RNA Genotype-pgenotype map (cont.) & RNA world: sequence - structure- interactions

Course computational biology 2025; Paulien Hogeweg;

Theoretical Biology and Bioinformatics Grp Utrecht University

# RNA genotype-phenotype mapping, summary

- "smoothness within ruggedness"
  single mutation can be neutral and can change 'everything'
- percolating and intercalating neutral networks
  from smooth-rugged towards neutral networks
- no local peaks: detours
- phenotypic vs genotypic information threshold
- diffusion on neutral networks (D prop.to λ)
- adaptive walk with majority of neutral mutations
- reconciliation neutral and adaptive evolution
- RNA landscape "ideal" for evolution
- Evolution of "robustness" (higher lamda) ("flattest")
- Evolution of evolvability (iff innovation along NN)

Implications evolution towards higher robustness

- more robustness $--> $ more exploration (D λ)
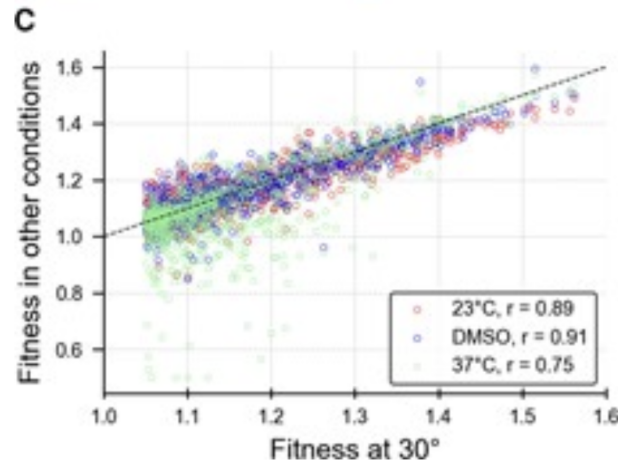
- evolution of evolvability *at level of population*

Experimental determination of tRNA GP map (23,284 muts robustness vs fitness: many more fit mutational neighbors
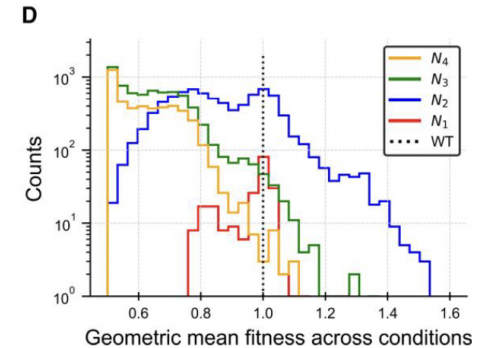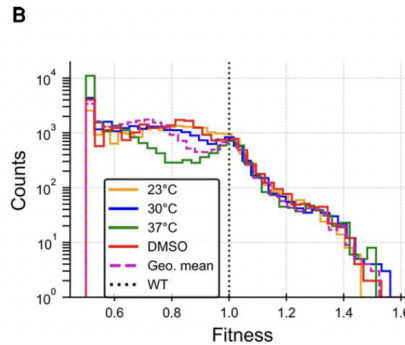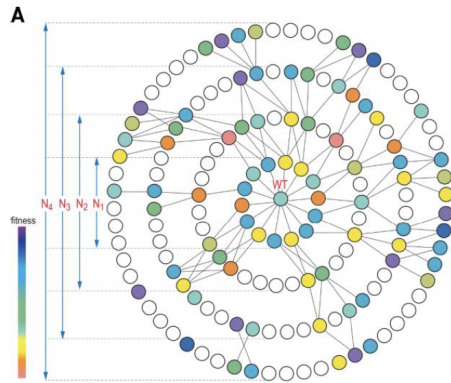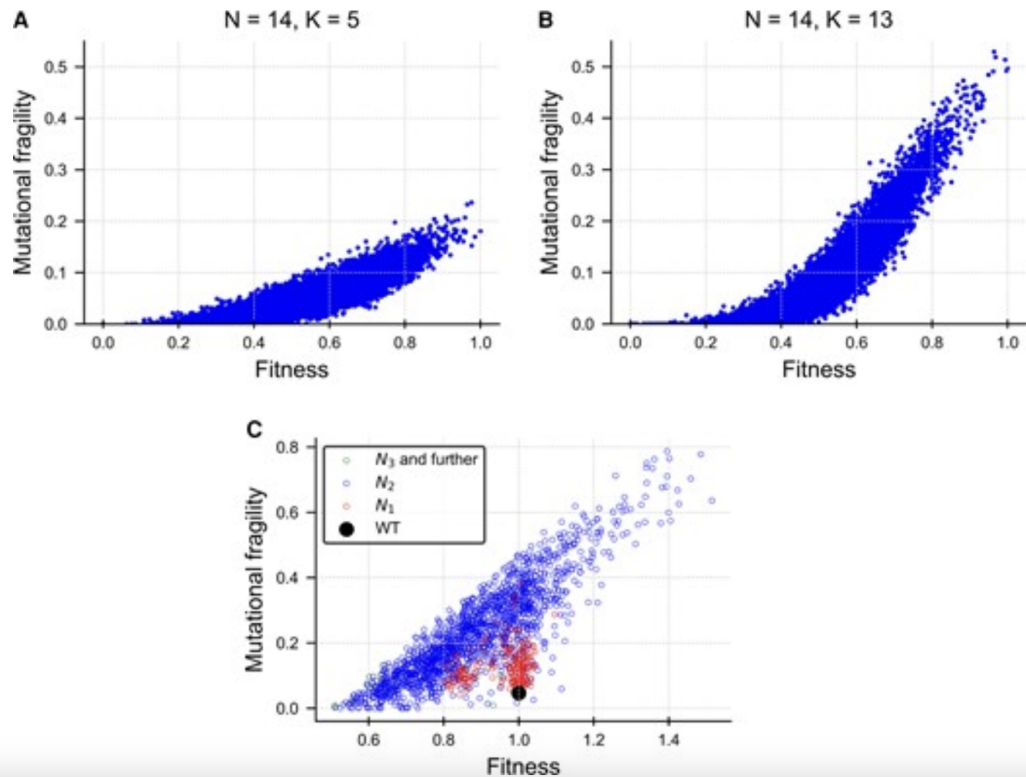

fitness = growth rate
in 4 conditions;

paths to fitter

genotypes exist!

Fitness Landscape Analysis of a tRNA Gene Reveals that the Wild Type Allele is Sub-optimal, Yet Mutationally Robust Gabzi, Pilpel, Friedlander, MBE 2022
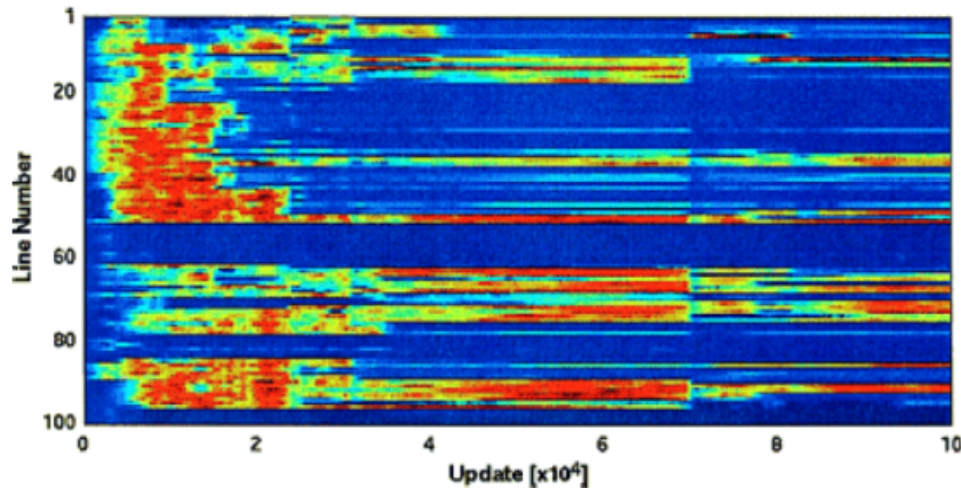
HOWEVER: Wildtype very robust!
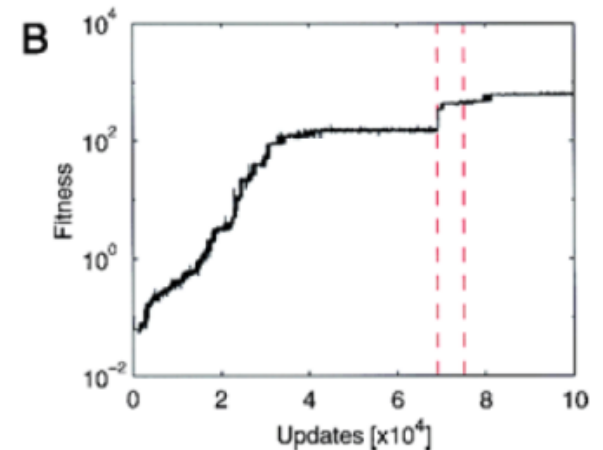
but note: less fit implies more robust

$$\phi_i = \frac{1}{|N_1(i)|} \sum_{j \in N_1(i)} |f_i - f_j|_+,$$

$$|x|_+ = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

AVIDA: Self-replicating computer program (Adami et al)
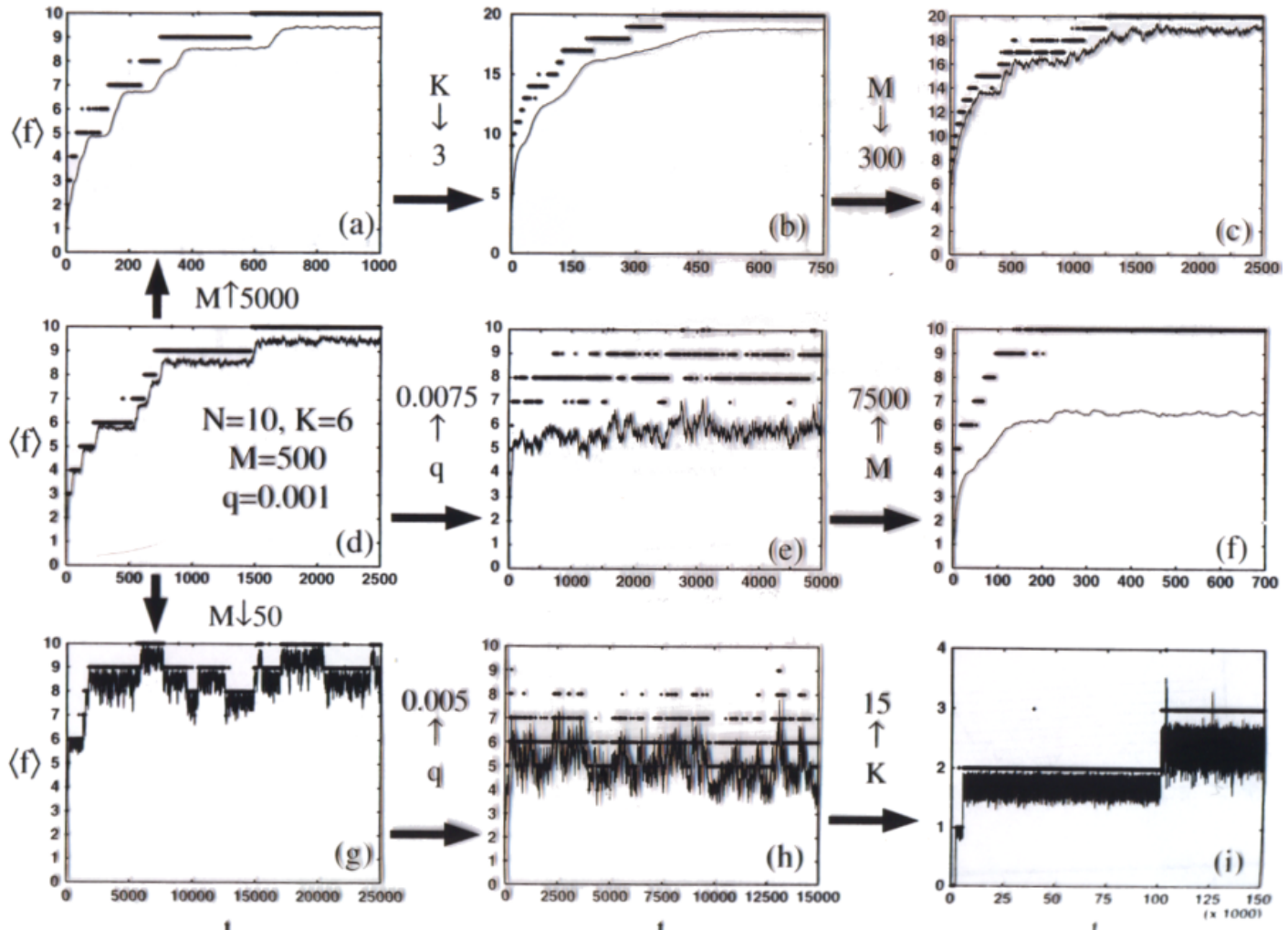

quasispecies

fitness

Population variability per position (gene) $p_i log(p_i)$

information accumulation upto information threshold..

