PNAS
www.pnas.org

Supplementary Information for

Genomic evidence of survival near ice sheet margins for some, but not all, North American trees

Jordan B. Bemmels, L. Lacey Knowles, Christopher W. Dick

Corresponding author: Jordan. B. Bemmels
Email: jbemmels@umich.edu

This PDF file includes:

Supplementary text
Figs. S1 to S8
Tables S1 to S3
Captions for Datasets S1 to S4
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Datasets S1 to S4

Supplementary Information Text

SNP discovery

Complete details of SNP genotyping have previously been described (1), but we report pertinent details here. SNPs were identified *de novo* from double-digest Restriction-Associated DNA (ddRAD) sequencing reads (2) generated on an *Illumina HiSeq* (Illumina; San Diego, CA). *Stacks* v. 1.44-1.46 (3, 4) was used to identify loci and call genotypes. Because all individuals included in the present study were previously genotyped (1), and all previously genotyped individuals across all populations were used to construct the catalogue of SNPs (i.e., no reference panel was used), our data are not subject to ascertainment bias. Only one SNP per ddRAD locus was retained, ensuring that SNPs are putatively unlinked and distributed randomly throughout the genome.

SNPs were filtered to a minimum minor allele frequency (MAF) of 3.3%. This is both the minimum detectable frequency in at least one population of each species, and an ideal minimum MAF for representing rangewide genetic structure (1). To determine the ideal minimum MAF, Bemmels and Dick (1) examined principal component (PC) plots of raw SNP genotypes filtered to a low, medium, and high MAF (1%, 3.3%, and 5%). There was little difference in PC plots between 3.3% and 5% minimum MAF, and patterns clearly distinguished groups of populations. However, at 1% minimum MAF, the PC plots behaved poorly in *C. cordiformis* and were prone to generating clusters of a few outlier individuals, suggesting that broad-scale geographic patterns were confounded by the close genetic relatedness among groups of individuals generated by more recent, local-scale processes (1, 5). The medium MAF of 3.3% was therefore the most appropriate for representing rangewide genetic structure while retaining the maximum

number of SNPs. Note that the minimum MAF in our simulations was also set to 3.3%, to match that of the empirical data.

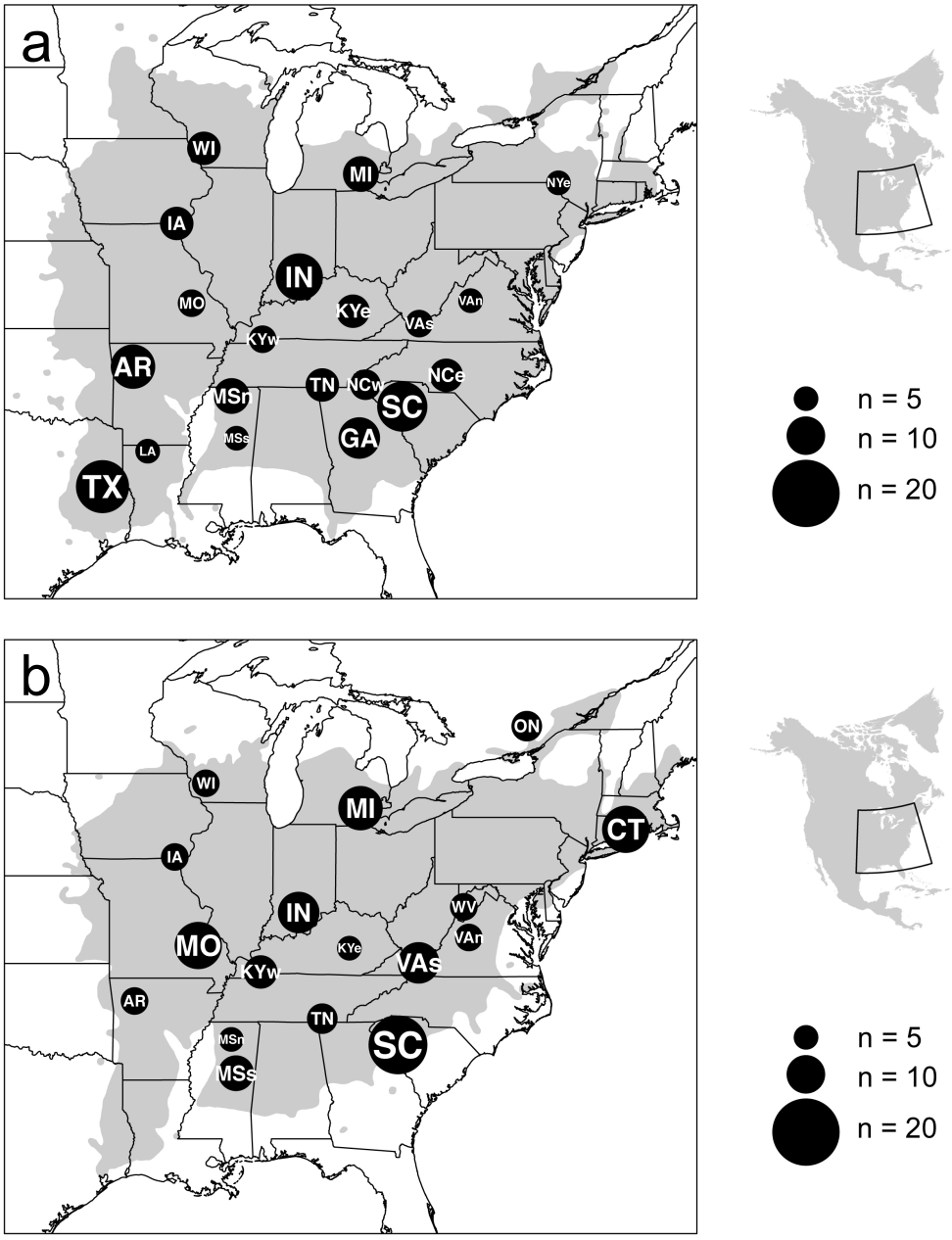

Fig S1. Geographic locations of sampled populations for (a) *Carya cordiformis* and (b) *Carya ovata*. The size of the black circle is proportional to the number of individuals sampled in each population, with population IDs written in white text. The geographic distribution of each species (6) is shown in grey shading.

