SCIENCE CHINA
Life Sciences

Evolution of genes and genomes in the genomics era

Song Ge and Ya-Long Guo

Citation: SCIENCE CHINA Life Sciences **63**, 602 (2020); doi: 10.1007/s11427-020-1672-0

View online: http://engine.scichina.com/doi/10.1007/s11427-020-1672-0

View Table of Contents: http://engine.scichina.com/publisher/scp/journal/SCLS/63/4

Published by the Science China Press

SCIENCE CHINA
Life Sciences

•INSIGHT•

April 2020 Vol.63 No.4: 602–605
https://doi.org/10.1007/s11427-020-1672-0

Evolution of genes and genomes in the genomics era

Song Ge[1,2*] & Ya-Long Guo[1,2]

[1]*State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, 100093 Beijing, China;*
[2]*University of Chinese Academy of Sciences, 100049 Beijing, China*

Received February 28, 2020; accepted March 2, 2020; published online March 13, 2020

Citation: Ge, S., and Guo, Y.L. (2020). Evolution of genes and genomes in the genomics era. Sci China Life Sci 63, 602–605. https://doi.org/10.1007/s11427-020-1672-0

During recent decades, we have witnessed substantial technological advances and revolutions in DNA sequencing. The range and scope of DNA sequencing applications has expanded tremendously, which have impacted all the disciplines of life sciences and even beyond (Shendure et al., 2017). The low-cost, high-throughput sequencing has brought vast new molecular and genomic data from both model and non-model organisms, and enables us to study the patterns and processes of evolution and the mechanisms underlying in an unprecedented scale and depth. In particular, genomic data plus sophisticated computational techniques and molecular biology have been combined in the dimension of evolution, thus changing the ways biologists view the world of life and providing the opportunity to answer long-standing questions about biodiversity and evolution (Wen et al., 2019). Here we present a brief summary of some recent investigations on the evolution of genes and genomes and highlight the new insights that these studies provided. We focus on four specific topics involving the evolution of gene gain and loss, transposable elements, RNA editing of organelle genomes, and phylogenetic reconstruction of major lineages, with emphasizing the need to unify strengths and insights across different disciplines of science to expand our understanding of evolution in general.

Genome evolution-gene gain or loss

Gene gain and loss events contribute to genome mutations and thus play important roles in evolution. The gain of genes or gene copies via the *de novo* gene formation and gene duplication has been extensively studies because the origin of new genes with novel functions is critical to the evolution of phenotypes and their adaptive potentials (Long et al., 2013; Albalat and Cañestro, 2016). Of various new genes, orphan genes are a special class of lineage-specific gene that lack detectable homologues in other lineages (Zhang et al., 2019). Although orphan genes might contribute to a variety of biological functions, their origination and function mechanisms remain largely elusive (Long et al., 2013). In a recent study on the *Caenorhabditis elegans* genome, Zhang et al. (2019) identified 893 orphan genes through a comprehensive and systematic computational pipeline and found that these orphan genes exhibited simple gene structures, including short in protein length, fewer exons and were frequently X-linked. Further RNA-seq data and gene ontology enrichment analyses indicated that these orphan genes were enriched with expression in embryo development and gonad, implying their potential function in early development. This study presented the first systematic evidence on the evolution of orphan genes and *de novo* origin of genes in nematodes, shedding new light on our understanding of these new genes. In another study on topological evolution of coexpression networks, Zu et al. (2019) analyzed the global structure and evolution of human gene coexpression net-

*Corresponding author (email: gesong@ibcas.ac.cn)

works driven by new gene integration. They found that the coexpression network comprised 334 small components and one "giant" connected subnet comprising of 6,317 interacting genes, and that the younger genes with a larger degree showed a property of hierarchical architecture. In particular, they indicated that gene duplication and orphan genes were two dominant evolutionary forces in shaping the coexpression network by developing new links through a "rich-gets-richer" mechanism.

In addition to gene gain, gene loss contributes to evolutionary change as well, as the "less-is-more" hypothesis proposed (Albalat and Cañestro, 2016). Loss-of-function (LoF) mutation is one of the main mechanisms underlying gene loss through the gain of a premature stop codon, splice site disruption, or disruption of a transcript reading frame (Albalat and Cañestro, 2016). LoF mutational variation provides a way for understanding of the phenotypic consequences of the loss of specific genes within diverse genomes while at the same time establishing their role in the evolutionary process. LoF mutations have recently gained attention owing to the improved detection power of advanced sequencing techniques (Narasimhan et al., 2016). Evidence showed that LoF mutations might affect important biological processes, such as development or resistance to stress in plants (Gujas et al., 2012), or intellectual disability in humans (Green et al., 2017). Recently, our study in *A. thaliana* based on population genomics, suggested that rapid evolution rate could be associated with a high frequency of LoF mutations (Xu et al., 2019). In particular, we found that 1% of *A. thaliana* genes with LoF variants are under positive selection, which probably correlated with local adaptation (Xu et al., 2019).

