Mutations arising in the wave front of an expanding population

Christopher A. Edmonds, Anita S. Lillie, and L. Luca Cavalli-Sforza*

Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305

Contributed by L. Luca Cavalli-Sforza, December 4, 2003

The ability to infer the time and place of origin of a mutation can be very useful when reconstructing the evolutionary histories of populations and species. We use forward computer simulations of population growth, migration, and mutation in an analysis of an expanding population with a wave front that advances at a constant slow rate. A pronounced founder effect can be observed among mutations arising in this wave front where extreme population bottlenecks arise and are followed by major population growth. A fraction of mutations travel with the wave front and generate mutant populations that are on average much larger than those that remain stationary. Analysis of the diffusion of these mutants makes it possible to reconstruct migratory trajectories during population expansions, thus helping us better understand observed patterns in the evolution of species such as modern humans. Examination of some historical data supports our model.

Certain cases of punctuated evolution may result from demographic growth causing geographic expansion of an initially small population into a large area, a phenomenon called demic diffusion (1). According to the standard model of human evolution (2), this is the way in which modern humans radiated from an initially small population in east Africa to the whole world over the last 50,000 years, a conclusion supported by archeology (3), by the genealogies reconstructed with polymorphisms of the nonrecombining portion of the Y chromosome (NRY) (4), by a similar analysis of mitochondrial DNA (mtDNA) data (5), and by examination of autosomal microsatellites (6). The geography of the major NRY haplogroups shows a particular pattern that has been used to give a qualitative interpretation of the times and places of occurrence of the major mutational events marking the human expansion (4) and to help reconstruct the dispersal routes.

A mathematical model by R. A. Fisher describes some major features of population demic expansions. The model was first generated for the wave of advance of advantageous genes (7) but can be extended to that of a growing population (1, 8, 9), under the joint effect of logistic growth of the population (with initial growth rate of g children per parent and a uniform saturation level of N_s individuals per deme), and of migration (by random diffusion with probability m). The major feature of this model is that, in a population expansion, a fairly narrow wave front is formed, defined as the edge of the wave of advance of the population, which moves forward at a constant radial rate determined by the geometric mean of m and g (7).

Here, we analyze by computer simulation possible rules relevant to the quantitative prediction of the place of origin of mutations. Combined with existing methods for inferring the time of origin, the procedure should help make it possible to describe accurately the trajectories of migration of haplotypes during the expansion, by joining with segments the places of origin of the mutations in the sequence in which they appear in the genealogy.

The Simulation Model

The experiments simulate uniparentally transmitted haploids (such as those of the NRY or mtDNA) that reproduce at a constant rate with discrete generations, with the number g of children per parent obeying a Poisson distribution. The simulation occurs on a 25×100 grid of demes, with progeny remaining at the same location as the parent with probability $1 - m$ and migrating with probability m/n to enter any of the n immediately adjacent demes. n varies from 2 to 4, depending on the position of each deme with respect to the grid boundaries. Each deme has a population of N individuals at any given generation, limited to a maximum of $N_s = 10$. Each iteration of the simulation represents one generation, and the cycle within each iteration is as follows: reproduction, migration, and finally random culling of individuals in demes where $N > N_s$.

Each simulation is started with a nonmutant individual seeded in a deme at the extreme left of the rectangle (longitude $j = 1$), with random choice of initial latitude i at time $t = 0$. As the simulation progresses, the population grows, and a wave front forms and moves toward the right at a relatively constant rate. Behind the wave front is a saturated area where the population of almost every deme is equal to the maximum, i.e., $N = N_s$. When the expanding population reaches a predetermined longitude of the grid (varied in steps of 10 from $j = 10$ to $j = 80$ in eight different sets of 8,000 simulations each), one individual located randomly in the wave front is mutated. The population continues to expand, with mutants multiplying, migrating, and dying out at the same rate as nonmutants. The simulation is stopped when either all of the mutants die out before the end of the simulation, the entire population dies out by chance (a rare event except at the very beginning), or the entire grid is saturated. Simulations are considered successful when, at the time the expansion has reached the right edge, there is still at least one mutant alive in the population. Fig. 1 shows an example of a successful simulation.

