Published in final edited form as:
Annu Rev Virol. 2020 September 29; 7(1): 63–81. doi:10.1146/annurev-virology-010320-061642.

Within-host Viral Diversity, a Window into Viral Evolution

Adam S. Lauring[1,2]

[1]Division of Infectious Diseases, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan

[2]Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan

Abstract

The evolutionary dynamics of a virus can differ within hosts and across populations. Studies of within-host evolution provide an important link between experimental studies of virus evolution and large-scale phylodynamic analyses. They can define the extent to which global processes are recapitulated on local scales and how accurately experimental infections model natural ones. They may also inform epidemiologic models of disease spread and reveal how host-level dynamics contribute to a virus' evolution at a larger scale. Over the last decade, advances in viral sequencing have enabled detailed studies of viral genetic diversity within hosts. Here, I review how within-host diversity is sampled, measured, and expressed, and how comparative studies of viral diversity can be leveraged to elucidate a virus' evolutionary dynamics. These concepts are illustrated with detailed reviews of recent work on the within-host evolution of influenza virus, Dengue virus and cytomegalovirus.

Keywords

diversity; evolution; sequencing; models; quasispecies

INTRODUCTION

It has been just over 40 years since Domingo and colleagues used T1 RNAse digestion and two- dimensional electrophoresis to demonstrate the extraordinary genetic diversity of an RNA phage population (1). With the advent of DNA sequencing, it was quickly recognized that most RNA viruses exist as genetically diverse populations, both in vitro and an vivo, e.g., (2–5). Over the last decade, so-called "next generation" sequencing technologies, have enabled detailed studies of viral diversity within infected hosts (6–8). These studies have in turn provided fundamental insights into viral evolutionary dynamics within hosts and their relationship to the epidemiology and evolution of viruses at larger scales (Figure 1).

Mutation is the ultimate source of genetic diversity. Nearly all viral RNA-dependent RNA polymerases (RdRp) lack proofreading or repair activities, and for most RNA viruses, this

alauring@med.umich.edu .

DISCLOSURE STATEMENT
The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

translates to a mutation rate of 10^{-6} to 10^{-4} substitutions per nucleotide per cellular infection (9, 10). While the cellular and viral polymerases involved in DNA virus replication have higher fidelity, DNA viruses can still exhibit significant mutational diversity, even within hosts (11). Cellular cytidine deaminases can also introduce specific mutations (e.g. C to U, G to A) into viral genomes (12, 13). Recombination among genomes is another important source of genetic diversity in plus stranded RNA viruses, retroviruses, and DNA viruses (14). It can lead to novel mutation combinations, gene amplification, and defective viral genomes. While recombination is much less common in negative stranded RNA viruses, viruses with segmented genomes can generate combinatorial diversity through reassortment (15).

Broadly speaking, the fate of mutations or genomic variants depends on both deterministic processes, such as selection, and stochastic processes, such as genetic drift (16–18). Natural selection will tend to increase the frequency of beneficial mutations in a population – positive selection – and decrease the frequency of detrimental ones – negative, or purifying, selection. Genetic drift refers to changes in the frequencies of variants in a population due to random sampling, which is particularly prominent in small populations. Recent work suggests that genetic drift plays an important role in shaping within-host viral diversity, as populations frequently experience transient reductions in population size, or bottlenecks (19, 20). The relative contribution of selection and drift is greatly affected by the effective population size, a model parameter that roughly corresponds to the number of individuals in a population that contribute mutations or variants to the next generation (16). If the effective population size of a virus is large, as in quasispecies models, evolution is largely deterministic and the frequency of a mutation can be predicted based on its starting frequency and selection coefficient. In small populations, selection is inefficient, and changes in mutation frequency are strongly influenced by bottlenecks and genetic drift.

Here, I review our current understanding of within-host viral diversity, with a heavy emphasis on studies performed over the last decade. I will first discuss how within-host diversity is sampled, identified, and quantified, as well as the strengths and pitfalls of various approaches. I will then describe how both cross-sectional and serial sampling of within-host viral populations can inform our understanding of natural selection and genetic drift. To illustrate these points, I will review recent work on the within-host diversity of influenza virus, Dengue virus, and cytomegalovirus (CMV), with a focus on data acquired from naturally-infected individuals. Finally, I will re-examine whether quasispecies-inspired hypotheses regarding the importance of genetic diversity are supported by recent empiric data on within-host populations.

ASSESSING VIRAL DIVERSITY

Over the last decade, advances in sequencing technology have revolutionized studies of viral diversity, particularly within hosts (Figure 2*a*). Most recent studies have used next generation sequencing platforms (NGS), which allow one to sequence a virus at sufficiently high depth of coverage to identify both the consensus sequence of the population and its minority sequence variants (8, 21). However, given that the field has been shaped by earlier work that relied on large-scale Sanger sequencing of virus populations, I will briefly cover how Sanger and NGS approaches can give different information. Regardless of the methodology used, it is important to amplify and/or sequence the viral nucleic acid directly, as even a single passage on cells can markedly alter the population's composition (22).

Sanger sequencing studies of viral population diversity rely on either reverse transcription polymerase chain reaction (RT-PCR) or PCR of individual RNA or DNA genomes, respectively. Clonal diversity is assessed by direct sequencing of amplicons from plaque purified virus or sequencing bacterial transformants that contain the amplified fragment cloned into a plasmid vector (4, 23, 24). Because of the labor involved in sequencing multiple clones, most Sanger studies have analyzed diversity in only a small region of the corresponding viral genome, typically one that is shown to be capable of maintaining a high level of sequence polymorphism. Properly controlled, Sanger sequencing can identify both single nucleotide variants (SNV) and recombinant haplotypes. Studies that interrogate the relationship between mutation rate and population diversity have tended to count only unique nucleotide substitutions, as mutations identified two or more times in the same population are likely to be at a higher frequency and subject to positive selection (10). For this same reason, many studies of viral diversity exclude singletons, which tend to be rare and can also arise through RT-PCR error (25). Importantly, Sanger sequencing can only capture a relatively small number of clones from a much larger population, and it is exceedingly hard to get a reliable estimate of a given variant's frequency.

With next generation sequencing, it is much easier to obtain sequence data across the entire genome and to sample a larger, and therefore representative, fraction of the overall population (26). While most studies have used RT-PCR or PCR to amplify viral genomes as fragments, hybridization-capture is increasingly used to enrich a sample for viral nucleic acid (8). By sequencing each base hundreds or thousands of times, one can identify mutations and their frequency within a population. Given that most NGS studies employ short reads (150–300 bases), they are not as reliable for detecting sequence haplotypes. A number of bioinformatic tools use overlap among sequence reads and the frequencies of variants to infer their linkage on a given genome (26). This type of inference can be challenging when the individual mutations are rare, and it is always possible that artifactual haplotypes can arise through RT and/or PCR recombination during genome amplification and library preparation.

Like Sanger sequencing, NGS can identify "false positive" sequence variants that result from RT-PCR and base-calling and are not actually present in the original population (27, 28). The rarer the variant, the more difficult it is to distinguish true from false positives, and most sequencing and bioinformatic pipelines cannot reliably detect variants present at <2% frequency. My laboratory has shown that the sensitivity and specificity of variant detection is also exquisitely sensitive to the number of input genomes (29). With lower nucleic acid input, one's ability to detect true positives goes down and the propagation of RT-PCR errors leads to an inflation of false positives. These effects have been noted by others and need to be accounted for when comparing diversity across samples with varying titer (30–32). When the viral input is adequate, the precision of frequency estimates generally scales with depth of coverage. A good rule of thumb is that the coverage should be ten times the reciprocal of a variant's frequency (i.e., 200x coverage to reliably estimate variants present at ≥5% frequency). Recent work suggests that polymorphisms within the binding sites for amplification primers can increase or decrease the measured frequency for any mutation in the corresponding amplicon by several fold (31).

Richness, Abundance, and Evenness

Ecologists and evolutionary biologists have developed a number of metrics to quantify the varying elements of biological diversity (33, 34). Each is valid in its own right, but subject to the caveat that they typically capture only some of these elements (Figure 2*b*). Classically, diversity is expressed as species richness, abundance, and evenness. Richness refers to the number of distinct species in an area or community. In Virology, richness is expressed as the number of unique single nucleotide variants or haplotypes in a population, irrespective of their frequency. Abundance is the number of individuals of each species in a community. Given that most studies capture only a subsample of a community, abundance is usually described in relative terms rather than as an absolute number per species. The most common graphical representation of relative abundance is a histogram with the number of identified species or variants on the y-axis and the number of times each was identified on the x-axis. In many NGS studies of viral diversity, the x-axis will be a frequency bin so as to indicate how many variants are present at 0–10%, 11–20% frequencies etc. Frequency histograms will also capture evenness, which refers to how similar the abundances of each species are in a given community. For example, a viral population with 10 different variants, each present at 10% would be very even. More commonly, the population is uneven, with an abundant “wild type” and a much larger number of minority variants.

