Kasia Bożek

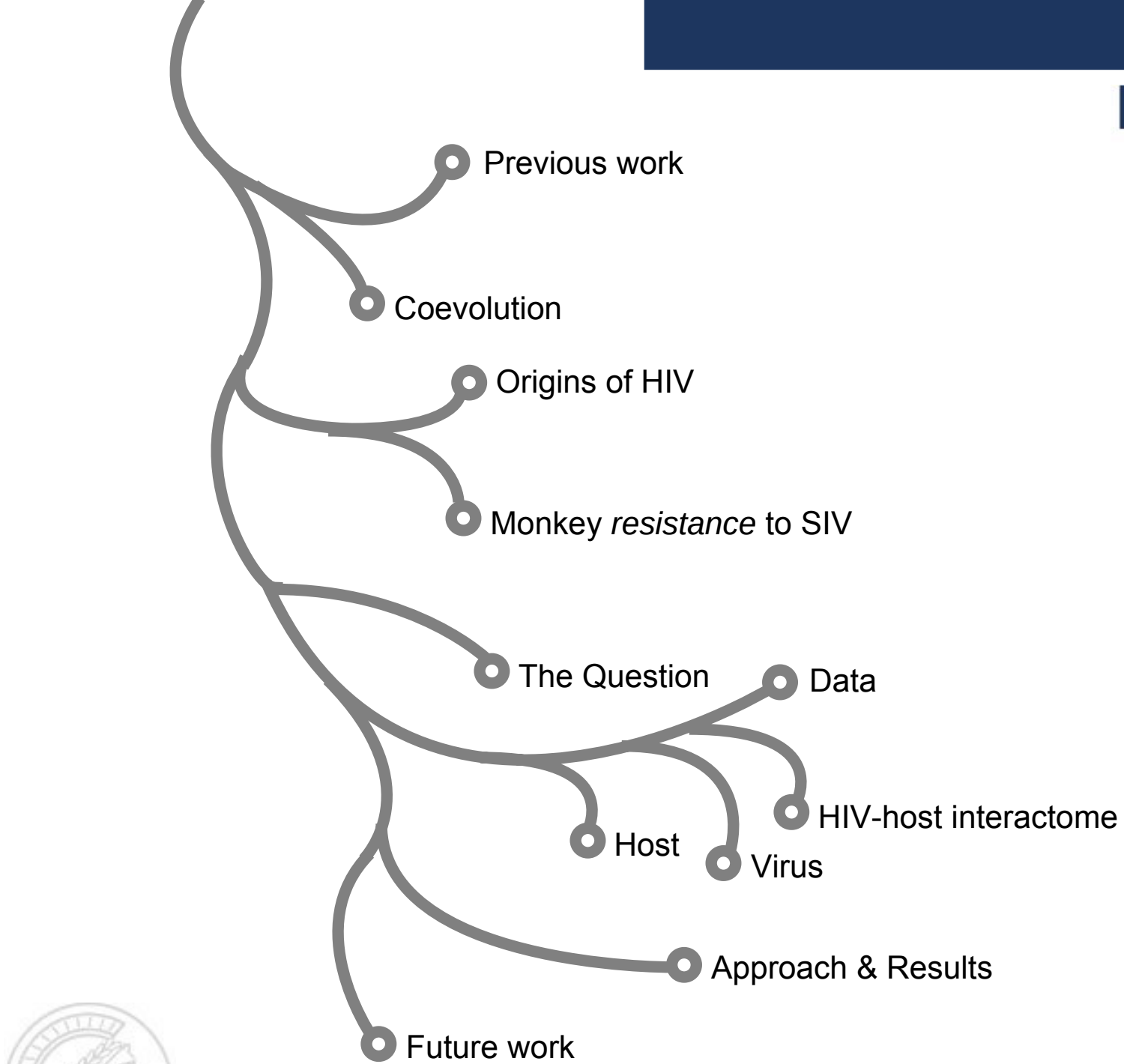
HIV-host coevolution

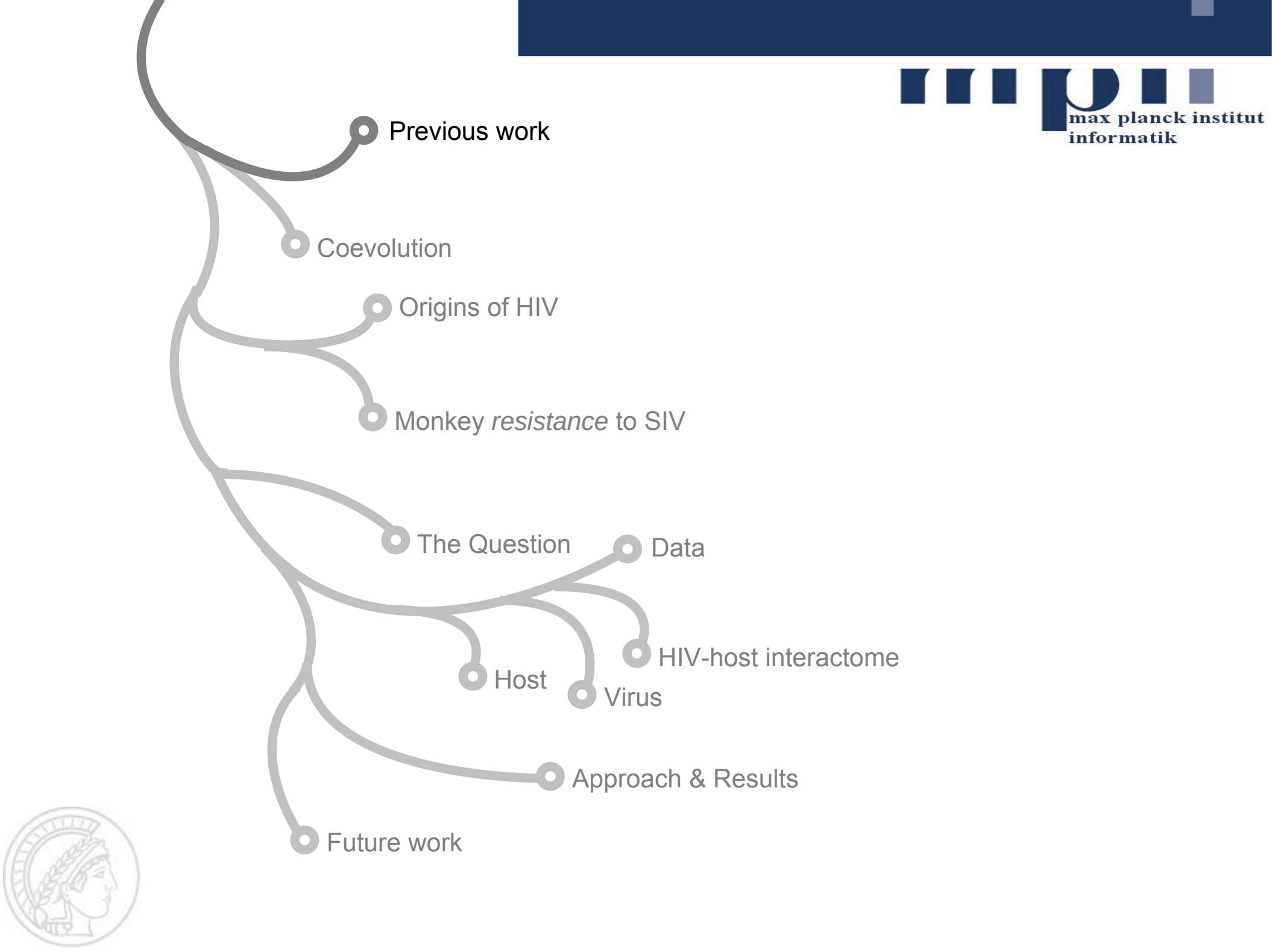
Max-Planck Institute for Computer Sciences

Computational Biology and Applied Algorithmics

Saarbrücken, Germany

KITP, 25.11.2008

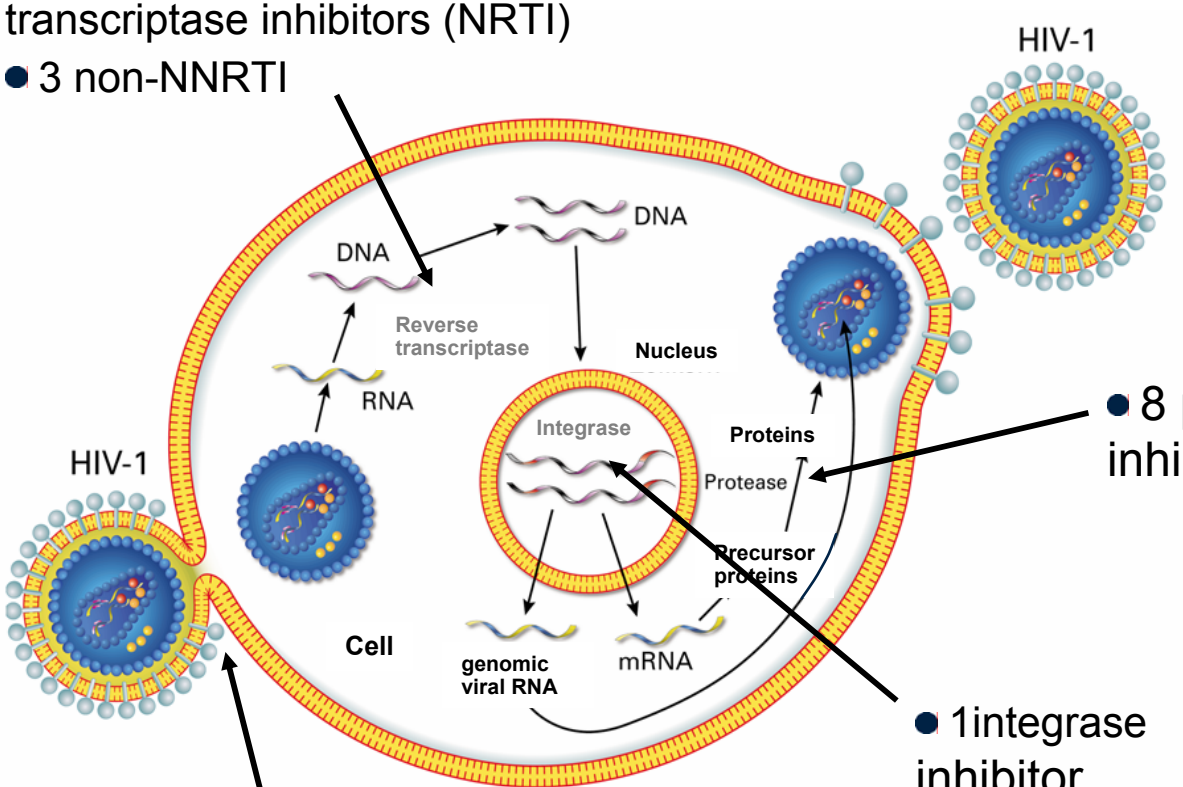

Previous work
Coevolution
Origins of HIV
Monkey *resistance* to SIV
The Question
Data
HIV-host interactome
Host
Virus
Approach & Results
Future work


Previous work
Coevolution
Origins of HIV
Monkey *resistance* to SIV
The Question
Data
HIV-host interactome
Host
Virus
Approach & Results
Future work

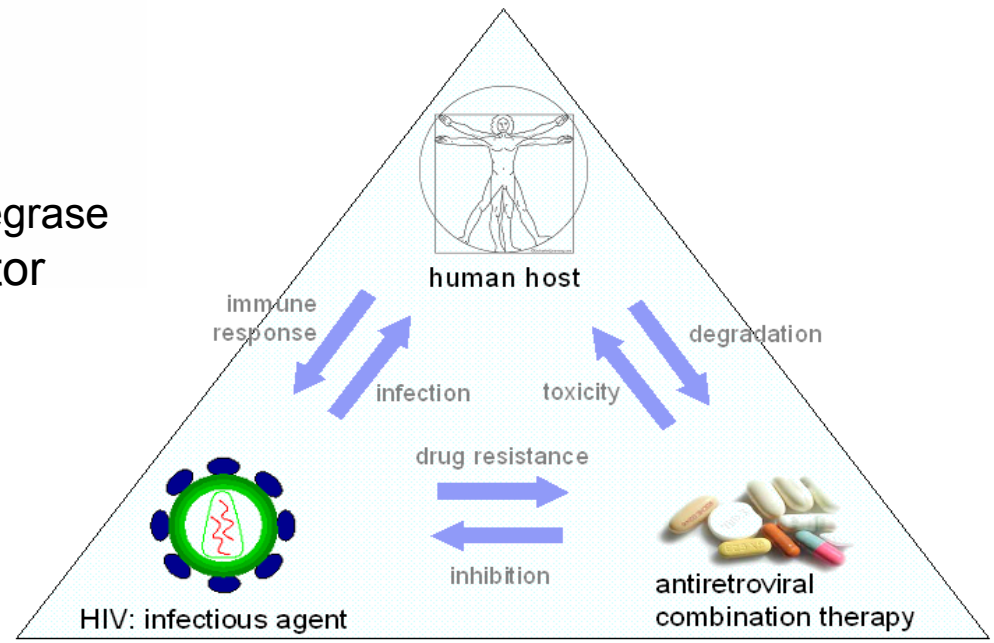
max planck institut informatik

HIV drug therapy

- 8 nucleoside reverse transcriptase inhibitors (NRTI)
- 3 non-NNRTI

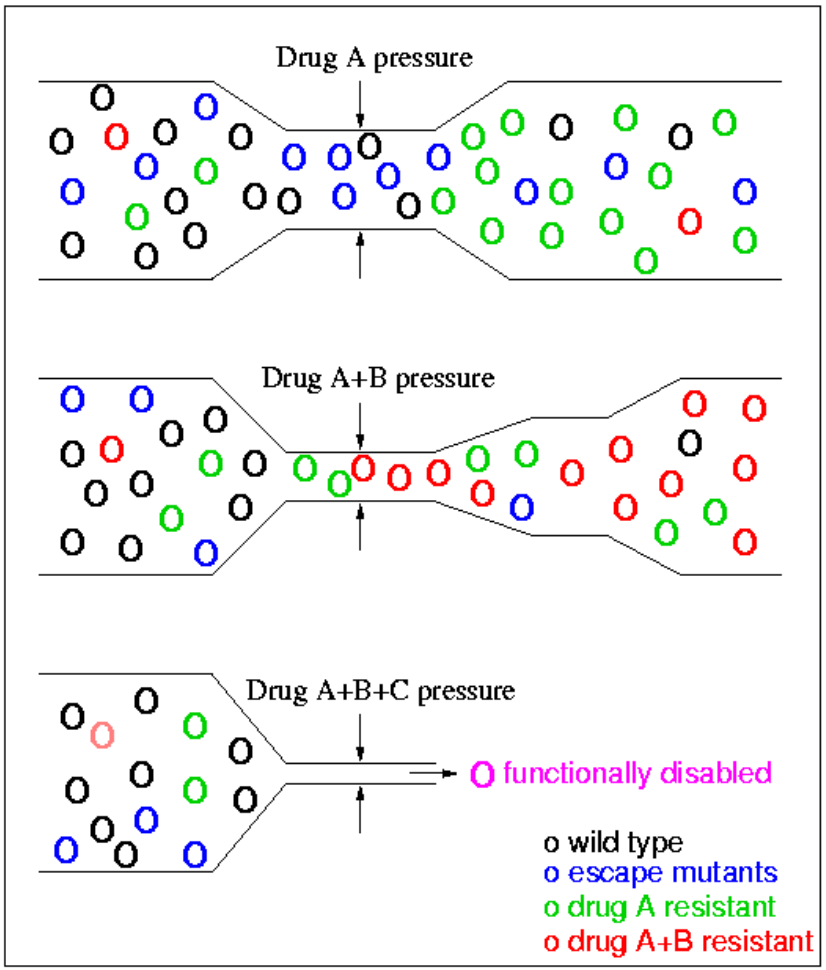


- 8 protease inhibitors (PI)
- 1integrase inhibitor
- 1 fusion inhibitor
- 1 coreceptor antagonist

