UNIVERSITY *of* York

This is a repository copy of *Towards a Holarctic synthesis of peatland testate amoeba ecology: development of a new continental-scale palaeohydrological transfer function for North America and comparison to European data.*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/137799/

Version: Accepted Version

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

1 **Towards a Holarctic synthesis of peatland testate amoeba ecology:**
2 **development of a new continental-scale palaeohydrological transfer function**
3 **for North America and comparison to European data**
4
5 Matthew J. Amesbury [a,b]*, Robert K. Booth [c], Thomas P. Roland [a], Joan Bunbury [d,e], Michael J.
6 Clifford [f], Dan J. Charman [a], Suzanne Elliot [g], Sarah Finkelstein [h], Michelle Garneau [i], Paul D. M.
7 Hughes [j], Alexandre Lamarre [k], Julie Loisel [l], Helen Mackay [m], Gabriel Magnan [i], Erin R. Markel
8 [c], Edward A. D. Mitchell [n,o], Richard J. Payne [p,q], Nicolas Pelletier [r], Helen Roe [g], Maura E.
9 Sullivan [c], Graeme T. Swindles [s], Julie Talbot [t], Simon van Bellen [i], Barry G. Warner [u]
10
11 * Corresponding author: Email: m.j.amesbury@exeter.ac.uk (M. J. Amesbury).
12
13 [a] Geography, College of Life and Environmental Sciences, University of Exeter, UK
14 [b] Environmental Change Research Unit (ECRU), Faculty of Biological and Environmental Sciences,
15 University of Helsinki, Finland
16 [c] Earth and Environmental Sciences, Lehigh University, USA
17 [d] Department of Geography and Earth Science, University of Wisconsin – La Crosse, USA
18 [e] Department of Geography, University of Toronto, Canada
19 [f] Desert Research Institute, Nevada, USA
20 [g] School of Natural and Built Environment, Queen's University Belfast, UK
21 [h] Department of Earth Sciences, University of Toronto, Canada
22 [i] Geotop, Université du Québec à Montréal, Canada
23 [j] Geography and Environment, University of Southampton, UK
24 [k] Environment and Climate Change Canada, Canada
25 [l] Department of Geography, Texas A&M University, USA
26 [m] School of Geography, Politics and Sociology, University of Newcastle, UK
27 [n] Institute of Biology, Faculty of Science, University of Neuchâtel, Switzerland
28 [o] Botanical Garden of Neuchâtel, Switzerland
29 [p] Environment, University of York, UK
30 [q] Department of Biology, Penza State University, Russia
31 [r] Department of Geography and Environmental Studies, Carleton University, Canada
32 [s] School of Geography, University of Leeds, UK
33 [t] Department of Geography, Université de Montréal, Canada
34 [u] Earth and Environmental Sciences, University of Waterloo, Canada
35
36 **Author contributions:** MJA and RKB conceived the work and compiled/archived the data. MJA
37 conducted data analysis and wrote the manuscript. All authors contributed data, actively discussed the
38 direction of the research and/or contributed to manuscript editing.
39
40 **Keywords:** North America, testate amoebae, peatland, water table, transfer function, ecology,
41 biogeography, cosmopolitanism.
42
43 **Highlights**
44 • Dataset of ca. 2000 North American peatland testate amoeba surface samples compiled
45 • Palaeohydrological transfer functions tested statistically, applied to independent palaeo data
46 • New model an effective tool for palaeohydrological reconstruction across North America
47 • Biogeographical patterns occur at taxonomic resolution applied; taxa hydrological optima
48   mostly robust across North America

49    • Combined North American and European dataset illustrates potential for Holarctic synthesis
50

51    **Abstract**
52    Fossil testate amoeba assemblages have been used to reconstruct peatland palaeohydrology for more
53    than two decades. While transfer function training sets are typically of local- to regional-scale in extent,
54    combining those data to cover broad ecohydrological gradients, from the regional- to continental- and
55    hemispheric-scales, is useful to assess if ecological optima of species vary geographically and therefore
56    may have also varied over time. Continental-scale transfer functions can also maximise modern
57    analogue quality without losing reconstructive skill, providing the opportunity to contextualise
58    understanding of purely statistical outputs with greater insight into the biogeography of organisms.
59    Here, we compiled, at moderate taxonomic resolution, a dataset of nearly 2000 modern surface
60    peatland testate amoeba samples from 137 peatlands throughout North America. We developed
61    transfer functions using four model types, tested them statistically and applied them to independent
62    palaeoenvironmental data. By subdividing the dataset into eco-regions, we examined biogeographical
63    patterns of hydrological optima and species distribution across North America. We combined our new
64    dataset with data from Europe to create a combined transfer function. The performance of our North-
65    American transfer function was equivalent to published models and reconstructions were comparable
66    to those developed using regional training sets. The new model can therefore be used as an effective
67    tool to reconstruct peatland palaeohydrology throughout the North American continent. Some eco-
68    regions exhibited lower taxonomic diversity and some key indicator taxa had restricted ranges.
69    However, these patterns occurred against a background of general cosmopolitanism, at the moderate
70    taxonomic resolution used. Likely biogeographical patterns at higher taxonomic resolution therefore do
71    not affect transfer function performance. Output from the combined North American and European
72    model suggested that any geographical limit of scale beyond which further compilation of peatland
73    testate amoeba data would not be valid has not yet been reached, therefore advocating the potential
74    for a Holarctic synthesis of peatland testate amoeba data. Extending data synthesis to the tropics and
75    the Southern Hemisphere would be more challenging due to higher regional endemism in those areas.
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97

98
99
**1. Introduction**

Testate amoebae are a polyphyletic group of microscopic, unicellular protists that occur globally in a range of soils, rivers, lakes and wetlands, including ombrotrophic peatlands – environments which are frequently used to reconstruct Holocene palaeohydrological change (Mitchell et al., 2008). These reconstructions are based on a transfer function approach, inherent in which is a robust understanding of peatland testate amoeba ecology, particularly as it relates to target reconstruction variables, principally water table depth (WTD).

Local- or regional-scale transfer function models around the globe have provided robust palaeohydrological reconstructions based on the strong empirical relationship between independent (i.e. WTD) and response (i.e. testate amoeba community change) variables (e.g. Swindles et al., 2009, 2014; van Bellen et al., 2014; Li et al., 2015). However, reconstruction magnitude can vary when local- and regional-scale models are applied outside of their training set range (Turner et al., 2013), providing a strong motivation to develop broader-scale, compiled models that provide standard reconstruction tools with directly comparable results.

Modern analogues from sites close to a core location may potentially better represent the contemporary environment than those from more distant sites in the present day. However, as we go back in time through different climate and ecological boundary conditions, as any transfer function reconstruction inevitably aims to do, local modern analogues may be less appropriate and it becomes more likely that better analogues will be found in different regions. Therefore, whilst local- and regional-scale models perform well statistically in cross-validation, the same performance cannot be assumed for downcore reconstructions and large-scale datasets/models should yield more reliable reconstructions for any given location given their greater range of modern analogues. However, as the geographical extent of training sets is increased to the continental-, hemispheric- or even global-scales, it may become more problematic to effectively model independent and response variable relationships due to the introduction of confounding factors such as latitude and temperature effects, secondary environmental gradients or non-linear processes such as permafrost and fire. Therefore, there may be a geographical range limit beyond which the compilation of modern testate amoeba and associated environmental data is no longer beneficial for model development. Recently, it has been shown that the development of continental-scale transfer functions may have advantages over the use of local- or regional-scale models (Amesbury et al., 2016). In particular, large training sets (i.e. > 1000 samples) capture a greater range of variability in target environmental variables and include more modern analogues for taxa that may be abundant in fossil records but that can be rare or absent in modern datasets (Charman et al., 2007; Swindles et al., 2015a), potentially improving palaeohydrological reconstructions.

In addition, although individual transfer functions statistically model the relationship between modern testate amoeba assemblages and WTD, a broader contextual understanding of peatland testate amoeba ecology and biogeography is important. For example, recent research has shown that certain taxa, or groups of taxa, may react to environmental change at different timescales (Sullivan and Booth, 2011; Marcisz et al., 2014) or have geographically restricted or skewed ranges (Booth and Zygmunt, 2005; Amesbury et al., 2016; Beyens and Bobrov, 2016), even extending to regional endemism, particularly in the tropics and Southern Hemisphere (Smith and Wilkinson, 2007; van Bellen et al., 2014; Reczuga et al., 2015). However, these ecological and biogeographical intricacies are not explicit in strictly statistical reconstructions. Indeed, research has shown that continental-scale geographical patterns of both genotypic and phenotypic diversity exist within seemingly cosmopolitan taxa (Heger et al., 2013; Roland

147  et al., 2017; Singer et al., 2018). In the case of *Hyalosphenia papilio* (Heger et al., 2013), this is a finding
148  potentially supported by a regionally skewed distribution of the same taxon across Europe (Amesbury
149  et al., 2016).

151  Greater insight into wide-scale biogeographical patterns of soil protists is a fundamental research
152  domain (Geisen et al., 2017). An increasing number of studies have used protists (Foissner, 2006, 2008),
153  and specifically testate amoebae (e.g. Smith and Wilkinson, 2007; Fournier et al., 2015; Fernández et
154  al., 2016; Lara et al., 2016), to argue against the 'everything is everywhere: but the environment selects'
155  paradigm (Baas Becking, 1934; see also De Wit and Bouvier, 2006; O'Malley, 2007, 2008). Research
156  communities can move towards a broader understanding of these patterns by working collaboratively
157  to make use of the ever-increasing amount of data from individual studies that can be compiled and
158  used to both develop new, robust reconstruction tools (Amesbury et al., 2016), as well as begin to
159  provide insight into important biogeographical debates (Smith et al., 2008). In addition, an
160  understanding of wide-scale peatland testate amoeba ecology and biogeography is important in order
161  to provide robust regional- to continental-scale palaeoclimatic interpretation of 1) palaeo-data
162  compilations (e.g. Swindles et al., 2013), 2) studies from individual sites or groups of sites aiming to
163  study the Holocene evolution of regional-scale climate systems (e.g. Magnan and Garneau, 2014; van
164  Bellen et al., 2016), 3) studies investigating anthropogenic drainage (e.g. Laggoun-Défarge et al., 2008;
165  Payne et al., 2010; Talbot et al., 2010) and 4) studies of ecohydrological changes from specific climatic
166  disturbances such as permafrost thaw (Lamarre et al., 2012; Swindles et al., 2015b; Pelletier et al., 2017;
167  Zhang et al., 2018).

169  Building on the development of a recent pan-European transfer function (Amesbury et al., 2016), here
170  we apply a similar approach to the North American continent. Through a wide collaborative network,
171  we have compiled almost 2000 published and unpublished testate amoeba surface samples from 137
172  sites covering 34° of latitude and almost 100° of longitude. Our objectives are to (1) describe the ecology
173  of peatland testate amoebae in North America in terms of the key environmental variable, water table
174  depth; (2) develop a new, rigorously tested continental-scale transfer function for reconstructing
175  ombrotrophic peatland palaeohydrology; and (3) as the first step towards a Holarctic synthesis, assess
176  the ecology and biogeography of testate amoeba across two continents by combining our new North-
177  American dataset with data from a pan-European synthesis (Amesbury et al., 2016) to create a
178  combined dataset with >3700 samples in total.

181  **2. Methods**

183  **2.1. Data compilation and taxonomy**
184  We collated a total dataset of 1956 modern surface samples from 137 peatlands located throughout
185  the Canada and the USA (Figure 1; Table 1). Peatlands included were predominately *Sphagnum*-
186  dominated and ombrotrophic, but data from other site types was also included (e.g. Booth and
187  Zygmunt, 2005). Data were compiled from 15 published sources in addition to four unpublished datasets
188  (Table 1). All original datasets, including previously unpublished datasets, in percentage data form and
189  a majority also as raw count data, have been archived and are freely available on the Neotoma
190  Paleoecology Database (www.neotomadb.org). Raw count data was available for 1844 of 1956 samples
191  (excluding the datasets of Warner, 1987 and Payne et al., 2006). For those samples, the mean count
192  was 148 individuals. Only 65 samples had a count <100, with a minimum count in any one sample of 52
193  individuals. All testate amoeba surface samples had an associated water table depth (WTD) value, with
194  a smaller number (n = 1859) also having an associated pH value. Although some datasets contained

197

198 [INSERT FIGURE 1]
199 **Figure 1:** Map showing the location of 137 sites included in the study (see Table 1 for full details). Sites
200 are colour-coded by United States Environmental Protection Agency (EPA) eco-region (see text for
201 details). Locations of two independent palaeo records at Lac Le Caron, Québec and Framboise, Nova
202 Scotia are also shown (black spots). For references to colour, readers are referred to the online version
203 of the article.

204

205 Since the multiple analysts involved in the original data collection (publication dates ranging over 30
206 years from 1987 – 2017) applied slightly varying taxonomic schemes, it was necessary to merge data
207 into a coherent taxonomy that was valid across all datasets. To do this, we followed the precedent set
208 by Amesbury et al. (2016) in a compilation of European peatland testate amoeba data, taking a low-
209 resolution approach and merging like taxa into groups wherever necessary. However, due to the
210 generally greater consistency between original taxonomies than was the case for Amesbury et al.
211 (2016), the relative extent of grouping required here was less, with a total of 72 taxa in the complete
212 merged dataset, of which 25 were groupings containing multiple taxa (Table 2). Despite the
213 improvement in taxonomic resolution from Amesbury et al. (2016), individual analysts should not count
214 samples to the taxonomic scheme used here, but rather define individual taxa to as great an extent as
215 possible (e.g. particularly within *Corythion – Trinema* type), grouping into this scheme for statistical
216 analysis only. Exceptions to the taxonomy of Amesbury et al. (2016), permitted by the relative
217 taxonomic consistency of the original datasets, were as follows:

218

219 1. *Arcella artocrea* was consistently identified in addition to *Arcella arenaria* type and *Arcella catinus*
220    type, so this division remained. However;
221 2. *Arcella catinus* type was identified in most datasets, whereas *Arcella arenaria* type was only
222    identified in two datasets, that did not include *Arcella catinus* type. Therefore, *Arcella arenaria* type
223    was merged into *Arcella catinus* type;
224 3. *Centropyxis cassis* type and *Centropyxis platystoma* type were consistently identified across all
225    datasets, so these divisions remained. *Centropyxis aerophila* type was only identified in two
226    datasets so was merged into *Centropyxis cassis* type;
227 4. *Cyclopyxis arcelloides* type was consistently split into *Cyclopyxis arcelloides* type, *Difflugia
228    globulosa* type and *Phryganella acropodia* type so these divisions remained;
229 5. *Difflugia pulex* type and *Hyalosphenia minuta* were consistently identified in addition to
230    *Cryptodifflugia sacculus*, so all taxa remained and were not grouped into *Cryptodifflugia sacculus*
231    type;
232 6. *Difflugia pristis* type, *Pseudodifflugia fulva* type and *Pseudodifflugia fasicularis* type were
233    consistently identified in addition to *Difflugia lucida* type, so all taxa remained and were not
234    grouped into *Difflugia lucida* type;
235 7. *Difflugia bacillifera* type and *Difflugia rubescens* type were consistently identified in addition to
236    *Difflugia oblonga* type, so all taxa remained and were not grouped into *Difflugia oblonga* type;
237 8. *Euglypha strigosa* type and *Euglypha tuberculata* type were consistently identified in addition to
238    *Euglypha ciliata* type, *Euglypha compressa* and *Euglypha cristata*. *Euglypha strigosa* type and
239    *Euglypha tuberculata* type therefore remained as separate taxa, with *Euglypha ciliata* type
240    containing only *Euglypha ciliata* type, *Euglypha compressa* and *Euglypha cristata*;
241 9. *Nebela penardiana* type was consistently identified in addition to *Gibbocarina (Nebela) tubulosa*
242    type, so this division remained;

243 10. *Heleopera petricola* was consistently identified in addition to *Heleopera sphagni*, so this division
244 remained;
245 11. *Padaungiella (Nebela) wailesi* type was consistently identified in addition to *Padaungiella (Nebela)*
246 *lageniformis*, so this division remained;
247 12. *Planocarina (Nebela) marginata* type was consistently identified in addition to *Planocarina*
248 *(Nebela) carinata* type, so this division remained.
249
250 Following a recent redefinition, we update the nomenclature of *Nebela militaris* to *Alabasta militaris*
251 (Duckert et al., 2018 in press). The complete dataset was screened for data quality and rare taxa. Eight
252 taxa were removed as they occurred in <1% of the total number of samples (i.e. <19 samples; Table S1).
253 Eleven samples were removed where taxonomic merging or deletion of taxa only identified to genera
254 (e.g. *Difflugia* sp., *Euglypha* sp. etc.) resulted in a percentage total <90% (Table S2). A further 18 samples
255 with extreme WTDs below the $0.5^{th}$ (-15 cm) or above the $99.5^{th}$ (67 cm) percentiles of all WTD
256 measurements were removed to avoid undue influence of these samples in subsequent transfer
257 function development (Table S3). This data management resulted in a useable dataset of 64 taxa and
258 1927 samples (Figure 2). Hereafter, this will be referred to as the full dataset. Of these 1927 samples,
259 all, with the exception of those from Warner and Charman (1994), had a known site location. It was not
260 possible to assign samples to exact sites for the Warner and Charman (1994) data due to unspecific
261 sample codes. Therefore, in all analyses requiring geographical information, reduced datasets were
262 used that excluded those samples.
263
264 [INSERT FIGURE 2]
265 **Figure 2:** Percentage distribution of all samples (n = 1927) and taxa (n = 64), ordered vertically by water
266 table depth (wettest samples at the top) and horizontally by water table depth optima from a WA-Tol
267 (inv) model of the full dataset (wettest optima at the left; see also Figure 5).
268
269 To investigate testate amoeba ecological and biogeographical patterns across North America and
270 Europe (Amesbury et al., 2016), we compiled the two continental-scale training sets to form a combined
271 dataset. We used the original merged datasets before the removal of rare taxa (n = 1799 for Europe, n
272 = 1956 for North America, total n = 3755), since combining the two may have resulted in additional
273 samples to characterise previously rare taxa. Given the greater degree of taxonomic grouping in the
274 European data, it was necessary to harmonise the North American data into the European taxonomic
275 scheme (Amesbury et al., 2016). Taxonomic changes made to facilitate this followed the reverse of the
276 12 numbered points above. In addition, three taxa not identified in Europe (*Conicocassis*
277 *pontigulasiformis*, *Nebela barbata* and *Waileseila eboracensis*) were added to the European taxonomy,
278 resulting in an initial 63 defined taxa/taxonomic groups. Once combined, we applied the same data
279 quality and rare taxa criteria as above and in Amesbury et al. (2016). We removed 11 rare taxa present
280 in <1% of the total number of samples (<37 samples), 35 samples with extreme WTDs below the $0.5^{th}$ (-
281 15 cm) or above the $99.5^{th}$ (82 cm) percentiles of all WTD measurements and a further 45 samples with
282 percentage totals <90% (see Tables S4 – S6 for details of removed samples/taxa). This resulted in a
283 workable combined dataset of 52 taxa and 3675 samples.
284
285 All biogeographical inferences resulting from this work should be set against the taxonomic resolution
286 applied. Light microscopic studies necessitate a practical, but relatively low-resolution approach when
287 set against known levels of cryptic and pseudo-cryptic diversity in testate amoebae (e.g. Heger et al.,
288 2013; Oliverio et al., 2014; Kosakyan et al., 2016; Singer et al., 2018). Therefore, where biogeographic
289 patterns are identified at the resolution applied, we can assume that these would still be valid at a higher
290 taxonomic resolution. However, if biogeographic patterns are absent for particular taxa, this does not
291 exclude possible evidence when true genetic diversity is examined.

