

A Bayesian Phylogenetic Model for Counting Recombination Events

Fang Fang^{1,2}, Vladimir N. Minin⁴, Marc A. Suchard⁴, Karin S. Dorman^{1,2,3}

¹Program in Bioinformatics and Computational Biology,

²Department of Genetics, Development and Cell Biology,

³Department of Statistics, Iowa State University

⁴Department of Biomathematics, University of California, Los Angeles

Abstract:

We describe a Bayesian method for counting recombination events in the evolutionary history of aligned viral sequences when recombination is rare. Previous recombination detection methods have focused on testing whether particular sequences are recombinant. Our method, in contrast, examines multiple recombinant sequences with the same recombinant structure and seeks to place a lower bound on the *number* of recombination events. Distinguishing the number of recombination events in the history of sequences with similar structure will be necessary to find recombination hotspots.

1. Introduction

Genetic recombination generates novel genetic mosaics that may be capable of rapidly adapting to changing environmental conditions (Felsenstein, 1974). Recombination in viruses is common and has been related to several medical concerns, including virulence (Liitsola et al., 1998), drug resistance (Kellam and Larder, 1995), vaccine development (Korber et al., 2001), and probably virus co-species transmission (Rest and Mindell, 2003).

Much effort has focused on understanding the mechanism of virus recombination through experimental approaches. Some have hypothesized that specific primary sequence or RNA secondary structure can promote recombination (Kohli, 1999; Moumen et al., 2003; Galetto et al., 2004). For example, recombination sites cluster within or close to AU-rich regions in brome mosaic bromovirus (Nagy and Bujarsi, 1997) and *in vitro* copy-choice recombination rates are high at several sites along the HIV-1 genome, including the R sequence (Moumen et al., 2001).

It may also be possible to deduce information about the mechanism of recombination by studying naturally occurring recombinant sequences. Although selection and evolution dilute sequence or structure signals that promote recombination, *in vivo* hotspots should be enriched for such signals when they exist. In highly recombinogenic viruses,

it has been possible to identify *in vivo* hotspots. Recombination occurs more frequently near the dimerization initiation sequence (Dykes et al., 2004) or in certain genes (Magiorkinis et al., 2003) of HIV-1. Preliminary evidence suggests a recombination hotspot in the HBV C gene (Fang et al., 2005).

Identification of *in vivo* hotspots relies on the ability to detect historical recombination events. Since recombination is easiest to detect when the recombining parents are genetically divergent, recombinants between different viral subspecies, called *genotypes*, are most often studied. Two recombinant sequences with recombination crossover points (COPs) located at the same position along a genome may be the consequence of a recombination hotspot. When the sequences are recombinants involving different parental genotypes, they are obviously the product of separate recombination events. It is common, however, to observe multiple recombinant sequences with the same COP locations and the same parental genotypes (Sugauchi et al., 2002; Fang et al., 2005). The conservative assumption is that these recombinants descend from a single event that spread by chance or selection (Robertson et al., 2003; Fang et al., 2005). However it is also possible, particularly when hotspots exist, that multiple, similar recombination events produced nearly identical mosaic structures (Fang et al., 2005).

Here, we report a Bayesian phylogenetic-based method for identifying multiple recombination events from two or more putative recombinant sequences with similar recombination structure. Our method is based on a dual multiple change-point (DMCP) recombination detection model (Minin et al., 2005). Although the DMCP model can detect recombinant sequences accurately, it is not designed to identify recombination events. In particular, it is impossible, using the current priors, to place appropriate prior probabilities on single vs. multiple recombination events. Here, we define a novel prior and test our methodology on simulated data. Finally, we analyze six HBV recombinants with similar mosaic structure and reveal at least two unique, underlying recombination events.

