



Identification of full-sibling families from natural single-tree ash progenies based on SSR markers and genome-wide SNPs

Melina Krautwurst¹ · Franziska Past² · Birgit Kersten¹ · Ben Bubner² · Niels A. Müller¹

Received: 18 July 2023 / Accepted: 2 July 2024
© The Author(s) 2024

Abstract

Common ash, *Fraxinus excelsior*, is threatened by the invasive pathogen *Hymenoscyphus fraxineus*, which causes ash dieback. The pathogen is rapidly spreading throughout Europe with severe ecological and economic consequences. Multiple studies have presented evidence for the existence of a small fraction of genotypes with low susceptibility. Such genotypes can be targets for natural and artificial selection to conserve *F. excelsior* and associated ecosystems. To resolve the genetic architecture of variation in susceptibility it is necessary to analyze segregating populations. Here we employed about 1000 individuals of each of four single-tree progenies from potentially tolerant mother trees to identify full-sibling (full-sib) families. To this end, we first genotyped all 4000 individuals and the four mothers with eight SSR markers. We then used the program COLONY to predict full-sibs without knowledge of the paternal genotypes. For each single-tree progeny, COLONY predicted dozens of full-sib families, ranging from 3–166 individuals. In the next step, 910 individuals assigned to full-sib families with more than 28 individuals were subjected to high-resolution genotyping using over one million genome-wide SNPs which were identified with Illumina low-coverage resequencing. Using these SNP genotyping data in principal component analyses we were able to assign individuals to full-sib families with high confidence. Together the analyses revealed five large families with 73–212 individuals. These can be used to generate genetic linkage maps and to perform quantitative trait locus analyses for ash dieback susceptibility. The elucidation of the genetic basis of natural variation in ash may support breeding and conservation efforts and may contribute to more robust forest ecosystems.

Keywords Common ash (*Fraxinus excelsior*) · Breeding-without-breeding · SSR markers · SNPs · Whole-genome resequencing · Full-sibling families · Ash dieback

Introduction

In the early 1990s the necrotrophic ascomycete fungus *Hymenoscyphus fraxineus*, causing the ash dieback (ADB) disease, was first observed in Europe (Coker et al. 2019; Evans 2019). The fungus spread from its first introduction to

Poland via wind-borne spores over most European ash populations (Kowalski 2006). The fungus can be traced back to Eastern Asia where it is associated with native *Fraxinus* species (Husson et al. 2011; McKinney et al. 2012; Zhao et al. 2013; Landolt et al. 2016). In Europe, the host of the pathogenic fungus is common ash (*Fraxinus excelsior*). Infected trees suffer from crown dieback and necrotic lesions. In the end, infection often leads to the death of the trees causing severe losses to European woodlands (Bakys et al. 2013; Coker et al. 2019).

Notably, *Fraxinus excelsior* exhibits natural variation in ADB susceptibility and several studies have shown that part of this variation is heritable with estimated heritabilities of 0.25–0.57 (Pliura and Baliuckas 2007; McKinney et al. 2011, 2012, 2014; Pliura et al. 2011; Lobo et al. 2014, 2015; Enderle et al. 2015; Harper et al. 2016; Muñoz et al. 2016; Plumb et al. 2020). Several different phenotypes are presumed to be connected to ADB susceptibility. For example,

Melina Krautwurst and Franziska Past have contributed equally to this work.

✉ Ben Bubner
ben.bubner@thuenen.de

✉ Niels A. Müller
niels.mueller@thuenen.de

¹ Thünen Institute of Forest Genetics, 22927 Grosshansdorf, Germany

² Thünen Institute of Forest Genetics, 15377 Waldsiedersdorf, Germany

the timing of bud burst and senescence may be important for variation in ADB although results of different studies are not always consistent (McKinney et al. 2011; Bakys et al. 2013; Stener 2013; Pliura et al. 2016; Nielsen et al. 2017).

Genotype–phenotype associations could reveal the genetic basis of variation in ADB and highlight candidate genes. For genetic mapping or pedigree studies, the selection of parents and artificial mating needs to be conducted. In natural tree populations it can be challenging to perform artificial mating or to identify the father to the naturally occurring seedlings, especially in species with a complex mating system as in *F. excelsior*. Common ash is a wind-pollinated and wind-dispersed, polygamous subdioecious tree species. The method “breeding-without-breeding” (BwB) overcomes the need for artificial mating, by working with paternally unknown but maternally known material. Mothers can be selected based on their genotype or phenotype. Paired with DNA markers, it is possible to reconstruct pedigree structures with BwB and to use the identified full-sib families for quantitative trait locus analyses or the assessment of various breeding values (El-Kassaby and Lstibůrek 2009; Lstibůrek et al. 2011, 2015). For pedigree prediction choosing a suitable downstream analysis for the sample set is important. Different molecular markers can be effective in assessing kinships, such as simple sequence repeat (SSR) and single-nucleotide polymorphism (SNP) markers (Amom et al. 2020; Jiang et al. 2020; Zeng et al. 2023).

SSR and SNP markers can be powerful tools in combination or separately (García et al. 2018; Capo-chichi et al. 2022; Zeng et al. 2023). SNPs offer the opportunity to identify single base changes between individuals, are mostly biallelic, as well as the most abundant source of genetic polymorphism (Agarwal et al. 2008). With new sequencing technologies high numbers of SNPs can be reliably identified (Howe et al. 2020). SSRs are multi-allelic, highly polymorphic and currently the cheaper option for kinship assessment compared to SNPs, for which sequencing needs to be performed (Ramesh et al. 2020).

For kinship identification within high numbers of individuals, an application of both marker types could be of advantage, because SSR markers offer a low-cost possibility with sufficient performance and can be used for preselecting large full-sib-families before sequencing and SNP calling. The purpose of this study is to analyze the feasibility and advantages of the application of two consecutive methodological steps for kinship assessment in common ash: (i) low-cost genotyping with SSRs to predict potential full-sib families, followed by (ii) high-resolution genotyping using whole-genome sequencing.

