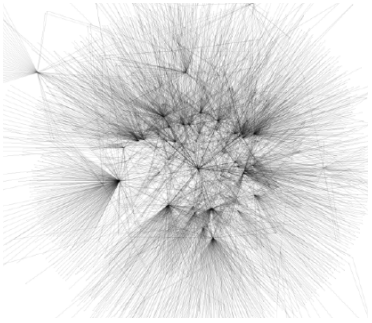
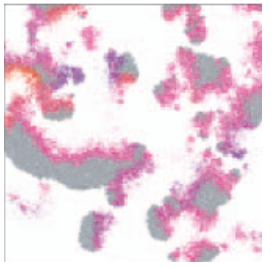
Modeling Biocomplexity: overview - review

challenges, themes , insights and methodology

Course Computational Biology 2025; Paulien Hogeweg;
Theoretical Biology and Bioinformatics Grp Utrecht University





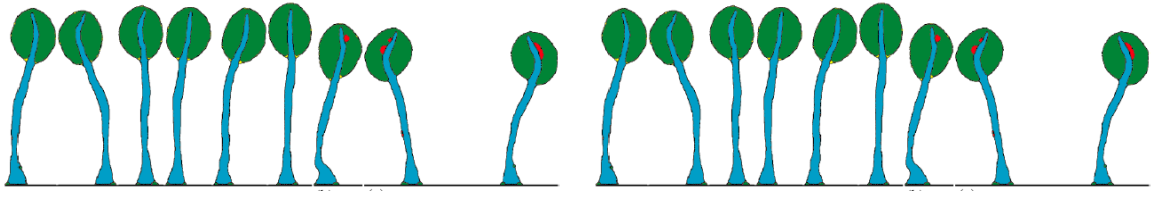
Biocomplexity

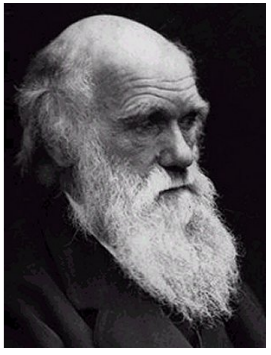
interactions among multiple space and time scales

How to model

How did it evolve

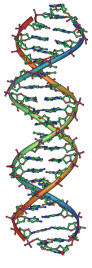
How to model how it evolves

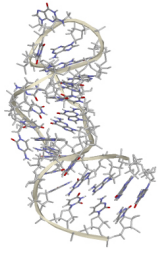


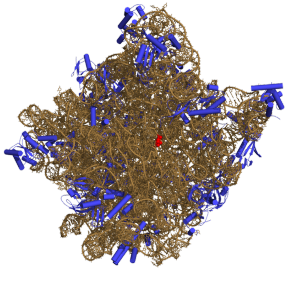


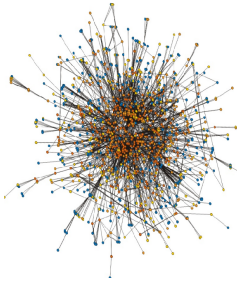
Darwin:
Theory of evolution
by mutation and
natural selection

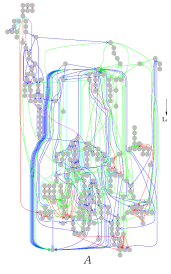
"**There is grandeur in this view of life**, with its several powers, having been originally breathed into a few forms or into one; and that, whilst thisi planet has gone cycling on according to the fixed law of gravity, **from so simple a beginning endless forms most beautiful and most wonderful have been, and are being, evolved**"

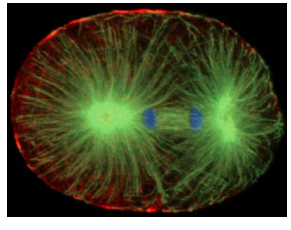


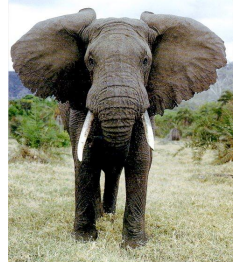












Modeling Biocomplexity
What can/cannot be done (sensibly)

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

Models should be as simple as possible but not simpler
Einstein

Empty cross-section? *"cherry picking"*

properties/processes,space/timescales,specicity/generality,Q/Q...

Experimental Theoretical/Computational Biology
(cf Ulam experimental mathematics)

A conversation with Paulien Hogeweg, Utrecht, Ni

“Thinking in the Most Interesting Simplification”

strategy as pursued in course

Use a "nonsupervised" modeling strategy:
do not make a model of some preconcieved phenomenon
but explore consequences of basic assumprions
"first principles"

Use individual based (agent based) models:
allow diversity within "population"s in time and space;
allow structure to evolve

Use spatial embedding, allowing spatial pattern formation
structuring interactions

Exploit self-organization (emergence of complexity)
and feedback of macro level to micro level
(first and second order emergence)

(Stepwize) increase degrees of freedom of modeled system

study special cases with multiple special models
'beacons' and generalize by convergence

Somedefining properties of biocomplexity: and consequences for modeling

---

- 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 and exploit these properties

Moreover: biotic systems are information accumulating, storing and processing systems

"Biological systems are distinguishable from chemical systems because they contain components that have many potential alternative e compositions but adopt a particular composition based on the history of the system. In this sense biological systems have a molecular memory (genotype), which is shaped by experience (selection) and maintained by self-reproduction"

Joyce (2012) Bit by Bit: The Darwinian Basis of Life:
"How many heritable bits are involved, and where did they come from
EVOEVO

Model 'requirements':

Redundant non-linear Genotype to Phenotype mapping
(many alternative codings of phenotypes)

Evolution of genome size and structure

("structured entities")

evolution of complexity, how?/why?/when?

classical models do not explain it (only reproduction rate)

Maynard Smith & Szathmary:
How DID it evolve?

