

**Route, Speed and Mode of Oak Postglacial Colonisation across the British Isles: Integrating Molecular Ecology, Palaeoecology and Modelling Approaches**

ANDREW LOWE,\* † CHARLES UNSWORTH,‡  
SOPHIE GERBER,§ SAM DAVIES\*¶, ROBERT  
MUNRO,\* COLIN KELLEHER,\*\* ANDY KING,††  
SIMON BREWER‡‡, ANDY WHITE¶ & JOAN  
COTTRELL||

\**Centre for Ecology & Hydrology, Edinburgh, Bush Estate, Penicuik EH26 0QB, UK; †School of Integrative Biology, University of Queensland, St Lucia, Brisbane QLD 4072, Australia; ‡Department of Engineering Science, University of Auckland, 20 Symonds Street, Auckland, New Zealand; §Biodiversité, Gènes et Écosystèmes, INRA, Université Bordeaux I, 69 Route d'Arcachon, 33612 Cestas cedex, France; ¶Departments of Mathematics, Heriot-Watt University, Edinburgh EH14 4AS, UK; \*\*Department of Botany, Trinity College, University of Dublin, Dublin 2, Ireland; ††IACR Long Ashton Research Station, Long Ashton, Bristol BS41 9AF, UK; ‡‡CEREGE CNRS 6635/Université Paul Cezanne, Europôle Méditerranéen de l'Arbois, BP80, F-13545 Aix-en-Provence cedex 4, France; ||Forest Research, Northern Research Station, Roslin EH25 9SY, UK*

**Summary**

This paper describes the route, speed and mode of colonisation of oaks by integrating a number of independent analyses using molecular ecology, palaeoecology and simulation modelling approaches. Using a synthetic map of the contemporary distribution of chloroplast DNA (integrating several published and unpublished data sets and describing variation in 1468 trees from 313 autochthonous stands of *Q. robur* and *Q. petraea* from Britain and Ireland), and considering the postglacial topographic landscape, the most

Correspondence: Department of Ecology and Evolution, School of Earth and Environmental Science, The University, Adelaide, SA 5005, Australia.

likely routes of postglacial colonisation across the British Isles are suggested. The overall pattern of these directions agrees with previous interpretations, but several routes, particularly those into Ireland, differ from previous interpretations and benefit here from using a single synthesised data set. Interestingly, the Atlantic oakwoods appear to have been colonised by individuals bearing a single haplotype (type 12). Two palaeoecology data sets, published separately for Britain and Ireland, are synthesised here and used to infer the timing of first arrival of oaks across the British Isles (between 9500 and 6000 years before present). The maximum observed colonisation speed within the British Isles is approximately  $500 \text{ m year}^{-1}$  in central and southern England. Outputs from a simulation model, which mimics postglacial colonisation processes, and which has been parameterised for the colonisation rate observed from the pollen core record and contemporary cpDNA structure, predict that the rapid colonisation rate observed, for at least the southern portion of the British Isles, can only be achieved via very rare (an approximate frequency 0.01%), very long distance seed dispersal events (up to 100 km). Potential agents of such dispersal events are birds or major meteorological disturbances, *e.g.* hurricanes. Additional simulation modelling and genetic analysis of latitudinally stratified populations indicate that non-synchronous colonisation fronts, topographic barriers and temperature related survival may also have had an effect on the speed of migration and resulting genetic structure. Finally, in an attempt to record predicted long distance seed dispersal events, a novel curve fitting technique is applied to molecular parentage assignment data for field established seedlings from a contemporary population. A notable discrepancy is recorded between contemporary field estimates (just over 1 km) and those predicted by simulation modelling, and is discussed in detail. A concluding section describes future research priorities.

## Introduction

In terms of the number of studies of genetic variation, phylogeography and gene flow, oaks are probably the most extensively investigated tree species in the world. This formidable volume of literature, which spans more than 50 years, has shed light on many aspects of historical dispersion, colonisation and contemporary gene flow of oak species. In particular, many workers have examined the chloroplast DNA phylogeography of oaks and these studies have shed tremendous light on the postglacial recolonisation routes of oaks across Europe. More recently, studies integrating palaeoecology, modelling approaches, population genetic and gene flow analyses have been used to further our understanding of the speed and dynamics of oak postglacial colonisation. It is the purpose of this paper to review existing data and to present several new pieces of evidence to extend our understanding of the route, speed and mode of postglacial colonisation across the British Isles. The following topics are covered:

1. Routes of colonisation – Integration of several published and unpublished data sets on the contemporary distribution of chloroplast DNA within the British Isles to suggest the most likely routes of postglacial colonisation with reference to regional topography and postglacial sea levels.
2. Speed of colonisation – Integration of two previously published palaeoecology data sets which record the first occurrence of oak pollen in sediment cores across the British Isles to infer a map of the timings of first arrival of oaks across the British Isles. The maximum observed colonisation speed within the British Isles is also calculated.

3. Mode of colonisation – Results of simulation modelling approaches which mimic postglacial processes are presented. These predict the maximum migration rate and ensuing population genetic structure within an advancing colonisation front of oaks. Considering the potential maximum colonisation rate observed from the pollen core record (2 above), the model-predicted cpDNA population genetic structure is compared with that from contemporary British Isles populations (1 above). This is in order to estimate the maximum distance and frequency of long distance dispersal events which occurred during the actual postglacial colonisation of the British Isles. Further simulation modelling results are also presented which consider the impact of non-synchronous colonisation fronts, topographic barriers to dispersal and temperature related survival on the speed of migration and resulting genetic structure within colonised populations. To test for this expected topographic impact, oak population genetic parameters are examined across a latitudinal cline within the British Isles.
4. Estimating maximum seed dispersal distances – A notable discrepancy occurs between the maximum distance of seed dispersal predicted by simulation models and that so far recorded from natural populations. Using a novel curve fitting technique, a new method of estimating maximum seed dispersal distance, using parentage analysis data from an adult tree and seedling population genotyped for polymorphic microsatellites, is presented. Reasons for discrepancies between modelled and actual dispersal estimations are discussed.
5. Future directions – The concluding section describes potential future research areas and highlights how oak postglacial data sets are being exploited in novel ways.

### **1. Routes of Colonisation**

Several investigators have used variation within chloroplast DNA (cpDNA) of contemporary populations to reconstruct the European postglacial colonisation routes using phylogeographic methods (Ferris *et al.*, 1993; Petit *et al.*, 1993; Ferris, 1996; Dumolin-Lapègue *et al.*, 1997; Ferris *et al.*, 1998; Petit *et al.*, 2002a; 2002b). In addition, these methods have been used to undertake detailed analysis of the British Isles, including Ireland (Ferris *et al.*, 1995; Cottrell *et al.*, 2002; Kelleher *et al.*, 2004; Muir *et al.*, 2004). Such is the volume of work conducted on European oaks that the scoring of cpDNA variants has now been standardised to allow comparisons to be made across studies (see Petit *et al.*, 2002a for an overview). Combining the information from these studies, but mainly relying on the more recently published syntheses, it appears that the British Isles were almost exclusively colonised by oak sources which survived the last glacial maximum in an Iberian refugium. For example the study of Cottrell *et al.*, 2002 noted that 98% of samples from what was believed to be autochthonous British oak bore one of three cpDNA haplotypes of Iberian refugial origin, (Petit *et al.*, 2002a, 2002b). Several studies have recorded strong longitudinal structuring of chloroplast variation within the British Isles, with haplotype 12 mainly found in Ireland and western Britain, haplotype 11 in eastern Britain and haplotype 10 in central and southern Britain (Ferris *et al.*, 1995; Ferris *et al.*, 1998; Cottrell *et al.*,

**Table 1.** Details of additional genotypes supplied for this study. Genotyping of the AK samples was done by RM and AL. Existing data from the studies of Cottrell *et al.* (2002), Petit *et al.* (2002a), Kelleher *et al.* (2004), Lowe *et al.* (2004) and Muir *et al.* (2004) have also been combined with this new data set for this paper.