A variety of summary statistics have been used to express viral diversity across populations, and the robustness of the most common ones have recently been evaluated (35). With comparative studies, it is important to ensure that populations have been equally sampled and that issues of sequencing bias and error have been addressed (33, 34). Many Sanger studies have used the number of SNV per thousand bases sequenced. This metric tends to be biased towards species richness, exclusively so if only unique SNV are included. In NGS studies, one can use either SNV per thousand bases or the number of SNV above a frequency cut-off. Ideally, studies that employ a variant threshold should justify this choice and demonstrate that the principle findings don’t change with subtle changes in the frequency cut-off (e.g., SNV identified at above 2%, 1%, and 0.5% frequency).

Shannon Entropy

Shannon entropy (abbreviated as H or D) is a diversity metric that accounts for both the number of variants present (richness) and their frequencies,

$$H = -\sum_{i=1}^{n} p_i(\ln p_i)$$

where pi is the frequency of ith allele and n is the number of allele. The allele can either be a SNV or haplotype. In some cases, Shannon entropy is measured on a per site level (with i = 4, for the 4 possible bases at a given position) and expressed as the average per site entropy across a genome. Zhao and Illingworth have shown that Shannon is highly sensitive to read depth and often underestimates true diversity (35).

Pairwise Nucleotide Diversity

Pairwise nucleotide diversity (**π**) is increasingly used as a summary statistic to compare across populations and express changes over time. It is easily calculated using variant call files generated in most NGS studies as

$$D_l = \frac{\sum_{i \neq j} n_i n_j}{\frac{1}{2} N(N-1)} = \frac{N(N-1) - \sum_i n_i (n_i - 1)}{N(N-1)}$$

where n_i and n_j are the number of copies of alleles i and j, respectively at a given locus, *l*, and N is the number of total number of alleles at a locus. Pairwise diversity across a genome can then be expressed as

$$\pi = \sum_{l=1}^{L} D_l / L$$

Where L is the number of sites, usually the length of the genome. Tajima’s D is similar to pairwise diversity but is adjusted for the number of segregating sites (36). Pairwise diversity appears to be more robust to read depth and provides a more accurate measurement of diversity than Shannon. A major drawback of both Shannon and **π** are that the raw values (e.g., Shannon of 4.71×10^{-5} or **π** of 0.035%) are somewhat abstract and less interpretable than a frequency histogram of relative abundance.

FROM WITHIN-HOST DIVERSITY TO WITHIN-HOST EVOLUTION

Measurements of within-host diversity are rarely the end goal of a study. They are more commonly a means to achieve a more complete understanding of how viral populations evolve (Figure 3*a*). This usually means identifying the sign (positive or negative) and strength of selection, particularly as they relate to viral adaptation to specific sites of replication, antiviral drugs, or host immune pressure. A major challenge in many such studies is to distinguish natural selection from stochastic processes like genetic drift. Comparative studies of diversity can also shed light on varying dynamics across viral species, the changes in a population over time, and spatial segregation across different sites within a host. As the observed differences in diversity can often be quite subtle, the strength of inference depends on the size of the study, the availability of control groups, and the use of an appropriate null model. This has been an area of innovation, and the field has shifted from “sequencing whatever I can get my hands on” to sequencing samples collected prospectively with associated host metadata.

The first order of business is usually to identify genes or sites under strong positive selection, and various methods have been used to detect selection in cross-sectional datasets (i.e., one sample per individual host) and in virus populations that are serially sampled over time. The dN/dS (or Ka/Ks) ratio compares the number of nonsynonymous substitutions per nonsynonymous site to the number of synonymous substitutions per synonymous site (37). The number of nonsynonymous and synonymous sites is based on the codons in a given open reading frame; given the genetic code, there is typically a twofold excess of nonsynonymous sites. As synonymous mutations are more likely to be selectively neutral and amino acid substitutions are more likely to be “seen” by natural selection, an excess of nonsynonymous mutations (dN/dS >1) suggests positive selection. If nonsynonymous mutations are under-represented (dN/dS <1), then negative, or purifying selection, is said to dominate. While these tests were initially developed for fixed (i.e., consensus) mutations over longer evolutionary timescales, they have also been applied to alignments of clonal viral sequences from within-host populations. A variant of this test, $\pi n/\pi s$, compares the pairwise diversity of nonsynonymous and synonymous mutations corrected for the number of sites and is often applied to NGS data (38). The dN/dS ratio is obviously biased against non-neutral synonymous mutations (e.g., those affecting translational efficiency or RNA structural elements) and does not capture noncoding mutations, which often have large phenotypic effects in viral systems (39, 40). A bigger issue is that the test was designed for populations that have significantly diverged over long time scales (41). Within hosts, dN/dS can be artificially inflated, as nonsynonymous mutations can accumulate through neutral processes, and selection can be fairly weak at purging deleterious mutations. Conversely, a threshold dN/dS of greater than one is poorly sensitive for detecting genes or sites under positive selection, particularly when purifying selection dominates – as it does within hosts (42).

Given these challenges, a number of other criteria have been applied to identify genes or mutations that are likely under positive selection, even in cross-sectional datasets (Figure 3*b*). First, mutations under positive selection are more likely to be present and at higher frequencies than other nonsynonymous mutations (43). For example, if host antibody positively selects viral escape variants in vivo, one would find enrichment of high frequency SNV in antigenic regions of a viral surface protein relative to other sites in the genome. One would be even more confident if the number and frequency of these variants were positively correlated with host antibody titer, a measure of selective pressure. Comparing mutations across sites and in the presence and absence of the proposed selective pressure are important, since mutations can also increase in frequency by chance (44). Second, newly arising mutations with a previously-characterized adaptive phenotype can reasonably be assumed to be positively selected. Classic examples would be a known drug resistance mutation identified in a treated individual or a mutation affecting human receptor usage that arises in a human host after a zoonotic transmission event (6, 45–47). Third, specific viral mutations that are repeatedly identified across individuals are suggestive of convergent evolution and selection for a given phenotype (43, 48). Of course, this only holds if the mutations truly arise independently in each population and are not present in multiple hosts within a transmission chain. Given that these parallel mutations can also arise due to stereotypical sequencing error or sample cross-contamination, it is important that additional controls are provided to exclude this possibility. Finally, parallelism between within-host mutations and those seen at larger (e.g., global scales) provides additional support for positive selection (48, 49). Various statistical models can be used to determine whether the appearance of viral mutations across hosts or across evolutionary scales is more than expected by chance.

Repeated sampling of individual hosts over the course of an infection is a powerful approach to explore within-host evolutionary dynamics. Identification of the same variant in multiple samples suggests that it is not a spurious result, and charting a steady upward or downward trajectory over time provides stronger evidence of positive or negative selection, respectively (43, 50, 51). Thus, one can use the criteria above with even more confidence. For example, time- series data on variant frequencies have been informative about the impact of antiviral- or immune-mediated "selective sweeps" on HIV diversity in vivo (52). Longitudinal data can also be used to fit population genetic models of within-host processes. These models provide estimates of effective population size and mutation rate, critical parameters for describing the diversification of viral populations in vivo (53–55). Wright-Fisher and Moran models assume neutrality, and are ideally parameterized using data on synonymous mutations or other sites known to evolve neutrally. The fact that some sites in viruses are under strong selection in vivo does not diminish the utility of neutral models. Because they capture how much diversification can happen due to random processes, they provide an important null model for testing hypotheses regarding the presence and strength of selection on a given variant. The difference between the effective population size and the true population size also highlights the relative importance of genetic drift and other stochastic forces on evolution in vivo (16, 56).

CASE STUDIES IN WITHIN-HOST DIVERSITY

Influenza, an Acute Infection with an RNA Virus

In humans, influenza A (IAV) and influenza B viruses (IBV) typically cause acute, self-limited infection of the respiratory tract. In some individuals, the virus can cause a clinically significant pneumonia, even in the absence of superinfection. While cases of mixed subtype (i.e., H3N2 and H1N1) and mixed strain (i.e., distinct genotypes) have been documented, work in animal models and studies of natural human infection suggest that the infecting population has little genetic diversity, and may be nearly clonal (50, 57). The virus undergoes explosive replication with peak titers 1–2 days after infection and a steady decline over the next 5–7 days. Given the virus' high mutation rate, new variants are rapidly generated with each cellular infection cycle (27).

Experimental infections of animals with molecularly barcoded viruses suggests that reassortment is remarkably common in vivo (58). However, the observed rate of reassortment is low in humans, likely because most reassortment events occur between viruses that are nearly identical (59).

My laboratory has used NGS to define temporal aspects of influenza virus diversity in both cross-sectional and longitudinal datasets with samples from a community-based cohort (44, 50, 60). Using a conservative 2% frequency cut-off, we have found that most IAV populations have fewer than 5 SNV, regardless of the day of sampling (44, 50). In more recent work, we have found that IBV exhibits considerably less diversity within hosts with fewer SNV and lower pairwise nucleotide diversity (60). Consistent with strong purifying selection, the vast majority of SNV are present at <10% frequency and the dN/dS ratio is approximately 0.2 for both IAV and IBV. Purifying selection has also been clearly demonstrated in H3 and H1 hemagglutinin sequences from a second community cohort

(61) and in a study of humans infected with avian H5N1 viruses (49). Sequences from 43 individuals who provided paired samples over a 6 day period indicate that within-host populations are highly dynamic with the rapid accumulation and elimination of SNV (50). This pattern of SNV turnover was similar for nonsynonymous and synonymous mutations, which suggests the influence of genetic drift over natural selection.

