

HIV-1 sequence alignment

- Mafft
 - option E-INS-I

```

○○○○○○○○○○XXX-----XXXX-----XXXXXXXXXXXX-XXXXXXXXXXXXXXXX○○○○○○○○○○
-----XXXXXXXXXXXXXXXX○○-----XXXXXXXXXXXXXXXXXXXX-XXXXXXXX-----
-----○○○○XXXXXX--XXXX○○○○○○○○-----XXXX--XXXXXXXXXXXXXXXXXXXXXXXX○○○○○○○○○○
-----XXXXX--XXXX○○○○○○○○○○○○○○○○○○○○○○○○○○○○XXXX-XXXXXXXXXXXX--XXXXXX-----
-----XXXXX--XXXX-----XXXXX--XXXXXXXXXXXX-XXXXXX○○○○-----

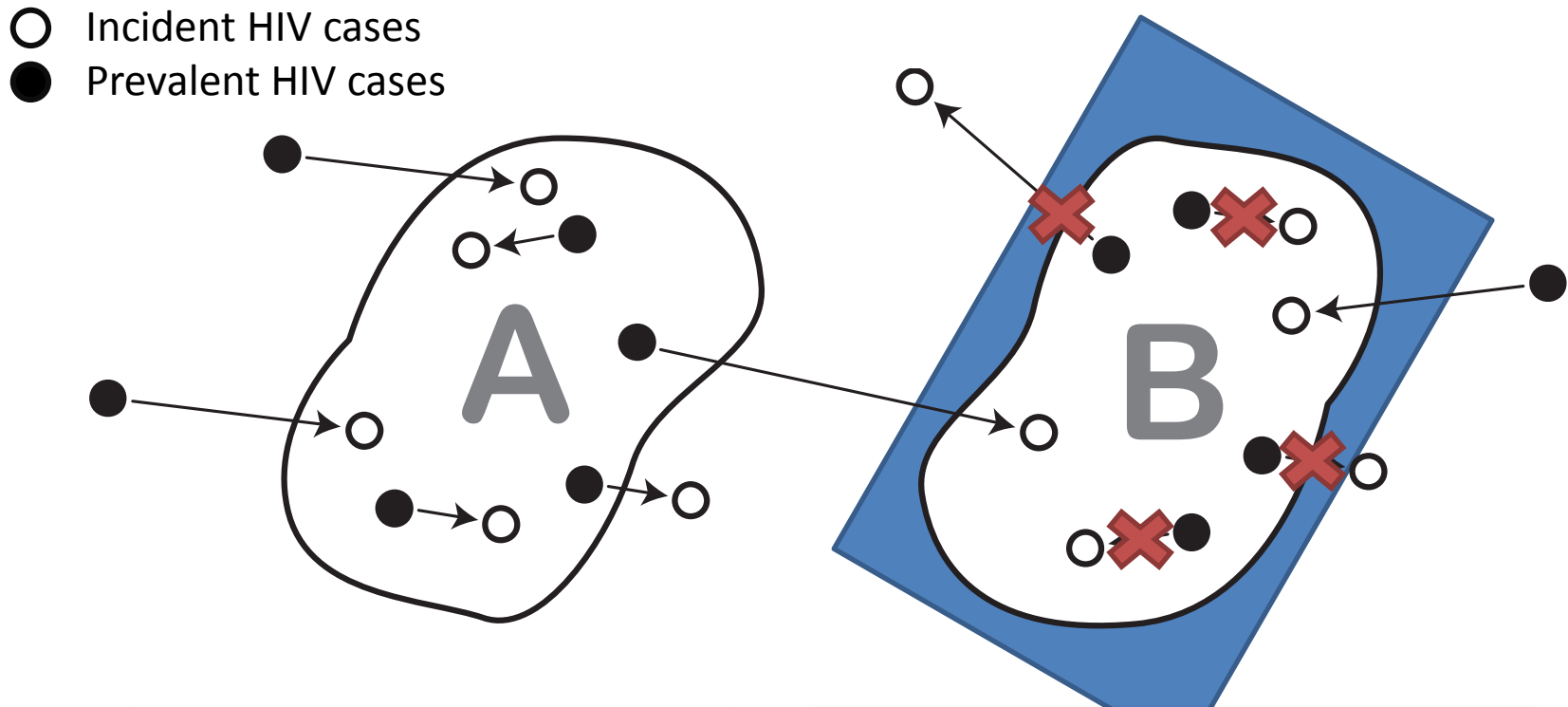
```

