Phylogenetic fate mapping

Stephen J. Salipante* and Marshall S. Horwitz*^{†‡}

Departments of *Genome Sciences and [†]Medicine, Division of Medical Genetics, University of Washington School of Medicine, Box 357720, Seattle, WA 98195

Communicated by Stanley M. Gartler, University of Washington, Seattle, WA, February 14, 2006 (received for review January 23, 2006)

Cell fate maps describe how the sequence of cell division, migration, and apoptosis transform a zygote into an adult. Yet, it is only in *Caenorhabditis elegans* where microscopic observation of each cell division has allowed for construction of a complete fate map. More complex, and opaque, animals prove less yielding. DNA replication, however, generates somatic mutations. Consequently, multicellular organisms comprise mosaics where most cells acquire unique genomes that are potentially capable of delineating their ancestry. Here we take a phylogenetic approach to passively retrace embryonic relationships by deducing the order in which mutations have arisen during development. We show that polyguanine repeat DNA sequences are particularly useful genetic markers, because they frequently change length during mitosis. To demonstrate feasibility, we phylogenetically reconstruct the lineage of cultured mouse NIH 3T3 cells based on mutations affecting the length of polyguanine markers. We then employ whole genome amplification to genotype polyguanine markers in single cells taken from a mouse and use phylogenetics to infer the developmental relationships of the sampled tissues. The result is consistent with the present understanding of embryogenesis and demonstrates the large scale potential of this method for producing a complete mammalian cell fate at the resolution of a single cell.

cell fate | development | genomic instability

The construction of the *Caenorhabditis elegans* cell fate map by Sulston *et al.* (1), who used a microscope to painstakingly record the lineage of each of the 671 cells produced during the worm's 12-h embryogenesis, represents a milestone of developmental biology. Mapping cell fate in higher organisms, such as mice, containing a trillion or more cells, also has proven valuable but is more difficult (2, 3) and has required manipulation of the embryo, in which cells are tagged via chimerism (4), dyes (5), radiation-induced cytogenetic abnormalities (6), or virally transferred (7), transgenic (8), or gene-targeted (9–11) markers. However, there are only a limited number of labels that can be applied to a single embryo. Moreover, labeling of cells can demonstrate a common ancestry but cannot define branches within lineages. Consequently, multiple experiments at various stages of embryogenesis may be necessary to establish developmental hierarchies. For example, a fate map of myotomes, using random recombination of a transgenic marker, required examining $\approx 3,000$ mouse embryos (12).

In mammals, however, new mutations arise with almost every mitosis (13), implying that most cells acquire unique genomes. Here we treat the cells of an adult mouse as members of an asexual population, individualized by somatic mutation, yet descended from a common founder (the zygote), and adapt phylogenetic methods used in evolutionary and microbial population biology (14) to trace their embryonic lineage. Previously, somatic hypermutation at Ig loci has partially elucidated lymphoid lineages (15), and postzygotic methylation patterns at CpG sites have allowed for the reconstruction of the ancestry of intestinal crypts (16). Derivation of cell fate via inherent genetic mosaicism could, in principle, offer a simple and passive means for generating a complete map of all lineages from just a single individual. Because it is presently impossible to sequence the genome of every cell, however, the challenge is to identify mutational hotspots sufficiently informative for inferring lineage relationships.

Short, repetitive sequences undergo insertion and deletion mutation at high frequency (17, 18). Occasional, meiotically arising length polymorphisms in short tandem repeat markers serve as the foundation for contemporary genetic linkage analysis used for mapping heritable traits (19) and find application in studies of population migration and evolution (20). Mitotically generated length variation in mononucleotide repeats, especially polyguanine tracts, has proven informative for reconstructing bacterial phylogenies (21). In mammalian cells, too, polyguanine tracts undergo mitotic deletion or insertion mutation at frequencies reported at $\approx 10^{-4}$ per cell per generation (17). Such a remarkable mitotic mutation rate raises the question of whether markers capable of demonstrating relationships between individuals within a population also may be used to infer relationships between individual cells within a multicellular organism.

Results

Test in Cell Culture. As an *in vitro* test, we manufactured a phylogenetic “tree” of mouse NIH 3T3 fibroblasts grown in tissue culture (Fig. 1A). We initially isolated a single NIH 3T3 cell and allowed it to divide 22 times. Then we extracted DNA from half of this clonal, “root” population. Seven cells from the remainder, each grown separately, underwent another 22 divisions, at which point DNA was again isolated from half of each population (A1–A7), and a third generation of single-cell isolates was seeded onto separate dishes (A1→B1–B5, A3→C1–C4, etc.). The cultured cells grew for a total of 133 days and should therefore reasonably approximate the length of time required for a zygote to become an adult mouse.

