bioRxiv preprint doi: https://doi.org/10.1101/009571. this version posted September 30, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Title

Human Cytomegalovirus Intrahost Evolution – A New Avenue for Understanding and Controlling Herpesvirus Infections

Nicholas Renzette[1], Laura Gibson[2], Jeffrey D. Jensen[3,4,5] and Timothy F. Kowalik[1,6,*]

1. Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, Worcester MA 01655, USA

2. Departments of Pediatrics and Medicine, Divisions of Infectious Diseases and Immunology, University of Massachusetts Medical School, Worcester MA 01655, USA

3. Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester MA 01655, USA

4. School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne 1015, Switzerland

5. Swiss Institute of Bioinformatics (SIB), Lausanne 1015, Switzerland

6. Immunology and Microbiology Program, University of Massachusetts Medical School, Worcester MA 01655, USA

*corresponding author: timothy.kowalik@umassmed.edu

Short Title

Human Cytomegalovirus Intrahost Evolution

bioRxiv preprint doi: https://doi.org/10.1101/009571. this version posted September 30, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Abstract

Human cytomegalovirus (HCMV) is exquisitely adapted to the human host, and much research has focused on its evolution over long timescales spanning millennia. Here, we review recent data exploring the evolution of the virus on much shorter timescales, on the order of days or months. We describe the intrahost genetic diversity of the virus isolated from humans, and how this diversity contributes to HCMV spatiotemporal evolution. We propose mechanisms to explain the high levels of intrahost diversity and discuss how this new information may shed light on HCMV infection and pathogenesis.

Introduction

Collectively, the cytomegaloviruses are ancient viruses that appear to infect the majority of vertebrates, but each particular cytomegalovirus infects a single host species with little evidence that the viruses jump between hosts. These traits have contributed to cytomegaloviruses adapting over long time scales (e.g. millennia) to the particular ecological niches in which they occupy. In particular, human cytomegalovirus is a β-herpesvirus that produces lifelong infections in most humans. In most individuals, HCMV infections result in a mild febrile illness, but can lead to severe symptoms in the immune-compromised or neonates. Nearly all HCMV infections result in widespread dissemination throughout the body, with diverse cell types such as epithelial, endothelial, fibroblast, and smooth muscle cells supporting productive viral infection. Further, the virus induces a myriad of immunomodulatory pathways to subvert the host innate and adaptive response [1]. Many excellent studies and reviews (for example, see [2] or [3]) have focused on the long term evolution of the virus and have highlighted the molecular characteristics that have helped create a highly successful pathogen. However, work mostly from the past decade has shown that the virus evolves on much shorter timescales, on the order of days or months, and within individual infected hosts. The purpose of this review is to discuss recent findings illustrating the complexity of short-term HCMV evolution and how it may relate to clinical manifestations of infection.

Diversity of Human Cytomegalovirus in Human Hosts

Work published more than 30 years ago based on restriction length polymorphisms (RFLP) has shown that HCMV exhibits unexpected levels of genetic diversity between individuals (i.e., *inter*host variability) [4]. These results were later confirmed with targeted re-sequencing of genetic loci, primarily focusing on glycoproteins such as gB, gN and g0, as well as whole genome data from both low passage isolates and clinical samples [5,6]. However, work reported more recently has shown that HCMV also exhibits significant levels of genetic diversity within a single individual (i.e., *intra*host diversity). Similar to interhost variability, most intrahost diversity data has focused on the HCMV glycoproteins. Various assays, such as PCR-RFLP [7], DNA sequencing [8], single strand conformation polymorphism [9] and heteroduplex mobility analysis [9], show that these regions can be variable within individuals. However, these studies were limited by the technologies available, and suffered from issues associated with low sampling depth and narrow breadth of the HCMV genome. Nevertheless, they regularly showed that mixed infections accounted for roughly one quarter to one half of HCMV infections over a wide range of human populations, including infants with congenital infection [10], people with HIV/AIDS [11,12], transplant recipients, and immunocompetent children [13] and adults [14]. An important paper from 2011 began to show the full extent of HCMV intrahost diversity by using ultra-deep pyrosequencing to study the gO, gN and gH loci from HCMV-infected transplant recipients [15]. They found that all patients studied had mixed infections, with as many as 6 genotypes observed in a single patient. Similarly, deep sequencing of 81,224 base pairs of the HCMV genome directly from a glioblastoma multiforme (GBM) tumor revealed high genetic diversity of the virus including an overabundance of hypervariable loci, in which all four nucleotides were observed at a single position in the viral genome. Work that combined genomics with high throughput sequencing showed that HCMV diversity was not limited to a subset of loci, but rather

bioRxiv preprint doi: https://doi.org/10.1101/009571. this version posted September 30, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

spanned the entire genome [16]. In fact, nearly every open reading frame (ORF) of the HCMV genome exhibited a measurable level of intrahost genetic diversity. Importantly, HCMV diversity was described quantitatively using various metrics and shown to rival that of some RNA viruses, the benchmark of highly diverse viral populations (Figure 1). From these studies, it has been become clear that HCMV is highly diverse within humans.

