

Alphaherpesvirus Genomics: Past, Present and Future

Chad V. Kuny and Moriah L. Szpara*

Departments of Biology, and Biochemistry and Molecular Biology, Center for Infectious Disease Dynamics, and the Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania, 16802, USA

*moriah@psu.edu, cvk108@psu.edu

DOI: <https://doi.org/10.21775/cimb.042.041>

Abstract

Alphaherpesviruses, as large double-stranded DNA viruses, were long considered to be genetically stable and to exist in a homogeneous state. Recently, the proliferation of high-throughput sequencing (HTS) and bioinformatics analysis has expanded our understanding of herpesvirus genomes and the variations found therein. Recent data indicate that herpesviruses exist as diverse populations, both in culture and *in vivo*, in a manner reminiscent of RNA viruses. In this review, we discuss the past, present, and potential future of alphaherpesvirus genomics, including the technical challenges that face the field. We also review how recent data has enabled genome-wide comparisons of sequence diversity, recombination, allele frequency, and selective pressures, including those introduced by cell culture. While we focus on the human alphaherpesviruses, we draw key insights from related veterinary species and from the beta- and gamma-subfamilies of herpesviruses. Promising technologies and potential future directions for herpesvirus genomics are highlighted as well, including the potential to link viral genetic differences to phenotypic and disease outcomes.

Introduction

Herpesviruses are ubiquitous worldwide, with nine species that infect humans and dozens more infecting other hosts (Davison, 2010). The lifelong nature of herpesvirus infection creates a burden for individual health as well as for public health policy (Gantt et al., 2016; Jansen et al., 2016; Looker et al., 2015a, 2015b). They also have a global impact through their effects on veterinary species involved in food production and those kept as companion animals (Davison, 2010; Loncoman et al., 2017). Within the *Herpesviridae* family, the *alphaherpesvirinae* subfamily is characterized by common genomic features, and epithelial and mucosal sites of active replication (Pellett and Roizman, 2013; Roizman et al., 2013). Most alphaherpesviruses establish lifelong latency in neurons, with a few non-neuronal exceptions such as the *Mardivirus* genus (e.g. Marek's disease virus (MDV) or *gallid alphaherpesvirus 2*). The human alphaherpesviruses consist of herpes simplex virus 1 and 2 (HSV-1,2 or *human herpesvirus 1,2*, HHV-1,2) and varicella zoster virus (VZV or HHV-3) (Davison, 2010). These viruses, as with all herpesviruses, have large dsDNA genomes that have presented challenges for genomic characterization, particularly in comparison to other viral families (Koonin et al., 2015).

Herpesviruses have co-existed with humanity for as long as we have written records. Ancient Egyptian hieroglyphics record the presence of herpetic lesions, implying that these viruses have been a notable health issue for millennia (Pellett and Roizman, 2013). However, detailed scientific studies of these viruses, especially those investigating the entirety of the viral genome, have occurred in quite recent times (Davison, 2010; Depledge et al., 2018a; Loncoman et al., 2017; Renner and Szpara, 2018). The progression of herpesvirus genomics studies has been greatly expanded by the advancement of DNA sequencing technologies. Sanger sequencing enabled the study of individual viral genes, or with painstaking effort, a composite full-length viral genome (Klupp et al., 2004; McGeoch et al., 1985, 1988). With the advent of high throughput sequencing (HTS), entire viral genomes can now be assembled from one sample and its sequencing "run", and in a matter of days rather than years (Greninger et al., 2018; Parsons et al., 2015). HTS approaches also yield

a depth of sequence information that provides insight into the genetic diversity within a viral sample. This relative wealth of data has revealed a substantial amount of genetic diversity between strains of any given alphaherpesvirus, as well as genetic diversity indicative of a population of viruses within a single sample. The effects that this level of sequence diversity may have on virus biology or clinical outcomes remain an area of active research (Houldcroft et al., 2017; Loncoman et al., 2017; Renner and Szpara, 2018).

Despite the recent expansion of alphaherpesvirus genomics, substantial challenges remain in the field. Alphaherpesvirus genomes have multiple features that are difficult to analyze with current short-read HTS technology, including highly repetitive areas, regions of high G+C content, and even the presence of genomic isomers. Handling these challenges is a key focus of the future of herpesvirus genomics. In this review, we will describe the progression of alphaherpesvirus genomics, including historical context, current areas of research, and potential future directions for the field.

The inherent (in)stability of herpesvirus genomes

A prevailing opinion among virologists is that DNA viruses are inherently stable and RNA viruses are inherently variable (Sanjuan et al., 2010). Indeed, RNA viruses are often described as existing as a quasispecies swarm of varying genomes, rather than as a defined genetic species (Andino and Domingo, 2015). Mechanistically, this view stems from the lack of error correction in most RNA-dependent RNA polymerases (Sanjuán and Domingo-Calap, 2016). By contrast, most DNA viruses have high fidelity polymerases with error correction, and have lower reported mutational rates (Hall and Almy, 1982; Drake and Hwang, 2005). Early studies of the mutation rate of HSV-1 examined single genes, and detected mutation rates on the order of 1×10^{-7} or 1×10^{-8} mutations per base per infectious cycle (Hall and Almy, 1982; Drake and Hwang, 2005). The rates identified in such studies are those that are then subsequently quoted in comparisons between RNA and DNA viruses (Drake, 1991; Sanjuan et al., 2010; Sanjuán and Domingo-Calap, 2016). While those mutation rates fit well with older restriction-fragment length polymorphism (RFLP) comparisons of herpesvirus genomes, these mutation rates fail to explain the ease with which

herpesvirus variants are selected or revealed under strong selective pressures. For example, analysis of multiple HSV-1 and HSV-2 strains revealed the presence of drug-resistance mutations in 1 out of 10^3 to 10^4 PFU in the overall virus population (Sarisky et al., 2000). This result suggests that alphaherpesviruses, even in culture, exist as populations of diverse genomes - a phenomenon otherwise known as standing variation (Firth et al., 2010; Renzette et al., 2014). An alternative explanation is that the rate of evolution observed under these laboratory conditions is substantially different than natural conditions, although the frequency with which drug-resistant viruses are detected *in vivo* suggests otherwise (Burrell et al., 2010; Sauerbrei et al., 2010).

Multiple lines of evidence support the idea that alphaherpesviruses exist in populations with standing variation. Genome-wide HTS comparisons of HSV-1 have revealed approximately 3-4% nucleotide variation genome-wide, with differences of 1-2% observed between plaque-picked subclones of a given strain (Parsons et al., 2015; Szpara et al., 2014; Bowen et al., 2016). Additionally, a study that inferred mutation rates from phylogenetic history found a higher than expected frequency of observed mutations in herpesvirus populations (Firth et al., 2010). There is also experimental evidence to support a hypothesis that genetic diversity is inherently beneficial to herpesviruses. In an investigation of Muller's ratchet - the hypothesis that small asexual populations will accumulate deleterious mutations - Jaramillo *et al.* subjected ten individual sub-clones of HSV-1 to sequential plaque-to-plaque transfers (Jaramillo et al., 2013). As a result of the extreme genetic bottlenecks induced by this experimental approach, two clonal lineages were completely lost, with the remaining clones exhibiting an attenuation of host mortality after intracerebral inoculation into mice (Jaramillo et al., 2013). Genetic analysis of these clones revealed a mutation frequency of 3.6×10^{-4} substitutions per base per plaque transfer, a far higher rate than those described in single gene studies (Hall and Almy, 1982; Drake and Hwang, 2005; Drake, 1991; Jaramillo et al., 2013). Additionally, while not the focus of this review, studies of the beta-herpesvirus human cytomegalovirus (HCMV) have described the existence and expansion of standing variation within immunocompromised hosts (Gorzer et al., 2010; Sijmons et al., 2015; Houldcroft et al., 2016; Hage et al., 2017). On the whole,

while studies of the mutation rate of the herpesvirus polymerases have been useful, the breadth of genetic diversity within herpesviruses is substantial and has relevance to both laboratory and clinical settings (Houldcroft et al., 2017; Loncoman et al., 2017; Renner and Szpara, 2018). With the advent of advanced sequencing technologies, researchers are beginning to appreciate the effects that genetic diversity can have on disease outcomes, drug resistance, and other viral phenotypes of interest.

Comparative genomics reveals the true diversity of herpesviruses

Two seminal examples of VZV genome analysis by Peters *et al.* 2006 and Tyler *et al.* 2007 predate the era of HTS, but nonetheless set the tone for many later analyses (Peters et al., 2006; Tyler et al., 2007). We applied many of the same comparative genomics analyses laid out in these papers in our first analyses of HSV-1 and pseudorabies virus (PRV) genomes (Szpara et al., 2010, 2011). These included phylogenetic and recombination analyses, as well as comparisons of coding diversity, tandem repeats, and specific genetic loci (Peters et al., 2006; Szpara et al., 2014; Tyler et al., 2007). Many subsequent studies have followed the same approach, albeit for different alphaherpesvirus species or for interesting subsets of viruses within a species (see for example (Depledge et al., 2014a; Newman et al., 2015; Bryant et al., 2018; Lewin et al., 2018)). Studies using these methods have illuminated the evolutionary history of several alphaherpesvirus species (Zell et al., 2012; Norberg et al., 2015; Kolb et al., 2015; Browning et al., 2016; Vaz et al., 2016a; Burrell et al., 2017; Koelle et al., 2017; Trimpert et al., 2017), demonstrated that recombination can occur between and within a single species (Loncoman et al., 2017; Lewin et al., 2018; Zell et al., 2012; Norberg et al., 2015; Kolb et al., 2015; Burrell et al., 2017; Koelle et al., 2017; Vaz et al., 2016b; Kolb et al., 2017) (as detailed below in "Mechanisms that drive genetic diversity"), and shown that vaccine strains can regenerate virulent viruses if given the opportunity (Lee et al., 2012; Ye et al., 2016). What remains under-represented in many HTS studies is a connection of the comparative genomics data to the measurement of biological phenotypes. These phenotypes could include classic virological measures, such as plaque morphology or replication fitness (Bowen et al., 2018; Parsons et al., 2015), or *in vivo* measures such as neuroinvasion and mortality rate (Bryant et al., 2018;

Pandey et al., 2017). The alphaherpesvirus literature reflects decades of research on the *in vitro* and *in vivo* phenotypic effects of individual gene deletions, mutations, or over-expression (Pellett and Roizman, 2013; Roizman et al., 2013). The next frontier for comparative genomics is to integrate these prior data with the insights from comparative genomics, as detailed below (see "Future Frontiers").

Mechanisms that drive genetic diversity

As we gain insight into the level of genetic diversity within the *alphaherpesvirinae* subfamily, one outstanding question is how such diversity is generated. One straightforward source of genetic diversity is DNA replication error by the viral polymerase (Hall and Almy, 1982; Drake and Hwang, 2005). However, it is unclear how much of a role DNA polymerase error plays in the generation of alphaherpesvirus genetic diversity. Studies of the error rate of herpesvirus polymerases have indicated a very low mutation rate, on the order of 1×10^{-7} or 1×10^{-8} mutations per base per infectious cycle, which would imply that many rounds of genomic replication are necessary if polymerase error were the sole or primary source of genetic diversity (Hall and Almy, 1982; Drake and Hwang, 2005). However, those studies were also done in the context of a single coding region, which may well underestimate polymerase error rates (Brown, 2004). Polymerases are known to have higher error rates in contexts of high G +C content or repetitive regions, and these sequence contexts are over-represented in alphaherpesvirus genomes with high inter-strain diversity, such as HSV-1 (Figure 1). In contrast the smaller VZV genome has an overall lower rate of inter-strain diversity. The VZV genome has reduced intergenic regions and ~50% G+C content as compared to ~68% G+C content for HSV-1 (peaking at ~80% average G+C in highly repetitive regions) (Figure 1) (Peters et al., 2006; Tyler et al., 2007; Szpara et al., 2011; Sijmons et al., 2015).

