

Center of Modeling, Simulation and Interactions
Université Côte d'Azur – June 4th 2020

Evolution of Complexity

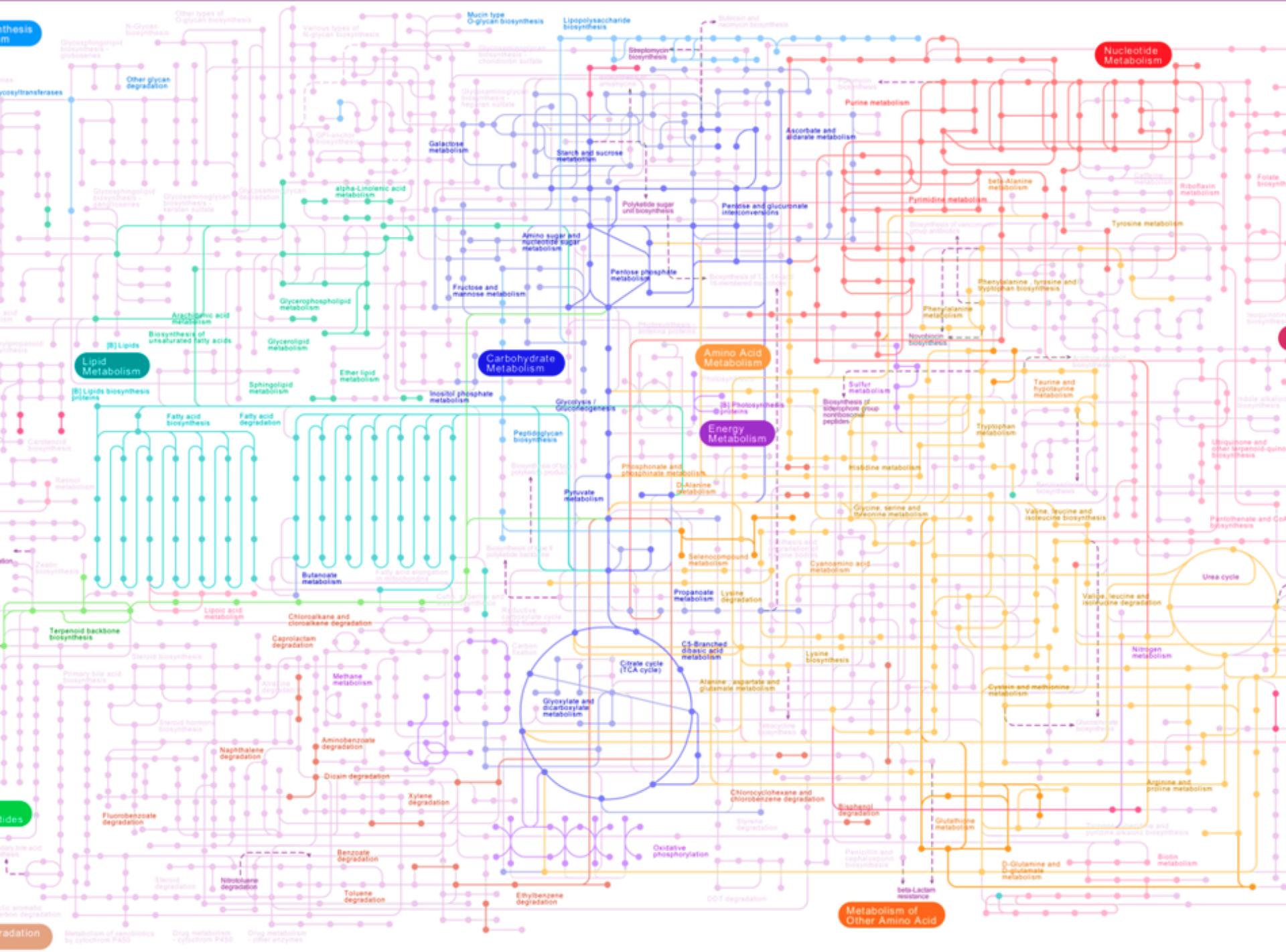
– a modelling exploration –

Guillaume Beslon

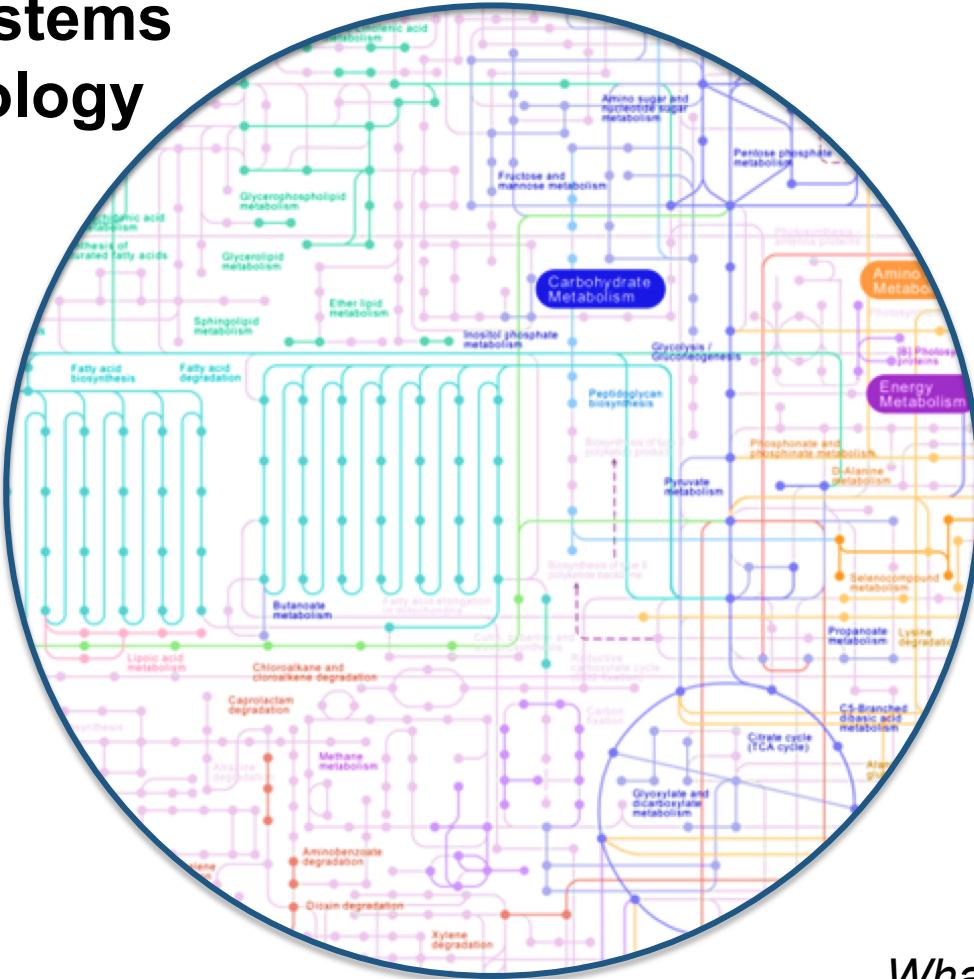
INRIA Beagle team, LIRIS, UMR CNRS 5205

Guillaume.Beslon@inria.fr





Proximal Causation: Systems Biology

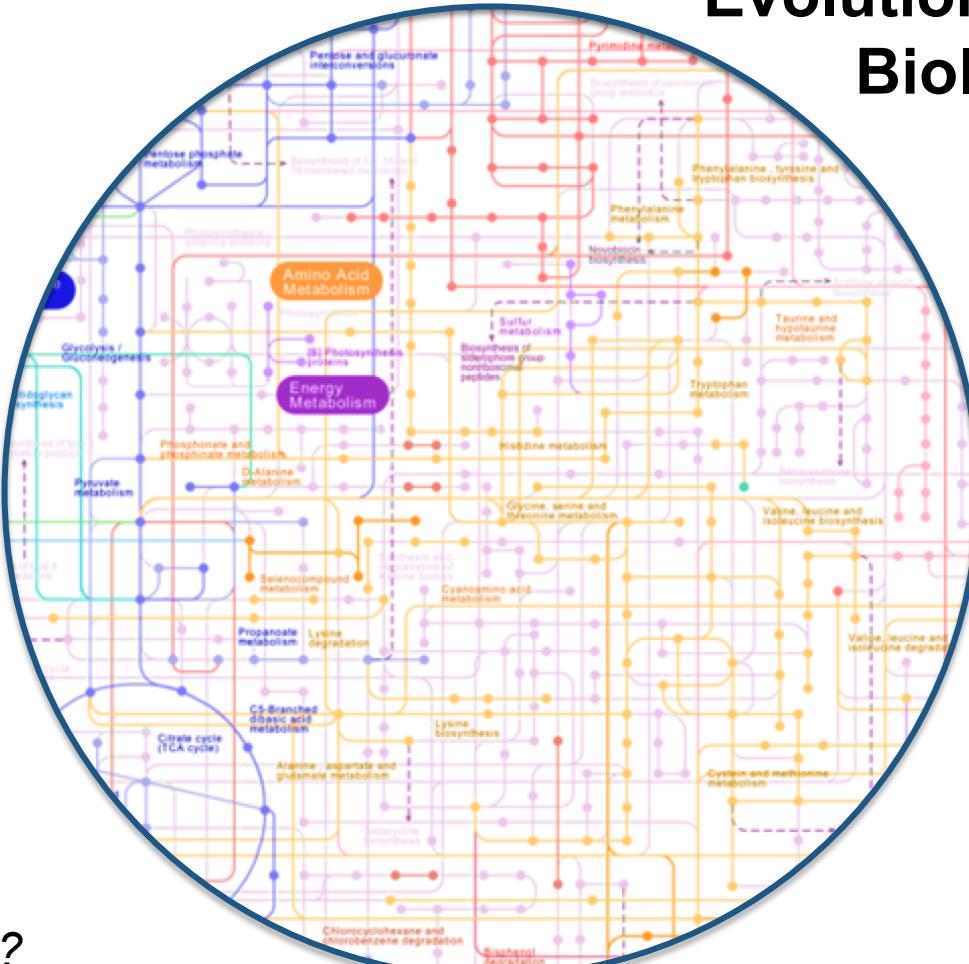


Concepts
Molecular networks,
Complex networks
Pathways and motifs
Multiscale systems
Modularity
Feedback

...
Questions

Are there design principles?
What is the robustness of this network?
How is information processed and linked to cell fate decisions?
Balance between Central/Distributed decision making?

Distal Causation: Evolutionary Biology



Concepts

Population and Drift

Selection and Fitness

Mutations and Epistasis

Fitness landscape

Genotype-to-Phenotype map

...

Questions

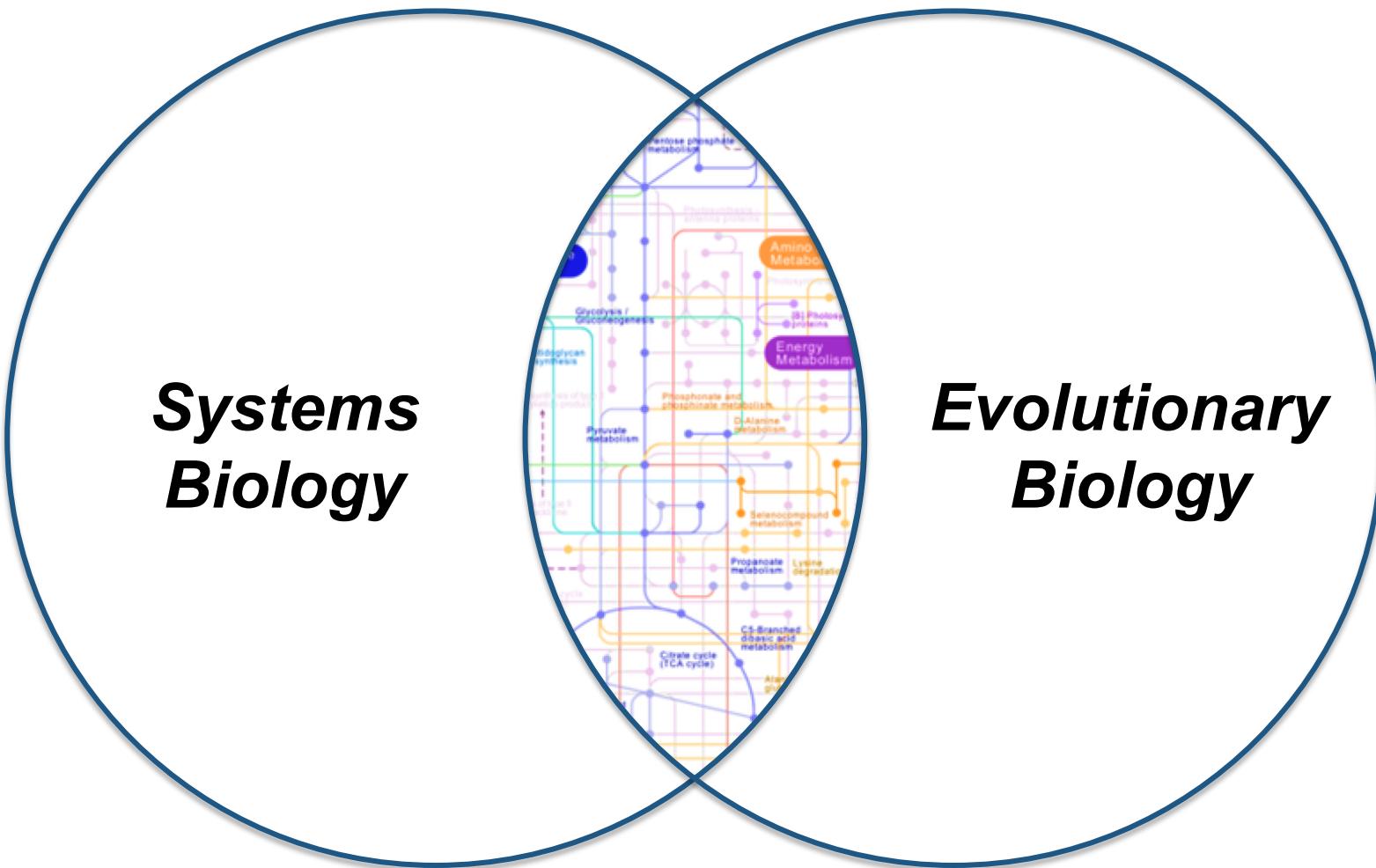
Which traits are under selection?

What is the fitness of this organism?

What is the distribution of mutational effects?

...

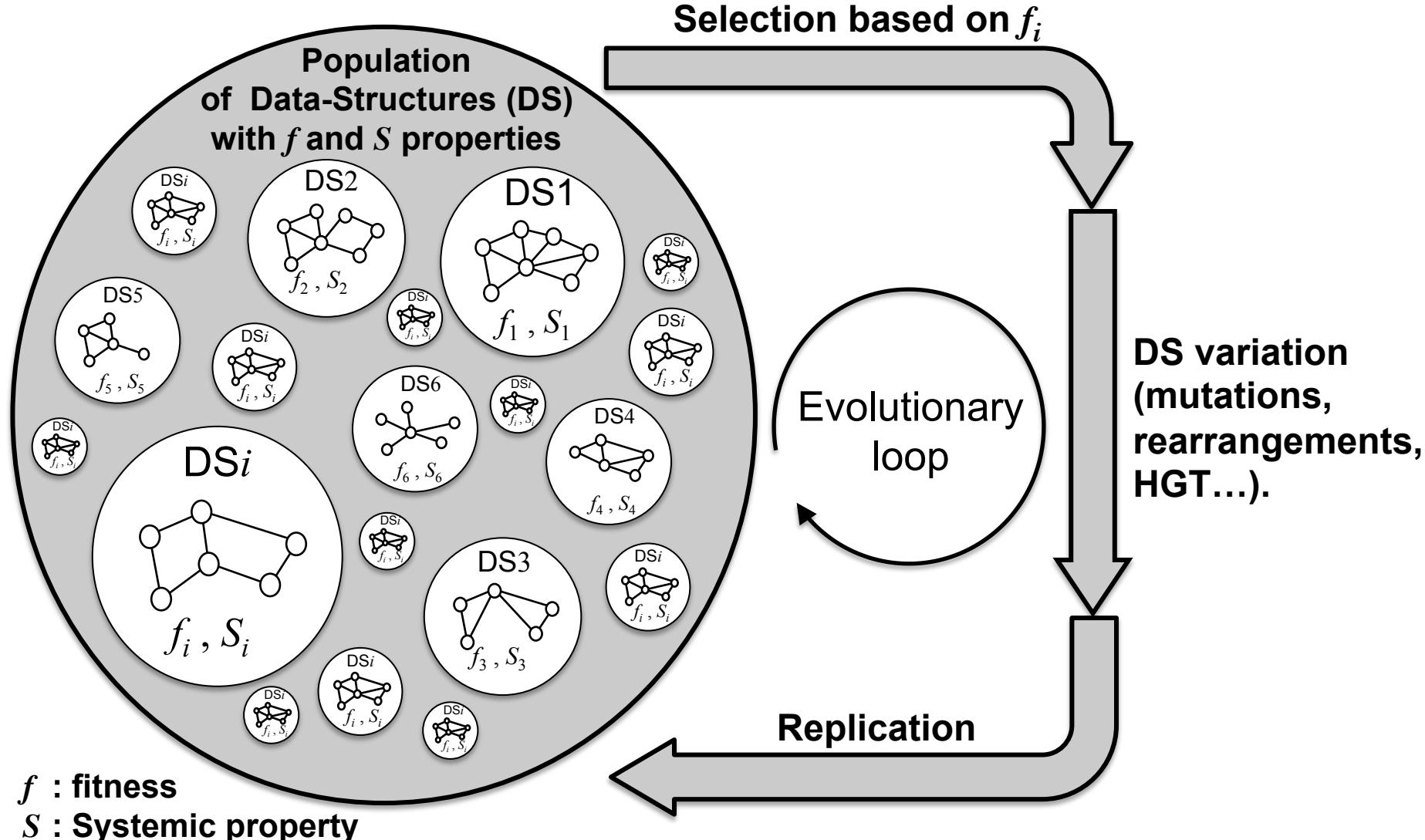
Reconciliating Proximal and Distal Causation? Evolutionary Systems Biology



