Forest Research

Assessment of Genetic Diversity and Population Structure in Welsh and Scottish Aspen (*Populus tremula*) Using Microsatellite DNA Markers

A report prepared for the Welsh Government to inform native aspen conservation and restoration planning in Wales

Stuart A'Hara and Christopher Hirst

The Research Agency of the
Forestry Commission

Forest Research is the Research Agency of the Forestry Commission and is the leading UK organisation engaged in forestry and tree related research.

The Agency aims to support and enhance forestry and its role in sustainable development by providing innovative, high quality scientific research, data, technical support and consultancy services.

Ymchwil Coedwigaeth yw Asiantaeth Ymchwil y Comisiwn Coedwigaeth a dyma'r sefydliad mwyaf blaenllaw yn y DU sy'n ymwneud ag ymchwil ym meysydd coedwigaeth a choed.

Nod yr Asiantaeth yw cynorthwyo a gwella coedwigaeth a'i rôl mewn datblygu cynaliadwy trwy ddarparu ymchwil wyddonol o ansawdd uchel, cymorth technegol a gwasanaethau ymgynghori.

Table of Contents

Table of Figures

Table of Tables

1 Executive Summary

This report presents the results of genetic analyses conducted on native aspen (*Populus tremula*) sampled from multiple sites in Wales and Scotland. Although the samples were collected independently for conservation-focused projects, the datasets have been integrated to facilitate assessment of genetic diversity and population structure across a broader geographic range than would be possible within individual studies. This is the first study to investigate Welsh aspen genetic diversity.

A total of 19 sampling locations, comprising 11 sites in Wales and eight in Scotland, were analysed. Following quality control, two triploid individuals and three non-*P. tremula* individuals were excluded, leaving 237 diploid *P. tremula* individuals for analysis. Samples were genotyped using a panel of ten microsatellite markers to quantify genetic diversity, highlight instances of clonal replication and investigate patterns of population structure. Two sampling locations with insufficient sample sizes were excluded from population-level analyses (Denbighshire, n = 1 and Knoydart, n = 2), giving a final analysis dataset of 190 individuals across 17 populations. Full methodology and results are presented in the main report.

Genetic diversity

Levels of genetic diversity were broadly comparable between Welsh and Scottish populations. Welsh populations exhibited marginally higher diversity across most measures; however, the observed differences were small. Observed heterozygosity indicates medium to high genetic diversity in Welsh and Scottish aspen. No consistent country-level pattern was evident in the distribution of private alleles among populations.

Clonal structure

A proportion of samples shared identical multilocus genotypes with other individuals from the same site, consistent with the well-established capacity of aspen to reproduce vegetatively through root suckering. No identical genotypes were detected across different sampling locations. All population genetic analyses were performed using a clone-corrected dataset, in which only a single representative of each unique genotype was retained.

Population structure

Statistically significant genetic differentiation was detected among sampling locations, indicating that populations are not genetically homogeneous. However, this differentiation was not associated with national origin. Welsh and Scottish populations could not be reliably distinguished on the basis of genetic data and

genetic variation between countries was not statistically significant. In addition, no evidence of isolation by distance was observed, as genetic differentiation did not increase with geographic separation across the sampled range.

Non-*P. tremula* individuals

Three samples were identified as genetically distinct from *P. tremula*, based on the presence of diagnostic alleles and their separation from the main dataset in multivariate analyses. These findings were consistent with field assessments indicating potential hybrid or non-native origin of these particular trees. Accordingly, these samples were excluded from the primary analyses. Within the remaining dataset, no evidence of introgression from non-*P. tremula* taxa was detected using the microsatellite marker panel employed in this study.

Individual genetic distance matrix

A pairwise genetic distance matrix was calculated across all 190 clone-corrected individuals and is provided as a supplementary deliverable alongside a population key.

2 Introduction

European aspen (*Populus tremula L.*) is one of the most widely distributed tree species in the Northern Hemisphere, occurring from western Europe to eastern Asia. As a pioneer species, it plays an important ecological role in woodland succession, ecosystem functioning, and biodiversity conservation. Aspen supports a rich assemblage of associated plants, fungi, lichens and invertebrates, many of which exhibit strong ecological associations with the species and may be vulnerable to its decline (von Wühlisch, 2009). *P. tremula* is a dioecious species with separate male and female trees with wind-dispersal of pollen and seeds (Easton, 1997).

Within the United Kingdom, *P. tremula* occurs throughout much of Wales, Scotland and parts of England, typically as a minor component of native woodland systems. Despite its wide distribution, many British populations are characterised by low stand densities, fragmented distributions and extensive clonal reproduction through root suckering. Sexual reproduction is thought to be infrequent in many populations, raising concerns regarding genetic connectivity, effective population size and long-

term adaptive potential (Worrell, 1995). These characteristics make *P. tremula* an important candidate for genetic assessment, particularly where management decisions require an understanding of population connectivity and diversity.

Genetic diversity is increasingly recognised as a fundamental component of biodiversity conservation. Genetic variation within and among populations underpins the capacity of species to adapt to environmental change such as emerging pests and pathogens, increased habitat fragmentation and climate change. Consequently, the conservation of forest genetic resources has become an important component of UK forestry and biodiversity policy. The UK Forest Genetic Resources Strategy identifies the protection, understanding and sustainable use of genetic diversity as a key requirement for maintaining resilient woodland ecosystems and supporting future woodland expansion (Trivedi *et al.,* 2019).

The importance of genetic evidence in informing policy has also been recognised within the UK Forestry Standard, which emphasises the need to maintain biological diversity and resilience in woodland management and establishment. Decisions relating to seed sourcing, restoration planting, translocation and the conservation of small or isolated populations increasingly require an understanding of patterns of genetic diversity and population structure. Genetic studies therefore provide a critical evidence base for balancing the maintenance of local adaptation with the need to preserve genetic diversity and facilitate adaptation to future environmental conditions.

Previous work on Scottish aspen populations demonstrated that, despite concerns surrounding extensive clonal reproduction, native populations retained substantial levels of genetic variation and showed only modest levels of population differentiation (Easton, 1997). These findings highlighted the importance of genetic studies for understanding the conservation status of British aspen and challenged assumptions that fragmented or predominantly clonal populations necessarily exhibit low diversity. However, much of the existing research has focused on Scotland, and relatively little

information has been available regarding patterns of genetic diversity across the wider range of the species in Britain. Advances in molecular genetic techniques, particularly the use of microsatellite markers, have greatly improved the ability to characterise clonal structure, assess genetic diversity and quantify population differentiation in *P. tremula*. Microsatellites, or simple sequence repeats (SSRs), are tandemly repeated DNA sequences consisting of short motifs ranging from 1 to 6 base pairs (bp) in length. Owing to their relatively high mutation rate, the number of repeat units can vary considerably among individuals, resulting in high levels of polymorphism. As codominant markers that frequently exhibit multiple alleles per locus, microsatellites provide substantial discriminatory power and can generate unique multilocus genotypes, facilitating accurate individual identification when an adequate number of loci are employed. Such approaches have been widely applied across Europe and have proven effective for distinguishing genetically unique individuals, identifying levels of vegetative reproduction and informing conservation management strategies of fragmented populations (e.g. Politov *et al.,* 2015). A recent paper (Provan *et al.,* 2026) has used a microsatellite marker approach to examine genetic diversity in populations from mainland Britain, Sweden and Italy and provides context for the present study which will include the analysis of Welsh aspen data for the first time. However, much of the existing research in Britain has focused on Scotland, and no genetic data have previously been available for Welsh populations specifically.

This report presents the first genetic assessment of Welsh *P. tremula* populations using microsatellite markers, incorporating Scottish populations as a reference dataset to contextualise Welsh genetic diversity within a broader British context. Specifically, we characterise clonal structure, genetic diversity, population differentiation and individual relatedness across 10 Welsh and 7 Scottish sampling locations. The resulting evidence contributes directly to ongoing efforts to conserve UK forest genetic resources and provides information relevant to woodland

restoration, genetic conservation planning, and evidence-based policy development for native aspen management in Britain.

3 Sampling and laboratory methods

3.1 Sampling design and study area

The *Populus tremula* samples in this study were collected from sites across Wales and Scotland between 2023 and 2025 with the exception of the Scottish Galloway samples, which were collected in 2014. In total, 242 samples were collected across 19 locations. The final dataset comprised of 153 samples from 11 Welsh locations and 84 samples from eight Scottish locations (Figure 1, Table 1). The location of each sample was captured by either a grid reference or GPS coordinates. If the exact location was not known, an approximation was used.

Samples from different sites were collected independently as part of several standalone projects and there was no overarching sampling design. The opportunistic nature of the sampling plan will be discussed later in the Limitations section of the report. Mar Lodge Estate samples comprised 96 individuals which resolved to 13 unique multilocus genotypes; only these 13 genotypes were taken forward for analysis (see Ramsden 2024 for further details).

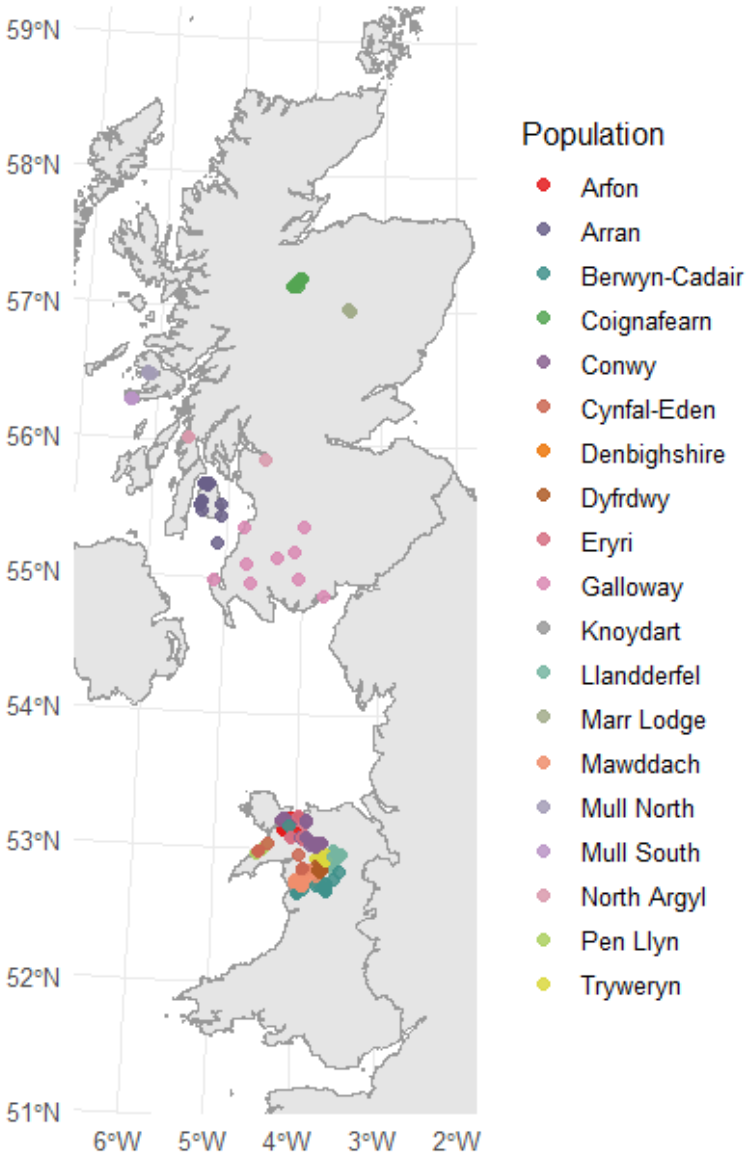

Figure 1 Sampling locations for *Populus tremula* individuals genotyped in this study. Points represent individual sampled trees coloured by population. Welsh populations are clustered in northwest Wales; Scottish populations span sites from Galloway in the southwest to Coignafearn in the Cairngorms. Knoydart and Denbighshire are indicated but were excluded from population-level analyses due to small sample sizes (n = 2 and n = 1 respectively).