293    **2.2. Statistical analyses**
294    Considering the key aim of this study to produce a transfer function for palaeohydrological
295    reconstruction, non-metric multidimensional scaling (NMDS) was used *sensu* Amesbury et al. (2016) to
296    objectively examine the dataset and potentially exclude outlying sites and/or samples that were non-
297    typical for the ombrotrophic peatlands commonly used in palaeoclimate research. NMDS was carried
298    out using the 'metaMDS' function in the R package vegan (Oksanen et al., 2015), with water table depth
299    and pH included as potential environmental drivers. Default settings and automatic data transformation
300    were applied and the analysis was conducted based on Bray-Curtis dissimilarity. Environmental variables
301    were passively fitted to the NMDS, and their significance tested, using the 'envfit' function in vegan.
302    Sample-specific values for the Shannon-Wiener diversity index (SWDI) were calculated using the
303    'diversity' function in vegan. Transfer function development was carried out using the R package rioja
304    (Juggins, 2015). Four common model types were applied; weighted averaging with and without
305    tolerance downweighting (WA/WA-Tol; both with potential for inverse (inv) or classical (cla)
306    deshrinking), weighted averaging with partial least squares (WAPLS), maximum likelihood (ML) and the
307    modern analogue technique (MAT; with the potential for a weighted-mean version, WMAT). For each
308    model, only the best performing version (judged by $r^2$ and leave-one-out root mean square error of
309    prediction; $RMSEP_{LOO}$) was selected for further analysis. Additional statistical testing followed
310    convention in the recent literature after Amesbury et al. (2013) and included calculation of leave-one-
311    site-out RMSEP (Payne et al., 2012; $RMSEP_{LOSO}$), segment-wise RMSEP (Telford and Birks, 2011;
312    $RMSEP_{SW}$) and a spatial autocorrelation test (Telford and Birks, 2005, 2009) using the 'rne' (random,
313    neighbour, environment) function in the R package palaeoSig (Telford, 2015). Also, following
314    convention, transfer function development followed a two-step process. We firstly developed models
315    using the full dataset and then, for each model type, performed a second, 'pruned', model run which
316    excluded samples with high residual values greater than 20% of the total range of measured water table
317    depths.
318
319    As part of the testing process, models were applied to independent (i.e. no modern samples from the
320    sites were included in the transfer function) palaeo-data from two *Sphagnum*-dominated ombrotrophic
321    peatlands: Framboise bog, Nova Scotia (45.7208°N, 60.5521°W; H. Mackay, unpublished) and Lac Le
322    Caron peatland, Québec (52.2884°N, 75.8393°W; van Bellen et al., 2011). Original taxonomies of the
323    independent palaeo datasets were merged and updated (i.e. in terms of nomenclature) as necessary in
324    order to match the taxonomic scheme of the new models produced here. Sample-specific errors for all
325    transfer function reconstructions were based on 1000 bootstrapping cycles. We compared our
326    reconstructions with output from the most geographically extensive published transfer function in
327    North America (Booth, 2008) and tested the significance of all new reconstructions (Telford and Birks,
328    2011b) using the 'randomTF' function in the R package palaeoSig (Telford, 2015; Payne et al., 2016).
329
330    To test for regional variability within the continental dataset, we developed regional models based on
331    United States Environmental Protection Agency (EPA) Level 1 eco-region sub-divisions
332    (www.epa.gov/eco-research/ecoregions), using the pruned WA-Tol (inv) model (excluding samples
333    from Warner and Charman, 1994; see Section 2.1) as the basis for regional model development (n =
334    1696). Our data covered eight different eco-regions but to ensure regional models were representative,
335    we only developed models for those regions containing >50 samples (Table 3). This resulted in five
336    regional models: taiga (samples n = 68, sites n = 7), northern forests (samples n = 978, sites n = 56),
337    northwestern forested mountains (samples n = 136, sites n = 16), marine west coast forest (samples n
338    = 266, sites n = 23) and eastern temperate forests (samples n = 377, sites n = 18). Models were not
339    developed for the tundra (samples n = 7, sites n = 1), Hudson Plain (samples n = 32, sites n = 7) and
340    North American deserts (samples n = 11, sites n = 1) eco-regions. The independent palaeo records were

341  located in the northern forest (Framboise) and taiga (Lac Le Caron) eco-regions. The programme PAST
342  (v3.20; Hammer et al., 2001) was used to run one-way PERMANOVA tests (9999 iterations, assessed by
343  Bray-Curtis dissimilarity) of the differences between samples from different eco-regions.
344
345  To test the validity of a combined North American and European transfer function, we applied the North
346  American (this study), European (Amesbury et al., 2016) and a pruned combined model (n = 3425; all
347  WA-Tol (inv)) to independent palaeo data from both North America (Framboise) and Europe (Tor Royal
348  Bog, UK *sensu*. Amesbury et al., 2016). We also qualitatively compared the modelled taxa water table
349  depth optima and relative optima ranking of the three models.
350
351
352  **3. Results**
353
354  **3.1. Ordination**
355  Summary NMDS results are shown in Figure 3. For the NMDS of the full dataset (Figures 3A and 3B),
356  square root transformation was automatically applied. WTD was significantly correlated to NMDS axis
357  1 (p < 0.001), although the WTD vector was at 45 degrees to the horizontal, with taxa favouring drier
358  conditions in the top left of the ordination space and taxa favouring wetter conditions in the bottom
359  right (Figure 3B). In general, all samples fell in a large, fairly coherent single swarm of data, with no
360  significant outliers, or groups of outliers, though samples tended to cluster more in the uppermost part
361  of the swarm (Figure 3A). When sub-divided by eco-region (Figure 3C), there was a high degree of
362  overlap, with no regions outlying all others. Only samples from the northwestern forested mountain
363  eco-region tended to cluster in a specific part of the sample swarm (towards the left; Figure 3C), whereas
364  samples from all other regions were located throughout the ordination space. This high degree of
365  overlap, coupled with the significant correlation of WTD to NMDS axis 1 provided a strong justification
366  to continue with transfer function development using the full dataset.
367
368  [INSERT FIGURE 3]
369  **Figure 3:** NMDS biplots for A) the full dataset (n = 1927) showing sample positions, B) the full dataset
370  showing taxa positions, C) a reduced dataset excluding samples from Warner and Charman (1994) (n =
371  1884) showing sample positions, colour-coded by EPA eco-region (see text for details). Vectors on all
372  plots show correlation with environmental drivers. Some taxa positions in B have been marginally
373  altered to improve legibility, but relative positions remain intact. Full names for taxa abbreviations can
374  be found in Table 2. For references to colour, readers are referred to the online version of the article.
375
376  **3.2. Transfer function development**
377  Performance statistics for the best-performing version of each model type are shown in Table 4. For the
378  full dataset, the range of WTD values was 82 cm (minimum WTD -15cm; maximum WTD 67 cm),
379  therefore a value of 16.4 cm (20% of range) was used to remove outlier samples during data pruning.
380  The number of pruned samples varied for each individual model type (Table 4). Following data pruning,
381  the WMAT K5 (weighted-mean modern analogue technique with five nearest neighbours) model was
382  the best performing when judged by $RMSEP_{LOO}$ and $R^2$. ML was the worst performing model (based on
383  $RMSEP_{LOO}$) when all samples were included but improved after data pruning (e.g. second best pruned
384  model for $r^2$ only), largely down to the removal of 410 samples with high residual values, more than
385  double the number of pruned samples than for the next highest model (WA-Tol (inv); n = 197). Pruned
386  samples for all models tended to be reasonably evenly distributed between wet and dry extremes,
387  however ML produced far more extreme negative residuals (i.e. samples where WTD was substantially
388  overestimated; Figure 4). Partly as a result, both ML and WMAT K5 suffered from comparatively high
389  average bias values when compared to the two WA-based models, although the WA-based models had

390    higher maximum bias (Table 4). WAPLS C2 and WA-Tol (inv) performed very similarly on all metrics, with
391    WA-Tol (inv) marginally outperforming WAPLS C2 after data pruning.
392
393    All models showed a decrease in relative performance for $RMSEP_{LOSO}$ and $RMSEP_{SW}$ compared to
394    $RMSEP_{LOO}$, though for ML, both reductions were marginal (Table 4). Relative performance ratings also
395    changed when different versions of RMSEP were applied. For example, when using $RMSEP_{LOSO}$, WA-
396    based models were the best performing, but using $RMSEP_{SW}$, both showed a higher relative decrease in
397    performance than WMAT K5 and ML (though $RMSEP_{SW}$ for the WA-based models was still lower than
398    for ML; Figure S1). Both WA-based models performed worse in 'dry' segments (i.e. WTDs greater than
399    40 cm) with comparatively less samples, whereas WMAT K5 and ML displayed more consistent RMSEP
400    values across all segments, suggesting that these models were not as affected by WTD and/or segment
401    sampling frequency (Figure S1). WMAT K5 performed relatively poorly for $RMSEP_{LOSO}$, suggesting that
402    this model type does not deal well with clustered training sets (Amesbury et al., 2016).
403
404    All models displayed evidence of spatial autocorrelation, given that $r^2$ declined much more rapidly when
405    geographically close compared to random samples were removed; this decline was sharpest for WMAT
406    K5 (Figure S2). $R^2$ values for WMAT K5 and ML declined steadily with the continued removal of
407    geographically close samples, whereas, interestingly, $r^2$ increased from previous values for both WA-
408    based models when all samples within 1000 km of each other were removed, such that the 1000 km
409    values were higher than for 100 km. Even values at 2500 km (i.e. the models would only contain samples
410    that are >2500 km apart) were comparable to 100 km. These results suggest that, despite the initial
411    decline in $r^2$, the WA-based models are more robust to spatial autocorrelation than WMAT K5 or ML.
412
413    Taxon-specific values for WTD optima and tolerance showed taxa ordered along a WTD gradient both
414    comparable to the diagonal WTD vector along axis 1 of the NMDS (Figure 3) and also to other studies.
415    In contrast to a European compilation (Amesbury et al., 2016), where taxonomic grouping resulted in
416    more similar tolerance ranges along the WTD gradient, tolerances in this dataset (Figure 5) showed a
417    more typical pattern of being wider for drier taxa and vice versa (Mariusz Lamentowicz and Mitchell,
418    2005). For example, the mean tolerance for the driest half of taxa, ranked by optima, was 12.5 cm,
419    whereas the same value for the wettest half was 8.4 cm (significantly different at $p < 0.05$ for a two-
420    sample, two-tailed t-test, df = 60).
421
422    [INSERT FIGURES 4 AND 5]
423    **Figure 4:** Biplots of leave-one-out cross-validated observed versus predicted water table depth (left)
424    and residual error plots (right) for the four model types under investigation. See text for model
425    abbreviations. Orange points are model runs with all data, black points are model runs after the removal
426    of samples with high residual values. For references to colour, readers are referred to the online version
427    of the article.
428
429    **Figure 5:** Water table depth optima and tolerances (cm) for 64 taxa based on the WA-Tol (inv) model
430    after the removal of outlying samples (n = 1730).
431
432    **3.3. Testing model application**
433    All four models were applied to independent palaeo records from Framboise and Lac Le Caron peatlands
434    (Figure 1) and compared to reconstructions using the transfer function of Booth (2008), the most
435    geographically extensive published transfer function from North America (Figure 6; Figure S3). All model
436    types, as well as the Booth (2008) reconstructions, showed broadly similar patterns of change (Figures
437    6 and S3). WMAT K5 and ML reconstructions tended to reconstruct somewhat more abrupt (i.e. change
438    over less samples) and extreme (i.e. higher and lower water table depth values at either end of the

439  scale) change, when compared to the WA-based models, a trend particularly evident in the Lac Le Caron
440  reconstructions. In general, all four model types tended to co-vary within a range of 5 – 10 cm, however
441  there were periods when particular reconstructions were more divergent. For example, at 10 cm depth
442  at Framboise, the WMAT K5 model reconstructed a significant dry shift not seen in other models (Figure
443  6A) and at Lac Le Caron, the ML model produced wet shifts at 36 cm and around 215 cm that were not
444  replicated by other models, as well as reconstructing generally wetter conditions for the lower ~150 cm
445  of the profile (Figure 6B). However, when viewed as z-scores (Figure 6C and D; Swindles et al., 2015a),
446  the majority of shifts were aligned, suggesting that all models are reliably reconstructing the main
447  patterns of palaeohydrological change in the profiles.

449  [INSERT FIGURE 6]
450  **Figure 6:** Comparison of transfer function reconstructions from four model types. A and B are raw water
451  table values (cm). C and D are z-scores. A and B include comparison to reconstruction from Booth (2008).
452  A and C for Framboise bog, B and D for Lac Le Caron peatland. For all plots, higher y-axis water table
453  depth or z-score values indicate drier conditions. Note varying y-axis scales.

455  The significance of all reconstructions was tested against models trained on random data (Telford and
456  Birks, 2011b; Table 5). Reconstructions at Framboise for all model types were either significant (i.e.
457  <0.05) or very close to significant (0.057 for ML). However, equivalent $p$ values at Lac Le Caron were
458  consistently insignificant, often to extremes. For example, ML had the lowest $p$ value at 0.287 (despite
459  having the highest value at Framboise), with both WA-based models having $p$ values > 0.98, meaning
460  their reconstructive ability was worse than almost all models trained on randomly generated data (see
461  also Payne et al., 2016).