“Major transitions in evolution”

common processes
”giving up selfsufficiency” and ”Division of labor”
”Multilevel evolution”

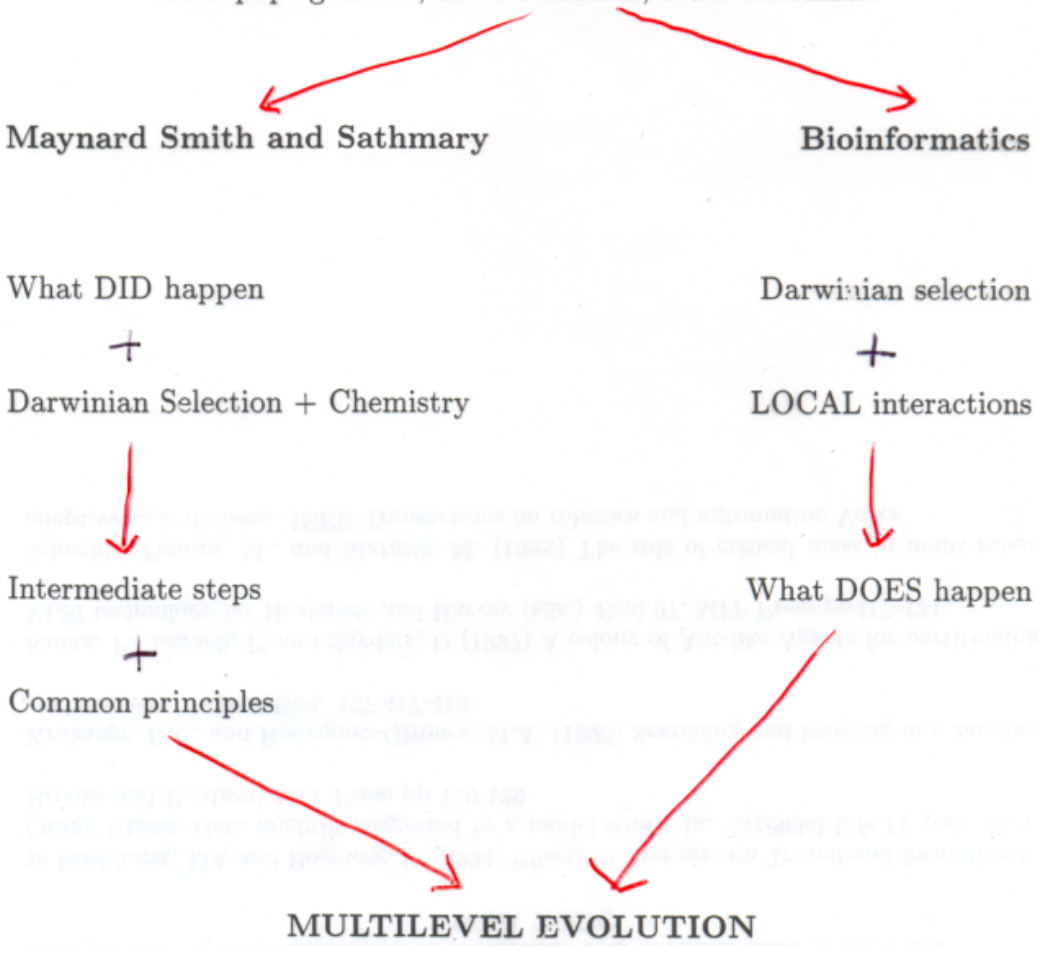
“Elephant, I believe you got about by random mutations

%

"Elephant, I believe you got about by random mutations"

Evolution of 'complexity' and 'novelty'

'Elephant I believe you got about by random mutations'



Maynard Smith & Szathmary 1997

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

- self-sufficient to "part of a whole'
  (e.g. symbiogenesis)
  Multiple levels + conflicts
- Division of labor
  DNA to DNA/RNA world, germ-line soma, social insects, human societies
- limited to unlimited inheritability
  attractor-based vs storage based (information based) inheritability

Emergence of these in "simple" models of Darwinian evolution

"levels" of multilevel evolution (1)

Replicators within replicators

- replicators and self-organized (replicating) spatial patterns automatic consequence of local interactions
- replicators in protocells (coupling of dynamics)

  cf stochastic corrector
- cells with dupdels/LCR/plasmids/transposons..
- HGT: cell level vs gene level replication/selection

  cf differential mobility (e.g. toxin-antitoxin)

"levels" of multilevel evolution (2)

- multilevel genotype-phenotype mapping
  - single molecule: RNA folding
  - GRN
  - "virtual cells" / "virtual microbes" (GRN/Metabolism)
  - regulation of mechanical properties of cells
  - development
- Evolution ON vs evolution OF GP maps
  evolution of coding structure (from RNA to Vmicrobes)

Note: Influence of [1] in [2]

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 evolutionary 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?
- * fitness as a time dependent function
  cf Savill & Hogeweg)
- Sparse fitness evaluation and information integration
- Genome expansion - long term evolutionary benefits
- Reconciliation adaptive and neutral evolution
  (presence before beneficial / loss of genes)
- Evolutionary vs (immediate) functional benefits

