# The poplar pathogen *Sphaerulina populicola* is genetically diverse and ubiquitously present in symptom-free host trees

**Fiona Paul** Corresp., [1] , **Imke Schmitt** [1, 2] , **Jürgen Otte** [2] , **Sunil Mundra** [2] , **Miklós Bálint** Corresp. [2]

[1] Institut für Ökologie, Evolution und Diversität, Johann Wolfgang Goethe Universität Frankfurt am Main, Frankfurt am Main, Germany

[2] Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main, Germany

Corresponding Authors: Fiona Paul, Miklós Bálint
Email address: fiona.paul@tuebingen.mpg.de, mbalint@senckenberg.de

Understanding the genetic variability and dispersal capabilities of plant pathogens is important to better predict pathogen spread and virulence, especially under global change conditions. The outcome of host-pathogen interactions directly depends on the dispersal capacity of the pathogens, on environmental conditions, and disease resistance of the host. Here, we analysed the distribution of finely resolved phylogenetic lineages of a poplar-specific fungal pathogen (*Sphaerulina populicola*) in western North America, across a 3,200 km latitudinal gradient of the host tree's (*Populus balsamifera*) distribution. We specifically tested whether dispersal limitations or environmental filters limit the spread and establishment of the pathogen into new areas. We assessed the genetic diversity of the pathogen with ITS1 oligotypes, recorded by metabarcoding leaf-associated fungal communities. The distribution of the recorded 16 *S. populicola* oligotypes showed no geographic patterns, indicating the lack of dispersal limitation of the pathogen throughout the investigated area. Climatic conditions also did not seem to restrict the occurrence and abundance of the pathogen oligotypes. Finally, we found strong variation in oligotype presence within single localities, which may suggest the importance of biotic factors in regulating the infection of this well-dispersing fungal pathogen. Our interpretation implies that the future prevalence of *S. populicola*-related disease will likely depend on the dynamics of host-pathogen interactions, since the pathogen does not seem dispersal-limited at a continental scale.

# The poplar pathogen *Sphaerulina populicola* is genetically diverse and ubiquitously present in symptom-free host trees

**Running title: Ubiquitous distribution of a poplar pathogen**

Fiona Paul[1+], Imke Schmitt[1,2], Jürgen Otte[2], Sunil Mundra[2]*, Miklós Bálint[2]*

[1]Institut für Ökologie, Evolution und Diversität, Goethe-Universität Frankfurt, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany.

[2] Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Senckenberganlage 25, 60325 Frankfurt am Main, Germany.

* shared senior authors. M. Bálint: mbalint@senckenberg.de; S. Mundra: sunil.mundra@unis.no, current address: Section for Genetics and Evolutionary Biology (EVOGENE), Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, 0316 Oslo, Norway

[+] current address: Department of Molecular Biology, Max Planck Institute for Developmental Biology, Max-Planck-Ring 5, 72076 Tübingen, Germany, fiona.paul@mpg.tuebingen.de

## 1. Abstract

Understanding the genetic variability and dispersal capabilities of plant pathogens is important to better predict pathogen spread and virulence, especially under global change conditions. The outcome of host-pathogen interactions directly depends on the dispersal capacity of the pathogens, on environmental conditions, and disease resistance of the host. Here, we analysed the distribution of finely resolved phylogenetic lineages of a poplar-specific fungal pathogen (*Sphaerulina populicola*) in western North America, across a 3,200 km latitudinal gradient of the host tree's (*Populus balsamifera*) distribution. We specifically tested whether dispersal limitations or environmental filters limit the spread and establishment of the pathogen into new areas. We assessed the genetic diversity of the pathogen with ITS1 oligotypes, recorded by metabarcoding leaf-associated fungal communities. The distribution of the recorded 16 *S. populicola* oligotypes showed no geographic patterns, indicating the lack of dispersal limitation of the pathogen throughout the investigated area. Climatic conditions also did not seem to restrict the occurrence and abundance of the pathogen oligotypes. Finally, we found strong variation in oligotype presence within single localities, which may suggest the importance of biotic factors in regulating the infection of this well-dispersing fungal pathogen. Our interpretation implies that the future prevalence of *S. populicola*-related disease will likely depend on the dynamics of host-pathogen interactions, since the pathogen does not seem dispersal-limited at a continental scale.



**Keywords:**

biogeography, ecological filters, endophyte, genetic diversity, metabarcoding, oligotyping,

## 2. Introduction

Disease outbreaks pose major threats to forest ecosystems (Sturrock et al., 2011). Forests affected by abiotic environmental stress, e.g., wind damage, drought, etc. are particularly vulnerable because weakened trees are more susceptible to pathogen infestation (Desprez-Loustau et al., 2007). Environmental stressors related to global change have intensified and increased in frequency during the last years, which has led to more frequent and more severe disease outbreaks (Sturrock et al. 2011, Pautasso et al. 2015, Desprez-Loustau et al. 2016). Climate change is predicted to affect the distribution and severity of forest diseases (Desprez-Loustau et al. 2007, Shaw & Osborne 2011, Sturrock et al. 2011). For example, diseases will have higher establishment probabilities, due to enhanced overwintering conditions or dispersal facilitated by human activities (Shaw & Osborne, 2011; Sturrock et al., 2011). In addition, in some areas pathogens will encounter more favourable conditions under climate change, such as mild winters (Desprez-Loustau et al., 2007).

Climate-related disease outbreaks already decimated forests (e.g., Solheim & Krokene 1998, Berg & Henry 2003). Overall, about 3.8 million hectares of forest per year (2000-2005) were affected by diseases and more than 36 million hectares by insects (FAO, 2010). The mountain pine beetle, *Dendroctonus ponderoseae*, destroyed more than 16 million ha of pine forest in Canada and the western United States over more than 10 years (Carroll et al. 2003, FAO 2010). Probably due to milder winter temperatures, the beetle was able to disperse into previously unaffected regions leading to an unprecedented epidemic (Carroll et al. 2003, Kurz et al. 2008). *Sphaerulina musiva* is another important forest fungal pathogen causing leaf spot disease in eastern cottonwood, *Populus deltoides.* Throughout the 20$^{th}$ century, this fungus has expanded its range and increased its host breadth to further species, such as *P. trichocarpa*, *P.*

*balsamifera* and *Salix lucida* (Newcombe et al. 2001, Callan et al. 2007, LeBoldus et al. 2009, Herath et al. 2016). *S. musiva* is also considered the most severe disease of hybrid poplars threatening poplar plantation areas of 44,128 ha in Canada and 45,000 ha in the US (FAO, 2012).

Disease-related damages to forest ecosystems depend on the ability of pathogens for efficient spread and infection. However, the biogeographic and environmental factors that influence pathogen distribution and establishment are not well understood. Theoretical frameworks suggest that a number of abiotic and biotic filters determine the distribution of species and the assembly of communities (Martiny et al. 2006, HilleRisLambers et al. 2012), and an organism may establish in the local species pool, if it can overcome these filters. The first filter acts on dispersal: an organism is present in a local community if it can disperse to there from the regional species pool. Despite small sizes and assumed high dispersal capacities many microbial taxa seem to be dispersal limited (Taylor et al. 2006, Hanson et al. 2012, Tedersoo et al. 2014). However, many pathogens may have large ranges when compared to other functional groups in fungi (Meiser et al. 2014, Tedersoo et al. 2014, Murray et al. 2015). Environmental conditions represent the second set of filters for local establishment. Even if an organism is able to arrive at a location, the local environmental conditions have to be favourable for survival. Finally, once the organism has overcome the dispersal and environmental filters, it has to deal with the other species already present in the local community. These interactions can lead to complex feedback mechanisms, by environmental conditions influencing the outcome of interactions (Chamberlain et al. 2014, Smith & Reynolds 2015).

The disease triangle (Agrios, 2005; McNew, 1960), another concept in phytopathology, states an additional factor, which contributes to the development of disease: a susceptible host (in addition to a suitable pathogen, and a favourable environment). Indeed, the genetic background

of the host influences fungal community composition in poplar (Bálint et al., 2013), and resistance genes have been described in poplar species, which enhance tolerance towards abiotic and biotic stressors (Newcombe & Bradshaw, 1996; Su et al., 2011). Host genetic composition could thus be viewed as additional filter that must be overcome by pathogens for successful development of disease.

In the present study, however, we focus only on the pathogen and the environment. We used the framework of ecological filters to investigate the biogeography of an opportunistic fungal pathogen, *Sphaerulina populicola* Peck (synonyms: *Septoria populicola, Mycosphaerella populicola*), associated with a continentally important boreal tree, *Populus balsamifera*. We hypothesized that local communities of the pathogen are filtered: i) by dispersal limitation caused by the same biogeographic barriers as in the host's case, ii) and/or by climatic factors including temperature and precipitation.

