The use of molecular markers and morphological leaf traits for species identification in the Future Trees Trust oak BSOs and plus tree collections

Interim Report

David Boshier[1,2] and Graham Muir[1]
[1]Department of Plant Sciences, University of Oxford
[2]Earth Trust

Introduction

This report covers specific research that provides information to ensure Future Trees Trust oak breeding work continues on a sound scientific footing. While selected plus trees were identified as either *Q. petraea* or *Q. robur* both species were included in the Future Trees Trust Breeding Seedling Orchards. The breeding strategy noted the need to split the two species. If the BSOs and any other orchards are to be registered under the National register as 'qualified' and later as 'tested' there needs to be complete removal of one of the species at each BSO at roguing. Although hybridisation between the two species maybe low it can be high at some sites, particularly where background levels of external pollen are high, as may be the case in the BSOs. The study aimed to clarify the IDs of selected plus trees, and the extent of any hybridisation or miss-collection in the BSO progeny. This will be crucial to the final roguing and selection within the BSOs.

The study used 8 nSSRs (microsatellites in 2 multiplexes) that had previously been shown to discriminate between *Q. robur* and *Q. petraea*. These were used to id FTT grafted oak plus tree selections and some of the related progeny arrays in the oak BSO at Paradise Wood. The results from the molecular study are correlated with a number of morphological measures of leaves that are also considered to be discriminatory using three different approaches (German-Degen http://software.bfh-inst2.de/software.html#Eiche, French-Kremer *et al*. 2002, British-Potter 1994). Sampling was from 32 plus trees + 14 progeny from each of 21 plus trees in Earth Trust's BSO at Paradise Wood.

The aim was to use the results to inform: a) future roguing of the Paradise Wood orchard; b) data analysis of growth and form traits from the BSOs to improve the accuracy of calculations of heritability and gain predictions; c) the accuracy of leaf morphology traits, so that they can be used in roguing of the other oak BSOs. This would save the need for use of molecular techniques for the other BSOs and result in a more cost effective and at the same time accurate way of roguing the orchards to one species.

The FTT selected oak plus trees falls into one of the following eight categories.

1) 94 provisional plus tree selections that were discarded (on vessel size), and are not present in any trials or clones bank.
2) 3 provisional plus tree selections that were discarded (on vessel size), but for unknown reasons are now present in clones banks.
3) 11 plus trees for which original id is doubtful - Jason Hubert marked them with a query. Some of these are held as clones, while 4 have progeny represented in the BSOs
4) 34 plus trees which have progeny in BSOs and also have clonal material from original plus trees
5) 22 plus trees with progeny in BSOs, but no clonal material from the original trees
6) 24 plus trees that are clonally propagated, but don't have progeny in the BSOs
7) 58 plus trees that are not propagated, neither clonally nor as progeny
8) 44 newer EMR plus tree selections (25 Q. robur, 3 Q. petraea, 16 unidentified; 1 region 10, 7 region 30, 36 region 40), most of which are available at Bradbourne, Kent (rows 249-283), and 44 were transferred to NRS. New selections have not been through the vessel screening cull use for the first plus tree selections.

Groups 3, 4 and 8 were the focus for this study using leaf material to compare morphology of the adult (from clonal material) with the morphology of the progeny at the Sotterly BSO to see how the characters segregate in the progeny and with the original plus tree id.

Materials and Methods

The original budget supplied by the FC was augmented by the following contributions. Some £1,250 of consumables was supplied from the hosting lab. 5 days of time were supplied by Earth Trust for measuring leaf samples. D. Boshier spent 6 days more than the original budget on project coordination, leaf sampling/ measuring data analysis and write up.

Leaf collection

Fresh leaves were collected in early July directly from the trees in the field trial and dried in ziplock bags with silica gel.

Microsatellite genotyping

Genomic DNA was isolated from leaves using the Qiagen DNeasy 96 Plant Kit (Qiagen; Hilden, Germany). Eight microsatellite loci: QrZAG7, QrZAG112, QrZAG20, QrZAG 96, QrZAG11, from Kampfer et al (1998); MsQ13 from Dow et al (1995) and QpZAG15, QpZAG110 from Steinkellner et al (1997), were amplified in a multiplex PCR following standard protocols (Guichoux et al 2011).

Briefly, fluorescently labeled PCR primers were used in a 10 µL reaction volume (0.4 mg/ml BSA, 1.5 mM $MgCl_2$, 400 µM dNTPs, 0.15-0.5 µM of each primer (concentrations for each primer pair in the premix are shown in the Appendix protocol), 20–100 ng template DNA and 2.0 U HotStar *Taq* DNA polymerase, (Qiagen; Hilden, Germany). The cycling profile consisted of an initial denaturation step of 15 min followed by 30 cycles of 30 sec at 94 °C, 1 min at 56 °C and 45 sec at 72 °C. A final extension at 72 °C for 45 min was used to assure a quantitative terminal transferase activity of the *Taq* polymerase. Labeled PCR fragments were detected on an ABI 3730xl DNA Analyzer (Applied Biosystems; Foster City, CA), alongside an internal size standard (LIZ 500), and fragment sizes were recorded using GeneMapper® v4.0 (Applied Biosystems).

Data analysis

Individuals from different families were pooled into two groups based on phenotypic classification, one for *Q. petraea* and one for *Q. robur*. Diversity indices and general statistics were calculated using MSAnalyser 4.05 (Dieringer and Schlötterer 2003). Using the same program, we determined Θ, an unbiased estimate of Wright's Fixation Index (Weir & Cockerham 1984), to identify the most informative loci separating the two species. The significance of pairwise Θ-values (referred to as F_{ST} in the text) was tested by permuting genotypes among groups. This method of permutation does not rely on Hardy-Weinberg assumptions (Goudet *et al*. 1996). A hierarchical analysis of molecular variance (AMOVA Excoffier *et al*. 1992) using the program ARLEQUIN 3.5.1.2 (Excoffier et al. 2005) was employed to examine the partitioning of microsatellite variance into components derived from species differences and those derived from families. The total variance was partitioned into covariance components resulting from: within populations (families), among populations (families) within species (taxa) and among species (taxa, including hybrids).

A model-based clustering method implemented in the program STRUCTURE (Pritchard et al. 2000) was used to assign individuals probabilistically to homogenous clusters (*K* populations) without consideration of sampling localities. Estimated posterior probabilities for the simulated model fitting the data were calculated assuming a uniform prior for *K*, where $K \in \{1, 2, 3, 4\}$. To minimize the effect of the starting configuration during the Monte Carlo simulation, we simulated 10^4 updates of the Markov chain (aka burn in) before data for the parameter estimation were collected from another 10^6 iterations. At least ten independent runs of the Markov chain were performed to assure convergence of the chain and homogeneity among runs for each prior of *K*. The program was run without population identifiers (USEPOPINFO = 0) and in the admixture mode (NOADMIX = 0). The best number of clusters was assessed according to the Δ*K* criterion proposed by (Evanno, Regnaut et al. 2005).

Results

Eight polymorphic microsatellites were screened in ˜22 open pollinated families of *Quercus petraea* and *Quercus robur* from the Paradise Wood BSO of the FTT. Genetic differentiation between both species was low (F_{ST} = 0.019), but highly significant (P = 0.0001) when all loci were considered jointly. A separate analysis of individual loci indicates that six (out of the eight) microsatellite loci significantly (P < 0.05) differentiate the two species (Table 1). To discount the possibility that this differentiation merely reflects familial substructure within species, we used the hierarchical sampling to partition the total variance according to species and family.

