Article

The Tree of Life: Tangled Roots and Sexy Shoots
Tracing the genetic pathway from the first Eukaryotes to Homo sapiens

Chris King[*]

ABSTRACT

The picture conveyed by the significance of endosymbiosis, genome fusion and horizontal transfer as key evolutionary processes complementing the vertical transmission of the tree of life, makes clear that evolution is not just a matter of competitive survival of the fittest gene, individual, or species, but of dynamic survival of genes in a surviving ecosystem. Although Dawkins' ([1978](#)) notion of the "selfish gene" was pivotal in drawing attention to the fact that it was the survival of genes and not organisms, or even species, that was the key evolutionary process, attributing the human sentiment of selfishness to a gene is somewhat of a self-serving advertising distraction on the part of the author, which diminishes the subtlety and complexity of the sometimes apparently paradoxical ways genes actually interact to bring about beneficial outcomes in the evolutionary dynamics of the ecosystem. Although the idea of selection of genes has been pivotal in defining the need to consider evolutionarily stable strategies under genetic variation in ways which have been subsequently confirmed time and time again in situations such as the sexual genetics of social insects such as bees and ants, social selection is by no means ineffectual, or much of sociobiology, including the biological basis of morality as an extension of reciprocal altruism, would cease to exist. Moreover, from what we have seen, particularly about horizontal gene transfer, and the capacity of mobile elements to induce modulated changes in nuclear genomes, it is not the 'selfishness' of a genetic element alone that results in survival of both a gene and its hosts, but dynamic feedbacks,, and relationships which ultimately contribute to a massive sharing of information in the manner of parallel genetic algorithms fundamental to the replicative genetic process, which enable global forms of genetic and genome optimization central to the overall viability of life as complex systems.

Key Words: tree of life, genetic pathway, Eukaryotes, Homo Sapien, DNA, RNA, evolution.

Introduction

This article is a fully referenced research review to overview progress in unraveling the details of the [evolutionary](#) [Tree of Life](#), from life's first occurrence in the hypothetical RNA-era, to humanity's own emergence and diversification, through migration and intermarriage, using research diagrams and brief discussion of the current state of the art, as of 2009, in response to summaries of some of these developments (Lawton [2009](#), Lane [2009](#)).

The Tree of Life, in biological terms, has come to be identified with the evolutionary tree of biological diversity. It is this tree which represents the climax fruitfulness of the biosphere and the genetic foundation of our existence, embracing not just higher [Eukaryotes](#), plants, animals and fungi, but [Protista](#), [Eubacteria](#), and [Archaea](#), the realm, including the extreme heat and salt-loving organisms, which appears to lie almost at the root of life itself.
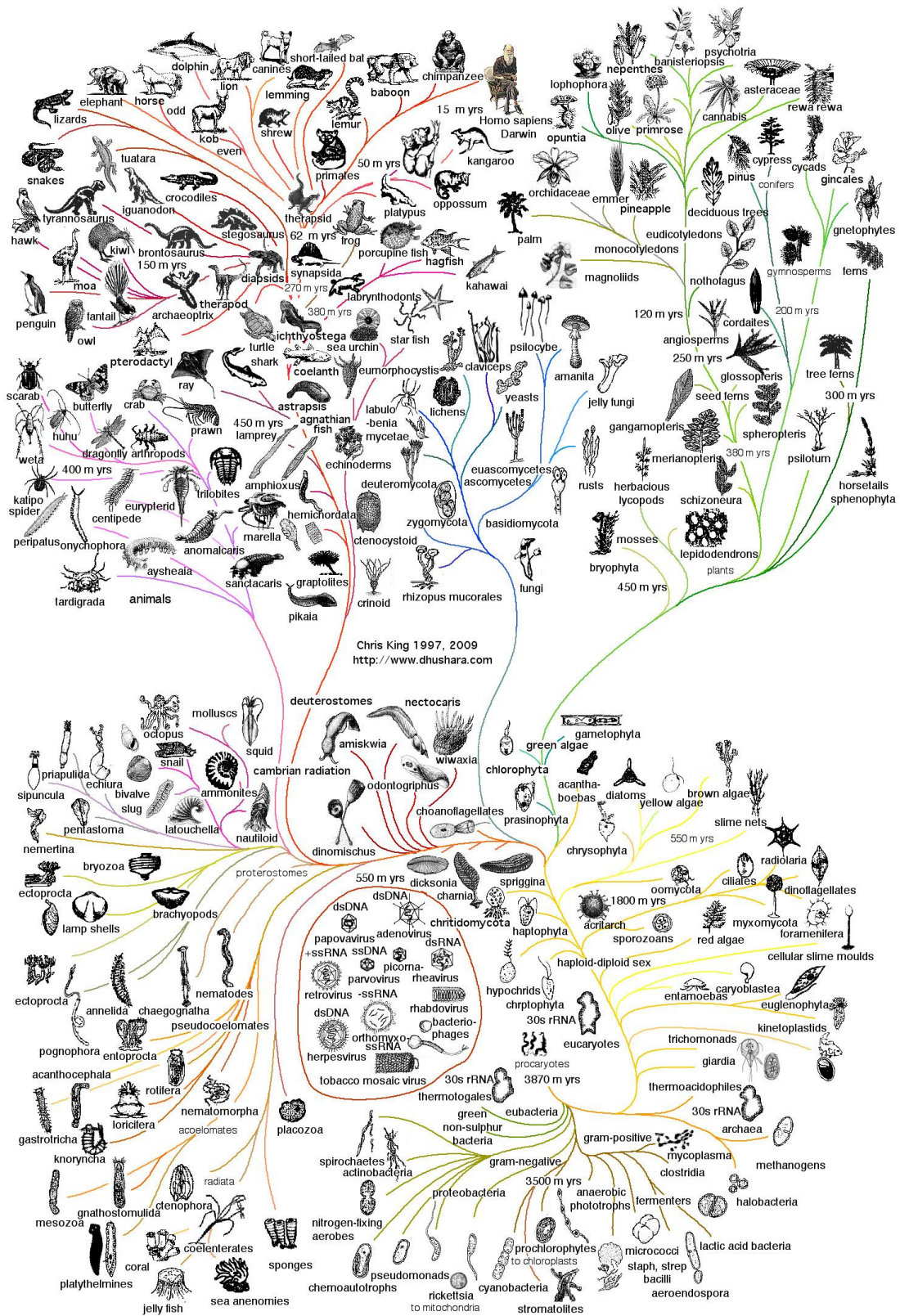
[*] Correspondence: Chris King http://www.dhushara.com E-Mail: chris@sexualparadox.org

Fig 1: The Tree of Life ([King](#) http://www.dhushara.com/book/unraveltree/life4CK.jpg)

The Universal Common Ancestor

Following a phase of biogenesis possibly based on cosmic symmetry-breaking (King 1978, 2004), based on new breakthroughs concerning RNA synthesis (Sutherland et. al. 2009) recent research suggests that the last universal common ancestor (LUCA) of all life on the planet may have arisen from a phase interface between alkaline hydrogen-emitting undersea vents and the archaic acidified iron-rich ocean (Martin and Russel 2003), giving rise to an active iron-sulphur reaction phase still present in living cells and associated with electron transport and some of the most ancient proteins, such as ferredoxin, in which differential dynamics in membranous micropores in the vents managed to concentrate polypeptides and polynucleotides to biologically sustainable levels (Baaske et. al. 2007, Budin et. al. 2009), giving rise to the RNA era, while at the same time providing a free energy source based on proton transport across membranous microcellular interfaces resulting from fatty acids also being concentrated above their critical aggregate concentration .
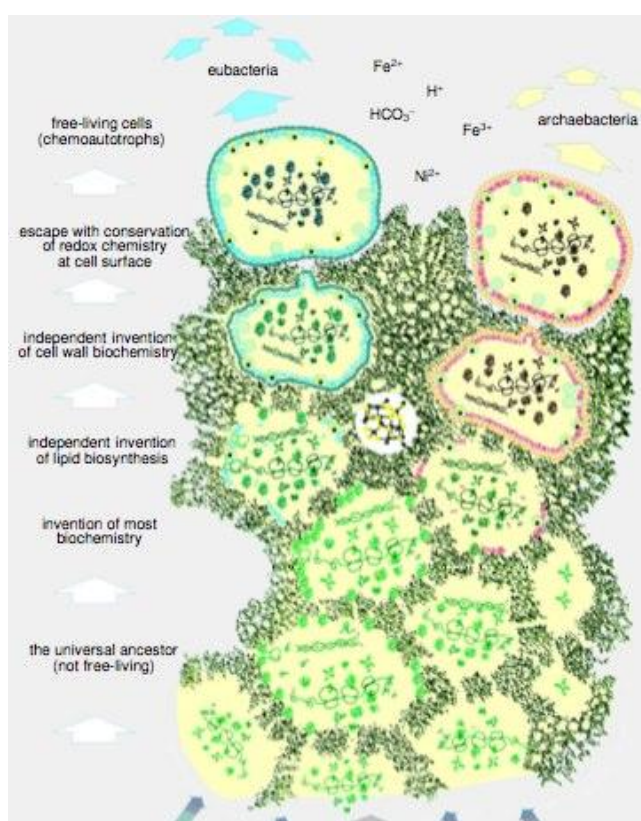

Fig 1a: Proposed scheme for the universal common ancestor (Martin and Russel 2003)

The universal common ancestor of the three domains of life may have thus been a proton-pumping membranous interface from which archaea and bacteria emerged as free-living adaptions. This is suggested by fundamental differences in their cell walls and other details of evolutionary relationships among some of the oldest genes.

It has also been proposed, on the basis of the highly-conserved commonality of transcription and translation proteins to all life, but the apparently independent emergence of distinct DNA replication enzymes in archaea/eucaryotes and eubacteria, that the last universal common ancestor had a mixed RNA-DNA metabolism based on reverse transcriptase, pinpointing it to the latter phases of the RNA era (Leipe et. al. 1999).

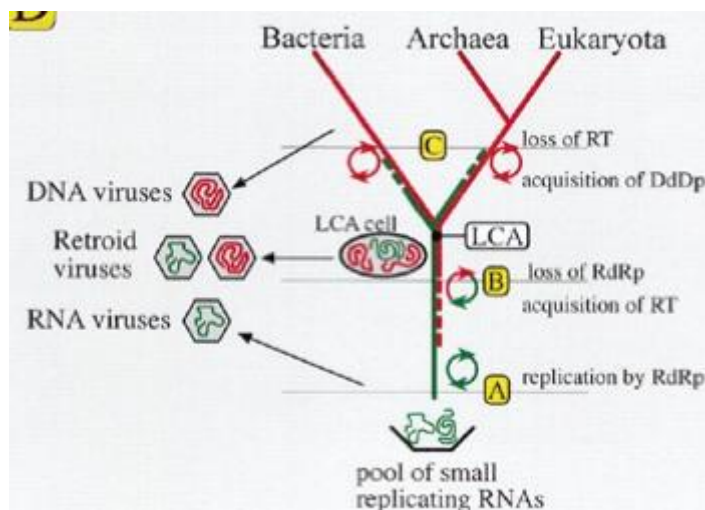Fig 1b: Hypothetical branching and evolution of RNA and DNA replication machinery (Leipe et. al. 1999).

Three Domains of Life

Life today is informationally based on the sequences of the four bases A, G, T and C in DNA, with messenger copies of the genetic sequence in mRNA forming intermediates in the assembly of proteins, as the cell's primary active chemical and structural agents. This is achieved through a process of translation at the ribosome - a supra-molecular complex composed of some 50 chaperoning proteins surrounding a core composed of three rRNA units, fed by amino-acid coupled tRNAs. The RNAs carry out the essential function, supporting the idea that translation was at first a purely RNA-based process of protein construction. In line with this and other RNA fossils found particularly in Eukaryotes, it is widely believed that life began based on RNA, which shares both the capacity for complementary replication of DNA and the formation of 3-dimensional chemically reactive conformations, similar to proteins, after which the ribosome evolved, transferring the reactive burden on to proteins sequenced through the genetic code. Some time later, the informational genome was consolidated into more stable DNA.
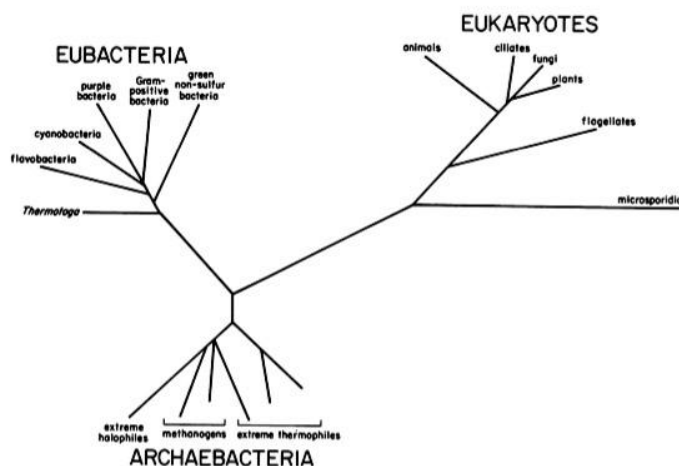

Fig 2: The initial tree of rRNAs shows three distinct founding domains (Woese 1987)

Originally the Bacteria and Archaea were thought to be one large diverse family of prokaryotes until Carl Woese (1977, 1978, 1987, 1990) and others investigated the evolutionary tree of ribosomal RNAs and found that there were three distinct founding evolutionary domains, then named eubacteria, archaebacteria along with the eukaryotes.

This gave the Eukaryotes a closer founding status as well, by contrast with the idea that the procaryotic bacteria came first and then, somehow the higher Eukaryote organisms with their complex cellular structures, including among others - the endoplasmic reticulum, along with the nuclear envelope and Golgi apparatus - all parts of a common complex of internal membranous partitions - and the architecture of microtubules, including centrioles, and the Eukaryote flagellum, as well as the Eukaryotes endosymbiont mitochondria and chloroplasts
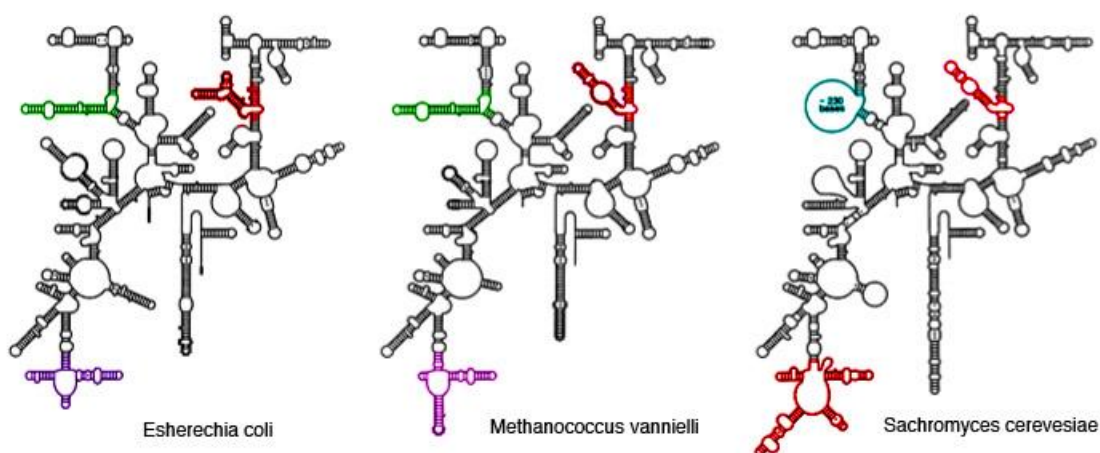

Fig 3: Key structural differences separating the larger rRNA units of the three domains (Woese 1987).

In addition to their evolutionary sequence divergence, the smaller 30s ribosomal RNAs of each domain, show distinct structural features characteristic of their own domain, but also emphasizing structural links between Bacteria and Archaea on the one hand and Archaea and Eukaryotes on the other, qualitatively confirming the central place of the Archaea in the divergence.
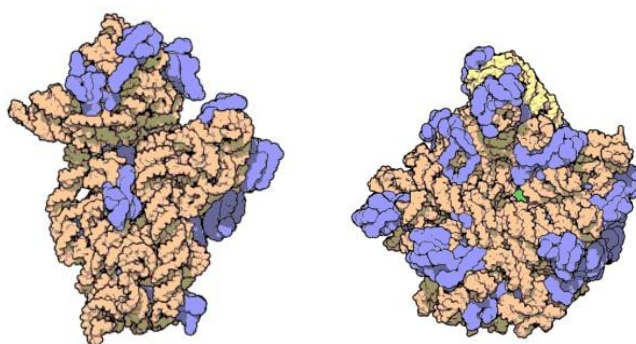

Fig 4: Small and large rRNA subunits of the eubacteia Thermus thermophilus and the archaeon Haloarcula marismortui. RNA orange and yellow, protein blue and active site green. (Wikipedia Ribosome)

The validity of the RNA-era concept and the capacity for RNAs to be both replicating informational and active ribo-enzymes is emphasized by the continuing dependence of the ribosome on rRNA

rather than the protein components is emphasized by the 3-dimensional realizations of the two subunits in fig 4, which show that the rRNA molecules are still carrying out the central task of protein assembly with only minor modification due to the 'chaperoning' proteins, despite 3.8 billion years of evolution.

Norman Pace subsequently enlarged the scope and accuracy of the rRNA tree, including a greater diversity of organisms. This tree has become the basis of several other studies (see e.g. fig 10).

