Assessing Genetic Variation in Flowering Genes in Selected Oak

Final report to Future Trees Trust, 15 December 2017

Colin T. Kelleher[1], Marc Delêtre[1], Rodrigo Olave[2]

[1]DBN Plant Molecular Laboratory, National Botanic Gardens of Ireland, Glasnevin, Dublin 9, Ireland
[2] Agri-Food and Biosciences Institute, Large Park, Hillsborough, Co. Down, BT26 6DR, Northern Ireland, United Kingdom

Summary

This project utilized molecular markers to investigate genetic diversity and genetic structure in experimental populations of oak (*Quercus robur*). An efficient multiplex reaction was used for undertaking Simple Sequence Repeat (SSR) analysis of 8 markers in the oak samples. This revealed high levels of genetic variation but no genetic structuring. DNA sequences were also used to test for genetic variation in coding regions of the genome. One region was identified to contain a number of potential polymorphisms for future studies. There is potential to extend the use of genetic markers to improve our understanding of flowering in oak.

Introduction

An experiment was established in 2001 to investigate potential of inducing early flowering in oak. Trials of oak (*Quercus robur*) seedlings were established at two experimental sites (Loughgall and Castlearchdale) in Northern Ireland. The experiment utilised different methods of seedling establishment -either Cell Grown (CG) or Root Production Method (RPM), and also thinning treatments to induce flowering. The report from these experiments suggest that growing oak trees with the RPM technique would induce flowering and acorn production in a shorter time by accelerated tree growth (Olave and Kelleher 2017).

As a part of these experiments samples were taken to assess if there was any genetic component to the results or if the results could be wholly attributed to the treatments applied. Separating genotypic and environmental influences is an important component in breeding programmes and assessing performance. This report provides the data from the genetic analysis. To undertake the genetic investigation we implemented a two stage analysis, first we assessed the populations using broad scale genetic markers to give an indication of any potential genetic clustering or genetic relatedness and then we assessed a subset of individuals for sequence variation in genes related to bud burst and flowering.

The aim was to assess if there was any evidence for genetic structuring within the samples and assess potential genetic markers for future studies.

Methods

Genotyping

Samples were collected from all individuals in Loughgall and Castelarchdale and stored dry in silica gel until used. A total of 196 samples (82 and 113 for the Loughgall and Castlearchdale populations, respectively) were analyzed using eight nuclear simple sequence repeat (SSR) markers (QrZAG112, QrZAG15, QrZAG7, QrZAG96, QrZAG20, QrZAG110, MsQ13, QrZAG11; Guichoux et al. 2011). Total genomic DNA was extracted from approximately 20 mg of lyophilized leaf tissue using NucleoSpin 96 Plant kits (Macherey-Nagel, Düren, Germany) following the manufacturer's instructions, except for the cell lysis incubation at 65°C, which was extended to 1 h. Purified DNA was eluted in a final volume of 100µl of TE solution and DNA concentrations were checked with on a Nanodrop ND-2000 spectrophotometer (ThermoScientific, USA).

Multiplex PCRs were carried out on a SimpliAmp™ thermal cycler (Applied Biosystems/Life Technologies, USA) with the following conditions: initial denaturing for 15 min at 95°C, followed by 30 cycles of 30 s at 94°C, 30 s at 56°C and 45 s at 72°C, with a final extension step at 60°C for 10 min. Genotyping was performed by Source BioScience on an ABI PRISM 3130XL Genetic Analyzer (Applied Biosystems/Life Technologies). Each genotyping plate was run along with six duplicate samples from previous runs to check for consistency across runs. Genotypes were extracted and analyzed using Peak Scanner v1.0 (Applied Biosystems).

