Why EvoSysBio?

Quantify the 5 fundamental factors of evolution by integrating rigorous models from across biology

Intra-Organism Biology

Rigorous models of molecules ... in organisms

molecular systems biology, biochemistry, cell biology, …

Trans-Organism Biology

Rigorous models of organisms ... in ecosystems

evolutionary biology, population genetics, ecology, …

Better predictions of evolution based on mechanistic insights

=

Evolutionary Systems Biology

Evolutionary **Sys**tems **Bio**logy

is

a trans-disciplinary **framework for constructing** reliable, testable, interactive **overviews of** nestable, dynamic, multi-dimensional

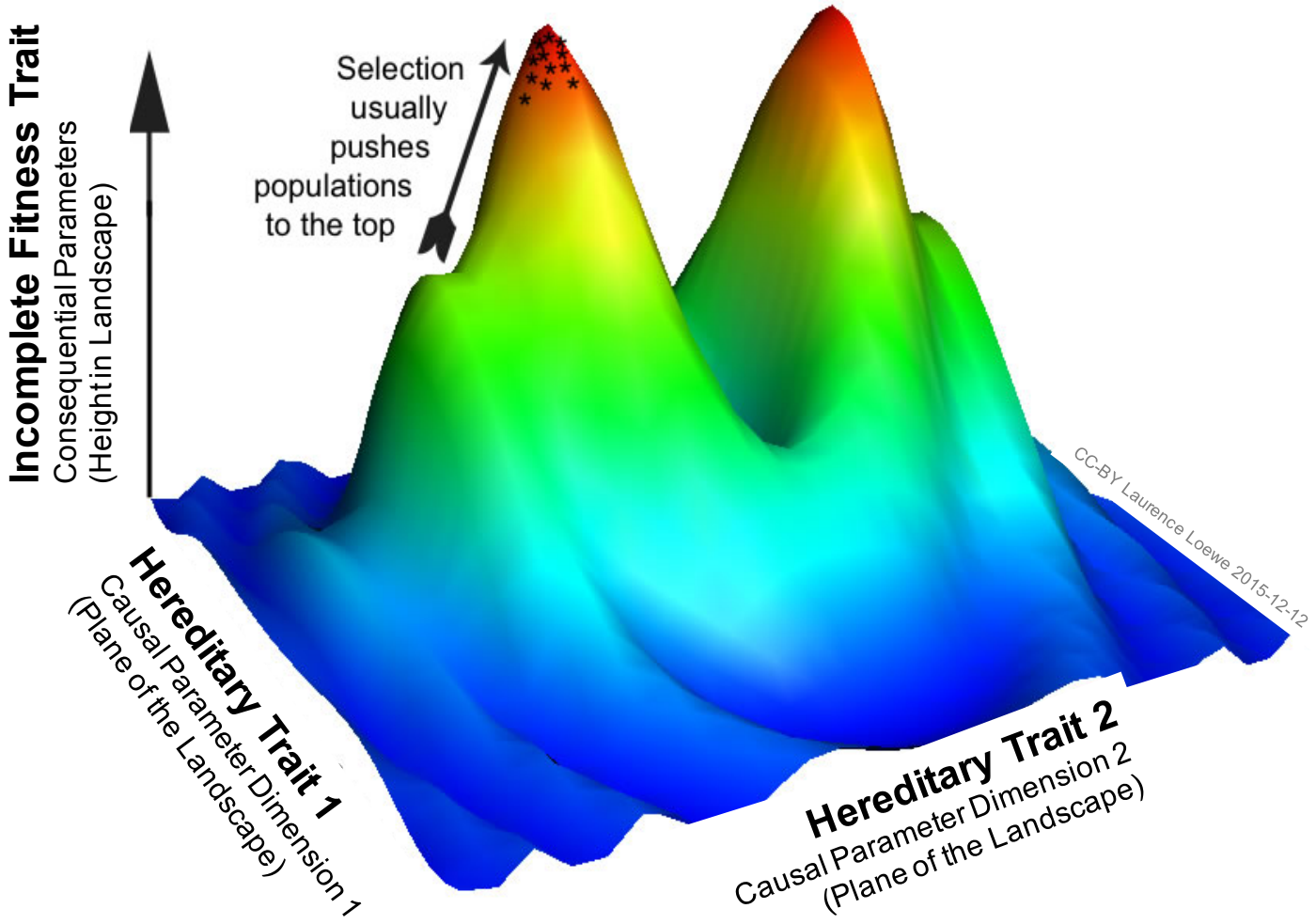
Fitness Landscapes

which mechanistically predict:

① **Changes in fitness of individual organisms** when their states and environments change;

② **How populations evolve** when organisms traverse fitness landscapes.

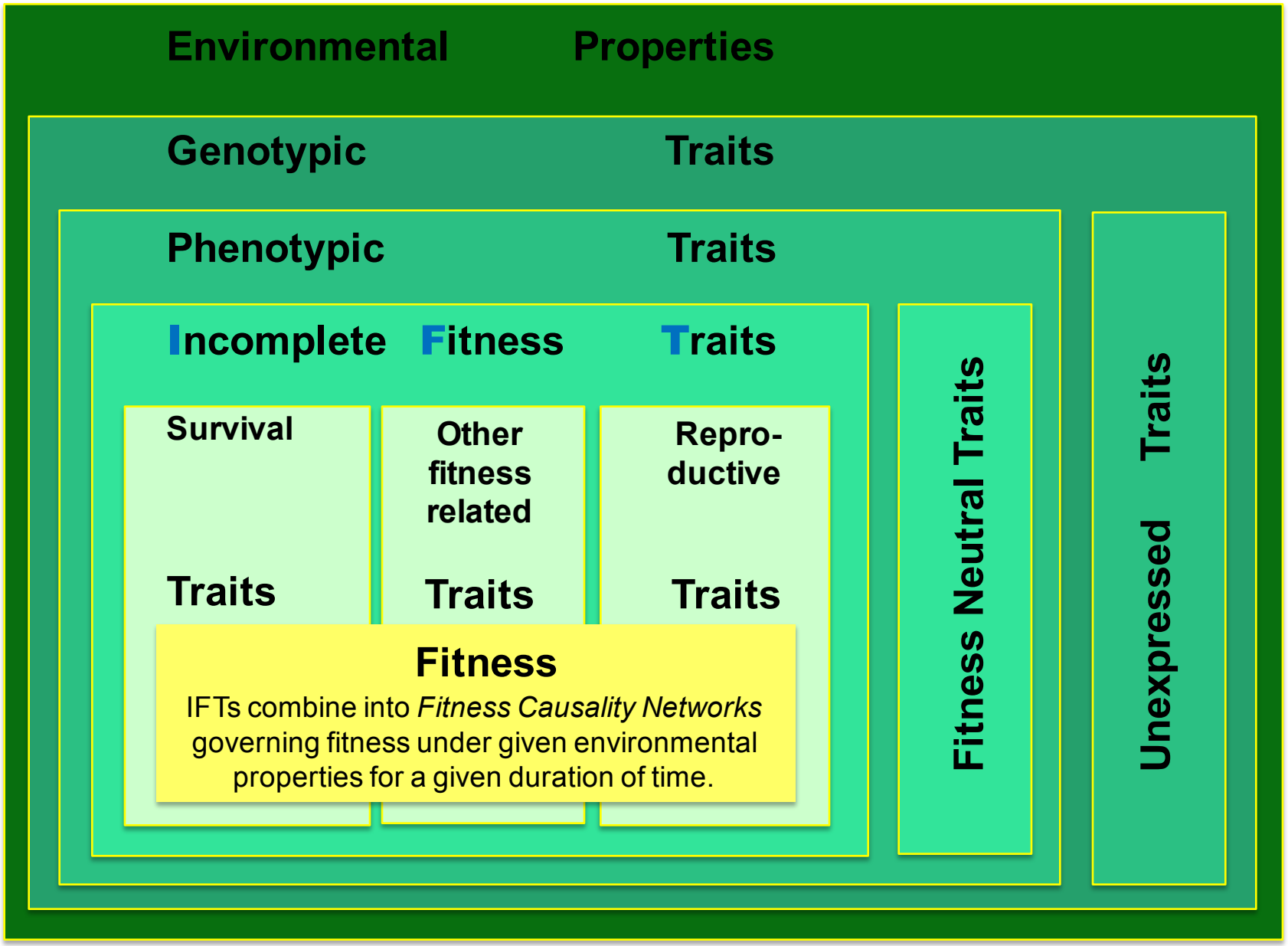
Massively **O**versimplified **C**artoonish **A**bstract **L**andscape of **I**ncomplete **F**itness **T**raits