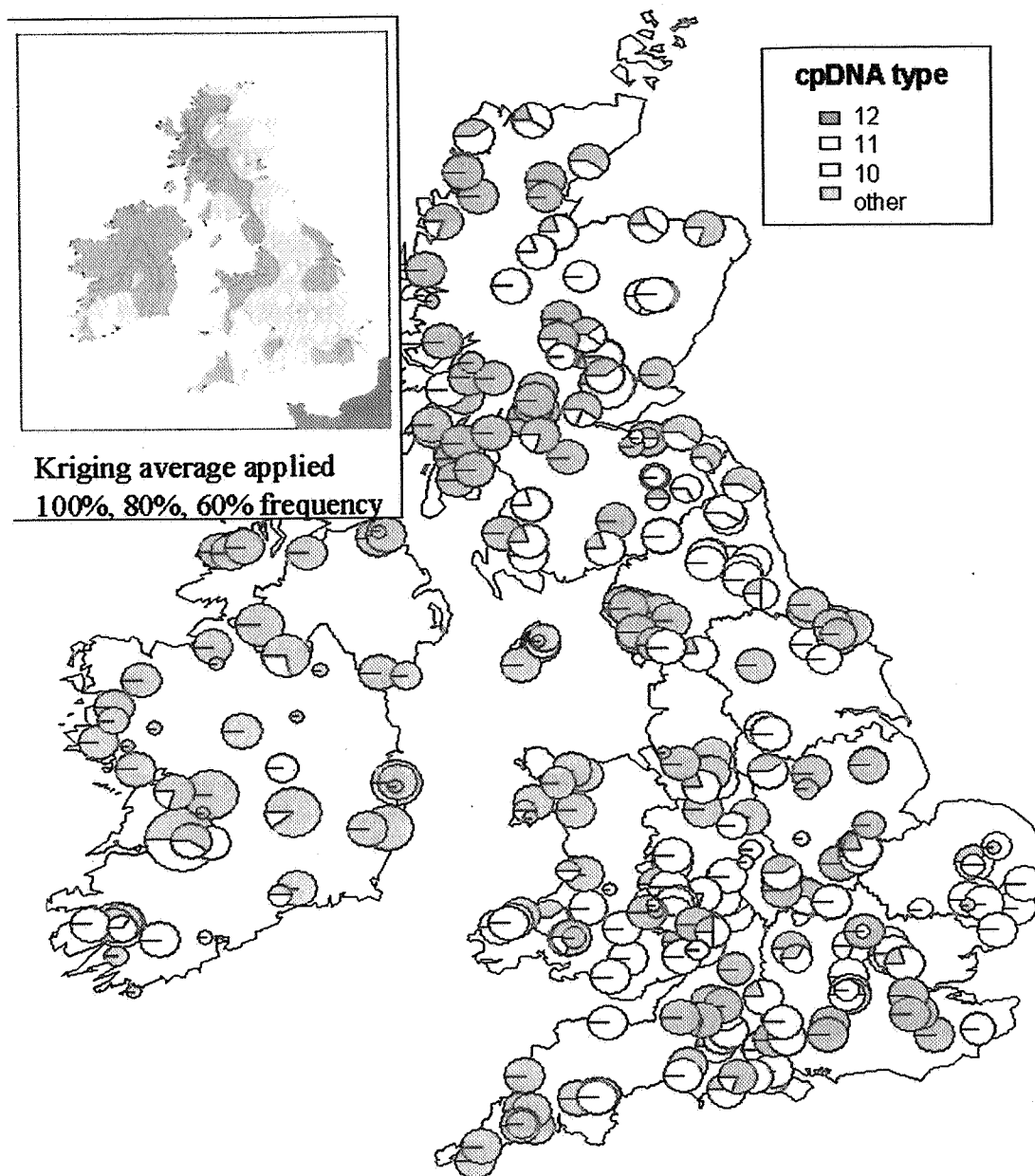
Site	Contributor	Species	Longitude (W)	Latitude (N)	Sample size	Haplotype 12	Haplotype 11	Haplotype 10	Other haplotypes
AIROR	Andy King	mixed	-5.761	57.08	1	0	1	0	0
ALL	Andy King	<i>robur</i>	-2.87	53.37	1	1	0	0	0
ASC	Andy King	<i>robur</i>	-0.65	51.4	4	2	2	0	0
BGP1	Andy King	<i>robur</i>	-1.2	52.67	1	0	0	1	0
BIAU	Andy King	<i>robur</i>	-1.67	54.65	4	1	2	1	0
BREEN	Andy King	mixed	-6.33	55.17	1	1	0	0	0
CAS	Andy King	<i>robur</i>	-1.81	52.58	2	0	0	2	0
CHC	Andy King	<i>petraea</i>	-4.57	52.9	2	1	0	1	0
CHY	Andy King	<i>robur</i>	-0.9	54.12	4	0	4	0	0
CKP	Andy King	<i>petraea</i>	-8.47	51.9	1	0	0	1	0
ELV	Andy King	<i>petraea</i>	-3.55	52.27	1	0	0	1	0
GUE	Andy King	<i>robur</i>	-0.45	51.92	0	0	1	0	0
HDM	Andy King	<i>robur</i>	0.25	52.1	2	0	2	0	0
INVER	Andy King	mixed	-5.65	57.03	1	1	0	0	0
Larling	Andy King	<i>robur</i>	0.9	52.45	2	0	1	1	0
LLA	Andy King	<i>robur</i>	-3.98	51.88	2	2	0	0	0
LLEYN	Andy King	<i>petraea</i>	-4.5	52.85	1	1	0	0	0
REE	Andy King	<i>robur</i>	-0.37	52.78	3	3	0	0	0
SHF	Andy King	<i>petraea</i>	-1.12	53.07	2	2	0	0	0
THM	Andy King	<i>robur</i>	0.83	52.13	1	1	0	0	0
TRAW	Andy King	mixed	-4.05	52.42	2	1	0	1	0
UCE	Andy King	<i>robur</i>	-1.88	52.48	1	0	0	1	0
WIND	Andy King	<i>robur</i>	-0.65	51.45	2	0	2	0	0
Cahir	Jim Martin	<i>petraea</i>	-7.56	52.23	2	1	1	0	0
Collooney	Jim Martin	<i>petraea</i>	-8.31	54.11	1	1	0	0	0
Cootchill	Jim Martin	<i>petraea</i>	-7.05	54.05	1	1	0	0	0
Croagh Patrick	Jim Martin	<i>petraea</i>	-9.4	53.45	1	1	0	0	0
Enniskerry	Jim Martin	<i>petraea</i>	-6.15	53.11	1	1	0	0	0
Foxford	Jim Martin	<i>petraea</i>	-9.06	53.59	1	1	0	0	0
Glengariff	Jim Martin	<i>petraea</i>	-9.35	51.46	1	1	0	0	0
Coed Gorllwyn	Dave Roberts	mixed	-4.04	52.96	5	5	0	0	0
Coed y Rhygen	Dave Roberts	mixed	-3.94	52.9	5	0	4	11	0
Llyn Clywedog	Dave Roberts	mixed	-3.59	52.47	5	5	0	0	0
Middleton	Dave Roberts	mixed	-4.16	51.86	4	4	0	0	0

2002; Kelleher *et al.*, 2004; Muir *et al.*, 2004). This longitudinal structuring has been used as evidence to suggest that there were at least three separate routes of colonisation from northern France into the British Isles, each distinguished by a different cpDNA type (Cottrell *et al.*, 2002). However, a number of studies have differed in their interpretation of colonisation routes into western Britain and Ireland (Petit *et al.*, 2002b; Kelleher *et al.*, 2004).

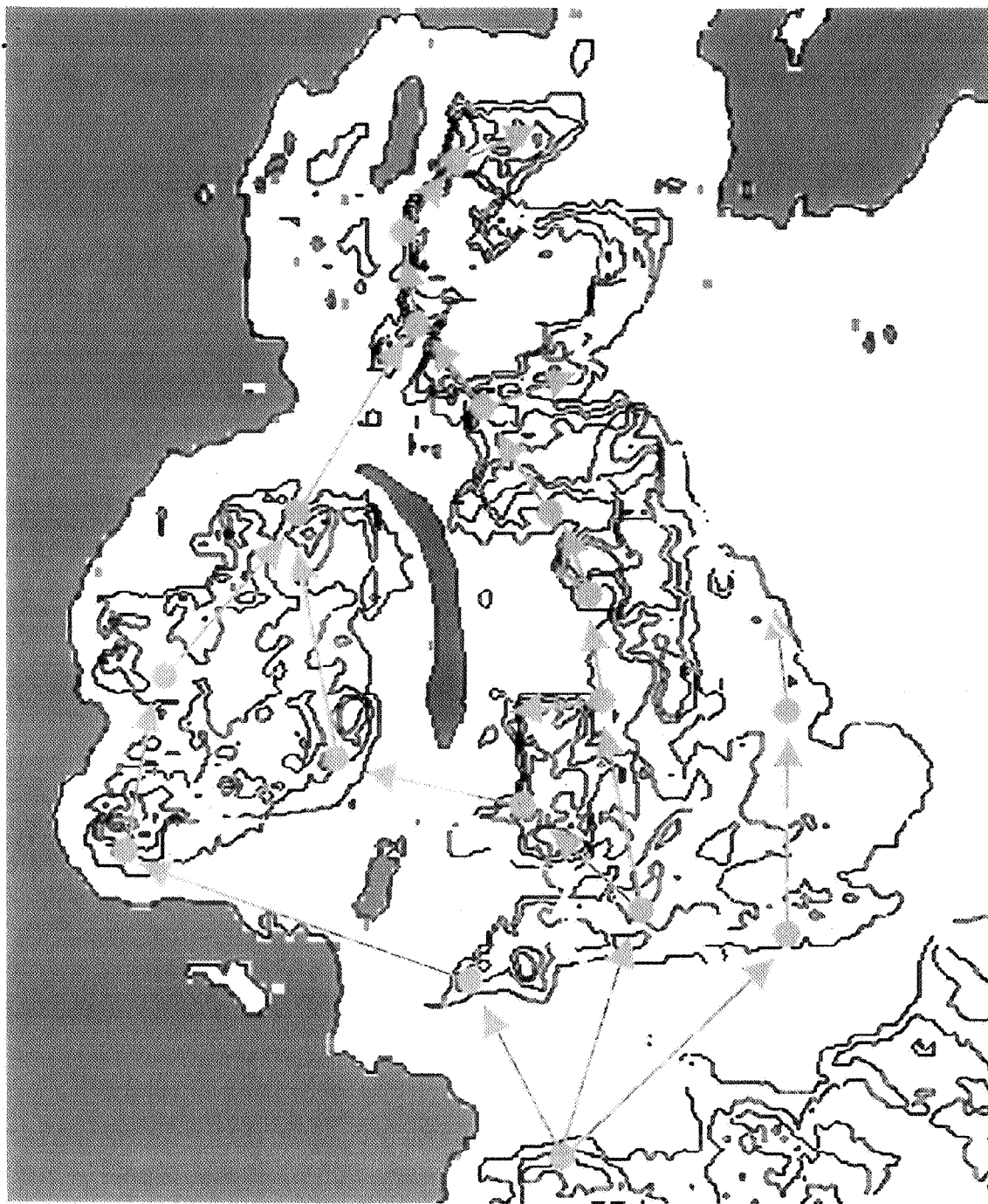
To try to resolve this issue more precisely we have synthesised here all available published cpDNA data for the British Isles (from Cottrell *et al.*, 2002; Petit *et al.*, 2002b; Kelleher *et al.*, 2004; Lowe *et al.*, 2004; Muir *et al.*, 2004), together with some previously unpublished work (Table 1), to produce the most

