Developing an Appropriate Evolutionary Baseline Model for the Study of Human Cytomegalovirus

Abigail A. Howell[1], John W. TerbotII[1,2], Vivak Soni (iD) [1], Parul Johri (iD) [1], Jeffrey D. Jensen[1,*], and Susanne P. Pfeifer (iD) [1,*]

[1]School of Life Sciences, Center for Evolution and Medicine, Arizona State University, Tempe
[2]Division of Biological Sciences, University of Montana, Missoula

*Corresponding authors: E-mails: jeffrey.d.jensen@asu.edu; susanne.pfeifer@asu.edu.

Abstract

Human cytomegalovirus (HCMV) represents a major threat to human health, contributing to both birth defects in neonates as well as organ transplant failure and opportunistic infections in immunocompromised individuals. HCMV exhibits considerable interhost and intrahost diversity, which likely influences the pathogenicity of the virus. Therefore, understanding the relative contributions of various evolutionary forces in shaping patterns of variation is of critical importance both mechanistically and clinically. Herein, we present the individual components of an evolutionary baseline model for HCMV, with a particular focus on congenital infections for the sake of illustration—including mutation and recombination rates, the distribution of fitness effects, infection dynamics, and compartmentalization—and describe the current state of knowledge of each. By building this baseline model, researchers will be able to better describe the range of possible evolutionary scenarios contributing to observed variation as well as improve power and reduce false-positive rates when scanning for adaptive mutations in the HCMV genome.

Key words: human cytomegalovirus, viral population genomics, evolutionary baseline model.

Significance

Human cytomegalovirus (HCMV) infection is a major cause of birth defects and can lead to severe effects in immuno-suppressed and immunonaïve individuals. Pathogenicity is likely driven by multiple factors, including the genetic diversity of the virus itself. Furthermore, the accurate identification of genomic loci underlying viral adaptation relies on an appropriate baseline model that accounts for constantly operating evolutionary processes shaping this genetic diversity. With this overview of the current understanding of these processes in HCMV, we provide the necessary details for researchers to implement such a baseline model for their own genomic analysis of patient samples.

Introduction

As the leading cause of infection-related birth defects—including cognitive and hearing impairments—human cytomegalovirus (HCMV) remains a major threat to global health, with a seroprevalence of more than 90% outside of the developed world (e.g., Boppana et al. 2013; Swanson and Schleiss 2013; Dreher et al. 2014). HCMV is also a primary cause of solid organ transplant failure (Balfour 1979) and often results in opportunistic infections in immunocompromised individuals or those with immature immune systems (e.g., Suárez et al. 2019, 2020). Additionally, primary infection or reactivation is implicated in a wide variety of health complications (Griffiths et al. 2015), and recent studies suggest that HCMV may play an active role in glioma pathogenesis in individuals with glioblastoma (Cobbs et al. 2002; Abdelaziz et al. 2019).

Moreover, along with human immunodeficiency virus type 1 (HIV-1), HCMV is the most common viral agent transmitted from mother to offspring and may itself contribute to the vertical transmission of HIV-1 (Johnson et al. 2015; Girsch et al. 2022).

HCMV is a β-herpesvirus in the Herpesviridae family with a relatively large double-stranded (ds) DNA genome of ~235 kb in size, including between 164 and 167 open reading frames (ORFs) (Dolan et al. 2004). Lytic infection is initiated by the expression of genes in a flow cascade, and DNA replication initiates 1–3 days postinfection (Weekes et al. 2014). The genome contains two unique regions—the unique long (U_L) and unique short (U_S) region—that are internally and externally flanked by repeats. The U_L region contains ORFs encoding gene products associated with latency and reactivation (Revello and Gerna 2010; Li et al. 2014); in laboratory passaged strains, cultures have been shown to accumulate large deletions in this region compared with clinically isolated viruses, likely owing to the relaxed selection in laboratory environments (Cha et al. 1996). In contrast, ORFs within the U_L region that encode envelope glycoproteins thought to be important for pathogenesis have been found to evolve under considerable constraint (He et al. 2006; Ji et al. 2006; Heo et al. 2008).

Multiple studies have suggested a link between pathogenesis and genomic variability (Meyer-König, Vogelberg, et al. 1998; Renzette et al. 2014; Wang et al. 2021), with high levels of diversity and multiple-strain infection found to be associated with higher viral loads (Pang et al. 2008; Sowmya and Madhavan 2009; Puchhammer-Stöckl and Görzer 2011). Furthermore, variation in the glycoproteins gO and gB, potentially generated through recombination (Meyer-König, Vogelberg, et al. 1998), has been proposed to influence cell tropism and dissemination (Hahn et al. 2004). Gaining a better understanding of the evolutionary forces that shape viral diversity is thus of critical importance both mechanistically and clinically. During the last decade, many efforts have been made to understand the relative contributions of admixture, positive and purifying selection, and infection-related bottlenecks in shaping HCMV inter-host and intrahost variation (Renzette et al. 2013, 2015, 2017; Pokalyuk et al. 2017). Relatedly, numerous efforts have focused on elucidating key evolutionary parameters including the underlying mutation and recombination rates, as well as the selective effects of newly arising mutations (the distribution of fitness effects [DFE]; Renzette et al. 2015, 2017; Morales-Arce et al. 2022).

Importantly, recent studies focused upon evolutionary inference procedures have simultaneously demonstrated the value of jointly estimating parameters of natural selection with population history, as a neglect of one to infer the other will often result in serious misinference (Johri et al. 2020, 2021). Moreover, only by first accounting for the constantly acting evolutionary processes of genetic drift (as shaped by the infection bottleneck and subsequent viral population growth, as well as the genetic structure associated with compartmentalization) and purifying and background selection (owing to the pervasive input of deleterious mutations) may one develop a meaningful baseline model of expected levels and patterns of genomic variation. This baseline model is critical for accurately detecting and quantifying rarer and episodic evolutionary processes, such as positive selection potentially leading to viral adaptation (Johri, Aquadro, et al. 2022; Johri, Eyre-Walker, et al. 2022). More specifically, owing to overlapping patterns between neutral and selective evolutionary processes (Jensen 2009; Bank et al. 2014), this baseline model is essential for defining rates of true positives and false positives associated with the detection of rare or episodic effects in any given population and for any given data set.

As such an evolutionary baseline model has yet to be fully described for HCMV, we here outline important components of such a model and review the current state of knowledge pertaining to each: mutation rates, recombination rates, the distribution of fitness effects, infection dynamics, and compartmentalization. We close with a series of recommendations for improving evolutionary inference in this important human pathogen and highlight key areas in need of further investigation.

Mutation Rate

The mutation rate quantifies the frequency at which spontaneous (*de novo*) mutations arise in a genome, as caused by a variety of factors including DNA replication errors and spontaneous DNA damage (see review of Pfeifer 2020). This rate is distinct from the substitution rate— that is, the rate at which mutations become fixed in a population—which is influenced not only by the *de novo* mutation rate but also by natural selection, genetic drift, as well as multiple other factors. However, for strictly neutral mutations, the rate of mutational input is equal to the rate of substitution (Kimura 1968), leading to a clock-like accumulation of mutations over time. Using a molecular clock (divergence)-based approach, recent studies have reported substitution rates of approximately 3.0×10^{-9} substitutions per nucleotide per year in HCMV (McGeoch et al. 2000)— one to two orders of magnitudes lower than the rate reported for a closely related virus, herpes simplex virus (HSV-1), which exhibits 3.0×10^{-8} (Sakaoka et al. 1994) and 1.4×10^{-7} (Kolb et al. 2013) substitutions per nucleotide per year. Mutation rates of both HCMV and HSV-1 have also been studied *in vitro*. For example, by scoring null mutations in the *tk* gene using ganciclovir, mutation rates in HSV-1 have similarly been estimated to range from 5.9×10^{-8} (Hwang et al. 2002; Drake and Hwang 2005) to 1.0×10^{-7} (Hall and Almy 1982) substitutions

per nucleotide per cell infection, where cell infection is an estimate of a viral generation.