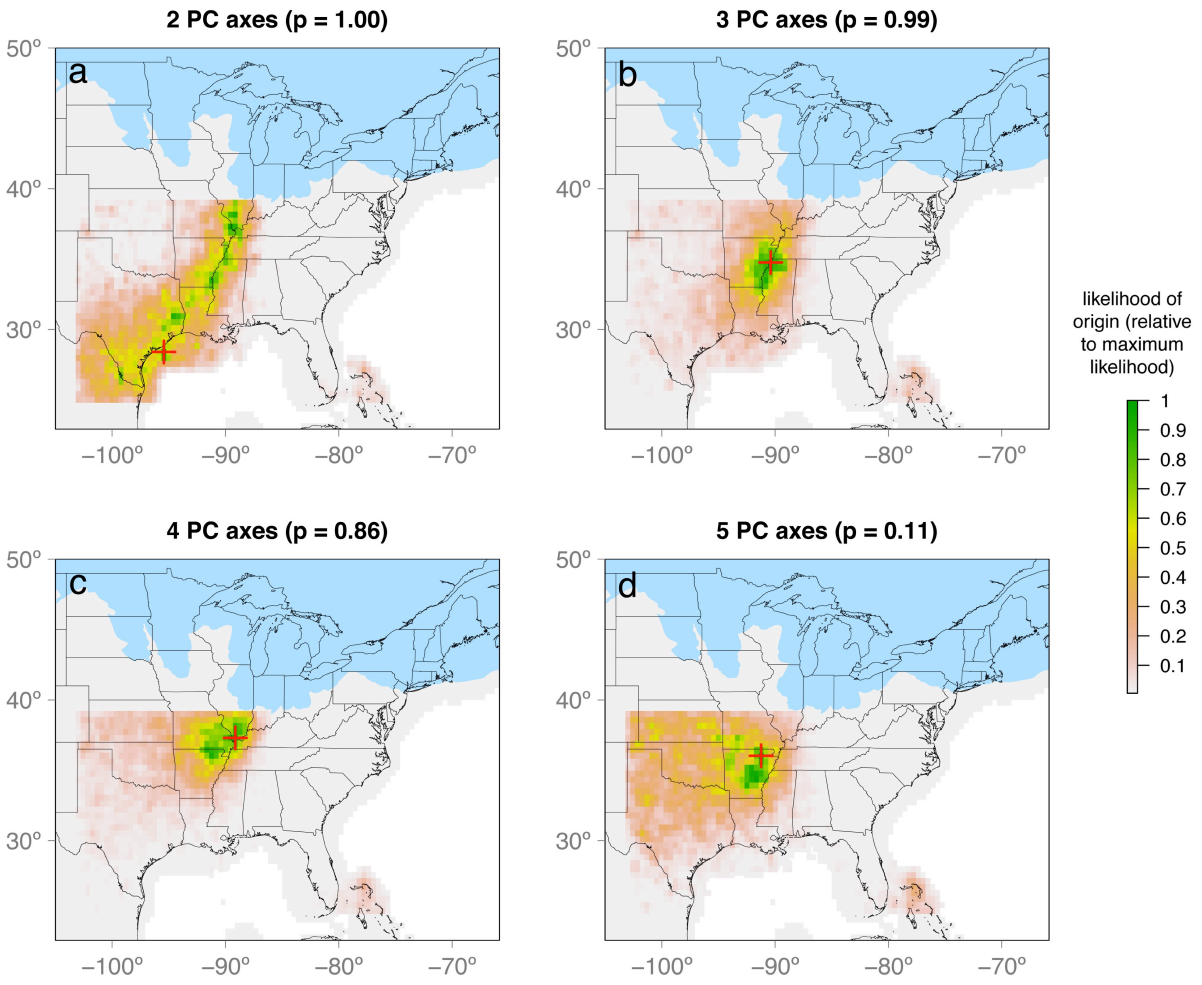

Fig. S2. Extended results for *Carya cordiformis*. Estimated expansion origin (Ω; red cross) in *Carya cordiformis* when different numbers of PC axes are retained. The shading of pixels depicts the kernel density showing the likelihood that each pixel served as the expansion origin, relative to the pixel with the maximum likelihood (i.e., Ω). Glaciated regions are shown in blue. Wegmann's *p*-value (7) is reported for each number of PC axes retained. Note that low *p*-values indicate the model is unable to easily generate the empirical data for this number of PC axes, and results for low *p*-values should be interpreted with caution.