Derived properties
JUST RNA?
or even just by wrongly computed (and 2 D) folding?

percolating neutral path; innovations
evolution toward robustness

NO......

similar (mutatis mutandis) properties in

Gene regulatory networks (A Wagner 2007a,b)

Protein folding (A Wagner 2010) (BUT! see later)

Metabolic networks (A.Wagner 2012)

see also books by A. Wagner

Manrubia et al From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics Physics of life 2021

From paradigm systems to general conclusions
vs
Studying "all" cases

NK landscapes (Kauffman):
Class of models to study impact of GP mapping on evolutionary dynamics.

N: number of properties (e.g. sequence length)
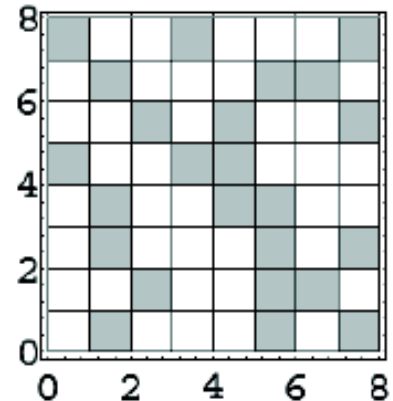K: number of "epistatic' interactions
most often 2 states per position

Fitness contribution of each $N.2^K$ states
chosen randomply. Fitness is sum of those

Calculate e.g pathlength to local peak
height of optima reached (etc.)

NO percolating, intercalating neutral paths
and its evolutionary consequences

versions include neutrality .....

Genotype-phenotype mapping: Coding structure

3 questions/answers:

Given code $->$ **which evolutionary dynamics?**

eg RNA folding: punctuated evolution etc.
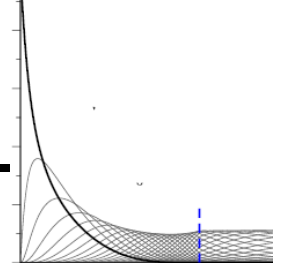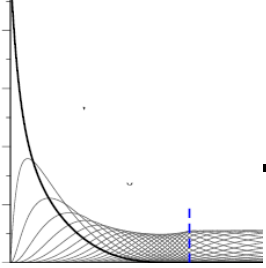
Given problem $->$ **how to code?**

expectation: smooth, non-redundant;

found intertwining neutral paths

Given evolutionary dynamics $->$ **which code?**

towards robustness, hence evolvability

2 images of RNA world

individual complexity

evolution toward robustness

interwinded neutral nets

1 sequence 2 ribozymes

innovation along neutral path
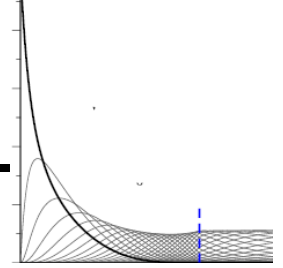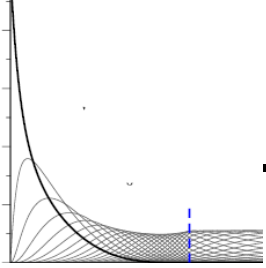
"RNA ideal evolvable molecule"

sequence to structure

ecosystem complexity

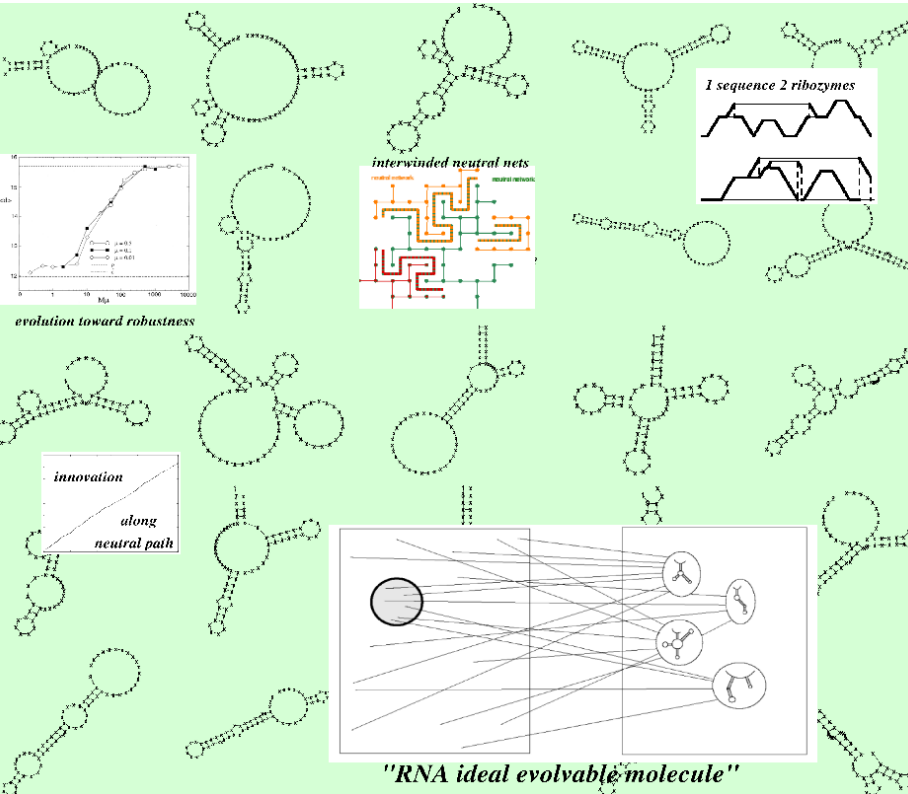$$X_{i+1} + X_i \longrightarrow 2X_{i+1} + X_i$$

X_i

t

replicator to wave/vesicle

the RNA world

individual complexity

evolution toward robustness

interwinded neutral nets

1 sequence 2 ribozymes

innovation along neutral path

"RNA ideal evolvable molecule"

ecosystem complexity

$$X_{i+1} + X_i \longrightarrow 2X_{i+1} + X_i$$

X_i

t

0   500

RNA (without world)
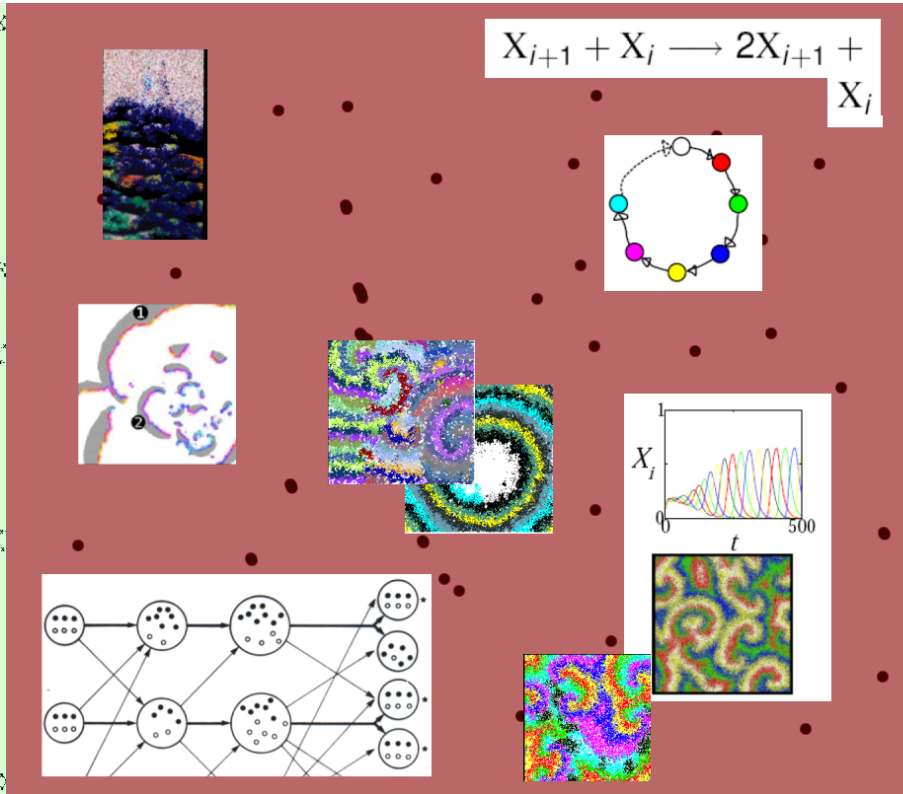
world (without RNA)

Today: RNA in space

Themes

Structured based modeling

Individual and/or ecosystem based complexity
ecosystem diversification and mutation rate

Evolution of coding structures (cont)
muliple coding
mutational neighborhood

RNA even more evolvable than seen so far

RNA world: Preconceived networks vs evolving individuals, emerging species, emerging interactions

---

- **structured individuals**
  here RNA sequences ($+$ and - strands)
  if folding in predefined structure: replicase

- **no predefined target or fitness**

- **no predefined interactions**
  but predefined reactions

DO SPECIES/ INTERACTION NETWORKS EVOLVE?

