

Giant viruses (Megavirales) : derived from mitochondrial spores?

Hervé Seligmann

Previously at:

Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UMR
MEPHI, Aix-Marseille Université, 13005 Marseille, France.

The National Natural History Collections, The Hebrew University of Jerusalem, 91904
Jerusalem, Israel

E-mail: varanuseremius@gmail.com

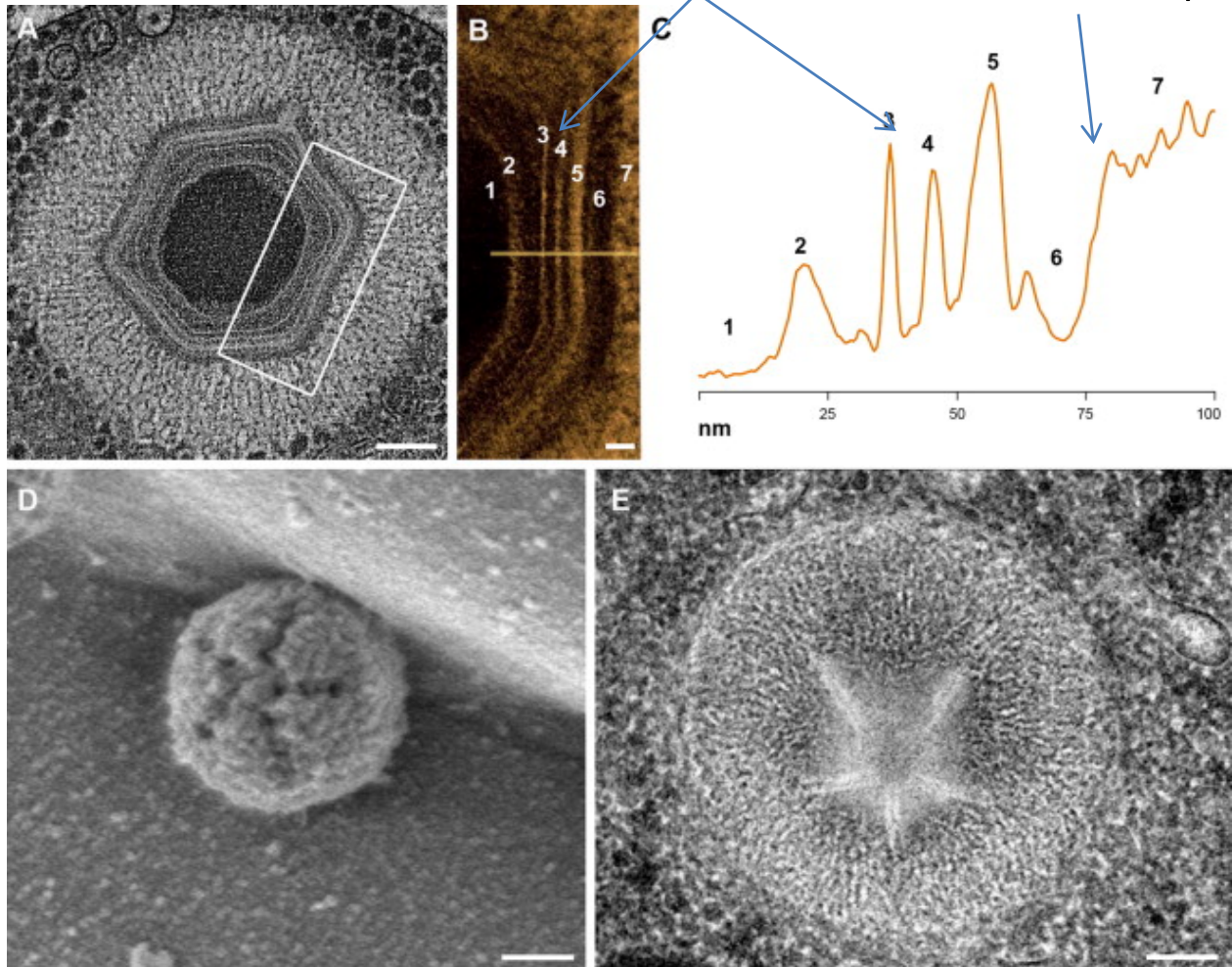
How the idea came up:

To study megaviral candidate homologues of mitochondrial rRNAs, candidate homologues mt CDs were examined as controls. BLASTn results indicated similar gene orders for mito- & giant virus genomes

- Mitochondria, bacterial spores and giant viruses have double membranes
- Spores are protein-coated, as are viruses
- Linear mitogenome-encoded DNA pol clusters with giant virus DNA pol
- Post-transcriptional RNA editing punctuated by secondary structure
- Numts: nuclear mitochondrial inserts like viruses?

