AIDS Rev. 2010;12:89-102

Covering All Bases in HIV Research: Unveiling a Hidden World of Viral Evolution

Bram Vrancken[1], Sébastian Lequime[2], Kristof Theys[1] and Philippe Lemey[1]

[1]Rega Institute for Medical Research, Katholieke Universiteit Leuven, Leuven, Belgium; [2]UFR Sciences Pharmaceutiques, Université François Rabelais de Tours, Tours, France

Abstract

The advent of next-generation sequencing technologies has greatly impacted genomic research by providing a means for increasing the amount of sequence information in a cost effective fashion. Among the major next-generation sequencing technologies, the Roche 454 platform has been widely adopted in HIV research. We discuss a broad range of applications of the 454 pyrosequencing platform in the HIV field, with a particular emphasis on antiretroviral therapy and virus-host interaction-related research. We also highlight some of the bioinformatics challenges, as well as advantages and potential limitations of this "deep" sequencing tool, and hint at future research applications. **(AIDS Rev. 2010;12:89-102)**

Corresponding author: Bram Vrancken, bram.vrancken@rega.kuleuven.be

Key words

HIV. 454 pyrosequencing. Minor variants. Next generation sequencing. Drug resistance. Antiretroviral treatment.

Introduction

Since its development in the 1970s, the dideoxy chain-termination method (or "Sanger" method[1,2]) has proven invaluable for genomic research and has set the benchmark in terms of read length and sequencing accuracy. Demands for high sequencing throughput, to a large extent driven by the Human Genome Project, stimulated considerable technical improvements and ultimately led to the contemporary 96-well automated capillary electrophoresis. These advances brought about substantial cost savings, but still struggle to meet the ever increasing demand for more affordable sequencing, exemplified by the 2004 NIH goal of the 1,000$ human genome by 2015[3].

Recent progress on this front has come from a number of high-throughput sequencing methods, commonly referred to as next-generation sequencing (NGS) platforms. Due to the parallelization of the sequencing process, they all offer an increased sequence data output, both at lower cost per base pair and at significantly higher speed compared to Sanger sequencing[4,5]. Initially employed for large-scale (re)sequencing projects, their use has been adopted in a diverse range of applications such as ChIP-sequencing, transcriptome and small RNA profiling, epigenetic studies, and metagenomics[6-8].

Three major players have now emerged on the NGS market: Roche 454, Illumina GA, and ABI SOLiD. As pioneer in the field, a great deal of experience has been acquired for the Roche 454 pyrosequencing platform[9], and because it excels in read length, it also appears very promising for increasing the sequence information of highly variable viruses such as HIV. We consider two important dimensions or goals in this increment: the accumulative sequence length through overlapping reads, and the coverage depth obtained by a large number of fragments representing the same genomic region (Fig. 1). While the latter can be considered as redundant information in many standard genomic projects, it provides a powerful population-targeted sequencing approach for those attempting to characterize heterogeneous (viral) populations.

Largely driven by drug resistance research, deep sequencing approaches have always been of great interest to investigate HIV evolutionary dynamics within hosts. As a member of the *Retroviridae* family,

Correspondence to:
Bram Vrancken
Kapucijnenvoer 33 blok i -7e verdiep
3000 Leuven, Belgium
E-mail: bram.vrancken@rega.kuleuven.be

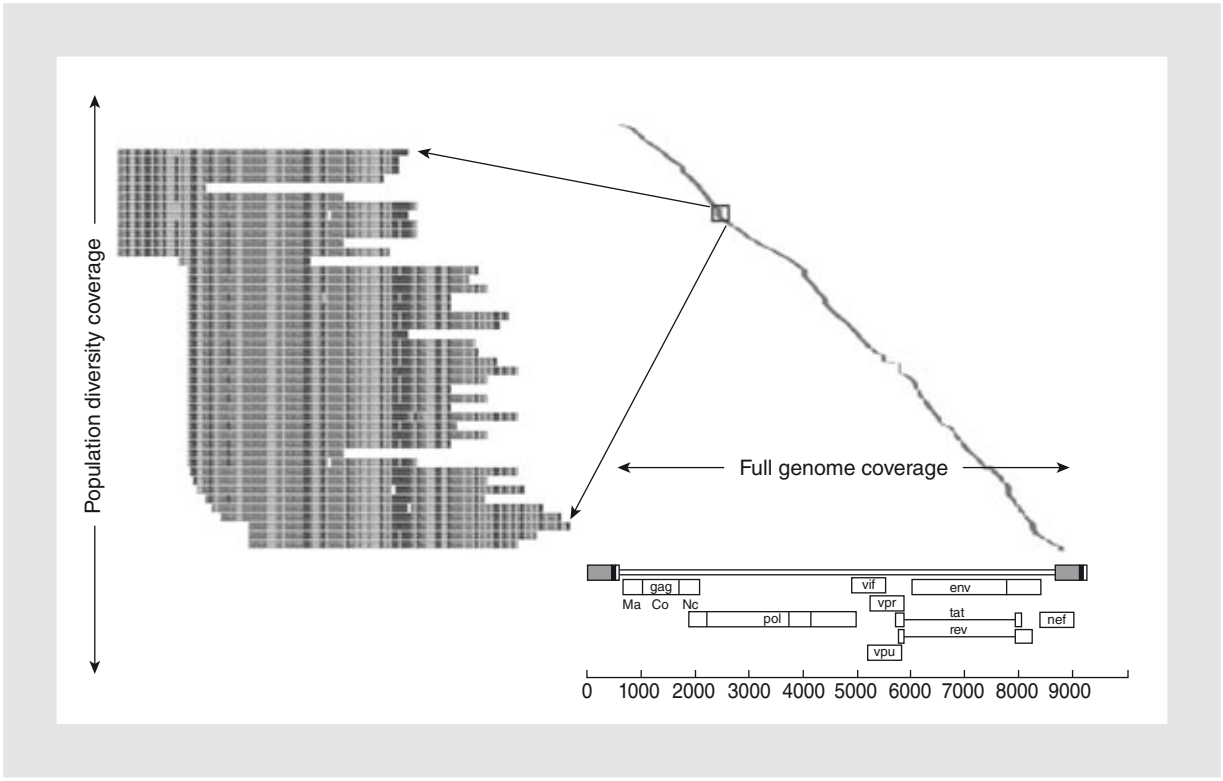


Figure 1. *Illustration of both dimensions of sequence increment. The rectangle marks an area in the HIV-1 genome alignment, for which the aligned reads and the coverage they achieve are presented in more detail. Results were obtained from a pilot GS20 run.*

the HIV lifecycle includes a step that converts the RNA genome into a proviral DNA. The virally encoded reverse transcriptase responsible for this task lacks proofreading activity and is the main contributor to the high error rate associated with the RNA to DNA to RNA transition (estimated at about $1/10^4$ to $1/10^5$ errors per base per cycle[10]). In combination with a rapid turnover and a large replication space, this provides the HIV virus with an enormous evolutionary potential to escape from host immune responses and antiretroviral therapy. Characterizing the viral population within hosts is therefore central to our understanding of HIV dynamics and how viral replication may be controlled.

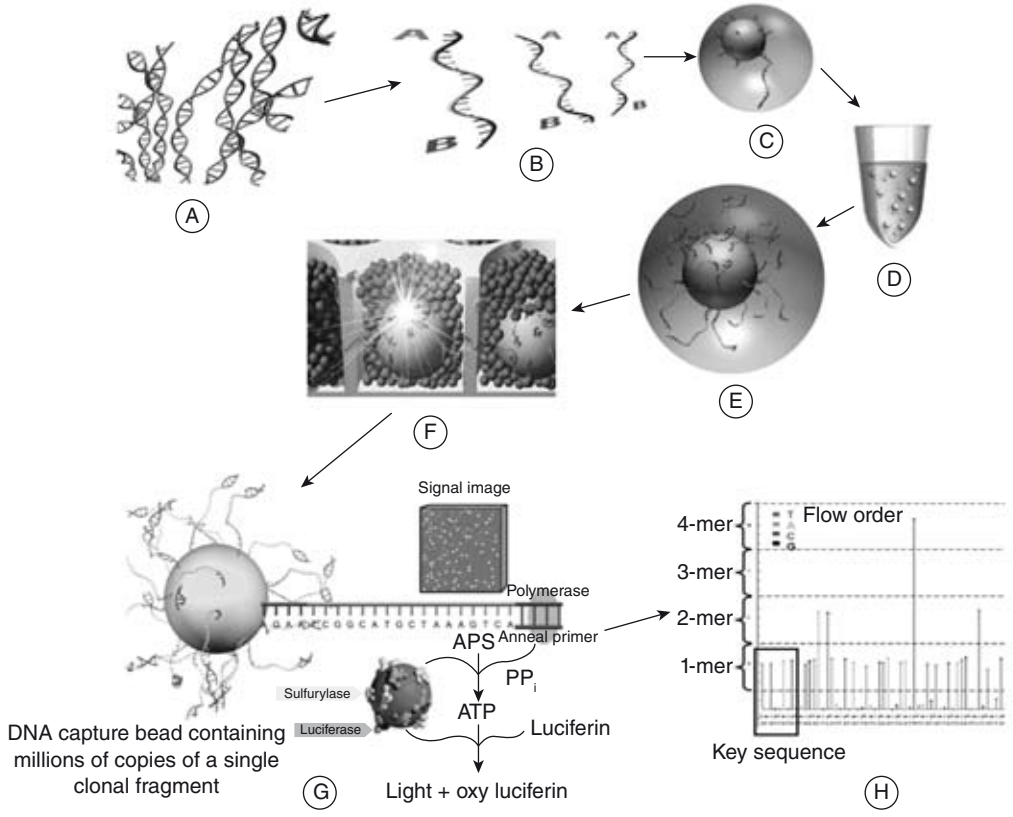
The Roche 454 technology is based on the sequencing-by-synthesis principle and consists of a number of preparatory steps to perform this in a large-scale parallel fashion. Briefly, a library of DNA molecules flanked with adaptors A and B at their opposing ends is created and fixed to beads under conditions whereby preferentially one molecule per bead will be amplified. The beads are emulsified and the DNA fragments are subsequently amplified by emulsion polymerase chain reaction (PCR). After breaking of the emulsion and an enrichment process, the beads carrying amplified DNA are loaded on a PicoTiterPlate™ and simultaneously subjected to sequencing by synthesis. Because an amplification step is required for RNA viruses before they can be subjected to the 454 pyrosequencing procedure, these viral sequencing experiments are most often designed to generate a DNA library that consists of amplicons incorporating the appropriate adaptor sequences at both ends. We provide a more detailed and illustrated description of the general Roche 454 procedure in Box 1. In Box 2 we highlight differences with other NGS platforms. Although 454 sequencing involves a cost per base that is about two orders of magnitude lower than traditional Sanger sequencing, the cost per individual run remains high[5,11].

