Constraint and opportunity in genome innovation

James A Shapiro
Department of Biochemistry and Molecular Biology; University of Chicago; Chicago, IL USA

Keywords: evolution, natural genetic engineering, horizontal DNA transfer, mobile genetic elements, viruses

The development of rigorous molecular taxonomy pioneered by Carl Woese has freed evolution science to explore numerous cellular activities that lead to genome change in evolution. These activities include symbiogenesis, inter- and intracellular horizontal DNA transfer, incorporation of DNA from infectious agents, and natural genetic engineering, especially the activity of mobile elements. This article reviews documented examples of all these processes and proposes experiments to extend our understanding of cell-mediated genome change.

Carl Woese was the most important evolutionary scientist of the 20th century. He converted evolution science from a descriptive and highly speculative subject into a field based on clear molecular evidence. In this tribute to Carl, I wish to show some of the ways he and other molecular biologists have opened our eyes to creative evolutionary possibilities unimagined in the pre-molecular vision of the Modern Synthesis merging Darwinism and Mendelian genetics.1,2

Woese's View of Core and Peripheral Cell Systems

Essential to Carl Woese revolutionizing phylogenomics was the recognition of how deeply embedded were the ribosome and associated translation functions into core information transfer functions of all cells. As Carl expressed this idea in a 2004 review:

“rRNA molecules are relatively large, universal in distribution, and constant in function. Importantly, their sequences are highly conserved overall, and, as central components of a complex and essential cellular mechanism, rRNAs arguably would be less subject to the vagaries of reticulate evolution than would other cellular components.”

The integrated nature of the cell translation apparatus made its central organelle, the ribosome, very stable in evolution, and thus, an ideal object for examining the deepest evolutionary relationships at the molecular level.

As this issue of RNA Biology demonstrates, the result of basing cell phylogenies on rRNA molecules was transformational for the life sciences. A whole new and unsuspected kingdom of life was uncovered,4 and core evolutionary relationships acquired a solid empirical basis.

What was true for the translational apparatus, also proved to be the case with other fundamental central features of molecular cell biology. Archaea, Bacteria, and Eukarya have complex, distinct, and conserved systems for DNA replication, transcription, and membrane biogenesis. The observations that the replication,5,6 transcription,7-9 and translation initiation10 systems of eukaryotic cells more closely resemble those of Archaea, while membrane biogenesis11,12 and other features of eukaryotic metabolism are more closely related to those of bacteria,13 provide enticing evidence in favor of the hypothesis that the two prokaryotic kingdoms preceded the origins of eukaryotic cells, which involved one or more fusion events.14-17

Molecular Phylogeny and Symbiogenesis: Evolutionary Innovation by Cell Mergers and DNA Transfers between Organelles and the Nucleus

Although the questions of eukaryotic origins are still actively debated, the molecular evidence for the symbiogenetic origins of the mitochondrion and the chloroplast/plastid are now incontrovertible.18,19 So cell fusions and the generation of cells with multiple genomic compartments in the nucleus and organelles is an established mechanism of genome innovation. When cells fuse, both the highly conserved and more variable segments of the genome contribute to the novel configuration. For example, our cells, and those of virtually all eukaryotes, contain both eukaryotic and bacterial ribosomes.

The genome record shows that endosymbiosis and symbiogenetic fusions are not extraordinary events and have occurred repeatedly.20 Both green and red algae have been involved in secondary and higher level fusions. The resulting photosynthetic cells (or their non-photosynthetic descendants) have at least four genome compartments: nucleus, mitochondrion, plastid, and nucleomorph (the former nucleus of the algal cell). The important question of how cell and organelle reproduction cycles become synchronized in symbiogenetic fusions remains an important subject for future research. Disruption of this synchronization in rapidly proliferating cancer cells may contribute to the Warburg effect through loss of mitochondrial function.21,22

Active DNA transfer between genome compartments is a key feature of symbiogenetic fusions and a major source of continuing variation for the resulting organism. Following cell fusions, DNA transfers occur from other genome compartments to the nucleus in all eukaryotic phyla.23-40 These transfers are ongoing.
and observed experimentally in real-time.\textsuperscript{36,40-45} (Less attention has been paid to transfers in the opposite direction.)