Beyond polymerase error, it is now clear that mutation and evolution in herpesviruses result not only from base substitutions, but also from recombination between strains of a given species, and to a less frequent extent, recombination between species (Loncoman et al., 2017; Norberg, 2010; Renner and Szpara, 2018). As with segmented RNA virus reassortment, recombination

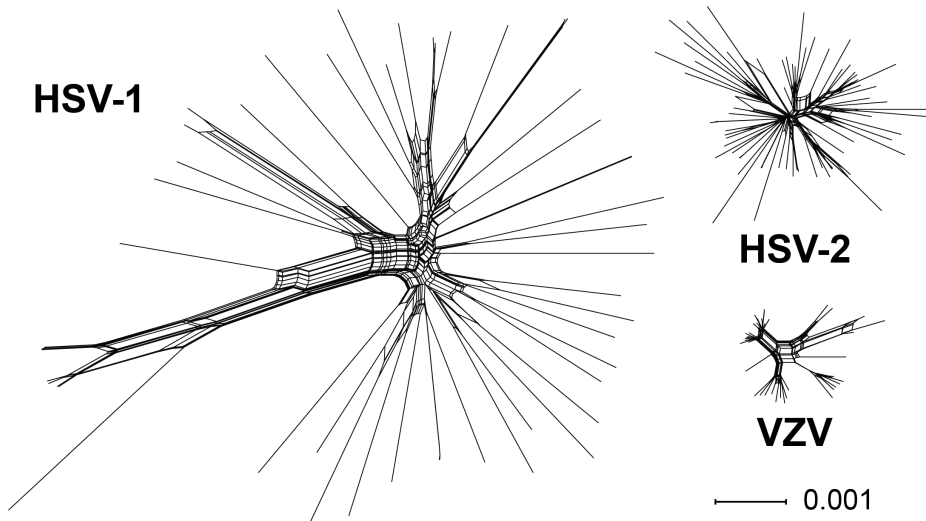


Figure 1. Comparison of the known genetic diversity within the human alphaherpesviruses. We generated SplitsTree diagrams (Huson, 1998) with recently published analyses of known genetic diversity in the three human alphaherpesviruses (Akhtar et al., 2019; Shipley et al., 2019; Zell et al., 2012). Each virus' tree is presented at the same scale, so that the size of the tree is proportional to the observed genetic diversity for each virus. The HSV-1 diagram includes 51 strains (Shipley et al., 2019), HSV-2 includes 68 strains (Akhtar et al., 2019), and VZV includes 42 strains (Zell et al., 2012).

in herpesviruses can provide a strong force for evolutionary shifts. Studies of lab-generated recombinants of HSV-1 have revealed a bias towards recombination breakpoints occurring in repetitive regions, areas of locally high G+C content, and intergenic regions (Lee et al., 2015). However, most studies of recombination within the alphaherpesviruses have focused on naturally circulating variants, and have inferred recombination and phylogenetic relationships for historical sites of recombination from a comparison of extant strains (Burrell et al., 2017; Koelle et al., 2017; Loncoman et al., 2017; Norberg et al., 2015). For example, as more strains of VZV have been sequenced, the number of phylogenetic clades have expanded and there is now evidence for ancient, inter-clade recombination in addition to modern recombination between strains (Norberg et al., 2015; Zell et al., 2012). Within circulating populations of

HSV-1 and HSV-2, there is evidence of rampant recombination among strains within each species, although it is unknown if the observed recombination is historical or recent (Szpara et al., 2014; Kolb et al., 2015; Norberg, 2010; Norberg et al., 2007). Recently, two separate groups have found evidence for inter-species recombination between HSV-1 and HSV-2 (Burrell et al., 2015, 2017; Casto et al., 2019; Koelle et al., 2017). These findings are based on the presence of several regions in the HSV-2 genome that have a high degree of sequence similarity to extant HSV-1 genomes, whereas the remainder of the HSV-2 genome is more closely related to a chimpanzee herpesvirus (ChHV) than to HSV-1 (Severini et al., 2013; Wertheim et al., 2014). For the non-human alphaherpesviruses, multiple examples of recombination both *in vivo* and *in vitro* have been reported, with intriguing results. Devlin and colleagues found that a virulent strain of Gallid herpesvirus 1 (infectious laryngotracheitis virus or ILTV) found in an agricultural setting was a spontaneous recombinant from two live-attenuated vaccine strains (Lee et al., 2012). As reviewed recently by Loncoman *et al.* (Loncoman et al., 2017), recombination of equine herpesviruses has been noted within-species (intrastrain, e.g. between EHV-1 isolates (Bryant et al., 2018; Kolb et al., 2017; Vaz et al., 2016b)) and between species (inter-strain, e.g. between EHV-1 and EHV-4 (Bryant et al., 2018)), and also within-species for feline herpesvirus (Kolb et al., 2017; Lewin et al., 2018), PRV (Kolb et al., 2017; Szpara et al., 2011; Ye et al., 2016), and MDV (Loncoman et al., 2017; Trimpert et al., 2017). The next challenge for understanding alphaherpesvirus recombination will be to determine where and how often this occurs *in vivo*. Since recombination requires the co-occurrence of two distinct viral lineages in one infected cell in the host, this may be rare and difficult to detect in clinical or field settings.

Other contributions to functional diversity

While single-nucleotide changes, recombination, and even horizontal gene transfer are generally accepted mechanisms for genetic variation within herpesviruses, data indicate that other mechanisms may also contribute to genetic diversity. These mechanisms include ribosome slippage and novel coding and non-coding RNAs revealed by transcriptome analysis or ribosome profiling. These genomic outputs cannot be predicted solely by analysis of

genomic DNA, and instead require approaches that integrate transcriptional and/or translational data in combination with a matched DNA genome comparison.

Several studies have applied sensitive RNA-sequencing-based approaches to demonstrate an expanded range of transcriptional impacts of HSV-1 infection on host cells. A recent series of papers used directional and nascent RNA-sequencing, in combination with ribosome profiling, to reveal previously undetected effects on the host transcriptome during HSV-1 replication (Rutkowski et al., 2015; Wyler et al., 2017; Hennig et al., 2018). These included the disruption of transcription termination (DoTT) or transcriptional read-through, which leads to the production of many run-on RNA transcripts (Rutkowski et al., 2015; Hennig et al., 2018). It also included a large number of antisense transcripts from the host genome (Wyler et al., 2017). More recently, these analyses have been applied to several alphaherpesvirus genomes, revealing many non-canonical transcripts and novel open reading frames (ORFs) expressed by these viruses, along with previously undetected transcript extensions, truncations, and splicing events (Tombácz et al., 2017a; Prazsák et al., 2018; Tombácz et al., 2019; Depledge et al., 2019; Whisnant et al., 2020). Future studies will need to explore the expression and potential function(s) of these transcripts in different cellular and *in vivo* models of infection, and their conservation across other herpesviruses (Arias et al., 2014; Tirosh et al., 2015).

Ribosome frameshifting is a necessary aspect of translation for retroviruses like HIV, where production of the nucleocapsid and polymerase proteins are directed by the same RNA transcript and ribosome frameshifting is required to obtain both products. While it occurs at a lower frequency, ribosome frameshifting has been demonstrated on thymidine kinase (TK) transcripts in HSV-1 (Griffiths, 2011; Griffiths et al., 2003; Pan and Coen, 2012). This can be detected in genomes with a homopolymer-based frameshift in the TK gene, which often arise in response to treatment with the antiviral drug acyclovir (Burrell et al., 2010; Griffiths, 2011; Sauerbrei et al., 2010). While a direct translation of the RNA encoded by such altered genes would lead to non-functional proteins, ribosomal frameshifting enables a low level of functional protein production in

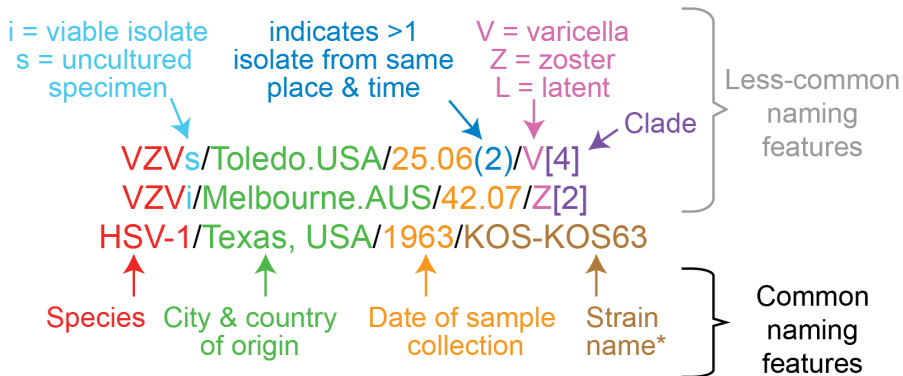
these drug-resistant viruses (Griffiths, 2011; Griffiths et al., 2003; Pan and Coen, 2012).

RNA editing is a mechanism by which viral transcripts may be recoded at the single-nucleotide level to differ from the existing viral genome. While host-defensive or therapeutic CRISPR-based RNA editing is a destructive process that may be used against herpesviruses (Nakaya et al., 2016; Oh et al., 2019; Suspene et al., 2011), there is also the potential for other outcomes of RNA editing. Potential pro-viral effects of RNA editing of selected transcripts have been explored for the gammaherpesviruses Epstein-Barr Virus, EBV (Lin et al., 2013) and Kaposi's Sarcoma-Associated Herpesvirus, KSHV (Arias et al., 2014). The only example of RNA editing thus far for alphaherpesviruses involves editing of a long non-coding RNA of MDV, which the authors propose could contribute to the viral lytic cycle via a role in regulating host innate immune defenses (Figueroa et al., 2016). Further analyses of alphaherpesvirus transcriptomes may reveal more such examples and enable further testing of their potential functional effects (Tombácz et al., 2017a; Prazsák et al., 2018; Tombácz et al., 2019; Depledge et al., 2019; Whisnant et al., 2020).