X: alignable residues

○: un-alignable residues

- : gaps

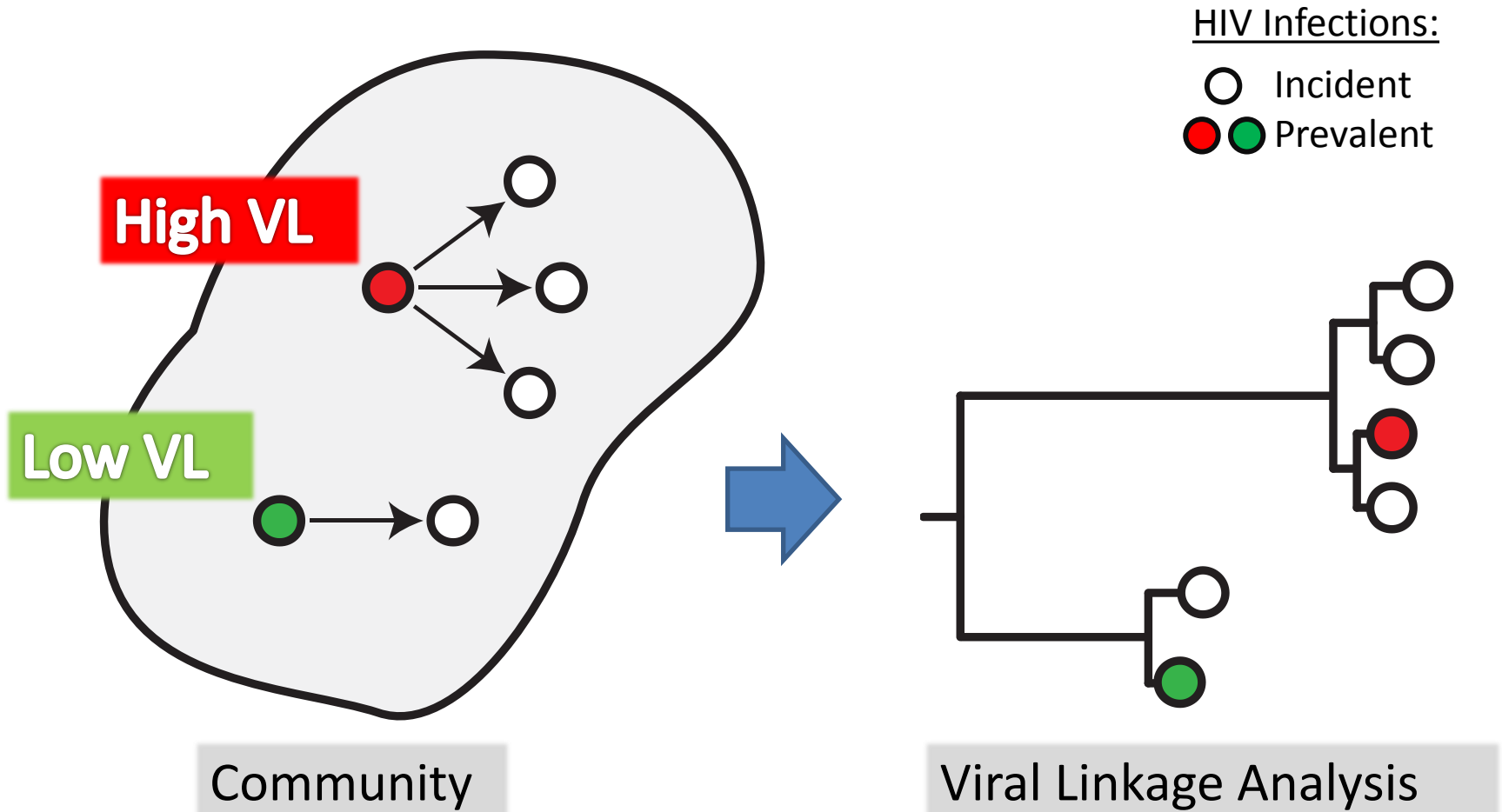
HIV-1 Transmission: Viral Linkage



Viral linkage:

- To identify the source (within vs. outside) of HIV transmission
- **To refine estimates of HIV incidence**