While positive selection appears to be a major driver of influenza globally (62–64), it does not appear to be strong within hosts. Positive selection has been demonstrated most clearly in the case of oseltamivir-resistance, particularly in immunocompromised hosts (6). These individuals often experience viral replication beyond the usual 7 day course, which allows more time for selection to drive newly arising variants to a level where they can be detected. Consistent with this model, Xue and colleagues identified strong evidence for positive selection in 4 immunocompromised hosts who were each sampled for over 50 days (48). Several mutations arose independently in multiple patients and a subset subsequently circulated globally at high frequency. In contrast, antibody-mediated selection of novel antigenic variants appears to be rare in most acutely infected individuals. Dinis et al. identified a number of antigenic variants in an observational study, but most were rare and there was no clear enrichment of SNV in antigenic epitopes (61). My laboratory also identified little evidence for selection in samples collected from a randomized trial of influenza vaccines (44). We found that SNV were no more common in antigenic sites and that diversity was not correlated with vaccination status or preseason antibody titer. Similarly, recent work suggests that while humans infected with H5N1 can harbor low frequency, mammalian-adaptive mutations, their spread is limited by purifying selection, genetic drift, and the relatively short time-scale of most infections (47, 49).

Dengue, an Acute Infection Cycling Through Host and Vector

Dengue fever is an acute systemic infection that is due to one of four serotypes of Dengue virus, which are vectored by *Aedes* mosquito species. Following inoculation, the virus replicates locally and in draining lymph nodes, with an asymptomatic incubation period of 4–7 days (49a). Symptoms usually coincide with viremia and replication in peripheral blood mononuclear cells (PBMC). The viremia typically lasts for 4–5 days, and viral clearance from peripheral tissues follows several days later. Individuals are most infectious while viremic, as mosquitoes acquire the virus again through a blood meal. The mosquito phase lasts about 10 days; the virus transits through the midgut then disseminates via the hemolymph and the fat body to the salivary glands. At this point, the mosquito is able to transmit the virus again to a naïve host. Severe Dengue disease – previously classified as Dengue hemorrhagic fever or Dengue shock syndrome – typically occurs with secondary Dengue virus infection with a different serotype (49b).

The within-mosquito dynamics of Dengue virus have been characterized through a number of elegant experiments in which mosquitoes were fed a blood meal and then dissected for viral population sequencing (65). Virus populations from the midgut on days 4 and 7 and the salivary gland on day 14, all exhibited increased diversity (Shannon entropy and π) relative to the blood meal. In all cases, the dN/dS values were between 0.2–0.3, suggestive of strong purifying selection. An analysis of within-host bottlenecks was performed using three synonymous SNV present in the blood meal and midgut samples. While the ingested blood meal was approximately 3000 infectious units, the founding population of the midgut was estimated to be just 5–42 genomes. Interestingly, this result closely matches an estimate obtained for Venezuelan equine encephalitis virus (66). Salivary gland samples showed occasional consensus level changes relative to the blood meal, but there was no convergence across mosquitoes. A separate study from the same group also found a lack of evolutionary convergence (i.e., positive selection) in viral sequences from saliva obtained from mosquito pools (67). Together, these data indicate that Dengue virus populations undergo one, and possibly several, bottleneck events in mosquitoes followed by rapid diversification under purifying selection. This combination of extreme genetic drift and purifying selection has also been observed in analogous experiments with Culex mosquitoes and West Nile virus (68, 69).

A large number of studies over the last 20 years have probed the diversity of Dengue viruses within human hosts. Early work employed Sanger sequencing of cloned amplicons derived from limited numbers of patients (70, 71). Viral diversity in the envelope was generally low, but variable across hosts, with a per-sample nucleotide diversity (number of SNV per nucleotides sequenced) of 0.10 – 0.84 and pairwise clonal diversity of 0.21–1.67 (70, 72). A follow-up study found that diversity within the capsid and envelope genes was lower in mosquitoes than in humans (73). Two larger studies identified similar diversity in 662 and 8458 envelope clones from 16 and 17 individuals, respectively (25, 74). Both found a within-host dN/dS of approximately 0.6; diversity was lower and the ratio was 0.23 when unique mutations, which are more likely to be RT-PCR error, were excluded. Two more recent NGS studies have explored genome-wide diversity using paired and cross-sectional sampling. Sessions et al. found low diversity in 12 paired samples, with few mutations above the 1% or 5% frequency thresholds (75). They estimated that SNV accumulate at a rate of 0.002 changes per position per day.

Parameswaran et al. used NGS to identify variants at >1% frequency in 31 plasma and 68 PBMC samples from 77 individuals (43). They found fewer SNV in secondary cases of Dengue and in the envelope, consistent with immune-mediated constraints. Variants were identified at marginally higher frequency across hosts in prM/M and NS3, which the authors suggested were due to immune selection. However, these variants were found to be less fit in a variety of culture based assays. Together, these data suggest that Dengue is subject to strong purifying selection in both mosquito and human hosts, with significant genetic drift in mosquitoes and limited evidence for positive selection in humans.

CMV, a Large dsDNA Virus with Cycles of Latency and Reactivation

While the vast majority of studies of within-host viral diversity have focused on RNA viruses and retroviruses, recent work has identified significant within-host diversity in CMV, a large double-stranded DNA virus (11). Cytomegalovirus replicates in mucosal epithelia and myeloid cells during acute infection and is shed in saliva, breast milk, and urine. The virus becomes latent in bone marrow myeloid precursors, with varying levels of reactivation and shedding over an individual’s lifetime. CMV, like many DNA viruses, has a low mutation rate, but it does exhibit a high level of recombination. The chronicity of the infection and the large population sizes within hosts, with cycles of reactivation, can lead to significant diversification. In immunocompetent hosts, many CMV infections are asymptomatic and the primary symptoms – fever, malaise, sore throat – are fairly nonspecific. As a result, most sampling and sequencing have been performed in congenitally infected infants or immunocompromised hosts, who experience severe disease and prolonged shedding. Here, patients have been found to harbor multiple genotypes by targeted sequencing and/or PCR-typing, e.g., (48, 76).

Given its large genome, studies of CMV genomic diversity have only really been feasible with NGS. Over the last decade, a number of studies have used either hybridization capture or long-range PCR to recover sufficient material for sequencing (8). The first systematic investigation of within-host diversity applied NGS to CMV amplified from the urine of 3 congenitally-infected infants who were sampled within 2 weeks of birth (77). Over 95% of the genome was covered and SNV were identified relative to a single, sample-specific consensus. Each sample had greater than 8000 SNV, and >90% were present at a frequency of less <10%. Pairwise nucleotide diversity was 0.18–0.22%, which is comparable to values for HIV and many RNA viruses. Targeted Sanger sequencing of four highly variable open reading frames provided evidence for both mixed infection and for clusters of within-host genotypes separated by one or two mutations. A much larger study of 84 samples from 43 immunocompromised children and adults also identified high pairwise diversity (78). Here, the distribution was bimodal, with a significant number of samples with very little diversity. Phylogenetic analysis and haplotype reconstruction of high-diversity samples from two individuals suggested that the signal was driven by mixed infection and subsequent recombination between these genetically distinct strains.

While within-host evolution of CMV is heavily influenced by stochastic events, there is also evidence for strong selection at specific loci. Renzette et al. explored these dynamics using serial plasma samples form one infant, serial urine samples from three infants, and serial urine and plasma samples from one infant (51). There was generally less viral diversity in urine relative to plasma. Within-host populations were stable in paired, longitudinal specimens; 88% of SNV had a frequency change of <10%. The divergence between plasma and urine samples in one infant was similar to the divergence observed between hosts. A population genetic model suggested that this combination of within-compartment stability and between-compartment divergence was largely driven by population bottlenecks, expansions, and splits. Subsequent studies from the same group have highlighted how within-host diversity in congenitally infected infants is shaped by multiple reinfection events and subsequent admixture among these often distinct populations (79). Recombination among homologous and genetically-distinct strains is prominent in CMV-infected individuals and there is evidence for positive selection of recombinants within hosts (78, 80). This parallels the prominent role of recombination in CMV genomic evolution over longer time scales.

