

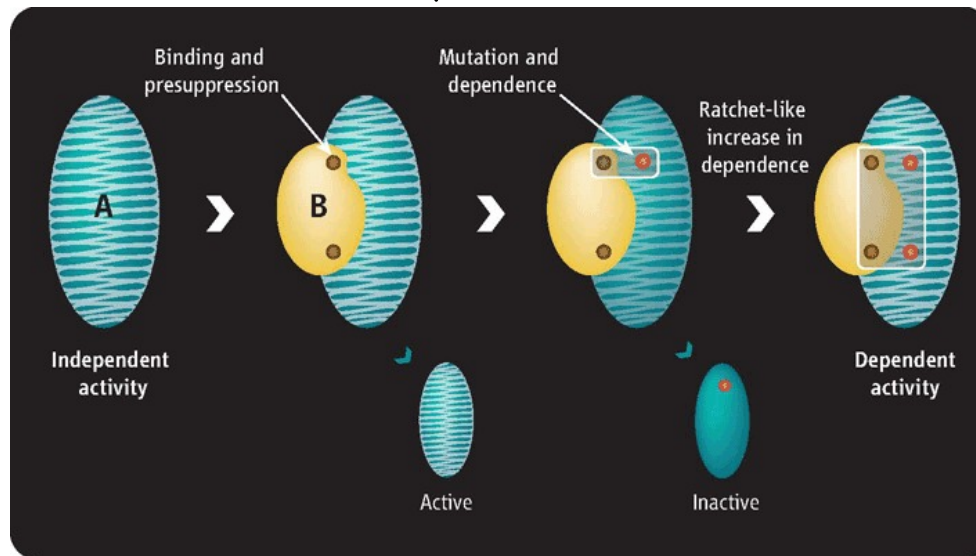
RNA world: sequence - structure- interactions

Course computational biology 2018/2019; Paulien Hogeweg;
Theoretical Biology and Bioinformatics Grp Utrecht University

Neutrality and evolution of complexity

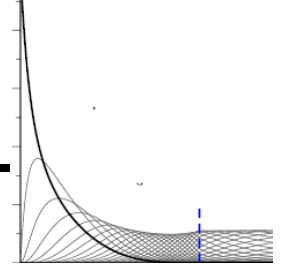
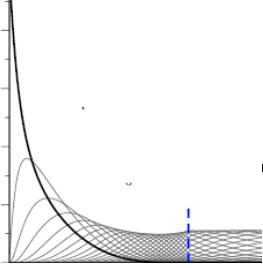
neutral ratchet/constructive neutral evolution/irremediable complexity

e.g. neutral binding / increase neutrality /
accumulation of mutations / indispensibility of binding

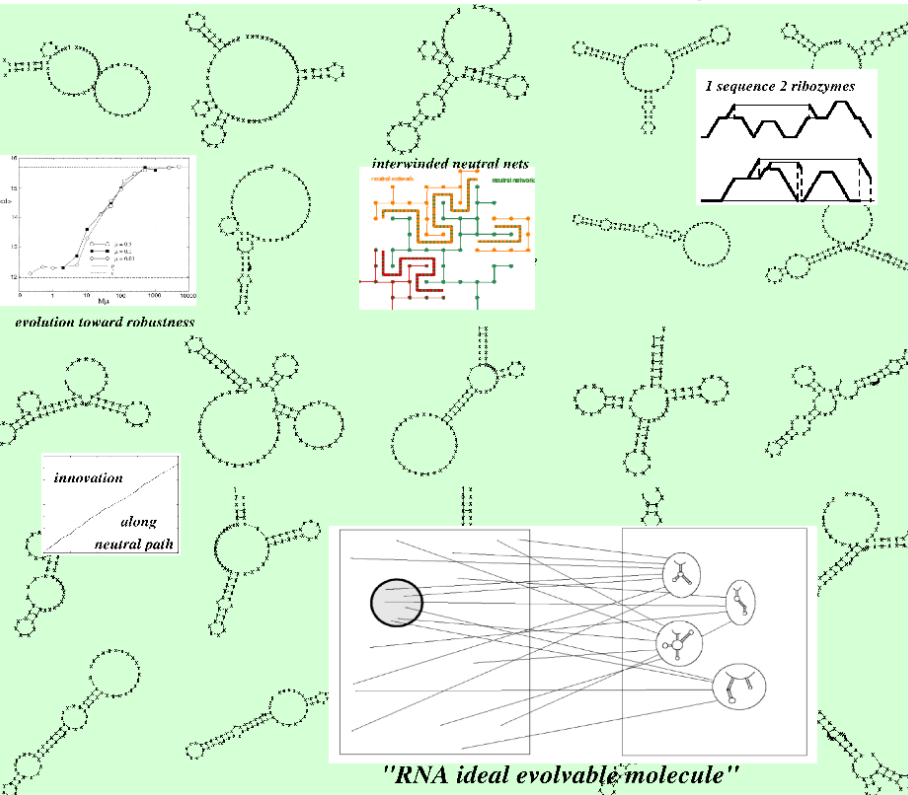


cf Covello & Gray 1993, Stoltzfus 1999, Lynch 2007, Gray et al 2010

2 images of RNA world

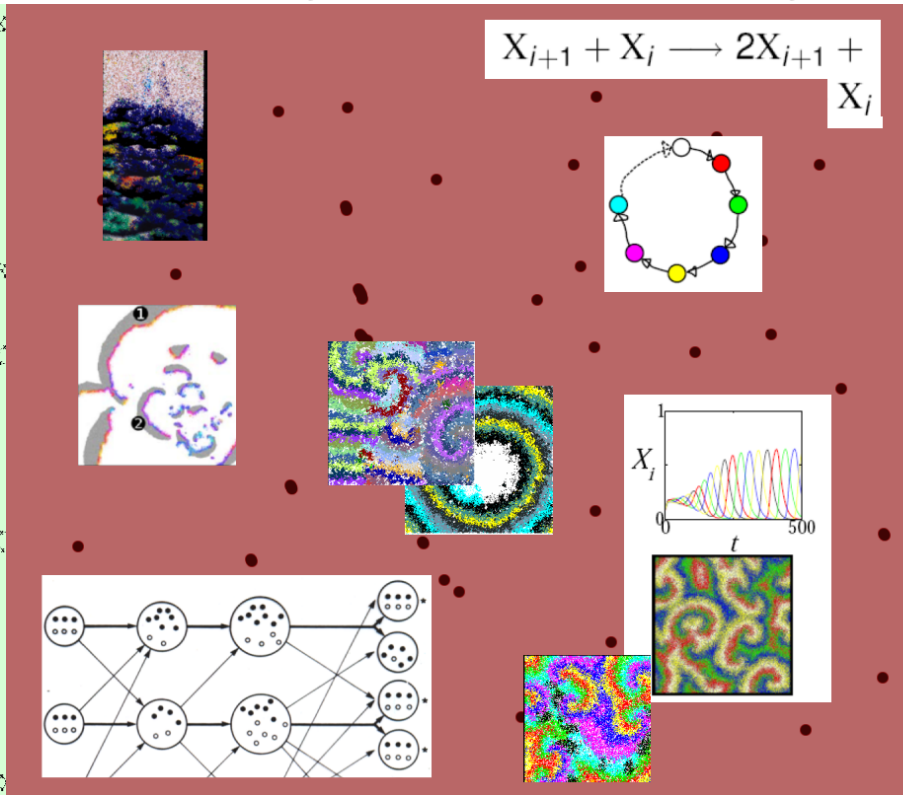


individual complexity



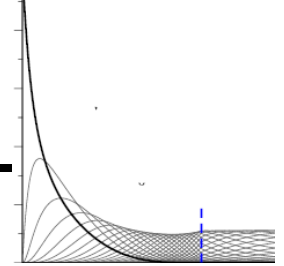
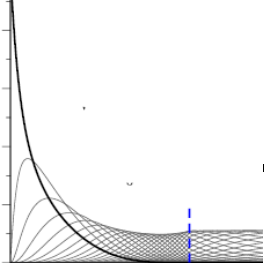
sequence to structure

