Infection, Genetics and Evolution 19 (2013) 349–360

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

# HIV-1 fitness landscape models for indinavir treatment pressure using observed evolution in longitudinal sequence data are predictive for treatment failure

Raphael Z. Sangeda [a], Kristof Theys [a], Gertjan Beheydt [a], Soo-Yon Rhee [a,b], Koen Deforche [c], Jurgen Vercauteren [a], Pieter Libin [a,c], Stijn Imbrechts [a], Zehava Grossman [d], Ricardo J. Camacho [e,f], Kristel Van Laethem [a,p], Alejandro Pironti [g], Maurizio Zazzi [h], Anders Sönnerborg [i], Francesca Incardona [j], Andrea De Luca [k], Carlo Torti [l], Lidia Ruiz [m], David A.M.C. Van de Vijver [n], Robert W. Shafer [b], Bianca Bruzzone [o], Eric Van Wijngaerden [p], Anne-Mieke Vandamme [a,f,*], for the Virolab and EuResist projects

[a] Rega Institute for Medical Research, Department of Microbiology and Immunology, KU Leuven, 3000 Leuven, Belgium
[b] Division of Infectious Diseases, Stanford University, Stanford, CA 94305, USA
[c] Mybiodata, Biomedical IT Solutions, 3110 Rotselaar, Belgium
[d] Central Virology Lab, MOH, Sheba Medical Center, Ramat-Gan 52621, Israel
[e] Molecular Biology Laboratory, Centro Hospitalar de Lisboa Ocidenta, 1169-050 Lisbon, Portugal
[f] Centro de Malária e outras Doenças Tropicais, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, 1349-008 Lisbon, Portugal
[g] Max Planck Institute for Informatics, 80804 Munich, Germany
[h] Department of Molecular Biology, University of Siena, 53100 Siena, Italy
[i] Division of Infectious Diseases, Department of Medicine, Karolinska Institute, Sweden
[j] EuResist Network GEIE – Informapro s.r.l., Rome, Italy
[k] University Division of Infectious Diseases, Siena University Hospital, Siena, Italy
[l] Institute of Infectious and Tropical Diseases, University of Brescia, 25123 Brescia, Italy
[m] Retrovirology Laboratory IrsiCaixa Foundation, 08916 Badalona, Spain
[n] Department of Virology, Erasmus Medical Centre, Erasmus University, Rotterdam, The Netherlands
[o] Hygiene Unit, S. Martino University Hospital, Genova, Italy
[p] University Hospitals, KU Leuven, 3000 Leuven, Belgium

ARTICLE INFO

*Article history:*
Available online 21 March 2013

*Keywords:*
HIV-1 drug resistance
Antiretrovirals
Treatment response
Evolution
Bioinformatics

ABSTRACT

We previously modeled the *in vivo* evolution of human immunodeficiency virus-1 (HIV-1) under drug selective pressure from cross-sectional viral sequences. These fitness landscapes (FLs) were made by using first a Bayesian network (BN) to map epistatic substitutions, followed by scaling the fitness landscape based on an HIV evolution simulator trying to evolve the sequences from treatment naïve patients into sequences from patients failing treatment.

In this study, we compared four FLs trained with different sequence populations. Epistatic interactions were learned from three different cross-sectional BNs, trained with sequence from patients experienced with indinavir (BNT), all protease inhibitors (PIs) (BNP) or all PI except indinavir (BND). Scaling the fitness landscape was done using cross-sectional data from drug naïve and indinavir experienced patients (Fcross using BNT) and using longitudinal sequences from patients failing indinavir (FlongT using BNT, FlongP using BNP, FlongD using BND). Evaluation to predict the failing sequence and therapy outcome was performed on independent sequences of patients on indinavir. Parameters included estimated fitness (LogF), the number of generations (GF) or mutations (MF) to reach the fitness threshold (average fitness when a major resistance mutation appeared), the number of generations (GR) or mutations (MR) to reach a major resistance mutation and compared to genotypic susceptibility score (GSS) from Rega and HIVdb algorithms.

In pairwise FL comparisons we found significant correlation between fitness values for individual sequences, and this correlation improved after correcting for the subtype. Furthermore, FLs could predict the failing sequence under indinavir-containing combinations. At 12 and 48 weeks, all parameters from

* Corresponding author at: Rega Institute for Medical Research, Department of Microbiology and Immunology, KU Leuven, Minderbroedersstraat 10, B-3000 Leuven, Belgium. Tel.: +32 16 332160; fax: +32 16 332131.
*E-mail address:* annemie.vandamme@uzleuven.be (A.-M. Vandamme).


http://dx.doi.org/10.1016/j.meegid.2013.03.014

all FLs and indinavir GSS (both for Rega and HIVdb) were predictive of therapy outcome, except MR for FlongT and FlongP. The fitness landscapes have similar predictive power for treatment response under indinavir-containing regimen as standard rules-based algorithms, and additionally allow predicting genetic evolution under indinavir selective pressure.



