Inferring genetic structure of European beech from observations of spectral phenotypes.

Ewa A. Czyż[1*], Bernhard Schmid[1], Maarten B. Eppinga[1], Marylaure de La Harpe[1,a], Aboubakr Moradi[1,2,b], Cheng Li[1], Michael E. Schaepman[1], Meredith C. Schuman[1,2]

[1] Department of Geography, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

[2] Department of Chemistry, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

[a] Current address: Amt für Natur und Umwelt, Graubünden, Ringstrasse 10, 7000 Chur, Switzerland

[b] Current address: Syngenta Crop Protection AG, Schaffhauserstrasse, 4334 Münchwilen, Switzerland

*To whom correspondence should be addressed; E-mail: ewa.czyz@uzh.ch. ORCID: 0000-0002-4505-0701.

Summary

Biodiversity loss presents a growing threat to the global environment and requires systematic and spatially contiguous monitoring. Monitoring of within-species genetic variation, a key factor when assessing biodiversity loss, is laborious and could be complemented by remote observations of phenotypes allowing inferences about genetic variation. We studied genetic and phenotypic variations in the common European beech (*Fagus sylvatica* L.) derived from whole-genome sequences and spectral phenotypes of more than 200 individuals at 22 sites across the species' natural range. The spectral phenotypes were collected under standardized illumination/observation conditions from the same top-of-canopy leaves used for nuclear DNA extraction. We found that information from spectral phenotypes positively contributes to predictions of genetic structure using environmental variables across the sites in our dataset. We therefore propose that future spaceborne satellites with sufficient spectral and spatial resolution, repeatedly acquiring spectral phenotypes globally across species ranges, improve our ability to rapid and spatially contiguous monitoring of genetic diversity in trees.

Keywords

biodiversity monitoring, *Fagus sylvatica*, intraspecific genetic variation, remote sensing, spectral phenotypes,

Introduction

Biodiversity loss presents a large and growing risk for the global environment and its ability to support life. Plant diversity loss can negatively affect ecosystem functioning (Purvis et al. 2000, Hector et al. 2007, Isbell et al. 2011, Cardinale et al. 2012, Reich et al. 2012,), ecosystem stability (Tilman et al. 1994, Hector et al. 2010, Isbell et al. 2015, van Moossel et al. 2021), and reduce the potential of populations and communities to adapt to environmental change (Des Roches et al. 2018). Biodiversity is under pressure from human activity (Johnson et al. 2009). To mitigate human impacts on biodiversity decline, targets have been set several times aiming to mitigate biodiversity loss worldwide (IPBES 2019) but many further steps need to be taken (Mora et al. 2006, Jenkins and Joppa 2009,). One important step towards biodiversity support is its continuous and standardized monitoring (Scholes et al. 2008, Turner 2014). Monitoring indicators of biodiversity would help to set conservation priorities (Scott et al. 1989), evaluate actions taken, and develop adaptive management (Sutherland et al. 2004) and policy adjustments (Perrings et al. 2011) to preserve and restore biodiversity. Monitoring schemes should span levels of biological organization from ecosystems to species and genes. Genetic diversity is especially challenging to monitor, but can provide valuable information about a species' contribution to ecosystem functioning (Fridley and Grim 2010, Tang et al. 2022) and its evolutionary potential, and thus is a key component when accounting for biodiversity loss (Wilson 1988, Exposit-Alonso et al. 2022).

The assessment of genetic diversity is generally based on field sample collection, with limited temporal and spatial coverage. Furthermore, genetic variation acquisitions are still commonly limited to a portion of whole genomes (Lou et al. 2021). These scale-induced limitations result in sparse and biased representations of genetic diversity, which lead to inaccurate representation of its current status, omission of unique but difficult-to-sample sites, and partially misses components of the variation that reflects total adaptation potential of species to environmental change (Hughes et al. 2008, Des Roches et al. 2018). Whole-genome sequencing provides a means to assess intraspecific genetic diversity without the bias of more restricted sets of molecular markers. However, the spatiotemporal limitations still apply. Phenotypic variation acquired over space and time may offer a useful indicator of a species' genetic diversity if a link between the two can be established across different test sites and environmental conditions.

To collect these phenotypic data, imaging spectrometers can provide repeated and spatially contiguous acquisitions of spectral phenotypes comprising integrated signatures of plant biochemical and structural traits (Guyot et al. 1992, Jacquemont and Ustin 2009, Schaepman et al. 2009, Schaepman et al. 2015). These plants traits result from the interaction of plant genes with the environment, and might affect the spectral

phenotypes of individuals from leaf to canopy levels (Meireles et al. 2020). Leaf spectral phenotypes are commonly estimated from spectroradiometer measurements under standardized illumination/observations conditions, while the spectral phenotypes of entire canopies at landscape-scale are derived from remote acquisitions of reflected radiation by airborne and spaceborne imaging spectrometers (Schaepman et al. 2009). Spectral phenotypes can indicate genetic groups and variants, and thus spectral acquisitions might facilitate monitoring of genetic diversity under natural conditions, if a link between spectral and genetic variation can be established (Cavender-Bares et al. 2016, Yamasaki et al. 2017). Until now, however, this potential remains uncertain as previous studies associating genetic and spectroscopy data collected under natural conditions have focused on relatively small numbers of sites covering a small proportion of the studied species' range, and did not integrate whole-genome information (Madricht et al. 2014, Czyż et al. 2020, Czyż et al. 2023).

Here, we demonstrate the potential of spectral phenotypes to contribute to assessing genetic structure of a plant species across its natural range. We use *Fagus sylvatica* L. (European beech) as our test species because it is a dominant, ecologically and economically important tree species in Europe that has gained interest as a model for ecological genomics of tree species under global change (Capblancq et al. 2020). We present the largest genomic sampling to date of the *F. sylvatica* species range using whole-genome data to derive genetic structure. We then use leaf reflectance data derived from top-of-canopy leaves with a contact probe, in combination with environmental variables, to predict the genetic structure. We thereby (1) contribute to current knowledge about genomic structure in *F. sylvatica*; and (2) indicate how spectral phenotypes acquired in natural environments may help to predict it. The results suggest that aerial and satellite-based imaging spectroscopy could be leveraged for genetic variation monitoring and conservation over larger areas and at higher temporal resolution than typically attainable when solely relying on field sampling.

Materials and Methods

Study areas and sample collection

We selected 22 sites dominated by *Fagus sylvatica* L. (common beech) across its natural range in Europe (Fig.**1a**). We abbreviate the site names to the first two letters referring to the country that the site is located in, and the third letter referring to the particular location in the country. Most of the sites are covered by protected, old-growth beech forests. The variation among the sites accounts for climatic (mean 50% of the range and 95% of the distribution of environmental values across the range, Supporting information Fig.

