PhyloPi: an affordable, purpose built phylogenetic pipeline for the HIV drug resistance testing facility

PA Bester,
A De Vries,
SJPK Riekert,
K Steegen,
G Van Zyl,
D Goedhals
INTRODUCTION
From specimen to sequence

Nucleic acid extraction → cDNA synthesis → 1st Round PCR → Nested PCR

Sequencing → PCR cleanup

T: +27(0)51 401 9111 | info@ufs.ac.za | www.ufs.ac.za
The n steps of phylogenetic inference

- Multiple sequence alignment
- Curation of multiple sequence alignment
  - Overhangs, gaps, missalignments
  - Programmatically or human
- Maximum likelihood calculation
- Tree rendering
PhyloPi: A block diagram

Software Used:
Well cited
Opensource
API or CLI

![Block Diagram of PhyloPi](image)
Phylopi: The hardware

- Broadcom BCM2837B0, Cortex-A53 (ARMv8) 64-bit SoC @ 1.4GHz
- 1GB LPDDR2 SDRAM
- 2.4GHz and 5GHz IEEE 802.11 b/g/n/ac wireless LAN, Bluetooth 4.2, BLE
- Gigabit Ethernet over USB 2.0 (maximum throughput 300 Mbps)
- Extended 40-pin GPIO header
- Full-size HDMI
- 4 USB 2.0 ports
- CSI camera port for connecting a Raspberry Pi camera
- DSI display port for connecting a Raspberry Pi touchscreen display
- 4-pole stereo output and composite video port
- Micro SD port for loading your operating system and storing data
- 5V/2.5A DC power input
- Power-over-Ethernet (PoE) support (requires separate PoE HAT)
PhyloPi: Considerations for software

- Opensource
  - Source code available for compilation
- CLI or API
- Well cited
  - High quality
- Light footprint yet accurate
PhyloPi: Initial setup
PhyloPi: Routine phylogenetics
PhyloPi: Results ....
Phylopi: Sanity check
Phylopi: Sanity check
Phylopi: Sanity check

Sanity check on your data

Fasta File: Browse... No file selected.

This is a crude test to determine whether your data is suitable for phylogenetics.

Each sequence in your fasta file will be aligned to references and be displayed graphically.
Phylopi: Sanity check

UFS UV

© Copyright reserved Kopiereg voorbehou
Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination antiretroviral therapy

Hiromi Imamichi\textsuperscript{a}, Robin L. Dewar\textsuperscript{b}, Joseph W. Adelsberger\textsuperscript{b}, Catherine A. Rehm\textsuperscript{a}, Una O’Doherty\textsuperscript{c}, Ellen E. Paxinos\textsuperscript{d}, Anthony S. Fauci\textsuperscript{a,1}, and H. Clifford Lane\textsuperscript{a}

\textsuperscript{a}Laboratory of Immunoregulation, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892; \textsuperscript{b}Clinical Services Program, Applied and Development Research Directorate, Leidos Biomedical Research, Inc., Frederick, MD 21072; \textsuperscript{c}Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA 19104; and \textsuperscript{d}Applications and Collaborations, Pacific Biosciences, Menlo Park, CA 94025

Contributed by Anthony S. Fauci, June 9, 2016 (sent for review April 23, 2016; reviewed by Scott Hammer and Nelson L. Michael)

Despite years of plasma HIV-RNA levels <40 copies per milliliter during combination antiretroviral therapy (cART), the majority of HIV-infected patients exhibit persistent seropositivity to HIV-1 and evidence of immune activation. These patients also show persistence of proviruses of HIV-1 in circulating peripheral blood mononuclear cells. Many of these proviruses have been characterized as defective and thus thought to contribute little to HIV-1 pathogenesis. By combining 5’LTR-to-3’LTR single-genome amplification and direct amplicon sequencing, we have identified the presence of “defective” proviruses capable of transcribing novel unspliced HIV-RNA (usHIV-RNA) species in patients at all stages of HIV-1 infection. Although these novel usHIV-RNA transcripts amplification of up to 8.9 kb HIV-1 DNA and 8.4 kb cell-associated unspliced HIV-RNA (usHIV-RNA), and sequencing of the resulting near full-length HIV-1 genome fragments. This approach allowed us to better characterize the genetic variability in HIV-1 proviral genomes and to determine which “defective” proviruses were transcribed and capable of encoding viral proteins.