Using a BLAST (22) search of the mouse genome, we identified uninterrupted polyguanine tracts of 40 bases or less as targets for PCR amplification with fluorescently labeled primers (Table 1, which is published as supporting information on the PNAS web site) and examined the length of the polyguanine tract at 28 such sites in each of the 31 clonal isolates, corresponding to the “nodes” shown in Fig. 1A. It is not possible to determine the true genotype of a particular single-cell isolate, but a sample of part of the clonally expanded population it has given rise to provides an estimate; here, we used half of the population for this purpose. Because the majority of cells will not be mutated at any given locus, DNA purified from the population should provide an average genotype, approximating that of its progenitor cell.

NIH 3T3 cells are diploid (23), and each marker displayed a distribution of intensities centered at either one or two maximal peaks (Fig. 2), corresponding to homozygosity or heterozygosity, respectively, for alleles of a particular autosomal marker. The observed distribution of nonmaximal peaks (“stutter”) largely results from slipped strand mispairing that creates insertions or deletions during PCR amplification (24). Despite the potential for confounding fragment size variation, genotypes prove reproducible between independent PCRs (using the “*D* value” metric, described in *Materials and Methods*).

Although derived from a clonal population, each NIH 3T3 isolate

Conflict of interest statement: No conflicts declared.

Freely available online through the PNAS open access option.

Abbreviation: SCOMP, single-cell comparative genomic hybridization.

[†]To whom correspondence should be addressed. E-mail: horwitz@u.washington.edu.

© 2006 by The National Academy of Sciences of the USA

Fig. 2. Examples of polyguanine marker genotype variation. Electropherograms of the root isolate of NIH 3T3 cell phylogeny aligned with derivatives obtained from various descendants. Allele lengths are indicated. Blue and green traces designate labeling with 6-FAM and HEX, respectively.

did develop mutations and exhibits a unique combination of polymorphisms across a total of 28 polyguanine markers (Table 2, which is published as supporting information on the PNAS web site). Twenty of the markers demonstrated length polymorphisms, with frequency varying between $<7.3 \times 10^{-4}$ and 1.0×10^{-2} mutations per cell per generation per allele (Table 1). We then determined whether the genetic variation was sufficient to allow for reconstruction of the phylogeny that we had manufactured. Two frequently used and complementary phylogenetic methods (25, 26) rely on neighbor-joining and Bayesian inference. Although the two phylogenies were similar in topology, the Bayesian method more accurately reconstructed the tissue culture tree and is shown in Fig. 1*B*. Posterior probabilities, which may be equated with confidence values for the proposed topology, are displayed along corresponding nodes. Branch lengths in the horizontal axis are proportionate to the number of cell divisions.

The topology of the deduced tree favorably compares to what is known. All isolates reliably group with known descendants, but there are several polytomies (B1–B3, E1–E2, D1, and C3), in which multiple related isolates are seen to arise from a single node. Polytomies appear when there are insufficient data to allow assignment of branching order. It is worth emphasizing, however, that the genetic data infers relationships between isolates to a greater resolution than could be known to us while actually constructing the tissue culture phylogeny. For example, among some related isolates (such as A1–A7), the reconstruction has introduced bifurcations and polytomies to more accurately portray the sequence of clonal expansion of a progenitor. Branch lengths correspond to the

number of cell divisions for each sampled node. The differences between the inferred and actual branch lengths may be partly attributable to an experimental inability to measure clonal variation in growth rates within the population of cells while constructing the tree. Again, the deduced tree may be a more accurate rendition given unavoidable uncertainties in the “actual” tree.

Predictably, the resolution decreases with removal of markers from the data set (data not shown), but a sampling limited to just the root and terminal nodes (Fig. 1C; analysis of E1–E2 without B1 and A1, etc.) reproduces the correct topology (Fig. 1D) surprisingly well, with the retention of just one polytomy (E1–E2) and introduction of only one error in branch order (at H1–H2). The method, therefore, is relatively robust, even without sampling ancestral cells, such as might occur with developmentally programmed apoptosis. (The root can be estimated by averaging sampled nodes.) Analogous to the situation of extinct progenitor species in evolutionary biology, the absence of access to embryonic tissue when sampling an adult organism should not be a limiting factor.

