VIRUS EVOLUTION

Virus Evolution, 2017, 3(1): vex015

doi: 10.1093/ve/vex015
Research article

Unexpected differences in the population genetics of phasmaviridids (*Bunyavirales*) from subarctic ponds

Matthew J. Ballinger,[1,*,†] Andrew S. Medeiros,[2] Jie Qin,[3] and Derek J. Taylor[1]

[1]The Department of Biological Sciences, State University of New York, Buffalo, NY, USA, [2]The Department of Geography, York University, Toronto, Canada and [3]College of Life Sciences, Capital Normal University, Beijing, China

*Corresponding author: E-mail: mattball@uvic.ca

†Present address: Department of Biology, University of Victoria, BC, Canada.

Abstract

Little is known of the evolution of RNA viruses in aquatic systems. Here, we assess the genetic connectivity of two bunyaviruses (*Kigluaik phantom orthophasmavirus* or KIGV and *Nome phantom orthophasmavirus* or NOMV) with zooplanktonic hosts from subarctic ponds. We expected weak genetic structure among populations as the hosts (phantom midges) have a terrestrial winged dispersal stage. To test whether their respective viruses mirror this structure, we collected and analyzed population datasets from 21 subarctic freshwater ponds and obtained sequences from all four genes in the viral genomes. Prevalence averaged 66 per cent for 514 host specimens and was not significantly different between recently formed thaw ponds and glacial ponds. Unexpectedly, KIGV from older ponds showed pronounced haplotype divergence with little evidence of genetic connectivity. However, KIGV populations from recent thaw ponds appeared to be represented by a closely related haplotype group, perhaps indicating a genotypic dispersal bias. Unlike KIGV, NOMV had modest structure and diversity in recently formed thaw ponds. For each virus, we found elevated genetic diversity relative to the host, but similar population structures to the host. Our results suggest that non-random processes such as virus–host interactions, genotypic bias, and habitat effects differ among polar aquatic RNA viruses.

Key words: climate; insect-specific; polar freshwater; dispersal; non-random gene flow.

1. Introduction

In recent years, it has become increasingly clear that the diversity of RNA viruses infecting arthropods in nature extends far beyond previous expectations (Cook et al. 2013; Ballinger et al. 2014; Li et al. 2015; Webster et al. 2015; Junglen 2016; Shi et al. 2016). With exceptions for the most notorious arboviruses, our understanding of the constraints imposed on virus evolution by host-specific life histories and ecologies are not well founded in examples from natural populations. Some predict that aspects of useful frameworks such as the quasispecies theory may prove incorrect in nature (Holland 2006), and indeed, in cases where data collected from natural populations have been examined in detail, unexpected discoveries have followed (Nichol et al. 1993; Weber de Melo et al. 2015).

In threatened and changing environments, these questions take on an additional urgency. Recent degradation of permafrost has both drained and formed countless ponds and lakes in polar regions (Yoshikawa and Hinzman 2003; Smith et al. 2005; Riordan et al. 2006; Smol and Douglas 2007; Bouchard et al. 2013; Hunt et al. 2013). If populations of insect-infecting RNA viruses in these waterbodies are typically connected by ongoing dispersal over vast distances, as is the case for DNA viruses of aquatic microbes (Laybourn-Parry et al. 2001; de Cárcer et al. 2015; Paez-Espino et al. 2016), then the consequences of these physical changes may be negligible for viral diversity and

biogeography. However, we need a better understanding of the processes that affect genetic connectivity for aquatic RNA viruses.

Our present study focuses on an unstudied family of RNA viruses, *Phasmaviridae* (genus *Orthophasmavirus*), of the order *Bunyavirales* (Adams et al. 2017), that were first described in polar zooplankton (Ballinger et al. 2014). The *Bunyavirales* are trisegmented, single-stranded, negative sense RNA viruses. Of the nine established families, four include emerging zoonotic agents of vertebrates, e.g. Rift Valley fever virus and Toscana virus (genus *Phlebovirus*, family *Phenuiviridae*), Crimean-congo hemorrhagic fever virus (genus *Nairovirus*, family *Nairoviridae*), Schmallenburg virus (genus *Orthobunyavirus*, family *Peribunyaviridae*), and Puumala virus and Dobrava–Belgrade virus (genus *Orthohantavirus*, family *Hantaviridae*). *Phasmaviridae* is phylogenetically positioned as sister to the *Hantaviridae*, and members have now been found infecting wild populations of mosquitos, cockroaches, water striders, psyllids, odonates and drosophilids (Ballinger et al. 2014; Li et al. 2015; Webster et al. 2015; Nouri et al. 2016; Shi et al. 2016). *In silico* searches for genes of phasmavirids yielded near complete virus genome sequences and a bewildering diversity of potential insect hosts including four major insect orders–Coleoptera, Hymenoptera, Lepidoptera and Diptera (Ballinger et al. 2014). Even so, little is known of the biology underlying phasmavirid–host interactions. The antiquity and widespread occurrence of phasmavirids and related viruses (Marklewitz et al. 2015) among insect taxa (while being undetected in other eukaryotes) is suggestive of an insect-specific tropism. However, dual tropisms and host jumps have repeatedly evolved in the bunyavirids (Junglen 2016), so the designation of an insect-specific tropism for phasmavirids remains tentative. Presently, a phasmavirid system amenable to laboratory study is lacking; however, detailed prevalence and diversity data may provide insight into some basic features underlying virus–host associations, e.g. vertical or horizontal transmission and persistent or acute infection.

The first proposed species of phasmavirids were associated with phantom midges (Diptera: Chaoboridae: Culicoidea)—*Kigluaik phantom orthophasmavirus* (KIGV) is associated with *Chaoborus trivittatus* and *Nome phantom orthophasmavirus* (NOMV) is associated with *C.* cf. *flavicans* (Ballinger et al. 2014). Initial evidence of high prevalence (45–77%) in larval stages and cophylogeny with putative hosts suggested vertical transmission of KIGV and NOMV. The genus *Chaoborus* has a near cosmopolitan geographic distribution with some species being used as food by humans (Ayieko and Oriaro 2008). The larvae are aquatic predators of zooplankton and can reach very high densities ($>50{,}000\ m^{-2}$; Xie et al. 1998). Phantom midges develop through four larval instars, and can spend nearly two years as a fourth instar in the zooplankton (Federenko and Swift 1972). Winged adults are non-feeding and survive for only a few days; eggs are deposited in ponds and lakes (Borkent 1979).

Between the summer of 2000 and 2015, we sampled more than 300 freshwater ponds in a hydrologically disturbed region of the Seward Peninsula, Alaska, to investigate the effects of disturbance processes on assemblage and loss of zooplankton communities, including species of *Chaoborus* (Taylor et al. 2016). We found that two *Chaoborus* species showed contrasting distribution patterns and habitat preferences. *C. trivittatus* was found primarily in large, permanent water bodies throughout the sampling interval; these habitats are ancient glacial lakes and thermokarst ponds (thousands of years old). *C.* cf. *flavicans*, on the other hand, was initially detected only near the boreal zone in the earliest sampling years, but appeared within and spread amongst small, recently formed ponds on the tundra (<50 years old) as sampling continued (Taylor et al. 2016). This species has now colonized new ponds across the entire sampling region, but remains absent from the older water bodies inhabited by *C. trivittatus*. In contrast, analyses of mandibles from sediment cores corresponding to the past 90 years revealed the continuous presence of *C. trivittatus* in one glacial pond on the Seward Peninsula (Medeiros et al. 2014), and in six glacial ponds near Iqlauit, Baffin Island (Medeiros and Quinlan 2011). Berendonk et al. (2009) concluded that populations of *Chaoborus* are strongly connected by dispersal, with species from larger permanent ponds showing weak within-region population genetic structure and species from smaller temporary ponds showing intermediate structure (presumably from frequent recolonization events). As such, we expected viruses associated with *C. trivittatus* (from small lakes and deep ponds) to exhibit weaker population genetic structure than those associated with the congeneric shallow pond specialist in Alaska, *C.* cf. *flavicans* (Taylor et al. 2016).