detailed map currently available of cpDNA variation within Britain and Ireland (describing variation in 1468 trees from 313 autochthonous stands of *Q. robur* and *Q. petraea*, Fig. 1). To facilitate identification of colonisation routes, the frequency of haplotypes 10, 11 and 12 have been kriged across the British Isles landscape (see methods described in Lowe *et al.*, 2004, Fig. 1 inset). This analysis identifies regions where one of the three haplotypes is dominant, occurring in 60, 80 and 100 % of surrounding trees. When considering colonisation route, the topography of a landscape is also important. Oak currently grows to an approximate altitude of 350 m at the latitude of the British Isles, and thus higher mountain ranges are likely to present a significant barrier to colonisation, including the Welsh mountains, Pennines, Scottish Borders and Grampians. Today the British Isles are a series of islands off the European continent, but for much of the postglacial period, Britain and Ireland were part of the European mainland. It is probable that sea depths which today are less than 100 m were land during postglacial colonisation (Brewer *et al.*, 2002; Petit *et al.*, 2002b), but those deeper than 100 m were probably significant water barriers to dispersal, especially the northern and central Irish Sea. Considering the potential mountain and marine barriers, and the minimum distance between Kriged foci dominated by the same chloroplast haplotype as likely colonisation paths, a series of colonisation routes are inferred for the British Isles (Fig. 2).

Using the method outlined above, the direction of the colonisation route of haplotype 11 is approximately south-north and it progresses up the eastern edge of England and Scotland from Kent, to East Anglia (this haplotype was previously identified as 'the East Anglian type' by Ferris and coworkers), to Humberside and up into Aberdeenshire, and is largely as reported by previous workers (Ferris *et al.*, 1998; Cottrell *et al.*, 2002). Haplotype 10 follows a central path through England and populates several regions in south central England, as reported by Cottrell *et al.* (2002). Further north, haplotype 10 has probably colonised east of the Pennines, through passes in the southern uplands of Scotland and around the eastern edge of the Grampians into the Inverness region. Interestingly, this haplotype (10) also appears to have taken a route through Devon, southern Wales and across the Irish Sea north of a deep trench in the Celtic Sea to colonise SE Ireland. Haplotype 12 also displays some interesting patterns which suggest three or four separate invasion routes. One route for haplotype 12 progresses up the eastern side of England, stopping at the East Ridings, and a second route progresses up the western side of England, through the Lake District, into SW Scotland, and up the western coast of Scotland, and is as previously reported by Cottrell *et al.* (2002). There appear to have been two potential colonisation routes into Ireland for haplotype 12, the first via an extreme westerly colonisation route which passed over Cornwall and through low sea level areas of the Celtic Sea into SW Ireland and up the west coast (which is suggested by Kelleher *et al.*, 2004). A second potential route of haplotype 12 is from the mid-west Welsh coast across a land bridge south of the deep Irish Sea trench into SE Ireland and up the east coast (suggested by Petit *et al.*, 2002b, but discredited by Kelleher *et al.*, 2004). From our analysis there appears to be no reason why both these routes could not have populated Ireland with oaks, and they are in addition to a third pathway described earlier, as traced by haplotype 10.



**Fig. 1.** Distribution and frequency of cpDNA haplotypes sampled from 1468 British and Irish oak trees from 313 autochthonous woodlands (Table 1). Size of pie indicates the number of individuals sampled from each population (where for the largest pie  $n = 8$ ). Inset, top left, is a Kriging average of cpDNA haplotype frequency, where green are regions of no overall dominance and the outer circles of the three representative haplotype colours are where that type is found within neighbouring populations at a frequency greater than 60%, additional concentric lines within these areas represent haplotype dominances of 80 and 100%.



**Fig. 2.** Approximate topographic map of the British Isles, showing marine depths greater than 100 m in blue and terrestrial upland areas greater than 350 m in pink (NOAA, 1988). Considering these topographic barriers and using the minimum distance between Kriged foci of haplotype dominance (circles, where haplotype 12 is orange, haplotype 11 is light grey and haplotype 10 is yellow, see Fig. 1), a series of likely terrestrial colonisation pathways is superimposed (arrows).

Finally, considering the topographically complex regions of western England and southern Scotland, which could potentially serve to slow the progress of oak colonisation, the western coast of Scotland could equally have been colonised via an Irish or an English route. These pathways are putative, and based solely on the pattern of cpDNA variation found within contemporary populations, they also do not consider the possibility of exchange across the North Sea. What is clear however, is that Atlantic oakwood populations share a different colonisation history to central and eastern oak populations in the British Isles.

## 2. Speed of Colonisation

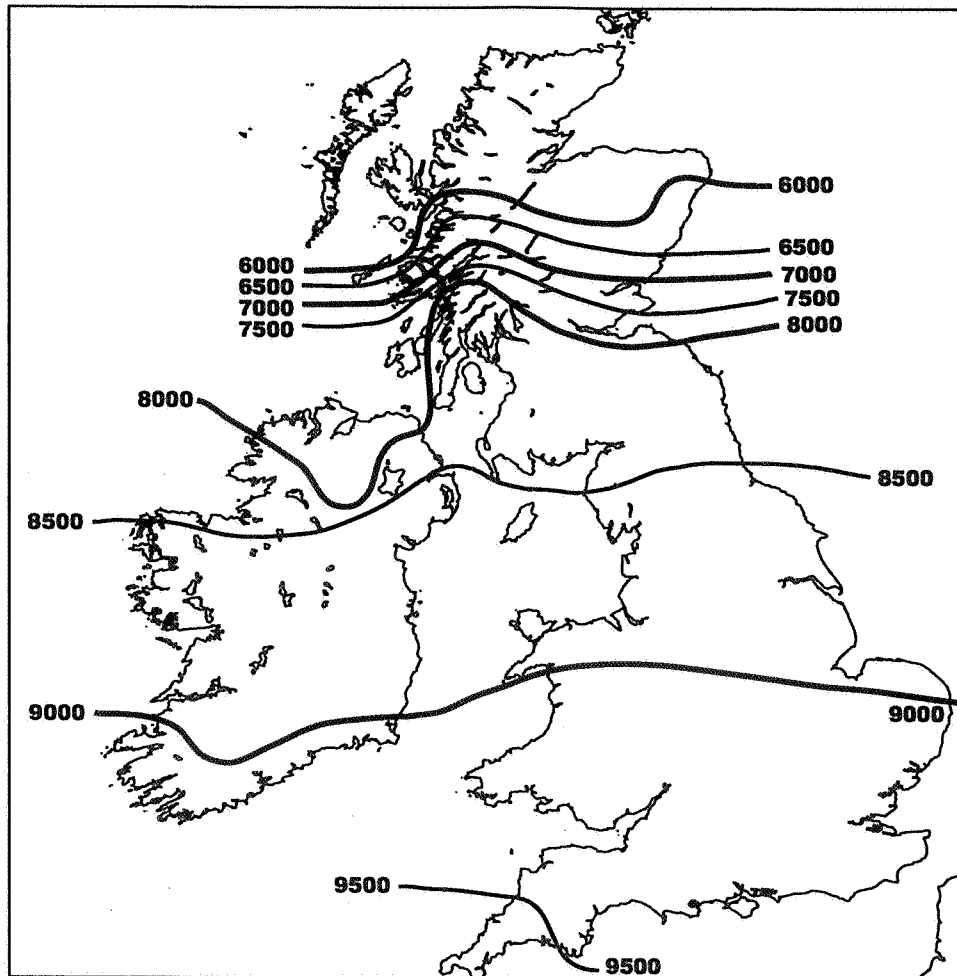
Analyses of sediment cores allow recording of the date at which oak pollen percentages first increase above background noise levels (considered a signature of proximate colonisation by oak populations, but may be an underestimation of the actual date of colonisation Brewer *et al.*, 2002). Using a series of cores, the timing of oak colonisation across Britain and Ireland has been inferred (Birks, 1989; Mitchell, 2002). Integrating these two analyses, we present here a first synthesised map of the timing of oak colonisation, to a 500 year resolution, for the British Isles (Fig. 3). From this map it appears that much of the British Isles was colonised very quickly (in less than 2000 years), with a maximum colonisation speed of 500 m year<sup>-1</sup> through south/central England. However, it took a further 2000 years to colonise Scotland, probably due to topographic and thermal limitations. Whilst a significant leading western edge effect was identified during a European scale analysis of oak pollen cores (Brewer *et al.*, 2002), the effect was only moderately expressed in the British Isles, including Cornwall being the first region to be colonised and a slight leading edge pattern on the west coast of Ireland and the west coast of Scotland. The pollen core data do not give a clear resolution on the likelihood of an English or Irish colonisation route of western Scotland.