Sequencing

Four nuclear ESTs (Expressed Sequence Tags) linked to bud burst and flowering (Oakbud2, Oakbud3, Oakbud8 and Oakbud10; Derory *et al.*, 2006) were also tested on four samples with distinct genotypes (2016.157, 159, 176, 197). PCR reactions were performed in 20µl final volume with 1µL (~100ng) of DNA template, 10µL of 2× MyTaq™ Red Mix (Bioline, USA), 8µL ddH2O, and 0.5µM each of forward and reverse primers (Eurofins, Germany).

To increase PCR specificity, a touch-down approach was used with the following cycling conditions: initial denaturing for 5 min at 94°C followed by 10 cycles of 30 s at 94°C, 30 s at $T°_A$ (-1.0°C/cycle) [66-56] and 1 min at 72°C, followed by 30 standard cycles of 30 s at 94°C, 30 s at 56°C and 1 min at 72°C. Amplification ended with a final extension at 72°C for 10 min. Amplifications were checked on 1.5% agarose gel stained with SybrSafe (Invitrogen). Positive PCRs were purified using the ExoSAP method (New England Biolabs) and sent to Macrogen Europe for sequencing in one direction.

Statistical Analyses

Summary statistics (mean number of alleles per locus, allelic richness corrected for sample size and unbiased expected heterozygosity) were computed using FSTAT 2.9.3.2 (Goudet, 1995).

Clustering analyses based on neighbor-joining (NJ) and Nei's minimum genetic distance D_A were carried out using the software POPULATIONS 1.2.31 (Langella, 1999). Bootstrap values were calculated using 1000 permutations of the loci. Trees were visualized and edited using FIGTREE 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

In order to test for the presence of genetic structure, analyses of molecular variance (AMOVA) were carried out in R (R Core Team, 2017) with the *poppr* 2.5.0 package (Kamvar *et al.*, 2014; Kamvar *et al.*, 2015). Significance tests were carried out based on 1000 permutations of the raw data.

Results and Discusion

Genotyping

The Simple Sequence Repeats (SSRs) used are neutral, non-coding regions of the DNA and are useful for assessing genetic variation. The clustering analysis was used to visually assess potential groups or relationships of genotypes in the data and the AMOVA was used to quantify any clustering or diversity partitioning. The clustering analysis indicates if there are any relationships between groups of samples or if there are genetically related individuals. There was no genetic structure apparent in Loughgall (Fig. 1) or in Castlearchdale (Fig. 2). This was confirmed by the AMOVAs, with inter-individual genetic variation representing 98.3% of the total variance in Loughgall (Table 1a) and 99.9% in Castlearchdale (Table 2a). In the former, approximately 3% of variation can be accounted for by between treatment differences. Though significant (P-value 0.008), this is most likely a consequence of differences in sample size between control (n = 13–16) and treatment (n = 6), resulting in higher genetic diversity in the control populations (although differences were not significant; Table 3). The high level of heterozygosity is consistent with other studies on oak genetics (e.g. Cottrell et al. 2003). However, it contrasts the results obtained in Olave and Kelleher (2017), in which inter-SSRs were used and showed low levels of diversity. This suggests that SSRs are better at estimating genetic diversity compared to the inter-SSR markers. A second series of AMOVAs with all replicates combined confirmed the lack of significant genetic structure in Loughgall (Table 1b) and Castlearchdale (Table 2b), dismissing the possibility of a genetic component to the phenotypic differences observed between treatments in both populations.

When all samples are combined into one clustering analysis, no obvious groupings are evident (Fig. 3). Although there are clusters due to the nature of the analysis, these clusters are not supported by bootstrap values, which give an indication of how reliable a grouping is.

