Distinct Circular Single-Stranded DNA Viruses Exist in Different Soil Types

Brian Reavy, Maud M. Swanson, Peter J. A. Cock, Lorna Dawson, Thomas E. Freitag, Brajesh K. Singh, Lesley Torrance, Arcady R. Mushegian, Michael Taliansky

The potential dependence of virus populations on soil types was examined by electron microscopy, and the total abundance of virus particles in four soil types was similar to that previously observed in soil samples. The four soil types examined differed in the relative abundances of four morphological groups of viruses. Machair, a unique type of coastal soil in western Scotland and Ireland, differed from the others tested in having a higher proportion of tailed bacteriophages. The other soils examined contained predominantly spherical and thin filamentous virus particles, but the Machair soil had a more even distribution of the virus types. As the first step in looking at differences in populations in detail, virus sequences from Machair and brown earth (agricultural pasture) soils were examined by metagenomic sequencing after enriching for circular Rep-encoding single-stranded DNA (ssDNA) (CRESS-DNA) virus genomes. Sequences from the family Microviridae (icosahedral viruses mainly infecting bacteria) of CRESS-DNA viruses were predominant in both soils. Phylogenetic analysis of Microviridae major coat protein sequences from the Machair viruses showed that they spanned most of the diversity of the subfamily Gokushovirinae, whose members mainly infect intracellular parasites. The brown earth soil had a higher proportion of sequences that matched the morphologically similar family Circoviridae in BLAST searches. However, analysis of putative replicase proteins that were similar to those of viruses in the Circoviridae showed that they are a novel clade of Circoviridae-related CRESS-DNA viruses distinct from known Circoviridae genera. Different soils have substantially different taxonomic biodiversities even within ssDNA viruses, which may be driven by physicochemical factors.

Virus particles are abundant in the environment and exceed the number of cellular organisms in marine and soil habitats by at least an order of magnitude (1, 2). The global population size is estimated to be >10^{30} virus particles (3), and the advent of deep sequencing of environmental samples has shown that virus genome sequences in sequence databases are a biased representation of virus diversity in the environment, as novel virus lineages are increasingly being discovered (4–7).

Many studies of virus populations in the environment have been performed in marine or aquatic habitats, which are amenable to concentration and purification of a relatively large virus biomass (8). Surveying viruses in soils has started recently, despite the recalcitrant character of many soil samples, and initial estimates of concentrations of viruses in soils are on the order of ~1.5 × 10^8 g^{-1} (wet weight) to ~10^9 virus particles g^{-1} (dry weight) (4, 9, 10). Data on changes to soil virus abundance and diversity caused by environmental factors have recently been reported; for example, it was shown that virus abundance and diversity are greater in wetland forest soil than in drier agricultural soils (4).

Most of the soil virus particles identified by electron microscopy were tailed phage types resembling Caudovirales, which account for 95% of all known bacteriophages (4, 11, 12). However, the complex morphology of soil debris makes it difficult to distinguish virus particles other than tailed phage (9, 10), and improved virus purification techniques showed that while the numbers of tailed bacteriophage (~4.8 × 10^6 g^{-1} [dry weight]) were similar to previous reports, they comprised only ~4% of the total virus particles present, and the remaining, overwhelming majority of particles consisted of spherical particles of various sizes, along with bacilliform, rod-shaped, and filamentous virus-like entities (10). Recent studies have also shown that 51 to 92% of virus particles in ocean samples are nontailed (13). One type of spherical virus that has been the subject of a number of recent studies is viruses containing circular Rep-encoding single-stranded DNA (ssDNA) (CRESS-DNA) genomes, which have been found to be more prevalent in a number of environments than previously thought (14–18).

In this work, we examined the viruses in several contrasting soil types, including agricultural and alpine soils and the rare coastal Machair soil (51), which is found only on the west coast of Scotland and in Ireland. Our analysis gives an initial outline of the virus loads in soils and reveals contrasts between viruses in different soils, which may be related to soil pH and/or water content. Taking into account the abundance and potential importance of ssDNA viruses (18), we have performed metagenomic analysis of...
### MATERIALS AND METHODS

**Soil samples.** Twenty samples of soil were obtained from the National Soil Inventory for Scotland (NSIS) archived collection (19) and from a supplementary collection of “rare soils”; both collections are housed by the James Hutton Institute (for more information, see Table S1 in the supplementary material). The soil samples were collected between 2007 and 2010 and stored at −80°C immediately after collection. In our experience, storage at −80°C and subsequent thawing does not affect the integrity of virus particles, and similar amounts of phage heads were observed in virus preparations from fresh soil samples and in frozen and thawed samples. The samples used in this study represent some of the main types of soil in Scotland. Machair is a coastal soil found only in the west of Scotland consisting of sea shell sand overlaid on peat and is used for low-intensity agriculture; the samples used in this study were from dune pastureland. Brown earths are higher input agricultural soils, and the samples analyzed here come from forest plantations or grassland. Alpine podzols are soils from mountain heaths, and humus-iron podzol samples were from heather moor, scrubland, or pine tree plantations. Five different representative samples of each soil type from different locations were used in the study, and details of the soil characteristics are given in Table 1 and Table S1 in the supplementary material.