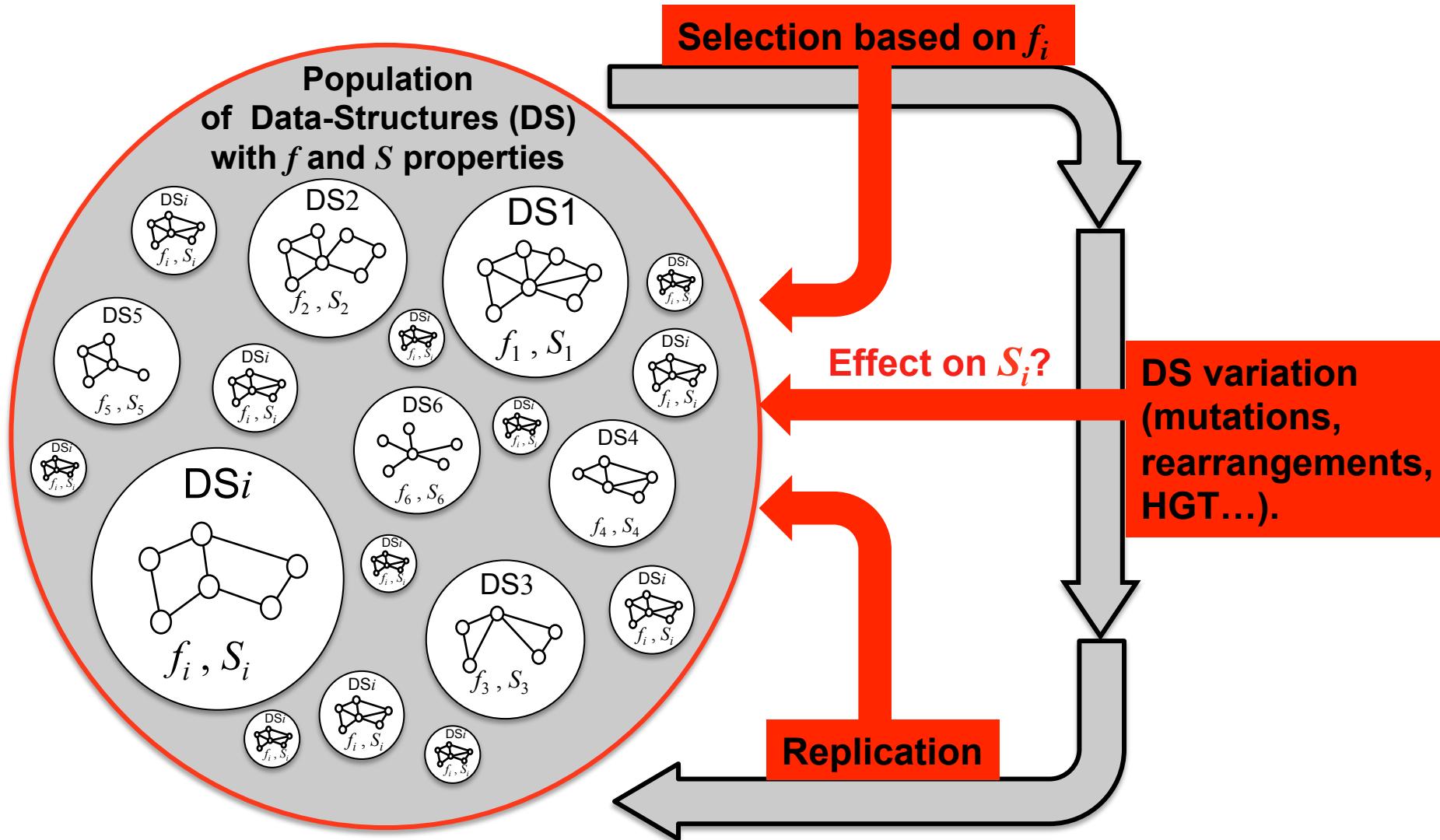
Concepts, Questions, and Tools in Evolutionary Systems Biology

- Concepts in Evolutionary Systems Biology
 - ESB mixes up systems biology and evolutionary biology concepts
 - Selection, complex networks, drift, fitness landscape, modularity, robustness, evolvability, mutation, epistasis...
- Questions in Evolutionary Systems Biology
 - What kind of systems are the most likely to result from an evolutionary process?
 - How does the structure of the system influence its evolution
- Tools for Evolutionary Systems Biology
 - Need tools to study (co)evolution of systems structure and systems properties
 - *Comparative genomics*
 - *in silico experimental evolution*

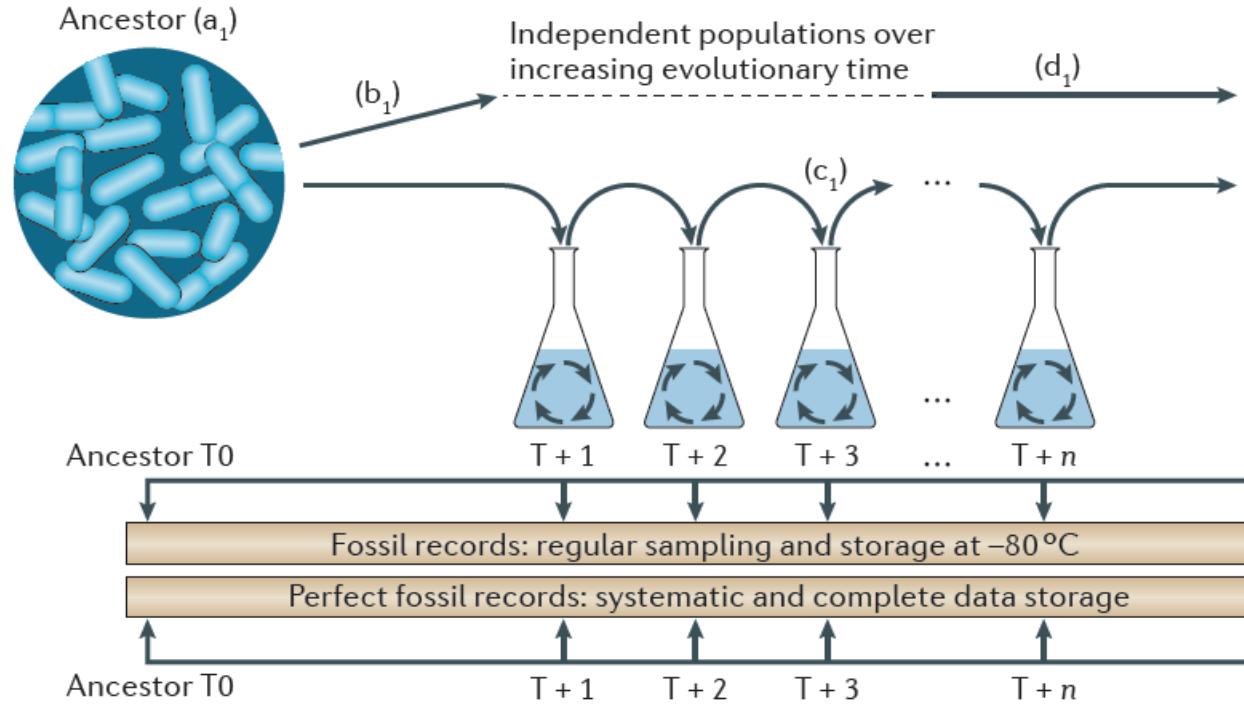
In Silico Experimental Evolution (ISEE)



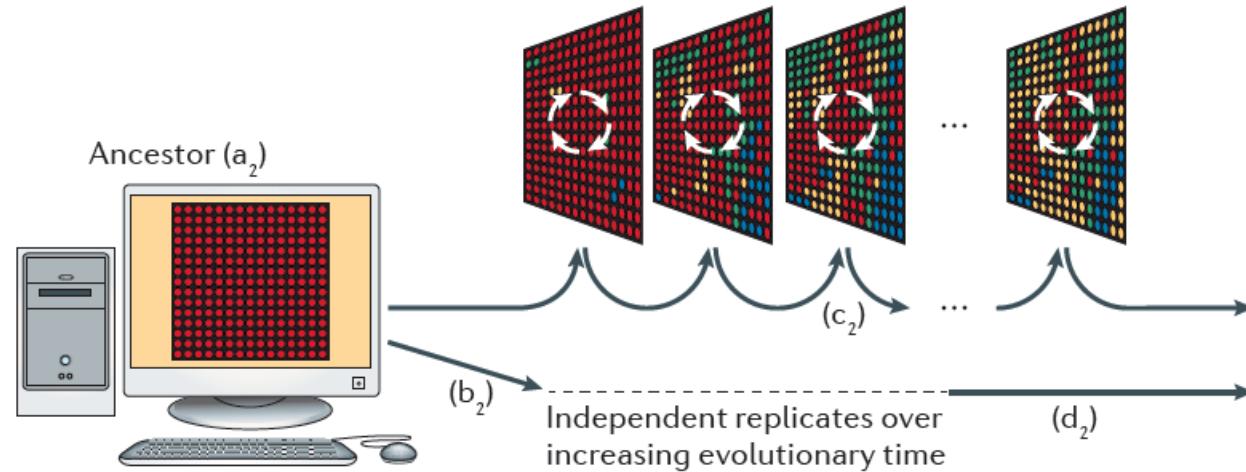
In Silico Experimental Evolution (ISEE)



in vitro experimental evolution



in silico experimental evolution



[Hindré, et al., *Nat. Rev. Microb.*, 2012]

What are these “data structures” we call “organisms”?

Table 2 | Genome formalisms in *in silico* experimental evolution

Formalism*	Description	Questions addressed
Program	The genome is a sequence of instructions in a programming language. The fitness of the program depends on its ability to create copies of itself in the computer's memory and/or to perform specific computations	<ul style="list-style-type: none">The emergence of parasites and hyperparasites³⁸The evolution of robustness, evolvability, complexity and modularity^{3,87,142,147,150}The adaptive radiation of species⁸⁴The information threshold (the maximum amount of information that can be evolutionarily maintained)¹⁵¹
Allelic	The genome is made up of a fixed gene number, n ; each gene can exist in a finite or infinite number of alleles; alleles are represented by integers or characters, and each individual is characterized by its n alleles	<ul style="list-style-type: none">The evolution of mutators^{135,136}Bacterial speciation in neutral conditions⁸⁶
Network	The individuals are characterized by a graph representing a gene-regulatory network, a neural network or even a logic circuit; there is no explicit DNA level, and mutations directly change the connections or the node numbers in the network	<ul style="list-style-type: none">The evolution of network evolvability and modularity^{122,123,152,153}The importance of post-transcriptional regulation¹²⁴The relationship of robustness to mutations and to noise¹⁵⁴The evolution of communication, cooperation and altruism^{89,155}
String-of-pearls	The genome is a variable-length string of ‘pearls’ of different types: phenotype genes, transcription factor genes, repeats, retrotransposons, binding sites, and so on; each pearl type can exist in a predefined number of variants; gene number, order and regulation can evolve through mutations and rearrangements	<ul style="list-style-type: none">Genome and network evolvability^{141,143}Resource processing in ecosystems⁸⁵Sympatric speciation¹⁵⁶
Sequence-of-nucleotides	The genome is a variable-length string of characters; predefined signal sequences, analogous to promoters, terminators or start-stop codons, are used to detect genes; point mutations, indels and rearrangements can be simulated in a realistic manner	<ul style="list-style-type: none">The evolution of non-coding DNA and gene number¹⁴⁰The evolution of the size and topology of gene networks^{126,127}Gene network inference^{157,158}

*Many formalisms have been proposed to represent the genome, each with strengths and weaknesses. The appropriate formalism strongly depends on the question of interest. Here, we focus on the approaches that are most directly comparable to *in vivo* microbial evolution experiments (that is, approaches for which the genome comprises several genes).

[Hindré, et al., *Nat. Rev. Microb.*, 2012]

What are these “data structures” we call “organisms”?

Table 2 | Genome formalisms in *in silico* experimental evolution

Formalism*	Description	Questions addressed
Program	The genome is a sequence of instructions in a programming language. The fitness of the program depends on its ability to create copies of itself in the computer and perform specific computations	<ul style="list-style-type: none">• The emergence of parasites and hyperparasites³⁸• The evolution of robustness, evolvability, complexity and modularity^{39,40}
Allelic	The genome is made up of a fixed set of alleles that can exist in a finite or infinite number of variants, represented by integers or characters, and each individual has a fixed set of its n alleles	
Network	The individuals are characterized by their position in a gene-regulatory network, a neural network or a metabolic network; there is no explicit DNA level, and nodes are identified by their connections or the node numbers in the network	<ul style="list-style-type: none">• The relationship of robustness to mutations and to noise¹⁵⁴• The evolution of communication, cooperation and altruism^{89,155}
String-of-pearls	The genome is a variable-length string of ‘pearls’ of different types: phenotype genes, transcription factor genes, repeats, retrotransposons, binding sites, and so on; each pearl type can exist in a predefined number of variants; gene number, order and regulation can evolve through mutations and rearrangements	<ul style="list-style-type: none">• Genome and network evolvability^{141,143}• Resource processing in ecosystems⁸⁵• Sympatric speciation¹⁵⁶
Sequence-of-nucleotides	The genome is a variable-length string of characters; predefined signal sequences, analogous to promoters, terminators or start-stop codons, are used to detect genes; point mutations, indels and rearrangements can be simulated in a realistic manner	<ul style="list-style-type: none">• The evolution of non-coding DNA and gene number¹⁴⁰• The evolution of the size and topology of gene networks^{126,127}• Gene network inference^{157,158}

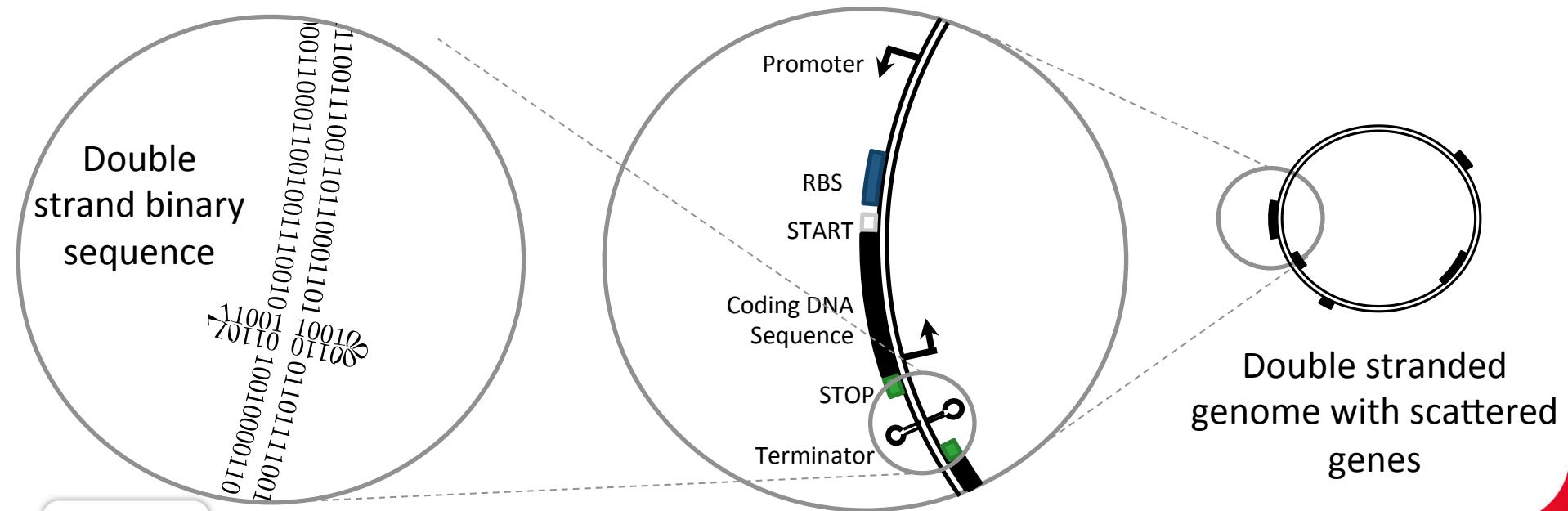
Aevol: The structure of the fitness landscape is (mainly? likely?) determined by the structure of the genotype-to-phenotype mapping and by the diversity of the mutational operators

*Many formalisms have been proposed to represent the genome, each with strengths and weaknesses. The appropriate formalism strongly depends on the question of interest. Here, we focus on the approaches that are most directly comparable to *in vivo* microbial evolution experiments (that is, approaches for which the genome comprises several genes).

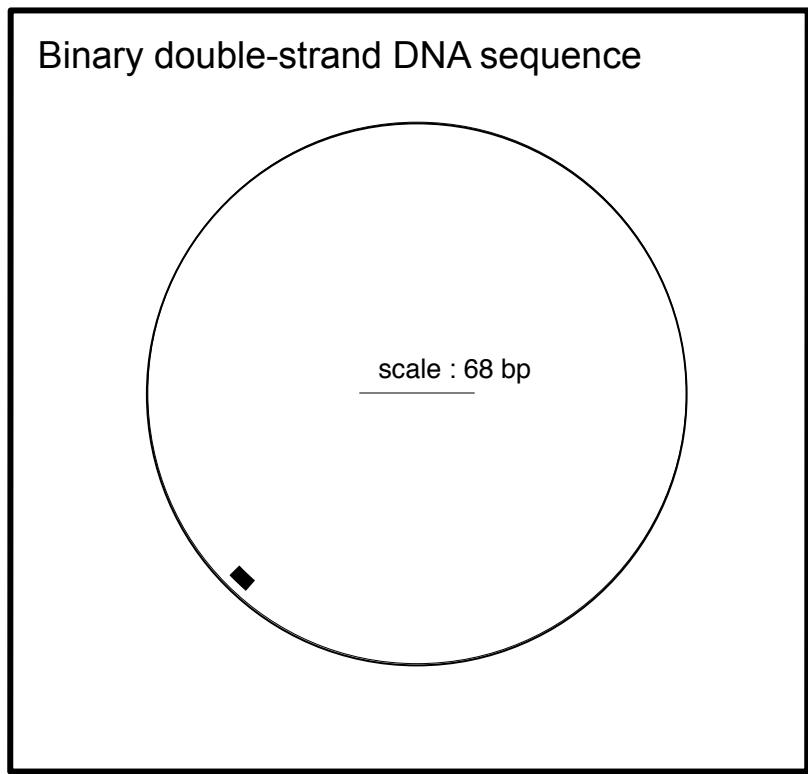
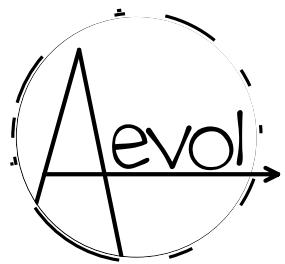
[Hindré, et al., *Nat. Rev. Microb.*, 2012]

Why the “sequence of nucleotides” formalism?

