Towards optimized viral metagenomes for double-stranded and single-stranded DNA viruses from challenging soils

Gareth Trubl [1, 2] , **Simon Roux** [3] , **Natalie Solonenko** [1] , **Yueh-Fen Li** [1] , **Benjamin Bolduc** [1] , **Josué Rodríguez-Ramos** [1, 4] , **Emiley A. Eloe-Fadrosh** [3] , **Virginia I. Rich** [Corresp., 1] , **Matthew B. Sullivan** [Corresp. 1, 5]

[1] Department of Microbiology, Ohio State University, Columbus, Ohio, United States

[2] Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California, United States

[3] Joint Genome Institute, Department of Energy, Walnut Creek, California, United States

[4] Department of Soil and Crop Sciences, Colorado State University, Fort Collins, Colorado, United States

[5] Civil, Environmental and Geodetic Engineering, Ohio State University, Columbus, Ohio, United States

Corresponding Authors: Virginia I. Rich, Matthew B. Sullivan
Email address: virginia.isabel.rich@gmail.com, mbsulli@gmail.com

Soils impact global carbon cycling and their resident microbes are critical to their biogeochemical processing and ecosystem outputs. Based on studies in marine systems, viruses infecting soil microbes likely modulate host activities via mortality, horizontal gene transfer, and metabolic control. However, their roles remain largely unexplored due to technical challenges with separating, isolating, and extracting DNA from viruses in soils. Some of these challenges have been overcome by using whole genome amplification methods and while these have allowed insights into the identities of soil viruses and their genomes, their inherit biases have prevented meaningful ecological interpretations. Here we experimentally optimized steps for generating quantitatively-amplified viral metagenomes to better capture both ssDNA and dsDNA viruses across three distinct soil habitats along a permafrost thaw gradient. First, we assessed differing DNA extraction methods (PowerSoil, Wizard mini columns, and cetyl trimethylammonium bromide) for quantity and quality of viral DNA. This established PowerSoil as best for yield and quality of DNA from our samples, though ~1/3 of the viral populations captured by each extraction kit were unique, suggesting appreciable differential biases among DNA extraction kits. Second, we evaluated the impact of purifying viral particles after resuspension (by cesium chloride gradients; CsCl) and of viral lysis method (heat vs bead-beating) on the resultant viromes. DNA yields after CsCl particle-purification were largely non-detectable, while unpurified samples yielded 1–2-fold more DNA after lysis by heat than by bead-beating. Virome quality was assessed by the number and size of metagenome-assembled viral contigs, which showed no increase after CsCl-purification, but did from heat lysis relative to bead-beating. We also evaluated sample preparation protocols for ssDNA virus recovery. In both CsCl-purified and non-purified samples, ssDNA viruses were successfully recovered by using the Accel-NGS 1S Plus Library Kit. While ssDNA viruses were identified in all three soil types, none were identified in the samples that used bead-beating, suggesting this lysis method may impact recovery. Further, 13 ssDNA vOTUs were identified compared to 582 dsDNA vOTUs, and the ssDNA vOTUs only accounted for ~4% of the assembled reads, implying dsDNA viruses were dominant in these samples. This optimized approach was combined with the previously published viral resuspension protocol into a sample-to-virome protocol for soils now available at protocols.io, where community feedback creates ‘living’ protocols. This collective approach will be particularly valuable given the high physicochemical variability of soils, which will may require considerable soil type-specific optimization. This optimized protocol provides a starting place for developing quantitatively-amplified viromic datasets and will help enable viral ecogenomic studies on organic-rich soils.

Introduction

Optimization of experimental methods to generate viral-particle metagenomes (viromes) from aquatic samples has enabled robust ecological analyses of marine viral communities (reviewed in Brum and Sullivan 2015; Sullivan, Weitz, and Wilhelm 2016; Hayes et al. 2017). In parallel, optimization of informatics methods to identify and characterize viral sequences has advanced viral sequence recovery from microbial-cell metagenomes, as well as virome analyses (Edwards and Rohwer 2005; Wommack et al. 2012; Roux et al. 2015; Brum & Sullivan, 2015; Roux et al. 2016; Bolduc et al. 2016; Ren et al. 2017; Amgarten et al. 2018). Application of these methods with large-scale sampling (Brum et al. 2015; Roux et al. 2016) has revealed viruses as important members of ocean ecosystems acting through host mortality, gene transfer, and direct manipulation of key microbial metabolisms including photosynthesis and central carbon metabolism during infection, via expression of viral-encoded 'auxiliary metabolic genes' (AMGs). More recently, the abundance of several key viral populations was identified as the best predictor of global carbon (C) flux from the surface oceans to the deep sea (Guidi et al. 2016). This finding suggests that viruses may play a role beyond the viral shunt and help form aggregates that may store C long-term. These discoveries in the oceans have caused a paradigm shift in how we view viruses: no longer simply disease agents, it is now clear that viruses play central roles in ocean ecosystems and help regulate global nutrient cycling.

In soils, however, viral roles are not so clear. Soils contain more C than all the vegetation and the atmosphere combined (between 1500–2400 gigatons; Lehmann and Kleber 2015), and soil viruses likely also impact C cycling, as their marine counterparts do. However, our knowledge about soil viruses remains limited due to the dual challenges of separating viruses from the highly heterogeneous soil matrix, while minimizing DNA amplification inhibitors (e.g. humics; reviewed in Williamson et al. 2017). For these reasons, most soil viral work is limited to direct counts and morphological analyses (i.e. microscopy observations), from which we have learned (i) there are 107–109 viruses/g soil, (ii) viral morphotype richness is generally higher in soils than in aquatic ecosystems, and (iii) viral abundance correlates with soil moisture, organic matter content, pH, and microbial abundance (reviewed in Williamson 2017; Narr et al. 2017). Thus, while sequencing data for soil viruses are hard to come by, such high particle counts and patterns suggest that viruses also play important ecosystems roles in soils.

The first barrier to obtaining sequence data for soil viruses is simply separating the viral particles from the soil matrix, and then accessing their nucleic acids. Viral resuspension is unlikely to be universally solvable with a single approach due to high variability of soil properties (e.g. mineral content and cation exchange capacity) impacting virus-soil interactions. There have been independent efforts to optimize virus resuspension methods tailored to specific soil types, and employing a range of resuspension methods (reviewed in Narr et al. 2017; Pratama and van Elsas, 2018). Once viruses are separated, extraction of their DNA must surmount the additional challenges of co-extracted inhibitors (hampering subsequent molecular biology, as previously described for soil microbes; Narayan et al. 2016; Zielińska et al. 2017), and low DNA yields.

While little empirical data are available for inhibitors in soil viral extractions, there have been a diversity of approaches to compensate for low DNA yields. Two widely used methods are multiple displacement amplification (MDA; 'whole genome' amplification using the phi29 polymerase) and random priming-mediated sequence-independent single-primer amplification (RP-SISPA). Both allow qualitative observations of viral sequences, but preclude quantitative ecological inferences. Specifically, MDA causes dramatic shifts in relative abundances of DNA templates, which impact subsequent estimates of viral populations diversity, and, most dramatically, over-amplify ssDNA viruses (Binga, Lasken, and Neufeld, 2008; Yilmaz, Allgaier, and Hugenholtz 2010; Kim, Whon, and Bae 2013; Marine et al. 2014). RP-SISPA is biased towards the most abundant viruses or largest genomes, and leads to uneven coverage along the amplified genomes (Karlsson, Belák, and Granberg 2013). More recently, quantitative amplification methods have emerged that use transposon-mediated tagmentation (Nextera, for dsDNA; Trubl et al. 2018; Segobola et al. 2018) or acoustic shearing to fragment and a custom adaptase (Accel-NGS 1S Plus, for dsDNA and ssDNA; Roux et al. 2016; Rosario et al. 2018) to ligate adapters to DNA templates, before PCR amplification is used to obtain enough material for sequencing. These approaches have successfully amplified as little as 1 picogram (Nextera XT; Rinke et al. 2016) and 100 nanograms (Accel-NGS 1S Plus; Kurihara et al. 2014) of input DNA for viromes while maintaining the relative abundances of templates.

We previously optimized a viral resuspension method for three soil habitats (palsa, bog, and fen, spanning a permafrost thaw gradient; Trubl et al. 2016). Given emerging quantitative, low-input DNA library construction options, we sought here to characterize how the choice of methods for viral particle purification, lysis and DNA extraction impacted viral DNA yield and quality, and resulting virome diversity. We tested three different DNA extraction methods, and then two virion lysis methods with and without further particle purification. The extracted DNA was prepared for sequencing using the Accel-NGS 1S Plus kit, generating quantitative soil viromes including both ssDNA and dsDNA viruses, enabling a robust comparison of the different protocols tested.

Methods

Field site and sampling

Stordalen Mire (68.35°N, 19.05°E) is a peat plateau in Arctic Sweden in a zone of discontinuous permafrost. Peat depth ranges from 1–3 meters (Johansson et al. 2006; Normand et al. 2017). Habitats broadly span three stages of permafrost thaw: palsa (drained soil, dominated by small shrubs, and underlain by intact permafrost), bog (partially inundated peat, dominated by Sphagnum moss, and underlain by partially thawed permafrost), and fen (fully inundated peat, dominated by sedges, and with no detectable permafrost at <1 m) (further described in Hodgkins et al. 2014). These soils vary chemically (Hodgkins et al., 2014; Normand et al. 2017; Wilson et al. 2017), hydraulically (Christensen et al. 2004; Malmer et al. 2005; Olefeldt et al. 2012; Jonasson et al. 2012), and biologically (Mondav et al. 2014; McCalley et al. 2014; Mondav et al. 2017; Woodcroft et al. 2018), creating three distinct habitats. Soil was collected with an 11 cm-diameter custom circular push corer at palsa sites, and with a 10 cm × 10 cm square Wardenaar corer (Eijkelkamp, The Netherlands) at the bog and fen sites. Three cores from each habitat were processed using clean techniques described previously (Trubl et al. 2016) and cut in five-centimeter increments from 1–40 cm for palsa and 1–80 cm for bog and fen cores. Samples were flash-frozen in liquid nitrogen and kept at –80°C until processing. The sampled palsa, bog, and fen habitats were directly adjacent, such that all cores were collected within a 120 m radius. For this work, viruses were analyzed from 20–24 cm deep peat, from three cores at each of the three habitats. For Experiment 1 (DNA extraction), 18 samples were used (9 bog and 9 fen), with 10 ± 1 g of soil per sample. For Experiment 2 (virion lysis and purification), 36 samples were used (12 palsa, 12 bog, and 12 fen) with 7.5 ± 1 g of soil per sample.

