



Life cycle of an n-globin pseudogene microsatellite locus

Jasmin H. Bavarva*, Hongseok Tae, Pawel Michalak and Harold R. Garner

Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

*Correspondence: jasmin.spu@gmail.com; jasminhb@vbi.vt.edu

Edited by:

Badri Padhukasahasram, Henry Ford Health System, USA

Reviewed by:

Gyanendra Singh, Kansas State University, USA

Keywords: microsatellites, n-globin, evolution, microsatellite life cycle, gorilla

Microsatellites are composed of tandemly repeated short motifs of 1–6 nucleotides. They are common in eukaryotic genomes; in humans they make up as much as 3% of the genome (Lander et al., 2001). These repeat-containing loci tend to be hyper-variable with variation occurring among individuals of the same species as well as between species. The origin and evolution of microsatellites remain a major puzzle. The birth of a tetranucleotide repeat (ATGT) in the lineage leading to African apes (gorilla, bonobo and chimpanzee) and humans was documented in the n-globin pseudogene (Messier et al., 1996). To test whether the locus is under further expansion or degeneration or if there are any variations as proposed in the life-cycle of the microsatellites (Buschiazzo and Gemmell, 2006), we analyzed positions chr11:5263801:5263831 (hg19) in 82 samples from the 1000 Genomes Project (Genomes Project et al., 2010). To obtain reliable read coverage for the locus, we selected Illumina exome sequencing data and visually inspected the INDEL length in the reads aligned to the locus. We found a repeat deletion (4 bases) of the 4-mer microsatellite locus in 46 samples (56%) relative to reference allele, as supported by at least two reads in each sample. We assume that the longer (reference) microsatellite allele is ancestral, as at least two mutations are reported in this allele (rs34312249 and rs147740082), suggesting a degeneration process of the pure microsatellite locus. The short allele had 28% frequency in analyzed samples.

The genus *Homo* is believed to have evolved 2.0 Myr (Curnoe, 2010) and modern humans 0.2 Myr (Vigilant et al., 1991) ago. Last evolutionary change in the n-globin pseudogene locus is documented

between chimpanzees and humans when the microsatellite locus was expanded in humans. Unfortunately, intraspecies variation in the locus is unknown in chimpanzees and gorillas and we were unable to find the locus in the Neanderthal genome, presumably due to its incompleteness (Green et al., 2006). The locus alteration from orangutan to gorilla took ~7.0 Myr followed by species evolution with no change in the locus (gorilla to chimpanzee). Similarly, the evolution of the chimpanzee to the genus *Homo* took ~3.0 Myr with an accompanying n-globin microsatellite locus expansion. As time required for subsequent stage of life cycle seems to decrease by

about 50% (Birth: ~7 Myr, Expansion: ~3 Myr, Degeneration: 1.8 Myr) (Figure 1A), it appears that the plausible death of the n-globin pseudogene microsatellite locus is expected within the next ~0.9 Myr (Figure 1B). However, human demography and unknown fitness effects of the microsatellite polymorphism can influence the locus evolutionary dynamics.

To test whether the scenario (life cycle of microsatellites) can be applied to other microsatellite loci, we reanalyzed our microsatellite data from multiple species (Galindo et al., 2009). We observed the birth, expansion, and degeneration of several microsatellite loci in four microsatel-