It should be noted, however, the mechanisms by which LoF mutations produce functional effects are complicated. A simple mechanism for a beneficial effect of a null mutation is the removal of a protein that is detrimental in the current environment (Gujas et al., 2012). It is also possible that LoF variants can act as dominant mutations, either act in a dominant-negative manner or dominant gain-of-function mutations (Xu et al., 2019). Nevertheless, the extent to which LoF mutations are correlated with adaptation and phenotypic diversification remain further investigation with multiple approaches (Albalat and Cañestro, 2016).

Evolutionary implications of transposable elements

A genome consists of a variety of elements, including protein-coding genes, non-coding RNAs, and diverse repeat elements. With the available of many genome sequences, the evolutionary patterns and dynamics of genomes have been the hot topics. Transposable elements (TEs) are the most variable components of the genomes, which can replicate and integrate into new genomic positions and are regarded as a major source of genomic mutations. Integration of TEs can provide novel genes or transcripts, provide regulatory elements and affect gene expression, generate genomic instability and rearrangements (Goerner-Potvin and Bourque, 2018). Therefore, TEs play a crucial role in shaping genomic architecture and phenotypic variation in diverse organisms. Changes in the environment can alter both the copy number of TEs and their effects on gene regulation, generating novel genetic and phenotypic variation of potential adaptive significance (Stapley et al., 2015). It has been demonstrated that TE mutations could affect adaptation to the environment in different organisms (Casacuberta and González, 2013; Li et al., 2018). In *Arabidopsis*'s relative *Capsella rubella*, for example, TE mutations drive rapid phenotypic variation, which could potentially help adapting to novel environments in species with limited genetic variation, to some extent reveals the mystery of genetic paradox of invasion (Niu et al., 2019).

To date, many lines of evidence demonstrated that TEs have the potential to quickly create abundant genetic diversity, thus being agents of shaping genomic architecture, phenotypic variation, and rapid adaptation (Casacuberta and González, 2013; Chuong et al., 2017). Nevertheless, the extents to which TEs can contribute either to rapid phenotypic variation and its mechanisms, or to the process of rapid adaptation, are unknown. In a recent study on transposable elements of human genome, Kellner and Makałowski (2019) scanned a large amount of data produced by ENCODE project for active transcription binding sites (TFBSs) located in TE-originated parts of polymerase II promoters. They found that of ca. 35,000 promoters in six different tissues, over 26,000 promoters harbored TEs. In particular, these TEs usually provide one or more of TFBSs in the host promoters, resulting in more than 6% of active TFBSs in these regions located in the TE-originated sequences. Thus, Kellner and Makałowski (2019) demonstrated that TEs contributed a large fraction of human TFBSs and showed that TEs played a significant role in shaping expression pattern in mammals and humans. Overall, such a study is not only informative for investigations on human genomes, but can also be of importance for our understanding of genome evolution in general.

RNA editing of organelle genomes

Mitochondria and chloroplasts are eukaryotic organelles and the organelle genomes work in concert with the nuclear genome to ensure proper organelle metabolism and biogenesis. RNA editing, a type of post-transcriptional modification, usually alters cytidine to uridine in plastids and mitochondria and plays important roles in various plant de-

velopmental processes such as organelle biogenesis, adaptation to environmental changes, and signal transduction (Ichinose and Sugita, 2017). Although RNA editing is a prevalent phenomenon in land plants and has been uncovered in all major plant lineages, the number of editing sites is variable, even within evolutionary lineages (Ichinose and Sugita, 2017). Previous studies showed that massive loss of mitochondrial RNA editing sites from *Welwitschia mirabilis* but the potential mechanisms of RNA site loss remain unclear (Fan et al., 2019). Fan et al. (2019) analyzed the pattern of RNA editing in the mitochondrial and plastid genomes of two gymnosperm plants, *W. mirabilis* and *Ginkgo biloba*, by comparing genomic sequences with transcripttomic and reverse-transcription PCR sequencing data. They found only 99 editing sites located in 13 protein-coding genes in the mitogenome and a complete loss of RNA editing from the plastome for *W. mirabilis*, which differed significantly from *G. biloba* in which 1,405 mitochondrial and 345 plastid editing sites were detected. Further investigation revealed that the editing loss from *W. mirabilis* was mainly due to the substitution of editable cytidines to thymidines at the genomic level, which might be caused by retroprocessing. Fan et al. (2019) was the first study to uncover extensive editing loss from both the mitogenome and plastome in a single evolutionary lineage and suggested that gene expression level and retroprocessing both contributed to the evolution of RNA editing in plant organellar genomes.