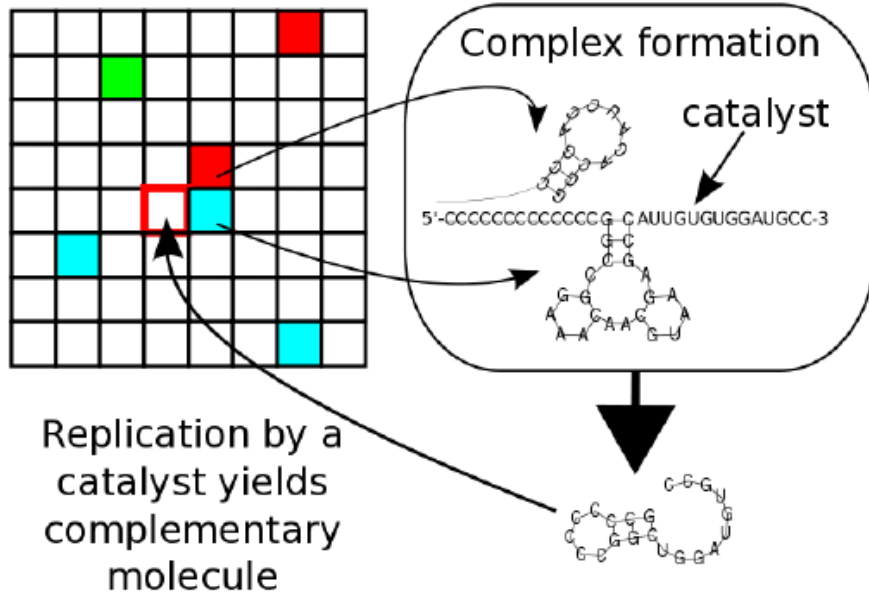**DOES EVOSYSTEM COMPLEXITY EVOLVE?**

genotype - phenotype - ecosystem mapping

Genotype → Phenotype → Interaction

sequence   structure   ecosystem

cat-C   cat-A

par-G   par-U
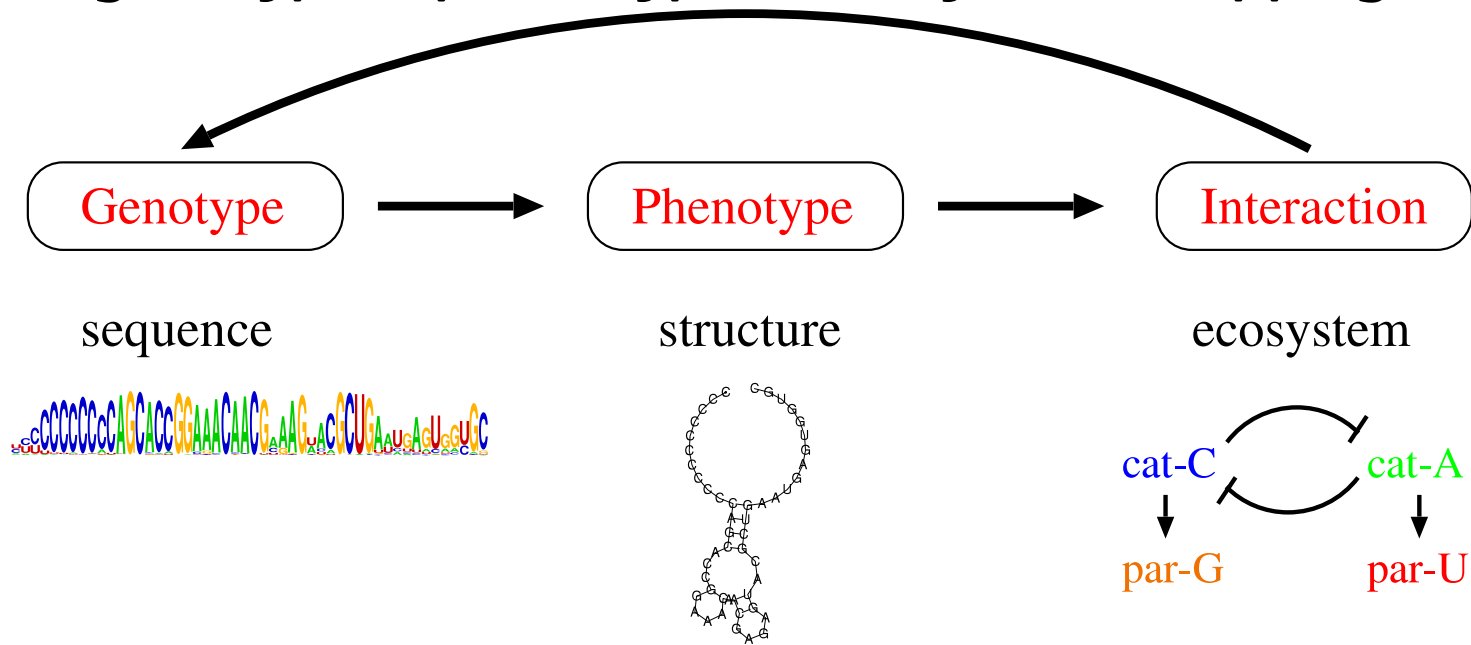
feedback from higher levels to lower levels in evolving system

interacting RNA's

Complex formation happens 5'-end $\rightarrow$ 3'-end

("strong" altruism)

Complex formation

catalyst

5'-CCCCCCCCCCCCCCG-C AUUGUGUGGAUGCC-3

Replication by a catalyst yields complementary molecule

1. $X + Y \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C_{X\sim Y}$ or $\underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} C_{Y\sim X}$

2. $C_{X\sim Y} + \Phi \xrightarrow{\kappa} X + Y + Y^{-1}$

3. $X \xrightarrow{d} \Phi$

only structure + reaction
no fitness function and no interaction predefined

Maximum mutation rates($\mu = .015$):
is only below informtion threshold
for evolved coding structure
ONE quasispecies

---

initial population

dynamics with mutation

= after stopping mutation
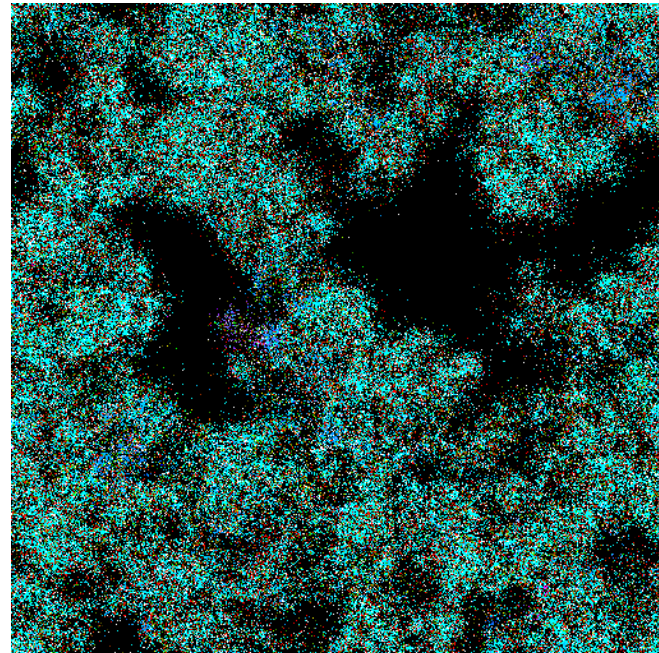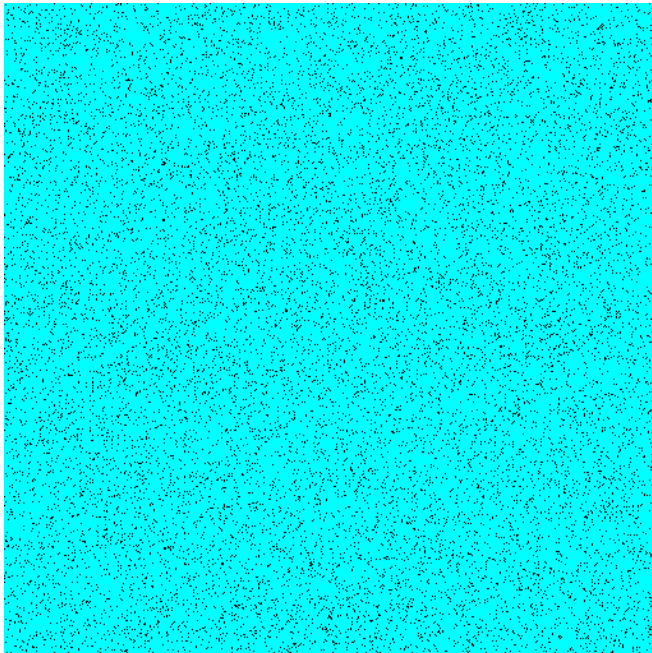
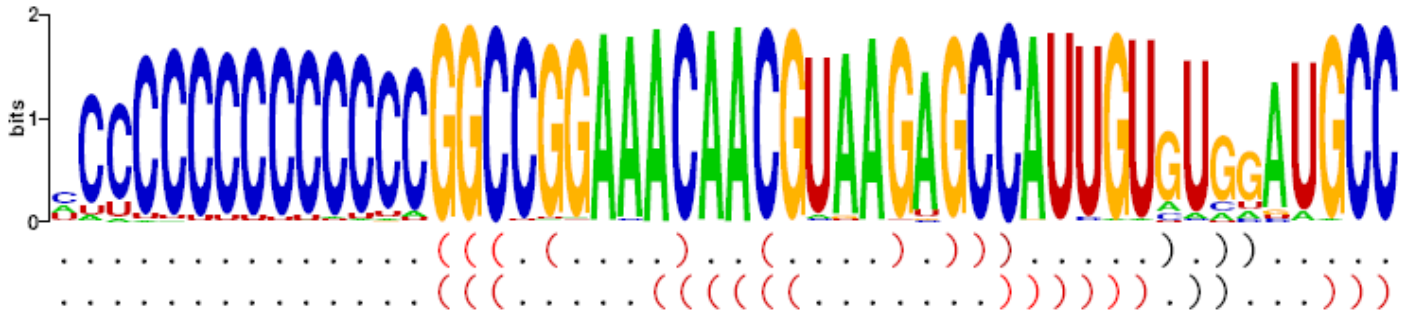# High mutation rate ($\mu = .015$) population structure of $+$ strands

■ Phylogeny reveals patterns in population of genotypes

Color	Types
Cyan	Catalyst
Red	Non-catalyst
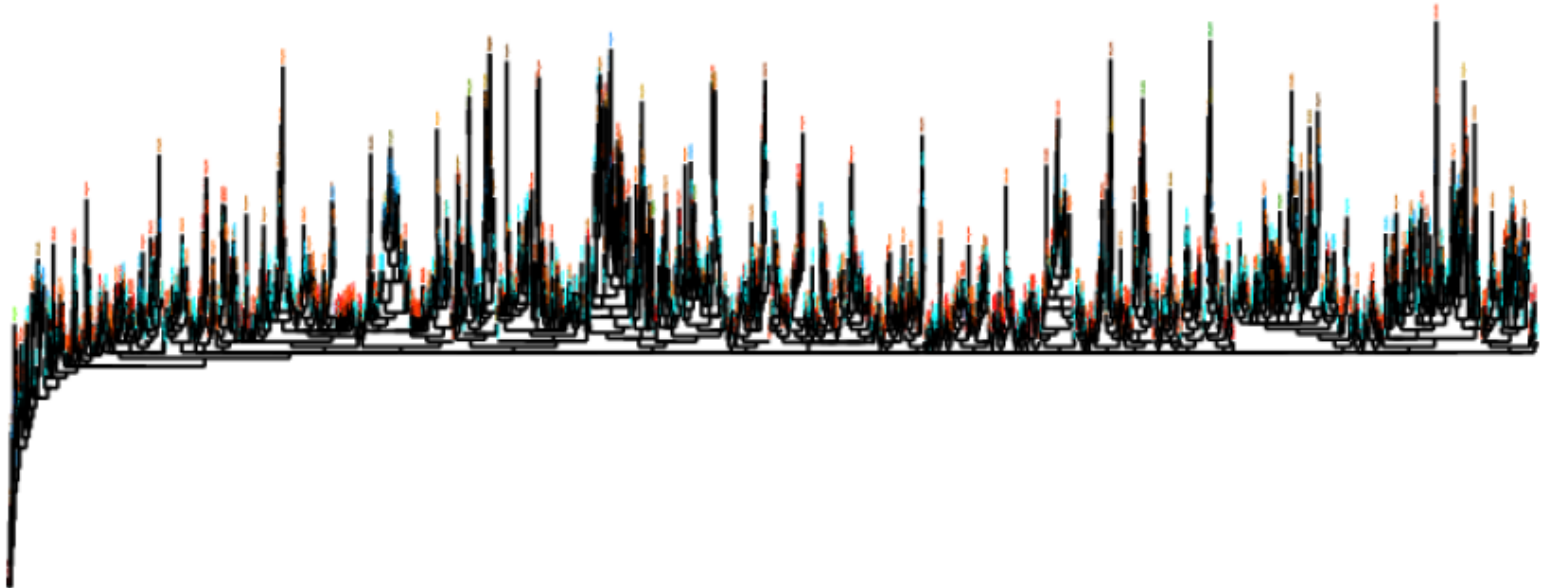
■ No clade patterns

■ Population is supported by various genotypes

■ One quasi-species

High mutation rates($\mu = .015$) sequence structure: symmetry breaking: only + strands catalytic

