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

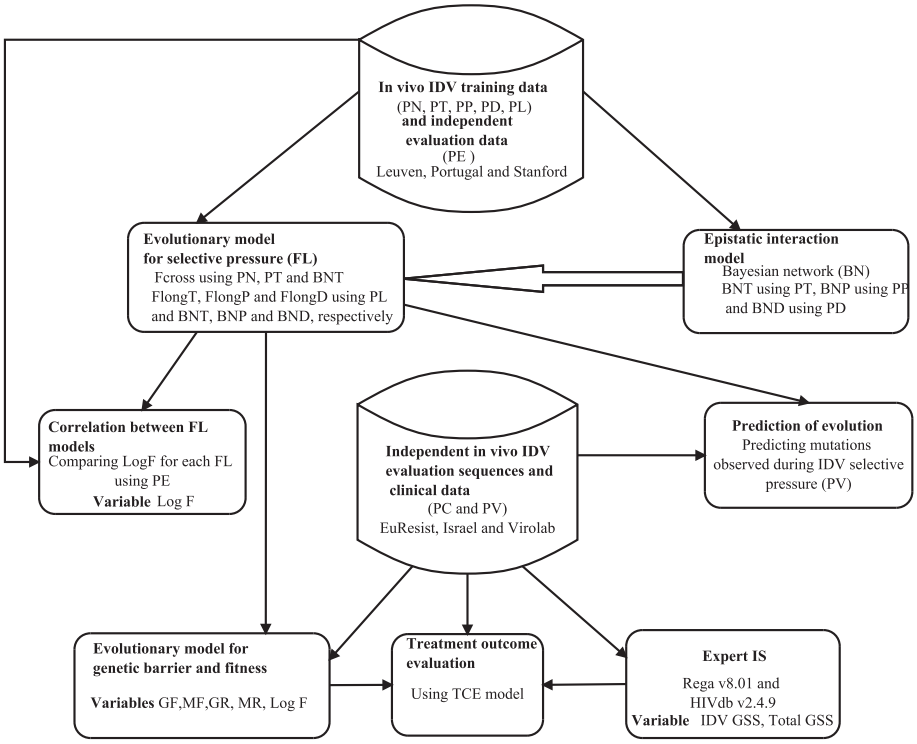

Fig. 1. Schema used in modeling fitness landscapes and its evaluation. Training data used to make epistatic interaction models using Bayesian Network Learning (BN) and to scale the Fitness Landscape (FL). Three types of BN (BNT, BNP and BND) were constructed trained with sequence populations PT, PP and PD respectively, which differ in protease inhibitor (PI) exposure. PT sequences were from patients with indinavir experience as the first PI, PP from patients experienced with any PI and PD from patients with any PI except indinavir. The population PN consisted of sequences naive to PI. To scale the fitness landscape Fcross, the evolution simulator was used to train Fcross such that evolution of PN over Fcross ended up in a population of sequences resembling PT. Population PL consisted of pairs of each a baseline and a follow-up sequence from the same patient before and after experience with indinavir, and these were used to train the scaling of all longitudinal FLs, such that baseline sequences evolved over the FL resulted in simulated follow-up sequences closely resembling the observed follow-up sequences after evolution over the FLs. Thus, four fitness functions Fcross, FlongT (using BNT), FlongP (using BNP) and FlongD (using BND) were scaled by simulating evolution of naïve or baseline sequence to treated sequence. Population PE, contained a set of sequences independent from the training set to evaluate the FL. Population PV, contained a set of longitudinal pairs of sequences independent of the training set to evaluate the prediction of evolution using the FL. Sequences of patients (PC) with treatment change episode (TCE) at week 12 and week 48 were used to derive selective pressure and genetic barrier parameters from each FL and its predictive power on virological failure determined and compared with that of Rega and HIVdb interpretation systems (IS). Selective pressure parameter: logarithm of absolute fitness (LogF) and genetic barrier parameters: number of viral generations to fitness threshold (GF), number of viral mutations to fitness threshold (MF), number of viral generations (GR) to a major drug resistance mutation (DRM) and number of viral mutations to major DRM (MR). These parameters are computed for each baseline sequence using the evaluation data over the respective FLs. For IS, genotypic susceptibility score (GSS) is calculated for each baseline sequence assigning a score for each drug. Summing up the individual drug scores gives total GSS for the background regimen.

had received indinavir before. Each TCE contained baseline parameters taken as close as possible to week 0. Both baseline genotype and viral load were taken not earlier than 12 weeks before or later than the start of TCE and a follow-up viral load at 12 (the closest value falling between 4 and 16, median 8.9 with interquartile range 5.7–12.6) weeks and/or 48 (16–52, median 40 with interquartile range 28.3–46.4) weeks (Supplementary Fig. 1). For endpoint one, endpoint two and both endpoints together, data from 281, 247 and 208 patients were available respectively (Supplementary Fig. 2).

2.2. Learning qualitative epistatic interactions between protease amino acids

Bayesian Network (BN) learning was performed in order to map the epistatic dependencies between mutations and/or polymorphisms due to selective pressure from a particular drug as described previously (Deforche et al., 2006; Deforche et al., 2007) and Fig. 1. BN learning used probability computations to learn the associations between mutations and the resulting BN can be represented graphically by nodes denoting amino acids and arcs (arrows) denoting direct dependencies between the joined nodes. BN learning was done using the B-course software (Myllymäki et al., 2002) and adapted by (Deforche et al., 2006), scoring models by maximizing the posterior probability of the model. Boolean variables were used to represent the presence or absence of amino acid mutations. Networks were learned including all polymorphic amino acids with prevalence greater than 1% in the sequence population of untreated patients. Mutations related to indinavir or PI treatment were selected after testing for significant independence from treatment using Mantel–Haenszel χ^2 test while correcting for multiple tests using Benjamini & Hochberg with a False Discovery Rate (FDR) of 0.05.

In total, three BNs were learned using the three training sequence populations from treatment experienced patients. The network BNT was learned using the population PT to map interactions between indinavir mutations, the network BNP was learned using the population PP to map interactions between PI mutations and the network BND was learned using the population PD to map interactions between PI mutations not selected by indinavir. All BNs are available upon request.

2.3. Construction and scaling of the fitness landscape models

HIV-1 fitness functions were estimated as described previously (Deforche et al., 2008a). The general method used to derive the fitness landscapes (FLs) is depicted in Fig. 1. The first step was modeling the qualitative epistatic interactions by the respective BNs as described above, which served as a template for the fitness

function structure. In a second step, the fitness landscape is scaled by simulating the evolution of viral sequences from naïve patients into sequences from treated patients. In an iterative procedure, sequences from naive patients (PN) are evolved over the FL, and simulated treated sequences are compared with the real observed sequences. Parameters of the fitness landscapes are incrementally improved to converge to the optimal fitness function. After convergence, the obtained fitness function provides a model for HIV-1 evolution under indinavir selective pressure. We therefore call these fitness functions, 'fitness landscapes', where the 'flat dimensions' of the landscape represent the genetic variability and the epistatic interactions between amino acids as described by the BN, and the scaling, the 'height of the peaks' or 'fitness' represents the predicted fitness of a particular sequence under the modeled selective pressure. A FL is normalized relative to a fitness value of 1 for reference sequence HIV-1 HXB2.

Here, we estimated one cross-sectional and three longitudinal indinavir fitness functions, using epistatic interactions that were learned by BNT, BNP and BND networks.

The cross-sectional FL model Fcross was estimated using cross-sectional sequences with indinavir experience by first deriving the network structure of BNT and second modeling the evolution of PN into PT. Given that the naive population (PN) included more sequences compared to the treated population (PT), a naive dataset was created by sampling sequences from PN as described before (Deforche et al., 2008a). This sampling was guided by a Neighbour Joining phylogenetic tree built with PAUP on the training data, and assigned more weight to sequences from the naïve population that were epidemiologically linked to the treated population. This procedure allowed for selection of sequences from naïve patients that more closely resemble the epidemiology background of the sequences from treated patients, thus learning amino acid changes that are mainly a consequence of evolution under drug selective pressure, thereby correcting for different epidemiological dependencies.