Table 1. Pairwise F_{ST} values between *Quercus petraea* and *Q. robur* for the eight microsatellite loci

	Wright's fixation index	
Locus	F_{ST}	***P*-value before & after correction** [a]
MSQ13	0.0127	0.0001*
Zag110	0.0096	0.0564
Zag15	0.0116	0.0228
Zag11	0.0268	0.0002*
Zag112	0.0288	0.0001*
Zag20	0.0184	0.0002*
Zag7	0.0206	0.0001*
Zag96	0.0294	0.0001*
All	0.0196	0.0001

[a] based on 10,000 replicates *indicates significance at the 5% level after Bonferroni correction for multiple tests.

Estimates of population structure using Φ_{ST} from AMOVA (Excoffier *et al. 1992*) suggest low but statistically significant subdivision both among populations (families) and between species (and hybrids) studied. Among families (within both parents and their hybrids), Φ_{ST} across loci at all sites is ~0.04 while among species and hybrids Φ_{ST} is ~0.02. Most of the variation (~96%) is partitioned within families and individuals (Table 2) while only approximately 1% was partitioned between species. This subdivision between species (treating hybrids as one of three 'species') was statistically significant (Φ_{ST} = 0.019; P < 0.001) as was the subdivision between families (Φ_{ST} = 0.044; P < 0.001). Excluding hybrids from this analysis altered the result only slightly: both subdivision among families and between species actually increased in statistical significance (Φ_{ST} = 0.044; P < 0.0001 and Φ_{ST} = 0.022; P < 0.0001, respectively).

Table 2. AMOVA analysis of population structure for eight microsatellite loci in *Quercus petraea*, *Quercus robur* and their hybrids.

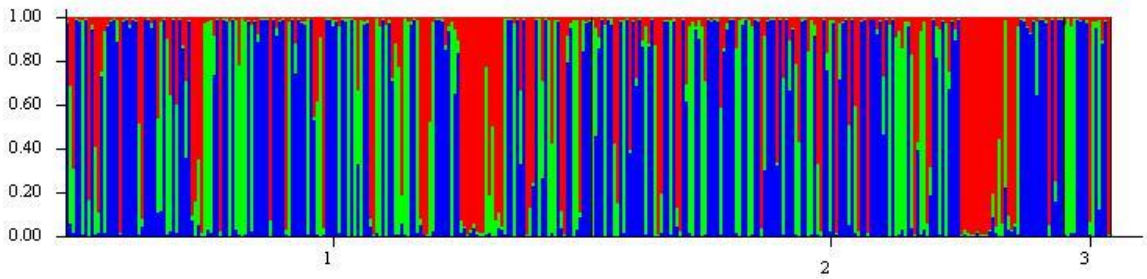
	Microsatellites			
Quercus petraea, Q. robur & their hybrids	**df**	**Sum of squares**	**% variation**	□ [a)]
Among species (taxa)	2	0.019	1.07	$□_{CT}$ = 0.019***
Among families within species (taxa)	19	0.044	2.46	$□_{SC}$ = 0.044***
Within families	610	1.745	96.48	$□_{ST}$ = 1.745***
Total	631	1.808		

a) □□is a fixation index similar to Wright's *Fis* statistic. It reflects the correlation of random pairs of haplotypes drawn from a group (species or family) relative to the correlation of pairs of random haplotypes drawn from the whole dataset (Excoffier *et al. 1992*). Significance is indicated with stars: *** P < 0.001;

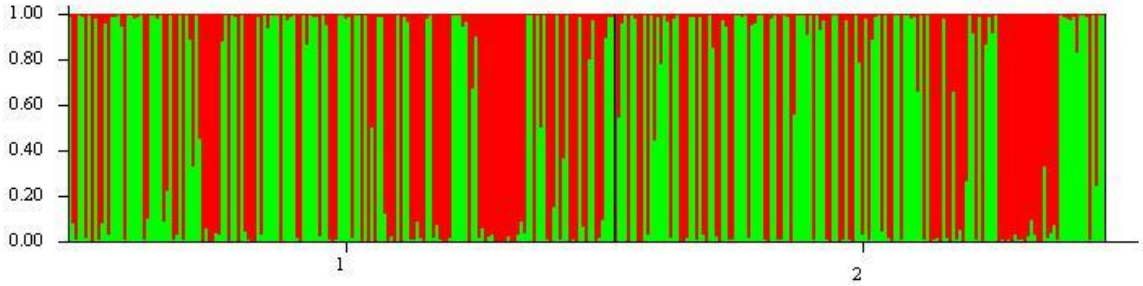
Despite this low, but statistically significant, hierarchical structure, an analysis of all taxa in STRUCTURE did not identify a most likely model of (*K*) genetic clusters and individuals did not cluster according to their population of origin. In the absence of a most likely number of genetic clusters, we set *K* = 3 to reflect the expected number of *a priori* biological groupings (*petraea*, *robur* and hybrids). However, no structure was apparent, and individuals from all three taxa were assigned to all three *a priori* clusters. A separate analysis including just the two parents (excluding the hybrids) did not improve this assignment.

Figure 1. Approximate Bayesian clustering (Pritchard et al. 2000) of *Quercus* individuals in the sample using eight microsatellite loci. In our analysis, Ln P(D) did not reach a maximum. A) *Cluster membership for K* = 3, all taxa, B) *Cluster membership for K* = 2, no hybrids, *petraea* and *robur* only.

A



B



Conclusions

The eight microsatellite loci genotyped in this study were apparently insufficient to assign individuals to either species (or to families), despite including two highly differentiating loci (*Zag96* & *Zag112*) detected between the two species in a European-wide sample of populations from *Quercus petraea* and *Quercus robur* (Muir and Schlötterer 2005). In this latter study, these two loci (and a larger battery of 20 microsatellites) had the power to assign individuals to taxonomic units.

The cause of the discrepancy between the two studies is not immediately clear. For example, the distribution of allele frequencies at the two most potentially differentiating loci (96 & 112) from Muir and Schlötterer (2005) indicate that one particular allele (of size 136 base pairs) was present at elevated frequencies in *Q. robur* (a frequency of nearly 80%). Conversely, the same allele occurred at a low frequency in *Q. petraea*. In the samples genotyped in this study, this same allele (136) occurs at a frequency of barely 30% in *Q. robur*. The picture is also not as expected (when compared to Muir and Schlötterer (2005) for locus 112, where the 86 allele (equivalent to our 88 allele) is equally frequent in both species. In both cases at these loci, alleles occur at lower than expected frequencies and without these frequency differences, discriminatory power appears to be weak.

There are two possible causes for the absence of these alleles occurring at high frequency across our sample, either: 1) the samples (phenotypes) used in this study do not reflect the underlying genotype. i.e. there is a high proportion of *Q. robur* phenotypes which are in fact *Q. petraea* or hybrid genotypes; or 2) the selective sweep which appears to have elevated the frequency of these alleles in mainland Europe did not occur in the U.K. (Muir and Schlötterer 2006). The morphological data suggests that

the former is the most likely. We believe the latter not to be the case, though if it were it would indicate that insufficient microsatellite loci have been genotyped at present.