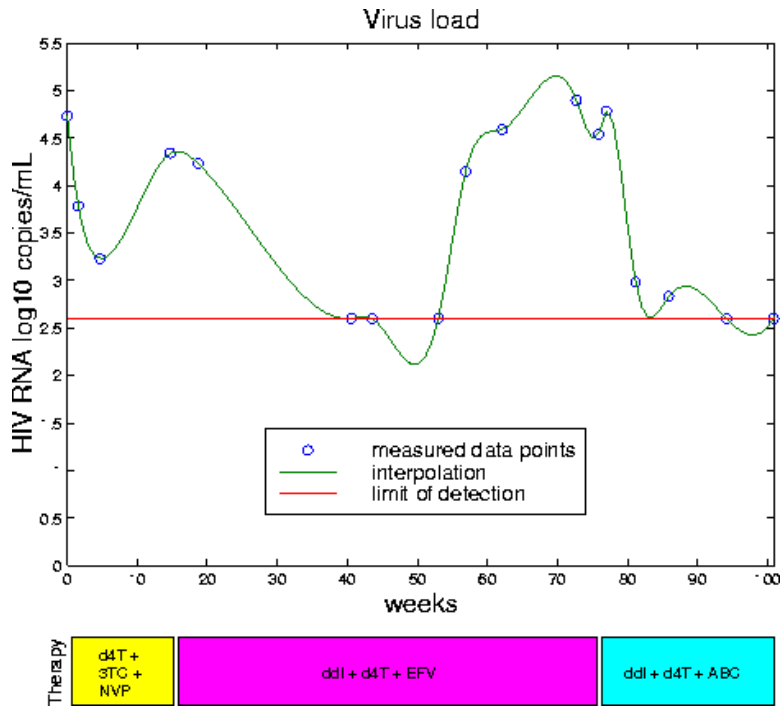


max planck institut informatik

Highly active antiretroviral therapy (HAART)



- ≥ 3 drugs, ≥ 2 classes



max planck institut informatik

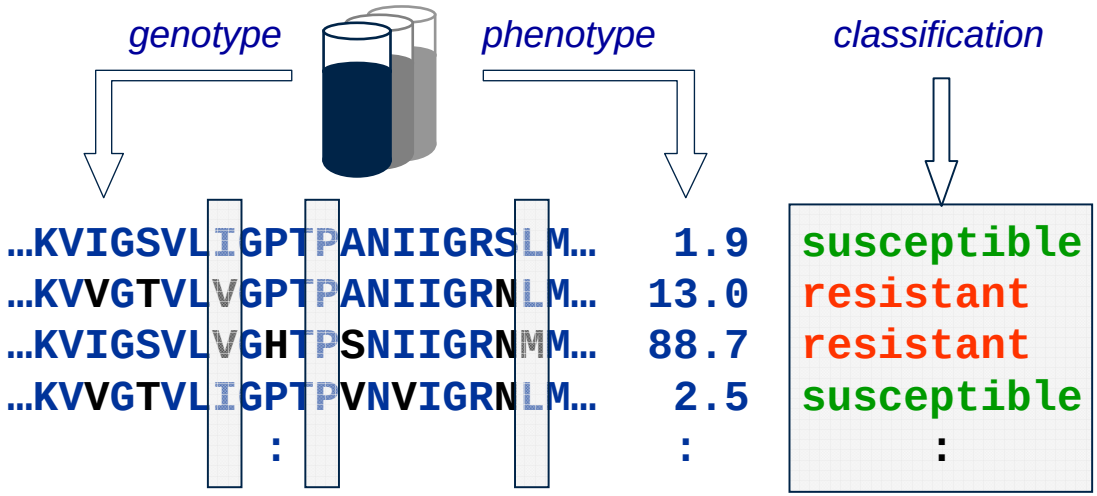
Highly active antiretroviral therapy (HAART)

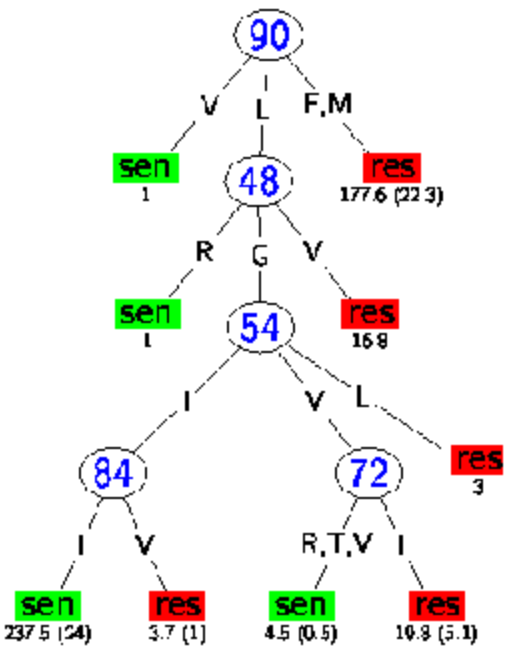
- resistance tables of International AIDS Society
- based on clinical experience
- biaised, cannot express mutation interdependence

max planck institut informatik

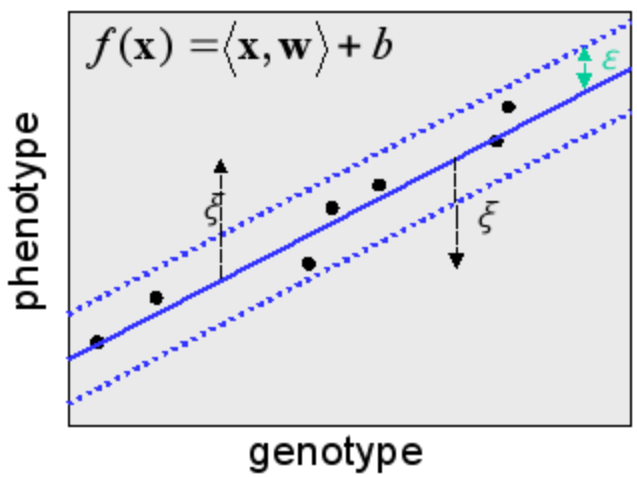
geno2pheno[resistance]

- based on phenotypic data of about 1000 viral variants





- uses different measures of information content
  - mutual information
  - distance from the decision boundary of Support Vector Machine (SVM) *(Beerenwinkel, 2002)*



max planck institut
matik

geno2pheno - version 3.0 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

geno2pheno [resistance]

geno2pheno [coreceptor]

Help

Statistics

References

Contact

Niko Beerenwinkel, Joach

On submitting below an HIV-
HXB2, a list of mutations and

NEW Features:

1. Identifier (optional):

2. Cutoffs:

N

NN

3. Pol-gene (PR and RT) nucleotide sequence:

4. Sequence ambiguities:

5. Output:

6. Action:

You will ma

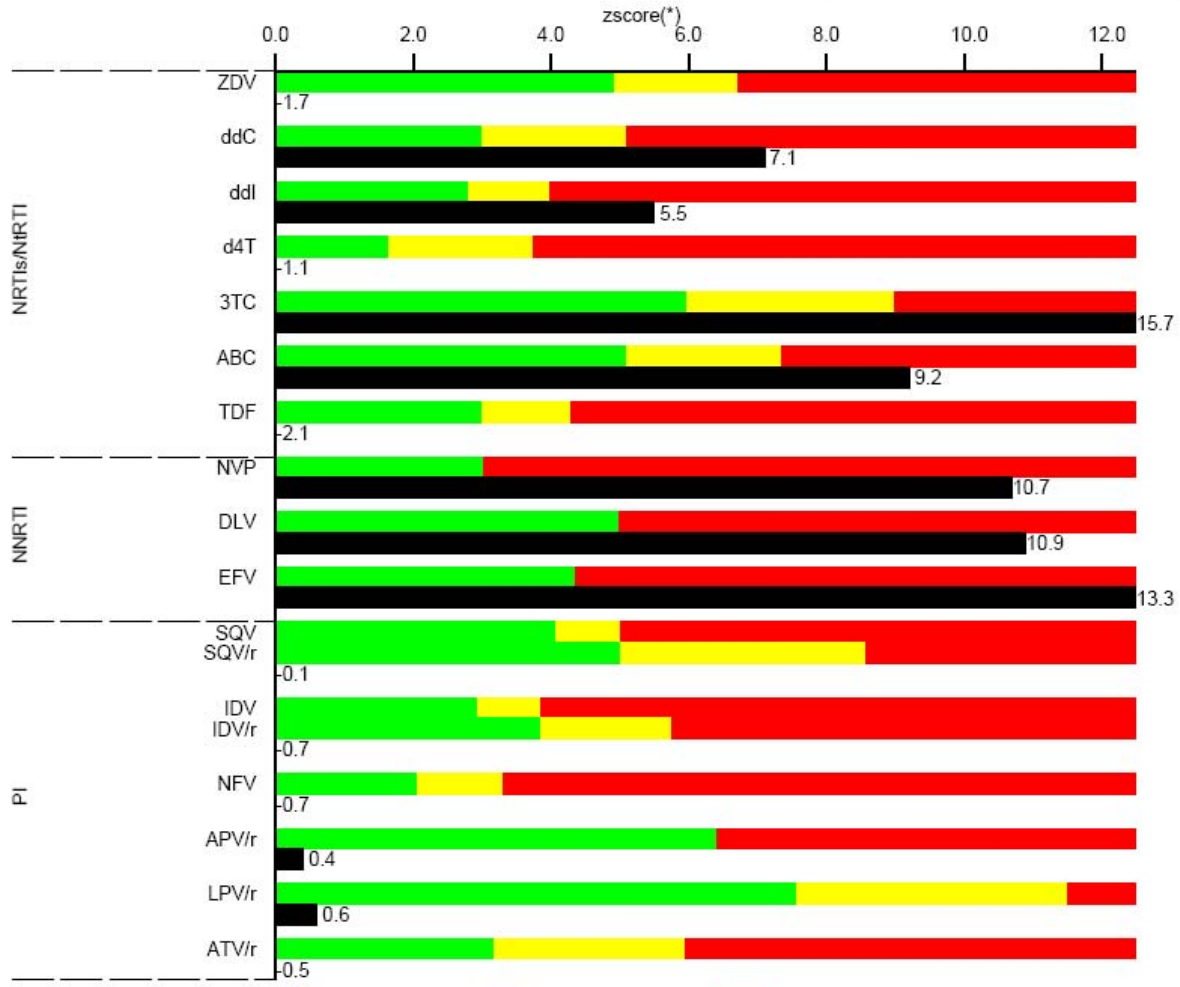
rg

geno2pheno®

phenotype prediction from genotype

IV. Interpretation

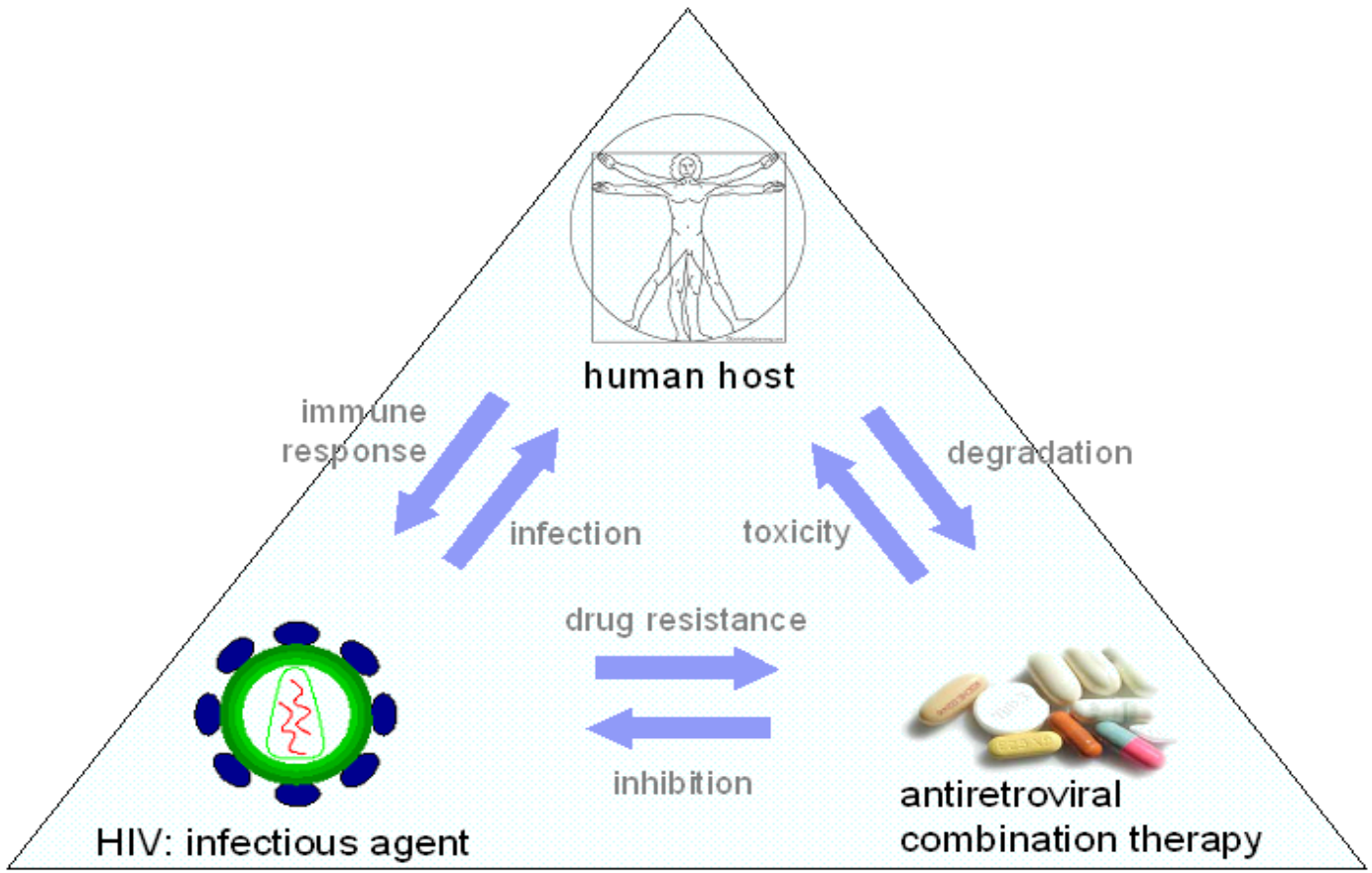
Patient:	Birth date:	Sampling date:
Current therapy:		Viral load:

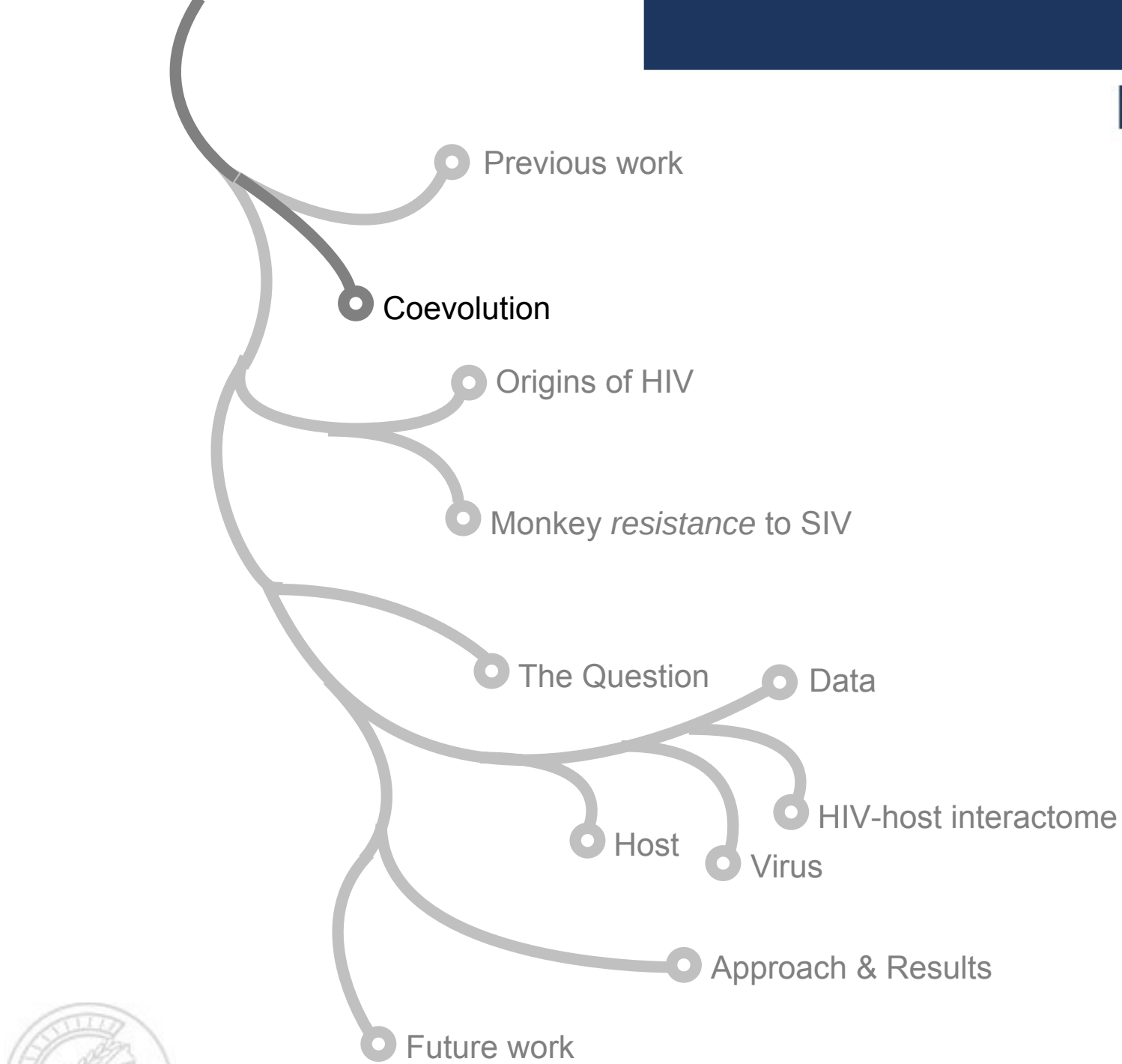


(*)number of standard deviations above mean of drug naive patients. Negative z-scores may indicate hypersusceptibility.

