

**The ecology of testate amoebae (Protista) in South-Central
Alaska peatlands: building transfer-function models for
palaeoenvironmental studies**

Richard J. Payne ^{1,2*}, Keiko Kishaba ³, Jeff J. Blackford ^{1,4}, and Edward A. D. Mitchell
^{3,5,6}

¹ Department of Geography, Queen Mary, University of London, Mile End Road,
London E14NS, United Kingdom

² Department of Biology, University of Bergen, Allègaten 41, N-5007 Bergen, Norway

³ Department of Biological Science, University of Alaska Anchorage, 3211
Providence Drive, Anchorage, AK, 99058, U.S.A.

⁴ School of Environment and Development, The University of Manchester, Oxford
Road, Manchester M13 9PL, United Kingdom

⁵ Laboratory of Ecological Systems (ECOS), ENAC, Swiss Federal Institute of
Technology (EPFL), Station 2, CH - 1015 Lausanne, Switzerland

⁶ Swiss Federal Research Institute WSL, Antenne Romande, Station 2, CH - 1015
Lausanne, Switzerland

* Current address: The British School at Athens, Odhós Soudhías 52, 10676 Athens,
Greece.

Abstract

Testate amoebae are valuable indicators of peatland hydrology and have been used in many palaeoclimatic studies in peatlands. Because the species' ecological optima may vary around the globe, the development of transfer function models is an essential prerequisite for regional palaeoclimatic studies using testate amoebae. We investigated testate amoebae ecology in nine peatlands covering a 250Km North-South transect in South-Central Alaska. Redundancy analysis and Mantel tests were used to establish the relationship between the measured environmental variables (water table depth and pH) and testate amoebae communities. Transfer function models were developed using WA, WA-PLS and ML techniques. Model prediction error was initially 15.8cm for water table depth and 0.3 for pH but this was reduced to 9.7cm and 0.2 by selective data exclusion. The relatively poor model performance compared to previous studies may be explained by the limitations of one-off water table measurements, the very large environmental gradients covered and by recent climatic change in the study area. The environmental preferences of testate amoebae species agree well with previous studies in other regions. This study supports the use of testate amoebae in palaeoclimate studies and provides the first testate amoebae transfer function from Alaska.

Keywords: Protozoa; *Sphagnum*; RDA; DCA; Mantel tests; community ecology

Introduction

Knowledge of how climate has changed in the past and how these changes have affected ecosystems is critical to understanding and predicting current and future climate change and its potential impact on ecosystems. Peatlands have been widely used for climate reconstruction because; although they do not offer the possibility of reaching annual resolution, they are ideal for studies at the decadal and centennial time scales which are relevant to humans and ecosystems. In such studies, a range of techniques is used including measures of peat humification, pollen and spores, and the macroscopic remains of plants. One of the most useful techniques in palaeo-climate studies from peatlands is testate amoebae analysis (Charman *et al.*, 2004; Blundell and Barber, 2005). Testate amoebae are a polyphyletic assemblage of heterotrophic and mixotrophic protists belonging to at least two separate taxonomic groups; the Arcellinida and the testate amoebae with filose pseudopodia (Meisterfeld, 2000a, 2000b). In peatlands, testate amoebae are very abundant and diverse and the species have well defined ecological preferences, particularly with respect to depth to water table (DWT), making them extremely useful hydrological indicators (Mitchell *et al.*, 1999; Booth, 2001, 2002). The test (shell) of testate amoebae can remain preserved and identifiable for millennia, making them a valuable tool in palaeoecological studies (Tolonen, 1986). Transfer functions derived from the study of the modern ecology of testate amoebae species allow reconstruction of changing mire surface wetness, which in ombrotrophic peatlands is primarily a function of climate. Indeed, palaeoclimate reconstructions using testate amoebae have been shown to compare favourably with measured water table depth

and climate for the recent past and with other proxy-climate records (Charman and Hendon, 2000; Charman *et al.*, 2001, 2004; Booth and Jackson, 2003; Schoning *et al.*, 2005).

Early studies of testate amoebae ecology were largely qualitative, and palaeoecological studies based upon them could only imply changing environmental conditions in vague terms (eg. Aaby 1976). However, in recent years quantitative techniques have been applied to testate amoebae studies, and it is now possible to investigate the ecology and palaeoecology of testate amoebae in a more statistically rigorous manner through the use of species-environment (transfer function) models. Previous quantitative ecological studies of testate amoebae have been geographically restricted, particularly to Europe, eastern North America, and New Zealand (Bobrov *et al.*, 1999; Booth, 2001, 2002; Charman and Warner, 1992, 1997; Lamentowicz and Mitchell, 2005; Mitchell *et al.*, 1999; Tolonen *et al.*, 1994; Warner and Charman, 1994; Wilmshurst *et al.*, 2003; Woodland *et al.*, 1998).

Although there is evidence that the hydrological preferences of many testate amoebae taxa are consistent between regions (Booth and Zygmunt, 2005) exceptions may exist for some taxa (Charman and Warner, 1997) and community structure often differs between regions. Therefore the study of testate amoebae ecology in South-Central Alaska and the development of new transfer function models based on these data is a prerequisite to their use in paleoecological studies in this region and will also add to our wider knowledge of testate amoebae ecology and biogeography. The aims of this study are to determine the hydrological and pH preferences of testate amoebae in Alaskan peatlands, to produce transfer function

models to allow palaeo-environmental reconstruction and to compare the ecology of testate amoebae in Alaska with elsewhere in the world.

