# Next generation sequencing to identify lacustrine haptophytes in the Canadian Prairies: Significance for temperature proxy applications

**Julien Plancq[1*], Jillian M. Couto[2], Umer Z. Ijaz[2], Peter R. Leavitt[3, 4], and Jaime L. Toney[1]**

[1]School of Geographical and Earth Sciences, University of Glasgow, Glasgow, G12 8QQ, UK.

[2]School of Engineering, University of Glasgow, Glasgow, G12 8QQ, UK.

[3]Limnology Laboratory, Department of Biology, University of Regina, Regina, S4S 0A2, Canada.

[4]Institute for Global Food Security, Queen's University Belfast, Belfast, BT9 7BL, UK.

*Corresponding author: Julien Plancq (julien.plancq@glasgow.ac.uk)

**Key Points:**

- Next generation sequencing (NGS) was used to identify alkenone-producing haptophyte species in five lakes of the Canadian Prairies.
- Group I and Group II haptophyte species were identified in oligohaline and oligohaline/saline lakes, respectively.
- Temperature dependency of alkenones is likely consistent within different taxa from Group I and Group II haptophytes.

## Abstract

The Great Plains of North America often experience prolonged droughts that have major economic and environmental impacts. Temperature reconstructions are thus crucial to help decipher the mechanisms responsible for drought occurrences. Long-chain alkenones (LCAs), lipids produced by three major phylogenetic groups (Group I, II and III) of haptophyte algae within the order Isochrysidales, are increasingly used for temperature reconstructions in lacustrine settings. However, to select the most appropriate calibration of the LCA-based temperature proxy, it is first essential to identify the LCA-producing haptophyte species present. Here we used next generation sequencing to target the 18S rRNA haptophyte gene from sediments with distinct LCA profiles to identify the LCA-producer(s) from five Canadian prairie lakes. In total, 374 operational taxonomic units (OTUs) were identified across the studied samples, of which 234 fell within the Phylum Haptophyta. Among the most abundant OTUs, three were characterized as LCA-producers, one falling within the Group I haptophytes and two within the Group II haptophytes. The OTU from Group I haptophytes was associated with a single, highly-specific LCA profile, whereas Group II OTUs showed higher variability in LCA distributions. Our study revealed that most of the LCA-producing OTUs thriving in the Canadian lakes are included within the genus *Isochrysis*, which helps guide selection of the most appropriate calibration for down-core temperature reconstructions. Our findings also suggest that the temperature dependency is likely consistent within different taxa from Group I and Group II haptophytes, but that other environmental parameters may influence the accuracy of the calibration.



## Plain Language Summary

The Great Plains of North America are extremely sensitive to changes in temperature and moisture and often experience prolonged periods of droughts and floods that have major impacts on agriculture production and ecosystem function. Temperature reconstructions are thus crucial to help decipher the climate mechanisms responsible for drought and flood occurrences. In the present study, we identified the algae that biosynthesize a very specific class of lipids, the long-chain alkenones (LCAs). LCAs have been successfully used for decades to reconstruct past temperature changes in marine environments, but it is important to identify the algae producing these lipids in lakes before we can use them as a palaeo-thermometer for aquatic environments. Here, we used new DNA-based techniques to identify three main distinct species in five lakes of the Canadian Prairies. One species produces a highly-specific LCA distribution and was found in freshwater lakes, while the two other species were found in more saline lakes. By comparing our findings with previous works, we also discovered that environmental parameters other than temperature may influence the accuracy of historical temperature reconstructions.


## 1. Introduction

Haptophytes are unicellular algae that possess a unique flagellum-like organelle, the haptonema, which might play a role in prey capture in some species (Kawachi et al., 1991). The Phylum Haptophyta is divided in two classes, the Pavlovophyceae, which include one order, and the Prymnesiophyceae, with 6 orders (Bendif et al., 2011; Jordan et al., 2004). While haptophytes are known principally from marine environments where they are one of the most

abundant groups of phytoplankton and significant primary producers, some representatives also thrive in inland waters, exhibiting a high degree of morphological, physiological and functional diversity (Jordan & Chamberlain, 1997).

Within the Prymnesiophyceae, a select group of species from the order Isochrysidales biosynthesize a class of $C_{35}$–$C_{42}$ di-, tri- and tetra-unsaturated methyl and ethyl ketones called long-chain alkenones (LCAs). These lipids are produced solely by three distinct phylogenetic groups within the Isochrysidales (Theroux et al., 2010). Group I includes haptophyte species restricted to freshwater (< 0.5 g/L) and oligohaline (0.5-5 g/L) lakes (e.g., Longo et al, 2016, 2018; Theroux et al., 2010), whereas Group II includes *Ruttnera lamellosa* and *Isochrysis galbana*, as well as related species found in brackish waters and saline lakes (e.g., Longo et al, 2016; Theroux et al., 2010; Toney et al., 2012). Group III includes the coccolithophores *Emiliania huxleyi* and *Gephyrocapsa oceanica*, which are the main LCA-producers in marine environments. Since the degree of unsaturation of $C_{37}$ LCAs has been shown to vary as a function of the environmental temperature experienced during haptophyte growth (Brassell et al., 1986; Prahl & Wakeham, 1987), LCA unsaturation indices ($U^{K}_{37}$, $U^{K'}_{37}$) have been developed and calibrated in both laboratory cultures and marine surface sediments as proxies for sea surface temperature (Conte et al., 2006; Müller et al., 1998; Prahl et al., 1988). These proxies have been successfully applied for decades to reconstruct past marine temperature records at various sites and timescales (e.g., Müller et al., 1998; Pahnke & Sachs, 2006; Rosell-Melé et al., 1995). More recently, LCA unsaturation has been demonstrated to vary with lake temperature in some sites, and a number of calibrations of the $U^{K}_{37}$ and $U^{K'}_{37}$ indices have been developed in cultures (Araie et al., 2018; Sun et al., 2007; Theroux et al., 2013; Toney et al., 2012; Zheng et al., 2016) and in situ in freshwater and saline lakes (Chu et al., 2005; D'Andrea et al., 2011, 2016; Longo et al., 2016; Toney et al., 2010; Zheng et al., 2016; Zink et al., 2001). However, unlike the open oceans where LCAs are mainly produced by *E. huxleyi* and *G. oceanica* that have a single linear relationship to temperature, multiple haptophyte species can be found in lakes, which may imply the need for site-specific LCA temperature calibrations (e.g., Randlett et al., 2014; Theroux et al., 2010). Consequently, achieving this goal requires the identification of both the haptophyte species and the LCAs present in each lake prior to historical temperature reconstructions in lacustrine environments (e.g., Castañeda & Schouten, 2011; Coolen et al., 2004; D'Andrea et al., 2006; Theroux et al., 2010).

The development of LCA-based temperature proxies is important for the Great Plains of North America, where there are few unambiguous metrics of past temperatures in inland waters before the instrumental period. This region is sensitive to changes in temperature and moisture and often experiences prolonged periods of droughts and floods that have major impacts on agriculture production and ecosystem function (Laird et al., 2003; Wheaton et al., 1988). Quantitative palaeotemperature reconstructions are thus crucial to help decipher the climate mechanisms responsible for drought occurrences. A recent study conducted on a suite of 106 lakes from the Canadian part of the Northern Great Plains (Canadian Prairies, southern Saskatchewan) has shown that salinity and stratification are the main environmental factors determining LCA presence and abundance, which in turn are likely linked to the life cycle of lacustrine haptophytes (Plancq et al., 2018a). In parallel, Araie et al. (2018) studied ten of these 106 lakes, and isolated seven haptophyte strains from three saline lakes. An analysis of both the 18S and 28S ribosomal RNA (rRNA) gene from each isolate revealed that they all belong to the Group II haptophyte, and, subsequently, the isolates were used to establish three culture

106 calibrations (Araie et al., 2018). However, four distinct LCA profiles were observed in the suite
107 of 106 survey lakes (Plancq et al., 2018a), suggesting that additional LCA-producing species
108 were yet to be found in the Canadian lakes.