Table 1 Number of individuals sampled per population and year.

Country	Population	2014	2023	2024	2025
Scotland	Arran	0	0	15	0
Scotland	Coignafearn	0	0	26	0
Scotland	Galloway	14	0	0	0
Scotland	Knoydart	0	0	2	0
Scotland	Mar Lodge †	0	0	13	0
Scotland	Mull North	0	0	0	5
Scotland	Mull South	0	0	0	6
Scotland	North Argyll	0	0	0	3
Wales	Arfon	0	4	11	4
Wales	Berwyn-Cadair	0	13	0	3
Wales	Conwy	0	15	2	6
Wales	Cynfal-Eden	0	3	3	5
Wales	Denbighshire	0	0	0	1
Wales	Dyfrdwy	0	10	1	4
Wales	Eryri	0	6	10	4
Wales	Llandderfel	0	12	0	2
Wales	Mawddach	0	12	7	1
Wales	Pen Llŷn	0	3	4	0
Wales	Tryweryn	0	5	0	2

† Mar Lodge Estate: 96 samples were collected and genotyped, resolving to 13 unique multilocus genotypes. Only these 13 genotypes were taken forward for analysis (see Ramsden 2024).

3.2 Sample collection and DNA extraction - Field collection methods, tissue type, storage, DNA extraction protocol

DNA was extracted from leaf tissue or leaf-bud material. Where leaves had flushed, an approximately 1 cm diameter disc was taken using the lid of a 2 ml microfuge tube; where leaves had not yet flushed, one or two buds were used with leaf scales removed using a scalpel blade. Tissue was homogenised by placing the sample in a 2 ml microfuge tube with two 3 mm steel ball bearings, freezing in liquid nitrogen,

and shaking at 25 beats sec^{-1} for 60 seconds using a Retsch MM300 mixer mill (Retsch GmbH, Germany).

DNA was extracted using a DNeasy Plant Pro Kit (Qiagen, Germany) following the manufacturer's protocol with minor modifications. Briefly, 500 µl of solution CD1 and 60 µl of solution PS were added to each homogenised sample and vortexed to facilitate cell lysis. Samples were centrifuged at 12,000 × g for 2 minutes, and the supernatant transferred to fresh 1.5 ml microcentrifuge tubes. 200 µl of solution CD2 was added and samples were centrifuged again at 12,000 × g for 1 minute. The supernatant was mixed with 500 µl of AW1 buffer, vortexed, and loaded onto an MB Spin Column in two 600 µl aliquots, each centrifuged at 12,000 × g for 1 minute. Columns were washed sequentially with 650 µl AW1 buffer and 650 µl AW2 buffer, followed by a final centrifugation at 16,000 × g for 2 minutes to remove residual wash buffer. DNA was eluted in 85 µl pre-warmed (65°C) elution buffer, reapplied to the column once to maximise yield, and stored at -20°C until required. Extraction success was confirmed by electrophoresis of a 5 µl aliquot on a 1.2% agarose gel. All extracted DNA is securely archived and available for future analyses.

3.3 Microsatellite Genotyping

DNA samples were genotyped using 11 microsatellite markers run as two multiplex mixes: PMGC2550 (Maksimovic *et al.,* 2014), PTR04 (Rahman *et al.,* (2000) PMS14, PMS15, PMS16, PMS18, PMS19, PMS20 (Smulders *et al.,* 2001), PMS05 (Van der Schoot *et al.*, 2000), and Asp302 and Asp322 (De Carvalho *et al.*, 2010). Primer sequences and multiplex compositions are provided in the Appendix.

Primers (Applied Biosystems, UK) were dissolved in 200 µl Tris-EDTA buffer (Sigma G3283) and incubated in the dark for 20 minutes. Multiplex primer mixes were prepared by combining 20 µl of each primer with Tris-EDTA buffer to a total volume of 500 µl. PCR amplifications were performed in 96-well plates using a Qiagen Multiplex PCR Kit. Each 50 µl reaction contained 48.2 µl Qiagen multiplex mix and 1.8 µl template DNA. Thermocycling conditions were as follows: initial denaturation

at 95°C for 15 minutes; 35 cycles of denaturation at 94°C for 30 seconds, annealing at 57°C for 90 seconds, and extension at 72°C for 90 seconds; followed by a final extension at 60°C for 30 minutes and a hold at 10°C.

Amplified products (15 µl) were submitted to the Dundee University Sequencing Service for fragment analysis on an ABI3500 capillary electrophoresis sequencer. Microsatellite lengths were determined and genotypes scored using Geneious Prime software (Figure 2). Electropherogram peaks were assessed visually and any inconsistencies or miscalls were manually corrected.

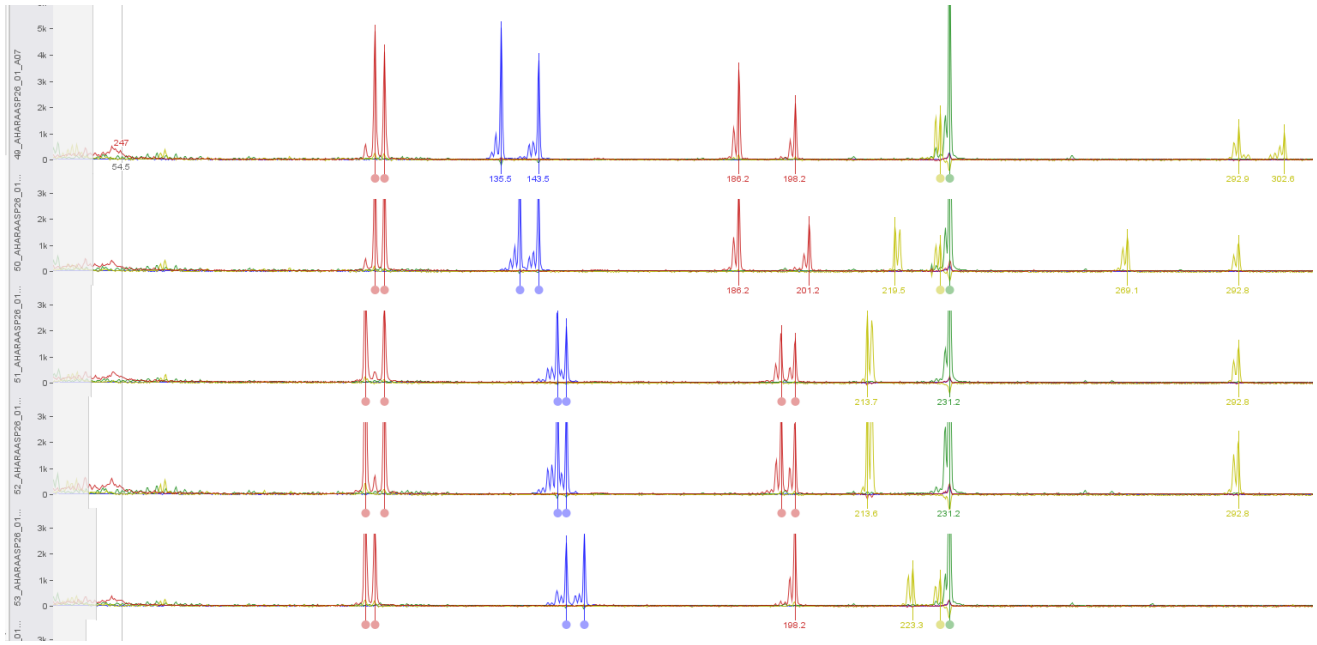

Figure 2 Example of multiplex PCR electropherogram results from five *Populus tremula* individuals across six microsatellite loci. Each row represents one individual; coloured peaks correspond to different fluorescent dye labels. The peaks visible on the far right of each panel correspond to locus PMS05, which was subsequently excluded from all analyses due to evidence of null alleles.

3.4 Sex Determination Assay

To support clonal analysis and generate information on any possible population-level gender bias, sex determination was performed following Pakull *et al.*, (2014), in which sex is inferred by the presence (male) or absence (female) of a PCR amplification product from the TOZ19 gene, run in conjunction with a control fragment that amplifies in both sexes to confirm reaction success.

A master mix for 100 reactions was prepared containing 1450 µl distilled water, 200 µl Tris-EDTA buffer (Sigma G3283), 160 µl dNTPs, 100 µl each of forward and reverse primers for both the TOZ19 locus and control fragment, and 10 µl HotStar Taq DNA

Polymerase (Bioron, Germany). Each reaction comprised 18.5 µl master mix and 1.5 µl template genomic DNA.

Amplification was carried out using a Techne Flexigene thermocycler under the following conditions as per the paper: initial denaturation at 95°C for 4 minutes; 38 cycles of 94°C denaturation for 45 seconds, 50°C annealing for 60 seconds, and 72°C extension for 60 seconds; with a final extension at 72°C for 7 minutes, followed by a hold at 10°C.

PCR products were separated by electrophoresis on a 1.5% agarose gel prepared in 1× TBE buffer with GelRed nucleic acid stain (5 µl) added prior to pouring. Electrophoresis was performed at 120 V for 30 minutes and bands were visualised using an iBright 750 imaging system (Thermo Fisher Scientific) (Figure 3).

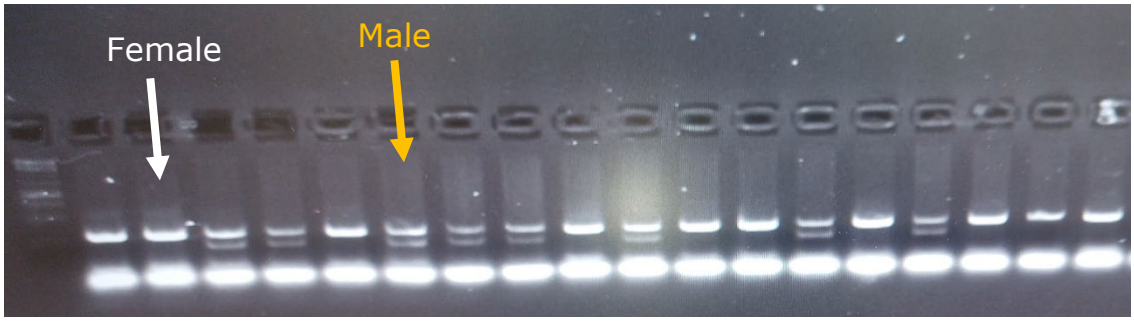

Figure 3 Example agarose gel showing results of the sex determination assay following Pakull *et al.*, (2014). A single band indicates a female individual; two bands indicate a male individual. Bright bands at the bottom of the gel are primer dimers. Selected lanes are annotated to illustrate the scoring criteria.