It is necessary here to highlight the various units being reported when comparing between the results described in different studies, with rates reported as substitutions per nucleotide per generation (s/n/g), substitutions per nucleotide per year (s/n/y), substitutions per nucleotide per cell infection (s/n/c), or substitutions per nucleotide per round of copying (s/n/r), if the mode of replication is known. The mode of replication of dsDNA viruses is likely limited to semiconservative replication, although RNA viruses by comparison are known to use a "stamping machine" model, where a single template is used for all progeny strands (Luria 1951). To compare between estimates using substitutions per nucleotide per cell infection and estimates using substitutions per nucleotide per year, we have used the number of viral cycles per year as a conversion factor (table 1). Specifically, conversion factors of 181.87 to 362.48 viral cycles per year were chosen to span lower and upper estimates for HCMV, while 1,946.67 viral cycles per year were used for closely related HSV-1 for comparison. These estimates are based on internalization times of 10 min (Bodaghi et al. 1999; Hetzenecker et al. 2016) and 30 min (Zheng et al. 2014), as well as eclipse times of 24–48 h (Jean et al. 1978) and 4 h (Nishide et al. 2019), for HCMV and HSV-1, respectively. Importantly, these conversions highlight the discrepancy between divergence and *in vitro* estimates of the substitution rate, demonstrating that molecular clock-based estimates primarily provide information about the rate of neutral and nearly neutral mutation, rather than estimating full mutational spectra

(as discussed in the below section). Additionally, the further analysis of future patient samples would be of great value in better characterizing the interhost variance in these rates.

Notably, these experimental and empirical measurements of the mutation rate based on genome-wide population genetic data neglect the substantial proportion of lethal and deleterious mutations that are removed from the population via purifying selection. Owing to this neglect, measurements obtained using these methods are likely an underestimate of the genuine genome-wide mutation rate (Peck and Lauring 2018). Mutation accumulation experiments provide a valuable (and less biased) alternative by subjecting a viral population to a series of bottlenecks that reduces the effective population size, thus minimizing the efficacy of selection. A similar strategy can be applied to natural, longitudinal population data. Using this approach, the mutation rate of HCMV was estimated by Renzette et al. (2015) as 2.0×10^{-7} mutations per nucleotide per generation using longitudinal samples obtained from 18 patients, where mutations were called if absent in earlier samples and present in all later samples. Importantly, however, evaluating such longitudinal data in the context of a mutation accumulation study comes with the qualification that selective pressures are expected to be much stronger in patient samples relative to traditional experimental mutation accumulation lines. In addition, the presence of a reinfection event during the longitudinal sampling—if not identified—would be expected to upwardly bias these estimates. It is also important to note that rate estimates of this sort are further complicated by practical limitations of clinical sampling. Specifically, previous studies have shown that deep sequencing through the use of polymerase chain reaction amplicons requires rare variants to be present at >1% frequency in order to be reliably detected (Fonager et al. 2015; Kyeyune et al. 2016)—though newer methods that utilize target enrichment protocols may improve upon this threshold (Hage et al. 2017). Given that the vast majority of variants are expected to be rare, such detection thresholds may be of considerable significance.

Mutation rates in viruses may evolve through both mutator and antimutator alleles, the fixations of which are thought to be governed by genome size and effective population size (Lynch et al. 2016). When effective population sizes are small, selection is weak and may be unable to prevent mutator alleles from fixing. To date, one hypermutator has been identified in HCMV (Chou et al. 2016). Mutator alleles are a double-edged sword for viruses, having important implications for the rate of adaption (Taddei et al. 1997; Travis and Travis 2002), but more significantly also create the possibility of mutational meltdown (Crotty et al. 2001; Beaucourt et al. 2011; Bank et al. 2016; Matuszewski et al. 2017; Ormond et al. 2017). Indeed, owing to interference between the greater input of deleterious mutations with the minor input of beneficial mutations, higher mutation

Table 1

In Vitro- and Divergence-Based Estimates of *De Novo* Mutation Rates in HCMV Compared with the Closely Related HSV-1

Virus	Approach	Original Unit[a]	Estimated Rate/Cycle	Reference
HCMV	*In vitro*	s/n/c	2.0×10^{-7}	Renzette et al. 2015
HCMV	Divergence	s/n/y	1.6×10^{-11} / 8.2×10^{-12}	McGeoch et al. 2000
HSV-1	Divergence	s/n/y	7.1×10^{-11}	Kolb et al. 2013
HSV-1	Divergence	s/n/y	4.1×10^{-11}	Sakaoka et al. 1994
HSV-1	*In vitro*	s/n/c	1.0×10^{-7}	Hall and Almy 1982
HSV-1	*In vitro*	s/n/c	5.9×10^{-8}	Hwang et al. 2002; Drake and Hwang 2005

NOTE.—To compare between estimates using substitutions per nucleotide per cell infection (s/n/c) and estimates using substitutions per nucleotide per year (s/n/y), we have used conversion factors of either 181.87 or 362.48 viral cycles per year to span uncertainty in HCMV, and 1,946.67 viral cycles per year for HSV-1.

[a]s = substitutions; n = nucleotide; c = cell infection; y = year.

rates may slow or stop the rate of adaptation (Pénisson et al. 2017; Jensen and Lynch 2020; Jensen et al. 2020). Other molecular determinants of viral mutation rates include postreplicative repair through interaction with DNA damage response pathways (Weitzman et al. 2010; Luftig 2014)—a particularly relevant mechanism for HCMV as herpesviruses are known to induce DNA damage responses (Xiaofei and Kowalik 2014).

As HCMV has been observed to be quite diverse compared with other DNA viruses—on the order of certain RNA viruses (Wang et al. 2002; Jerzak et al. 2005)—one formal possible explanation for the high levels of nucleotide diversity observed in HCMV is an exceptionally high mutation rate (i.e., as levels of neutral variation are expected to be a factor of the effective population size as well as the underlying mutation rate). This hypothesis was recognized as unlikely by Renzette et al. (2011), owing, among other reasons, to the proofreading activity of HCMV's DNA polymerase (Nishiyama et al. 1983). Although Cudini et al. (2019) recently rediscussed this possibility (and see the response of Jensen and Kowalik 2020), there appears to be general agreement that RNA virus-like levels of variation in HCMV are not due to RNA virus-like mutation rates. Specifically, following multiple studies on HCMV interhost and intrahost variation (Renzette et al. 2013, 2015, 2017; Pokalyuk et al. 2017; and see the below sections), it has been demonstrated that observed diversity is likely generated by a combination of mutation, recombination, reinfection, compartmentalization, selection, and infection population size histories (Jensen 2021)—with a mutation rate of 2.0×10^{-7} mutations per nucleotide per generation appearing consistent with the data (Renzette et al. 2015). More specifically, the observed high levels of variation appear to more likely be related to the population dynamics related to compartmentalization, gene flow, and reinfection, rather than to particularly elevated rates of mutation (e.g., Pokalyuk et al. 2017; Jensen and Kowalik 2020). Renzette et al. additionally identified a weak but highly significant positive correlation between estimated mutation rates and single nucleotide polymorphism (SNP) density across the HCMV genome, as may be expected. Heterogeneity in mutation rates across the genome was additionally proposed as a contributing factor underlying the observed correlations between intraspecies variation and recombination rates, as well as of that between variation and divergence (Renzette et al. 2016).