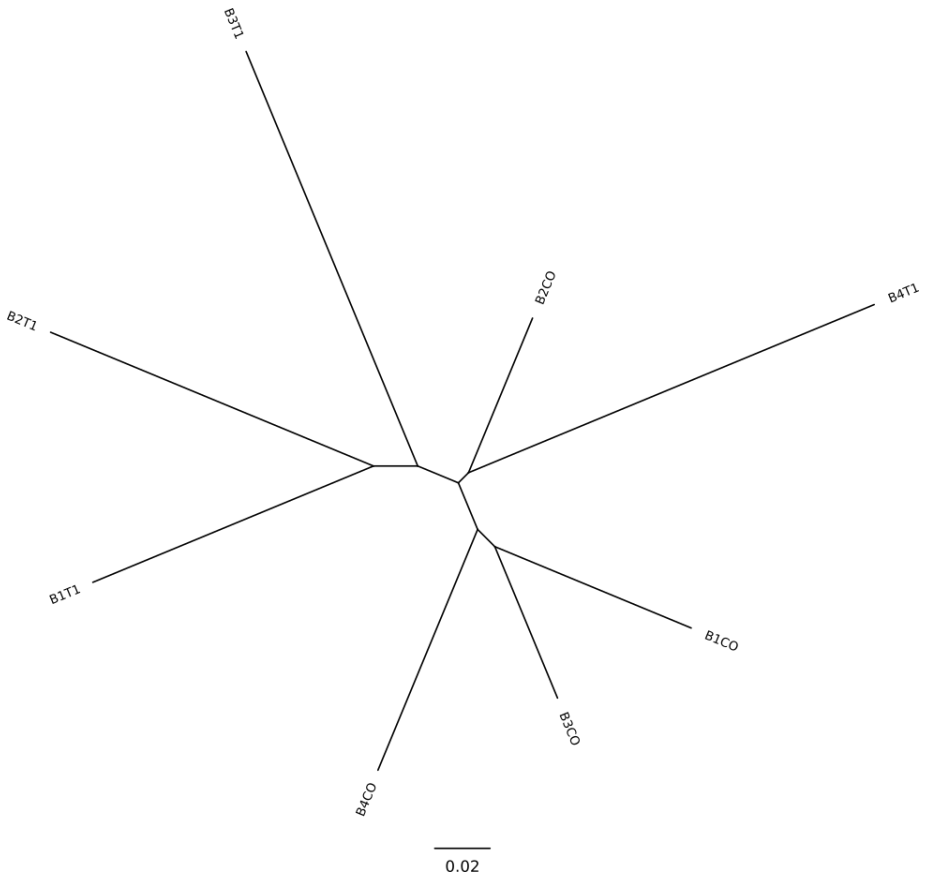

Fig. 1 Unrooted NJ tree of Loughgall (tree removal experiment) based on D_A (Nei *et al.*, 1983). No bootstrap value was >50. There was no significant genetic differentiation between subpopulations or between control and treatment when combining all replications (F_{ST} = 0.007, *NS*).

Table 1. (a) Analysis of molecular variance (AMOVA) of Loughgall samples; **(b)** AMOVA with all four replicates combined.

(a)	Source of Variation	Degrees of freedom	Sum of squares	Variance components	Percentage variation (%)	*P*-value
	Among Blocks	3	14.769	-0.079	-1.687	0.990
	Among Treatments within Blocks	4	24.004	0.158	3.341	0.008
	Within Treatments	74	344.777	4.659	98.345	0.010
	Total	81	383.551	4.738		

(b)	Source of Variation	Degrees of freedom	Sum of squares	Variance components	Percentage variation (%)	*P*-value
	Among Treatments	1	5.874	0.034	0.714	0.103
	Within Treatments	80	377.677	4.721	99.286	
	Total	81	383.551	4.755		