The longitudinal FLs (FlongT, FlongP and FlongD) were estimated using the same population (PL) of longitudinal sequence pairs of patients on indinavir treatment to scale the fitness landscape in the second step. No sampling procedure to correct for epidemiological bias was needed because sequence pairs were derived from the same patient. The respective longitudinal models differed in the sequence population from which the epistatic interactions were learned in the first step. The model FlongT used the network structure of BNT, the model FlongP used the network structure of the BNP and the model FlongD used the network structure of BND.

2.4. Calculating parameters for genotypic prediction

2.4.1. From longitudinal and cross-sectional fitness landscapes

For each sequence and per fitness landscape, one parameter denoting viral fitness and four parameters related to the genetic barrier to resistance were computed (Fig. 1). Predicted viral fitness taken as the common logarithm scale (LogF), was considered as the quantification of drug susceptibility. The genetic barrier represents the evolutionary distance required for the virus to become a drug resistant virus. The genetic barrier to indinavir resistance was calculated by simulating the evolution over a FL until it was considered resistant, and defined as the average time of 100 simulations. The criterion for indinavir resistance was the selection of a Major Resistance Mutations (MRM) 46I, 46L, 82A, 82F, 82T or 84V (Johnson et al., 2011). Measures of the genetic barrier were the number of generations (GR) and number of mutations (MR) to the appearance of any indinavir-MRM. Another measure of genetic barrier to resistance was the number of generations and the number of mutations during simulated evolution to reach a fitness threshold (GF and MF respectively). The fitness threshold was defined as the average fitness after the appearance of a major resistance mutation during the 100 simulations and over all sequences.

2.4.2. GSS from Rega V8.0.1 and HIVdb V2.4.9

Genotypic Susceptibility Scores (GSS) were calculated using two publicly available genotypic resistance interpretation system (IS) (Rega v8.0.1 and HIVdb v4.2.9). According to Rega v8.0.1 (http://regaweb.med.kuleuven.be/software/rega_algorithm), a baseline sequence was assigned a weighted score for each drug based on the 3 level categories. Susceptible (S): GSS = 1 for all drugs except 1.5 for a boosted PI; intermediate resistant (I): GSS = 0.5 (unboosted PI, nucleoside reverse transcriptase inhibitors (NRTIs) and nucleotide reverse transcriptase inhibitors (NtRTIs) and etravirine) or 0.75 (boosted PI) or 0.25 (Non-nucleoside reverse transcriptase inhibitors (NNRTIs) except etravirine or entry inhibitors (EI)); otherwise assigned resistant (R): GSS = 0. According to HIVdb (http://hivdb.stanford.edu/DR/asi/index.html), a baseline genotype was assigned a GSS according one of the five levels that are included in the algorithm. The score was 1 for susceptible, 0.75 for potential low-level resistance, 0.5 for low-level resistance, 0.25 for intermediate resistance and 0 for high resistance. The drug zalcitabine (DDC) is not featured in the current Rega and HIVdb versions since it was removed from the market due to safety concerns. However, since this drug was used by patients in our evaluation dataset the scores from the last Rega algorithm (Rega v6.4.1) that included DDC were used. Similarly, the score for the drug amprenavir, also no longer supported by the current algorithms, was substituted by scores of its pro-drug fosamprenavir (Figs. 1–3).

2.5. Evaluating the fitness landscapes

The performance of the four fitness models was evaluated in different validation experiments performed on three evaluation sequence populations that were not included in the training data. These experiments encompassed the comparison of estimated fitness, predicting evolution and prediction of virological outcome.

2.5.1. Pairwise comparison of estimated fitness values

The three longitudinal and one cross-sectional FLs were applied to the PI experienced sequence population PE to compute viral fitness (LogF). Pairwise comparisons of fitness values were done by a linear regression model and corrected for the effect of subtype. Goodness of fit for the FL models was assessed by the squared correlation coefficient (R^2) and further explored by scatter plots. Descriptive statistics was used to analyze results from each of the fitness function.

2.5.2. Evaluating evolution predicted by the fitness landscape

The evaluation population PV of longitudinal sequences was used to predict mutation evolution. The population PV consisted of pairs of 626 longitudinal sequences from 313 patients. The ability to predict observed evolution during indinavir treatment, according to models Fcross, FlongT, FlongP and FlongD was explored using the baseline sequence from the dataset PV. Correlation between predicted evolution and the observed evolution in the follow-up sequence from PV was done using a linear model correcting for multiple testing.

2.5.3. Prediction of treatment outcome in short and long term

A third evaluation was done based on the parameters LogF, GR, MT, GR and GF of the cross-sectional and longitudinal FLs compared to Rega IS and HIVdb IS GSS parameters (Fig. 1) for their predictive performance on a risk of virological failure at short (week 12) and long term (week 48) using logistic regression. A baseline

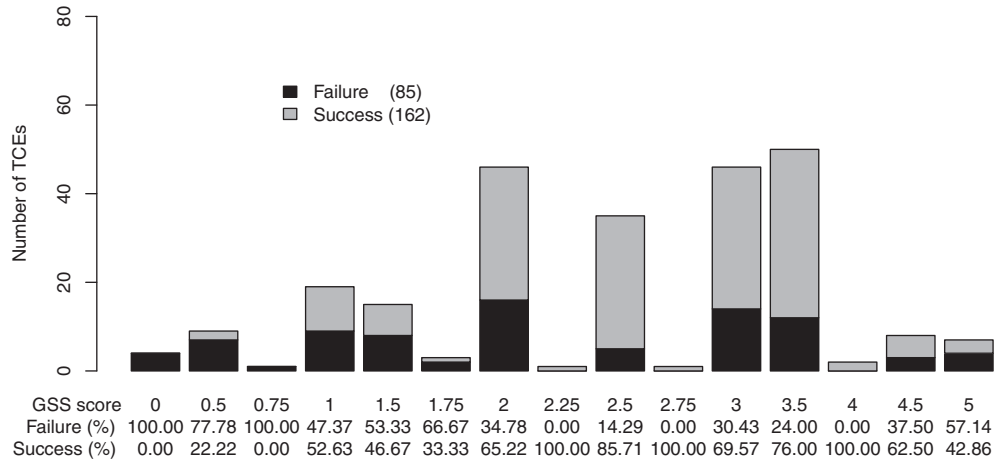

Fig. 2. Description of the total regimen genotypic susceptibility score as determined by the Rega v8.0.1 weighted algorithm for all treatment change episodes (TCEs) used for prediction of week 48 virological outcome. The distribution was similar to that of week 12 TCEs (Supplementary Fig. 9).

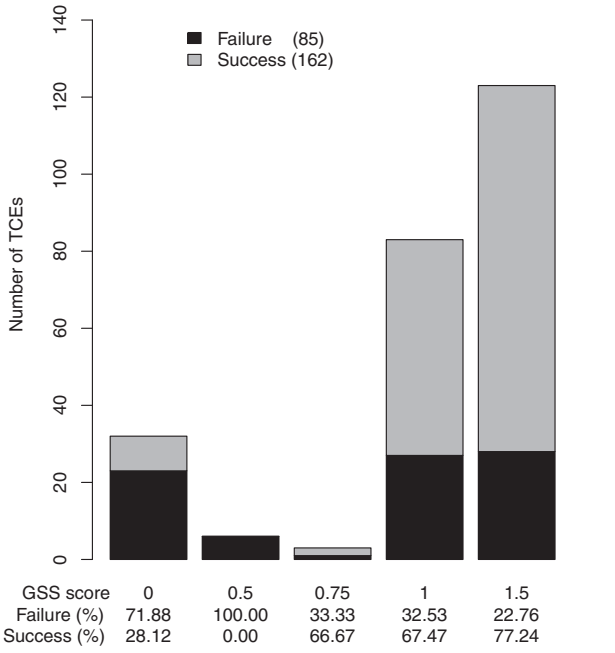

Fig. 3. Description of the indinavir genotypic susceptibility score (GSS) as determined by the Rega v8.0.1 weighted algorithm for all treatment change episodes (TCEs) used for prediction of week 48 virological outcome. A score of 0, 0.5, 0.75, 1 and 1.5 is assigned for resistant, intermediate, intermediate-boosted indinavir, susceptible and susceptible-boosted, respectively. The distribution was similar for the week 12 TCEs (Supplementary Fig. 10).

genotype from the treatment change episodes (sequence population PC) was used to compute these parameters. For both time points, virological success was defined as achievement of a viral load less than 500 copies/ml or 2 log drop versus baseline viral load (Zazzi et al., 2011). In addition, any therapy change under a detectable viral load was considered a treatment failure. Logistic regression models were corrected for the possible confounders; baseline viral load, previous PI exposure, the time lag between the start of therapy and the follow-up viral load and GSS of the background regimen excluding the indinavir GSS.