To date most studies have investigated microbial diversity patterns using global sequence similarity thresholds clustering sequences into OTUs (Huse, Welch, Morrison, & Sogin, 2010). This however can group potentially unrelated taxa and skew interpretations since the OTUs do not necessarily represent biologically meaningful units (Eren et al., 2013; Patin, Kunin, Lidström, & Ashby, 2013; Ryberg, 2015). Recent studies indicate the significance of investigating below OTU-level diversity, revealing ecologically or even clinically meaningful patterns previously overlooked (Eren et al., 2013; Eren, Morrison, et al., 2015). For example, the oligotyping approach revealed habitat preference of bacteria in the human oral cavity, host-specificity of bacteria in the genus *Blautia* in animal and human faeces or epidemiological significance of oligotypes of *Gardnerella vaginalis* causing bacterial vaginosis (Eren et al., 2011; Eren, Borisy, Huse, & Mark Welch, 2014; Eren, Sogin, et al., 2015). This approach has been

used in bacterial 16S studies, however it remains largely untested in ITS-based fungal community studies. Here we analysed the geographic distribution of *S. populicola* genetic diversity using oligotyping with a taxonomic unit resolution within conventional 97% sequence similarity OTUs.

## 3. Material & Methods

### 3.1. Study system

Balsam poplar is a North-American foundation tree species covering temperate and boreal habitats and ranging from Northern USA, over Canada to Northern Alaska (Burns & Honkala, 1990). It is an economically important species, harvested for wood, pulp, paper, fuel and possibly for new forms of bioenergy (Polle, Janz, Teichmann, & Lipka, 2013). Studies of the genetic diversity revealed different genotypes in different geographic demes reflecting adaptation to climatic conditions along the latitudinal gradient of its range (Keller et al. 2010, Keller et al. 2011). *S. populicola* is the causal agent of leaf spot disease in balsam poplar and black cottonwood (*Populus trichocarpa*) (Zalasky, 1978). Its distribution is assumed to be sympatric with the two host species (Zalasky 1978, Newcombe et al. 2001). Infections lead to losses through defoliation and eventually death (Newcombe et al. 2001, Busby et al. 2014). Poplars are known to harbour asymptomatic infections of *Sphaerulina sp* (Ostry & McNabb, 1983). *Sphaerulina populicola* is able to survive in poplar leaves as an asymptomatic endophyte. Such latent pathogens may cause disease once certain conditions are met (Johnson & Oelmüller, 2009), e.g., when environmental conditions lead to a stressed and less defensive host plant (Desprez-Loustau et al., 2016) and/or if the pathogen encounters a genetically disease susceptible

host plant. An earlier study indicated that *S. populicola* could occur in geographic localities more than 2,000 km apart (Bálint et al., 2015).

### 3.2 Sampling

We sampled trees of balsam poplar at 12 locations along a latitudinal gradient of about 3,200 km from Southern Canada to Northern Alaska (see Table S1 in Supporting Information). Three subarctic sampling sites are located in Alaska (see Table S1 in Supporting Information). Trees were haphazardly sampled within an area of about 0.5-2.0 $km^2$. At all sampling locations 30 trees were sampled except Arctic Village (10 trees) and Portage (25 trees). For each tree, one symptomless leaf was collected and put into a paper envelope in a zip lock bag containing silica gel for drying.

### 3.3 DNA extraction, PCR and sequencing

From the collected leaves, discs of 8mm in diameter were cut out using a pritchel. The leaf surface was sterilized for 1 min in 4% sodium hydrochloride solution (NaClO) and it was subsequently washed twice for 1min in sterile water containing 0.1% of Tween 20 to break surface tension. Subsequently, the leaf material was dried in the hood for 3 hours. The leaf material was grinded in a TissueLyser (Qiagen, Hilden, Germany). The DNA extraction was performed according to the CTAB method (Cubero, Crespo, Fatehi, & Bridge, 2002). Three technical replicates as well as three biological replicates were ran for two samples. Additionally, 23 negative DNA extraction controls were run in parallel.

In a first PCR reaction, the Internal Transcribed Spacer (ITS1) marker gene region was amplified with primers ITS1FI2 (Schmidt et al., 2013) and ITS2 (White, Bruns, Lee, & Taylor, 1990). These primers contained the Illumina sequencing primers (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG and

GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG). The 25 µL reactions contained 0.65 U TaKaRa ExTaq polymerase (Clontech Laboratories, Inc. USA), 2.5 µL buffer (10 x), 15 µL water, 2.0 µL dNTP mixture (2.5 mM each), ~ 10 ng genomic DNA (0.5 µL), 0.5 µL (0.22 µM) forward and reverse primer and 2 µL (0.86 µM) blocking primer Poplar-b1-F. The blocking primer represses the amplification of poplar ITS1 products (Bálint et al., 2015). The cycling conditions were an initial denaturation for 4min at 95°C, then 35 cycles of 95°C for 20s, 51°C and 56°C respectively for 20s, 72°C for 20 s followed by a final elongation of 3min at 72°C. We used two different annealing temperatures in two separate PCR reactions for each sample to capture a higher proportion of the biodiversity (Schmidt et al., 2013). After this step, the two products were pooled. We performed SPRI-purification of the PCR product with the Agencourt AMPure XP beads (Beckman Coulter GmbH, Krefeld, Germany).

In a second PCR, primers containing part of the Illumina sequencing primer, an 8 bp nucleotide index as well as the Illumina plate adapter were amplified to both ends of the amplicons. This was done similarly as described in Bálint et al. (2017) and following the Illumina 16S metabarcoding protocol (Illumina, 2016). This way the amplicons contain a unique 8 bp index on each end, which eliminates index jumping events during the library preparation (Schnell, Bohmann, & Gilbert, 2015). The 25 µL PCR reactions contained 0.65 U TaKaRa ExTaq polymerase, 2.5 µL buffer (10 x), 9 µL water, 2.0 µL dNTP mixture (2.5 mM each), 8 µL purified amplicon and 0.75 µL of each forward and reverse primer. PCR conditions were as follows: initial denaturation for 4min at 95°C, then 8 cycles with 30s at 95°C, 30s at 55°C, 30s at 72°C and a final extension for 3 min at 72°C. The PCR products were purified using SPRI-beads. The DNA concentration was measured with the Qubit™ 3.0 Fluorometer (Life Technologies, Germany), then the DNA concentration was normalized, and samples were

pooled. We performed paired-end sequencing (2 x 300 bp) on an Illumina MiSeq sequencer at StarSEQ® GmbH, Mainz, Germany.

**3.4 Sequence processing**

From the sequencing run, we received the reads from 344 trees at 12 sampling locations, demultiplexed into one forward and reverse read file for each sample. The reads were processed based on the pipeline developed by Bálint et al. (2014) (see Table S2 in Supporting Information).

We trimmed all sequences to 200 base pairs (bp) using the FASTX-TOOLKIT 0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/index.html; for summary of processing steps, see Table S2 in Supporting Information). Further, all reads with a minimum average quality below 26 phred score were discarded. Paired-end reads were assembled with PANDASEQ 2.9 (Masella, Bartram, Truszkowski, Brown, & Neufeld, 2012) using a minimum overlap of 25bp. The reads were stripped of forward and reverse primers with FQGREP 0.4.4 (https://github.com/indraniel/fqgrep) allowing no mismatches in the primer sequences. We removed reads containing primer artefacts, i.e., further primer sequences contained in the reads after removal of the primers. The fungal ITS1 region was extracted with ITSX 1.0.11 (Bengtsson-Palme et al., 2013), and subsequently sequences shorter than 100 bp were removed. We retained a set of high quality fungal ITS1 amplicon reads, from which we extracted the subset of reads relevant to the study purpose.

Target reads for this study are sequences assigned to *S. populicola* at 97% sequence similarity through the UNITE database (Kõljalg et al., 2013). A total of 291 *S. populicola* sequences are currently stored in the UNITE database (release date 22 August 2016) and were used to search against the query sequences with usearch_global from VSEARCH 2.0.2 (Rognes, Flouri, Nichols, Quince, & Mahé, 2016), an algorithm that employs global pairwise alignment.

### 3.5 Oligotyping

Most metabarcoding studies use global sequence similarity thresholds and clustering algorithms to group sequences into taxonomic units (Huse et al., 2010). These OTUs often consist of homogeneous taxa. Higher resolution of taxonomic units can be achieved by investigating variable positions in the sequenced locus. Haplotypes are often delimited based on SNPs occurring within multiple loci. Another approach that can be used in high-throughput metabarcoding approaches of microbial communities promising higher resolution is oligotyping. For the analysis of intraspecific diversity of *S. populicola* we used MINIMUM ENTROPY DECOMPOSITION (MED, Eren et al. 2015), which is an automated algorithm based on oligotyping (Eren et al., 2013). The algorithm iteratively searches for highly variable nucleotide positions within an alignment of marker gene amplicon reads and decomposes the dataset into nodes that contain reads with the same nucleotides at these highly variable positions. The algorithm works independently of global sequence similarity thresholds and concentrates on phylogenetically informative positions. Sequencing or PCR errors should be randomly distributed throughout the gene region and are therefore not considered for the taxonomic grouping of sequences. We obtained oligotypes by running MINIMUM ENTROPY DECOMPOSITION 2.2 with default settings.