In certain cases, organelle--nucleus transfers accompany DNA break repair in both plants and animals.\textsuperscript{35,46-48} The sequence evidence indicates this repair occurs by non-homologous end-joining (NHEJ).\textsuperscript{34,47} Like all cell-mediated DNA changes, organelle--nucleus DNA transfers display non-random patterns,\textsuperscript{49} notably with respect to introns\textsuperscript{44} in regions of open chromatin configuration.\textsuperscript{50} In addition, as is true of other genome changes, stress events activate organelle--nucleus DNA transfers.\textsuperscript{35}

Non-homologous incorporation of organelle DNA provides a mechanism for the generation of genomic novelties, including new coding sequences.\textsuperscript{32,34,50,51} Transfer into introns sometimes results in novel splicing patterns and incorporation of new exons into the mRNA product. This and other "exonization" processes solve an important evolutionary problem, the rapid origination of novel protein domains (http://shapiro.bsd.uchicago.edu/Origin_of_New_Protein_Domains.html). Integrated organelle DNA can have other effects on nuclear genome function. In yeast, for example, mitochondrial DNA inserts serve as sites for activation of DNA replication.\textsuperscript{52} There is particular interest in the role of "numts" (nuclear mitochondrial DNA)\textsuperscript{53} in our own evolutionary history.\textsuperscript{34,54,55}

### Molecular Phylogeny and Horizontal DNA Transfer Encoding Peripheral Systems: Evolutionary Innovation by Accumulation of External Coding Sequences

Woese's insightful distinction between core and peripheral functions led to recognition of widespread horizontal DNA transfer between prokaryotic cells (http://shapiro.bsd.uchicago.edu/ExtraRefs.AntibioticResistanceAndHorizontalTransfer.shtml). This recognition resolved the problem that the phylogenetic trees computed for certain proteins agreed with the rRNA taxonomy while others did not:

"The many protein trees that differ in topology from the rRNA tree also differ in topology from one another, the hallmark of HGT. Moreover, some protein-based trees do exhibit topologies in agreement with that of the universal rRNA tree ... Nearly all of the universal components of translation and transcription do so, as do a small number of other proteins, e.g., HSP-60 ... Cellular components can be roughly classified according to the degree to which it is connected to the rest of the cell. Loosely connected, or modular, elements define one extreme of the spectrum. Such components tend to be largely self-defining in their structure/function, interacting minimally with other elements in the cell, and are, therefore, obvious candidates for horizontal gene displacement by alien homology. At the other extreme are the tightly coupled elements, which have extensive, specific, and constraining physical and chemical ties to others of the cellular componentry and, therefore, could seldom, if ever, be sufficiently mimicked by an alien homolog to be displaced by it. The remarkable difference between the HGT profiles of the aminoacyl-tRNA synthetases and others of the translation componentry is thus explained by the loosely coupled, modular nature of the former and the tightly coupled nature of the latter ..."\textsuperscript{56}

Horizontal DNA transfer between cells is another process for rapid genome innovation and acquisition of essential functions needed in changing ecologies. Recognized since the early 1960s as central to the rapid evolution and dissemination of multiple antibiotic resistance in bacteria,\textsuperscript{57} the general role of horizontal transfer in adaptation of bacteria and archaea to the multifarious ecologies on our planet came to be firmly established by the turn of the 21st century.\textsuperscript{58-66}

Although many "eukaryote chauvinists" wish to adhere to strictly vertical inheritance and believe that horizontal transfer is exclusively a prokaryotic phenomenon, it has proven to be important in the evolution of eukaryotic genomes as well.\textsuperscript{67} For example, diverse plant parasitic nematodes owe their vegan lifestyle to hydrolytic enzymes acquired from bacteria and fungi, which enable them to digest plant materials.\textsuperscript{68,74} Evidently, it proved more efficient to adapt to a new food source by borrowing enzymes from distant taxa rather than evolving them internally from the pre-existing nematode genome. It is noteworthy that each lineage of plant parasitic nematodes acquired these essential functions from different fungi and bacteria. So the horizontal acquisition strategy was used many times.

Chinese workers have recently reported parallel bacteria- and fungus-to-shrimp transfers.\textsuperscript{75} A diversity of prokaryotic donors for similar functions has also been found in eukaryotic microbial parasites;\textsuperscript{76} eukaryotic microbes are prone to acquire DNA across taxonomic barriers from both prokaryotic and eukaryotic donors;\textsuperscript{77-80} there is evidence of extensive horizontal transfers from endosymbiotic bacteria to their animal hosts;\textsuperscript{81-85} and diverse adaptive biochemical pathways in multicellular organisms appear to have originated in bacteria, fungi, and other microbes.\textsuperscript{86-88}

Direct horizontal transfer between multicellular eukaryotes is well-documented for mobile genetic elements.\textsuperscript{89-91} It is harder to find examples of horizontal transfer of DNA that is not intrinsically mobile, but examples have been reported.\textsuperscript{92} They include sequences encoding glyoxylate cycle enzymes in metazoa,\textsuperscript{97} photosynthetic carbon cycles,\textsuperscript{98,99} anti-freeze proteins in fish,\textsuperscript{100} mimicry pattern determinants in butterflies,\textsuperscript{101} and acquisition of diverse expressed functions by a parasitic plant from its host.\textsuperscript{102}

