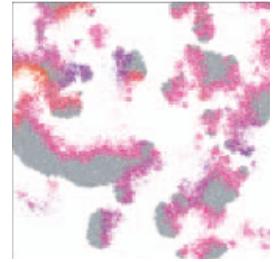
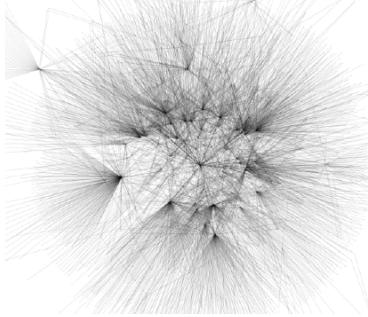


1

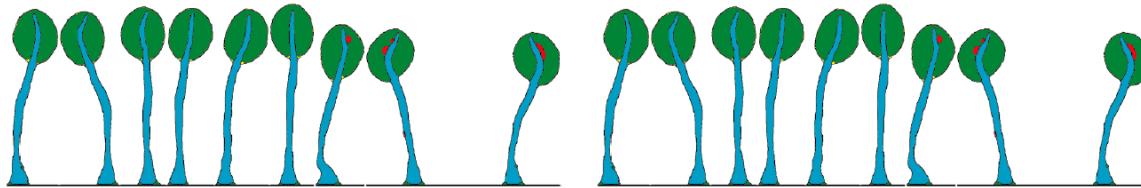
Modeling Biocomplexity: overview - review

challenges, themes , insights and methodology



Modeling Biocomplexity

interactions among multiple space and
time scales



Modeling Biocomplexity

What can/cannot be done (sensibly)

Models should be simple *ENOUGH*
(parameters, understanding)

Models should be as simple as possible but not simpler
Einstein

Empty cross-section?
**develop modeling approaches such that
modeling biocomplexity becomes feasible**

exploit Selforganization (maximize flexibility)

exploit evolution (generic-nongeneric properties)

alternative caricatures



A conversation with Paulien Hogeweg, Utrecht, NL

“Thinking in the Most Interesting Simplification”

some defining properties of biocomplexity:

- Local interactions, many different entities in small numbers
spatial individual based modeling
- 'leaky' multiple levels of organization (feedbacks)
self-organizing (recognition!)
(partially) predefined (*CPM*))
- interlocking time scales
 - ecology, regulation and evolution
- evolved (evolving) systems
neutrality
not 'simplest' implementation
evolutionary signatures (evolvability)

Investigate these properties

**evolution of complexity, how?/why?/when?
classical models do not explain it (only reprod. rate)
information threshold**



“Elephant, I believe you got about by random mutations”

Evolution of 'complexity' and 'novelty'

'Elephant I believe you got about by random mutations'

NOT pop. genetics, RNA evolution; *NOT* automatic

Maynard Smith and Sathmary

Bioinformatics

What DID happen

七

Darwinian Selection + Chemistry

Darwinian selection

+

LOCAL interactions

Intermediate steps

What DOES happen

+

Common principles

MU & Szathmary 1997

MULTILEVEL EVOLUTION

Table 1.2 The major transitions

Replicating molecules	→	Populations of molecules in compartments
Independent replicators	→	Chromosomes
RNA as gene and enzyme	→	DNA + protein (genetic code)
Prokaryotes	→	Eukaryotes
Asexual clones	→	Sexual populations
Protists	→	Animals, plants, fungi (cell differentiation)
Solitary individuals	→	Colonies (non-reproductive castes)
Primate societies	→	Human societies (language)

Table 1.3 Conflict between selection at different levels

Form of cooperation	Exceptions
A fair meiosis	Meiotic drive, transposition
Sexual reproduction	Parthenogenesis
Differentiation of somatic cells	Escape from growth control
Non-reproductive castes of social insects	Egg-laying worker bees

MS-S: Processes in evolution of complexity common principles

- selfsufficient to “part of a whole”
(e.g. symbiogenesis)
Multiple levels + conflicts
- Division of labor
DNA to DNA/RNA world, germline soma, social insects,
human societies
- limited to unlimited inheritability
attractor-based vs storage based (information based) inher-
itability

“levels” of multilevel evolution

- replicators and self-organized (replicating) spatial patterns
automatic consequence of local interactions
- replicators in protocells (coupling of dynamics)
- cells with dupdels/LCR/plasmids/transposons..
- HGT: cell level vs gene level replication/selection
- multilevel genotype-phenotype mapping
evolution of coding structure (from RNA to Vmicrobes)

and combinations...

Time scales

Maynard Smith: methodological assumption/requirement

We cannot hope to explain these transitions in terms of the ultimate benefits they conferred. For example, it may be that, in the long run, the most important difference between prokaryotes and eukaryotes is that the latter evolved a mechanism for chromosome segregation at cell division that permits DNA replication to start simultaneously at many origins, whereas prokaryotes have only a single origin of replication. At the very least, this was a necessary precondition for the subsequent increase in DNA content, without which complexity could not increase. But this is not the reason why the change occurred in the first place: as we explain in Chapter 6, the new segregation mechanism was forced on the early eukaryotes by the loss of a rigid cell wall, which plays a crucial role in the segregation of prokaryotic chromosomes. Or to take a second example, meiotic sex was an important preadaptation for the subsequent evolution of radiation of the eukaryotes, but it could not have originated for that reason.

The transitions must be explained in terms of immediate selective advantage to individual replicators: we are committed to the gene-centred approach outlined by Williams (1966), and made still more explicit by Dawkins (1976). There is, in fact, one feature of the transitions listed in Table 1.2 that leads to this conclusion. At some point in the life cycle, there is only one copy, or very few copies, of

HOWEVER:

Multilevel evolution --> multiple timescales

long term information integration

- “immediate” benefits - for whom? (cf death rate)
- benefits vs neutrality and side-effects
(cf presence before beneficial / loss of genes)
- Genome expansion - long term evolutionary benefits
- * fitness as a time dependent function
cf Savill & Hogeweg)
- * Sparse fitness evaluation and information integration