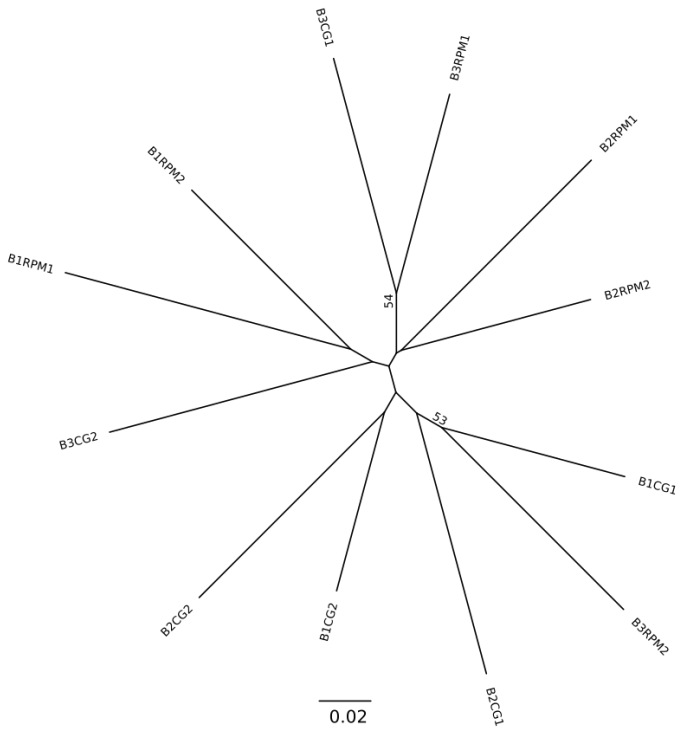

Fig. 2 Unrooted NJ tree of Castlearchdale (growth container experiment) based on D_A (Nei *et al.*, 1983). Only bootstrap values >50 are shown. There was no significant genetic differentiation between subpopulations when treated individually, but CG2 and RPM1 showed very low but significant differentiation when replications were combined (F_{ST} = 0.006, $P < 0.05$).

Table 2. (a) Analysis of molecular variance (AMOVA) of Castlearchdale samples; **(b)** AMOVA with all three replicates combined.

(a)	**Source of Variation**	**Degrees of freedom**	**Sum of squares**	**Variance components**	**Percentage variation (%)**	***P*-value**
	Between Blocks	2	10.94	0.000	0.009	0.505
	Between Treatments within Blocks	9	49.05	0.003	0.058	0.481
	Within Treatments	101	547.45	5.420	99.932	0.430
	Total	112	607.44	5.424		

(b)	**Source of Variation**	**Degrees of freedom**	**Sum of squares**	**Variance components**	**Percentage variation (%)**	***P*-value**
	Among Treatments	3	18.154	0.023	0.421	0.152
	Within Treatments	109	589.290	5.406	99.579	
	Total	112	607.444	5.429		