Experiment 1: Optimizing DNA extraction

Viruses were resuspended using a previously optimized method for these soils (Trubl et al. 2016) with minor adjustments. Briefly, 10 ml of a 1% potassium citrate resuspension buffer amended with 10% phosphate buffered-saline, 5 mM ethylenediaminetetraacetic acid, and 150 mM magnesium sulfate was added to 10 ± 0.5 g peat. Viruses were physically dispersed via 1 min of vortexing, 30 s of manual shaking, and then 15 min of shaking at 400 rpm at 4 °C. The samples were then centrifuged for 20 min at 1,500 ×g at 4 °C to pellet debris, and the supernatant was transferred to new tubes. The resuspension steps above were repeated two more times and the supernatants were combined, and then filtered through a 0.2 µm polyethersulfone membrane filter to remove particles and cells and transferred into a new 50 ml tube. The filtrate was then purified via overnight treatment with DNase I (ThermoFisher, Waltham, Massachusetts) at a 1:10 dilution at 4°C, inactivated by adding a final concentration of 10 mM EDTA and EGTA and mixing for 1 hour. All viral particles were further purified by CsCl density gradients, established with five CsCl density layers of ρ 1.2, 1.3, 1.4, 1.5, and 1.65 g/cm3; we included a 1.3 g/cm3 CsCl layer to collect ssDNA viruses (Thurber et al. 2009). After density gradient centrifugation of the viral particles, we collected and pooled the 1.3-1.52 g/cm3 range from the gradient for viral DNA extraction. The viral DNA was extracted (same elution volume) using one of three methods: Wizard mini columns (Wizard; Promega, Madison, WI, products A7181 and A7211), cetyl trimethylammonium bromide (CTAB; Porebski, Bailey, and Baum 1997), or DNeasy PowerSoil DNA extraction kit with heat lysis (10 min incubation at 70˚C, vortexing for 5 s, and 5 min more of incubation at 70˚C) (PowerSoil; Qiagen, Hilden, Germany, product 12888). The extracted DNA was further cleaned up with AMPure beads (Beckman Coulter, Brea, CA, product A63881). DNA purity was assessed with a Nanodrop 8000 spectrophotometer (Implen GmbH, Germany) by the reading of A260/A280 and A260/A230, and quantified using a Qubit 3.0 fluorometer (Invitrogen, Waltham, Massachusetts). DNA sequencing libraries were prepared using Swift Accel-NGS 1S Plus DNA Library Kit (Swift BioSciences, Washtenaw County, Michigan), and libraries were determined to be ‘successful’ if there was a smooth peak on the Bioanalyzer with average fragment size of <1kb (200–800 bp ideal) and minimal-to-no secondary peak at ~200 bp (representing concatenated adapters) (Fig.

S1), and <20 PCR cycles were required for sequencing. Six libraries were successful (two from bog and four from fen) and required 15 PCR cycles. The successful libraries were sequenced using Illumina HiSeq (300 million reads, 2 x 100 bp paired-end) at JP Sulzberger Columbia Genome Center.

Experiment 2: Optimizing particle lysis and purification

Viromes were generated as in Experiment 1 with minor changes. First, viruses were resuspended as described for Experiment 1, except half of the samples were not purified with CsCl density gradient centrifigation. Second, DNA was extracted from all samples using the PowerSoil method, but the physical method of particle lysis was tested by half of the samples undergoing the standard heat lysis as above and the other half undergoing the alternative PowerSoil bead-beating step (with 0.7 mm garnet beads). Third, the extracted DNA was further cleaned up with DNeasy PowerClean Pro Cleanup Kit (Qiagen, Hilden, Germany, product 12997), instead of AMPure beads. Assessment of microbial contamination was done via qPCR (pre and post-cleanup) with primer sets 1406f (5′-GYACWCACCGCCCGT-3′) and 1525r (5′-AAGGAGGTGWTCCARCC-3′) on 5 µl of sample input to amplify bacterial and archaeal 16S rRNA genes as previously described (Woodcroft et al. 2018). Finally, the 12 palsa samples were sequenced at the Joint Genome Institute (JGI; Walnut creek, CA), where library preparation was performed using the Accel-NGS 1S Plus kit. All viromes required 20 PCR cycles, except –CsCl, bead-beating which required 18. All libraries were sequenced using the Illumina HiSeq-2000 1TB platform (2 x 151 bp paired-end).

Bioinformatics and statistics

The same informatics and statistics approaches were applied to viromes from Experiments 1 and 2. The sequences were quality-controlled using Trimmomatic (Bolger, Lohse, and Usadel 2014), adaptors were removed, reads were trimmed as soon as the average per-base quality dropped below 20 on 4 nt sliding windows, and reads shorter than 50 bp were discarded, with an additional 10 bp removed from the beginning of read pair one and the end of read pair two to remove the low complexity tail specific to the Accel-NGS 1S Plus kit, per the manufacturer's instruction. Reads were assembled using SPAdes (Bankevich et al. 2012; single-cell option, and k-mers 21, 33, and 55), and the contigs were processed with VirSorter to distinguish viral from microbial contigs (virome decontamination mode; Roux et al. 2015).

Contigs that were selected as VirSorter categories 1 and 2 were used to identify dsDNA viral contigs (as in Trubl et al. 2018). ssDNA viruses, due to short genomes and highly divergent hallmark genes, can frequently be missed by automatic viral sequence identification tools (e.g. VirSorter from Roux et al. 2015 or VirFinder in Ren et al. 2017). We therefore applied a two-step approach to ssDNA identification. First, we identified circular contigs that matched ssDNA marker genes from the PFAM database (Viral_Rep and Phage_F domains), using hmmsearch (Eddy, 2009; HMMER v3; cutoffs: score ≥ 50 and e-value ≤ 0.001). This identified four Phage_F-encoding and five Viral_Rep-encoding circular contigs, i.e. presumed complete genomes. Second, 2 new HMM profiles were generated, using the protein sequences from the nine identified circular viral contigs, and used to search (hmmsearch with the same cutoffs) the viromes' predicted proteins. This resulted in a final set of 23 predicted ssDNA contigs identified across nine viromes (Table S1).

The viral contigs were clustered at 95% average nucleotide identify (ANI) across 85% of the contig (Roux et al. 2018a) using nucmer (Delcher, Salzberg, and Phillippy 2003). The same contigs were also compared by BLAST to a pool of potential laboratory contaminants (i.e. Enterobacteria phage PhiX17, Alpha3, M13, Cellulophaga baltica phages, and Pseudoalteromonas phages), and any contigs matching a potential contaminant at more than 95% ANI across 80% of the contig were removed. Viral operational taxonomic units (vOTUs) were defined as non-redundant (i.e. post-clustering) viral contigs >10kb for dsDNA viruses (from VirSorter categories 1 or 2; Roux et al. 2015) and circular contigs from 4–8 kb for Microviridae viruses or 1–5 kb for circular replication-associated protein (Rep)-encoding ssDNA (CRESS DNA) viruses. The vOTUs represent populations that are likely species-level taxa and there is extensive literature context supporting this new standard terminology, which is summarized in a recent consensus paper (Roux et al. 2018a). The relative abundance of vOTUs was estimated based on post-QC reads mapping at ≥90% ANI and covering >10% of the contig (Paez-Espino et al. 2016; Roux et al. 2018a) using Bowtie2 (Langmead and Salzberg 2012). Figures were generated with R, using packages Vegan for diversity (Oksanen et al. 2016) and ggplot2 (Wickham 2016) or pheatmap (Kolde 2012) for heatmaps. Hierarchical clustering (function pvclust; method.dist="euclidean" and method.hclust="complete") was conducted on Bray-Curtis dissimilarity matrices using 1000 bootstrap iterations and only the approximately unbiased (AU) bootstrap values were reported.

Data availability

The 18 viromes from Experiments 1 and 2 are available at the IsoGenie project database under data downloads at https://isogenie.osu.edu/ and at CyVerse (https://www.cyverse.org/) file path /iplant/home/shared/iVirus/Trubl_Soil_Viromes. Data was processed using The Ohio Supercomputer Center (Ohio Supercomputer Center 1987). The final optimized protocol can be accessed here: https://www.protocols.io/view/soil-viral-extraction-protocol-for-ssdna-amp-dsdna-tzzep76.

Results and Discussion

Two experiments were performed to optimize the generation of quantitatively-amplified viromes from soil samples (Fig. 1). Experiment 1 evaluated three different DNA extraction methods for DNA yield, purity, and successful virome generation on the challenging humic-laden bog and fen soils. Experiment 2 compared two viral particle purification methods (with or without CsCl) and two virion lysis methods (heat vs bead-beating), for DNA yield, microbial DNA contamination, and successful virome generation for all three site habitats (palsa, bog and fen). An optimized virome generation protocol was determined for these palsa, bog and fen soils.

Different DNA extraction methods display variable efficiencies and recover distinct vOTUs

In Experiment 1, three DNA extraction methods were evaluated for DNA yield and purity: PowerSoil DNA extraction kits, Wizard mini columns, and a classic molecular biological approach using cetyl trimethylammonium bromide (CTAB). The PowerSoil kit was designed for humic-rich soils, which dominate our site (Hodgkins et al. 2014; Normand et al. 2017), and has performed well previously for viral samples (Iker et al. 2013). Wizard mini columns were used previously to generate viromes from these soils (Trubl et al. 2018). CTAB performs well on polysaccharide-rich samples (Porebski, Bailey, and Baum 1997), such as our site's peat soils.

Overall, the PowerSoil kit performed best, with the highest DNA yields and increased purity which led to more successful libraries and identification of more vOTUs. Specifically, the PowerSoil kit generally yielded the most DNA in the bog and fen, although the increase was only significant in the fen habitat (one-way ANOVA, α 0.05, and Tukey's test with p-value <0.05; Fig. 2A). DNA purity, which is also essential to virome generation (since proteins, phenols, and organics can inhibit amplification; reviewed in Alaeddini 2012), was examined via A260:280 (Fig 2B; for proteins and phenol contamination; Maniatis et al. 1982) and A260:230 ratios (Fig 2C; for carbohydrates and phenols; Maniatis et al. 1982; Tanveer, Yadav, and Yadav 2016). We posited that A260:280 is a more robust predictor of virome success, since previous work showed that A260:230 of DNA extracts had limited correlation to amplification success (Costa et al. 2010; Ramos-Gómez et al. 2014), and is highly variable for low DNA concentrations typical for soil viral extracts. For bog samples, at least one replicate from each DNA extraction method had a clean sample based on A260:280 (defined as 1.6–2.1), and PowerSoil extracts consistently exhibited the highest A260:230 ratios (i.e. inferred to be cleanest). For the fen, the same trend was recapitulated (PowerSoil having the cleanest ratios). One bog PowerSoil sample, and one fen CTAB sample, had unusually high A260:280 ratios, suggesting the presence of leftover extraction reagents in the sample.