Given the high initial prevalence data we previously found for KIGV (Ballinger et al. 2014) and the contrasting patterns of host demography in our study region (Taylor et al. 2016), we reasoned that *Chaoborus* phasmavirids may serve as an informative system to test the influence of host-specific ecological factors on virus diversity in natural populations. Here, we compare the prevalence and spatial genetic structure of two subarctic phasmavirids and their hosts from two pond age classes, aiming to assess genetic connectivity and non-random processes (e.g., virus–host effects). We find high prevalence for both KIGV and NOMV in our study sites on the Seward Peninsula, Alaska, and on Baffin Island, Nunavut, but unexpected contrasts in virus diversity and genetic structure, and interestingly, our results suggest these contrasts are largely attributable to habitat age. This work implicates an important interaction between host-specific ecologies, such as dispersal behavior, and habitat history for structuring virus diversity in natural populations, and suggests that greater attention toward these constraints on virus evolution and diversity are warranted as climate effects continue to dramatically alter freshwater habitats in arctic and subarctic regions.

2. Materials and Methods

2.1 Field collections and sample preservation

Chaoborus larvae and other zooplankton were collected from freshwater ponds of the southern Seward Peninsula and the vicinity of Iqaluit, Baffin Island, Arctic Canada (Fig. 1) in late July or early August of 2011–5 by multiple oblique tows using a 200–250 μm throw or dip net (Wildco Scientific) and stored in 100 per cent ethanol at −20 °C. Populations of *Chaoborus* were preserved in common vials such that specimens from the same pond had identical preservation conditions. Species determinations were made based on morphological characteristics of the larval mandible and the labral blade (Borkent 1979). Here, we refer to the species whose larval mandibles superficially resemble those of the lake-dwelling species *Chaoborus* (*Chaoborus*) *flavicans* as *C.* cf, *flavicans*. Note that *C.* cf, *flavicans* is genetically and ecologically similar to *Chaoborus* (*Chaoborus*) *crystallinus*, a species that is highly successful in smaller, temporary, fishless ponds (Dupuis et al. 2008). Site information is listed in Supplementary Table S1. We attempted to establish a laboratory culture from wild caught larvae of both host species in order to directly observe transmission from adult to eggs under lab conditions. A flight

cage with aquarium was constructed based on the design outlined by Moore (1986).

For virus screens of individual larvae, second to fourth instar larvae were cut in half using a fresh microscope slide coverslip. The posterior tissues were dried briefly to evaporate the ethanol prior to RNA extraction and the anterior tissues were stored in 100 per cent ethanol at −20 °C. Dried tissues were individually ground in 1.5 ml microfuge tubes with QuickExtract (Epicentre) using sterilized pestles and incubated at 62 °C while shaking at 200 RPM for 30 min. cDNA was generated using GoScript reverse transcriptase (Promega), GoTaq master mix (Promega) was used for PCR amplification. PCR primer sequences and thermal cycling programs are listed in Supplementary Table S2. All PCR products were assessed for amplification success and correct size on a 1 per cent agarose gel stained with ethidium bromide. Genomic locations for amplicons are shown in Supplementary Fig. S1. We screened eleven populations of *C. trivittatus* from Alaska and two populations from lakes near Iqaluit, Baffin Island. Eight populations of *C.* cf. *flavicans* from Alaska were screened for NOMV. About 514 subarctic specimens of *Chaoborus* were screened for phasmavirids. We subsampled the invertebrate communities from the southern Seward ponds examined in this study, with a focus on populations in which other arthropod predators are present, and screened by RT-PCR for the respective phasmavirid L segment. As a further assessment of host specificity, we screened eleven larvae of a third species of *Chaoborus* (*Chaoborus americanus*) from a pond where it co-occurs with *C. trivittatus*. Initial RNA quality of field preserved specimens was verified by successful RT-PCR and sequencing of actin mRNA from twenty specimens of *Chaoborus*. We also examined the mitochondrial population genetic structure for the host species by sequencing two mitochondrial genes, *COI* and *ND4*.

Unpurified PCR products were sent to the high throughput DNA sequencing facility at the University of Washington or Tacgen (California) for Sanger sequencing. Sequence chromatograms were assembled and examined in Geneious R7. Primer regions were trimmed and consensus sequences were generated based on highest quality. Each amplicon was sequenced in both directions and the assembled reads were examined manually for internal peaks. Sequence alignments were created in Geneious using the MAFFT algorithm (Katoh and Standley 2014) plugin. We sequenced the RT-PCR products from each virus and host marker.

2.2 Haplotype analysis and phylogenetic inference

Median joining haplotype networks were inferred and visualized using the software PopArt v1.7 beta (Leigh and Bryant 2015). Host mitochondrial gene sequences were concatenated and treated as a single locus for haplotype analyses. Association tests between KIGV and *C. trivittatus* haplotypes used the r_d index (Agapow and Burt 2001) and were carried out with the R poppr package with 9,999 permutations (Kamvar et al. 2014). This index is a modification of the index of association (I_A), which was developed to test for linkage disequilibrium, i.e. statistical association between alleles at different loci (Brown et al. 1980; Maynard Smith et al. 1993). It describes the extent to which individuals that possess the same or different haplotypes at one locus are more likely than random to possess the same haplotype or different haplotypes at another. However, because I_A is sensitive to sample size (number of loci being tested), we used the modified r_d (Agapow and Burt 2001), which is not. We estimated population genetic structure (Φ_{ST}) and diversity (π) with MEGA 7 (Kumar et al. 2008).

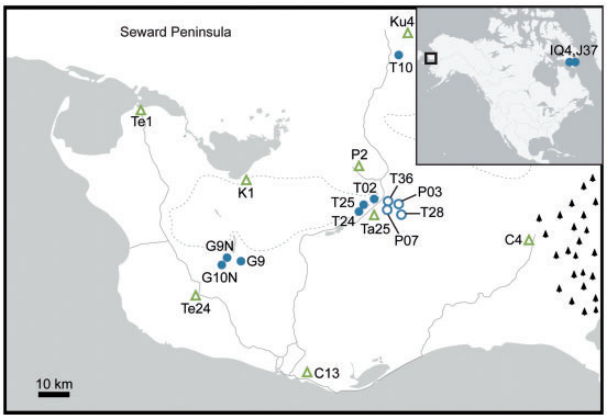

Figure 1. Geographic locations of subarctic pond habitats for phasmavirids (*Bunyavirales*) in the present study. The inset shows the North American study regions of the southern Seward Peninsula, Alaska (open rectangle) and Baffin Island, Arctic Canada (closed blue circles). Open symbols are thaw ponds formed since the 1950s while solid symbols are glacial or ancient thermokarst ponds. Blue circles indicate pond populations of *Kigluaik phantom orthophasmavirus* (KIGV) and green triangles represent pond populations of *Nome phantom orthophasmavirus* (NOMV). Black tree symbols represent the tree line region while the dashed region indicates the Kigluaik Mountains and Bendeleben Mountains (beneath inset).