**Virus purification.** A method for the extraction of viruses from 1-g samples of soils was adapted from the method described by Swanson et al. (10). One-gram soil samples were divided equally between two 2-ml Eppendorf tubes, and 1 ml of 0.067 M phosphate buffer, pH 7.5, and 250 µl of 1-mm-diameter sterile glass beads were added to each tube, which was shaken in a Qiagen Retsch tissue lyser for 5 min at 30 revolutions s⁻¹. The tubes were centrifuged at 10,000 relative centrifugal force (RCF) for 5 min, and the supernatant was collected. The pellets were resuspended in 1 ml phosphate buffer, and the extraction procedure was repeated twice. The supernatants for each sample were pooled, passed through a 0.45-µm Millex-HA low-protein-binding filter (Millipore, Ireland), and stored at −20°C before being thawed. The pellets were resuspended overnight in 1 ml of 0.067 M phosphate buffer, pH 7.5. The samples were centrifuged at 10,000 rpm for 5 min to clarify the supernatants, which were then centrifuged at 22,000 rpm through a 20% sucrose cushion in a 10% (vol/vol) CHCl₃ and centrifuged as described above. The supernatants were centrifuged at 22,000 rpm through a 20% sucrose cushion in a SW55 rotor for 2 h in an Optima L80-XL ultracentrifuge (Beckman Coulter). The pellets were resuspended overnight in 1 ml 0.033 M phosphate buffer, pH 7.5. The samples were centrifuged at 10,000 RCF for 5 min to clarify the supernatants, which were then centrifuged at 70,000 rpm (229,947 × g) in an MLA130 rotor for 1 h in an Optima MAX ultracentrifuge (Beckman Coulter), and the pellets were resuspended as described above.

**Electron microscopy.** Electron microscopy was performed as described by Swanson et al. (10). Approximately 20 digital photographs (magnification, ×15,000) were taken across the center line of selected gridholes for each sample. Virus particles in the photographs were counted, measured, and recorded as four morphotypes: tailed phage, spherical particles, thin filamentous particles, and thick filamentous particles. The totals for each soil sample were summed to give a total for each soil type. The percent abundance of each morphotype was then calculated for each soil type.

---

### Table 1 Details of soil samples used in this study

<table>
<thead>
<tr>
<th>Soil type</th>
<th>Soil sample</th>
<th>Location</th>
<th>pH</th>
<th>Organic matter (% C)</th>
<th>H₂O content (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpine podzol</td>
<td>Ben Vuirich no. 1</td>
<td>Mountain heath</td>
<td>3.90</td>
<td>20.3</td>
<td>8.05</td>
</tr>
<tr>
<td>Alpine podzol</td>
<td>Ben Vuirich no. 12</td>
<td>Mountain heath</td>
<td>4.08</td>
<td>28.7</td>
<td>2.56</td>
</tr>
<tr>
<td>Alpine podzol</td>
<td>Coire nan Clach no. 1</td>
<td>Mountain heath</td>
<td>4.13</td>
<td>30.7</td>
<td>7.04</td>
</tr>
<tr>
<td>Alpine podzol</td>
<td>Lochluichart no. 1</td>
<td>Mountain heath</td>
<td>4.36</td>
<td>20.3</td>
<td>5.42</td>
</tr>
<tr>
<td>Alpine podzol</td>
<td>Sgoran Dhub Mor no. 1</td>
<td>Mountain heath</td>
<td>4.18</td>
<td>17.6</td>
<td>2.86</td>
</tr>
<tr>
<td>Brown earth</td>
<td>Bohally Wood no. 1</td>
<td>Silver fir (Abies alba) plantation</td>
<td>4.35</td>
<td>5.13</td>
<td>2.66</td>
</tr>
<tr>
<td>Brown; Earth</td>
<td>Brae and Oape no. 2</td>
<td>Sitka spruce (Picea sitchensis) plantation</td>
<td>4.60</td>
<td>5.03</td>
<td>2.20</td>
</tr>
<tr>
<td>Brown earth</td>
<td>Chanlock 84 no. 2</td>
<td>Grassland</td>
<td>5.50</td>
<td>10.6</td>
<td>4.54</td>
</tr>
<tr>
<td>Brown earth</td>
<td>Craigton Row</td>
<td>Permanent and old ley pastures of rye grass and crested dogs tail (Lolio-Cynosoretum)</td>
<td>5.15</td>
<td>13.6</td>
<td>6.08</td>
</tr>
<tr>
<td>Brown earth</td>
<td>Hoardweel no. 1</td>
<td>Rotational ley pastures of rye grass and crested dogs tail (Lolio-Cynosoretum)</td>
<td>5.58</td>
<td>9.5</td>
<td>4.77</td>
</tr>
<tr>
<td>Humus-iron podzol</td>
<td>Balholmish no. 1</td>
<td>Dry boreal heather (Vaccinio-Eriecetum cinereae) moor</td>
<td>3.91</td>
<td>45.5</td>
<td>13.23</td>
</tr>
<tr>
<td>Humus-iron podzol</td>
<td>Cowal no. 21</td>
<td>Base-rich bracken scrub (Ranunculus ficaria-Pteridium aquilinum)</td>
<td>3.53</td>
<td>7.8</td>
<td>2.96</td>
</tr>
<tr>
<td>Humus-iron podzol</td>
<td>Drumguish no. 1</td>
<td>Scots pine (Pinus sylvestris) plantation</td>
<td>4.11</td>
<td>44.6</td>
<td>11.52</td>
</tr>
<tr>
<td>Humus-iron podzol</td>
<td>Edraderou no. 1</td>
<td>Lodgepole pine (Pinus contorta) plantation</td>
<td>4.13</td>
<td>45.8</td>
<td>13.36</td>
</tr>
<tr>
<td>Humus-iron podzol</td>
<td>Lower Dell no. 1</td>
<td>Native pinewood (Pinetum scoticae)</td>
<td>3.32</td>
<td>46.9</td>
<td>12.72</td>
</tr>
<tr>
<td>Machair</td>
<td>Coll no. 1</td>
<td>Eyebright-red fescue (Euphrasio-Festucetum arenariae) dune pasture</td>
<td>7.41</td>
<td>7.5</td>
<td>0.94</td>
</tr>
<tr>
<td>Machair</td>
<td>Coll no. 3</td>
<td>Eyebright-red fescue (Euphrasio-Festucetum arenariae) dune pasture</td>
<td>7.49</td>
<td>7.15</td>
<td>1.85</td>
</tr>
<tr>
<td>Machair</td>
<td>Eriskay no. 3</td>
<td>Eyebright-red fescue (Euphrasio-Festucetum arenariae) dune pasture</td>
<td>7.52</td>
<td>4.9</td>
<td>0.70</td>
</tr>
<tr>
<td>Machair</td>
<td>Hough no. 1</td>
<td>Eyebright-red fescue (Euphrasio-Festucetum arenariae) dune pasture</td>
<td>7.35</td>
<td>13.3</td>
<td>2.38</td>
</tr>
<tr>
<td>Machair</td>
<td>Hough no. 2</td>
<td>Northern marram grass (Elymo-Ammophiletum) dune</td>
<td>7.65</td>
<td>11.7</td>
<td>1.58</td>
</tr>
</tbody>
</table>

