SBMLDock

Docker Driven Systems Biology Tool Development and Usage

E. Z. Gnimpieba, M. Thavappiragasam, A. Chango, B. Conn,
C. M. Lushbough

Introduction

- Simplify researchers workflows
- Docker is a new emerging technology
- Work with our SBML tools

SBMLDock

- Collects 7 CLI tools to manipulate SBML files
- We packaged in a special way
  - Each tool has it’s needed libraries
  - We wrap java execution of all tools
  - We provide test data for each tool

Tool specifics

- ParaABioS
- SBMLChecker
- SBMLCompare
- SBMLMerge

Tool specifics

- SBMLSplit
- SBMLModeler
- SBMLAnnotate

Docker specifics

- A self contained system
- Interacts with the kernel of the host system
- Repeatable and shareable
- Runs on Linux
    - Windows and Mac can work

Docker specifics

- Mount volumes in the container

  -v /tmp/sbmldock:/tmp

- Tell the container where to run

  -w /tmp

- Run our image
- Call one of our programs

Run a tool

- We've included test data
- This is what SBMLAnnotate would produce

```
wjconn@ted:/tmp/sbmldock$ docker run -v /tmp/sbmldock:/tmp -w /tmp usdbioinforma
tics/sbmldock SBMLAnnotate /opt/SBMLAnnotate/one.xml out.xml
start..
wjconn@ted:/tmp/sbmldock$ ls
hsperfdata_root  jsbml.log  out.xml
wjconn@ted:/tmp/sbmldock$
```

Tools in a docker container

- Tools run the same on every system
- Docker images have unique numeric ids
- Containers run the same every time you start them

Conclusion

- Collected SBML specific tools
- Docker provides:
  - Usability
  - Reproducibility
  - Unique identification

- **Acknowledgement:** This work has been partially supported by the National Science Foundation/EPSCoR Award No. IIA-1355423 and by the state of South Dakota, through BioSNTR.

- **References.**

- 1. Beasley, J.M., Coronado, G.D., Livaudais, J., Angeles-Llerenas, A., Ortega-Olvera, C., Romieu, I., Lazcano-Ponce, E., Torres-Mejía, G.: Alcohol and risk of breast cancer in Mexican women. Cancer Causes Control. 21, 863–870 (2010).
- 2. Cock, P.J.A., Grüning, B.A., Paszkiewicz, K., Pritchard, L.: Galaxy tools and workflows for sequence analysis with applications in molecular plant pathology. PeerJ. 1, e167 (2013).
- 3. Hucka, M.: Systems Biology Markup Language (SBML). In: Dubitzky, W., Wolkenhauer, O., Cho, K.-H., and Yokota, H. (eds.) Encyclopedia of Systems Biology SE - 1091. pp. 2057–2063. Springer New York (2013).
- 4. Thavappiragasam, M., Lushbough, C.M., Gnimpieba, E.Z.: Heuristic parallelizable algorithm for similarity based biosystems comparison. In: Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics - BCB '14. pp. 782–789 (2014).
- 5. Thavappiragasam, M., Lushbough, C., Gnimpieba, E.: SBMLChecker, a Semantic approach for SBML model reliability evaluation. worldcomp-proceedings.com. 2–5 (2014).
- 6. Thavappiragasam, M., Lushbough, C.M., Gnimpieba, E.Z.: Automatic biosystems comparison using semantic and name similarity. In: Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics - BCB '14. pp. 790–796 (2014).
- 7. Dräger, A., Rodriguez, N., Dumousseau, M., Dörr, A., Wrzodek, C., Le Novère, N., Zell, A., Hucka, M.: JSBML: a flexible Java library for working with SBML. Bioinformatics. 27, 2167–8 (2011).

Opinion Paper

Evolutionary Constraint-Based Formulation Requires New Bi-level Solving Techniques

Marko Budinich, Jérémie Bourdon, Abdelhalim Larhlimi, Damien Eveillard
ComBi Team, LINA,UMR 6241 CNRS, EMN, Université de Nantes

Computer Methods in Systems Biology, September 2015

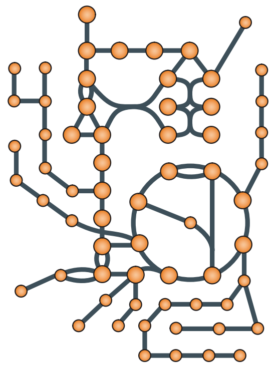
Constraint Based Models

Constraint Based Models

- Constraint Based Models are a type of metabolic models

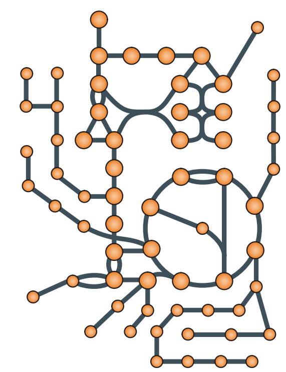
Constraint Based Models

- Constraint Based Models are a type of metabolic models



Constraint Based Models

- Constraint Based Models are a type of metabolic models
- We are interested into rates of reactions **v** (fluxes)

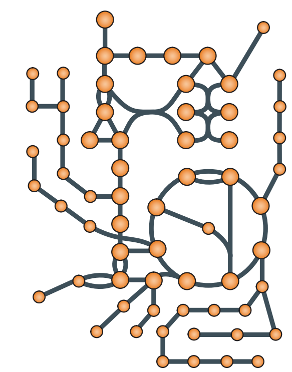


Constraint Based Models

- Constraint Based Models are a type of metabolic models

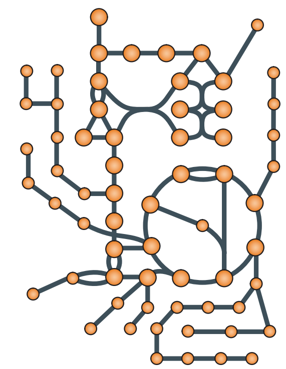
- We are interested into rates of reactions **v** (fluxes)

- Metabolic network is translated to a Stoichiometric Matrix



Constraint Based Models

- Constraint Based Models are a type of metabolic models
- We are interested into rates of reactions **v** (fluxes)
- Metabolic network is translated to a Stoichiometric Matrix



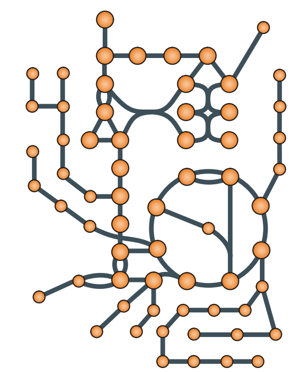
Metabolites

−1	0	0	0	0	1	0	0	0
−1	0	0	0	0	0	1	0	0
0	0	0	−1	−1	0	0	1	0
1	−1	0	1	0	0	0	0	0
0	1	−1	0	0	0	0	0	0
0	0	−1	0	1	0	0	0	0
0	0	1	0	0	0	0	0	−1

Reactions

Constraint Based Models

- Constraint Based Models are a type of metabolic models
- We are interested into rates of reactions **v** (fluxes)
- Metabolic network is translated to a Stoichiometric Matrix
- Steady State Mass Balance and bound constraints defines a solution space



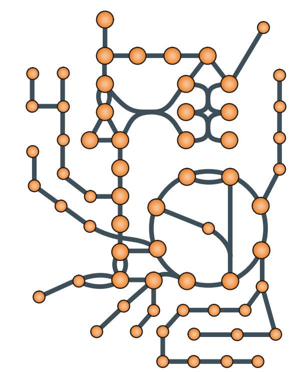
Metabolites

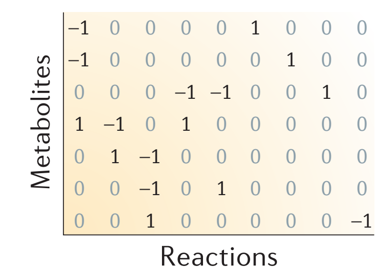
−1	0	0	0	0	1	0	0	0
−1	0	0	0	0	0	1	0	0
0	0	0	−1	−1	0	0	1	0
1	−1	0	1	0	0	0	0	0
0	1	−1	0	0	0	0	0	0
0	0	−1	0	1	0	0	0	0
0	0	1	0	0	0	0	0	−1

Reactions

Constraint Based Models

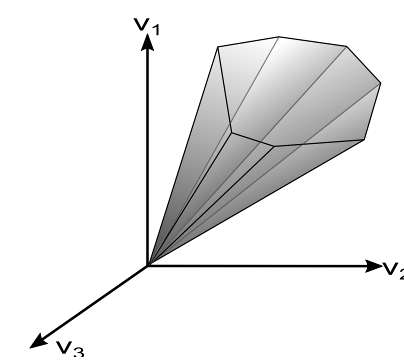
- Constraint Based Models are a type of metabolic models
- We are interested into rates of reactions **v** (fluxes)
- Metabolic network is translated to a Stoichiometric Matrix
- Steady State Mass Balance and bound constraints defines a solution space