Tajima's D and Fay and Wu's normalized H estimates were calculated with DNAsp (Librado and Rozas 2009). Isolation by distance effects were estimated by regressing linearized γ_{ST} against log distance among ponds.

Phylogenetic trees were inferred using Mr. Bayes 3.2.6 (with a chain length of five million a subsampling frequency of 1,000 and four heated chains; Ronquist et al. 2012). The Burn in length was 500,000 and the effective sample sizes (ESS) were assessed in Geneious 7. Trees were visualized with FigTree v1.4 (Rambaut 2012). Specimens with sequence for only one segment or putative reassortments were omitted for this analysis. Reassortment or recombination was assessed with GARD (Kosakovsky Pond et al. 2006). GARD is used to test for evidence of recombination at one or more positions within a nucleotide sequence alignment, we extended it to test for both recombination and reassortment among virus genome segments simply by concatenating the virus gene sequences prior to performing the analysis. Tests for the detection of natural selection on codon positions were done using the Fast Unconstrained Bayesian AppRoximation (FUBAR; Murrell et al. 2013), SLAC, FEL and MEME methods of the HyPhy package (Pond and Muse 2005) available on the Datamonkey webserver (Delport et al. 2010). While testing for evidence of positive selection, we concatenated the gene sequences (after removing the putative reassorted sequences) and examined the consensus report of methods with a significance level of 0.05. Temporal signal and the suitability of the data for molecular clock analyses were assessed using TempEst (Rambaut et al. 2016) to regress root to tip distances based on a concatenated complete gene alignment over a five year (2011–5) annual sampling window for KIGV.

3. Results

3.1 Phasmavirids are detected in single hosts and maintain high prevalence in recent subarctic ponds

We detected KIGV RNA markers in 73.4 per cent of 342 larvae (range 40–100%) of *C. trivittatus* sampled from eleven

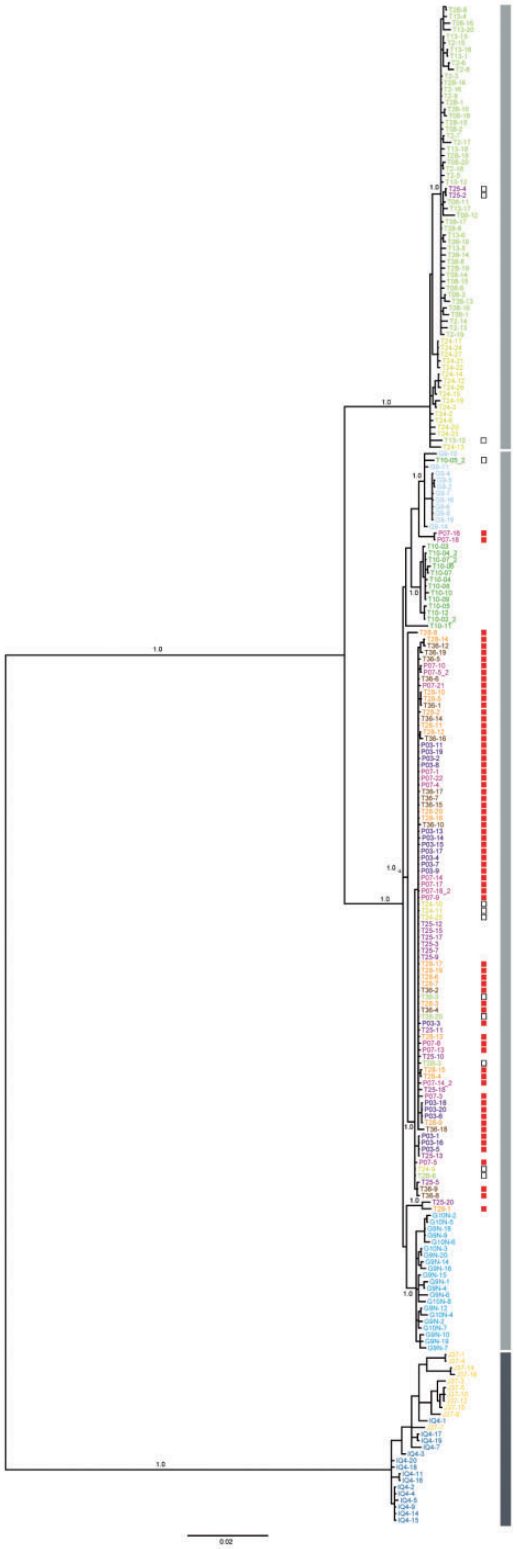

Figure 2. Bayesian summary tree of subarctic phasmavirids (KIGV:*Bunyavirales*). The alignment is based on a concatenation of the four gene sequences in the genome and is midpoint rooted. Node values indicate posterior probability values for strongly supported groupings. Acronyms represent specimens from ponds (see Fig. 1). Specimen names are colored by pond population. Ponds G10N and G9N have been shaded the same, as they are previously connected basins. Red filled blocks indicate sequences from ponds that formed since the 1950s. Open blocks represent proposed sequences from viruses that have recently dispersed from ancient ponds. Gray rectangles show the major clades detected.

populations on the Seward Peninsula and two on Baffin Island (Supplementary Table S1). The association of pond age class with prevalence (using the median value for the site with multiyear samples) was not significant (Fisher's exact test, $P = 0.3365$). Multiyear analysis (2011–2105) of prevalence in the same pond revealed the maximum prevalence in 2011 at 70 per cent and the minimum value of 40 per cent in 2015 for KIGV in *C. trivittatus*. We also screened eight *C.* cf. *flavicans* populations for NOMV and found 51.2 per cent of 172 larvae were positive. In total, we screened 514 subarctic *Chaoborus* larvae from 21 freshwater ponds and found at least 66 per cent were infected with a phasmavirid. Mitochondrial genes from hosts were amplified from each of these extracts. We identified no virus loci at which polymorphisms were supported in both strands, i.e., we found no evidence for infection by multiple strains. We tested for evidence of recombination and genome segment reassortment in KIGV and NOMV virus sequences using GARD (Kosakovsky Pond et al. 2006) and found that four sequences of KIGV (or ~1.2%) and three sequences of NOMV were putative reassortants or assembly errors as their component gene sequences were placed in separate divergent clades. GARD detected no evidence of recombination after the removal of these sequences from the input alignment, and we performed subsequent analyses without them. We also carried out a TempEst analysis on KIGV sequences collected from site T02 each year for five consecutive years (2011–5) in order to test for a signature of clocklike sequence evolution in this virus population and found no detectable temporal signal (R^2=0.0157, slope = −0.0001). We therefore used sequences from all sample dates of site T02 for subsequent analyses.

To confirm that KIGV and NOMV are specifically associated with the two *Chaoborus* host species, we screened freshwater invertebrates collected from ponds in which we sampled phasmavirid-infected *Chaoborus* hosts. We failed to detect sequences of phasmavirids from any of these RNA extracts (Supplementary Table S3). We also screened a third species of *Chaoborus*, *C. americanus*, which we found co-occurring in a recently formed pond with *C. trivittatus*. KIGV amplified from 64 per cent of *C. trivittatus* specimens from this pond, site P7 (18 positive of 28 screened), but was not detected in any of the eleven congeneric *C. americanus* larvae we screened (this is significantly less prevalence using Fisher's exact test, $P = 0.0006$).