463  **3.4. Eco-region variability and models**
464  One-way PERMANOVA tests showed that there were significant differences in testate amoeba
465  assemblages between factor eco-regions ($p < 0.0001$, F = 15.16). Species richness (i.e. number of taxa
466  present) differed among eco-regions whereas eco-region mean species diversity values were typically
467  more consistent (in the range 1.9 – 2, overall range for all samples 0.19 – 3.07), with the exception of
468  the northwestern forested mountains and North American deserts regions (Table 3). Those eco-regions
469  characterised by only a single site (tundra, North American deserts) had expectedly lower species
470  richness, but, in the case of the tundra region, still displayed species diversity equivalent to the total
471  dataset and to other regions (e.g. Swindles et al., 2009, 2014), despite being based on only seven
472  samples. There were notable differences in species richness among regions with more comparable
473  numbers of sites and samples. For example, eastern temperate forests (18 sites, 346 samples) contained
474  62 different taxa, whereas northwestern forested mountains (16 sites, 131 samples) contained only 42
475  taxa. In this case, the 20 fewer taxa in the northwestern forested mountains were often relatively rare
476  taxa that were absent from multiple regions, however also included the absence of other relatively
477  abundant taxa present in all other regions (e.g. *Amphitrema wrightianum*, *Bullinularia indica*, *Difflugia
478  oblonga*, *Difflugia pulex*). Inter-regional assemblage differences were investigated by plotting taxa
479  distributions across the eight eco-regions (Figure S5). Only eight of 64 taxa occurred in all eight eco-
480  regions: *Assulina muscorum*, *Nebela tincta* type, *Euglypha tuberculata*, *Hyalosphenia papilio*, *Corythion-
481  Trinema* type, *Centropyxis cassis* type, *Euglypha rotunda* type and *Difflugia globulosa* type. However,
482  allowing for the reduced species richness in the less-sampled tundra and North American desert eco-
483  regions (Table 3; regions 1 and 8 on Figure S5), 42 of 64 taxa occurred in six or more eco-regions. Of the
484  22 taxa that were present in five or less eco-regions, 14 were absent from the less-sampled tundra and
485  North American desert regions, as well as the relatively taxa-poor northwestern forested mountains
486  region. Of those 14 taxa, the most frequent were *Physochila griseola* (619 (out of 1884) occurrences),
487  *Bullinularia indica* (519 occurrences) and *Centropyxis ecornis* type (376 occurrences). Taxa present in

488     only two or three eco-regions were largely driven by low occurrence, with only *Nebela minor* of that
489     group occurring in >100 samples (Figure S5).
490
491     Taxa tended to occur on a continuum where their presence in samples and eco-regions, and local
492     abundance, were broadly aligned (Figure S6). For example, common, locally abundant key hydrological
493     indicator taxa such as *Assulina muscorum*, *Trigonopyxis arcula* type and *Archerella flavum* tended to
494     occur in multiple (i.e. 6-8) eco-regions, whereas rare (in both occurrence and abundance) taxa such as
495     *Pseudodifflugia fasicularis* type and *Padaungiella lageniformis* type tended to have more restricted
496     ranges, occurring in only a few (i.e. 2-3) eco-regions. However, this overall trend also encompassed
497     considerable variability. Relatively rare taxa could still be locally abundant and occur across multiple
498     eco-regions. For example, *Arcella hemispherica* occurred in only 3% of samples, but had a maximum
499     abundance of 63% and was present in six eco-regions. Taxa that were rare in both occurrence and
500     abundance could still occur across multiple eco-regions. For example, *Quadrulella symmetrica* was
501     present in only 3.6% of samples, never above 9% abundance, but was only absent from one eco-region
502     (North American desert). The absence of *Q. symmetrica* from the North American desert eco-region is
503     noteworthy given the recent discovery of a *Quadrulella* taxon in a Mexican desert (Pérez-Juárez et al.,
504     2017). Finally, locally abundant taxa could also have more restricted ranges. For example, *Difflugia*
505     *oviformis* type occurred in only two eco-regions (NF and ETF), but formed more than half of the
506     assemblage in some samples in those regions.
507
508     To further test overall model performance and to further investigate internal biogeographical variability,
509     we developed models for five eco-regions, using only the WA-Tol (inv) model type (*cf*. Amesbury et al.,
510     2016; see Section 2.2). Eco-region model reconstructions varied notably (Figures 7 and S4), in terms of
511     value, pattern and, less often, direction of change. Most notably, the northwestern forested mountains
512     (NFM) model produced reconstructions, for both palaeo datasets, that had notably drier reconstructed
513     water table depth values when compared to all other models (including the pruned WA-Tol (inv) model
514     based on the full dataset; Figure 7A and B). Mean WTD for individual samples and the overall range of
515     WTDs sampled for the NFM model were equivalent to other models (Table 3), suggesting these
516     differences were not simply driven by poor characterisation of 'wet' conditions in the model. Even when
517     viewed as z-scores, the NFM model appeared anomalous, with often conflicting patterns and direction
518     of change, despite having lower $RMSEP_{LOO}$ and higher $R^2$ values than all other regional models (Table 3).
519     The eastern temperate forest (ETF) model most closely resembled the WA-Tol (inv) reconstruction at
520     both sites; this model had the highest number of samples and joint highest number of taxa of the eco-
521     region models (Table 3). Despite having a much lower $R^2$, and higher maximum and average bias values
522     (Table 3), the taiga reconstructions were broadly in line with the WA-Tol (inv) reconstruction at both
523     sites. The northern forest (NF) model reconstructions, despite being based on a high number of species
524     and taxa compared to most other eco-region models (Table 3), were variable in the extent to which they
525     aligned with other models. At Framboise, a site within the NF eco-region, the NF reconstruction was
526     closely aligned to the ETF and WA-Tol (inv) reconstructions, whereas at Lac Le Caron, it often exhibited
527     differing patterns and direction of change, suggesting potential problems with supra-regional
528     application of eco-region models (Turner et al., 2013). All eco-region model reconstructions at both sites
529     were tested for significance against models trained on random data (Telford and Birks, 2011b; Table 5).
530     *P* values for only two of ten reconstructions fell under the suggested significance cut-off value of 0.05
531     (NF and ETF models at Framboise, 0.015 and 0.007 respectively). Other *p* values ranged from 0.471 to
532     0.997 and no reconstructions at Lac Le Caron were significant. <span style="color:red">Taxa optima for the five eco-region</span>
533     <span style="color:red">models were compared to those of the full dataset (Figure S7). Taxa optima rank, *cf*. Figure 8, was not</span>
534     <span style="color:red">compared due to variable n for taxa in each eco-region model. Optima for individual taxa varied from</span>
535     <span style="color:red">0.8 cm (*A. wrightianum*) to 42.5 cm (*Difflugia rubescens*), with a mean value of 11.6 cm and standard</span>
536     <span style="color:red">deviation of 7.3 cm.</span>

537

538 [INSERT FIGURE 7]

539 **Figure 7:** Comparison of transfer function reconstructions from five eco-region models. All panels
540 include comparison to pruned continental WA-Tol (inv) model. A and B are raw water table values (cm).
541 C and D are z-scores. A and C for Framboise bog, B and D for Lac Le Caron peatland. Model abbreviations
542 in key: NF = northern forests; NFM = northwestern forested mountains; MWCF = marine west coast
543 forest; ETF = eastern temperate forests. For all plots, higher y-axis water table depth or z-score values
544 indicate drier conditions. Note varying y-axis scales.

545

546 **3.5. Combined North American and European compilation**
547 Model development began with a dataset of 52 taxa and 3675 samples. Data pruning removed 250
548 samples with residual values >20% of the range of measured water table depths (19.4 cm). The pruned,
549 combined WA-Tol (inv) model (n = 3425) had an $RMSEP_{LOO}$ of 7.94 cm and $R^2$ of 0.65, similar to
550 continental North American (RMSEP 7.42 cm, $R^2$ 0.72; Table 4) and European (RMSEP 7.72 cm, $R^2$ 0.59;
551 Amesbury et al., 2016) models, despite an almost doubling of sample numbers. Reconstructions using
552 North American, European (Amesbury et al., 2016) and combined models for independent palaeo
553 records in both Europe and North America were very similar when viewed as both reconstructed water
554 table depth values and z-scores (Figure 8). Individual taxa water table depth optima, whether in cm or
555 rank order, were not significantly different across the three models (comparing the European and North
556 American data as the only completely independent variables; for optima r = 0.80, $p$ <0.05, for rank $r_s$ =
557 0.82, $p$ <0.05), confirming WTD as an ecologically important environmental variable (*cf*. Juggins, 2013).
558 Despite the general consistency in WTD optima, some variability was evident. Most notably, *Difflugia*
559 *labiosa* type was the wettest ranked taxa in the North American and combined models, with negative
560 values indicative of standing water, but ranked 45 of 46 taxa in the European model (Amesbury et al.,
561 2016) with a dry water table depth optima of ~25 cm. However, in the European dataset, *D. labiosa* was
562 only found in one sample in one eco-region (Scandinavia) at very low abundance (0.7%; Amesbury et
563 al., 2016) and was therefore poorly characterised. Its more extensive presence in the North American
564 dataset (as *D. oviformis* type) in 60 samples spanning two eco-regions (northern forests and eastern
565 temperate forests) with a maximum abundance of >50% provides better quality ecological data for this
566 taxon and highlights the value of large, ecologically diverse datasets in improving modern analogue
567 quality, as well as the persistent issue of taxonomic consistency (Payne et al., 2011). Other differences
568 in optima and rank were less extreme, but varied by up to 15 ranking positions or approximately 10 cm
569 in terms of water table optima (e.g. *Nebela penardiana* type, *Assulina seminulum* type). On average,
570 taxa optima varied by 4.3 cm and 5.2 ranking positions across all three models (equivalent values of 3.8
571 cm and 4.4 ranking positions excluding the anomalous *D. labiosa* type).

572

573 [INSERT FIGURES 8 AND 9]

574 **Figure 8:** Comparison of transfer function reconstructions from European, North American and
575 combined models. All models are pruned WA-Tol (inv). A and B are raw water table values (cm). C and
576 D are z-scores. A and C for Framboise, B and D for Tor Royal Bog. For all plots, higher y-axis water table
577 depth or z-score values indicate drier conditions. Note varying y-axis scales.

578

579 **Figure 9:** Comparison of water table depth optima (cm) for 46 taxa from European, North American and
580 combined models, ordered by the combined model. Top panel: rank order of taxa (coloured spots) with
581 range between highest and lowest rank (grey bars). Bottom panel: modelled optima (coloured spots)
582 with range between highest and lowest optima (grey bars). For references to colour, readers are
583 referred to the online version of the article.

584

585

## 4. Discussion

### 4.1. A new continental-scale transfer function for North America

Performance statistics (i.e. $r^2$ and $RMSEP_{LOO}$) for the North American model were equivalent to a recent European compilation (Amesbury et al., 2016) and a range of other regional-scale models within North America and Europe (Table S7, i.e. within mean ± 1σ of 20 reviewed models). Further statistical tests (*sensu*. Amesbury et al., 2013) showed that potential issues such as clustered sampling design and spatial autocorrelation had only minimal effect on model performance and output. Reconstructions on independent palaeo-data were comparable to a less sample-dense and more geographically restricted North American model (Booth, 2008). Taken together, these multiple strands of evidence suggest that our new model can be used as an effective tool to reconstruct peatland palaeohydrology throughout the North American continent. More ecologically bespoke regional models may still provide reliable reconstructions within and potentially even outside of their geographical limits (Turner et al., 2013; Willis et al., 2015) and have an important role to play in ongoing research where they suit specific research questions. However, the volume of data in the continental-scale compiled dataset provides improved modern analogues for many taxa. All four models tested performed similarly in statistical terms, however due to potential issues with spatial autocorrelation and the somewhat abrupt nature of some reconstructed shifts in WTD with WMAT K5 and ML (Figure 6), as well as the differing magnitude of ML reconstructions (Figure 6A and B), we favour WA-based models. A pruned WA-Tol (inv) model marginally outperformed WAPLS C2 and produced reconstructions slightly more aligned to the model of Booth (2008), so WA-Tol (inv) is our preferred model choice.

The variability of $p$ values for significance testing of transfer function reconstructions (Telford and Birks, 2011b) between sites and models trained on the same data brings the use of the 'randomTF' test into further question (Amesbury et al., 2016; Payne et al., 2016). Across all model types and eco-regions, no reconstructions from Lac Le Caron had $p$ values <0.05. Eco-region model reconstructions for Framboise varied widely between 0.007 and 0.782. At Framboise, NF and ETF model reconstructions were both significant (0.015 and 0.007 respectively), whereas MWCF and taiga model reconstructions were non-significant (0.471 and 0.485 respectively), despite showing extremely similar patterns of change (Figure 7; Table 5). Given that the efficacy of the 'randomTF' method has been recently reviewed and questioned (Amesbury et al., 2016; Payne et al., 2016), these results further call into question the usefulness of this test.

To make transfer functions more openly accessible to the research community and to facilitate users' own application of the model to fossil data, we provide the full compiled North American dataset along with R code to apply the WA-Tol (inv) model as supplementary information to this manuscript. In addition, site-based datasets preserving the original taxonomy of the individual studies are available through the Neotoma Paleoecological Database (www.neotomadb.org).

### 4.2. North-American regional variability

Biogeographical patterns that are evident in the dataset appear to be largely driven either by sampling frequency or restricted to particular eco-regions (e.g. NFM). Outside of these limitations, biogeographical differences occurred against a background of relative cosmopolitanism across the dataset. For example, the finding that *H. papilio* occurred across all eco-regions is particularly interesting given recent evidence suggesting that it displays geographical patterns of both genotypic and phenotypic diversity in other regions (Heger et al., 2013; Amesbury et al., 2016). This observation is reassuring for the broader applicability of transfer functions across large geographical areas. Although we used only a moderate taxonomic resolution (higher than for a European compilation; Amesbury et al., 2016) that did not allow us to identify the full complexity of biogeographic patterns, and hence may

635  have caused us to miss some discrepancies in taxa hydrological optima, this potential loss of information
636  did not result in poor model performance. Even though a higher taxonomic resolution may permit the
637  development of more precise transfer function models it appears that the taxonomic scheme currently
638  used by ecologists and palaeoecologists is robust and the resulting model performance satisfactory.
639
640  Taxa optima were broadly consistent across eco-region models (Figure S7), with mean variability
641  between eco-region models of 11.6 cm. Highly variable water table depth optima for individual taxa
642  tended to be driven by poor characterisation. For example, *Difflugia rubescens* and *Arcella gibbosa* both
643  had tightly clustered optima across a majority of models, with outlying values for the MCWF model only
644  driving the two highest values for overall optima range; in that model, the optima for both taxa were
645  driven by occurrence in a single sample (from 191 samples in the model; Table 3), both at <2%. Likewise,
646  the outlying optima for *Padaungiella lageniformis* in the NF model was driven by only 6 occurrences
647  (from 915 samples in the model; Table 3), never at >1% abundance. Thus, whilst the variability evident
648  between taxa optima across eco-regions may be partly driven by genuine biogeographical differences,
649  the overall consistency provides support for continental-scale compilation. The variability in eco-region
650  model reconstructions illustrates a potential advantage of sample dense continental-scale compilations.
651  In particular, the NFM model produced anomalous reconstructions that were significantly drier (i.e.
652  higher WTD) than other eco-region models and also anomalous when viewed as z-scores. Although not
653  clearly outlying and therefore included in the data analysis, this was the only region to be partly
654  clustered in the NMDS analysis (section 3.1) and was relatively taxa poor. All sites in this eco-region
655  were from the Rocky Mountains and, although *Sphagnum*-dominated, were typically fens and included
656  sites that were fed by highly acidic and iron-rich groundwater (Booth and Zygmunt, 2005), which
657  appears to have driven a distinct local flora. Although this produced anomalous results for a regional-
658  scale model, when these data were included in a continental-scale model, potentially confounding
659  factors did not overtly effect the magnitude or direction of change of the reconstruction (Figure 7),
660  whilst valuable ecological information was maintained.
661
662  **4.3. Spatial scales for transfer function development**
663  Continental-scale models from North America and Europe, as well as a combined model, reconstructed
664  independent test data from both continents identically, in terms of palaeohydrological interpretation.
665  Despite some variability, water table depth optima were also broadly aligned between continents
666  (Figure 9). Modelled water table depth optima and taxa distribution may vary on local or regional scales
667  due to differences in peatland type (e.g. forested vs. non-forested; Beaulne et al., 2018), climatic context
668  (Booth and Zygmunt, 2005) or the influence of non-linear processes such as permafrost (e.g. at the
669  unpublished Scotty Creek and Toolik sites (Table 1); see also Lamarre et al., 2013). However, the relative
670  position and ecological niches of taxa in respect to surface wetness is often more consistent (Booth and
671  Zygmunt, 2005; Beaulne et al., 2018). Our comparison of water table depth optima and rank between
672  European and North American compilations showed that, with the exception of a single anomalous
673  taxon, differences were relatively small (Figure 9) and importantly, did not have a noticeable effect on
674  reconstructions of test data from both within and outside of each continent, especially when viewed as
675  residual values (Figure 8; Swindles et al., 2015a). Therefore, it appears that potentially confounding
676  climatic or peatland type differences that may exist within continental-scale compiled datasets are
677  compensated by more in-depth characterisation of the average ecological niche of each taxon.
678
679  These findings also support the recent assertion that transfer functions trained on very large datasets
680  perform well due to the vastly increased number of analogues contained within them, which provide a
681  better representation of testate amoeba ecology than less populated local- or regional-scale models
682  where modern analogues may be either lacking or based on low numbers of samples for some taxa
683  (Amesbury et al., 2016). In a test of regional and supra-regional application of transfer functions, the

684     only difference between reconstructions was in the absolute magnitude of change (i.e. in reconstructed
685     water table depth values), with direction of change, the most reliable variable for palaeoclimatic
686     interpretation, consistent (Turner et al., 2013; Swindles et al., 2015a). Given that the recent suggestion
687     to display transfer function output as standardised z-scores (Swindles et al., 2015a; Amesbury et al.,
688     2016) tends to mitigate magnitude differences, we therefore advocate the further development and
689     use of continental-scale models that maximise modern analogue quality without losing reconstructive
690     skill and provide standard reconstruction tools with directly comparable results.
691
692     Continental-scale models from North America (this study) and Europe (Amesbury et al., 2016) have
693     similar performance statistics than regional- or local-scale models, but potentially offer additional
694     scientific benefit. The compilation and taxonomic harmonisation of such datasets, alongside their
695     archiving in freely available repositories (e.g. Neotoma), permits further community-driven effort
696     towards a better understanding of wide-scale testate amoeba ecology and biogeography that can
697     contextualise and enhance the interpretation of transfer function results. In addition, the lack of
698     significant reduction in performance statistics and similarity of reconstructions produced by a combined
699     North American and European model, when compared to the two continental-scale models suggests
700     that any geographical limit of scale beyond which further compilation of peatland testate amoebae data
701     would not be beneficial, has not yet been reached. We therefore call for the continued synthesis of
702     modern testate amoeba data, particularly for Eurasia. Our results show that harmonisation of this data
703     alongside existing datasets may permit a future Holarctic synthesis of peatland testate amoeba ecology
704     and biogeography. A potential challenge when extending beyond the Holarctic realm is that additional
705     Southern Hemisphere specific taxa (e.g. *Alocodera*, *Apodera*, *Certesella)* will need to be considered,
706     while many characteristic taxa of northern peatlands will be missing (e.g. *Planocarina*, *Hyalosphenia*
707     *papilio*).
708

**5. Conclusions**

710     We compiled a dataset of 1956 modern surface testate amoeba samples from 137 peatlands located
711     throughout the United States of America and Canada. Following data management procedures, a
712     reduced dataset of 1927 samples was used, following exploratory NMDS analysis, to develop
713     continental-scale transfer function models. These models were tested using a range of statistical
714     approaches and application to independent palaeo-data. Though all pruned models tested were valid,
715     WA-Tol (inv) was selected as our preferred model choice. Our new model provides an effective,
716     standardised tool for reconstructing peatland palaeohydrology throughout the North American
717     continent. Using the compiled dataset and models based on EPA Level 1 eco-regions, we examined
718     regional variability within the dataset. Some regions and taxa demonstrated differences in taxonomic
719     diversity and/or biogeography, but these were set against a broad pattern of cosmopolitanism across
720     the dataset. To examine the potential for a future Holarctic synthesis of peatland testate amoeba data,
721     we combined our new North American dataset with an existing recent compilation from Europe to form
722     a combined dataset of >3700 samples. Model outputs from this compilation were equivalent to the
723     North American and European models independently, suggesting that there is scope for further merging
724     of data to develop a future Holarctic dataset and transfer function. Such a dataset would be an
725     extremely valuable community resource, allowing collaborative examination of peatland testate
726     amoeba ecological and biogeographical patterns to an extent not currently possible.
727

**References**

Amesbury, M.J., Mallon, G., Charman, D.J., Hughes, P.D.M., Booth, R.K., Daley, T.J., Garneau, M., 2013. Statistical testing of a new testate amoeba-based transfer function for water-table depth reconstruction on ombrotrophic peatlands in north-eastern Canada and Maine, United States. J. Quat. Sci. 28, 27–39.