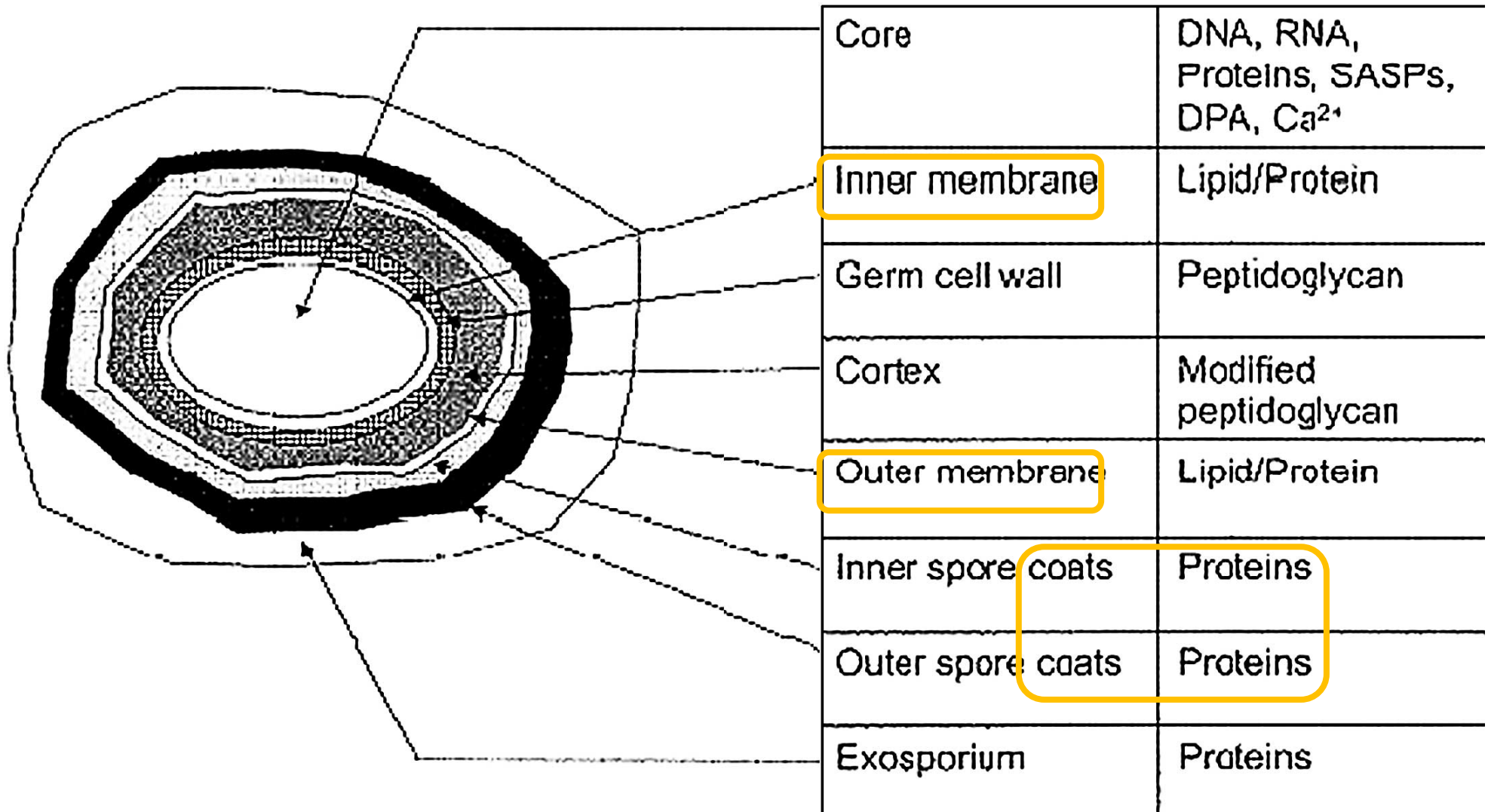
Similar morphological organizations

Mimivirus: 3+4, membranes; 5+6, capsids



From: Mutsafi et al 2014. Virology 466-467, 3-14

Bacterial endospore structure



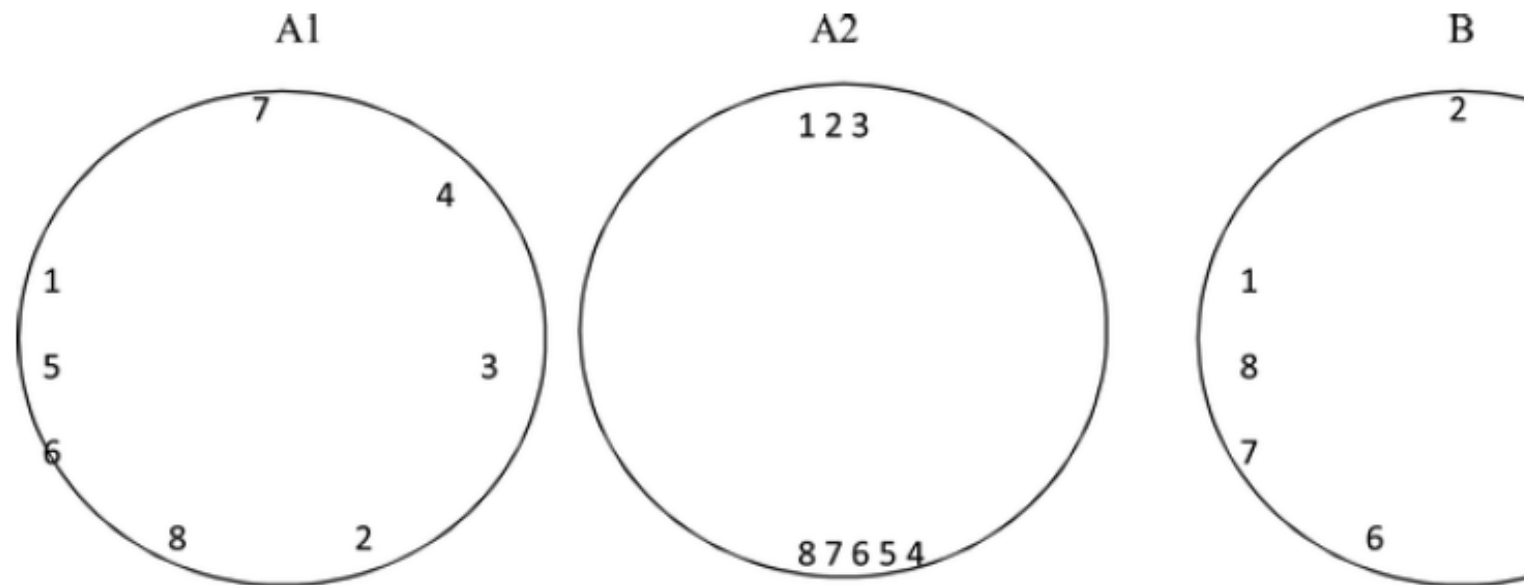
Ancestral synteny:

On average, alignments of amoeban mitogenes with giant virus genomes map across the giant virus genome in an order similar to their order on the amoeban mitogenome.