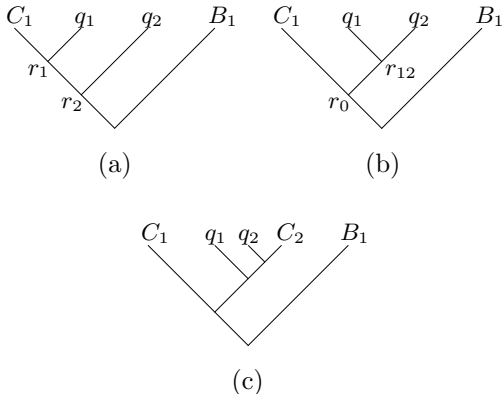


Figure 1: Polyphyletic and monophyletic recombinants. Compared to reference genotype sequences C_1 and B_1 , the recombinant sequences q_1 and q_2 are (a) polyphyletic indicating they were from different recombination events, or (b) monophyletic, indicating either one or two recombination events. (c) In the second case, if another representative sequence C_2 is added, it may be possible to establish that there were two recombination events.

2. Methods

2.1 The phylogeny of recombination events

We examine the phylogenetic relationship of multiple recombinants by aligning them to nonrecombinant genotype sequences representing possible parental genotypes. We borrow the terms *monophyly* and *polyphyly* from general molecular phylogenetics to characterize the resulting phylogenies. In our model, the term monophyly is used to describe trees where all the recombinants form a monophyletic group. If the group of recombinants is disrupted by representative sequences the trees are polyphyletic.

Monophyly is consistent with either single or multiple recombination events, and polyphyly can only imply multiple events. We use an example to demonstrate the argument. Let C_1 and B_1 be representative sequences for genotypes C and B . Suppose q_1 and q_2 are two sequences that share a COP where their genotype switches from C to B . Figure 1 shows the evolutionary relationships between the recombinants and representative sequences for the portion of the genome where both recombinants belong to genotype C . In Figure 1(a), q_1 and q_2 are polyphyletic. It is obvious that two distinct parents produced the recombinants, and thus there were two distinct recombination events. In Figure 1(b) another possible phylogeny is displayed where the two recombinants are monophyletic. Unfortunately,

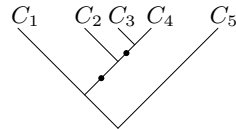


Figure 2: Selecting representative genotype sequences. C_1, \dots, C_5 are genotype C non-recombinant sequences. The filled circles represent the most recent ancestors of two parents involved in separate recombination events.

given the current reference sequences, it is impossible to determine how many recombination events occurred in Figure 1(b).

2.2 Choosing representative sequences

The choice of representative genotype sequences determines whether multiple recombination events can be detected. If representative sequence C_1 is unable to confirm that the recombinants in Figure 1(b) resulted from two events, then another choice of representative C sequence, e.g. C_2 in Figure 1(c), may be able to resolve the events. Thus, including more representative sequences can increase the chance of detecting multiple recombination events when they exist.

Unfortunately, both computational difficulty and statistical uncertainty rise rapidly as more representative sequences are included. In recombination detection, a collection of aligned representative sequences is summarized most commonly by a single consensus sequence consisting of the majority nucleotide at each position. However, the consensus sequence is not a real sequence and since the branching order completely determines monophyly vs. polyphyly, the consensus sequence cannot be used to identify recombination events. Instead, we select a limited number of nonrecombinant, exemplar sequences to represent each genotype.

While the number of exemplars per genotype may be limited by computational burden, exemplars can be selected to increase the chance of detecting multiple events. Consider Figure 2, where the phylogenetic relationships of five exemplars for genotype C are shown. Let two polyphyletic recombinant sequences descend from the ancestors marked by circles on the tree. Polyphyly can be detected with fewer representative sequences as long as at least one descendent of the more ancient of the two ancestors, i.e. at least one of C_2, C_3 , or C_4 , is selected.

Generally the recombinant branch points will not be known prior to the analysis. In this case, a simplistic strategy is to select exemplars that cover as much of the tree structure as possible. Given a topology as in Figure 2, we select one exemplar from each of the two subclades of the root, i.e. select one from $\{C_1, C_2, C_3, C_4\}$ and one from $\{C_5\}$. When recombinant branch points are distributed randomly in the genotype tree, this strategy is the most likely to detect polyphyly. In Figure 2, given sufficient statistical power to resolve phylogenies, this strategy would detect polyphyly 75% of the time.