Amesbury, M.J., Swindles, G.T., Bobrov, A., Charman, D.J., Holden, J., Lamentowicz, M., Mallon, G., Mazei, Y.A., Mitchell, E.A.D., Payne, R.J., Roland, T.P., Turner, T.E., Warner, B.G., 2016. Development of a new pan-European testate amoeba transfer function for reconstructing peatland palaeohydrology. Quat. Sci. Rev. 152, 132–151.

Baas Becking, L.G.M., 1934. *Geobiologie of Inleiding tot de Milieukunde*. The Hague: Van Stockum & Zoon.

Beaulne, J., Magnan, G., Garneau, M., 2018. Evaluating the potential of testate amoebae as indicators of hydrological conditions in boreal forested peatlands. Ecol. Indic. 91, 386-394.

Beyens, L., Bobrov, A., 2016. Evidence supporting the concept of a regionalized distribution of testate amoebae in the Arctic. Acta Protozool. 55, 197–209.

Bobrov, A.A., Charman, D.J., 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.

Booth, R.K., 2001. Ecology of testate amoebae (Protozoa) in two Lake Superior coastal wetlands: implications for paleoecology 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. J. Paleolimnol. 28, 329–348.

Booth, R.K., 2008. Testate amoebae as proxies for mean annual water-table depth in Sphagnum-dominated peatlands of North America. J. Quat. Sci. 23, 43–57.

Booth, R.K., Sullivan, M.E., Sousa, V.A., 2008. Ecology of testate amoebae in a North Carolina pocosin and their potential use as environmental and paleoenvironmental indicators. Écoscience 15, 277–289.

Booth, R.K., 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. Divers. Distrib. 11, 577–590.

Charman, D.J., Warner, B.G., 1997. The ecology of testate amoebae (Protozoa: Rhizopoda) in oceanic peatlands in Newfoundland, Canada: Modelling hydrological relationships for palaeoenvironmental reconstruction. Écoscience 4, 555–562.

Charman, D.J., Blundell, A., ACCROTELM Members, 2007. A new European testate amoebae transfer function for palaeohydrological reconstruction on ombrotrophic peatlands. J. Quat. Sci. 22, 209–221.

776  Charman, D.J., Warner, B.G., 1992. Relationship between testate amoebae (Protozoa: Rhizopoda) and
777  microenvironmental parameters on a forested peatland in northeastern Ontario. Can. J. Zool. 70, 2474–
778  2482.

779  De Wit, R., Bouvier, T., 2006. "Everything is everywhere, but, the environment selects"; what did Baas
780  Becking and Beijerinck really say? Environ. Microbiol. 8, 755–758.

781  Duckert, C., Blandenier, Q., Kupferschmid, F.A.L., Kosakyan, A., Mitchell, E.A.D., Lara, E., Singer, D.
782  2018 in press. En garde! Redefinition of *Nebela militaris* (Arcellindia, Hyalospheniidae) and erection of
783  *Alabasta* gen. nov. Eur. J. Protistol.

784  Fernández, L.D., Fournier, B., Rivera, R., Lara, E., Mitchell, E.A.D., Hernández, C.E., 2016. Water–energy
785  balance, past ecological perturbations and evolutionary constraints shape the latitudinal diversity
786  gradient of soil testate amoebae in south-western South America. Glob. Ecol. Biogeogr. 25, 1216–1227.

787  Foissner, W., 2006. Biogeography and dispersal of micro-organisms: A review emphasizing protists. Acta
788  Proto 45, 111–136.

789  Foissner, W., 2008. Protist diversity and distribution: some basic considerations. Biodivers. Conserv. 17,
790  235–242.

791  Fournier, B., Coffey, E.E.D., van der Knaap, W.O., Fernández, L.D., Bobrov, A., Mitchell, E.A.D., 2015. A
792  legacy of human-induced ecosystem changes: Spatial processes drive the taxonomic and functional
793  diversities of testate amoebae in Sphagnum peatlands of the Galápagos. J. Biogeogr. 43, 533–543.

794  Geisen, S., Mitchell, E.A.D., Wilkinson, D.M., Adl, S., Bonkowski, M., Brown, M.W., Fiore-Donno, A.M.,
795  Heger, T.J., Jassey, V.E.J., Krashevska, V., Lahr, D.J.G., Marcisz, K., Mulot, M., Payne, R., Singer, D.,
796  Anderson, O.R., Charman, D.J., Ekelund, F., Griffiths, B.S., Rønn, R., Smirnov, A., Bass, D., Belbahri, L.,
797  Berney, C., Blandenier, Q., Chatzinotas, A., Clarholm, M., Dunthorn, M., Feest, A., Fernández, L.D.,
798  Foissner, W., Fournier, B., Gentekaki, E., Hájek, M., Helder, J., Jousset, A., Koller, R., Kumar, S., La Terza,
799  A., Lamentowicz, M., Mazei, Y., Santos, S.S., Seppey, C.V.W., Spiegel, F.W., Walochnik, J., Winding, A.,
800  Lara, E., 2017. Soil protistology rebooted: 30 fundamental questions to start with. Soil Biol. Biochem.
801  111, 94–103.

802  Hammer, Ø., Harper, D.A.T., Ryan, P.D., 2001. PAST: paleontological statistics software package for
803  education and data analysis. Palaeontol. Electron. 4, 9pp. http://palaeo-
804  electronica.org/2001_1/past/issue1_01.htm.

805  Heger, T.J., Mitchell, E.A.D., Leander, B.S., 2013. Holarctic phylogeography of the testate amoeba
806  Hyalosphenia papilio (Amoebozoa: Arcellinida) reveals extensive genetic diversity explained more by
807  environment than dispersal limitation. Mol. Ecol. 22, 5172–5184.

808  Juggins, S., 2013. Quantitative reconstructions in palaeolimnology: New paradigm or sick science? Quat.
809  Sci. Rev. 64, 20–32.

810  Juggins, S., 2015. Rioja: Analysis of Quaternary Science Data. R Package Version 0.9-5. https://cran.r-
811  project.org/web/packages/rioja/index.html.

812  Kosakyan, A., Lahr, D.J.G., Mulot, M., Meisterfeld, R., Edward, A., Mitchell, D., Lara, E., 2016.
813  Phylogenetic reconstruction based on COI reshuffles the taxonomy of hyalosphenid shelled testate
814  amoebae and reveals the convoluted evolution of shell plate shapes. Cladistics 32, 606-623.

815  Laggoun-Défarge, F., Mitchell, E., Gilbert, D., Disnar, J.R., Comont, L., Warner, B.G., Buttler, A., 2008.
816  Cut-over peatland regeneration assessment using organic matter and microbial indicators (bacteria and
817  testate amoebae). J. Appl. Ecol. 45, 716–727.

818  Lamarre, A., Garneau, M., Asnong, H., 2012. Holocene paleohydrological reconstruction and carbon
819  accumulation of a permafrost peatland using testate amoeba and macrofossil analyses, Kuujjuarapik,
820  subarctic Québec, Canada. Rev. Palaeobot. Palynol. 186, 131–141.

821  Lamarre, A., Magnan, G., Garneau, M., Boucher, É., 2013. A testate amoeba-based transfer function for
822  paleohydrological reconstruction from boreal and subarctic peatlands in northeastern Canada. Quat.
823  Int. 306, 88–96.

824  Lamentowicz, L., Lamentowicz, M., Gałka, M., 2008. Testate amoebae ecology and a local transfer
825  function from a peatland in western Poland. Wetlands 28, 164–175.

826  Lamentowicz, M., Mitchell, E.A.D., 2005. The ecology of testate amoebae (protists) in *Sphagnum* in
827  north-western Poland in relation to peatland ecology. Microb. Ecol. 50, 48–63.

828  Lara, E., Roussel-Delif, L., Fournier, B., Wilkinson, D.M., Mitchell, E.A.D., 2016. Soil microorganisms
829  behave like macroscopic organisms : patterns in the global distribution of soil euglyphid testate
830  amoebae. J. Biogeogr. 43, 520–532.

831  Li, H., Wang, S., Zhao, H., Wang, M., 2015. A testate amoebae transfer function from *Sphagnum*-
832  dominated peatlands in the Lesser Khingan Mountains, NE China. J. Paleolimnol. 54, 189–203.

833  Magnan, G., Garneau, M., 2014. Evaluating long-term regional climate variability in the maritime region
834  of the St. Lawrence North Shore (eastern Canada) using a multi-site comparison of peat-based
835  paleohydrological records. J. Quat. Sci. 29, 209–220.

836  Marcisz, K., Lamentowicz, Ł., Słowińska, S., Słowiński, M., Muszak, W., Lamentowicz, M., 2014. Seasonal
837  changes in Sphagnum peatland testate amoeba communities along a hydrological gradient. Eur. J.
838  Protistol. 50, 445–455.

839  Markel, E.R., Booth, R.K., Yangmin Qin, 2010. Testate amoebae and $^{13}$C of Sphagnum as surface-
840  moisture proxies in Alaskan peatlands. The Holocene 20, 463–475.

841  Mitchell, E.A.D., Charman, D.J., Warner, B.G., 2008. Testate amoebae analysis in ecological and
842  paleoecological studies of wetlands: Past, present and future. Biodivers. Conserv. 17, 2115–2137.

843  Oliverio, A.M., Lahr, D.J.G., Nguyen, T., Katz, L. a., 2014. Cryptic diversity within morphospecies of
844  testate amoebae (Amoebozoa: Arcellinida) in New England bogs and fens. Protist 165, 196–207.

845  O'Malley, M.A., 2008. "Everything is everywhere: but the environment selects": ubiquitous distribution
846  and ecological determinism in microbial biogeography. Stud. Hist. Philos. Sci. Part C Stud. Hist. Philos.
847  Biol. Biomed. Sci. 39, 314–325.

848  O'Malley, M.A., 2007. The nineteenth century roots of "everything is everywhere." Nat. Rev. Microbiol.
849  5, 647–651.

850  Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., Hara, R.B.O., Simpson, G.L., Solymos,
851  P., Stevens, M.H.H., Wagner, H., 2015. Package "vegan" version 2.3-1.

852  Payne, R.J., Babeshko, K. V., van Bellen, S., Blackford, J.J., Booth, R.K., Charman, D.J., Ellershaw, M.R.,
853  Gilbert, D., Hughes, P.D.M., Jassey, V.E.J., Lamentowicz, L., Lamentowicz, M., Malysheva, E.A., Mauquoy,
854  D., Mazei, Y.A., Mitchell, E.A.D., Swindles, G.T., Tsyganov, A., Turner, T.E., Telford, R.J., 2016. Significance
855  testing testate amoeba water table reconstructions. Quat. Sci. Rev. 138, 131–135.

856  Payne, R.J., Kishaba, K., Blackford, J.J., Mitchell, E.A.D., 2006. Ecology of testate amoebae (Protista) in
857  south-central Alaska peatlands: building transfer-function models for palaeoenvironmental studies. The
858  Holocene 16, 403–414.

859  Payne, R.J., Mitchell, E.A.D., 2007. Ecology of Testate Amoebae from Mires in the Central Rhodope
860  Mountains, Greece and Development of a Transfer Function for Palaeohydrological Reconstruction.
861  Protist 158, 159–171.

862  Payne, R.J., Charman, D.J., Matthews, S., Eastwood, W.J., 2008. Testate amoebae as
863  palaeohydrological proxies in Sürmene Ağaçbaşi Yaylasi peatland (northeast Turkey). Wetlands 28,
864  311–323.

865  Payne, R.J., Ryan, P.A., Nishri, A., Gophen, M., 2010. Testate amoeba communities of the drained Hula
866  wetland (Israel): Implications for ecosystem development and conservation management. Wetl. Ecol.
867  Manag. 18, 177–189.

868  Payne, R.J., Lamentowicz, M., Mitchell, E.A.D., 2011. The perils of taxonomic inconsistency in
869  quantitative palaeoecology: Experiments with testate amoeba data. Boreas 40, 15–27.

870  Payne, R.J., Telford, R.J., Blackford, J.J., Blundell, A., Booth, R.K., Charman, D.J., Lamentowicz, M.,
871  Mitchell, E.A.D., Potts, G., Swindles, G.T., Warner, B.G., Lamentowicz, L., Woodland, W.A., 2012. Testing
872  peatland testate amoeba transfer for clustered training-sets. The Holocene 22, 819–825.

873  Pelletier, N., Talbot, J., Olefeldt, D., Turetsky, M.R., Blodau, C., Sonnentag, O., Quinton, W.L., 2017.
874  Influence of Holocene permafrost aggradation and thaw on the paleoecology and carbon storage of a
875  peatland complex in northwestern Canada. The Holocene 27, 1391-1405.

876  Pérez-Juárez, H., Serrano-Vázquez, A., Kosakyan, A., Mitchell, E.A.D., Rivera Aguilar, V.M., Lahr, D.J.G.,
877  Hernández Moreno, M.M., Cuellar, H.M., Eguiarte, L.E., Lara, E., 2017. Quadrulella texcalense sp. nov.
878  from a Mexican desert: An unexpected new environment for hyalospheniid testate amoebae. Eur. J.
879  Protistol. 61, 253–264.

880  Reczuga, M.K., Swindles, G.T., Grewling, Ł., Lamentowicz, M., 2015. Arcella peruviana sp. nov.
881  (Amoebozoa: Arcellinida, Arcellidae), a new species from a tropical peatland in Amazonia. Eur. J.
882  Protistol. 51, 437–449.

883  Roe, H.M., Elliott, S.M., Patterson, R.T., 2017. Re-assessing the vertical distribution of testate amoeba
884  communities in surface peats: implications for palaeohydrological studies. Eur. J. Protistol.

885  Roland, T.P., Amesbury, M.J., Wilkinson, D.M., Charman, D.J., Convey, P., Hodgson, D.A., Royles, J.,
886  Clauß, S., Völcker, E., 2017. Taxonomic implications of morphological complexity within the testate
887  amoeba genus Corythion from the Antarctic Peninsula. Protist 168, 565-585.

888  Singer, D., Kosakyan, A., Seppey, C.V.W., Pillonel, A., Fernández, L.D., Fontaneto, D., Mitchell, E.A.D.,
889  Lara, E., 2018. Environmental filtering and phylogenetic clustering correlate with the distribution
890  patterns of cryptic protist species. Ecology 99, 904–914.

891  Smith, H.G., Bobrov, A., Lara, E., 2008. Diversity and biogeography of testate amoebae. Biodivers.
892  Conserv. 17, 329–343.

893  Smith, H.G., Wilkinson, D.M., 2007. Not all free-living microorganisms have cosmopolitan distributions
894  – the case of Nebela (Apodera) vas Certes (Protozoa : Amoebozoa : Arcellinida). J. Biogeog. 34, 1822–
895  1831.

896  Sullivan, M.E., Booth, R.K., 2011. The potential influence of short-term environmental variability on the
897  composition of testate amoeba communities in Sphagnum peatlands. Microb. Ecol. 62, 80–93.

898  Swindles, G.T., Charman, D.J., Roe, H.M., Sansum, P.A., 2009. Environmental controls on peatland
899  testate amoebae (Protozoa: Rhizopoda) in the North of Ireland: Implications for Holocene palaeoclimate
900  studies. J. Paleolimnol. 42, 123–140.

Swindles, G.T., Lawson, I.T., Matthews, I.P., Blaauw, M., Daley, T.J., Charman, D.J., Roland, T.P., Plunkett, G.M., Schettler, G., Gearey, B.R., Turner, T.E., Rea, H.A., Roe, H.M., Amesbury, M.J., Chambers, F.M., Holmes, J., Mitchell, F.J.G., Blackford, J.J., Blundell, A., Branch, N., Holmes, J., Langdon, P.G., McCarroll, J., McDermott, F., Oksanen, P.O., Pritchard, O., Stastney, P., Stefanini, B., Young, D., Wheeler, J., Becker, K., 2013. Centennial-scale climate change in Ireland during the Holocene. Earth-Science Rev. 126, 300–320.

Swindles, G.T., Reczuga, M., Lamentowicz, M., Raby, C.L., Turner, T.E., Charman, D.J., Gallego-Sala, A., Valderrama, E., Williams, C., Draper, F., Honorio Coronado, E.N., Roucoux, K.H., Baker, T., Mullan, D.J., 2014. Ecology of testate amoebae in an Amazonian peatland and development of a transfer function for palaeohydrological reconstruction. Microb. Ecol. 68, 284–298.

