An investigation into effects of long-distance seed dispersal on organelle population genetic structure and colonization rate: a model analysis

S Davies[1,2], A White[2] and A Lowe[1,3]

[1]CEH Edinburgh, Bush Estate, Penicuik, Midlothian EH26 OQB, UK; [2]Departments of Mathematics, Heriot-Watt University, Edinburgh EH14 4AS, UK; [3]School of Life Sciences, University of Queensland, Brisbane, Queensland 4072, Australia

A simulation-based modelling approach is used to examine the effects of stratified seed dispersal (representing the distribution of the majority of dispersal around the maternal parent and also rare long-distance dispersal) on the genetic structure of maternally inherited genomes and the colonization rate of expanding plant populations. The model is parameterized to approximate postglacial oak colonization in the UK, but is relevant to plant populations that exhibit stratified seed dispersal. The modelling approach considers the colonization of individual plants over a large area (three 500 km × 10 km rolled transects are used to approximate a 500 km × 300 km area). Our approach shows how the interaction of plant population dynamics with stratified dispersal can result in a spatially patchy haplotype structure. We show that while both colonization speeds and the resulting genetic structure are influenced by the characteristics of the dispersal kernel, they are robust to changes in the periodicity of long-distance events, provided the average number of long-distance dispersal events remains constant. We also consider the effects of additional physical and environmental mechanisms on plant colonization. Results show significant changes in genetic structure when the initial colonization of different haplotypes is staggered over time and when a barrier to colonization is introduced. Environmental influences on survivorship and fecundity affect both the genetic structure and the speed of colonization. The importance of these mechanisms in relation to the postglacial spread and genetic structure of oak in the UK is discussed.
Heredity (2004) **93,** 566–576. doi:10.1038/sj.hdy.6800555
Published online 22 September 2004

Keywords: colonization; seed dispersal; population structure; *Quercus*

Introduction

Plant species exhibit a variety of mechanisms of seed dispersal. The majority of seeds will be dispersed locally around the maternal parent in relation to the speed and direction of wind for wind-dispersed seeds, the crown size for gravity-dispersed seeds and disperser behaviour for animal dispersed seeds. There is also potential for long-distance dispersal through a variety of mechanisms. These include rare climatic events (eg hurricanes), long-distance animal dispersal (eg migratory birds) or long-distance river transport.

Long-distance events are not normally measured in empirical studies of seed dispersal. Even with the recent increase in molecular studies targeting maternally inherited, and therefore seed-dispersed, genomes (eg chloroplast DNA (cpDNA) in Angiosperms), sample sizes are usually too small to reveal detailed information on the dynamics of these potentially very rare events. However, the importance of rare long-distance dispersal events to the colonizing ability of a species has been highlighted by a number of studies and has been influential in determining colonization patterns following glacial periods (Skellam, 1951; Hewitt, 1996; Ibrahim

et al, 1996; Le Corre *et al*, 1997; Cain *et al*, 1998, 2000; Clark, 1998; Clark *et al*, 1999; Petit *et al*, 2000). For forest trees, the timings and direction of postglacial colonization have been investigated using analysis of fossil pollen data and simulation studies (eg Birks, 1989; Hewitt, 1996, 1999; Le Corre *et al*, 1997). These studies indicate that expansion from refugial centres following glacial retreat is often rapid and so must have involved long-distance dispersal events that set up colonies well ahead of the main species advance. The maternally inherited genome of such colonists can then come to dominate leading populations, since later migrants will need to invade established populations at carrying capacity (Hewitt, 1993; Ibrahim *et al*, 1996; Le Corre *et al*, 1997).

The postglacial colonization of oaks across Britain provides a good case study to examine the influence of long-distance dispersal on migration rates. Within the British Isles there are no recorded refugial oak populations and there is a very good pollen record against which the timing of events can be set. Colonization speed has been determined from pollen core analysis and shows that the spread of oak was particularly rapid, reaching speeds up to 500 m/yr (Huntley and Birks, 1983; Birks, 1989; Bennett, 1997). Oaks, like other forest trees, are long-lived and therefore their present genetic variability is likely to reflect ancient population and colonisation events, and the seed dispersal history can be uncovered by examining cpDNA variation since it is nonrecombinant and maternally transmitted in oaks

(Ferris *et al*, 1993; Dumolin *et al*, 1995; Ferris *et al*, 1995). Recent studies have provided detailed spatial information on the distribution of cpDNA variability in European oak populations (eg Ferris *et al*, 1993, 1995; Cottrell *et al*, 2002; Petit *et al*, 2002a). The results of phylogeographical analysis of 2613 European populations using cpDNA variation show a strong phylogeographic structure with an east–west division of variation with multiple clines relating to the colonisation routes through Europe from three refugial populations in the southern peninsulas of Iberia, Italy and the Balkans (Hewitt, 1999; Brewer *et al*, 2002; Petit *et al*, 2002b). In addition, most mutations appear to have been generated prior to postglacial expansion (Ferris *et al*, 1993, 1995; Hewitt, 1999; Petit *et al*, 2002b), so present cpDNA structure can be seen as a consequence of seed dispersal only. This detailed information about the spread of oak makes it a system amenable to mathematical modelling.

Models for the range expansion of oak, and other plant species, historically used normal or short-tailed dispersal functions fitted to local data on seed dispersal (Skellam, 1951; Birks, 1989; Cain *et al*, 1998; Higgins and Richardson, 1999), yet these studies failed to capture the rate of postglacial expansion observed in plant populations. More recently, simulation models have been used to evaluate the importance of long-distance dispersal in the spread of plants. Comparisons of different dispersal kernels show that leptokurtic dispersal kernels (where the tail of the distribution is positive at large distances) are required to produce the colonization speeds seen in plant populations (Shigesada *et al*, 1995; Hewitt, 1996; Ibrahim *et al*, 1996; Le Corre *et al*, 1997; Clark, 1998; Clark *et al*, 1999; Cain *et al*, 1998, 2000). Stratified dispersal kernels use two components to describe dispersal, one represents the distribution of the majority of seed dispersal around the maternal parent and the second component represents rare long-distance dispersal. It is argued that it is better to represent dispersal by two different mechanisms so the central tendency and the tail can be varied independently (Clark, 1998; Clark *et al*, 1999; Pakeman, 2001). Stratified dispersal kernels have been used to simulate realistically the ecto/endozoochorous dispersal of woodland herbaceous species (Pakeman 2001); the rapid colonization of the forest herb *Asarum canadense* (Cain *et al*, 1998); the behaviour of *Hieracium pilosella* (Winkler and Stocklin, 2002); and the dispersal of *Calluna vulgaris* and *Erica cinera* by wind (Bullock and Clark, 2000). Simulation models based on the oak system concluded that long-distance dispersal events not only allowed for rapid expansion but they were also the most important factor determining the spatial genetic structure of maternally inherited genes (Le Corre *et al*, 1997; Petit *et al*, 2000). What is now required is an understanding of how the dynamics produced by leptokurtic seed dispersal interact with other physical and environmental factors to influence the genetic structure within populations.

