Current Opinion in **Genetics & Development**
Editors **Marnie E. Halpern** USA and **Steven Leach** USA

Volume 35 (2015)

Editorial Inquiries

Current Opinion in Genetics & Development,
Elsevier
Radarweg 29
1043 NX Amsterdam
The Netherlands
Tel: +31 (0)20 485 3044
Fax: +31 (0)20 485 2370
e-mail: ja.bakker@elsevier.com

Current Opinion in Genetics & Development
ISSN 0959-437X
is published every two months by
Elsevier
www.elsevier.com

The Aims of the journal can be found at
www.elsevier.com/locate/gde

Current Opinion in Genetics & Development
is indexed and/or abstracted by
BIOSIS
CAB Abstracts International
CAB Health
Chemical Abstracts
EMBASE
Medline
Also covered in the abstract and citation database SCOPUS®
Full text available on ScienceDirect®

Editorial Board

2015 Contents

The subject of genetics and development is divided into six major sections, each of which is reviewed once a year.

ELSEVIER

Amsterdam – Boston – London – New York – Oxford – Paris – Philadelphia – San Diego – St. Louis

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Volume 35, December 2015

CONTENTS

Abstracted/indexed in: BIOSIS, CAB Abstracts International, CAB Health, Chemical Abstracts, EMBASE, Medline. Also covered in the abstract and citation database SCOPUS®. Full text available on ScienceDirect®

Cover blurb

Convergence towards a similar genome architecture in phylogenetically unrelated plant pathogens. The flanking distance between neighboring genes provides a measurement of local gene density and is displayed as a color-coded heat map based on a whole genome analysis of the fungus *Leptosphaeria maculans* and the oomycete *Phytophthora infestans*. In addition, the figure displays the distribution of Avr effector genes of L. maculans and P. infestans according to the length of their 50 and 30 flanking intergenic regions. Note how in both cases the Avr effector genes primarily occupy the gene sparse regions of the genome. (See the paper of Suomeng Dong, Sylvain Raffaele and Sophien Kamoun.)

DOI 10.1016/S0959-437X(15)00137-9

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Editorial overview: Genomes and evolution: ''Seq-ing'' answers in life's genomes

Antonis Rokas and Pamela S Soltis

Current Opinion in Genetics & Development 2015, **35**:iv–vi

For a complete overview see the Issue

Available online 9th December 2015

http://dx.doi.org/10.1016/j.gde.2015.11.005

Antonis Rokas

Department of Biological Sciences, Department of Biomedical Informatics, and Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN 37235, USA
e-mail: antonis.rokas@vanderbilt.edu

Antonis Rokas is the Cornelius Vanderbilt Chair of Biological Sciences at Vanderbilt University. He received his Ph.D. in 2001 from Edinburgh University, where he trained with Graham Stone. After his Ph.D., he worked as a Human Frontiers Science Program Long-Term postdoctoral fellow with Sean B. Carroll at the University of Wisconsin-Madison. He has been a faculty member at Vanderbilt University since 2007. Research in the Rokas lab focuses on the study of the DNA record to gain insight into the patterns and processes of evolution. Through a combination of computational and experimental approaches, his current research aims to understand the evolution of human pregnancy, the molecular foundations of the fungal lifestyle, and the reconstruction of the tree of life.

Pamela S Soltis

Florida Museum of Natural History and the Genetics Institute, University of Florida, Gainesville, FL 32611 USA
e-mail: psoltis@flmnh.ufl.edu

Pamela S. Soltis is a Distinguished Professor and Curator of Molecular Systematics and Evolutionary Genetics in the Florida Museum of Natural History at the University of Florida. She received her Ph.D. from the University of Kansas in 1986. Following graduation, she joined the faculty at Washington State University and moved to the University of Florida in 2000. Her research interests are in plant diversity, with an emphasis on the use of phylogenies for understanding broad questions of plant evolution. She is particularly interested in the evolutionary

Ever since the dawn of biology, scientists have harnessed biodiversity to discover life's fundamental biological processes and mechanisms. Mendel discovered the laws of genetics by studying peas, and Darwin got insights on how natural selection works through the study of pigeons. Frogs and chickens facilitated great advances in embryology, and the ''one gene, one enzyme'' hypothesis was formulated through studying the bread mold *Neurospora crassa*. Similarly, the gene responsible for the human metabolic disease alkaptonuria, Garrod's first inborn error of metabolism, was discovered studying the mold *Aspergillus nidulans*, and the discovery that RNA has catalytic, in addition to replicative, functions was made by studying the ciliated protist *Tetrahymena*.