In addition to nuclear sequences, whole organelle genomes are subject to transfer between plants and animals.\textsuperscript{103-106}

Besides interspecific hybridization between closely related species, microbial or arthropod parasites, viruses, and bacterial endosymbionts are assumed to be vectors for DNA transfer between multicellular organisms.\textsuperscript{107-110} Endosymbionts transfer between different host species.\textsuperscript{111-113} Large DNA viruses carry a mixture of DNA sequences from all domains of life, and some can infect both protists and multicellular hosts (http://shapiro.bsd.uchicago.edu/Viral_Composites.html).\textsuperscript{114-117}

Amoebae are common hosts for many of these large DNA viruses and constitute an evolutionary "melting pot,"\textsuperscript{118} where sequences from all domains can be combined and then packaged into delivery particles (http://shapiro.bsd.uchicago.edu/Amoebal_Viruses.html). Some of the hosts for these viruses are phagocytic and therefore likely to acquire sequences from engulfed cells.\textsuperscript{119} These large viruses have satellite "virophages," which can infect cells carrying diverse viral hosts,\textsuperscript{120} and they...
even have their own transposable elements ("transpovirons") specific to the viruses and their virophages.\textsuperscript{121,122} So there appear to be abundant molecular tools available for rearranging the DNA sequences in the evolutionary melting pot.\textsuperscript{123}

Significantly, many bacteria known as vertebrate pathogens also infect amoeba.\textsuperscript{124} \textit{Legionella pneumophila} is an example.\textsuperscript{153-156} \textit{Legionella} is also capable of taking up DNA from its environment.\textsuperscript{127,128} Thus, this normally aquatic bacterium has the cell tropism and DNA transfer capabilities needed to transmit DNA segments across virtually the whole eukaryotic lineage. In addition to \textit{Legionella}, other bacteria infect amoebal protists, such as \textit{Salmonella}, \textit{Mycobacterium}, \textit{Klebsiella}, \textit{Yersinia enterocolitica}, \textit{Pseudomonas aeruginosa}, \textit{Stenotrophomonas cenocepacia}, \textit{Vibrio cholerae}, \textit{Bacillus cereus}, \textit{Enterococcus faecalis}, \textit{Enteropathogenic Escherichia coli} (EPEC), \textit{Enterobacter aerogenes}, \textit{Aeromonas hydrophilia}, and \textit{Neisseria meningitides}.\textsuperscript{126,129-131} There is even evidence of conjugal transfer within amoebae between animal and plant pathogenic bacteria.\textsuperscript{132} In other words, the amoebal melting pot, containing sequences from all three domains of life, has numerous infectious links to more complex eukaryotes.

In addition to providing evolutionary vectors and melting pots, viruses of all kinds (including RNA viruses) insert their genomes into eukaryotic host genomes with surprising frequency.\textsuperscript{133-136} Integration can occur by retroviral integrase functions, sometimes followed by recombination with other viral sequences,\textsuperscript{151} or by NHEJ at DNA breaks.\textsuperscript{152,153} Note that integration events at DNA breaks have the same potential to generate novel sequence configurations as the repair events involving organelle DNA cited previously.\textsuperscript{129}

Not surprisingly, viral functions have been recruited, or "exapted,"\textsuperscript{91,134} for cell biology.\textsuperscript{144,146,155-159} The most extensively investigated case is the role retroviruses have played in the evolution of cell fusions proteins (syncytins) and the placenta, a critical step in mammalian evolution (http://shapiro.bsd.uchicago.edu/Exon_Shuffling.html).\textsuperscript{194,195} This combinatorial process is far more efficient than protein evolution by individual amino acid changes because domain shuffling puts together established functionalities in new arrangements. The existence of shared domains means there must be NGE processes for domain amplifications and rearrangement. A number of these exon shuffling processes involve mobile elements.\textsuperscript{196-205}

\textbf{Origination of novel coding sequences by reverse transcription and sequence fusions}

Reverse transcription of processed and edited RNA molecules generates novel cDNA coding sequences subject to genome integration.\textsuperscript{206-209} The cDNAs can be integrated as independent intron-free coding sequences or inserted into existing genetic loci to generate novel fusion protein determinants.\textsuperscript{210,212} Transposons can also generate novel chimeric coding sequences directly at the DNA level.\textsuperscript{213}