WITHIN-HOST DIVERSITY AND VIRAL QUASISPECIES

In the years after Domingo's pioneering work on phage diversity, Domingo, Holland and others explored the impact of low replicative fidelity and high diversity on viral evolutionary dynamics (81, 82). This body of work led to the development of what is often described as “viral quasispecies theory.” Here, RNA viruses are best represented as a diverse collection of mutants – or alternatively as a “quasispecies,” “mutant spectrum,” “swarm,” or “cloud”(82a-g). More provocatively, it has been proposed that the collection of mutants –not the individual viruses themselves – is a determinant of viral phenotype and the target of natural selection. Foundational experimental work in the poliovirus system further suggested that viral diversity is a determinant of virulence and that rare variants within the quasispecies are major drivers of pathogenesis in vivo (Figure 4) (4). The validity and explanatory power of viral quasispecies theory have been much debated in the literature (83–85). Here, I will focus on the evidence linking within-host diversity to virulence.

Work in experimental systems has clearly established that viral diversity is a correlate of virulence and attenuation. Both Pfeiffer and Kirkegaard and Vignuzzi et al. discovered that a poliovirus harboring a single substitution in the polymerase, $3D^{G64S}$, had a lower mutation rate and generated populations with reduced genetic diversity (4, 86, 87). Vignuzzi et al. went on to show that the less diverse $3D^{G64S}$ populations were attenuated in a mouse model of infection (4). Since then, high fidelity variants have been identified in a number of other viral systems, and when evaluated, a correlation between diversity and virulence is consistently observed (88–94). However, recent work from my laboratory suggests that this relationship is confounded by the fact that viral polymerases are subject to a speed-fidelity trade-off, which means that high fidelity variants also have a significant replication defect (95–97). We found that $3D^{G64S}$ polioviruses replicated much more slowly than the corresponding wild type (WT) and used a variety of mutants to uncouple replicative speed from mutation rate and population diversity. In contrast to earlier models, our work showed that replicative speed was a determinant of virulence –not population diversity (Figure 4). This same speed-fidelity trade-off has been identified in other RNA viruses and may confound other studies linking viral mutation rates and population diversity to pathogenesis (96, 98).

Another issue in relating diversity to virulence is reconciling the relatively small differences in diversity across virus populations with larger scale differences in infectious outcome. Much of the early work on viral quasispecies relied on Sanger sequencing of clones to quantify diversity in vitro and in vivo (4). As detailed above, this approach can be biased toward rare variants and the levels of diversity are far removed from their biological impact. For example, Sanger sequencing of the capsid region of WT and $3D^{G64S}$ populations identified mutation frequencies of 0.026% and 0.004%, respectively (4). My laboratory used NGS to show that the differences between WT and $3D^{G64S}$ were trivial and largely confined to the fraction of variants present at <0.1% frequency (97). A more recent study of PB1-V43I, a putative high-fidelity variant of influenza virus, measured a mutation frequency of 0.048% for this mutant relative to 0.071% for the WT in vitro; in mice infected with these viruses, the frequencies were 0.048% for PB1-V43I and 0.082% for the WT (93). While NGS and multidimensional scaling are able to resolve small differences in the type and level of diversity across populations, a robust mechanistic framework to explain how variants that are often present at <1% frequency can substantially influence pathogenesis is lacking (99).

Finally, the concept of a large and diverse within-host mutant swarm does not fit well with experimental work in animal models and empiric data from natural infections. Pfeiffer and colleagues have used barcoded viruses to demonstrate that poliovirus and Coxsackievirus populations are subject to a series of strong bottlenecks that dramatically reduce viral diversity in vivo (100–102). Similar results have been obtained in experimental infections of mosquitoes with barcoded Venezuelan Equine Encephalitis viruses and of mosquitoes and macaques with barcoded Zika viruses (66, 103). As detailed above, virus populations in natural infections are subject to frequent bottlenecks, population splits, and transient expansions and contractions (50, 51). Even though the total number of viruses in a host may be quite large, these stochastic processes make the population behave as if it is a much smaller one. Thus, the available data do not seem to be compatible with the existence of an intact mutant swarm, much less a model in which the composition of the mutant spectrum is itself optimized by natural selection.

CONCLUSIONS

Advances in sequencing technology have enabled a large number of studies of viral diversity within naturally-infected hosts. Given the sheer volume of sequence data and multiple sources for error, attention to quality control and validation is critical. This is particularly true for studies of natural infections, where inference is entirely based on secondary analyses of sequence data and follow-up experiments are often not possible. While I have focused mainly on issues related to single nucleotide variants, the same principles apply to studies of recombinant haplotypes and defective viral genomes within hosts. Similarly, it is important to control for sampling bias in comparative studies and to recognize the advantages and disadvantages of various diversity metrics. The explanatory power of comparative studies can be improved through serial sampling or by studying virus populations across carefully selected hosts with appropriate controls. The available data from multiple viral systems suggests that within-host populations are subject to strong purifying selection and that most mutations are rare. Changes in variant frequency are often due to stochastic processes like genetic drift, and there is limited evidence for strong positive selection. These empiric data stand in contrast to what has been suggested from experimental infections of laboratory animals and viral quasispecies models. Improvements in sequencing methods, including molecular barcoding (104, 105), may allow for reliable detection of positively selected variants, which may be present at lower frequencies. Applied in the appropriate epidemiological context, these approaches may elucidate phenotypes under selection within hosts and how within-host processes relate to larger-scale viral dynamics.

ACKNOWLEDGMENT

Work on viral diversity in the author’s laboratory is funded by NIH R01 AI118886.

LITERATURE CITED

1. Domingo E, Sabo D, Taniguchi T, Weissmann C. 1978. Nucleotide sequence heterogeneity of an RNA phage population. Cell 13:735–44 [PubMed: 657273]

2. Parvin JD, Moscona A, Pan WT, Leider JM, PALESEl P. Measurement of the Mutation Rates of Animal Viruses: Influenza A Virus and Poliovirus Type. J. Virol, p. 7

3. Crotty S, Maag D, Arnold JJ, Zhong W, Lau JYN, et al. 2000. The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. Nat. Med 6(12):1375–79 [PubMed: 11100123]
4. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature 439(7074):344–48 [PubMed: 16327776]
5. Aaskov J. 2006. Long-Term Transmission of Defective RNA Viruses in Humans and Aedes Mosquitoes. Science 311(5758):236–38 [PubMed: 16410525]
6. Ghedin E, Laplante J, DePasse J, Wentworth DE, Santos RP, et al. 2011. Deep Sequencing Reveals Mixed Infection with 2009 Pandemic Influenza A (H1N1) Virus Strains and the Emergence of Oseltamivir Resistance. J. Infect. Dis 203(2):168–74 [PubMed: 21288815]
7. Kao RR, Haydon DT, Lycett SJ, Murcia PR. 2014. Supersize me: how whole-genome sequencing and big data are transforming epidemiology. Trends Microbiol 22(5):282–91 [PubMed: 24661923]
8. Houldcroft CJ, Beale MA, Breuer J. 2017. Clinical and biological insights from viral genome sequencing. Nat. Rev. Microbiol 15(3):183–92 [PubMed: 28090077]
9. Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R. 2010. Viral Mutation Rates. J. Virol 84(19):9733–48 [PubMed: 20660197]
10. Peck KM, Lauring AS. 2018. Complexities of Viral Mutation Rates. J. Virol 92(14):e01031–17, /jvi/92/14/e01031–17.atom [PubMed: 29720522]
11. Renner DW, Szpara ML. 2017. Impacts of Genome-Wide Analyses on Our Understanding of Human Herpesvirus Diversity and Evolution. J. Virol 92(1):e00908–17 [PubMed: 29046445]
12. Samuel CE. 2011. Adenosine deaminases acting on RNA (ADARs) are both antiviral and proviral. Virology 411(2):180–93 [PubMed: 21211811]
13. Harris RS, Dudley JP. 2015. APOBECs and virus restriction. Virology 479–480:131–45
14. Pérez-Losada M, Arenas M, Galán JC, Palero F, González-Candelas F. 2015. Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. Infect. Genet. Evol 30:296–307 [PubMed: 25541518]
15. McDonald SM, Nelson MI, Turner PE, Patton JT. 2016. Reassortment in segmented RNA viruses: mechanisms and outcomes. Nat. Rev. Microbiol 14(7):448–60 [PubMed: 27211789]
16. Rouzine IM, Rodrigo A, Coffin JM. 2001. Transition between Stochastic Evolution and Deterministic Evolution in the Presence of Selection: General Theory and Application to Virology. Microbiol. Mol. Biol. Rev 65(1):151–85 [PubMed: 11238990]
17. Moya A, Holmes EC, González-Candelas F. 2004. The population genetics and evolutionary epidemiology of RNA viruses. Nat. Rev. Microbiol 2(4):279–88 [PubMed: 15031727]
18. Frost SDW, Magalis BR, Kosakovsky Pond SL. 2018. Neutral Theory and Rapidly Evolving Viral Pathogens. Mol. Biol. Evol 35(6):1348–54 [PubMed: 29688481]
19. McCrone JT, Lauring AS. 2018. Genetic bottlenecks in intraspecies virus transmission. Curr. Opin. Virol 28:20–25 [PubMed: 29107838]
20. Kennedy DA, Dwyer G. 2018. Effects of multiple sources of genetic drift on pathogen variation within hosts. PLOS Biol 16(3):e2004444 [PubMed: 29590105]
21. Metzker ML. 2010. Sequencing technologies — the next generation. Nat. Rev. Genet 11(1):31–46 [PubMed: 19997069]
22. McWhite CD, Meyer AG, Wilke CO. 2016. Sequence amplification via cell passaging creates spurious signals of positive adaptation in influenza virus H3N2 hemagglutinin. Virus Evol 2(2):vew026 [PubMed: 27713835]
23. Crotty S, Cameron CE, Andino R. 2001. RNA virus error catastrophe: Direct molecular test by using ribavirin. PNAS 98(12):6895–900 [PubMed: 11371613]
24. Levi LI, Gnädig NF, Beaucourt S, McPherson MJ, Baron B, et al. 2010. Fidelity Variants of RNA Dependent RNA Polymerases Uncover an Indirect, Mutagenic Activity of Amiloride Compounds. PLOS Pathog 6(10):e1001163 [PubMed: 21060812]
25. Thai KTD, Henn MR, Zody MC, Tricou V, Nguyet NM, et al. 2012. High-Resolution Analysis of Intrahost Genetic Diversity in Dengue Virus Serotype 1 Infection Identifies Mixed Infections. J. Virol 86(2):835–43 [PubMed: 22090119]