91 S1), edaphic and phylogeographic variation (Caudullo et al. 2017). Data were collected *in situ* during August–
92 September 2020. At each site we sampled 10 *F. sylvatica* trees that occupied canopy space along two ca.
93 10 x 200 m transects separated by 50 m. Each transect comprised 5 dominant trees separated by at least
94 30 m. At some sites we sampled fewer than 10 trees due to a lack of *F. sylvatica* trees in the canopy within
95 the transect (4 trees at RSB, 8 trees at ESP) and in some cases the sampled material was of poor quality for
96 the DNA extraction (1 tree at PLM, 1 tree at ITP). When possible, we selected transects on south-west to
97 south-east-facing slopes to normalize for local adaptation to exposition to solar radiation. From each tree,
98 we collected a top-of-canopy branch of 20 to 100 cm length with the help of tree climbers or, when tree
99 height did not exceed 15 m, with a telescoping carbon fiber pruner (Takeni Trading Co., Osaka, Japan). From
100 each branch, we randomly selected three undamaged leaves separated by at least 5 cm, from which we
101 acquired spectral data after harvesting (see Materials and Methods, Spectral data) and which we
102 subsequently stored on silica gel for later genetic analysis (see Materials and Methods, Genetic data).

103 ## Genetic data

104 The collected leaves were taken back to the laboratory where we took 50 mg of undamaged leaf lamina and
105 then stored these samples in tubes at –20 °C until grinding and DNA extraction. The DNA was purified using
106 a Norgen Plant and Fungi genomic DNA extraction kit (Norgen Biotek, Thorold, ON, Canada), modified to
107 reduce RNA contamination and to improve lysis, yield, and purity. Specifically, samples were flash-frozen in
108 liquid nitrogen immediately prior to grinding and ground using 3 mm glass beads at 30 Hz for 3 min in a
109 TissueLyser II (Qiagen) to a fine powder. The powder was briefly centrifuged before adding lysis buffer and
110 vortexed to mix with the buffer. One µl of RNAse A (DNAse-free, 100,000 units/ml in 50% glycerol, 10 mM
111 Tris-HCl, pH 8.0) and 3 µl Proteinase K (>600 µl ~ 20 mg/ml in 10 mM Tris-HCl pH 7.5, containing calcium
112 acetate and 50% v/v glycerol) were added to the lysis buffer prior to 10 min incubation at 56°C with 1000
113 rpm agitation. An additional 1 µl of RNAse A was subsequently added to the reaction, followed by a 10 min
114 incubation at 65°C with 2000 rpm agitation. The samples were then processed according to the
115 manufacturer's protocol with the modification that 70% ethanol was mixed with lysate by pipetting ca.
116 10x/sample instead of vortexing. The DNA from clarified lysate was captured on spin columns
117 (centrifugation until all lysate passed) and columns were washed three times with Solution A prior to elution
118 with Elution Buffer B (100 µl, 5 min RT incubation, spinning until all elution buffer was recovered).

119 The purified DNA was used for sequencing library preparation (Novogene, Cambridge, UK): genomic DNA
120 was randomly sheared into short fragments, which were end repaired, A-tailed, and ligated with Illumina
121 adapters. The fragments with adapters were PCR amplified, size selected, and purified. Libraries were

122 sequenced by Illumina NovaSeq PE150 (Novogene, Cambridge, UK) to produce ca. 11 Gbp/sample, resulting
123 in a mean ca. 20x coverage of the 541 Mbp nuclear genome (Mishra et al. 2022). The Illumina adapters
124 were trimmed from the raw reads, and reads with phred scores below 3 and windows of 4 base pair sliding
125 window with average phred scores below 15 were also removed using *trimmomatic* v. 0.40. We aligned
126 reads against the chromosome-level reference genome (Mishra et al. 2022) in *bowtie2* v. 2.4.5 and removed
127 alignment duplicates and unaligned reads in *picard* v. 2.0.1. Sequentially, we derived sets of alleles per each
128 individual tree by calling variants against the chromosome-level reference genome (Mishra et al. 2022)
129 using the function *HaplotypeCaller* in *GATK4*. We consolidated the derived alleles and obtained the joint
130 SNPs among all individuals using the function *GenotypeGVCFs* in *GATK4*. From the derived SNPs, we selected
131 those with phred scores above 30 (i.e., accuracy higher than 99.9%), minor allele frequencies of more than
132 0.1, and minimum and maximum allele depths of 10 and 50, respectively. We pruned the derived SNPs for
133 linkage disequilibrium by removing SNPs that had a correlation coefficient of greater than 0.1 with each
134 other in *plink* v. 1.9 software. This resulted in 1,380,310 SNPs across all sampled individuals used in further
135 analyses. Patterns of *F. sylvatica* genetic diversity and structure have been studied with marker-based
136 approaches including allozymes (Gallois et al. 1998), AFLP (Pluess and Weber 2012, Guevara et al. 2022),
137 microsatellites and, more recently, SNPs in candidate genes (Csilléry et al. 2014, Pfenninger et al. 2020,
138 Meger et al. 2021, Cuervo-Alarcon et al. 2021) and whole-genome sequences from limited area of the specie
139 range (Stefanini et al. 2023). However, no study so far used whole-genome sequences from across the range
140 of *F. sylvatica*.

141 ## Spectral data

142 Spectral data comprised leaf reflectance spectra collected in the wavelength range from 350–2500 nm using
143 an ASD FieldSpec 4 spectroradiometer (serial n° 18739, ASD Inc., Boulder, CO, USA), coupled to a plant
144 probe with leaf clip attachment, and equipped with a standardized halogen light source and a set of
145 standard black and white reference backgrounds (serial n°445, ASD Inc., Boulder, CO, USA). Measurements
146 of three leaves harvested from 20 cm – 2 m long branches (1 branch/tree) were acquired within minutes to
147 up to 3 h (for ITE, ITM, 5 trees at ITV, and 5 trees at PLB) from the time of branch harvest using the contact
148 probe positioned in the middle of the leaf and to the right of the midrib as viewed from the adaxial side (i.e.
149 the side of the leaf facing the incident radiation (e.g. sun and light source). Measurements comprised five
150 successive readings per leaf for four types of acquisitions: a white reference R_w sample, a leaf with white
151 reference as background T_w, a black reference R_b sample, and the leaf with black reference as background
152 T_b (cf., eq. 1). For each reading, five scans, each with an integration time of 8.5 ms, were recorded and