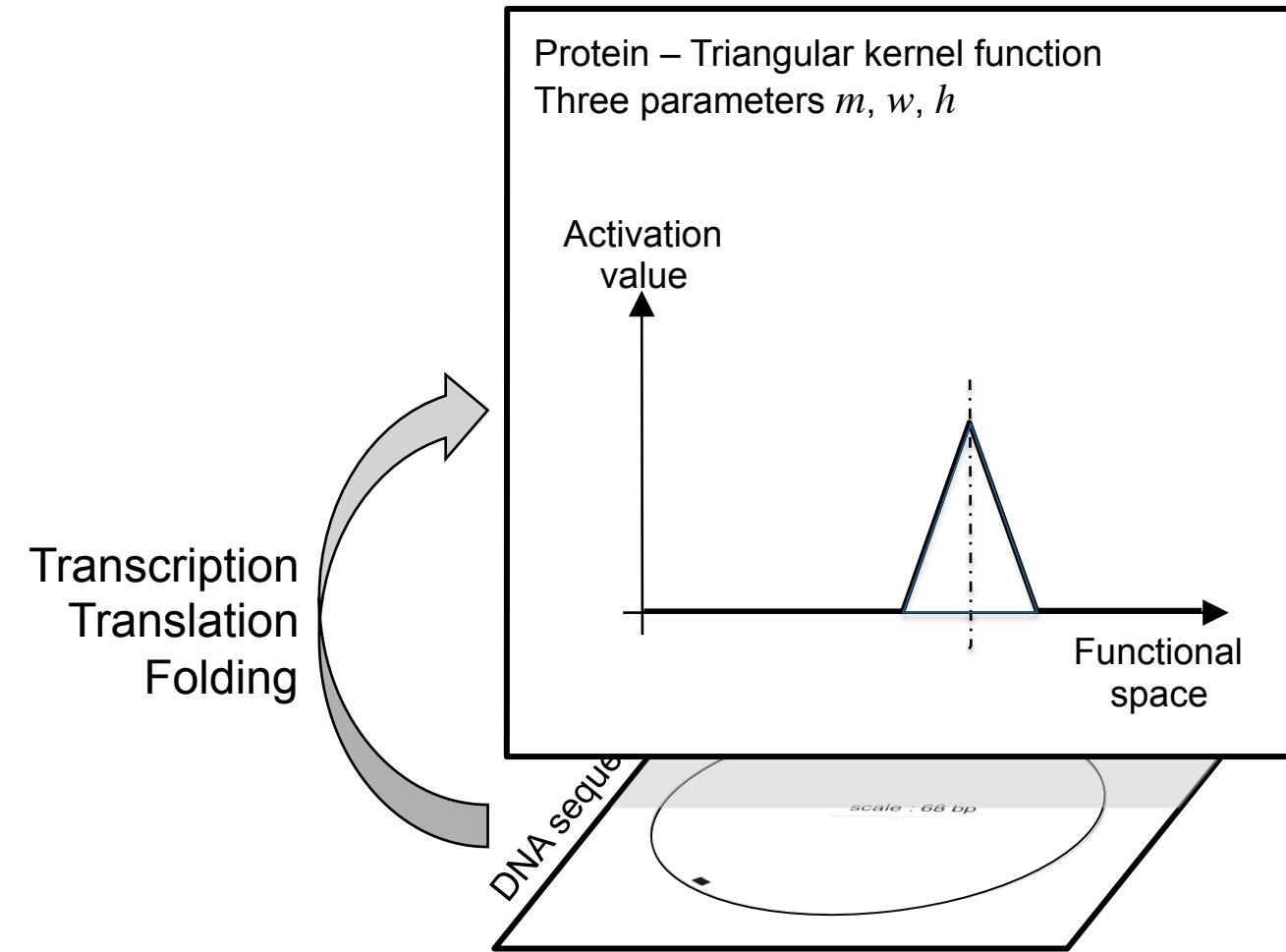
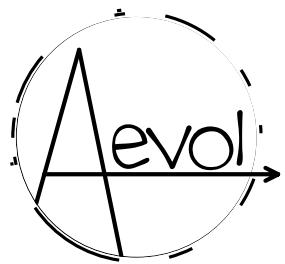
- The fate of evolution depends on the way information is stored and on the way it is modified
 - In the sequence of nucleotides formalism, the genotype-to-phenotype map is directly inspired from the “central dogma”
 - Mutations can be accurately modeled and their effect is filtered by the “genotype-to-phenotype map”; Hence many mutations are neutral



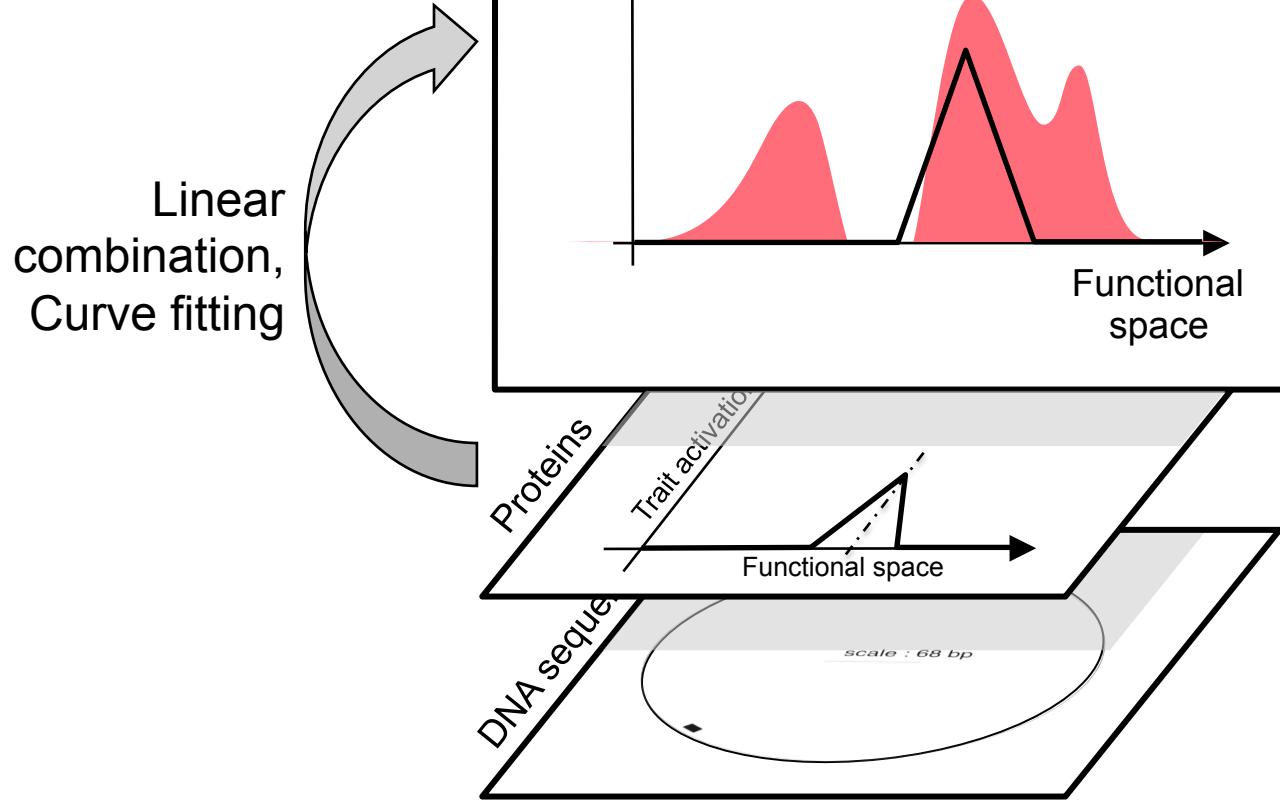
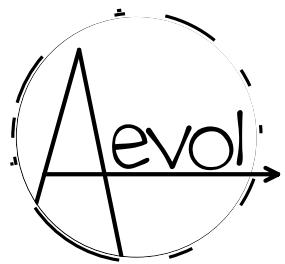
The Aevol platform



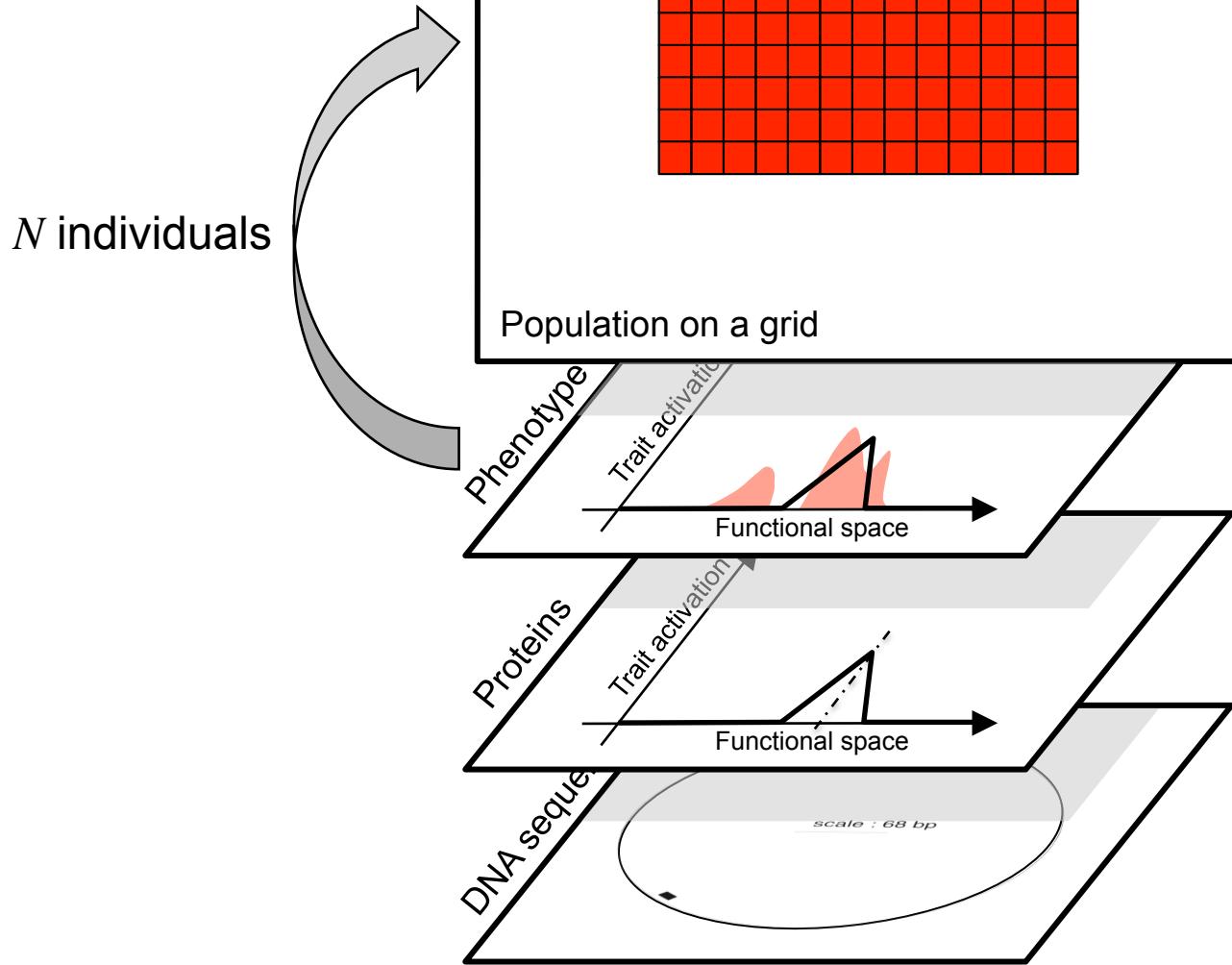
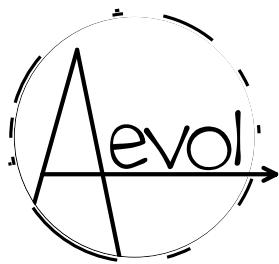
The Aevol platform

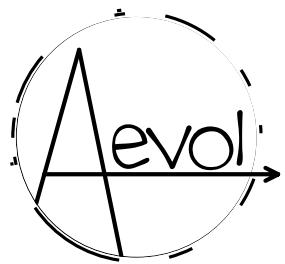


The Aevol plate



The Aevol plate





The Aevol platform

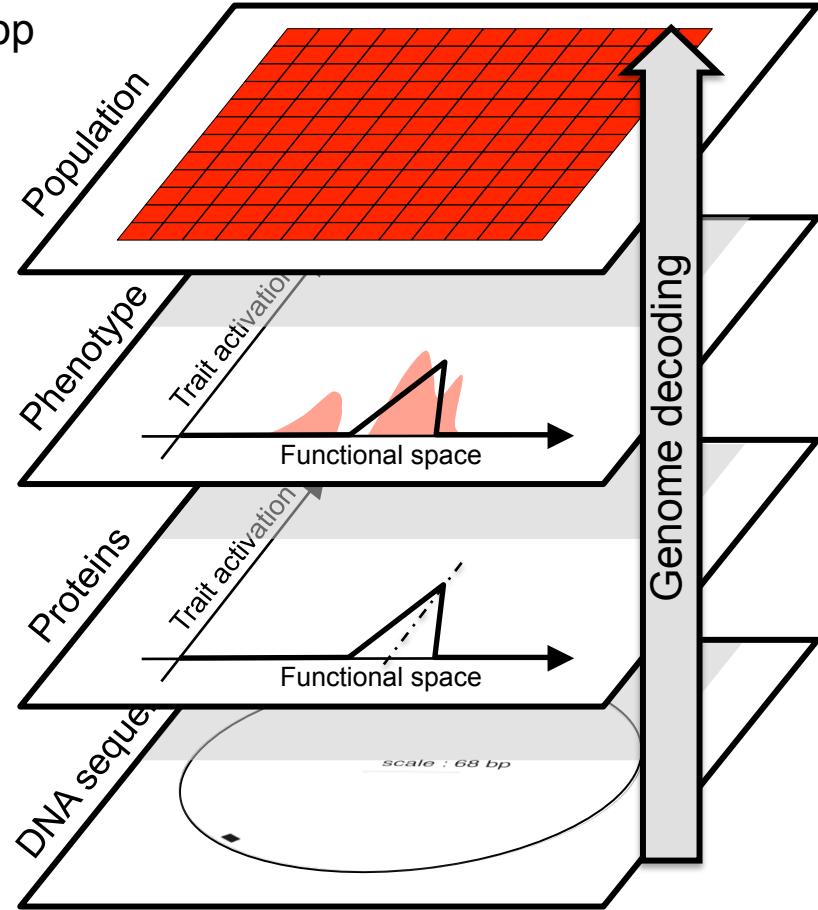
Generation 0

Clonal population

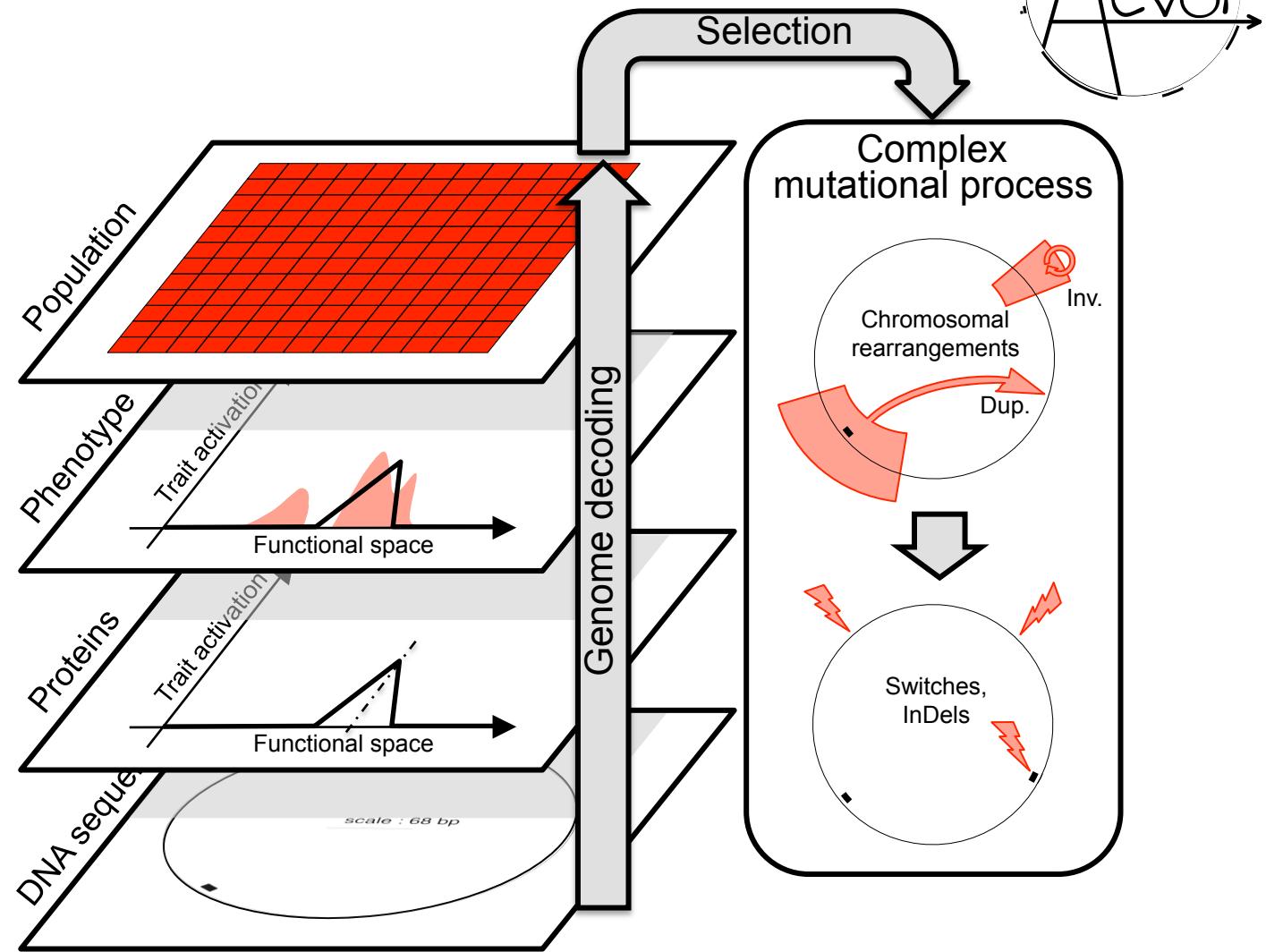
Genome size: 1000 bp

Gene number: 1

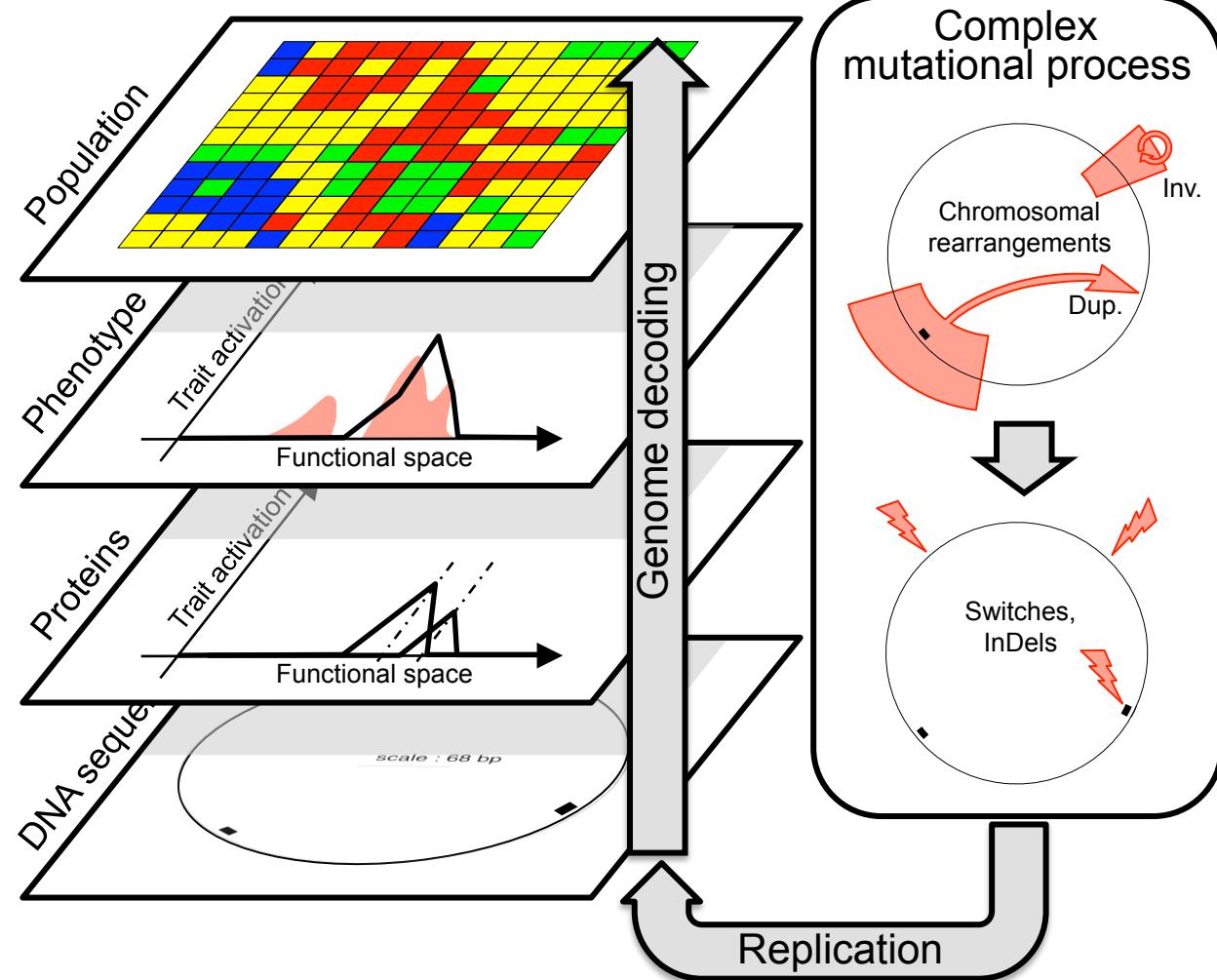
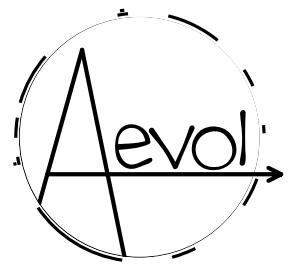
Fitness: 10^{-129}