## 3. Mode of Colonisation

Many workers have been struck by the apparent 'Reid's paradox' of oaks, which was first articulated as: 'The oak, to gain its present most northerly position in North Britain after being driven out by the cold probably had to travel fully six hundred miles, and this without external aid would take something like a million years' (Reid, 1899). In other words, oak cannot have colonised over the distance and with the speeds it did by gravity dispersal alone. There must have been some other vector of long distance dispersal. To examine these colonisation issues several workers have used a simulation modelling approach (Hewitt, 1996; Ibrahim *et al.*, 1996; Le Corre *et al.*, 1997; Petit *et al.*, 2001; Davies *et al.*, 2004). Model approaches have the advantage of testing *in silico* processes that would take very long time periods in nature or would be simply unobservable due to issues of scale, and are ideal for examining colonisation processes over large distances or time scales. We present here a simplified model of the one used by Davies *et al.* (2004), to describe the main findings of oak colonisation modelling work.

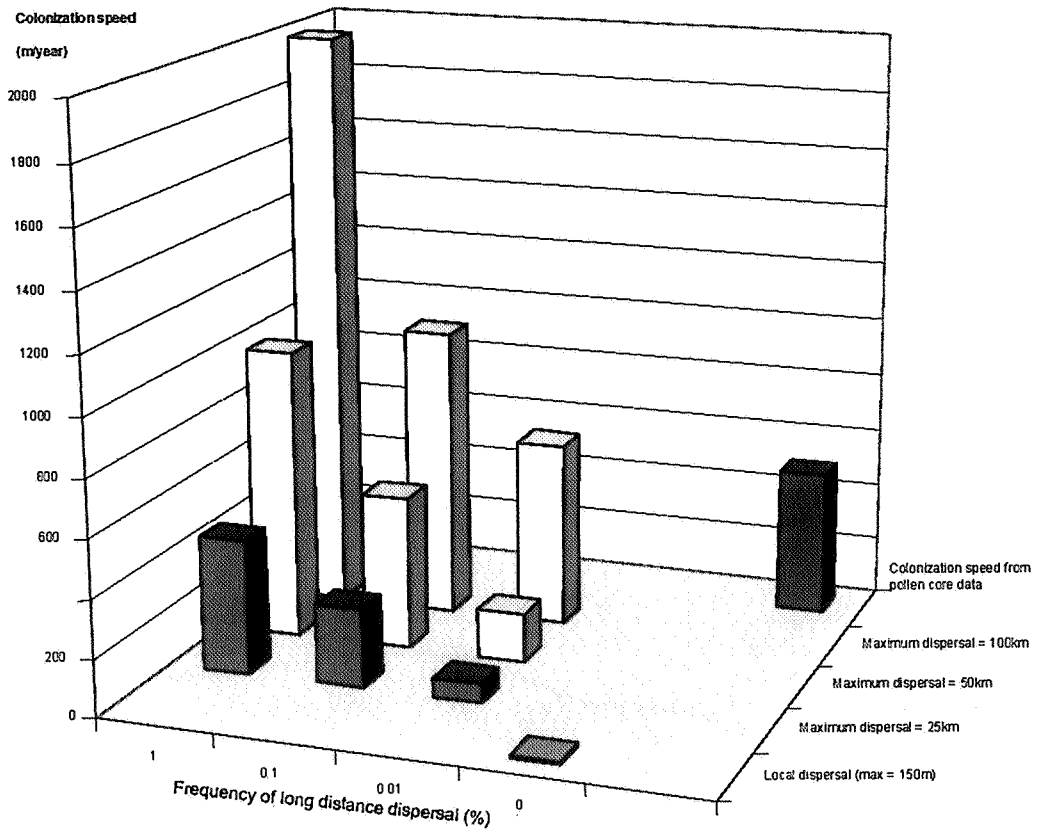
The model landscape exists as a spatial array of cells, 'rolled' so that the bottom edge is continuous to the uppermost edge. Each cell represents a 50 m





**Fig. 3.** Isochrone map showing recolonisation of *Quercus* in the British Isles. The map shows the potential limits of the distribution of the oak through time, with each contour representing its limit for a particular period. The contours cluster toward the north indicating a slowing of the rate of spread. The contours are in 1000 years before present (after Birks, 1989; Mitchell, 2002).

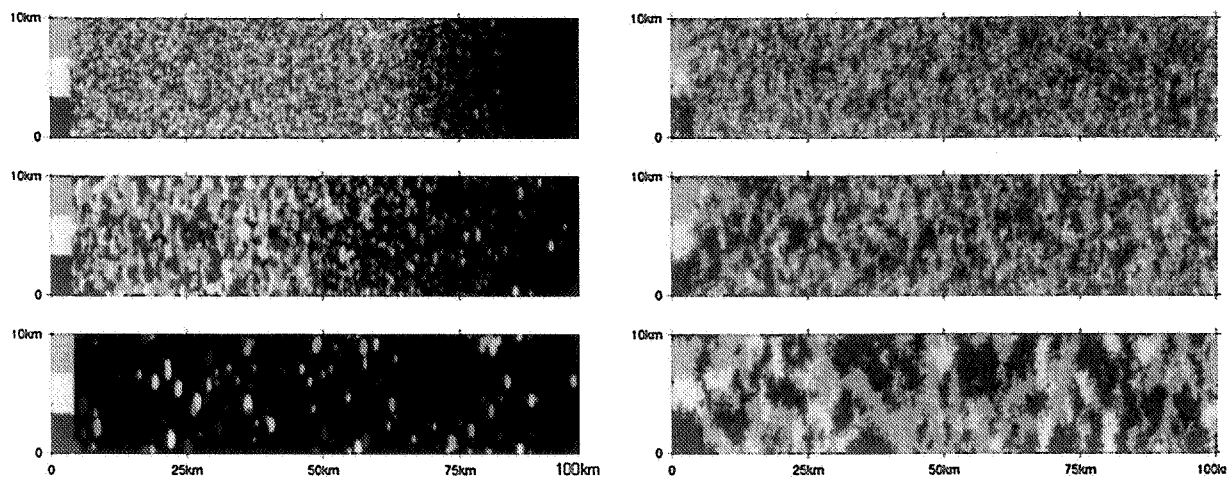
square and can be colonised by a single tree. Each tree has its maturity set at 15 years with an average lifespan of 200 years and a maximum lifespan of 400 years. Seed dispersal occurs annually and is split into two components, local and long distance. The proportion of local ( $p$ ) to long distance dispersal ( $1-p$ ) can be altered as can the maximum distance for dispersal ( $d_{max}$ ). For each mature tree the model determines the number of locally dispersed offspring based on a Poisson distribution with the mean number of local offspring  $pNO$  and the number of long distance offspring with mean  $(1-p)NO$  (where  $NO$  is the average number of offspring per tree). The position of offspring is determined by a dispersal curve



**Fig. 4.** Estimates of colonisation speed made for different modelling scenarios, and compared to the maximum colonisation speed estimated from the pollen core data. For comparison the colonisation speed achievable under a gravity-only dispersal model (the Reid's paradox scenario) is also indicated.

derived from field estimates of seed dispersal (see below). If more than one offspring is predicted to occur in any cell then the resulting a single offspring is randomly chosen from those predicted. Offspring have the same cpDNA as their maternal parent. For full details of the model see Davies *et al.* (2004).

Several simulations were run to test which modelled frequency and distance of long distance dispersal events provide the closest approximation to the observed speed of colonisation obtained from the pollen core record (Fig. 4). Using parameters which simulate purely local gravity dispersal (maximum dispersal distance of 150 m), only very slow colonisation speeds are possible (approx 10 m year<sup>-1</sup>). Under this gravity-only dispersal, oaks would have taken approximately 100,000 years just to colonise from the southern tip to the most northerly extent of the British Isles. This gravity-only estimate is much closer to Reid's prediction than the rate observed in the pollen core, therefore some form of long distance dispersal must have been involved. Under the conditions of the

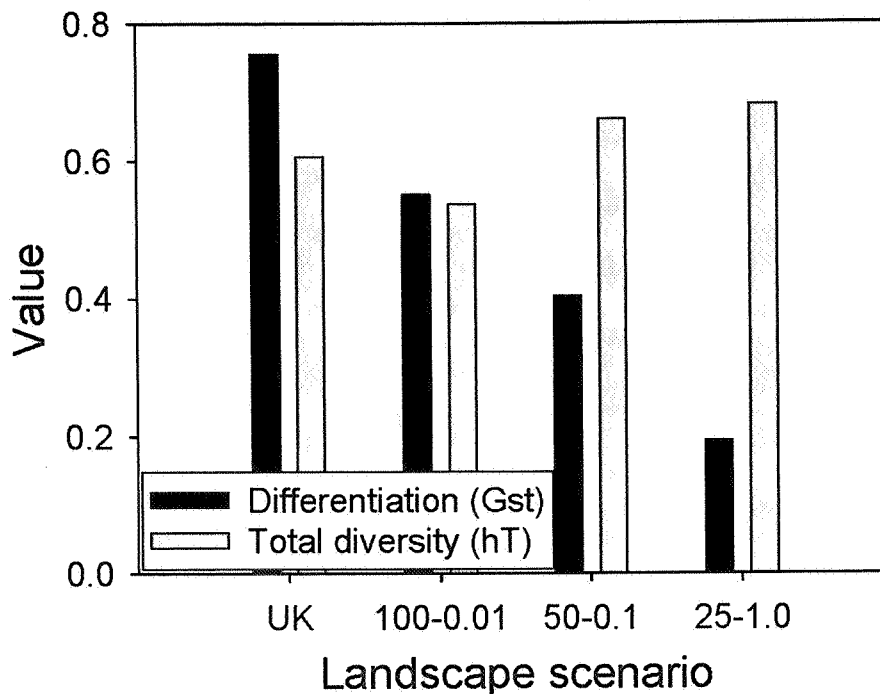


**Fig. 5.** Landscape colonisation starting from left hand block colonising open cells according to frequency and distance. Topmost plots disperse 1 % of seeds up to 25 km, middle plots 0.1 % up to 50 km and bottom plots 0.01 % up to 100 km. The plots have been left to run for 200 years (left) and 10,000 years (right).