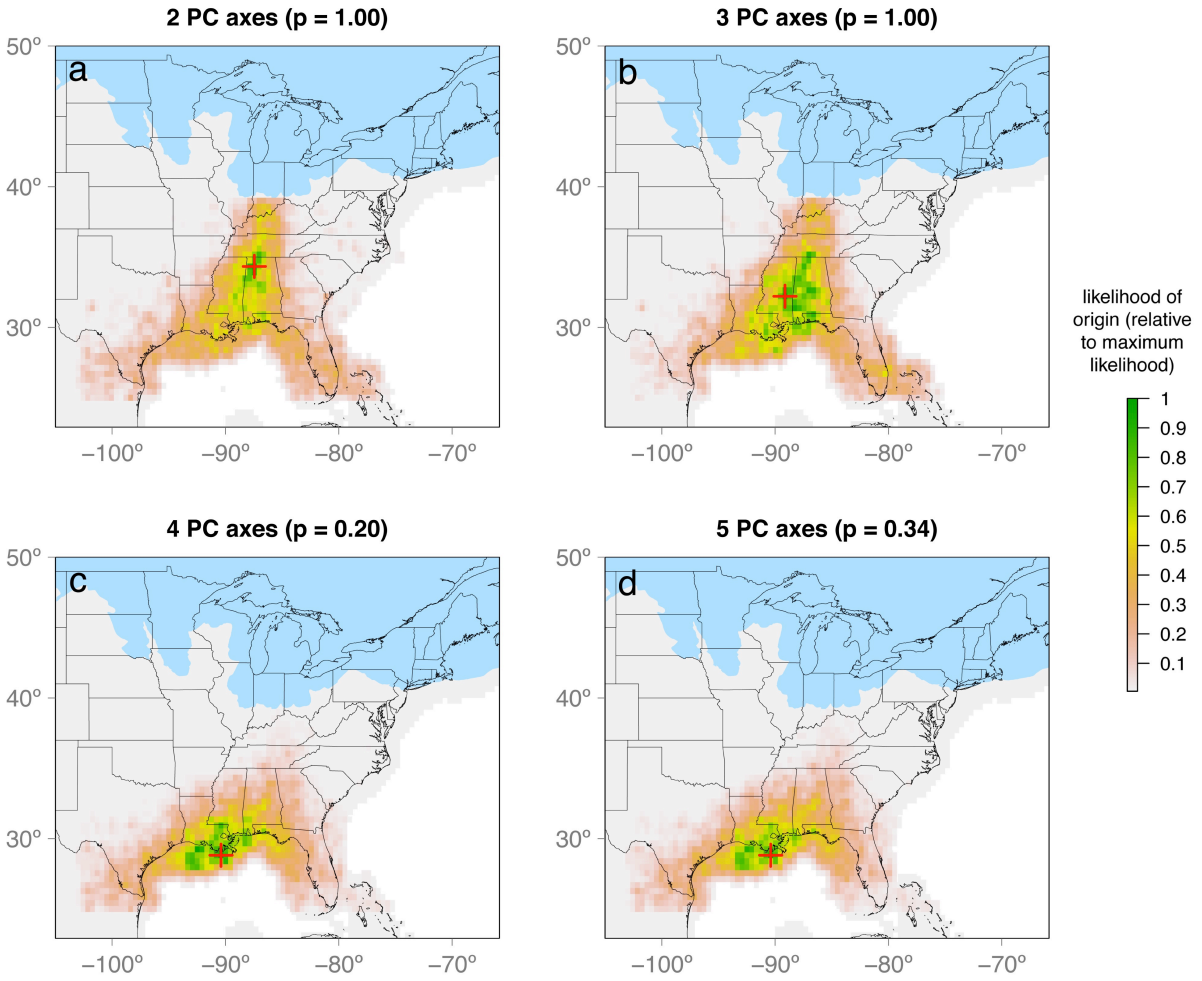

Fig S3. Extended results for *Carya ovata.* Estimated expansion origin (Ω; red cross) in *Carya ovata* when different numbers of PC axes are retained. The shading of pixels depicts the kernel density showing the likelihood that each pixel served as the expansion origin, relative to the pixel with the maximum likelihood (i.e., Ω). Glaciated regions are shown in blue. Wegmann's *p*-value (7) is reported for each number of PC axes retained. Note that low *p*-values indicate the model is unable to easily generate the empirical data for this number of PC axes, and results for low *p*-values should be interpreted with caution.

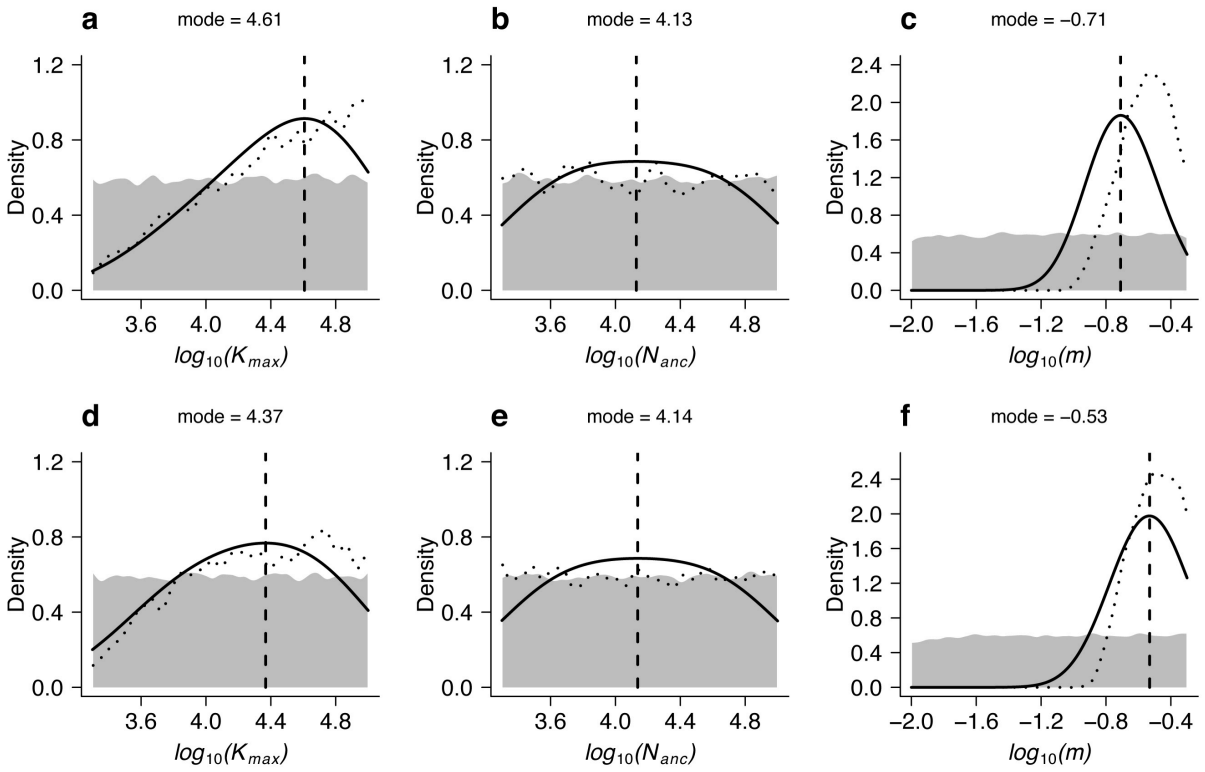

Fig S4. Demographic parameter estimates (dashed vertical lines) and prior and posterior distributions for (a-c) *Carya cordiformis* and (d-f) *Carya ovata*. Grey shading: prior distribution; dotted line: posterior distribution of raw retained simulations; solid black line: posterior distribution after ABC-GLM regression adjustment (8); K_{max}: maximum carrying capacity; N_{anc}: ancestral population size, m: migration rate.