In this study, we will construct a spatial simulation model to investigate the effect of long-distance dispersal on the spread and genetic structure of plant populations under different dispersal and landscape scenarios. These scenarios have been developed with particular reference to the colonisation of oak in the British Isles, but will be relevant to other species with leptokurtic seed dispersal. A key difference between our study and previous

simulations of plant dispersal (Hewitt, 1996; Ibrahim *et al*, 1996; Le Corre *et al*, 1997; Petit *et al*, 2000) is the spatial scale employed. Utilisation of spatial models in previous studies simulated large grid areas (100 km × 300 km in Le Corre *et al*, 1997, 100 km × 1320 km in Petit *et al*, 2000), but the number of cells within this area was relatively small, chosen for computational efficiency, and represented demes rather than individual trees. These studies therefore made *a priori* assumptions about genetic dynamics within each deme. Our study considers a similar landscape area (300 km × 500 km) but divides this area into patches that are 50 m × 50 m in size (and assumed to be occupied by a single tree, Cottrell *et al*, 2002), and so the genetic dynamics and patterning are produced from individual behaviour. This is possible in our study due to the increased processor power now available and by approximating one of the spatial dimensions by using rolled transects (see discussion of Figure 1 later). In addition, our study will investigate a number of other potentially important influences on dispersal (and resulting colonization speed and genetic structure), which remain untested by previous work (eg catastrophic climatic events, barriers to dispersal and environmental effects on fecundity and survivorship). The model is parameterized to approximate the dynamics for oak populations in Britain and a number of scenarios are considered and assessed for genetic structure and colonization rate.

1. *Default*: continuous long-distance dispersal occurring at low probability.
2. *Hurricanes*: long-distance dispersal operates in certain years only, but with the same average probability of long-distance dispersal over time as in the default run. This represents freak climatic events that may be capable of dispersal over long-distances (eg rare storms, hurricanes) and can be used to test the robustness of genetic structure and colonization rate to the period of long-distance dispersal events.
3. *Staggered start*: where a time lag operates between the initial seeding of different haplotypes. Analysis of the pollen core record (Birks, 1989) and current cpDNA distribution (Cottrell *et al*, 2002) indicates that oak with a single haplotype first entered south west Britain around 9500 BP (Brewer *et al*, 2002; Cottrell *et al*, 2002) with the other two haplotypes colonizing the south east later.
4. *Barrier*: part of the landscape is made uncolonizable. Many of the mountainous areas in Britain are likely to have presented significant barriers to the dispersal of oak, for example, those greater than 350 m in height such as the Welsh mountains, Pennines, Borders and Grampians. Today the British Isles is a series of islands off the European continent; however, for much of the postglacial period, Britain and Ireland were part of the mainland. It is probable that with seabed depths of less than 100 m the land was colonizable at this time, but where the seabed depth was deeper than 100 m this was a significant barrier to dispersal, especially the northern and central Irish Seas.
5. *Latitudinal climate change*: to model the effects of climate on fitness we reduce fecundity or seed survival as latitude increases. This simulates a decrease in fitness as environmental conditions

become less favourable as a species reaches its range limits. This will allow us to test whether latitudinal climate effects on fitness can account for the lower diversity and lower differentiation observed in the more northerly, slower colonised populations (Cottrell et al, 2002). We acknowledge that climatic conditions may have varied at all latitudes during the colonization of oak and that this may have impacted on the genetic composition currently observed. We do not consider this here but focus on how the difference in climatic conditions between northerly and southerly latitudes affects fitness.

These experiments allow us the unique opportunity to test the effects of different mechanisms on colonization speed and population diversity and differentiation. By conducting these tests at a scale roughly equivalent to the size of England, and with dispersal attributes approximating those of oak, it will be possible to hypothesise which mechanisms were most influential in producing the distribution of cpDNA diversity observed in British oak populations. However, the results should only be considered qualitatively to assess trends in the effects of different mechanisms and should not be considered a wholly accurate representation of the oak system.

Methods

Population dynamics and dispersal
We assumed that seed dispersal occurred annually and it is represented by a stratified two-component dispersal kernel (Le Corre et al, 1997; Clark, 1998) to approximate local and long-distance dispersal. The proportion of local dispersal (p) to long-distance dispersal ($1-p$) can be altered as can the maximum distance for dispersal (d_{max}). We assume local dispersal occurs within a 150 m radius and is represented by a linear (decreasing) probability generating function (see Equation (1)). It is difficult to find estimates for dispersal parameters for colonizing oak populations, and so the estimates for local dispersal are derived from cpDNA data from established populations (Cottrell et al, 2002; J Cottrell and S Gerber, personal communication). Long-distance dispersal is represented by a uniform probability generating function between 150 m and d_{max} (see Equation (1)). There was no data on the maximum distance that an acorn may be dispersed, and so d_{max} is estimated around a maximum of 25–130 km (Petit et al, 1997) and accounts for various mechanisms that may be responsible for rare seed dispersal up to this distance. The probability generating function, y, with distance, x, from the parental plant is represented as follows:

$$y(x) = \begin{cases} \frac{2p}{150} - \frac{1-p}{d_{max}-150} - \frac{x}{150}\left(\frac{2p}{150} - \frac{2(1-p)}{d_{max}-150}\right) & x \leq 150 \\ \frac{1-p}{d_{max}-150} & 150 < x \leq d_{max} \end{cases}$$
(1)

