10

Individual and/or ecosystem based adaptation

Genome evolution

Course computational biology 2025; Paulien Hogeweg;

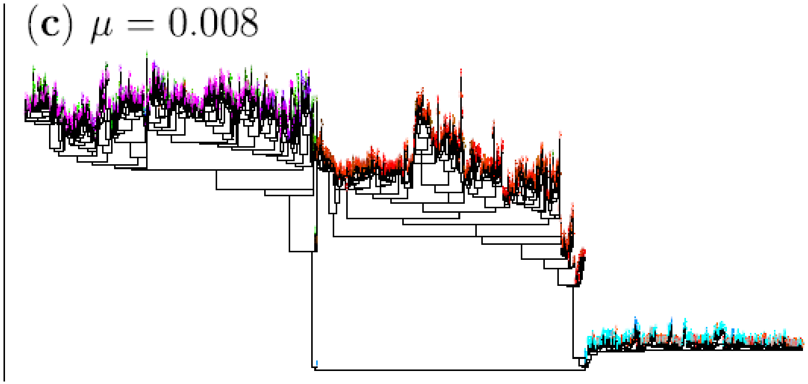
Theoretical Biology and Bioinformatics Grp Utrecht University

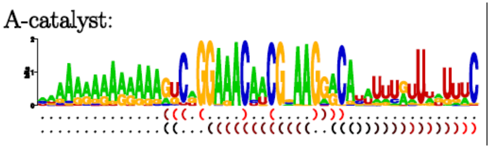
Last time:

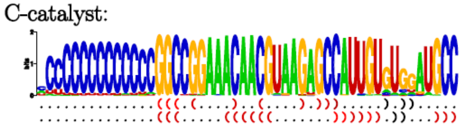
Eco-evolutionary dynamics of RNA -
sequence-structure-interaction: Coding strategies

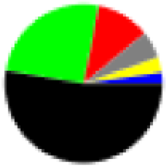
- Symmetry breaking + and - string like in minimodels ($->$ not model artefact of mini model $->$ not due to coding limitations of RNA)
- Coding evolves to to cope with high mutations rates
- Steep quasispecies with "functional" mutational NB (vs Eigen: no functionality to be expected in quasispecies)
- *individually coded but ecosystem based diversity evolves and persists close to the Information Threshold"*
- Lower mutation rates: speciation: replicators and parasites
- Coding adapted to evological 'role'
- BUT Very different coding of functional similar competing replicators.

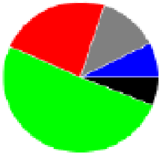
Mutational neighborhood of 2 functionally equivalent RNA’s

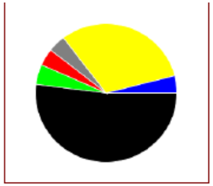


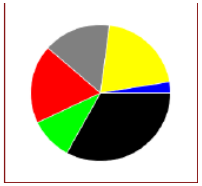












optimal replicator

random replicator

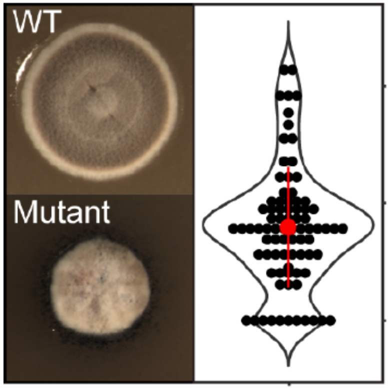
-s,

black replicator; yellow parasite; green helper; red staller

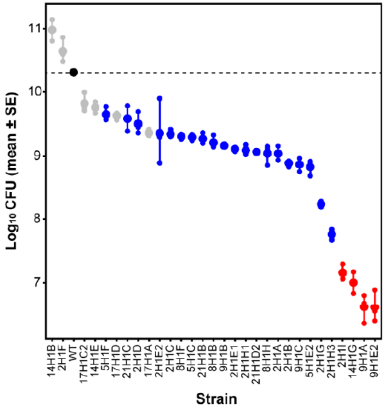
Takeuchi & Hogeweg 2008, Colizzi & Hogeweg 2014

Quasispecies based division of labour:

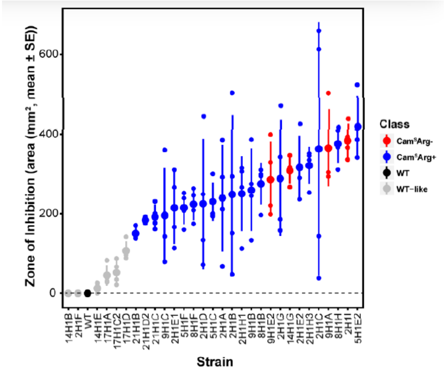
Antibiotic production is organized by division of labour in Streptocyces
Zheren Zhang...Daniel E Rozen,Science advances, 6(3) 2020.



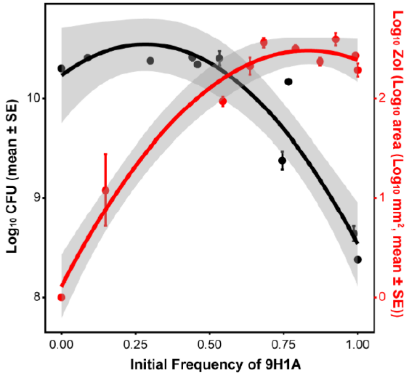
High Mut. Rate



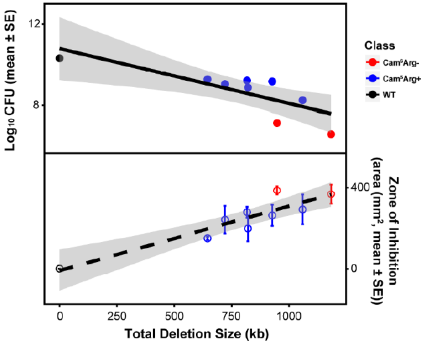
Mut. fitness

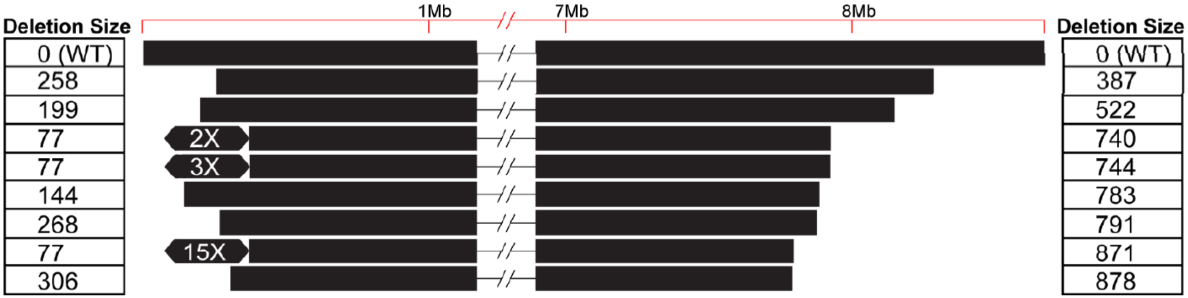


antibiotics production



QS fitness
only > 50





Deletion Size	Deletion Size
0 (WT)	0 (WT)
258	387
199	522
77	740
77	744
144	783
268	791
77	871
306	878

Genome structure and targeted mutations

TODAY:

What does ecology do for evolution to predefined target

Evolution of coding structure to cope with high mutation rates
- multiple coding

To cope with complex ”target”
- sparse fitness evaluation
- ecosystem based ”problem solving”

combatting parasites
- and in evolution of computation

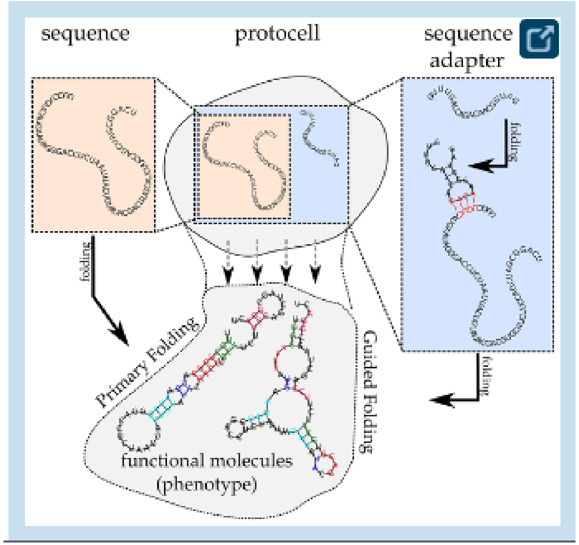
Genome GRN evolution

Evolution of coding structure cont.
versatility of RNA structure: multiple coding

Evolve towards target == set of (25) RNA structures .

ALL other structures (Shapiro) TOXIC

define possible interaction of RNA's:
adaptors (=single hairpin)
can bind to other RNA
bound (modified) nucl not 'available'
for folding



fitness of cell: set of struct.
cells compete in space

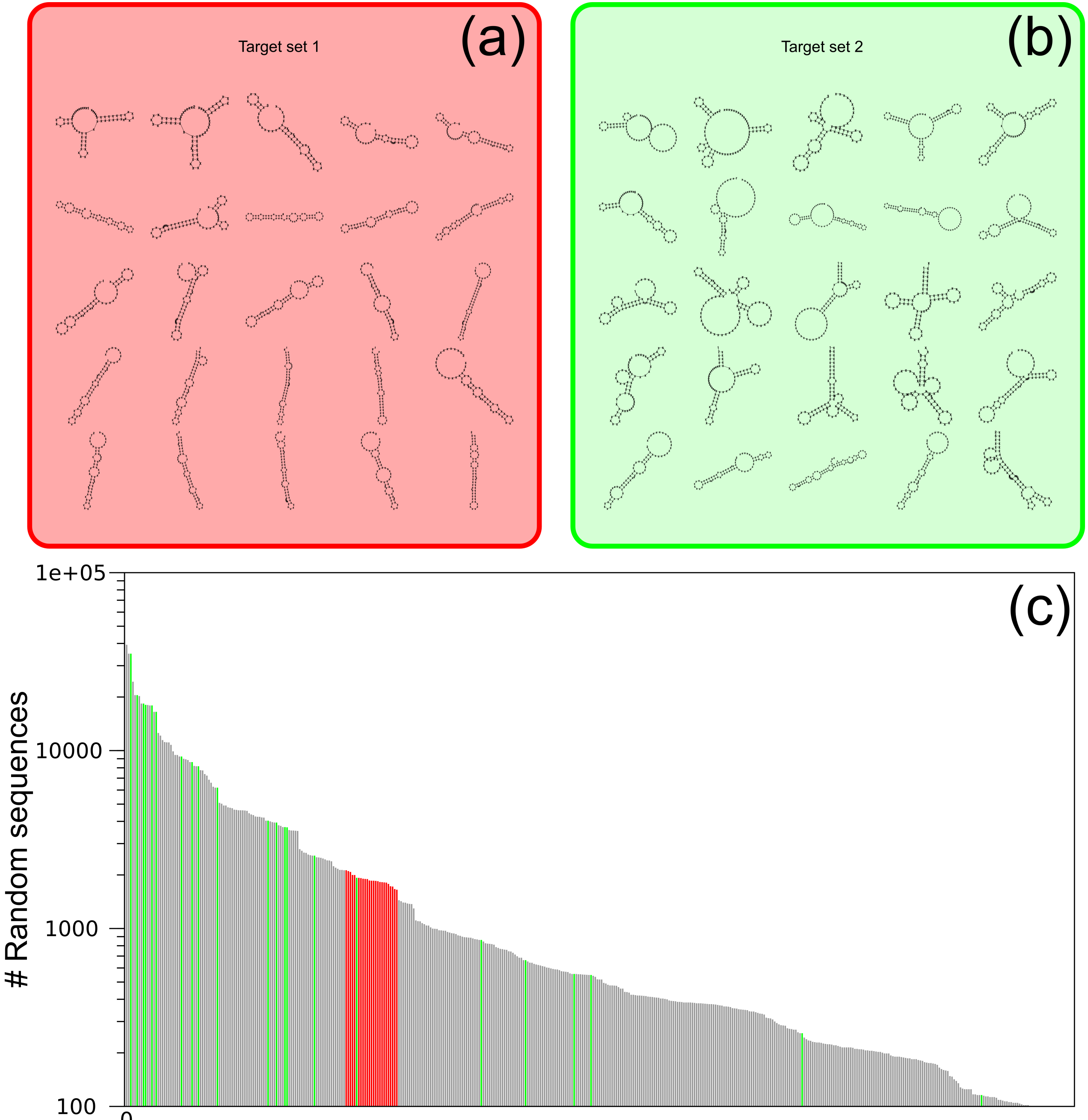
replication of cells and RNA's coordinated
mutations: point mutations;
dupdels of RNA's and parts of RNA's
single bases; point mutations

How to cope with high mutation rates?