The Copernican principle asserts that the Earth is a typical rocky planet in a typical planetary system, located in an unexceptional region of a common barred-spiral galaxy, hence it is probable that the universe teems with complex life. This is supported to a reasonable extent by the discovery of an increasing number of planets including some putative "Goldilocks" zone planets where water would be liquid and life as we know it could potentially exist. Set against this, the Rare Earth hypothesis argues that the emergence of complex life requires a host of fortuitous circumstances including a galactic habitable zone, a central star and planetary system having the requisite character, the circumstellar habitable zone, the size of the planet, the advantage of a large satellite, conditions needed to assure the planet has a magnetosphere and plate tectonics, the chemistry of the lithosphere, atmosphere, and oceans, the role of "evolutionary pumps" such as massive glaciation and rare bolide impacts, and whatever led to the still mysterious Cambrian explosion of animal phyla. This might mean that planets able to support a bacterial level of life are not so uncommon, but those supporting complex multicellular life might be.
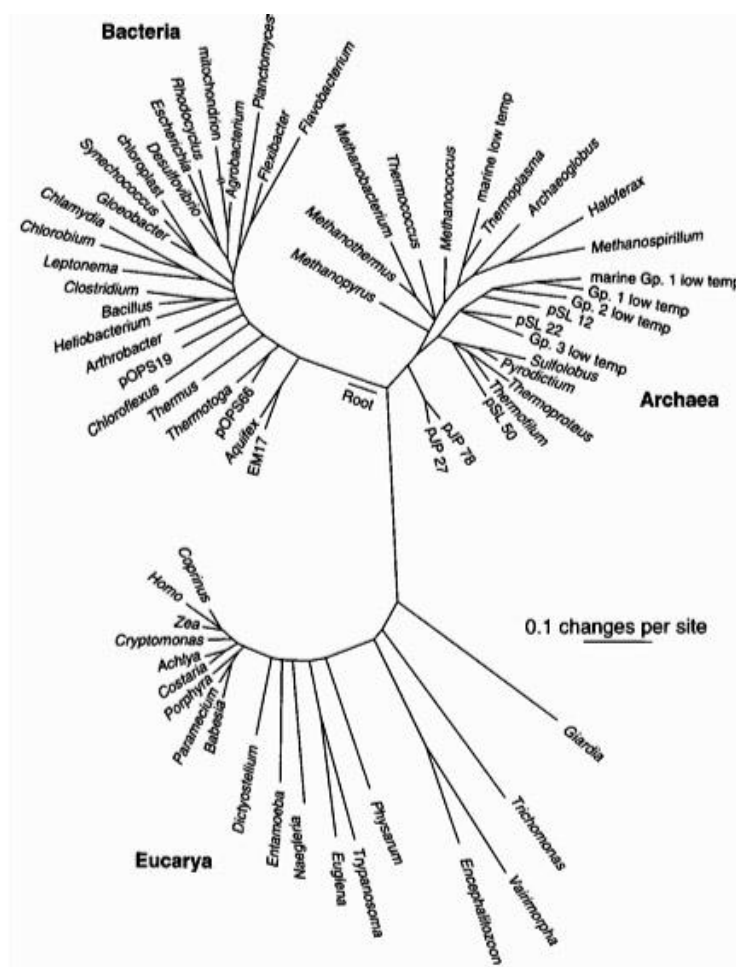

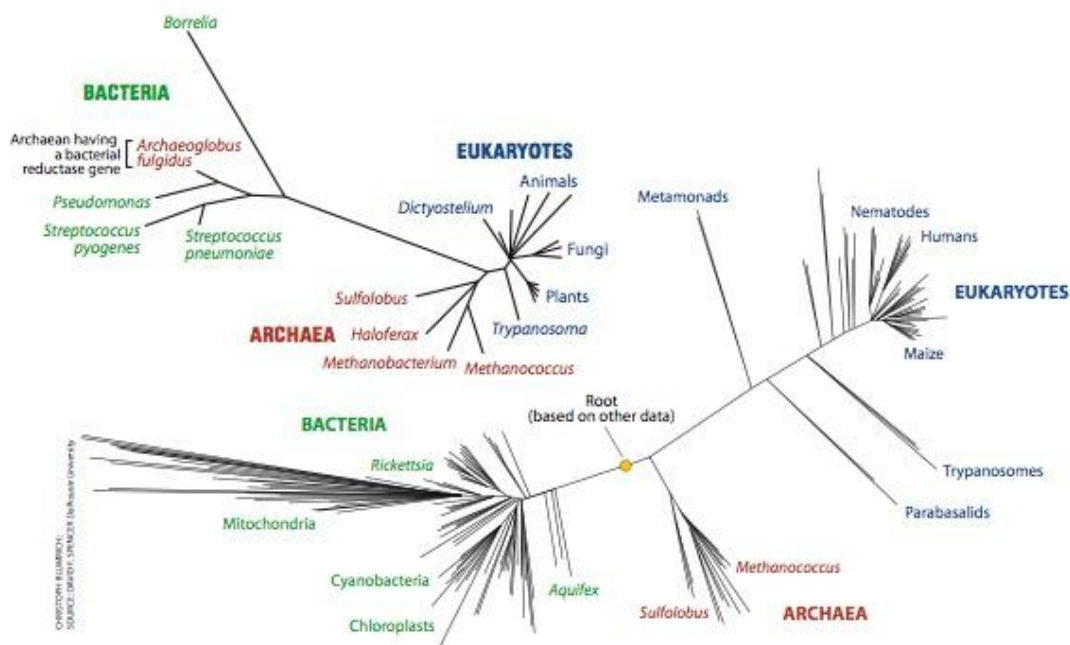Fig 5: Further elaboration of the rRNA tree (Pace 1997)

Fig 6: Lower right: A third rRNA tree which suggests Archaea lie very close to the root is contrasted with that for the enzyme HMGCoA reductase, which also shows evidence of horizontal transfer to an Archaean (ex Doolittle 2000).

Bringing this question to a pivotal crux in our context, the emergence of mitochondria as endosymbionts has been proposed to be a critical bottleneck which allowed complex life to evolve only once, because, only in this effectively fractal cellular architecture, can the membrane surface areas necessary to support the chemical reactions enabling the vastly larger number of genes in a complex organism's genome to maintain metabolic stability (Lane and Martin 2010). Whether such endo-symbiosis is rare. or a common extreme of parasitic relationships would then determine how likely or unlikely complex life might be.

Offset against both the uniqueness of the mitochondrial endo-symbiosis and the closely linked, but independent question of the origin of the nucleus and nuclear envelope, has been the discovery of mimiviruses and mamaviruses infecting amoeba (Raoult et, al. 2004) and related very large aquatic viruses such as CroV infecting single celled plankton species, which despite their recent discovery, appear from ocean gene analyses to be potentially ubiquitous and widespread in the oceans and possibly playing a crucial role in regulating the atmospheric-oceanic pathways, such as carbon sequestration (Fisher, Allen, Wilson and Suttle 2010). These form an intermediate genetic position between viruses and cells, having the largest genomes, with extensive cellular machinery and larger than the smallest completely autonomous bacterial and archaeal genomes.

As an illustration of genes in mimivirus normally appearing only in cellular genomes, the mimivirus has genes for central protein-translation components, including four amino-acyl transfer RNA synthetases, peptide release factor 1, translation elongation factor EF-TU, and translation initiation factor 1. The genome also exhibits six tRNAs. Other notable features include the presence of both type I and type II topoisomerases, components of all DNA repair pathways, many polysaccharide synthesis enzymes, and one intein-containing gene. Inteins are protein-splicing domains encoded by mobile intervening sequences (IVSs). They self-catalyze their excision from the host protein, ligating their former flanks by a peptide bond. They have been found in all domains of life (Eukaria, Archaea, and Eubacteria), but their distribution is highly sporadic. Only a few instances of viral inteins have been described. Self-splicing type I introns are a different type of mobile IVS, self-

excising at the mRNA level. They are rare in viruses. Mimivirus exhibits four instances of self-excising intron, all in RNA polymerase genes.
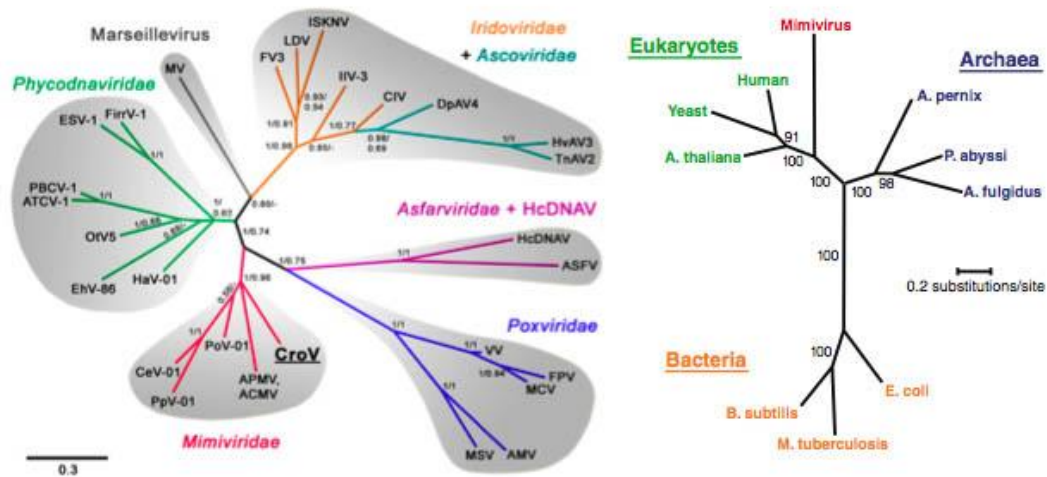
Evolutionary diversification of Mimiviruses from nucleocytoplasmic large DNA viruses (Fisher et. al.) and in relation to the three domains of cellular life based on the concatenated sequences of seven universally conserved protein sequences (Raoult et. al.)

Mimiviruses also host parasitic virophages, affectionately named sputnik as viral satellites, which piggy back on the metabolism of the large viral factories set up by these giant viral genomes causing the mimiviruses to sicken, and these virophages also contains genes that are linked to viruses infecting each of the three domains of life Eukarya, Archaea and Bacteria (La Scola et. al. 2008). It has thus been suggested that they have a primary role in the establishment of cellular life and that they may have been instrumental in the emergence of the nuclear envelope.
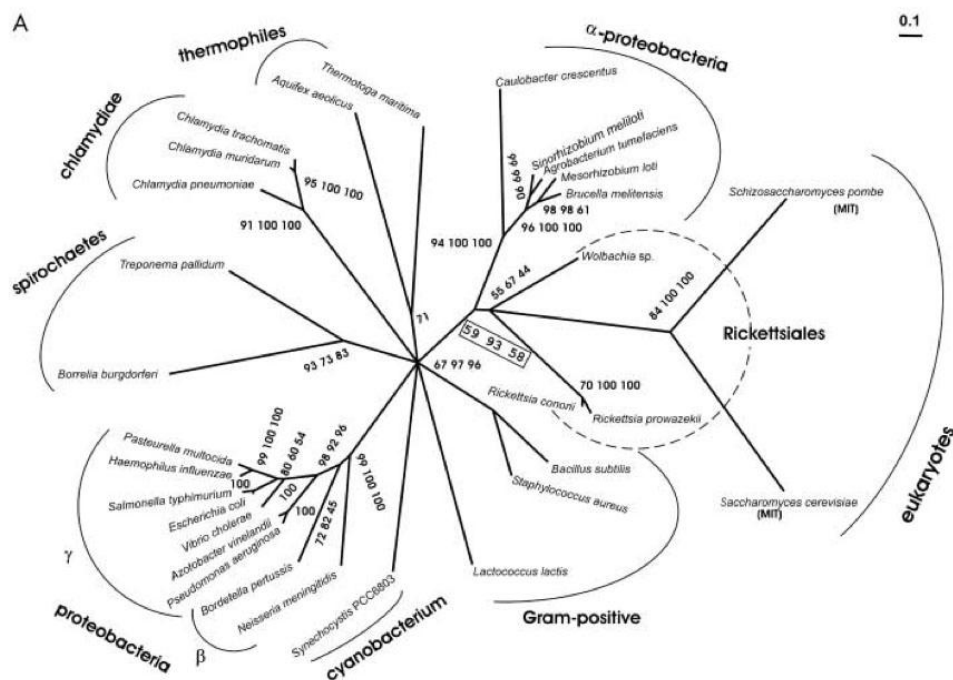

Fig 7: Evolution of iron-sulphur cluster proteins of mitochondria is linked to α-proteobacteria (Emelyanov 2003)

Tangled Roots of Horizontal Transfer

Despite the division into three domains, further investigations of proteins in the three domains began to reveal a much more confused and complicated picture. Firstly the ribosomal proteins, like the rRNAs show distinct, easily differentiated morphologies with some correspondences linking one pair of domains and other another pair (Forterre 2006b, Woese 2000). Secondly, the proteins in Eukaryotes appear to have a mixed origin with the informational ones having an evolutionary relationship with Archaea but the metabolic enzymes appearing to have a bacterial origin. This suggests that the Eukaryote genome has either resulted from one, or more symbiotic fusions e.g. an Archaeal and a bacterial genome and/or that there has been a high degree of horizontal gene transfer between bacteria and Eukaryotes.

The evidence for symbiotic inclusions is clear from the fact that all Eukaryotes, except for a few primitive anaerobic varieties, such as the metamonad human gut parasite Giardia lamblia, all have endosymbiotic respiring mitochondria, which are evolutionarily related to α-proteobacteria such as Rickettsiae (Emelyanov 2003). Plants also have photosynthetic chloroplasts derived from cyanobacteria. α-proteobacteria, including Rickettsiae (and related Wollbachia and Agrocbacterium), obligately live in the cytoplasm of other cells and so are naturally adapted to becoming an endo-symbiont of a glycolytic organism by providing respiring energy to the host's metabolism resulting in the mitochondrion. Giardia still retains traces of mitochondrial proteins so appears to have lost its respiring organelles, rather than occupying a place in the tree before mitochondria were incorporated into eucarya (Adam 2000).

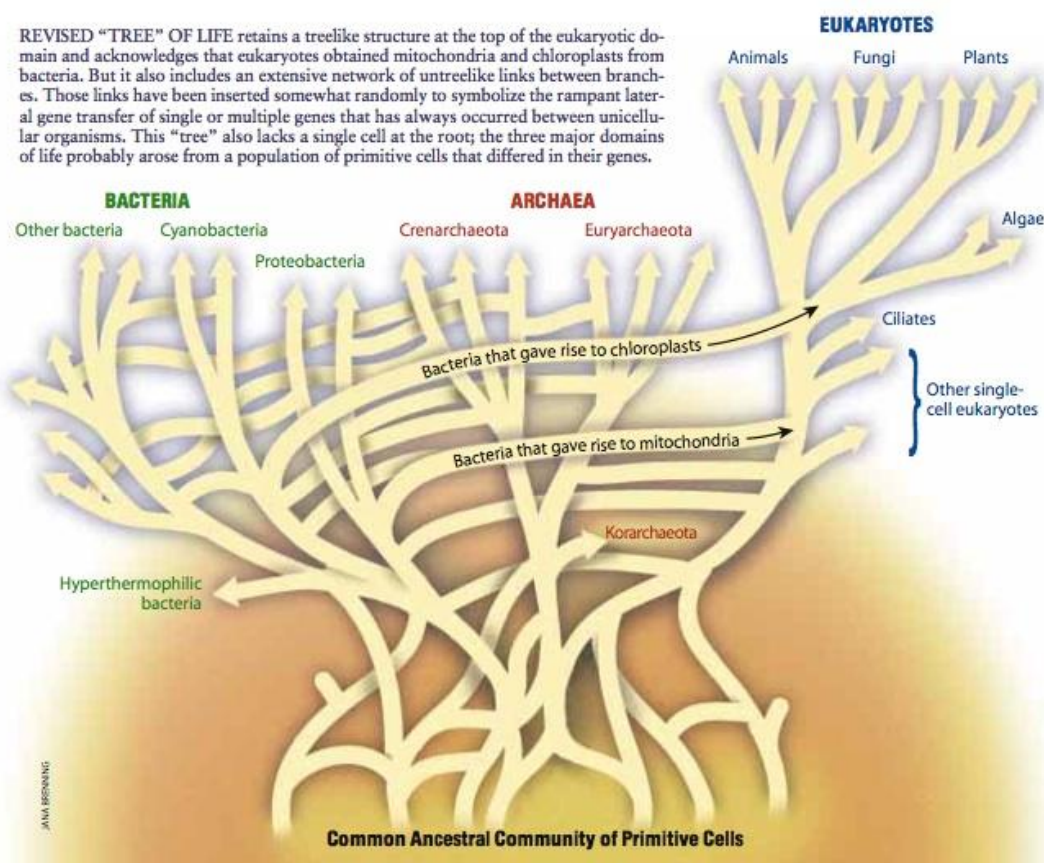Fig 8: Tangled roots (Doolittle 2000)

The picture of horizontal transfer is even more tangled in bacterial and archaeal genomes, which contain a great number of shared and exchanged genes, probably through viral transfer.

This has led generally to Doolittle (1998), Woese (2002), and others, proposing a tangled root to the tree of life involving a transition from a regime in which thre was a much higher rate of horizontal exchange and effective global optimization of genomes, to tree-like vertical evolution of genomes, once the more complex genomes of the Eukaryote domains became established.

