Diversity and ice nucleation activity of *Pseudomonas syringae* in drone-based water samples from eight lakes in Austria

Regina Hanlon [1], **Celia Jimenez-Sanchez** [1], **James Benson** [1], **Ken Aho** [2], **Cindy Morris** [3], **Teresa M Seifried** [4], **Philipp Baloh** [4], **Hinrich Grothe** [4], **David Schmale** [Corresp. 1]

[1] School of Plant and Environmental Sciences, Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, Virginia, United States

[2] Department of Biological Sciences, Idaho State University, Pocatello, Idaho, United States

[3] Institut National de la Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE), Montfavet, France

[4] Faculty of Technical Chemistry, Vienna University of Technology, Institute of Materials Chemistry, Vienna, Austria

Corresponding Author: David Schmale
Email address: dschmale@vt.edu

Bacteria from the *Pseudomonas syringae* complex (comprised of at least 15 recognized species and more than 60 different pathovars of *P. syringae sensu stricto*) have been cultured from clouds, rain, snow, streams, rivers, and lakes. Some strains of *P. syringae* express an ice nucleation protein (hereafter referred to as ice+) that catalyzes the heterogeneous freezing of water. Though *P. syringae* has been sampled intensively from freshwater sources in the U.S. and France, little is known about the genetic diversity and ice nucleation activity of *P. syringae* in other parts of the world. We investigated the haplotype diversity and ice nucleation activity at -8°C (ice+) of strains of *P. syringae* from water samples collected with drones in eight freshwater lakes in Austria. A phylogenetic analysis of citrate synthase (*cts*) sequences from 271 strains of bacteria isolated from a semi-selective medium for *Pseudomonas* revealed that 69% (188/271) belonged to the *P. syringae* complex and represented 32 haplotypes in phylogroups 1, 2, 7, 9, 10, 13, 14 and 15. Strains within the *P. syringae* complex were identified in all eight lakes, and seven lakes contained ice+ strains. Partial 16S rDNA sequences were analyzed from a total of 492 pure cultures of bacteria isolated from non-selective medium. Nearly half (43.5%; 214/492) were associated with the genus *Pseudomonas*. Five of the lakes (ALT, GRU, GOS, GOL, and WOR) were all distinguished by high levels of *Pseudomanas* ($p \leq 0.001$). HIN, the highest elevation lake, had the highest percentage of ice+ strains. Our work highlights the potential for uncovering new haplotypes of *P. syringae* in aquatic habitats, and the use of robotic technologies to sample and characterize microbial life in remote settings.

Diversity and ice nucleation activity of *Pseudomonas syringae* in drone-based water samples from eight lakes in Austria

Regina Hanlon[1], Celia Jimenez-Sanchez[1], James Benson[1], Ken Aho[2], Cindy E. Morris[3], Teresa Seifried[4], Philipp Baloh[4], Hinrich Grothe[4], David G. Schmale III[1]*

[1]School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, Virginia 24061, USA

[2]Department of Biological Sciences, Idaho State University, Pocatello, Idaho 83209, USA

[3]INRAE, Pathologie Végétale, F-84140, Montfavet, France

[4]Institute of Materials Chemistry, Faculty of Technical Chemistry, Vienna University of Technology, Austria

*Corresponding author; dschmale@vt.edu / Phone 540-231-6943 / Fax 540-231-7477

ABSTRACT

Bacteria from the *Pseudomonas syringae* complex (comprised of at least 15 recognized species and more than 60 different pathovars of *P. syringae sensu stricto*) have been cultured from clouds, rain, snow, streams, rivers, and lakes. Some strains of *P. syringae* express an ice nucleation protein (hereafter referred to as ice+) that catalyzes the heterogeneous freezing of water. Though *P. syringae* has been sampled intensively from freshwater sources in the U.S. and France, little is known about the genetic diversity and ice nucleation activity of *P. syringae* in other parts of the world. We investigated the haplotype diversity and ice nucleation activity at -8°C (ice+) of strains of *P. syringae* from water samples collected with drones in eight freshwater lakes in Austria. A phylogenetic analysis of citrate synthase (*cts*) sequences from 271 strains of bacteria isolated from a semi-selective medium for *Pseudomonas* revealed that 69% (188/271) belonged to the *P. syringae* complex and represented 32 haplotypes in phylogroups 1, 2, 7, 9, 10, 13, 14 and 15. Strains within the *P. syringae* complex were identified in all eight lakes, and seven lakes contained ice+ strains. Partial 16S rDNA sequences were analyzed from a total of 492 pure cultures of bacteria isolated from non-selective medium. Nearly half (43.5%; 214/492) were associated with the genus *Pseudomonas*. Five of the lakes (ALT, GRU, GOS, GOL, and WOR) were all distinguished by high levels of *Pseudomanas* ($p \leq 0.001$). HIN, the highest elevation lake, had the highest percentage of ice+ strains. Our work highlights the potential for uncovering new haplotypes of *P. syringae* in aquatic habitats, and the use of robotic technologies to sample and characterize microbial life in remote settings.

Subjects Biodiversity, Freshwater Biology

38 **Keywords** *Pseudomonas syringae,* Ice nucleation, Diversity, Species Richness, Drone, Bacteria,

39 Austria, Lake[1]

40

INTRODUCTION

Bacteria in the genus *Pseudomonas* are ubiquitous in natural and managed environments (Berge et al., 2014). Members of the *Pseudomonas syringae* complex have been cultured from clouds, rain, snow, streams, rivers, and lakes. Some strains of *P. syringae* express an ice nucleation protein (hereafter referred to as ice+) that catalyzes the heterogeneous freezing of water (Morris et al., 2014). The ice+ phenotype has been used to compare *P. syringae* populations in a freshwater lake in Virginia, USA during different seasons (Pietsch, Vinatzer & Schmale, 2017). The *P. syringae* complex (sometimes referred to as *P. syringae sensu lato* (Bull et al., 2011)) is comprised of at least 15 recognized species and more than 60 different pathovars of *P. syringae sensu stricto* (Gomila et al., 2017; Gutiérrez-Barranquero, Cazorla & De Vicente, 2019). In this work, we will use the name *P. syringae* to refer to members of the complex.

Benson et al. (2019) reported a new method to collect microorganisms from freshwater lakes in Austria using a DrOne Water Sampling SystEm (DOWSE). Briefly, a sterile conical 50mL tube was inserted into a custom 3D-printed water sampler and carried by the drone on a 4.5m long nylon tether to target locations in each lake. Samples were collected by drone from three different distances from shore (1, 25 and 50 m), and microorganisms were cultured on two types of agar media (one was semi-selective for *Pseudomonas spp.*, and the other was a general growth medium). Here, we expand the work of Benson et al. (2019) by analyzing the haplotype diversity (based on the partial sequence of the citrate synthase housekeeping gene used to determine the phylogenetic context of strains (Berge et al 2014)) and ice nucleation activity of strains collected during the drone-based water sampling missions in eight lakes in Austria. It is important to note that the Benson et al. (2019) study focused on drone-based water sampling as a new method and approach for aquatic microbiology.

Based on the preliminary observations in Benson et al. (2019), we hypothesized that (1) that the haplotype structure of *P. syringae* and other culturable bacteria varies among the eight Austrian lakes and (2) greater numbers of ice-nucleating strains of *P. syringae* are observed in colder Austrian lakes at higher altitudes. To test these hypotheses, strains from Benson et al. (2019) were sub-cultured, sequenced, subjected to haplotype analyses, and tested for ice nucleation activity at -8°C. The specific objectives of this study were to: (1) investigate the haplotype structure of strains of *P. syringae* from the eight Austrian lakes based on partial sequences of the citrate synthases (*cts*) gene, (2) determine the ice nucleation activity of strains of *P. syringae* from the eight Austrian lakes using a droplet freezing assay, (3) examine the diversity of total culturable bacteria from the eight Austrian lakes based on partial sequences of 16s rDNA, and (4) investigate potential associations of populations of culturable bacteria among the eight Austrian lakes using a non-metric multidimensional diversity analysis (NMDS). Since the bacteria analyzed in this study were collected with drones, our work highlights the potential for robotic systems to monitor the diversity and life history of microorganisms in lake environments, particularly in remote alpine settings.

