

Neutral Theory and Rapidly Evolving Viral Pathogens

Simon D.W. Frost,^{1,2} Brittany Rife Magalis,³ and Sergei L. Kosakovsky Pond^{*,3}

¹Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom

²The Alan Turing Institute, London, United Kingdom

³Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA

*Corresponding author: E-mail: spond@temple.edu.

Associate editor: Sudhir Kumar

Abstract

The evolution of viral pathogens is shaped by strong selective forces that are exerted during jumps to new hosts, confrontations with host immune responses and antiviral drugs, and numerous other processes. However, while undeniably strong and frequent, adaptive evolution is largely confined to small parts of information-packed viral genomes, and the majority of observed variation is effectively neutral. The predictions and implications of the neutral theory have proven immensely useful in this context, with applications spanning understanding within-host population structure, tracing the origins and spread of viral pathogens, predicting evolutionary dynamics, and modeling the emergence of drug resistance. We highlight the multiple ways in which the neutral theory has had an impact, which has been accelerated in the age of high-throughput, high-resolution genomics.

Key words: HIV, Influenza A virus, neutral theory, natural selection, population structure, molecular clock.

Introduction

Rapidly evolving pathogens, especially RNA viruses such as HIV-1, Influenza A virus (IAV), and Hepatitis C virus (HCV), are subject to some of the strongest evolutionary forces that have been reported in evolutionary biology. Within individual hosts, viruses undergo adaptive change at genomic sites that are targeted by the humoral (Bush et al. 1999; Frost et al. 2005), cellular (Kawashima et al. 2009), and innate immune responses (Rehermann 2009), sometimes sweeping to fixation in a matter of weeks (Henn et al. 2012), and subject to selective coefficients as high as 0.03 (on average) for individual escape mutations (Liu et al. 2006). Establishing infection in new hosts often requires numerous adaptive changes, such as receptor specificity adjustments in IAV (Taubenberger and Kash 2010) or SARS coronavirus (Li et al. 2005), or longer-term evolutionary arms races with host antiviral defense systems (Sawyer et al. 2004). Viruses are often subject to multiple, sometimes conflicting forces, for example imposed by the constraints within vectors, (Woelk and Holmes 2002) or the disparities between population and individual level fitness effects (Poon et al. 2007). The development of drug resistance is a key example of anthropogenic selection (Blair et al. 2015) that has been extensively studied and modeled (Hughes and Andersson 2015). Initial escape mutations almost invariably carry a fitness cost (Mammano et al. 2000), but are frequently compensated for by subsequent fitness-restoring mutations (Maisnier-Patin and Andersson 2004). Many seminal methods for detecting signatures of natural selection in molecular data were first applied to viral data (Nielsen and Yang 1998) or host immune loci (Hughes et al. 1990). Yet, despite tremendous observed divergence in many such systems (e.g., with HIV-1 isolates differing from each other by as much as

40%), discernible phenotypic differences between them in terms of virulence, transmissibility, and pathogenicity, while detectable (Pant Pai et al. 2012), are not dramatic. Moreover, within-host diversity is often surprisingly limited (Lemey et al. 2006). Here, we provide an overview of how much of the variation observed may be neutral, not necessarily in spite of, but because of, complex and dynamic selection pressures. We focus on HIV-1, a chronic infection, and influenza A virus (IAV), an acute infection, to illustrate these concepts, driven by the availability of large sequence databases and the research interest in these recalcitrant significant public health burdens.

Viral Diversity

RNA virus populations represent a large, yet finite set of replicating nucleotide sequences, for example, there are over 10^7 infected cells in a typical HIV infection (Chun et al. 1997). Combined with a short generation time and a high per-generation mutation rate, genetic variation is generated at a high rate, even in IAV infection, where the duration of infection is typically short (Poon et al. 2016). This variation can be harnessed to make both qualitative (e.g., clinical) and quantitative inferences from sampled viral gene sequence data. In this context, classical population genetics theory has been well developed, with the frequency of an allele, and thus diversity within a population, influenced by a combination of mutation, recombination or reassortment, natural selection, and population dynamics. RNA-dependent RNA polymerases are highly error-prone during replication, with misincorporation, or mutation, on the order of approximately one per 10^3 – 10^5 nucleotides polymerized (Ward et al. 1988). Infection of a cell by multiple, heterogeneous sequences can produce hybrid progeny as the result of recombining parental

sequences as frequently as once per generation for HIV (Schlub et al. 2010). Furthermore, when the viral genome consists of multiple, distinct segments of RNA, as in the case of influenza, reassortment of segments may occur in a dual infection leading to the production of new genotypes. This mechanism is postulated to account for the major antigenic shift of IAV, and is responsible for major pandemics (Vijaykrishna et al. 2015). Neutral theory treats these processes as stochastic in nature and assumes that most mutations are selectively neutral or slightly deleterious. Although considerable effort has gone into debating whether an underlying neutral model is applicable to rapidly evolving viruses (see Leigh Brown [1997] and citations thereof), the assumption of selective neutrality for the majority of mutations has served as a useful, and even necessary null hypothesis for evolutionary hypothesis testing in the context of highly diverse viral populations.