max planck institut informatik

- modeling viral evolution (Beerenwinkel, 2003)
- mutagenetic trees (Beerenwinkel, 2005)
- therapy optimization (Altmann, 2007)
- debiasing clinical databases (Altmann, 2007)
- prediction of coreceptor usage (Sander, 2007)

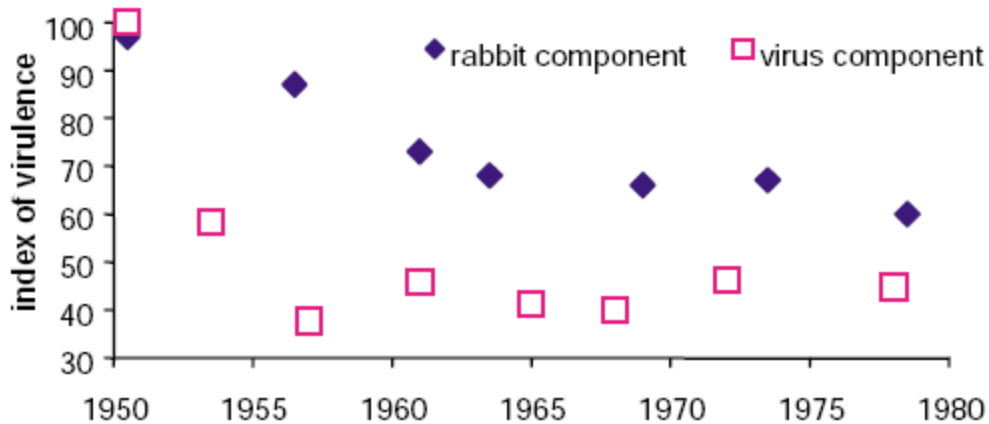



Previous work
Coevolution
Origins of HIV
Monkey *resistance* to SIV
The Question
Data
HIV-host interactome
Host
Virus
Approach & Results
Future work

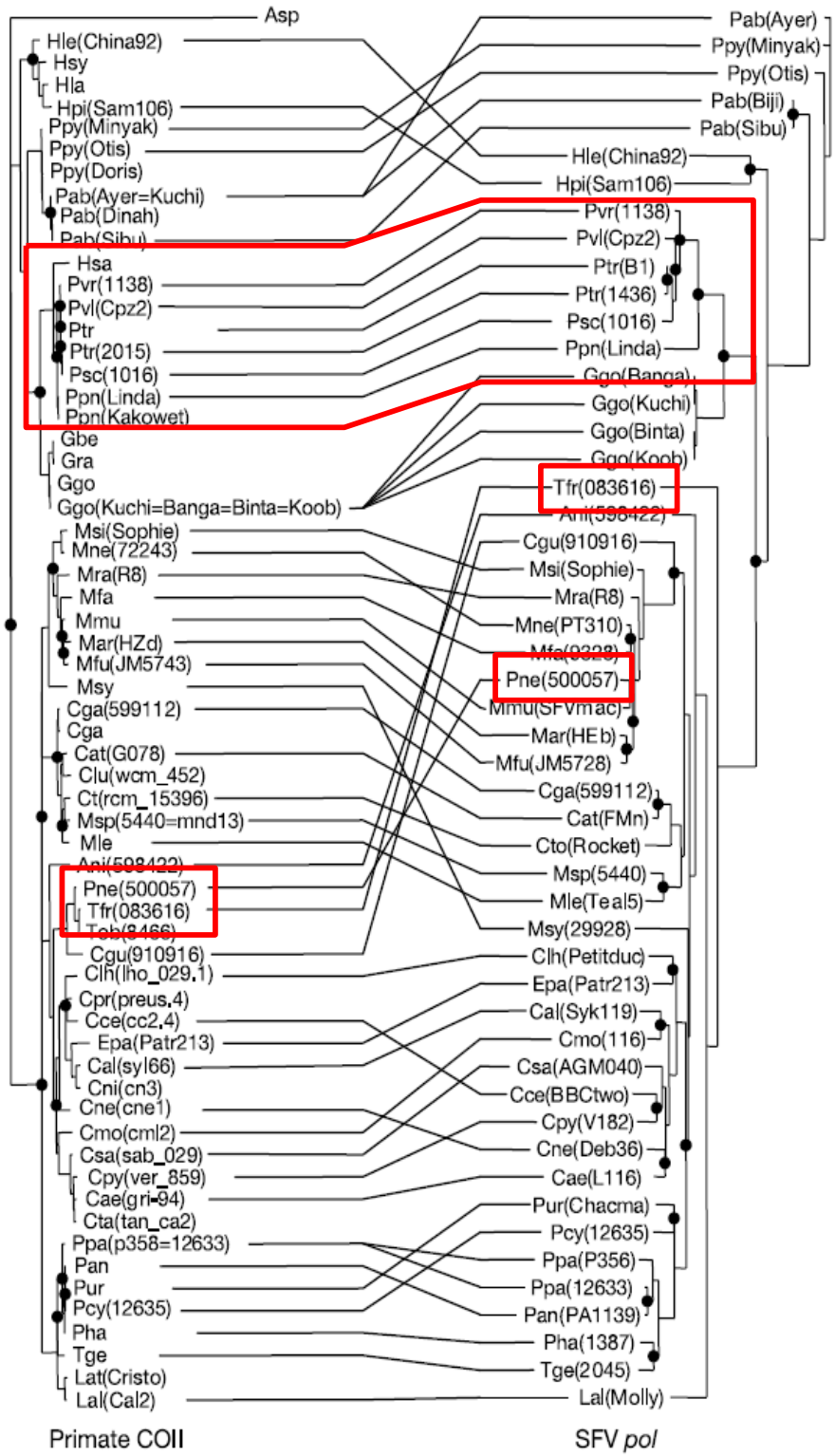
Coevolution is
process of reciprocal adaptive genetic change in two or more species.

Coevolution requires
- genetic variation
- reciprocal effects on the fitness of the 2 populations
- dependence of the outcome interaction on the genotypes involved

Coevolution can be studied in terms of
- paired phenotypic traits (resistance – infectivity)
- interacting host and pathogen molecules
- genes, nucleotide sequences



Woolhouse et al. 2002



max planck institut informatik

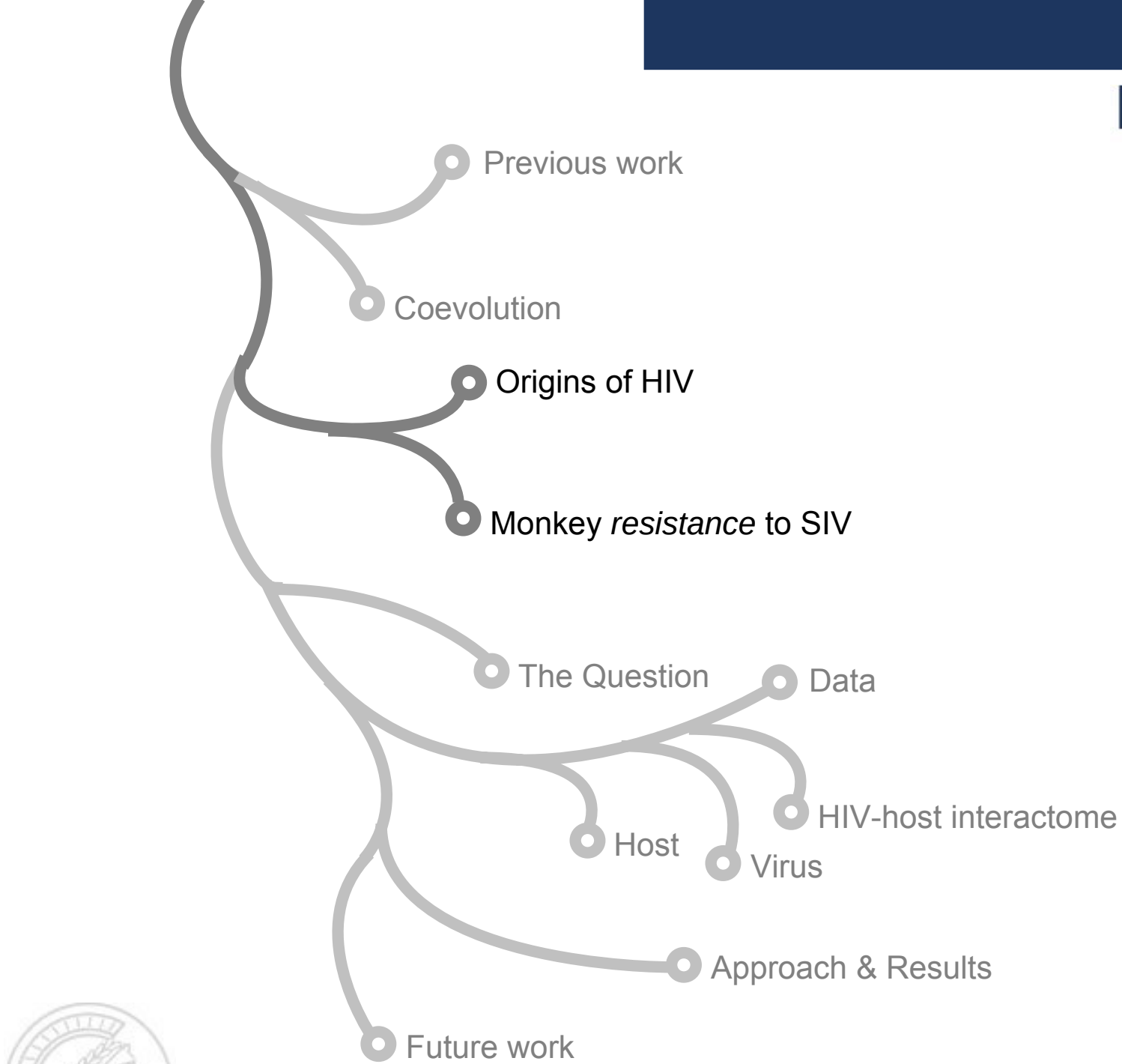
Simian foamy viruses (SFV)

ubiquitous, non-pathogenic retroviruses that infect all primates.

Phylogenetical comparison of *SFV polymerase* and *mitochondrial cytochrome oxidase subunit II* from African and Asian monkeys and apes.

- congruent trees
- extremely low rate of SFV evolution: 10^{-8} substitutions per site per year (10^{-3} for HIV)
- cospeciation during 30 million years

Switzer et al. 2005

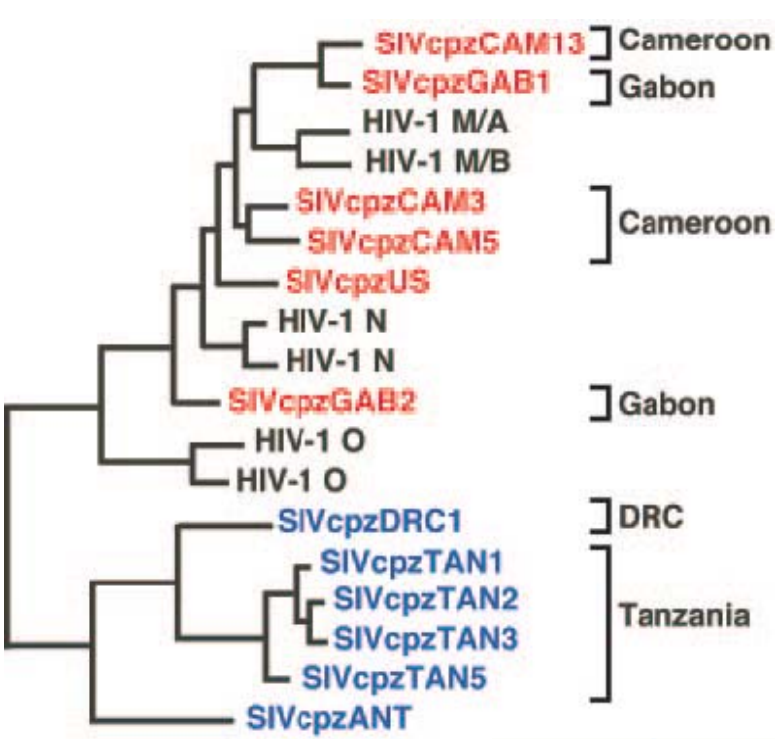

Previous work
Coevolution
Origins of HIV
Monkey *resistance* to SIV
The Question
Data
HIV-host interactome
Host
Virus
Approach & Results
Future work

Simian Immunodeficiency Viruses (SIVs)

primate lentiviruses infecting more than 36 different nonhuman primate species in Africa.

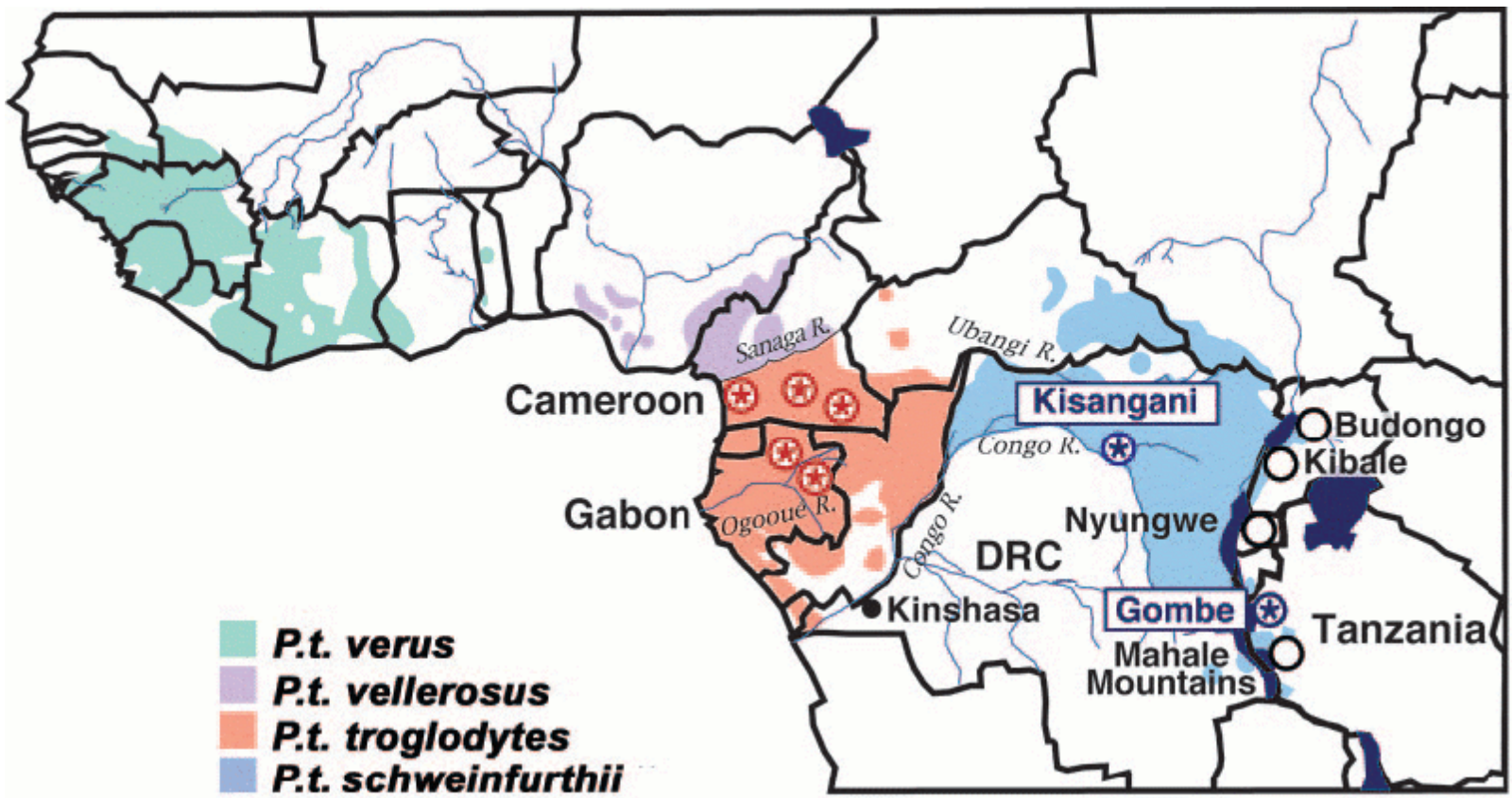
Cross-species transmission of *chimpanzee SIVcpz* and *sooty mangabey SIVsmm* generated HIV type 1 and 2 in human population.