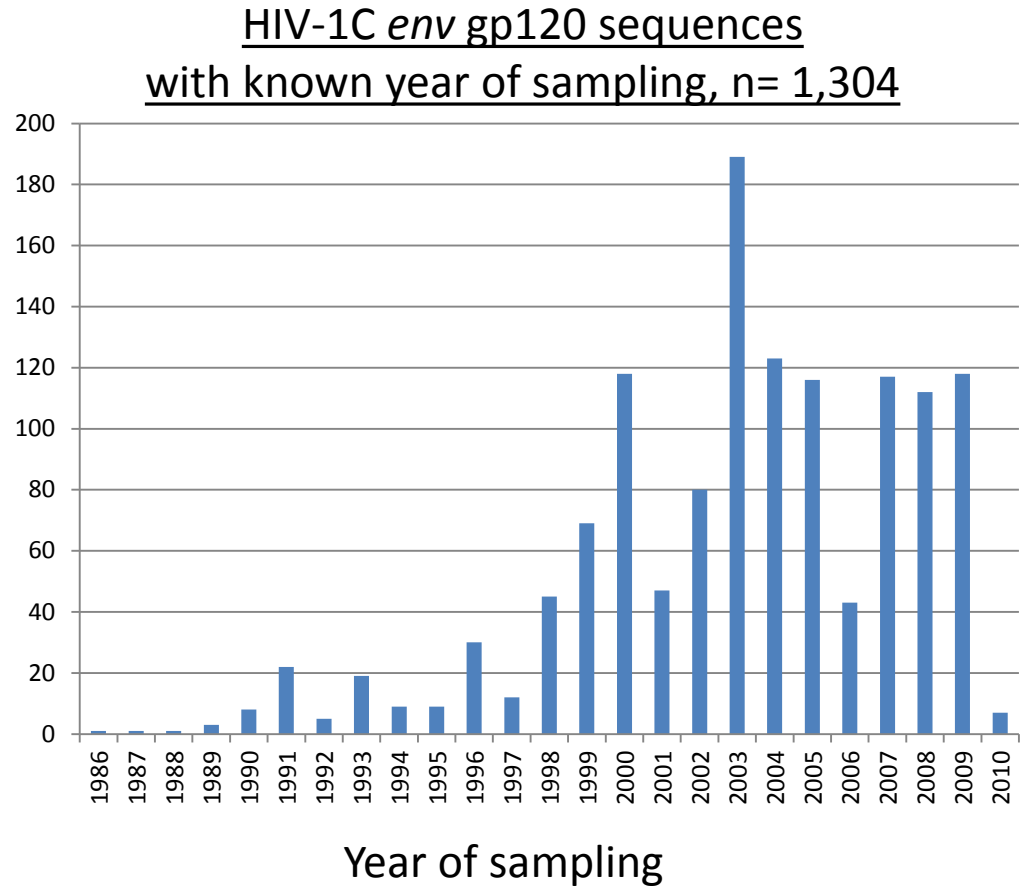
HIV-1 Transmission: Viral Linkage



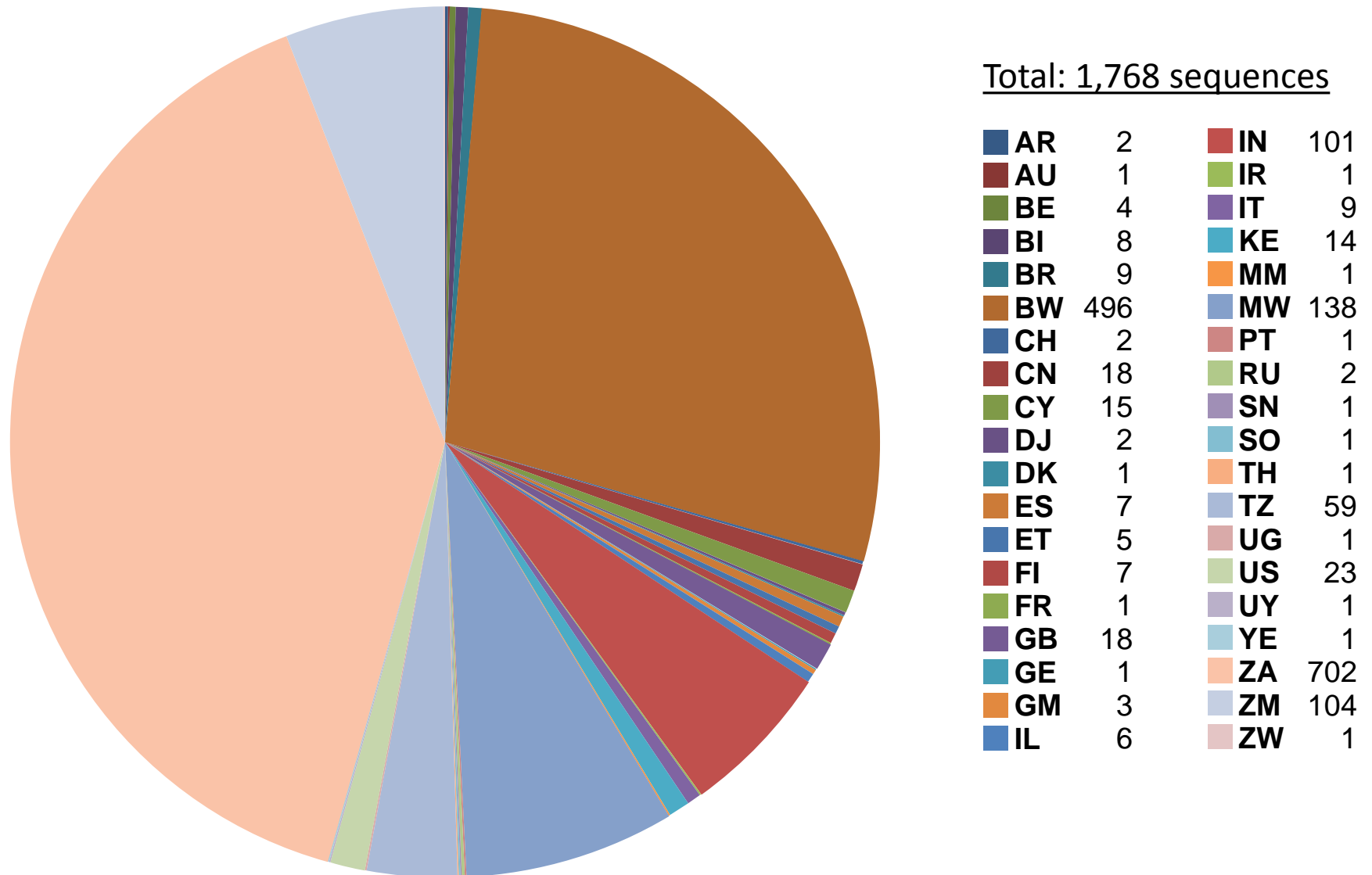
HIV-1C *env* gp120: Distribution by year

LANL/GenBank:

- HIV-1C *env* gp120
- $\geq 1,000$ bp
- N: 8,352
- Single sequence per subject: 1,379
- Country info: 1,366
- Year of sampling: 1,304