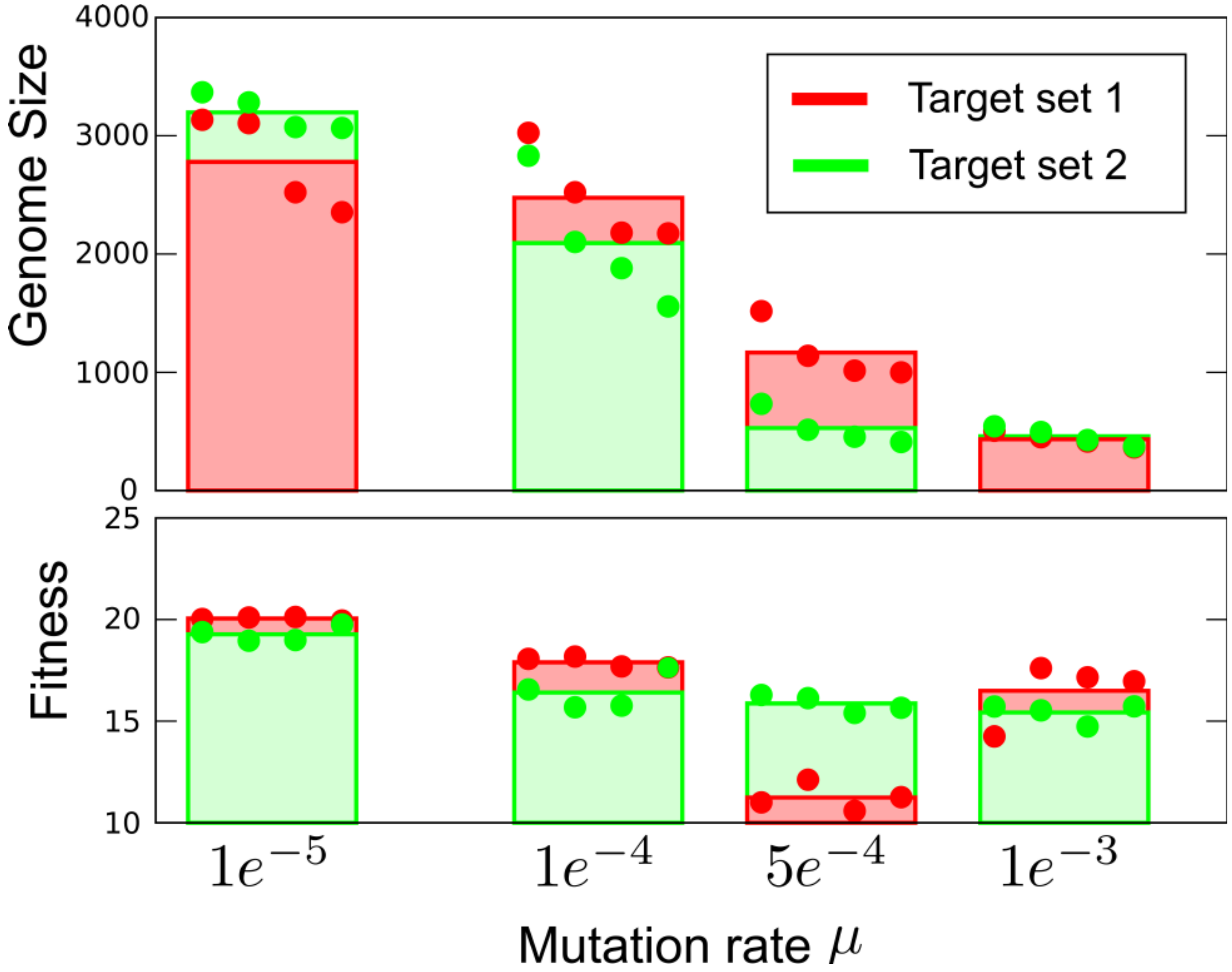
How to cope with high mutation rates?

.

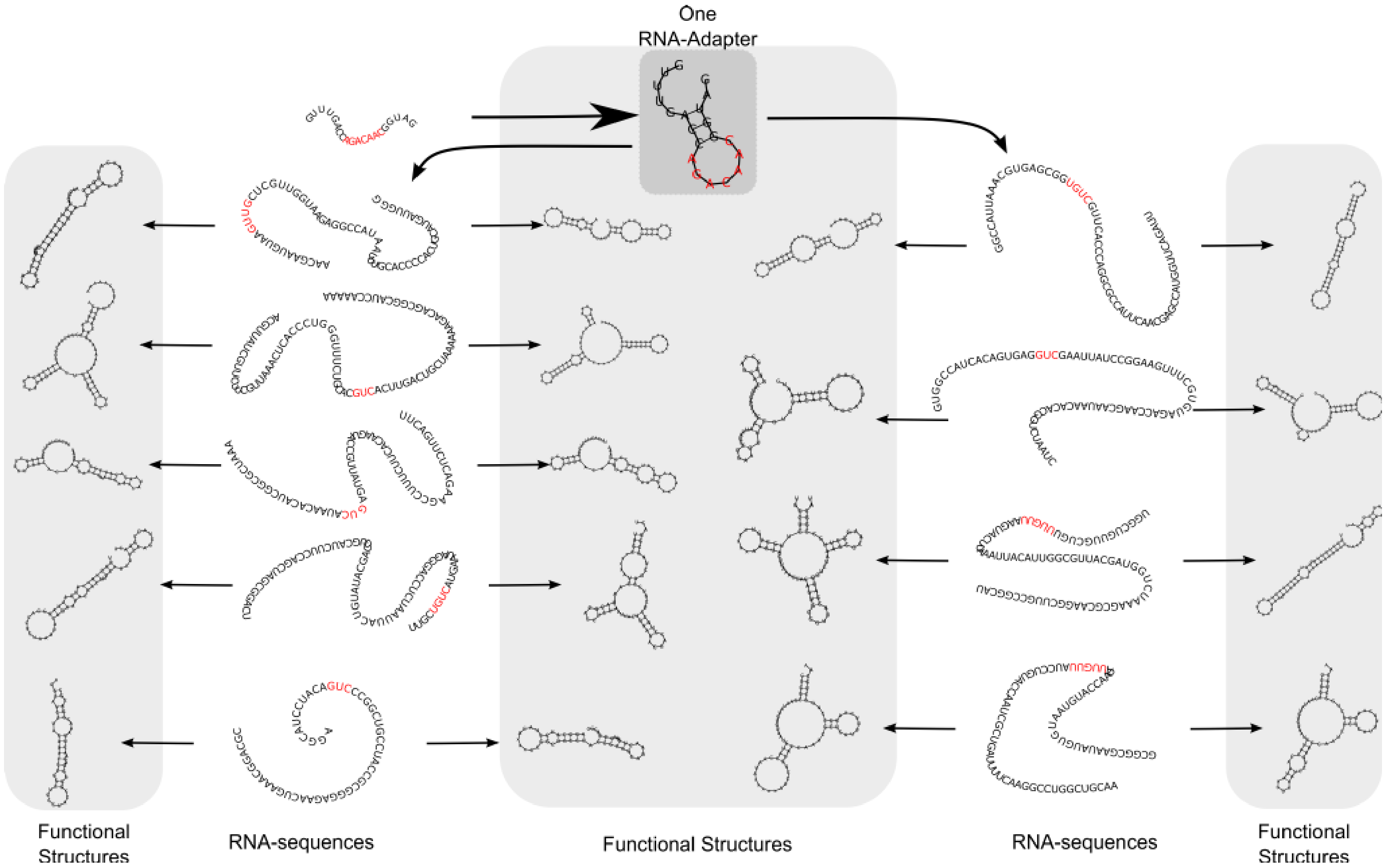
de Boer & H. PLOS-One 2012


(a)
Target set 1
(b)
Target set 2
(c)
1e+05
10000
1000
100
0
# Random sequences

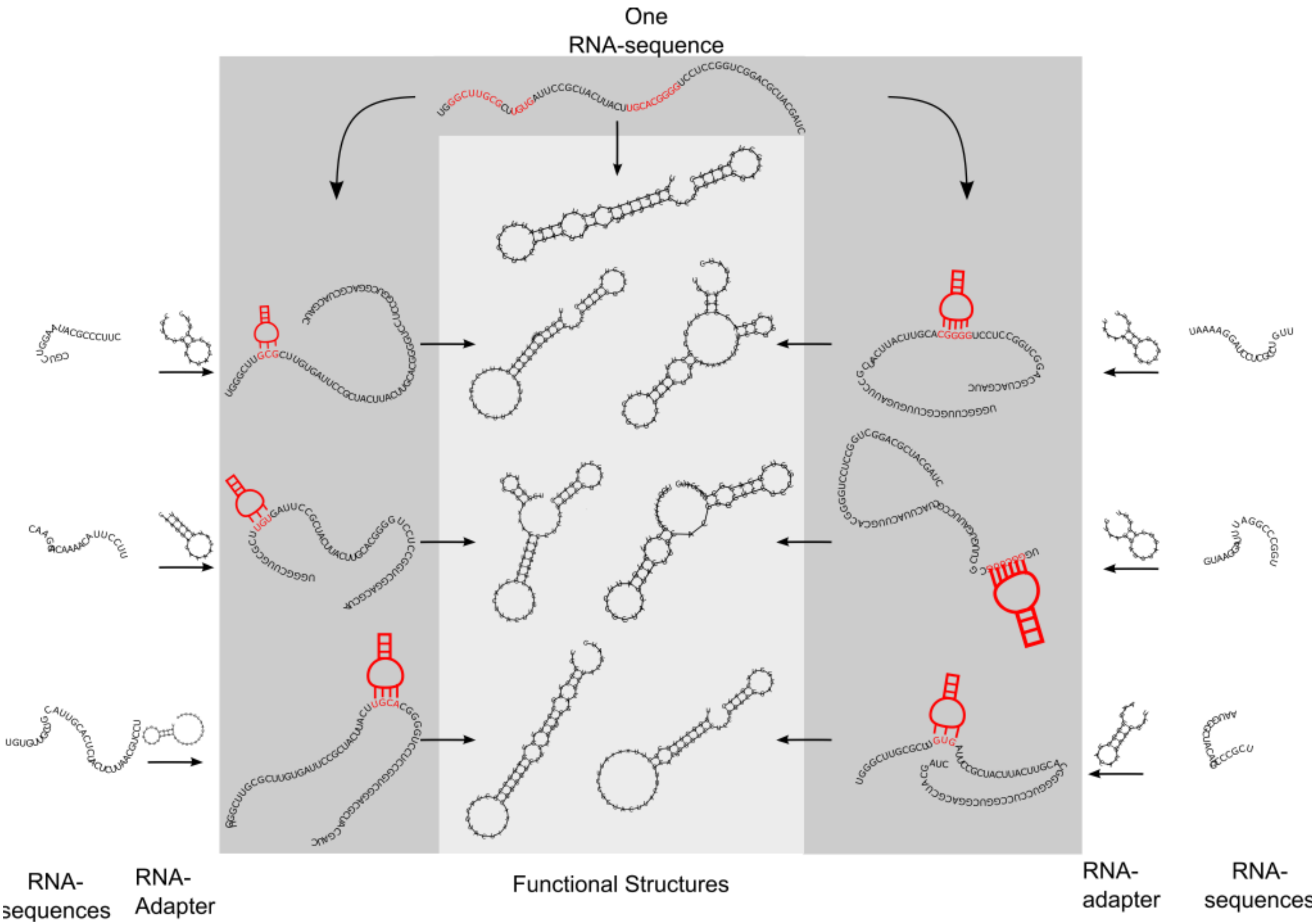
high mutation rate - short genome - same functionality



one adaptor used by all sequences



many adaptors used by 1 sequence



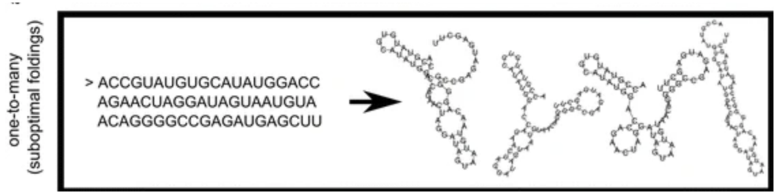
Conclusion: multitple coding

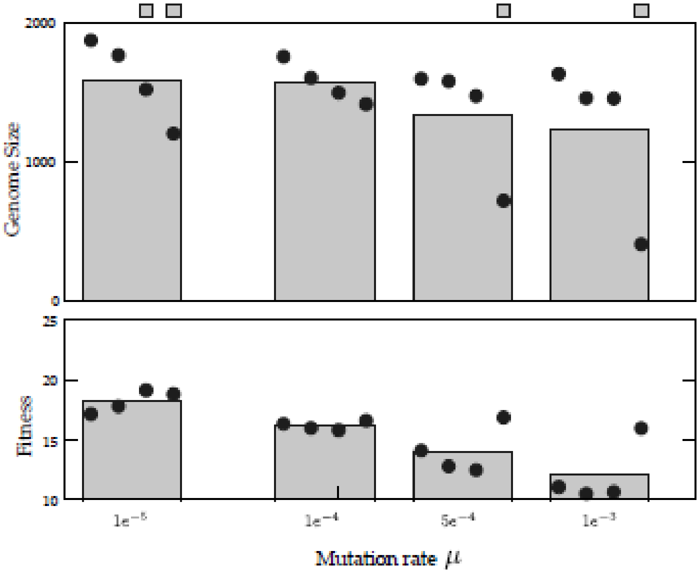
RNA even more an “ideal evolvable molecule”

information threshold shapes coding structure
multiple coding arises and alleviates information threshold

information threshold does not (necessarily) limit functionality

(Similar effects seen with alternative (non-minimal energy) foldings)





Also in this case:
local competition in space helps!

. well mixed:

Conclusions

Coding structure adapts to mutation rate
Coding length, selection strength

Result:

Evolution converges to being

Close to Information Threshold

“Doing something” in spatial eco-evolutionary systems
Life as “function optimization”
individual vs ecosystem based problem solving

- NON 'trivial' task (constructive evolution))
  Problem solving (modeling trick)
- Local vs Global 'fitness'
  sparse fitness evaluation
- Study longterm information integration
- competition and/or cooperation

Spatial pattern formation and speciation enable evolving complex problem solving

Local Competition and Co-evolution as Optimization Strategy (Hillis)

- Evolution of FAST sorter
- Coding:
  Diploid Shuffles
- Fitness:
  # of correctly sorted (side effect)
- FAST SORTERS

Green's Sorter

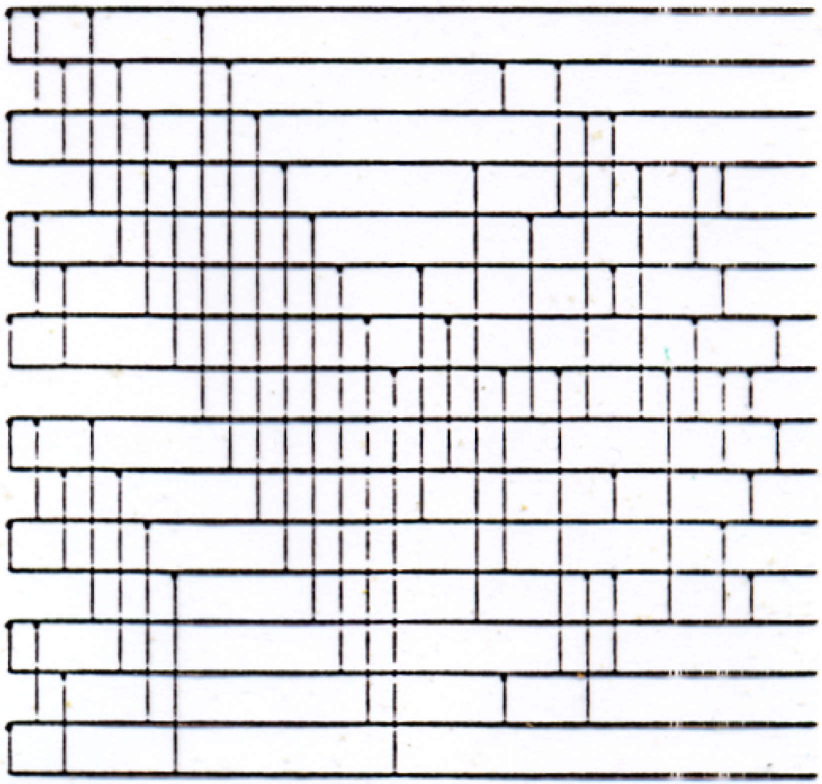


Fig. 2. Green's 60-comparison sorter.