EvoSysBio's Journey

Mostly MOCA-LIFTs → **Reliable LIFTs**

Environments, Genes, Phenotypes, Chance ➔ Fitness



IFTs of biological data in *Context*

Mean fitness of organisms in groups
(or other population statistics)

Organism survival, reproduction
(RIFTs network of tradeoffs)

Real-world Incomplete Fitness Traits
(RIFTs are reality calibrated SIFTs)

Simulated Incomplete Fitness Traits
(SIFTs are entirely computational)

Time series of phenotypic traits
(nested: ... in cells in tissues in …)

Molecular functions network
(collection of interaction rates)

Molecular structures collection
(structures ➔ govern functions)

Hereditary materials, genetic code,...
(sequences ➔ govern structures)

LIFTs are Methods for Mapping IFT points in *Planes* to *Heights*

7: **Summarizing** stats of relevant groups in specified environments and times

6: **Balancing** trade-offs in IFT networks of organism life history and physiology

5. **Mapping** *in silico* IFTs to relevant observed real-world IFTs

4: **Extracting** fitness-related traits (IFTs) from analyses of time series

3: **Simulating** dynamic time series of parts in systems biology networks

2: **Abstracting** the structural biology of structure-function relationships

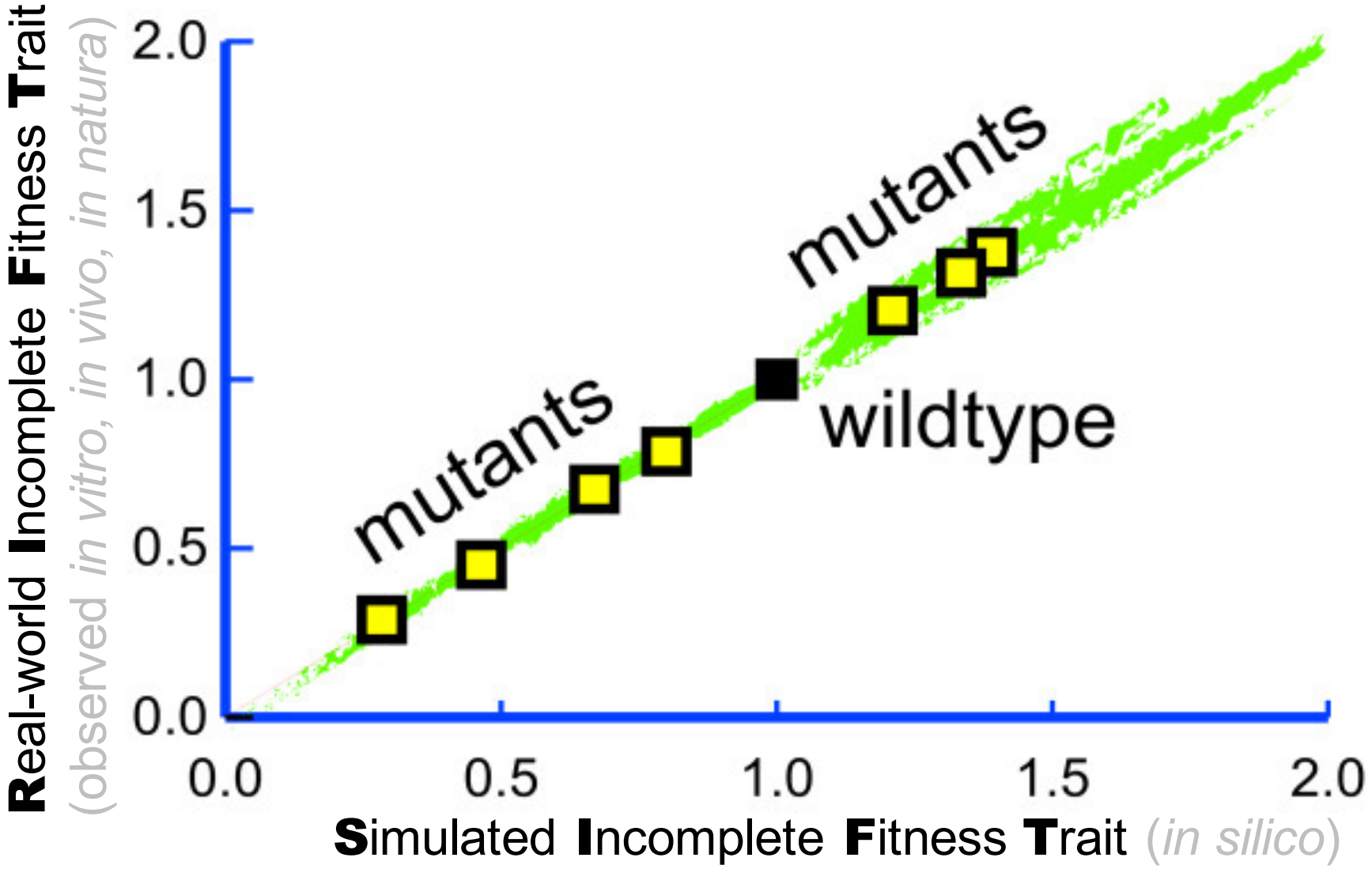
1: **Folding** of expressed sequences into dynamic 3D structures

Mechanistic Maps: Mutations to Phenotype Changes

compute from full Fitness Causality Networks of **L**andscapes of **I**ncomplete **F**itness **T**raits

Simulated **R**eal-world **I**ncomplete **F**itness **T**rait Maps

Simulate SIFTs and measure RIFTs for the same wildtype. If linear extrapolations of SIFTs predict observed RIFTs, the SRIFT Map is usable. If not, statistical approximations may help with mapping.