- Very high C frequency in 5'-end
- High G frequence in 3'-end
  - → many GC pairs
- Many interspersing U in 3'-end
  - → prevents base-pair formation *in homo*
- No 5'-end in template strand
  - → prevents non-functional complex formation
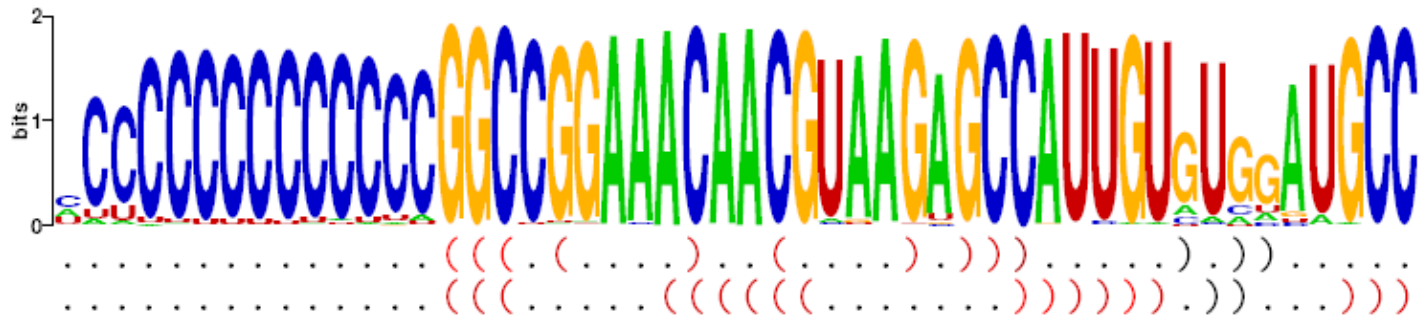
- Sequence is delicately tuned up
  - Almost all base-pairs are GC
  - Many other G and C that should not pair
    - → Difficult to form correct base-pairs
  - High sequence conservation in all positions
    - Loop region must be tuned too

lowering mutation rates ($\mu - .13$) : SPECIATION
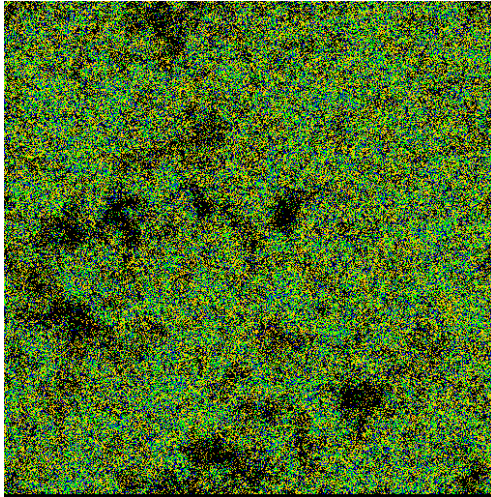
m=.014 $->$ m=.013
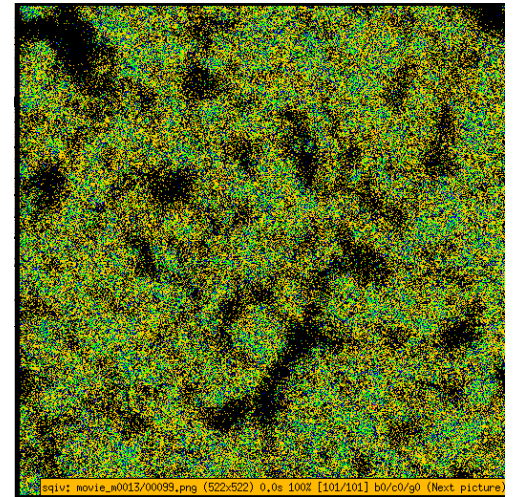
lowering mutation rates: ($\mu = .13$) population structure
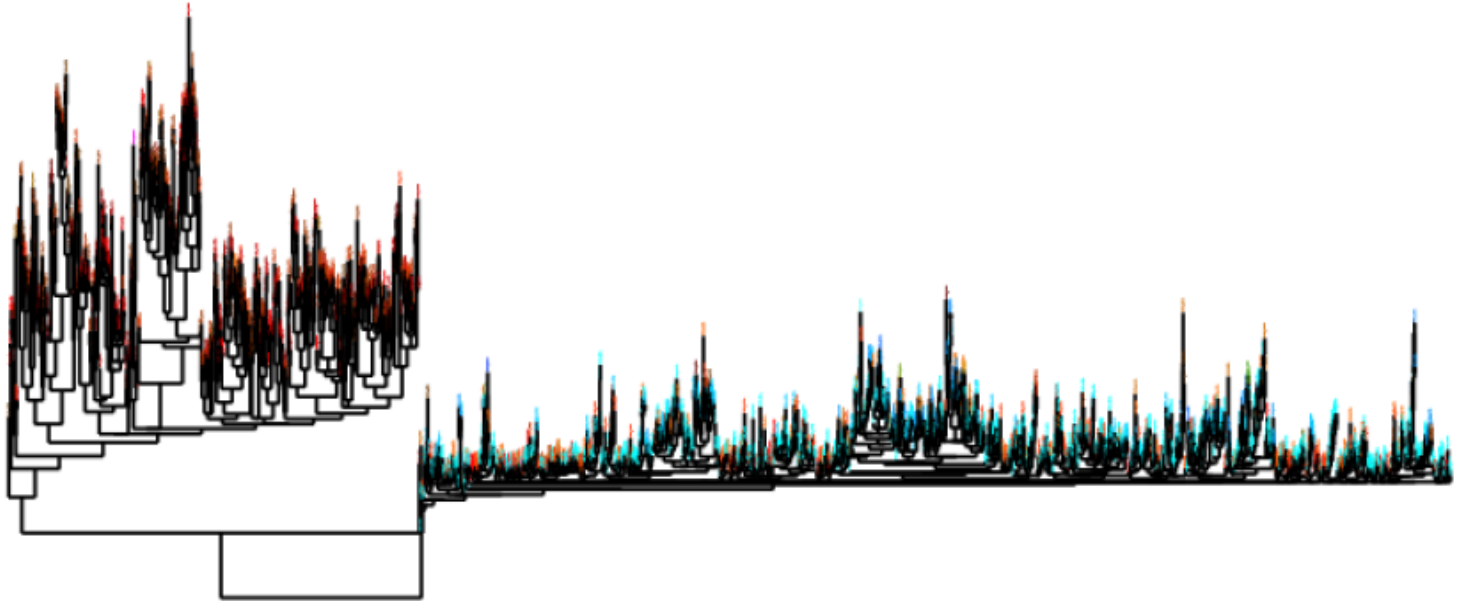
Color	Types
Cyan	Catalyst
Red	Non-catalyst

- Two quasi-species
  - distinct sequence classes
- Catalyst & Non-catalyst

lowering mutation rates: ($\mu = .13$) sequence structure

■ Catalyst

■ Non-catalyst →Parasite

■ No catalytic structure in both strands
■ Long 3'-end with many G
■ No 5'-end in both strands
■ Sequence conservation patterns & secondary structure

Parasite invades in periphery of QS

■ Population of Sequences
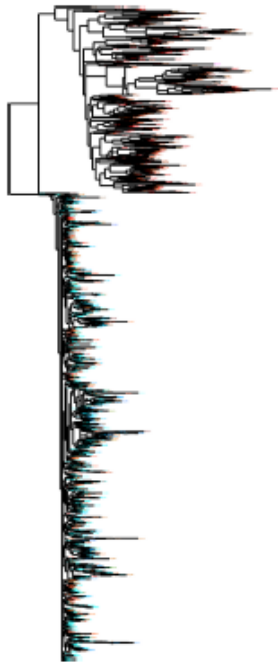
■ Genotype & Phenotype
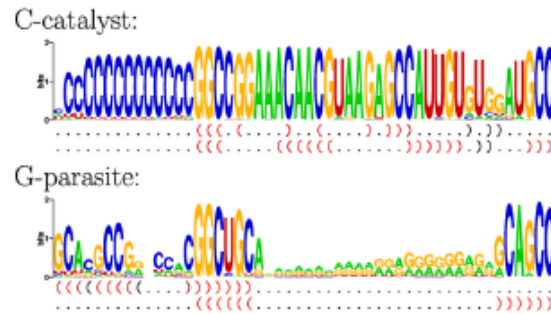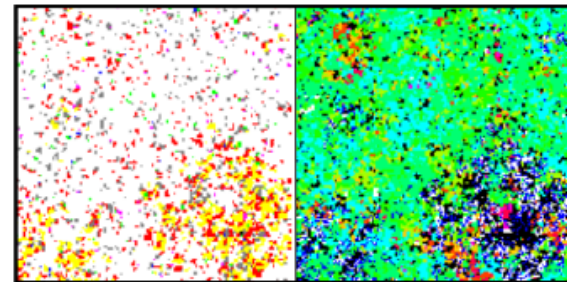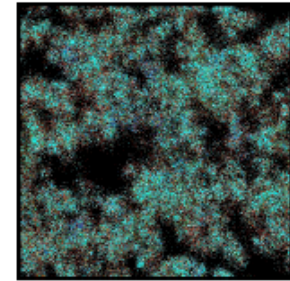
■ Space & Time

Hamm. dist. from master sequence
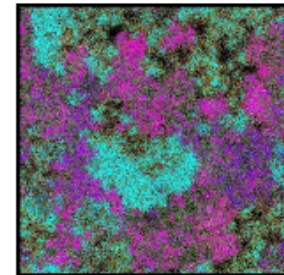
Lower mutation rate $\mu = .008$: 3 quasispecies
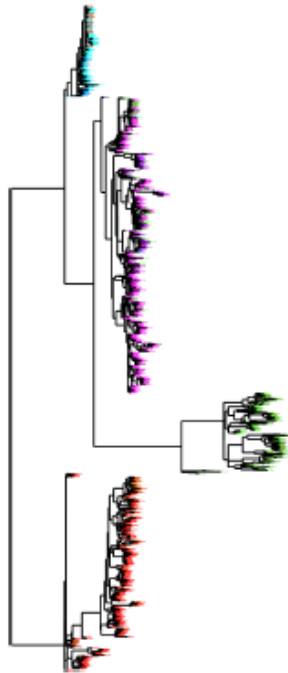
■ Population of Sequences

■ Genotype & Phenotype

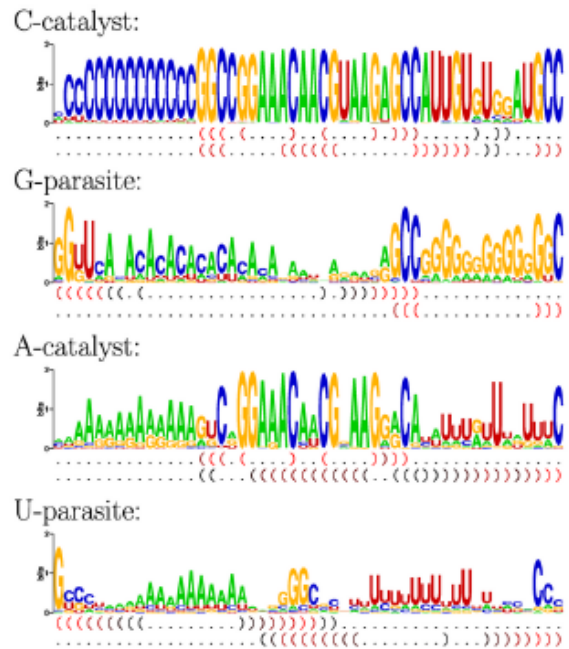■ Space & Time

C-catalyst:

G-parasite:
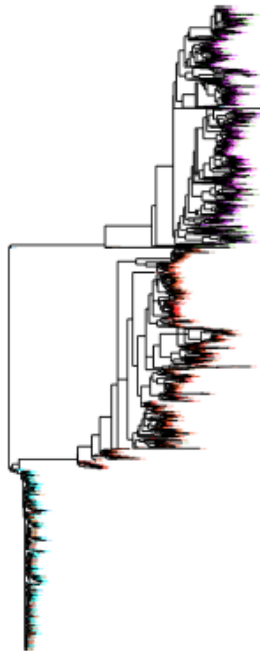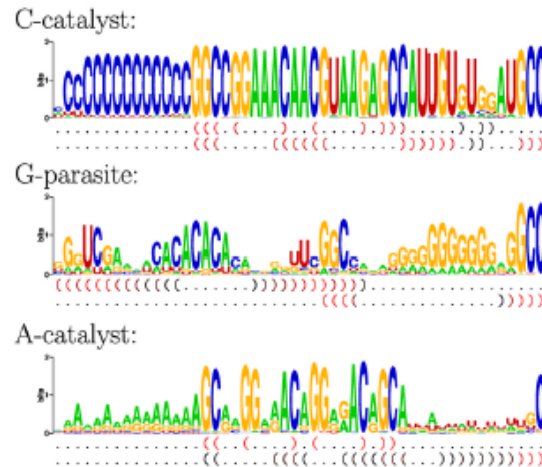
A-catalyst:

A catalyst: HIGH neutrality (ca 50%

Lower mutation rate $\mu = .004$: 4 quasispecies

■ Population
of Sequences

■ Genotype
& Phenotype

■ Space
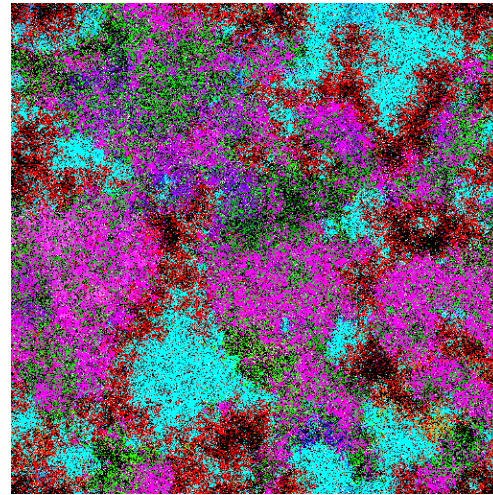& Time

C-catalyst:

G-parasite:

A-catalyst:
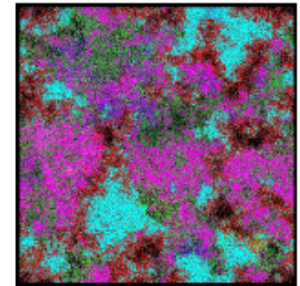
U-parasite:

evolved 4 species system; evolved interaction topology
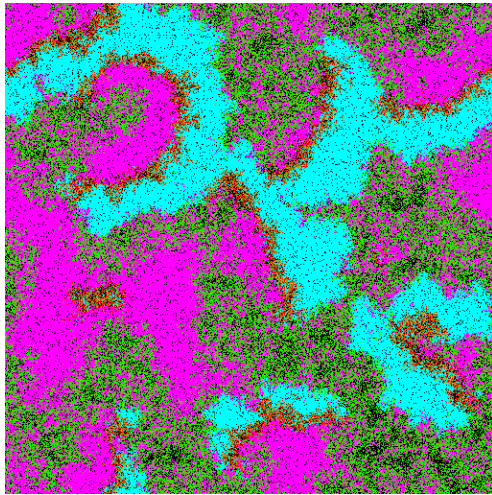
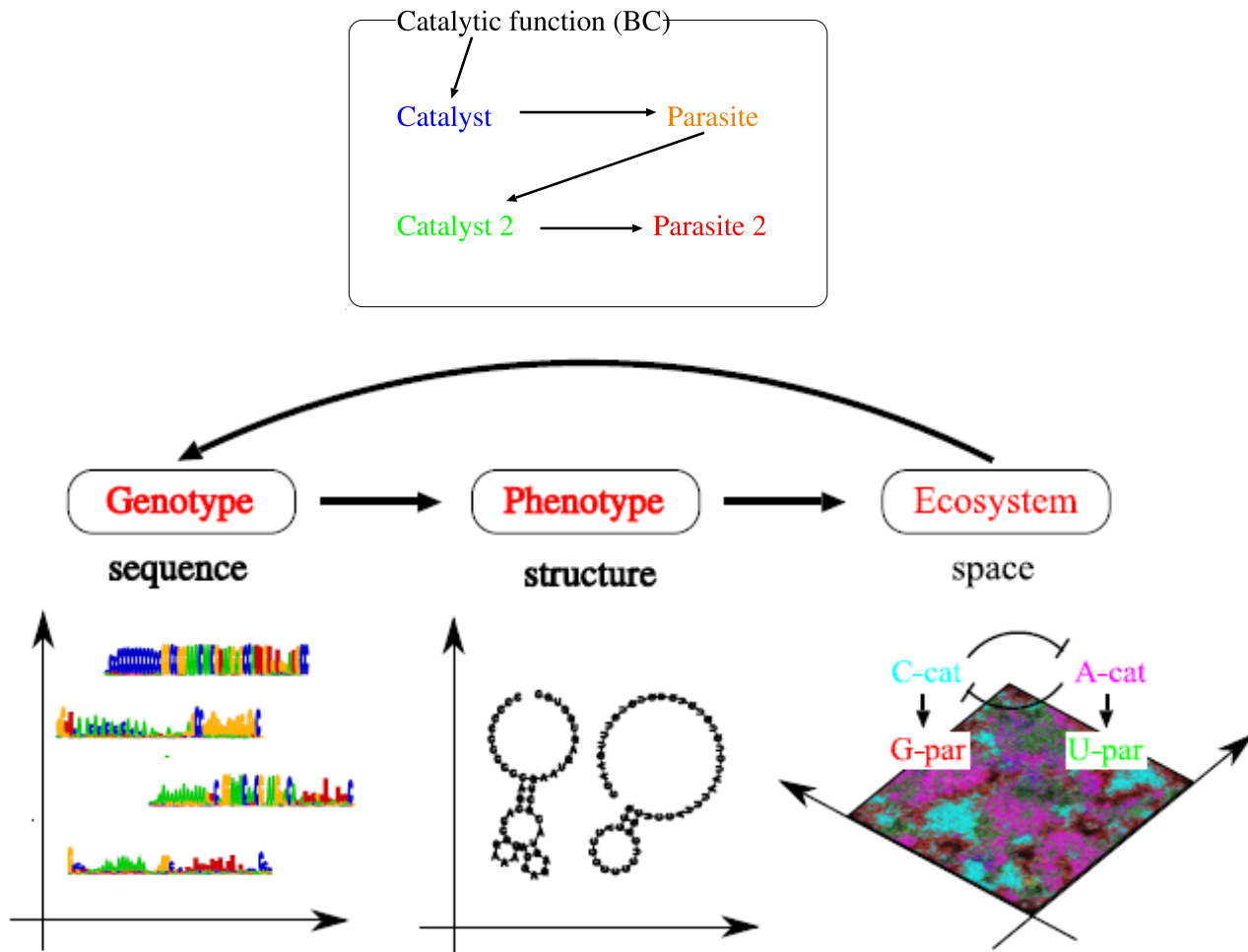ECOsystem ($\mu = 0$)   EVOL. system ($\mu = .004$)

Direct Interaction structure

	C-catalyst CYAN		A-catalyst MAGENTA		G-parasite RED		U-parasite GREEN	
	cat. str.	comp.	cat. str.	comp.	logo str.	comp.	logo str.	comp.
C-cat	**0.52**	**0.87**	0.36	0.45	**0.81**	**0.65**	0.26	0.36
A-cat	0.39	0.05	**0.50**	**0.77**	0.14	0.48	**0.63**	**0.55**

From Coding structure to ecosystem based information accumulation

Catalytic function (BC)

Catalyst → Parasite

Catalyst 2 → Parasite 2

Genotype → Phenotype → Ecosystem

sequence   structure   space

C-cat   A-cat

G-par   U-par

Conclusion

Very stable multi-(quasi)species systems evolves

Interaction topology different from anything studied before.

Variability increases with decreasing mutation rate
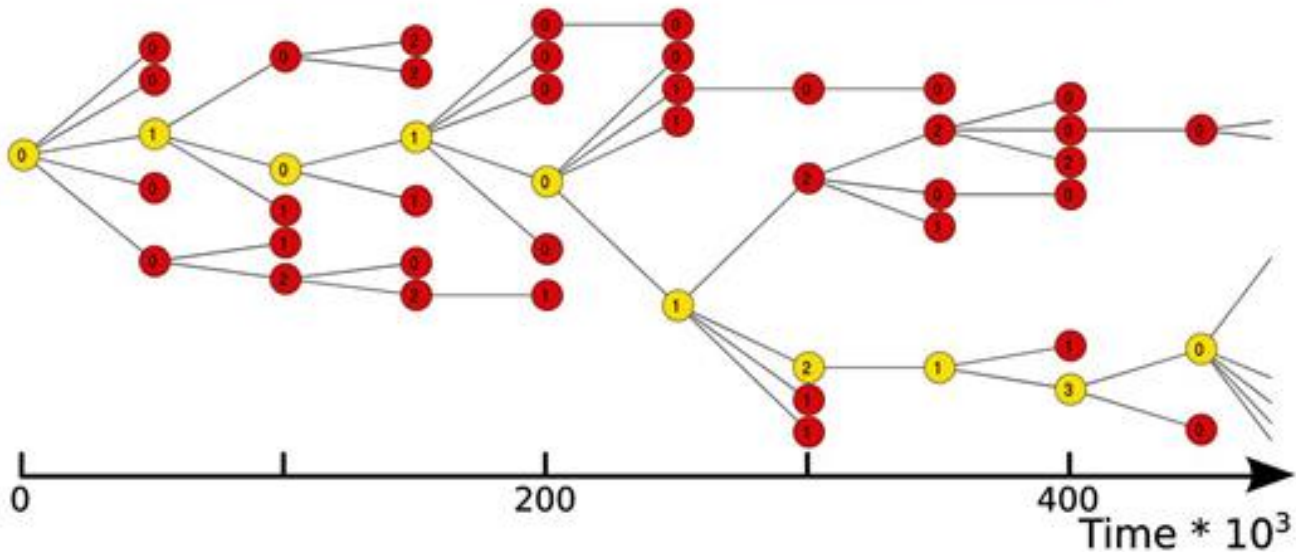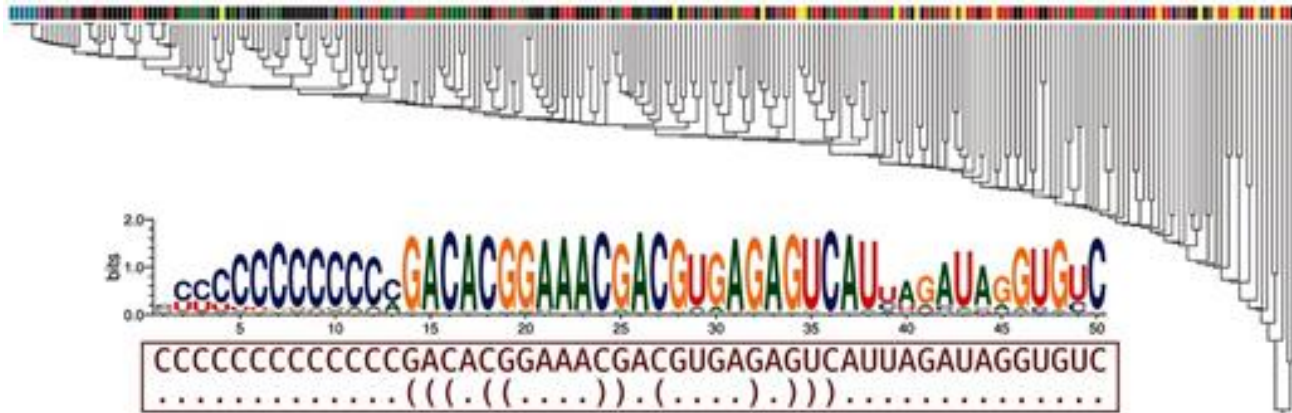speciation

Ecosystem based "solution" only at lower mutation rates

EVOLVED genotype-phenotype-interaction-spatial structure
mutual dependent (and "make sense" in relation to each
other)