-> mitogenome turned into viral genome? Opposite?
Convergence? Chance?

The effect is strongest for:
conserved mitogenes, and short giant virus genomes.

Seligmann 2018 Virus Research 253, 77-88.



3 scenarios:

C random recent inserts ancestral synten

Gene	Order	1	2	3	4	5	6	7	8
Gene1	1	1	2	3	4	5	6	7	8
Gene2	2	2	3	4	5	6	7	8	1
Gene3	3	3	4	5	6	7	8	1	2
Gene4	4	4	5	6	7	8	1	2	3
Gene5	5	5	6	7	8	1	2	3	4
Gene6	6	6	7	8	1	2	3	4	5
Gene7	7	7	8	1	2	3	4	5	6
Gene8	8	8	1	2	3	4	5	6	7
A1		-0.26	0.41	0.31	0.02	-0.64	-0.17	0.12	0.21
A2		<u>1</u>	0.33	-0.14	-0.43	-0.52	-0.42	-0.14	0.33
B		0.31	0.98	0.31	-0.17	-0.45	-0.55	-0.26	-0.17

Fig. 1. Schematic representation of distribution of mitogenome-like sequences in megaviral genomes (imaginary example), along hypotheses a1 reflect numerous random short insertions, hypothesis a2 few single insertion(s). The pattern for hypothesis b reflects a mitogenome-like co sequences from other origins inserted between ancestral sequences with mitochondrial origin. Hypotheses a and b are not exclusive, a could

Mt gene order similar to order of mt alignments on giant virus genomes

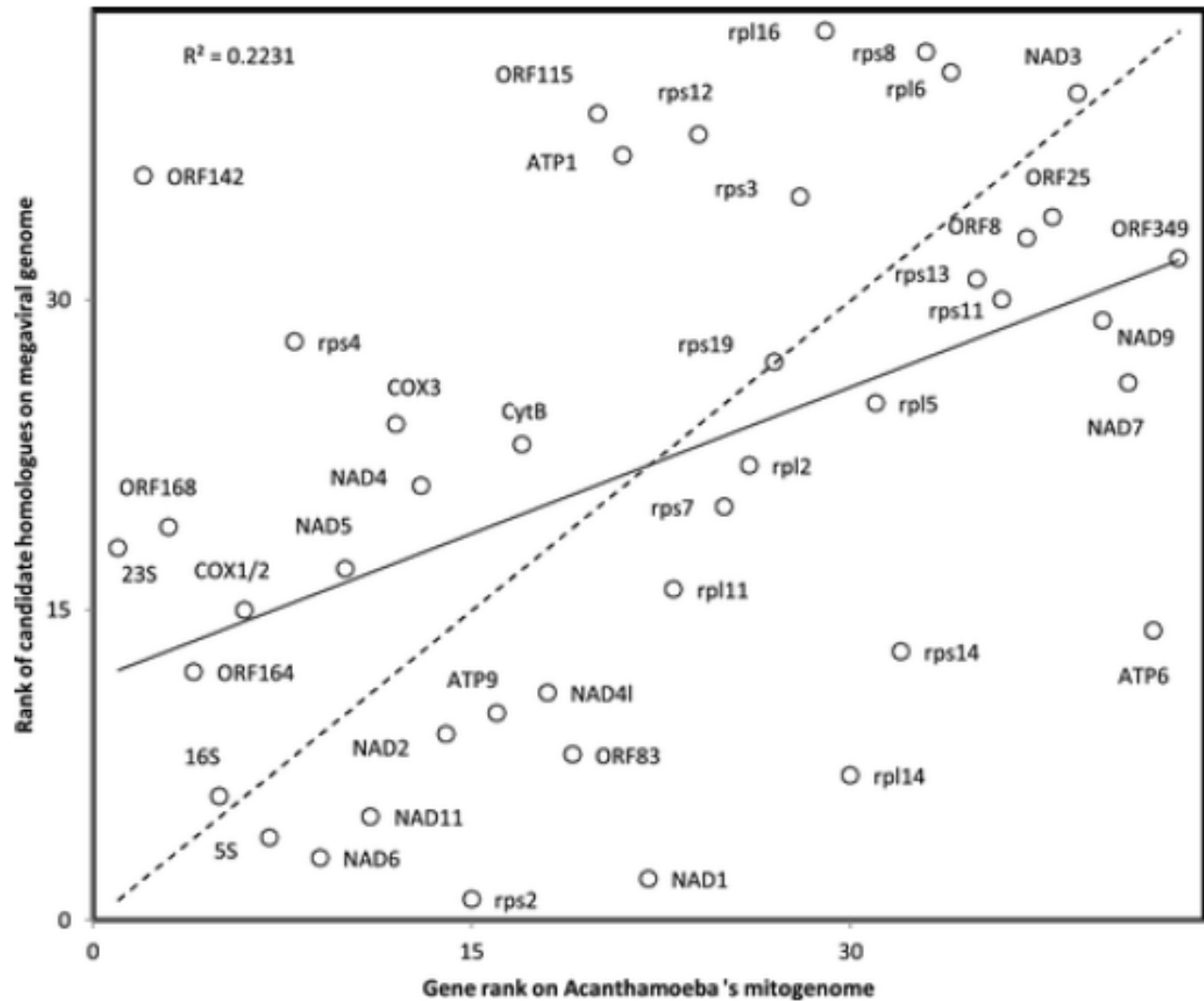


Fig. 2. Megaviral genome order of candidate megaviral homologues of mitogenes as a function of their order in *Acanthamoeba*'s mitogenome. The dashed line indicates $y = x$, the continuous line represents the linear regression between megaviral and mitogenome orders (Pearson correlation $r = 0.472$, one tailed $P = 0.0007$).

