A three-part origin saga

1: Monday, 03/05
Origin of life…and before

2: Tuesday, 03/06
Origin of first cells

3: Friday, 03/09

NOW

BIG BANG
~13 GYA

Earth
~4.5 GYA

EUK
~4 GYA

~2 GYA
On the origin of viruses and cells: the ancient virus world as the cradle of cellular life

Eugene V. Koonin
National Center for Biotechnology Information, NIH, Bethesda, MD, USA

KITP, Santa Barbara, March 9, 2007
• Comparative genomics and evolution of viruses: the ancient Virus World

• The nature of the Last Universal Cellular Ancestor (LUCA) and the origin of cells

• An attempt on a grand synthesis: the viral model of the early evolution of life
“The reputedly intractable problem of the origin of viruses has long been neglected. In the modern literature, ‘virus evolution’ has come to refer to studies more akin to population genetics, such as the worldwide scrutiny of new polymorphisms appearing daily in the H5N1 avian flu virus [1], than to the fundamental question of where viruses come from. This is now rapidly changing, as a result of the coincidence of bold new ideas (and the revival of old ones), the unexpected spectacular features of some recently isolated viruses [2,3], as well as the steady increase in the numbers of genomic sequences for ‘regular’ viruses and cellular organisms, which enhances the power of comparative genomics [4].”

The first paradox of virus evolution:
The metagenomes of viruses and cellular organisms have comparable complexities

1 cm³ of seawater contains $10^6$-$10^9$ virus particles;
~10 per each cell


Millions of distinct bacteriophage species in the water, soil, and gut


~ 90% of phage genes have no homologs in bacterial genomes with large proportion of unique genes

Liu *et al.* (2005) *Virus Res.* **117**:68
The second paradox of virus evolution:
the diversity of viral genetic cycles
as opposed
to the single genetic cycle of cellular organisms
A general transfer is one which can occur in all cells. The obvious cases are
DNA -> DNA
DNA --> RNA
RNA --> Protein

A special transfer is one which does not occur in most cells, but may occur *in special circumstances*.

RNA --> RNA
RNA --> DNA
DNA --> Protein

*It [classification] was intended to apply only to present-day organisms, and not to events in the remote past, such as the origin of life or the origin of the code*

Natural history of viral genes: a one-page summary of *viral comparative genomics*

I. *Genes with readily detectable homologs from cellular life forms:*
   1. Genes with closely related homologs from cellular organisms (typically, the host of the given virus) present in a narrow group of viruses
   2. Genes that are conserved within a virus lineage or even several lineages and have moderately close cellular homologs
   *Origin: relatively recent (1) or ancient (2) acquisition from host*

II. *Virus-specific genes*
   3. ORFans, i.e., genes without detectable homologs except, possibly, in closely related viruses
   4. Virus-specific genes that are conserved within a virus lineage
   ??
   *Acquisition from host but with rapid divergence from ancestor once within viral genomes?*

III. *Viral hallmark genes*
   5. Genes shared by many diverse virus lineages, with only very distant homologs in cellular organisms
Contributions of different classes of viral genes to the genomes of different classes of viruses: strong dependence on genome size

Genome size (log) 0 1 2 3

Most RNA viruses, retroelements, RCR replicons

Large RNA viruses, adeno, tailed phages

NCLDV, herpes, large phages

[Diagram]

- Recent acquisitions
- Old acquisitions
- Virus-specific ORFans
- Virus-specific conserved
- Virus hallmark
Natural history of viral genes: Viral Hallmark Genes

Shared by many diverse groups of viruses

Strong support for monophyly of all viral members of the respective gene families

Only distant homologs in cellular organisms

Play major roles in genome replication, packaging and assembly

Can be viewed as distinguishing characters of the ‘virus state’
The hallmark viral genes/proteins

Proteins crucial for virus replication and structure that are shared by a broad variety of virus lineages but are only represented by very distant homologs (except for obvious acquisitions from viruses) in cellular life forms