Intrahost Population Dynamics

Despite the capacity of viral populations within infected individuals to generate genetic variation, viral diversity at any one time may be relatively low. Viral infections are typically founded by a small number of infectious particles, resulting in a bottleneck that can dramatically reduce genetic variation (McCrone and Lauring 2018). In addition, selectively neutral or even somewhat deleterious mutations can arise to high frequency during these bottlenecks. Albeit less dramatic, bottlenecks may even be present within infected individuals, due to spatial structuring of target cells (Frost et al. 2001; Salemi and Rife 2016), intrinsic variation in the number of progeny released by an infected cell, and extrinsic noise driving temporal variation in viral production.

In immunologically naïve individuals who have not previously encountered the virus, selection pressures exerted by the adaptive immune response may be absent during the early stages of infection. During acute HIV-1 infection, evolution is very well approximated by the neutral process (Lee et al. 2009). However, even following the establishment of potent antiviral responses, mutations with little or no impact on fitness may change in frequency; selective sweeps can result in mutations rising to high frequency due to hitchhiking (Zanini and Neher 2013) and may drive down diversity at linked mutations, as has been demonstrated during the emergence of drug resistance mutants in HIV infection (Pennings et al. 2014). If sweeps are frequent, clonal interference may occur such that even selected mutations may evolve in a manner more typical of neutral mutations (Gerrish and Lenski 1998)—a phenomenon dubbed “genetic draft” (Gillespie 2000).

Between-Host Dynamics

In contrast to within-host dynamics, evidence of selection at the between-host level is weaker. Viruses such as HIV-1 and HCV exhibit phylogenies that closely resemble those generated by a time-varying coalescent model. Once corrections are made for changing generation time over the course of the epidemic, the level of genetic variation at the population level

is broadly consistent with the number of infected individuals (Frost and Volz 2013). In addition to the structure of the phylogenetic tree being consistent with neutral expectations, even selected mutations may evolve at the population level as if affected by drift. Over multiple infections, viruses experience different host environments, and the resulting fluctuating selection pressure can also result in additional “noise” more consistent with a small census population size (Poon et al. 2007; Fryer et al. 2010). Consequently, HIV-1 has adapted rather slowly to humans since its emergence as a pandemic pathogen (Fryer et al. 2012).

Molecular Clock and Divergence Dating

The divergence of viral molecular sequences over time, whether within or among hosts, becomes a powerful tool in investigating the temporal origins of viral strains or outbreaks, particularly when this divergence exhibits a clock-like behavior (Gojobori et al. 1990). The early observation that proteins evolve at a constant, or clock-like, rate over time (Zuckerandl and Pauling 1965) was a natural argument in support of neutral theory. The divergence of HIV-1 and IAV has been repeatedly found to conform to the molecular clock hypothesis, both at the level of the entire viral population (Buonagurio et al. 1986; Korber et al. 2000), and within individual hosts (Poon et al. 2012). The original strict clock estimates of the time of introduction for the HIV-1 group M pandemic (Korber et al. 2000) were only slightly refined using much more complex models (Salemi et al. 2001; Faria et al. 2014), and studies of known transmission histories demonstrated the applicability of the molecular clock (Leitner and Albert 1999) to HIV-1 transmission studies, albeit with some caveats due to the within-host evolution and transmission dynamics. A different overall rate of nonsynonymous relative to synonymous substitutions is often observed due to differing degrees of selection pressure (Gojobori et al. 1990; Frost et al. 2005), but the clock based on synonymous substitutions appears to be a good proxy for within-host replication dynamics (Lemey et al. 2007).

The action of some types of episodic natural selection is compatible with clock-like evolution (Gillespie 1986). However, several important selective regimes deviate from the strict (or even relaxed) molecular clocks. Selective regimes are usually very different between interior branches of viral phylogenies, reflective of long-term evolutionary pressures informed by transmission and largely purifying selection, and terminal branches that represent within-host evolution that is more neutral or even maladaptive at the level of the population (Pond et al. 2006; Pybus et al. 2007). This may lead to strong deviations from neutral clock-like behavior (Wertheim and Kosakovsky Pond 2011). Differing population dynamics in different host species or geographical location may result in highly correlated evolutionary rates in the phylogeny. This can confound estimates of the molecular clock, but can be corrected by incorporating external information, such as host range (Worobey et al. 2014; Frost et al. 2015). Indeed, incorporating detailed sequence information such as risk group, location of isolation, and even relative geographical

distances into a relaxed molecular clock model, a growing field known as phylodynamics (Grenfell et al. 2004), has been used effectively to expand on previous molecular epidemiology studies to include estimates of spatial origins, migration and transmission rates, and predictors of disease spread (Lemey et al. 2009; Faria et al. 2014; Rife et al. 2017). In HIV infection, there is evidence to suggest that the rate of the evolution at the population level is intimately associated with the rate of transmission. Rapid transmission restricts the time for the virus to diversify within the host, which in turn leads to a low rate of the molecular clock at the between-host level (Maljkovic Berry et al. 2007).

Drug Resistance Development and Reversion

Rapid fixation of escape mutations in response to treatment with incompletely suppressive therapy or a therapy with a low genetic barrier of resistance is a classical example of natural selection in action (Belshe et al. 1989; Larder and Kemp 1989). However, evolutionary dynamics leading to the rise of drug resistance and following the acquisition of drug resistance mutations are anything but simple. Although some mutations can dramatically reduce fitness of resistance variants in comparison with wildtype when the drug is not present (Brenner et al. 2002), many have rather minor effects on fitness (Kühnert et al. 2018). Upon transmission to a drug-naïve host, many drug resistance mutations revert to wildtype alleles, although at markedly different rates (Little et al. 2008), but some remain fixed in new hosts (Castro et al. 2013). There has been considerable debate as to whether drug-resistant mutations are sourced from standing genetic variation (i.e., preexisting mutations at low frequency) or from new mutations that occur during therapy. Studies have reported many instances of low-frequency variants being swept to fixation by the action of a drug (Tsibris et al. 2009). However, there is also evidence that drug resistance mutations are repeatedly generated and turn over during error-prone replication (Gianella et al. 2011).