Phylogenetically Mapping Single Mouse Cells. We then attempted to construct a cell fate map of a mouse by examining tissues obtained from a single adult specimen. An initial consideration is that only mutations that have arisen early enough in development to have been inherited by a majority of the cells in the sampled tissue can be detected. Another issue is that tissues are composed of multiple parenchymal cell types and may also contain vascular, nervous, lymphatic, and connective components, each having different embryological origins. For these reasons, we focused on genotyping individual cells. We therefore purified single types of cells from different tissues of a mouse.

We used enzymatic digestion to release individual cells from tissue samples, and isolated neurons from the cerebral cortex and cerebellum, tubular epithelial cells from the renal cortices, myocytes from the apex of the heart's left ventricle, serous cells from the parotid glands, and hepatocytes from the liver, all from the same mouse. Because each of these cell types constitutes a majority in the tissue from which it was obtained, and because cell dissociation adequately preserves histology (27), we felt that cell types could be confidently assigned on the basis of microscopic appearance. We collected a large number of hepatocytes because of the liver's well defined anatomy and the ease with which cells could be isolated.

To obtain multiple genotypes from a single cell containing just one diploid copy of the genome, however, it is necessary to first generate larger quantities of that cell's DNA. We therefore evaluated different methods of whole genome amplification before selecting “single-cell comparative genomic hybridization” (SCOMP), a form of ligation-mediated PCR (28). Whole genome amplification does not faithfully copy all sequences (29), so to evaluate its fidelity at markers containing polyguanine repeats, we compared the SCOMP-amplified genotypes of individual cells that had recently divided in culture. We clonally isolated NIH 3T3 cells, observed each cell as it proceeded through two or three cell divisions, then separated each of the daughter cells and subjected each cell to SCOMP whole genome amplification. We then compared the genotypes for different polyguanine markers between the cells. Because the mutation frequency at polyguanine markers in NIH 3T3 cells is no greater than 10^{-2} per allele per mitosis, the probability of a mutation occurring over the course of two or three cell divisions is low ($< 1 - 0.99^3 = 0.03$), and we can assume that differences between cells within the same small colony represent artifacts of whole genome amplification. Taking the most commonly observed genotype in a population of cells as the actual genotype for the colony, we estimated the frequency of errors introduced by SCOMP whole genome amplification by comparing genotypes among cells from the same colony. Fig. 4, which is published as supporting information on the PNAS web site, represents results with four different polyguanine markers, where each genotype is replicated six times for each cell in a four-cell colony,

and the resulting 24 electropherograms for each marker are superimposed. Artifacts are limited; of a total of 156 alleles genotyped across eight different polyguanine markers (with multiple replications per sample), only 13 (8%) were inconsistent with the predicted true genotype of the colony.

We next genotyped single cells obtained from the mouse. Not all of the markers used for NIH 3T3 cells proved to be suitable. Some did not yield PCR products from mouse tissues, presumably because of strain differences in the genomic sequence surrounding repeats. Some markers, and a few cells, failed to consistently amplify with SCOMP whole genome amplification. Therefore, we needed to evaluate additional polyguanine repeat loci to generate a panel of 31 usable markers (Table 1). Finally, based on the above estimated error rate of 8% for SCOMP, we discarded six markers which displayed $< 10\%$ polymorphism across samples.

We then genotyped the resulting DNA from each of 84 mouse cells with each of the 31 markers (although a genotype could not be determined for every marker from every cell). As was the case for the manufactured NIH 3T3 phylogeny, each mouse cell displayed a unique combination of genotypes across the panel of markers tested. We used phylogenetics to explore the interrelatedness of each mouse cell. We reasoned that analysis of a subset of the most divergent cells would produce the most accurate tree. Therefore, we have presented a Bayesian tree (Fig. 3) constructed from all genotypes (Table 3, which is published as supporting information on the PNAS web site) of all 11 cells from outside the liver and a subset consisting of about half of the most genotypically diverse hepatocytes. The resulting phylogeny divides into two high probability groups (posterior probability = 0.99), one composed entirely of hepatocytes and the other containing cells from the rest of the body along with hepatocytes. It should be noted that the topology of this tree agrees well with that produced by the neighbor-joining approach with all samples. We also constructed a Bayesian tree from all 84 cell samples (data not shown). This tree had a similar overall topology but was less structured in its terminal branches.

