

## **Appendix S1**

Ecological Monographs (ESA)

### **Glacial refugia, post-glacial range dynamics and hybridization zones of Pinaceae in Eurasia captured from sedimentary ancient DNA**

Stefano Meucci, Kathleen R. Stoof-Leichsenring, Yanrong Zhang, Andrei Andreev, Konstantin V. Krutovsky, Boris K. Biskaborn, Inger G. Alsos, Laura Parducci, Kevin Nota, Sandra Garcés-Pastor, Petr Kuneš, Walter Finsinger, Eleonora Cagliero, Daniel Vondrák, Jaroslav Kukla, Darrell Kaufman, Barbara Wohlfarth, Heikki Seppä, Ulrike Herzschuh



**Figure S1.** Absolute read counts of captured DNA sequences from Lake Imandra. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



**Figure S2.** Absolute read counts of captured DNA sequences from Lake Ezzhinitsa. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



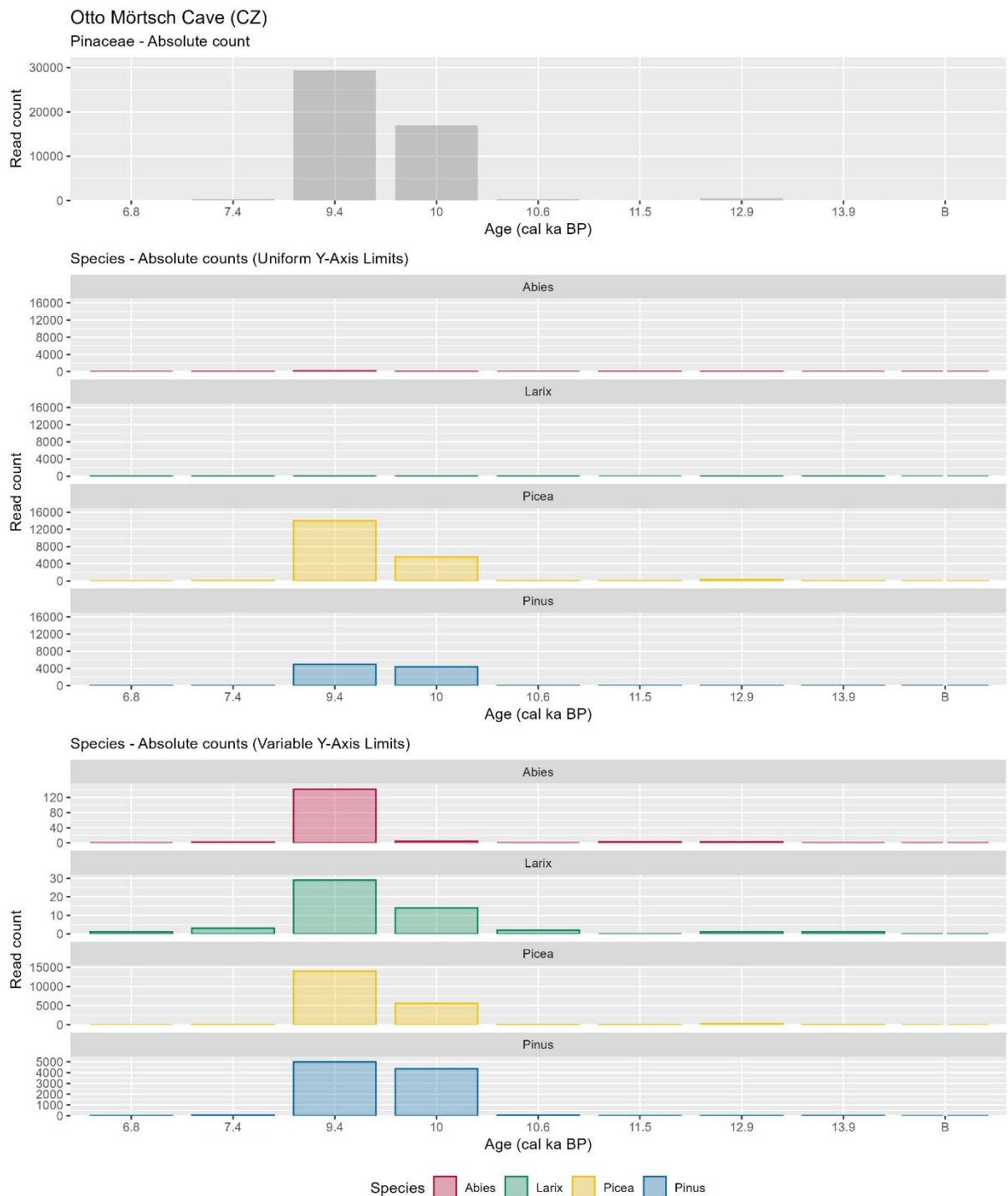
**Figure S3.** Absolute read counts of captured DNA sequences from Atteköp wetland. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



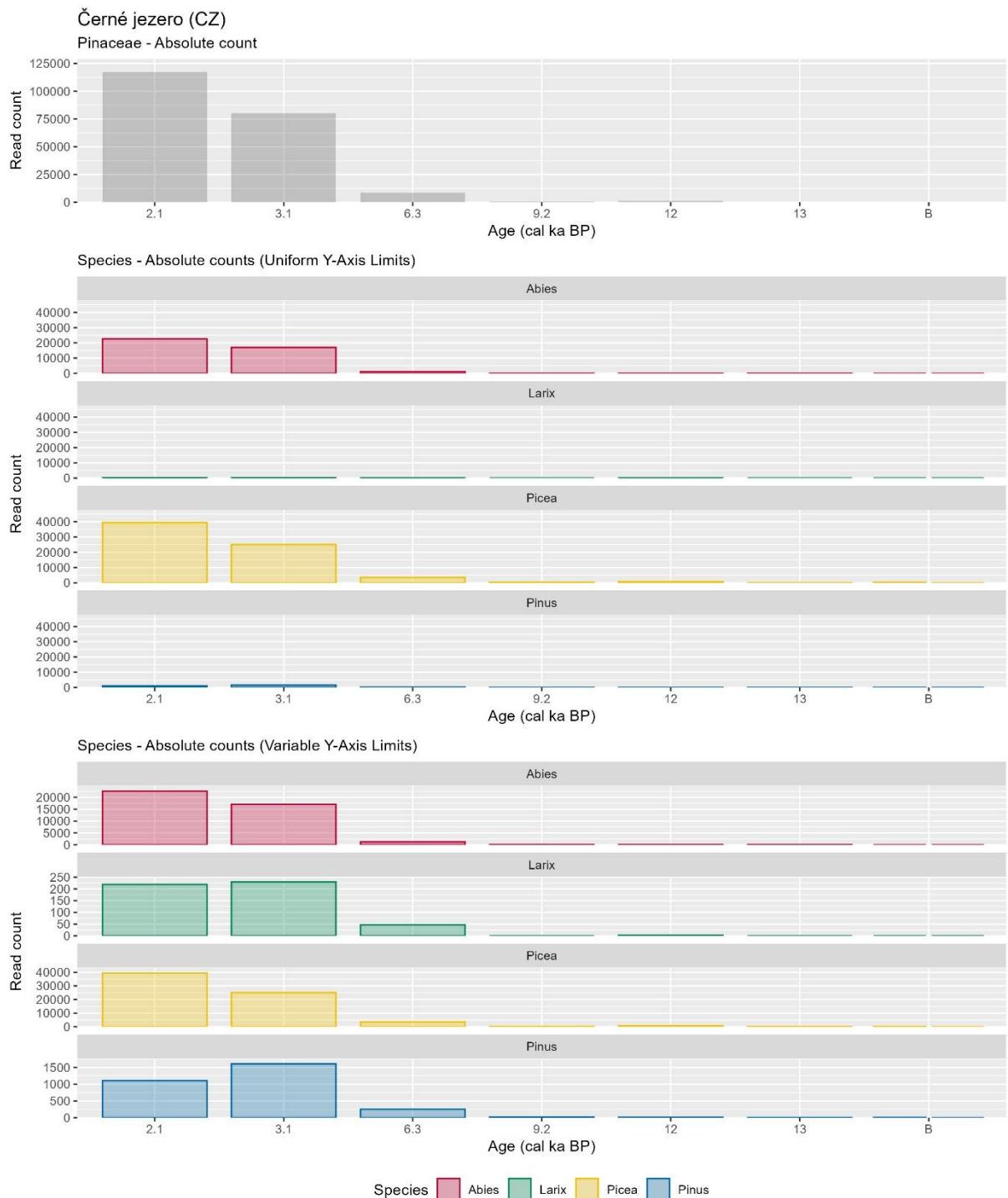
**Figure S4.** Absolute read counts of captured DNA sequences from Hässeldala Port wetland. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



**Figure S5.** Absolute read counts of captured DNA sequences from Lake Nautajärvi. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



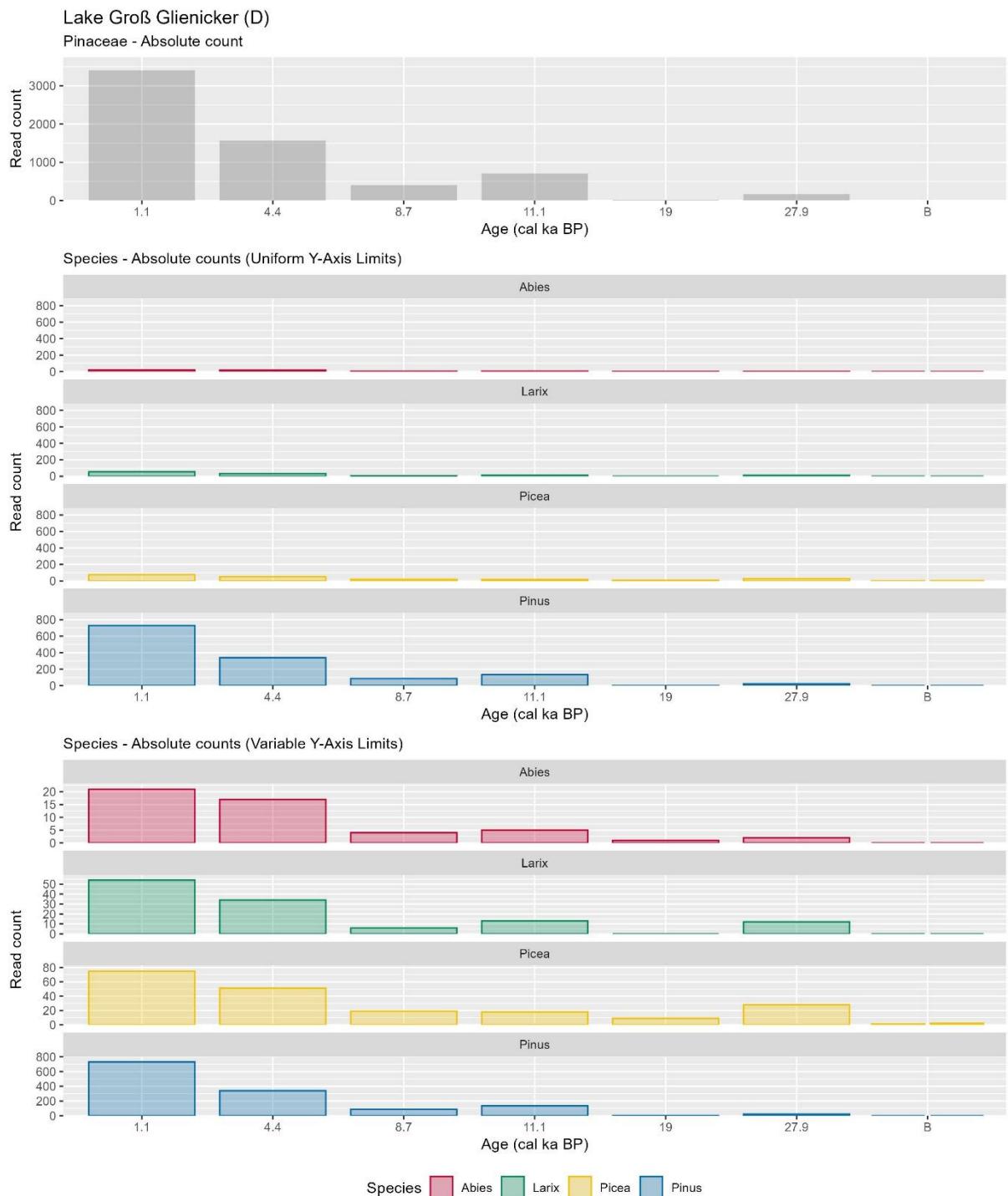
**Figure S6.** Absolute read counts of captured DNA sequences from Otto Mörtsch Cave. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