Subsequently, to try to clarify the taxonomic relationships founding the tree of life, Peer Bork and his team produced a refined evolutionary tree by selecting only universal proteins that had not been subjected to horizontal transfer, providing the most detailed tree root diagram to date, although admittedly on only a skeleton gene set comprising some 1% of the respective genomes. The phylogenetic tree has its basis in a cleaned and concatenated alignment of 31 universal protein families and covers 191 species whose genomes have been fully sequenced.
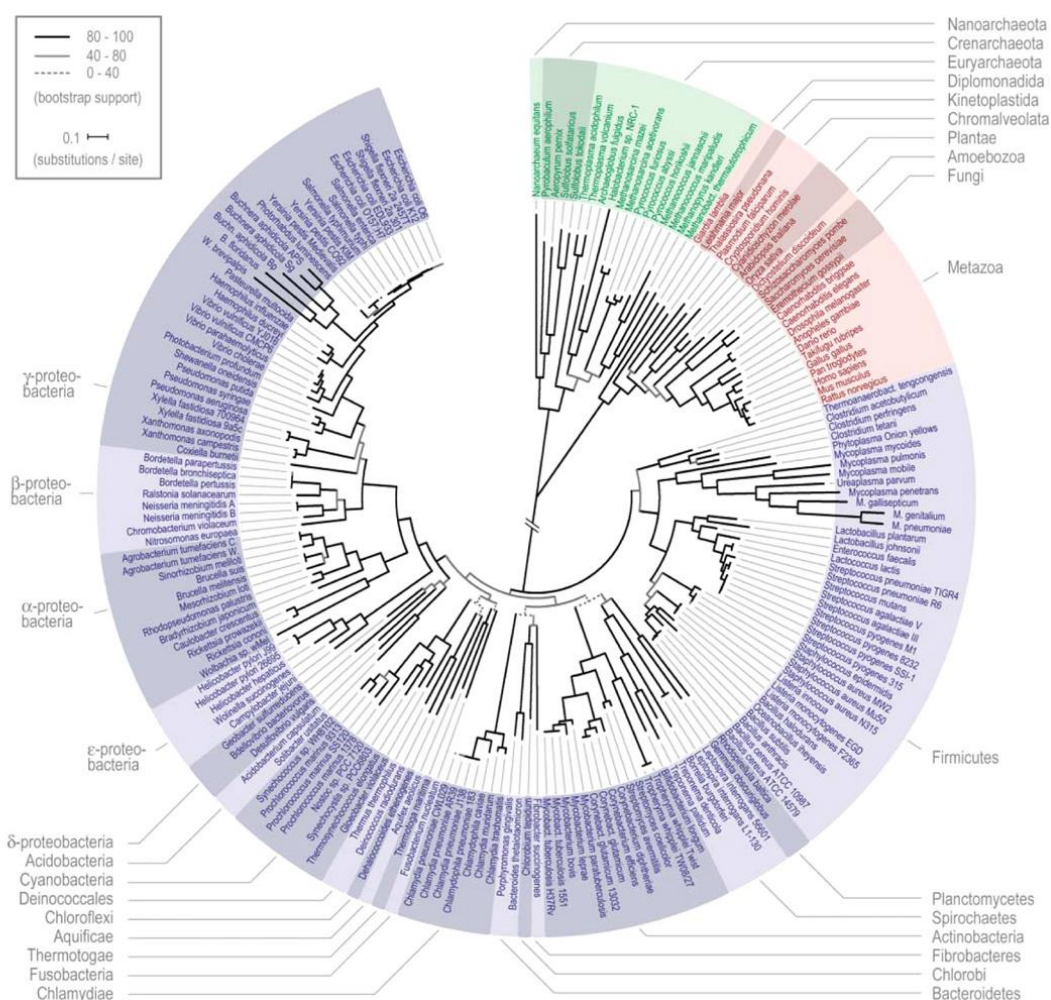

Fig 9: High resolution tree of the three domains of life (Ciccarelli, Bork et. al. 2006).Purple eubacteria, green archaea, red Eukaryotes.
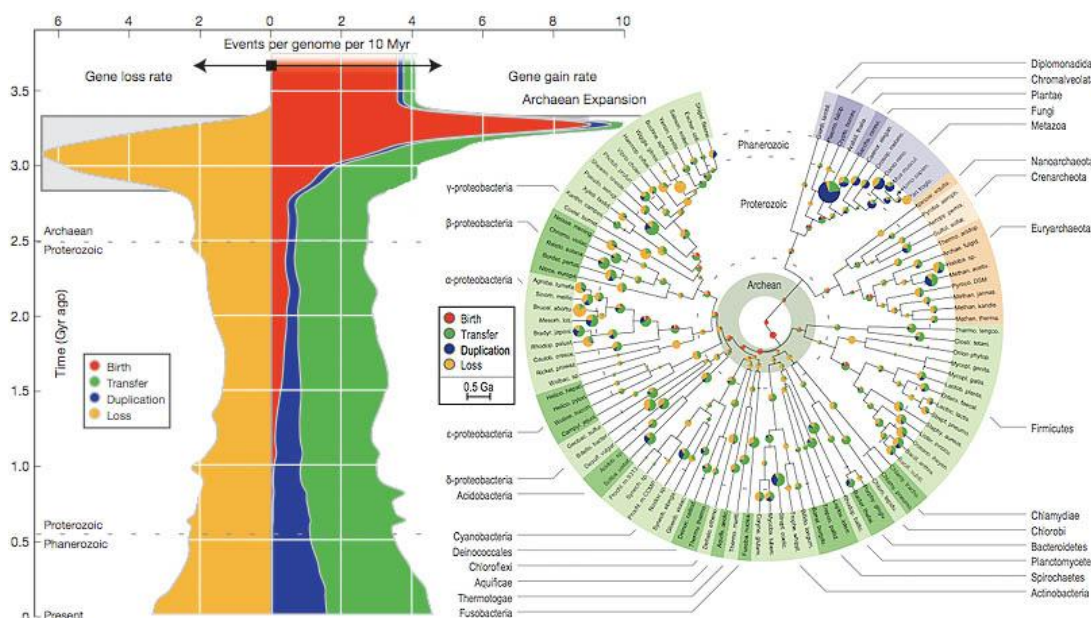
Fig 9b: Tree diagram of the birth, transfer, duplication and loss of key genes in the redox and electron transport pathways, in a founding burst of gene evolution between 3.3 and 2.7 billion years ago (David and Alm 2010).

Then Lawrence David and Eric Alm (2010) produced the above tree investigating the central genes common to a wide spectrum of life forms, involving the founding steps of redox reactions and electron transport, demonstrating a rapid evolutionary innovation during an Archaean genetic expansion between 3.3 and 2.7 billion years ago. They mapped the evolutionary history of 3983 gene families that occur in a wide range of modern species. They were able to show that 27 per cent of these gene families appeared in a short evolutionary burst. Many of the genes from this time were involved in electron transport - a key step in respiration and photosynthesis, which ultimately led to oxygen-producing photosynthesis and the "great oxygenation event" 2.4 billion years ago, when the atmosphere became oxygen rich.

This lends support to the idea that the collective primordial genome functioned as a supercomputer (King 2010) based on parallel genetic algorithms combined with horizontal genetic transfer, whose bit computation rate through mutation and recombination is sufficient to generate the functional conformations, through protein folding, to solve the key metabolic pathways over a period no longer than 300 million years.

Bacteria engage in much more radical forms of pan-sexuality than higher organisms, involving viruses and plasmids, themselves separate mobile genetic elements, acting as agents of genetic transfer, accelerating the pace of bacterial evolution (Maxmen 2010). This enables the genetic sequences of bacteria, archaea and protists to move around in the genome and to be exchanged between cells, and even between different species. Sexual exchange of material can happen both through viral exchange and through a conjugation plasmid, which can spool DNA from one bacterium into another, resulting in a net donation of genes from one strain or species to another, which ensures a broad exchange of genetic material throughout bacterial ecosystems, resulting in rapid accumulation of advantageous genes exemplified by plasmid borne infectious drug resistance.

To give a very rough idea of the computing power of the combined bacterial genome alone, taking into account bacterial soil densities ($\sim 10^9$/g), effective surface area ($\sim 10^{18}$ cm^2), genome sizes ($\sim 10^6$), combined reproduction and mutation rates ($\sim 10^{-3}$/s) gives a combined presentation rate of new combinations of up to 10^{30} bits per second, roughly 10^{13} times greater than the current fastest computer at 2 petaflops or about 10^{17} bit ops per second. Corresponding rates for complex life forms would be much lower, at around 10^{17} per second because they are fewer in total number and have lower reproduction rates and longer generation times, but they are still vying with the computation rates of the fastest supercomputer on earth.
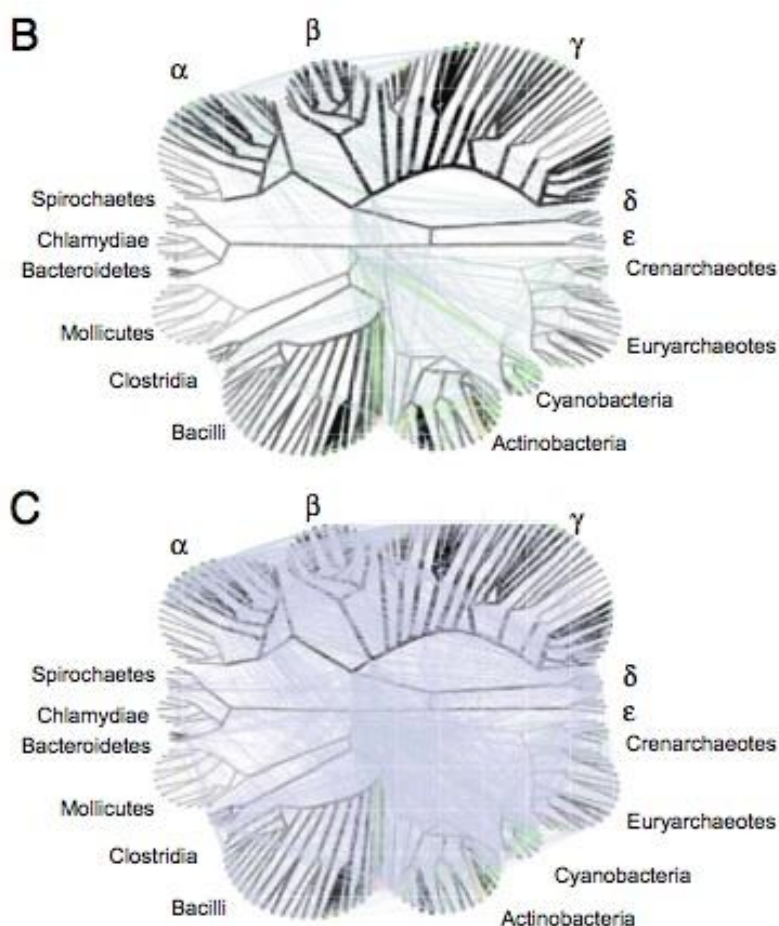

Fig 10: Horizontal transfers across the bacterial tree under two thresholds 10, 5 genes (Dagan et. al.).

This picture of bit rates coincides closely with the Archaean expansion scenario noted above and suggests that evolution has been a two-phase process in which the much higher bit rates of the collective single-celled genome under promiscuous sexuality and horizontal transfer has arrived at a global genetic solution to the protein folding problems of the central metabolic, electro-chemical and even root developmental pathways, which are then later capitalized on by multi-celled organisms, through gene duplication and loss as well as the creation of new specialized genes at a much lower rate.

The massive extent of horizontal transfer in eubacteria, as well as archaea has also become clear suggesting large components of procaryote genomes are effectively globally optimized for their niches by frequent genetic transfer. Dagan et. al. have characterized the extent of horizontal transfer

for a series of thresholds as well as establishing specific modularity of horizontal transfer of functions   between
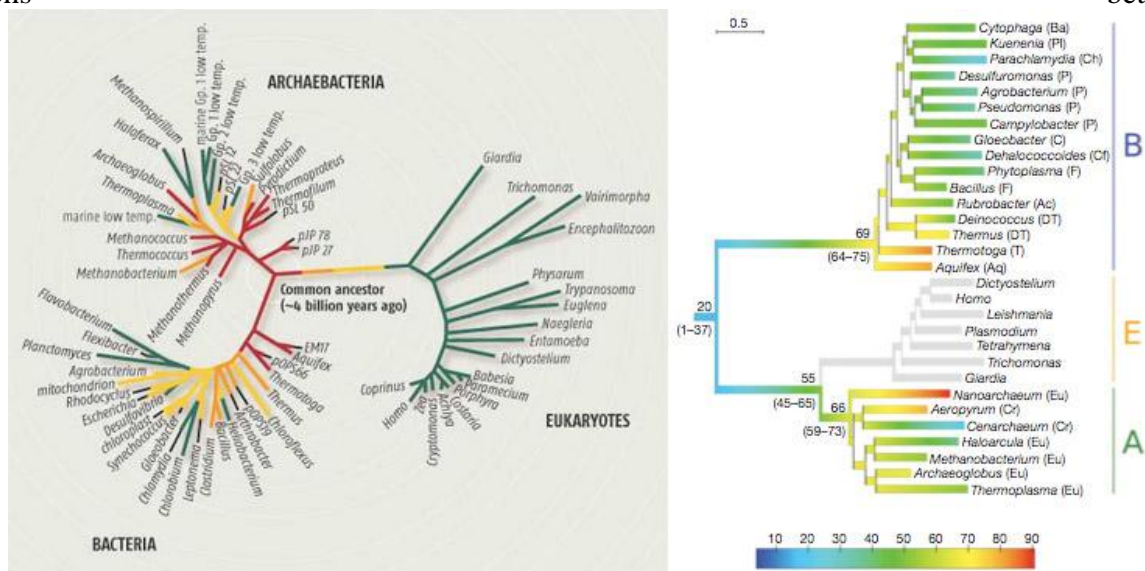
groups.

Fig 11: Evolutionary root of the tree of life and its diversification into archaea, bacteria and Eukaryotes appears to have gone through an early period of cool temperature consistent with an RNA era, followed by a hot period (Anathaswamy, Boussau et. al.)

Surviving Archaea are known to inhabit extreme environments, including hot volcanic pools, hydrothermal vents and extreme salty environments and several arrangements of the root of the tree, including Bork's team's work suggest a hot origin for life. However other research (Brochier and Philippe, Boussau et. al.), concludes the base root may have been at about 25°C, a more viable temperature for a simple RNA metabolism, with a succeeding period of high temperature adaptions shortly after the differentiation of the three domains in evolutionary time.
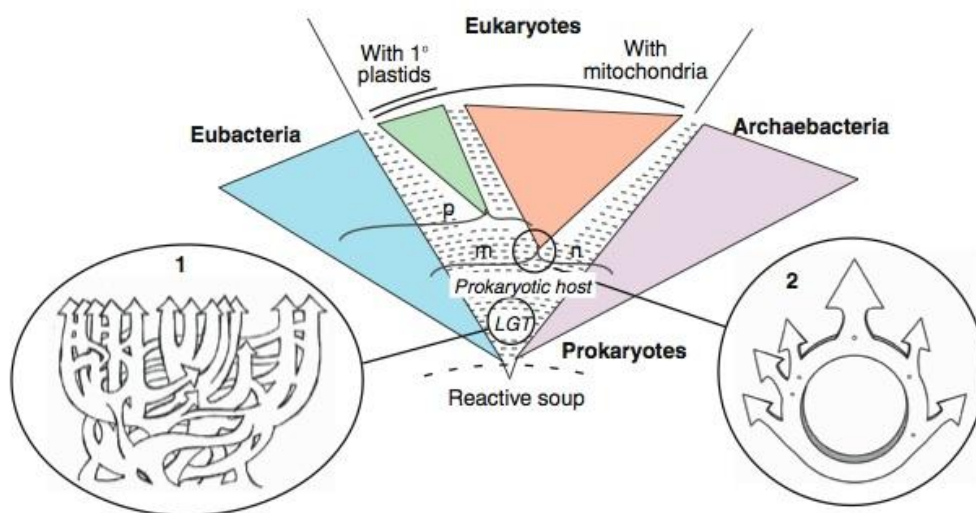

Fig 12: Genetic diffusion at the root of the tree (Dagan and Martin)

Critics of the validity of the tree root concept, such as Dagan and Martin emphasize the small proportion (1%) of the genome used in Bork's study and stress both the lateral (or horizontal) gene transfer events uniting the prokaryote realms and the apparent chimaeric nature of the Eukaryote

genome, which appears to contain both archaea-related informational genes and eubacterial metabolic ones, in addition to obvious endosymbiont events of the mitochondrion and chloroplast.
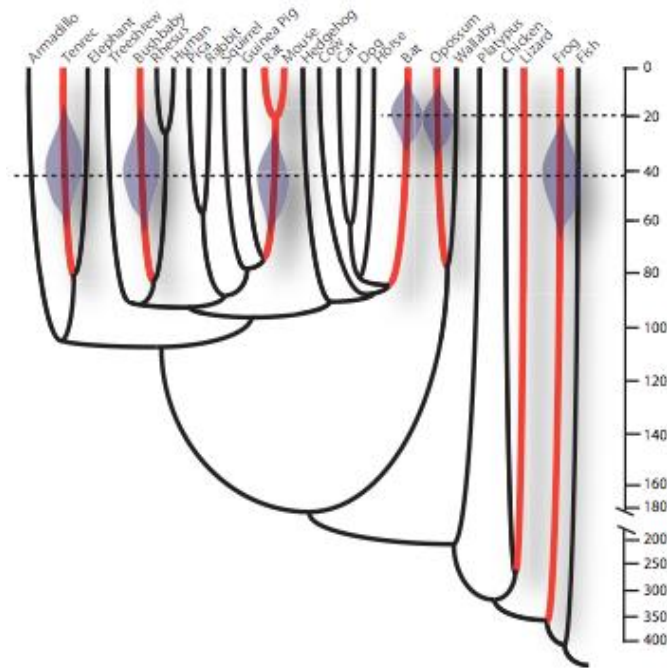

Fig 13: Pattern of invasions of the SPIN element (Lisch)

The case for horizontal transfer of genes between unrelated Eukaryote species through infectious elements invading new and hence non-resistant species is also well established. The SPIN element is present in a diverse unrelated set of species, spanning amphibians, reptiles, marsupials and mammals while absent from closely related species (Lisch).