Recombination Rate

Recombination not only contributes genetic variation through the generation of novel genotypic combinations, but it may also improve the efficacy of selection through the reduction of interference effects between and among beneficial and deleterious variants (Hill and Robertson 1966; Felsenstein 1974; Lynch et al. 1995; Pénisson et al. 2017). Studies examining the intergenic variability of HCMV glycoprotein loci (Meyer-König, Haberland, et al. 1998; Haberland et al. 1999; Yan et al. 2008) provided the initial evidence for homologous recombination in the HCMV genome. Nearly two decades later, Renzette et al. (2015) estimated a genome-wide recombination map using a population genetic approach, reporting a mean recombination rate of ~0.23 crossover events per genome per generation, based on observed patterns of linkage disequilibrium (LD) (i.e., by assessing the extent to which observed haplotype distributions may be explained by variable rates of recombination; and see the review of Stumpf and McVean (2003) for a discussion on estimating recombination rates from population genetic data). The authors further reported a correlation between recombination rate and SNP density, consistent with widespread purifying selection, as has been observed in multiple diverse species (e.g., Begun and Aquadro 1992; Pfeifer and Jensen 2016; Renzette et al. 2017; and see the review of Charlesworth and Jensen 2021). However, as with mutation rates, recombination rate estimates can also be misinferred, for example, due to unaccounted for progeny skew, which is known to increase levels of LD in highly skewed populations relative to standard Wright–Fisher expectations (and as such may downwardly bias recombination rate estimation if unaccounted for; Eldon and Wakeley 2008; Birkner et al. 2013). This observation highlights the need for further computational method development of mutation and recombination rate estimators for the type of generalized progeny skew distributions applicable to viruses and other human pathogens (Morales-Arce et al. 2020; Sabin et al. 2022).

In addition to LD–based approaches, studies have also characterized recombination in the HCMV genome using a combination of phylogenetic and population-level analyses. By constructing "phylogenetic trees" for each gene in the HCMV genome and correcting for recombination breakpoints with the genetic algorithm GARD, Kosakovsky Pond et al. (2006) found that the majority of loci showed no consistent phylogenetic patterns, indicating that recombination occurs often enough that whole genomes can behave as "gene-scale mosaics." In other words, what certain authors refer to as variable phylogenetic trees are in fact better described as variable coalescent histories. Further, like the Renzette et al. studies, Sijmons et al. (2015) also observed a correlation between recombination rate and nucleotide diversity using a phylogenetic approach. However, phylogenetic-based approaches are generally poorly suited for the study of recombination compared with the coalescent-based approaches utilized in population genetics—and multiple studies suffer from these limitations when trying to distinguish between recombination and competing evolutionary processes in a

phylogenetic framework (e.g., Houldcroft et al. 2016; Cudini et al. 2019). Specifically, coalescent theory provides a sophisticated framework for the study of variable gene genealogies owing to recombination (Wakeley 2009) and avoids the pretense of searching for a single (and non-existent) "phylogenetic tree' to describe within-population variation (e.g., Cudini et al. 2019; and see Rosenberg and Nordborg 2002 for a discussion).

The Distribution of Fitness Effects (DFE)

HCMV is characterized by a large genome relative to other human viruses. Although the set of protein-coding genes in HCMV experiences constant revision, there are 45 core genes that are conserved across all herpesviruses and ~117 noncore genes that are more specific to the CMVs, many of which are still being functionally characterized (Van Damme and Van Loock 2014; Mozzi et al. 2020). Although it is clear that protein-coding regions occupy the majority of the HCMV genome, these uncertainties mean that the precise fraction of the genome that experiences direct purifying selection is not yet fully defined—though roughly 25% of the genome has been observed to be nearly devoid of variation, potentially suggesting strong constraint (Renzette et al. 2015). Interestingly, within-patient nucleotide diversity in noncoding regions of the genome has generally been observed to be on the same order as less-constrained coding regions (Renzette et al. 2011), suggesting the presence of functionally important regions interspersed across the genome and/or widespread background selection effects (Renzette et al. 2016). This combination of factors renders the identification of neutrally evolving sites challenging.

Previous studies have used comparisons of sequence evolution at nonsynonymous versus synonymous sites at various evolutionary scales to quantify selective forces acting on protein-coding regions in the HCMV genome. A comparative genomic analysis across multiple CMV species found pervasive purifying selection in most protein-coding regions (as indicated by low levels of d_N/d_S; Mozzi et al. 2020), as would be expected. Similarly, comparisons of sequence polymorphism within hosts to the divergence among hosts (i.e., using the McDonald and Kreitman 1991 test) also indicated the action of widespread purifying selection (Renzette et al. 2011). In contrast, evidence for positive selection was limited to specific regions, including the glycoproteins (Renzette et al. 2013). Thus, although glycoproteins and their linked regions will likely be additionally impacted by recurrent selective sweeps, the majority of the genome is expected to be largely affected by the direct and linked effects of purifying selection.

As selection against harmful mutations at functionally important sites in the genome can affect patterns of variation at linked neutral alleles (i.e., background selection; Charlesworth et al. 1993) and as this effect has been suggested to be a primary determinant of genomic variation in HCMV (Renzette et al. 2016), it is important to characterize the DFE of newly arising mutations across the genome. A recent study by Morales-Arce et al. (2022) used an approximate Bayesian computation (ABC) framework to infer the DFE of deleterious mutations from a within-patient sample of HCMV. This study accounted for the specific demographic history of the within-patient population as associated with viral infection dynamics (as previously inferred by Renzette et al. 2013), non-Wright–Fisher replication dynamics, as well as background selection. They inferred that roughly 50% of all new mutations were effectively neutral ($-1 < 2N_e s \leq 0$), 24% were mildly deleterious ($-10 < 2N_e s \leq -1$), 12% were moderately deleterious ($-100 < 2N_e s \leq -10$), and 13% were strongly deleterious ($2N_e s \leq -100$), where N_e refers to the effective population size and s to the selection coefficient against the homozygote (fig. 1A). As these estimates were obtained for all sites comprising the functional region (i.e., the inference was not restricted to nonsynonymous sites) and ~30% of all sites in coding regions are likely to have little or no fitness costs upon mutation (e.g., synonymous changes), the DFE at functionally important sites in HCMV is probably closer to 30% effectively neutral, 34% weakly deleterious, 17% moderately deleterious, and 19% lethal mutations (fig. 1B). Importantly, although such a correction naturally depends on the fraction of synonymous sites that are behaving neutrally, these estimates are in fact quite consistent with multiple previous random mutagenesis studies that measured the proportion of lethal mutations in DNA viruses to be ~20% (e.g., Sanjuán 2010). While Morales-Arce et al. (2022) accounted for a number of factors that add complexity to within-patient populations of HCMV (including an extremely strong bottleneck corresponding to the infection), they simulated only a single population of HCMV. As there is strong evidence of HCMV populations being structured within patients (Pokalyuk et al. 2017; Sackman et al. 2018; and see the section on Compartmentalization below), current estimates of the deleterious DFE might still be biased, and future inference incorporating both compartmentalization and reinfection will be important in this regard.

