Limited Lifespan of Fragile Regions in Mammalian Evolution

Max A. Alekseyev1 and Pavel A. Pevzner2

1 University of South Carolina, Columbia, SC, U.S.A.
2 University of California, San Diego, CA, U.S.A.

Abstract. An important question in genome evolution is whether there exist fragile regions (rearrangement hotspots) where chromosomal rearrangements are happening over and over again. Although nearly all recent studies supported the existence of fragile regions in mammalian genomes, the most comprehensive phylogenomic study of mammals (Ma et al. (2006) Genome Research 16, 1557-1565) raised some doubts about their existence. We demonstrate that fragile regions are subject to a “birth and death” process, implying that fragility has limited evolutionary lifespan. This finding implies that fragile regions migrate to different locations in different mammals, explaining why there exist only a few chromosomal breakpoints shared between different lineages. The birth and death of fragile regions phenomenon reinforces the hypothesis that rearrangements are promoted by matching segmental duplications and suggests putative locations of the currently active fragile regions in the human genome.

1 Introduction

In 1970 Susumu Ohno [35] came up with the Random Breakage Model (RBM) of chromosome evolution, implying that there are no rearrangement hotspots in mammalian genomes. In 1984 Nadeau and Taylor [34] laid the statistical foundations of RBM and demonstrated that it was consistent with the human and mouse chromosomal architectures. In the next two decades, numerous studies with progressively increasing resolution made RBM the de facto theory of chromosome evolution.

RBM was refuted by Pevzner and Tesler, 2003 [38] who suggested the Fragile Breakage Model (FBM) postulating that mammalian genomes are mosaics of fragile and solid regions. In contrast to RBM, FBM postulates that rearrangements are mainly happening in fragile regions forming only a small portion of the mammalian genomes. While the rebuttal of RBM caused a controversy [7, 43, 44], Peng et al., 2006 [36] and Alekseyev and Pevzner, 2007 [2] revealed some flaws in the arguments against FBM. Furthermore, the rebuttal of RBM was followed by many studies supporting FBM [6, 8, 9, 12, 14, 17, 19, 20, 21, 24, 27, 28, 29, 30, 31, 39, 40, 41, 46, 47, 49, 51].

Comparative analysis of the human chromosomes reveals many short adjacent regions corresponding to parts of several mouse chromosomes [16]. While such
a surprising arrangement of synteny blocks points to potential rearrangement hotspots, it remains unclear whether these regions reflect genome rearrangements or duplications/assembly errors/alignment artifacts. Early studies of genomic architectures were unable to distinguish short synteny blocks from artifacts and thus were limited to constructing large synteny blocks. Ma et al., 2006 [25] addressed the challenge of constructing high-resolution synteny blocks via the analysis of multiple genomes. Remarkably, their analysis suggests that there is limited breakpoint reuse, an argument against FBM, that led to a split among researchers studying chromosome evolution and raised a challenge of reconciling these contradictory results. Ma et al., 2006 [25] wrote: “a careful analysis [of the RBM vs FBM controversy] is beyond the scope of this study” leaving the question of interpreting their findings open.

Various models of chromosome evolution imply various statistics and thus can be verified by various tests. For example, RBM implies exponential distribution of the synteny block sizes, consistent with the human-mouse synteny blocks observed in [34]. Pevzner and Tesler, 2003 [38] introduced the “pairwise breakpoint reuse” test and demonstrated that while RBM implies low breakpoint reuse, the human-mouse synteny blocks expose rampant breakpoint reuse. Thus RBM is consistent with the “exponential length distribution” test [34] but inconsistent with the “pairwise breakpoint reuse” test [37]. Both these tests are applied to pairs of genomes, not taking an advantage of multiple genomes that were recently sequenced. Below we introduce the “multispecies breakpoint reuse” test and demonstrate that both RBM and FBM do not pass this test. We further propose the Turnover Fragile Breakage Model (TFBM) that extends FBM and complies with the multispecies breakpoint reuse test.

Technically, findings in [25] (limited breakpoint reuse between different lineages) are not in conflict with findings in [38] (rampant breakpoint reuse in chromosome evolution). Indeed, Ma et al., 2006 [25] only considered inter-reuse between different branches of the phylogenetic tree and did not analyze intra-reuse within individual branches of the tree. TFBM reconciles the recent studies supporting FBM with the Ma et al., 2006 [25] analysis. We demonstrate that data in [25] reveal rampant but elusive breakpoint reuse that cannot be detected via counting repeated breakages between various pairs of branches of the evolutionary tree. TFBM is an extension of FBM that reconciles seemingly contradictory results in [6, 8, 12, 14, 17, 19, 20, 21, 24, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 39, 40, 41, 46, 47, 49, 51] and [25] and explains that they do not contradict to each other. TFBM postulates that fragile regions have a limited lifespan and implies that they can migrate between different genomic locations. The intriguing implication of TFBM is that few regions in a genome are fragile at any given time raising a question of finding the currently active fragile regions in the human genome.

While many authors have discussed the causes of fragility, the question what makes certain regions fragile remains open. Previous studies attributed fragile regions to segmental duplications [5, 18, 42, 50], high repeat density [32], high recombination rate [33], pairs of tRNA genes [10, 23], inhomogeneity of gene distribution [36], and long regulatory regions [17, 30, 36]. Since we observed