Soil microbial metagenome protocols commonly include further DNA clean-up after extraction to remove inhibitory substances commonly seen in soil (summarized in Roose-Amsaleg, Garnier-Sillam, and Harry 2001; Roslan, Mohamad, and Omar 2017), therefore we evaluated the potential improvement in viral DNA purity from clean-up by AMPure beads. Purity (measured via A260:280) improved significantly in the bog Wizard and PowerSoil extracts, and the fen CTAB extracts, and improved in both bog and fen CTAB and PowerSoil extracts. For A260:230, all post-clean-up DNAs were still below the standard minimum threshold (1.6–2.2, Fig.2C).

Although DNA extract yield and purity metrics are useful indicators of extract quality, the goal is successful library preparation and sequencing. Thus, we used the cleaned up DNA to attempt virome generation, which revealed that PowerSoil-derived DNA was more amenable to library construction than the other extracts. Specifically, five of six PowerSoil extracts successfully generated libraries, whereas only one of the Wizard and none of the CTAB extracts led to successful library construction (threshold for success described in methods). Presumably, the success of the PowerSoil extraction methods was increased due to the kit having been optimized for humic-laden soils (specific reagents proprietary to Qiagen).

Where sequencing library construction was successful, we then sequenced and analyzed the resultant viromes to assess whether the vOTUs captured varied across replicate PowerSoil viromes and between the PowerSoil and Wizard viromes. In total, the 6 viromes produced 1,311 dsDNA viral contigs (VirSorter categories 1 and 2; Roux et al. 2015), which clustered into 516 vOTUs (see methods; Roux et al. 2018a). There were dramatic changes in the presence and relative abundance of vOTUs across the two DNA extraction kits evaluated, the biological replicates, and the soil habitats, which is partially the result of uneven coverage due to the 15 rounds of PCR performed to amplify the DNA. While PCR amplification is a powerful tool that permits ecological interpretation of resulting viral data (Duhaime and Sullivan 2012; Solonenko and Sullivan 2013; Solonenko et al. 2013), library amplification can lead to an enrichment in short inserts, resulting in uneven coverage, a bias that scales with the number of PCR cycles performed (Roux et al. 2018b). The differences in vOTU presence/absence among viromes decreased but remained noticeable even when using the most sensitive thresholds proposed for the detection of a vOTU in a metagenome (Roux et al. 2018b, Fig. S2). This suggests bias from the DNA extraction method (as reported previously for microbial populations; Delmont et al. 2011; Zielińska et al. 2017), and/or haphazard detection of low-abundance vOTUs due to inadequate sampling and/or sequencing depth.

Heat-based lysis of non-CsCl-purified virus particles provides the most comprehensive viromes

With PowerSoil identified as the optimal DNA extraction kit (yielding the most successful viromes), in Experiment 2 we next evaluated whether density-based particle purification and/or alternative virion lysis methods could increase viral DNA yield, as previously suggested (Delmont et al. 2011; Zielińska et al. 2017). We reasoned that purification by cesium-chloride (CsCl) density gradients could result in viral loss (as previously described in Trubl et al. 2016), but also lead to reduced microbial DNA contamination by removing ultra-small (<0.2um) cells that survive the filtration step. For lysis methods, we compared the two suggested in the PowerSoil protocol and posited that heat lysis would work better because it has been used previously on viruses (reviewed in McCance 1996) and the bead-beating method was previously shown to cause ~27% more viral loss than not using beads with PowerSoil extraction kit on diverse soils (Iker et al. 2013).

To assess this, viruses were resuspended from three palsa, bog, and fen samples as previously described (Trubl et al. 2016), and then the samples were split with half undergoing particle purification via CsCl gradients and half not, and each purification treatment lysed by each of the two lysis methods (heat and bead beating) for a total of 4 treatments, all followed by PowerSoil extraction (Fig. 1). We found significant differences in DNA yield due to purification and lysis method choice (Fig. 4, one-way ANOVA, α 0.05, and Tukey's test with p-value <0.05). CsCl purification had the most impact: yield was higher without it than with it for all but one sample (Palsa, –CsCl[BB]). Lysis method also mattered, with heat producing significantly higher DNA yield than bead-beating (t test, p-value <0.05), for the –CsCl samples in the palsa and fen samples (not significant in the bog) (Fig. 4). These findings suggest that DNA yields are best when not purifying the resuspended viral particles and when lysed using heat.

Higher DNA yields could result from contaminating (i.e. non-viral) DNA, so we quantified microbial DNA in all extracts via 16S rRNA gene qPCR (Fig. 5). Surprisingly, we generally observed higher microbial contamination in the CsCl-purified samples (Fig. 5, one-way ANOVA, α 0.05, and Tukey's test with p-value <0.05), and this varied along the thaw gradient with palsa contamination being higher than that of bog and fen samples. Since residual soil organics can interfere with PCR (Kontanis and Reed, 2006), we repeated the qPCR assay after DNA purification with the PowerClean kit. Generally, microbial contamination increased for –CsCl samples (Fig. 5), suggesting that their previously low microbial contamination was due to PCR inhibition, and +CsCl samples had mixed results, but in each habitat +CsCl[BB] samples had a significant increase in measurable contamination (Fig. 5). All treatments had higher qPCR-based microbial contamination after PowerClean, except +CsCl[H] samples which averaged a 1.5–26-fold reduction. Overall there was still no consistent, or significant, improvement in microbial contamination from inclusion of a CsCl purification step, even after PowerClean treatment.

Since we sequenced bog and fen viromes to characterize treatment effects on viral signal in Experiment 1, we opted in Experiment 2 to do this evaluation on the 12 Palsa samples, which were all sequenced. We found that the higher DNA yields in the –CsCl samples led to ~3-fold more viral contigs, which were also an average of 2.3-fold larger than +CsCl samples (Fig. 6A). The results from heat-lysis samples were more modest as they resulted in only ~33% more viral contigs, and statistically indistinguishable contig sizes across treatments (Fig. 6B; unequal variance t-test, p-value >0.05). These findings suggest that the optimal combination for recovering virus genomes from these soils is to skip CsCl purification and lyse the resultant viral particles using heat.

We next evaluated whether vOTU representation and diversity estimates from the same samples varied across the purification and lysis methods tested here. In total, we identified 66 vOTUs from these 12 palsa viromes, with 100% of the vOTUs identified in –CsCl samples, 89% (59) identified in the +CsCl samples, and vOTUs identified by both datasets displaying an average of 30-fold more coverage (Fig. 7) in –CsCl viromes. This indicates that the CsCl purification step reduced the samples to a subset of the initial viral community and did not help recover virus genomes that would be missed otherwise. Profiles of the recovered communities clustered first by soil core (AU branch supports >76), then mostly by purification (AU branch supports >66), and lastly by lysis, and did not change after varying the threshold for considering a lineage present (Fig. S3). Collectively this suggests that differences introduced by sample preparation were outweighed by the distinctiveness of each core's viral community. We proceeded to use diversity metrics to evaluate the different methods' impacts. The alpha diversity metrics paralleled treatment DNA yields where –CsCl samples were on average 56% more diverse than the +CsCl samples, and heat samples were on average 83% more diverse than the bead-beating samples (Fig. S4A). A comparison of dissimilarities among samples suggested the lysis method had more of an impact, although this effect was variable between samples and thus not statistically significant overall (Fig. S4B).

ssDNA viruses are recovered in all 3 habitats

Viromes have previously either neglected ssDNA viruses or qualitatively described them, but with the onset of the Accel-NGS 1S Plus kit, we leveraged the viromics data produced here to investigate the diversity and relative abundance of ssDNA viruses in our soil samples. ssDNA viruses are known from culture collections to commonly infect plants as opposed to bacteria, but their distributions in nature remain poorly explored outside of aquatic systems (Labonté and Suttle 2013). Notably, the first quantitative ssDNA/dsDNA viromes suggested that identifiable ssDNA viruses represent a few percent of the viruses observed in marine and freshwater systems (Roux et al. 2016).

To assess this biological signal in soils, we investigated the recovery and relative abundance of ssDNA viruses across our different soil habitats and sample preparations. Overall, we identified 35 putative ssDNA viruses, 11 from the Microviridae family and 24 CRESS DNA viruses (Fig. 8), which clustered into 13 vOTUs (3 Microviridae and 10 CRESS DNA). These ssDNA vOTUs were only a small fraction of the total vOTUs identified in each habitat (1% in bog and fen, and 8% in palsa) and only bog and fen samples included both types (Microviridae and CRESS-DNA), while palsa samples included exclusively CRESS-DNA viruses (Table S1). This suggests that, as for dsDNA viruses, the composition of the ssDNA virus community varies along the thaw gradient, potentially as a result of known changes in the host communities (Trubl et al. 2018), both microbial (Mondav et al. 2017; Woodcroft et al. 2018) and plant (Hodgkins et al. 2014; Normand et al. 2017). Notably, bead-beating-lysis samples did not include any ssDNA viruses. We posit that this was likely due to the heterogeneity of soil, because ssDNA viruses have previously been identified from experiments that used a bead-beating lysis (Hopkins et al. 2014). Finally, ssDNA viruses represented on average 4% of the community in the samples where ssDNA and dsDNA viruses were detected, which suggests that ssDNA viruses are not the dominant type of virus in these soils.