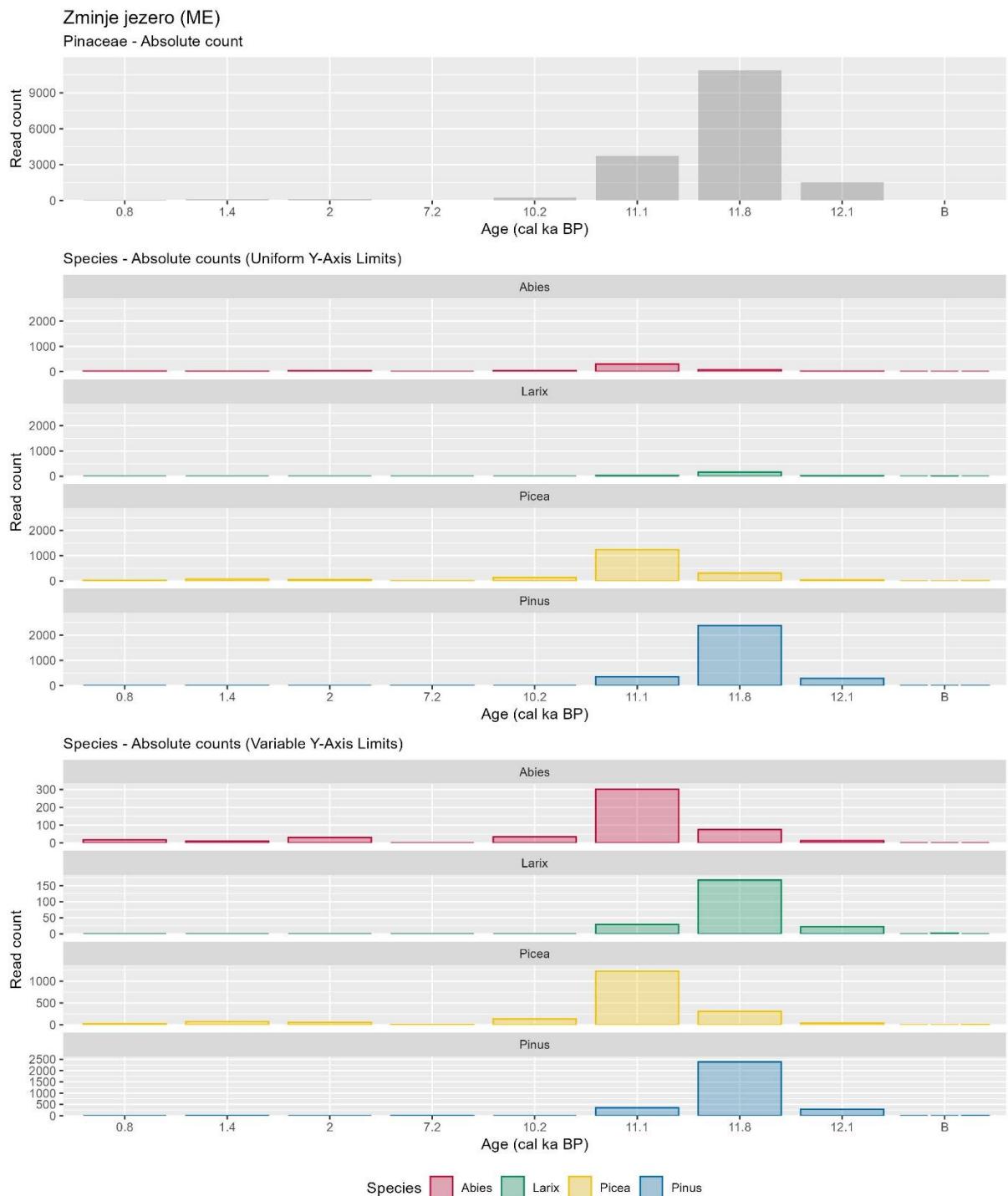
**Figure S7.** Absolute read counts of captured DNA sequences from Černé jezero. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



**Figure S8.** Absolute read counts of captured DNA sequences from Lake Sulsseewli. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



**Figure S9.** Absolute read counts of captured DNA sequences from Lake Groß Glienicker. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



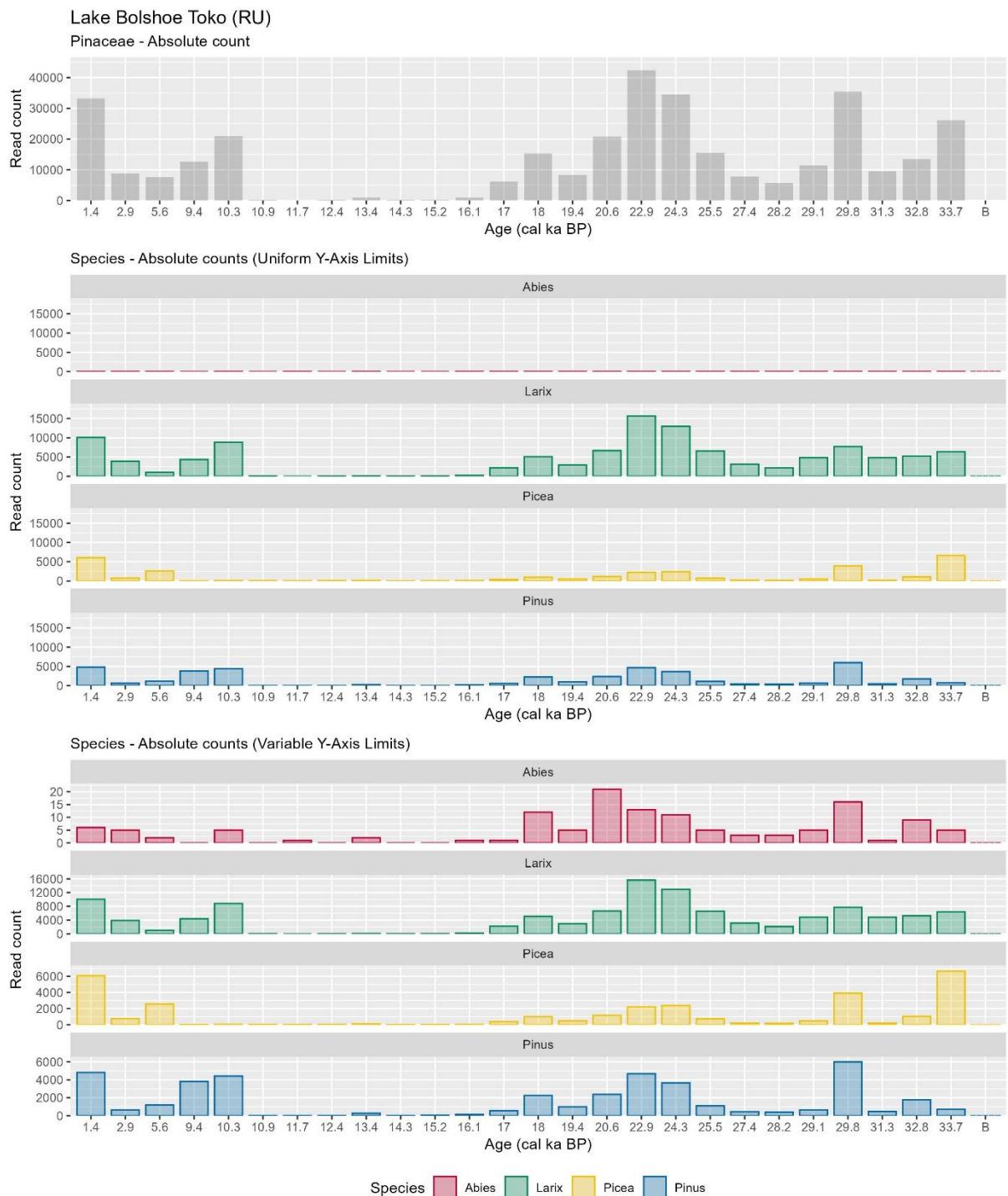
**Figure S10.** Absolute read counts of captured DNA sequences from Zminje jezero. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



**Figure S11.** Absolute read counts of captured DNA sequences from Lake Lama. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



**Figure S12.** Absolute read counts of captured DNA sequences from Lake Khamra. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



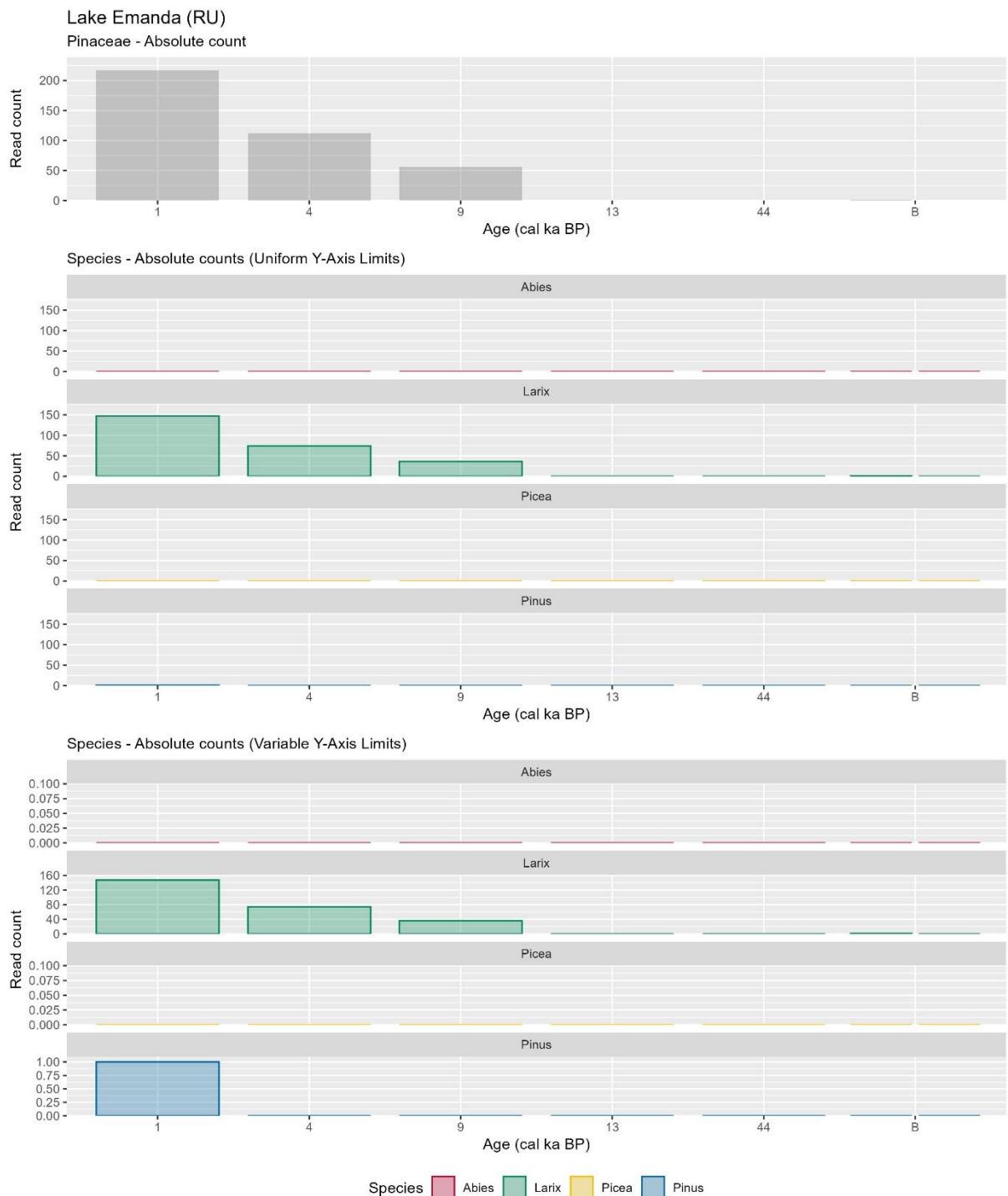
**Figure S13.** Absolute read counts of captured DNA sequences from Lake Bolshoe Toko. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