Multiplexing different samples on one plate has therefore been proposed as a cost effective strategy. This can be easily achieved by physically dividing the PicoTiterPlate™ through the use of gaskets, bearing in mind though that such a procedure leads to substantial capacity loss[12] and may suffer more severely from errors due to edge-effects[13]. To avoid these problems, Hoffmann, et al.[14] took an approach of adding unique tags for each sample to the primers used in the PCR amplification step, making it possible to attribute each read to its original sample *a posteriori* – a strategy referred to as barcoding.

Box 1. The Roche 454 sequencing procedure

The pre-processing of samples for the current Roche 454 platform, the GS FLX, follows the same steps as for its predecessor, the GS20. The procedure starts by creating a double-stranded DNA-library (Fig. 2 A) in which each fragment is flanked by adaptors A and B at the opposing ends. These adaptors provide the priming sites for amplification and sequencing. Subsequently, the library is denatured (Fig. 2 B) and the single-stranded DNA strands are annealed to beads coated with oligomers complementary to one of both adaptors (Fig. 2 C), depending on the library preparation method. The beads are then subjected to an emulsion PCR (Fig. 2 D) (emPCR). This step serves to clonally amplify the templates in water-in-oil micro reactors, bypassing the need to clone the fragments in any other laborious way. Following the emPCR, each template-carrying bead, now with approximately 10^7 clonal copies of the initial template (Fig. 2 E), is transferred into a well on the PicoTiterPlate™ (Fig. 2 F). These contain approximately 3.4 *10^6 wells on the GS FLX[7], thus facilitating the parallel sequencing of many templates. The actual sequencing process (Fig. 2 G) starts by delivering the different deoxyribonucleotide triphosphates (dNTP) in fixed order. A pyrophosphate is released with each incorporation of a dNTP and is used by adenosine triphosphate (ATP) sulfurylase to release ATP. This in turn serves as an energy source for the conversion of luciferine to oxyluciferine, a luminescent reaction. Relying on the equimolar stoichiometry of the reactions, the resulting light signal intensity should be proportional to the amount of dNTP incorporated. This feature allows accurate detection of homopolymers, albeit only over a limited range. All light signals in the PicoTiterPlate™ are detected by a charge coupled device, and the sequences are reconstructed by converting the sequentially detected light signals to the appropriate dNTP or homopolymers of the corresponding dNTP (Fig. H).

Illustration of the 454 pyrosequencing procedure.



Courtesy of Roche Diagnostics.

In this review, we present an overview of the applicability of NGS in HIV research, which mainly amounts to characterization of the viral intra-host population and the use of vast amounts of sequence data in virus-host interaction studies. We do not attempt to give an exhaustive overview of all advances made in the rapidly evolving field of HIV research, but we focus on the use of the Roche 454 platform for a number of clinically important issues. We start by discussing bioinformatics challenges that are particularly important for HIV research and position the NGS platform in the evolution of HIV genotyping research. We review the ground covered by the new deep-sequencing approaches in drug resistance research, but also touch upon HIV

Box 2. Illumina and SOLiD versus Roche 454

Although conceptually based on the same principle of massively parallel sequencing, the Illumina, SOLiD, and 454 technologies differ in sequencing chemistry. As for Roche 454, the input for both Illumina and SOLiD are adaptor-flanked DNA fragments. On the SOLiD platform the DNA is clonally amplified through an emulsion PCR, whereas on the Illumina platform a procedure called bridge-PCR generates a cluster of clonal sequences[4] . The Illumina platform adheres to a sequencing-by-synthesis principle. In brief, each cycle consists of the incorporation of a terminally labeled dNTP, followed by the detection and subsequent cleavage of the dye. In contrast, the SOLiD technology builds on the sequencing-by-ligation principle. Here, differentially labeled octamers are used as probes and sequencing is performed through various rounds of annealing and ligation[4,124]. The main difference with the Roche 454 platform is that on both of these NGS platforms, each nucleotide is queried one at a time, as opposed to determining the length of a homopolymer stretch. As a result, the most commonly found errors are substitutions instead of indels as for the Roche 454 platform[4]. Also, whereas Roche 454 obtains long reads, on average 400 bases on the GS FLX with Titanium chemistry, both Illumina and SOLiD technologies can be classified as "short read" technologies, with current read lengths of 75-100 and 50 bases respectively[8]. However, the output of the latter platforms is considerably larger: depending on the experimental setup, up to 35 Gb (Illumina) and 50 Gb (SOLiD) output per run can be achieved[8], which reduces the cost per base even further than for the Roche platform[4,5]. For a more detailed explanation on these and other NGS platforms, we refer to a number of excellent reviews on this topic [4,6-8,11,124].

integration and immune escape studies. For every aspect, we attempt to highlight the relative strengths of the NGS platform compared to other assays, but also point at potential pitfalls.

Bioinformatics challenges

Next-generation sequencing platforms generate vast amounts of data that consist of many short, overlapping sequencing reads. In the case of Roche 454, 400 to 600 million high-quality bases can be produced in a single run. The analysis of data with these and other platform-specific characteristics poses particular bioinformatics challenges; here, we consider base-calling, read sorting and alignment, haplotype reconstruction, comparative analyses, and visualization.

When the aim is to characterize the HIV intrapatient population in great detail, or to detect single nucleotide polymorphisms in general, read accuracy and accurate base calling (i.e. reliably recognizing the identity of a base) are essential parameters. In this regard, it is important to note that individual bases on the GS20 system are of lower quality than the bases obtained through Sanger sequencing[9]. Further innovations leading to the GS FLX system, the successor of the GS20 platform, now ensure that single-read accuracies at least equivalent to Sanger sequencing can be obtained[15]. In parallel, efforts have been undertaken to improve the base quality scoring algorithm. The original proprietary algorithm provided an estimate of the probability that the detected homopolymer contains a wrongly called additional base, referred to as an "overcall"[9,13]. Improved algorithms that more reliably estimate the actual base quality were published almost simultaneously by Quinlan, et al.[16] and a group at the Broad Institute[17]. The current proprietary scoring algorithm was constructed in collaboration with the latter institute, and as it is based on the methodology developed by Ewing and Green[18], it provides consistency with previously obtained phred-based quality scores.

Read quality also becomes a critical issue when resorting to barcode procedures. Erroneous assignments of reads due to sequencing errors in the barcode cannot be detected nor corrected with the methodology introduced by Hoffmann, et al.[14]. This motivated Hamady, et al.[19] to refine the barcode approach by basing its design on the Hamming distance, a method that allows the identification and correction of errors, thus minimizing sample assignment errors. This strategy of error-correcting barcodes was further elaborated on by Roche, whereby they designed barcodes (called "Multiplex IDentifiers" or "MIDs") based on the minimum edit distance. These have the advantage that they not only apply to strings of the same length (like Hamming codes), but also take into account insertions and deletions, which are more common errors on the 454 platform than substitutions.

Following the read sorting according to their barcodes (when present), the next step in the analysis of (viral) data generally involves their alignment against a reference sequence. Here, the quantity of data and the variability of HIV pose a particular challenge. Traditional heuristic mapping solutions like BLAST may still be impractically slow and they may fail to align a number of sequences in case no "word hit" is found to initiate the alignment extension[6]. On the other hand, faster algorithms designed for efficient mapping, with a mismatch constraint or maximum divergence from the reference sequence[20], may result in unnecessary

data loss when applied to highly divergent data. This problem was addressed by Archer, et al.[21], who minimized data loss due to sequence divergence by making use of a data-specific consensus sequence as the template for read-mapping and subsequent pairwise alignment.