Infection Dynamics

The demographic history of a population is an important determinant of both genetic variation and potential selective outcomes and therefore an appropriate starting point for evolutionary analysis, particularly in light of the high levels of HCMV diversity observed within patients (Drew et al. 1984; Spector et al. 1984; Haberland et al. 1999; Faure-Della Corte 2010; Renzette et al. 2011, 2013, 2015, 2016, 2017; Hage et al. 2017; Pokalyuk et al. 2017). The expected intrahost population dynamics involve
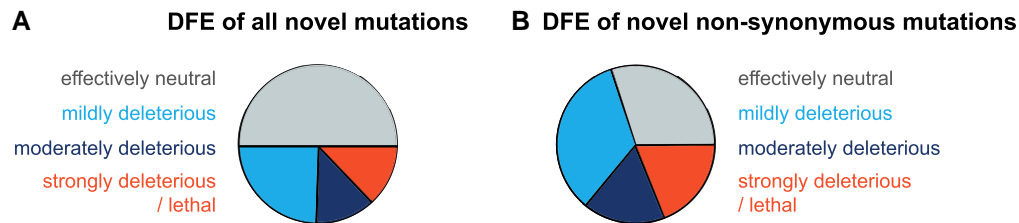
Fig. 1.—Distribution of fitness effects (DFE) of all new and new nonsynonymous mutations. (A) Using an approximate Bayesian framework to account for the specific demographic history of their within-patient population, Morales-Arce et al. (2022) inferred the DFE of all new mutations in human cytomegalovirus as roughly 50% effectively neutral ($-1 < 2N_e s \leq 0$; gray), 24% mildly deleterious ($-10 < 2N_e s \leq -1$; light blue), 12% moderately deleterious ($-100 < 2N_e s \leq -10$; dark blue), and 13% strongly deleterious/lethal ($2N_e s \leq -100$; red), where N_e refers to the effective population size and s to the selection coefficient against the homozygote. (B) Assuming that ~30% of all sites in coding regions likely have little or no fitness costs upon mutation, the DFE at functionally important sites corresponds to roughly 30% effectively neutral, 34% mildly deleterious, 17% moderately deleterious, and 19% strongly deleterious/lethal mutations.

a strong population bottleneck (a temporary reduction in population size) at the point of infection, followed by rapid population expansion (see review of Jensen 2021). The level of intrahost genetic variation that is present at the point of infection will in part be determined by the severity of the bottleneck. If the transmission bottleneck is wide, then there may be numerous virions founding the initial infection, resulting in greater genetic variation and an increased probability that beneficial variants may be transferred from the founding population. Conversely, a narrow bottleneck can result in a severe loss of genetic variation, with low-frequency variants being eliminated regardless of their fitness effects. This process is known as a founder effect (see Zwart and Elena 2015, for a discussion of this effect in viral populations).

In the case of congenital infections, demographic modeling approaches have shown support for a population bottleneck associated with the initial transplacental infection (transmission of virions from the maternal compartment to the fetal plasma compartment), followed by additional bottlenecks associated with compartmental infections (fig. 2; and Renzette et al. 2013; for a detailed discussion regarding the population structure dynamics between compartments, see the section below). Importantly, the initial bottleneck was shown to involve potentially hundreds of unique HCMV genomes, which helps to explain the relatively high levels of genetic diversity observed at the point of infection, as compared with certain RNA viruses in which a single (or very few) virions are thought to be involved in infection (Keele et al. 2008; Fischer et al. 2010; Renzette et al. 2013, 2014). Furthermore, Renzette et al. (2013) found support for gene flow between urine and plasma compartments (the two compartments sampled in that study). Their results further suggested that plasma may serve as a "route" for gene flow within the host, with preliminary evidence indicating that it carries compartment-specific variants from other compartments; this process may thus also be an important determinant of within-host variation.

Further evidence for admixture between compartments (this time including plasma, urine, and saliva compartments) was found by Pokalyuk et al. (2017), suggesting that reinfection postbirth is possible via, for instance, breast milk (Numazaki 1997; Enders et al. 2011; also see the review of Bardanzellu et al. 2019). In other words, maternal compartment-specific variants appeared to be transmitted to the infant postbirth. Although the above examples are focused upon congenital infections, related work has similarly highlighted the importance of multistrain infections in immunosuppressed adults and particularly the relationship between this infection status and the emergence of antiviral resistance mutations (e.g., in transplant recipients; Suárez et al. 2019, 2020).

To date no method exists to prevent maternal–fetal transmission or to reduce the severity of fetal infection (Britt 2017). Therefore, the characterization of population dynamics is likely to be integral to future therapeutic strategies. For example, clinically imposing a more severe population bottleneck during pregnancy may reduce genetic variation in the HCMV infecting population, limiting the pool of variation on which natural selection may subsequently act, thereby potentially improving treatment outcomes. Finally, it has been shown that host immune suppression can reactivate dormant viruses, restarting production of viral progeny; this switch from latent to productive life cycles can induce temporary or sustained CMV replication (Porter et al. 1985; Dupont and Reeves 2016).

Demographic inference in HCMV is inherently challenging due to the genome-wide impact of selection (see the DFE section above), which will in turn bias common demographic estimators which are based on neutrality (see the discussion of Ewing and Jensen 2016; Pouyet et al. 2018). Namely, neutral demographic estimators require sufficiently large nonfunctional regions and high rates of recombination, such that assumptions of strict neutrality hold (Gutenkunst et al. 2009; Excoffier et al. 2013; Kelleher et al. 2019; Steinrücken et al. 2019). These criteria ensure that variants can be chosen that are not experiencing
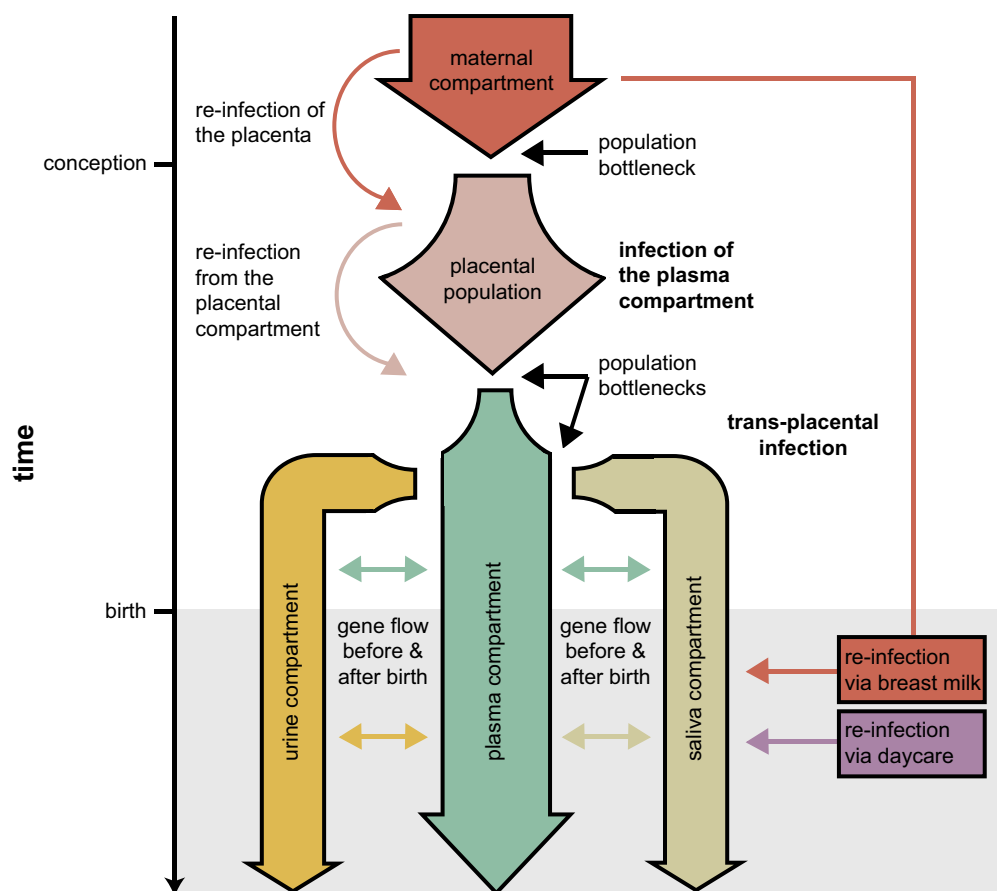
FIG. 2.—Demographic dynamics of congenital human cytomegalovirus (HCMV) infection. Demographic scenarios of infection and reinfection in HCMV likely contributing to the high levels of observed interhost and intrahost diversity, including a population bottleneck associated with the initial transplacental infection (transmission of virions from the maternal compartment [red]/plasma [pink] to the fetal plasma compartment [green]), followed by additional bottlenecks associated with compartmental infections (urine [yellow] and saliva [olive]), as well as gene flow between compartments and reinfection of compartments during pregnancy and after birth (e.g., via breast milk [red] and/or daycare [purple]).

background selection. For example, Renzette et al. (2013) utilized ∂a∂i, a neutral demographic inference approach based on the site frequency spectrum (Gutenkunst et al. 2009), to build and parameterize HCMV infection models (and see Sackman et al. 2018; Jensen and Kowalik 2020).

