A population-epigenetic model to infer site-specific methylation rates from double-stranded DNA methylation patterns

Diane P. Genereux*†‡, Brooks E. Miner†, Carl T. Bergstrom†, and Charles D. Laird†§¶

*Program in Population Biology, Ecology, and Evolution in the Graduate Division of Biological and Biomedical Sciences, Emory University, Atlanta, GA 30322; and Departments of †Biology and §Genome Sciences and ¶Center on Human Development and Disability, University of Washington, Seattle, WA 98195

Communicated by Stanley M. Gartler, University of Washington, Seattle, WA, March 11, 2005 (received for review December 22, 2004)

Cytosine methylation is an epigenetic mechanism in eukaryotes that is often associated with stable transcriptional silencing, such as in X-chromosome inactivation and genomic imprinting. Aberrant methylation patterns occur in several inherited human diseases and in many cancers. To understand how methylated and unmethylated states of cytosine residues are transmitted during DNA replication, we develop a population-epigenetic model of DNA methylation dynamics. The model is informed by our observation that *de novo* methylation can occur on the daughter strand while leaving the opposing cytosine unmethylated, as revealed by the patterns of methylation on the two complementary strands of individual DNA molecules. Under our model, we can infer site-specific rates of both maintenance and *de novo* methylation, values that determine the fidelity of methylation inheritance, from double-stranded methylation data. This approach can be used for populations of cells obtained from individuals without the need for cell culture. We use our method to infer cytosine methylation rates at several sites within the promoter of the human gene *FMR1*.

bisulfite genomic sequencing | epigenetic fidelity | fragile X syndrome | mathematical modeling | population epigenetics

DNA methylation is an important epigenetic mechanism in eukaryotes, where it occurs primarily in the form of 5-methylcytosine (1, 2). Cytosine methylation often is involved in stable transcriptional inactivation, such as in X-chromosome inactivation and genomic imprinting, and sometimes is transmitted through sexual reproduction, producing phenotypic variation (3). In addition, aberrant cytosine methylation patterns are associated with several inherited human diseases, including fragile X (4–6) and ICF (immune deficiency, centrometric heterochromatin, and facial abnormalities) syndromes (7, 8), and many cancers (9). Understanding how methylated and unmethylated states of cytosine residues are preserved through cell division, therefore, will lend crucial insight into numerous biological processes.

CpG/CpG dyads are the principal units of cytosine methylation in vertebrates; these dyads consist of the dinucleotide CpG on one strand and the complementary CpG dinucleotide on the opposing DNA strand. The symmetry of this arrangement provides a means whereby cytosine methylation patterns can be maintained from parent to daughter strands of DNA (10–12). Methyl groups are incorporated into DNA by two types of methylation events. Maintenance methylation occurs when the pattern of methylation on the parent DNA strand serves as the signal for methylation on the newly synthesized daughter DNA strand (13); *de novo* methylation is defined as the addition of methylation at unmethylated dyads that occurs without regard to template pattern (14). These methylation processes are mediated by a class of enzymes known as DNA methyltransferases. Maintenance methyltransferase exhibits a preference for hemimethylated CpG/CpG dyads (methylated on one strand only), is thought to operate principally during the S (DNA replication) phase of the cell cycle, and is localized to the replication fork (15). *De novo* methyltransferases do not exhibit a preference for hemimethylated sites and are thought to be active over a broader temporal and spatial range in mammalian cells (16, 17).

Our fundamental question is this: How do maintenance and *de novo* methylation interact to perpetuate distinct methylation profiles in cellular populations? At loci for which there is no evidence of active demethylation (2, 18), observed methylation frequencies must be determined primarily by the relative rates of maintenance and *de novo* methylation, as previously noted (19–22). Unless maintenance methylation occurs with perfect fidelity, a *de novo* methylation process is an essential component of any system in which methylation frequencies do not decline over repeated rounds of DNA replication (19, 21).

Otto and Walbot (19) and Pfeifer *et al.* (21) used population models to investigate the factors that determine the density of DNA methylation. Both of these useful models were limited by the lack of information on the relative frequencies of methylated, hemimethylated, and unmethylated CpG/CpG dyads. In particular, information was unavailable for the frequency and configuration of hemimethylated dyads that are present, at least transiently, during DNA replication. The existence of this hemimethylated class was revealed by Bird (23) by using methyl-sensitive restriction enzyme analysis; Liang *et al.* (24) improved on this approach to estimate the frequencies of the dyad classes at two CpG sites in cultured cells.