model, the maximum rate of colonisation observed in the pollen core record for the British Isles ( $500 \text{ m year}^{-1}$ ) is matched by either relatively common (1%), shorter distance dispersal (up to 25 km), moderately rare dispersal (0.1%) over moderate distances (up to 50 km), or very rare dispersal (0.01%) over long distances (up to 100 km).

To discern between these three scenarios, which differ in the frequency and distance of long distance dispersal, it is useful to examine the pattern of cpDNA variation within contemporary populations (Fig. 1). The distribution of cpDNA variation in the British Isles presents large patches in which populations are dominated by a single haplotype. In some cases these patches can be tens of km across. This pattern was first observed in a regional cpDNA survey of north-western France (Petit *et al.*, 1997), but has been subsequently found in all areas of European oak that have been surveyed in sufficient detail (*e.g.* descriptions of the British Isles, Cottrell *et al.*, 2002; Italy, Fineschi *et al.*, 2002; Scandinavia, Jensen *et al.*, 2002; and the Iberian Peninsula, Olalde *et al.*, 2002). Petit *et al.*, (2001) describe this pattern as 'a patch-work quilt' and speculate that it is due to the dynamics of seed dispersal, with the dominant haplotype patches being 'footprints of colonisation'. To test this idea further the cpDNA genetic structure established following landscape colonisation by the three model scenarios was examined (Fig. 5). Only the most infrequent (0.01%), longest distance (100 km) dispersal strategy produces large patch-like patterns of cpDNA variation. In addition, the diversity statistics calculated for 10 populations of 25 individuals sampled from the colonised landscape after 10,000 years under this scenario (and repeated for the other two scenarios) were most similar to the actual UK data (Fig. 6).

Additional modelling work by Davies *et al.* (2004) has found that a number



**Fig. 6.** Comparison of diversity and differentiation statistics calculated for different modelling scenarios and compared to actual values from the contemporary British Isles oak sample

of environmental features can further influence speed of colonisation and genetic structuring within simulated landscapes. In particular, cool temperatures that cause fruiting failure or inhibit seed germination and establishment lead to slower colonisation speeds, reduced genetic diversity and increased patch size in areas that had been colonised under these conditions. In addition, large topographic barriers and an asymmetrical colonising front can cause further loss of diversity. To examine if any of these factors had impacted on the diversity or structure of British Isles populations, the total cpDNA data set was reanalysed after stratification according to date of colonisation of populations based on the pollen core data (Fig. 7. inset). A trend towards lower diversity and particularly lower differentiation was found for the more northerly, slower colonised populations, and so some of these environmental features could be impacting on genetic diversity during colonisation.

#### 4. Estimating Maximum Seed Dispersal Distances

Whilst modelling studies undertaken so far appear to tie in well with colonisation processes inferred from phylogeographic analyses, there is still a requirement for long distance dispersal over some considerable distance (50-100 km depending on the study, Petit *et al.*, 2001; Davies *et al.*, 2004). Since oaks are predominantly gravity dispersed, this magnitude of dispersal distance is surprising. Even secondary dispersal by squirrels and ground mammals along the forest floor

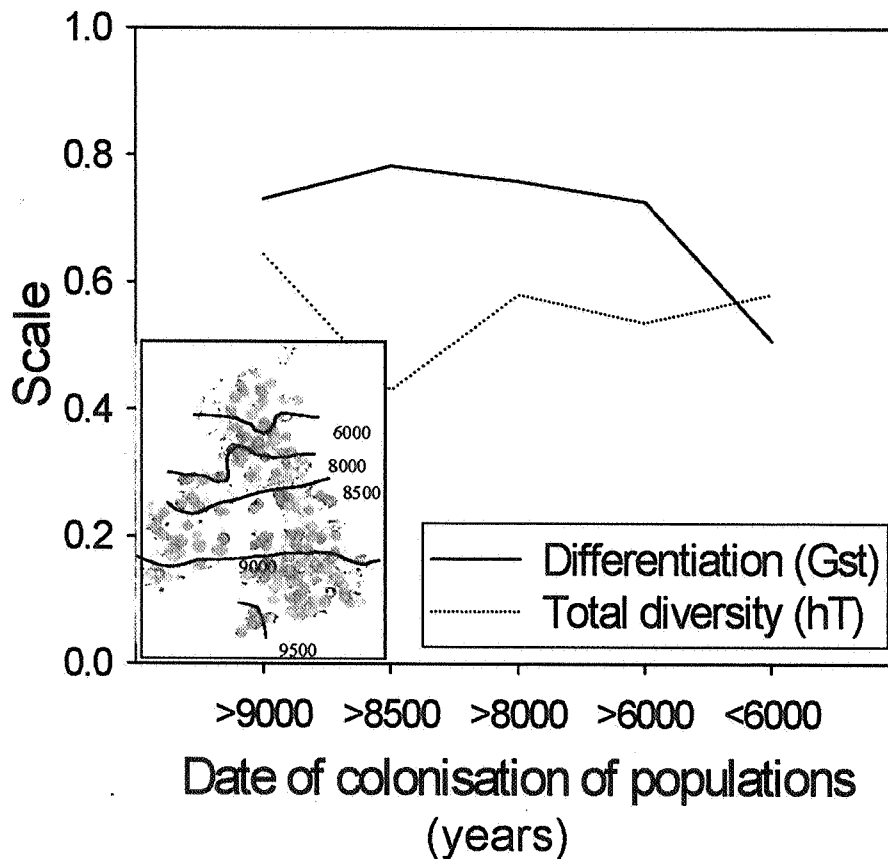
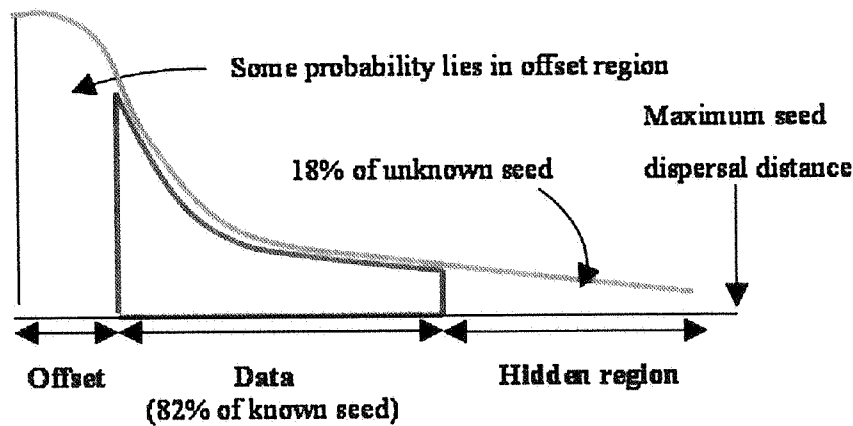


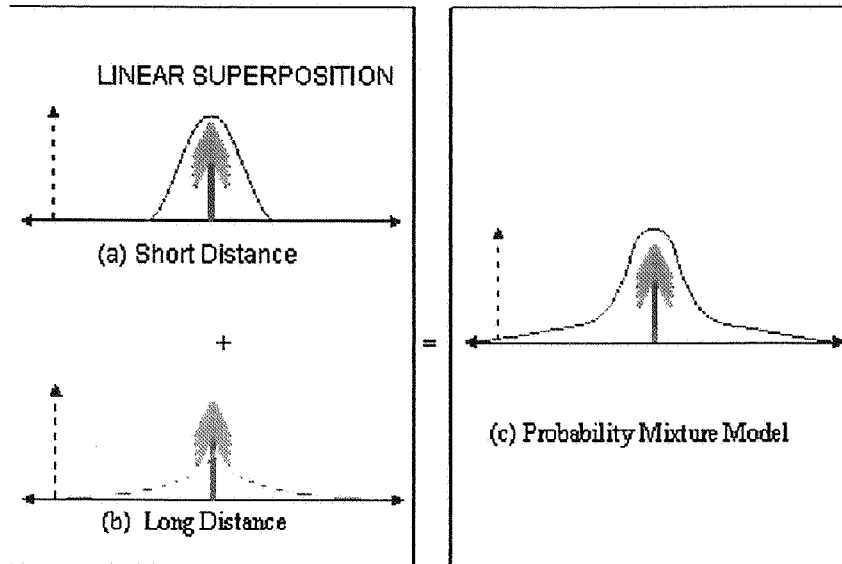
Fig. 7. Regional estimates of population genetic diversity and differentiation for British Isles populations stratified by colonisation timing, see inset of superimposed first date of colonisation estimated from the pollen core record.

would only be likely to redistribute acorns a maximum of one kilometer from the source tree, with the majority of dispersals being much shorter than this (Sork *et al.*, 1983; Sork, 1984). Perhaps the only agents capable of really long distance dispersal are birds and rare climatic storms. Bossema (1979) reports that jays have been observed to carry acorns more than one kilometer and hurricanes are widely known to be able to carry small and medium sized objects over several, and even tens of kilometers. In this respect, the simulation results of Davies *et al.* (2004) found that a similar speed of colonisation and cpDNA population genetic diversity and structure resulted from either very rare, single long distance dispersal events that can occur randomly every fruiting season, or following periodic mass dispersal (every 100 seasons) by a number of individuals, as long as, averaged over time, the overall frequency of dispersal events was identical. Thus, it seems that both birds and/or hurricanes could potentially be the very rare, very long distance dispersal agents of oaks.



