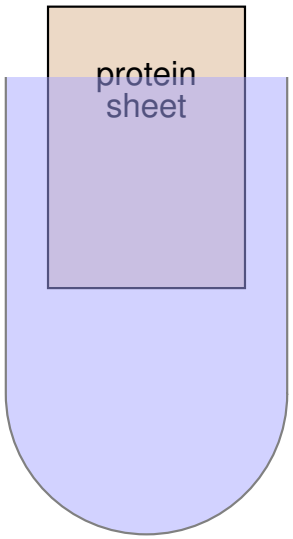
In-vitro selection algorithm

- for finding RNA strand that binds to protein

Make tube of *random* RNA strands



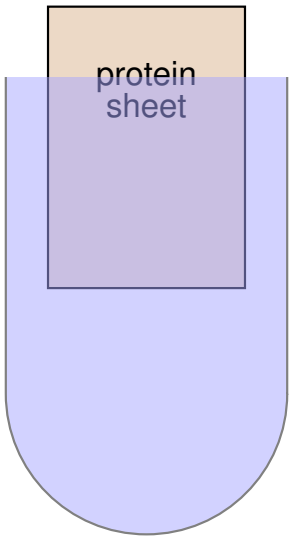
for t=1 to 20 do

1. Dip sheet coated with protein into tube
2. Pull out and wash off RNA that stuck
3. Multiply washed off RNA back to original amount (normalization)

In-vitro selection algorithm

- for finding RNA strand that binds to protein

Make tube of *random* RNA strands



for t=1 to 20 do

1. Dip sheet coated with protein into tube
2. Pull out and wash off RNA that stuck
3. Multiply washed off RNA back to original amount (normalization)

Can't be done with computer
because 10^{15} RNA strands per liter of soup

Basic scheme

Start with unit amount of random RNA

Loop

1. Functional separation into
   good RNA and bad RNA
2. Amplify good RNA to unit amount

Duplicating DNA with PCR

needed for normalization step

- Invented by Kary Mullis 1985
- Heat - double stranded DNA comes apart
- Cool - Short primers hybridize at the ends
- Taq Polymerase runs along DNA strand and complements bases
- Back to double stranded DNA

Duplicating DNA with PCR

needed for normalization step

- Invented by Kary Mullis 1985
- Heat - double stranded DNA comes apart
- Cool - Short primers hybridize at the ends
- Taq Polymerase runs along DNA strand and complements bases
- Back to double stranded DNA

- Ideally all DNA strands multiplied by factor of 2

The caveat

- Not many interesting/specific functional separations found
- Need high throughput for functional separation

Mathematical description of in-vitro selection

- Number the unknown RNA strands as $1, 2, \ldots n$ where $n \approx 10^{15}$ (# of strands in 1 liter)
- $s_i \geq 0$ is share of RNA i

Mathematical description of in-vitro selection

- Number the unknown RNA strands as $1, 2, \ldots n$ where $n \approx 10^{15}$ (# of strands in 1 liter)
- $s_i \geq 0$ is share of RNA i
- Contents of tube represented as share vector $\boldsymbol{s} = (s_1, s_2, \ldots, s_n)$
- When tube has unit amount, then $s_1 + s_2 + \cdots + s_n = 1$

Mathematical description of in-vitro selection

- Number the unknown RNA strands as $1, 2, \ldots n$ where $n \approx 10^{15}$ (# of strands in 1 liter)
- $s_i \geq 0$ is share of RNA i
- Contents of tube represented as share vector $\boldsymbol{s} = (s_1, s_2, \ldots, s_n)$
- When tube has unit amount, then $s_1 + s_2 + \cdots + s_n = 1$
- $W_i \in [0, 1]$ is the fraction of one unit of RNA i that is good
- W_i is fitness of RNA strand i for attaching to protein
- Protein represented as fitness vector $\boldsymbol{W} = (W_1, W_2, \ldots, W_n)$

Mathematical description of in-vitro selection

- Number the unknown RNA strands as $1, 2, \ldots n$ where $n \approx 10^{15}$ (# of strands in 1 liter)
- $s_i \geq 0$ is share of RNA i
- Contents of tube represented as share vector $\boldsymbol{s} = (s_1, s_2, \ldots, s_n)$
- When tube has unit amount, then $s_1 + s_2 + \cdots + s_n = 1$
- $W_i \in [0, 1]$ is the fraction of one unit of RNA i that is good
- W_i is fitness of RNA strand i for attaching to protein
- Protein represented as fitness vector $\boldsymbol{W} = (W_1, W_2, \ldots, W_n)$
- Vectors $\boldsymbol{s}$, $\boldsymbol{W}$ unknown: Blind Computation!