1) The markers show separation principally into two groups, but these do not correspond to the original species IDs of the seed trees, nor are they consistent within the individual families.
2) the morphometric data suggest: a) some families (up to 66% with 5-17% mixing) may contain a mixture of both species - this maybe due to mixing of acorns during the original collections (the acorns were picked off the ground under the plus tree - starting from the trunk of the tree and spiralled outwards aiming to get the required quantity before the edge of the canopy). b) some plus trees maybe incorrectly identified and may require reclassification in the register. c) some material in the clonal archive may not have come from the originally selected plus tree (large disagreement between progeny IDs and maternal ID).

The study has revealed that the issues are more complex than originally thought and require more time than originally envisaged/budgeted to come to more concrete answers that will be directly useful to advance the FTT work. The data suggests that more work needs to be done to follow up these issues. This requires considerably more time going through the detailed records (and talking to Jason Hubert who made the original selections) to look at the stand composition from which the original plus-tree selections were made, to see if the 'mixed' families are from mixed stands and vice versa, and the non-problematic families are in stands regarded as pure *Q. robur* or *Q. petraea*. Such information will also allow a more sophisticated re-run of the molecular and morphological data to give more conclusive results that will help plan the long-term management/roguing of the oak BSOs and identify the future research needed prior to roguings.

Acknowledgements
We thank Liam Dolan for hosting and partly funding the molecular biology part of this study in his lab and H. Prescott for technical assistance. Christina Vinson answered many questions concerning GeneMapper. Suzanne Schmitt for assistance with german translation. Steve Gregory, Peter Mobbs, and Tom for leaf measurements. Ed Eaton for access to data from his MSc study.

Literature cited
Dieringer D, Schlötterer C, 2003. microsatellite analyser (MSA): a platform-independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167–169.
Dow, B. D., M. V. Ashley, and H. F. Howe. 1995. Characterization of highly variable $(GA/CT)_n$ microsatellites in the bur oak, *Quercus marcrocarpa*. *Theoretical and Applied Genetics*, **91**, 137–141.
Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
Excoffier, L., P. E. Smouse and J. M. Quattro, 1992 Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.
Excoffier, L. G. Laval, and S. Schneider, 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47-50.
Goudet J, Raymond M, de Meeüs T, Rousset F, 1996. Testing differentiation n diploid populations. *Genetics*, **144**, 1933–1940.
Guichoux E, Lagache L, Wagner S, Léger P, Petit RJ, 2011. Two highly validated multiplexes (12-plex and 8-plex) for species delimitation and parentage analysis in oaks (*Quercus* spp.). *Molecular Ecology Resources*, **11**, 578-85.
Kampfer S, Lexer C, Glössl J, Steinkellner H, 1998. Characterization of $(GA)_n$ microsatellite loci from *Quercus robur*. *Hereditas*, **129**, 183–186.
Kremer, A., Dupouey, J. -L., Deans, J. D., Cottrell, J., Csaikl, U., Finkeldey, R., Espinel, S., Jensen, J., Kleinschmit, J., van Dam, B., Ducousso, A., Forrest, I., Lopez de Heredia, U., Lowe, A. J., Tutkova, M., Munro, R. C., Steinhoff, S. and Badeau, V. 2002. Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. *Annals of Forest Science* **59**, 777-787.

Muir G, Schlötterer C. 2005. Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Molecular Ecology*, **14**, 549–561.

Muir G, Schlötterer C .2006. Moving beyond single-locus studies to characterize hybridization between oaks (*Quercus* ssp.). *Molecular Ecology* **15**, 2301–2304.

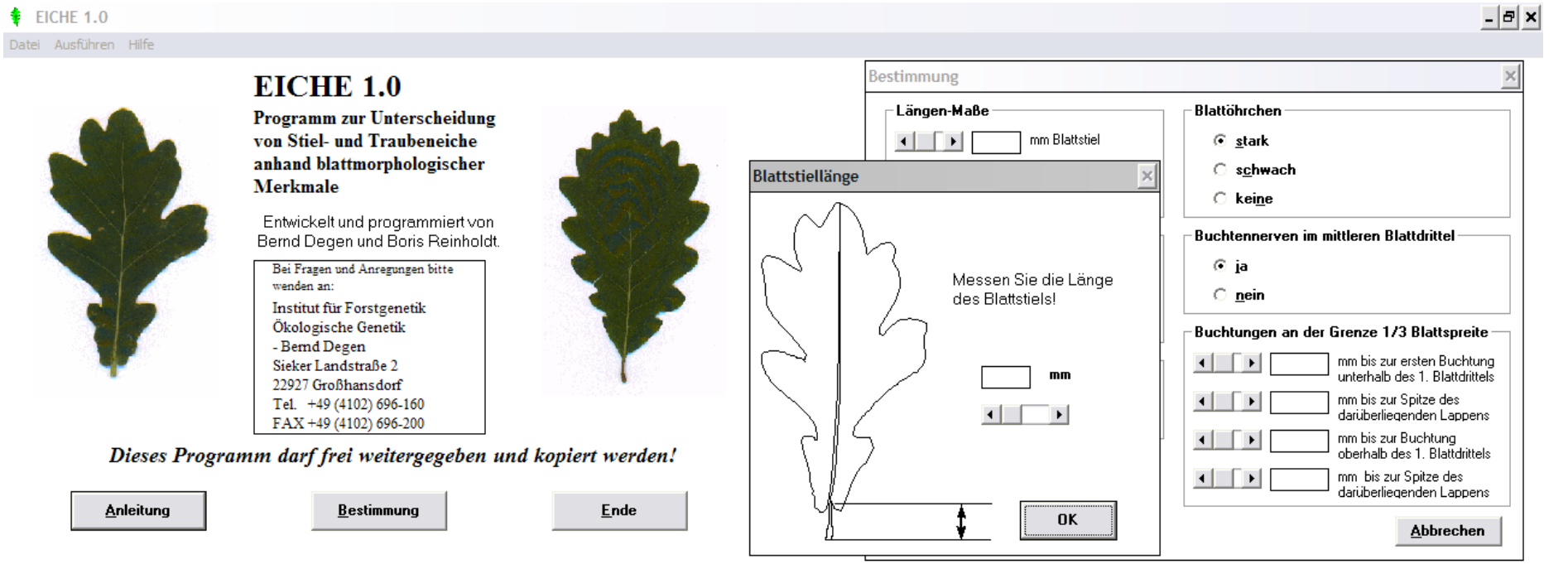
Potter, S.M. 1994. A guide to the identification of Pedunculate and Sessile Oaks and their hybrids *Quarterly Journal of Forestry* 88:

Pritchard, J K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959

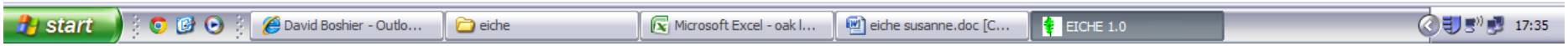
Steinkellner H, Fluch S, Turetschek E *et al.* 1997. Identification and characterization of $(GA/CT)_n$ – microsatellite loci from *Quercus petraea*. *Plant Molecular Biology*, **33**, 1093–1096.