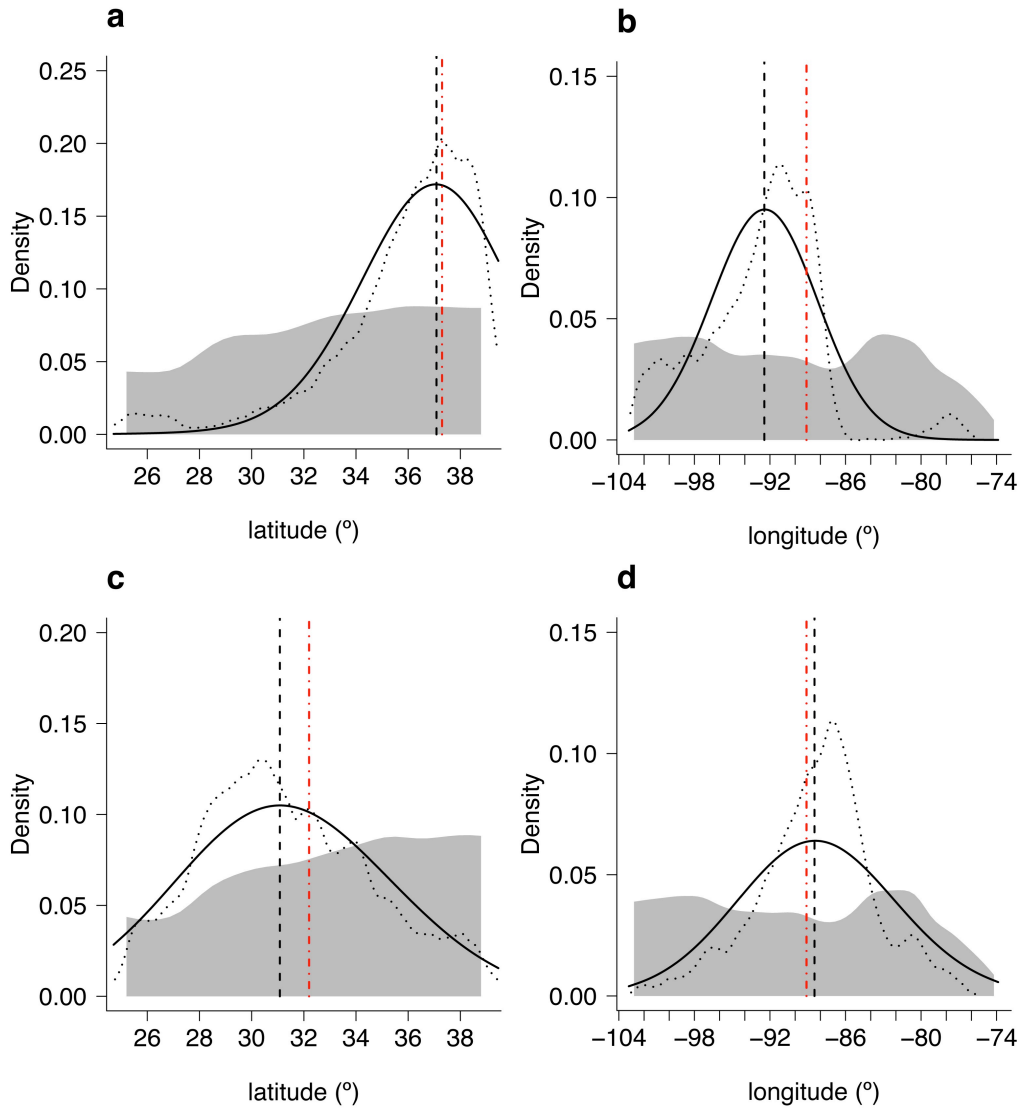

Fig S5. Geographic parameter estimates and prior and posterior distributions for (a-b) *Carya cordiformis* and (c-d) *Carya ovata*. Grey shading: prior distribution; dotted line: posterior distribution of raw retained simulations; solid black line: posterior distribution after ABC-GLM regression adjustment (8). The dashed vertical black line represents the mode of the adjusted posterior distribution; however, posterior distributions of latitude and longitude considered in isolation should be interpreted with caution due to the two-dimensional nature of the simulations. We therefore also plot the estimated expansion origin (Ω) from the two-dimensional kernel density for comparison (dashed and dotted vertical red line; see Fig. 2). Note that the shape of the uniform prior is warped because different areas are covered by land at each latitude and longitude, and demographic simulations cannot be initiated from oceans. Nonetheless, each two-dimensional latitude-longitude combination corresponding to a land pixel has an equal probability of serving as the expansion origin in the simulations.