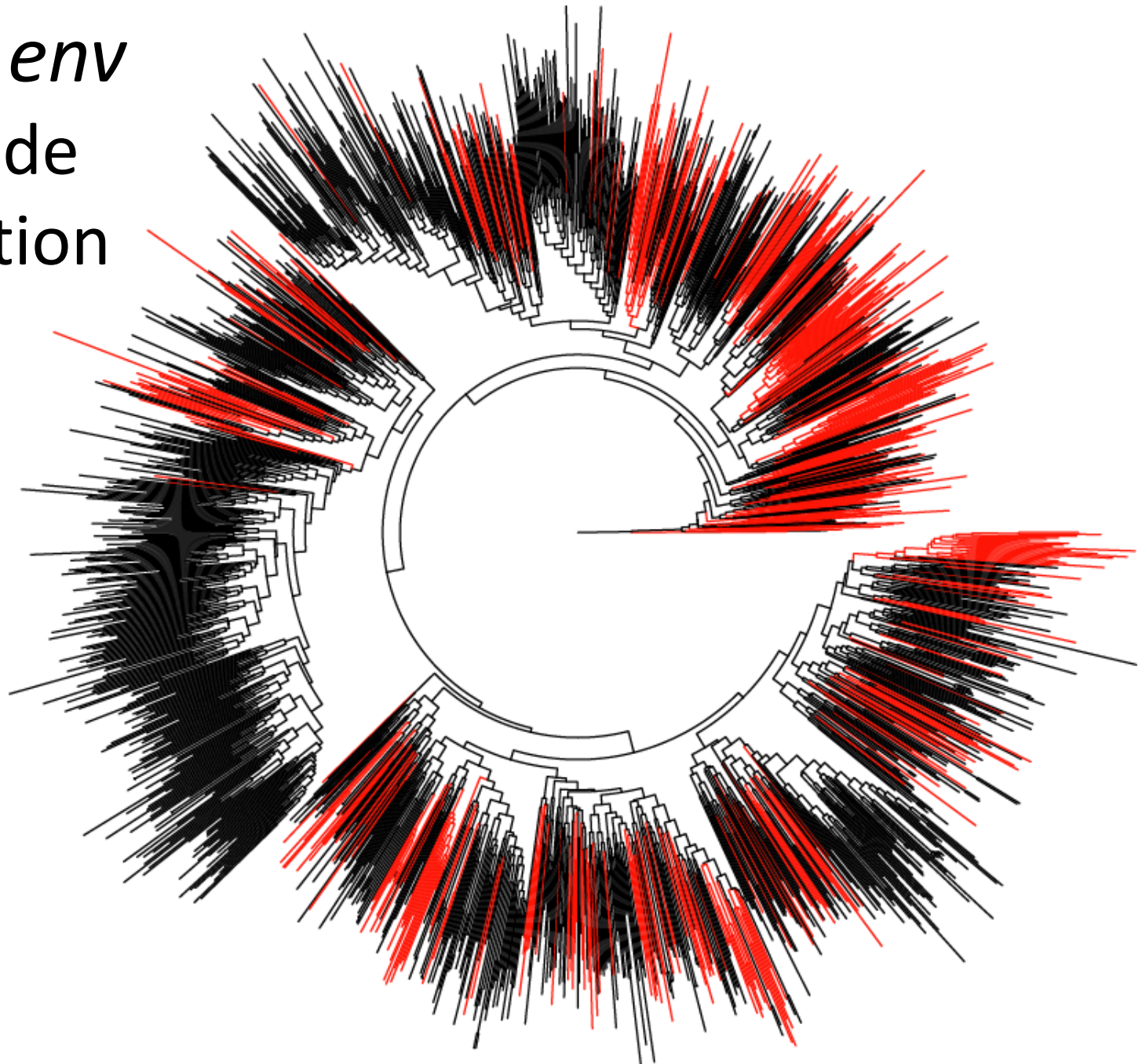


HIV-1C *env* gp120: Distribution by country



HIV-1C *env* worldwide distribution

ML tree
FT2-018M

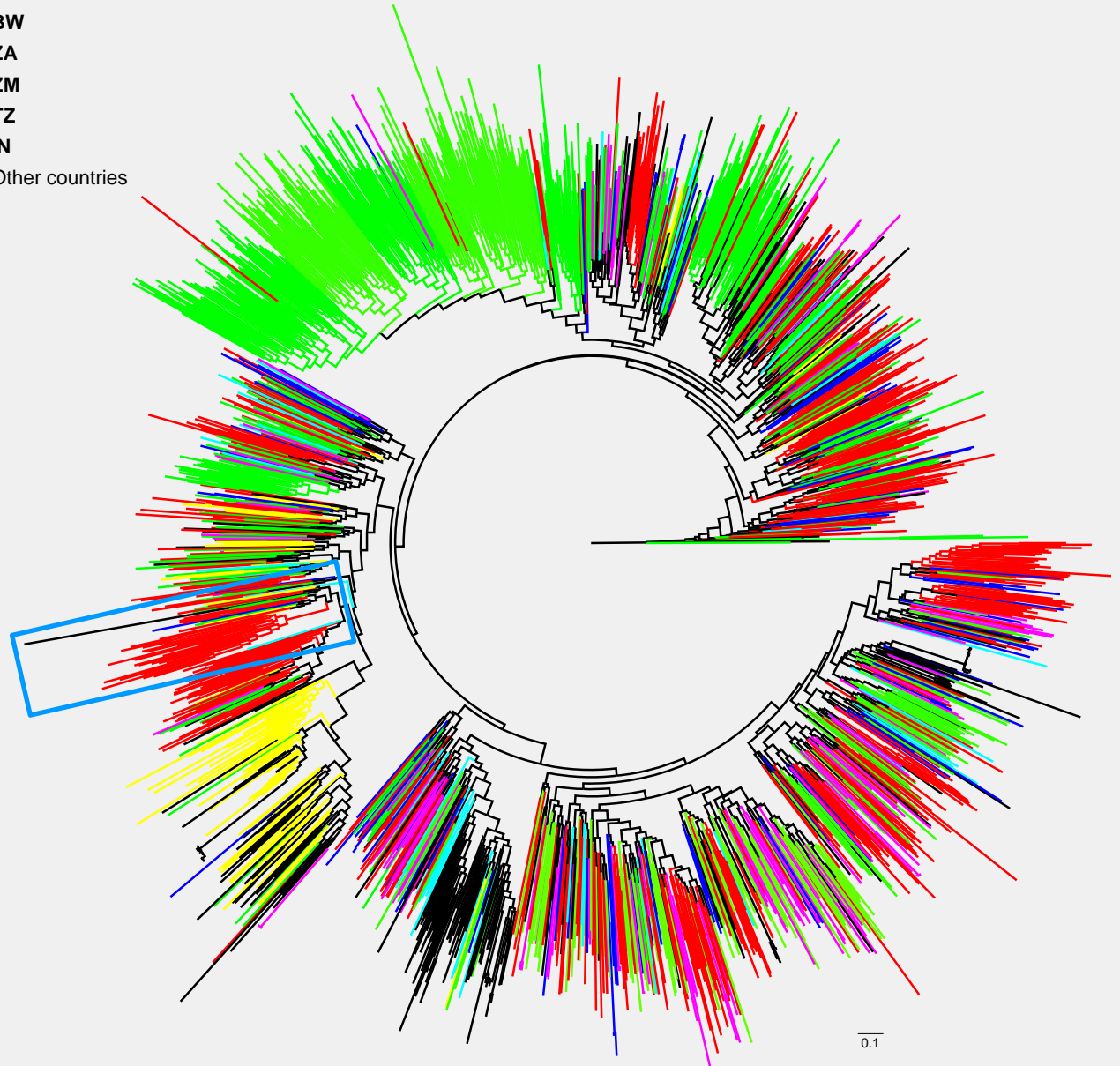


Sequences
from Botswana
are in red