Mechanisms of HCMV Diversity

Given the high levels of HCMV intrahost diversity, examination of its source is warranted. Currently, there is no clear mechanism to explain the observed diversity, though the literature does offer several non-mutually exclusive mechanisms that could be playing a role. High levels of replication during primary infection could contribute to the generation of *de novo* mutations in each host, though the total number of mutations generated may be limited by the proofreading activity of the viral DNA polymerase [17]. The excess of low frequency mutations in HCMV intrahost populations [15] as well as the star phylogeny of clones sampled from the populations [16] are consistent with this pattern (Figure 2A). However, clones from HCMV clinical samples also reveal that some members of the population are highly divergent from the most common sequence, a result consistent with reinfection (Figure 2B). HCMV reinfections appear to be common in both immunocompromised and healthy individuals [4,18-20]. Indeed, reinfections in some populations may occur at an annualized rate of 10%, comparable to the annualized rate of new infections [21]. Thus, reinfections offer a route for repeatedly introducing diversity into HCMV intrahost populations. However, this mechanism does not immediately appear to address the diversity observed in very young patients, specifically congenitally infected neonates. Here, the explanation may lie in the transplacental route of transmission associated with congenital infection. One recent study showed that maternal-to-fetal transmission may represent 10s to 100s of unique HCMV virions [22]. In contrast, transmission associated with RNA viruses is associated with a much lower number of virions, with bottlenecks so restrictive that only a single sequence transmits as in the case of HIV [23,24]. In this extreme case of a single virion transmission, all pre-existing viral population diversity is lost and must be generated *de novo* during replication in the newly infected host. More relaxed bottlenecks would result in 100s of HCMV virions being transmitted during congenital infections. In this circumstance, a portion of the pre-existing diversity present in the mother may be transmitted to the fetus. Whether this mechanism contributes to HCMV diversity in other transmission contexts, such as in healthy children or adults, is unknown.

The two mechanisms of recombination and natural selection may also alter HCMV intrahost diversity. For other organisms, a positive correlation has been shown between recombination rate and diversity [25]. Both positive selection of beneficial alleles and negative selection against deleterious alleles will reduce diversity at a locus due to the fixation or clearance of mutations, respectively [26,27]. By uncoupling neutral variation from the beneficial or deleterious alleles, recombination can allow for higher levels of neutral variation to remain in the population. There is ample evidence of recombination in the HCMV genome [28-31], so this mechanism may play a role in shaping HCMV diversity. However, to date the recombination rate has not been quantified across the HCMV genome, while only a limited number of studies have analyzed the strength or frequency of selection targeting HCMV intrahost populations.

Change of Populations in Human Hosts

With such high levels of intrahost diversity, researchers began to ask whether this leads to highly dynamic viral populations. For example, do the frequencies of specific mutations or genotypes change over time, and if so, how rapidly? Much research suggests that HCMV populations are generally stable over time. Longitudinal evolution studies of HCMV have predominantly focused on solid organ transplant recipients. In this patient population, genotyping via qPCR and sequencing of HCMV loci have shown mixed viral populations, but the composition of the mixed populations remains nearly constant in most patients over

bioRxiv preprint doi: https://doi.org/10.1101/009571. this version posted September 30, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

time [32-35]. A longitudinal study of congenitally infected infants showed similarly stable populations [22]. The predominant pattern from this work shows that HCMV populations contain a dominant sequence (which may represent > 90% of the total viral population) while the remaining population is occupied by a diverse array of largely stable sequences (i.e. minor variants) [15,16].

However, some caveats should be considered with this statement of longitudinal genetic stability. First, there may exist many thousands of stable mutations in an HCMV population [15,16,22,36] that do not rise in frequency over time, but importantly, are also not cleared from the population. These stable mutations (i.e., standing variation) provide a mechanism by which the virus could rapidly adapt to changing environments, such as new host compartments, immune pressure or antiviral drugs. Second, even if most mutations are stable, the change of only a few can dramatically alter viral fitness or pathogenesis. The increase in frequency of antiviral drug resistance mutations [37-39] clearly illustrates this point. Interestingly, antiviral treatment has also been shown to cause changes in non-drug resistance associated loci, such as gB, presumably due to linkage with a drug resistance mutation [33]. Lastly, HCMV populations that undergo spatial compartmentalization, as discussed in the next section, appear to undergo rapid evolution.