Laird *et al.* (25) used hairpin-bisulfite PCR to determine patterns of these three dyad classes on CpG islands in DNA obtained from human subjects. The observed patterns enable us to resolve an issue that is key to the calculation of site-specific methylation rates and our understanding of the molecular mechanisms of methylation. Does *de novo* methylation occur on the daughter strand, and can *de novo* methylation occur at a single cytosine residue within a CpG/CpG dyad, leaving the opposing complementary cytosine unmethylated? Here, we describe our insights on this issue and use them to build a population-epigenetic model for the dynamics of DNA methylation. We demonstrate the utility of this model by using maximum likelihood to estimate the rates of maintenance and *de novo* methylation at several CpG sites within the *FMR1* promoter on the human inactive X chromosome.

Methods

Hemimethylated CpG/CpG Dyads Can Arise by *de Novo* Methylation on the Daughter Strand, and Perhaps also on the Parent Strand, While the Opposing Cytosine Remains Unmethylated. Although it is reasonable to assume that *de novo* methylation can occur on a single CpG within a CpG/CpG dyad while leaving the complementary cytosine unmethylated, existing evidence in support of this

‡To whom correspondence should be addressed. E-mail: dgenere@emory.edu.

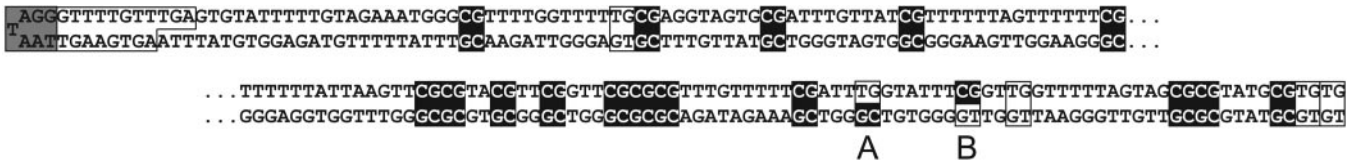

Fig. 1. A double-stranded sequence from the promoter of the hypermethylated *FMR1* of a normal human female recovered by using hairpin-bisulfite PCR. Unmethylated cytosines are converted to uracil during the bisulfite reaction and appear as thymine after PCR amplification. Methylated cytosines are not converted. Unconverted (methylated) CpGs are shown in black, and converted (unmethylated) CpGs are boxed. Among these 22 CpG/CpG dyads, 16 were methylated, 2 were hemimethylated, and 4 were unmethylated. The two sites labeled A and B are hemimethylated sites of opposed polarity, with the methylated cytosine on the top strand at one site and on the bottom strand at the other site. The 26-nucleotide hairpin linker is boxed at the far left, with the randomized 7-nucleotide variable barcode shaded. This double-stranded sequence has been separated into two halves for ease of presentation.

assumption for *in vivo* (26) and *in vitro* (27) systems is limited, and none of this evidence distinguishes between *de novo* methylation on the parent strand and *de novo* methylation on the daughter strand. Here, we use hairpin-bisulfite PCR data obtained from unstimulated somatic cells (leukocytes) to provide direct molecular evidence in support of the conclusion that *de novo* methylation can occur at a CpG on the daughter strand while leaving the CpG on the parent strand unmethylated.

In data from individual alleles at both the hypermethylated *FMR1* and LINE-1 transposable element loci, we observed pairs of hemimethylated dyads within which the methylated cytosines were opposed in polarity (25). For example, the hemimethylated CpG/CpG dyad A in Fig. 1 had a methylated cytosine on the bottom strand, whereas dyad B had a methylated cystosine on the top strand. Regardless of which DNA strand was the parent and which was the daughter in the replication event that gave rise to the double-stranded molecule represented by this sequence, at least one of these two hemimethylated dyads must have been produced by a *de novo* methylation event on the daughter strand, while the complementary cytosine on the parent strand remained unmethylated. The two hemimethylated dyads could not both have been produced by the failure of maintenance methylation.

Hemimethylated dyads occur at significant frequencies in hypermethylated alleles analyzed by hairpin-bisulfite PCR (25). For published sequences, the frequencies of hemimethylated dyads are 6% for the single-copy gene *FMR1* and 12% for LINE-1 transposable element sequences (25). In these sequences, hemimethylated dyads occurred at frequencies that were, on average, >20 times the frequency expected from the failure of bisulfite conversion (25), a technical artifact that would incorrectly indicate the presence of a methyl group on a cytosine that, in fact, lacked one. In addition, sequences often contain two hemimethylated dyads opposed in polarity (25); in our more recent data, such pairs were found to occur in 24 of 110 hypermethylated sequences from CpG sites 1–22 of the hypermethylated *FMR1* promoter in normal human female leukocytes (B.E.M., unpublished data; see also *Note Added in Proof*). Methods to detect hemimethylated dyads were initially validated by using model oligonucleotides (25); recently, the hairpin-bisulfite PCR method used to identify these dyads has been verified through an *in vitro* study of epigenetic processes (28). Observed hemimethylated dyads thus result principally from biological processes, rather than from methodological problems.