Clear nomenclature to document viral strains and genome sources

As increasing numbers of alphaherpesvirus isolates of a given species are studied at the genomic level, it has become increasingly important to document the different sources of material used to obtain viral genomes. While descriptive terms such as "clinical isolate" and "laboratory strain" are often used to denote either low or high numbers of cell culture passages for a given viral isolate, these terms are interpreted variably by different research groups (Kuhn et al., 2013; Wilkinson et al., 2015). There are no standards for how many times a viral sample can be passaged in culture before it is no longer considered a "clinical isolate", nor are there standards as to whether a viral population should be plaque purified before calling it a strain. The introduction of sensitive DNA isolation and amplification techniques has also enabled whole viral genomes to be collected and sequenced directly from hosts, avoiding cell culture entirely (Greninger et al., 2018; Depledge et al., 2011; Watson et al., 2013; Johnston et al., 2017a; Shipley et al., 2018, 2020). It is important that researchers

A. Recommended nomenclature usage based on VZV:



B. Examples of variations in nomenclature usage:

HSV-1/0116209/India/2011
 HSV-2/US/BID-G19094/89-390/1989
 (VZV) Var/Cli/Ves/UK/321/2015

Figure 2. Viral nomenclature presents an opportunity to record crucial details about strain origin and history. (A) Common aspects of viral strain identification are shown here, based on the 2008 recommendations of VZV Nomenclature Meeting (Breuer et al., 2010). The recommended nomenclature includes the following data for each strain (moving from left to right in the diagram): viral status as a cultured isolate or an uncultured sequence, the geographic city and country of origin, collection date, whether or not any related sequences were collected, the disease presentation (e.g. varicella, zoster or latent), and the strain clade based on phylogenetic clustering or single-nucleotide polymorphism (SNP) profile (Breuer et al., 2010). The nomenclature components are color-coded consistently throughout the diagram. In a review article recommending this approach (Renner and Szpara, 2018), we suggested the addition of strain name to this nomenclature, in keeping with recommended practices for RNA virus nomenclature (Kuhn et al., 2013). **(B)** However, the current usage of this nomenclature is limited, with variable applications whose content is not always straightforward to interpret. A selection of recent variations in nomenclature usage are shown from the following GenBank records and publications: KJ847330 (Bondre et al., 2016), KR135321 (Newman et al., 2015), and MG764307 (Depledge et al., 2018b). A more consistent application of the recommended viral strain nomenclature would provide clear benefits for sample data retrieval and the accurate linking of sequence data with published results. Figure adapted from Breuer *et al.* 2008 (Breuer et al., 2010).

generating new genomic data, and those using sequences from the GenBank databases for purely computational studies, appreciate these distinctions. The VZV community has recommended a viral naming system that is similar to those used for RNA viruses or bacteria to help record and clarify these details for each new genome (Figure 2A) (Breuer et al., 2010). This standard emphasizes defining not only a name for each strain, but also its geographic origin, collection date, disease type (e.g. varicella, zoster or latent), and its status as a cultured isolate or an uncultured sequence (Breuer et al., 2010). Whenever possible the strain clade, based on phylogenetic clustering or single-nucleotide polymorphism (SNP) profile, is included as well. While not yet widely or consistently adopted (Figure 2B), this nomenclature approach would clearly be beneficial if applied to other alphaherpesviruses as well. This would improve data-accessibility across the primary literature and sequence databases, as well as enabling easier comparison of data across research groups.

Genomes: consensus vs. population

Most prior studies that catalog herpesvirus diversity have produced a single, consensus genome for each new sample. The consensus genome is representative of the most common allele or nucleotide at each position in the genome (Figure 3). In the simplest scenario, the consensus genome from a sample is built from the most common nucleotide at each position throughout the genome (Figure 3A). However, this scenario is complicated by evidence that herpesvirus isolates generally exist as populations that contain a degree of sequence diversity. Because more than one nucleotide may exist at a given location across the viral population in a sample, the creation of a single consensus genome necessarily removes the details of underlying components of the viral genetic population. The consensus genome for a given sample may also be an amalgamation of frequent but non-contiguous alleles (Figure 3B). Genetic variants or alleles that exist at a sub-consensus level are often referred to as minor variants. Depending on the intended purpose of a study, an analysis of the presence and identity of minor variants may be crucial to a proper interpretation of the data, or that information may be viewed as a luxury. For example, studies of transmission or evolution of HSV-1 (Greninger et al., 2018; Pandey et al., 2017; Shipley et al., 2018; Lassalle et al., 2020), HSV-2 (Akhtar

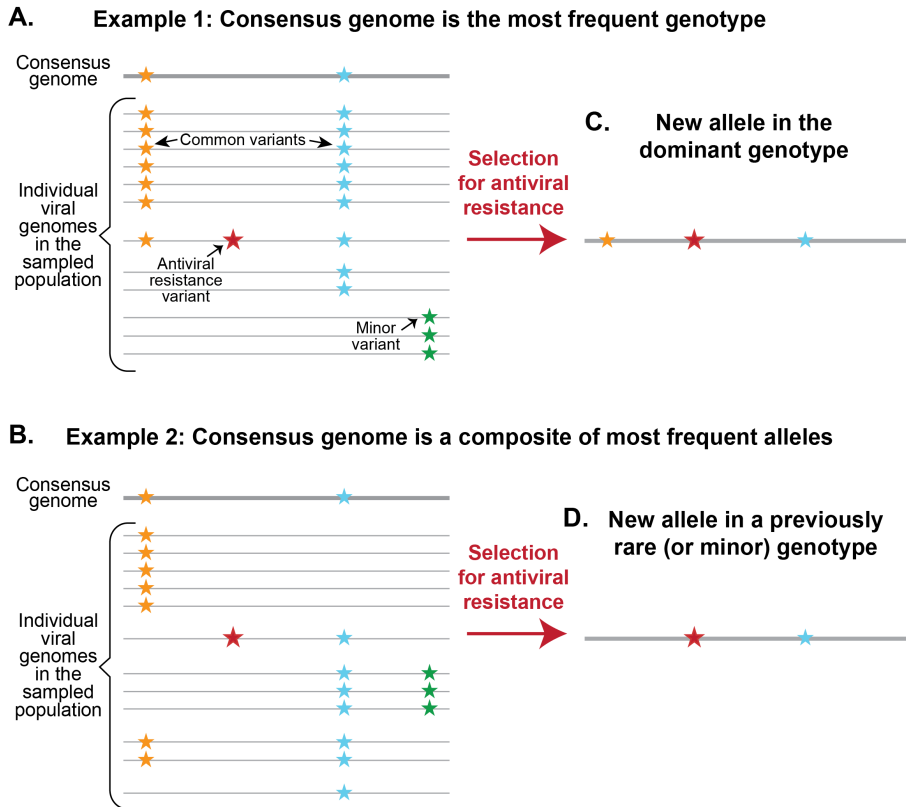


Figure 3. A sample of virus contains a population of viral genomes, which includes both common and rare or low-frequency (minor) genetic variants. A viral consensus genome represents the most common allele or nucleotide every position in the genome. In this diagram, lines represent individual genomes, with alleles or loci that vary between genomes denoted as stars. The thicker line represents the consensus genome that results from deep sequencing and assembly of the most common (or most frequently observed) alleles. **(A)** In example 1, the consensus genome is the most common or frequent genotype present in the viral sample. **(B)** In example 2, the consensus genome contains all of the most common alleles, but this amalgamation of alleles does not exist at high frequency in the viral sample. **(C-D)** In either case, selection pressures such as the application of an antiviral drug can lead to a genome that carries a rare drug-resistance variant (denoted by the red star) being positively selected. The drug-resistant genotype may then become the most commonly observed allele. Depending on the genomic context of the drug-resistance allele, this allele may appear in the same genetic context as the previously dominant genotype **(C)**, or it may appear alongside other variants that were previously rare **(D)**. Figure adapted from Renner and Szpara 2018 (Renner and Szpara, 2018).

et al., 2019; Minaya et al., 2017a), and VZV (Depledge et al., 2016a, 2016b, 2018b) have compared minor variants detected in samples of related virus populations to draw their conclusions. From a clinical perspective, minor variants may include drug-resistant variants in the viral population that could flourish following treatment (Figure 3CD) (Burrel et al., 2010; Houldcroft et al., 2017; Sauerbrei et al., 2010). Phylogenetic studies, by contrast, would have little use for analyzing the minor variants in their samples (Zell et al., 2012; Kolb et al., 2017; Johnston et al., 2017a; Pfaff et al., 2016). So, when designing a genetic study of herpesviruses, careful consideration must be given to the analysis of minor variants and the depth of sequence coverage needed to do so.

Another consideration that must be made when analyzing consensus sequences from deep sequencing data relates to the large size of alphaherpesvirus genomes. Since the sequence read length of commonly used HTS or deep sequencing approaches are substantially shorter than alphaherpesvirus genomes (e.g. ~300 bases vs hundreds of kilobases), it is difficult to determine which genetic variants coexist within one intact genome. If each variant is the most common at its location, it will be included as an amalgamation of several genotypes as the consensus genome (see Figure 3). For smaller viruses, researchers have the ability to clone and sequence individual genomes, which has enabled the measurement of viral genotypes in a given population and the development of software to infer likely haplotypes from the HTS data (Lou et al., 2013; Töpfer et al., 2014; Jayasundara et al., 2015). Barcoded methods offer the potential to improve the linking of variant alleles into haplotypes (Lauring and Andino, 2011; Keys et al., 2015), but these have not yet been widely applied to herpesviruses. Trimpert *et al* recently applied a combination of HTS-based assessment, with bacterial artificial chromosome (BAC)-based genome capture, to validate MDV population diversity in a series of polymerase mutants (Trimpert et al., 2019). This combination enabled them to confirm the co-location of minor variants observed by HTS, in single BAC-cloned viral genomes (Trimpert et al., 2019). Other technologies such as nanopore-based (MinION®, Oxford Nanopore) or single-molecule real-time sequencing (SMRT®), Pacific Biosystems (PacBio)) offer the promise of longer

read lengths and the resulting ability to link minor alleles, although they have not yet been widely applied to analyze minor variants in herpesvirus populations (Tombácz et al., 2017a; Prazsák et al., 2018; Tombácz et al., 2019; Depledge et al., 2019; Whisnant et al., 2020). However, both long-read technologies currently exhibit a degree of sequencing error that is sufficient to limit their useful application at this point in their development. Thus short-read HTS platforms such as Illumina® (e.g. MiSeq or HiSeq) remain the most common choice for herpesvirus genomics, where the technology is sufficient to identify the location and prevalence of minor variants, but not to evaluate their co-occurrence on single genomes (Hage et al., 2017; Shipley et al., 2018; Depledge et al., 2016b; Pandey et al., 2016). Newer hybrid sequencing approaches may help to bridge this gap, by employing both the sensitivity of Illumina short-read HTS, and the longer scaffold accuracy of long-read Nanopore or PacBio HTS.

The path to obtaining a viral genome

Most herpesvirus sequencing projects, whether from laboratory experiments or clinical samples, begin with infection of a cell line to generate a sufficient level of DNA for the preparation of sequencing libraries. This approach enables the generation of large quantities of DNA, including the amplification of DNA from a potentially scarce clinical sample. However, growing viruses in cell culture is inherently different from how the virus would replicate *in vivo*, with the potential to introduce unexpected genetic bottlenecks or selective pressures (Parsons et al., 2015; Wilkinson et al., 2015; Kuny et al., 2020). The potential for cell culture-induced bias in sequencing results has led to the development of oligonucleotide-based enrichment methods, with the goal of sequencing uncultured samples to the same quality as cultured stocks (Greninger et al., 2018; Depledge et al., 2011; Johnston et al., 2017a; Shipley et al., 2018, 2020). This "oligo-enrichment" approach uses synthetic RNA or DNA probes, also known as baits, that are designed to hybridize with sparse amounts of the targeted viral genome(s) using sequence complementarity. These are mixed into solution with the sample of interest. Once they have hybridized with viral genetic material, the baits can be isolated from the mixed sample by virtue of an attached tag, such as biotin. The sample that results from this isolation is greatly

enriched for the targeted DNA and can be amplified and sequenced using standard HTS approaches. Oligo-enrichment has enabled the sequencing of alphaherpesviruses directly from samples such as skin swabs, saliva, blood, vesicle fluid, and tissue samples, among others (Greninger et al., 2018; Depledge et al., 2011; Johnston et al., 2017a; Shipley et al., 2018). The sensitivity of viral genome capture with this method has been used to gain new insights into viral population diversity in infected humans. This has included tracking changes at the consensus genome level in multiple reactivation episodes (Shipley et al., 2018; Greninger et al., 2018; Casto et al., 2019; Depledge et al., 2014b), or during transmission between human hosts (Depledge et al., 2016b; Shipley et al., 2019), and examining how specific allelic variants (i.e. minor variants) change in frequency over time within individual hosts (Minaya et al., 2017b; Shipley et al., 2018).