The $N_s = 10$ maximum value per deme may seem low when compared with the expectation for a stationary population at the end of the expansion of modern humans. Considering, however, that only one haploid gender is active and that each generation exists at only one time in our model, the actual number of individuals per deme is ≈60, including the other gender and other generations, and is therefore rather similar to that of a hunting band (10). In an asexual system, there is no need to consider larger social units, and thus the amount of simulated drift is not at all unrealistic for the NRY or mtDNA of a hunting–gathering population.

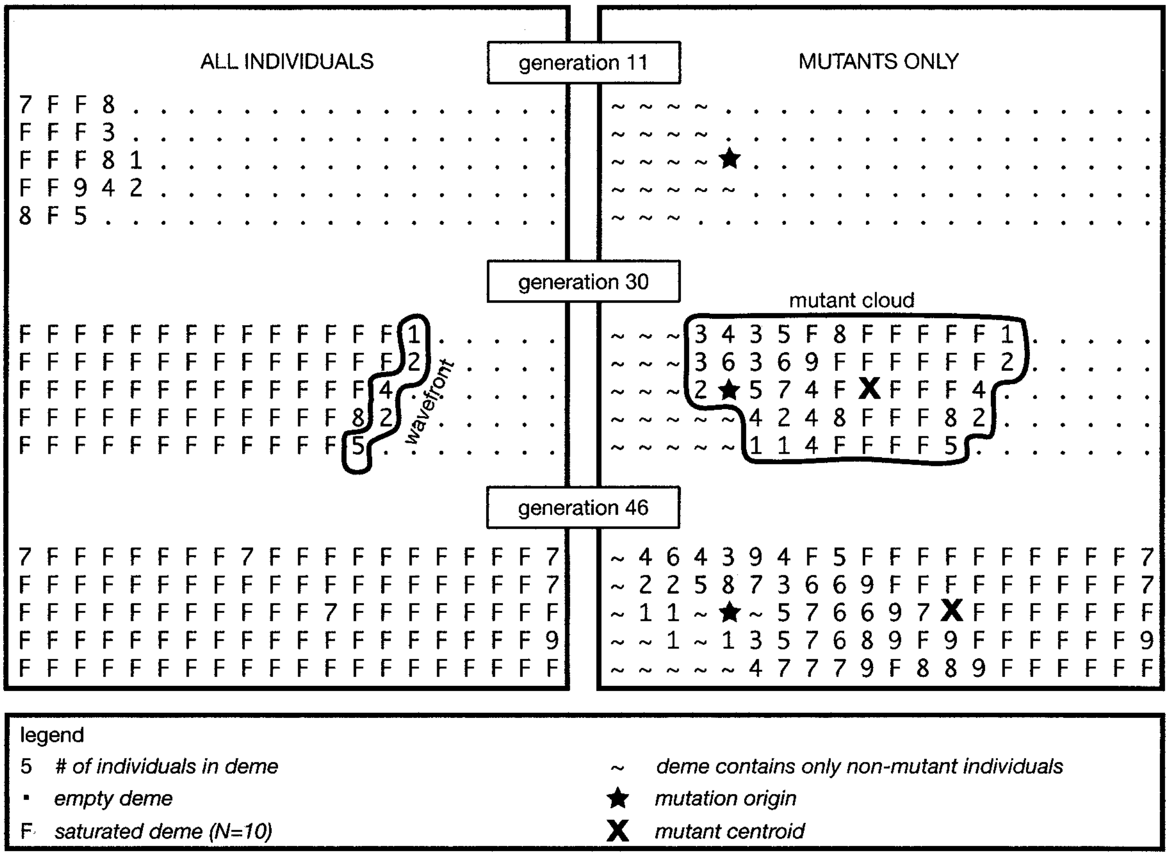
The Mutation Traveling Phenomenon

Fig. 2 shows the end results of six successful simulations; in each the mutation was seeded when the wave front had reached longitude $j = 20$. The six simulations were chosen to reflect some of the common results of the experiment and exemplify the variety of evolutionary success and the variation in the geographic pattern of the mutants. The star indicates the origin of the mutation, X the centroid (the deme whose coordinates are

Abbreviation: NRY, nonrecombining portion of the Y chromosome.