109 In the present study, we sought to identify these additional LCA-producing haptophyte
110 species via genomic analyses targeting the haptophyte 18S rRNA gene. Although Sanger
111 sequencing is commonly used in environmental genomic studies to identify haptophyte species
112 in lacustrine environments (e.g., Araie et al., 2018; Coolen et al., 2004; D'Andrea et al., 2016;
113 Theroux et al., 2010), we opted for next generation sequencing (NGS). NGS has been previously
114 employed when studying haptophytes, providing many novel insights into the taxonomy,
115 phylogeny, diversity and ecology of these algae (Egge et al., 2015; Endo et al., 2018; Gran-
116 Stadniczeñko et al., 2017; Shalchian-Tabrizi et al., 2011; Theroux et al., 2012). Here we chose
117 five Canadian prairie lakes from which distinct LCA profiles and 18S rRNA data were generated
118 from sediment to identify the LCA-producer(s). We identified 234 different haptophyte rRNA
119 signatures, referred to as operational taxonomic units (OTUs) herein, of which three were found
120 to be LCA-producing species. These new findings, compared with previous studies, are
121 discussed to determine which calibration of the $U_{37}^{K}$ index would be most appropriate for
122 temperature reconstructions in the Canadian Prairies.

123

124 **2. Material and Methods**

125 2.1. Sampling

126 Following on a previous study of 106 lakes from the Canadian Prairies (southern
127 Saskatchewan, Canada) (Plancq et al., 2018a), five lakes were selected based on their LCA
128 profiles and concentrations (Figure 1). Manitou Lake and Middle Lake feature a LCA profile
129 dominated by $C_{37:4}$ LCA, whereas Dewey Lake exhibits a profile mainly composed of the $C_{37:3}$
130 LCA. Richmond Lake displays a predominance of $C_{38}$ LCAs over $C_{37}$ LCAs, with particularly
131 elevated concentrations of the $C_{38:3}$ LCA. Finally, Lenore Lake shows the presence of tri-
132 unsaturated LCA isomers, including the $C_{37:3b}$ isomer, which has been shown to be characteristic
133 of Group I haptophytes (Longo et al., 2013, 2016, 2018). These lakes were sampled between
134 June 22$^{nd}$ and June 30$^{th}$ 2016. In situ lake parameters for each lake, including salinity,
135 conductivity, pH, depth and temperature, were measured at 0.5-m depth intervals (lakes <10 m
136 depth) or 1-m depth intervals (lakes >10 m depth) for the entire water column using a YSI Pro
137 Plus meter (YSI Inc., Yellow Springs, Ohio, USA) following Pham et al. (2008). Surface
138 sediments were collected from each lake using an Ekman grab sampler and the uppermost 1-cm
139 interval was kept at 4 °C and in the dark in sealed bags until processing to avoid any cross
140 contamination. Sediment sample from each lake was split for DNA and LCA analyses.

141 2.2. DNA extraction

142 DNA was extracted from 0.5 g of sediment from each lake using FastDNA Spin Kit for soils
143 and sediments (MP Biomedicals, Santa Ana, California, USA) according to the manufacturer's
144 instructions. DNA was also extracted from surface sediment of Deadmoose Lake collected in
145 2013 and from a culture sample of a strain of LCA-producing haptophyte algae isolated from
146 Lake George, USA (*Isochrysis* sp.; RCC 4054) for a positive control. Negative controls (DNA

147 extraction blanks) were also used during the DNA extraction process to check for any potential
148 contamination. Total extracted DNA yields were quantified using a QUBIT fluorescence assay
149 (Invitrogen, Carlsbad, California, USA).

150 2.3. DNA amplification and sequencing

151 For PCR amplification we used the primer pair Hap454 described in Egge et al. (2013), with
152 universal eukaryote forward primer 528Flong: 5'-GCGGTAATTCCAGCTCCAA-3', and
153 haptophyte-specific reverse primer PRYM01 + 7: 5'-GATCAGTGAAAACATCCCTGG-3'.
154 This primer pair targets the 18S V4 region of the SSU rDNA of haptophyte species and was
155 designed to give nucleotide fragments of about 400 bp (Egge et al., 2013) suitable for Illumina
156 MiSeq (Illumina, San Diego, California, USA) sequencing.

157 Fusion primers for sequencing on the Illumina MiSeq platform were designed by adding
158 adaptors: 528Flong(F): 5'-
159 ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGCGGTAATTCCAGCTCCAA-
160 3' (where N are degenerate bases); PRYM01+7(R): 5'-
161 GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATCAGTGAAAACATCCCTGG-3'.
162 The fusion primers were HPLC-purified after synthesis (Eurofins, Ebersbeg, Germany), and
163 Polymerase chain reactions (PCRs) were directly performed with the fusion primers on a
164 BIOTAQ PCR kit (bioline BIO-2107), with the following conditions: 3 min initial denaturing at
165 95 °C, 25 or 30 cycles of denaturing for 20 s at 98 °C, 15 s primer annealing at 56 °C and 15 s of
166 primer extension at 72 °C, with a final extension of 1 min at 72 °C. Each sample was run at
167 optimal dilution (1:10 or 1:100) for maximum product yield. PCR reactions were run in 25 µl
168 volume using Q5 High-Fidelity DNA Polymerase (New England Biolabs, Ipswich,
169 Massachusetts, USA). Sextuplicate PCR products were pooled for each sample, and templates
170 were purified using a Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, California,
171 USA) according to manufacturing guidelines. Sequencing was performed at the Centre for
172 Genomic Research (University of Liverpool, UK), using the Illumina MiSeq platform with v2
173 chemistry to yield 2x250bp paired-end reads.

174 2.4. Bioinformatics and phylogenetic reconstructions

175 We followed the bioinformatics pipeline according to the recommendations given in recent
176 publications (D'Amore et al., 2016; Schirmer et al., 2015) that results in significant reduction of
177 substitution errors. Briefly, we performed quality trimming (Sickle) combined with error
178 correction (BayesHammer) followed by read overlapping (PANDAseq). The paired-end reads
179 were then trimmed and filtered using Sickle v1.21 (Joshi & Fass, 2011) using the default
180 settings, i.e., by applying a sliding window approach and trimming regions where the average
181 base quality drops below 20. SPAdes was used to perform error-correction as it is bundled with
182 BayesHammer, which is used for pre-correction of reads before performing assembly (Nikolenko
183 et al., 2013). We then used PANDAseq v2.4 (Masella et al., 2012) with a minimum overlap of
184 10 bp to assemble the forward and reverse reads into a single sequence spanning the entire V4
185 region. After obtaining the consensus sequences from each sample, we used the UPARSE
186 pipeline employing usearch v7.0.1001 (Edgar, 2013) slightly modified after Schirmer et al.,
187 (2015) (https://bitbucket.org/umerijaz/amplimock/src) for Operational Taxonomic Unit (OTU)
188 construction. The reads were de-replicated and singletons were discarded. In the next step, the

reads were clustered based on 97% similarity (e.g., Liu et al., 2009; Tragin et al., 2018). Even though the *cluster_otu* command in usearch removes reads that have chimeric models built from more abundant reads, a few chimeras may be missed, especially if they have parents that are absent from the reads or are present with very low abundance. Therefore, in the next step, we used a reference-based chimera-filtering step using a gold database (http://drive5. com/uchime/uchime_download.html) that is derived from the ChimeraSlayer reference database in the Broad Microbiome Utilities (http://microbiomeutil.sourceforge. net/). A total of 374 OTUs comprising all samples were generated. The number of sequences recovered per OTU for each sample is listed in the Supplementary Table S1. The *assign_taxonomy.py* script from the Qiime workflow (Caporaso et al., 2010) was then used to taxonomically classify the representative OTUs against the SILVA SSU Ref NR database release v123 database (Quast et al., 2013). The biom file for the OTUs was then generated by combining the abundance table with taxonomy information using *make_otu_table.py* from the Qiime workflow.

To find the phylogenetic distances between OTUs, we first aligned the OTUs against each other using MAFFT v7.040 (Katoh & Standley, 2013). We then constructed the phylogeny of 18S rRNA sequences using GTR + Γ + I (mixed invariable sites and gamma-distributed rates; Lanave et al., 1984; Rodriguez et al., 1990; Tavaré, 1986) in MrBayes version 3.2.2 (Ronquist et al., 2012). Although there are many models for nucleotide substitution, our selection of GTR+ Γ+ I model was based on recommendations given in Zhou et al. (2011) and Buckley and Cunningham (2002), where in the latter case, GTR + Γand GTR + Γ +I model led to the selection of the correct topology for 18S rRNA data. The chain length for our analysis was 1,000,000 generations with trees sampled every 100 generations using MCMC (Markov Chain Monte Carlo) analysis. The first 10,000 trees were discarded as burn-in for the tree topology and posterior probability. OTU representative sequences and full-length 18S rRNA gene sequences from reference taxa were analysed to infer OTU species' identities (from Araie et al., 2018; D'Andrea et al., 2016; Plancq et al., 2018b; Theroux et al., 2010). We selected *Cyclonexis annularis*, *Chrysoxys* sp., *Ochromonas danica*, *Odontella sinensis* and *Thraustochytrium multirudimentale* as outgroups for our Bayesian analyses after de Vargas et al. (2007).