The Eukaryote Nuclear Genome as a Genetic Fusion

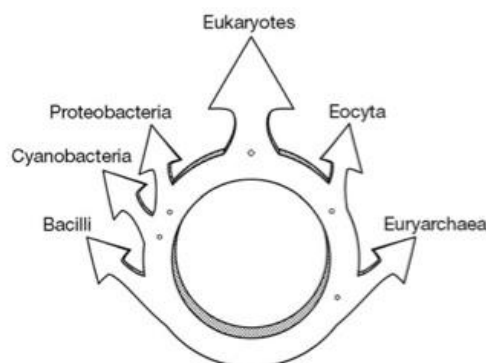Fig 14: The proposed ring of life (Rivera and Lake)

In a further study, Rivera and Lake used a new algorithm to take account of possible genetic fusion events by forming a genetic ring through matching partial trees into a most parsimonious whole, inferring that the Eukaryote genome has arisen from a fusion of an archaeal (possibly eocyte) genome with that of either a cyanobacterium or possibly a γ-proteobacterium.

The method used cannot definitively determine whether or not the eubacterial genome could have come from the mitochondrial event, which, to an even greater extent than the more recent chloroplast, has resulted in a high net transfer of genes from the mitochondrial chromosome to the nucleus, leaving open the possibility that in addition to the mitochondrial symbiosis, and the later chloroplast one, there may have been an additional genetic fusion. Lane and Archibald have cited further major endosymbiosis events involving complex three genome interaction in protista, where both green and red algae have been incorporated by endosymbiosis into other protists, which demonstrate both that endosymbiois has occurred many times and the genomic complexity of nuclear symbiont gene exchange.
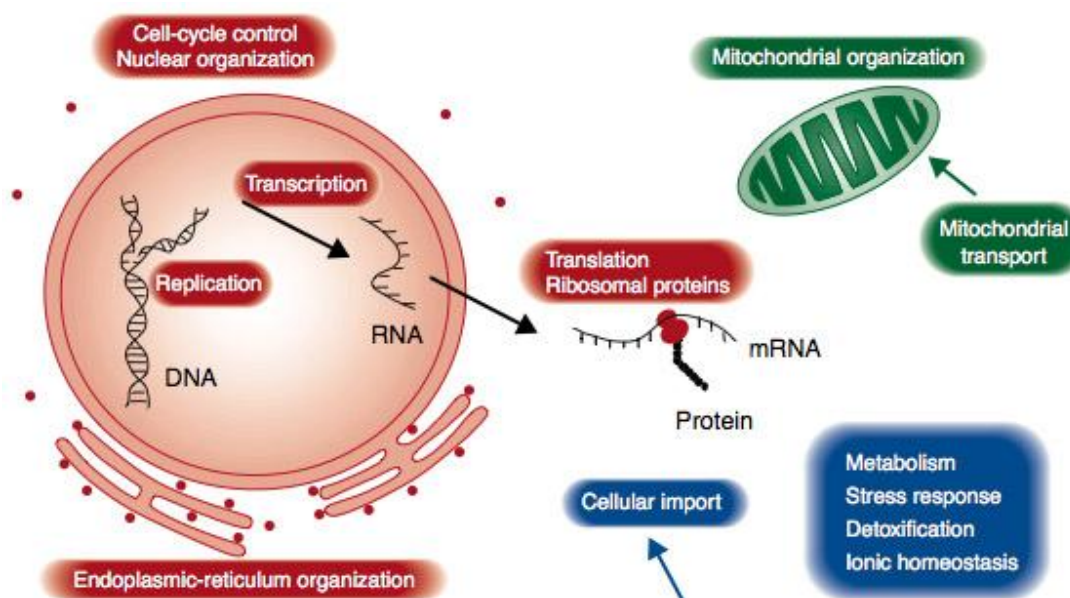

Fig 15: Proposed fusion between two genomes - informational, from Archaea (red), and metabolic, from Eubacteria (blue) as well as mitochondrial genes migrating to the nucleus (green) (Horiike et. al.). Up to 75% of nuclear genes whose ancestry has been elucidated may come from bacteria (Lane).

The idea of a genetic fusion between an archaea and a γ-proteobacterium is also supported by several other lines of research including evolution of glycolytic enzymes (Emelyanov), 'homology-hit analysis' of non-mitochondrial genes determined the number of yeast orthologous ORFs in each functional category to the ORFs in six Archaea and nine Bacteria at several thresholds, suggesting an archaeal parasite engulfed by a eubacterium (Horiike et. al.) the proposal that close association between a central methanogenic archaebacterium (archaea) and a close-knit surrounding clump of ancestral sulfate-respiring δ-proteobacteria could have also led to the nucleus and endoplasmic reticulum (Moreira and Lopez-Garcia).
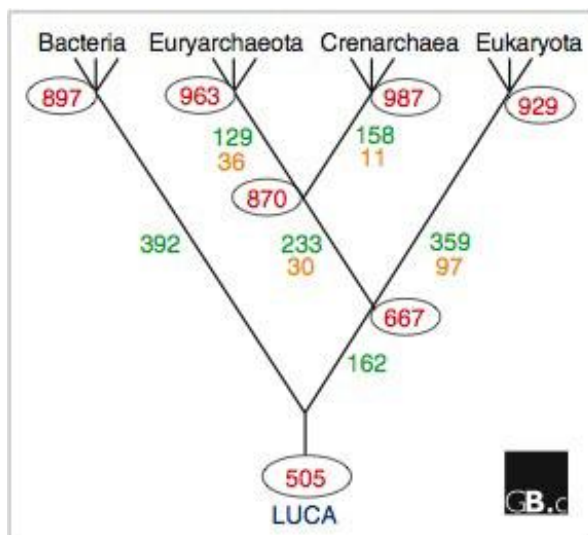
The most parsimonious scenario for the evolution of the main lineages of life. The red numbers in ovals near the internal nodes show the size of the reconstructed gene sets of the respective ancestral forms. Green numbers show gene gains and brown numbers gene losses assigned to each of the branches in the tree. LUCA, last universal common ancestor.

Fig 16: Gene replacement tree root (Makarova) Hartman and Federov's list of putative chronocyte genes correspond to the 359 above.

Other theories stick to the three domain paradigm and propose that a primitive Eukaryote precursor possibly still retaining an RNA-based genome, as suggested by Woese (1998) might be the case for the progenote (first root life form) to be the last universal common ancestor of all three domains, possibly including genes for endoplasmic reticulum and microtubules, engulfed both and archaea and a eubacterium. Hartman and Federov cite a collection of such genes, including those for ribosomal proteins as well, naming the organism as a chronocyte.

This is also consistent with the much greater complexity of use of RNA in Eukaryotes, including alternative splicing, the use of introns, interfering-RNAs in gene regulation, micro-RNAs and the use of the nucleus to contain a diversely functioning RNA informational metabolism not unlike that of a putative progenote.
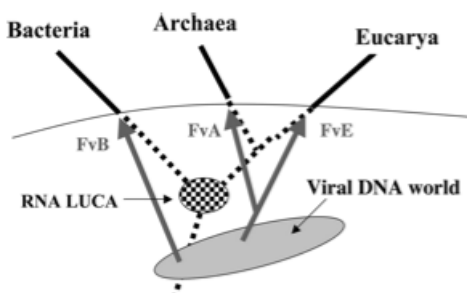
Viral Influences on the Nuclear Genome

Fig 17: Proposed viral contribution of DNA polymerases FvA etc. founder viruses (Forterre). Other cellular and viral genealogies are possible and the scheme is merely representative.

Forterre looks likewise to a three component origin, but his emphasis is on the idea that viruses have contributed major components to the genome of all three groups, possibly providing each of three RNA-based cell lineages with independent transitions to DNA-based genomes by contributing DNA-polymerases, thus radically improving the stability and competitiveness of these cell lines who became the eventual survivors. In addition to the ribosomal proteins and rRNAs having distinct qualitative features in each domain, many DNA informational proteins exist in different non-homologous families (usually with several versions for one family). There are already six known non-homologous families of cellular DNA polymerases. In the case of DNA polymerases of the B family, there is one version in Bacteria (only found in some proteobacteria), one in Archaea, and several in Eukarya. The distribution of the different versions and families of cellular DNA informational proteins among domains is erratic most of the time and does not fit with any of the models proposed for the universal tree, suggesting abrupt insertion into the cellular genomes by viral transfer.
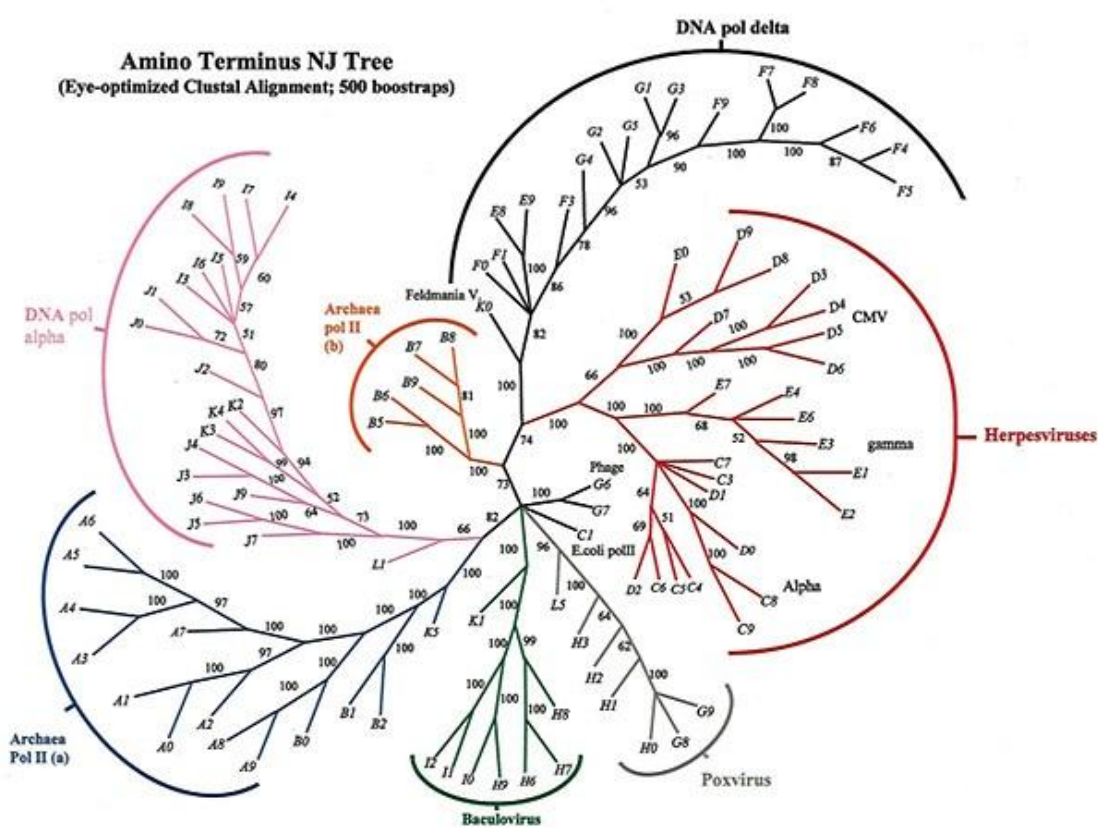
Fig 18: Evolutionary tree of DNA polymerase amino termini (Villareal and Defilippis).

Villareal and Defilippis have likewise investigated the idea that DNA viruses are the origin of DNA replication proteins, by investigating the amino terminus and constructing an evolutionary tree which shows DNA polymerases of DNA viruses, eukaryotes (α,δ), archaea, E coli and two phages rooted in a tree consistent with a viral origin.

This idea has a great deal of plausibility because viruses are now know to have a potentially primal origin, rather than being recent escapees from cellular genomes which have undergone reductive parasitic changes to their genome. Viruses clearly also have retained both RNA-RNA, DNA-DNA and retrotranscription DNA-RNA-DNA using both RNA and DNA stages in their capsid viral forms, so they retain all the transitional states between RNA and DNA-based replication.

Fig 19: Bacterial DNA polymerases also show viral members (<u>underlined</u>) close to the root of the tree (File et. al.).

Furthermore the retroviruses and related mobile genetic elements have a common ancient evolutionary origin, which is rel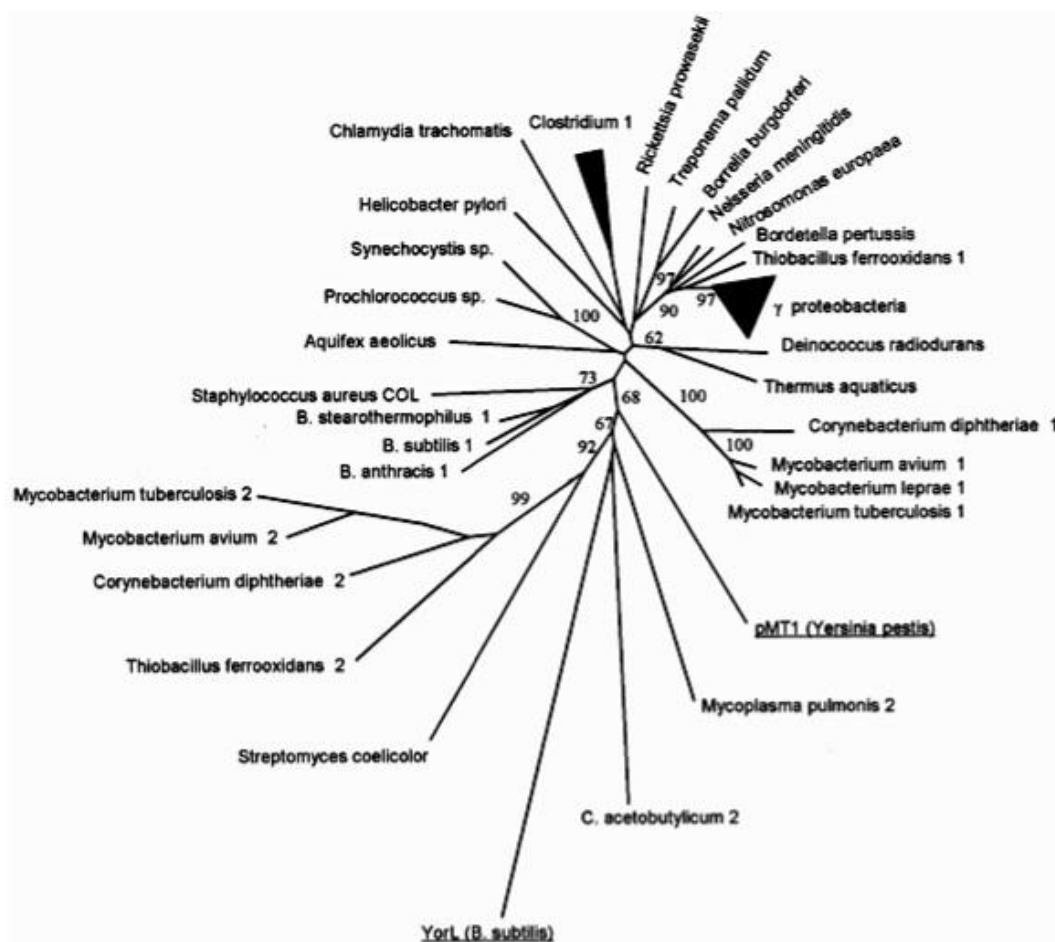ated to telomerase, which itself uses an RNA primer to initiate chromosome duplication. There is thus a plausible case that telomerase is in fact a biological fossil of a retroviral conversion of the founding Eukaryote cell line to a DNA genome.

The Symbiotic Face of Eukaryote Mobile Elements

In 1978, following the work of Darryl Reanny (1974-6), I proposed (1978, 1992) that viruses and transposable elements, far from just being selfish genes (Dawkins 1976), formed part of a dynamical system of genetic symbiosis between the hosts and the mobile genetic elements, because the mobile elements permitted forms of coordinated gene expression and the formation of new genes in a modular manner, which would otherwise be impossible, achieving in return perpetuation of their own genomes over evolutionary time scales. Most of the details of this proposal have proved to be realized.

