Methods S1 – *P. glauca* SNP atlas characterization.

The TP rate was high whatever the depth, reaching up to 91.7% (Table 1), and varied little among SNPs with a depth of 10 reads or more (TP rate=0.018 x (depth) + 85.1; R^2=0.23). Thus, a minimum alignment depth of 10 reads was used for subsequent analyses and resulted in a global TP rate of 88.1%. The VarScan p-values were found to clearly discriminate the true positive from false positive SNPs (Table 1), as shown by the rapid and linear decrease in the TP rate with increasing VarScan p-values (TP rate=-8.98 × (p-value) + 102.73; R^2=0.99). Predicted SNPs with a maximum VarScan p-value cutoff of 0.10 had a global minimum TP rate of 92.1%. The TP rate also increased with the MAF (TP rate=0.47 × (MAF) + 89.49; R^2=0.48) but it was high notwithstanding the MAF (Table 1). Since SNPs with a MAF<0.01 had been discarded (see Methods), no further restriction was applied based on this parameter (Table 1).

Methods S2 -Sequence processing

The P. *glauca* ESTs were previously described (Supplementary Table 3). Identical 454 reads derived from one amplification reaction are technical replicates which may generate false positive SNPs. Therefore, their removal had been recommended before any quantitative analysis (Gomez-Alvarez et al, 2009; Teal & Schmidt, 2010; Ueno et al, 2010). Replicates were searched within each batch of sequences derived from one given PCR amplification with the 454 Replicate Filter software (Teal & Schmidt 2010; http://microbiomes.msu.edu/replicates/). Reads originating from the same PCR emulsion were first grouped together and were then processed by using a two-step approach. For the first step, only exact duplicates of each group were filtered out by using those criteria:

1.0 for the sequence identity cutoff, 1.0 for the length difference requirement and 20 for the initial base-pair requirement. The resulting groups were filtered out again using less restrictive criteria: 0.98 for the sequence identity cutoff, 0.98 for the length difference requirement and 20 for the initial base-pair requirement. Finally, remaining reads originating from the same group were pooled back together to their initial group. When a cluster of duplicated reads was identified, we kept only the longest sequence. In total, the 454 GS-FLX sequencing procedure produced 8.2% of duplicates, while the 454 GS-FLX-Titanium sequencing technology resulted in 1.5% of duplicates (Supplementary Table 4).

MosaikAligner, a module of the Mosaik software (version1.1.0021 modified to process Illumina 1.3+ format; http://bioinformatics.bc.edu/marthlab/Mosaik) was used for read alignment since it enables to analyse and merge data from Sanger, 454 and Illumina GAII technologies. The alignment parameters were: "-hs 15 -mmp 0.03 -minp 0.90 -act 55 -m unique –mmal" (for the Sanger and 454 GS-FLX-Titanium reads), "-hs 15 -mmp 0.03 -minp 0.90 -act 26 -m unique –mmal" (for the GS-FLX reads) and "-mm 3 -act 35 -bw 29 -m unique" (for the Illumina GAII reads). In the alignments, the maximum percentage of mismatches allowed on the aligned read length was 3%. Reads were incorporated in the alignments only if at least 90% of the length of the read could be aligned to the reference sequence.

The MosaikBuild module converted the read formats into Mosaik native read format and automatically deleted trimming and lagging unresolved bases ('N') for all reads. It also automatically filtered out 58,865 reads with 5 or more interior unresolved bases ('N') (0.05% of the quality reads). The MosaikAligner module left unassembled the reads which could not be mapped to the reference database unambiguously. In total, 3.4 million

of such reads (5.4% of the quality reads) were excluded from the alignment because they matched two or more sequences from this database. Most of the reads (41.0% of the quality reads) discarded from the final alignments were excluded because they did not comply with the alignment parameters.