Function fitting as model for (complex) TODO
coevolution as optimization strategy
sparse fitness evaluation through co-evolving problems
cf Pagie and Hogeweg 1997

ISSUES

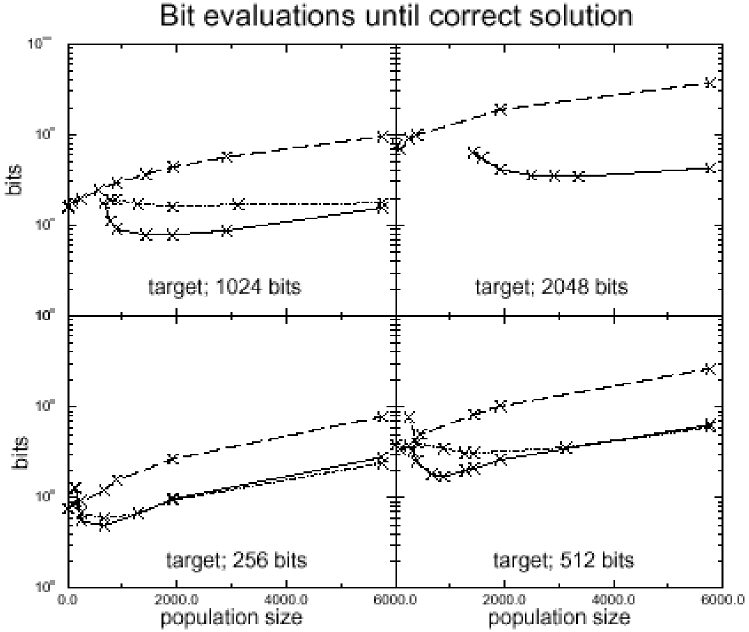
- How gets the 'complete' problem solved
  - Information integration
  - generalizability (never 'seen' cases)
- What type of solution is generated
  - complexity of solution
  - mutational robustness
  - generalizability

compare sparse vs complete fitness evaluation

in SPACE (local competition)

Individual based problem solving: Information integration sparse fitness evaluation more efficient (< number of evaluations)

Linear model: bitstring match

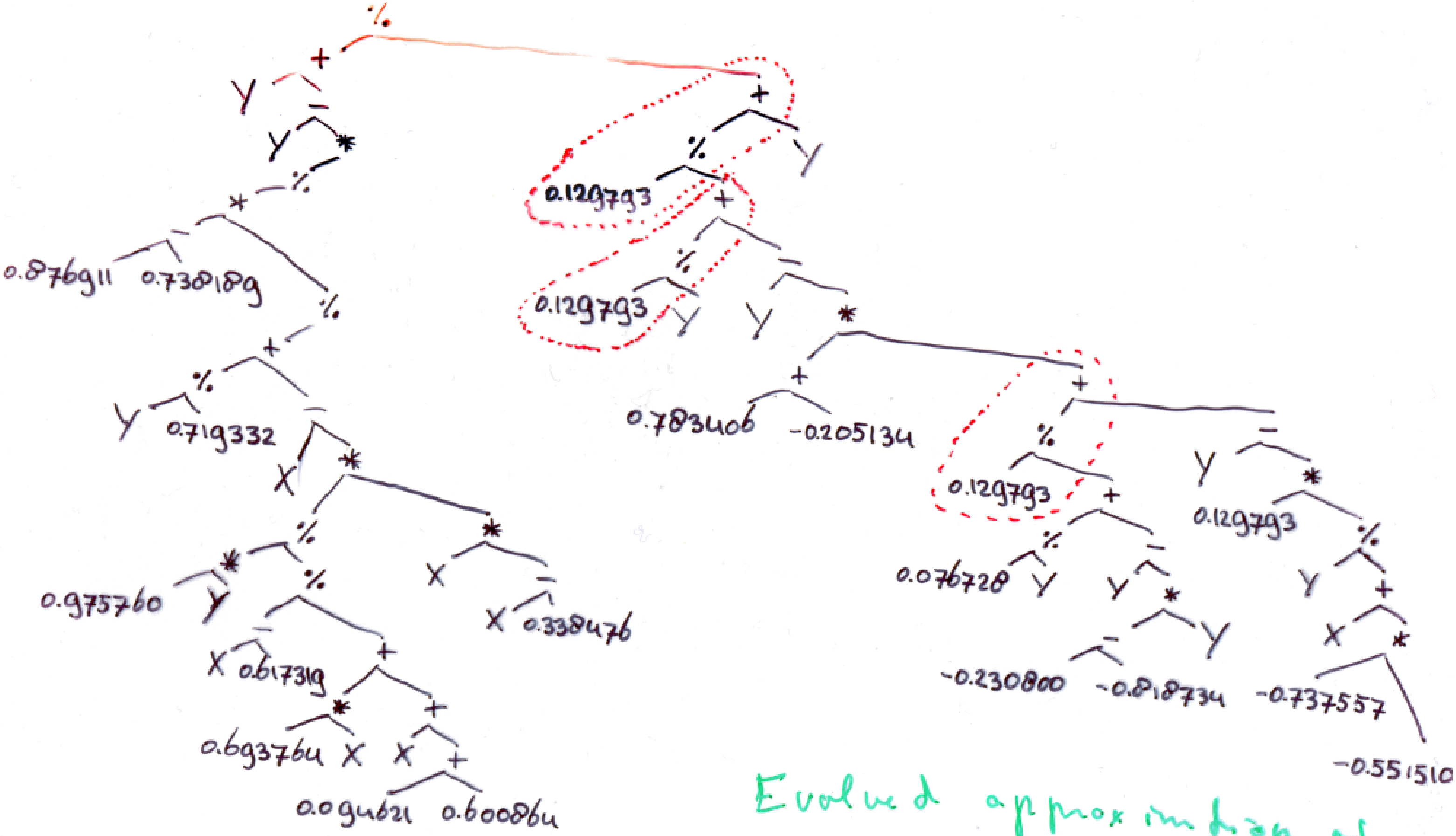


Individual problem solving Type of solution: complexity, generalizability, mutational robustness

- function: $f = 1/(1 + x^{-4}) + 1/(1 + y^{-4})$
- Genetic programming without minimum atom / function set
  -> alternative solutions
- co-evolved solutions and sampled points in space
- compare sparse fitness evaluation:
  (only some values (8) seen per lifetime) and coevolution
  with 'complete' fitness evaluation (many values seen (here 26^2))
- fitness distance to target functions

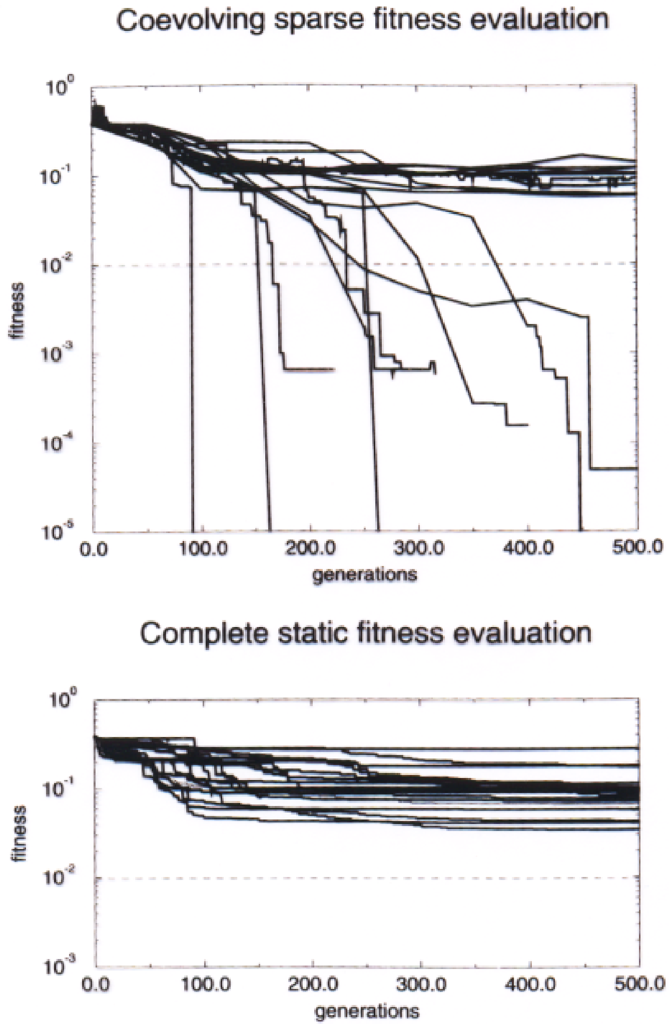
sparse fitness evolution
better fit (distance)
“*better* fit (simpler function)
“*better* fit (more generalizable)
LOWER mutational robustness

Pagie and Hogeweg 1997



Evolved approximation of

$$f = \frac{1}{1+x^{-4}} + \frac{1}{1+y^{-4}}$$



Coevolving sparse fitness evaluation

Complete static fitness evaluation

Figure 2. Fitness curves of the best-of-generation solution for coevolving (a) and complete static problem evaluation (b). Fitness is based on the complete problem set that consists of 26×26 problems. The fitness curves that drop below 10^{-5} go to values between 10^{-15} and 10^{-17}. The horizontal dotted lines give the value of the hit criterion (see text).

Looking at solutions

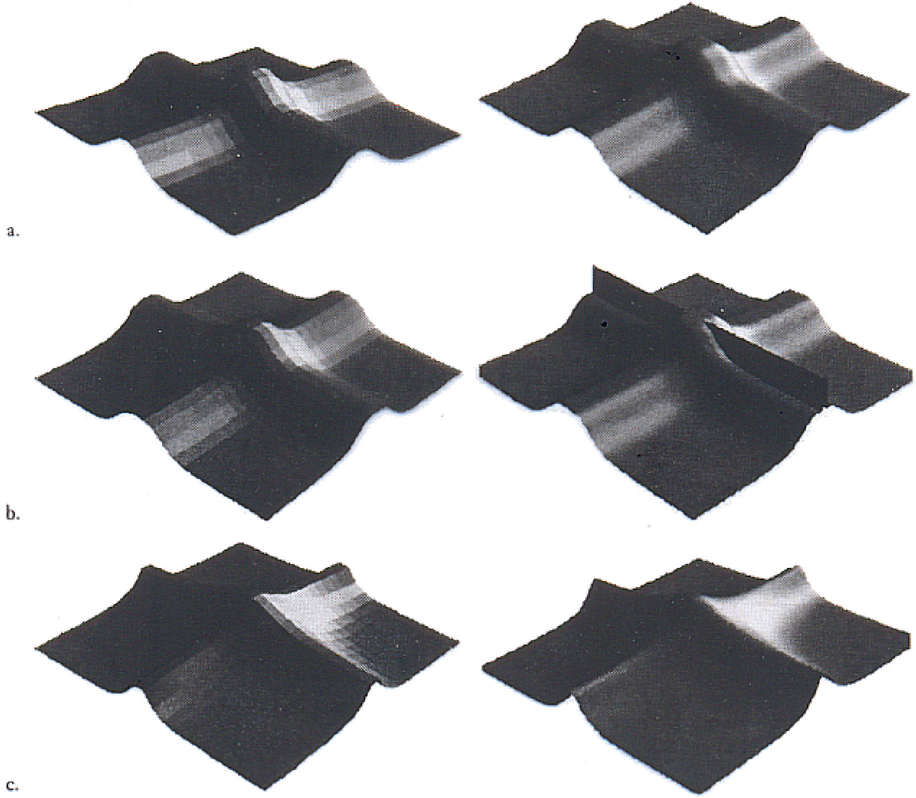


Figure 3. Three typical final solutions produced by coevolving fitness evaluation. The left plots are based on 26 × 26 evaluated problems, the right plots on 100 × 100 evaluated problems. Two correct solutions that approximate the target function are shown in (a) and (b); an incorrect solution is shown in (c). All solutions generalize well on the 100 × 100 problems.

Static vs. sparse fitness evaluation (unseen data)

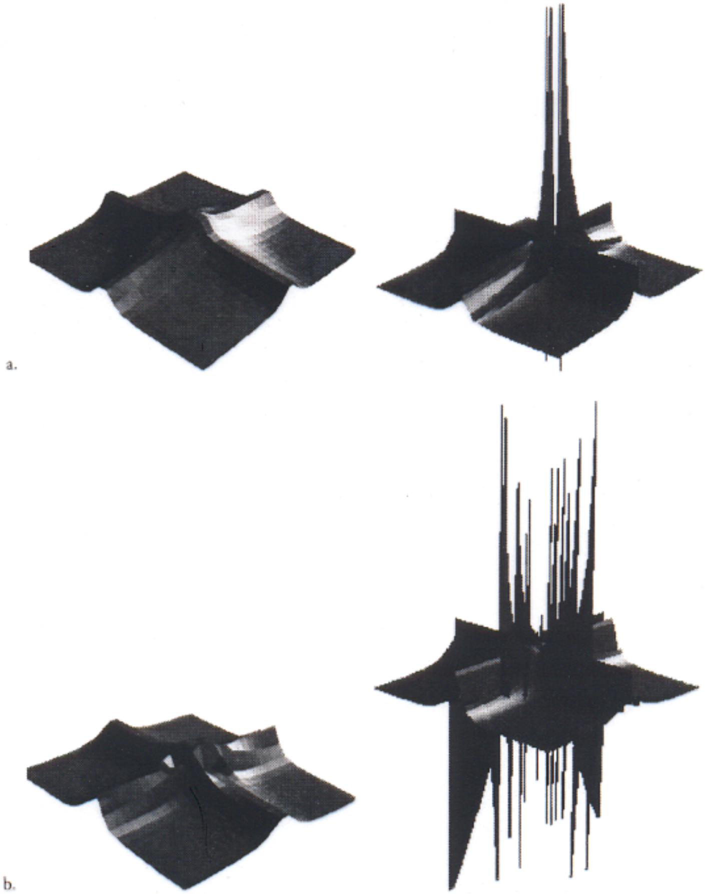


Figure 4. Two typical final solutions produced by static fitness evaluation. The left plots are based on 26 × 26 evaluated problems, the right plots on 100 × 100 evaluated problems. Neither solution is correct.