Once viral genomic DNA is prepared for sequencing, the next step for a researcher is to decide which platform they will choose to generate sequence data. The most common approach is to use short-read HTS platforms, such as Illumina®. A detailed examination of how HTS data is processed into full-length consensus genomes and/or minor variants has been described in depth elsewhere (Houldcroft et al., 2017; Posada-Céspedes et al., 2016). The basic approach is for the short sequencing reads to be stitched, via bits of overlapping sequence, into longer and longer contiguous pieces that eventually form a new consensus genome. In a second quality-control step, individual sequencing reads are then aligned (or "mapped") back to their matching locations on the newly constructed consensus genome, usually with a great number of sequencing reads present per nucleotide (i.e. the sequencing or coverage depth). When sequencing coverage is deep (e.g. >300-fold coverage), genomes can be created with greater confidence than with lesser coverage. Deep sequence coverage also allows low-frequency minor variants to be identified with greater confidence. Publications generally only report minor variants above a certain minimum frequency threshold, such as 2% (Greninger et al., 2018; Hage et al., 2017; Shipley et al., 2018; Depledge et al., 2014b). It is important to appreciate that with 100X coverage, a 2% threshold for minor variant detection would mean that only 2 sequencing reads were sufficient to identify (or "call")

the minor variant. For this reason, we recommend a sequencing depth far exceeding 100-fold for the confident and accurate detection of minor variants. Greater coverage depth means that more sequence reads must support the minor variant for it to be detected or called, thus minimizing the chance of an incorrectly identified minor variant. The cost of high coverage depth is that fewer viral strain genomes can be assembled from a given sequencing "run" (e.g. one iteration of operating an Illumina MiSeq). Again, this is where the priority of the researcher factors into decisions made in the sequencing process. If the priority is to sequence as many genomes as possible in one sequencing run, the limitations of existing technology mean that such genomes will have lower coverage than a sequencing run that involves only a limited number of unique samples. However research that is focused solely on species identification or serotype determination may benefit from increased efficiency and cost-savings by foregoing coverage depth and the detection of minor variants (Greninger et al., 2018; Zell et al., 2012; Vaz et al., 2016a; Pfaff et al., 2016).

Since the analysis of sequenced genomes often depends on how accurately individual sequencing reads can be overlapped to create the scaffold of a full-length genome, repetitive regions present a distinct problem (Figure 4) (Treangen and Salzberg, 2011). If a given sequencing read consists entirely of a perfectly repeating sequence, it may be unclear how the read overlaps with its neighbors, i.e. how a long a given repeating array of sequences truly is. Many alphaherpesvirus genomes have multiple extremely long repetitive arrays, such as the classic "reiterations" recognized in early studies of HSV-1 and VZV (Davison and Wilkie, 1981; Moss et al., 1981). Length variations in these tandem repeats have been recognized since the era of restriction fragment length polymorphism (RFLP) analyses, and these fluctuations further confound the true length determination of these repeats (Moss et al., 1981; Simon et al., 1989; Sakaoka et al., 1994). In HTS genome assemblies, the issue of uncertain overlaps in repetitive sequence arrays can cause the termination of *de novo* assembled fragments, creating gaps in the resulting consensus genome. It may also confound the ability to map sequence reads back to repeated regions in alignment-based methods of genome analysis (Parsons et al., 2015; Treangen and Salzberg, 2011). In addition to having large structural repeats at internal

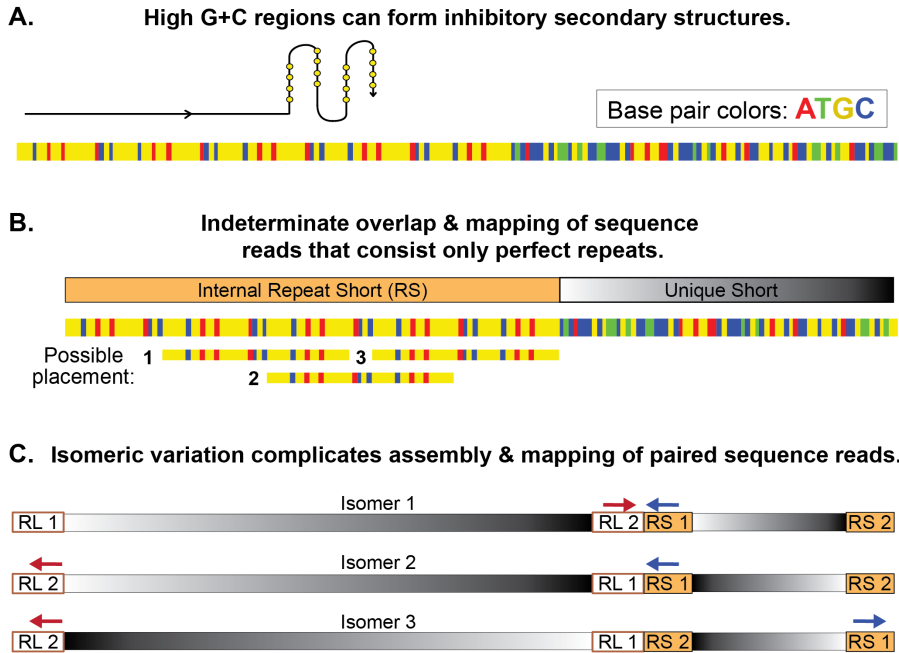


Figure 4. Repetitive and/or G+C rich regions complicate the process of genome sequencing and assembly. (A) Challenges in the physical sequencing of G+C-rich regions and/or areas containing DNA secondary structures such as the stem-loop at the origin of replication (Ori) can lead to a lower quantity of raw sequence-read data being produced in during high-throughput sequencing (HTS). (B) During *de novo* assembly or alignment of viral genome sequence data, challenges arise from ambiguities in the overlap and/or mapping of sequence reads that contain tandem repeats, homopolymers, or adjacent mixtures of these motifs. (C) Alphaherpesvirus genomes generally have two unique regions (termed unique long and unique short; shown as gradients of gray), each of which are flanked by repeats (known as repeat long (RL) or repeat (RS); shown as white and orange boxes). RL and RS exist in two copies, one terminal and one internal and inverted, on each genome. Isomers of these genomes - which appear to have either the left or right unique and repeated regions in the reverse orientation - exist in viral populations (Mahiet et al., 2012; Roizman, 1979). Sequence read mapping issues can result from ambiguities in mapping or placement of paired-end sequences that span the isomerization boundary of the long and short regions of alphaherpesvirus genomes (gradients of gray denote orientation of the unique regions). The combination of decreased sequence quantity (A) and ambiguities in sequence read mapping (B) combine to create regions of low-coverage and gaps in the final consensus genome produced by HTS (C). This phenomenon is frequently observed in the terminal and internal repeats of HSV-1 and HSV-2 consensus genomes (Greninger et al., 2018; Parsons et al., 2015; Szpara et al., 2011; Newman et al., 2015; Norberg et al., 2011).

and terminal regions, and the classically-defined large reiterations, many alphaherpesviruses are rich in short tandem repeated sequences (also known as short sequence repeats or SSRs, and variable number tandem repeats, or VNTRs) (Szpara et al., 2011). The G+C content tends to be even higher within these repeats (Szpara et al., 2011, 2014). This leads to secondary structures such as stem-loops and G-quadruplexes that decrease the overall yield of sequence data in these areas and further complicates the assembly of these high G+C repetitive regions (Figure 4) (Szpara et al., 2011; Parsons et al., 2015; Artusi et al., 2015; Biswas et al., 2016).

Alphaherpesviruses often have latency- or neurovirulence-associated genes encoded in the large structural repeated regions of the genome (e.g. LAT and ICP34.5 in HSV-1), in addition to the key transcriptional regulator ICP4 (also known as IE180 or IE62, depending on the virus). Therefore, researchers must balance the desire to study the biological functions associated with these regions of the genome, with the technical difficulties of sequencing them. Based on the priorities of each study, some groups have decided that omission of the repeats was the most useful (Greninger et al., 2018; Newman et al., 2015; Norberg et al., 2011). Other studies include an analysis of the repetitive regions, with certain caveats given to the analysis of the data (Szpara et al., 2014, 2011; Kolb et al., 2015; Pandey et al., 2016). Other groups seek to address the technical difficulties of using HTS approaches in repetitive areas by using Sanger sequencing to resolve troublesome regions of the genome (Minaya et al., 2017a). The challenges surrounding the repetitive regions of these genomes, and how to deal with them, remain one of the most challenging aspects of alphaherpesvirus genomics.

High quality genomes result from dedicated bioinformatics and quality control

While the decisions that go into choosing a sequencing method are not trivial, they mean little without proper bioinformatic efforts to assemble viral genomes and analyze them. The ability to rapidly sequence and assemble large DNA virus genomes like the alphaherpesviruses has been facilitated by advances in software and computational workflows, although the diversity of approaches to

the task has led to a wide variety in the quality of finished genomes. One of the first choices in the bioinformatic analysis of viral HTS data is whether to construct a genome by alignment of the sequencing reads to a prior reference genome, or by *de novo* assembly (Posada-Céspedes et al., 2016). Each approach has its own strengths and weaknesses that should be considered. Alignment-based approaches use a previously determined reference genome to achieve faster results, but the end result is biased by the reference genome. As a result, these approaches are prone to mis-calling of minority and structural variants, and may have repeat-lengths that are missing (gapped) or biased by the reference genome chosen for alignment. *De novo* approaches are unbiased by prior data, but they are significantly more computationally intensive, and can require more curation following the initial assembly to obtain optimal results. However, the effort required by *de novo* assembly is often rewarded by a greater ability to detect structural changes in genomes, such as insertions or deletions, or minor variants in the viral population. We have found that a combination of iterative *de novo* assembly and reference-guided ordering of the resulting sequence blocks, as first demonstrated for herpesviruses by Cunningham *et al.* (Cunningham et al., 2010), has been the most useful and robust approach thus far - whether for cultured viral genomes, sparse field samples, or newer oligo-enrichment derived genomes (Parsons et al., 2015; Szpara et al., 2014; Shipley et al., 2018; Pandey et al., 2016)

The tools for performing the steps of genome assembly and analysis are often open-source and include web-based platforms as well as others using command-line (Unix) interfaces. The vast majority of viral *de novo* assembly programs have been developed and tested only for RNA viruses. We have developed both web-based (VirAmp; (Wan et al., 2015)) and command-line based (VirGA; (Parsons et al., 2015)) workflows for *de novo* assembly of herpesvirus genomes. VirAmp is unique in using the Galaxy framework of web-based bioinformatics tools (Blankenberg et al., 2010; Giardine et al., 2005). Most options for viral genome assembly, alignment, annotation and comparison utilize a command-line interface, which has a more difficult learning curve. These Unix-based software options, such as our *de novo* genome assembly suite VirGA (Parsons et al., 2015), the alignment packages Bowtie and BWA

(Langmead and Salzberg, 2012; Posada-Céspedes et al., 2016), and similar script-based workflows (Greninger et al., 2018; Mangul et al., 2014), are usually open source and are generally available through repositories like BitBucket or GitHub. If these programs prove too difficult to use, researchers can also choose from commercial packages (ex. Geneious© Biomatters and CLC Bio© Qiagen) for alignment or *de novo* assembly of genomes.

The proliferation of genomics software has been a boon for the production and publication of genomics data. However, this increased accessibility comes with a potential loss of quality, if these "easier-to-use" programs are used improperly. The rationale for the generation of high-quality genomes has been established by the human genome project and by many microbial genome projects (Fraser et al., 2002). For those that are new to genomics, the question of what constitutes an optimal, "high quality" genome may be a very difficult one to answer. Without delving into a true point-by-point definition of what may be considered a high-quality genome assembly, there are a few guidelines that are helpful to follow, particularly for alphaherpesvirus genomics. Genomes should be fully assembled, ideally without gaps or unfinished regions in the final sequence. This includes the large structural repeat regions, and the genes contained within these. These genomes should also be annotated to the full ability that current research allows and deposited into an international repository such as GenBank or the European Nucleotide Archive (ENA). The genome annotation should include genes and protein coding regions, along with details on the assembly and annotation methods, in addition to strain source and handling (see Figure 2). These details enable future comparisons with other genomes and gene-sequence data. Raw sequence read data should be deposited in an international Short Read Archive (SRA), for others to validate and/or re-analyze in the future. As discussed above ("Clear nomenclature..."), high quality genomes with detailed annotation and nomenclature (Figure 2) enable future scientists to accurately trace the origins of the data and utilize these data in further studies.