Tissue compartmentalization of HCMV in Human Hosts

A hallmark of HCMV infections is dissemination to a wide range of host tissue compartments and cell types [40]. Due to the high levels of diversity associated with HCMV infections, researchers have begun to study whether viral diversity sampled from compartments differ, both in total level of variation and in specific genetic or genomic sequences observed. Indeed, early reports of diverse HCMV infections also demonstrated that HCMV sampled from different organs of healthy individuals could differ in gB genotype composition [14,41]. This finding has been supported in subsequent work studying other genomic loci [10,33,42]. Quantitative descriptions of the diversity have demonstrated significant differences in the level of diversity between compartments, with blood populations appearing more diverse than urine or intraocular populations [7,22]. However, the most striking example of compartmentalization may be differences in drug resistance genotypes observed in some patients. Several studies have shown that ganciclovir (GCV) resistance genotypes are not evenly distributed in patient compartments during treatment [43-46], with, for example, drug sensitive strains present in the cerebrospinal fluid and drug resistant strains in the blood [46].

Although the mechanism(s) explaining compartmentalization are not clear, there are at least two hypotheses that can be proposed. First, compartmentalization could result from the stochasticity of dissemination. Bottlenecking of the viral populations during dissemination will increase the proportion of mutations in the population that are governed by stochastic fluctuations in frequency (i.e., genetic drift) [47]. These stochastic changes can drastically skew the composition of populations in the distal relative to the original compartment [48]. Second, compartmentalization could result from natural selection during dissemination. Selection for mutants more fit in a compartment, for example due to cellular tropism, could lead to distinct populations in that tissue. Few studies have tested between these two models. There is evidence that selection [49] or both bottlenecks and selection can explain the observed compartmentalization [22] but the phenomenon needs to be studied in a wider range of patient populations to distinguish between these possibilities.

HCMV Diversity, Evolution and Clinical Disease: Is there a connection?

We are now developing a clearer understanding of HCMV intrahost genetic diversity and evolution, but there is still the looming question of how these observations inform our understanding of HCMV disease.

bioRxiv preprint doi: https://doi.org/10.1101/009571. this version posted September 30, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

While this question is clinically relevant, work connecting diversity and pathogenesis is not fully developed, so discussion of its implications remains more speculative than others.

Many conflicting reports have examined the correlation between variation of HCMV genotype sequences and risk of transmission or disease progression and severity [41,50-54]. From this work, no clear conclusion can be reached about a possible effect of any single genotype on disease. However, subsequent research has shown that mixed genotype infections do correlate with increased viral load, HCMV disease severity, and even progression to AIDS in people with HIV infection [7,8,55,56]. Further, studies of intrauterine transmission of mothers with preconceptional immunity showed a higher rate of transmission and symptomatic congenital infections in mothers who were reinfected during pregnancy as compared to those who were not reinfected [18,19]. From these data, it is not clear if reinfection *per se* or merely the presence of multiple strains lead to the clinical outcomes. It has been demonstrated in a mouse model of cytomegalovirus infection that in vivo complementation occurs due to co-infection, allowing for the replication and dissemination of an attenuated strain due to transcomplementation from a wild type strain [57]. Thus, it is possible that multiple strain infections alter HCMV fitness by allowing a largely haploid virus to act as a polyploid.

Two specific areas where HCMV intrahost diversity may alter pathogenesis are dissemination and immune evasion. The possible role in dissemination is based on studies of the viral glycoproteins, proteins important in cell-to-cell spread and tissue tropism. In one report, many gB sequences were isolated from urine samples, but only a subtype of sequences were associated with plasma, suggesting that genetic diversity influences tissue tropism of the virus [41]. Similarly, variation in the gO protein alters the ratio of the endothelial-tropic gH/gL/UL128-131 complex to the fibroblast-tropic gH/gL/gO complex and thus was proposed to alter tropism and viral dissemination. From these data, it is possible to speculate that a diverse population of HCMV may harbor a subset of mutants that better replicate in distinct host compartments, and that viral diversity therefore enhances viral dissemination. In addition to dissemination, the possible effect of immune evasion on the relationship between viral intrahost diversity and pathogenesis has been suggested but not clearly proven. HCMV infections are regulated by both innate and adaptive immune responses. Variation in the *UL40* gene product modulates NK-cell mediated immunosurveillance by altering interaction of HLA-E with the CD94-NKG2A inhibitory receptor [58]. In studies of congenital infections, the intrahost diversity of *UL40,* particularly in the HLA-E binding domain, is significantly higher than the genome wide average [16]. Thus, mutants in the population likely display varied phenotypes in terms of NK-cell evasion, which may have profound effects on viral pathogenesis and the balance between productive and latent infections. Similar to NK cell evasion, variation in $CD8^+$ T cell epitopes can alter the host adaptive immune response. Primary infection with mixed strains leads to the production of cross-reactive T cells, though exposure to some viral variants may limit the range of epitopes and viruses recognized by the T cell population [59]. Further, epitope recognition of the host T cell response has been shown to broaden over time in congenital infections [60], and thus may select against a wide range of viral mutations. Whether viral diversity is altered by and/or drives the adaptive immune response is unclear. Future work is needed to elucidate the directionality of this dynamic relationship between HCMV and its human host and the contribution of pre-existing and *de novo* viral mutants in the process.