*To whom correspondence should be addressed. E-mail: cavalli@stanford.edu.

www.pnas.org/cgi/doi/10.1073/pnas.0308064100



ALL INDIVIDUALS

MUTANTS ONLY

generation 11

```
7 F F 8 . . . . . . . . . . . . . . . .    ~ ~ ~ ~ . . . . . . . . . . . . . . . .
F F F 3 . . . . . . . . . . . . . . . .    ~ ~ ~ ~ . . . . . . . . . . . . . . . .
F F F 8 1 . . . . . . . . . . . . . . .    ~ ~ ~ ~ ★ . . . . . . . . . . . . . . .
F F 9 4 2 . . . . . . . . . . . . . . .    ~ ~ ~ ~ ~ . . . . . . . . . . . . . . .
8 F 5 . . . . . . . . . . . . . . . . .    ~ ~ ~ . . . . . . . . . . . . . . . . .
```

generation 30

wavefront

mutant cloud

```
F F F F F F F F F F F F F F 1 . . . . .    ~ ~ ~ 3 4 3 5 F 8 F F F F F 1 . . . . .
F F F F F F F F F F F F F F 2 . . . . .    ~ ~ ~ 3 6 3 6 9 F F F F F F 2 . . . . .
F F F F F F F F F F F F F 4 . . . . . .    ~ ~ ~ 2 ★ 5 7 4 F X F F F 4 . . . . . .
F F F F F F F F F F F F 8 2 . . . . . .    ~ ~ ~ ~ ~ 4 2 4 8 F F F 8 2 . . . . . .
F F F F F F F F F F F F 5 . . . . . . .    ~ ~ ~ ~ ~ 1 1 4 F F F F 5 . . . . . . .
```

generation 46

```
7 F F F F F F F 7 F F F F F F F F F F 7    ~ 4 6 4 3 9 4 F 5 F F F F F F F F F F 7
F F F F F F F F F F F F F F F F F F F 7    ~ 2 2 5 8 7 3 6 6 9 F F F F F F F F F 7
F F F F F F F F F F F 7 F F F F F F F F    ~ 1 1 ~ ★ ~ 5 7 6 6 9 7 X F F F F F F F
F F F F F F F F F F F F F F F F F F F 9    ~ ~ 1 ~ 1 3 5 7 6 8 9 F 9 F F F F F F 9
F F F F F F F F F F F F F F F F F F F F    ~ ~ ~ ~ ~ 4 7 7 7 9 F 8 8 9 F F F F F F
```

legend

5	*# of individuals in deme*	~	*deme contains only non-mutant individuals*
·	*empty deme*	★	*mutation origin*
F	*saturated deme (N=10)*	X	*mutant centroid*

Fig. 1. Three successive simulations (expressed in number of generations since the beginning) in a single experiment. This figure illustrates the expansion of the population and the subsequent dispersion of the mutant subpopulation within it. For graphic clarity, this example uses a smaller grid size. Note the success of the mutant population as it propagates with the wave front, as well as the movement of the mutant centroid with respect to the mutant origin.

the average latitude and longitude of the mutants). Block shadings indicate the relative mutant frequency in each deme. In three of the simulations shown here, mutations stayed near their origin or moved only slightly; in the other three, they moved with the wave front from their origin. One was relatively unsuccessful whereas the remaining two had greater degrees of mutant

