C. R. Biologies xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Comptes Rendus Biologies

www.sciencedirect.com

Trajectories of genetics, 150 years after Mendel/Trajectoires de la génétique, 150 ans après Mendel

Non-functional genes repaired at the RNA level

Gènes non fonctionnels réparés au niveau de l'ARN

Gertraud Burger

Department of Biochemistry, Robert-Cedergren Centre for Bioinformatics and Genomics, Université de Montréal, Montréal, Canada

ARTICLE INFO

Article history:
Received 14 March 2016
Accepted after revision 12 April 2016
Available online xxx

Keywords:
Gene fragmentation
Trans-splicing
Insertion and substitution RNA editing
Evolution of complexity
Protists
Diplonemids

ABSTRACT

Genomes and genes continuously evolve. Gene sequences undergo substitutions, deletions or nucleotide insertions; mobile genetic elements invade genomes and interleave in genes; chromosomes break, even within genes, and pieces reseal in reshuffled order. To maintain functional gene products and assure an organism's survival, two principal strategies are used – either repair of the gene itself or of its product. I will introduce common types of gene aberrations and how gene function is restored secondarily, and then focus on systematically fragmented genes found in a poorly studied protist group, the diplonemids. Expression of their broken genes involves restitching of pieces at the RNA-level, and substantial RNA editing, to compensate for point mutations. I will conclude with thoughts on how such a grotesquely unorthodox system may have evolved, and why this group of organisms persists and thrives since tens of millions of years.

Mots clés :
Gènes fragmentés
Épissage en *trans*
Édition de l'ARN par insertion et substitution
Évolution de la complexité
Protistes
Diplonémides

RÉSUMÉ

Les génomes et les gènes évoluent d'une façon permanente. Les séquences des gènes subissent des substitutions, des délétions ou des insertions de nucléotides ; des éléments génétiques mobiles envahissent les génomes et s'intègrent dans des gènes ; des chromosomes cassent, même à l'intérieur des gènes, et les morceaux se raboutent dans un ordre différent. Afin de maintenir des produits de gènes fonctionnels et d'assurer la survie d'un organisme, deux stratégies principales sont utilisées, la réparation du gène ou celle de son produit. Les types communs d'altération des gènes et la façon dont leur fonction est rétablie secondairement seront décrits. Ensuite sera détaillé le cas de gènes fragmentés systématiquement, que l'on retrouve dans un groupe de protistes peu étudié, les Diplonémides. L'expression de leurs gènes morcelés implique la jonction des morceaux au niveau de l'ARN, ainsi que l'édition des séquences de ces ARN afin de compenser les mutations ponctuelles. La question de savoir comment un système aussi peu habituel peut avoir évolué et de comprendre pourquoi ce groupe d'organismes persiste et prospère depuis une dizaine de millions d'années sera discutée en conclusion.

1. Genes that no longer tell their story

Text-book genes have a simple, straightforward make-up. They are encoded on a contiguous stretch of DNA, and

Email address: Gertraud.Burger@umontreal.ca.

http://dx.doi.org/10.1016/j.crvi.2016.04.004
1631-0691/

may include introns of the various types (Fig. 1). If a standard gene specifies a protein, the amino acid sequence of the gene product can be readily inferred from the nucleotide sequence of the coding regions using a codon-to-amino acid translation table. If the gene specifies a structural RNA, then the gene sequence is essentially identical with that of the transcript, only that thymidines (Ts) in DNA are replaced by uridines (Us) in RNA.

However, the past decades of genomics research have unearthed an unexpected range of aberrant genes whose coding sequence differs substantially from that of their functional product (i.e. protein or RNA). Differences between a gene and its product may exist at a 'microscopic' scale, affecting single nucleotides, or at a 'macroscopic' scale concerning length or structure. How do aberrant, seemingly non-functional genes give rise to functional proteins or structural RNAs? In other words, how does the cell compensate such genetic defects?