Predefined vs “emergent” fitness

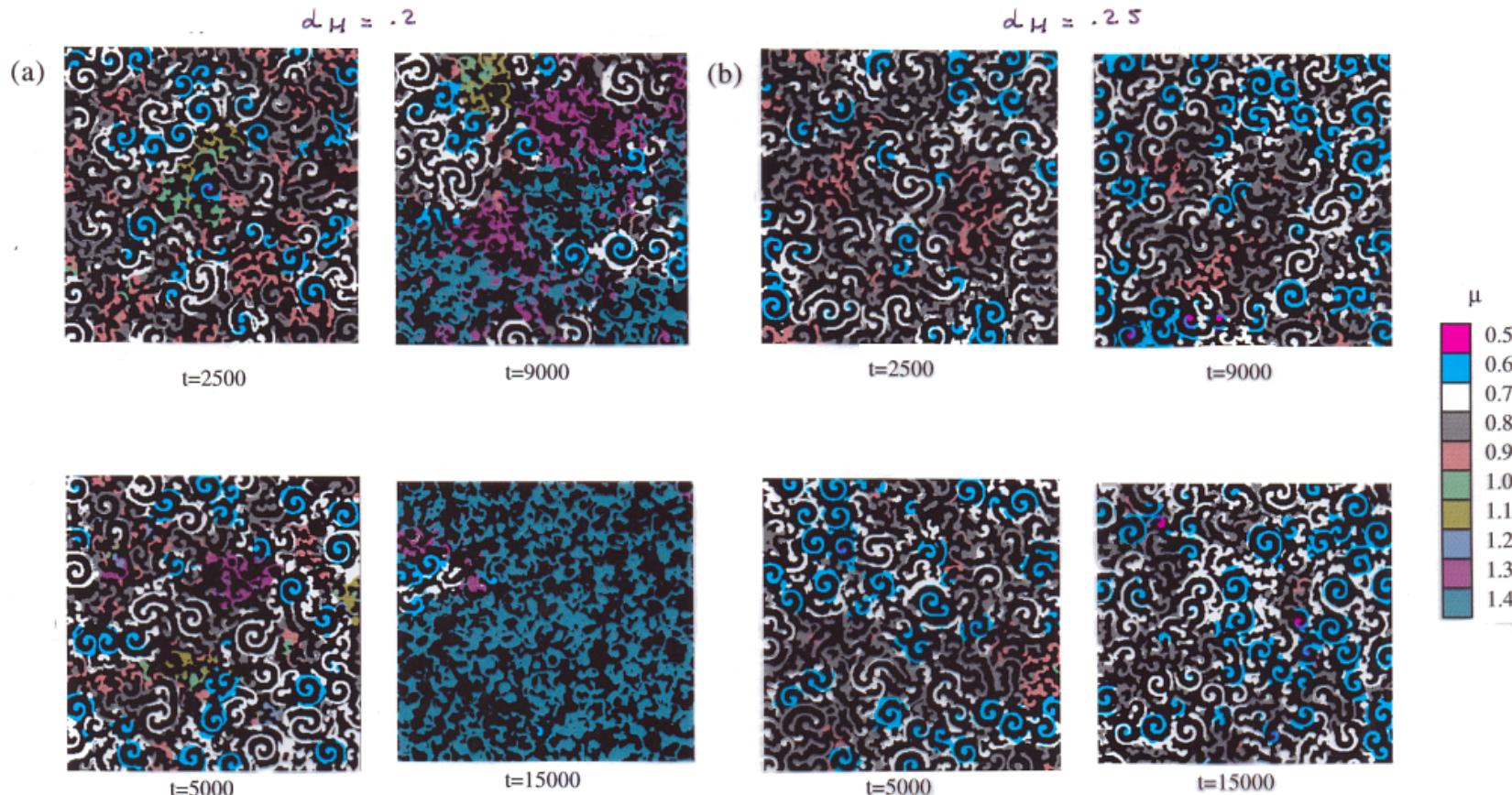
Fitness as Time-dependent function

Nicholson Baily Host-parasitoid Lattice Map model; evolution of directed parasitoid migration (β) (Savill et al JTB 1997)

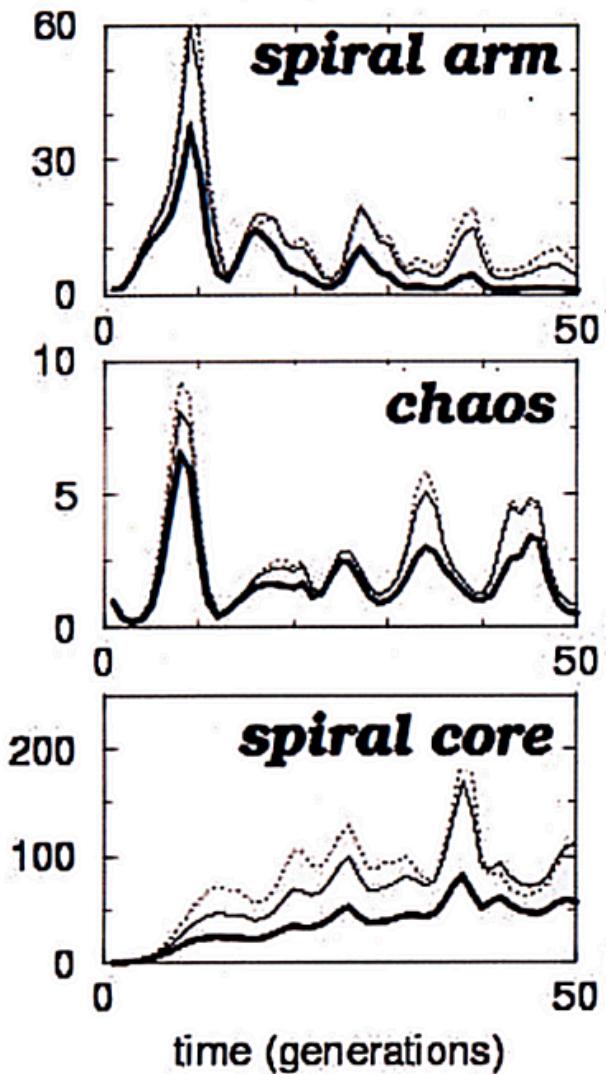
- 3 levels of selection:
 - host / parasitoids
 - spiral waves / chaotic waves
 - regions of spiral waves and chaotic waves
- 'Life history' of spiral
 - 'born' at edge of spiral and chaotic region with high β values
 - increases domain (faster rotation)
 - decreases differential migration (β) (core as germline)
 - because of lower β shrinks and dies
- outcome depends on spiral birth/death rate
 - (which depends on host migration)
- **Multiple timescales**