This inference problem of estimating demography in the presence of selection is indeed somewhat circular, as the estimation of selection will also be biased by unaccounted for demographic dynamics (Rousselle et al. 2018; Johri et al. 2020). This fact highlights the importance of performing joint, simultaneous inference of selection with demography, rather than taking the more common stepwise approach of first estimating one and then the other (see review of Johri, Eyre-Walker, et al. 2022). Recently proposed ABC approaches that jointly estimate population history and the DFE of deleterious mutations perform such joint inference and importantly do not require the a priori identification of neutrally evolving sites (Johri et al. 2020). Explicitly accounting for viral infection dynamics, Morales-Arce et al. (2020) incorporated progeny skew into the joint ABC

inference scheme of Johri et al. (2020)—an important extension to this framework as the assumption of small progeny distributions utilized by a majority of population genetic inference approaches is likely violated in many pathogens, as noted above (see reviews of Tellier and Lemaire 2014; Irwin et al. 2016). The authors demonstrated that their tailoring of this ABC inference approach specifically to viral populations avoided misinference resulting from a neglect of this consideration. Other recent inference approaches have also relaxed the assumption of small progeny skew, demonstrating an ability to coestimate parameters related to the biology of progeny skew together with those of demographic and selective histories (e.g., Matuszewski et al. 2018; Sackman et al. 2019).

Compartmentalization

The final consideration of note impacting intrahost population dynamics of viral infections is population structure between different areas of infection, commonly referred to as

compartmentalization (Zárate et al. 2007). Compartmentalization may be relevant for any virus not localized to a single organ or cell type (Di Liberto et al. 2006; Zárate et al. 2007; Renzette et al. 2014; Sackman et al. 2018)—including HCMV, known to infect several cells and organs throughout the body.

As a long-studied virus, HCMV has been well documented to infect a wide variety of cells including the epithelial cells of gland and mucosal tissue, smooth muscle cells, fibroblasts, macrophages, dendritic cells, hepatocytes, and vascular endothelial cells (Sinzger et al. 2008; Jean Beltran and Cristea 2014). Unsurprisingly given this broad cellular tropism, evidence of infection in specific organs is similarly extensive and includes the brain and peripheral nerves, the eyes, the placenta, the lungs, the gastrointestinal tract from the esophagus to the colon, the liver, the lymph nodes, the heart, the peripheral blood, and the kidneys (Plachter et al. 1996). Of these areas, viral shedding from salivary glands, the ductal epithelium of mammary glands and the kidney, and the syncytiotrophoblasts (placenta) is thought to be critical to interhost transmission (Mocarski 2004; Kinzler and Compton 2005). However, because of potential gene flow between compartments within a host, other sites of infection are nonetheless important for understanding the intrahost dynamics of this virus.

Another necessary consideration is the location of regions that can harbor the latent stage—these areas are likely important for the maintenance of genetic diversity that may otherwise be lost in actively replicating lineages (Chou 1989; Frange et al. 2013). While infections can occur across the body, the latent, and importantly nonreproducing, stage of the virus seems to be limited in cell tropism. Specifically, HCMV has been found to use endothelial and select myeloid lineages as well as monocytes, macrophages, and their progenitors (i.e., cells found in the circulating plasma population) as latency sites (Jarvis and Nelson 2002; Yatim and Albert 2011).

Given the wide range of potential sites of infection, it is crucial to resolve observed levels of intrahost population structuring that are indicative of compartmentalization. Several studies have observed considerable genomic diversity (Renzette et al. 2011, 2013; Mayer et al. 2017; Pokalyuk et al. 2017; Cudini et al. 2019; Pang et al. 2020), while others have found intrahost populations to be comparatively invariant (Hage et al. 2017). The comparison of patients with single-versus multiple-infection histories is likely one important source of disparity in these observed levels of variation (Mayer et al. 2017; Pokalyuk et al. 2017; Sackman et al. 2018; Cudini et al. 2019; Jensen and Kowalik 2020; Houldcroft et al. 2020; Pang et al. 2020). It should also be noted that the importance of multiple infections in shaping intrahost diversity of infants may still rely on compartmentalization within the maternal infection (e.g., with primary infections arising from the cervical population and secondary infections being associated with the mammary gland population; Sackman et al. 2018; Pang et al. 2020).

Compartmentalization has also been implicated as a clinically important factor in the development of a multi-drug resistant lineage within the chronic infections of immunocompromised patients (Frange et al. 2013; Renzette et al. 2014; Suárez et al. 2019, 2020). Furthermore, multiple population genetic studies using longitudinally sampled patient data concluded that compartmentalization is an important factor in explaining intrahost diversity of fetal and infant infections (Renzette et al. 2013, 2015). Models developed from these studies focused on three subpopulations corresponding to source sites of samples: salivary glands/saliva, blood/plasma, and kidney/urine (Renzette et al. 2014, 2015; Pokalyuk et al. 2017; Sackman et al. 2018). Generally, these models attribute plasma as the circulating population that serves as an intermediary for spread between the distal compartments of salivary glands and kidney (fig. 2). Of particular note, levels of genetic divergence between compartments of a single patient were found to be as great as those observed between the same compartment sampled from unrelated patients (Renzette et al. 2013), suggesting limited between-compartment gene flow within a single host. Yet, the extent to which these considerable levels of differentiation are attributable to localized, compartment-specific adaptation, or simply the constant operation of neutral evolutionary processes, remains unresolved—and this continues to stand as one of the most pressing and interesting evolutionary questions in the HCMV system.

Closing Thoughts

When developing an evolutionary baseline model of HCMV, special consideration should be given to the demographic processes that shape genetic diversity and the sampling methods that generate clinical data sets, including the ability to detect low-frequency variants, as well as the level of progeny skew, bottleneck severity during infection and reinfection, and the degree of compartmental admixture. Correctly modeling these processes and accounting for various ascertainment biases will allow researchers to better describe the relative contributions of each evolutionary force in shaping observed levels and patterns of variation, as well as quantify uncertainty in model choice and in the identification of adaptive loci. In addition, gaining a better understanding of when and how HCMV diversity is generated has important implications for vaccine development as well as antiviral therapy, both for determining the timing of drug delivery and for combating resistance evolution.

Acknowledgments

Literature Cited

Abdelaziz MO, et al. 2019. Development of a human cytomegalovirus (HCMV)-based therapeutic cancer vaccine uncovers a previously unsuspected viral block of MHC class I antigen presentation. Front Immunol. 10:1776.

Balfour HH. 1979. Cytomegalovirus: the troll of transplantation. Arch Intern Med. 139(3):279–280.