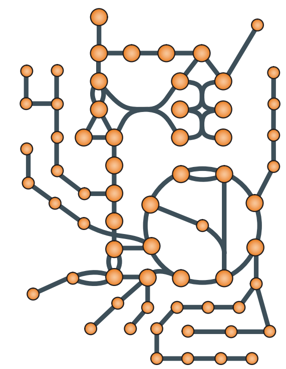
$$\frac{d[C]}{dt} = Sv = 0$$

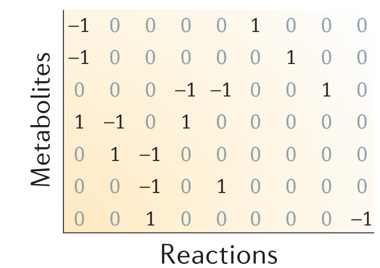
$$lb \leq v \leq ub$$



Constraint Based Models

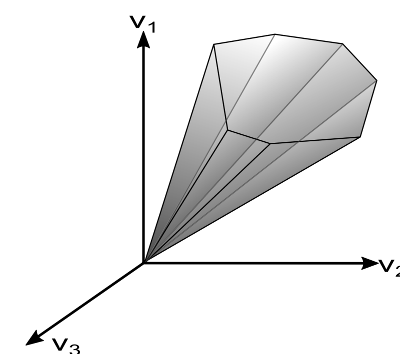
- Constraint Based Models are a type of metabolic models
- We are interested into rates of reactions **v** (fluxes)
- Metabolic network is translated to a Stoichiometric Matrix
- Steady State Mass Balance and bound constraints defines a solution space
- Usually formalized as optimization problem





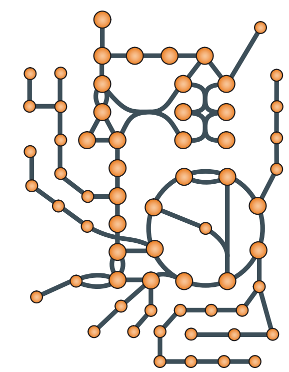
$$\frac{d[C]}{dt} = Sv = 0$$

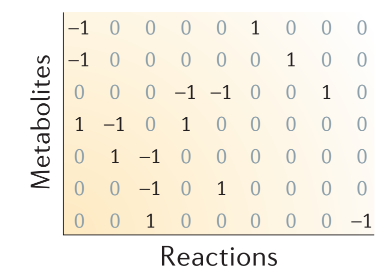
$$lb \leq v \leq ub$$



Constraint Based Models

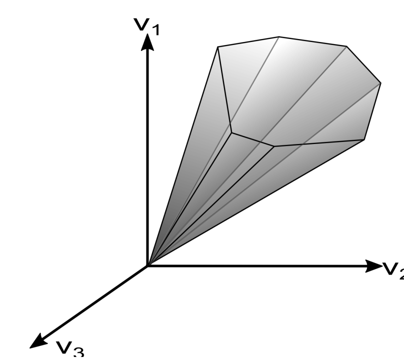
- Constraint Based Models are a type of metabolic models
- We are interested into rates of reactions **v** (fluxes)
- Metabolic network is translated to a Stoichiometric Matrix
- Steady State Mass Balance and bound constraints defines a solution space
- Usually formalized as optimization problem





$$\frac{d[C]}{dt} = \mathbf{Sv} = 0$$

$$\mathbf{lb} \leq \mathbf{v} \leq \mathbf{ub}$$



$$\max \; c^T v$$

s.t

$$\mathbf{Sv}=0$$

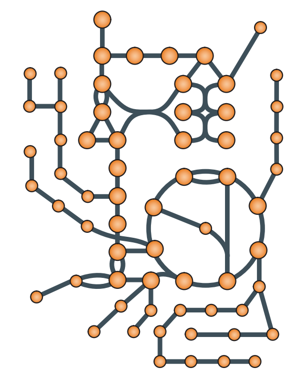
$$\mathbf{lb} \leq \mathbf{v} \leq \mathbf{ub}$$

$$v_i = \mathbb{E}, \; i \in \mathcal{L}$$

...

Constraint Based Models

- Constraint Based Models are a type of metabolic models
- We are interested into rates of reactions **v** (fluxes)
- Metabolic network is translated to a Stoichiometric Matrix
- Steady State Mass Balance and bound constraints defines a solution space
- Usually formalized as optimization problem

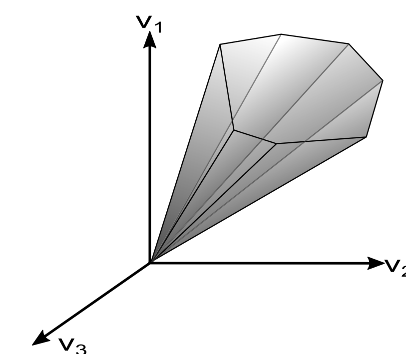


Metabolites × Reactions

−1	0	0	0	0	1	0	0	0
−1	0	0	0	0	0	1	0	0
0	0	0	−1	−1	0	0	1	0
1	−1	0	1	0	0	0	0	0
0	1	−1	0	0	0	0	0	0
0	0	−1	0	1	0	0	0	0
0	0	1	0	0	0	0	0	−1

$$\frac{d[C]}{dt} = \mathbf{Sv} = 0$$

$$\mathbf{lb} \leq \mathbf{v} \leq \mathbf{ub}$$



$$\max \; c^T v$$

s.t

$$\mathbf{Sv}=0$$

$$\mathbf{lb} \leq \mathbf{v} \leq \mathbf{ub}$$

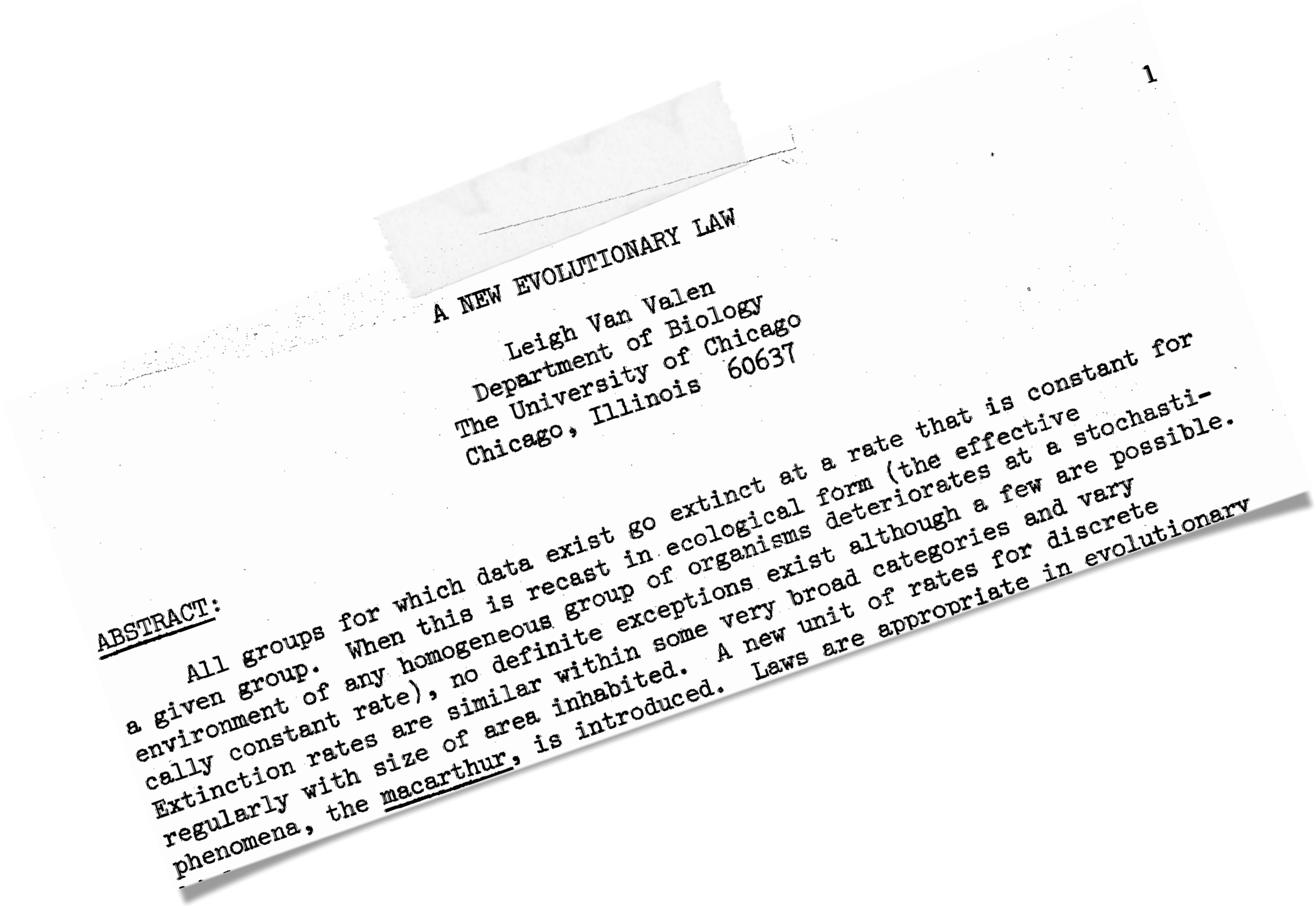
$$\mathbf{v_i} = \mathbf{E}, \; \mathbf{i} \in \mathcal{L}$$

...