MATERIALS AND METHODS

Lake locations and sampling

Freshwater sampling was conducted at eight different lakes in Austria in June 2018, Altausseersee (ALT), Grundlsee (GRU), Toplitzsee (TOP), Vorderer Gosausee (GOS), Gosaulacke (GOL), Hinterer Gosausee (HIN), Ossiachersee (OSS) and Wörthersee (WOR), (Figure 1) (Benson et al., 2019). Sampling details and design of the Drone Water SamplEr (DOWSE) were previously described in Benson et al. (2019). Briefly, seven of the eight lakes

were sampled with a drone 1 m, 25 m, and 50 m from the shoreline. Samples were collected from a 30 x 50 m grid of nine collection sites in the lake. Due to the narrow width of GOL, this lake was sampled at 1 and 25 m from shore in four locations to form a 45 x 25 m grid of eight sample collections (Benson et al., 2019). Lake area (km^2) and elevation are listed in Table 1.

Lake sampling permissions and safety of drone operations

Permissions to sample the lakes were granted by the Austrian Federal Forests AG, DI Martin Heinz Stürmer, on 3 April 2018. A formal field collection permit was not required for this work. Sites for drone operations were carefully selected to be minimally intrusive to people in the area during the time of sampling. The drones used as part of this work were registered with the Federal Aviation Administration (FAA). The UAS pilot for the missions reported in this manuscript was a certified FAA Remote Pilot under Part 107, Certificate Number 4038906.

Collection and processing of microorganisms

Lake samples from Benson et al. (2019) were processed for storage on the same day of collection. Briefly, 100 mL from each location was filtered through a 0.2µm single use Pall filter funnel (VWR International, Radnor, PA, USA, #28143-542). Filters were transferred to 15 mL tubes and stored at 4°C. All samples were shipped on ice to VA, USA for culturing, storage, and downstream analyses. Details concerning the culturing of microorganisms for this study are described in Benson et al. (2019). Briefly, KBC agar plates were used to select for bacteria in the genus *Pseudomonas* and TSA agar plates were used to grow culturable bacteria (including *Pseudomonas*). While not all bacteria collected from the environment are culturable in a laboratory setting, we chose an agar plate composition that had been tested with precipitation

samples. Previous studies explored various agar plate compositions (including TSA) with regard to the culturability of environmental rain (Failor et al., 2017), and simulated rain samples (Hanlon et al., 2017).

Sequence analysis and identification (*cts* and 16S)

Pure strains were subjected to sequence analyses of the citrate synthase (*cts*) housekeeping gene (for strains on the semi-selective medium, KBC) or 16S rDNA (for strains cultured on TSA). Berge et al. (2014) used Multi Locus Sequence Typing (MLST) analysis of 216 *Pseudomonas syringae* strains to identify 23 clades and describe 13 phylogroups in the *Pseudomonas syringae* genetic complex. These authors showed that the *cts* housekeeping gene alone was 97% accurate in predicting phylogeny. Prior to sequencing, a 5 min. template boil was performed in 100 µL of water with a 1 mm toothpick stab of frozen stock or from a fresh colony streak plate. A 25 µL PCR reaction was carried out with 2 µL of template from the quick boil tube. The *cts* gene was amplified using GoTaq® Green Master Mix (Promega M712) with primers ctsFor (5′ CCC GTC GAG CTG CCA ATW CTG A 3′) and ctsRev (5′ ATC TCG CAC GGS GTR TTG AAC ATC 3′). The *cts* gene sequence data was used to confirm that strains belong to the *Pseudomonas syringae* complex and to determine the phylogroups within the complex. Conditions of the *cts* PCR reaction were previously described in Pietsch, Vinatzer & Schmale (2017). PCR products were confirmed on an 1% TBE agarose gel with SYBRTM Safe DNA Gel Stain (Invitrogen by Thermo Fisher Scientific). Reactions were cleaned enzymatically with rSAP and ExoI and Sanger sequencing reactions were performed by Eton Biosciences (AB1 3730 x 1 DNA Sequencer, Eton Biosciences, 104 T.W. Alexander Drive, Bldg 4A, RTP, NC 27709, USA). Sanger generated sequences were trimmed (using Phred Program, Version 0.020425.c;

Copyright (C) 1993-2002 by Phil Green and Brent Ewing) and subsequently queried against the NCBI nr (non-redundant) database. The same PCR parameters were used to amplify a portion of the 16S rDNA gene. The forward and reverse primers used were 518 F (5' CCA GCA GCC GCG GTA ATA CG 3') and 1491 R (5' ATC GGY TAC CTT GTT ACG ACT TC 3') (Turner et al., 1999). The identification of culturable bacteria was determined to the classification level of genus by using trimmed sequences as described above and sequences were submitted to GenBank.

Phylogenetic analysis of *P. syringae* complex strains in lakes

To specifically determine which of the strains isolated from KBC medium were within the *P. syringae* complex and to situate them phylogenetically, we compared, via BLAST, the *cts* sequences for these strains with the overall NCBI data base which contains the full diversity of this bacterium reported by Berge and colleagues (Berge et al., 2014). Strains for which the *cts* sequence did not have at least 90% sequence similarity with that of strains in the *P. syringae* complex were eliminated from subsequent analysis thereby facilitating alignment of the *cts* sequences along a greater number of bases than if all strains were included in the analysis. Four additional strains were eliminated because we could not sequence the *cts* gene to obtain a sufficient number of bases (>300). The resulting numbers of *P. syringae* complex sequences for each lake were 10 for ALT, 30 for GRU, 37 for TOP, 39 for GOS, 14 for GOL, 18 for HIN, 39 for OSS and 1 for WOR. The phylogenetic relationships among strains was determined as previously reported (Berge et al., 2014) based on 311 base pairs of the retained sequences and for 34 reference strains via the neighbor-joining algorithm in MEGA7 with 1,000 bootstrap

replicates. The total number of confirmed *P. syringae* complex strains from each lake is shown in Table 2.

Ice-nucleation activity of confirmed *P. syringae* complex strains

Confirmation of positive ice+ *Pseudomonas* samples that were tentative *P. syringae* (271 strains) was verified in three independent ice-nucleation assays and reported in Supplemental Table 1. Positive ice+ strains that were part of the *P. syringae* complex are listed in Table 2. The assays were performed on a boat made with PARAFILM® M (Sigma P6543, 20 in. x 50 ft) floating in a Lauda Alpha RA 12 (LCKD 4908) cooling bath (LAUDA-Brinkmann, LP, Delran, NJ, 08075) with ethylene glycol coolant fluid (Air gas RAD64000246). This parafilm boat-based assay has been described previously (Pietsch et al., 2016; Garcia et al., 2019). Pure cultures of bacteria were transferred from culture plates with a sterile toothpick to 140 µL of DI water into a 96-well plate. Plates were vortexed for 30 seconds and incubated at 4°C for 1 hour. Twelve-microliter droplets of the microbial suspension were loaded in duplicate at -6°C. The temperature of the cooling bath was decreased to -7°C, held for 2 minutes, then decreased to -8°C. After a final incubation time of 10 minutes at -8°C, all drops that froze were considered to be INA+ samples. The freezing temperature was recorded for each drop on the cryofloat.