ecosystem complexity

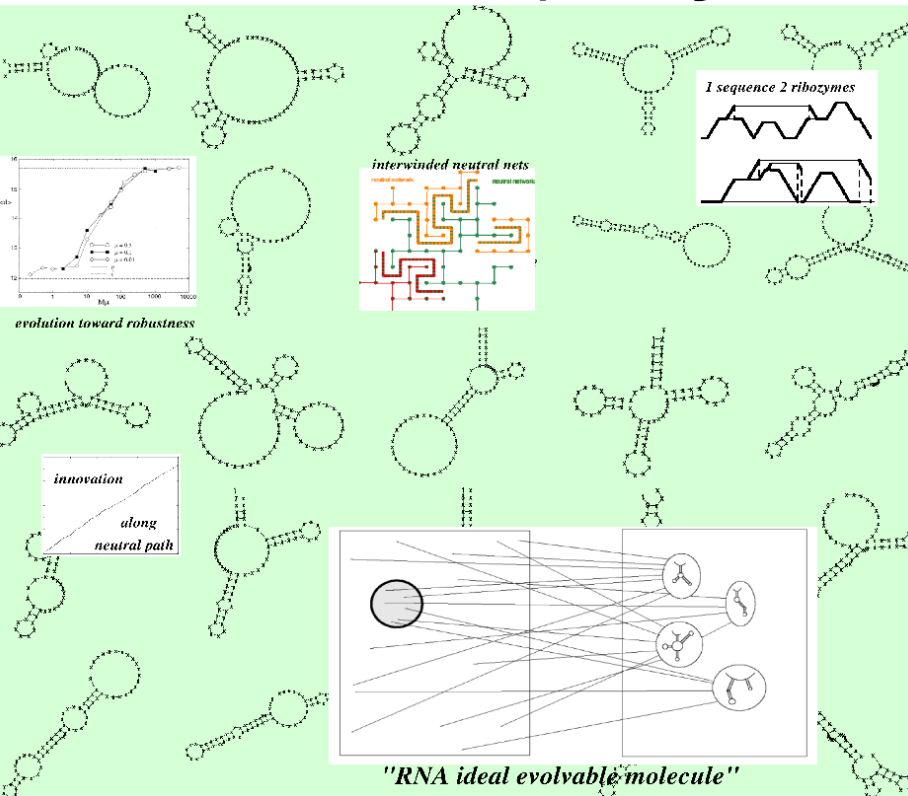


replicator to wave/vesicle

the RNA world

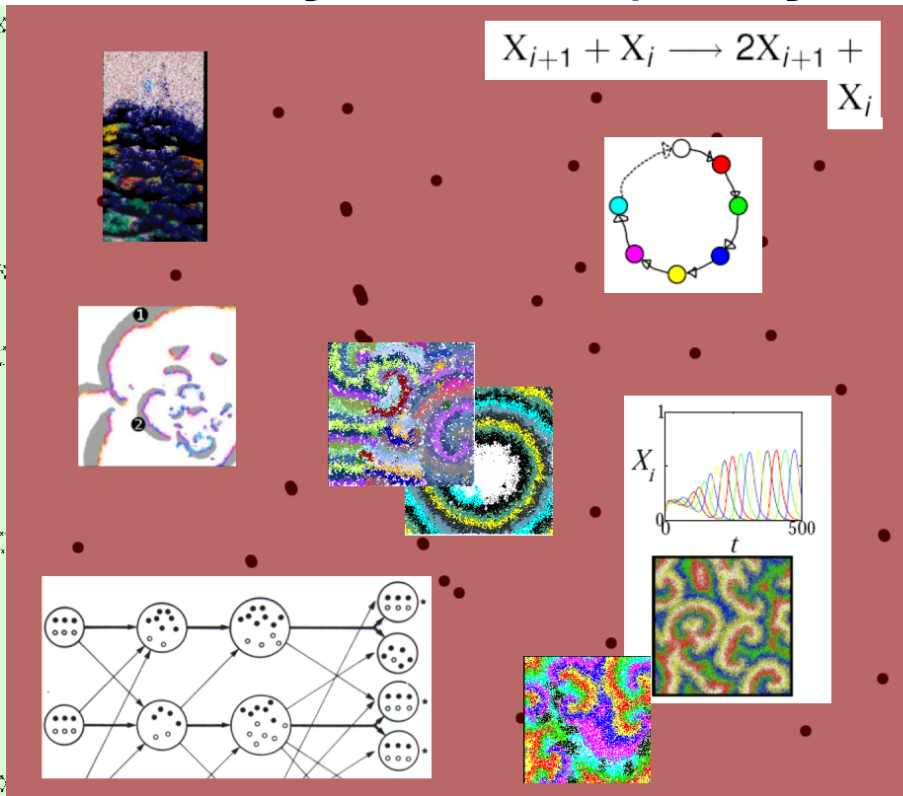


individual complexity



RNA (without world)

ecosystem complexity



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)

multiple coding

mutational neighborhood

RNA even more evolvable than seen so far

flexible coding AND multilevel selection

individual vs/and ecosystem based complexity

RNA replicases in space

(replicase function requires a structure (shapiro))

–mutation rate and shape of quasispecies

–mutation rate and speciation

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**

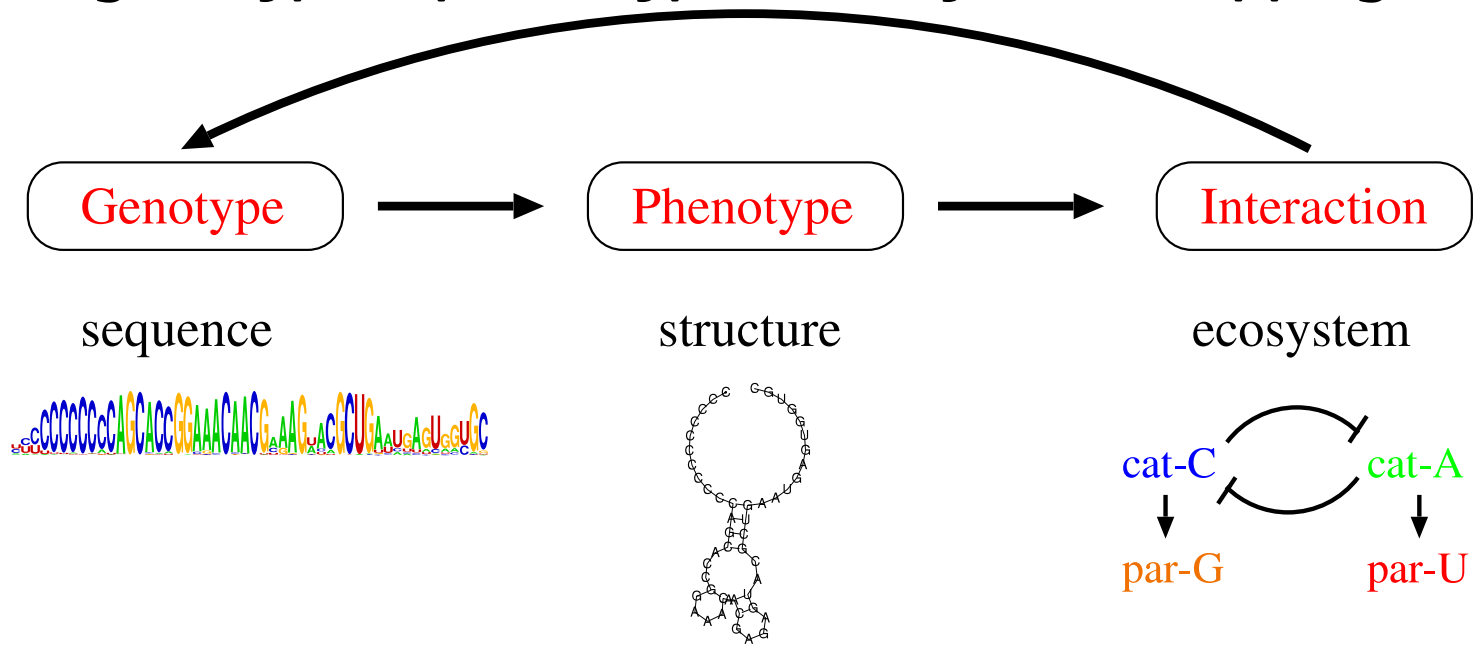
- **no predefined interactions**

but predefined reactions

DO SPECIES/ INTERACTION NETWORKS EVOLVE?

DOES EVOSYSTEM COMPLEXITY EVOLVE?

genotype - phenotype - ecosystem mapping

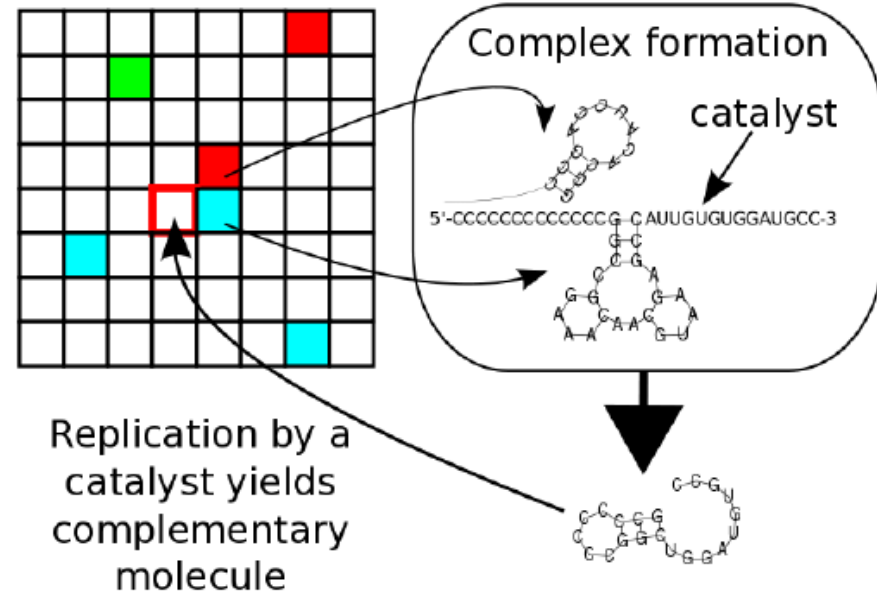


feedback from higher levels to lower levels in evolving system

interacting RNA's

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

(*“strong” altruism*)