Conclusions

The development of a sample-to-sequence pipeline for ssDNA and dsDNA viruses in soils is crucial for characterizing viruses and their impact in these ecosystems. Our work here built upon previous work that optimized virus resuspension from soils by evaluating DNA extraction and lysis methods to increase DNA yields and purity. Additionally, this is the first evaluation of the Accel-NGS 1S Plus kit to capture ssDNA viruses in soils. Although these efforts have made inroads towards characterizing the soil virosphere, several challenges remain. Initial challenges arise from lack of data on which fraction of the free virus particles are being recovered from soils, and how to achieve a holistic sampling of the virus community (i.e. dsDNA, ssDNA, and RNA viruses). Beyond these, the presence of non-viral DNA in capsids or vesicles, e.g. gene transfer agents, can dilute viral signal in viromes and complicate interpretation (reviewed in Roux et al. 2013; Hurwitz, Hallam and Sullivan 2013; Lang and Beatty 2010), although new methods are being developed to identify and characterize these contaminating agents (reviewed in Lang, Westbye, Beatty 2017). The advent of long-read sequencing technologies have recently been applied to viromics and can improve contig generation for regions of genome with high similarity or complexity (summarized in Roux et al. 2017; Karamitros et al. 2018) and prevent formation of chimeric contigs. Longer-read viromes can thereby not only increase vOTU recovery but also provide resolution of hypervariable genome regions with niche-defining genes, and help capture micro-diverse populations missed by short-read assemblies (Warwick-Dugdale et al. 2018). Next, inferences of viral impacts on microbial communities and C cycling will require predicting hosts both in silico (Edwards et al. 2015; Paez-Espino et al. 2017) and in vitro (Deng et al. 2014; Brum & Sullivan 2015; Cenens et al. 2015), approaches to which are emerging. Finally, identification of the active viral community and characterization of their roles in biogeochemical processes can be better resolved with techniques like stable isotope-based approaches linked with nanoscale secondary ion mass spectrometry (NanoSIP; Pacton et al. 2014; Pasulka et al. 2018; Gates et al. 2018). Application of these and other approaches to soil viromics will increase and diversify publicly available viral datasets, advance our understanding of soil viral ecology, and improve our knowledge of viral roles in soil ecosystems.

Acknowledgments

We thank Olivier Zablocki for his suggestions and comments. We thank Moira Hough, Sky Dominguez, and Nicole Raab for collecting the soil cores and geochemical data, and the Abisko Naturvetenskapliga Station for field support. Bioinformatics were supported by The Ohio Supercomputer Center and by the National Science Foundation under Award Numbers DBI-0735191 and DBI-1265383; URL: www.cyverse.org. This study was funded by the Genomic Science Program of the United States Department of Energy Office of Biological and Environmental Research (grants DE-SC0010580 and DE-SC0016440), The Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program, and by the Gordon and Betty Moore Foundation Investigator Award (GBMF#3790 to MBS). The SCGSR program is administered by the Oak Ridge Institute for Science and Education (ORISE) for the DOE. ORISE is managed by ORAU under contract number DE-SC0014664.

References

Alaeddini, R., 2012. Forensic implications of PCR inhibition—a review. Forensic Science International: Genetics, 6(3), pp.297-305.

Amgarten, D.E., Braga, L.P.P., Da Silva, A.M. and Setubal, J.C., 2018. MARVEL, a Tool for Prediction of Bacteriophage Sequences in Metagenomic Bins. Frontiers in genetics, 9, p.304.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D. and Pyshkin, A.V., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology, 19(5), pp.455-477.

Binga, E.K., Lasken, R.S. and Neufeld, J.D., 2008. Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. The ISME journal, 2(3), p.233

Bolger, A.M. Lohse, M. and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, p.btu170.

Brum, J.R. and Sullivan, M.B., 2015. Rising to the challenge: accelerated pace of discovery transforms marine virology. Nature Reviews Microbiology, 13(3), p.147.

Cenens, W., Makumi, A., Govers, S.K., Lavigne, R. and Aertsen, A., 2015. Viral transmission dynamics at single-cell resolution reveal transiently immune subpopulations caused by a carrier state association. PLoS genetics, 11(12), p.e1005770.

Costa, J., Mafra, I., Amaral, J.S. and Oliveira, M.B.P., 2010. Detection of genetically modified soybean DNA in refined vegetable oils. European Food Research and Technology, 230(6), pp.915-923.

Delcher, A.L., Salzberg, S.L. and Phillippy, A.M., 2003. Using MUMmer to identify similar regions in large sequence sets. Current protocols in bioinformatics, (1), pp.10-3.

Delmont, T.O., Robe, P., Cecillon, S., Clark, I.M., Constancias, F., Simonet, P., Hirsch, P.R. and Vogel, T.M., 2011. Accessing the soil metagenome for studies of microbial diversity. Applied and Environmental Microbiology, 77(4), pp.1315-1324.

Deng, L., Ignacio-Espinoza, J.C., Gregory, A.C., Poulos, B.T., Weitz, J.S., Hugenholtz, P. and Sullivan, M.B., 2014. Viral tagging reveals discrete populations in Synechococcus viral genome sequence space. Nature, 513(7517), p.242.

Duhaime, M.B. and Sullivan, M.B., 2012. Ocean viruses: rigorously evaluating the metagenomic sample-to-sequence pipeline. Virology, 434(2), pp.181-186.

Eddy, S.R., 2009. A new generation of homology search tools based on probabilistic inference. In Genome Informatics 2009: Genome Informatics Series Vol. 23 (pp. 205-211).

Edwards, R.A., McNair, K., Faust, K., Raes, J. and Dutilh, B.E., 2015. Computational approaches to predict bacteriophage–host relationships. FEMS microbiology reviews, 40(2), pp.258-272.

Fierer, N., 2017. Embracing the unknown: disentangling the complexities of the soil microbiome. Nature Reviews Microbiology, 15(10), p.579.

Gates, S.D., Condit, R.C., Moussatche, N., Stewart, B.J., Malkin, A.J. and Weber, P.K., 2018. High Initial Sputter Rate Found for Vaccinia Virions Using Isotopic Labeling, NanoSIMS, and AFM. Analytical chemistry, 90(3), pp.1613-1620.

Han, L., Sun, K., Jin, J. and Xing, B., 2016. Some concepts of soil organic carbon characteristics and mineral interaction from a review of literature. Soil Biology and Biochemistry, 94, pp.107-121.

Hayes, S., Mahony, J., Nauta, A. and van Sinderen, D., 2017. Metagenomic approaches to assess bacteriophages in various environmental niches. Viruses, 9(6), p.127.

Hodgkins, S.B., Tfaily, M.M., McCalley, C.K., Logan, T.A., Crill, P.M., Saleska, S.R., Rich, V.I. and Chanton, J.P., 2014. Changes in peat chemistry associated with permafrost thaw increase greenhouse gas production. Proceedings of the National Academy of Sciences, p.201314641.

Hopkins, M., Kailasan, S., Cohen, A., Roux, S., Tucker, K.P., Shevenell, A., Agbandje-McKenna, M. and Breitbart, M., 2014. Diversity of environmental single-stranded DNA phages revealed by PCR amplification of the partial major capsid protein. The ISME journal, 8(10), p.2093

Hurwitz, B.L., Hallam, S.J. and Sullivan, M.B., 2013. Metabolic reprogramming by viruses in the sunlit and dark ocean. Genome biology, 14(11), p.R123.

Iker, B.C., Bright, K.R., Pepper, I.L., Gerba, C.P. and Kitajima, M., 2013. Evaluation of commercial kits for the extraction and purification of viral nucleic acids from environmental and fecal samples. Journal of virological methods, 191(1), pp.24-30.

Johansson, T., Malmer, N., Crill, P.M., Friborg, T., Aakerman, J.H., Mastepanov, M. and Christensen, T.R., 2006. Decadal vegetation changes in a northern peatland, greenhouse gas fluxes and net radiative forcing. Global Change Biology, 12(12), pp.2352-2369.

Karamitros, T., van Wilgenburg, B., Wills, M., Klenerman, P. and Magiorkinis, G., 2018. Nanopore sequencing and full genome de novo assembly of human cytomegalovirus TB40/E reveals clonal diversity and structural variations. BMC genomics, 19(1), p.577.

Karlsson, O.E., Belák, S. and Granberg, F., 2013. The effect of preprocessing by sequence-independent, single-primer amplification (SISPA) on metagenomic detection of viruses. Biosecurity and bioterrorism: biodefense strategy, practice, and science, 11(S1), pp.S227-S234

Kim, M.S., Whon, T.W. and Bae, J.W., 2013. Comparative viral metagenomics of environmental samples from Korea. Genomics & informatics, 11(3), pp.121-128.

Kolde, R., 2012. Pheatmap: pretty heatmaps. R package version, 61

Kontanis, E.J. and Reed, F.A., 2006. Evaluation of real-time PCR amplification efficiencies to detect PCR inhibitors. Journal of forensic sciences, 51(4), pp.795-804.

Labonté, J.M. and Suttle, C.A., 2013. Previously unknown and highly divergent ssDNA viruses populate the oceans. The ISME journal, 7(11), p.2169.

Lang, A.S., Westbye, A.B. and Beatty, J.T., 2017. The distribution, evolution, and roles of gene transfer agents in prokaryotic genetic exchange. Annual review of virology, 4, pp.87-104.

Langmead, B. and Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature methods, 9(4), p.357.

Lehmann, J. and Kleber, M., 2015. The contentious nature of soil organic matter. Nature, 528(7580), p.60.

Maniatis T., Fritsch E.F., Sambrook J. Molecular cloning: a laboratory manual. Cold Spring Harbor: Cold Spring Harbor Laboratory; 1982.

Marine, R., McCarren, C., Vorrasane, V., Nasko, D., Crowgey, E., Polson, S.W. and Wommack, K.E., 2014. Caught in the middle with multiple displacement amplification: the myth of pooling for avoiding multiple displacement amplification bias in a metagenome. Microbiome, 2(1), p.3

Mondav, R., McCalley, C.K., Hodgkins, S.B., Frolking, S., Saleska, S.R., Rich, V.I., Chanton, J.P. and Crill, P.M., 2017. Microbial network, phylogenetic diversity and community membership in the active layer across a permafrost thaw gradient. Environmental microbiology, 19(8), pp.3201-3218

Narayan, A., Jain, K., Shah, A.R. and Madamwar, D., 2016. An efficient and cost-effective method for DNA extraction from athalassohaline soil using a newly formulated cell extraction buffer. 3 Biotech, 6(1), p.62.

Narr, A., Nawaz, A., Wick, L.Y., Harms, H. and Chatzinotas, A., 2017. Soil Viral Communities Vary Temporally and along a Land Use Transect as Revealed by Virus-Like Particle Counting and a Modified Community Fingerprinting Approach (fRAPD). Frontiers in microbiology, 8, p.1975.

Normand, A.E., Smith, A.N., Clark, M.W., Long, J.R. and Reddy, K.R., 2017. Chemical composition of soil organic matter in a subarctic peatland: influence of shifting vegetation communities. Soil Science Society of America Journal, 81(1), pp.41-49.

