ELSEVIER

Start-up entities in the origin of new genes

Vincent Daubin* and Howard Ochman

The remarkable diversity in the contents of genomes raises questions about how new genes and new functions originate. Recent evidence indicates that parasitism, particularly the molecular interactions between phage and their bacterial hosts, is a likely mechanism for generating new genes. This invention of such novel functions seems to be founded on a strategy that secures the short-term survival of parasitic elements and thereby contributes to the renovation of gene repertoires in their host.

Addresses
Department of Biochemistry & Molecular Biophysics, University of Arizona, Tucson, Arizona, 85721, USA
*e-mail: daubin@email.arizona.edu

Current Opinion in Genetics & Development 2004, **14**:616–619

This review comes from a themed issue on
Genomes and evolution
Edited by David Haig and Steve Henikoff

Available online 3rd October 2004

0959-437X/$ – see front matter

DOI 10.1016/j.gde.2004.09.004

Introduction

The genomes of cellular organisms typically contain thousands of genes, but at most 100 are common to all life forms. Even within a bacterial species such as *Escherichia coli*, the differences in gene content among strains can be enormous [1]. This implies that the renovation of gene repertoires through the invention of new genes may have played a larger role than the modification of ancestral sequences in generating evolutionary novelty.

The processes by which new genes are produced are not fully known; however, the most broadly held view is that regions of the genome are duplicated and that subsequent functional diversification can produce genes conferring novel properties [2]. Because the invention of useful genes through the modification of existing sequences is a slow and tentative process, many organisms have drawn on an alternative means — namely, the enlistment of established traits from unrelated organisms. There are obvious advantages to this strategy in that the genes from other organisms have been already refined by selection and the benefits can be instantaneous. Numerous lineages have successfully exploited novel or previously unsuitable environments by such gene-acquisition events, ranging from those leading to antibiotic resistance [3,4] and thermophily [5] in some bacterial species, to photosynthesis [6–8,9•] and aerobiosis [10] in eukaryotes.

Here we provide evidence that parasitism, particularly the molecular interactions between phage and their bacterial hosts, is a likely mechanism for generating new genes. In particular, mobile sequences, such as phages in bacteria and transposable elements in eukaryotes, might be actively devising novel traits for their host organisms.

Co-opting genes from selfish elements

The alien genes that can be adopted most easily by an organism are often those of parasitic or selfish elements that are already present in the genome of that organism. For example, in *Pseudomonas aeruginosa*, genes encoding the tail of two different bacteriophages have been converted into bacteriocins used by the bacteria to kill its competitors [11]. In eukaryotes, it also seems that transposable elements have served as substrates for new genes: in *Drosophila*, the extension of chromosome ends involves proteins similar to those encoded by two long interdispersed element (LINE)-like retrotransposons, TART (telomere-associated retrotransposon) and HeT-A, suggesting that, in the fly lineage, genes from transposable elements have assumed the function usually achieved by telomerases [12,13•].

Even more striking is the nucleus-encoded mitochondrial RNA polymerase of eukaryotes, which has a high degree of similarity to the RNA polymerase of bacteriophage T3 and T7 [14]. Apparently, this gene was recruited by an ancestral eukaryote for mitochondrial transcription, and a duplicated form functions in transcription in chloroplasts [15]. In addition, similar events in which viral genes have displaced host genes have been proposed for the replicative helicase DnaB and the DNA polymerase γ of mitochondria (reviewed in [16••]). Recently, Mallet *et al.* [17••] have reported that a gene from a human endogenous retrovirus, which is restricted to the hominoid lineage has taken on an active role in formation of the human placenta, possibly by favoring cell fusion.

In the above examples, the recruited genes maintain significant similarity to genes that are present within selfish elements, but this need not be the case. In vertebrates, one of the mechanisms responsible for generating new antigen-binding domains of immunoglobulins is reminiscent of that of transposases within DNA transposons, suggesting that this system was recruited from a selfish element in the ancestor of vertebrates [18].

Despite the operational similarities, the genes controlling this activity in vertebrates, *RAG1* and *RAG2*, are not homologous to known transposons or to any other genes for that matter.

Parasites as patrons

Novel traits might sometimes arise from genes that originally functioned in a parasite or symbiont [16••], but the number and scope of such traits, particularly those encoded by bacterial accessory elements, are somewhat limited. However, features of these elements enable them to capture and to disseminate genes from previous hosts, and in most bacterial genomes there is evidence that plasmids, phage or transposons have been responsible for bringing in genes from other organisms [19]. Thus, despite their relatively limited coding potential, mobile elements offer a vast repertoire of user-ready genes that can potentially benefit their hosts.