1. $X + Y \xrightleftharpoons[k_1]{k_2} C_{X \sim Y} \text{ or } \xrightleftharpoons[k_2]{k_1} C_{Y \sim X}$
2. $C_{X \sim Y} + \Phi \xrightarrow{\kappa} X + Y + Y^{-1}$
3. $X \xrightarrow{d} \Phi$

only structure + reaction

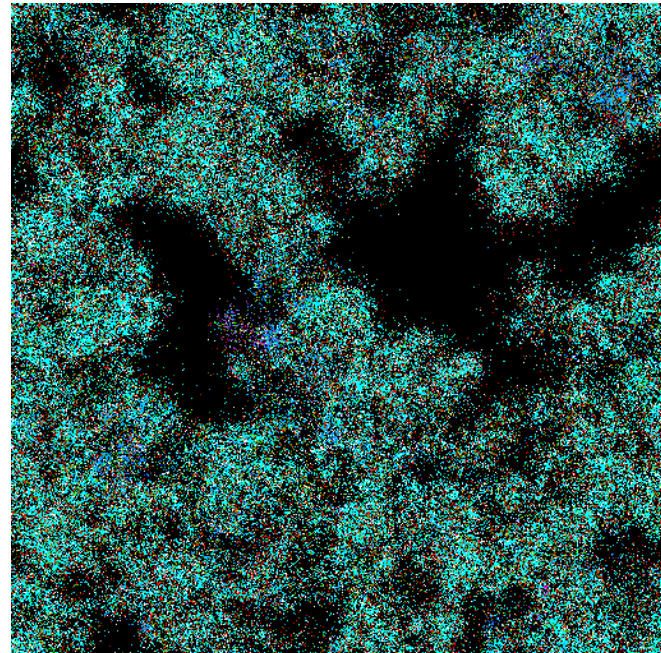
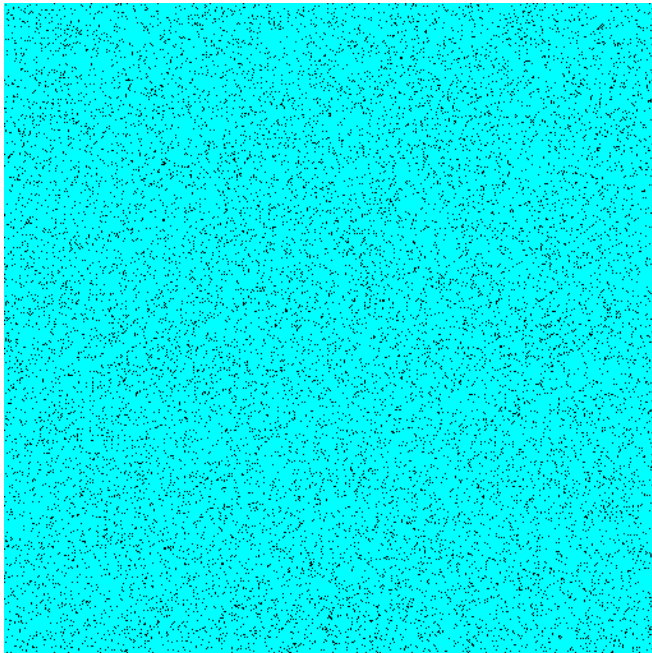
no fitness function and no interaction predefined

close to information threshold($\mu = .015$)

initial population

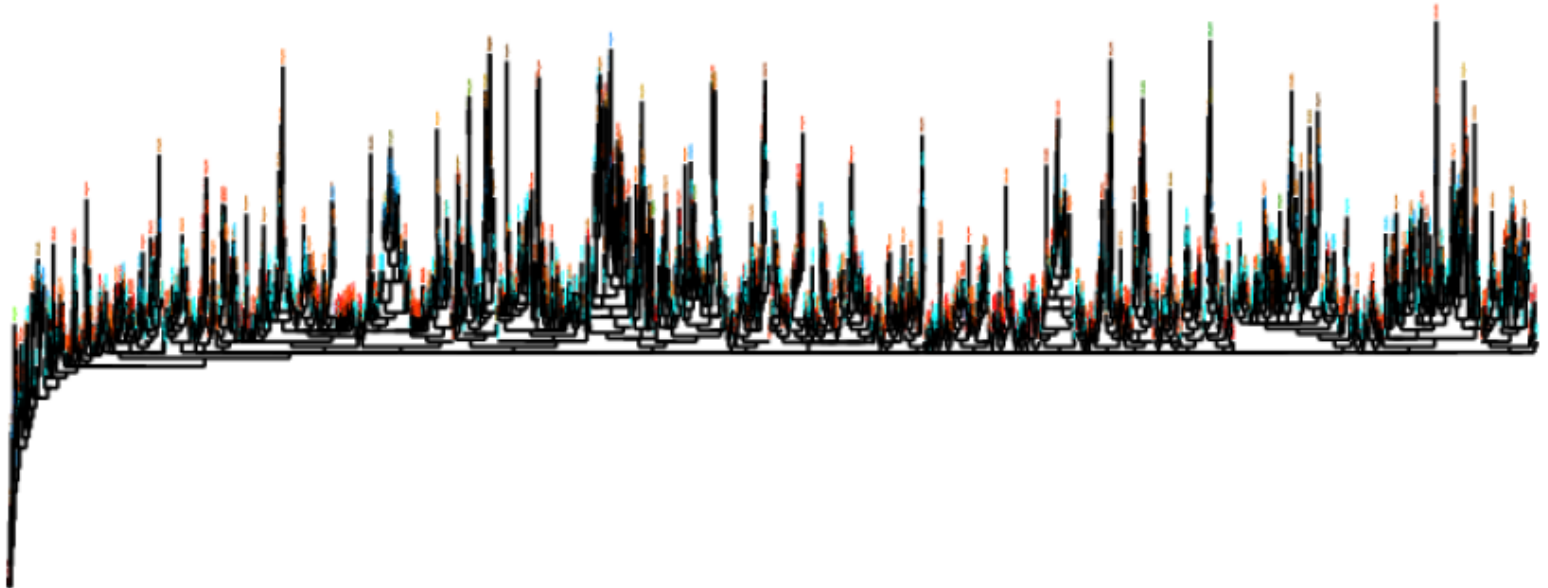
dynamics with mutation

= after stopping mutation



close to information threshold ($\mu = .015$)
population structure

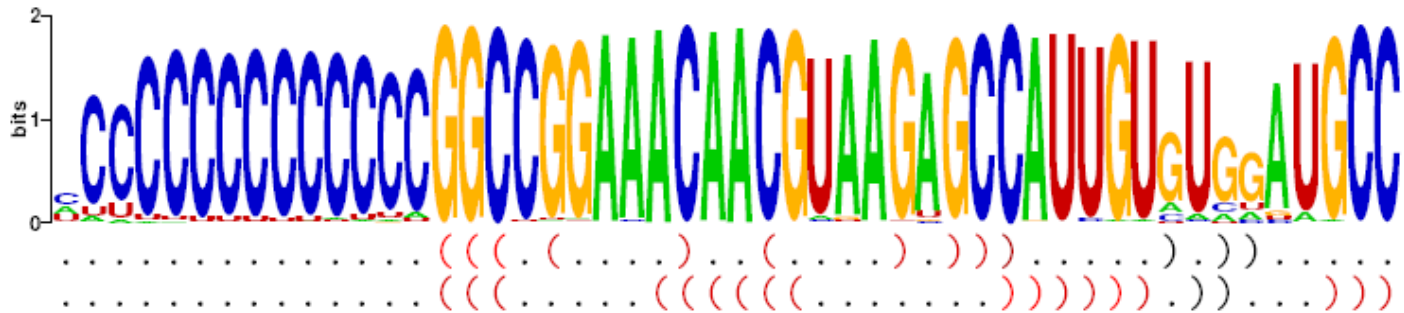
- Phylogeny reveals patterns in population of genotypes



- No clade patterns
- Population is supported by various genotypes
- One quasi-species

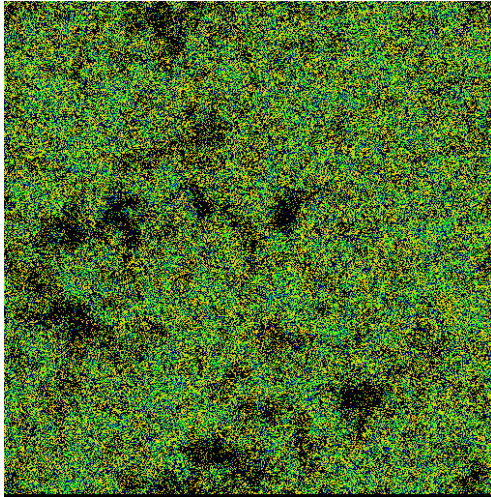
Color	Types
Cyan	Catalyst
Red	Non-catalyst

close to information threshold ($\mu = .015$)
sequence structure

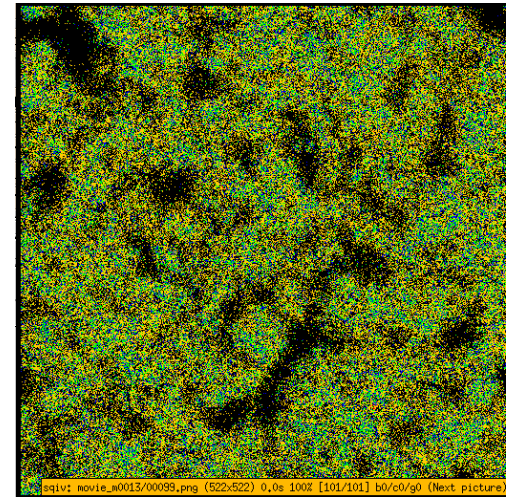


- Very high C frequency in 5'-end
- High G frequency in 3'-end
 - → many GC pairs
- Many interspersing U in 3'-end
 - → prevents base-pair formation *in homo*
- No 5'-end in template strand
 - → prevents non-functional complex formation