In recent years, the advent of high-throughput sequencing technologies has revolutionized the study of organismal and molecular diversity, and of the mechanisms and processes that have given rise to it. This issue is dedicated to reviews on the latest advances in the making of biodiversity, with special emphasis on organisms that have traditionally been invisible or difficult to study, on important biological processes that have for some reason or other until recently been neglected, and on approaches that aim to integrate knowledge from life's diversity to advance understanding of the evolutionary process as well as of the human condition.

New perspectives on life's organismal and molecular diversity

Life, at least as we know it, would be impossible without microbes. It is thus surprising to think that, only in the last decade or so, have we truly begun to explore microbial diversity. As Gibbons and Gilbert discuss, the millions of microbial ''species'' so far identified through sequencing form distinct ecological communities in coordination with, as well as in response to, specific environments and conditions. The outcome of this interplay across time and space is that microbes end up engineering their environments, an attribute that could be harnessed to improve human health or even Earth's atmosphere.

What's true of microbes is also true of proteins—it's hard to imagine life without them. And yet, most proteins' functions remain unknown or hypothetical. One intriguing approach is to consider life's proteins as a language whose grammar and syntax are still being deciphered; if protein domains are its words, what rules can be inferred from studying the repertoire of protein domain architectures found in life's genomes? As Scaiewicz and Levitt discuss, examining the ever-increasing amount of

consequences of polyploidy, from the genomic to the macroevolutionary levels. Using museum specimens as a source of biodiversity data, she is addressing the response of plant communities to climate change and developing tools to integrate specimen data with phylogenies, genetics, and computational resources.

genome sequence data has led to a nearly full catalogue of the words (domains) that make up the protein language, but we are still learning a great deal about its sentences; the number of protein domain combinations is increasing on par with the number of protein sequences. This likely reflects the functional versatility that many protein domains exhibit as well as the speed with which evolutionary mechanisms can generate novel architectures.

Another, complementary, approach to understanding protein function and evolution is to examine families of protein domains that are the exception rather than the rule. Das, Dawson and Orengo focus on that 7% of unusually structurally and functionally diverse domain superfamilies and summarize the several different molecular mechanisms by which diversity can arise. Importantly, as these mechanisms appear to generate functional diversity through altering the domains' ligands, their substrates or their contexts but not the underlying chemistry, general functions can be inferred even for domains from the most diverse superfamilies.

A great illustration of how diversity in protein domain architecture translates to biological diversity is offered by Wang, Sivonen and Fewer's review on polyketide synthases (PKSs) and nonribosomal peptide synthases (NRPSs), families of biosynthetic enzymes that are key to producing an amazing variety of secondary metabolites. Sleuthing the diversity of life's genomes for secondary metabolic pathways is revealing a large amount of hidden diversity in domain architecture that not only challenges existing paradigms of PKS and NRPS biosynthesis but also provides a wonderful new source of candidate drugs.

The making of biodiversity

Genome sequences are superb markers of diversity and rich archives of history and function, so it is hardly surprising that they are leading the way to advance understanding of the drivers of biodiversity. As Hittinger and colleagues discuss, the remarkable range of ecologies that yeasts exhibit is associated with metabolic diversity. Although the genomics of most yeast biodiversity remains unexplored, a slew of projects aimed at dramatically increasing the genomic sampling of yeast populations and species, coupled with the wealth of knowledge on many aspects of yeast metabolism and the availability of unparalleled genetic dissection tools, promise to create a superb model lineage for studying the evolution of eukaryotic biodiversity.

Genomes are also paving the way toward understanding the patterns and processes that underlie the evolutionary diversification of animals, the topic reviewed by Dunn and Ryan. Efforts to gain insights into the genetic drivers of animal diversification have been perennially derailed by (scientifically unfounded) expectations that early-branching lineages should be the least genomically and phenotypically complex and that our lineage, the vertebrates, should be the most genomically and phenotypically complex. These expectations are finally giving way to a more nuanced understanding of the relationship between genotype and phenotype, as Simonti and Capra discuss in their review on the evolution of the human genome. This is largely due to the increasing appreciation of the contribution of genetic regulatory mechanisms in generating phenotypic diversity as well as to the advent of several transformative sequencing-based technologies for the genome-wide identification of different types of genetic elements that modulate gene expression.