Mutational Stability

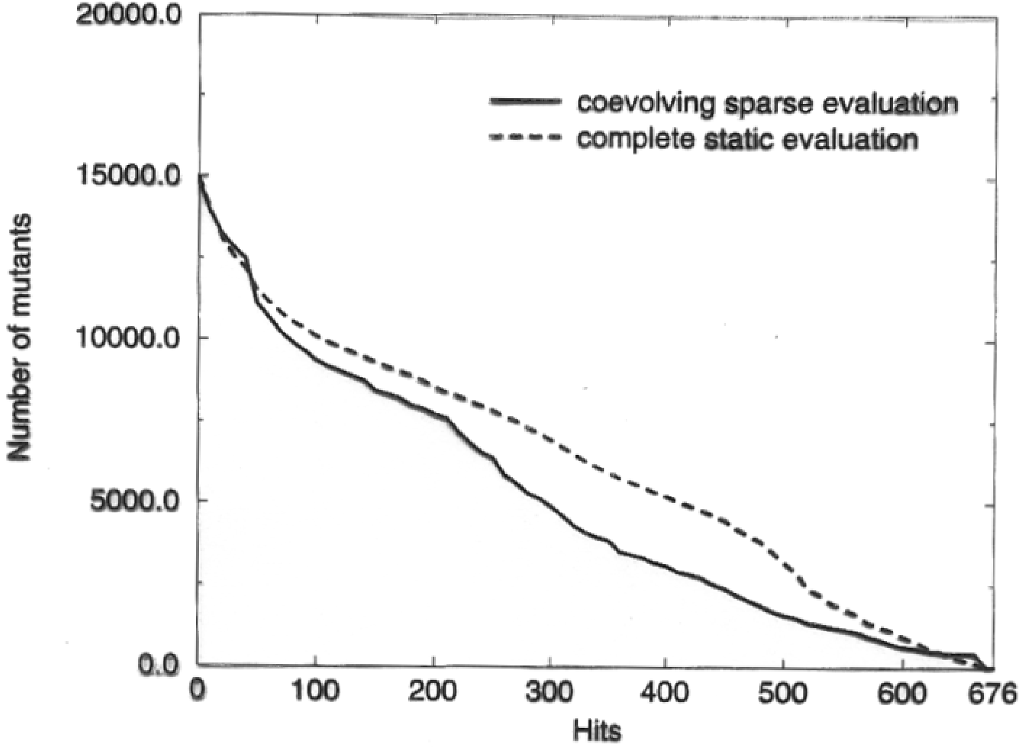
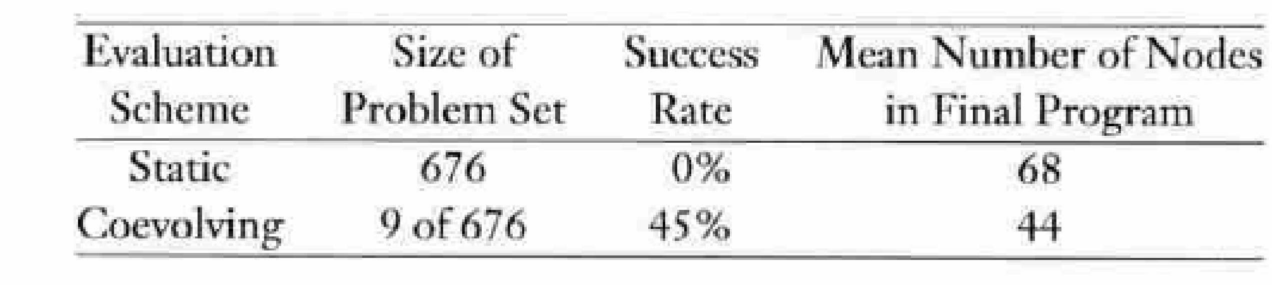


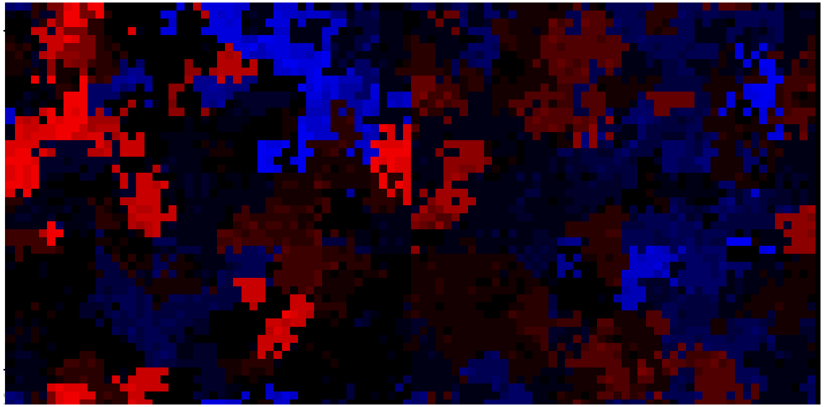
Figure 5. Histogram of the number of one-point mutants having at least *x* number of hits.

Genome size

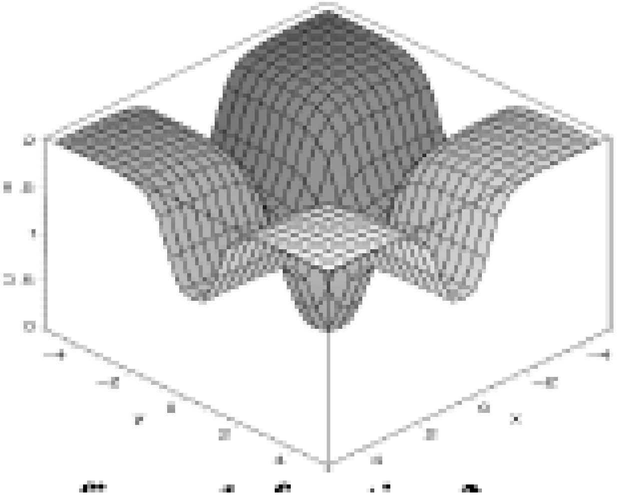


Evaluation Scheme	Size of Problem Set	Success Rate	Mean Number of Nodes in Final Program
Static	676	0%	68
Coevolving	9 of 676	45%	44

Spatial pattern formation and PARASITE speciation



$$G = 1/(1 + x^{-4}) + 1/(1 + y^{-4})$$



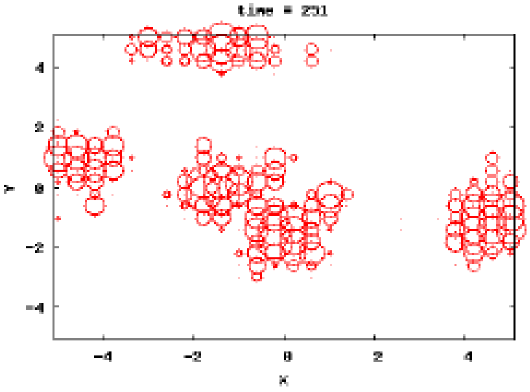
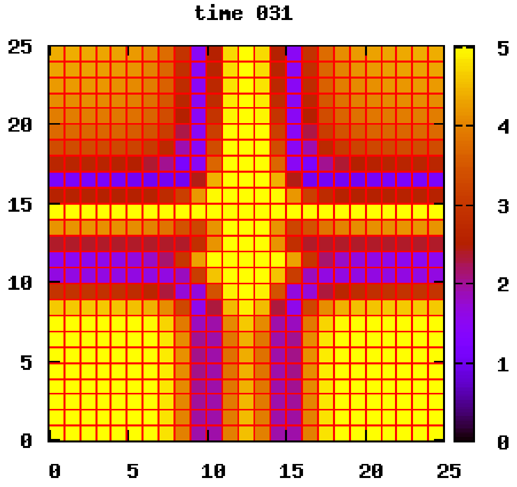
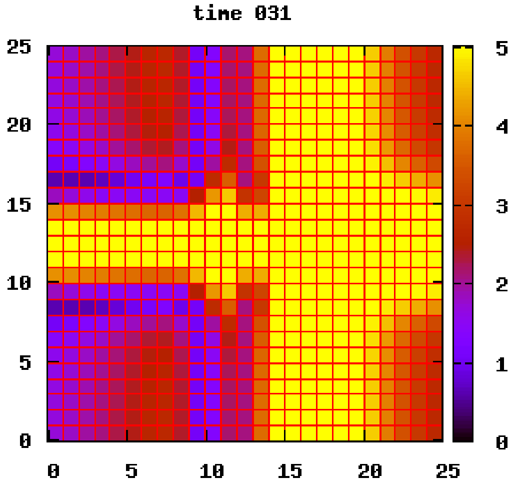


figure 5: parasite speciation in evaluating

Differentiation of host phenotypes: "good at eating different prey"





Spatial pattern formation and speciation vs red queen evolution HOST ANCESTOR TRACE

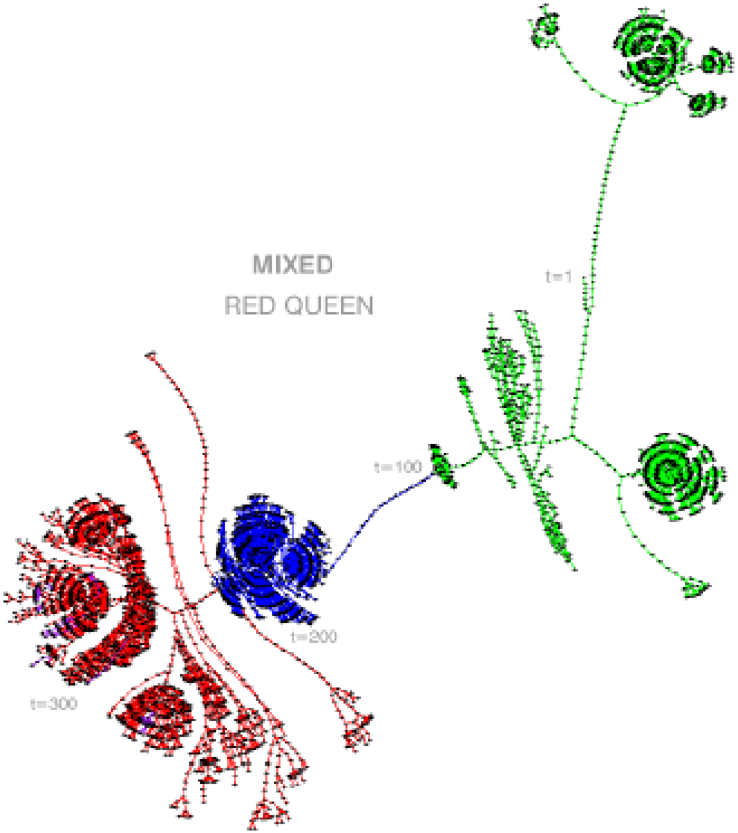


figure 6: combined ancestor trace in well mixed system

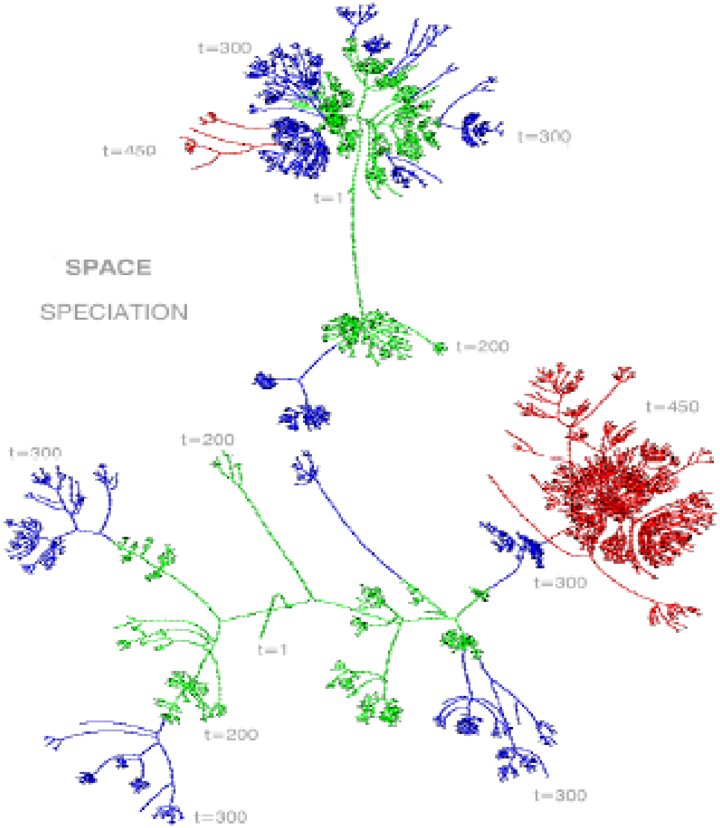


Figure 7: combined ancestor trace with pattern formation

conclusions

Evolutionary “setting” influences evolved structures at multiple levels even if identical “target”.

Sparse fitness evaluation:

better results

mutational less robust

more generalisable
(=== more robust to changes in environment

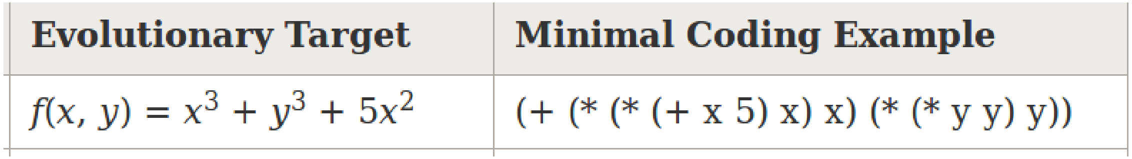
smaller genome

Individual based vs ecosystem based "problem solving" predator-prey-scavenger coevolution

FK de Boer, P Hogeweg 2010

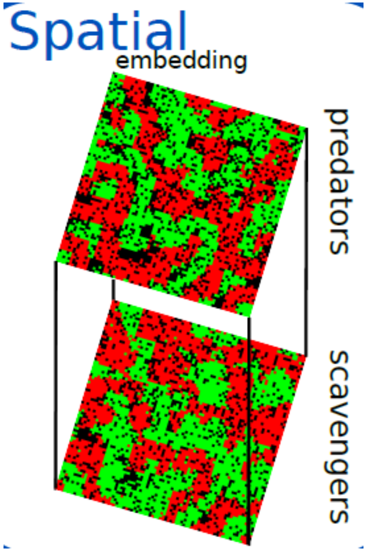
Problem: solve "function" - fully digest all possible prey
prey 2 continuous properties:0 $< X, Y < K$
Fully eaten when predator calculates f(X,Y) correctly

e.g.