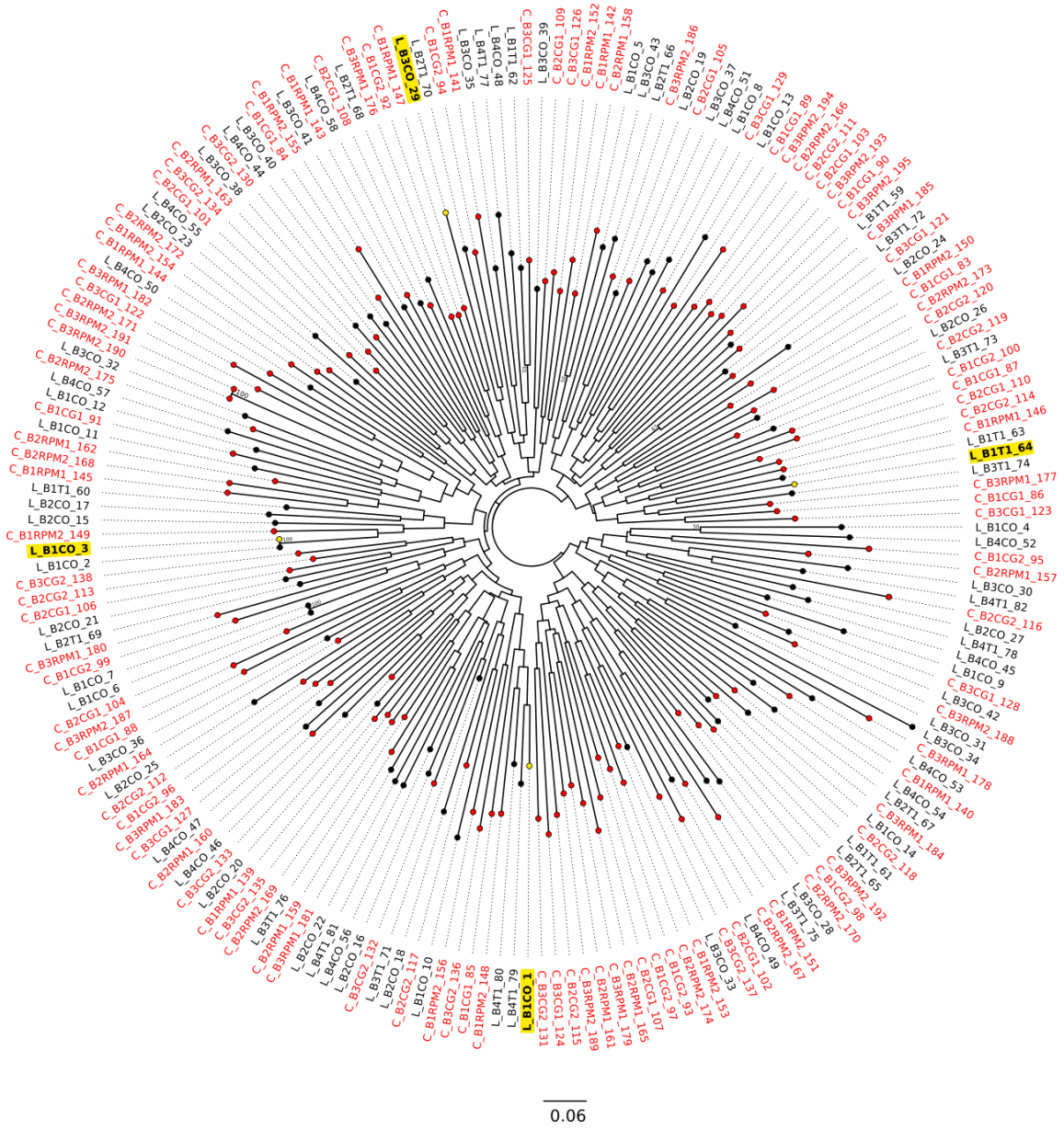

Fig. 3. Unrooted NJ tree of Loughgall (black) and Castlearchdale (red) samples based on D_A (Nei *et al.*, 1983). Only bootstrap values >50 are shown. The four samples used for sequencing analyses are highlighted in yellow. Samples from the two populations are well mixed and dispersed across the tree, with no real apparent clustering. Only 7 bootstrap values (based on 1000 permutations) were >50%.

Table 3. Sample size (*N*) and gene diversity (H_E) among (a) Loughgall and (b) Castlearchdale subpopulations. (CO – control, T – Treatment; CG – Cell Grown, RPM – Root Production Method).

(a)

	Treatments			
	CO		T	
Blocks	*N*	H_E	*N*	H_E
B1	14	0.781 (0.142)	6	0.688 (0.314)
B2	13	0.744 (0.142)	6	0.771 (0.167)
B3	16	0.791 (0.138)	6	0.693 (0.286)
B4	15	0.798 (0.101)	6	0.739 (0.141)

(b)

	Treatments							
	CG1		CG2		RPM1		RPM2	
Blocks	*N*	H_E	*N*	H_E	*N*	H_E	*N*	H_E
B1	9	0.784 (0.170)	9	0.767 (0.201)	9	0.780 (0.147)	9	0.787 (0.142)
B2	10	0.767 (0.179)	10	0.752 (0.188)	9	0.741 (0.207)	10	0.770 (0.171)
B3	9	0.767 (0.235)	9	0.719 (0.204)	10	0.794 (0.135)	10	0.833 (0.094)