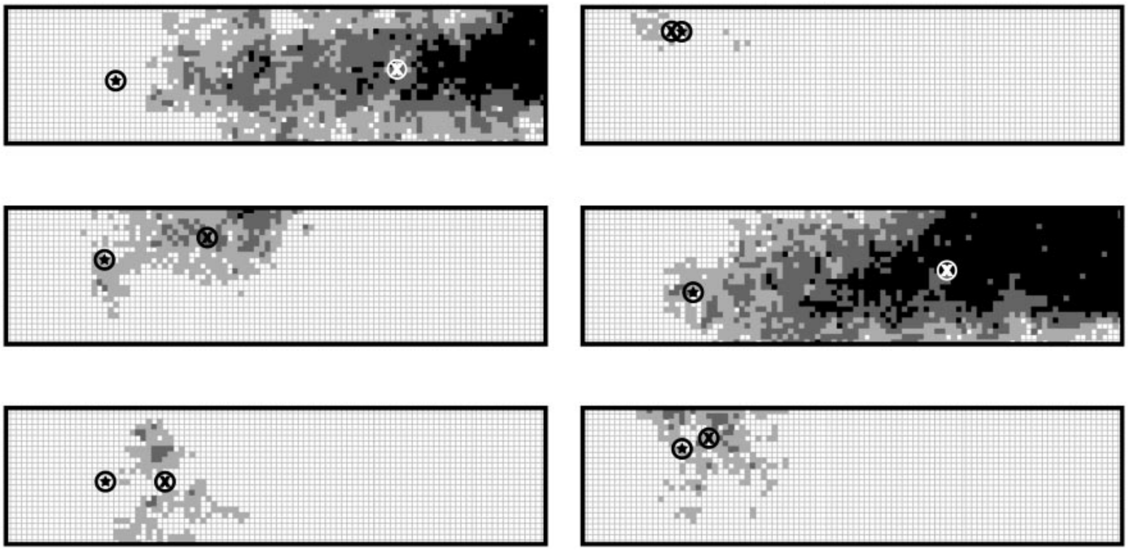


Fig. 2. End results of six independent simulations. The star indicates the place of origin of the mutation, and X indicates the centroid of mutants at the end of the simulation. The block shading indicates the proportion of mutants in that deme, subdivided in four classes (white, 0–25%; black, 75–100%). The data for this graph are from simulations seeded with one mutation at a random position in the wave front when it had reached longitude $j = 20$. Parameters used were migration rate $m = 0.4$, growth rate $g = 1.8$ expected children per parent, and saturation $N_s = 10$ per deme.

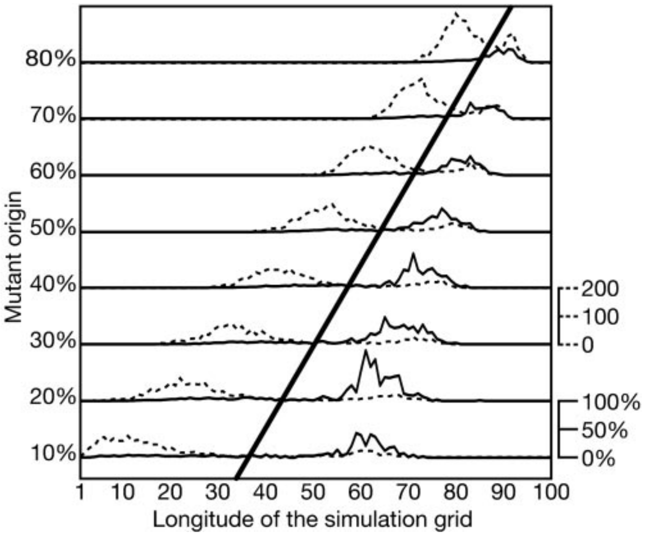


Fig. 3. Distribution of mutants from eight experiments of 8,000 simulations each. Each experiment tested a different mutant origin by varying the longitude of the origin from $j = 10$ to $j = 80$ on a grid of width 100. The ordinates of the dotted curves indicate the numbers of simulations with centroids at that longitude, and the ordinates of the solid curves indicate the percentages of mutants. The oblique straight line separates the mutant populations that are stationary (at the left) from those that are traveling (at the right) and is fitted from the mean of the centroid means given in Table 2.

success. The latter two show high mutant frequencies at the end of the expansion.

Our major interest was to compare the place of origin of mutation with the location of the centroid at the end of the simulation, to predict the former on the basis of the latter. Of the 64,000 simulations, ≈18% were successful, and Fig. 3 shows different distributions of the longitudes of the centroids of mutants at the simulations' ends. All of the distributions are strongly bimodal, with a left mode centered around the place of origin of the mutation and a right mode approximately midway between the place of origin of the mutation and the right edge of the grid, i.e., the end of the expansion. The ordinates expressed in the scales at the right of the graphs indicate the quantities measured: the dotted curves represent the total numbers of successful simulations (and therefore of mutations) that made it to the end, the solid curves the total percentage of mutants observed in those simulations. The bimodality shows that each mutation followed one of two sharply different behaviors: it either remained close to the place of origin (the left modes) or traveled toward the end of the expansion (the right modes). We found no evidence in any of our experiments of mixed behavior. Statistics of the numbers of mutants and their average locations are given in Tables 1 and 2. Altogether, the majority of mutations remained near the origin (≈78%) whereas the remainder (≈22%) traveled in the direction of the expansion, and on average their centroid can be found about midway between the place of origin of the mutation and the end of the expansion.