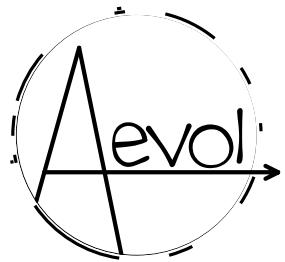
The Aevol platform



The Aevol platform



The Aevol platform



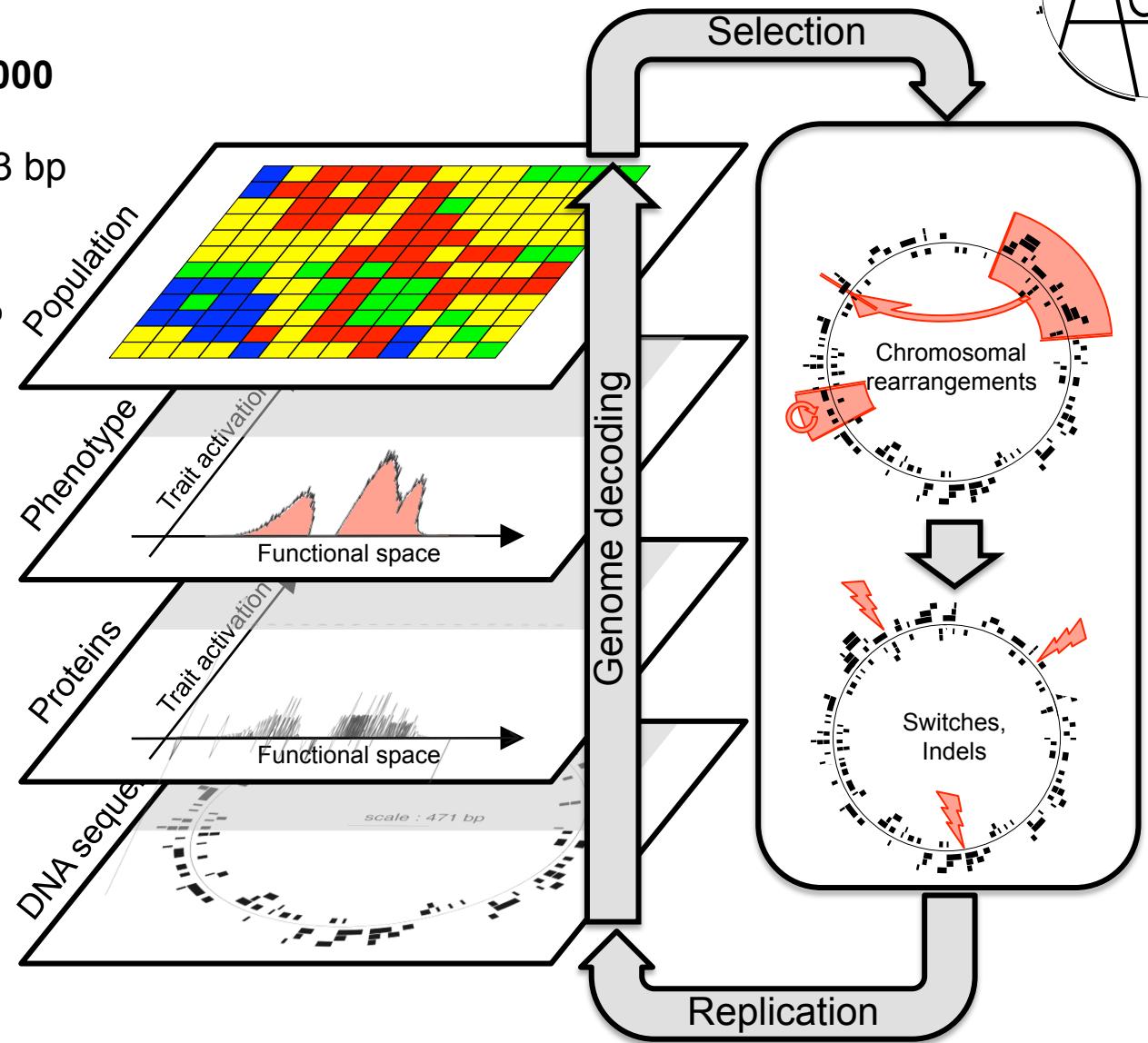
Generation 10,000,000

Genome size: 14 623 bp

Gene number: 137

Fitness: 0.012

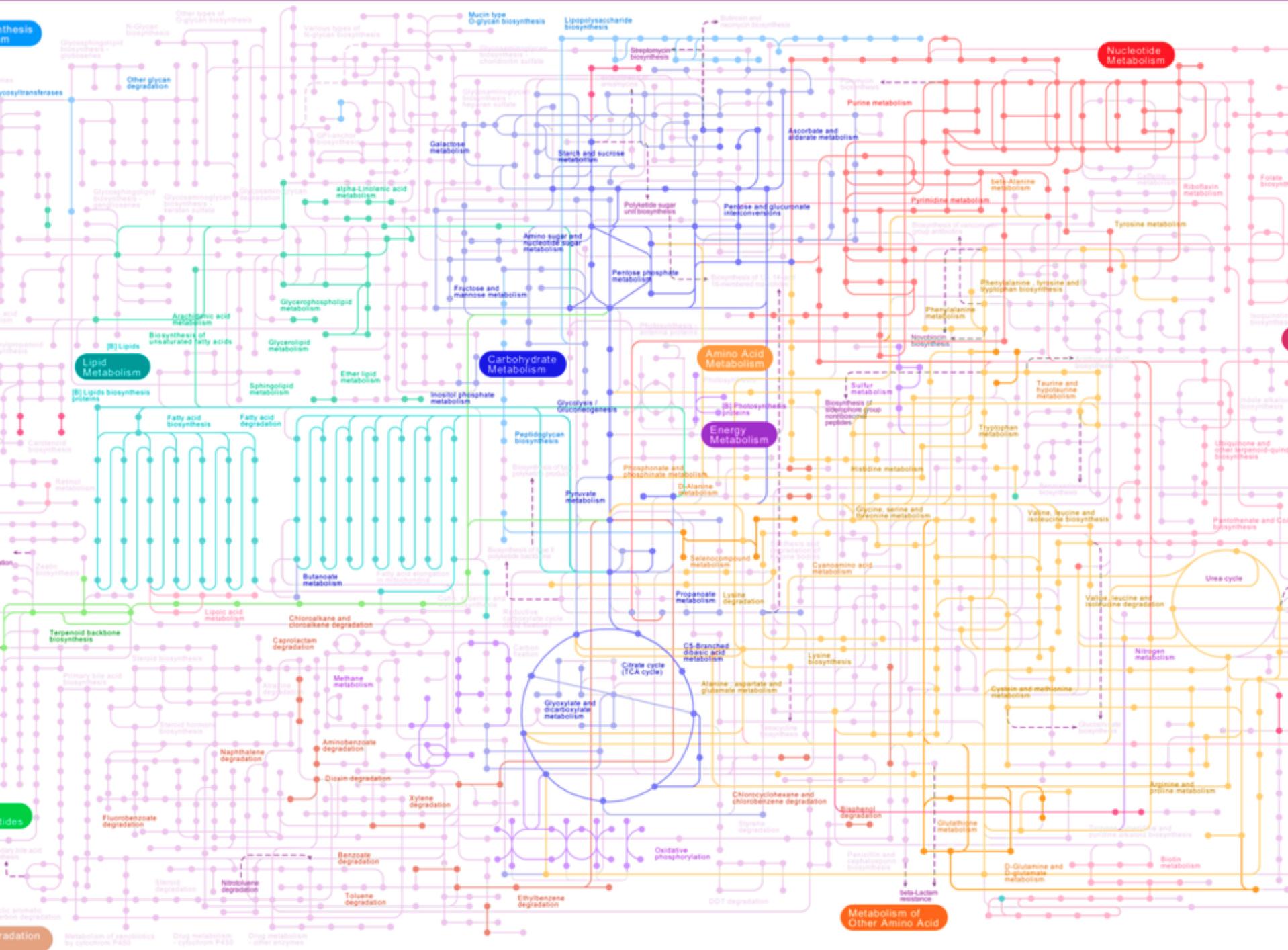
Coding fraction: 71%



Demo time (1)

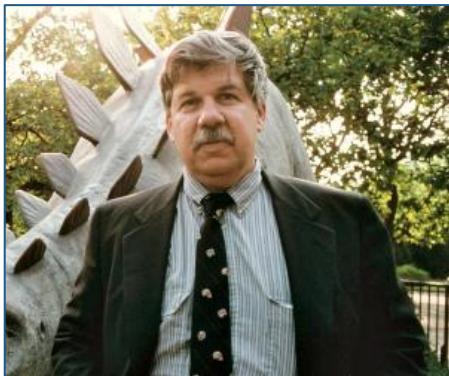
Conclusion: Aevol enable to study

- Evolution of genome and regulation networks (with R-aevol)
 - Under the influence of mutations rates and patterns, population size, environnemental variations, ...
- Using Aevol we (and others) have been able to study
 - Effect of mutation rate on genome robustness (Knibbe et al., 2007)
 - Effect of mutation rate on the complexity of regulation networks (Beslon et al., 2010)
 - Influence of genome structure on the maintenance of cooperative strategies (Frenoy et al., 2013)
 - Influence of mutator phenotype on genome structure (Rutten et al., 2019)
 - Undirect effect of population size on the amount of non-coding sequences (*in prep*)
 - Dynamics of innovation in viruses (*in prep*)
 - ...



Origin of biological complexity

- A recurrent matter of debate
- Two main families of explanations
 - Passive (the simple consequence of mutation accumulation)
 - Selective (complex organisms are better adapted)



Stephen Jay Gould:
“Drunkards walk model”
(or more recently Brandon
& McShea “ZFEL”)

→ How can we discriminate both explanations? Or propose others?



Richard Dawkins:
“Progressive increase in [...] complexity is to be expected only in taxa [that benefit from] complexity.”

Origin of biological complexity

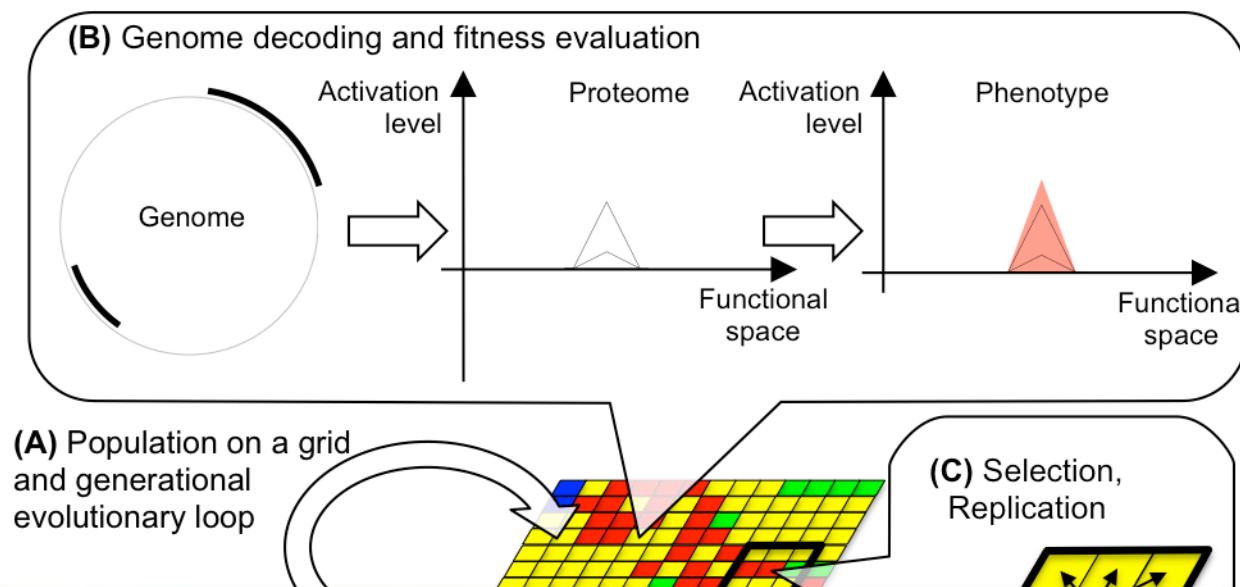
- Why is it important?
 - The origin of complexity is the most basic question in Evolutionary Systems Biology
 - Understanding the evolutionary origin of complexity can provide an analysis framework to understand extant organisms
- Why is it difficult?
 - Complexity is ill-defined
 - *What to measure and how to measure it?*
 - Complexity arose a very long time ago (millions/billions of years)
 - *phylogenetic studies are (at least) very difficult*
 - “All” extant organisms are complex
 - *Comparative approaches are useless*
 - All theories are based on thought experiments
 - *We need experimental tools to test our thought experiments...*

An “impossible experiment” to study the evolution of complexity

- Let populations of simple organisms evolve in situations where there is no selective pressure for complexity
 - “Impossible experiment” (with real organisms)
- Let's turn to artificial organisms and to Aevol
 - In Aevol, complexity can evolve thanks to rearrangements and gene duplication
 - In Aevol, complexity can be measured at different levels:
 1. Sequence level → genome
 2. Functional level 1 → proteome
 3. Functional level 2 → phenotype
 4. Selective level → fitness

Experimental setup

- Modified Aevol: the phenotypic target is a “simple” function
 - i.e. a function that can be fitted by a single gene/protein
→ the target is a single triangle with $m = 0.5$, $w = 0.1$, $h = 0.5$
 - Three different mutation rates (10^{-4} ; 10^{-5} ; 10^{-6} mut.bp $^{-1}$.gen $^{-1}$)
 - 100 repetitions per mutation rate, 250,000 generations
- Control experiment (the target is a Gaussian function)
 - Control requires an “infinite” complexity



Complexity measures

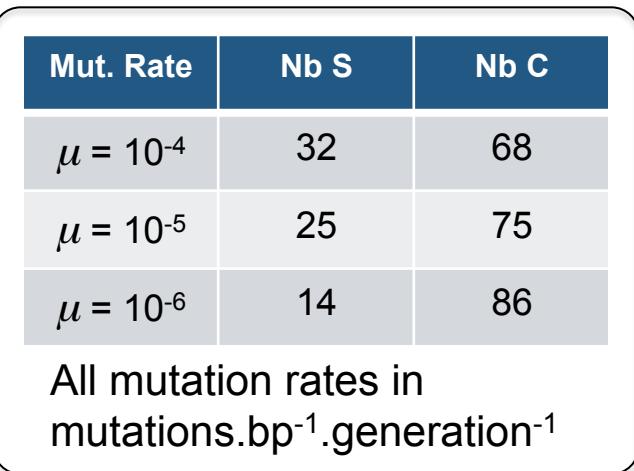
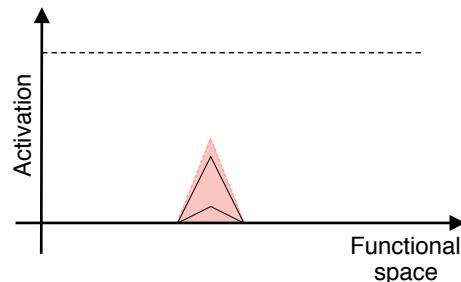
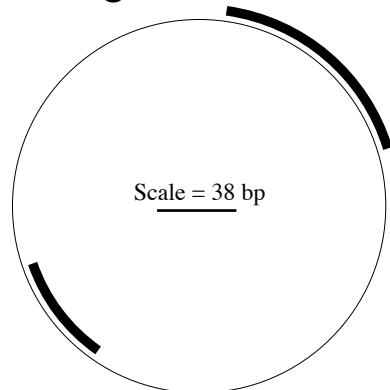
- Following Adami (2000), we measure complexity by the quantity of information stored in a biological structure
- Quantitative measures:
 - **Genomic complexity:** C_G
The quantity of information is equal to the number of bp which mutation would change the phenotype (“essential genome”)
 - **Functional complexity:** C_P
The quantity of information is equal to the number of parameters needed to describe the proteome (i.e. number of different m , w , h values)
- Qualitative measure:
 - **Simple (S) vs Complex (C) organisms**
An organism is considered “simple” if all its proteins contribute to the same function (possibly with different activation levels) – in mathematical terms, when the sum of its triangular kernels is a triangular kernel

Demo time (2)

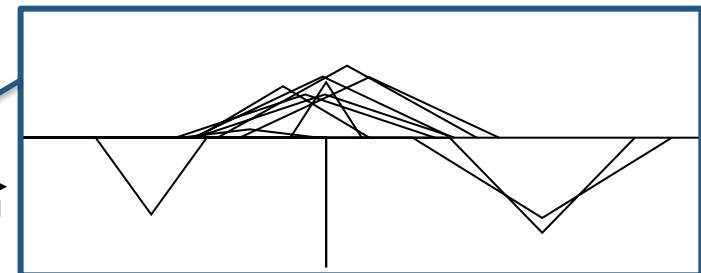
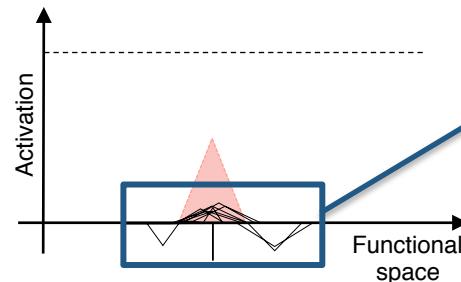
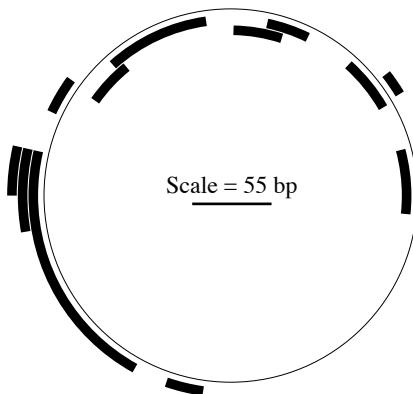
Results

Even in a simple environment, ~2/3 of the populations evolve complex functional structures