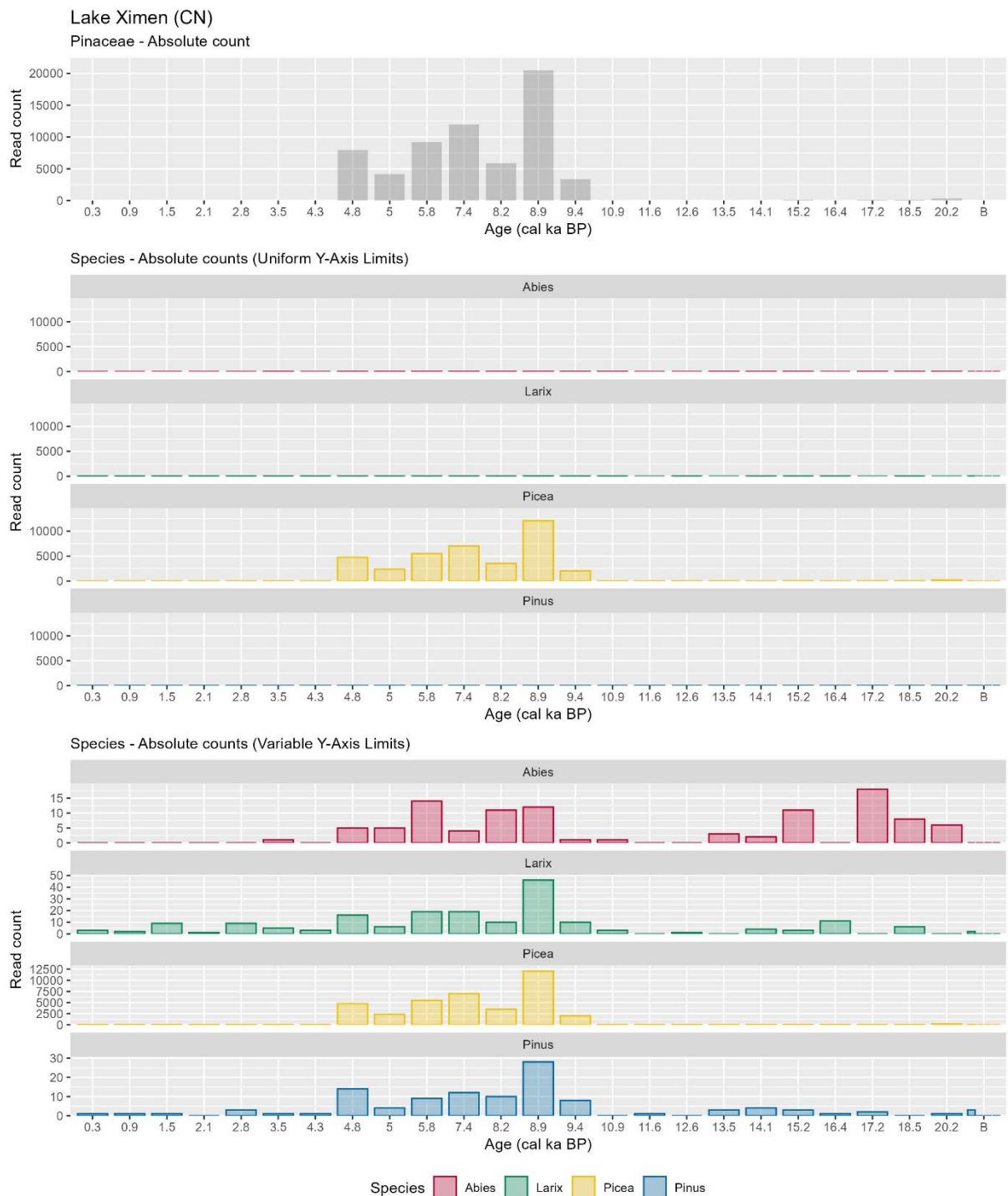
**Figure S14.** Absolute read counts of captured DNA sequences from Lake Ulu. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



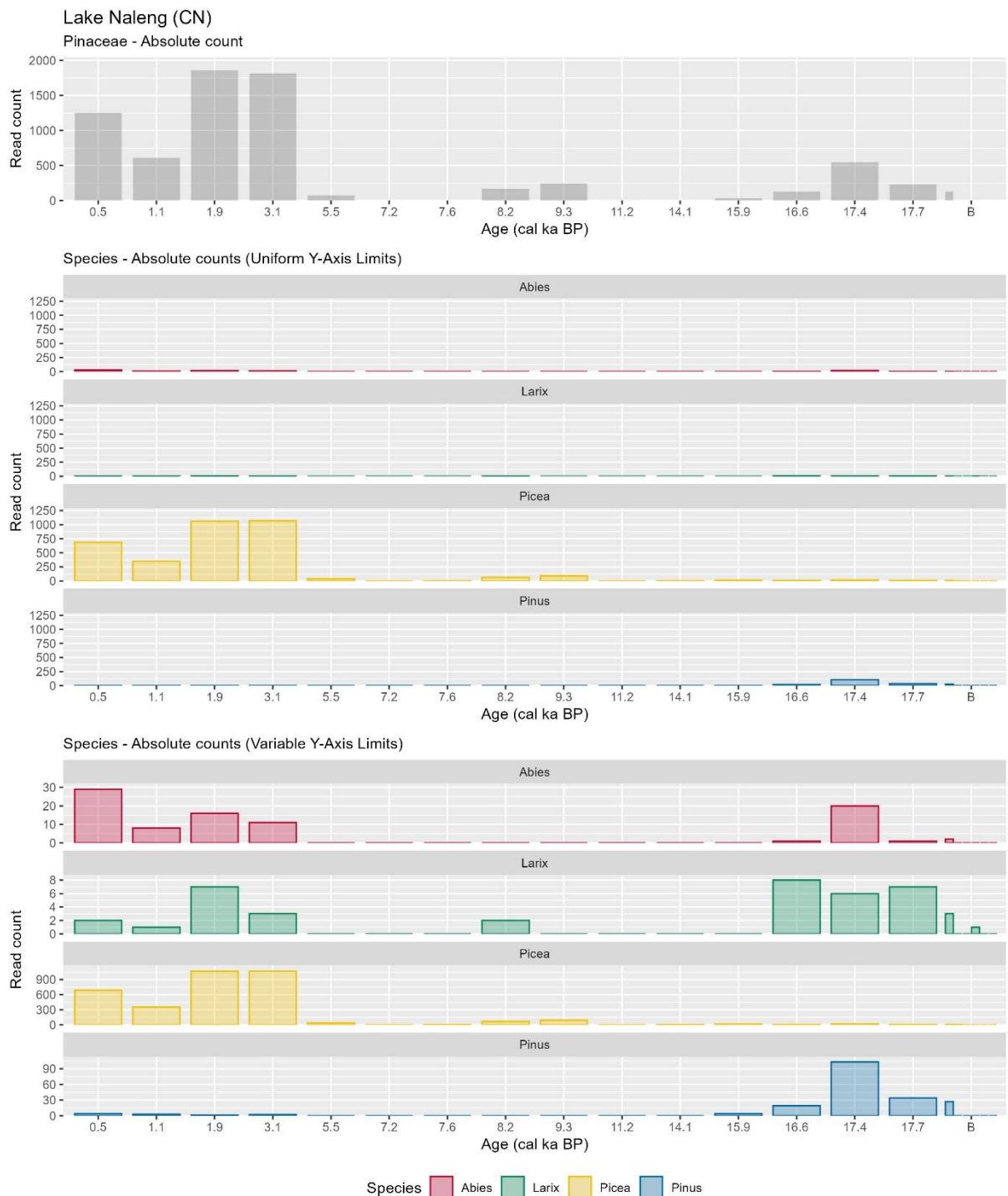
**Figure S15.** Absolute read counts of captured DNA sequences from Batagay megaslump. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



**Figure S16.** Absolute read counts of captured DNA sequences from Lake Emanda. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



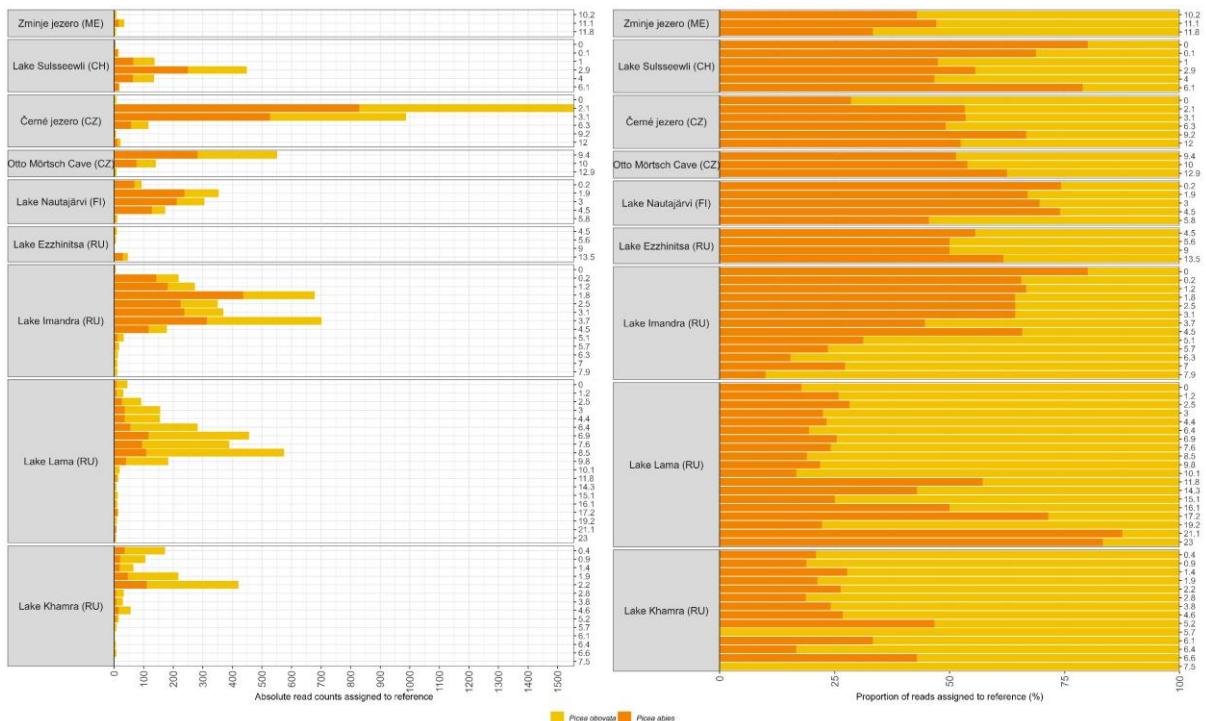
**Figure S17.** Absolute read counts of captured DNA sequences from Lake Ximen. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



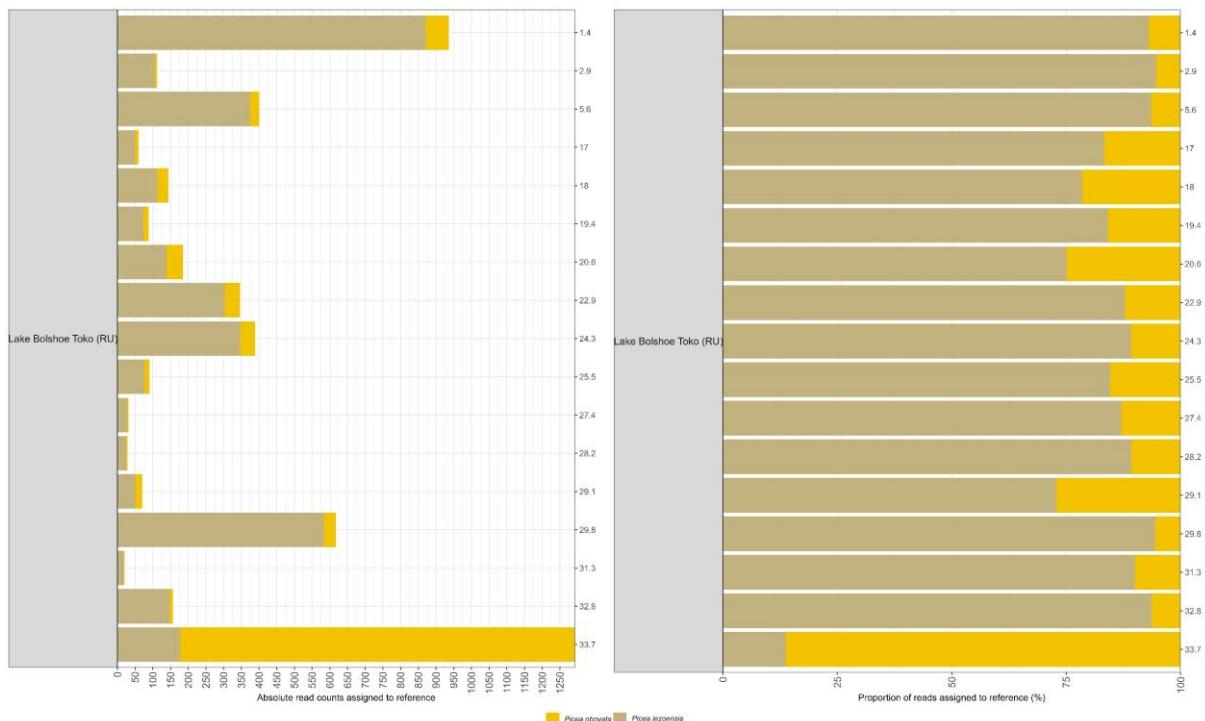
**Figure S18.** Absolute read counts of captured DNA sequences from Lake Naleng. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