Swindles, G.T., Holden, J., Raby, C.L., Turner, T.E., Blundell, A., Charman, D.J., Menberu, M.W., Kløve, B., 2015a. Testing peatland water-table depth transfer functions using high-resolution hydrological monitoring data. Quat. Sci. Rev. 120, 107–117.

Swindles, G.T., Morris, P.J., Mullan, D., Watson, E.J., Turner, T.E., Roland, T.P., Amesbury, M.J., Kokfelt, U., Schoning, K., Pratte, S., Gallego-Sala, A., Charman, D.J., Sanderson, N., Garneau, M., Carrivick, J., Woulds, C., Holden, J., Parry, L., Galloway, J.M., 2015b. The long-term fate of permafrost peatlands under rapid climate warming. Nat. Sci. Reports. 5:17951. doi: 10.1038/srep17951.

<span style="color:red">Swindles, G.T., Amesbury, M.J., Turner, T.E., Carrivick, J.L., Woulds, C., Raby, C.L., Mullan, D.J., Roland, T.P., Galloway, J.M., Parry, L.E., Kokfelt, U., Garneau, M., Charman, D.J., Holden, J., 2015c. Evaluating the use of testate amoebae for palaeohydrological reconstruction in permafrost peatlands. Palaeogeogr. Palaeoclimatol. Palaeoecol. 424, 111–122.</span>

Talbot, J., Richard, P.J.H., Roulet, N.T., Booth, R.K., 2010. Assessing long-term hydrological and ecological responses to drainage in a raised bog using paleoecology and a hydrosequence. J. Veg. Sci. 21, 143–156.

Telford, R.J., Birks, H.J.B., 2005. The secret assumption of transfer functions: Problems with spatial autocorrelation in evaluating model performance. Quat. Sci. Rev. 24, 2173–2179.

Telford, R.J., Birks, H.J.B., 2009. Evaluation of transfer functions in spatially structured environments. Quat. Sci. Rev. 28, 1309–1316.

Telford, R.J., Birks, H.J.B., 2011a. Effect of uneven sampling along an environmental gradient on transfer-function performance. J. Paleolimnol. 46, 99–106.

Telford, R.J., Birks, H.J.B., 2011b. A novel method for assessing the statistical significance of quantitative reconstructions inferred from biotic assemblages. Quat. Sci. Rev. 30, 1272–1278.

Telford, R.J., 2015. palaeoSig: Significance Tests for Palaeoenvironmental Reconstructions. R Package Version 1.1-3. https://cran.r-project.org/web/packages/palaeoSig/.

Turner, T.E., Swindles, G.T., Charman, D.J., Blundell, A., 2013. Comparing regional and supra-regional transfer functions for palaeohydrological reconstruction from Holocene peatlands. Palaeogeogr. Palaeoclimatol. Palaeoecol. 369, 395–408.

van Bellen, S., Garneau, M., Booth, R.K., 2011. Holocene carbon accumulation rates from three ombrotrophic peatlands in boreal Quebec, Canada: Impact of climate-driven ecohydrological change. Holocene 21, 1217–1231.

van Bellen, S., Mauquoy, D., Payne, R.J., Roland, T.P., Daley, T.J., Hughes, P.D.M., Loader, N.J., Street-Perrott, F.A., Rice, E.M., Pancotto, V. a., Bellen, S.V.A.N., 2014. Testate amoebae as a proxy for reconstructing Holocene water table dynamics in southern Patagonian peat bogs. J. Quat. Sci. 29, 463–474.

944    van Bellen, S., Mauquoy, D., Hughes, P.D., Roland, T.P., Daley, T.J., Loader, N.J., Street-Perrott, F.A., Rice,
945    E.M., Pancotto, V.A., Payne, R.J., 2016. Late-Holocene climate dynamics recorded in the peat bogs of
946    Tierra del Fuego, South America. The Holocene 26, 489–501.

947    Warner, B.G., 1987. Abundance and Diversity of Testate Amoebae (Rhizopoda, Testacea) in Sphagnum
948    Peatlands in Southwestern Ontario, Canada. Arch. fur Protistenkd. 133, 173–189.

949    Warner, B.G., Charman, D.J., 1994. Holocene changes on a peatland in northwestern Ontario
950    interpreted from testate amoebae (Protozoa) analysis. Boreas 23, 270–279.

951    Willis, K.S., Beilman, D., Booth, R.K., Amesbury, M.J., Holmquist, J., MacDonald, G., 2015. Peatland
952    paleohydrology in southern West Siberian Lowlands: Comparison of multiple testate amoeba transfer
953    functions, sites, and Sphagnum $^{13}$C values. The Holocene 25, 1425–1436.

954    Woodland, W.A., Charman, D.J., Sims, P.C., 1998. Quantitative estimates of water tables and soil
955    moisture in Holocene peatlands from testate amoebae. The Holocene 8, 261–273.

956    Zhang, H., Piilo, S.R., Amesbury, M.J., Charman, D.J., Gallego-Sala, A. V., Väliranta, M.M., 2018. The
957    role of climate change in regulating Arctic permafrost peatland hydrological and vegetation change
958    over the last millennium. Quat. Sci. Rev. 182, 121–130.

Ecoregions:
- Tundra
- Taiga
- Hudson plain
- Northern forests
- Northwestern forested mountains
- Marine west coast forest
- Eastern temperate forest
- North American desert
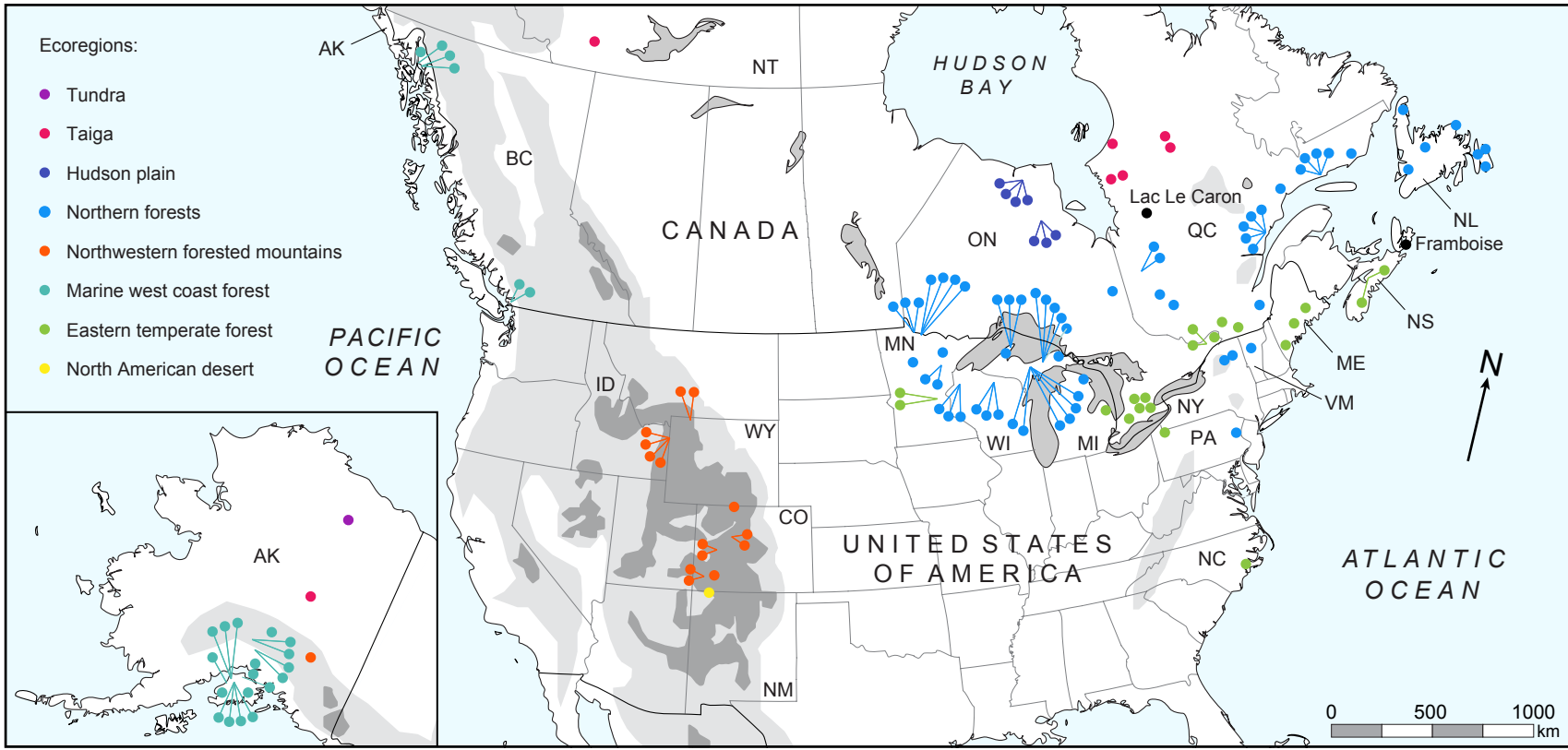
All taxa proportions are % values

1900 1800 1700 1600 1500 1400 1300 1200 1100 1000 900 800 700 600 500 400 300 200 100 0
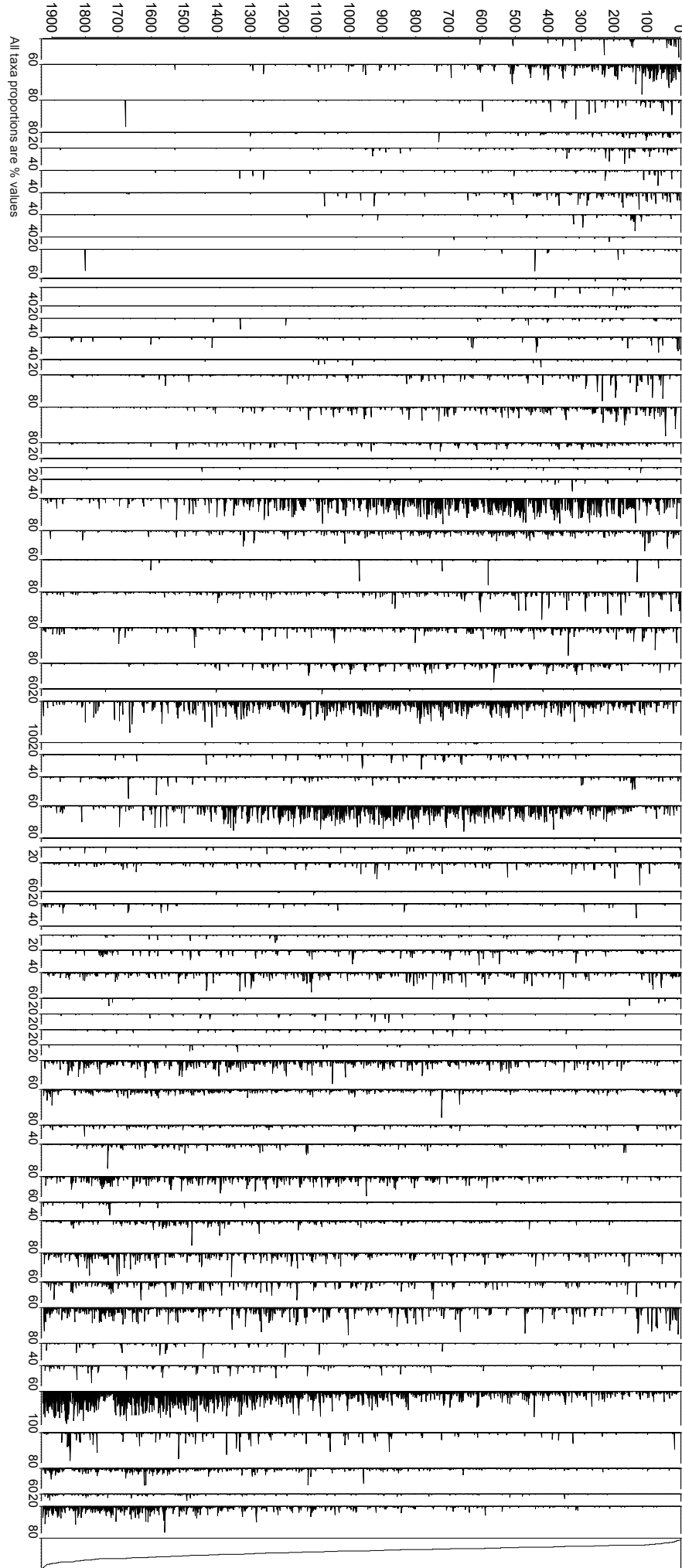
*Difflugia oviformis* type
*Difflugia globulosa* type
*Arcella vulgaris* type
*Difflugia bacilliarium* type
*Planocarina carinata* type
*Difflugia lucida* type
*Amphitrema wrightianum* type
*Difflugia rubescens*
*Difflugia urceolata* type
*Arcella hemispherica* type
*Arcella gibbosa* type
*Difflugia oblonga* type
*Difflugia leidyi*
*Difflugia acuminata* type
*Difflugia pristis* type
*Planocarina marginata* type
*Pseudodifflugia fulva* type
*Arcella discoides* type
*Placocista spinosa*
*Quadrulella symettrica*
*Centropyxis platystoma* type
*Tracheleuglypha dentata*
*Archerella flavum*
*Heleopera sphagni* type
*Difflugia pulex*
*Centropyxis aculeata* type
*Phryganella acropodia* type
*Physochila griseola* type
*Gibbocarina tubulosa* type
*Hyalosphenia papilio*
*Padaungiella wailesi* type
*Argynnia dentistoma* type
*Heleopera petricola* type
*Hyalosphenia elegans*
*Padaungiella lageniformis* type
*Sphenoderia lenta*
*Euglypha strigosa* type
*Nebela barbata*
*Euglypha ciliata* type
*Pseudodifflugia fascicularis* type
*Heleopera rosea*
*Nebela collaris* type
*Centropyxis cassis* type
*Cryptodifflugia sacculus* type
*Nebela minor*
*Nebela penardiana* type
*Nebela flabellulum*
*Nebela tincta* type
*Euglypha tuberculata* type
*Euglypha rotunda* type
*Arcella catinus* type
*Alabasta militaris* type
*Arcella artocrea* type
*Heleopera sylvatica*
*Cyclopyxis arcelloides* type
*Hyalosphenia subflava*
*Corythion - Trinema* type
*Hyalosphenia minuta*
*Centropyxis ecornis* type
*Assulina muscorum* type
*Cryptodifflugia oviformis* type
*Assulina seminulum* type
*Bullinularia indica*
*Trigonopyxis arcula* type
Water-table depth (cm)
(Scale range -20 to 60 cm)