Multiple simian-human transmissions are estimated for the beginning of the 20th century.

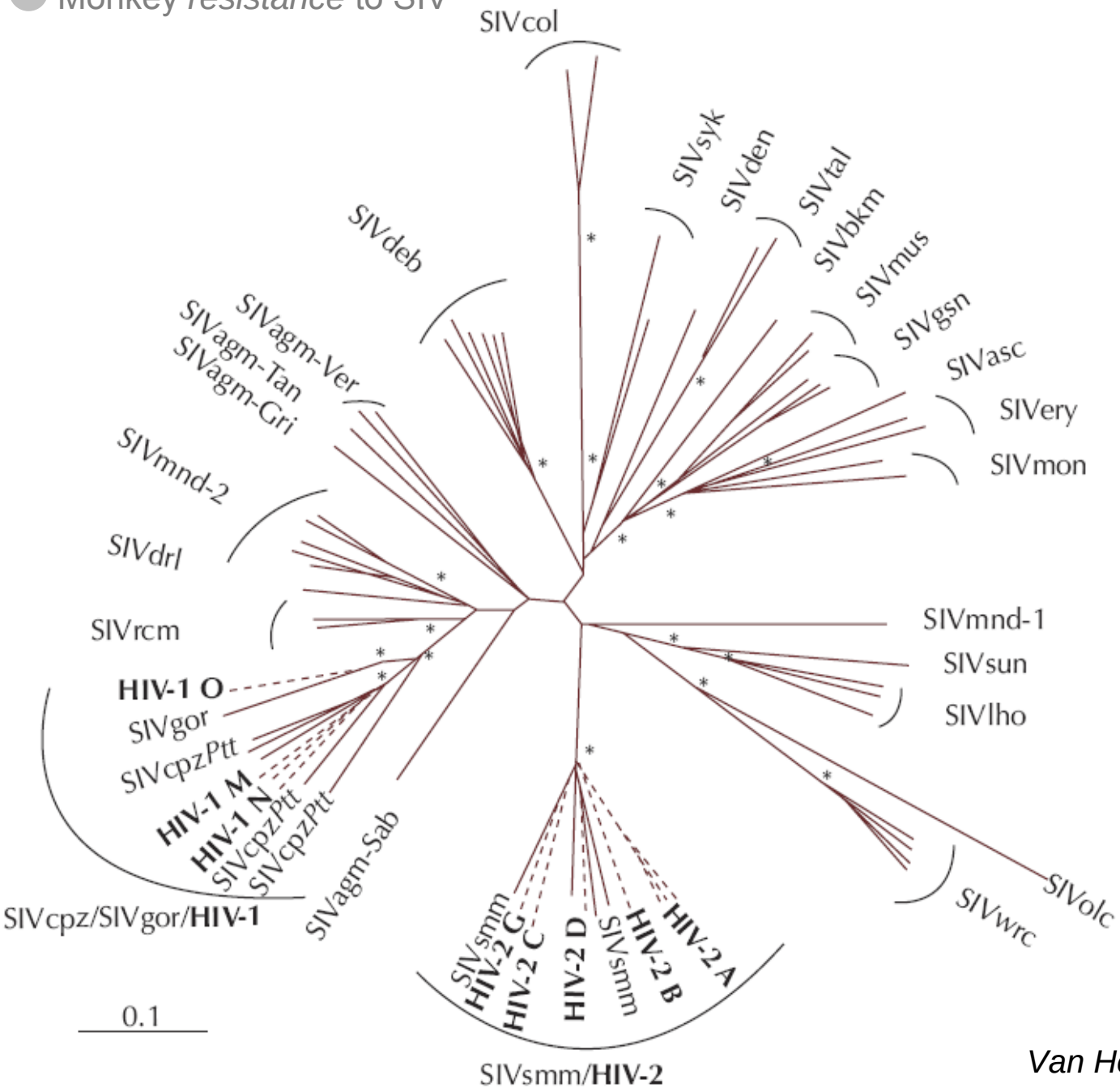


max planck institut informatik

Sharp et al. 2005



Monkey *resistance* to SIV



Van Heuverswyn, 2007

SIV-host coevolution?

- geographical distribution of infections
- phylogenetical evidence for cross-species simian-simian transmission events
- closely related monkey species harbor closely related SIVs
- all naturally infected monkey species do not develop AIDS

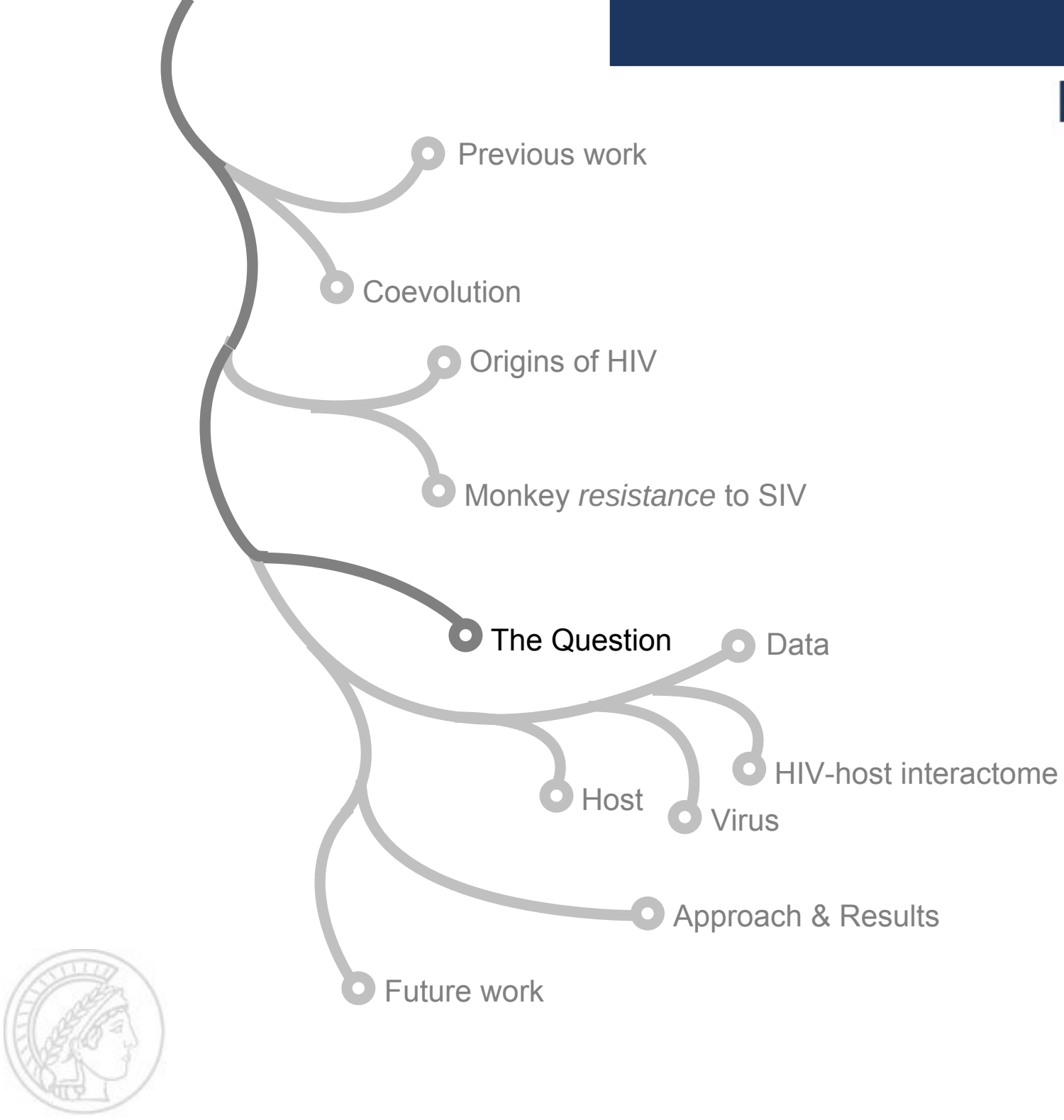
max planck institut informatik

Monkey *resistance* to SIV

All naturally infected monkey species do not develop AIDS.

Differences in immune response:

- progressive loss of CD4+ T lymphocytes, chronic immune activation, gradual destruction of immune function in humans
- SIVs do not cause immunodeficiency in their natural hosts
- not associated with CD4 cell decline or function loss, no degenerative changes in lymph node architecture
- SIV establishes a persistent infection and high replication rate
- Long Term Non-Progressors – low replication; monkeys – high viremia but no immune activation


Previous work
Coevolution
Origins of HIV
Monkey *resistance* to SIV
The Question
Data
HIV-host interactome
Host
Virus
Approach & Results
Future work

max planck institut informatik

SIV/HIV – host coevolution

compare virus-host interacting proteins in SIV-primate and HIV-human infection in the search for:

- immune system footprints on the viral genome, adaptive changes that followed the interspecies transmission
- host interacting factors evolution
- potential reasons of HIV pathogenicity in humans

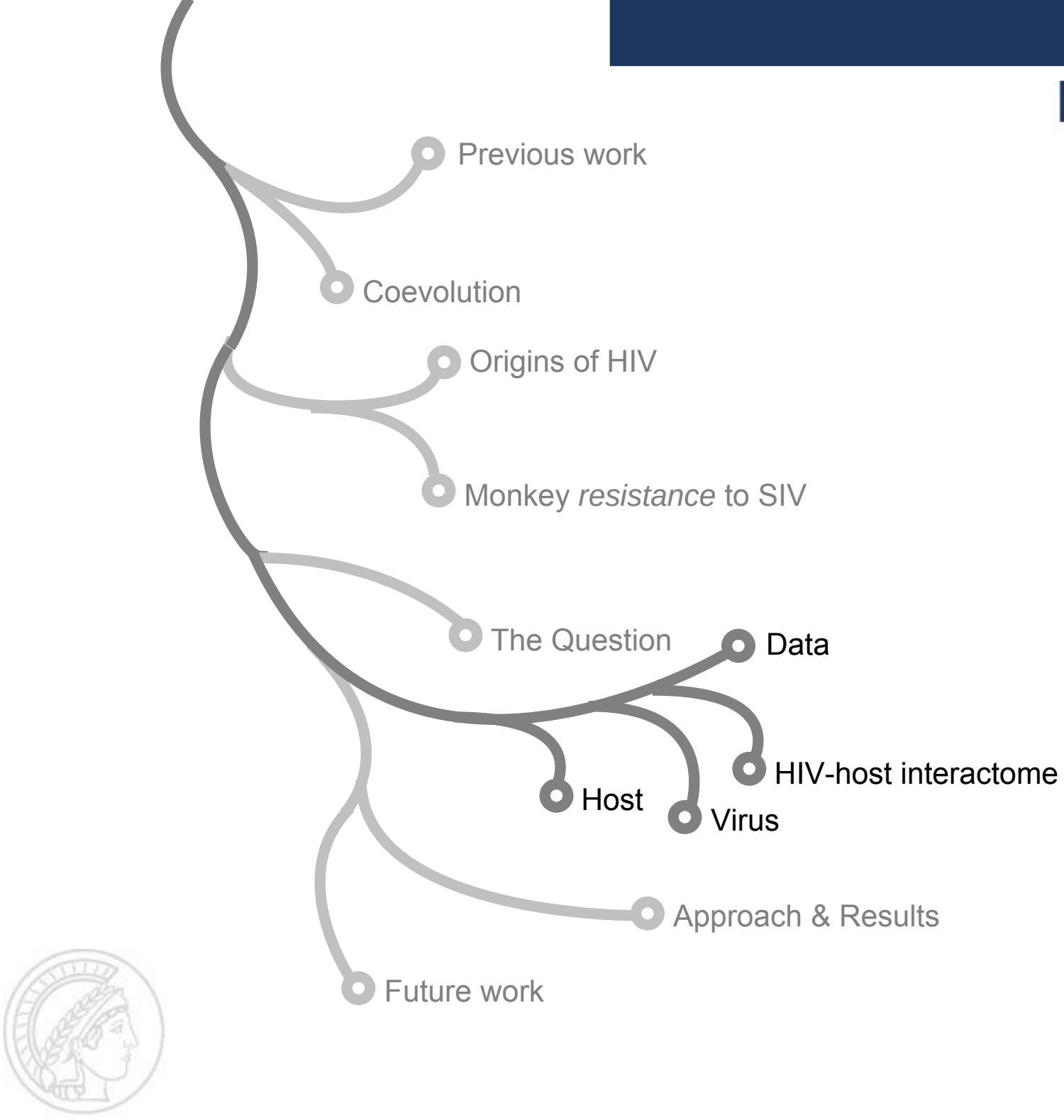
basing on:

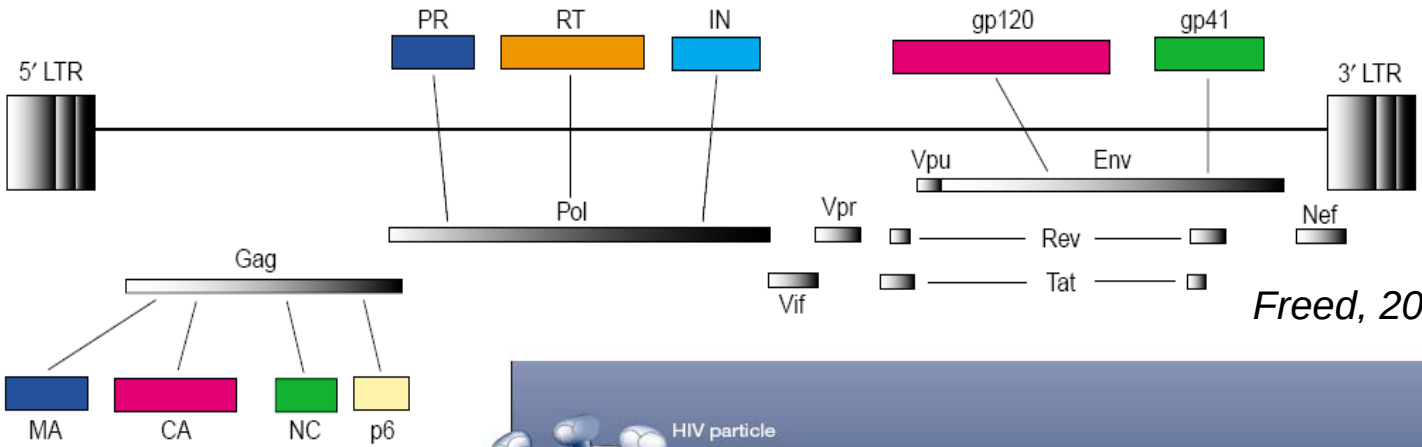
- big similarity of the primate genomes (98% for chimp-human)

analysis of:

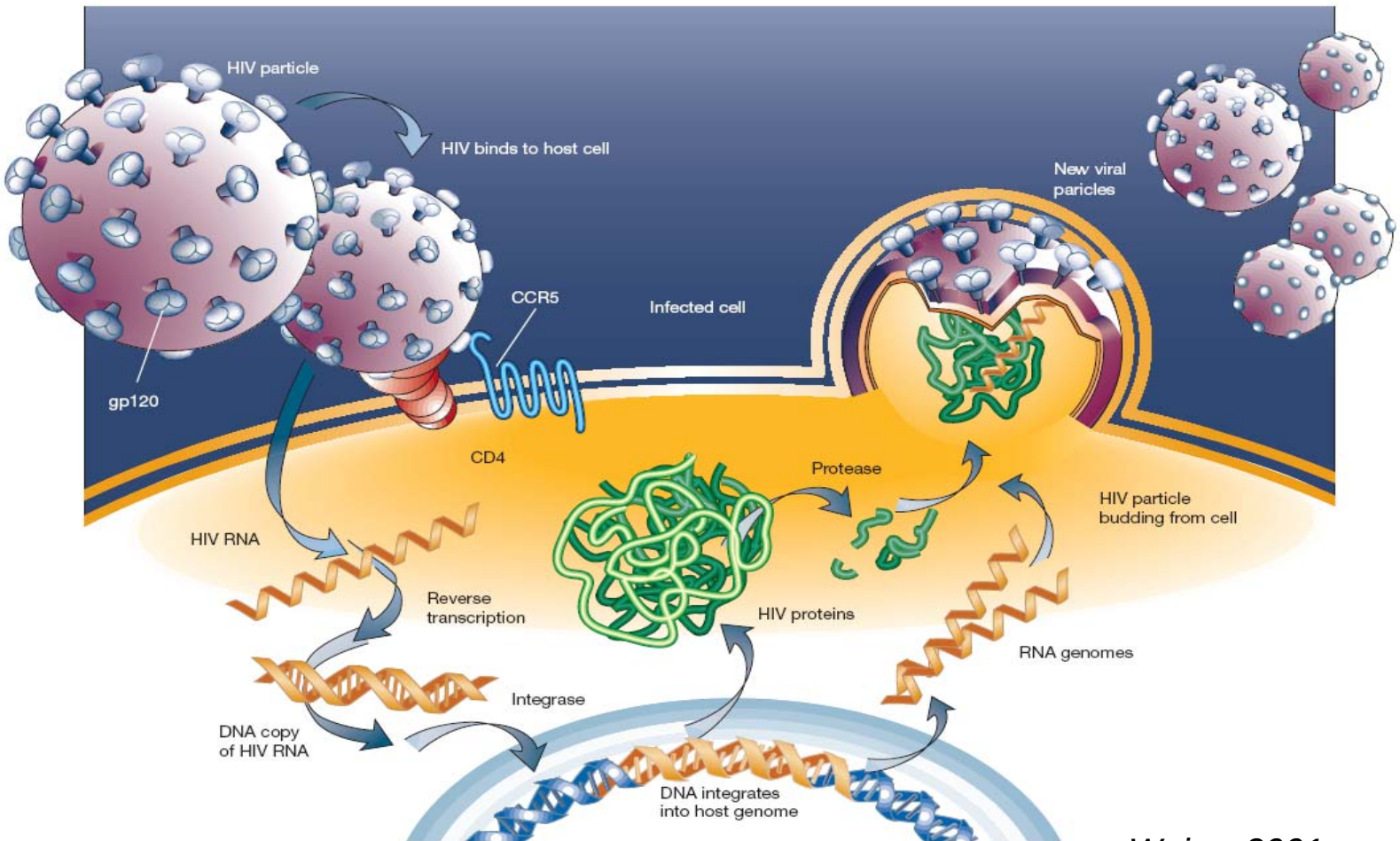
- molecular coadaptation
- evolution of genes implicated in host-pathogen interaction

using coevolution as a *null hypothesis*


Previous work
Coevolution
Origins of HIV
Monkey *resistance* to SIV
The Question
Data
HIV-host interactome
Host
Virus
Approach & Results
Future work



Freed, 2004



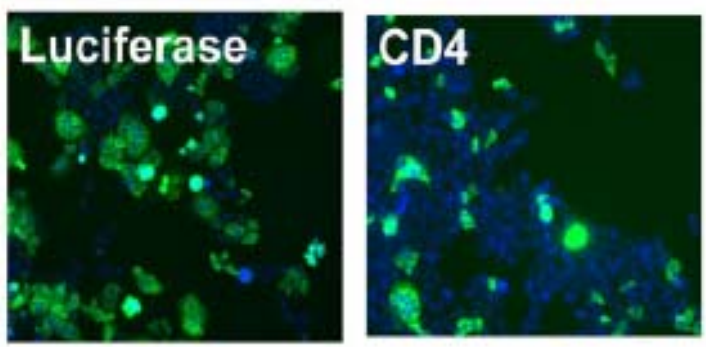
Weiss, 2001

max planck institut informatik

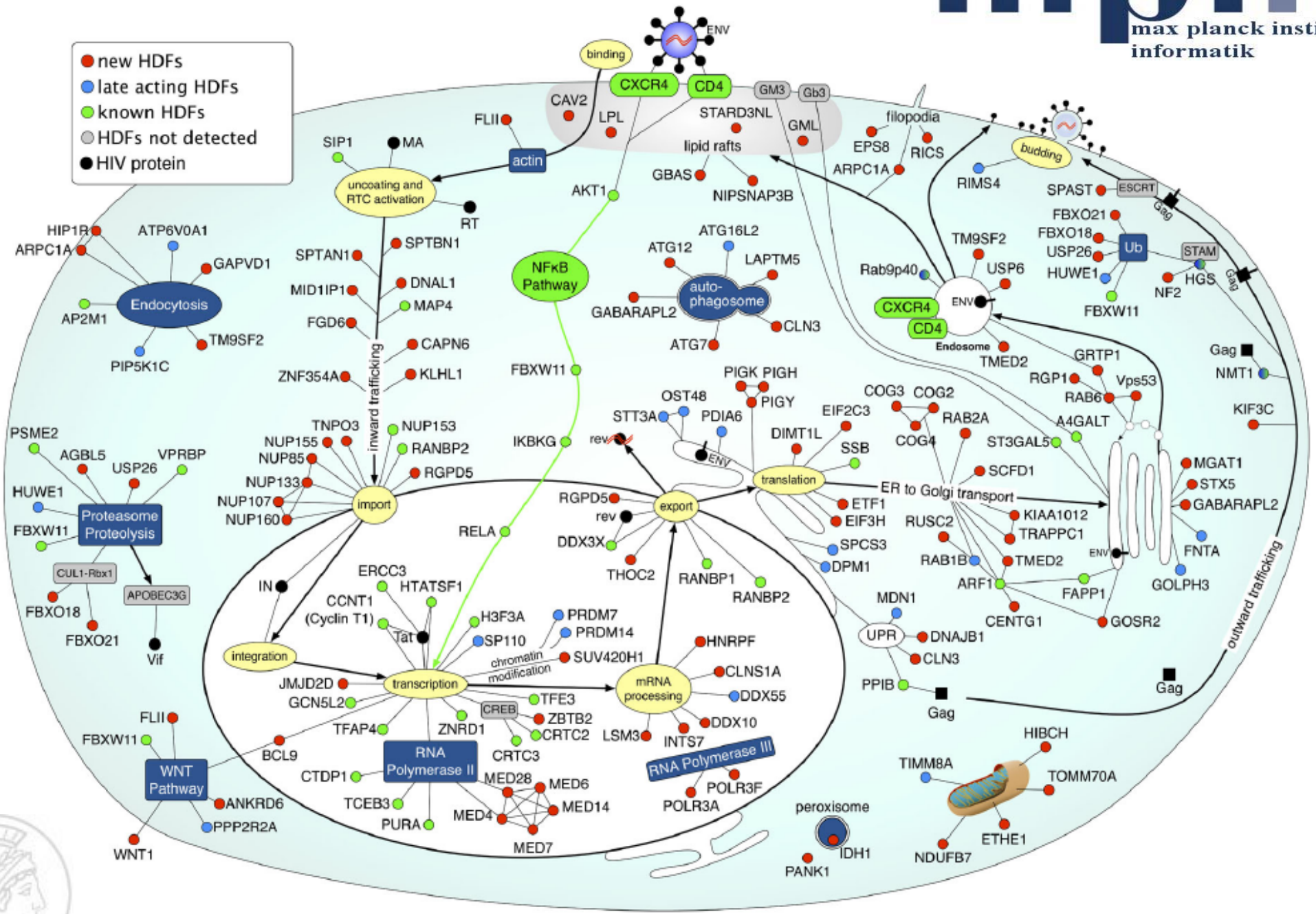
"Identification of Host Proteins Required for HIV Infection Through a Functional Genomic Screening" Brass AL et al. Science 2008

- large scale siRNA screen (over 21 000 siRNA pools)
- over 250 HIV-dependency factors (HDFs) identified (13% previously known)
- subcellular localization, biological process and function
- extensive exploitation of host cell functions by HIV
- potential therapeutic targets

HIV interactome – dependencies between detected HDFs and HIV proteins established using NCBI HIV interaction database, protein-ptorein interaction and homology information.



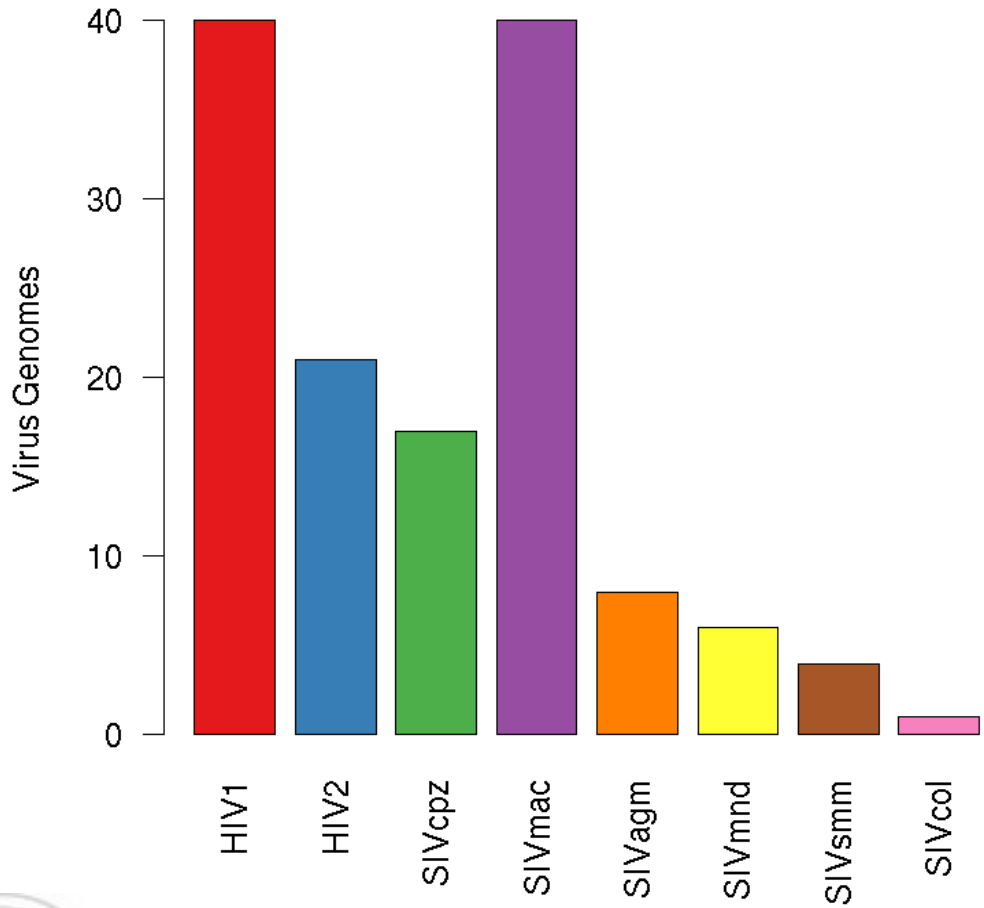
max planck institut informatik



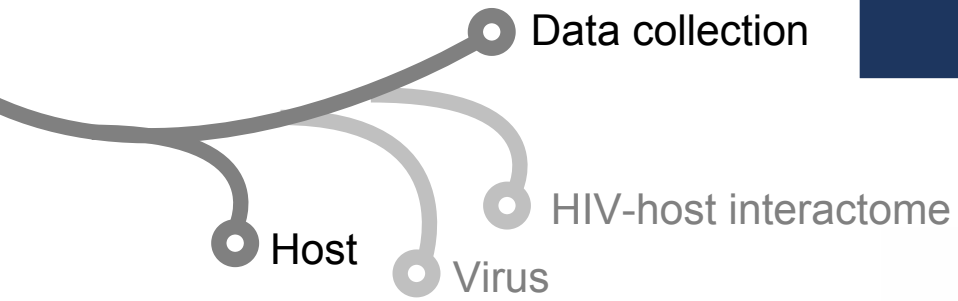
HIV-1, Human Protein Interaction Database

National Institute of Allergy & Infectious Diseases

- comprehensive database of the described interactions between HIV-1 and cellular proteins
- interactions reported in the scientific literature
- concise but detailed summary of all known interactions containing RefSeq, NCBI accession numbers, description of the protein-protein interaction, pubmed reference



- 109 complete SIV genomes in GenBank
- 22 HIV-2 genomes
- 1232 HIV-1 genomes
- filter:
  - species of interest
  - therapy-naive HIV strains
  - time and geography representative sample