\textbf{Origination of novel exons}

There is no mechanism for the rapid appearance of novel exons in conventional theory—and no mechanism that does not build upon pre-existing coding sequences. Nonetheless, many examples have been documented where segments of mobile element or viral insertions contain the appropriate transcription and splicing signals to encode totally novel exons (http://shapiro.bsd.uchicago.edu/Origin_of_New_Protein_Domains.html).\textsuperscript{214} Since the mobile element content of each lineage is distinct, we can expect different exons and protein domains to appear in different lineages.\textsuperscript{213} This expectation fits with the existence of lineage-specific regulatory proteins and protein families.\textsuperscript{217,218}
Origination of coordinately regulated multilocus networks

Even more complex than producing a single new transcriptional complex is the formation of multi-locus networks coordinated by shared regulatory signals. This latter process is likely to have a vanishingly small probability of success based on independent changes at each locus before any integrated network functionality emerges. However, we know that activation of specific families of mobile elements can result in non-independent insertions at multiple loci, rapidly generating networks that can be coordinately regulated. Genetic analysis tells us that mobile elements have indeed introduced common regulatory elements during network rewiring (http://shapiro.bsd.uchicago.edu/Table5C-1.MobileElementsFoundtoBeExaptedas cis-RegulatoryControlSitesinAnimals.html). Coordination of expression often occurs to activate or inhibit the expression of one another, resulting in the origination of novel biochemical capabilities by exon functions.

We know that distinct stress regimes not only stimulate the actions of NGE operators like SOS mutators and transposons but also stimulate the accumulation of different intracellular second messengers. In E. coli, carbohydrate starvation raises the level of cAMP, while amino acid starvation raises the level of (p)ppGpp. Can these distinct intracellular conditions alter the specificity of genome changes?

The relevant experiments are straightforward. Experimenters can isolate mutant clones of stress-activated cells using various selections (antibiotic resistance, carbon source utilization, reversion of biosynthetic deficits) and then screen those clones for hypermutable and unselected mutations by whole genome sequencing. If the activating stress influences the spectrum of resulting genome changes, then different patterns should emerge in mutation type (e.g., point mutation vs. insertions) and location (e.g., mutations preferentially in biosynthetic vs. catabolic COGs).

Failure to find any stress-induced biases would validate the conventional view that biological inputs do not influence genome change. If biases do occur, then it will be possible to investigate both the underlying mechanisms and their adaptive utility.

Targeting of NGE by diverse molecular interactions

We have become accustomed to adapting natural targeting processes to our own genome rewriting goals. It is reasonable to hypothesize that the molecular targeting processes already identified will be found to serve adaptive functions. They certainly do so when NGE has evolved to become part of the normal lifecycle (http://shapiro.bsd.uchicago.edu/ExtraRefs.NaturalGeneticEngineeringPartNormalLifeCycle.shtml). Specific examples of functionally targeted genome restructuring include yeast mating-type switches, microbial antigenic variation (http://shapiro.bsd.uchicago.edu/Antigenic_Variation.html), and the adaptive immune system (http://shapiro.bsd.uchicago.edu/ExtraRefs.ImmuneSystemChanges.shtml). The adaptive utilization of genome targeting mechanisms can be investigated in systems such as bacterial transposon Tn7, yeast retrotransposons, or the Drosophila gypsy retrovirus, where the molecular basis for specificity is well documented and amenable to genetic modification. Mutant elements lacking targeting specificity can be tested for the ability to generate adaptive responses to stress as compared with the targeted parent element. Selections can include the ability to mobilize resistance determinants through bacterial populations (Tn7), activation of protein expression in yeast, or establishment of chromatin boundaries to recover functions silenced by position-effect in Drosophila (gypsy).

Real-time observations on domain shuffling and origination of novel functional domains (exonization)

The genome sequence record indicates that novel biochemical functions arise through domain accretion, domain shuffling, and the origination of novel domains by exonization of non-coding DNA and reverse-transcribed RNA. These processes have been documented historically and by synthetic model systems in the laboratory.

But there have not been real-time experiments to examine the generation of novel biochemical capabilities by exonization.

---

Table 1. Some biochemical activities involved in natural genetic engineering

| Activity                                      |
|----------------------------------------------|
| Nucleases (cutting)                          |
| Ligases (splicing)                           |
| DNA Polymerases (replicative, proofreading, and error-prone “mutator”) |
| Excisionases (remove improper/damaged bases) |
| Helicases (unwinding proteins)               |
| Annealing proteins (e.g., RecA)              |
| Site-specific recombinases (combined cutting and splicing) |
| Resolvases (cutting homologous recombination intermediates) |
| Reverse transcriptases (RNA → DNA)          |
| Transposases and integrases (cutting and splicing) |
| Sequence-specific, structure-specific DNA/RNA binding |

An extended and fully referenced version of this table is available online at http://shapiro.bsd.uchicago.edu/Table4A.CellBiochemicalActivitiesUsedin NaturalGeneticEngineering(NGE).html.