A

B

C
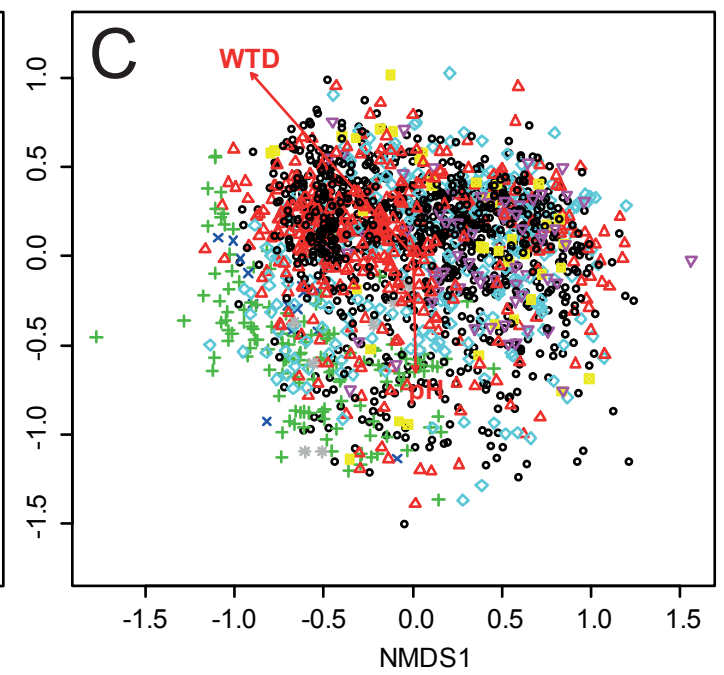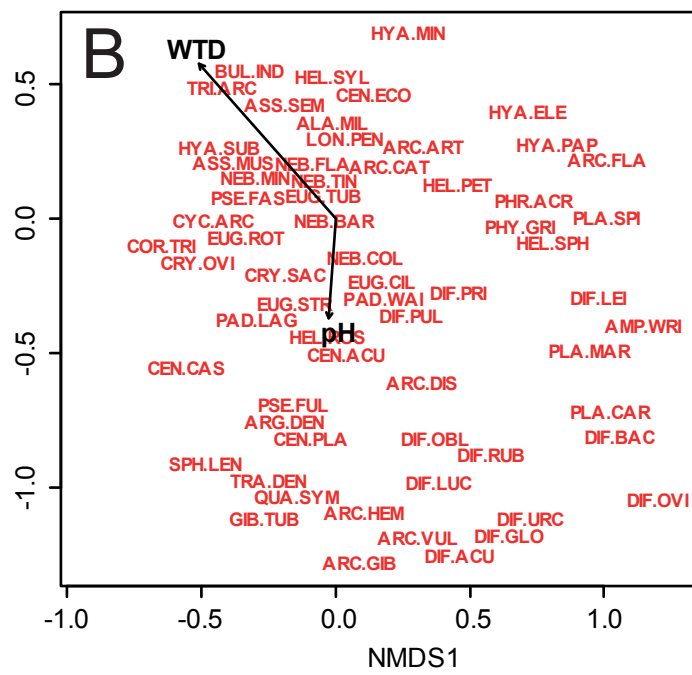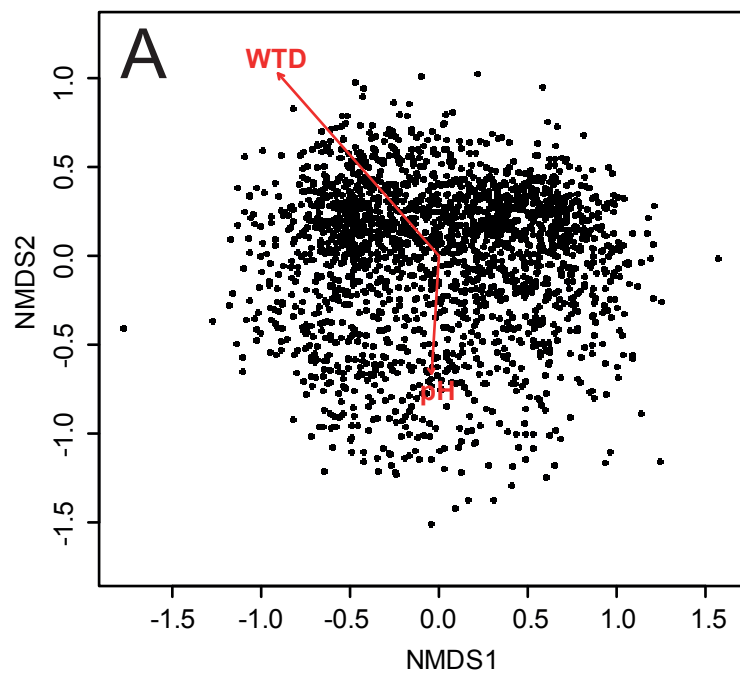
- ○ Northern forests
- △ Eastern temperate forest
- + Northwestern forested mountains
- × North Amercian desert
- ◇ Marine west coast forest
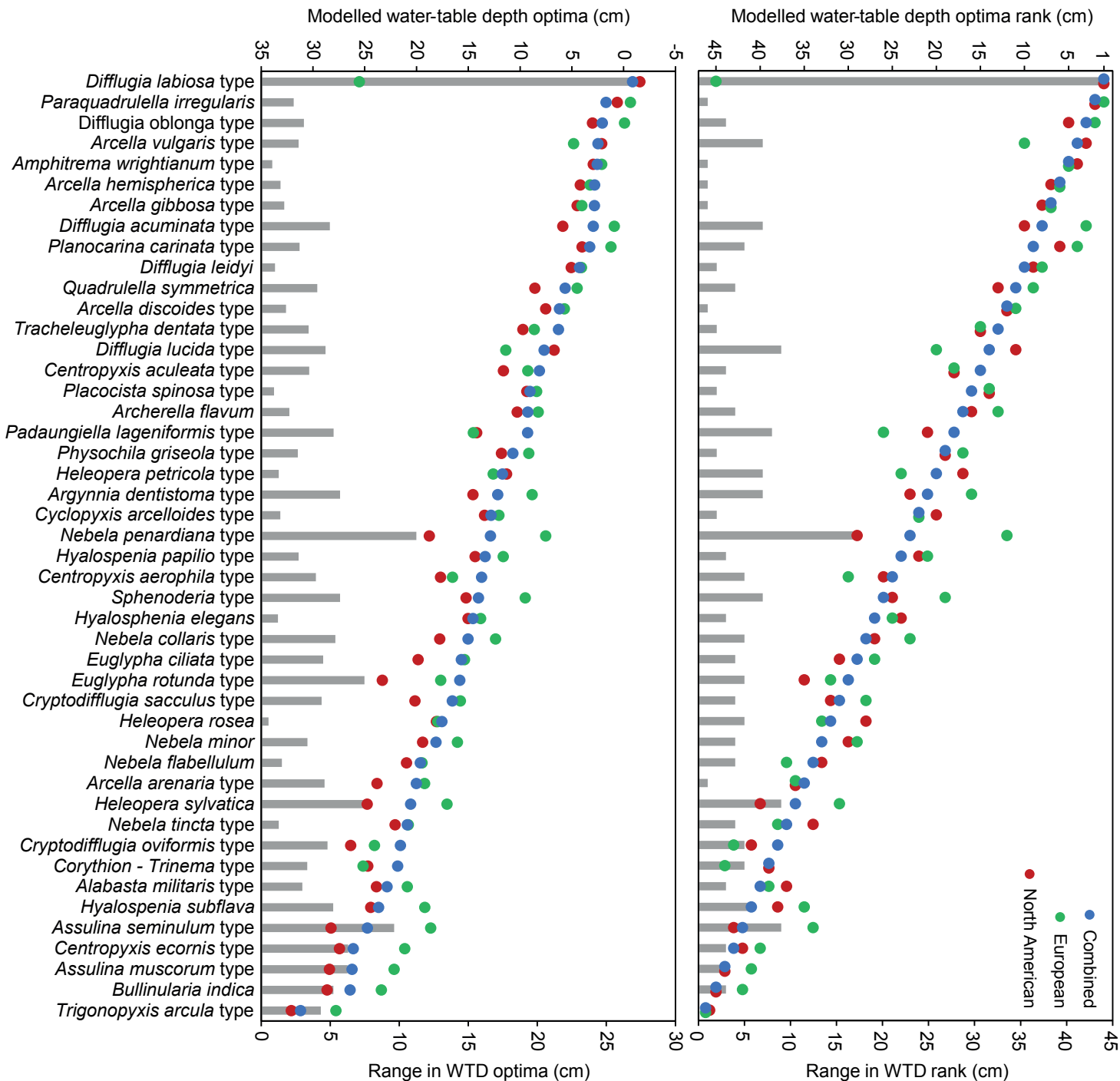- ▽ Taiga
- ■ Hudson plain
- ＊ Tundra

Water table depth (cm)

Legend: ML, WMAT K5, WAPLS C2, WA-Tol (inv), Booth (2008) WA-Tol (inv)

Legend: Taiga — NF — NFM — MWCF — ETF — Pruned WA-Tol (inv) (full dataset)

Legend: — European — North American — Proto-NH

Modelled water-table depth optima (cm)

Modelled water-table depth optima rank (cm)

Range in WTD optima (cm)

Range in WTD rank (cm)

*Difflugia labiosa* type
*Paraquadrulella irregularis*
Difflugia oblonga type
*Arcella vulgaris* type
*Amphitrema wrightianum* type
*Arcella hemispherica* type
*Arcella gibbosa* type
*Difflugia acuminata* type
*Planocarina carinata* type
*Difflugia leidyi*
*Quadrulella symmetrica*
*Arcella discoides* type
*Tracheleuglypha dentata* type
*Difflugia lucida* type
*Centropyxis aculeata* type
*Placocista spinosa* type
*Archerella flavum*
*Padaungiella lageniformis* type
*Physochila griseola* type
*Heleopera petricola* type
*Argynnia dentistoma* type
*Cyclopyxis arcelloides* type
*Nebela penardiana* type
*Hyalospenia papilio* type
*Centropyxis aerophila* type
*Sphenoderia* type
*Hyalosphenia elegans*
*Nebela collaris* type
*Euglypha ciliata* type
*Euglypha rotunda* type
*Cryptodifflugia sacculus* type
*Heleopera rosea*
*Nebela minor*
*Nebela flabellulum*
*Arcella arenaria* type
*Heleopera sylvatica*
*Nebela tincta* type
*Cryptodifflugia oviformis* type
*Corythion - Trinema* type
*Alabasta militaris* type
*Hyalospenia subflava*
*Assulina seminulum* type
*Centropyxis ecornis* type
*Assulina muscorum* type
*Bullinularia indica*
*Trigonopyxis arcula* type

North American
European
Combined

**Table 1:** Site details and meta-data. Data ordered by year of publication/surname and listed alphabetically by site name. For Warner and Charman (1994), sample codes could not be confidently linked to specific sites, so only the total number of samples in the dataset is shown.

| Reference | State/Province, Country | Site name | Lat (°) | Long (°) | Number of samples |
|---|---|---|---|---|---|
| Warner (1987) | Ontario, Canada | Chesney | 43.226 | -80.644 | 4 |
| Warner (1987) | Ontario, Canada | Ellice Huckleberry Marsh | 43.481 | -80.949 | 2 |
| Warner (1987) | Ontario, Canada | Kossuth | 43.449 | -80.359 | 3 |
| Warner (1987) | Ontario, Canada | Sifton | 42.970 | -81.323 | 6 |
| Warner (1987) | Ontario, Canada | Summit Muskeg Preserve | 43.231 | -80.052 | 7 |
| Charman and Warner (1992) | Ontario, Canada | Wally Creek | 49.076 | -80.537 | 108 |
| Warner and Charman (1994) | Minnesota, USA | Beckman Lake | 45.421 | -93.186 | |
| Warner and Charman (1994) | Ontario, Canada | Buriss | 48.666 | -93.632 | |
| Warner and Charman (1994) | Minnesota, USA | Cedar Lake | 45.412 | -93.198 | |
| Warner and Charman (1994) | Ontario, Canada | Emo | 48.702 | -93.825 | |
| Warner and Charman (1994) | Ontario, Canada | Fort Frances Airport | 48.545 | -93.398 | 49 |
| Warner and Charman (1994) | Ontario, Canada | Kehl Lake | 48.579 | -93.576 | |
| Warner and Charman (1994) | Ontario, Canada | Pinewood | 48.754 | -94.290 | |
| Warner and Charman (1994) | Ontario, Canada | Rainy River | 48.758 | -94.533 | |
| Warner and Charman (1994) | Ontario, Canada | Stratton | 48.680 | -94.124 | |
| Booth (2001), Booth (2002) | Michigan, USA | Grand Traverse Bay | 47.166 | -88.255 | 40 |
| Booth (2001), Booth (2002) | Michigan, USA | Tahquamenon | 46.475 | -84.995 | 55 |
| Booth (2002), Booth (2008) | Michigan, USA | Adam's Trail | 46.547 | -86.06 | 21 |
| Booth (2002) | Michigan, USA | Au Train | 46.548 | -86.059 | 23 |
| Booth (2002) | Michigan, USA | Independence | 47.419 | -88.01 | 29 |
| Booth (2002), Booth (2008) | Michigan, USA | Minden | 43.61 | -82.836 | 43 |
| Booth (2002), Booth (2008) | Michigan, USA | North Rhody | 46.583 | -86.073 | 24 |
| Booth (2002) | Michigan, USA | Partridge | 47.42 | -88.008 | 22 |

| | | | | | |
|---|---|---|---|---|---|
| Booth (2002), Booth (2008) | Michigan, USA | South Rhody | 46.565 | -86.075 | 32 |
| Booth (2002) | Michigan, USA | Uncle Sam | 47.419 | -88.016 | 8 |
| Booth (2002) | Michigan, USA | West Adam's Trail | 46.553 | -86.105 | 4 |
| Booth and Zygmunt (2005) | Colorado, USA | Burro Bridge | 37.846 | -107.727 | 8 |
| Booth and Zygmunt (2005) | New Mexico, USA | Chatanooga | 36.961 | -107.693 | 14 |
| Booth and Zygmunt (2005) | Wyoming, USA | East Lily | 44.954 | -109.696 | 10 |
| Booth and Zygmunt (2005) | Colorado, USA | Henderson A | 39.367 | -106.583 | 5 |
| Booth and Zygmunt (2005) | Colorado, USA | Henderson B | 39.354 | -106.603 | 8 |
| Booth and Zygmunt (2005) | Colorado, USA | Keystone | 38.867 | -107.04 | 9 |
| Booth and Zygmunt (2005) | Wyoming, USA | Lilypad | 44.164 | -111.008 | 15 |
| Booth and Zygmunt (2005) | Wyoming, USA | Little Moose | 44.976 | -109.757 | 12 |
| Booth and Zygmunt (2005) | Idaho, USA | Lizard | 44.156 | -111.071 | 5 |
| Booth and Zygmunt (2005) | Wyoming, USA | Robinson | 44.167 | -111.071 | 15 |
| Booth and Zygmunt (2005) | Colorado, USA | Shafer | 40.866 | -106.605 | 9 |
| Booth and Zygmunt (2005) | Colorado, USA | South Mineral | 37.816 | -107.723 | 8 |
| Booth and Zygmunt (2005) | Colorado, USA | Splain's Gulch | 38.834 | -107.076 | 7 |
| Booth and Zygmunt (2005) | Colorado, USA | Wager Gulch | 37.878 | -107.369 | 9 |
| Booth and Zygmunt (2005) | Idaho, USA | West Robinson | 44.167 | -111.075 | 5 |
| Payne et al. (2006) | Alaska, USA | Bicentennial Park | 61.160 | -149.743 | 28 |
| Payne et al. (2006) | Alaska, USA | Clam Gulch | 60.239 | -151.375 | 3 |
| Payne et al. (2006) | Alaska, USA | Houston | 61.640 | -149.863 | 9 |
| Payne et al. (2006) | Alaska, USA | Jigsaw Lake | 60.739 | -150.490 | 18 |
| Payne et al. (2006) | Alaska, USA | Kachemak | 59.792 | -151.169 | 8 |
| Payne et al. (2006) | Alaska, USA | Moose Pass | 60.509 | -149.442 | 8 |
| Payne et al. (2006) | Alaska, USA | Sheep Creek | 62.003 | -150.055 | 9 |
| Payne et al. (2006) | Alaska, USA | Sterling | 60.528 | -150.518 | 8 |
| Booth (2008) | Michigan, USA | Big Red Pine | 46.253 | -86.623 | 10 |
| Booth (2008) | New York, USA | Bloomingdale | 44.383 | -74.139 | 19 |

| | | | | | |
|---|---|---|---|---|---|
| Booth (2008) | Wisconsin, USA | Crystal | 46.007 | -89.605 | 13 |
| Booth (2008) | Minnesota, USA | Ely | 47.446 | -92.436 | 9 |
| Booth (2008) | Wisconsin, USA | Falliston | 45.995 | -89.613 | 15 |
| Booth (2008) | Michigan, USA | Grimes | 46.276 | -86.65 | 10 |
| Booth (2008) | Michigan, USA | Herron | 45.102 | -83.639 | 10 |
| Booth (2008) | Michigan, USA | Hole-in-bog | 47.3 | -94.249 | 14 |
| Booth (2008) | Wisconsin, USA | Hornet | 46.077 | -92.12 | 15 |
| Booth (2008) | Michigan, USA | Island Camp | 46.272 | -86.647 | 11 |
| Booth (2008) | Michigan, USA | Kentucky Trail | 46.247 | -86.601 | 10 |
| Booth (2008) | Maine, USA | Orono | 44.871 | -68.727 | 18 |
| Booth (2008) | Vermont, USA | Peacham | 44.292 | -72.873 | 12 |
| Booth (2008) | Michigan, USA | Small Red Pine | 46.238 | -86.617 | 5 |
| Booth (2008) | Wisconsin, USA | Small Sarracenia | 46.093 | -92.115 | 10 |
| Booth (2008) | New York, USA | Spring Pond | 44.368 | -74.499 | 26 |
| Booth (2008) | Wisconsin, USA | Spruce | 46.095 | -92.112 | 8 |
| Booth (2008) | Minnesota, USA | Toivola A | 46.927 | -92.795 | 13 |
| Booth (2008) | Minnesota, USA | Toivola B | 46.986 | -92.888 | 6 |
| Booth (2008) | Wisconsin, USA | Trout Lake | 46.004 | -89.594 | 68 |
| Booth (2008) | Michigan, USA | Upper Twin | 46.276 | -86.647 | 10 |
| Booth et al. (2008) | North Carolina, USA | Catfish Pocosin | 34.919 | -77.088 | 35 |
| Markel et al. (2010) | Alaska, USA | Brown's Lake | 60.471 | -150.728 | 38 |
| Markel et al. (2010) | Alaska, USA | Canaday | 64.519 | -146.911 | 8 |
| Markel et al. (2010) | Alaska, USA | Caswell | 62.004 | -150.060 | 5 |
| Markel et al. (2010) | Alaska, USA | Funny River | 60.415 | -150.901 | 8 |
| Markel et al. (2010) | Alaska, USA | Gate Creek | 62.333 | -150.545 | 8 |
| Markel et al. (2010) | Alaska, USA | Jake Lake | 62.430 | -150.704 | 7 |
| Markel et al. (2010) | Alaska, USA | Gas Field | 60.445 | -151.246 | 24 |
| Markel et al. (2010) | Alaska, USA | Marathon B | 60.603 | -151.211 | 5 |