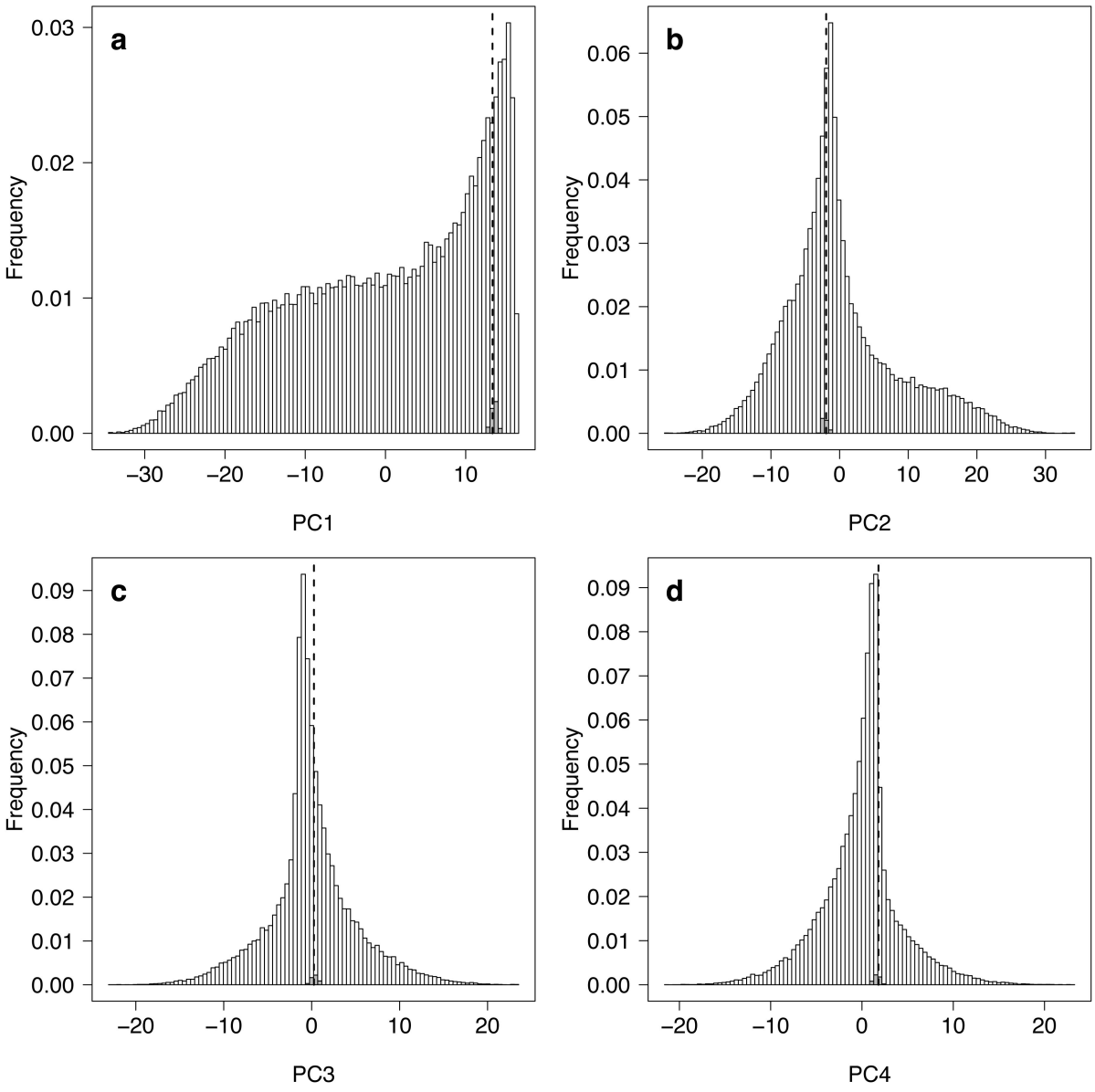

Fig. S6. Fit between empirical and simulated summary statistics for *Carya cordiformis*. Each panel (a-d) represents a different principal component (PC) axis of variation in the summary statistics (population pairwise directionality index Ψ and pairwise F_{ST}). White bars: distribution of PC values among all simulations; small solid grey bars: distribution of PC values among retained simulations; vertical dashed line: PC values of the empirical data.