Largely consistent patterns of escape mutations (Paredes et al. 2017) together with strong phenotypic effects would suggest a predominant role for deterministic evolution (Coffin 1995). However, the large variation in the timing and development of drug resistance among patients has suggested a notable stochastic component (Frost et al. 2000; Pennings et al. 2014), further reinforced by the complex pharmacokinetics of antiviral drugs in different tissues (Lorenzo-Redondo et al. 2016). The neutral theory enhanced our understanding of this variation and its relationship to epistatic interactions between resistance sites and the conditionally neutral genetic background, that is, sites for which mutation does not alter the fitness of the pathogen, but may alter the fitness effects of subsequent mutations. Resistance mutations can become highly advantageous (or entrenched) despite being destabilizing in the wildtype background, suggesting that anticipation of epistatic effects is important for the design of future therapies (Flynn et al. 2017).

Evolutionary Plasticity and Neutral Fitness Landscapes

A central goal of evolutionary biology is to understand how genetic variation underpins phenotypic variation, and ultimately fitness. The way pleiotropic and epistatic gene effects are organized within the genotype–phenotype (G–P) map is expected to play a pivotal role in the ability of viruses to adapt and evolve. For example, HIV-1 resistance mutations confer different degrees of resistance (Petropoulos et al. 2000; Rhee et al. 2004), vary in their degree of cross-resistance to different drugs or drug classes (Harrigan et al. 2001), and differ in the fitness costs incurred in the absence of treatment (Croteau et al. 1997; Martinez-Picado et al. 1999; Mammano et al. 2000). The fitness landscape of these resistance mutations, particularly the effect of their epistatic interactions on this landscape, has been a challenge to characterize quantitatively, although recent progress has been made thanks to high-throughput data generation (e.g., Hinkley et al. 2011). The finding by Kouyos et al. (2012) of an increasing ability of populations to move across the fitness landscape without changing their fitness with increasing magnitude of epistatic effects was an influential factor in understanding the role of the conditionally neutral genetic background in phenotypic variation, specifically the connectivity of neutral networks (Huynen 1996; Huynen et al. 1996), within a real system. The observed fitness landscape for drug-resistant HIV-1, though rugged, is exclusively due to the cancelling out of selective effects and the large fraction of relatively equally fit viral variants with significantly differing genotypic profiles (Kouyos et al. 2012). Fontana et al. (1993a, 1993b) found that landscapes in which fitness is predicted by RNA secondary structure combine neutrality and ruggedness similarly to the landscapes described specifically for HIV drug resistance.

Similar to HIV-1 drug resistance, IAV is capable of evading immune recognition through antigenic drift of its surface proteins, hemagglutinin (HA) and neuraminidase (NA), complicating long-term control of the disease through vaccination (Smith et al. 1999). One of the most striking complications for the reconstruction of the IAV G–P map for HA is that, although genetic change is gradual, antigenic change is punctuated. HA inhibition assays show that H3N2 sequences can be clustered, each with unique antigenic properties. Between 1968 and 2003, these clusters emerged and replaced each other within as little as 2–8 years (Smith et al. 2004). Empirical evidence suggests that there is immunity against one strain confers almost complete immunity against other strains from the same antigenic cluster (Gill and Murphy 1977), whereas cross-immunity is as low as 60–85% between clusters adjacent in time (Gill and Murphy 1977; Meiklejohn et al. 1978). Interpreting influenza clusters in terms of intertwined neutral networks that map both main mutations and epistatic interactions corroborated phylogenetic evidence (also modeled under neutral evolution) (Grenfell et al. 2004) in that weak within-cluster selection and the selective sweeps that accompany cluster transitions are sufficient to explain the restricted interpandemic diversity of HA in light of large antigenic changes (Koelle et al. 2006).

Next-Generation Sequencing-Based Studies

The introduction of next-generation sequencing (NGS) technologies afforded an opportunity to measure intrahost viral diversity with remarkable precision and resolution. In addition to relatively rapid adaptive change driven by the immune system (Henn et al. 2012) or drug pressure (Flynn et al. 2015), HIV-1 undergoes longer term evolutionary changes that include reversion of substitutions at positions where changes acquired in previous hosts are no longer advantageous (Zanini et al. 2015)—a pattern also reported for HCV (Ray et al. 2005).

The ability to measure intrahost populations using NGS more reliably has also revealed that, whereas continual positive selection of antigenically drifted variants drives global patterns of IAV population dynamics, stochastic processes such as strain migration and within-clade reassortment dominate IAV evolution within the human host, with minority variants rarely shared among individuals (McCrone and Lauring 2018). Single-cell analysis of IAV infection has demonstrated that high variation in the number of progeny released by an infected cell as well as extrinsic noise affecting viral replication can drive down genetic variation and increase the importance of stochastic effects, particularly early on in infection (Heldt et al. 2015). The same is not said to be true in pigs, however, as diversity within even partially immune swine is significant and highly dynamic over the course of infection (Diaz et al. 2015), underscoring the role played by within-host evolution in antigenic shift and the need for further NGS studies of genetic plasticity in intermediate hosts.