Conclusions and Future Directions

Here we highlight a few areas in which additional study could provide valuable information for ongoing and future efforts to lessen the burden of HCMV disease. An increasing body of evidence has demonstrated that HCMV is genetically diverse in human hosts, and there is tantalizing evidence that this diversity may contribute to pathogenesis. As such, HCMV intrahost diversity will remain as an important avenue to explore when developing novel antivirals and effective vaccines against this virus, as well as in

bioRxiv preprint doi: https://doi.org/10.1101/009571. this version posted September 30, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

designing treatment regimens - with previous results in other viruses suggesting that reduced diversity may indeed be associated with decreased pathogenesis [61,62]. Yet, in order for HCMV diversity to become a therapeutic target, it is necessary to better characterize the evolutionary processes responsible for this diversity. Relatedly, the need remains to better explore the observed correlative relationship between HCMV diversity and disease in order to disentangle cause from effect, and in so doing characterize the number of mutations important for the ultimate disease phenotype. Finally, relatively little attention has been given thus far to virus-host interactions, and further illuminating the role of variable host genetics as well as the co-evolutionary relationship between these organisms will likely become fruitful as sequencing costs continue to fall, thus enabling large scale paired human-HCMV sequencing duos. Encouraging, with these continued advances on the intersection of virology and population genetics, there is every reason to be optimistic that these aims will result in new tools to aid in our ongoing struggle with HCMV.

Acknowledgments

We wish to thank our many colleagues who provided specimens from HCMV-infected patients that contributed to our understanding of intrahost evolution. This publication was supported by a grant from the National Institutes of Health (R01HD061959). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

bioRxiv preprint doi: https://doi.org/10.1101/009571. this version posted September 30, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

References

1. Mocarski ES (2002) Immunomodulation by cytomegaloviruses: manipulative strategies beyond evasion. Trends in Microbiology 10: 332-339.
2. McGeoch DJ, Rixon FJ, Davison AJ (2006) Topics in herpesvirus genomics and evolution. Virus Res 117: 90-104.
3. Davison AJ (2011) Evolution of sexually transmitted and sexually transmissible human herpesviruses. Annals of the New York Academy of Sciences 1230: E37-E49.
4. Huang ES, Huong SM, Tegtmeier GE, Alford C (1980) Cytomegalovirus: genetic variation of viral genomes. Ann N Y Acad Sci 354: 332-346.
5. Bradley AJ, Lurain NS, Ghazal P, Trivedi U, Cunningham C, et al. (2009) High-throughput sequence analysis of variants of human cytomegalovirus strains Towne and AD169. J Gen Virol 90: 2375-2380.
6. Cunningham C, Gatherer D, Hilfrich B, Baluchova K, Dargan DJ, et al. (2009) Sequences of complete human cytomegalovirus genomes from infected cell cultures and clinical specimens. J Gen Virol 91: 605-615.
7. Sowmya P, Madhavan HN (2009) Analysis of mixed infections by multiple genotypes of human cytomegalovirus in immunocompromised patients. J Med Virol 81: 861-869.
8. Coaquette A, Bourgeois A, Dirand C, Varin A, Chen W, et al. (2004) Mixed Cytomegalovirus Glycoprotein B Genotypes in Immunocompromised Patients. Clin Infect Dis 39: 155-161.
9. Trincado DE, Scott GM, White PA, Hunt C, Rasmussen L, et al. (2000) Human cytomegalovirus strains associated with congenital and perinatal infections. J Med Virol 61: 481-487.
10. Ross SA, Novak Z, Pati S, Patro RK, Blumenthal J, et al. (2011) Mixed infection and strain diversity in congenital cytomegalovirus infection. J Infect Dis 204: 1003-1007.
11. Spector SA, Hirata KK, Neuman TR (1984) Identification of Multiple Cytomegalovirus Strains in Homosexual Men with Acquired Immunodeficiency Syndrome. Journal of Infectious Diseases 150: 953-956.
12. 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: 952-953.
13. Numazaki K, Ikehata M, Asanuma H, Chiba S (1998) Simultaneous infection of immunocompetent individuals with multiple cytomegalovirus strains. The Lancet 352: 1710.
14. Meyer-König U, Ebert K, Schrage B, Pollak S, Hufert FT (1998) Simultaneous infection of healthy people with multiple human cytomegalovirus strains. The Lancet 352: 1280-1281.
15. Görzer I, Guelly C, Trajanoski S, Puchhammer-Stockl E (2010) Deep sequencing reveals highly complex dynamics of human cytomegalovirus genotypes in transplant patients over time. J Virol: JVI.00475-00410.
16. 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: e1001344.
17. Chou S, Marousek GI (2008) Accelerated evolution of maribavir resistance in a cytomegalovirus exonuclease domain II mutant. J Virol 82: 246-253.
18. Boppana SB, Rivera LB, Fowler KB, Mach M, Britt WJ (2001) Intrauterine Transmission of Cytomegalovirus to Infants of Women with Preconceptional Immunity. New England Journal of Medicine 344: 1366-1371.
19. Yamamoto AY, Mussi-Pinhata MM, Boppana SB, Novak Z, Wagatsuma VM, et al. (2010) Human cytomegalovirus reinfection is associated with intrauterine transmission in a highly cytomegalovirus-immune maternal population. American Journal of Obstetrics and Gynecology 202: 297.e291-297.e298.