A second benefit of the strategy is that it can be applied recursively. If, for example C_1 and C_5 are selected, the recombinants would appear as monophyletic, closest neighbors of C_1 . A recursive analysis could now select two sequences, one each from the two clades, $\{C_1\}$ and $\{C_2, C_3, C_4\}$. Assuming the correct phylogeny is inferred, there is now a 100% chance to detect polyphyletic recombinants.

The recursive strategy outlined above relies on a properly rooted genotype tree. We root genotype trees by including one exemplar sequence from another genotype, but it is worth noting that rooting is notoriously error-prone. Other taxon-sampling strategies are possible, but we attempt no thorough analysis of this problem here.

2.3 Choosing recombinant sequences

Before we can test for multiple recombination events, it is necessary to identify a group of recombinants with identical recombinant structure. Any method that can identify recombinants, including the parental genotypes and crossover point locations, can be used to find groups. A method for clustering recombinant sequences based on inferred recombinant structure is described in Fang et al. (2005).

In some cases there may be many more recombinants with the same structure than can be feasibly included in the analysis. If these recombinants result from multiple events, then the descendants of each event should form separate clades in a phylogenetic tree. Thus, to choose recombinants, we align all recombinants, infer a phylogenetic tree with statistical support (e.g. bootstrap support or posterior branch probabilities), and randomly select one recombinant from some or all of the statistically supported clades.

2.4 Multiple recombination events model (MREM)

Following the DMCP model (Minin et al., 2005), MREM assumes the alignment is divided into an

unknown number of independent, non-overlapping segments separated by change points. An unknown number M of these change points represent crossover points where the topology changes because of recombination. Within the same segment, sites are independent and identically distributed.

The goal is to test for multiple recombination events by searching for polyphyly among the recombinant sequences. If any region of the alignment supports polyphyletic recombinants, then we conclude that at least two recombination events have occurred. To quantitate the support for multiple events, we compute the Bayes' factor in favor of polyphyletic recombinants

$$B_{PM} = \frac{p(\text{polyphyletic} \mid D)}{p(\text{monophyletic} \mid D)} \times \frac{q(\text{monophyletic})}{q(\text{polyphyletic})},$$

where $p(\cdot \mid D)$ is the estimated posterior probability, $q(\cdot)$ is the prior probability, and the hypothesis "polyphyletic" means the recombinants are polyphyletic anywhere along the aligned data.

The data D are N representative genotype sequences aligned with R recombinants. Like the DMCP model, MREM, assumes a fixed genotype topology for computational reasons, but allows the R recombinants to branch out anywhere, producing

$$T_{\text{total}} = \frac{(N-2)! [2(N+R)-5]!}{(2N-4)! 2^{R-1} (N+R-3)!}$$

possible topologies for $N > 2$ and $R > 0$. For $R > 2$ there are

$$T_{\text{mono}} = \frac{(2N-3)(2R-5)!}{(R-3)! 2^{R-3}}$$

topologies with monophyletic recombinants. Otherwise, $T_{\text{mono}} = 2N - 3$. These formulae make it clear that polyphyletic phylogenies radically dominate monophyletic phylogenies as N or R increases.

Bayes factors can be sensitive to choice of priors (Kass and Raftery, 1989). To minimize the sensitivity, the prior odds should not stray far from one (Kass and Raftery, 1995; Suchard et al., 2005). In the original DMCP model (Minin et al., 2005), a uniform prior on topologies leads to a high prior probability on polyphyly and multiple events. For example, when $N = 6$ and $R = 2$, there are $T_{\text{total}} = 99$ total topologies possible and $T_{\text{mono}} = 9$ are monophyletic. When there are $M = 3$ crossover points, the prior probability of polyphyly is > 0.9999 , producing prior odds well above $10^4 \gg 1$.