Having built a read alignment, mutational patterns in the viral population can be readily analyzed, with the caveat, however, that mutational linkage is largely lost. By exploiting two useful characteristics of 454 data, i.e. high coverage and overlapping clonal reads, Eriksson, et al.[22] have undertaken an effort to reconstruct the viral population structure starting from pyro-reads. In short, after alignment and error correction, haplotypes were reconstructed and their frequency was estimated using both a parsimony and maximum-likelihood approach. Analysis of the most complex sample with GS20-data at their disposal, which represents the entire protease and the first 241 codons of the reverse transcriptase, revealed that all haplotypes present at a frequency of at least 2.2% could be reconstructed with 99% confidence. Zagordi, et al.[23] employed a different approach to tackle the haplotype reconstruction problem. They considered the observed reads as a sample of the actual viral population, and developed a Bayesian model to assign reads to their respective haplotypes, as well as to distinguish biological variants from sequencing errors. The latter constitutes an important consideration in haplotype reconstruction, and on a simulated data set, the Bayesian model outperformed the method of Eriksson, et al.[22] in terms of error detection and correction. In contrast to the former methodology, however, the Bayesian inference is limited to local haplotype reconstruction. Nonetheless, given the read length on the GS FLX Titanium series (400 bp), this already permits the haplotype analysis of entire genes of interest, such as the HIV protease gene (297 bp). A confounding factor that needs to be considered for such reconstructions is recombination, *in vivo* as well as *in vitro*, because a considerable amount of recombination can be introduced during the PCR amplification step[24]. Taking this into account, together with further improvements in read accuracy and read length, will be necessary for more accurate haplotype reconstruction and linkage analysis.

Unfortunately, no tailor-made NGS software has been made available yet that fully addresses the needs of viral evolutionary studies. By providing summary statistics of individual mutations and their frequencies, software like Segminator[21], has made important steps towards tracking the evolution at individual sites in response to therapy and host selective pressure. However, accommodating automated comparative analyses across multiple samples and visualization remains an important research direction. When focusing on reads of a single amplicon, phylogenetic analysis has been shown to be a powerful tool to answer evolutionary questions using deep-sequencing samples (Fig. 3). Complemented by haplotype reconstruction methods, this may even be extended to study the evolutionary pathways over larger genome regions. Because read redundancy may be collapsed to a limited number of haplotypes, haplotype networks may also prove useful to visualize the temporal evolution of the viral population. Finally, we note that in the absence of linkage, many evolutionary processes can still be investigated using classical population genetic estimators, some of which have been specifically adapted for viral evolutionary processes[25].

Minor variant detection

Despite the hope offered by antiretroviral therapy to suppress viral replication, even through targeting several steps in the replication cycle simultaneously, drug resistance remains an insurmountable challenge to maintain therapeutic success. Regardless of the administered antiretroviral drug cocktail, the enormous evolutionary capacity of the virus ensures that all possible drug resistance mutations (DRM) may already be present, or may be generated relatively quickly due to residual viral replication[26]. As these mutations confer a selective advantage to the viral variants involved, their fixation in the population is inevitably associated with therapy failure.

To support medical care providers faced with these issues, resistance testing has become a useful tool to assess the potential success of therapeutic decisions[27,28]. The gold standard in resistance testing involves interpreting the viral genotype, which is routinely determined through direct sequencing of the RT-PCR products using the dideoxy chain termination method. This procedure, often referred to as “population” or “bulk” sequencing, has also been indispensable in characterizing the major HIV resistance pathways under drug selective pressure. Despite these efforts, treatment failure in response to HAART remains a common feature.

In recent years, evidence has accumulated that drug-resistant variants present at low frequency in the population may have a profound impact on therapy outcome[29-35]. However, not every study can associate

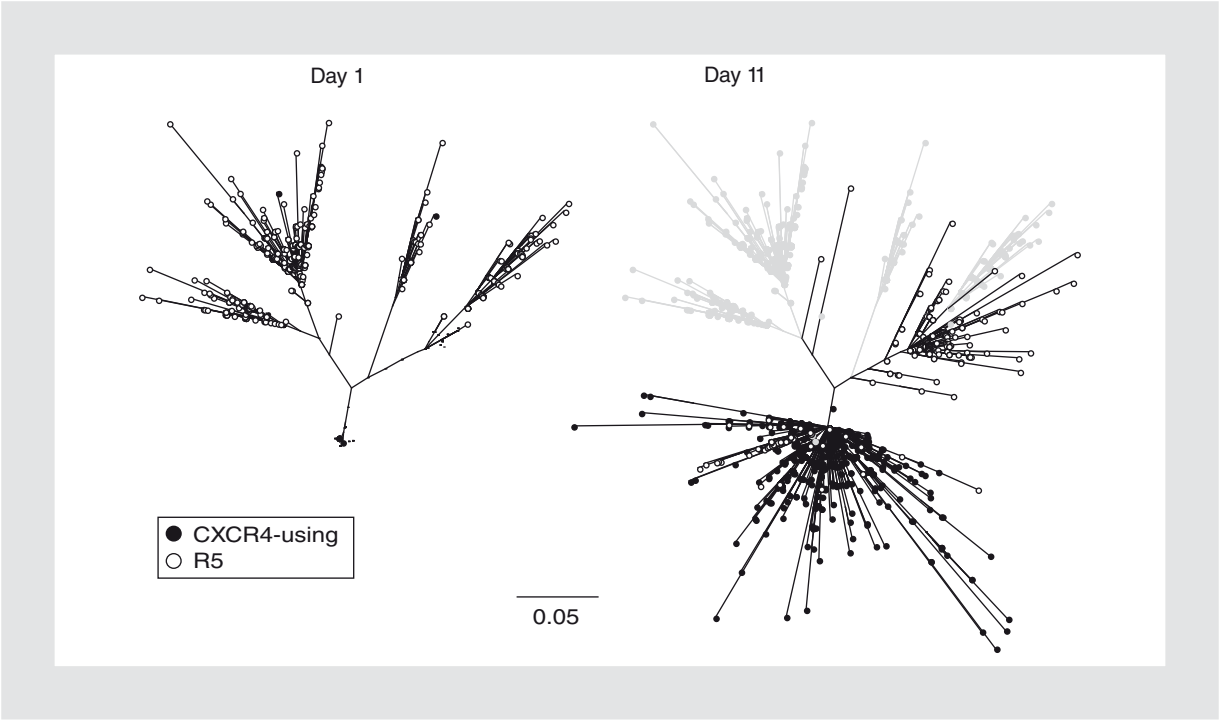


Figure 3. *Phylogenetic trees showing the short-term evolutionary history of a within-host viral population. The trees were inferred from all HIV-1 V3-loop sequences from one patient on maraviroc monotherapy obtained at two different time points (day 1 and day 11); each tip represents a unique V3 sequence. For day 11, we represent a tree including the unique sequences for both time points, but gray out the day 1 sequences. The taxon symbol represents the predicted coreceptor usage (see key). The scale bar represents substitutions per site. The dramatic expansion of CXCR4-using variants from day 1 to day 14 in the lower branch of the tree indicates that pretherapy low-level CXCR4-using variants can be selected for under maraviroc monotherapy* (adapted from Archer, et al.[21], with permission from Wolters Kluwer Health [provided by Copyright Clearance Center]).

their presence with therapy failure (e.g. Peuchant, et al.[36]), questioning the clinical significance of pre-therapy minority DRM; which low-level variants in what selective circumstances affect therapy response is still a matter of ongoing research.

Population sequencing may be of limited use in this respect because it cannot reliably detect mutations present in less than about 20-25% of the viral population[37-39]. Consequently, several methods have been developed to examine the HIV population for low-abundance variants. As pointed out by Shafer[40], these generally belong to one of two main approaches: clonal sequencing and point mutation assays.

Some point mutation assays can achieve a detection limit at least as low as 454 sequencing[41-44]. However, their use in routine genotyping remains limited due to the inherent need to take into account the sequence variation around the position under investigation[45]. This makes their validation on clinical samples a daunting task, considering the high diversity of the circulating HIV types and the large number of amino acid positions that can affect resistance. Clonal sequencing approaches, on the other hand, can profit from targeting longer sequence lengths compared to 454 sequencing, which facilitates the characterization of linkage amongst resistance mutations (~750 bp with Sanger sequencing vs. ~400 bp on the GS FLX with Titanium chemistry). However, detailed quantification of minor variants requires time-consuming, costly, and laborious efforts[46]. With these limitations in mind, Roche 454 pyrosequencing offers an attractive alternative by capitalizing on high sensitivity and reasonably long read lengths. As discussed below, this combination makes it a promising tool to gain more quantitative insights into the kinetics of viral evolution in various research settings.

