Rodrigo *BMC Res Notes (2017) 10:393*
DOI 10.1186/s13104-017-2741-3

BMC Research Notes

RESEARCH NOTE

Open Access

Evolutionary impact of copy number variation rates

Guillermo Rodrigo*

Abstract

Objective: Copy number variation is now recognized as one of the major sources of genetic variation among individuals in natural populations of any species. However, the relevance of these unexpected observations goes beyond diagnosing high diversity.

Results: Here, it is argued that the molecular rates of copy number variation, mainly the deletion rate upon variation, determine the evolutionary road of the genome regarding size. Genetic drift will govern this process only if the effective population size is lower than the inverse of the deletion rate. Otherwise, natural selection will do.

Keywords: Birth–death process, Gene duplication, Genome size, Neutral evolution, Population genetics

Introduction

The advent of genomic systems biology is leading, in very recent years, to the discovery of widespread genetic features, previously unrecognized in complex organisms like humans; they can then have strong implications in biomedicine. One of these fascinating features is copy number variation [1], which is already considered one of the major sources of genetic variation. Thereby, in natural populations, some individuals have significant portions of the genome repeated, even entire genes, something until now believed to occur at a large scale only in microbes [2]. Recently, genomic studies with populations of different model organisms are serving to estimate copy number variation rates [1, 3], revealing great differences among them. However, the generation of genetic diversity, here genome rearrangements, cannot be fully understood without accounting for an evolutionary perspective [4]. In this regard, what is the impact of these rates?

In this short piece, it is argued that these rates greatly determine the way by which the genome can increase its size, i.e., the evolutionary force that controls this process. Indeed, the acquisition of genetic redundancy is believed to be the major mechanism to increase genome size, and then genome complexity [5, 6]. For that, duplications have to be fixed in the population, a process that mainly occurs, according to the classical theory, by random genetic drift under effectively neutral selective conditions thanks to a reduction in effective population size [6]. However, the balance between duplication and deletion, at a given locus, regulates the power of drift in fixation, as it is illustrated here with a simple quantitative analysis.

Main text

Theory

The complex process of copy number variation in a single organism, which involves widely different mechanisms [7], can be simplified to a birth–death process. This allows creating a toy model from which to make predictions (see "Limitations"). If μ denotes the duplication rate of a locus and λ the deletion rate upon duplication, the frequency of the genotype with two copies (x) in a population of size N is governed by the following stochastic differential equation

$$\frac{dx}{dt} = \mu - (\mu + \lambda)x + \sqrt{\frac{x(1-x)}{N}}z(t), \qquad (1)$$

where t is measured in generations, and $z(t)$ is a stochastic process with mean zero and correlation delta, having assumed a Wright-Fisher reproduction model and

*Correspondence: guillermo.rodrigo@csic.es
Institute for Integrative Systems Biology, CSIC-UV, 46980 Paterna, Spain

strictly neutral selective conditions [8]. This means that the stationary solution will be $x = \mu/(\mu + \lambda)$, and that the eventual fixation of genetic variants will only occur transiently. This is an important consideration, implying that duplicates will be preserved for long time if they quickly accumulate, upon fixation, beneficial [9] or complementary, degenerative mutations [10], escaping from the birth–death process.

The system has two different time scales, one given by $1/(\mu + \lambda)$, associated with copy number variation, the other by N, associated with genetic drift. Certainly, if $(\mu + \lambda)N \ll 1$, the system can be assumed dominated by genetic drift at short times. Accordingly, the typical fluctuation amplitude in frequency (Δx, around the stationary solution) can follow the Einstein's theory of Brownian particles [11]. A fixation time of $t \approx 6N$ can be derived if we integrate over $[0, 1]$ the variance of the stochastic process, $x(1 - x)/N$, to have constant diffusion, in good tune with the Kimura's calculation [8]. But, in general, we have

$$\Delta x = \frac{1}{\sqrt{12(\mu + \lambda)N}}. \quad (2)$$

Fixation will occur in displacements that reach $x = 1$ [i.e., $\Delta x = \lambda/(\mu + \lambda)$], which yields the condition of $\lambda N < 1/12$, when $\lambda \gg \mu$ (typically in nature) [1]. Fluctuations can even be three times the typical value, although they will occur sporadically. This yields the soft condition of $\lambda N < 1$ to have chances for fixation. By contrast, if $(\mu + \lambda)N \gg 1$, the system is mostly dominated by the balance between duplication and deletion. Therefore, $\Delta x \ll 1$, which entails that duplications cannot be fixed.