3.5 Statistical analysis

All analyses were conducted in R version 4.5.2 (R Core Team 2025) using the following packages: adegenet v2.1.11 (Jombart, 2008; Jombart & Ahmed, 2011), poppr v2.9.8 (Kamvar *et al.*, 2014; 2015), hierfstat v0.5.11 (Goudet, 2005), pegas

v1.4 (Paradis, 2010), ape v5.8 (Paradis & Schliep, 2019), vegan v2.6.8 (Oksanen *et al.*, 2024), sf v1.0.23 (Pebesma, 2018) and tidyverse v2.0.0 (Wickham *et al.*, 2019). Specific analytical methods are described in the relevant results sections

4 Data Preparation and Quality Control

4.1 Missing Data

A total of 242 individuals were genotyped using 11 microsatellite loci. Prior to the final analysis, missing data were assessed at both the locus and population level. Ten of the eleven loci showed no missing data. Locus PMS19 had a missing data rate of 0.42% across all individuals. Locus PMS05 showed the highest missing data rate at 11.39%, with both allele columns affected equally.

At the population level, overall missing data rates were low. Mean missing rate per population ranged from 0% (Arran, Coignafearn, Dyfrdwy, Mar Lodge, Mull South) to 3.9% (Tryweryn). Maximum individual-level missing rate across all populations was 9.1%, corresponding to a single missing locus (PMS05) in affected individuals.

4.2 Locus Quality – Hardy-Weinberg Equilibrium Testing

All 11 loci were tested for deviation from Hardy-Weinberg Equilibrium (HWE) using exact tests with 1000 Monte Carlo permutations, applied to the clone-corrected dataset. Ten of eleven loci showed no significant deviation from HWE expectations ($p > 0.05$; Table 2). Locus NED PMS05 showed a highly significant deviation ($\chi^2 = 683.16$, $df = 28$, $p < 0.001$), with observed heterozygosity ($Ho = 0.26$) substantially lower than expected ($He = 0.77$). This pattern, combined with the high missing data rate at this locus (11.39%) and discrepancies in clone assignment (see Section 5.1), is consistent with the presence of null alleles - alleles that consistently fail to amplify, or cause true heterozygotes to be scored as homozygotes. Locus NED PMS05 was

therefore excluded from all subsequent analyses. All remaining analyses are based on the ten-locus dataset.

Table 2 Hardy-Weinberg Equilibrium (HWE) test results for 11 microsatellite loci genotyped in *Populus tremula* (n = 237, clone-corrected). Observed (Ho) and expected (He) heterozygosity are shown per locus. P-values are from exact tests with 1000 Monte Carlo permutations.

Locus	Ho	He	p
6- FAM Asp322	0.795	0.785	0.314
VIC PMS14	0.487	0.537	0.272
NED PMS18	0.672	0.75	0.174
NED PMS05	0.25	0.77	< 0.001 ***
PET PTR04	0.574	0.557	0.944
PET PMS15	0.718	0.724	0.643
6-FAM 2550	0.744	0.747	0.147
6-FAM PMS19	0.753	0.785	0.651
NED Asp02	0.8	0.799	0.668
PET PMS16	0.738	0.792	0.063
PET PMS20	0.421	0.454	0.494

4.3 Non-*P. tremula* Individuals

Three individuals (J2426, Jen2524 and RHYD2406) were identified as potentially non-*P. tremula* based on field annotations, genotypic screening and multivariate analysis, and were excluded from all subsequent analyses. These individuals are retained in a separate dataset for reference.

4.4 Triploid Individuals

Aspen is a diploid species (2n =38) which means that there are in two copies of each allele at a given locus. Triploid aspen was first documented by Muntzing (1936) and reported to be a vigorous, large clone with large leaves. Two individuals in this study (AR2512, North Argyll and CL2, Coignafearn) were identified as putative triploid individuals based on the presence of three alleles at one or more loci and excluded from further analyses. They are retained in a separate dataset for reference. Provan

et al., (2026) similarly reported low levels of triploid clone occurrence in Irish samples with two occurrences - including the largest clone in their study. There is significant interest in the micropropagation of triploid aspen clones due to their potential fast growth and timber quality properties (P. Livingstone, pers. comm.). Notably, DNA fingerprinting examines only a limited number of loci, and the alleles detected at these sites may, by chance, display a monomorphic or diploid profile, even in a truly triploid individual. Consequently, microsatellite marker-based genotyping cannot unequivocally establish ploidy. Flow cytometry, which measures nuclear DNA content relative to a known diploid control or internal standard, remains the most robust and reliable approach for confirming ploidy status.

4.5 Clonal Structure

4.5.1 Clone detection and correction.

A total of 237 diploid *P. tremula* individuals were taken forward for genotyping across 19 sampling locations in Wales (n = 153) and Scotland (n = 84), following pre-filtering of Mar Lodge Estate samples to 13 unique multilocus genotypes from 96 collected samples (see Section 3.1). Multilocus genotype (MLG) analysis using ten microsatellite loci (PMS05 excluded) identified 192 unique genotypes. Seventy-five individuals (31.6%) were members of clonal groups - genets represented by two or more sampled ramets. Following clone correction, 45 clonal ramets were removed, retaining one representative per unique genotype per population. Clonal replicates were more prevalent in Welsh populations (41 redundant ramets removed, 26.8% of Welsh samples) than in the Scottish populations included in this analysis (4 redundant ramets removed, 4.8% of Scottish samples). However, this comparison excludes the pre-filtering of Mar Lodge Estate samples and does not reflect an unbiased estimate of clonality across Scottish populations.

Two populations with insufficient sample sizes for population-level analyses were excluded following clone correction - Denbighshire (n = 1, Wales) and Knoydart (n = 2, Scotland) - giving a final analysis dataset of 190 individuals across 17 populations.

No multilocus genotypes were shared across different sampling locations, indicating that all detected clonal groups were confined to single sites.

Clone assignments for Welsh samples were cross-referenced against field-based clone identifications provided by J. Wong (pers. comm.). Five discrepancies were identified where field records assigned individuals to the same clone, but molecular analysis recovered different MLGs (Table 3). Two of these discrepancies (Clone codes 29 and 93) were resolved following exclusion of locus NED PMS05, with individuals converging to single MLGs consistent with field assignments. The remaining three discrepancies (Clone codes 66, 80 and 86) persisted after PMS05 exclusion. Of these, Clone code 80 (individuals J2415 and J2416, Pen Llŷn) showed differences at three loci (NED PMS18, NED PMS05, PET PTR04) and may represent genuinely distinct genets; the remaining two differed at single loci consistent with somatic mutation or scoring error. A somatic mutation is a replication-induced gain or loss of repeat units at a microsatellite locus that arises in a body cell which can become fixed in a branch, root sucker, or sector of the clone and be propagated through vegetative growth.

Following clone correction, one representative per MLG per population was retained for all subsequent analyses, giving a final clone-corrected dataset of 190 individuals across 17 populations (Knoydart n=2 and Denbighshire n=1 excluded from population-level analyses).

Table 3 Discrepancies between molecular multilocus genotype (MLG) assignments and field-based clone identifications. Pairs of individuals assigned to the same field clone code but recovering different MLGs in molecular analysis are shown. Loci differing refers to the number of loci at which allele profiles differed between paired individuals. Clone codes 29 and 93 were resolved following PMS05 exclusion (see Section 4.2).

Clone code	Individual 1	Individual 2	Loci differing	Loci affected
29	J2406	J2408	0	None – confirmed clone
29	J2406	JEN05	1	NED PMS05
29	J2406	Jen2506	1	NED PMS05
93	RHYD2401	RHYD2402	0	None – confirmed clone
93	RHYD2401	RHYD2403	1	NED PMS05
66	J2421	J2422	1	NED PMS18
80	J2415	J2416	3	NED PMS18, NED PMS05, PET PTR04
86	J2423	J2424	1	PET PMS15

4.5.2 Spatial extent of clonal genets

The spatial extent of clonal genets was assessed for Welsh populations using convex hull analysis of georeferenced individuals sharing identical MLGs (Figure 4). A total of 26 multi-ramet clonal groups were identified in Wales. The largest clone by geographic extent was Clone code 9 (Conwy, MLG 26) comprising four ramets spanning a maximum distance of 30,746 m and a convex hull area of 3,304 ha. The second largest by area was Clone code 48 (Mawddach, MLG 91) with six ramets spanning 15,349 m and 3,624 ha. The majority of clonal groups (22 of 26) comprised only two ramets. It should be noted that the spatial extent of clonal genets is dependent on sampling design - where sampling deliberately targeted multiple ramets within known clonal patches, patch size reflects field knowledge rather than an unbiased estimate of natural clone size. Furthermore, where individuals sharing identical MLGs are separated by considerable distances, human-mediated movement of plant material cannot be excluded as an explanation for their shared genotype. In such cases, spatial extent reflects the history of planting activity as much as natural

clonal spread, and local site knowledge will be essential for interpreting individual clone extents.

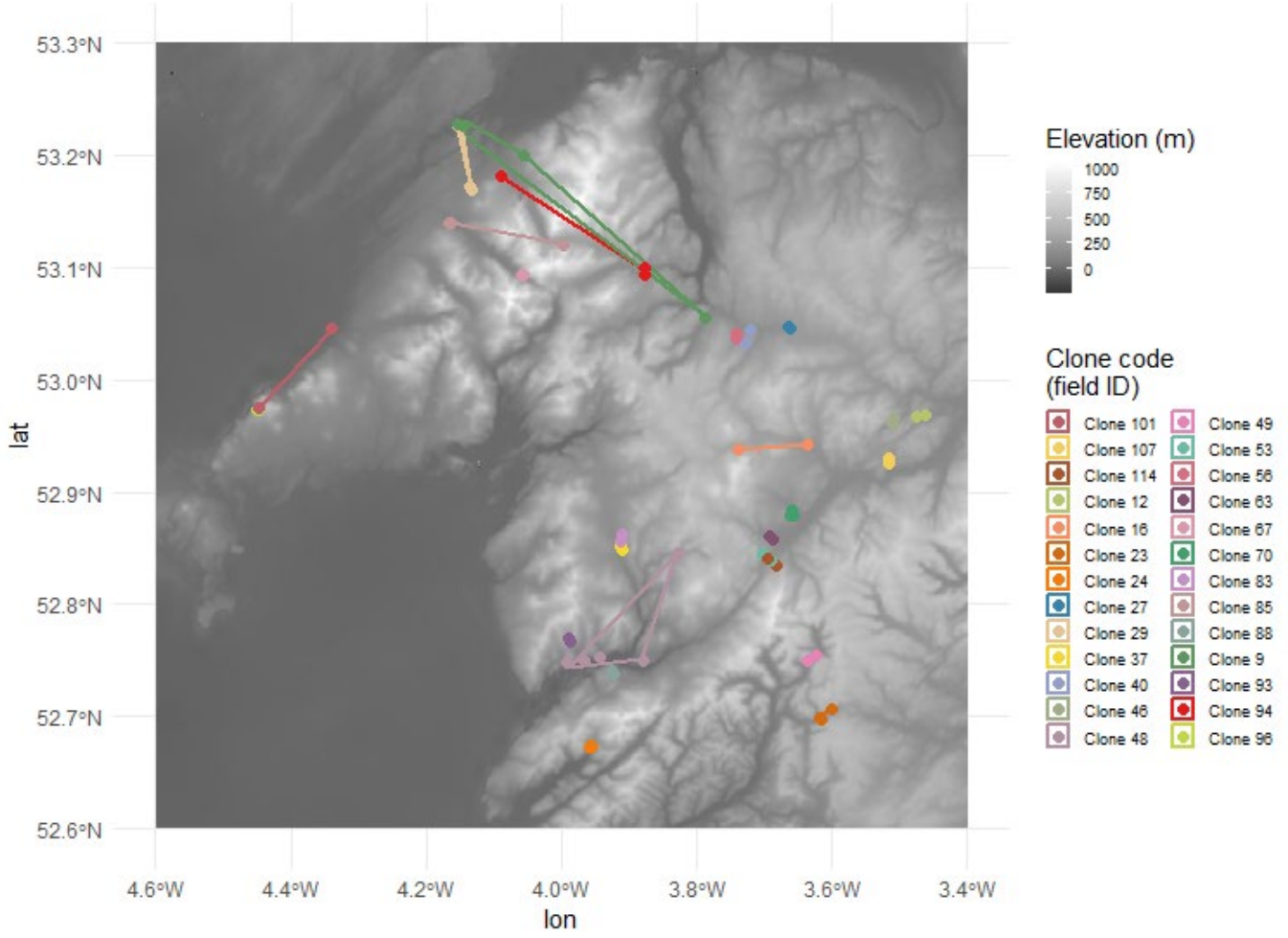

Figure 4 Spatial extent of clonal genets in Welsh *Populus tremula* populations. Lines and polygons show convex hulls connecting ramets sharing identical multilocus genotypes. Points represent individual sampled trees. Background shows topographic elevation (m). Only clonal groups comprising two or more individuals are shown.