Bank C, et al. 2016. An experimental evaluation of drug induced mutational meltdown as an antiviral strategy. Evolution 70(11): 2470–2484.

Bank C, Foll M, Ferrer-Admetlla A, Ewing G, Jensen JD. 2014. Thinking too positive? Revisiting current methods in population genetic statistical inference. Trends Genet. 30(12):540–546.

Bardanzellu F, Fanos V, Reali A. 2019. Human breast milk-acquired cytomegalovirus infection: certainties, doubts and perspectives. Curr Pediatr Rev. 15(1):30–41.

Beaucourt S, et al. 2011. Isolation of fidelity variants of RNA viruses and characterization of virus mutation frequency. J Vis Exp. 52:2953.

Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlation with recombination rates in *D. melanogaster*. Nature 356(6369):519–520.

Birkner M, Blath J, Eldon B. 2013. An ancestral recombination graph for diploid populations with skewed offspring distribution. Genetics 193(1):255–290.

Bodaghi B, et al. 1999. Entry of human cytomegalovirus into retinal pigment epithelial and endothelial cells by endocytosis. Investig Ophthalmol Vis Sci. 40(11):2598–2607.

Boppana SB, Ross SA, Fowler KB. 2013. Congenital cytomegalovirus infection: clinical outcome. Clin Infect Dis. 57(Suppl 4):S178–S181.

Britt WJ. 2017. Congenital human cytomegalovirus infection and the enigma of maternal immunity. J Virol. 91(15):e02392-16.

Cha T-A, et al. 1996. Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. J Virol. 70(1):78–83.

Charlesworth B, Jensen JD. 2021. Effects of selection at linked sites on patterns of genetic variability. Annu Rev Ecol Evol. 52:177–197.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134(4):1289–1303.

Chou SW. 1989. Reactivation and recombination of multiple cytomegalovirus strains from individual organ donors. J Infect Dis. 160(1):11–15.

Chou S, Ercolani RJ, Lanier ER. 2016. Novel cytomegalovirus UL54 DNA polymerase gene mutations selected *in vitro* that confer brincidofovir resistance. Antimicrob Agents Chemother. 60(6):3845–3848.

Cobbs CS, et al. 2002. Human cytomegalovirus infection and expression in human malignant glioma 1. Cancer Res. 62(12): 3347–3350.

Crotty S, Cameron CE, Andino R. 2001. RNA Virus error catastrophe: direct molecular test by using ribavirin. Proc Natl Acad Sci U S A. 98(12):6895–6900.

Cudini J, et al. 2019. Human cytomegalovirus haplotype reconstruction reveals high diversity due to superinfection and evidence of within-host recombination. Proc Natl Acad Sci U S A. 116(12): 5693–5698.

Di Liberto G, et al. 2006. Clinical and therapeutic implications of hepatitis C virus compartmentalization. J Gastroenterol. 131(1):76–84.

Dolan A, et al. 2004. Genetic content of wild-type human cytomegalovirus. J Gen Virol. 85(Pt 5):1301–1312.

Drake JW, Hwang CBC. 2005. On the mutation rate of herpes simplex virus type 1. Genetics 170(2):969–970.

Dreher AM, et al. 2014. Spectrum of disease and outcome in children with symptomatic congenital cytomegalovirus infection. J Pediatr. 164(4):855–859.

Drew WL, Sweet ES, Miner RC, Mocarski ES. 1984. Multiple infections by cytomegalovirus in patients with acquired immunodeficiency syndrome: documentation by Southern blot hybridization. J Infect Dis. 150(6):952–953.

Dupont L, Reeves MB. 2016. Cytomegalovirus latency and reactivation: recent insights into an age old problem. Rev Med Virol. 26(2):75–89.

Eldon B, Wakeley J. 2008. Linkage disequilibrium under skewed offspring distribution among individuals in a population. Genetics 178(3):1517–1532.

Enders G, Daiminger A, Bäder U, Exler S, Enders M. 2011. Intrauterine transmission and clinical outcome of 248 pregnancies with primary cytomegalovirus infection in relation to gestational age. J Clin Virol. 52(3):244–246.

Ewing GB, Jensen JD. 2016. The consequences of not accounting for background selection in demographic inference. Mol Ecol. 25(1): 135–141.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. PLoS Genet. 9(10):e1003905.

Faure-Della Corte M, et al. 2010. Variability and recombination of clinical human cytomegalovirus strains from transplantation recipients. J Clin Virol. 47(2):161–169.

Felsenstein J. 1974. The evolutionary advantage of recombination. Genetics 78(2):737–756.

Fischer W, et al. 2010. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. PLoS One 5(8):e12303.

Fonager J, et al. 2015. Identification of minority resistance mutations in the HIV-1 integrase coding region using next generation sequencing. J Clin Virol. 73:95–100.

Frange P, et al. 2013. Temporal and spatial compartmentalization of drug-resistant cytomegalovirus (CMV) in a child with CMV meningoencephalitis: implications for sampling in molecular diagnosis. J Clin Microbiol. 51(12):4266–4269.

Girsch JH, et al. 2022. Host-viral interactions at the maternal-fetal interface. What we know and what we need to know. Front Virol. 2:16.

Griffiths P, Baraniak I, Reeves M. 2015. The pathogenesis of human cytomegalovirus. J Pathol. 235(2):288–297.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5(10): e1000695.

Haberland M, Meyer-König U, Hufert FT. 1999. Variation within the glycoprotein B gene of human cytomegalovirus is due to homologous recombination. J Gen Virol. 80(Pt 6):1495–1500.

Hage E, et al. 2017. Characterization of human cytomegalovirus genome diversity in immunocompromised hosts by whole-genome sequencing directly from clinical specimens. J Infect Dis. 215(11): 1673–1683.

Hahn G, et al. 2004. Human cytomegalovirus UL131-128 genes are indispensable for virus growth in endothelial cells and virus transfer to leukocytes. J Virol. 78(18):10023–10033.

Hall JD, Almy RE. 1982. Evidence for control of herpes simplex virus mutagenesis by the viral DNA polymerase. Virology 116(2): 535–543.

He R, et al. 2006. Sequence variability of human cytomegalovirus UL146 and UL147 genes in low-passage clinical isolates. Intervirology 49(4):215–223.

Heo J, et al. 2008. Polymorphisms within human cytomegalovirus chemokine (UL146/UL147) and cytokine receptor genes (UL144) are not predictive of sequelae in congenitally infected children. Virology 378(1):86–96.

Hetzenecker S, Helenius A, Krzyzaniak MA. 2016. HCMV Induces macropinocytosis for host cell entry in fibroblasts. Traffic 17(4): 351–368.

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet Res. 8:269–294.

Houldcroft CJ, et al. 2016. Detection of low frequency multi-drug resistance and novel putative maribavir resistance in immunocompromised pediatric patients with cytomegalovirus. Front Microbiol. 7:1317.

Houldcroft CJ, Cudini J, Goldstein RA, Breuer J. 2020. Reply to Jensen and Kowalik: consideration of mixed infections is central to understanding HCMV intrahost diversity. Proc Natl Acad Sci U S A. 117(2):818–819.

Hwang YT, Liu BY, Hwang CBC. 2002. Replication fidelity of the supF gene integrated in the thymidine kinase locus of herpes simplex virus type 1. J Virol. 76(8):3605–3614.

Irwin KK, et al. 2016. On the importance of skewed offspring distributions and background selection in virus population genetics. Heredity 117:393–399.

Jarvis MA, Nelson JA. 2002. Mechanisms of human cytomegalovirus persistence and latency. Front Biosci. 7:d1575–d1582.

Jean JH, Yoshimura N, Furukawa T, Plotkin SA. 1978. Intracellular forms of the parental human cytomegalovirus genome at early stages of the infective process. Virology 86(1):281–286.