Discussion

We have phylogenetically reconstructed cell lineages in cultured mouse NIH 3T3 fibroblasts and in isolated cells from different tissues of an adult mouse on the basis of spontaneously arising mitotic mutations in polyguanine repeat sequences.

We have assembled a proof-of-principal cell fate map from a subset of mouse cells displaying the greatest genetic heterogeneity. The map demonstrates separation of cells into two broad groups, one composed entirely of hepatocytes and the other containing different types of cells and a few hepatocytes. The tree contains a number of polytomies, indicating, not surprisingly, greater complexity to development than can be addressed with the limited number of sampled cells and markers used in this preliminary study. Nevertheless, fine-scale resolution is evidenced by the association of smaller groups of isolates at the terminal branches of the tree and among smaller groups of liver cells radiating from the polytomy into which most hepatocytes group. Most hepatocytes from the same lobe of the liver form high probability clades, reflecting a common heritage for groups of anatomically related cells. Some clades, however, contain hepatocytes from more than one lobe, consistent with observations that embryonic cells giving rise to the liver undergo several cell divisions before the lobes segregate (30). The phylogeny demonstrates significant admixture of hepatocytes from the left and middle lobes, suggesting that those lobes segregate later in development than the right and caudal lobes. However, this observation will have to be validated by further studies. It is not unexpected that some hepatocytes should cluster with other cell types, because stages of mouse development as advanced as late gastrulation (31) are marked by pluripotency and cell migration, to the extent that somatic tissues contain mixtures of ancestral cells (32, 33). It is noteworthy that all nonhepatic cells group together within the same major clade, rather than being scattered randomly

Fig. 3. Cell fate map showing lineage relationships between different types of cells sampled from different anatomic regions of an adult mouse, depicted as an unrooted, rectangular, Bayesian-derived phylogram. Each terminal node is a single cell. Cells taken from the same site are denoted by a dash followed by a number. Cells from the kidney are colored in pink, heart in red, neurons in gray, parotid in yellow, and liver in green. The shading corresponds to laterality and subdivisions within each organ. Confidence values are indicated alongside their corresponding nodes. Scale bar records mitotic divisions along the horizontal axis. R., right; L., left.

throughout the tree, although too few cells were sampled from other tissues to explore their detailed relationships. Overall, the phylogeny is consistent with present models of development and supports the validity of the approach.

The studies with the cultured NIH 3T3 cells indicate that the accuracy of the reconstruction continues to improve with the addition of informative markers. We estimate that ≈ 80 to 280 polyguanine markers would provide a 50% to 90% probability, respectively, of detecting a mutation in a single-cell division (see *Materials and Methods*). The mutation rates that we observed in polyguanine sequences are generally consistent with those measured in other mammalian cells, as determined by assays requiring insertions or deletions in polyguanine tracts to restore ORFs in selectable marker genes (17); however, some markers are more mutable than others, and a fewer number of selected markers might suffice. It may be possible to drive the mutation rate higher by exposing cells or organisms to mutagens or through the use of

strains accelerating somatic mutation, such as mismatch repair deficiency, in which repeats are particularly unstable (34). Although we have used a relatively small number of markers for this study, searches of metazoan genomes (including fruit fly, chicken, mouse, and human) indicate that the frequency of polyguanine tracts is not a limiting factor. The SCOMP method produces enough DNA from a single cell to allow for 510 PCR amplifications, which allows for genotyping ≈ 120 markers with current requirements for internal validation, but an improved statistical understanding of PCR-induced variation may allow accurate interpretation of genotypes based on fewer repeat amplifications.

It may be possible to extract more information from each mutational event. Different alleles of similar length may be more closely related than alleles with greater differences in the length of their polyguanine tracts. The mutations usually consisted of a single-base expansion or contraction, although shifts by two or more bases also occurred. A better understanding of the mutational

mechanism might afford refinement of the evolutionary model with use of “ordered character states.”

There is also room for improvement in the methods of cell sampling before whole genome amplification and genotyping. One approach is laser capture microscopy (35), which would permit the isolation of single cells from tissue sections and thereby allow for greater correlation of tissue architecture with cell lineage. Cell sorting techniques (36) based on surface marker expression might further allow for better assignment of cell type.

While preparing our manuscript, Frumkin *et al.* (37) reported similar work. Still, there are significant differences between the two approaches.