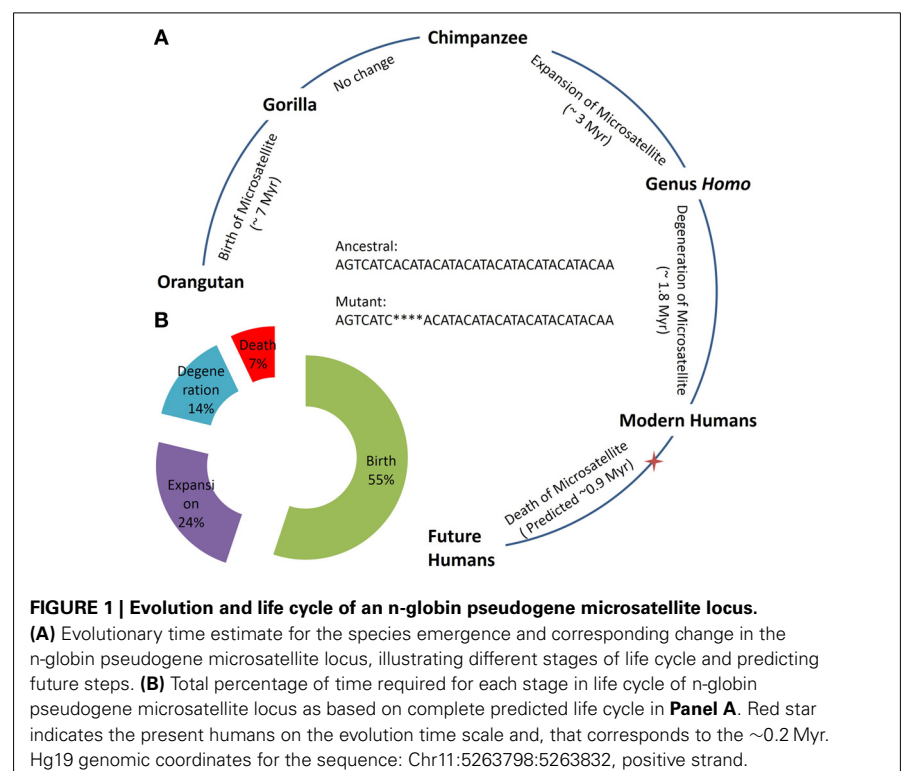


FIGURE 1 | Evolution and life cycle of an n-globin pseudogene microsatellite locus.

(A) Evolutionary time estimate for the species emergence and corresponding change in the n-globin pseudogene microsatellite locus, illustrating different stages of life cycle and predicting future steps. (B) Total percentage of time required for each stage in life cycle of n-globin pseudogene microsatellite locus as based on complete predicted life cycle in Panel A. Red star indicates the present humans on the evolution time scale and, that corresponds to the ~0.2 Myr. Hg19 genomic coordinates for the sequence: Chr11:5263798:5263832, positive strand.

lite motif families in human, chimpanzee, orangutan, rhesus macaque, marmoset and mouse genomes. Although, there were few loci that showed deviations from the expected life-cycle, such as AATGG repeat in *DMD* and *HEPH* gene that showed stability for an extended period of time, and in *PCNXL2* where it showed one repeat diminution in chimpanzee compared to orangutan and human, majority of the repeats arguably followed the proposed life cycle.

Overall, it would be of great scientific significance to constantly monitor a number of marker microsatellite loci for genetic deviations as time progresses and more samples are being sequenced.

ACKNOWLEDGMENTS

This work was supported by the Medical Informatics and Systems Division director's fund at Virginia Bioinformatics Institute.

REFERENCES

- Buschiazzo, E., and Gemmell, N. J. (2006). The rise, fall and renaissance of microsatellites in eukaryotic genomes. *Bioessays* 28, 1040–1050. doi: 10.1002/bies.20470
- Curnoe, D. (2010). A review of early homo in southern africa focusing on cranial, mandibular and dental remains, with the description of a new species (*Homo gautengensis* sp. nov.). *Homo* 61, 151–177. doi: 10.1016/j.jchb.2010.04.002
- Galindo, C. L., Mciver, L. J., McCormick, J. F., Skinner, M. A., Xie, Y., Gelhausen, R. A., et al. (2009). Global microsatellite content distinguishes humans, primates, animals, and plants. *Mol. Biol. Evol.* 26, 2809–2819. doi: 10.1093/molbev/msp192
- Genomes Project, C., Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., et al. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073. doi: 10.1038/nature09534
- Green, R. E., Krause, J., Ptak, S. E., Briggs, A. W., Ronan, M. T., Simons, J. F., et al. (2006). Analysis of one million base pairs of Neanderthal DNA. *Nature* 444, 330–336. doi: 10.1038/nature05336
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062
- Messier, W., Li, S. H., and Stewart, C. B. (1996). The birth of microsatellites. *Nature* 381, 483. doi: 10.1038/381483a0
- Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., and Wilson, A. C. (1991). African populations and the evolution of human mitochondrial DNA. *Science* 253, 1503–1507. doi: 10.1126/science.1840702

Received: 01 October 2013; accepted: 16 November 2013; published online: 04 December 2013.

Citation: Bavarva JH, Tae H, Michalak P and Garner HR (2013) Life cycle of an n-globin pseudogene microsatellite locus. *Front. Genet.* 4:267. doi: 10.3389/fgene.2013.00267

This article was submitted to *Evolutionary and Population Genetics*, a section of the journal *Frontiers in Genetics*.

Copyright © 2013 Bavarva, Tae, Michalak and Garner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.