Methods

Study sites

Nine sampling sites were selected to cover the range of peatland types found throughout South-Central Alaska. The sites are distributed along a 250 Km North-South transect from the southern Kenai Peninsula to the Susitna River valley (Figure 1) and span a range of climatic regimes from a moist oceanic climate in the south to a drier and more continental regime further north. **Kachemak** (59°47'N, 151° 09'W) is the most southerly site in this study. The peatland occupies a terrace approximately 200x100 m in size in the hills overlooking Kachemak Bay on the southern edge of the Kenai Peninsula. The site is on a slight south-facing slope and clearly receives some drainage from hillsides to the North. The mire is characterized by deep hollows and steep hummocks and is probably the most mesotrophic site in this study. The vegetation of the site is sedge-dominated, with abundant *Trichophorum caespitosum* and *Sphagnum* spp. mostly restricted to hollows. **Ninilchik** (60°00'N, 151° 42'W) is a large, flat peatland occupying a basin approximately 500x200 m in size near the eastern shore of Cook Inlet. The site is surrounded by coniferous forest but is largely treeless. In common with many peatlands in this region, the site appears to be minerotrophic and may represent the infilling of a lake basin. When sampled the site was very dry, with many plants appearing desiccated. Vegetation of the site includes *Sphagnum* spp., *Empetrum nigrum*, and sedges. No water table depth

measurements are available from this site. **Clam Gulch** peatland (60°14'N, 151° 22'W) is approximately 800x250m in size and occupies a shallow basin on the western edge of the Kenai Peninsula. When sampled the site was quite wet, with three large pools occupying the centre of the site; parts of the mire may be floating on these water bodies. The site is flat, with some low hummock and hollow microforms. The site is largely free of trees, and the surrounding *Picea* woodland extends little beyond the edge. Vegetation of the site is heavily dominated by *Sphagnum* with *E. nigrum*, *Ledum groenlandicum*, and *Cladonia portentosa* on hummocks. **Sterling** peatland (60°31'N, 150° 31'W) is approximately 300x200 m in size and, like several other studied sites, may represent lake infill; the site receives little obvious drainage. The mire has notable holes in the surface and has steep hummocks; the vegetation is dominated by *Sphagnum*. **Jigsaw Lake** peatland (60°45'N, 150° 30'W) is adjacent to Jigsaw Lake, an oligotrophic, closed-basin lake located within the Kenai National Wildlife Refuge (KNWR). The mire is located in a small cove facing northwest and surrounded by *Picea* forest. The peatland edge floats on the water-body, but most of its surface is approximately 1 m higher than the lake level. The vegetation of this site is dominated by *Sphagnum* spp., *Betula* spp., *Andromeda polifolia*, *Picea mariana*, and *Ledum groenlandicum*.

The Moose Pass and Bicentennial Park sites lie to the east of the Kenai lowlands and at a higher elevation into the Kenai and Chugach Mountains. **Moose Pass** peatland (60°30'N, 149° 26'W) is located in the steep glacial valley of Moose Creek, towards the western edge of the Kenai Mountains. The mire is relatively small (approximately 100 m across) and is surrounded by *Picea* forest. The vegetation of the site includes *Sphagnum* spp, scattered *Picea*, *Ledum groenlandicum* and *Empetrum nigrum*. The site appears more ombrotrophic than others in this study.

The sampling area lies towards the centre of the site. **Bicentennial Park** peatland (61°06'N, 149° 44'W) is located in the Bicentennial Park of Anchorage, in the foothills of the Chugach Mountains. It is composed of two open bodies of water surrounded by a floating peat mat, gradually becoming firmer towards the periphery. At the contact between the surrounding forest and the peatland is a lagg, with extensive *Sphagnum* cover. The vegetation includes *Sphagnum* spp., *Ledum groenlandicum*, *Chamaedaphne calyculata*, *Myrica gale*, and *Vaccinium* spp.

The Houston and Sheep Creek sites are the most northerly of those in this study. These peatlands lie in the extensive floodplain lowlands of the Susitna River, which covers several thousand Km² of land adjacent to the Susitna and Yentna rivers to the north of Cook Inlet. **Houston** peatland (61°38'N, 149°51'W) forms part of a large wetland complex covering more than 20 Km². The sampled area occupies two adjoining treeless clearings within a region of scrubby *Betula* woodland approximately 300 m from the eastern edge of the site. The vegetation of the sampling area is dominated by *Sphagnum* spp, other species including *Ledum groenlandicum*, *Andromeda polifolia* and sedges occur on hummocks. **Sheep Creek** (62°00'N, 150°03'W) is the most northerly of the sites studied here, the peatland lies approximately 3 Km from the eastern bank of the Susitna River, approximately 2 Km north of its confluence with Sheep Creek. The mire is extensive and forested in areas. The sampling site occupies an area approximately 200 m west of the George Parks Highway. The vegetation of this area consists of scattered *Picea* and *Betula* with *Sphagnum*-dominated lawns and hollows, with sedges, *Ledum groenlandicum* and *Empetrum nigrum* on hummocks.

Sampling and laboratory procedures

Initial fieldwork was conducted in April 2003. The study sites were selected, and surface samples were collected for analysis in the laboratory. To provide some measure of heterogeneity, at least three locations from each microform type on the mire surface (hummock, hollow, lawn etc) were sampled; between nine and thirty samples were taken from each site. Surface samples approximately 5x5x10cm were removed and placed in plastic bags sealed for return to the laboratory. At each sampling site the surface vegetation was described and a GPS reading taken. To provide an estimate of water table depth variation over the growing season, PVC rods were inserted in the ground at each sampling point, following the method of Belyea (1999) and Bragazza (1996). Subsequent fieldwork was conducted in September 2003 when additional peat samples were taken and the depth to water table (DWT) measured by digging a small hole and measuring the depth from the surface after at least 30 minutes. The trial with PVC rods was unsuccessful because of disturbance by animals and, at some sites, insufficient length of rods. Peat pH was measured on the samples taken at the second sampling. A 20 ml volume of moss sample was placed in a 100 ml beaker, the beakers were filled with distilled water to the 100 ml level and left for one hour. The pH was measured with a Thermo Orion Portable pH/ISE Meter, model 250 A plus.

Testate amoebae were extracted from the surface samples removed in the first fieldwork. In all samples, the dominant moss species was used in the preparation. The upper green part of each moss was removed and the section from around 3 to 5cm depth used in the preparation, as this is believed to be most representative of the death assemblage. Testate amoebae sample preparation is

based on the method of Hendon & Charman (1997). Moss samples were cut into fine pieces and boiled in distilled water for 10 minutes to release the amoebae. The boiled samples were then filtered at 300µm and back-filtered through a 20µm mesh. The fraction remaining on the 20µm filter was stored in 5ml vials with glycerol. Testate amoebae were identified and counted with the use of light and phase-contrast microscopy using several identification guides (Charman *et al.*, 2000; Deflandre, 1929, 1936; Grospietsch, 1958; Lüftenegger *et al.*, 1988; Ogden, 1983; Ogden & Hedley, 1980). We aimed to reach a minimum count of 150 individuals. Along with testate amoebae, the rotifer *Habrotrocha angusticollis* was also counted and tallied for numerical analysis. The data used for statistical analyses are based on percentages rather than absolute abundances to permit easier comparison with the palaeoecological record.