DNA Sequencing

Four nuclear ESTs linked to bud burst and flowering time were tested on a subset of four individuals with distinct multi-locus genotypes (DNA accessions 2016.D157, 159, 176, 197). The ESTs are coding regions of the genome and so any variation in these regions could potentially manifest in phenotypic variation. The individuals chosen were representatives of the diversity within the samples (see Fig. 3). Sequences were relatively short (200-400 bp) and of very good quality and revealed little polymorphism overall, with the exception of Oakbud3, which resulted in 7 SNPs (Single Nucleotide Polymorphisms) (Table 4). However, the polymorphism observed in Oakbud3 is potentially useful for future screening to assess variability of genes for flowering and bud burst. It shows that, even in a small population, there is variation in some of the key genes responsible for bud burst and flowering. In addition, for future studies it would be beneficial to assess gene transcript levels in the trees that have early flowering. In particular genes related to flowering or growth hormones could be targeted. This would aid our understanding of what genes the various treatments are impacting on.

Table 4. Sequencing results from the test of four ESTs associated with bud burst (Accessions from Derory *et al.*, 2006).

Locus	Accession	Sequence length (bp)	No. of individuals	Variation - SNPs/Indels
Oakbud2	CR627759	221	4	A/G substitution at position 210
Oakbud3	CR627918	437	4	C/A substitution at positions 58 and 316 A/T substitution at positions 114, 186 and 304 A/G substitution at positions 138 and 165
Oakbud8	CR627822	347/1122*	4	C/T substitution at position 315 A/G substitution at position 318
Oakbud10	CR627933	231	2	None

*2016.D159 1122 bp (probably due to the amplification of an intron)

Conclusions

The overall aim of this part of the project was to test for genetic structuring in the experimental population. Although high levels of genetic diversity were found in the individuals, no genetic structuring was found. The majority of the diversity found in the populations was due to individual differences rather than any other grouping. Thus, we can confidently attribute the variation in the flower inducing experiment (Olave and Kelleher 2017) to differences in treatments or seedling production method rather than to genetic differences. Continued assessment of the flowering and acorn production at the sites in Loughgall and Castlearchdale should be undertaken to gather more long-term data.

The SSR markers used showed greater and more reliable variability estimates than the inter-SSR markers used previously. It is recommended they be used for any future tests on assessing genetic variation in oak. We used a multiplex reaction to minimize consumables used and to increase efficiency.

The DNA sequencing revealed one region of potential use for future studies. This region showed SNPs that could be used for screening populations for variation. A future project investigating gene transcripts in, for example, in genes coding for flowering or growth hormones would give a greater understanding of the potential improvement in performance.

References

Cottrell, J.E., Munro, R.C., Tabbener, H.E., Milner, A.D., Forrest, G.I and Lowe, A.J. (2003) Comparison of fine scale genetic structure using nuclear microsatellites within two British oakwoods differing in population history. *Forest Ecology and Management* 176: 287-303.

Derory J, Léger P, Garcia V, Schaeffer J, Hauser MT, Salin F, Luschnig C, Plomion C, Glössl J, Kremer A (2006) Transcriptome analysis of bud burst in sessile oak (*Quercus petraea*). *New Phytologist* 170, 723–38.

Goudet J (1995) FSTAT (version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* 86 485–486.

Guichoux, E., et al. (2011) Two highly validated multiplexes (12-plex and 8-plex) for species delimitation and parentage analysis in oaks (*Quercus* spp.). *Molecular Ecology Resources* 11(3): 578-585.

Kamvar ZN, Brooks JC and Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics* 6, 208. doi: 10.3389/fgene.2015.00208

Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2:e281. doi: 10.7717/peerj.281

Langella O (1999) Populations 1.2.32. http://bioinformatics.org/tryphon/populations/

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution* 19, 153–170.

Olave, RJ & Kelleher, CT (2017). *Methods of inducing early flowering in oak (Quercus robur* L*)*. Report prepared for Future Trees Trust.

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/