3.2 Phylogeny, genetic diversity, and population genetic structure differ with pond age and species

The phylogeny of KIGV revealed three divergent clades: an eastern arctic, and two western subarctic clades (Fig. 2). Viral sequences from older ponds were associated with pond-specific clades supported by large posterior probability values. In contrast, 68 of 70 viral genotypes from new ponds were placed within or closely related to a clade formed by sequences from a single glacial pond (T25). The remaining two viral genotypes from recent ponds formed a unique group. Twelve viral sequences from older ponds were detected within clades of viral sequences from other old ponds, suggesting recent dispersal among old ponds. However, as with the viral sequences from new ponds, a bias appeared to exist as eight of these putatively dispersed sequences were also nested in the clade of sequences from site T25. Similar population patterns to the total evidence phylogeny are apparent from median joining networks based on individual genes of KIGV (Fig. 3). Unlike KIGV, NOMV lacked pond-specific clades (Fig. 4a). In addition, unlike KIGV, recent thaw ponds containing NOMV contained several divergent

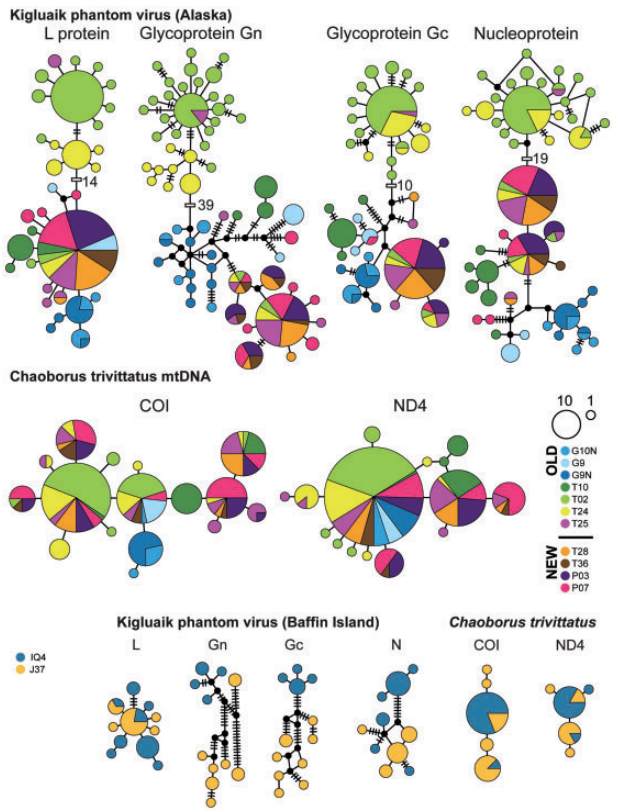

Figure 3. Median joining haplotype networks for pond populations of KIGV (*Bunyavirales*) and their phantom midge host (*Chaoborus trivittatus*). Networks built from haplotypes of four virus genes, L, Gn, Gc, and N, from host populations on the Seward Peninsula in Alaska are shown above networks of the host mitochondrial genes COI and ND4 from the same region. Below, networks constructed from haplotypes of the same virus and host genes collected on Baffin Island, in Iqaluit, Nunavut are shown. Pie graphs are scaled as shown in the key and colored by waterbody. Lines and vertical bars show single mutations.

clusters—that is a clade-specific bias for recent ponds was unapparent for NOMV.

For KIGV in Alaska, all measures of genetic diversity (among populations, within populations and total diversity) were markedly lower for recently formed ponds compared to the measures from older glacial ponds (Table 1). Note that the same population genetic patterns for KIGV exist when using the smaller data set from fewer genes (Table 1). Indeed, the mean among population diversity for recent ponds was 0. Glacial ponds on Baffin Island also had a greater mean total diversity of KIGV than recent Alaskan ponds (by a factor of three). In contrast to KIGV, NOMV from recent ponds had a moderately high mean total diversity (0.016) and the highest mean diversity within populations for any group. For the KIGV-*C. trivittatus* system, the mean genetic diversity of the host was less than 1/10 of the viral diversity for glacial ponds. But, for recent ponds, the diversity of KIGV was identical to the host diversity (0.003). We estimated that NOMV had a 16-fold greater diversity than the mean total diversity of the host's mtDNA from recent ponds.

We tested for association of divergence with distance for each virus by regressing GammaST to distance among sites. There was also no association for KIGV (R^2=0.0051, P = 0.582); however, we found a significant effect of distance with population divergence for NOMV (R^2=0.2166, P = 0.018). To examine virus and host population genetic structure, we estimated the Φ index for groups of ponds by age class. The Φ index is very similar to the classic F_{ST} but it incorporates the nucleotide distance among alleles rather than treating all alleles as equally different to one another (Excoffier et al. 1992). KIGV populations were strongly structured in ancient (glacial) ponds (Φ = 0.772, P < 0.001) and showed no significant genetic structure in recent thaw ponds (Φ < 0.001, n.s.). The only older pond pairwise comparisons that lacked clear haplotype divergence for the four genes were between ponds G10N and G9N with coefficient of differentiation not significantly different from 0 (d=−0.017; SE = 0.037). Glacial pond populations of the KIGV host, *C. trivittatus*, had remarkably strong genetic structure in Alaska (Φ = 0.442, P < 0.001) but showed only two major lineages (compared to three in the virus; Fig. 5). However, as with their viruses, the host populations from recent thaw ponds showed a lack of significant genetic structure (Φ = 0.0009, P = 0.36). NOMV had a significant but modest degree of population genetic structure (Φ = 0.305, P < 0.001). For the NOMV host, *C.* cf. *flavicans*, the mitochondrial genes showed weak but significant differentiation (Φ = 0.127, P < 0.001) as a single haplotype dominated in frequency across the entire region (Fig. 4b). Therefore, we report a clear contrast in genetic structure between the viruses and between host populations from glacial and recent thaw ponds.

We were not successful in attempting to establish cultures of *Chaoborus* in the lab to test for vertical transmission, these consistently arrested at the adult stage—no egg rafts were deposited. However, for each gene of KIGV we rejected the null hypothesis of no association of virus haplotypes with those of the host mtDNA (Fig. 6). The test statistics (r_d=0.04–0.11) significantly exceeded the permutated expectation under no linkage. Association tests for NOMV and its host were not carried out because the host mtDNA was largely composed of a single haplotype.

3.3 Purifying selection is a dominant selective force on phasmavirid evolution

Each of the four KIGV genetic markers (L, Gn, Gc, and N) had a significantly negative Tajima's test scores at P < 0.05 in the recently formed ponds (48 sequences per marker distributed across four of these ponds), but these were not significant in the older ponds (98 sequences per marker distributed across seven old ponds on the Seward Peninsula, AK and an additional 16 sequences from two lakes near Iqaluit on Baffin Island). As a negative Tajima's D does not distinguish between purifying selection and population expansion, we also calculated Fay and Wu's normalized H statistic for each marker, which is not sensitive to demographic changes, and found the distribution was not significantly different from zero, t(3)=−2.305, P = 0.104. The same markers sampled from old ponds is different from zero, t(7)=−4.210, P = 0.004. Since these genomes are segmented, we also examined gene-specific H in recently formed and old ponds. In recently formed ponds, only the Gn gene region exhibited a significant negative departure from neutrality (i.e. a normalized H of 0), one-sample t-test with Holm–Bonferroni P value correction: Gn, t(3)=−11.67, P = 0.005 (non-significant test results were: L, t(3)=2.888, P = 0.189; Gc, t(3)=−2.626, P = 0.189; N, t(3)=−1.204, P = 0.315), and in older ponds, both of the genes encoded on the M segment departed from neutrality, Gn: t(6)=−5.211, P = 0.008; Gc: t(6)=−3.454, P = 0.040 (non-significant tests L, t(6)=−2.2824, P = 0.095; N, t(6)=−2.4819, P = 0.095).