• RNA-dependent RNA polymerase – most RNA viruses – eukaryotes and bacteria
• Reverse transcriptase – retroid viruses/retroelements/group II introns – eukaryotes and bacteria (very few instances in archaea)
• Superfamily III helicase – numerous groups of RNA, small DNA, and large DNA viruses – eukaryotes and bacteria
  • Distinct family of archaeo-eukaryotic primases – larger DNA viruses and plasmids - eukaryotes and archaea
• Rolling circle replication initiator nuclease – a variety of small DNA viruses, plasmids, and transposons – bacteria, archaea, and eukaryotes
• Distinct (A32-like) family of DNA-pumping ATPase involved in packaging – various large and small DNA viruses – bacteria and eukaryotes
• Icosahedral capsid protein (jelly-roll domain) – most icosahedral viruses, from smallest RNA viruses to mimivirus

For each of these proteins, homologs with the same fold are found in cellular life forms but monophyly of viral proteins is beyond reasonable doubt
A few examples of viral hallmark proteins

- Jelly-roll capsid protein
- Superfamily 3 helicase
- RNA-dependent RNA polymerase and Reverse transcriptase
- Rolling circle replication initiation endonuclease
Jelly-roll capsid proteins are found in an astonishing variety of RNA and DNA viruses.
Superfamily 3 DNA helicases from diverse viruses are very similar.

Helicase superfamily 3 is a monophyletic, deep-branching viral clade within a major class of P-loop NTPases present in a vast variety of viruses: from the smallest (+) RNA viruses (picornaviruses) to the largest DNA viruses (e.g., mimivirus). Iyer et al. (2004) *J. Struct. Biol.* **146**:11-31.
What is the evolutionary explanation for the striking features of Viral Hallmark Genes?

- Artifact of the rapid sequence divergence of viral genes
- Heritage of the last universal common ancestor of viruses
- Spread by horizontal gene transfer
- **Origin from a pre-cellular phase of life’s evolution: A primordial gene pool**

**Shared by many diverse groups of viruses**

**Strong support for monophyly of all viral members of the respective gene families**

**Only distant homologs in cellular organisms**

**Play major roles in genome replication, packaging and assembly**

**Can be viewed as distinguishing characters of the virus state**
The origin and evolutionary history of hallmark viral genes - the primordial gene pool hypothesis:

The hallmark genes AND, by implication, the major lineages of modern viruses descend directly from a primordial gene pool
A crucial corollary: If viruses come directly from a primordial gene pool, then, origin of viruses is inextricably linked to the origin of cells
Competing concepts of the origin of viruses

Cell degeneration

- CELL
- SMALL PARASITIC CELL
- VERY SMALL VIRUS

Escaped genes

- CHROMOSOME
- mRNA
- PLASMID
- VIRUS

Primordial genetic systems

- PRE-CELLULAR LIFE FORMS
- RNA
- DNA
- VIRUS

? 

CELLS
Last Universal Common Ancestor: **LUCA**

• The “uniformitarian” view (Hutton-Lyell-Darwin): LUCA must have been like modern prokaryotes, perhaps, somewhat simpler

• The radical view: LUCA was dramatically different from any extant cell…at least in some respects
LUCA: complexity versus “cellularity”

- Complex but unlike modern cells
- Primitive with no cellular organization
- Complex and cell-like (a prokaryote)
- Primitive but cell-like
Different versions of LUCA

Modern-type cell
- dsDNA
- RNA
- membrane

RNA cell
- RNA
- dsDNA
- ssDNA
- membrane

Pre-cellular stage, mixture of RNA segments

Pre-cellular stage, mixture of diverse genetic elements
What does comparative genomics have to say about LUCA?