bioRxiv preprint doi: https://doi.org/10.1101/009571. this version posted September 30, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

20. Bale Jr JF, Petheram SJ, Souza IE, Murph JR (1996) Cytomegalovirus reinfection in young children. The Journal of Pediatrics 128: 347-352.
21. Ross SA, Arora N, Novak Z, Fowler KB, Britt WJ, et al. (2010) Cytomegalovirus Reinfections in Healthy Seroimmune Women. Journal of Infectious Diseases 201: 386-389.
22. Renzette N, Gibson L, Bhattacharjee B, Fisher D, Schleiss MR, et al. (2013) Rapid Intrahost Evolution of Human Cytomegalovirus Is Shaped by Demography and Positive Selection. PLoS Genet 9: e1003735.
23. Fischer W, Ganusov VV, Giorgi EE, Hraber PT, Keele BF, et al. (2010) Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. PLoS One 5: e12303.
24. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, et al. (2008) Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proceedings of the National Academy of Sciences 105: 7552-7557.
25. Begun DJ, Aquadro CF (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. Nature 356: 519-520.
26. Maynard-Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet Res 23: 23-35.
27. Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. Genetics 134: 1289-1303.
28. Yan H, Koyano S, Inami Y, Yamamoto Y, Suzutani T, 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: 2275-2279.
29. 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: 1162-1169.
30. Haberland M, Meyer-König U, Hufert FT (1999) Variation within the glycoprotein B gene of human cytomegalovirus is due to homologous recombination. Journal of General Virology 80: 1495-1500.
31. Steininger C, Schmied B, Sarcletti M, Geit M, Puchhammer-Stockl E (2005) Cytomegalovirus genotypes present in cerebrospinal fluid of HIV-infected patients. AIDS 19: 273-278.
32. Garrigue I, Corte MF-D, Magnin N, Recordon-Pinson P, Couzi L, et al. (2008) UL40 Human Cytomegalovirus Variability Evolution Patterns Over Time in Renal Transplant Recipients. Transplantation 86: 826-835 810.1097/TP.1090b1013e3181859edd.
33. Görzer I, Kerschner H, Jaksch P, Bauer C, Seebacher G, et al. (2008) Virus load dynamics of individual CMV-genotypes in lung transplant recipients with mixed-genotype infections. J Med Virol 80: 1405-1414.
34. Puchhammer-Stockl E, Gorzer I, Zoufaly A, Jaksch P, Bauer CC, et al. (2006) Emergence of multiple cytomegalovirus strains in blood and lung of lung transplant recipients. Transplantation 81: 187-194.
35. Stanton R, Westmoreland D, Fox JD, Davison AJ, Wilkinson GW (2005) Stability of human cytomegalovirus genotypes in persistently infected renal transplant recipients. J Med Virol 75: 42-46.
36. Bhattacharjee B, Renzette N, Kowalik TF (2012) Genetic Analysis of Cytomegalovirus in Malignant Gliomas. Journal of Virology 86: 6815-6824.
37. Springer KL, Chou S, Li S, Giller RH, Quinones R, et al. (2005) How Evolution of Mutations Conferring Drug Resistance Affects Viral Dynamics and Clinical Outcomes of Cytomegalovirus-Infected Hematopoietic Cell Transplant Recipients. Journal of Clinical Microbiology 43: 208-213.
38. Chou S, Marousek G, Guentzel S, Follansbee SE, Poscher ME, et al. (1997) Evolution of Mutations Conferring Multidrug Resistance during Prophylaxis and Therapy for Cytomegalovirus Disease. Journal of Infectious Diseases 176: 786-789.

bioRxiv preprint doi: https://doi.org/10.1101/009571. this version posted September 30, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