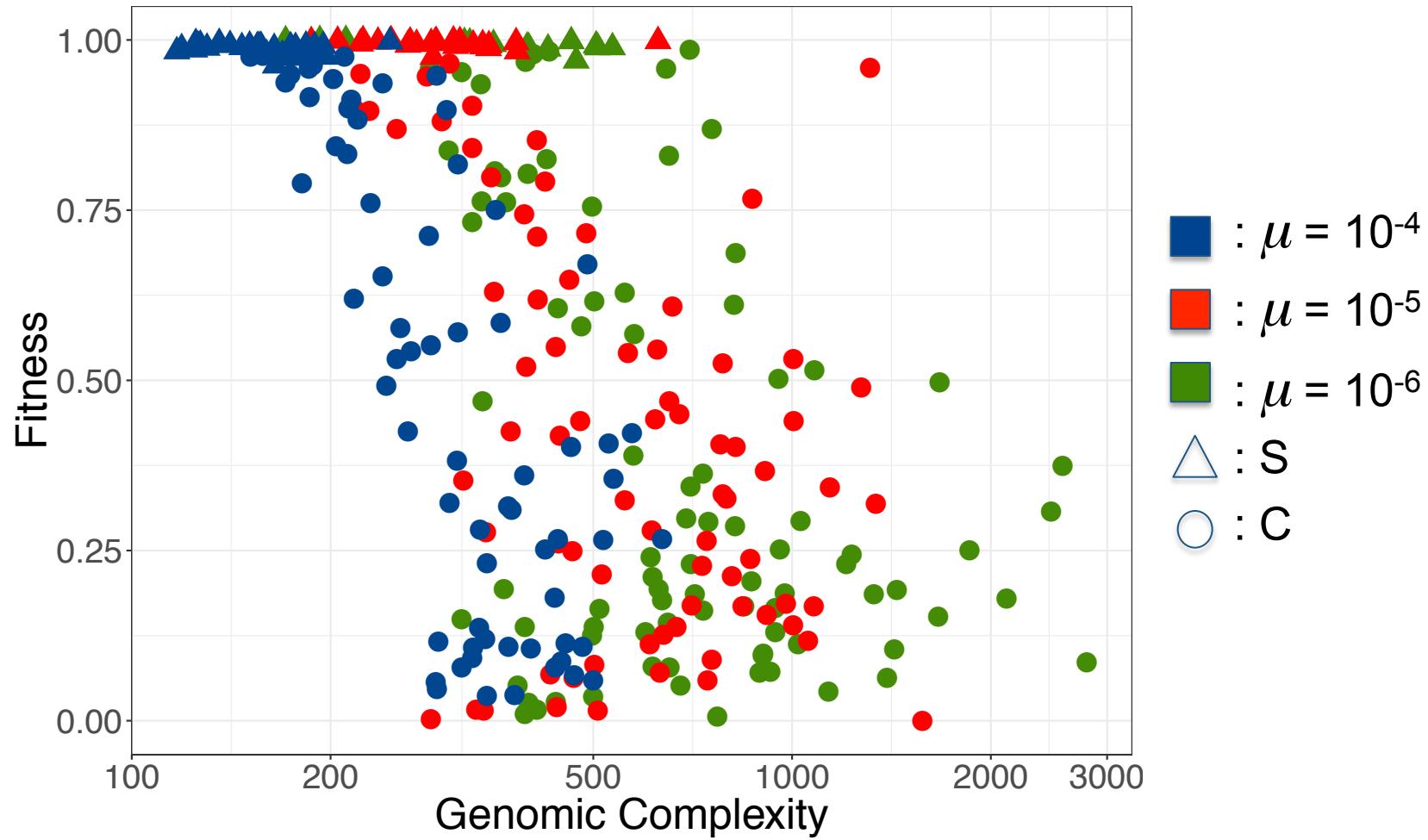
Simple organism



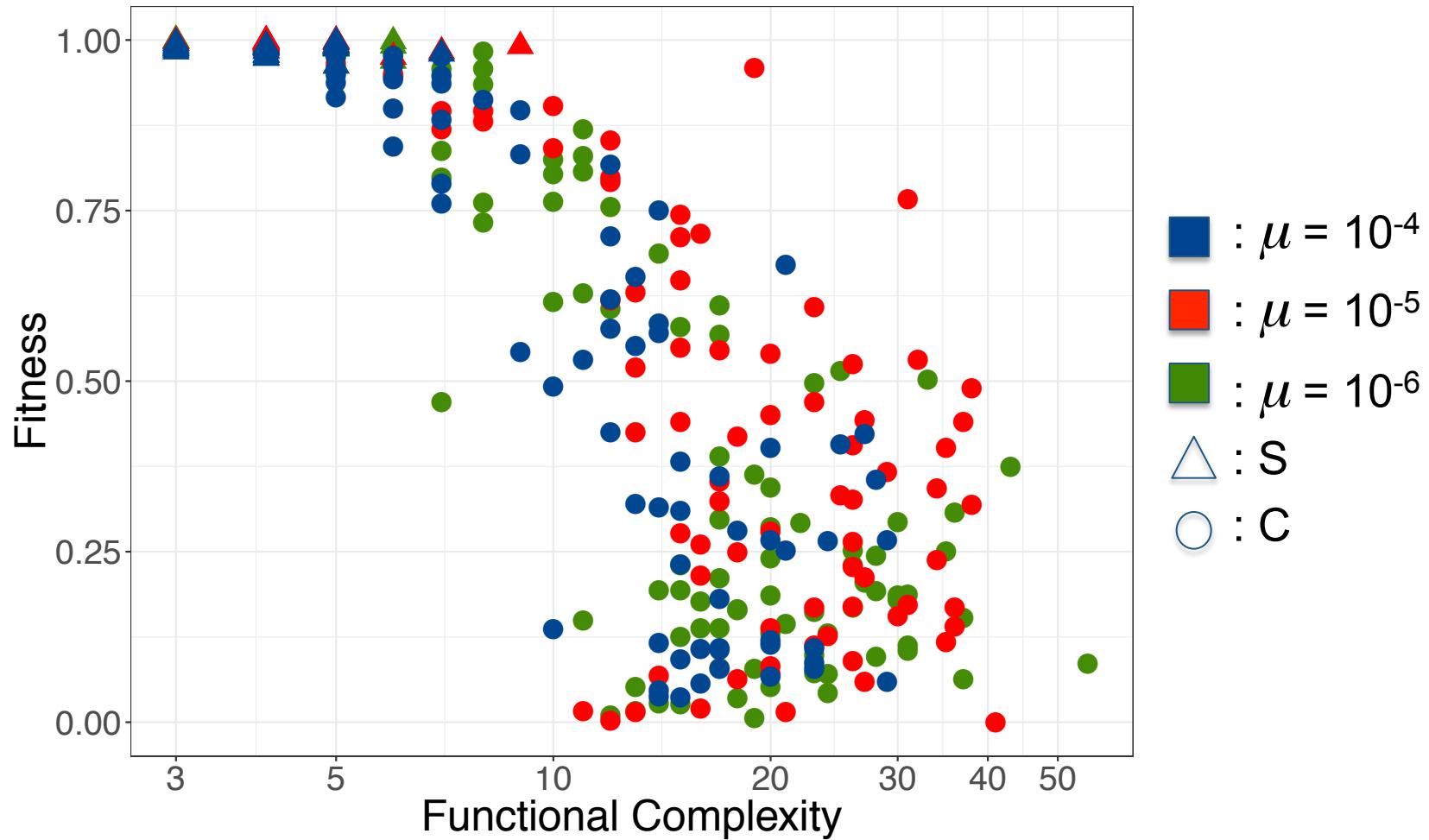
Complex organism



But simple organisms are better than complex ones whatever the complexity measure



But simple organisms are better than complex ones whatever the complexity measure

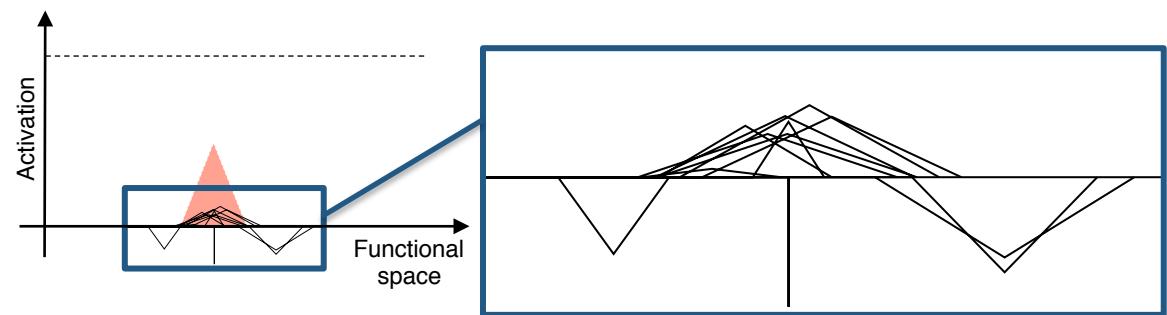
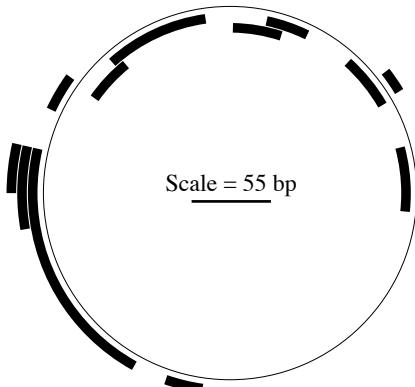


What if the environment requires complexity?

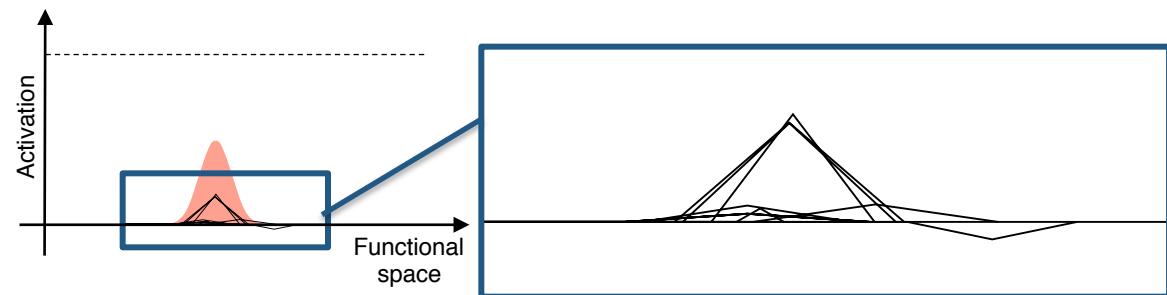
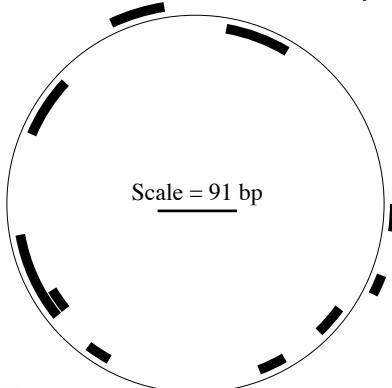
What if organisms evolve in complex environments?

- Comparison between organisms evolved in a simple environment (triangle target) and in a complex environment (Gaussian target)

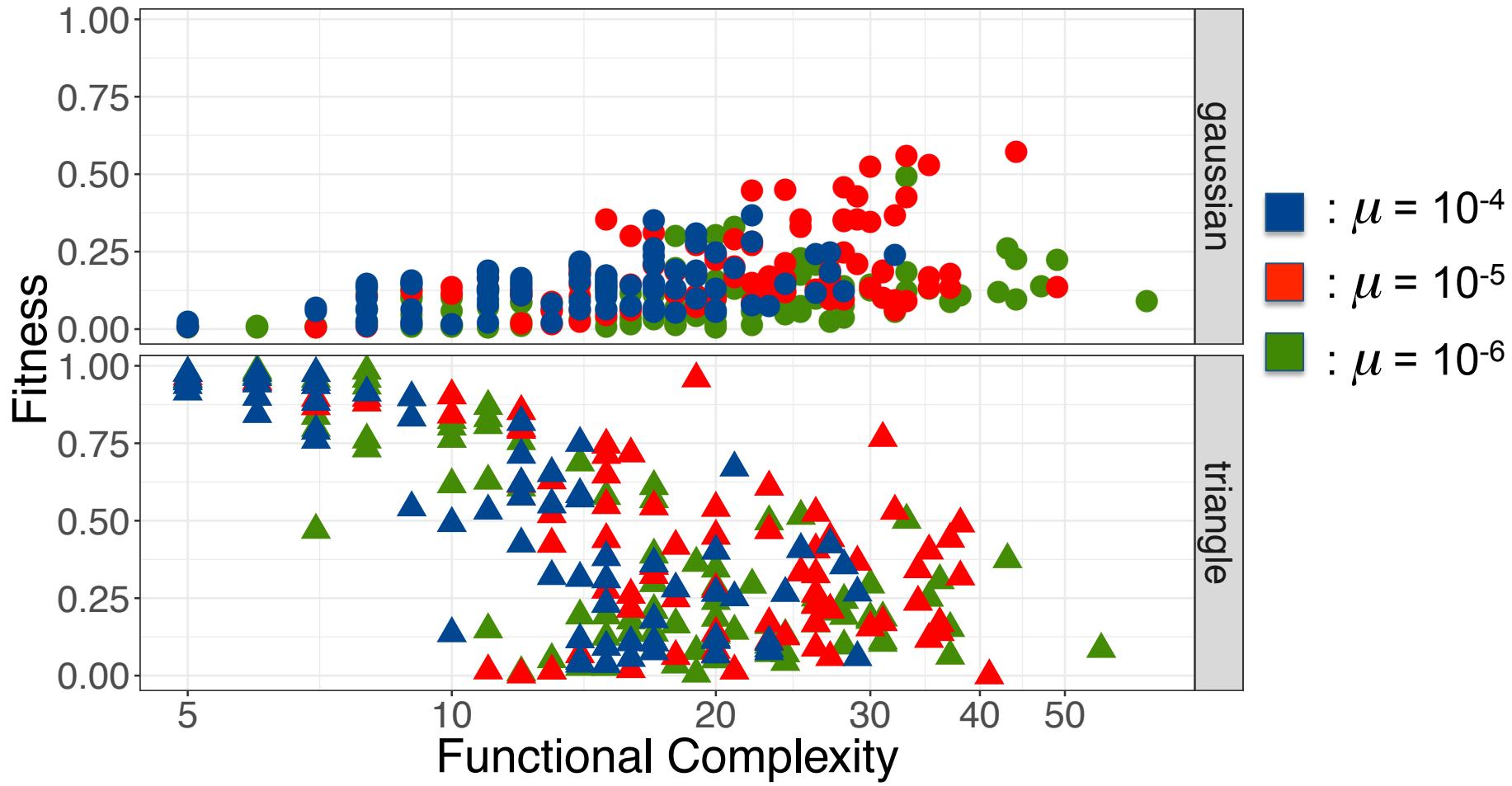
Simple environment (triangle target)



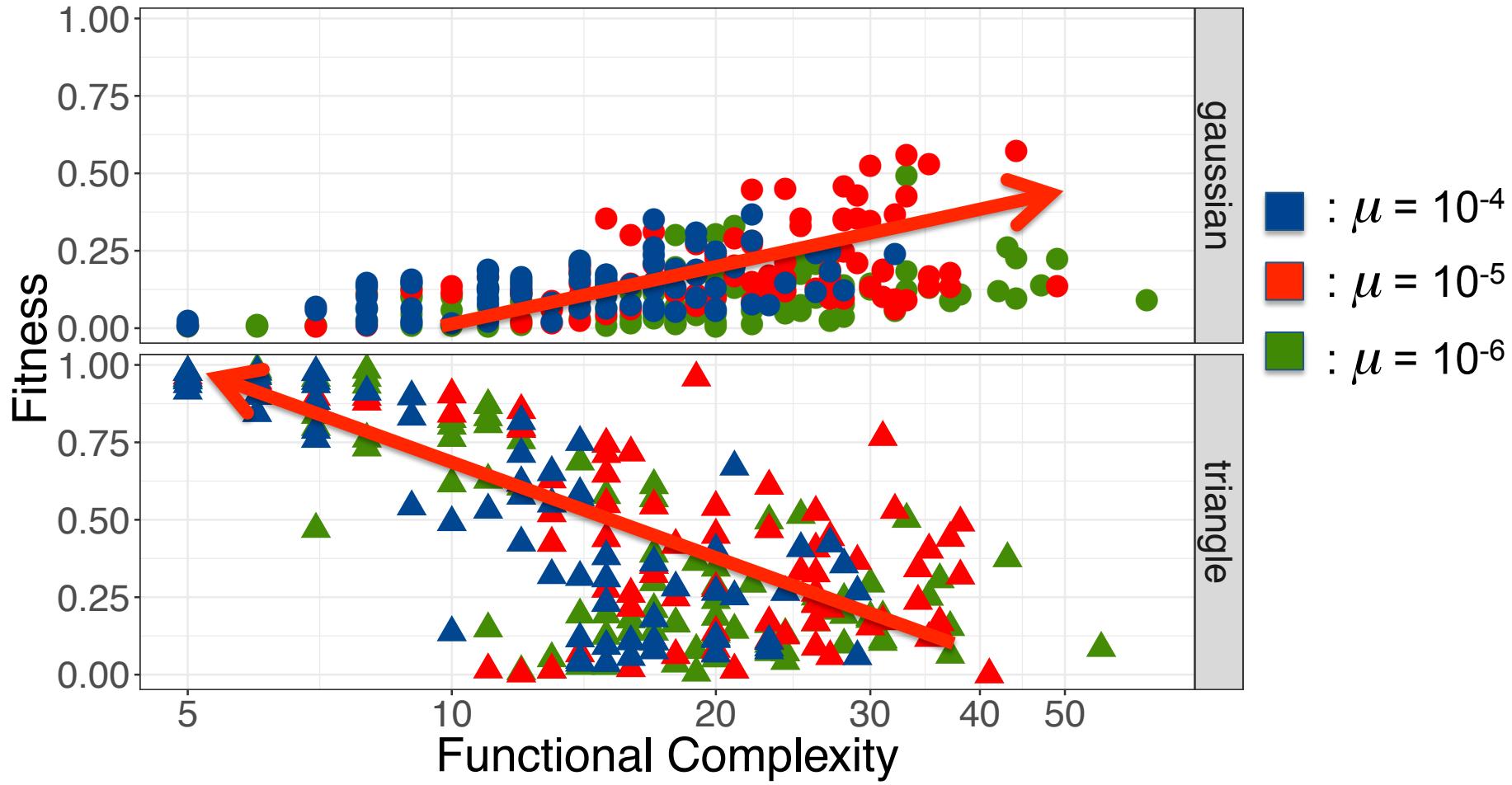
Complex environment (Gaussian target)