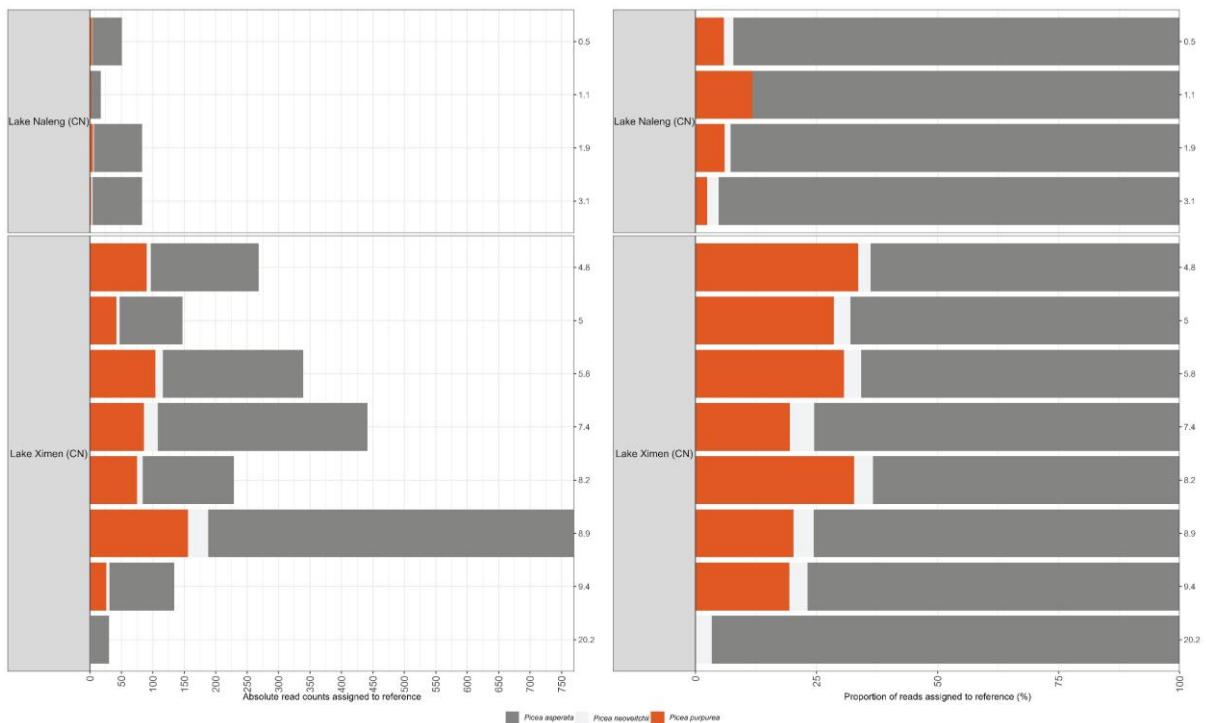
**Figure S19.** Absolute read counts of captured DNA sequences from Hidden Lake. The upper panel shows the number of taxonomically assigned reads to Pinaceae. The lower panels display absolute counts of genus-specific high-quality reads resulting from competitive mapping against the concatenated reference genome, composed of the four complete genus-specific chloroplast genomes used for hybridization bait design (**Supplementary Data 3a**). These are shown using both uniform y-axis limits (for better comparison of species abundances) and variable y-axis limits (for better visualization of low-abundance taxa). Data are ordered by sample age.



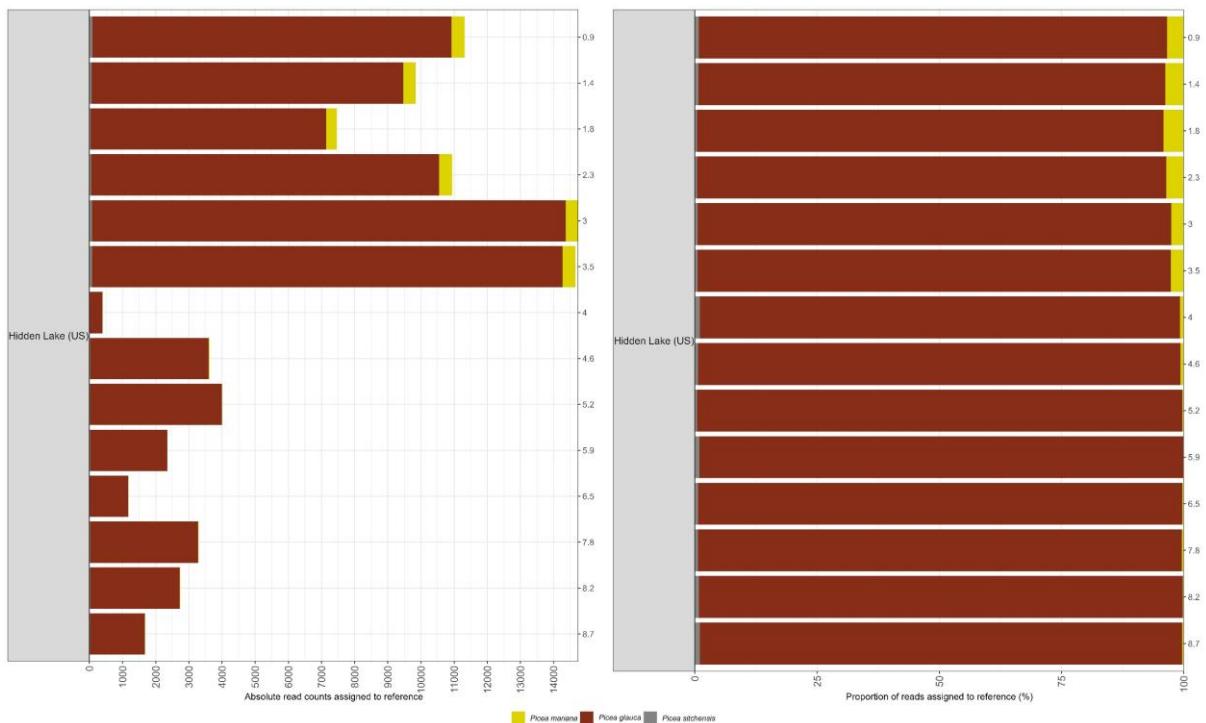
**Figure S20.** Absolute counts and proportions of read counts assigned to reference by SNP-based species identification of *Picea*-specific reads mapped to the cp reference genome of the phylogenetically and geographically distant species *Picea pungens* (NC\_067714.1). Reads were assigned based on single-species variants ascertained from the alignment of species-specific cp reference genomes to *Picea pungens* (**Supplementary Data 3d**). Each bar represents a sample, ordered by calibrated age (cal ka BP) within each site.



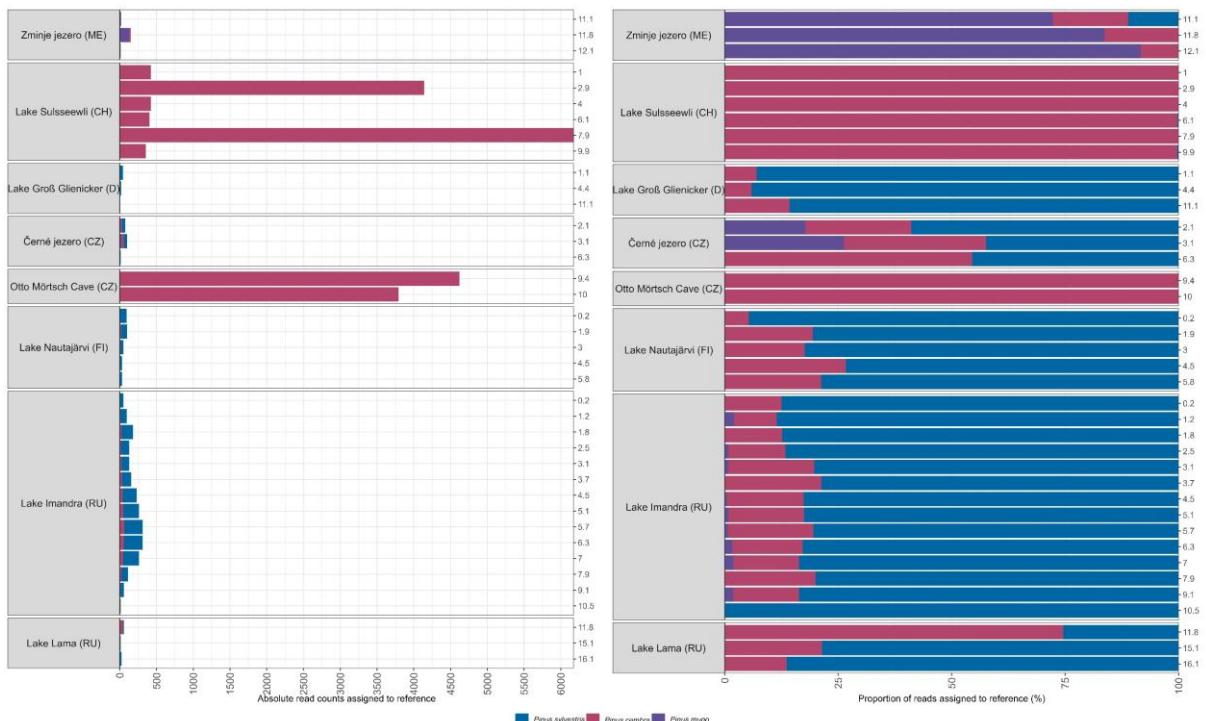
**Figure S21.** Absolute counts and proportions of read counts assigned to reference by SNP-based species identification of *Picea*-specific reads mapped to the cp reference genome of the phylogenetically and geographically distant species *Picea pungens* (NC\_067714.1). Reads were assigned based on single-species variants ascertained from the alignment of species-specific cp reference genomes to *Picea pungens* (**Supplementary Data 3d**). Each bar represents a sample, ordered by calibrated age (cal ka BP) within each site.