Material and methods

Plant material

The four mother trees are distributed across the state of Mecklenburg-Western Pomerania in the north-east of Germany (Table 1). These trees were selected using an assessment scheme developed in a previous project (ResEsche, FNR project number “FKZ 22019915”). This scheme should ensure the vitality and silvicultural quality of the selected trees. The vitality criteria are assessed on the basis of foliage, shoot and trunk damage. The quality is recorded with parameters such as diameter at breast height (DBH), height and trunk shape. The selected mother trees all showed no or only a few dieback symptoms in the crown area (no more than 10% crown defoliation, no more than 15% replacement shoot proportion). The upright growth was of perfect or at least of normal quality (straight/upright growth, weak twisted growth, solid woodiness, etc.). Three of the four mother trees showed no signs of ADB, only Dar-18 had a slight stem necrosis. Around 3000 seeds per tree were collected in 2018 as green seeds. All seeds were sown in a nursery bed within two weeks after harvesting. After germinating in Spring 2019, 960 seedlings per spring progeny were planted in 6 × 4 QuickPot plates (24 pots, 16 cm deep, HerkuPlast Kubern GmbH, Ering, Germany). They were first cultivated in a greenhouse and then transferred under a shading net in the nursery. In September 2019 plants were re-potted in 4 × 3 QuickPot plates (12 pots, 18 cm deep, HerkuPlast Kubern GmbH, Ering, Germany) and stayed in the nursery until planting. From 15 to 17th of April 2021 the seedlings were planted in a semi-randomized block design at a trial site near Schulzendorf (Brandenburg, Germany; Table 1). The area around the trial site is characterized by agriculture and small forests. Infected trees of *F. excelsior* with ash dieback symptoms were observed adjacent to the trial site.

Table 1 Location information of mother trees and trial site

Name	Forest district	Forest area	Coordinates
<i>Mother trees</i>			
Fri-8	Radelübbe	Friedensthal	53.6135667, 11.3055333
Kar-4	Jaegerhof	Karlsburg	54.0674667, 13.4514500
Eve-2	Grevesmühlen	Eversdorf	53.8784500, 11.2492833
Dar-18	Dargun	Dargun	53.8537899, 12.8458667
<i>Trial site</i>			
Schulzendorf	Strausberg	Wriezen	52.686818, 14.125038

Sampling, DNA extraction and SSR genotyping

Sampling for SSR genotyping was conducted in the nursery in late summer 2019 (Fri-8) and early summer 2020 (other progenies). The QuickPots were arranged in a 96-well-plate-like format. This arrangement allowed sampling in 96-well-plates without the need for time-consuming individual labeling. The 96-well-plates intended for DNA extraction were filled with two ceramic beads (1.4 mm Omni Beads, Omni International, Kennesaw; United States) per well using the customized Brendan bead dispenser (<https://customlabinstitute.wordpress.com>). The plate was cooled during the sampling process with a plate fitting ice pack. The sample, a 2 × 3 mm piece of the youngest, fully developed leaflet, was taken with forceps, which were cleaned with Ethanol (70%) between each sample. After sampling, the plates were stored at – 80 °C. Because of the large number of samples, a “quick and dirty”-method for DNA extraction was chosen (Hu et al. 2014). The samples in the frozen sampling plates were homogenized (30 Hz; MM400, Retsch, Haan, Germany) for at least four minutes in a precooled container. Depending on the homogenization grade, additional homogenization was performed. After adding 200 µl buffer (50 mM Tris, pH 8; 300 mM NaCl, 0, 1 g/ml saccharose), the plates were again homogenized for another minute. After centrifuging the plate (5889 × g; 5 min), the upper phase, containing the DNA, was directly used for polymerase chain reaction (PCR) after diluting 1:10 on the same day. Plates with the “quick and dirty” DNA extract were stored at

– 20 °C and, after another round of centrifugation, could be used for another PCR.

The PCR was performed with the Multiplex PCR Kit (Type-It Microsatellite master mix; Qiagen, Hilden, Germany). The SSR primer mix consisted of eight primers (see Tables 2, S1). With this SSR multiplex, a touchdown procedure was performed (Table S2). The first progeny (Fri-8) was analyzed with F24 which was later replaced by F12, because F12 was more variable and more reliable.

The PCR products were analyzed with capillary electrophoresis (GenomeLab™ GeXP, Beckman Coulter) and the corresponding chemical kit (SCIEX, Framingham U.S.A). The peak scoring was done with the provided software (GenomeLab) to obtain a list of the alleles (Tables S3–S7).

COLONY analysis

In order to determine genetic sample relationships, the SSR-genotyping output was transferred to the COLONY Software (Jones and Wang 2010) Version: 2.0.6.5/2.0.6.6. In addition to the input data, input parameters for the analysis of all progenies had to be defined (see Supplement: Supplemental materials on COLONY parameters for family estimations with SSR markers). In contrast to other methods, the software can be used with monoecious and dioecious species and is not restricted to codominant markers without genotyping error. Also, instead of pairwise comparisons, it uses a full-pedigree likelihood approach, which takes into account the likelihood of the entire pedigree structure and allows the simultaneous inference of parentage and sibship (Jones and

Table 2 SSR primers used for genotyping of ash seedlings in a multiplex assay. Primer sequences are based on the reference genome BATG0.5

Primer	Locus name	Primer sequence	Motif	Conc	Label (color)
F12	FEMSATL12C	F: TTTTGAACCCCTTGATTTTG R: CTATATACACCTACCTCTC	(GA) ₆ CA (GA) ₈	0.2 µM	DY-751 (blue)
F23	Contig184_179564_179439	F: GCCATTGTTGGGTTTCACTC R: CCGGGCTTAGTATCCAACCTG	(AG) ₁₀	0.2 µM	DY-682 (green)
F25	Contig1992_131610_131822	F: CTCGGAGGTGGTTGATGAGT R: AGAGCTCCCAACGCTCAATA	(TA) ₇	0.2 µM	DY-682 (green)
F26	Contig918_100265_100431	F: TGTTAGTGGTATGGTGGAGGC R: TTGCAGGCACAATTACATAAAG	(CT) ₁₃	0.2 µM	DY-751 (blue)
F27	Contig5418_11372_11513	F: TTATTGCCGATGTCTGTGGA R: CGGGAAGTTTCACCTCAGAT	(TG) ₉	0.2 µM	Cy5 (black)
F28	Contig3670_10994_10705	F: CACTTCAATGTCAGCATTCCA R: GAAAGTTCACCCAGTCAAAAT	(CT) ₁₅	0.4 µM	DY-682 (green)
F29	Contig4738_19258_19142	F: GCAAGGGAAGTAGCAACGAC R: TGTGACTCAAAATAGGGCTCAGA	(AG) ₂₁	0.2 µM	DY-751 (blue)
F30	Contig344_81629_81375	F: GGAGCAGTACCATATGCGCT R: AGGAAGATCTAACGCTGCTTG	(AG) ₁₄	0.2 µM	Cy5 (black)
F24	Contig5748_109580_109480	F: ACATGCCTTCTCTTCGCC R: TTCTAGGGCCTGCAAACAAC	(TA) ₈	0.1 µM	DY-751 (blue)