**Fig. 8.** Schematic of known and hidden components of dispersal as calculated from parentage exclusion analysis of seedlings from an exhaustively sampled plot of oak trees. Also indicated are the necessary data adjustment procedures, including resetting the zero to offset the hidden probability distribution function (HPDF), and normalising the area beneath the curve according to the proportion of known seed dispersal (in this case 82% of the data), and creating the HPDF.

Quantifying such components of long distance dispersal has proven difficult (Bullock & Clarke, 2000). The two parts of dispersal curves that are important for ecologists are the shapes at the origin and in the tails (Williamson, 2002). Some authors have assumed two types of dispersal curves to combine seed dispersal near and far, convexity near the source and fat tail far from the source (Clark *et al.*, 1999; Bullock & Clarke, 2000). This is a stratified movement model approach (Gomez, 2003), where local movement occurs by one mechanism but other mechanisms are invoked for long distance dispersal (Cain *et al.*, 2000). To examine this area further, we present a method for estimating very long seed dispersal distances within a contemporary oak population, and is similar to other recently published models (Austerlitz *et al.*, 2004; Grivet *et al.*, 2005). Using a geo-referenced population of adult oak trees and seedlings, which has undergone molecular parentage assignment, the method estimates the maximum short and long distance components of seed dispersal using a normalised linear superposition of normal and exponential probability distribution functions (PDF) creating a bimodal probability mixture model. This method allows an estimation of seed dispersal beyond the scale of the plot used for the parentage assignment. First a decay curve is fitted to the within-plot dispersal data (seedlings with one or two parental assignments involving trees inside the plot, the 'known' component). Then by using the proportion of seed dispersal which has occurred from outside the plot (seedlings with no parental assignments for trees inside the plot, the 'hidden component' of dispersal), an intersect with the x axis is calculated according to the 'hidden' area under the curve, see schematic in Fig. 8

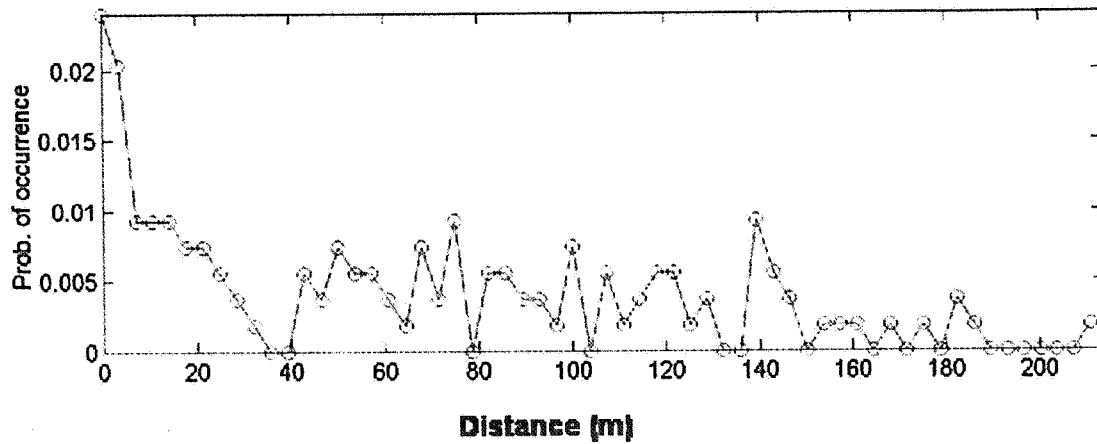


**Fig. 9.** Schematic of the linear superimposition of (a) short distance and (b) long distance dispersal processes to attain (c) the probability mixture model.

To test the method, seed dispersal data were obtained for seedlings growing on the floor of a stand located in the middle of *La Petite Charnie* forest, a 700 ha state forest in the northwest of France, close to Le Mans. This stand is 5 ha in area comprising a mixture of 296 mature oak trees (167 *Quercus robur*, 124 *Q. petraea* trees, five trees with intermediate morphology) and 161 seedlings sampled along a regular grid. All trees were genotyped for six microsatellite loci, Qpzag104, Qpzag36, Msq4, Qpzag1/5, Qpzag9, and Msq13 (Streiff *et al.*, 1998, 1999). The exclusion probabilities are 100% for single parent and parent pairs. Assignment of single parents and parent pairs for each seedling was made with the maximum likelihood method described by Gerber *et al.* (2000), using the software FaMoz (Gerber *et al.*, 2003; 2004; unpublished data). Of the 161 seedlings, 131 had at least one parent identified within the 5 ha plot area. In this instance the closest parent was assumed to be the mother and the distance to that parent was calculated from coordinate data.

#### *The model*

The model is built around two main dispersal processes, short and long distance seed dispersal. Short distance dispersal by gravity results in seeds falling randomly around the maternal tree, and can be approximated to a normal distribution. Whereas the long distance component is best equated to an exponential function. The model combines these two functions into a probability mixture model (Fig. 9), for precise details of the mathematical functions (see Appendix 1). Probability mixture models can then be fitted against the shifted



**Fig. 10.** Shifted HPDF, distance (m) against probability of occurrence of original data with the percentage of the known data distribution = 82% (an area of 0.820) and offset = 2.75.

hidden probability density function of the original data (Fig. 10). A large number of probability mixture models can be generated, but the one with the best fit to the field data (with ks-statistic = 0.2667 and area = 0.82), was a normalised linear superposition of a normal distribution (short distance dispersal)  $\sigma_E = 61$  and an exponential distribution (long distance dispersal)  $\mu_N = 161$ . The optimum distribution of the model is plotted in Fig. 11.

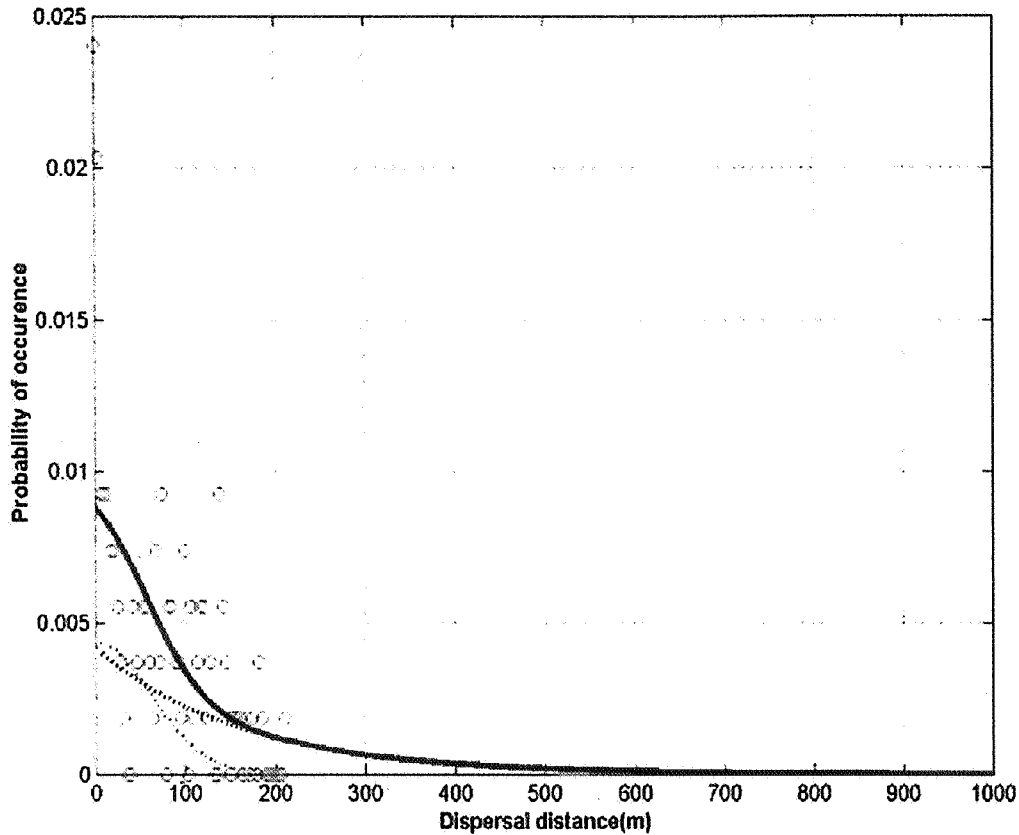
Table 2 lists the maximum seed dispersal distances predicted for a range of probabilities under the best fitting normal and exponential functions. Considering the number of seeds tested, it is reasonable to consider a probability of  $1 \times 10^{-5}$  for the location of the maximum dispersal distance. Under this model, maximum short-distance dispersal, estimated by fitting a normal distribution (gravity dispersal) is 215 m ( $\pm 57$  m), and maximum long-distance dispersal, estimated by fitting an exponential function is 972 m ( $\pm 259$  m). It was notable that the combined model fitted the data better than the exponential function alone, indicating that a combination of dispersal components provides a better descriptor of field data than a single component. This observation also supports the idea that there are a number of seed dispersal mechanisms operating in the field, each of which has a different dispersal profile.