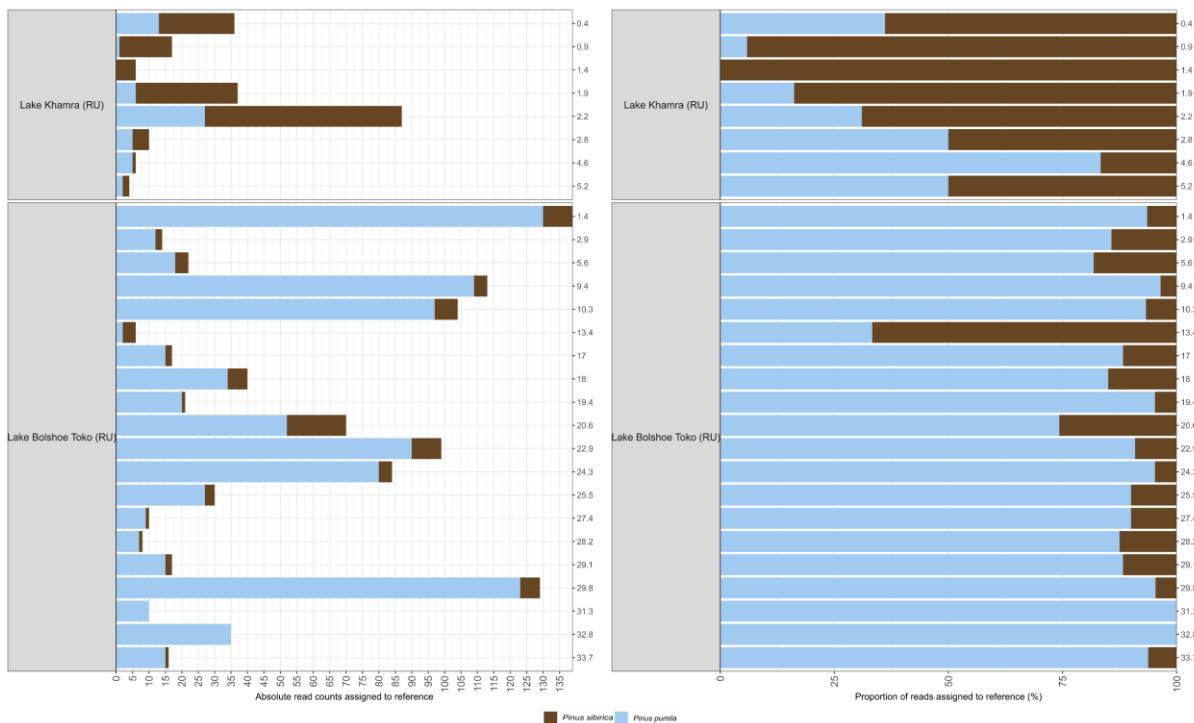
**Figure S22.** Absolute counts and proportions of read counts assigned to reference by SNP-based species identification of *Picea*-specific reads mapped to the cp reference genome of the phylogenetically and geographically distant species *Picea pungens* (NC\_067714.1). Reads were assigned based on single-species variants ascertained from the alignment of species-specific cp reference genomes to *Picea pungens* (**Supplementary Data 3d**). Each bar represents a sample, ordered by calibrated age (cal ka BP) within each site.



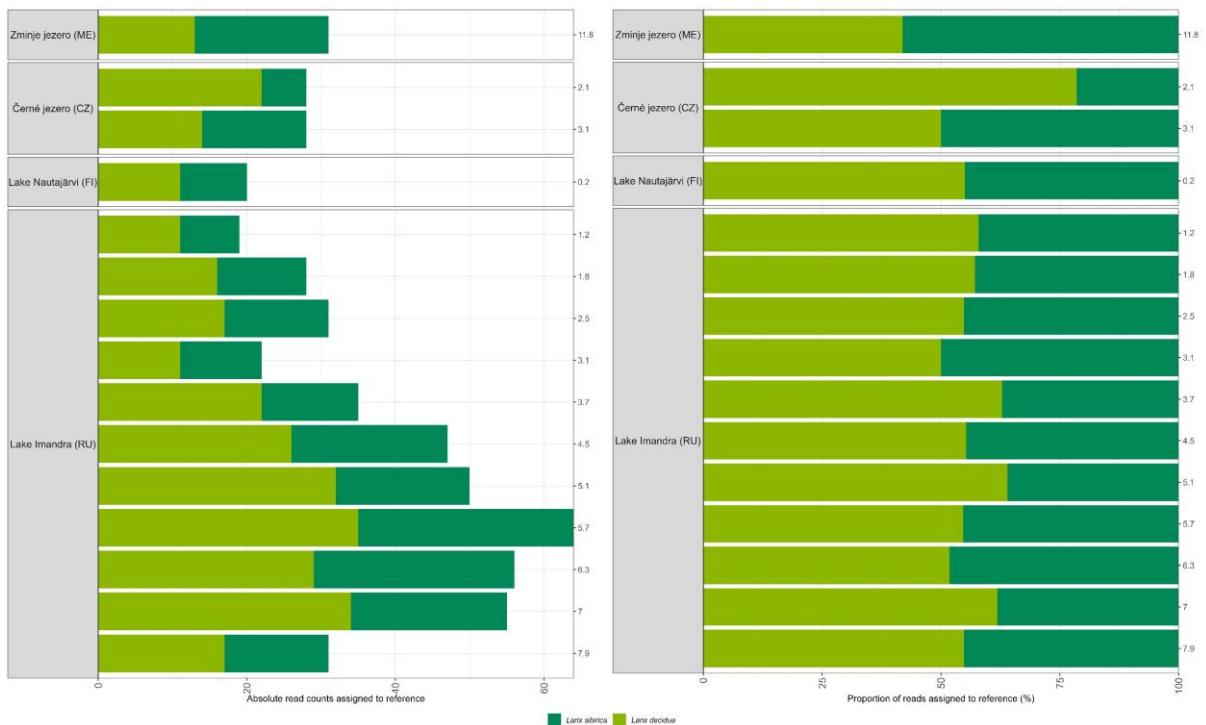
**Figure S23.** Absolute counts and proportions of read counts assigned to reference by SNP-based species identification of *Picea*-specific reads mapped to the cp reference genome of the phylogenetically and geographically distant species *Picea pungens* (NC\_067714.1). Reads were assigned based on single-species variants ascertained from the alignment of species-specific cp reference genomes to *Picea pungens* (**Supplementary Data 3d**). Each bar represents a sample, ordered by calibrated age (cal ka BP) within each site.



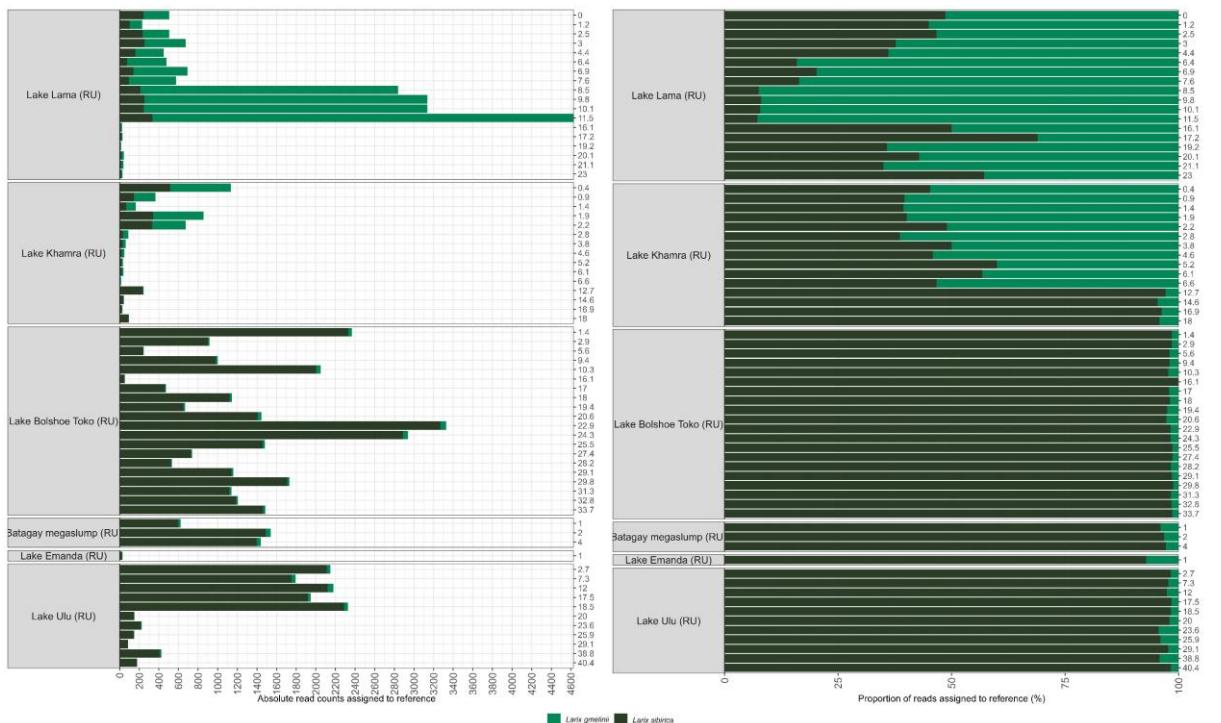
**Figure S24.** Absolute counts and proportions of read counts assigned to reference by SNP-based species identification of *Pinus*-specific reads mapped to the cp reference genome of the phylogenetically and geographically distant species *Pinus taeda* (NC\_021440.1). Reads were assigned based on single-species variants ascertained from the alignment of species-specific cp reference genomes to *Pinus taeda* (**Supplementary Data 3d**). Each bar represents a sample, ordered by calibrated age (cal ka BP) within each site.



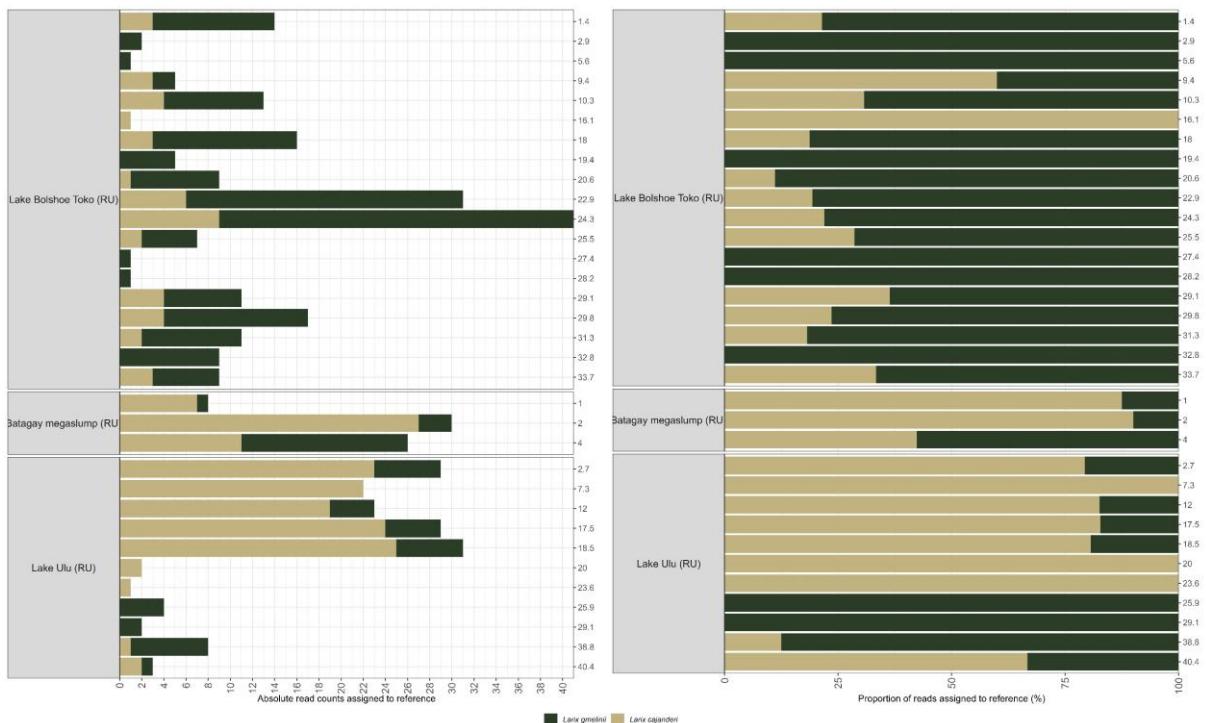
**Figure S25.** Absolute counts and proportions of read counts assigned to reference by SNP-based species identification of *Pinus*-specific reads mapped to the cp reference genome of the phylogenetically and geographically distant species *Pinus taeda* (NC\_021440.1). Reads were assigned based on single-species variants ascertained from the alignment of species-specific cp reference genomes to *Pinus taeda* (**Supplementary Data 3d**). Each bar represents a sample, ordered by calibrated age (cal ka BP) within each site.