Mathematical description of in-vitro selection

- Number the unknown RNA strands as $1, 2, \ldots n$ where $n \approx 10^{15}$ (# of strands in 1 liter)
- $s_i \geq 0$ is share of RNA i
- Contents of tube represented as share vector $\boldsymbol{s} = (s_1, s_2, \ldots, s_n)$
- When tube has unit amount, then $s_1 + s_2 + \cdots + s_n = 1$
- $W_i \in [0, 1]$ is the fraction of one unit of RNA i that is good
- W_i is fitness of RNA strand i for attaching to protein
- Protein represented as fitness vector $\boldsymbol{W} = (W_1, W_2, \ldots, W_n)$
- Vectors $\boldsymbol{s}$, $\boldsymbol{W}$ unknown: Blind Computation!

Strong assumption:

- Fitness W_i independent of share vector $\boldsymbol{s}$

Update

Good RNA in tube $\boldsymbol{s}$:

$$s_1 W_1 + s_2 W_2 + \cdots + s_n W_n = \boldsymbol{s} \cdot \boldsymbol{W}$$

Bad RNA:

$$s_1(1 - W_1) + s_2(1 - W_2) + \cdots + s_n(1 - W_n) = 1 - \boldsymbol{s} \cdot \boldsymbol{W}$$

- Amplification:
  - Good share of RNA i is $s_i W_i$ - multiplied by factor F
  - If precise, then all good RNA multiplied by same factor F
  - Final tube at end of loop

$$F\ \boldsymbol{s} \cdot \boldsymbol{W}$$

  - Since final tube has unit amount of RNA

$$F\ \boldsymbol{s} \cdot \boldsymbol{W} = 1 \text{ and } F = \frac{1}{\boldsymbol{s} \cdot \boldsymbol{W}}$$

  - Update in each loop

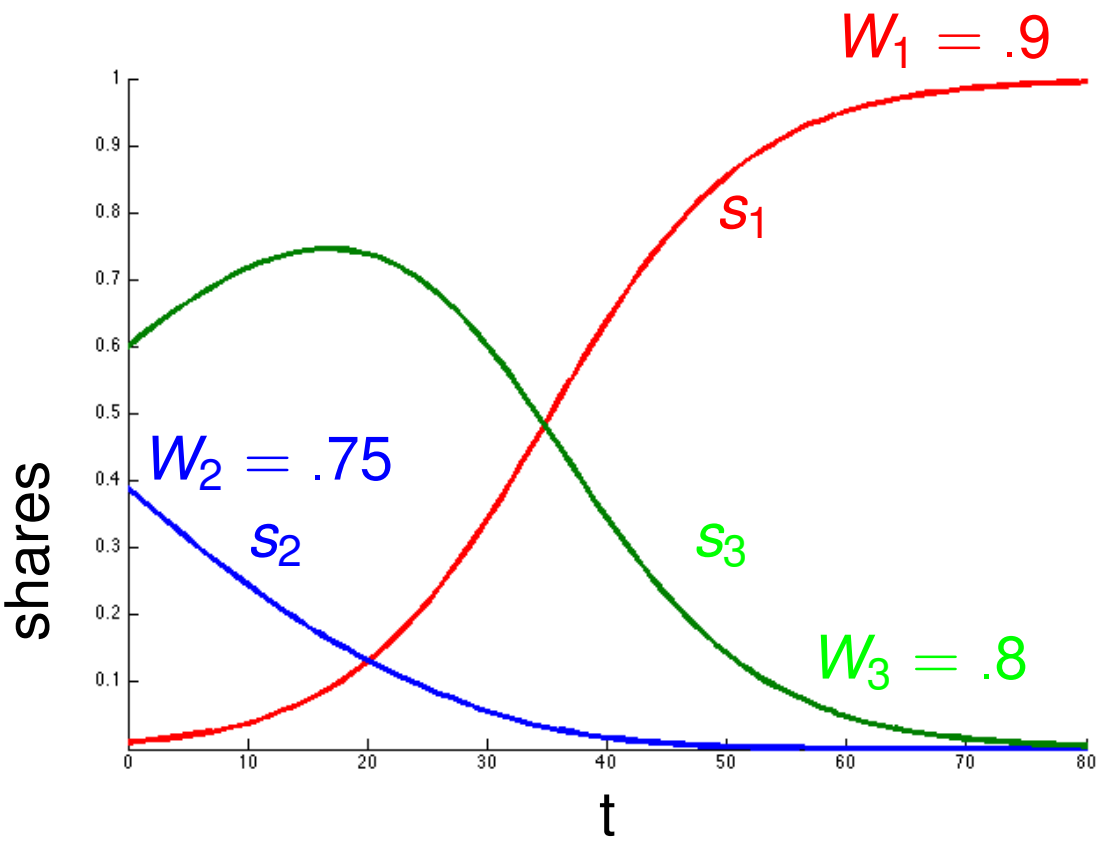
$$s_i := \frac{s_i W_i}{\boldsymbol{s} \cdot \boldsymbol{W}}$$