Frumkin *et al.* (37) genotyped short tandem repeat markers, whereas we have exclusively relied on mononucleotide repeat (in particular, polyguanine) markers. Short tandem repeats are comprised of larger repeat units and produce less stutter, making them easier to interpret than polyguanine repeats, but short tandem repeats mutate less frequently. As a result, Frumkin *et al.* (37) needed more markers, as well as mismatch repair-deficient cell lines and plant strains, to detect sufficient somatic diversity. The *MSH2*-deficient strain of *Arabidopsis* used by them accumulates mutations throughout the genome and becomes morphologically abnormal (38), making it a somewhat less suitable model of normal development.

As we have done here with cultured mismatch repair proficient mouse NIH 3T3 cells, Frumkin *et al.* (37) reconstructed a phylogeny of cultured mismatch repair-deficient human adenocarcinoma cells. Their reconstructed trees appear more accurate than ours, but they made no provisions for correlating branch lengths with the number of cell divisions, nor do they consider the possibility of branching within a sampled node. (For example, nodes A1, A5, and A3 in Fig. 1A would be placed at the intersection of the branches, rather than as a separate branch, if the trees were rendered similarly to Frumkin *et al.*; ref. 37). Their cultured trees also had as many as 127 cell divisions between each generation, whereas ours grew for just 22 divisions between each generation. Given that Frumkin *et al.* (37) estimate a “ballpark” figure of ≈ 40 cell divisions between the zygote and a newborn mouse, our *in vitro* results establish the method as tractable within the constraint of the limited number of mitoses occurring during embryogenesis.

Frumkin *et al.* (37) proposed the use of whole genome amplification from single cells but have not yet incorporated that method. Their studies were consequently restricted to the use of bulk plant tissues to obtain sufficient DNA quantities for genotype analysis. The limitation of that method is that a mutation would need to occur early enough developmentally to be present in the majority of the sampled cells, thus limiting the power to map cell fate to earlier points in development. Although they were able to correlate genetic differences with regional anatomy, they did not produce enough information to allow for the phylogenetic reconstruction of the cellular lineages of the plants that they studied.

Nevertheless, our work is complementary to Frumkin *et al.* (37), and we share their optimism that phylogenetic analysis of a limited number of somatically mutable markers ultimately can be scaled to generate retrospectively complete fate maps of unlimited resolution from just a single specimen and that this method will illuminate development of complex multicellular organisms in the same way that the fate map of *C. elegans* has (39). We envision that this approach will prove useful for both broad and focused questions of embryogenesis and may have application in the study of other mitotically dividing cell populations, such as cancer.

Materials and Methods

NIH 3T3 Cell Phylogeny. A single NIH 3T3 (American Type Culture Collection) cell was isolated by limiting dilution (“root”) and clonally expanded for 22 divisions, measuring cell numbers with a hemocytometer. Cell cultures were maintained in DMEM/10% BCS with penicillin (100 units/ml), streptomycin (100 μ g/ml), and

ciprofloxacin (100 μ g/ml) at 37°C in 5% CO₂. After 22 divisions, half of the clonal population was harvested for DNA by using the QIAamp DNA Micro Kit (Qiagen, Valencia, CA). A separate fraction of cells was subjected to limiting dilution to obtain new clonal isolates for subculture, and the process was repeated twice.