### 3.6 Statistical analysis

All statistical analysis were performed with R 3.3.3 (R Core Team, 2017). We used the 'plyr' package (Wickham 2011; for all statistical analysis, see GitHub repository) to subtract the maximum read count of an oligotype of any negative control from the read count of each sample for the oligotype. We also removed rare observations in two steps: first, we set oligotype observations to zero in a sample if the read count of the oligotype fell below 0.1% of the highest

count of that oligotype (Valentini et al., 2016). Second, we set all oligotype observations in a sample to zero if they fell below a threshold of 0.05% of the overall highest read count of the dataset (Pansu et al., 2015). Additionally, we discarded all oligotypes if they occurred in less than five trees. Oligotype observations per tree fit well the expected increase in richness with increasing sequencing depth (Bálint et al. 2016) except for a few outlier samples, which we subsequently removed.

Climatic data for the sampling locations was retrieved from the WorldClim database (http://www.worldclim.org/). We used the current climate dataset with monthly means from 1950-2000 at 30 arc sec resolution. Extraction of climatic data and formatting was done using the R scripts provided by Kathryn G Turner (https://gist.github.com/kgturner/6643334, https://gist.github.com/kgturner/6644150). We used the mean temperature and precipitation for the months of January and July and calculated the mean annual temperature and precipitation (see Table S1 in Supporting Information). Possible correlations between explanatory variables were checked with the 'corrplot' package (Wei & Simko, 2016). This revealed that especially latitude strongly correlated with many other climatic variables, such as temperature and precipitation (see Fig. S1 in Supporting Information). Thus, we retained only latitude as explanatory variable in the exploratory models (for all statistical analysis, see GitHub repository).

We used the Hill's series of diversity, as implemented in the 'vegan' package (Oksanen et al., 2016), as response variables to evaluate oligotype diversity patterns. The Hill's series of diversity consists of three indices, differently scaling rare and abundant taxa (Hill, 1973). Hill's N0 is species richness, meaning the total number of present taxa. Hill's N1 is the antialgortihm of the Shannon diversity index and Hill's N2 is the inverse Simpson diversity, giving more

weight to more abundant species and therefore measuring community evenness. Single variables influencing diversity patterns were first investigated with generalized linear models assuming negative binomial distributions (glm.nb) as implemented in the 'MASS' package (Venables & Ripley 2002). We built a final mixed effects generalized linear model assuming a negative binomial distribution (glmer.nb) of the response variable using the 'lme4' package (Bates, Mächler, Bolker, & Walker, 2015). We included sequencing depth, sampling location, number of replicates per sampling location, geographic deme and latitude as explanatory variables and tested their significance in the models. The diversity models can be written as follows: Hill.1 ~ (scale(reads)) + replicates/region/latitude + (1|site). Sequencing depth and sampling location accounted for a large part of the variation in preliminary generalized linear models, and were therefore added as fixed and random effect to the mixed effects models respectively (hill.1 ~ reads + site, reads: Df= 1, deviance= 18.242, res. Df= 59, res Dev.= 66.545, $p < 0.001$, site: Df= 11, deviance= 35.824, res. Df= 48, res. Dev.= 30.721, $p < 0.001$). We performed model selection using ANOVA and AIC criteria.

We investigated the abundance patterns of oligotypes using the 'mvabund' package (Wang, Naumann, Wright, Eddelbuettel, & Warton, 2016). The same variables as mentioned for the diversity patterns were tested for significance in influencing the abundance patterns. The abundance models can be written as follows: abundance ~ reads/site/replicates/region/latitude. We used manyglm generalized linear models with negative binomial distribution to evaluate compositional differences in oligotype assemblages. This fits a model on each oligotype and summarizes model results with consideration to potential correlations among oligotypes (Wang et al., 2016). We tested the models using ANOVA with likelihood-ratio test and 1000 bootstrap

iterations. We used non-metric multidimensional scaling to visualise the oligotype compositional differences among the trees as implemented in the 'vegan' package (Oksanen et al., 2016).

For visualization of the genetic distances of the oligotypes, we constructed a median-joining network using the most abundant sequence of each oligotype. We first calculated pairwise differences as distances based on the Kimura model with the 'ape' package (Paradis, Claude, & Strimmer, 2004). The median-joining network was generated based on the calculated differences using the 'pegas' package (Paradis, 2010).

## 4. Results

Of 9,074,044 raw reads, we retained 5,524,137 reads after sequence processing (see Table S2 in Supporting Information). Of these reads, 116,603 were assigned to *S. populicola* at 97% global sequence similarity (see Table S2 in Supporting Information). After filtering of rare sequence types, the oligotype abundance table consisted of 16 oligotypes in 61 tree samples, covering all 12 different sampling locations (see GitHub repository, also see Fig. S2 in Supporting Information). This corresponds to 36,144 total reads contained in the abundance table used for statistical analysis. Two oligotypes were highly abundant, accounting for more than 90% of the total sequences (see Fig. S2 in Supporting Information). Four oligotypes had abundances of 300 to 800 reads throughout the studied locations, while the others were very rare (see Fig. S2 in Supporting Information).

The variation in the diversity measures of the Hill's series was explained by sequencing depth and sampling location. Any variables that were further included in the model did not increase the variation explained and the influence of the variables were not statistically significant (Table 1). None of the explanatory variables explained community composition and

the abundances of any of the oligotypes statistically significantly (Table 2). Sequencing depth did not explain variation in abundance, but sampling location showed a marginally significant influence on abundance in the community model (Table 2).

The oligotype diversity (Fig. 2, Table 1) and assemblage composition (Table 2, also see Fig. S3 in Supporting Information) is not significantly different between the Alaskan population and the Western Canadian population. Most oligotypes are present in many sampling locations, and especially in geographically distant localities along the latitudinal gradient (Fig. 1a).

When including latitude to the mixed effects GLM including sampling location as a random effect and sequencing depth as a fixed effect, the variation explained did not increase and the influence of latitude was not statistically significant (Table 1). When investigating abundance patterns of oligotypes, adding latitude did not increase the explanatory power of the model (Table 2).

We observed high variation of oligotype composition within sampling locations as well as among sampling locations (Fig. 1a, also see Fig. S4 in Supporting Information). Distances between samples of the same sampling locations were often greater than distances of samples originating from different locations (see Fig. S4 in Supporting Information).

## 5. Discussion

Perhaps the most striking result of the study is the almost complete lack of phylogeographic patterns in *S. populicola* oligotypes. This is in spite of the species' huge range, the presence of biogeographic barriers and stark environmental differences, and well-known phylogeographic structures (S. R. Keller et al., 2010) and local adaptation (S. R. Keller et al.,

316 2011) of the balsam poplar hosts. We found evidence for the presence of the pathogen at each of
317 the investigated locations and observed no clear phylogeographic structures in the intraspecific
318 genetic diversity of the pathogen (Fig. 1). Although there were differences in oligotype
319 abundance among some locations (Table 2), these are likely not the results of dispersal limitation
320 (due to distance or biogeographic barriers), since every haplotype is present in both south-central
321 Canada and about 3,000 km north-west in Alaska.

322 One explanation for the striking lack of biogeographic patterns is an advanced capability
323 for dispersal. This is not rare for microorganisms - these were once considered to possess
324 unlimited dispersal capabilities (Beijerinck, 1913). Indeed, some tree pathogens show high
325 within-population genetic variation coupled with low genetic differentiation among
326 geographically distant populations (Gérard et al. 2006, Cabral et al. 2016). This is generally
327 discussed as a result of efficient gene flow resulting from excellent dispersal capacities over
328 large geographic distances (Gérard et al. 2006, Cabral et al. 2016). However, the dispersal of
329 many other microorganisms (including pathogens) may be limited by diverse factors: low host
330 tree abundance in the agricultural Great Plains area represents an important gene flow barrier for
331 the white pine blister rust, *Cronartium ribicola* (Hamelin et al. 2000, Brar et al. 2015). A similar
332 barrier exists also for the Sapstain fungus, *Ceratocystis resinifera*, again due to the lack host
333 trees in an agricultural area (Morin, Breuil, & Bernier, 2004). We expected that areas southeast
334 of the Mackenzie Mountains (at latitude N60 and longitude -E125 on Fig. 1) will act as
335 biogeographic barriers for *S. populicola* just as they do for the balsam poplar hosts (S. R. Keller
336 et al., 2010). However, neither of the median-joining network (Fig. 1), nor the statistical test
337 (Table 2) show support for this barrier. As another explanation, we cannot exclude the possibility
338 that the observed lack of differentiation in the used marker is due to human activities. The

exchange of tree material may possibly initiate/sustain gene flow between geographically distant pathogen populations, even if the pathogen is incapable of dispersing over long distances (Sakalidis, Feau, Dhillon, & Hamelin, 2016). For *Septoria musiva*, a pathogen causing stem cankers as well as leaf spot disease (and a close relative of *S. populicola*), a range expansion was facilitated by human activities (Sakalidis et al., 2016). However, our results do not support this scenario: we neither find evidence of populations that may have acted as sources of colonization, nor do we find pronounced genetic distances between geographically close sites, both of which could be interpreted as signatures of human-assisted colonization (Sakalidis et al., 2016). Finally, there might not be enough variation in the ITS marker itself to record phylogeographic patterns. Studies mentioned above that found evidence for gene flow barriers in fungal pathogens (Brar et al., 2015; Hamelin et al., 2000; Morin et al., 2004) relied on multiple loci. A solution may be genotyping with faster evolving markers, such as microsatellites in a metabarcoding framework (De Barba et al., 2017), but to our knowledge this approach remains to be tested in fungal ecology.