As with all PCR-based methods, the sensitivity that can be achieved depends on the number of available target templates[47,48], in this case for the creation of the dsDNA library. This is in turn limited by the patient's viral load and the efficiency of the preparatory procedures. The results of Poon, et al.[49] indicate that RNA extraction from the sample represents the most critical step, and therefore they suggest a minimum of two

replicates of the RNA extraction step as a standard for the use of NGS in genotypic resistance testing.

In addition to sensitivity, biases in template-to-product ratio are an important factor to consider because any deviation in the proportionality of the templates will confound the characterization of the viral population composition. This bias may be introduced by the experimenter, but also by PCR drift and PCR selection. The PCR drift is a consequence of stochastic variations in the early cycles of the PCR reaction and will become particularly important at low viral loads and when low-level variants are involved[50,51]. Therefore, it is advised to pool the products of a number of independent RT-PCR reactions in order to reduce the founder effect of the RT-PCR procedure[52].

Although it remains difficult to separate PCR drift from experimenter-induced variations, Tsibris, et al.[53] clearly revealed experimental consistency issues. After 4-replicate amplification of the same HIV-extract, analysis of the proportion of CXCR4-using minor variants revealed that this fraction (mean ± standard deviation of 2.42 ± 0.55%) varied within 22.8% of the mean.

In order to minimize the effect of PCR selection (inherently favored amplification of some templates over others due to differential annealing efficiency of the primers), Bracho, et al.[54] advise to utilize a certain amount of primer degeneracy reconcilable with specific amplification.

Finally, we note that errors introduced in all stadia of the experiment compromise the detection limit, which advocates for the use of high fidelity polymerases in all amplification steps. A statistical approach that accounts for this source of error was developed by Wang, et al.[55]. For this, they relied on sequence coverage redundancy, the specific error pattern in homo- and heteropolymer stretches, and a reference sequence as a control in order to discriminate between experimentally introduced errors and true variants.

Minor variants and therapy

The first reports demonstrating that minor variants can readily be detected down to < 1% by exploiting 454 pyrosequencing for "deep" interrogation of almost the entire protease and reverse transcriptase genes were published nearly simultaneously[14,55]. These studies drew a clear picture of the applicability of 454 sequencing in detecting minor variants: all DRM identified by population sequencing were also detected by 454 pyrosequencing in both studies, and in addition, a variable number of less abundant DRM (4[14] to 18[55]) were only detected by the "deep" sequencing approach. More recently, 454 pyrosequencing was used in a larger study to assess the impact of baseline low-abundance DRM on therapy outcome in a subset of patients of the FIRST study[56]. The comprehensive survey confirmed that 454 pyrosequencing identifies significantly more DRM than standard genotyping, with the majority being low-level variants (here defined as < 20% prevalence). Importantly, the low-frequency nonnucleoside reverse transcriptase inhibitor (NNRTI) DRM that were missed by the standard sequencing approach had a significantly negative impact on therapy outcome in these treatment-naive patients. This is in line with results from other studies[32,34], but not all[57], which demonstrates the need for further research of this topic[58].

Minor variants may also be of clinical importance in the course of changing therapy regimens because successive treatments may lead to the accumulation of multiclass resistance[40]. This view is further substantiated by the results of Le, et al.[59]. Here, the excellent sensitivity of "deep" pyrosequencing allowed the detection of low-abundance DRM in all 22 samples as opposed to only in three subjects with bulk sequencing. Overall, the standard genotyping method failed to detect 95% of the low-abundance drug resistance variants, whereas 454 pyrosequencing accurately detected all mutations identified by ABI Sanger sequencing. Importantly, most of the low-frequency DRM missed by standard genotyping correlated with therapy history. This led the authors to suggest that the 454 platform could be of value as a "deep" genotyping tool after therapy failure, providing potentially useful information for future therapy regimens.

Roche 454 sequencing is also the method of choice to query the viral population in patients with regimens containing recently approved drugs such as the integrase inhibitor raltegravir[60]. In order to provide a detailed characterization of the kinetics of drug resistance under raltegravir selective pressure, Ceccherini, et al.[61,62] applied "deep" sequencing on a subset of six patients, of whom four experienced treatment failure. Only in the latter could low-level variants with DRM be detected at baseline, but they were not necessarily associated with evolution at the same amino acid position at therapy failure, which is in agreement with a recent study by the same group on a larger dataset[63]. Because no primary mutations were detected at baseline, a consistent feature in the before-mentioned studies, they suggested that a more complex dynamics than straightforward drug-driven selection of primary DRM present at baseline

may influence resistance to raltegravir. In addition, since no correlation between baseline secondary raltegravir DRM and therapy outcome was found, it was concluded that integrase genotyping before raltegravir administration may, at this time, not be cost effective[64]. However, both the recent report by Codoñer, et al.[65] on baseline minority variants with major raltegravir resistance mutations that were selected under a raltegravir-containing regimen, and the divergent results concerning the prevalence of low-level variants harboring major DRM to integrase inhibitors in integrase strand inhibitor-naive patients[66], indicate further research on this topic is warranted.

Transmitted drug resistance

As a consequence of widespread antiretroviral therapy, transmitted drug resistance (TDR) has become a relevant problem. The circulation of viral variants with decreased susceptibility to widely administered drugs poses an obvious threat to the success of first-line therapy regimens[67,68]. Current estimates of TDR, around 9% in Europe[69], result from analyses of the samples using Sanger-based population sequencing techniques and may therefore be prone to underestimation[70]. Not surprisingly, the application of more sensitive techniques, such as allele-specific PCR[36,71], which only queries one position at a time, yields a higher level in TDR prevalence. A potential fitness cost associated with many DRM may also contribute to this underestimation. As noted by Deeks[72], this means that by the time HIV manifests clinically, the proportion of resistant variants has most likely already dropped under the detection limit of standard sequencing approaches. As a consequence, little is known about the impact of TDR on the success of first-line therapy regimens, which leaves open the possibility that more sensitive genotyping tools may be used to support medical care providers on treatment initiation decisions. Furthermore, Johnson and Geretti[73] noted that a more precise evaluation of TDR may also serve a public health purpose by aiding in the identification of transmission networks and populations on which to focus preventive measures.

The potential of the 454 sequencing platform to assess TDR prevalence on a population scale in a cost effective manner was demonstrated by Ji, et al.[74]. To this end, they adopted a population-targeted sequencing approach to provide a detailed characterization of an unlabeled sequence collection pooled across multiple patients. In comparison to the bulk sequencing results, an additional 133 nucleotide and 87 amino acid differences (prevalence cutoff $\geq$ 0.2%) relative to the HXB2 reference sequence were identified in 96 samples. Moreover, all TDR mutations detected via Sanger sequencing were also recognized by 454 sequencing, with the additional ones situated at the lower extreme end (< 5%) of the frequency spectrum.

In addition to screening, studies evaluating the impact of TDR on first-line therapy response now also resort to the 454 platform. Lataillade, et al.[75] studied TDR mutations in antiretroviral therapy-naive patients in the CASTLE study. They reported that TDR was common in a representative subset of these subjects: 30.5% of them had one or more TDR mutations at baseline and 15.6% only had TDR mutation(s) at levels > 20%. In accordance with other studies[76,77], low-frequency DRM had no impact on the success of first-line protease inhibitor-based therapy.

Since the massive sequence output of NGS still comes at a considerable cost, a prudent strategy is to consider conventional genotyping methods in cases where they can provide sufficient information about the optimal therapy choice. In the following two studies, viral populations were scrutinized for minor variants in patients with a DRM that, if present in newly infected patients, may be attributed to TDR.

Varghese, et al.[78] investigated the prevalence of low-abundance etravirine DRM in therapy naive patients where standard population sequencing could only detect the K103N mutation. This mutation has, on its own, no effect on etravirine efficacy, but is associated with high-level resistance against the commonly used NNRTI efavirenz and nevirapine. In multiple samples, one or more additional TDR mutations were discovered, but, as no major etravirine DRM was detected in levels above 0.5%, they concluded that etravirine may be effective in such a subset of therapy naive subjects. Based on a similar analysis in NNRTI-experienced patients, however, they discourage the use of etravirine in NNRTI-failing patients in whom K103N is the only detectable DRM by population sequencing.

Mitsuya, et al.[79] hypothesized that the presence of T215Y/F revertants, detected with population sequencing, could be a surrogate marker for minor variants with T215Y/F. These revertant mutations have a prevalence of about 3% in treatment-naive patients and are weakly associated with therapy failure. No evidence was found in support of this hypothesis, but two of the patients were found to experience therapy failure, possibly linked to other low-level mutations. Based on this and the additional low-level DRM detected, they suggest

that first-line therapy for patients with T215Y/F revertants should aim for a higher genetic barrier to resistance than usual.

Entry inhibitors

The advent of entry inhibitors further highlighted the need to characterize HIV populations prior to therapy initiation. For a CCR5 inhibitor like maraviroc, and possibly others enrolled in clinical studies[80], CXCR4-using variants remain uninhibited. Determining the coreceptor usage before administering drugs from this class has therefore become a therapeutic guideline[81].