153 averaged per leaf. The averaged reflectance spectra per tree was calculated from the mean of leaf
154 reflectance R using a formula for background-differentiated measurements (Miller et al. 1987):

155

156 $$R = \frac{T_b * R_w - T_w * R_b}{R_w - R_b} * 100\ [\%] \qquad \text{[eq. 1]}$$

157

158 The ASD spectroradiometer provides convolved sampling intervals of 1 nm spectral sampling, although the
159 spectral resolution of the instrument ranges from 3–10 nm sampling interval. To minimize collinearity
160 effects of neighboring spectral bands, we resampled in postprocessing the spectral signal from the original
161 2151 spectral bands to 216 spectral bands by linearly interpolating the original signal using *interp1* function
162 in *Matlab* (v. R2017b) followed by a multiplication with a Spectral Response Function (SRF) of Gaussian
163 shape with Full Width at Half Maximum (FWHM) of 10 nm.

164 ## Climatic data

165 Climatic data comprised 19 bioclimatic variables derived from *WorldClim Version2* (Fick et al. 2017) at a
166 resolution of ~1 km with global coverage. We extracted the value of each environmental variable per site
167 based on the geographic locations of sampled sites and applied min-max normalization of extracted values
168 for each variable.

169 ## Statistical analysis

170 First, we used the filtered SNPs (see Materials and Methods, Genetic data) to estimate genetic structure
171 and relatedness of individuals. Genetic structure was calculated in *ADMIXTURE* v. 1.3 (Alexander et al. 2015)
172 and the number of genetic clusters was set to 4 based on the smallest cross-validation error (Supporting
173 information Fig. **S2**). To derive the relationships among trees, we calculated the SNPs-based pairwise
174 distance between individuals using *plink* v. 1.9, and applied the *cmdscale* function of *Matlab* (v. R2017b) to
175 perform Principal coordinate analysis (PCoA) and to derive the genetic Principal Coordinates (PCOs). Based
176 on pairwise distances between individuals, we additionally calculated within-site standard deviations of
177 genetic relatedness.

178 Furthermore, we estimated independent and common contributions of climatic and spectral information
179 to model genetic structure among sampled *F. sylvatica* sites. We used redundancy analysis (RDA) as a
180 multivariate ordination method to simultaneously account for multiple response variables (Capblancq and

181 Forester 2021). We used Akaike Information Criterion (AIC) values for forward variable selection and a
182 threshold of 2.5 for Variance Inflation Factor (VIF) scores to reduce the impact of collinearity of variables
183 (Johnston et al. 2018). We limited the response variables to *F. sylvatica* distribution along the first and the
184 second genetic PCO, which explained together 7.7% of total variation in derived genetic structure
185 (Supporting information Fig. S3). Further genetic PCOs (3-6) did not correlate strongly with environmental
186 and spectral variation (Supporting information Fig. S4).

187 We constructed RDA models from environmental and spectral information separately, which resulted in the
188 construction of environment-based (including 4 out of 19 explanatory environmental variables) and
189 spectral-based (including 5 out of 216 explanatory spectral variables) models. For the full model, we
190 considered all 9 variables selected for the environment- and spectral-based models to construct RDA
191 models including both environmental and spectral information as explanatory variables. Following the
192 variable selection procedures described above yielded a full model that include three environmental and
193 five spectral explanatory variables. We derived the independent contributions of each environmental and
194 spectral variable to the full model and we partitioned the genetic variance (as variance along genetic PCO1
195 and PCO2) to within- and between-site variation using analysis of variance (ANOVA) with explanatory terms
196 for environment and spectral information and for site identity. Additionally, we conducted an analysis based
197 on the five first spectral Principal Components (PCs) derived from PCA on the resampled spectral data to
198 compare how well these corresponded to variation along genetic PCOs, and found the explanatory power
199 to be lower than for our main analysis using RDA (Supporting information Tab. S1). We used the *rda* function
200 of the *vegan* package (v. 2.6-4) in *R* (v. 4.1.3) to construct the models and to find the model fits, and *pca*
201 function of *Matlab* (v. R2017b) to derive the spectral PCs.

202 To determine the contribution of spectral information to genetic structure predictions, we conducted leave-
203 one-out validation. Accordingly, we created a model of 21 populations to predict the distribution of
204 individuals of the 22nd population along genetic PCOs from (1) environmental information only, and (2)
205 environmental information merged with spectral information. We present the contribution of spectral
206 information to genetic structure prediction as the percentage change in predicted versus actual distribution
207 along genetic PCO1 and PCO2 from the model based on (1) environmental and (2) environmental with
208 spectral information, respectively.