In constructing an earlier mathematical model, Otto and Walbot (19) lacked data on the existence and configuration of hemimethylated dyads and assumed that *de novo* methylation of one cytosine within a CpG/CpG dyad immediately leads to methylation of its complementary cytosine. Our inference that *de novo* methylation occurs at single cytosines within CpG/CpG dyads without concomitant methylation at the complementary cytosine allows us to construct a model that estimates the rate of *de novo* methylation in light of the molecular observations described above.

Population-Epigenetic Model for the Dynamics of CpG Methylation. Here, we develop a population-epigenetic model for the dynamics of CpG methylation based on hairpin-bisulfite PCR data on the frequencies of dyad classes, the insights from these data concerning *de novo* methylation (as described above), and the assumption that methylation profiles are at equilibrium in populations of normal adult cells. This assumption is consistent with the finding that individuals' methylation profiles were remarkably similar in blood samples collected 5 years apart (29). At equilibrium, the relative frequencies of methylated, hemimethylated, and unmethylated CpG/CpG dyads at each CpG site are also assumed to be constant over time, although the specific pattern of methylated and unmethylated single CpGs across sites may differ among alleles at a single time point.

The model is designed to track the frequencies of methylated, hemimethylated, and unmethylated CpG/CpG dyads at a given CpG site across a population of cells from a single tissue of a single individual. We assume that the production of hemimethylated CpG/CpG dyads from methylated parent CpGs occurs by failure to transmit the methylated state of the parent strand and that the probability of loss or active removal of the methyl group from an already-methylated CpG is negligible. We also assume that each parent CpG gives rise to a daughter CpG on the newly synthesized complementary strand, neglecting the very small although nonzero rate of mutation at the DNA sequence level.

Let μ represent the probability of maintenance methylation: the net probability that a methylated parent strand enters the DNA replication process and gives rise to a daughter strand that, by any mechanism, becomes methylated at the complementary CpG some time before the beginning of the next round of DNA replication. Because cytosine rather than methylcytosine is incorporated into the daughter strand during DNA replication, each methylated parent strand that enters DNA replication at first produces a new, hemimethylated dyad. When maintenance methylation occurs, a methyl group is added to the cytosine at the complementary CpG site on the daughter strand, producing a methylated dyad consisting of two strands, each of which is methylated at a given CpG site by the time it enters the next replication phase.

Let δ_p represent the probability of CpG *de novo* methylation on the parent strand, which we define as the net probability that an unmethylated CpG on a parent strand that enters one round of DNA replication becomes methylated by any mechanism before its entry into the next round of DNA replication. Let δ_d represent the probability of CpG *de novo* methylation on the daughter strand, which we define as the net probability that a CpG on the daughter of an unmethylated parent becomes methylated by any mechanism before its entry into the next replication phase.

We assume for simplicity, and because existing data do not inform a more specific assumption, that when $\delta_p \neq 0$, the two strands of an unmethylated dyad are independently susceptible to *de novo* methlyation. Hence, the occurrence of a *de novo* event