Herpesviruses can occasionally present unique challenges to these guidelines, as completion of the sequencing of complex tandem repeats or G+C-rich

sequences can be difficult. However, the absence of such sequences hinders future comparative studies, because partial genomes represent an incomplete source of data. This is particularly worrisome when these regions are a source of genetic diversity and rapid evolution, as with the herpesviruses. In a recent comparison of neonatal HSV-2 genomes, for example, we found almost sixty adult-derived HSV-2 genomes in the GenBank sequence database (Akhtar et al., 2018). However less than ten of these GenBank-deposited HSV-2 genomes included the key genes found in the internal and terminal repeat regions: the key transcriptional regulator ICP4 (encoded by the RS1 gene), the neurovirulence protein ICP34.5 (RL1), and the immediate-early protein ICP0 (RL2) (Johnston et al., 2017a, 2017b; Koelle et al., 2017; Kolb et al., 2015; Newman et al., 2015). This missing data prevents a full genetic comparison of coding diversity in these genes, all of which have been demonstrated to be crucial for *in vivo* pathogenesis and spread (Roizman and Campadelli-Fiume, 2007). Since a sequenced viral genome forms the basis for all hybrid sequencing applications like RNAseq, chromatin-immunoprecipitation (ChIP) sequencing, and chromatin conformational capture (CCC or 3C) sequencing, deficiencies in the starting reference genome can cause many unintentional downstream problems. The tremendous insights that hybrid genomic approaches can provide therefore rely upon a strong foundation in the initial deciphering of viral genome populations.

Future Frontiers

Throughout this article we have discussed the progression of alphaherpesvirus genomics, from the first assembled genomes to the current state of the art, and some of the insights that these studies have revealed. The technological advancements that have occurred in recent years have driven a substantially greater understanding of the genetic diversity and evolution of herpesviruses. We now have an appreciation of the variations that can be seen within herpesvirus genomes, and that these viruses often exist as populations which contain heterogeneous genomes. This heterogeneity has been documented via the *in vivo* detection of minor variants during genital infection by HSV-1 (Shipley et al., 2018) or HSV-2 (Minaya et al., 2017a), and the selection of drug-resistant variants during antiviral therapy (Burrell et al., 2010; Depledge et al., 2016b;

Houldcroft et al., 2017; Sauerbrei et al., 2010). Studies using newer oligo-enrichment approaches have begun to reveal the genetic diversity present *in vivo*, in natural human herpesvirus infections (Depledge et al., 2016b; Greninger et al., 2018; Johnston et al., 2017a; Shipley et al., 2018), without any potential bias from expanding the virus population in cell culture. The plasticity of alphaherpesvirus populations has also been observed in cultured virus stocks (Greninger et al., 2018; Parsons et al., 2015), and examined over time, as seen in our recent application of *in vitro* evolution to HSV-1 (Kuny et al., 2020). With a renewed appreciation of the connection between herpesvirus genetics and biology, it is only natural to look to the future for the next revelations.

First, the potential for improvement within third-generation sequencing (long-read) technology is tantalizing, and its application to herpesvirus genomics could be revolutionary. Early applications of MinION and SMRT long-read sequencing in herpesvirology have already been useful, providing many new insights into novel transcriptional networks (Depledge et al., 2018a, 2018c; Tombácz et al., 2017b, 2017a). The combination of this approach with approaches discussed above (see "Other contributions to functional diversity" above), such as ribosome profiling or the analysis of novel transcripts, will further illuminate the functional genetic outputs of these viral genomes (Depledge et al., 2018a). Oldfield *et al.* recently demonstrated a completely synthetic approach to the generation of an alphaherpesvirus genome, which provides an opportunity for vast improvements in the efficiency of generating multiple precise mutations in a single viral genome, albeit at a currently higher cost (Oldfield et al., 2017). Maturation and improvement in the accuracy of MinION and SMRT technologies will allow for improved detection of recombinant genomes and structural variants, as well as the ability to define genetic haplotypes in mixed populations. These technologies may also allow for sequencing through the entirety of the large internal and terminal repeat regions of alphaherpesviruses, allowing improved assembly and a higher-throughput analysis of those areas of the genome. These methods may also be able to detect methylated bases or secondary structures in the viral DNA, or even read substrates besides DNA and RNA (Depledge et al., 2018a).

Another intriguing future area of alphaherpesvirus genomics research will be to link viral genetics to observable phenotypes, particularly virulence (Dutilh et al., 2013; Power et al., 2016). As current-generation HTS technologies have matured, genome wide association studies (GWAS) and quantitative trait locus (QTL) studies have been applied to this question for herpesviruses. Brandt and colleagues recently used QTL analyses to examine the contribution of specific viral genotypes to ocular phenotypes of HSV-1 infection in mice (Kolb et al., 2016; Lee et al., 2016). This study used the recombinant progeny of two strains of HSV-1, where the individual recombinants were completely sequenced with HTS and analyzed with comparative genomics (Lee et al., 2015). Examining the phenotypes of randomly generated or naturally occurring variants is also known as forward genetics, and these approaches can complement previous decades of reverse genetic approaches (i.e. disrupting specific genes and studying the results). Reverse genetics was crucial to discover the function(s) of many herpesvirus genes, but the types of mutations that are introduced in most reverse genetic approaches (e.g. entire gene deletion or complete disruption of enzymatic function) are rarely seen in field or clinical samples (Bryant et al., 2018; Depledge et al., 2018b; Pandey et al., 2016; Johnston et al., 2017b). The more subtle changes observed in natural settings may have substantial impacts in human or animal clinical outcomes (Loncoman et al., 2017; Bryant et al., 2018; Mechelli et al., 2015; Arav-Boger, 2015). Extending GWAS analyses to naturally isolated viral variants, particularly in comparison to matched human samples, will shed light on the intersection of human and viral genetics that leads to a spectrum of observed disease outcomes (Thompson et al., 2014; Ramchandani et al., 2019).

Acknowledgements

For the purposes of this review of alphaherpesvirus genomics, we have focused primarily on references using high-throughput and deep sequencing-based approaches. However, these data would not be possible without the foundational research of many scientists whose work is not cited here for space reasons. We thank members of the Szpara lab and the Genomics Methods discussion groups at the International Herpesvirus Workshops for insightful conversations and feedback. This work was supported by the Eberly College of

Science and the Huck Institutes for the Life Sciences at the Pennsylvania State University (MLS), and by NIH grant support from R01 AI132692.

References

- Akhtar, L.N., Bowen, C.D., Renner, D.W., Pandey, U., Della Fera, A.N., Kimberlin, D.W., Prichard, M.N., Whitley, R.J., Weitzman, M.D., and Szpara, M.L. (2018). Genotypic and phenotypic diversity within the neonatal HSV-2 population. *bioRxiv* <http://biorxiv.org/lookup/doi/10.1101/262055>.
- Akhtar, L.N., Bowen, C.D., Renner, D.W., Pandey, U., Della Fera, A.N., Kimberlin, D.W., Prichard, M.N., Whitley, R.J., Weitzman, M.D., and Szpara, M.L. (2019). Genotypic and Phenotypic Diversity of Herpes Simplex Virus 2 within the Infected Neonatal Population. *MSphere* *4*, e00590-18.
- Andino, R., and Domingo, E. (2015). Viral quasispecies. *Virology* *479-480*, 46-51.
- Arav-Boger, R. (2015). Strain Variation and Disease Severity in Congenital Cytomegalovirus Infection. *Infect. Dis. Clin. North Am.* *29*, 401-414.
- Arias, C., Weisburd, B., Stern-Ginossar, N., Mercier, A., Madrid, A.S., Bellare, P., Holdorf, M., Weissman, J.S., and Ganem, D. (2014). KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. *PLoS Pathog* *10*, e1003847.
- Artusi, S., Nadai, M., Perrone, R., Biasolo, M.A., Palù, G., Flamand, L., Calistri, A., and Richter, S.N. (2015). The Herpes Simplex Virus-1 genome contains multiple clusters of repeated G-quadruplex: Implications for the antiviral activity of a G-quadruplex ligand. *Antiviral Res.* *118*, 123-131.
- Biswas, B., Kandpal, M., Jauhari, U.K., and Vivekanandan, P. (2016). Genome-wide analysis of G-quadruplexes in herpesvirus genomes. *BMC Genomics* *17*.
- Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* *Chapter 19*, Unit 19.10.1-21.
- Bondre, V.P., Sankararaman, V., Andhare, V., Tupekar, M., and Sapkal, G.N. (2016). Genetic characterization of human herpesvirus type 1: Full-length

- genome sequence of strain obtained from an encephalitis case from India. *Indian J. Med. Res.* *144*, 750-760.
- Bowen, C.D., Renner, D.W., Shreve, J.T., Tafuri, Y., Payne, K.M., Dix, R.D., Kinchington, P.R., Gatherer, D., and Szpara, M.L. (2016). Viral forensic genomics reveals the relatedness of classic herpes simplex virus strains KOS, KOS63, and KOS79. *Virology* *492*, 179-186.
- Bowen, C.D., Paavilainen, H., Renner, D.W., Palomäki, J., Lehtinen, J., Vuorinen, T., Norberg, P., Hukkanen, V., and Szpara, M. (2018). HSV-1 strains circulating in Finland demonstrate uncoupling of geographic and phenotypic variation.
- Breuer, J., Grose, C., Norberg, P., Tipples, G., and Schmid, D.S. (2010). A proposal for a common nomenclature for viral clades that form the species varicella-zoster virus: summary of VZV Nomenclature Meeting 2008, Barts and the London School of Medicine and Dentistry, 24-25 July 2008. *J. Gen. Virol.* *91*, 821-828.
- Brown, J. (2004). Effect of gene location on the evolutionary rate of amino acid substitutions in herpes simplex virus proteins. *Virology* *330*, 209-220.
- Browning, G.F., Vaz, P.K., Devlin, J.M., Hartley, C.A., Studdert, M.J., Gilkerson, J.R., Horsington, J., and Ficorilli, N.P. (2016). Evidence of widespread natural recombination among field isolates of equine herpesvirus 4 but not among field isolates of equine herpesvirus 1. *J. Gen. Virol.* *97*, 747-755.
- Bryant, N.A., Wilkie, G.S., Russell, C.A., Compston, L., Grafham, D., Clissold, L., McLay, K., Medcalf, L., Newton, R., Davison, A.J., et al. (2018). Genetic diversity of equine herpesvirus 1 isolated from neurological, abortigenic and respiratory disease outbreaks. *Transbound. Emerg. Dis.* *65*, 817-832.
- Burrel, S., Deback, C., Agut, H., and Boutolleau, D. (2010). Genotypic Characterization of UL23 Thymidine Kinase and UL30 DNA Polymerase of Clinical Isolates of Herpes Simplex Virus: Natural Polymorphism and Mutations Associated with Resistance to Antivirals. *Antimicrob. Agents Chemother.* *54*, 4833-4842.
- Burrel, S., Désiré, N., Marlet, J., Dacheux, L., Seang, S., Caumes, E., Bourhy, H., Agut, H., and Boutolleau, D. (2015). Genetic Diversity within Alphaherpesviruses: Characterization of a Novel Variant of Herpes Simplex Virus 2. *J. Virol.* *89*, 12273-12283.