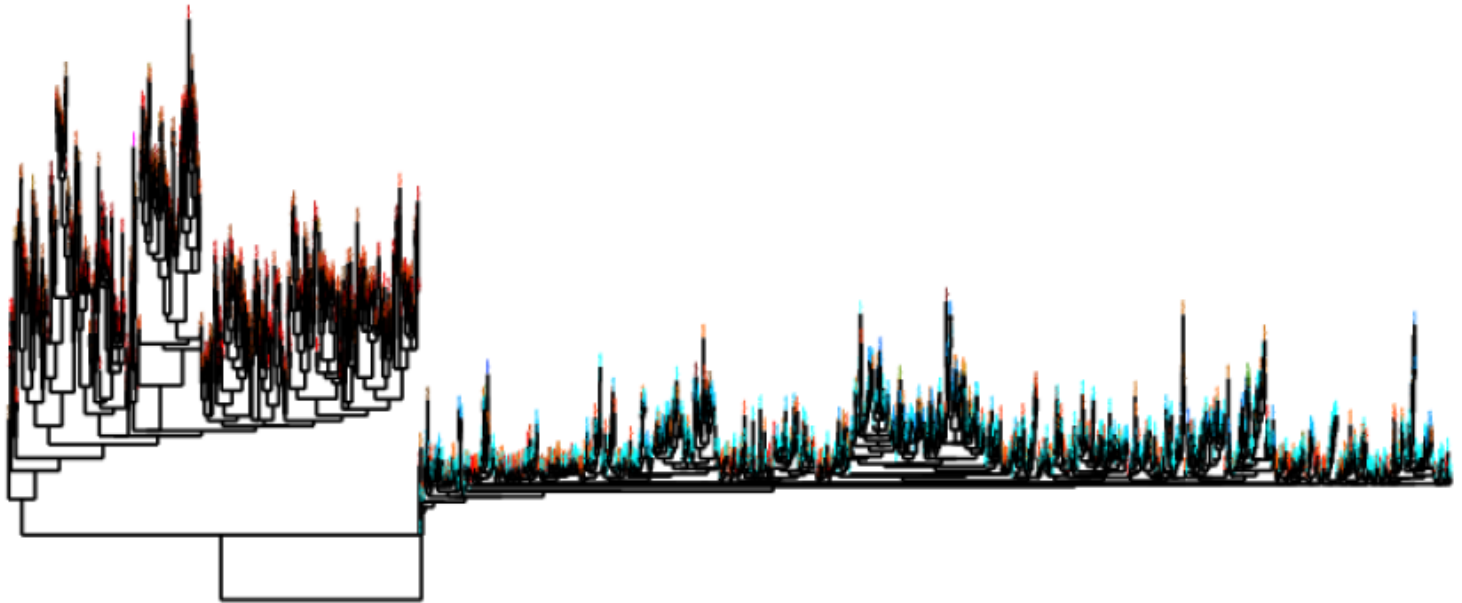
lowering mutation rates ($\mu = .13$) : SPECIATION



$m=.014 - > m=.013$



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

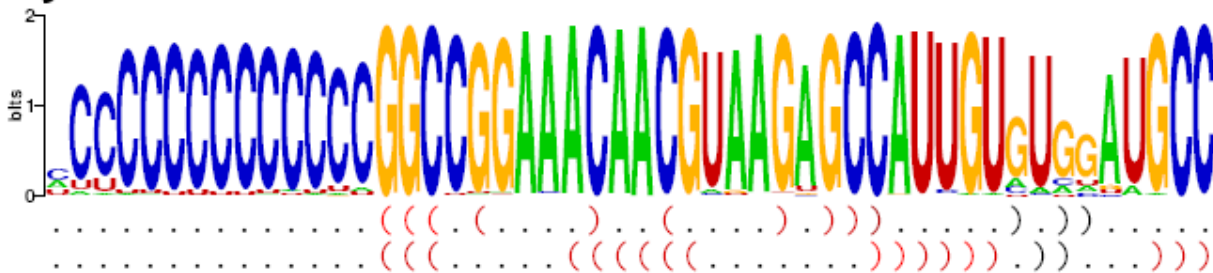


Color	Types
Cyan	Catalyst
Red	Non-catalyst

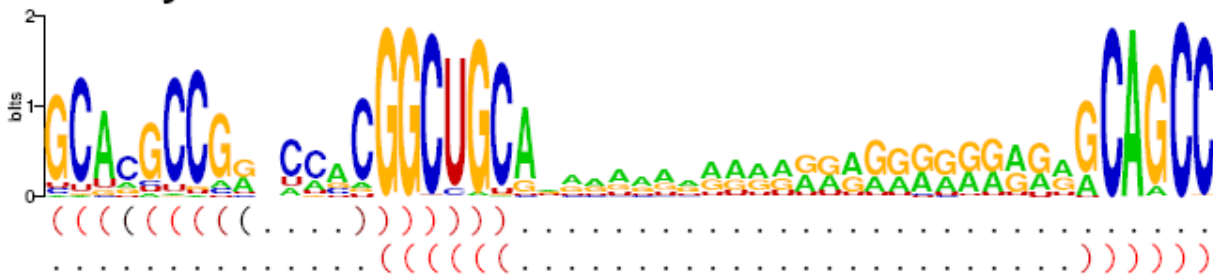
- Two quasi-species
 - distinct sequence classes
- Catalyst & Non-catalyst

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

- Catalyst



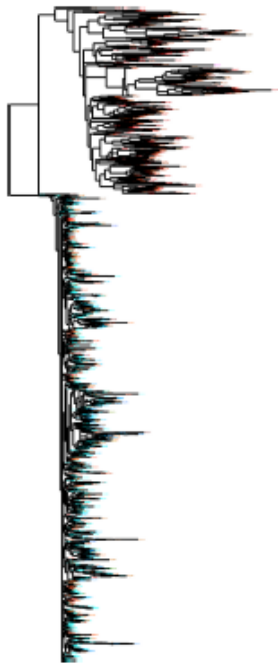
- Non-catalyst → Parasite



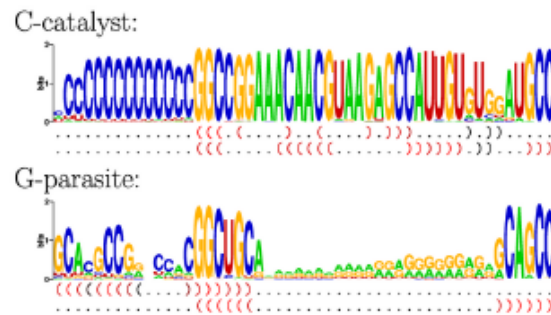
- No catalytic structure in both strands
- Long 3'-end with many G
- No 5'-end in both strands
- Sequence conservation patterns & secondary structure

Parasite invades in periphery of QS

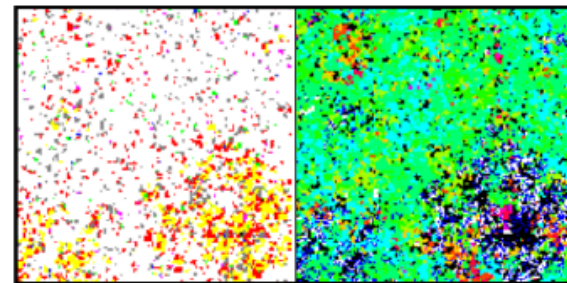
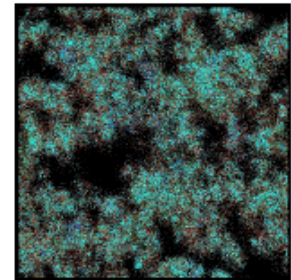
■ Population of Sequences