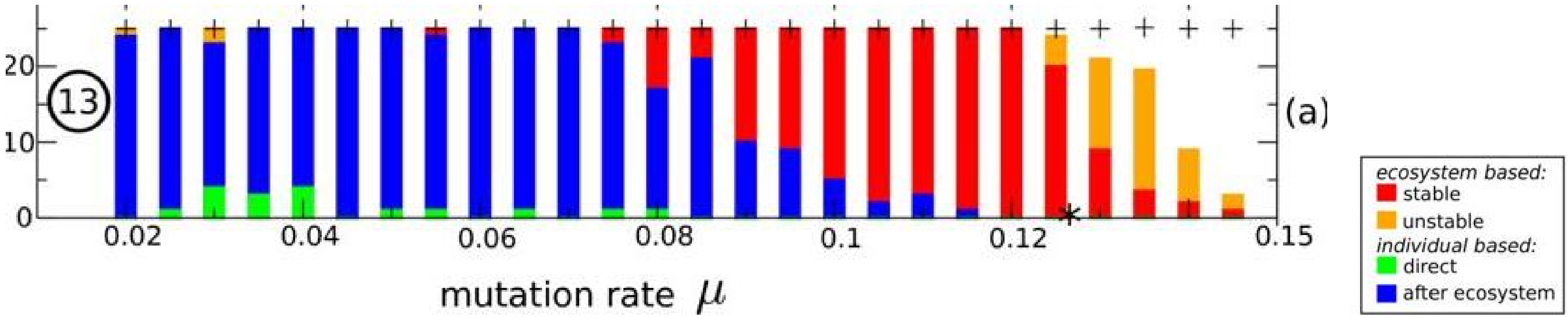
Evolutionary Target	Minimal Coding Example
$f(x, y) = x^3 + y^3 + 5x^2$	(+ (* (* (+ x 5) x) x) (* (* y y) y))

Fitness predator: how well it solves "its" prey
Fitness prey: how badly predator solves it
Fitness scavenger: How it solves "what is left"

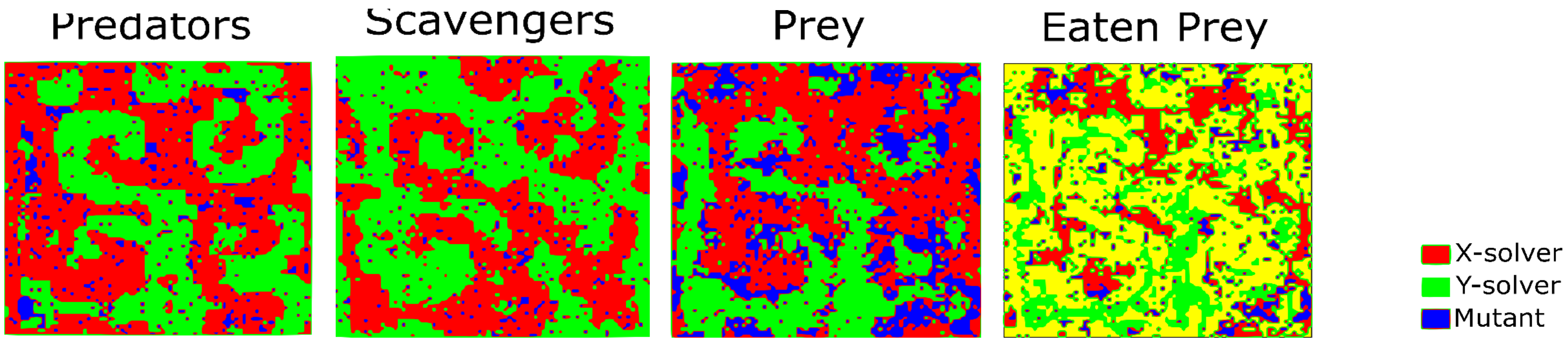


do individual predators, or does the ecosystems solve it

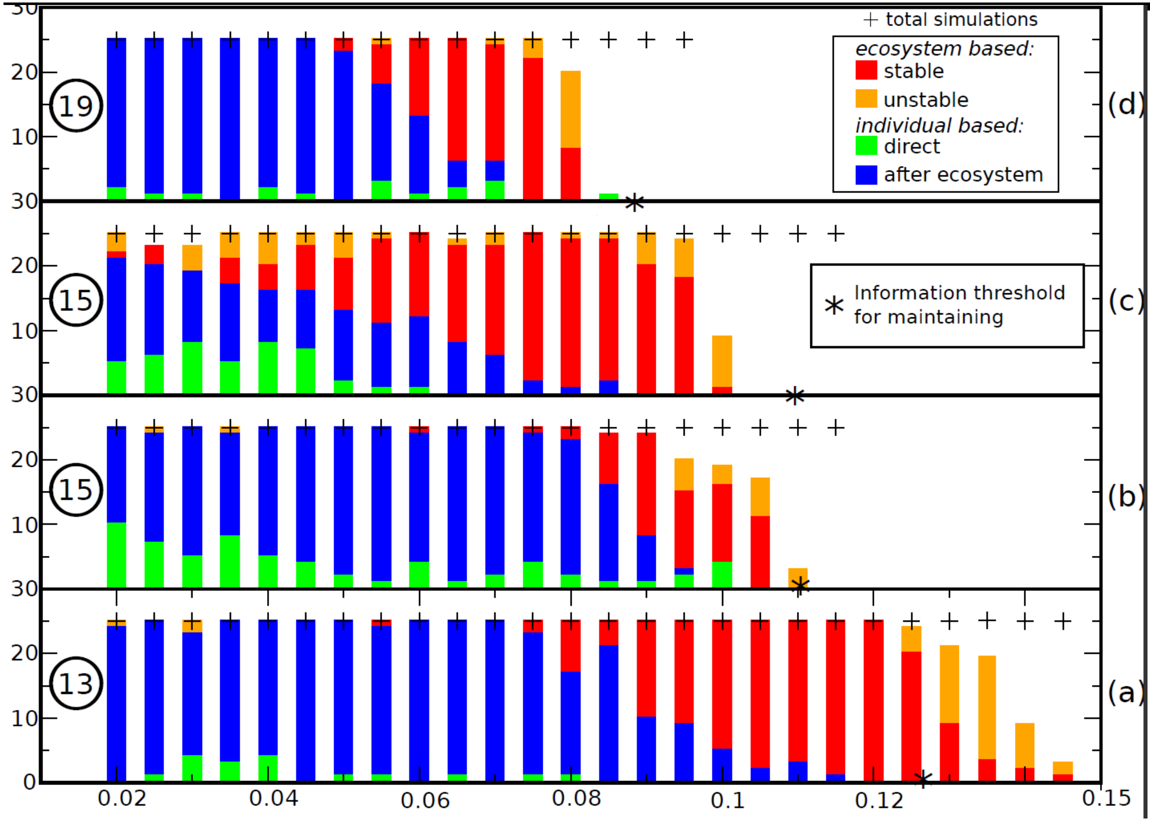
Ecosystem based solution 'easier' to evolve preceeds individual based solution

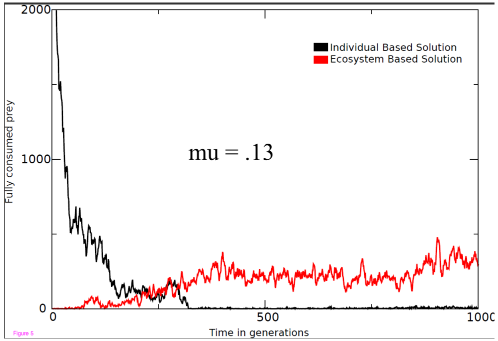


Two predator populations specializing on X or Y
Two scavenger populations specializing on X or Y
Two prey populations with high X or high Y values
Self organize in spiral waves,
X predator and Y scavenger pairs together digest prey fully
(i.e. encode the target function correctly)



high mutation rates lead to ecosystem based solutions (cf hypercycles, RNA model)





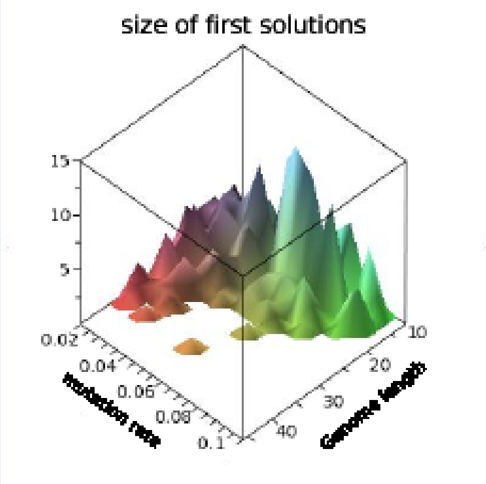
evolving information processing

maintaining information processing

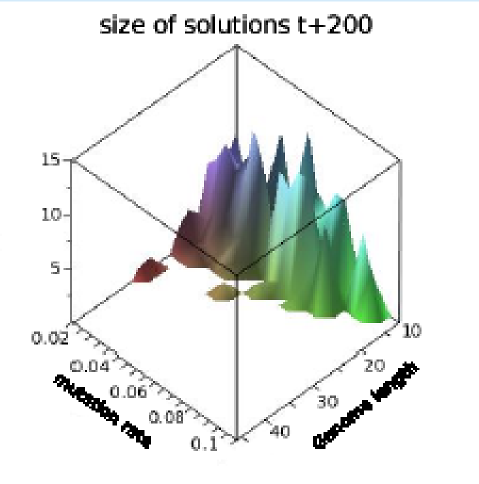
Conclusions

Ecosystem based solution feasible when mutation rate too high for individual to "fit in"

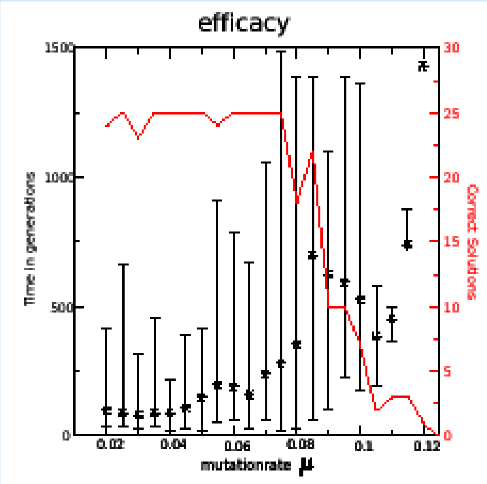
Difference between generation and maintainance of individual based solution



High mutation rates prevent large genomes



After prolonged evolution streamlining of genomes



genome inflation facilitates evolvability

high mutation rate prevents genome expansion and compromises evolvability

conclusion

“Cooperation” (getting something done together) through spatial selforganization

Division of labor among predators

Coerces prey into certain types

See less $->$ can do more:
Cooperative solution of “all” problems, by “seeing” only a subset of problems

No direct or indirect fitness benefit for predators to give scavengers an eatable bite.

ecosystem based solution precedes individual based solution
ecosystem based solution stable at high mutation rates

conclusions

Individual information integration (smart individuals)
only if sparse fitness evaluation
however integrated on all neighbours
Host - parasite system

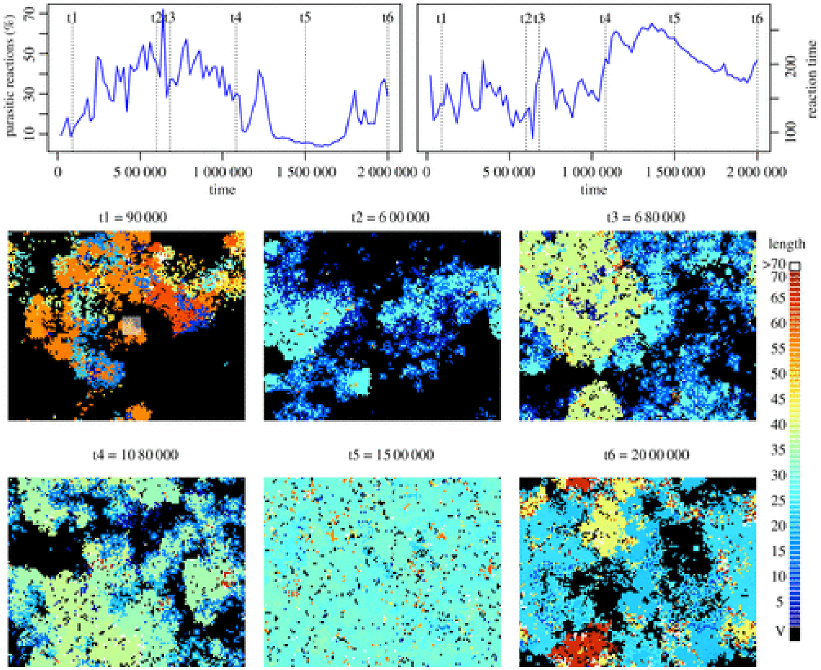
Collective problem solving: complex/smart ecosystems
selection for specialization
e.g only profit when relative to neighbours on which you are better
Predator - prey

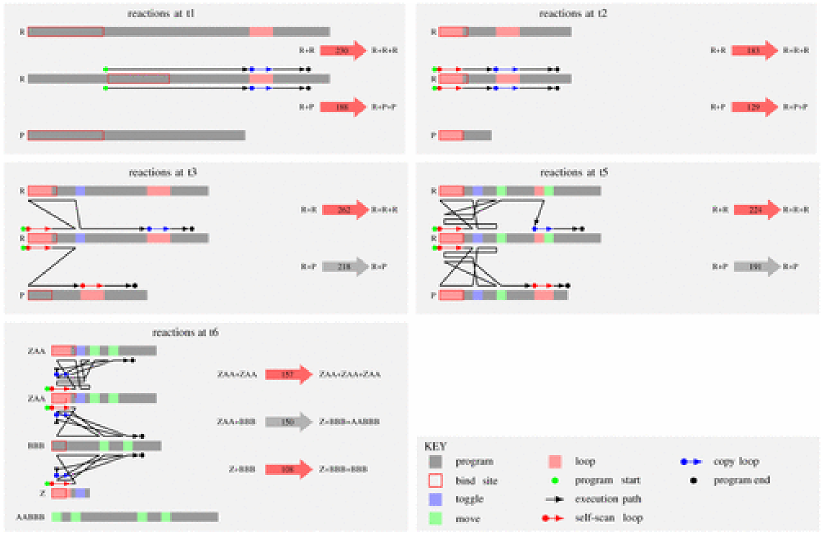
“generalized ’immune’ system” vs “specialized predators”

Evolution of defense systems

“nothing in biology makes sense except in the light of parasites”

Hickinbotham et al 2021





Evolve RNA-like computer programs(STRINGMOL).