F12 is adapted from marker FEMSATL12 (Lefort et al. 1999). All other primers are from Sollars et al. (2017)

Conc. final concentration

Wang 2010). This likelihood approach can lead to different outputs if the analysis is repeated. To ensure the reliability of the results, the software was run twice. The two runs were compared and the individuals that were observed in both runs were chosen. The chosen families are listed in Table 3 and named according to the COLONY output (e.g., FS 05).

DNA extraction and Illumina low-coverage resequencing

The first batch of 167 samples with young leaf material was collected in September 2020, and immediately placed on ice in 1.5 ml Eppendorf safe-lock tubes (Eppendorf, Wesslingen, Germany). In addition, leaves from the four mother tree clones (Table 1) were collected. All samples were stored at -80°C until DNA extraction. Frozen samples were homogenized using pestle and mortar in liquid nitrogen. All following steps were conducted following the DNA extraction protocol by Bruegmann et al. (2022). The second batch of 751 samples was collected in June 2021. For the 751 samples, we used the MagMAX™ Plant DNA Isolation Kit (Thermo Fisher Scientific, Germany) following purification using the KingFisher™ Apex (Thermo Fisher Scientific, Germany) with a 96 deep-well head. DNA sample QC and library preparation for sequencing were performed by Novogene (UK) Ltd. (Cambridge, UK) for both sample batches. In the first batch, 163 of 172 samples passed the quality control, including the four mother tree samples. In the second batch, 747 of 751 samples passed the quality control. Sequencing data (2×150 bp reads) were generated on the Novaseq 6000 platform for both batches. The first batch

was sequenced to an average sequencing depth of $10.8 \times$ and the second batch to $11.3 \times$ according to the ash reference genome (Sollars et al. 2017). Both batches together comprise 906 samples plus the four mother trees.

Mapping and variant calling

For both batches, sequencing data were mapped against the common ash reference genome (Sollars et al. 2017) using bwa-mem (version bwa-0.7.17.tar.bz2 (Li and Durbin 2009)). Grouping of the reads and duplicates were marked using Picard tool's (v2.26.2) (<http://broadinstitute.github.io/picard/>). Joint variant calling was performed for batch one with GATK (version 4.2.3.0), following the best practices for germline short variant discovery wherever possible (Poplin et al. 2017). For the second batch, the variant calling was performed by Novogene using Sentieon (Aldana and Freed 2022). For generating gVCFs the `HaplotypeCaller` from GATK was used. After combining the gVCFs with GATK's `GenomicsDBImport`, the `GenotypeGVFs` tool was used for the joint genotyping. For batch one GATK version 4.2.3.0 and for batch two GATK version 4.0.5.1 were used (Kemp 2003).

Variant filtering

The two sample sets were filtered separately, with the same filtering options. For hard filtering, we mostly followed the documentation on 'Hard-filtering germline short variants' on the GATK website. We filtered indels and SNPs separately. We removed variants based on strand bias (FisherStrand 'FS' > 60 & StrandOddsRatio 'SOR' > 3) and mapping quality (RMSMappingQuality 'MQ' < 40 , MappingQualityRankSumTest 'MQRankSum' < -1). Based on the distribution of the variant confidence score QualByDepth 'QD' we chose a more stringent cutoff of $QD > 10$, to remove any low-confidence variants. Filtering was performed with bcftools v1.7 (Li 2011). We then extracted the variant sequencing depth values 'DP' and minor allele frequency `frq2` using vcftools v0.1.15 (Danecek et al. 2011). To visualize the DP and choose the parameters we used R (R Core Team 2022). Minimal mean DP was 5.2 and max-mean DP 10.3. Non-biallelic SNPs were excluded and SNPs with more than 10% missing data were removed. Filtering resulted in a set of 14.42 million SNPs for the first batch and 11.78 million SNPs for the second batch. Before the resulting VCF (Variant Call Format) files could be merged, an intersect was calculated using the `isec` function of bcftools (Danecek et al. 2021) to identify common SNPs in both VCF files. Then, the `merge` function was used to create one multi-sample file. Further, the minor allele frequency was filtered 0.08–0.50 with PLINK v1.9 (Purcell et al. 2007).

Table 3 Summary of all full-sibling families identified by SSR and SNP marker

Progeny	Abbreviation	SSR full-sibling family name	Size	SNP full-sibling family name	Size
Friedrichsthal	Fri-8	FS 10	81	Cluster 1	212
		FS 17	90		
		FS 22	63		
Karlsburg	Kar-4	FS 29	113	Cluster 1	164
		FS 05	37	Cluster 2	25
		FS 21	30		
		FS 41	33		
Eversdorf	Eve-2	FS 20	166	Cluster 2	138
		FS 08	85	Cluster 1	73
		FS 30	39	Cluster 3	35
Dargun	Dar-18	FS 19	123	Cluster1	116
		FS 14	28		

The description of Cluster indicates the groups of individuals clustering in the PCA (Figs. 1, 2)

FS Full-sib families identified by COLONY

The merged file of both batches comprised 5.87 million SNPs for 910 individuals.

We used PLINK for principal component analyses (PCA) to determine family clusters in the SNP datasets. Further, we used the R package ‘SNPRelate’ (Zheng et al. 2012) for the genome-wide identity-by-state analysis to create a dendrogram based on the SNP data. The dendrogram implements the formed clusters of the PCA and does not include the outliers. Only for Fri-8 outliers were included.

