Genomics of drug resistance

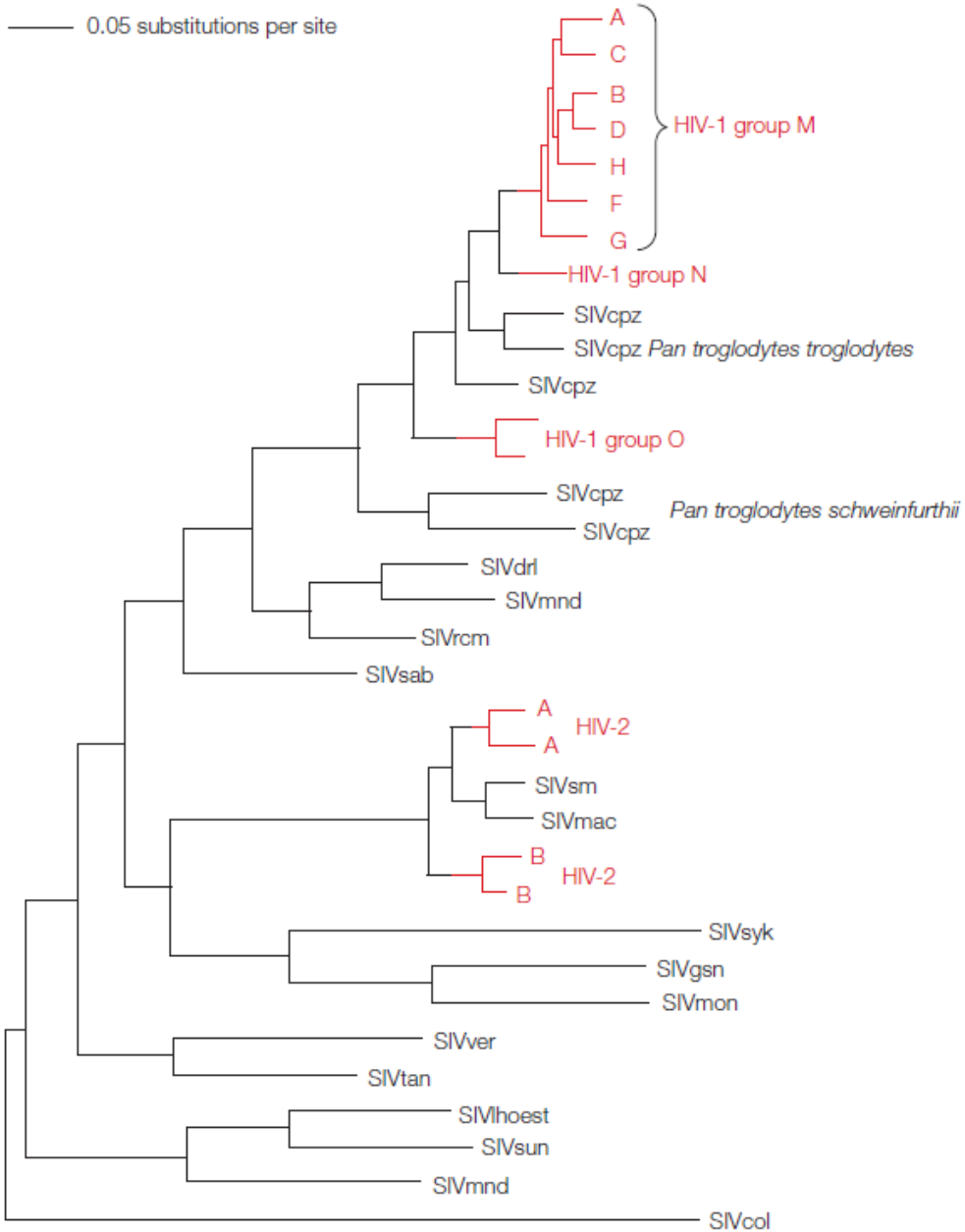
Alan Moses

Some work with Maria Safi and Ryan Lilien

Motivation

- Drug resistance is prevalent in bacterial and viral diseases such as malaria, tuberculosis and AIDS
- What is drug resistance?
- I’m going to use examples from HIV

Origin of the HIV virus(es)



ML Tree based on HIV pol gene, from Rambaut et al. 2004

- HIV-1 and HIV-2 probably emerged from monkeys independently
- Evidence for multiple transfers of HIV from other primates to humans Hahn et al. *Science 2000*
- Most infections around the world are HIV-1 type M, which came from chimpanzee
- Other strains are primarily found in West Africa
- HIV-1 Probably emerged around 1930 in Africa

How does a bloodborne virus get from monkeys to humans?

Hahn et al. *Science 2000*

Fig. 6. Bushmeat market in west central Africa. A chimpanzee (separated into skull, rib cage, limbs, and various internal organs) is shown in the middle of the photo, along with other smoked or fresh meat, including two blue duikers and a spot-nosed guenon in the upper right corner. [Photograph courtesy of Karl Ammann]

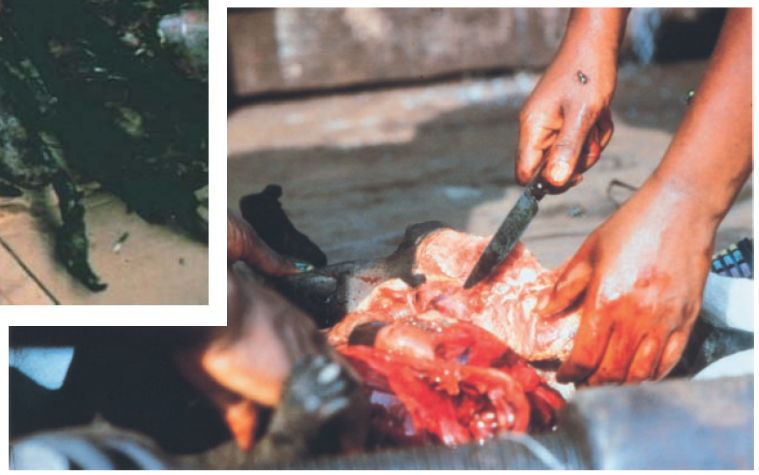
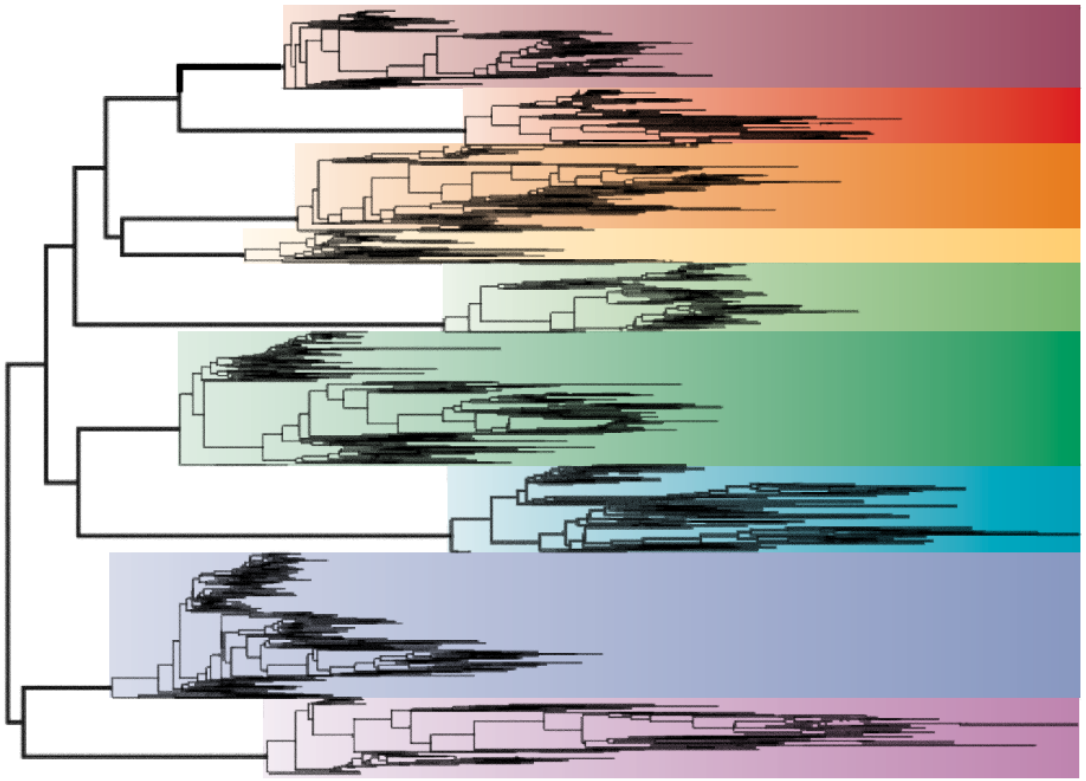


Fig. 4. Human exposure to primate blood during food preparation. [Photograph courtesy of Karl Ammann]