■ Genotype & Phenotype



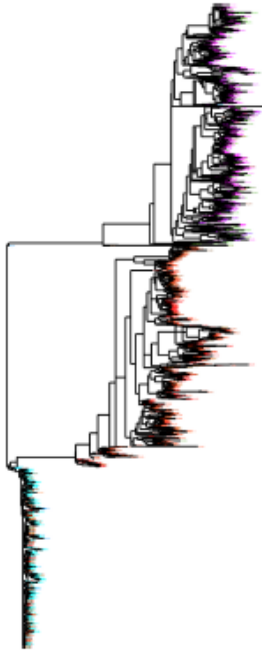
■ Space & Time



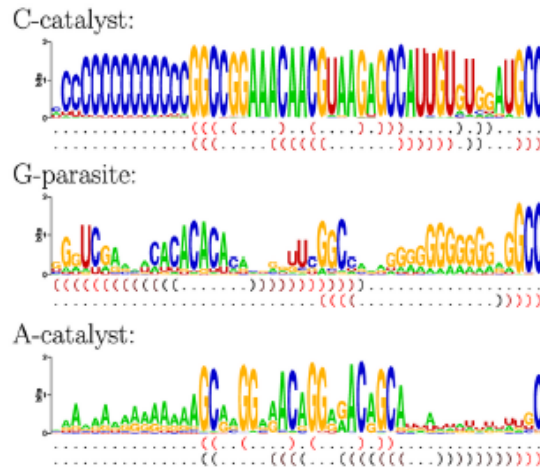
Hamm.
dist. from
master
sequence

Lower mutation rate $\mu = .008$: 3 quasispecies

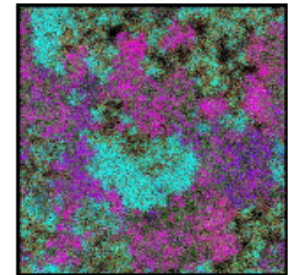
■ Population of Sequences



■ Genotype & Phenotype



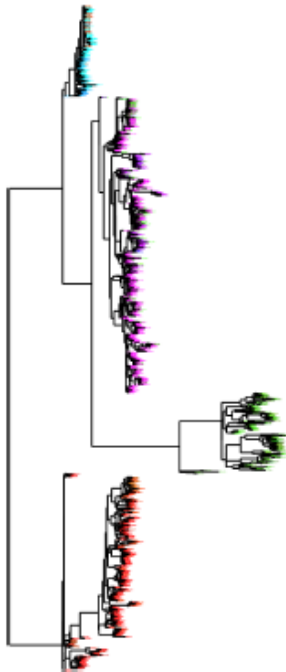
■ Space & Time



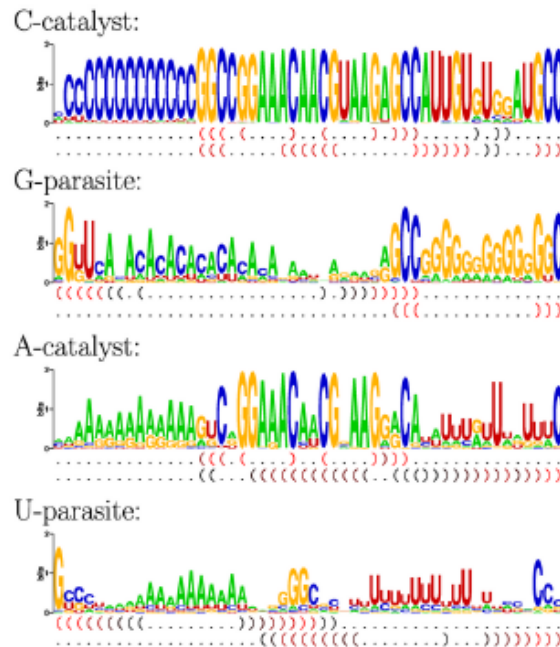
A catalyst: HIGH neutrality (ca 50%)

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

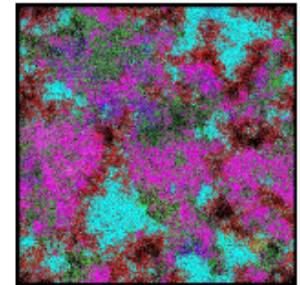
Population of Sequences



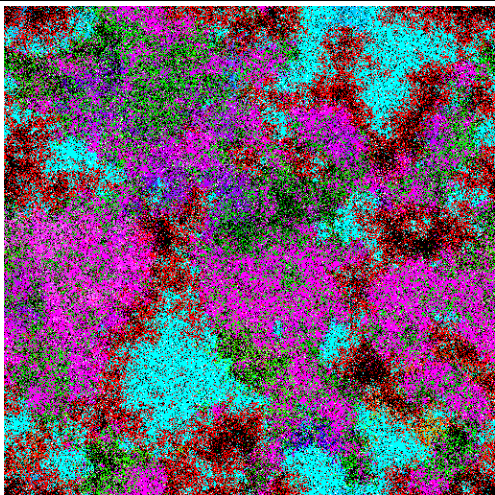
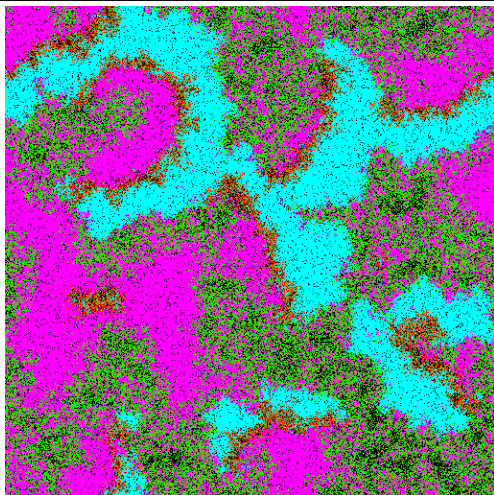
Genotype & Phenotype



Space & Time



evolved 4 species system; evolved interaction topology



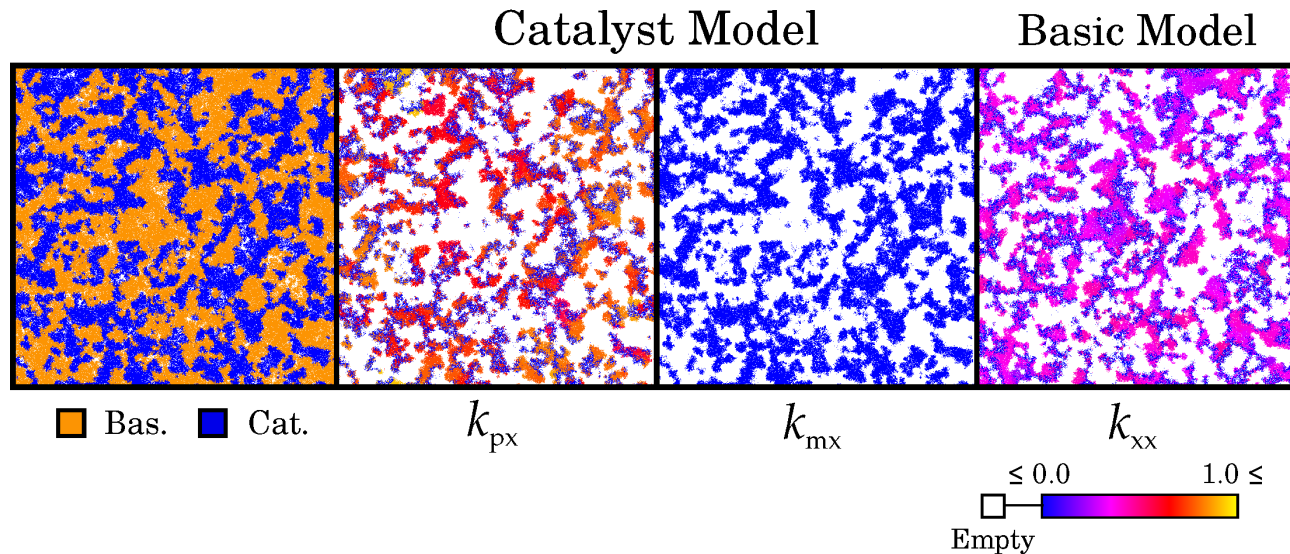
ECOsystem ($\mu = 0$)

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

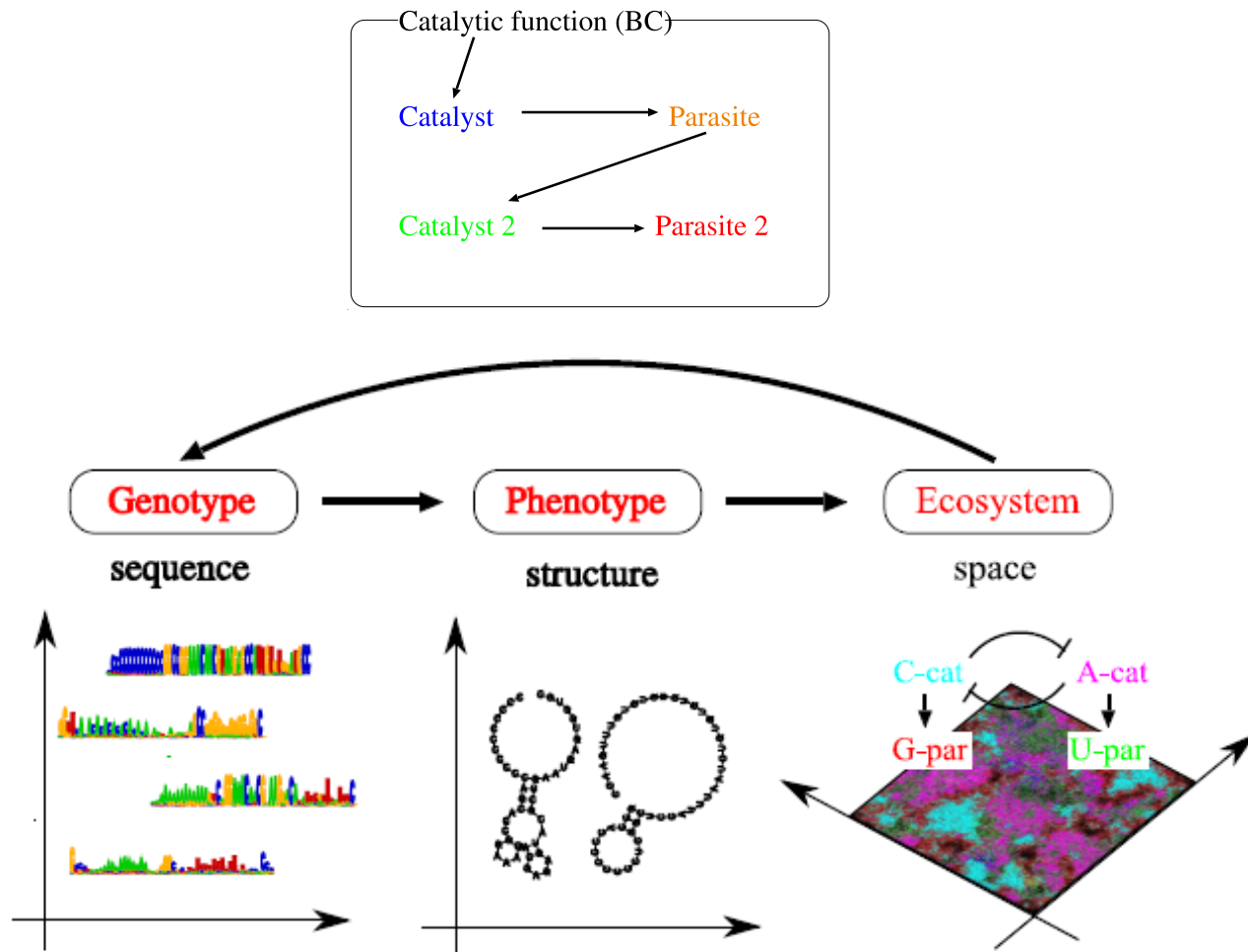
Similar ecology as in minimal RP system with complementary replication with 2 separate species (von den Dunk et al 1017)