The maturity of trees was set at 15 years with an average lifespan of 200 years and a maximum life span of 400 years. We assumed that each mature tree has a mean total number of offspring (N_0) each year (seeds that will successfully germinate), if unaffected by competition from other individuals for free grid cells. The proportion of seeds dispersed locally was p; therefore, the mean number of local offspring was pN_0 and the mean number of long-distance offspring was $(1-p)N_0$. The dispersal kernel then determined the distance at which an offspring was located from its parent, and this, when combined with the dispersal direction (chosen at random between 0 and 360°), determined the grid cell where the offspring potentially grew. If the selected grid cell was vacant, then the seed became a tree. If more than one successful seed was predicted to reach an empty cell, then the resulting offspring was chosen at random, based on proportions of the seeds present. All offspring had the same chloroplast haplotype as their parent.
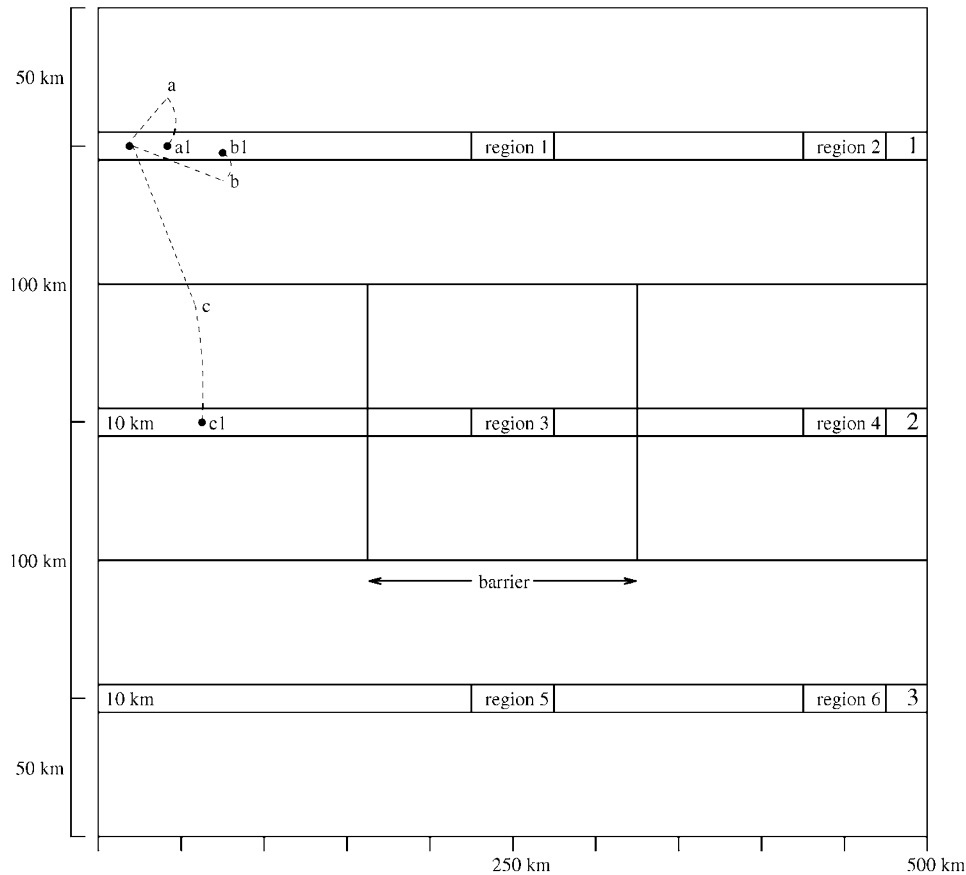
The spatial set-up
We used an individual-based spatial simulation model with a spatial scale relevant to the scale of England (300 km × 500 km). This scale was too large to cover with grid cells so, instead, we modelled cells along three 10 km wide by 500 km long 'latitudinal transects' (see Figure 1), divided into 50 m square grid cells that could be colonised by a single tree. The three latitudinal transects were positioned horizontally in the centre of three adjacent 100 km × 500 km sections. If a dispersal event moves a seed outside of a latitudinal transect, but not out of the section along which the transect runs, the seed is repositioned into the original latitudinal transect (Figure 1, seeds a and b). If a long-distance dispersal event occurred that was sufficient to cross the boundary into the adjacent section, the seed was repositioned into the latitudinal transect in that section (see seed c in Figure 1). This spatial set-up allowed a representation of a very large area while still retaining the detail of an individual-based model. The level of competition experienced by individuals when using this method compares with using a 'full' version of the model (where the whole area contains grid cells), since although migrants are not considered from individuals outside of the three transects, this is counterbalanced by reinserting the migrants which leave the transects (by the rolling process). The transect method described here has been compared with a 'full' version of the model (albeit over a smaller 300 km × 150 km scale) and the results for the genetic statistics and colonisation speed in the full model lay within the range for the five replicates using the rolled transect method (with the default parameters) and close to the average values over the five replicates.

Initial conditions
In a nationwide study of British oaks (Cottrell et al, 2002), 98% of all trees sampled had one of three haplotypes of Iberian origin with the remaining 2% consisting of two different haplotypes of possibly non-natural origin; therefore, we used three haplotypes in the model. Each latitudinal transect was seeded with one haplotype to randomly fill 2% of the first 10 km approximating the clinal chloroplast distribution in the north of France/ south of England prior to colonization (Cottrell et al, 2002; Petit et al, 2002b). These trees were assigned a random age up to the maximum of 400 years and N_0 was chosen to be 1.

The following terms are used throughout the analysis and are defined below for clarity.

Figure 1 Spatial set-up and positioning of the three latitudinal transects. The figure shows the effect of rolling the edge of the latitudinal transects on dispersal, the regions used for data extraction (region 1–6) and position of the barrier to colonization (see later). Rolling the transects allows for representation of dispersal from trees within the 100 km × 500 km sections but outside the 10 km × 500 km latitudinal transects. For example, if a seed is dispersed to point a, it will be dispersed into the latitudinal transect at point a1 (similarly for b to b1). If a seed is dispersed sufficiently far that it reaches the adjacent section such as point c, it will be dispersed to the latitudinal transect in that section at point c1. Seeds that disperse outside the 300 km × 500 km arena are assumed lost. The left-hand side of the figure represents south and the right-hand side north; thus, latitude increases from left to right.

Region: We sampled haplotypes in six regions as outlined in Figure 1.

Population: In all, 20 populations were sampled from different parts of each region.

Latitudinal transect: The three, rolled 10 km × 500 km transects (see Figure 1).

Column 1 (Column 2): The combination of regions 1, 3, 5 (2, 4, 6) (see Figure 1). Statistics were either averaged over these regions or the populations from these regions were pooled and statistics determined for the combined populations.
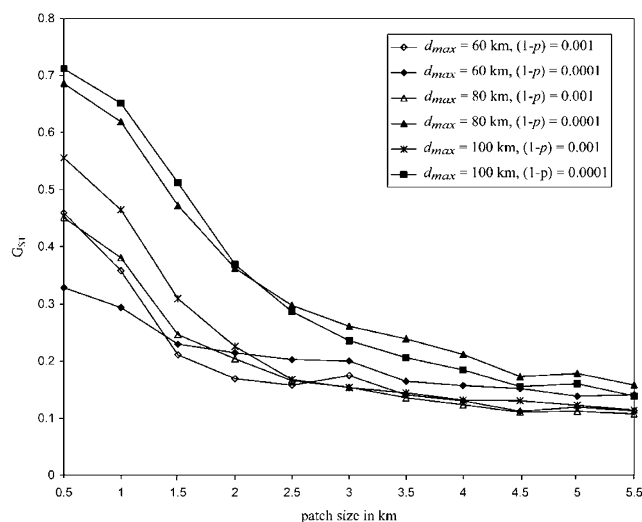
Preliminary study

Initial tests were performed on the default set-up of the model to determine the values of d_{max} and p that result in colonization speed and genetic structure that best represent British oak populations. We appreciate that many dispersal factors will have influenced the current genetic structuring of oak populations (not just those in the default set-up) and therefore focus the analysis in this study on understanding how the results that include different dispersal factors compare with the default scenario rather than how the different scenarios compare

to current data. For the initial tests, we used a single run of the model for 2000 years with three different lengths for maximum long-distance dispersal ($d_{max} = 60$, 80 and 100 km) in combination with two different probabilities for long-distance dispersal ($(1-p) = 0.0001$ and 0.001). The speed of colonisation was calculated by determining how long it takes for a column in the grid to become 10% colonised, which approximately corresponds to the likelihood of a pollen grain being present in a pollen core sample. Diversity statistics were determined (according to Nei, 1973, see also Pons and Petit, 1995) for average population (H_S) and total (H_T) genetic diversity and differentiation (G_{ST}) at a number of patch sizes to determine which was the most appropriate sample scale. Model data were extracted from regions 2, 4 and 6 (Figure 1), with 20 populations of 10 individuals randomly sampled for 10 different patch sizes in each region.