self reinforcing spatial patterns



fitness as function of time



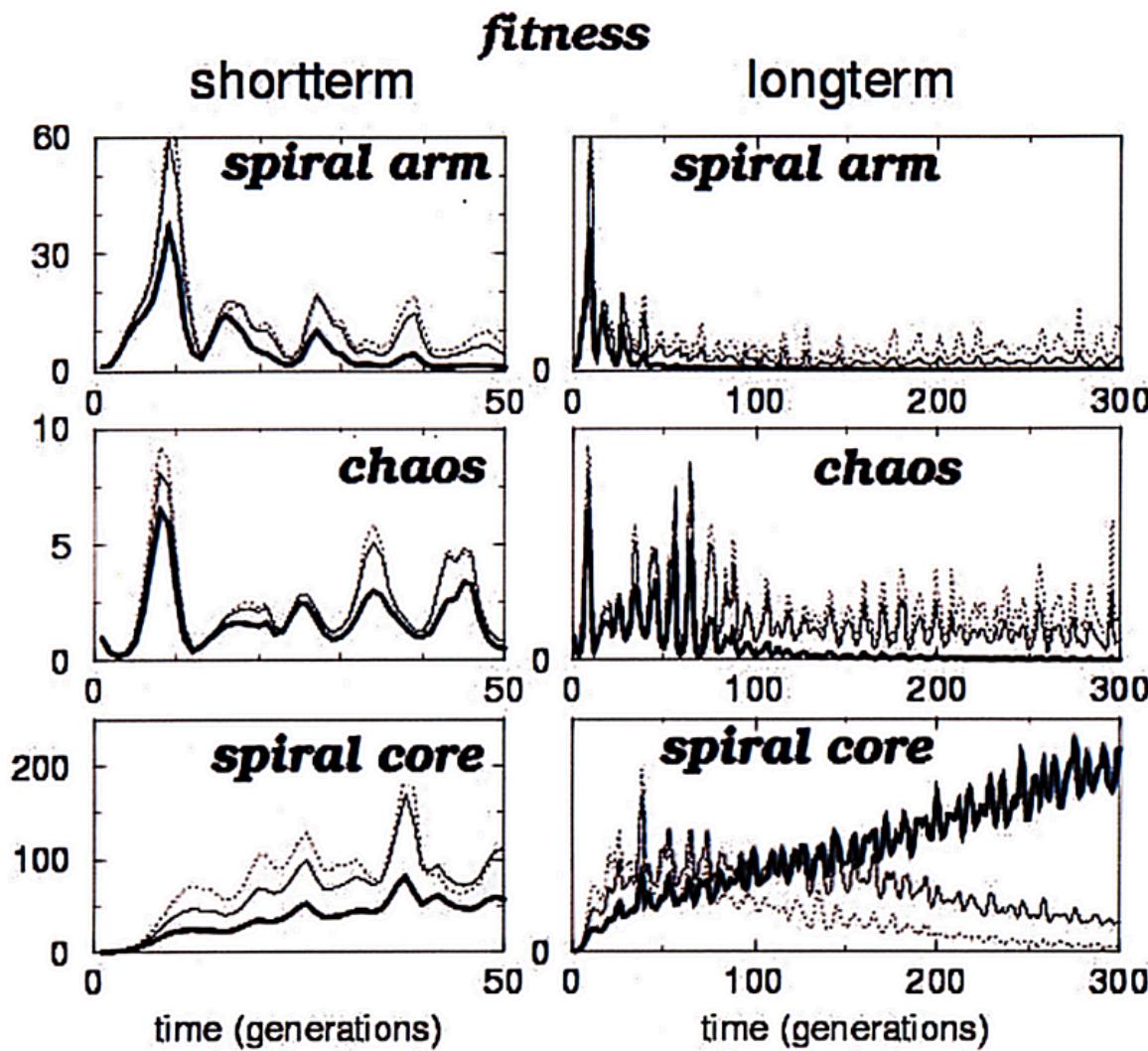
fitness as # offspring

through time

from 1 type at 1 location

to anywhere (any type)

fitness as function of time



Simple case of individual vs population based diversity

RM system: bacteria, plasmids, phages

“Bag of plasmids”

Restriction modification systems

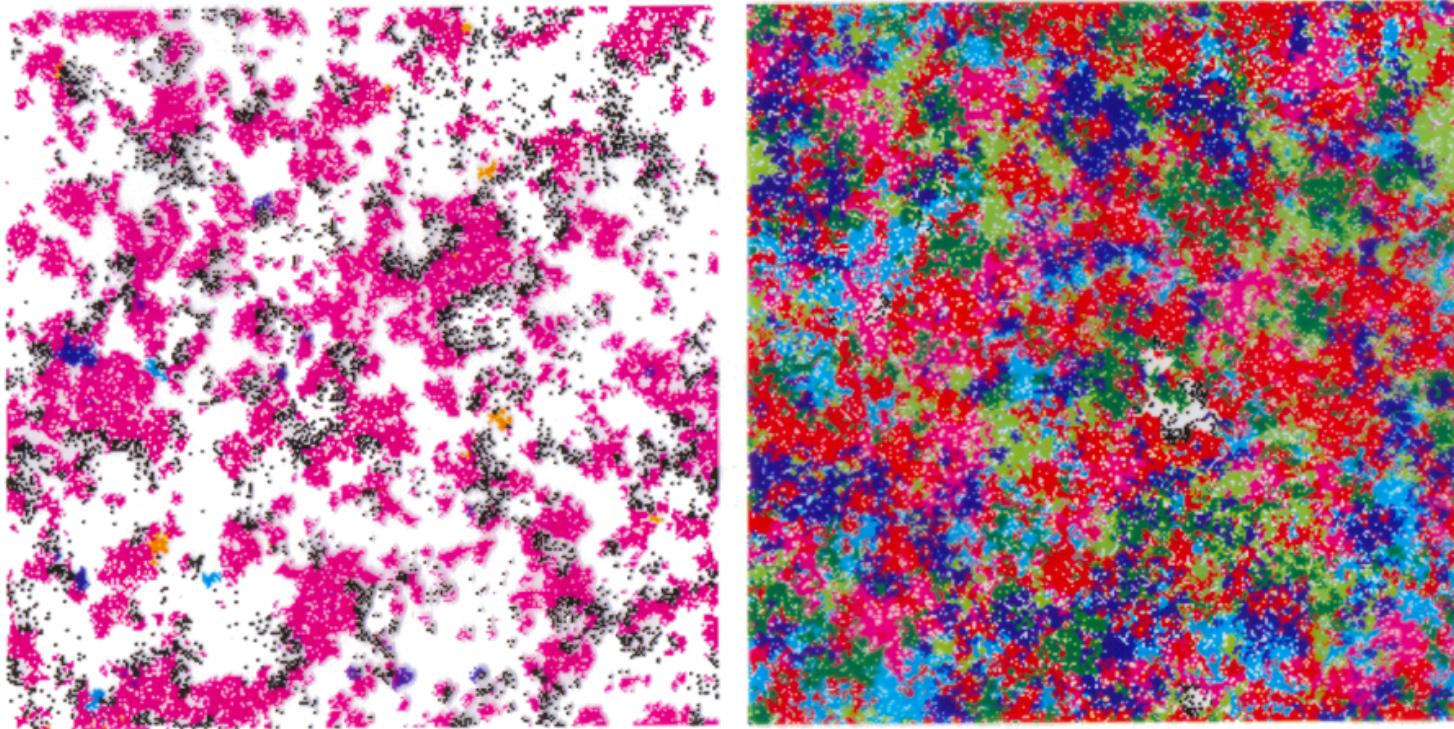
on plasmids

Cut DNA on specific sequence,
methylate DNA so that not cut

HGT

Function(?)

antiviral defense



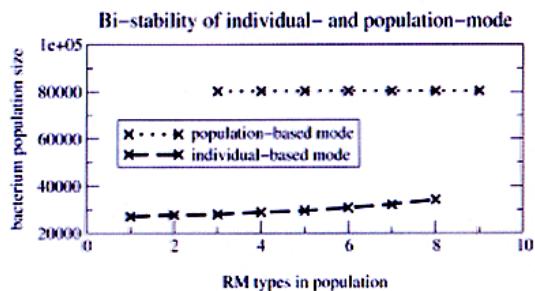
Two alternative attractors

Individual based diversity': waves of virus infection
population based diversity: static pattern, few phages