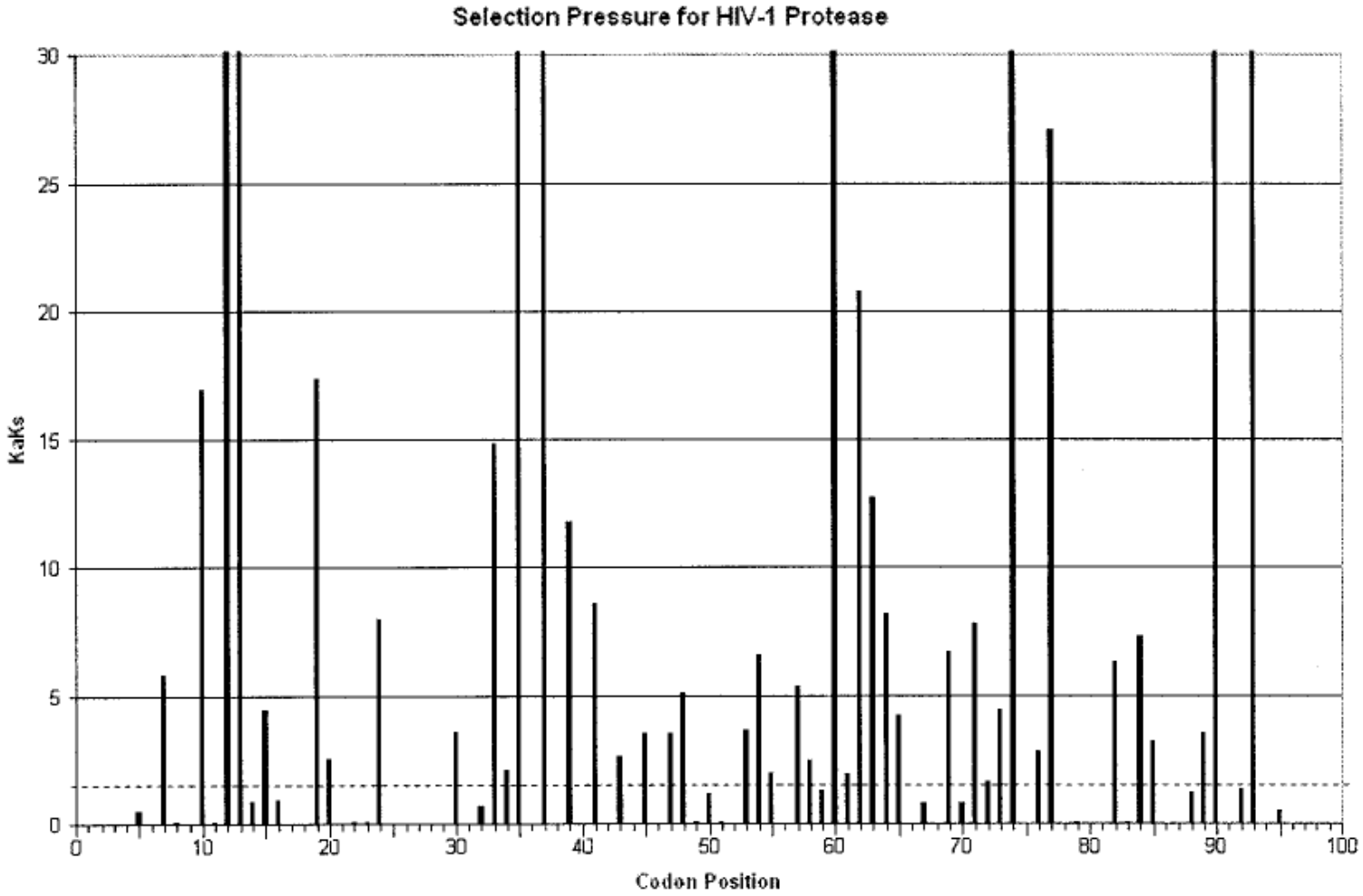
Evolution doesn’t stop!



- Each colour represents sequences from one patient
- Natural selection induced by the immune system drives rapid evolution
- Sequence diversity at any time is low

Tree based on HIV env gene, from Rambaut et al. 2004

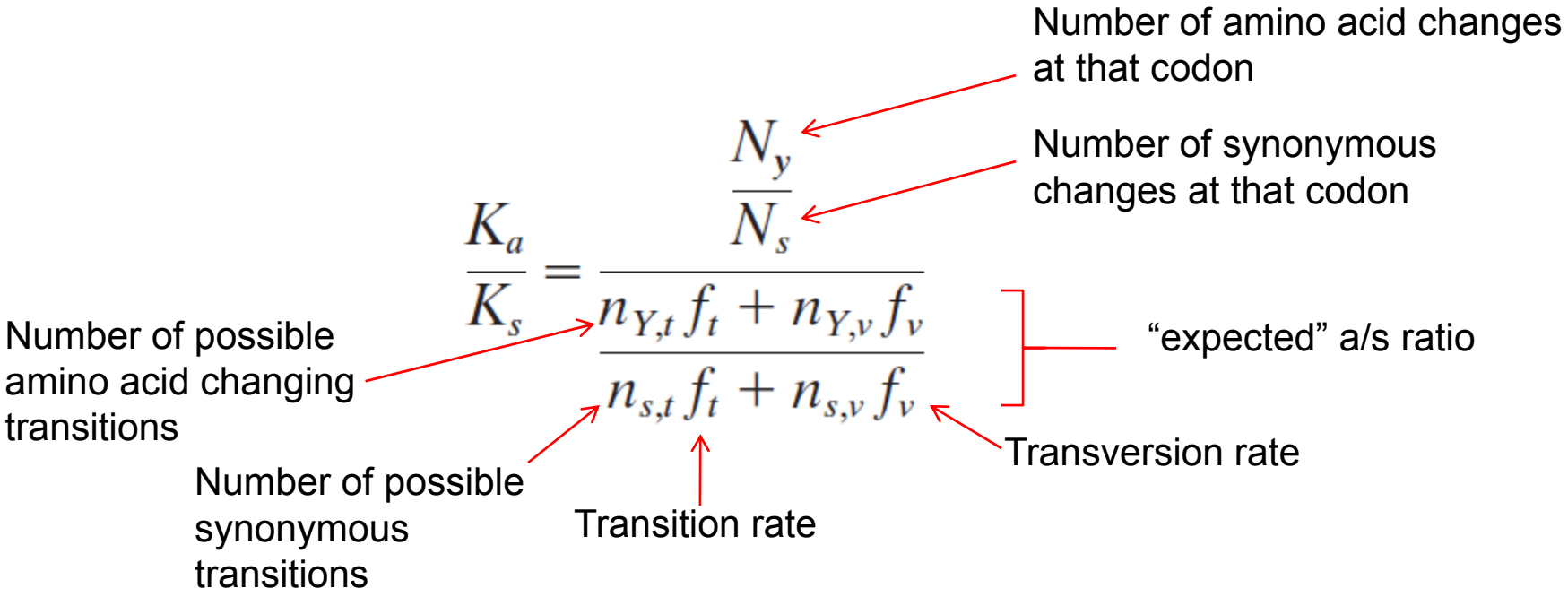
Evolution doesn't stop!



Position specific estimates of Ka/Ks from 40,000 patient samples from Chen et al. 2004

- Strong evidence for natural selection at key positions in HIV proteins

Position specific a/s ratio



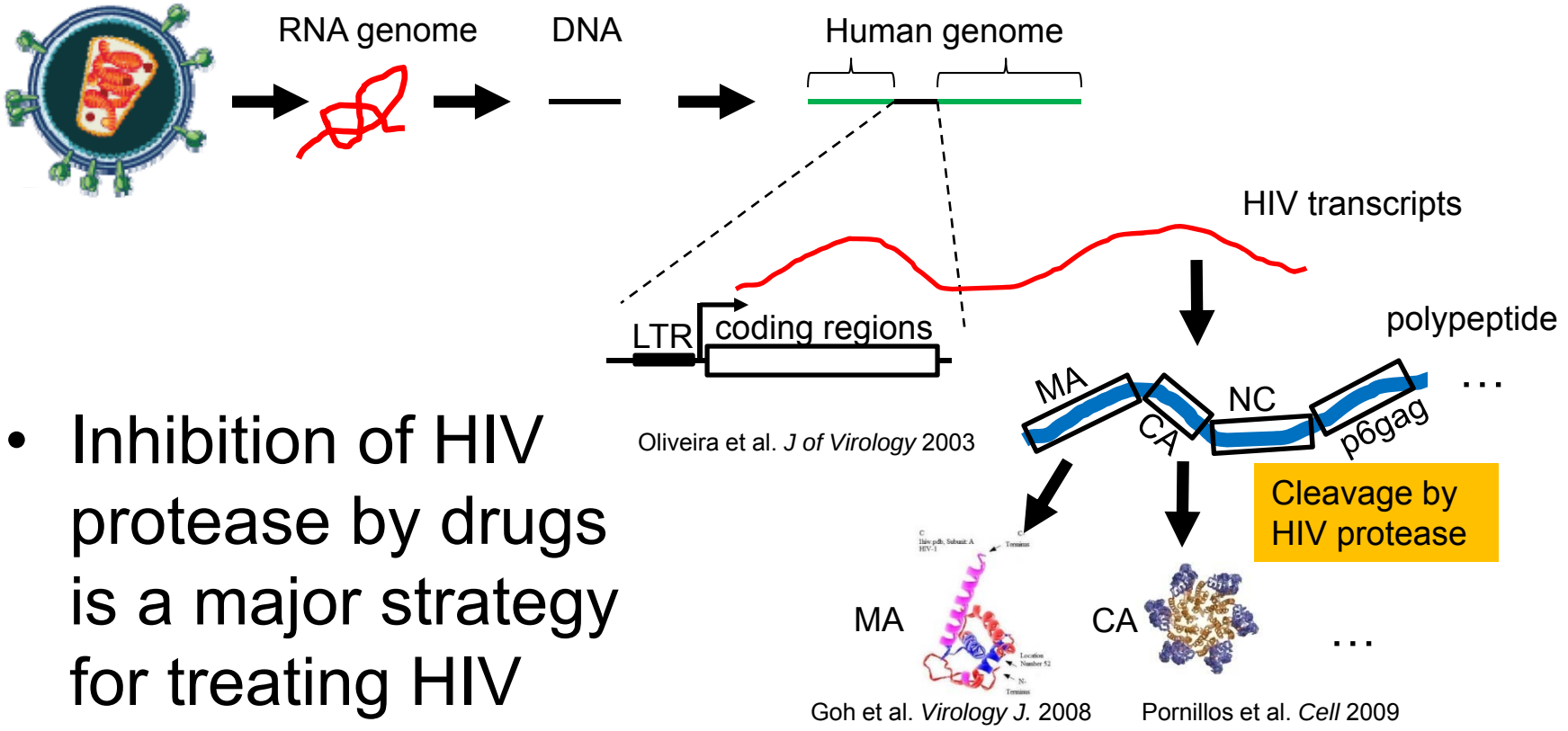
- Usually Ka/Ks (or dN/dS) is computed for a whole gene/protein.
- Several programs now exist to compute position specific estimates using Bayesian or ML inference
- Beware the use of a/s ratios in population samples:

we study the expected dN/dS ratio for samples drawn from a single population under selection, and we find that in this context, dN/dS is relatively insensitive to the selection coefficient. Moreover, the hallmark signature of positive selection over divergent lineages, dN/dS>1, is violated within a population. For population samples, the relationship between selection and dN/dS does not follow a monotonic function, and so it may be impossible to infer selection pressures from dN/dS.