Although in general the identification of testate amoebae is relatively easy based upon shell characteristics, the criteria for differentiating some species are unclear. As three of the authors were responsible for the analysis of samples from different sites we adopted a conservative taxonomic approach. We use the following grouping of species: *Amphitrema wrightianum* s.l. includes *A. wrightianum* and *A. stenostoma*. *Centropyxis aerophila* s.l. includes *Centropyxis aerophila*, *Centropyxis aerophila* var. *sphagnicola*, and *Centropyxis cassis*. *Euglypha rotunda* s.l. includes *Euglypha laevis* and *Euglypha rotunda*. *Nebela tincta* s.l. includes *Nebela minor*, *Nebela tincta*, and *Nebela tincta* var. *major*. *Phryganella acropodia* s.l. includes *Phryganella acropodia* and *Cyclopyxis arcelloides*. *Placocista spinosa* s.l. includes *Placocista spinosa* and *Placocista spinosa hyalina*. *Trinema* sp. includes *Trinema complanatum*, *Trinema enchelys*, and *Trinema lineare*. After this selection 36 species plus the rotifer *Habrotrochoa angusticollis* were left in the data set.

Numerical analyses

For the species-environment correlations, 27 species with less than five occurrences were omitted prior to percentage calculation. Samples with either less than 100 individuals counted or with missing pH or DWT data were also omitted from the data set. Two outliers, BB27 and BB 32, were omitted due to very high percentage for *Arcella discoides* and *Centropyxis laevigata* respectively. With this selection, 101 of the initial 121 samples remained in the data set.

In order to describe the general structure of the testate amoebae data, we performed a DCA on the percentage species data transformed using the natural logarithm [$x' = \ln(x + 1)$]. Triplots with representation of the first three axes were produced using the R-package (Casgrain & Legendre, 2004). Following the general description of the structure of the data, mantel tests and redundancy analyses (RDA) were used to explore the relationships between testate amoebae and their environment:

We used Mantel tests to test the general relationship between the testate amoebae community data and 1) the measured environmental variables (pH and DWT) and 2) the sites. For the testate amoebae data, a similarity matrix was computed with the Steinhaus asymmetrical index. For the environmental variables, a Euclidian distance was calculated on normalized data, and then a similarity matrix was calculated from the distance matrix ($S=1-D$). For the binary site variables a Euclidian distance was calculated, and then a similarity matrix was calculated from the distance matrix ($S=1-$

D). Mantel tests were then performed between pairs of similarity matrices, these analyses were carried out using the R-package (Casgrain & Legendre, 2004).

In order to analyse and quantify the relationships between the species and their environment, we performed several redundancy analyses (RDAs). The species data were transformed prior to the analysis with the Hellinger distance (Rao, 1995). This transformation allows the use of Euclidian-based methods such as RDA rather than Chi-square distance-based methods such as CCA (Legendre & Gallagher, 2001). The variation in the species data was partitioned to determine the fraction of variance explained by a) the sites, b) both the sites and the environmental variables, c) only the environmental variables, and d) none of these (Borcard *et al.*, 1992). To do this we performed two RDAs with either the sites binary variables or the pH and DWT variables. We then did a partial RDA in which pH and DWT were used as covariables and the binary site variables were used as explanatory variables. This analysis allowed us to determine the fraction of the variance that was explained by both sets of explanatory variables. These analyses were performed using the program Canoco (Ter Braak, 1988-1992).

The last step in the numerical analyses was the species-environment modelling. As the RDA showed that there are strong relationships between both pH and DWT and testate amoebae community composition (details are presented in the results), there is therefore a valid basis to relate these properties using species-environment models (transfer functions). As an initial DCA showed the gradients to be long, it was most appropriate to use models based on a unimodal distribution. Linear methods such as Partial Least Squares (PLS) are avoided as these have been shown to be

problematic when used for palaeo-environmental reconstruction with testate amoebae data (Wilmshurst *et al.*, 2003). Three unimodal models are tested here: 1) weighted averaging (WA), 2) weighted average partial least squares (WAPLS), and 3) maximum likelihood (ML).

Weighted averaging (WA) is the simplest of the three models. This technique works on the assumption that a species will be most abundant at its optimum for an environmental variable and therefore it is possible to estimate a species' optimum by taking an average of the values for an environmental variable at each site the species occurs and weighting this average by the species abundance at each of those sites (Birks, 1995). Weighted averaging has been the preferred or only model used in the majority of testate amoebae transfer function studies to date (Charman and Warner, 1992, 1997; Mitchell *et al.*, 1999; Warner and Charman, 1994; Wilmshurst *et al.*, 2003; Woodland *et al.*, 1998). Weighted averaging has the advantages of being both conceptually and computationally simple. Modifications of WA were tested with tolerance down-weighting (WA-Tol) and with both inverse and classical deshrinking. Weighted average partial least squares (WA-PLS) is a modified form of WA, which is essentially a unimodal equivalent of PLS. It is an improvement over WA in that it takes account of residual correlations in the biological data (ter Braak and Juggins, 1993; Birks, 1995). This method has been used in some studies of testate amoebae ecology and found to outperform simple WA (Booth, 2001, 2002; Bobrov *et al.*, 1999). The number of components required to produce optimal performance varies among data sets. Maximum likelihood (ML) is the most statistically rigorous technique used in this study. This method fits a parabolic response curve to the data for each species and uses this to estimate species

optima; this approach is distinct from WA, which assumes that this can be approximated by using a weighted average. No previous testate amoebae studies have investigated the contribution that ML models may make.

To assess the relative performance of these three models, two measures are used. The root mean square error of prediction (RMSEP), which assesses the random differences between observed and predicted values, and the maximum bias, which assesses the maximum error in any section of the environmental gradient. Errors are routinely under-estimated if the training set used for prediction includes those samples that are used to assess the errors in this prediction. Therefore a cross-validation method needs to be used, in this study both jack-knifing and bootstrapping using 1000 cycles were employed. RMSEP and Maximum bias estimated by both of these methods are considered (denoted $RMSEP_{jack}$, $RMSEP_{boot}$, $Max\ Bias_{jack}$ and $Max\ Bias_{boot}$). For the data-filtering exercise $RMSEP_{jack}$ has been used as the primary criterion to allow easier comparison with the results of previous studies. All species-environment modelling was carried out using the program C2 version 1.3 (Juggins, 2003).

