Tang and Lyons, Int J Evol 2012, 1:1
http://dx.doi.org/10.4172/ievj.1000e105

International Journal of Evolution

Editorial A SCITECHNOL JOURNAL

The Evolution of Genome Structure

Haibao Tang[1]* and Eric Lyons[2]

Genomes are often referred to as the "operating system" of living cells and organisms [1]. Analogous to the operating system of a computer or a smart phone, the architecture and the layout of the components within the genome are crucial to the coordination of gene expression and regulation. The genome structures of all living organisms have constantly been under modification over the course of evolution, through both small-scale (base substitutions, indels) and large-scale events (genome rearrangements, duplications). By studying the impact of these changes, we have come to understand the functional importance of various genetic elements as well as the regulatory hierarchy among these elements. Recent advances in massively parallel sequencing technology offer power to accurately assemble and reconstruct the genomes from a wide variety of cells and organisms. Perhaps some of the most intriguing questions are: What are the functionally important elements in the genomes? How they are organized? How do they change over time? These questions are addressed through the comparisons of multiple genomes in order to identify the conservation of nucleotides, genes and gene arrays. Sequences or genes that resist changes (i.e. under purifying selection) are presumed to carry out important biological functions. Likewise, sets of linked genes, or patches of genomic regions can also be conserved, which have often been referred to as 'domains' or 'blocks' by different researchers.

Domains are organizing principles of chromosome structure in nature, both in prokaryotes and eukaryotes [2]. Some genomic domains are not merely structural units but are presumed to be functional entities and were found to be highly conserved across many species. The genomes of prokaryotic cells are compact and multiple genes are often contained in cassettes called "operons". The genes within the same operons are transcribed onto the same messenger RNA and therefore tightly co-regulated. Additionally, bacterial chromosomes are found to be organized into independent topological domains such as the individual supercoiled loops seen under electron microscope [2]. In eukaryotic cells, gene sets that require tight co-regulation are no longer functionally constrained to be in the same 'neighborhood', possibly due to the asynchronous nature of transcription through the existence regulatory motifs on individual genes rather than on linked gene sets. However, there are still strong evidences that at least in some well-studied instances, "operon-like" structures do exist in complex eukaryotic organisms with gene orders conserved over a wide range of taxa, e.g. *Hox* gene clusters controlling vertebrate body plan development [3] or plant metabolic pathways [4].

Even without the functional constraints, genomes retain most of the structural features simply through common descent, known as conserved synteny blocks. Patterns of shared synteny (or lack thereof) are often used as relatively stable phylogenetic characters to classify species taxonomy as well as the dating of evolutionary events [5]. Although occurring at a rate much less frequent than base substitutions, genomic rearrangement events tend to disrupt the otherwise collinear ordering of matching genes and sequences, especially if the rearrangements are not selected against. Many genomic rearrangement events leave unique patterns or signatures that can be clearly identified from genomic comparisons. Recognizing these patterns on syntenic dotplots is a popular method to infer these events. For example, X-alignments between bacterial genomes reflect the high likelihood of genomic inversions occurring symmetrically around the origin of replication (Figure 1). Since most linear representations of bacterial genomes start at the origin, many such inversion events between two bacterial genomes result in their syntenic dotplot having an "X"-like pattern [6].

Over the course of evolution, genes of certain types and families proliferate and expand in numbers, which are often correlated with functional innovation and increasing organismal complexity. Duplicate genes typically originate via various mechanisms – proximal or tandem duplicates arise through unequal crossover, single gene duplication through retroposition and transposition, and whole genome duplication through polyploidy events. Whole genome duplication events have occurred during the evolution across all domains of life, but most strikingly in the lineage of the flowering plants, which experienced rampant polyploidy events including both shared as well as lineage-specific events [5]. Following the whole genome duplication, genomes became much more enriched in certain functional classes that are involved in transcriptional and signal

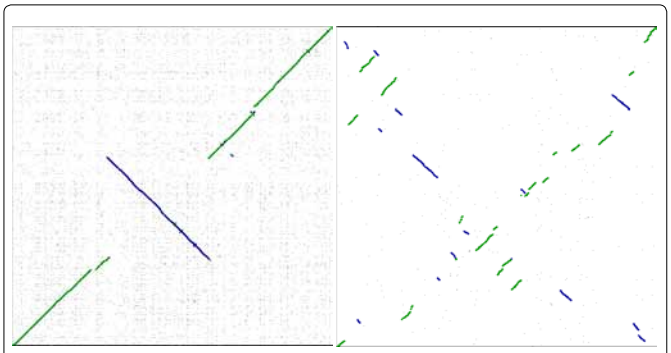

Figure 1: Syntenic dotplots showing varying degrees of genomic inversion events in bacterial genome comparisons, known as 'X-alignments'. Dot plots are generated using online comparative genomics suite CoGe SynMap utility [10]. Each dot represents a matching gene pair. Colored blocks represent synteny blocks. Left: comparison between *Anaeromyxobacter dehalogenans* strain 2CP-1 on x-axis vs. strain 2CP-1C on y-axis [11]. Right: *Prochlorococcus marinus* strain MIT9211 on x-axis vs. strain MIT9313 on y-axis [12].