Ohio Supercomputer Center. 1987. Ohio Supercomputer Center. Columbus OH: Ohio Supercomputer Center.

Oksanen, J., Blanchet, F., Kindt, R., Legendre, P. and O’Hara, R., 2016. Vegan: community ecology package. R package 2.3-3.

Olefeldt, D., Roulet, N.T., Bergeron, O., Crill, P., Bäckstrand, K. and Christensen, T.R., 2012. Net carbon accumulation of a high-latitude permafrost palsa mire similar to permafrost-free peatlands. Geophysical Research Letters, 39(3).

Pacton, M., Wacey, D., Corinaldesi, C., Tangherlini, M., Kilburn, M.R., Gorin, G.E., Danovaro, R. and Vasconcelos, C., 2014. Viruses as new agents of organomineralization in the geological record. Nature communications, 5, p.4298.

Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N., Rubin, E., Ivanova, N.N. and Kyrpides, N.C., 2016. Uncovering Earth’s virome. Nature, 536(7617), p.425.

Paez-Espino, D., Pavlopoulos, G.A., Ivanova, N.N. and Kyrpides, N.C., 2017. Nontargeted virus sequence discovery pipeline and virus clustering for metagenomic data. nature protocols, 12(8), p.1673.

Pasulka, A.L., Thamatrakoln, K., Kopf, S.H., Guan, Y., Poulos, B., Moradian, A., Sweredoski, M.J., Hess, S., Sullivan, M.B., Bidle, K.D. and Orphan, V.J., 2018. Interrogating marine virus-host interactions and elemental transfer with BONCAT and nanoSIMS-based methods. Environmental microbiology, 20(2), pp.671-692.

Phan, T.G., Mori, D., Deng, X., Rajindrajith, S., Ranawaka, U., Ng, T.F.F., Bucardo-Rivera, F., Orlandi, P., Ahmed, K. and Delwart, E., 2015. Small circular single stranded DNA viral genomes in unexplained cases of human encephalitis, diarrhea, and in untreated sewage. Virology, 482, pp.98-104

Porebski, S., Bailey, L.G. and Baum, B.R., 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant molecular biology reporter, 15(1), pp.8-15

Ramos-Gómez, S., Busto, M.D., Perez-Mateos, M. and Ortega, N., 2014. Development of a method to recovery and amplification DNA by real-time PCR from commercial vegetable oils. Food Chemistry, 158, pp.374-383.

Ren, J., Ahlgren, N.A., Lu, Y.Y., Fuhrman, J.A. and Sun, F., 2017. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. Microbiome, 5(1), p.69.

Roose-Amsaleg, C.L., Garnier-Sillam, E. and Harry, M., 2001. Extraction and purification of microbial DNA from soil and sediment samples. Applied Soil Ecology, 18(1), pp.47-60.

Rosario, K., Fierer, N., Miller, S., Luongo, J. and Breitbart, M., 2018. Diversity of DNA and RNA viruses in indoor air as assessed via metagenomic sequencing. Environmental science & technology, 52(3), pp.1014-1027.

Roslan, M.A.M., Mohamad, M.A.N. and Omar, S.M., 2017. High quality DNA from peat soil for metagenomic studies a minireview on dna extraction methods. Science, 1(2), pp.01-06.

Roux, S., Emerson, J.B., Eloe-Fadrosh, E.A. and Sullivan, M.B., 2017. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. PeerJ, 5, p.e3817.

Roux, S., Krupovic, M., Debroas, D., Forterre, P. and Enault, F., 2013. Assessment of viral community functional potential from viral metagenomes may be hampered by contamination with cellular sequences. Open biology, 3(12), p.130160.

Roux, S., Adriaenssens, E.M., Dutilh, B.E., Koonin, E.V, Kropinski, A.M., Krupovic, M., Kuhn, J.H. Lavigne, R., Brister, J.R., Varsani, A., Amid, C., Aziz, R.K., Bordenstein, S.R., Bork, P., Breitbart, M., Cochrane, G.R., Daly, R.A., Desnues, C., Duhaime, M.B., Emerson, J.B., Enault, F., Fuhrman, J.A., Hingamp, P., Hugenholtz, P., Hurwitz, B.L., Ivanova, N.N., Labonté, J.M., Lee, K-B., Malmstrom, R.R., Martinez-Garcia, M., Mizrachi, I.K., Ogata, H., Páez-Espino, D., Petit, M-A., Putonti, C., Rattei, T., Reyes, A., Rodriguez-Valera, F., Rosario, K., Schriml, L., Schulz, F., Steward, G.F., Sullivan, M.S., Sunagawa, S., Suttle, C.A., Temperton, B., Tringe, S.G., Thurber, R.V., Webster, N.S., Whiteson, K.L., Wilhelm, S.W., Wommack, K.E., Woyke, T., Wrighton, K.C., Yilmaz, P., Yoshida, T., Young, M.J., Yutin, N., Allen, L.Z., Kyrpides, N.C., Eloe-Fadrosh, E.A. 2018a. Minimum Information about an Uncultivated Virus Genome (MIUViG). Nature Biotechnology, 37(1), 29–37.

Roux, S., Trubl, G., Goudeau, D., Nath, N., Couradeau, E., Ahlgren, N.A., Zhan, Y., Marsan, D., Chen, F., Fuhrman, J.A. and Northen, T.R., 2018b. Optimizing de novo genome assembly from PCR-amplified metagenomes (No. e27453v1). PeerJ Preprints.Roux, S., Solonenko, N.E., Dang, V.T., Poulos, B.T., Schwenck, S.M., Goldsmith, D.B., Coleman, M.L., Breitbart, M. and Sullivan, M.B., 2016. Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. PeerJ, 4, p.e2777.

Roux, S., Enault, F., Hurwitz, B.L. and Sullivan, M.B., 2015. VirSorter: mining viral signal from microbial genomic data. PeerJ, 3, p.e985 Segobola, J., Adriaenssens, E., Tsekoa, T., Rashamuse, K. and Cowan, D., 2018. Exploring viral diversity in a unique South African soil habitat. Scientific reports, 8(1), p.111.

Solonenko, S.A., Ignacio-Espinoza, J.C., Alberti, A., Cruaud, C., Hallam, S., Konstantinidis, K., Tyson, G., Wincker, P. and Sullivan, M.B., 2013. Sequencing platform and library preparation choices impact viral metagenomes. BMC genomics, 14(1), p.320.

Solonenko, S.A. and Sullivan, M.B., 2013. Preparation of metagenomic libraries from naturally occurring marine viruses. In Methods in enzymology (Vol. 531, pp. 143-165). Academic Press.

Tanveer, A., Yadav, S. and Yadav, D., 2016. Comparative assessment of methods for metagenomic DNA isolation from soils of different crop growing fields. 3 Biotech, 6(2), p.220.

Warwick-Dugdale, J., Solonenko, N., Moore, K., Chittick, L., Gregory, A.C., Allen, M.J., Sullivan, M.B. and Temperton, B., 2018. Long-read metagenomics reveals cryptic and abundant marine viruses. bioRxiv, p.345041.

Wickham, H., 2016. ggplot2: elegant graphics for data analysis. Springer.

Wommack, K.E., Bhavsar, J., Polson, S.W., Chen, J., Dumas, M., Srinivasiah, S., Furman, M., Jamindar, S. and Nasko, D.J., 2012. VIROME: a standard operating procedure for analysis of viral metagenome sequences. Stand Genomic Sci 6: 427–439.

Woodcroft, B.J., Singleton, C.M., Boyd, J.A., Evans, P.N., Emerson, J.B., Zayed, A.A., Hoelzle, R.D., Lamberton, T.O., McCalley, C.K., Hodgkins, S.B. and Wilson, R.M., 2018. Genome-centric view of carbon processing in thawing permafrost. Nature, p.1.

Yilmaz, S., Allgaier, M. and Hugenholtz, P., 2010. Multiple displacement amplification compromises quantitative analysis of metagenomes. Nature methods, 7(12), p.943.

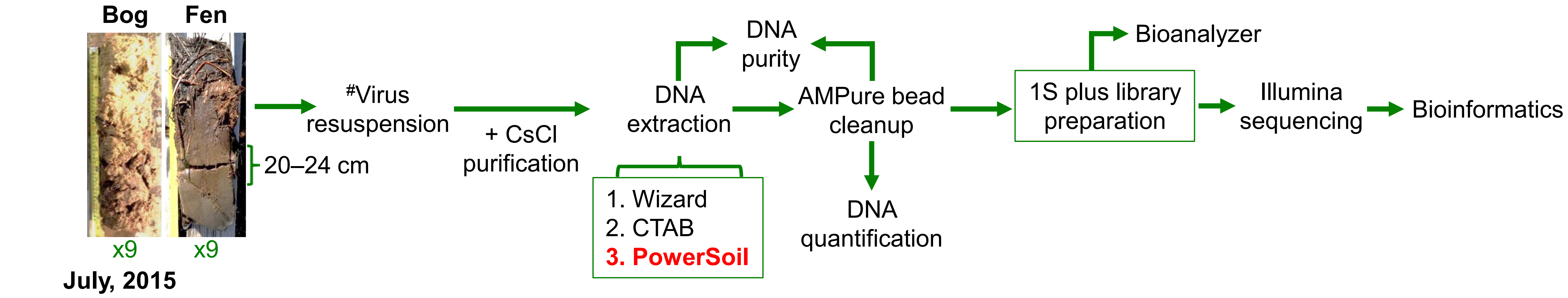
Zielińska, S., Radkowski, P., Blendowska, A., Ludwig-Gałęzowska, A., Łoś, J.M. and Łoś, M., 2017. The choice of the DNA extraction method may influence the outcome of the soil microbial community structure analysis. MicrobiologyOpen, 6(4), p.e00453.

Figure 1(on next page)

Overview of experiments to optimize methods for virome generation.

Two experiments (Experiment 1 in green and Experiment 2 in blue) evaluated three DNA extraction methods, two different virion lysis methods, and CsCl virion purification, for optimizing virome generation from three peats soils along a permafrost thaw gradient. Nine soil cores were collected in July 2015, three from each habitat, and used to create 18 samples (9 bog and 9 fen) with 10 ± 1 g of soil in each sample for Experiment 1 and 36 samples (12 palsa, 12 bog, and 12 fen) with 7.5 ± 1 g of soil in each sample for Experiment 2; representative photos of cores were taken by Gary Trubl. Viruses were resuspended as previously described in Trubl et al. (2016), but with the addition of a DNase step and a 1.3 g/ml layer for CsCl purification. Red font color indicates the best-performing option within each set. $^{\#}$ denotes adapted protocol from Trubl et al. 2016. $^{\#\#}$ indicates that only 12 palsa samples proceeded to library preparation.