209
Results

210
Derived genetic structure of *F. sylvatica* across the species natural range

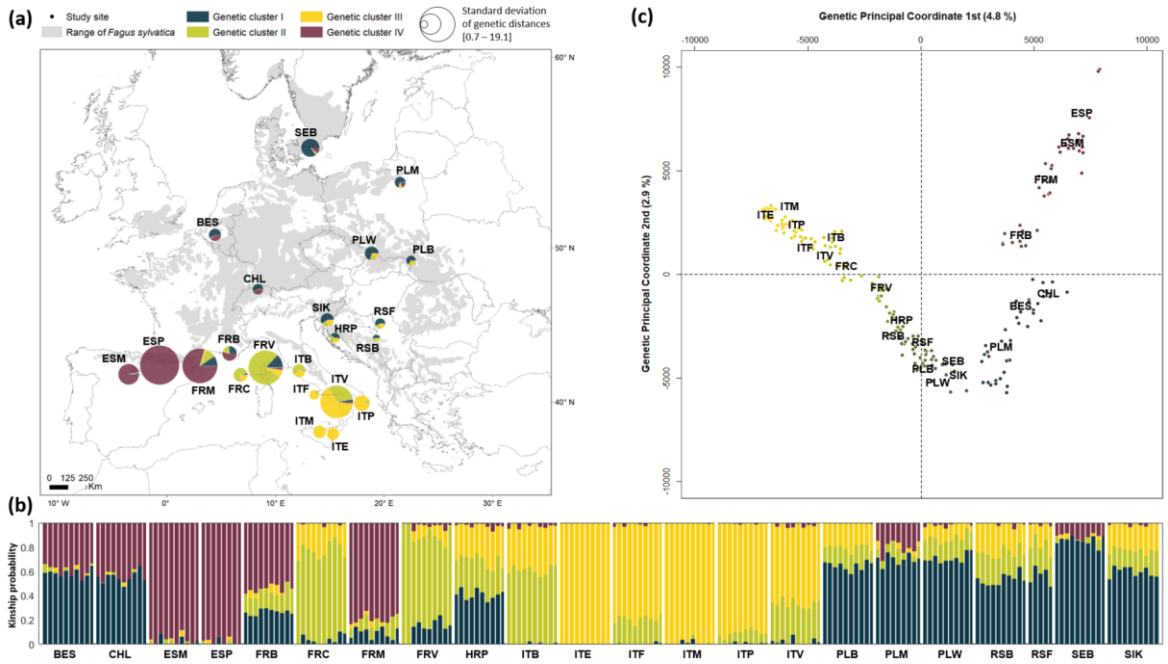

Fig. 1. Genetic structure among 22 F. sylvatica *sites presented as kinship probabilities of belonging to K=4 genetic clusters with the smallest cross-validation error in* ADMIXTURE *inferred from 1,380,310 SNPs across the nuclear genome (see Materials and Methods, Genetic data). (**a**) Averaged kinship probability and the standard deviation of genetic variation among the individuals from each sampled site (N=10 except for RSB, N=4, ESP, N=8, PLM and ITP, N=9). (**b**) The kinship probability among 209 sampled* F. sylvatica *individuals. (**c**) Relationship between* F. sylvatica *individuals plotted along the first and second Principal Coordinates (PCOs) derived from genetic distances among individuals inferred from 1,380,310 SNPs across the nuclear genome (see Materials and Methods, Genetic data). The horseshoe-like distribution of the individuals along the first and second PCOs does not affect the conclusions of the presented analysis (Podani and Miklós 2002). Individuals are colored by combined kinship probabilities to four genetic clusters.*

We present, to our knowledge, the first assessment of genetic structure of a broad-leaf tree species across its natural range derived from whole-genome sequences of nuclear DNA. Using *ADMIXTURE* cross-validation, we identified four genetic clusters representing the current *Fagus sylvatica* L. genetic structure (Fig. **1**). We recognized the high abundance of one genetic cluster in southern Italy (ITE, ITM, ITP, ITV; cluster indicated in yellow), one genetic cluster in the southern Pyrenees (ESP, ESM, FRM; cluster indicated in maroon), and one genetic cluster centered in Corsica (FRV, FRC; cluster indicated in green). The green genetic cluster intermixes with maroon genetic cluster along French Mediterranean basin (FRB, FRM), with yellow genetic cluster along the Appenninan Peninsula (ITB, ITV, ITF, ITP), and with blue genetic cluster along Balkans (HRP, RSB, RSF, SIK) and Carpathian mountain range (PLB, PLW). The blue (4th of the recognized genetic clusters) dominates our sampling across Europe and is the most widespread in central and northern part of the continent (Fig. **1a**).

The sampled sites in central and norther Europe are also characterized by relatively small standard deviation of genetic distances within sites (Fig. 1b). In contrast, per-individual kinship probabilities to different genetic clusters, and pairwise genetic dissimilarities based on allele frequencies between individuals, demonstrate relatively higher within and between-population variation of southern than northern samples (Fig. 1b,c, Supporting information Fig. S5).

Observed genetic structure explained by environmental and spectral information

Together, environmental and spectral information explained 77.7% of the variance in the distribution of sampled *F. sylvatica* along the first and second genetic PCOs (Fig. 2). Environmental variables independently of spectral information explained 39.8% of the variation. The environmental variables contributing most were Mean Temperature of Driest Quarter (BIO 09), Isothermality (BIO 03), Precipitation Seasonality (BIO 15) and Precipitation of Driest Quarter (BIO 17). These variables define environmental niches which may determine adaptive patterns in European beech (Fang and Lechowicz 2006, Bolte et al. 2007). In contrast, variables from leaf spectra explained 12.2% of variation in genetic structure among trees (genetic PCO1 and PCO2) independently of environmental information. The spectral variation explained 12.1% of genetic variation between sites and only 0.1% of genetic variation within sites (Supporting information Fig. S6). Portions of spectra which were most informative about the genetic structure correspond to: pigments absorption, related to electronic transitions induced by light at 410 nm and 530 nm (Gitelson et al. 2006), leaf surface morphological properties affecting the reflection of UV light at 350 nm, leaf inner structure related to scattering of electromagnetic radiation at 790 nm (Gausman et al. 1971), and leaf water content related to vibrational absorption of photons at 790 nm and 1500 nm. Environmental and spectral variation combined contributed an additional 25.6% to predicting genetic variation along genetic PCO1 and PCO2.

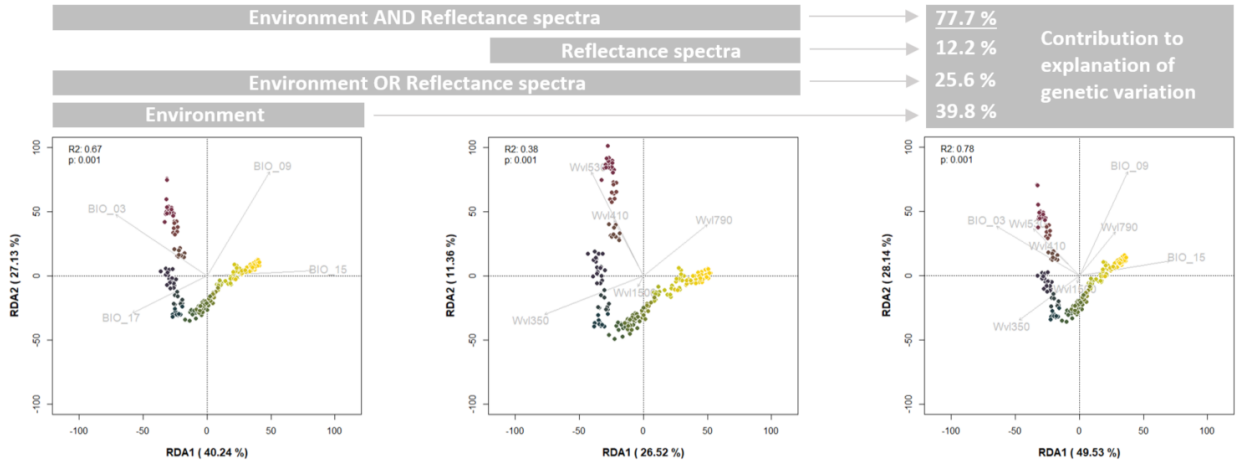

Fig. 2. RDA models for prediction of F. sylvatica *genetic variation as captured by the first (PCO1) and second (PCO2) genetic Principal Coordinate across the species range derived from environmental information, spectral information, and environmental with spectral information combined. The variables were chosen based on AIC and VIF <2.5. Out of 19 environmental variables, four were identified to be indicative of the genetic structure between sites, and out of 216 spectral bands, five were indicative of the genetic structure between and within sites. The full model combined three out of four selected environmental variables and five out of the five spectral variables selected for the spectral-based model. The incorporation of spectral information added 12.2% explanatory power to the model created based on environmental (with contribution of 39.8%) and spectral information, which together explained 77.7% of the variation in genetic structure among sampled* F. sylvatica *individuals (mean of R2 for PCO1 and PCO2). The* F. sylvatica *individuals in RDA space are colored by combined kinship probabilities to four genetic clusters.*