Analysis of microbial richness and similarity

16S rDNA sequences were analyzed from a total of 492 pure cultures of bacteria isolated from TSA. Richness and Shannon diversity were characterized at all levels of taxa for 492 strains cultured on general enriched TSA agar. Variation in richness and diversity with distance to shorelines and lakes levels were analyzed with mixed effect models with lake treated as a

178 random effect. Random effects were tested using likelihood ratio tests whereas fixed effects were
179 tested using F-tests with Satterthwaite adjusted degrees of freedom.
180
181 **Population analyses**
182 Variation in the composition of genera of bacteria among and within lakes was analyzed using
183 non-metric multidimensional scaling (NMDS) (Shepard, 1962a,b; Kruskal, 1964a,b) and
184 PERMutational Multivariate ANOVA (PERMANOVA; Anderson, 2001). Dissimilarity matrices
185 for NMDS and PERMANOVA were created using Bray-Curtis dissimilarity (Bray & Curtis,
186 1957). The resulting 2D NMDS ordination was adequate for drawing inferences (stress = 0.14).
187 All analyses were performed using the R statistical environment (R core team 2019) with heavy
188 reliance on the packages vegan (Oksanen et al., 2019), lme4 (Bates et al., 2015), lmerTest
189 (Kuznetsova, Brockhoff & Christensen, 2017), and asbio (Aho, 2023).
190
191 **RESULTS**
192 **Culturing and identification of strains in the *P. syringae* complex**
193 415 strains recovered from KBC were identified as *Pseudomonas*. Sequences were submitted to
194 GenBank and assigned accession numbers (MW857572 - MW857985 and MW892633). Based
195 on specific sequence criteria relative to the *cts* gene, 271 strains were defined as tentative *P.*
196 *syringae* based on significant sequence homology. These strains were then analyzed using a
197 *Pseudomonas* specific database to determine if they were part of the *P. syringae* complex
198 defined by Berge et al. (2014). 188 strains were associated with the *P. syringae* complex and
199 represented 32 haplotypes in eight phylogroups (Fig. 2). Workflow is shown in Supplemental
200 Fig. 1.

The numbers of tentative *P. syringae* sequences for each lake were 26 for ALT, 53 for GRU, 53 for TOP, 41 for GOS, 24 for GOL, 21 for HIN, 49 for OSS and 4 for WOR for a total of 271 strains. A phylogenetic analysis of these strains showed that 188 strains grouped with *P. syringae* as part of the *P. syringae* complex, and the remaining 83 grouped with reference strain PAO1 for *P. fluorescens* (data not shown). The 188 strains in the *P. syringae* complex represented 32 haplotypes in phylogroups PG01, PG02, PG07, PG09, PG10, PG13, PG14 and PG15 of the *P. syringae* complex, for the portion of the *cts* gene that was sequenced. Among the 32 haplotypes, 11 were common to at least two lakes and five lakes had a total of 21 unique haplotypes, i.e haplotypes represented by only one strain in one lake (Figure 2).

The structure and size of the *P. syringae* complex phylogroups varied among lakes, and we found few correlations among the factors measured here. For example, neither the concentration nor the abundance of any of the haplotypes or phylogroups were significantly correlated with lake altitude (Spearman's rank correlation, $p<0.05$). PG01 and PG02 were the only phylogroups whose fractions in the population were significantly correlated with total concentrations of *P. syringae* complex strains in lakes; they increased as total concentrations increased (Spearman's rank correlation, $p<0.05$).

Ice nucleation activity of *P. syringae* strains

The numbers of tentative *Pseudomonas* strains that were tested for ice nucleation activity at -8°C (and % that froze) in the droplet freezing assay for each lake are included in Supplemental Table 1. There were 26 for ALT (12%), 50 for GRU (38%), 44 for TOP (18%), 40 for GOS (35%), 23 for GOL (26%), 20 for HIN (65%), 44 for OSS (30%) and 4 for WOR (0%) (Supplemental Table 1). The frequency of tentative *Pseudomonas* strains that were ice nucleation active (at -8°C)

varied among lakes (0 to 65%) and was significantly positively correlated with the fraction of phylogroup 2 strains in the population but with no other phylogroup component nor with the total abundance of tentative *Pseudomonas* strains in the lake (Spearman's Rank correlation, $p<0.01$).

Microbial richness and similarity among the eight lakes

Thirty-five unique genera of culturable bacteria were identified from the non-selective TSA plates across the eight lakes (7.34x10^4 cfu/ mL mean of 8 lakes). These strains were assigned accession numbers MN490845 to MN491336. Over forty percent of the strains (43.5% [214/492]) were associated with the genus *Pseudomonas*. Population richness (based on genera of bacteria) and Shannon diversity were analyzed for the set of 492 bacteria identified from the eight lakes. The highest richness was detected in TOP and OSS lakes, while the lowest richness was reported in GOS and GOL (Figure 3). *Pseudomonas* was the most abundant genus, and was present in all eight lakes at a mean percentage of 43.5% (214/492). The percentage of *Pseudomonas* ranged from 23% in HIN, to 64% in GOS. ALT, GRU, GOS, GOL and OSS had 50% or more *Pseudomonas* in the strains analyzed (Figure 4). The Shannon diversity analysis confirmed TOP and OSS to have the highest diversity. The analysis also showed more diversity in HIN relative to ALT, whereas the richness data did not distinguish a difference between HIN and ALT.

Population analysis of culturable bacteria

A PERMANOVA of samples within lakes indicated that bacterial assemblages varied strongly by lake (R2 = 0.58, $F_{!,\#\$} = 2.81$, p = 0.001, 999 permutations). Bacterial assemblages did not vary with distance to lake shoreline (R2 = 0.04, $F_{\%,\#\$} = 0.73$, p = 0.745, 999 permutations). An NMDS analysis showed relationships of bacterial assemblages within and among lakes (Figure 4). Particularly distinct were assemblages of TOP (lower right), HIN (center right) and OSS (top). In the 2D ordination for genus, the vector fitting p-values less than or equal to 0.004 ($p \leq 0.004$) showed the strongest association to the ordination as marked by arrows in Figure 5. ALT, GRU, GOS, GOL, and WOR were all distinguished by high levels of *Pseudomonas* ($p \leq 0.001$). WOR also showed a correlation to *Exiguobacterium* ($p \leq 0.036$). TOP was indicated by *Janthinobacterium* ($p \leq 0.001$), while OSS was indicated by *Pedobacter* ($p \leq 0.001$) and *Brevundimonas* ($p \leq 0.003$). HIN was correlated to two genera, *Flavobacterium* ($p \leq 0.015$) and *Massilia* ($p \leq 0.016$).

DISCUSSION

Knowledge of microbial ecology and functions in aquatic systems, and in particular alpine lakes, can be hampered by the difficulty to access lakes for sampling. Benson et al. (2019) reported a new method and approach to collect microorganisms from freshwater lakes in Austria using a Drone Water Sampling SystEm (DOWSE). Here, we expand the work of Benson et al. (2019) through a comprehensive analysis of the diversity and ice nucleation activity of strains collected during those drone-based missions. The work presented in this manuscript expands our understanding of the aquatic microbiology of remote alpine lakes in Austria, and adds to a growing body of literature on the biogeography and ubiquity of *Pseudomonas* in aquatic habitats

(Christner et al., 2008; Morris et al., 2008, 2010; Pietsch, Vinatzer & Schmale, 2017; Morris et al., 2022a).

High levels of *Pseudomonas* in five of the lakes supports the ubiquitous nature of *P. syringae* in aquatic environments. The set of 188 strains confirmed to belong to the *P. syringae* complex represented 32 haplotypes in eight phylogroups. Phylogroup 02 (PG02) is the most ubiquitous group, and includes strains that have been collected from a range of habitats, including rain, trees, and an irrigation basin (Berge et al., 2014). For our dataset, 43% (80/188) of the identified strains belong to PG02 and were collected from six of the eight lakes. The ice+ phenotype was reported at 85 % for non-pathogenic PG02 environmental strains tested by Berge et al. (2014). In our study, PG09 and PG07 were represented by 45 and 27 strains, respectively. PG09 is exclusively composed of aquatic habitat collections while PG07 represents strains responsible for pathogenicity in potato (Berge et al., 2014). PG13 was the fourth most abundant pathogroup with 24 strains represented by three lakes (OSS, GRU, TOP). Interestingly, PG13 is defined by strains isolated from non-plant sources, from wild alpine plants (Berge et al., 2014) and from five Icelandic habitats (Morris et al., 2022b).