Lewis, N. E., et al. (2012). Constraining the metabolic genotype– phenotype relationship using a phylogeny of in silico methods. Nature Reviews: Microbiology, 10(4), 291–305
Bordbar, A., Monk, J. M., King, Z. A., & Palsson, B. O. (2014). Constraint-based models predict metabolic and associated cellular functions. Nature Reviews Genetics, 15(2), 107–120.

Evolution

Evolution



1

A NEW EVOLUTIONARY LAW

Leigh Van Valen
Department of Biology
The University of Chicago
Chicago, Illinois 60637

ABSTRACT:

All groups for which data exist go extinct at a rate that is constant for a given group. When this is recast in ecological form (the effective environment of any homogeneous group of organisms deteriorates at a stochastically constant rate), no definite exceptions exist although a few are possible. Extinction rates are similar within some very broad categories and vary regularly with size of area inhabited. A new unit of rates for discrete phenomena, the macarthur, is introduced. Laws are appropriate in evolutionary

Evolution

1

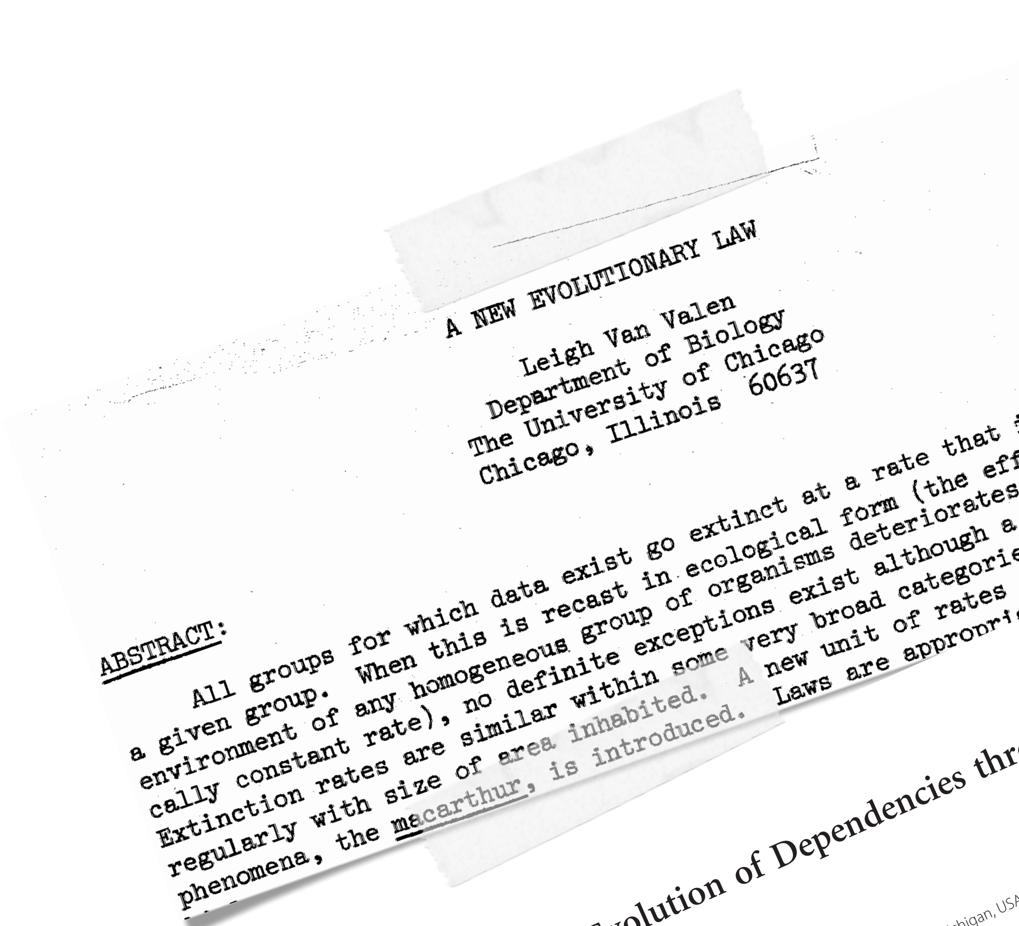
A NEW EVOLUTIONARY LAW

Leigh Van Valen
Department of Biology
The University of Chicago
Chicago, Illinois 60637

ABSTRACT:

All groups for which data exist go extinct at a rate that i
a given group. When this is recast in ecological form (the effe
environment of any homogeneous group of organisms deteriorates
cally constant rate), no definite exceptions exist although a few
Extinction rates are similar within some very broad categories and vary
regularly with size of area inhabited. A new unit of rates for discrete
phenomena, the macarthur, is introduced. Laws are appropriate in evolutionary

Evolution



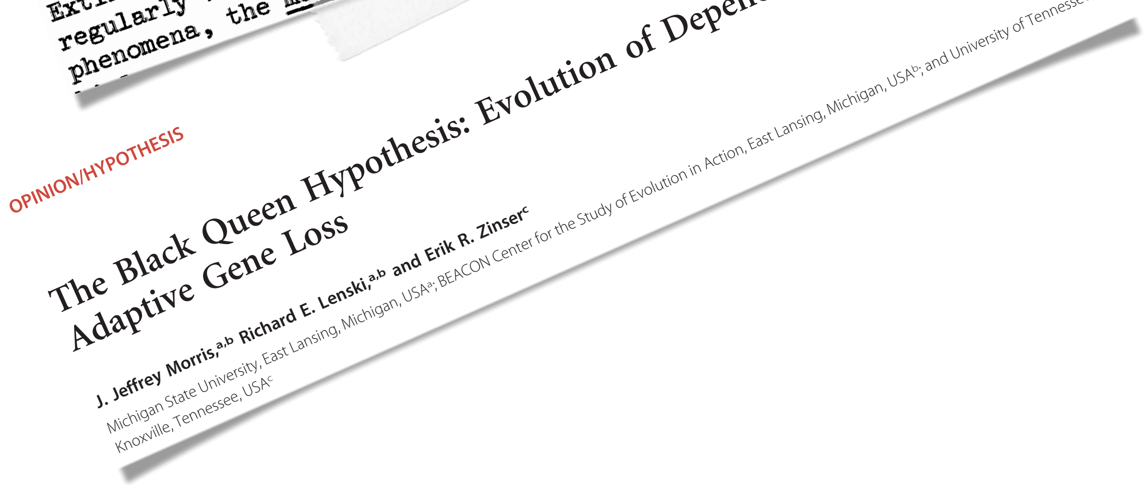
1

A NEW EVOLUTIONARY LAW

Leigh Van Valen
Department of Biology
The University of Chicago
Chicago, Illinois 60637

ABSTRACT:

All groups for which data exist go extinct at a rate that i... a given group. When this is recast in ecological form (the eff... environment of any homogeneous group of organisms deteriorates ... cally constant rate), no definite exceptions exist although a few ... Extinction rates are similar within some very broad categories and vary regularly with size of area inhabited. A new unit of rates for discrete phenomena, the macarthur, is introduced. Laws are appropriate ... evolutionary ...