At least 2 major problems

Apparently unrelated in archaea and bacteria:

• membrane chemistry/biogenesis
• core DNA replication systems
A radical …but parsimonious solution: LUCA had neither

- modern-type membranes
- modern-type DNA replication

Leipe DD, Aravind L, Koonin EV. Did DNA replication evolve twice independently? Nucleic Acids Res. 1999 Sep 1;27(17):3389-401


Koonin EV, Martin W. On the origin of genomes and cells within inorganic compartments. Trends Genet. 2005 Dec;21(12):647-54
A membrane-less, non-cellular but spatially confined LUCA: a hypothesis to explain the disparity of membranes in bacteria and archaea

Martin W, Russell MJ. On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells.


- We propose that life evolved in structured iron monosulphide precipitates in a seepage site hydrothermal mound at a redox, pH and temperature gradient between sulphide-rich hydrothermal fluid and iron(II)-containing waters of the Hadean ocean floor. The naturally arising, three-dimensional compartmentation observed within fossilized seepage-site metal sulphide precipitates indicates that these inorganic compartments were the precursors of cell walls and membranes found in free-living prokaryotes.
Iron sulfides (FeS – ancient or Fe$_2$S$_3$ -modern); NiS
3D compartments, not just 2D surfaces

The crucible of Life?

Early evolution from monomers to first cells

Koonin, Martin, Trends Genet. 2005 Dec;21(12):647-54
• The Martin-Russell chimney – a system of hydrothermal compartments - is a perfect home for a multipartite, virus-like genome of LUCA postulated by Leipe et al.

Koonin EV, Martin W. On the origin of genomes and cells within inorganic compartments. Trends Genet. 2005 Dec;21(12):647-54
Early evolution from monomers to first cells

Russell-Martin chimney

Transfer of genetic content between compartments (infection)

LUCA

Koonin, Martin, Trends Genet. 2005 Dec;21(12):647-54
Salient features of early evolution: first ensembles of genetic elements – LUCA - archaeal and bacterial cells

- LUCA: community of genetic elements inhabiting the Russell-Martin mound – more a *state* than an organism - *LUCAS*
- RNA replication+retrovirus-like cycle
- **Selfish cooperatives**: ensembles of virus-like genetic elements coding not only for replication but also for accessory functions (including translation, precursor synthesis etc)
- Advantage of physical connection between functionally coupled genes – origin of (super)operons from selfish cooperators – driving force for subsequent transition to DNA genomes
- No cell division but effective selection for compartment contents via “infection” of neighboring compartments – “horizontal transfer” is an intrinsic feature of LUCA’s lifestyle
- Independent invention of at least two machineries for membrane biogenesis and DNA replication
- Probably, many escape attempts but only two successful ones: archaea and bacteria
Archaea and bacteria escape with their own complements of viruses

Persistence of purely selfish replicons parasitizing on selfish cooperatives

Early virus world evolves

Russell-Martin chimney

Viruses coming out of the chimney...
RNA-protein world

DNA; diverse replication/expression systems

Advanced membranes, protocells

archaea

bacteria

Cellular escape

escape

With complements of viruses

networks of inorganic compartments

RNA-protein world

RNA world

First genetic elements

DNA; diverse replication/expression systems

Further diversification; major classes of viruses; membrane-containing virions

Virus-like entities; ensembles of selfish cooperators

LUCAS

Origin of cells and viruses: sketch of a synthesis

time
The Virus World

- There is a small pool of essential genes involved in viral replication and structure and shared by extremely diverse groups of viruses - **virus hallmark genes**

- ...and much larger gene pools shared by related groups of viruses, e.g., tailed phages

- All major classes of viruses emerged within the primordial gene pool (pre/post-LUCA/S)

- The Virus World has existed as a continuous gene flow since the pre-LUCA/S times

- Under this concept, **viruses and cells co-evolved**, ancestors of viral and cellular genomes were indistinguishable at very early stages of evolution, and virus-like agents antedate fully-fledged cells

- **Unification of the cellular and viral branches of life’s evolution** at the earliest stages analogous to the unification of physical interactions in the first seconds after the Big Bang

Koonin, Dolja, 2006, Virus Res.
Koonin, Senkevich, Dolja, 2006, Biology Direct
On the origin of different types of cells