4.6 Sex ratio

Sex was successfully determined for 150 of the 153 Welsh individuals analysed; sex could not be assigned to three individuals due to PCR amplification failure. Among the successfully genotyped samples, 82 individuals were identified as male (54.7%)

and 68 as female (45.3%), indicating a broadly balanced sex ratio with a slight male bias (Table 4).

All 26 multi-ramet clonal groups identified within the Welsh dataset comprised individuals of a single sex. This finding is consistent with vegetative reproduction via root suckering, whereby genetically identical ramets arise from a common parent and therefore share the same sex.

Table 4 Sex ratio per Welsh *Populus tremula* sub-population. Individuals with unknown sex (n = 3) are excluded from proportional calculations. Female and male counts include clonal ramets - individuals sharing identical multilocus genotypes are not removed prior to sex ratio calculation.

Location	Female	Male	Total	Prop Male	Prop Female
Cynfal-Eden	3	7	10	0.7	0.3
Mawddach	6	14	20	0.7	0.3
Conwy	7	14	21	0.667	0.333
Dyfrdwy	6	9	15	0.6	0.4
Eryri	8	12	20	0.6	0.4
Berwyn-Cadair	6	8	14	0.571	0.429
Arfon	9	10	19	0.526	0.474
Llandderfel	8	5	13	0.385	0.615
Tryweryn	5	2	7	0.286	0.714
Pen Llŷn	6	1	7	0.143	0.857
Denbighshire	1	0	1	0	1

4.7 Genetic Diversity

4.7.1 Allelic diversity

Across the ten microsatellite loci retained for analysis, the number of alleles per locus ranged from three (PTR04) to sixteen (PMS19), with a mean of 8.6 alleles per locus (Table 5). PMS19 and PMS14 were the most polymorphic loci with 16 and 12 alleles respectively. PTR04 was the least polymorphic with three alleles across the full dataset.

Table 5 Number of alleles observed heterozygosity (Ho) and expected heterozygosity (He) per microsatellite locus in the clone-corrected *P. tremula* dataset (n = 190). Loci ordered by number of alleles. PMS05 excluded due to null alleles (see Section 4.2).

Locus	N alleles	Ho	He
PMS19	16	0.757	0.787
PMS14	12	0.489	0.535
L2550	11	0.742	0.746
Asp322	9	0.8	0.785
PMS18	8	0.663	0.751
PMS16	8	0.737	0.796
PMS15	7	0.711	0.721
Asp02	7	0.805	0.799
PMS20	5	0.421	0.452
PTR04	3	0.579	0.557

4.7.2 Heterozygosity

Expected heterozygosity (He) measures the probability that two randomly drawn alleles at a locus are different and provides an estimate of genetic diversity independent of sample size. Observed heterozygosity (Ho) measures the proportion of individuals that are heterozygous at a given locus. Comparison of Ho and He provides information on whether a population is mating randomly - under Hardy-Weinberg equilibrium Ho and He are expected to be approximately equal, with consistent deficits of observed heterozygosity suggesting inbreeding or the presence of null alleles, and consistent excesses suggesting scoring errors or balancing selection.

Mean observed heterozygosity (Ho) across all loci was 0.670 and mean expected heterozygosity (He) was 0.693 in the clone-corrected dataset, indicating a slight overall deficit of observed heterozygosity. At the population level, Ho ranged from 0.517 (Mull South) to 0.780 (Mull North) and He ranged from 0.630 (Mull South) to 0.771 (Mull North) across all populations (Table 6). Welsh populations showed a narrower range of He (0.665–0.726) than Scottish populations (0.630–0.771), with

Welsh populations showing consistently moderate diversity across all sub-populations. At the country level, Wales showed higher mean Ho (0.686) and mean He (0.705) than Scotland (Ho = 0.648, He = 0.677) after clone correction and rarefaction to equal sample sizes (n = 79 per country), indicating marginally greater allelic diversity in Welsh populations.

The inbreeding coefficient (Fis) was calculated per population as 1 - (Ho/He), providing a measure of the deficit or excess of observed heterozygotes relative to Hardy-Weinberg expectation (Table 6). Fis values were close to zero across most populations, consistent with random mating following clone correction. Mull South showed the highest positive Fis value (0.179), indicating a notable deficit of observed heterozygotes relative to expectation - consistent with the low diversity and elevated differentiation observed at this population across multiple metrics. Several populations showed slightly negative Fis values (Eryri -0.047, Mawddach -0.062, Cynfal-Eden -0.009, Tryweryn -0.033, Mull North -0.012, Mar Lodge -0.004), indicating a marginal excess of heterozygotes, which is not uncommon in outbreeding populations.

Table 6 Genetic diversity statistics per population in the clone-corrected *P. tremula* dataset. Genotypes = total individuals sampled; MLGs = unique multilocus genotypes retained after clone correction; Prop clonal = proportion of sampled individuals sharing a genotype with at least one other individual; He = expected heterozygosity (Nei's gene diversity);Ho = observed heterozygosity; Allelic richness = rarefied allelic richness standardised to n = 5 (North Argyll excluded, n = 3); Private alleles = number of alleles unique to each population; rbarD = index of association after clone correction; Fis = inbreeding coefficient, calculated as 1 − (Ho/He), where positive values indicate a deficit of observed heterozygotes relative to Hardy-Weinberg expectation and negative values indicate an excess.

Country	Population	Genotypes	MLGs	Prop clonal	He	Ho	Allelic richness	Private alleles	rbarD	Fis
Scotland	Arran	15	15	0	0.658	0.653	2.87	0	0.029	0.008
Scotland	Coignafearn	26	24	0.077	0.677	0.633	2.89	1	0.036	0.065
Scotland	Galloway	14	13	0.071	0.659	0.631	2.84	0	-0.011	0.043
Scotland	Mar Lodge	96†	13	0	0.682	0.685	2.93	2	-0.001	-0.004
Scotland	Mull North	5	5	0	0.771	0.780	3.30	1	0.029	-0.012
Scotland	Mull South	6	6	0	0.630	0.517	2.70	0	-0.071	0.179
Scotland	North Argyll	3	3	0	0.760	0.700	NA	1	-0.012	0.079
Wales	Arfon	19	14	0.263	0.695	0.657	2.98	2	-0.025	0.055
Wales	Berwyn-Cadair	16	11	0.312	0.717	0.691	3.04	1	-0.005	0.036
Wales	Conwy	23	15	0.348	0.715	0.687	3.05	0	0.008	0.039
Wales	Cynfal-Eden	11	7	0.364	0.665	0.671	2.87	1	-0.055	-0.009
Wales	Dyfrdwy	15	9	0.4	0.706	0.656	2.96	0	-0.029	0.071
Wales	Eryri	20	19	0.05	0.677	0.709	2.87	1	0.037	-0.047
Wales	Llandderfel	14	11	0.214	0.697	0.645	2.92	0	-0.025	0.075
Wales	Mawddach	20	12	0.4	0.690	0.733	2.97	1	-0.017	-0.062
Wales	Pen Llŷn	7	7	0	0.692	0.657	2.92	1	0.066	0.051
Wales	Tryweryn	7	6	0.143	0.726	0.750	3.19	0	0.036	-0.033

† 96 samples genotyped; 13 unique MLGs identified and taken forward for analysis (see Section 3.1 and Ramsden 2024).

4.7.3 Allelic richness

Rarefied allelic richness, standardised to n = 5 (Mull North, the smallest eligible population), ranged from 2.704 (Mull South) to 3.297 (Mull North). Welsh populations ranged from 2.871 (Eryri) to 3.193 (Tryweryn). Scottish mainland populations ranged from 2.841 (Galloway) to 2.927 (Mar Lodge). Mull South had the lowest rarefied allelic richness of any population. North Argyll was excluded from the rarefied allelic richness analysis due to insufficient sample size (n = 3).

Overall, rarefied allelic richness was broadly comparable across Welsh and Scottish populations. The absence of a clear country-level difference in allelic richness contrasts slightly with the marginally higher He observed in Welsh populations, suggesting the diversity difference between countries is driven by allele frequency distributions rather than the number of distinct alleles present.

4.7.4 Private alleles

A total of 12 private alleles were identified across 10 populations in the clone-corrected dataset. Private alleles were absent from seven populations (Llandderfel, Dyfrdwy, Tryweryn, Conwy, Mull South, Galloway and Arran). Arfon and Mar Lodge each had the highest number of private alleles (n = 2). Single private alleles were recorded in Cynfal-Eden, Berwyn-Cadair, Mawddach, Eryri, Pen Llŷn , Mull North, North Argyll and Coignafearn. Raw private allele counts should be interpreted with caution given the unequal sample sizes across populations - sampling larger population grants a greater probability of capturing rare alleles. Private alleles were

distributed across both Welsh and Scottish populations with no country-level pattern, suggesting that unique allelic diversity is not concentrated in either country.

4.7.5 Linkage disequilibrium

The standardised index of association (rbarD) measures the degree to which alleles at different loci are non-randomly associated across individuals in a population. In a randomly mating sexual population, alleles at different loci are expected to be independent of each other, giving rbarD values near zero. Positive rbarD values indicate that certain allele combinations occur together more often than expected by chance, which can reflect clonal reproduction, inbreeding, or small effective population size. In clonal organisms such as aspen, elevated rbarD in uncorrected datasets is expected because clonal individuals share entire multilocus genotypes, artificially inflating the association between alleles at different loci.

After clone correction, rbarD values were low across all populations, ranging from -0.071 (Mull South) to 0.066 (Pen Llŷn), with all values close to zero. No population showed substantial positive rbarD after clone correction, indicating that the weak positive LD signal present in the uncorrected dataset was attributable to the inclusion of clonal individuals rather than genuine non-random allelic associations. This pattern is consistent with Fis scores, indicating predominantly sexual reproduction in the sampled populations following removal of clonal ramets.