Seven of the eight lakes had tentative *P. syringae* strains that were ice+. Lake locations can be seen in Figure 1, and lake characteristics including size, depth, and elevation can be seen in Table 1. Our hypothesis that samples from higher elevations would contain more ice+ bacteria was supported by our data; the lake with the highest elevation (HIN) had the most ice+ strains of all the lakes. HIN was also one of the coldest lakes at 17°C, only 1 degree warmer than GOL, the coldest lake. It is worth noting that GOL is only 0.04 km^2 in area, so temperature changes might be observed at different rates in smaller lakes such as GOL compared to larger lakes such an HIN (area of 0.31 km^2). Two of the three deepest lakes (GRU and GOS) had over 40% ice+

strains. These lakes were also more than double the altitude of WOR, the deepest lake and at the lowest elevation of the lakes studied.

A PERMANOVA of samples within lakes indicated that bacterial assemblages varied strongly by lake. This variation among lakes supports the notion that microbe composition is linked to sample source and immediate environmental conditions, including land-use of the surrounding area (Bowers et al., 2011). GOS, GOL, and HIN are located close to each other in Upper Austria (Figure 1). The straight-line distance between GOS and HIN is 4 kilometers with GOS and GOL below 1,000 meters in altitude and HIN above 1,100 meters (Table 1). HIN had a greater bacterial diversity than GOS and GOL.

Microbial assemblages did not vary with distance from shoreline ($R2 = 0.04$, $F_{2,13}=0.73$, $p = 0.745$, 999 permutations). However, high concentrations of microbes were observed in surface "hot spots" within a 30m X 50m sampling grid on the eight lakes (Benson et al., 2019). A "hot spot" on the surface included one or more of nine sampling locations in the 150m^2 sampling grid that had high concentrations of microbes. Wind across these hotspots could engender aerosolization and transport of microbes at elevated frequencies (Pietsch et al., 2018). Moreover, improved imaging and tracking techniques could be utilized in future work to predict hot spots or guide collection efforts to the center of the hot spots, such as visible scum on the water surface associated with high concentrations of toxic cyanobacteria commonly referred to as harmful algal blooms (HABs) (Tian et al., 2017; Ma et al., 2021).

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

314

315 **Author Contributions**

316 RH and CJ-S cultured microbes, collected and analyzed 16S sequence data, and led the writing
317 of the manuscript. RH, JB, PB, and DS participated in field collections. RH, CJ-S, and JB
318 conducted ice nucleation assays, managed the strain collection. KA conducted statistical analyses
319 using the 16S data and assisted in writing the manuscript. TMS helped organize permissions for
320 sampling lakes. CM conducted haplotype analyses from *cts* data, and assisted in writing the
321 manuscript. DS managed the project, designed experiments, was the pilot in command (PIC) for
322 all drone missions, analyzed data, and assisted in writing the manuscript. HG and DGS
323 participated in conceptualization, investigation, writing, supervision, project administration, and
324 funding acquisition.

325

326 **Funding Sources**

327 This research was supported by grants to D. Schmale from the Institute of Critical Technology
328 and Applied Science (ICTAS) at Virginia Tech (#177220), the College of Agriculture and Life
329 Sciences at Virginia Tech (#137605), and the National Science Foundation (NSF) under Grant
330 Numbers DEB-1241068 [Dimensions: Collaborative Research: Research on Airborne Ice-
331 Nucleating Species (RAINS)], AGS-1520825 (HAZARDS SEES: Uncovering the Hidden
332 Skeleton of Environmental Flows: Advanced Lagrangian Methods for Hazard Prediction,
333 Mitigation, and Response), and IIS-1637915 (NRI: Coordinated Detection and Tracking of
334 Hazardous Agents with Aerial and Aquatic Robots to Inform Emergency Responders). This
335 research was supported by grants to H. Grothe from the Austrian Science Fund (FWF) under
336 Grant P26040, and bridge project 850689 [EarlySnow] from the Austrian Research Promotion

Agency (FFG). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors of this research. The authors would like to thank Martin Heinz Stürmer (Österreichische Bundesforste AG) and the Austrian Federal Forests AG for the support and permission to sample the lakes.

REFERENCES

Aho K. 2023. asbio: A Collection of Statistical Tools for Biologists.

Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance: NON-PARAMETRIC MANOVA FOR ECOLOGY. *Austral Ecology* 26:32–46. DOI: 10.1111/j.1442-9993.2001.01070.pp.x.

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software* 67. DOI: 10.18637/jss.v067.i01.

Benson J, Hanlon R, Seifried T, Baloh P, Powers C, Grothe H, Schmale D. 2019. Microorganisms Collected from the Surface of Freshwater Lakes Using a Drone Water Sampling System (DOWSE). *Water* 11:157. DOI: 10.3390/w11010157.

Berge O, Monteil CL, Bartoli C, Chandeysson C, Guilbaud C, Sands DC, Morris CE. 2014. A User's Guide to a Data Base of the Diversity of Pseudomonas syringae and Its Application to Classifying Strains in This Phylogenetic Complex. *PLoS ONE* 9:e105547. DOI: 10.1371/journal.pone.0105547.

Bowers RM, McLetchie S, Knight R, Fierer N. 2011. Spatial variability in airborne bacterial communities across land-use types and their relationship to the bacterial communities of potential source environments. *The ISME Journal* 5:601–612. DOI: 10.1038/ismej.2010.167.

360 Bray JR, Curtis JT. 1957. An Ordination of the Upland Forest Communities of Southern
361 Wisconsin. *Ecological Monographs* 27:325–349. DOI: 10.2307/1942268.
362 Bull CT, Clarke CR, Cai R, Vinatzer BA, Jardini TM, Koike ST. 2011. Multilocus sequence
363 typing of Pseudomonas syringae sensu lato confirms previously described genomospecies
364 and permits rapid identification of P. syringae pv. coriandricola and P. syringae pv. apii
365 causing bacterial leaf spot on parsley. *Phytopathology* 101:847–858.
366 Christner BC, Morris CE, Foreman CM, Cai R, Sands DC. 2008. Ubiquity of Biological Ice
367 Nucleators in Snowfall. *Science* 319:1214. DOI: 10.1126/science.1149757.
368 Failor KC, Schmale DG, Vinatzer BA, Monteil CL. 2017. Ice nucleation active bacteria in
369 precipitation are genetically diverse and nucleate ice by employing different mechanisms.
370 *The ISME Journal* 11:2740–2753. DOI: 10.1038/ismej.2017.124.
371 Garcia EB, Hanlon R, Makris MR, Powers CW, Jimenez-Sanchez C, Karatum O, Marr LC,
372 Sands DC, Schmale DG. 2019. Microbial diversity of individual raindrops collected from
373 simulated and natural precipitation events. *Atmospheric Environment* 209:102–111. DOI:
374 10.1016/j.atmosenv.2019.04.023.
375 Gomila M, Busquets A, Mulet M, Garcia-Valdes E, Lalucat J. 2017. Clarification of taxonomic
376 status within the Pseudomonas syringae species group based on a phylogenomic analysis.
377 *Frontiers in microbiology* 8:2422.
378 Gutiérrez-Barranquero JA, Cazorla FM, De Vicente A. 2019. Pseudomonas syringae pv.
379 syringae associated with mango trees, a particular pathogen within the “hodgepodge” of
380 the Pseudomonas syringae complex. *Frontiers in Plant Science* 10:570.

Hanlon R, Powers C, Failor K, Monteil CL, Vinatzer BA, Schmale DG. 2017. Microbial ice nucleators scavenged from the atmosphere during simulated rain events. *Atmospheric Environment* 163:182–189. DOI: 10.1016/j.atmosenv.2017.05.030.

Kruskal JB. 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29:1–27. DOI: 10.1007/BF02289565.

Kruskal JB. 1964b. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29:115–129. DOI: 10.1007/BF02289694.

Kuznetsova A, Brockhoff PB, Christensen RHB. 2017. **lmerTest** Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82. DOI: 10.18637/jss.v082.i13.

Ma J, Jin S, Li J, He Y, Shang W. 2021. Spatio-Temporal Variations and Driving Forces of Harmful Algal Blooms in Chaohu Lake: A Multi-Source Remote Sensing Approach. *Remote Sensing* 13:427. DOI: 10.3390/rs13030427.

