Genomes and evolution
Population genetics and molecular evolution of whole genomes
Editorial overview

Andrew G Clark and Sarah P Otto

Current Opinion in Genetics & Development 2002, 12:631–633

0959-437X/02/$ – see front matter

Andrew G Clark

Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York, USA; e-mail: ac347@cornell.edu

Andrew Clark's laboratory works on molecular population genetics in *Drosophila* and in humans. The *Drosophila* systems include an analysis of genetic variation responsible for variation and interspecific divergence in sperm competition, bacterial resistance, and in flight metabolism. His lab has also been instrumental in the discovery and analysis of genes on the *Drosophila* Y chromosome. His work in human genetics centers on methods for testing association between genotypic and phenotypic variation in complex traits, with an emphasis on cardiovascular disease.

Sarah P Otto

Department of Zoology, University of British Columbia, Vancouver BC V6T 1Z4, Canada; e-mail: otto@zoology.ubc.ca

Sarah Otto is an Associate Professor in the Department of Zoology at the University of British Columbia. Her research focuses on the evolutionary forces that act on genomic structure. Using theoretical models and experiments in yeast, she has investigated the evolution of chromosome number, the evolution of recombination rates and the influence of genomic structure on rates of adaptation. She is currently an NSERC Steacie Fellow.

Introduction

This issue of *Current Opinion in Genetics & Development* is split into two major themes. The first is an overview of genomic landscapes, focusing on the evolutionary processes that shape the patterns observed in various genomic features, such as codon usage and gene density. The second is an overview of the evolution of functional sequences.

Genomic landscapes

The great bulk of eukaryotic genomes is noncoding, and it is of considerable interest to learn what, if anything, this DNA is doing. Evolutionary comparisons both within and among species provide a powerful test of the simplest hypothesis: that noncoding DNA is mutating at random and is subject to drift alone. Ludwig (pp 634–639) provides a clear overview of this area and marshals the evidence that noncoding sequences face a number of evolutionary constraints that have an effect on the changes observed over time. Similarly, Duret (pp 640–649) reviews recent studies indicating that weak selective forces do shape codon usage bias — the preferential use of certain codons to encode an amino acid — although the evidence for this is clearer for invertebrates than vertebrates. Duret is, however, cautious in concluding that selection is responsible for codon bias as most of the studies are based on simple models assuming homogeneous mutation rates and unbiased gene conversion, simplifications that are known to be invalid. Indeed, as Duret points out, mutation rates vary substantially depending on the level of gene expression and on the base composition at flanking sites and within the genomic region.

To the extent that mutations are by-products of cell division, there should also be differences between mutation rates in males and females. This is particularly true for mammals, where the formation of sperm and the formation of eggs are profoundly different cellular processes, because eggs are arrested at meiosis I for most of a female's life whereas sperm undergo continuous mitotic division. By comparing rates and patterns of evolutionary change on the X chromosome, which spends two-thirds of its time in females, to the Y chromosome and to autosomes, investigators have attempted to estimate the relative rates of mutation in males and females. Li, Yi and Makova (pp 650–656) review this topic, bringing us up to date on the latest results garnered from analyses of large genomic regions. They conclude that there is strong support for an elevated mutation rate in males of primates, rodents and birds.

Just as heterogeneity in mutation rates becomes obvious upon close inspection, extensive variation in recombination rates has also been found. It has long been recognized that the rate of recombination per unit physical length, expressed in centiMorgans per megabase pair, varies by at least a factor of ten within the genomes of both humans and *Drosophila*. The paper by Nachman (pp 657–663) underscores this heterogeneity and reviews recent studies that document variation in recombination at a much finer level of resolution.

Nachman also describes the impact of this variation on patterns of evolutionary change. Selection is expected to be less efficient at removing deleterious mutations in genomic regions with very low rates of recombination, and consequently, a wide variety of mutations and transposable elements are expected to accumulate in these regions. This expectation is taken to an extreme on the Y chromosome, which is largely nonrecombining. The classic model for the evolution of the Y chromosome hypothesizes that it originated from an X chromosome, which then evolved lower recombination rates, loss of gene function, and dosage compensation, retaining only the few genes necessary for male fertility. Genomic analyses are, however, revealing that there is much more to Y chromosome evolution, as reviewed by Carvalho (pp 664–668). Y chromosomes of humans and flies appear to have undergone a series of translocations from various autosomes, which has served as an ongoing source of functional sex-linked genes. Indeed, this transfer of DNA from autosomes may even account for the origin of the Y chromosome in some species. As Carvalho argues, the Y chromosome of *Drosophila melanogaster* shows no trace of homology with the X and may have evolved instead from a supernumerary or B chromosome.