Data visualization

The results were further analyzed and visualized using the R packages ‘ggplot2’ (Wickham 2016), ‘VCFR’ (Knaus and Grünwald 2017), ‘MASS’ (Venables and Ripley 2003) and ‘dendextend’ (Galili 2015).

Results

A total of 960 samples per single-tree progeny of each of four mother trees, potentially tolerant against ash dieback, were collected and analyzed with eight SSR markers (Tables S3–S7). In the end, 3476 of 3840 (90.5%) samples could be successfully genotyped. The program COLONY predicted between 151 and 179 different fathers per progeny (Figs. 1, 2; Table S8) giving rise to single trees without full-sib-family membership up to full-sib families with 166 members. In total, 910 individuals, which were assigned to the nine largest predicted full-sib families, were selected for high-resolution genotyping using Illumina whole-genome resequencing (two sets). Only families with at least 28 individuals were chosen. With a sequencing depth of $10\times$ we were able to identify a total of 5.87 million SNPs after read mapping to the *F. excelsior* reference genome (Sollars et al. 2017) and SNP filtering. Employing only largely independent SNPs ($r^2 < 0.2$) within each progeny, we were able to reliably assign full-sib families. The final file included six samples that were sequenced and analyzed in set 1 and set 2. With these ‘twins’ we were able to compare the two downstream analyses. The ‘twins’ are represented in the Fri-8 dendrogram (Fig. 2), which shows the same results for both downstream analyses.

Table 3 summarizes all results of the full-sibling family identification using SSR and SNP markers. Full-sib families identified by COLONY are named with FS (for full-siblings) and the number given by COLONY. Full-sib families assigned with the PCA analysis are named after the number of clusters formed in the PCA. The dendrogram is based on identity-by-state analysis, showing the genetic distance as the proportion of loci where the alleles

are identical (Figs. 1, 2e, f). The vertical axis shows the proportion of the genetic distance between the individuals, with longer branches indicating greater genetic distance and shorter branches indicating closer genetic similarity. The height of the nodes corresponds with the genetic distance at which the identity-by-state algorithm decided to merge or split.

For the two single-tree progenies from Eve-2 and Dar-18, the family structure predicted by COLONY using the eight SSR markers was largely consistent with the SNP data analyses (Fig. 1). Only a few outliers were detected. In Eve-2, family sizes were close to the predictions, with the largest full-sib family comprising 138 individuals (Table 3). This demonstrates the general feasibility of performing breeding-without-breeding in ash with the eight described SSR markers. For Dar-18, a single dominant pollen donor gave rise to a single large full-sib family, of which 116 individuals could be confirmed with the SNP markers. Similar to Eve-2, Dar-18 also demonstrates the feasibility of the SSR markers approach.

The other two single-tree progenies from Kar-4 and Fri-8 showed discrepancies between the SSR and SNP marker classifications (Fig. 2). Fri-8 was predicted to be composed of three full-sib families using SSR markers. However, the SNP markers revealed it to be a single large family of 212 individuals, with additional outliers representing other pollen donors. The progenies of Fri-8 had the lowest genetic distance (0.20 individual dissimilarity) compared to the other dendrograms (0.25–0.30 individual dissimilarity). Further, the outliers of Fri-8 are placed at a higher level and show that they are genetically less similar. The lower individual dissimilarity of Fri-8 compared to the other progenies can be an indicator that the unknown fathers of the full-sib family are more closely related than the fathers of other full-sib families, especially compared with Dar-18, which shows the highest dissimilarity in all four dendrograms.

In Kar-4, the SNP marker analysis showed two relatively large families (164 and 25 individuals) instead of four predicted by SSR markers. Full-sib families identified by COLONY for Kar-4, indicated by blue and yellow dots, formed one cluster in the SNP PCA (Fig. 2), demonstrating these are one family instead of two. The predicted fathers of the four families identified by COLONY in both runs showed strong similarity (Table S9). The assigned offspring changed between the predicted fathers when comparing the runs. SNP analysis defined these families (predicted by COLONY) as one family.

To ensure the reliability of our full-sib determinations, we suggest considering only those individuals identified as full-sibs by both SSR and SNP methods. This combined approach enhances confidence, thus improving the robustness and accuracy of the family structure predictions.