4.8 Population Structure

4.8.1 Pairwise FST

Genetic differentiation between populations was quantified using pairwise FST (Weir and Cockerham 1984), calculated on the clone-corrected ten-locus dataset. FST measures the proportion of total genetic variation attributable to differences between populations, with values ranging from 0 (no differentiation) to 1 (complete differentiation).

Pairwise FST values ranged from 0.025 to 0.160 across all population pairs (Figure 5). Mean within-Wales FST was 0.058 and mean within-Scotland FST was 0.027, indicating slightly greater differentiation among Welsh sub-populations than among Scottish populations. The highest pairwise FST values involved Mull South, which showed greater differentiation from all other populations than any other pairwise comparison (maximum FST = 0.160, Mull South vs North Argyll). Welsh and Scottish populations were interleaved throughout the FST matrix with no consistent country-level clustering, indicating that genetic differentiation between populations is not structured by country of origin.

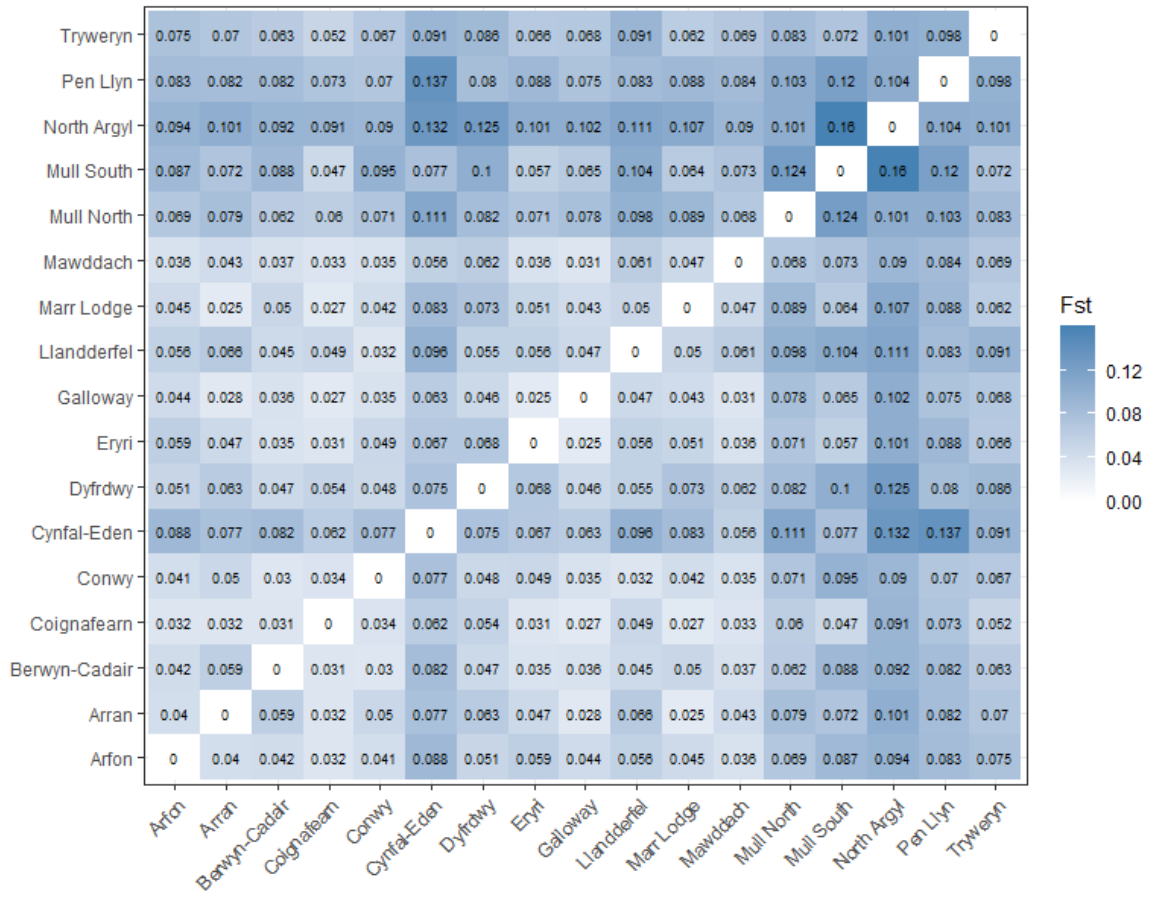

Figure 5 Pairwise FST between *Populus tremula* populations based on ten microsatellite loci calculated on the clone-corrected dataset (Weir and Cockerham 1984). Darker shading indicates greater genetic differentiation.

4.8.2 Principal coordinates analyses

Multivariate ordination of genetic distances was performed using Principal Coordinates Analysis (PCoA) on pairwise Dps (proportion of shared alleles) distances, conducted on the clone-corrected ten-locus dataset. Dps is a standard distance metric for microsatellite data that measures the proportion of alleles not shared between pairs of individuals.

In the PCoA, PC1 explained 9.8% and PC2 explained 8.9% of total variance (Figure 6). The low variance explained on the first two axes reflects weak overall population structure across the dataset. The 95% confidence ellipses for Welsh and Scottish individuals overlapped substantially, with no country-level separation detectable along either axis. No discrete genetic clusters corresponding to geographic origin were apparent.

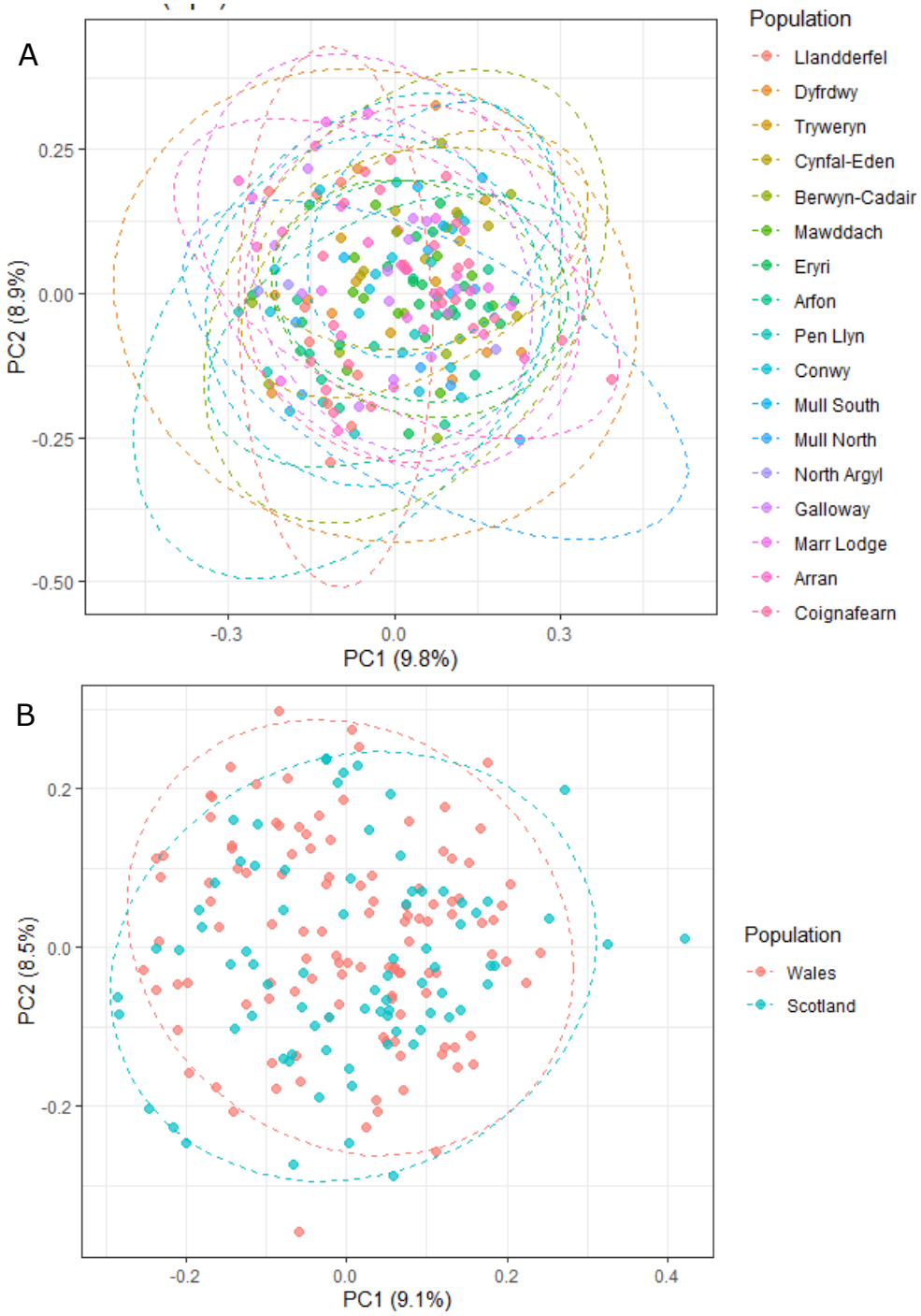

Figure 6 Principal Coordinates Analysis (PCoA) of pairwise Dps (proportion of shared alleles) distances between clone-corrected *Populus tremula* individuals (n = 190). Points represent individual genotypes. (A) individuals coloured by sampling

population; (B) individuals coloured by country of origin. Dashed ellipses show 95% confidence ellipses per group.

4.8.3 Analysis of molecular variance

Hierarchical analysis of molecular variance (AMOVA) was used to partition total genetic variance across three hierarchical levels: between countries (Scotland vs Wales), between populations within countries, and within individuals. Genetic distances were calculated as Euclidean distances on allele frequencies with quasieuclid correction applied to ensure a positive definite distance matrix. Differentiation at each hierarchical level is expressed as a phi statistic (Φ), analogous to FST, representing the proportion of variance attributable to that level relative to total variance. Significance at each level was assessed by 999 permutations. Results are summarised in Table 7.

Table 7 Analysis of molecular variance (AMOVA) partitioning genetic variation in clone-corrected *Populus tremula* across three hierarchical levels (n = 190, PMS05 excluded). Significance assessed by 999 permutations. Bold values indicate statistical significance ($p < 0.05$).

Source of variation	Variance component	% variation	p value
Variations between countries	0.0172	0.25	0.079
Between populations within countries	0.0514	0.74	**0.013**
Between individuals within populations	0.1874	2.69	**0.03**
Within individuals	6.7168	96.33	**0.006**
Total	6.9728	100	NA

The large majority of genetic variation (96.33%) was distributed within individuals rather than between populations or countries. Variation between countries accounted for 0.25% of total variance and was not statistically significant (p = 0.079). Variation

between populations within countries accounted for 0.74% of total variance and was statistically significant (p = 0.013), indicating detectable but modest genetic differentiation between sampling locations. Variation between individuals within populations accounted for 2.69% of total variance and was also significant (p = 0.030).

4.8.4 Isolation by distance

The relationship between genetic and geographic distance was assessed using a Mantel test comparing pairwise genetic distances against log-transformed geographic distances between individuals. No significant correlation between genetic and geographic distance was detected (Mantel r = 0.014, p = 0.240, 9999 permutations) (Figure 7). Genetic differentiation between populations is therefore not organised by geographic distance across the sampled range.