- Burrell, S., Boutolleau, D., Ryu, D., Agut, H., Merkel, K., Leendertz, F.H., and Calvignac-Spencer, S. (2017). Ancient Recombination Events between Human Herpes Simplex Viruses. *Mol. Biol. Evol.* *34*, 1713-1721.
- Casto, A.M., Roychoudhury, P., Xie, H., Selke, S., Perchetti, G.A., Wofford, H., Huang, M.-L., Verjans, G.M.G.M., Gottlieb, G.S., Wald, A., et al. (2019). Large, stable, contemporary interspecies recombination events in circulating human herpes simplex viruses. *J. Infect. Dis.* *jiz199*.
- Cunningham, C., Gatherer, D., Hilfrich, B., Baluchova, K., Derrick, J., Thomson, M., Griffiths, P.D., Wilkinson, G.W.G., Schulz, T.F., Dargan, D.J., et al. (2010). Sequences of complete human cytomegalovirus genomes from infected cell cultures and clinical specimens. *J. Gen. Virol.* *91*, 605-615.
- Davison, A.J. (2010). Herpesvirus systematics. *Vet. Microbiol.* *143*, 52-69.
- Davison, A.J., and Wilkie, N.M. (1981). Nucleotide sequences of the joint between the L and S segments of herpes simplex virus types 1 and 2. *J. Gen. Virol.* *55*, 315-331.
- Depledge, D.P., Palser, A.L., Watson, S.J., Lai, I.Y.-C., Gray, E.R., Grant, P., Kanda, R.K., Leproust, E., Kellam, P., and Breuer, J. (2011). Specific Capture and Whole-Genome Sequencing of Viruses from Clinical Samples. *PLoS ONE* *6*, e27805.
- Depledge, D.P., Gray, E.R., Kundu, S., Cooray, S., Poulsen, A., Aaby, P., and Breuer, J. (2014a). Evolution of Cocirculating Varicella-Zoster Virus Genotypes during a Chickenpox Outbreak in Guinea-Bissau. *J. Virol.* *88*, 13936-13946.
- Depledge, D.P., Kundu, S., Jensen, N.J., Gray, E.R., Jones, M., Steinberg, S., Gershon, A., Kinchington, P.R., Schmid, D.S., Balloux, F., et al. (2014b). Deep Sequencing of Viral Genomes Provides Insight into the Evolution and Pathogenesis of Varicella Zoster Virus and Its Vaccine in Humans. *Mol. Biol. Evol.* *31*, 397-409.
- Depledge, D.P., Yamanishi, K., Gomi, Y., Gershon, A.A., and Breuer, J. (2016a). Deep Sequencing of Distinct Preparations of the Live Attenuated Varicella-Zoster Virus Vaccine Reveals a Conserved Core of Attenuating Single-Nucleotide Polymorphisms. *J. Virol.* *90*, 8698-8704.
- Depledge, D.P., Brown, J., Macanovic, J., Underhill, G., and Breuer, J. (2016b). Viral Genome Sequencing Proves Nosocomial Transmission of Fatal Varicella. *J. Infect. Dis.* *214*, 1399-1402.

- Depledge, D.P., Mohr, I., and Wilson, A.C. (2018a). Going the distance: optimizing RNA-Seq strategies for transcriptomic analysis of complex viral genomes. *J. Virol.*
- Depledge, D.P., Cudini, J., Kundu, S., Atkinson, C., Brown, J.R., Haque, T., Houldcroft, C.J., Koay, E.S., McGill, F., Milne, R., et al. (2018b). High Viral Diversity and Mixed Infections in Cerebral Spinal Fluid From Cases of Varicella Zoster Virus Encephalitis. *J. Infect. Dis.*
- Depledge, D.P., Kalanghad Puthankalam, S., Sadaoka, T., Beady, D., Mori, Y., Placantonakis, D., Mohr, I., and Wilson, A. (2018c). Native RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen.
- Depledge, D.P., Srinivas, K.P., Sadaoka, T., Bready, D., Mori, Y., Placantonakis, D.G., Mohr, I., and Wilson, A.C. (2019). Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nat. Commun.* *10*, 1-13.
- Drake, J.W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl. Acad. Sci.* *88*, 7160-7164.
- Drake, J.W., and Hwang, C.B.C. (2005). On the mutation rate of herpes simplex virus type 1. *Genetics* *170*, 969-970.
- Dutilh, B.E., Backus, L., Edwards, R.A., Wels, M., Bayjanov, J.R., and van Hijum, S.A.F.T. (2013). Explaining microbial phenotypes on a genomic scale: GWAS for microbes. *Brief. Funct. Genomics* *12*, 366-380.
- Figuroa, T., Boumart, I., Coupeau, D., and Rasschaert, D. (2016). Hyperediting by ADAR1 of a new herpesvirus lncRNA during the lytic phase of the oncogenic Marek's disease virus. *J. Gen. Virol.* *97*, 2973-2988.
- Firth, C., Kitchen, A., Shapiro, B., Suchard, M.A., Holmes, E.C., and Rambaut, A. (2010). Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Mol. Biol. Evol.* *27*, 2038-2051.
- Fraser, C.M., Eisen, J.A., Nelson, K.E., Paulsen, I.T., and Salzberg, S.L. (2002). The Value of Complete Microbial Genome Sequencing (You Get What You Pay For). *J. Bacteriol.* *184*, 6403-6405.
- Gantt, S., Orem, J., Krantz, E.M., Morrow, R.A., Selke, S., Huang, M.-L., Schiffer, J.T., Jerome, K.R., Nakaganda, A., Wald, A., et al. (2016). Prospective Characterization of the Risk Factors for Transmission and

- Symptoms of Primary Human Herpesvirus Infections Among Ugandan Infants. *J. Infect. Dis.* jiw076.
- Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., et al. (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* *15*, 1451-1455.
- Gorzer, I., Guelly, C., Trajanoski, S., and Puchhammer-Stockl, E. (2010). Deep sequencing reveals highly complex dynamics of human cytomegalovirus genotypes in transplant patients over time. *J. Virol.* *84*, 7195-7203.
- Greninger, A.L., Roychoudhury, P., Xie, H., Casto, A., Cent, A., Pepper, G., Koelle, D.M., Huang, M.-L., Wald, A., Johnston, C., et al. (2018). Ultrasensitive Capture of Human Herpes Simplex Virus Genomes Directly from Clinical Samples Reveals Extraordinarily Limited Evolution in Cell Culture. *MSphere* *3*, e00283-18.
- Griffiths, A. (2011). Slipping and sliding: Frameshift mutations in herpes simplex virus thymidine kinase and drug-resistance. *Drug Resist. Updat.* *14*, 251-259.
- Griffiths, A., Chen, S., Horsburgh, B.C., and Coen, D.M. (2003). Translational Compensation of a Frameshift Mutation Affecting Herpes Simplex Virus Thymidine Kinase Is Sufficient To Permit Reactivation from Latency. *J. Virol.* *77*, 4703-4709.
- Hage, E., Wilkie, G.S., Linnenweber-Held, S., Dhingra, A., Suárez, N.M., Schmidt, J.J., Kay-Fedorov, P.C., Mischak-Weissinger, E., Heim, A., Schwarz, A., et al. (2017). Characterization of Human Cytomegalovirus Genome Diversity in Immunocompromised Hosts by Whole-Genome Sequencing Directly From Clinical Specimens. *J. Infect. Dis.* *215*, 1673-1683.
- Hall, J.D., and Almy, R.E. (1982). Evidence for control of herpes simplex virus mutagenesis by the viral DNA polymerase. *Virology* *116*, 535-543.
- Hennig, T., Michalski, M., Rutkowski, A.J., Djakovic, L., Whisnant, A.W., Friedl, M.-S., Jha, B.A., Baptista, M.A.P., L'Hernault, A., Erhard, F., et al. (2018). HSV-1-induced disruption of transcription termination resembles a cellular stress response but selectively increases chromatin accessibility downstream of genes. *PLOS Pathog.* *14*, e1006954.
- Houldcroft, C.J., Bryant, J.M., Depledge, D.P., Margetts, B.K., Simmonds, J., Nicolaou, S., Tutill, H.J., Williams, R., Worth, A.J.J., Marks, S.D., et al. (2016).

- Detection of Low Frequency Multi-Drug Resistance and Novel Putative Maribavir Resistance in Immunocompromised Pediatric Patients with Cytomegalovirus. *Front. Microbiol.* 7.
- Houldcroft, C.J., Beale, M.A., and Breuer, J. (2017). Clinical and biological insights from viral genome sequencing. *Nat. Rev. Microbiol.* 15, 183-192.
- Huson, D.H. (1998). SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14, 68-73.
- Jansen, M.A.E., van den Heuvel, D., Bouthoorn, S.H., Jaddoe, V.W.V., Hooijkaas, H., Raat, H., Fraaij, P.L.A., van Zelm, M.C., and Moll, H.A. (2016). Determinants of Ethnic Differences in Cytomegalovirus, Epstein-Barr Virus, and Herpes Simplex Virus Type 1 Seroprevalence in Childhood. *J. Pediatr.* 170, 126-134.e6.
- Jaramillo, N., Domingo, E., Munoz-Egea, M.C., Tabares, E., and Gadea, I. (2013). Evidence of Muller's ratchet in herpes simplex virus type 1. *J. Gen. Virol.* 94, 366-375.
- Jayasundara, D., Saeed, I., Maheswararajah, S., Chang, B.C., Tang, S.-L., and Halgamuge, S.K. (2015). ViQuaS: an improved reconstruction pipeline for viral quasispecies spectra generated by next-generation sequencing. *Bioinformatics* 31, 886-896.
- Johnston, C., Magaret, A., Roychoudhury, P., Greninger, A.L., Reeves, D., Schiffer, J., Jerome, K.R., Sather, C., Diem, K., Lingappa, J.R., et al. (2017a). Dual-strain genital herpes simplex virus type 2 (HSV-2) infection in the US, Peru, and 8 countries in sub-Saharan Africa: A nested cross-sectional viral genotyping study. *PLOS Med.* 14, e1002475.
- Johnston, C., Magaret, A., Roychoudhury, P., Greninger, A.L., Cheng, A., Diem, K., Fitzgibbon, M.P., Huang, M., Selke, S., Lingappa, J.R., et al. (2017b). Highly conserved intragenic HSV-2 sequences: Results from next-generation sequencing of HSV-2 U L and U S regions from genital swabs collected from 3 continents. *Virology* 510, 90-98.
- Keys, J.R., Zhou, S., Anderson, J.A., Eron, J.J., Rackoff, L.A., Jabara, C., and Swanstrom, R. (2015). Primer ID Informs Next-Generation Sequencing Platforms and Reveals Preexisting Drug Resistance Mutations in the HIV-1 Reverse Transcriptase Coding Domain. *AIDS Res. Hum. Retroviruses* 31, 658-668.