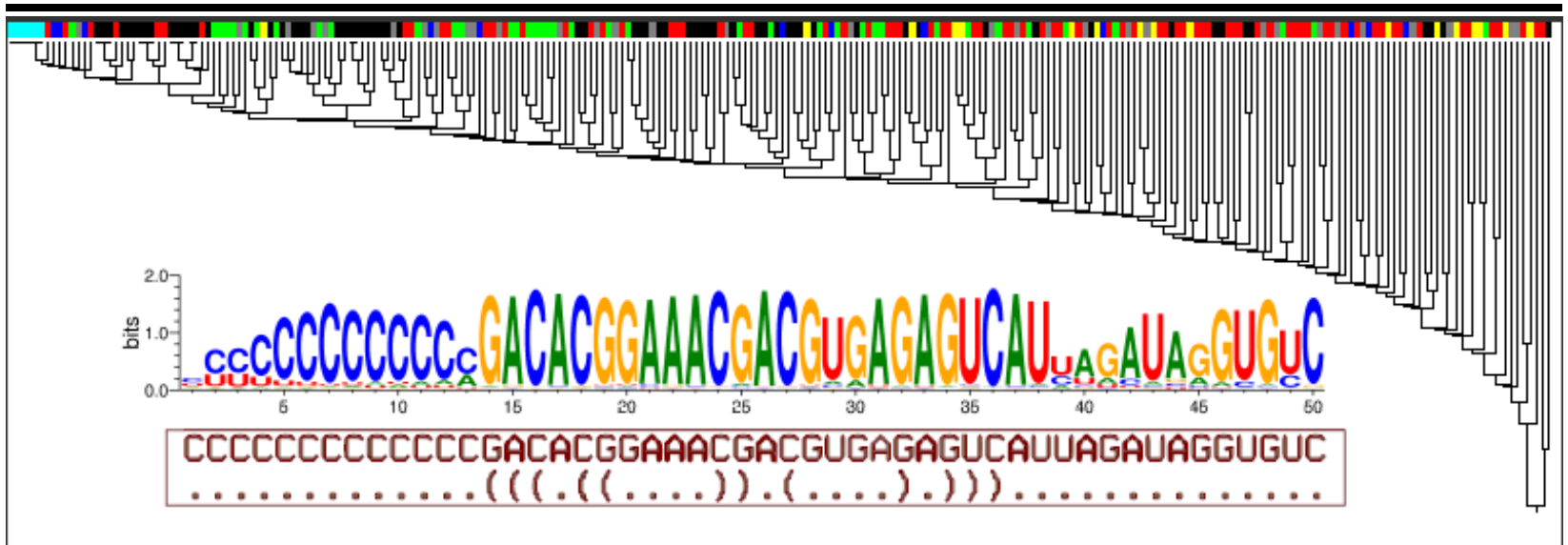**Evolved, niche dependent mutational landscape**
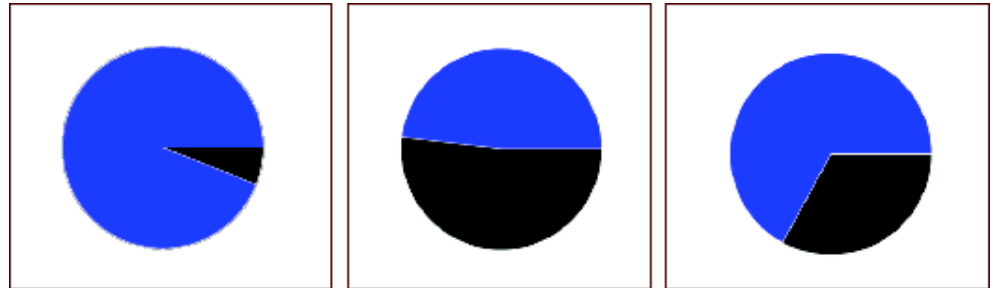
Evolution of coding structure at high mutation rates
Mutational neighborhood

High mut. rate: 1 quasispesies: mutations along line(s) of descent

High mut. rate: 1 quasispesies LOW variability

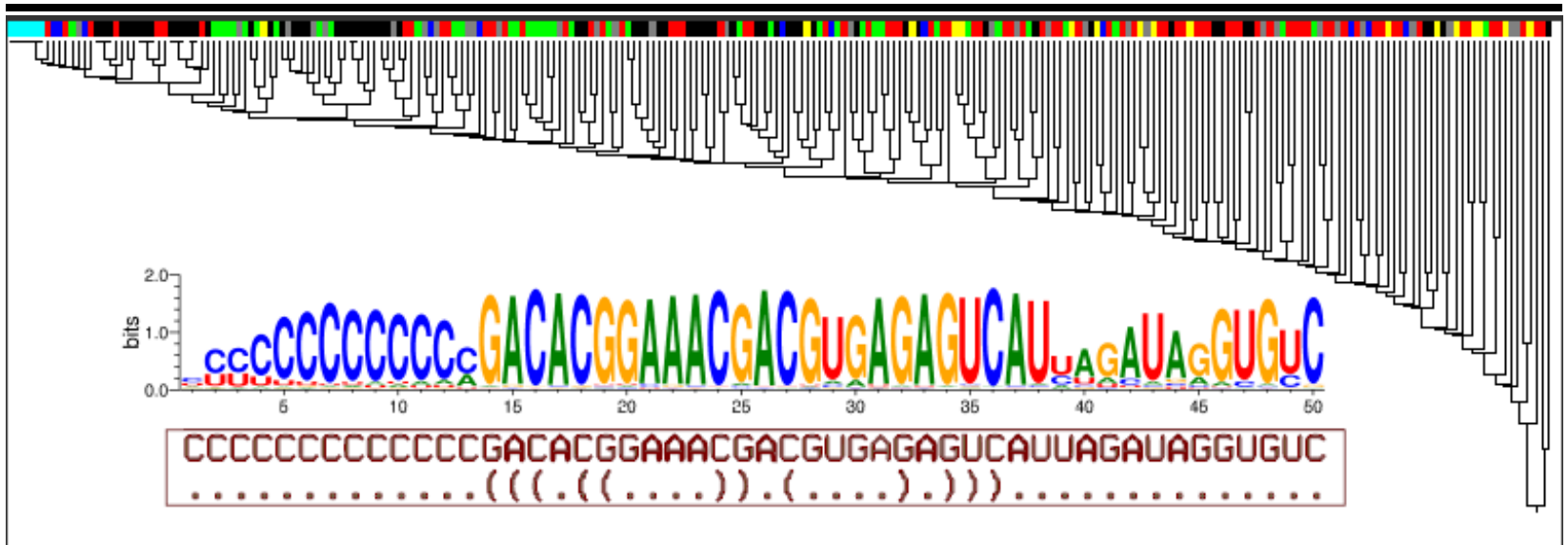

mutational neighborhood of master seq.: STEEP

EVOLVED   optimal repl   av. random

black replicator; blue rest

High mut. rate: 1 quasispesies LOW variability

mutational NB: STEEP and "special"

EVOLVED   optimal repl   av. random

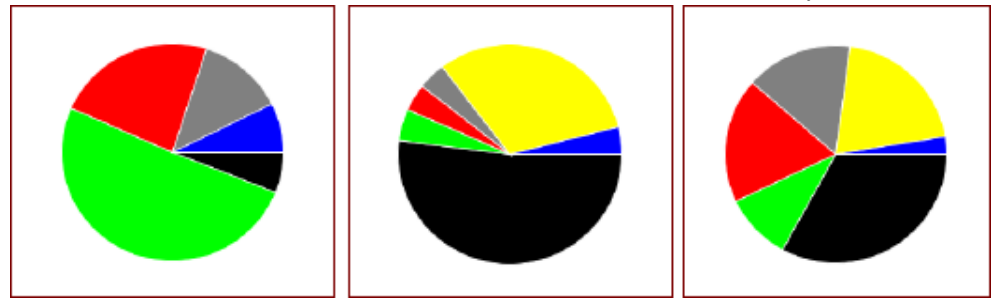black replicator; blue rest; yellow parasites

1 quasispesies: codes for multiple functions
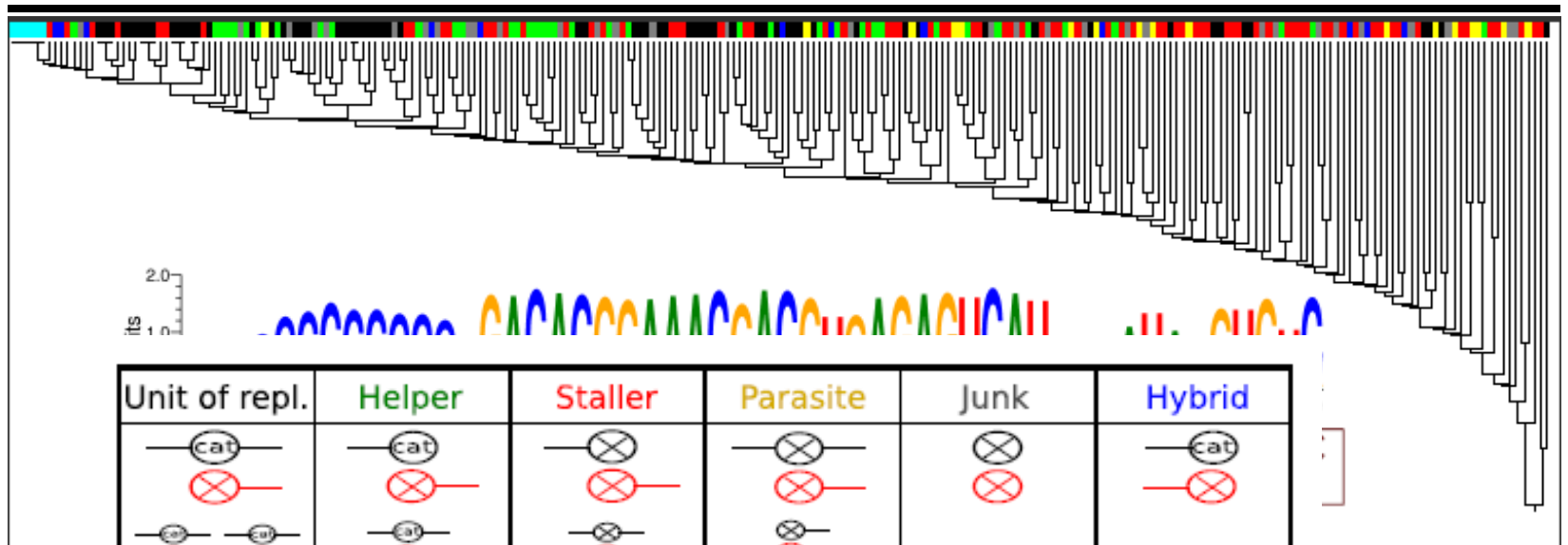
Unit of repl.	Helper	Staller	Parasite	Junk	Hybrid
VIABLE	NOT VIABLE	NOT VIABLE	VIABLE	NOT VIABLE	NOT VIABLE

mutational NB: STEEP and "special"