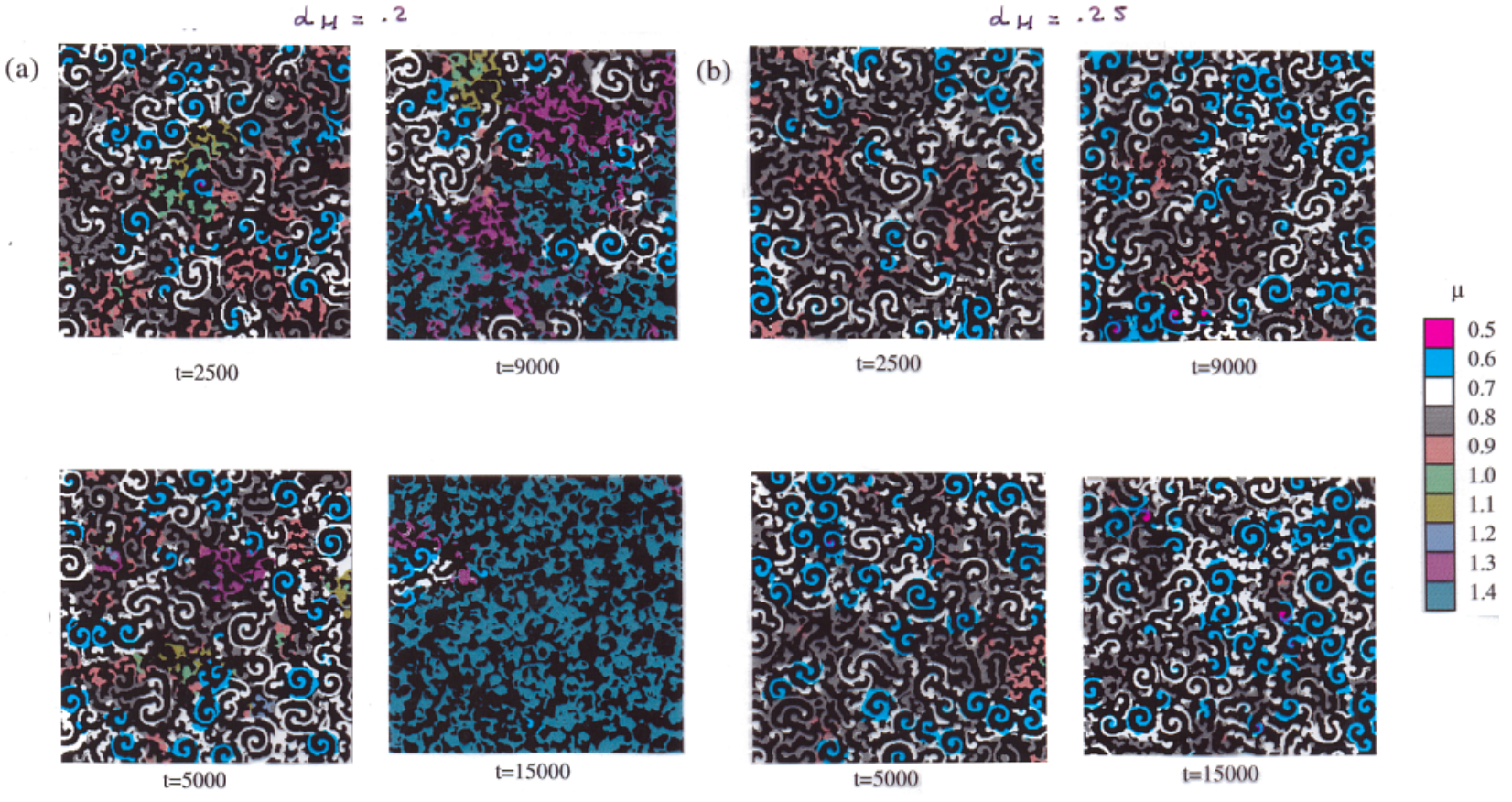
Predefined vs "emergent" fitness

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

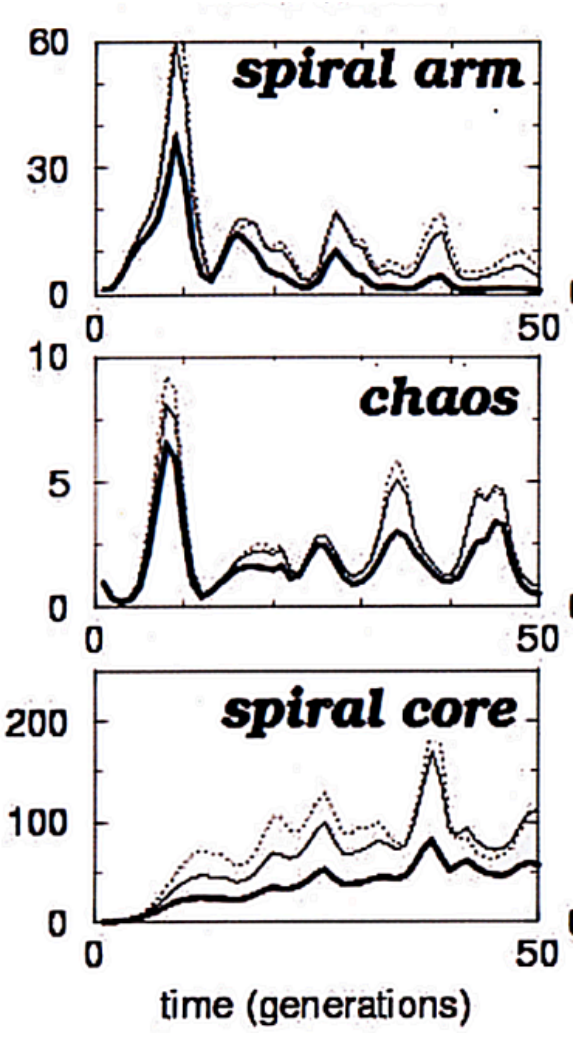
- 3 levels of selection:
  - host / parasitoids
  - spiral waves / chaotic waves
  - regions of spiral waves and chaotic waves
- 'Life history' of spiral
  - 'born' at edge of spiral and chaotic region with high β values increases domain (faster rotation)
  - *decreases differential migration (β) (core as germ-line)*
  - *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