Parasites emerge. kill system UNLESS in space

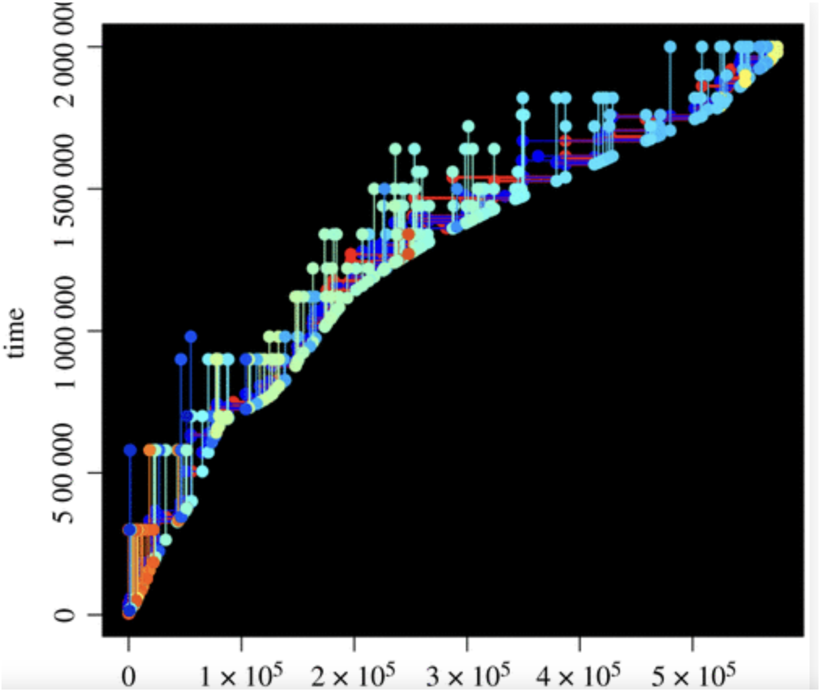
Programs evolve self-nonself recognition

Parasites mimic replicators, but are recognized again

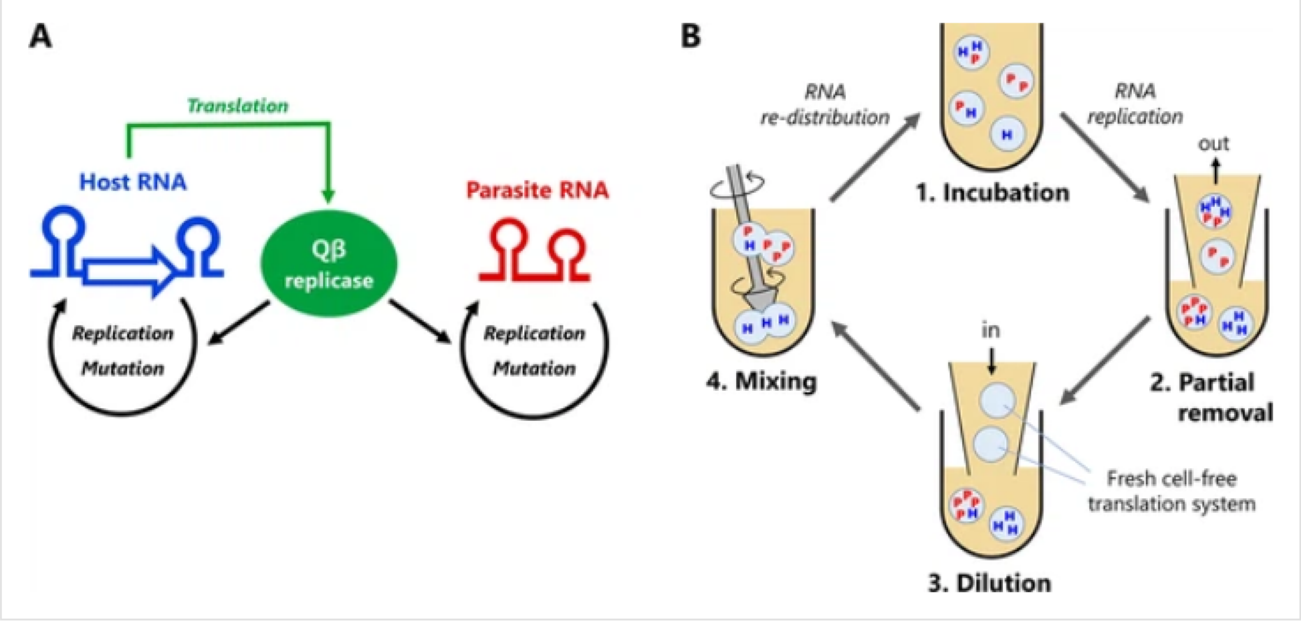
After parasites are extinguished, simplication of replicators

New parasites emerge

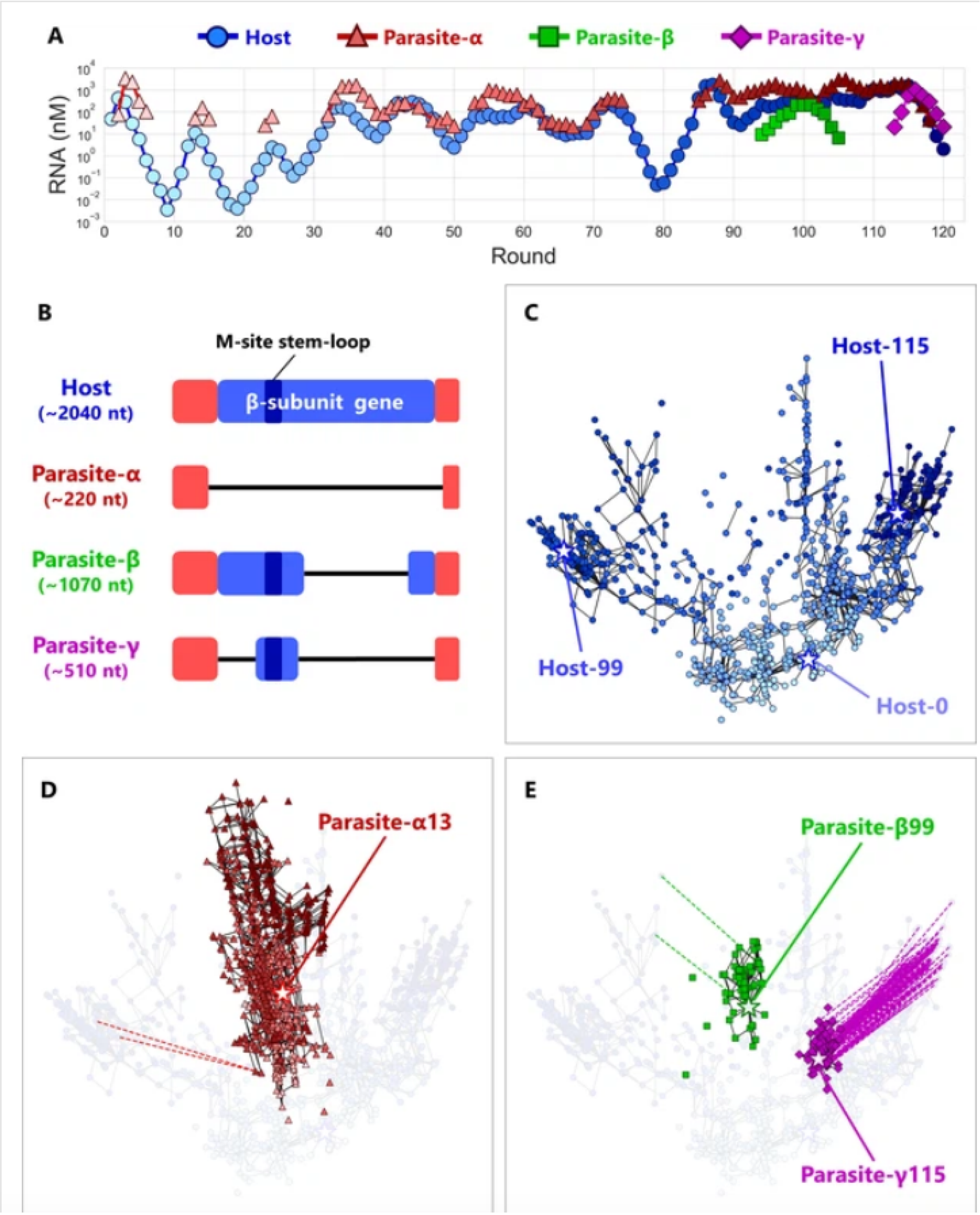
No long-lived parasitic lineages: directly derived from replicators

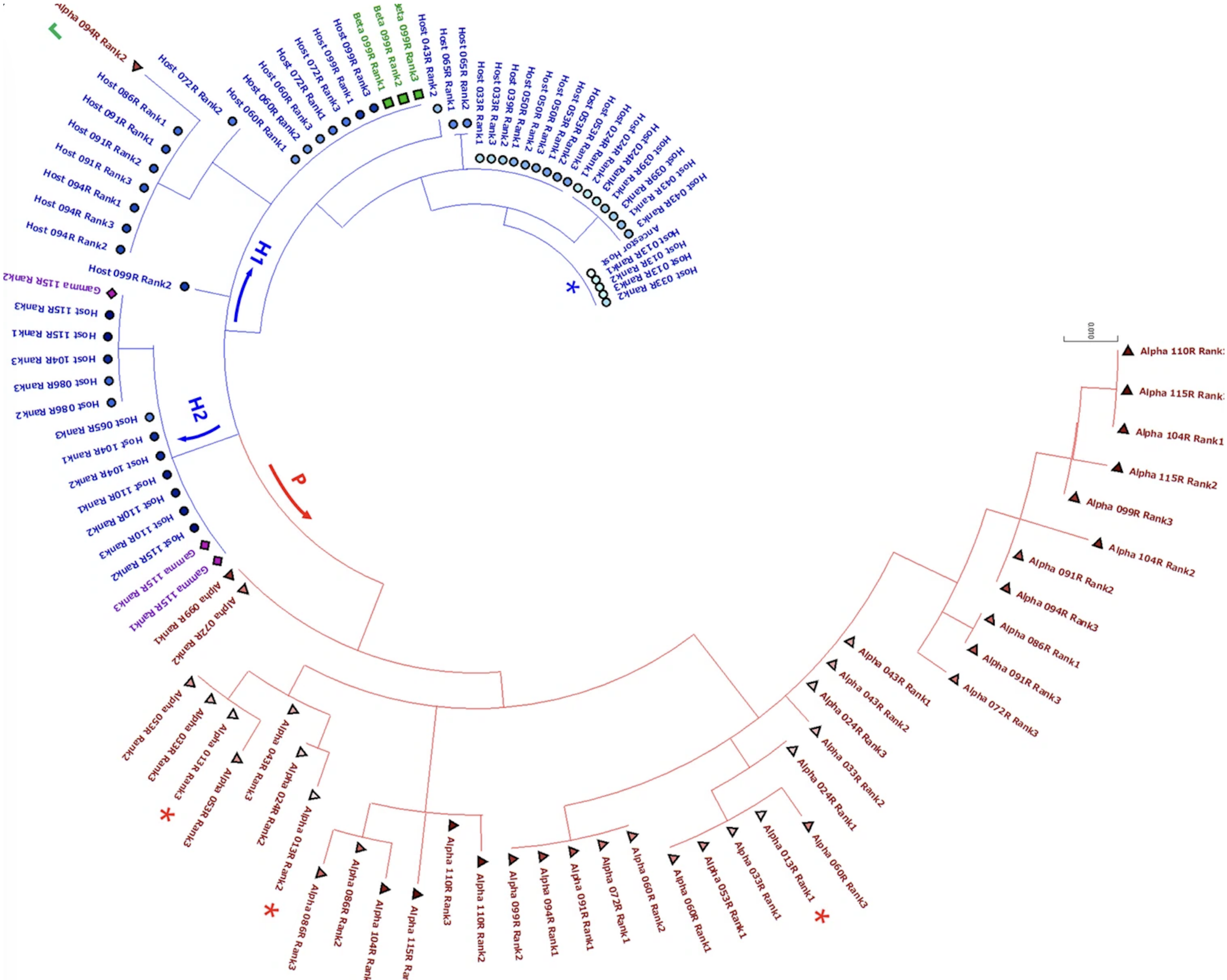


Somewhat similar to experimental RNA evolution Emergence and diversification of a host-parasite RNA ecosystem through Darwinian evolution Taro Furubayashi ....Norikazu Ichihashi 2020



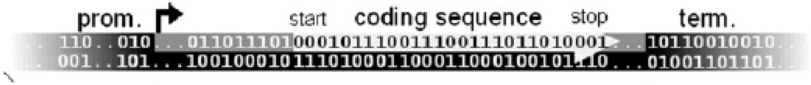
One parasitic lineage; others arise by single deletion mutations from hosts, to be less ”recognisable’

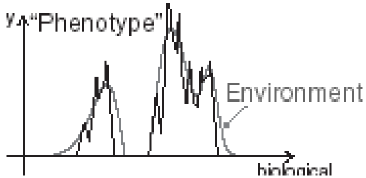

A
Host
Parasite-α
Parasite-β
Parasite-γ
RNA (nM)
10^4
10^3
10^2
10^1
10^0
10^-1
10^-2
10^-3
0
10
20
30
40
50
60
70
80
90
100
110
120
Round
B
M-site stem-loop
Host
(~2040 nt)
β-subunit gene
Parasite-α
(~220 nt)
Parasite-β
(~1070 nt)
Parasite-γ
(~510 nt)
C
Host-115
Host-99
Host-0
D
Parasite-α13
E
Parasite-β99
Parasite-γ115