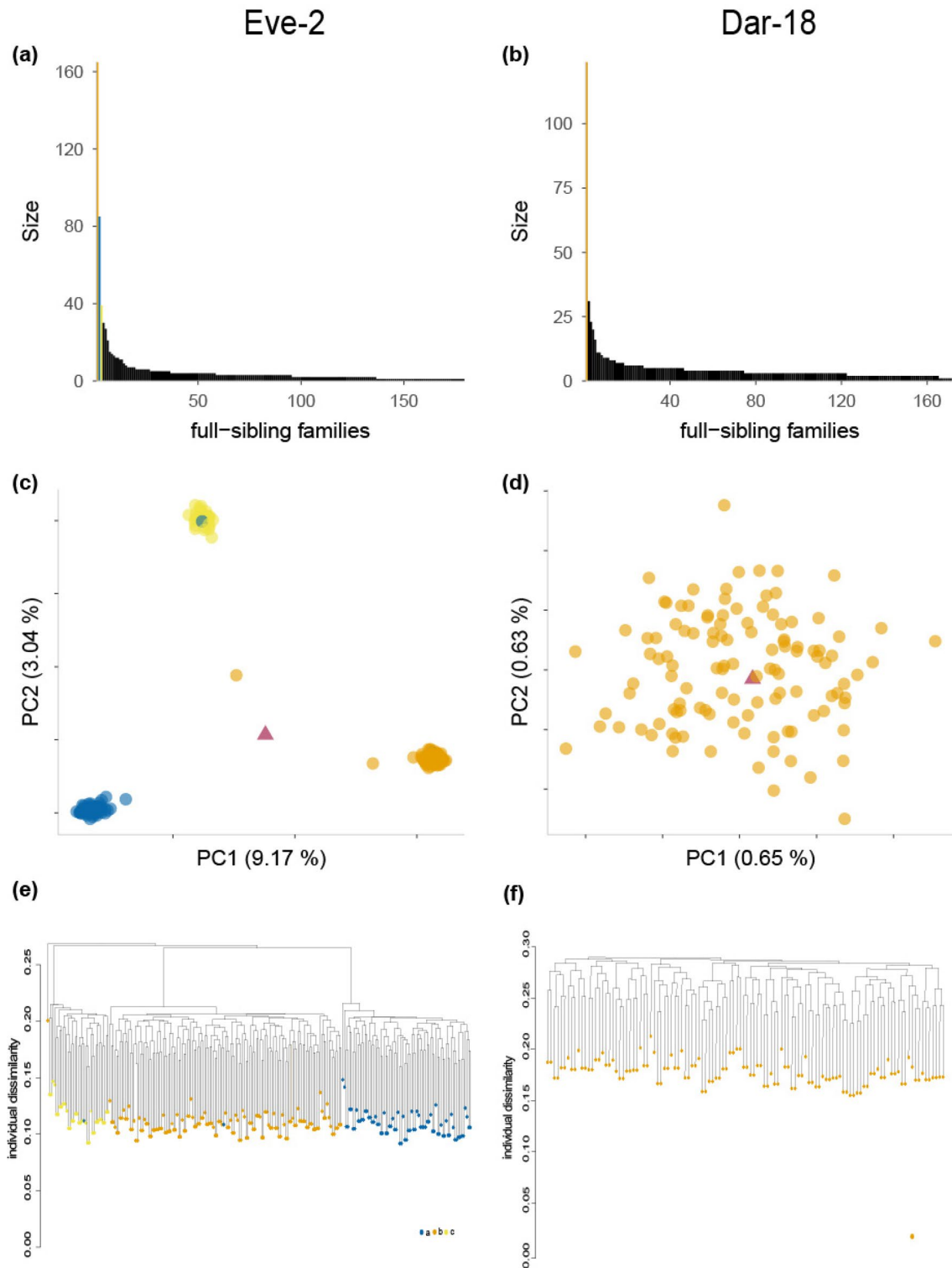


Fig. 1 Comparison of SSR and SNP full-sibling identification from the mother trees Eve-2 and Dar-18. The bar plots **a** and **b** show all progenies that were analyzed with the SSR markers and assigned to full-sibling families (ranging from 1 to 165 siblings). Each bar represents one predicted full-sibling family. Selected families with more than 30 individuals are indicated by colors, that is three families for Eve-2 **a** and one family for Dar-18 **b**. The principal component analyses **c** and **d** show the SNP marker results. The color scheme of the

dots represents the results of the SSR markers. The red triangle represents the mother tree. Panel **e** and **f** show the results of genome-wide identity-by-state analyses using the SNP marker results in dendrograms. The y-axis represents the individual dissimilarity and the x-axis represents the individual samples being clustered. The dots represent the SSR results and the dendrogram clustering represents the SNP markers

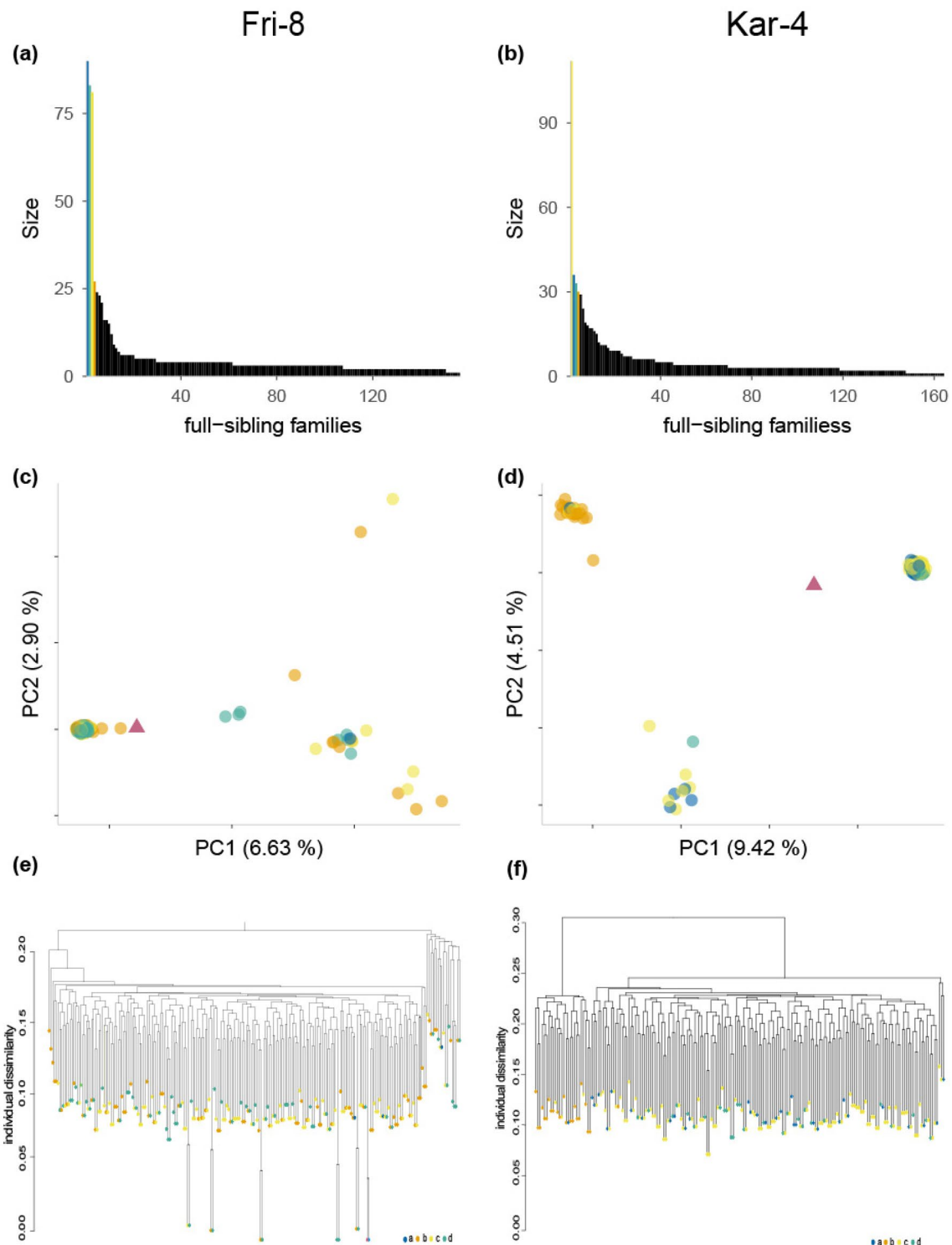


Fig. 2 Comparison of SSR and SNP full-sibling identification from the mother trees Fri-8 and Kar-4. The bar plots **a** and **b** show all progenies that were analyzed with the SSR markers and assigned to full-sibling families (ranging from 1 to 112 siblings). Each bar represents one predicted full-sibling family. Selected families with more than 30 individuals are indicated by colors, that is four families for Fri-8 **a** and four families for Kar-4 **b**. The principal component analyses **c**,

