RESEARCH ARTICLES

The Fate of Mutations Surfing on the Wave of a Range Expansion

Seraina Klopfstein, Mathias Currat, and Laurent Excoffier

Computational and Molecular Population Genetics Lab, Institute of Zoology, University of Bern, Bern, Switzerland

Many species, including humans, have dramatically expanded their range in the past, and such range expansions had certainly an impact on their genetic diversity. For example, mutations arising in populations at the edge of a range expansion can sometimes surf on the wave of advance and thus reach a larger spatial distribution and a much higher frequency than would be expected in stationary populations. We study here this surfing phenomenon in more detail, by performing extensive computer simulations under a two-dimensional stepping-stone model. We find that the probability of survival of a new mutation depends to a large degree on its proximity to the edge of the wave. Demographic factors such as deme size, migration rate, and local growth rate also influence the fate of these new mutations. We also find that the final spatial and frequency distributions depend on the local deme size of a subdivided population. This latter result is discussed in the light of human expansions in Europe as it should allow one to distinguish between mutations having spread with Paleolithic or Neolithic expansions. By favoring the spread of new mutations, a consequence of the surfing phenomenon is to increase the rate of evolution of spatially expanding populations.

Introduction

Many species including humans have gone through massive range expansions during the Pleistocene and the Holocene, mainly due to climatic changes that modified the distribution of suitable habitats (e.g., postglacial re-expansions in Europe, Hewitt 1996; Housley et al. 1997, Bocquet-Appel and Demars 2000). But range expansions can also follow speciation events or result from the artificial introduction of individuals into new habitats as in the case of invasive species (e.g., Estoup et al. 2004). Interestingly, range expansions leave a distinct pattern in the genetic structure of the population (Austerlitz et al. 1997, 2000; Ray, Currat, and Excoffier 2003; Excoffier 2004). Genetic diversity initially decreases due to a series of local bottlenecks caused by subsequent founder events (Nei, Maruyama, and Chakraborty 1975; Austerlitz et al. 1997) occurring during the range expansion. The genetic distance to the ancestral population thereby increases, and the level of intrademe diversity depends on the demographic parameters driving the range expansion (Austerlitz et al. 1997; Excoffier 2004). Moreover, clines of allele frequencies often appear along the major migrational direction (Barbujani, Sokal, and Oden 1995; Fix 1997). Recently, a simulation study (Edmonds, Lillie, and Cavalli-Sforza 2004) showed that new mutations occurring at the front of a two-dimensional range expansion could occasionally travel with the wave of advance and be carried over long distances. However, a majority of these new mutations usually do not travel at all and remain at low frequencies or are lost by genetic drift. The successful "surfing mutations" can reach very high frequencies and eventually occupy a large area.

In the present study, we examine the surfing phenomenon in more detail through extensive computer simulations, focusing on the fate of surfing mutations. We investigate the importance of the initial location of the mutation relative to the edge of the expansion on its ability to catch the wave. We also study the effects of various demographic properties of the local population subdivisions (carrying capacity, intrinsic rate of growth, and migration rates between neighboring demes) on the fate of the new mutation. Finally, we discuss the implications of our results for the study of the settlement of Europe by modern humans.

Key words: range expansion, stepping-stone, neutral mutation, genetic drift, simulation, Europe.

E-mail: laurent.excoffier@zoo.unibe.ch.

Mol. Biol. Evol. 23(3):482–490. 2006
doi:10.1093/molbev/msj057
Advance Access publication November 9, 2005

Materials and Methods

We used the SPLATCHE program (Currat, Ray, and Excoffier 2004) to simulate a range expansion of haploid individuals on a rectangular two-dimensional stepping-stone grid (25 × 100 demes) identical to that used in Edmonds, Lillie, and Cavalli-Sforza (2004). A range expansion was initiated in a deme located in the middle of the left side of the grid (position $<0, 13>$). The initial population was assumed to be at carrying capacity and made up of haploid individuals, all carrying the same allele. In each generation, we simulated two distinct demographic events: (1) the density of all demes was logistically regulated, with carrying capacity K and an intrinsic growth rate r and (2) all nonempty demes then sent migrants at rate m to the adjacent demes (resulting either in the colonization of new demes or in gene flow between two nonempty demes). This process was continued for a predefined number of generations.

In order to simulate the occurrence and the fate of a new neutral mutation, we made two modifications to the program SPLATCHE. First, we offered the possibility to the user to specify the $<x, y>$ coordinates of a deme where a new mutant allele could appear, as well as the number of generations elapsed between the first colonization of this deme and the occurrence of the new mutation in that deme. This last parameter is referred to as Δt in the rest of the paper and represents the age of the deme when the mutation first appeared. In our simulations, the mutation appears in the central deme at x-coordinate 10, as shown by the dotted line in figure 1*A*. Individuals carrying the mutant allele then experience the same series of events than the rest of the population, with demographic regulation preceding