NIH 3T3 Cell Genotypes. Oligonucleotide (Operon) sequences for each polyguanine marker are listed (Table 1). Each forward primer contained the indicated 5' fluorescent dye (6-FAM, HEX, or TET). Ten-microliter PCR amplifications with 18 ng of template DNA each were carried out for 40 cycles by using TaqDNA polymerase (Qiagen), and PCR fragments were resolved with an ABI PRISM 310 Genetic Analyzer by using GENESCAN ANALYSIS 3.1.2 and GENOTYPER 2.1 (Applied Biosystems). Most polyguanine markers produced electropherograms exhibiting a normal distribution of smaller peaks surrounding a single peak of maximal intensity, a pattern expected for homozygous alleles. However, some exhibited two adjacent maximal peaks of nearly equal intensity, which could represent heterozygous alleles differing in length by a single base pair. To differentiate between homozygosity and heterozygosity for alleles with one base pair insertions or deletions, seven such markers were selected for further analysis, and for each, DNA from eight NIH 3T3 cell isolates was amplified by using four separate PCRs and genotyped. The difference between the heights of each marker's two tallest peaks was quantified and expressed as a percentage (*D*) of the tallest peak's height. The mean *D* was calculated for each marker/isolate pair and formed a bimodal distribution with a statistically significant difference (two-sided χ^2 , $P = 0.014$) between samples with an average $D \leq 5.4\%$ and samples with $D > 5.4\%$. Genotypes with *D* of $\leq 5.4\%$, therefore, are assigned as heterozygotes with a one-base-pair difference in allele length. There was variability present in *D* among separate amplifications for each marker/isolate pair, with an average deviation from the mean *D* of 1.43 ($\sigma = 0.679$). To determine the origin of this variability, a single PCR was genotyped 16 separate times, and an average deviation from the mean *D* was calculated to be 0.52 ($\sigma = 0.744$). This observation suggests that the differences are largely the consequence of variability introduced by the PCR process. We therefore repeated any single genotype with *D* between 2.61 and 8.19 [$5.4 \pm (\text{average deviation from the mean } D + 2\sigma)$] at least two additional times to better assess the true value of *D* for that marker. Genotypes with $D < 2.61$ or $D > 8.19$ could be confidently ($\approx 97.5\%$) considered to represent isolates exhibiting double (heterozygosity) or single (homozygosity) peaks, respectively.

Evaluation of Whole Genome Amplification Fidelity. NIH 3T3 cells were plated at clonal density, and well isolated single cells were identified microscopically 8 h later. The cells were permitted to divide for 3 days, after which two colonies (one containing six cells, and the other containing four cells) were selected for analysis. None of the cells in either colony appeared mitotic. We added 0.25% trypsin/1 mM EDTA to the culture plate, and by using a micropipette under a microscope, we isolated each cell from the colonies and separately transferred each into a sterile 96-well PCR plate in a volume of 1 μ l of trypsin/EDTA solution. Cells were digested with proteinase K (Roche Diagnostics) and subjected to whole genome amplification by using the SCOMP method as described in ref. 28. To evaluate the fidelity of whole genome amplification at polyguanine tracts, SCOMP products from each of the 10 cells were genotyped at eight different polyguanine loci (markers 115, 119, 148, 162, 166, 184, 194, and 210) six times each. The average *D* value was calculated (as described above) and used to compare loci across isolates from the same colony.

Mouse Cell Isolation. Tissue samples obtained from one 7-month-old female 129X1/SVJ mouse were freshly harvested and stored under liquid nitrogen until use. Samples of ≈ 0.5 mm³ were dissected from the cerebral cortex and cerebellum, the renal cortex, the apex of the

left ventricle of the heart, and the parotid glands. Sampling of the liver was performed by macerating each lobe with a razor blade. To dissociate tissues into single cells, samples were incubated at 37°C in 5% CO₂ in 150 μ l of PBS containing 2.8 Wünsch units/ml Liberase Blendzyme 3 (Roche Applied Science) and 1 mg/ml hyaluronidase (Sigma) for 30 min. After incubation, tissue digests were agitated by gentle pipetting, and single cells were isolated by diluting digests into culture plates containing 0.25% trypsin and 1 mM EDTA solution. Each cell was examined by phase contrast microscopy for appropriate histology. In all, 79 hepatocytes, four renal tubular cells, one cardiac myocyte, six neurons, and two parotid serous cells were collected.

Single-Cell Genotypes. Oligonucleotides (ABI for NED-labeled primers, Operon for all others) are listed in Table 1. Each isolated mouse cell was processed and subjected to the SCOMP method of whole genome amplification (as described above). Five-microliter PCR amplifications containing 0.1 μ l of SCOMP product each were carried out for 42 cycles by using TaqDNA polymerase (Qiagen), and PCR fragments were resolved with an ABI PRISM 3100 Genetic Analyzer by using GENESCAN ANALYSIS 3.7 and GENOTYPER 3.7 (Applied Biosystems). To test the reproducibility of genotyping for all markers, eight amplifications from a single SCOMP amplification (from a hepatocyte) were evaluated for each marker. One marker (154), previously working satisfactorily with NIH 3T3 genomic DNA, yielded inconsistently sized products when using SCOMP-amplified DNA template and, consequently, was excluded. Variability in the *D* value between adjacent maximal peaks (as described above) sometimes is greater for DNA amplified by SCOMP from a single cell than for genomic DNA from NIH 3T3 cells. Therefore, additional reactions were performed as needed (between three and six repetitions per genotype) to better characterize the average *D*.