Mini model The 2 predifines species do not catalyse each other. They both differentiate in catalysts and parasites. Only + strand catalytic in complementary replicating species

RNA model One species differentiates into 4 species: 2 with catalytic + strands and non-catalytic - strands, and 2 parasites (non catalytic in both strands) with some cross catalysis maintained (unavoidable?)

From Coding structure to ecosystem based information accumulation



Conclusion

Very stable multi-(quasi)species systems evolves

Interaction topology different from anything studied before.

Variability increases with decreasing mutation rate
speciation

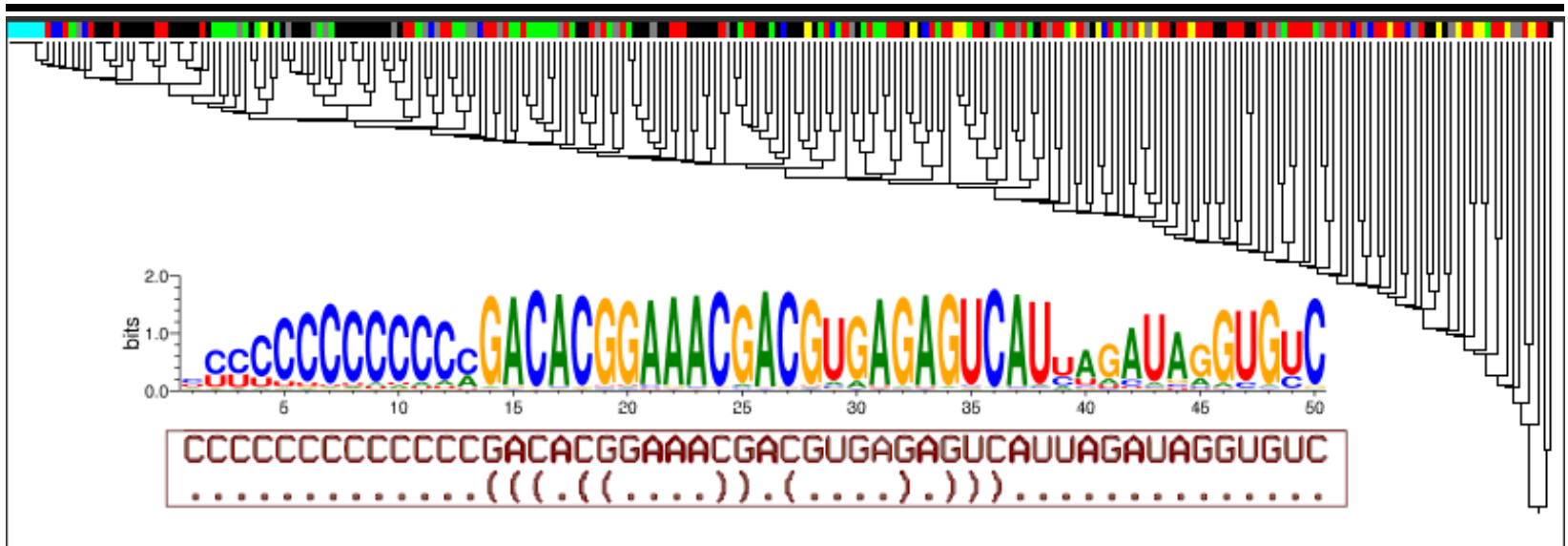
Ecosystem based “solution” only at lower mutation rates

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

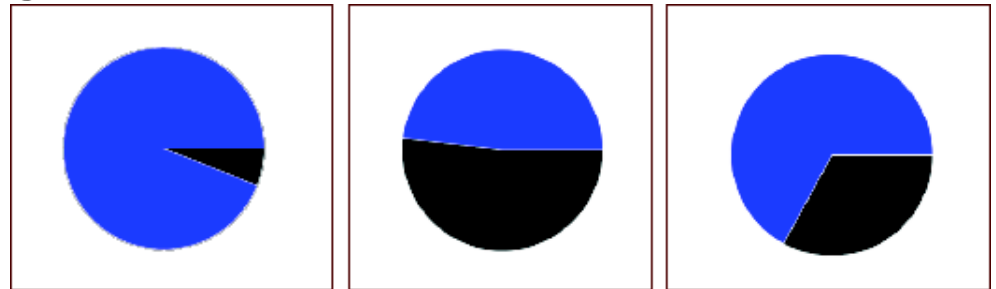
Evolved, niche dependent mutational landscape

Evolution of coding structure at high mutation rates

High mut. rate: 1 quasispecies LOW variability



mutational neighborhood of master seq.: STEEP

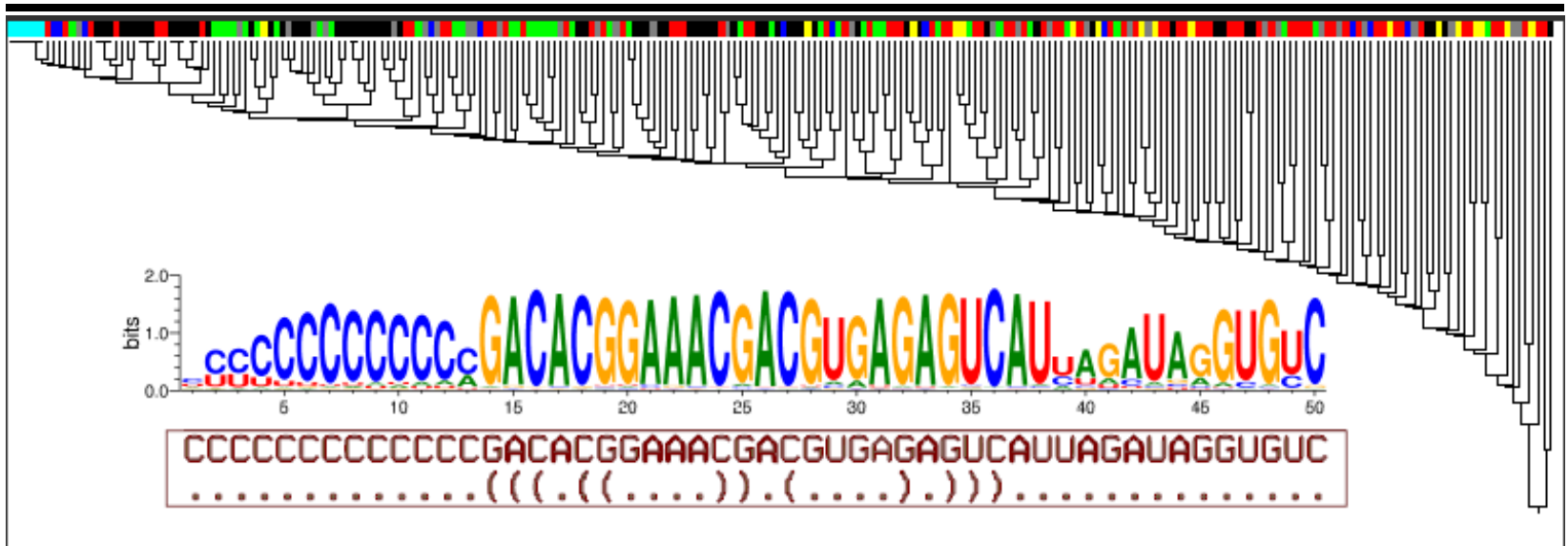


EVOLVED

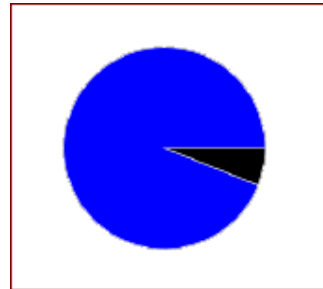
optimal repl

av. random

High mut. rate: 1 quasispecies LOW variability



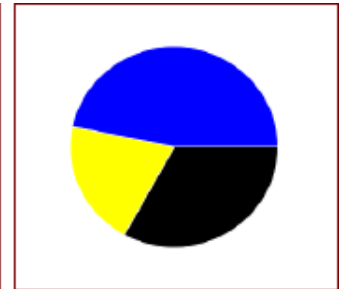
mutational NB: STEEP and “special”