By some reckonings, 40 to 50 per cent of the human genome consists of DNA imported horizontally by viruses, some of which has taken on vital biological functions. Coding sequences comprise less than 5% of the human genome, whereas repeat sequences account for at least 50% and probably

much more. Transposable LINE or long-intermediate repeat retroelements common to vertebrates, with a history running back to the Eukaryote origin are specifically activated in both sperms and eggs during meiosis (Branciforte and Martin, Tchénio et. al., Trelogan and Martin). These replicate from transcribed RNA copies of themselves thus using RNA to instruct DNA copies, indicating an origin in RNA-based life, as does the active RNA processing of our own Eukaryote cells. Their RNA-based reverse transcriptase shows homologies with the telomerase essential for maintaining immortality in our germ line, indicating a common and symbiotic origin. 100,000 partially defective LINEs and their 300,000 dependent smaller fellow traveler Alu SINEs make up a significant portion of the human and mammalian genomes, along with pseudogenes, apparently defective copies of existing genes translocated by elements such as LINEs.
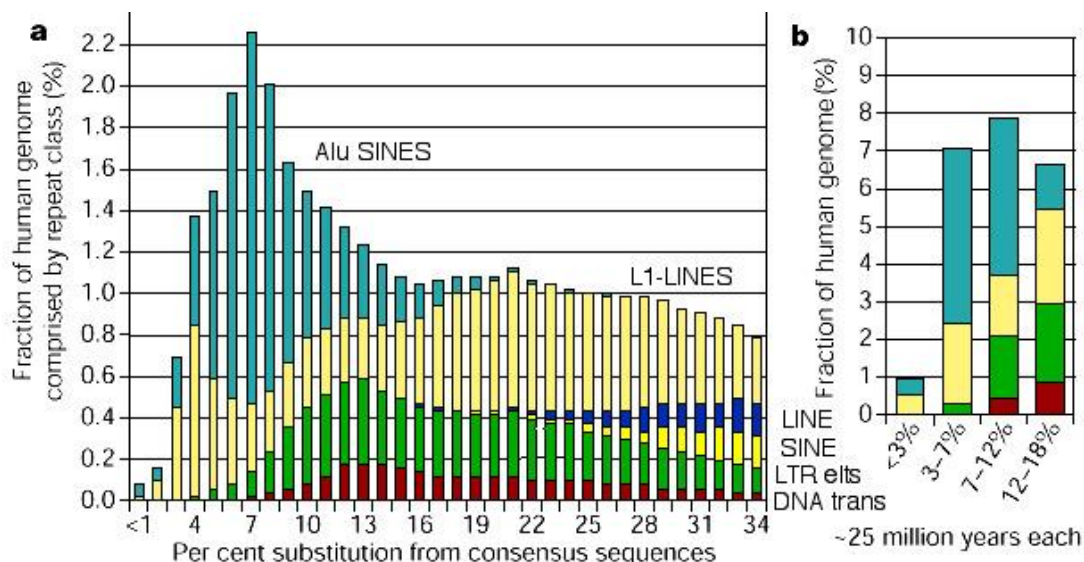

Fig 20: Human transposable element evolutionary history of L1-LINEs (cream), Alu elements (lt. blue), retrovirus-like LTR (long-terminal repeat) elements (green) and DNA transposons (dk. brown). Older LINEs and SINEs are in yellow and dk. blue. This history extends back over 200 million years indicating the very ancient basis of this potentially symbiotic relationship (Human Genome Consortium).

These elements travel passively down the germ line with chromosomal DNA, so their specific activation during meiosis suggests they may perform a role of coordinated regulatory mutation. This suggests that the type of symbiotic sexuality embraced by bacteria and plasmids also continues to function in higher organisms in a form of sexual symbiosis between our chromosomes and transposable genetic elements. This is consistent with the 1.4% point mutation divergence between humans and chimps, being overshadowed by an additional 3.9% divergence to 5.4% overall (Britten), when insertions and deletions are accounted.

SINEs, such as human Alu, a free-rider on the LINE reverse transcriptase derived from the small cellular RNA used to insert nascent proteins through the membrane, are in turn implicated in active functional genes (Reynolds, Schmid) particularly some involved in cellular stress reactions, again suggesting genetic symbiosis. Humans have about 13 times as many RNA edits as non-primate species, including inosine insertions associated with Alu elements, as well as intron deletions (Holmes) and newly inserted exons (Ast), which may differentiate humans from other apes through alternative splicing of genes expressed in the brain. RNA editing is abundant in brain tissue, where editing defects have been linked to depression, epilepsy and motor neuron disease. There is a new

Alu insert about every 100 births. As many as three quarters of all human genes are subject to alternative splice editing.
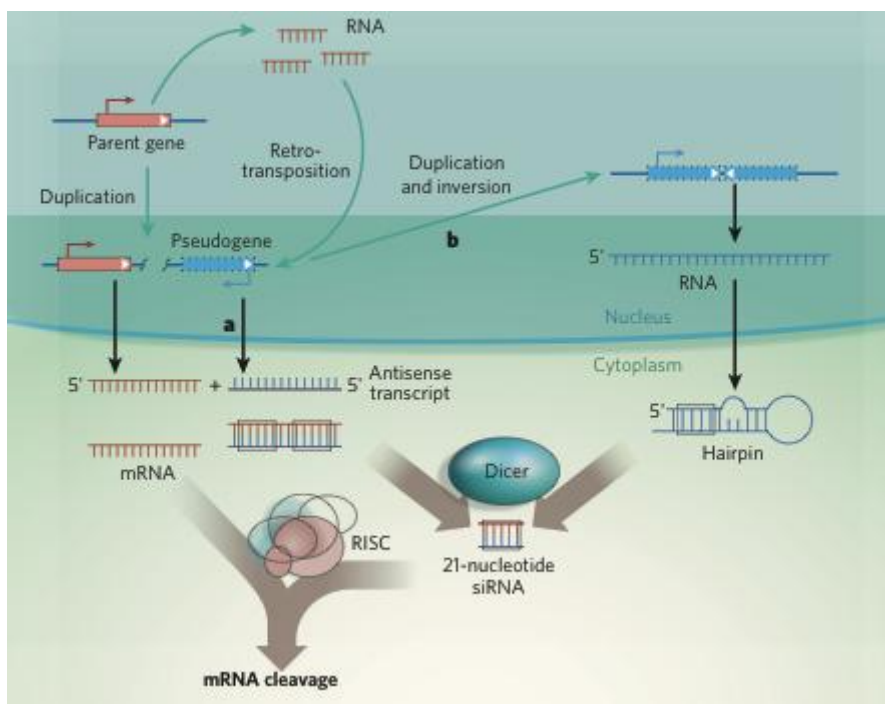

Fig 21: Pseudogene-mediated production of endogenous small interfering RNAs (endo-siRNAs). Pseudogenes can arise through the copying of a parent gene (by duplication or by retrotransposition). (a) An antisense transcript of the pseudogene and an mRNA transcript of its parent gene can then form a double-stranded RNA. (b) Pseudogenic endo-siRNAs can also arise through copying of the parent gene as in a and then nearby duplication and inversion of this copy. The subsequent transcription of both copies results in a long RNA, which folds into a hairpin, as one half of it is complementary to its other half. In both a and b, the double-stranded RNA is cut by Dicer into 21-nucleotide endo-siRNAs, which are guided by the RISC complex to interact with, and degrade, the parent gene's remaining mRNA transcripts. The mRNA from genes is in red and that from pseudogenes is in blue. Green arrows indicate DNA rearrangements (Sasidharan and Gerstein).

Recent explosion of the area of interfering miRNAs as regulatory elements in gametogenesis and development (Großhans) has provided an explanation of how pseudogenes, including those retrotransposed via LINE elements, can gain functional regulatory significance even though they do not produce translatable mRNAs.

Although the data from the human genome project indicated that human LINEs are becoming less active as a group by comparison with the corresponding elements in the more rapidly evolving mouse genome, there remain about 60 active human LINE elements which are known to be responsible for mutations in humans. More recent investigation (Boissinot et. al.) shows that the most recent families are highly active. Around four million years ago shortly after the chimp-human split, a new family Ta-L1 LINE-1 emerged and is still active, with about half the Ta insertions being polymorphic, varying across human populations. Moreover 90% of Ta-1d, the most recent subfamily are polymorphic, showing highly active lines remain present. LINEs are more heavily distributed on the sex chromosomes with X chromosomes containing 3 times as many full length potentially active elements and the Y chromosome 9 times as many! This is consistent with a continuing mutational

load on humans which is removed more slowly from the sex chromosomes by crossing over in proportion to the degree to which crossing over is inhibited in each (i.e. totally on the Y and largely in males in the X but not in females). Sexual recombination is a protection from mutational error in a process called Muller's ratchet.
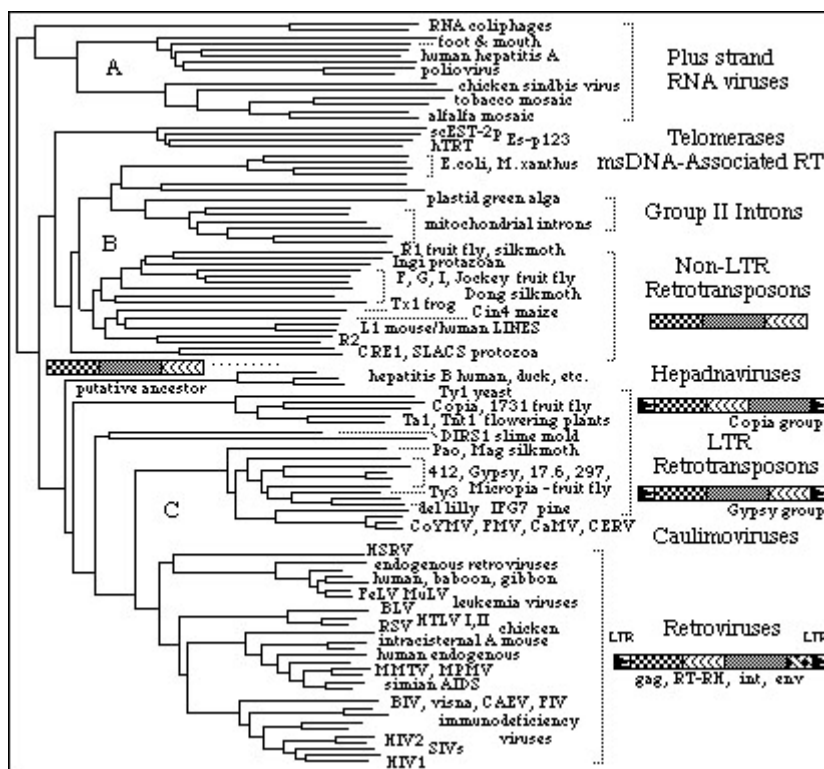

Fig 22: Evolution of reverse transcriptases from a common ancestor bearing a LINE archetype (Xiong and Eickbush, Nakamura et. al.). The root of their evolution goes back to the transfer from RNA to DNA at the beginning of life. They form a complementary evolutionary tree to that of cellular life as genetic symbionts of metazoa travelling down the germ line. Their group includes telomerases essential to the reproductive cycle.

LINEs are preferentially expressed in both steriodogenic and germ-line tissues in mice (Branciforte and Martin, Trelogan and Martin), suggesting stress could interact with meiosis. L1 expression occurs in embryogenesis, at several stages of spermatogenesis including leptotene, and in the primary oocytes of females poised at prophase 1. Conversely the SRY-group male determining gene SOX has been found to regulate LINE retrotransposition (Tchénio et. al.). Similarly LINE elements have been proposed to be 'boosters' in the inactivation of one X chromosome that happens in female embryogenesis (Lyon). This could enable somatic stress to have a potential effect on translocation in the germ-line which might enable form of genetic adaption in long-lived species such as humans.

Both L1 and Alu elements may be able to self-regulate rates of replication, through the existence of stealth drivers, viable elements which maintain a low transcription rate of active elements, with little genomic impact and hence little negative selection. These occasionally seed daughter master elements, which may replicate actively to form new families when conditions permit. This picture is consistent with long periods of quiescence, punctuated by bursts of 'saltatory' replication leading to large copy numbers (Han et. al.).

DNA Decipher Journal | January 2011 | Vol. 1 | Issue 1 | pp. 073-109
King, C., *The Tree of Life: Tangled Roots and Sexy Shoots*

94

Endogenous Retroviruses and the Placenta

Endogenous retroviruses, or ERVs, which also travel down the germ line as free-riders, although some may retain infectious capacity, may be essential for placental function, as every mammal tested has placental blooms of endogenous retroviruses which appear to both aid the formation of the syncytium, the super-cellular fused membrane that enables diffusion from the mother to the baby and the immunity suppression, which prevents rejection of the embryo, both characteristics of retroviruses such as HIV.

Mi and colleagues (2000) found a placental gene whose sequence was homologous to several retroviral envelope proteins. The sequence, now called syncytin, is identical to the envelope protein of the HERV-W retrovirus (Blond et. al.) which exists in around 40 apparently defective viral copies, including those in which the two syncytin viral *env* genes are fully functional (Mi et. al.). Syncytin is expressed at high levels in the syncytiotrophoblast (and at low levels in the testes) and nowhere else. Most of the other genes of the provirus have been mutated, suggesting that the envelope glycoprotein function was specifically selected. If cultured cells are made to express syncytin, they will fuse together, and this fusion can be blocked with antibodies against syncytin. HERV-W is only found in primates, but mice have similar retroviral blooms and ERV-related Syncytin genes have also been found in them (Dupressoir et. al.). The ability of mammals and thus ourselves to form a viable placenta and give birth to live young may thus depend on the mammals having harnessed a viral gene somewhere in our evolutionary lineage.

The defective copies of endogenous retroviruses may also serve to protect the host against further infection by becoming transcribed and causing incorporation of defective elements into the replicating virus (Best et. al.).
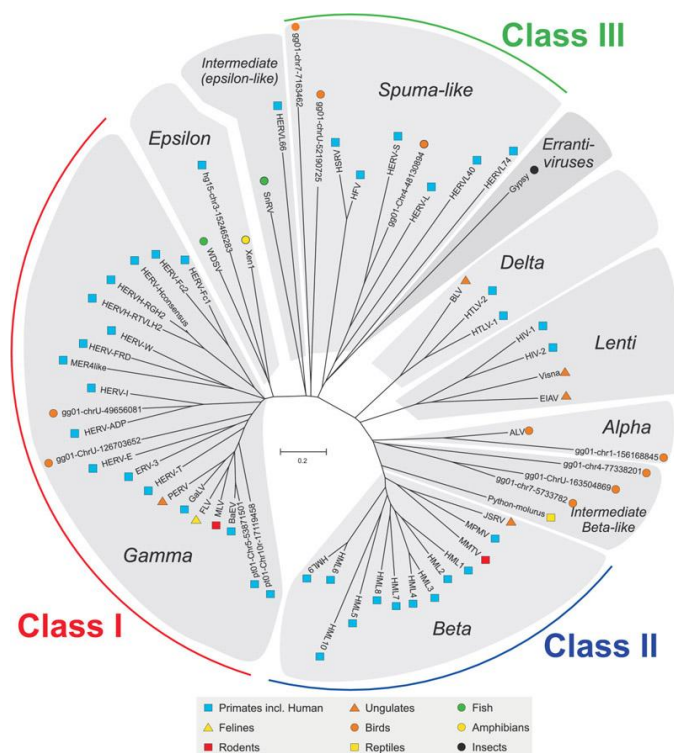

Fig 23: The evolutionary tree of endogenous retrovirus classes along with other infectious retroviruses such as HIV span virtually all metazoa. When retrotranspons are included (fig 22) they extend to all Eukaryote realms.

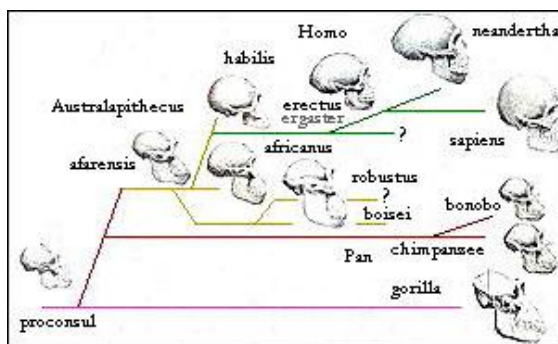## Emergence and Diversification of Modern Humans

Fig 24: One hypothetical evolutionary tree for humans and related apes. There is much debate about the actual form of such a tree.

Homo sapiens appears to have evolved into a single dominant species on the planet, after preceding period in which fossil evidence suggests there were several different anthropoid species coexistent.

The final stage of this process was the disappearance of Homo erectus and Neanderthal, the latter after a well-defined period of coexistence in Europe at the end of the last ice age.

Our own evolutionary and cultural roots appear to lie in Africa, with evidence of culture and cosmetics running back over 100,000 years, in addition to evidence for tools and weapons. An alternative regional development theory has proposed that humans evolved through a considerable amount of interbreeding over the whole African and Asian continental region, however genetic evidence is coming to point towards an African origin with only at most very occasional cross fertilization with related species.

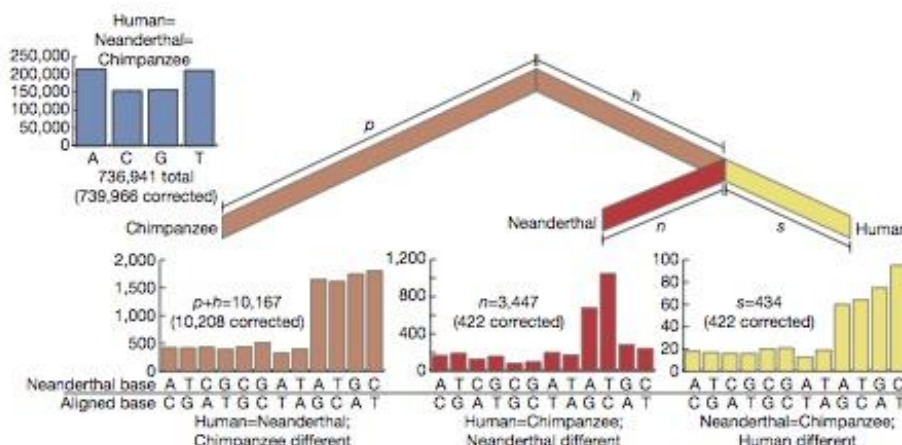Fig 25: Human-Neanderthal-Chimp divergences (Green et. al.)