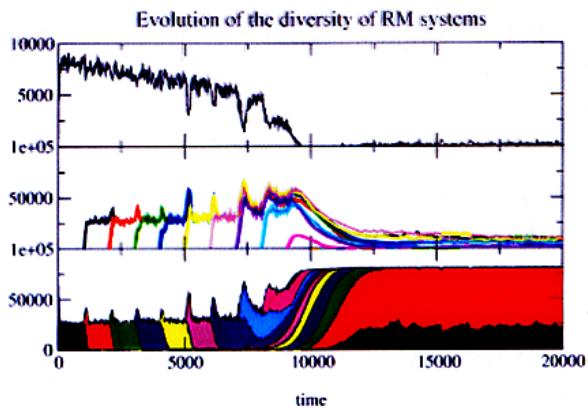
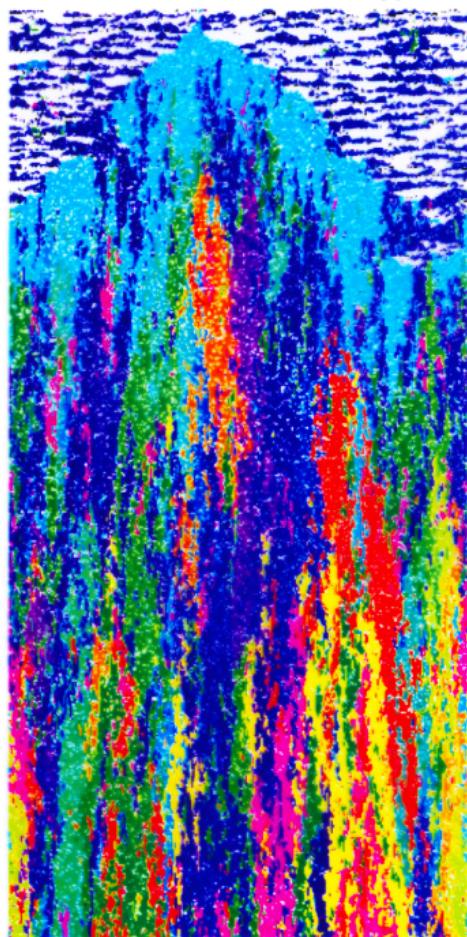
“smart” bacteria benefits only the plasmids and the viruses

transition from individual based to population based diversity

RM systems: from individual diversity to population diversity



Types



bacteria RM systems plagues



conclusion RM systems

First: Large “genomes”

Later: Reduction of Genomes

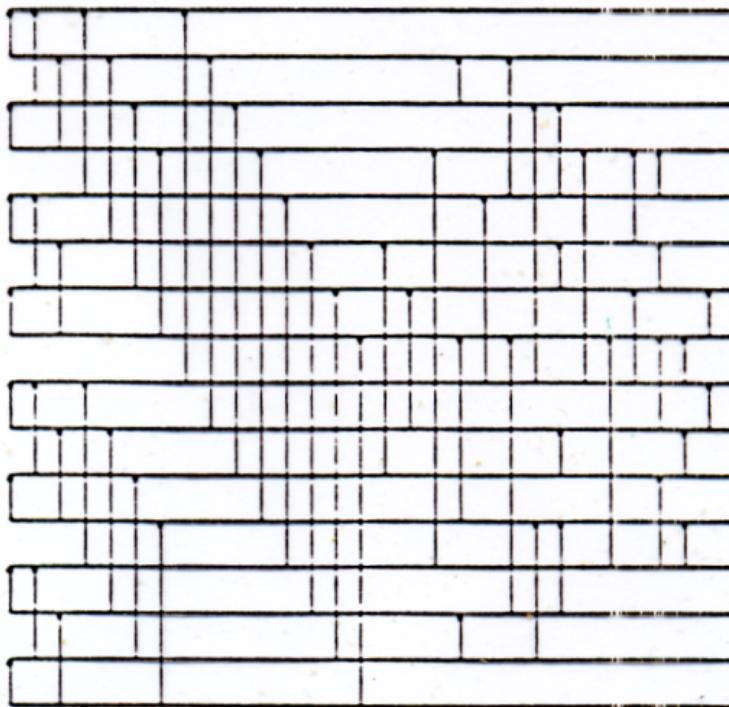
Decline of Virus population

Cycle

beneficial for cells vs viruses.

Sparse fitness evaluation as Optimization Strategy (Hillis)

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



Green's Sorter

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

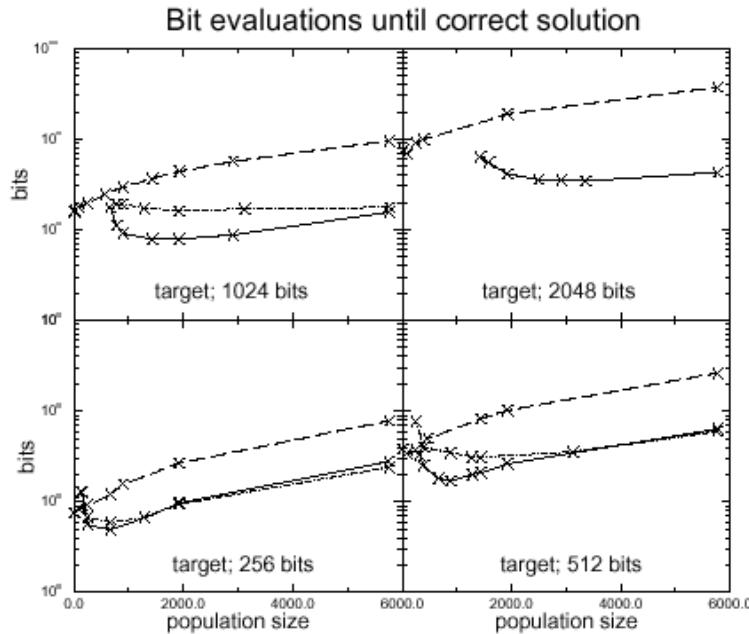
**Function fitting as model for
coevolution as optimization strategy
sparse fitness evaluation through co-evolving problems
cf Pagie and Hogeweg 1997**

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)

Information integration sparse fitness evaluation more efficient (< number of evaluations)