2.5.4. Performance of logistic regression models

In order to evaluate the accuracy of the prediction of short and long term virological failure, performance of each FL parameter was estimated using a 10-fold receiver operating characteristic (ROC) curve analysis wherein the area under the curve (AUC) was used as performance measurement. The FL parameters were compared to the Rega and HIVdb indinavir GSS after correcting for GSS of the backbone regimen.

2.5.5. Survival analysis

A final analysis constituted a Kaplan–Meier analysis of the TCEs to test how GSS, LogF, GR, MT, GR and GF could be useful in predicting time to viral suppression. The proportions of TCEs achieving treatment success were plotted against time to suppression of viral load. For this purpose, both the GSS from the Rega and HIVdb were grouped into three; 0–1, >1–2 and >2. For the genotypic parameters from the FLs, the values were grouped per quartile and the difference in median survival time tested. Further, the Cox proportional hazards model was applied to adjust for other covariates including baseline viral load and the time lag between the start of therapy and the follow-up viral load.

3. Results

3.1. Sequence data and subtype distribution

In total 17,539 viral sequences were used in this study for model training or evaluation. More than half of the sequence dataset belonged to HIV-1 subtype B (74%), followed by subtype G (11%). The detailed subtype distribution of all eight sequence populations is shown in Supplementary Fig. 3. When testing for equal population proportions, subtype B distribution was not significantly different (p-value = 0.08) in the training sequence data sets used for each of the Bayesian Networks (PT, PP and PD) but significantly different to the naïve population (PN) (p-value = 0.04). The sequence distributions of the evaluation datasets (PE, PC and PV) were however significantly different from the training datasets (p-value < 0.05).

3.2. Epistatic interactions learned from the Bayesian Networks

The networks BNT, BNP and BND mapped 89 mutations at 48 sites, 90 mutations at 47 sites and 90 mutations at 47 sites, respectively. A BN was not constructed from the longitudinal sequence data alone as in this population mutations were not sufficiently abundant to model all the epistatic interactions of different amino acid positions in protease. Mutations 20T/V/M, 30N, 33F, 35N/G, 36V, 48V, 53L, 69RI/Y, 71I, 74A, 88D/S and 92R/K were present in BNP and BND but absent in BNT.

Table 1
Coefficient of determination (R^2 value) from the comparison of fitness values for each of the sequences from population PE. Per sequence, fitness values calculated according to the respective fitness models were compared. R^2 values were determined using a linear regression model. All correlations were significant with p-value of less than 0.01. Introducing subtype as correction factor improves the R^2 value considerably.

Model	Fcross vs. FlongT	Fcross vs. FlongP	Fcross vs. FlongD	FlongT vs. FlongP	FlongT vs. FlongD	FlongP vs. FlongD
Before subtype correction	0.55	0.42	0.69	0.88	0.74	0.73
After subtype correction	0.77	0.68	0.78	0.89	0.81	0.78

3.3. Fitness landscape models and their comparison

The fitness values, LogF, were obtained by applying Fcross, FlongT, FlongP and FlongD to each sequence in the PE dataset. The mean of LogF for FlongT, FlongP and FlongD of the PE dataset was not significantly different. However, LogF of the cross-sectional fitness function had significantly different mean compared to the longitudinal fitness functions (Supplementary Fig. 4). We also found a significant correlation between fitness derived from the fitness function learned entirely from cross-sectional data, Fcross, with the fitness according to the three landscapes learned from indinavir longitudinal sequences, FlongT, FlongP and FlongD (Table 1). Surprisingly, the best correlation was found between Fcross and FlongD, which used the epistatic interactions learned by excluding sequences from indinavir experienced patients. The pairwise correlation coefficients (R^2) ranged from 0.42 to 0.88.

Correlation values increased when the subtype factor was taken into account in the linear regression model (R^2 values ranging from 0.68 to 0.89). All pairwise comparisons, with and without subtype correction, are shown in Fig. 4 and Supplementary Figs. 5–7. Even after correcting for subtype, there still appeared some banding patterns in the correlation curves. This was more evident comparing Fcross versus FlongT and FlongP than in the comparison among the other FLs. The difference could be explained by resistance mutations which were present in different proportion in the sequence populations used for the different landscapes (Table 2).

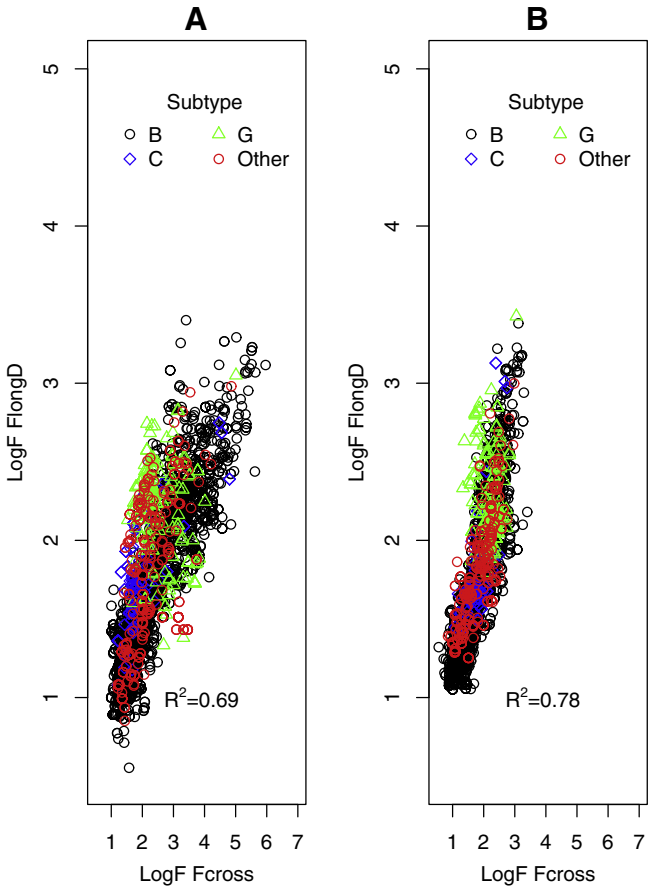

Fig. 4. Evaluation of linear relationship between predicted fitness of the evaluation sequence population PE, comparing Fcross with FlongD. Before (Panel A) and after (Panel B) correction for subtype as indicated.

3.4. Evaluating evolution predicted by fitness landscape

The ability of each FL to predict viral evolution under indinavir selective pressure was evaluated using an independent dataset of longitudinal sequences (PV) sampled before and after treatment with indinavir. Predicted evolution according to Fcross, FlongT FlongP and FlongD correlated significantly with the observed evolution during indinavir treatment for 22, 25, 16 and 18 mutations, respectively (Table 3). Most major mutations included in the IAS-USA list were significantly predicted for all FLs. Mutations 10V, 46L, 71T, 73S and 82F for Fcross, 71T and 73S for FlongT, 46L and 71T and 73S for FlongP and 10V by FlongD were not significantly predicted although the predictions were significant before correcting the p-value for multiple testing. Negative correlations were not found for any mutation in Fcross. For a few mutations in longitudinal landscapes, negative correlations were found, meaning that these mutations were more often predicted than observed. This was the case for mutations; 55R in FlongT; 24I, 55R and 71V in FlongP and 54V, 60E, 71V, 73S, 74S, 84V and 90M in FlongD. Most minor IAS mutations were also significantly predicted, and several other mutations, not in the IAS-USA list were also significantly predicted.