• Substantial genetic and metabolic complexity at a pre-cellular, virus-like stage – where LUCA/S (not a prokaryote!), probably, belongs

• Independent origin of two types of membrane organization, escape of two types of cells (archaea and bacteria) capturing two independently evolved DNA replication systems from the pre-cellular pool, probably, amidst many failed attempts

• Origin of eukaryotic cell through endosymbiotic collision of the two types of prokaryotic cellular organization; origin of the nucleus and some other major eukaryotic novelties as defense against the Invasion of Group II introns, the progenitors of spliceosomal introns

Koonin, Martin, Trends Genet. 2005 Dec;21(12):647-54
Koonin, Senkevich, Dolja, 2006, Biology Direct
There is little new under the sun…

"...life may have remained in the virus stage for many millions of years before a suitable assemblage of elementary units was brought together in the first cell"

Haldane JBS: The Origin of Life. 
Rationalist Annual 1928, 148:3-10.
Acknowledgments

Bill Martin (U. Duesseldorf, Germany)

Tania Senkevich (NIAID, NIH)

Valerian Dolja (Oregon State)
Welcome to Biology Direct
An online journal published by BioMed Central

Biology Direct considers original research articles, hypotheses and reviews in selected subject areas, and will eventually cover the full spectrum of biology. Subject areas already launched include Genomics, Bioinformatics and Systems Biology, and Immunology.

Biology Direct aims to provide authors and readers of research articles with a novel system of peer review. This will include making the author responsible for obtaining reviewers' reports, via the journal's Editorial Board; making the peer review process open rather than anonymous; and publishing the reviewers' reports along with the articles, thus increasing both the responsibility and the reward of the referees and eliminating sources of abuse in the refereeing process. See about Biology Direct and the Biology Direct instructions for authors for further information.

Submit your article online

Research  Open Access
Retrospective analysis of protein kinase C-beta (PKC-β) expression in lymphoid malignancies and its association with survival in diffuse large B-cell lymphomas
Shuyu Li, Mark Phong, Michael Lahn, Leslie Brall, Susan Sutton, Boris K Lin, Donald Thornton, Birong Liao
Biology Direct 2007, 2:8 (21 February 2007)
[Abstract] [Full Text] [PDF] [PubMed] [Related articles]

Research  Open Access
Conserved intron positions in ancient protein modules
Albert DG de Roos
Biology Direct 2007, 2:7 (8 February 2007)
[Abstract] [Full Text] [PDF] [PubMed] [Related articles]
Top 10 most accessed articles for last 30 days / past year / all time

1. **Research**  Open Access  Highly accessed
   **Origin and evolution of the peroxisomal proteome**
   Toni Gabaldón, Berend Snel, Frank van Zijlmeren, Wieger Hemrika, Henk Tabak, Martijn A Huynen
   Biology Direct 2006, 1:8 (23 March 2006)
   [Abstract] [Full Text] [PDF] [PubMed] [Related articles] [F1000 Biology] [Cited on BioMed Central]

2. **Hypothesis**  Open Access  Highly accessed
   **The ancient Virus World and evolution of cells**
   Eugene V Koonin, Tatiana G Senkevich, Valerian V Dolja
   Biology Direct 2006, 1:29 (19 September 2006)
   [Abstract] [Full Text] [PDF] [PubMed] [Related articles] [F1000 Biology] [Cited on BioMed Central]

3. **Research**  Open Access  Highly accessed
   **A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action**
   Kira S Makarova, Nick V Grishin, Svetlana A Shabalina, Yuri I Wolf, Eugene V Koonin
   Biology Direct 2006, 1:7 (16 March 2006)
   [Abstract] [Full Text] [PDF] [PubMed] [Related articles] [F1000 Biology] [Cited on BioMed Central]

4. **Research**  Open Access  Highly accessed
   **Rooting the tree of life by transition analyses**
   Thomas Cavalier-Smith
   Biology Direct 2006, 1:19 (11 July 2006)
   [Abstract] [Full Text] [PDF] [PubMed] [Related articles] [Cited on BioMed Central]