Spectral information contributed most to prediction of the first genetic PCO (Fig. **3a**). The introduction of predictors derived from leaf spectral phenotypes improved the genetic structure prediction by 59.7% along genetic PCO1 in comparison to the prediction based on environmental variables only (R2 of 0.79 versus 0.63) (Supporting information Fig. **S7**). The spectral indicators contributing most to prediction of the distribution of individuals along this genetic PCO1 relate to leaf interaction with spectral bands centered at 530 nm and 410 nm in the visible region, influenced by plant pigments. The contribution of spectral information to prediction of the second genetic PCO produced a mean 7.9% improvement in comparison to the prediction based on environmental variables only (R2 of 0.76 versus 0.74) (Fig. **3b**, Supporting information Fig. **S7**). The smaller contribution of spectral information to prediction of the second genetic PCO might correspond to the shared explanatory power of wavelengths related to leaf structural properties (i.e spectral bands centered at 350 nm and 790 nm), which contribute most to the second genetic PCO axis of the full RDA model (Fig. **2**).

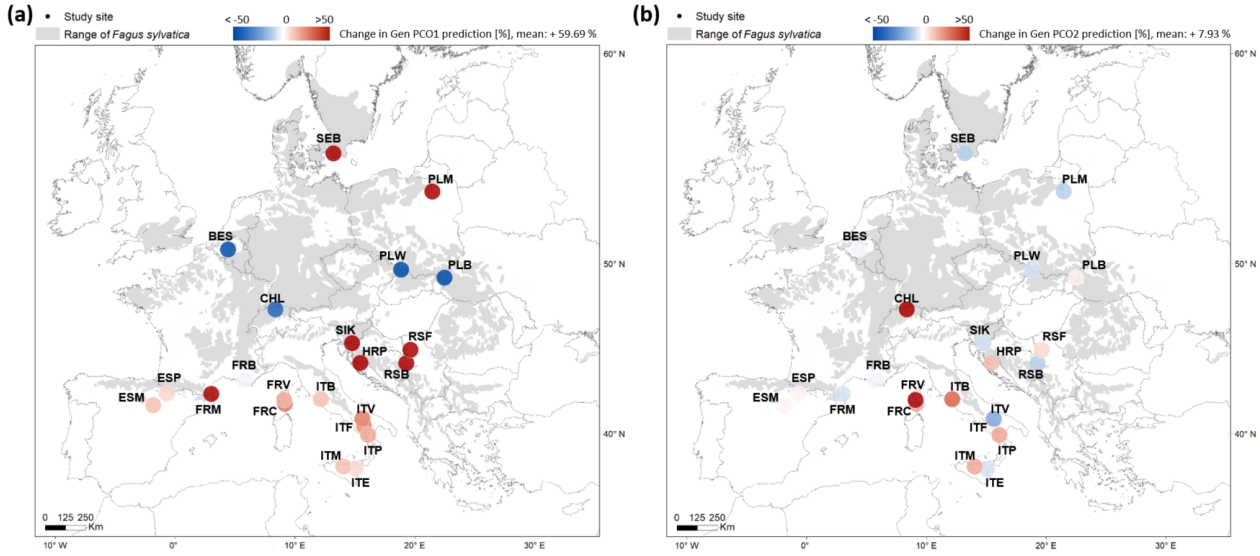

Fig. 3. Per site contribution of spectral information to predict F. sylvatica *genetic variation as a change (%) in predicted versus actual distribution along (**a**) first genetic Principal Coordinate (mean change of + 59.7 %)*

281 *and (**b**) second genetic Principal Coordinate (mean change of + 7.9 %). The change compare the prediction*
282 *accuracy of RDA models based on (1) environmental information only and (2) environmental with spectral*
283 *information combined. The improvement (in red) or decrease (in blue) in the genetic variation prediction*
284 *represents the mean per site change derived from leave-one-out cross validation of a model based on 21*
285 *sites to predict the genetic distribution of individuals of the 22nd site. The within-site variation in prediction*
286 *accuracy is presented in Supporting information Fig. **S8**.*

287 # DISCUSSION

288 We present contributions of spectral phenotype data in combination with environmental data to improve
289 the prediction of intraspecific genetic structure derived from whole-genome sequences of a forest tree
290 species across its natural range. The spectral predictions of spatio-temporal patterns in genetic variation
291 may be further improved by incorporation of the species phylogeographic history, and may be further
292 guided by remotely acquired spectral information at landscape level.

293 ## Species phylogeographic history

294 While spectra carry information about current genetic composition, further incorporation of information
295 about the phylogeographic history of the species would constrain model predictions to the available gene
296 pool of particular areas. The genetic clusters indicated in yellow and maroon may correspond to previously
297 reported glacial refugia (Magri et al. 2006), while further genetic and phylogeographic studies would be
298 valuable in Corsica to determine whether the green genetic cluster may represent intermixing between
299 refugia or an additional refugium (Petit et al. 2003). The origin of the most widespread genetic cluster
300 (indicated in blue) in central and northern Europe remains uncertain and might be assigned to the eastern
301 range of the Alps (Gomory et al. 2003) or southern Moravia (Magri et al. 2006). The phylogenetic structure
302 across the range of *F. sylvatica* is thought to result from post-glacial expansion: from the west across the
303 central European plains (maroon cluster), from the south across Carpathian mountain range (yellow
304 cluster), and from Corsica and nearby sites at the French Mediterranean basin (FRB, FRM) across the
305 Appenninan Peninsula (ITB, ITV, ITF, ITP), Balkans (HRP, RSB, RSF, SIK) and Carpathian mountain range (PLB,
306 PLW), likely permitted by terrestrial connections between the Adriatic and Mediterranean coasts during the
307 last glaciation (green cluster) (Hazler et al. 1997). In contrast to lower mountain ranges, the Alps likely act
308 as a barrier for gene flow between southern and northern populations (Paule 1995). Additionally, the
309 continuous distribution of the species in the northern part of its range likely allows greater gene flow across
310 the landscape, while the patchier distribution in the south limits gene flow and maintains genetic distance
311 between populations. This may results in higher genetic differences between Southern sites, which may
312 also be enhanced by post-glacial climatic changes causing migration to higher altitudes in the southern part

313 of the species range and environmental pressures that drive the diversification of *F. sylvatica* populations
314 at their southern limit (Thiébaut et al. 1982). Merging the current genetic structure and the phylogeographic
315 history of the species with regularly monitored spectral (phenotypic) variation will allow more accurate
316 prediction of genetic structure within tree species across their range.