Even if organisms can disperse over long distances, they may still not be able to establish in local communities if they cannot pass the second set of ecological filters: the constraints of the abiotic environment. Environmental filters may also act on the genetic variants of the same species, as we see in the climatic adaptation of the balsam poplar hosts of *S. populicola* (Soolanayakanahally et al. 2009, Keller et al. 2011, Olson et al. 2013). However, our results show no support for environmental limitations regarding the diversity and abundance of recorded oligotypes of *S. populicola* (Table 1, 2): we see no broad-scale patterns of oligotype diversity nor oligotype composition along major climatic gradients (which are correlates of latitude). This may mean that beside being able to disperse over vast geographic distances, *S. populicola* oligotypes

may also overcome the filter of environmental conditions for local community assembly (although we cannot exclude the existence of further, not investigated abiotic filters, or co-adaptation to a locally adapted host, which may be responsible for the significant differences in oligotype abundances among some localities, Fig. 1, Table 2). This may indicate that pathogen species may share common distribution patterns with their hosts (supported also by a study on *Armillaria* species on conifers in Japan (Hasegawa, Ota, Hattori, Sahashi, & Kikuchi, 2011)). Continental-scale population genetic structure of tree pathogens sometimes shows little differentiation (Hamelin et al. 2000, Morin et al. 2004, Brar et al. 2015), but in other cases fungal tree pathogens are clearly influenced by the environment (e.g. those of *Populus angustifolia,* (Busby et al., 2014)). An alternative explanation is provided by metacommunity theory (Leibold et al., 2004): strong immigration of an excellent disperser into unfavourable, but not completely unsuitable environments would also lead to the observed lack of environmental pattern – and this explanation would also be in line with the excellent dispersal abilities of *S. populicola*. Finally, neutral markers such as ITS are suboptimal to record the signs of environmental filtering since they accumulate variation over a longer period of time, but may fail to record the signature of recent selective events. Such genes may be targeted in metabarcoding studies, but need to be identified first. This means that there might be variation in *S. populicola* lineages in loci related to climatic adaptation (as recorded by single nucleotide polymorphisms – SNPs – in its balsam poplar host (Stephen R. Keller, Levsen, Olson, & Tiffin, 2012), but this is not reflected in variation found in the ITS region.

Finally, we found that the infection of *S. populicola* oligotypes was highly variable among symptom-free balsam poplar leaves at each location: we did not record any oligotypes from many leaves at any given location, but several leaves contained more than one oligotype at

the same location (see Fig. S4 in Supporting Information). One explanation is that leaf-specific microhabitat conditions define the local infection patterns of oligotypes. Endophyte infections are known to depend on microhabitat conditions influenced by location of leaves (Scholtysik, Unterseher, Otto, & Wirth, 2013), but we consider this explanation unlikely since all leaves were picked from lower branches, about 1.5 – 2 m from the ground. A more likely explanation in our opinion is the effects of biotic interactions as the last filters of community assembly, and particularly the possible genetically defined capability of hosts to inhibit or favour colonization by certain oligotypes. Indeed, earlier works found that tree hosts may influence the composition of foliar endophyte communities (Saunders & Kohn 2009, Bálint et al. 2013, Busby et al. 2014). Such genetically defined resistance then may result in patchy infection patterns among hosts, which would be in agreement with our results and fit the disease triangle framework (Agrios, 2005; McNew, 1960). Priority effects provide another explanation: early colonizers may inhibit the establishment of latecomers (Kennedy & Bruns 2005, Fukami 2015). This would also be in full agreement with the observed patchy patterns, even if the local colonization of *S. populicola* is completely random. The test of these hypotheses will require genotype information from all tree hosts, in addition to the collection of multiple leaves from each tree, none of which was available for the current study. Further, such test will also need to investigate leaves symptomatically infected with *S. populicola*, while here we investigated only asymptomatic leaves.

## 6. Conclusions

In summary, we do not find evidence either for dispersal limitations for *S. populicola* and its oligotypes among south-central Canada and north-central Alaska, or for environmental filtering of *S. populicola* oligotypes across this vast and climatically diverse range. We interpret these results mainly as indications for high dispersal capabilities of this forestry pathogen, although the limitations of the neutral ITS marker call for further investigations with markers of higher resolution and/or indicative of selection. In contrast with the continental-scale homogeneity of genetic diversity patterns, we found highly patchy local infections among balsam poplar hosts, at least in healthy leaves. This may indicate the importance of host resistance in shaping *S. populicola* prevalence on a local scale. This is in line with the actual concepts about host-pathogen interactions, but further advances will need genotype-level information about the disease resistance of hosts, in addition to information about pathogen microhabitats, and/or the temporal sequence of infection.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26985v1 | CC BY 4.0 Open Access | rec: 12 Jun 2018, publ: 12 Jun 2018

## 7. References

Agrios, G. N. (2005). Plant pathology. 5th. *Elsevier Acad. Press. Burlington. Mass, EU*, 251–262.

Bálint, M., Bahram, M., Eren, A. M., Faust, K., Fuhrman, J. A., Lindahl, B., … Tedersoo, L. (2016). Millions of reads, thousands of taxa: microbial community structure and associations analyzed via marker genes. *FEMS Microbiology Reviews*, *40*(5), 686–700. http://doi.org/10.1093/femsre/fuw017

Bálint, M., Bartha, L., O'Hara, R. B., Olson, M. S., Otte, J., Pfenninger, M., … Schmitt, I. (2015). Relocation, high-latitude warming and host genetic identity shape the foliar fungal microbiome of poplars. *Molecular Ecology*, *24*(1), 235–48. http://doi.org/10.1111/mec.13018

Bálint, M., Márton, O., Schatz, M., Düring, R.-A., & Grossart, H.-P. (2017). Proper experimental design requires randomization/balancing of molecular ecology experiments. *bioRxiv*. http://doi.org/http://dx.doi.org/10.1101/109280

Bálint, M., Tiffin, P., Hallström, B., O'Hara, R. B., Olson, M. S., Fankhauser, J. D., … Schmitt, I. (2013). Host Genotype Shapes the Foliar Fungal Microbiome of Balsam Poplar (Populus balsamifera). *PLoS ONE*, *8*(1), e53987. http://doi.org/10.1371/journal.pone.0053987

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using {lme4}. *Journal of Statistical Software*, *67*(1), 1–48. http://doi.org/10.18637/jss.v067.i01

Beijerinck, M. W. (1913). Jaarboek van de Koninklijke Akademie voor Wetenschappen.

*Wetenschappen. Amsterdam, The Netherlands: Müller*.

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., … Nilsson, R. H. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in Ecology and Evolution*, *4*(10), 914–919. http://doi.org/10.1111/2041-210X.12073

Berg, E. E., & Henry, J. D. (2003). The history of spruce bark beetle outbreak in the Kluane region as determined from the dendrochronology of selected forest stands. *Parks Canada Report*.