Gene order most conserved short giant virus genomes

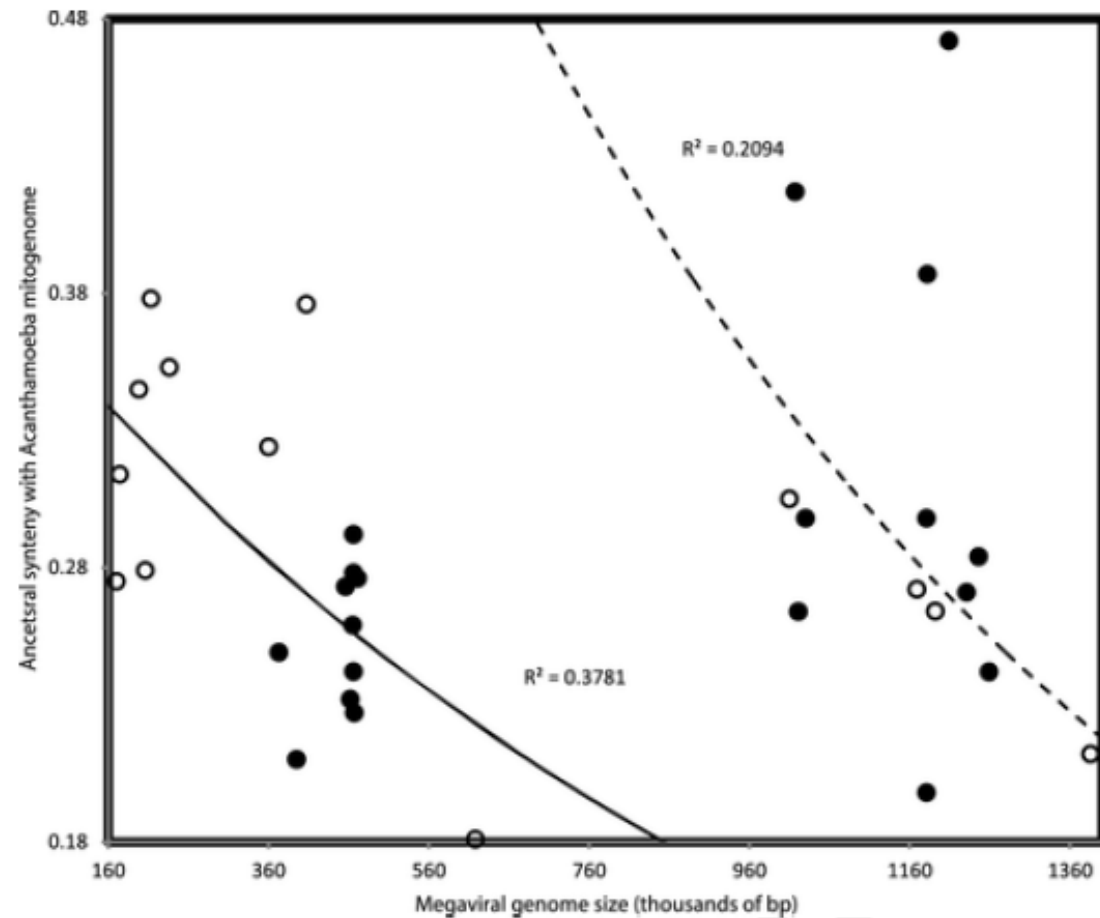
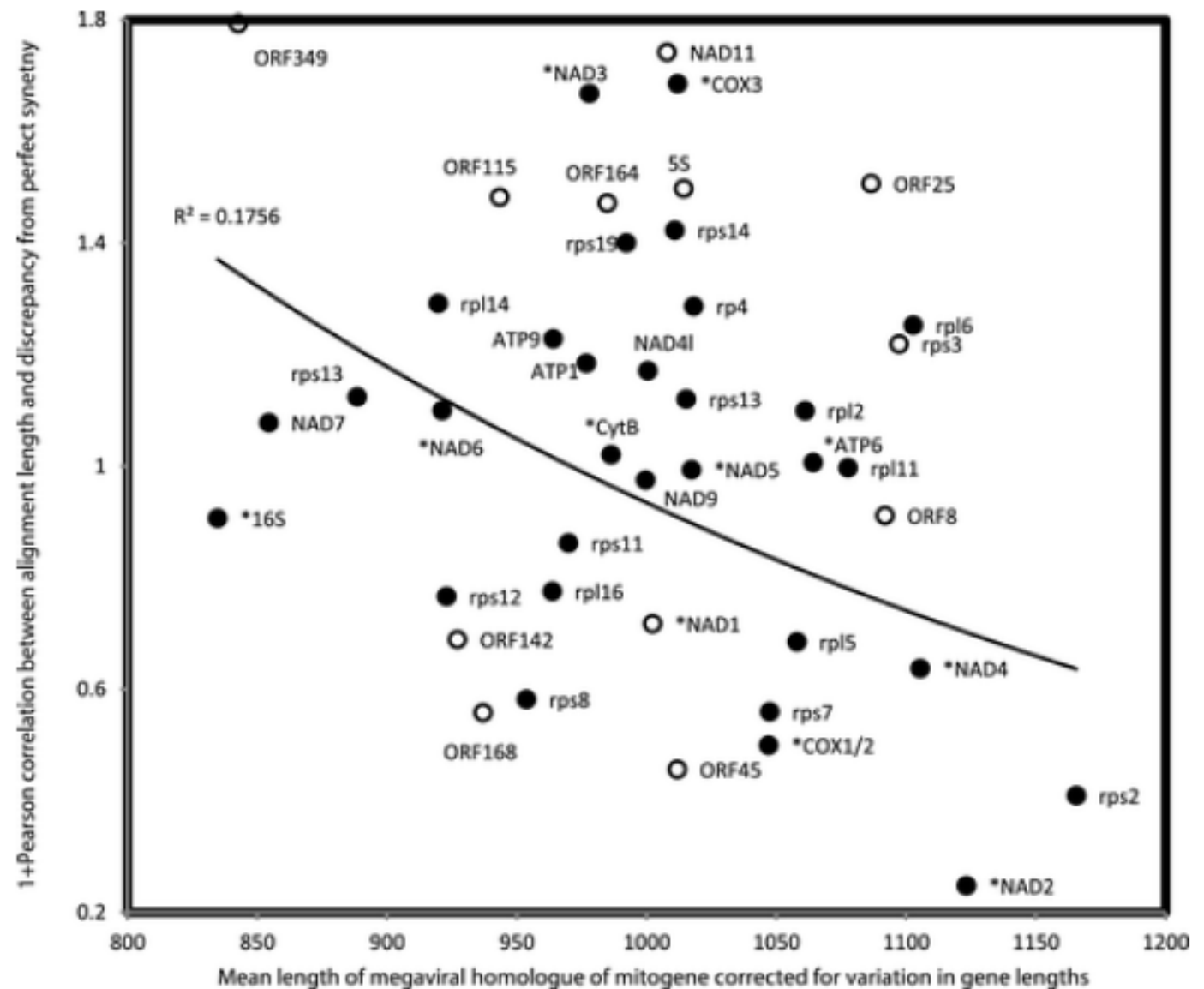


