**Pathogen without borders: *Sphaerulina populicola* is genetically diverse and spreads freely through the host tree’s range**

Fiona Paul, Imke Schmitt, Sunil Mundra, Miklós Bálint

**Supporting information**

**Appendix S1: Additional tables and figures**

**Table S1. Geographic, climatic and sampling information for the 12 sampling locations.** Listed are the geographic location, i.e. longitude and latitude, the altitude for the sampling sites of balsam poplar in Canada and Alaska. Climatic parameters include mean annual precipitation as well as mean precipitation in the months January and July. The same parameters are given for the temperatures. Further, the geographic region of the trees, as well as the average read number per site coming from the sequencing run before cleaning, the number of replicates in the cleaned abundance table and the location code as used in Fig.1.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sampling location | Longitude(° W) | Latitude(° N) | Altitude(m alsm) | Mean precipitation (mm) | Mean temperature (° C) | Geographic region | Average raw read number | Nbr of replicates | Location Code |
| January | July | Annual | January | July | Annual |
| Arctic Village | -145.572 | 68.119 | 610 | 10 | 26 | 165 | -29.1 | 11 | -10.3 | Alaska | 39,764 | 4 | ava |
| Boyle | -112.376 | 54.289 | 712 | 21 | 88 | 460 | -14.9 | 16 | 1.4 | Canada | 25,281 | 7 | boy |
| Cadomin | -117.483 | 53.143 | 1446 | 34 | 87 | 534 | -12.2 | 12.4 | 0.8 | Canada | 39,636 | 2 | cad |
| Carnduff | -101.51 | 49.133 | 496 | 13 | 63 | 396 | -16.2 | 19.6 | 3.1 | Canada | 31,875 | 13 | car |
| Denali N. Park | -148.917 | 63.396 | 654 | 20 | 78 | 432 | -17.6 | 11.6 | -3.4 | Alaska | 28,363 | 3 | den |
| Fairbanks | -146.967 | 64.406 | 206 | 21 | 60 | 342 | -23.5 | 15.4 | -3.4 | Alaska | 35,766 | 4 | fbk |
| Fort McMurray | -111.325 | 56.588 | 388 | 23 | 82 | 453 | -20.7 | 16.4 | 0 | Canada | 27,225 | 8 | ftm |
| Grande Prairie | -118.569 | 54.452 | 983 | 36 | 81 | 553 | -12.8 | 14 | 1.6 | Canada | 19,040 | 3 | gpr |
| Hay River | -115.514 | 60.737 | 173 | 19 | 47 | 339 | -25.1 | 15.2 | -3.6 | Canada | 31,553 | 7 | hay |
| Love | -104.857 | 53.409 | 454 | 20 | 71 | 451 | -20.9 | 17.3 | 0.1 | Canada | 22,418 | 4 | lov |
| Melville | -102.363 | 50.811 | 529 | 20 | 60 | 424 | -18.6 | 18.1 | 1.5 | Canada | 25,551 | 3 | mel |
| Portage | -98.17 | 49.561 | 284 | 20 | 72 | 484 | -17.2 | 19.9 | 2.9 | Canada | 28,265 | 3 | por |

**Table S2. Read loss in the sequence processing steps.** For each step in the processing pipeline for a dataset of ITS1 rDNA sequences extracted from balsam poplar leaves, the number of remaining reads is given, as well as the proportion of reads lost from the previous step.

|  |  |  |
| --- | --- | --- |
| Sequence processing step | Remaining number of reads | Percentage lost to previous step |
| Trimming to 200bp | 9,074,044 | 0.00% |
| Quality filtering | 8,101,308 | 10.72% |
| Paired end assembly | 7,177,744 | 11.40% |
| Removal of primers | 6,807,044 | 5.16% |
| Removal of primer artefacts | 6,776,726 | 0.45% |
| ITS extractions | 5,620,169 | 17.07% |
| Removal of short sequences | 5,524,137 | 1.71% |
| Assignment to *S. populicola* | 116,603 | 97.89% |



**Figure S1.** **Correlation matrix of explanatory variables of 12 sampling locations.** This shows the correlation of different variables used as explanatory variables in the models explaining *S. populicola* oligotype diversity and distribution along a latitudinal gradient of the balsam poplar distribution in Canada and Alaska. Given are latitude of the sampling location, raw read number, number of replicates per location, longitude of the sampling location, the mean temperature in July, the mean precipitation in July, the mean precipitation annually, the mean temperature in January, the mean annual temperature, altitude of the study location and the mean precipitation in January. Numbers show correlation coefficients. Circles depict the strength of correlation. The colours stand for the direction of the relationship, i.e. red shows a negative correlation and blue indicates a positive correlation. Circle size and colour intensity are proportional to the correlation coefficient.



**Figure S2. Total abundances of *S. populicola* oligotypes in balsam poplar.** This pie chart displays the square root-transformed abundances of the 16 oligotypes found in this study within the OTU of *S. populicola* in the host tree balsam poplar along an altitudinal gradient of the tree’s range in Canada and Alaska.



**Figure S3. NMDS plot displaying differences in samples from different geographic regions.** The dots represent samples, i.e. samples trees, that are placed in the two-dimensional space according to each other’s differences. Dots are jittered to tease apart directly overlaying sample dots. The colours of the dots represent the geographic region of the sampling location, i.e. Canada representing the Southern geographic deme and Alaska representing the Northern geographic deme of balsam poplar.



**Figure S4.** **NMDS plot displaying differences in samples from different locations.** Each dot represents a sample, i.e. a sampled balsam poplar tree, which is placed in the two-dimensional space according to the rank difference amongst all the samples. The dots are coloured according to the sampling location. Dots are jittered to tease apart directly overlaying sample dots.