d show SNP marker results. The color scheme of the dots represents the results of the SSR markers. The red triangle represents the mother tree. Panels **e**, **f** show the results of genome-wide identity-by-state analysis using the SNP marker in dendrograms. The y-axis represents the individual dissimilarity and the x-axis represents the individual samples being clustered. The dots represent the SSR results and the dendrogram clustering represents the SNP markers

Discussion

For genotype–phenotype association studies it is important to maximize the statistical power by employing large mapping populations with a high number of individuals. In BwB approaches, the reliable determination of full-sib families is critical. Although whole-genome resequencing and genome-wide SNP detection provided robust data for this purpose, sequencing costs are still prohibitive with thousands of samples, such as the 4,000 individuals in our study. The preselection of individuals prior to SNP genotyping was thus important, and SSR markers provided the required simplicity and relatively low costs.

Our results show that implementing a two-step genotyping process with SSR and whole-genome sequencing is an effective way to achieve the identification of full-sib families in ash. Similar methodologies have been employed in other forestry studies, highlighting the effectiveness of the BwB approach. For example, in a study on Douglas-fir, Slavov et al. (2005) used two paternity assignment methods to reconstruct full-sib families within 16 half-sibling families. The program CERVUS (version 2.0), based on likelihood-based assignment, identified parentage with high accuracy. This was complemented by the PFL (Pedigree-Free Likelihood) program, which assigns paternity by genotypic exclusion. Both methods demonstrated the feasibility and effectiveness of the BwB approach (El-Kassaby et al. 2007). Additionally, Lstibůrek et al. (2015) conducted a simulation study examining different sizes of three conceptual populations using the BwB approach. The study demonstrated that it is possible to obtain genetic information and select superior genotypes from commercial forest plantations without the need for controlled breeding. These findings were further supported by studies from Čepl et al. (2018) which also underlined the potential of the BwB strategy for genetic analysis and breeding in forestry.

SSRs can be used to do a preselection of individuals for resequencing based on full-sib prediction, thus avoiding sequencing of less relevant individuals, the genome-wide SNP data can be employed for full-sib validation and construction of genetic maps. Regarding the discrepancies, especially in Fri-8, it becomes clear how crucial the accurate scoring of the SSR markers is. Two SSR markers showed inconclusive peaks in some cases in this population. This further points to scoring errors and/or null alleles being responsible for the incongruence between SSR and SNP markers. It would be interesting to re-evaluate the SSR genotypes in the resequencing data to better understand the underlying cause. Additionally, strong similarity of predicted fathers may be a further indication of unreliable progeny reconstruction and should be

carefully checked especially when relying on SSR markers alone. Despite some contradiction between SSR and SNP marker classification, the preselection of individuals by SSR analysis allowed us to identify relatively large full-sib families for deeper analysis by SNP genotyping. Similar studies have shown comparable results (van Inghelandt et al. 2010). In a study by Zavinon et al. (2020) family identification in Beninese pigeon pea (*Cajanus cajan* (L.)) populations was conducted with 30 informative SSR loci and 794 genotyping-by-sequencing (GBS) derived SNPs. The results with both marker sets were similar, but the PCA based on SNP markers showed the more accurate results.

For genotype–phenotype association studies, crossing and back-crossing generations are essential. For tree species with long generation times, this can be challenging. With the BwB technique we were able to identify full-sib families within the F_1 generation of four different tolerant mother trees. The combination of SSR and SNP markers enabled the successful identification of families that are a valuable resource to perform quantitative trait locus analyses for susceptibility to *H. fraxineus* in future studies.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s41348-024-00966-2>.

Acknowledgements We thank our technical assistants Viktoria Blunk, Annika Eikhof, Katrin Groppe, Marlies Karas, Heidrun Mattauch, and Regina Zimmermann for their help with laboratory and fieldwork. We thank members of the FraxForFuture project and the Thünen Institute of Forest Genetics for experimental advice and discussion on the project and data analysis. The FraxForFuture research network is funded by the German Federal Ministry of Food and Agriculture and the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety. The projects of the sub-networks are funded by the Waldklimafonds (funding code: 2219WK21A4) and the Thünen Institute. The project executing agency is the Fachagentur für Nachhaltige Rohstoffe e.V. (FNR).

Funding Open Access funding enabled and organized by Projekt DEAL. Bundesanstalt für Landwirtschaft und Ernährung, Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz, Waldklimafonds, Fachagentur Nachhaltige Rohstoffe