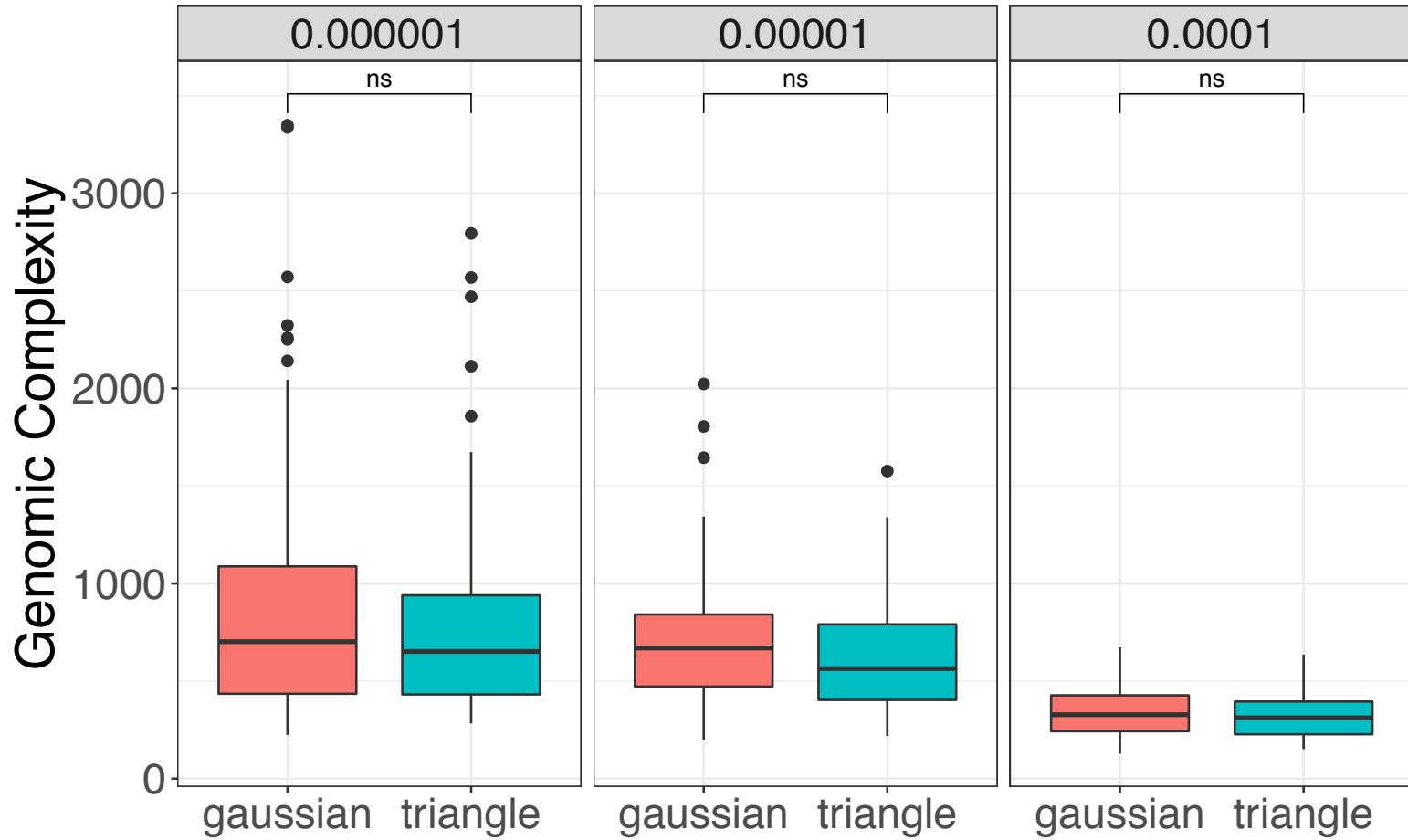
In complex conditions being more complex gives a fitness advantage (unlike simple conditions)



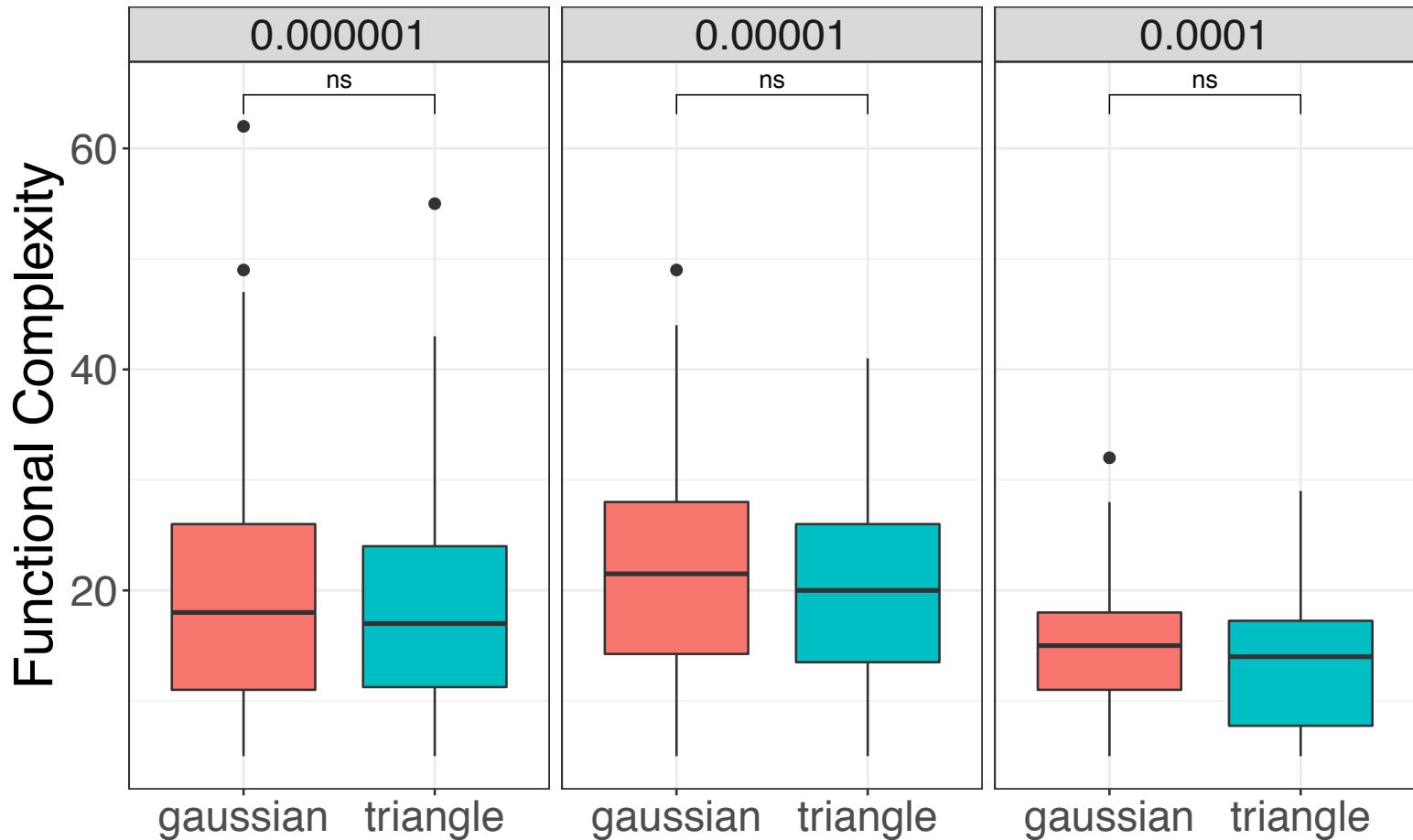
In complex conditions being more complex gives a fitness advantage (unlike simple conditions)



... but organisms evolving in complex environments are not more complex than those envolving in simple ones!



... but organisms evolving in complex environments are not more complex than those evolving in simple ones!



Complexity is NOT driven by selection...

→ *Is it driven by indirect selection*?*

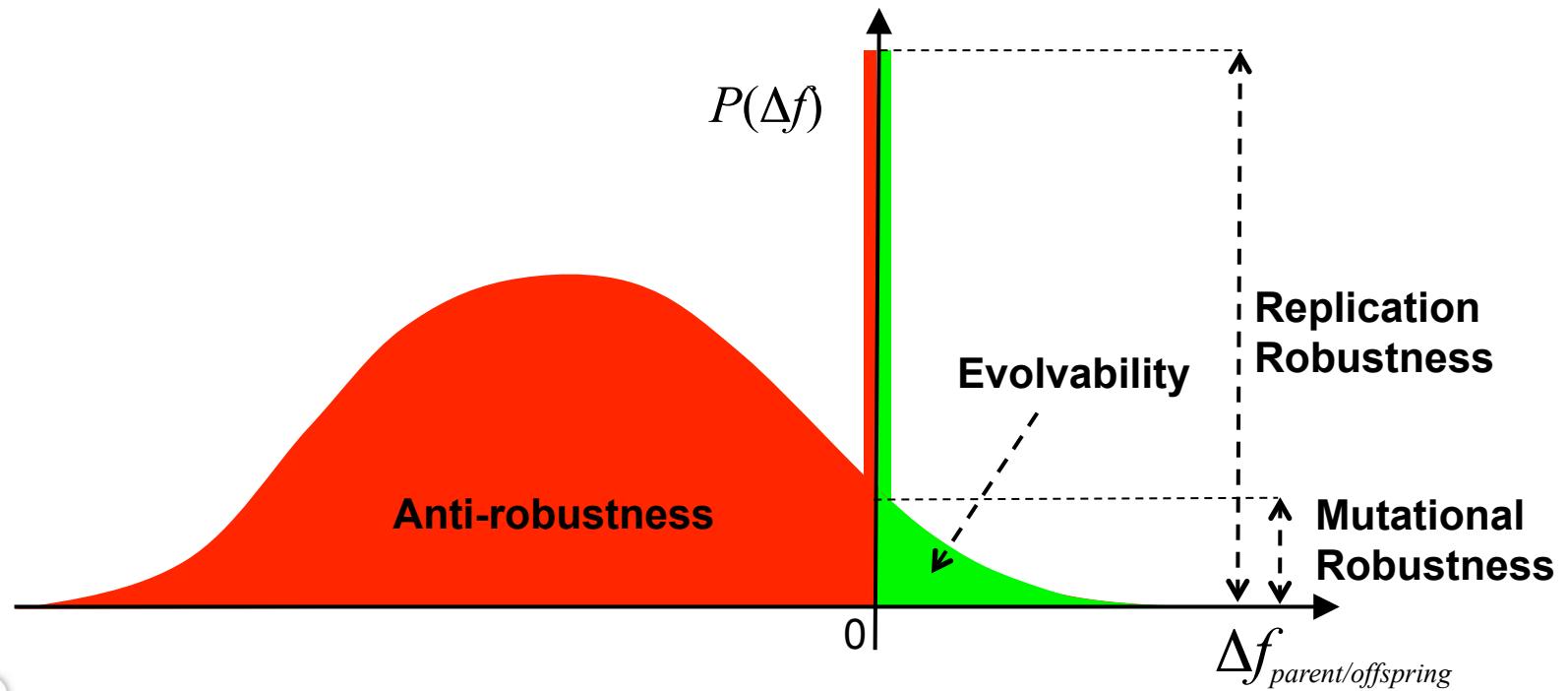
*** i.e. selection for robustness or selection for evolvability**

Evolvability, Robustness, Anti-robustness: Tentative definitions

- **Evolvability:**
 - Expected degree to which a lineage will increase in fitness after evolving for a certain time in a particular environment (Woods et al., 2011) → Expectancy of fitness gain
- **Mutational Robustness:**
 - Expected degree to which a lineage will remain constant in fitness after evolving for a certain time in a particular environment → Proportion of neutral mutants
- **Replication Robustness:**
 - Expected degree to which a lineage will remain constant in fitness after evolving for a certain time in a particular environment → Proportion of neutral offspring
- **Anti-Robustness:**
 - Expected degree to which a lineage will loose fitness after evolving for a certain time in a particular environment → Expectancy of fitness loss

Evolvability, Robustness, Anti-robustness: Tentative definitions

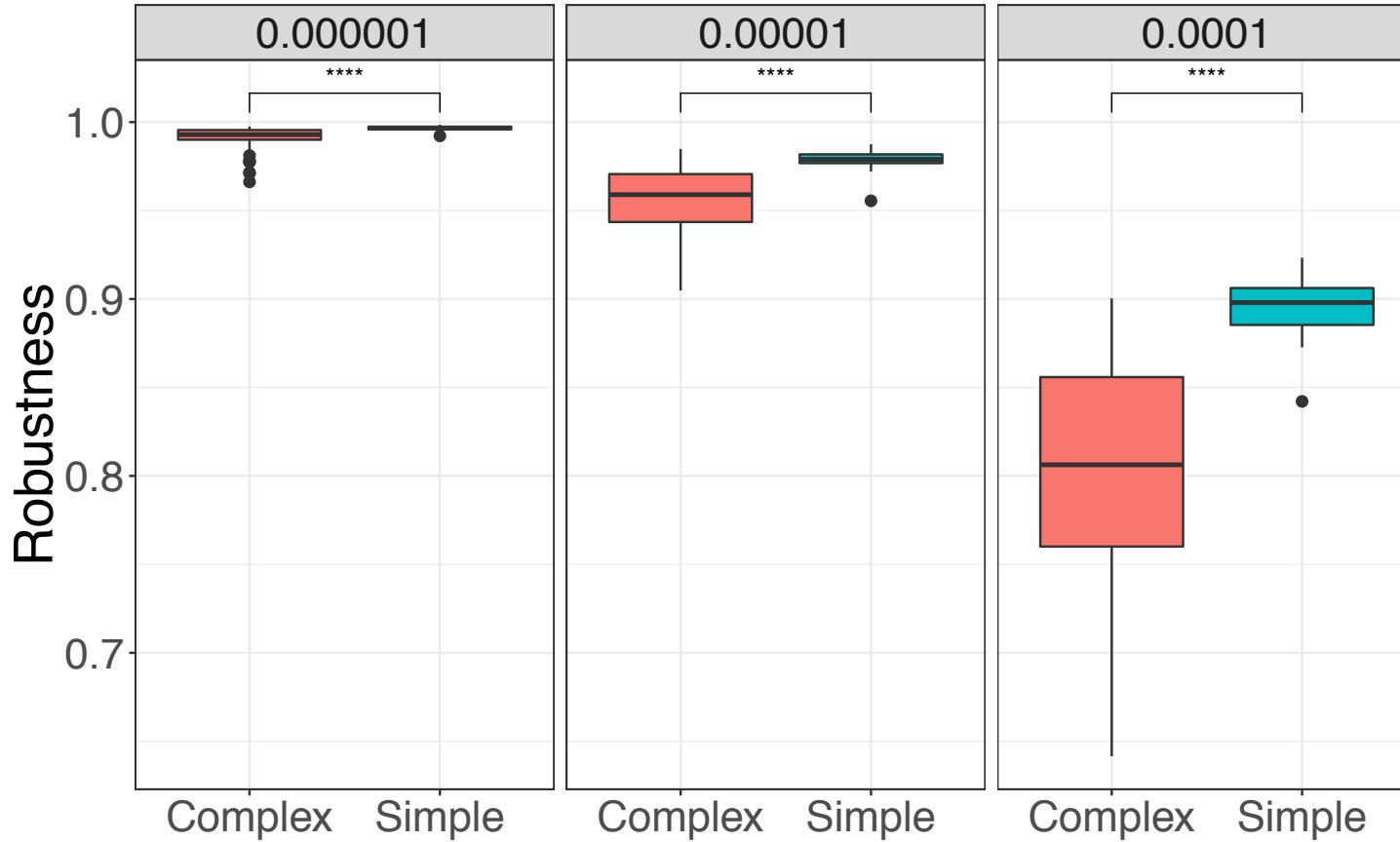
- Evolvability, Robustness and anti-robustness can be estimated from the distribution of offspring fitness compared to parents fitness
 - Almost impossible to measure *in vivo*
 - Can be “easily” estimated *in silico* by monte-carlo sampling



Estimation of evolvability and robustness in Aevol

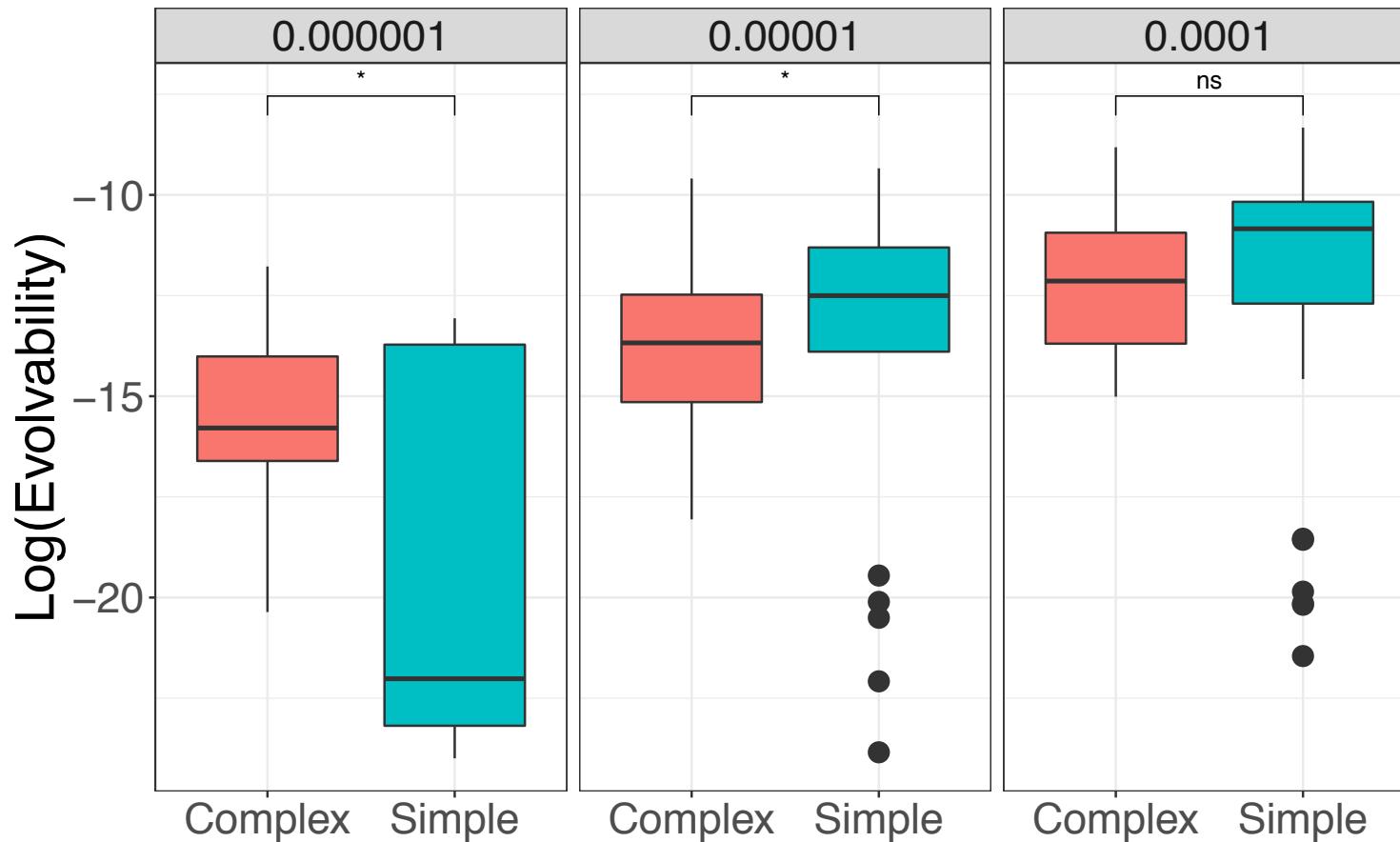
- For each of the 300 evolved genotypes, we generated 10,000,000 offspring in the same experimental conditions
- From the 10,000,000 offspring we can estimate
 - Fraction of neutral offspring → Replication Robustness
 - Fraction of neutral mutants → Mutational Robustness
 - Expectation of fitness loss → Anti-robustness
 - Expectation of fitness gain → Evolvability
- This procedure gives a sketch of the local structure of the fitness landscape around the tested genome...
 - In Aevol the fitness landscape has very high number of dimensions and the number of dimensions changes as the number of genes and the size of the genome change (i.e. as complexity changes)

Simple organisms are more robust than complex ones*



*: Replication robustness.