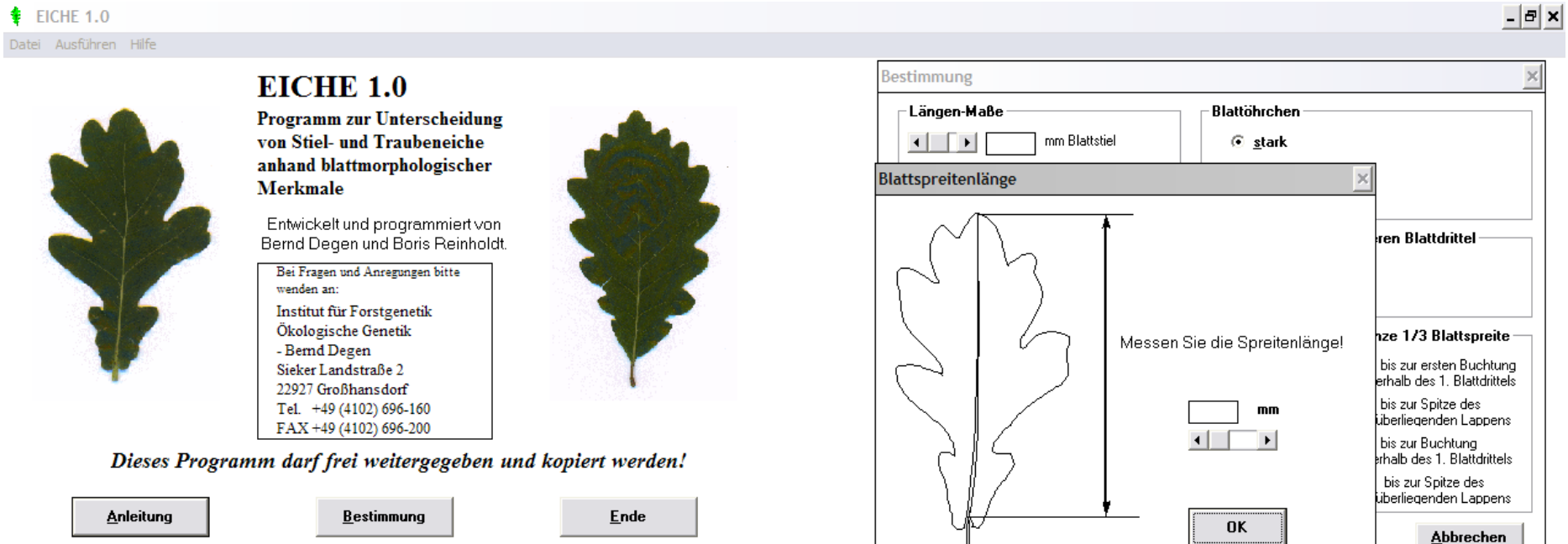
Weir BS, Cockerham CC, 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Appendix 1 Leaf traits measured

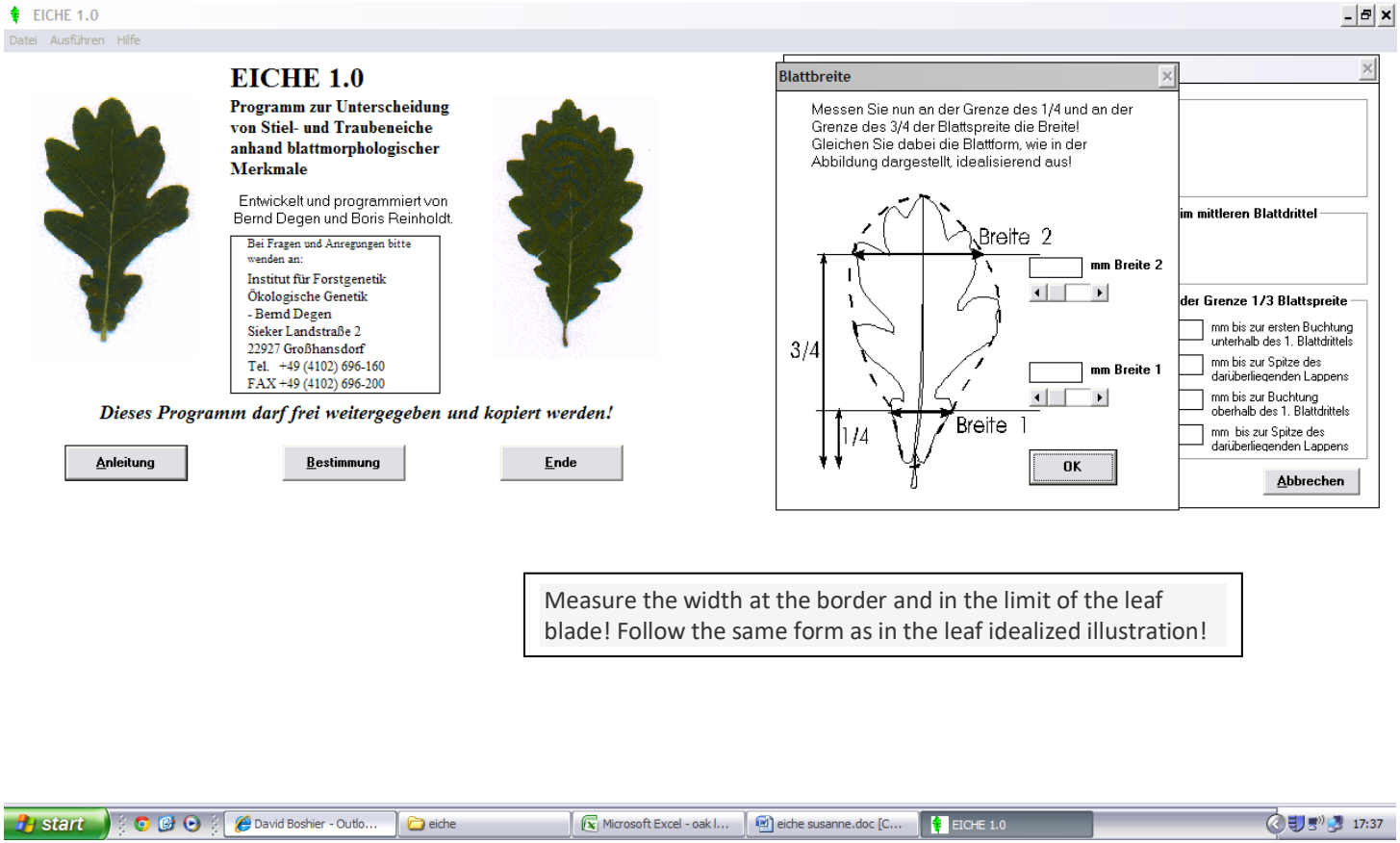


Measure the length of the petiole as in the illustration!