- Klupp, B.G., Hengartner, C.J., Mettenleiter, T.C., and Enquist, L.W. (2004). Complete, Annotated Sequence of the Pseudorabies Virus Genome. *J. Virol.* **78**, 424-440.
- Koelle, D.M., Norberg, P., Fitzgibbon, M.P., Russell, R.M., Greninger, A.L., Huang, M.-L., Stensland, L., Jing, L., Magaret, A.S., Diem, K., et al. (2017). Worldwide circulation of HSV-2 × HSV-1 recombinant strains. *Sci. Rep.* **7**, 44084.
- Kolb, A.W., Larsen, I.V., Cuellar, J.A., and Brandt, C.R. (2015). Genomic, Phylogenetic, and Recombinational Characterization of Herpes Simplex Virus 2 Strains. *J. Virol.* **89**, 6427-6434.
- Kolb, A.W., Lee, K., Larsen, I., Craven, M., and Brandt, C.R. (2016). Quantitative Trait Locus Based Virulence Determinant Mapping of the HSV-1 Genome in Murine Ocular Infection: Genes Involved in Viral Regulatory and Innate Immune Networks Contribute to Virulence. *PLOS Pathog.* **12**, e1005499.
- Kolb, A.W., Lewin, A.C., Moeller Trane, R., McLellan, G.J., and Brandt, C.R. (2017). Phylogenetic and recombination analysis of the herpesvirus genus varicellovirus. *BMC Genomics* **18**.
- Koonin, E.V., Krupovic, M., and Yutin, N. (2015). Evolution of double-stranded DNA viruses of eukaryotes: from bacteriophages to transposons to giant viruses: Evolution of dsDNA viruses of eukaryotes. *Ann. N. Y. Acad. Sci.* **1341**, 10-24.
- Kuhn, J.H., Bao, Y., Bavari, S., Becker, S., Bradfute, S., Brister, J.R., Bukreyev, A.A., Chandran, K., Davey, R.A., Dolnik, O., et al. (2013). Virus nomenclature below the species level: a standardized nomenclature for natural variants of viruses assigned to the family Filoviridae. *Arch. Virol.* **158**, 301-311.
- Kuny, C.V., Bowen, C.D., Renner, D.W., Johnston, C.M., and Szpara, M.L. (2020). In vitro evolution of herpes simplex virus 1 (HSV-1) reveals selection for syncytia and other minor variants in cell culture. *Virus Evol.* **6**.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359.
- Lassalle, F., Beale, M.A., Bharucha, T., Williams, C.A., Williams, R.J., Cudini, J., Goldstein, R., Haque, T., Depledge, D.P., and Breuer, J. (2020). Whole genome sequencing of Herpes Simplex Virus 1 directly from human

- cerebrospinal fluid reveals selective constraints in neurotropic viruses. *Virus Evol.* **6**.
- Lauring, A.S., and Andino, R. (2011). Exploring the Fitness Landscape of an RNA Virus by Using a Universal Barcode Microarray. *J. Virol.* **85**, 3780-3791.
- Lee, K., Kolb, A.W., Sverchkov, Y., Cuellar, J.A., Craven, M., and Brandt, C.R. (2015). Recombination Analysis of Herpes Simplex Virus 1 Reveals a Bias toward GC Content and the Inverted Repeat Regions. *J. Virol.* **89**, 7214-7223.
- Lee, K., Kolb, A.W., Larsen, I., Craven, M., and Brandt, C.R. (2016). Mapping Murine Corneal Neovascularization and Weight Loss Virulence Determinants in the Herpes Simplex Virus 1 Genome and the Detection of an Epistatic Interaction between the UL and IRS/US Regions. *J. Virol.* **90**, 8115-8131.
- Lee, S.-W., Markham, P.F., Coppo, M.J.C., Legione, A.R., Markham, J.F., Noormohammadi, A.H., Browning, G.F., Ficorilli, N., Hartley, C.A., and Devlin, J.M. (2012). Attenuated Vaccines Can Recombine to Form Virulent Field Viruses. *Science* **337**, 188-188.
- Lewin, A.C., Kolb, A.W., McLellan, G.J., Bentley, E., Bernard, K.A., Newbury, S.P., and Brandt, C.R. (2018). Genomic, Recombinational and Phylogenetic Characterization of Global Feline Herpesvirus 1 Isolates. *Virology* **518**, 385-397.
- Lin, Z., Wang, X., Strong, M.J., Concha, M., Baddoo, M., Xu, G., Baribault, C., Fewell, C., Hulme, W., Hedges, D., et al. (2013). Whole-Genome Sequencing of the Akata and Mutu Epstein-Barr Virus Strains. *J. Virol.* **87**, 1172-1182.
- Loncoman, C.A., Vaz, P.K., Coppo, M.J., Hartley, C.A., Morera, F.J., Browning, G.F., and Devlin, J.M. (2017). Natural recombination in alphaherpesviruses: Insights into viral evolution through full genome sequencing and sequence analysis. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* **49**, 174-185.
- Looker, K.J., Magaret, A.S., May, M.T., Turner, K.M.E., Vickerman, P., Gottlieb, S.L., and Newman, L.M. (2015a). Global and regional estimates of prevalent and incident herpes simplex virus type 1 infections in 2012. *PLoS ONE* **10**, e0140765.
- Looker, K.J., Magaret, A.S., Turner, K.M.E., Vickerman, P., Gottlieb, S.L., and Newman, L.M. (2015b). Global estimates of prevalent and incident herpes simplex virus type 2 infections in 2012. *PloS One* **10**, e114989.

- Lou, D.I., Hussmann, J.A., McBee, R.M., Acevedo, A., Andino, R., Press, W.H., and Sawyer, S.L. (2013). High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. *Proc. Natl. Acad. Sci.* *110*, 19872-19877.
- Mahiet, C., Ergani, A., Huot, N., Alende, N., Azough, A., Salvaire, F., Bensimon, A., Conseiller, E., Wain-Hobson, S., Labetoulle, M., et al. (2012). Structural Variability of the Herpes Simplex Virus 1 Genome In Vitro and In Vivo. *J. Virol.* *86*, 8592-8601.
- Mangul, S., Wu, N.C., Mancuso, N., Zelikovsky, A., Sun, R., and Eskin, E. (2014). Accurate viral population assembly from ultra-deep sequencing data. *Bioinformatics* *30*, i329-i337.
- McGeoch, D.J., Dolan, A., Donald, S., and Rixon, F.J. (1985). Sequence determination and genetic content of the short unique region in the genome of herpes simplex virus type 1. *J. Mol. Biol.* *181*, 1-13.
- McGeoch, D.J., Dalrymple, M.A., Davison, A.J., Dolan, A., Frame, M.C., McNab, D., Perry, L.J., Scott, J.E., and Taylor, P. (1988). The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. *J. Gen. Virol.* *69*, 1531-1574.
- Mechelli, R., Manzari, C., Policano, C., Annese, A., Picardi, E., Umeton, R., Fornasiero, A., D'erchia, A.M., Buscarinu, M.C., Agliardi, C., et al. (2015). Epstein-Barr virus genetic variants are associated with multiple sclerosis. *Neurology* *84*, 1362-1368.
- Minaya, M.A., Jensen, T., Goll, J., Korom, M., Datla, S.H., Belshe, R.B., and Morrison, L.A. (2017a). Molecular evolution of herpes simplex virus 2 complete genomes: Comparison between primary and recurrent infections. *J. Virol.* *JVI.00942-17*.
- Minaya, M.A., Jensen, T.L., Goll, J.B., Korom, M., Datla, S.H., Belshe, R.B., and Morrison, L.A. (2017b). Molecular evolution of herpes simplex virus 2 complete genomes: comparison between primary and recurrent infections. *J. Virol.* *91*, e00942-17.
- Moss, B., Winters, E., and Cooper, N. (1981). Instability and reiteration of DNA sequences within the vaccinia virus genome. *Proc. Natl. Acad. Sci. U. S. A.* *78*, 1614-1618.

- Nakaya, Y., Stavrou, S., Blouch, K., Tattersall, P., and Ross, S.R. (2016). In Vivo Examination of Mouse APOBEC3- and Human APOBEC3A- and APOBEC3G-Mediated Restriction of Parvovirus and Herpesvirus Infection in Mouse Models. *J. Virol.* *90*, 8005-8012.
- Newman, R.M., Lamers, S.L., Weiner, B., Ray, S.C., Colgrove, R.C., Diaz, F., Jing, L., Wang, K., Saif, S., Young, S., et al. (2015). Genome Sequencing and Analysis of Geographically Diverse Clinical Isolates of Herpes Simplex Virus 2. *J. Virol.* *JVI.01303-15*.
- Norberg, P. (2010). Divergence and genotyping of human alpha-herpesviruses: an overview. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* *10*, 14-25.
- Norberg, P., Kasubi, M.J., Haarr, L., Bergstrom, T., and Liljeqvist, J.-A. (2007). Divergence and recombination of clinical herpes simplex virus type 2 isolates. *J. Virol.* *81*, 13158-13167.
- Norberg, P., Tyler, S., Severini, A., Whitley, R., Liljeqvist, J.-A., and Bergstrom, T. (2011). A genome-wide comparative evolutionary analysis of herpes simplex virus type 1 and varicella zoster virus. *PLoS One* *6*, 1-8.
- Norberg, P., Depledge, D.P., Kundu, S., Atkinson, C., Brown, J., Haque, T., Hussaini, Y., MacMahon, E., Molyneaux, P., Papaevangelou, V., et al. (2015). Recombination of Globally Circulating Varicella-Zoster Virus. *J. Virol.* *89*, 7133-7146.
- Oh, H.S., Neuhauser, W.M., Eggan, P., Angelova, M., Kirchner, R., Eggan, K.C., and Knipe, D.M. (2019). Herpesviral lytic gene functions render the viral genome susceptible to novel editing by CRISPR/Cas9. *ELife* *8*.
- Oldfield, L.M., Grzesik, P., Voorhies, A.A., Alperovich, N., MacMath, D., Najera, C.D., Chandra, D.S., Prasad, S., Noskov, V.N., Montague, M.G., et al. (2017). Genome-wide engineering of an infectious clone of herpes simplex virus type 1 using synthetic genomics assembly methods. *Proc. Natl. Acad. Sci.* *201700534*.
- Pan, D., and Coen, D.M. (2012). Net -1 frameshifting on a noncanonical sequence in a herpes simplex virus drug-resistant mutant is stimulated by nonstop mRNA. *Proc. Natl. Acad. Sci.* *109*, 14852-14857.
- Pandey, U., Bell, A.S., Renner, D.W., Kennedy, D.A., Shreve, J.T., Cairns, C.L., Jones, M.J., Dunn, P.A., Read, A.F., and Szpara, M.L. (2016). DNA from dust:

- Comparative genomics of large DNA viruses in field surveillance samples. *MSphere* 1, e00132-16.
- Pandey, U., Renner, D.W., Thompson, R.L., Szpara, M.L., and Sawtell, N.M. (2017). Inferred father-to-son transmission of herpes simplex virus results in near-perfect preservation of viral genome identity and in vivo phenotypes. *Sci. Rep.* 7, 13666.
- Parsons, L.R., Tafuri, Y.R., Shreve, J.T., Bowen, C.D., Shipley, M.M., Enquist, L.W., and Szpara, M.L. (2015). Rapid genome assembly and comparison decode intrastrain variation in human alphaherpesviruses. *MBio* 6, e02213-14.
- Pellett, P.E., and Roizman, B. (2013). Herpesviridae. In *Fields Virology*, D.M. Knipe, and P.M. Howley, eds. (Philadelphia, PA: Lippincott Williams and Wilkins), pp. 1802-1822.
- Peters, G. a, Tyler, S.D., Grose, C., Severini, A., Gray, M.J., Upton, C., and Tipples, G. a (2006). A full-genome phylogenetic analysis of varicella-zoster virus reveals a novel origin of replication-based genotyping scheme and evidence of recombination between major circulating clades. *J. Virol.* 80, 9850-9860.
- Pfaff, F., Groth, M., Sauerbrei, A., and Zell, R. (2016). Genotyping of herpes simplex virus type 1 by whole-genome sequencing. *J. Gen. Virol.* 97, 2732-2741.
- Posada-Céspedes, S., Seifert, D., and Beerenwinkel, N. (2016). Recent advances in inferring viral diversity from high-throughput sequencing data. *Virus Res.* pii: S0168-1702(16)30413-0.
- Power, R.A., Parkhill, J., and de Oliveira, T. (2016). Microbial genome-wide association studies: lessons from human GWAS. *Nat. Rev. Genet.* 18, 41-50.
- Prazsák, I., Moldován, N., Balázs, Z., Tombácz, D., Megyeri, K., Szűcs, A., Csabai, Z., and Boldogkői, Z. (2018). Long-read sequencing uncovers a complex transcriptome topology in varicella zoster virus. *BMC Genomics* 19.
- Ramchandani, M.S., Jing, L., Russell, R.M., Tran, T., Laing, K.J., Magaret, A.S., Selke, S., Cheng, A., Huang, M.-L., Xie, H., et al. (2019). Viral Genetics Modulate Orolabial Herpes Simplex Virus Type 1 Shedding in Humans. *J. Infect. Dis.* 219, 1058-1066.

