



Doctoral Thesis

Studies in Viral Quasispecies Reconstruction

Author(s):

Töpfer, Armin

Publication Date:

2014

Permanent Link:

<https://doi.org/10.3929/ethz-a-010280492> →

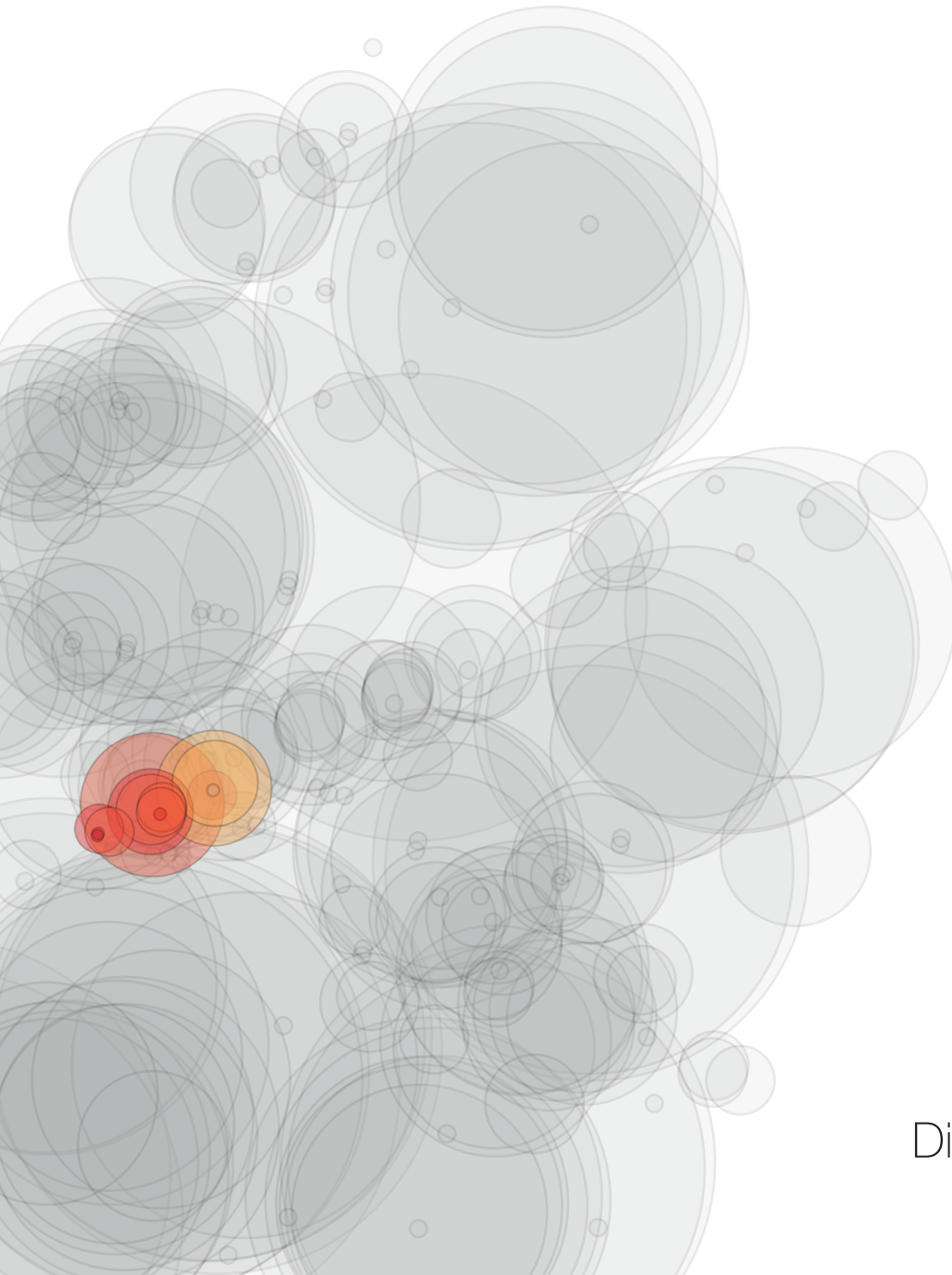
Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

STUDIES IN

VIRAL QUASISPECIES RECONSTRUCTION



Armin Töpfer
Diss ETH N° 22143

Abstract

Every year, more than one million deaths are caused by HIV-1. In contrast to other viruses, such as influenza, there is no vaccine or cure against HIV-1. A major reason is that the genetic diversity of HIV-1 within a single host is as high as the genetic diversity of influenza in the entire human population in one season. Every HIV-1 infected individual hosts billions of virions that can be summarized to a set of genetically different strains, called haplotypes, with their respective frequencies. Such distribution of haplotypes, also referred to as a viral quasispecies, evolves by point mutations, recombination, and selection. The study of viral quasispecies is of clinical importance, as low-frequency haplotypes may harbor drug resistance mutations that can lead to treatment failure and escape the immune system. Quasispecies diversity is also known to affect virulence and pathogenesis. The genetic fingerprint of a quasispecies is identified by sequencing. Although single-cell sequencing has been established, single-virion sequencing is not yet possible. Therefore, quasispecies samples are probed by next-generation sequencing technologies that allow to read millions of fragments in a single sequencing run. Each fragment, also called read, is much shorter than the genome length of interest and subject to substitutions, deletions, and insertions by the library preparation and the sequencing process. In addition to these limitations, in this process, the assignment of read to haplotype is lost. The goal is to assemble the viral quasispecies, i.e., reconstruct individual haplotypes and estimate their frequencies, from a jigsaw puzzle of overlapping reads. This is in particular a hard problem, as genetic distances between haplotypes can be very close, low-frequency variants have to be distinguished from sequencing errors, and recombinant structures are highly abundant in HIV-1 quasispecies.