Deep Mutational Scanning

A high mutation rate, short generation time, and strong selection of both single-point and epistatic mutations certainly play a role in the capacity of HIV and IAV evolutionary escape from immunity. However, other viruses with comparable mutation rates, such as measles, show little propensity for antigenic change (Sheshberadaran et al. 1983; Duffy et al. 2008), despite obvious selective benefits. Several explanations have been offered to account for these differences (Koelle et al. 2006; Lipsitch and O'Hagan 2007; Heaton et al. 2013), but the impact of evolutionary plasticity has been difficult to test prior to developments in NGS, as the full understanding of mutational tolerance requires in-depth consideration of all possible mutations at each site. This type of analysis cannot be done simply by examining variability among observed naturally occurring viruses, since the filtering lens of selection cannot be removed or corrected for. Moreover, only a fraction of theoretically tolerable mutations have been fixed in natural viral populations due to the finite timespan during which evolution has been exploring possible sequences (Kondrashov et al. 2010). The application of NGS techniques to *in vitro* functional selection of large mutant protein libraries ($\sim 10^5$), referred to as deep mutational scanning (DMS, Fowler et al. [2010]; Araya and Fowler [2011]), has proven to be a rapid and inexpensive methodology for exploring individual viral fitness landscapes that would be difficult to perform using structural or evolutionary analyses alone. DMS

investigations of IAV demonstrated the ability of HA to tolerate a remarkably wide array of single-point mutations without loss of function, particularly at antigenic sites (Thyagarajan and Bloom 2014), providing support for the role of a vast nearly neutral landscape in IAV evolution as well, even in the absence of epistatic compensation. A similar study involving the rapidly adapting HIV envelope glycoprotein (Env) (Haddock et al. 2016) uncovered a reduced tolerance to amino acid mutations in broadly neutralizing antibody (nAb) epitopes, consistent with the conserved nature of these regions and crucial for further studies of the role of nAbs in HIV prevention. Despite an overall tolerance of synonymous mutations in Env, reduced tolerance in the region of the Rev-Response element (RRE), one of several conserved HIV-1 RNA structures (Watts et al. 2009), was observed, emphasizing the dangers of overestimating the neutral effects of synonymous mutations. The stark contrast in RRE synonymous tolerance to the rest of Env also suggests that while there may be strong local RNA-structure constraints, many genomic regions are quite tolerant to change (Watts et al. 2009; Pollom et al. 2013).

The results of the described DMS studies have presented evidence comparable to those of evolutionary methods deeply rooted in neutral theory, while also offering opportunities for the development of experimentally informed evolutionary models as improvements to modern phylogenetic approaches (Bloom 2014). Further examination could reveal additional sites targeted by the wide array of drugs available and by other arms of the host immune response. Additional studies using viral replication in alternative cell populations should also make it possible to isolate the specific role of these selective pressures in shaping HIV and IAV evolution in the context of evolutionary plasticity.

Conclusions

The use of models and predictions derived from the neutral theory of evolution continues to provide critical information as to the evolutionary and population dynamics of pathogens and the forces driving these dynamics. Additionally, analyses of neutral networks connecting genotypic and phenotypic variation have given detailed descriptions of the role of single-point mutations and epistatic interactions in shaping the viral fitness landscape, particularly for HIV and IAV. Advanced sequencing techniques applied to these and similar studies continue to provide deeper understanding of viral adaptation, while uncovering new evidence as to how evolutionary plasticity can vary among rapidly evolving RNA viruses that differ so much with respect to antigenic change.

Deep sequencing and other high throughput technologies are poised to vastly broaden our understanding of viral evolution. We are dramatically expanding the catalog of known viruses across species (Shi et al. 2018), mapping the interplay between viral evolution and immune responses in a single host (Liao et al. 2013), studying the dynamics of individual infected cells (Heldt et al. 2015), and developing much more sensitive diagnostic tools (Van Laethem et al. 2015). When coupled with error-reducing experimental techniques

(Jabara et al. 2011), deep sequencing is directly measuring population properties that could only be estimated before, and offer new avenues of investigation through the application of machine learning approaches to the “big data” generated by these techniques. These experimental data will serve as invaluable checks of the validity of theoretical predictions, including the neutral theory, and, as is often the case, prompt their refinement and reassessment.

Acknowledgments

This study was supported in part by grants R01 AI134384 (NIH/NIAID), R01 GM093939 (NIH/NIGMS), and U01 GM110749 (NIH/NIGMS).