39. Eckle T, Lang P, Prix L, Jahn G, Klingebiel T, et al. (2002) Rapid development of ganciclovir-resistant cytomegalovirus infection in children after allogeneic stem cell transplantation in the early phase of immune cell recovery. Bone Marrow Transplant 30: 433-439.
40. Sinzger C, Grefte A, Plachter B, Gouw ASH, The TH, et al. (1995) Fibroblasts, epithelial cells, endothelial cells and smooth muscle cells are major targets of human cytomegalovirus infection in lung and gastrointestinal tissues. Journal of General Virology 76: 741-750.
41. Meyer-König U, Vogelberg C, Bongarts A, Kampa D, Delbrück R, et al. (1998) Glycoprotein B genotype correlates with cell tropism in vivo of human cytomegalovirus infection. Journal of Medical Virology 55: 75-81.
42. Beyari MM, Hodgson TA, Kondowe W, Molyneux EM, Scully C, et al. (2005) Inter- and intra-person cytomegalovirus infection in Malawian families. J Med Virol 75: 575-582.
43. Hamprecht K, Eckle T, Prix L, Faul C, Einsele H, et al. (2003) Ganciclovir-Resistant Cytomegalovirus Disease after Allogeneic Stem Cell Transplantation: Pitfalls of Phenotypic Diagnosis by In Vitro Selection of an UL97 Mutant Strain. Journal of Infectious Diseases 187: 139-143.
44. Prix L, Hamprecht K, Holzhüter B, Handgretinger R, Klingebiel T, et al. (1999) Comprehensive Restriction Analysis of the UL97 Region Allows Early Detection of Ganciclovir-Resistant Human Cytomegalovirus in an Immunocompromised Child. Journal of Infectious Diseases 180: 491-495.
45. Eckle T, Prix L, Jahn G, Klingebiel T, Handgretinger R, et al. (2000) Drug-resistant human cytomegalovirus infection in children after allogeneic stem cell transplantation may have different clinical outcomes. Blood 96: 3286-3289.
46. Frange P, Boutolleau D, Leruez-Ville M, Touzot F, Cros G, et al. (2013) Temporal and Spatial Compartmentalization of Drug-Resistant Cytomegalovirus (CMV) in a Child with CMV Meningoencephalitis: Implications for Sampling in Molecular Diagnosis. Journal of Clinical Microbiology 51: 4266-4269.
47. Monsion B, Froissart R, Michalakis Y, Blanc S (2008) Large bottleneck size in Cauliflower Mosaic Virus populations during host plant colonization. PLoS Pathog 4: e1000174.
48. Nei M, Maruyama T, Chakraborty R (1975) Bottleneck Effect and Genetic Variability in Populations. Evolution 29: 1-10.
49. McWhorter AR, Smith LM, Masters LL, Chan B, Shellam GR, et al. (2013) Natural Killer Cell Dependent Within-Host Competition Arises during Multiple MCMV Infection: Consequences for Viral Transmission and Evolution. PLoS Pathog 9: e1003111.
50. Humar A, Kumar D, Gilbert C, Boivin G (2003) Cytomegalovirus (CMV) Glycoprotein B Genotypes and Response to Antiviral Therapy, in Solid-Organ-Transplant Recipients with CMV Disease. J Infect Dis 188: 581-584.
51. Tanaka K, Numazaki K, Tsutsumi H (2005) Human cytomegalovirus genetic variability in strains isolated from Japanese children during 1983-2003. J Med Virol 76: 356-360.
52. Revello MG, Campanini G, Piralla A, Furione M, Percivalle E, et al. (2008) Molecular epidemiology of primary human cytomegalovirus infection in pregnant women and their families. J Med Virol 80: 1415-1425.
53. Mao ZQ, He R, Sun M, Qi Y, Huang YJ, et al. (2007) The relationship between polymorphisms of HCMV UL144 ORF and clinical manifestations in 73 strains with congenital and/or perinatal HCMV infection. Arch Virol 152: 115-124.
54. Gilbert C, Handfield J, Toma E, Lalonde R, Bergeron MG, et al. (1999) Human cytomegalovirus glycoprotein B genotypes in blood of AIDS patients: lack of association with either the viral DNA load in leukocytes or presence of retinitis. J Med Virol 59: 98-103.
55. Leach CT, Detels R, Hennessey K, Liu Z, isscher BR, et al. (1994) A Longitudinal Study of Cytomegalovirus Infection in Human Immunodeficiency Virus Type I-Seropositive Homosexual

bioRxiv preprint doi: https://doi.org/10.1101/009571. this version posted September 30, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Men: Molecular Epidemiology and Association with Disease Progression. Journal of Infectious Diseases 170: 293-298.

56. 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: 4004-4010.
57. Čičin-Šain L, Podlech J, Messerle M, Reddehase MJ, Koszinowski UH (2005) Frequent Coinfection of Cells Explains Functional In Vivo Complementation between Cytomegalovirus Variants in the Multiply Infected Host. Journal of Virology 79: 9492-9502.
58. Heatley SL, Pietra G, Lin J, Widjaja JM, Harpur CM, et al. (2013) Polymorphism in human cytomegalovirus UL40 impacts on recognition of human leukocyte antigen-E (HLA-E) by natural killer cells. J Biol Chem 288: 8679-8690.
59. Smith C, Gras S, Brennan RM, Bird NL, Valkenburg SA, et al. (2014) Molecular Imprint of Exposure to Naturally Occurring Genetic Variants of Human Cytomegalovirus on the T cell Repertoire. Sci Rep 4.
60. Gibson L, Dooley S, Trzmielina S, Somasundaran M, Fisher D, et al. (2007) Cytomegalovirus (CMV) IE1- and pp65-specific CD8+ T cell responses broaden over time after primary CMV infection in infants. J Infect Dis 195: 1789-1798.
61. Pfeiffer JK, Kirkegaard K (2005) Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. PLoS Pathog 1: e11.
62. Pfeiffer JK, Kirkegaard K (2003) A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity. Proc Natl Acad Sci U S A 100: 7289-7294.