Changes in differentiation for varying patch size indicate the scale of haplotype aggregation. An ideal sampling strategy is one that picks up the differences between the diversity and differentiation measures, by sampling on a scale within haplotype patches and comparing samples on a scale that is greater than the

distance between haplotype patches (an area containing trees of a single haplotype). Changes in G_{ST} for different d_{max} and p values (Figure 2) show that G_{ST} was still increasing at the smallest patch size in all runs; therefore, haplotype patch size was approximately $0.5\,km^2$ or below (sampling from a population in a patch size smaller than this would have been problematic due to the small number of trees) and differentiation levelled out for a patch size greater than 3 km. Therefore, we sampled results using a patch size of $0.5\,km \times 0.5\,km$ with greater than 3 km between populations sampled. This is also roughly equivalent to the scale on which experimental populations were sampled (Cottrell *et al*, 2002).

Comparing the results in Table 1 to those for the British oak population (Cottrell *et al*, 2002), we conclude that values of $d_{max} = 100\,km$ and $(1-p) = 0.0001$ provide the most concordance for colonization speed and diversity measures between experimental and modelled populations and were used for the remaining simulations (unless otherwise stated).
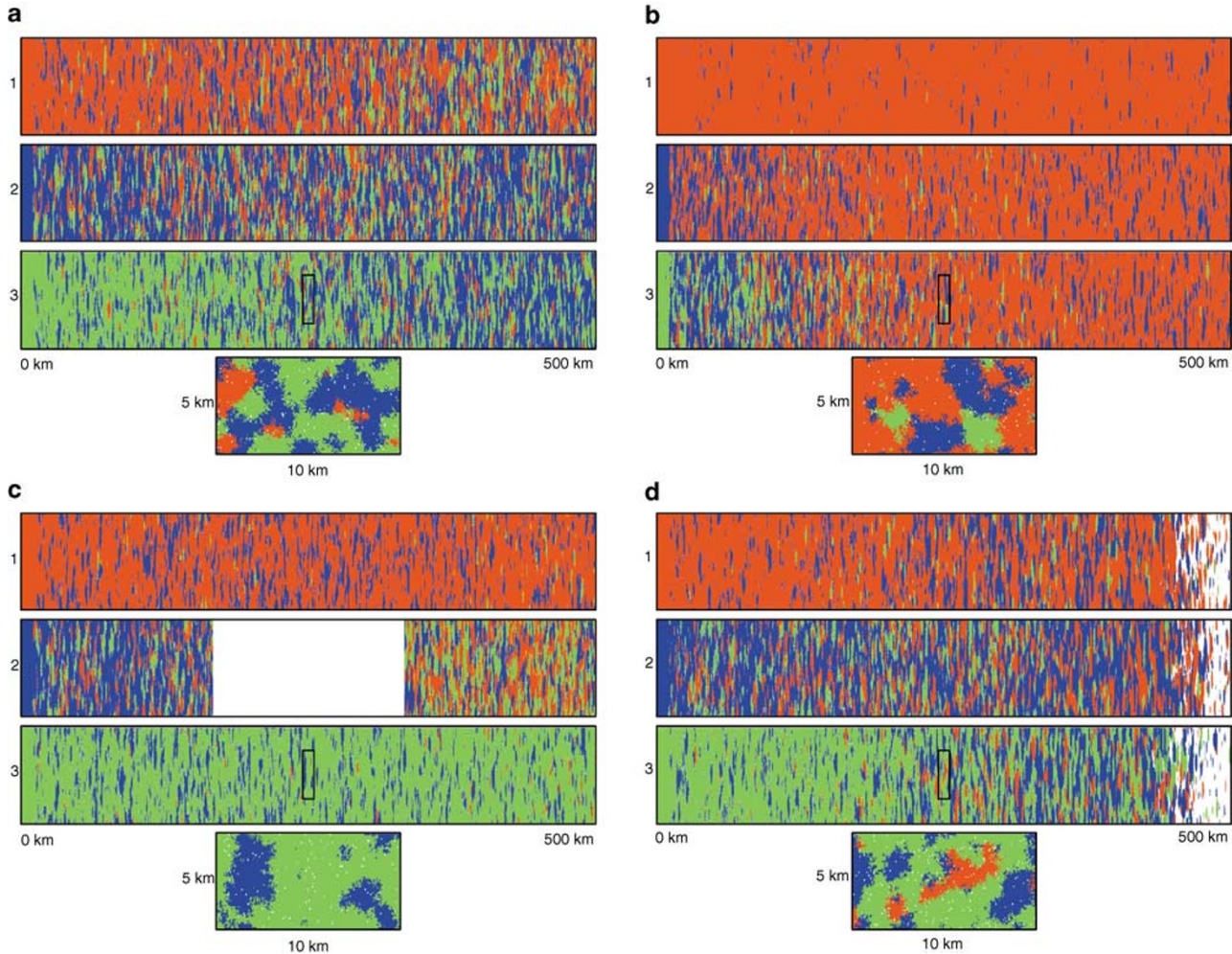
Results

For each scenario detailed below, five replicates of the model were run for 2000 years.

Default scenario

Figure 3a shows that the haplotypes form a patchy structure with dominance of the initially seeded haplotype decreasing with latitude (over the five runs the average $H_S = 0.16$ in column 1 and 0.17 in column 2. The range over the five replicates around these average values is shown in Figure 4a). Within-population diversity is low, indicating that a single haplotype dominates within each $0.5\,km \times 0.5\,km$ sample patch. Average regional diversity is relatively high, indicating that the regions themselves contain a mixture of the three haplotypes. There is a slight trend of increasing within-population and average regional diversity with latitude (average regional diversity in column 1 is 0.5 and 0.56 in column 2, Figure 4a). Total diversity for the pooled populations in columns 1 and 2 are close to its maximum value, indicating a near equal proportion of the three haplotypes over the three transects. The individual regions and pooled populations both show genetic structure due to the spatially patchy nature of the populations with the average within-region $G_{ST} = 0.68$ and pooled $G_{ST} = 0.75$. The sample populations in each region are dominated by one haplotype but the regions themselves contain a mixture of all haplotypes; so differentiation remains high within regions.

