Theoretical Population Biology 117 (2017) 43–50

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

Forward and backward evolutionary processes and allele frequency spectrum in a cancer cell population

Hisashi Ohtsuki, Hideki Innan *

SOKENDAI, The Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan

ARTICLE INFO

Article history:
Received 9 May 2017
Available online 1 September 2017

Keywords:
Allele frequency spectrum
Cancer
Population genetics
Branching theory
Coalescent theory

ABSTRACT

A cancer grows from a single cell, thereby constituting a large cell population. In this work, we are interested in how mutations accumulate in a cancer cell population. We provide a theoretical framework of the stochastic process in a cancer cell population and obtain near exact expressions of allele frequency spectrum or AFS (only continuous approximation is involved) from both forward and backward treatments under a simple setting; all cells undergo cell divisions and die at constant rates, b and d, respectively, such that the entire population grows exponentially. This setting means that once a parental cancer cell is established, in the following growth phase, all mutations are assumed to have no effect on b or d (i.e., neutral or passengers). Our theoretical results show that the difference from organismal population genetics is mainly in the coalescent time scale, and the mutation rate is defined per cell division, not per time unit (e.g., generation). Except for these two factors, the basic logic is very similar between organismal and cancer population genetics, indicating that a number of well established theories of organismal population genetics could be translated to cancer population genetics with simple modifications.

1. Introduction

A tumor grows from a single cell, as has been well recognized for several decades (Muller, 1950; Nowell, 1976; Fidler, 1978; Dexter et al., 1978; Merlo et al., 2006). Through the growth process, cells accumulate various kinds of mutations, from simple point mutations to more drastic changes at the chromosomal level, such as deletions and amplifications (Sjöblom et al., 2006; Wood et al., 2007; The Cancer Genome Atlas Research Network, 2008, 2012, 2014; Garraway and Lander, 2013; Vogelstein et al., 2013). There are two major categories of mutations in cancer cells, driver and passenger mutations. The former are generally cell autonomous, that is, they increase the reproductive ability of the carrier cell (i.e., adaptive), while the latter have no effect on the reproductive ability (i.e., neutral). A new technology for genome sequencing from a single cell opened a new window in cancer genetics, because sequencing a number of cells from a single tumor makes it possible to identify heterogeneity in the catalog of driver and passenger mutations between cells, from which we are able to infer when and how the tumor has grown (Navin, 2015). Even without such desirable data available, the frequencies of mutations in bulk-sequencing data are informative to infer the history of a tumor (Williams et al., 2016).

Population genetics provides a solid theoretical framework for a wide variety of such inference methods (e.g., Nielsen and Slatkin, 2013; Wakeley, 2009). The coalescent (Kingman, 1982; Hudson, 1983; Tajima, 1983) plays the central role in providing theoretical predictions of the pattern of genetic variation, which can be used to compute the likelihood of the observed variation data (Donnelly, 1996; Tavaré et al., 1997). It concerns the history of the sampled individuals, by tracing their ancestral lineages up to the MRCA, most recent common ancestor (e.g., Nielsen and Slatkin, 2013; Wakeley, 2009).

One might think that the coalescent theory can be directly applied to cancer cells due to the obvious analogy; all cancer cells should follow a simple genealogy up to their MRCA. However, the direct application of the standard population genetics (i.e., organismal population genetics) to a cancer cell population may not be exactly correct because of some fundamental differences in the propagation system, as we explain below (see also Sidow and Spies, 2015).

In organismal population genetics, the process can be specified by the expected number of offsprings for each individual, namely, the fitness (e.g., Crow and Kimura, 1970; Ewens, 1979). In the Wright–Fisher model with N haploids (Fisher, 1930; Wright, 1931), all individuals are randomly replaced every generation, and individuals with higher fitness likely produce more offsprings. In the Moran model (Moran, 1962), individuals are replaced one by one, that is, one step consists of a coupling event of birth and death; one dead individual is replaced by the offspring of one randomly

* Correspondence to: Graduate University for Advanced Studies, Hayama, Kanagawa 240-0193, Japan.
E-mail address: innan_hideki@soken.ac.jp (H. Innan).

http://dx.doi.org/10.1016/j.tpb.2017.08.006
0040-5809/

chosen individual from the population allowing self-replacement. Consequently, all individuals are on average replaced in N steps, which roughly correspond to one generation in the Wright–Fisher model. It has been well known that theoretical results under the two models are nearly identical in various cases (e.g., Crow and Kimura, 1970; Ewens, 1979; Wakeley, 2009; Bhaskar and Song, 2009). Through this random reproduction process either in the Wright–Fisher or Moran model, mutations that arise in the population will fix or get extinct by the joint action of random genetic drift and selection. A mutation is defined as adaptive when it increases the fitness of the carrier individual.