Results
Characterization of HIV-1 Proviruses in Patients with HIV-1 Infection. The 5’LTR-to-3’LTR single genome amplification and direct amplicon sequencing of HIV-1 proviruses was performed for
Phylopi: Search features

Search the master fasta file for samples
Type your sample IDs in the box below, if more than one use commas

Search previous results
Name | Primary sample | Secondary sample
If all fields are left empty, all previous results will be retrieved

Search
Return to input
Phylopi: Flexing its muscle ...

```
hiv.lang.gov
POL CDS (11 337)
>______________________
>______________________
>______________________
>______________________
>______________________
>______________________
>______________________

n = 1
while n <= 50:
    select n random sequences using PhyloPi WI upload
    n += 1
```
N sequences retrieved by BLAST

Time used by BLAST vs. N input sequences

Time used by MAFFT vs. total sequences

Time used by FastTree vs. total sequences

\[ y = 4.628x \]
\[ R^2 = 0.998 \]

\[ y = 11.02x \]
\[ R^2 = 1 \]

\[ y = 0.153x^{1.87} \]
\[ R^2 = 0.993 \]
N sequences retrieved by BLAST

Time used by BLAST vs. N input sequences

\[ y = 4.628x \]
\[ R^2 = 0.998 \]

Time used by MAFFT vs. total sequences

\[ y = 11.02x \]
\[ R^2 = 1 \]

Time used by FastTree vs. total sequences

\[ y = 0.652x \]
\[ R^2 = 0.993 \]
N sequences retrieved by BLAST

Time used by BLAST vs. N input sequences

Time used by MAFFT vs. total sequences

Time used by FastTree vs. total sequences

Time used by MAFFFT vs. total sequences

Time used by FastTree vs. total sequences
N sequences retrieved by BLAST

Time used by BLAST vs. N input sequences

Time used by MAFFT vs. total sequences

Time used by FastTree vs. total sequences
N sequences retrieved by BLAST

Time used by BLAST vs. N input sequences

Only MAFFT: O(N^2)
A note on FastTree:

gcc -DUSE_DOUBLE -O3 -finline-functions -funroll-loops -Wall -o FastTree FastTree.c -lm
N sequences retrieved by BLAST

Time used by BLAST vs. N input sequences

Time used by MAFFT vs. total sequences

Time used by FastTree vs. total sequences
Choosing colours: Intra- and inter distances ...

<table>
<thead>
<tr>
<th></th>
<th>Inter sequences</th>
<th>Intra clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subtype B</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR_RT</td>
<td>1969</td>
<td>359</td>
</tr>
<tr>
<td>INT</td>
<td>2461</td>
<td>405</td>
</tr>
<tr>
<td><strong>Subtype C</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR_RT</td>
<td>1753</td>
<td>255</td>
</tr>
<tr>
<td>INT</td>
<td>1171</td>
<td>217</td>
</tr>
</tbody>
</table>
...
PhyloPi: Utility

• PhyloPi beats transcription errors
  – Names, DOB
• PhyloPi beats transcription errors
  – Names, DOB
• Patients may use different first names
• PhyloPi beats transcription errors
  – Names, DOB
• Patients may use different first names
• Patients may marry and change their surname
• PhyloPi beats transcription errors
  – Names, DOB
• Patients may use different first names
• Patients may marry and change their surname
• PhyloPi sees vertical / horizontal transmission
• PhyloPi beats transcription errors
  – Names, DOB
• Patients may use different first names
• Patients may marry and change their surname
• PhyloPi sees vertical transmission
  – We are not convicting a dentist
• PhyloPi beats transcription errors
  – Names, DOB
• Patients may use different first names
• Patients may marry and change their surname
• PhyloPi sees vertical
  – We are not convicting a dentist
• PhyloPi has found patient/folder mixups in clinics
  – We have been accused of sorcery
Conclusion

• The software is free, the hardware is cheap
  – $35.00
• Portable and standalone
  – WiFi hotspot, self contained
• Fast, but accurate
• A safety net
• Human insight still required
The grey matter which mattered
Chat with us:
Poster session
Thank You

Dankie