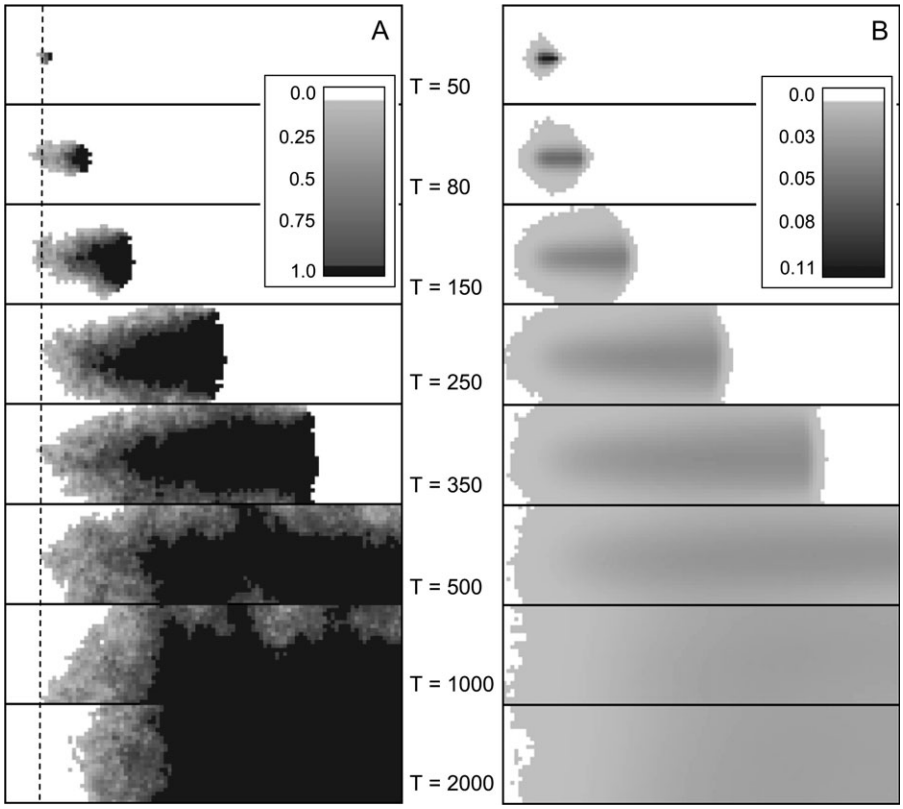

FIG. 1.—Spatial distribution of the frequency of a new mutation. (*A*) Spread of a random (but surviving) mutation. (*B*) Average frequency over all successful simulations. Simulations conditions were $K = 50$, $m = 0.2$, and $r = 0.5$. The time T is expressed in generations since the start of the expansion. The geographic origin of the expansion is in deme at position <0, 13> on the left side of the grid, while the new mutation appears in deme at position <10, 13>. We have drawn a dashed line at *x*-coordinate 10 on Fig. 1*A* to represent the *x*-origin of the mutation.

the exchange of migrants. Second, we extended the demographic model implemented in SPLATCHE in order to account for the genetic drift of mutant and nonmutant allele frequencies. While the density of the whole population within a deme (including mutants and nonmutants) is logistically regulated, the absolute number of mutants followed a binomial distribution with probability parameter equal to its relative frequency in the previous generation.

We performed many simulations by varying the growth rate r (between 0.2 and 0.8, which includes most values estimated for human populations, see e.g., Fort, Pujol, and Cavalli-Sforza 2004), the migration rate m (range: 0.1–0.4), and the local carrying capacity K (range: 10–500) to study their influence on the diffusion of the mutation. The various combinations of m and K we used allowed us to study a broad range of migrations between neighboring demes (Nm varied between 1 and 200). These migration rates cover most of the situations encountered for humans (see e.g., Excoffier 2004) and should hold for other species colonizing uniform habitats, as modeled here. Additionally, we varied the parameter Δt from 0 to 50 generations to see the influence of the age of the deme where the mutation appears on its ability to surf.

We record descriptive statistics such as the frequency and the spatial distributions of the mutation at different time points T after the onset of the colonization. The frequency of the occurrence of the traveling behavior (which will be called the "surfing" behavior thereafter) of the new mutations was computed as the proportion of simulations in which the centroid of the mutant spatial distribution (calculated as the average *x*- and *y*-coordinate of all demes in which at least one mutant allele was present) moved more than 20 demes in the direction of the expansion. This distance threshold is similar to that used by Edmonds, Lillie, and Cavalli-Sforza (2004). Simulations (results not shown) have revealed that the position of the mutant centroid along the *x* axis was clearly bimodally distributed and that the width of the first mode (representing mutants that do not move far from their origin) was about 20 demes for a large range of K values. This result thus provides an empirical justification of our criterion to distinguish the surfing mutations from the others on our simulation grid, but it is clear that this criterion could be different in the real world (e.g., when range expansions occur on a much smaller surface or when there is environmental heterogeneity).

Simulations were considered as "successful" if the mutation was still found in the gene pool after a given time T after the onset of the colonization. Descriptive statistics reported in the *Result* were computed on a minimum of 1,000 successful simulations per simulation conditions. Because the survival rate of the mutation varied a lot according to the simulation parameter values, the total number of simulations including those that were unsuccessful was limited to 500,000 (a number that was only reached in the case of Δt values of 10 and 50). All simulations were conducted on a 40-node Linux cluster. Simulation series

Table 1
Frequency Distribution of the New Mutations at Two Times *T* After the Start of the Colonization and for Various Carrying Capacities *K*

				Frequency in Colonized Area				Frequency in Whole Area			
	K	Survival Probability	Probability of Surfing	0%–5%	5%–20%	20%–50%	50%–100%	0%–5%	5%–20%	20%–50%	50%–100%
$T = 500$	10	0.15	0.91	0	0.003	0.03	0.96	0.11	0.12	0.38	0.39
	50	0.06	0.74	0.01	0.3	0.35	0.33	0.31	0.16	0.38	0.15
	100	0.05	0.45	0.27	0.33	0.29	0.11	0.53	0.13	0.28	0.06
	500	0.13	0.05	0.94	0.04	0.02	0.001	0.96	0.02	0.02	0.001
$T = 1{,}500$	10	0.13	0.98	0	0	0.009	0.99	0.05	0.14	0.48	0.33
	50	0.04	0.99	0.002	0.09	0.45	0.45	0.13	0.22	0.39	0.26
	100	0.02	0.96	0.02	0.23	0.54	0.21	0.15	0.22	0.52	0.11
	500	0.02	0.30	0.69	0.18	0.12	0.006	0.77	0.10	0.12	0.006

required hours to days on a 2.8-GHz Pentium 4, depending on the survival probability of the mutation.