## 2.5. LCA extraction and analysis

Sediments were freeze-dried at 0.01 Pa for 36 hours, homogenized and extracted with dichloromethane (DCM):methanol (MeOH) (9:1, v:v) using a Dionex model ASE350 accelerated solvent extractor. Following evaporation of the solvent, the total lipid extracts were separated into neutral and acid fractions by elution through a LC-$NH_2$ SPE column using DCM:isopropyl alcohol (1:1, v:v) followed by ether with 4% acetic acid (v:v) as eluents, respectively. The neutral fractions were further separated into four fractions of increasing polarity by chromatography over a silica gel column packed with 35-70 μm particles using hexane, DCM, ethyl acetate:hexane (1:3, v:v) and MeOH as eluents.

LCAs are contained in the second fraction (DCM) and were detected and quantified using gas chromatography with a flame-ionization detector (GC-FID). To remove compounds that can co-elute with LCAs such as alkenoates, LCA fractions were saponified by heating at 60 °C overnight in 1N KOH in MeOH:$H_2O$ (95:5, v:v). After heating, the reaction mixture was cooled to room temperature, quenched with NaCl (5% by volume), extracted using hexane, and rerun on the GC-FID. GC-FID analyses were performed on an Agilent 7890B Series GC system

configured with an Agilent VF-200 ms capillary column (60 m length, 0.25 mm internal diameter, 0.10 µm film thickness) (Longo et al., 2013). Hydrogen was used as the carrier gas at a 36 cm/s column flow rate. The GC method used splitless injection (320 °C), and the oven temperature was programmed from 50 °C (hold for 1 min) to 255 °C at 20 °C/min, then to 300 °C at 3 °C/min, followed to 10 °C/min increase to 320 °C and hold for 10 min. Samples were also run with the same temperature program on an Agilent 7890B Series GC coupled with a 5977A GC-EI Mass Spectrometer (GC-MS) to confirm the identity the LCAs using the known ion chromatograms and by comparison of mass spectral data and GC retention times with published data (de Leeuw et al., 1980; Marlowe et al., 1984). LCAs were quantified using hexatriacontane (*n*-$C_{36}$ alkane) as an internal standard added to the sample before injection.

## 3. Results

### 3.1. LCA profiles and concentrations

Consistent with the study of Plancq et al. (2018a), four distinct LCA profiles were observed in the surface sediments from the studied lakes (Figure 2). Manitou, Middle and Deadmoose lakes all featured a LCA profile dominated by $C_{37:4}$ alkenone, whereas Dewey Lake featured a profile mainly composed of the $C_{37:3}$ alkenone. Richmond Lake displayed a predominance of $C_{38}$ alkenones over $C_{37}$ alkenones, with particularly elevated concentrations of the $C_{38:3}$ alkenone. Finally, Lenore Lake showed the presence of tri-unsaturated alkenone isomers, including the $C_{37:3b}$ isomer, which has been shown to be characteristic of the Group I of LCA-producing haptophytes (Dillon et al., 2016; Longo et al., 2013, 2016, 2018). Total concentrations of LCAs ranged from 12.9 to 2184 µg/g dry sediment, with the highest concentrations (>1000 µg/g dry sediment) recorded in Manitou and Middle lakes (Table 1).

### 3.2. OTU diversity and phylogenetic analyses

Illumina sequencing revealed a total of 374 OTUs across the sample studied (Supplementary Table S1). Based on bioinformatic taxonomic assignment (SILVA database), 234 OTUs were identified within the Phylum Haptophyta, whereas 83 OTUs fell within the Eukaryota Domain, 1 to the Bacteria Domain and 56 OTUs did not correspond to any known sequences referenced within the SILVA database. Out of the 83 OTUs falling in the Eukaryota Domain, 11 belonged to the Archaeplastida group (including mainly Chlorophyta), 22 to the Alveolata group (including Apicomplexa and Dinoflagellata), 22 to the Opisthokonta group (including mainly Fungi), 19 to the Stramenopiles group (including Diatomea and Chrysophycea), 1 to the Amoebozoa group, and 8 as “uncultured eukaryotes”. These eukaryotic groups are commonly found in these and other regional Canadian lakes (Rawson & Moore, 1944; Vogt et al., 2011). Overall, the number of reads for these non-haptophyte sequences was very low; in average 2120 (0.13% of the total reads), with a maximum for OTU_54 (46548 reads; 2.51% of the total reads). The abundance for the 234 OTUs from the Phylum Haptophyta for each lake is shown in a circular phylogenetic tree in Supplementary Figure S1. We selected the most abundant OTUs for each sample by choosing those with more than 2% of the total reads, resulting in 20 OTUs which represented at least 70% of the total reads in each sample (Figure 3). The DNA sequences for the 20 OTUs are available on GenBank under accession numbers MK092726 to MK092745.

Overall, the most common OTUs were: OTU_16 (46.4%, 51.2%, 59.3%, 22.9%, 15.5% in Middle, Manitou, Dewey, Richmond and Deadmoose lakes, respectively), OTU_29 (61.4%, 21.4%, 39.6% in RCC 4054, Richmond and Deadmoose lakes, respectively), OTU_21 (80.4% in Lenore Lake), and OTU_2 (11.3%, 24.6% in Middle and Richmond lakes, respectively).

To infer nearest-neighbor relationships with previously published sequences, a phylogenetic tree was constructed using the 20 most abundant OTUs from the present study along with publicly available haptophyte 18S rRNA genes in the GenBank database (Figure 4). Out of the 20 most abundant OTUs, one (OTU_54) did not fall within the Phylum Haptophyta and grouped instead with the outgroup species. This OTU was only found in low abundance (2.5%) in Deadmoose Lake. Three OTUs belonged to the order Pavlovales of the Phylum Haptophyta (OTU_2, 5 and 19). OTU_2 was particularly abundant in Richmond (24.6%) and Middle (11.6%) lakes, while it was less abundant in Manitou Lake (3.3%) (Figure 3). OTU_19 was mainly found in Manitou Lake (8.1%), and in similar low proportions (2.1%) in Middle and Deadmoose lakes. OTU_5 was found in low abundances in Lenore (3.4%) and Deadmoose lakes (2.3%).

All other OTUs grouped within the order Isochrysidales of the Phylum Haptophyta (Figure 4). OTU_21, along with OTU_1, 7, 10, and 14, clustered in the Group I of LCA-producing haptophytes. Specifically, OTU_21 was closely related to the OTU7 from Lake Toyoni (Japan), the Lake Vikvatnet sequence (Norway), the OTU5 (from lakes in Greenland, China, USA, Canada), and the Greenland sequences (from BrayaSø, HundeSø, LimnaeSø lakes). OTU_14 was also closely related to the Greenland sequences, whereas OTU_1, 7 and 10 were included in the EV clade (sequences from France). All these Group I OTUs were only found in the oligohaline Lenore Lake, where OTU_21 largely dominated (80.4%) while the other OTUs were less abundant (2.5 to 3.4%) (Figure 3). OTU_16 and OTU_29 clustered in the Group II of LCA-producing haptophytes, along with OTU_6, 8, 9, 13, 15, 18, 27, 38, and 57 (Figure 4). More precisely, OTU_16, 6, 8, 15, 18, and 57 were closely related to the OTU8 (from lakes in China, USA and Canada) and *Isochrysis galbana* (Figure 4). OTU_29, 13, 38, and 9 were close to OTU6 from Lake Tso Ur (China), *Pseudoisochrysis paradoxa*, and haptophyte isolates from Lake Success and Lake Deadmoose (Canada). OTU_27 was closely related to *Ruttnera lamellosa* (Figure 4). These Group II OTUs were found in different proportions in all the studied lakes, except Lenore Lake (Figure 3). In Middle, Manitou and Dewey lakes, OTU_16 was the most abundant (between 46.4 and 49.3%), while all other OTUs were found in lower abundances (1.7 to 8.8%). In Deadmoose Lake, OTU_29 (39.6%) dominated over OTU_16 (15.5%), whereas both OTUs were found in similar proportions (21.4% and 22.9%) in Richmond Lake. The other OTUs were present in low proportions (1.1 to 4.0%).