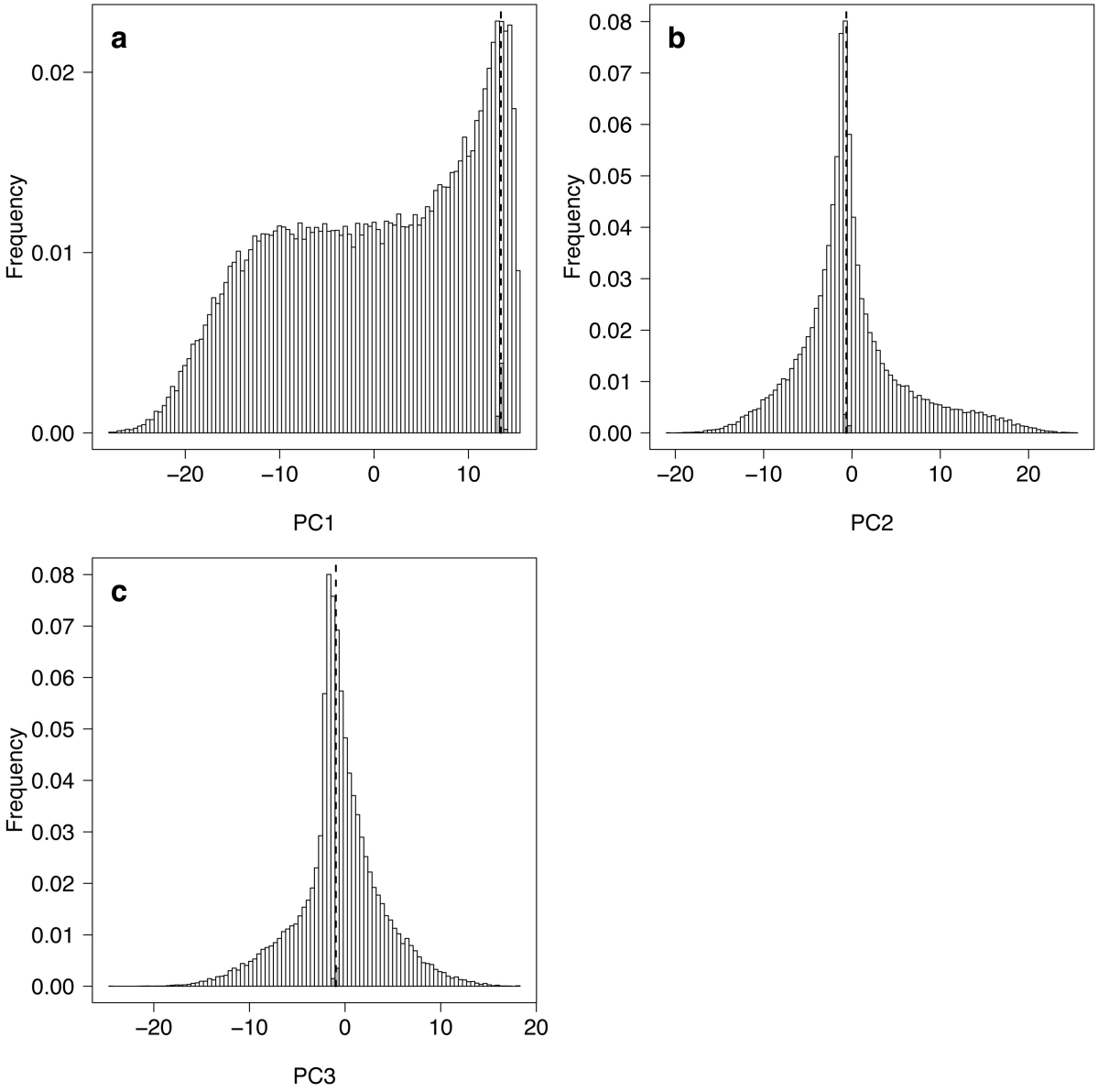

Fig. S7. Fit between empirical and simulated summary statistics for *Carya ovata*. Each panel (a-c) represents a different principal component (PC) axis of variation in the summary statistics (population pairwise directionality index Ψ and pairwise F_{ST}). White bars: distribution of PC values among all simulations; small solid grey bars: distribution of PC values among retained simulations; vertical dashed line: PC values of the empirical data.

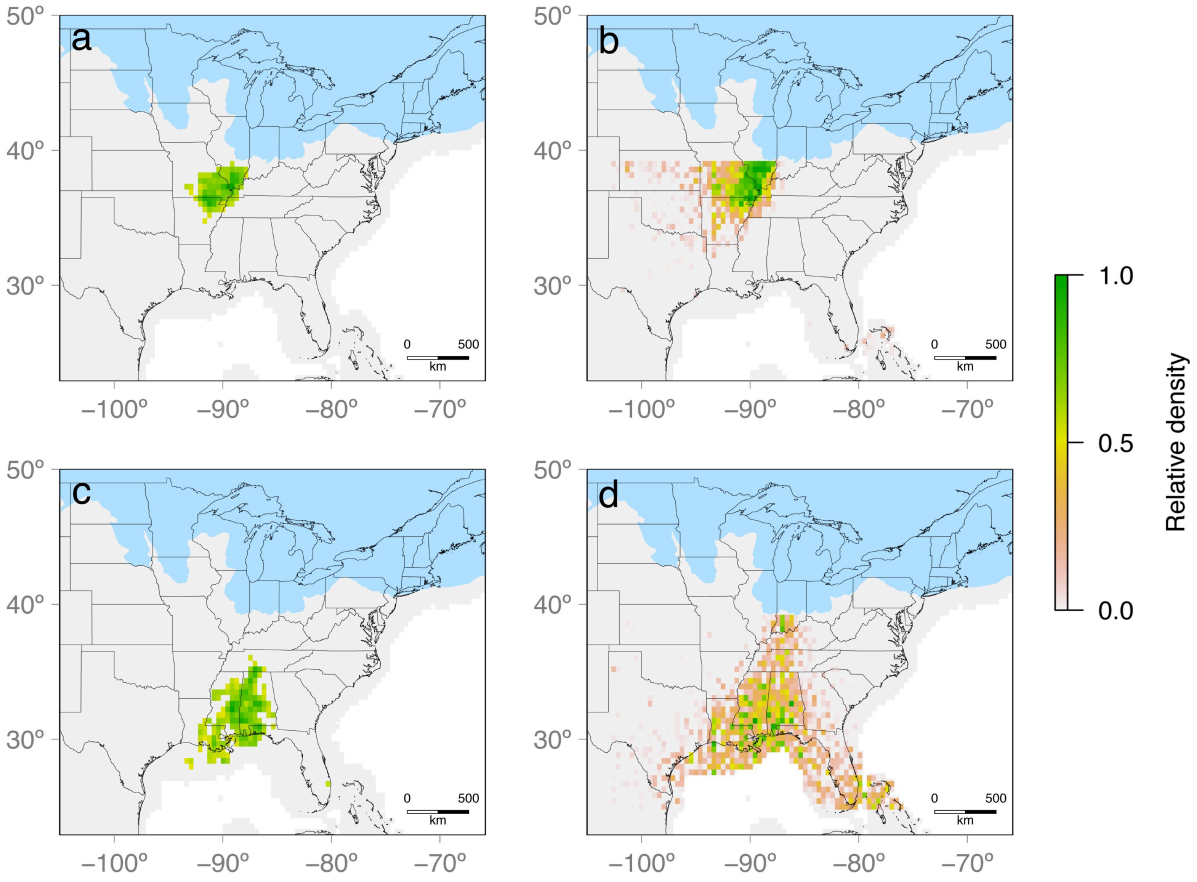

Fig. S8. True expansion origins of pseudo-observed datasets (PODs) whose inferred origins match those of the empirical data. Panels (a) and (c) show all areas with a high likelihood (≥0.5) of serving as the expansion origin in the empirical data for (a) *Carya cordiformis* and (c) *Carya ovata.* Corresponding panels (b) and (d) show the true expansion origins of all PODs that resulted in inferred expansion origins contained within the areas in panels (a) and (c). Densities in panels (b) and (d) were generated by first weighting each POD relative to the likelihood of its corresponding inferred expansion origin from panels (a) and (c), and then summing weighted values across all PODs.