Tables 1 and 2 show that among simulations there is extreme variation among the number of mutants, measured as the distribution of the percentages of mutants with respect to the total population at the end of each simulation. There are two major differences between mutants that remain near the origin (termed stationary) and those that travel, in addition to the relative position of their centroids. The stationary ones have, on average, very few mutants; in the majority of simulations, the number of mutants remains below the level of polymorphism (conventionally set at 1% of the total population) whereas, in the traveling ones, mutants may even become the majority and are concentrated at the end of the expansion. The location of the centroids, moreover, is more variable for the stationary ones (by a factor >2). This result is reminiscent of what is observed in the geographic distribution of the oldest NRY and mtDNA mutants in maps of the continents and of the world (4, 5) where several early mutations reach their highest frequencies (as much as 100%) at the extreme periphery of the expansion. This behavior is largely determined by the population size of the deme in which the mutant appears. If the mutation happens to arise in a wave front deme with just one individual, the mutant frequency in that deme will be 100% for that generation, and the chance that the mutant will survive and prosper is high. In general, if there is one mutation in a population of N individuals, its chance of final fixation is $1/N$. Clearly, the lowest chances of success are for mutations arising in demes saturated by nonmutant individuals. Fig. 4 summarizes this finding, showing that the probability of success is strictly controlled by the population size of the deme on the wave front in which the initial mutant appears, and not by the migration rate m. From the same figure, however, it is clear that the migration rate has some effect on the structure of the wave front, as expected from the theory of the wave of the advance (7): a higher migration rate tends to increase the proportion of small size demes at the edge of the wave front, thus causing an overall increase in the success rate of traveling mutations.

Consideration of the wave front structure also explains why mutations arising in the wave front tend to move in the direction of the expansion. Progeny of the mutants migrating to adjacent areas located in the direction of the expansion, which are unoccupied or occupied by a few nonmutant individuals, are more likely to be at a higher proportion with respect to nonmutants than in the deme in which they were originally located and have a reasonable chance of even reaching 100% frequency

Table 1. Simulations classed by size of mutant population

Longitude of mutation origin	Percentage of mutants at end of simulation					Total
	<1%	1–3.2%	>3.2–10%	>10–31.6%	>31.6–100%	
10	759 (0.0%)	261 (0.4%)	104 (11.5%)	98 (82.7%)	135 (100%)	1357 (15.6%)
20	725 (0.0%)	245 (1.2%)	115 (12.2%)	101 (92.1%)	103 (100%)	1289 (15.6%)
30	699 (0.0%)	208 (2.4%)	127 (37.8%)	140 (97.9%)	63 (100%)	1237 (18.5%)
40	754 (0.0%)	214 (4.2%)	102 (49.0%)	168 (98.8%)	27 (100%)	1265 (18.2%)
50	911 (0.5%)	192 (6.3%)	95 (71.6%)	161 (100%)	4 (100%)	1363 (17.8%)
60	1029 (1.4%)	200 (16.0%)	169 (87.6%)	114 (100%)	0	1512 (20.4%)
70	1200 (2.3%)	227 (43.2%)	204 (97.1%)	56 (100%)	0	1687 (25.0%)
80	1390 (7.2%)	288 (77.4%)	209 (99.5%)	2 (100%)	0	1889 (36.7%)

Data from each of the eight experiments shown in Fig. 3. Each column represents the number of successful simulations with mutant populations in that class. The percentages of those simulations where the mutant population traveled are given in parentheses.