Allele frequency clines (AFCs) can result from spatially varying selective forces (Costa et al. 1992), from admixture between genetically distinct populations (Ammerman and Cavalli-Sforza 1973; Menozzi, Piazza, and Cavalli-Sforza 1978; Currat and Excoffier 2005), or from subsequent founder events during a range expansion (Barbujani, Sokal, and Oden 1995; Fix 1997; Currat and Excoffier 2005). To investigate if the surfing phenomenon leads to clinal patterns of allele frequencies, we tested the formal presence of AFCs by regressing the mutant frequency (averaged over demes at a given position on the x axis) plotted against its x-coordinate. As described in Currat and Excoffier (2005), clines were considered significant if they have a significantly positive slope (corresponding to an increase of the frequency of the mutant in the direction of the expansion) at the 5% level.

Results

Factors Affecting the Success of New Mutations

An example of the spatial and frequency distributions of a new mutation surfing with the wave is shown in figure 1 for different times after the onset of the simulations. As illustrated in figure 1*A*, surfing mutations travel with the wave of advance. The speed of the wave and the exact arrival time of the mutations depend on the migration rate (m), on the logistic growth rate within the deme (r), as well as on the deme carrying capacities (K). For instance, when $m = 0.2$ and $r = 0.5$, the colonization of the whole simulated world takes on average 580 generations for $K = 10$, 420 generations for $K = 50$, and 350 generations for $K = 500$. The colonization proceeds faster in larger populations because the absolute number of emigrants Nm is larger and thus allows for a faster foundation of new populations at the edge of the expansion. The spatial frequency distribution of a new mutation averaged over all simulations reported in figure 1*B* is quite different as it is more widespread, more uniform, and shows overall lower frequencies. This difference is mainly due to the large variance in the success of new mutations (i.e., a majority of mutations do not surf on the wave and remain around their place of origin and/or disappear, as reported in table 1 for different values of K), but it should be stressed that a common feature of all successful mutations is to show the highest frequencies at the latitude of their origin and close to the front of the wave. The persistence of a new mutation in a population is therefore strongly related to the fact that this mutation has been carried by the wave and is thus a surfing mutation.

While Edmonds, Lillie, and Cavalli-Sforza (2004) only described the surfing phenomenon in very small demes ($K = 10$), we find here that new mutations can also spread in a population with much larger demes but that there is a negative relationship between deme carrying capacity and the probability of a mutation to travel with the wave (table 1). Also, when the wave spreads in larger demes, the average frequency of the mutation remains much lower than when it occurs in smaller demes. For instance, for $K = 10$, 96% of the surviving mutations have a frequency above 50% in the colonized area after 500 generations, whereas for $K = 500$, 94% of the surviving mutants stay below a frequency of 5%. For intermediate carrying capacities ($K = 50$ and $K = 100$), about 60% of the new mutations will have frequencies between 5% and 50%. As stated before, the probability to surf and thus to diffuse away from the place of origin is much lower for subdivided populations with larger carrying capacities. Indeed, if $K \leq 100$, more than 95% of the surviving mutations after 1,500 generations will have surfed away from their place of origin, but only 30% of the mutations will have surfed if $K = 500$. Table 1 also shows that for $K = 10$, almost all (99%) mutants surviving after 1,500 generations have an average frequency higher than 50% in the colonized area, whereas for $K = 500$, less than 1% of the mutants reach this frequency.

When we look at the average mutation frequency over the whole simulated area instead of only in the area colonized by the mutation, the patterns are similar, but the average frequency of the mutation is much lower (table 1). This reduction in frequency is stronger for smaller deme sizes than for larger ones. This pattern can be understood from figure 2, which relates the mutant frequency in the colonized area with the colonized surface. The relationship between these two variables is clearly sigmoid but drastically different for various carrying capacities. It appears that for small K ($K = 10$), a majority of new mutations will both occupy a relatively large portion of the simulated grid $>30\%$ and reach large frequencies ($>80\%$) in keeping with previous results (Edmonds, Lillie, and Cavalli-Sforza 2004), while for $K = 500$, a majority of new mutations will be at low frequencies ($<10\%$) and have colonized less than 40% of the simulated grid.

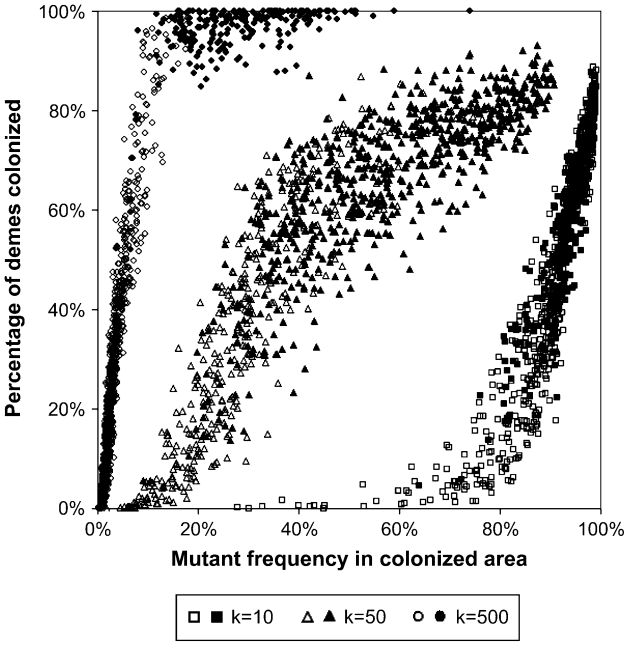

FIG. 2.—Relationship between mutant frequency and colonized area for different carrying capacities K 500 generations after the onset of the colonization. The percentage of demes colonized by mutant genes is plotted against the average mutant frequency in these demes. In all simulations $m = 0.2$, $r = 0.5$, and $T = 500$. The closed symbols represent simulations where a significant cline of mutant allele frequency was observed.