317 Potential of remote observations to monitor genetic variation

318 Our spectral-genetic associations were spatially and temporally limited to single *in situ* sampling campaigns
319 per site, whereas spectral phenotypes at canopy level acquired from remote observations of airborne and
320 spaceborne imaging spectrometers will allow for repeated acquisitions that capture dynamic phenotypic
321 responses on the landscape scale. Data from existing spaceborne imaging spectrometers, and the wealth
322 of planned and upcoming imaging spectrometers missions (including NASA's SBG (Cawse-Nicholson et al.
323 2021) and ESA's CHIME (Celesti et al. 2022)), will bring spatially contiguous and temporally dense
324 information about spectral properties of the Earth surface resolved at spectral intervals which are
325 sufficiently narrow to be interpreted in terms of molecular absorption features. Spectrally resolved
326 information acquired repeatedly over landscapes may further help to calibrate predictive models of
327 intraspecific genetic structure across species ranges and predict genetic composition from unsampled
328 areas. Such spectral-based modelling might be further applied across different phylogenetic groups
329 (Meireles et al. 2020) and might facilitate spatial mapping of gene pools, beyond existing trait-based
330 approaches.

331 Here we have shown the contribution of spectral phenotypes to predict genetic structure without
332 discriminating the adaptive component. Common garden experiments from seeds collected at the sites may
333 help to separate genetic and environmental influences on spectral variation and support its association with
334 genes of known function. Such associations could facilitate research on selective processes that underlay
335 genetic adaptations across wide spatial and temporal scales.

336 Spectrally complete predictions of genetic structure and gene frequency distributions across space and time
337 will help to target environmentally, phenotypically and genetically underrepresented sites or hotspots for
338 sampling of genetic resources, whose sampling will further contribute to calibration of predictive models.
339 Improved predictions of genetic variation using spectral phenotypes could indicate changes in genetic
340 diversity and adaptation over time when combined with repeat monitoring, and could deliver information
341 for the adjustment of current efforts to safeguard biodiversity at the global scale, at unprecedented speed
342 and spatial completeness.

Acknowledgments

This study is supported the University of Zurich Research Priority Program on Global Change and Biodiversity (URPP-GCB), and funded by the NOMIS grant on "Remotely Sensing Ecological Genomics". We express our gratitude to Emmanuelle Argenti, Ventura Gennaro, Hugot Laetitia, Elodie Magnanou, Joseph Garrigue, Kris Vandekerkhove, Arne Verstraeten, Nilsson Jorgen, Piotr Małż, Katarzyna Gierad, Dejan Stojanović and the Institute of Lowland Forestry and Environment; University of Novi Sad, Gordan Lukač, Katja Konečnik, Janez Šubic, Damen Kocjan, Eustaquio Gil Pelegrín, Juan Pedro Ferrio, Piotr Poloczek and Stanisław Kucharzyk for allowing access to their sampling sites and support on investigation and to Romeo Galiano, Joseph Garrigue, Zbigniew Stebel, Helgessons TrädTjänst AB company and Michał Szydłowski for assistance in sample collection. We express our gratitude to Reinhard Furrer, Anna K. Schweiger, Benjamin Dauphin, Katalin Csilléry and Ingmar Staude for helpful discussions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions

Conceptualization: EAC (lead), MES (lead), BS, MBE, MH, MCS. Methodology: EAC (lead), MES, MCS, BS, MBE, AM, MH. Software: EAC (lead), BS, CL. Validation: EAC, BS, MBE. Formal analysis: EAC (lead), BS, MBE. Investigation: EAC (lead), BS, MBE, MH, CL, MCS, MES. Resources: MES (lead), MCS. Data Curation: EAC. Writing - Original Draft: EAC. Writing - Review and Editing: EAC (lead), BS, MH, MCS, CL, MES. Visualization: EAC. Supervision: MES, MCS, BS. Project administration: MES, MCS. Funding acquisition: MES.

Data availability

The raw short-read whole genome sequences genomic data that support the findings in this study are deposited at NCBI Sequence Read Archive (SRA) under the project with submission number SUB13301963. The leaf spectroscopic data are archived in the Specchio Spectral Information System (https://specchio.ch/). All data will be publicly available upon publication.

Code availability

The bash codes developed for deriving genetic structure, and *R* and *Matlab* codes developed for the statistical analysis are deposited at GitHub (*to be established upon the publication*) and will be publicly available upon publication.

References

1 Alexander, D. H., Shringarpure, S. S., Novembre, J., & Lange, K. (2015). Admixture 1.3 software manual. Los Angeles: UCLA Human Genetics Software Distribution.

2 Bolte, A., Czajkowski, T., & Kompa, T. (2007). The north-eastern distribution range of European beech—a review. Forestry, 80(4), 413-429.

3 Capblancq, T., & Forester, B. R. (2021). Redundancy analysis: A Swiss Army Knife for landscape genomics. Methods in Ecology and Evolution, 12(12), 2298-2309.

4 Capblancq, T., Morin, X., Gueguen, M., Renaud, J., Lobreaux, S., & Bazin, E. (2020). Climate - associated genetic variation in *Fagus sylvatica* and potential responses to climate change in the French Alps. Journal of Evolutionary Biology, 33(6), 783-796.

5 Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., ... & Naeem, S. (2012). Biodiversity loss and its impact on humanity. Nature, 486(7401), 59-67.

6 Caudullo, G., Welk, E., & San-Miguel-Ayanz, J. (2017). Chorological maps for the main European woody species. Data in brief, 12, 662-666.

7 Cavender-Bares, J., Meireles, J. E., Couture, J. J., Kaproth, M. A., Kingdon, C. C., Singh, A., ... & Townsend, P. A. (2016). Associations of leaf spectra with genetic and phylogenetic variation in oaks: prospects for remote detection of biodiversity. Remote Sensing, 8(3), 221.

