Mutation Research 789 (2016) 48–56

Contents lists available at ScienceDirect

Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis

journal homepage: www.elsevier.com/locate/molmut
Community address: www.elsevier.com/locate/mutres

Simple sequence repeat variations expedite phage divergence: Mechanisms of indels and gene mutations☆

Tiao-Yin Lin*

Department of Biological Science and Technology, Institute of Molecular Medicine and Bioengineering, National Chiao Tung University, Hsin Chu, Taiwan

ARTICLE INFO

Article history:
Received 20 November 2015
Received in revised form 26 March 2016
Accepted 11 April 2016
Available online 14 April 2016

Key words:
Simple sequence repeat
Insertion and deletion
Genome evolution
Molecular mechanism
Phage divergence
Human gene mutations

ABSTRACT

Phages are the most abundant biological entities and influence prokaryotic communities on Earth. Comparing closely related genomes sheds light on molecular events shaping phage evolution. Simple sequence repeat (SSR) variations impart over half of the genomic changes between T7M and T3, indicating an important role of SSRs in accelerating phage genetic divergence. Differences in coding and noncoding regions of phages infecting different hosts, coliphages T7M and T3, *Yersinia* phage ϕYeO3-12, and *Salmonella* phage ϕSG-JL2, frequently arise from SSR variations. Such variations modify noncoding and coding regions; the latter efficiently changes multiple amino acids, thereby hastening protein evolution. Four classes of events are found to drive SSR variations: insertion/deletion of SSR units, expansion/contraction of SSRs without alteration of genome length, changes of repeat motifs, and generation/loss of repeats. The categorization demonstrates the ways SSRs mutate in genomes during phage evolution. Indels are common constituents of genome variations and human diseases, yet, how they occur without preexisting repeat sequence is less understood. Non-repeat-unit-based misalignment-elongation (NRUBME) is proposed to be one mechanism for indels without adjacent repeats. NRUBME or consecutive NRUBME may also change repeat motifs or generate new repeats. NRUBME invoking a non-Watson-Crick base pair explains insertions that initiate mononucleotide repeats. Furthermore, NRUBME successfully interprets many inexplicable human di- to tetranucleotide repeat generations. This study provides the first evidence of SSR variations expediting phage divergence, and enables insights into the events and mechanisms of genome evolution. NRUBME allows us to emulate natural evolution to design indels for various applications.

1. Introduction

Phages have recently been utilized in targeted gene delivery, biocontrol of foodborne pathogens, and therapeutics for animal and human infections [1–3]. They are the most abundant entities in the biosphere with vastly diverse sequences [4]. The mechanisms by which the genetic diversity arises have been reviewed [4,5]. Genomic mosaicism is a pervasive feature of phage genomes [6]. Genetic exchanges by different recombination processes contribute to the genomic mosaicism [7–11]. In addition, duplication of genes, inversion of genomic segments, insertion of mobile elements, and point mutations also cause phage evolution [5,12–14]. Whether there are other mechanisms important to phage evolution awaits ascertainment.

Eukaryotic genomes harbor widespread simple sequence repeats (SSRs) of 1–6 base repeat units [15]. Prokaryotes have high genomic gene density and significantly less SSR repeat numbers and length relative to eukaryotes [16,17]. SSRs have been discovered in different eukaryotic viruses and utilized for virus genotyping [18–20]. For prokaryotic viruses, repeat sequences in the lysozyme gene of T4 phage have been employed for studying indel mutations [21]. Because bacteriophages have compact genomes with many functional constraints, the occurrence, variety, significance, and potential applications of SSRs in whole genomes need in-depth investigation.

Nucleotide indels are common in genomes and may affect human diseases [22]. Indels of repeat sequences have been proposed to occur by Streisinger slippage [21,23]. For deletions of nonrepeat nucleotides, mechanisms of misincorporation misalignment [24], dNTP stabilized misalignment [25], and transient

Abbreviations: SSR, simple sequence repeat; NRUBME, non-repeat-unit-based misalignment-elongation; T7M, T7 Meselson; DTRs, direct terminal repeats; RA, relative abundance; RD, relative density; RCRs, recombinant regions; NRCRs, non-recombinant regions; nWC, non-Watson-Crick; PDH, pyruvate dehydrogenase; XP, xeroderma pigmentosum.

☆ Data deposition the complete genome sequence of T7M was deposited to GenBank ID: JX421753.