## 4. Discussion

### 4.1. Haptophyte diversity in the Canadian lakes

In the present study, we analysed sediment from five Canadian lakes for LCA-producing haptophytes to assist with temperature reconstruction. We employed next generation sequencing (NGS) with the Illumina Miseq platform to capture a 400 bp amplicon of the 18S ribosomal

RNA gene, which enabled us to identify 20 major OTUs (with > 2% abundance). Within this group of 20, just four abundant (> 10%) OTUs dominated, with the rest existing as relatively rare (referred to as "rare subset" herein) OTUs (Figure 3). Overall, this 400 bp fragment of the 18S gene is very conserved amongst the Phylum Haptophyta (Egge et al., 2015), and indeed all the OTUs in the rare subset were phylogenetically similar to one of the four abundant OTUs (Figure 4) suggesting either the possibility of rare false positives, or the presence of a new, undiscovered diversity of haptophytes within these lakes.

To explore this further, we considered the methods used in our study in more detail. NGS offers the advantage of immensely high resolutions compared with its predecessor Sanger sequencing and would therefore enable us to identify a greater diversity of rare organisms residing in environmental samples. Although this is now a routine occurrence in the prokaryotic world (Land et al., 2015), NGS is still fairly new to the study of haptophytes and rare variants could be proven to be real if replicated in future independent studies. Nonetheless, the NGS results from sample RCC 4054 were perplexing as they should have contained a pure culture of *Isochrysis* sp. In concordance, NGS showed that OTU_29 was the most abundant (61.4%), but others from the rare subset were also present (Figure 3), with OTU_9 [6.2%], OTU_38 [3.1%] and OTU_13 [2.4%] being extremely related to OTU_29, while OTU_27 [4.2%] was more closely related to *R. lamellosa* (Figure 4). To the best of our knowledge this is the first time that a pure culture has undergone next generation amplicon sequencing. Given that RCC 4054 was, at one time, sub-cultured from an environmental sample, it is plausible that a new, high-resolution method could have picked up additional rare (sub)groups. However, the present analysis of a 400 bp amplicon shows that OTU_9, 38 and 13 are very closely related to OTU_29. These amplicons were all prepared using a PCR step and despite the use of high-fidelity polymerase, low numbers of amplification cycles and well-optimized bioinformatics pipelines that reduce substitution error rates by 93% (Schirmer et al., 2015), it is still possible that rare errors could have occurred very early on during PCR in a subset of molecules and persisted through the process. If this were the case, then these would appear as rare, highly related OTUs. Additional NGS of whole genomes (metagenomics) forgoes the PCR step and would provide a suitable solution, but is beyond the scope of this study and would be better placed in future work. In the current work, we applied a conservative cut-off of 10% and thus, our subsequent discussion of each lake will mainly focus on the OTUs that occur at an abundance of more than 10% (OTU_16, 29, 21 and 2).

The distribution of OTUs varied among sites, with some lakes mainly exhibiting a single haptophyte OTU, whereas other basins presented a greater diversity (Figure 3). For examples, Manitou, Dewey and Middle lakes were all characterized mainly by OTU_16 and other closely related Group II Isochrysidales OTUs (OTU_6, 15, 57, 18), while Lenore Lake was clearly characterized by OTU_21 (and other Group I Isochrysidales OTUs, OTU_14, 7, 1 and 10). Deadmoose Lake was dominated by OTU_29 and secondarily OTU_16, both of which are included in Group II Isochrysidales. Interestingly, OTU_16 and 29 and OTU_2 (order Pavlovales) were found in the same proportions within Richmond Lake. While the presence of Group I Isochrysidales OTU_21 (and OTU_14, 7, 1 and 10) in Lake Lenore only can be linked to salinity (see section 4.2 below), it is more difficult to explain why Group II Isochrysidales OTUs (mainly OTU_16 and 29) and Pavlovales OTUs (OTU_2, 5 and 19) can be found in different proportions among lakes. These discrepancies in diversity do not seem to be related to differences in the environmental parameters measured during the sampling of the lakes (salinity, pH, conductivity, stratification; Table 1), even though the low sample size (n = 6) is insufficient

to conclusively determine which environmental parameter may play a role. A study by Theroux (2013) at Lake George (North Dakota, USA), which is close to our study region, has shown that OTU8 haptophyte (Hap A) blooms while nutrient concentrations are low in the spring, whereas OTU7 haptophyte (Hap B) blooms when nutrient concentrations are higher in the late spring/early summer. Interestingly, OTU_16 and OTU_29 seem to be closely related to OTU8 and OTU7, respectively (Figure 4). The differences in their proportions among the studied lakes could be then linked in different nutrient conditions. Further studies would be however necessary to test and confirm this hypothesis.

### 4.2. LCA-producers in Canadian prairie lakes

Three main LCA-producers have been identified in the five lakes studied. Two (OTU_16 and OTU_29) belong to the Group II of LCA-producing haptophytes and are found in oligohaline and saline lakes (Manitou, Middle, Deadmoose, Richmond, Dewey), while one (OTU_21) segregates within the Group I of LCA-producing haptophytes and was found only in oligohaline Lenore Lake (Figures 3 and 4). Interestingly, Richmond and Deadmoose lakes each contained two main LCA-producing haptophyte species, while the other studied lakes only contained a single dominant LCA-haptophyte species. These results are consistent with previous studies that have shown that Group I haptophytes only occur in freshwater and oligohaline lakes, whereas Group II haptophytes occur in oligohaline to hyperhaline lakes (Longo et al., 2016, 2018; Plancq et al., 2018b; Simon et al., 2013; Theroux et al., 2010; Toney et al., 2012). When looking at the LCA profiles (Figure 2), OTU_21 is associated with a LCA profile with tri-unsaturated isomers and the presence of $C_{38}$Me LCAs, which has been shown to be characteristic of Group I haptophytes (Longo et al., 2013, 2018). In contrast, Group II OTU_16 and OTU_29 appear to be associated with different and/or similar LCA profiles (Figure 2). For example, OTU_16 is dominant in both Middle and Dewey lakes, but the LCA profile is characterised by high abundance of $C_{37:4}$ LCA in Middle Lake, whereas it is dominated by the $C_{37:3}$ LCA in Dewey Lake (Figures 2 and 3). This finding makes it difficult to access species effects for these distributions. In addition, although the LCA profiles recorded in the surface sediments probably mainly reflect the production of these dominant LCA-producers, the other low abundant (less than 10% of reads) Group I (OTU_14, 1, 7 and 10) and Group II OTUs (OTU_6, 8, 15, 18, 57, 13, 38, 9 and 27) also contributed to the LCA record. Thus, LCA profiles alone often do not reflect the entire extent of haptophyte species diversity known in lakes, as similar LCA profiles may be produced by different species and/or different LCA profiles may be derived from similar haptophyte species (Theroux et al., 2010). This pattern implies that the different LCA profiles recorded in the surface sediments of Deadmoose, Manitou, Middle, Dewey and Richmond lakes might have been controlled by environmental conditions rather than differences in Group II haptophyte species. This may be especially highlighted in Richmond Lake where the LCA profile is characterised by a predominance of $C_{38}$ LCAs over $C_{37}$ LCAs. Culture studies on the LCA-producers *E. huxleyi*, *G. oceanica* and *I. galbana* have confirmed that nutrient, light stress or cell physiological state can affect the LCA distribution such as the ratio between $C_{37}$ and $C_{38}$ LCAs (e.g., Conte et al., 1998; Prahl et al., 2003; Versteegh et al., 2001). Genomic analyses are thus important to unambiguously identify LCA-producing haptophyte species in the studied lake.

Using 18S and 28S rRNA gene analyses on haptophyte isolates and environmental samples, Araie et al. (2018) previously identified four distinct LCA-producing OTUs included in the Group II haptophytes from four lakes (Snakehole, Deadmoose, Success and Waldsea; Figure 1)

of the Canadian Prairies. One OTU from Snakehole Lake clustered with *Isochrysis litoralis* and *Isochrysis nuda*, while two OTUs from Deadmoose and Success lakes were included in the *I. galbana* clade. Finally, one OTU from Waldsea Lake clustered with *R. lamellosa*. Our genomic analyses reveal the presence of at least three additional LCA-producing OTUs. OTU_29 is actually closely related to the haptophyte sequences from Deadmoose and Success lakes, while OTU_16 is more closely related to the OTU8 which has been found in lakes from USA and China (Theroux et al., 2010). OTU_21 occupied a unique phylogenetic position within Group I haptophytes. All together, these data show that at least seven LCA-producing OTUs thrive in the Canadian lakes, even though five of them are included within the genus *Isochrysis*.