8 Cawse-Nicholson, K., Townsend, P. A., Schimel, D., Assiri, A. M., Blake, P. L., Buongiorno, M. F., ... & SBG Algorithms Working Group. (2021). NASA's surface biology and geology designated observable: A perspective on surface imaging algorithms. Remote Sensing of Environment, 257, 112349.

9 Celesti, M., Rast, M., Adams, J., Boccia, V., Gascon, F., Isola, C., & Nieke, J. (2022, July). The Copernicus Hyperspectral Imaging Mission for the Environment (Chime): Status and Planning. In IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium (pp. 5011-5014). IEEE.

10 Csilléry, K., Lalagüe, H., Vendramin, G. G., González - Martínez, S. C., Fady, B., & Oddou - Muratorio, S. (2014). Detecting short spatial scale local adaptation and epistatic selection in

climate - related candidate genes in European beech (*Fagus sylvatica*) populations. Molecular Ecology, 23(19), 4696-4708.

11 Cuervo-Alarcon, L., Arend, M., Müller, M., Sperisen, C., Finkeldey, R., & Krutovsky, K. V. (2021). A candidate gene association analysis identifies SNPs potentially involved in drought tolerance in European beech (Fagus sylvatica L.). Scientific Reports, 11(1), 2386.

12 Czyż, E. A., Guillén Escribà, C., Wulf, H., Tedder, A., Schuman, M. C., Schneider, F. D., & Schaepman, M. E. (2020). Intraspecific genetic variation of a Fagus sylvatica population in a temperate forest derived from airborne imaging spectroscopy time series. Ecology and evolution, 10(14), 7419-7430.

13 Czyż, E. A., Schmid, B., Hueni, A., Eppinga, M. B., Schuman, M. C., Schneider, F. D., ... & Schaepman, M. E. (2023). Genetic constraints on temporal variation of airborne reflectance spectra and their uncertainties over a temperate forest. Remote Sensing of Environment, 284, 113338.

14 Des Roches, S., Post, D. M., Turley, N. E., Bailey, J. K., Hendry, A. P., Kinnison, M. T., ... & Palkovacs, E. P. (2018). The ecological importance of intraspecific variation. Nature ecology & evolution, 2(1), 57-64.

15 Exposito-Alonso, M., Booker, T. R., Czech, L., Gillespie, L., Hateley, S., Kyriazis, C. C., ... & Zess, E. (2022). Genetic diversity loss in the Anthropocene. Science, 377(6613), 1431-1435.

16 Fang, J., & Lechowicz, M. J. (2006). Climatic limits for the present distribution of beech (*Fagus* L.) species in the world. Journal of Biogeography, 33(10), 1804-1819.

17 Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1 - km spatial resolution climate surfaces for global land areas. International journal of climatology, 37(12), 4302-4315.

18 Fridley, J. D., & Grime, J. P. (2010). Community and ecosystem effects of intraspecific genetic diversity in grassland microcosms of varying species diversity. Ecology, 91(8), 2272-2283.

19 Gallois, A., Audran, J. C., & Burrus, M. (1998). Assessment of genetic relationships and population discrimination among *Fagus sylvatica* L. by RAPD. Theoretical and applied genetics, 97, 211-219.

20 Gausman, H. W., Allen, W. A., Escobar, D. E., Rodriguez, R. R., & Cardenas, R. (1971). Age Effects of Cotton Leaves on Light Reflectance, Transmittance, and Absorptance and on Water Content and Thickness 1. Agronomy Journal, 63(3), 465-469.

21 Gitelson, A. A., Keydan, G. P., & Merzlyak, M. N. (2006). Three - band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. Geophysical research letters, 33(11).

22 Gomory, D., Paule, L., Shvadchak, I. M., Popescu, F., Sulkowska, M., Hynek, V., & Longauer, R. (2003). Spatial patterns of the genetic differentiation in European beech (*Fagus sylvatica* L.) at allozyme loci in the Carpathians and the adjacent regions. Silvae Genetica, 52(2), 78-83.

23 Guevara, M. Á., Sánchez-Gómez, D., Vélez, M. D., de María, N., Díaz, L. M., Ramírez-Valiente, J. A., ... & Cervera, M. T. (2022). Epigenetic and Genetic Variability in Contrasting Latitudinal *Fagus sylvatica* L. Provenances. Forests, 13(12), 1971.

24 Guyot, G., Baret, F., & Jacquemoud, S. (1992). Imaging spectroscopy for vegetation studies (Vol. 2, pp. 145-165). Kluwer Academic Publishers: Norwell, MA, USA.

25 Hazler, K., Comps, B., Sugar, I., Melovski, L., Tashev, A., & Gracan, J. (1997). Genetic structure of *Fagus sylvatica* L. populations in Southeastern Europe. Silvae Genetica, 46(4), 229-235.Paule, L. (1995). Gene conservation in European beech (*Fagus sylvatica* L.). Forest Genetics, 2(3), 161-170.

26 Hector, A., & Bagchi, R. (2007). Biodiversity and ecosystem multifunctionality. Nature, 448(7150), 188-190.

27 Hector, A., Hautier, Y., Saner, P., Wacker, L., Bagchi, R., Joshi, J., ... & Loreau, M. (2010). General stabilizing effects of plant diversity on grassland productivity through population asynchrony and overyielding. Ecology, 91(8), 2213-2220.

28 Hughes, A. R., Inouye, B. D., Johnson, M. T., Underwood, N., & Vellend, M. (2008). Ecological consequences of genetic diversity. Ecology letters, 11(6), 609-623.

29 IPBESIntergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES). (2019). Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.

30 Isbell, F., Calcagno, V., Hector, A., Connolly, J., Harpole, W. S., Reich, P. B., ... & Loreau, M. (2011). High plant diversity is needed to maintain ecosystem services. Nature, 477(7363), 199-202.

31 Isbell, F., Craven, D., Connolly, J., Loreau, M., Schmid, B., Beierkuhnlein, C., ... & Eisenhauer, N. (2015). Biodiversity increases the resistance of ecosystem productivity to climate extremes. Nature, 526(7574), 574-577.

32 Jacquemoud, S., & Ustin, S. (2019). Leaf optical properties. Cambridge University Press.

33 Jenkins, C. N., & Joppa, L. (2009). Expansion of the global terrestrial protected area system. Biological conservation, 142(10), 2166-2174.

34 Johnson, C. N., Balmford, A., Brook, B. W., Buettel, J. C., Galetti, M., Guangchun, L., & Wilmshurst, J. M. (2017). Biodiversity losses and conservation responses in the Anthropocene. Science, 356(6335), 270-275.