Complex organisms have no (no clear?) evolvability advantage*



*: Evolutionary potential.

Complexity is not driven by selection nor by indirect selection...

→ ***What is the long term fate of complex organisms in simple environments?***

Simplicity and complexity are stable identities...

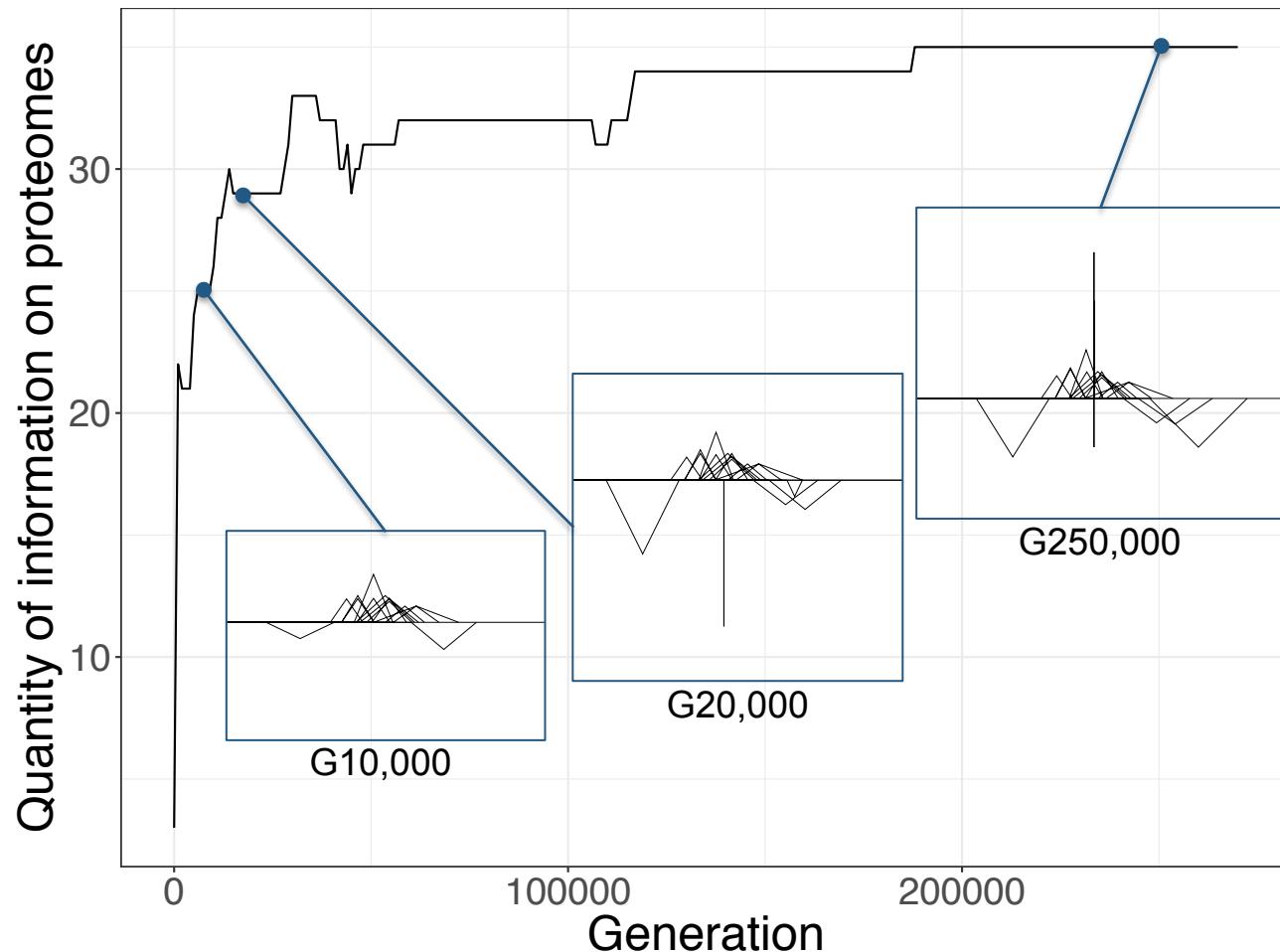
- Fraction of individuals that retain their S/C identity between generations 10,000 and 250,000

	$\mu = 10^{-4}$	$\mu = 10^{-5}$	$\mu = 10^{-6}$
$P_{S \rightarrow S}$	100% [100% – 87.9%] (28/28)	100% [100% – 85.7%] (23/23)	92.3% [98.6% – 66.7%] (12/13)
$P_{C \rightarrow C}$	94.4% [97.8% – 86.6%] (68/72)	97.4% [99.3% – 91%] (75/77)	97.7% [99.4% – 92%] (85/87)

→ Identity is acquired early during evolution and conserved afterward...

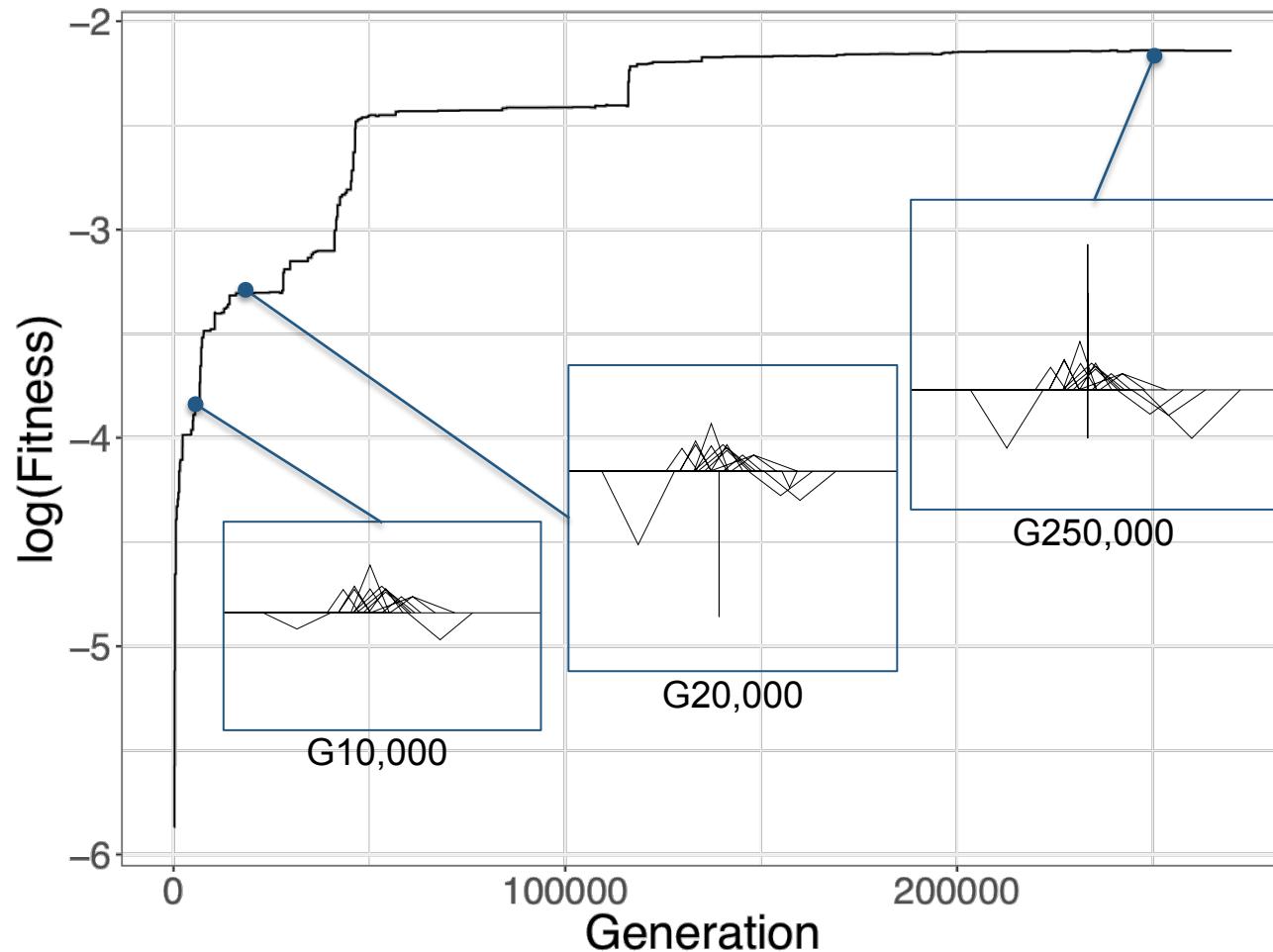
Complex individuals grow in functional complexity

- Long-term evolution of a complex individual



Complex individuals gain fitness meanwhile

- Long-term evolution of a complex individual



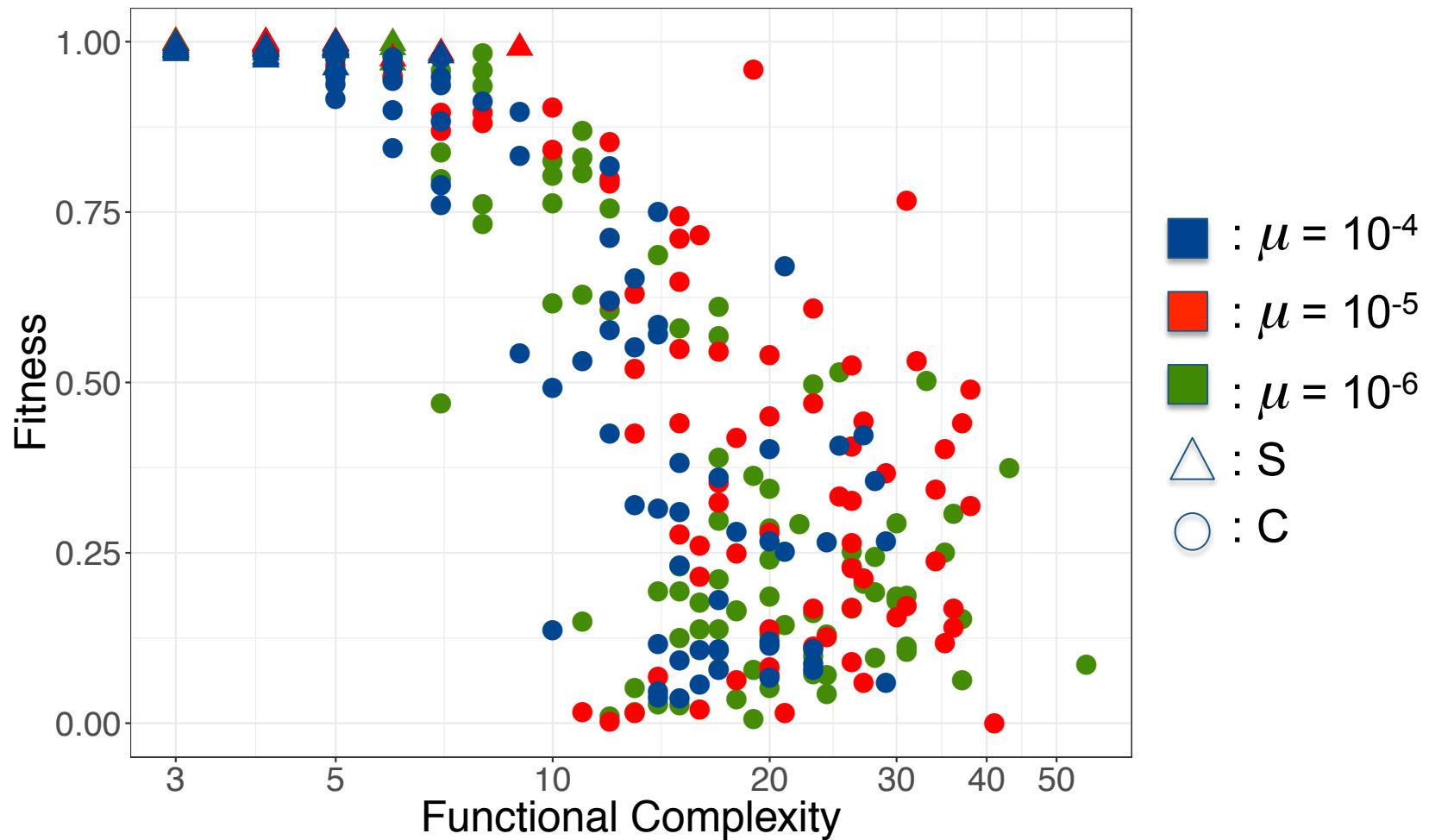
Complex individuals grow in fitness and in complexity

- Variation of complexity and fitness for individuals that retain their S/C identity between generations 10,000 and 250,000

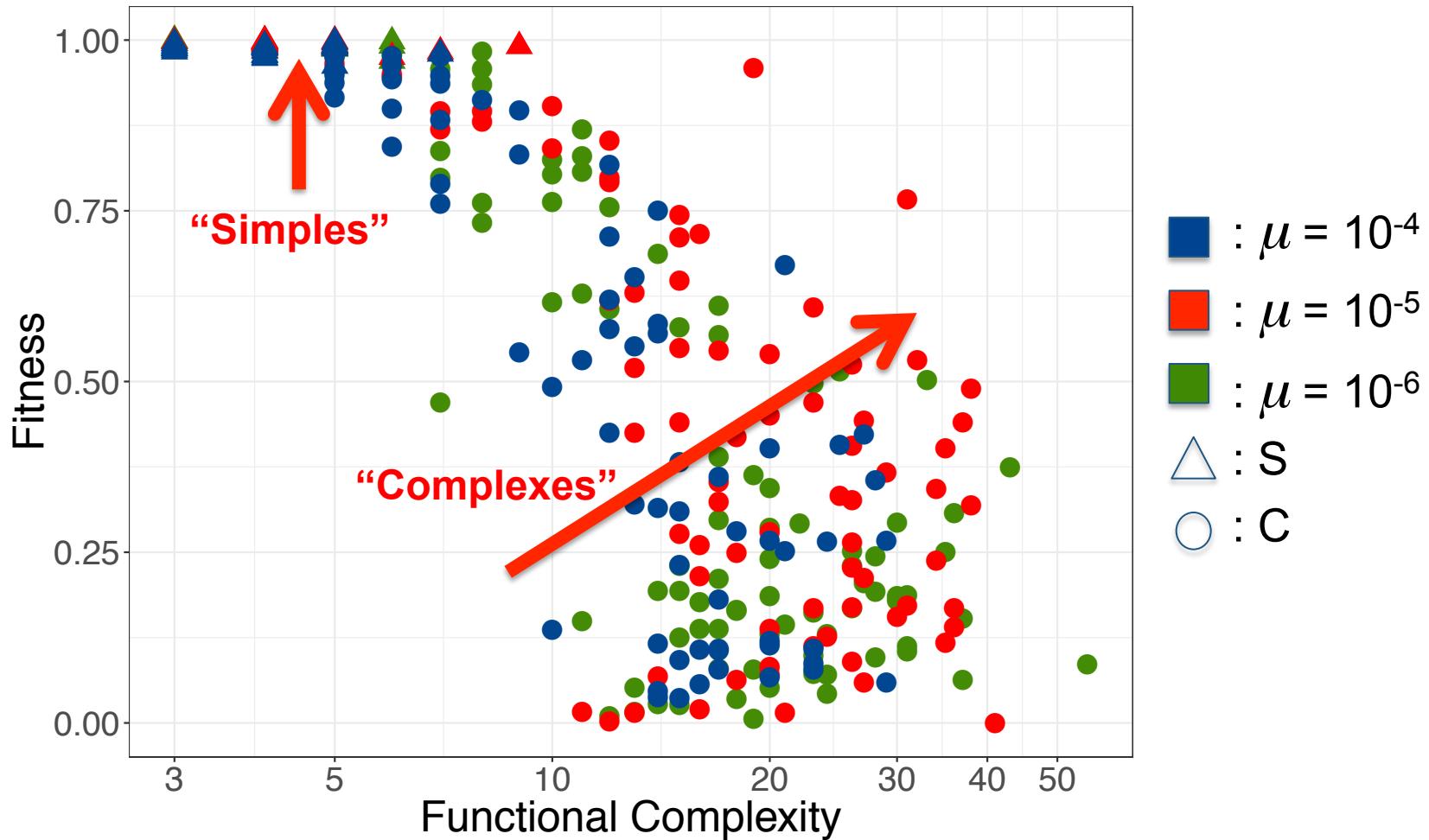
	$\Delta\mathcal{C}_G$	$\Delta\mathcal{C}_P$	$\Delta\text{Fitness}$
$S \rightarrow S$	-43.8 ± 2.2	-0.32 ± 0.31	$+0.09 \pm 0.1$
$C \rightarrow C$	$+25.3 \pm 1.62$	$+3.58 \pm 0.27$	$+0.16 \pm 0.05$

- Complex individuals grow in genomic and functional complexity
- Simple individuals decrease in genomic complexity
- Complex individuals gain fitness

But in simple environments simple organisms are better than complex ones whatever the complexity measure



But in simple environments simple organisms are better than complex ones whatever the complexity measure



Discussion

- In Aevol we observe a “complexity ratchet”
 - *Complex organisms increase their complexity even in simple environments*
 - In Aevol the complexity ratchet is stronger than selection
 - *Complex organisms have lower fitness and lower robustness than simple ones*
 - But in Aevol the complexity ratchet is fueled by selection
 - *Complex organisms improve while growing in complexity*
- **The complexity ratchet is enforced by sign-epistasis**
- Mutations/genes/proteins that would be favorable in the a *Simple* context are highly deleterious in a *Complex* context
- But...