3.5. Evaluation of fitness landscapes with clinical outcome

3.5.1. Characteristics of TCEs used

Descriptive analysis of used TCEs is shown in Table 4. The majority of patients were treated with a combination of indinavir, stavudine and didanosine (10%) or indinavir, lamivudine and

Table 2
Mutations which were more significantly prevalent in the treated longitudinal sequences used to scale the fitness landscapes FlongT, FlongP and FlongD, than the sequence used to scale Fcross.

Position	% With mutation		
	2nd Sequence of PL	PT	p-value
10I	46.12	23.2	<0.01
46L	16.44	6.6	<0.01
46I	30.14	13.12	0.000
54V	23.74	11.6	<0.01
55R	5.02	1.95	0.02
71T	12.79	7.28	0.02
71V	26.03	12.45	<0.01
73S	8.68	2.03	<0.01
74S	8.68	3.39	<0.01
77I	35.16	24.64	0.02
82A	28.31	14.14	<0.01
84V	10.96	3.22	<0.01
85V	9.13	2.2	<0.01
90M	28.77	10.16	<0.01
93L	44.75	33.36	0.03

Table 3
Mutations for which predicted variation correlated significantly with observed evolution during indinavir treatment (correcting for false discovery rate of 0.05), in the evaluation sequence pairs (PV) using Fcross, FlongT, FlongP and FlongD fitness functions. The number (*N*) of baseline sequences without the mutation (*M*) of which *n* developed the mutation during treatment. P: *p*-values for correlation between predicted probability for selection of the mutation and observed selection. Indinavir IAS-USA mutations are shown in bold; the major indinavir IAS-USA mutations are in bold italics. NS: mutation not significantly predicted and -; mutation not present in this particular fitness landscape.

M	*N*	*N*	Fcross *P*	FlongT *P*	FlongP *P*	FlongD *P*
10F	309	9	NS	1.31E-04	NS	7.21E-03
10I	**251**	**67**	**9.23E-16**	**2.21E-11**	**8.37E-09**	**1.99E-08**
10V	**302**	**13**	**NS**	**5.99E-04**	**1.47E-03**	**NS**
13V	262	24	1.60E-02	NS	NS	NS
15V	251	21	9.52E-05	1.06E-02	6.57E-03	NS
20I	303	18	1.35E-02	1.98E-10	1.66E-09	3.16E-07
20R	**297**	**19**	**1.36E-03**	**2.73E-09**	**4.20E-07**	**2.95E-06**
24I	**305**	**17**	**6.78E-07**	**1.02E-02**	**2.10E-03**	–
33I	312	5	NS	NS	NS	2.41E-03
35D	247	18	1.09E-02	NS	4.21E-03	NS
36L	308	2	1.57E-04	NS	NS	NS
37D	246	17	2.37E-04	3.34E-03	7.52E-05	NS
37E	305	5	NS	7.42E-07	NS	NS
43T	312	15	1.35E-02	NS	–	–
46I	***281***	***63***	***2.42E-09***	***7.52E-07***	***1.14E-11***	***6.64E-10***
46L	**294**	**23**	**NS**	**5.01E-02**	**NS**	**1.05E-03**
54V	**288**	**46**	**7.56E-16**	**3.68E-02**	**9.64E-09**	**2.02E-02**
55R	307	20	4.22E-07	1.47E-06	4.97E-05	6.09E-07
57K	285	11	–	2.40E-05	4.73E-05	2.02E-02
58E	305	12	1.03E-02	NS	3.48E-05	5.34E-02
60E	284	15	NS	NS	NS	3.34E-02
62V	236	54	4.91E-08	1.01E-09	3.43E-10	6.64E-10
63H	310	2	4.02E-03	NS	NS	NS
63P	107	34	4.32E-03	4.55E-03	NS	7.21E-03
63T	291	3	NS	6.80E-03	3.28E-03	7.21E-03
64V	264	20	NS	NS	NS	NS
65D	303	4	NS	NS	NS	3.15E-02
66F	310	3	NS	4.72E-02	NS	NS
69Q	308	4	NS	5.39E-02	NS	NS
71T	**282**	**15**	**NS**	**NS**	**NS**	**1.20E-02**
71V	**266**	**59**	**9.52E-16**	**3.24E-17**	**8.62E-05**	**9.23E-09**
72L	304	18	1.79E-04	3.00E-03	–	–
72T	300	5	NS	NS	NS	2.63E-02
73S	**293**	**36**	**NS**	**NS**	**NS**	**3.15E-02**
74S	311	13	NS	NS	NS	3.15E-02
82A	***278***	***58***	***2.66E-08***	***7.00E-09***	***2.04E-06***	***4.53E-06***
82F	***308***	***4***	***NS***	***6.02E-04***	–	–
84V	***297***	***21***	***8.65E-07***	***1.57E-04***	***2.25E-05***	***6.77E-04***
85V	307	15	NS	NS	NS	4.53E-06
90M	**255**	**73**	**2.36E-14**	**1.22E-05**	**2.25E-05**	**1.08E-09**
93L	216	32	NS	5.01E-02	NS	NS

zidovudine (9%) at the first time-point (Supplementary Fig. 8). These therapies were administered in the period 1996–2006.

The frequency of the regimen-specific GSS made by combining GSS of the backbone regimen with that of indinavir is shown in Fig. 2 and Supplementary Fig. 9. The majority of patients had a regimen-specific GSS greater than 2 and responded well to therapy. Baseline sequences were mainly predicted as susceptible to indinavir and were responders (Fig. 3 and Supplementary Fig. 10). There was a significant but weak correlation between regimen-specific GSS and the indinavir GSS ($R^2 = 0.43$; p-value < 0.01 and $R^2 = 0.33$; p-value < 0.01 for Rega and HIVdb, respectively).

3.5.2. Prediction of virological failure in the 12 weeks analysis

For the FL parameters used, the median score (Interquartile range) for the baseline sequences of the week12 dataset are shown in Table 5.

A logistic regression model was used to evaluate the performance of each of these FL parameters and the indinavir GSS in predicting therapy failure at week 12. A higher Rega indinavir and HIVdb indinavir GSS at baseline was associated with lower odds of virological failure at week 12, i.e. 0.44 (0.27–0.70, p-value < 0.01) and 0.24 (0.11–0.53, p-value < 0.01), respectively. Concerning the FL parameters, for all models, a higher fitness was associated with higher odds of virological failure (Table 6). A higher genetic barrier to resistance, measured as GR, GF, MF and MR were significantly associated with lower odds of virological failure in all FLs (except MR from FlongT).

3.5.3. Prediction of virological failure in the 48 weeks analysis

Similarly, for patients who had a follow-up viral load at endpoint two, we analyzed the association between the parameters above with risk of therapy failure at week 48.

A unit increase in both Rega indinavir GSS and HIVdb indinavir GSS was associated with lower probability of virological failure, with an odds ratio of 0.22 (0.11–0.39 p-value < 0.01) and 0.12 (0.05–0.32 p-value < 0.01), respectively.

For the FL parameters, a unit change of LogF, GF, MF and GR were predictive of virological failure for all models (Table 6). With these models a higher fitness was associated with higher odds of virological failure, and an increase in genetic barrier parameters GF, MF and GR were associated with lower odds ratio of virological failure (Table 6).

In case of genetic barrier to resistance measured as MR, they were only predictive in the models Fcross, FlongD (Table 6). MR from FlongT and FlongP could not significantly predict virological failure, although the trend was there (p-value = 0.07).

3.6. Performance of logistic regression models

For week 48, the mean of the AUC of 10-fold cross-validation for each FL parameter in the four FLs model, and indinavir GSS for Rega

Table 4
Description of TCE parameters used for evaluation of fitness landscape models.

