)
\]

\[
\left(\alpha - 1\right)^r \binom{l}{r}
\]
Effect of the mutation radius $r$

\[ \mathbb{E}_P \{ m \mid n, r \} = n + \left( 1 - \frac{n}{l(1 - 1/\alpha)} \right) r \]
Optimal Control of Mutation Rate

The fact that $P_+$ varies with mutation radius and distance to optimum means that we can find optimal mutation rate control functions $\mu(n)$, which are solutions to the following optimization problems (Belavkin, 2010, 2012):

- Maximum adaptation in no more than $\lambda$ generations

$$\text{maximize} \quad \mu(x) \quad \mathbb{E}\{f_{s+t}\} \quad \text{subject to} \quad t \leq \lambda$$

- Maximum adaptation in no more than $\lambda$ bits between $p_s$ and $p_{s+t}$:

$$\text{maximize} \quad \mu(x) \quad \mathbb{E}\{f_{s+t}\} \quad \text{subject to} \quad I(p_{s+t}, p_s) \leq \lambda$$

- Cumulative criterion:

$$\sup_{\mu(x)} \sum_{t=0}^{\lambda} \mathbb{E}\{f_{s+t}\} \leq \sum_{t=0}^{\lambda} \sup_{\mu(x)} \mathbb{E}\{f_{s+t}\}$$
Optimal mutation rate control functions in $\mathcal{H}_4^{10}$

Expected Fitness in Time

\[ d_n(T, a) = \max_\mu P_\mu(m < n | n) \]

Constant $1/l$
Step
Linear $n/l$

\[ P_0(m < n | n) \]

Distance to optimum $n = d(T, a)$

Generation $t$
Evolution as an Information Dynamic System

- EPSRC Sandpit ‘Math of Life’ (July, 2009):

- Three year project (2010–13)
- Followed by two BBSRC project.

  Middlesex University : Roman Belavkin
  University of Warwick : John Aston
  University of Keele : Alastair Channon & Elizabeth Aston
  University of Manchester : Chris Knight, Rok Krašovec & Danna Gifford
Mutation Rate Control in *E. coli*

- Used strains of *Escherichia coli* K-12 MG1665
- Fluctuation test using media $50 \mu g/ml$ of Rifamipicin
- Estimated mutation rates $\mu$ in *E. coli* strains grown in Davis minimal medium with different amount of glucose.
Experimental Results (Krašovec et al., 2014)

- Strong relationship between $\mu$ and density of cells ($p < .0001$).
- No such relationship in the $luxS$ quorum sensing mutant ($p = .0234$).

**Experimental Results (Krašovec et al., 2014)**

- Strong relationship between $\mu$ and density of cells ($p < .0001$).
- No such relationship in the *luxS* quorum sensing mutant ($p = .0234$).

Experimental Results (Krašovec et al., 2014)

- Strong relationship between $\mu$ and density of cells ($p < .0001$).
- No such relationship in the luxS quorum sensing mutant ($p = .0234$).

Plastic mutation rates in bacteria (Krašovec et al., 2017)

Plastic mutation rates in yeast (Krašovec et al., 2017)

Plastic rates in all domains of life (Krašovec et al., 2017)

>70 years of published data (1943–2016), 67 studies, 26 species.

Adaptation rate and AMR

Evolutionary Dynamics

Mutation

Recombination
Recombination

Mechanisms of adaptation

Mutation $a \mapsto b$

\[
\begin{align*}
a &= \text{ATAGGACTCA} \\
b &= \text{ATG GGATTCA}
\end{align*}
\]

- Mutation rate $\mu$
- Fitness

Recombination $a \mapsto ab$

\[
\begin{align*}
a &= \text{ATAGGACTCA} \\
b &= \text{GTAGGCTAGT} \\
ab &= \text{ATAGGCTAGA} \\
ba &= \text{GTAGGACTCT}
\end{align*}
\]

- Recombination rate $\mu$
- Fitness of parents
- Similarity of parents

Horizontal gene transfer (HGT)

A form of recombination — part of a DNA string is picked up from another cell or from extracellular environment.

May occur via different mechanisms (e.g. conjugation, transduction and natural transformation).
Problem

Find \( P(d(\top, ab) = m, d(\top, ba) = m') \mid d(\top, a) = n, d(\top, b) = k, d(a, b) = h, r) \)

\[
P(m, m') = \sum_{n=0}^{l} \sum_{k=0}^{l} \left( \sum_{h=0}^{l} \sum_{r=0}^{l} P(m.m' \mid n, k, h, r) \right) P(r, h \mid n, k) P(n, k)
\]
Probability of recombination onto $S(T, m)$

**Theorem**

$$P(m \mid n, k, h, r) = \frac{\sum_{n_+} (\alpha - 2)^{n_0} \binom{n-n_+}{n_0} (\alpha - 1)^{n_-} \binom{l-n}{n_-} \binom{n}{n_+} \sum_{r_+} \binom{l-n_+-n_-}{r_-} \binom{n_-}{r_-} \binom{n_+}{r_+}}{\binom{l}{r} \sum_{n_+} (\alpha - 2)^{n_0} \binom{n-n_+}{n_0} (\alpha - 1)^{n_-} \binom{l-n}{n_-} \binom{n}{n_+}}$$

where

$$n_+ - n_- = n - k \Rightarrow n_- = n_+ - (n - k) \geq 0$$
$$n_+ + n_- + n_0 = h \Rightarrow n_0 = h - 2n_+ + (n - k) \geq 0$$
$$r_+ - r_- = n - m \Rightarrow r_- = r_+ - (n - m) \geq 0, \quad r - r_+ - r_- \geq 0$$

**Triangle inequalities**

$$|n - h| \leq k \leq n + h, \quad |n - \min\{r, h\}| \leq m \leq n + \min\{r, h\}$$
Effect of the second parent’s distance $k$

\[
\mathbb{E}_P \{ m \mid n, k, h, r \} = n + \frac{(k - n)}{l} r
\]
Effect of the recombination height $h$

\[
\mathbb{E}_P\{m \mid n, k, h, r\} = n + \frac{(k - n)}{l} r
\]
Effect of the recombination radius $r$

\[
\mathbb{E}_P\{m \mid n, k, h, r\} = n + \frac{(k - n)}{l} r
\]
Acknowledgements

U. of Manchester: Christopher Knight, Rok Krašovec, Danna Gifford
U. of Keele: Alastair Channon & Elizabeth Aston
U. of Warwick: John Aston (now in Cambridge)

EPSRC  SANDPIT: Evolution as an Information Dynamic System (EP/H031936/1)
BBSRC  The theory and practice of evolvability: Effects and mechanisms of mutation rate (BB/L009579/1)
BBSRC  Adaptive landscapes of antibiotic resistance: population size and 'survival-of-the-flattest' (BB/M021106/1)

Thank you!
Adaptation rate and AMR

Evolutionary Dynamics

Mutation

Recombination


References
rifampicin resistance depends on escherichia coli cell-cell interactions. *Nature Communications*, 5(3742).