Wildtypes instantiate LIFTs to DMEs

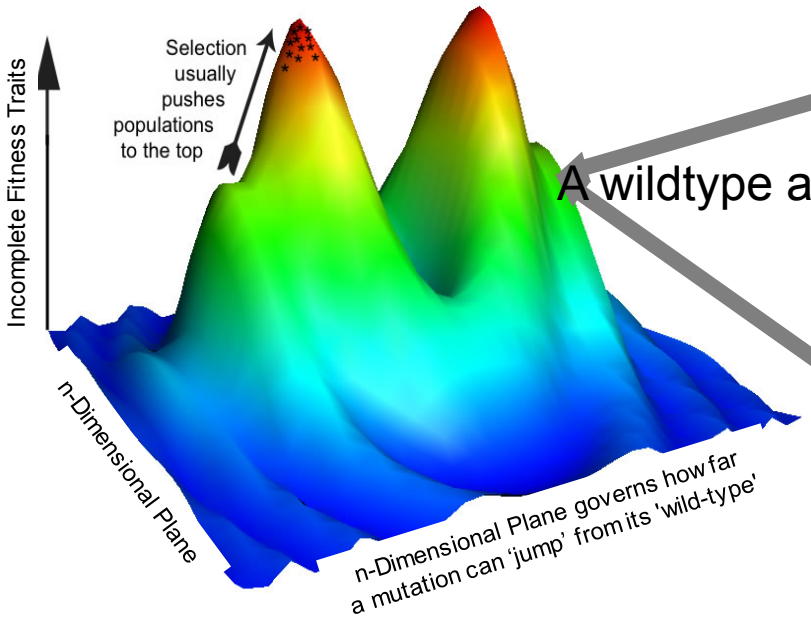
Landscape of
Incomplete
Fitness
Traits

Instantiate

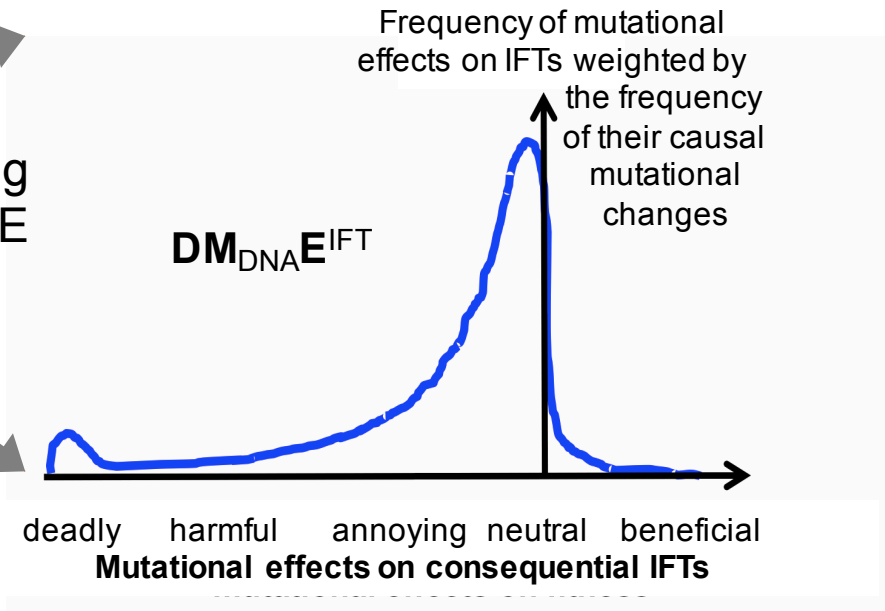
Wildtype

Abstract

Distribution
of increasing or decreasing
Mutational
changes in the causal
input parameter plane with
Effects
on consequential
Incomplete Fitness Traits