Comment [EM1]: Instead of this long section, you could just state that we used two measures, JK and BS with 1000 cycles and give a reference for the details.

Previous studies have improved the performance of their transfer functions by selectively removing species and samples; here we investigate the effect of this data filtering and use these methods to optimise model performance. One common data treatment is the exclusion of species that occur in only a small number of samples. The underlying assumption is that the model will be inadequately able to characterize the optima and tolerances of species that only occur a few times, and that overall model performance may be improved by eliminating them. Although this has been

regularly applied in testate amoebae studies, the precise cut-off point has varied considerably among authors, from species with a single occurrence through species with as many as three (Booth, 2001), four (Booth, 2002), and even six occurrences (Charman and Warner, 1997). In this study we investigate the effect that increasing this cut-off point makes to model performance and the number of species remaining in the data set.

Another commonly used method to improve model performance is to remove those samples that have a high residual, possibly due to unusual testate amoebae communities or inaccurate environmental measurements. Woodland *et al.* (1998) and Wilmshurst *et al.* (2003) have used this strategy in testate amoebae studies. In these studies a single cut-off point was assigned, and removing samples with high residuals was shown to improve model performance. Here we assign a series of cut-off points to see how increasingly stringent filtering affects model performance relative to the number of samples included. Initially this was achieved simply by applying a series of cut-off points to residuals produced with the entire data set. However, it was found that model performance could be further improved by using these same cut-off points but applying them to residuals produced for each successively filtered data set. This iterative methodology produces a small but meaningful improvement in model performance for both pH and DWT reflecting the impact that removing each sample has upon the residuals of the others. Another possible method, which may be used to improve model performance, is to remove those species that have the broadest tolerances and that are therefore less useful as bioindicators. In this study this was undertaken using species standard errors

produced through weighted averaging, by setting a series of cut-off points and using a similar iterative method to the filtering for high residuals.

Results

Site characteristics

The depth to water table and peat pH of the sampling locations are shown in Figure 2. Depth to water table (DWT) was highly variable among sites and samples, ranging from 93cm to 5cm and averaging 35.7cm (σ 18.9cm). The driest site was Moose Pass (average 55.9cm), and the wettest was Kachemak (average 18.3cm). Peat pH ranged from 3.8 to 5.8 and averaged 4.6 (σ 0.4). The most acidic site was Clam Gulch (average 4.4), and the least acidic was Kachemak (average 5.5).

General structure of the testate amoebae data

A total of 62 species were found, and the average species richness per sample was 14.3 (σ 3.8). The five most abundant species in decreasing order of abundance were *Assulina muscorum*, *Amphitrema flavum*, *Hyalosphenia papilio*, *Phryganella acropodia*, and *Euglypha ciliata*, and these five species accounted for 51.4% of the total. The position of samples in the DCA (Figure 3) illustrates two main patterns. The first observation is that clear differences exist in testate amoebae communities among the sites. This is also attested by the significant, although low, correlation between the similarity matrices based on species data and site dummy variables

(Mantel test, $r=0.19$, $P<0.001$, 999 permutations). The second observation is that within most sites an important scattering of sampling points can be seen, thus confirming that we succeeded in sampling a wide range of environmental conditions at each site. A corollary of this observation is the very high similarity of several groups of samples that were supposed to represent similar micro-environmental conditions and were sampled in triplicates. Several such cases can be seen, for example for Bicentennial Bog groups of three samples, e.g. 1-2-3, 4-5-6, etc. represent the same microtopography. In such cases our data confirm that the structure of testate amoebae communities is predictable on the basis of macroscopic characteristics of the site (vegetation and microtopography). The position of the species in the ordination (Figure 3) suggests that the first axis is primarily explained by the moisture gradient. Species generally associated with wet habitats, such as *Nebela marginata*, *N. griseola*, and *Amphitrema wrightianum* have low scores on the first axis, while species generally associated with dry conditions, such as *Corythion dubium*, *Assulina muscorum*, *Trinema lineare*, *Euglypha rotunda*, or *Centropyxis aerophila* have high scores on the first axis.

Species-environment correlations and variance partitioning

In the Mantel tests all pairs of similarity matrices were significantly correlated (species vs. pH & DWT, species vs. site binary variables, and pH & DWT vs. site binary variables; $P<0.001$; $r=0.18-0.19$; 999 permutations).

In the RDA (Figure 4), the site variables and the two quantitative variables (pH and DWT) respectively explained 31.2 and 12.9% of the variation in the species data.

Both sets of variables were significant (Monte-Carlo permutation test, 999 permutations, $P < 0.001$).

The partial RDA revealed that 7.1% of the variation was explained by both sets of variables. Therefore the fraction of the variance explained by the site variables alone was 24.1% (31.2-7.1%), while the fraction explained by the two quantitative variables (pH and DWT) alone was 5.8% (12.9-7.1%). Finally 63% of the variation was left unexplained.

Species-environment models

The performance of WA, WA-PLS and ML models was initially assessed using all data except for species occurring in only a single sample (Table 1). Several variants of the WA model were tested; the best performing was simple weighted averaging with inverse deshrinking. The overall best performing model in terms of $RMSEP_{jack}$ is a two-component WA-PLS model which produces values of 15.8cm for DWT and 0.3 for pH (Table 1; Figure 5). Errors assessed using bootstrapping are generally greater than those estimated using jack-knifing, consistent with the results of other studies. It is notable that while the ML model performs poorly in terms of $RMSEP$, it performs best for maximum bias.

Model performance may be improved by selective exclusion of the data; the first method we attempted was by removing those samples with only a small number of occurrences. Results are shown in Figure 6; removing species with low occurrences clearly reduces $RMSEP_{jack}$ particularly at higher levels where large numbers of

species are removed. Removing species occurring in as many as seven samples reduces the total number of species by around a third. However, this decrease in $RMSEP_{jack}$ is not uniform, so for instance removing species with three occurrences from the DWT data produces a greater $RMSEP_{jack}$ than removing species with only two occurrences. To decide what degree of filtering is to be used a balance needs to be reached between model performance and the number of species remaining and therefore the applicability of the model to palaeo-testate communities, this is an essentially subjective judgement. In this study it was decided that the optimum filtering would be to remove species with two or less occurrences from the DWT data set and a single occurrence from the pH data set. These filtered data sets were used for the subsequent stage of data filtering.