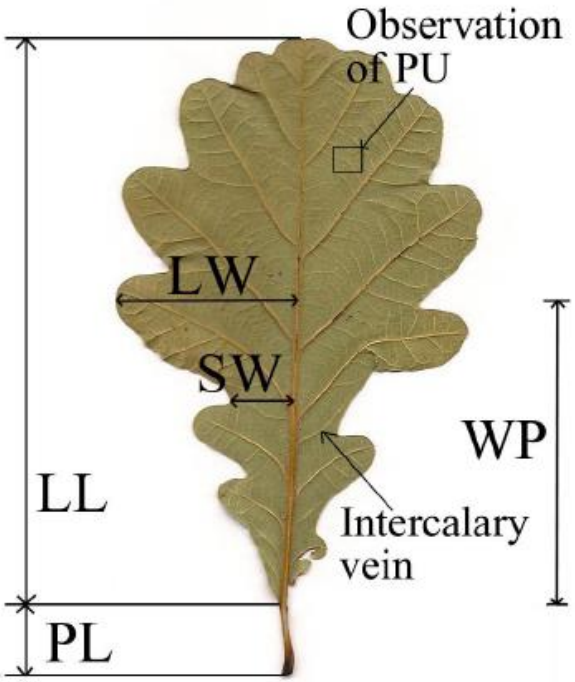


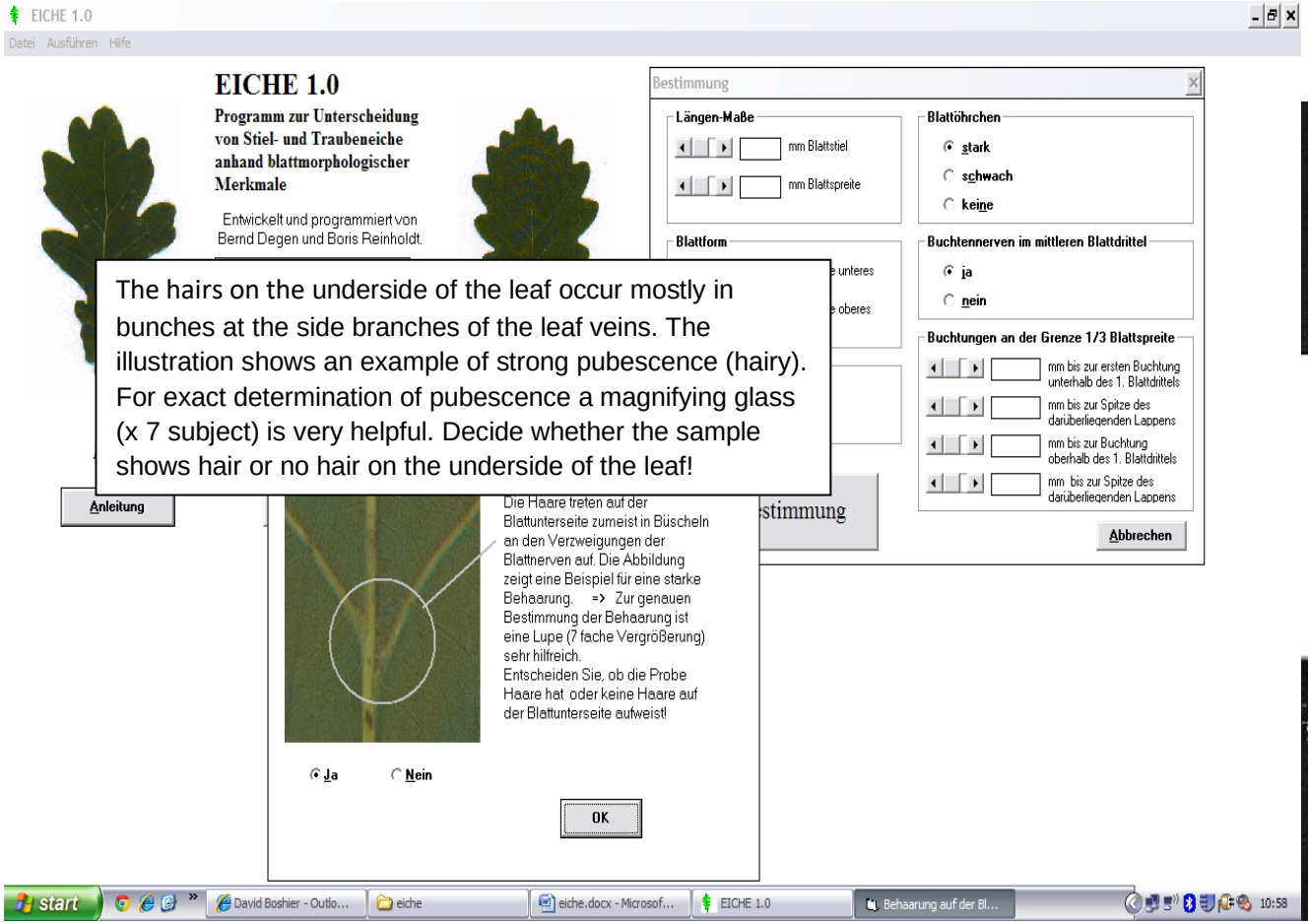


Measure the length of the leaf blade as in the illustration!