Analogous to resistance testing, coreceptor usage can be assessed using phenotypic or genotypic tests. Genotypic tests offer the advantage of speed and lower cost compared to phenotypic assays[82] and thus several genotype interpretation tools have been developed. Although a broader genomic context may influence coreceptor usage, genotyping is currently restricted to the V3 loop sequence[83]. A disadvantage of these methods, keeping them for now in the research realm, is their limited sensitivity, especially when bulk sequence data are used[84,85]. This problem inspired a search for improved sensitivity by applying interpretation algorithms to the population data generated by 454 sequencing. As sequences can be misclassified when only the V3 loop sequence is taken into account, further improvements in predictive power may be achieved by including non-V3 loop (clonal) sequences in the analysis[83].

The pilot studies by Archer, et al.[21] and Tsibris, et al.[53] on the use of 454 pyrosequencing in combination with V3 loop-based interpretation tools both hint at the kinetics of pretherapy minor X4 variants and minor *de novo* generated resistant variants under the selective pressure of CCR5 antagonists. In the patient examined by Archer, et al.[21], X4 strains were detected at low frequency (~ 0.5%) before maraviroc treatment initiation. Using phylogenetic analysis, they established that the X4 variants rising to high frequency after a 10-day maraviroc monotherapy evolved from the preexisting X4-using minor population (Fig. 3), and thus do not necessarily evolve *de novo* under the drug selective pressure of CCR5 antagonists. The latter was consistent with the study of Tsibris, et al.[53], who found that in three out of four patients failing vicriviroc therapy, the resistant variants already existed at levels as low as 0.8-2.8% in the baseline samples.

The insights into viral dynamics based on the analysis of circulating viruses[21,53] are elegantly complemented by the results of Rozera, et al.[86,87]. By combining the capability of cell lineage marker detection with the analysis of clonal V3 loop amplicon sequences produced by 454 pyrosequencing, they realized an unprecedented comprehensive comparison between coreceptor usage of proviral and viral sequences, both originating from the same cell type. Phylogenetic analysis suggested that the proviral reservoirs in the monocyte-macrophage lineage and $CD4^+$ T-lymphocytes are a source of rebounding virus after therapy interruption. Furthermore, archived X4 provirus was identified as a putative source of replication-competent virus, which was confirmed in a later study[88]. As such, these may be selected under the drug selective pressure of CCR5 antagonists and lie at the basis of therapy failure.

Swenson, et al.[89] further expanded on the advantage of large-scale 454 sequencing as an alternative to standard sequence analysis in detecting X4 variants. The 454 pyrosequencing detected CXCR4-using variants in 202 samples, whereas 28% of them were declared X4-negative by population sequencing. As expected, this was primarily attributed to the low X4 prevalence in these samples.

In addition to outperforming standard sequencing as a "deep" interrogation tool, 454 sequencing may also prove more sensitive than some phenotypic tropism tests. Van 't Wout, et al.[90,91] reported that the 454 pyrosequencing approach detected X4 variants at levels above 5% in samples declared negative for CXCR4-using variants with the MT2 assay, and different studies demonstrated the ability to characterize samples by deep sequencing that failed the Trofile™ and/or Virco phenotypic tropism assay[52,92]. Likewise, Swenson, et al.[93] and Pou, et al.[94] reported that quantitative deep sequencing outperforms the standard Trofile™ assay and obtains similar results as the enhanced Trofile™ assay. Taken together, these findings indicate that the combination of 454 pyrosequencing and coreceptor prediction algorithms could provide a valuable alternative to phenotyping.

Another advantage over tropism assays is the possibility to quantify 454 sequences very accurately. Swenson, et al.[95] used this property to stratify the patient population according to the measured X4 prevalence, leading to the finding that a low prevalence of X4 is associated with improved treatment response. The accurate determination of the appropriate clinical cut-off could be simply performed by submitting all patient samples to deep sequencing. However, a more economical approach is proposed by Vandekerckhove, et al.[96].

They recommend a first screening by population sequencing and tropism prediction tools to retain only the samples with low levels of X4 variants for further investigation on the 454 platform.

The genetics of HIV integration

The clonal output and low cost per-base of the Roche 454 platform opens up attractive opportunities for studies requiring a large amount of sequence information such as HIV integration-site studies. As the current therapies are unable to eliminate the long-lived latent reservoirs, HIV patients need to resort to life-long HAART treatment. Therefore, if complete HIV eradication in infected patients is the goal, it is essential to gain more insights into HIV integration and latency establishment[97-99]. A better understanding of HIV integration site selection also has practical implications for gene therapy using lentiviral vectors[100]. In this research field, 454 sequencing is contributing to the analysis of adverse events associated with gene therapy[101,102].

Wang, et al.[55] made use of the massive 454 output to produce a dataset exceeding previous ones by one to two orders of magnitude. Complemented by state-of-the-art bioinformatics analyses, they were able to associate *in vivo* integration with a number of post-translational histone modifications linked to active transcription. Inhibitory histone modifications, on the other hand, were shown to negatively impact integration.

Marshall, et al.[103] further substantiate that PSIP1/LEDGF/p75 (further referred to as LEDGF) is important, though not essential, in integration site targeting. In their experimental setup, the barcode approach was applied in order to sequence many integration site populations in parallel, which enabled them to identify 3,566 unique integration site sequences. Through the mapping of integration site sequences derived by 454 pyrosequencing from cells depleted of various nuclear factors, Ocwieja, et al.[104] identified several proteins in addition to LEDGF that may play a role in integration site targeting.

Also profiting from the increased quantity of information, Ferris, et al.[105] and Ronen, et al.[106] recently used 454 pyrosequencing to demonstrate that fusions of the LEDGF integrase-binding domain with an alternative chromatin-binding domain can redirect HIV-1 integration to different locations, which can have important implications in lentiviral vector-based gene therapy.

Employing a multiplexing strategy analogous to Marshall, et al.[103], Brady, et al.[107] addressed the question whether HIV preferentially integrates at different locations in resting versus activated CD4+ T-cells. They revealed that *de novo* integration occurs more often in expression-repressive environments in resting cells, leading them to suggest that differential target site selection for integration results more frequently to latency establishment in resting cells.

Immune-driven adaptation

The exact mechanisms by which HIV circumnavigates and eventually destroys the immune system are not completely understood. In contrast, it is well established that the adaptation of the viral genome to its host environment plays an important role in disease progression. Unveiling this adaptive evolution is therefore critical for a better understanding of AIDS progression and may assist in the development of an effective vaccine[108].

The interactions between the evolutionary patterns, host genotype, and disease progression are difficult to disentangle based on limited markers of the viral genome and incomplete knowledge of the host genetics involved. Unfortunately, comprehensive studies designed to overcome these hurdles are cost prohibitive. The combination of high-throughput human leukocyte antigen (HLA) typing with the sensitive detection of adaptive changes over the complete genome may, however, offer the opportunity for examination of large cohorts at manageable costs.

Different groups have made progress on the HLA typing front by applying 454 pyrosequencing to PCR-amplified HLA exons, which can be almost entirely covered by the current read length, demonstrating its potential for economical high-throughput HLA typing[109-112].

Recently, Bimber, et al.[113] reported on the use of 454 pyrosequencing to track the CD8+ T lymphocyte (CTL)-mediated adaptation of SIV in macaques. They revealed a previously unrecognized epitope complexity and, after having validated their approach on two CTL epitopes, also demonstrated that their method can be extended to the complete viral genome through the use of overlapping amplicons. This sequence data set was used by Hughes, et al.[114] to infer a relationship between the frequency of the infecting variant and the time after infection, which permits the accurate reconstruction of the transmitted viral sequence based on data from two or more postinfection time points. The

same dataset was expanded by Love, et al.[115], whereby the depth of coverage permitted them to more accurately model the kinetics of CTL-mediated adaptation. They provide further evidence for the CTL response as an important driving factor of viral evolution during the acute phase of SIV infection. The authors also speculate that their results may imply that the initial infectious viral reservoir is preferentially established by the transmitted variant, which would provide a possible explanation of why the CTL response cannot completely eliminate the infecting variants.

Furthermore, correct mapping of immune escape pathways will require the identification of co-evolutionary events[116], a research area where, similar to the aforementioned analysis performed by Rozera, et al.[86], the combination of phylogenetics with 454 sequence data may be of value, as demonstrated by Poon, et al.[49]. Interestingly, their results suggest that inter-host variability estimated from population-based sequencing results can provide sufficient information to aid in the selection of the most informative regions for further "deep" investigation of epistatic interactions at the intra-host level.

As the tools for comprehensively assessing escape pathways are now available, our knowledge on HIV dynamics in response to HLA class I and class II selective pressure and other host responses will undoubtedly be refined in the near future, hopefully with implications for rational vaccine design and ultimately also for patient care[116].