HIV-1C *env* worldwide distribution

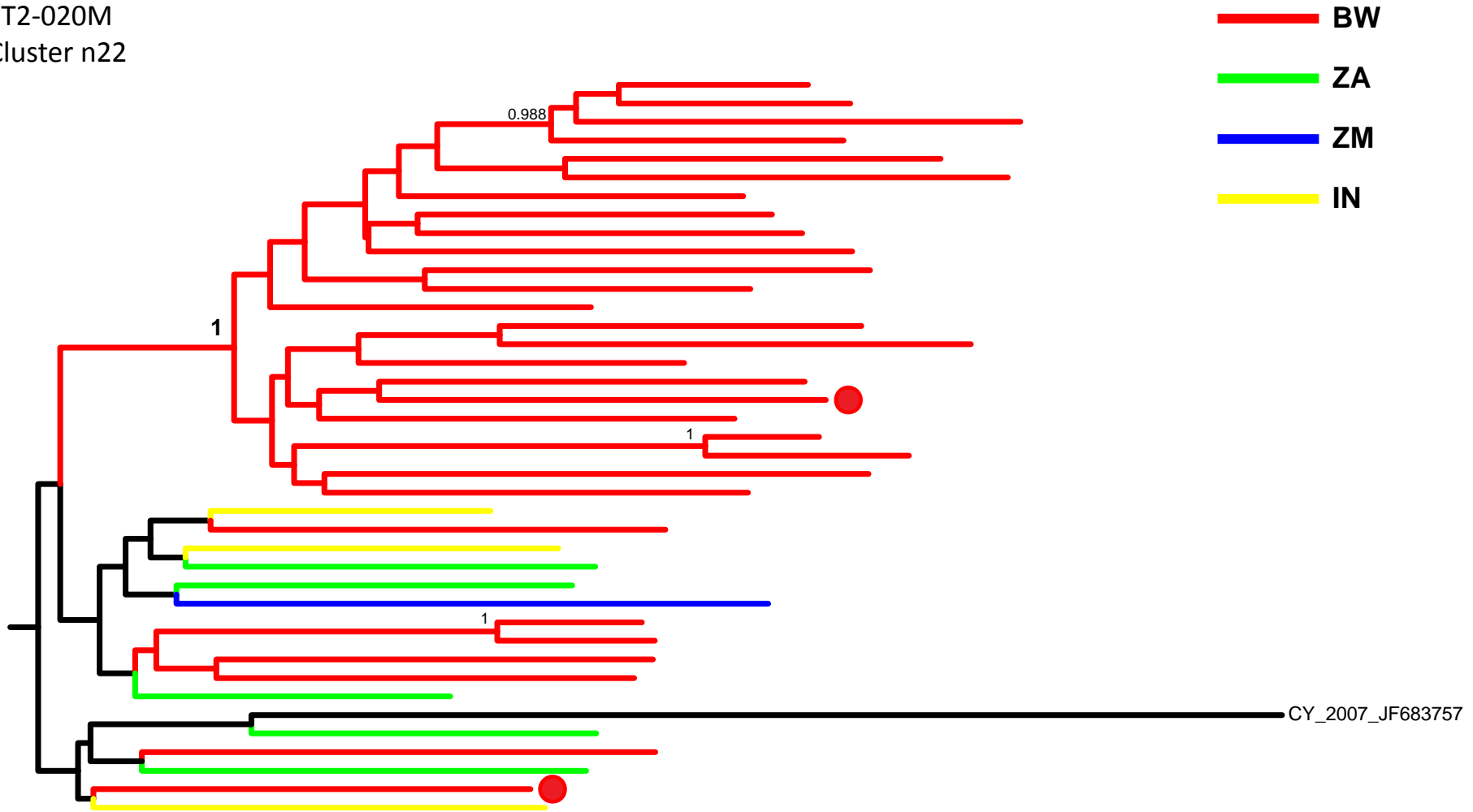
ML tree
FT2-020M

— BW
— ZA
— ZM
— TZ
— IN
— Other countries



HIV-1C *env* worldwide distribution

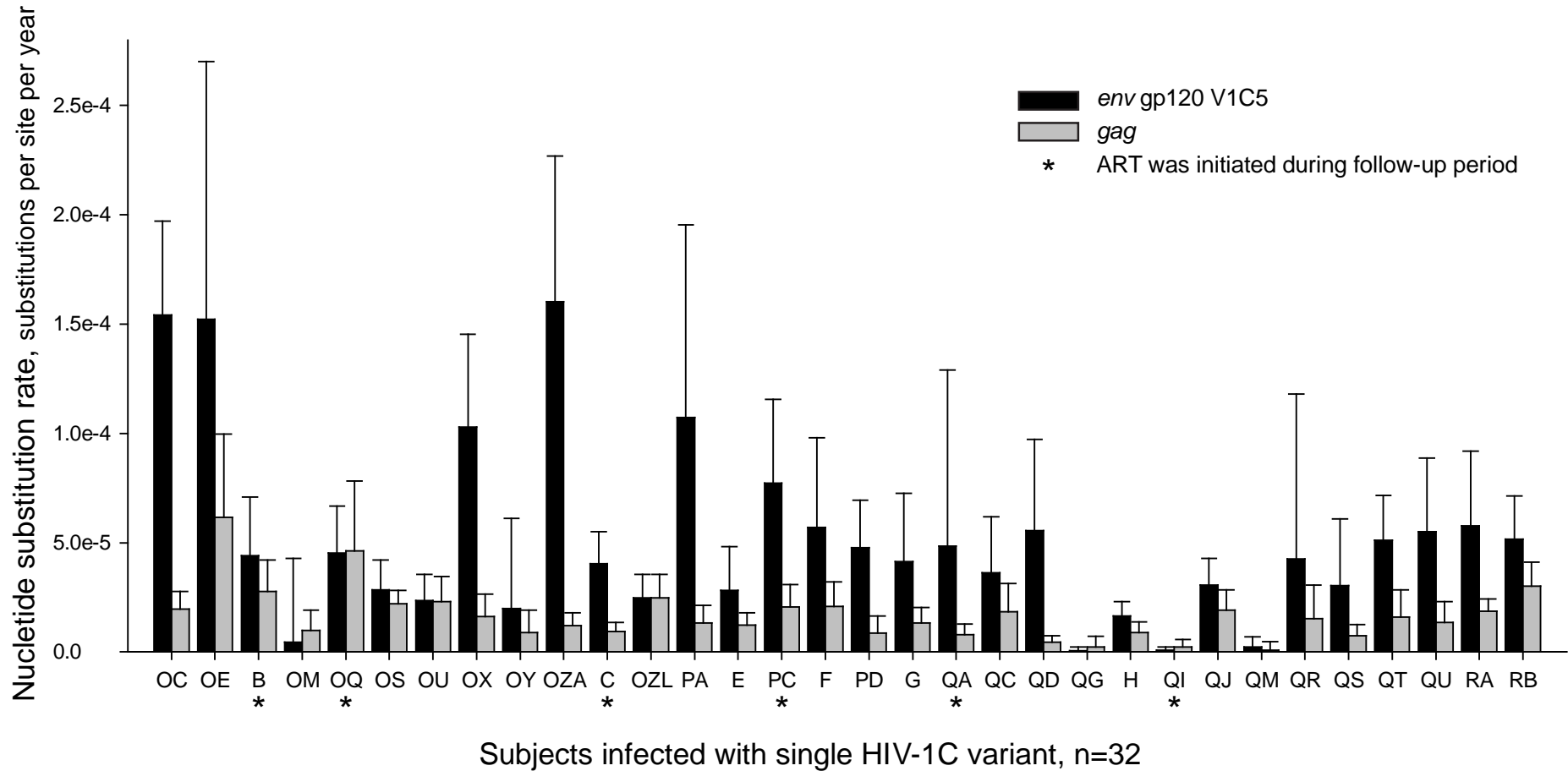
ML tree