Recent advances in genome-editing technologies provide promising tools for crop improvement and these technologies, particularly the CRISPR–Cas9 system, have been successfully used in many crop species, such as rice, wheat, sorghum, maize, tomato, potato, and soybean (Li et al., 2019). However, targeted modification of the mitochondrial genome has not yet been achieved. A recent study developed a new method (mitoTALENs) to knock out CMS-associated genes in rice and rapeseed using transcription activator-like effector nucleases (TALENs) with mitochondria localization signals, and demonstrated its usefulness in modifying the mitochondrial genome in plants (Kazama et al., 2019). To understand the RNA editing, therefore, is of importance in both theory and practice because it not only facilitates the functional study of the genes involved in organelle biogenesis and development, but may also provides a valuable technique in practice of agricultural application (Kazama et al., 2019).

Exploring evolutionary history with integrative approaches

Evolutionary biology mainly concerns historical events that happened in the past and thus scientists in this area often make inferences about invisible patterns and processes. In recent decades, evolutionary biologists have made great strides in uncovering evolutionary history of almost all taxonomic groups across the tree of life. However, a clear understanding of exact patterns and processes of evolution for a specific lineage remains a central challenge and requires integrative investigations involving phylogeny, biogeography, paleontology, and geology (Wen et al., 2019). Such endeavors have been successfully performed focusing on either a specific lineage (e.g., Zou et al., 2008; Lu et al., 2018) or a given geographic region (e.g., Price and Wagner, 2018). As an important group of freshwater fish, the Cypriniformes comprise ca. 4,200 species accounting for 25% of the diversity of all freshwater fish and are widely distributed across the world's continents except Antarctica, South America, and Australia (Tao et al., 2019). Despite these, the evolutionary history of this major freshwater fish group remains largely unresolved. Tao et al. (2019) reconstructed the phylogenetic relationship of this group using sequences of one mitochondrial and 15 nuclear genes, with the hope to gain insights into the evolutionary history of Cypriniformes. Their phylogenetic and biogeographical analyses confirmed the monophyly of Cypriniformes and seven constituent subclades. They further indicated that the origin of the Cypriniformes was about 193 Mya during the early Jurassic, coinciding with the onset of the Pangaea breakup, suggesting that the burst in species diversity in Cyprinidae was possibly in response to the plasticity of pharyngeal dentition. This study helps to improve our understanding of the evolutionary history of this diverse and important freshwater fish group.

Phylogenetic reconstruction has now advanced into the age of phylogenomics that takes full use of sequence information from the whole genomes and showed its great power in resolving difficult phylogenetic problems. Phylogenomic approaches has also been successfully used to resolve problems involving the evolution of genomes, gene families and gene function (Fan et al., 2019; Zhang et al., 2019), as highlighted above. With more and more organisms with whole genome sequenced, it has been an increasing practice to apply multiple genes, particularly sequences from multiple genomes, to reconstruct phylogenies at different hierarchical levels. Moreover, the importance of integrative research involving different disciplines has been widely acknowledged. We believe that this set of work would stimulate empirical studies that simultaneously incorporate evidence from phylogenetic relationships, divergence times, geographic distribution, functional traits, and ecology.

Future perspectives

Over the past decades, we have made great progress on the evolution of genomes, gene families and gene function as well as the organism history, with increasingly easy access to

full genome sequences. Nevertheless, many fundamental questions in evolutionary biology remain unanswered or have not explained clearly in terms of conceptual dimension and mechanism, including the tree of life, speciation and extinction, natural selection and adaptation, evolution of cooperation and conflict, origin of sex and mating system, evolution of gene functions and gene networks, etc. To achieve these goals, we need exploration at different time scales, from the long time scale for divergence of major lineages to short time scale among populations within species. Additionally, the cooperation and interaction of scientists with diverse expertise and from different disciplines such as molecular biology, genome biology, ecology, and bioinformatics are necessary and highly encouraged to ultimately to solve challenging scientific problems in life sciences.

Compliance and ethics *The author(s) declare that they have no conflict of interest.*

Acknowledgements *This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (XDB31000000) and the National Natural Science Foundation of China (91731301).*

References

Albalat, R., and Cañestro, C. (2016). Evolution by gene loss. Nat Rev Genet 17, 379–391.

Casacuberta, E., and González, J. (2013). The impact of transposable elements in environmental adaptation. Mol Ecol 22, 1503–1517.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet 18, 71–86.

Fan, W., Guo, W., Funk, L., Mower, J.P., and Zhu, A. (2019). Complete loss of RNA editing from the plastid genome and most highly expressed mitochondrial genes of *Welwitschia mirabilis*. Sci China Life Sci 62, 498–506.