26. Beerenwinkel N, Zagordi O. 2011. Ultra-deep sequencing for the analysis of viral populations. Curr. Opin. Virol 1(5):413–18 [PubMed: 22440844]

27. Pauly MD, Procario MC, Lauring AS. 2017. A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses. eLife 6:e26437 [PubMed: 28598328]

28. Robasky K, Lewis NE, Church GM. 2014. The role of replicates for error mitigation in next-generation sequencing. Nat. Rev. Genet 15(1):56–62 [PubMed: 24322726]

29. McCrone JT, Lauring AS. 2016. Measurements of Intrahost Viral Diversity Are Extremely Sensitive to Systematic Errors in Variant Calling. J. Virol 90(15):6884–95 [PubMed: 27194763]

30. Gallet R, Fabre F, Michalakis Y, Blanc S. 2017. The Number of Target Molecules of the Amplification Step Limits Accuracy and Sensitivity in Ultradeep-Sequencing Viral Population Studies. J. Virol 91(16):e00561–17, /jvi/91/16/e00561–17.atom [PubMed: 28566384]

31. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, et al. 2019. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol 20(1):8 [PubMed: 30621750]

32. Illingworth CJR, Roy S, Beale MA, Tutill H, Williams R, Breuer J. 2017. On the effective depth of viral sequence data. Virus Evol 3(2):

33. Magurran A, McGill B. 2011. Biological diversity: Frontiers in measurement and Assessment Oxfod University Press. First ed.

34. Gregori J, Salicrú M, Domingo E, Sanchez A, Esteban JI, et al. 2014. Inference with viral quasispecies diversity indices: clonal and NGS approaches. Bioinformatics 30(8):1104–11 [PubMed: 24389655]

35. Zhao L, Illingworth CJR. 2019. Measurements of intrahost viral diversity require an unbiased diversity metric. Virus Evol 5(1):

36. Tajima F. Statistical Methodfor Testing the Neutral Mutation Hypothesis by DNA Polymorphism p. 11

37. Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol 3(5):418–26 [PubMed: 3444411]

38. Nelson CW, Hughes AL. 2015. Within-host nucleotide diversity of virus populations: Insights from next-generation sequencing. Infect. Genet. Evol 30:1–7 [PubMed: 25481279]

39. Cuevas JM, Domingo-Calap P, Sanjuán R. 2012. The Fitness Effects of Synonymous Mutations in DNA and RNA Viruses. Mol. Biol. Evol 29(1):17–20 [PubMed: 21771719]

40. Visher E, Whitefield SE, McCrone JT, Fitzsimmons W, Lauring AS. 2016. The Mutational Robustness of Influenza A Virus. PLOS Pathog 12(8):e1005856 [PubMed: 27571422]

41. Kryazhimskiy S, Plotkin JB. 2008. The Population Genetics of dN/dS. PLOS Genet 4(12):e1000304 [PubMed: 19081788]

42. Lin J-J, Bhattacharjee MJ, Yu C-P, Tseng YY, Li W-H. 2019. Many human RNA viruses show extraordinarily stringent selective constraints on protein evolution. PNAS 116(38):19009–18 [PubMed: 31484772]

43. Parameswaran P, Wang C, Trivedi SB, Eswarappa M, Montoya M, et al. 2017. Intrahost Selection Pressures Drive Rapid Dengue Virus Microevolution in Acute Human Infections. Cell Host Microbe 22(3):400–410.e5 [PubMed: 28910637]

44. Debbink K, McCrone JT, Petrie JG, Truscon R, Johnson E, et al. 2017. Vaccination has minimal impact on the intrahost diversity of H3N2 influenza viruses. PLOS Pathog, p. 18

45. Illingworth CJR. 2015. Fitness Inference from Short-Read Data: Within-Host Evolution of a Reassortant H5N1 Influenza Virus. Mol. Biol. Evol 32(11):3012–26 [PubMed: 26243288]

46. Sobel Leonard A, McClain MT, Smith GJD, Wentworth DE, Halpin RA, et al. 2016. Deep Sequencing of Influenza A Virus from a Human Challenge Study Reveals a Selective Bottleneck and Only Limited Intrahost Genetic Diversification. J. Virol 90(24):11247–58 [PubMed: 27707932]

47. Imai H, Dinis JM, Zhong G, Moncla LH, Lopes TJS, et al. 2018. Diversity of Influenza A(H5N1) Viruses in Infected Humans, Northern Vietnam, 2004–2010. Emerg. Infect. Dis 24(7):1128–238 [PubMed: 29912683]

48. Xue KS, Stevens-Ayers T, Campbell AP, Englund JA, Pergam SA, et al. 2017. Parallel evolution of influenza across multiple spatiotemporal scales. eLife 6:e26875 [PubMed: 28653624]

49. Moncla LH, Bedford T, Dussart P, Horm SV, Rith S, et al. 2019. Quantifying within-host evolution of H5N1 influenza in humans and poultry in Cambodia. Microbiology

50. McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauring AS. 2018. Stochastic processes constrain the within and between host evolution of influenza virus. eLife 7:e35962 [PubMed: 29683424]

51. Renzette N, Gibson L, Bhattacharjee B, Fisher D, Schleiss MR, et al. 2013. Rapid Intrahost Evolution of Human Cytomegalovirus Is Shaped by Demography and Positive Selection. PLOS Genet 9(9):e1003735 [PubMed: 24086142]

52. Feder AF, Rhee S-Y, Holmes SP, Shafer RW, Petrov DA, Pennings PS. 2016. More effective drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. eLife 5:e10670 [PubMed: 26882502]

53. Foll M, Shim H, Jensen JD. 2015. WFABC: a Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. Mol. Ecol. Resour 15(1):87–98 [PubMed: 24834845]

54. Kimura M Solution of a process of random genetic drift with a continuous model. PNAS 41:144–50

55. Williamson EG, Slatkin M. Using Maximum Likelihood to Estimate Population Size From Temporal Changes in Allele Frequencies, p. 8

56. Kouyos RD, Althaus CL, Bonhoeffer S. 2006. Stochastic or deterministic: what is the effective population size of HIV-1? Trends Microbiol 14(12):507–11 [PubMed: 17049239]

57. Varble A, Albrecht RA, Backes S, Crumiller M, Bouvier NM, et al. 2014. Influenza A Virus Transmission Bottlenecks Are Defined by Infection Route and Recipient Host. Cell Host Microbe 16(5):691–700 [PubMed: 25456074]

58. Tao H, Steel J, Lowen AC. 2014. Intrahost Dynamics of Influenza Virus Reassortment. J. Virol 88(13):7485–92 [PubMed: 24741099]

59. Sobel Leonard A, McClain MT, Smith GJD, Wentworth DE, Halpin RA, et al. 2017. The effective rate of influenza reassortment is limited during human infection. PLOS Pathog 13(2):e1006203 [PubMed: 28170438]

60. Valesano AL, Fitzsimmons WJ, McCrone JT, Petrie JG, Monto AS, et al. 2019. Influenza B viruses exhibit lower within-host diversity than influenza A viruses in human hosts. J. Virol JVI.01710–19, jvi;JVI.01710–19v1

61. Dinis JM, Florek NW, Fatola OO, Moncla LH, Mutschler JP, et al. 2016. Deep Sequencing Reveals Potential Antigenic Variants at Low Frequencies in Influenza A Virus-Infected Humans. J. Virol 90(7):3355–65 [PubMed: 26739054]

62. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. 2008. The genomic and epidemiological dynamics of human influenza A virus. Nature 453(7195):615–19 [PubMed: 18418375]

63. Bedford T, Suchard MA, Lemey P, Dudas G, Gregory V, et al. 2014. Integrating influenza antigenic dynamics with molecular evolution. eLife 3:e01914 [PubMed: 24497547]