***Corresponding author:** Haibao Tang, J. Craig Venter Institute, 9704 Medical Center Dr., Rockville, MD, USA, E-mail: tanghaibao@gmail.com

Received: June 29, 2012 **Accepted:** June 30 2012 **Published:** July 03, 2012

doi:http://dx.doi.org/10.4172/ievj.1000e105

transduction pathways, both upstream elements like protein kinases, and downstream elements like transcription factors [7]. Some of the new genes derived from the genome duplication events have become the vital parts in the regulation of flower development [7].

Perhaps the most exciting aspect of studying genome evolution is the ability to generalize and predict the evolutionary outcomes of how genomes behave when they duplicate themselves or multiple genomes merge into one. When similar genomes merge and co-exist in the same nucleus, many genetic loci share redundant molecular function among which some gene copies quickly become dispensable and are no longer under strong purifying selection, except for the genes that have adapted and diversified in their functions (Figure 2). Therefore we can consider these genome merger events as "nature's genetic experiments". By studying the fate of individual genes following the genome mergers, researchers have found that one genome often out-competes the other genome(s) through disproportionate gene loss among different genomes, sometimes happening as quickly as several generations [8]. This is known as the "genome dominance" effect and is perhaps related to the heterosis (hybrid vigor) [8]. Many questions remain though: how do the genomes remain their identities after entering in the same nucleus? How do the regulatory networks re-organize after the merger of two separate genomes?

The integration of high throughput data and adoption of genomic tools will continue to reveal interesting patterns and provide predictive models for evolutionary biologists and geneticists. Our evolutionary models are becoming more accurate as more genome data come on line. With each new genome sequenced, studied and compared, we gain more insight on how genomes work and how they retain their composition and structure, and what the likely changes are over time under various environmental cues. Understanding the nature and evolutionary behavior of the individual genetic elements as well as the global architecture within the natural genomes also lay a firm foundation for the future engineering of synthetic genomes [9].

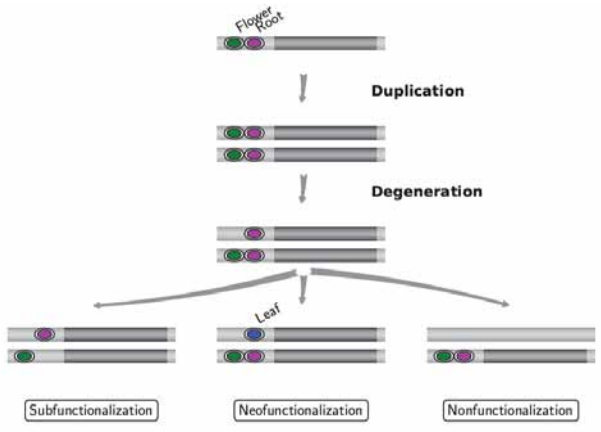

Figure 2: Interactions of multiple genetic loci lead to the shuffling of genetic elements and reorganization of genomes. In this case the initial gene contains two cis-regulatory elements that respond to activation signal in different tissues. The locus is then duplicated, resulting in two exact copies of the initial gene along with the control elements. Three evolutionary outcomes follow: either the initial two functions get partitioned in the two daughter copies (subfunctionalization), acquiring new regulatory elements (neofunctionalization), or loss of a redundant copy (nonfunctionalization).

References

1. Yan KK, Fang G, Bhardwaj N, Alexander RP, Gerstein M (2010) Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. Proc Natl Acad Sci U S A 107: 9186-9191.
2. Postow L, Hardy CD, Arsuaga J, Cozzarelli NR (2004) Topological domain structure of the Escherichia coli chromosome. Genes Dev 18: 1766-1779.
3. Fraser P, Bickmore W (2007) Nuclear organization of the genome and the potential for gene regulation. Nature 447: 413-417.
4. Field B, Osbourn AE (2008) Metabolic diversification--independent assembly of operon-like gene clusters in different plants. Science 320: 543-547.
5. Tang H, Bowers JE, Wang X, Ming R, Alam M, et al. (2008) Synteny and collinearity in plant genomes. Science 320: 486-488.
6. Eisen JA, Heidelberg JF, White O, Salzberg SL (2000) Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biol 1: research0011.1-11.9
7. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, et al. (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473: 97-100.
8. Freeling M, Woodhouse MR, Subramaniam S, Turco G, Lisch D, et al. (2012) Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. Curr Opin Plant Biol 15: 131-139.
9. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, et al. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329: 52-56.
10. Lyons E, Freeling M (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J 53: 661-673.
11. http://genomevolution.org/r/4y5i
12. http://genomevolution.org/r/4y5k

Author Affiliations

Top

[1]*J. Craig Venter Institute, 9704 Medical Center Dr., Rockville, MD, USA*
[2]*iPlant Collaborative, School of Plant Sciences, University of Arizona, Tucson, AZ, USA*