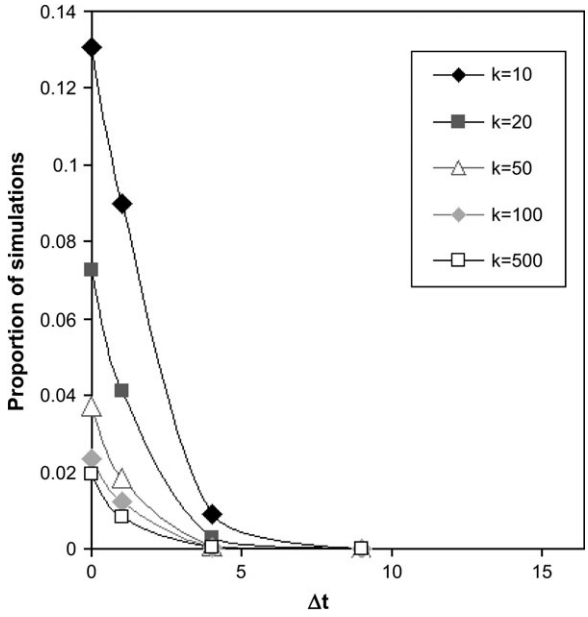

FIG. 4.—Influence of the age of the deme Δt on the success of new mutants. We report here the proportion of simulation for which the mutant was still observed after 500 generations as a function of the age of the deme where the mutant initially appeared, for different deme sizes K. Simulation conditions were $m = 0.2$ and $r = 0.5$.

The influence of other demographic parameters on the success of a new mutation is summarized in figure 3, where we report this probability for all combinations of the following parameter values: carrying capacity K: 10, 50, 100, and 500; migration rate m: 0.1, 0.2, and 0.4; and growth rate r: 0.2, 0.4, 0.5, and 0.8. We find that the mutation success (and thus the surfing phenomenon) can indeed occur for all parameter combinations but that the proportion of successful mutations varies according to the particular values of the demographic parameters K, m, and r. Overall, the success of a mutation is positively correlated with r and negatively with K and m, such that the composite parameter $r/(Km)$ is a good predictor of the success of the mutation, which was measured as its final frequency averaged over all demes and simulations (fig. 3).

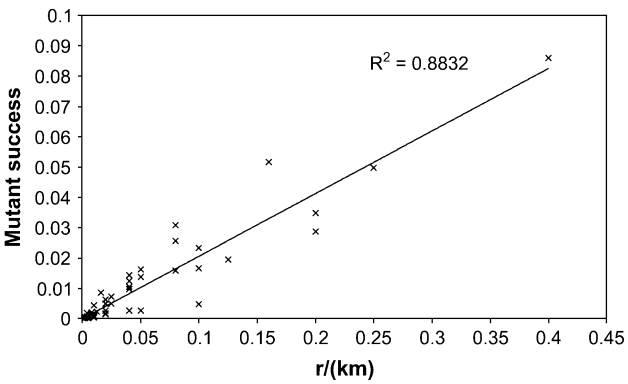

FIG. 3.—Success of a mutation measured as a function of the composite parameter $r/(Km)$. Mutant success at time $T = 500$ is measured here as the average frequency over all demes and simulations. K, m, and r were varied as described in the text. As shown by R^2 value, the linear regression explains more than 88% of the mutant success ($P < 0.0001$).

The Importance of Being Close to the Wave

The age Δt of the deme at the time when the mutation first appears is also a crucial factor for explaining the success of the mutation. In figure 4, we show the sharp decrease in the proportion of successful mutations after 500 generations as a function of Δt. From the analysis of 500,000 simulations, it is clear that when $\Delta t > 10$, new mutations have virtually no chance to survive, confirming that a mutation appearing in an already colonized area has virtually no chance to spread if it is not selected (Slatkin 1980). Except for very small deme sizes ($K = 10$), the new mutation thus needs to appear in demes colonized less than four generations ago to have some chance to survive and surf. However, the spatial and frequency distributions of surviving mutations were found overall similar, irrespective of Δt (results not shown). Thus, even though mutations that appear in older demes have a lower probability to surf on the wave, and if they do, they will spread similarly to those having appeared on the wave front, implying that the surfing behavior is not restricted to new mutations but also to any standing variant.

Mutant AFCs

The frequency of AFCs is listed in table 2 for different deme sizes K at different times. When considering all simulations, significant AFCs are overall very rare, in keeping with previous results (Currat and Excoffier 2005), but are slightly more likely to occur for low K values because

Table 2
Frequency of Mutant Frequency Clines After a Range Expansion

	Generation 500		Generation 1,500		Generation 5,000	
K	Among All Simulations	Among Successful Simulations	Among All Simulations	Among Successful Simulations	Among All Simulations	Among Successful Simulations
10	0.032	0.216	0.041	0.317	0.042	0.379
50	0.032	0.557	0.025	0.674	0.023	0.840
500	0.008	0.069	0.003	0.245	0.002	0.594

Note.—The proportion of simulations showing a significant AFC is listed for three different time points. Simulation conditions were $m = 0.2$ and $r = 0.5$ in all cases.

founder effects are stronger. The proportion of AFCs is larger among successful mutations than among all simulations, and it is largest for intermediate K values (K = 50). As shown in figure 2, AFCs occur predominantly in simulations where new mutations have successfully colonized a large area. For $K = 50$, almost all mutations having colonized 60% of the simulated grid show significant clines, whereas for $K = 500$, mutations need to have colonized almost always more than 90% of the grid to show significant AFCs. But note that for large K, the distribution of significant clines is clearly bimodal because many mutations having occupied a small area at low frequencies also present significant clines (fig. 2). These local clines probably represent cases of isolation by distance and are not found for lower K values, probably due to stronger local drift. The shape of the clines depends to a large extent on the deme size (K), clines being steeper, more ragged, and usually occurring on a smaller distance for lower than for higher K values (fig. 5).

