Modeling the evolution of Alu repeats in human genome

Marek Kimmel, kimmel@rice.edu

Department of Statistics, Rice University, 6100 Main Street, Houston, TX 77005, USA

Keywords: Branching process, infinite alleles model, Alu repeats, evolution

AMS: 60J80

Abstract

Alu elements occupy about eleven percent of the human genome and are still increasing in copy numbers. Since Alu elements substantially impact the shape of our genome, there is a need for modeling the amplification, mutation and selection forces of these elements. Our proposed theoretical neutral model follows a continuous-time branching process described by Pakes [3], or its equivalent discrete counterpart as described by [1]. From the model, we derived a limiting frequency spectrum of the Alu element distribution, which serves as the theoretical, neutral frequency to which real Alu insertion data can be compared to through statistical goodness of fit tests. Departures from the neutral frequency spectrum may indicate different types of selection. A comparison of the Alu sequence data (obtained by courtesy of Dr. Jerzy Jurka) with our model shows that the distribution of Alu sequences in subfamilies AluYa5 and AluYb8 does not follow the expected distribution derived from the branching process; for a thorough biological discussion see Ref. [2]. This observation suggests that Alu sequences do not evolve neutrally and might be under selection.

References

[1] Griffiths, R.C., Pakes, A.G. (1988). *An infinite-alleles version of the simple branching process.* Adv. in Appl. Probab., 20, 489-524.

[2] Kimmel, M., Mathaes, M. (2009). *Modeling the evolution of Alu repeats and testing their neutrality.* Biol. Direct, submitted.

[3] Pakes, A.G. (1989). *An infinite alleles version of the Markov branching process.* J. Austral. Math. Soc. Ser. A, 46, 146-169.