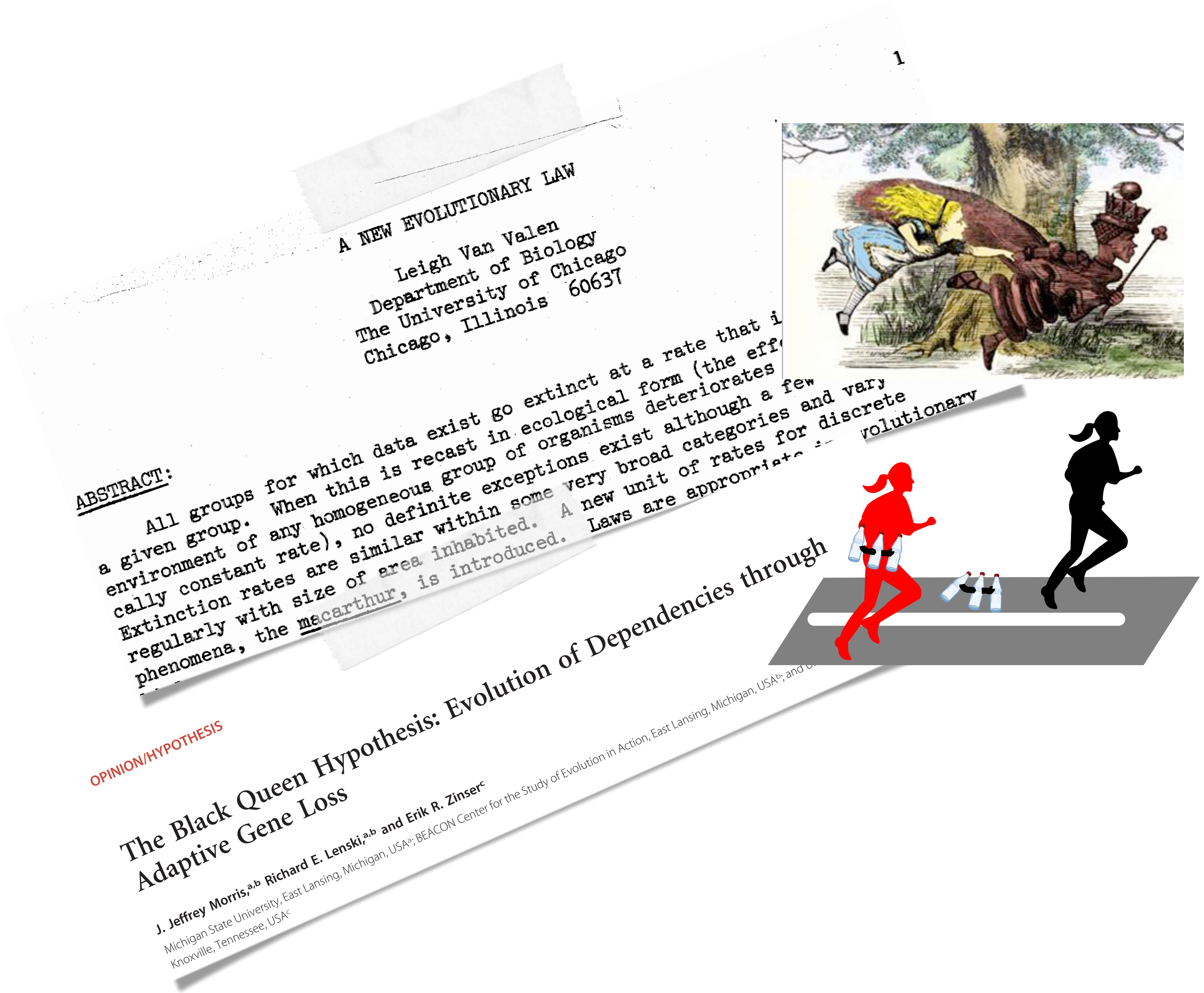
OPINION/HYPOTHESIS

The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss

J. Jeffrey Morris,[a,b] Richard E. Lenski,[a,b] and Erik R. Zinser[c]

Michigan State University, East Lansing, Michigan, USA[a]; BEACON Center for the Study of Evolution in Action, East Lansing, Michigan, USA[b]; and University of Tennessee, Knoxville, Tennessee, USA[c]

Evolution



1

A NEW EVOLUTIONARY LAW

Leigh Van Valen
Department of Biology
The University of Chicago
Chicago, Illinois 60637

ABSTRACT:

All groups for which data exist go extinct at a rate that i
a given group. When this is recast in ecological form (the eff
environment of any homogeneous group of organisms deteriorates
cally constant rate), no definite exceptions exist although a few
Extinction rates are similar within some very broad categories and vary
regularly with size of area inhabited. A new unit of rates for discrete
phenomena, the macarthur, is introduced. Laws are appropriate i evolutionary

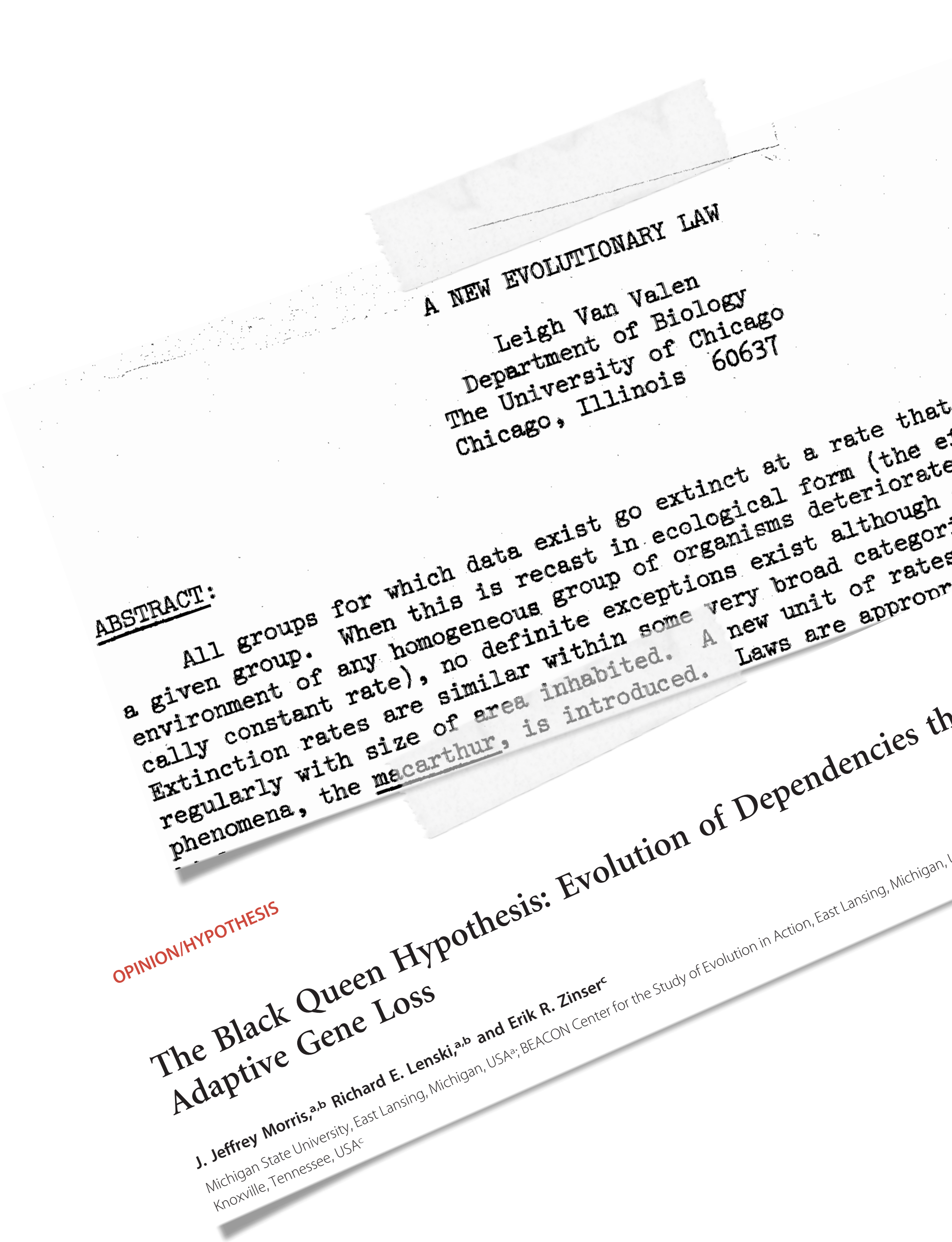
OPINION/HYPOTHESIS

The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss

J. Jeffrey Morris,[a,b] Richard E. Lenski,[a,b] and Erik R. Zinser[c]

Michigan State University, East Lansing, Michigan, USA[a]; BEACON Center for the Study of Evolution in Action, East Lansing, Michigan, USA[b]; and U
Knoxville, Tennessee, USA[c]

Evolution



1

A NEW EVOLUTIONARY LAW

Leigh Van Valen
Department of Biology
The University of Chicago
Chicago, Illinois 60637

ABSTRACT:

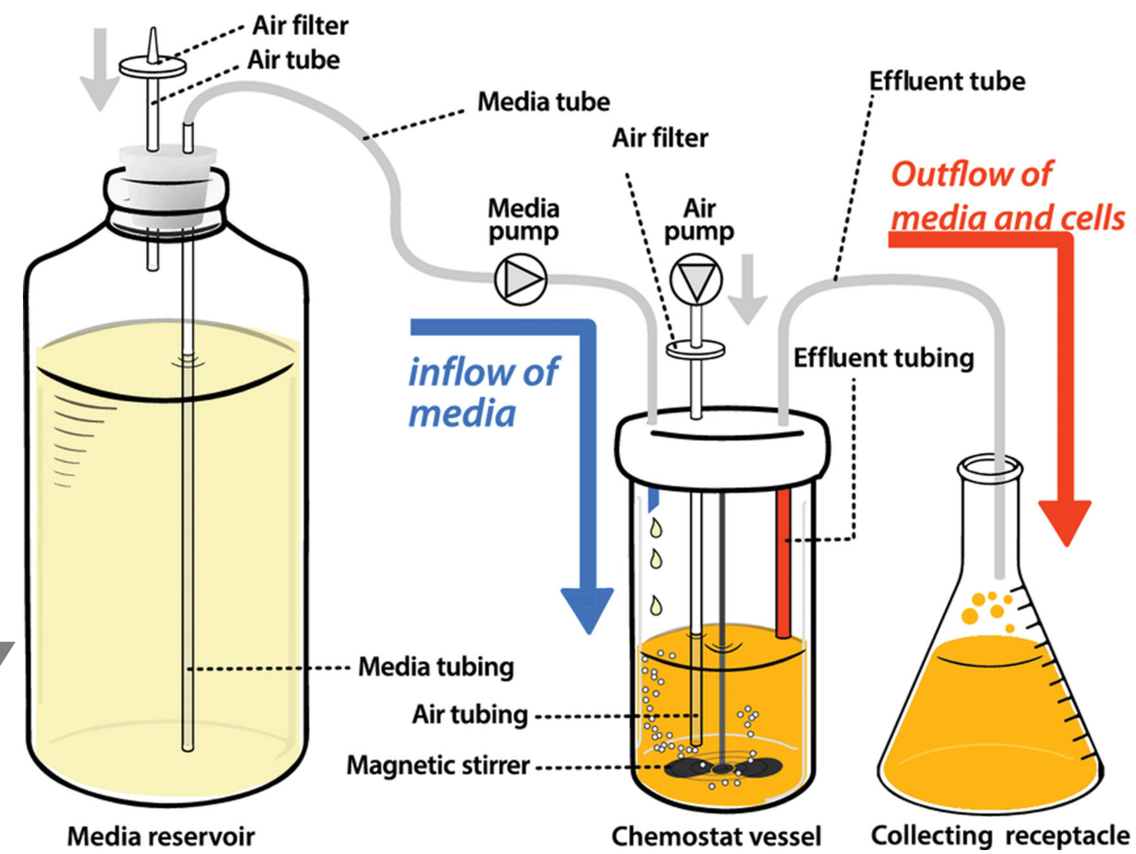
All groups for which data exist go extinct at a rate that i... a given group. When this is recast in ecological form (the eff... environment of any homogeneous group of organisms deteriorates ...cally constant rate), no definite exceptions exist although a few ... Extinction rates are similar within some very broad categories and vary regularly with size of area inhabited. A new unit of rates for discrete phenomena, the macarthur, is introduced. Laws are appropriate ... evolutionary

OPINION/HYPOTHESIS

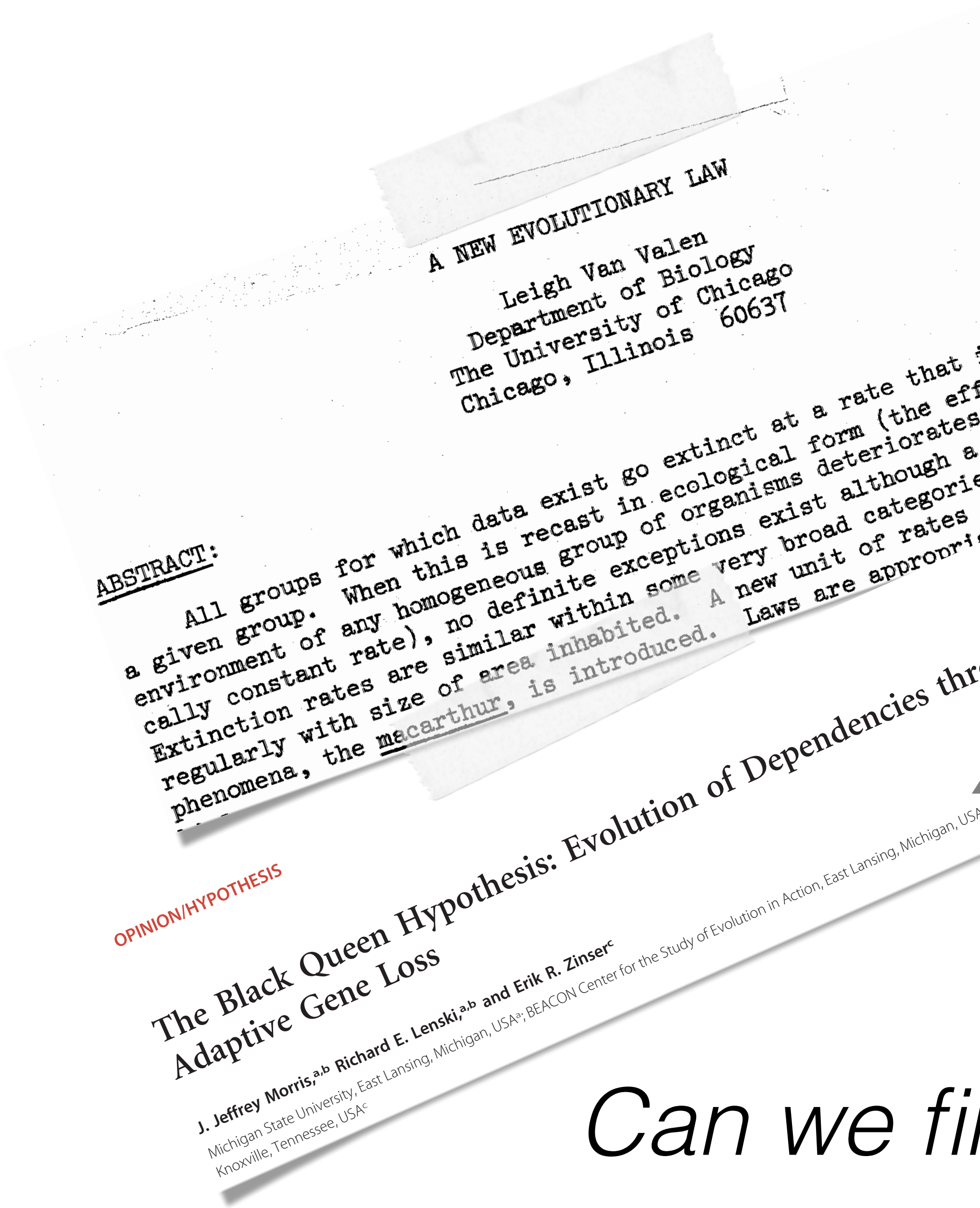
The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss

J. Jeffrey Morris,[a,b] Richard E. Lenski,[a,b] and Erik R. Zinser[c]

Michigan State University, East Lansing, Michigan, USA[a]; BEACON Center for the Study of Evolution in Action, East Lansing, Michigan, USA[b]; and U... Knoxville, Tennessee, USA[c]



Evolution



1

A NEW EVOLUTIONARY LAW

Leigh Van Valen
Department of Biology
The University of Chicago
Chicago, Illinois 60637

ABSTRACT:

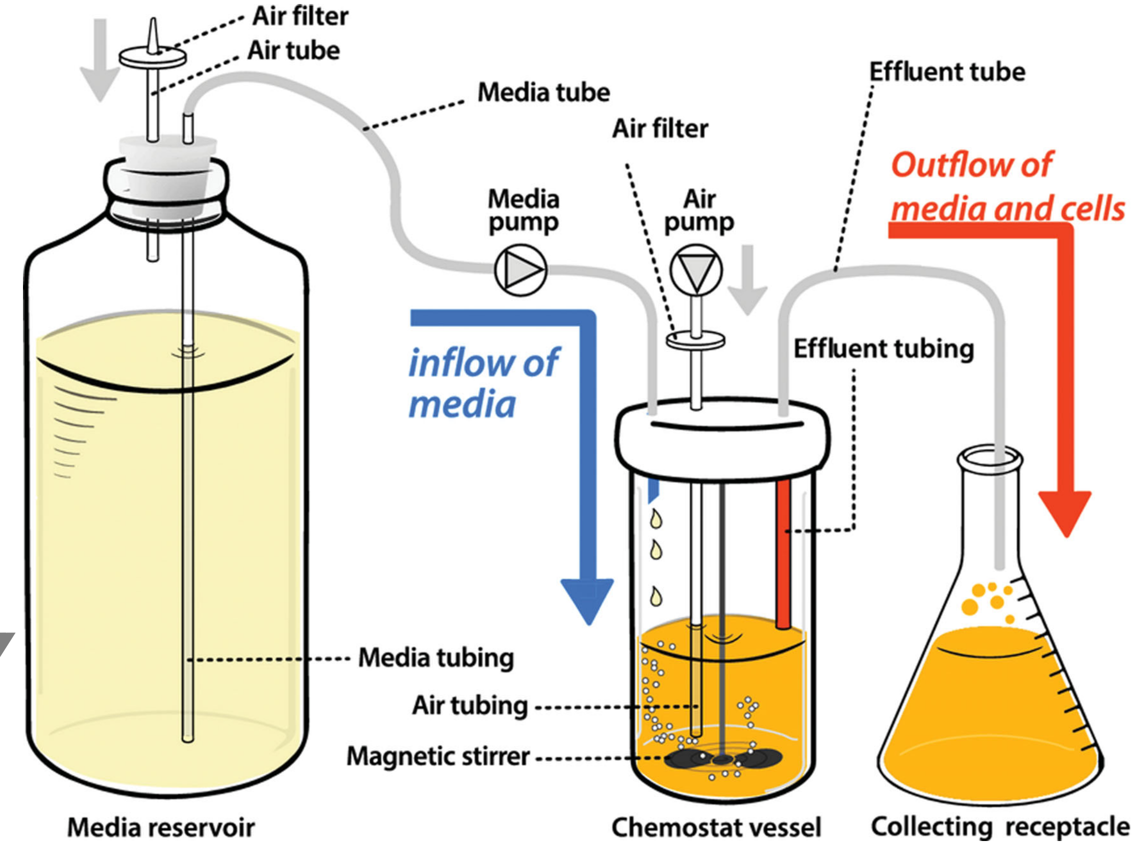
All groups for which data exist go extinct at a rate that i… a given group. When this is recast in ecological form (the eff… environment of any homogeneous group of organisms deteriorates… cally constant rate), no definite exceptions exist although a few… Extinction rates are similar within some very broad categories and vary… regularly with size of area inhabited. A new unit of rates for discrete… phenomena, the macarthur, is introduced. Laws are appropriate … volutionary

OPINION/HYPOTHESIS

The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss

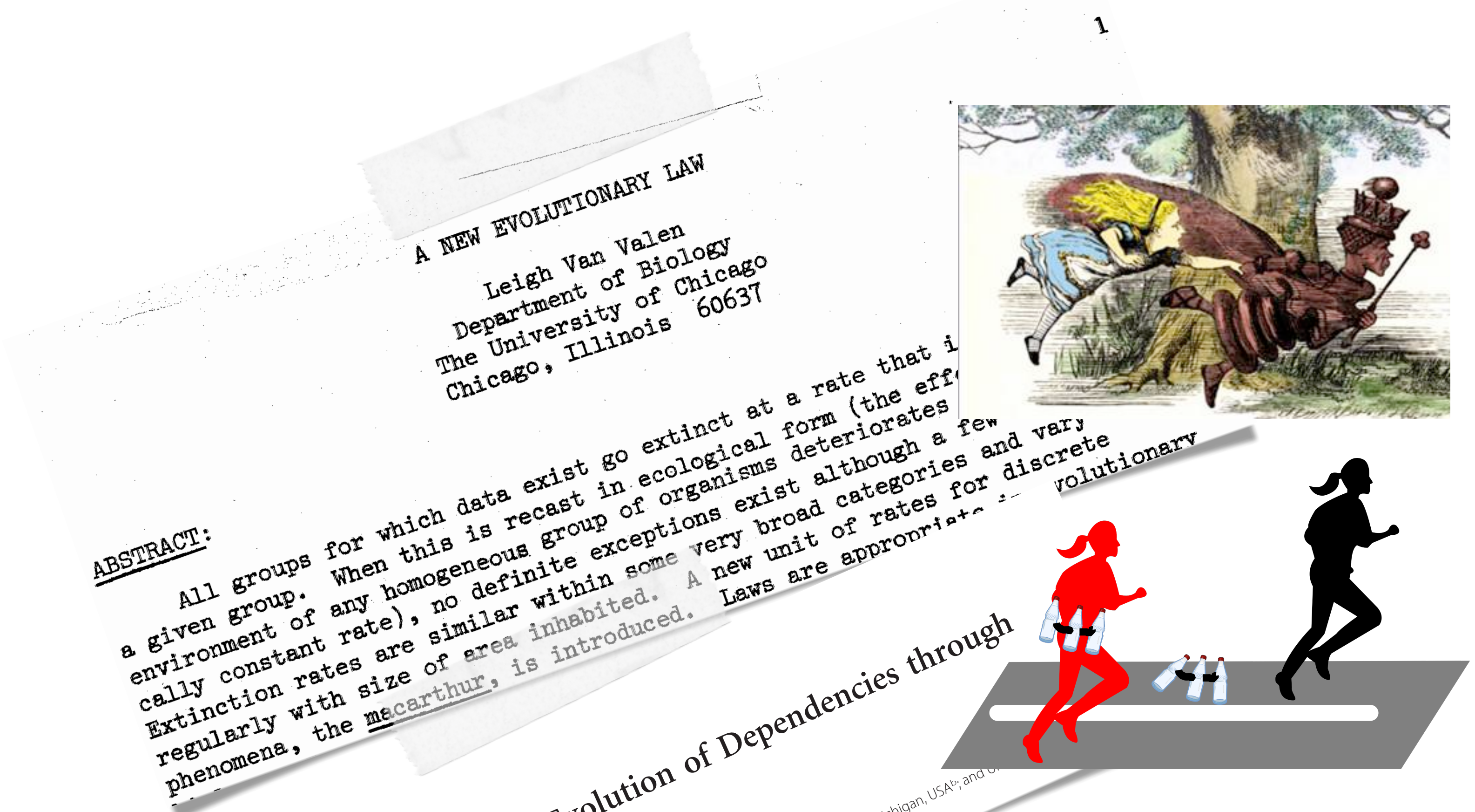
J. Jeffrey Morris,[a,b] Richard E. Lenski,[a,b] and Erik R. Zinser[c]

Michigan State University, East Lansing, Michigan, USA[a]; BEACON Center for the Study of Evolution in Action, East Lansing, Michigan, USA[b]; and … Knoxville, Tennessee, USA[c]



Can we find candidate genes for these theories?

Evolution



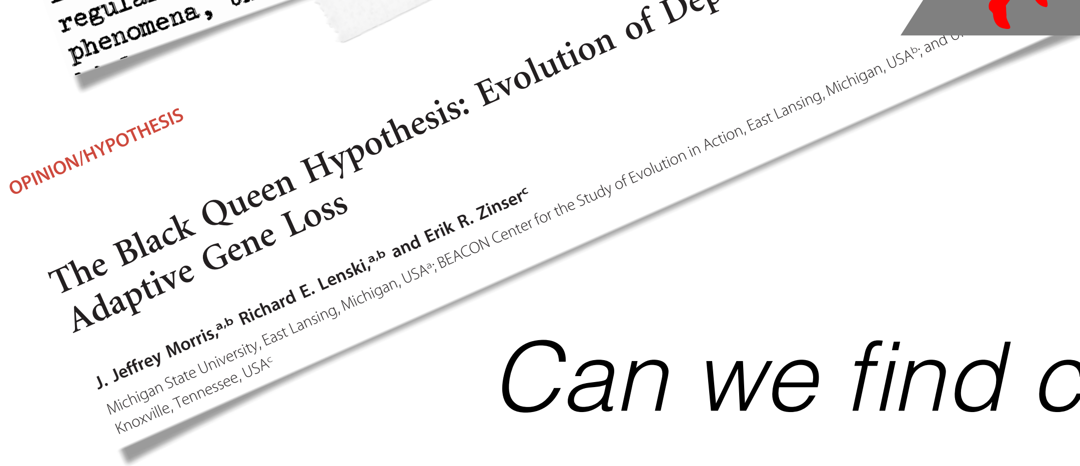
1

A NEW EVOLUTIONARY LAW

Leigh Van Valen
Department of Biology
The University of Chicago
Chicago, Illinois 60637

ABSTRACT:

All groups for which data exist go extinct at a rate that i
a given group. When this is recast in ecological form (the eff
environment of any homogeneous group of organisms deteriorates
cally constant rate), no definite exceptions exist although a few
Extinction rates are similar within some very broad categories and vary
regularly with size of area inhabited. A new unit of rates for discrete
phenomena, the macarthur, is introduced. Laws are appropriate ... volutionary