Brar, S., Tsui, C. K. M., Dhillon, B., Bergeron, M.-J., Joly, D. L., Zambino, P. J., … Hamelin, R. C. (2015). Colonization history, host distribution, anthropogenic influence and landscape features shape populations of white pine blister rust, an invasive alien tree pathogen. *PloS One*, *10*(5), e0127916. http://doi.org/10.1371/journal.pone.0127916

Burns, R. M., & Honkala, B. H. (1990). Sylvics of North America. Volume 2 : Hardwoods. *United States Department of Agriculture (USDA), Forest Service, Washington, D.C.*

Busby, P. E., Newcombe, G., Dirzo, R., & Whitham, T. G. (2014). Differentiating genetic and environmental drivers of plant-pathogen community interactions. *Journal of Ecology*, *102*(5), 1300–1309. http://doi.org/10.1111/1365-2745.12270

Cabral, P. G. C., Maciel-Zambolim, E., Oliveira, S. A. S., Caixeta, E. T., & Zambolim, L. (2016). Genetic diversity and structure of *Hemileia vastatrix* populations on *Coffea* spp. *Plant Pathology*, *65*(2), 196–204. http://doi.org/10.1111/ppa.12411

Callan, B. E., Leal, I., Foord, B., Dennis, J. J., & van Oosten, C. (2007). *Septoria musiva* isolated from cankered stems in hybrid poplar stool beds, Fraser Valley, British Columbia. *North American Fungi*, *2*, 1–9. http://doi.org/10.2509/pnwf.2007.002.007

Carroll, A. L., Taylor, S. W., Regniere, J., & Safranyik, L. (2003). Effect of climate change on range expansion by the mountain pine beetle in British Columbia. *Pages 223-232 in T.L Shore et Al. (Eds) Mountain Pine Beetle Symposium: Challenges and Solutions, Oct. 30-31, 2003. Kelowna BC. Natural Resources Canada, Infromation Report BC-X-399, Victoria*. Retrieved from http://digitalcommons.usu.edu/barkbeetles/195

Chamberlain, S. A., Bronstein, J. L., & Rudgers, J. A. (2014). How context dependent are species interactions? *Ecology Letters*, *17*(7), 881–890. http://doi.org/10.1111/ele.12279

Cubero, O. F., Crespo, A., Fatehi, J., & Bridge, P. D. (2002). Isolation of nucleic acids from lichens. In *Protocols in Lichenology. Kranner I, Beckett R, Varma A (eds). Springer Berlin Heidelberg* (pp. 381–391). http://doi.org/10.1007/978-3-642-56359-1

De Barba, M., Miquel, C., Lobréaux, S., Quenette, P. Y., Swenson, J. E., & Taberlet, P. (2017). High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Molecular Ecology Resources*, *17*(3), 492–507. http://doi.org/10.1111/1755-0998.12594

Desprez-Loustau, M.-L., Aguayo, J., Dutech, C., Hayden, K. J., Husson, C., Jakushkin, B., … Vacher, C. (2016). An evolutionary ecology perspective to address forest pathology challenges of today and tomorrow. *Annals of Forest Science*, *73*(1), 45–67. http://doi.org/10.1007/s13595-015-0487-4

Desprez-Loustau, M.-L., Robin, C., Reynaud, G., Déqué, M., Badeau, V., Piou, D., … Marçais,

B. (2007). Simulating the effects of a climate-change scenario on the geographical range and activity of forest-pathogenic fungi. *Canadian Journal of Plant Pathology*, *29*(2), 101–120. http://doi.org/10.1080/07060660709507447

Eren, A. M., Borisy, G. G., Huse, S. M., & Mark Welch, J. L. (2014). Oligotyping analysis of the human oral microbiome. *Proceedings of the National Academy of Sciences*, *111*(28), E2875–E2884. http://doi.org/10.1073/pnas.1409644111

Eren, A. M., Maignien, L., Sul, W. J., Murphy, L. G., Grim, S. L., Morrison, H. G., & Sogin, M. L. (2013). Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods in Ecology and Evolution*, *4*(12), 1111–1119. http://doi.org/10.1111/2041-210X.12114

Eren, A. M., Morrison, H. G., Lescault, P. J., Reveillaud, J., Vineis, J. H., & Sogin, M. L. (2015). Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *The ISME Journal*, *9*(4), 968–979. http://doi.org/10.1038/ismej.2014.195

Eren, A. M., Sogin, M. L., Morrison, H. G., Vineis, J. H., Fisher, J. C., Newton, R. J., & McLellan, S. L. (2015). A single genus in the gut microbiome reflects host preference and specificity. *The ISME Journal*, *9*(1), 90–100. http://doi.org/10.1038/ismej.2014.97

Eren, A. M., Zozaya, M., Taylor, C. M., Dowd, S. E., Martin, D. H., & Ferris, M. J. (2011). Exploring the diversity of Gardnerella vaginalis in the genitourinary tract microbiota of monogamous couples through subtle nucleotide variation. *PLoS ONE*, *6*(10), e26732. http://doi.org/10.1371/journal.pone.0026732

FAO. (2010). Global Forest Resources Assessment. Main report. *FAO Forestry Paper 163,*

*Rome*.

FAO. (2012). Improving lives with poplars and willows. Synthesis of country progress reports. In *24th Session of the International Poplar Commission, Dehradun, India. Working Paper IPC/12, Rome. Available at www.fao.org/forestry/ipc2012/en*.

Fukami, T. (2015). Historical contingency in community assembly : integrating niches, species pools, and priority effects. *Annual Review of Ecology, Evolution and Systematics*, *46*, 1–23. http://doi.org/10.1146/annurev-ecolsys-110411-160340

Gérard, P. R., Husson, C., Pinon, J., & Frey, P. (2006). Comparison of Genetic and Virulence Diversity of *Melampsora larici-populina* Populations on Wild and Cultivated Poplar and Influence of the Alternate Host. *Phytopathology*, *96*(9), 1027–36. http://doi.org/10.1094/PHYTO-96-1027

Hamelin, R. C., Hunt, R. S., Geils, B. W., Jensen, G. C., Jacobi, V., & Lecours, N. (2000). Barriers to gene flow between eastern and western population of *Cronartium ribicola* in North America. *Phytopathology*, *90*(10), 1073–1078.

Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C., & Martiny, J. B. H. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology*, *10*(7), 497–506. http://doi.org/10.1038/nrmicro2795

Hasegawa, E., Ota, Y., Hattori, T., Sahashi, N., & Kikuchi, T. (2011). Ecology of *Armillaria* species on conifers in Japan. *Forest Pathology*, *41*(6), 429–437. http://doi.org/10.1111/j.1439-0329.2010.00696.x

Herath, P., Beauseigle, S., Dhillon, B., Ojeda, D. I., Bilodeau, G., Isabel, N., … Feau, N. (2016).

529 Anthropogenic signature in the incidence and distribution of an emerging pathogen of
530 poplars. *Biological Invasions*, *18*(4), 1147–1161. http://doi.org/10.1007/s10530-015-1051-8

531 Hill, M. O. (1973). Diversity and evenness: a unifying notation and its consequences. *Ecology*,
532 *54*(2), 427–432. http://doi.org/10.2307/1934352

533 HilleRisLambers, J., Adler, P. B., Harpole, W. S., Levine, J. M., & Mayfield, M. M. (2012).
534 Rethinking Community Assembly through the Lens of Coexistence Theory. *Annual Review*
535 *of Ecology, Evolution, and Systematics*, *43*, 227–48. http://doi.org/10.1146/annurev-ecolsys-
536 110411-160411

537 Huse, S. M., Welch, D. M., Morrison, H. G., & Sogin, M. L. (2010). Ironing out the wrinkles in
538 the rare biosphere through improved OTU clustering. *Environmental Microbiology*, *12*(7),
539 1889–1898. http://doi.org/10.1111/j.1462-2920.2010.02193.x

540 Illumina. (2016). 16S Metagenomic Sequencing Library Preparation. Part # 15044223 Rev. A pp
541 28. Accessed on April 24 2017. Retrieved from
542 http://support.illumina.com/content/dam/illumina-
543 support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-
544 prep-guide-15044223-b.pdf

545 Johnson, J. M., & Oelmüller, R. (2009). Mutualism or parasitism : life in an unstable continuum.
546 What can we learn from the mutualistic interaction between *Piriformospora indica* and
547 *Arabidopsis thaliana*? *Endocytobiosis and Cell Research: Journal of the International*
548 *Society of Endocytobiology*, *19*, 81–111. Retrieved from
549 http://pubman.mpdl.mpg.de/pubman/faces/viewItemOverviewPage.jsp?itemId=escidoc:158
550 7443

Keller, S. R., Levsen, N., Olson, M. S., & Tiffin, P. (2012). Local adaptation in the flowering-time gene network of balsam poplar, populus balsamifera L. *Molecular Biology and Evolution*, *29*(10), 3143–3152. http://doi.org/10.1093/molbev/mss121

Keller, S. R., Olson, M. S., Salim, S., William, S. A., & Peter, T. (2010). Genomic diversity, population structure, and migration following rapid range expansion in the Balsam Poplar, *Populus balsamifera*. *Molecular Ecology*, *19*(6), 1212–1226. http://doi.org/10.1111/j.1365-294X.2010.04546.x

Keller, S. R., Soolanayakanahally, R. Y., Guy, R. D., Silim, S. N., Olson, M. S., & Tiffin, P. (2011). Climate-driven local adaptation of ecophysiology and phenology in balsam poplar, *Populus balsamifera* L. (Salicaceae). *American Journal of Botany*, *98*(1), 99–108. http://doi.org/10.3732/ajb.1000317

Kennedy, P. G., & Bruns, T. D. (2005). Priority effects determine the outcome of ectomycorrhizal competition between two Rhizopogon species colonizing Pinus muricata seedlings. *New Phytologist*, *166*(2), 631–638.

Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., … Larsson K-H. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology*, *22*(21), 5271–5277. http://doi.org/10.1111/mec.12481

Kurz, W. A., Dymond, C. C., Stinson, G., Rampley, G. J., Neilson, E. T., Carroll, A. L., … Safranyik, L. (2008). Mountain pine beetle and forest carbon feedback to climate change. *Nature*, *452*(7190), 987–90. http://doi.org/10.1038/nature06777

LeBoldus, J. M., Blenis, P., Thomas, B. R., Feau, N., & Bernier, L. (2009). Susceptibility of *Populus balsamifera* to *Septoria musiva* : A Field Study and Greenhouse Experiment. *Plant*

573 *Disease*, *93*(11), 1146–1150. http://doi.org/10.1094/PDIS-93-11-1146

574 Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., …
575 Gonzalez, A. (2004). The metacommunity concept: A framework for multi-scale
576 community ecology. *Ecology Letters*, *7*(7), 601–613. http://doi.org/10.1111/j.1461-
577 0248.2004.00608.x

578 Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L.,
579 … Staley, J. T. (2006). Microbial biogeography: putting microorganisms on the map.
580 *Nature Reviews Microbiology*, *4*(2), 102–112. http://doi.org/10.1038/nrmicro1341

581 Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G., & Neufeld, J. D. (2012).
582 PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics*, *13*(1), 31.
583 http://doi.org/10.1186/1471-2105-13-31

584 McNew, G. L. (1960). The nature, origin and evolution of parasitism. *In: HORSFALL, J. G.;*
585 *DIMOND, A.E. (Ed.). Plant Pathology: An Advanced Treatise. New York, Academic Press.*
586 *2*, 19–69.

587 Meiser, A., Bálint, M., & Schmitt, I. (2014). Meta-analysis of deep-sequenced fungal
588 communities indicates limited taxon sharing between studies and the presence of
589 biogeographic patterns. *New Phytologist*, *201*(2), 623–635.
590 http://doi.org/10.1111/nph.12532

591 Morin, C., Breuil, C., & Bernier, L. (2004). Genetic Variability and Structure of Canadian
592 Populations of the Sapstain Fungus *Ceratocystis resinifera*. *Phytopathology*, *94*(12), 1323–
593 1330. http://doi.org/10.1094/PHYTO.2004.94.12.1323

Murray, K. A., Preston, N., Allen, T., Zambrana-Torrelio, C., Hosseini, P. R., & Daszak, P. (2015). Global biogeography of human infectious diseases. *Proceedings of the National Academy of Sciences*, *112*(41), 12746–12751. http://doi.org/10.1073/pnas.1507442112

Newcombe, G., & Bradshaw, H. D. (1996). Quantitative trait loci conferring resistance in hybrid poplar to Septoria populicola, the cause of leaf spot. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, *26*(11), 1943–1950. http://doi.org/Doi 10.1139/X26-219

Newcombe, G., Ostry, M., Hubbes, M., Périnet, P., & Mottet, M. (2001). Poplar diseases. In *Dickmann DI, Isebrands JG,Eckenwalder JE, Richardson J (eds) Poplar Culture in North America. National Research Council of Canada Research Press, Ottawa* (pp. 249–276).

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., … Wagner, H. (2016). vegan: Community Ecology Package. Retrieved from http://cran.r-project.org/package=vegan

Olson, M. S., Levsen, N., Soolanayakanahally, R. Y., Guy, R. D., Schroeder, W. R., Keller, S. R., & Tiffin, P. (2013). The adaptive potential of *Populus balsamifera* L. to phenology requirements in a warmer global climate. *Molecular Ecology*, *22*(5), 1214–1230. http://doi.org/10.1111/mec.12067

Ostry, M. E., & McNabb, H. S. (1983). Diseases of intensively cultured hybrid poplars: A summary of recent research in the North Central Region. In *Hansen, E. A. (eds) Intensive planting culture: 12 years research general technologies. U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station, St. Paul, MN* (pp. 102–109).

Pansu, J., Giguet-Covex, C., Ficetola, G. F., Gielly, L., Boyer, F., Zinger, L., … Choler, P.

(2015). Reconstructing long-term human impacts on plant communities: An ecological approach based on lake sediment DNA. *Molecular Ecology*, *24*(7), 1485–1498. http://doi.org/10.1111/mec.13136

Paradis, E. (2010). pegas: an {R} package for population genetics with an integrated--modular approach. *Bioinformatics*, *26*, 419–420.

Paradis, E., Claude, J., & Strimmer, K. (2004). A{PE}: analyses of phylogenetics and evolution in {R} language. *Bioinformatics*, *20*, 289–290.

Patin, N. V., Kunin, V., Lidström, U., & Ashby, M. N. (2013). Effects of OTU Clustering and PCR Artifacts on Microbial Diversity Estimates. *Microbial Ecology*, *65*(3), 709–719. http://doi.org/10.1007/s00248-012-0145-4

Pautasso, M., Schlegel, M., & Holdenrieder, O. (2015). Forest Health in a Changing World. *Microbial Ecology*, *69*(4), 826–842. http://doi.org/10.1007/s00248-014-0545-8

Polle, A., Janz, D., Teichmann, T., & Lipka, V. (2013). Poplar genetic engineering: Promoting desirable wood characteristics and pest resistance. *Applied Microbiology and Biotechnology*, *97*(13), 5669–5679. http://doi.org/10.1007/s00253-013-4940-8

R Core Team. (2017). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from https://www.r-project.org/

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. http://doi.org/10.7287/peerj.preprints.2409v1

Ryberg, M. (2015). Molecular operational taxonomic units as approximations of species in the

637 light of evolutionary models and empirical data from Fungi. *Molecular Ecology*, *24*(23),
638 5770–5777. http://doi.org/10.1111/mec.13444

639 Sakalidis, M. L., Feau, N., Dhillon, B., & Hamelin, R. C. (2016). Genetic patterns reveal
640 historical and contemporary dispersal of a tree pathogen. *Biological Invasions*, *18*(6), 1781–
641 1799. http://doi.org/10.1007/s10530-016-1120-7

642 Saunders, M., & Kohn, L. M. (2009). Evidence for alteration of fungal endophyte community
643 assembly by host defense compounds. *New Phytologist*, *182*(1), 229–238.
644 http://doi.org/10.1111/j.1469-8137.2008.02746.x

645 Schmidt, P.-A., Bálint, M., Greshake, B., Bandow, C., Römbke, J., & Schmitt, I. (2013).
646 Illumina metabarcoding of a soil fungal community. *Soil Biology and Biochemistry*, *65*,
647 128–132. http://doi.org/10.1016/j.soilbio.2013.05.014

648 Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated - reducing
649 sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology*
650 *Resources*, *15*(6), 1289–1303. http://doi.org/10.1111/1755-0998.12402

651 Scholtysik, A., Unterseher, M., Otto, P., & Wirth, C. (2013). Spatio-temporal dynamics of
652 endophyte diversity in the canopy of European ash (Fraxinus excelsior). *Mycological*
653 *Progress*, *12*(2), 291–304. http://doi.org/10.1007/s11557-012-0835-9

654 Shaw, M. W., & Osborne, T. M. (2011). Geographic distribution of plant pathogens in response
655 to climate change. *Plant Pathology*, *60*(1), 31–43. http://doi.org/10.1111/j.1365-
656 3059.2010.02407.x

657 Smith, L. M., & Reynolds, H. L. (2015). Plant-soil feedbacks shift from negative to positive with

658 decreasing light in forest understory species. *Ecology*, *96*(9), 2523–2532.
659 http://doi.org/10.1890/14-2150.1

660 Solheim, H., & Krokene, P. (1998). Growth and virulence of mountain pine beetle associated
661 blue-stain fungi, *Ophiostoma clavigerum* and *Ophiostoma montium*. *Canadian Journal of*
662 *Botany*, *76*(4), 561–566. http://doi.org/10.1139/cjb-76-4-561

663 Soolanayakanahally, R. Y., Guy, R. D., Silim, S. N., Drewes, E. C., & Schroeder, W. R. (2009).
664 Enhanced assimilation rate and water use efficiency with latitude through increased
665 photosynthetic capacity and internal conductance in balsam poplar (*Populus balsamifera*
666 L.). *Plant, Cell and Environment*, *32*(12), 1821–1832. http://doi.org/10.1111/j.1365-
667 3040.2009.02042.x