| | | | | | |
|---|---|---|---|---|---|
| Markel et al. (2010) | Alaska, USA | No Name | 60.640 | -151.08 | 5 |
| Markel et al. (2010) | Alaska, USA | Parks Highway | 63.081 | -149.527 | 7 |
| Markel et al. (2010) | Alaska, USA | Richardson | 62.604 | -145.448 | 3 |
| Markel et al. (2010) | Alaska, USA | Teardrop Lake | 60.472 | -150.700 | 6 |
| Sullivan and Booth (2011) | Pennsylvania, USA | Tannersville | 41.037 | -75.266 | 47 |
| Sullivan and Booth (2011) | Pennsylvania, USA | Titus | 41.952 | -79.759 | 20 |
| Amesbury et al. (2013) | Newfoundland, Canada | Burnt Village | 51.126 | -55.927 | 25 |
| Amesbury et al. (2013) | Nova Scotia, Canada | Colin | 45.108 | -63.895 | 27 |
| Amesbury et al. (2013), Charman and Warner (1997) | Newfoundland, Canada | Crooked | 49.139 | -56.111 | 14 |
| Amesbury et al. (2013) | Quebec, Canada | Kegaska | 50.211 | -61.25 | 4 |
| Amesbury et al. (2013), Lamarre et al. (2013) | Quebec, Canada | Lebel | 49.123 | -68.292 | 19 |
| Amesbury et al. (2013), Lamarre et al. (2013) | Quebec, Canada | Manic | 49.123 | -68.292 | 30 |
| Amesbury et al. (2013) | Newfoundland, Canada | Nordan's Pond | 49.161 | -53.598 | 20 |
| Amesbury et al. (2013), Charman and Warner (1997) | Newfoundland, Canada | Ocean Pond | 47.427 | -53.438 | 9 |
| Amesbury et al. (2013) | Nova Scotia, Canada | Petite | 45.144 | -63.937 | 16 |
| Amesbury et al. (2013), Lamarre et al. (2013) | Quebec, Canada | Plaine | 50.27 | -63.54 | 21 |
| Amesbury et al. (2013) | Maine, USA | Saco | 43.553 | -70.471 | 26 |
| Amesbury et al. (2013), Charman and Warner (1997) | Newfoundland, Canada | Sam's River | 46.731 | -53.519 | 6 |
| Amesbury et al. (2013) | Quebec, Canada | Sept Isles | 50.202 | -66.601 | 6 |
| Amesbury et al. (2013) | Maine, USA | Sidney | 44.388 | -69.788 | 24 |
| Amesbury et al. (2013), Charman and Warner (1997) | Newfoundland, Canada | Stephenville | 48.55 | -58.434 | 6 |
| Amesbury et al. (2013) | Quebec, Canada | Tourbiere de l'Aeroport | 50.275 | -63.614 | 7 |
| Amesbury et al. (2013), Charman and Warner (1997) | Newfoundland, Canada | Witless Bay | 47.337 | -53.017 | 4 |
| Lamarre et al. (2013) | Quebec, Canada | Abeille | 54.115 | -73.327 | 6 |
| Lamarre et al. (2013) | Quebec, Canada | Baie-M | 49.093 | -68.243 | 11 |
| Lamarre et al. (2013) | Quebec, Canada | Baie Commeau | 49.097 | -68.244 | 7 |
| Lamarre et al. (2013) | Quebec, Canada | Frontenac | 45.966 | -71.139 | 10 |
| Lamarre et al. (2013) | Quebec, Canada | Kuujuarapik | 55.226 | -77.696 | 10 |

| | | | | | |
|---|---|---|---|---|---|
| Lamarre et al. (2013) | Quebec, Canada | La Grande Bog | 53.65 | -77.726 | 10 |
| Lamarre et al. (2013) | Quebec, Canada | La Grande Fen | 53.677 | -78.216 | 10 |
| Lamarre et al. (2013) | Quebec, Canada | Lebel-G | 49.101 | -68.219 | 10 |
| Lamarre et al. (2013) | Quebec, Canada | Matagami 1 | 49.752 | -77.657 | 6 |
| Lamarre et al. (2013) | Quebec, Canada | Matagami 2 | 49.685 | -77.732 | 10 |
| Lamarre et al. (2013) | Quebec, Canada | Morts | 50.264 | -63.669 | 17 |
| Lamarre et al. (2013) | Quebec, Canada | Pylone | 53.797 | -73.327 | 3 |
| Lamarre et al. (2013) | Quebec, Canada | Roma | 50.298 | -63.693 | 25 |
| Lamarre et al. (2013) | Quebec, Canada | Val d'Or | 48.187 | -77.595 | 10 |
| Lamarre et al. (2013) | Quebec, Canada | Verendrye | 47.303 | -76.855 | 4 |
| Roe et al. (2017) | Ontario, Canada | Alfred | 45.494 | -74.805 | 47 |
| Roe et al. (2017) | Ontario, Canada | Mirabel | 45.685 | -74.045 | 40 |
| Roe et al. (2017) | Quebec, Canada | Mer Bleue | 45.402 | -75.486 | 58 |
| Clifford (unpublished) | British Colombia, Canada | Burns | 49.123 | -122.970 | 12 |
| Clifford (unpublished) | British Colombia, Canada | Eagle Creek | 58.284 | -134.521 | 12 |
| Clifford (unpublished) | British Colombia, Canada | Fish Creek | 58.336 | -134.554 | 12 |
| Clifford (unpublished) | British Colombia, Canada | Point Bridget | 58.647 | -134.934 | 12 |
| Clifford (unpublished) | British Colombia, Canada | Point Lena | 58.386 | -134.746 | 12 |
| Clifford (unpublished) | British Colombia, Canada | Surrey | 49.204 | -122.747 | 12 |
| Finkelstein/Bunbury (unpublished) | Ontario, Canada | Attawapiskat River 1 | 52.845 | -83.927 | 16 |
| Finkelstein/Bunbury (unpublished) | Ontario, Canada | Attawapiskat River 2 | 52.720 | -83.940 | 11 |
| Finkelstein/Bunbury (unpublished) | Ontario, Canada | Attawapiskat River 3 | 52.712 | -84.174 | 5 |
| Finkelstein/Bunbury (unpublished) | Ontario, Canada | Sutton River 1 | 55.752 | -84.525 | 2 |
| Finkelstein/Bunbury (unpublished) | Ontario, Canada | Sutton River 2 | 54.676 | -84.599 | 5 |
| Finkelstein/Bunbury (unpublished) | Ontario, Canada | Sutton River 3 | 54.610 | -84.609 | 5 |
| Finkelstein/Bunbury (unpublished) | Ontario, Canada | Sutton River 4 | 54.599 | -84.640 | 3 |
| Swindles (unpublished) | Alaska, USA | Toolik | 68.624 | -149.580 | 7 |
| Talbot (unpublished) | Ontario, Canada | Folly | 45.272 | -75.466 | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Talbot (unpublished) | | Quebec, Canada | Mont St. Bruno | 45.335 | -73.215 | 2 |
| Talbot (unpublished) | | Northwest Territories, Canada | Scotty Creek | 61.307 | -121.298 | 21 |
| | | | | | **TOTAL:** | **1956** |

**Table 2:** Full taxonomy applied in this study.

| # | Abbreviation | Full name | Also includes |
|---|---|---|---|
| 1 | ALA MIL | Alabasta (Nebela) militaris type | |
| 2 | AMP WRI | Amphitrema wrightianum type | Amphitrema stenostoma |
| 3 | ARC ART | Arcella artocrea | |
| 4 | ARC CAT | Arcella catinus type | Arcella arenaria type |
| 5 | ARC DIS | Arcella discoides type | |
| 6 | ARC GIB | Arcella gibbosa type | |
| 7 | ARC HEM | Arcella hemispherica type | Arcella rotundata |
| 8 | ARC VUL | Arcella vulgaris type | Arcella crenulata |
| 9 | ARC FLA | Archerella flavum | |
| 10 | ARG DEN | Argynnia dentistoma type | Nebela dentistoma<br>Nebela vitraea type |
| 11 | ASS MUS | Assulina muscorum | |
| 12 | ASS SEM | Assulina seminulum type | Assulina scandinavica |
| 13 | BUL IND | Bullinularia indica | |
| 14 | CEN ACU | Centropyxis aculeata type | |
| 15 | CEN CAS | Centropyxis cassis type | Centropyxis aerophila type<br>Centropyxis constricta type |
| 16 | CEN PLA | Centropyxis platystoma type | |
| 17 | CEN ECO | Centropyxis ecornis type | Centropyxis delicatula type<br>Centropyxis laevigata type |
| 18 | CON PON | Conicocassis pontigulasiformis | |
| 19 | COR TRI | Corythion – Trinema type | Corythion dubium |

| | | | Corythion complanatum<br>Trinema enchelys<br>Trinema lineare<br>Trinema sp. |
|---|---|---|---|
| 20 | CRY OVI | Cryptodifflugia oviformis type | |
| 21 | CRY SAC | Cryptodifflugia sacculus | |
| 22 | CYC ARC | Cyclopyxis arcelloides type | Cyclopyxis kahli type |
| 23 | CYP AMP | Cyphoderia ampulla type | Cyphoderia calceolus |
| 24 | DIF ACU | Difflugia acuminata type | Difflugia bacillariarum type |
| 25 | DIF BAC | Difflugia bacillifera type | |
| 26 | DIF GLO | Difflugia globulosa type | |
| 27 | DIF LEI | Difflugia leidyi | |
| 28 | DIF LUC | Difflugia lucida type | |
| 29 | DIF OBL | Difflugia oblonga type | Difflugia lanceolata |
| 30 | DIF OVI | Difflugia oviformis type | Difflugia lithophila type<br>Netzelia (Difflugia) tuberculata type |
| 31 | DIF PRI | Difflugia pristis type | |
| 32 | DIF PUL | Difflugia pulex type | |
| 33 | DIF RUB | Difflugia rubescens type | |
| 34 | DIF URC | Difflugia urceolata type | |
| 35 | EUG CIL | Euglypha ciliata type | Euglypha compressa<br>Euglypha cristata |
| 36 | EUG ROT | Euglypha rotunda type | |
| 37 | EUG STR | Euglypha strigosa type | |
| 38 | EUG TUB | Euglypha tuberculata type | |
| 39 | GIB TUB | Gibbocarina (Nebela) tubulosa type | |
| 40 | HEL ROS | Heleopera rosea | |
| 41 | HEL PET | Heleopera petricola | |
| 42 | HEL SPH | Heleopera sphagni | |
| 43 | HEL SYL | Heleopera sylvatica | |

| 44 | HYA ELE | Hyalosphenia elegans | |
|---|---|---|---|
| 45 | HYA MIN | Hyalosphenia minuta | |
| 46 | HYA PAP | Hyalosphenia papilio | |
| 47 | HYA SUB | Hyalosphenia subflava | |
| 48 | LES EPI | Lesquereusia epistomum | |
| 49 | LES MOD | Lesquereusia modesta type | Lesquereusia spiralis type |
| 50 | NEB BAR | Nebela barbata | |
| 51 | NEB COL | Nebela collaris type | Hyalosphenia ovalis |
| 52 | NEB FLA | Nebela flabellulum | |
| 53 | NEB MIN | Nebela minor | |
| 54 | NEB PEN | Nebela penardiana type | |
| 55 | NEB TIN | Nebela tincta type | Nebela bohemica<br>Nebela parvula |
| 56 | PAD LAG | Padaungiella (Nebela) lageniformis | |
| 57 | PAD WAI | Padaungiella (Nebela) wailesi | |
| 58 | PAR IRR | Paraquadrula irregularis | Paraquadrula sp. |
| 59 | PHR ACR | Phryganella acropodia type | |
| 60 | PHY GRI | Physochila griseola type | (Nebela griseola type) |
| 61 | PLA CAL | Plagiopyxis callida type | |
| 62 | PLA SPI | Placocista spinosa type | Placocista lens |
| 63 | PLA CAR | Planocarina (Nebela) carinata type | Gibbocarina (Nebela) galeata type |
| 64 | PLA MAR | Planocarina (Nebela) marginata type | |
| 65 | PSE FUL | Pseudodifflugia fulva type | |
| 66 | PSE FAS | Pseudodifflugia fasicularis type | |
| 67 | PYXI | Pyxidicula type | Pyxidicula patens<br>Pyxidicula sp. |
| 68 | QUA SYM | Quadrulella symmetrica | |
| 69 | SPH LEN | Sphenoderia lenta | |
| 70 | TRA DEN | Tracheleuglypha dentata | |

| 71 | TRI ARC | Trigonopyxis arcula type | Trigonopyxis minuta type |
|----|---------|--------------------------|--------------------------|
| 72 | WAI EBO | Waileseila eboracensis   |                          |

**Table 3:** Numbers of sites, samples and taxa in different EPA eco-regions, along with information on water table depth (WTD) values for all samples within each region and basic performance statistics for eco-region transfer functions derived from the pruned WA-Tol (inv) model dataset (n = 1696). Comparable data for all samples is shown in top row for reference. SD = standard deviation. SWDI = Shannon Wiener diversity index. Dashed lines indicate where no performance statistics are available as models were not run for all eco-regions due to low sample sizes.

| Ecoregion | Number of samples | Number of sites | Number of taxa | Mean WTD ± 1 SD (cm) | WTD range (min – max) (cm) | SWDI ± 1 SD | RMSEP$_{LOO}$ | $R^2$ | Maximum bias | Average bias |
|-----------|-------------------|-----------------|----------------|----------------------|----------------------------|-------------|---------------|-------|--------------|--------------|
| *ALL SAMPLES* | *1696* | *126* | *64* | *18.3 ± 14.1* | *72 (-15 – 57)* | *1.95 ± 0.44* | *7.42* | *0.72* | *13.87* | *0.02* |
| Tundra | 7 | 1 | 25 | 11.7 ± 10 | 29 (1 – 30) | 1.96 ± 0.24 | - | - | - | - |
| Taiga | 64 | 6 | 56 | 12 ± 10.1 | 49 (-3 – 46) | 2.01 ± 0.49 | 8.06 | 0.36 | 32.52 | 0.71 |
| Hudson plain | 32 | 7 | 44 | 15 ± 14.2 | 47 (-4 – 43) | 1.90 ± 0.43 | - | - | - | - |
| Northern forests | 915 | 55 | 62 | 18.2 ± 14.2 | 71 (-15 – 56) | 2.01 ± 0.41 | 6.90 | 0.76 | 15.53 | 0.46 |
| Northwestern forested mountains | 130 | 16 | 42 | 18.4 ± 15.2 | 55 (-5 – 50) | 1.70 ± 0.55 | 6.87 | 0.80 | 13.89 | 0.29 |
| Marine west coast forest | 191 | 22 | 59 | 18.1 ± 13.7 | 69 (-15 – 54) | 1.91 ± 0.36 | 8.30 | 0.63 | 18.37 | 0.19 |
| Eastern temperate forests | 346 | 18 | 62 | 20 ± 13.8 | 65 (-8 – 57) | 1.94 ± 0.46 | 7.12 | 0.73 | 15.96 | 0.05 |
| North American deserts | 11 | 1 | 18 | 29.5 ± 17 | 50 (0 – 50) | 1.47 ± 0.62 | - | - | - | - |

**Table 4:** Performance statistics for all transfer function models for water-table depth based on leave-one-out (RMSEP$_{LOO}$), leave-one-site-out (RMSEP$_{LOSO}$) and segment-wise (RMSEP$_{SW}$) cross validation methods. Results are in order of performance as assessed by RMSEP$_{LOO}$ of the pruned model run . Results for RMSEP$_{LOO}$ are given for both pre- and post-removal of samples with high residual values (see text for details; figures in parentheses show performance statistics after the removal of outlier samples). Results for RMSEP$_{LOSO}$ and RMSEP$_{SW}$ are only given for pruned models. Relative decreases in model performance are between RMSEP$_{LOO}$ of the pruned models to RMSEP$_{LOSO}$ and RMSEP$_{SW}$. SD is the standard deviation of all water-table measurements included in each model after the removal of outliers. Results are only shown for the best performing version of each model type. WA-Tol (inv) = weighted averaging with tolerance downweighting and inverse deshrinking; WAPLS C2 = second component of weighted averaging partial least squares; ML = maximum likelihood; WMAT K5 = weighted averaging modern analogue technique with five nearest neighbours.

| Model type | RMSEP$_{LOO}$ | R$^2$ (LOO) | Maximum bias (LOO) | Average bias (LOO) | RMSEP$_{LOSO}$ | Relative decrease LOO - LOSO | RMSEP$_{SW}$ | Relative decrease LOO - SW | Number of outlier samples removed | n for post-outlier removal model | SD |
|---|---|---|---|---|---|---|---|---|---|---|---|
| WMAT K5 | 9.38 (6.83) | 0.64 (0.79) | 22.97 (10.84) | -0.79 (-0.73) | 8.21 | 0.201 | 7.07 | 0.034 | 161 | 1767 | 14.85 |
| WA-Tol (inv) | 10.20 (7.42) | 0.57 (0.72) | 28.48 (13.87) | 0.03 (0.02) | 7.70 | 0.037 | 8.06 | 0.086 | 197 | 1730 | 14.10 |
| WAPLS C2 | 10.10 (7.53) | 0.58 (0.71) | 27.38 (14.38) | 0.02 (0.02) | 7.85 | 0.042 | 8.18 | 0.086 | 190 | 1737 | 14.11 |
| ML | 14.76 (8.60) | 0.53 (0.77) | 10.31 (6.95) | -3.00 (-1.18) | 8.67 | 0.008 | 8.75 | 0.017 | 410 | 1517 | 15.72 |

**Table 5:** PalaeoSig p values for the significance of reconstructions against models based on randomly generated data for the pruned versions of four model types and five eco-region models, against two independent test sets. See Tables 3 and 4 for further statistics on all models.

| Model | Framboise | Lac Le Caron |
|---|---|---|
| WA-Tol (inv) | 0.013 | 0.987 |
| WAPLS C2 | 0.015 | 0.984 |
| WMAT K5 | 0.002 | 0.709 |
| ML | 0.057 | 0.287 |
| Taiga | 0.485 | 0.943 |
| Northern forests | 0.015 | 0.729 |
| Northwestern forested mountains | 0.782 | 0.997 |
| Marine west coast forest | 0.471 | 0.871 |
| Eastern temperate forests | 0.007 | 0.940 |

**Supplementary figure captions**

**Figure S1:** Summary results for the calculation of segment-wise RMSEP. Histograms show sampling distribution in 12 segments for the full dataset (n = 1927). Coloured lines show RMSEP for each individual segment for the four main model types under investigation.

**Figure S2:** Spatial autocorrelation diagrams for the four model types under investigation. Plots show effect on $r^2$ of deleting sites at random (open circles), from the geographical neighbourhood of the test site (filled circles labelled with corresponding distances) or that are most environmentally similar (crosses) during cross-validation. Note variable y-axis scales.