Morris CE, Conen F, Alex Huffman J, Phillips V, Pöschl U, Sands DC. 2014. Bioprecipitation: a feedback cycle linking Earth history, ecosystem dynamics and land use through biological ice nucleators in the atmosphere. *Global Change Biology* 20:341–351. DOI: 10.1111/gcb.12447.

Morris CE, Lacroix C, Chandeysson C, Guilbaud C, Monteil C, Piry S, Rochelle-Newall E, Fiorini S, Van Gijsegem F, Barny M-A, Berge O. 2022a. *Comparative seasonal abundance and diversity of populations of the Pseudomonas syringae and Soft Rot Pectobacteriaceae species complexes throughout the Durance River catchment from its French Alps sources to its delta*. Ecology. DOI: 10.1101/2022.09.06.506731.

Morris CE, Ramirez N, Berge O, Lacroix C, Monteil C, Chandeysson C, Guilbaud C, Blischke A, Sigurbjörnsdóttir MA, Vilhelmsson OÞ. 2022b. Pseudomonas syringae on Plants in

Iceland Has Likely Evolved for Several Million Years Outside the Reach of Processes That Mix This Bacterial Complex across Earth's Temperate Zones. *Pathogens* 11:357. DOI: 10.3390/pathogens11030357.

Morris CE, Sands DC, Vanneste JL, Montarry J, Oakley B, Guilbaud C, Glaux C. 2010. Inferring the Evolutionary History of the Plant Pathogen Pseudomonas syringae from Its Biogeography in Headwaters of Rivers in North America, Europe, and New Zealand. *mBio* 1:e00107-10. DOI: 10.1128/mBio.00107-10.

Morris CE, Sands DC, Vinatzer BA, Glaux C, Guilbaud C, Buffière A, Yan S, Dominguez H, Thompson BM. 2008. The life history of the plant pathogen Pseudomonas syringae is linked to the water cycle. *The ISME Journal* 2:321–334. DOI: 10.1038/ismej.2007.113.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H. 2019.Vegan: community ecology package. *Available at https://www.mcglinnlab.org/publication/2019-01-01_oksanen_vegan_2019/* (accessed July 26, 2023).

Pietsch RB, Grothe H, Hanlon R, Powers CW, Jung S, Ross SD, Schmale Iii DG. 2018. Wind-driven spume droplet production and the transport of *Pseudomonas syringae* from aquatic environments. *PeerJ* 6:e5663. DOI: 10.7717/peerj.5663.

Pietsch RB, Hanlon R, Bohland C, Schmale DG. 2016. Turning into Ice. *Science Teacher* 83:37–43. DOI: 10.2505/4/tst16_083_09_37.

Pietsch RB, Vinatzer BA, Schmale DG. 2017. Diversity and Abundance of Ice Nucleating Strains of Pseudomonas syringae in a Freshwater Lake in Virginia, USA. *Frontiers in Microbiology* 8. DOI: 10.3389/fmicb.2017.00318.

Shepard RN. 1962a. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika* 27:125–140. DOI: 10.1007/BF02289630.

Shepard RN. 1962b. The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika* 27:219–246. DOI: 10.1007/BF02289621.

Tian D, Xie G, Tian J, Tseng K-H, Shum CK, Lee J, Liang S. 2017. Spatiotemporal variability and environmental factors of harmful algal blooms (HABs) over western Lake Erie. *PLoS ONE* 12:e0179622. DOI: 10.1371/journal.pone.0179622.

Turner S, Pryer KM, Miao VPW, Palmer JD. 1999. Investigating Deep Phylogenetic Relationships among Cyanobacteria and Plastids by Small Subunit rRNA Sequence Analysis. *The Journal of Eukaryotic Microbiology* 46:327–338. DOI: 10.1111/j.1550-7408.1999.tb04612.x.

FIGURE AND TABLE LEGENDS (IN CITATION ORDER)

Figure 1. Map showing eight lakes sampled in Austria. ALT, GRU, and TOP were sampled on field day 1, GOS, GOL, HIN on field day 2, and OSS and WOR on field day 3. These lake groups are shown as blue, red, and green label location pins, respectively.

Table 1. Lake name, abbreviation, GPS location, size, altitude, depth at 50m from shore, and temperature for all lakes sampled during the campaign.

449 **Table 2.** Lake name abbreviation, collection date, number of confirmed strains belonging to the
450 *P. syringae* complex from each lake, number of strains assayed for ice+ (163 of 188), number of
451 frozen samples, and percent of frozen strains confirmed to belong to the *P. syringae* complex.
452
453 **Figure 2.** Haplotype diversity of 188 strains in the *P. syringae* complex from eight lakes in
454 Austria. Solid colored bars indicate the fraction of haplotypes (HT) in each lake. Colors represent
455 the haplotypes of different phylogroups (PG) that were in common to at least two lakes.
456 Haplotypes in black (a total of 21 HT) were only present in one of the eight lakes. Hashed bars
457 indicate the total population concentration of colonies selected on KBC in each lake (10^3 cfu/
458 mL).
459
460 **Figure 3.** Richness (a) and Shannon diversity (b) of lakes. Lake locations were significantly
461 different with respect to both richness (4.41, $p = 0.036$) and Shannon diversity (4.2, $p = 0.04$)
462 after controlling for distance to shoreline. The trend for level of richness and diversity was the
463 same for the eight lakes. Error bars are SEMs.
464
465 **Figure 4.** Bar graph of diversity at the classification level of genus for microbial growth on
466 general growth TSA media (total 492 sequences with 35 unique genera identified). *P. syringae*
467 was the only microbe present in all eight lakes. Taxa in bar plots are stacked from most abundant
468 (at the bottom of bars) to least abundant (at the tops of bars), and taxa names in legend also
469 follow this convention.
470

471 **Figure 5.** Two dimensional NMDS projection of lake genera data (stress = 0.14). Vector fitting
472 of genera with significant correlations with the projection p-values ($p \leq 0.04$) are marked by
473 arrows. The arrows point toward the region of most rapid increase in the projection and arrow
474 length is scaled by the strength of the correlation.
475

Figure 1

Figure 1.

Map showing eight lakes sampled in Austria. ALT, GRU, and TOP were sampled on field day 1, GOS, GOL, HIN on field day 2, and OSS and WOR on field day 3. These lake groups are shown as blue, red, and green label location pins, respectively.