To control the prior probability of polyphyly, we modify the DMCP model prior on topologies $\tau = (\tau_1, \dots, \tau_{M+1})$ with the help of a supplemental vector $G = (G_1, \dots, G_{M+1})$. Let $G_k = 1$ when the

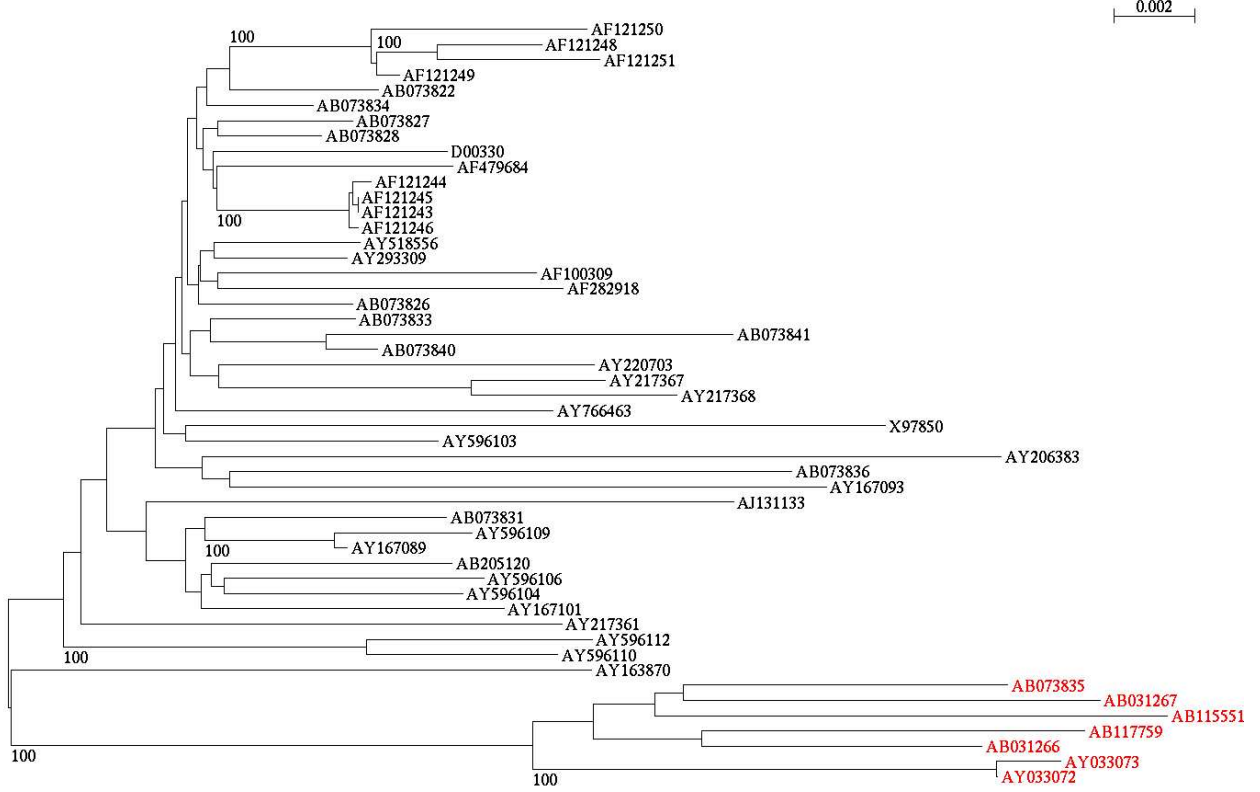


Figure 3: HBV B/C recombinants. Neighbor-joining tree of 50 B/C recombinants with the same recombinant structure. Branches supported by all 100 bootstrap datasets are labeled. The tested branch is the one separating the accessions in red from the accessions in black.

topology τ_k for region k is monophyletic. Otherwise, $G_k = 0$. A priori, we assume segments are independently assigned monophyly status with probability β_M . To control the prior probability of complete monophyly at level a , we set

$$\beta_M = e^{-\frac{\log(a)}{M+1}},$$

for all $M = 0, 1, \dots$. Given G , topologies τ are selected uniformly from the appropriate class (polyphyletic or monophyletic) of trees with the constraint that topologies differ across topology change points. The resulting (G, τ) can be viewed as the realization of a Markov chain with state (G_i, τ_i) in region i . The initial state distribution is uniform

$$p(\tau_1, G_1) = \begin{cases} \frac{\beta_M}{T_{\text{mono}}} & \text{if } G_1 = 1 \\ \frac{1 - \beta_M}{T_{\text{total}} - T_{\text{mono}}} & \text{if } G_1 = 0, \end{cases}$$

over all possible monophyletic or polyphyletic trees.

