Viral Sampling Design and Inference of Adaptation to the Host

Tae-Kun Seo(1,2), Jeffery L. Thorne(3), Masami Hasegawa(1,2), Hirohisa Kishino(4)

(1) Department of Biosystems Science, The Graduate University for Advanced Studies, Hayama, Kanagawa, 240-0193, Japan; (2) The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu Minato-ku, Tokyo 106-8569, Japan; (3) Program in Statistical Genetics, Statistics Department, Box 7566, North Carolina State University, Raleigh NC 27695-7566 U.S.A.; (4) Laboratory of Biometrics, Graduate School of Agriculture and Life Sciences, University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-8657 Japan

contact : seo@ism.ac.jp; thorne@statgen.ncsu.edu; hasegawa@ism.ac.jp; kishino@wheat.ab.a.u-tokyo.ac.jp

1. Introduction

The process by which viral RNA is copied into DNA by the viral enzyme reverse transcriptase is error-prone and forms the basis for high genetic variability and for a high rate of evolution. In this paper, experimental design techniques for serially sampled data are employed. The question is how to choose the isolation date for the next sample to be sequenced in order to yield a more powerful test of the molecular clock hypothesis(MCH) or in order to improve estimates of evolutionary rates or divergence times. Also, we observed negative correlation between evolutionary rate and effective population size of HIV-1 within nine patients. We expect it gives us a clue to unravel the adaptation process of HIV and the variation in the length of asymptomatic period.

2. Detecting rate variation and inferring population size

[A] Detecting rate variation

(1) Detecting deterministic trend of evolutionary rate

If we assume linearly increasing rate, $r(t) = a(t - t_0) + r$, and Poisson process as evolutionary event, the likelihood function is $\prod_{k=1}^{p} e^{-\bar{r}_k t_k N} \frac{\bar{r}_k \tau_k N}{x_k!}$. Power to reject null hypothesis($H_0 : a = 0$) can be calculated as $P \approx \int_{1+\lambda/1+2\lambda c^2_{0.05}}^{\infty} d\chi^2 (1 + \frac{\lambda^2}{1+2\lambda})$, where $\lambda = -a^2 V_{aa}^{-1}$.

(2) Detecting random fluctuation of evolutionary rate

If we assume that the rate in each branch follows gamma distribution(mean: μ variance: $\delta\mu^2$), the evolutionary process follows negative binomial distribution. Power to reject null hypothesis($H_0 : \delta = 0$) can be calculated as in the previous section with $\lambda = -\delta^2 V_{dd}^{-1}$

(3) Estimating evolutionary rate and time of origin under MCH

The asymptotic variance of m.l.e. can be obtained with Fisher information matrix. As a function of sampling time, the asymptotic variance enable us to find the next optimal sampling time for more accurate inference.

[B] Inferring effective population size

We used coalescent theory to infer effective population size. The coalescent likelihood is $P(t) = \frac{n(n-1)}{2N} \exp\left(-\frac{n(n-1)}{2N} t \right)$ where N is effective population size and n is the number of samples. This is applied to the estimated dates of coalescent events from the serially sampled data. The evolutionary rate is estimated by the maximum likelihood method([1]) and the variance of estimated population sized were obtained by Monte-Carlo simulation.

3. Result and Discussion

To detect linearly increasing evolutionary change, the additional sampling should be done before the oldest sample or after the latest sample. This means that, in general, the more disperse the sampling times, the higher is the power to reject null hypothesis. In detecting gamma-distributed rate, the longer is the additional branch, the higher is the power to reject null hypothesis. Under the assumption of rate constancy, we obtained same result as the case of linearly increase rate. As an example, we analyzed the data set published by Leitner and Albert([2]).

Assuming generation time as 1.2 day([3]), we estimated internal node times and population size with nine patients(data were obtained from [4]). We observed negative correlation.

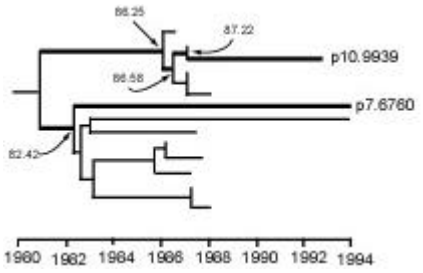

Figure 1 13 sequences from [3]. Additional sequences is added to thick line

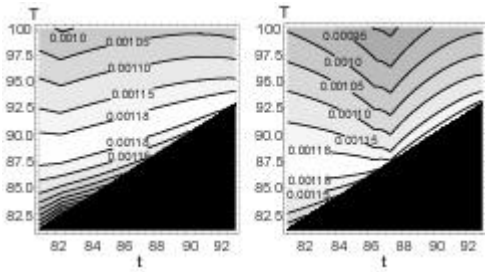

Figure 2 std error of estimate rate assuming additional sample is added to P7.6760(left) or P10.9939

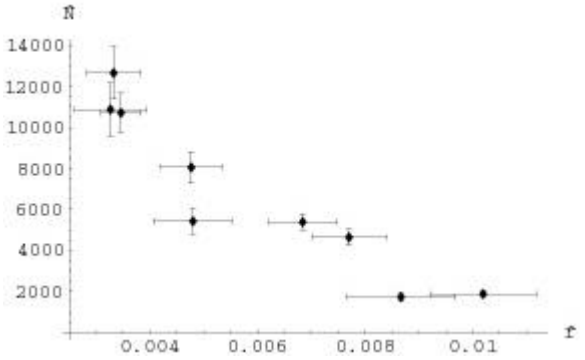

Figure 3 negative correlation between estimated evolutionary rate and effective population size in nine patients

Reference

[1] Rambaut,A. (2000) *Bioinformatics* 16:395-399

[2] Leitner,T. and Albert,J.(1999) *Proc. Natl. Acad. Sci. USA* 96:10752-10757

[3] Rodrigo,A.G. *et al.*(1999) *Proc. Natl. Acad. Sci. USA* 96:2187-2191

[4] Shankarappa,R. *et al.*(1999) *J. Virol.* 73:10489-10502.

Résumé

La puissance d'examiner l'horloge moléculaire est étudiée et la corrélation négative entre le taux évolutionnaire et la taille efficace de population de Hiv-1 a été observée.