Table S1. Empirical summary statistics for *Carya cordiformis*. Below the diagonal: population pairwise F_{ST}; above the diagonal: population pairwise Ψ. For geographic locations of populations, see Fig. S1.

	AR	GA	IA	IN	KYe	KYw	LA	MI	MO	MSn	MSs	NCe	NCw	NYe	SC	TN	TX	VAn	VAs	WI
AR	–	0.00	-0.01	0.00	-0.03	-0.02	-0.03	-0.02	-0.04	-0.01	-0.04	-0.04	-0.03	-0.10	0.02	-0.03	0.01	-0.06	-0.04	-0.02
GA	0.04	–	-0.02	0.00	-0.03	-0.03	-0.03	-0.02	-0.05	-0.02	-0.05	-0.04	-0.03	-0.11	0.02	-0.03	0.01	-0.05	-0.05	-0.02
IA	0.03	0.05	–	0.01	-0.02	-0.01	-0.02	0.00	-0.03	0.00	-0.03	-0.03	-0.02	-0.09	0.03	-0.02	0.03	-0.05	-0.03	-0.01
IN	0.02	0.04	0.03	–	-0.03	-0.02	-0.04	-0.02	-0.04	-0.02	-0.05	-0.04	-0.03	-0.10	0.02	-0.03	0.02	-0.06	-0.04	-0.02
KYe	0.03	0.06	0.04	0.03	–	0.01	0.00	0.01	-0.02	0.02	-0.02	-0.01	0.00	-0.07	0.05	0.00	0.04	-0.03	-0.02	0.00
KYw	0.02	0.02	0.02	0.02	0.03	–	-0.01	0.01	-0.02	0.00	-0.02	0.00	-0.01	-0.08	0.04	-0.01	0.04	-0.04	-0.02	0.01
LA	0.05	0.03	0.07	0.06	0.06	0.03	–	0.01	-0.01	0.01	-0.01	0.00	0.00	-0.08	0.05	0.00	0.05	-0.02	-0.01	0.02
MI	0.02	0.04	0.02	0.01	0.03	0.02	0.07	–	-0.03	0.00	-0.03	-0.01	-0.01	-0.08	0.04	-0.01	0.04	-0.04	-0.03	0.00
MO	0.02	0.04	0.04	0.02	0.04	0.02	0.05	0.03	–	0.03	0.01	0.01	0.01	-0.05	0.06	0.01	0.06	-0.02	0.00	0.03
MSn	0.03	0.04	0.05	0.03	0.07	0.03	0.04	0.05	0.05	–	-0.03	-0.02	-0.02	-0.09	0.03	-0.02	0.03	-0.05	-0.03	0.00
MSs	0.04	0.04	0.06	0.03	0.07	0.04	0.05	0.05	0.06	0.05	–	0.00	0.02	-0.06	0.06	0.00	0.06	-0.02	0.00	0.03
NCe	0.05	0.04	0.06	0.06	0.08	0.05	0.08	0.07	0.08	0.08	0.08	–	0.01	-0.06	0.06	0.00	0.06	-0.02	-0.01	0.01
NCw	0.04	0.05	0.05	0.04	0.06	0.04	0.08	0.03	0.05	0.07	0.09	0.08	–	-0.06	0.05	0.00	0.05	-0.03	-0.02	0.01
NYe	0.08	0.09	0.07	0.06	0.09	0.08	0.12	0.06	0.08	0.12	0.14	0.15	0.09	–	0.12	0.06	0.12	0.04	0.05	0.08
SC	0.02	0.02	0.04	0.02	0.05	0.01	0.04	0.03	0.03	0.03	0.03	0.05	0.06	0.11	–	-0.05	0.00	-0.08	-0.06	-0.04
TN	-0.04	-0.06	-0.03	-0.05	0.00	-0.03	0.00	-0.04	-0.02	0.01	0.01	0.05	-0.01	0.06	0.01	–	0.05	-0.03	-0.02	0.01
TX	0.02	0.01	0.04	0.03	0.05	0.02	0.03	0.04	0.04	0.04	0.03	0.07	0.06	0.11	0.03	0.01	–	-0.07	-0.06	-0.04
VAn	0.04	0.06	0.05	0.04	0.05	0.03	0.08	0.04	0.04	0.07	0.06	0.08	0.05	0.08	0.06	-0.03	0.07	–	0.01	0.04
VAs	0.05	0.07	0.05	0.04	0.04	0.05	0.09	0.04	0.05	0.07	0.08	0.08	0.06	0.09	0.06	-0.01	0.05	0.06	–	0.02
WI	0.02	0.05	0.01	0.02	0.04	0.02	0.07	0.01	0.03	0.05	0.07	0.08	0.05	0.06	0.04	0.00	0.06	0.04	0.04	–

Table S2. Empirical summary statistics for *Carya ovata*. Below the diagonal: population pairwise F_{ST}; above the diagonal: population pairwise Ψ. For geographic locations of populations, see Fig. S1.