The transition probabilities are

$$p(\tau_k, G_k | \tau_{k-1}, G_{k-1}) = \begin{cases} \frac{\beta_M}{T_{\text{mono}} - 1} & \text{if } G_k = 1, G_{k-1} = 1 \\ \frac{1 - \beta_M}{T_{\text{total}} - T_{\text{mono}}} & \text{if } G_k = 0, G_{k-1} = 1 \\ \frac{\beta_M}{T_{\text{mono}}} & \text{if } G_k = 1, G_{k-1} = 0 \\ \frac{1 - \beta_M}{T_{\text{total}} - T_{\text{mono}} - 1} & \text{if } G_k = 0, G_{k-1} = 0, \end{cases}$$

which leads to the joint prior

$$\begin{aligned} q(\tau, G | M) &= \frac{\beta_M^{G_1} (1 - \beta_M)^{1 - G_1}}{T_{\text{mono}}^{G_1} (T_{\text{total}} - T_{\text{mono}})^{1 - G_1}} \\ &\times \left(\frac{\beta_M}{T_{\text{mono}} - 1} \right)^{n_{MM}} \times \left(\frac{1 - \beta_M}{T_{\text{total}} - T_{\text{mono}}} \right)^{n_{MP}} \\ &\times \left(\frac{\beta_M}{T_{\text{mono}}} \right)^{n_{PM}} \times \left(\frac{1 - \beta_M}{T_{\text{total}} - T_{\text{mono}} - 1} \right)^{n_{PP}}, \end{aligned}$$

where counting over all $k > 1$,

$$\begin{aligned} n_{MM} &= \# \text{ changepoints with } G_{k-1} = 1, G_k = 1 \\ n_{MP} &= \# \text{ changepoints with } G_{k-1} = 1, G_k = 0 \\ n_{PM} &= \# \text{ changepoints with } G_{k-1} = 0, G_k = 1 \\ n_{PP} &= M - n_{MM} - n_{MP} - n_{PM}. \end{aligned}$$

2.5 Simulation study

Carr et al. (2001) report 16 HIV B/F recombinants sampled from three South American countries. Two of these recombinants (accessions AY037275 and AY037277) have identical recombinant structures and were isolated from transmission-linked sexual partners. It is very likely that these recombinants result from the same recombination event. Analysis with our model and additional phylogenetic tests found the two transmission-linked recombinants are monophyletic throughout the genome.

We extract a 2000 base pair region from the alignment where both recombinants belong to genotype F and use this data set to test the sensitivity of MREM. Artificial polyphyletic data sets are generated by swapping a middle segment of one representative F sequences with one recombinant to produce a small region of polyphyly. The advantage of using these artificial data sets is that the region of polyphyly is known a priori, but the properties of the sequences are otherwise authentically HIV-like. Ten data sets with a simulated polyphyletic segment varying in length from 100 to 180 were tested using MREM.

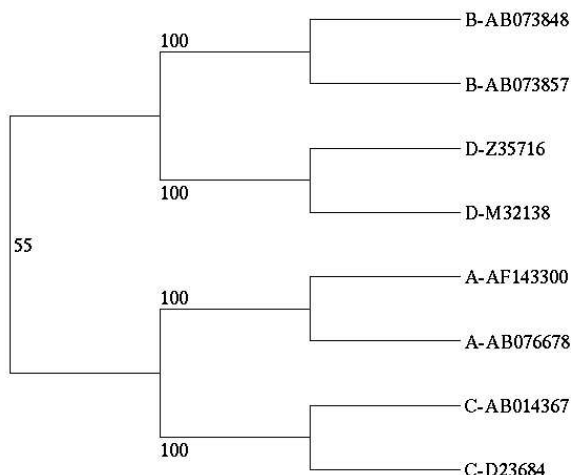


Figure 4: HBV fixed parental tree. Neighbor-joining tree of the six selected non-recombinant HBV genotype sequences. Bootstrap support for each branch is included.