*a Sample was selected for metagenomic sequencing. More details are given in Table S1 in the supplemental material.*
using the Ninth Report of the International Committee on Taxonomy of Viruses (20).

**Virus DNA purification.** DNA was purified from virus particles following DNase digestion to remove contaminating host DNA as described by Swanson et al. (21).

**DNA sequencing.** Viral DNA was amplified using a GenomiPhi V2 DNA Amplification kit (GE Healthcare Life Sciences). Duplicate reactions were performed and pooled for each sample to minimize bias introduced by the phi29 DNA polymerase (22, 23). The DNA was purified using a QIAexII gel extraction kit (Qiagen) and resuspended at an appropriate concentration for sequencing. DNA sequencing was performed on a Roche 454 GS FLX+ platform by the University of Liverpool Centre for Genomic Research using one quadrant of a plate for each soil virus library.

**Microarray data accession numbers.** Initial analysis of the sequences was performed using MG-RAST v3 (24), which provided quality control and automated annotation for the individual sequences. Further analysis of the sequences was performed using METAGENassist (25), with the taxonomic assignments produced by MG-RAST. The sequences were also assembled into contigs using MIRA v3.2 (26) in single-threaded accurate de novo mode. The metagenomes are available on the MG-RAST server (http://metagenomics.anl.gov/) with the accession numbers 4552830.3 (Machair sequences), 4552818.3 (brown earth sequences), 4537815.3 (Machair contigs), and 4537814.3 (brown earth contigs). Complete amino acid sequences of circovirus-like replicases and of Microviridae major coat proteins (CPs) were aligned using MEGA6 (27) with the HKY85 substitution model. Maximum-likelihood phylogenetic trees were constructed from these alignments using PhyML 3.0 (28), with the HKY85 substitution model. One hundred bootstrap replicates were performed for each tree to assess the statistical support of the clades.

**RESULTS**

**Abundances and different morphologies of virus particles in soils.** Filtrates enriched with virus particles from the four soil types in this study were examined by electron microscopy. The four soil types included the unique Machair soils and three other types of soil: brown earths, which are commonly cultivated and are generally found in the drier, warmer low-lying areas, and podzols which are generally infertile soils and are less often cultivated. The podzols were subdivided based on topography into humus-iron podzols, which are found predominantly in forested areas, and alpine podzols, which are located in mountainous areas at high altitude. The Machair soil and alpine podzol are both classified as rare soils, and the brown earth and humus-iron podzols can be regarded as their respective nearest common equivalents within Scotland. Virus particles were identified by shape and size and characterized into four morphological types: tailed bacteriophage-like particles, spherical particles, thin filaments, and thick filaments (Fig. 1). There are \( \sim 1.2 \times 10^9 \) virus particles g\(^{-1}\) (dry weight) in soils in the Dundee area of Scotland (10), which is in agreement with virus abundances reported for other temperate soils (4, 9, 29). Here, we used a Dundee soil sample (a freely drained brown earth) as a reference to assess total virus particle abundances in the different soils studied here. We found that all four soil types contained virus particles in amounts comparable to that found in the Dundee reference soil, i.e., between \( 0.8 \times 10^9 \) and \( 1.9 \times 10^9 \) virus particles g\(^{-1}\) (dry weight).