Kryazhimskiy & Plotkin *PLoS Genetics* 2008

Background on HIV protease

- HIV protease is a very well-studied protein where we can understand resistance mutations



Oliveira et al. *J of Virology* 2003

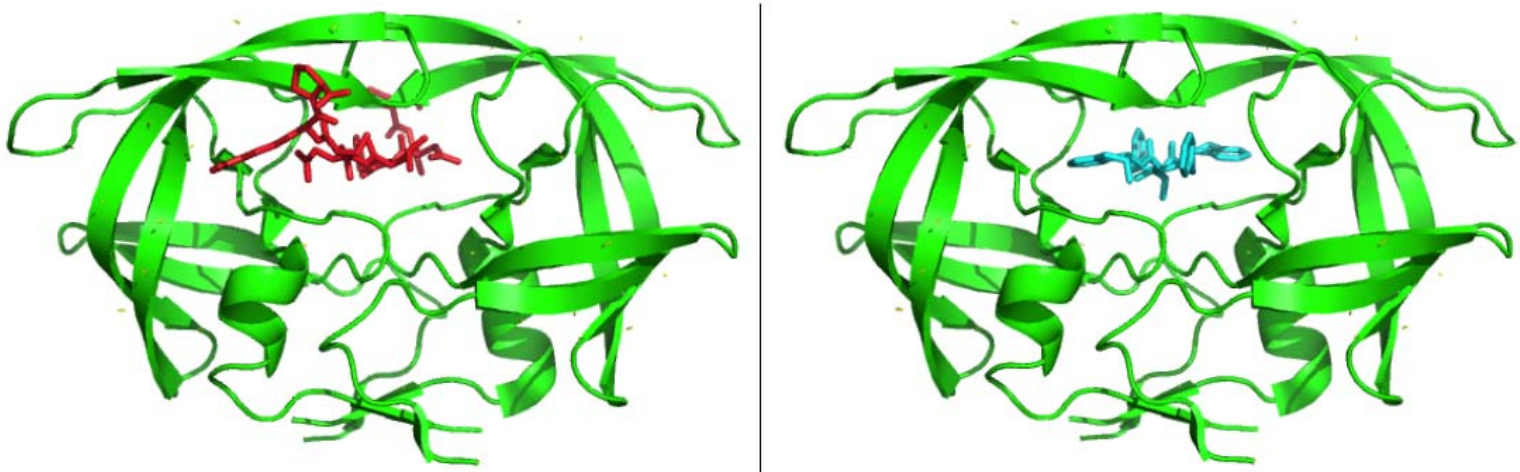
Goh et al. *Virology J.* 2008

Pornillos et al. *Cell* 2009

- Inhibition of HIV protease by drugs is a major strategy for treating HIV

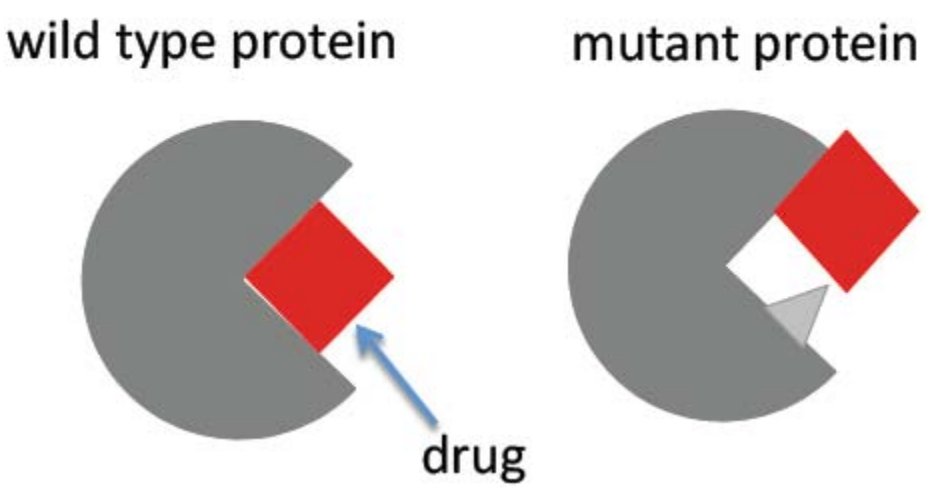
Kohl et al. *PNAS* 1988

How do mutations confer drug resistance?



Prabhu-Jeyabalan et al., 2002 Andersson et al., 2003

HIV protease bound to native substrate (left) and drug (right)



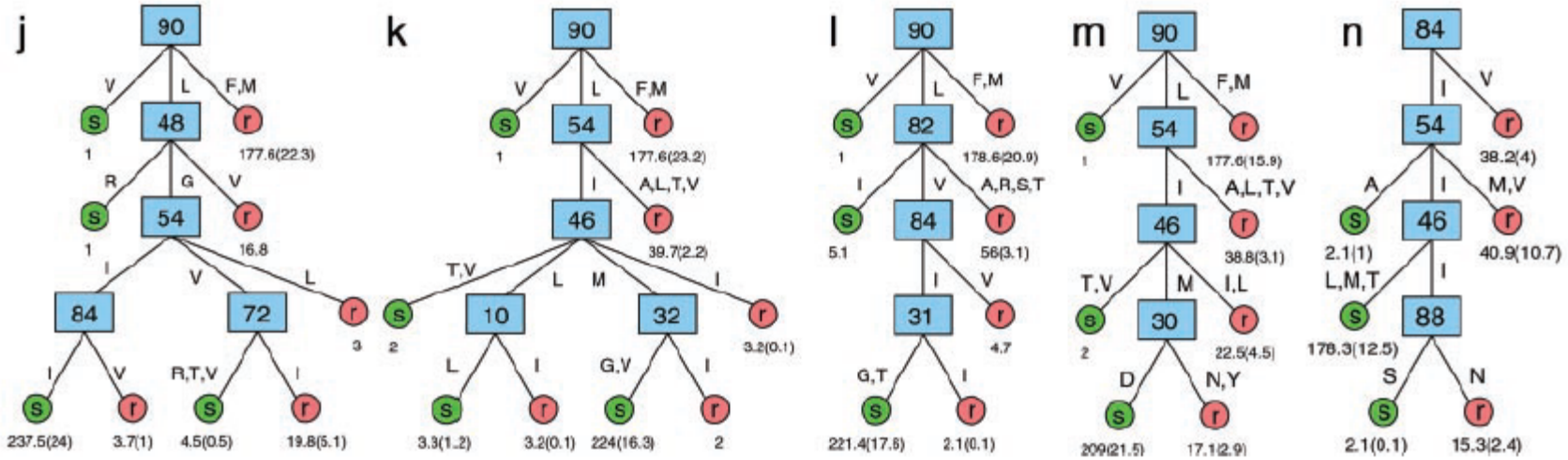
- E.g., Mutation in the active site that prevent drug binding

Predicting drug resistance

- Knowing when/how drug resistance would evolve could be useful:
  - Treat patients with drugs that will work on their HIV strains (personalized medicine)
  - Develop new drugs to which resistance can't evolve (easily)
- Several possible ways to predict evolution of drug resistance
  - Learn from patient sequencing data
  - Learn from lab experiments
  - Predict from first principles?

Diversity and complexity of HIV-1 drug resistance: A bioinformatics approach to predicting phenotype from genotype *PNAS* 2002

Niko Beerenwinkel*†‡, Barbara Schmidt†§, Hauke Walter§, Rolf Kaiser¶, Thomas Lengauer*‡, Daniel Hoffmann‖, Klaus Korn§, and Joachim Selbig*,**



- 471 samples, 14 different HIV drugs
- Each position in the protease or reverse transcriptase gene is a potential predictor, with 20 possible states
- Train decision trees for each drug

				Learning phase			Leave-one-out experiments			
Drug	Cutoff	No. of samples	Resistant fraction, %	Minimal split*	No. of interior vertices	Training error, %	Prediction error, %	Sensitivity, %	Specificity, %	Observed resistance factor
ZDV	8.5	456	58.1	2	5	8.8	10.7	92.1	85.8	419
ddC	2.5	456	43.0	7	5	23.7	26.3	58.2	85.4	1
ddI	2.5	456	49.1	7	4	25.7	32.0	73.7	62.5	3
d4T	2.5	456	38.6	7	4	21.5	25.4	63.1	81.8	2
3TC	8.5	452	54.4	2	4	7.7	10.4	87.4	92.2	13
ABC	2.5	445	66.3	5	5	13.5	15.5	92.5	68.7	4
NVP	8.5	457	45.1	2	7	7.0	9.6	82.0	97.2	407
DLV	8.5	455	36.5	2	5	8.1	10.5	77.7	96.2	168
EFV	8.5	443	35.9	2	6	7.7	10.2	79.9	95.4	7
SQV	3.5	465	46.7	2	5	11.2	12.5	87.6	87.5	39
IDV	3.5	469	48.8	2	5	11.2	10.9	89.5	88.8	32
RTV	3.5	469	50.1	2	4	9.0	10.2	89.8	89.7	33
NFV	3.5	468	53.6	2	4	9.6	11.5	89.6	87.1	93
APV	3.5	277	32.9	2	4	10.5	12.6	82.4	89.8	3