Jean Beltran PM, Cristea IM. 2014. The life cycle and pathogenesis of human cytomegalovirus infection: lessons from proteomics. Expert Rev Proteomics 11(6):697–711.

Jensen JD. 2009. On reconciling single and recurrent hitchhiking models. Genome Biol Evol. 1:320–324.

Jensen JD. 2021. Studying population genetic processes in viruses: from drug-resistance evolution to patient infection dynamics. 4th ed. Encyclopedia of Virology 5:227–232.

Jensen JD, Kowalik TF. 2020. A consideration of within-host human cytomegalovirus genetic variation. Proc Natl Acad Sci. 117(2): 816–817.

Jensen JD, Lynch M. 2020. Considering mutational meltdown as a potential SARS-CoV-2 treatment strategy. Heredity 124:619–620.

Jensen JD, Stikeleather RA, Kowalik TF, Lynch M. 2020. Imposed mutational meltdown as an antiviral strategy. Evolution 74(12): 2549–2559.

Jerzak G, Bernard KA, Kramer LD, Ebel GD. 2005. Genetic variation in West Nile virus from naturally infected mosquitoes and birds suggests quasispecies structure and strong purifying selection. J Gen Virol. 86(Pt 8):2175–2183.

Ji YH, et al. 2006. Polymorphisms of human cytomegalovirus UL148A, UL148B, UL148C, UL148D genes in clinical strains. J Clin Virol. 37(4):252–257.

Johnson EL, et al. 2015. Cytomegalovirus upregulates expression of CCR5 in central memory cord blood mononuclear cells, which may facilitate in utero HIV type 1 transmission. J Infect Dis. 211(2):187–196.

Johri P, et al. 2021. The impact of purifying and background selection on the inference of population history: problems and prospects. Mol Biol Evol. 38(7):2986–3003.

Johri P, Aquadro CF, et al. 2022. Recommendations for improving statistical inference in population genomics. PLoS Biol. 20: e3001669.

Johri P, Charlesworth B, Jensen JD. 2020. Toward an evolutionarily appropriate null model: jointly inferring demography and purifying selection. Genetics 215(1):173–192.

Johri P, Eyre-Walker A, Gutenkunst RN, Lohmueller KE, Jensen JD. 2022. On the prospect of achieving accurate joint estimation of selection with population history. Genome Biol Evol. 14(7): evac088.

Keele BF, et al. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A. 105(21):7552–7557.

Kelleher J, et al. 2019. Inferring whole-genome histories in large population datasets. Nat Genet. 51(9):1330–1338.

Kimura M. 1968. Evolutionary rate at the molecular level. Nature 217: 624–626.

Kinzler ER, Compton T. 2005. Characterization of human cytomegalovirus glycoprotein-induced cell-cell fusion. J Virol. 79(12): 7827–7837.

Kolb AW, Ané C, Brandt CR. 2013. Using HSV-1 genome phylogenetics to track past human migrations. PLoS One 8(10):e76267.

Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW. 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics 22(14):3096–3098.

Kyeyune F, et al. 2016. Low-frequency drug resistance in HIV-infected Ugandans on antiretroviral treatment is associated with regimen failure. Antimicrob Agents Chemother. 60(6):3380–3397.

Li G, et al. 2014. An epistatic relationship between the viral protein kinase UL97 and the UL133-UL138 latency locus during the human cytomegalovirus lytic cycle. J Virol. 88(11):6047–6060.

Luftig MA. 2014. Viruses and the DNA damage response: activation and antagonism. Annu Rev Virol. 1(1):605–625.

Luria SE. 1951. The frequency distribution of spontaneous bacteriophage mutants as evidence for the exponential rate of phage reproduction. Cold Spring Harb Symp Quant Biol. 16:463–470.

Lynch M, et al. 2016. Genetic drift, selection and the evolution of the mutation rate. Nat Rev Genet. 17(11):704–714.

Lynch M, Conery J, Burger R. 1995. Mutation accumulation and the extinction of small populations. Am Nat. 146(4):489–518.

Matuszewski S, Hildebrandt ME, Achaz G, Jensen JD. 2018. Coalescent processes with skewed offspring distributions and non-equilibrium demography. Genetics 208(1):323–338.

Matuszewski S, Ormond L, Bank C, Jensen JD. 2017. Two sides of the same coin: a population genetics perspective on lethal mutagenesis and mutational meltdown. Virus Evol. 3(1):vex004.

Mayer BT, et al. 2017. Transient oral human cytomegalovirus infections indicate inefficient viral spread from very few initially infected cells. J Virol. 91(12):e00380-17.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351(6328):652–654.

McGeoch DJ, Dolan A, Ralph AC. 2000. Toward a comprehensive phylogeny for mammalian and avian herpesviruses. J Virol. 74(22): 10401–10406.

Meyer-König U, Haberland M, Von Laer D, Haller O, Hufert FT. 1998. Intragenic variability of human cytomegalovirus glycoprotein B in clinical strains. J Infect Dis. 177(5):1162–1169.

Meyer-König U, Vogelberg C, et al. 1998. Glycoprotein B genotype correlates with cell tropism in vivo of human cytomegalovirus infection. J Med Virol. 55(1):75–81.

Mocarski ES. 2004. Immune escape and exploitation strategies of cytomegaloviruses: impact on and imitation of the major histocompatibility system. Cell Microbiol. 6(8):707–717.

Morales-Arce AY, Harris RB, Stone AC, Jensen JD. 2020. Evaluating the contributions of purifying selection and progeny-skew in dictating within-host Mycobacterium tuberculosis evolution. Evolution 74(5):992–1001.

Morales-Arce AY, Johri P, Jensen JD. 2022. Inferring the distribution of fitness effects in patient-sampled and experimental virus populations: two case studies. Heredity 128(2):79–87.

Mozzi A, et al. 2020. Past and ongoing adaptation of human cytomegalovirus to its host. PLoS Pathog. 16(5):e1008476.

Nishide M, et al. 2019. Antiviral and virucidal activities against herpes simplex viruses of umesu phenolics extracted from Japanese apricot. Microbiol Immunol. 63(9):359–366.

Nishiyama Y, Maeno K, Yoshida S. 1983. Characterization of human cytomegalovirus-induced DNA polymerase and the associated 3′-to-5′, exonuclease. Virology 124(2):221–231.

Numazaki K. 1997. Human cytomegalovirus infection of breast milk. FEMS Microbiol Immunol. 18(2):91–98.

Ormond L, et al. 2017. The combined effect of oseltamivir and favipiravir on influenza A virus evolution. Genome Biol Evol. 9(7): 1913–1924.

Pang J, et al. 2020. Mixed cytomegalovirus genotypes in HIV-positive mothers show compartmentalization and distinct patterns of transmission to infants. eLife 9:e63199.

Pang X, Humar A, Preiksaitis JK. 2008. Concurrent genotyping and quantitation of cytomegalovirus gB genotypes in solid-organ-transplant recipients by use of a real-time PCR assay. J Clin Microbiol. 46(12):4004–4010.

Peck KM, Lauring AS. 2018. Complexities of viral mutation rates. J Virol. 92(14):e01031-17.

Pénisson S, Singh T, Sniegowski P, Gerrish P. 2017. Dynamics and fate of beneficial mutations under lineage contamination by linked deleterious mutations. Genetics 205(3):1305–1318.

Pfeifer SP. 2020. Spontaneous mutation rates. In: Ho, S.Y.W. (eds), *The molecular evolutionary clock. Theory and practice*. Cham: Springer Nature. p. 35–44. https://doi.org/10.1007/978-3-030-60181-2_3

Pfeifer SP, Jensen JD. 2016. The impact of linked selection in chimpanzees: a comparative study. Genome Biol Evol. 8(10):3202–3208.