Experiment 1: identify best DNA extraction method



Bog Fen

x9 x9

July, 2015

20–24 cm

#Virus resuspension

+ CsCl purification

DNA extraction

1. Wizard
2. CTAB
3. **PowerSoil**

DNA purity

AMPure bead cleanup

DNA quantification

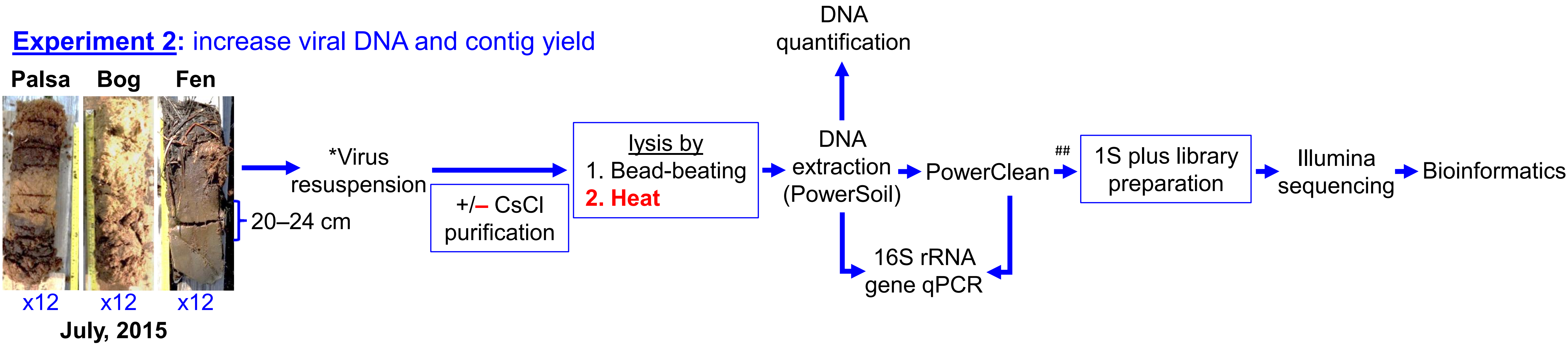
1S plus library preparation

Bioanalyzer

Illumina sequencing

Bioinformatics

Experiment 2: increase viral DNA and contig yield



Palsa Bog Fen

x12 x12 x12

July, 2015

20–24 cm

*Virus resuspension

+/– CsCl purification

lysis by
1. Bead-beating
2. **Heat**

DNA extraction (PowerSoil)

DNA quantification

16S rRNA gene qPCR

PowerClean

##

1S plus library preparation

Illumina sequencing

Bioinformatics

Figure 2(on next page)

Impact of extraction methods on DNA yields and purity (Experiment 1).

Bog samples are shown on the left of each panel, fen samples on the right. DNA extraction methods are color-coded: purple for CTAB, blue for Wizard, and green for PowerSoil. * denotes significant difference via one-way ANOVA, α 0.05, and Tukey's test with p-value <0.05. † denotes significant difference for t test, p-value <0.05; †† = p-value <0.01; ††† = p-value <0.001. A) The DNA concentration (ng/µl) after AMPure purification for the three DNA extraction methods. B) DNA extract purity via A260/A280. Dotted lines are purity thresholds: Acceptable range in yellow shading and preferred range in red shading. C) DNA extract purity via A260/A230.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27640v1 | CC BY 4.0 Open Access | rec: 6 Apr 2019, publ: 6 Apr 2019

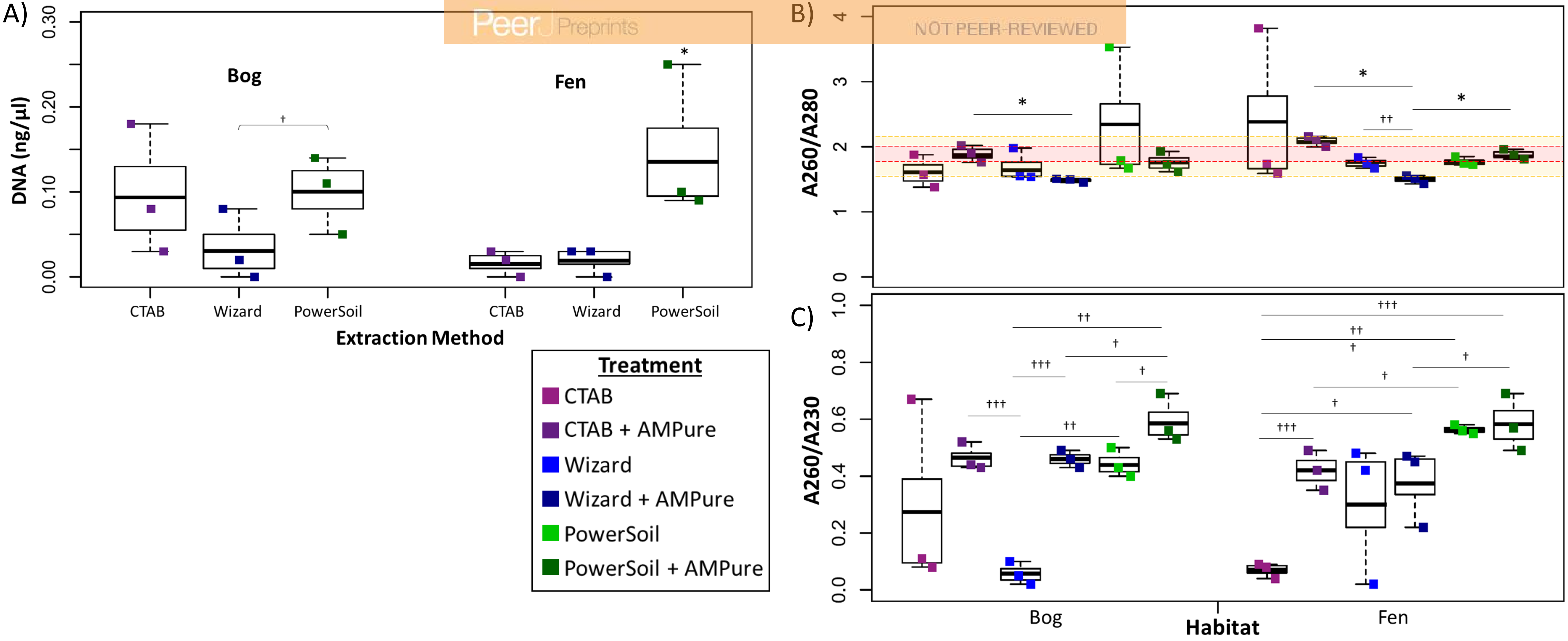
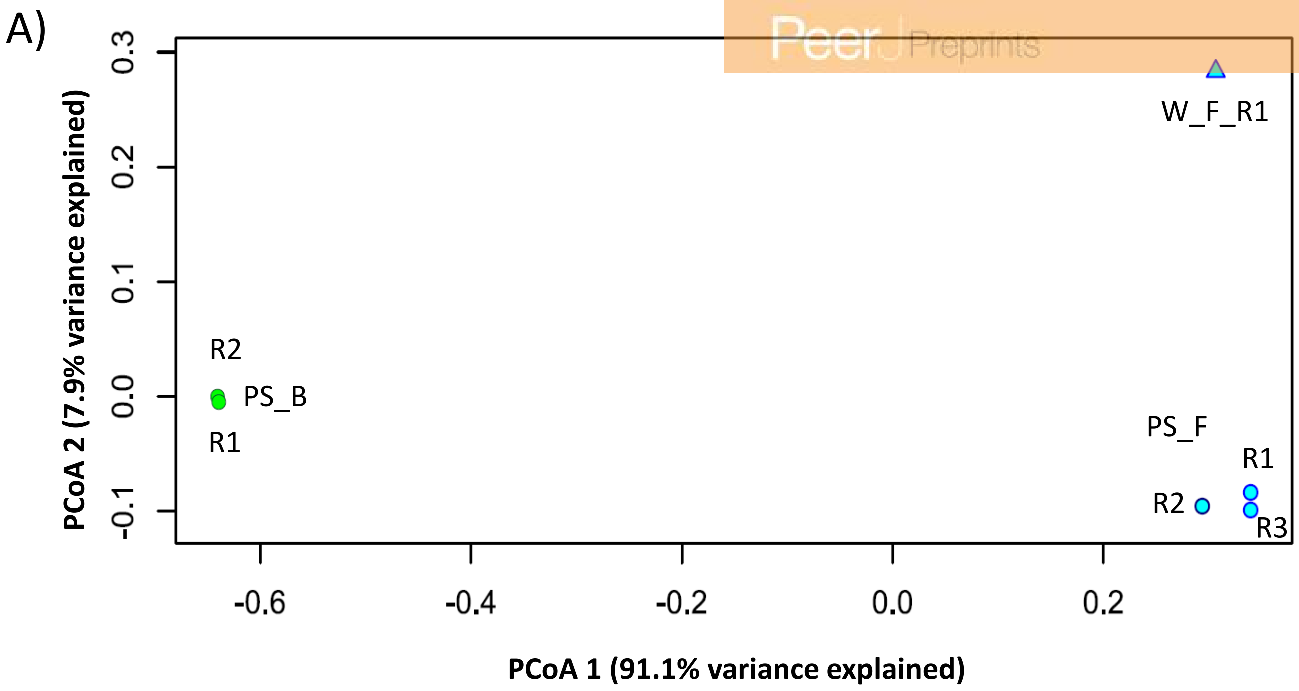

A)
DNA (ng/µl)
0.00
0.10
0.20
0.30
Bog
Fen
†
*
CTAB
Wizard
PowerSoil
CTAB
Wizard
PowerSoil
Extraction Method
B)
A260/A280
0
1
2
3
4
*
*
††
*
C)
A260/A230
0.0
0.2
0.4
0.6
0.8
1.0
†††
††
†
†††
†
††
†††
††
†
†
†
†
†††
Bog
Habitat
Fen
Treatment
🟪 CTAB
🟪 CTAB + AMPure
🟦 Wizard
🟦 Wizard + AMPure
🟩 PowerSoil
🟩 PowerSoil + AMPure

Figure 3(on next page)

Impact of extraction methods on recovery and abundance of vOTUs (Experiment 1).