Conclusion/future perspectives

Next-generation sequencing is truly transforming genetic research, and in the field of HIV, the 454 sequencing platform has opened the doors to study the hidden world of minor variants in unprecedentedly comprehensive ways. Parallelization of the sequencing process in combination with the high sensitivity and good read length offers an invaluable tool for increasing the sequence information of viral populations. Not only does it render their characterization more cost effective, but Roche 454 also allows a more accurate quantification of TDR[74] and, by aiming at a minimal coverage of the targeted genomic region, many samples can be interrogated simultaneously[117].

Not surprisingly, our review has a particular focus on HIV drug resistance applications; this research has pioneered amplicon sequencing using 454 because of the clear benefits of its deep-sequencing capacity. The second dimension we initially outlined, the complete genome perspective, has been comparatively less exploited for HIV. However, since further cost reductions are expected in the coming years, the potential advantages for epidemiological studies, which are often limited by genomic marker size, are enormous. A glimpse of the opportunities on this level is provided by the near full-length characterization of a BF recombinant[118]. For other viral pathogens such as influenza, epidemiological surveillance is already benefitting from the next generation sequencing era[119].

In the absence of a vaccine, HAART remains the only option in the near future to delay the onset of AIDS. The advent of drugs from new classes, and the concomitant need to sequence additional genomic regions in genotyping assays, will render genotyping an increasingly time-consuming and labor-intensive task. Although many technical hurdles may still need to be overcome, this may present the opportunity for the 454 pyrosequencing strategy to be adapted for simultaneously "deep" genotyping of all regions of interest for a multitude of samples in a single experiment.

More generally, 454 pyrosequencing offers a suitable technique to gain more profound insights into the processes that govern the genesis and preservation of genetic diversity in viral populations under different selective pressures, both host-and therapy-imposed. With respect to the latter, it is important to keep in mind that drugs differ in their genetic barrier to resistance[120]. This implies that the clinical relevance of minority DRM can differ between regimens[58,121] and diverse prevalence cutoffs may need to be determined. The NNRTI, for example, are known to have a relatively low genetic barrier to resistance and NNRTI-based regimens can therefore expected to be relatively sensitive to minority DRM, as indicated by recent findings[56,77,122].

As noted by Bushman, et al.[5], 454 pyrosequencing can be applied to all viruses that conceptually face the same problems as in HIV research. The hepatitis C virus (HCV) represents an medically important candidate that has an intra-host population structure at least as complex as HIV. Drugs targeting specific viral protein functions are under investigation, but given the high error rate of the HCV RNA polymerase, resistance is expected to develop in a similar fashion as for HIV. A similar picture is also emerging regarding the adaptation of HCV to the immune response, where the relative role of the humoral and cellular immunity in controlling viral replication is yet to be fully disentangled[123].

In conclusion, we anticipate that the advantages of increasing sequencing output will find fertile ground in a multidisciplinary environment, and complemented with dedicated bioinformatics developments, the explosion of new quantitative and qualitative analyses promises more detailed insights in the relationship between infectious disease, pathogen evolution, and epidemiological processes.

Acknowledgments

B. Vrancken and K. Theys were supported by a Ph.D. grant from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT Vlaanderen). P. Lemey was supported by a postdoctoral fellowship from the Fund for Scientific Research (FWO) Flanders.

We thank A. Rambaut for data compilation of a GS20 pilot run and the permission to use the resulting alignment in figure 1.

We thank D. Robertson for providing the original version of figure 3.

454, GS 20, GS FLX, GS FLX Titanium are trademarks of Roche.

References

1. Sanger F, Coulson A. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol. 1975;94:441-8.
2. Sanger F, Nicklen S, Coulson A. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA. 1977;74:5463-7.
3. Bayley H. Sequencing single molecules of DNA. Curr Opin Chem Biol. 2006;10:628-37.
4. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008;26:1135-45.
5. Bushman F, Hoffmann C, Ronen K, et al. Massively parallel pyrosequencing in HIV research. AIDS. 2008;22:1411-5.
6. MacLean D, Jones J, Studholme D. Application of 'next-generation' sequencing technologies to microbial genetics. Nat Rev Microbiol. 2009;7:287-96.
7. Voelkerding K, Dames S, Durtschi J. Next-generation sequencing: from basic research to diagnostics. Clin Chem. 2009;55:641-58.
8. Metzker ML. Sequencing technologies -the next generation. Nat Rev Genet. 2010;11:31-46.
9. Margulies M, Egholm M, Altman W, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437:376-80.
10. Mansky LM. Forward mutation rate of human immunodeficiency virus type 1 in a T lymphoid cell line. AIDS Res Hum Retroviruses. 1996;12: 307-14.
11. Hert D, Fredlake C, Barron A. Advantages and limitations of next-generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. Electrophoresis. 2008;29:4618-26.
12. Parameswaran P, Jalili R, Tao L, et al. A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. Nucleic Acids Res. 2007;35:e130.
13. Huse S, Huber J, Morrison H, Sogin M, Welch D. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol. 2007;8:R143.
14. Hoffmann C, Minkah N, Leipzig J, et al. DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. Nucleic Acids Res. 2007;35:e91.
15. Droege M, Hill B. The Genome Sequencer FLX System–longer reads, more applications, straight forward bioinformatics and more complete data sets. J Biotechnol. 2008;136:3-10.
16. Quinlan A, Stewart D, Strömberg M, Marth G. Pyrobayes: an improved base caller for SNP discovery in pyrosequences. Nat Methods. 2008;5:179-81.
17. Brockman W, Alvarez P, Young S, et al. Quality scores and SNP detection in sequencing-by-synthesis systems. Genome Res. 2008;18:763-70.
18. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 1998;8:186-94.
19. Hamady M, Walker J, Harris J, Gold N, Knight R. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat Methods. 2008;5:235-7.
20. Horner D, Pavesi G, Castrignanò T, et al. Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. Brief Bioinform. 2010;11:181-97.
21. Archer J, Braverman M, Taillon B, et al. Detection of low-frequency pretherapy chemokine (CXC motif) receptor 4 (CXCR4)-using HIV-1 with ultra-deep pyrosequencing. AIDS. 2009;23:1209-18.
22. Eriksson N, Pachter L, Mitsuya Y, et al. Viral population estimation using pyrosequencing. PLoS Comput Biol. 2008;4:e1000074.
23. Zagordi O, Geyrhofer L, Roth V, Beerenwinkel N. Deep sequencing of a genetically heterogeneous sample: local haplotype reconstruction and read error correction. J Comput Biol. 2010;17:417-28.
24. Shao W, Boltz V, Kearney M, et al. Characterization of HIV-1 sequence artifacts introduced by bulk PCR and detected by 454 sequencing. Antiviral Therapy. 2009;14(Suppl 1):A123.
25. Williamson S. Adaptation in the env gene of HIV-1 and evolutionary theories of disease progression. Mol Biol Evol. 2003;20:1318-25.
26. Coffin J. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science. 1995;267:483-9.
27. Vandamme A, Sönnerborg A, Ait-Khaled M, et al. Updated European recommendations for the clinical use of HIV drug resistance testing. Antivir Ther. 2004;9:829-48.
28. Hirsch M, Günthard H, Schapiro J, et al. Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society-USA panel. Clin Infect Dis. 2008;47:266-85.
29. Charpentier C, Dwyer D, Mammano F, Lecossier D, Clavel F, Hance AJ. Role of minority populations of human immunodeficiency virus type 1 in the evolution of viral resistance to protease inhibitors. J Virol. 2004;78:4234-47.
30. Lecossier D, Shulman N, Morand-Joubert L, et al. Detection of minority populations of HIV-1 expressing the K103N resistance mutation in patients failing nevirapine. J Acquir Immune Defic Syndr. 2005;38:37-42.
31. Briones C, de Vicente A, Molina-París C, Domingo E. Minority memory genomes can influence the evolution of HIV-1 quasispecies in vivo. Gene. 2006;384:129-38.
32. Metzner K, Giulieri S, Knoepfel S, et al. Minority quasispecies of drug-resistant HIV-1 that lead to early therapy failure in treatment-naive and -adherent patients. Clin Infect Dis. 2009;48:239-47.
33. Balduin M, Oette M, Däumer M, Hoffmann D, Pfister H, Kaiser R. Prevalence of minor variants of HIV strains at reverse transcriptase position 103 in therapy-naïve patients and their impact on the virological failure. J Clin Virol. 2009;45:34-8.
34. Johnson J, Li J, Wei X, et al. Minority HIV-1 drug resistance mutations are present in antiretroviral treatment-naïve populations and associate with reduced treatment efficacy. PLoS Med. 2008;5:e158.
35. Rowley C, Boutwell C, Lee E, et al. Ultrasensitive detection of minor drug-resistant variants for HIV after nevirapine exposure using allelespecific PCR: clinical significance. AIDS Res Hum Retroviruses. 2010;26:293-300.
36. Peuchant O, Thiébaut R, Capdepont S, et al. Transmission of HIV-1 minority-resistant variants and response to first-line antiretroviral therapy. AIDS. 2008;22:1417-23.
37. Günthard H, Wong J, Ignacio C, Havlir D, Richman D. Comparative performance of high-density oligonucleotide sequencing and dideoxynucleotide sequencing of HIV type 1 pol from clinical samples. AIDS Res Hum Retroviruses. 1998;14:869-76.
38. Hance A, Lemiale V, Izopet J, et al. Changes in human immunodeficiency virus type 1 populations after treatment interruption in patients failing antiretroviral therapy. J Virol. 2001;75:6410-17.
39. Van Laethem K, Van Vaerenbergh K, Schmit J, et al. Phenotypic assays and sequencing are less sensitive than point mutation assays for detection of resistance in mixed HIV-1 genotypic populations. J Acquir Immune Defic Syndr. 1999;22:107-18.
40. Shafer RW. Low-abundance drug-resistant HIV-1 variants: finding significance in an era of abundant diagnostic and therapeutic options. J Infect Dis. 2009;199:610-12.
41. Cai F, Chen H, Hicks C, Bartlett J, Zhu J, Gao F. Detection of minor drug-resistant populations by parallel allele-specific sequencing. Nat Methods. 2007;4:123-5.
42. Halvas E, Aldrovandi G, Balfe P, et al. Blinded, multicenter comparison of methods to detect a drug-resistant mutant of human immunodeficiency virus type 1 at low frequency. J Clin Microbiol. 2006;44:2612-14.
43. Lalonde M, Troyer R, Syed A, et al. Sensitive oligonucleotide ligation assay for low-level detection of nevirapine resistance mutations in human immunodeficiency virus type 1 quasispecies. J Clin Microbiol. 2007;45:2604-15.