#### *Limitations of the model analysis*

The maximum seed dispersal distance estimated from the field data using the mixture probability model, even including the maximum error limit, is only 1231 m, and this falls far short of simulation predictions of the likely distances involved during oak postglacial colonisation (between 50 and 100 km; Petit *et al.*, 2001; Davies *et al.*, 2004).

There may be several reasons for a discrepancy between the estimates of





**Fig. 11.** Best fit hidden probability distribution function (HPDF) for the evaluated seed dispersal data. The HPDF is a normalised linear superimposition of a normal distribution (short distance dispersal,  $\sigma E = 61$ ) and an exponential distribution (long distance dispersal,  $\mu N = 161$ ), with an area = 0.82 and ks-statistic = 0.2667. Where circles represent the HPDF of the field data, the solid curve represents the optimum probability mixture model, and dotted lines are the normal and exponential distributions that make up the optimum model.

seed flow data from contemporary populations and dispersal distances predicted by simulation models:

1. The number of seedlings used to estimate the dispersal distance is relatively small and many more seedlings, spread over a wider spatial scale ought to be sampled to obtain an improved estimate of long distance dispersal. The very long seed dispersal events predicted by the modelling procedures are so rare that during a single sampling season very few such events are likely to take place. Thus, at the scale of plot we have sampled, we may not be able to accurately estimate a curve fitting function for a long distance component. Field validation of the very long seed dispersal distances predicted by

**Table 2.** Probability of finding a seed up to a distance from a particular tree for the single best fitting HDPF solution, for a range of probabilities ( $Pr$ ) the distance (to the nearest m) and estimation error (0.2667, equal to the ks-fit) are indicated separately for both short distance (normal function,  $\sigma_E = 61$ ) and long distance (exponential function,  $\mu_N = 161$ ) dispersal distributions.

$Pr$	Short distance (error)		Long distance (error)	
$10^{-3}$	105	$\pm 28$	299	$\pm 61$
$10^{-4}$	168	$\pm 45$	599	$\pm 160$
$10^{-5}$	213	$\pm 57$	970	$\pm 259$
$10^{-6}$	250	$\pm 67$	1341	$\pm 358$

simulation models may simply be too labour intensive to be viable. However, a number of landscape-scale genotyping projects are underway for oaks, data from which could be used to tackle this issue.

2. Within contemporary ecosystems, where many forest niches are already filled, long distance dispersal that leads to successful establishment is much rarer than in the more open habitats which were typical during periods of maximal post glacial colonisation. As a result, these long distance dispersal events may only lead to colonisation events during periods of mass migration (*e.g.* following rapid climatic change), or during colonisation of open, but suitable landscapes (*e.g.* following ecological disruptions).
3. A third possibility is that we may be inaccurately estimating long distance seed dispersal from the modelling results. It is possible that a finer-scale cpDNA patch structure exists in the UK which we have so far failed to identify. However, a much finer scale survey of cpDNA variation (Petit *et al.*, 1997) found a similar scale of patch structure to the one identified by this study. But it is certainly true that further cpDNA mapping and modelling work would help improve these estimations.

## 5. Future Directions

The sheer number of scientists working on the postglacial colonisation and population scale seed dispersal of oaks has stimulated much scientific advancement. New methods for phylogenetic analysis, parentage analysis and simulation modelling have been developed by various workers to tackle questions relating to the ecology and evolution of oaks, and these have been widely taken up by researchers outside the oak group.

Perhaps one of the most important direct products of the oak research is a European-wide integrated map of cpDNA variation. This cpDNA map has already been used to: validate the origin of oak wood used in the cooperage industry (Deguilloux *et al.*, 2003, 2004); examine whether superior-form oaks and seed orchards are of local origin (Lowe *et al.*, 2004); and test whether there has been a change in seed sourcing strategy during rotational planting in the Forest of Dean (Cottrell *et al.*, 2004). Some of the most exciting developments in this area lie in the possibility of using this map for molecular archeological

interpretation, where the historical usage of oak can be inferred by accessing DNA from ancient timbers found at important archeological sites. The techniques already exist to perform this analysis (Dumolin-Lapègue *et al.*, 1999; Deguilloux *et al.*, 2002), and are being applied to a number of projects. One such project is the identification of the source origin of timbers used in the main and refit constructions of Henry VIII's flag ship, the Mary Rose (A.J. Lowe, unpublished data), which was raised from the sea bed in a remarkably well-preserved state in 1982.

### Acknowledgements

The authors would like to thank David Roberts and Jim Martin for contributing material and data, and Peter Levy, Graham Muir and Stephen Cavers for discussion and comments on the work. This work was part funded by the EU (Grant Number FAIR1-CT95-0297), Forestry Commission and Natural Environment Research Council.

### References

- Austerlitz, F., Dick, C.W., Dutech, C., Klein, E.K., Oddou-Muratorio, S., Smouse, P.E. & Sork, V.L. (2004). Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology* **13**, 937-954.
- Birks, H.J.B. (1989). Holocene isochrone maps and patterns of tree-spreading in the British Isles. *Journal of Biogeography* **16**, 503-540.
- Bossema, I. (1979). Jays and oaks: an eco-ethological study of a symbiosis. *Behaviour* **70**, 1-117.
- Brewer, S., Cheddadi, R., de Beaulieu, J.-L., Reille, M. & Data Contributors (2002). The spread of deciduous *Quercus* throughout Europe since the last glacial period. *Forest Ecology and Management* **156**, 27-48.
- Bullock, J.M. & Clarke, R.T. (2000). Long distance seed dispersal by wind: measuring and modelling the tail of the curve. *Oecologia (Berlin)* **124**, 506-521.
- Cain, M.L., Milligan, B.G. & Strand, A.E. (2000). Long-distance seed dispersal in plant populations. *American Journal of Botany* **87**, 1217-1227.
- Clark, J.S., Silman, M., Kern, R., Macklin, E. & Hillerislambers, J. (1999). Seed dispersal near and far: patterns across temperate and tropical forests. *Ecology* **80**, 1475-1494.
- Cottrell, J., Munro, R.C., Tabbener, H., Gillies, A.C.M., Deans, D. & Lowe, A.J. (2002). Distribution of Chloroplast DNA variation in British Oaks (*Quercus robur* and *Q. petraea*): the influence of postglacial colonisation and human management. *Forest Ecology and Management (Special Issue)* **156**, 181-196.
- Cottrell, J.E., Samuel, C.J.A. & Sykes, R. (2004). The effect of planting date on the species and chloroplast DNA haplotype composition of oakwoods in the Forest of Dean. *Forestry* **77**, 99-106.
- Davies, S., White, A. & Lowe, A.J. (2004). An investigation into effects of long-distance seed dispersal on organelle population genetic structure and colonization rate: a model analysis. *Heredity* **93**, 566-576.
- Deguilloux, M.F., Pemonge, M.H. & Petit, R.J. (2002). Novel perspectives in

- wood certification and forensics: dry wood as a source of DNA. *Proceedings of the Royal Society of London Series B-Biological Sciences* **269**, 1039-1046.
- Deguilloux, M.F., Pemonge, M.H., Bertel, L., Kremer, A. & Petit, R.J. (2003). Checking the geographical origin of oak wood: molecular and statistical tools. *Molecular Ecology* **12**, 1629-1636.
- Deguilloux, M.F., Pemonge, M.H. & Petit, R.J. (2004). DNA-based control of oak wood geographic origin in the context of the cooperage industry. *Annals of Forest Science* **61**, 97-104.
- Dumolin-Lapegue, S., Demesure, B., Fineschi, S., Le Corre, V. & Petit, R.J. (1997). Phylogeographic structure of white oaks throughout the European continent. *Genetics* **146**, 1475-1487.
- Dumolin-Lapègue, S., Pemonge, M.H., Gielly, L., Taberlet, P. & Petit, R.J. (1999). Amplification of oak DNA from ancient and modern wood. *Molecular Ecology* **8**, 2137-2140.
- Ferris, C. (1996). Ancient history of the common oak. *Tree News (Autumn)*, 12-13.
- Ferris, C., Oliver, R.P., Davy, A.J. & Hewitt, G.M. (1993). Native oak chloroplasts reveal an ancient divide across Europe. *Molecular Ecology* **2**, 337-344.
- Ferris, C., Oliver, R.P., Davy, A.J. & Hewitt, G.M. (1995). Using Chloroplast DNA to trace postglacial migration routes of oaks into Britain. *Molecular Ecology* **4**, 731-738.
- Ferris, C., King, R.A., Väinölä, R. & Hewitt, G.M. (1998). Chloroplast DNA recognizes three refugial sources of European oaks and suggests independent eastern and western immigrations to Finland. *Heredity* **80**, 584-593.
- Fineschi, S., Turchini, D., Grossoni, P., Petit, R.J. & Vendramin, G.G. (2002). Chloroplast DNA variation of white oaks in Italy. *Forest Ecology and Management* **156**, 103-114.
- Gerber, S., Chabrier, P. & Kremer, A. (2003). FaMoz: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. *Molecular Ecology Notes* **3**, 479-481.
- Gerber, S. *et al.* (2004). Mesure directe des flux de gènes en forêt. Actes du 4ème Colloque National *Le patrimoine génétique: la diversité et la ressource*, La Châtre, 14-16 Octobre 2002: pp. 349-368.
- Gerber, S., Mariette, S., Streiff, R., Bodénès, C. & Kremer, A. (2000). Comparison of microsatellites and AFLP markers for parentage analysis. *Molecular Ecology* **9**, 1037-1048.
- Gomez, J.M. (2003). Spatial patterns in long-distance dispersal of *Quercus ilex* acorns by jays in a heterogeneous landscape. *Ecography* **26**, 573-584.
- Grivet, D., Smouse, P.E. & Sork, V.L. (2005). A novel approach to an old problem: tracking dispersed seeds. *Molecular Ecology* **14**, 3585-3595.
- Hewitt, G.M. (1996). Some genetic consequences of ice ages and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247-276.
- Ibrahim, K.M., Nichols, R.A. & Hewitt, G.M. (1996). Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* **77**, 282-291.