Figure 2 Changes in G_{ST} with patch size at different lengths of long-distance dispersal ($d_{max} = 60$, 80 or 100 km) and with a probability of 0.001 or 0.0001.

Hurricane scenario

The hurricane scenario uses the same parameters as the default scenario, except that the proportion of long-distance dispersal $(1-p) = 0$ for 99 years and 0.01 for

Table 1 Speed, diversity and differentiation for different probabilities of long-distance dispersal and at different lengths of maximum dispersal

Maximum length of dispersal (d_{max})		Probability of long-distance dispersal $(1-p)$			
		0.001		0.0001	
		Average	Pooled	Average	Pooled
60 km	Speed (m/year)	523		278	
	H_S	0.18	0.18	0.05	0.05
	H_T	0.27	0.65	0.14	0.65
	G_{ST}	0.27	0.71	0.22	0.92
80 km	Speed (m/year)	713		348	
	H_S	0.24	0.24	0.13	0.13
	H_T	0.41	0.65	0.39	0.65
	G_{ST}	0.40	0.63	0.69	0.80
100 km	Speed (m/year)	937		450	
	H_S	0.32	0.32	0.20	0.20
	H_T	0.55	0.63	0.58	0.61
	G_{ST}	0.41	0.49	0.65	0.67

Populations sampled at regions 2, 4 and 6 (between 400–450 km) after 2000 years. Colonization to Britain had an average speed of approximately 500 m/year and current oak populations have an average H_T of 0.6 with a G_{ST} of 0.75 (Cottrell *et al.*, 2002).

Figure 3 Spatial arrangement of the three haplotypes in the latitudinal transects after 2000 years. The relative scale of the transects (10 km × 500 km) gives a compressed appearance and so a small area is enlarged with the true scale on each axis showing the patchy genetic structure. Here, (**a**) default scenario; (**b**) 100 year staggered start scenario; (**c**) barrier scenario; and (**d**) decreased seed survival with latitude.

1 year in every 100. Thus, the amount of long-distance dispersal over a 100-year period is the same as in the default scenario, but the actual event occurs at a periodicity of 100 years. The diversity and differentiation statistics for runs with this hurricane effect were in close agreement with those of the default scenario (Figure 4b). This indicates that rare, periodic instances of long-distance dispersal of the same additive total magnitude have a similar effect on genetic structure as the continuous low-probability of long-distance dispersal events of the default scenario.

Staggered start scenario

The set-up is the same as in the default runs, except that in the first year, only the uppermost latitudinal transect is seeded (2% of the first 10 km of the transect are randomly seeded with haplotype 1, as in the default run). After 100 years the central latitudinal transect is seeded with haplotype 2 and after a further 100 years the bottom latitudinal transect is seeded with haplotype 3. Figure 3b shows the spatial arrangement of the three haplotypes with this 100-year stagger. Clearly, the first haplotype

given a 'head start' can dominate in all transects and the last haplotype to be seeded exists only in a few isolated patches. A staggered start results in a reduction of within-population and total diversity well below that exhibited in the default runs (Figure 4c). There is also a trend of decreasing diversity with latitude (a reversal of the default scenario trend), as the haplotype seeded first becomes more dominant with latitude (Figure 4c). Tests were also undertaken with reduced stagger times. As the stagger is reduced the dominance of the initially seeded haplotype is reduced and genetic structure tends to that of the default runs.

Barrier scenario

A barrier to dispersal was simulated by making one area in the array uncolonizable (the central third of the central transect, see Figure 1). In this area, trees could not grow but if seeds were dispersed far enough and could pass over the barrier, they could establish on the other side (eg mountain ranges, seas). This scenario shows genetic structure similar to the default runs up until the barrier (Figure 3c). The barrier then greatly reduced the
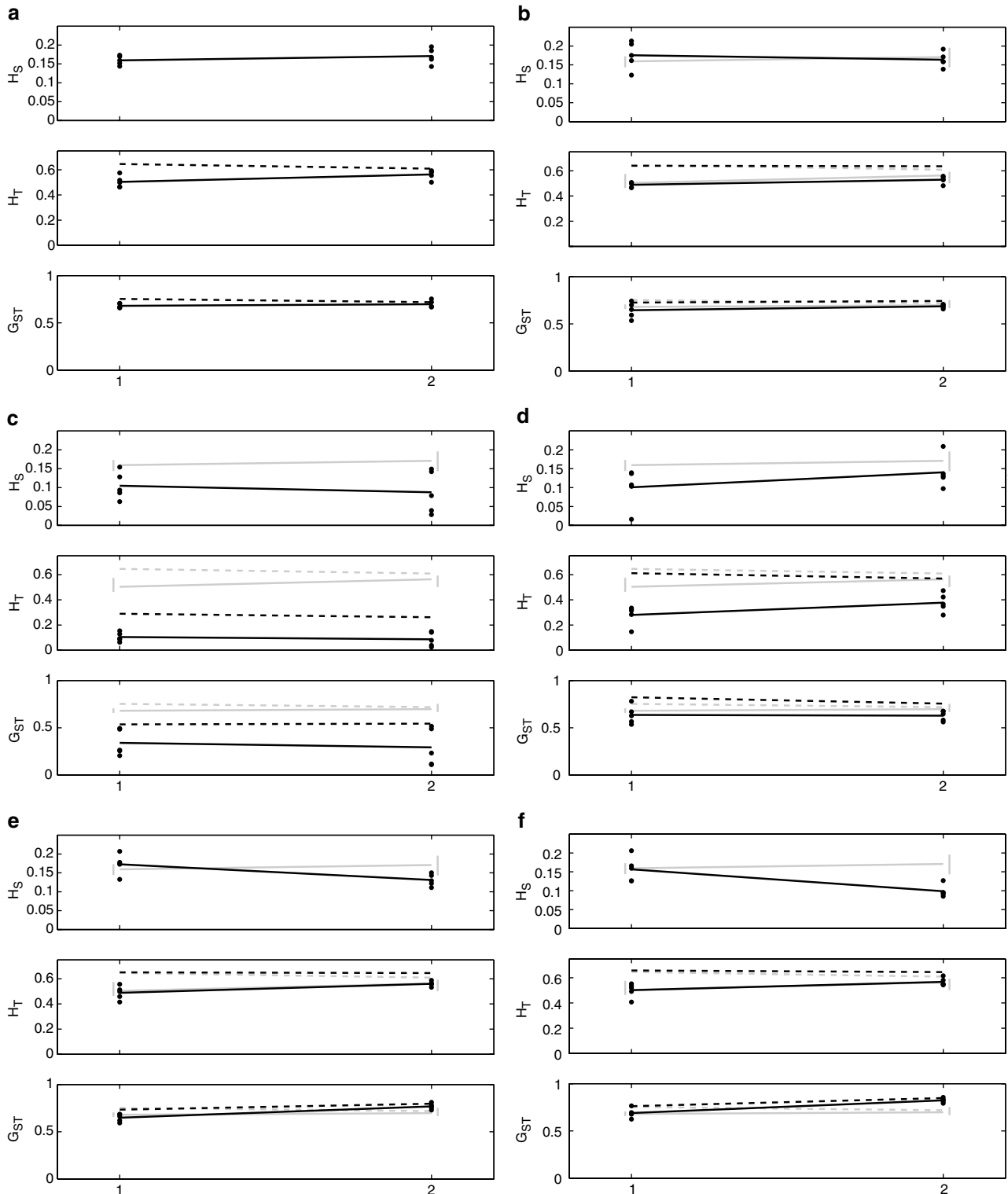
Figure 4 H_S, H_T and G_{ST} values for the average within-region and pooled populations from all three regions in columns 1 and 2. The regional average results for each of the five replicated simulations are represented by a point (●) and the average is shown by a solid line (—). The average of the five pooled statistics is shown by a dotted line (– –). Here, (**a**) default; (**b**) hurricane; (**c**) 100 year stagger; (**d**) barrier; (**e**) decreased fecundity with latitude; and (**f**) increased seed mortality with latitude. In (**b**) to (**f**) the default results are added in gray for comparison. The sidebars show the range of variation in the default runs. For each individual replicate (in all scenarios) the 95% confidence range was small (and within ± 0.01 for H_S, ± 0.005 for H_T and ± 0.02 for G_{ST}).