A principal coordinate analysis of the viromes by normalized relative abundance of the 516 vOTUs based on their Bray-Curtis dissimilarity. Viromes distinguished by habitat (bog colored green, fen blue) and DNA extraction method (PowerSoil as circle, Wizard as triangle).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27640v1 | CC BY 4.0 Open Access | rec: 6 Apr 2019, publ: 6 Apr 2019



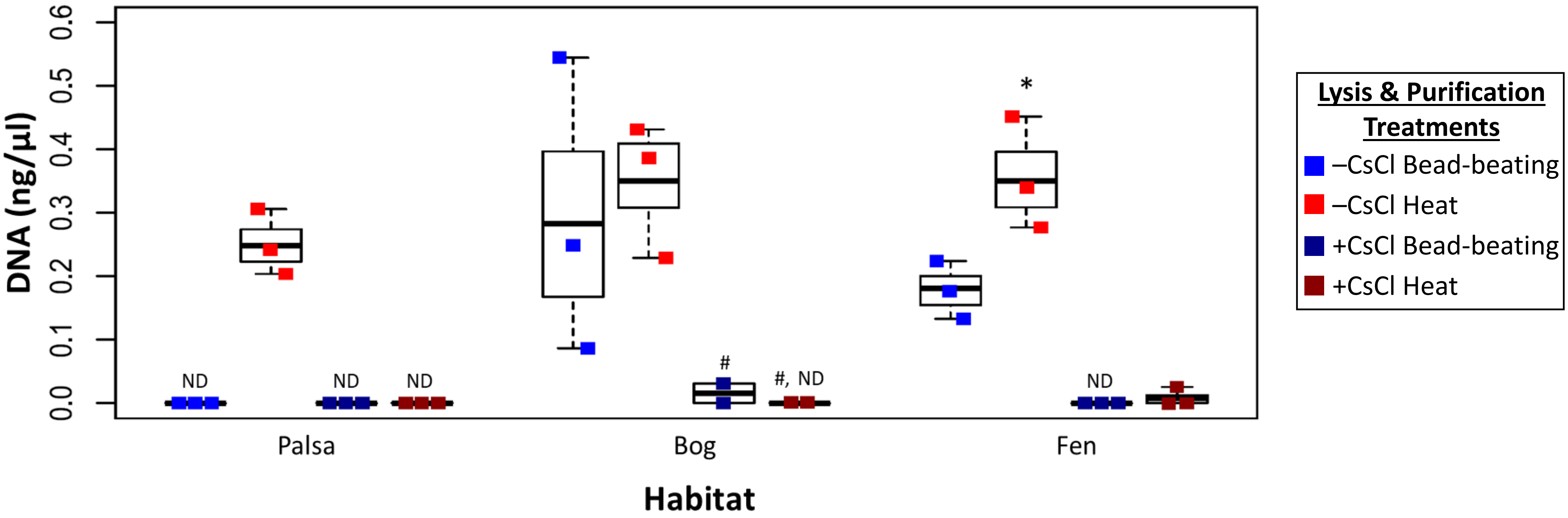
Nomenclature

PS=PowerSoil
W=Wizard
B=Bog
F=Fen
R1/R2/R3=Replicate

Figure 4(on next page)

Impact of lysis and purification methods on DNA yields (Experiment 2).

The DNA concentration (ng/µl) is given for the two virion lysis methods used, with or without CsCl purification, for all three habitats. The four treatments are color coded with blue for bead-beating, red for heat lysis and a darker shade if also purified with CsCl. * denotes significant difference via one-way ANOVA, α 0.05, and Tukey's test with p-value <0.05. # denotes n=2. N/D denotes non-detectable DNA concentration.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27640v1 | CC BY 4.0 Open Access | rec: 6 Apr 2019, publ: 6 Apr 2019

Figure 5(on next page)

Evaluation of microbial contamination (Experiment 2).

The 16S rRNA gene contamination (square root) is indicated for each virome grouped by habitat before (left) and after (right) clean up with PowerClean. The four treatments are color coded with blue for bead-beating and red for heat lysis and a darker shade after CsCl purification. # denotes no data available. 16S qPCR primers were 1406F-1525R (Woodcroft et al. 2018). †denotes significant difference for t test, p-value <0.05; ††= p-value <0.01; †††= p-value <0.001.

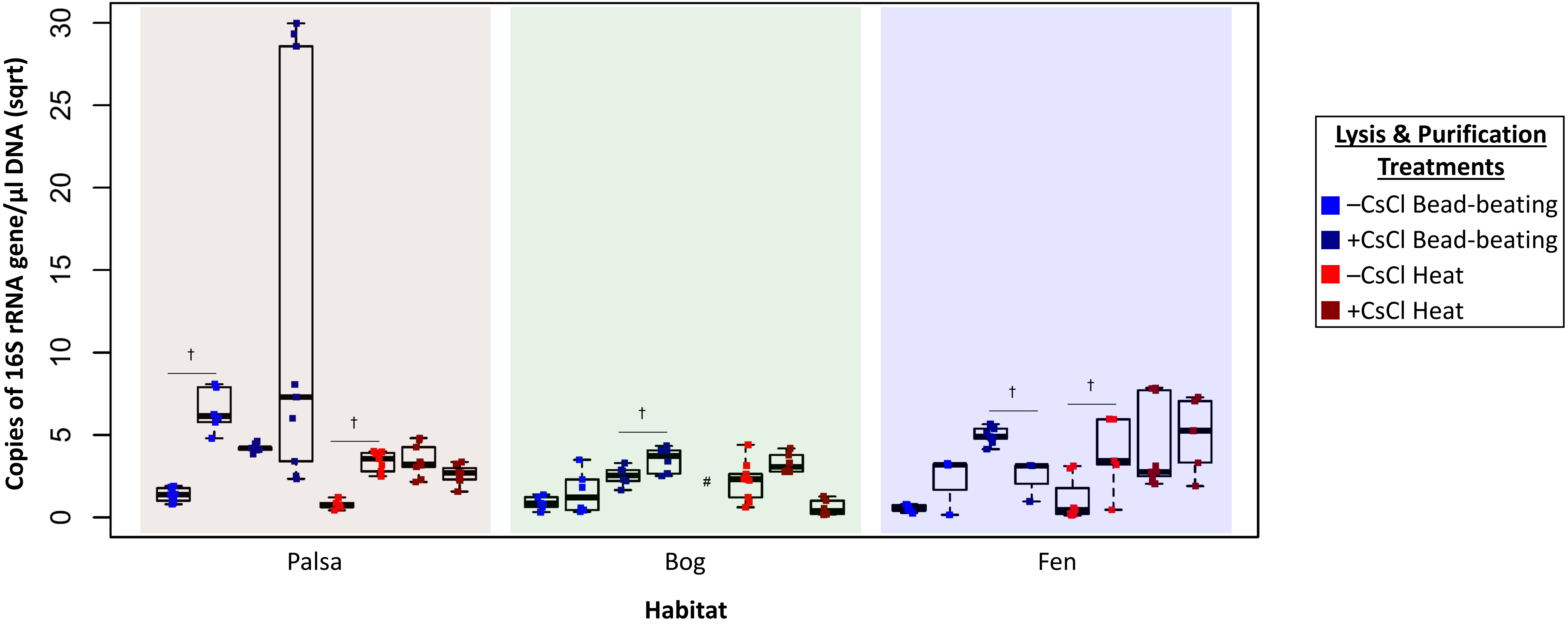

Copies of 16S rRNA gene/µl DNA (sqrt)
0
5
10
15
20
25
30
†
†
†
#
†
†
Palsa
Bog
Fen
Habitat
Lysis & Purification Treatments
🟦 –CsCl Bead-beating
🟦 +CsCl Bead-beating
🟥 –CsCl Heat
🟥 +CsCl Heat

Figure 6(on next page)

Number and size of assembled viral contigs (Experiment 2).

Boxplots show the number of viral contigs assembled, and those > 10 kb, for each treatment. Viral contigs were identified by two approaches: the “conservative” one included only contigs in VirSorter categories 1 & 2 for which a viral origin is very likely, while the “sensitive” one also included contigs in VirSorter category 3, for which a viral origin is possible but unsure.

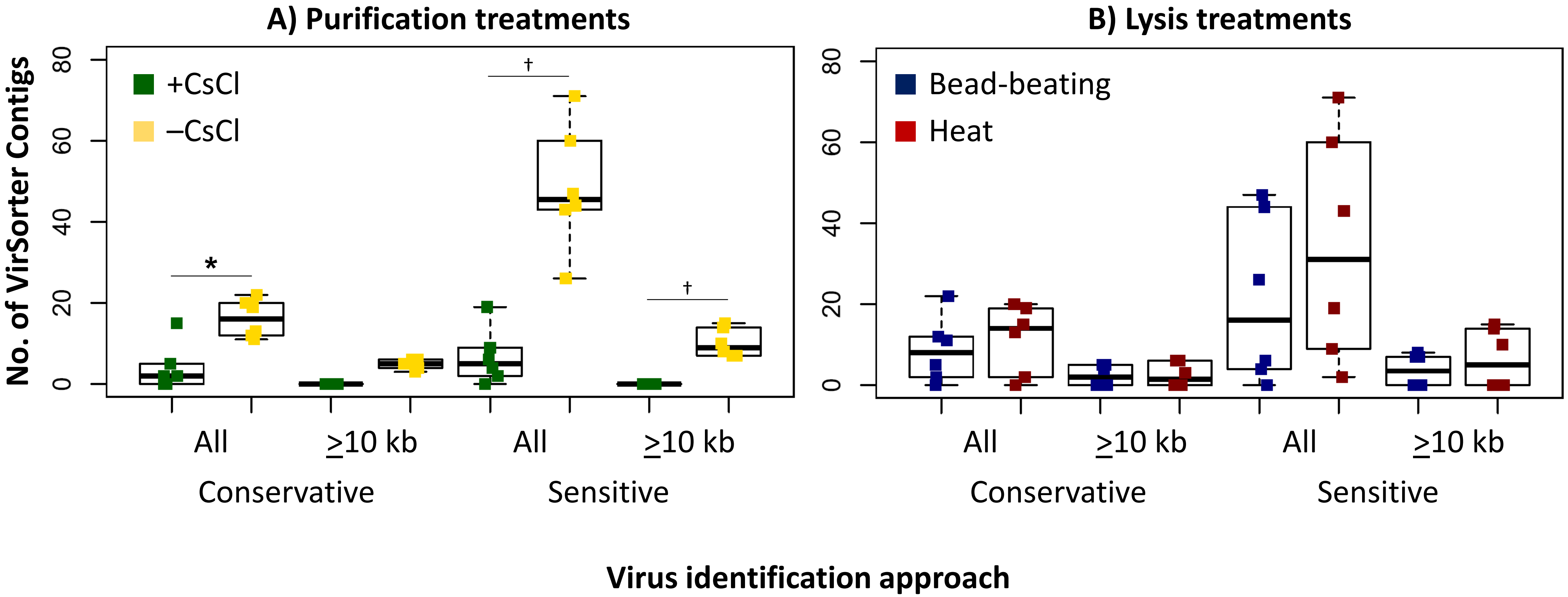

A) Purification treatments
B) Lysis treatments
No. of VirSorter Contigs
0
20
40
60
80
+CsCl
−CsCl
Bead-beating
Heat
*
†
†
All
≥10 kb
All
≥10 kb
Conservative
Sensitive
Virus identification approach

Figure 7(on next page)

Relative abundance of vOTUs across 12 palsa viromes (Experiment 2).