64. Bedford T, Riley S, Barr IG, Broor S, Chadha M, et al. 2015. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. Nature 523(7559):217–20 [PubMed: 26053121]

65. Lequime S, Fontaine A, Ar Gouilh M, Moltini-Conclois I, Lambrechts L. 2016. Genetic Drift, Purifying Selection and Vector Genotype Shape Dengue Virus Intra-host Genetic Diversity in Mosquitoes. PLOS Genet 12(6):e1006111 [PubMed: 27304978]

66. Forrester NL, Guerbois M, Seymour RL, Spratt H, Weaver SC. 2012. Vector-Borne Transmission Imposes a Severe Bottleneck on an RNA Virus Population. PLOS Pathog 8(9):e1002897 [PubMed: 23028310]

67. Lequime S, Richard V, Cao-Lormeau V-M, Lambrechts L. 2017. Full-genome dengue virus sequencing in mosquito saliva shows lack of convergent positive selection during transmission by Aedes aegypti. Virus Evol 3(2):

68. Grubaugh ND, Weger-Lucarelli J, Murrieta RA, Fauver JR, Garcia-Luna SM, et al. 2016. Genetic Drift during Systemic Arbovirus Infection of Mosquito Vectors Leads to Decreased Relative Fitness during Host Switching. Cell Host Microbe 19(4):481–92 [PubMed: 27049584]
69. Grubaugh ND, Fauver JR, Rückert C, Weger-Lucarelli J, Garcia-Luna S, et al. 2017. Mosquitoes Transmit Unique West Nile Virus Populations during Each Feeding Episode. Cell Rep 19(4):709–18 [PubMed: 28445723]
70. Wang W-K, Lin S-R, Lee C-M, King C-C, Chang S-C. 2002. Dengue Type 3 Virus in Plasma Is a Population of Closely Related Genomes: Quasispecies. J. Virol 76(9):4662–65 [PubMed: 11932434]
71. Holmes EC. 2003. Patterns of Intra- and Interhost Nonsynonymous Variation Reveal Strong Purifying Selection in Dengue Virus. J. Virol 77(20):11296–98 [PubMed: 14512579]
72. Tu Z, He Y, Lu H, Xu L, Yang Z, et al. 2013. Mutant spectrum of dengue type 1 virus in the plasma of patients from the 2006 epidemic in South China. Int. J. Infect. Dis 17(11):e1080–81 [PubMed: 23827321]
73. Lin S-R, Hsieh S-C, Yueh Y-Y, Lin T-H, Chao D-Y, et al. 2004. Study of Sequence Variation of Dengue Type 3 Virus in Naturally Infected Mosquitoes and Human Hosts: Implications for Transmission and Evolution. J. Virol 78(22):12717–21 [PubMed: 15507664]
74. Descloux E, Cao-Lormeau V-M, Roche C, De Lamballerie X. 2009. Dengue 1 Diversity and Microevolution, French Polynesia 2001–2006: Connection with Epidemiology and Clinics. PLOS Negl. Trop. Dis 3(8):e493 [PubMed: 19652703]
75. Sessions OM, Wilm A, Kamaraj US, Choy MM, Chow A, et al. 2015. Analysis of Dengue Virus Genetic Diversity during Human and Mosquito Infection Reveals Genetic Constraints. PLOS Negl. Trop. Dis 9(9):e0004044 [PubMed: 26327586]
76. Gorzer I, Guelly C, Trajanoski S, Puchhammer-Stockl E. 2010. Deep Sequencing Reveals Highly Complex Dynamics of Human Cytomegalovirus Genotypes in Transplant Patients over Time. J. Virol 84(14):7195–203 [PubMed: 20463084]
77. Renzette N, Bhattacharjee B, Jensen JD, Gibson L, Kowalik TF. 2011. Extensive Genome-Wide Variability of Human Cytomegalovirus in Congenitally Infected Infants. PLOS Pathog 7(5):e1001344 [PubMed: 21625576]
78. Cudini J, Roy S, Houldcroft CJ, Bryant JM, Depledge DP, et al. 2019. Human cytomegalovirus haplotype reconstruction reveals high diversity due to superinfection and evidence of within-host recombination. PNAS 116(12):5693–98 [PubMed: 30819890]
79. Pokalyuk C, Renzette N, Irwin KK, Pfeifer SP, Gibson L, et al. 2017. Characterizing human cytomegalovirus reinfection in congenitally infected infants: an evolutionary perspective. Mol. Ecol 26(7):1980–90 [PubMed: 27988973]
80. Lassalle F, Depledge DP, Reeves MB, Brown AC, Christiansen MT, et al. 2016. Islands of linkage in an ocean of pervasive recombination reveals two-speed evolution of human cytomegalovirus genomes. Virus Evol 2(1):vew017 [PubMed: 30288299]
81. Domingo E, Martín V, Perales C, Grande-Pérez A, García-Arriaza J, Arias A. 2006. Viruses as Quasispecies: Biological Implications. In Quasispecies: Concept and Implications for Virology, ed. Domingo E, 299:51–82. Berlin/Heidelberg: Springer-Verlag
82. Holland J, Spindler K, Horodyski F, Grabau E, Nichol S, VandePol S. 1982. Rapid Evolution of RNA Genomes. Sci. New Ser 215(4540):1577–85
83. Moya A, Elena SF, Bracho A, Miralles R, Barrio E. 2000. The evolution of RNA viruses: A population genetics view. PNAS 97(13):6967–73 [PubMed: 10860958]
84. Jenkins GM, Worobey M, Woelk CH, Holmes EC. 2001. Evidence for the Non-quasispecies Evolution of RNA Viruses. Mol. Biol. Evol 18(6):987–94 [PubMed: 11371587]
85. Wilke CO. 2005. Quasispecies theory in the context of population genetics. BMC Evol. Biol 5(1):44 [PubMed: 16107214]
86. Pfeiffer JK, Kirkegaard K. 2003. A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity. PNAS 100(12):7289–94 [PubMed: 12754380]
87. Pfeiffer JK, Kirkegaard K. 2005. Increased Fidelity Reduces Poliovirus Fitness and Virulence under Selective Pressure in Mice. PLOS Pathog 1(2):e11 [PubMed: 16220146]

88. Gnadig NF, Beaucourt S, Campagnola G, Borderia AV, Sanz-Ramos M, et al. 2012. Coxsackievirus B3 mutator strains are attenuated in vivo. PNAS 109(34):E2294–2303 [PubMed: 22853955]

89. Zeng J, Wang H, Xie X, Yang D, Zhou G, Yu L. 2013. An increased replication fidelity mutant of foot-and-mouth disease virus retains fitness in vitro and virulence in vivo. Antivir. Res 100(1):1–7 [PubMed: 23880348]

90. Zeng J, Wang H, Xie X, Li C, Zhou G, et al. 2014. Ribavirin-Resistant Variants of Foot-and-Mouth Disease Virus: the Effect of Restricted Quasispecies Diversity on Viral Virulence. J. Virol 88(8):4008–20 [PubMed: 24453363]

91. Sadeghipour S, McMinn PC. 2013. A study of the virulence in mice of high copying fidelity variants of human enterovirus 71. Virus Res 176(1–2):265–72 [PubMed: 23856384]

92. Rozen-Gagnon K, Stapleford KA, Mongelli V, Blanc H, Failloux A-B, et al. 2014. Alphavirus Mutator Variants Present Host-Specific Defects and Attenuation in Mammalian and Insect Models. PLOS Pathog 10(1):e1003877 [PubMed: 24453971]

93. Cheung PPH, Watson SJ, Choy K-T, Fun Sia S, Wong DDY, et al. 2014. Generation and characterization of influenza A viruses with altered polymerase fidelity. Nat. Commun 5(1):4794 [PubMed: 25183443]

94. Coffey LL, Beeharry Y, Borderia AV, Blanc H, Vignuzzi M. 2011. Arbovirus high fidelity variant loses fitness in mosquitoes and mice. PNAS 108(38):16038–43 [PubMed: 21896755]

95. Regoes RR, Hamblin S, Tanaka MM. 2013. Viral mutation rates: modelling the roles of within-host viral dynamics and the trade-off between replication fidelity and speed. Proc. R. Soc. B Biol. Sci 280(1750):20122047

96. Furio V, Moya A, Sanjuan R. 2005. The cost of replication fidelity in an RNA virus. PNAS 102(29):10233–37 [PubMed: 16006529]

97. Fitzsimmons WJ, Woods RJ, McCrone JT, Woodman A, Arnold JJ, et al. 2018. A speed-fidelity trade-off determines the mutation rate and virulence of an RNA virus. PLOS Biol 16(6):e2006459 [PubMed: 29953453]

98. Furió V, Moya A, Sanjuán R. 2007. The cost of replication fidelity in human immunodeficiency virus type 1. Proc. R. Soc. B Biol. Sci 274(1607):225–30

99. Xiao Y, Dolan PT, Goldstein EF, Li M, Farkov M, et al. 2017. Poliovirus intrahost evolution is required to overcome tissue-specific innate immune responses. Nat. Commun 8(1):375 [PubMed: 28851882]