For KIGV, 38.9 per cent of 658 codon positions across 230 sequences had significant evidence of having evolved under purifying selection (Supplementary Table S4). Ten codon positions showed evidence of significant positive selection. Five of these

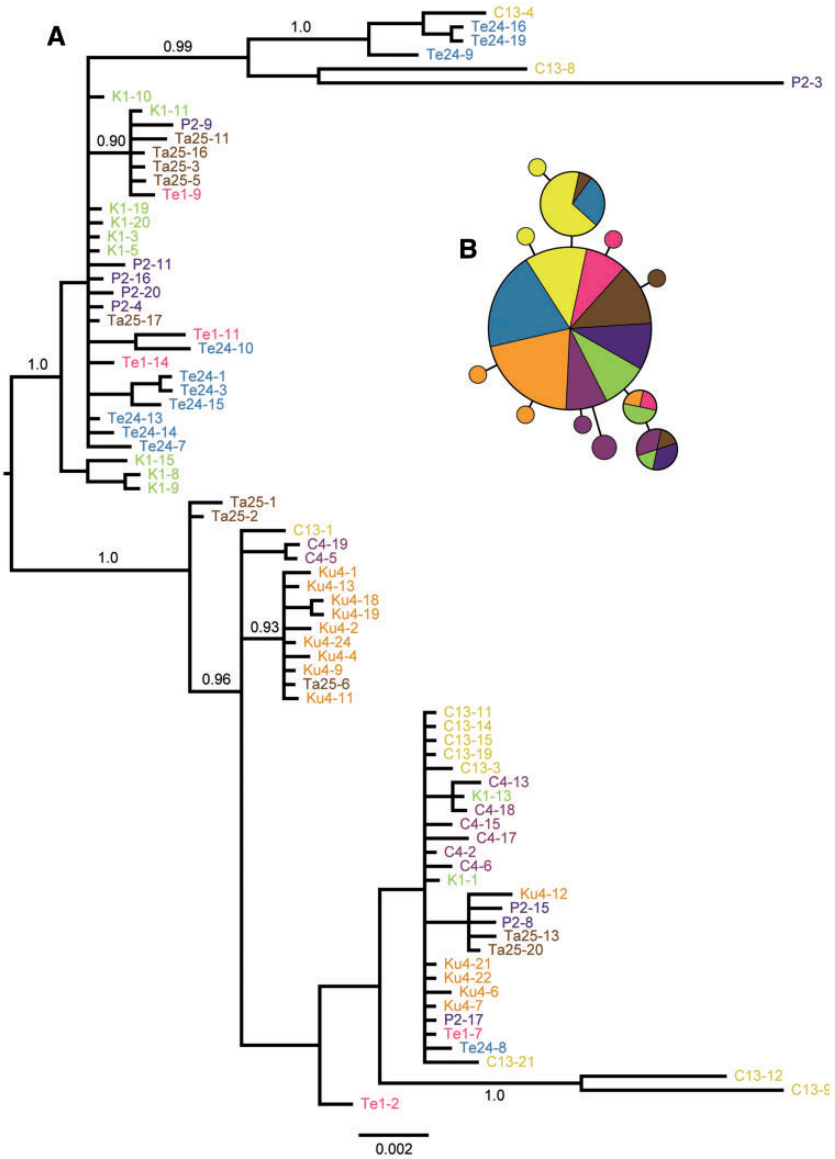

Figure 4. A. Bayesian phylogenetic tree for *Nome phantom orthophasmavirus* (NOMV:*Bunyavirales*) populations in northwest Alaska. Posterior probabilities > 0.90 are shown. B. Median joining haplotype networks for the host's (*Chaoborus* cf. *flavicans*) mitochondrial DNA for phasmavirids (NOMV:*Bunyavirales*). Pie graph sizes are proportional to sample sizes. Lines and vertical bars represent single mutations.

were from the GnGc gene and five were from the N gene region. Nine of the ten significant values were detected using MEME (at a conservative posterior probability value of $P < 0.05$). Two sites had evidence of positive selection for at least three methods. For NOMV, 8.2 per cent of 410 codons across 79 sequences had evidence for significant purifying selection and one site showed evidence for positive selection (only the MEME method proved significant for positive selection).

4. Discussion

Our evidence shows that aquatic RNA viruses can exhibit the opposite biogeographic pattern to that of aquatic DNA viruses. Namely, there can be a lack of population connectivity over evolutionary time despite high prevalence, flying hosts, and a close proximity on the scale of hundreds of meters. Non-random processes, such as virus–host interactions, local selection, and genotypic dispersal bias appear to markedly affect the population genetic structure of related RNA viruses in subarctic ponds. Our results suggest that future predictions of the effects of recent climate change on RNA virus genetic connectivity will need to be specifically tailored to the aquatic virus–host system.

Perhaps the most unexpected finding was the pronounced difference in the degree of genetic structure between old and recently formed ponds for KIGV. Most likely, this pattern is a reflection of the host's dispersal history. Even though the host, *C. trivittatus*, is among the largest of the phantom midges with high population densities, our results suggest that their populations appear much less connected by dispersal than populations of other phantom midges (Berendonk et al. 2009). As most regional pond populations of viruses that we examined were within dispersal range of flying hosts (evidenced by colonization of recently formed ponds), dispersal appears to be a weak determinant of population genetic structure. This pattern is consistent with evolution-based priority effects whereby resident lineages have a local fitness advantage over non-resident