EVOLVED   optimal repl   av. random
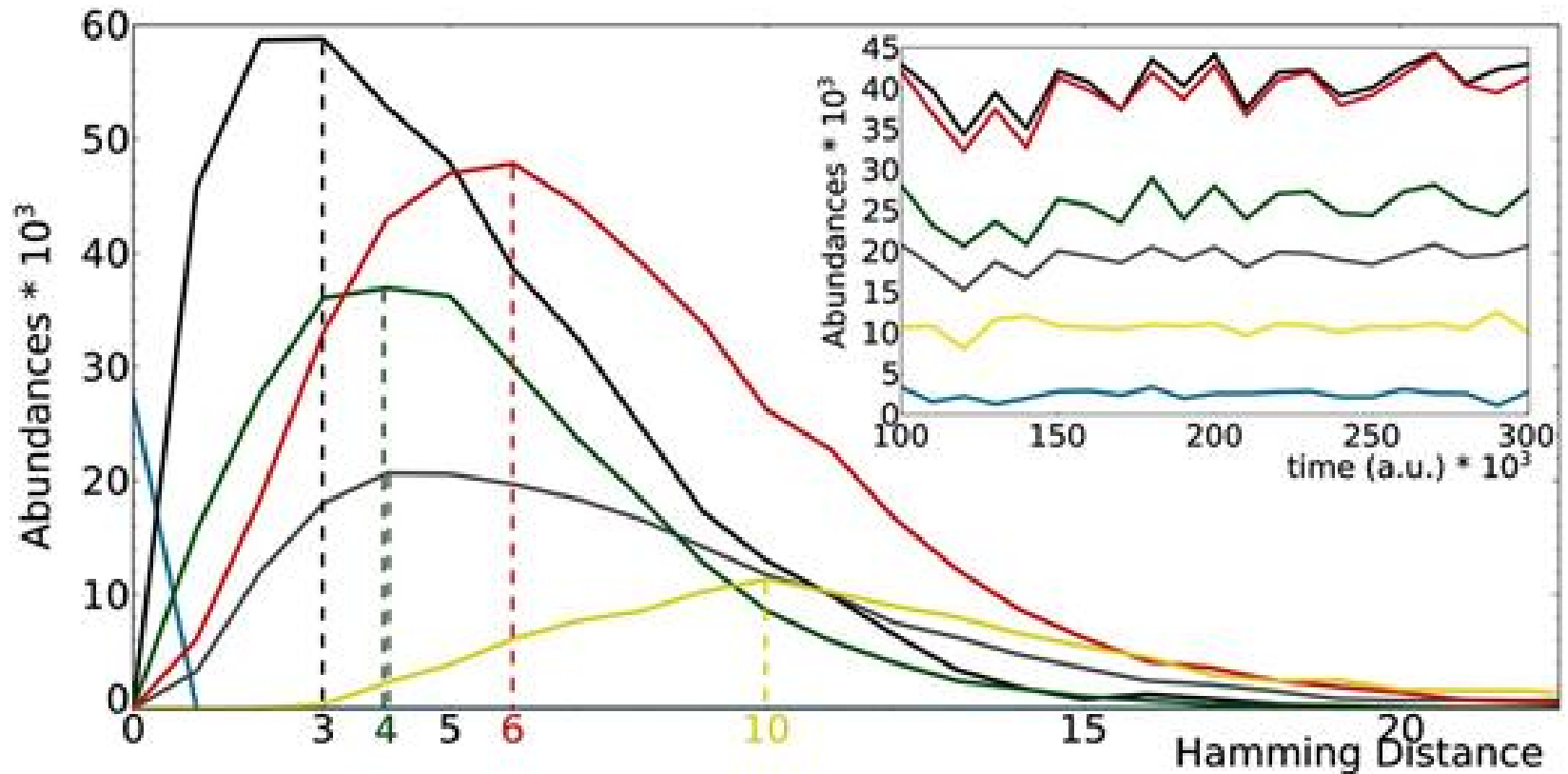
black repl.; blue rest; yellow parasites; green helpers; red stallers; gray junk
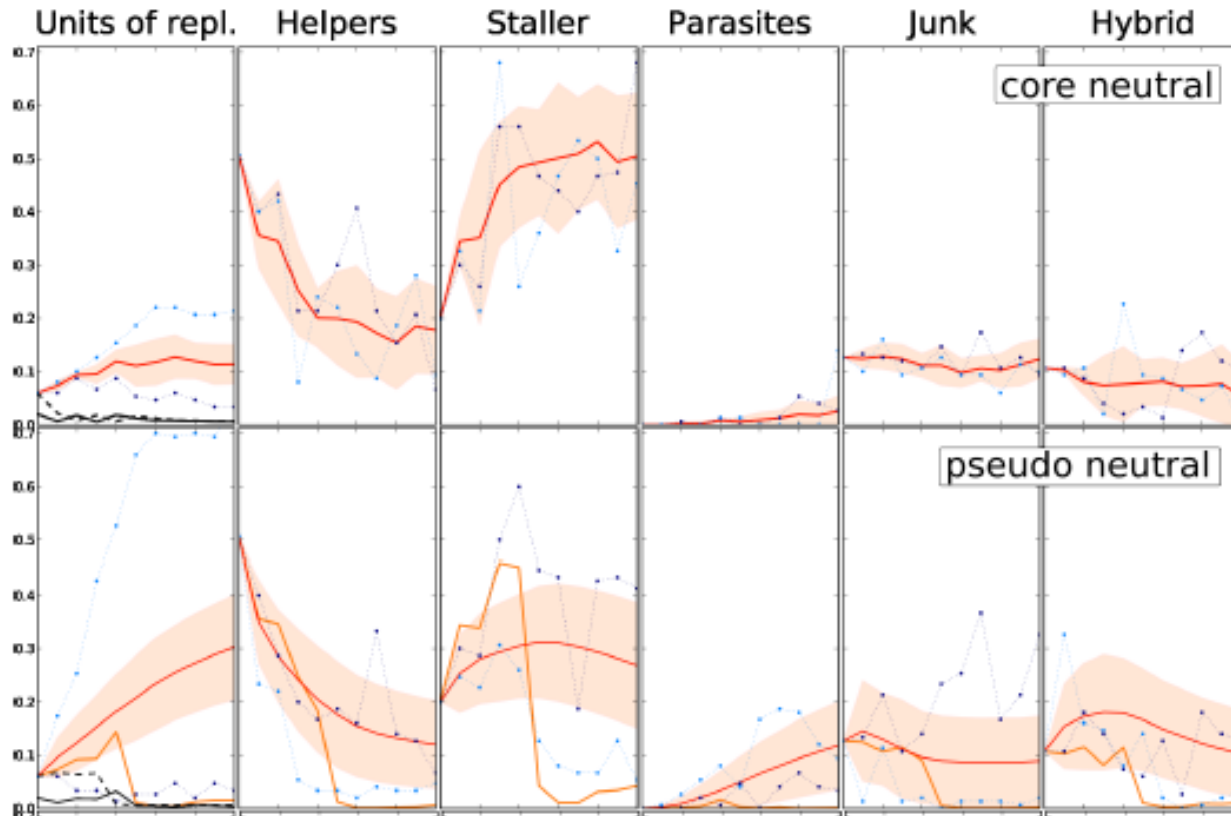
Top follow replicases with $>=$ replic rates masterseq. bottom follow replicases with $<$ replic rates masterseq.

# Abundance of functional types
at Hamming distance to master sequence

quasispecies composition in field

weakly reflects mutational neighbourhood
more replicators (becuse of replication)
, less helpers , more stallers
(like neigborhood of other replicases)
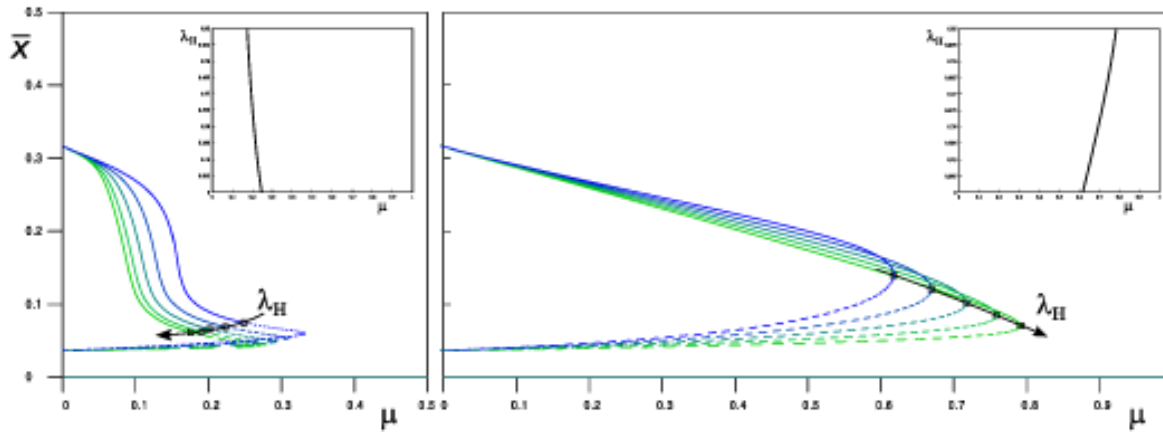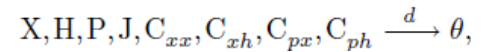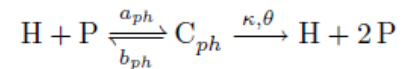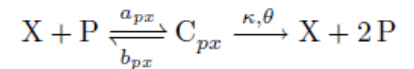
Replicases with 'good' MN
overrepresented.

size = frequency; yellow core replicase

Master Seq.   Average random   Average field
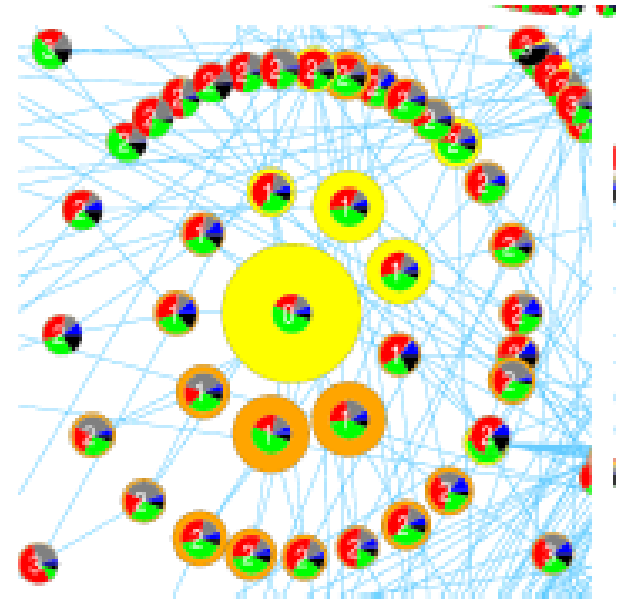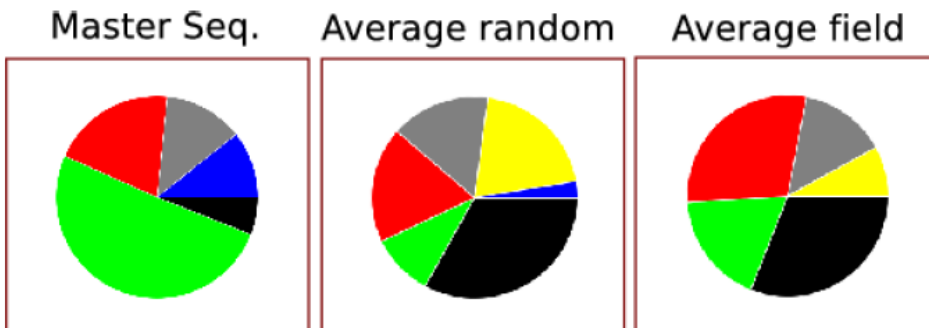
Helpers "help"

change in junk $->$ extinction
change in empty $->$ extinction

in simplified ODE model:
increases max μ without parasites
decreases max μ with parasites

.

$$2\,X \underset{b_{xx}}{\overset{a_{xx}}{\rightleftharpoons}} C_{xx} \begin{cases} \xrightarrow{\kappa,\theta,(1-\mu)} 3\,X \\ \xrightarrow{\kappa,\theta,\mu\cdot\lambda_H} 2\,X+H \\ \xrightarrow{\kappa,\theta,\mu\cdot\lambda_P} 2\,X+P \\ \xrightarrow{\kappa,\theta,\mu\cdot\lambda_J} 2\,X+J \end{cases}$$

$$X+H \underset{b_{xh}}{\overset{a_{xh}}{\rightleftharpoons}} C_{xh} \begin{cases} \xrightarrow{\kappa,\theta,(1-\mu)} 2\,X+H \\ \xrightarrow{\kappa,\theta,\mu\cdot\lambda_H} X+2\,H \\ \xrightarrow{\kappa,\theta,\mu\cdot\lambda_P} X+H+P \\ \xrightarrow{\kappa,\theta,\mu\cdot\lambda_J} X+H+J \end{cases}$$

$$X+P \underset{b_{px}}{\overset{a_{px}}{\rightleftharpoons}} C_{px} \xrightarrow{\kappa,\theta} X+2\,P$$

$$H+P \underset{b_{ph}}{\overset{a_{ph}}{\rightleftharpoons}} C_{ph} \xrightarrow{\kappa,\theta} H+2\,P$$

$$X,H,P,J,C_{xx},C_{xh},C_{px},C_{ph} \xrightarrow{d} \theta,$$
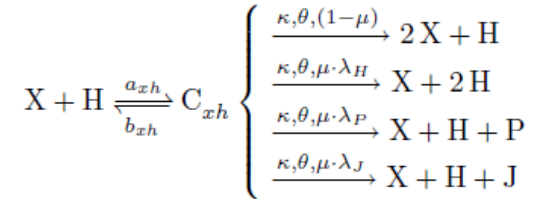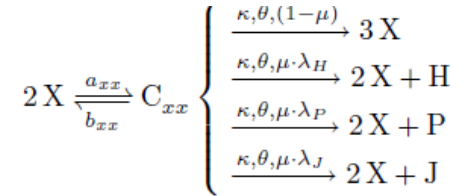
parasites   without parasites

Stallers "stall"