The relative abundances of the different virus morphotypes were assessed for each type of soil. Spherical particles and thin filaments predominated in all four soil types, and tailed phages constituted less than 15% of the total particles in all the soils (Fig. 2). The four soil types have different distributions of virus morphotypes, but the Machair soil is markedly different from the other three soils in having a more even distribution of virus types, with significantly higher numbers of both tailed phages and thick filaments than the other soils. This difference in the Machair soil compared to the other three soils examined was confirmed by canonical variant analysis using Genstat (VSN International, United Kingdom) (Fig. 3).

Tailed bacteriophages (order Caudovirales) had relative abundances ranging between 2.2% and 14.4%, with the highest abundance in the Machair soils and the lowest in the alpine podzols. Tailed phages were sometimes seen as complete particles (Fig. 1) but were more often present as separate heads and tails. This is possibly due to the preparation procedures for microscopy, as we do not routinely see great differences between frozen samples and those that have been processed without freezing. Detached phage heads could be easily identified due to their dark interior, which is presumably the result of stain penetration through damaged areas of the capsid. These heads were spherical, ranging in diameter from \( \sim 30 \) nm to 100 nm, or elongated, with the longest side ranging in size from 50 to 200 nm. The tails appeared to be either siphoviruses or myoviruses. The tails of Siphoviridae are long and flexuous and often appeared striated. The tails of bacteriophages belonging to the family Myoviridae are shorter, rigid, and contractile; we saw several forms differing in overall shape, size, and extent of contraction. Bacteriophages of the family Podoviridae, which have short, contractile tails, were not seen as intact particles, and short intact tails could not be identified in the virus preparations. However, it is likely that some of the phage heads belonged to podovirus particles; podovirus sequences were identified in the samples (see below).

Spherical virus particles varying in diameter from 20 to 120 nm were seen in each of the four soil types; abundances ranged from 32.3% to 67.9% of the total particles, with the highest abundance in the humus-iron podzols and the lowest in the Machair soils. Known RNA-containing spherical viruses have sizes from 22 to 150 nm; therefore, many of the virus particles seen may have had RNA genomes. However, only DNA genomes were sequenced in this study, and in general, we obtained sequence matches to virus families that had capsids smaller than 30 nm. We did not detect sequences from any virus families whose expected virion sizes would be in the 30- to 120-nm size range. This could be for a number of reasons: the observed particles may be from known families of viruses whose genomes were not represented because of the DNA preparation methods we used, they could be novel groups of viruses whose sequences did not generate BLAST hits, or they could be morphological variants of virus groups that typically have different size ranges.

Thin filamentous particles with diameters of 5 to 10 nm and lengths of 50 nm to 1,000 nm were also found in all four soils; abundances ranged from 27.5% to 55.5% of the total particles observed, with the highest abundance in the alpine podzols and the lowest abundance in the humus-iron podzols. This morphology is most reminiscent of the ssDNA viruses in the genus Inovirus (7 nm in diameter and 700 to 2,000 nm long), which have bacterial hosts. Moreover, sequences related to inoviruses were found in the Machair Hough no. 1 soil sample (see Table 3), suggesting that at least some of these particles may be inovirus related. Other known viruses with this kind of flexuous filamentous particles, such as Potyviridae, Closteroviridae, and Flexviridae, all have ssRNA genomes, which would not have been sequenced by our protocols.

Thick filamentous particles were found in all soils; abundances
ranged from 1.6% to 25.0% of the total particles observed, with the highest abundance in Machair soils and the lowest abundance in the alpine podzols. These particles were especially well preserved in the Machair soils. They varied considerably in width and length, but most were 15 to 50 nm in diameter and 100 to 600 nm long and resembled the double-stranded DNA (dsDNA)-containing Lipothrixviridae (24 to 38 nm in diameter and 410 to 2,200 nm long), which infect thermophilic Archaea species. We have no sequence evidence of the Lipothrixviridae in any of our samples and so cannot be certain whether the particles seen in the electron microscope are actually virions.

Oval and bacilliform particles were also seen in the soils, but in smaller numbers than the other morphotypes. They varied in size, but most particles were 50 to 80 nm by 35 to 75 nm. Although there are some virus genera with bacilliform particles, their dimensions are not in the same range.

Some physicochemical characteristics of the soil samples obtained from the NSIS database were examined to determine if any of the characteristics could be correlated with the differences in

![FIG 2](image-url)

FIG 2 Relative abundances of different types of virus particles in four Scottish soil samples. Viruses were purified from each soil sample listed in Table 1, and the particles observed by electron microscopy were counted for each sample as described in Materials and Methods. The totals for all the samples in each soil type were summed, and the abundances of the four types of virus particles are expressed as percentages of the total number of particles counted for each soil type. The total numbers of particles counted for each soil type were as follows: Machair, 2,885; brown earth, 1,794; humus-iron podzol, 1,234; and Alpine podzol, 1,406.
viruses populations observed in the Machair soils. Both the Machair and the brown earth soils had low organic matter contents (percent C) compared to the podzols (Table 1). The Machair soils differed significantly from the other soil types in having low water content and high pH values, which could contribute to the stability of virus particles.