668 Sturrock, R. N., Frankel, S. J., Brown, A. V., Hennon, P. E., Kliejunas, J. T., Lewis, K. J., …
669 Woods, A. J. (2011). Climate change and forest diseases. *Plant Pathology*, *60*(1), 133–149.
670 http://doi.org/10.1111/j.1365-3059.2010.02406.x

671 Su, X., Chu, Y., Li, H., Hou, Y., Zhang, B., Huang, Q., … Tian, Y. (2011). Expression of
672 multiple resistance genes enhances tolerance to environmental stressors in transgenic poplar
673 (Populus × euramericana "Guariento"). *PLoS ONE*, *6*(9), 1–11.
674 http://doi.org/10.1371/journal.pone.0024614

675 Taylor, J. W., Turner, E., Townsend, J. P., Dettman, J. R., & Jacobson, D. (2006). Eukaryotic
676 microbes, species recognition and the geographic limits of species: examples from the
677 kingdom Fungi. *Philosophical Transactions of the Royal Society of London B: Biological*
678 *Sciences*, *361*(1475), 1947–63. http://doi.org/10.1098/rstb.2006.1923

679 Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., … Abarenkov,

680 K. (2014). Global diversity and geography of soil fungi. *Science*, *346*(6213), 1256688.
681 http://doi.org/10.1126/science.aaa1185

682 Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., … Dejean, T.
683 (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA
684 metabarcoding. *Molecular Ecology*, *25*(4), 929–942. http://doi.org/10.1111/mec.13428

685 Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth). New York:
686 Springer. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4

687 Wang, Y., Naumann, U., Wright, S., Eddelbuettel, D., & Warton, D. (2016). mvabund: Statistical
688 Methods for Analysing Multivariate Abundance Data. Retrieved from http://cran.r-
689 project.org/package=mvabund

690 Wei, T., & Simko, V. (2016). corrplot: Visualization of a Correlation Matrix. R package version
691 0.77.

692 White, T. J., Bruns, T., Lee, S. J. W. T., & Taylor, J. W. (1990). Amplification and Direct
693 Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. *PCR Protocols: A Guide*
694 *to Methods and Applications*, *18*(1), 315–322. http://doi.org/10.1016/B978-0-12-372180-
695 8.50042-1

696 Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of*
697 *Statistical Software*, *40*(1), 1–29. Retrieved from http://www.jstatsoft.org/v40/i01/

698 Zalasky, H. (1978). Stem and leaf spot infections caused by *Septoria musiva* and *S. populicola*
699 on poplar seedlings. *Phytoprotection*, *59*(1), 43–50.

700

**Acknowledgements**

For discussions on methodology and interpretations of results, we thank the participants of the Endophyte Reading Group of the Goethe University in Frankfurt am Main.

**Author contributions**

FP, IS and MB designed the research. FP and MB performed the research. SM contributed analytic tools. FP analysed the data. All authors contributed to writing the MS.

**Supporting information**

Additional Supporting Information may be found in the online version of this article:

Additional tables and figures

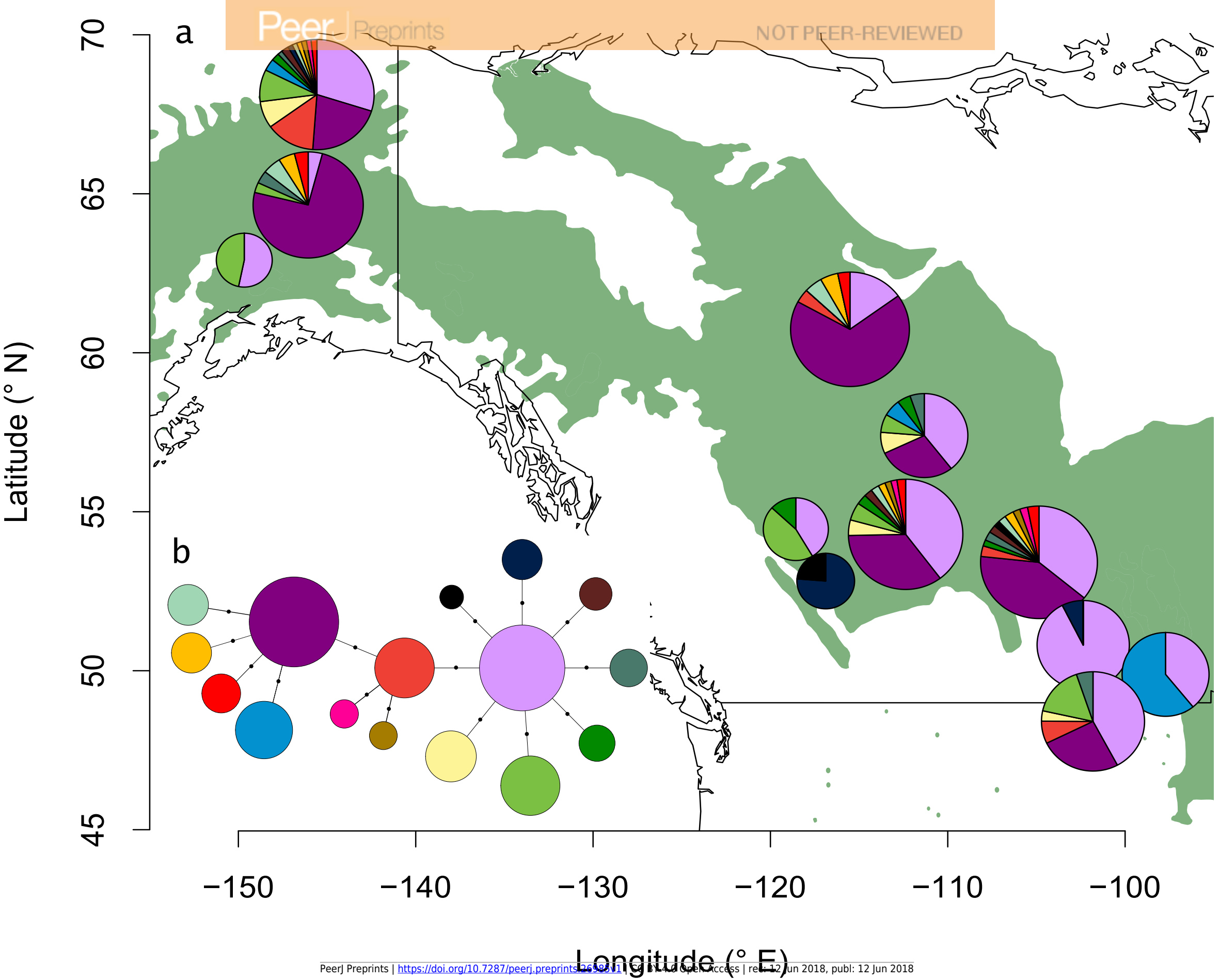
**Data availability**

The raw data of the study are deposited at figshare (xxx). Representative sequences of the oligotypes are deposited in GenBank under accession numbers xxx. The R script and all files used for the analysis and the generation of graphics (raw and cleaned abundance table and files for the generation of the map) are deposited in the open GitHub repository here: https://github.com/FionaPaul/Sphaerulina.

## Figure 1(on next page)

Figure 1. Distribution of *S. populicola* oligotypes and genetic distances of the 16 oligotypes found in the study area.

(a) The distribution of the host tree balsam poplar in Northern North America is indicated in green and the elevation in grey shades. Each pie chart indicates the frequency of the different *S. populicola* oligotypes found at each sampling location. Oligotype abundances are square root-transformed and pie chart sizes reflect log-transformed summed abundance of all oligotypes at the given location. (b) Median-joining network computed with the representative sequences of each of the 16 oligotypes. Pairwise differences are calculated based on the Kimura model. The radius of the circles represent the log-transformed abundance of each oligotype across the study area. Each black dot on a connecting line between two oligotypes indicates a substitution, i.e., a change in nucleotides at one position in the alignment.