**Figure S3:** Water table depth reconstructions, with error, from two independent test sites for four different model types and the WA-Tol (inv) model of Booth (2008). Left-hand column is for Framboise bog, right-hand column is for Lac Le Caron peatland.

**Figure S4:** Water table depth reconstructions, with error, from two independent test sites for five different eco-region models, compared to the pruned WA-Tol (inv) model based on the full dataset. Left-hand column is for Framboise bog, right-hand column is for Lac Le Caron peatland. Model abbreviations in key: NF = Northern forests; NFM = Northwestern forested mountains; MWCF = Marine west coast forest; ETF = Eastern temperate forests.

**Figure S5:** Distribution of all taxa (ordered alphabetically) across eight eco-regions. Eco-regions codes (x-axes): 1 = Tundra, 2 = Taiga, 3 = Hudson Plain, 4 = Northern forests, 5 = Northwestern forested mountains, 6 = Marine west coast forest, 7 = Eastern temperate forests, 8 = North American deserts. Y-axes show % abundance of the named taxon in all samples in which it occurs. Number of occurrences in 1884 samples shown in brackets after each taxon code. For taxa name abbreviations, see Table 2.

**Figure S6:** Comparison of occurrence (i.e. % presence in all samples; black bars), abundance (i.e. maximum % abundance in any one sample; grey bars) and presence in n eco-regions (red line) for all 64 taxa included in the full dataset.

**Figure S7:** Comparison of water table depth optima (cm) for 64 taxa from the full dataset and five eco-region models (see text for details), ordered by the full dataset. Coloured spots represent optima of individual taxa in different models, grey bars represent range between the highest and lowest optima for each taxon. For references to colour, readers are referred to the online version of the article.

% presence in all samples
% maximum abundance in any sample
Present in n eco-regions

Modelled water-table depth optima (cm)

Range in water-table depth optima (cm)

**Table S1:** Rare taxa removed from complete dataset, occurring in <1% of the total number of samples (i.e. <19).

| Name | Abbreviation | Number of occurrences | Maximum abundance (%) |
|---|---|---|---|
| Conicocassis pontigulasiformis | CON PON | 2 | 1.5 |
| Cyphoderia ampulla | CYP AMP | 8 | 1.9 |
| Lesquereusia epistomum | LES EPI | 1 | 0.5 |
| Lesquereusia modesta type | LES MOD | 6 | 12.4 |
| Paraquadrula irregularis | PAR IRR | 4 | 13.4 |
| Plagiopyxis callida type | PLAGIO | 2 | 1.1 |
| Pyxidicula sp. | PYXI | 9 | 10.1 |
| Waileseila eboracensis | WAI EBO | 2 | 0.8 |

**Table S2:** Samples removed where taxonomic merging or deletion of results only identified to species level resulted in a percentage total <90%.

| Reference | Sample ID | Site | Remaining % |
|---|---|---|---|
| Finkelstein/Bunbury (unpublished) | M4 | Attawapiskat River 1 | 89.1 |
| Finkelstein/Bunbury (unpublished) | M17 | Attawapiskat River 1 | 87.5 |
| Finkelstein/Bunbury (unpublished) | M23 | Attawapiskat River 1 | 85.6 |
| Finkelstein/Bunbury (unpublished) | M24 | Attawapiskat River 1 | 86.3 |
| Finkelstein/Bunbury (unpublished) | M25 | Attawapiskat River 1 | 79.8 |
| Finkelstein/Bunbury (unpublished) | M28 | Attawapiskat River 1 | 88.9 |
| Finkelstein/Bunbury (unpublished) | M11 | Attawapiskat River 2 | 89.0 |
| Finkelstein/Bunbury (unpublished) | M42 | Sutton River 2 | 88.2 |
| Charman and Warner (1992) | 1142 | Wally Creek | 89.2 |
| Charman and Warner (1992) | 1212 | Wally Creek | 88.4 |
| Charman and Warner (1992) | 12C12 | Wally Creek | 84.3 |

**Table S3:** Samples removed with extreme water-table depths (WTDs) (i.e. those below the 0.5$^{th}$ or above the 99.5$^{th}$ percentiles).

| Reference | Sample ID | Site | WTD (cm) |
|---|---|---|---|
| Warner and Charman (1994) | NWO-BB | Unknown | 82.0 |
| Warner and Charman (1994) | NWO-N | Unknown | 68.0 |
| Warner and Charman (1994) | NWO-D | Unknown | 100.0 |
| Warner and Charman (1994) | NWO-E | Unknown | 100.0 |
| Warner and Charman (1994) | NWO-F | Unknown | 100.0 |
| Warner and Charman (1994) | NWO-G | Unknown | 74.0 |
| Roe et al. (2017) | MER17 | Mer Bleue | 82.0 |
| Roe et al. (2017) | ALF18 | Alfred | 72.0 |
| Sullivan and Booth (2011) | TA28 | Tannersville | -16.5 |
| Markel et al. (2010) | AK93 | Brown's Lake | -17.0 |
| Markel et al. (2010) | AK84 | Brown's Lake | -16.5 |
| Markel et al. (2010) | AK48 | Gas field | -18.0 |
| Booth (2002) | 1020B | Au Train Bay | -27.0 |
| Booth (2001) | 967ada | Tahquamenon Bay | -21.0 |
| Booth (2001) | 968ada | Tahquamenon Bay | -26.0 |
| Booth (2001) | 832ada | Grand Traverse Bay | -21.0 |
| Booth (2001) | 832bda | Grand Traverse Bay | -37.0 |
| Booth (2008) | AL06 | Herron | 75.0 |

**Table S4:** Rare taxa removed from the combined North American and European dataset, occurring in <1% of the total number of samples (i.e. <37).

| Name | Abbreviation | Number of occurrences | Maximum abundance (%) |
|---|---|---|---|
| Arcella mitrata | ARC MIT | 2 | 0.9 |
| Campascus minutus | CAM MIN | 11 | 19.7 |
| Conicocassis pontigulasiformis | CON PON | 2 | 1.5 |
| Difflugia gramen type | DIF GRA | 24 | 4.1 |
| Lagenodifflugia vas | LAG VAS | 8 | 3.7 |
| Lesquereusia epistomum | LES EPI | 15 | 3.3 |
| Plagiopyxis type | PLAGIO | 10 | 11.0 |
| Pontigulasia elisa | PON ELI | 1 | 0.7 |
| Pseudodifflugia gracilis | PSE GRA | 3 | 2.2 |
| Pyxidicula type | PYXI | 31 | 12.1 |
| Waileseila eboracensis | WAI EBO | 2 | 0.8 |

**Table S5:** Samples removed from the combined North American and European dataset where taxonomic merging resulted in a percentage total <90%.

| Reference | Sample ID | Site | Remaining % |
|---|---|---|---|
| Bobrov et al. (1999) | KATIN_6 | Katin Moch | 53.4 |
| Lamentowicz et al. (2013);  Jassey et al. (2014) | WEK14 | Wierzchołek | 53.8 |
| Lamentowicz et al. (2008) | CHL23 | Chlebowo | 59.1 |
| Lamentowicz et al. (2013);  Jassey et al. (2014) | WEK6 | Wierzchołek | 59.6 |
| Lamentowicz et al. (2013);  Jassey et al. (2014) | WEK11 | Wierzchołek | 59.8 |
| Lamentowicz et al. (2008) | CHL22 | Chlebowo | 68.8 |
| Finkelstein/Bunbury (unpublished) | M25 | Attawapiskat River 1 | 69.7 |

| Lamentowicz et al. (2013);  Jassey et al. (2014) | WEK2 | Wierzchołek | 70.2 |
|---|---|---|---|
| Lamentowicz et al. (2008) | CHL7 | Chlebowo | 72.6 |
| Lamentowicz et al. (2013);  Jassey et al. (2014) | RUR23 | Rurzyca | 75.6 |
| Lamentowicz et al. (2010) | 72 | Lej Nair | 75.8 |
| Lamentowicz et al. (2010) | 71 | Lej Nair | 78.1 |
| Finkelstein/Bunbury (unpublished) | M23 | Attawapiskat River 1 | 78.8 |
| Lamentowicz et al. (2013);  Jassey et al. (2014) | WEK13 | Wierzchołek | 79.0 |
| Lamentowicz et al. (2013);  Jassey et al. (2014) | WEK5 | Wierzchołek | 79.3 |
| Finkelstein/Bunbury (unpublished) | M28 | Attawapiskat River 1 | 79.6 |
| Bobrov et al. (1999) | STAR_1 | Staroselia | 80.1 |
| Lamentowicz et al. (2008) | CHL1 | Chlebowo | 80.4 |
| Bobrov et al. (1999) | KATIN_7 | Katin Moch | 80.7 |
| Finkelstein/Bunbury (unpublished) | M24 | Attawapiskat River 1 | 81.2 |
| Swindles et al. (2015) | C1 | Crash | 82.4 |
| Charman and Warner (1992) | 12C12 | Wally Creek | 84.3 |
| Lamentowicz et al. (2013);  Jassey et al. (2014) | WEK10 | Wierzchołek | 85.4 |
| Turner et al. (2013) | TM85 | Malham Tarn Moss | 85.5 |
| Mallon (unpublished) | GAL_3 | Gallseredsmossen | 85.7 |
| Finkelstein/Bunbury (unpublished) | M44 | Sutton River 2 | 87.0 |
| Lamentowicz et al. (2013);  Jassey et al. (2014) | WEK12 | Wierzchołek | 87.2 |
| Finkelstein/Bunbury (unpublished) | M17 | Attawapiskat River 1 | 87.5 |
| Lamentowicz et al. (2013);  Jassey et al. (2014) | KAZ3 | Kazanie | 87.9 |
| Finkelstein/Bunbury (unpublished) | M42 | Sutton River 2 | 88.2 |
| Lamentowicz et al. (2013);  Jassey et al. (2014) | RUR10 | Rurzyca | 88.4 |
| Charman and Warner (1992) | 1212 | Wally Creek | 88.4 |
| Bobrov et al. (1999) | Q92_1 | Tverskaya Reserve | 88.6 |
| Mallon (unpublished) | S11_8 | South | 89.0 |
| Finkelstein/Bunbury (unpublished) | M11 | Attawapiskat River 2 | 89.0 |
| Swindles et al. (2009) | MO6 | Moninea | 89.0 |

| | | | |
|---|---|---|---|
| Lamentowicz (unpublished) | K 09 | Kuźnik | 89.1 |
| Finkelstein/Bunbury (unpublished) | M4 | Attawapiskat River 1 | 89.1 |
| Charman and Warner (1992) | 1142 | Wally Creek | 89.2 |
| Swindles et al. (2009) | MO32 | Moninea | 89.3 |
| Lamentowicz et al. (2013);  Jassey et al. (2014) | RUR15 | Rurzyca | 89.3 |
| Lamentowicz et al. (2013);  Jassey et al. (2014) | RUR22 | Rurzyca | 89.3 |
| Lamentowicz (unpublished) | K 07 | Kuźnik | 89.7 |
| Bobrov et al. (1999) | KATIN_5 | Katin Moch | 89.7 |
| Swindles et al. (2009) | MO27 | Moninea | 89.9 |

**Table S6:** Samples removed from the combined North American and European dataset with extreme water-table depths (WTDs) (i.e. those below the 0.5th or above the 99.5th percentiles).

| Reference | Sample ID | Site | WTD (cm) |
|---|---|---|---|
| Turner et al. (2013) | TM91 | Malham Tarn Moss | -60 |
| Turner et al. (2013) | TM61 | Malham Tarn Moss | -50 |
| Booth (2001) | 832bda | Grand Traverse Bay | -37 |
| Booth (2002) | 1020B | Au Train Bay | -27 |
| Booth (2001) | 968ada | Tahquamenon Bay | -26 |
| Lamentowicz (unpublished) | K 14 | Kuźnik | -25 |
| Booth (2001) | 832ada | Grand Traverse Bay | -21 |
| Booth (2001) | 967ada | Tahquamenon Bay | -21 |
| Lamentowicz et al. (2013);  Jassey et al. (2014) | CZS14 | Czarci Staw | -20 |
| Lamentowicz et al. (2010) | 30 | Mauntschas | -20 |
| Turner et al. (2013) | TM62 | Malham Tarn Moss | -20 |
| Markel et al. (2010) | AK48 | Gas field | -18 |
| Lamentowicz et al. (2013);  Jassey et al. (2014) | WAG22 | Wagowo | -17 |

| Markel et al. (2010) | AK93 | Brown's Lake | -17 |
| Markel et al. (2010) | AK84 | Brown's Lake | -16.5 |
| Sullivan and Booth (2011) | TA28 | Tannersville | -16.5 |
| Turner et al. (2013) | TM69 | Malham Tarn Moss | 83 |
| Lamentowicz (unpublished) | SŁ26 | Słowińskie Błoto | 84 |
| Lamentowicz (unpublished) | SŁ27 | Słowińskie Błoto | 84 |
| Lamentowicz (unpublished) | SŁ28 | Słowińskie Błoto | 84 |
| Lamentowicz (unpublished) | SŁ29 | Słowińskie Błoto | 84 |
| Turner et al. (2013) | ThM189 | Thornton Moor | 85 |
| Mitchell et al. (1999) | S 22 | Les Pontins | 100 |
| Mitchell et al. (1999) | S 23 | Les Pontins | 100 |
| Mitchell et al. (1999) | S 44 | Le Bois des Lattes | 100 |
| Warner and Charman (1994) | NWO-D | Unknown | 100 |
| Warner and Charman (1994) | NWO-E | Unknown | 100 |
| Warner and Charman (1994) | NWO-F | Unknown | 100 |
| Mitchell et al. (1999) | S 16 | Le Cachot | 120 |
| Mitchell et al. (1999) | S 17 | Le Cachot | 120 |
| Mitchell et al. (1999) | S 18 | Le Cachot | 120 |
| Mitchell et al. (1999) | S 19 | Le Cachot | 120 |
| Mitchell et al. (1999) | S 20 | Le Cachot | 120 |
| Mitchell et al. (1999) | S 21 | Le Cachot | 150 |
| Mitchell et al. (1999) | S 41 | Le Bois des Lattes | 150 |

**Table S7:** Summary of published peatland testate amoeba transfer functions in North America and Europe. $RMSEP_{LOO}$ and $R^2$ values are for the best performing model as selected by the authors of each publication (as shown in model type column). Data for number of sites and samples is for data-pruned version of model if this information is available, if not, then values are for the complete dataset. For model type: WA-Tol (inv) = weighted averaging with tolerance downweighting and inverse deshrinking; WA-Tol = weighted averaging with tolerance downweighting, no deshrinking method specified; WAPLS C2 = second component of weighted averaging partial least squares; WA = weighted-averaging, no deshrinking method specified, WA (inv) = weighted averaging with inverse deshrinking, ML = maximum likelihood.

| Continent | Location | Number of sites | Number of samples | Model type | $RMSEP_{LOO}$ (cm) | $R^2$ | Reference |
|---|---|---|---|---|---|---|---|
| **North America** | **North America** | **126** | **1730** | **WA-Tol (inv)** | **7.42** | **0.72** | **This study** |
| North America | Newfoundland | 14 | 60 | WA-Tol | 6.32 | 0.71 | Charman and Warner (1997) |
| North America | Michigan | 2 | 74 | WAPLS C2 | 7.48 | 0.61 | Booth (2001) |
| North America | Michigan | 11 | 139 | WAPLS C2 | 7.51 | 0.74 | Booth (2002) |
| North America | Alaska | 8 | 101 | WAPLS C2 | 9.7 | Not given | Payne et al. (2006) |
| North America | Mid-continental and eastern USA | 31 | 369 | WAPLS C2 | 6.6 | 0.83 | Booth (2008) |
| North America | North Carolina | 1 | 33 | WA | 5.8 | 0.62 | Booth et al. (2008) |
| North America | Alaska | 12 | 126 | WAPLS C2 | 4.6 | 0.86 | Markel et al. (2010) |
| North America | Atlantic Canada and NE USA | 18 | 225 | WA-Tol (inv) | 5.66 | 0.87 | Amesbury et al. (2013) |
| North America | Québec | 18 | 165 | WA (inv) | 5.44 | 0.8 | Lamarre et al. (2013) |
| Europe | Europe | 113 | 1302 | WA-Tol (inv) | 7.72 | 0.59 | Amesbury et al. (2016) |
| Europe | Europe | 8 | 119 | WAPLS C2 | 5.6 | 0.71 | Charman et al. (2007) |
| Europe | Great Britain | 9 | 163 | WA-Tol | 3.93 | Not given | Woodland et al. (1998) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Europe | Western Russia | 6 | 31 | WAPLS C2 | 5.16 | 0.83 | Bobrov et al. (1999) |
| Europe | North-western Poland | 3 | 45 | WA-Tol | 9.89 | 0.91 | Lamentowicz and Mitchell (2005) |
| Europe | Greece | 4 | 57 | ML | 2.48 | 0.68 | Payne and Mitchell (2007) |
| Europe | Turkey | 1 | 42 | ML | 7.1 | 0.81 | Payne et al. (2008) |
| Europe | Western Poland | 1 | 30 | WA | 6.99 | 0.4 | Lamentowicz et al. (2008) |
| Europe | Northern Ireland | 3 | 95 | WA-Tol | 4.99 | 0.83 | Swindles et al. (2009) |
| Europe | England | 6 | 116 | WA-Tol (inv) | 7.36 | 0.71 | Turner et al. (2013) |
| Europe | Northern Sweden | 9 | 59 | WA-Tol (inv) | 5.25 | 0.87 | Swindles et al. (2015c) |
| | | | | **MEAN:** | 6.33 | 0.74 | |
| | | | | **STANDARD DEVIATION:** | 1.76 | 0.13 | |