**Sequence analysis of virus metagenomes.** One sample of each of the Machair (Hough no. 1) and brown earth (Hoardweel no. 1) soil samples was chosen for sequence analysis, as the virus population in the Machair soil was significantly different from those in the other soils and the brown earth is considered the closest common equivalent to the Machair sample. DNA was extracted from purified virus particles from these samples, and multiple-displacement amplification was performed before sequencing in order to preferentially amplify circular ssDNA genomes, as it was expected that the abundant spherical morphotype (32% of the total particle) preferentially amplify circular ssDNA genomes, as it was expected that the abundant spherical morphotype (32% of the total particle) would contain a large percentage of this type of genome, as had been shown in earlier studies (30, 31).

Over 200,000 sequence reads were generated for each metagenome, giving a total of ~126 Mbp of sequence for the Machair sample and ~77 Mbp of sequence for the brown earth sample (Table 2). The sequences were initially analyzed using the MG-RAST pipeline (24). Most of the predicted protein sequences (87.9% of Machair and 97.3% of brown earth sequences) were orphans, i.e., had no significant matches to translated genes in a BLASTX search of the SEED nonredundant database (32), with an E value cutoff of 1e−3, or in parallel searches of accessory databases used by MG-RAST using appropriate algorithms and significance selection criteria. These figures for orphans are within the range of 72 to 99% reported for viromes and comparable to the average of 93.7% for aquatic environments (33). Also, only 0.08% of the annotated sequences were identified as RNA features, figures that are comparable to those from other virus metagenomic analyses (33).

The taxonomic distributions of the assignable sequences differed markedly between the two soil samples. The predominant identifiable sequences in the Machair sample were of viral origin (65% of the assigned sequences), whereas sequences annotated as bacterial were predominant in the brown earth sample (74% of the assigned sequences) (Fig. 4). The taxonomic profiles generated by MG-RAST were examined further using METAGENAssist, which characterized the recognizable sequences into 304 separate taxonomic groups at the family level. The distribution of sequences among these groups was used to calculate Simpson’s index of biodiversity, which was 0.91 for the brown earth virome and 0.63 for the Machair virome, indicating that the brown earth virome was more taxonomically diverse than the Machair virome.

Most of the sequences of viral origin were, as expected for both samples, derived from ssDNA (CRESS-DNA) viruses. The distributions of ssDNA viruses differed between the two samples, with Microviridae sequences (represented by small spherical viruses mainly infecting bacteria) comprising 84.6% of the virus sequences in the Machair sample but only 50.3% in the brown earth sample (Table 3). Conversely, sequences related to viruses of the family Circoviridae were more common in the brown earth virome (see below), where they accounted for 18.3% of the identifiable viral sequences compared to 0.7% in the Machair virome. This dissimilarity between the two soils is likely not only to reflect differences in virus populations in these habitats, but also indirectly to suggest different relative abundances of their prokaryotic and eukaryotic hosts.

**Diversity within ssDNA viruses in soils.** In the viromes of surface seawater, CRESS-DNA virus genomes have been preferentially amplified by techniques similar to those used here, and this approach led to high percentages of sequences (66 to 73%) that could be assembled into nearly full genomes (34). In contrast, we found that only a relatively small number of the sequences in our soil viromes assembled into contigs (Table 2), with only ~12% of Machair sequences assembling into 10,985 contigs and 11.7% of the brown earth sequences assembling into 10,928 contigs, suggesting that there is greater diversity of CRESS-DNA virus sequences in our soil samples than was found in the study of Kim and Bae (34). To analyze the diversity of viruses in the soil samples, we aligned the Microviridae major coat protein sequences from 48 contigs assembled from the Machair soil virome with those of viruses of enterobacteria and 81 Microviridae sequences derived from various environmental viromes (16) and built phylogenetic trees. Figure 5 shows an abbreviated version of the tree for clarity of display; the complete tree can be seen in Fig. S1 in the supplemental material. Almost all major CP sequences from the Machair soil grouped within the subfamily Gokushovirinae of the Microviridae close to a subfamily, Pichovirinae, recently identified by Roux et al. (16). The subfamily Pichovirinae has previously contained only sequences assembled from aquatic environments (Bourget sequences, Pavin110, and JCVI001). One Machair soil sequence (M1738) actually clustered within this aquatic environment group, possibly because the coastal nature of the Machair soil means it is an interface between aquatic and terrestrial environments. Isolates in the Gokushovirinae have been identified as infecting intracellular parasites, such as Chlamydia, Spiroplasma, or Bdellovibrio (16). Chlamydia and Spiroplasma are obligate parasites of eukaryotic organisms, but Bdellovibrio parasitizes bacteria that are abundant in the environment. Pyrosequencing analysis of 16S rRNA genes was performed in order to identify putative hosts of the Gokushovirinae in soils. BLAST searches of the 16S rRNA sequences against the NCBI database indicated the presence of Bdellovibrio-associated sequence types in Machair but not in brown earth soils (see Table S2 in the supplemental material). Chlamydia-associated sequence types were retrieved from both Machair and brown earth soils, with the majority of sequences retrieved from Machair soils. BLAST matches were.

**FIG 3** Canonical variant analysis of virus morphological types in Scottish soils. +, Machair; ○, humus-iron (H.I.) podzol; □, Alpine (A.) podzol; ×, brown earth.
mostly to uncultured *Chlamydia* or *Bdellovibrio* soil clones or to bacterial sequences isolated from soil-inhabiting organisms, such as the worm *Eisenia fetida* or amoebas. No sequences with BLAST match similarity to *Spiroplasma* were retrieved from the sequencing libraries.