Application to the Colonization of Europe

In order to investigate the propagation of mutant genes in an irregular grid, we simulated the spread of mutations during the Paleolithic or during the European Neolithic transition using a digital map of Europe and the Near East (as previously used in Currat and Excoffier 2005). In this model, the continental surface of Europe was divided into square demes of 2,500 km^2. The origin of the expansions was set in the Near East as indicated by a gray cross in figure 6, and we investigated the fate of new mutations occurring early in northwestern Anatolia (see black cross in fig. 6). Simulation conditions were $K = 50, m = 0.2, r = 0.5, T = 1{,}500$ for

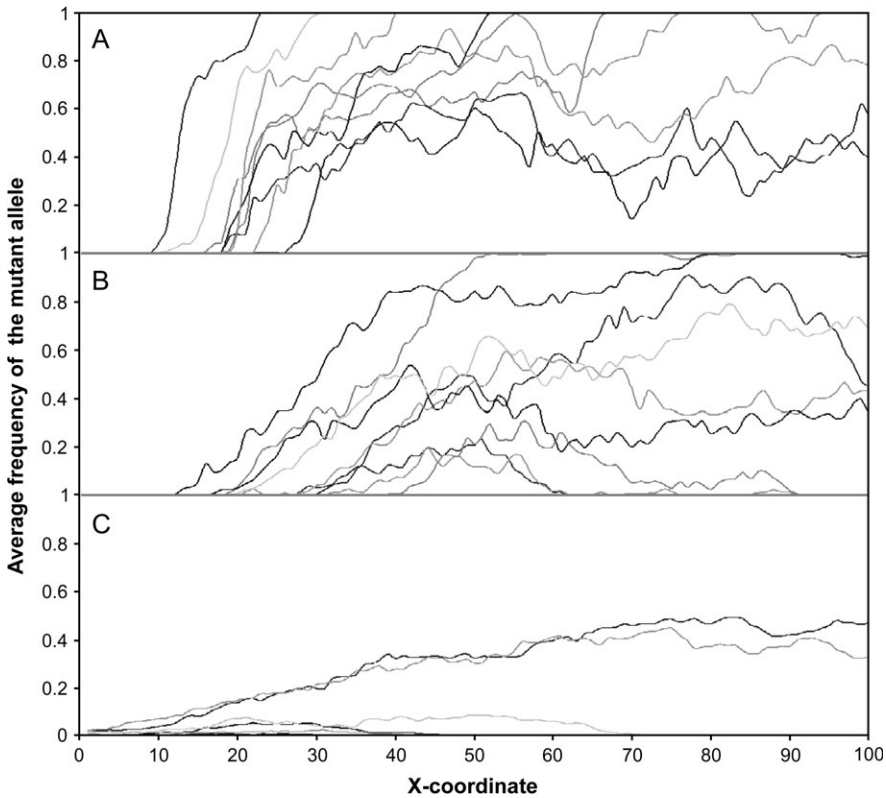

FIG. 5.—Profile of mutant frequency along the x axis. The average frequency of the mutant allele along the x axis is shown for 10 randomly chosen successful simulations for different deme sizes. (*A*) $K = 10$. (*B*) $K = 50$. (*C*) $K = 500$. Some simulations clearly show a clinal pattern with an increase in mutant frequency toward the right part of the range. In all cases, the mutant initially appeared at position 10. Simulation conditions were $m = 0.2$ and $r = 0.5$.

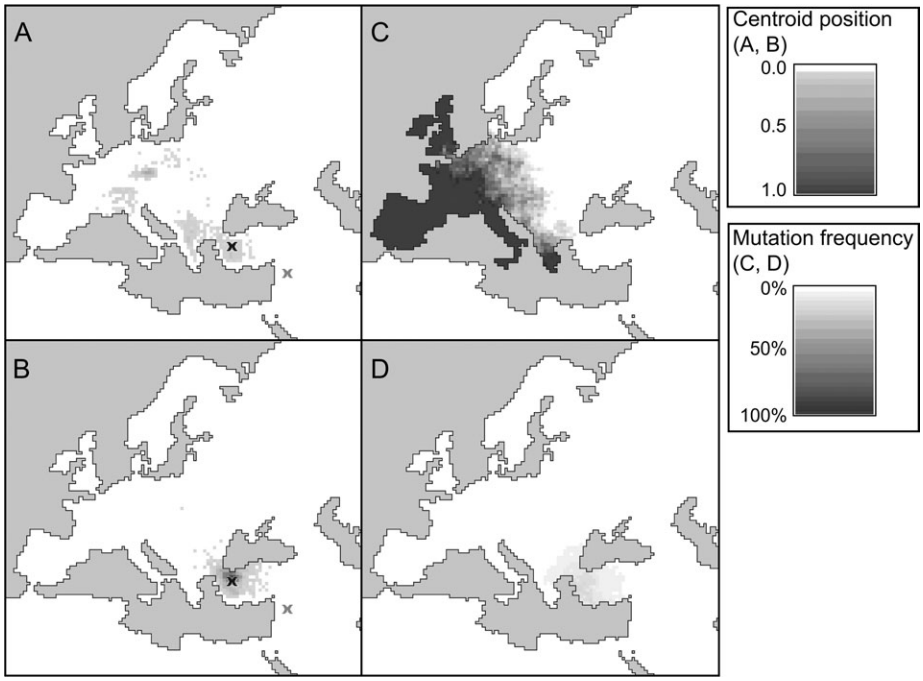

FIG. 6.—Simulation of the spread of new mutants during the colonization of Europe. Position of the centroid of the mutant distribution after the colonization of Europe for a Paleolithic expansion having started 1,500 generations ago (*A*) or a Neolithic expansion having started 500 generations ago (*B*). On panes *C* and *D* we report two representative cases of the frequency distributions of new mutations having appeared during a Paleolithic (*C*) or a Neolithic (*D*) expansion. The origin of the expansion is shown in panes *A* and *B* as a gray cross, while the origin of the mutation is shown as a black cross.

Paleolithic simulations and $K = 500$, $m = 0.2$, $r = 0.5$, $T = 500$ for Neolithic simulations. These parameters correspond approximately to the expansion of early modern humans about 40,000 years ago and an expansion of farmers about 10,000 years ago in Europe (see Currat and Excoffier 2005 for details). For the sake of simplicity, Paleolithic and Neolithic expansions were simulated independently, and each expansion was assumed to occur in an empty territory, even though some interbreeding has certainly occurred between Paleolithic and Neolithic populations (Dupanloup et al. 2004; Currat and Excoffier 2005).