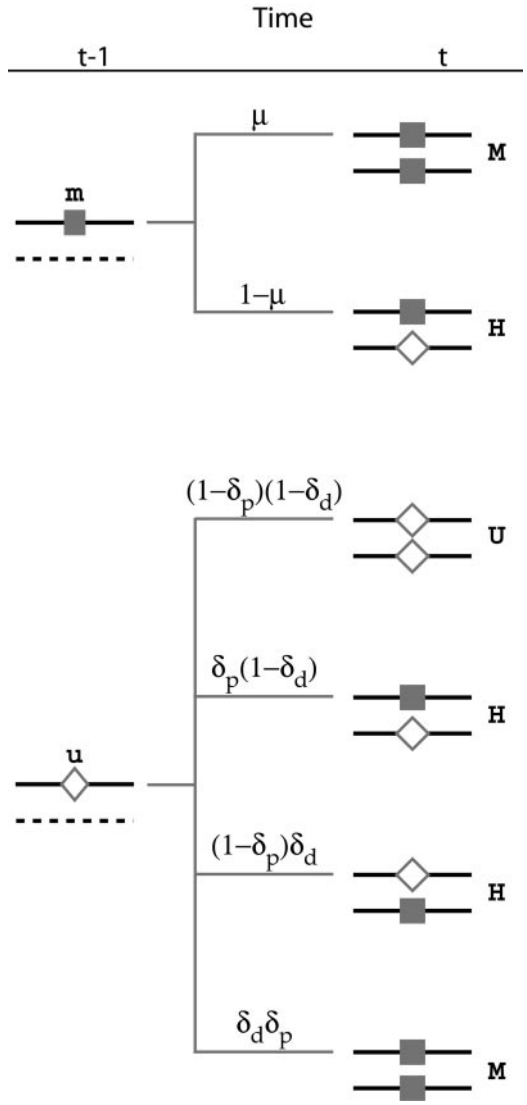

Fig. 2. The transition from methylated and unmethylated single CpGs tallied at time $t - 1$, just before one round of DNA replication, to methylated, hemimethylated, and unmethylated CpG/CpG dyads tallied at time t, just before the next round of DNA replication. The transition probabilities are determined by the probabilities of maintenance (μ) and *de novo* (δ_p and δ_d) methylation events. *De novo* methylation probabilities are considered separately for parent (δ_p) and daughter (δ_d) strands, under the assumption that they are independently susceptible to *de novo* methylation. Unmethylated single CpGs entering the replication process are represented by white diamonds and u, and methylated single CpGs entering the replication process are represented by filled squares and m. Methylated CpG/CpG dyads produced by the replication process are represented by M, hemimethylated dyads are represented by H, and unmethylated dyads are represented by U.

on either the parent or the daughter strand does not alter the probability of a subsequent *de novo* event on the other strand.

The parameters μ, δ_p, and δ_d do not refer to the rates of action of specific maintenance and *de novo* methyltransferases. Rather, μ, $(1 - \delta_p)$, and $(1 - \delta_d)$ represent the net rates at which strands beget like strands, and, thus, they determine equilibrium methylation frequencies. Transition probabilities from single CpGs tallied just before one round of DNA replication to CpG/CpG dyads tallied just before the next round of DNA replication are summarized in Fig. 2.

We denote the frequency of methylated CpG/CpG dyads by M, hemimethylated dyads by H, and unmethylated dyads by U. We denote the frequencies of methylated and unmethylated single CpGs by m and u, respectively.

Our model tallies M, H, and U, and m, and u just before DNA replication (i.e., after existing dyads have undergone maintenance and *de novo* methylation and just before the two strands of the dyads separate to serve as parent strands in the next round of DNA replication). By formulating our model to tally single CpGs and CpG/CpG dyads just before DNA replication, we accommodate *de novo* methylation events that occur at any point during the cell cycle.

The following equations relate the frequencies of single CpGs m and u at time $t - 1$, just before one round of DNA replication, to the observed frequencies of the three types of dyads M, H, and U at time t, just before the next round of DNA replication.

$$\begin{aligned} M_t &= \mu m_{t-1} + \delta_p\delta_d u_{t-1} \\ H_t &= \delta_d(1 - \delta_p)u_{t-1} + \delta_p(1 - \delta_d)u_{t-1} + (1 - \mu)m_{t-1} \\ U_t &= (1 - \delta_p)(1 - \delta_d)u_{t-1}. \end{aligned} \quad [1]$$

The frequencies of methylated and unmethylated single CpGs, m_t and u_t, follow directly from the CpG/CpG dyad frequencies: $m_t = M_t + H_t/2$ and $u_t = U_t + H_t/2$. Substituting these expressions into the relation **1**, we derive a recursion for the CpG/CpG dyad frequencies at time t in terms of the dyad frequencies at time $t - 1$.

$$\begin{aligned} M_t &= \mu\left(M_{t-1} + \frac{H_{t-1}}{2}\right) + \delta_p\delta_d\left(U_{t-1} + \frac{H_{t-1}}{2}\right) \\ H_t &= \delta_p(1 - \delta_d)\left(U_{t-1} + \frac{H_{t-1}}{2}\right) \\ &\quad + \delta_d(1 - \delta_p)\left(U_{t-1} + \frac{H_{t-1}}{2}\right) \\ &\quad + (1 - \mu)\left(M_{t-1} + \frac{H_{t-1}}{2}\right) \\ U_t &= (1 - \delta_p)(1 - \delta_d)\left(U_{t-1} + \frac{H_{t-1}}{2}\right). \end{aligned} \quad [2]$$

At the dyad equilibria, the frequencies of methylated, hemimethylated, and unmethylated dyads do not change with each round of DNA replication ($M_{t+1} = M_t$, $H_{t+1} = H_t$, and $U_{t+1} = U_t$).

Solving recursion **2**, we find the equilibrium dyad frequencies:

$$\begin{aligned} \hat{M} &= \frac{\mu(\delta_p + \delta_d - \delta_p\delta_d) + \delta_p\delta_d}{1 + \delta_p + \delta_d - \mu} \\ \hat{H} &= \frac{2(\delta_p + \delta_d - \delta_p\delta_d)(1 - \mu)}{1 + \delta_p + \delta_d - \mu} \\ \hat{U} &= \frac{(1 - \delta_p)(1 - \delta_d)(1 - \mu)}{1 + \delta_p + \delta_d - \mu}. \end{aligned} \quad [3]$$

The corresponding equilibrium frequencies of the two classes of single CpGs are

$$\begin{aligned} \hat{m} &= \frac{\delta_p + \delta_d}{1 + \delta_p + \delta_d - \mu} \\ \hat{u} &= \frac{1 - \mu}{1 + \delta_p + \delta_d - \mu}. \end{aligned} \quad [4]$$