The second method of data-filtering tested was removing samples with high residuals. This significantly improves model performance (Figure 6). For DWT, removing samples with a residual greater than 10 cm reduces the $RMSEP_{jack}$ to 7.41 cm, a decrease of over 50%. The samples removed are generally located at the ends of the environmental gradients, particularly from the very dry sites for DWT. Therefore removing these samples from the data set restricts the environmental gradient for which the transfer function has predictive power when applied to palaeoecological data, and a balance needs to be achieved between this and model performance. It was decided that an optimal compromise would be to use the DWT data set with residuals greater than 20 cm excluded and the pH data set with residuals over 0.4 excluded, these data sets retain 91 and 103 samples respectively. This cut-off point is approximately 20% of the environmental gradient for both variables.

The final data-filtering method tested was removing those species with highest standard errors; results of these trials are shown in Figure 6. Moderate filtering does reduce $RMSEP_{jack}$ for DWT, although it makes little difference for pH, but at higher levels of filtering $RMSEP_{jack}$ is significantly increased. Species initially removed are those with few observations such as *C. platystoma* and *T. dentata*. Under more rigorous filtration, more common species such as *T. arcula* and *C. oviformis* are also removed. It was decided not to use this filtering for either DWT or pH as even at low filtering, where RMSEP may be decreased slightly, Max Bias is increased.

Discussion

This study fills an important geographical gap in the knowledge of the autoecology of peatland testate amoebae and its potential geographical variability. Both of these aspects are critical to the use of these organisms as proxy indicators in paleoecological studies. Similar to previous studies, we found that testate amoebae communities were significantly correlated with water table and pH. The species coefficients for Alaskan peatlands show reasonable similarity with the results of other quantitative and qualitative studies of testate amoebae ecology in peatlands of other regions (Warner and Charman, 1994; Charman 1997; Charman and Warner 1997; Woodland et al. 1998; Bobrov et al. 1999; Booth, 2001). At the wettest end of the DWT gradient are species such as *Heleopera sphagni* and *Placocista spinosa* and at the driest are species such as *Trigonopyxis arcula*, *Nebela militaris*, and *Bulinularia indica*. These results are in keeping with the recognized moisture preferences of these species.

Direct comparison among studies is difficult because most studies have used one-time DWT measurement, with the notable exception of Woodland et al. (1998). To date, only one study has compared two sites using exactly the same methods (Booth and Zygmunt, 2005). But in this study too, the DWT measurements were not done on the same day, and even if they had, given the distance among sites (Rocky Mountains and Great Lakes regions respectively) the sampling date would be less important than the time since the last significant rain event. Given the above-mentioned methodological differences in water table depth measurements, it is difficult to compare species optima with previous results, as most studies have only measured DWT on a single sampling occasion and therefore do not take account of annual variations in water table. Booth and Zygmunt (2005) compared the DWT optima for testate amoebae between two distant study sites in North America. They found that although the communities were sometimes quite different, the species' optima were consistent between the two regions. For comparisons among three sites or more, one approach to this is to compare the relative position of species optima on the overall DWT gradient for each study (Booth, 2001; Lamentowicz and Mitchell, 2005). The plot in Figure 7 compares our results with those of six previous studies, yielding good overall agreement. The optima for any given species in our study are generally towards the middle of the range of optima for that species across the seven studies. Thus overall the ecological patterns of testate amoebae in South-Central Alaska generally compare well with observations from previous studies in other parts of the world. This provides further support for their use as bioindicators over broad geographical ranges.

Future work should aim at finding a reliable methodology for obtaining information on the water table depth (or other related micro-environmental variable such as the oxic / anoxic boundary, range of water table fluctuation, etc.). This would make direct comparison among studies more relevant, and would also allow an assessment of how other factors such as the general climate influence community structure. It is likely for example that a given species may be restricted to wet micro-habitats in regions where the frequency of summer precipitation is low (continental climate) but could survive in drier micro-habitats (lower DWT) under a more oceanic climate. The PVC discoloration method might hold such promises, while being very cost-effective. Unfortunately we were unsuccessful in using it in this study. It has recently been shown that even within a given brand and type of PVC tape, some rolls fail to change colour for unknown reasons; further tests are therefore required (Booth et al. 2005).

The proportion of total variance explained by depth to water table and pH was lower in our data than in comparable studies. In this study 5.8% of the variation was explained by the measured environmental variables, other studies have found values of 9.1% (Charman, 1997), 9.7 % (Booth, 2001), and 11.7% (Booth, 2002). The proportion of variance explained by the sites (24.1%) was much higher than that explained by pH and DWT (5.8%). At first this would suggest that the sites differed significantly with respect to some important variable to which testate amoebae are sensitive. Indeed 7.1% of the variance is explained by both sets of variables. However, some non-measured variables differ among sites and account for about one quarter of the variation in the species data. Given the relatively large area covered by this study and the important climatic differences among the sites, ranging from sub-oceanic to sub-continental, it is likely that inter-site differences in the

extremes or the range of temperature, moisture, or other variables may have a strong influence on testate amoebae community structure. Recording such patterns, however, was beyond the scope of this study, and no meteorological data were available.

Model performances have been substantially improved by selective filtering of the data, but this improvement is not uniform. To understand the data structure and the impacts of data removal a step-wise approach is preferable to a single cut-off point as used in previous studies. The methodology used in this study may thus be useful for improving model performance in other studies. After the selected data-filtering the $RMSEP_{jack}$ for the transfer function is reduced to 9.73 cm for DWT (Table 1). This is a considerable improvement over the full data-set, but still compares poorly with the results of other studies which have produced values for $RMSEP_{jack}$ as low as 3.93cm (Woodland *et al.*, 1998), 6.3cm (Charman and Warner, 1997), and 7.5cm (Booth, 2001, 2002; Wilmshurst *et al.*, 2003). When applied to palaeoecological data this standard of model performance will allow differentiation of wet from dry phases of peat growth but might not detect subtler surface-moisture changes.

A possible cause of this poor performance is a taxonomic bias due to the fact that the analyses were divided among three individuals (RP, KK, and EM). To minimize this possible bias we used a conservative taxonomic approach, similar to the one proposed by Charman *et al.* (Charman *et al.*, 2000). Significant differences were nevertheless found among the three sets of data that were pooled in this study, but these differences also certainly reflect true differences among the sites. For example, in Bicentennial Bog a wide range of microhabitats was sampled, but overall this is

where many of the wetter microhabitats occurred. Jigsaw Lake was the only site adjacent to a true lake, as opposed to small pools in the center of peatlands. The sites also differ in terms of vegetation, which may also affect testate amoebae communities.