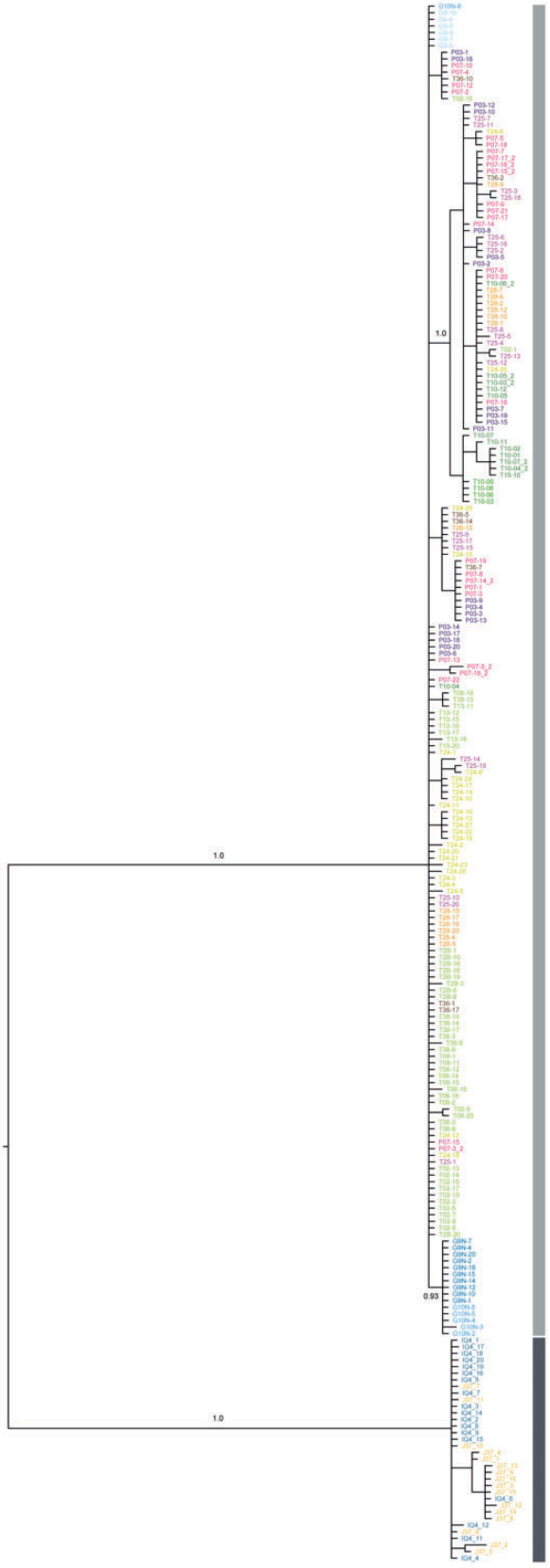

Figure 5. Bayesian phylogenetic tree of host mitochondrial DNA. Specimen names for phantom midge larvae (*Chaoborus trivittatus*) are colored by pond population. Note that the scale is different than for the viral tree and that identical sequences for the mtDNA tree have been given short tips by the software used. Gray rectangles show the major clades detected.

lineages (De Meester et al. 2016). Genotype-associated dispersal bias also may be contributing to the population genetic structure of KIGV. The only two older class ponds (G10N and G9N) that lacked differentiation for KIGV are former basins of a single, previously larger pond. These basins may retain a physical connection during spring melt. The finding that a single genotype is prevalent in the four recently formed ponds and in site T25 is consistent with a dispersal bias. Indeed, a similar bias appears to occur for older ponds. The shallow depth of the genetic variation in site T25 indicates that it too has likely undergone recent colonization by the T25 clade despite being an older pond. The mechanism behind this non-random dispersal is unclear but it is not merely an artifact of geographic distance as the largest pond in the sample is ~500 m closer than site T25 to the recently formed ponds but shows almost no evidence of recent dispersal into these ponds. In addition, T25 is <125 m from site T24 but these ponds contain divergent arrays of KIGV haplotypes with almost no shared haplotypes. Nor is variation in prevalence a factor as prevalence is universally high in ponds. Our mtDNA evidence indicates that genetically diverse host females are laying egg rafts in recently formed ponds. Non-random gene flow has been detected before in RNA viruses (Duggal et al. 2015) where a clade of West Nile Virus has been spreading at a rate 10 per cent greater than other clades. For influenza A, gene flow at the intra-continental scale is strongly restricted by geographic distance but there is surprising variation in levels of gene flow among pairs of nearby localities, suggesting ecological factors play a role here as well (Lam et al. 2012). In the present case, the apparent bias is much more severe as nearly all the recently dispersed genotypes are from a single genotype and clade.

Strong genetic structure in viruses in the absence of host genetic structure has been reported for both vertebrate and for invertebrate hosts (Biek et al. 2006; Stenger et al. 2016). In the present study, the viruses generally showed an elevated population genetic structure compared to their hosts. Although mitochondrial and bunyavirid genomes are both cytoplasmic elements, the viral genomes have a reduced N_e and elevated mutation rate relative to the host genes. As such, near fixation among viral populations in this system may presage a lack of connectivity of host zooplankton populations in older pond habitats. Indeed, RNA viruses have been used as proxies for insect host population questions (Stenger et al. 2016). The use of such viral proxies is relevant only when there is a close virus–host association as in the case of vertical transmission. In the present case, high viral prevalence (up to 100%, Supplementary Table S1) in early larval stages in old and newly formed ponds (where eggs must be deposited to colonize), the strong host specificity (even under coexistence with congeneric larvae with similar life histories), the association of viral genes and host mtDNA genes and the mirroring of host population genetic structure are consistent with vertical transmission via egg rafts and larvae. Still, we do not rule out the existence of horizontal transmission in these systems and transmission from adult to egg has yet to be directly observed. The single deep split in the Alaskan KIGV virus gene tree that is absent from the host mtDNA may be explained by a horizontal infection from a host maternal lineage that is unrepresented in the present data. Finally, host switching has apparently occurred in the *Orthophasmaviruses*, as KIGV and NOMV are not sister species.

The host for NOMV, *C.* cf. *flavicans*, appears adapted to shallow (<1 m) ponds that are less permanent than the glacial ponds that are typical of KIGV-*C. trivittatus* system (Taylor et al. 2016). *C.* cf. *flavicans* has a shorter generation time than *C.*

Table 1. Measures of diversity (π) for populations of subarctic phasmavirids (*Bunyavirales*) and their hosts. Standard error estimates (from bootrapping) are shown in parentheses. Viruses are *Kigluaik phantom orthophasmavirus* (KIGV) and *Nome phantom orthophasmavirus* (NOMV), while hosts are species of phantom midges (*Chaoborus*). KIGV analyses are separated into age classes and regions (Alaska and Baffin). In addition, KIGV has a second row of measurements with the same genes used for NOMV estimates (that is, excluding the Gn gene region) to permit a more direct comparison between viral species.

Populations	Mean total diversity	Mean interpop diversity	Mean diversity within populations	Coefficient of differentiation
KIGV-glacial	0.035 (0.003)	0.024 (0.002)	0.010 (0.001)	0.708 (0.011)
	0.025 (0.003)	0.018 (0.002)	0.007 (0.001)	0.722 (0.011)
KIGV-recent	0.003 (0.001)	0.000 (0.000)	0.003 (0.001)	−0.016 (0.009)
	0.002 (0.001)	0.000 (0.000)	0.002 (0.001)	−0.024 (0.009)
KIGV Baffin	0.009 (0.001)	0.001 (0.001)	0.008 (0.001)	0.146 (0.059)
	0.008 (0.002)	0.002 (0.001)	0.005 (0.001)	0.278 (0.054)
KIGV host glacial	0.003 (0.001)	0.001 (0.001)	0.001 (0.000)	0.472 (0.070)
KIGV host recent	0.003 (0.001)	0.000 (0.000)	0.003 (0.001)	0.020 (0.038)
KIGV host Baffin	0.002 (0.001)	0.000 (0.000)	0.001 (0.001)	0.185 (0.081)
NOMV	0.016 (0.002)	0.003 (0.001)	0.013 (0.002)	0.207 (0.021)
NOMV host recent	0.001 (0.000)	0.000 (0.000)	0.001 (0.000)	−0.008 (0.162)

trivittatus, and may be better adapted to frequent local extinctions and recolonizations. Unlike the KIGV system, we presently lack older reference populations for NOMV and the host *C. cf. flavicans*. Thus, the diversity pattern that we see in recent thaw ponds may be typical for a system where most populations are frequently recolonized from multiple sources.