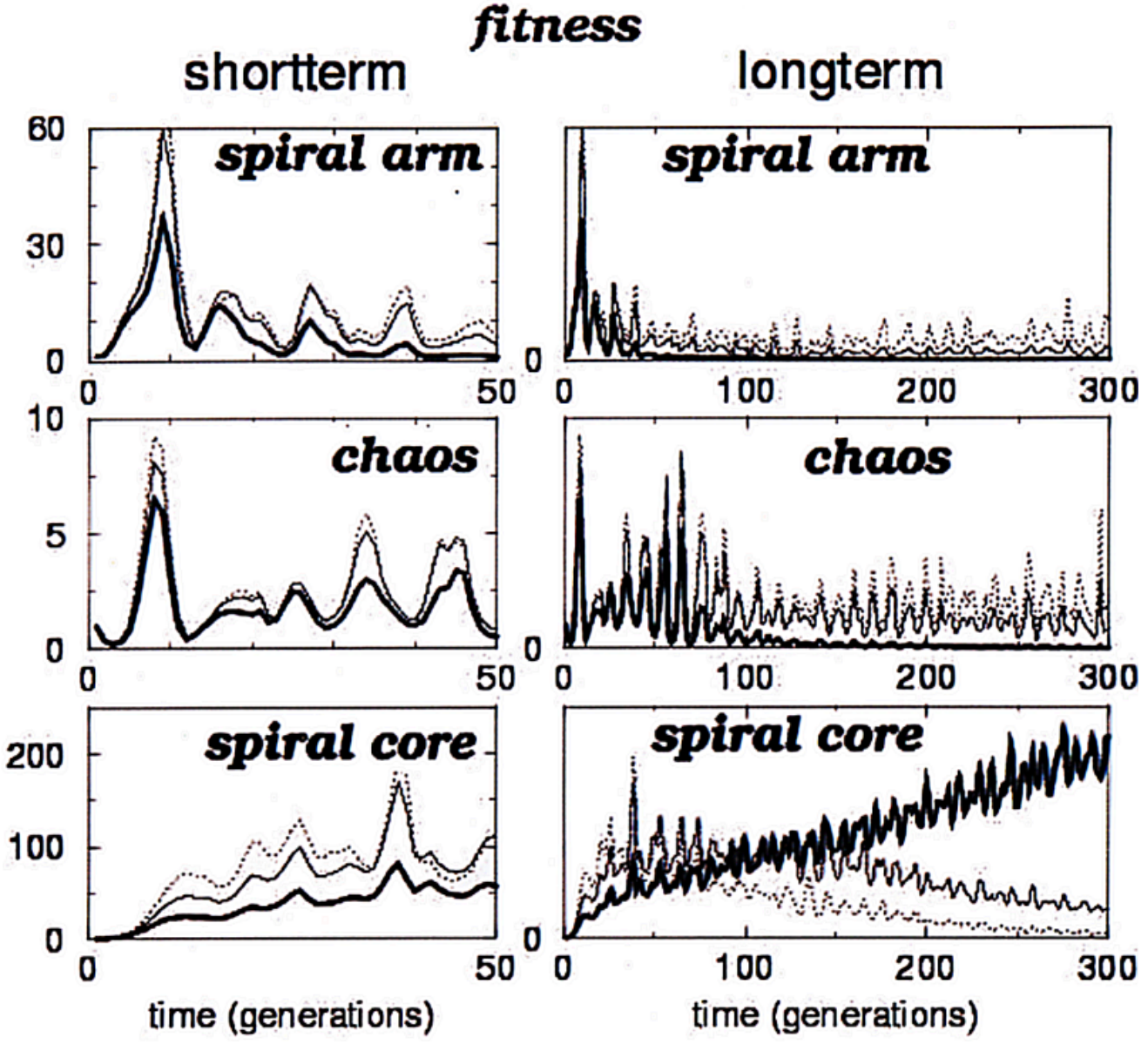


fitness as # offspring

through time

from 1 type at 1 location

to anywhere (any type)