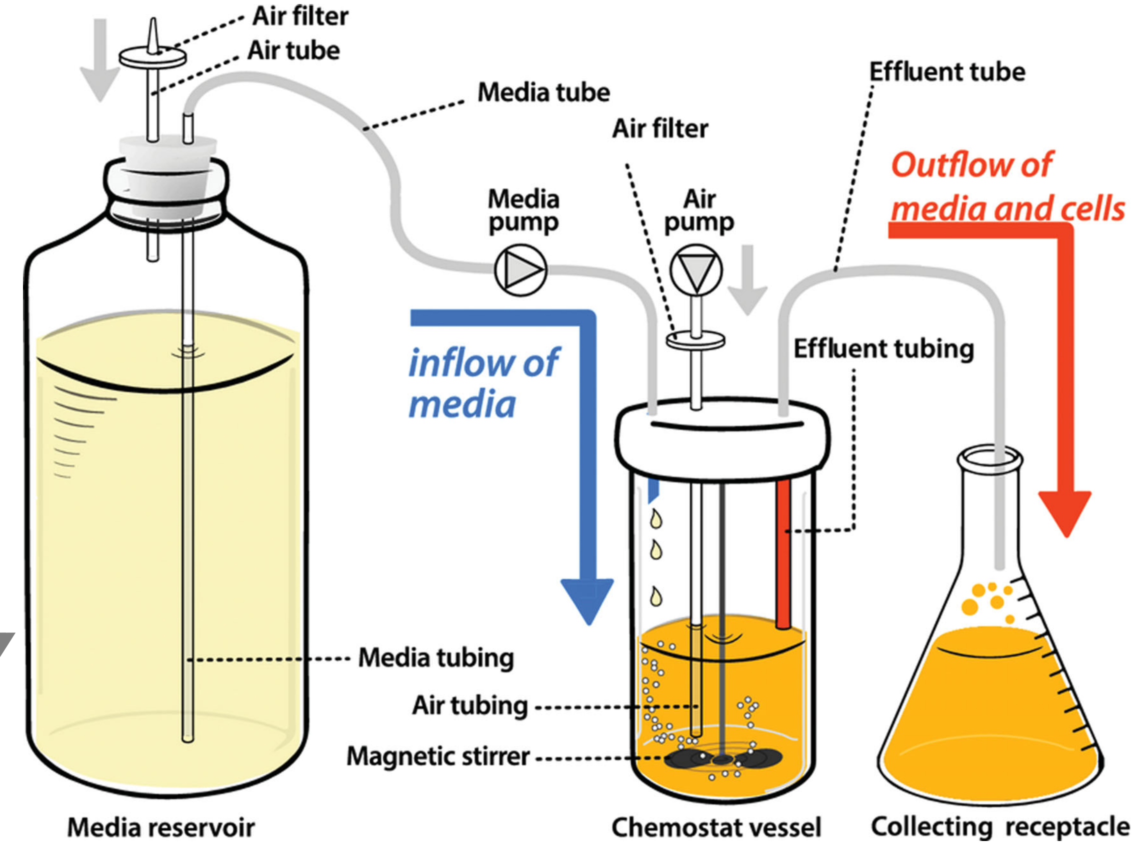


OPINION/HYPOTHESIS

The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss

J. Jeffrey Morris,[a,b] Richard E. Lenski,[a,b] and Erik R. Zinser[c]

Michigan State University, East Lansing, Michigan, USA[a]; BEACON Center for the Study of Evolution in Action, East Lansing, Michigan, USA[b]; and U...
Knoxville, Tennessee, USA[c]



Can we find candidate genes for these theories?

Van Valen, L. (1973). A new evolutionary law. Evolutionary Theory, 1, 1–30.
Van Valen, L. (1974). Molecular evolution as predicted by natural selection. Journal of Molecular Evolution, 3(2), 89–101.
Morris, J. J., Lenski, R. E., & Zinser, E. R. (2012). The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. mBio, 3(2)
Gresham, D., & Hong, J. (2014). The functional basis of adaptive evolution in chemostats. Fems Microbiology Reviews

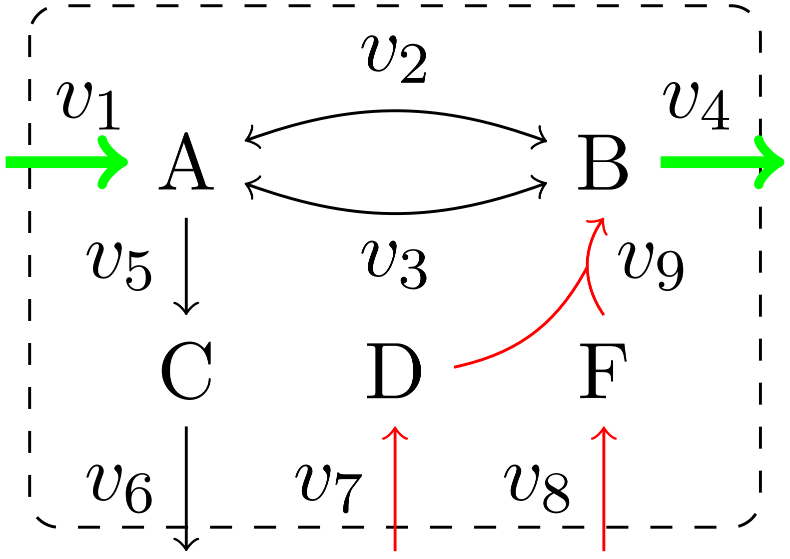
Gene Loss using CBM

Gene Loss using CBM

- Blocked reactions are reactions that **can not** carry flux under given conditions

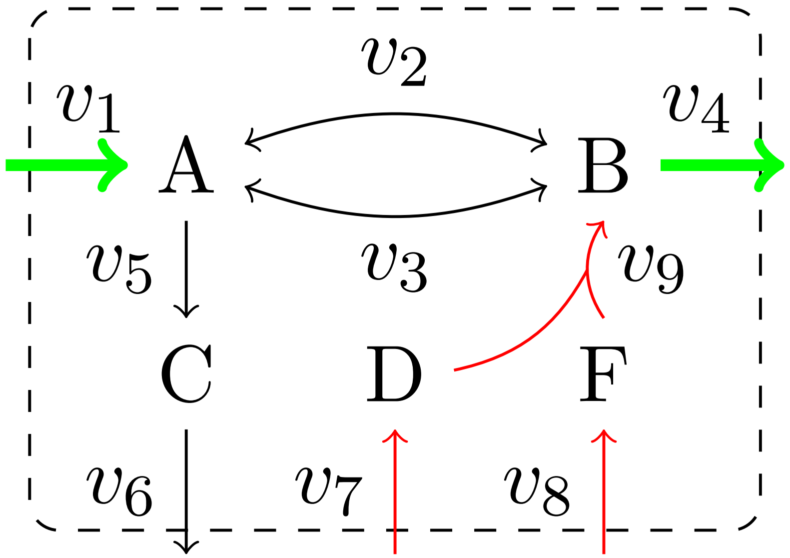
Gene Loss using CBM

- Blocked reactions are reactions that **can not** carry flux under given conditions



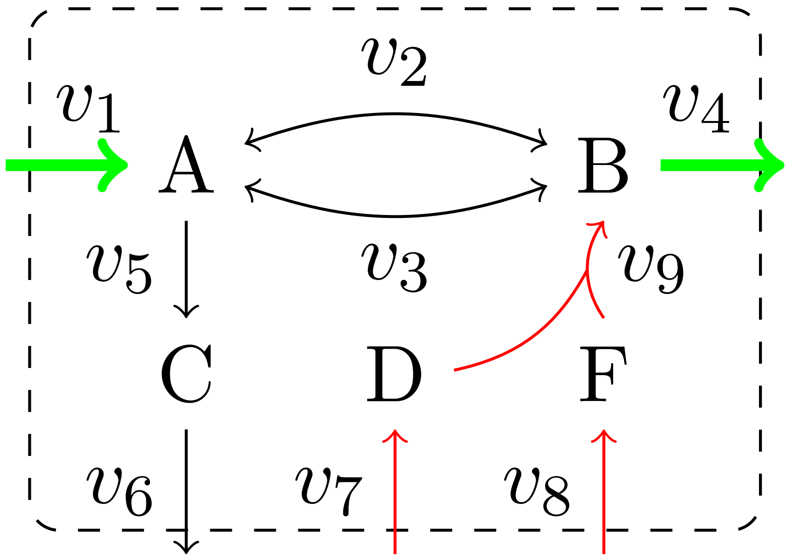
Gene Loss using CBM

- Blocked reactions are reactions that **can not** carry flux under given conditions
- We expect that these reactions will be loosed through evolution



