ELSEVIER

Genomes and evolution
Deciphering the genomic palimpsest

Editorial overview

David Haig and Steve Henikoff

Current Opinion in Genetics & Development 2004, **14**:599–602

0959-437X/$ – see front matter

DOI 10.1016/j.gde.2004.10.002

David Haig

Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA
e-mail: dhaig@oeb.harvard.edu

David Haig is a theoretical biologist interested in intragenomic and intergenerational conflicts. He has a current interest in the evolutionary biology of huddling.

Steve Henikoff

Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109-1024, USA
e-mail: steveh@fhcrc.org

Steve Henikoff studies epigenetic mechanisms in *Drosophila* and *Arabidopsis* and develops tools for genomic research.

Introduction

A genome is a great compendium of interwoven texts written at different times by different authors (we speak, of course, metaphorically). Some genomes are prolix whereas others are terse. The original text may be supplemented by epigenetic annotations or overwritten by the graffiti of parasitic elements. Sometimes an insertion from another story becomes incorporated into a new and unintended message. An organism's evolutionary history lies hidden within the multiple layers of this document if only we knew how to decipher a text without a unifying narrative or single coherent plot. Several of these compendia have now been published, and many more are in press, but the problem of interpretation has just begun. This issue of *Current Opinion in Genetics and Development* provides an anthology of commentaries on these documents and their interpretation. A single story-line may be hard to discern, but there are several recurrent themes.

Selfish genetic elements and genome evolution

Although large-scale genomic sequencing projects were motivated mostly by the desire to precisely map genes, a by-product has been the identification of multitudes of genomic parasites that are much more abundant than genes in genomes like ours. Among the insights to come from sequencing efforts are the interesting consequences of selfish genetic elements for gene and protein evolution.

Sequences derived from L1 retrotransposons make up ≈ 20% of the human genome. *Alu* elements that piggy-back on the replication machinery of L1s make up an additional 10%. **Jurka** describes how *Alu* elements are responsible for much of the current action in our genomes. Ectopic recombination between *Alu*s is a source of chromosomal rearrangements and segmental duplications, and recent evidence suggests that *Alu* 'exonization' has led to new gene functions.

''First, do no harm'' is not only the physician's credo, but also a rule adhered to by homing endonucleases. Once they find a good home, they stay there. **Burt** and **Koufopanou** argue that homing endonucleases should degenerate once they reach fixation, having filled all available niches. Therefore, finding a new home — either within the host or elsewhere — must occur. The extreme specificity of homing endonucleases has made them increasingly popular for genomic targeting applications.

Where do new genes come from? According to **Daubin** and **Ochman**, duplication and divergence is a slow conservative process, but novelty might

arise from outside acquisitions. In some cases, immediate benefits accrue. Bacteriophages may be important vectors for the horizontal transfer of genes among bacterial genomes, sometimes providing a genetic capability otherwise lacking from the host genome. In other cases, today's parasite is a potential source of future novelty. An initially successful bacteriophage might find itself domesticated, with the host retaining some useful function and discarding the rest.

Expanding and contracting genomes

Vinogradov discusses the evolution of genome size. Why do our genomes contain so much DNA when so little of it has identifiable function, and why do many related eukaryotes show wide variation in DNA amount? Whereas some of the old 'C-value paradox' was addressed by the documentation of selfish elements and their carcasses (see the paper by **Jurka**), a lively discussion has been raging as to whether mutation or selection underlies changes in genome size, and just how these changes occur.

Some of the best-studied examples of extreme genome reduction are long-established endosymbiotic bacteria of eukaryote hosts. What are the processes that lead to this genome reduction? To address this question, **Moran** and **Plague** review evidence from genome comparisons between free-living bacteria and close relatives that can replicate only within eukaryotic hosts. The transition to strict host-dependence has repeatedly been accompanied by an increase in the frequency of mobile elements. This may contribute to the rapid loss of functional genes from the bacterial genome both directly — because of the disruption of previously functional genes by the insertion of mobile elements — and indirectly — because of an increased frequency of large deletions resulting from homologous recombination among the multiple elements inserted at different genomic locations. Selection to maintain many gene functions is less effective in host-restricted bacteria because the new environment is less demanding and because the effective population size of the bacteria is greatly reduced. This means that a mobile element insertion is more likely to be fixed, even though the gene it disrupts has some useful function.