2.6 Analysis of HBV B/C recombinants

Sugauchi et al. (2002) identified 41 HBV recombinants with the same mosaic structure C-B. Fang et al. (2005) not only confirmed the earlier result, but identified 68 additional recombinants with the same structure. A phylogenetic analysis of all these recombinants identifies several clades with 100% bootstrap support. Figure 3 shows a pruned version of the tree including only 50 of the total 109 B/C recombinants for clarity. We select the longest branch with 100% bootstrap support and use MREM to test the hypothesis that this branch defines two, distinct recombination events.

There are too many recombinants to include in a single analysis, so we randomly select four recombinants (AY206380, AB073836, AB073835, AY033073), two from each group. We also select four recombinant sequences (AB073836, AY206380, AY206383, AB073841) from the larger group as a rough control for monophyly. Both data sets were aligned to HBV genotypes *A*, *B*, *C*, and *D*. We chose two representative sequences from each genotype as described in the methods. The fixed parental tree relating these selected representative sequences is shown in Figure 4.

3. Results

3.1 Simulation study

We generate six data sets, five with a simulated polyphyletic regions of varying length 100 to 180 and one fully monophyletic. Table 1 reports the Bayes factor in favor of polyphyletic recombinants, as well as the bias and length of the 95% Bayesian confidence intervals for the estimated COP locations. The significance of Bayes factors are assessed according to the scale presented by Kass and Raftery (1995). As the length of the polyphyletic region increases above 120, there is decisive evidence of polyphyly. The accuracy of crossover point estimation lags behind sensitivity for polyphyly. Confidence intervals for the COP locations did not vary much for the range of simulated data.

3.2 Analysis of HBV B/C recombinants

Recombinants involving genotypes B and C with a crossover point near the end of the S gene have been observed frequently in HBV-infected individuals (Sugauchi et al., 2002). To determine whether these recombinants might have descended from multiple events, we select four recombinants and six representative genotype sequences for MREM analysis.

Segment Length	$\log_{10} B_{PM}$	Significance	COP 1		COP 2	
			Bias	BCI	Bias	BCI
none	-0.98	strong support monophyly	NA	NA	NA	NA
100	-0.25	barely in favor	NA	NA	NA	NA
120	-0.18	barely in favor	NA	NA	NA	NA
140	1.34	strongly support polyphyly	28	123	9	75
160	1.33	strongly support polyphyly	28	123	10	75
180	2.57	decisively support polyphyly	5	126	0	44

Table 1: Simulation results. We report the length of the simulated polyphyletic segment, the log base 10 Bayes factors in favor of polyphyly, and the Kass and Raftery (1995) interpretation of the Bayes factor significance. Columns 4 and 6 give the average posterior bias for the first and second COP, respectively. Columns 5 and 7 give the widths of the 95% Bayesian confidence intervals for the first and second COP, respectively.