	AR	CT	IA	IN	KYe	KYw	MI	MO	MSn	MSs	ON	SC	TN	VAn	VAs	WI	WV
AR	–	0.03	-0.01	0.02	-0.01	0.02	0.04	0.03	-0.02	0.03	-0.03	0.05	0.01	-0.03	0.02	-0.03	-0.01
CT	0.04	–	-0.03	0.00	-0.04	-0.01	0.00	0.00	-0.05	0.00	-0.06	0.01	-0.02	-0.05	-0.01	-0.05	-0.04
IA	0.02	0.03	–	0.04	0.00	0.02	0.04	0.03	-0.02	0.04	-0.02	0.05	0.02	-0.02	0.03	-0.02	-0.01
IN	0.06	0.02	0.05	–	-0.03	-0.01	0.01	0.00	-0.04	0.00	-0.05	0.01	-0.02	-0.04	0.00	-0.06	-0.04
KYe	0.05	0.03	0.04	0.02	–	0.03	0.04	0.04	-0.02	0.04	-0.02	0.05	0.02	-0.02	0.03	-0.01	-0.01
KYw	0.04	0.03	0.04	0.02	0.02	–	0.02	0.01	-0.03	0.02	-0.04	0.02	-0.01	-0.04	0.00	-0.04	-0.03
MI	0.02	0.01	0.01	0.02	0.03	0.01	–	0.00	-0.06	-0.01	-0.06	0.01	-0.03	-0.06	-0.01	-0.06	-0.05
MO	0.03	0.02	0.03	0.02	0.03	0.02	0.01	–	-0.05	0.00	-0.05	0.02	-0.02	-0.05	-0.01	-0.06	-0.04
MSn	0.05	0.04	0.07	0.04	0.03	0.04	0.04	0.05	–	0.06	-0.01	0.07	0.03	-0.01	0.04	-0.01	0.00
MSs	0.04	0.03	0.04	0.02	0.03	0.02	0.02	0.03	0.05	–	-0.06	0.01	-0.03	-0.06	-0.01	-0.06	-0.05
ON	0.00	-0.02	-0.01	0.02	0.03	0.01	0.00	0.01	0.04	0.03	–	0.07	0.04	0.00	0.05	0.00	0.01
SC	0.05	0.02	0.05	0.01	0.02	0.01	0.03	0.02	0.04	0.02	0.04	–	-0.03	-0.07	-0.02	-0.07	-0.06
TN	0.07	0.04	0.07	0.02	0.04	0.02	0.04	0.04	0.06	0.04	0.04	0.02	–	-0.04	0.01	-0.03	-0.03
VAn	0.07	0.03	0.07	0.03	0.05	0.03	0.04	0.03	0.06	0.05	0.04	0.02	0.04	–	0.04	0.00	0.01
VAs	0.05	0.02	0.04	0.01	0.02	0.02	0.03	0.02	0.04	0.01	0.04	0.02	0.02	0.03	–	-0.04	-0.03
WI	0.04	0.07	0.05	0.08	0.08	0.08	0.05	0.07	0.08	0.09	0.02	0.08	0.10	0.08	0.08	–	0.01
WV	0.05	0.02	0.04	0.02	0.02	0.02	0.03	0.02	0.04	0.03	0.06	0.03	0.03	0.03	0.02	0.08	–

Table S3. Summary of results for different PC axes retained. Results are shown for (a) *Carya cordiformis* and (b) *Carya ovata* when different numbers of principal component (PC) axes of spatial summary statistics are retained in estimating *Ω*. The cumulative % variation refers to the cumulative amount of variation explained by the retained PC axes. Also reported are the marginal density of the model, Wegmann's *p*-value (7), and the estimated expansion origin (*Ω*). The optimal number of PC axes to retain based on changes in Wegmann's *p* (assessed *post hoc*) is indicated with an arrow.

(a)

Retained PC axes	**Cumulative % variation explained**	**marginal density**	***p*-value**	***Ω***	
2	56.5	1.24×10^{-2}	1.00	28.4ºN, 95.4ºW	
3	62.8	3.92×10^{-3}	0.99	34.8ºN, 90.4ºW	
4	67.7	1.65×10^{-3}	0.86	37.3ºN, 89.1ºW	←
5	72.0	1.36×10^{-4}	0.11	36.0ºN, 91.2ºW	

(b)

Retained PC axes	**Cumulative % variation explained**	**marginal density**	***p*-value**	***Ω***	
2	56.3	3.05×10^{-2}	1.00	34.3ºN, 87.4ºW	
3	63.4	4.12×10^{-2}	1.00	32.2ºN, 89.1ºW	←
4	67.9	3.68×10^{-3}	0.20	28.8ºN, 90.4ºW	
5	72.2	6.69×10^{-3}	0.34	28.8ºN, 90.4ºW	

Additional Dataset S1 (separate file)

Species distribution model of *Carya cordiformis* for the current time period, provided in ESRI ASCII Raster file format. Habitat suitability values range from 0 (unsuitable) to 10 (maximum suitability). For details of model construction see (1).

Additional Dataset S2 (separate file)

Species distribution model of *Carya cordiformis* for the Last Glacial Maximum (21.5 ka), provided in ESRI ASCII Raster file format. Habitat suitability values range from 0 (unsuitable) to 10 (maximum suitability). All unglaciated regions were assigned a minimum habitat suitability value of 1 so that demographic simulations could be initiated from areas potentially containing climatic microrefugia. For details of model construction see (1).

Additional Dataset S3 (separate file)

Species distribution model of *Carya ovata* for the current time period, provided in ESRI ASCII Raster file format. Habitat suitability values range from 0 (unsuitable) to 10 (maximum suitability). For details of model construction see (1).

Additional Dataset S4 (separate file)

Species distribution model of *Carya ovata* for the Last Glacial Maximum (21.5 ka), provided in ESRI ASCII Raster file format. Habitat suitability values range from 0 (unsuitable) to 10 (maximum suitability). All unglaciated regions were assigned a minimum habitat suitability value of 1 so that demographic simulations could be initiated from areas potentially containing climatic microrefugia. For details of model construction see (1).

References

1. Bemmels JB, Dick CW (2018) Genomic evidence of a widespread southern distribution during the Last Glacial Maximum for two eastern North American hickory species. *J Biogeogr* 45(1739–1750).
2. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7(5):e37135.
3. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3 Genes, Genomes, Genet* 1(3):171–182.
4. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* 22(11):3124–3140.
5. De la Cruz O, Raska P (2014) Population structure at different minor allele frequency levels. *BMC Proc* 8(Suppl 1):S55.
6. Little EL (1971) *Atlas of United States trees, volume 1. Conifers and important hardwoods. USDA Miscellaneous Publication 1146.* (Washington, DC).
7. Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics* 11:116.
8. Leuenberger C, Wegmann D (2010) Bayesian computation and model selection without likelihoods. *Genetics* 184(1):243–252.