EVOLVED



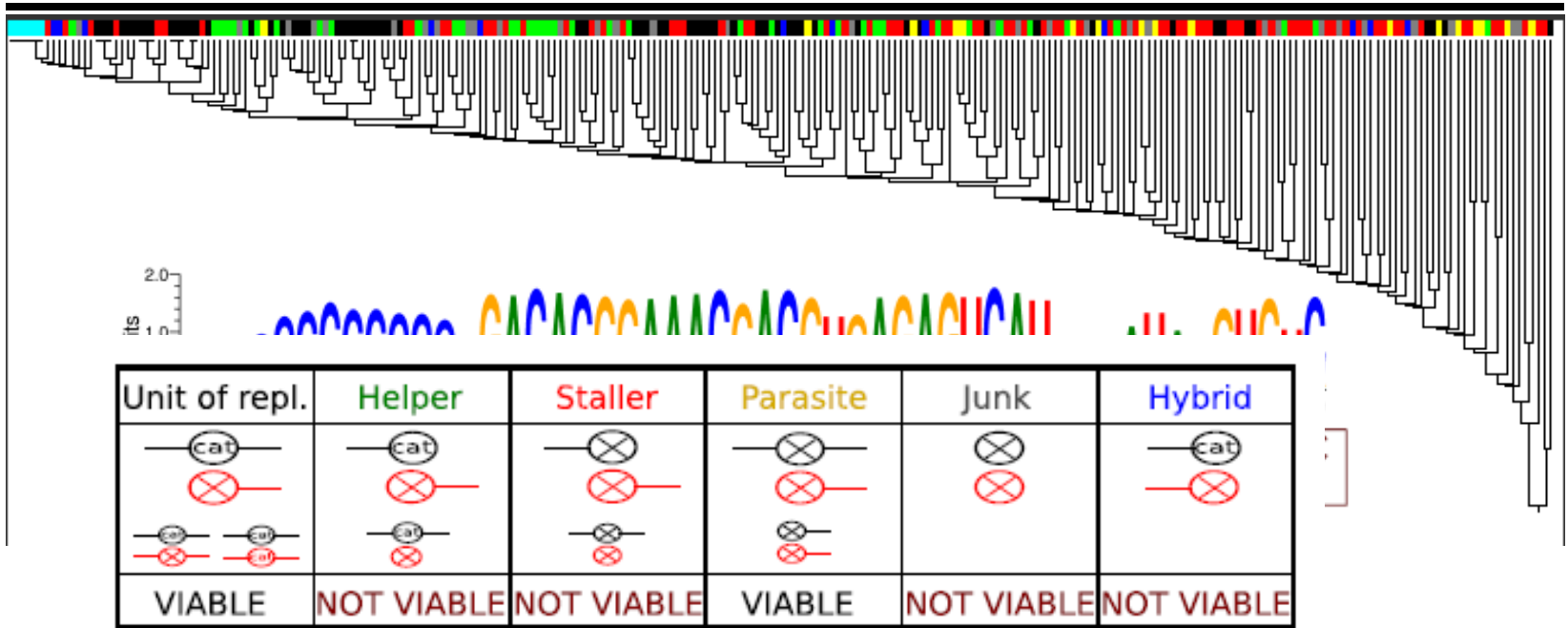
optimal repl



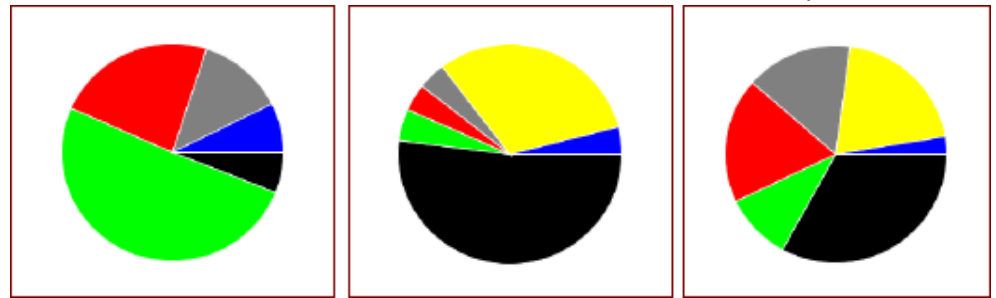
av. random

black replicator; blue rest; yellow parasites

1 quasispecies: codes for multiple functions



mutational NB: STEEP and "special"



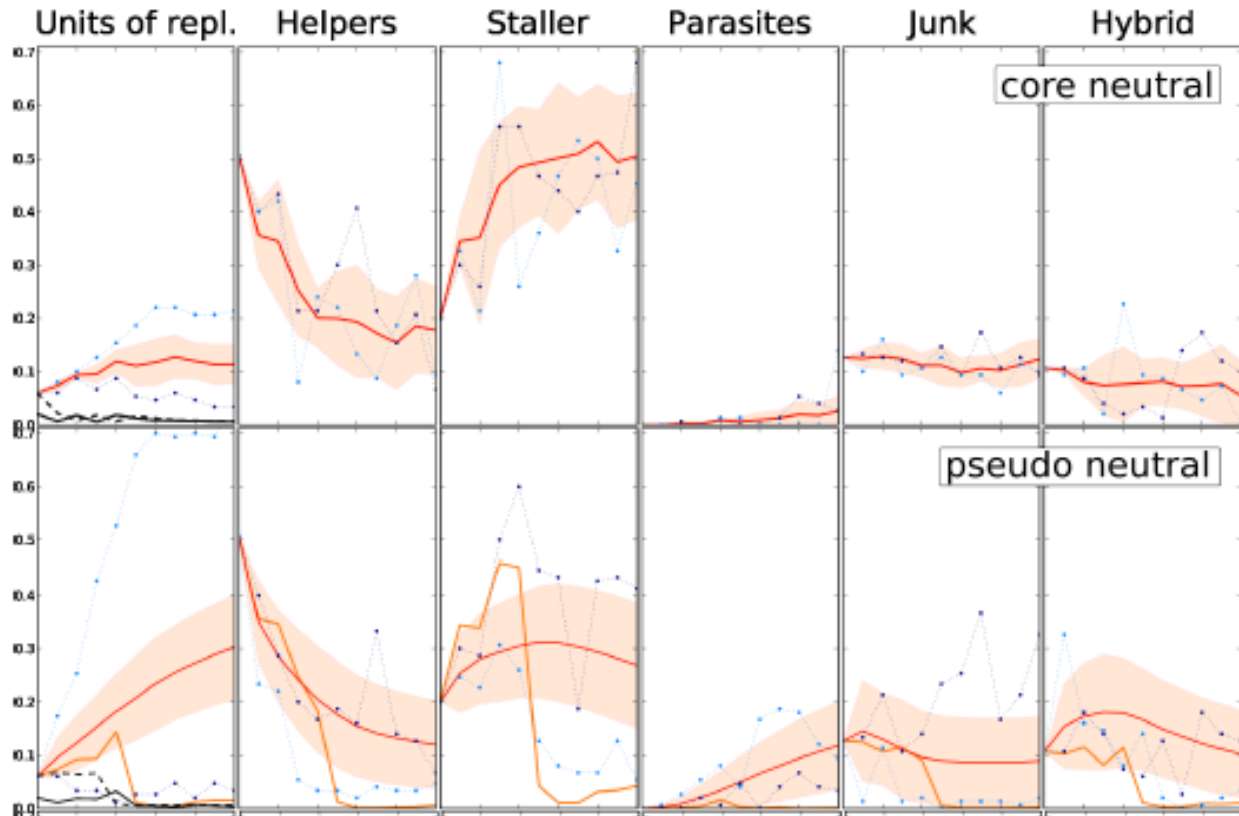
EVOLVED

optimal repl

av. random

black repl.; blue rest; yellow parasites; green helpers; red stallers; gray junk

mutational neighborhood at larger Hamming distances

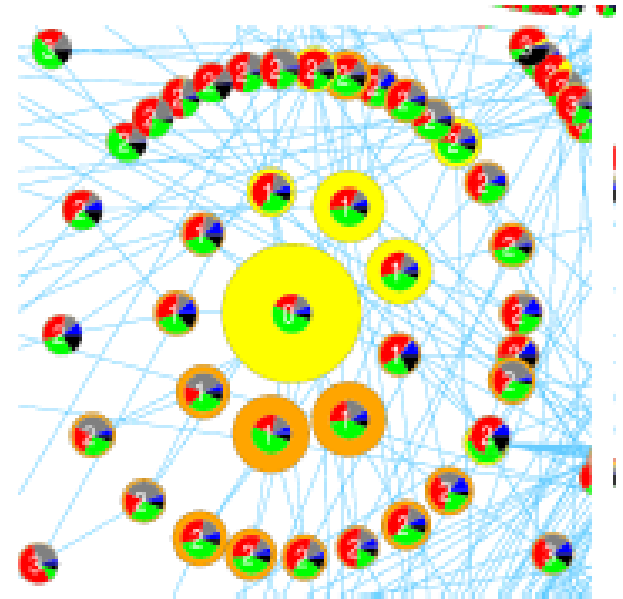
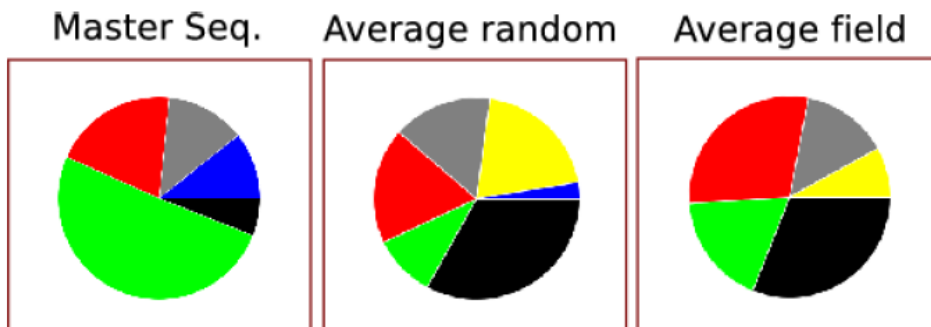