Alpha 094R Rank2
Host 072R Rank2
Host 086R Rank1
Host 091R Rank1
Host 091R Rank2
Host 091R Rank3
Host 094R Rank1
Host 094R Rank3
Host 094R Rank2
Host 060R Rank1
Host 060R Rank2
Host 060R Rank3
Host 072R Rank1
Host 072R Rank3
Host 099R Rank1
Host 099R Rank3
Beta 099R Rank1
Beta 099R Rank2
eta 099R Rank3
Host 043R Rank 2
Host 065R Rank1
Host 065R Rank2
Host 033R Rank1
Host 033R Rank3
Host 039R Rank2
Host 039R Rank1
Host 050R Rank2
Host 050R Rank3
Host 050R Rank1
Host 053R Rank2
Host 053R Rank3
Host 053R Rank1
Host 024R Rank2
Host 024R Rank3
Host 024R Rank1
Host 039R Rank3
Host 043R Rank1
Host 043R Rank3
Ancestor Host
Host 013R Rank1
Host 013R Rank3
Host 013R Rank2
Host 033R Rank2
*
H1
Host 099R Rank2
Gamma 115R Rank2
Host 115R Rank3
Host 115R Rank1
Host 104R Rank3
Host 086R Rank3
Host 086R Rank2
H2
Host 065R Rank3
Host 104R Rank1
Host 104R Rank2
Host 110R Rank1
Host 110R Rank2
Host 110R Rank3
Host 115R Rank2
Gamma 115R Rank3
Gamma 115R Rank1
Alpha 099R Rank1
Alpha 072R Rank2
P
0.010
Alpha 110R Rank
Alpha 115R Rank
Alpha 104R Rank1
Alpha 115R Rank2
Alpha 099R Rank3
Alpha 104R Rank2
Alpha 091R Rank2
Alpha 094R Rank3
Alpha 086R Rank1
Alpha 091R Rank3
Alpha 072R Rank3
Alpha 043R Rank1
Alpha 043R Rank2
Alpha 024R Rank3
Alpha 033R Rank2
Alpha 024R Rank1
Alpha 060R Rank3
Alpha 013R Rank1
Alpha 033R Rank1
Alpha 053R Rank1
Alpha 060R Rank1
*
Alpha 060R Rank2
Alpha 072R Rank1
Alpha 091R Rank1
Alpha 094R Rank1
Alpha 099R Rank2
Alpha 110R Rank2
Alpha 110R Rank3
Alpha 115R Ran
Alpha 104R Rank
Alpha 086R Rank2
Alpha 086R Rank3
*
Alpha 013R Rank2
Alpha 024R Rank2
Alpha 043R Rank3
Alpha 053R Rank3
*
Alpha 013R Rank3
Alpha 033R Rank3
Alpha 053R Rank2

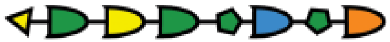
GENOME EVOLUTION: 3 (4) modeling frameworks genome structure and genotype to phenotype/fitnessmap

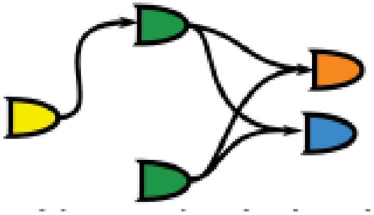
AEVOL bit (nucleotide) level coding of genome allows evolution of new genes!

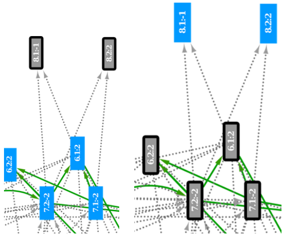




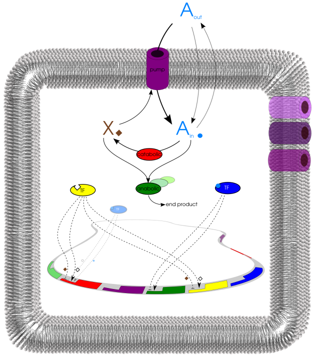
PoaS course grained genome allows multilevel (evolving) GPf mapping

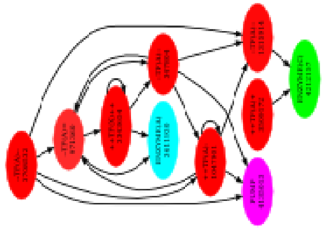






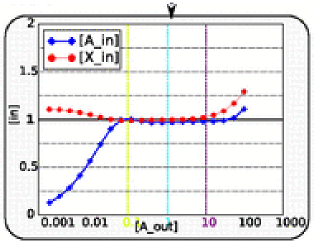
PoaS + metabolism (Virtual cell) allows *indirect* sensing environment





$$\begin{pmatrix} \frac{d[A]}{dt} = \frac{[A]_{out}[X]Vmax_p[Prot]_p}{([A]_{out}+KA_p)([X]+KX_p)} & (2) \\ \frac{d[X]}{dt} = \frac{-d[A]}{dt} & (3) \end{pmatrix}$$

$$\begin{pmatrix} \frac{d[A]}{dt} = \frac{-[Prot]_c[A]Vmax_c}{[A]+KA_c} & (4) \\ \frac{d[X]}{dt} = \frac{-d[A]}{dt}\ E_{corr} & (5) \end{pmatrix}$$



Evolution of (observed) structure in Gene Regulatory Networks (GRN) role of mutational operators

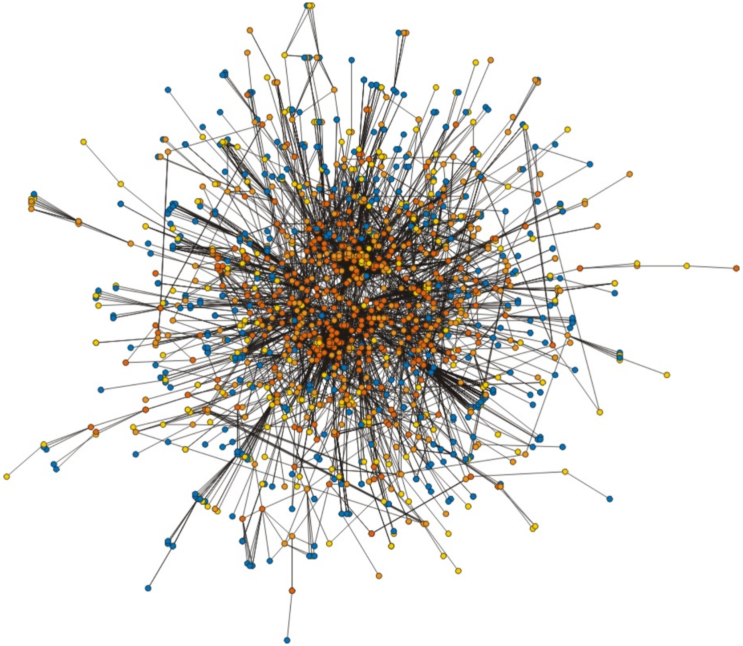
Structural features of transcription regulation networks

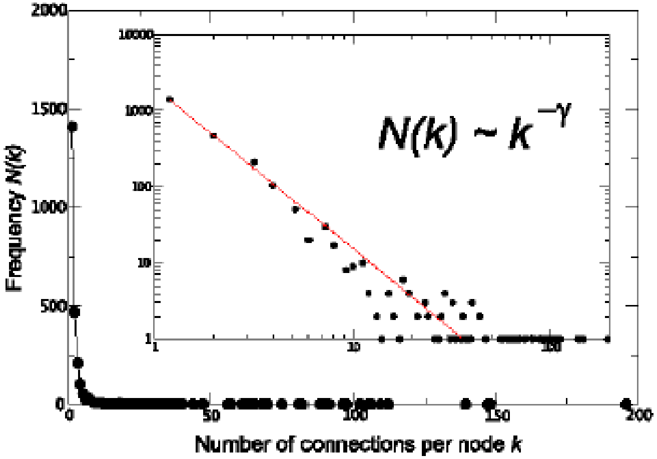
powerlaw and FFL

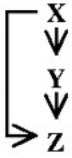
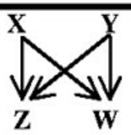
- Characterizing topology of GRN
- *What do we mean with "over-representation of..."*
- *Random mutations = / = randomization*

Importance of coding structure

Observed properties of the GRN of Yeast





Network	Nodes	Edges	N_{real}	N_{rand} ± SD	Z score	N_{real}	N_{rand} ± SD	Z score
Gene regulation (transcription)					**Feed-forward loop**			**Bi-fan**
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13
*S. cerevisiae**	685	1,052	70	11 ± 4	14	1812	300 ± 40	41

(also many other networks: neural networks, computer networks, (but not Eco-networks))

Milo & Alon 2002

Genomic encoding of GRN
Modeling Mutational Dynamics

bag of genes with binding sites (BS)

BS deletion 8x10^{-3}

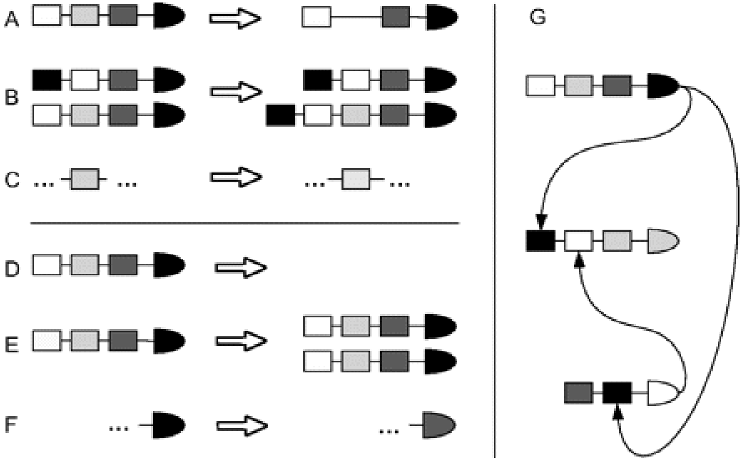
BS duplication 8x10^{-3}

Bs mutation 8x10^{-4}

Gene deletion 1x10^{-3}

Whole gene duplication $1x10^{-3}$

Protein divergence/innovation 5x10^{-3}

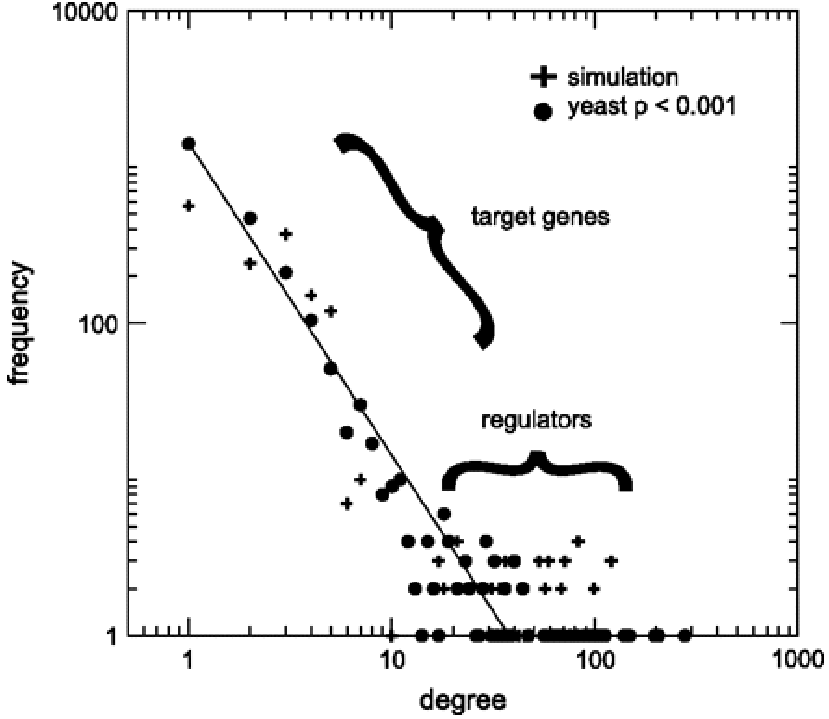


Parameters loosely chosen from literature, NOT FITTED.