At equilibrium, the ratio of hemimethlyated to unmethylated dyads is independent of the rate of maintenance methylation, μ,

and is determined strictly by the rates of *de novo* methylation, δ_p and δ_d.

$$\frac{\hat{H}}{\hat{U}} = \frac{2(\delta_p + \delta_d - \delta_p\delta_d)(1 - \mu)}{(1 - \delta_p)(1 - \delta_d)(1 - \mu)} = \frac{2}{(1 - \delta_p)(1 - \delta_d)} - 2. \quad [5]$$

As the rates of *de novo* methylation, δ_p and δ_d, both approach 1, the frequency of unmethylated dyads, *U*, goes to zero, and the ratio of *H* to *U* becomes undefined. This result is not surprising: if *de novo* methylation were perfectly efficient, every unmethylated CpG on a parent molecule would become methlyated by *de novo* methylation and would give rise to a daughter CpG that also became methylated by *de novo* methylation. The absence of unmethylated CpG/CpG dyads from a population in which unmethylated single CpGs were present would lead us to infer a *de novo* methylation rate of 1; all unmethylated single CpGs would be created by failure of maintenance at rate $(1 - \mu)$. In a sample of finite size, the absence of unmethylated dyads could, in principle, result from a *de novo* rate of 1 but could also result from sampling effects. In the next section, we use maximum-likelihood parameter inference to address this and other issues of sample size.

Estimating Rate Parameters and Their Confidence Limits by Maximum Likelihood. Given "true" dyad frequencies of $\hat{M}$, $\hat{H}$, and $\hat{U}$ under a given pair of maintenance and *de novo* rates, the likelihood of observing dyad counts of *M*, *H*, and *U* is given by:

$$\mathcal{L}(M, H, U \mid \hat{M}, \hat{H}, \hat{U}) = \frac{(M + H + U)!}{M!H!U!}\hat{M}(\mu, \delta_p, \delta_d)^M\hat{H}(\mu, \delta_p, \delta_d)^H\hat{U}(\mu, \delta_p, \delta_d)^U.$$

Assuming that dyad frequencies are at equilibrium in populations of normal, differentiated cells, we can use expression **3** to derive an approximate expression for the likelihood as a function of μ, δ_p, and δ_d:

$$\mathcal{L}(M, H, U \mid \mu, \delta_p, \delta_d) = \frac{(M + H + U)!}{M!H!U!}\left(\frac{\mu(\delta_p + \delta_d - \delta_p\delta_d) + \delta_p\delta_d}{1 + \delta_p + \delta_d - \mu}\right)^M \cdot\left(\frac{2(\delta_p + \delta_d - \delta_p\delta_d)(1 - \mu)}{1 + \delta_p + \delta_d - \mu}\right)^H \cdot\left(\frac{(1 - \delta_p)(1 - \delta_d)(1 - \mu)}{1 + \delta_p + \delta_d - \mu}\right)^U.$$

Results and Discussion

We used hairpin-bisulfite PCR (25) to assess the frequencies of the three dyad classes at the *FMR1* locus in peripheral blood leukocytes collected from a normal human female. Experimental conditions were as described in ref. 30; sequencing analysis took place at the Comparative Genomics Center, University of Washington. Each of the recovered alleles was then categorized as either hypermethylated or hypomethylated according to the classification guidelines described in ref. 25. Hypermethylated alleles represent those from the inactive X chromosome. The distinct barcode in the hairpin of each of the 33 sequences verified its independent cellular origin (30). We processed sequence data by using a simple PERL script (available at http://protist.biology.washington.edu/lairdlab/sequenceprocess.htm). Fig. 1 shows an example of one such sequence, with 16 methylated dyads, 2 hemimethylated dyads, and 4 unmethylated dyads.

We examined seven CpG sites within the promoter of the fragile X gene, *FMR1*, in a normal human female (Fig. 3). The