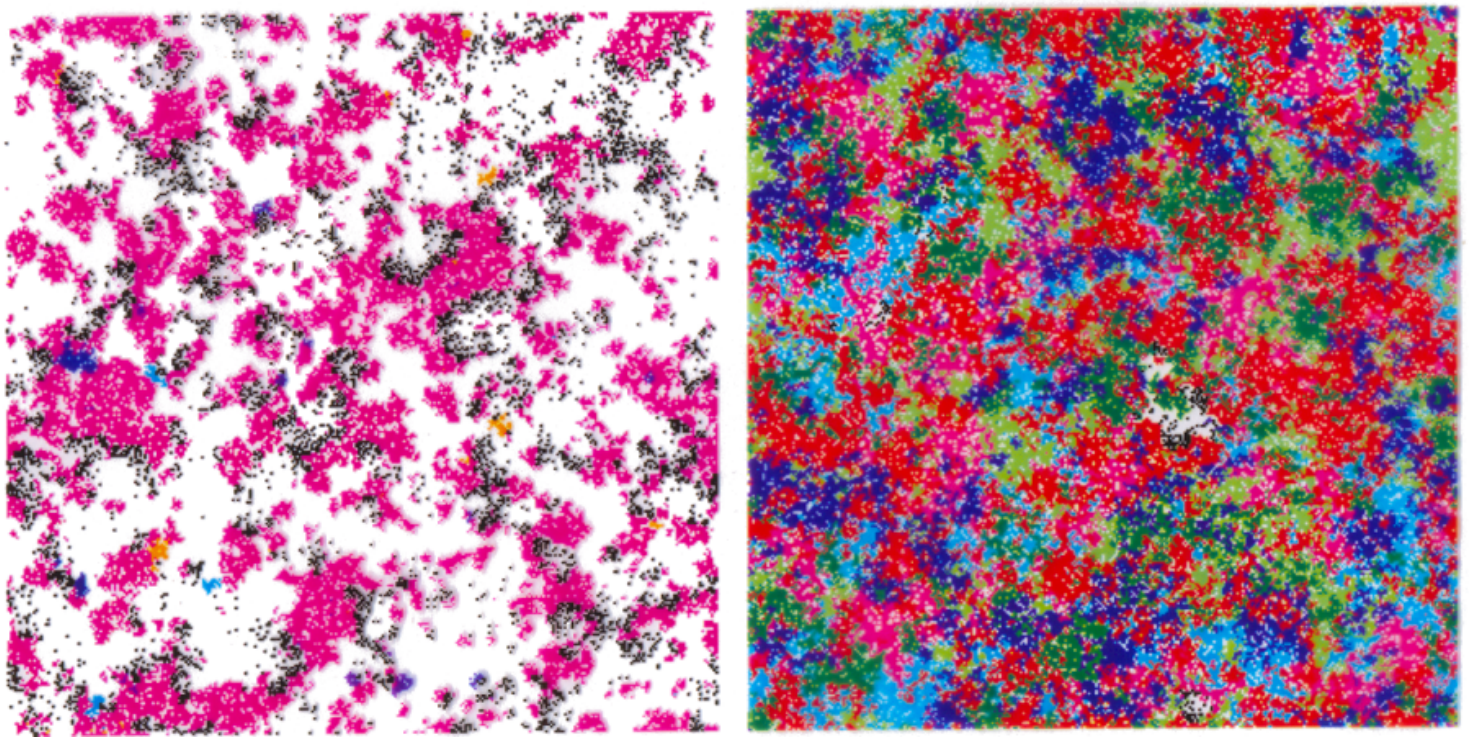

fitness
shortterm
longterm
spiral arm
spiral arm
chaos
chaos
spiral core
spiral core
60
30
0
10
5
200
100
0
50
100
200
300
time (generations)
time (generations)

Immediate fitness for whom?
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

Funtion(?)
antiviral defense

cf Pagie and Hogeweg 2000



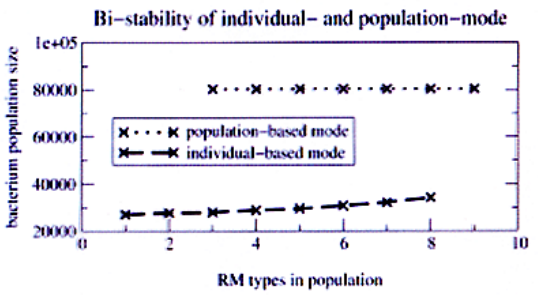
Two alternative attractors

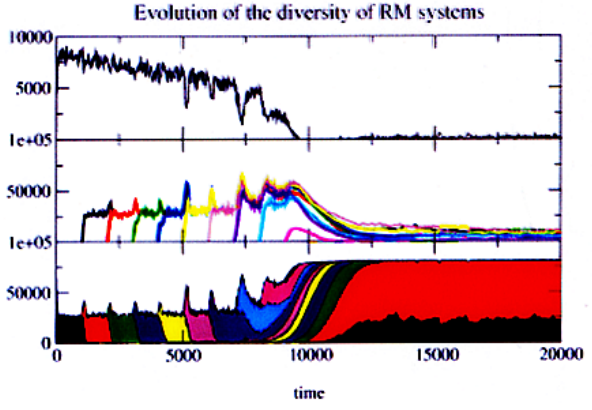
Individual based diversity': waves of virus infection

population based diversity: static pattern, few phages

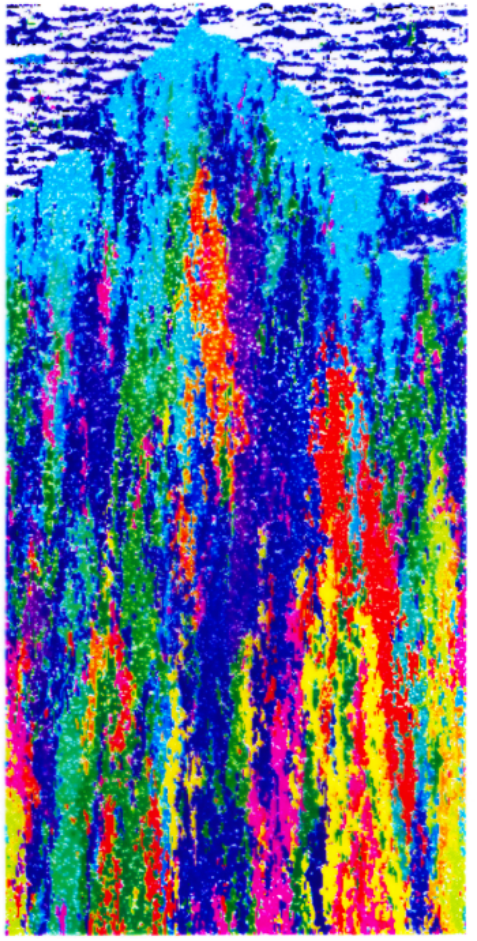
transition from individual based to population based diversity

RM systems: from individual diversity to population diversity

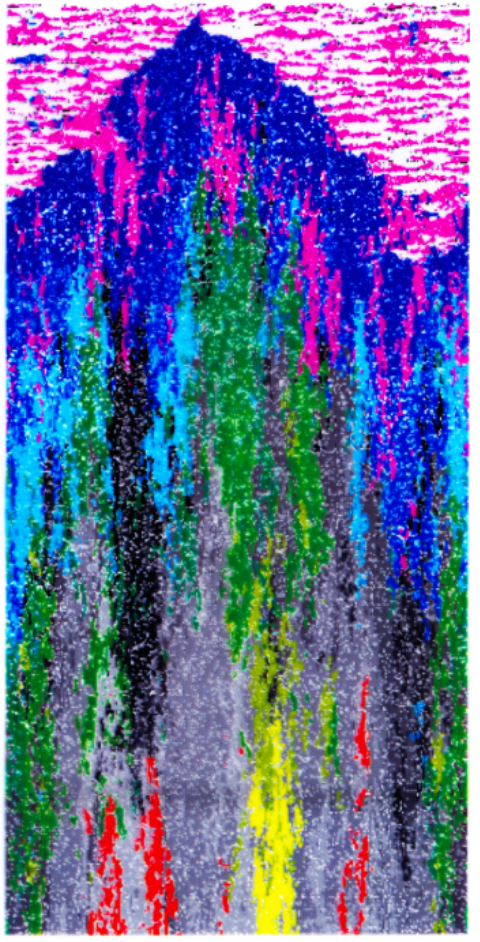




Types



numbers



conclusion RM systems

First: Large "genomes"

Later: Reduction of Genomes

Decline of Virus population

Cycle

beneficial for cells vs viruses.