These data support and expand recent observations describing the cosmopolitan nature of the subfamily *Gokushovirinae* in sediments, sewage, seawater, and freshwater environments (22) to soil environments. Such a remarkable representation of gokushoviruses presumably indicates their significant ecological role; this remains an important topic for future studies. In addition, two Machair soil sequences, M2198 and M4578 (Fig. 5, gray shading), formed a pair that clustered close to the viruses of the *Pichovirinae* in the abbreviated phylogenetic analysis (Fig. 5) but close to the genus *Microvirus*, whose members infect *Enterobacteria*, in the more complete analysis (see Fig. S1 in the supplemental material).

We performed a similar analysis with sequences annotated in the MG-RAST analysis as resembling the replicase protein of *Circoviridae* and compared them with the replicates of a wide range of known circoviruses and circovirus-related metagenomic samples (35). Interestingly, none of the full-length replicase sequences from the soil samples clustered with viruses of the known *Circoviridae* genera (such as *Circovirus* or *Cyctovirus*). Instead, all the soil sequences were in a group of recently described novel CRESS-DNA viruses (35). Figure 6 shows an abbreviated version of the tree for clarity of display; the complete tree can be seen in Fig. S2 in the supplemental material. Most of the sequences (BEU125 to BEELKF5) from the brown earth virome formed a single cluster within the CRESS-DNA group, but three other sequences (BEDUBR2, BEBRITL, and BECOVSX) were dispersed within it. The sequences from the Machair virome were more dispersed within the CRESS-DNA virus group than those from the brown earth virome. These results indicate that varied ssDNA viruses, with as yet uncharacterized host ranges, occur in soils.

**Other virus sequences present in soils.** Sequences matching other ssDNA virus families were also present in the two soil viromes, but in smaller amounts, and included *Geminiviridae*, *Nanoviridae*, and *Parvoviridae* (Table 3). Interestingly, sequences related to viruses in the family *Inoviridae* of filamentous ssDNA bacteriophages were also found in the Machair virome. *Inoviridae*-related sequences have not been commonly found in metagenomic studies of virus populations, possibly due to their loss during purification (22). It is not evident to us what was different about our purification protocol, and some Machair habitats possibly retain more *Inoviridae*-related viruses than other soils.

As expected with multiple-displacement amplification, dsDNA virus sequences were present in small amounts: 0.2% of total virus sequences in the Machair soil and 2.06% in the brown earth. Virus sequences from the order *Caudovirales*, especially the family *Siphoviridae*, were the most common among these dsDNA virus genes.

**Metabolic subsystems present in the viromes.** A wide range of molecular functions are commonly found in virus metagenomes (8). Some of these functions are required for the reproduction of viruses themselves, while others are likely to be included in phage DNA as the result of lysogeny. Distributions of gene products by predicted function in the two soil viromes were generated using MG-RAST, which annotated 12.1% of the Machair and 2.7% of the brown earth predicted protein fragments by similarity to proteins with known functions; 70.8% of Machair and 60.3% of brown earth annotated protein fragments were assigned to functional categories, the most numerous of which are shown in Fig. 7. Calculation of Simpson’s index of diversity using these profiles gave similar values for both viromes (0.86 for the Machair virome).

### Table 2 Details of sequencing of two Scottish soil virus metagenomes

<table>
<thead>
<tr>
<th>Soil</th>
<th>Total no. of reads</th>
<th>Total bp read</th>
<th>rRNA sequences</th>
<th>Orphans</th>
<th>Total bp assembled into contigs</th>
<th>No. of contigs assembled</th>
<th>Longest contig (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machair (Hough no. 1)</td>
<td>214,880</td>
<td>126 million</td>
<td>0.08</td>
<td>87.9</td>
<td>15 million</td>
<td>10,985</td>
<td>13,949</td>
</tr>
<tr>
<td>Brown earth (Hoardweel no. 1)</td>
<td>211,172</td>
<td>77 million</td>
<td>0.08</td>
<td>97.3</td>
<td>9 million</td>
<td>10,928</td>
<td>8,792</td>
</tr>
</tbody>
</table>

### Table 3 Virus sequences identified in soil viromes

<table>
<thead>
<tr>
<th>Group</th>
<th>Order</th>
<th>Family</th>
<th>% of total sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unclassified</td>
<td></td>
<td></td>
<td>Machair 11.76 Brown earth 26.33</td>
</tr>
<tr>
<td>dsDNA</td>
<td>Caudovirales</td>
<td>Myoviridae</td>
<td>0.02 0</td>
</tr>
<tr>
<td></td>
<td>Caudovirales</td>
<td>Podoviridae</td>
<td>0.01 0.13</td>
</tr>
<tr>
<td></td>
<td>Caudovirales</td>
<td>Siphoviridae</td>
<td>0.18 1.29</td>
</tr>
<tr>
<td>ssDNA</td>
<td>Microviridae</td>
<td></td>
<td>84.6 50.32</td>
</tr>
<tr>
<td></td>
<td>Circoviridae</td>
<td></td>
<td>0.73 18.32</td>
</tr>
<tr>
<td></td>
<td>Geminiviridae</td>
<td></td>
<td>0.35 1.03</td>
</tr>
<tr>
<td></td>
<td>Inoviridae</td>
<td></td>
<td>0.01 0</td>
</tr>
<tr>
<td></td>
<td>Nanoviridae</td>
<td></td>
<td>2.32 1.55</td>
</tr>
<tr>
<td></td>
<td>Parvoviridae</td>
<td></td>
<td>0.02 0.39</td>
</tr>
</tbody>
</table>