### 4.3. Implications for temperature proxy applications

Previous studies have used in situ calibrations of the $U_{37}^{K}$ proxy and successfully reconstructed lacustrine temperatures during the Holocene (e.g., D'Andrea et al., 2011; He et al., 2013; van der Bilt et al., 2018; Wu et al., 2018). Such success suggests that, as in the marine environments, quantitative palaeotemperature reconstructions using modern $U_{37}^{K}$ calibrations can be produced in lakes, even though the LCA-producers probably changed through time. However, extensive background research is needed for the development of site- and species-specific calibrations before LCA can be confidently used in lacustrine ecosystems (Castañeda & Schouten, 2011). Unfortunately, there are no site-specific calibrations developed for the studied Canadian prairie lakes. However, we can use the information from genomic analyses to try and determine the most appropriate temperature calibration available in the literature. In the absence of their own in situ calibrations, some studies that successfully reconstructed past lacustrine temperatures have actually applied previously published calibrations (e.g., van der Bilt et al., 2018; Wu et al., 2018). When looking at the most abundant OTUs, OTU_29 is closely related to haptophyte isolates from Deadmoose and Success lakes, and thus the culture calibrations established for those isolates (Araie et al., 2018) could be used for temperature reconstructions at Deadmoose and Richmond lakes. Similarly, OTU_16 is closely related to the OTU8 which has been found in Lake George, USA, for which in situ and culture calibrations are available (Toney et al., 2012). Temperature calibrations from Lake George could thus be used at Middle, Manitou and Dewey lakes. For lakes with the presence of Group I haptophyte producers, three in situ temperature calibrations have been developed from Lake BrayaSø (Greenland; D'Andrea et al., 2011), Lake Vikvatnet (Norway; D'Andrea et al., 2016) and Lake Toolik (Alaska; Longo et al., 2016). Recently, Richter et al. (2019) showed that Group I haptophyte sequences from Toolik Lake (Alaska) were closely related to haptophyte sequences from Lake BrayaSø (Greenland) and previously defined OTUs in lakes from Canada, China, France, Greenland, and the United States, and all included in the "Greenland phylotype". These findings suggest that LCA production in response to temperature may be consistent among Group I haptophytes (Longo et al., 2018; Richter et al., 2019). This pattern further implies that, although OTU_21 appears to be closely related to sequences from Lake Vikvatnet (Norway) and Lake BrayaSø (Greenland), the in situ calibrations of D'Andrea et al. (2011, 2016) and Longo et al. (2016) are likely to be appropriate for Lenore Lake.

In principle, the presence of multiple species in a single lake complicates the use of the $U_{37}^{K}$ proxy, as the relationship between temperature and the LCA unsaturation index could differ between taxa (Randlett et al., 2014; Theroux et al., 2010; Toney et al., 2010). Lakes containing a single LCA-producing species would thus be ideal for palaeotemperature reconstructions since

they would be immune to "species mixing effects" (Longo et al., 2016). Among our study lakes, each seems susceptible to such "species mixing effects", even though Middle, Manitou, Dewey and Lenore lakes each contain one predominant LCA-producer, but several other OTUs. Nevertheless, a study by Theroux (2013) at Lake George, USA, showed that, even though OTU7 (Hap B) and OTU8 (Hap A) occur during different seasonal blooms and have distinct LCA profiles, their culture calibrations are statistically indistinguishable and the combined linear calibration throughout the blooms is equally robust and very similar to the in situ calibration by Toney et al. (2010). Thus, despite having two different taxa in the same lake, the in situ calibration appears to be robust from the perspective of a paleoclimate reconstruction.

Comparison of different culture calibrations for Group II haptophyte species (Araie et al., 2018; Nakamura et al., 2014, 2016; Sun et al., 2007; Theroux et al., 2013; Toney et al., 2012; Versteegh et al., 2001; Zheng et al., 2016) reveals that the slope of the regressions, which is temperature dependent, is highly variable (0.015 - 0.059; Figure 5). The slopes can actually be different between highly related OTUs (for example Sc2 and Dm2; Figure 5) or strains from the same species (for example *I. galbana*; Figure 5). Interestingly, when comparing different in situ calibrations from lakes containing Group I (D'Andrea et al., 2011, 2016; Longo et al., 2016; Zink et al., 2001) and Group II haptophytes (Toney et al., 2012; Wang & Liu, 2013), the slopes converge on a smaller range (0.02 - 0.028), suggesting that the temperature dependency is likely more consistent in the natural environment and among different taxa from Group I and Group II haptophyte species. Differences in temperature dependency in cultures versus the natural environment have been observed for both marine and non-marine haptophyte taxa, which may be linked to differences in physiological conditions/status (e.g., Araie et al., 2018; Ono et al., 2012; Sun et al., 2007; Theroux et al., 2010). In addition, the y-intercept varies between culture and environmental calibrations (Figure 5). For example, the y-intercept for the Hap A culture calibration is -0.76 while it is -0.85 for the in situ calibration of Lake George (Figure 5). This pattern could imply that other lake properties such as nutrients or salinity interfere (Araie et al., 2018). Further studies will be needed to determine how exactly environmental factors control the y-intercept of the LCA temperature calibrations.

Taken together, these observations suggest that in situ calibrations might be better suited than culture calibrations for palaeotemperature reconstructions as they integrate natural variations in LCA distributions and in LCA-producers within a lake. For the Canadian lakes, the in situ calibration developed for Lake George is likely a suitable calibration for lakes hosting Group II haptophytes, while the in situ calibrations from Greenland, Alaska or Norway are all likely appropriate for lakes with Group I haptophytes. However, the development of site-specific calibrations for our studied lakes might still be important before confidently use the $U^{K}_{37}$ for temperature reconstructions.

## 5. Conclusions

In the present study, we used next generation sequencing (NGS) to identify the LCA producing haptophyte species in five lakes from the Canadian Prairies. Even though NGS can introduce substitution-type miscalls, three distinct LCA-producing OTUs were identified. Combined with previous data, our study shows that Group I LCA-producing haptophytes are

likely to be found in freshwater/oligohaline lakes in the Canadian Prairies, whereas Group II LCA-producing haptophytes, mainly included within the genus *Isochrysis*, are present in oligohaline/saline lakes. The Group I haptophytes produce a highly-specific LCA profile, while Group II haptophytes appear to produce multiple LCA profiles. This study thus highlights the importance of genomic analyses to unambiguously identify LCA-producing haptophyte species. While these data help selecting the most appropriate calibration for the $U_{37}^{K}$ proxy for down-core temperature reconstructions in the Canadian Prairies, our observations suggest that temperature dependency is likely consistent within different taxa from Group I and Group II haptophytes and that in situ calibrations might be better suited for palaeotemperature reconstructions. Further studies will be however needed to confirm this observation and to determine how other environmental factors control the LCA temperature calibration.

## Acknowledgments and Data

We thank two anonymous reviewers for their constructive comments that helped to significantly improve the quality of the manuscript. We thank Deirdre Bateson and Heather Haig for her help and assistance during the fieldtrip and Julie Russell for her help in the genomic laboratory. Mike Zwick is thanked for his help and insight during first stages of the genomic work. John Kenny is also thanked for useful discussions during the sequencing. This research was funded by the ERC project (637776) ALKENoNE to JLT. UZI is funded by a NERC Independent Research Fellowship (NE/L011956/1) and JMC by the EPSRC (EP/K038885/1). PRL was supported by the Canada Research Chair program. DNA sequences for the 20 OTUs are on the GenBank under accession numbers MK092726 to MK092745. The raw sequencing data are available on the European Nucleotide Archive under the study accession number: PRJEB32131 (http://www.ebi.ac.uk/ena/data/view/PRJEB32131). All the other data can be found in the manuscript or the Supplement Information.