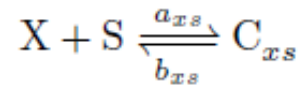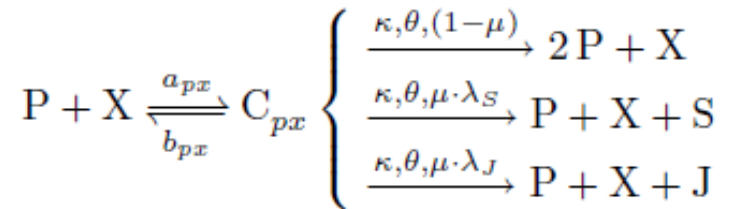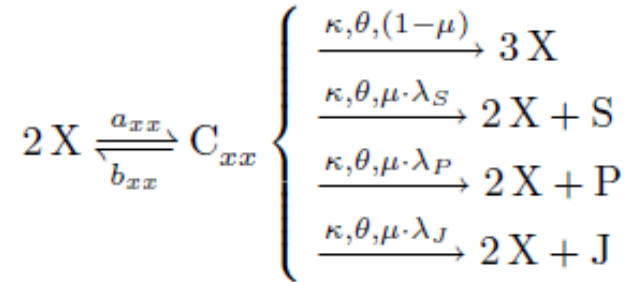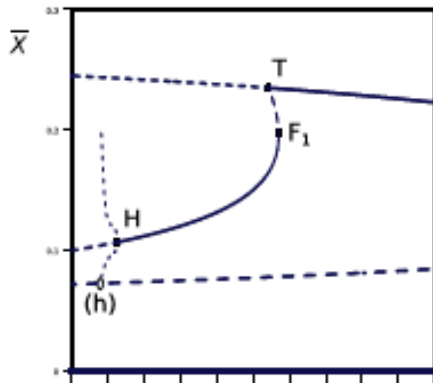change in junk $->$ increases density

BUT master seq. replaced

'pseudo stallers' evolve

change into empty space

parasite lineage evolves!

in simplified ODE model:

protects against parasites

(h)

$$2X \underset{b_{xx}}{\overset{a_{xx}}{\rightleftharpoons}} C_{xx} \begin{cases} \xrightarrow{\kappa,\theta,(1-\mu)} 3X \\ \xrightarrow{\kappa,\theta,\mu\cdot\lambda_S} 2X+S \\ \xrightarrow{\kappa,\theta,\mu\cdot\lambda_P} 2X+P \\ \xrightarrow{\kappa,\theta,\mu\cdot\lambda_J} 2X+J \end{cases}$$

$$P+X \underset{b_{px}}{\overset{a_{px}}{\rightleftharpoons}} C_{px} \begin{cases} \xrightarrow{\kappa,\theta,(1-\mu)} 2P+X \\ \xrightarrow{\kappa,\theta,\mu\cdot\lambda_S} P+X+S \\ \xrightarrow{\kappa,\theta,\mu\cdot\lambda_J} P+X+J \end{cases}$$

$$X+S \underset{b_{xs}}{\overset{a_{xs}}{\rightleftharpoons}} C_{xs}$$

$$P+S \underset{b_{ps}}{\overset{a_{ps}}{\rightleftharpoons}} C_{ps}$$

$$X,S,P,J,C_{xx},C_{px},C_{xs},C_{ps} \xrightarrow{d} \theta,$$
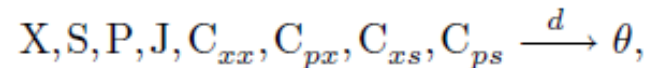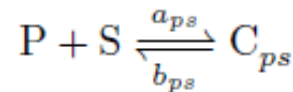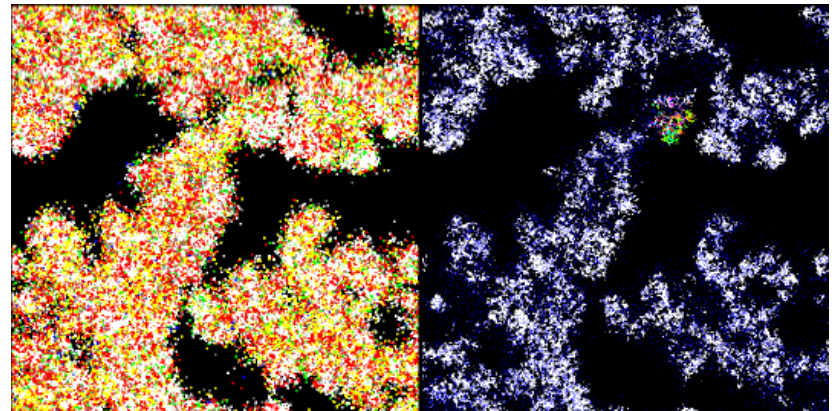
with parasites: x-axis: fraction staller-mutants

Variability of evolved quasispacies

|	mut. nei.	μ_{max}	replic. rates +/+	+/-	competition 7	8
			Steep quasispecies			
1.		0.0164	0.902 0.914	0.858 0.831	✓	✓
2.		0.0154	1.000 0.932	0.878 0.854	✓	✓
3.		0.0151	1.000 0.870	0.777 0.744	✓	✓
4.		0.0145	1.000 0.866	0.817 0.777	✓	✓
5.		0.0151	1.000 0.818	0.777 0.731	✓	✓
6.		0.0143	1.000 0.858	0.777 0.729	x	x
			Flat quasispecies			
7.		0.0154	0.725 0.892	0.817 0.798		
8.		0.0149	0.902 0.872	0.817 0.792		
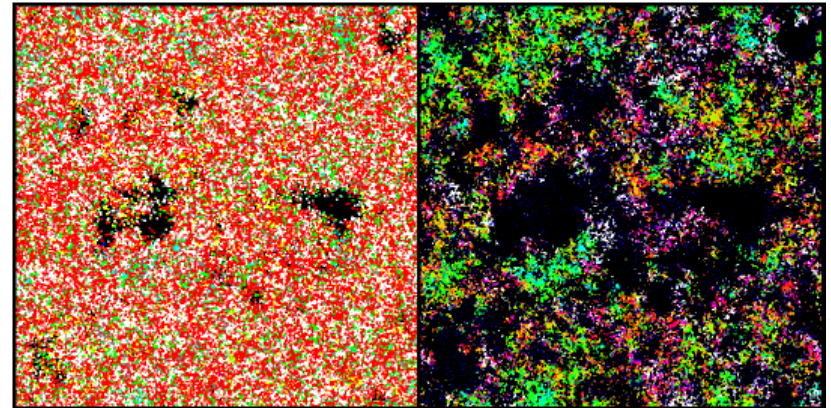
Steep quasispecies

Flat quasispecies

random

colors ' majority function' - dist. from masterseq

cyan-green-yellow-red-magenta-blue

(c) $\mu = 0.008$

A-catalyst:

C-catalyst:
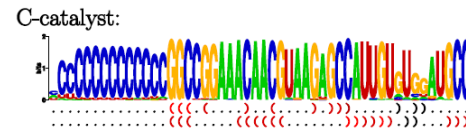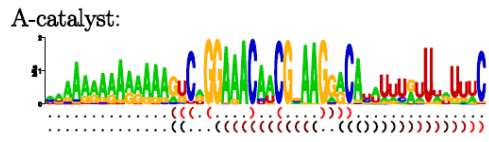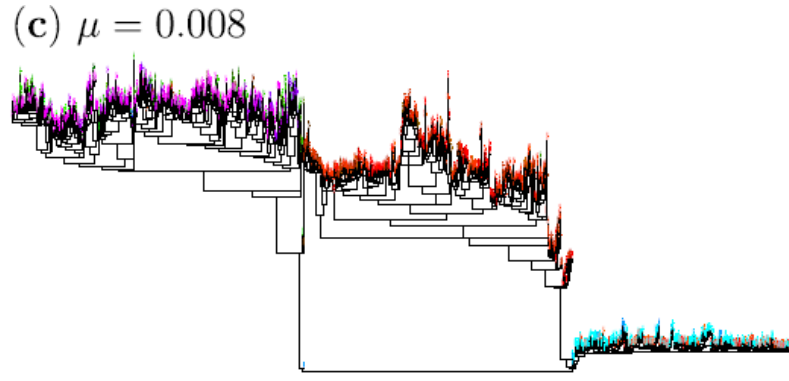
optimal replicator
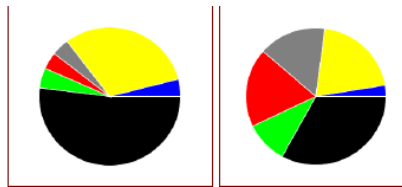
random replicator

-s,

black replicator; yellow parasite; green helper; red staller

Takeuchi & Hogeweg 2008, Colizzi & Hogeweg 2014

Conclusions RNA world at high mutation rates

Evolution of very specific coding structure.

One mastersequence codes for functional diverse ecosystem

Decoded by mutations (hence clearest at high mutation rates)
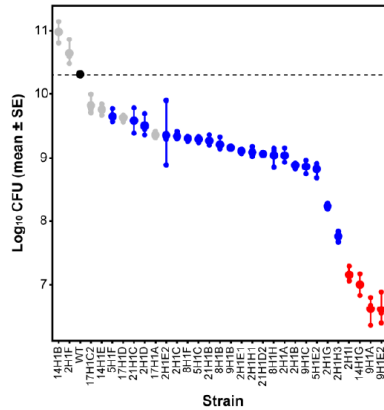
In steep quasispecies most pronounced (best 'control')

individually coded but ecosystem based diversity
evolves and persists close to the Information Threshold

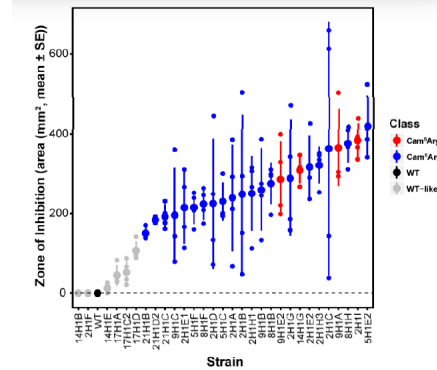# Quasispecies based division of labour:
Antibiotic production is organized by division of labour in Streptocyces
Zheren Zhang...Daniel E Rozen,Science advances, 6(3) 2020.
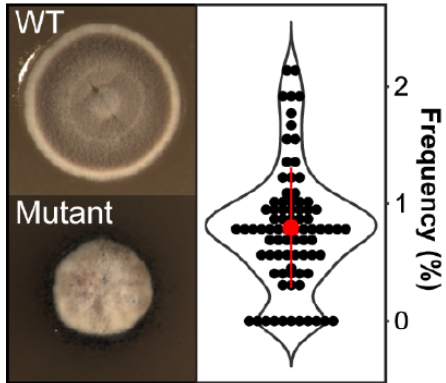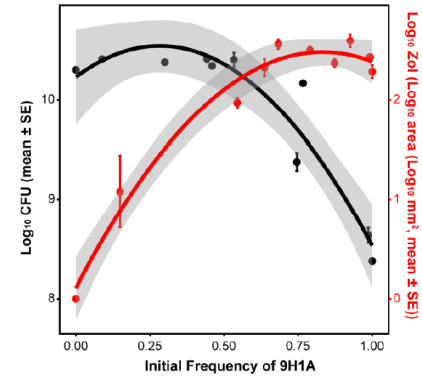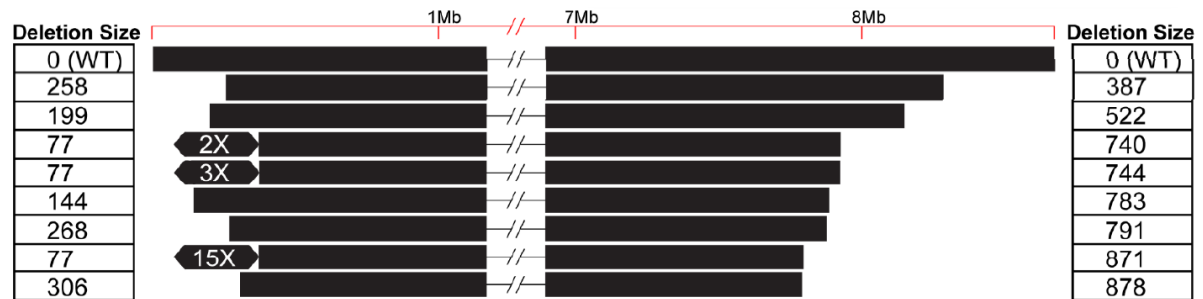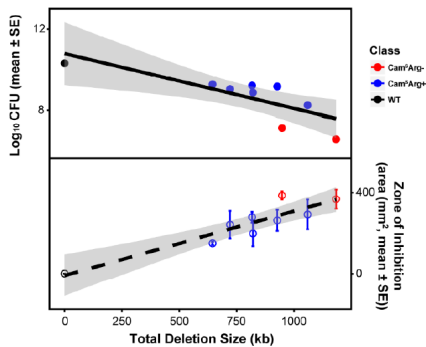

High Mut. Rate

Mut. fitness

antibiotics production

QS fitness
only > 50

Genome structure and targeted mutations