According to genetic analysis, Neanderthals diverged from homo sapiens ~500,000 years ago. There has been no major interbreeding, but possibly some transfer of genes e.g. from human males to Neanaderthal females, although candidate human genes conferring natural advantage do have a profile consistent with transfer from Neanderthals (Green et. al.).
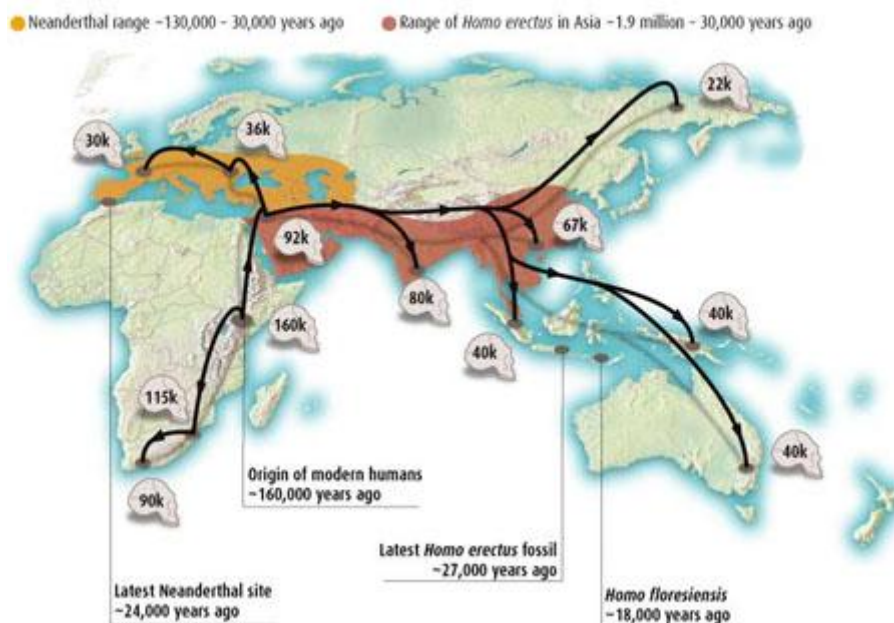
Fig 26: The "Out of Africa" hypothesis may be consistent with a degree of regional development involving some sexual interbreeding with Neanderthals and Homo erectus (New Scientist).

Specific genes such a s PDHA1 consist of two families with the last common ancestor 1.8 million years ago, and microcephalin variants appearing 40,000 years ago also have differences suggesting an original divergence 1 million years ago suggesting 'introgression' from Neanderthals (Jones). An even more ancient divergence in the pseudogene RRM2P4 in East Asian people suggests interbreeding with Homo Erectus. Some evidence from skeletons is also consistent with this picture. However more recent sequencing of the Neanderthal nuclear genome (Callaway) suggests little or no interbreeding with *Homo sapiens* and has cast doubt on the existence of the microcephalin variant in Neanderthals, as well as a gene associated with increased fertility in Icelanders also attributed to transfer from Neanderthals.
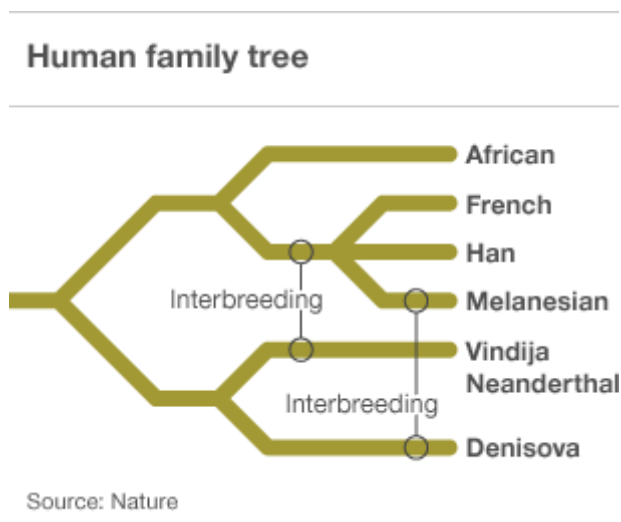
Comprehensive investigation of the Neanderthal genome (Green, R. E. et al. (2010) Science 328, 710-722., Nature | doi:10.1038/news.2010.225) suggests that there was a period of interbreeding between Neanderthals and humans in the Near East around the time of the first migration out of

Africa, rather than more recently in Europe, as the putative sequences are shared by non-African French, Han and Papuan, but not by the African Yoruba or San. It is estimated that among the former, 1-4% of the genome derives from Neanderthal sequences, although there is little evidence for these corresponding to the specific genes suggested by Lahn's team. Other transfers could have occurred but are not apparent in the research.

The situation has been complicated by two finds. Firstly we have the 'hobbit' human remains found on Flores, named Homo floresiensis. These are variously claimed to be a Separate human species possibly related to Homo erectus, or disclaimed as microcephalic human pygmy peoples. More recently we have the discovery of remains of Denisovians, a further species branching off from the Neanderthals. Genetic analysis of the remains indicates a significant interbreeding specifically with Melanesian people of some 6% (Reich D, et. al. 2010 Genetic history of an archaic hominin group from Denisova Cave in Siberia Nature doi:10.1038/nature09710).

Chromosomes contain a variety of markers that can be used to compare diverse populations and infer an evolutionary relationship between them. These include the slowly varying protein polymorphisms of coding regions which are useful for long-term trends, single nucleotide polymorphisms, and non-coding region changes (mutation rates about 2.5×10^{-8} per base pair per generation and useful for reconstructing evolutionary history only over millions of years) insertion and deletion events (about 8% of polymorphisms, extending from one to millions of nucleotides), particularly those driven by transposable elements such as the LINEs and even more frequent SINEs, non-coding micro-satellites (mutation rate 10^{-5} - 10^{-2} due to repeat slippage) and mini-satellite regions of repeating DNA (mutation rates as high as 2×10^{-1} due to meiotic recombination in sperm) that both evolve rapidly and are not subject to the strong selection of coding regions which can differentiate changes over the much shorter time scales of modern human migration.

The insertions and deletions of the million or so Alu elements in the human genome are particularly useful, as the most active sub-population of about 1000 Alu is actively transcribing and undergoing rapid change. A subpopulation of Alu are capable of generating new coding regions (exons), when inserted into non-coding introns between spliced sections of a translated mRNA, because one base-pair change within Alu leads to formation of a new exon reading into the surrounding DNA. This is not necessarily deleterious because alternative splicing still allows the original protein to be made as well. We have the highest number of introns per gene of any organism, and thus have to have gained an advantage from this costly error-prone process. Alus may have given rise, through alternative splicing, to new proteins that drove primates' divergence from other mammals. Recent studies have shown that the nearly identical genes of humans and chimps produce essentially the same proteins in most tissues, except in parts of the brain, where certain human genes are more active and others generate significantly different proteins through alternative splicing of gene transcripts. Our divergence from other primates may thus be due in part to alternative splicing.

If we consider the likely effects of the out of Africa hypothesis, we would expect that founding African populations not subject to active expansion and migration would have greater genetic diversity and that the genetic makeup of other world populations would come from a subset of the African diversity, consisting of those subgroups who migrated.

In the case of mitochondrial mtDNA (mutation rate about 2.5×10^{-7}) and its hyper-variable D-loop (mutations rates as high as 4×10^{-3}), which is transmitted only down the maternal line (see Tishkoff and Verrelli for caveat) and the non-recombining majority of the Y-chromosome which is transmitted only down the paternal line, each with no recombination, we would expect greater diversity going deeper into the historical tree of divergence, with certain existing groups who have

retained the founding patterns of survival and have not undergone rapid population expansions to retain an increasingly diverse source variation. All these features are broadly observed in the genetic data to date.
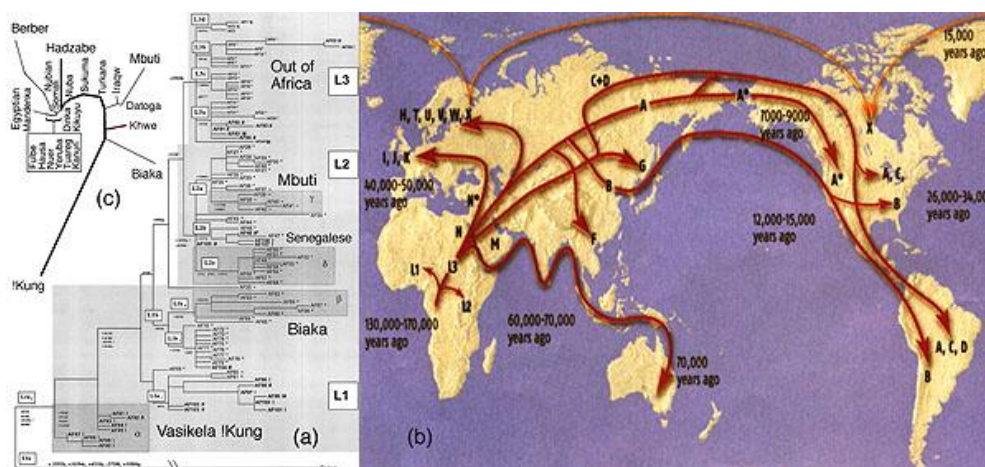

Fig 27: (a) MtDNA tree for African groups showing haplotypes of !Kung, Mbuti and Biaka as well as the line coming out of Africa (Chen et. al.). (b) Diagram of world migration and regional differentiation of successive mtDNA haplotypes (Gilbert). (c) mtDNA distances between founding African groups including Hadza (clicks) Khwe is from (Knight et. al.). Recent mtDNA evidence suggests a first wave of migration down the coast of Asia all the way to Australia (Forster et. al.).

Most studies of non-coding regions of autosomal, X-chromosome, and mitochondrial mtDNA genetic variation (which are desirable markers because they are not so subject to selection and thus have relatively neutral drift) show higher levels of genetic variation in African populations compared to non-African populations, using many types of markers. Although some studies of Y-chromosome variation have observed higher heterozygosity levels in non-African populations, the African populations have higher levels of pairwise sequence differences, consistent with these populations being ancestral. High levels of diversity in African populations alone do not prove that African populations are ancestral. A recent bottleneck event and/or colonization and extinction events among non-African populations, or a more recent onset of population growth in non-Africans, could also cause a decrease in genetic diversity (Tishkoff and Verrelli). In fact the complete inter-fertility of all human populations and the relative lack of genetic divergence by comparison with the few remaining chimp colonies in the wild (Hrdy 183) does indicate a significant bottleneck. The genetic data is consistent with a human emergence from a population of only 10,000 around 100,000 years ago. This is also consistent with the delayed maturation, long birth spacings as a result of prolonged lactation and high infant mortality seen in gather-hunter populations such as the !Kung. At such low growth rates a population of 100 would take 50,000 years to reach 10,000 (Hrdy 183).

Fig 28: Patterns of male migration. The Genographic Project - a partnership between National Geographic and IBM - will collect DNA samples from over 100,000 people worldwide to provide a high-resolution genetic map of human migration.

However studies of protein polymorphisms as well as mtDNA haplotypes, X-chromosome and Y-chromosome haplotypes, autosomal microsatellites and minisatellites, Alu elements, and autosomal haplotypes indicate that the roots of the population trees constructed from these data are composed of African populations and/or that Africans have the most divergent lineages, as expected under a recent African origin rather than a multi-regional emergence model. Additionally, studies of autosomal, X-chromosomal haplotype and mtDNA variation indicate that Africans have the largest number of population-specific alleles and that non-African populations harbor a subset of the genetic diversity that is present in Africa, as expected if there was a genetic bottleneck when modern humans migrated out of Africa. Analysis of genetic variation among ethnically diverse human populations indicates that populations cluster by geographic region (i.e., Africa, Europe/Middle East, Asia, Oceania, New World) and that African populations are highly divergent. The mtDNA studies hypothesize a primal female ancestor - the African Eve - around 150,000 years ago (Chen et. al.) while the Y-chromosome Adam is more recent, at around 90,000 years ago (Underhill et. al.) consistent with the greater reproductive variance of males than females. Differences between the Y- and mtDNA distributions indicate how migration, intermarriage and female exogamy have affected the gene pool.
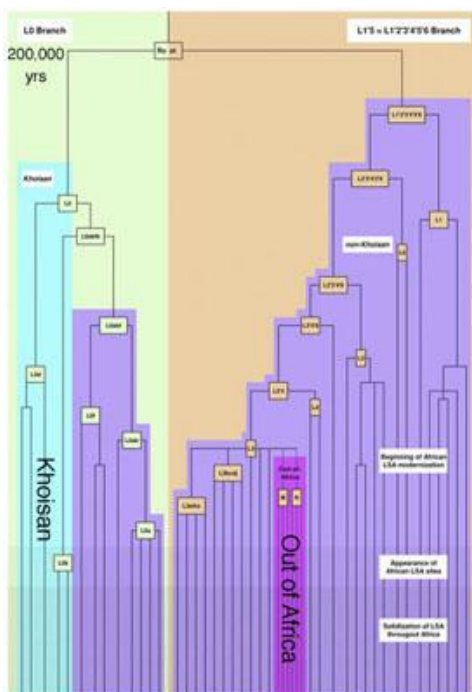
Fig 29: (Right) Genographic project study of mitochondrial origins shows a deep split separating Khoisan mitochondrial inheritance from other groups, including those migrating out of Africa, suggesting a separation of some 100,000 years possibly caused by long term drought in Africa (Behar et al.)

The genetic patterns of both these and autosomal microsatellites (Zhivotovsky et. al.) are consistent with founding African diversity with migratory radiations to form other world populations, with deep founding radiations to the forest people such as the Biaka and Mbuti, Khoisan click-language speaking !Kung-san bushmen of Botswana and the Sandawe of Tanzania, and possibly the Hadzabe,

as well as the forest people such as the Mbuti and Biaka 'pygmies' who have adopted the Bantu languages of the farming neighbours with which they now share semi-symbiotic relationships. Along with some Ethiopian and Sudanese sub-populations, these groups may represent some of the oldest and deeply diversified branches of modern humans. Such recent genetic evidence has laid bare the relationships between some of the founding human groups spread across Africa from the 'Cushite' horn of Ethiopia to the southern Kalahari. Mitochondrial DNA studies have highlighted the ancient origin of the !Kung San and of pygmy peoples of the Congo Basin such as the Mbuti and the Biaka.

Y-chromosome studies have shown the !Kung share a most ancient haplotype with sub-populations from Ethiopia and the Sudan. According to an overall survey of genetic research by Sarah Tishkoff of the University of Maryland, the most deeply ancestral known human DNA lineages may be those of East Africans, such as the Sandawe, who share many phenotypic features and a click language with the !Kung. This suggests southern Khoisan-speaking peoples originated in East Africa. The most ancient populations are now believed to also include the Sandawe, Burunge, Gorowaa and Datog people of Tanzania. The Burunge and Gorowaa migrated to Tanzania from Ethiopia within the last 5,000 years consistent with an ancient founding population in this area. Echoes of the earliest language spoken by ancient humans tens of thousands of years ago may have been preserved in the distinctive clicking sounds still spoken by some existing African tribes.
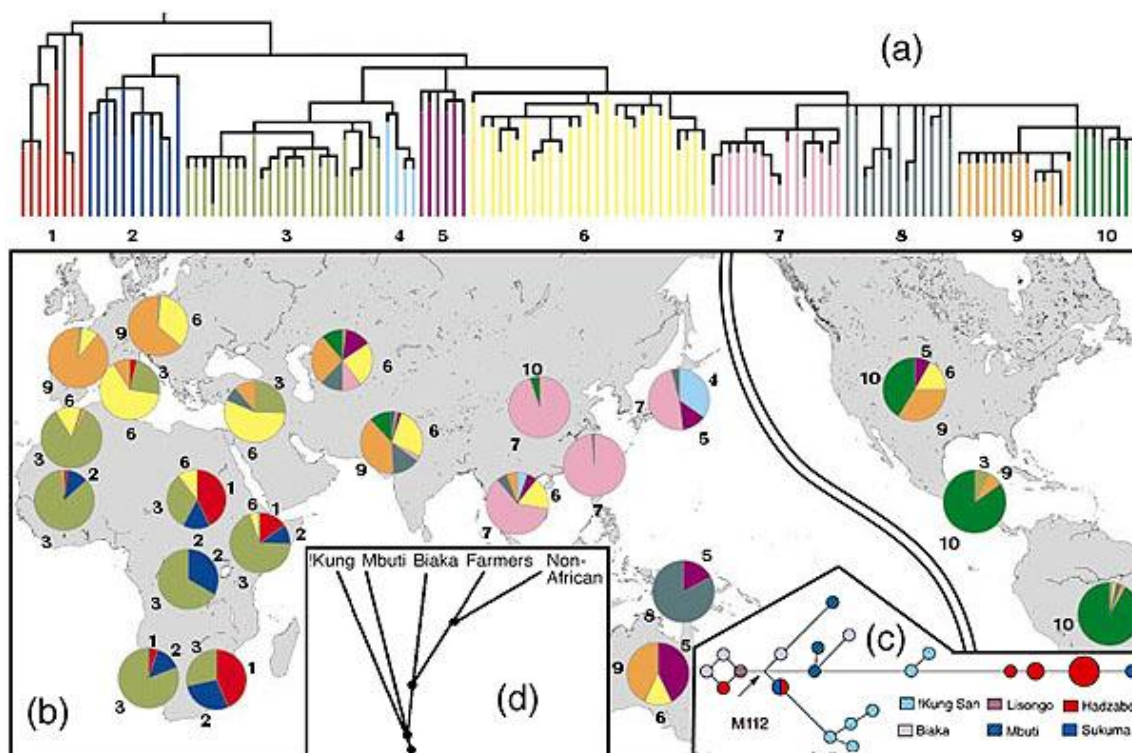
Fig 30: (a) Non-recombining Y-chromosome evolutionary tree (Underhill et. al.) (b) Geographical distribution showing the ancient haplotype shared by the San and Ethiopian and Sudanese sub-populations. (c) Genetic distances between Khoisan and forest peoples sharing M112 a Y-chromosome allele common only in these groups showing great genetic distance between Hadzabe and San peoples (Knight et. al.) . (d) Autosome satellite analysis confirming ancient divergence of San and forest peoples leading to migration from Africa (Zhivotovsky et. al.).