44. Detsika M, Chandler B, Khoo S, et al. Detection and quantification of minority HIV isolates harbouring the D30N mutation by real-time PCR amplification. J Antimicrob Chemother. 2007;60:881-4.
45. Kapoor A, Jones M, Shafer R, Rhee S, Kazanjian P, Delwart E. Sequencing-based detection of low-frequency human immunodeficiency virus type 1 drug-resistant mutants by an RNA/DNA heteroduplex generator-tracking assay. J Virol. 2004;78:7112-23.
46. Liu S, Rodrigo A, Shankarappa R, et al. HIV quasispecies and resampling. Science. 1996;273:415-16.
47. Hughes J, Totten P. Estimating the accuracy of polymerase chain reaction-based tests using endpoint dilution. Biometrics. 2003;59: 505-11.
48. Paredes R, Marconi V, Campbell T, Kuritzkes D. Systematic evaluation of allelespecific real-time PCR for the detection of minor HIV-1 variants with pol and env resistance mutations. J Virol Methods. 2007;146: 136-46.
49. Poon A, Swenson L, Dong W, et al. Phylogenetic analysis of population-based and deep sequencing data to identify coevolving sites in the nef gene of HIV-1. Mol Biol Evol. 2010;27:819-32.
50. Karrer E, Lincoln J, Hogenhout S, In situ isolation of mRNA from individual plant cells: creation of cell-specific cDNA libraries. Proc Natl Acad Sci USA. 1995;92:3814-18.
51. Elnifro E, Ashshi A, Cooper R, Klapper P. Multiplex PCR: optimization and application in diagnostic virology. Clin Microbiol Rev. 2000; 13:559-70.
52. Vandenbroucke I, Marck H, Mostmans W, et al. HIV-1 V3 envelope deep sequencing for clinical plasma specimens failing in phenotypic tropism assays. AIDS Res Ther. 2010;7.
53. Tsibris A, Korber B, Arnaout R, et al. Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo. PLoS One. 2009;4:e5683.
54. Bracho M, García-Robles I, Jiménez N, Torres-Puente M, Moya A, González-Candelas F. Effect of oligonucleotide primers in determining viral variability within hosts. Virol J. 2004;1:13.
55. Wang C, Mitsuya Y, Gharizadeh B, Ronaghi M, Shafer RW. Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. Genome Res. 2007;17:1195-201.
56. Simen B, Simons J, Hullsiek K, et al. Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. J Infect Dis. 2009;199: 693-701.
57. Moorthy A, Kuhn L, Coovadia A, et al. Plasma frequencies of nevirapine resistance influence virologic responses to nevirapine maintenance therapy in single-dose exposed HIV-infected children initially treated with lopinavir HAART. 17th CROI - Program & Abstracts. 2010; San Fransisco [abstract 159].
58. Paredes R, Clotet B. Clinical management of HIV-1 resistance. Antiviral Res. 2010;85:245-65.
59. Le T, Chiarella J, Simen B, et al. Low-abundance HIV drug-resistant viral variants in treatment-experienced persons correlate with historical antiretroviral use. PLoS One. 2009;4:e6079.
60. FDA. FDA notifications. FDA approves raltegravir for HIV-1 treatment-naive patients. AIDS Alert. 2009;24:106-7.
61. Ceccherini-Silberstein F, Armenia D, D'Ariggo R, et al. Baseline variability of HIV-1 integrase in multi-experienced patients treated with raltegravir: a refined analysis by pyrosequencing. 16th CROI - Program & Abstracts. 2009; Alexandria [abstract 682].
62. Ceccherini-Silberstein F, Armenia D, Vandenbroucke I, et al. Selection and coevolution of mutational patterns in the HIV-1 integrase characterized by pyrosequencing analysis. Antiviral Therapy. 2009;14(Suppl 1):A125.
63. Ceccherini-Silberstein F, Armenia D, Vandenbroucke I, et al. Primary mutations associated with resistance to raltegravir are not detectable by pyrosequencing in integrase inhibitors-naive patients. 17th CROI -Program & Abstracts. 2010; San Fransisco [abstract 566].
64. Ceccherini-Silberstein F, Van Baelen K, Armenia D, et al. Secondary HIV-1 integrase resistance mutations, found as minority quasispecies in integrase therapy naive patients, have little or no effect on susceptibility to integrase inhibitors. Antimicrob Agents Chemother. 2010. [Epub ahead of print].
65. Codoñer F, Pou C, Thielen A, et al. Dynamic escape of pre-existing raltegravir-resistant HIV-1 from raltegravir pressure. 8th European HIV Drug Resistance Workshop [poster]. 2010; Sorrento, Italy.
66. Charpentier C, Laureillard D, Piketty C, et al. High frequency of integrase Q148R minority variants in HIV-infected patients naive of integrase inhibitors. AIDS. 2010;24:867-73.
67. Fox J, Dustan S, McClure M, Weber J, Fidler S. Transmitted drug-resistant HIV-1 in primary HIV-1 infection; incidence, evolution and impact on response to antiretroviral therapy. HIV Med. 2006 Oct;7:477-83.
68. Little S, Holte S, Routy J, et al. Antiretroviral drug resistance among patients recently infected with HIV. N Engl J Med. 2002;347:385-94.
69. SPREAD programme. Transmission of drug-resistant HIV-1 in Europe remains limited to single classes. AIDS. 2008;22:625-35.
70. Booth C, Geretti A. Prevalence and determinants of transmitted antiretroviral drug resistance in HIV-1 infection. J Antimicrob Chemother. 2007;59:1047-56.
71. Carr J, Green T, Shaw D, et al. Application of an allelespecific PCR to clinical HIV genotyping samples detects additional K103N mutations in both therapy naïve and experienced patients. J Med Virol. 2009;81: 1983-90.
72. Deeks S. Transmitted minority drug-resistant HIV variants: a new epidemic? PLoS Med. 2008;5:e164.
73. Johnson J, Geretti A. Low-frequency HIV-1 drug resistance mutations can be clinically significant but must be interpreted with caution. J Antimicrob Chemother. 2010;65:1322-6.
74. Ji H, Massé N, Tyler S, et al. HIV drug resistance surveillance using pooled pyrosequencing. PLoS One. 2010;5:e9263.
75. Lataillade M, Chiarella J, Yang R, et al. Prevalence and clinical significance of transmitted drug-resistant (TDR) HIV mutations by ultra-deep sequencing (UDS) in HIV-infected ARV-naive subjects in CASTLE Study. Antiviral Therapy. 2009;14(Suppl 1):A44.
76. Paredes R, Lalama C, Ribaudo H, et al. Preexisting minority drug-resistant HIV-1 variants, adherence, and risk of antiretroviral treatment failure. J Infect Dis. 2010;201:662-71.
77. Metzner K, Rauch P, von Wyl V, et al. Efficient suppression of minority drug-resistant HIV type 1 (HIV-1) variants present at primary HIV-1 infection by ritonavir-boosted protease inhibitor-containing antiretroviral therapy. J Infect Dis. 2010;201:1063-71.
78. Varghese V, Shahriar R, Rhee S, et al. Minority variants associated with transmitted and acquired HIV-1 nonnucleoside reverse transcriptase inhibitor resistance: implications for the use of second-generation nonnucleoside reverse transcriptase inhibitors. J Acquir Immune Defic Syndr. 2009;52:309-15.
79. Mitsuya Y, Varghese V, Wang C, et al. Minority human immunodeficiency virus type 1 variants in antiretroviral-naive persons with reverse transcriptase codon 215 revertant mutations. J Virol. 2008;82: 10747-55.
80. Esté J, Telenti A. HIV entry inhibitors. Lancet. 2007;370:81-8.
81. Hammer S, Eron J, Reiss P, et al. Antiretroviral treatment of adult HIV infection: 2008 recommendations of the International AIDS Society-USA panel. JAMA. 2008;300:555-70.
82. Vercauteren J, Vandamme A. Algorithms for the interpretation of HIV-1 genotypic drug resistance information. Antiviral Res. 2006;71: 335-42.
83. Lengauer T, Sander O, Sierra S, Thielen A, Kaiser R. Bioinformatics prediction of HIV coreceptor usage. Nat Biotechnol. 2007;25:1407-10.
84. Low A, Dong W, Chan D, et al. Current V3 genotyping algorithms are inadequate for predicting X4 co-receptor usage in clinical isolates. AIDS. 2007;21:F17-24.
85. Rose J, Rhea A, Weber J, Quiñones-Mateu M. Current tests to evaluate HIV-1 coreceptor tropism. Curr Opin HIV AIDS. 2009;4:136-42.
86. Rozera G, Abbate I, Bruselles A, et al. Massively parallel pyrosequencing highlights minority variants in the HIV-1 env quasispecies deriving from lymphomonocyte sub-populations. Retrovirology. 2009;6:15.
87. Rozera G, Abbate I, Bruselles A, et al. Archived HIV-1 minority variants detected by ultra-deep pyrosequencing in provirus may be fully replication competent. AIDS. 2009;23:2541-3.
88. Abbate I, Rozera G, Bruselles A, et al. Quantitative deep sequencing reveals hidden lineages and shifts of viral population in Archived and Replicating HIV-1 Genomes. 17th CROI - Program & Abstracts. 2010; San Fransisco [abstract 269].
89. Swenson L, Dong W, Mo T, et al. Quantification of HIV tropism by «deep» sequencing shows a broad distribution of prevalence of X4 variants in clinical samples that is associated with virological outcome. Global Antivir J. 2008;4(Suppl 1):88.
90. van 't Wout A, Swenson L, Welkers M, Dong W, Schuitemaker H, Harrigan P. Detection of CXCR4-using HIV-1 variants in longitudinally obtained paired plasma and PBMC samples using 454-sequencing. Antiviral Therapy. 2009;14(Suppl 1):A88.
91. van 't Wout A, Swenson L, Dong W, Schuitemaker H, Harrigan R. Detection of predicted CXCR4-using HIV-1 variants in longitudinally obtained paired plasma and peripheral blood mononuclear cells samples using 454-sequencing. 17th CROI - Program & Abstracts. 2010; San Fransisco [abstract 276].
92. Van Baelen K, Rondelez E, Wesenbeeck L, et al. Population and clonal tropism phenotyping of clinical isolates and comparison with next generation sequencing and prediction tools. Antiviral Therapy. 2009;14(Suppl 1):A134.
93. Swenson L, Dong W, Mo T, et al. Large-scale application of deep sequencing using 454 technology to HIV tropism screening. 17th CROI -Program & Abstracts. 2010; San Fransisco [abstract 545].
94. Pou C, Codoner F, Thielen A, et al. High resolution tropism kinetics by quantitative deep sequencing in HIV-1-infected subjects initiating suppressive first-line ART. 17th CROI -Program & Abstracts. 2010; San Fransisco [abstract 544].
95. Swenson L, Dong W, Mo T, et al. Quantification of HIV tropism by «deep» sequencing shows a broad distribution of prevalence of X4 variants in clinical samples that is associated with virological outcome. 16th CROI -Program & Abstracts. 2009; Alexandria [abstract 680].