As expected from our previous results, the average final frequency of the mutation is found much higher after the Paleolithic expansion than after the Neolithic expansion (44% and 2% in the colonized area, respectively), and the spatial distributions of the centroid of the mutation geographic distribution clearly differ between the Paleolithic (fig. 6*A*) and Neolithic (fig. 6*B*) range expansions. In keeping with previous simulations in the rectangular area, new Neolithic mutations are much more likely to stay around their place of origin than Paleolithic mutations because they occur in larger populations. New Paleolithic mutations are thus more likely to surf on the range expansion. It is also interesting to note that a larger proportion of mutations show significant AFCs after the Paleolithic expansion than after the Neolithic expansion (45% and 8% of all successful simulations, respectively).

Discussion

Scope of the Surfing Phenomenon

Our results underline two important aspects of the surfing mutation phenomenon in shaping the genetic diversity of a population during and after a range expansion. First, we have shown that the surfing phenomenon not only occurs in small populations, as previously reported (Edmonds, Lillie, and Cavalli-Sforza 2004), but also in populations of much larger local size. In the human context, it implies that this phenomenon could have not only occurred in hunter-gatherer populations but also in larger populations of farmers. Second, we find that the survival rate of a mutation decreases drastically if it does not occur on the edge of the range expansion. If the carrying capacity of the demes is larger than 10, the probability of survival of a new mutation is indeed almost negligible if the deme in which it appears has been colonized more than five generations ago (fig. 4). This result suggests that a large proportion of new mutations arising during range expansions could persist and reach high frequencies in populations with subdivisions of small sizes or remain at quite low frequencies in populations with larger local sizes. It therefore implies that times of population range expansions are very important evolutionary periods, where mutations could predominantly accumulate, potentially contributing to well-known lineage-specific differences in rates of evolution (Li 1997). Indeed, because the surfing phenomenon is expected to lead to a reduced rate of loss of neutral mutations (see fig. 4), mutations could accumulate and even go to fixation "after" rather than before or during the speciation events, as previously assumed (Gavrilets 2003). Therefore, a speciation event followed by a range expansion could lead to an accelerated rate of neutral evolution, which would be independent from potential adaptive mutations linked to the speciation event (see e.g., Gillespie 1991). Note, however, that the power of the surfing phenomenon

in speeding up evolution could even be stronger if not only neutral but also positively selected mutations were considered. Their advantage relative to the ancestral allele could enable them to spread even more easily than neutral mutations (Otto and Whitlock 1997). Advantageous and neutral mutations could thus accumulate at the edge of the population range, potentially producing new multilocus combinations, some of which could reveal highly adaptive (Eswaran, Harpending, and Rogers 2005).

The potential of the surfing phenomenon to enhance the probability of fixation of slightly deleterious mutations remains to be assessed. However, this phenomenon could help explaining the unusually large spread of some negatively selected recessive alleles, like for instance, those involved in human genetic diseases throughout Europe, which could have been propagated during Paleolithic or Neolithic expansions in the heterozygous and therefore neutral state. Finally, the recent observation of several haplotypes in brain genes found at much higher frequencies outside Africa than in sub-Saharan Africa (e.g., Evans et al. 2005; Mekel-Bobrov et al. 2005) could also be explained by the surfing mechanism. Indeed, standing variants present in populations leaving Africa could have been driven to very high frequencies by the surfing mechanism rather than by selection acting on the brain as previously thought (Balter 2005).

Demographic Parameters Influencing Mutation Success

The large influence of demography on mutational success has important implications. While only population size and growth rate are considered important in influencing the success of neutral mutations in stationary populations (Kimura 1962; Maruyama 1970), a combination of these two factors with the migration rate determines the success of mutations in populations under range expansion. A more exact analytical relationship between these parameters and mutation success remains to be determined in the context of range expansions; nevertheless, the strong influence of the coefficient $r/(Km)$ on the mutant success can be understood as follows: high r values should allow the new mutant to rapidly reach high local frequencies, before nonmutant genes immigrate (at rate Km) from other demes.

Deme size does not only affect the probability of a mutation to surf but also determines its final frequency and spatial distribution. Very small and very large deme sizes strongly condition the behavior of a new mutation. If K is very small ($K = 10$) a mutation is able to spread only if it reaches a very high frequency, whereas if K is very large ($K = 500$) a mutation is able to reach a considerable frequency only if it attains a large spatial distribution (see fig. 2). The difficulty for a mutation to spread when K is small is due to strong local genetic drift, leading to a fast extinction of the mutation if it does not reach a considerable frequency in every colonized deme. If K is large, there is less genetic drift, many migrations occur (Nm is large), and mutations can thus colonize a large portion of the grid. On the other hand, a large deme size and a large number of exchanged migrants between neighboring demes prevent mutations from reaching high local frequencies. In these two extreme cases, the relationship between colonized area and mutant average frequency shows little variation, which is not the case for intermediate values of K (e.g., $K = 50$ in fig. 2). In that case, while there is also a positive relationship between mutant frequency and colonized area, the mutant average frequency can cover a much wider range (5%–90%) than for lower or higher K values (see fig. 2).

The fact that deme size has such a high influence on the behavior of new mutations suggest that it should be possible to distinguish between neutral mutations that have arisen in range expansions of populations with small or large densities, like in the case of human Paleolithic and Neolithic expansions in Europe, as discussed below.