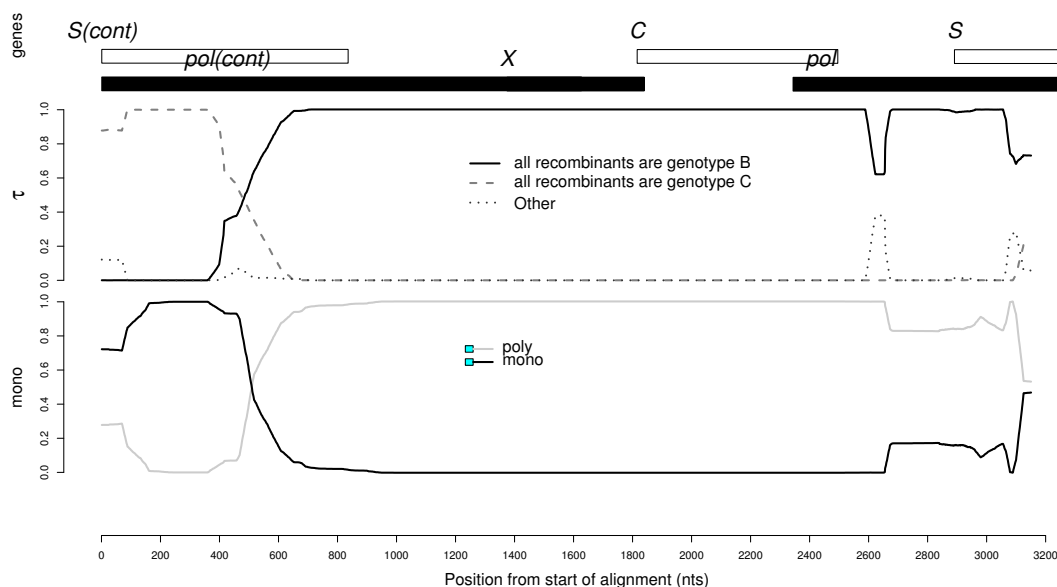


Figure 5: MREM analysis of four HBV B/C recombinants from different subgroups. Gene locations are plotted at the top. In the top plot, we report the combined marginal posterior probability of all topologies assigning genotype B (solid line), genotype C (dashed line), or other genotypes (dotted line) to the four recombinants at each position along the genome. The bottom plot shows the posterior support for polyphyly (gray line) or monophyly (black line) at each genome position.

Figure 5 plots a rough map of the genes along the HBV genome as well as the marginal posterior distribution of τ and the posterior probability of polyphyly or monophyly at each position along the genome. The plot of τ clearly indicates that all four recombinants share the C-B recombinant structure, with a crossover point near position 500. The plot

of polyphyly vs. monophyly indicates that all four recombinants are monophyletic in the 5' C region, but distinctly polyphyletic in the 3' B region. The same data when analyzed with the DMCP model are consistently 100% polyphyletic throughout the alignment, including the 5' region (data not shown).

We also test whether four recombinants selected

from the larger group (upper clade in Figure 3) show any evidence of polyphyly. The \log_{10} Bayes factor $B_{PM} = -0.873$, supporting monophyly. Further analysis with additional recombinants and representative sequences may produce support for polyphyly, but we did not undertake an exhaustive search.

4. Discussion

MREM tests for the possibility that multiple recombination events produced a group of identical recombinant structures. Our simulations show that polyphyletic regions as small as 140 nucleotides are detectable with high support for data derived from HIV-1 sequences. In general, however, sequence diversity and relationships will determine the power to detect monophyly.

We examine six HBV B/C recombinants with identical structures and find strong support for polyphyly. Closer analysis of the phylogeny supported in the polyphyletic region indicates two recombinant clusters, thus at least *two* recombinant events. The appearance of at least two events in the history of these sequences supports the observation that the location of the single crossover point in this recombinant is a possible recombination hotspot (Fang et al., 2005). Assuming these are the only two recombination events in this dataset, it is clear that the descendants of one of these events have been inordinately successful, having been sampled 102 times in the public databases.

MREM controls the prior probability of polyphyly at any arbitrary level. Given a uninformative prior on polyphyly, the method supports monophyly in the genotype C region of the HBV B/C recombinants shown elsewhere to descend from multiple events. Evidently, there is unresolved polyphyly in this region that the selected genotype C representative sequences are unable to detect. Interestingly, the DMCP model, which generously favors polyphyly *a priori*, exclusively supports polyphyly at all positions of the alignment. It is impossible to compute a Bayes factor given this kind of result. Even for cases where the data allow computation of a Bayes factor, DMCP results should be treated with caution. Since the DMCP prior odds for polyphyly are often very large, Bayes factors computed by the DMCP model are expected to be less accurate than those computed by MREM (Kass and Raftery, 1995).

