1 : ***Duffy,S. et al. (2008),*** show that the high rate of nucleotide substitution in RNA viruses is matched by some DNA viruses, suggesting that evolutionary rates in viruses are explained by diverse aspects of viral biology, such as genomic architecture and replication speed, and not simply by polymerase fidelity.

2 : ***Lauring and Andino, 2010***, This diversity allows a viral population to rapidly adapt to dynamic environments and evolve resistance to vaccines and antiviral drugs. quasispecies theory has provided a population-based framework for understanding RNA viral evolution. Here, we discuss basic principles of quasispecies theory and describe its relevance for our understanding of viral fitness, virulence, and antiviral therapeutic strategy.

3 : ***Rozera et al., 2014*** , Seventeen HIV-1-infected combined antiretroviral therapy naive patients were enrolled . As markers of immune activation, plasma sCD14 and soluble tumour necrosis fact or receptor II levels were measured. Median diversity of HIV RNA was lower in patients with early infection versus chronic infection patients. The loss of gut/PBMC compartmentalization in more advanced stages of HIV infection was confirmed by longitudinal observation.

4: ***Kuroda et al., 2010,*** This study demonstrated that de novo sequencing can comprehensively detect pathogens, and such in-depth investigation facilitates the identification of influenza A viral heterogeneity. To better characterize the A/H1N1/2009 virus, unbiased comprehensive techniques will be indispensable for the primary investigations of emerging infectious diseases.

5: *Gaschen, 2002,* To contend with the diversity, country-specific vaccines are being considered, but evolutionary relationships may be more useful than regional considerations, Consensus or ancestor sequences could be used in vaccine design to minimize the genetic differences between vaccine strains and contemporary isolates, effectively reducing the extent of diversity by half .

6: ; ***Capobianchi et al., 2013,*** NGS has been applied to metagenomics-based strategies for the discovery of novel viruses and the characterization of viral communities. These applications are particularly suitable for viruses such as human immunodeficiency virus, hepatitis B virus, and hepatitis C virus, whose error-prone replication machinery, combined with the high replication rate, results, in each infected individual, in the formation of many genetically related viral variants referred to as quasi-species. With traditional approaches, it is difficult to detect and quantify minority genomes present in viral quasi-species that, in fact, may have biological and clinical relevance.

7: *Goodwin et al., 2016*. By contrast, long-read approaches provide read lengths that are well suited for de novo genome assembly applications and full-length isoform sequencing. Recent improvements in chemistry, costs, throughput and accessibility are driving the emergence of new, varied technologies to address applications that were not previously possible. These include integrated long-read and short-read sequencing studies, routine clinical DNA sequencing, real-time pathogen DNA monitoring and massive population-level projects. Many new instruments with varied chemistries and applications are being released or being developed

8: *Beerenwinkel et al., 2012,* e discuss sample preparation, including reverse transcription and amplification, and the effect of experimental conditions on diversity estimates due to in vitro base substitutions, insertions, deletions, and recombination. The use of different NGS platforms and their sequencing error profiles are compared in the context of various applications of diversity estimation, ranging from the detection of single nucleotide variants to the reconstruction of whole-genome haplotypes .

9: *Ho and Tzanetakis, 2014,*) Next generation sequencing has revolutionized virus discovery. The pipeline was used to process more than 30 samples resulting in the detection of all viruses known to infect the processed samples, the extension of the genomic sequences of others, and the discovery of several novel viruses. VirFind was tested by four external users with datasets from plants or insects, demonstrating its potential as a universal virus detection and discovery tool.

10: *Li et al., 2016,* Here they describe VIP , a one-touch computational pipeline for virus identification and discovery from metagenomic NGS data. they validated the feasibility and veracity of this pipeline with sequencing results of various types of clinical samples and public datasets. VIP has also contributed to timely virus diagnosis in acutely ill patients, demonstrating its potential in the performance of unbiased NGS-based clinical studies with demand of short turnaround time

11: *Maarala et al., 2018,* they propose ViraPipe, a scalable metagenome analysis pipeline that is able to analyze thousands of human microbiomes in parallel in tolerable time. They show the scalability of ViraPipe by running experiments on mining virus related genomes from NGS datasets in a distributed Spark computing cluster.

12: *Wan et al., 2015:* their pipeline allows users to assemble, analyze, and interpret high coverage viral sequencing data with an ease and efficiency that was not possible previously. their software makes a large number of genome assembly and related tools available to life scientists and automates the currently recommended best practices into a single, easy to use interface. Their tested our pipeline with three different datasets from human herpes simplex virus.

13: ; *Taylor et al*, All HIVDR mutations identifiable by SS were detected by the MiSeq-HyDRA protocol, while LADRVs at frequencies of 1~15% were detected by MiSeq-HyDRA only.

14:) *Wymantet al., 2018,* *D*e novo assembly avoids this bias by aligning the reads to themselves, producing a set of sequences called contigs. However contigs provide only a partial summary of the reads, misassembly may result in their having an incorrect structure, and no information is available at parts of the genome where contigs could not be assembled. To address these problems we developed the tool shiver to pre-process reads for quality and contamination, then map them to a reference tailored to the sample using corrected contigs supplemented with the user's choice of existing reference sequnces.

15: *lee2020,*) Next generation sequencing is a trending new standard for genotypic HIV-1 drug resistance testing. Many NGS HIVDR data analysis pipelines have been independently developed, each with variable outputs and data management protocols. Here we compared the performance of five NGS HIVDR pipelines using proficiency panel samples from NIAID Virology Quality Assurance program. All pipelines detected amino acid variants at full range of frequencies and demonstrated good linearity as compared to the reference frequency values.

16:*Eliseev et al., 2020,* We simulated coalescent-based populations that spanned known levels of viral genetic diversity, including mutation rates, sample size and effective population size, to test the limits of the haplotype reconstruction methods and to ensure coverage of predicted intra-host viral diversity levels . However, under higher levels of diversity , haplotype reconstruction quality was highly variable and, on average, poor. All haplotype reconstruction tools, except QuasiRecomb and ShoRAH, greatly underestimated intra-host diversity and the true number of haplotypes.

17: *) Di Giallonardo,F. et al,* To study complex heterogeneous virus populations are comprehensively, novel methods are required that allow for complete reconstruction of the individual viral haplotypes. Here, we show that assembly of whole viral genomes of ~8600 nucleotides length is feasible from mixtures of heterogeneous HIV-1 strains derived from defined combinations of cloned virus strains and from clinical samples of an HIV-1 superinfected individual. Haplotype reconstruction was achieved using optimized experimental protocols and computational methods for amplification, sequencing and assembly.

18: *Zanini et al.2015,* **.(** We show that patterns of minor diversity are reproducible between patients and mirror global HIV-1 diversity, suggesting a universal landscape of fitness costs that control diversity. Frequent recombination limits linkage disequilibrium

to about 100bp in most of the genome, but strong hitch-hiking due to short range linkage limits diversity