Transposable elements are a ubiquitous feature of most eukaryotic genomes and determining their patterns of abundance and age can be very illuminating. Eickbush and Furano (pp 669–674) describe the vast difference, in distribution and natural history, of retrotransposable elements between human and Drosophila genomes. There are over a thousand times more retrotransposable elements in human genomes than in *Drosophila*, and human genomes also contain half a million copies of SINE elements, which are completely lacking in *Drosophila*. Interestingly, retrotransposable elements are also much less polymorphic in humans than in flies. Thus we can infer that retrotransposons are more easily fixed in mammalian populations than in fly populations, which suggests that they are better tolerated in mammals. A possible explanation for such a difference in tolerance is that the rate of ectopic homologous recombination is much lower in humans. Hard proof that ectopic exchange is the major source of selection against retrotransposons in flies is not available, but the contrast between the patterns of retrotransposon abundance and age in flies and mammals is truly striking and provides tantalizing evidence for the ectopic exchange model.

Patterns of molecular diversity across genomic landscapes provide the opportunity to make inferences about a population's evolutionary past. Excoffier (pp 675–682) reviews the growing literature that has attempted to distinguish between two major models of human demographic history. The first model ('Multiregional Evolution') postulates that today's humans have descended from populations throughout the world that were in place hundreds of thousands of years ago. The second model ('Recent African Origin') postulates that we are all descended from relatively small numbers of Africans, some of whom migrated out of Africa within the last 200,000 years. To distinguish between these two possibilities, researchers have searched for clues in the levels of molecular variation and linkage disequilibria observed among human genomes sampled from throughout the world. Excoffier concludes that the evidence leans towards a Recent African Origin of humans, although he notes that more sophisticated alternatives must be considered to incorporate population sub-division, the possibility of selection, and migration back to Africa.

The evolution of functional sequences

In addition to information gathered about genomic landscapes, there is much interest in understanding the evolution and genomic organization of functional genes. For decades now, population geneticists have been devising tests to identify genes that show a signature of natural selection. A popular method, suggested by McDonald and Kreitman [1], compares levels of variability within species to differences between species for silent and replacement changes. Previous applications of this test have been limited to single genes, but recent analyses have scaled the comparisons up to the genomic level. As reviewed by Schlötterer (pp 683–687), these more powerful studies lead to the conclusion that amino acid changes are much more frequently driven by selection than was previously thought. Schlötterer then points out that genome-wide analyses will be even more useful for discovering how local populations adapt to their environment. This idea can be traced back to Lewontin and Krakauer [2], who pointed out that, if there are environmental conditions that select for locally-adapted alleles in some populations, then the affected genes should exhibit greater among-population differentiation in allele frequencies than non-affected genes. Schlötterer describes a number of ways that this idea can be tested and reviews the few studies that have examined local adaptation using genomic data. Such studies provide a valuable filtering mechanism for identifying gene regions that are putative targets of local adaptation. In the following article, Yang (pp 688–694) provides an excellent introduction to methods that infer the action of natural selection by fitting evolutionary models to sequence data from multiple species. The real strength of Yang's modeling approach is that one can pinpoint which sites and which taxa show evidence of selection. These same tools are readily scaled up to genomic comparisons, provided that one can properly identify and align orthologous gene sets across the species in question. Combining genome-wide data with these powerful new methods promises to reveal many more of the signatures of adaptive evolutionary changes that have been left in molecular sequence data.

Genomes have complex pathways of regulation in order to ensure the appropriate timing and tissue specificity of gene expression during development. These mechanisms of regulation are themselves products of evolution, and Gibson and Honeycutt (pp 695–700) tackle the question of how one can make inferences about this. Because the inference of regulatory pathways requires considerable lab

bench work, our knowledge of the complete set of regulatory rules for any organism lags far behind our inventories of genes. Nevertheless, there are clear examples from comparative studies indicating that whole branches of regulatory pathways have been either lost or gained between one species and another. Gibson and Honeycutt also review the challenges of using bioinformatic methods to identify the targets of regulatory pathways and provide an entrée into the literature on network theory as applied to gene regulatory pathways.