dispersal of haplotype 2 (blue) and prevented colonization of haplotype 1 into transect 3 (and *vice versa*) so that transects 1 and 3 were largely dominated by the original

seeded haplotype. Haplotype 2 was maintained, in isolated patches, in transects 1 and 3. After the barrier, the central transect was dominated by a mixture of

haplotypes 1 and 3. The statistics highlight this with greatly reduced diversity in column 1 (since populations were sampled from transects 1 and 3 and both were dominated by a single haplotype), with an increase in diversity in column 2, due to the mixing of haplotypes in the central transect after the barrier (H_T in column 1 was 0.28 and 0.38 in column 2). Throughout, the diversity is greatly reduced from the default scenario (Figure 4c, in default runs H_T was 0.5 in column 1 and 0.56 in column 2).

Latitudinal scenarios

Two potential effects of climate change with latitude are modelled. In the first, fecundity decreased with latitude, with N_0 decreasing linearly from 1 to 0.1 over the latitude range. In the second effect, survival of seeds decreases with latitude. Here an additional survival probability for successfully dispersed seeds was applied and varied linearly from 1 to 0.1 over the latitude range. Genetic diversity and structure for these scenarios are in close agreement with each other and the default scenario (Figures 3d and 4e, f). There is, however, a key difference for within-population diversity; the two scenarios compare closely with the default run in column 1 (H_S in default runs was 0.16 and was 0.16 with the fecundity effect and 0.17 with the seed survival effect) but both show a clear decreasing trend with latitude and a reduction of diversity in column 2 (in default runs within-population diversity increased to 0.17 but decreased to 0.13 with the fecundity effect and to 0.1 with the seed survival effect).

Speed of colonization

Figure 5 shows the position of the leading edge of the dispersing tree population over time. Dispersal speed is similar for the default, hurricane, staggered start and barrier scenarios, and in all of these scenarios speed of advancement is constant and approximately 500 m/yr. Thus, these factors do not affect the rate of spread. The rate of population spread is reduced in the scenarios where latitude affects fecundity and seed survival. Here, the colonization speed decreases with latitude. The decrease in speed is most severe when latitude affects seed survival.

Discussion

Changes in colonization speed and genetic structure with varying long-distance dispersal parameters

The individual-based model allowed assessment of how rare long-distance dispersal events affect genetic structure and colonization rates in plant populations. Our study showed that increasing the maximum length of dispersal and the probability of a long-distance dispersal event substantially increased colonization speed (Table 1). These effects allow more dispersal between latitudinal transects and so also increase average diversity within a population (H_S) and the diversity of the region (H_T) (Table 1). When within-population diversity is low and differentiation within a region is high, it indicates that the genetic structure of the region contains patches dominated by a single haplotype and this patch size is greater than the area of the population sampled. Le Corre et al (1997) proposed that patch size is
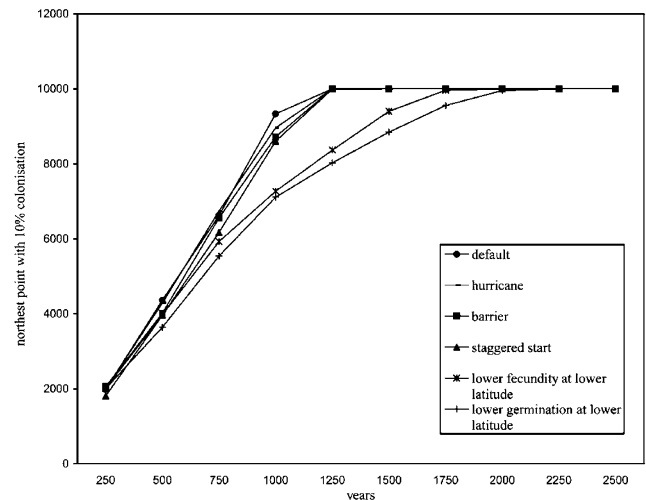
Figure 5 Position of the leading edge of the dispersing tree population over time for the different scenarios (in units of grid cells, 1 cell = 50 m).

indicative of the distance of long-distance dispersal and Ibrahim et al (1996) proposed that this was a result of a greater lag between populations formed by longer dispersed migrants and the advancing front of colonization. Our results confirm this but also suggest that patch size is most sensitive to the proportion of long-distance dispersal. A lower proportion of long-distance dispersal generated a cpDNA structure with larger-scale patchiness since successful long-distance migrants will dominate local populations, with less chance of competition from other long-distance dispersers.

Scenario effects on colonization speed

The analysis in this study shows that long-distance dispersal is critical for producing speeds of colonization on the same scale of oak recolonization to Britain (see Table 1). The speed of colonization remains constant and is determined by the frequency and length of rare long-distance dispersal events. This is in agreement with other models of seed dispersal (Hewitt, 1996; Ibrahim et al, 1996; Le Corre et al, 1997; Clark, 1998; Clark et al, 1999; Clark et al, 2001; Bullock and Clark, 2000; Pakeman 2001; Petit et al, 2000). The results from this study additionally show that the speed of colonization is robust to changes in the periodicity of dispersal, a barrier of seedling establishment and a staggered start of colonization. We have also shown that colonization speed is reduced when a latitudinal gradient of decreasing fecundity or decreasing seedling survival is applied. Both decreased fecundity and decreased seedling survival effectively reduce the number of successful long-distance dispersal events and further demonstrate the strong relationship of rare long-distance dispersal to colonization speed. The pollen core data show a slowing in the speed of recolonization of oak to Britain (Birks, 1989). These results therefore suggest that this is most likely due to the effects on fitness of individuals close to the northerly limit of the species range. This could arise since climate conditions become less favourable with latitude or due to climate change slowing during the period of colonization.