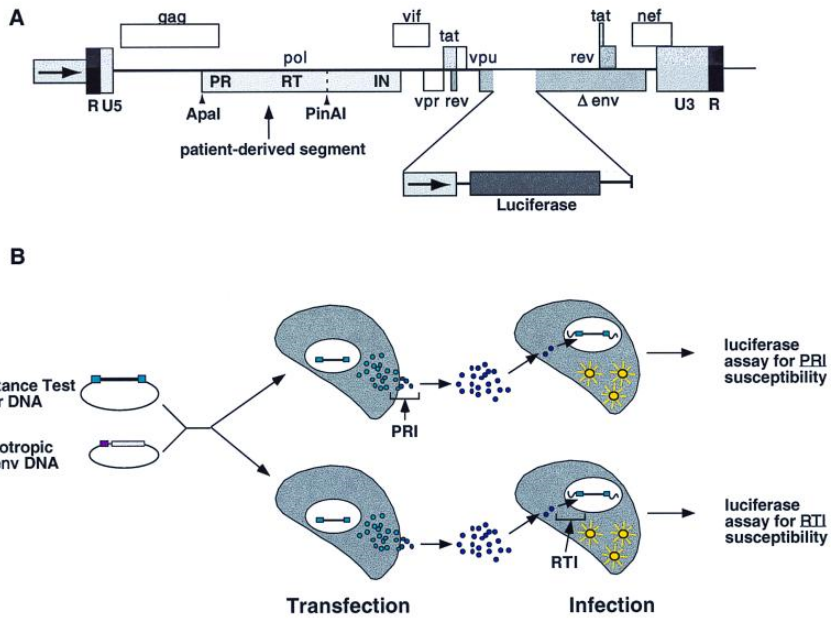
LETTERS

nature genetics

Nature Genetics 2011

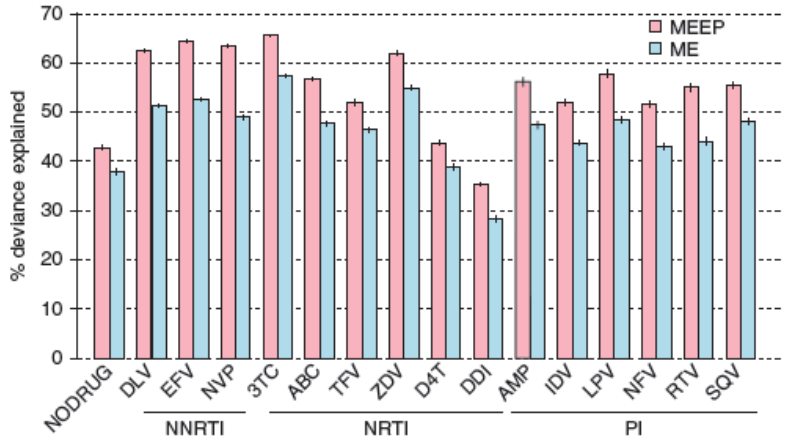
A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase

Trevor Hinkley[1], João Martins[1], Colombe Chappey[2,3], Mojgan Haddad[2], Eric Stawiski[2,3], Jeannette M Whitcomb[2], Christos J Petropoulos[2] & Sebastian Bonhoeffer[1]



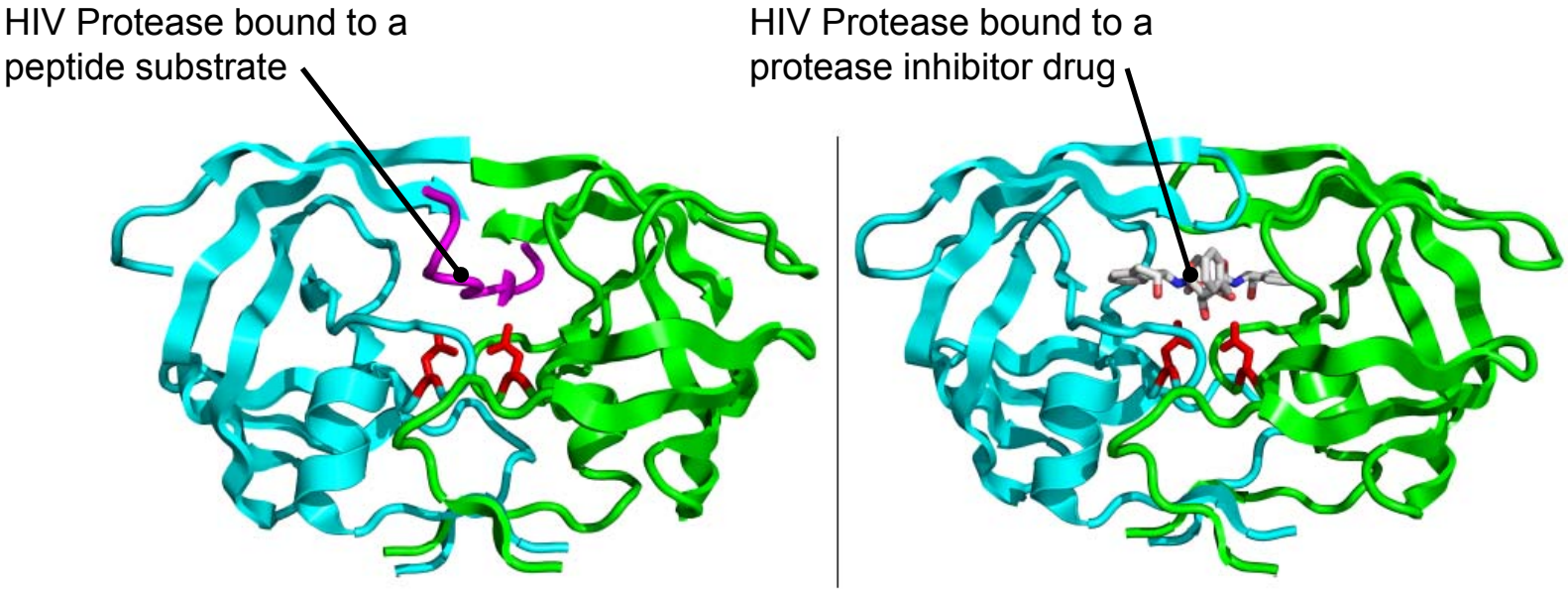
Petropoulos et al. 2000 virologic.com

- ~70,000 virus samples from patients
- Genes from these viruses are inserted into a test strain, and the effect on replicative capacity is measured in the lab
- “population sequencing” to get estimates of genotypes for viral proteins in each strain
- Replicative capacity is predicted from genotype using regularized regression model
- Predictive power assessed using cross-valedation
- “fitness landscape” for HIV



Introduction

- HIV evolves resistance to protease inhibitors through mutations in the active site of the enzyme



Prabu-Jeyabalan et al. *Structure* 2002

Andersson et al. *Eur. J. Biochem.* 2003

- "Resistant" mutations prevent binding of the drugs, but allow binding of the peptide substrates

A simple adaptive process

- Imagine that HIV has reached some “equilibrium” in its evolution on a (possibly dynamic) fitness landscape (including immune system, etc.)

- Adding protease inhibitors greatly reduces the fitness of the “wt” genotype

- Over time the population adapts, heading towards a new equilibrium

A simple adaptive process

- Assume Protease is a Michaelis-Menten enzyme, and viral growth is proportional to the rate of cleavage

$$v_0 = \frac{k_{cat}[S]\,[\text{Protease}]}{k_M\,(1+[\text{Drug}]/k_I) + [S]}$$

Compare relative fitness of genotypes a and b

$$s_{ab} \equiv \frac{\text{growth rate b} - \text{growth rate a}}{\text{growth rate a}} \approx \frac{k_{Ma}\,k_{Ib}}{k_{Mb}\,k_{Ia}} - 1 \approx e^{\Delta G_{Sa}-\Delta G_{Sb}-(\Delta G_{Da}-\Delta G_{Db})} - 1$$

Selection coefficient

e.g., Wylie & Shakhnovich *PNAS* 2011

Binding energies of the substrate and drug

These are determined by the 3D conformation of the amino acid residues in the active site

In general, the selection co-efficients will depend on the genotype in a complicated way.

A simple adaptive process

- Assume Protease is a Michaelis-Menten enzyme, and viral growth is proportional to the rate of cleavage

$$v_0 = \frac{k_{cat}[S]\,[\text{Protease}]}{k_M\,(1+[\text{Drug}]/k_I) + [S]}$$

Compare relative fitness of genotypes a and b

$$s_{ab} \equiv \frac{\text{growth rate b - growth rate a}}{\text{growth rate a}} \approx \frac{k_{Ma}\,k_{Ib}}{k_{Mb}\,k_{Ia}} - 1 \approx e^{\Delta G_{Sa}-\Delta G_{Sb}-(\Delta G_{Da}-\Delta G_{Db})} - 1$$

Selection coefficient

e.g., Wylie & Shakhnovich *PNAS* 2011

Binding energies of the substrate and drug

Fixation probability

$$F_{ab} = \frac{1 - e^{-2s_{ab}}}{1 - e^{-2Ns_{ab}}}$$

Kimura 1962

→ No population genetics needed

How do we get binding energies?

- Most resistant genotypes that have been observed contain fewer than 3 amino acid differences from the wt
- Evaluate the energy of each amino acid sequence by re-optimizing the co-crystal structures of the protease with drug and native substrate
- Protease active site is ~15 amino acid positions, and several are required for catalysis, so they are invariant

→ "only" $\binom{11}{2}$ x 20 x 20 = 22,000 possible amino acid sequences

- Too hard (for us) to do detailed molecular dynamics simulations for all these protease sequences

How do we get binding energies for a large fitness lanscapes?