“ giving up self-sufficiency”
= exploiting opportunities

- Evolution of cross-feeding in Vmicrobe
  (”black queen” evolution)
- Emergence of not self-sustaining cheaters/parasites in single replicator systems
- Local interactions lead to more global inter-dependencies through spatial pattern formation and the emergence of novel levels of selection (evolutionary interdependence)
- Individual and/vs ecosystem based ‘problem solving’
  (cf original purpose of hypercycles)
- Evolution of DNA in RNA world

division of labor - non-reproducing "helpers"

- non-heritable phenotypic differentiation
  regulation Dictyostelium Spore/Stalk

- division of labor in quasispecies (mutational decoded division of labor) cf RNA, Streptomyces

- Spatial determined division of labor
  All offspring from core of spiral (location decoded)
  (compare germline and soma)

- Symmetry breaking of plus/minus strands of complementary replicators

- information storage vs information usage

- individual based vs ecosystem based problem solving

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

Evolution of DNA in RNA world

Example for:

Common processes in major transitions (MS-S)

Giving up selfsufficiency

Division of labor

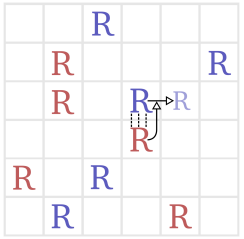
attractor based vs storge based inheritance

conflict resolution between levels of selection

multiple specific models to discover generic properties

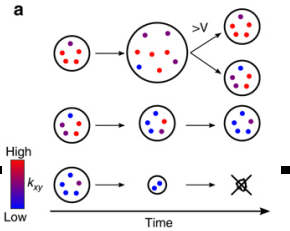
Convergence to same pattern

but in different ways



Multiple models; convergent evolution

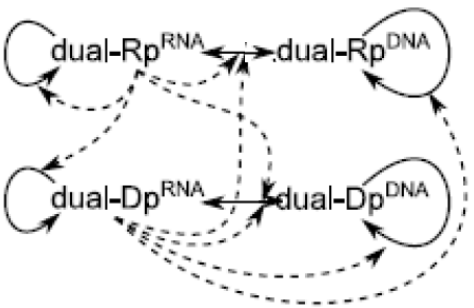
Nobuto Takeuchi et al 2011.2017,2019

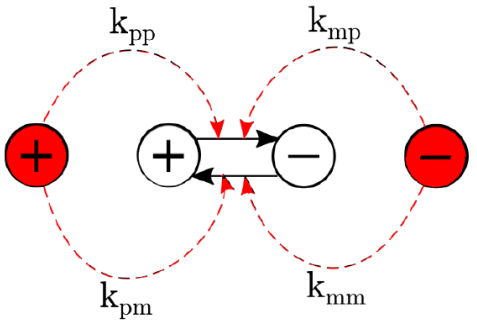


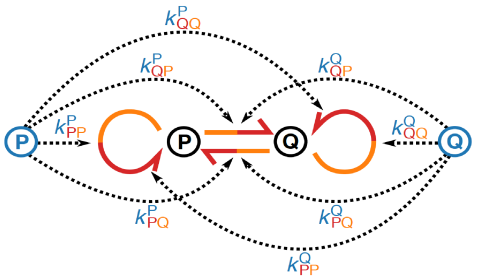
SPACE

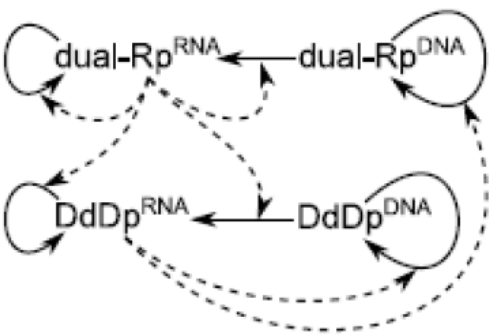
Protocells

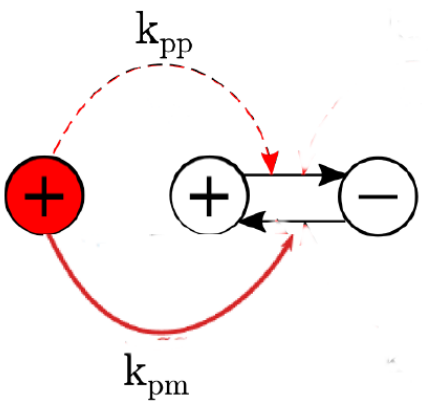
Invasion DNA in RNA world **complementary replication** **symmetry breaking**