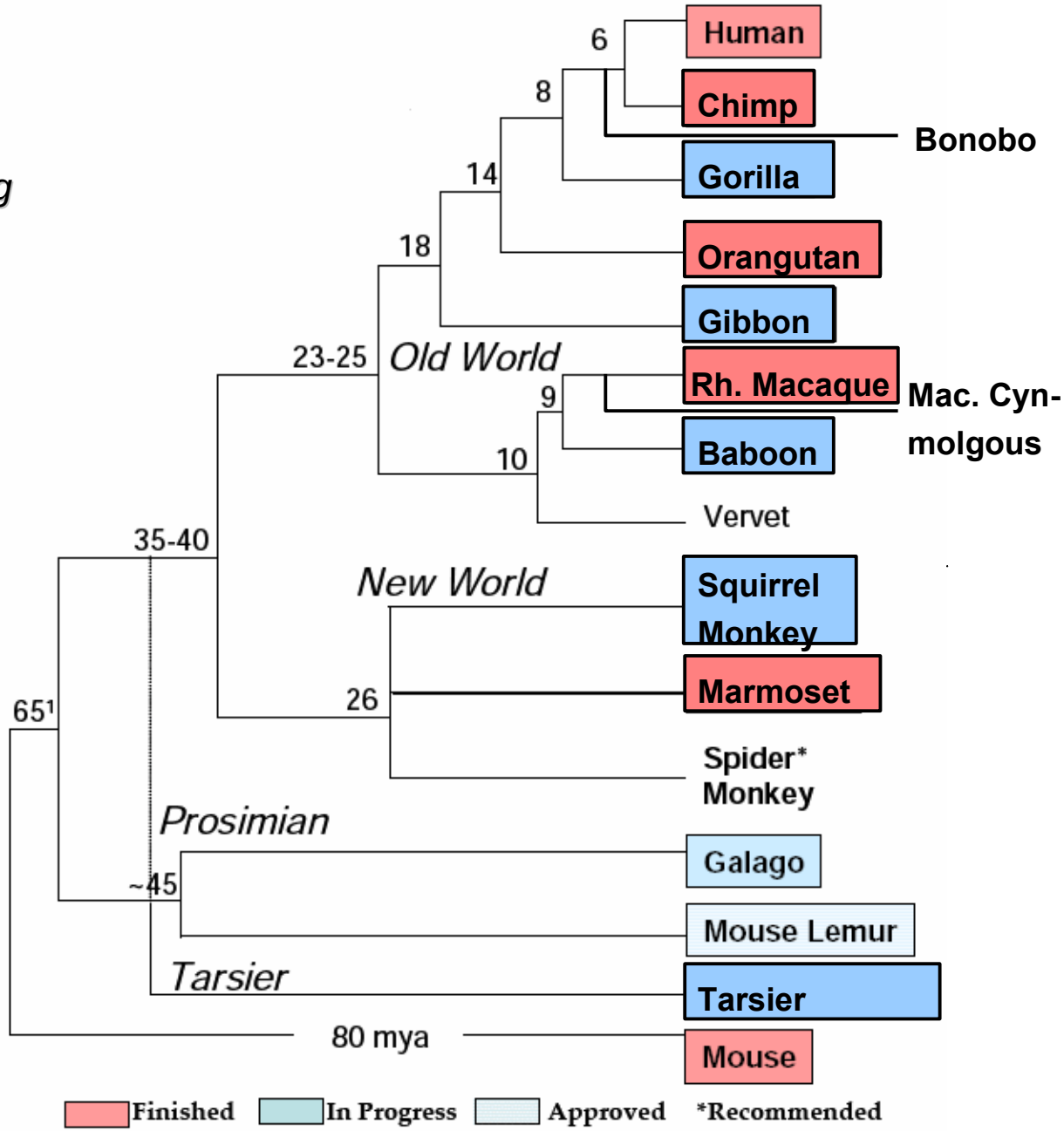


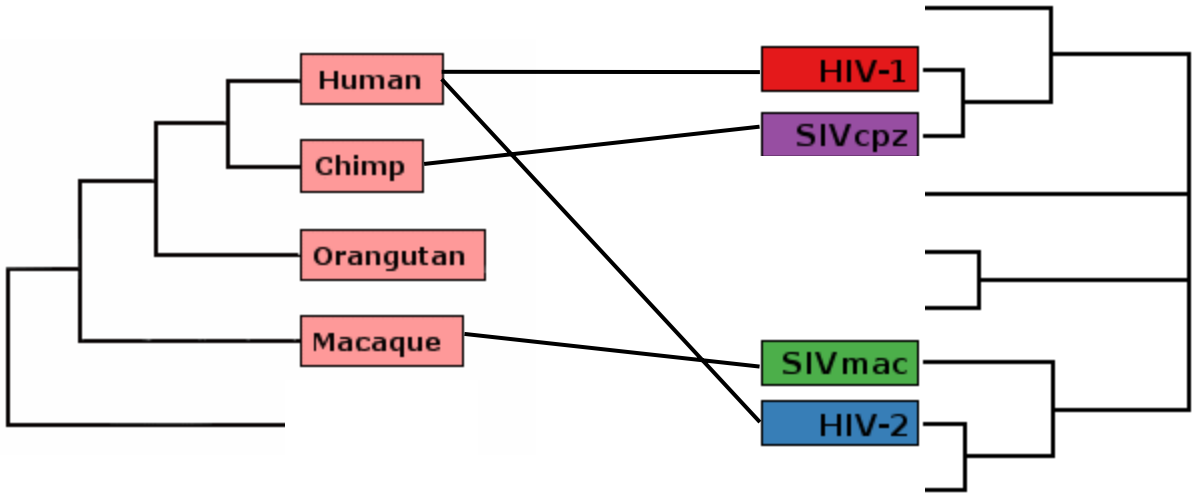
max planck institut informatik

Non-human primate genome sequencing project launched in 2006

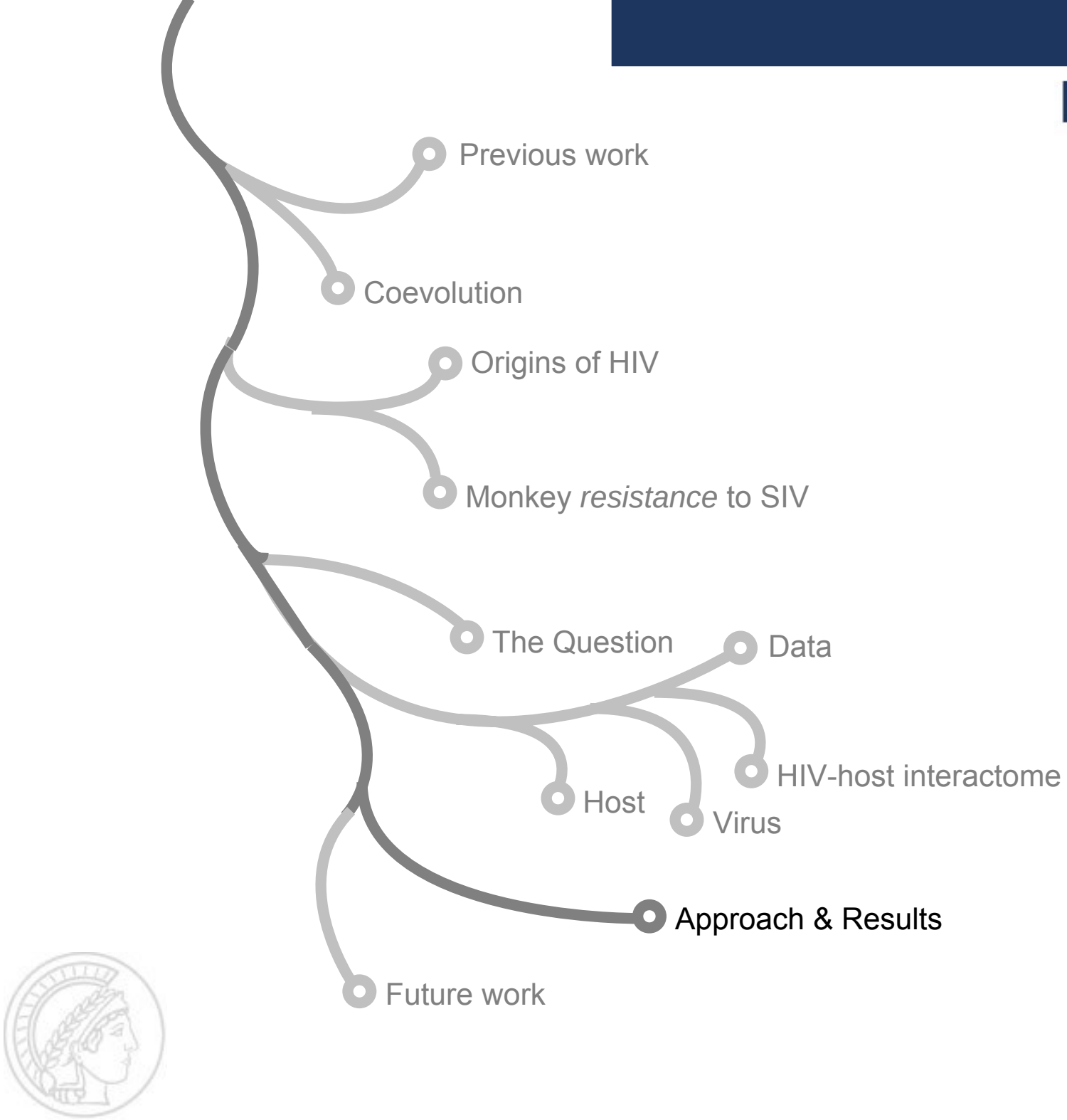
- chimp 2003
- macaque 2006
- marmoset 2007
- orangutan 2008

9 other primate sequencing projects mentioned by ISC





- 3 sequenced species known to be infected by the virus
- 1439 host genes (NIAID)
- corresponding virus sequences (118 genomes of 4 species)
- protein-protein interactions


Previous work
Coevolution
Origins of HIV
Monkey *resistance* to SIV
The Question
Data
HIV-host interactome
Host
Virus
Approach & Results
Future work

Comparison of evolution of interacting host and viral proteins:

- conservation of primate genomes
- variability among and within viral genomes

"*The degree of genomic diversity that HIV generates in a single infected individual can be greater than the worldwide diversity of influenza A virus during an epidemic.*" (Korber et al. 2001)

- 2 different measures of divergence

max planck institut informatik

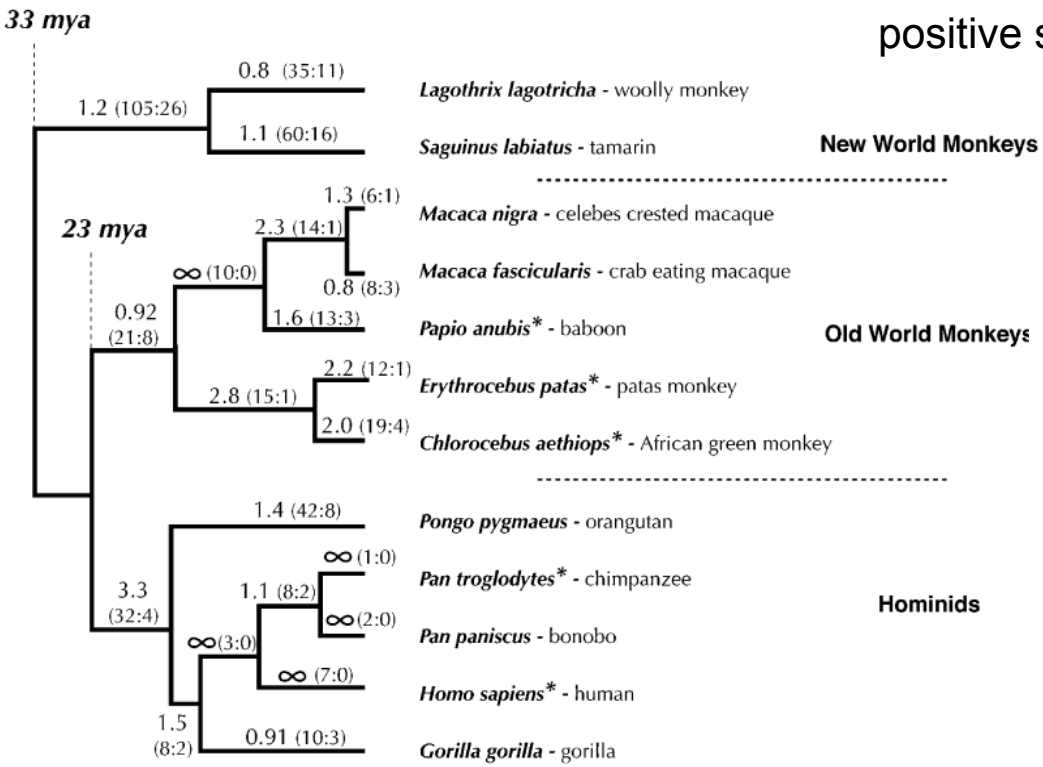
Open access, freely available online PLoS BIOLOGY

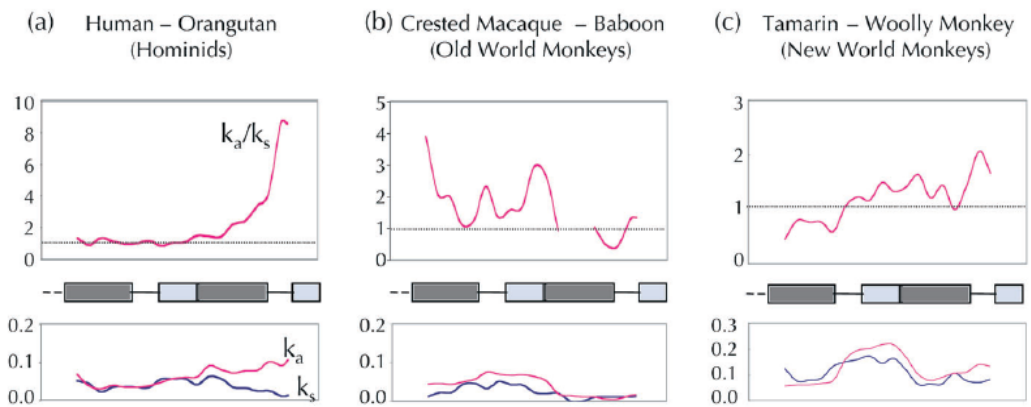
Ancient Adaptive Evolution of the Primate Antiviral DNA-Editing Enzyme APOBEC3G

Sara L. Sawyer[1], Michael Emerman[1,2], Harmit S. Malik[1*]

Host – APOBEC protein

- single-stranded DNA-editing enzyme causes hypermutation of HIV
- HIV-encoded virion infectivity factor (Vif) protein targets APOBEC3G for destruction
- species-specific interaction
- different members of APOBEC protein family subject to positive selection throughout history of primate evolution

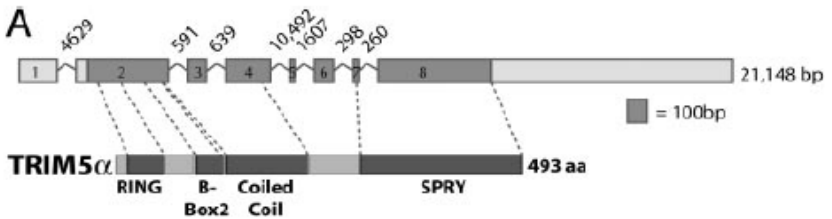


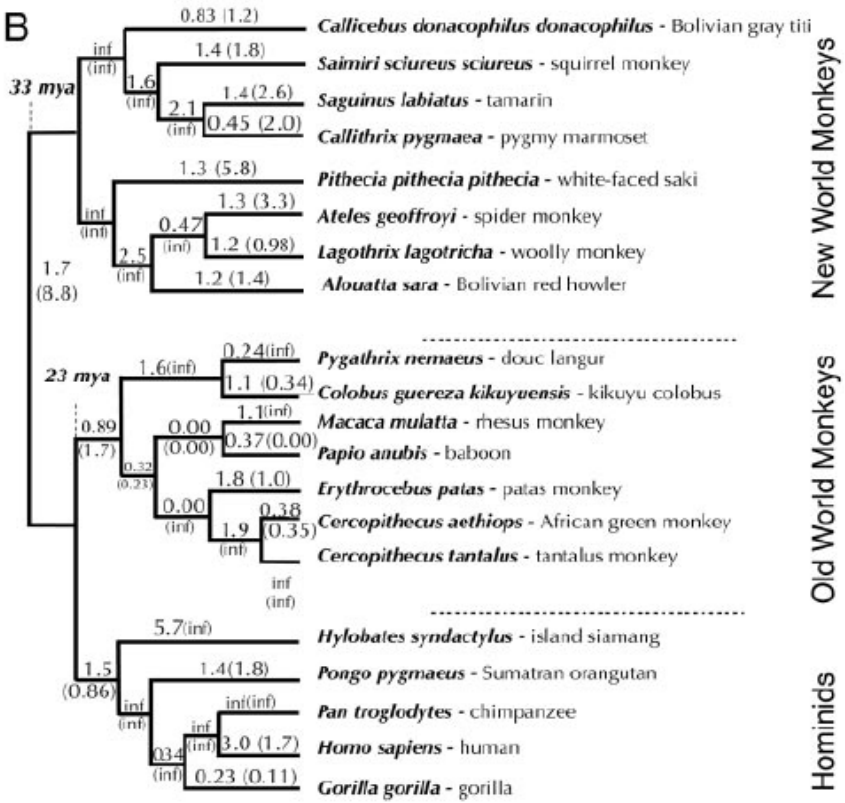


max planck institut informatik

Positive selection of primate *TRIM5α* identifies a critical species-specific retroviral restriction domain