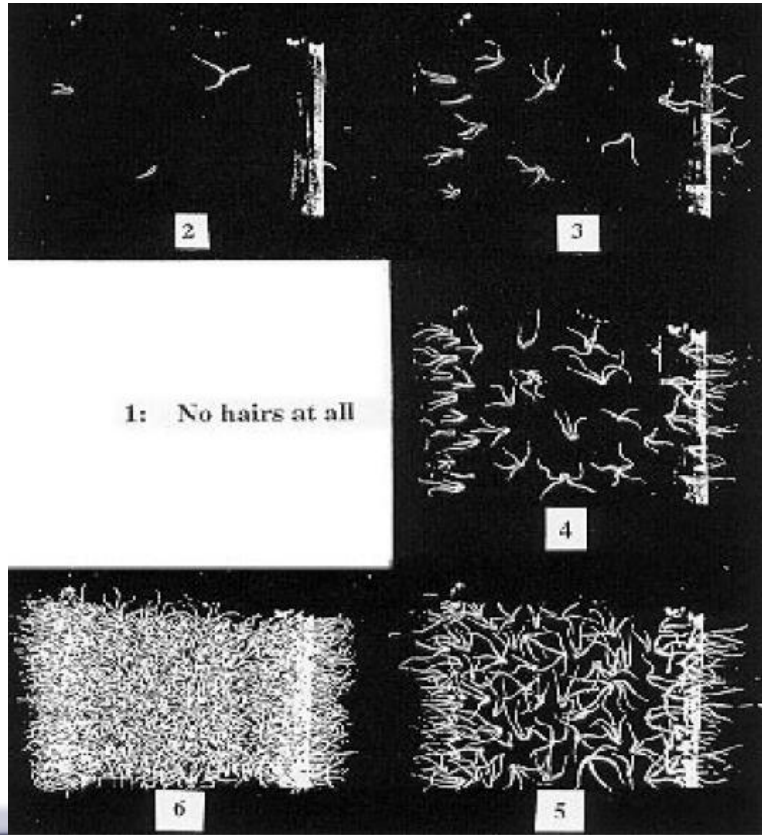


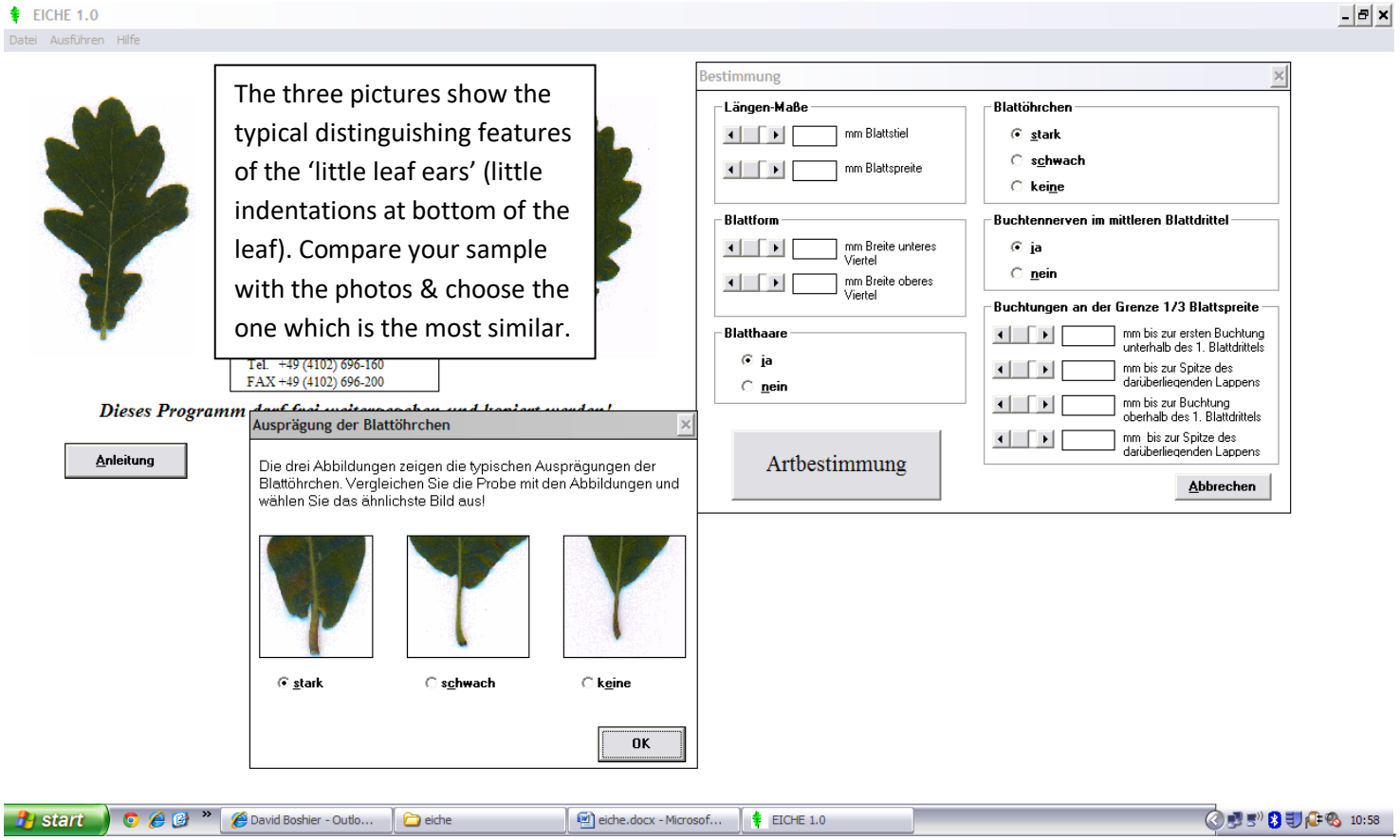
Measure the width at the border and in the limit of the leaf blade! Follow the same form as in the leaf idealized illustration!