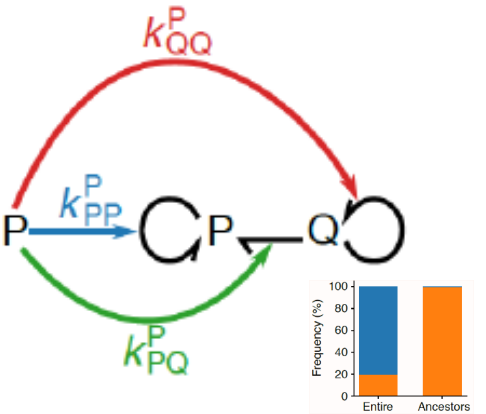








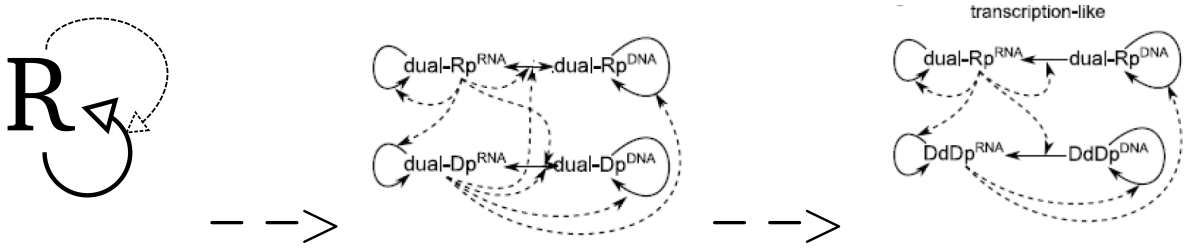




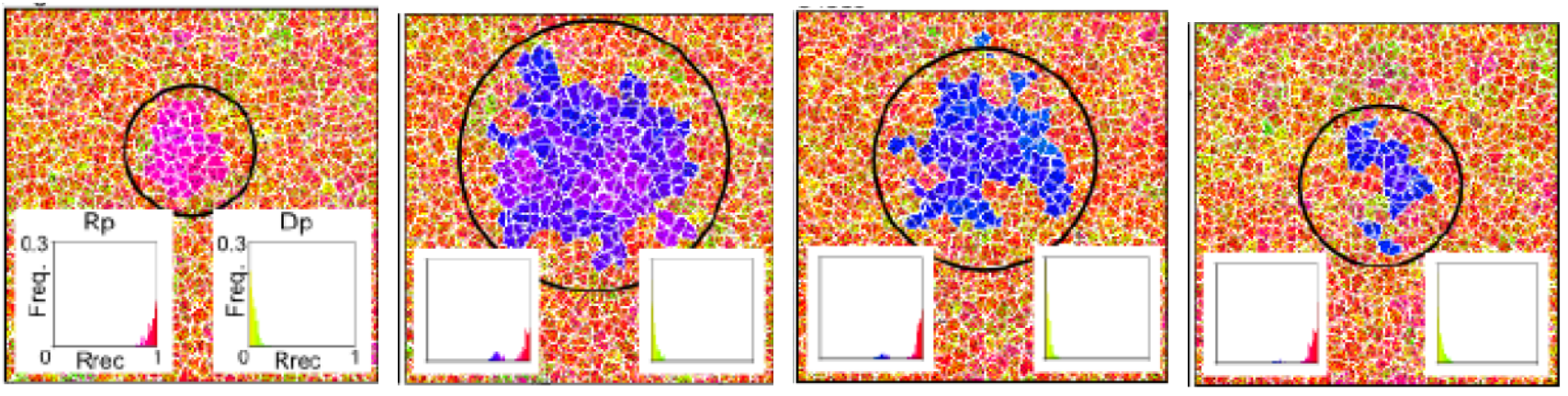
Division of labor; Unidirectional Information Flow

Crick's Dogma

Unidirectional information flow evolves for evolution: Negative immediate functionality



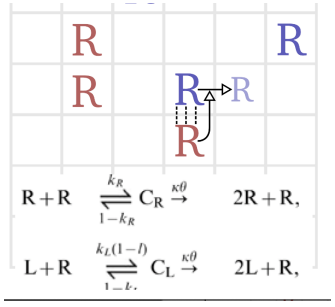
- DNA polymerase can invade and specificity evolves such that information flows from DNA to RNA (no reversed transciptase)
- This prevents RNA evolution to lower catalysis
- Loss of DNA in cell $--> $ expansion followed by extinction

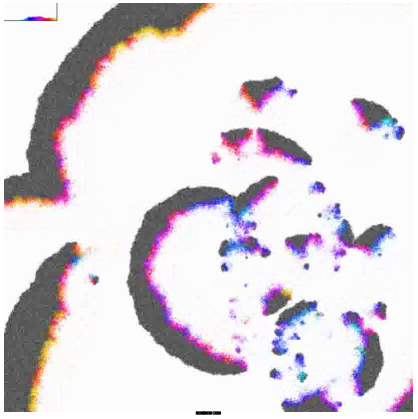


Takeuchi, Hogeweg & Koonin 2011;

Major transitions as generic properties of Darwinian evolution
Multilevel Evolution: Conflict Resolution (− > complexity)

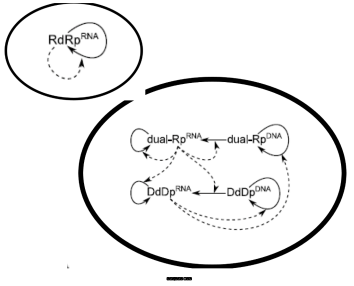
Emergence
waves as Darwinian entities

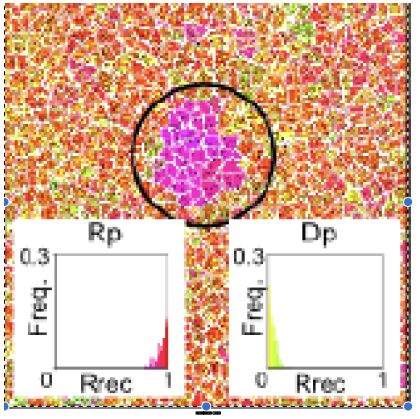




birth/aging/death
mutation/selection
max. birthrate
of waves

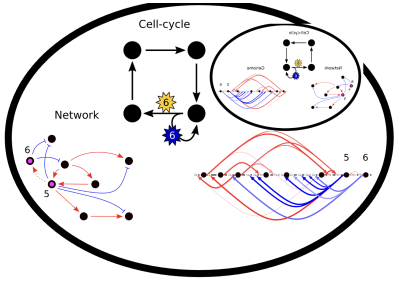
unlimited inheritance
DNA in RNA world
symmetry breaking