- Since the structures of the mutants are expected to be very similar to the wt, we can rapidly rule out 'bad' structures using fast DEE-based algorithms

Safi & Lilien *J. Comput. Chem.* 2010 Safi & Lilien *J. Chem. Inf. Model.* 2012

- Don't actually compute accurate energies for these genotypes, just assign them all to a qualitatively 'bad' state
- Use MM-PBSA implemented in AMBER package to evaluate energies for structures that remain

Case et al. *AMBER 12 UCSF*

- Only 787 of 22,000 amino acid sequences remain We can actually compute this fitness landscape!

What does a fitness landscape look like?

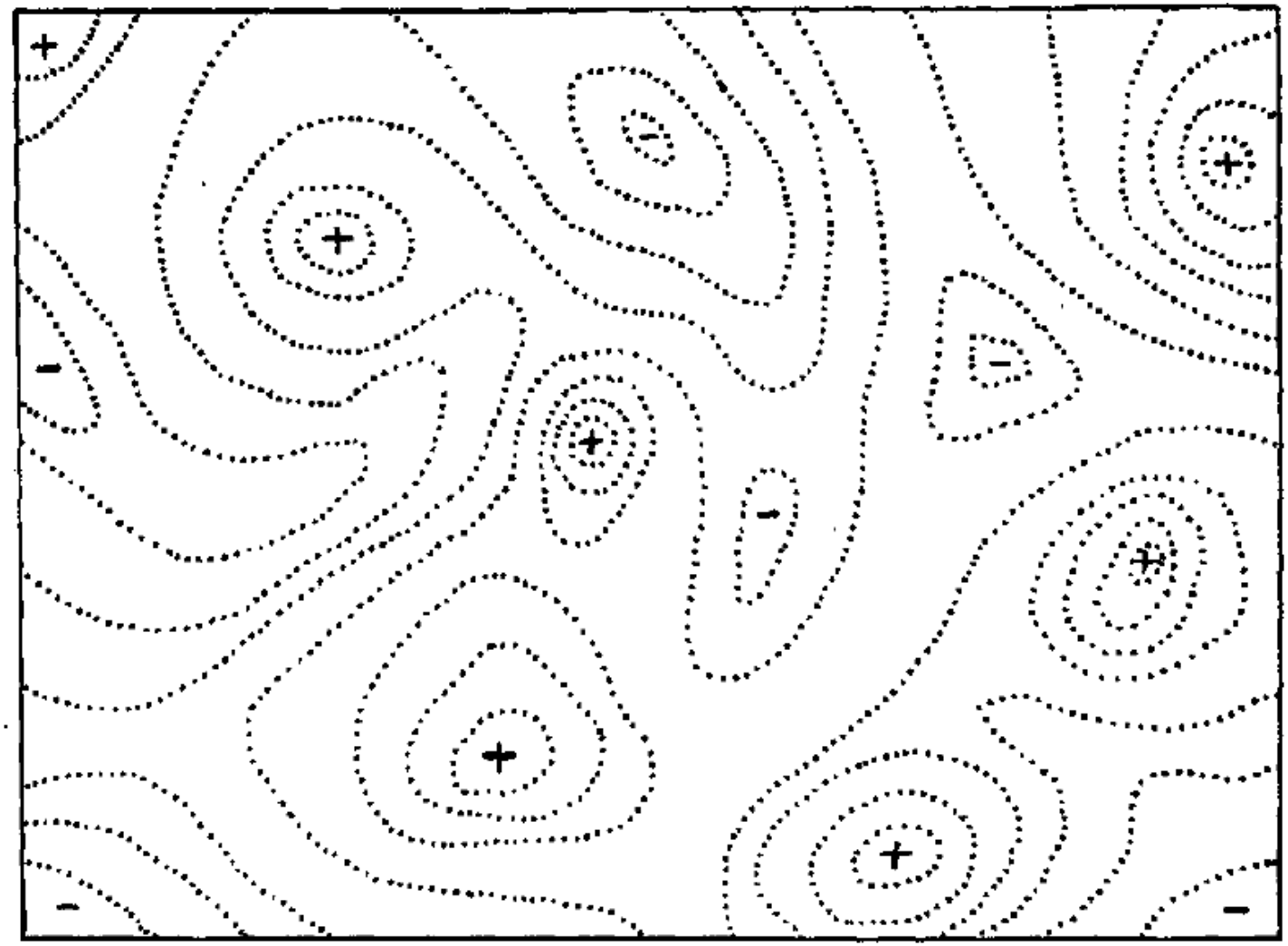
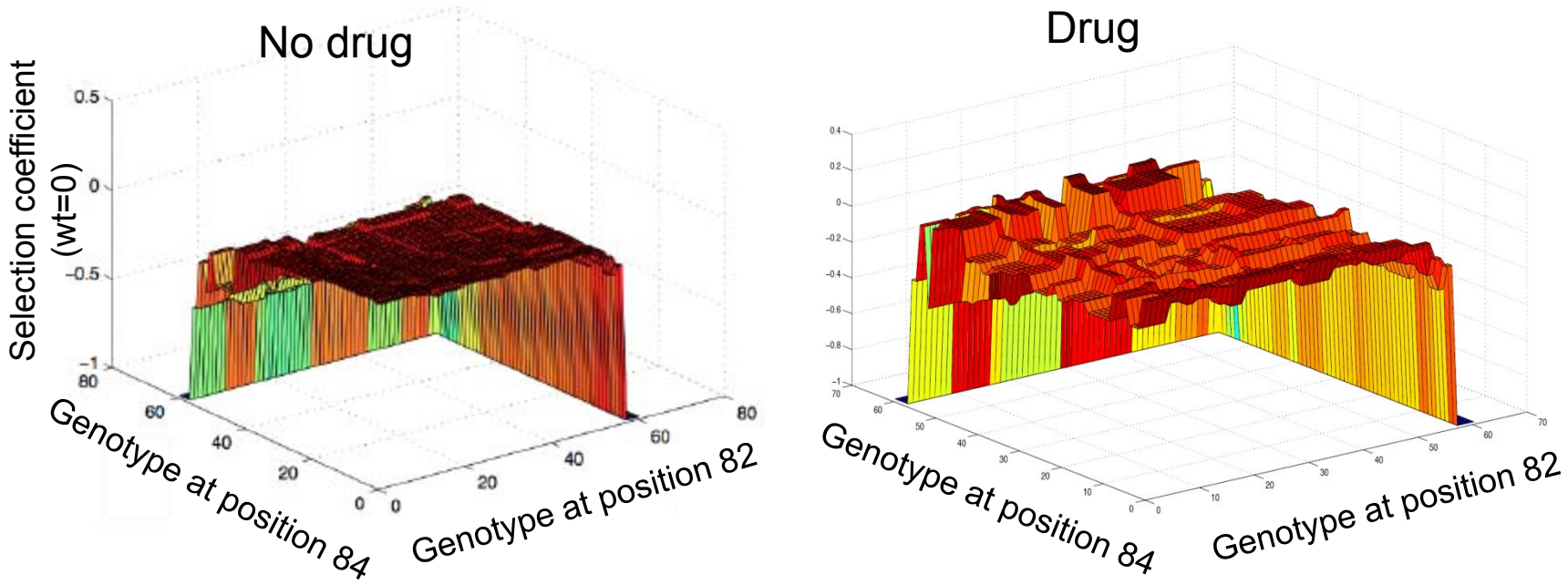


FIGURE 2.—Diagrammatic representation of the field of gene combinations in two dimensions instead of many thousands. Dotted lines represent contours with respect to adaptiveness.

Wright 1932

A smaller problem

- Consider only 2 positions in the active site, V82 and I84
- These positions contain many of the known resistant mutations



→ These representations of fitness landscapes are not actually useful, because the ordering of genotypes on the X-Y plane is arbitrary

