

## Minireview

# The where and wherefore of evolutionary breakpoints

David Sankoff

Address: Department of Mathematics and Statistics, University of Ottawa, 585 King Edward Avenue, Ottawa K1N 6N5, Canada. Email: sankoff@uottawa.ca

---

### Abstract

The 'action' in genome-level evolution lies not in the large gene-containing segments that are conserved among related species, but in the breakpoint regions between these segments. Two recent papers in *BMC Genomics* detail the pattern of repetitive elements associated with breakpoints and the epigenetic conditions under which breakage occurs.

---

For many years, dating back to well before the genomics era, there have been numerous observations and hypotheses of associations between the presence or absence of breakpoints of chromosomal evolution and prominent features of the genomic landscape: telomeres, centromeres, recombination hotspots, gene deserts or gene-rich regions, isochores, cytogenetically fragile sites, oncological rearrangements, segmental duplications, transposons and other repetitive elements. Two recent papers in *BMC Genomics* take somewhat different tacks on this subject. Longo *et al.* [1] capitalize on new sequencing resources for the tamar wallaby, *Macropus eugenii*, to substantiate the links between the rapid and complex patterns of evolution of centromeric sequence and recurrent rearrangement activity in marsupials, and to discover one evolutionary breakpoint region in humans that has repetitive element similarity to corresponding regions in marsupials. Lemaitre *et al.* [2] combine a high-resolution breakpoint localization procedure with specialized data that they have calculated or obtained on DNase sensitivity, CG content, hypomethylation and replication origins [3] to dispel some of the most widespread folklore in the field. They show that propensity to breakage is not favored in gene deserts but, on the contrary, is closely related to transcriptional activity and DNA accessibility in a region, a conclusion that lends a decidedly epigenetic flavor to our understanding of rearrangement.

### The ephemeral breakpoint

A breakpoint or breakpoint region is not a tangible physical entity in a genome; it is an analytical construct arising only in the comparison of two genomes and, as such, exists or not, and has one set of characteristics or another, depending on the assumptions and methodology of this comparison. When we can identify two contiguous chromosomal

segments in one genome, each of which seems orthologous to a different segment in another genome, and these latter segments are not contiguous, we can say that there is a breakpoint. When one of the segments is small (according to a threshold of anywhere from  $10^2$  to  $10^6$  base pairs), we might wish to consider the two breakpoints delimiting the segment as reflecting a single breakpoint. If the two segments are actually contiguous in the second genome but one is inverted compared with its orientation in the first genome, we might want to count the breakpoint or not. Normally, the DNA alignment of the two genomes will not be such that the breakpoint can be pinpointed as separating two specific adjacent base pairs, but rather there will be a more or less lengthy region in the middle of the segment on the first genome that does not align well to either of the two segments of the second genome or their flanking sequences. Instead of break 'point', we have a break 'region' with its own particular characteristics [4].

To complete the deconstruction of the breakpoint terminology, we can naïvely imagine the free ends of two or more double-stranded breaks in DNA molecules flailing around inside the nucleus until they are repaired (incorrectly), resulting in a rearrangement within a chromosome or involving two chromosomes. This does indeed happen as a result of radiation, toxic or mechanical stress or, as is clearly demonstrated by Lemaitre *et al.* [2], following normal cellular activity that requires regions of open chromatin. It should be emphasized, however, especially where breakpoints are associated with repetitive elements, rearrangements do not derive from any actual DNA breakage, but from nonhomologous recombination caused by faulty alignment of repetitive elements during meiosis.

The Longo *et al.* article [1] contains a carefully executed and controlled analysis of the distribution of different kinds of repetitive elements in selected segments from three kinds of genomic region in the tamar wallaby: centromeric regions, breakpoint regions (actually three locations in one breakpoint region) and euchromatic regions not containing a breakpoint. They showed a dramatic enrichment in the breakpoint region of sequence characteristic of endogenous retroviruses (ERVs) and LINE1 transposable elements, and a deficiency of SINE

and CR1 transposable element sequences, when compared with the euchromatic regions, with the centromeric regions falling in between the two other patterns. In addition, in a human genomic region containing parts homologous to the marsupial breakpoint region and parts homologous to one of the euchromatin selections, the pattern of repetitive elements makes a transition from ERVs and LINE1s to SINEs. This is suggestive of an association between neocentromeric tendencies, regional instabilities around evolutionary breakpoints and the incorporation of specific kinds of repetitive elements. Although the authors' [1] longstanding interest in marsupial evolution and the role of centromeres in genomic rearrangements, as well as the availability of new sequence resources on the tammar wallaby, are certainly sufficient motivation for the study of repetitive elements in this context, and given the very different patterns known for human and primate pericentromeric evolution, it will now be important to generalize this work to genomes for which sequencing is essentially complete and to undertake a more comprehensive survey of repetitive elements in regions of each kind.