References

- Araya CL, Fowler DM. 2011. Deep mutational scanning: assessing protein function on a massive scale. *Trends Biotechnol.* 29(9): 435–442.
- Belshe RB, Burk B, Newman F, Cerruti RL, Sim IS. 1989. Resistance of influenza A virus to amantadine and rimantadine: results of one decade of surveillance. *J Infect Dis.* 159(3): 430–435.
- Blair JMA, Webber MA, Baylay AJ, Ogbolu DO, Piddock LJV. 2015. Molecular mechanisms of antibiotic resistance. *Nat Rev Microbiol.* 13(1): 42–51.
- Bloom JD. 2014. An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Mol Biol Evol.* 31(8): 1956–1978.
- Brenner BG, Routy J-P, Petrella M, Moisi D, Oliveira M, Detorio M, Spira B, Essabog V, Conway B, Lalonde R, et al. 2002. Persistence and fitness of multidrug-resistant human immunodeficiency virus type 1 acquired in primary infection. *J Virol.* 76(4): 1753–1761.
- Buonagurio DA, Nakada S, Parvin JD, Krystal M, Palese P, Fitch WM. 1986. Evolution of human influenza A viruses over 50 years: rapid, uniform rate of change in NS gene. *Science* 232(4753): 980–982.
- Bush RM, Fitch WM, Bender CA, Cox NJ. 1999. Positive selection on the H3 hemagglutinin gene of human influenza virus A. *Mol Biol Evol.* 16(11): 1457–1465.
- Castro H, Pillay D, Cane P, Asboe D, Cambiano V, Phillips A, Dunn DT, Aitken C, Asboe D, Webster D, UK Collaborative Group on HIV Drug Resistance, et al. 2013. Persistence of HIV-1 transmitted drug resistance mutations. *J Infect Dis.* 208(9): 1459–1463.
- Chun T-W, Carruth L, Finzi D, Shen X, DiGiuseppe JA, Taylor H, Hermankova M, Chadwick K, Margolick J, Quinn TC. 1997. Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* 387(6629): 183–188.
- Coffin JM. 1995. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science* 267(5197): 483–489.
- Croteau G, Doyon L, Thibeault D, McKercher G, Pilote L, Lamarre D. 1997. Impaired fitness of human immunodeficiency virus type 1 variants with high-level resistance to protease inhibitors. *J Virol.* 71(2): 1089–1096.
- Diaz A, Enomoto S, Romagosa A, Sreevatsan S, Nelson M, Culhane M, Torremorell M. 2015. Genome plasticity of triple-reassortant H1N1 influenza A virus during infection of vaccinated pigs. *J Gen Virol.* 96(10): 2982–2993.
- Duffy S, Shackleton LA, Holmes EC. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet.* 9(4): 267.
- Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, Tatem AJ, Sousa JD, Arinaminpathy N, Pèpin J, et al. 2014. HIV epidemiology: the early spread and epidemic ignition of HIV-1 in human populations. *Science* 346(6205): 56–61.
- Flynn JK, Ellenberg P, Duncan R, Ellett A, Zhou J, Sterjovskij J, Cashin K, Borm K, Gray LR, Lewis M, et al. 2017. Analysis of clinical HIV-1 strains with resistance to maraviroc reveals strain-specific resistance mutations, variable degrees of resistance, and minimal cross-resistance to other CCR5 antagonists. *AIDS Res Hum Retroviruses* 33(12): 1220–1235.
- Flynn WF, Chang MW, Tan Z, Oliveira G, Yuan J, Okulicz JF, Torbett BE, Levy RM. 2015. Deep sequencing of protease inhibitor resistant HIV patient isolates reveals patterns of correlated mutations in Gag and protease. *PLoS Comput Biol.* 11(4): e1004249.
- Fontana W, Stadler PF, Bornberg-Bauer EG, Griesmacher T, Hofacker IL, Tacker M, Tarazona P, Weinberger ED, Schuster P. 1993. RNA folding and combinatorial landscapes. *Phys Rev E* 47(3): 2083.
- Fontana W, Konings DA, Stadler PF, Schuster P. 1993. Statistics of RNA secondary structures. *Biopolymers* 33(9): 1389–1404.
- Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, Fields S. 2010. High-resolution mapping of protein sequence-function relationships. *Nat Methods* 7(9): 741.
- Frost SD, Volz EM. 2013. Modelling tree shape and structure in viral phylodynamics. *Philos Trans R Soc B.* 368(1614): 20120208.
- Frost SD, Nijhuis M, Schuurman R, Boucher CA, Brown AJ. 2000. Evolution of lamivudine resistance in human immunodeficiency virus type 1-infected individuals: the relative roles of drift and selection. *J Virol.* 74(14): 6262–6268.
- Frost SD, Dumaourier MJ, Wain-Hobson S, Brown AJ. 2001. Genetic drift and within-host metapopulation dynamics of HIV-1 infection. *Proc Natl Acad Sci U S A.* 98(12): 6975–6980.
- Frost SDW, Wrin T, Smith DM, Pond SLK, Liu Y, Paxinos E, Chappey C, Galovich J, Beauchaine J, Petropoulos CJ, et al. 2005. Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. *Proc Natl Acad Sci U S A.* 102(51): 18514–18519.
- Frost SDW, Pybus OG, Gog JR, Viboud C, Bonhoeffer S, Bedford T. 2015. Eight challenges in phylodynamic inference. *Epidemics* 10:88–92.
- Fryer HR, Frater J, Duda A, Roberts MG, Investigators ST, Phillips RE, McLean AR. 2010. Modelling the evolution and spread of HIV immune escape mutants. *PLoS Pathog.* 6(11): e1001196.
- Fryer HR, Frater J, Duda A, Palmer D, Phillips RE, McLean AR. 2012. Cytotoxic T-lymphocyte escape mutations identified by HLA association favor those which escape and revert rapidly. *J Virol.* 86(16): 8568–8580.
- Gerrish PJ, Lenski RE. 1998. The fate of competing beneficial mutations in an asexual population. *Genetica* 102-103:127–144.
- Gianella S, Delpont W, Pacold ME, Young JA, Choi JY, Little SJ, Richman DD, Kosakovsky Pond SL, Smith DM. 2011. Detection of minority resistance during early HIV-1 infection: natural variation and spurious detection rather than transmission and evolution of multiple viral variants. *J Virol.* 85(16): 8359–8367.
- Gill P, Murphy A. 1977. Naturally acquired immunity to influenza type A: a further prospective study. *Med J Aust.* 2(23): 761–765.
- Gillespie JH. 1986. Natural selection and the molecular clock. *Mol Biol Evol* 3(2): 138–155.
- Gillespie JH. 2000. Genetic drift in an infinite population. The pseudo-hitchhiking model. *Genetics* 155:909–919.
- Gojbori T, Moriyama EN, Kimura M. 1990. Molecular clock of viral evolution, and the neutral theory. *Proc Natl Acad Sci U S A.* 87(24): 10015–10018.
- Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, Holmes EC. 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303(5656): 327–332.
- Haddox HK, Dingens AS, Bloom JD. 2016. Experimental estimation of the effects of all amino-acid mutations to HIV’s envelope protein on viral replication in cell culture. *PLoS Pathog.* 12(12): e1006114.
- Harrigan PR, Montaner JS, Wegner SA, Verbiest W, Miller V, Wood R, Larder BA. 2001. World-wide variation in HIV-1 phenotypic susceptibility in untreated individuals: biologically relevant values for resistance testing. *Aids* 15(13): 1671–1677.
- Heaton NS, Sachs D, Chen C-J, Hai R, Palese P. 2013. Genome-wide mutagenesis of influenza virus reveals unique plasticity of the hemagglutinin and NS1 proteins. *Proc Natl Acad Sci U S A.* 110(50): 20248–20253.
- Heldt FS, Kupke SY, Dorl S, Reichl U, Frensing T. 2015. Single-cell analysis and stochastic modelling unveil large cell-to-cell variability in influenza A virus infection. *Nat Commun.* 6(1): 8938.