100. Pfeiffer JK, Kirkegaard K. 2006. Bottleneck-mediated quasispecies restriction during spread of an RNA virus from inoculation site to brain. PNAS 103(14):5520–25 [PubMed: 16567621]

101. Kuss SK, Etheredge CA, Pfeiffer JK. 2008. Multiple Host Barriers Restrict Poliovirus Trafficking in Mice. PLOS Pathog 4(6):e1000082 [PubMed: 18535656]

102. McCune BT, Lanahan MR, tenOever BR, Pfeiffer JK. 2019. Rapid dissemination and monopolization of viral populations in mice revealed using a panel of barcoded viruses. J. Virol JVI.01590–19, jvi;JVI.01590–19v1

103. Aliota MT, Dudley DM, Newman CM, Weger-Lucarelli J, Stewart LM, et al. 2018. Molecularly barcoded Zika virus libraries to probe in vivo evolutionary dynamics. PLOS Pathog 14(3):e1006964 [PubMed: 29590202]

104. Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R. 2011. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. PNAS 108(50):20166–71 [PubMed: 22135472]

105. Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA. 2012. Detection of ultra- rare mutations by next-generation sequencing. PNAS 109(36):14508–13 [PubMed: 22853953]

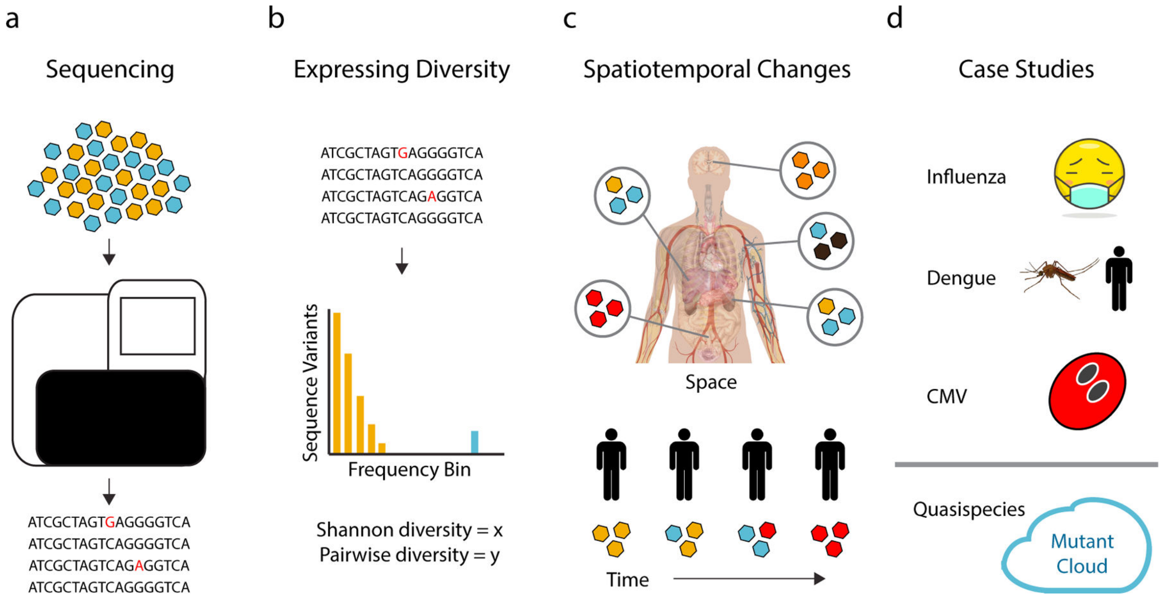

Figure 1.

Within-host diversity and virus evolution. (a) Advances in sequencing technology have revolutionized the study of within-host viral diversity. The type and frequency of mutations (red letters) in a population are easily obtained from NGS (b) Viral diversity can be measured and compared using a variety of metrics. Ideally, these metrics capture the varying impact of mutations present at high (blue bars) vs. low (yellow bars) frequency. (c) Differences in diversity across tissues (top, different colored viruses) or changes over time (bottom) can be used to model within-host viral dynamics due to selection and genetic drift. (d) Studies of within-host viral diversity have provided insights into the evolution of influenza virus, Dengue virus, and cytomegalovirus and the extent to which within-host virus populations can be accurately described as “quasispecies” or “mutant clouds.”

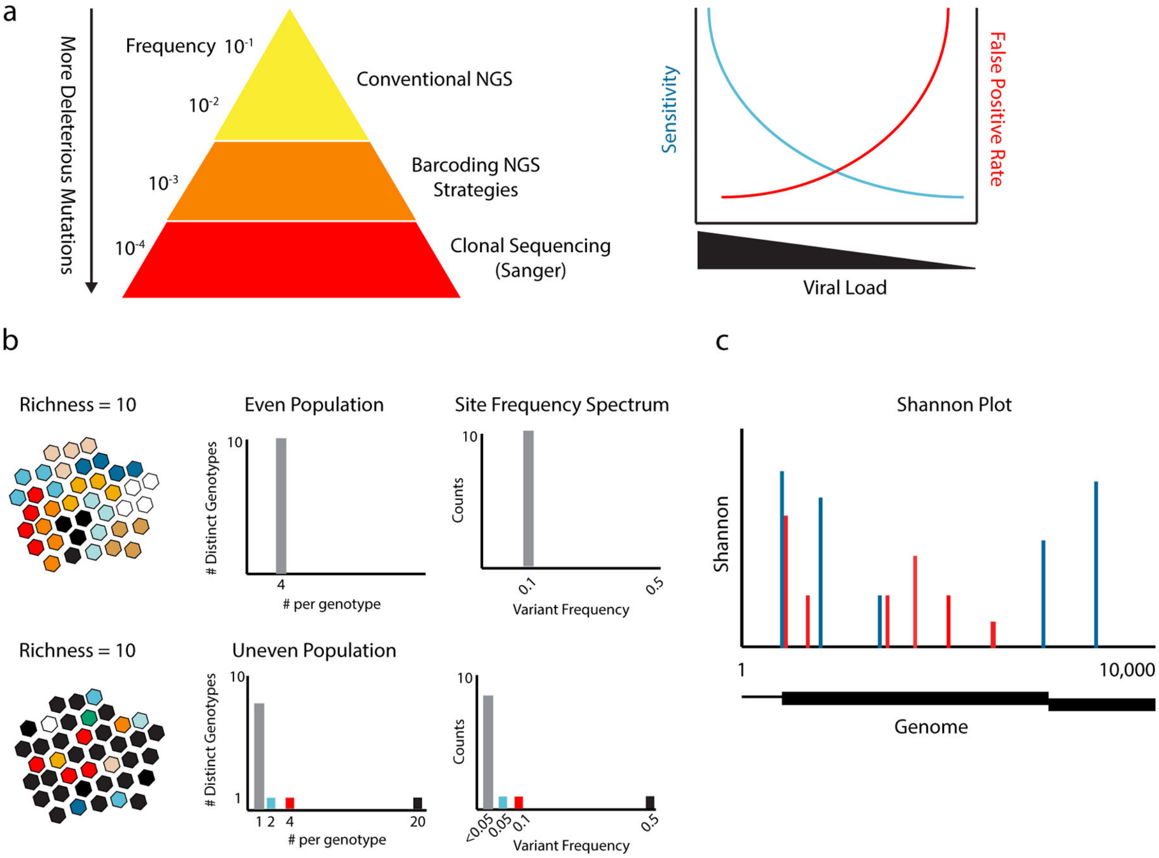

Figure 2.

Viral sequencing and diversity metrics. (a) Due to the negative impact of most mutations, the vast majority of sequence variants are relatively rare. The sensitivity and specificity of NGS for rare variant detection is highly dependent on the number of genomes sequenced and largely independent of read depth. Sensitivity drops at lower inputs, because rare variants can only be detected if the population is completely sampled. Smaller populations require more amplification, which propagates RT-PCR error. As a result, variants identified in low input populations are more likely to be false positives than true positives. (b) The diversity of two populations (different color viruses) expressed as richness, the number of variants or genotypes; evenness, the relative abundances of each variant in the population; and the site frequency spectrum, the numbers of different mutants and their respective frequencies. Both have a richness of 10 genotypes. The even population has equal numbers (n=4) of each genotype, and the 10 genotypes are present at a frequency of 0.1 (grey bars). The uneven population has the same 10 genotypes, but one is present at a frequency of 0.5 (black), one at 0.1 (red), and one at 0.05 (blue). The rest are singletons (summed as grey bars). (c) Shannon entropy at sites across a hypothetical 10 kilobase viral genome for two populations. In the genome, the noncoding regions are represented as lines and two different reading frames on the coding region are represented as boxes. The Shannon entropy at polymorphic sites for the two populations are shown as red and blue bars. Nonpolymorphic sites have an entropy of zero.