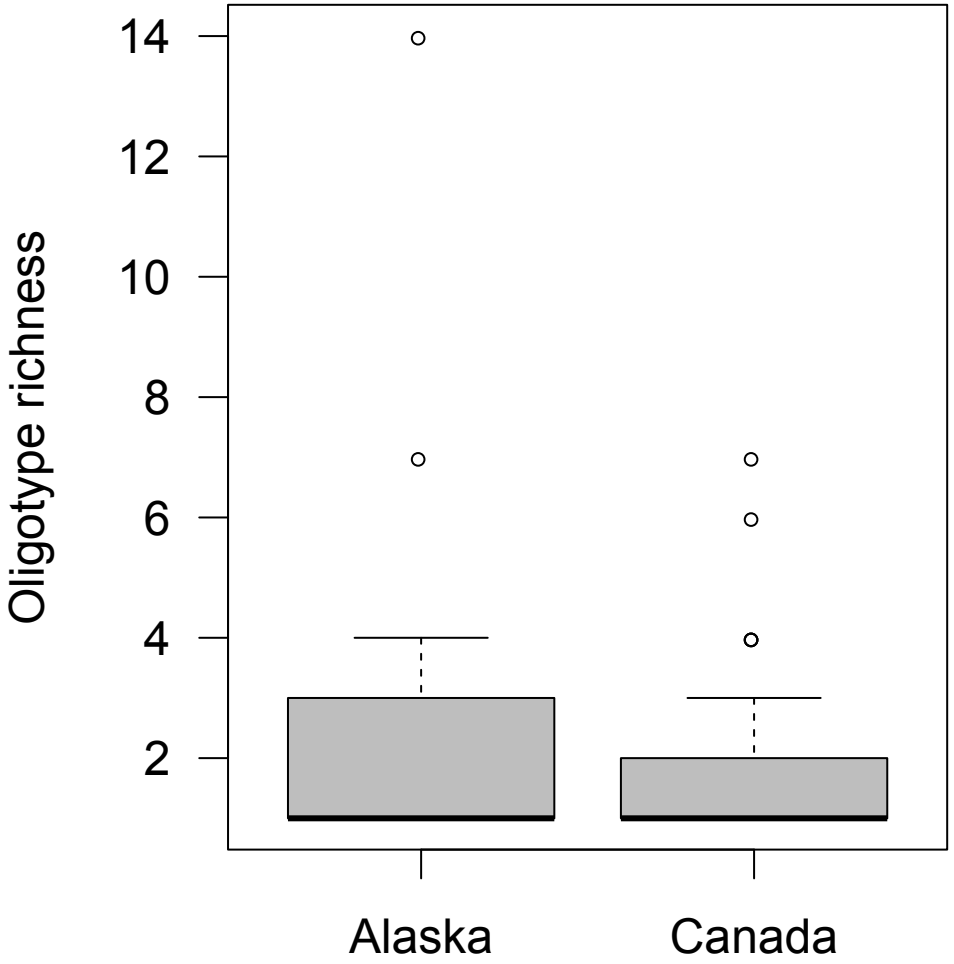



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26982v1 | CC BY 4.0 Open Access | rec: 12 Jun 2018, publ: 12 Jun 2018

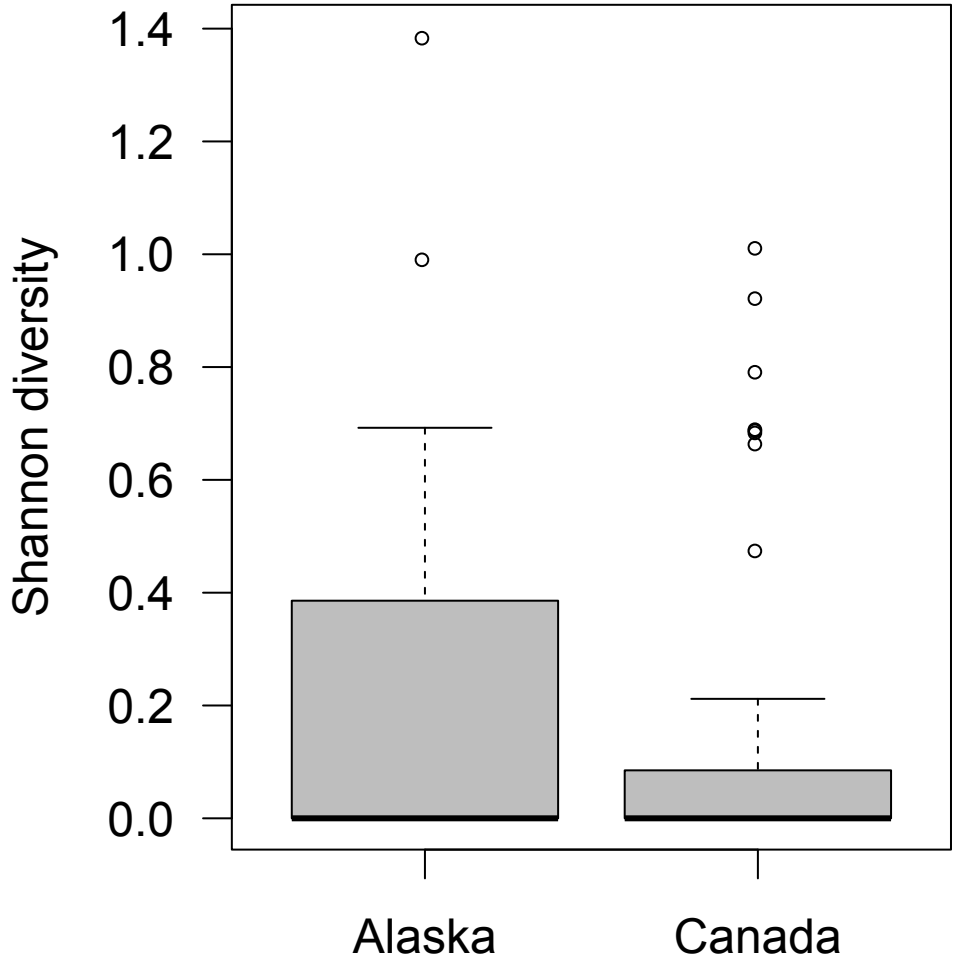
## Figure 2(on next page)

Figure 2. Richness and diversity of two geographic regions.

(a) Boxplots of *S. populicola* oligotype richness of the two main geographic regions of the balsam poplar distribution, Canada and Alaska. (b) Boxplots displaying the Shannon diversity of *S. populicola* oligotypes in the two regions, Canada and Alaska.




a




b

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26985v1 | CC BY 4.0 Open Access | rec: 12 Jun 2018, publ: 12 Jun 2018

## Table 1(on next page)

Table 1. Summarized model results for diversity patterns .

Summary of general linear mixed effects models for factors influencing diversity measures of oligotypes of *S. populicola* distributed throughout the range of its host, balsam poplar (*P. balsamifera*) in Canada and Alaska. Here exemplary results are given for the Hill's N0, which is species richness. Each variable was added to the null model consisting of sampling location as a random effect and sequencing depth as a fixed effect (see Rscript, Supplementary Material). The latitude is a substitute variable for highly autocorrelated climatic conditions (temperature and precipitation, Fig. S1 in Supporting Information).

**Table 1. Summarized model results for diversity patterns.** Summary of general linear mixed effects models for factors influencing diversity measures of oligotypes of *S. populicola* distributed throughout the range of its host, balsam poplar (*P. balsamifera*) in Canada and Alaska. Here exemplary results are given for the Hill's N0, which is species richness. Each variable was added to the null model consisting of sampling location as a random effect and sequencing depth as a fixed effect (see Rscript, Supplementary Material). The latitude is a substitute variable for highly autocorrelated climatic conditions (temperature and precipitation, Fig. S1 in Supporting Information).

| Fixed effects | Degrees of freedom | Deviance | Chi-square | p-value |
|---|---|---|---|---|
| **Sequencing depth** | 4 | 197,57 | | |
| **Replication number** | 5 | 197,54 | 0,0329 | 0,856 |
| **Geographic deme** | 5 | 197,10 | 0,4744 | 0,491 |
| **Latitude** | 5 | 196,46 | 1,1201 | 0,2899 |

## Table 2(on next page)

Table 2. Summarized model results for abundance patterns.

Summary of manyglm models investigating oligotype abundances of *S. populicola* in its host tree balsam poplar throughout 12 sampling locations across its range throughout Canada and Alaska. Multivariate generalized linear models were fitted for each oligotype individually accounting for correlations among oligotypes. Each variable was added individually to the model and its significance in the model was tested with ANOVA (likelihood-ratio test, 1000 bootstraps). The latitude is a substitute variable for highly autocorrelated climatic conditions (temperature and precipitation, Fig. S1 in Supporting Information).

**Table 2. Summarized model results for abundance patterns.** Summary of manyglm models investigating oligotype abundances of *S. populicola* in its host tree balsam poplar throughout 12 sampling locations across its range throughout Canada and Alaska. Multivariate generalized linear models were fitted for each oligotype individually accounting for correlations among oligotypes. Each variable was added individually to the model and its significance in the model was tested with ANOVA (likelihood-ratio test, 1000 bootstraps). The latitude is a substitute variable for highly autocorrelated climatic conditions (temperature and precipitation, Fig. S1 in Supporting Information).

| Factor | Residual degrees of freedom | Degrees of freedom | Test statistics | P-value |
|---|---|---|---|---|
| **Sequencing depth** | 59 | 1 | 22,81 | 0,125 |
| **Sampling location** | 49 | 11 | 201,3 | **0,044** |
| **Replication number** | 59 | 1 | 24,5 | 0,108 |
| **Geographic deme** | 59 | 1 | 7,268 | 0,866 |
| **Latitude** | 59 | 1 | 12,56 | 0,453 |