## References

Araie, H., Nakamura, H., Toney, J. L., Haig, H. A., Plancq, J., Shiratori, T., et al. (2018). Novel alkenone-producing strains of genus *Isochrysis* (Haptophyta) isolated from Canadian saline lakes show temperature sensitivity of alkenones and alkenoates. *Organic Geochemistry*, *121*, 89–103. https://doi.org/10.1016/j.orggeochem.2018.04.008

Archer, J., Baillie, G., Watson, S., Kellam, P., Rambaut, A., & Robertson, D. (2012). Analysis of high-depth sequence data for studying viral diversity: a comparison of next generation sequencing platforms using Segminator II. *BMC Bioinformatics*, *13*(47). https://doi.org/10.1186/1471-2105-13-47

Bendif, E., Probert, I., Herve, A., Billard, C., Goux, D., Lelong, C., Cadoret, J. P., & Veron, B. (2011). Integrative taxonomy of the Pavlovophyceae (Haptophyta): A reassessment. *Protist*, *162*, 738–761. https://doi.org/10.1016/j.protis.2011.05.001

Brassell, S. C., Brereton, R. G., Eglinton, G., Grimalt, J., Liebezeit, G., Marlowe, I. T., et al. (1986). Palaeoclimatic signals recognized by chemometric treatment of molecular stratigraphic data. *Organic Geochemistry*, *10*, 649–660. https://doi.org/10.1016/S0146-6380(86)80001-8

Buckley, T. R., & Cunningham, C. W. (2002). The effects of nucleotide substitution model assumptions on estimates of nonparametric bootstrap support. *Molecular biology and evolution, 19,* 394-405.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods, 7*, 335–336. https://doi.org/10.1038/nmeth.f.303

Castañeda, I. S., & Schouten, S. (2011). A review of molecular organic proxies for examining modern and ancient lacustrine environments. *Quaternary Science Reviews, 30*, 2851–2891. https://doi.org/10.1016/j.quascirev.2011.07.009

Chu, G., Sun, Q., Li, S., Zheng, M., Jia, X., Lu, C., et al. (2005): Long-chain alkenone distributions and temperature dependence in lacustrine surface sediments from China. *Geochimica et Cosmochimica Acta*, *69*, 4985–5003. https://doi.org/10.1016/j.gca.2005.04.008

Conte, M. H., Sicre, M. A., Ruhlemann, C., Weber, J. C., Schulte, S., Schulz-Bull, D., & Blanz, T. (2006). Global temperature calibration of the alkenone unsaturation index ($U^{K'}_{37}$) in surface waters and comparison with surface sediments. *Geochemistry Geophysics Geosystems*, *7*(2), Q02005. https://doi.org/10.1029/2005GC001054

Coolen, M. J. L., Muyzer, G., Rijpstra, W. I. C., Schouten, S., Volkman, J. K., & Sinninghe Damsté, J. S. (2004). Combined DNA and lipid analyses of sediments reveal changes in Holocene haptophyte and diatom populations in an Antarctic lake. *Earth and Planetary Science Letters*, *223*, 225–239. https://doi.org/10.1016/j.epsl.2004.04.014

D'Amore, R., Ijaz, U. Z., Schirmer, M., Kenny, J. G., Gregory, R., Darby, A. C., et al. (2016). A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genomics*, *17*(55). https://doi.org/10.1186/s12864-015-2194-9

D'Andrea, W. J., Huang, Y., Fritz, S. C., & Anderson, N. J. (2011). Abrupt Holocene climate change as an important factor for human migration in West Greenland. *Proceedings of the National Academy of Sciences USA*, *108*(24), 9765–9769. https://doi.org/10.1073/pnas.1101708108

D'Andrea, W. J., Lage, M., Martiny, J. B. H., Laatsch, A. D., Amaral-Zettler, L. A., Sogin, M. L., & Huang, Y. (2006). Alkenone producers inferred from well-preserved 18S rDNA in Greenland lake sediments. *Journal of Geophysical Research Biogeosciences, 111*, G03013. https://doi:10.1029/2005JG000121

D'Andrea, W. J., Theroux, S., Bradley, R. S., & Huang, X. (2016). Does phylogeny control $U^{K}_{37}$-temperature sensitivity? Implications for lacustrine alkenone paleothermometry. *Geochimica et Cosmochimica Acta*, *175*, 168–180. https://doi.org/10.1016/j.gca.2015.10.031

de Leeuw, J. W., van der Meer, F. W., & Rijpstra, W. I. C. (1980). On the occurrence and structural identification of long chain unsaturated ketones and hydrocarbons in sediments. In A. D. Douglas & J. R. Maxwell (Eds.), *Advances in Organic Geochemistry* (pp. 211–217). Pergamon Press, Oxford.

de Vargas, C., Aubry, M.-P., Probert, I., & Young, J. (2007). Origin and Evolution of Coccolithophores: From Coastal Hunters to Oceanic Farmers. In G. F. Paul & H. K. Andrew (Eds.), *Evolution of Primary Producers in the Sea* (pp. 251-285). Academic Press, Burlington.

Dillon, J. T., Longo, W. M., Zhang, Y., Torozo, R., & Huang, Y. (2016). Identification of double-bond positions in isomeric alkenones from a lacustrine haptophyte. *Rapid Communications in Mass Spectrometry*, *30*, 112–118. https://doi.org/10.1002/rcm.7414

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, *10*, 996-998. https://doi:10.1038/nmeth.2604

Endo, H., Ogata, H., & Suzuki, K. (2018). Contrasting biogeography and diversity patterns between diatoms and haptophytes in the central Pacific Ocean. *Scientific Reports*, *8*, 10916. https://doi.org/10.1038/s41598-018-29039-9

Egge, E., Bittner, L., Andersen, T., Audic, S., de Vargas, C., & Edvardsen, B. (2013). 454 Pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: a test for marine haptophytes. *PLoS ONE*, *8*, e74371. https://doi.org/10.1371/journal.pone.0074371

Egge, E. S., Johannessen, T. V., Andersen, T., Eikrem, W., Bittner, L., Larsen, A., et al. (2015). Seasonal diversity and dynamics of haptophytes in the Skagerrak, Norway, explored by high-throughput sequencing. *Molecular Ecology*, *24*, 3026–3042. https://doi.org/10.1111/mec.13160

Gran-Stadniczeñko, S., Šupraha, L., Egge, E. D., & Edvarsen, B. (2017). Haptophyte Diversity and Vertical Distribution Explored by 18S and 28S Ribosomal RNA Gene Metabarcoding and Scanning Electron Microscopy. *Journal of Eukaryotic Microbiology*, *64*, 514–532. https://doi.org/10.1111/jeu.12388

He, Y. X., Liu, W. G., Zhao, C., Wang, Z., Wang, H. Y., Liu, Y., et al. (2013). Solar influenced late Holocene temperature changes on the northern Tibetan Plateau. *Chinese Science Bulletin*, *58*, 1053–1059. https://doi.org/10.1007/s11434-012-5619-8

Jordan R. W., & Chamberlain, A. H. L. (1997). Biodiversity among haptophyte algae. *Biodiversity and Conservation*, *6*, 131–152.

Jordan, R. W., Cros, L., & Young, J. R. (2004). A revised classification scheme for living haptophyte. *Micropaleontology*, *50*, 55–79. https://www.jstor.org/stable/4097103

Joshi, N. A., & Fass, J. N. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for fastq files (version 1.21), available at: https://github.com/najoshi/sickle.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*, 772–780. https://doi.org/10.1093/molbev/mst010

Kawachi, M., Inouye, I., Maeda, O., & Chihara, M. (1991). The haptonema as a food-capturing device: observations on *Chrysochromulina hirta* (Prymnesiophyceae). *Phycologia*, *30*, 563–573. https://doi.org/10.2216/i0031-8884-30-6-563.1

Kircher, M., Stenzel, U., & Kelso, J. (2009). Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biology*, *10*, R83. https://doi.org/10.1186/gb-2009-10-8-r83

Laird, K., Cumming, B., Wunsam, S., Rusak, J., Oglesby, R., Fritz, S. & Leavitt, P. R. (2003). Lake sediments record large-scale shifts in moisture regimes across the northern prairies of North America during the past two millennia. *Proceedings of the National Academy of Sciences*, *100*, 2483–2488. https://doi.org/10.1073/pnas.0530193100

Lanave, C., Preparata, G., Saccone, C., & Serio, G. (1984). A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution*, *20*(1), 86–93. https://doi.org/10.1007/BF02101990

Land, M., Hauser, L., Jun, S. R., Nookaew, I., Leuze, M. R., Ahn, T.-H., et al. (2015). Insights from 20 years of bacterial genome sequencing. *Functional & Integrative Genomics*, *15*, 141–161. https://doi.org/10.1007/s10142-015-0433-4

Liu, H., Probert, I., Uitz, J., Claustre, H., Aris-Brosou, S., Frada, M., et al. (2009). Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans.