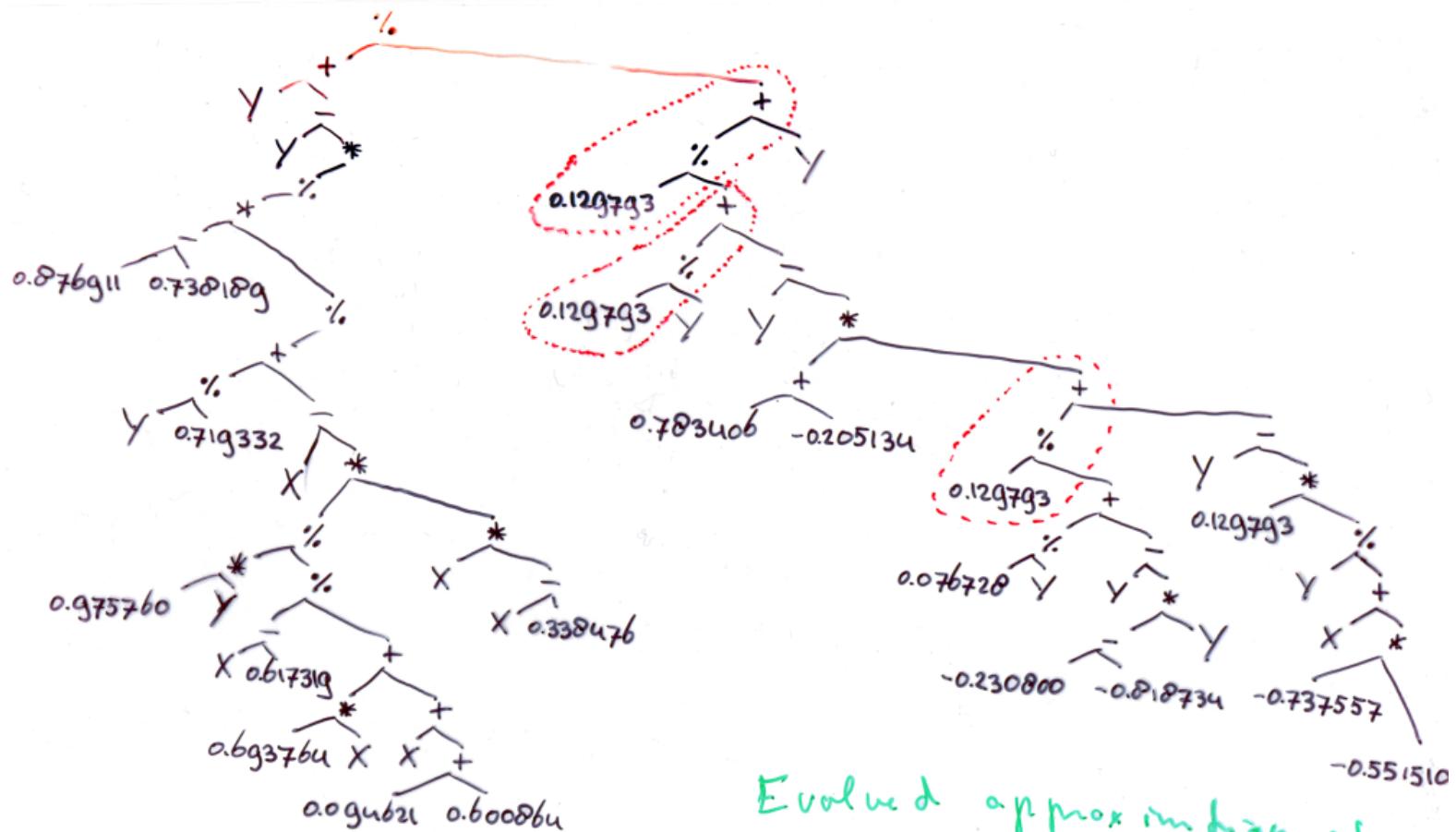
Linear model: bitstring match



Individual problem solving Type of solution: complexity, generalizable, 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

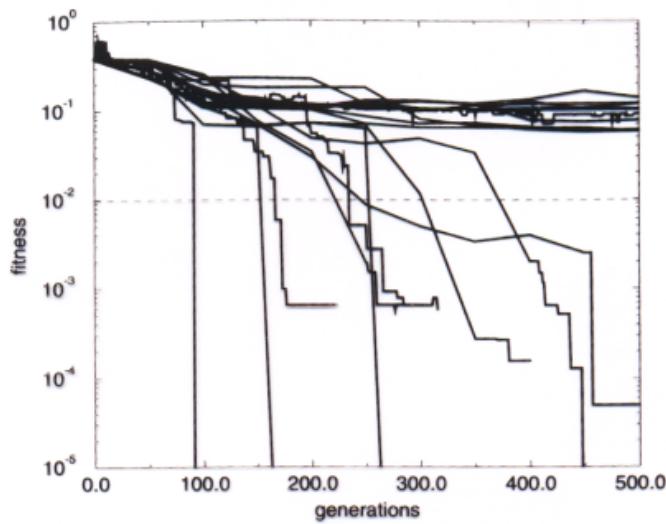
sparse fitness evolution
better fit (distance)
"better fit (simpler function)
"better fit (more generalizable)
LOWER mutational robustness



Evolved approximation of

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

Coevolving sparse fitness evaluation



Complete static fitness evaluation

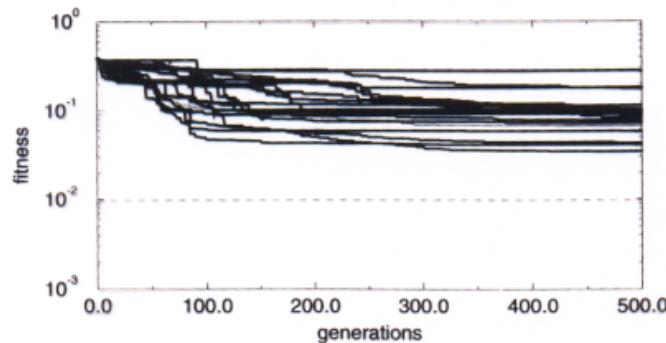


Figure 2. Fitness curves of the best-of-generation solution for coevolving (a) and complete static problem evaluation (b). Fitness is based on the complete problem set that consists of 26×26 problems. The fitness curves that drop below 10^{-5} go to values between 10^{-15} and 10^{-17} . The horizontal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

Looking at solutions

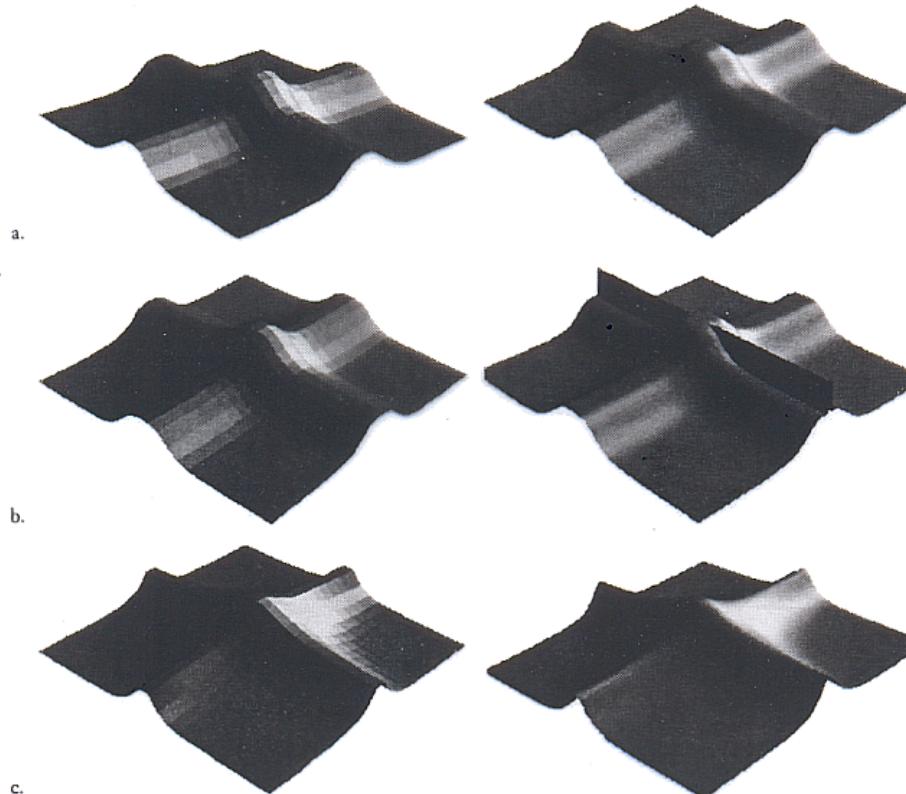


Figure 3. Three typical final solutions produced by coevolving fitness evaluation. The left plots are based on 26×26 evaluated problems, the right plots on 100×100 evaluated problems. Two correct solutions that approximate the target function are shown in (a) and (b); an incorrect solution is shown in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