The genome sequence record indicates that novel biochemical functions arise through domain accretion, domain shuffling, and the origination of novel domains by exonization of non-coding DNA and reverse-transcribed RNA. These processes have been documented historically and by synthetic model systems in the laboratory.

But there have not been real-time experiments to examine the generation of novel biochemical capabilities by exonization.

---

Some Ideas for Evolutionary Research in the 21st Century

The research agenda for the present century must include recreating in real-time the innovative NGE processes we infer from the genomic record. That is the only way we can achieve a solid empirical understanding of the molecular mechanisms that produce functional genomic novelties.

Do cell control circuits play any role in facilitating the efficiency of the genomic search process? That these circuits control NGE activation and have the capacity to target NGE processes is well-documented (http://shapiro.bsd.uchicago.edu/Table1I.7.shtml and http://shapiro.bsd.uchicago.edu/Table1I.11.shtml). What seems difficult for many biologists to conceive is that NGE can be biased or “informed” by cell networks in a way that is adaptively useful. In order to initiate this line of evolution science research, I suggest the following topics:

- Specificity of mutagenic events following activation of NGE functions by distinct life history or stress events

---

www.landesbioscience.com RNA Biology 189
shuffling or exonization. The key to such experiments is to devise selections for novel reactions that will lead to proliferation on exotic substrates not metabolizable by known biochemistry. Organisms with known metabolic diversity can be tested for their ability to incorporate carbon, nitrogen, or other elemental nutrients from currently refractory substrates. It is critical to give the selected organisms long periods of incubation to produce the novel activities.245-247 This is an in vivo alternative to the design of proteins based on existing biochemical knowledge.248 The in vivo approach may lead to the discovery of unpredictable or totally novel exons and domain combinations, which can be revealed by cloning and sequencing the DNA that confers novel metabolic capacity.

Investigation of potential for regulatory coordination of DNA changes in network activation

Mobile elements provide a solution for the problem of integrating regulatory signals at multiple genetic loci.222 How rapidly and efficiently can this happen? That question can be answered experimentally by deleting the transcription signals from dispersed sequences encoding different steps in a catabolic or biosynthetic pathway in bacteria or yeast. The ability of mobile elements to activate transcription of each coding sequence alone and in pairs or higher combinations can be measured. If the frequency of multiple activations is significantly higher than the product of the individual activations, then there is prima facie evidence for coordinated insertion events. The underlying mechanisms are then available to investigation by modifying both the mobile element(s) involved in coordinate activation and the DNA sequences associated with each coding sequence. Essential controls include determining whether activation of a single coding sequence leads to slow growth on the selective medium, thereby facilitating the independent occurrence of insertions activating other sequences.249

These and other more ambitious evolution experiments are now practical because the products of selection are amenable to rapid whole genome sequencing. In addition, thanks to pioneers like Carl Woese, the conceptual environment for evolution science has changed dramatically. The establishment of clear molecular taxonomy and discovery of a distinct life form less than 40 years ago have proved central to liberating evolution science from pre-DNA ideas and prejudices. We should all be grateful.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Huxley J. Evolution: the modern synthesis. London: Allen & Unwin, 1942.
2. Gould SJ. The Structure of Evolutionary Theory. Cambridge: Harvard University Press, 2002.
3. Woese CR. A new biology for a new century. Microbiol Mol Biol Rev 2004; 68:173-86; PMID:15187180; http://dx.doi.org/10.1128/MMBR.68.2.173-186.2004
4. Woese CR, Magrum LJ, Fox GE. Archaeabacteria. J Mol Evol 1978; 11:245-51; PMID:691075; http://dx.doi.org/10.1007/BF01734485
5. Ishino Y, Ishino S. Rapid progress of DNA replication studies in Archaea, the third domain of life. Sci China Life Sci 2012; 55:386-403; PMID:22645083; http://dx.doi.org/10.1007/s11427-012-4324-9
6. O’Donnell M, Langston L, Stillman B. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. Cold Spring Harb Perspect Biol 2013; 5;5; PMID:2381849
7. Puhler G, Leffers H, Gropp F, Palm P, Klenk HP, Lottspeich F, Garratt RA, Zillig W. Archaeabacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. Proc Natl Acad Sci U S A 1989; 86:4569-73; PMID:2499884; http://dx.doi.org/10.1073/pnas.86.12.4569
8. Langer D, Hain J, Thuriaux P, Zillig W. Transcription in archaea: similarity to that in eucarya. Proc Natl Acad Sci U S A 1995; 92:5768-72; PMID:7597027; http://dx.doi.org/10.1073/pnas.92.13.5768
9. Bartlett MS. Determinants of transcription initiation by archaeal RNA polymerase. Curr Opin Microbiol 2011; 14:677-84;  PMID:21249191 ; http://dx.doi.org/10.1016/j.mib.2011.05.016
10. Benelli D, Londei P. Transcription initiation in Archaea: conserved and domain-specific features. Biochem Soc Trans 2011; 39:89-93; PMID:21265752; http://dx.doi.org/10.1042/BST0390089
27. Noutsos C, Kleine T, Armbruster U, DalCorso G. Nucleolar intercompartmental DNA transfer from the plastid to the nucleus. Bioessays 2008; 30:556-66; PMID:18478535; http://dx.doi.org/10.1002/bies.20761