Data availability All genomic variants have been deposited in the European Variant Archive (EVA) under the accession number ‘PRJEB64325’.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated

otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agarwal M, Shrivastava N, Padh H (2008) Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep* 27:617–631. <https://doi.org/10.1007/s00299-008-0507-z>
- Aldana R, Freed D (2022) Data processing and germline variant calling with the sentieon pipeline. *Methods Mol Biol* 2493:1–19. https://doi.org/10.1007/978-1-0716-2293-3_1
- Amom T, Tikendra L, Apana N, Goutam M, Sonia P, Koijam AS et al (2020) Efficiency of RAPD, ISSR, iPBS, SCoT and phytochemical markers in the genetic relationship study of five native and economical important bamboos of North-East India. *Phytochemistry* 174:112330. <https://doi.org/10.1016/j.phytochem.2020.112330>
- Bakys R, Vasaitis R, Skovsgaard JP (2013) Patterns and severity of crown dieback in young even-aged stands of European ash (*Fraxinus excelsior* L.) in relation to stand density, bud flushing phenotype, and season. *Plant Protect Sci* 49:120–126. <https://doi.org/10.17221/70/2012-PPS>
- Brueggemann T, Fladung M, Schroeder H (2022) Flexible DNA isolation procedure for different tree species as a convenient lab routine. *Silvae Genetica* 71:20–30. <https://doi.org/10.2478/sg-2022-0003>
- Capo-chichi LJA, Elakhdar A, Kubo T, Nyachiro J, Juskiw P, Capettini F et al (2022) Genetic diversity and population structure assessment of Western Canadian barley cooperative trials. *Front Plant Sci* 13:1006719. <https://doi.org/10.3389/fpls.2022.1006719>
- Čepl J, Stejskal J, Lhotáková Z, Holá D, Korecký J, Lstibůrek M et al (2018) Heritable variation in needle spectral reflectance of Scots pine (*Pinus sylvestris* L.) peaks in red edge. *Remote Sens Environ* 219:89–98. <https://doi.org/10.1016/j.rse.2018.10.001>
- Coker TLR, Rozsypálek J, Edwards A, Harwood TP, Butfoy L, Buggs RJA (2019) Estimating mortality rates of European ash (*Fraxinus excelsior*) under the ash dieback (*Hymenoscyphus fraxineus*) epidemic. *Plants People Planet* 1:48–58. <https://doi.org/10.1002/ppp3.11>
- Core Team R (2022) R language and environment for statistical computing. R Foundation for statistical. Vienna, Austria
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA et al (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO et al (2021) Twelve years of SAMtools and BCFtools. *Gigascience*. <https://doi.org/10.1093/gigascience/giab008>
- El-Kassaby YA, Lstibůrek M, Liewlaksaneeyanawin C, Slavov GT, Howe GT eds. (2007) Breeding without breeding: approach, example, and proof of concept. In: Proceedings of the IUFRO division 2 joint conference: low input breeding and conservation of forest genetic resources. Antalya, Turkey
- El-Kassaby YA, Lstibůrek M (2009) Breeding without breeding. *Genet Res* 91:111–120. <https://doi.org/10.1017/S001667230900007X>
- Enderle R, Nakou A, Thomas K, Metzler B (2015) Susceptibility of autochthonous German *Fraxinus excelsior* clones to *Hymenoscyphus pseudoalbidus* is genetically determined. *Ann for Sci* 72:183–193. <https://doi.org/10.1007/s13595-014-0413-1>
- Evans MR (2019) Will natural resistance result in populations of ash trees remaining in British woodlands after a century of ash dieback disease? *R Soc Open Sci* 6:190908. <https://doi.org/10.1098/rsos.190908>
- Galili T (2015) dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* 31:3718–3720. <https://doi.org/10.1093/bioinformatics/btv428>
- García C, Guichoux E, Hampe A (2018) A comparative analysis between SNPs and SSRs to investigate genetic variation in a juniper species (*Juniperus phoenicea* ssp. turbinata). *Tree Genet Genomes* 14:1–9. <https://doi.org/10.1007/s11295-018-1301-x>
- Harper AL, McKinney LV, Nielsen LR, Havlickova L, Li Y, Trick M et al (2016) Molecular markers for tolerance of European ash (*Fraxinus excelsior*) to dieback disease identified using associative transcriptomics. *Sci Rep* 6:19335. <https://doi.org/10.1038/srep19335>
- Howe GT, Jayawickrama K, Kolpak SE, Kling J, Trappe M, Hipkins V et al (2020) An axiom SNP genotyping array for douglas-fir. *BMC Genomics* 21:9. <https://doi.org/10.1186/s12864-019-6383-9>
- Hu J-Y, Zhou Y, He F, Dong X, Liu L-Y, Coupland G et al (2014) miR824-regulated AGAMOUS-LIKE16 contributes to flowering time repression in Arabidopsis. *Plant Cell* 26:2024–2037. <https://doi.org/10.1105/tpc.114.124685>
- Husson C, Scala B, Caël O, Frey P, Feau N, Ioos R et al (2011) *Chalara fraxinea* is an invasive pathogen in France. *Eur J Plant Pathol* 130:311–324. <https://doi.org/10.1007/s10658-011-9755-9>
- Jiang K, Xie H, Liu T, Liu C, Huang S (2020) Genetic diversity and population structure in *Castanopsis fissa* revealed by analyses of sequence-related amplified polymorphism (SRAP) markers. *Tree Genet Genomes* 16:1–10. <https://doi.org/10.1007/s11295-020-01442-2>
- Jones OR, Wang J (2010) COLONY: a program for parentage and sibship inference from multilocus genotype data. *Mol Ecol Resour* 10:551–555. <https://doi.org/10.1111/j.1755-0998.2009.02787.x>
- Kemp F (2003) Modern applied statistics with S. *J Royal Stat Soc D* 52:704–705. https://doi.org/10.1046/j.1467-9884.2003.t01-19-00383_22.x
- Knaus BJ, Grünwald NJ (2017) VCFR: a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour* 17:44–53. <https://doi.org/10.1111/1755-0998.12549>
- Kowalski T (2006) *Chalara fraxinea* sp. nov. associated with dieback of ash (*Fraxinus excelsior*) in Poland. *For Path* 36:264–270. <https://doi.org/10.1111/j.1439-0329.2006.00453.x>
- Landolt J, Gross A, Holdenrieder O, Pautasso M (2016) Ash dieback due to *Hymenoscyphus fraxineus*: what can be learnt from evolutionary ecology? *Plant Pathol* 65:1056–1070. <https://doi.org/10.1111/ppa.12539>
- Lefort F, Brachet S, Frascaria-Lacoste N, Edwards KJ, Douglas GC (1999) Identification and characterization of microsatellite loci in ash (*Fraxinus excelsior* L.) and their conservation in the olive family (Oleaceae). *Mol Ecol* 8:1088–1089. https://doi.org/10.1046/j.1365-294X.1999.00655_8.x
- Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Lobo A, Hansen JK, McKinney LV, Nielsen LR, Kjær ED (2014) Genetic variation in dieback resistance: growth and survival of *Fraxinus excelsior* under the influence of *Hymenoscyphus pseudoalbidus*. *Scand J for Res* 29:519–526. <https://doi.org/10.1080/02827581.2014.950603>
- Lobo A, McKinney LV, Hansen JK, Kjær ED, Nielsen LR (2015) Genetic variation in dieback resistance in *Fraxinus excelsior*