* Corresponding author at: Department of Biological Science and Technology, National Chiao Tung University, 75 Poai St., Hsin Chu, Taiwan.
E-mail address: tylin@cc.nctu.edu.tw

http://dx.doi.org/10.1016/j.mrfmmm.2016.04.001
0027-5107/

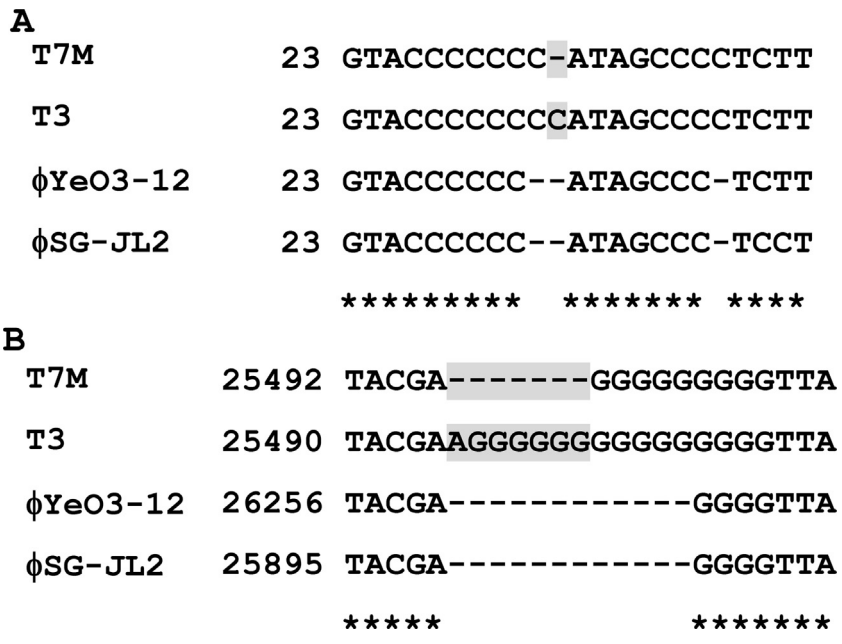

Fig. 1. Indels in the noncoding regions of phages. (A) Indels of the C repeats in the terminal repeats. (B) Indels of the G repeats and the A between ϕ*13* and gene *13*. The sequences of four phages are aligned. The numbers shown on the left side are the nucleotide numbers in the genome of each phage. The differences between T7M and T3 are shown in gray shade.

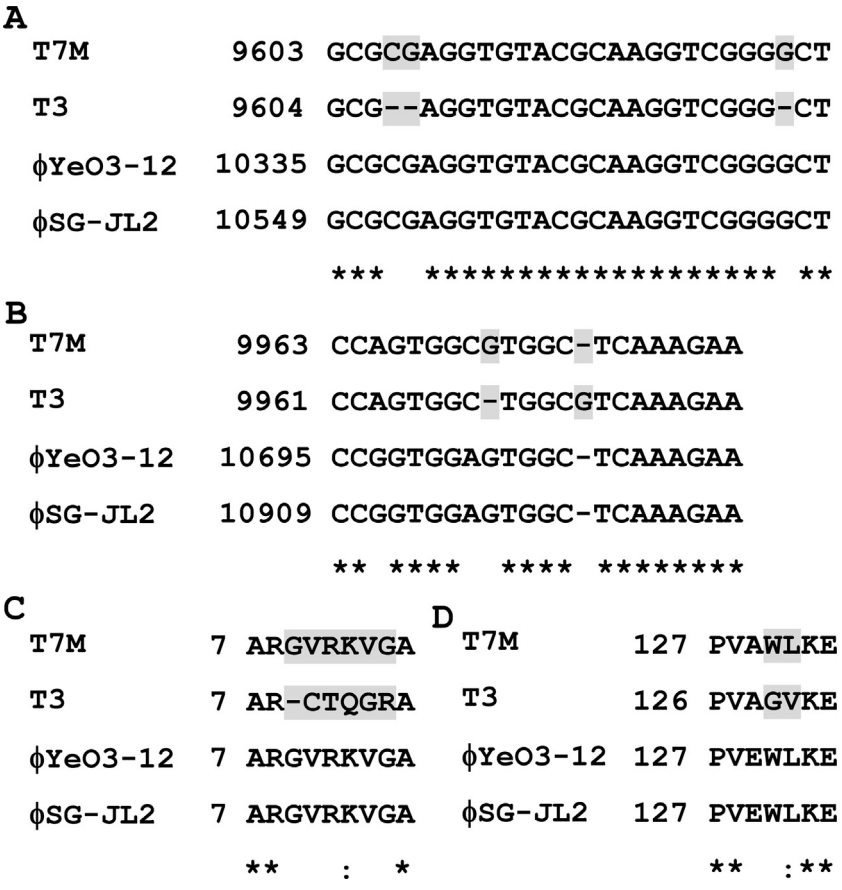

Fig. 2. Sequence changes in gene *3* of T7M and T3. (A) Indels in the CG and G repeats. (B) Indels of the nonrepeat nucleotide G. (C) Amino acid sequence change caused by (A). (D) Amino acid sequence change caused by (B). Sequences of gene *3* of ϕYeO3-12 and ϕSG-JL2 are also aligned for comparison. Positions of interest are shaded in grey for T7M and T3. The numbers shown on the left side are the genomic nucleotide numbers or gp3 residue numbers of each phage.

dislocation [26] have been proposed. The mechanism of nucleotide insertion at a position that does not manifest adjacent repeat sequence is poorly understood [27]. Furthermore, how a nonrepeat sequence generates its first repeated copy is unclear.