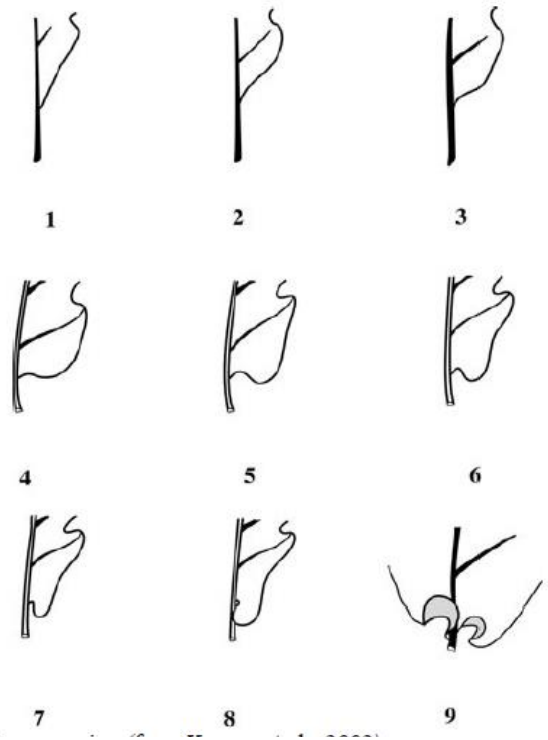


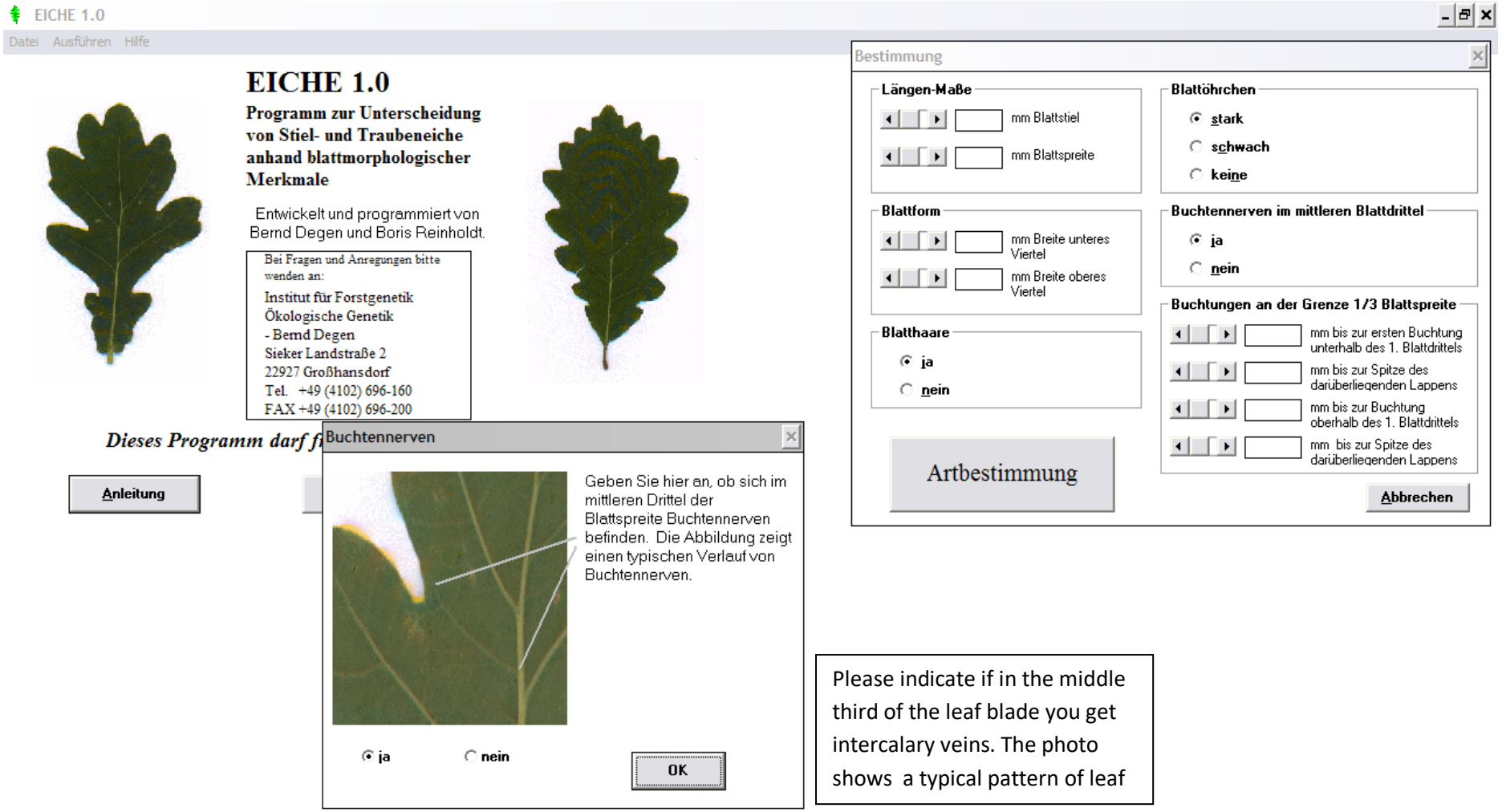


The hairs on the underside of the leaf occur mostly in bunches at the side branches of the leaf veins. The illustration shows an example of strong pubescence (hairy). For exact determination of pubescence a magnifying glass (x 7 subject) is very helpful. Decide whether the sample shows hair or no hair on the underside of the leaf!

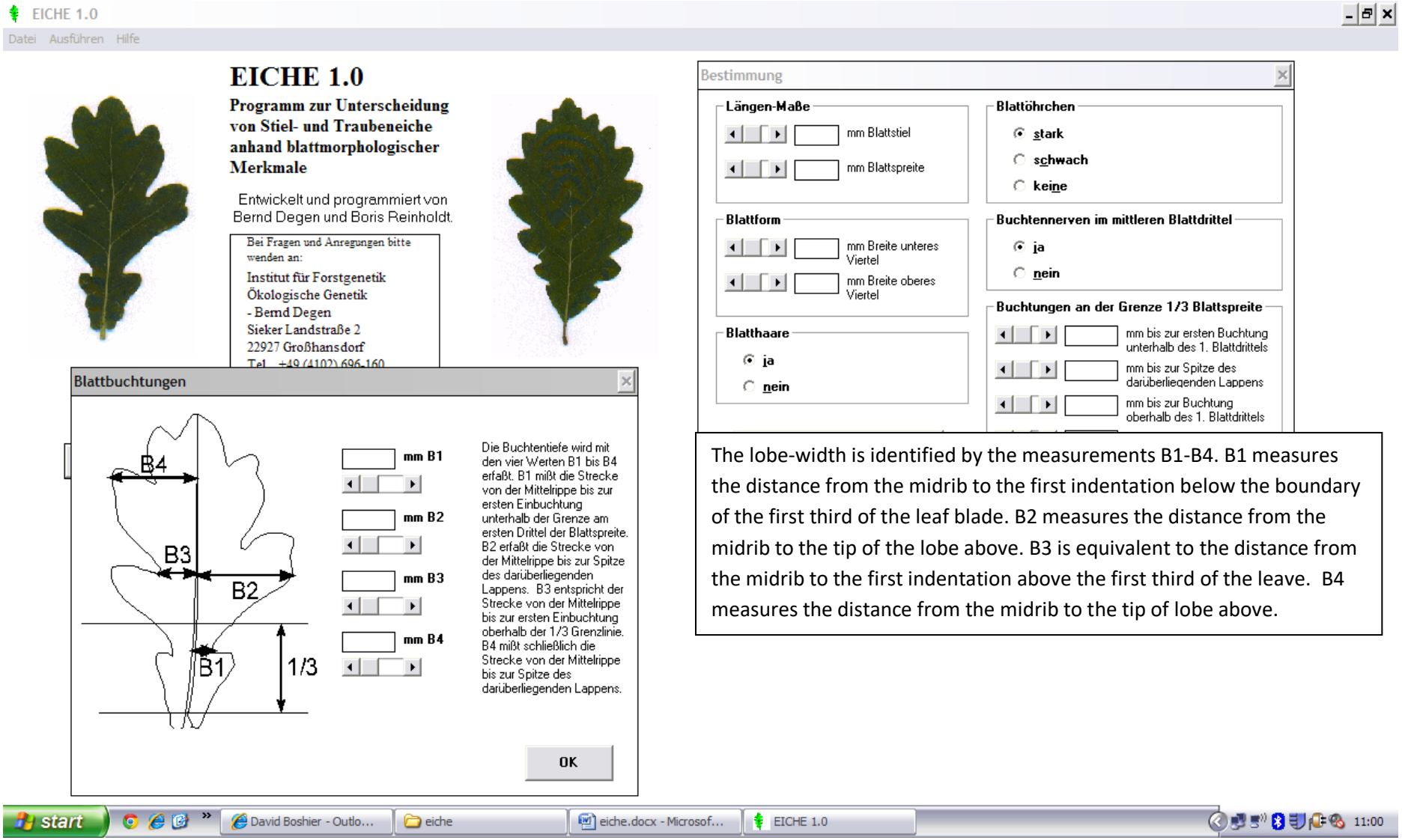



EICHE 1.0
Datei Ausführen Hilfe
The three pictures show the typical distinguishing features of the 'little leaf ears' (little indentations at bottom of the leaf). Compare your sample with the photos & choose the one which is the most similar.
Tel. +49 (4102) 696-160
FAX +49 (4102) 696-200
Anleitung
Ausprägung der Blattöhrchen
Die drei Abbildungen zeigen die typischen Ausprägungen der Blattöhrchen. Vergleichen Sie die Probe mit den Abbildungen und wählen Sie das ähnlichste Bild aus!
stark
schwach
keine
OK
Bestimmung
Längen-Maße
mm Blattstiel
mm Blattspreite
Blattform
mm Breite unteres Viertel
mm Breite oberes Viertel
Blatthaare
ja
nein
Artbestimmung
Blattöhrchen
stark
schwach
keine
Buchtennerven im mittleren Blattdrittel
ja
nein
Buchtungen an der Grenze 1/3 Blattspreite
mm bis zur ersten Buchtung unterhalb des 1. Blattdrittels
mm bis zur Spitze des darüberliegenden Lappens
mm bis zur Buchtung oberhalb des 1. Blattdrittels
mm bis zur Spitze des darüberliegenden Lappens
Abbrechen
start
David Boshier - Outlo...
eiche
eiche.docx - Microsof...
EICHE 1.0
10:58


1
2
3
4
5
6
7
8
9



Please indicate if in the middle third of the leaf blade you get intercalary veins. The photo shows a typical pattern of leaf



The lobe-width is identified by the measurements B1-B4. B1 measures the distance from the midrib to the first indentation below the boundary of the first third of the leaf blade. B2 measures the distance from the midrib to the tip of the lobe above. B3 is equivalent to the distance from the midrib to the first indentation above the first third of the leave. B4 measures the distance from the midrib to the tip of lobe above.

A guide to the identification of Pedunculate and Sessile Oaks and their hybrids

This is based on a method by S.M. Potter, a version of which was published in the "Quarterly Journal of Forestry", Vol.88, No.1, 1994.