Static vs. sparse fitness evaluation (unseen data)

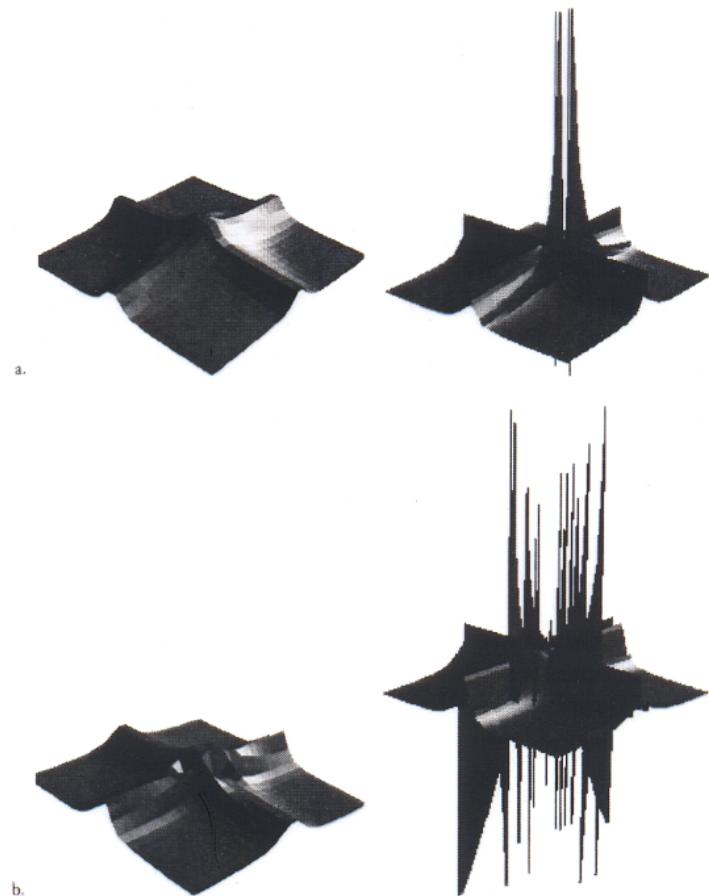


Figure 4. Two typical final solutions produced by static fitness evaluation. The left plots are based on 26×26 evaluated problems, the right plots on 100×100 evaluated problems. Neither solution is correct.

Mutational Stability

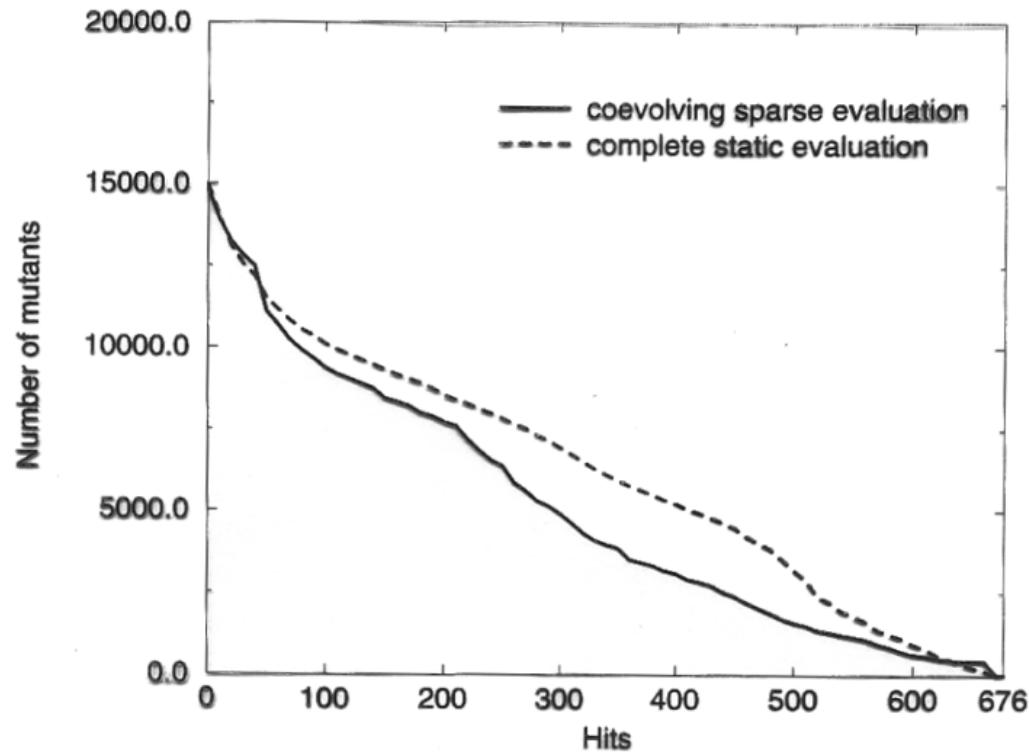


Figure 5. Histogram of the number of one-point mutants having at least x number of hits.

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

conclusion

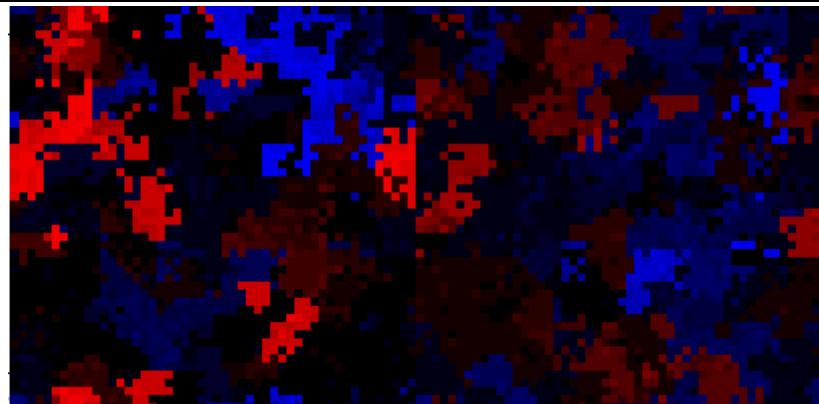
Sparse fitness evaluation benefits evolutionary adaptation