In common with most previous transfer function studies, we used one-off water table measurements. These measurements may be unrepresentative of the actual conditions, which have influenced the testate amoebae community over the period of several years represented by the examined tests. It is notable that the transfer function study with the lowest RMSEP (Woodland *et al.* 1998) was also the only one to use long-term water table monitoring data. In addition, our DWT measurements were taken at the end of the summer, when the peatlands are probably at their driest. It is possible that measurements at this time of year may not be truly representative, particularly as many species could be inactive in such dry conditions. It is therefore possible that the relatively high RMSEP_{jack} of the transfer function could be due, at least in part to the hydrological state of the sampling points being poorly characterised.

Perhaps a more ecologically meaningful explanation for the relatively high RMSEP_{jack} of the models is the unusually dry conditions that characterize many peatlands in South-Central Alaska. In most previous studies the maximum depth to water table recorded ranges between 30 and 50cm (Booth, 2001, 2002; Bobrov *et al.*, 1999; Charman and Warner, 1997; Wilmshurst *et al.*, 2003). In this study DWT values as high as 93cm were encountered. Although these measurements were taken in late summer when higher DWT values would be expected, it seems clear that many of

the sites in this study are unusually dry. At the highest DWT values model fit is poorest and it is these samples that are preferentially removed when filtering for samples with high residuals. Poor model fit at the driest sites has been observed in several previous studies (Booth, 2002; Lamentowicz and Mitchell, 2005; Mitchell *et al.*, 1999). It is therefore possible that the presence of some extremely dry sites may have impacted model performance. A plot of DWT gradient against model performance for several studies shows a generally good relationship (Figure 8). When assessed by this criterion, model performance in this study seems more reasonable. Figure 8 also suggests a good relationship between the restriction in environmental gradient and improvement in $RMSEP_{jack}$ with data filtering; the improvement in model performance may therefore have been primarily achieved at the expense of the environmental gradient considered. A balance needs to be reached between model performance and model applicability; this decision remains an essentially subjective one.

A further possible cause of poor model performance is that current and recent climatic change may be causing the study sites to change. Indeed, at the sampling time, many of the peatlands on the Kenai lowlands were very dry with many plants appearing drought-affected. Much research has illustrated recent climatic change in the Arctic and sub-Arctic regions (Moritz *et al.*, 2002; Sturm *et al.*, 2001). In south-central Alaska, Klein *et al.* (submitted) suggested a variety of evidence for wetland drying caused by climatic change over the last 50 years. If the climate of the region, and therefore the hydrology of the peatlands, is in a state of transition then it is possible that the testate amoebae community may not have fully adjusted to the change, especially since the sampled depth (3-5cm) integrates conditions over

several years. This might depend on the location, hydrology and climate of the individual peatlands and could have decreased the model fit.

This study provides the first testate amoebae transfer function for southern Alaska, and indeed the entire North Pacific region. Results demonstrate that DWT and pH are important controls on testate amoebae communities and that the ecological preferences of testate amoebae species are broadly similar to those of other regions of the world. Overall model performance is somewhat below that of previous studies, this may be due to a combination of recent climatic change, the large environmental ranges encountered in this study, and the limitations related to the one-time DWT measurement. This work will allow quantified reconstruction of pH and DWT changes in palaeoecological studies underway from peatlands in the region.

Acknowledgements

RJP was supported by a EU Marie Curie doctoral fellowship and a QMUL Westfield studentship. This study was partly funded by two grants to EADM from the University of Alaska Anchorage. Fieldwork funding to RJP and JJB was provided by the Royal Society, the Bill Bishop Memorial Trust, the Knowle Hill School Fund and the University of London Central Research Fund. Becky Morris generously assisted with the fieldwork. Thanks to John Birks and Richard Telford for helpful discussions. Ed Oliver helped with preparation of some of the figures. This study was initiated while EADM was at the University of Alaska Anchorage and finished while he was at EPFL&WSL and supported by EU project RECIPE, which is partly funded by the European Commission (n° EVK2-2002-00269) and partly, for the Swiss partners

EPFL and WSL-AR, by the OFES (Swiss Federal Office for Education and Science),
Switzerland. A constructive review by Bob Booth helped improve the manuscript.

Table 1. Comparative model performance for full data-set assessed using RMSEP and Maximum Bias values produced by both boot-strapping and jack-knifing for DWT (cm) and pH.

| Variable | Model | RMSE | RMSE | Max | Max |
|----------|------------|-------------------|-------------------|-----------------------------|-----------------------------|
| | | P_{jack} | P_{boot} | $\text{Bias}_{\text{jack}}$ | $\text{Bias}_{\text{boot}}$ |
| DWT | WA | 15.9 | 16.4 | 44.5 | 44.7 |
| DWT | WA- PLS | 15.8 | 16.6 | 43.3 | 43.2 |
| DWT | ML | 24.5 | 21.8 | 26.7 | 31.2 |
| pH | WA | 0.30 | 0.31 | 0.60 | 0.65 |
| pH | WA- PLS | 0.30 | 0.33 | 0.56 | 0.66 |
| pH | ML | 0.40 | 0.40 | 0.49 | 0.62 |

Table 2. Model performance pre- and post-filtering using optimal 2-component WA-PLS model for DWT(cm) and pH.

| Model | DWT | DWT | pH | pH |
|--------------------------|-------------|-----------------|-------------|-----------------|
| | (full data) | (filtered data) | (full data) | (filtered data) |
| RMSEP _{jack} | 15.8 | 9.7 | 0.30 | 0.21 |
| RMSEP _{boot} | 16.6 | 10.2 | 0.33 | 0.22 |
| Max Bias _{jack} | 43.3 | 14.0 | 0.56 | 0.29 |
| Max Bias _{boot} | 43.2 | 14.2 | 0.66 | 0.20 |

Figure 3 DCA of 101 testate amoebae samples with representation of samples (a) and species (b). The species data was log-transformed using $\ln(x+1)$. Axes 1 and 2 represent 16.2 and 10.9% of the variation in the species data respectively. Eight site binary variables were projected passively in the ordination space. Species names are in italics and passively projected variables are in bold. The size of the symbols is proportional to the scores on the third axis. Negative values are represented by filled symbols and positive values by open symbols.