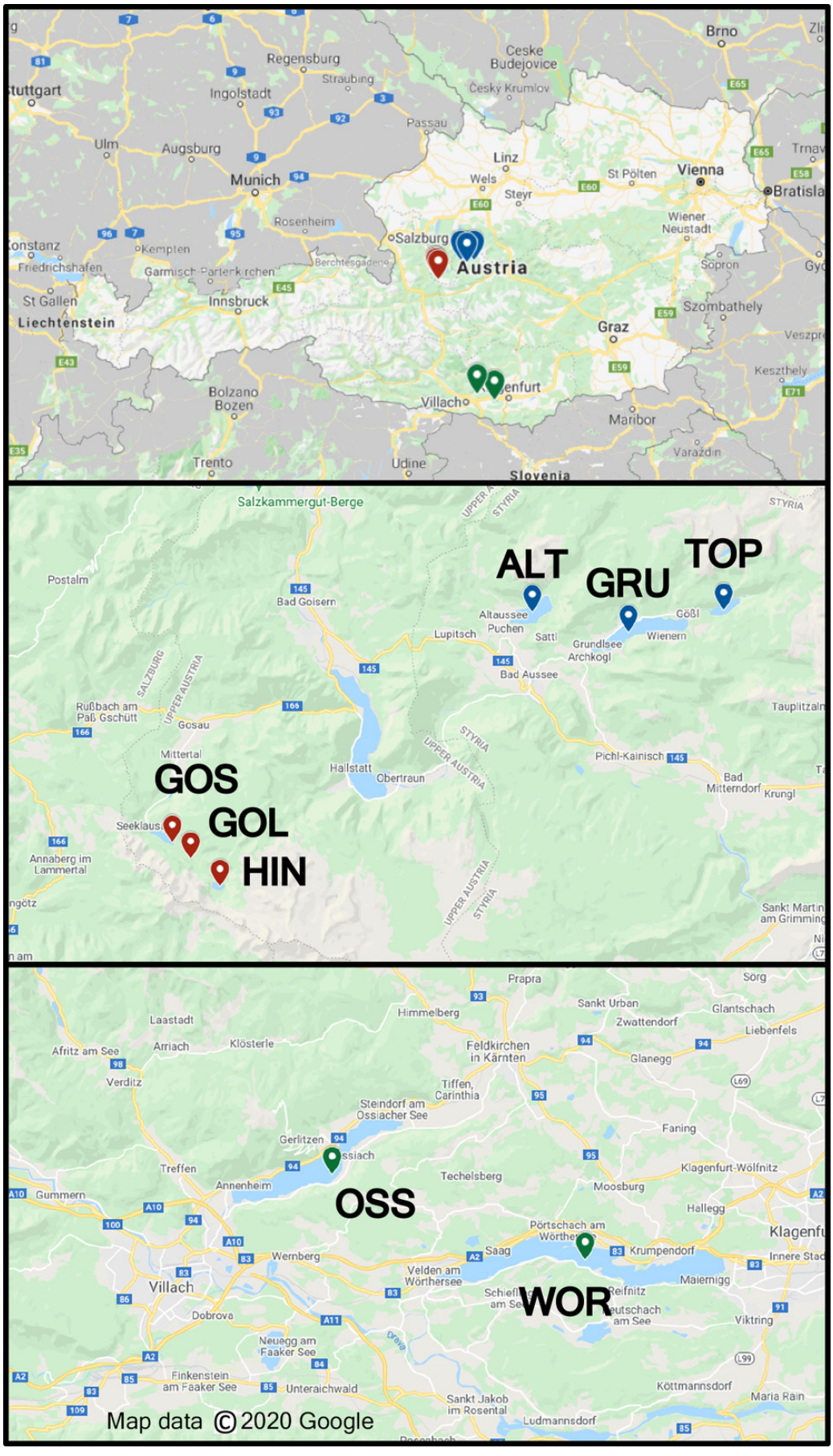
Brno
Regensburg
Ceske Budejovice
Straubing
Ingolstadt
Český Krumlov
Passau
Ulm
Augsburg
Linz
Munich
Wels
St Pölten
Vienna
Steyr
Rosenheim
Wiener Neustadt
Salzburg
Austria
Sopron
Innsbruck
St Gallen
Liechtenstein
Szombathely
Graz
Bolzano Bozen
Villach
Maribor
Trento
Udine
Slovenia
Salzkammergut-Berge
ALT
GRU
TOP
Postalm
Bad Goisern
Altaussee
Lupitsch
Puchen
Sattl
Grundlsee
Archkogl
Wienern
Gößl
Bad Aussee
Rußbach am Paß Gschütt
Gosau
Mittertal
GOS
Hallstatt
Obertraun
Pichl-Kainisch
Bad Mitterndorf
Krungl
GOL
HIN
Annaberg im Lammertal
Prapra
Sörg
Himmelberg
Sankt Urban
Glantschach
Zwattendorf
Laastadt
Arriach
Klösterle
Feldkirchen in Kärnten
Liebenfels
Afritz am See
Glanegg
Verditz
Tiffen, Carinthia
Steindorf am Ossiacher See
Faning
Gerlitzen
Treffen
Annenheim
Techelsberg
Klagenfurt-Wölfnitz
Gummern
OSS
Moosburg
Hallegg
Pörtschach am Wörther
Wernberg
Saag
Krumpendorf
Velden am Wörthersee
Maiernigg
Villach
Reifnitz
WOR
Dobrova
Viktring
Neuegg am Faaker See
Finkenstein am Faaker See
Unteraichwald
Köttmannsdorf
Sankt Jakob im Rosental
Maria Rain
Ludmannsdorf
Map data ©2020 Google

Figure 2

Figure 2.

Haplotype diversity of 188 strains in the *P. syringae* complex from eight lakes in Austria. Solid colored bars indicate the fraction of haplotypes (HT) in each lake. Colors represent the haplotypes of different phylogroups (PG) that were in common to at least two lakes. Haplotypes in black (a total of 21 HT) were only present in one of the eight lakes. Hashed bars indicate the total population concentration of colonies selected on KBC in each lake (10^3 cfu/ mL).

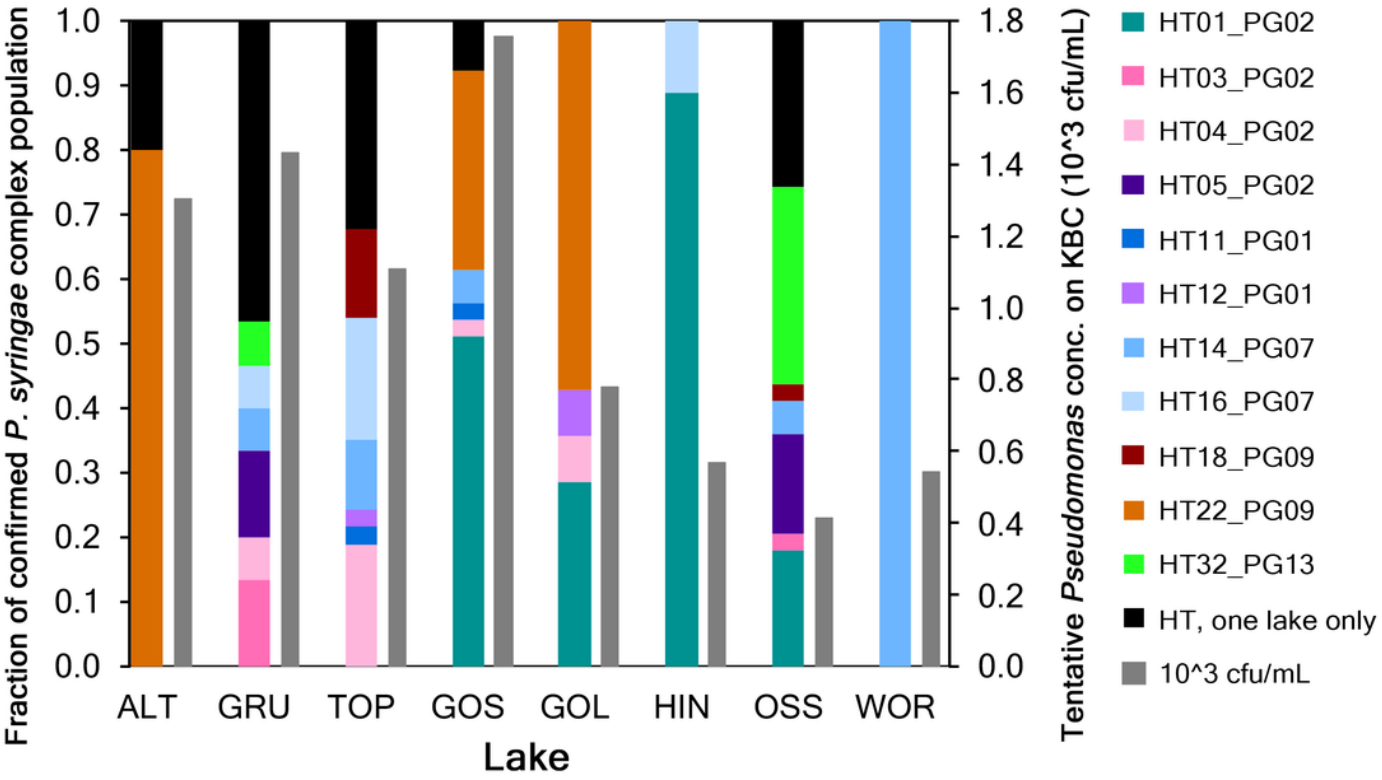

Figure 3

Figure 3.

Richness (a) and Shannon diversity (b) of lakes. Lake locations were significantly different with respect to both richness (4.41, p = 0.036) and Shannon diversity (4.2, p = 0.04) after controlling for distance to shoreline. The trend for level of richness and diversity was the same for the eight lakes. Error bars are SEMs.