Insights from evolutionary comparisons

Inferences about evolutionary processes benefit greatly from being able to make comparisons among different organisms, especially when their underlying phylogeny is understood. Such comparisons provide important information about the order in which characters were acquired — if a feature is present in two organisms it was probably present in their common ancestor — and about the existence of alternative solutions to similar problems.

Sex determination is a conserved process, but the switch that directs development down the male or female path is not. **Schartl** discusses the evolution of sex determination and sex chromosomes in non-mammalian vertebrates. Our XX/XY system is only one of several used by fishes, reptiles, birds and amphibians. ZW/ZZ chromosomal sex determination (where the female is heterogametic and the male homogametic), systems with multiple sex chromosomes, and temperature and hormonal sex determination represent some of the sex determination processes that are found in non-mammalian vertebrates. Male-determining Y chromosomes have evolved repeatedly, each time recruiting a different gene to be the sex-determining switch. The study of fish sex-chromosomes promises to provide an important perspective on stability and plasticity of sex determination.

The deepest branch in the phylogenetic tree of extant mammals separates the egg-laying monotremes from live-bearing mammals. **Grützner** and **Graves** review the limited data available about monotreme genomes. Among other unusual features is the existence of a meiotic translocation chain of sex chromosomes. Questions arise concerning the origin of dosage compensation and parental imprinting, which might be addressed in this small group of mammals. Monotreme genomics is still in its infancy, but expect this to change quickly as soon as large-scale sequencing gets underway. These are exciting times. Now, the genome of almost any species can be sequenced (given the will and a cooperative funding agency). If the platypus genome is sequenced as planned, we may soon have an entire monotreme genome sequence and yet still know very little about many fundamental features of monotreme biology.

The use of chromosome-specific hybridization probes ('chromosome paints') has been used in clinical practice to identify the origin of chromosome fragments involved in rearrangements. But paints from one species can also be used to recognize homologous sequences in another species and identify the small subset of chromosome rearrangements that have become fixed differences between species. **Wienberg** discusses recent use of chromosome painting to reconstruct the evolutionary history of eutherian chromosomes. The method detects major interchromosomal events (fusions, fissions or reciprocal translocations) but does not detect inversions (unless the inversion fortuitously overlaps the boundary of a previous translocation) or[l2] small translocations. These studies reveal remarkable chromosome stability in most mammalian lineages. Fusions or fissions of entire chromosomes appear to have been fixed more frequently than have reciprocal translocations. Despite this background of evolutionary conservatism, some lineages (including mice and rats) have undergone major 'reshuffling' of chromosomes.

The success of comparative genomics has provided ample justification for large-scale sequencing projects, for example, providing important insights into gene function. Can

it be used to tell us how humans differ from their closest relatives? **Ruvolo** discusses progress in sequencing and analysing the chimpanzee genome. The first large chunk of chimp sequence is out and the rest is on the way, and expression analysis is in full swing. So far, no 'IQ gene' has emerged — in fact, brains shows fewer interspecific differences than liver both in number of genes differentially expressed and in the magnitude of the differences.

Harding and **McVean** make good use of comparisons to chimpanzees in an attempt to understand the ancestral population of modern humans. With the accumulation of genotyping data concerning humans, their commensals and chimps, the evolutionary history of our species becomes better refined. Was there a bottleneck followed by a population expansion from a founder group that continues today? Current evidence appears to be more consistent with a rather complex ancestral population structure for the species.