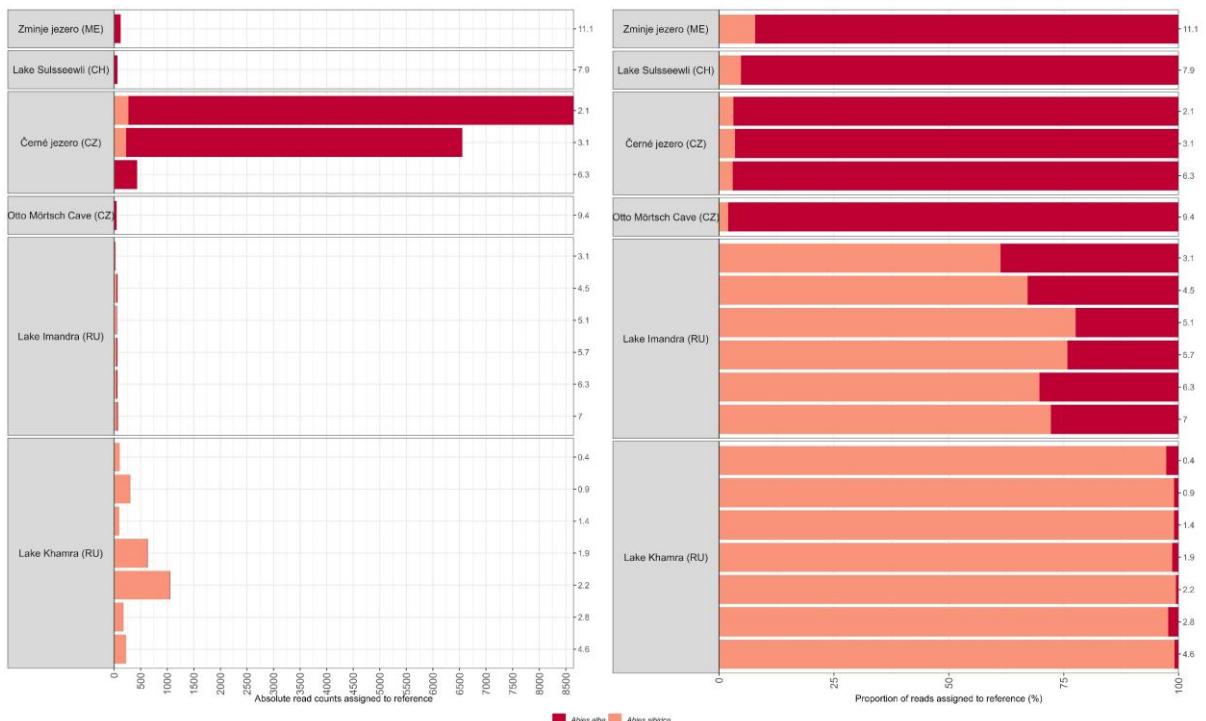
**Figure S26.** Absolute counts and proportions of read counts assigned to reference by SNP-based species identification of *Larix*-specific reads mapped to the cp reference genome of the phylogenetically and geographically distant species *Larix potaninii* (KX880508.1). Reads were assigned based on single-species variants ascertained from the alignment of species-specific cp reference genomes to *Larix potaninii* (**Supplementary Data 3d**). Each bar represents a sample, ordered by calibrated age (cal ka BP) within each site.



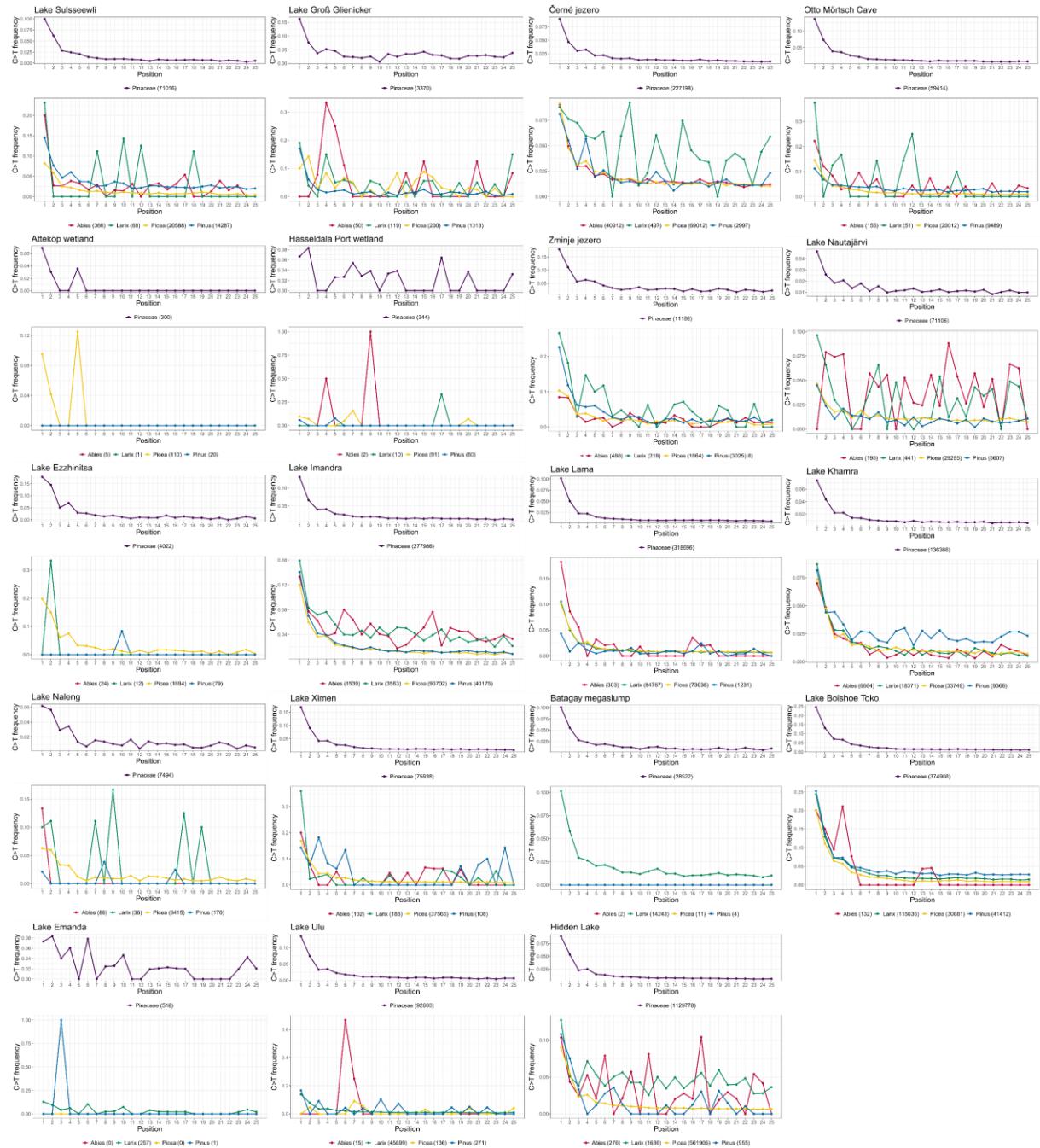
**Figure S27.** Absolute counts and proportions of read counts assigned to reference by SNP-based species identification of *Larix*-specific reads mapped to the cp reference genome of the phylogenetically and geographically distant species *Larix potaninii* (KX880508.1). Reads were assigned based on single-species variants ascertained from the alignment of species-specific cp reference genomes to *Larix potaninii* (**Supplementary Data 3d**). Each bar represents a sample, ordered by calibrated age (cal ka BP) within each site.



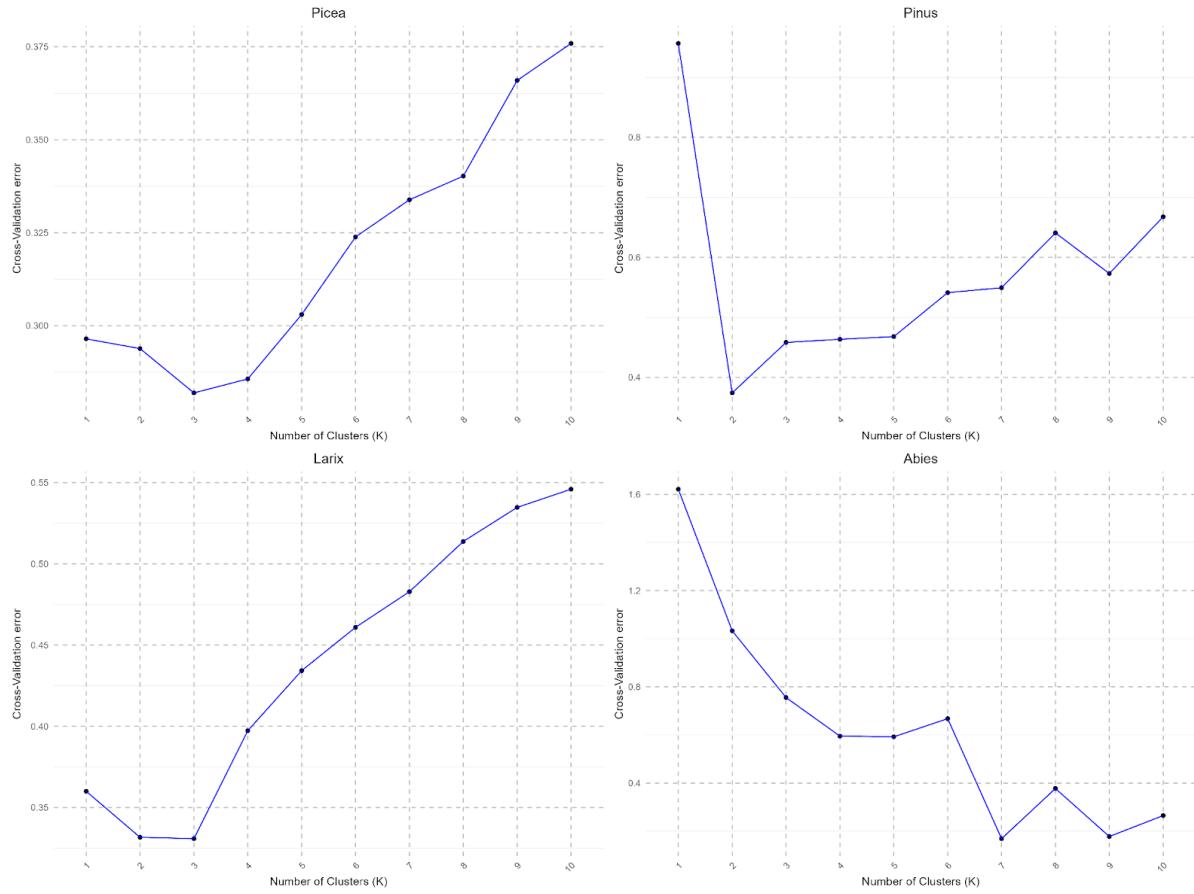
**Figure S28.** Absolute counts and proportions of read counts assigned to reference by SNP-based species identification of *Larix*-specific reads mapped to the cp reference genome of the phylogenetically and geographically distant species *Larix potaninii* (KX880508.1). Reads were assigned based on single-species variants ascertained from the alignment of species-specific cp reference genomes to *Larix potaninii* (**Supplementary Data 3d**). Each bar represents a sample, ordered by calibrated age (cal ka BP) within each site.