Sara L. Sawyer*, Lily I. Wu†, Michael Emerman*†, and Harmit S. Malik*‡



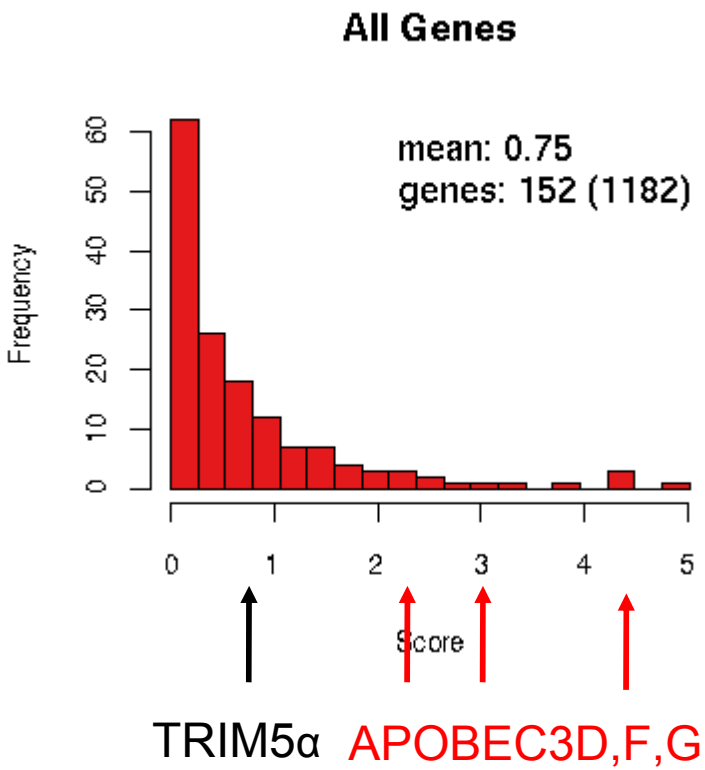


Host – TRIM5α protein

- restricts the virus in a species-specific manner
- interacts with viral capsid
- strong evidence of ancient positive selection in primate TRIM5α gene
- positive selection concentrated in the SPRY domain of the protein

max planck institut informatik

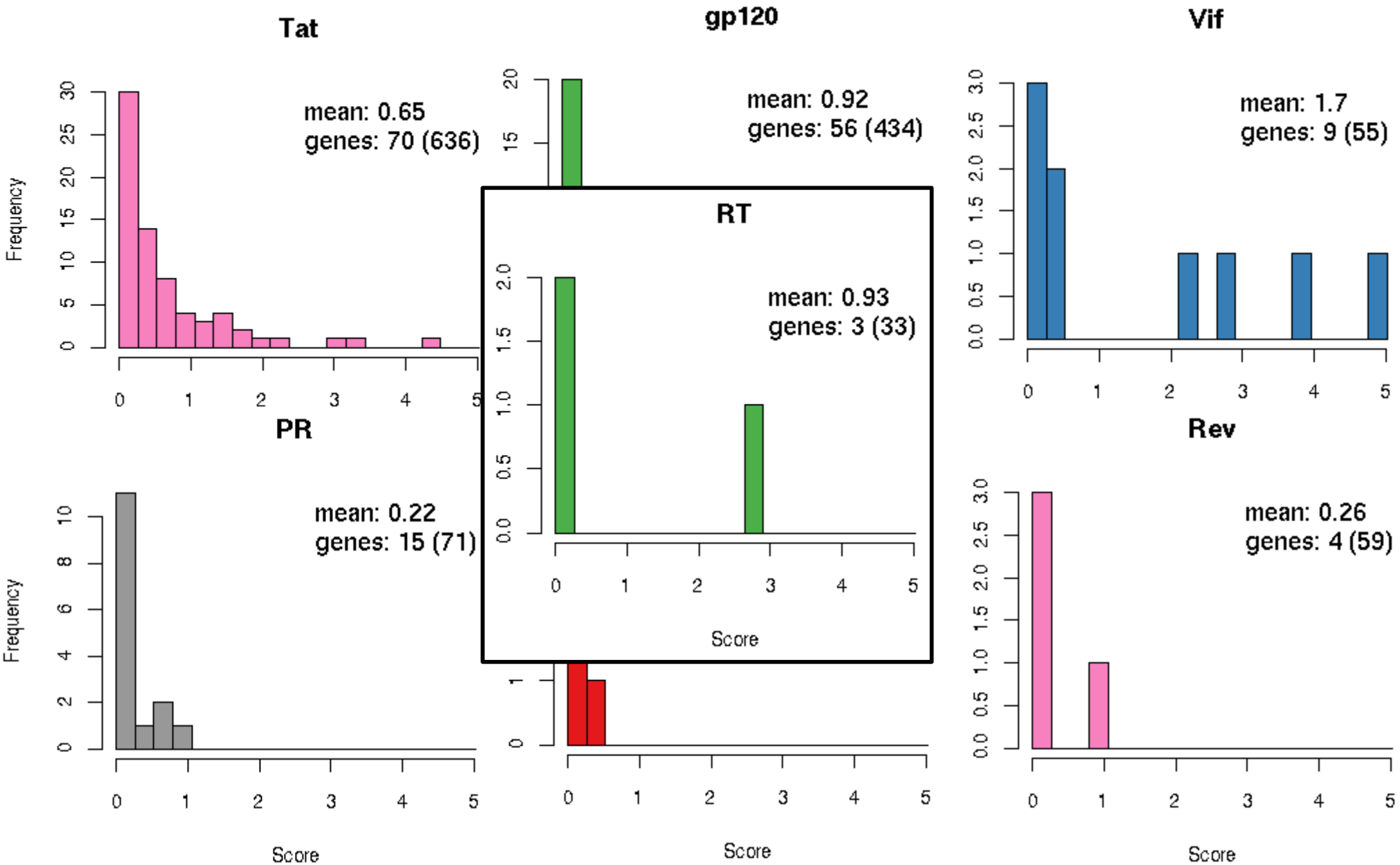
Host – search for sites under positive selection



- big genome conservation and long coding sequences
- inference of sites under positive selection (*PAML, Zhang et al. 2005*)
  - statistical measure of dN/dS (ω) variation among sites allowing a subset of sites to have ω > 1
  - maximum likelihood parameter fit
  - likelihood ratio test (LTR) to compare a null model that does not allow ω > 1 with an alternative model that does
- 2 LRTs
  - M1a (Nearly-Neutral) – M2a (Positive Selection)
  - M7 (beta) – M8 (beta&ω)
- score based on the number of positively selected sites
- 152 out of 1182 genes under positive selection

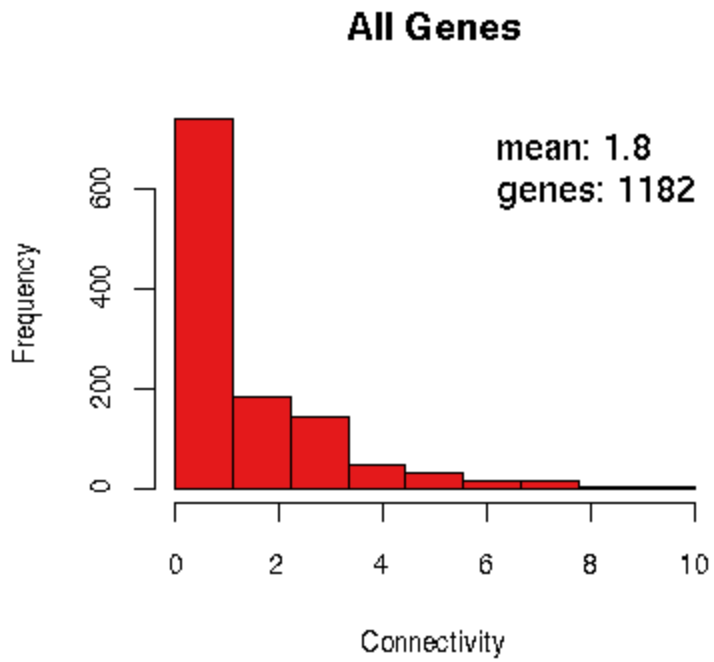
max planck institut informatik

Host – positive selection vs. interaction grouping



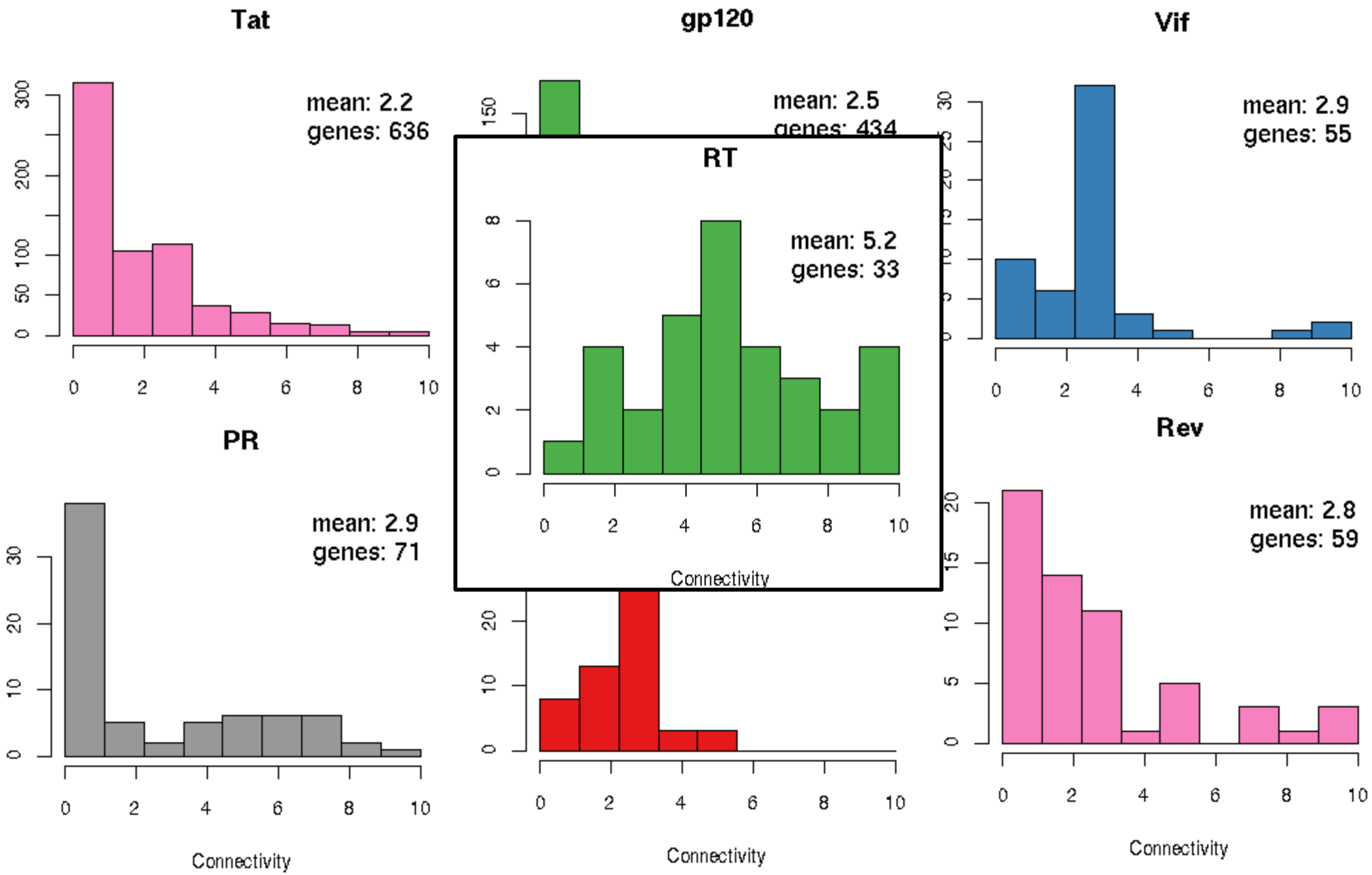
max planck institut informatik

Host – number of interactions



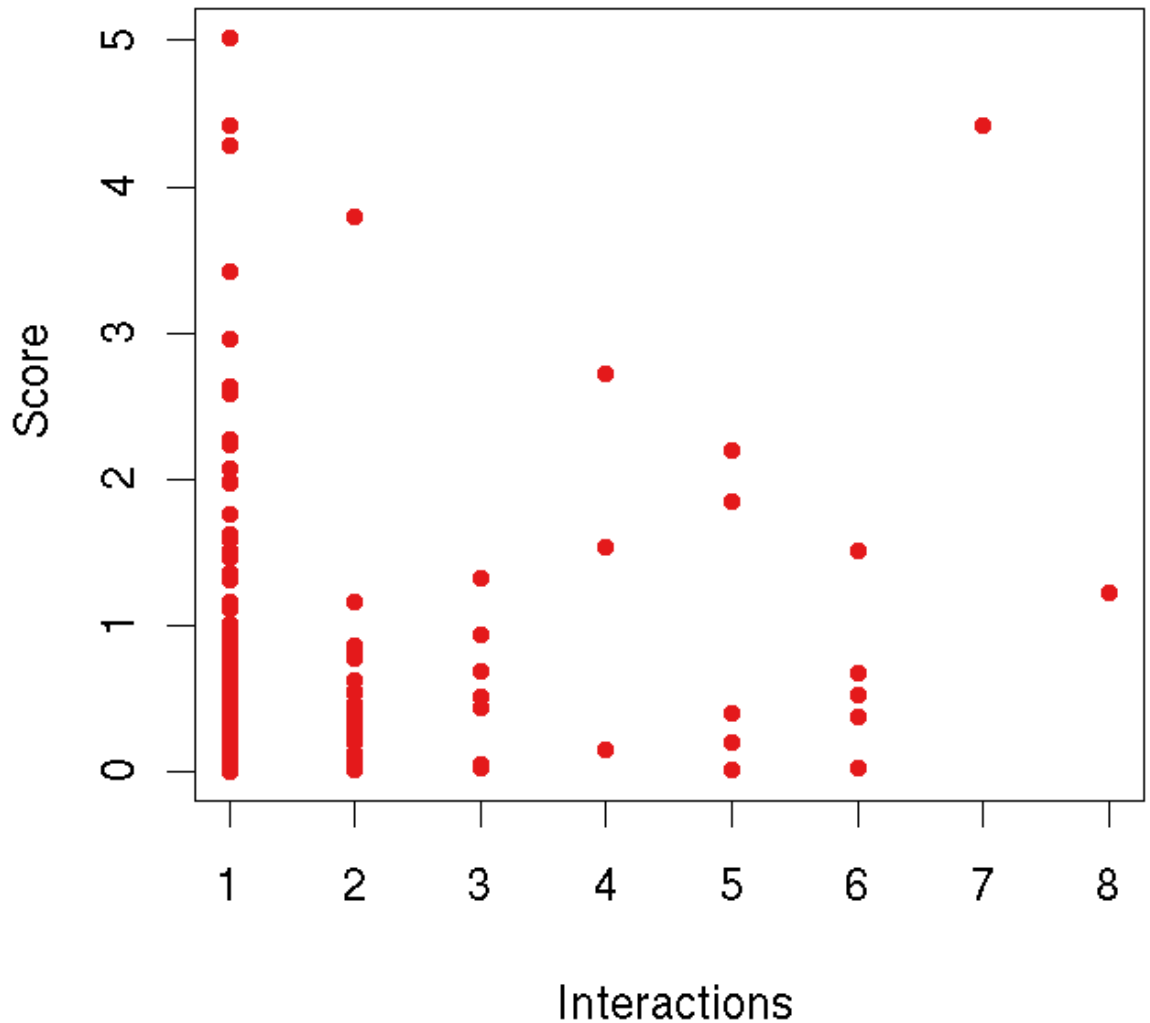
max planck institut informatik

Host – number of interactions



max planck institut informatik

Host – positive selection vs. number of interactions



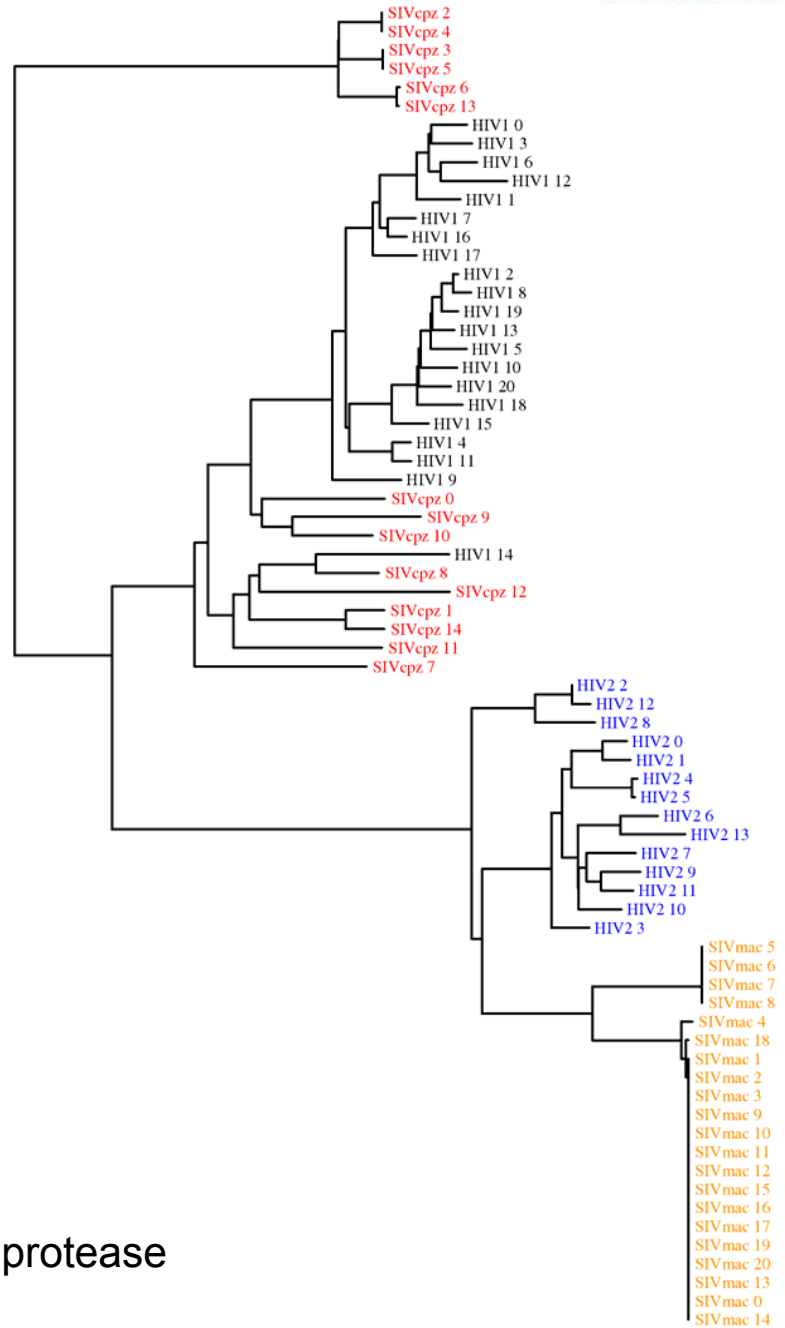
Host – positive selection in entire CDS

- estimation of synonymous and nonsynonymous substitution rates (*Yang and Nielsen, 2000*)
- conservative test genes positively selected among all 4 species
- out of 1152 genes:
  - 22 genes with $\omega > 1$
  - 769 genes with $\omega > 0$
  - TRIM5α $\omega > 1$
  - APOBEC genes $0.83 < \omega < 0.93$