Perhaps the greatest advances in our understanding of the genetic drivers of biodiversity have been made by studying lifestyles related to human affairs, such as pathogenic and domesticated ones. As Leckenby and Hall argue, genome evolution of eukaryotic pathogens and parasites that infect animals is characterized by gene family expansion, gene loss and episodes of genomic rearrangement. All these processes are also at play in fungi and oomycetes that have independently evolved into plant parasites, as described by Dong, Raffaele and Kamoun. Intriguingly, the genomes of these plant parasites have both gene-dense, repeat-poor regions that contain housekeeping genes as well as gene-sparse, repeat-rich regions that contain virulence factors. The two types of regions differ dramatically in their rates of evolutionary change, raising the hypothesis that such ''two-speed'' genome architectures may have been adaptive for survival following frequent jumps between hosts. Whereas filamentous plant parasites have genomes that are bigger than their free-living relatives (Dong, Raffaele and Kamoun), the genomes of domesticated microbes discussed by Gibbons and Rinker bear several hallmarks of genomic specialization. There is more than one path to this genomic specialization, however, as it sometimes stems from the pseudogenization or loss of metabolic genes and pathways and other times through their selective acquisition or retention through hybridization or horizontal gene transfer.

The mechanisms of genome evolution

Genomic comparisons across the tree of life, such as those described in the reviews of this volume, show abundant evidence that several different processes such as hybridization, horizontal gene transfer, genomic rearrangement, gene loss and pseudogenization sculpt genome evolution. But arguably the most important source of gene innovation is gene duplication.

A special, but particularly influential, case of gene duplication is whole-genome duplication. As discussed by Soltis, Marchant, Van de Peer and Soltis, whole-genome duplications or polyploidization events have had a huge impact on plant genome evolution, yet what follows these wholesale doublings or multiplications of entire genomes is only beginning to be understood. Insights obtained from characterization of the genetic changes in genomes that recently became polyploid might be key for devising strategies to detect the weak signals left by polyploidization events that transpired deeper in the past. Scanlon, Fowler and Freeling focus on the non-random patterns of loss of genes duplicated during polyploidization; because the genes retained tend to encode proteins with regulatory functions, the authors provocatively argue that they inadvertently ''drive'' genome evolution toward increased regulatory complexity.

Polyploidy is one explanation for the remarkable differences in the sizes of plant genomes; the other one is repeat content. As discussed by Dodsworth, Leitch and Leitch, repeats have a profound impact on gene evolution, both through their influence on the regulation of nearby genes as well as through elevating mutational rates associated with gene copy number and structural variation. Repeat content, especially in the form of transposable elements, might have also been the underlying cause for the evolution of epigenetic mechanisms in eukaryotes. As argued by Maurer-Alcalá and Katz, these epigenetic mechanisms were subsequently coopted in other genomic functions and are a major contributor to the remarkable diversity of regulatory landscapes observed across eukaryotes.

Although the bulk of insights into the mechanisms of genome evolution stems from comparative studies, Jerison and Desai cogently argue why laboratory microbial evolution experiments coupled with high-throughput sequencing can be a powerful complementary approach to understanding the mechanisms of genome evolution. Using epistasis, in which the fitness effects of mutations are background-dependent, as an example, they provide several striking examples of how a given mutation potentiates or precludes subsequent ones in a gene, pathway or the genome.

Utilizing evolution to understand human disease

Aside from their intrinsic value for describing and understanding life's diversity and the mechanisms that give rise to it, genome comparisons are also proving invaluable for understanding human disease. As discussed by McWhite, Liebeskind and Marcotte, identification of conserved genetic networks can be exploited to identify non-obvious models of human disease and new disease gene candidates; conversely, lack of conservation at the genetic level can be invaluable in avoiding inferences based on phenotypic conservation. As technologies for sequencing and functionally testing genomes continue to expand their reach beyond a few model systems, so will biodiversity's potential to improve out understanding of the processes and mechanisms that give rise to life, humanity and disease.