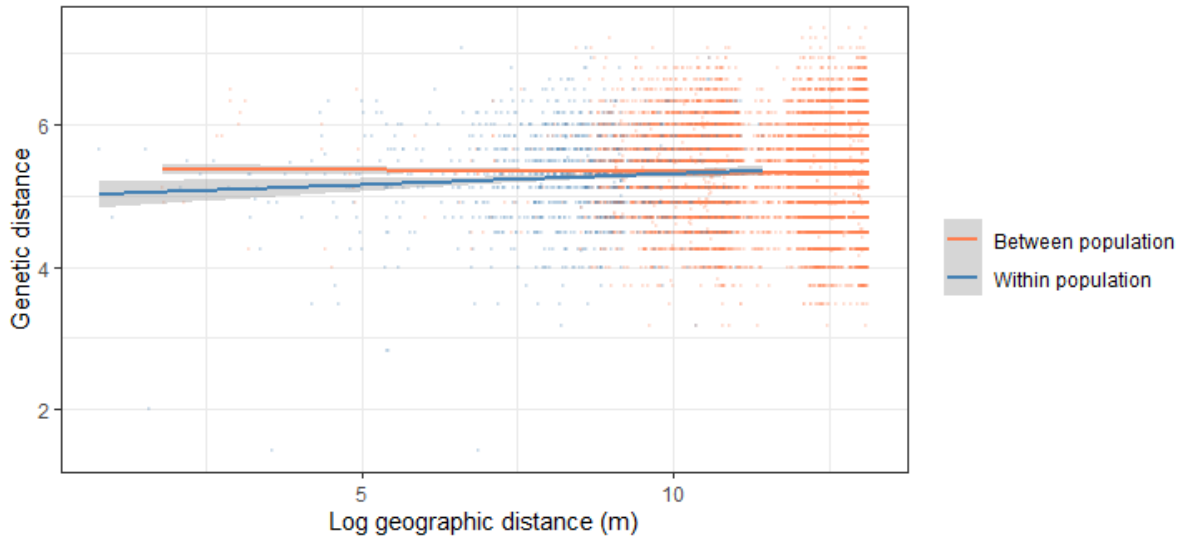

Figure 7 Relationship between log geographic distance and genetic distance for all pairwise individual comparisons in the clone-corrected *P. tremula* dataset (n = 190). Blue points and regression line show within-population pairs; coral points and regression line show between-population pairs. Shaded areas show 95% confidence intervals. Mantel r = 0.014, p = 0.240 (9999 permutations).

4.9 Individual genetic distance matrix

Reliable kinship inference using moment-based estimators typically requires a greater number of markers than employed in this study to achieve adequate precision, particularly for distinguishing between relationship categories such as first-degree relatives and more distant relatives. With ten loci, individual pairwise estimates carry wide confidence intervals that result in substantial overlap between relationship categories, making formal kinship assignment unreliable at the individual level. The Dps distance matrix is therefore provided as a tool for identifying broadly similar or dissimilar individuals rather than for formal kinship inference. Individuals with unusually low pairwise Dps values relative to the population mean may warrant further investigation using higher-density marker panels.

A pairwise genetic distance matrix was calculated across all 190 clone-corrected individuals using Dps (proportion of shared alleles). Dps measures the proportion of alleles that are not shared between two individuals across all loci, ranging from 0 (identical genotypes) to 1 (no shared alleles). This metric is consistent with the distance measure used in the PCoA ordination and is provided as a deliverable for downstream use. The full pairwise matrix is provided as a supplementary file (pairwise_Dps_matrix.csv) alongside a population key (pairwise_Dps_population_key.csv).

Mean pairwise Dps per individual was calculated to screen for genetic outliers - individuals with unusually high mean distance from all other sampled individuals may represent non-native material, or genuinely unusual genotypes.

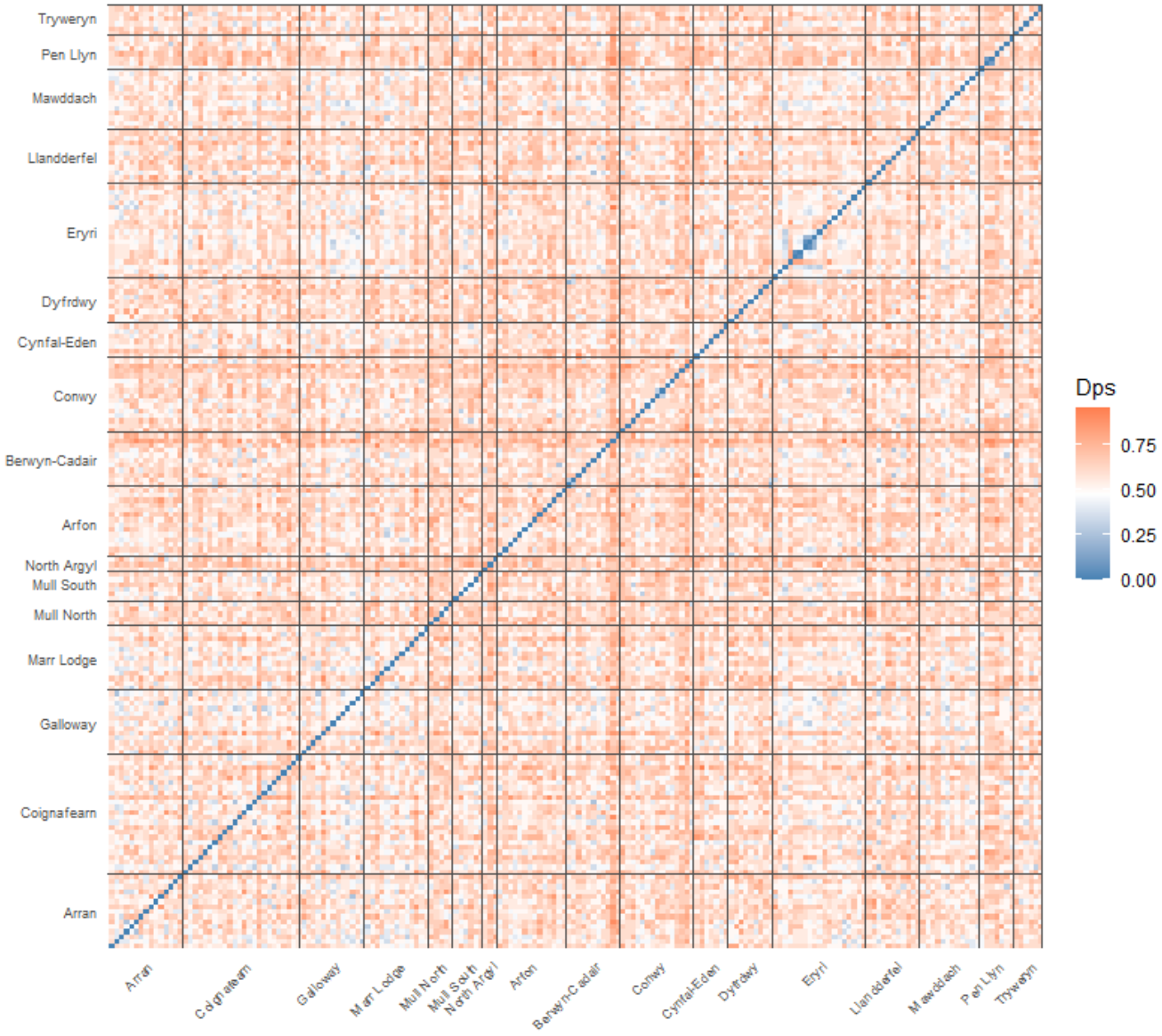

Figure 8 Pairwise Dps (proportion of shared alleles) distance matrix for all clone-corrected *P. tremula* individuals (n = 190, PMS05 excluded). Individuals are ordered by country then population alphabetically. Blue indicates low genetic distance (high allele sharing); coral indicates high genetic distance. Black lines delineate population boundaries. The diagonal represents self-comparisons (Dps = 0).

Outlier screening identified two individuals with mean Dps greater than 2.5 standard deviations above the dataset mean: RHYD02 (Berwyn-Cadair, Z = 3.21, mean Dps

= 0.717) and SND10 (Conwy, Z = 2.63, mean Dps = 0.694). RHYD02 is from Gwastadfryn, Bro Dysynni at 75m altitude (Clone code 24, Origin not recorded). SND10 is from A5, Afon Merddwr, Pentrefoelas at 234m altitude (Clone code 27, Origin not recorded). Neither individual carried novel alleles absent from the main *P. tremula* dataset, indicating their outlier status reflects unusual allele frequency combinations within the *P. tremula* gene pool rather than introgression from non-*P. tremula* sources. Both individuals are retained in the main dataset but flagged for further investigation.

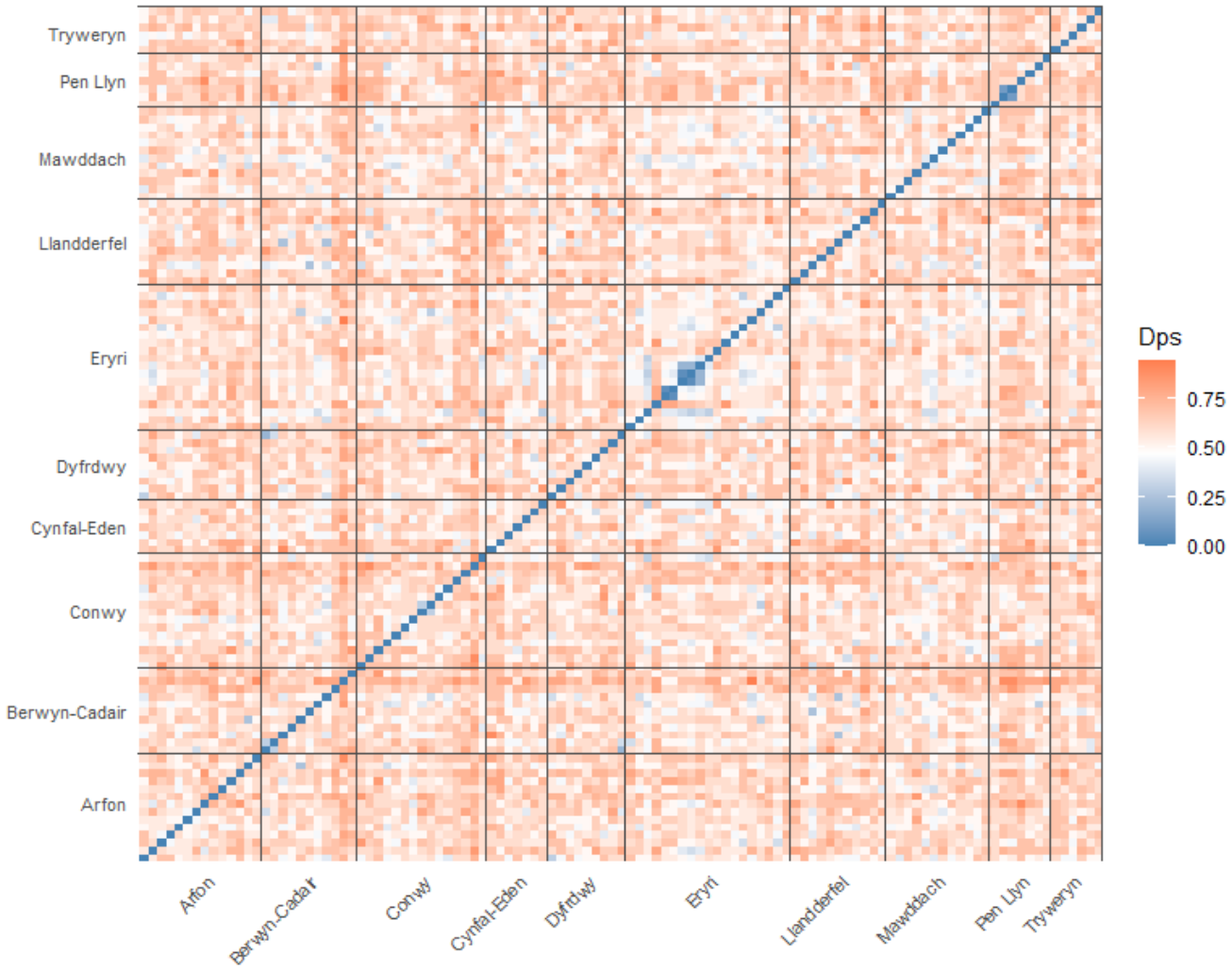

Figure 9 Pairwise Dps distance matrix for Welsh *P. tremula* populations only (n = 111, PMS05 excluded). Individuals ordered by population alphabetically. Colour scale

and layout as Figure 8. The Eryri population shows a notably lighter within-population block relative to other Welsh populations, indicating greater allele sharing among Eryri individuals.