Fig. 3. Estimate of ancestral synteny between *Acanthamoeba*'s mitogenome and candidate megaviral homologues in 36 megaviral genomes (syn in Table 1) as a function of megaviral genome size (in thousands of bp). Filled and hollow symbols indicate amoeban and non-amoeban host, respectively. Correlation analyses done separately for small and large genomes (continuous and dashed lines, respectively: Pearson correlation $r = -0.62$, $P = 0.002$ and $r = -0.46$, $P = 0.043$, one tailed tests).

The longer the alignment, the closer it is in the viral genome to its predicted position according to the mt gene order, and this more for conserved mt genes



From mitochondria to giant viruses or from giant viruses to mitochondria?

A different approach using tRNA secondary structure and the order of integration of amino acids in the genetic code

Numerous hypotheses suggest AA integration order: AA chemical inertness, AA structural complexity, Miller's experiment, meteorite AAs, Juke's neutral evolution, Wong's metabolic coevolution, etc.

Trifonov's reviews (2000, 2004) list 41 hypotheses, show that they overall converge, justifying a consensus order of inclusion of amino acids in the genetic code:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
G A V D P S E L T R N K Q I C H F M Y W
Simple----->complex

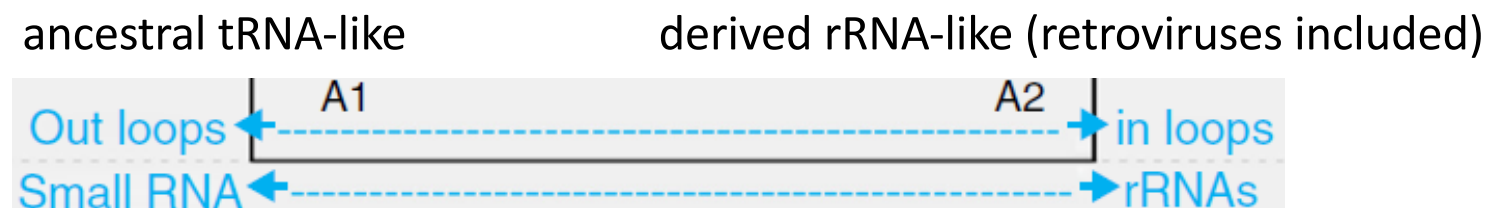
Interestingly, this order is on average recovered when averaging the positions of each AA species in a random sample of proteins (Seligmann 2014 Protein sequences recapitulate the genetic code's evolution. Comput Struct Biotechnol J 16, 177-180)

The tRNA->rRNA axis of RNA evolution

Classifying diverse RNAs by secondary structure produces 2 main RNA groups:
presumed primitive tRNA-like, and presumed derived rRNA-like (Seligmann & Raoult 2016)

Among others, theoretical 22-nucleotide RNA rings designed to mimick ancestral RNAs
(Demongeot 1978, Demongeot & Moreira 2007) are mainly tRNA-like (Seligmann & Raoult 2018)

Presumably, tRNA accretion produced rRNAs (Bloch et al 1983; Agmon 2009; Root-Bernstein & Root-Bernstein 2015; Barthélémy & Seligmann 2016), tRNA primitive->rRNA derived



