Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Universität Zürich[UZH]

UNI BASEL

SIB Swiss Institute of Bioinformatics

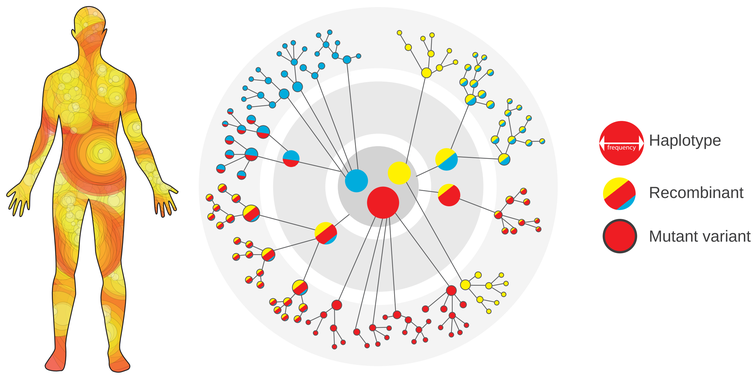
Full-length HIV-1 Haplotype Reconstruction from Heterogeneous Virus Populations

Armin Töpfer[1,2,ϰ], Francesca Di Giallonardo[3,4,ϰ], Melanie Rey[5], Sandhya Prabhakaran[5], Yannick Duport[3], Christine Leemann[3], M. Rita Lecca[6], Andrea Patrignani[6], Martin Däumer[7], Christian Beisel[1], Huldrych F. Günthard[3*], Volker Roth[5,*], Niko Beerenwinkel[1,2,*] and Karin J. Metzner[3,*]

[1] Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, Basel, Switzerland
[2] SIB Swiss Institute of Bioinformatics, Basel, Switzerland
[3] Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, University of Zurich, Zurich, Switzerland
[4] Life Science Zurich Graduate School, University of Zurich, Zurich, Switzerland
[5] Department of Mathematics and Computer Science, University of Basel, Basel, Switzerland
[6] Functional Genomics Center Zurich (FGCZ), University of Zurich, ETH Zurich, Zurich, Switzerland
[7] Institut für Immunologie und Genetik, Kaiserslautern, Germany
[ϰ] Equal contributions

1 Introduction

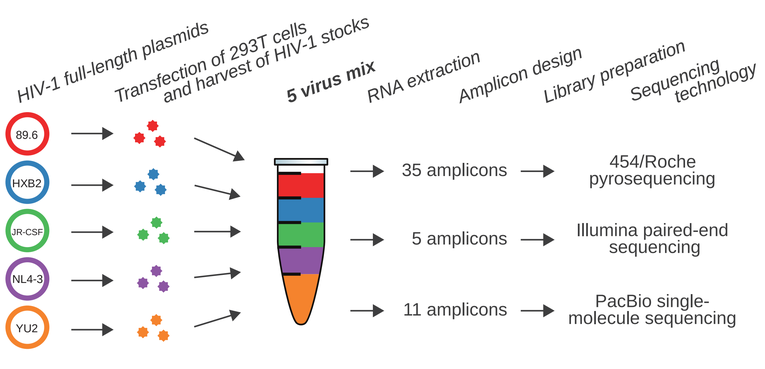
Next-generation sequencing (NGS) can provide in-depth resolution of heterogeneous intra-patient viral populations, composed of one or more genetically distinct haplotypes. However, such analyses are complicated by small read length and by the error rate of amplification and sequencing. Identification of co-occurring resistance mutations, composition, and ongoing evolution of intra-patient viral populations on the basis of full-length genomes are of great interest, but remain unsolved. Our protocols and methods can decompose a heterogeneous population into **individual full-length haplotypes** to **identify quasispecies structures**, **superinfections**, and **study evolution of viral populations**.



Typical viral population with a few dominant haplotypes and a swarm of lower frequent variants, generated by mutation and recombination.

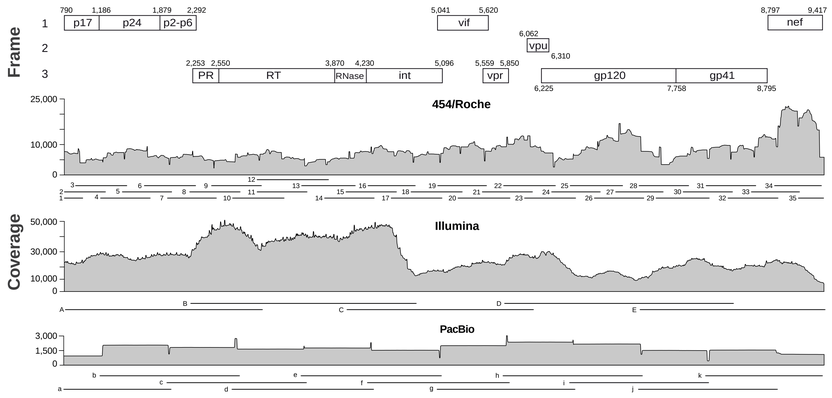
2 Experimental design

Every sequencing technology demands its own amplicon design and library preparation.