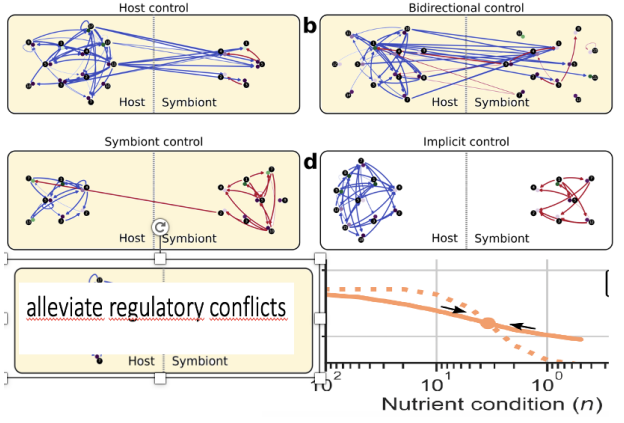




unidirectional
information flow
Crick's "dogma"
for evolution

Eukaryogenesis
complexity by endosymbiosis





implicit control − >
novel adaptaion mechanism
homeostasis
r vs K selected

evolution of complexity by relaxing "fast replication curse"

Multilevel evolution (2)

Evolution **ON** *vs* **OF** *GP maps*

IFF

predefined genome structure (e.g. RNA of fixed length)
predefined mutations (e.g. point mutations)

GP map well definable, can possibly be studied globally
Evolution: walk **ON** GP($-x->$ fitness) landscape
(e.g. towards high neutrality)

OTHERWISE

Evolution **OF** GP map e.g. because

Indels/LCR change genome structure and dimensionality of landscape
genome structure changes which mutations occur
(non-random random mutations)
GP map only locally 'sensible' as *Mutational Neighborhood*

mapping phenotype to fitness variable in space and time

evolution ON GP maps: RNA GP landscape

genotypic vs phenotypic information threshold

neutral vs adaptive evolution
VS
neutral AND adaptive evolution

evolution of robustness (neutrality)

– – – >

evolution of evolvability (population level)

Multilevel evolution: Evolution OF GP map and Evolution of evolution
random mutation =/= random

Random mutation/selection leads to:

mutational priming

"functional" mutational neighborhood

U-shape mutational neighborhood: neutrality and selection
(minimization of "slightly deleterious mutations"

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

evolution of regulatory network (GP mapping):
effect of random mutations biased to beneficial ones

evolvability and/vs regulation

evolution of (frequency of)mutational operators
(e.g. DupDels/ HGT (of cells, of genes)

.....

Predictability of evolution *(of what?)*

- Generic patterns (as discussed above)
- Diverse "implementations"
- early towards multiple alternative attractors
- "Developmental Drift"
- Depends on prior evolution
  "predictibility is an unpredicable outcome of evolution"

Specific models often more general then general models

qualitatively same phenomena reoccur in many different models
– evolving genome size/structure (INDELS, LCR etc)
– evolving GP mapping
– no predefined fitness (those that survive (or very general fitness criterion like homeostasis)

- RNA GP map vs NK fitness landscapes
- Early genome inflation followed by streamlining
  fitness gain by gene loss
- U shaped mutational neighborhood
  skewed to neutral or deleterious evolves as response to (changing) conditions
- mutational neighborhood 'special' and 'beneficial'
  reflects evolutionary memory
- evolutionary memory retained despite high neutrality.
- substitution mutation and INDELS, LCR very different role in evolution

Multilevel evolution allows evolution of complexity
Novelty for conflict resolution

From Modeling evolution from "first principles" to modeling evolution of evolved special cases

Use evolution as tool to:

(1)alleviate parameter curse
(debug models and experiments) (cf Lac-operon)

(2)understand A regulatory network/ genome structure
(cf Ribosomal RNA yeast)

(3) Understand WGD in yeast

supervised modeling:
how to reproduce predefined pattern/structure

Simplify to minimal/general mechanisms/requirements

e.g.

Turing patterns/Positional Information/Clock and Wavefront

How implemented; evolutionary conserveration vs reinvention at different levels: molecules/pathways/mechanisms

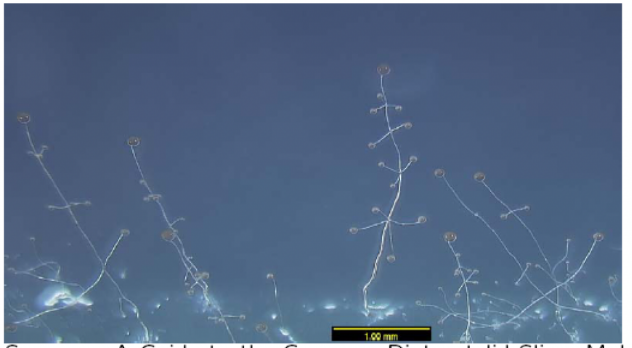
Large scale models: implement "all" prior knowledge sufficient to match?

stripes too easy(?) (degenerate)

find sufficient conditions (for "hard problems")

"how to compute an organism"
minimal mechanisms - multiscale coordination
novel sensing, novel behavior
multiple observables
evolutionary testing of the model (?)

Polysphondylium violaceum



A.R. Swanson, A Guide to the Common Dictyostelid Slime Molds of Great Smoky Mountains National Park

continuous redifferentiation prestalk-stalk
sidebranches (polyshondinium)

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
  – does not need explanation (cf Lymphnode)
- bend "OPTIMAL"
  - cf walking in circles
  - cf "optimal foraging", defense etc
  - cf for WHOM (in multilevel selection)
  –'clever' bacteria benefit viruses and plasmid
  -cf cost/benefits (trade-offs)
  – positive selection early death
- Side-effect of evolution to MANY banana's (+gravity)
- Side-effect of growth
- side-effect of mutational operators/drift (cf FFL, modularity)

- Knock out genes - till non-bend phenotype
- Make detailed model of “real” banana’s....

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