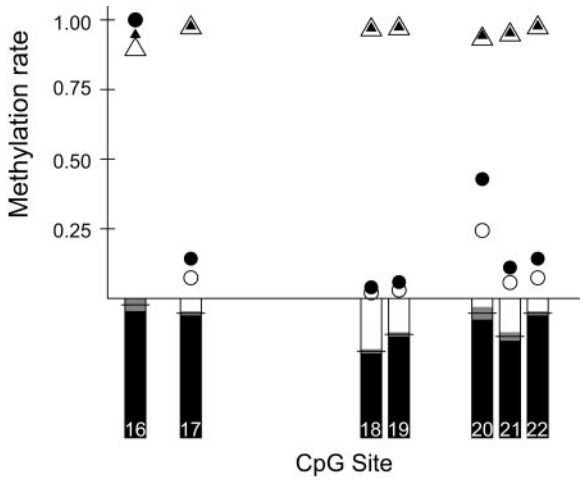

Fig. 3. Observed CpG/CpG dyad proportions and inferred methylation rates for seven CpG sites of the *FMR1* promoter on the inactive X chromosome of a normal human female. (*Lower*) The bar graphs represent the proportions of methylated (black), hemimethylated (gray), and unmethylated (white) CpG/CpG dyads at each site; the horizontal line that divides each bar into two parts illustrates the frequency of methylated (below the line) and unmethylated (above the line) single CpGs at each site. (*Upper*) Maximum-likelihood point estimates for methylation rates. Triangles represent maintenance methylation rates (μ), and circles represent *de novo* methylation rates (δ). Open circles and triangles represent rates inferred under the assumption that parent and daughter strand *de novo* methylation rates are equal ($\delta_p = \delta_d$); filled circles and triangles represent rates inferred under the assumption that *de novo* methylation occurs only on the daughter strand ($\delta_p = 0$). The inferred *de novo* methylation rate for site 16 is 1.0 under both assumptions, because the sampled sequences lacked CpG/CpG dyads of the unmethylated class. The values represented were calculated from the methylation patterns of 33 distinct hypermethylated sequences from different cells, as determined by their molecular barcodes. CpG site numbers correspond to those used in ref. 29.

distributions of the three classes of CpG/CpG dyads at each CpG site displayed intersite heterogeneity ($\chi^2 = 26.67; P < 0.01$). The zero value for unmethylated dyads at site 16 precludes the use of a likelihood ratio (*G*) test; however, our data exceed conservative criteria for the use of the χ^2 test in the case of small expected frequencies (31).

Given the significant heterogeneity among CpG sites, we applied our likelihood method separately to data from each of these seven CpG sites. Maximum-likelihood point estimates of μ, δ_p, and δ_d were calculated under each of two alternative assumptions: (*i*) that the rates of *de novo* methylation were equal for parent and daughter strands ($\delta_p = \delta_d$) and (*ii*) that *de novo* methylation occurred exclusively on the daughter strand ($\delta_p = 0$). A third alternate assumption, that *de novo* methylation occurs exclusively on the parent strand, is excluded by our observation that *de novo* methylation occurs on the daughter strand (see above). We used the likelihood ratio λ to compute approximate 95% confidence regions for the pair of rates for each CpG site (examples shown in Fig. 4). These approximate 95% regions indicate all those μ and δ parameter value pairs with log likelihood values within 3.0 units of the maximum-likelihood estimate (32). The test statistic $-2\log\lambda$ is asymptotically distributed as χ^2_d with *d* degrees of freedom, so our approximate 95% confidence interval covers $-\log\lambda < \chi^2_{.05,2}/2 = 3.0$.

We note five features of the data revealed by the present analysis. (*i*) The sites we analyzed exhibit large variation in the point estimates of the *de novo* methylation rates; under the assumption that $\delta_p = \delta_d$, sites 18 and 20 differ by a factor of 12 in our estimates of their *de novo* methylation rates (0.02 versus 0.24, respectively) (Fig. 3). Some of this variation may arise from