Remark

The deletion rate that was considered here is the rate at which a repeated portion of the genome is deleted. Certainly, duplication imposes a genetic instability that is generally resolved by deletion [12]; sometimes by other means, like relocation [13]. Experimentally, such a deletion rate needs to be estimated from populations with individuals carrying duplications. The deletion rate of significant, but unique fragments is expected to be only a lower bound. Despite, this has already been proposed as a determinant of genome size [14].

Application

This simple theory can be applied to analyze the fixation ability in different organisms (Fig. 1). For *Salmonella enterica*, $\mu \approx 10^{-4}$/locus/gen. and $\lambda \approx 2{\cdot}10^{-2}$/locus/gen. [12], with $N \approx 10^{8}$. In this case, $(\mu + \lambda)N \approx \lambda N \approx 2{\cdot}10^{6} \gg 1$, which entails that this bacterium cannot acquire genetic redundancy, at least by drift. Similar is the case for the lower eukaryote *Saccharomyces cerevisiae*, where $\mu \approx 3{\cdot}10^{-6}$/locus/gen. and $\lambda \approx 2{\cdot}10^{-6}$/locus/gen. [3], with $N \approx 10^{7}$, give $(\mu + \lambda)N \approx 50 \gg 1$. However, the scenario is different in higher eukaryotes. For *Drosophila melanogaster*, $\mu \approx 2{\cdot}10^{-7}$/locus/gen. and $\lambda \approx 10^{-6}$/locus/gen. [15], with $N \approx 10^{6}$. This results in $(\mu + \lambda)N \approx \lambda N \approx 1$, the soft limit, suggesting that transient fixation of duplications could occur. Better is the case for *Caenorhabditis elegans*, as $\mu \approx 10^{-7}$/locus/gen. and $\lambda \approx 2{\cdot}10^{-7}$/locus/gen. [16], with $N \approx 10^{5}$, lead to $(\mu + \lambda)N \approx 0.03 \ll 1$. For *Homo sapiens*, $\mu + \lambda \approx 10^{-6}$/locus/gen. [1], with $N \approx 10^{4}$, ensures many momentary fixations by drift, as $(\mu + \lambda)N \approx 0.01 \ll 1$.

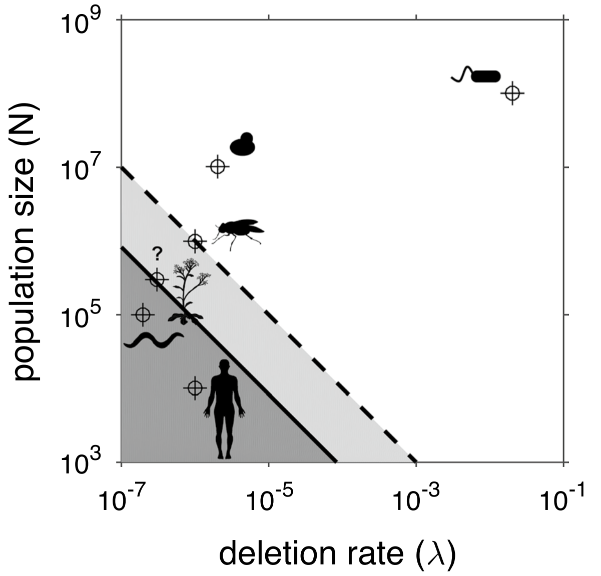

Fig. 1 Phase space between effective population size (*N*) and deletion rate upon duplication (λ) showing the region where genome size can increase by genetic drift (*shaded region*). The *solid line* corresponds to the limit $\lambda N = 1/12$ (strict), whilst the *dashed line* to $\lambda N = 1$ (soft). $\lambda \gg \mu$ is assumed. Six different organisms are contextualized (*S. enterica*, *S. cerevisiae*, *D. melanogaster*, *C. elegans*, *A. thaliana*, and *H. sapiens*). All λ values correspond to empirical estimations, except the value for the plant, which is a prediction