28. Bock R, Timmis JN. Reconstructing evolution: gene transfer from plastids to the nucleus. Bioessays 2008; 30:556-66; PMID:18478535; http://dx.doi.org/10.1002/bies.20761

29. Kleine T, Maier UG, Leister D. Transfer of intercompartmental DNA from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annu Rev Plant Biol 2009; 60:115-38; PMID:19014347; http://dx.doi.org/10.1146/annurev.arplant.030408.092119

30. Harkazi-Covo E, Zeller RM, Martin W. Molecular poteugrains: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet 2010; 6:e1000834; PMID:20168995; http://dx.doi.org/10.1371/journal.pgen.1000834

31. Ahmadinejad N, Dagan T, Greunheit N, Martin W, Grubald T. Evolution of splanichosomal insertions following endosymbiotic gene transfer. BMC Evol Biol 2010; 10:57; PMID:20785877; http://dx.doi.org/10.1186/1471-2148-10-57

32. Popa O, Harkazi-Covo E, Landau G, Martin W, Dagan T. Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. Genome Res 2011; 21:599-609; PMID:21270712; http://dx.doi.org/10.1101/gr.115992.110

33. Leister D, Kleine T. Role of intercompartmental DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annu Rev Plant Biol 2009; 60:115-38; PMID:19014347; http://dx.doi.org/10.1146/annurev.arplant.030408.092119

34. Lloyd AH, Timmis JN. The origin and characterization of new genome origins containing a cytoplasmic organelle DNA. Mol Biol Evol 2011; 28:2019-28; PMID:21252282; http://dx.doi.org/10.1093/molbev/msr201

35. Wang D, Lloyd AH, Timmis JN. Nuclear genome diversity in somatic cells is accelerated by environmental stress. Plant Signal Behav 2012; 7:595-7; PMID:22356183; http://dx.doi.org/10.4161/psb.19871

36. Wang D, Timmis JN. Cytoplasmic organelle DNA preferentially inserts into open chromatin. Genome Biol Evol 2013; 5:1060-4; PMID:23665654; http://dx.doi.org/10.1093/gbe/evt200

37. Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile GH, Hirakawa Y, et al. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature 2012; 492:59-63; PMID:22957150; http://dx.doi.org/10.1038/nature11759

38. Curris BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile GH, Hirakawa Y, et al. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature 2012; 492:59-63; PMID:22957150; http://dx.doi.org/10.1038/nature11759

39. Song S, Jiang F, Yuan J, Guo W, Mao Y. Exceptionally high cumulative percentage of NUMTs originat- ing from linear mitochondrial DNA molecules in the Hydra magnipapillata genome. BMC Genomics 2013; 14:447; PMID:23826818; http://dx.doi.org/10.1186/1471-2164-14-447

40. Gould SB. Evolutionary genomics: Alga’s complex origins. Nature 2012; 492:46-8; PMID:23201689; http://dx.doi.org/10.1038/nature11759

41. Jones C, Richle E, Leister D. Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. Genome Res 2005; 15:616-28; PMID:15867426; http://dx.doi.org/10.1101/gr.378705

42. Noutsos C, Richle E, Leister D. Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. Genome Res 2005; 15:616-28; PMID:15867426; http://dx.doi.org/10.1101/gr.378705

43. Thorsness PE, Fox TD. Escape of DNA from mitochondria to the nucleus: frequent gene transfer from the chloroplast to the nucleus. Proc Natl Acad Sci U S A 1975; 72:2418-22; PMID:806081; http://dx.doi.org/10.1073/pnas.72.4.2418

44. Noutsos C, Kleine T, Armbruster U, DalCorso G. Leister D. Nuclear insertions of organelar DNA can create novel patches of functional exon sequences. Trends Genet 2007; 23:597-601; PMID:17981356; http://dx.doi.org/10.1016/j.tig.2007.08.016