The evolutionary features of KIGV and NOMV in the present study are more clearly aligned with persistent viruses than with acute viruses (Holmes 2009). Whereas acute infections require connectivity to avoid extinction, we found little evidence of connectivity or of population extinction with older insular habitats. Evolutionary rates are normally observable among years in acute RNA viruses but the studied viruses lack observable evolution over five or more years. Timescale estimation may be hindered by the action of purifying selection that we found. Virulence is often high with acute infections while prevalence is low (< 5%). Under high virulence with some vertical inheritance we expect newly founded populations to have reduced prevalence because healthy hosts should be more effective at dispersing and reproducing than infected individuals. However, the universally high prevalence of KIGV and NOMV (which includes populations of newly formed ponds; Supplementary Table S1), suggests weak virulence. We did find one observation that may be more indicative of acute than persistent infection—positive selection in KIGV genomes. The Gn region, for example, has been implicated in mediating viral attachment to host receptors in bunyavirids, so we may expect ongoing selection at this gene with acute viruses. Still, there is nothing known about the functionality of these residues in phasmavirids. Moreover, the prevailing evolutionary signal across the genomes of these viruses is purifying selection.

While we have studied just two natural virus–host systems, the arthropod hosts (phantom midges) are important predators of zooplankton in aquatic ecosystems worldwide. Moreover, there is likely to be a vast diversity of unexplored RNA virus–arthropod zooplankton associations (Ballinger et al. 2013; Shi et al. 2016). Our results suggest that the ongoing losses and gains of tundra ponds associated with climate change and landscape evolution can have differing and unexpected effects on related aquatic virus–host systems. For KIGV, the formation of thaw ponds appears to markedly reduce haplotype diversity. We expect that older pond loss will be associated with loss of private clades of viruses that we found in this system. In contrast, NOMV expanded with its host into recent shallow recent thaw tundra ponds with substantial diversity. In neither case did we find that aquatic viruses in similar habitats are connected by dispersal regardless of proximity as was found for DNA viruses. Instead, we identified prevalent viruses with flighted hosts showing little or no apparent genetic connectivity. Our results indicate that sampling across lake/pond age will be important for understanding the evolutionary dynamics of aquatic RNA viruses. Finally, understanding the effects of climate change on these aquatic viruses will be likely complicated by differing genotypic responses and RNA virus–host interactions.

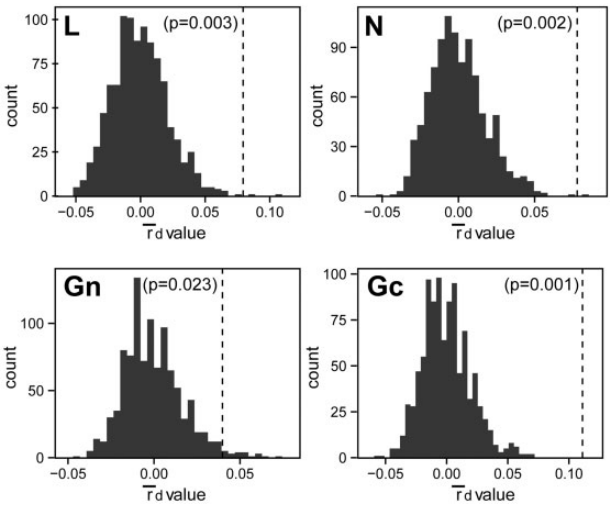

Figure 6. Association tests between gene sequences for *Kigluaik phantom orthophasmavirus* (*Bunyavirales*) and host (phantom midges, *Chaoborus trivittatus*) mitochondrial DNA sequences. The test statistic is r_d with significance (in parentheses) determined using 9,999 permutations. Gene abbreviations are given for each test of association.

Acknowledgements

We are grateful to Art Borkent, Alexey Kotov, William Nichols, Skyler Resendez for assistance collecting *Chaoborus* samples and Jack Zhan for help with coding. We also thank the Nunavut Research Institute for logistical support and two anonymous reviewers for suggestions that improved

the manuscript. This work was supported by the National Science Foundation (ARC 1023334).

Conflict of interest: None declared.

Data availability

Chaoborus and *Orthophasmavirus* nucleotide sequences generated during this work have been submitted to GenBank under accession numbers KY566224–KY567977.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

References

Adams, M. J. et al. (2017) 'Changes to Taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses', *Archives of Virology*. doi: 10.1007/s00705-017-3358-5.

Agapow, P.-M., and Burt, A. (2001) 'Indices of Multilocus Linkage Disequilibrium', *Molecular Ecology Notes*, 1: 101–2.

Ayieko, M. A., and Oriaro, V. (2008) 'Consumption, Indigenous Knowledge and Cultural Values of the Lakefly Species within the Lake Victoria Region', *African Journal of Environmental Science and Technology*, 2: 282–6.

Ballinger, M. J. et al. (2014) 'Discovery and Evolution of Bunyavirids in Arctic Phantom Midges and Ancient Bunyavirid-Like Sequences in Insect Genomes', *Journal of Virology*, 88: 8783–94.

—— et al. (2013) 'Selectively Maintained Paleoviruses in Holarctic Water Fleas Reveal an Ancient Origin for Phleboviruses', *Virology*, 446: 276–82.

Berendonk, T. U., Spitze, K., and Kerfoot, W. C. (2009) 'Ephemeral Metapopulations Show High Genetic Diversity at Regional Scales', *Ecology*, 90: 2670–5.

Biek, R., Drummond, A. J., and Poss, M. (2006) 'A Virus Reveals Population Structure and Recent Demographic History of Its Carnivore Host', *Science*, 311: 538–41.

Borkent, A. (1979) 'Systematics and Bionomics of the Species of the Subgenus Schadonophasma Dyar and Shannon (Chaoborus, Chaoboridae, Diptera)', *Quaestiones Entomologicae*, 15: 122–255.

Bouchard, F. et al. (2013) 'Vulnerability of Shallow Subarctic Lakes to Evaporate and Desiccate When Snowmelt Runoff is Low', *Geophysical Research Letters*, 40: 6112–7.

Brown, A. H. D., Feldman, M. W., and Nevo, E. (1980) 'Multilocus Structure of Natural Populations of *Hordeum spontaneum*', *Genetics*, 96: 523–36.

Cook, S. et al. (2013) 'Novel Virus Discovery and Genome Reconstruction from Field RNA Samples Reveals Highly Divergent Viruses in Dipteran Hosts', *PLoS One*, 8/11: 1–22.

de Cárcer, D. A. et al. (2015) 'Biodiversity and Distribution of Polar Freshwater DNA Viruses', *Science Advances*, 1: e1400127.

De Meester, L. et al. (2016) 'Evolving Perspectives on Monopolization and Priority Effects', *Trends in Ecology & Evolution*, 31: 136–46.

Delport, W. et al. (2010) 'Datamonkey 2010: A Suite of Phylogenetic Analysis Tools for Evolutionary Biology', *Bioinformatics*, 26: 2455–7.

Duggal, N. K. et al. (2015) 'Genotype-Specific Variation in West Nile Virus Dispersal in California', *Virology*, 485: 79–85.

Dupuis, D. D., Svensson, J.-E., and Taylor, D. J. (2008) 'The Cryptic Origins of Environment-Indicating Phantom Midges (Chaoborus)', *Limnology and Oceanography*, 53: 236–43.

Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992) 'Analysis of Molecular Variance Inferred from Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data', *Genetics*, 131: 479–91.

Federenko, A. Y., and Swift, M. C. (1972) 'Comparative Biology of *Chaoborus americanus* and *Chaoborus trivittatus* in Eunice Lake, British Columbia', *Limnology and Oceanography*, 17: 721–30.