Sparse fitness evaluation: more generisable solutions

Sparse fitness evolution: less mutational robustness

long term information integration

Spatial pattern formation and PARASITE speciation



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

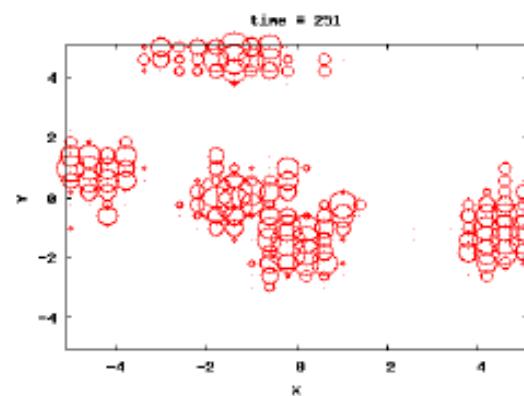
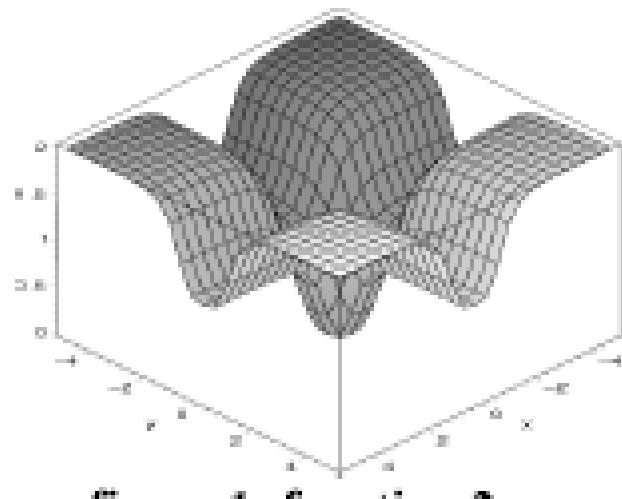
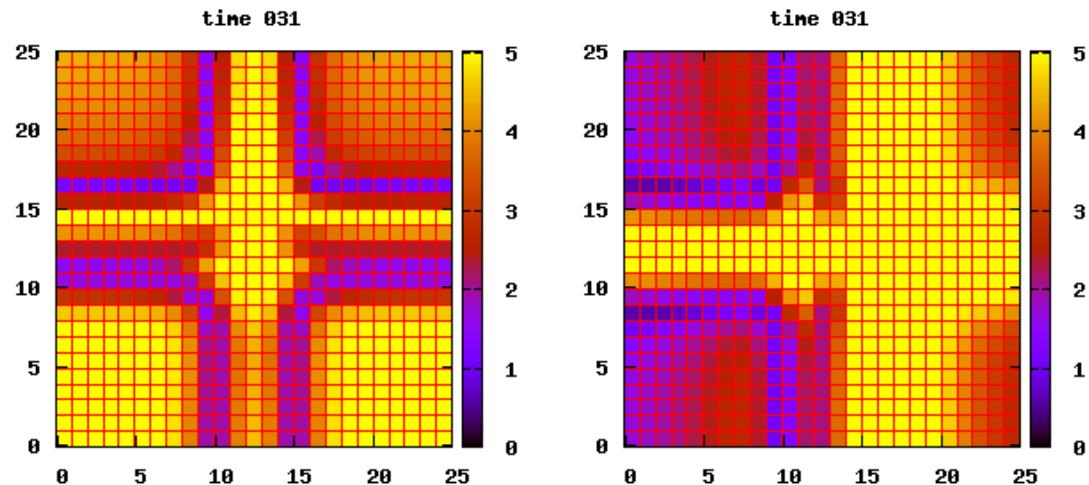


figure 5: parasite speciation in evaluating

Differentiation of host phenotypes: “good at eating different prey”



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

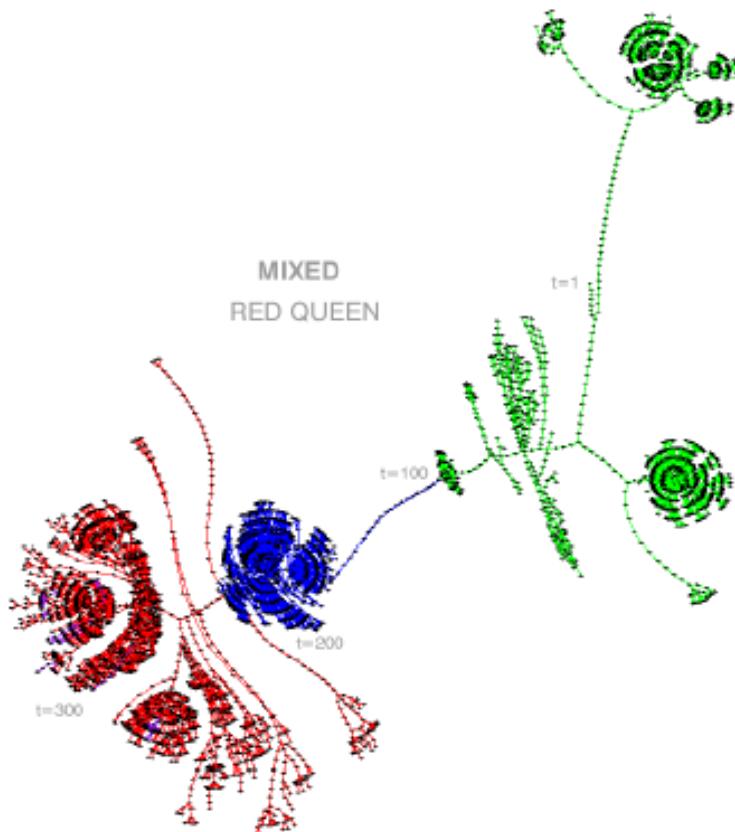


figure 6: combined ancestor trace in well mixed system

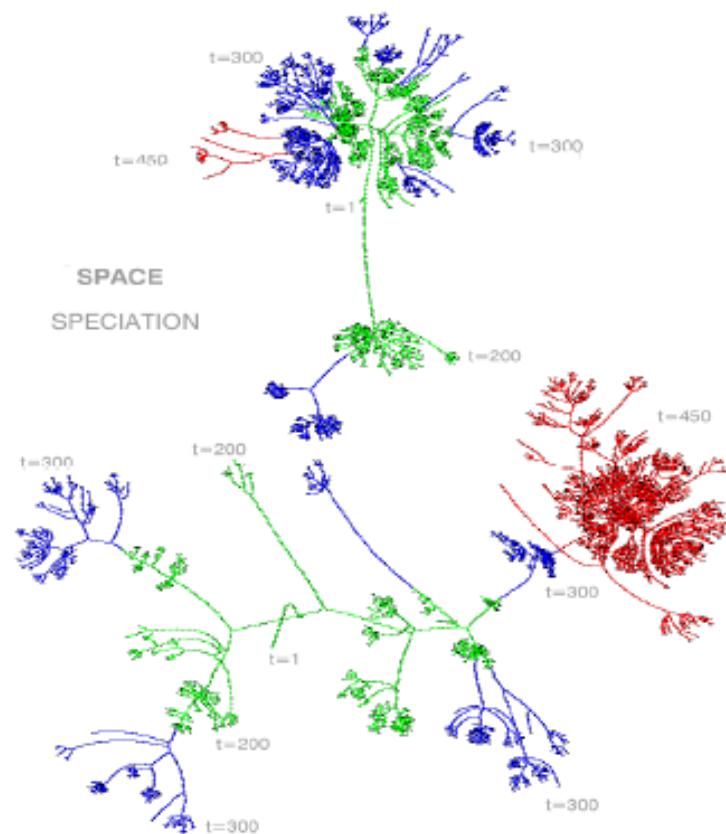


Figure 7: combined ancestor trace with pattern formation

conclusions

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

Sparse fitness evaluation:

better results

mutational less robust

more generalizable

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

smaller genome

“ giving up self-sufficiency” = exploiting opportunities

- Evolution of cross-feeding by mutational cascade in Vmicrobe
- Emergence of not selfsustaining cheaters/parasites in single replicator systems (population selfsufficientcy) --> more efficient selfreplicators of 'others'
- Local interactions lead to more global interdependencies through spatial pattern formation and the emergence of novel levels of selection (evolutionary interdependence)
- Conflicting multilevel selection and innovation (“smart solutions”)
- Evolution of DNA in RNA world

division of labor - non-reproducing “helpers”