FT2-020M
Cluster n22



Evolutionary Rates

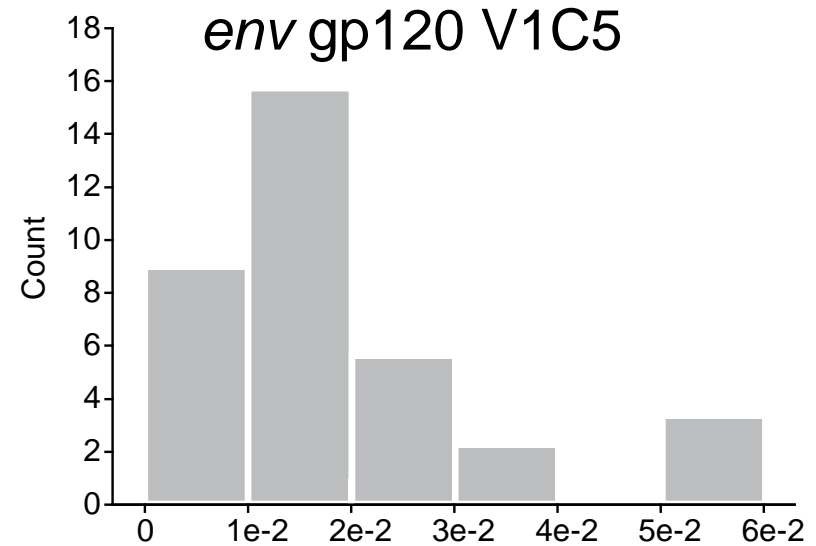
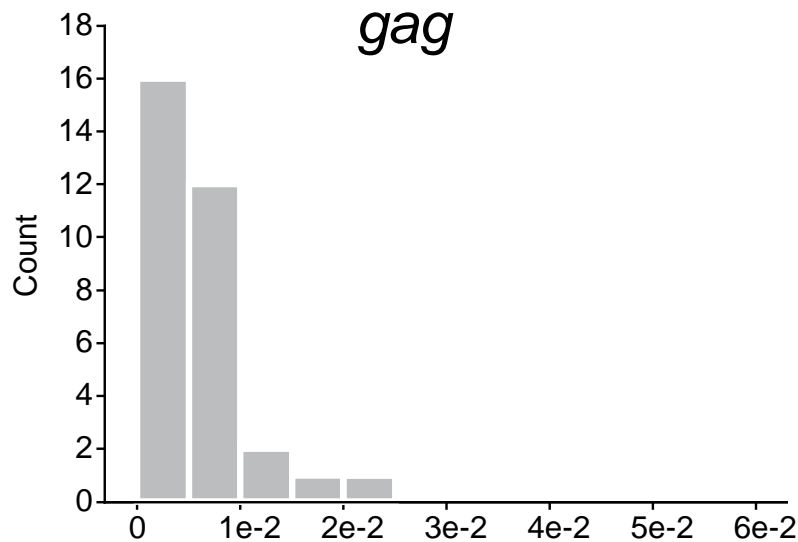
**HIV-1 subtype C substitution rates in
gag and *env***

Intra-host Evolutionary Rates in HIV-1C *gag* & *env*