Table 2. Statistics of mutant centroids

Longitude of mutation origin	No. of successful simulations		Mean of mutant centroid longitudes		SD of mutant centroid longitudes		Mean of % of mutants		SD of % of mutants	
	Stationary	Traveled	Stationary	Traveled	Stationary	Traveled	Stationary	Traveled	Stationary	Traveled
10	1,145	212	13.2	58.6	16.5	7.4	1.5	42.9	2.5	20.9
20	1,088	201	24.1	63.6	15.9	6.8	1.3	32.9	1.9	15.1
30	1,008	229	33.8	67.3	13.1	7.3	1.2	23.7	1.9	13.5
40	1,035	230	43.1	71.6	12.2	6.0	1.0	19.8	1.4	10.2
50	1,120	243	52.5	77.1	11.2	4.9	0.6	14.6	0.8	7.8
60	1,204	308	62.0	81.0	9.3	4.2	0.6	9.1	0.8	6.0
70	1,266	421	71.5	85.8	7.2	3.2	0.4	5.2	0.5	3.9
80	1,196	693	80.3	90.4	5.2	2.2	0.2	2.4	0.3	2.1

Columns 2 and 3 show the number of successful simulations divided according to the behavior of the mutants (stationary or traveling), as decided on the basis of the separation line of Fig. 3. Columns 4 and 5 show the means and 6 and 7 show the SD of the mutant centroid longitudes for simulations divided in the same manner. Similarly, columns 8 and 9 show the means and 10 and 11 show the SD of the percentage of mutants in the population, also divided according to mutant behavior.

when moving to the periphery of the expansion. One therefore expects, and indeed observes in Fig. 4, that the probability of success of the mutant is higher when the number of individuals in the deme in which the mutation originated is smaller. The existence of the wave front will generate frequent extreme bottlenecks and therefore strong drift effects, resulting in high mutant frequencies, especially in demes located in the expanding edge of the population. Moreover, mutants that move in the direction of the expansion will still be in demes with fewer individuals where they may be in relatively higher proportions and therefore have a greater chance of final success. This asymmetry causes the traveling phenomenon we have described.

The centroid of traveling mutants is located approximately midway between the end of the expansion and the place of origin of the mutation (expansions usually proceeded until they met the continental edge although, on occasion, they did cross bodies of water when there was some assumption that the opposite side was reachable via the means of transportation available). This finding suggests that the origin of a traveling mutation can be estimated to be between the origin of the expansion (if known) and the centroid, at a distance from the end of the expansion equal to about twice that between the centroid and the expansion end. We will refer to this rule as "twice the distance," noting, however, that it might be a little more than twice (≈2.2 times in the conditions of the simulation shown in Table 1) when judging from the precise relationship between the centroid and the place of origin of the mutation shown in Fig. 3 and Table 2. In practice, one would need to know the origin of the expansion and draw a straight line through it and the centroid of the mutants being considered. The straight line could then be continued until the geographic end of the expansion, presumably the continental edge or other insurmountable geographic barrier. The most likely place of origin would then be determined by the twice-the-distance rule. This result assumes that the migratory path was straight, which is only an average, but this assumption may be compensated for by knowledge of the terrain. If the reciprocal order (the time sequence) of many mutations is known from the genealogy, one could trace segments from one mutation to the next. This statement depends on the assumption that the mutation traveled, which is known only in probability. There is other knowledge, however, that may be used to overcome this uncertainty.

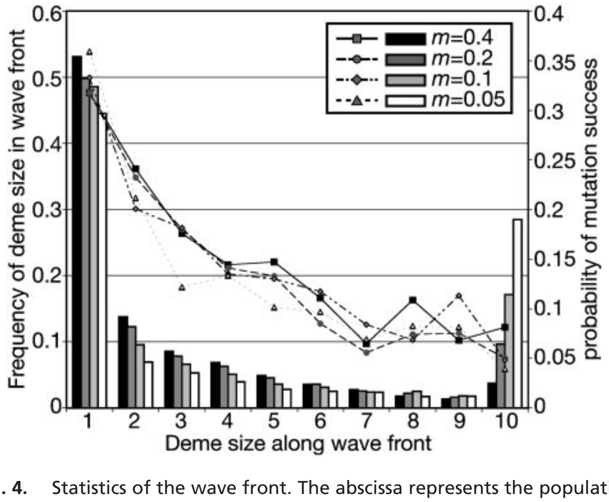


Fig. 4. Statistics of the wave front. The abscissa represents the population sizes of the demes located along the wave front. The ordinate of the curves for four different migration rates *m* are the probabilities of success of mutations as a function of the deme size (scale on the right ordinate). The rectangles (scale on the left ordinate) show the effects of migration rates on the average relative frequency of deme size and give an indication of the shape of the wave front in terms of deme population.