The evolutionary process of a cancer cell population does not follow such a simple replacement system. Fig. 1 illustrates the process from cancer initiation, progression to the following rapid growth, which may be roughly divided into two major phases, and the applicability of organismal population genetics may differ depending on the phase. The first phase (Phase I) from cancer initiation to initial progression could be well handled under the organismal population genetic framework (Komarova et al., 2003; Iwasa et al., 2004; Michor et al., 2004). This phase is commonly modeled in a constant-size population of cells. Most theoretical models for cancer initiation suppose that a tissue consists of a number of small compartments of cells and that cancer initiation can occur in a compartment. The system starts with a normal compartment with a certain number of asexually reproducing normal cells, which is denoted by N_0. N_0 is usually assumed to be constant because the number of cells in a healthy tissue is maintained roughly constant by homeostatic systems, that is, cell division occurs when needed. The Moran model is more suitable to apply to this process than the Wright–Fisher model because it can be modeled such that one cell death asks for one cell division. Indeed, the Moran model has been frequently used to explore a number of problems on cancer initiation (reviewed in Michor et al., 2004). One of the major problems is how a cancer initiates. A compartment of a normal tissue could become a cancer when oncogenes are activated and/or tumor-suppressor genes (TSGs) are inactivated. It is believed that at least several mutational alternations in cancer genes (oncogenes and TSGs) are required for the formation of a parental cancer cell. Such accumulation of mutations in cancer genes could allow a cell to acquire typical behaviors of cancer cells, for example, avoiding apoptosis (programmed cell death) that makes it difficult to maintain the equilibrium between birth and death in the compartment, thereby shifting towards uncontrolled proliferation (neoplasia). There are a large body of theory only for the fixation process of mutations in cancer genes, especially for the inactivation of TSGs, perhaps because the problem is mathematically too simple for the activation of oncogenes (Michor et al., 2004). Inactivation of a TSG involves the fixation of a double-mutant, that is, both alleles have to be silenced according to Knudson's two-hit model (Knudson, 1971). This situation is very similar to the fixation process of a pair of compensatory mutations in organismal population genetics (Innan and Stephan, 2001), and the results are indeed in good agreement (Iwasa et al., 2004). Thus, it can be considered that the applicability of organismal population genetics is quite good in Phase I because the assumption of a constant-size population roughly holds so that the stochastic process through random genetic drift works as organismal population genetics predicts.

By contrast, in the second phase (Phase II) where cells have acquired extraordinary high proliferative ability, the population grows very rapidly, and the stochastic process is less important for changing allele frequencies because most cells have very low death rates by avoiding apoptosis and their cell divisions occur independently of each other. As a consequence, a fixation of adaptive mutation hardly occurs in a cancer cell population because the spread of an adaptive mutation does not necessarily kill other cells with lower reproductive rates, as has been pointed out by Sidow and Spies (2015). This reproducing system is quite different from that organismal population genetics supposes.

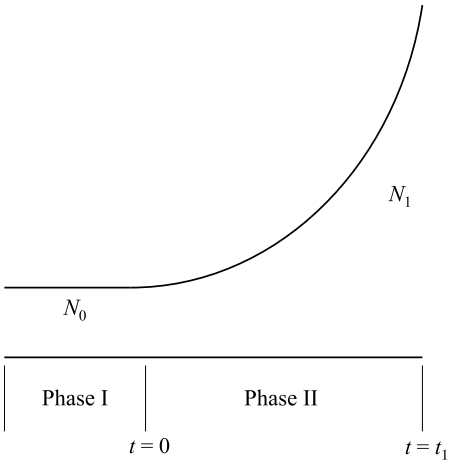

Fig. 1. Illustrating the model of the growth of a cancer cell population.

The behavior of mutations in an exponentially growing has been well studied since Luria and Delbrück (1943) who investigated the evolutionary process of resistance mutations in a bacterial population (see also Kessler and Levine, 2013). The model handles neutral mutations in an exponentially growing population, which will confer selective advantages after an environmental change (e.g., viral infection). Models with stochastic processes taken into account have been explored by Kessler and Levine (2013) and Antal and Krapivsky (2011). The Luria–Delbrück model thus provides the basis for exploring the behavior of driver and passenger mutations in a cancer cell population (e.g., Kansal et al., 2000; Haeno et al., 2007; Antal and Krapivsky, 2011; Durrett et al., 2011; Foo and Leder, 2013; Bozic et al., 2016). Most of these works focus on the number of mutations per cell, the evolutionary "waves" of driver mutations or more complicated tree structure (but see Durrett, 2013, 2015), which may not be straightforward to apply cancer genomic data, especially when single cell-based sequences are not available.

To be more applicable to recent cancer genomic data, we here ask how the well established theory of organismal population genetics can be applied to Phase II assuming an exponential growth. In particular, we are interested in the allele frequency spectrum (AFS, or SFS: site frequency spectrum) of passenger mutations in a cancer cell population. AFS is the summarized information of genotype data that are frequently used in organismal population genetics. Under the basic neutral theory of the coalescent for a constant size population (Kingman, 1982; Hudson, 1983; Tajima, 1983) with the assumption of infinitely many sites (Kimura, 1969), the expected AFS can be described in a simple form (Fu, 1995), but for a non-constant size population, it is not very straightforward to obtain the expected AFS in a simple closed form. Even with any complicated demographic setting, the expected AFS can be written as a function of the expectations of coalescent times (Griffiths and Tavaré, 1994, 1998), but these expectations are not easy to derive in a simple form in many cases although possible computationally (Williamson et al., 2005; Polanski and Kimmel, 2003; Polanski et al., 2003). AFS provides substantial information on the past demography, making it possible to infer various demographic parameters including population size changes and migration rates (Nielsen, 2000; Adams and Hudson, 2004; Williamson et al., 2005; Gutenkunst et al., 2009; Bhaskar et al., 2015; Gao and Keinan, 2016).

In this article, we consider a model of a rapidly growing cancer cell population for exploring how mutations accumulate within