45. Noutsos C, Kleine T, Armbruster U, DalCorso G. Leister D. Nuclear insertions of organelar DNA can create novel patches of functional exon sequences. Trends Genet 2007; 23:597-601; PMID:17981356; http://dx.doi.org/10.1016/j.tig.2007.08.016

46. Ricchetti M, Tekaiya F, Dujon B. Continued colonization of the human genome by mitochondrial DNA. PLoS Biol 2004; 2:e273; PMID:15369197; http://dx.doi.org/10.1371/journal.pbio.0020273

47. Stemmman S, Bock R. Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. Plant Cell 2006; 18:2869-78; PMID:17085684; http://dx.doi.org/10.1105/tpc.106.046466

48. Ricchetti M, Fairhead C, Dujon B. Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. Nature 1999; 402:96-100; PMID:10575425; http://dx.doi.org/10.1038/47076

49. Harkazi-Covo E, Covo S. Numt-mediated double-strand break repair mitigates deleterious during primate genome evolution. PLoS Genet 2008; 4:e1000237; PMID:18949041; http://dx.doi.org/10.1371/journal.pgen.1000237

50. Lloyd AH, Timmis JN. Endosymbiotic evolution in action: Real-time observations of chloroplast to nucleus gene transfer. Mob Genet Elements 2011; 1:216-20; PMID:22479609; http://dx.doi.org/10.1016/j.mge.2011.07.137947

51. Rousseau-Gueutin M, Ayliffe MA, Timmis JN. Plastid DNA in the nucleus: new genes for old. Plant Signal Behav 2012; 7:2609-72; PMID:22415049; http://dx.doi.org/10.4161/psb.22415049

52. Thorsness PE, Fox TD. Escape of DNA from mitochondria to the nucleus: frequent gene transfer from the chloroplast to the nucleus. Proc Natl Acad Sci U S A 1975; 72:2418-22; PMID:806081; http://dx.doi.org/10.1073/pnas.72.4.2418

53. Thorsness PE, Fox TD. Escape of DNA from mitochondria to the nucleus: frequent gene transfer from the chloroplast to the nucleus. Proc Natl Acad Sci U S A 1975; 72:2418-22; PMID:806081; http://dx.doi.org/10.1073/pnas.72.4.2418

54. Jensen-Seaman MI, Wildschutte JH, Soto-Calderón ID, Anthony NM. A comparative approach shows differences in patterns of numt insertion during hominoid evolution. J Mol Evol 2009; 68:688-99; PMID:19471988; http://dx.doi.org/10.1007/s00229-009-9243-4

55. Tekaia F, Dujon B. Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. Nature 1999; 402:96-100; PMID:10575425; http://dx.doi.org/10.1038/47076

56. Wooley CR. Endosymbiotic evolution: organelle genomes of flowering plants. Genome Res 2005; 15:616-28; PMID:15867426; http://dx.doi.org/10.1101/gr.378705

57. Wooley CR. Endosymbiotic evolution: organelle genomes of flowering plants. Genome Res 2005; 15:616-28; PMID:15867426; http://dx.doi.org/10.1101/gr.378705
115. Filée J. Lateral gene transfer, lineage-specific gene transfer of ecologically important traits. Annu Rev Ecol Evol Syst 2011; 42:217-28; PMID:21366522; http://dx.doi.org/10.1146/annurev-ecolsys-110810-134158.

124. Chien M, Morozova I, Shi S, Sheng H, Chen J, Gomez SM, Asamani G, Hill K, Nuara J, Feder M, et al. The genomic sequence of the accidental pathogen Legionella pneumophila. J Bacteriol 2011; 193:1114-21; PMID:21169481; http://dx.doi.org/10.1128/JB.01789-10.

135. Roiz D, Vázquez A, Seco MP, Tenorio A, Rizzoli A. Filoviruses are ancestral single-stranded DNA viruses in vertebrate genomes. PLoS Pathog 2013; 9:e1003759; http://dx.doi.org/10.1371/journal.ppat.1003759.

137. Taylor DJ, Leach RW, Fishcotte C, Schuweck M, Schäfer C, Pace JK 2nd, Brindley PJ, Feschotte C. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. Nature 2010; 464:1347-50; PMID:20428170; http://dx.doi.org/10.1038/nature08939.

139. Barteneva NS, Maltez N, Vorobjev IA. Microvesicles and intercellular communication in the context of parasitism. Front Cell Infect Microbiol 2013; 3:49; PMID:24032108; http://dx.doi.org/10.3389/fcimb.2013.00049.