Perhaps most notable among the sequences that have been transferred by such elements are microbial pathogenicity islands, which are chromosomally encoded clusters of virulence genes that are absent from the genomes of related non-pathogenic strains and species. Such islands have been detected in many bacterial genomes, and it is now clear that they have played a substantial role in the adaptation of bacteria to new environments and symbiotic interactions (reviewed in [19]). Features of these islands, including their atypical base compositions, their occurrence at known phage integration sites, and/or the presence of characteristic repeat motifs, suggest that they have been introduced into the chromosome by mobile elements. Furthermore, the bacteriophage or plasmids bearing these genes can be sometimes transmitted between organisms [19].

Although such events of lateral gene transfer are commonly viewed as exchanges between bacteria, most involve some parasitic intermediate that serves as the vehicle for spreading genes among organisms. Thus, novel genes recruited in this manner will not only benefit the recipient organism but also assure the survival of the parasitic donor. Thus, by providing useful genes, the inevitable conflict between parasites and hosts can be converted into a mutually beneficial symbiosis, in which the parasite and host coexist.

Elaborating genetic novelty

Maintaining this molecular symbiosis requires the parasitic elements to safeguard their existence by supplying genes that confer an advantage to their host. It seems that the selfish elements themselves may be devising new genes that are only of use to their host. Because viruses and phage typically have high rates of evolution, nonhomologous recombination and gene exchange, they possess the tools necessary for quickly creating and disseminating novel sequences [20,21••,22].

This mechanism seems to be a very powerful means for innovation and may have contributed significantly to the diversity among bacterial genomes. Although the ancestries of many genes within a genome can be established by their similarity to previously characterized sequences, there exists a large class of genes, constituting up to 20% of the genes in a genome, for which there are no homologs in any other organism. These 'ORFan' genes are unique and represent the only members of their protein families; they are even found in species for which genome sequences are available from close relatives, suggesting that they can originate very quickly. For example, the many sequenced strains of *E. coli* diverged relatively recently, but each contains a unique set of ORFan genes that are not present in their closely related strains [1].

These ORFans display many characteristics of phage genes, namely, they are short, (A+T)-rich and quickly evolving, and occur in clusters near phage integration sites [23,24,25•]. In addition, their lack of homologs probably reflects the vast amounts of unexplored phage diversity [21••,26]. A *de novo* origin of ORFans in phage implies that some fraction of the new and potentially useful genes in bacteria are being created by noncellular organisms; in this regard, one might view bacteriophages as start-up entities whose existence is based on creating an innovation that has been overlooked by other organisms.

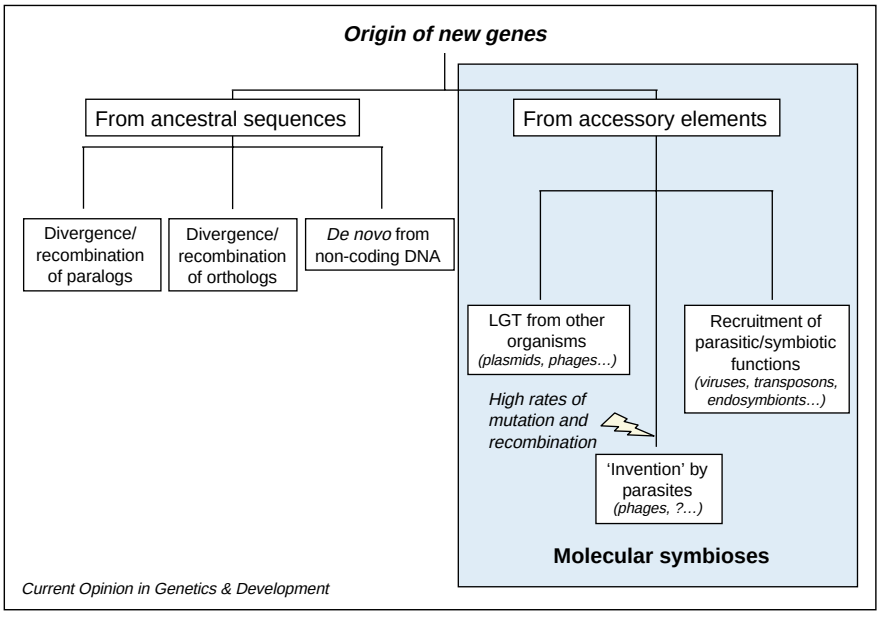
Many ORFans in the *E. coli* genome are functional, and some are already known to be involved in key cellular functions such as translation and replication [25•]. Although it is premature to generalize these observations to eukaryotes, the ORFan genes of *Drosophila* are also short, (A+T)-rich and quickly evolving [27••], suggesting their possible link to transposable elements, which are known to be (A+T)-biased in most eukaryotes [28]. Thus, several mechanisms may account for the generation of novel genes in cellular organisms (Figure 1). Although evolution from ancestral sequences probably remains the main contributor of new genes in eukaryotes, accessory elements may have had a significant role in the evolution of new genes in prokaryotes.