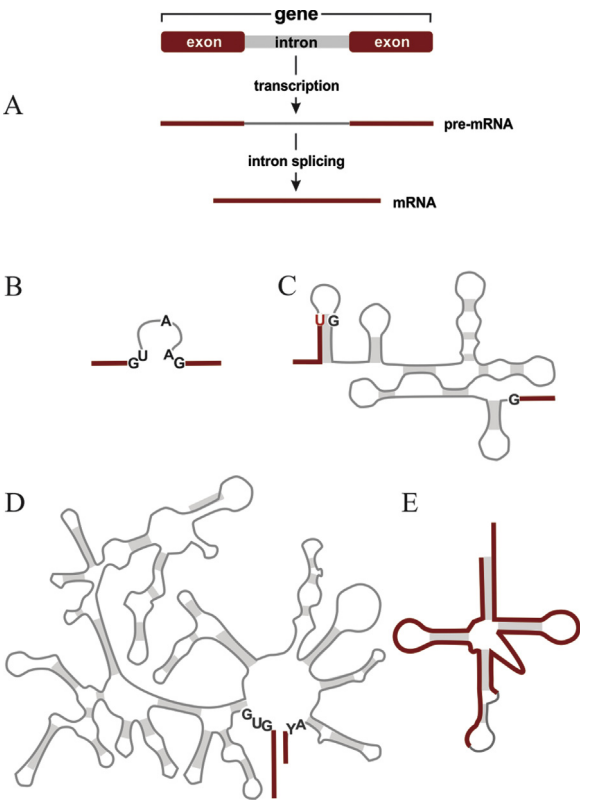

Fig. 1. Typical gene structure and common intron types. Red bars and lines, coding sequence; gray bars and lines, intron sequence. **(A)** Schema of a protein-coding gene that is interrupted by an intron. The separated coding regions are referred to as exons. The exons plus the intron are transcribed. The precursor mRNA (pre-mRNA) is then cleaved at the exon-intron boundaries, the intron is released, and the exons are religated by intron splicing, yielding a contiguous messenger RNA (mRNA; or a mature ribosomal or transfer RNA if the gene codes for a structural RNA). **(B–E)** Common intron types. Gray shading, regions held together by base pairing (secondary structure interactions). Conserved nucleotide positions that are important for splicing are marked black (positions in introns) or red (positions in exons). (B), spliceosomal introns predominant in the eukaryotic nucleus. **(C)**, Group I intron, **(D)**, Group II intron. Both are found in organelles, and less frequently in bacteria and in the nucleus. Group II also occurs in archaea. **(E)**, tRNA intron (also called 'archaeal' intron). It occurs in tRNA genes of archaea, many eukaryotes, but not bacteria. For a review on intron distribution, see [4].

Most spectacular cases of genes with nucleotide substitutions occur in mitochondria and plastids of dinoflagellates – alveolate algae including taxa notorious for toxic marine blooms. In certain species, gene and messenger RNA (mRNA) sequences differ by up to 10% of nucleotide positions. Apparently, these sequence edits occur at the RNA level, but it is still unknown whether during or after transcription [for a review on RNA editing, see 1].

Genes with large numbers of nucleotide deletions and insertions (indels) have been discovered first in mitochondria of trypanosomes, a protist group including important human and life stock pathogens. Many of the genes have remained unrecognized for a long time, because conceptual translation of their nucleotide sequence into protein yields reading frame shifts and premature stop codons. Only the comparison between DNA and mRNA sequences revealed numerous insertions and deletions in RNA, all of a single nucleotide type, U. It has been demonstrated experimentally that Us are removed or added after transcription, from the precursor-RNA. Depending on the gene, changes of this type can be massive, with nearly 80% nucleotides in an RNA generated post-transcriptionally.

Genes may be multiple times longer than their product due to insertion of mobile genetic elements. For this, nature has invented an arsenal of remedies. Some work at the level of RNA, others at the genomic and even protein level. For example, classical introns (spliceosomal, archaeal, Group I and Group II; see Fig. 1) are removed post-transcriptionally by RNA splicing. In contrast, transposable elements that populate the germline nucleus of certain ciliates, are eliminated via DNA rearrangements during formation of a working-copy nucleus (the somatic nucleus [2]). A third stratagem is employed to remove *hop* and *byp* insertion elements that are found in certain bacteriophages and in yeast mitochondria. These insertions are retained in mature mRNA and only eliminated during translation via ribosome hopping, also referred to as programmed translational bypassing [3].

Another most striking gene aberration is fragmentation – with pieces situated in distant chromosomal locations or even on different chromosomes – while its RNA or protein product is in one piece. There are a number of examples where the gene breakpoints are within introns, so that fragments contain exons plus adjacent intron halves. Such fragments are transcribed separately, but exons are still joined together accurately, because the intron splicing machinery is able to perform splicing across multiple pre-RNA molecules [4]. An instance of unusually massive gene fragmentation is described in more detail below. In this case, salvation does not rely on introns, but involves a process not encountered elsewhere before.

2. A case of massively fragmented genes – specifying perfectly conventional products

A system where the sequence of genes differs in an unprecedented degree from that of their products is found in a barely known protist group, the diplonemids. Diplonemids are unicellular flagellates thriving predominantly in marine,

Please cite this article in press as: G. Burger, Non-functional genes repaired at the RNA level, C. R. Biologies (2016), http://dx.doi.org/10.1016/j.crvi.2016.04.004