### Reuse and recurrence

The term 'breakpoint reuse' is used in the rearrangements literature to cover two rather different concepts. In its original algorithmic use [5], it denoted the excess of the number of rearrangements necessary to transform one genome into another compared with half the number of breakpoints induced by the comparison of two genomes (given that inversions and reciprocal translocations normally create two breakpoints each). This was accounted for by assuming that some breakpoints (without specifying which ones) were used more than once in the transformation. Soon afterwards, its most frequent meaning became the recurrence of the same breakpoint in two lineages but not their common ancestor with respect to an outgroup lineage [6]. Despite the attractiveness of these concepts to many authors (such as Longo *et al.* [1]), neither breakpoint reuse nor breakpoint recurrence is solidly established as a major evolutionary phenomenon, in contrast to well-known disease-causing somatic cell rearrangements. The original concept of reuse, which did not pertain to particular breakpoints but only their aggregates, has rarely if ever been systematically and quantitatively documented at the level of all the individual breakpoints induced by a pair of genomes. Indeed, the algorithmic results suggesting breakpoint reuse are not only wildly variable depending on how telomeric breakpoints are weighted [7], but are in any case predictable artifacts of highly constrained models of evolution through rearrangement [8] (models that permit no deletion of chromosome segments, no chromosome or chromosomal arm duplication, no segmental duplication, no transpositions, no jumping translocations and no deletion of paralogous syntenic blocks or interleaving deletions of duplicated blocks), and of the levels of resolution used in

defining syntenic blocks and breakpoint regions [9,10]. In the breakpoint definition above, if two breakpoints are collapsed when the small segment between them is below threshold size (a common practice), this mistakenly shows up as an increase in breakpoint reuse. As for the phylogenetic recurrence of breakpoints, the major source in this field [6] actually shows that 80% of the breakpoints in their mammalian phylogeny are not recurrent, and that almost all of the remaining ones affect the syntenically unstable rodent lineage. The tiny proportion of apparently recurrent breakpoints in the rest of the phylogeny would be hard to distinguish from coincidence, given the resolution of the syntenic block construction.

The connection between the 'fragile sites' in traditional cytogenetics and evolutionary breakpoints is exceedingly weak [11] and, indeed, statistically insignificant except through a heuristically contrived categorization of the data. The same may be said for the oft-cited attempt [6] to associate cancer breakpoints with evolutionary breakpoints by selectively comparing only two of the reported frequency categories of neoplastic breakpoints.

### Accident and selection

An evolutionary breakpoint is the product not only of some meiotic accident at a site predisposed to breakage or nonhomologous recombination. It is also a configuration that has managed to do all of the following: make it through steps of abnormal chromosome alignment and segregation to the gamete stage; participate in creating a viable heterokaryotypic zygote that eventually develops into reproductive maturity; endure generations of likely negative selection; and emerge through genetic drift as a homokaryotypic feature of some presumably small bottleneck population. Predisposition to breakage at the cellular level is just the first step on the road to fixation, and phenotypic selection operating at the meiotic, embryonic, adult and population levels has a more important role. Somatic cells presumably have many of the same predispositions to physical breakage, although not of course to nonhomologous recombination, but cancer cells do not have to survive meiosis or life outside the affected individual, and that may be a large part of the reason why the repertoire and quantitative distribution of rearrangements in tumor genomes are very different from those in evolution [12].