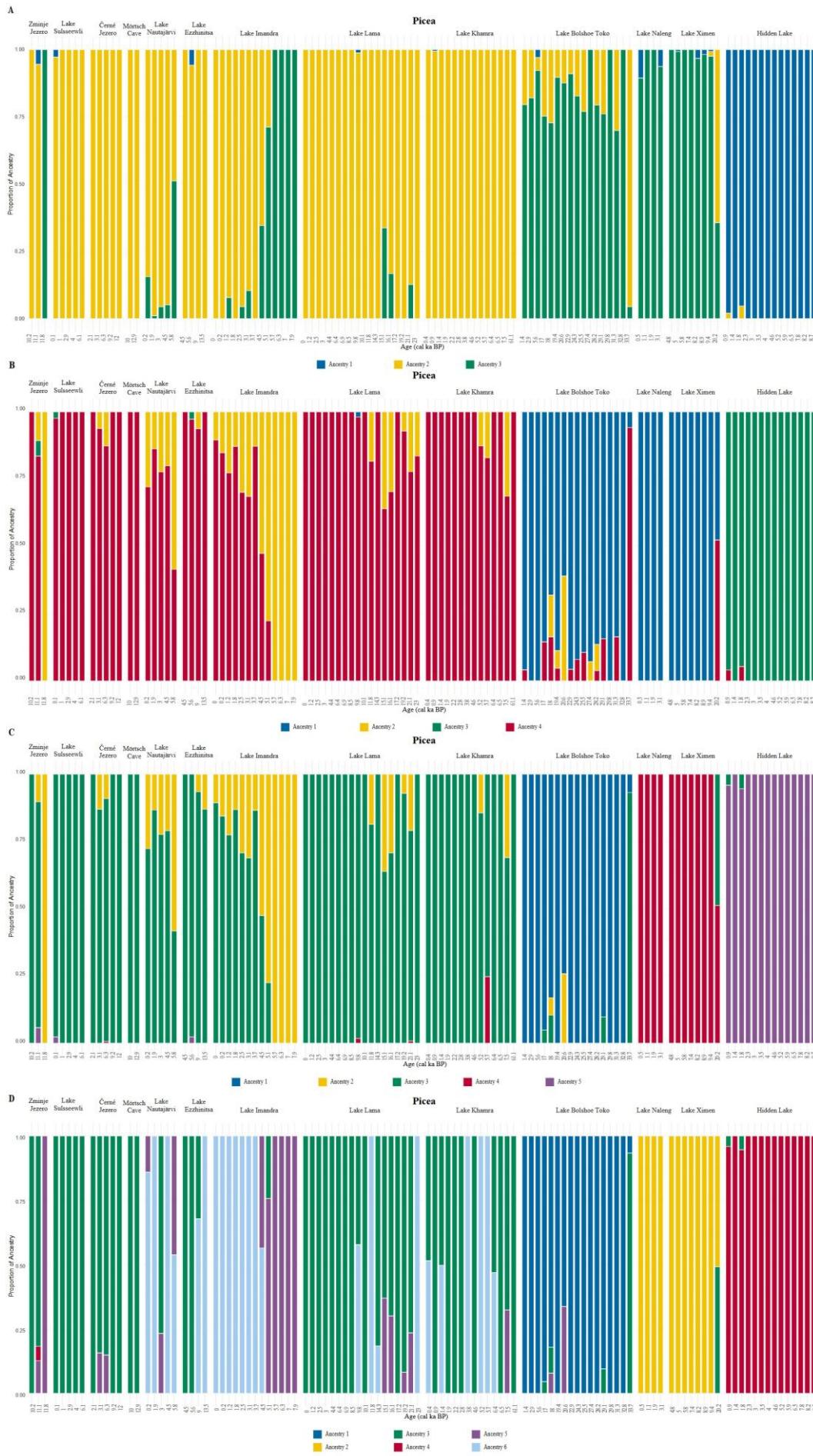
**Figure S29.** Absolute counts and proportions of read counts assigned to reference by SNP-based species identification of *Abies*-specific reads mapped to the cp reference genome of the phylogenetically and geographically distant species *Abies balsamea* (NC\_042778.1). Reads were assigned based on single-species variants ascertained from the alignment of species-specific cp reference genomes to *Abies balsamea* (**Supplementary Data 3d**). Each bar represents a sample, ordered by calibrated age (cal ka BP) within each site.



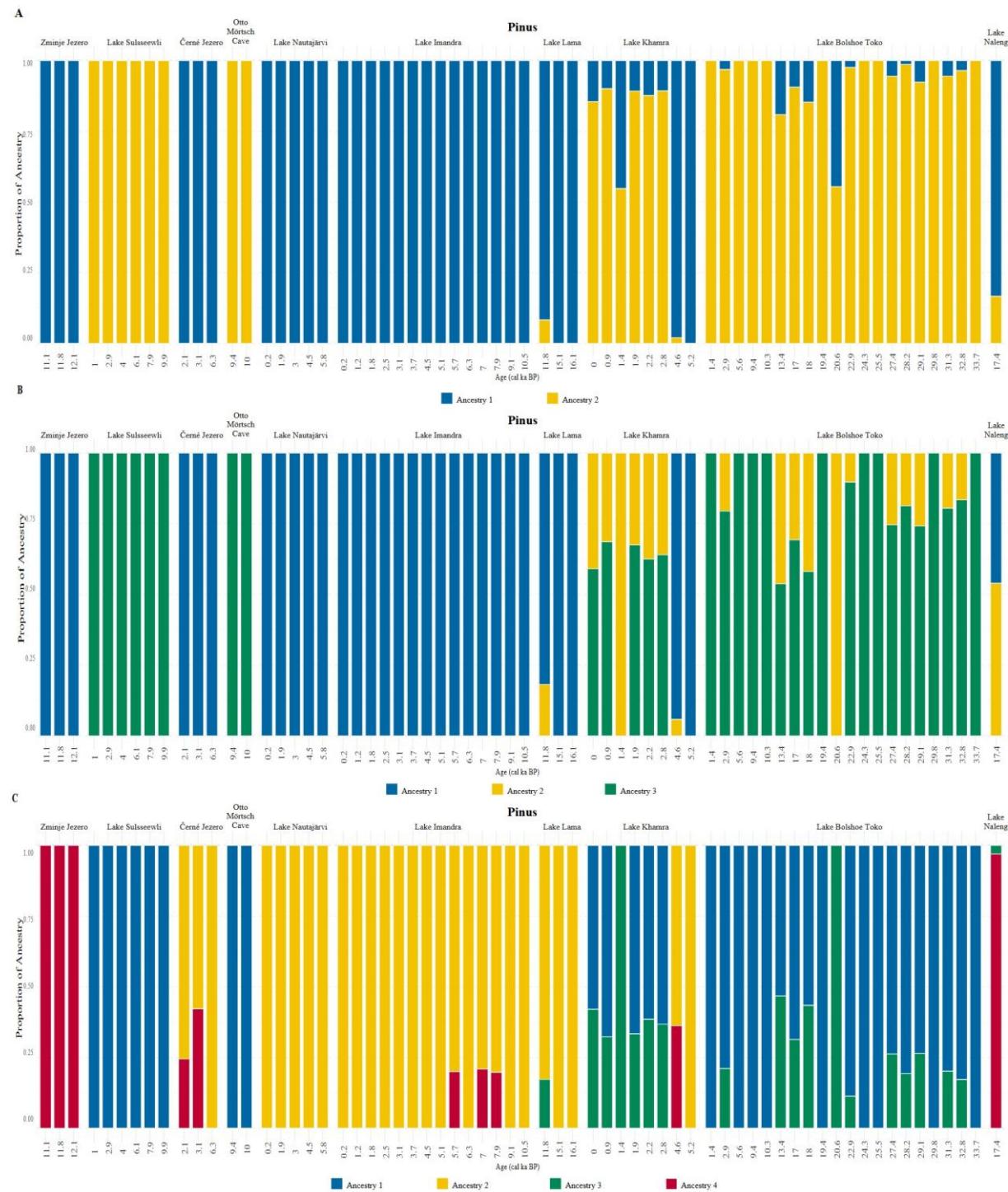
**Figure S30.** The upper panels display the C>T substitution frequencies across the first 25 positions at the 5' end of Pinaceae assigned reads, merged from all samples at each site. The legend indicates the Pinaceae total number of reads. The lower panels show the C>T substitution frequencies across the first 25 positions at the 5' end of genus-specific reads, merged from all samples at each site. The legend indicates the colors for each genus, along with the total number of merged reads for each genus.



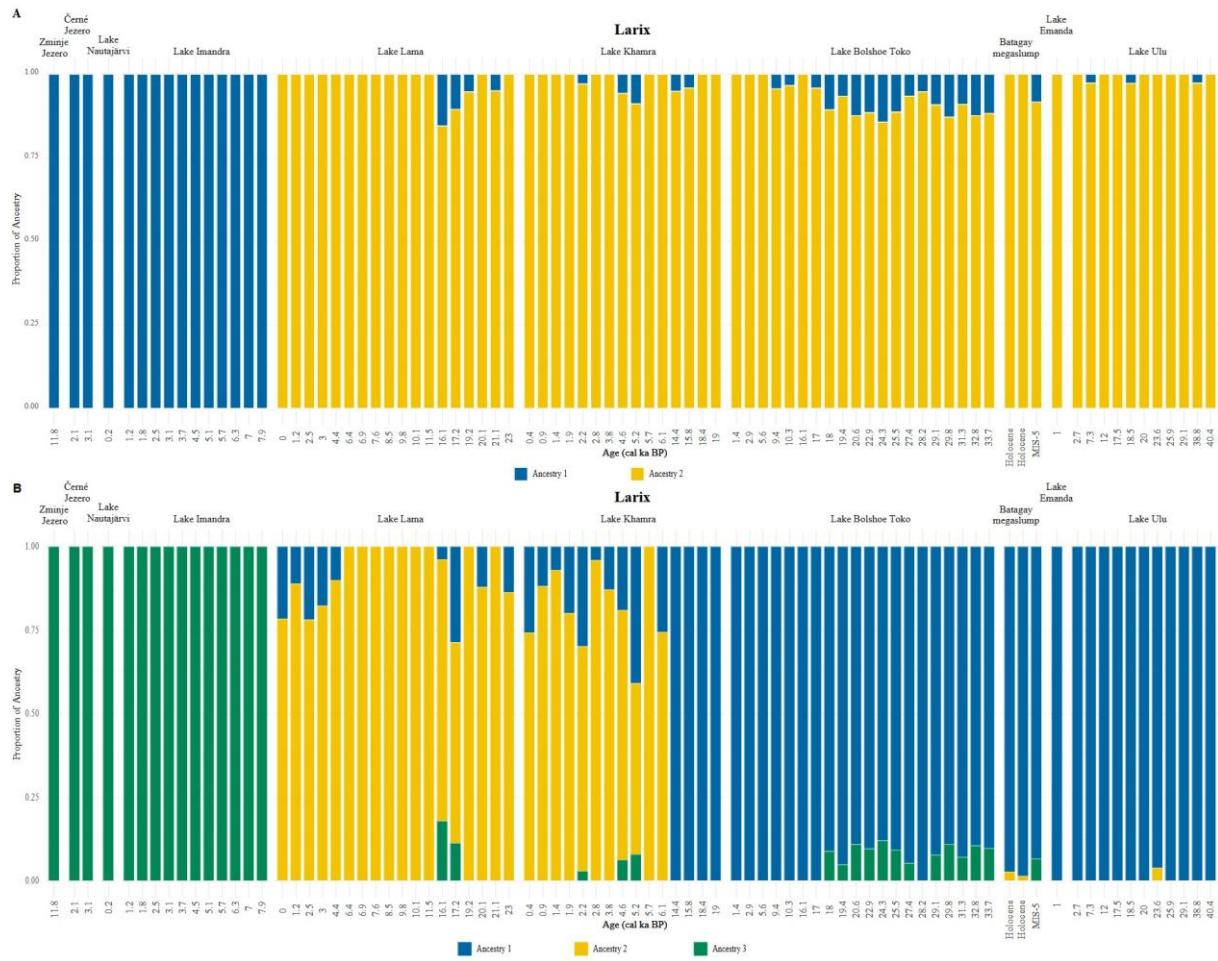
**Figure S31** Cross-validation error plots for determining the optimal number of clusters (K) in the genetic structure based on the variations detected in the genus-specific reads of (a) *Picea*, (b) *Pinus*, (c) *Larix*, and (d) *Abies*. Points are connected by solid blue lines, highlighting trends across different K values. The optimal K, where the cross-validation error is minimized, suggests the most appropriate number of clusters for each genus.



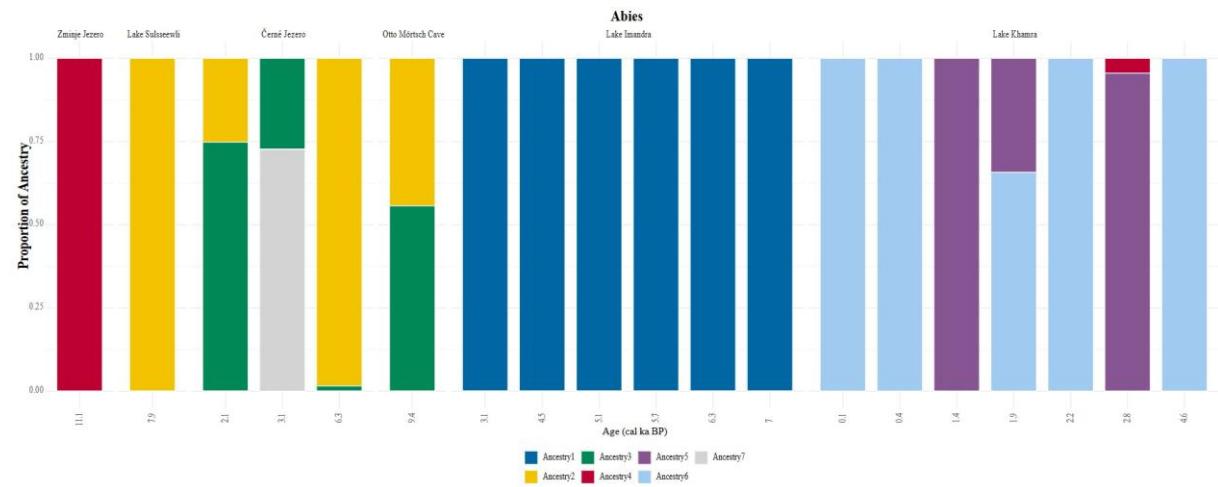
**Figure S32** Admixture plot showing the proportions of **(A)** three, **(B)** four, **(C)** five, and **(D)** six distinct genetic ancestries based on *Picea* SNP variants. Each vertical bar represents a sample, ordered by age within each site, with sites separated by empty bars and arranged by longitude. Colored segments indicate the contribution of each genetic ancestry.