A heatmap showing the Euclidean-based hierarchical clustering of a Bray-Curtis dissimilarity matrix calculated from vOTU relative abundances within each virome with an approximately unbiased (AU) bootstrap value (n=1000). The relative abundances were normalized by contig length and per Gbp of metagenome and were $\log_{10}$ transformed. Reads were mapped to contigs at ≥ 90% nucleotide identity and the relative abundance was set to 0 if reads covered <10% of the contig. Heatmaps with alternative genome coverage thresholds are presented in Fig. S3. Abbreviations: H, heat lysis; BB, bead-beating; +/– CsCl, with or without cesium chloride purification; C, core.

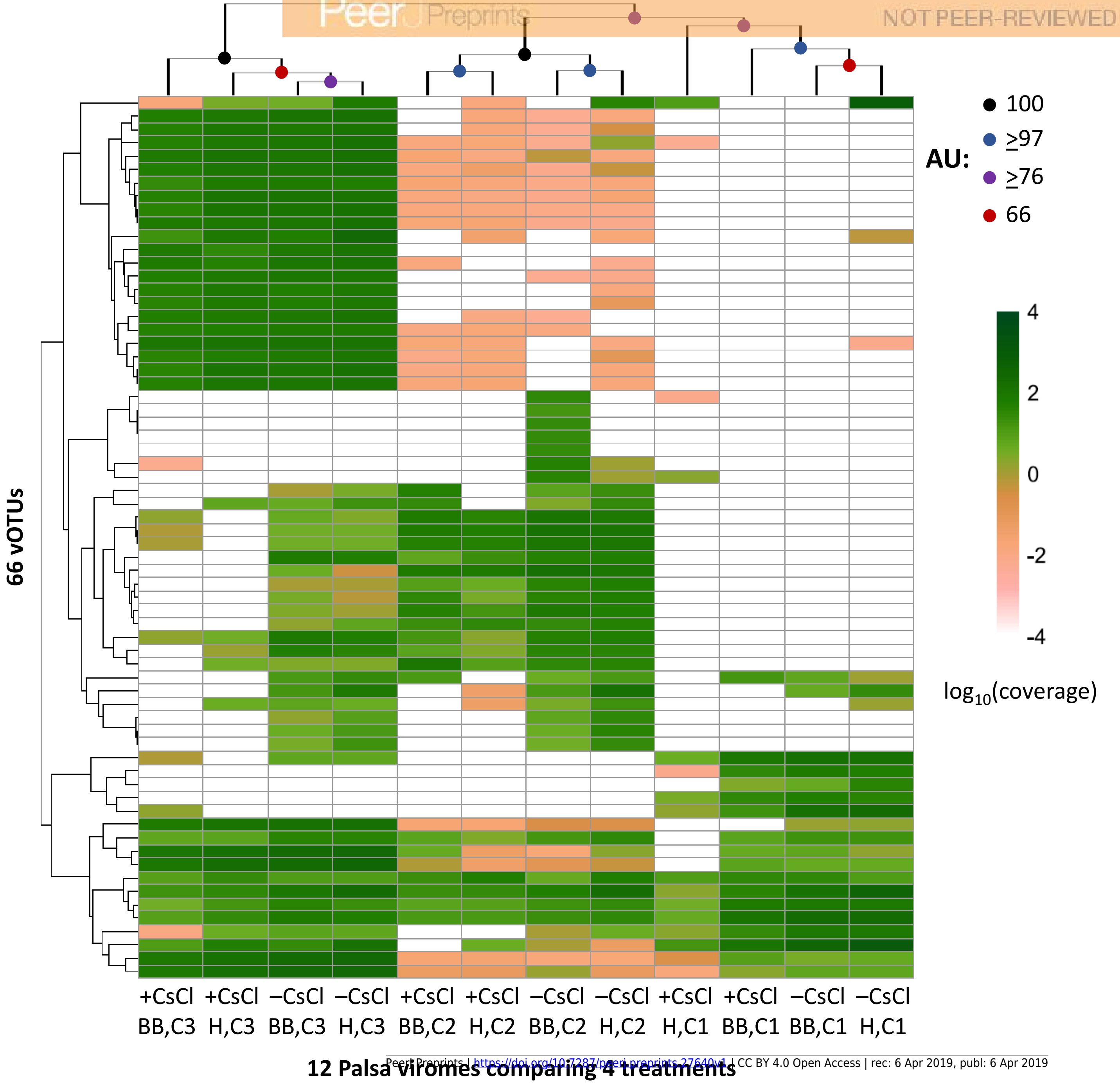
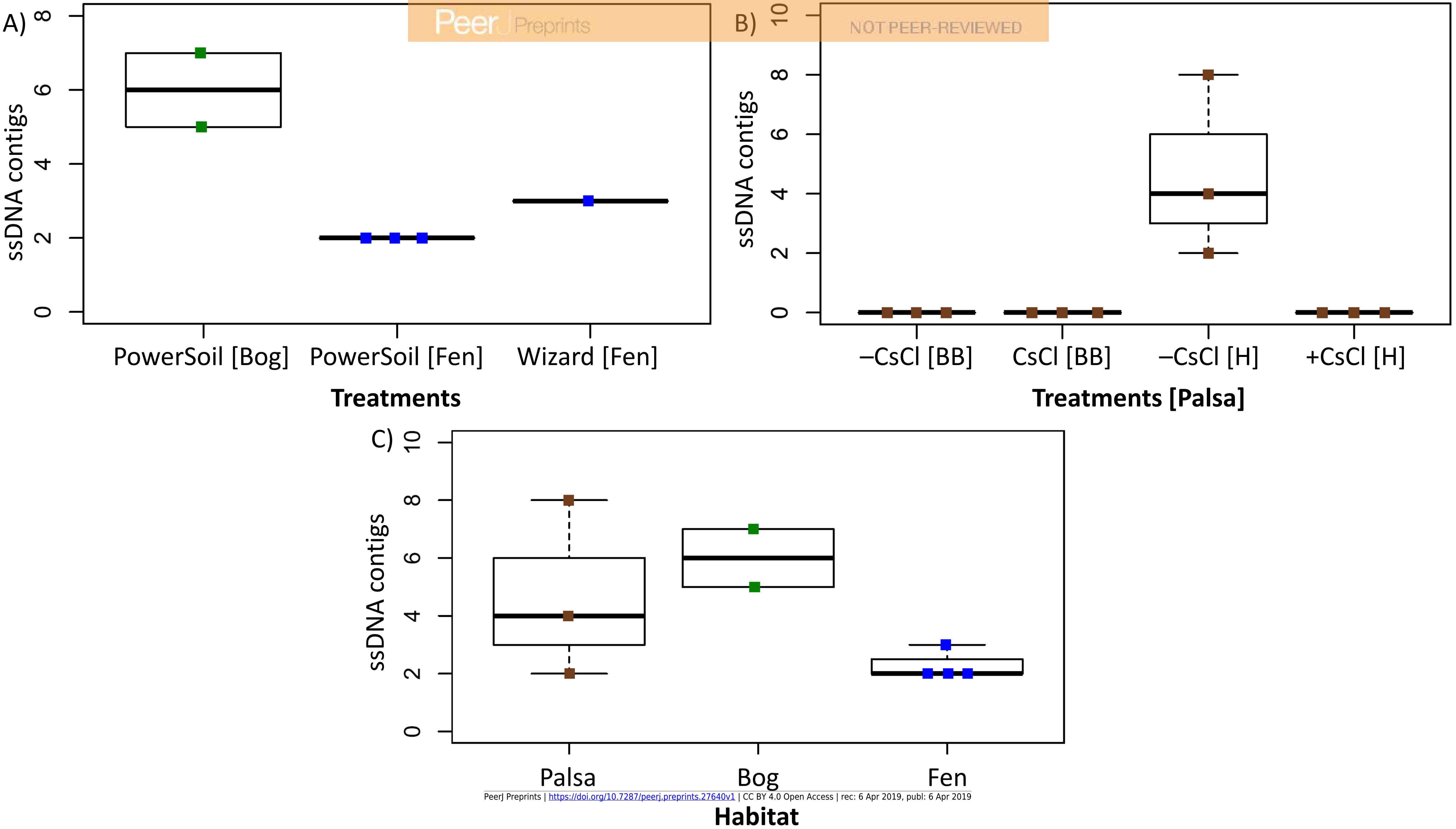

AU:
● 100
● ≥97
● ≥76
● 66
4
2
0
-2
-4
log$_{10}$(coverage)
66 vOTUs
+CsCl BB,C3
+CsCl H,C3
−CsCl BB,C3
−CsCl H,C3
+CsCl BB,C2
+CsCl H,C2
−CsCl BB,C2
−CsCl H,C2
+CsCl H,C1
+CsCl BB,C1
−CsCl BB,C1
−CsCl H,C1
12 Palsa viromes comparing 4 treatments

Figure 8(on next page)

Recovery of ssDNA viruses across habitats and methods.

A) ssDNA viral contigs from viromes in Experiment 2. The PowerSoil bog samples are grouped, as are the PowerSoil fen samples. The single Wizard virome from the fen habitat is also shown. B) ssDNA viral contigs from viromes in Experiment 2 grouped by the four treatments: +/– CsCl and bead-beating [BB] or heat [H] virion lysis method. C) ssDNA viruses from both Experiments are shown and grouped by habitat.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27640v1 | CC BY 4.0 Open Access | rec: 6 Apr 2019, publ: 6 Apr 2019