A wildtype anchoring a DME



Estimation of Mutational Effects *in Silico*

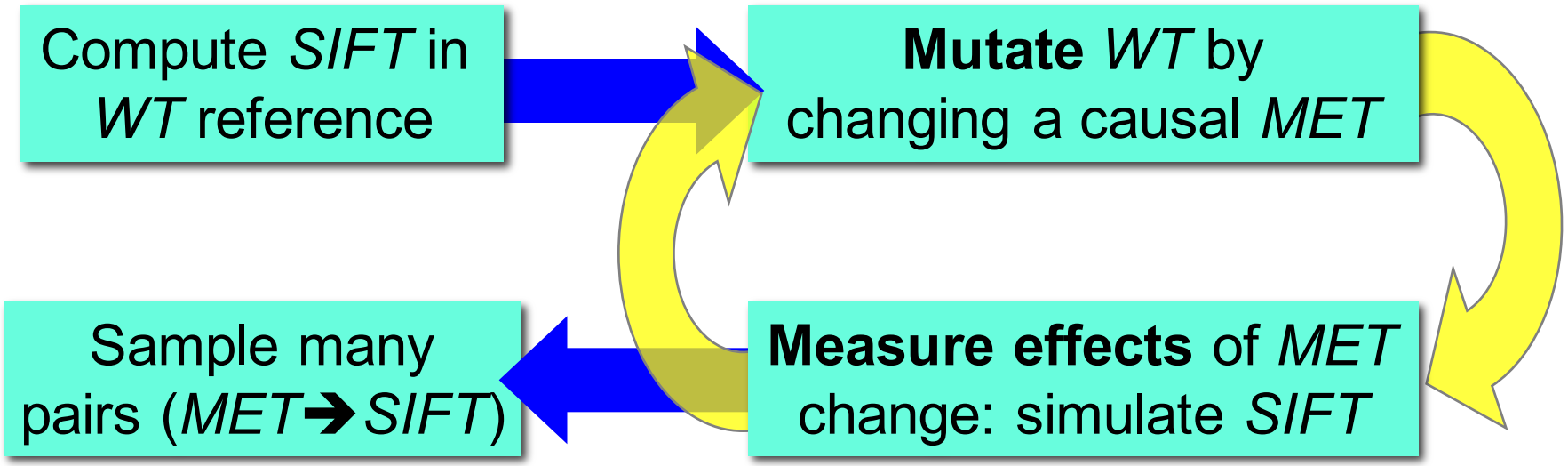
Find: **D**istribution of **M**utational **E**ffects of a given *MET* on a given *SIFT*, short:

$DM_{MET}E^{SIFT}$

1. Define a *WT:* a reasonable systems biology model of a **WildType** system
2. Define a *SIFT:* a consequential **Simulated Incomplete Fitness Trait**

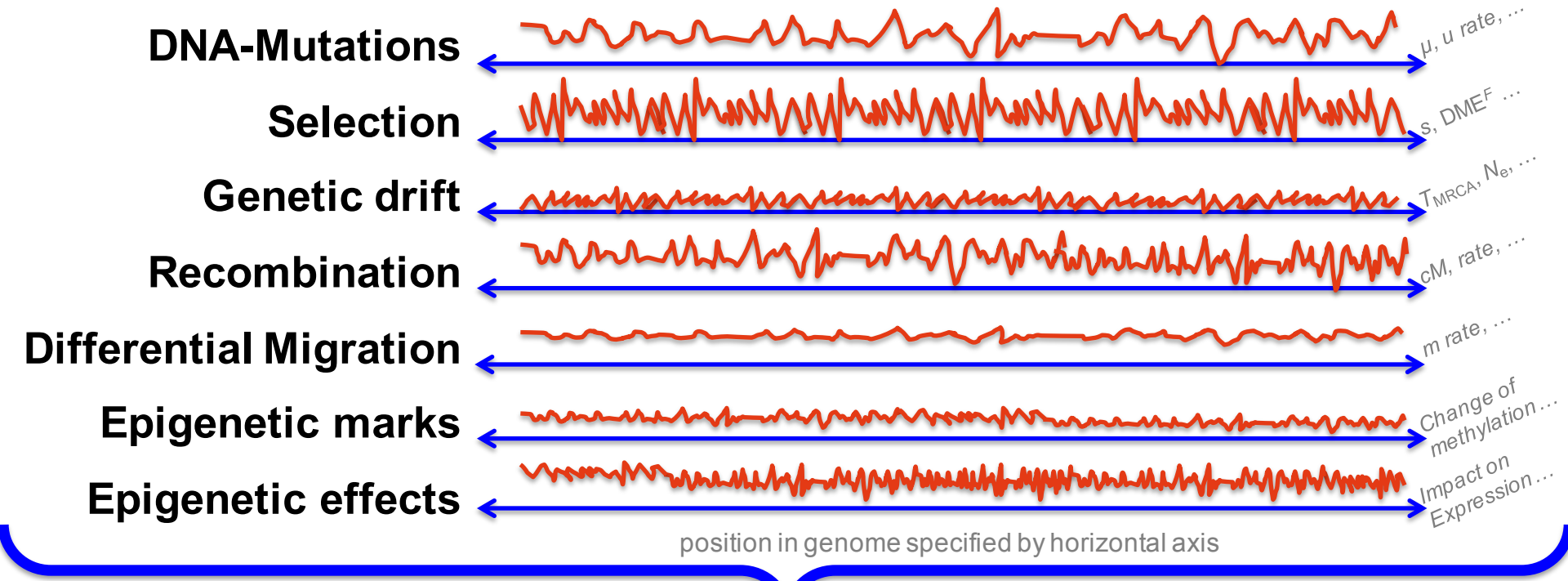
   SIFTs are computable properties of a *WT* with a plausible effect on IFTs that are known to affect fitness in reality at least under some conditions. Use expert intuition to define candidates and expert judgment to test their quality. Examples: reproduction, survival, energy cost, gene regulation accuracy, speed, stability ...
3. Define a *MET:* a credible causal **Mutation Encoding Trait**