The fate of start-up entities

Although ORFans are abundant and harbor numerous features of phage genes, there is scant evidence of the actual elements that brought these genes into existence or into the genome. If the invention of new genes is founded on a strategy that secures survival of selfish elements, why are the ORFan genes in bacterial genomes so rarely associated with identifiable phage sequences?

Bacteria possess a mechanism whereby sequences under relaxed selective constraints are gradually eroded and eliminated from the genome [29,30]. This process of deletional bias has contributed to the massive reduction

Figure 1



Mechanisms of the origin of new genes. Eukaryotic genome repertoires are usually thought to evolve from ancestral sequences that are already present in the genome (left), although contributions by accessory elements (right), such as transposons, have been reported (see text). In prokaryotes, selfish elements (right), and particularly bacteriophages, are likely to have played a much more significant role. Abbreviation: LGT, lateral gene transfer.

in the genome size of endosymbiotic and parasitic bacteria, which derived from free-living bacteria with much larger gene inventories [31,32].

In eukaryotes, the ancestral α-Proteobacteria, which was recruited for aerobiosis, is now reduced to a dedicated organelle [10]. Similarly, in bacteria, accessory elements can be eliminated once their few beneficial genes have been integrated into the host genome. By devising traits with latent potential to bacteria, phage can initially promote their own survival in host genomes, although eventually only the beneficial genes, and not those of phage, are retained. Borrowing an analogy from the corporate world again, phage might be viewed as start-up companies that are initially very successful, but ultimately disappear or become subsumed.

Conclusions

The manner in which organisms develop new traits has been the subject of continued speculation and exploration over the past few decades. Full genome sequences can now provide comprehensive information about gene repertoires, but comparative analyses are necessary to elucidate the origins, evolution and dynamics of genome contents. The finding that most genomes contain sequences with no homologs, even in close relatives, has prompted a search for both diversity in previously unexplored areas and the ways in which these sequences can originate.

On the basis of their features, mobile elements seem to be among the most dynamic and inventive gene resources for cellular genomes. Current efforts to sequence phage genomes should aid our understanding of the mechanisms underlying genetic innovation, and we predict that these sequences will provide insight into the origin of ORFans in bacterial genomes.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Welch RA, Burland V, Plunkett G III, Redford P, Roesch P, Rasko D, Buckles EL, Liou SR, Boutin A, Hackett J *et al.*: **Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli***. *Proc Natl Acad Sci USA* 2002, **99**:17020-17024.

2. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes**. *Science* 2000, **290**:1151-1155.

3. Carattoli A, Filetici E, Villa L, Dionisi AM, Ricci A, Luzzi I: **Antibiotic resistance genes and *Salmonella* genomic island 1 in *Salmonella enterica* serovar *Typhimurium* isolated in Italy**. *Antimicrob Agents Chemother* 2002, **46**:2821-2828.

4. Deng W, Liou SR, Plunkett G III, Mayhew GF, Rose DJ, Burland V, Kodoyianni V, Schwartz DC, Blattner FR: **Comparative genomics of *Salmonella enterica* serovar *Typhi* strains Ty2 and CT18**. *J Bacteriol* 2003, **185**:2330-2337.

5. Forterre P, Bouthier De La Tour C, Philippe H, Duguet M: **Reverse gyrase from hyperthermophiles: probable transfer of a thermoadaptation trait from archaea to bacteria**. *Trends Genet* 2000, **16**:152-154.

6. Moreira D, Le Guyader H, Philippe H: **The origin of red algae and the evolution of chloroplasts**. *Nature* 2000, **405**:69-72.

7. Raven JA, Allen JF: **Genomics and chloroplast evolution: what did cyanobacteria do for plants?** *Genome Biol* 2003, **4**:209.

8. Douglas SE: **Plastid evolution: origins, diversity, trends**. *Curr Opin Genet Dev* 1998, **8**:655-661.

9. • Archibald JM, Rogers MB, Toop M, Ishida K, Keeling PJ: **Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans***. *Proc Natl Acad Sci USA* 2003, **100**:7678-7683.

A study of the gene transfers that accompanied the secondary endosymbiosis of a photosynthetic eukaryote in the alga *Bigelowiella natans*.

10. Andersson SG, Kurland CG: **Origins of mitochondria and hydrogenosomes**. *Curr Opin Microbiol* 1999, **2**:535-541.

11. Nakayama K, Takashima K, Ishihara H, Shinomiya T, Kageyama M, Kanaya S, Ohnishi M, Murata T, Mori H, Hayashi T: **The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to λ phage**. *Mol Microbiol* 2000, **38**:213-231.