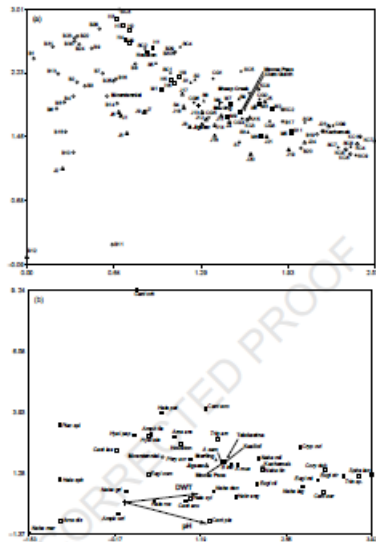


Figure 4 RDA of 101 testate amoebae samples. Species data were log-transformed using $\ln(x+1)$. Together pH and DWT explain 12.9% of the variation in the species data (both significant, $P < 0.001$, Monte-Carlo test, 999 permutations). For clarity, species scores were multiplied by 4.5 and environmental variables scores by 3.

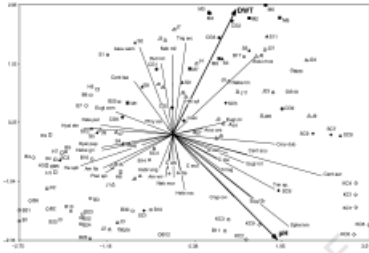
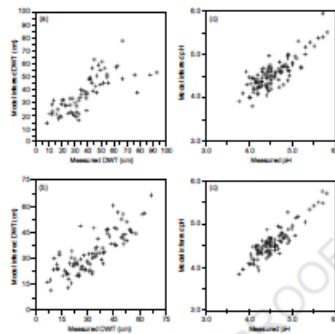
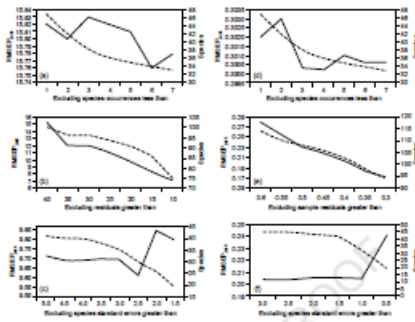


Figure 5 WA-PLS observed versus predicted values for DWT (a&b) and pH



(c&d), pre- and post-filtering.

Figure 6 Data filtering plots. Filtering for species counts, residuals and standard error for pH (a-c) and DWT (d-f) showing change in $RMSEP_{jack}$ (solid line) and consequent change in number of samples or species (dashed line).



line).

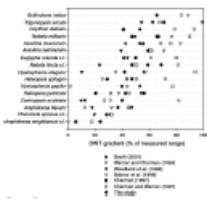


Figure 7 Relative position of species optima on the DWT gradient for major species in this study and selected previous studies. Results for this study are produced by WA using the filtered data set, results for other studies are as reported.

Figure 8 DWT gradient length versus $RMSEP_{jack}$ from previously published studies and this study pre- and post-filtering. DWT range figures do not include DWT values recorded as negative. Where the studies used data-filtering to improve model performance, the filtered $RMSEP_{jack}$ values are included here.

References

Aaby, B. 1976: Cyclic variations in climate over the past 5500 years reflected in raised bogs. *Nature* 263, 281-284.

Birks, H.J.B. 1995: Quantitative palaeoenvironmental reconstructions, In Maddy, D. and Brew, J.S., editors, *Statistical modelling of quaternary science data*, Technical Guide 5, Cambridge: Quaternary Research Association.

Belyea, L.R. 1999: A novel indicator of reducing conditions and water-table depth in mires. *Functional Ecology* 13, 431-434.

Blundell, A., and Barber, K. 2005: A 2800-year palaeoclimatic record from Tore Hill Moss, Strathspey, Scotland: the need for a multi-proxy approach to peat-based climate reconstructions. *Quaternary Science Reviews* 24, 1261-1277.

Bobrov, A., Charman, D.J. and Warner, B.G. 1999: Ecology of testate amoebae (Protozoa: Rhizopoda) on peatlands in Western Russia with special attention to niche separation in closely related taxa. *Protist* 150, 125-136.

Borcard, D., Legendre, P., and Drapeau, P. 1992: Partialling out the spatial component of ecological variation. *Ecology* 73, 1045-1055.

Booth, R.K. 2001: Ecology of testate amoebae (protozoa) in two Lake superior coastal wetlands: implications for palaeoecology and environmental monitoring. *Wetlands* 21, 564-576.

Booth, R.K. 2002: Testate amoebae as paleoindicators of surface-moisture changes on Michigan peatlands: modern ecology and hydrological calibration. *Journal of Paleolimnology* 28, 129-348.

Booth, R.K. and Jackson, S.T. 2003: A high-resolution record of late Holocene moisture variability from a Michigan raised bog. *The Holocene* 13, 865-878.

Booth, R.K. and Zygmunt, J.R. 2005: Biogeography and comparative ecology of testate amoebae inhabiting *Sphagnum*-dominated peatlands in the Great Lakes and Rocky Mountain regions of North America. *Diversity and Distributions* 11, in press

Booth, R.K., Hotchkiss, S.C., and Wilcox, D.A. 2005: Discoloration of polyvinyl chloride (PVC) tape as a proxy for water-table depth in peatlands: validation and assessment of seasonal variability. *Functional Ecology*, in press

Bragazza, L. 1996: Delimitation of the aerobic peat layer in a *Sphagnum* mire on the southern Alps. *Oecologia Montana* 5, 41-46.

Casgrain, P. and Legendre, P. 2004: *The R Package for Multivariate and Spatial Analysis*. Département de sciences biologiques, Université de Montréal, Montréal, Canada.

Charman, D.J. and Warner, B.G. 1997: The ecology of testate amoebae (Protozoa: Rhizopoda) and microenvironmental parameters in Newfoundland, Canada: modeling hydrological relationships for palaeoenvironmental reconstruction. *Ecoscience* 4, 555-562.

Charman, D.J. and Warner, B.G. 1992: Relationship between testate amoebae (Protozoa: Rhizopoda) and microenvironmental parameters on a forested peatland in northeastern Ontario. *Canadian Journal of Zoology* 7, 2474-2482.