140. Yutin N, Colson P, Raoult D, Koonin EV. Mimiviridae: clusters of orthologous genes, reconstruction of gene repertoire evolution and proposed expansion of the giant virus family. Virol J 2013; 10:106; PMID:23557328; http://dx.doi.org/10.1186/1743-422X-10-106.

141. Iskra-Caruana ML, Baurens FC, Gayral P, Chabannes P, et al. The genomic sequence of the accidental pathogen Legionella pneumophila. J Bacteriol 2011; 193:1131-44; PMID:21450636; http://dx.doi.org/10.1128/JB.00395-10.

142. Kapoor A, Simmonds P, Lipkin WI. Discovery and characterization of mammalian endogenous parvoviruses. J Virol 2010; 84:12628-35; PMID:20943966; http://dx.doi.org/10.1128/JVI.00732-10.

143. Katzourakis A, Gifford RJ. Endogenous viral elements in animal genomes. PLoS Genet 2010; 6:e1001191; http://dx.doi.org/10.1371/journal.pgen.1001191.

144. Huws SA, Morley RJ, Jones MV, Brown MR, Smith AM. Widespread endogenization of cellular genes in eukaryotic nuclear genomes. J Virol 2010; 84:11876-90; PMID:20861255; http://dx.doi.org/10.1128/JVI.01039-10.

145. Horn P, Honda T, Suzuki Y, Kobayashi Y, Daito T, Oshida T, Iriu K, Jern P, Gojobori T, Coffin JM, et al. Endogenous non-retroviral RNA virus elements in mammalian genomes. Nature 2010; 463:84-9; PMID:20504395; http://dx.doi.org/10.1038/211248a0; http://dx.doi.org/10.1159/000287809.

146. Horie M, Honda T, Suzuki Y, Kobayashi Y, Daito T, Oshida T, Iriu K, Jern P, Gojobori T, Coffin JM, et al. Endogenous non-retroviral RNA virus elements in mammalian genomes. Nature 2010; 463:84-9; PMID:20504395; http://dx.doi.org/10.1038/211248a0; http://dx.doi.org/10.1159/000287809.
151. Gruking MB, Weber J, Dewannieux M, Goetzlik E, Heidtmann H, Zinkernagel RM, Hangartner L. Combinaison de retron- 

poson and exogenous RNA virus results in non- 

retroviral dNA integration. Science 2009; 323:163; 

PMID:19590848; http://dx.doi. 

.org/10.1126/science.1167375 

152. Bill CA, Summers J. Genomic DNA double-strand 

breaks are targets for headpavidna DNA integra- 

Proc Natl Acad Sci U S A 2004; 101:11139- 

40; PMID:15258209; http://dx.doi. 

.org/10.1073/pnas.0403925101 

153. Bill CA, Summers J. Genomic DNA double-strand 

breaks are targets for headpavidna DNA integra- 

Proc Natl Acad Sci U S A 2004; 101:11139- 

40; PMID:15258209; http://dx.doi. 

.org/10.1073/pnas.0403925101 

154. Gould SJ, Vrba ES. Exaptation—a missing term in 

the evolution of novel fun - 

ctional entities. Proc Natl Acad Sci U S A 2000; 97: 

585-601; PMID:10825806; http://dx.doi. 

.org/10.1073/pnas.97.585587 

155. Blond JL, Lavielle D, Cheynet V, Bouton O, Oriol 

gki236 

gki236
245. Campbell JH, Lengyel JA, Langridge J. Evolution of a second gene for beta-galactosidase in Escherichia coli. Proc Natl Acad Sci U S A 1973; 70:1841-5; PMID:4124306; http://dx.doi.org/10.1073/pnas.70.6.1841

246. Shapiro JA. Observations on the formation of clones containing araB-lacZ cistron fusions. Mol Gen Genet 1984; 194:79-90; PMID:6233472; http://dx.doi.org/10.1007/BF00383501

247. Hall BG, Hauer B. Acquisition of new metabolic activities by microbial populations. Methods Enzymol 1993; 224:603-13; PMID:8264414; http://dx.doi.org/10.1016/0076-6879(93)24045-V

248. Davids T, Schmidt M, Botcher D, Bornscheuer UT. Strategies for the discovery and engineering of enzymes for biocatalysis. Curr Opin Chem Biol 2013; 17:215-20; PMID:23523243; http://dx.doi.org/10.1016/j.cbpa.2013.02.022

249. Reeb JR, Kugelberg E, Reams AB, Kofoid E, Andersson DI. Origin of mutations under selection: the adaptive mutation controversy. Annu Rev Microbiol 2006; 60:477-501; PMID:16761951; http://dx.doi.org/10.1146/annurev.micro.60.080805.142045