4.10 Hybrid screening

4.10.1 Identification of non-*P. tremula* individuals

Three individuals were retained in a separate flagged dataset following data preparation (Section 4.3): J2426 and RHYD2406 (phenotypically recorded as putative hybrids) and Jen2524 (recorded as *P. alba* based on morphological assessment). Genotypic screening confirmed that all three individuals carry alleles absent from all 237 individuals in the main *P. tremula* dataset. Allele 164 at locus Asp322 and allele 117 at locus PTR04 were present in all three flagged individuals but absent from the main *P. tremula* dataset. Individual J2426 additionally carried novel alleles at locus PMS19 (allele 195). The consistent occurrence of the same novel alleles across multiple flagged individuals suggests these alleles may be diagnostic of non-*P. tremula* hybridisation rather than scoring artefacts.

Principal components analysis (PCA) on allele frequencies was used for hybrid screening rather than PCoA, as PCA is more sensitive to novel alleles absent from the reference *P. tremula* dataset. Both flagged genets were positioned far outside the *P. tremula* cloud on PC1 (5.3% variance explained), confirming substantial genetic differentiation from *P. tremula* (Figure 10).

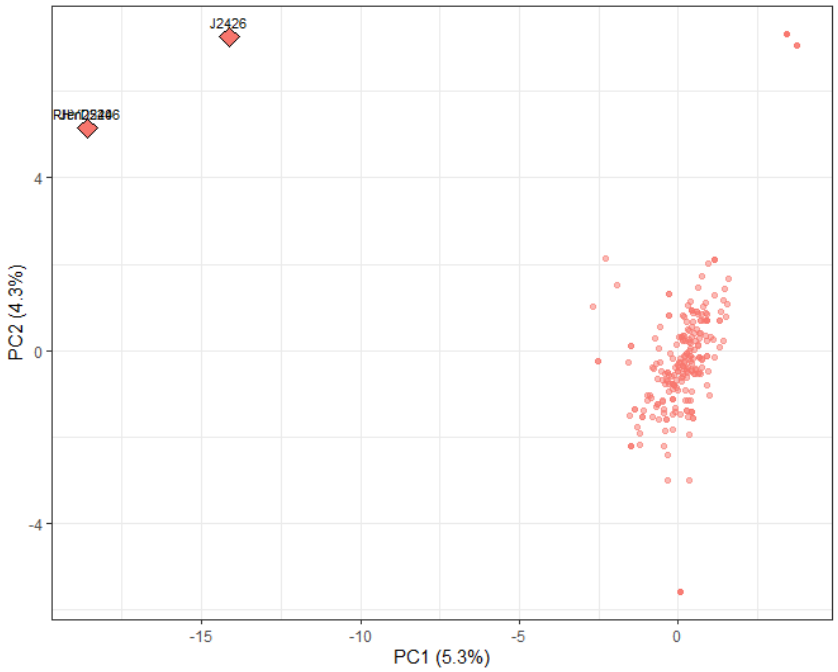

Figure 10 Principal Components Analysis (PCA) of allele frequencies including three flagged non-tremula individuals (diamonds) projected alongside the main *P. tremula* dataset. Both flagged genets (J2426 and Jen2524/RHYD2406) are positioned far outside the tremula cloud on PC1 (5.3% variance explained), confirming genetic differentiation from *P. tremula*. Note that Jen2524 and RHYD2406 overlap, consistent with their identical genotypic profiles.

Mean pairwise Queller-Goodnight relatedness to the main *P. tremula* dataset was 0.193 for J2426 and 0.092 for both Jen2524 and RHYD2406, compared to a mean within-population relatedness of 0.255 across the main tremula dataset. The lower relatedness of Jen2524 and RHYD2406 to *P. tremula* relative to J2426 may reflect differences in the degree of introgression or hybrid generation between individuals.

Jen2524 and RHYD2406 showed shared the same genotype, suggesting that they may represent the same non- *P. tremula* genet sampled on separate occasions.

4.10.2 Screening of main dataset for introgression

No individuals in the main *P. tremula* dataset carried the novel alleles identified in the flagged individuals at loci Asp322, PTR04 or PMS19 that appear indicative of *P. alba*. This indicates no detectable introgression of *P.alba* alleles into the main dataset at these loci. It should be noted that this screening is obviously limited to alleles absent from the *P.tremula* dataset at the ten loci genotyped - introgression of alleles shared between species, or introgression detectable only at unsampled loci, cannot be excluded on the basis of this analysis alone. Interestingly, when Wilson (2019) screened 140 individuals from Scotland with 10, 465 SNP markers and compared them with published whole-genome sequence he found evidence of hybridisation with the closely related species *Populus alba* (6 individuals) and *Populus trichocarpa (3 individuals)*.

Whilst *P. nigra* is a riparian species and perhaps not expected to come in contact with aspen in the UK, there is always the chance of human-mediated planting of *P. nigra* and *P. tremula* within pollination distance. Substantial genotypic data exist for UK black poplar populations, with considerable overlap in the microsatellite marker panels used across species (A'Hara, pers. comm.). Looking at the data there is a clear species-specific allele size range for one of the markers: marker Asp322 (136-154bp in *P. tremula* and 158-183pb in *P. nigra*) and very nearly species-specific range in marker PMS16 (157-193bp in aspen and 133-160bp in black poplar).

If hybridisation between *P. tremula* and *P. nigra* had occurred, we would expect F1 individuals to retain black poplar-specific alleles at one and probably both these diagnostic loci. No such alleles were observed in the Welsh samples, indicating an absence of detectable black poplar genetic contribution. When the Welsh aspen dataset is assessed in this context, there is no evidence of black poplar hybridisation in any of the samples analysed.

5 Limitations of the study

The analyses presented in this report are based on ten microsatellite loci, following exclusion of one locus (PMS05) due to evidence of null alleles. Whilst microsatellite markers are well-established tools for population genetic analysis, the relatively small number of loci limits the precision of individual-level estimates, particularly pairwise relatedness coefficients which carry wide confidence intervals at this marker density. Population-level statistics such as FST, AMOVA variance components and allelic diversity metrics are more robust to marker number and should be interpreted with greater confidence than individual-level estimates.

Microsatellite loci are selectively neutral markers located in non-coding regions of the genome. The analyses presented here characterise neutral genetic diversity and demographic history - patterns of gene flow, genetic drift and clonal reproduction. They do not provide information on adaptive genetic variation or local adaptation. Conclusions regarding the genetic similarity of Welsh and Scottish populations are therefore limited to neutral genetic processes and cannot be extended to inferences about any adaptive variation between populations.

The sampling design was not fully standardised across populations - sample sizes ranged from 3 (North Argyll) to 26 (Coignafearn) individuals per population. Comparisons of diversity metrics and clonality rates between populations should therefore be interpreted with caution, as these reflect sampling design as much as underlying biological differences.

6 Discussion

Despite the limitations outlined above, this study provides the first genetic characterisation of Welsh *Populus tremula* populations and offers new insights into the genetic diversity and population structure of British aspen. The principal finding is that no significant genetic differentiation was detected between Welsh and Scottish

populations at the country level. Whilst statistically significant differentiation was detected among sampling locations (AMOVA, $p = 0.013$), this structure was not organised by geographic distance or country of origin, indicating that Welsh and Scottish aspen do not represent genetically discrete units at the markers examined.

The lack of a correlation between geographic distance and genetic differentiation - that is, the absence of an isolation-by-distance pattern - can arise for several non-mutually exclusive reasons:

1. Historical connectivity and gene flow. Aspen has strong dispersal potential through wind-dispersed pollen and seeds. If populations were more connected historically, past gene flow may have homogenised genetic variation across large areas, leaving a persistent signal that persists even where populations are now fragmented.

2. Clonal reproduction buffering differentiation. Extensive clonal spread via root suckering can maintain local genotypes over long periods, potentially reducing the apparent impact of limited contemporary gene flow and obscuring spatial patterns of differentiation.

3. Recent fragmentation. If habitat fragmentation is relatively recent, there may not have been sufficient time for genetic drift to generate a clear pattern of increasing differentiation with geographic distance.

4. Long-distance dispersal events. Even rare long-distance dispersal events can disrupt isolation-by-distance patterns by introducing genetic material between geographically distant populations.

5. Small number of sampling locations. With only 17 populations included in this analysis, statistical power to detect a distance–differentiation relationship is constrained, particularly where variation among sites is uneven.

6. Landscape and ecological factors. Barriers and corridors such as mountain ranges, river systems and woodland continuity may matter more than straight-line

geographic distance. Genetic structure may reflect these landscape features rather than distance per se.

In this study, the combination of strong dispersal ability, clonal persistence, and likely historical connectivity of aspen populations across Britain provides a plausible explanation for why genetic differentiation exists among sites but does not increase predictably with geographic distance.

6.1 Genetic diversity

Welsh populations showed moderate to high levels of genetic diversity across all metrics examined. Expected heterozygosity ranged from 0.665 to 0.726 across Welsh sub-populations, comparable to values reported for Irish aspen (Buckley *et al.*, 2025) and consistent with the broader pattern of moderate microsatellite diversity reported in European *P. tremula* populations. At the country level, Wales showed marginally higher mean He (0.705) and Ho (0.686) than Scotland (He = 0.677, Ho = 0.648), though the difference was small and both countries showed broadly similar diversity across all metrics. Rarefied allelic richness showed no consistent country-level difference, suggesting the marginally higher heterozygosity in Welsh populations reflects differences in allele frequency distributions rather than the number of distinct alleles present. Private alleles were distributed across both Welsh and Scottish populations with no country-level pattern, indicating that unique allelic diversity is not concentrated in either country. Overall, the genetic diversity results are consistent with populations that have maintained reasonable levels of diversity despite concerns regarding fragmentation and predominantly clonal reproduction in many British aspen stands.

6.2 Clonal replication

The identification of clonal patches sharing identical multilocus genotypes is consistent with the known biology of aspen, which reproduces vegetatively through root suckering and can maintain clonal genets over long periods. The ability to

identify clonal replicates using molecular markers has practical value for conservation management, enabling targeted selection of unique individuals for clone bank or seed orchard creation, and avoiding inadvertent over-representation of a single genet in restoration plantings.