- Renner, D.W., and Szpara, M.L. (2018). The impacts of genome-wide analyses on our understanding of human herpesvirus diversity and evolution. *J. Virol.* **92**, e00908-17.
- Renzette, N., Gibson, L., Jensen, J.D., and Kowalik, T.F. (2014). Human cytomegalovirus intrahost evolution—a new avenue for understanding and controlling herpesvirus infections. *Curr. Opin. Virol.* **8**, 109-115.
- Roizman, B. (1979). The structure and isomerization of herpes simplex virus genomes. *Cell* **16**, 481-494.
- Roizman, B., and Campadelli-Fiume, G. (2007). Alphaherpes viral genes and their functions. In *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*, (Cambridge), pp. 70-83.
- Roizman, B., Knipe, D.M., and Whitley, R. (2013). Herpes Simplex Viruses. In *Fields Virology*, D.M. Knipe, and P.M. Howley, eds. (Philadelphia, PA: Lippincott Williams and Wilkins), pp. 1823-1897.
- Rutkowski, A.J., Erhard, F., L'Hernault, A., Bonfert, T., Schilhabel, M., Crump, C., Rosenstiel, P., Efstathiou, S., Zimmer, R., Friedel, C.C., et al. (2015). Widespread disruption of host transcription termination in HSV-1 infection. *Nat. Commun.* **6**, 7126.
- Sakaoka, H., Kurita, K., Iida, Y., Takada, S., Umene, K., Kim, Y.T., Ren, C.S., and Nahmias, A.J. (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**, 513-527.
- Sanjuán, R., and Domingo-Calap, P. (2016). Mechanisms of viral mutation. *Cell. Mol. Life Sci.* **73**, 4433-4448.
- Sanjuan, R., Nebot, M.R., Chirico, N., Mansky, L.M., and Belshaw, R. (2010). Viral Mutation Rates. *J. Virol.* **84**, 9733-9748.
- Sarisky, R.T., Nguyen, T.T., Duffy, K.E., Wittrock, R.J., and Leary, J.J. (2000). Difference in incidence of spontaneous mutations between Herpes simplex virus types 1 and 2. *Antimicrob. Agents Chemother.* **44**, 1524-1529.
- Sauerbrei, A., Deinhardt, S., Zell, R., and Wutzler, P. (2010). Phenotypic and genotypic characterization of acyclovir-resistant clinical isolates of herpes simplex virus. *Antiviral Res.* **86**, 246-252.

- Severini, A., Tyler, S.D., Peters, G.A., Black, D., and Eberle, R. (2013). Genome sequence of a chimpanzee herpesvirus and its relation to other primate alphaherpesviruses. *Arch. Virol.* *158*, 1825-1828.
- Shiple, M.M., Renner, D.W., Ott, M., Bloom, D.C., Koelle, D.M., Johnston, C., and Szpara, M.L. (2018). Genome-wide surveillance of genital herpes simplex virus type 1 from multiple anatomic sites over time. *J. Infect. Dis.* *218*, 595-605.
- Shiple, M.M., Renner, D.W., Pandey, U., Ford, B., Bloom, D.C., Grose, C., and Szpara, M.L. (2019). Personalized viral genomic investigation of herpes simplex virus 1 perinatal viremic transmission with dual fatality. *Mol. Case Stud.* *5*, a004382.
- Shiple, M.M., Rathbun, M.M., and Szpara, M.L. (2020). Oligonucleotide Enrichment of HSV-1 Genomic DNA from Clinical Specimens for Use in High-Throughput Sequencing. In *Herpes Simplex Virus*, R.J. Diefenbach, and C. Fraefel, eds. (New York, NY: Springer New York), pp. 199-217.
- Sijmons, S., Thys, K., Mbong Ngwese, M., Van Damme, E., Dvorak, J., Van Loock, M., Li, G., Tachezy, R., Busson, L., Aerssens, J., 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*, 7673-7695.
- Simon, A., Mettenleiter, T.C., and Rziha, H.J. (1989). Pseudorabies virus displays variable numbers of a repeat unit adjacent to the 3' end of the glycoprotein gII gene. *J. Gen. Virol.* *70*, 1239-1246.
- Suspene, R., Aynaud, M.-M., Koch, S., Padeloup, D., Labetoulle, M., Gaertner, B., Vartanian, J.-P., Meyerhans, A., and Wain-Hobson, S. (2011). Genetic editing of herpes simplex virus 1 and Epstein-Barr herpesvirus genomes by human APOBEC3 cytidine deaminases in culture and in vivo. *J. Virol.* *85*, 7594-7602.
- Szpara, M.L., Parsons, L., and Enquist, L.W. (2010). Sequence variability in clinical and laboratory isolates of herpes simplex virus 1 reveals new mutations. *J. Virol.* *84*, 5303-5313.
- Szpara, M.L., Tafuri, Y.R., Parsons, L., Shamim, S.R., Verstrepen, K.J., Legendre, M., and Enquist, L.W. (2011). A wide extent of inter-strain diversity in virulent and vaccine strains of alphaherpesviruses. *PLoS Pathog.* *7*, 1-23.

- Szpara, M.L., Gatherer, D., Ochoa, A., Greenbaum, B., Dolan, A., Bowden, R.J., Enquist, L.W., Legendre, M., and Davison, A.J. (2014). Evolution and diversity in human herpes simplex virus genomes. *J. Virol.* **88**, 1209-1227.
- Thompson, R.L., Williams, R.W., Kotb, M., and Sawtell, N.M. (2014). A Forward Phenotypically Driven Unbiased Genetic Analysis of Host Genes That Moderate Herpes Simplex Virus Virulence and Stromal Keratitis in Mice. *PLoS ONE* **9**, 1-10.
- Tirosh, O., Cohen, Y., Shitrit, A., Shani, O., Le-Trilling, V.T.K., Trilling, M., Friedlander, G., Tanenbaum, M., and Stern-Ginossar, N. (2015). The Transcription and Translation Landscapes during Human Cytomegalovirus Infection Reveal Novel Host-Pathogen Interactions. *PLoS Pathog.* **11**, e1005288.
- Tombácz, D., Csabai, Z., Szűcs, A., Balázs, Z., Moldován, N., Sharon, D., Snyder, M., and Boldogkői, Z. (2017a). Long-Read Isoform Sequencing Reveals a Hidden Complexity of the Transcriptional Landscape of Herpes Simplex Virus Type 1. *Front. Microbiol.* **8**.
- Tombácz, D., Balázs, Z., Csabai, Z., Moldován, N., Szűcs, A., Sharon, D., Snyder, M., and Boldogkői, Z. (2017b). Characterization of the Dynamic Transcriptome of a Herpesvirus with Long-read Single Molecule Real-Time Sequencing. *Sci. Rep.* **7**, 43751.
- Tombácz, D., Moldován, N., Balázs, Z., Gulyás, G., Csabai, Z., Boldogkői, M., Snyder, M., and Boldogkői, Z. (2019). Multiple Long-Read Sequencing Survey of Herpes Simplex Virus Dynamic Transcriptome. *Front. Genet.* **10**.
- Töpfer, A., Marschall, T., Bull, R.A., Luciani, F., Schönhuth, A., and Beerenwinkel, N. (2014). Viral Quasispecies Assembly via Maximal Clique Enumeration. *PLoS Comput. Biol.* **10**, e1003515.
- Treangen, T.J., and Salzberg, S.L. (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.*
- Trimpert, J., Groenke, N., Jenckel, M., He, S., Kunec, D., Szpara, M.L., Spatz, S.J., Osterrieder, N., and McMahon, D.P. (2017). A phylogenomic analysis of Marek's disease virus reveals independent paths to virulence in Eurasia and North America. *Evol. Appl.* **10**, 1091-1101.

- Trimpert, J., Groenke, N., Kunec, D., Eschke, K., He, S., McMahon, D.P., and Osterrieder, N. (2019). A proofreading-impaired herpesvirus generates populations with quasispecies-like structure. *Nat. Microbiol.* 1-9.
- Tyler, S.D., Peters, G.A., Grose, C., Severini, A., Gray, M.J., Upton, C., and Tipples, G.A. (2007). Genomic cartography of varicella-zoster virus: a complete genome-based analysis of strain variability with implications for attenuation and phenotypic differences. *Virology* 359, 447-458.
- Vaz, P.K., Job, N., Horsington, J., Ficorilli, N., Studdert, M.J., Hartley, C.A., Gilkerson, J.R., Browning, G.F., and Devlin, J.M. (2016a). Low genetic diversity among historical and contemporary clinical isolates of feline herpesvirus 1. *BMC Genomics* 17.
- Vaz, P.K., Horsington, J., Hartley, C.A., Browning, G.F., Ficorilli, N.P., Studdert, M.J., Gilkerson, J.R., and Devlin, J.M. (2016b). Evidence of widespread natural recombination among field isolates of equine herpesvirus 4 but not among field isolates of equine herpesvirus 1. *J. Gen. Virol.* 97, 747-755.
- Wan, Y., Renner, D.W., Albert, I., and Szpara, M.L. (2015). VirAmp: a galaxy-based viral genome assembly pipeline. *GigaScience* 4.
- Watson, S.J., Welkers, M.R.A., Depledge, D.P., Coulter, E., Breuer, J.M., de Jong, M.D., and Kellam, P. (2013). Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philos. Trans. R. Soc. B Biol. Sci.* 368, 1-9.
- Wertheim, J.O., Smith, M.D., Smith, D.M., Scheffler, K., and Pond, S.L.K. (2014). Evolutionary Origins of Human Herpes Simplex Viruses 1 and 2. *Mol. Biol. Evol.* 1-26.
- Whisnant, A.W., Jürges, C.S., Hennig, T., Wyler, E., Prusty, B., Rutkowski, A.J., L'hernault, A., Djakovic, L., Göbel, M., Döring, K., et al. (2020). Integrative functional genomics decodes herpes simplex virus 1. *Nat. Commun.* 11.
- Wilkinson, G.W.G., Davison, A.J., Tomasec, P., Fielding, C.A., Aicheler, R., Murrell, I., Seirafian, S., Wang, E.C.Y., Weekes, M., Lehner, P.J., et al. (2015). Human cytomegalovirus: taking the strain. *Med. Microbiol. Immunol. (Berl.)* 204, 273-284.
- Wyler, E., Menegatti, J., Franke, V., Kocks, C., Boltengagen, A., Hennig, T., Theil, K., Rutkowski, A., Ferrai, C., Baer, L., et al. (2017). Widespread

activation of antisense transcription of the host genome during herpes simplex virus 1 infection. *Genome Biol.* **18**.

Ye, C., Guo, J.-C., Gao, J.-C., Wang, T.-Y., Zhao, K., Chang, X.-B., Wang, Q., Peng, J.-M., Tian, Z.-J., Cai, X.-H., et al. (2016). Genomic analyses reveal that partial sequence of an earlier pseudorabies virus in China is originated from a Bartha-vaccine-like strain. *Virology* **491**, 56-63.

Zell, R., Taudien, S., Pfaff, F., Wutzler, P., Platzer, M., and Sauerbrei, A. (2012). Sequencing of 21 Varicella-Zoster Virus Genomes Reveals Two Novel Genotypes and Evidence of Recombination. *J. Virol.* **86**, 1608-1622.