- Henn MR, Boutwell CL, Charlebois P, Lennon NJ, Power KA, Macalalad AR, Berlin AM, Malboeuf CM, Ryan EM, Gnerre S, et al. 2012. Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog.* 8(3): e1002529.
- Hinkley T, Martins J, Chappey C, Haddad M, Stawiski E, Whitcomb JM, Petropoulos CJ, Bonhoeffer S. 2011. A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. *Nat Genet.* 43(5): 487–489.
- Hughes AL, Ota T, Nei M. 1990. Positive darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. *Mol Biol Evol.* 7(6): 515–524.
- Hughes D, Andersson DI. 2015. Evolutionary consequences of drug resistance: shared principles across diverse targets and organisms. *Nat Rev Genet.* 16(8): 459–471.
- Huynen MA. 1996. Exploring phenotype space through neutral evolution. *J Mol Evol.* 43(3): 165–169.
- Huynen MA, Stadler PF, Fontana W. 1996. Smoothness within ruggedness: the role of neutrality in adaptation. *Proc Natl Acad Sci U S A.* 93(1): 397–401.
- 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. *Proc Natl Acad Sci U S A.* 108(50): 20166–20171.
- Kawashima Y, Pfafferoth K, Frater J, Matthews P, Payne R, Addo M, Gatanaga H, Fujiwara M, Hachiya A, Koizumi H, et al. 2009. Adaptation of HIV-1 to human leukocyte antigen class I. *Nature* 458(7238): 641–645.
- Koelle K, Cobey S, Grenfell B, Pascual M. 2006. Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* 314(5807): 1898–1903.
- Kondrashov AS, Povolotskaya IS, Ivankov DN, Kondrashov FA. 2010. Rate of sequence divergence under constant selection. *Biol Direct.* 5:5.
- Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, Hahn BH, Wolinsky S, Bhattacharya T. 2000. Timing the ancestor of the HIV-1 pandemic strains. *Science* 288(5472): 1789–1796.
- Kouyos RD, Leventhal GE, Hinkley T, Haddad M, Whitcomb JM, Petropoulos CJ, Bonhoeffer S. 2012. Exploring the complexity of the HIV-1 fitness landscape. *PLoS Genet.* 8(3): e1002551.
- Kühnert D, Kouyos R, Shirreff G, Pečerska J, Scherrer AU, Böni J, Yerly S, Klimkait T, Aubert V, Günthard HF. 2018. Quantifying the fitness cost of HIV-1 drug resistance mutations through phylodynamics. *PLoS Pathog.* 14(2): e1006895.
- Larder BA, Kemp SD. 1989. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science* 246(4934): 1155–1158.
- Lee HY, Giorgi EE, Keele BF, Gaschen B, Athreya GS, Salazar-Gonzalez JF, Pham KT, Goepfert PA, Kilby JM, Saag MS, et al. 2009. Modeling sequence evolution in acute HIV-1 infection. *J Theor Biol.* 261(2): 341–360.
- Brown AJL. 1997. Analysis of HIV-1 env gene sequences reveals evidence for a low effective number in the viral population. *Proc Natl Acad Sci U S A.* 94(5): 1862–1865.
- Leitner T, Albert J. 1999. The molecular clock of HIV-1 unveiled through analysis of a known transmission history. *Proc Natl Acad Sci U S A.* 96(19): 10752–10757.
- Lemey P, Rambaut A, Pybus OG. 2006. HIV evolutionary dynamics within and among hosts. *AIDS Rev.* 8(3): 125–140.
- Lemey P, Kosakovsky Pond SL, Drummond AJ, Pybus OG, Shapiro B, Barroso H, Taveira N, Rambaut A. 2007. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. *PLoS Comput Biol.* 3(2): e29.
- Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. *PLoS Comput Biol.* 5(9): e1000520.
- Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, Wong S-K, Huang I-C, Xu K, Vasilieva N, et al. 2005. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* 24(8): 1634–1643.
- Liao H-X, Lynch R, Zhou T, Gao F, Alam SM, Boyd SD, Fire AZ, Roskin KM, Schramm CA, Zhang Z, et al. 2013. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* 496(7446): 469–476.
- Lipsitch M, O'Hagan JJ. 2007. Patterns of antigenic diversity and the mechanisms that maintain them. *J R Soc Interface* 4(16): 787–802.