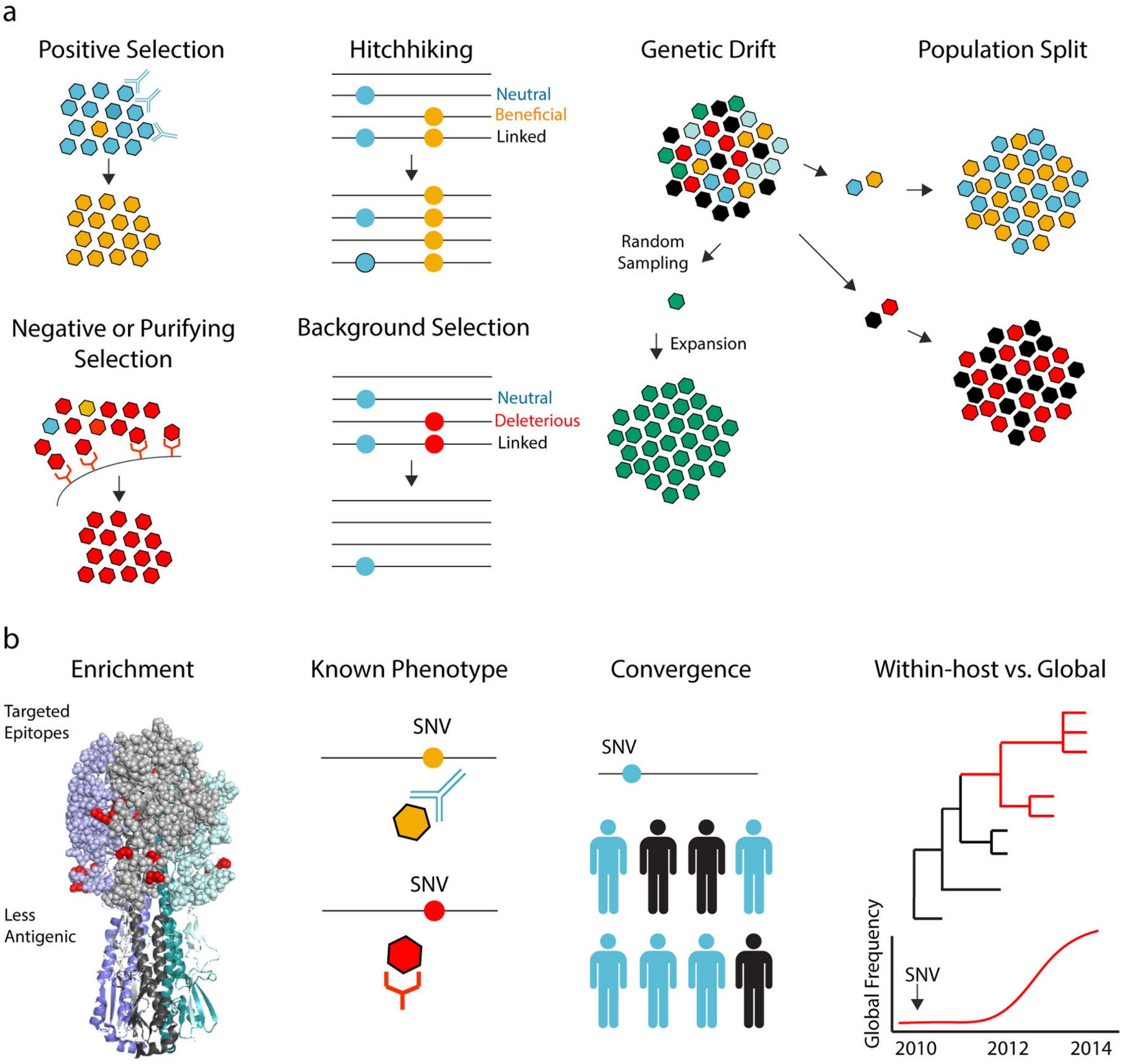

Figure 3.

Using within-host data to elucidate evolutionary dynamics. (a) The frequency of a mutation can increase or decrease due to selective and nonselective processes. Positive selection will increase the frequency of a beneficial mutation (e.g., a mutation that leads to antibody escape) and will decrease the frequency of a detrimental mutation (e.g., a surface protein mutant that can no longer bind its receptor). A neutral mutation (blue) can increase in frequency if it is linked to a beneficial mutation on the same genome (hitchhiking) or decrease in frequency if it is linked to a detrimental one (background selection). Genetic drift is a change in a mutation's frequency due to stochastic processes. It typically occurs in small populations due to random sampling. Genetic drift also occurs during population bottlenecks or splits followed by expansions. (b) Criteria for identifying within-host variants potentially under positive selection. From left to right: enrichment of high frequency variants in viral proteins or certain protein domains, shown here as nonsynonymous single nucleotide variants (SNV) in antigenic sites (red) modeled on the structure of the influenza hemagglutinin trimer; identification of a variant already known to mediate a certain phenotype, such as antibody escape or receptor binding; identification of the same mutation in multiple individual hosts not linked by transmission; identification of a within-host variant that is subsequently seen in different host populations or at different evolutionary scales.

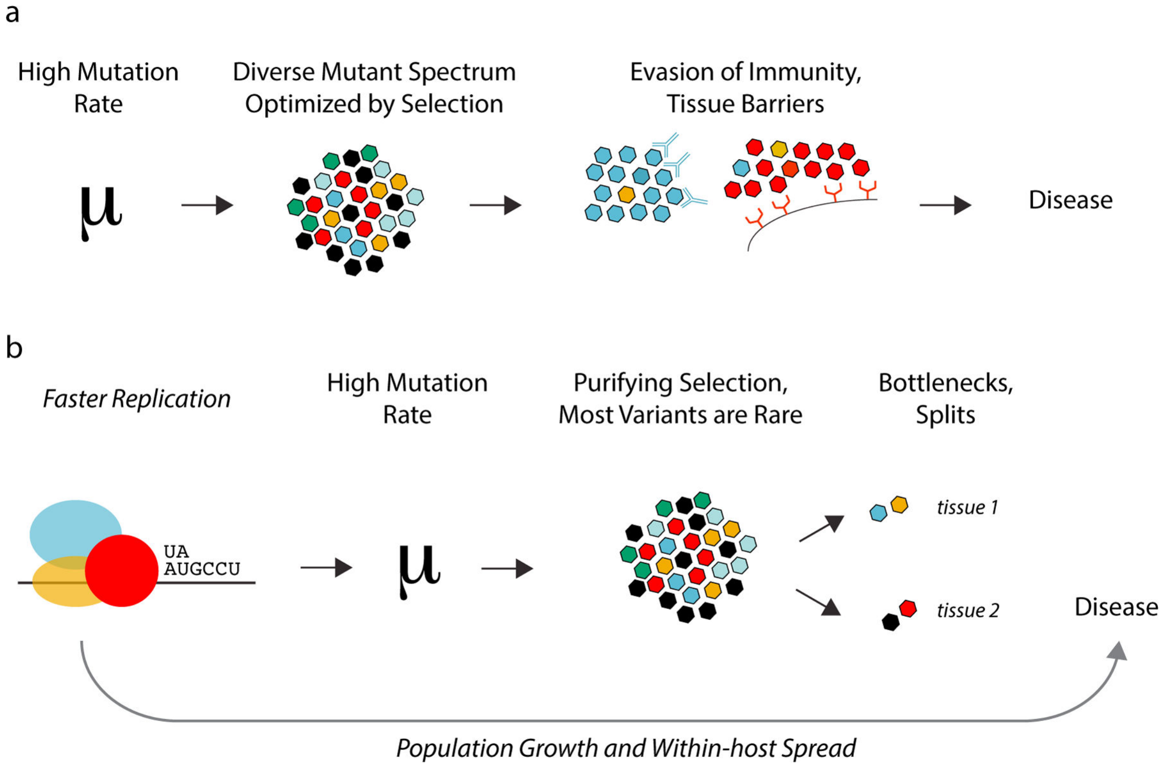
a
High Mutation Rate
Diverse Mutant Spectrum Optimized by Selection
Evasion of Immunity, Tissue Barriers
μ
Disease
b
Faster Replication
High Mutation Rate
Purifying Selection, Most Variants are Rare
Bottlenecks, Splits
UA
AUGCCU
μ
tissue 1
tissue 2
Disease
Population Growth and Within-host Spread

Figure 4.

Quasispecies diversity and virulence. (a) Diverse RNA virus populations are often referred to as quasispecies, mutant swarms, or mutant clouds. According to viral quasispecies theory, diversity is a determinant of viral phenotype and virulence. Here, high mutation rates lead to a diverse population, and the mutant spectrum is optimized by natural selection. Maintenance of a diverse population within hosts enables rapid adaptation and ultimately virulence. (b) An alternative model. Recent work from my laboratory suggests that the evolution of RNA dependent RNA polymerases is shaped by a speed-fidelity trade-off. Selection favors faster polymerases and faster polymerases are inherently more error-prone. Faster replication also leads to more rapid within-host spread and increased virulence. Empiric data across viral systems demonstrates that within-host diversity is rapidly lost or partitioned due to purifying selection, bottlenecks, and population splits. Thus, within-host diversity is neither maintained nor optimized by selection.