96. Vandekerckhove L, Verhofstede C, Vandenbroucke I, et al. Comparison of population and 454 genotyping and Trofile phenotyping for tropism determination in patients selected for maraviroc initiation. Antiviral Therapy. 2009;14(Suppl 1):A131.
97. Gilden D, Kort R. XVII International AIDS Conference: From Evidence to Action -Basic science. J Int AIDS Soc. 2009;12(Suppl 1):S3.
98. Redel L, Le Douce V, Cherrier T, et al. HIV-1 regulation of latency in the monocyte-macrophage lineage and in CD4+ T lymphocytes. J Leukoc Biol. 2010;87:575-88.
99. Coiras M, Lopez-Huertas M, Alcami J. HIV-1 latency and eradication of long-term viral reservoirs. Discov Med. 2010;9:185-91.
100. Ciuffi A, Ronen K, Brady T, et al. Methods for integration site distribution analyses in animal cell genomes. Methods. 2009;47:261-8.
101. Wang G, Garrigue A, Ciuffi A, et al. DNA bar coding and pyrosequencing to analyze adverse events in therapeutic gene transfer. Nucleic Acids Res. 2008;36:e49.
102. Wang G, Levine B, Binder G, et al. Analysis of lentiviral vector integration in HIV+ study subjects receiving autologous infusions of gene modified CD4+ T cells. Mol Ther. 2009;17:844-50.
103. Marshall H, Ronen K, Berry C, et al. Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. PLoS One. 2007;2:e1340.
104. Ocwieja K, Konen R, Brady T, Berry C, Bushman F. Novel host factors in HIV integration site targeting. 17th CROI -Program & Abstracts. 2010; San Fransisco [abstract 251].
105. Ferris A, Wu X, Hughes C, et al. Lens epithelium derived growth factor fusion proteins redirect HIV-1 DNA integration. Proc Natl Acad Sci USA. 2010;107:3135-40.
106. Ronen K, Gijsbers R, Vets S, et al. Use of LEDGF/p75 fusion proteins to retarget lentiviral integration outside of genes. 17th CROI -Program & Abstracts. 2010; San Fransisco [abstract 145].
107. Brady T, Agosto L, Malani N, Berry C, O'Doherty U, Bushman F. HIV integration site distributions in resting and activated CD4+ T cells infected in culture. AIDS. 2009;23:1461-71.
108. Goulder PJR, Watkins DI. Impact of MHC class I diversity on immune control of immunodeficiency virus replication. Nat Rev Immunol. 2008;8:619-30.
109. Wiseman R, Karl J, Bimber B, et al. Major histocompatibility complex genotyping with massively parallel pyrosequencing. Nat Med. 2009;15:1322-6.
110. Gabriel C, Danzer M, Hackl C, et al. Rapid high-throughput human leukocyte antigen typing by massively parallel pyrosequencing for high-resolution allele identification. Hum Immunol. 2009;70:960-4.
111. Bentley G, Higuchi R, Hoglund B, et al. High resolution, high-throughput HLA genotyping by next-generation sequencing. Tissue Antigens. 2009;74:393-403.
112. Jia X, Lennon R, Henn M, et al. Next-generation sequencing for HLA-typing and MHC sequence interrogation for host genetics of HIV infection. 17th CROI -Program & Abstracts. 2010; San Fransisco [abstract 452].
113. Bimber B, Burwitz B, O'Connor S, Ultradeep pyrosequencing detects complex patterns of CD8+ T-lymphocyte escape in simian immunodeficiency virus-infected macaques. J Virol. 2009;83:8247-53.
114. Hughes A, O'Connor S, Dudley D, Burwitz B, Bimber B, O'Connor D. Dynamics of haplotype frequency change in a CD8+TL epitope of simian immunodeficiency virus. Infect Genet Evol. 2010;10:555-60.
115. Love T, Thurston S, Keefer M, Dewhurst S, Lee H. Mathematical modeling of ultradeep sequencing data reveals that acute CD8+ T-lymphocyte responses exert strong selective pressure in simian immunodeficiency virus-infected macaques, but still fail to clear founder epitope sequences. J Virol. 2010; [in press].
116. Carlson J, Brumme Z. HIV evolution in response to HLA-restricted CTL selection pressures: a population-based perspective. Microbes Infect. 2008;10:455-61.
117. Djikeng A, Spiro D. Advancing full length genome sequencing for human RNA viral pathogens. Future Virol. 2009;4:47-53.
118. Bruselles A, Rozera G, Bartolini B, et al. Use of massive parallel pyrosequencing for near full-length characterization of a unique HIV Type 1 BF recombinant associated with a fatal primary infection. AIDS Res Hum Retroviruses. 2009;25:937-42.
119. Höper D, Hoffmann B, Beer M. Simple, sensitive, and swift sequencing of complete H5N1 avian influenza virus genomes. J Clin Microbiol. 2009;47:674-9.
120. Clavel F, Hance A. HIV drug resistance. N Engl J Med. 2004;350: 1023-35.
121. Ceccherini-Silberstein F, Svicher V, Van Houtte M. New horizons in resistance interpretation. Eur Infect Dis. 2009;3:36-40.
122. Halvas E, Wiegand A, Boltz V, et al. Low frequency nonnucleoside reverse-transcriptase inhibitor-resistant variants contribute to failure of efavirenz-containing regimens in treatment-experienced patients. J Infect Dis. 2010;201:672-80.
123. Timm J, Roggendorf M. Sequence diversity of hepatitis C virus: implications for immune control and therapy. World J Gastroenterol. 2007;13:4808-17.
124. Morozova O, Marra M. Applications of next-generation sequencing technologies in functional genomics. Genomics. 2008;92:255-64.