The overall balance between male and female individuals in the Welsh samples (54.7% male, 45.3% female) contrasts with the situation reported in Ireland (Buckley *et al.*, 2025), where single-sex clonal stands were found to predominate. All 26 multi-ramet clonal groups identified in Wales were single-sex, consistent with the clonal reproduction mechanism of root suckering which produces ramets that are genetically and therefore sexually identical. Variation in sex ratio was observed among Welsh sub-populations, though interpretation of these patterns requires caution given the potential influence of targeted sampling of known clonal individuals on the recorded sex ratios.

The identification of clonal replicates separated by considerable geographic distances should be interpreted cautiously. Whilst such findings may reflect genuine clonal expansion, other explanations including human-mediated movement of plant material should be considered. Local knowledge of site history and planting records will be important in evaluating these cases.

6.3 Future research directions

The results of this study provide a foundation for understanding the genetic diversity and population structure of native *Populus tremula* in Britain, but several areas warrant further investigation.

Expanded geographic sampling. The current study included populations from a range of locations in Wales and Scotland, but sampling effort was not standardised and coverage of the full British range of the species is incomplete. Additional sampling from further Welsh and Scottish locations, and inclusion of English populations, would provide a more comprehensive assessment of genetic diversity and connectivity and

help determine whether the patterns observed here are representative of British aspen as a whole.

Increased marker resolution. Microsatellite markers are well-established tools for assessing neutral genetic diversity and population structure, but their resolution is limited relative to genome-wide approaches. Future studies employing high-density SNP markers or whole-genome sequencing would provide greater power to detect fine-scale population structure, identify signatures of local adaptation, and assess historical demographic processes. Wilson *et al.*, (2019) used a panel of over 10,000 SNP markers to examine population structure in Scottish and Eurasian aspen and reported no clear structure among Scottish populations in PCA, consistent with the findings of the present study.

Investigation of adaptive variation. The present study characterised neutral genetic variation only. Understanding adaptive genetic variation is likely to be important for developing resilient conservation and restoration strategies, particularly in the context of climate change, emerging pests and pathogens, and ongoing habitat fragmentation. Common garden provenance trials using genotypes from across the Welsh and Scottish sampling range would provide complementary phenotypic data to inform seed sourcing decisions.

Long-term monitoring. Aspen populations comprise a mixture of long-lived clonal patches and sexually derived individuals. Long-term monitoring of selected populations would help quantify rates of clonal expansion, sexual recruitment and population turnover, improving understanding of the demographic processes maintaining genetic diversity and informing management interventions aimed at promoting natural regeneration.

6.4 Non-*P. tremula* individuals and introgression

Three individuals identified as potentially non-*P. tremula* on the basis of field annotations were confirmed as genetically distinct through novel allele detection and principal components analysis. All three carried alleles at loci Asp322 and PTR04 absent from all 237 individuals in the main *P. tremula* dataset and were positioned far outside the *P. tremula* genetic cloud in multivariate space. No individuals in the main tremula dataset carried these diagnostic alleles, providing no evidence of introgression of non- *P. tremula* genetic material into the sampled populations at the markers examined. Jen2524 and RHYD2406 showed identical genotypic profiles and relatedness signatures, suggesting they may represent the same non- *P. tremula* genet sampled on separate occasions — a finding worth investigating further with local site records. Whilst the microsatellite panel employed here has limited power for definitive hybrid identification, the consistency of the novel allele signal across multiple flagged individuals and its complete absence from the main dataset provides reasonable confidence in these conclusions.

7 Conclusion

This study provides the first genetic characterisation of Welsh *Populus tremula* populations using microsatellite markers, and the first direct comparison of Welsh and Scottish aspen genetic diversity. The principal findings are that Welsh and Scottish populations are not genetically discrete at the country level, that genetic diversity is broadly comparable between countries, and that population structure exists among sampling locations but is not organised by geographic distance or national origin. Clonal structure was detected in Welsh populations, with all multi-ramet clonal groups confined to single sites and all clonal groups found to be single sex. Three non- *P. tremula* individuals were identified and confirmed as genetically distinct, with no evidence of introgression detected in the main dataset. Together these results provide a useful genetic evidence base to inform native aspen conservation and restoration planning in Wales.

Appendix

Primers used in the microsatellite multiplex mixes (1 and 2).

Primer	Sequence	5’ dye	Multiplex mix
PMGC2550F	AGGTTACAAACTTTGTTGTAGC	FAM	1
PMGC2550R	GAACAAACTCTCACTGTGGTC		
PMS19F	AGCCACAGCAAATTCAGATGATGC	FAM	1
PMS19R	CCTGCTGAGAAGACTGCCTTGACA		
Asp302F	GGCAGCGTCTTCATTCCTAT	NED	1
Asp302R	AAAGCCATTGAAGTGGAGGA		
PMS16F	CTCGTACTATTTCCGATGATGACC	PET	1
PMS16R	AGATTATTAGGTGGGCCAAGGACT		
PMS18F	CTTCACATAGGACATAGCAGCATC	NED	1
PMS18R	CACCAGAGTCATCACCAGTTATTG		
PMS14F	CAGCCGCAGCCACTGAGAAATC	VIC	2
PMS14R	GCCTGCTGAGAAGACTGCCTTGAC		
PMS15F	CAACAAACCATCAATGAAGAAGAC	PET	2
PMS15R	AGAGGGTGTTGGGGGTGACTA		
PMS20F	GTGCGCACATCTATGACTATCG	PET	2
PMS20R	ATCTTGTAATTCTCCGGGCATCT		
PMS05F	TTCTTTTTCAACTGCCTAACTT	NED	2
PMS05R	TGATCCAATAACAGACAGAACA		
Asp322F	CATTAACGCCCCATTTCAGT	FAM	2
Asp322R	GTGAGGCACCACCCTGATAG		
PTR04F	TTGCTTTGTTTGATAAAATGTCC	PET	2
PTR04R	GTGCTCTAGCCATGCTCAAA		

Acknowledgements

We are grateful to the following organisations and individuals for contributing samples included in this study: Mar Lodge Estate, Coignafearn Estate, Eadha Enterprises, Andy Robinson, Basil Ramsden and Shaila Rao. Welsh sampling was undertaken with the support of Awdurdod Parc Cenedlaethol Eryri/Snowdonia National Park Authority.

We thank Tom Jenkins and Gail Merriman for commissioning this work and for helpful discussions during its development. We also thank Jenny Wong for her contributions to the project.

References

Buckley, D., Kelleher, C. and Carey, B. (2025). The conservation genetics of Aspen (25/GR/03). Report for DAFM Conservation of Genetic Resources Grant Aid Scheme for Food and Agriculture (GRGAS).

de Carvalho, D., Ingvarsson, P.K., Joseph, J., Suter, L., Sedivy, C., Macaya-Sanz, D., Cottrell, J., Heinze, B., Schanzer, I. and Lexer, C. (2010). Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Molecular Ecology*, 19, 1638–1650.

Easton, E.P. (1997). *Genetic Variation and Conservation of the Native Aspen (Populus tremula L.) Resource in Scotland*. PhD Thesis, University of Edinburgh.

Goudet, J. (2005). hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5, 184–186.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405.

Jombart, T. and Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*. doi:10.1093/bioinformatics/btr521.

Kamvar, Z.N., Tabima, J.F. and Grünwald, N.J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281.

Kamvar, Z.N., Brooks, J.C. and Grünwald, N.J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, 6, 208.

Maksimović, Z., Čortan, D., Ivetić, V., Mladenović-Drinić, S. and Šijačić-Nikolić, M. (2014). Genetic structure of black poplar (*Populus nigra* L.) population in the area of Great War Island. *Genetika*, 46(3), 963–973.

Müntzing, A. (1936). The chromosomes of a giant *Populus tremula*. *Hereditas*, 21, 383–393.

Pakull, B., Groppe, K., Mecucci, F., Gaudet, M., Sabatti, M. and Fladung, M. (2014). A simple PCR-based marker to determine sex in aspen. *Plant Biology*, 17(1), 256–261.

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*, 26, 419–420.

Paradis, E. and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35, 526–528.

Pebesma, E. (2018). Simple features for R: standardized support for spatial vector data. *The R Journal*, 10(1), 439–446.

Politov, D.V., Belokon, M.M., Belokon, Y.S., Polyakova, T.A., Shatokhina, A.V., Mudrik, E.A., Azarova, A.B., Filippov, M.V. and Shestibratov, K.A. (2015). Application of microsatellite loci for molecular identification of elite genotypes, analysis of clonality, and genetic diversity in aspen *Populus tremula* L. (Salicaceae). *International Journal of Plant Genomics*, 2015, 261518.

Provan, J., Brown, J.A., Beatty, G.E., McKeown, N.J., Montgomery, S.S. and Montgomery, W.I. (2026). Ghosts of giants: high levels of clonality may represent massive remnant clones in European aspen (*Populus tremula*). *Conservation Genetics*, 27(2), 45.

R Core Team (2025). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rahman, M.H., Dayanandan, S. and Rajora, O.P. (2000). Microsatellite DNA markers in *Populus tremuloides*. *Genome*, 43, 293–297.

Ramsden, B. (2024). *The Aspen Often Falls Far from the Tree: Unusual Reproductive Dynamics and High Diversity within a Population of Scottish Aspen (Populus tremula L.)*. MSc Thesis, University of Aberdeen.

Smulders, M.J.M., van der Schoot, J., Arens, P. and Vosman, B. (2001). Trinucleotide repeat microsatellite markers for black poplar (*Populus nigra* L.). *Molecular Ecology Notes*, 1(3), 188–190.

Trivedi, C., Cavers, S., Atkinson, N., Clark, J. and Cottrell, J. (2019). *A Strategy for UK Forest Genetic Resources: Protecting the UK's Unique Diversity of Trees and Shrubs*. Royal Botanic Gardens, Kew.

van der Schoot, J., Pospíšková, M., Vosman, B. and Smulders, M.J.M. (2000). Development and characterization of microsatellite markers in black poplar (*Populus nigra* L.). *Theoretical and Applied Genetics*, 101, 317–322.

Weir, B.S. and Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358–1370.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi,

K., Vaughan, D., Wilke, C., Woo, K. and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.

Wilson, J.R. (2019). *Evolutionary History and Local Adaptation in Scottish Populations of Populus tremula*. MSc Thesis, Swedish University of Agricultural Sciences (SLU), Uppsala.

von Wühlisch, G. (2009). *Eurasian Aspen (Populus tremula) — Technical Guidelines for Genetic Conservation and Use*. EUFORGEN, Bioversity International.

Worrell, R. (1995). European aspen (*Populus tremula* L.): a review with particular reference to Scotland. I. Distribution, ecology and genetic variation. *Forestry*, 68, 93–105.

Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Solymos, P., Stevens, M.H.H., Szoecs, E. and Wagner, H. (2024). vegan: Community Ecology Package. R package version 2.6-8. URL https://CRAN.R-project.org/package=vegan.

Alice Holt Lodge
Farnham
Surrey, GU10 4LH, UK
Tel: **0300 067 5600**

Northern Research Station
Roslin
Midlothian, EH25 9SY, UK
Tel: **0300 067 5900**

Forest Research in Wales
Talybont Research Office
Cefn Gethiniog
Talybont-on-Usk
Brecon
Powys, LD3 7YN, UK
Tel: **0300 067 5709**

info@forestresearch.gov.uk
www.forestresearch.gov.uk