Top follow replicates with \geq replic rates masterseq. bottom follow replicates with $<$ replic rates masterseq.

quasispecies composition in field

weakly reflects mutational neighbourhood
more replicators (because of replication
, less helpers , more stallers (like neighborhood of other repli-
cases)

Replicases with 'good' mutational neighborhood overrepre-
sented.



Helpers “help”

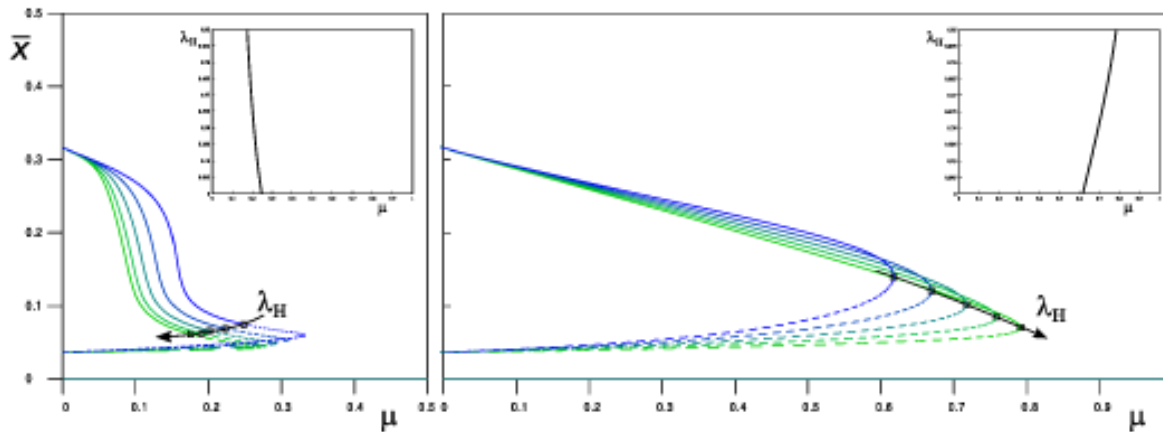
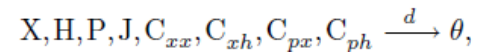
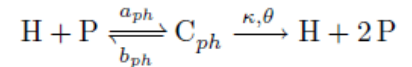
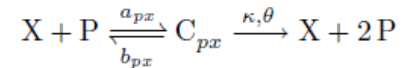
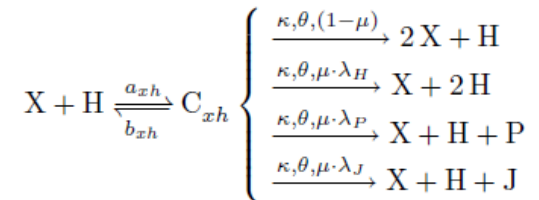
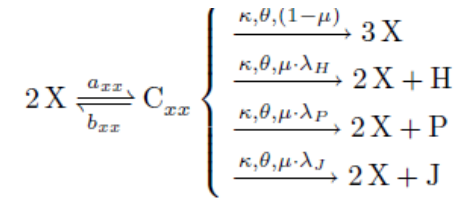
change in junk – > extinction

change in empty – > extinction

in simplified ODE model:

increases max μ without parasites

decreases max μ with parasites



with parasites

without parasites

Stallers “stall”

change in junk – > increases density

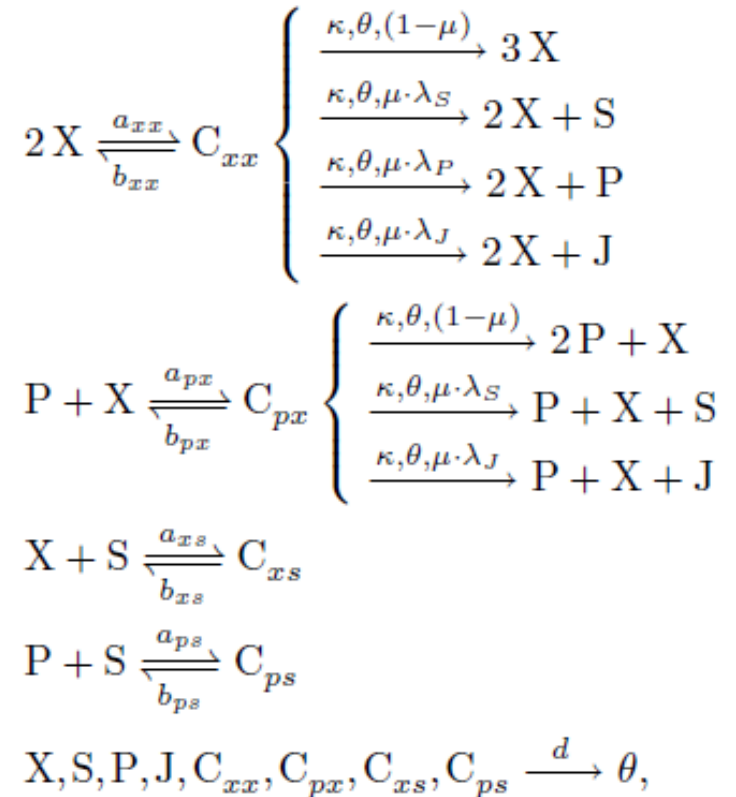
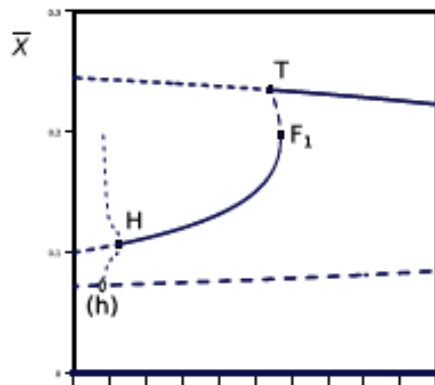
BUT master seq. replaced

‘pseudo stallers’ evolve

change into empty space









parasite lineage evolves!

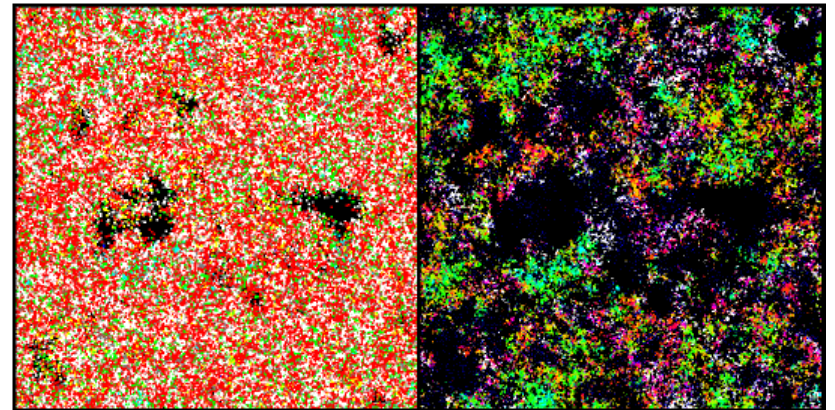
in simplified ODE model:
protects against parasites



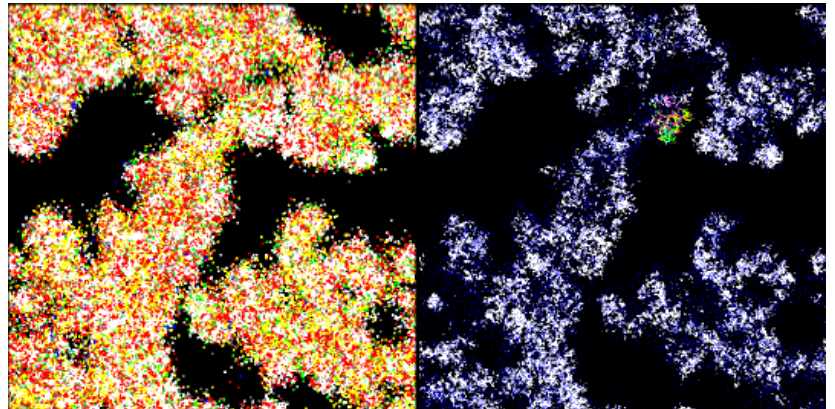
with parasites: x-axis: fraction staller-mutants

Variability of evolved quasispecies

	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

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*