Implementing Bayes rule in RNA

- s_i is prior $P(i)$
- $W_i \in [0..1]$ is probability $P(Good|i)$
- $\boldsymbol{s} \cdot \boldsymbol{W}$ is probability $P(Good)$
-

$$s_i := \frac{s_i W_i}{\boldsymbol{s} \cdot \boldsymbol{W}} \quad \text{is Bayes rule} \quad \underbrace{P(i|Good)}_{\text{posterior}} = \frac{\overbrace{P(i)}^{\text{prior}} P(Good|i)}{P(Good)}$$

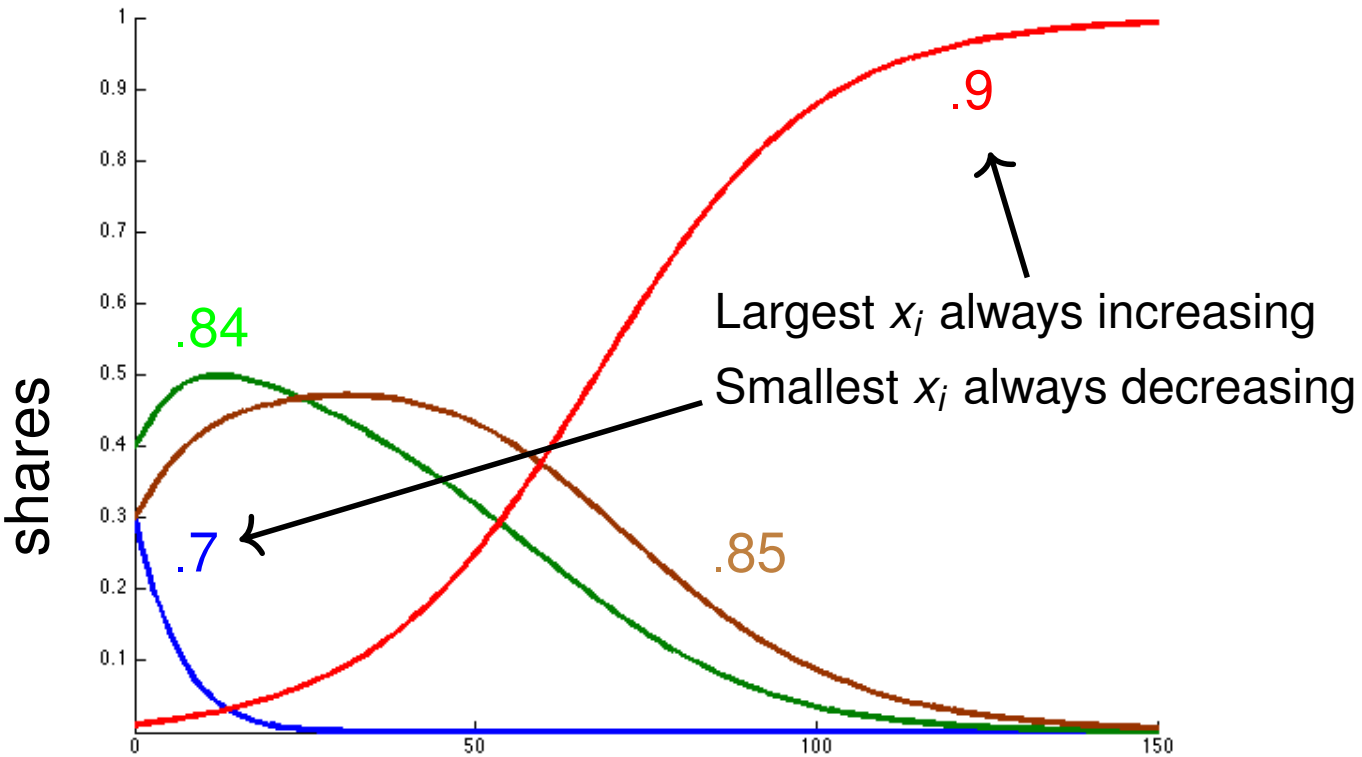
Iterating Bayes Rule with same data likelihoods



Initial $\boldsymbol{s} = (.01, .39, .6)$

$\boldsymbol{W} = (.9, .75, .8)$

$\max_i W_i$ eventually wins



$$s_{t,i} = \frac{s_{0,i} W_i^t}{\text{normalization}}$$

t is speed parameter

The best are *too* good

- Multiplicative update

$$s_i \sim s_i \underbrace{\text{fitness factor}_i}_{\geq 0}$$

The best are *too* good

- Multiplicative update

$$s_i \sim s_i \underbrace{\text{fitness factor}_i}_{\geq 0}$$

- Blessing

  Best get amplified exponentially fast
  Can do this with 10^{15} variables

The best are *too* good

- Multiplicative update

$$s_i \sim s_i \underbrace{\text{fitness factor}_i}_{\geq 0}$$

- Blessing

  Best get amplified exponentially fast
  Can do this with 10^{15} variables

- Curse

  The best wipe out the others
  Loss of diversity

View of evolution

Simple view

- Inheritance
  - mutation
  - selection for the fittest with multiplicative update