35 Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. Quality & quantity, 52, 1957-1976.

36 Lou, R. N., Jacobs, A., Wilder, A. P., & Therkildsen, N. O. (2021). A beginner's guide to low - coverage whole genome sequencing for population genomics. Molecular Ecology, 30(23), 5966-5993.

37 Madritch, M. D., Kingdon, C. C., Singh, A., Mock, K. E., Lindroth, R. L., & Townsend, P. A. (2014). Imaging spectroscopy links aspen genotype with below-ground processes at landscape scales. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1643), 20130194.

38 Magri, D., Vendramin, G. G., Comps, B., Dupanloup, I., Geburek, T., Gömöry, D., ... & De Beaulieu, J. L. (2006). A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. New phytologist, 171(1), 199-221.

39 Meger, J., Ulaszewski, B., & Burczyk, J. (2021). Genomic signatures of natural selection at phenology-related genes in a widely distributed tree species *Fagus sylvatica* L. BMC genomics, 22(1), 1-20.

40 Meireles, J. E., Cavender - Bares, J., Townsend, P. A., Ustin, S., Gamon, J. A., Schweiger, A. K., ... & O'Meara, B. C. (2020). Leaf reflectance spectra capture the evolutionary history of seed plants. New Phytologist, 228(2), 485-493.

41 Miller, J. R., Hare, E. W., Hollinger, A. B., & Sturgeon, D. R. (1987, January). Imaging spectrometry as a tool for botanical mapping. In Imaging spectroscopy II (Vol. 834, pp. 108-113). SPIE.

42 Mishra, B., Ulaszewski, B., Meger, J., Aury, J. M., Bodénès, C., Lesur-Kupin, I., ... & Thines, M. (2022). A Chromosome-level genome assembly of the European Beech (*Fagus sylvatica*) reveals anomalies for organelle DNA integration, repeat Content and distribution of SNPs. Frontiers in genetics, 12, 691058.

43 Mora, C., Andréfouët, S., Costello, M. J., Kranenburg, C., Rollo, A., Veron, J., ... & Myers, R. A. (2006). Coral reefs and the global network of marine protected areas. Science, 312(5781), 1750-1751.

44 Perrings, C., Duraiappah, A., Larigauderie, A., & Mooney, H. (2011). The biodiversity and ecosystem services science-policy interface. Science, 331(6021), 1139-1140.

45 Petit, R. J., Aguinagalde, I., de Beaulieu, J. L., Bittkau, C., Brewer, S., Cheddadi, R., ... & Vendramin, G. G. (2003). Glacial refugia: hotspots but not melting pots of genetic diversity. science, 300(5625), 1563-1565.

46 Pfenninger, M., Reuss, F., Kiebler, A., Schönnenbeck, P., Caliendo, C., Gerber, S., ... & Feldmeyer, B. (2020). Genomic basis of drought resistance in *Fagus sylvatica*. bioRxiv, 2020-12.

47 Pluess, A. R., & Weber, P. (2012). Drought-adaptation potential in Fagus sylvatica: linking moisture availability with genetic diversity and dendrochronology. PloS one, 7(3), e33636.

48 Podani, J., & Miklós, I. (2002). Resemblance coefficients and the horseshoe effect in principal coordinates analysis. Ecology, 83(12), 3331-3343.

49 Purvis, A., & Hector, A. (2000). Getting the measure of biodiversity. Nature, 405(6783), 212-219.

50 Reich, P. B., Tilman, D., Isbell, F., Mueller, K., Hobbie, S. E., Flynn, D. F., & Eisenhauer, N. (2012). Impacts of biodiversity loss escalate through time as redundancy fades. Science, 336(6081), 589-592.

51 Schaepman, M. E., Jehle, M., Hueni, A., D'Odorico, P., Damm, A., Weyermann, J., ... & Itten, K. I. (2015). Advanced radiometry measurements and Earth science applications with the Airborne Prism Experiment (APEX). Remote Sensing of Environment, 158, 207-219.

52 Schaepman, M. E., Ustin, S. L., Plaza, A. J., Painter, T. H., Verrelst, J., & Liang, S. (2009). Earth system science related imaging spectroscopy—An assessment. Remote Sensing of Environment, 113, S123-S137.

53 Scholes, R. J., Mace, G. M., Turner, W., Geller, G. N., Jürgens, N., Larigauderie, A., ... & Mooney, H. A. (2008). Toward a global biodiversity observing system. Science, 321(5892), 1044-1045.

54 Scott, J. M., Csuti, B., Estes, J. E., & Anderson, H. (1989). Status assessment of biodiversity protection. Conservation Biology, 3(1), 85-87.

55 Stefanini, C., Csilléry, K., Ulaszewski, B., Burczyk, J., Schaepman, M. E., & Schuman, M. C. (2023). A novel synthesis of two decades of microsatellite studies on European beech reveals decreasing genetic diversity from glacial refugia. Tree Genetics & Genomes, 19(1), 3.

56 Sutherland, W. J., Pullin, A. S., Dolman, P. M., & Knight, T. M. (2004). The need for evidence-based conservation. Trends in ecology & evolution, 19(6), 305-308.

57 Tang, T., Zhang, N., Bongers, F. J., Staab, M., Schuldt, A., Fornoff, F., ... & Liu, X. (2022). Tree species and genetic diversity increase productivity via functional diversity and trophic feedbacks. Elife, 11, e78703.

58 Thiébaut, B., Lumaret, R., & Vernet, P. H. (1982). The bud enzymes of beech (*Fagus sylvatica* L.) genetic distinction and analysis of polymorphism in several French populations.

59 Tilman, D., & Downing, J. A. (1994). Biodiversity and stability in grasslands. Nature, 367(6461), 363-365.

60 Turner, W. (2014). Sensing biodiversity. Science, 346(6207), 301-302.

61 van Moorsel, S. J., Hahl, T., Petchey, O. L., Ebeling, A., Eisenhauer, N., Schmid, B., & Wagg, C. (2021). Co - occurrence history increases ecosystem stability and resilience in experimental plant communities. Ecology, 102(1), e03205.

62 Wilson, E. O. (1988). Biodiversity.

63 Yamasaki, E., Altermatt, F., Cavender-Bares, J., Schuman, M. C., Zuppinger-Dingley, D., Garonna, I., ... & Shimizu, K. K. (2017). Genomics meets remote sensing in global change studies: monitoring and predicting phenology, evolution and biodiversity. Current opinion in environmental sustainability, 29, 177-186.