In the Field

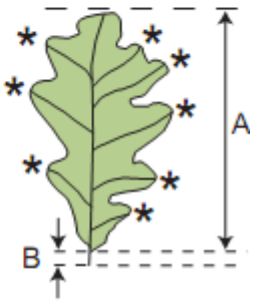
Collect fifteen leaves from the tree to be identified. These should be taken from the most accessible part of the lower crown, but epicormic shoots or coppice growth should be avoided. Young trees, less than about 4m in height may still have predominantly juvenile foliage, which cannot be reliably identified by this method.

If possible the fifteen leaves should be made up of three samples of five, taken from equidistant points around the crown. Select whole, undamaged leaves from the middle part of the shoot; avoid leaves which have been produced during the second (Lammas) flush of growth from late July onwards, as these can be of abnormal shape.

Pack the leaves in polythene bags.

At Home

If the leaves cannot be measured immediately they should be stored, in their polythene bags, in a refrigerator or preferably in a freezer, to prevent mildew or depredations by the insects that will inevitably be attached to them.



Take one leaf at a time from the bag and make the following measurements:

1. Measure the length of the leaf blade or lamella - 'A' on Diagram
2. Measure the length of the leaf stalk or petiole - 'B' on Diagram.
3. *The petiole percentage*, 'P' , for that leaf is calculated as

$$P = \frac{100B}{A}$$

Make a note of the value of P.

4. *Calculate the number of pairs of leaf lobes.*
As the leaf may not be symmetrical the easiest way to do this is to count the number of lobes and divide by two. On the diagram above, each lobe is marked by an asterisk: there are 3 on one side of the leaf and 5 on the other, so the number of pairs of lobes, 'L' , is

$$L = \frac{3+5}{2} = 4$$

Note that the tip of the leaf does not count as a lobe. It may sometimes be difficult to distinguish between a small lobe and a wavy margin to the leaf, especially near the petiole. If in doubt, note whether a vein extends from the mid-rib to the edge of the leaf at that point: if it does, this should be counted as a lobe. **Make a note of the value of L.**

5. *Assess the depth and regularity of lobing.*
A more or less symmetrical leaf (as in diagram opposite) has regular lobing, whilst that shown in above would count as irregular. The depth of lobing can usually be assessed by eye, but if in doubt the measurement shown opposite can be used.
If X = O.5Y or less the lobing is shallow; if more than O.5Y it is deep.

- If the lobes are shallow and regular, give the leaf a score of +1;
- If deep and irregular score -l;
- If intermediate, i.e. shallow and irregular or deep and regular, score 0.

(N.B. The leaf in diagram above would score -l and that in diagram opposite +1).

6. Assess whether auricles (small ear-like appendages) are present or absent at the base of the leaf blade, near the petiole. The diagram below gives some examples of "strong" and "weak" auricles.

Strong auricles Weak auricles

- If auricles are absent or there is a single very weak one the leaf scores +1;
- If two strongly developed auricles score -l;
- If intermediate, i.e. two very weak auricles or one, well developed, score 0.

7. Look at the underside of the leaf.

- If there are clusters of silvery-white hairs on each side of the mid-rib, especially near its base, visible with the naked eye the leaf will score +1.
- If hairs are quite obviously absent score -1.
- If there are hairs present but they are thinly scattered and scarcely visible without a lens, score 0.

8. Add together the three scores obtained in steps 5, 6 and 7 to give a total *Leaf Character Score*, C, which can vary from +3 to -3. **Make a note of C.**

9. This completes the measurement of the first leaf. The process is now repeated for each of the remaining fourteen leaves and each of the values recorded above summed and the totals divided by 15 to arrive at average values of P, L and C.

10. Plot the values of P, L and C (as P against L and P against C) on the graph, the following rules apply to its interpretation:—

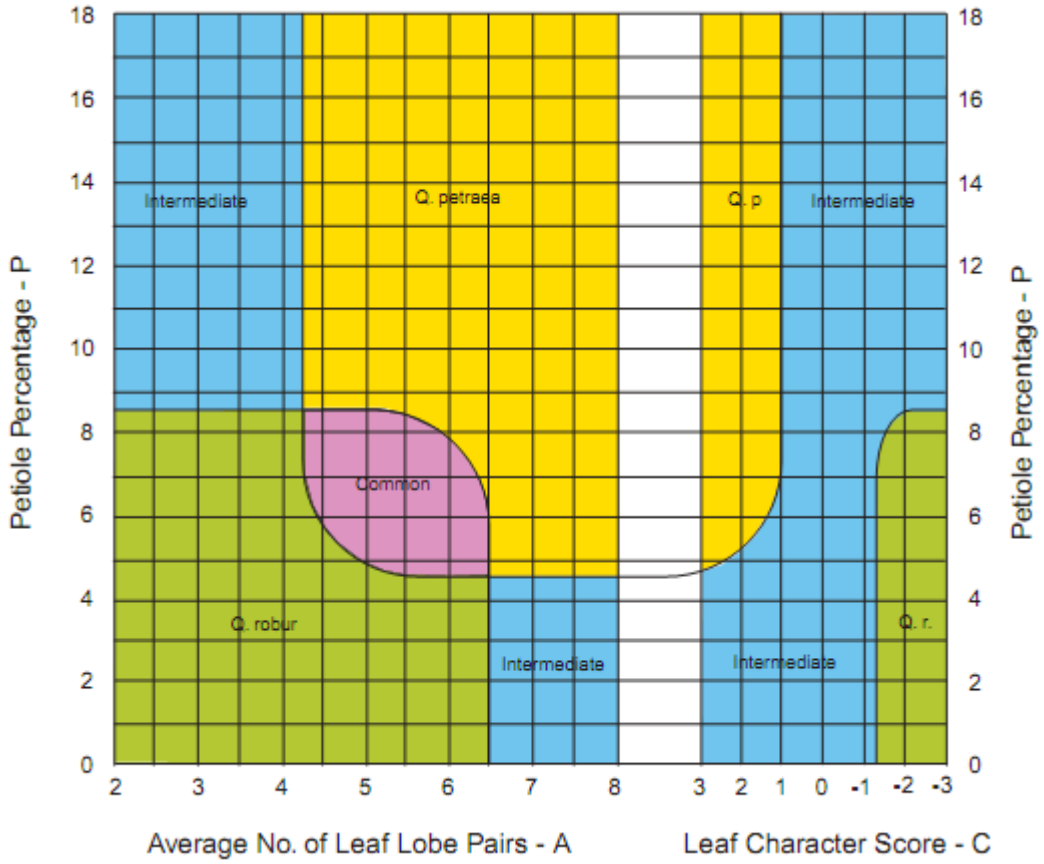
(i) If the tree falls within the "Q. petraea" area on both parts of the diagram, it can be regarded as *Quercus petraea*, Sessile Oak.

(ii) If the tree falls within the "Q. robur" area on both parts of the diagram, it can be regarded as *Quercus robur*, Pedunculate Oak.

(iii) If the tree falls within the "Q. robur" or "Q. petraea" area on one part of the diagram but in the "Intermediate" area on the other part it can be regarded as a hybrid.

(iv) If the tree falls within the "Intermediate" area on both parts of the diagram it can be regarded as a hybrid.

(v) If the tree falls within the "Common" area on the left hand side of the diagram refer to its position on the right hand side to identify it.


Petiole Percentage - P
Petiole Percentage - P
Intermediate
Q. petraea
Q. p
Intermediate
Common
Q. robur
Intermediate
Intermediate
Q. r.
Average No. of Leaf Lobe Pairs - A
Leaf Character Score - C