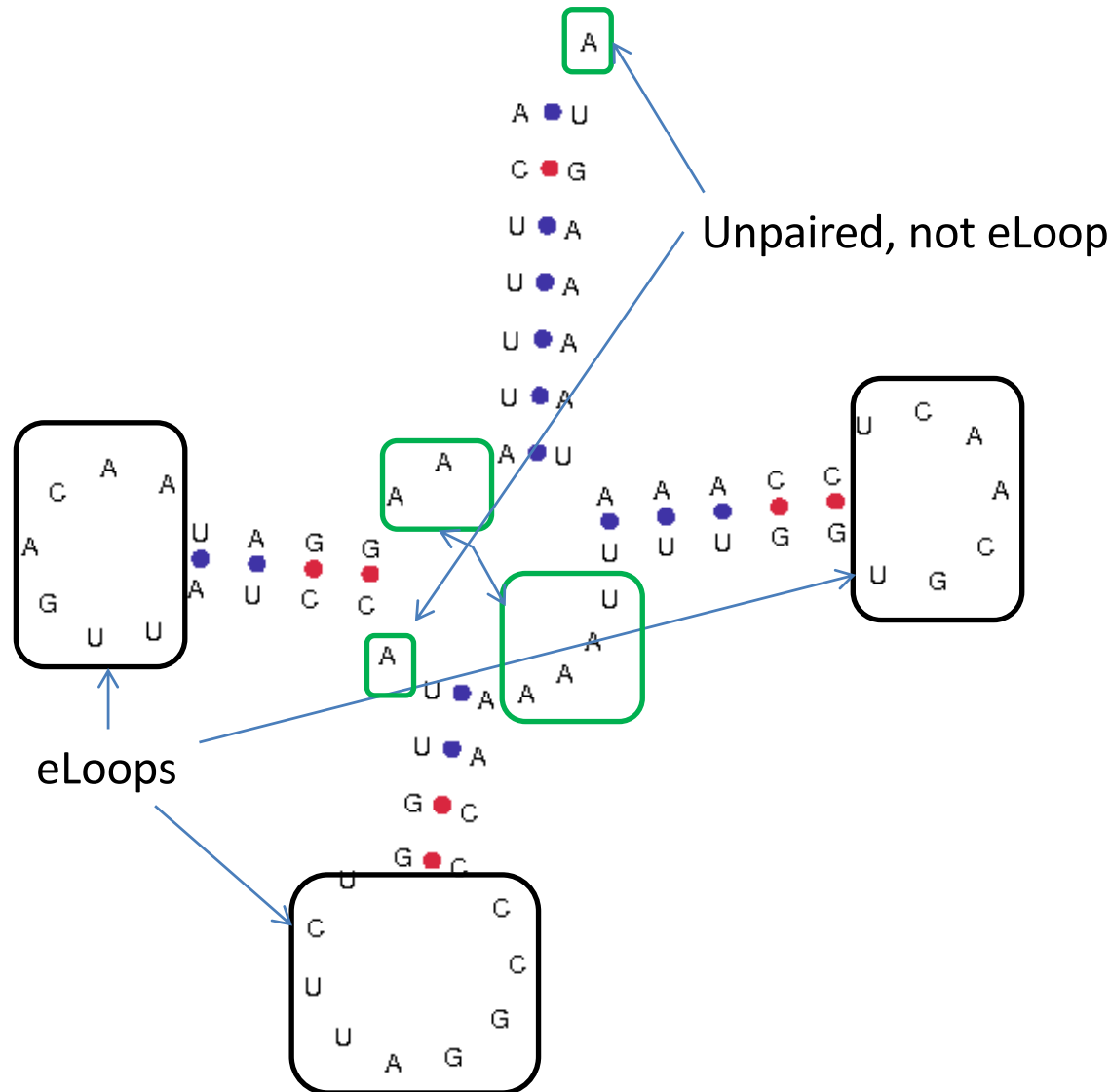
Each tRNA from the human mitogenome, Tupanvirus (Abrahão et al 2018), chloroplast, *Rickettsia monacensis*, *E. coli* and Yeast was scored along the tRNA->rRNA axis:
Its highest similarity to a tRNA-like RNA subtracted from its highest similarity to a rRNA-like RNA

Presumably, tRNAs with 'recent' cognate AAs should be more rRNA-like
Especially so for genetic code evolution hypotheses derived from other tRNA properties

Length: 71 nucleotides. 20 complementary nucleotide pairs, %loop = $100 \times (71 - 2 \times 20) / 71 = 43.66$