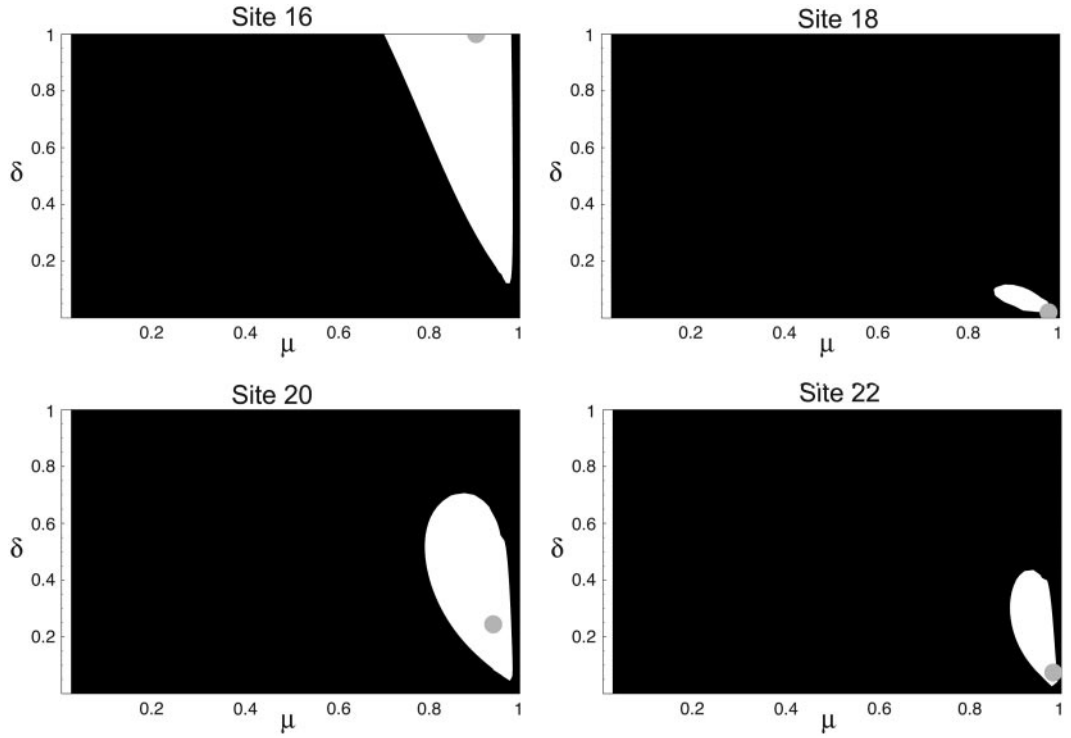

Fig. 4. Examples of maximum-likelihood estimates of site-specific methylation rates and their approximate 95% confidence regions. Point estimates of maintenance and *de novo* methylation rates were inferred by maximum likelihood under the model given in Fig. 2 by using the data on CpG/CpG dyad classes represented in Fig. 3 *Lower*. The estimates shown here are based on the assumption that parent and daughter strand *de novo* methylation rates are equal ($\delta_p = \delta_d$). The approximate 95% confidence regions are shown in white; the point estimates are indicated by gray dots. The four sites for which estimates are shown here represent the range of outcomes from the analysis of the seven CpG sites shown in Fig. 3.

the comparatively small number of sequences used, especially when some dyad classes are few in number; some of it may represent biological differences among sites. (*ii*) Point estimates of the rate of *de novo* methylation for individual sites can differ without comparable differences in the point estimates of the rate of maintenance methylation; under the same assumption that $\delta_p = \delta_d$, point estimates of μ at sites 18 and 20 differed by only 0.03 (0.97 and 0.94, respectively) (Fig. 3). (*iii*) Two sites can have different point estimates of maintenance and *de novo* methylation even when they have similar overall proportions of methylated single CpGs, raising the possibility that they are maintained at equilibrium by different contributions from the two processes (see sites 20 and 22 in Fig. 4). (*iv*) Sites for which the unmethylated class is absent from the data set yield point estimates of δ_p and/or $\delta_d = 1$, with exceptionally large confidence intervals, reflecting the possibility that unmethylated dyads are absent from the data set as a result of sampling effects rather than reflecting biological processes (see site 16 in Figs. 3 and 4). (*v*) Although a modest-size data set like the one analyzed here can provide site-specific rate estimates, larger data sets will typically be required for powerful tests of hypotheses regarding intersite rate variation.

Our estimates under the assumption that *de novo* methylation occurs only on the daughter strand ($\delta_p = 0$) may be compared with those from a previous analysis by Laird *et al.* (25). (E_m as used in the previous study is analogous our μ, and E_d is analogous to our δ_d where $\delta_p = 0$.) Here, we estimate site-specific methylation rates that range from 0.90 to 0.98 for μ and from 0.02 to 1.0 for δ_p and/or δ_d. When considering a larger number of CpG sites in aggregate, Laird *et al.* 2004 (25) arrived at estimates of $\mu = 0.96$ and $\delta = 0.17$. These previous calculations were not written in terms of epigenetic population dynamics and did not provide estimates of site-specific rates. Other researchers have also inferred rates within the broad range of our estimates but without the advantage of direct, multisite information on the frequencies of all three dyad classes (21, 22, 26). In contrast to the method presented here, previous studies also required tissue culture and other experimental manipulation of cellular samples.

Concluding Remarks

We illustrate the powerful synergy between mathematical modeling and a unique molecular method that reveals the cytosine methylation patterns on double-stranded DNA molecules. The specific example we present here, calculation of site-specific rates of *de novo* and maintenance methylation, illustrates the value of methods that can be applied to populations of cells obtained from individuals without need for further manipulation such as growth in tissue culture or molecular labeling. We have raised the possibility of differences in site-specific rates of maintenance and *de novo* methylation and have provided statistical tools with which to test for such differences. Our approach will enable more direct exploration of the fundamental issue of methylation dynamics and the perturbation of these dynamics in normal development and human disease.

Note Added in Proof. Recent and more extensive data on double-stranded methylation patterns in human Line-1 sequences confirm this conclusion (33).

We thank Josh Akey, Alice Burden, Stan Gartler, Scott Hansen, Megan McCloskey, Sally Otto, Nicole Putney, Karen Smith, Reinhard Stöger, and reviewers for thought-provoking discussions and comments on the manuscript. D.P.G., B.E.M., C.T.B., and C.D.L. also thank each other for donning our fUn hats and joining our friends Em and Ed in the productive discussions that followed. This work was

supported by a National Science Foundation Graduate Research Fellowship (to D.P.G.) and National Institutes of Health Grants GM53805 and HD02247 (to C.D.L.). The Comparative Genomics Center in the Department of Biology at the University of Washington is funded in part by Major Research Instrumentation Grant 002236 from the Murdock Foundation.