Inference of the Place of Origin

In a population that is largely stationary and has not undergone a demographic and geographic expansion, the mutation's place of origin would normally be inferred from its centroid. It can be seen, however, that, even for stationary mutations, there is a random error for this type of inference, indicated by the standard deviations of Table 2, which increase with mutation age (a phenomenon investigated more generally by S. Gimelfarb, unpublished results). It is clear, however, that, in an expanding population, the centroid of a mutation cannot be accepted as an unbiased estimate of the place of origin because it tends to be displaced toward the end of the expansion by the traveling phenomenon.

On initial examination, it may seem that the difficulty of deciding whether a mutation did or did not travel makes the problem of identifying its place of origin insoluble. Several facts will usually help to make the inference safe. The first consideration is that the great majority of mutants that remain near their place of origin end up having very low frequency. In the conditions of our simulations, they are nonpolymorphic in the majority of the cases and had frequency >10% only 0.5% of the time. In practice, the most interesting mutations are those that reach high frequencies; if one can establish their location and if we have information on enough mutations, we can reconstruct the early paths of migratory streams that helped settle the area inhabited by the species: the whole world in the case of humans.

An alternative method (for estimating the place of origin of a mutation) that ignores the indeterminacy problem is to place the mutation on the straight line between the origin and the end of the expansion, passing through the mutant's centroid, at a location calculated from the time of origin of the mutation as deduced from the genealogy. This approach is based on the assumption that the expansion proceeded at a constant rate. In some cases (e.g., the spread of farming to Europe), this assumption is shown to be correct by archeological dates of the first arrival of agriculture (1) although the rate varied somewhat in the Mediterranean and in central Europe. The statistical error of times of origin of mutations is still rather high, but improvements in data and methods will reduce this source of error. One can test this method against the use of the rule we have proposed here, but, in general, knowledge of the genealogy will allow us to recognize mutations that are early and frequent as having a very high probability of having traveled. It should also be noted that it is usually difficult to exclude a contribution by natural selection although, in some cases, there may be evidence in favor of it for a few mutations (5).

Conclusions

The most interesting conclusion is that it is possible to estimate the place of origin of a mutation by a study of the location of the mutants, but there is an unexpected uncertainty derived from a strong stochastic effect, whereby a mutation "decides" shortly after its origin whether it will travel fast and generate more mutants at the end of the expansion or remain stationary and be on average much less successful. For this reason, the mutant's centroid cannot be used directly for inferring the place where the mutations originated; rather, it is necessary to shift the origin toward the beginning of the expansion. The systematic error that arises from using the observed centroid as a place of origin is greatest for the early and most productive mutations in the genealogy of a new species, and thus the centroid can supply a reasonable estimate only for very late and infrequent mutations.

A successful species like that of modern humans has had many major demographic and geographic expansions. The geographic and temporal trajectories of its haplotypes can be reconstructed if one can calculate the times and places of origin of the mutations that appear in their genealogies. The results will be especially informative if the genealogy is rich with mutations and can thus offer a satisfactory picture of the migratory streams occurring in the expansion. It is also necessary to have an appropriate set of geographic data. The haploid asexual transmission of the NRY and mtDNA can generate particularly useful information for this analysis. It is probably easier to reconstruct trajectories in the NRY because the low mutation rate avoids uncertainties associated with mutation recurrence. Here, it is clear that many of the earliest mutations must have occurred closer to the expansion origin than their centroid would imply. This finding must be true, for instance, of well known mutations such as *YAP*, *RPS4Y*, *M89*, *M9*, *M45*, *M3*, and others. Qualitative rules were used in earlier maps (4), but the data available so far are still inadequate for a quantitative analysis.