Genotypes connected in these diagrams are ***equidistant***

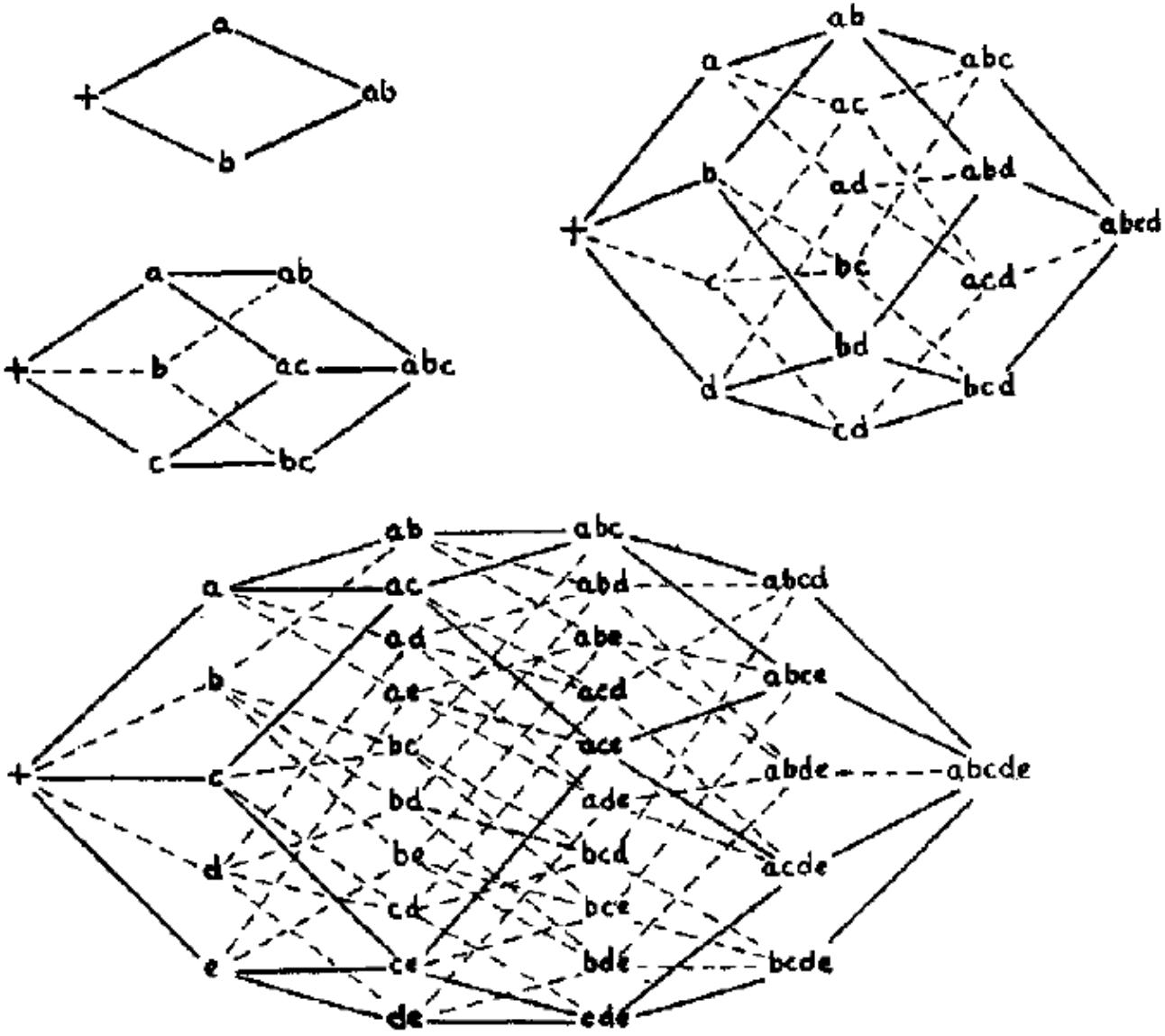


FIGURE 1.—The combinations of from 2 to 5 paired allelomorphs.

Wright 1932

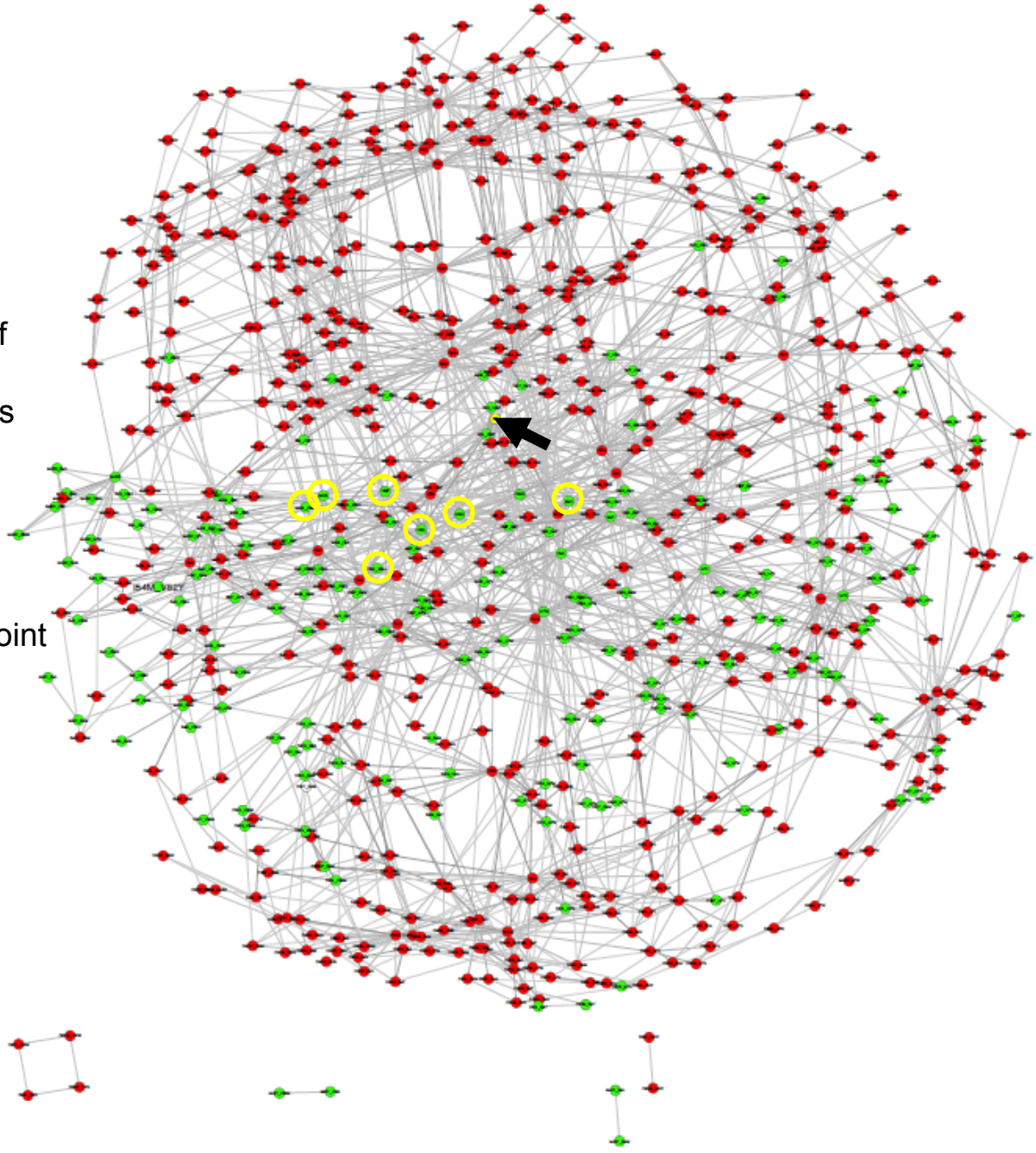
Genotype network visualized with cytoscape

Fitness (from the perspective of the virus) of genotypes is indicated by colours: Red nodes are deleterious, green are beneficial

Connections between nodes indicate the number of single point mutations between them

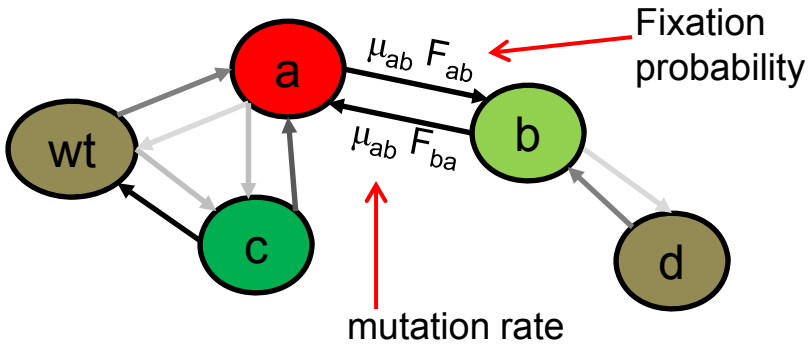
wt

E.g., of known drug resistant mutants



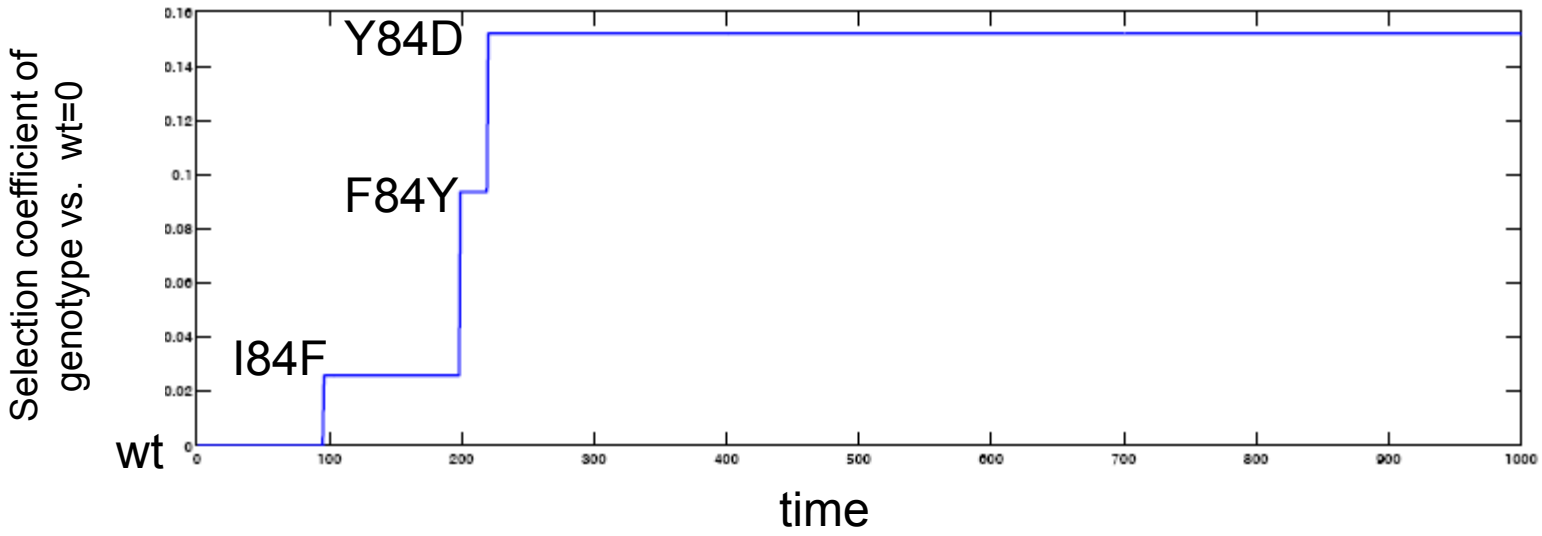
Evolution is a random walk on the genotype graph

Each pair of nodes separated by a single point mutation has bi-directional connections