1. Russo, V., Martienssen, R. & Riggs, A. (1996) *Epigenetic Mechanisms of Gene Regulation* (Cold Spring Harbor Lab. Press, Plainview, NY).
2. Bird, A. (2002) *Genes Dev.* **16,** 6–21.
3. Cubas, P., Vincent, C. & Coen, E. (1999) *Nature* **401,** 157–161.
4. Oberlé, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanuer, A., Boué, J., Bertheas, M. & Mandel, J. (1991) *Science* **24,** 1097–1102.
5. Bell, M., Hirst, M., Nakahori, Y., MacKinnon, R., Roche, A., Flint, T., Jacobs, P., Tommerup, N., Tranebjaerg, L., Froster, I., *et al.* (1991) *Cell* **64,** 861–866.
6. Heitz, D., Rousseau, R., Devys, D., Saccone, S., Abderrahim, H., LePalsier, D., Cohen, D., Vincent, A., Toniolo, D., Della, V., *et al.* (1991) *Science* **251,** 1236–1239.
7. Hansen, R., Wijmenga, C., Luo, P., Stanek, A., Canfield, T., Weemaes, C. & Gartler, S. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 14412–14417.
8. Xu, G., Bestor, T., Bourćhis, D., Hsieh, C., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J. & Vieggas-Pequignot, E. (1999) *Nature* **402,** 187–191.
9. Jones, P. & Baylin, S. (2002) *Nat. Rev. Genet.* **3,** 415–428.
10. Sager, R. & Kitchin, R. (1975) *Science* **189,** 426–433.
11. Holliday, R. & Pugh, J. (1975) *Science* **187,** 226–232.
12. Riggs, A. (1975) *Cytogenet. Cell. Genet.* **14,** 9–25.
13. Bestor, T. (1992) *EMBO J.* **11,** 2611–2617.
14. Okano, M., Bell, D., Haber, D. & Li, E. (1999) *Cell* **99,** 247–257.
15. Leonhardt, H., Page, A., Weier, H. & Bestor, T. (1992) *Cell* **71,** 865–873.
16. Yokochi, T. & Robertson, K. (2002) *J. Biol. Chem.* **277,** 11735–11745.
17. Hsieh, C. (1999) *Mol. Cell. Biol.* **19,** 8211–8218.
18. Bruniquel, D. & Schwartz, R. (2003) *Nat. Immunol.* **4,** 235–240.
19. Otto, S. & Walbot, V. (1990) *Genetics* **124,** 429–437.
20. Riggs, A. & Xiong, Z. (2004) *Proc. Natl. Acad. Sci. USA* **6,** 4–5.
21. Pfeifer, G., Steigerwald, S., Hansen, R., Gartler, S. & Riggs, A. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 8252–8256.
22. Wigler, M., Levy, D. & Perucho, M. (1981) *Cell* **24,** 33–40.
23. Bird, A. (1978) *J. Mol. Biol.* **118,** 49–60.
24. Liang, G., Chan, M., Tomigahara, Y., Tsai, Y., Gonzales, F., Li, E., Laird, P. & Jones, P. (2002) *Mol. Cell. Biol.* **22,** 480–491.
25. Laird, C., Pleasant, N., Clark, A., Sneeden, J., Hassan, K., Manley, N., Vary, J. C., Jr., Morgan, T., Hansen, R. & Stöger, R. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 204–209.
26. Ushijima, T., Wantanabe, N., Okochi, E., Kaneda, A., Sugimura, T. & Miyamoto, K. (2003) *Genome Res.* **13,** 868–874.
27. Lin, I., Han, L., Taghava, A., O'Brien, L. & Hsieh, C. (2002) *Mol. Cell. Biol.* **22,** 704–723.
28. Vilkaitis, G., Suetake, I., Klima-Auskas, S. & Tajima, S. (2005) *J. Biol. Chem.* **280,** 64–72.
29. Stöger, R., Kajimura, T., Brown, W. & Laird, C. (1997) *Hum. Mol. Genet.* **6,** 1791–1801.
30. Miner, B., Stöger, R., Burden, A., Laird, C. & Hansen, R. (2004) *Nucleic Acids Res.* **32,** e135.
31. Roscoe, J. & Byars, J. (1971) *J. Am. Stat. Assoc.* **66,** 755–759.
32. Meeker, W. & Escobar, L. (1995) *Am. Stat.* **49,** 48–53.
33. Burden, A. F., Manley, N. C., Clark, A. D., Gartler, S. M., Laird, C. D. & Hansen, R. S. (Feb. 14, 2005) *J. Biol. Chem.*, 10.1074/jbc.M413836200.