Phylogenetic Reconstruction. Genotypes for the two alleles of each marker were recorded as the length of the amplicon (in base pairs), with each number occupying one of two positions in a table (Tables 2 and 3). We assumed that the fewest number of mutations possible had occurred and assigned alleles of a particular size into either the first or second position, so as to maximize the number of identical alleles across all isolates. Phylogenies were constructed by using the

Bayesian method as implemented by MRBAYES 3.1.1 (40), setting the morphological evolutionary model with a gamma-distributed rate model. The option of unordered characters was selected, because mutations at mononucleotide tracts can occur in units of greater than one nucleotide (17), rather than in a predictable stepwise fashion. Under the morphological model, MRBAYES employs a symmetric Dirichlet distribution to estimate the frequency of transition between character states. Given that polyguanine tracts appear to persist in the genome over the course of many generations, we used a fixed (infinite) Dirichlet distribution to express an equal probability of expansion or contraction at all loci. The tissue culture reconstruction was run for 4.8×10^6 generations with a sampling frequency of 100 generations and a burn-in value of 2.4×10^3 trees, approximately twice the number required for convergence (split frequency SD < 0.01). Similarly, a reconstruction of the NIH 3T3 phylogeny containing only the initial (root) and terminal isolates (E1→H2) was run for 5×10^6 generations with a burn-in of 2.5×10^3 trees, and the mouse cell phylogeny was run for 5×10^6 generations with a burn-in of 3×10^4 trees, both with a sampling frequency of 100 generations. The consensus trees produced by MRBAYES were exported into TREEVIEW (41) and converted to rectangular phylograms. Phylograms were edited for clarity by using Adobe Systems (San Jose, CA) ILLUSTRATOR 10. The scale of all phylogenies, initially expressed in units of “mutations per site,” was converted into “cell divisions” by (mutations per site of scale bar) $\times$ (number of sites in input file) $\times$ ($\log_2(1/\Sigma$ (mutation per locus in mutations per cell per generation per allele))), where mutation rates below the threshold for detection were taken as zero, and the mutation frequency in the mouse was estimated to be equal to that of the NIH 3T3 cells. Exploratory analysis used the neighbor-joining method as implemented by SPLITSTREE4 (42). The probability of detecting a mutation during a single mitosis was calculated by $(1 - (\text{average mutation per locus in mutations per cell per generation per allele}))^{(\text{number of markers})}$, using the same simplifying assumptions about mutation rate described above.

We thank J. Felsenstein and B. Hall for advice regarding phylogenetics, H.-H. Lee for help with the mouse, D. Ehlert for help with figures, and R. Waterston for use of the ABI 3100 machine. This work was supported by National Institutes of Health (NIH) Grant R01DK58161 (to M.S.H.) and NIH Grant T32GM007266 (to S.J.S.) for the Medical Scientist Training Program.

1. Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. (1983) *Dev. Biol.* **100**, 64–119.
2. Clarke, J. D. & Tickle, C. (1999) *Nat. Cell Biol.* **1**, E103–E109.
3. Stern, C. D. & Fraser, S. E. (2001) *Nat. Cell Biol.* **3**, E216–E218.
4. Bowen, J., Hinchliffe, J. R., Horder, T. J. & Reeve, A. M. (1989) *Anat. Embryol.* **179**, 269–283.
5. Honig, M. G. & Hume, R. I. (1989) *Trends Neurosci.* **12**, 333–335, 340–341.
6. Nesbitt, M. N. & Gartler, S. M. (1971) *Annu. Rev. Genet.* **5**, 143–162.
7. Sanes, J. R., Rubenstein, J. L. & Nicolas, J. F. (1986) *EMBO J.* **5**, 3133–3142.
8. Zernicka-Goetz, M., Pines, J., Ryan, K., Siemerling, K. R., Haseloff, J., Evans, M. J. & Gurdon, J. B. (1996) *Development (Cambridge, U.K.)* **122**, 3719–3724.
9. Sauer, B. (1998) *Methods* **14**, 381–392.
10. Zong, H., Espinosa, J. S., Su, H. H., Muzumdar, M. D. & Luo, L. (2005) *Cell* **121**, 479–492.
11. Matsuoaka, T., Ahlberg, P. E., Kessaris, N., Iannarelli, P., Dennehy, U., Richardson, W. D., McMahon, A. P. & Koentges, G. (2005) *Nature* **436**, 347–355.
12. Eloy-Trinquet, S., Mathis, L. & Nicolas, J. F. (2000) *Curr. Top. Dev. Biol.* **47**, 33–80.
13. Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. (1998) *Genetics* **148**, 1667–1686.
14. Felsenstein, J. (1988) *Annu. Rev. Genet.* **22**, 521–565.
15. Michael, N., Martin, T. E., Nicolae, D., Kim, N., Padjen, K., Zhan, P., Nguyen, H., Pinkert, C. & Storb, U. (2002) *Immunity* **16**, 123–134.
16. Kim, K. M. & Shibata, D. (2004) *BMC Gastroenterol.* **4**, 8.
17. Boyer, J. C., Yamada, N. A., Roques, C. N., Hatch, S. B., Riess, K. & Farber, R. A. (2002) *Hum. Mol. Genet.* **11**, 707–713.
18. Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. & Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31**, 77–84.
19. Weber, J. L. (1990) *Curr. Opin. Biotechnol.* **1**, 166–171.
20. Zhivotovsky, L. A., Rosenberg, N. A. & Feldman, M. W. (2003) *Am. J. Hum. Genet.* **72**, 1171–1186.
21. Diamant, E., Palti, Y., Gur-Arie, R., Cohen, H., Hallerman, E. M. & Kashi, Y. (2004) *Appl. Environ. Microbiol.* **70**, 2464–2473.
22. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402.
23. Nigro, S., Geido, E., Infusini, E., Orecchia, R. & Giaretti, W. (1996) *Int. J. Cancer* **67**, 871–875.
24. Clarke, L. A., Rebelo, C. S., Goncalves, J., Boavida, M. G. & Jordan, P. (2001) *Mol. Pathol.* **54**, 351–353.
25. Hall, B. G. (2005) *Mol. Biol. Evol.* **22**, 792–802.
26. Douady, C. J., Delsuc, F., Boucher, Y., Doolittle, W. F. & Douzery, E. J. (2003) *Mol. Biol. Evol.* **20**, 248–254.
27. Mojica, W. D., Rapkiewicz, A. V., Liotta, L. A. & Espina, V. (2005) *Cancer* **105**, 483–491.
28. Klein, C. A., Schmidt-Kittler, O., Schardt, J. A., Pantel, K., Speicher, M. R. & Riethmuller, G. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 4494–4499.
29. Dickson, P. A., Montgomery, G. W., Henders, A., Campbell, M. J., Martin, N. G. & James, M. R. (2005) *Nucleic Acids Res.* **33**, e119.
30. Wareham, K. A. & Williams, E. D. (1986) *J. Embryol. Exp. Morphol.* **95**, 239–246.
31. Lawson, K. A., Meneses, J. J. & Pedersen, R. A. (1991) *Development (Cambridge, U.K.)* **113**, 891–911.
32. McMahon, A., Fosten, M. & Monk, M. (1983) *J. Embryol. Exp. Morphol.* **74**, 207–220.
33. Soriano, P. & Jaenisch, R. (1986) *Cell* **46**, 19–29.
34. Kim, M., Trinh, B. N., Long, T. I., Oghamian, S. & Laird, P. W. (2004) *Nucleic Acids Res.* **32**, 5742–5749.
35. Dietmaier, W., Hartmann, A., Wallinger, S., Heinmoller, E., Kerner, T., Endl, E., Jauch, K. W., Hofstadter, F. & Ruschoff, J. (1999) *Am. J. Pathol.* **154**, 83–95.
36. Schubert, E. L., Hsu, L., Cousens, L. A., Glogovac, J., Self, S., Reid, B. J., Rabinovitch, P. S. & Porter, P. L. (2002) *Am. J. Pathol.* **160**, 73–79.
37. Frumkin, D., Wasserstrom, A., Kaplan, S., Feige, U. & Shapiro, E. (2005) *PLoS Comput. Biol.* **1**, e50.
38. Hoffman, P. D., Leonard, J. M., Lindberg, G. E., Bollmann, S. R. & Hays, J. B. (2004) *Genes Dev.* **18**, 2676–2685.
39. Sulston, J. E. (2003) *Chembiochem* **4**, 688–696.
40. Ronquist, F. & Huelsenbeck, J. P. (2003) *Bioinformatics* **19**, 1572–1574.
41. Page, R. D. (1996) *Comput. Appl. Biosci.* **12**, 357–358.
42. Huson, D. H. (1998) *Bioinformatics* **14**, 68–73.