- Little SJ, Frost SDW, Wong JK, Smith DM, Pond SLK, Ignacio CC, Parkin NT, Petropoulos CJ, Richman DD. 2008. Persistence of transmitted drug resistance among subjects with primary human immunodeficiency virus infection. *J Virol.* 82(11): 5510–5518.
- Liu Y, McNevin J, Cao J, Zhao H, Genowati I, Wong K, McLaughlin S, McSweyn MD, Diem K, Stevens CE, et al. 2006. Selection on the human immunodeficiency virus type 1 proteome following primary infection. *J Virol.* 80(19): 9519–9529.
- Lorenzo-Redondo R, Fryer HR, Bedford T, Kim E-Y, Archer J, Pond SLK, Chung Y-S, Penugonda S, Chipman J, Fletcher CV, et al. 2016. Persistent HIV-1 replication maintains the tissue reservoir during therapy. *Nature* 530(7588): 51–56.
- Maisnier-Patin S, Andersson DI. 2004. Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. *Res Microbiol.* 155(5): 360–369.
- Berry IM, Ribeiro R, Kothari M, Athreya G, Daniels M, Lee HY, Bruno W, Leitner T. 2007. Unequal evolutionary rates in the human immunodeficiency virus type 1 (HIV-1) pandemic: the evolutionary rate of HIV-1 slows down when the epidemic rate increases. *J Virol.* 81(19): 10625–10635.
- Mammano F, Trouplin V, Zennou V, Clavel F. 2000. Retracing the evolutionary pathways of human immunodeficiency virus type 1 resistance to protease inhibitors: virus fitness in the absence and in the presence of drug. *J Virol.* 74(18): 8524–8531.
- Martinez-Picado J, Savara AV, Sutton L, Richard T. 1999. Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. *J Virol.* 73(5): 3744–3752.
- McCrone JT, Lauring AS. 2018. Genetic bottlenecks in intraspecies virus transmission. *Curr Opin Virol.* 28:20–25.
- Meiklejohn G, Eickhoff TC, Graves P, IJ. 1978. Antigenic drift and efficacy of influenza virus vaccines, 1976–1977. *J Infect Dis.* 138(5): 618–624.
- Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148(3): 929–936.
- Pant Pai N, Shivkumar S, Cajas JM. 2012. Does genetic diversity of HIV-1 non-B subtypes differentially impact disease progression in treatment-naïve HIV-1-infected individuals? A systematic review of evidence: 1996–2010. *J Acquir Immune Defic Syndr.* 59(4): 382–388.
- Paredes R, Tzou PL, van Zyl G, Barrow G, Camacho R, Carmona S, Grant PM, Gupta RK, Hamers RL, Harrigan PR, et al. 2017. Collaborative update of a rule-based expert system for HIV-1 genotypic resistance test interpretation. *PLoS One* 12(7): e0181357.
- Pennings PS, Kryazhimskiy S, Wakeley J. 2014. Loss and recovery of genetic diversity in adapting populations of HIV. *PLoS Genet.* 10(1): e1004000.
- Petropoulos CJ, Parkin NT, Limoli KL, Lie YS, Wrinn T, Huang W, Tian H, Smith D, Winslow GA, Capon DJ, Whitcomb JM. 2000. A novel phenotypic drug susceptibility assay for human immunodeficiency virus type 1. *Antimicrob Agents Chemother.* 44(4): 920–928.
- Pollom E, Dang KK, Potter EL, Gorelick RJ, Burch CL, Weeks KM, Swanstrom R. 2013. Comparison of SIV and HIV-1 genomic RNA structures reveals impact of sequence evolution on conserved and non-conserved structural motifs. *PLoS Pathog.* 9(4): e1003294.
- Pond SLK, Frost SDW, Grossman Z, Gravenor MB, Richman DD, Brown AJL. 2006. Adaptation to different human populations by HIV-1 revealed by codon-based analyses. *PLoS Comput Biol.* 2(6): e62.
- Poon AFY, Kosakovsky Pond SL, Bennett P, Richman DD, Leigh Brown AJ, Frost SDW. 2007. Adaptation to human populations is revealed by within-host polymorphisms in HIV-1 and hepatitis C virus. *PLoS Pathog.* 3(3): e45.
- Poon AFY, Swenson LC, Bunnik EM, Edo-Matas D, Schuitemaker H, van 't Wout AB, Harrigan PR. 2012. Reconstructing the dynamics of HIV evolution within hosts from serial deep sequence data. *PLoS Comput Biol.* 8(11): e1002753.

