Available online at www.sciencedirect.com
ScienceDirect
C. R. Biologies 332 (2009) 254–266
COMPTES RENDUS BIOLOGIES
http://france.elsevier.com/direct/CRASS3/

Evolution / Évolution

A prominent role for segmental duplications in modeling Eukaryotic genomes

Romain Koszul [a,d,*], Gilles Fischer [b,c,d]

[a] *Department of Molecular and Cellular Biology, Harvard University, 7, Divinity Avenue, Cambridge, MA 02143, USA*
[b] *Unité de Génétique moléculaire des levures, Institut Pasteur, 25, rue du Docteur Roux, 75724 Paris cedex 15, France*
[c] *Université Pierre et Marie Curie-Paris, 6, UFR927, F.75005 Paris cedex 05, France*
[d] *CNRS, URA2171, F.75724 Paris cedex 15, France*

Accepted after revision 12 July 2008

Available online 2 December 2008

Presented by Claude Combes

Abstract

Segmental duplications (SDs) are a major element of eukaryotic genomes. Whereas their quantitative importance vary among lineages, SDs appear as a fundamental trait of the recent evolution of great-apes genomes. The chromosomal instability generated by these SDs has dramatic consequences both in generating a high level of polymorphisms among individuals and in originating numerous human pathogenic diseases. However, even though the importance of SDs has been increasingly recognized at the genomic level, some of the molecular pathways that lead to their formation remain obscure. Here we review recent evidences that the interplay between several mechanisms, some conservative, some based on replication, explains the complex SDs patterns observed in many genomes. Recent experimental studies have indeed partially unveiled some important aspects of these mechanisms, shedding interesting and unsuspected new lights on the dramatic plasticity of eukaryotic genomes. ***To cite this article: R. Koszul, G. Fischer, C. R. Biologies 332 (2009).***

Résumé

Impact des duplications segmentales sur l'évolution des génomes Eucaryotes. L'objectif de cette revue est de souligner l'importance du rôle structurant joué par les duplications segmentales (DS) au sein des génomes eucaryotes, au regard des découvertes récentes sur leurs mécanismes de formation. Alors que leurs proportions varient suivant les especes, les DS ont joué un rôle majeur dans l'évolution récente des génomes des grands singes. L'instabilité chromosomique associée à ces structures répétées induit un niveau élevé de polymorphisme au sein des populations humaines et est associée à de nombreux cancers et maladies géniques. Les mécanismes de formation des DS sont multiples et certains demeurent méconnus. Les différentes voies permettant de former des DS peuvent être distinguées suivant qu'elles impliquent ou non une étape de réplication et donc un gain net de matériel génétique. Des études récentes ont notamment permis de proposer un nouveau mécanisme basé sur la réplication des chromosomes et *a priori* susceptible de dupliquer n'importe quel locus, soulignant par là l'impressionnante plasticité des génomes eucaryotes. ***Pour citer cet article : R. Koszul, G. Fischer, C. R. Biologies 332 (2009).***

* Corresponding author.
E-mail address: rkoszul@mcb.harvard.edu (R. Koszul).

1631-0691/$ – see front matter
doi:10.1016/j.crvi.2008.07.005

Keywords: Segmental duplication; Replication; Replication factory; Genome evolution; Chromosome rearrangements; CNV; Genomic disease

Mots-clés : Duplication segmentale ; Réplication ; Usine de réplication ; Évolution génomique ; Réarrangements chromosomiques ; CNV ; Maladies géniques

1. Introduction

From the time chromosomes were shown to be the physical basis of heredity, scientists have expressed interest in the consequences of physical aberrations resulting in loss and gain of genetic material. Most of these early remarkable achievements and hypotheses originated from members of the group of Thomas H. Morgan and were developed through genetic analysis of *Drosophila melanogaster* mutants. A notable contribution came from Calvin B. Bridges who, soon after publishing his 1916 key PhD article on chromosomes, support of heredity [1], emphasized the interest duplicating of chromosomal regions as a mean to generate copies of identical genes as early as 1918. Such copies would be prone to accumulate independent mutations and thus acquire new functions, eventually contributing to speciation events [2]. Bridges' research on mutant flies eventually converged with his early interest in duplications to a remarkable climax, providing the basis for the discovery of HOX gene clusters [3]. This cluster appeared, indeed, to result from a series of gene and genome duplications [4]. A very early contribution to the study of segmental duplications (SDs) was made by another member of Morgan's group, Hermann J. Muller, who described X-ray irradiated flies in which a segment from the X chromosome had been duplicated and translocated onto chromosome II [5]. By observing apparent innocuousness from the presence in the same genome of these rearranged chromosomes II with intact X chromosomes, he elegantly proposed that such small duplications, eventually associated with translocations, were less likely to be deleterious than full aneuploidies, and therefore were probably an important source of genetic diversity. As we will see, his hypothesis appears amazingly accurate 70 years later. Early genetic works on filamentous fungi also revealed large duplications of chromosome segments resulting from non-reciprocal terminal translocations [6–8].

The importance of redundancy in eukaryotic genomes was nonetheless dramatically underestimated until the advent of genome sequencing and the revolution of the genomic era. Comparative genomic studies have described a diversity of genomic structures reminiscent of a variety of gene duplication mechanisms. Interestingly, whereas most eukaryotic genomes present a relatively homogenous level of genome redundancy (between 40% and 60% of genes in a given genome belong to gene families), these mechanisms appear somehow lineage-specific among species. Whole genome duplication (WGD), or polyploidy, is a phenomena originally proposed to account for the presence of two sets of identical chromosomes in maize [9] and commonly observed in plants lineages [10]. Such events have since been detected in the evolutionary history of many others species (see Jaillon, Aury and Wincker [11], this issue). However, in addition to these dramatic, relatively rare events, other mechanisms have contributed to generate the genetic redundancy observed in most eukaryotic genomes. Tandem gene duplications are easily recognizable by the physical proximity between the repeated gene copies. On the other hand, dispersed duplicates generally result from the retroposition of mRNA with the limitation that this mechanism often results in the formation of pseudogenes. Finally, a large proportion of eukaryotic genomes appear to have been extensively reshaped by duplications of large DNA segments. Initially considered as peculiar features of genome architecture, interest in SDs intensified with the release of complete genome sequences. A remarkable breakthrough occurred when it was established that recent duplications appear to constitute more than 5% of the human genome [12,13].

2. Segmental duplications among eukaryotic genomes

2.1. Segmental duplications and copy number variation regions in the human genome

The recent interest in segmental duplications as structural elements of eukaryotic genomes has without doubt been magnified by the discovery of their abundance in the human genome sequence. By considering the overrepresentation of some regions among whole-genome shotgun sequence reads, Eichler and collaborators identified potential duplicated regions and were thus able to generate the first genomic map of SDs [12]. With a later refinements, the original predictions proved to be close to the actual map, with approximately 5% of the genome consisting in duplications of DNA segments $>$1–5 kb sharing $>$90% identity, usually $<$300 kb al-