%loop (boxes) = $100 \times 23 / 31 = 74.2$

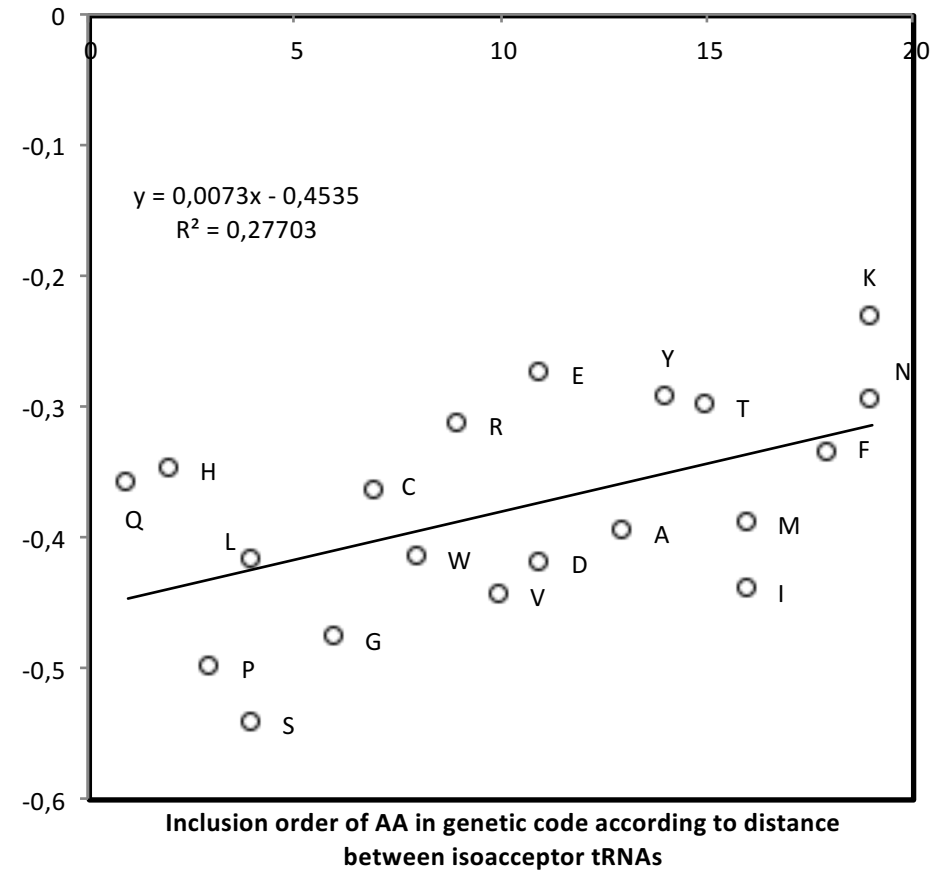
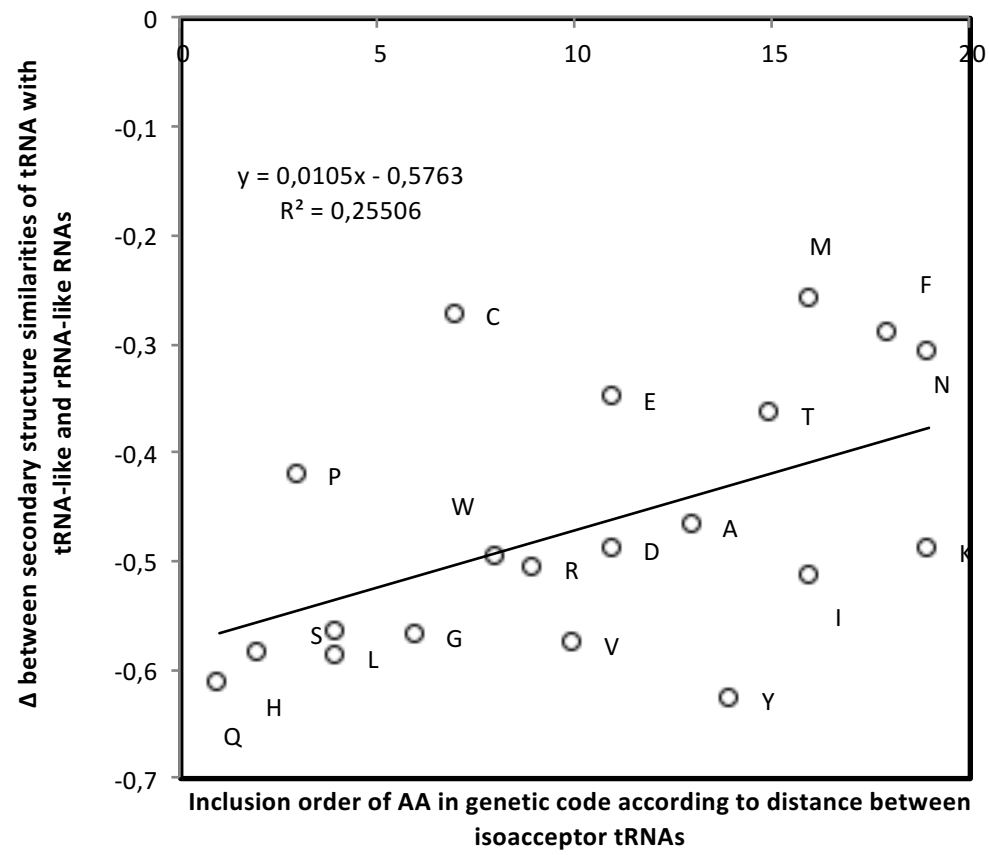
There are 7 G-C pairs in stems, hence %stem GC = $2 \times 7 / 40 = 35$ and 10 unpaired GCs, %loop GC = $10 / 31 = 32.26$



Correlations between tRNA-rRNA axis and AA inclusion order strongest for tRNA-derived hypotheses for *Tupanvirus*, mitochondria, *Nanoarchaeum*