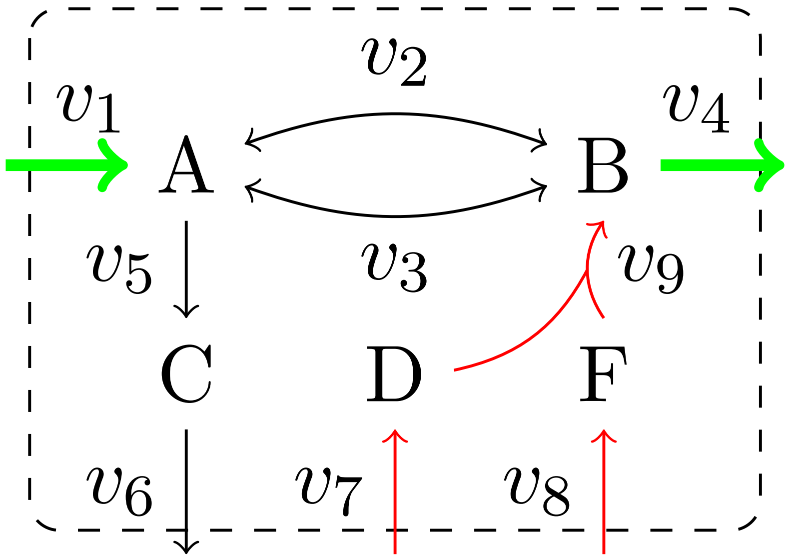
Gene Loss using CBM

- Blocked reactions are reactions that **can not** carry flux under given conditions
- We expect that these reactions will be loosed through evolution
- Given a set of conditions $\mathbf{E}$, we can find blocked reactions through simulation



Gene Loss using CBM

- Blocked reactions are reactions that **can not** carry flux under given conditions
- We expect that these reactions will be loosed through evolution
- Given a set of conditions $\mathbf{E}$, we can find blocked reactions through simulation



$$\max\ c^T v$$

s.t

$$\mathbf{Sv=0}$$

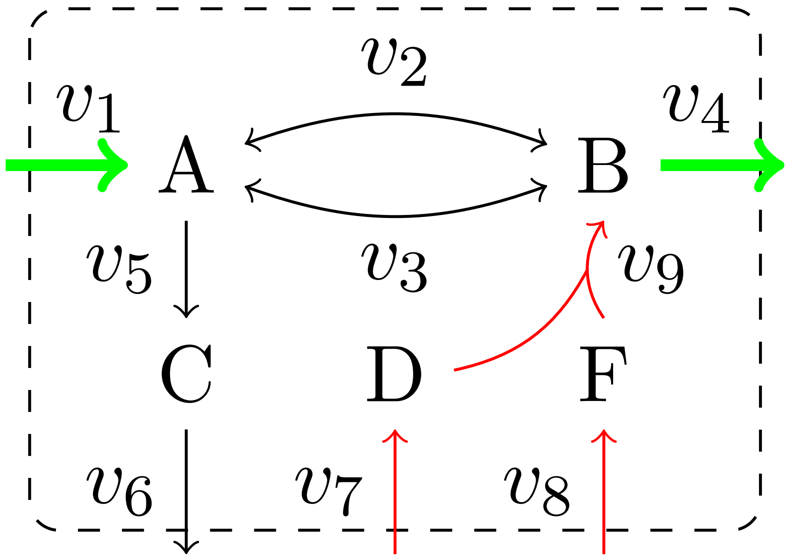
$$\mathbf{lb \leq v \leq ub}$$

$$\mathbf{v_i = E_1,\ i \in \mathcal{L}}$$

...

Gene Loss using CBM

- Blocked reactions are reactions that **can not** carry flux under given conditions
- We expect that these reactions will be loosed through evolution
- Given a set of conditions $\mathbf{E}$, we can find blocked reactions through simulation



$$\max \; c^T v$$

s.t

$$\mathbf{Sv=0}$$

$$\mathbf{lb \leq v \leq ub}$$

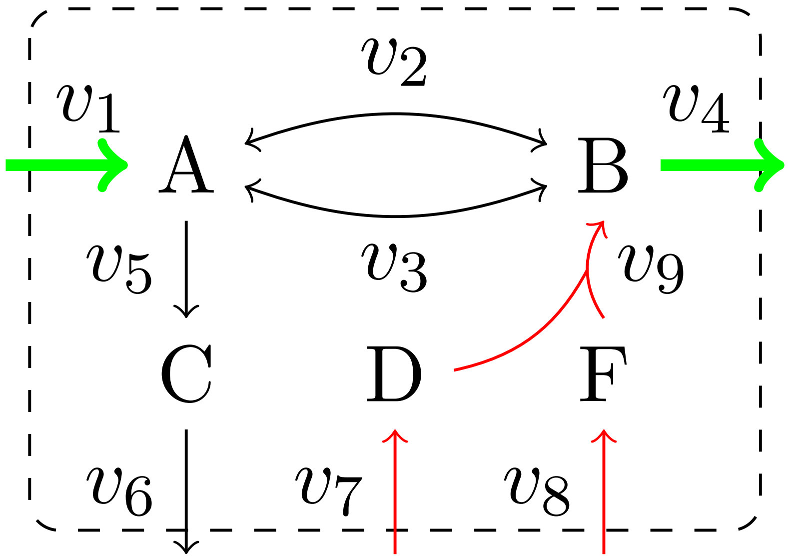
$$\mathbf{v_i = E_1, i \in \mathcal{L}}$$

...

...

Gene Loss using CBM

- Blocked reactions are reactions that **can not** carry flux under given conditions
- We expect that these reactions will be loosed through evolution
- Given a set of conditions $\mathbf{E}$, we can find blocked reactions through simulation



$\mathit{max}\ c^T v$
s.t
Sv=0
lb ≤ v ≤ ub
$v_i = \mathbf{E}_1, i \in \mathcal{L}$
...

...

$\mathit{max}\ c^T v$
s.t
Sv=0
lb ≤ v ≤ ub
$v_i = \mathbf{E}_n, i \in \mathcal{L}$
...

It is possible to determine conditions $\mathbf{E}$ to obtain most of blocked reactions?

A Bi-Level Optimization Approach

A Bi-Level Optimization Approach

First, we find number of blocked reactions

A Bi-Level Optimization Approach

First, we find number of blocked reactions

$$max \sum_{i=1}^{n} f_i^+ + f_i^-$$

subject to

$$\mathbf{Sv} = 0 \qquad (1)$$

$$\mathbf{lb} \leq \mathbf{v} \leq \mathbf{ub} \qquad (2)$$

$$\mathbf{v}_L = \mathbf{E},\ L \in \mathcal{L} \qquad (3)$$

$$f_i^+,\ f_i^- \in \{0, 1\} \qquad (4)$$

$$f_i^+ + f_i^- \leq 1 \qquad (5)$$

$$v_i \geq \epsilon f_i^+ - M f_i^- \qquad (6)$$

$$v_i \leq -\epsilon f_i^- + M f_i^+ \qquad (7)$$

A Bi-Level Optimization Approach

First, we find number of blocked reactions

$$max \sum_{i=1}^{n} f_i^+ + f_i^-$$

subject to

$$\mathbf{Sv} = 0 \quad (1)$$

$$\mathbf{lb} \leq \mathbf{v} \leq \mathbf{ub} \quad (2)$$

$$\mathbf{v}_L = \mathbf{E},\ L \in \mathcal{L} \quad (3)$$

$$f_i^+,\ f_i^- \in \{0,1\} \quad (4)$$

$$f_i^+ + f_i^- \leq 1 \quad (5)$$

$$v_i \geq \epsilon f_i^+ - M f_i^- \quad (6)$$

$$v_i \leq -\epsilon f_i^- + M f_i^+ \quad (7)$$

Next, we pick settings $\mathbf{E}$ to maximize blocked reactions

A Bi-Level Optimization Approach

First, we find number of blocked reactions

$$max \sum_{i=1}^{n} f_i^+ + f_i^-$$

subject to

$$\mathbf{Sv} = 0 \qquad (1)$$

$$\mathbf{lb} \leq \mathbf{v} \leq \mathbf{ub} \qquad (2)$$

$$\mathbf{v}_L = \mathbf{E},\ L \in \mathcal{L} \qquad (3)$$

$$f_i^+,\ f_i^- \in \{0,1\} \qquad (4)$$

$$f_i^+ + f_i^- \leq 1 \qquad (5)$$

$$v_i \geq \epsilon f_i^+ - M f_i^- \qquad (6)$$

$$v_i \leq -\epsilon f_i^- + M f_i^+ \qquad (7)$$

Next, we pick settings $\mathbf{E}$ to maximize blocked reactions

$$min\ z$$

subject to

$$\mathbf{lb_L} \leq \mathbf{E} \leq \mathbf{ub_L}$$

$$z = max \sum f_i^+ + f_i^-$$

subject to

$$\mathbf{Sv}=0 \qquad (1)$$

$$\mathbf{lb}\leq\mathbf{v}\leq\mathbf{ub} \qquad (2)$$

$$\mathbf{v}_L=\mathbf{E},\ L\in\mathcal{L} \qquad (3)$$

... ...

$$v_i\leq-\epsilon f_i^- + M f_i^+ \qquad (7)$$

What is next?

- In general, Mixed Integer Bi-Level problems are considered as an “still unsolved”
- Links environmental conditions to evolution processes
- Developing algorithms to solve these type of problems are of interest in System Biology

“Nothing in Biology Makes Sense Except in the Light of Evolution”

– Theodosius Dobzhansky (1900-1975)

Thanks for your attention!