   METs are causal traits of a *WT* with some stability ('heritable') that change independently from the SIFTs they impact. Examples: DNA sequence, promoter binding strength, kinetic parameter of enzyme, speed, ...
4. Define a causal **$DM_{MET}E$**, then sample its consequential **$DM_{MET}E^{SIFT}$**



Genomic Landscapes ...

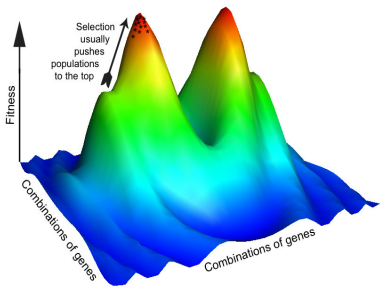
define genome-wide distributions of evolutionary factors that summarize Fitness Causality Networks incorporating effects from biochemistry to ecology



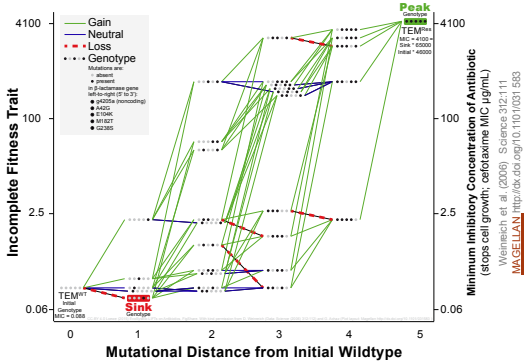
… Pose **Huge Computing Challenges** for Simulating Evolution:

- **Biological Species**: extremely diverse, often nested, different details matter
- **Population sizes**: can be huge (10^{14} bacteria @ 1KB / human gut = 100PB)
- **Genome sizes**: can be huge (*E.coli*: 4 million bp; *H.sapiens*: 6 billion bp)
- **Time heterogeneity**: can be huge (biochemistry: <ns; evolution: >1000My)

Where does **EvoSysBio** need to go from here?



EvoSysBio's Journey



Mostly MOCA-LIFTs

Reliable LIFTs

1. **Embrace Trans-Disciplinarity**
   Develop a shared semantics and language to integrate existing mountains of diffuse physical, bio, and simulation data using the best math and stats methods.
2. **Use Computational Data Science**
   Put computers to work. Reduce inefficient '*silicon digging*' and '*data shoveling*'. Automate!
3. **Improve Models: reproducible, reliable, usable, documented, …**
   Write models for efficiently extending them without the need for reimplementing.
4. **Improve Tools: more accuracy, speed, usability, documentation, …**
   Improve versatility and reliability of simulation tools for more diverse complex problems.
5. **Allow Models to Inspire Biological Wet-Lab Experiments**
   Collect relevant raw time series data; use it to estimate better values for key parameters.

More Details Loewe 2009 *A Framework for Evolutionary Systems Biology* BMC Systems Biology 3:27.
Soyer 2012 *Evolutionary Systems Biology*, Adv Exp Med Biol, vol 751, New York, Springer.