Most genes in plants and animals are broken into exons and introns. The observation that tissue-specific alternative splicing occurs suggests that a key evolutionary advantage of introns lies in the finer degree of regulation and wider array of proteins produced per gene that intron splicing and transcript processing allow. Lynch and Richardson (pp 701–710) begin their review of the evolution of spliceosomal introns by pointing out the rather serious cost that the presence of introns imposes. The presence of introns increases the risk that a gene will be mis-expressed and almost certainly increases the number of mutations that lead to loss of function. Despite this, spliceosomal introns are ubiquitous, and Lynch and Richardson review the mechanisms that may have led to selection for the proliferation of spliceosomal introns.

A fundamental process that organisms must carry out is the fair distribution of genetic material during cell division. Any selfish genetic element that can seize control of this process and become over-represented among surviving daughter cells can readily take over within a population. Thus the very stability of Mendelian rules relies on the evolution of features of chromosomes that guarantee their proper segregation. Malik and Henikoff have published a series of intriguing papers on the evolutionary changes in the centromere-specific histone CenH3 (*Cid* in *Drosophila*). Here (pp 711–718), they review evidence for a delicate co-evolutionary interplay between centromere-binding proteins and centromeric DNA. Because of the serious distortions in chromosome behavior that occur if centromeric histones and the DNA sequences they bind are not compatible, Malik and Henikoff point out that divergence over time in these sequences could form a genetic basis for reproductive isolation between species.

Population genetics has traditionally focused on point mutations as the source of genetic variation upon which selection acts. Now that >200 bacterial genomes have been fully sequenced, however, it is clear that changes in gene content are a key source of variation in fitness, as reviewed by Whittam and Bumbaugh (pp 719–725). Many of these changes in gene content represent lateral gene transfer events, and related strains vary by as much as 20% in gene content. The transferred genes are frequently critical to pathogenesis and often implicate toxin-bearing phage as the agents of transfer. Interestingly, evidence has also accumulated that genomic segments can be rapidly and repeatedly lost from different strains of a bacterial species. These comparative analyses have painted an extremely dynamic picture of how bacterial genomes evolve.

This issue of *Current Opinion in Genetics & Development* is brought to a close with a discussion of the costs of genomic sequencing projects. These costs have been borne by funding agencies and companies largely because they have promised that genomic information will be of value in understanding, curing and preventing human diseases. A subsequent project with the same promised benefits has recently been launched to study variation among human genomes: the so-called haplotype mapping project (or 'HapMap'). This represents by far the largest and most expensive experiment in population genetics to date. It is an attempt to quantify genome-wide patterns of genetic variation and linkage disequilibrium in the hope of mapping genes underlying complex traits. Both the allele frequency spectrum and the patterns of linkage disequilibrium reflect past evolutionary events, including selection associated with complex diseases, and so can be used to infer sites of selection. Terwilliger *et al.* (pp 726–734) examine the issues and come to the pointed conclusion that this approach to finding genes for complex disease is not as powerful as linkage methods that already exist. Part of the problem with this issue is that the genetic basis of complex diseases may vary enormously from one disease to another. Linkage methods do work better than association methods if there are relatively few genes of large effect that recurrently mutate, generating a variety of disease-causing alleles. At the other extreme, many diseases may be so complex that there is little overlap in etiology from one case to another. Finding the genes responsible for such diseases is beyond the scope of any scientific method that relies on replicable results. While the HapMap project will provide a huge wealth of data, making the best use of this data to understand complex diseases and, more generally, our evolutionary history remains a challenge that will undoubtedly spawn more discoveries in this burgeoning field, which may be aptly named 'population genomics'.

Finally, as the genomics community is recognizing the power of comparative analysis and is laying plans for complete genomic sequencing of numerous non-model organisms, the opportunity to describe genome-wide variation within species will expand enormously, allowing population genomics to extend its vision beyond the rather narrow focus on humans and flies.

Acknowledgement

We would like to thank Mike Lynch for his help in planning and reviewing the articles for this issue.

References

1. McDonald JH, Kreitman M: Adaptive protein evolution at the ***Adh*** locus in ***Drosophila***. *Nature* 1991, 351:652-654.

2. Lewontin RC, Krakauer J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 1973, 74 :175-195.