- non-heritable phenotypic differentiation
TODO based behaviour / regulation Bumbles - Dicty
- division of labor in quasispecies (mutational decoded division of labor)
- All offspring from core of spiral (location decoded) (compare germline and soma)
- Individual and/vs ecosystem based ‘problem solving’
(cf original purpose of hypercycles)
- Symmetry breaking of plus/minus strands of complementary replicators
“proto-genome” (**side-effects** viability at larger volumes/larger diffusion rates

limited vs unlimited inheritability

attractor based vs storage based inheritance

- metabolism first vs evolution first scenario of origin of life
autocatalytic sets, Evolvable?
- In RNA world both intertwined (ligation based models)
- “BIT by BIT” (Joyce): first largely attractor based.
- evolution of DNA in RNA world: separation of ‘work’ and information storage
(immediate disadvantage, evolutionary advantage)
- redundant multilevel GP mapping
 - evolution of coding structure: robustness/evolvability
(NB population vs individual evolvability)

Multilevel evolution conflicts and the evolution of complexity/novelty

- Novel mechanisms when macrolevel and micro level selection similar
 - conflict resolution
 - symmetry breaking plus/minus strands (reduces intracellular mutation pressure)
- In space: 'super'symmetry breaking advantageous at individual level (availability of templates)
AND wave level through strong catalysis, enhanced by evolvability
AND spatially different selection pressures lead to speciation
- Evolution of DNA in RNA world (evolutionary stabilization)
- Gene proliferation (INDELS, transposons) enhance GP mapping evolution
robustness, special and (individual) evolvability, U-shape
- “WGD and irremediable complexity”
- HGT rates compensate rare positive cell level selection.

Multilevel evolution and Evolution of evolution

random mutation =/= random

Random mutation/selection leads to:

mutational priming, mutational neighborhood

U-shape mutational neighborhood: neutrslity and selection
(minimization of “slighly deleterious mutations”)

genome structuring through transposons:
some mutations become more frequent then others
increases evolvability

evolution of regulatory network (G/P mapping):
effect of random mutations biased to beneficial ones

evolvability and/vs regulation

Multilevel modeling: computing an organism

from emergent structures and evolutionary consequences
to defining multiple levels

from studying evolution to studying evolved systems

from studying evolution to using evolution to
(1)alleviate parameter curse (debug models and experiments)
(2)understand regulatory networks and morphogenesis

heuristics for modeling complex systems

Simplify

- including structure leads simpler models (root reflux)
- find sufficient conditions (for “hard problems”)

cf Dd lifecycle vs stripes

- use known simple models to interpret behavior of more complex networks

e.g. Turing patterns for fish-stripes or for digits

complexify

why not a preconceived simple solution?

- (Turing patterns and Dros. stripes)
- PIP network: include structure of other levels

data fitting

use detailed low level knowledge + detailed high level knowledge: FIT!?!
cf Drosophila stripes: developmental dynamics of expression

Model-fitting

Model repository of known interactions, sufficient?
what else do we need?

Use evolution to alleviate parameter curse

RESULTS ++

What are models for...

- Proof of principles
There is a system which ... (game of Life)
- base line expectation (attractors, pattern formation)
- Expectation exploration what happens if we assume....
emergent behavior
- Paradigm systems (flags for navigation)
- Debugging (e.g. Lac operon)
- Predictive models
NB parameters, initial conditions!
(compare aerodynamics and weather prediction)
of WHAT
(compare Newton's apple)
(synthetic biology)
(iff nothing else interferes...)
- *understanding...*

modes of explanation

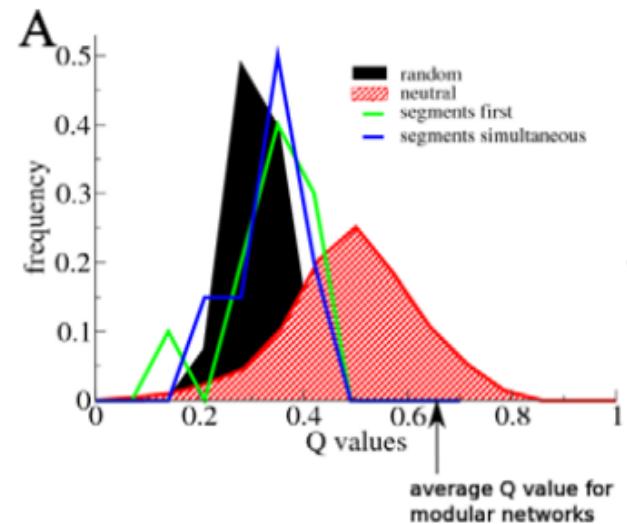
“Waarom, waarom, waarom zijn de bananen krom”

- NOT bend – > NOT banana. **tautology**
 - cf “survival of the fittest”
 - error threshold (AND too low mutation rate/ small population)
 - fitness as time dependent function

(re)define concepts such as most informative

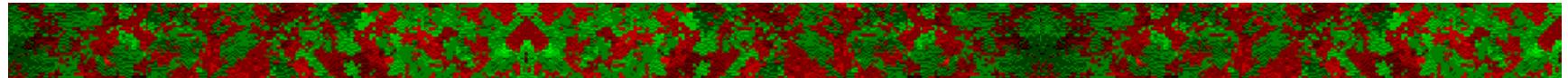
- “Almost all cases” BEND (straight pathological case)
 - However straight simplest (linear) case
 - cf order parameter (lambda in CA, diversity, NK networks)
 - generic properties
 - WHY lymph-node structuring, spirals etc
- bend “OPTIMAL”
 - cf walking in circles
 - cf “optimal foraging”, defense etc
 - TODO vs textbooks
 - cf for WHOM (in multilevel selection)
 - ‘clever’ bacteria benefit viruses and plasmid
 - cf cost/benefits (trad-offs)
 - positive selection early death
 - virus infectivity

- WHY not (better for individual)
-cf Dicty - stalk cells
- Side-effect of evolution to MANY banana's (+gravity)
- Side-effect of growth
- side-eft of mutationl operators/drift (cf FFL, modularity)
- Knock out genes - till non-bend phenotype
- Make detailed model of “real” banana's....



evolved networks modular? Yes/No

Elephant, I believe....



You got about by
random mutations
local interactions
multilevel selection
genome structuring
mutational priming (non-random mutation)

.....

and I can only understand you by

Simplification
but not to one level