Parameter	Median (interquartile range)	
	Week 12	Week 48
Baseline total GSS (Rega)	2.50 (2.00–3.50)	2.50 (2.00–3.50)
Baseline IDV GSS (Rega)	1.50 (1.00–1.50)	1.00 (1.00–1.50)
Baseline total GSS (HIVdb)	2.00 (1.50–3.00)	2.25 (1.50–3.00)
Baseline IDV GSS (HIVdb)	1.00 (0.75–1.00)	1.00 (1.00–1.00)
Baseline VL	23,110 (4239.00–129,000) copies/ml	23,110 (3,395–112,216) copies/ml
Follow-up VL	400 (80–2,169) copies/ml	100 (50–6,965) copies/ml
Time lag between start of TCE and follow-up VL	8.86 (5.71–12.57) weeks	40.00 (28.29–46.43) weeks
Follow-up duration	91 (64–107) days	300.00 (231–353) days
Therapy duration	315 (144–573) days	413 (276–728) days
Therapy year	2000 (1999–2002)	2000 (1999–2002)

GSS, genotypic susceptibility score; IDV, indinavir; VL, viral load; TCE, treatment change episode.

Table 5
Median values (interquartile range (IQR)) for fitness landscape (FL) parameters derived from treatment change episodes used to predict virological failure at 12 and 48 weeks.

Parameter	FL Week	Fcross Median (IQR)	FlongT Median (IQR)	FlongP Median (IQR)	FlongD Median (IQR)
Log F	12	1.66 (1.44–2.11)	1.24 (1.12–1.48)	1.28 (1.15–1.52)	1.37 (1.18–1.73)
	48	1.63 (1.43–2.03)	1.23 (1.11–1.45)	1.27 (1.15–1.50)	1.36 (1.18–1.68)
GR	12	163.22 (67.98–201.10)	191.8 (104.8–245.5)	183.1 (142.60–228.70)	170.9 (137.2–198.7)
	48	166.75 (78.23–201.26)	199.9 (109.20–187.7)	181.1 (140.1–228.7)	170.60 (141.80–197.40)
MR	12	3.86 (1.92–4.62)	3.01 (1.96–3.78)	2.93 (2.37–3.68)	3.22 (2.64–3.81)
	48	3.92 (2.02–4.55)	2.99 (1.93–3.70)	2.89 (2.35–3.63)	3.04 (2.71–3.76)
GF	12	166.01 (74.96–209.37)	559.8 (372.7–673.5)	587.40 (406.40–695.30)	390.4 (257.0–468.70)
	48	170.07 (91.04–209.56)	559.2 (394.70–686.70)	592.50 (430.0–696.70)	405.30 (281.70–475.60)
MF	12	4.06 (2.27–5.08)	8.90 (6.23–10.32)	9.12 (6.64–10.43)	7.5 (5.01–8.80)
	48	4.19 (2.58–5.11)	8.99 (6.63–10.47)	9.33 (6.88–10.49)	7.68 (5.38–8.84)

Table 6
Logistic regression analysis for predicting risk of virological failure at week 12 and 48 using each parameter of the fitness landscapes (FL) separately. The odds ratio (OR) and its 95% confidence interval (CI) are calculated per unit increase in LogF, MR and MF or per 100 generations in case of GR and GF. The model was adjusted for the confounders, baseline viral load, time lag between the start of treatment change episode, previous protease inhibitor exposure and the time of follow-up viral load and for the total genotypic susceptibility score of the background regimen.

FL		Fcross		FlongT		FlongP		FlongD	
Parameter	Week	OR (95% CI)	*p*-value	OR (95% CI)	*p*-value	OR (95% CI)	*p*-value	OR (95% CI)	*p*-value
Log F	12	1.74 (1.12–2.69)	0.01	2.13 (1.07–4.35)	0.03	2.09 (1.11–4.12)	0.03	2.02 (1.08–3.77)	0.03
	48	2.63 (1.53–4.51)	<0.01	3.77 (1.73–8.64)	<0.01	4.16 (1.89–9.86)	<0.01	3.18 (1.52–6.64)	<0.01
GR	12	0.60 (0.44–0.83)	<0.01	0.78 (0.62–0.98)	0.04	0.67 (0.50–0.88)	<0.01	0.68 (0.49–0.95)	0.02
	48	0.58 (0.40–0.83)	<0.01	0.77 (0.59–0.99)	0.05	0.73 (0.54–0.98)	0.04	0.65 (0.45–0.94)	0.02
MR	12	0.79 (0.69–0.92)	<0.01	0.87 (0.75–1.01)	0.07	0.82 (0.69–0.96)	0.02	0.82 (0.69–0.98)	0.03
	48	0.77 (0.66–0.91)	<0.01	0.86 (0.72–1.01)	0.07	0.84 (0.71- 1.01)	0.07	0.80 (0.67–0.97)	0.02
GF	12	0.82 (0.71–0.94)	<0.01	0.86 (0.76–0.97)	0.01	0.85 (0.75–0.96)	<0.01	0.78 (0.64–0.94)	0.01
	48	0.56 (0.39–0.80)	<0.01	0.82 (0.72–0.93)	<0.01	0.80 (0.70–0.92)	<0.01	0.69 (0.56–0.87)	<0.01
MF	12	0.82 (0.71- 0.94)	<0.01	0.89 (0.82–0.97)	0.01	0.89 (0.82–0.97)	<0.01	0.88 (0.79–0.97)	0.01
	48	0.77 (0.66–0.90)	<0.01	0.86 (0.78–0.95)	<0.01	0.85 (0.78–0.94)	<0.01	0.82 (0.72–0.92)	0.02

Table 7
Area under the curve (AUC) for prediction models from the Receiver Operating Characteristic (ROC) analysis for week 12 and 48 (a) the expert systems and (b) genotypic parameters of each fitness landscape (FL). Wilcoxon rank test was done between Rega IDV GSS with HIVdb IDV GSS (A) or Rega IDV GSS with Log F, GR, MR, GF and MF. All models were not significantly different in their AUC. IDV, indinavir; GSS, genotypic susceptibility score).

A

Algorithm	Week	AUC for IDV GSS
Rega	12	0.67
	48	0.71
HIVdb	12	0.67
	48	0.69

B

	Week	Fcross	FlongT	FlongP	FlongD
LogF	12	0.65	0.67	0.66	0.66
	48	0.65	0.67	0.66	0.66
GR	12	0.66	0.65	0.6	0.65
	48	0.65	0.66	0.62	0.65
MR	12	0.66	0.64	0.63	0.65
	48	0.65	0.66	0.63	0.65
GF	12	0.67	0.66	0.66	0.66
	48	0.65	0.65	0.66	0.66
MF	12	0.68	0.65	0.65	0.66
	48	0.65	0.65	0.66	0.65

and HIVdb are shown in Table 7 and ROC curve in Supplementary Figs. 11 and 12. All models were predictive, but neither model was better than the other in predicting treatment response in both week 12 (data not shown) and week 48 datasets.

3.7. Survival analysis

To further illustrate the genotypic characteristics of the data used for predicting virological outcome, we used Kaplan–Meier curves (Fig. 5). The median time to reach viral-suppression below 500 copies/ml or a decrease of 2 or more logs of viral load during follow-up was 10 weeks. The median survival time of the three GSS groups according to Rega indicated that patients with higher baseline GSS reached an undetectable viral load significantly earlier than those with a lower GSS (Table 8). A similar significant trend was shown by the interquartile groups of the GR and MR FL parameters indicating that the higher the remaining genetic barrier the earlier undetectable viral load was reached. For the GF, the less fit viruses required a shorter time to be suppressed. The discriminative power of the interquartile groups of LogF and MF with *p*-value 0.08 and 0.09, respectively was not significant.

The overall models remained significant after adjusting for covariates using the Cox proportional hazard model, although the individual FL parameters became insignificant when combined in the model (data not shown).