Genetic deduction appealing to selection-based arguments at the gene expression level, together with indirect and anecdotal evidence, has recently prompted speculation about prohibition of rearrangement breakage in short intergenic regions in mammals [13]. These claims, however, have effectively been demolished by Lemaitre *et al.* [2], who measured directly and systematically, at a high level of resolution, the connections between both high rate of breakage and short intergenic distances and four strong

correlates of transcriptional activity: GC content, proximity to origins of replication (as inferred from 'N-domains' [3]), hypomethylation (based on CpG ratios) and DNase sensitivity. This innovative and convincing work, to which the authors added support ranging from the classic Bernardi theory of isochores [14] to the more recent mammalian replicon model, overturns the conventional genetic wisdom and reopens evolutionary questions about mechanisms promoting neutral variation at the karyotypic level. It adds a weighty contribution to the accumulating body of results, such as those on the gibbon *Nomascus leucogenys leucogenys* [15] and those previously produced by the O'Neills-Graves collaboration on marsupials, cited in the Longo *et al.* article [1], on the epigenetic conditioning of evolutionary chromosome rearrangement.

### Acknowledgments

This work was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC). DS holds the Canada Research Chair in Mathematical Genomics.

### References

1. Longo MS, Carone DM, NISC Comparative Sequencing Program, Green E, O'Neill MJ, O'Neill RJ: **Distinct retroelement classes define evolutionary breakpoints demarcating sites of evolutionary novelty.** *BMC Genomics* 2009, **10**:334.
2. Lemaitre C, Zaghoul L, Sagot MF, Gautier C, Arneodo A, Tannier E, Audit B: **Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organization.** *BMC Genomics* 2009, **10**:335.
3. Huvet M, Nicolay S, Touchon M, Audit B, d'Aubenton-Carafa Y, Arneodo A, Thermes C: **Human gene organization driven by the coordination of replication and transcription.** *Genome Res* 2007, **17**:1278-1285.
4. Trinh P, McLysaght A, Sankoff D: **Genomic features in the breakpoint regions between syntenic blocks.** *Bioinformatics* 2004, **20**(Suppl 1):i318-i325.
5. Pevzner P, Tesler G: **Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution.** *Proc Natl Acad Sci USA* 2003, **100**:7672-7677.
6. Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, Auvil L, Beever JE, Chowdhary BP, Galibert F, Gatzke L, Hitte C, Meyers SN, Milan D, Ostrander EA, Pape G, Parker HG, Raudsepp T, Rogatcheva MB, Schook LB, Skow LC, Welge M, Womack JE, O'Brien SJ, Pevzner PA, Lewin HA: **Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps.** *Science* 2005, **309**: 613-617.
7. Bergeron A, Mixtacki J, Stoye J: **On computing the breakpoint reuse rate in rearrangement scenarios.** In *Proceedings of Recomb Workshop on Comparative Genomics*. Edited by Nelson CE, Vialette S. Berlin: Springer-Verlag; 2008: 226-240. [Istrail S, Pevzner P, Waterman M (Series Editors): *Lecture Notes in Bioinformatics*, volume 5267.]
8. Sankoff D: **The signal in the genomes.** *PLoS Comput Biol* 2006, **2**:e35.
9. Mazowita M, Haque L, Sankoff D: **Stability of rearrangement measures in the comparison of genome sequences.** *J Comput Biol* 2006, **13**:554-566.
10. Sinha AU, Meller J: **Sensitivity analysis for reversal distance and breakpoint reuse in genome rearrangements.** *Pac Symp Biocomput* 2008, **13**:37-48.
11. Ruiz-Herrera A, Castresana J, Robinson T: **Is mammalian chromosomal evolution driven by regions of genome fragility?** *Genome Biol* 2006, **7**:R115.
12. Ozery-Flato M, Shamir R: **On the frequency of genome rearrangement events in cancer karyotypes.** In *The First Annual RECOMB Satellite Workshop on Computational Cancer Biology*, 2007 [http://www.cs.tau.ac.il/~ozery/cancerGR\_11b.pdf]
13. Peng Q, Pevzner PA, Tesler G: **The fragile breakage versus random breakage models of chromosome evolution.** *PLoS Comput Biol* 2006, **2**:e14.
14. Mouchiroud D, D'Onofrio G, Aïssani B, Macaya G, Gautier C, Bernardi G: **The distribution of genes in the human genome.** *Gene* 1991, **100**:181-187.
15. Carbone L, Harris RA, Vessere GM, Mootnick AR, Humphray S, Rogers J, Kim SK, Wall JD, Martin D, Jurka J, Milosavljevic A, de Jong PJ: **Evolutionary breakpoints in the gibbon suggest association between cytosine methylation and karyotype evolution.** *PLoS Genet* 2009, **5**:e1000538.

---

Published: 24 July 2009  
doi:10.1186/jbiol62  
© 2009 BioMed Central Ltd