This dissertation comprises four published studies and one book chapter on viral quasispecies reconstruction. This work has been performed during the author's doctorate in the Computational Biology Group at the Department of Biosystems Science and Engineering, ETH Zürich. For the structure of this dissertation, the book chapter serves as an introduction into the field of viral quasispecies reconstruction, including the required biological background, next- and third-generation sequencing, and an overview of probabilistic viral haplotype reconstruction methods and their relation to human genetics. Two studies present state-of-the-art haplotype reconstruction methods, *QuasiRecomb* and *HaploClique*. For the application of viral quasispecies reconstruction, two additional studies have been performed to shed light on the role of viral quasispecies in classical swine fever virus infected pigs and elaborate the feasibility of viral quasispecies assembly in clinical practice. Each study is shortly motivated and introduced into its field, accompanied by the main findings. The original papers are included in the corresponding sections, labeled with an indicator header. A final discussion summarizes the findings of this thesis, puts it into short and long term perspective, and presents possible applications outside its field.

Kurzfassung

Jährlich sterben mehr als eine Million Menschen an HIV-1. Im Gegensatz zu anderen Viren, wie die Influenza, gibt es keinen Impfstoff oder Heilmittel gegen HIV-1. Einer der Hauptgründe ist, dass die genetische Diversität von HIV-1 in einem Patienten gleich hoch ist wie die weltweite Variabilität von Influenza in einer Saison. In jedem HIV-1 Infizierten befinden sich Milliarden von Viruspartikel. Diese können zu einer Menge von genetisch verschiedenen Stämmen, auch als Haplotypen bekannt, zusammengefasst werden. Eine solche Verteilung von Haplotypen, auch als virale Quasispezies bezeichnet, entwickelt sich durch Punktmutationen, Rekombinationen und Selektion. Die Erforschung von viralen Quasispezies ist von klinischer Relevanz, da niedrigfrequente Haplotypen dem Immunsystem entweichen und Medikamentenresistenzen aufweisen können, die zum Fehlschlag der Therapie führen können. Zudem hat die Diversität einer Quasispezies einen Einfluss auf die Virulenz und den Krankheitsverlauf. Der genetische Fingerabdruck einer Quasispezies wird durch Sequenzierung ermittelt. Die Sequenzierung von einzelnen Zellen ist zwar bereits möglich, aber nicht einzelner Viruspartikel. Daher werden Quasispeziesproben, mit tausenden von Viruspartikeln pro Milliliter, mit neuesten Technologien analysiert, welche Millionen von Fragmenten sequenzieren können. Jedes Fragment, auch Read genannt, ist um ein vielfaches kürzer als der originale Haplotyp und ist durch Substitutionen, Deletionen und Insertionen der Probenaufbereitung und Sequenzierung verrauscht. In diesem Prozess geht die Zugehörigkeit von Read zu Haplotyp verloren. Das Ziel ist die Rekonstruktion von individuellen Haplotypen und die Schätzung derer Frequenzen von einem Puzzle überlappender Reads. Dies ist ein sehr schweres Problem, da die einzelnen Haplotypen sehr eng verwandt sein können, niedrigfrequente Varianten von Sequenzierfehlern unterschieden werden müssen und rekombinante Strukturen in HIV-1 Quasispezies reichlich vorhanden sind.

Diese Dissertation umfasst vier publizierte Studien und ein Buchkapitel über die Rekonstruktion von viralen Quasispezies. Diese Arbeit wurde während des Doktorats des Autors in der Computational Biology Gruppe am Department für Biosystems Science and Engineering an der ETH Zürich durchgeführt. In dieser Dissertation dient das Buchkapitel als Einleitung in das Forschungsgebiet der viralen Quasispeziesrekonstruktion, liefert das benötigte biologische Hintergrundwissen und Informationen über Sequenziertechnologien. Zwei Studien präsentieren Methoden, die den aktuellen Stand der Forschung im Bereich der Haplotyprekonstruktion repräsentieren, **QuasiRecomb** und **Haplo-Clique**. Bezüglich der Anwendung werden zwei weitere Studien vorgestellt, die die Rolle von viralen Quasispezies in schweinepestinfizierten Schweinen analysiert und sich mit der Umsetzbarkeit der Haplotyprekonstruktion in klinischen Anwendungen befasst. Jede Studie wird kurz motiviert und ins Gesamtthema eingeführt. Die originalen Veröffentlichungen sind in dieser Dissertation eingebunden und durch eine Kopfzeile erkenntlich gemacht. Eine abschließende Diskussion fasst die Kernpunkte dieser Dissertation zusammen, bietet kurz- und langfristige Ideen für die Weiterentwicklung und beschreibt Anwendungen außerhalb der Quasispeziesrekonstruktion.