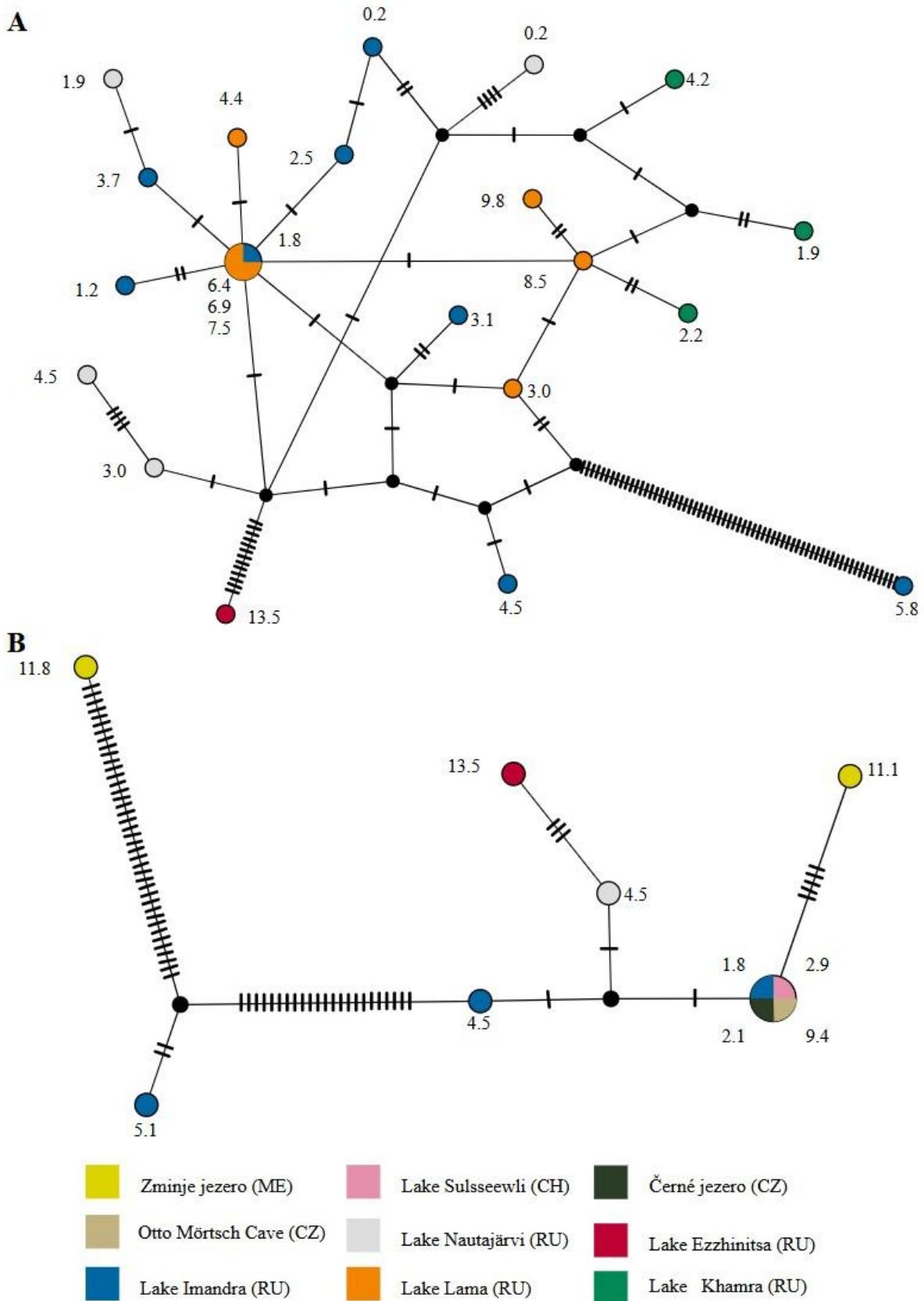
**Figure S33** Admixture plot showing the proportions of **(A)** two, **(B)** three and **(C)** four distinct genetic ancestries based on *Pinus* SNP variants. Each vertical bar represents a sample, ordered by age within each site, with sites separated by empty bars and arranged by longitude. Colored segments indicate the contribution of each genetic ancestry.



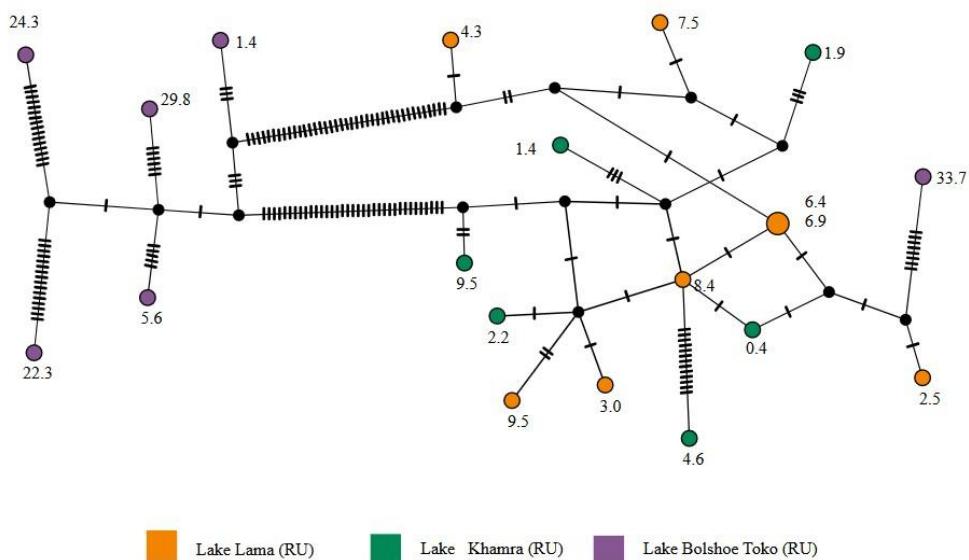
**Figure S34** Admixture plot showing the proportions of (A) two and (B) three distinct genetic ancestries based on *Larix* SNP variants. Each vertical bar represents a sample, ordered by age within each site, with sites separated by empty bars and arranged by longitude. Colored segments indicate the contribution of each genetic ancestry.



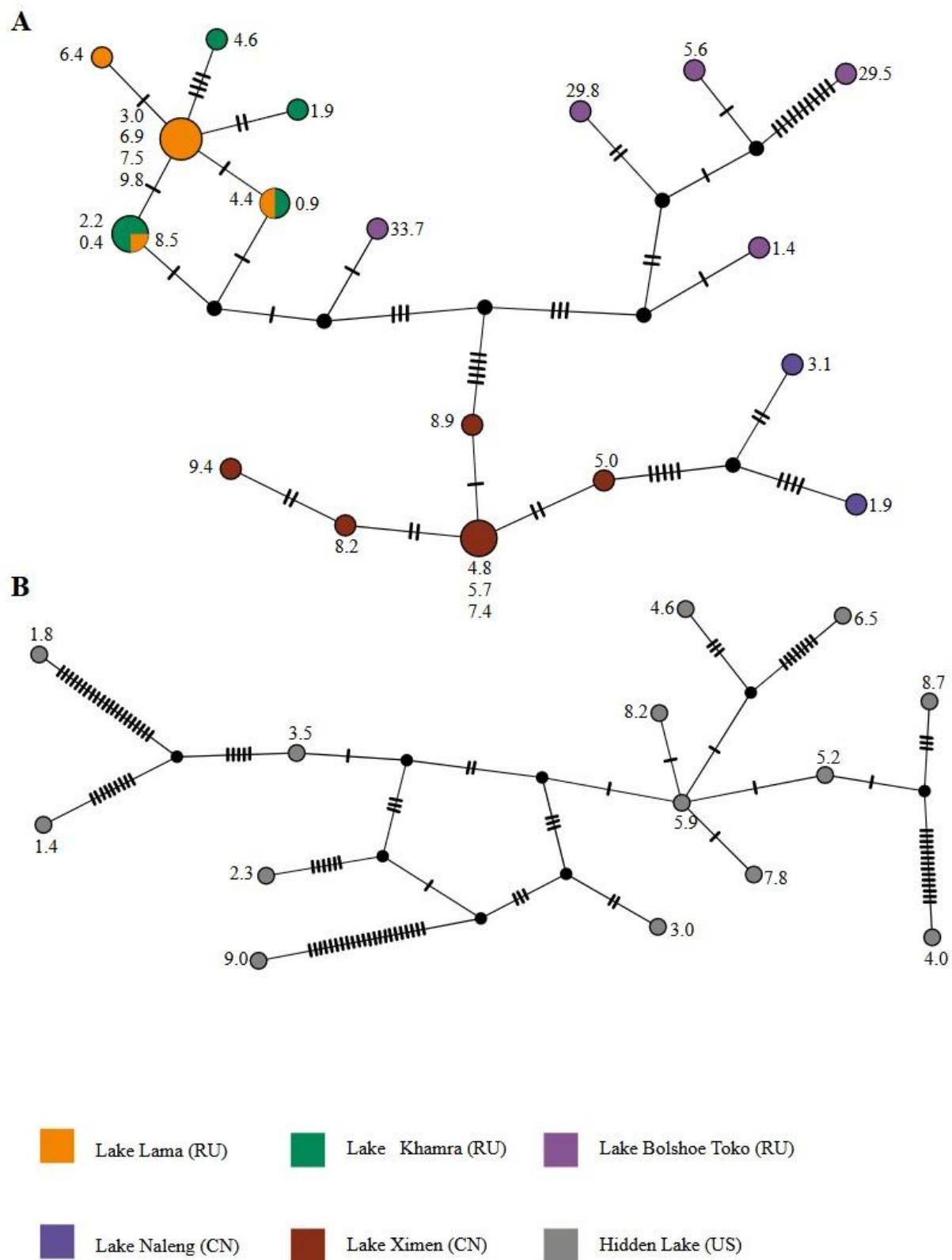
**Figure S35** Admixture plot showing the proportions of seven distinct genetic ancestries based on *Abies* SNP variants. Each vertical bar represents a sample, ordered by age within each site, with sites separated by empty bars and arranged by longitude. Colored segments indicate the contribution of each genetic ancestry.



**Figure S36** TCS haplotype networks based on *Picea* variants: (A) 96 segregating sites and (B) 70 segregating sites. Each node represents a haplotype; larger nodes indicate multiple samples sharing the same haplotype. Sample ages (cal ka BP) are shown next to nodes. Hatch marks on edges indicate inferred mutations (bp differences). Node colors correspond to sampling sites (see legend); black internal nodes represent inferred ancestral haplotypes.



**Figure S37** TCS haplotype networks based on *Picea* variants: 95 segregating sites. Each node represents a haplotype; larger nodes indicate multiple samples sharing the same haplotype. Sample ages (cal ka BP) are shown next to nodes. Hatch marks on edges indicate inferred mutations (bp differences). Node colors correspond to sampling sites (see legend); black internal nodes represent inferred ancestral haplotypes.



**Figure S38.** TCS haplotype networks based on *Picea* variants: (A) 49 segregating sites and (B) 82 segregating sites. Each node represents a haplotype; larger nodes indicate multiple samples sharing the same haplotype. Sample ages (cal ka BP) are shown next to nodes. Hatch marks on edges indicate inferred mutations (bp differences). Node colors correspond to sampling sites (see legend); black internal nodes represent inferred ancestral haplotypes.