- Jensen, J.S., Gillies, A.C.M., Csaikl, U., Munro, R.C., Madsen, S.F., Roulund, H. & Lowe, A.J. (2002). Chloroplast DNA variation within the Nordic countries. *Forest Ecology and Management (Special Issue)* **156**, 167-180.
- Kelleher, C.T., Hodkinson, T.R., Kelly, D.L. & Douglas, C.G. (2004). Characterisation of chloroplast DNA haplotypes to reveal the provenance and genetic structure of oaks in Ireland. *Forest Ecology and Management* **189**, 123-131.
- Le Corre, V., Machon, N., Petit, R.J. & Kremer, A. (1997). Colonization with long-distance seed dispersal and genetic structure of maternally inherited genes in forest trees: a simulation study. *Genetic Research, Cambridge* **69**, 117-125.
- Lowe, A.J., Munro, R.C., Samuel, S. & Cottrell, J. (2004). The utility and drawbacks of chloroplast DNA for identifying native British oak stands. *Forestry*, **77**, 335-347.
- Mitchell, F. (2002). Natural invaders: the postglacial tree colonisation of Ireland. In: *Biological Invaders: the Impact of Exotic Species* (ed. C. Moriarty & D.A. Murray), pp. 2-12. Dublin: Royal Irish Academy.
- Muir, C., Lowe, A.J., Fleming, C.C. & Vogl, C. (2004). High nuclear genetic diversity, high levels of outcrossing and low differentiation among remnant populations of *Quercus petraea* at the margin of its range in Ireland. *Annals of Botany* **93**, 691-697.
- NOAA. National Geophysical Data Center (1988). Underlying topography from ETOPO5 – Data Announcement 88-MGG-02, Digital relief of the Surface of the Earth, Boulder, Colorado.
- Olalde, M., Herrán, A., Espinel, S. & Goicoechea, P.G. (2002). White oaks phylogeography in the Iberian Peninsula. *Forest Ecology and Management* **156**, 89-102.
- Petit, R.J., Kremer, A. & Wagner, D.B. (1993). Geographic structure of chloroplast DNA polymorphisms in European oaks. *Theoretical and Applied Genetics* **87**, 122-128.
- Petit, R.J., Pineau, E., Demesure, B., Bacilieri, R., Ducouso, A. & Kremer, A. (1997). Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences, USA* **94**, 9996-10001.
- Petit, R.J., Bialozyt, R., Brewer, S., Cheddadi, R. & Comps, B. (2001). From spatial patterns of genetic diversity to postglacial migration in forest trees. In: *Integrating Ecology and Evolution in a Spatial Context* (ed. J. Silvertown & J. Antonovics), pp. 295-318. Oxford: Blackwell Science.
- Petit, R.J. *et al.* (2002a). Chloroplast DNA variation in European white oaks: Phylogeography and patterns of diversity based on data from over 2,600 populations. *Forest Ecology and Management (Special Issue)* **156**, 5-26.
- Petit, R.J. *et al.* (2002b). Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management (Special Issue)* **156**, 49-74.
- Reid, C. (1899). *Origin of the British Flora*. London: Dalau.
- Sork, V.L., Stacey, P. & Averett, J.E. (1983). Utilization of red acorns in non-bumper crop year. *Oecologia (Berlin)* **59**, 49-53.
- Sork, V.L. (1984). Examination of seed dispersal and survival in red oak, *Quercus*

- rubra* (Fagaceae), using metal tagged acorns. *Ecology* **65**, 1020-1022.
- Streiff, R., Labbe, T., Bacilieri, R., Steinkellner, H., Glossl, J. & Kremer, A. (1998). Within-population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. *Molecular Ecology* **7**, 317-328.
- Streiff, R., Ducouso, A., Lexer, C., Steinkeller, H., Glossl, J. & Kremer, A. (1999). Pollen dispersal inferred from paternity analysis in a mixed stand of *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. *Molecular Ecology* **8**, 831-841.
- Williamson, M. (2002). Overview and synthesis: the tale of the tail. In *Dispersal Ecology* (ed. J.M. Bullock, R.E. Kenward & R.S. Hails), pp. 431-443. Oxford: Blackwell Science Publishers.

### Appendix 1

A short distance dispersal function can best be described by a normal 'probability density function' (PDF, Fig. 9a). Seeds can fall either side of the tree, located at the origin, at negative and positive distance values of ( $x$ ). The probability that a seed falls a distance ( $x$ ) from the tree reduces as the distance is increased following a normal distribution. The area beneath the normal probability density function is unity and so the total probability of finding a seed within the ( $x$ ) bounds of the normal curve is one. The equation for the normal distribution is:

$$\Pr(x | \mu_N, \sigma_N) = \frac{1}{\sigma_N \sqrt{2\pi}} e^{-\frac{(x-\mu_N)^2}{2\sigma_N^2}} \quad \text{Equation 1}$$

Integrating Equation 1) between zero and a distance  $x_1$  gives the probability of finding a seed within this region for a normal distribution with parameters  $\mu_N$  and  $\sigma_N$ . The parameter  $\mu_N$  is a location parameter and determines the location of the central peak. Thus,  $\mu_N = 0$  will locate the normal distribution at the origin.

Actual seed dispersal data are incorporated as the distance from the seedling to the foot of the tree to which it has been assigned genetically, in any direction, and is represented as a positive value. Thus, the negative part of the normal distribution is not required for the model. In addition, the trunk of the tree would be located at the origin (see Fig. 8), thus setting  $\mu_N = 0$ , in equation 1, and simplifying equation 1 to that below:

$$\Pr(x | 0, \sigma_N) = \frac{1}{\sigma_N \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_N^2}} \quad \text{where } x \geq 0 \quad \text{Equation 2}$$

Integrating Equation 2 between zero and a distance  $x_1$  is proportional to the probability of finding a seed within this region, which is greater than or equal to zero, with a location parameter,  $\mu_N = 0$ , and a spread parameter of  $\sigma_N$ . It should also be noted that the total area under the curve is 0.5 since it only represents half of the normal distribution, where  $x \geq 0$  is being used.

The second physical process modelled here is long distance. An

'Exponential distribution' can be used to model 'rare events' and is suitable for approximating such long distance events. The two dimensional schematic in Fig. 9b shows how a seed would be distributed for an exponential process. The probability of a seed falling a distance ( $x$ ) decreases exponentially as the distance increases. The exponential distribution is given in Equation 3 and is defined for positive values of ( $x$ ). As mentioned in section 2.2, the seeds fall at  $x \geq 0$  so it is not necessary to adapt this function.

$$f(x | \mu_E) = \frac{1}{\mu_E} e^{-\frac{x}{\mu_E}} \quad \text{where, } x \geq 0 \quad \text{Equation 3}$$

The exponential distribution also has a parameter ( $\mu_E$ ) which describes the spread of the distribution at the origin. The area under the exponential PDF is unity.

The short and long distance dispersal components can be linearly superimposed as probabilities, for the same values of ( $x$ ), of each of the separate models, to generate a probability mixture model (Fig. 9c). Equation 4 describes the probability of finding a seed within the bimodal probability mixture model described above.

$$\Pr(x | \mu_N = 0, \sigma_N, \mu_E) = \frac{\frac{1}{\sigma_N \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_N^2}} + \frac{1}{\mu_E} e^{-\frac{x}{\mu_E}}}{1.5} \quad \text{where } x \geq 0 \quad \text{Equation 4}$$

Where ( $\mu_N$ ) identifies the  $\mu$  parameter of the normal distribution which is zero. And ( $\mu_E$ ) is the  $\mu$  parameter of the exponential distribution, and is calculated for the original seed dispersal data (Fig. 10).

For 30 seedlings, no potential parent was identified within the plot and so a 'hidden probability density function' (HPDF) method was developed to take account of seed dispersal distances greater than the sample plot dimensions (Fig. 8). A Kolmogorov-Smirnoff test was used to compare a range of bimodal probability mixture models with the HPDF to estimate 'goodness of fit'.

1  
2  
3