![FIG 4 Phylogenetic distribution of sequences in soil viromes. The sequences were analyzed in MG-RAST, and the best BLAST hits with E values of <10^-4 are indicated for the Machair and brown earth viromes.](http://aem.asm.org/.../images/figure4.png)
FIG 5 Phylogenetic tree of Microviridae major CP sequences. Complete Microviridae-related major coat protein amino acid sequences were aligned using MEGA6 with the MUSCLE algorithm. The alignment was edited manually, and PhyML was used to generate the tree. Bar = 1.0 amino acid substitution/site. Microviridae subfamilies are indicated on the right and are shaded in different colors; the Machair virus sequences that grouped close to the Pichovirinae are shaded in gray. The colored squares next to the sequence names indicate the environments they were obtained from, where known, according to the ecosystem color key. Bootstrap values of 80 or greater are indicated on the tree by colored circles, and values of 60 to 79 are indicated by open circles.
FIG 6 Phylogenetic analysis of Circoviridae replicate protein sequences. Complete Circoviridae-related replicate sequences were aligned, and a phylogenetic tree was constructed as described for Fig. 5. Bar = 1.0 amino acid substitution/site. Virus subfamilies are indicated on the right and are shaded in different colors. The colored squares next to the sequence names indicate the environments they were obtained from, where known, according to the ecosystem color key. Bootstrap values of 80 or greater are indicated on the tree by colored circles, and values of 60 to 79 are indicated by open circles.
and 0.88 for the brown earth virome), suggesting that the two viromes were similarly diverse with respect to the metabolic functions they contained. The main metabolic functions, such as carbohydrate metabolism, amino acid and derivatives metabolism, virulence, protein metabolism, DNA and RNA metabolism, and nucleoside and nucleotide metabolism, previously identified in other viral metagenomes (36), were also the most abundant in the Machair and brown earth viromes. Despite the similarity between the two soil viromes, some differences in the relative amounts of the different functions were apparent. The two soil virus metagenomes had different amounts of the amino acids and derivatives subsystem (22.68% in the Machair virome compared to 9.55% in the brown earth virome). The mean percentage for amino acids and derivatives in 42 viromes studied earlier is 10.13% (36), i.e., much more similar to that found in the brown earth virome. METAGENassist was also used to generate metabolic profiles using 106 identified metabolic subsystems, the most prominent of which are shown in Fig. 7. In this analysis, putative factors of sulfur metabolism were overrepresented in the brown earth virome compared to the Machair virome in the METAGENassist analysis (Fig. 7B).

**DISCUSSION**

Our electron microscopy observations showed that the abundance of virus-like particles in four soil types in this study is of the same order as that found previously in brown earth soil from the Dundee area (10). However, sometimes differences in virus abundances in different soils can be seen, which appear to correlate with soil moisture and organic matter content (4), and this may be due to specific environmental conditions pertinent to those soils. We also found that spherical particles were approximately twice as abundant as tailed phages and slightly more abundant than either thick or thin filamentous particles; all this is in broad agreement with observations that nontailed viruses constitute more than half of the total viruses in marine samples (13). The distribution of virus morphologies in Machair soil samples was significantly different from those in the other three soil types, and the four morphological types we used for classification were more equally rep-
resented there than in other types. Soil virus communities may reflect the microbial host populations, and a greater diversity of microorganisms is found in low-input agricultural systems such as Machair than in more intensely managed farmland (37). Soil physicochemical characteristics, such as pH, can also influence microbial community structures (38). The Machair soils thus differed from the other soils analyzed in having higher pH, lower water content, and presumably greater microbial diversity, which influence the virus population structure. Physicochemical attributes could also act to select certain virus types based on particle stability under certain conditions (e.g., pH), and this selection of the virus population could subsequently affect microbial populations in the soil by selective predation of the hosts that those viruses infect. The relationships between virus and microbial host populations are thus likely to be dynamic and complex, with soil physicochemical characteristics affecting both virus and host populations in different ways and virus and host populations affecting each other. Changing land use practices or other environmental factors that alter soil physicochemical characteristics could thus affect soil microbial communities directly or indirectly by altering virus population structures.