4. Discussion

In this study, novel longitudinal fitness landscapes (FLs) were compared to the more conventional cross-sectional FL for patient receiving indinavir-containing combination treatment, and found to have the power to predict both the failing genotype and virological outcome. The technique of Bayesian Network (BN) Learning was used to investigate which mutations and polymorphisms appear together more than due to chance and epidemiology alone, a characteristic assumed to reflect epistatic interactions between the amino acids at these different positions of the HIV protein in question. Consequently, the BN denotes the qualitative representation of epistatic interactions mapped from statistical associations between amino acids. Longitudinal sequences from patients failing indinavir were then used to scale the fitness landscape. The fitness function representing the fitness landscape allows obtaining a quantitative estimate of the fitness of a particular sequence. Thus,

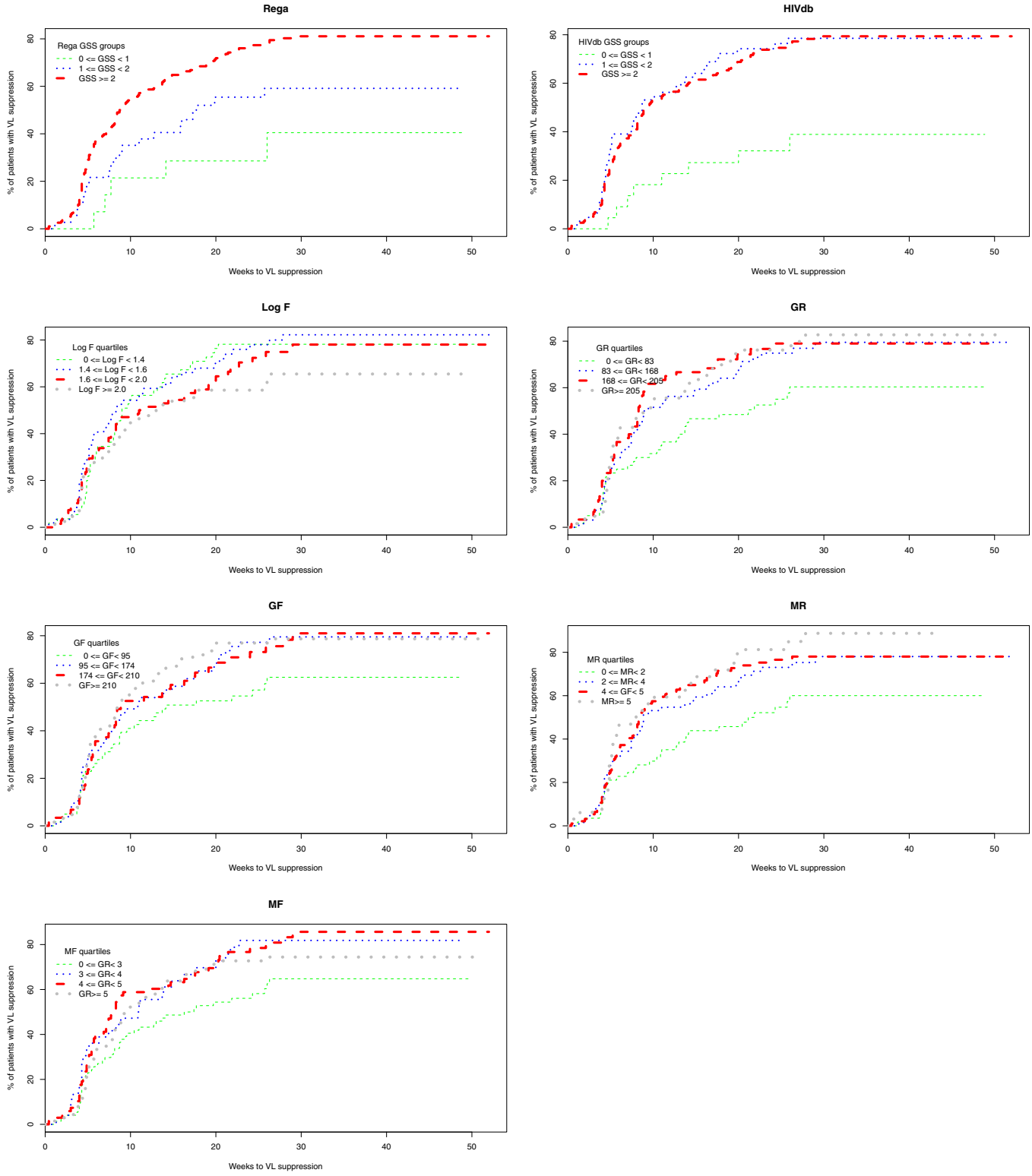

Fig. 5. Kaplan–Meier curve to show associations between the three indinavir genotypic susceptibility score (GSS) groups from Rega and HIVdb derived indinavir GSS or the quartile-groups of the indinavir cross-sectional FL parameters (LogF, GR, MR, GF and MF) derived from the treatment change episodes at 48 weeks, versus the time to viral-suppression below 500 copies per ml or a decrease in viral load of 2 logs or more. Adding all the FL parameters to a Cox proportional hazard model resulted in an overall significant model, although the individual parameters became insignificant.

the fitness landscape has to be viewed as a map representing this fitness for the entire sequence space; it also allows quantifying the genetic barrier to resistance.

HIV-1 fitness is a parameter that defines the replicative capacity in the context of environmental conditions like the presence of drug in a patient. The terms epistasis and fitness in this paper refer to statistical properties as conceptualized in population genetics. Our estimated *in vivo* fitness value is calculated based on the assumption that the increase in the prevalence of a mutation or pattern of mutations in sequences from treated compared to drug

Table 8
Time to virological suppression below 500 copies/ml or a decrease of 2 or more logs of viral load for groups of parameters. For Rega and HIVdb genotypic susceptibility score (GSS), the cut off was 0–1, >1–2 and >2 for group 1, 2 and 3, respectively. FL parameters were divided into 4 groups according to interquartile values. For Rega and HIVdb GSS 4th group was not applicable (NA).

Parameters	Time to virological suppression in weeks				*p*-value
	Group1	Group 2	Group 3	Group 4	
Rega	–	17.71	8.79	NA	<0.01
HIVdb	–	8.71	9.14	NA	0.01
Log F	8.14	5.71	7.71	8.71	0.08
GR	21.14	9.14	8.29	9.00	0.01
GF	14.14	10.86	8.86	8.71	0.05
MR	21.86	8.86	8.36	7.86	<0.01
MF	16.86	10.93	8.00	9.71	0.09

naive patients reflects an increase in fitness of the virus under drug selective pressure. The drug-pressure exerted on the virus drives adaptive evolution and selects for resistant viruses with higher fitness. In the past, using a dataset of genotype–phenotype pairs, we have shown that *in vivo* fitness estimated from a fitness landscape for the drug nelfinavir correlated well with *in vitro* resistance fold change phenotype (Deforche et al., 2008a). Not only major drug resistance mutations, but also polymorphic mutations contribute to this fitness, since they can increase fitness both in the presence and absence of drug. In the past we also showed that the estimated fitness in our protease fitness landscape models correlated significantly with viral load in drug naïve patients, and this correlation was linked to the presence of such polymorphic mutations (Theys et al., 2012). Thus, our fitness estimate can be considered an *in vivo* resistance phenotype, capturing both phenotypic resistance and intrinsic replication capacity (in the absence of drug). The fitness landscapes also allow quantifying the genetic barrier to resistance. A FL is therefore a good tool for evaluating both the susceptibility of the virus for a drug and the individualized genetic barrier of that virus for that drug, both of which can be predictive of therapy failure, because drugs with lower genetic barrier are more likely to cause therapy failure. We have previously reported that a higher genetic barrier derived from such a FL was significantly associated with higher viral load reduction in the short term and with lower odds of virological failure on long term (Deforche et al., 2008b; Theys et al., 2010).

Our previous FLs were constructed using cross-sectional data, comparing sequences from drug-naïve patients with sequences from patients experienced with the drug being investigated. To avoid a bias from the epidemiological connection between mutations rather than epistatic fitness interaction, a phylogenetic tree can be used to sample sequences from naïve patients that are epidemiologically related to sequences from treated patients. However, to completely remove this bias, FLs should be learned from longitudinal sequences (Deforche et al., 2006). These sequences are more informative than cross-sectional data because all differences in the longitudinal sequence pairs represent mutations that survived under the selective pressure of the therapy, even though they might arise in two ways, through stochastic effects or through positive selective pressure from the drugs. With such sequences coming from the same patient, the analysis was not subjected to epidemiological bias, e.g. a difference in evolutionary history. Differences in cross-sectional data also include the natural variation (which can be considered noise in this context) of the entire drug naïve HIV-1 population and hence modeling the fitness function requires a lot more sequences to separate significant signal from the noise. Nevertheless, longitudinal sequences are more difficult to obtain compared to the cross-sectional sequences due to the inherent difficulty in following the patient and the confidential nature of patient data.