References

- Carr, J. K., Avila, M., Carrillo, M. G., Salomon, H., Hierholzer, J., Watanaveeradej, V., Pando, M. A., Negrete, M., Russell, K. L., Sanchez, J. B., Deborah, L., Andrade, R., Vinales, J., and McCutchan, F. E. (2001), "Diverse BF recombinant have spread widely since the introduction of HIV-1 into South America," *AIDS*, 19, 41–47.
- Dykes, D., Balakrishnam, M., Planelles, V., Zhu, Y., Bambara, R. A., and Demeter, L. M. (2004), "Identification of a preferred region for recombination and mutation in HIV-1 gag," *Virology*, 326, 262–279.
- Fang, F., Suchard, M. A., Minin, V. N., and Dorman, K. S. (2005), "Large-scale recombination detection in HBV full-length sequences," In preparation.
- Felsenstein, J. (1974), "The evolutionary advantage of recombination," *Genetics*, 78, 737–756.
- Galetto, R., Moumen, A., Giacomoni, V., Veron, M., Charneau, P., and Negroni, M. (2004), "The structure of HIV-1 genomic RNA in the gp120 gene determines a recombination hot spot in vivo," *Journal of Biological Chemistry*, 279, 36625–36632.
- Kass, R. E. and Raftery, A. E. (1989), "Comment on "Investigating therapies of potentially great benefit: ECMO, by ware",", *Statistical Science*, 4, 310–317.
- (1995), "Bayes factors," *Journal of American Statistical Association*, 90, 773–795.
- Kellam, P. and Larder, B. A. (1995), "Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance," *Journal of Virology*, 69, 669–674.
- Kohli, A. (1999), "Molecular characterization of transforming plasmid rearrangement in transgenic rice reveals a recombination hotspot in the CaMV promoter and confirms the predominance of microhomology mediated recombination," *The Plant Journal*, 17, 591–601.
- Korber, B., Gaschen, B., Yustm, K., Thakallapally, R., Kesmir, C., and V., D. (2001), "Evolutionary and immunological implications of contemporary HIV-1 variation," *British Medical Bulletin*, 58, 19–42.

- Liitsola, K., Tashkinova, I., Laukkanen, T., Korovina, G., Smolskaja, T., Momot, O., Mashkilleysen, N., Chaplinskias, S., S., B.-K., Vanhatalo, J., Leinikki, P., and Salminen, M. (1998), "HIV-1 genetic subtype A/B recombinant strain causing an explosive epidemic in injecting drug users in Kaliningrad," *AIDS*, 12, 1907–1919.
- Magiorkinis, G., Paraskevis, D., Vandamme, A., Magiorkinis, E., Sypsa, V., and Hatzakis, A. (2003), "In vivo characteristics of Human Immunodeficiency Virus type 1 intersubtype recombination: Determination of hot spots and correlation with sequence similarity," *Journal of General Virology*, 84, 2715–2722.
- Minin, V. N., Dorman, K. S., Fang, F., and Suchard, M. A. (2005), "Dual multiple change-point model leads to more accurate recombination detection," *Bioinformatics*, 21, 3034–3042.
- Moumen, A., Polomack, L., Roques, B., Buc, H., and Negroni, M. (2001), "The HIV-1 repeated sequence R as a robust hot-spot for copy-choice recombination," *Nucleic Acids Research*, 29, 3814–3821.
- Moumen, A., Polomack, L., Unge, T., Veron, M., Buc, H., and Negroni, M. (2003), "Evidence for a mechanism of recombination during reverse transcription dependent on the structure of the acceptor RNA," *Journal of Biological Chemistry*, 278, 15973–15978.
- Nagy, P. D. and Bujarsi, J. J. (1997), "Engineering of homologous recombination hotspots with AU-rich sequences in Brome Mosaic Virus," *Journal of Virology*, 71, 3799–3810.
- Rest, J. S. and Mindell, D. P. (2003), "SARS associated Coronavirus has a recombinant polymerase and Coronaviruses have a history of host-shifting," *Infection, Genetics and Evolution*, 3, 219–225.
- Robertson, D. L., Monumen, A., and Negroni, M. (2003), "HIV recombination breakpoints: Evidence for hot-spots and cold-spots," in *HIV Dynamics and Evolution 10th International Workshop*, Lake Arrowhead, California.
- Suchard, M. A., Weiss, R. E., and Sinsheimer, J. S. (2005), "Models for estimating bayes factors with applications to phylogeny and tests of monophyly," *Biometrics*, 61, 665–673.
- Sugauchi, F., Orito, E., and Mizokami, M. (2002), "Hepatitis B virus of genotype B virus with or without recombination with genotype C over the precore region plus the core gene," *Journal of Virology*, 76, 5985–5992.