Holland, J. J. (2006) 'Transitions in Understanding of RNA Viruses: A Historical Perspective', *Current Topics in Microbiology and Immunology*, 299: 371–401.

Holmes, E. C. (2009) *The Evolution and Emergence of RNA Viruses*. New York, NY: Oxford University Press.

Hunt, S., Yu, Z., and Jones, M. (2013) 'Lateglacial and Holocene Climate, Disturbance and Permafrost Peatland Dynamics on the Seward Peninsula, Western Alaska', *Quaternary Science Reviews*, 63: 42–58.

Junglen, S. (2016) 'Evolutionary Origin of Pathogenic Arthropod-Borne Viruses – A Case Study in the Family Bunyaviridae', *Current Opinion in Insect Science*, 16: 81–6.

Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. (2014) 'Poppr: An R Package for Genetic Analysis of Populations with Clonal, Partially Clonal, and/or Sexual Reproduction', *PeerJ*, 2: e281.

Katoh, K., and Standley, D. M. (2014) 'MAFFT: iterative refinement and additional methods', in Russell D. J. (Ed.) *Multiple Sequence Alignment Methods*, pp. 131–46. New York, NY: Humana Press.

Kosakovsky Pond, S. L. et al. (2006) 'GARD: A Genetic Algorithm for Recombination Detection', *Bioinformatics*, 22: 3096–8.

Kumar, S. et al. (2008) 'MEGA: A Biologist-Centric Software for Evolutionary Analysis of DNA and Protein Sequences', *Briefings in Bioinformatics*, 9: 299–306.

Lam, T. T.-Y. et al. (2012) 'Migratory Flyway and Geographical Distance are Barriers to the Gene Flow of Influenza Virus Among North American Birds', *Ecology Letters*, 15: 24–33.

Laybourn-Parry, J., Hofer, J. S., and Sommaruga, R. (2001) 'Viruses in the Plankton of Freshwater and Saline Antarctic Lakes', *Freshwater Biology*, 46: 1279–87.

Leigh, J. W., and Bryant, D. (2015) 'popart: Full-Feature Software for Haplotype Network Construction', *Methods in Ecology and Evolution*, 6: 1110–6.

Li, C.-X. et al. (2015) 'Unprecedented Genomic Diversity of RNA Viruses in Arthropods Reveals the Ancestry of Negative-Sense RNA Viruses', *eLife*, 4: e05378.

Librado, P., and Rozas, J. (2009) 'DnaSP v5: A Software for Comprehensive Analysis of DNA Polymorphism Data', *Bioinformatics*, 25: 1451–2.

Marklewitz, M. et al. (2015) 'Evolutionary and Phenotypic Analysis of Live Virus Isolates Suggests Arthropod Origin of a Pathogenic RNA Virus Family', *Proceedings of the National Academy of Sciences of the United States of America*, 112: 7536–41.

Maynard Smith, J. et al. (1993) 'How Clonal are Bacteria?', *Proceedings of the National Academy of Sciences of the United States of America*, 90: 4384–8.

Medeiros, A. S., and Quinlan, R. (2011) 'The Distribution of the Chironomidae (Insecta: Diptera) Along Multiple Environmental Gradients in Lakes and Ponds of the Eastern Canadian Arctic', *Canadian Journal of Fisheries and Aquatic Sciences*, 68: 1511–27.

—— et al. (2014) 'Biological and Nutrient Responses to Catchment Disturbance and Warming in Small Lakes Near the Alaskan Tundra–Taiga Boundary', *The Holocene*, 24: 1308–19.

Moore, M. V. (1986) 'Method for Culturing the Phantom Midge, *Chaoborus* (Diptera: Chaoboridae), in the Laboratory', *Aquaculture*, 56: 307–16.

Murrell, B. et al. (2013) 'FUBAR: A Fast, Unconstrained Bayesian Approximation for Inferring Selection', *Molecular Biology and Evolution*, 30: 1196–205.

Nichol, S. T., Rowe, J. E., and Fitch, W. M. (1993) 'Punctuated Equilibrium and Positive Darwinian Evolution in Vesicular Stomatitis Virus', *Proceedings of the National Academy of Sciences of the United States of America*, 90/22: 10424–8.

Nouri, S. et al. (2016) 'Diverse Array of New Viral Sequences Identified in Worldwide Populations of the Asian Citrus Psyllid (*Diaphorina citri*) Using Viral Metagenomics', *Journal of Virology*, 90: 2434–45.

Paez-Espino, D. et al. (2016) 'Uncovering Earths Virome', *Nature*, 536: 425–30.

Pond, S. L. K., and Muse, S. V. (2005) 'Statistical methods in molecular evolution', in Nielsen R. (Ed.) *HyPhy: Hypothesis testing using phylogenies*, pp. 125–81. New York, NY: Springer.

Rambaut, A. (2012) *FigTree v1. 4. Molecular Evolution, Phylogenetics and Epidemiology*. Edinburgh: University of Edinburgh, Institute of Evolutionary Biology.

—— et al. (2016) 'Exploring the Temporal Structure of Heterochronous Sequences using TempEst (Formerly Path-O-Gen)', *Virus Evolution*, 2: vew007.

Riordan, B., Verbyla, D., and McGuire, A. D. (2006) 'Shrinking Ponds in Subarctic Alaska Based on 1950–2002 Remotely Sensed Images', *Journal of Geophysical Research*, 111: G04002.

Ronquist, F. et al. (2012) 'MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space', *Systematic Biology*, 61: 539–42.

Shi, M. et al. (2016) 'Redefining the Invertebrate RNA Virosphere', *Nature*, 540: 539–43.

Smith, L. C. et al. (2005) 'Disappearing Arctic Lakes', *Science*, 308: 1429.

Smol, J. P., and Douglas, M. S. (2007) 'Crossing the Final Ecological Threshold in High Arctic Ponds', *Proceedings of the National Academy of Sciences of the United States of America*, 104: 12395–7.

Stenger, D. C. et al. (2016) 'Sequence Polymorphism in an Insect RNA Virus Field Population: A Snapshot from a Single Point in Space and Time Reveals Stochastic Differences Among and Within Individual Hosts', *Virology*, 498: 209–17.

Taylor, D. J. et al. (2016) 'Climate-Associated Tundra Thaw Pond Formation and Range Expansion of Boreal Zooplankton Predators', *Ecography*, 39: 43–53.

Weber de Melo, V., Sheikh, A. H. et al. (2015) 'Spatiotemporal Dynamics of Puumala Hantavirus Associated with Its Rodent Host, *Myodes glareolus*', *Evolutionary Applications*, 8/6: 545–59.

Webster, C. L. et al. (2015) 'The Discovery, Distribution, and Evolution of Viruses Associated with *Drosophila melanogaster*', *PLoS Biology*, 13: e1002210.

Xie, P., Iwakuma, T., and Fujii, K. (1998) 'Studies on the Biology of *Chaoborus flavicans* (Meigen)(Diptera: Chaoboridae) in a Fish-Free Eutrophic Pond, Japan', *Hydrobiologia*, 368: 83–90.

Yoshikawa, K., and Hinzman, L. D. (2003) 'Shrinking Thermokarst Ponds and Groundwater Dynamics in Discontinuous Permafrost Near Council, Alaska', *Permafrost and Periglacial Processes*, 14: 151–60.