We hypothesized that in the absence of sufficient cross-sectional or longitudinal sequence data from patients failing a new drug, we can map epistatic interactions in the targeted protein by training a BN based on cross sectional sequences from patients experienced with other drugs targeting the same protein, while using the scarce longitudinal sequences from patients treated with the new drug under investigation to scale the FL. The essence of using these sequence populations was that more amino acid interactions can be learned from drug selective pressure with other drugs acting on the same HIV protein, than from the limited number of longitudinal data available of only the drug under investigation. This is a valuable option for new drugs for which sequence data are limited but that inhibit a protein for which extensive data from other drugs is available. The drug indinavir was chosen as a model drug, because sufficient data were available for both cross-sectional and longitudinal FL construction and evaluation.

For all models, epistatic interactions were learned from three sets of cross-sectional data of HIV-1 sequences: from patients (i) failing the drug under investigation, in this case indinavir (BNT used in Fcross and FlongT), (ii) failing any drug that targets the same gene as the drug under investigation (BNP used in FlongP), or (iii) failing any drug that targets the same gene except data from patients failing the drug under investigation (BND used in FlongD). In the cross-sectional fitness landscape, cross-sectional sequences were then used to scale Fcross, while for all the longitudinal landscapes, we used the same small set of longitudinal data (Flong) for this scaling step. The FL FlongD mimics the situation of a salvage drug for which not sufficient cross-sectional data are available to build a BN. Although most of the mutations known to be involved in indinavir drug resistance are included in our longitudinal or cross-sectional FLs, the mutation lists are slightly different. This can be explained by the different mutations included in the BN. Essentially our three BNs modeled the same mutations, except that some mutations present in BNP and BND were absent in BNT. These extra mutations correspond to mutations known to be associated with resistance to PIs other than indinavir for instance 20T/V/M, 33F, 53L and 71Y are selected by the drug atazanavir, 30N is predominantly a major and 88D/S a minor nelfinavir resistance mutation, 48V is selected by atazanavir and saquinavir. This shows that more interactions can be captured when pooling data from more than one protease inhibitor.

Cross-sectional and longitudinal models were compared to each other in terms of the estimated fitness predicted from the same set of sequences. Since the BN maps epistatic interactions among resistance mutations and polymorphic amino acids including from different subtypes, we anticipated that the FL should be able to capture the effect of drug selective pressure in different subtypes. To investigate how reproducible the FLs were across subtypes, we used an evaluation sequence set with a different treatment history and subtype distribution than the training dataset. Despite a significant subtype difference between the training and evaluation sequences the estimated fitness according to the different FLs correlated well with each other. However, at log fitness values above 1.5 there appeared to be some banding pattern in the plots (Supplementary Fig. 6). Separate correlation bands for fitness values of subtype B sequences versus other subtypes clearly indicate that the linear regression model needs to be corrected for a subtype factor (right panels of Fig. 4 and Supplementary Figs. 5–7). Because our FL models inferred fitness for mutation patterns of evolutionary information, the inter-subtype variability may introduce a bias. We assume this is because one subtype will not evolve into a different subtype due to drug selective pressure only, and within a patient in a short time interval, such that the impact of subtype dependent polymorphisms that are not selected for during therapy (for example because they are the wildtype in that subtype) will not be present in our fitness landscapes. We were

however unable to compare the predictive power within each subtype because except for subtype B, there were not enough data available.

All four FLs could predict the accumulation of mutations comparing the baseline sequence with a follow-up sequence. Of the correctly predicted mutations, 10I/V, 20R, 24I, 46I/L, 54V, 71T/V, 82A/F, 84V and 90M correspond to those listed in the updated mutation list for indinavir (Johnson et al., 2011). The appearance of major indinavir mutation 84V was significantly more predicted (negatively correlated) than observed using FL model for which the epistatic interactions (BND) were trained on PI treatment data that excluded patients failing on indinavir (PD). The following indinavir IAS mutations, 20M, 32I, 36I, 76V and 82T were present in some of the fitness landscapes but not significantly predicted. The following indinavir IAS mutations; 10R and 73A were not present in any of the fitness landscapes.

Using an independent set of clinical data, FLs were evaluated for predictive performance in the short term (12 weeks) and long term (48 weeks). A unit increase in both Rega and HIVdb indinavir GSS at baseline was significantly associated with lower odds of virological failure in both short and long term. Likewise, the fitness parameters predictive of treatment failure in all longitudinal models were those pertaining to the fitness and genetic barrier to a fitness threshold (LogF, GF, and MF) while the measure of genetic barrier to resistance, GR and MR were predictive for all FLs except for GR the FL model FlongT (both short and long term) and FlongP (only long term). FlongT, the fitness landscape for which both the epistatic interactions and the fitness scaling were modeled using sequences that had experienced indinavir selective pressure, is not performing as good at the other fitness landscapes. We can only speculate that this may be due to the subtype effect, or to the fact that there were fewer training data to catch the epistatic interactions, in comparison with the other longitudinal fitness landscapes.

We analyzed the data used to evaluate the performance of the FL and GSS from expert systems for its ability to predict median survival time across interquartile groups of FL parameters and GSS cut-offs. The median time to maximal suppression was around 10 weeks. We could significantly distinguish time to suppression based on the various GSS groups and interquartile parameters of the FLs. However, we could not rule out the influence of other factors that determine virological failure such as adherence and suboptimal doses.

This study is a proof of concept, showing how longitudinal indinavir data can be used to estimate the *in vivo* HIV-1 fitness function under indinavir selective pressure. Because drugs within the same drug class select mutations which show similar interactions, the epistatic interactions may be modeled with interactions induced by other drugs from the same class on the same protein. Mutations derived from the sparse longitudinal data are then used to scale the FL, and it is this scaling that would then be useful to derive individualized genetic barrier for that new drug and hence provide a model to predict therapy success given a baseline sequence from a patient.

Our model of a FL, given its power to predict the failing genotype, can be a useful tool to predict therapy outcome in resource-limited settings, albeit in the presence of at least one genotype at some point before the failing treatment. The combination of FL models with other confounders such as adherence, viral load and CD4+ T-cell count history, age, drug metabolic markers and treatment history should improve the prediction of therapy response and resistance development in general. The cost-effectiveness of this approach in resource-limited settings compared to routine genotyping deserves further investigation.

Further studies are needed to validate the clinical applicability of these models to contemporary ART including PIs with a high genetic barrier to resistance. For these new models an improved simulator is planned to enable pairwise sequence comparison from longitudinal dataset, thereby restricting within-pair evolution.

5. Conclusion

This study suggests that fitness landscapes (FL) estimated from longitudinal data perform well as compared to FLs using cross-sectional data or the expert systems in predicting treatment response. FL can be modeled from longitudinal data, even when epistatic interactions are learned from cross sectional data and mainly from selective pressure with other drugs acting on the same HIV protein. With this approach we can model FL for drugs with limited sequence information such as those used in salvage therapy. Therefore, estimated fitness and genetic barrier derived from longitudinal FLs can contribute to an improvement of predicted treatment outcome. This could be used along with expert systems such as Rega IS to improve its predictive power.

Acknowledgment

RZS receives a PhD grant from the Belgian Technical Cooperation (BTC). KD and GB were funded by a PhD grant from the Institute for the Promotion of Innovation through Sciences and Technology in Flanders (IWT). KT received funding from the Research Fund of the KU Leuven (PDMK/10/204) and from the Fonds voor Wetenschappelijk Onderzoek (FWO) Flanders. This work was supported in part by the KU Leuven through grant OT/04/43, the Interuniversity Attraction Poles (IAP) P6/41, the Fonds voor Wetenschappelijk Onderzoek (FWO) Flanders (grant G.0611.09N) and the EU funded projects Virolab (grant IST STREP project 027446), EuResist (grant IST-4-027173-STP). The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under the project "Collaborative HIV and Anti-HIV Drug Resistance Network (CHAIN)" – grant agreement no 223131.