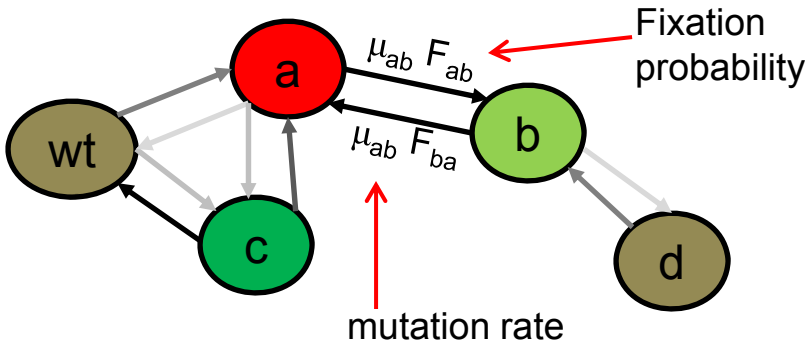
simulate evolution on our fitness landscapes

Simulation is done at the codon level, so this is a Markov chain with 1000s of states

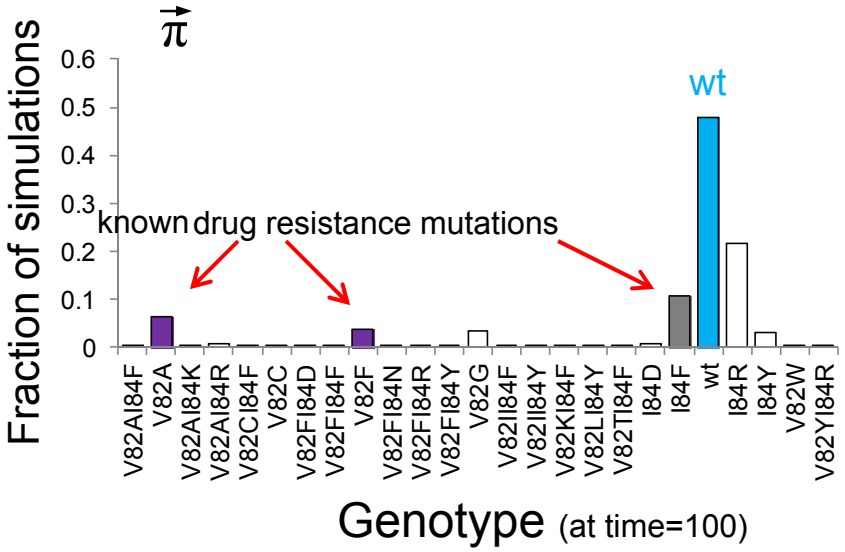


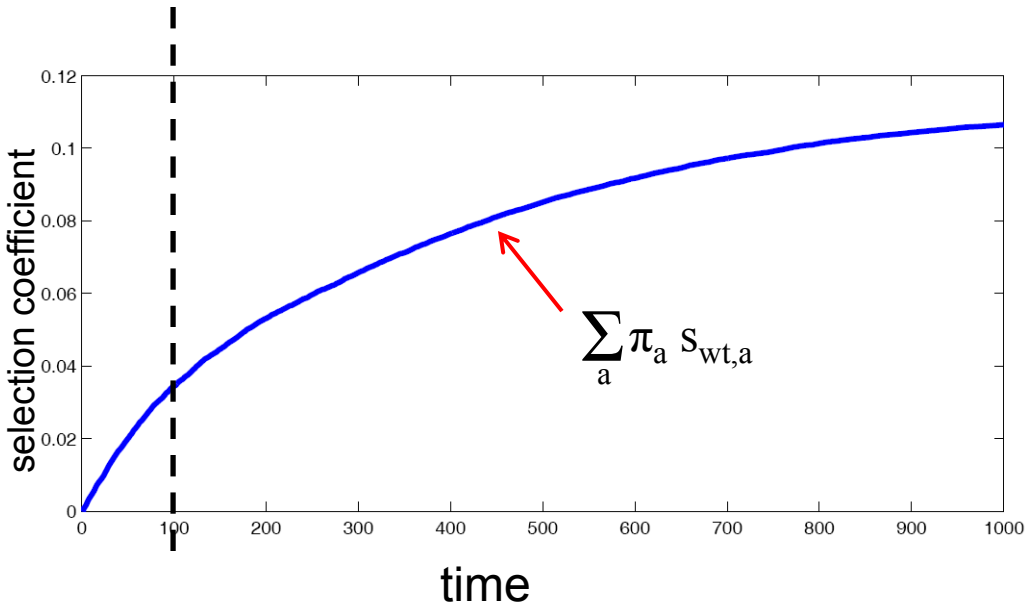
Evolution is a random walk on the genotype graph

Each pair of nodes separated by a single point mutation has bi-directional connections