bioRxiv preprint doi: https://doi.org/10.1101/009571. this version posted September 30, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

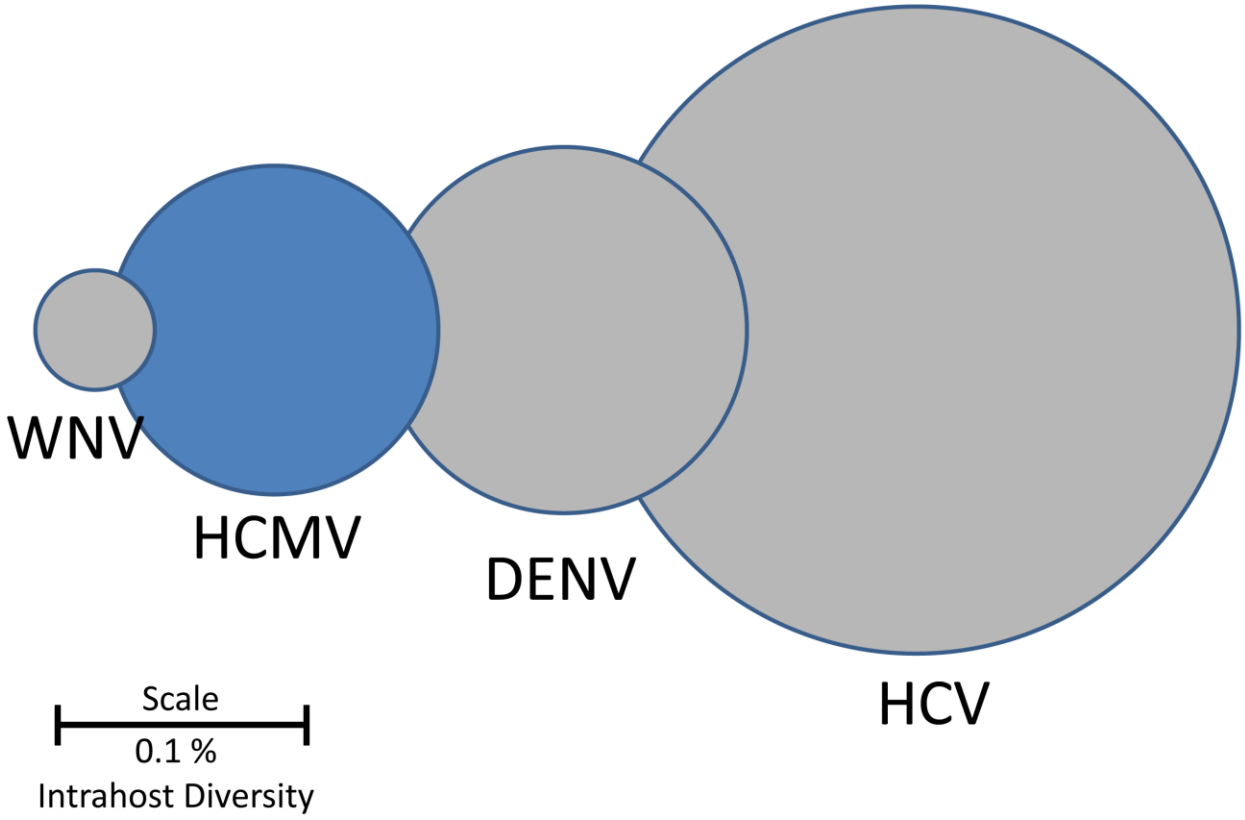

Figure 1: HCMV Intrahost genetic diversity as compared to RNA viruses. Viral intrahost diversities are represented as circles with the diameters, drawn on a log scale, representing reported values of diversity. The genetic diversity of HCMV populations is comparable to those of RNA viruses, an unexpected result given that HCMV is a large dsDNA virus. Values were obtained from [16] and references therein. Abbreviations are as follows: WNV: West Nile Virus, HCMV: human cytomegalovirus, DENV: dengue virus, HCV: hepatitis C virus. Scale bar represents the diameter of a genetic diversity value of 0.1%.

bioRxiv preprint doi: https://doi.org/10.1101/009571. this version posted September 30, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