12. Pardue ML, DeBaryshe PG: **Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres**. *Annu Rev Genet* 2003, **37**:485-511.

13. • Abad JP, De Pablos B, Osoegawa K, De Jong PJ, Martin-Gallardo A, Villasante A: **TAHRE, a novel telomeric retrotransposon from *Drosophila melanogaster*, reveals the origin of *Drosophila* telomeres**. *Mol Biol Evol* 2004, **21**:1620-1624.

The description of TAHRE, a new retrotransposon related to TART (telomere-associated retrotransposon) and HeT-A, that highlights the evolutionary origin of telomeres.

14. Cermakian N, Ikeda TM, Cedergren R, Gray MW: **Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage**. *Nucleic Acids Res* 1996, **24**:648-654.

15. Hedtke B, Borner T, Weihe A: **Mitochondrial and chloroplast phage-type RNA polymerases in *Arabidopsis***. *Science* 1997, **277**:809-811.

16. •• Filee J, Forterre P, Laurent J: **The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies**. *Res Microbiol* 2003, **154**:237-243.

This paper reviews how viruses seem to have contributed to the replication and translation machinery of cellular organisms. An atypical and exciting view of the evolutionary relationships between parasites and their hosts is presented.

17. •• Mallet F, Bouton O, Prudhomme S, Cheynet V, Oriol G, Bonnaud B, Lucotte G, Duret L, Mandrand B: **The endogenous retroviral locus *ERVWE1* is a *bona fide* gene involved in hominoid placental physiology**. *Proc Natl Acad Sci USA* 2004, **101**:1731-1736.

A striking example of how the human endogenous retrovirus HERV has been recruited to the formation of the placenta during primate evolution. An evolutionary and functional analysis demonstrates that the retrovirus-derived gene and its function have been conserved in primates.

18. Agrawal A, Eastman QM, Schatz DG: **Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system**. *Nature* 1998, **394**:744-751.

19. Dobrindt U, Hochhut B, Hentschel U, Hacker J: **Genomic islands in pathogenic and environmental microorganisms**. *Nat Rev Microbiol* 2004, **2**:414-424.

20. Hendrix RW: **Bacteriophages: evolution of the majority**. *Theor Popul Biol* 2002, **61**:471-480.

21. •• Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, Jacobs-Sera D, Falbo J, Gross J, Pannunzio NR *et al.*: **Origins of highly mosaic mycobacteriophage genomes**. *Cell* 2003, **113**:171-182.

A glimpse at the overwhelming diversity of bacteriophage genomes.

22. Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML, Brüssow H: **Phage as agents of lateral gene transfer**. *Curr Opin Microbiol* 2003, **6**:417-424.

23. Daubin V, Lerat E, Perriere G: **The source of laterally transferred genes in bacterial genomes**. *Genome Biol* 2003, **4**:R57.

24. Charlebois RL, Clarke GD, Beiko RG, St Jean A: **Characterization of species-specific genes using a flexible, web-based querying system**. *FEMS Microbiol Lett* 2003, **225**:213-220.

25. • Daubin V, Ochman H: **Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli***. *Genome Res* 2004, **14**:1036-1042.

An analysis of ORFan genes in the *E. coli* genome that suggests that most of these genes are derived from bacteriophages and that some have assumed key functions in the bacteria.

26. Rohwer F: **Global phage diversity**. *Cell* 2003, **113**:141.

27. •• Domazet-Loso T, Tautz D: **An evolutionary analysis of orphan genes in *Drosophila***. *Genome Res* 2003, **13**:2213-2219.

This paper analyses the genes that seem to be restricted to the fly lineage and offers a complete analysis of their sequence characteristics and evolutionary rates. The authors propose a model to explain the origin of ORFans that invokes the duplication of ancestral genes and the acceleration of evolutionary rates in one of the resulting paralogs.

28. Lerat E, Capy P, Biemont C: **Codon usage by transposable elements and their host genes in five species**. *J Mol Evol* 2002, **54**:625-637.

29. Mira A, Ochman H, Moran NA: **Deletional bias and the evolution of bacterial genomes**. *Trends Genet* 2001, **17**:589-596.

30. Lawrence JG, Hendrix RW, Casjens S: **Where are the pseudogenes in bacterial genomes?** *Trends Microbiol* 2001, **9**:535-540.

31. Moran NA, Wernegreen JJ: **Lifestyle evolution in symbiotic bacteria: insights from genomics**. *Trends Ecol Evol* 2000, **15**:321-326.

32. Moran NA, Mira A: **The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola***. *Genome Biol* 2001, **2**: research0054.1-research0054.12.