Intra-host Evolutionary Rates in HIV-1C *gag* & *env*

Distribution of HIV-1C *gag* and *env* substitution rates, n=32



Substitutions per site per year

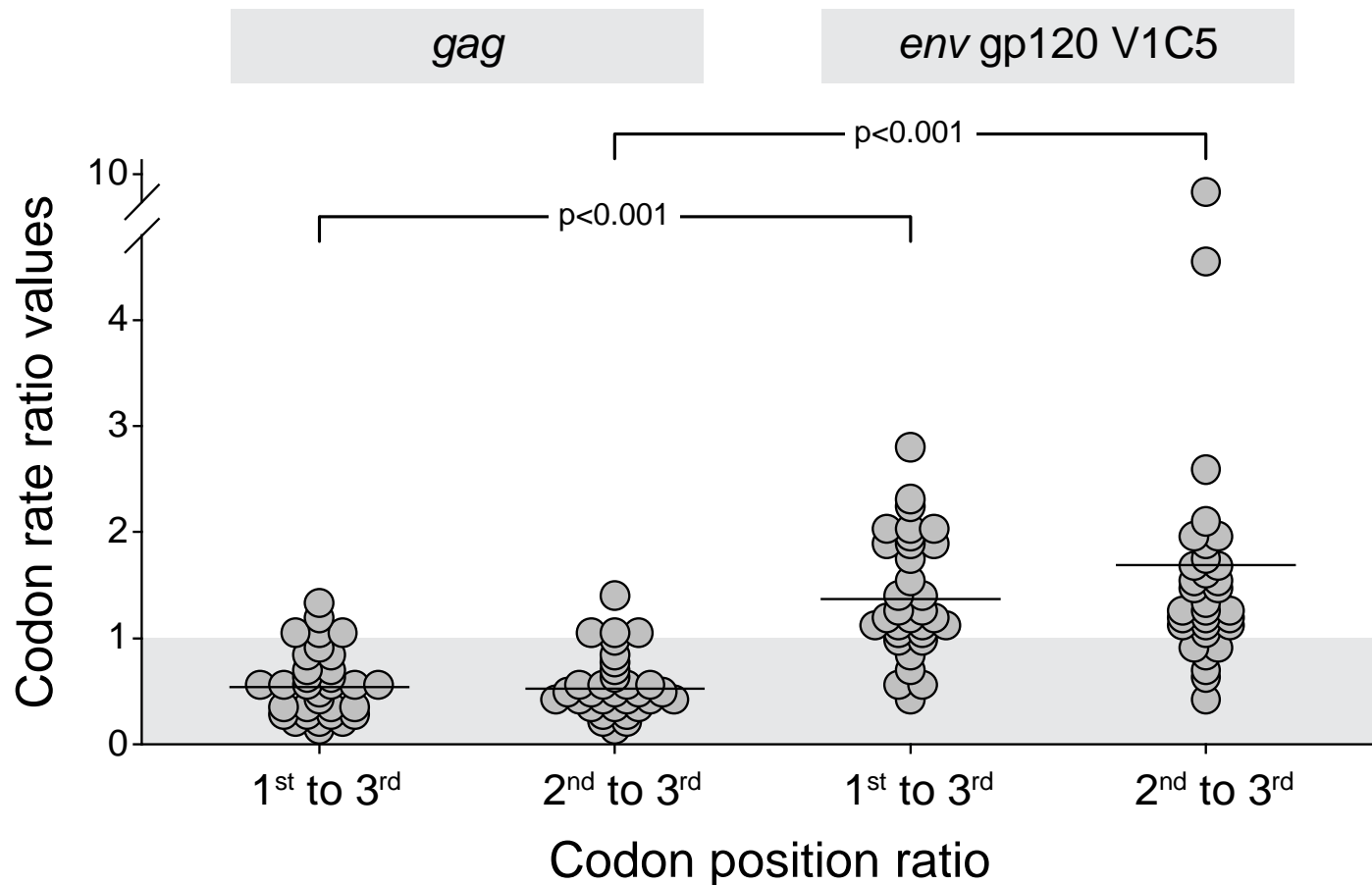
Median (IQR):