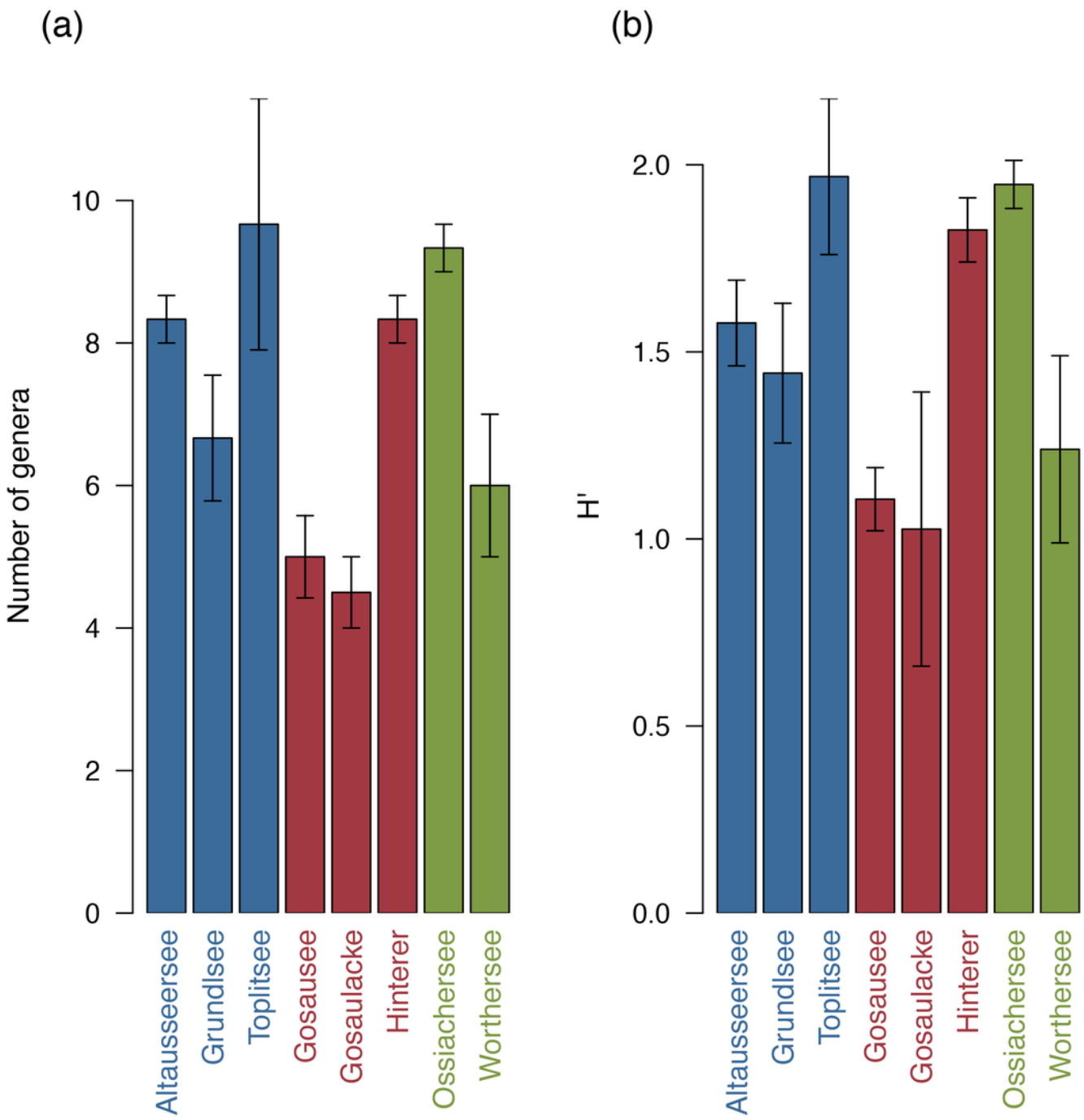
(a)
Number of genera
0
2
4
6
8
10
Altausseersee
Grundlsee
Toplitsee
Gosausee
Gosaulacke
Hinterer
Ossiachersee
Worthersee
(b)
H'
0.0
0.5
1.0
1.5
2.0
Altausseersee
Grundlsee
Toplitsee
Gosausee
Gosaulacke
Hinterer
Ossiachersee
Worthersee

Figure 4

Figure 4.

Bar graph of diversity at the classification level of genus for microbial growth on general growth TSA media (total 492 sequences with 35 unique genera identified). *P. syringae* was the only microbe present in all eight lakes. Taxa in bar plots are stacked from most abundant (at the bottom of bars) to least abundant (at the tops of bars), and taxa names in legend also follow this convention.

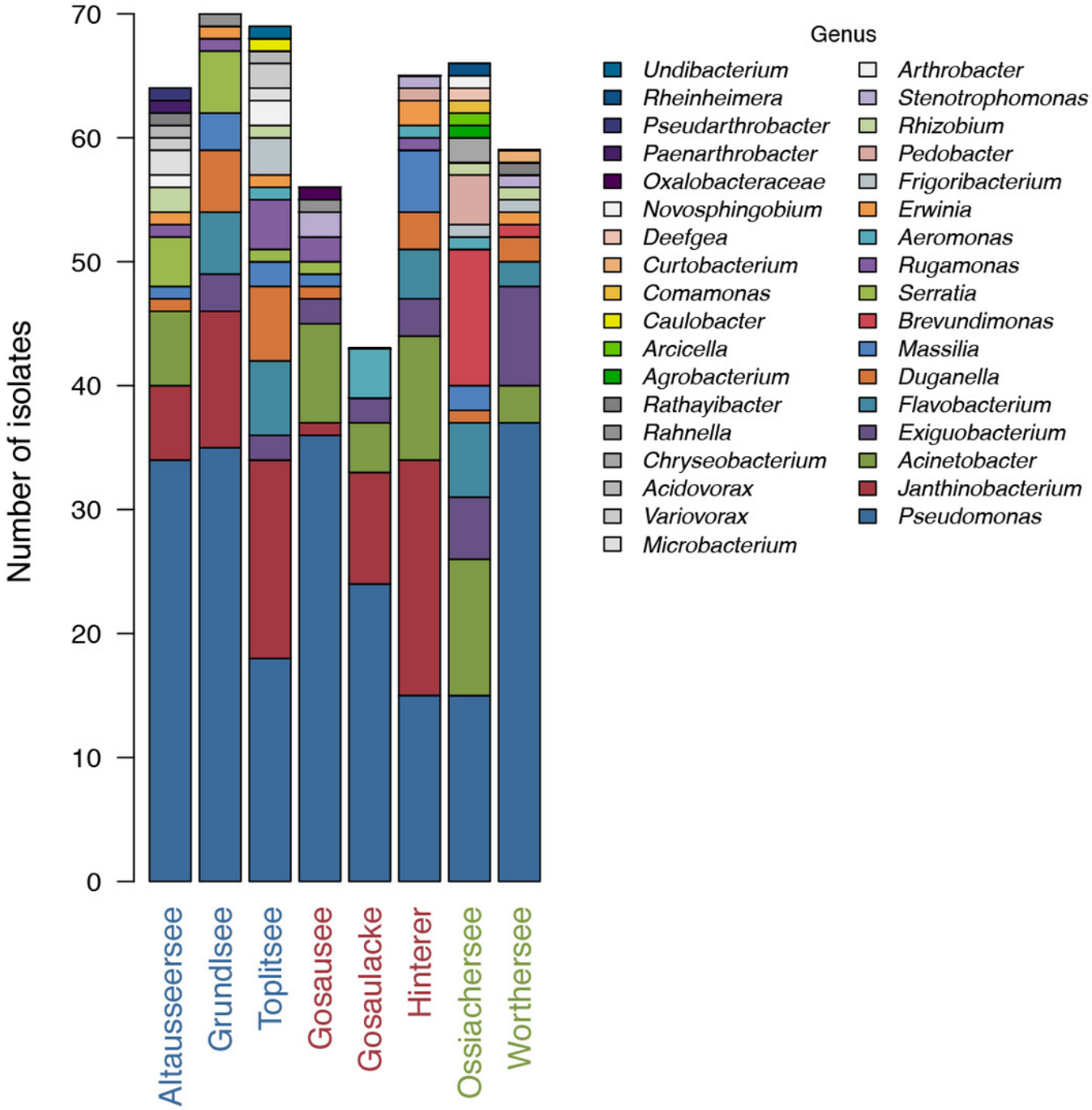
Number of isolates
0
10
20
30
40
50
60
70
Altausseersee
Grundlsee
Toplitsee
Gosausee
Gosaulacke
Hinterer
Ossiachersee
Worthersee
Genus
Undibacterium
Rheinheimera
Pseudarthrobacter
Paenarthrobacter
Oxalobacteraceae
Novosphingobium
Deefgea
Curtobacterium
Comamonas
Caulobacter
Arcicella
Agrobacterium
Rathayibacter
Rahnella
Chryseobacterium
Acidovorax
Variovorax
Microbacterium
Arthrobacter
Stenotrophomonas
Rhizobium
Pedobacter
Frigoribacterium
Erwinia
Aeromonas
Rugamonas
Serratia
Brevundimonas
Massilia
Duganella
Flavobacterium
Exiguobacterium
Acinetobacter
Janthinobacterium
Pseudomonas