The development of next-generation sequencing platforms has been followed by a large number of metagenomic studies in many environments, but relatively little work has been done in soils, despite the fact that soil forms the biogeochemical basis upon which most terrestrial life depends. Close to 90% of the world’s bacterial population and most of its diversity is contained in soils (39), and the same is true for many eukaryotic organisms (40). Moreover, there remains a substantial gap in our understanding of virus communities in soils. Sequencing of soil virus populations has been performed on a pilot scale, with at most ~2,000 sequences analyzed for each soil type (30, 40). In this work, we have extended our knowledge of viruses in soils, particularly for viruses with circular ssDNA genomes, and have shown that populations of these viruses differ between soil samples. Gokushovirus-related sequences were more prominent in the Machair soil. Gokushoviruses are traditionally recognized as narrowly targeted, niche-specific viruses infecting obligate parasitic bacteria, such as Chlamydia and Spiroplasma, which parasitize eukaryotic organisms, or Bdellovibrio, which parasitizes bacteria, but gokushoviruses have recently been found in other environmental samples, such as water or sediments (18). This work expands these observations and shows the presence of gokushoviruses in soils. Remarkably, data from a 454 pyrosequencing analysis of 16S rRNA genes indicated the presence of Bdellovibrio-associated sequence types in Machair but not in brown earth soils, and this may at least partly contribute to the relative preponderance of gokushoviruses in the Machair soil.

A greater percentage of Circoviridae-related novel CRESS-DNA virus sequences were present in the brown earth soil than in the Machair soil. Circoviridae-related viruses commonly infect eukaryotic hosts, and it has been shown that the soil moisture content can affect certain soil-inhabiting invertebrates, with greater numbers being found at higher moisture levels (41). It is thus likely that the higher water content in the brown earth soil (see Table S1 in the supplemental material) may mean that there are more eukaryote hosts, such as invertebrates, present in that soil than in the Machair soils, and this could account for the greater prevalence of Circoviridae-related novel CRESS-DNA virus sequences in that soil. These differences in the virus communities in the Machair and brown earth soils are consistent with previous metagenomic studies on different kinds of soils, which also showed that virus population profiles in different soils are genetically unique (40). Factors contributing to these differences may include highly local adaptation of phage populations in soils to their specific host bacteria (42) and rapid coevolution of phages and their bacterial hosts in soil (43, 44). We have probably analyzed only a small fraction of the virus diversity in these soil samples, and it would be of interest to determine if, with deeper sequencing, the profiles of these viruses could be extended and, if they prove to be soil specific, may even prove to be suitable as diagnostics of soil traces for agricultural, environmental, or forensic purposes.

The Microviridae major CP sequences that we analyzed from soils grouped almost exclusively with gokushoviruses. It is becoming clear that gokushoviruses are probably ubiquitous in the environment (18), and the results described here extend their presence to soils. Two of the Microviridae major CPs (M2198 and M4578) group close to the genus Microviridae and may represent a novel group within the Microviridae. Some of the Circoviridae-related novel CRESS-DNA virus sequences may have originated from viruses infecting larger organisms, such as invertebrates or mammals. Many of these hosts are presumably soil inhabiting, but a subset of the Circoviridae-related novel CRESS-DNA virus sequences may have originated from viruses that were deposited from organisms that live above the soil. The extent to which soils could act as reservoirs of viruses for organisms not normally inhabiting soil but which come into contact with soil is worth further investigation. The evidence from both the Microviridae-related major coat protein sequences and the Circoviridae-related novel CRESS-DNA virus replica sequences, along with the low percentage of sequences that assembled into contigs, indicates that the soils examined contain different populations of viruses.

Viruses in soils may have at least two important roles. First, as pathogens, they have the ability to modulate populations of soil microorganisms and therefore to change the biology and functions of the soils. Second, by being agents of horizontal gene transfer, viruses can facilitate the rapid adaptation of soil bacteria to changing conditions by transferring beneficial genes. Horizontal gene transfer has often been regarded as being mediated largely by phages with dsDNA genomes, but Microviridae also integrate into host chromosomes (45) and acquire host genes (16). The Microviridae have smaller genomes and probably transfer genes at a lower frequency than transducing dsDNA phages, but the abundance and variety of microviruses in different environments observed by us and others (14, 17, 18, 30, 33, 46–49) suggest that they may play a more important role in gene transfer than had previously been suspected.

The viromes in the Machair and brown earth soils had different taxonomic compositions, but most categories of metabolic functions were represented in both viromes. Previous studies on viromes from various environments have also shown that virus assemblages that differ taxonomically carry genes involved in similar metabolic functions and that there is some adaptation of the metabolic functions to the environment from which the virome originates (36). Viromes, therefore, appear to act as environment-specific reservoirs of genes for metabolic functions (50). The two soil virus metagenomes studied here may differ in amino acid metabolism and sulfur metabolism.

This study reveals high representation of ssDNA viruses in
soils, with differences in morphology and taxonomic profiles in different soils appearing to be related to the physicochemical characteristics of the soils and possibly also to microbial biodiversity. However, there was full representation of metabolic functions in both soil viromes examined, suggesting that CRESS-DNA viruses may play more pronounced ecological roles, including pathogenicity and gene transfer, than was recognized before.

ACKNOWLEDGMENTS

This work (B.R., M.M.S., P.J.A.C., L.D., L.T., T.E.F., and M.T.) was funded by the Rural and Environment Science and Analytical Services Division of the Scottish Government (RESAS). A.R.M.’s work on the manuscript was supported in part by the Independent Research and Development Program of the U.S. National Science Foundation (NSF).

The statements and opinions expressed in this work are made in a personal capacity and do not constitute endorsement by the NSF or the government of the United States.

We thank Lucinda Robinson, the James Hutton Institute, Aberdeen, and Waleed Abu Al-Soud, Department of Biology, University of Copenhagen, Copenhagen, Denmark, for help with 16S RNA gene analysis.

REFERENCES