## 1. Introduction

Prediction of human immunodeficiency virus-1 (HIV) drug resistance is useful to clinicians caring for HIV-1 patients, given the multitude of possible highly active antiretroviral therapy combinations using the more than 25 available antiretroviral drugs and taking into account the drug resistance profiles (Altmann et al., 2007; Shafer and Schapiro, 2008). Prospective controlled studies have shown that patients whose physicians have access to drug resistance data, particularly genotypic-resistance data, respond better to therapy than patients of physicians without such access (Van Laethem and Vandamme, 2006; Liu and Shafer, 2006). This kind of data has led several experts in North America and Europe to recommend drug resistance testing in the management of HIV-1 infected patients (Liu and Shafer, 2006; Hirsch et al., 2008; Vandamme et al., 2011).

Genotyping is preferentially used to detect resistance related mutations and the resistance pattern is then interpreted using several publicly available algorithms (Van Laethem et al., 2002; Van Laethem and Vandamme, 2006; Liu and Shafer, 2006; Vercauteren and Vandamme, 2006). The currently available interpretation systems are however subject to variability and discordances which may affect the choice of the proposed therapy and ultimately the treatment success. Another drawback of genotypic drug resistance testing is the difficulty to accurately predict the effect of complex interactions among the many mutations that contribute to drug resistance and inability to detect minor, but clinically relevant, drug-resistant variants in a patient's virus quasispecies (Liu and Shafer, 2006; Vercauteren and Vandamme, 2006; Shafer and Schapiro, 2008). Thus, it is important to update the current interpretation algorithms to correctly predict virological response to treatment.

We previously described a method that models mutational resistance pathways and estimates a fitness landscape (FL) based on *in vivo* virus genetic data and treatment information. The modeled FLs were made by using first a Bayesian network (BN) to map epistatic substitutions, followed by scaling the fitness landscape based on an HIV evolution simulator trying to evolve the sequences from treatment naïve patients into sequences from patients failing treatment. We showed that this fitness function significantly predicts resistance development and virological response (Deforche et al., 2008b; Theys et al., 2010). However, the current method requires a large amount of genotypic data to model a FL. Especially in the case of newly approved drugs that are initially administered in salvage regimens, viral sequences from patients treated with one of these drugs as the only drug in its drug class are rare or not yet available. Longitudinal sequence data obtained from patients treated with new drugs in salvage therapy offer a valuable solution. Since these sequences reduced the problem of inter-patient variability, they are more informative and therefore can overcome the need for more sequences.

This study aimed to develop longitudinal FLs and to compare the different designs to the conventional cross-sectional FL for the protease inhibitor (PI) indinavir with respect to their clinical applicability. Three strategies for the FL model design were evaluated with respect to how robust they can be used in the prediction of treatment outcome. The three strategies differ in how epistatic mutational interaction is learned in a first step from extensive cross sectional data available in an entire drug class, while using in a second step a limited set of longitudinal data for a potential new drug in the class to scale the fitness landscape. Indinavir was chosen because sufficient cross-sectional and longitudinal data of patients failing indinavir with resistance mutations in first and salvage therapy are available to both construct and evaluate the different models.

## 2. Materials and methods

### 2.1. Data and sequence populations

Two data sets of HIV-1 clinical data were used in this analysis. The first data set was primarily used to build the fitness models and was pooled from the Stanford HIV drug resistance Database (Kantor et al., 2001), Hospital Egas Moniz Lisbon, Portugal and the University Hospitals, Leuven, Belgium. The second data set provided independent data to evaluate the performance of the models in predicting resistance evolution and therapy response and was obtained from the European research consortiums EuResist and Virolab and from Israel's HIV Reference Laboratory (Fig. 1). Sequence data were locally stored in a RegaDB instance to facilitate data management and analysis (Libin et al., 2007). Different sources were chosen to investigate the robustness of the models with respect to the varying treatment strategies and the varying prevalence of HIV-1 subtypes. For each sequence, the subtype was determined using the Rega HIV-1 Subtyping Tool v2 (de Oliveira et al., 2005).

In total, 8 protease sequence populations were extracted from these two sources. From the first source, the following 5 training and 1 evaluation populations were derived. Training populations were as follows; the population PN represented 9116 sequences from PI-naive patients, the population PT represented 1181 sequences from patients treated with indinavir as their first PI, the population. PP represented 2883 sequences from patients treated with any PI and only included for each patient the last available sequence after PI experience, the population PD represented 1726 sequences from patients treated with any PI except indinavir and included only the last available sequenced per patient. The population PL represented pairs of 438 longitudinal sequences from 219 patients consisting of a baseline sequence before and a follow-up sequences after indinavir treatment. There was no overlap of sequences from this latter population with the other sequence populations. The evaluation population PE represented 3690 PI experienced sequences independent from those described above, which were used for comparison of absolute fitness values derived from various fitness landscapes.

From the second source, two non-overlapping evaluation populations were derived from indinavir treated patients, irrespective of previous PI experience. The population PV consisted of 626 longitudinal sequences paired from 313 patients and was used for evaluating predicted evolution against observed evolution. The population PC represented 320 baseline sequences from indinavir treated patients for whom a treatment change episode (TCE) accompanied by baseline genotype, baseline and follow-up viral load was available. TCEs were from patients receiving an indinavir-containing treatment regimen whether or not the patient was failing or