Goerner-Potvin, P., and Bourque, G. (2018). Computational tools to unmask transposable elements. Nat Rev Genet 19, 688–704.

Green, C., Willoughby, J., and Balasubramanian, M. (2017). De novo *SETD5* loss-of-function variant as a cause for intellectual disability in a 10-year old boy with an aberrant blind ending bronchus. Am J Med Genet 173, 3165–3171.

Gujas, B., Alonso-Blanco, C., and Hardtke, C.S. (2012). Natural Arabidopsis brx loss-of-function alleles confer root adaptation to acidic soil. Curr Biol 22, 1962–1968.

Ichinose, M., and Sugita, M. (2017). RNA editing and its molecular mechanism in plant organelles. Genes 8, 5.

Kazama, T., Okuno, M., Watari, Y., Yanase, S., Koizuka, C., Tsuruta, Y., Sugaya, H., Toyoda, A., Itoh, T., Tsutsumi, N., et al. (2019). Curing cytoplasmic male sterility via TALEN-mediated mitochondrial genome editing. Nat Plants 5, 722–730.

Kellner, M., and Makałowski, W. (2019). Transposable elements significantly contributed to the core promoters in the human genome. Sci China Life Sci 62, 489–497.

Li, G., Liu, Y.G., and Chen, Y. (2019). Genome-editing technologies: the gap between application and policy. Sci China Life Sci 62, 1534–1538.

Li, Z.W., Hou, X.H., Chen, J.F., Xu, Y.C., Wu, Q., González, J., and Guo, Y.L. (2018). Transposable elements contribute to the adaptation of *Arabidopsis thaliana*. Genome Biol Evol 10, 2140–2150.

Long, M., VanKuren, N.W., Chen, S., and Vibranovski, M.D. (2013). New gene evolution: little did we know. Annu Rev Genet 47, 307–333.

Lu, L.M., Ickert-Bond, S., and Wen, J. (2018). Recent advances in systematics and evolution of the grape family Vitaceae. J Syst Evol 56, 259–261.

Narasimhan, V.M., Hunt, K.A., Mason, D., Baker, C.L., Karczewski, K.J., Barnes, M.R., Barnett, A.H., Bates, C., Bellary, S., Bockett, N.A., et al. (2016). Health and population effects of rare gene knockouts in adult humans with related parents. Science 352, 474–477.

Niu, X.M., Xu, Y.C., Li, Z.W., Bian, Y.T., Hou, X.H., Chen, J.F., Zou, Y.P., Jiang, J., Wu, Q., Ge, S., et al. (2019). Transposable elements drive rapid phenotypic variation in *Capsella rubella*. Proc Natl Acad Sci USA 116, 6908–6913.

Price, J.P., and Wagner, W.L. (2018). Origins of the Hawaiian flora: Phylogenies and biogeography reveal patterns of long-distance dispersal. J Syst Evol 56, 600–620.

Shendure, J., Balasubramanian, S., Church, G.M., Gilbert, W., Rogers, J., Schloss, J.A., and Waterston, R.H. (2017). DNA sequencing at 40: past, present and future. Nature 550, 345–353.

Stapley, J., Santure, A.W., and Dennis, S.R. (2015). Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. Mol Ecol 24, 2241–2252.

Tao, W., Yang, L., Mayden, R.L., and He, S. (2019). Phylogenetic relationships of Cypriniformes and plasticity of pharyngeal teeth in the adaptive radiation of cyprinids. Sci China Life Sci 62, 553–565.

Wen, J., Nie, Z.-., and Ickert-Bond, S.M. (2019). Advances in biogeography in the age of a new modern synthesis. J Syst Evol 57, 543–546.

Xu, Y.C., Niu, X.M., Li, X.X., He, W., Chen, J.F., Zou, Y.P., Wu, Q., Zhang, Y.E., Busch, W., and Guo, Y.L. (2019). Adaptation and phenotypic diversification in Arabidopsis through loss-of-function mutations in protein-coding genes. Plant Cell 31, 1012–1025.

Zhang, W., Gao, Y., Long, M., and Shen, B. (2019). Origination and evolution of orphan genes and *de novo* genes in the genome of *Caenorhabditis elegans*. Sci China Life Sci 62, 579–593.

Zou, X.H., Zhang, F.M., Zhang, J.G., Zang, L.L., Tang, L., Wang, J., Sang, T., and Ge, S. (2008). Analysis of 142 genes resolves the rapid diversification of the rice genus. Genome Biol 9, R49.

Zu, J., Gu, Y., Li, Y., Li, C., Zhang, W., Zhang, Y.E., Lee, U.J., Zhang, L., and Long, M. (2019). Topological evolution of coexpression networks by new gene integration maintains the hierarchical and modular structures in human ancestors. Sci China Life Sci 62, 594–608.