But the rates (μ and λ) confidently estimated until now (in mutation accumulation experiments) are really scarce, only available for some model organisms [17]. In addition, the deletion rates might be underestimated (see "Remark"). For *Arabidopsis thaliana*, e.g., only bioinformatic estimates have been produced, although these give values that differ from experimental estimates in several orders of magnitude. Based on the values of *D. melanogaster* and *C. elegans*, one can predict $\lambda \approx 10^{-7}$–10^{-6}/locus/gen. for *A. thaliana*, resulting in $\lambda N \approx 0.03$–$0.3 < 1$, as $N \approx 3{\cdot}10^{5}$. Higher eukaryotes have indeed more chances to transiently fix duplications by drift due to a reduced effective population size [6].

Conclusion

Definitely, $\lambda N < 1$ has to be satisfied in order to reach transient fixation of duplications by genetic drift. Otherwise, the population remains stably polymorphic regarding copy number. If this were the case, positive selective conditions should be invoked to explain an increase in genome size. After all, the precise characterization at the molecular level of the genome rearrangement rates, especially the deletion rate upon duplication, will shed much light to recognize how fortuitous was the path to reach the life that today we see on the Earth [18].

Limitations

The following limitations associated with the mathematical model were identified:

- Simplification to a birth–death process, while genome rearrangements may be more complex processes (e.g., gene relocation in the chromosome to stabilize a duplicate).
- No consideration of high-order variations, such as gene triplications or quadruplications, while these are found in nature. No consideration of individuals with zero copies, assuming they are deleterious and then quickly diluted.
- Population size assumed constant, while this may vary with time due to multiple environmental factors.

Acknowledgements
Not applicable.

Competing interests
The author declares that he has no competing interests.

Availability of data and materials
Copy number variation rates in different model organisms obtained from Refs. [1, 3, 12, 15, 16].

Consent to publish
Not applicable.

Ethics approval and consent to participate
Not applicable.

Funding
Work funded by the Spanish Ministry of Economy (MINECO/FEDER), Grant BFU2015-66894-P. Support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 23 June 2017 Accepted: 8 August 2017
Published online: 10 August 2017

References

1. Conrad DF, et al. Origins and functional impact of copy number variation in the human genome. Nature. 2010;464:704–12.
2. Anderson P, Roth J. Spontaneous tandem genetic duplications in *Salmonella typhimurium* arise by unequal recombination between rRNA (rrn) cistrons. Proc Natl Acad Sci USA. 1981;78:3113–7.
3. Lynch M, et al. A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc Natl Acad Sci USA. 2008;105:9272–7.
4. Ayala FJ. Darwin's greatest discovery: design without designer. Proc Natl Acad Sci USA. 2007;104:8567–73.
5. Ohno S. Evolution by gene duplication. New York: Springer Verlag; 1970.
6. Lynch M, Conery JS. The origins of genome complexity. Science. 2003;302:1401–4.
7. Hastings PJ, et al. Mechanisms of change in gene copy number. Nat Rev Genet. 2009;10:551–64.
8. Kimura M, Ohta T. The average number of generations until fixation of a mutant gene in a finite population. Genetics. 1969;61:763–71.
9. Zhang J, et al. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci USA. 1998;95:3708–13.
10. Force A, et al. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 1999;151:1531–45.
11. Einstein A. On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat. Ann d Phys. 1905;17:549–60.
12. Reams AB, et al. Duplication frequency in a population of *Salmonella enterica* rapidly approaches steady state with or without recombination. Genetics. 2010;184:1077–94.
13. Wong S, Wolfe KH. Birth of a metabolic gene cluster in yeast by adaptive gene relocation. Nat Genet. 2005;37:777–82.
14. Petrov DA, et al. Evidence for DNA loss as a determinant of genome size. Science. 2000;287:1060–2.
15. Schrider DR, et al. Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster*. Genetics. 2013;194:937–54.
16. Lipinski KJ, et al. High spontaneous rate of gene duplication in *Caenorhabditis elegans*. Curr Biol. 2011;21:306–10.
17. Katju V, Bergthorsson U. Copy-number changes in evolution: rates, fitness effects and adaptive significance. Front Genet. 2013;4:273.
18. Lynch M. The frailty of adaptive hypotheses for the origins of organismal complexity. Proc Natl Acad Sci USA. 2007;104:8597–604.