*Proceedings of the National Academy of Sciences, 106*, 12803–12808. https://doi.org/10.1073/pnas.0905841106

Longo, W. M., Dillon, J. T., Tarozo, R., Salacup, J. M., & Huang, Y. (2013). Unprecedented separation of long-chain alkenones from gas chromatography with a poly(trifluoropropylmethylsiloxane) stationary phase. *Organic Geochemistry*, *65*, 94–102. https://doi.org/10.1016/j.orggeochem.2013.10.011

Longo, W. M., Huang. Y., Yao, Y., Zhao, J., Giblin, A. E., Wang, X., et al. (2018). Widespread occurrence of distinct alkenones from Group I haptophytes in freshwater lakes: Implications for paleotemperature and paleoenvironmental reconstructions. *Earth and Planetary Science Letters*, *492*, 239-250. https://doi.org/10.1016/j.epsl.2018.04.002

Longo, W. M., Theroux, S., Giblin, A. E., Zheng, Y., Dillon, J. T., & Huang, Y. (2016). Temperature calibration and phylogenetically distinct distributions for freshwater alkenones: Evidence from northern Alaskan lakes. *Geochimica et Cosmochimica Acta*, *180*, 177–196. https://doi.org/10.1016/j.gca.2016.02.019

Marlowe, I. T., Brassell, S. C., Eglinton, G., & Green, J. C. (1984). Long-chain unsaturated ketones and esters in living algae and marine sediments. *Organic Geochemistry*, *6*, 135–141. https://doi.org/10.1016/0146-6380(84)90034-2

Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G., & Neufeld, J. D. (2012). PANDAseq: paired-end assembler for Illumina sequences. *BMC Bioinformatics*, *13*(31). https://doi.org/10.1186/1471-2105-13-31

Müller, P. J., Kirst, G., Ruhland, G., von Storch, I., & Rosell-Melé, A. (1998). Calibration of the alkenone paleotemperature index based on core-tops from the eastern South Atlantic and the global ocean (60°N–60°S). *Geochimica et Cosmochimica Acta*, *62*(10), 1757–1772. https://doi.org/10.1016/S0016-7037(98)00097-0

Nakamura, H., Sawada, K., Araie, H., Shiratori, T., Ishida, K., Suzuki, I., & Shiraiwa, Y., 2016. Composition of long chain alkenones and alkenoates as a function of growth temperature in marine haptophyte *Tisochrysis lutea*. *Organic geochemistry*, *99*, 78–89. https://doi.org/10.1016/j.orggeochem.2016.06.006

Nakamura, H., Sawada, K., Araie, H., Suzuki, I., & Shiraiwa, Y. (2014). Long chain alkenes, alkenones and alkenoates produced by the haptophyte alga *Chrysotila lamellosa* CCMP1307 isolated from a salt marsh. *Organic geochemistry*, *66*, 90–97. https://doi.org/10.1016/j.orggeochem.2013.11.007

Nikolenko, S. I., Korobeynikov, A. I., & Alekseyev, M. A. (2013). BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics*, *14*, S7. https://doi.org/10.1186/1471-2164-14-S1-S7

Ono, M., Sawada, K., Shiraiwa, Y., & Kubota, M. (2012). Changes in alkenone and alkenoate distributions during acclimatization to salinity change in *Isochrysis galbana*: implication for alkenone-based paleosalinity and paleothermometry. *Geochemical Journal*, *46*, 235–247. https://doi.org/10.2343/geochemj.2.0203

Pahnke, K., & Sachs, J. P. (2006). Sea surface temperatures of southern midlatitudes 0–160 kyr B.P. *Paleoceanography*, *21*, PA2003. https://doi.org/10.1029/2005PA001191

Pham, S. V., Leavitt, P. R., McGowan, S., & Peres-Neto, P. (2008). Spatial variability of climate and land-use effects on lakes of the northern Great Plains. *Limnology and Oceanography*, *53*, 728–742. https://doi.org/10.4319/lo.2008.53.2.0728

Plancq, J., Cavazzin, B., Juggins, S., Haig, H. A., Leavitt, P. R., & Toney, J. L. (2018a). Assessing environmental controls on the distribution of long-chain alkenones in the Canadian

668 Prairies. *Organic Geochemistry*, *117*, 43–55.
669 https://doi.org/10.1016/j.orggeochem.2017.12.005
670 Plancq, J., McColl, J. L., Bendle, J. A., Seki, O., Couto, J. M., Henderson, A. C. G., Yamashita.
671 Y., Kawamura, K., & Toney, J. L. (2018b). Genomic identification of the long-chain alkenone
672 producer in freshwater Lake Toyoni, Japan: implications for temperature reconstructions.
673 *Organic Geochemistry*, *125*, 189–195. https://doi.org/10.1016/j.orggeochem.2018.09.011
674 Prahl, F. G. Muehlhausen, L. A., & Zahnle, D. L. (1988). Further evaluation of long-chain
675 alkenones as indicators of paleoceanographic conditions. *Geochimica et Cosmochimica Acta*,
676 *52*(9), 2303–2310. https://doi.org/10.1016/0016-7037(88)90132-9
677 Prahl, F. G. & Wakeham, S. G. (1987). Calibration of unsaturation patterns in long-chain ketone
678 compositions for paleotemperature assessment. *Nature*, *330*, 367–369.
679 https://doi.org/10.1038/330367a0
680 Prahl, F. G., Wolfe, G. V., & Sparrow, M. A. (2003). Physiological impacts on alkenone
681 paleothermometry. *Paleoceanography*, *18*(2), 1–7. https://doi.org/10.1029/2002PA000803
682 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA
683 ribosomal RNA gene database project: improved data processing and web-based tools.
684 *Nucleic Acids Research*, *41*, D590–D596. https://doi.org/10.1093/nar/gks1219
685 Randlett, M.-E., Coolen, M. J. L., Stockhecke, M., Pickarski, N., Litt, T., Balkema, C., et al.
686 (2014). Alkenone distribution in Lake Van sediment over the last 270 ka: influence of
687 temperature and haptophyte species composition. *Quaternary Science Reviews*, *104*, 53–62.
688 https://doi.org/10.1016/j.quascirev.2014.07.009
689 Rawson, D. S., & Moore, J. E. (1944). The saline lakes of Saskatchewan. *Canadian Journal of*
690 *Research*, *22d*, 141–201. https://doi.org/10.1139/cjr44d-011
691 Richter, N., Longo, W. M., George, S., Shipunova, A., Huang, Y., & Amaral-Zettler, L. (2019).
692 Phylogenetic diversity in freshwater-dwelling Isochrysidales haptophytes with implications
693 for alkenone production. *Geobiology*, 1-9. https://doi.org/10.1111/gbi.12330
694 Rodriguez, F., Oliver, J. L., Marin, A., & Medina, J. R. (1990). The general stochastic-model of
695 nucleotide substitution. *Journal of Theoretical Biology*, *142*(4), 485–501.
696 https://doi.org/10.1016/S0022-5193(05)80104-3
697 Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Hohna, S., et al. (2012).
698 MrBayes3.2: Efficient Bayesian phylogenetic inference and model choice across a large
699 model space. *Systematic Biology*, *61*(3), 539–542. https://doi.org/10.1093/sysbio/sys029
700 Rosell-Melé, A., Eglinton, G., Pflaumann, U., & Sarnthein, M. (1995). Atlantic core-top
701 calibration of the $U^{K}_{37}$ index as a sea-surface palaeotemperature indicator. *Geochimica et*
702 *Cosmochimica Acta*, *59*(15), 3099–3107. https://doi.org/10.1016/0016-7037(95)00199-A
703 Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., Sloan, W. T., & Quince, C. (2015). Insight into
704 biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform.
705 *Nucleic Acids Research*, *43*, e37. https://doi.org/10.1093/nar/gku1341
706 Shalchian-Tabrizi, K., Reier-Røberg, K., Ree, D. K., Klaveness, D., & Bråte, J. (2011). Marine
707 freshwater colonizations of Haptophytes inferred from phylogeny of environmental 18S
708 rDNA sequences. *Journal of Eukaryotic Microbiology*, *58*, 315–318.
709 https://doi.org/10.1111/j.1550-7408.2011.00547.x
710 Simon, M., López-García, P., Moreira, D., & Jardillier, L. (2013). New haptophyte lineages and
711 multiple independent colonizations of freshwater ecosystems. *Environmental Microbiology*
712 *Reports*, *5*, 322-332. https://doi.org/10.1111/1758-2229.12023

Sun, Q., Chu, G., Liu, G., Li, S., & Wang, X. (2007). Calibration of alkenone unsaturation index with growth temperature for a lacustrine species, *Chrysotila lamellosa* (Haptophyceae). *Organic Geochemistry, 38*, 1226–1234. https://doi.org/10.1016/j.orggeochem.2007.04.007

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lecture Notes on Mathematical Modelling in the Life Sciences, 17*, 57–86.