Discussion

- In AevoL we observe a “complexity ratchet”
 - Complex organisms increase their complexity even in simple environments
- In AevoL the complexity ratchet is stronger than selection
 - Complex organisms have lower fitness and lower robustness than simple ones
- But in AevoL the complexity ratchet is fueled by selection
 - Complex organisms improve while growing in complexity
 - The complexity ratchet is enforced by negative epistasis
 - Mutations/genes/proteins that would be favorable in the a *Simple* context are highly deleterious in a *Complex* context
 - **But... can we generalize results obtained on a model?**

Discussion

- In Aevol we observe a “complexity ratchet”
 - Complex organisms increase their complexity even in simple environments

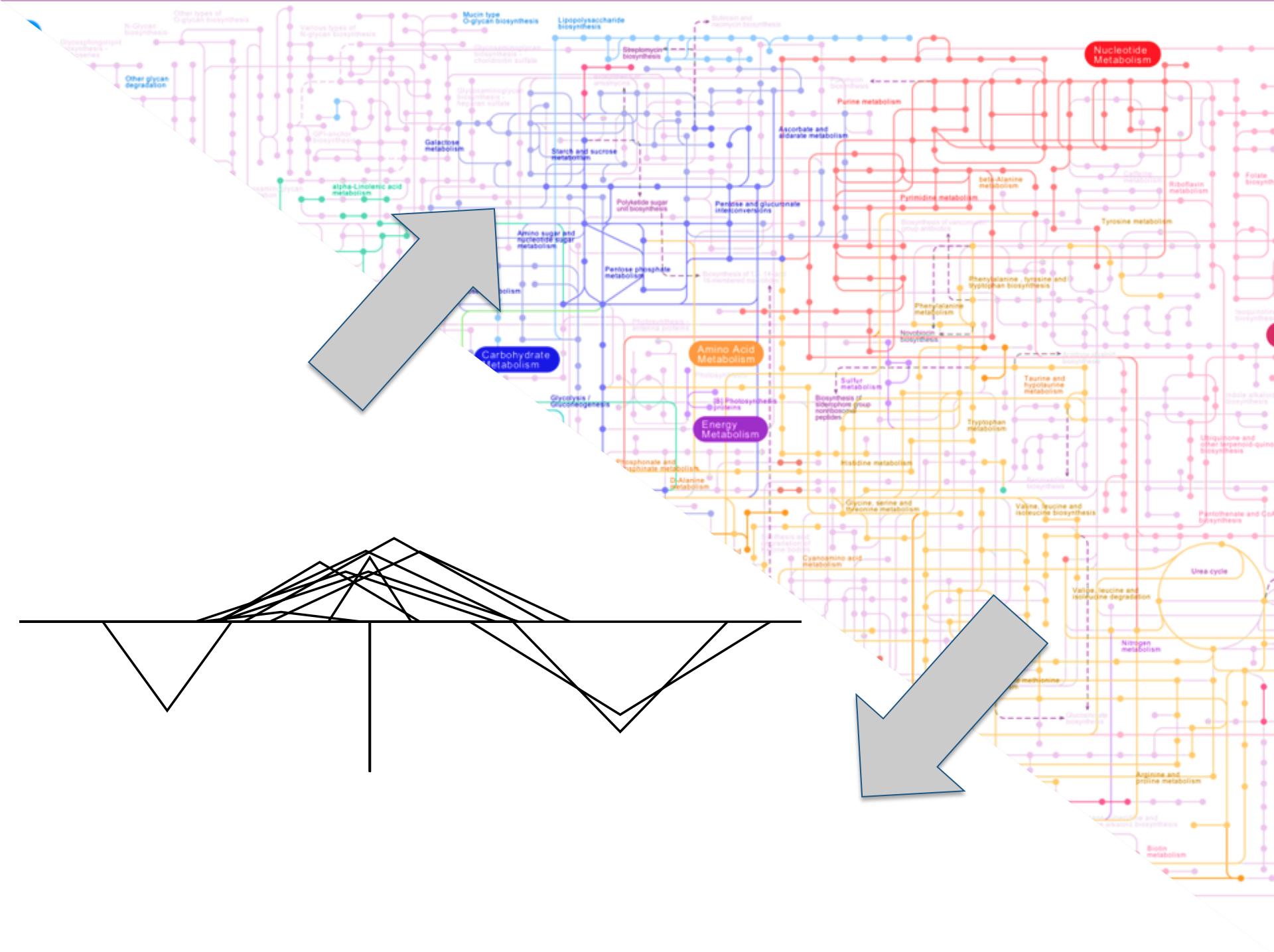
- In A

“The decisive thing with modeling is not the model per se, but what the model and working with the model does to our mind. [...] It could be argued that a criterion to determine good models is that they are no longer needed afterward”

[V. Grimm, *Ecological modelling*, 1999]

→ Mutations/genes/proteins that would be favorable in the a Simple context are highly deleterious in a Complex context

→ But... can we generalize results obtained on a *model*?



Conclusion

1. *in silico* experimental evolution enables experimental studies of macro-evolutionary phenomenon at the system level
 - *Evolution of complexity, of modularity, of genome structure...*
2. Complexity is likely to be due to the historical nature of the evolutionary process (“complexity ratchet”)
 - *Is it ubiquitous to all historical processes (eg. the making of the laws)? Is it similar to “path-dependence” (eg. QWERTY)*
3. Selection for complexity and/or mutational biases are not mandatory for complexity to evolve
 - *But this does not imply that there is no selection for complexity nor mutational biases toward complexity*
4. Complex (biological?) structures can flourish in conditions where complexity is not needed
 - *The global function of these complex structures could very well be simple*
→ *What are the consequences for systems biology?*

Further readings

- Liard, V., Parsons, D. P., Rouzaud-Cornabas, J., & Beslon, G. (2020). The complexity ratchet: Stronger than selection, stronger than evolvability, weaker than robustness. *Artificial Life*, 26(1), 38-57.
- Beslon, G., Liard, V., Parsons, D. P., & Rouzaud-Cornabas, J. (2020) "Of Evolution, Systems and Complexity." To appear in: "Evolutionary Systems Biology" Vol. 2
- Rutten, J. P., Hogeweg, P., & Beslon, G. (2019). Adapting the engine to the fuel: mutator populations can reduce the mutational load by reorganizing their genome structure. *BMC Evolutionary Biology*, 19(1), 191.
- Liard, V., Parsons, D., Rouzaud-Cornabas, J., & Beslon, G. (2018). The Complexity Ratchet: Stronger than selection, weaker than robustness. In: Artificial Life Conference Proceedings (pp. 250-257).
- Biller, P., Knibbe, C., Beslon, G., & Tannier, E. (2016). Comparative genomics on artificial life. In: Conference on Computability in Europe (pp. 35-44).
- Batut, B., Parsons, D. P., Fischer, S., Beslon, G., & Knibbe, C. (2013). In silico experimental evolution: a tool to test evolutionary scenarios. *BMC bioinformatics*, 14(S15):S11.
- Hindré, T., Knibbe, C., Beslon, G., & Schneider, D. (2012). New insights into bacterial adaptation through in vivo and in silico experimental evolution. *Nature Reviews Microbiology*, 10(5), 352-365.
- Knibbe, C., Coulon, A., Mazet, O., Fayard, J. M., & Beslon, G. (2007). A long-term evolutionary pressure on the amount of noncoding DNA. *Molecular biology and evolution*, 24(10), 2344-2353.

Beagle Team (Lyon)

Carole Knibbe, David Parsons,
Stephan Fischer, Bérénice Batut,
Eric Tannier, Christophe Rigotti,
Jonathan Rouzaud-Cornabas,
Charles Rocabert, Vincent Liard,
Théotime Grohens, Jaap Rutten,
Marco Foley, Paul Banse,
Juliette Luiselli, ...

Dracula Team (Lyon)

Samuel Bernard

LBBE (Lyon)

Gabriel Marais, Vincent Daubin

INSERM (Paris)

Dusan Misevic, Antoine Frénoy

FP7 EvoEvo partners

Dominique Schneider, Paulien Hogeweg,
Santiago Elena, Susan Stepney



www.aevol.fr

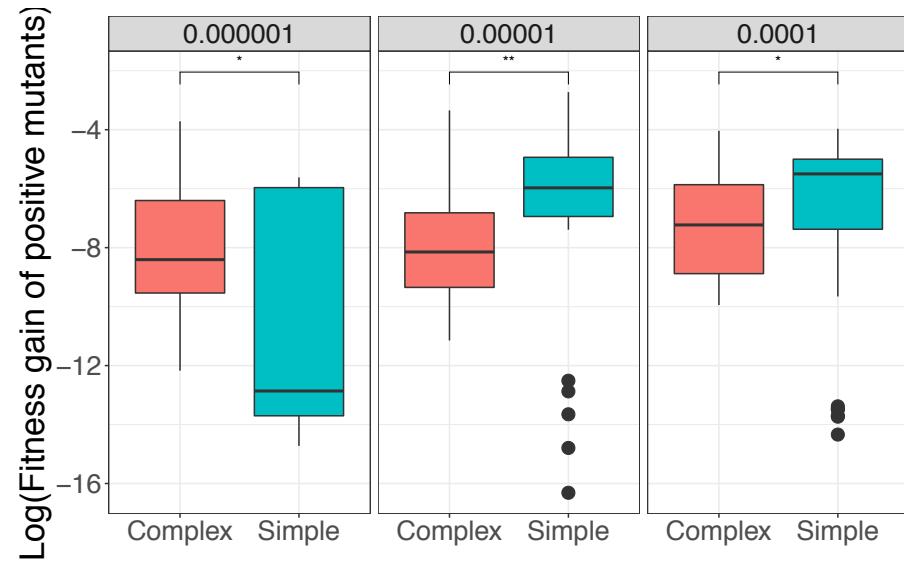
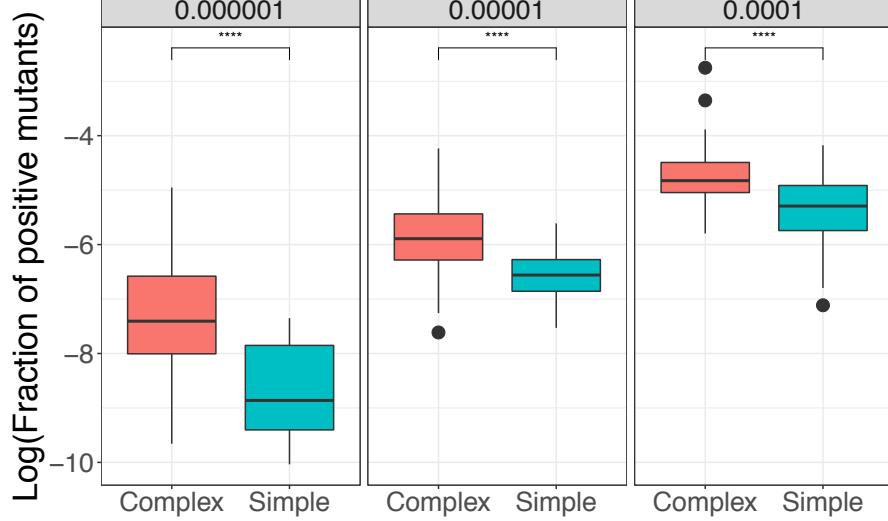
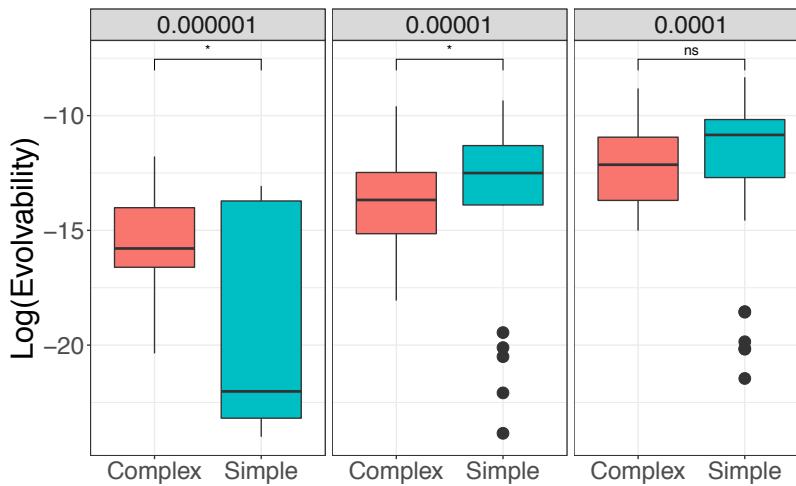
Can robustness constraints invert the complexity ratchet?

- Mutation rates constrain the genome information content
 - (Eigen and Schuster, 1977; Knibbe et al., 2007; Fischer et al., 2014)
 - What if the organisms undergo a strong increase in mutation rate?
- Fraction of C→S transitions under an elevated mutation rate

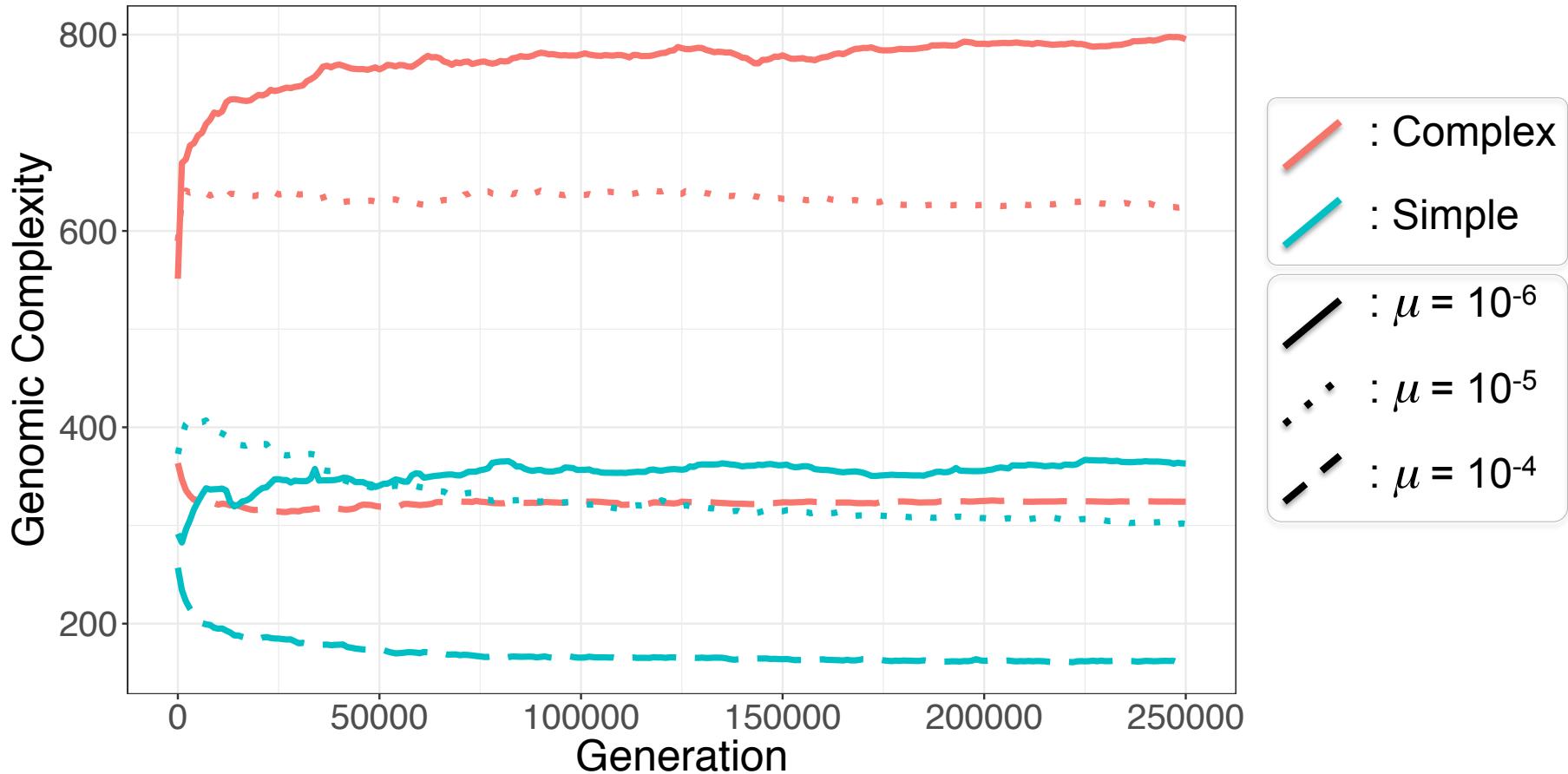
	$\mu = 10^{-4}$	$\mu = 10^{-5}$	$\mu = 10^{-6}$
$\mu_{new} = 10^{-3}$	45.9% [58.3% – 34%] (28/61)	64.4% [74.4% – 52.9%] (47/73)	81.2% [88.1% – 71.6%] (69/85)
$\mu_{new} = 10^{-4}$	/	2.7% [9.3% – 0.7%] (2/74)	10.6% [18.9% – 5.7%] (9/85)
$\mu_{new} = 10^{-5}$	/	/	1.2% [6.4% – 0.2%] (1/85)

→ A harsh mutational pressure can circumvent the complexity ratchet...

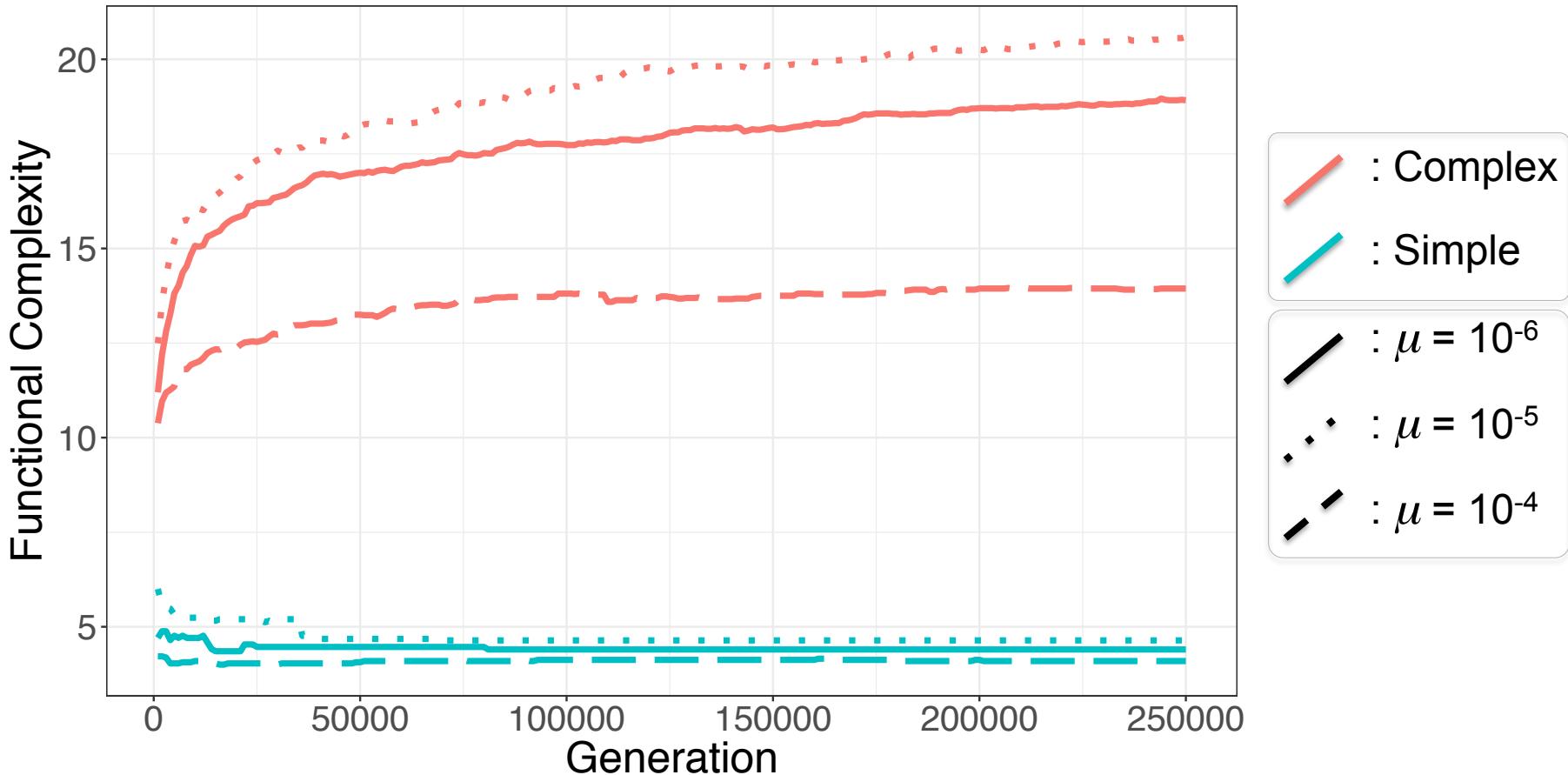
A short discussion on evolvability ...



Complex individuals gain fitness and functional complexity. Only low mutators gain genomic complexity



Complex individuals gain fitness and functional complexity



Complex individuals gain fitness...