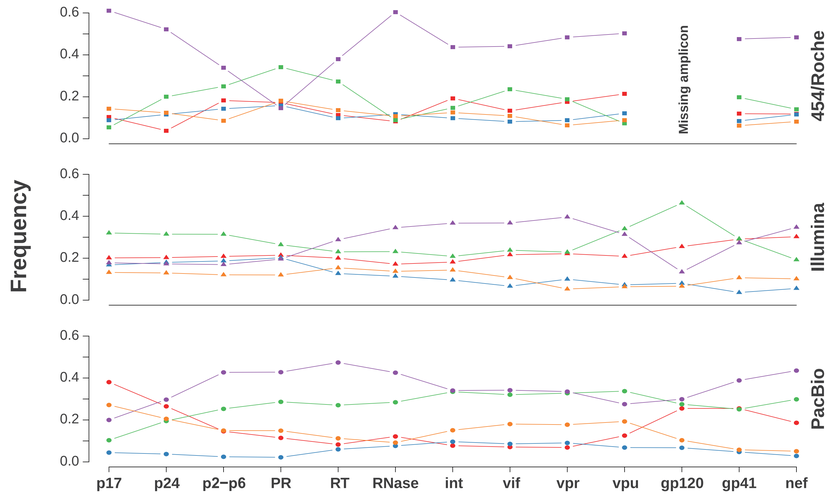
3 Sequencing

Sufficient amplicon overlap is a prerequisite for haplotype phasing. Coverage cannot compensate for too short reads.



4 Gene-wise reconstruction

Fluctuating gene-wise haplotype frequency estimates.

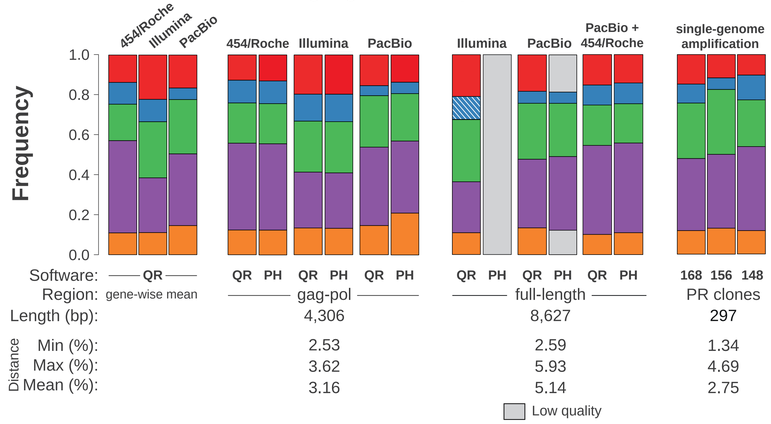


	p17	p24	p2−p6	PR	RT	RNase	int	vif	vpr	vpu	gp120	gp41	nef
Length (bp):	396	693	413	297	1,320	350	866	578	291	248	1,533	1,037	620
Distance Min (%):	3.77	1.72	2.17	1.34	2.49	2.49	1.73	3.45	3.08	2.82	1.89	0.96	1.44
Max (%):	6.80	3.60	4.34	4.69	3.33	5.54	2.88	5.52	6.50	16.93	11.60	8.09	13.36
Mean (%):	5.23	2.82	3.30	2.75	2.88	4.62	2.29	4.81	4.96	12.82	9.50	6.29	8.06

Gene-wise haplotype reconstruction was performed with QuasiRecomb [1].

5 Full-length reconstruction

Frequency estimates are more robust for larger regions and similar to single-genome amplification results.



Haplotype reconstruction was performed with QuasiRecomb **(QR)** [1] and PredictHaplo **(PH)** [2].

6 Conclusion

Whole-genome viral haplotype reconstruction is technically possible, but depends critically on:

- amplicon layout and successful amplification,
- sufficient read length, and
- homogeneous coverage.

This work has been supported by the Swiss National Science Foundation under grants No. CR32I2_127017 and 146331 (to NB, KJM, VR, and HFG).

[1] Armin Töpfer, Osvaldo Zagordi, Sandhya Prabhakaran, Volker Roth, Eran Halperin, and Niko Beerenwinkel. **Probabilistic inference of viral quasispecies subject to recombination.** Journal of Computational Biology. February 2013
[2] Sandhya Prabhakaran, Melanie Rey, Osvaldo Zagordi, Niko Beerenwinkel, and Volker Roth. **HIV Haplotype Inference using a propagating Dirichlet Process Mixture Model.** Computational Biology and Bioinformatics, IEEE/ACM Transactions. January 2014

7 Contact

Armin Töpfer
Computational Biology Group, ETH Zürich

armin.toepfer@bsse.ethz.ch
@XLR