Theroux, S. (2013). Diversity and Ecology of Lacustrine Haptophyte Algae: Implications for Paleothermometry, (Doctoral dissertation). Retrieved from Brown Digital Repository . (https://repository.library.brown.edu/studio/item/bdr:320529/). Brown University Library. https://doi.org/10.7301/Z0K35S0S

Theroux, S., D'Andrea, W. J., Toney, J. L., Amaral-Zettler, L., & Huang, Y. (2010). Phylogenetic diversity and evolutionary relatedness of alkenone-producing haptophyte algae in lakes: Implications for continental paleotemperature reconstructions. *Earth and Planetary Science Letters, 300*, 311–320. https://doi.org/10.1016/j.epsl.2010.10.009

Theroux, S., Huang, Y. & Amaral-Zettler, L. (2012). Comparative molecular microbial ecology of the spring haptophyte bloom in a Greenland arctic oligosaline lake. *Frontiers in Microbiology, 3*. https://doi.org/10.3389/fmicb.2012.00415

Theroux, S., Toney, J. L., Amaral-Zettler, L., & Huang, Y. (2013). Production and temperature sensitivity of long chain alkenones in the cultured haptophyte *Pseudoisochrysis paradoxa*. *Organic Geochemistry, 62*, 68–73. https://doi.org/10.1016/j.orggeochem.2013.07.006

Toney, J. L., Huang, Y., Fritz, S. C., Baker, P. A., Grimm, E., & Nyren, P. (2010). Climatic and environmental controls on the occurrence and distributions of long-chain alkenones in lakes of the interior United States. *Geochimica et Cosmochimica Acta, 74*, 1563–1578. https://doi.org/10.1016/j.gca.2009.11.021

Toney, J. L., Theroux, S., Andersen, R. A., Coleman, A., Amaral-Zettler, L., & Huang, Y. (2012). Culturing of the first 37:4 predominant lacustrine haptophyte: Geochemical, biochemical, and genetic implications. *Geochimica et Cosmochimica Acta, 78*, 51–64. https://doi.org/10.1016/j.gca.2011.11.024

Tragin, M., Zingone, A., & Vaulot, D. (2018). Comparison of coastal phytoplankton composition estimated from the V4 and V9 regions of the 18S rRNA gene with a focus on photosynthetic groups and especially Chlorophyta. *Environmental Microbiology, 20*, 506–520. https://doi.org/10.1111/1462-2920.13952

van der Bilt, W. G. M., D'Andrea, W. J., Bakke, J., Balascio, N. L., Werner, J. P., Gjerde, M., & Bradley, R. S. (2018). Alkenone-based reconstructions reveal four-phase Holocene temperature evolution for High Arctic Svalbard. *Quaternary Science Reviews, 183*, 204–213. https://doi.org/10.1016/j.quascirev.2016.10.006

Versteegh, G. J. M., Riegman, R., de Leeuw, J. W., & Jansen, J. H. F. (2001). $U^{K'}_{37}$ values for *Isochrysis galbana* as a function of culture temperature, light intensity and nutrient concentrations. *Organic Geochemistry, 32*, 785–794. https://doi.org/10.1016/S0146-6380(01)00041-9

Vogt, R. J., Rusak, J. A., Patoine, A., & Leavitt, P. R. (2011). Differential effects of energy and mass influx on the landscape synchrony of lake ecosystems. Ecology, 92, 1104–1114. https://doi.org/10.1890/10-1846.1

Wang, Z., & Liu, W. G. (2013). Calibration of the $U^{K'}_{37}$ index of long-chain alkenones with the in-situ water temperature in Lake Qinghai in the Tibetan Plateau. *Chinese Science Bulletin, 58*, 803–808. https://doi.org/10.1007/s11434-012-5527-y

Wheaton, E. E, Arthur, L. M., Chorney, B., Shewchuk, S., Thorpe, J., Whiting, J., & Wittrock, V. (1992). The prairies drought of 1988. *Climatological Bulletin*, *26*, 188–205.

Wu, X., Liu, X., Wang, J., & Yamamoto, M. (2018). Climatic and ecological changes of the past 1900 years inferred from long-chain alkenones in Kusai Lake, northern Qinghai-Tibetan Plateau. *Quaternary International*, *487*, 33–42. https://doi.org/10.1016/j.quaint.2018.02.007

Zheng, Y., Huang, Y., Andersen, R. A., & Amaral-Zettler, L. A. (2016). Excluding the di-unsaturated alkenone in the $U^{K}_{37}$ index strengthens temperature correlation for the common lacustrine and brackish-water haptophytes. *Geochimica et Cosmochimica Acta*, *175*, 36–46. https://doi.org/10.1016/j.gca.2015.11.024

Zhou, J., Liu, X., Stones, D. S., Xie, Q., & Wang, G. (2011). MrBayes on a graphics processing unit. *Bioinformatics*, *27*, 1255-1261. https://doi.org/10.1093/bioinformatics/btr140

Zink, K.-G., Leythaeuser, D., Melkonian, M., & Schwark, L. (2001). Temperature dependency of long-chain alkenone distributions in recent to fossil limnic sediments and in lake waters. *Geochimica et Cosmochimica Acta*, *65*(2), 253–265. https://doi.org/10.1016/S0016-7037(00)00509-3

**Table 1.** In situ parameters and total LCA concentration measured at the studied lakes. Lake temperature, conductivity, salinity and pH values are mean values for the entire water column. Please note that data for Deadmoose Lake are from a survey conducted in 2013 (Plancq et al., 2018a).

| Lake | Lat. (°N) | Long. (°W) | Depth (m) | Lake Temperature (°C) | Dissolved Oxygen (%) | Conductivity (mS/cm) | Salinity (g/L) | pH | Total LCA (µg/g sed) |
|---|---|---|---|---|---|---|---|---|---|
| **Deadmoose** | 52.3 | 105.17 | 15.1 | 12.6 | 54.2 | 15.1 | 12.3 | 8.87 | 31.6 |
| **Dewey** | 50.58 | 107.07 | 2.7 | 20.8 | 6.73 | 11647.9 | 7.29 | 8.44 | 16.2 |
| **Lenore** | 52.45 | 105 | 10 | 18.4 | 9.95 | 2.15 | 1.27 | 7.35 | 12.9 |
| **Manitou** | 52.76 | 109.75 | 16.9 | 9.58 | 2.44 | 28132.5 | 25.2 | 9.51 | 1079.7 |
| **Middle** | 52.56 | 105.19 | 9.6 | 18.8 | 8.61 | 2.73 | 1.67 | 6.87 | 2184.5 |
| **Richmond** | 52.01 | 108.02 | 2.1 | 21.4 | 7.01 | 12704.1 | 7.89 | 8.48 | 20.8 |

## Figure captions

**Figure 1.** Map of the southern Saskatchewan, Canada showing the location of the studied lakes. The lakes indicated in grey are the lakes previously studied by Araie et al. (2018) and from which DNA data on haptophyte isolates and/or environmental samples are available. Deadmoose Lake was also studied by Araie et al. (2018).

**Figure 2.** LCA profiles found in the surface sediments of the studied Canadian lakes. Please note that the LCA profile from Deadmoose Lake was similar to the one from Manitou Lake and is thus not reported here.

**Figure 3.** Bar chart showing the abundance (in % of the total number of reads) of the 20 most abundant OTUs for each studied lakes and for the positive control 4054 (culture sample RCC 4054 *Isochrysis* sp.). These are the OTUs representing more than 2% of the total number of reads. OTUs were grouped according to nearest-neighbour relationships inferred from their phylogenetic tree placement (Figure 4).

**Figure 4.** A phylogenetic tree depicting 18S rRNA gene-inferred relationships among haptophyte algae. A Bayesian inference was used to generate this consensus tree from publicly available partial length sequence fragments. GenBank accession numbers for publically available sequences follow all species names and are based on NCBI taxonomy. The evolutionary distance for the number of changes per site is represented by the scale bar. Brackets mark the LCA-producing haptophyte Groups I, II, III, and the outgroup. The phylogenetic positions of the OTUs with more than 2% of the total number of reads in the studied lakes are shown in bold. The four most abundant OTUs (with more than 10% of the total reads) that are considered in the discussion are highlighted in bold and grey. The names of the lakes where each OTU is found are indicated in the brackets besides the OTU name. MID: Middle; LEN: Lenore; MAN: Manitou; DEW: Dewey; RIC: Richmond; DEA: Deadmoose; 4054: culture sample RCC 4054 *Isochrysis* sp.

**Figure 5.** Comparison of culture and in situ $U_{37}^{K}$ temperature calibrations reported in the literature. The equation for each calibration is reported to compare the slopes (see discussion).