Speciation can be viewed as the process by which a single genome becomes two genomes with different evolutionary histories. What underlies the origin of species? A direct experimental attack on this question would seem to require a 100 000 year grant to observe the emergence of post-zygotic reproductive isolation. Fortunately, geneticists have made important progress in identifying genes that are responsible for sterility and inviability that characterize species hybrids (see the paper by **Orr**, **Masly** and **Presgraves**). Remarkably, the first few speciation genes to be identified show clear evidence for positive Darwinian selection. The next step is to understand how the biology of speciation genes results in reduced hybrid fitness.

Epigenetics: mystery giving way to mechanism

As developmental programs unfold, cellular changes are inherited despite the fact that the DNA sequences remain the same. In most complex organisms, DNA methylation provides one mechanism for epigenetic inheritance, although most developmental differences are maintained at the level of chromatin. Understanding how these differences are established and maintained has made epigenetics one of the most exciting areas in biology over the past few years.

The biggest recent surprise is evidence for the involvement of small double-stranded RNA in the establishment of epigenetic states. RNA interference is now well-established as a general post-transcriptional regulatory mechanism, and the elucidation of this phenomenon has been in the forefront of eukaryotic biology, both because of its biological importance and its utility in gene silencing. Very recent work reviewed by **Hodgetts** suggests that a similar as-yet mysterious process occurs in the nucleus whereby small interfering RNAs target heterochromatin formation. This new paradigm for developmental regulation continues to reveal how ignorant we have been, and still are, of basic genetic mechanisms in the 'post-genomic' era.

Many epigenetic phenomena, such as X chromosome inactivation, involve DNA methylation. Interest in DNA methylation is further intensified by evidence that it is responsible for much of what goes awry in human ills, such as silencing of tumor-suppressor genes. However, favorite model systems for studying other basic genetic processes — including flies, worms and yeast — lack DNA methylation. Fortunately, *Arabidopsis* has emerged as a powerful model for studying DNA methylation. **Rangwala** and **Richards** describe recent progress, both in understanding the role of methylation in biological processes and the mechanism whereby some cytosines are methylated and others are not. Recent evidence suggests that small double-stranded RNAs trigger DNA methylation, in some cases via a histone modification.

During somatic development, programmed changes of epigenetic state provide a useful set of switches that allow genetically identical cells to acquire differentiated functions and allow facultative responses of a genotype to environmental changes. **Chong** and **Whitelaw** review recent evidence that epigenetic changes may also be transmitted from one generation to the next via gametes. This raises the interesting possibility of transgenerational 'differentiation' of phenotypes, allowing organisms to make adaptive responses to information about past environments as well as about the current environment.

Progress in epigenetics research will depend in large part on technological advances to move the field into 'epigenomics', in the same way that DNA sequencing technology enabled genetics to develop into genomics. Among the most desirable epigenomic technologies are those that can suggest how genes are regulated. DNA microarrays are especially promising, because they provide a large-scale platform for mapping potential gene regulatory interactions. **Hanlon** and **Lieb** report on recent advances in mapping protein-binding to DNA by combining chromatin immunoprecipitation with DNA microarrays ('ChIP-chip'). Results so far are exciting but the field still suffers from difficulties because chromatin is complex and inherently inhomogeneous and because comparing results from different arrays is often not straightforward. Despite these challenges, it is clear that this and other genomics technologies will greatly facilitate research into epigenetic mechanisms and developmental programs.

A *palimpsest* is a parchment on which one text has been written over another, partially erased, text. Sometimes, the earliest extant copy of a significant document can be

found written beneath a more recent text. For example, the earliest known copy of a treatise by Archimedes (copied in the 10th century) is found hidden beneath a 12th century Greek liturgical text [1]. Similarly, genomes have multiple layers of text with significant information about the evolutionary past, if only the tools can be developed to read the message. Prospects look bright. The exponential accumulation of sequence data increases the power of comparative analyses both between and within species. Meanwhile, new genomic technologies parlay sequence information, enabling researchers to gain a deeper understanding of basic biological mechanisms.

References

1. Wilson NG: **Archimedes: the palimpsest and the tradition**. *Byzantinische Zeitschrift* 1999, **92**:89-101.