Feed-Forward Loop Circuits as a Side Effect of Genome Evolution

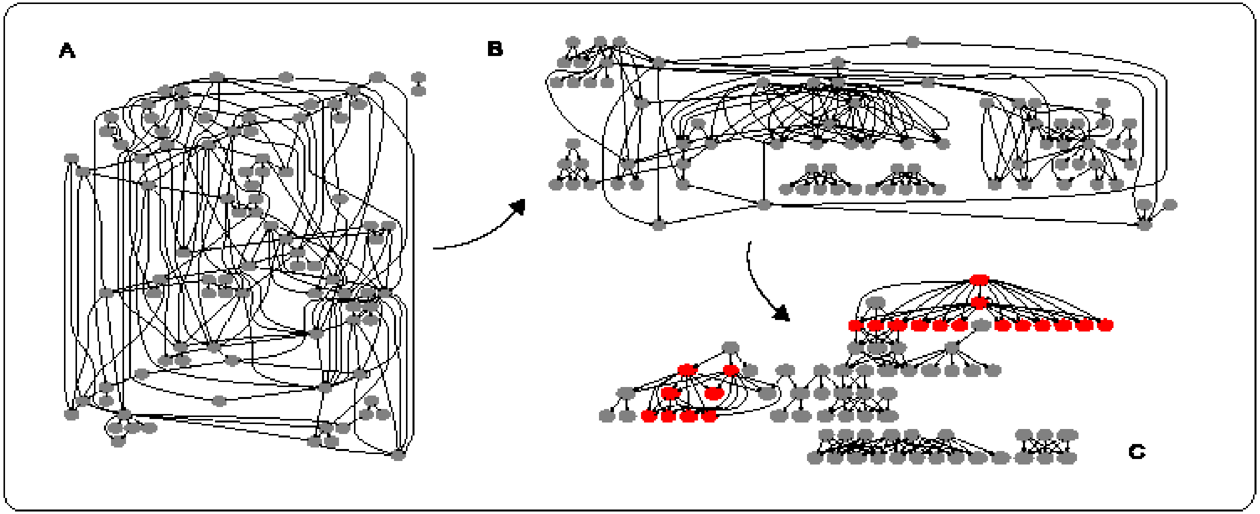
Otto X. Cordero, Paulien Hogeweg MBE 2006

mutational dynamics WITHOUT selection leads to Powerlaw distribution of connections with similar γ as Yeast GRN for similar number of genes and TF (2000, 100)



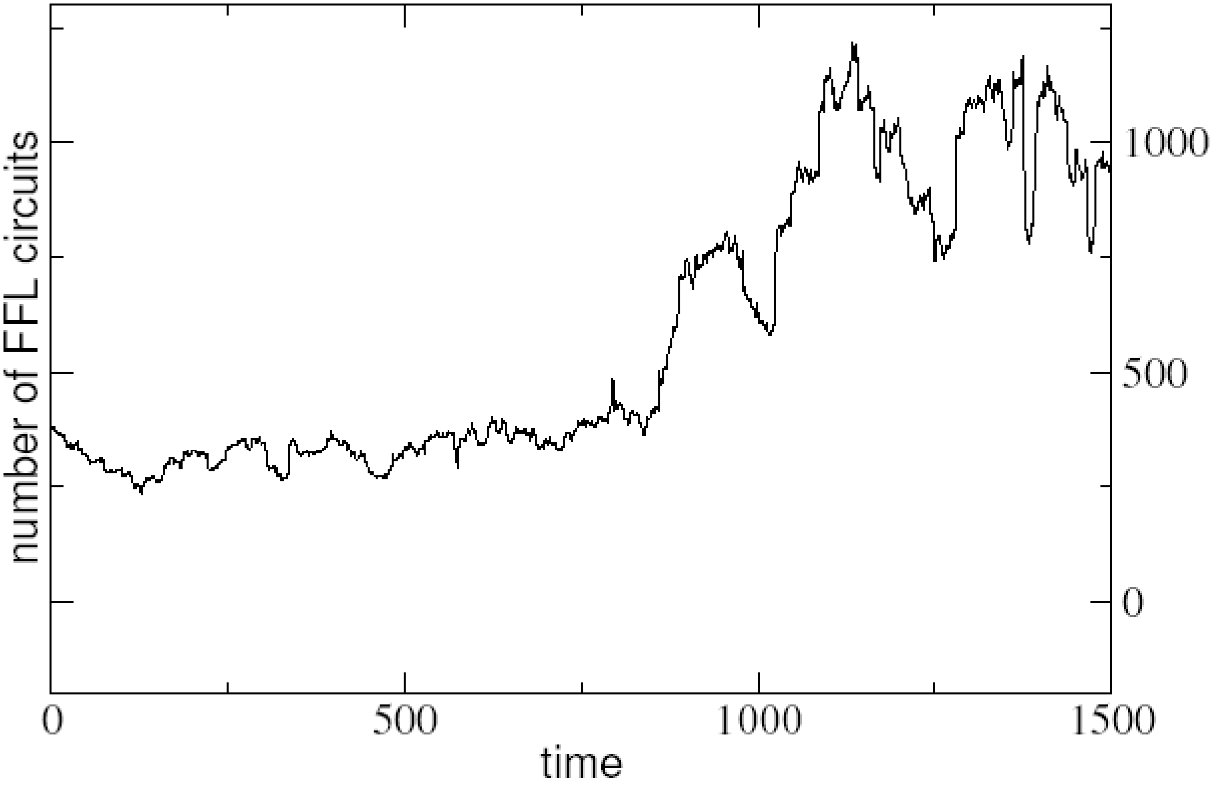
Toy model: Visualization of network restructuring during neutral evolution (hierarchical) structured network for free!

- Visualization of the network evolution (toy example):



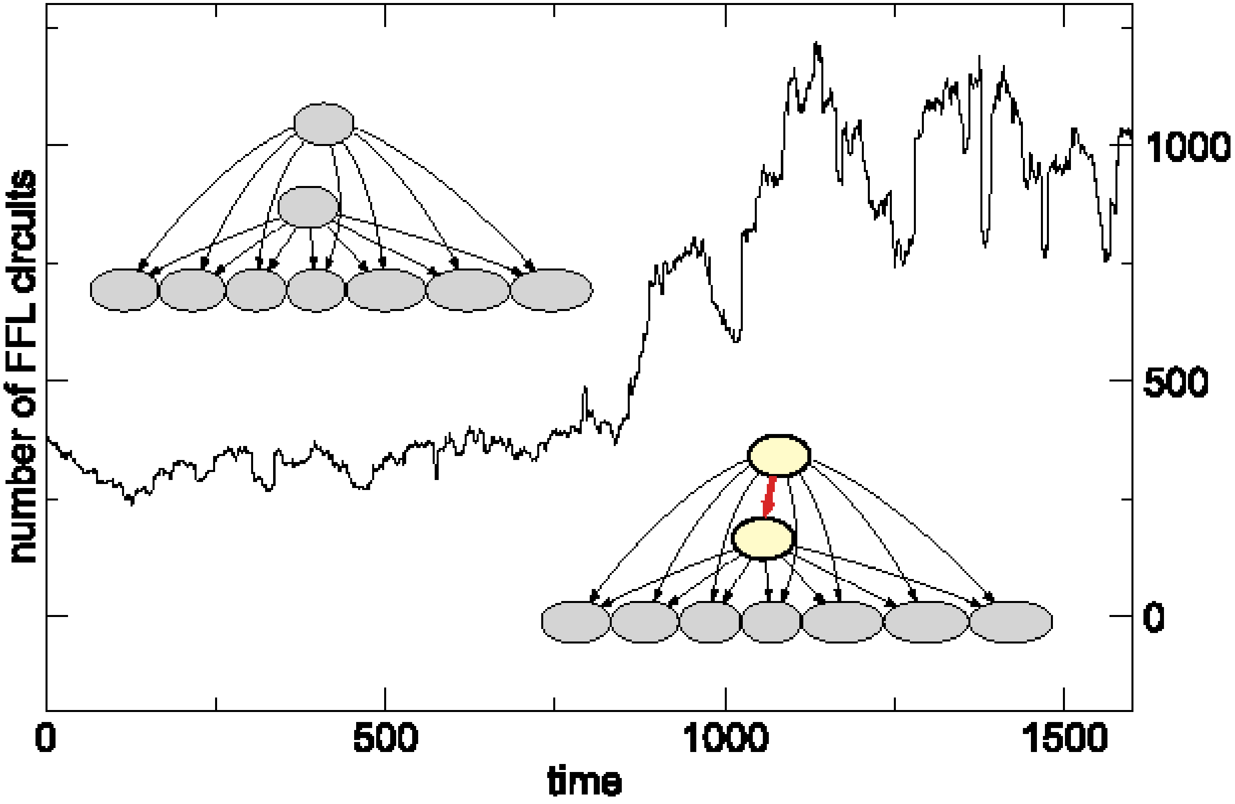
During evolution Sudden increase of FFL motifs: FFL as mutational signature

- Results of mutational dynamics at the microlevel:



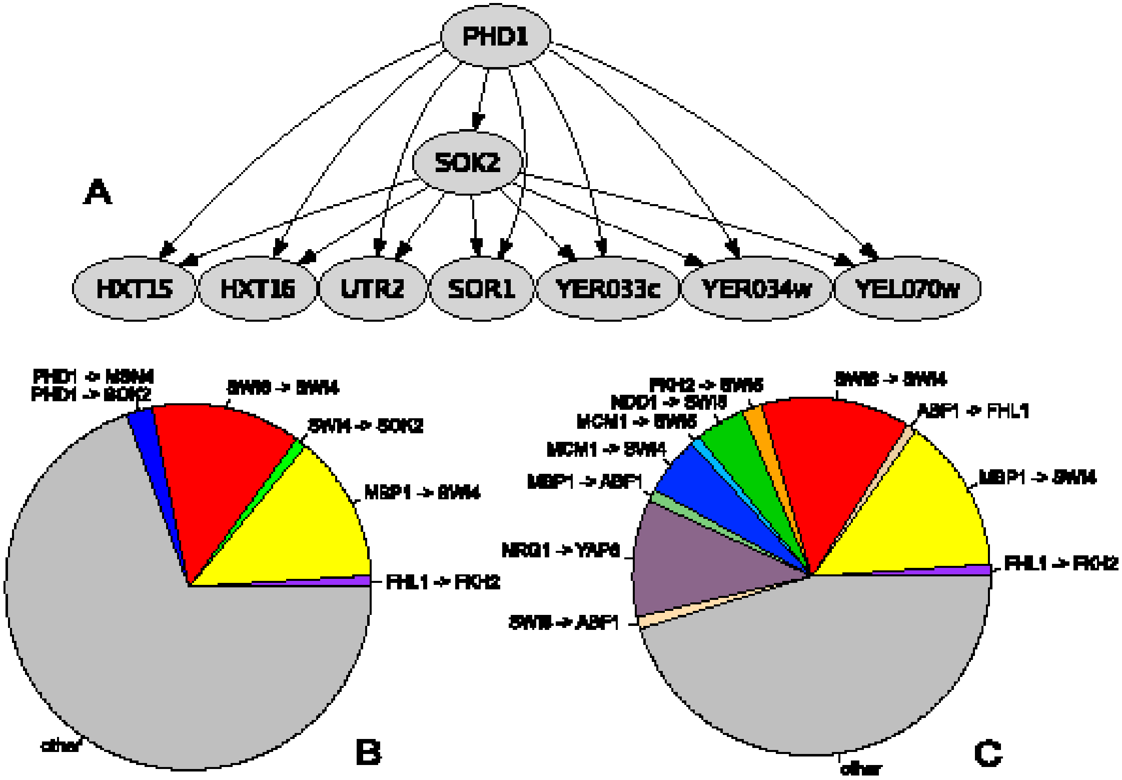
large increase of FFL motifs: Originate through duplication of hub + new connection

- Mechanics of massive FFL formation:



Over-representation of FFL motifs in Yeast: Duplication + connection of Hub genes

- Evidence in the yeast network:

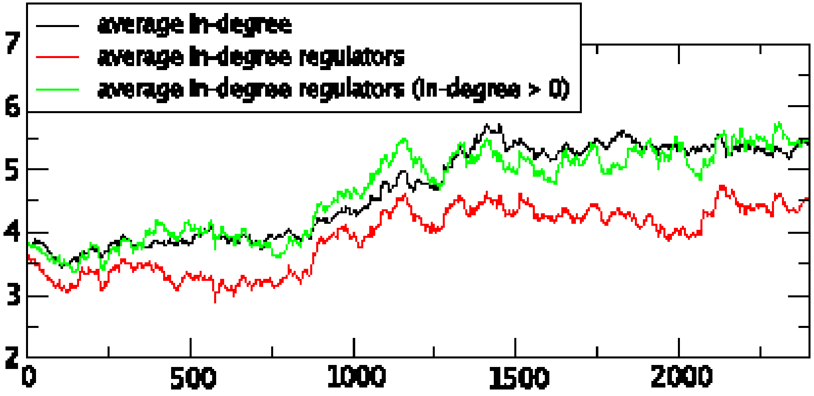


random mutations vs randomization

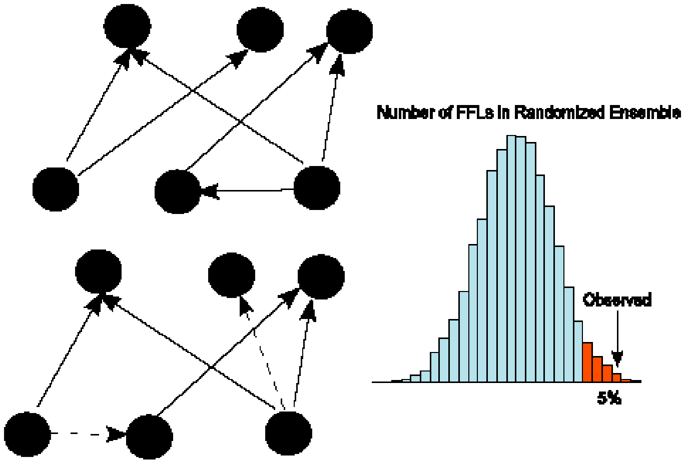
Randomization tests: keep everything the same EXCEPT feature to be tested

here: keep degree distribution
— test for FFL

BUT random mutations
do not conserve degree distribution



- Randomization test: swapping connections.

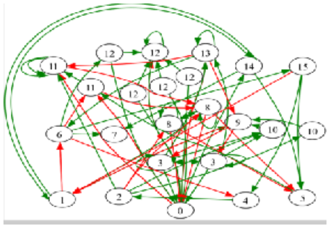


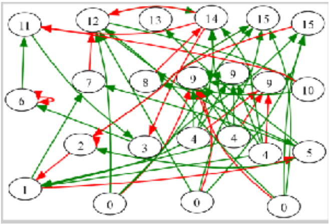
discussion/conclusions
Assessing relevance of observed evolved network structure in models or “real” life

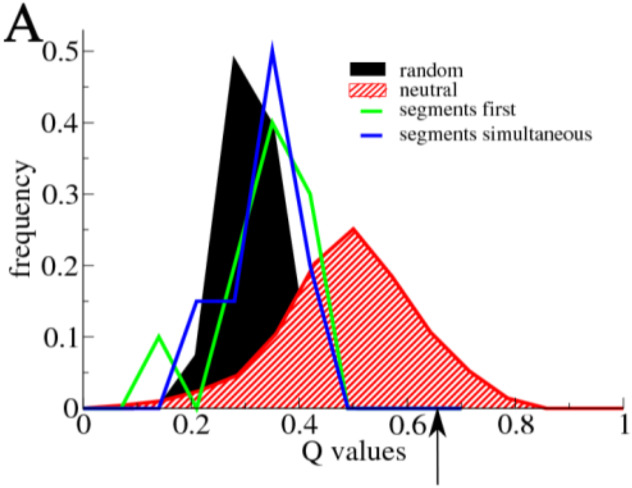
Random mutations lead to non-random structure
With or without selection of “something”

Random mutations as stochastic dynamical system
goes to attractor

Example: Modularity in evolved GRN model
Drosophila segmentation/differentiation







average Q value for
modular networks

ten Tusscher & Hogeweg 2011: Evolution of Networks for Body Plan Patterning; Interplay of Modularity, Robustness and Evolvability

random networks as null model?