Charman, D.J. and Hendon, D. 2000: Long-term changes in soil water tables over the past 4500 years: relationships with climate and North Atlantic atmospheric circulation and sea surface temperatures. *Climatic Change* 47, 45-59.

Charman, D. J., Hendon, D. and Woodland, W. A. 2000: *The identification of testate amoebae (Protozoa: Rhizopoda) in peats*, Cambridge: Quaternary Research Association.

Charman, D.J., Caseldine, C., Baker, A., Gearey, B., Hatton, J. and Proctor, C. 2001: Paleohydrological records from peat profiles and speleothems in Sutherland, Northwest Scotland. *Quaternary Research* 55, 223-234.

Charman, D.J., Brown, A.D., Hendon, D. and Karofeld E. 2004: Testing the relationship between Holocene peatland palaeoclimate reconstructions and instrumental data at two European sites. *Quaternary Science Reviews* 23, 137-143.

Clarke, K. J. 2003: *Guide to the Identification of Soil Protozoa-Testate Amoebae*,
Freshwater Biological Association. Kendal: Titus Wilson & Son.

Corbett, S.A. 1973. An illustrated introduction to the Testate Rhizopods in
Sphagnum, with special reference to the area around Malham Tarn, Yorkshire. *Field
Studies* 3, 801-838.

Deflandre, G. 1929: Le genre *Centropyxis* Stein. *Archiv für Protistenkunde* 67, 322-
375.

Deflandre, G. 1936: Etude monographique sur le genre *Nebela* Leidy. *Annales de
Protistologie* 5, 201-286.

Grospietsch, T. 1958: *Wechseltierchen (Rhizopoden)*, Stuttgart: Kosmos Verlag.

Hendon, D. and Charman, D.J. 1997: The preparation of testate amoebae
(Protozoa:Rhizopoda) samples from peat. *The Holocene* 7, 199-205.

Juggins, S. 2003: *C2 user guide. Software for ecological and palaeoecological data
analysis and visualisation*, Newcastle Upon Tyne: University of Newcastle.

Klein, E., Berg, E. and Dial, R. Submitted: Wetland drying and succession across the
Kenai Peninsula Lowlands, Southcentral Alaska. *Canadian Journal of Forest
Research*.

Lamentowicz, M. and Mitchell, E.A.D. 2005: The ecology of testate amoebae (protists) in *Sphagnum* in north-western Poland in relation to peatland ecology. *Microbial Ecology* 50, 48-63.

Legendre, P. and Gallagher, E.D. 2001: Ecologically meaningful transformations for ordination of species data. *Oecologia* 129, 271-280.

Lüftenegger, G., Petz, W., Berger, H., Foissner, W., and Adam, H. 1988: Morphologic and Biometric Characterization of 24 Soil Testate Amebas (Protozoa, Rhizopoda). *Archiv für Protistenkunde* 136, 153-189.

Meisterfeld, R. 2002a: Order Arcellinida Kent, 1880. In Lee, J.J., Leedale, G.F. and Bradley, P. editors, *The illustrated guide to the protozoa Vol. 2*, Lawrence, Kansas: Society of Protozoologists, 827-860.

Meisterfeld, R. 2002b: Testate amoebae with filopodia. In Lee, J.J., Leedale, G.F. and Bradley, P. editors, *The illustrated guide to the protozoa Vol. 2*, Lawrence, Kansas: Society of Protozoologists, 1054-1084.

Mitchell, E.A.D., Warner, B.G., Buttler, A. and Gobat, J-M. 1999: Ecology of testate amoebae (Protozoa:Rhizopodia) in *Sphagnum* peatlands in the Jura mountains, Switzerland and France. *Ecoscience* 6, 565-576.

Moritz, R.E., Bitz, C.M., and Steig, E.J. 2002: Dynamics of Recent Climate Change in the Arctic. *Science* 297, 1497-1502.

Nikolaev, S.I., Mitchell, E.A.D., Petrov, N.B., Berney, C., Fahrni, J. and Pawlowski, J. 2005: The testate lobose amoebae (order Arcellinida Kent, 1880) finally find their home within Amoebozoa. *Protist* 156, (in press)

Ogden, C.G. 1983: Observations on the systematics of the genus *Diffugia* in Britain (Rhizopoda, Protozoa). *Bulletin of the Natural History Museum. Zoology series* 44, 1-73.

Ogden, C.G. and Hedley, R.H. 1980: *An atlas to freshwater testate amoebae*, Oxford: Oxford University Press.

Rao, C.R. 1995: A review of canonical coordinates and an alternative to correspondence analysis using Hellinger distance. *Qüestió* 19, 23-63.

Schoning, K., Charman, D.J and Wastegård, S. 2005: Reconstructed water tables from two ombrotrophic mires in eastern central Sweden compared with instrumental meteorological data. *The Holocene* 15, 111-118.

Sturm, M., Racine, C., and Tape, K. 2001: Climate change: Increasing shrub abundance in the Arctic. *Nature* 411, 546 - 547.

Ter Braak, C.J.F. 1988-1992: *CANOCO - a FORTRAN program for Canonical community Ordination (version 2.1)*. Ithaca, New York: Microcomputer Power.

Ter Braak, C.J.F. and Juggins, S. 1993: Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia* 269, 485-502.

Tolonen, K. 1986: Rhizopod analysis. In Berglund, B. E., editor, *Handbook of Holocene Palaeoecology and Palaeohydrology*, Chichester: John Wiley, 645-666.

Tolonen, K., Warner, B.G. and Vasander, H. 1994: Ecology of testaceans (protozoa:rhizopoda) in mires in Southern Finland: II multivariate analysis. *Archiv fur Protistenkunde* 144, 97-112.

Warner, B.G. and Charman, D.J. 1994: Holocene changes on a peatland interpreted from testate amoebae (Protozoa) analysis. *Boreas* 23, 270-280.

Wilmshurst, J., Wiser, S. and Charman, D.J. 2003: Reconstructing Holocene water tables in New Zealand using testate amoebae: differential preservation of tests and implications for the use of transfer functions. *The Holocene* 13, 61-72.

Woodland, W.A., Charman, D.J., & Simms, P.C. 1998: Quantitative estimates of water tables and soil moisture in Holocene peatlands from testate amoebae. *The Holocene* 8, 261-273.