Autosomal data also lend themselves to this study, provided mutation rates are low and natural selection minimal. As an example, Fisher offered a hypothesis on the evolution of the RH gene system, according to which the ancestral haplotype *Dce* underwent three mutations, producing haplotypes, *DCe*, *DcE*, and *dce*, which have the highest frequencies (as described in ref. 11). Later, the other four major haplotypes, still now in lesser frequencies than the first four, were produced, three from single crossovers between the first four (*DCe*, *dCe*, *dcE*). The very rare *dCE* haplotype was hypothesized by Fisher to have originated from a second crossover between two haplotypes of single crossovers. The geographic distribution is in agreement. Modern humans started developing in east Africa and first colonized the rest of Africa where the original haplotype *Dce* predominates. They then spread first to central Asia, where a major demographic explosion began shortly after 50,000 years ago, and started expanding in all directions. Interestingly, the center of Asia, which was most probably an early nexus of expansion of modern humans, is one of the very few places in the world outside Africa that has a nonnegligible frequency of the ancestral haplotype, in agreement with the idea that the major mutations occurred after the expansion from the center. The mutation $c \rightarrow C$ occurred in the expansion to southeast Asia, and the resulting haplotype *DCe* is predominant in southeast Asia and is at its maximum in Oceania. The mutation $e \rightarrow E$ took place in the expansion to the north of Asia, and *DcE* has a maximum frequency in the north of Siberia. In the expansions toward west Asia and eventually Europe, $D \rightarrow d$ took place, generating *dce* (the standard RH negative haplotype). Like the other two, *dce* also reaches its maximum value at the farthest point from the Asian center of expansion, in the Basque region (12). Probably, the *d* mutation was affected more than any other by adverse natural selection because of the well known immunological incompatibility system affecting it (11) but was also helped by drift in the demographic bottleneck during the last glaciation.

Mathematical investigations by Vlad *et al*. (13) have shown that diffusion equations incorporating expansion models, e.g., those described by Fisher's differential equation, are common to the theory of hydrodynamic flow and to that of genetic expansions. They describe phenomena very similar to those shown here by simulation and may possibly be used to predict more exactly the location of mutational events. Probabilistic analysis of the drift component may continue to need an additional treatment similar to Markov chain Monte Carlo methods.

We are grateful for comments and suggestions by L. Excoffier, M. Feldman, S. Gimelfarb, J. Mountain, J. Ross, P. Underhill, and M. Vlad. This work has been supported in part by National Institute of General Medical Sciences Grant 28428.

1. Ammerman, A. A. & Cavalli-Sforza, L. L. (1984) *The Neolithic Transition and the Population Genetics of Europe* (Princeton Univ. Press, Princeton).
2. Cavalli-Sforza, L. L. & Feldman, M. W. (2003) *Nat. Genet. Supp.* **33,** 266–275.
3. Klein, R. G. (1999) *The Human Career* (Univ. of Chicago Press, Chicago), 2nd Ed.
4. Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Foley, R. A., Mirazón Lahr, M., Oefner, P. J. & Cavalli-Sforza, L. L. (2001) *Ann. Hum. Genet.* **65,** 43–62.
5. Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G., Hosseini, S., Brandon, M., Easley, K., Chen, E., Brown, M. D., *et al.* (2003) *Proc. Natl. Acad. Sci. USA* **100,** 171–176.
6. Zhivotovsky, L. A., Rosenberg, N. A. & Feldman, M. W. (2003) *Am. J. Hum. Genet.* **72,** 1171–1186.
7. Fisher, R. A. (1937) *Ann. Eugen.* **7,** 355–369.
8. Skellam, J. (1951) *Biometrika* **38,** 196–218.
9. Rendine, S., Piazza, A. & Cavalli-Sforza, L. L. (1986) *Am. Nat.* **128,** 762–768.
10. Cavalli-Sforza, L. L., ed. (1986) *African Pygmies* (Academic, Orlando, FL).
11. Cavalli-Sforza, L. L. & Bodmer., W. F. (1999) *The Genetics of Human Populations* (Dover, New York).
12. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1994) *History and Geography of Human Genes* (Princeton Univ. Press, Princeton).
13. Vlad, M. O., Moran, F., Tsuchiya, M., Cavalli-Sforza, L. L., Oefner, P. J. & Ross, J. (2002) *Phys. Rev. E.* **65,** 061110-1–061110-17.