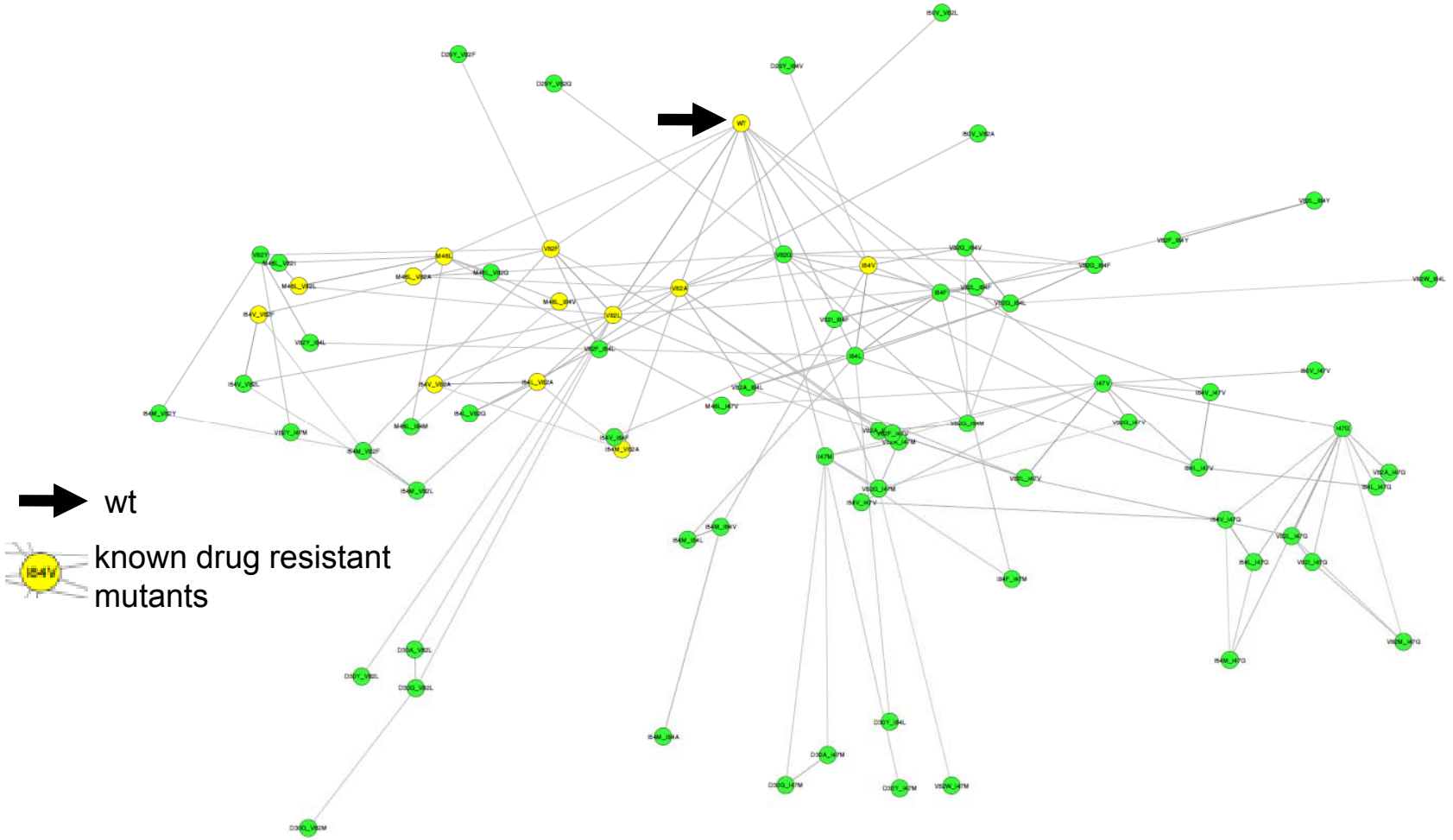


simulate evolution on our fitness landscapes



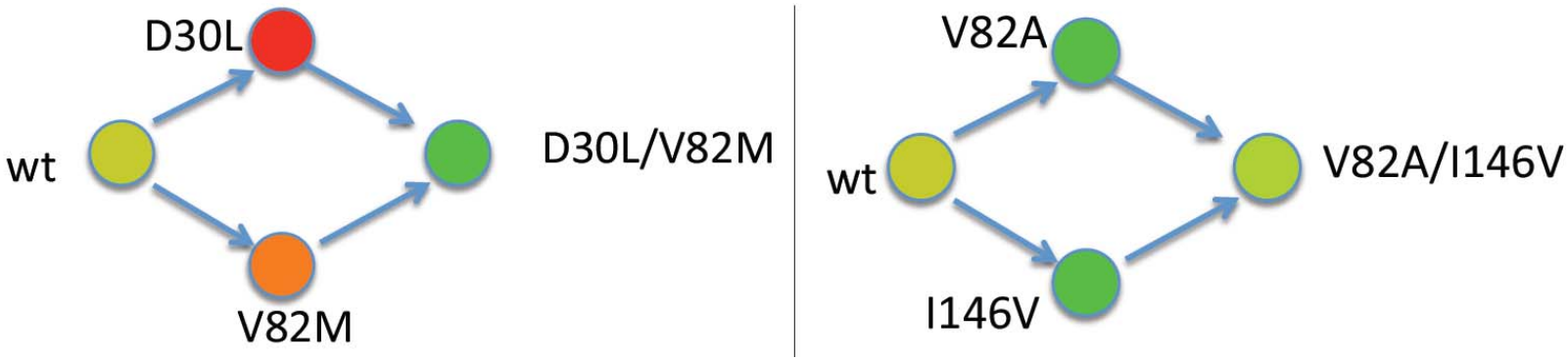


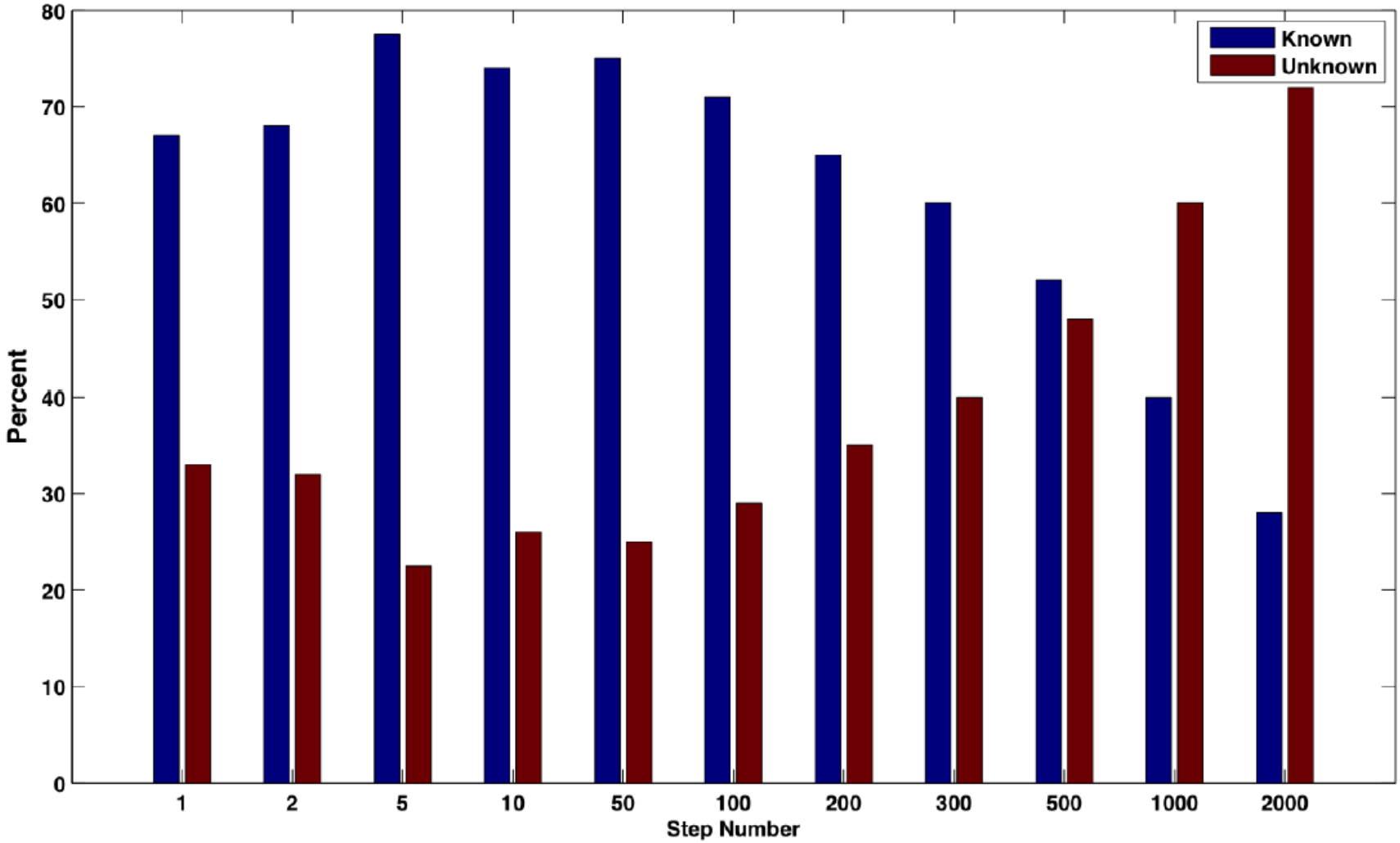
Most genotypes are not visited by the evolutionary simulation



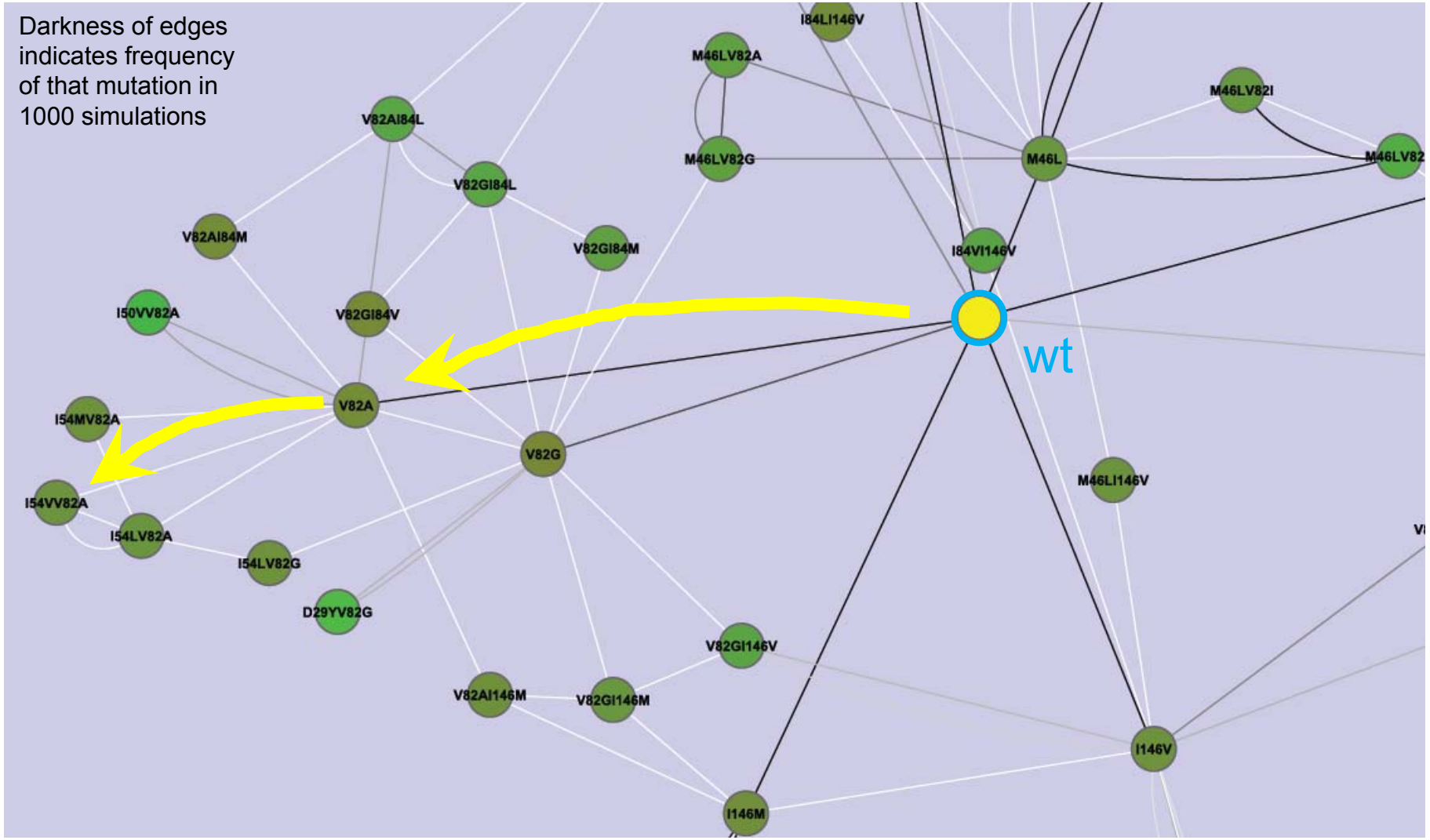
MI HIVdb =

Mutation (alone)	Structure (alone)	Evolution (Structure+Mutation)
0.01	0.22	0.32


D30L
wt
V82M
D30L/V82M
V82A
wt
I146V
V82A/I146V


Known
Unknown
Percent
Step Number
0
10
20
30
40
50
60
70
80
1
2
5
10
50
100
200
300
500
1000
2000

Compare to observed adaptive trajectories



This sequence of mutations was reported in a patient that developed resistance to this drug Eshleman et al. *J Infect Dis.* 2001

Predicting drug resistance

- Knowing when/how drug resistance would evolve could be useful:
  - Treat patients with drugs that will work on their HIV strains (personalized medicine)
  - Develop new drugs to which resistance can't evolve (easily)
- Several possible ways to predict evolution of drug resistance
  - Learn from patient sequencing data
  - Learn from lab experiments
  - Predict from first principles

Predict adaptation on computed genotype networks

Mutation is a non-negligable force in HIV evolution?

Acknowledgements etc.

Maria Mirza Safi

Ryan Lilien

Moses Lab

Alex Nguyen Ba

Louis-Francois Handfield

Gelila Tilahun

Bob Strome

Taraneh Zarin

Funding

Ontario

NSERC CRSNG

UNIVERSITY OF TORONTO

Canada Foundation for Innovation
Fondation canadienne pour l'innovation

Canadian Institutes of Health Research
Instituts de recherche en santé du Canada

Political Message

Open access: “real scientists do it in public”

www.plos.org www.biomedcentral.com