- Poon LLM, Song T, Rosenfeld R, Lin X, Rogers MB, Zhou B, Sebra R, Halpin RA, Guan Y, Twaddle A, et al. 2016. Quantifying influenza virus diversity and transmission in humans. *Nat Genet.* 48(2): 195–200.
- Pybus OG, Rambaut A, Belshaw R, Freckleton RP, Drummond AJ, Holmes EC. 2007. Phylogenetic evidence for deleterious mutation load in RNA viruses and its contribution to viral evolution. *Mol Biol Evol.* 24(3): 845–852.
- Ray SC, Fanning L, Wang X-H, Netski DM, Kenny-Walsh E, Thomas DL. 2005. Divergent and convergent evolution after a common-source outbreak of hepatitis C virus. *J Exp Med.* 201(11): 1753–1759.
- Rehermann B. 2009. Hepatitis C virus versus innate and adaptive immune responses: a tale of coevolution and coexistence. *J Clin Invest.* 119(7): 1745–1754.
- Rhee S-Y, Liu T, Ravela J, Gonzales MJ, Shafer RW. 2004. Distribution of human immunodeficiency virus type 1 protease and reverse transcriptase mutation patterns in 4, 183 persons undergoing genotypic resistance testing. *Antimicrob Agents Chemother.* 48(8): 3122–3126.
- Rife BD, Mavian C, Chen X, Ciccozzi M, Salemi M, Min J, Prosperi MC. 2017. Phylodynamic applications in 21st century global infectious disease research. *Glob Health Res Policy* 2(1): 13.
- Salemi M, Rife B. 2016. Phylogenetics and phyloanatomy of HIV/SIV intra-host compartments and reservoirs: the key role of the central nervous system. *Curr HIV Res.* 14(2): 110–120.
- Salemi M, Strimmer K, Hall WW, Duffy M, Delaporte E, Mboup S, Peeters M, Vandamme AM. 2001. Dating the common ancestor of SIVcpz and HIV-1 group M and the origin of HIV-1 subtypes using a new method to uncover clock-like molecular evolution. *FASEB J.* 15(2): 276–278.
- Sawyer SL, Emerman M, Malik HS. 2004. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol.* 2(9): E275.
- Schlub TE, Smyth RP, Grimm AJ, Mak J, Davenport MP. 2010. Accurately measuring recombination between closely related HIV-1 genomes. *PLoS Comput Biol.* 6(4): e1000766.
- Sheshberadaran H, Chen S-N, Norrby E. 1983. Monoclonal antibodies against five structural components of measles virus i. characterization of antigenic determinants on nine strains of measles virus. *Virology* 128(2): 341–353.
- Shi M, Lin X-D, Chen X, Tian J-H, Chen L-J, Li K, Wang W, Eden J-S, Shen J-J, Liu L, et al. 2018. The evolutionary history of vertebrate RNA viruses. *Nature* 556(7700): 197–202.
- Smith DJ, Forrest S, Ackley DH, Perelson AS. 1999. Variable efficacy of repeated annual influenza vaccination. *Proc Natl Acad Sci U S A.* 96(24): 14001–14006.
- Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA. 2004. Mapping the antigenic and genetic evolution of influenza virus. *Science* 305(5682): 371–376.
- Taubenberger JK, Kash JC. 2010. Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host Microbe* 7(6): 440–451.
- Thyagarajan B, Bloom JD. 2014. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *Elife* 3. doi: 10.7554/eLife.03300.
- Tsibris AMN, Korber B, Arnaout R, Russ C, Lo C-C, Leitner T, Gaschen B, Theiler J, Paredes R, Su Z, et al. 2009. Quantitative deep sequencing reveals dynamic HIV-1 escape and large population shifts during CCR5 antagonist therapy in vivo. *PLoS One* 4(5): e5683.
- Van Laethem K, Theys K, Vandamme A-M. 2015. HIV-1 genotypic drug resistance testing: digging deep, reaching wide?. *Curr Opin Virol.* 14:16–23.
- Vijaykrishna D, Mukerji R, Smith GJD. 2015. RNA virus reassortment: an evolutionary mechanism for host jumps and immune evasion. *PLoS Pathog.* 11(7): e1004902.
- Ward CD, Stokes MA, Flanagan JB. 1988. Direct measurement of the poliovirus RNA polymerase error frequency in vitro. *J Virol.* 62(2): 558–562.
- Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Jr, Swanstrom R, Burch CL, Weeks KM. 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460(7256): 711.
- Wertheim JO, Kosakovsky Pond SL. 2011. Purifying selection can obscure the ancient age of viral lineages. *Mol Biol Evol.* 28(12): 3355–3365.
- Woelk CH, Holmes EC. 2002. Reduced positive selection in vector-borne RNA viruses. *Mol Biol Evol.* 19(12): 2333–2336.
- Worobey M, Han G-Z, Rambaut A. 2014. A synchronized global sweep of the internal genes of modern avian influenza virus. *Nature* 508(7495): 254–257.
- Zanini F, Neher RA. 2013. Quantifying selection against synonymous mutations in HIV-1 env evolution. *J Virol.* 87(21): 11843–11850.
- Zanini F, Brodin J, Thebo L, Lanz C, Bratt G, Albert J, Neher RA. 2015. Population genomics of inpatient HIV-1 evolution. *Elife* 4:e11282.
- Zuckerkindl E, Pauling L. 1965. Evolutionary divergence and convergence in proteins. In: *Evolving genes and proteins*. Elsevier. p. 97–166.