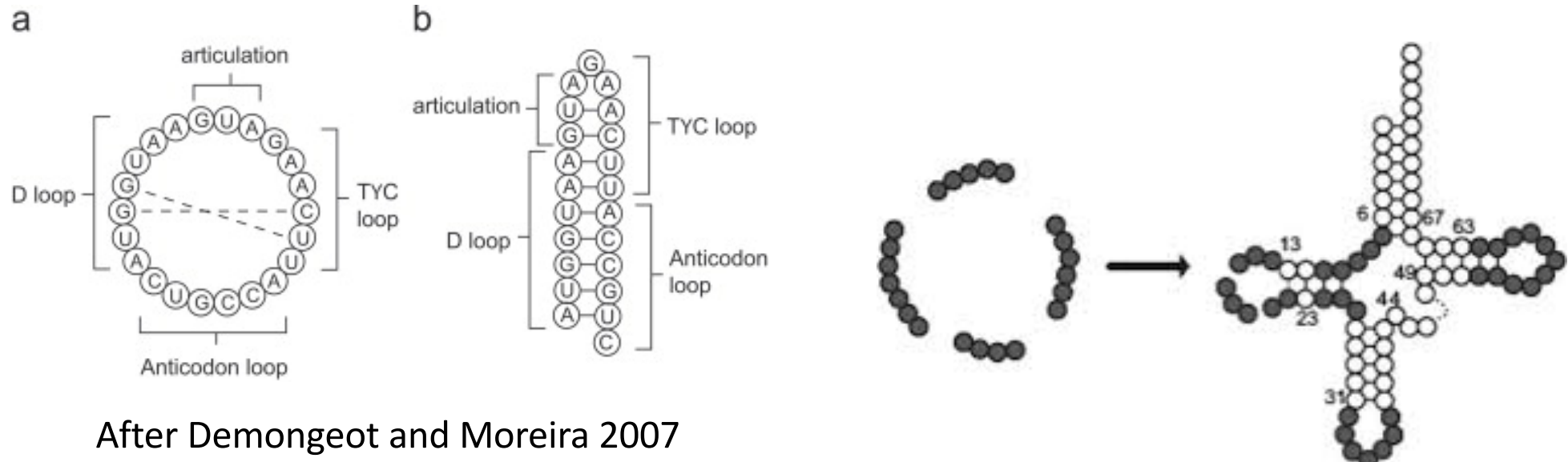
Mitochondrial tRNAs

Tupanvirus tRNAs



Theoretical RNA rings as proto-tRNAs

25 rings of 22 nucleotides, designed in silico, to form stem-loop hairpin, one codon per amino acid and a stop, presumed minimal ancestral gene. They have several realistic coding/structural properties. RNA ring consensus resembles tRNA consensus.



After Demongeot and Moreira 2007

Genome	+/42	P<0.05	Iso-t	Best	Best hypothesis
Chloroplast	31*	0	-0.18	0.37	Cavalier-Smith hypothesis
<i>E. coli</i>	25	2	-0.24	0.45**	Fox's proteinoid composition
Mitochondrion	33**	3	0.51**	0.51**	Evolutionary distance among isoacceptor tRNAs
<i>Nanoarchaeum</i>	18	1	0.55**	0.55**	Evolutionary distance among isoacceptor tRNAs
<i>Rickettsia</i>	22	1	-0.18	0.44**	Juke's neutral evolution theory
<i>Tupanvirus</i>	33**	15	0.53**	0.56**	Amino acid chemical inertness
<i>Yeast</i>	14-	3	-0.03	0.49**	Algebraic model of Hornos & Hornos
RNA rings	37**	10	-0.26	0.47**	Wong's metabolic coevolution theory

General remarks

Synteny between order of megaviral candidate homologues of mitochondrial genes and mitogene order suggests common ancestry/ies of mitogenomes and giant viruses

Association between tRNA-rRNA axis score of tRNAs and their cognate amino acid inclusion order in the genetic code confirms RNA evolution from tRNA-like to rRNA-like

Analysis of tRNA secondary structures according to the tRNA-rRNA axis are compatible with common ancestry/ies of mitogenomes and giant viruses

Theoretical RNA rings as a group seem realistic origin of life RNA population/network

Thanks to Jacques Demongeot, Eric Faure and Christian Michel for constructive discussions

References

- Abrahão J et al 2018 Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nat Commun* 9, 749.
- Agmon I 2009 The dimeric proto-ribosome: structural details and possible implications on the origin of life. *Int J Mol Sci* 10, 2921-2934.
- Barthélémy RM, Seligmann H 2016 Cryptic tRNAs in chaetognath mitochondrial genomes. *Comput Biol Chem* 62:119-32.
- Bloch DP, McArthur B, Widdowson R, Spector D, Guimaraes RC, Smith J 1983 tRNA-rRNA sequence homologies: evidence for a common evolutionary origin? *J Mol Evol* 19, 420-428.
- Cavalier-Smith T 2001 Obcells as proto-organisms: membrane heredity, lithophosphorylation, and the origins of the genetic code, the first cells, and photosynthesis. *J Mol Evol* 53, 555-595.
- Chaley et al 1999 Relationships among isoacceptor tRNAs seem to support the co-evolution theory of the origin of the genetic code. *J Mol Evol* 48, 168-177.
- Demongeot J 1978 Sur la possibilité de considérer le code génétique comme un code à enchaînement. *Revue de Biomaths* 62, 61-66.
- Demongeot J, Besson J 1983 Genetic-code and cyclic codes. *Comptes R Acad Sci III Life Sci* 296, 807-810.
- Demongeot J, Moreira A 2007 A possible circular RNA at the origin of life. *J Theor Biol* 249, 314-324.
- Fox SW, Waehneltd TV 1968 The thermal synthesis of neutral and basic proteinoids. *Bioch Bioph Acta* 160, 246-249.
- Hornos JEM, Hornos YMM 1993 Algebraic model for the evolution of the genetic code. *Phys Rev Lett* 71, 4401-4404.
- Juke TH 2000 The neutral theory of molecular evolution. *Genetics* 154, 956-958.
- Root-Bernstein M, Root-Bernstein R 2015 The ribosome as a missing link in the evolution of life. *J Theor Biol* 367, 130-158.
- Seligmann H 2018a Protein sequences recapitulate genetic code evolution. *Comp Struct Biotechnol J* 16, 177-189.
- Seligmann H 2018b Giant viruses as protein-coated mitochondria? *Virus Res* 253, 77-86.
- Seligmann H, Raoult D 2016 Unifying view of stem-loop hairpin RNA as origin of current and ancient parasitic and non-parasitic RNAs, including in giant viruses. *Curr Opin Microbiol* 31, 1-8.
- Seligmann H, Raoult D 2018 Stem-loop RNA hairpins in giant viruses: invading rRNA-like repeats and a template free RNA. *Front Microbiol* 9, 101.
- Trifonov EN 2000 Consensus temporal order of amino acids and evolution of the triplet code. *Gene* 261, 139-151.
- Trifonov EN 2004 The triplet code from first principles. *J Biomol Struct Dyn* 22, 1-11.
- Wong JTF 1975 A co-evolution theory of the genetic code. *Proc Natl Acad Sci U S A* 72, 1909-1912.