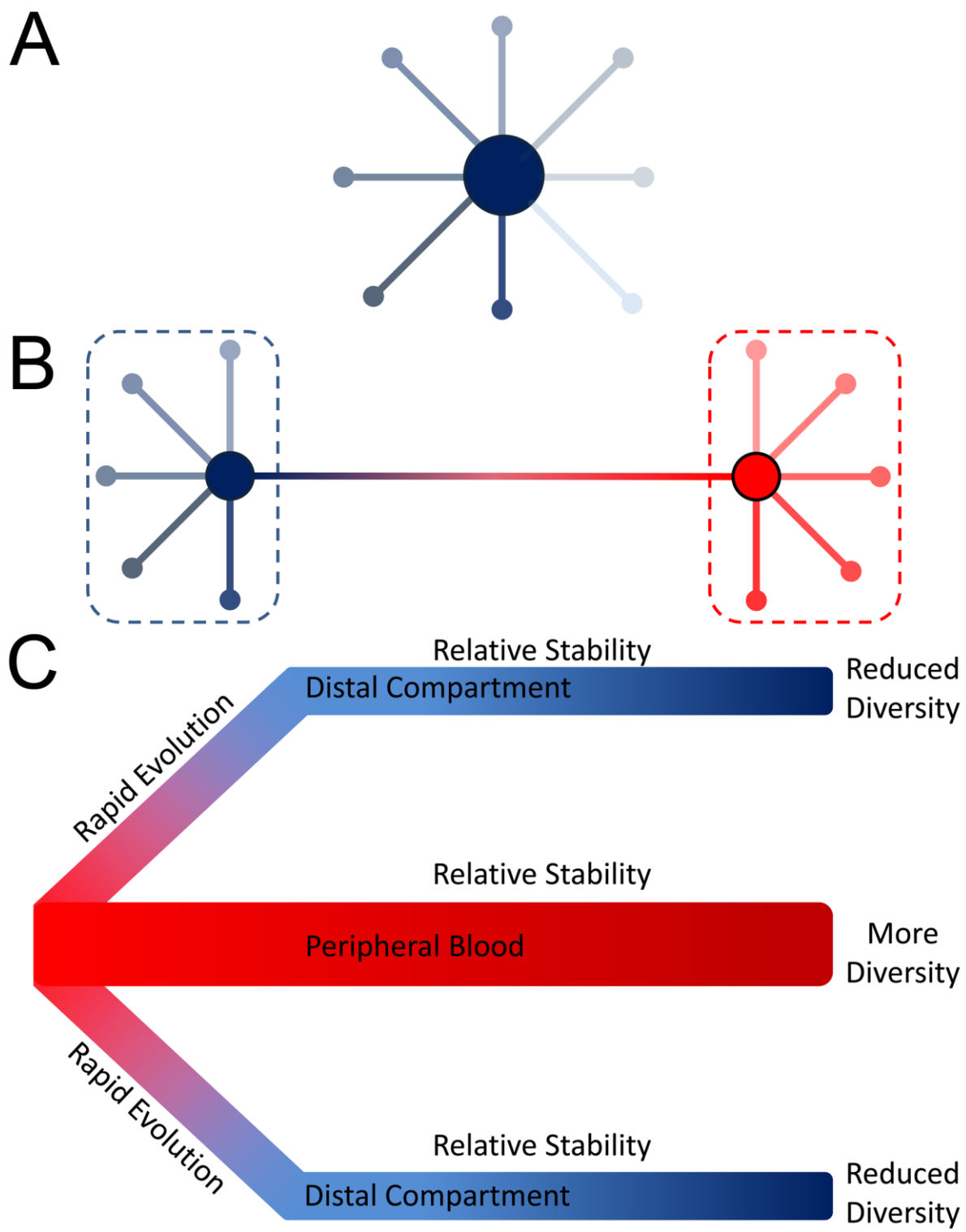

Figure 2: Models of HCMV Diversity and Evolution. Panel A shows an example of the structure of an HCMV intrahost population. Minor variants in the population are distributed around a central sequence. Most variants are only a few mutational steps away from this central sequence. This distribution of sequences is known as a star phylogeny and can be formed by the high levels of replication associated with primary infection, where most de novo mutations will be rare and shared by few members of the population. **Panel B** shows a more complicated population structure which could result from reinfection.

bioRxiv preprint doi: https://doi.org/10.1101/009571. this version posted September 30, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

In this example, the blue and red dashed lines represent strains, with minor variants radiating from the central sequence of each strain. The genetic distance between strains is significantly greater than the genetic distance of minor variants from their respective central sequences. **Panel C** depicts a model of HCMV evolution derived from congenital infection data. In this model, the peripheral blood harbors the most diverse sub-population of HCMV within the body. During dissemination to distal compartments, the viral population can rapidly evolve, either due to natural selection or stochastic mechanisms such as population bottlenecks, leading to populations in the distal tissues or organs being genetically differentiated from the peripheral blood compartment. This phenomenon is known as compartmentalization. The populations in the distal compartments are also less genetically diverse than those of the peripheral blood. Once HCMV has disseminated to compartments or if the virus remains in the peripheral blood, the populations become relatively stable and many mutations remain at similar frequencies over time.