5.22E-03 (3.28E-03–7.55E-03)

1.58E-02 (9.99E-03–2.04E-02)

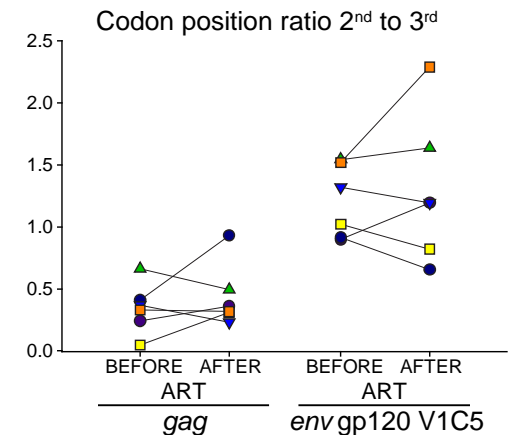
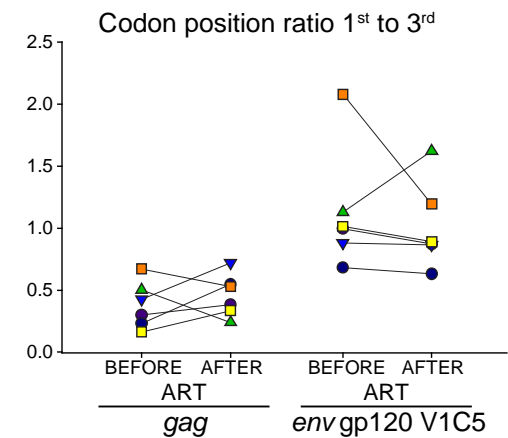
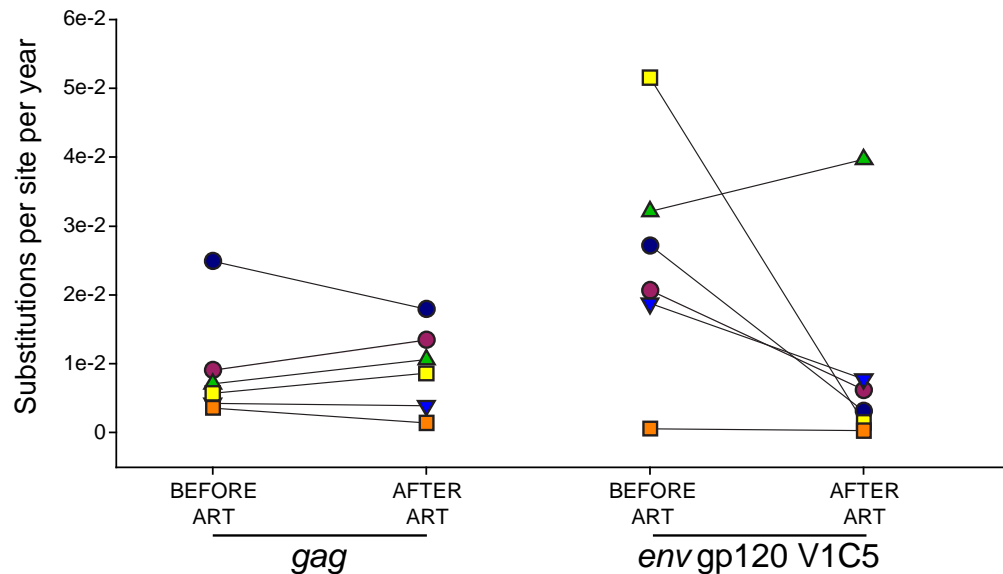
Intra-host Evolutionary Rates in HIV-1C *gag* & *env*

Relative substitution rates



Intra-host Evolutionary Rates in HIV-1C *gag* & *env*

HIV-1C substitution rates before and after ART



Vlad's points to success: suggestions to young TanZamBo scientists:

- **Maintain high level of science in your research:**
 - **Generate new ideas & hypothesis-driven research:**
 - Read papers in professional journals;
 - Think creatively:
 - What would you did differently? What would you do better?
 - Challenge existing paradigms;
 - Come up with a new question and develop a plan/approach: how to address this question?
 - **Methodological innovations:**
 - Learn and apply novel methods and techniques:
 - Visit other labs – short visits are fine;
 - Collaborate with other labs;
 - If you're going to buy equipment, consider used;
 - Explore collaborative usage: try to create core facility.
 - Try to reproduce new methods from the papers;
 - Set up a goal: to submit 2-4 abstracts per year;
 - Visit professional conferences:
 - Present and discuss your work;
 - Use comments and discussions in your paper;
 - Communicate with other researchers.
- **Statistical and bioinformatics support (Sample size? Power?):**
 - Take classes on biostatistical methods in research;
 - Communicate with the experts.