Figure 5

Figure 5.

Two dimensional NMDS projection of lake genera data (stress = 0.14). Vector fitting of genera with significant correlations with the projection p-values ($p \leq 0.04$) are marked by arrows. The arrows point toward the region of most rapid increase in the projection and arrow length is scaled by the strength of the correlation.

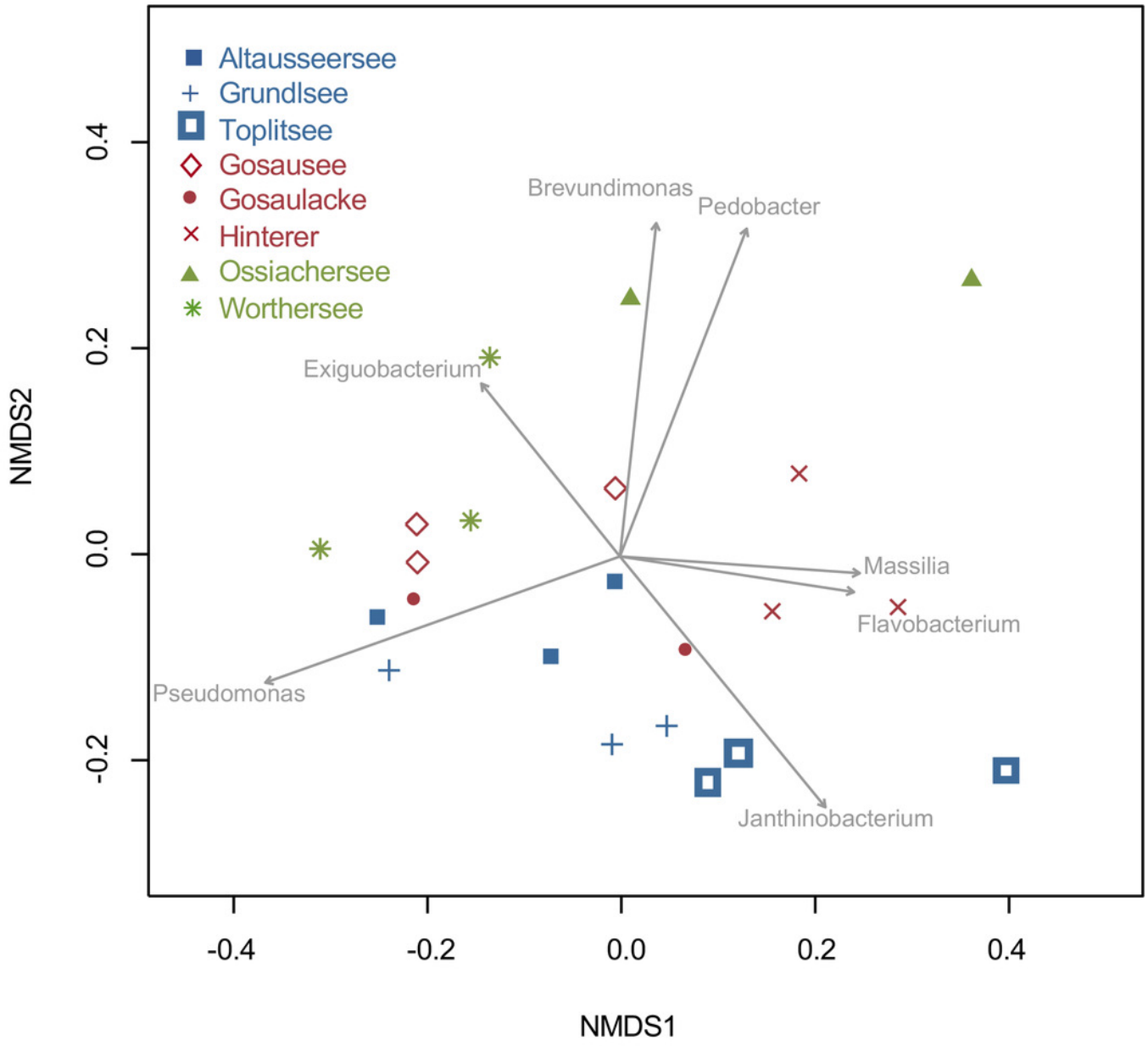
Altausseersee
Grundlsee
Toplitsee
Gosausee
Gosaulacke
Hinterer
Ossiachersee
Worthersee
Brevundimonas
Pedobacter
Exiguobacterium
Massilia
Flavobacterium
Pseudomonas
Janthinobacterium
NMDS1
NMDS2
-0.4
-0.2
0.0
0.2
0.4

Table 1(on next page)

Table 1.

Lake name, abbreviation, GPS location, size, altitude, depth at 50m from shore, and temperature for all lakes sampled during the campaign.

Lake Name	Abbreviation	Latitude and Longitude	Approximate Area (km2)	Elevation Above Sea Level (meters)	Depth 50m from shore (m)	Temperature (C)
Altaussee	ALT	47°38'27.3"N 13°47'04.2"E	2.1	712	3.4	21
Grundlsee	GRU	47°37'51.1"N 13°51'23.2"E	4.1 (5)	708	21.3	19
Toplitzsee	TOP	47°38′30.0″N 13°55′40.0″E	0.54	718	16	22
Vorderer Gosausee	GOS	47°31'28.1"N 13°30'52.7"E	0.67	917	16.2	21
Gosaulacke	GOL	47°30'58.3"N 13°31'42.8"E	0.04	978	4.2	16
Hinterer Gosausee	HIN	47°30'07.8"N 13°33'00.1"E	0.31	1151	11.2	17
Ossiachersee	OSS	46°39'54.7"N 13°58'01.8"E	10.8	502	8.6	25
Wörthersee	WOR	46°37'19.6"N 14°09'06.0"E	19.4	440	24.8	26

1

Table 2(on next page)

Table 2.

Lake name abbreviation, collection date, number of confirmed strains belonging to the *P. syringae* complex from each lake, number of strains assayed for ice+ (163 of 188), number of frozen samples, and percent of frozen strains confirmed to belong to the *P. syringae* complex.

Abbreviation	Date of Collection	Confirmed as part of the *P. syringae* complex based on HT analysis	Strains Tested for ice+	Total Ice+	Percent Ice+
ALT	June 7 2018	10	10	1	10%
GRU	June 7 2018	30	30	13	43%
TOP	June 7 2018	37	28	8	29%
GOS	June 8 2018	39	30	13	43%
GOL	June 8 2018	14	13	4	31%
HIN	June 8 2018	18	17	13	76%
OSS	June 10 2018	39	34	13	38%
WOR	June 10 2018	1	1	0	0%
All Lakes	Total	188	163	65	

1