max planck institut informatik

Virus

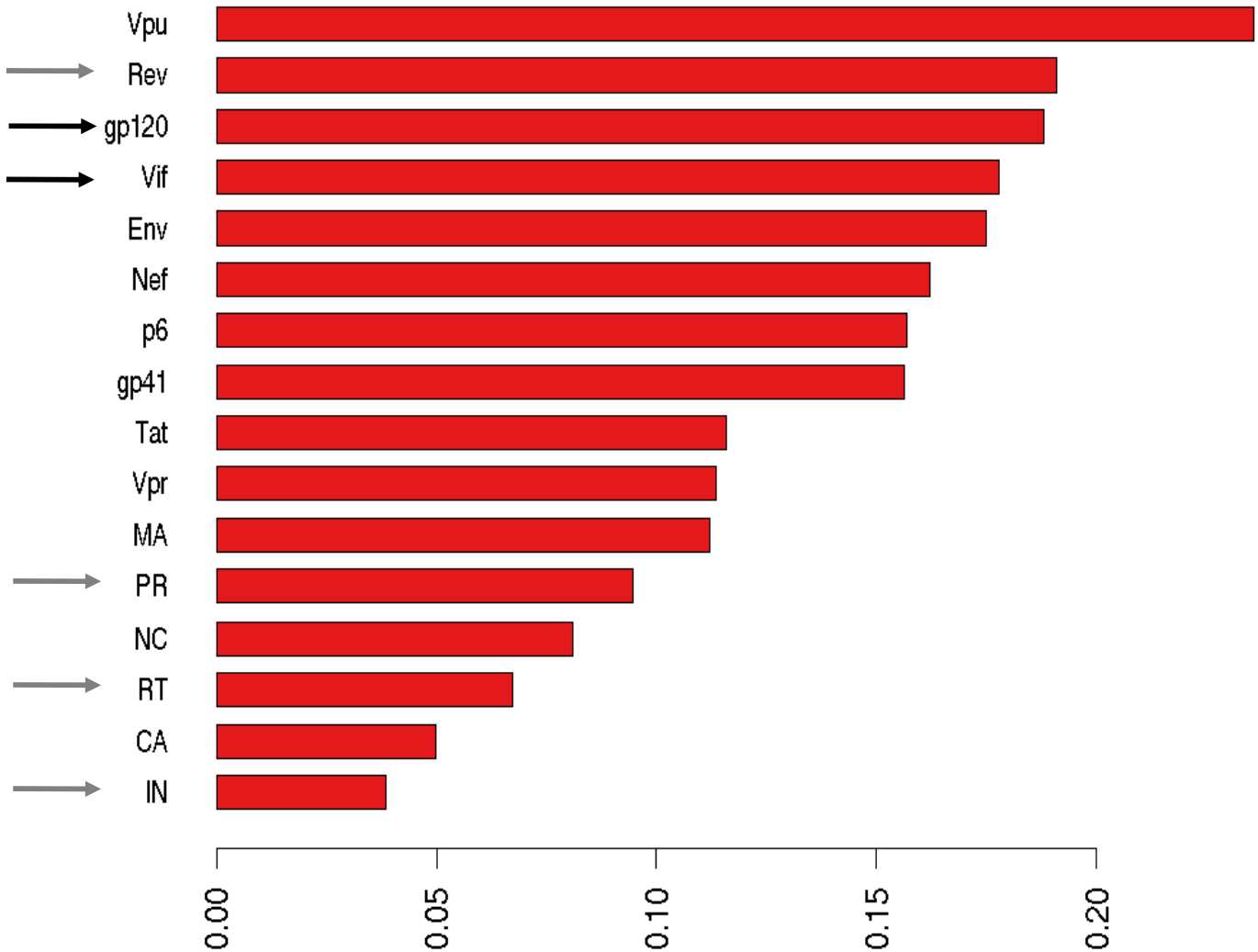
- more genomes
- large divergence within and among species
- short coding sequences



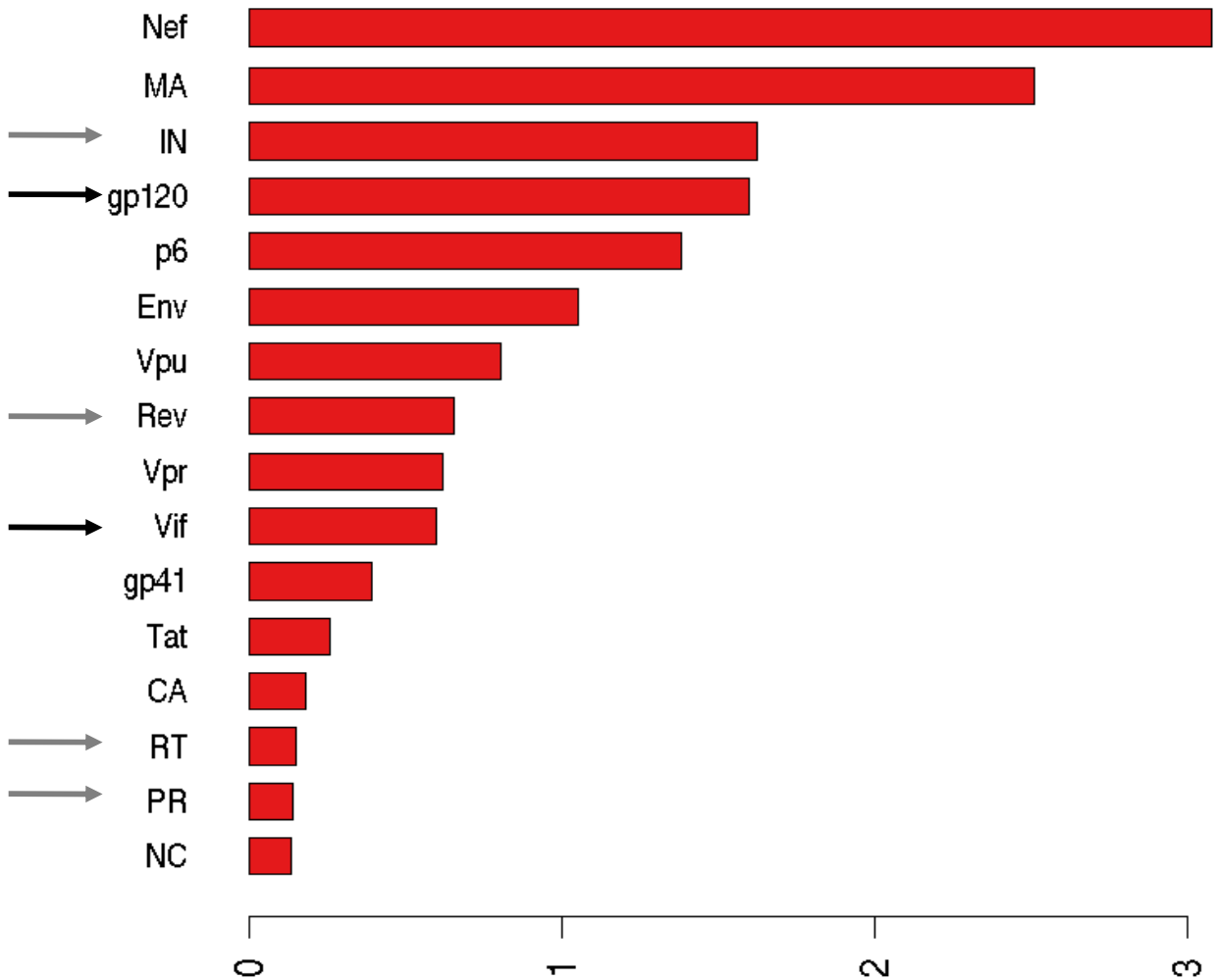
protease

max planck institut informatik

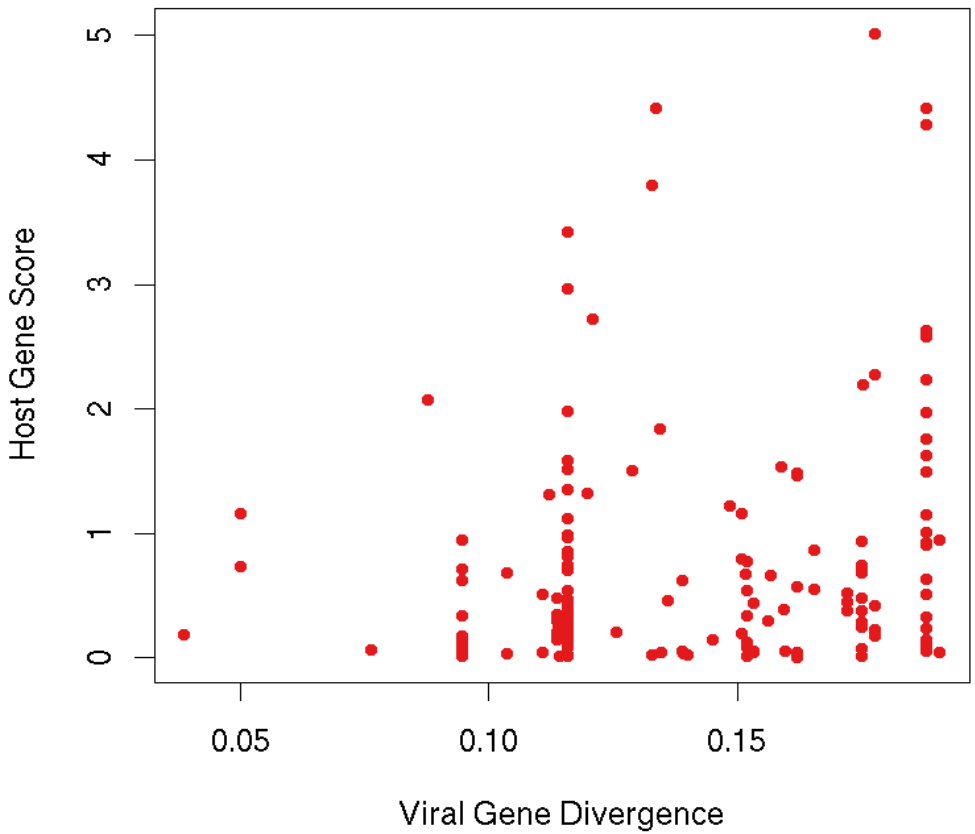
Virus – interspecies divergence

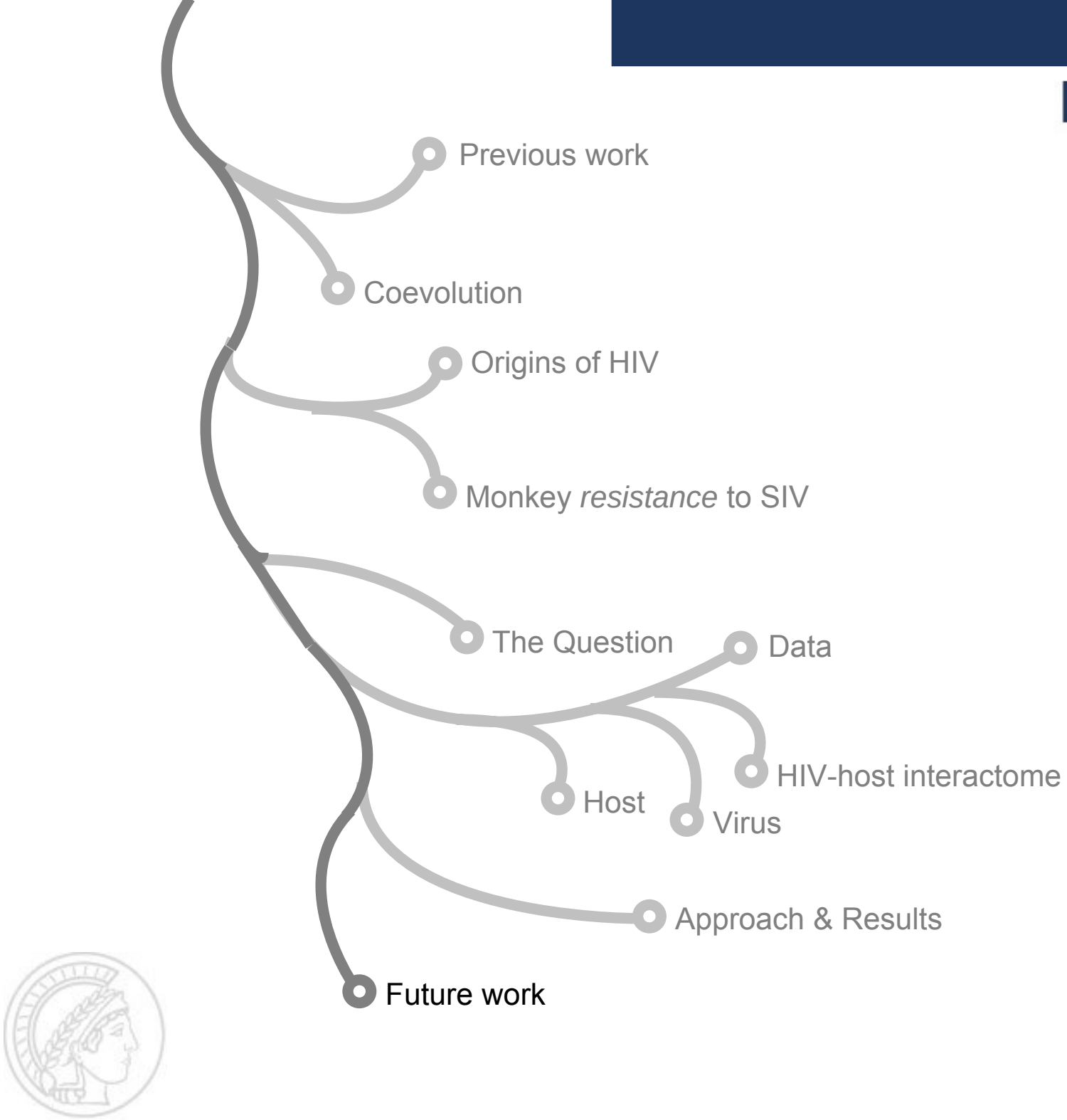


Virus – positive selection among species



Host and Virus combined




Previous work
Coevolution
Origins of HIV
Monkey *resistance* to SIV
The Question
Data
HIV-host interactome
Host
Virus
Approach & Results
Future work

max planck institut informatik

- coherent model of host-virus coevolution based on evolution measures and interaction data
  - quantification of the extent of HIV/SIV-host coevolution
  - characterization of know examples of lack of adaptation (TRIM5α – CA)
  - new potential “unadapted” interactions
  - quantification of the lack of adaptation

- How much time do the HIV and humans need to coadapt?
- To what extent can the HIV drive human evolution?
- Can we estimate viral pathogenicity from sequence-interaction-virulence data?
- …

max planck institut informatik

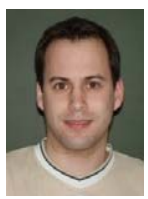
Acknowledgements

Christoph Bock

Oliver Sander

Alice McHardy

Alexander Thielen



Andre Altmann

Thomas Lengauer