Highlighting unique features of human genetic evolution, are two key genes whose mutations cause microcephaly, consistent with increased brain size, whose rapid spread through the human population may coincide with spurts in human culture. Microcephalin (Evans et. al.) appeared ~37,000 years ago coinciding with the birth of culture and ASPM spread from the Near East around 5000 years ago (Mekel-Bobrov et. al.). However studies linking these variants have failed to find differences in intelligence and results remain highly controversial (DOI:10.1126/science.314.5807.1872). Nevertheless, these results are consistent with an overall examination of linkage disequilibrium in single nucleotide polymorphisms (Moyzis et. al.) which indicate that about 7% of our genes have been subject to selection in the last 50,000 years, a figure similar to domestication of maize, including genes for protein metabolism, disease resistance and brain function.

In a counterpoint to these studies, (Hein, Rohde et. al.) estimate that the repeated spreading of family trees by sexually recombining mobile populations and differences in reproductive rates leads to an estimate of the most recent common ancestor of our global populations existing just 3,500 years ago, excepting these most isolated groups.
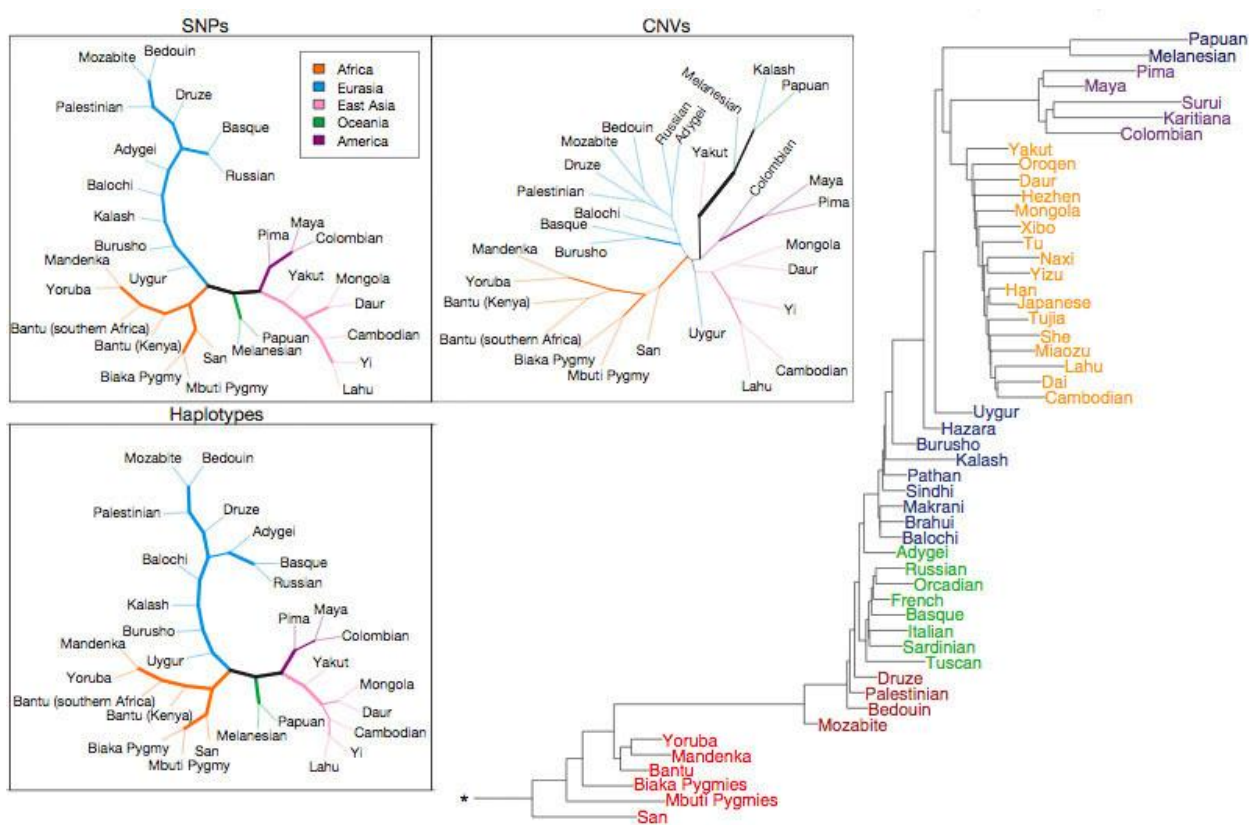
Fig 31: Human divergence trees calculated by single nucleotide polymorphisms (SNPs) top left (Li et. al.) lower right (Jakobsson et. al.) and for haplotypes and copy number variation between populations (Li et. al.).

Further studies of the nuclear genome, using SNPs (single nucleotide polymorphisms), CNVs (copy number variation) and haplotype have thrown up reasonably consistent maps of regional divergence of principal human groups, demonstrating correspondence to the "Out of Africa" hypothesis and consistent with major patterns of migration.

The evolutionary tree of human ethnic and migratory peoples bears an interesting relationship with the corresponding tree of languages, which has been resolved at least as far as the founding Indo-European languages (Gray and Atkinson).
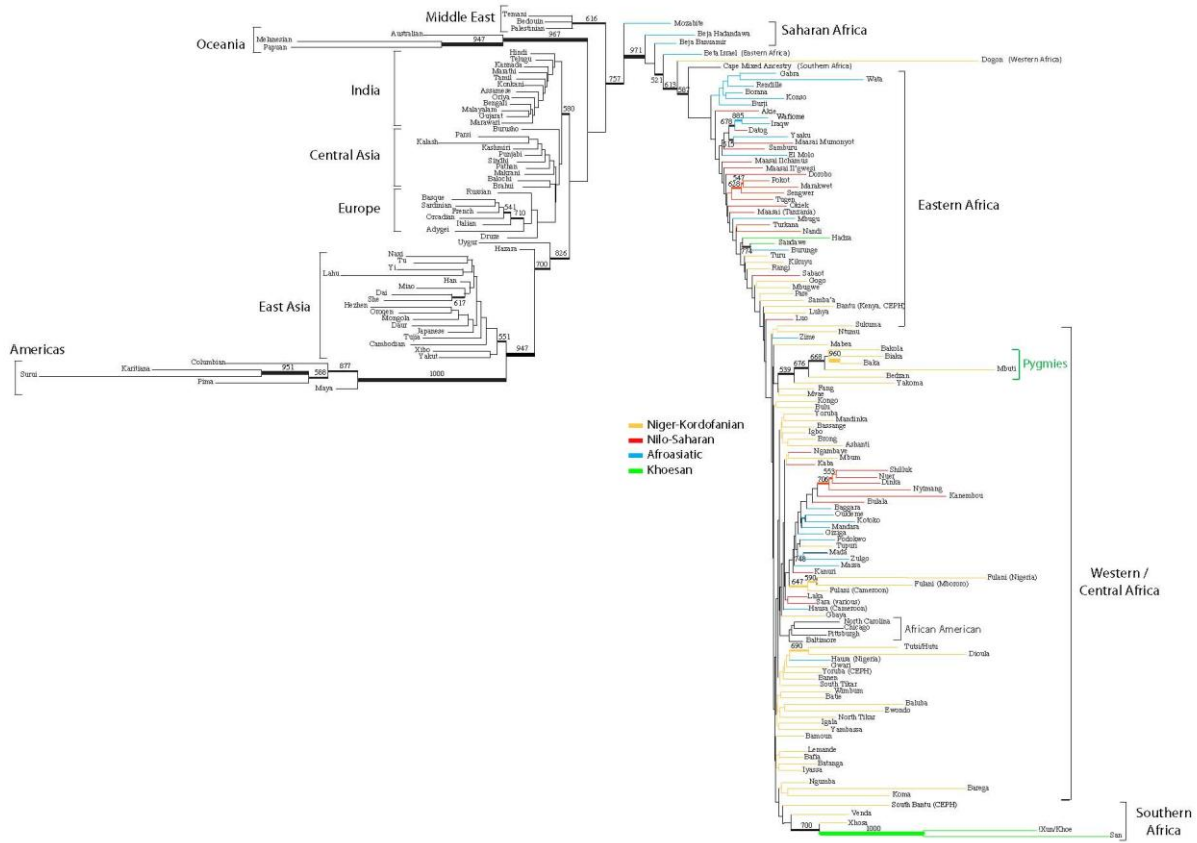
Fig 31a: In 2009, Tishkoff et. al. reported on a major study of African and African American evolution containing the most detailed information on African diversity to date.
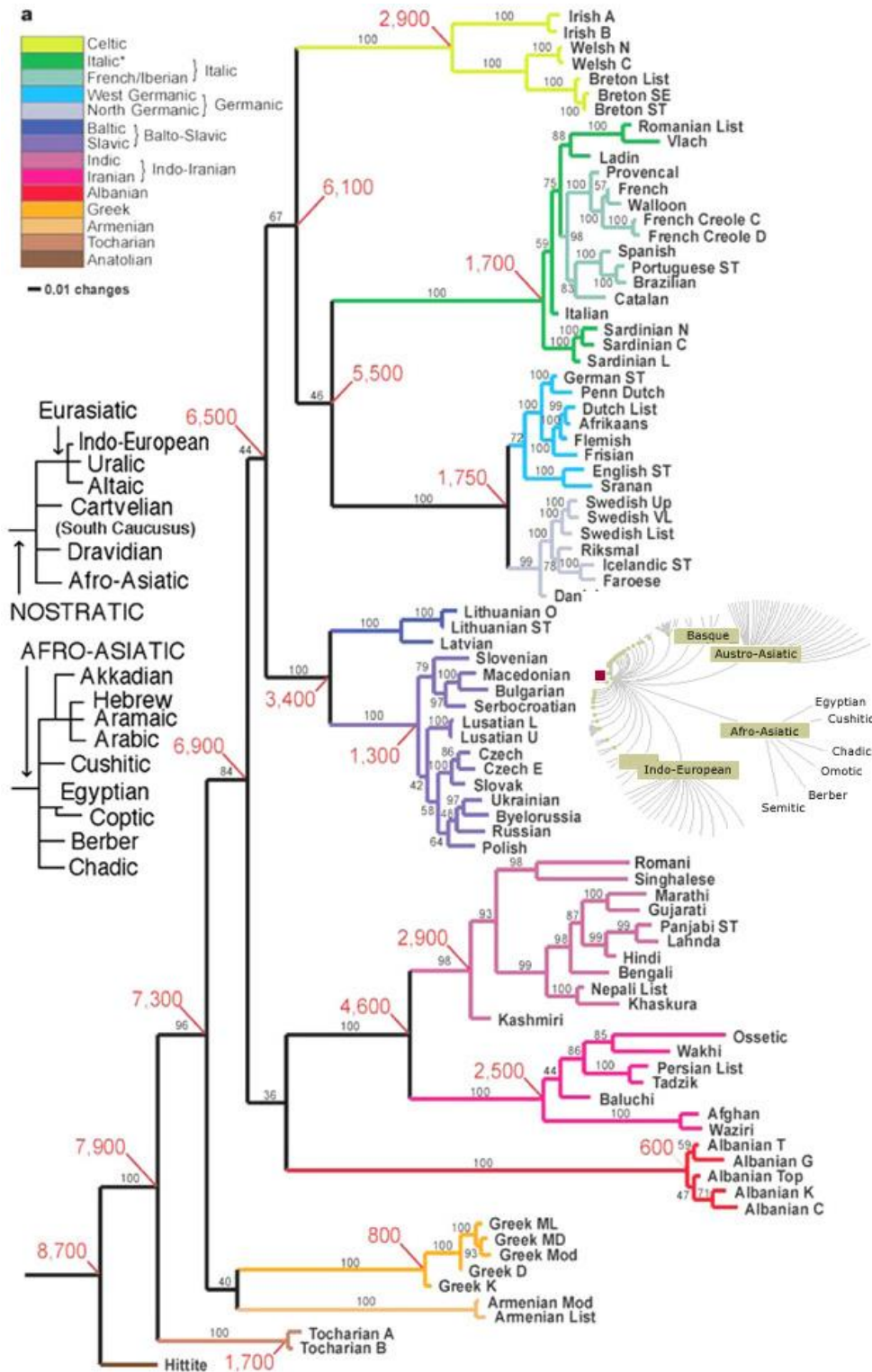
Fig 32: Evolutionary tree of Indo-European languages suggests a possible radiation corresponding to the Kurgans occurred around 4,900 BC (6,900 BP) and that they were preceded by Hittite migrations into Anatolia. Time scales in red are BP (Gray and Atkinson). Significantly Tocharian appears in Buddhist writings from China's Xinjiang province, indicating early far-eastern spread. Inset: hypothetical relationship between Indo-European and wider language groups such as Afro-Asiatic.
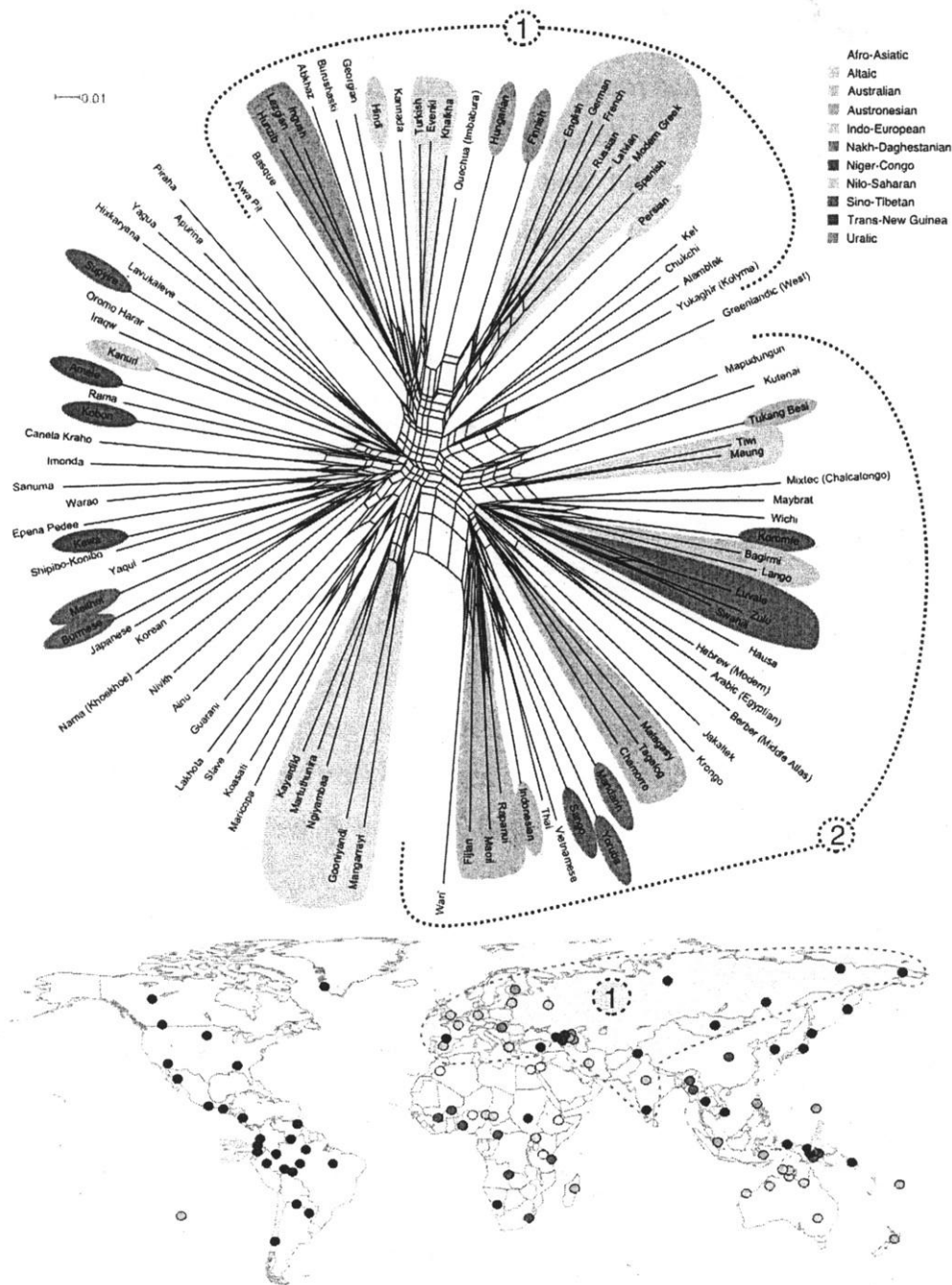
Figure 1: NeighborNet for the 99 most well-attested languages in the WALS database. This network is based on 138 typological characters and shows the signals grouping languages. The length of the branches on the network are proportional to amount of divergence between languages, and the box-like structures reflect conflicting signal. Family groups are color-coded, and potential language areas are marked with dashed lines and numbered as described in the text. The dashed area and arrows on the map show the extent of the third Eurasian cluster.