Scenario effects on genetic structure

The scenarios tested differed in the magnitude with which genetic structure deviated from the default runs; however, genetic structural patterns were still predominantly determined by rare long-distance dispersal. In model populations, the genetic structure brought about by long-distance dispersal consists of mosaics of patches of a single haplotype. This occurs as long-distance dispersal allows the establishment of individuals ahead of the main colonization front, followed by secondary local dispersal resulting in permanent patches of migrant haplotypes. Other simulated populations show patches of highly inbred demes that persist for hundreds of generations in spite of gene flow into patches (Ibrahim et al, 1996; Le Corre et al, 1997; Petit et al, 2002a, b). The stability of these migrant patches results in highly differentiated populations. High levels of differentiation are observed in oak populations (Petit et al, 2002b) and are therefore in agreement with a genetic structure dominated by the successive founder events of long-distance dispersal.

At more northerly latitudes, there was less dominance of the initially seeded haplotype and the genetic structure of the array became more influenced by long-distance dispersal, either directly or secondarily from the stable patches of migrant haplotypes. Therefore, distance between refugial populations and established colonized populations may be important in determining genetic characteristics in postcolonization populations. Cottrell et al (2002) found notable clumping of British populations with 66% of mixed species woods fixed for a single haplotype. This clumped pattern of haplotypes has also been found in populations in western France where areas of up to 50 km across are fixed for a single haplotype (Petit et al, 1997, 2002b). The genetic structure in model populations also remained stable over time, where populations in southerly latitudes maintain significant differentiation between transects after 2000 years (Figure 4).

Simulations changing the periodicity of long-distance dispersal showed no real change in genetic structure. This implies that the genetic structuring resulting from long-distance dispersal is robust to the frequency of dispersal events, providing the overall amount of dispersal remains the same, and allows a greater number of vectors of long-distance dispersal including birds, animals and also rare events such as hurricanes, to be considered as important for rapid colonization.

Staggering the start time of different haplotypes colonizing the arena caused a dramatic change in genetic structure, with a decrease in population diversity as the earliest colonizing haplotype dominated in all transects. The haplotype given a 'head start' will usually remain dominant in its transect and has a greater opportunity to disperse to free colonizable space in the neighbouring latitudinal transect and can then establish populations. These populations can then generate further dispersal events and allow that haplotype to dominate all transects, leading to low total diversity that decreases with latitude. However, late colonizers still maintain a presence after 2000 years through formation of stable fixed patches occurring from long-distance dispersal events, even with a 200-year lag after initial colonization. The haplotype structure in Britain shows a dominance of one haplotype along the western edge and in patches across the country increasing its dominance in northern England and Scotland (Cottrell et al, 2002). This is the haplotype that dominates Cornish populations and it may be that this haplotype originated from the initial colonization to Britain and that has given it an increasing advantage in colonization.

A barrier to colonization restricted the spread of haplotype 2 in model populations (the haplotype which is seeded in the transect containing the barrier, Figure 1), which only remained in small numbers due to the establishment of small stable patches through long-distance dispersal (Figure 3c). Haplotypes 1 and 3 become restricted to their own transects, as they cannot disperse across the barrier. In British populations, one haplotype is more abundant in English populations than in Scottish populations. In the area around the Pennines the other two haplotypes dominate with one haplotype to the west and one to the east (Cottrell et al, 2002). It is possible that this land barrier also acts as a significant barrier to establishment and may partially account for the decrease in one haplotype in northerly populations and the predominantly east/west division of the two other haplotypes.

In model populations with a latitudinal climate effect lower within-population diversity resulted, since either lower fecundity lessened the frequency of long-distance dispersal or because the extra mortality effect means long-distance seed dispersal was less successful. The lower probability of long-distance dispersal also affects the resulting haplotype patch size, as any successful seed that disperses ahead of the main front, even though local dispersal is also reduced by lower fecundity or increased seed mortality, is able to establish a local population with much less competition from other long-distance dispersers. Therefore, patches of a single haplotype tend to become larger, particularly in northerly regions where the mortality effect is most severe. The average and pooled regional diversity is similar to the default runs reflecting the fact that although the local haplotype patch sizes has increased there is still a mixture of the three haplotypes across the transects. Therefore, latitudinal effects on seed fecundity or mortality do not alter regional diversity but do affect within-population diversity through the increased size of local patches. This could explain the results for British oak populations, which show that all haplotypes are present in Scottish populations, but there is a lowering of genetic diversity within populations with increasing latitude (Cottrell et al, 2002).

Leptokurtic seed dispersal effects on genetic diversity

During colonization it is predicted that multiple founding events, a characteristic of leptokurtic dispersal, will show a strong decrease in allelic richness with distance from source that may not be mitigated by pollen dispersal in dense, fixed populations (Nei et al, 1975; Hewitt, 1996). These founder effects may be particularly severe in oak, as after the successful establishment of a colonizing individual, most acorns dispersed by the tree germinate within its neighbourhood (König et al, 2002). In our model, population genetic structure was the product of multiple founder events from long-distance dispersal with local dispersal confined to those cells surrounding the mother tree, yet within-population

diversity increased as the proportion of long-distance seeds dispersal increased. Our study indicates that the creation of a patch structure from long-distance dispersal of individual trees can maintain genetic diversity of populations because, although the founder events of long-distance colonization may result in low within-population diversity, the stability of these patches allows maintenance of migrant haplotypes within the total population. The presence of long-distance dispersal in model populations allowed all haplotypes to be present to some degree in the most northerly regions, even when there was some disadvantage to a haplotype due to late colonization or a landscape barrier. European oak populations also show this maintenance of cpDNA diversity where haplotypes are not lost during the spread of a few thousand kilometres (Petit *et al*, 2002b) and all three haplotypes found in Britain are present to some degree in the most northerly populations (Cottrell *et al*, 2002).

Decreases in haplotype diversity have, however, been reported for populations of forest trees following post-glacial colonization. For oaks, while three haplotypes of Iberian origin are found in Britain, there are 12 haplotypes in the potential source area (Cottrell *et al*, 2002; Petit *et al*, 2002a). The phylogeographic work on oaks (Cottrell *et al*, 2002; Petit *et al*, 2002a, b) indicates that there has been relatively little loss of genetic diversity with northward migration of oaks. Most diversity is lost within the first stages of colonization, which coincides with negotiation of large physical barriers (Pyrenees, Alps, Carpathians) but once past these, almost all cpDNA types are then transferred to northern latitudes. Studies of cpDNA variation in other European trees have found 13 haplotypes in refugial populations of *Alnus glutinosa* with only two found in northern Europe (King and Ferris, 1998), and samples from *Fagus sylvatica* consisted of 11 haplotypes in refugia but only one in northern populations (Demesure *et al*, 1996). Therefore, long-distance dispersal may not be able to mitigate loss of alleles and diversity in certain circumstances. For instance, if environmental factors that can decrease diversity (such as a staggered start or a barrier) are severe (eg the Pyrenean mountain range, which probably caused an extreme bottleneck of Iberian haplotype diversity; Petit *et al*, 2002b), or if southern populations are not able to spread beyond the northern leading edge (Hewitt, 1999), as may be the case for *F. sylvatica*, which had a slower rate of postglacial colonization than oak (Birks, 1989; Hewitt, 1999) and may therefore have experienced less long-distance dispersal. There could also potentially be decreased diversity as a species colonizes new habitat due to differences in colonization ability or fitness differences associated with certain haplotypes. Therefore, the dispersal of oak to Britain illustrates that long-distance dispersal has the potential to alleviate loss of alleles and therefore loss of diversity during colonization; however, this is not evident in all postglacial colonizing species and so must be considered in the context of biological, physical and environmental factors relevant to any given species.