We acknowledge contributions of the Virolab project by the following colleagues: Professor Nino Manca, Dr. Franco Gargiulo, Dr. Eugenia Quiros-Roldan, Dr. Laura Albini, Dr. Giuseppe Paraninfo and Dr. Giuseppe Lapadula.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.meegid.2013.03.014.

References

Altmann, A., Beerenwinkel, N., Sing, T., Savenkov, I., Doumer, M., Kaiser, R., Rhee, S.-Y., Fessel, W.J., Shafer, R.W., Lengauer, T., 2007. Improved prediction of response to antiretroviral combination therapy using the genetic barrier to drug resistance. Antivir. Ther. (Lond.) 12, 169–178.

de Oliveira, T., Deforche, K., Cassol, S., Salminen, M., Paraskevis, D., Seebregts, C., Snoeck, J., van Rensburg, E.J., Wensing, A.M.J., Van de Vijver, D.A., Boucher, C.A., Camacho, R.J., Vandamme, A.-M., 2005. An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics 21, 3797–3800.

Deforche, K., Silander, T., Camacho, R.J., Grossman, Z., Soares, M.A., Van Laethem, K., Kantor, R., Moreau, Y., Vandamme, Anne-M., 2006. Analysis of HIV-1 pol sequences using Bayesian Networks: implications for drug resistance. Bioinformatics 22, 2975–2979.

Deforche, K., Camacho, R.J., Grossman, Z., Silander, T., Soares, M.A., Moreau, Y., Shafer, R.W., Van Laethem, K., Carvalho, A.P., Wynhoven, B., Cane, P., Snoeck, J., Clarke, J., Sirivichayakul, S., Ariyoshi, K., Holguin, A., Rudich, H., Rodrigues, R., Bouzas, M.B., Cahn, P., Brigido, L.F., Soriano, V., Sugiura, W., Phanuphak, P., Morris, L., Weber, J., Pillay, D., Tanuri, A., Harrigan, P.R., Shapiro, J.M., Katzenstein, D.A., Kantor, R., Vandamme, A.-M., 2007. Bayesian network analysis of resistance pathways against HIV-1 protease inhibitors. Infect. Genet. Evol. 7, 382–390.

Deforche, K., Camacho, R.J., Van Laethem, K., Lemey, P., Rambaut, A., Moreau, Y., Vandamme, A.-M., 2008a. Estimation of an *in vivo* fitness landscape experienced by HIV-1 under drug selective pressure useful for prediction of drug resistance evolution during treatment. Bioinformatics 24, 34–41.

Deforche, K., Cozzi-Lepri, A., Theys, K., Clotet, B., Camacho, R.J., Kjaer, J., Van Laethem, K., Phillips, A., Moreau, Y., Lundgren, J.D., Vandamme, A.-M., 2008b. Modelled *in vivo* HIV fitness under drug selective pressure and estimated genetic barrier towards resistance are predictive for virological response. Antivir. Ther. (Lond.) 13, 399–407.

Hirsch, M.S., Günthard, H.F., Schapiro, J.M., Brun-Vézinet, F., Clotet, B., Hammer, S.M., Johnson, V.A., Kuritzkes, D.R., Mellors, J.W., Pillay, D., Yeni, P.G., Jacobsen, D.M., Richman, D.D., 2008. Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society-USA panel. Clin. Infect. Dis. 47, 266–285.

Johnson, V.A., Calvez, V., Günthard, H.F., Paredes, R., Pillay, D., Shafer, R., Wensing, A.M., Richman, D.D., 2011. 2011 update of the drug resistance mutations in HIV-1. Top. Antivir. Med. 19, 156–164.

Kantor, R., Machekano, R., Gonzales, M.J., Dupnik, K., Schapiro, J.M., Shafer, R.W., 2001. Human immunodeficiency virus reverse transcriptase and protease sequence database: an expanded data model integrating natural language text and sequence analysis programs. Nucleic Acids Res. 29, 296–299.

Libin, P., Deforche, K., van Laethem, K., Camacho, R.J., Vandamme, A.-M., 2007. Regadb: An open source, community-driven HIV data and analysis management environment. In 13th International Bioinformatics Workshop on Virus Evolution and Molecular Epidemiology, Lisbon, Portugal.

Liu, T.F., Shafer, R.W., 2006. Web resources for HIV type 1 genotypic-resistance test interpretation. Clin. Infect. Dis. 42, 1608–1618.

Myllymäki, P., Silander, Tomi., Tirri, H., Uronen, P., 2002. B-course: a web-based tool for Bayesian and causal data analysis. Int. J. Artif. Intel. Tools 11, 369–387.

Shafer, R.W., Schapiro, J.M., 2008. HIV-1 drug resistance mutations: an updated framework for the second decade of HAART. AIDS Rev. 10, 67–84.

Theys, K., Deforche, K., Beheydt, G., Moreau, Y., van Laethem, K., Lemey, P., Camacho, R.J., Rhee, S.-Y., Shafer, R.W., Van Wijngaerden, E., Vandamme, A.-M., 2010. Estimating the individualized HIV-1 genetic barrier to resistance using a nelfinavir fitness landscape. BMC Bioinformatics 11, 409.

Theys, K., Deforche, K., Vercauteren, J., Libin, P., Van de Vijver, D.A.M.C., Albert, J., Asjö, B., Balotta, C., Bruckova, M., Camacho, R.J., Clotet, B., Coughlan, S., Grossman, Z., Hamouda, O., Horban, A., Korn, K., Kostrikis, L.G., Kücherer, C., Nielsen, C., Paraskevis, D., Poljak, M., Puchhammer-Stockl, E., Riva, C., Ruiz, L., Liitsola, K., Schmit, J.-C., Schuurman, R., Sönnerborg, A., Stanekova, D., Stanojevic, M., Struck, D., Van Laethem, K., Wensing, A.M., Boucher, C.A.B., Vandamme, A.-M., 2012. Treatment-associated polymorphisms in protease are significantly associated with higher viral load and lower CD4 count in newly diagnosed drug-naive HIV-1 infected patients. Retrovirology 9, 81.

Van Laethem, K., De Luca, A., Antinori, A., Cingolani, A., Perna, C.F., Vandamme, A.-M., 2002. A genotypic drug resistance interpretation algorithm that significantly predicts therapy response in HIV-1-infected patients. Antivir. Ther. (Lond.) 7, 123–129.

Van Laethem, K., Vandamme, A.-M., 2006. Interpreting resistance data for HIV-1 therapy management-know the limitations. AIDS Rev. 8, 37–43.

Vandamme, A.-M., Camacho, R.J., Ceccherini-Silberstein, F., de Luca, A., Palmisano, L., Paraskevis, D., Paredes, R., Poljak, M., Schmit, J.-C., Soriano, V., Walter, H., Sönnerborg, A., 2011. European recommendations for the clinical use of HIV drug resistance testing: 2011 update. AIDS Rev. 13, 77–108.

Vercauteren, J., Vandamme, A.-M., 2006. Algorithms for the interpretation of HIV-1 genotypic drug resistance information. Antivir. Res. 71, 335–342.

Zazzi, M., Kaiser, R., Sönnerborg, A., Struck, D., Altmann, A., Prosperi, M., Rosen-Zvi, M., Petroczi, A., Peres, Y., Schülter, E., Boucher, C.A, Brun-Vezinet, F., Harrigan, P.R., Morris, L., Obermeier, M., Perno, C.-F., Phanuphak, P., Pillay, D., Shafer, R.W., Vandamme, A.-Mieke, Van Laethem, K., Wensing, A.M.J., Lengauer, T., Incardona, F., 2011. Prediction of response to antiretroviral therapy by human experts and by the EuResist data-driven expert system (the EVE study). HIV Med. 12, 211–218.