The shape and fabric of language evolution

Fig 33: Hypothetical core of all human languages from "The Shape and Fabric of Language Evolution" further work associated with Gray and Atkinson's research.

Conclusion: The Tree of Life, the Selfish Gene, and Climax Genetic Diversity

The picture conveyed by the significance of endosymbiosis, genome fusion and horizontal transfer as key evolutionary processes complementing the vertical transmission of the tree of life, makes clear that evolution is not just a matter of competitive survival of the fittest gene, individual, or species, but of dynamic survival of genes in a surviving ecosystem. Although Dawkins' (1978) notion of the "selfish gene" was pivotal in drawing attention to the fact that it was the survival of genes and not organisms, or even species, that was the key evolutionary process, attributing the human sentiment of selfishness to a gene is somewhat of a self-serving advertising distraction on the part of the author, which diminishes the subtlety and complexity of the sometimes apparently paradoxical ways genes actually interact to bring about beneficial outcomes in the evolutionary dynamics of the ecosystem.

The Mandala of Evolution (Dion Wright) (detail).

Although the idea of selection of genes has been pivotal in defining the need to consider evolutionarily stable strategies under genetic variation in ways which have been subsequently confirmed time and time again in situations such as the sexual genetics of social insects such as bees and ants, social selection is by no means ineffectual, or much of sociobiology, including the biological basis of morality as an extension of reciprocal altruism, would cease to exist.

Moreover, from what we have seen, particularly about horizontal gene transfer, and the capacity of mobile elements to induce modulated changes in nuclear genomes, it is not the 'selfishness' of a genetic element alone that results in survival of both a gene and its hosts, but dynamic feedbacks,, and relationships which ultimately contribute to a massive sharing of information in the manner of parallel genetic algorithms fundamental to the replicative genetic process, which enable global forms of genetic and genome optimization central to the overall viability of life as complex systems.

Just as a predator, such as a lion survives, not because it is a selfish beast thinking only of eating the next gazelle, but because the predator, although it is surviving by killing individual antelopes, is maintaining a degree of stability in population dynamics, without which, the herbivores might multiply causing a massive famine, leading to cycles of boom and bust and the potential extinction or attrition of antelopes, lions and the grasslands.

Likewise, although we may think of individual genes, transposable elements, or viruses as 'selfish' for reproducing sufficiently to ensure their own survival, and sometimes behaving as noxious parasites, the overall effects of this process, in evolution can be to enrich the genetic potential of many unrelated organisms along the way, changing forever the face of the ecosystems in which they exist, enabling organisms of far greater complexity to evolve and to survive in the closing circle of the biosphere.

References

1. Adam R 2000 *The Giardia lamblia Genome* Int. J. Parasitology 30 474-84.
2. Anathaswamy Anil 2003 *The thermal history of life is revealed* New Scientist 31 May 20.
3. Arnold M, Sapir Y, Martin N 2008 *Genetic exchange and the origin of adaptations: prokaryotes to primates* Phil. Trans. R. Soc. B 363 2813–2820.
4. Ast G 2005 *The alternative genome* Scientific American 292/4.
5. Baaske P, Weinert F, Duhr S, Lemke K, Russell M, Braun D (2007) *Extreme accumulation of nucleotides in simulated hydrothermal pore systems* PNAS 104/22 9346-9351.
6. Behar et al. 2008 *The Dawn of Human Matrilineal Diversity, The American Journal of Human Genetics* 2008 doi:10.1016/j.ajhg.2008.04.002.
7. Best S, Le Tissier P, Stoye J 1997 *Endogenous retroviruses and the evolution of resistance to retroviral infection* Trends in Microbiology 5/8 313-8.
8. Blond J, Beseme F, Duret L, Bouton O, Bedin F , Perron H, Mandrand B, Mallet F 1999 *Molecular Characterization and Placental Expression of HERV-W, a New Human Endogenous Retrovirus Family* J. Virol. 73/2 1175-1185.
9. Boissinot, Stephanie, Entezam, A, Furano, A. 2001 *Selection Against Deleterious LINE-1-Containing Loci in the Human Lineage* Mol. Biol. Evol. 18(6) 926–935.
10. Boussau B, Blanquart S, Necsulea A, Lartillot N, Gouy M 2008 *Parallel adaptations to high temperatures in the Archaean eon* Nature 456 942-6.
11. Branciforte D., Martin S. 1994 *Developmental and Cell-type specificity of LINE-1 Expression in Mouse Testis: Implications for Transposition* Molecular and Cellular Biology 14/4 2584-92.
12. Britten Roy J 2002 *Divergence between samples of chimpanzee and human DNA sequences is 5% counting indels*. Proc. Nat. Acad. Sci. 99 13633-5.
13. Brochier C, Philippe H 2002 *A non-hyperthermophilic ancestor for Bacteria* Nature 417 244.
14. Budin I, Bruckner R, Szostak J (2009) *Formation of Protocell-like Vesicles in a Thermal Diffusion Column* JACS 131 9628-9629.
15. Callaway, E 2008 *Neanderthal genome already giving up its secrets* New Scientist 10 Dec.
16. Chen Y, Olckers A, Schurr T, Kogelnik A, Huoponen ,Wallace D 2000 *mtDNA Variation in the South African Kung and Khwe—and Their Genetic Relationships to Other African Populations* Am. J. Hum. Genet. 66:1362–1383.
17. Ciccarelli F, Doerks T, von Mering C, Creevey C, Snel B, Bork, Peer 2006 *Toward Automatic Reconstruction of a Highly Resolved Tree of Life* Science 311 1283-7.
18. Dagan T, Martin W 2006 *The tree of one percent* Genome Biology 7 118.
19. Dagan T, Artzy-Randrup Y, Martin W 2006 *Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution* PNAS 105/29 10039 –10044.
20. David L, Alm E 2010 *Rapid evolutionary innovation during an Archaean genetic expansion* Nature doi:10.1038/nature09649
21. Dawkins, Richard 1976. The Selfish Gene. New York City: Oxford University Press. ISBN 0-19-286092-5
22. Doolittle W 1998 *You Are What You Eat: A Gene Transfer Rachet Could Account for Bacterial Genes in Eukaryotic Nuclear Genomes* Trends in Genetics 14/8 307–311.
23. Doolittle W 1999 *Phylogenetic Classification and the Universal Tree* Science 284 2124–8.
24. Doolittle W 2000 *Uprooting the Tree of Life* Sci. Am. Feb 90-5.
25. Dupressoir A, Marceau G, Vernochet C, Be´nit L, Kanellopoulos C, Sapin V, Heidmann T 2005 *Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae* PNAS 102/3 725-30.

26. Emelyanov V 2003 *Mitochondrial connection to the origin of the eukaryotic cell* Eur. J. Biochem. 270, 1599–1618.

27. Evans P, Gilbert S, Mekel-Bobrov N, Vallender, E, Anderson J, Vaez-Azizi L, Tishkoff S, Hudson R, Lahn B 2007 *Microcephalin, a Gene Regulating Brain Size, Continues to Evolve Adaptively in Humans* Science 309 1717-20.

28. File J, Forterre P, Sen-Lin T, Laurent J 2002 *Evolution of DNA Polymerase Families: Evidences for Multiple Gene Exchange Between Cellular and Viral Proteins* J Mol Evol 54 763–773.

29. Fisher, Allen, Wilson and Suttle 2010 *Giant virus with a remarkable complement of genes infects marine zooplankton* PNAS doi: 10.1073/pnas.1007615107

30. Forster P, Matsumura S 2005 *Did early humans go north or south?* Science 308 965-6.

31. Forterre P 2006 *The origin of viruses and their possible roles in major evolutionary transitions* Virus Research 117 5-16.

32. Forterre P 2006b *Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: A hypothesis for the origin of cellular domain* PNAS 103/10 3669 –3674.

33. Gilbert Tom 2003 *Death and Destruction* New Scientist 31 May 32.

34. Gray Russell, Atkinson Quentin 2003 *Language-tree divergence times support the Anatolian theory of Indo-European origin* Nature 426 435 - 439

35. Green R, Krause J, Ptak S, Briggs A, Ronan M, Simons J, Du L, Egholm M, Rothberg J, Paunovic M, Paabo S 2006 *Analysis of one million base pairs of Neanderthal DNA* Nature 444 330-336. doi:10.1038/nature05336

36. Green et. al., Science 314, 1068-71

37. Großhans H, Filipowicz W 2008 *The expanding world of small RNAs* Nature 451 414-6, doi:10.1038/nature06863, 06642, 06908, 07015

38. Han K, Xing J, Wang, Hedges D, Garber R, Cordaux R, Batzer M 2005 *Under the genomic radar: The Stealth model of Alu amplification* Genome Research 15 655-64.

39. Hartman H, Fedorov A 2002 *The origin of the eukaryotic cell: A genomic investigation* PNAS 99/3 1420-1425.

40. Hein J 2004 *Pedegrees for all humanity* Nature 431, 518-9.

41. Holmes Bob 2005 *Does RNA editing make us brainy?* New Scientist 29 Jan 13.

42. Horiike T, Hamada K, Kanaya S, Shinozawa T 2001 *Origin of eukaryotic cell nuclei by symbiosis of Archaea in Bacteria is revealed by homology-hit analysis* Nature Cell Biology 3 210-4.

43. Hrdy, Sarah Blaffer 1999 Mother Nature : A History of Mothers, Infants, and Natural Selection Pantheon New York.

44. International Human Genome Sequencing Consortium 2001 *Initial sequencing and analysis of the human genome* Nature 409 15.

45. Jakobsson, Mattias et. al. 2008 *Genotype, haplotype and copy-number variation in worldwide human populations* Nature 451 998 doi:10.1038/nature06742

46. Jones, Dan 2007 *The Neanderthal Within* New Scientist 3 Mar.

47. King C. C. 1978 *Unified Field Theories and the Origin of Life* Report Series 134 Department of Mathematics, University of Auckland.

48. King C. C. 1992 *Modular Transposition and the Structure of Eucaryote Regulatory Evolution* Genetica 86 127-142.

49. King C. C. 2004 *Cosmic Symmetry-breaking, Bifurcation, Fractality and Biogenesis* Neuroquantology 3 149-185 http://www.dhushara.com/pdf/bifurc.pdf (See: http://www.dhushara.com/book/bchtm/biocos.htm)

50. King C. C. 2010 *Holy War against Science: Natural Evolution versus Intelligent Design* "SGJ", ISSN: 2153-831X http://scigod.com/index.php/sgj/article/view/75/84

51. King C. C. 2011 *The Tree of Life: Tangled Roots and Sexy Shoots: Tracing the genetic pathway from the first Eukaryotes to Homo sapiens* http://www.dhushara.com/book/unraveltree/unravel.htm

52. Knight A, Underhill P, Mortensen H, Zhivotovsky L, Lin A, Henn B, Louis D, Ruhlen M, Mountain J 2003 *African Y Chromosome and mtDNA Divergence Provides Insight into the History of Click Languages* Current Biology, 13, 464–473.

53. Lane C, Archibald J 2008 *The eukaryotic tree of life: endosymbiosis takes its TOL* Trends in Ecology and Evolution 23/5 268-274.

54. Lane Nick 2009 *Was our oldest ancestor a proton-powered rock?* New Scientist 19 Oct.

55. Lane N, Martin W 2010 *The energetics of genome complexity* Nature 467 929-934 doi:10.1038doi:10.1038/nature09486

56. La Scola B et. al. 2008 *The virophage as a unique parasite of the giant mimivirus* Nature 455 100-104 doi:10.1038/nature07218

57. Lawton G 2009 *Why Darwin was wrong about the tree of life* New. Sci. 2692 21 Jan.

58. Leipe D, Aravind L, Koonin E (1999) *Did DNA Replication evolve twice independently?* Nucleic Acids Research 27/17 3389-3401.

59. Li, Jun et. al. 2008 *Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation* Science 319 1100 DOI: 10.1126/science.1153717.

60. Lisch D 2008 *A new SPIN on horizontal transfer* PNAS 105/44 16827-16828.

61. Lyon Mary F. 2000 *LINE-1 elements and X chromosome inactivation: A function for "junk" DNA?* PNAS June 6 97/12.

62. Makarova K, Koonin E 2003 *Comparative genomics of archaea: how much have we learned in six years, and what's next?* Genome Biology 4 115.

63. Martin, William and Russell Michael J. (2003) *On the origins of cells: a hypothesis for the evolutionary chemoautotrophictransitions from abiotic geochemistry to prokaryotes, and from prokaryotes to nucleated cells* Phil. Trans. R. Soc. Lond. B 358 59-85.

64. Maxmen A 2010 *Virus-like particles speed bacterial evolution* Nature doi:10.1038/news.2010.507 http://www.nature.com/news/2010/100930/full/news.2010.507.html

65. Mekel-Bobrov N, Gilbert S, Evans P, Vallender E, Anderson J, Hudson R, Tishkoff S, Lahn B 2007 *Ongoing Adaptive Evolution of ASPM, a Brain Size Determinant in Homo sapiens* Science 309 1720-2

66. Mi, S. et. al. 2000 *Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis.* Nature 403: 785-789.

67. Moreira D, Lopez-Garcia P 1998 *Symbiosis Between Methanogenic Archaea and δ-Proteobacteria as the Origin of Eukaryotes: The Syntrophic Hypothesis* J Mol Evol 47 517-530.

68. Moyzis Robert 2005 Proc. of the Nat. Acad. Sci. DOI: 10.1073/pnas.0509691102)

69. Nakamura et al., 1997 *Telomerase Catalytic Subunit Homologs from Fission Yeast and Human* Science 277 955-959.

70. Pace J, Gilbert C, Clark M, Feschotte C 2008 *Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods* PNAS 105/44 17023–17028.

71. Pace N 1997 *A Molecular View of Microbial Diversity and the Biosphere* Science 276 734-40.

72. Powner, M. W., Gerland, B., Sutherland, J. D. 2009 Nature 459, 239-242 (See also Wade N 2009 *Chemist Shows How RNA Can Be the Starting Point for Life* http://www.nytimes.com/2009/05/14/science/14rna.html)

73. Raoult D et, al. 2004 *The 1.2-Mb Genome Sequence of Mimivirus* Science 306 1344-50. doi:10.1126/science.1101485

74. Reanney D 1974 *Viruses and evolution* Int. Rev. Cytol. 37 21-55.

75. Reanney D 1975 *A regulatory role for viral RNA in eukaryotes* J. Theor. Biol. 49 461-92.

76. Reanney D 1976 *Extrachromosomal elements as possible agents of adaption and development* Bact. Rev. 40 552-90.

77. Reynolds Wanda 1995 PNAS 92, 8229.

78. Rivera M, Lake J 2002 The ring of life provides evidence for a genome fusion origin of eukaryotes Nature 431 152-5.

79. Rohde D, Olson S, Chang J 2004 *Modelling the recent common ancestry of all living humans* Nature 431 562- 565.

80. Sasidharan R, Gerstein M 2008 *Protein fossils live on as RNA* Nature 453/5 729-32

81. Schmid C. 1998 *Does SINE evolution preclude Alu function?* Nucleic Acids Research 26 4541-4550.

82. Tchénio T., Casella J-F, Heidmann T., 2000 *Members of the SRY family regulate LINE retrotransposons* Nuc. Acid Res. 28/2 411-425.

83. Tishkoff Sarah, Verrelli Brian 2003 *Patterns of Human Genetic Diversity: Implications for Human Evolutionary History and Disease* Annu. Rev. Genomics Hum. Genet. 4:293–340.

84. Trelogan Stephanie, Martin Sandra 1995 *Tightly regulated developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis* Proc Nat Acad Sci 92 1520 .

85. Underhill P, Passarino G, LinA, Shen P, Mirazo M, Lahr N, Foley R, Oefner P, Cavalli-Sforza L, 2001 *The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations* Ann. Hum. Genet., 65, 43-62.

86. Villareal L, Defilippis V 2000 *A Hypothesis for DNA Viruses as the Origin of Eukaryotic Replication Proteins* Journal of Virology 74/15 7079-84.

87. Villareal L *The Viruses That Make Us: A Role For Endogenous Retrovirus In The Evolution Of Placental Species* http://www.dbc.uci.edu/~faculty/villarreal/new1/erv-placental.html

88. Woese C, Magrum L, Fox G 1977 *Phylogenetic structure of the prokaryotic domain: The primary kingdoms* PNAS 74/11 5088-90,

89. Woese C, Magrum L, Fox G 1978 *Archaebacteria* J. Mol. Evot. 11 245-52

90. Woese C. 1987 Bacterial Evolution Microbiological Reviews 51/2 221-271

91. Woese C, Kandler O, Wheelis M 1990 *Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya* PNAS 87 4576-9.

92. Woese C 1998 *The universal ancestor* PNAS 95 6854-9.

93. Woese C, Olsen G, Ibba M, Soll D 2000 Microbiol . Mol . Biol . Rev. 64, 202–236.

94. Woese C 2002 *On the evolution of cells* PNAS 99/13 8742-7.

95. Xiong Y., Eickbush T.H. 1990 *Origin and evolution of retroelements based upon their reverse transcriptase sequences* EMBO J. 9/10 3353-62.

96. Zhivotovsky, L., Rosenberg N, Feldman M 2003 *Features of Evolution and Expansion of Modern Humans, Inferred from Genomewide Microsatellite Markers* Am. J. Human Genetics May

97. Tishkoff, Sarah A.; Reed, Floyd A. , et. al. (2009) *The Genetic Structure and History of Africans and African Americans* www.sciencexpress.org 30 April DOI:10.1126/science.1172257 .