In summary, this study has confirmed findings of previous studies that long-distance seed dispersal is crucial for producing the colonization rates and the genetic structure observed in postglacial populations of European deciduous oaks. However, in addition, we have shown how combinations of physical and environmental mechanisms may explain the patterns of genetic diversity and haplotype structuring found in oak populations in Britain. The results can be generalized to other plant systems in which rare long-distance dispersal may be a feature.

References

Bennett KD (1997). *Evolution and Ecology: The Pace of Life*. Cambridge University Press: Cambridge.

Birks HJB (1989). Holocene isochrone maps and patterns of tree spreading in the British Isles. *J Biogeogr* **16**: 503–540.

Brewer S, Cheddadi R, Beauliu JL de, Reille M (2002). The spread of deciduous *Quercus* throughout Europe since the last glacial period. *Forest Ecol Manage* **156**: 27–48.

Bullock JM, Clark R.T (2000). Long distance seed dispersal by wind: measuring and modelling the tail of the curve. *Oecologia* **124**: 506–521.

Cain ML, Damman H, Muir A (1998). Seed dispersal and the Holocene migration of woodland herbs. *Ecol Monogr* **68**: 325–327.

Cain ML, Milligan BG, Strand AE (2000). Long distance seed dispersal in plant populations. *Am J Bot* **87**: 1217–1227.

Clark JS (1998). Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. *Am Nat* **152**: 204–224.

Clark JS, Silman M, Kern R, Macklin E, HilleRisLambers J. (1999). Seed dispersal near and far: patterns across temperate and tropical forests. *Ecology* **80**: 1475–1494.

Clark JS, Silman M, Kern R, Macklin E, HilleRisLambers J (2001). Invasion by extremes: population spread with variation in dispersal and reproduction. *Am Nat* **157**: 537–554.

Cottrell JE, Munro RC, Tabbener HE, Gillies ACM, Forrest GI, Deans JD *et al* (2002). Distribution of chloroplast DNA variation in British oaks (*Quercus robur* and *Q-petraea*): the influence of postglacial colonization and human management. *Forest Ecol Manage* **156**: 181–195.

Demesure B, Comps B, Petit RJ (1996). Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution* **50**: 2515–2520.

Dumolin S, Demesure B, Petit RJ (1995). Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theor Appl Genet* **91**: 1253–1256.

Ferris C, Oliver RP, Davy AJ, Hewitt GM (1993). Native oak chloroplasts reveal an ancient divide across Europe. *Mol Ecol* **2**: 337–344.

Ferris C, Oliver RP, Davy AJ, Hewitt GM (1995). Using chloroplast DNA to trace postglacial migration routes of oaks into Britain. *Mol Ecol* **4**: 731–738.

Hewitt GM (1993). Postglacial distribution and species substructure: lessons from pollen, insects and hybrid zones. In: Lees DR, Edwards D (eds) *Evolutionary Patterns and Processes, Linnean Society Symposium Series*. Academic Press: London, **14**: 97–123.

Hewitt GM (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol J Linn Soc* **58**: 247–276.

Hewitt GM (1999). Post-glacial re-colonization of European biota. *Biol J Linn Soc* **68**: 87–112.

Higgins SL, Richardson DM (1999). Predicting plant migration rates in a changing world: the role of long distance dispersal. *Am Nat* **153**: 464–475.

Huntley B, Birks HJB (1983). *An Atlas of Past and Present Pollen Maps for Europe*. Cambridge University press: Cambridge.

Ibrahim KM, Nichols RA, Hewitt GM (1996). Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* **77**: 282–291.

King RA, Ferris C (1998). Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Mol Ecol* **4**: 95–103.

König AO, Ziegenhagen B, van Dam B, Csaikl UM, Coart E, Degen B *et al* (2002). Chloroplast DNA variation in western Central Europe and genetic consequences of human influences. *Forest Ecol Manage* **156**: 147–166.

Le Corre V, Machon N, Petit RJ, Kremer A (1997). Colonization with long-distance seed dispersal and genetic structure of maternally inherited genes in forest trees: a simulation study. *Genet Res (Cambridge)* **68**: 117–125.

Nei M (1973). Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* **70**: 3321–3323.

Nei M, Maruyama T, Chakraborty R (1975). The bottleneck effect and genetic variability in populations. *Evolution* **29**: 1–10.

Pakeman RJ (2001). Plant migration rates and seed dispersal mechanisms. *J Biogeogr* **28**: 795–800.

Petit RJ, Bialozyt R, Brewer S, Cheddadi R, Comps B (2000). From spatial patterns of genetic diversity to postglacial migration processes in forest trees. In: Silvertown J, Antonovics J (eds) *Integrating Ecology and Evolution in a Spatial Context*. Blackwell Science: Oxford pp 295–318.

Petit RJ, Brewer S, Bordacs S, Burg K, Cheddadi R, Coart E *et al* (2002b). Identification of refugia and post-glacial colonization routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecol Manage* **156**: 49–74.

Petit RJ, Csaikl, UM, Bordacs S, Burg K, Coart E, Cottrell JE *et al.* (2002a). Chloroplast DNA variation in European white oaks phlogeography and patterns of diversity based on data from 2600 populations. *Forest Ecol Manage* **156**: 5–26.

Petit RJ, Pineau E, Demesure B, Bacilieri R Ducousso A (1997). Chloroplast DNA footprints of postglacial recolonization by oaks. *Proc Natl Acad Sci USA* **94**: 9996–10001.

Pons O, Petit RJ (1995). Estimation, variance and optimal sampling of gene diversity I. Haploid locus. *Theor Appl Genet* **90**: 462–470.

Shigesada N, Kawasaki K, Takeda Y (1995). Modeling stratified diffusion in biological invasions. *Am Nat* **146**: 229–251.

Skellam JG (1951). Random dispersal in theoretical populations. *Biometrika* **38**: 196–218.

Winkler E, Stocklin J (2002). Sexual and vegetative reproduction of *Hieracium pilosella* L. under competition and disturbance: a grid based simulation model. *Ann Bot* **89**: 525–526.