Surfing Mutations Lead to AFCs

We find that AFCs can be obtained for all deme sizes but are most frequent for $K = 50$ when conditioning on the observation of the mutation (table 2). The lower proportion of successful simulations showing AFCs with $K = 10$ is certainly due to stronger genetic drift leading to more irregular frequency profiles of the mutation in low density demes (fig. 5). The decrease in cline frequency observed from $K = 50$ to $K = 500$ is on the other hand linked to the small frequencies reached by a large proportion of mutations when $K = 500$ (table 1) and their tendency to remain close to or to radiate around their place of origin. The increase in the proportion of successful simulations showing clines over time (table 2) can be explained by the fact that mutations remaining close to their place of origin, and thus not showing AFC patterns, are found at lower frequency and have thus a larger probability to go extinct, while surfing mutations will be more successful over time and will be distributed over larger areas, providing more opportunities to show clinal patterns.

Europe has been colonized relatively late (~40,000 years ago) by modern humans who totally replaced Neanderthals (Mellars 2004). A second colonization wave occurred during the Neolithic with the spread of farming, but the exact contribution of the Paleolithic genes to the current gene pool is still under discussion (Richards et al. 1996; Chikhi et al. 1998*a*). Clines of allele frequency along the colonization routes into Europe have been found for a variety of markers, including classical markers (Menozzi, Piazza, and Cavalli-Sforza 1978; Cavalli-Sforza, Menozzi, and Piazza 1993), nuclear polymorphisms (Chikhi et al. 1998*b*), and Y-chromosome markers (Hill, Jobling, and Bradley 2000; Rosser et al. 2000). It has been proposed that the progressive admixture of Neolithic and Paleolithic populations could be responsible for the creation of these continent-wide AFCs, a process known as demic diffusion (Ammerman and Cavalli-Sforza 1973; Menozzi, Piazza, and Cavalli-Sforza 1978). The observation of AFCs in Europe has thus been taken as a strong indication of the importance of the Neolithic component into the current European gene pool (Chikhi et al. 1998*a*), but it has been shown that AFCs can also be obtained in range expansions without any admixture, as a consequence of a succession of bottlenecks (Barbujani, Sokal, and Oden 1995; Fix 1997; Currat and Excoffier 2005).

The surfing phenomenon also includes a series of founder events, and it leads to a continual increase of

the allele frequency during the successful spread of the new mutation (see fig. 1), with its maximal frequency being reached at the edge of the expansion. This suggests that AFCs resulting from the surfing phenomenon could be differentiated from those resulting from admixture events for markers whose ancestral state can be reconstructed. Our study further suggests that mutations having arisen during Paleolithic range expansions should show larger absolute frequency differences than those having occurred during a pure Neolithic expansion (see fig. 5). Conversely, mutations that are found today at very low frequencies and nevertheless show a clinal pattern, which can be observed in many recessive disease alleles such as phenylketonuria (Tighe et al. 2003), hemochromatosis (Lucotte 2001), or cystic fibrosis (Mateu et al. 2002), are much more likely to have been spread during the Neolithic than during the Paleolithic expansion. Finally, we would predict that new mutations being highly localized and at relatively low frequencies are more likely to have spread during the Neolithic expansion. It should thus be possible to link the spatial distribution of a mutation to a major expansion event. The spatial pattern of mutations could then possibly be used in the future to estimate the genetic contribution of successive expansions.

While all European AFCs certainly cannot be attributed to the surfing phenomenon, it is very likely that some of them result from this mechanism. It is indeed important to note here that while we have focused our analysis on the spread of new mutations, any allele present at the front of a wave of advance could surf on the wave. Thus our result should apply equally well to standing variants and to new mutants.

Further simulations including environmental information and its temporal change since the first colonization of Europe could lead to a better understanding of how mutations could have spread into Europe during the expansion phases in human history. There could indeed be a strong influence of spatial heterogeneity (including spatial bottlenecks) on the ability of new mutations to catch the wave, while temporal changes in the environment could help to date the appearance of currently observed mutations.

As an extension of our approach, the behavior and surfing abilities of positively and negatively selected genes remain to be investigated. Because the local growth abilities of a mutation have been shown to positively affect its ability to surf (as shown by the high influence of the local growth rate r in fig. 3), mutations that are under positive selection could even be more strikingly influenced by the surfing phenomenon. On the other hand, it would be interesting to study how intense negative selection must be to prevent deleterious mutations from spreading with an expansion wave. Given that recessive mutants behave essentially as neutral alleles at low frequencies, recessive deleterious alleles could still surf on a wave of advance and diffuse in a large area but be prevented from reaching high frequencies by negative selection. This phenomenon could explain why so many rare recessive diseases are found at low frequencies in Europe, and it would make it unnecessary to invoke complex selection mechanisms to explain their wide distribution. The surfing phenomenon has thus probably strongly influenced both neutral and selected genetic diversity of many species and populations, as well as their ability to adapt to newly colonized environments.

Acknowledgments

Thanks to Pierre Berthier for programming and computing assistance. This study was supported by a Swiss National Science Foundation grant number 3100A0-100800 to L.E.

Literature Cited

Ammerman, A., and L. Cavalli-Sforza. 1973. A population model for the diffusion of early farming in Europe. Pp. 343–357 *in* C. Renfrew, ed. The explanation of culture change. Duckworth, London.

Austerlitz, F., B. Jung-Muller, B. Godelle, and P.-H. Gouyon. 1997. Evolution of coalescence times, genetic diversity and structure during colonization. Theor. Popul. Biol. **51**:148–164.

Austerlitz, F., S. Mariette, N. Machon, P. H. Gouyon, and B. Godelle. 2000. Effects of colonization processes on genetic diversity: differences between annual plants and tree species. Genetics **154**:1309–1321.

Balter, M. 2005. Evolution. Are human brains still evolving? Brain genes show signs of selection. Science **309**:1662–1663.

Barbujani, G., R. R. Sokal, and N. L. Oden. 1995. Indo-European origins: a computer-simulation test of five hypotheses. Am. J. Phys. Anthropol. **96**:109–132.

Bocquet-Appel, J.-P., and P. Y. Demars. 2000. Neanderthal contraction and modern human colonization of Europe. Antiquity **74**:544–552.

Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza. 1993. Demic expansions and human evolution. Science **259**:639–646.

Chikhi, L., G. Destro-Bisol, G. Bertorelle, V. Pascali, and G. Barbujani. 1998*a*. Clines of nuclear DNA markers suggest a largely neolithic ancestry of the European gene pool. Proc. Natl. Acad. Sci. USA **95**:9053–9058.

Chikhi, L., G. Destro-Bisol, V. Pascali, V. Baravelli, M. Dobosz, and G. Barbujani. 1998*b*. Clinal variation in the nuclear DNA of Europeans. Hum. Biol. **70**:643–657.

Costa, R., A. A. Peixoto, G. Barbujani, and C. P. Kyriacou. 1992. A latitudinal cline in a Drosophila clock gene. Proc. R. Soc. Lond. B **250**:43–49.

Currat, M., and L. Excoffier. 2005. The effect of the Neolithic expansion on European molecular diversity. Proc. Biol. Sci. **272**:679–688.

Currat, M., N. Ray, and L. Excoffier. 2004. SPLATCHE: a program to simulate genetic diversity taking into account environmental heterogeneity. Mol. Ecol. Notes **4**:139.

Dupanloup, I., G. Bertorelle, L. Chikhi, and G. Barbujani. 2004. Estimating the impact of prehistoric admixture on the genome of Europeans. Mol. Biol. Evol. **21**:1361–1372.

Edmonds, C. A., A. S. Lillie, and L. L. Cavalli-Sforza. 2004. Mutations arising in the wave front of an expanding population. Proc. Natl. Acad. Sci. USA **101**:975–979.

Estoup, A., M. Beaumont, F. Sennedot, C. Moritz, and J. M. Cornuet. 2004. Genetic analysis of complex demographic scenarios: spatially expanding populations of the cane toad, Bufo marinus. Evolution **58**:2021–2036.

Eswaran, V., H. Harpending, and A. R. Rogers. 2005. Genomics refutes an exclusively African origin of humans. J. Hum. Evol. **49**:1–18.

Evans, P. D., S. L. Gilbert, N. Mekel-Bobrov, E. J. Vallender, J. R. Anderson, L. M. Vaez-Azizi, S. A. Tishkoff, R. R. Hudson, and B. T. Lahn. 2005. Microcephalin, a gene regulating brain

size, continues to evolve adaptively in humans. Science **309**:1717–1720.
Excoffier, L. 2004. Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island models. Mol. Ecol. **13**:853–864.
Fix, A. G. 1997. Gene frequency clines produced by kin-structured founder effects. Hum. Biol. **69**:663–673.
Fort, J., T. Pujol, and L. L. Cavalli-Sforza. 2004. Palaeolithic populations and waves of advance. Camb. Archaeol. J. **14**: 53–61.
Gavrilets, S. 2003. Perspective: models of speciation: what have we learned in 40 years? Evolution Int. J. Org. Evolution **57**: 2197–2215.
Gillespie, J. H. 1991. The causes of molecular evolution. Oxford University Press, Oxford.
Hewitt, G. M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biol. J. Linn. Soc. **58**:247–276.
Hill, E. W., M. A. Jobling, and D. G. Bradley. 2000. Y-chromosome variation and Irish origins. Nature **404**:351–352.
Housley, R. A., C. S. Gamble, M. Street, and P. Pettit. 1997. Radiocarbon evidence for the lateglacial human recolonisation of northern Europe. Proc. Prehist. Soc. **63**:25–54.
Kimura, M. 1962. On the probability of fixation of mutant genes in a population. Genetics **47**:713–719.
Li, W. H. 1997. Molecular evolution. Sinauer Associates, Inc., Sunderland, Mass.
Lucotte, G. 2001. Frequency analysis and allele map in favor of the celtic origin of the C282Y mutation of hemochromatosis. Blood Cells Mol. Dis. **27**:549–556.
Maruyama, T. 1970. On the fixation probability of mutant genes in a subdivided population. Genet. Res. **15**:221–225.
Mateu, E., F. Calafell, M. D. Ramos, T. Casals, and J. Bertranpetit. 2002. Can a place of origin of the main cystic fibrosis mutations be identified? Am. J. Hum. Genet. **70**: 257–264.
Mekel-Bobrov, N., S. L. Gilbert, P. D. Evans, E. J. Vallender, J. R. Anderson, R. R. Hudson, S. A. Tishkoff, and B. T. Lahn. 2005. Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. Science **309**:1720–1722.
Mellars, P. 2004. Neanderthals and the modern human colonization of Europe. Nature **432**:461–465.
Menozzi, P., A. Piazza, and L. L. Cavalli-Sforza. 1978. Synthetic maps of human gene frequencies in Europeans. Science **201**:786–792.
Nei, M., T. Maruyama, and R. Chakraborty. 1975. The bottleneck effect and genetic variability in populations. Evolution **29**:1–10.
Otto, S. P., and M. C. Whitlock. 1997. The probability of fixation in populations of changing size. Genetics **146**:723–733.
Ray, N., M. Currat, and L. Excoffier. 2003. Intra-deme molecular diversity in spatially expanding populations. Mol. Biol. Evol. **20**:76–86.
Richards, M., H. Corte-Real, P. Forster, V. Macaulay, H. Wilkinson-Herbots, A. Demaine, S. Papiha, R. Hedges, H. J. Bandelt, and B. Sykes. 1996. Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am. J. Hum. Genet. **59**:185–203.
Rosser, Z. H., T. Zerjal, M. E. Hurles et al. (60 co-authors). 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am. J. Hum. Genet. **67**:1526–1543.
Slatkin, M. 1980. The distribution of mutant alleles in a subdivided population. Genetics **95**:503–524.
Tighe, O., D. Dunican, C. O'Neill et al. (25 co-authors). 2003. Genetic diversity within the R408W phenylketonuria mutation lineages in Europe. Hum. Mutat. **21**:387–393.

Lisa Matisoo-Smith, Associate Editor

Accepted October 19, 2005