- confirmed by progeny inoculation assay. *For Path* 45:379–387. <https://doi.org/10.1111/efp.12179>
- Lstibůrek M, Ivanková K, Kadlec J, Kobliha J, Klápště J, El-Kassaby YA (2011) Breeding without breeding: minimum fingerprinting effort with respect to the effective population size. *Tree Genet Genomes* 7:1069–1078. <https://doi.org/10.1007/s11295-011-0395-1>
- Lstibůrek M, Hodge GR, Lachout P (2015) Uncovering genetic information from commercial forest plantations—making up for lost time using “Breeding without Breeding.” *Tree Genet Genomes* 11:1–12. <https://doi.org/10.1007/s11295-015-0881-y>
- McKinney LV, Nielsen LR, Hansen JK, Kjær ED (2011) Presence of natural genetic resistance in *Fraxinus excelsior* (Oleraceae) to *Chalara fraxinea* (Ascomycota): an emerging infectious disease. *Heredity* 106:788–797. <https://doi.org/10.1038/hdy.2010.119>
- McKinney LV, Thomsen IM, Kjaer ED, Nielsen LR (2012) Genetic resistance to *Hymenoscyphus pseudoalbidus* limits fungal growth and symptom occurrence in *Fraxinus excelsior*. *For Path* 42:69–74. <https://doi.org/10.1111/j.1439-0329.2011.00725.x>
- McKinney LV, Nielsen LR, Collinge DB, Thomsen IM, Hansen JK, Kjær ED (2014) The ash dieback crisis: genetic variation in resistance can prove a long-term solution. *Plant Pathol* 63:485–499. <https://doi.org/10.1111/ppa.12196>
- Muñoz F, Marçais B, Dufour J, Dowkiw A (2016) Rising out of the ashes: additive genetic variation for crown and collar resistance to *Hymenoscyphus fraxineus* in *Fraxinus excelsior*. *Phytopathology* 106:1535–1543. <https://doi.org/10.1094/PHYTO-11-15-0284-R>
- Nielsen LR, McKinney LV, Kjær ED (2017) Host phenological stage potentially affects dieback severity after *Hymenoscyphus fraxineus* infection in *Fraxinus excelsior* seedlings. *Balt for* 23:229–232
- Pliura AL, Baliuckas VI (2007) Genetic variation in adaptive traits of progenies of Lithuanian and western European populations of *Fraxinus excelsior* L. *Balt for* 13:28–38
- Pliura A, Lygis V, Suchockas V, Bartkevicius E (2011) Performance of twenty four European *Fraxinus excelsior* populations in three Lithuanian progeny trials with a special emphasis on resistance to *Chalara Fraxinea*. *Balt for* 17:17–34
- Pliura A, Lygis V, Marčiulyniene D, Suchockas V, Bakys R (2016) Genetic variation of *Fraxinus excelsior* half-sib families in response to ash dieback disease following simulated spring frost and summer drought treatments. *iForest* 9:12–22. <https://doi.org/10.3832/ifer1514-008>
- Plumb WJ, Coker TLR, Stocks JJ, Woodcock P, Quine CP, Nemesio-Gorriz M et al (2020) The viability of a breeding programme for ash in the British Isles in the face of ash dieback. *Plants People Planet* 2:29–40. <https://doi.org/10.1002/ppp3.10060>
- Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, van der Auwera GA et al (2017) Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*. <https://doi.org/10.1101/201178>
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575. <https://doi.org/10.1086/519795>
- Ramesh P, Mallikarjuna G, Sameena S, Kumar A, Gurulakshmi K, Reddy BV et al (2020) Advancements in molecular marker technologies and their applications in diversity studies. *J Biosci* 45:1–15. <https://doi.org/10.1007/s12038-020-00089-4>
- Slavov GT, Howe GT, Adams WT (2005) Pollen contamination and mating patterns in a douglas-fir seed orchard as measured by simple sequence repeat markers. *Can J for Res* 35:1592–1603. <https://doi.org/10.1139/x05-082>
- Sollars ESA, Harper AL, Kelly LJ, Sambles CM, Ramirez-Gonzalez RH, Swarbreck D et al (2017) Genome sequence and genetic diversity of European ash trees. *Nature* 541:212–216. <https://doi.org/10.1038/nature20786>
- Stener L-G (2013) Clonal differences in susceptibility to the dieback of *Fraxinus excelsior* in southern Sweden. *Scand J for Res* 28:205–216. <https://doi.org/10.1080/02827581.2012.735699>
- van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor Appl Genet* 120:1289–1299. <https://doi.org/10.1007/s00122-009-1256-2>
- Venables WN, Ripley BD (2003) *Modern applied statistics with S*. Springer, New York
- Wickham H (2016) *ggplot2: elegant graphics for data analysis*. Springer, Cham
- Zavinon F, Adoukonou-Sagbadja H, Keilwagen J, Lehnert H, Ordon F, Perovic D (2020) Genetic diversity and population structure in Beninese pigeon pea [*Cajanus cajan* (L.) Huth] landraces collection revealed by SSR and genome wide SNP markers. *Genet Resour Crop Evol* 67:191–208. <https://doi.org/10.1007/s10722-019-00864-9>
- Zeng W, Su Y, Huang R, Hu D, Huang S, Zheng H (2023) Insight into the complex genetic relationship of chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook.) advanced parent trees based on SSR and SNP datasets. *Forests* 14:347. <https://doi.org/10.3390/f14020347>
- Zhao Y-J, Hosoya T, Baral H-O, Hosaka K, Kakishima M (2013) *Hymenoscyphus pseudoalbidus*, the correct name for *Lambertella albida* reported from Japan. *Mycotaxon* 122:25–41. <https://doi.org/10.5248/122.25>
- Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28:3326–3328. <https://doi.org/10.1093/bioinformatics/bts606>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.