Plachter B, Sinzger C, Jahn G. 1996. Cell types involved in replication and distribution of human cytomegalovirus. Adv Virus Res. 46: 195–261.

Pokalyuk C, et al. 2017. Characterizing human cytomegalovirus reinfection in congenitally infected infants: an evolutionary perspective. Mol Ecol. 26(7):1980–1990.

Porter KR, Starnes DM, Hamilton JD. 1985. Reactivation of latent murine cytomegalovirus from kidney. Kidney Int. 28(6):922–925.

Pouyet F, Aeschbacher S, Thiéry A, Excoffier L. 2018. Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. eLife 7:e36317.

Puchhammer-Stöckl E, Görzer I. 2011. Human cytomegalovirus: an enormous variety of strains and their possible clinical significance in the human host. Future Virol. 6(2):259–271.

Renzette N, et al. 2013. Rapid intrahost evolution of human cytomegalovirus is shaped by demography and positive selection. PLoS Genet. 9(9):e1003735.

Renzette N, et al. 2015. Limits and patterns of cytomegalovirus genomic diversity in humans. Proc Natl Acad Sci U S A. 112(30): E4120–E4128.

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.

Renzette N, Gibson L, Jensen JD, Kowalik TF. 2014. Human cytomegalovirus intrahost evolution—a new avenue for understanding and controlling herpesvirus infections. Curr Opin Virol. 8:109–115.

Renzette N, Kowalik TF, Jensen JD. 2016. On the relative roles of background selection and genetic hitchhiking in shaping human cytomegalovirus genetic diversity. Mol Ecol. 25(1):403–413.

Renzette N, Pfeifer SP, Matuszewski S, Kowalik TF, Jensen JD. 2017. On the analysis of intrahost and interhost viral populations: human cytomegalovirus as a case study of pitfalls and expectations. J Virol. 91(5):1976–1992.

Revello MG, Gerna G. 2010. Human cytomegalovirus tropism for endothelial/epithelial cells: scientific background and clinical implications. Rev Med Virol. 20(3):136–155.

Rosenberg N, Nordborg M. 2002. Genealogical trees, coalescent theory, and the analysis of genetic polymorphisms. Nat Rev Genet. 3(5):380–390.

Rousselle M, Mollion M, Nabholz B, Bataillon T, Galtier N. 2018. Overestimation of the adaptive substitution rate in fluctuating populations. Biol Lett. 14(15):20180055.

Sabin S, Morales-Arce AY, Pfeifer SP, Jensen JD. 2022. The impact of frequently neglected model violations on bacterial recombination rate estimation: a case study in *Mycobacterium canettii* and *Mycobacterium tuberculosis*. G3 (Bethesda) 12(5):jkac055.

Sackman AM, Harris RB, Jensen JD. 2019. Inferring demography and selection in organisms characterized by skewed offspring distributions. Genetics 211(3):1019–1028.

Sackman AM, Pfeifer SP, Kowalik TF, Jensen JD. 2018. On the demographic and selective forces shaping patterns of human cytomegalovirus variation within hosts. Pathogens 7(1):16.

Sakaoka H, et al. 1994. Quantitative analysis of genomic polymorphism of herpes simplex virus type 1 strains from six countries: studies of molecular evolution and molecular epidemiology of the virus. J Gen Virol. 75(Pt 3):513–527.

Sanjuán R. 2010. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. Philos Trans R Soc Lond B Biol Sci. 365(1548):1975–1982.

Sijmons S, et al. 2015. High-throughput analysis of human cytomegalovirus genome diversity highlights the widespread occurrence of gene-disrupting mutations and pervasive recombination. J Virol. 89(15):7673–7695.

Sinzger C, Digel M, Jahn G. 2008. Cytomegalovirus cell tropism. Curr Top Microbiol Immunol. 325:63–83.

Sowmya P, Madhavan HN. 2009. Analysis of mixed infections by multiple genotypes of human cytomegalovirus in immunocompromised patients. J Med Virol. 81(5):861–869.

Spector SA, Hirata KK, Neuman TR. 1984. Identification of multiple cytomegalovirus strains in homosexual men with acquired immunodeficiency syndrome. J Infect Dis. 150(6):953–956.

Steinrücken M, Kamm J, Spence JP, Song YS. 2019. Inference of complex population histories using whole-genome sequences from multiple populations. Proc Natl Acad Sci U S A. 116(34):17115–17120.

Stumpf MPH, McVean GAT. 2003. Estimating recombination rates from population-genetic data. Nat Rev Genet. 4(12):959–968.

Suárez NM, et al. 2019. Human cytomegalovirus genomes sequenced direction from clinical material: variation, multiple-strain infection, recombination, and gene loss. J Infect Dis. 220(5):781–791.

Suárez NM, et al. 2020. Whole-genome approach to assessing human cytomegalovirus dynamics in transplant patients undergoing antiviral therapy. Front Cell Infect Microbiol. 10:267.

Swanson EC, Schleiss MR. 2013. Congenital cytomegalovirus infection: new prospects for prevention and therapy: for pediatric clinics of North America: advances in evaluation, diagnosis and treatment of pediatric infectious disease. Pediatr Clin North Am. 60(2):335–349.

Taddei F, et al. 1997. Role of mutator alleles in adaptive evolution. Nature 387(6634):700–702.

Tellier A, Lemaire C. 2014. Coalescence 2.0: a multiple branching of recent theoretical developments and their applications. Mol Ecol. 23(11):2637–2652.

Travis JMJ, Travis ER. 2002. Mutator dynamics in fluctuating environments. Philos Trans R Soc Lond B Biol Sci. 269(1491):591–597.

Van Damme E, Van Loock M. 2014. Functional annotation of human cytomegalovirus gene products: an update. Front Microbiol. 5:218.

Wakeley J. 2009. Coalescent theory: an introduction. Greenwood Village: Roberts & Company Publishers.

Wang HY, et al. 2021. Common polymorphisms in the glycoproteins of human cytomegalovirus and associated strain-specific immunity. Viruses 13(6):1106.

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):4662–4665.

Weekes MP, et al. 2014. Quantitative temporal viromics: an active approach to investigate host-pathogen interaction. Cell 157: 1460–1472.

Weitzman MD, Lilley CE, Chaurushiya MS. 2010. Genomes in conflict: maintaining genome integrity during virus infection. Annu Rev Microbiol. 64:61–81.

Xiaofei E, Kowalik TF. 2014. The DNA damage response induced by infection with human cytomegalovirus and other viruses. Viruses 6(5):2155–2185.

Yan H, et al. 2008. Genetic linkage among human cytomegalovirus glycoprotein N (gN) and gO genes, with evidence for recombination from congenitally and post-natally infected Japanese infants. J Gen Virol. 89(Pt 9):2275–2279.

Yatim N, Albert ML. 2011. Dying to replicate: the orchestration of the viral life cycle, cell death pathways, and immunity. Immunity 35(4):478–490.

Zárate S, Pond SLK, Shapshak P, Frost SDW. 2007. Comparative study of methods for detecting sequence compartmentalization in human immunodeficiency virus type 1. J Virol. 81(12):6643–6651.

Zheng K, et al. 2014. Epidermal growth factor receptor-PI3K signalling controls cofilin activity to facilitate herpes simplex virus 1 entry into neuronal cells. mBio 15(1):e00958-13.

Zwart MP, Elena SF. 2015. Matters of size: genetic bottlenecks in virus infection and their potential impact on evolution. Annu Rev Virol. 2(1):161–179.

Associate editor: Wenfeng Qian