Why the shared mutations in the Hominidae exon X GULO pseudogene are not evidence for common descent

Royal Truman and Peer Terborg

The GULO X pseudogene has been used for years as evidence for evolutionary theory. Since Hominidae are claimed to have evolved from a rat common ancestor, the modern rat GULO sequence was used as the outgroup in phylogenetic tree building. Our analysis shows that the sophisticated mathematical treatment and the conclusion that differences from the rat sequence fits perfectly with an evolutionionary model for random neutral mutations was never warranted, even had the rat sequence been representative of the intact GULO gene.

Examination of the nucleotide sequences of an expanded dataset of functional GULO genes revealed that the rat gene has undergone exceedingly rapid mutations (or reflects a separate design). Many papers on exon X point out that evolutionary theory predicts mutations on *pseudogenes* to be far more rapid than on *genes* since mutations on functionally important genes would often lead to proteins with new undesirable amino acids and therefore be subject to purifying selection. Our expanded dataset contradicts this assumption, since many putative nucleotide mutations found only in the rat genome did not lead to new amino acids. In the absence of novel amino acids accelerated positive selection is precluded, leaving no rational evolutionary reason as to why these unique changes should have fixed throughout huge rat populations.

We examined the expanded dataset from both an evolutionary and a creation science point of view. Since evolutionary theory assumes a common ancestor for all the organisms in this dataset, we used a nucleotide consensus sequence of intact GULO exon X instead of the rat sequence to re-analyse the available data. Much of the data reflects statistical coincidences, and we explain with Bayes' rule how such artefacts are misleading our evolutionist colleagues. Clusters of separate designs and the presence of informative nucleotide patterns for regulatory purposes provide an alternative to a common identical ancestral GULO gene.

GULO pseudogenes have been intensively investigated and reported[1–4] by Professor Nishikimi and his colleagues. Exon X (number 10) is often presented as providing strong support for neo-Darwinian theory. Three apparently airtight arguments have been advanced:

1. 'Eighteen out of the 164 nucleotides compared are common to the primate species but different from the respective corresponding nucleotides of the rat sequences, indicating that the nucleotide substitutions at these positions occurred after the divergence of those primates from the rat.'[1]
   The authors then used the presumed branching dates for various organisms and the assumption of mutations free from natural selection to estimate rates of nucleotide (nt) substitutions.[2] These were calculated to be 2.75×10^{-9}, 1.2×10^{-9} and 2.15×10^{-9} substitutions/site/year for various lineages, and agreed with the evolutionary estimate obtained from a comparison of six kinds of genes between humans and Old World monkeys: 2.3×10^{-9}.
2. The authors then examined the proportions of synonymous and nonsynonymous substitutions of the amino acids in the sequences of their dataset. They pointed out that 'In the case of functional genes, nonsynonymous substitutions generally occur less frequently than the synonymous substitutions, because substitutions in the former are restricted by the selective pressure during evolution.'[5]
3. Finally, 'The result showed that many of the amino acid substitutions are nonconservative.'[5] 'These findings indicate again that the mutations in the primate GULO genes occurred without functional restriction after the loss of its function.'[5]

In an earlier draft of this paper the original data was re-analysed and it was argued that neutral mutation would not lead to the pattern reported. Fortunately, we realized that we would tacitly be assuming that in the distant past both rat and Hominidae (humans, macaques, orangutans, gorillas and chimpanzee) all indeed had identical GULO genes. In reading the literature, it was clear that during the twenty years or so of studying the GULO pseudogene, scientists and reviewers were so sure evolutionary theory was true that no one even thought to question something so basic. An identical, common ancestral GULO gene must have existed for the organisms studied, according to evolutionary theory, but distinct creation does not share this compulsion.

We decided to collect more data, and discovered everyone was wrong in their interpretations of the exon X sequences.[6] We offer here a new approach. Perhaps new data in the future will require a re-evaluation of our best efforts to date. This is the nature of science, especially involving non-reproducible, non-testable facts from ancient history.

The evolutionary interpretation is wrong

We collected exon X sequences reported for the GULO pseudogene of orangutan, human, chimpanzee, macaque and guinea pig genomes and discovered (table 1) that all these sequences shared the same nucleotide at nine positions which differed from that of the rat, whose GULO is functional. This fact has to our knowledge not been published until now.

A review of the papers which analysed this pseudogene revealed that the phylogenetic relationships shown in figure 1[1–3] was assumed. The authors are aware[2] that others claim the guinea pig lineage branched off preceding the 80 Ma (million years) common ancestor shown in figure 1, meaning that rats would be considered by these evolutionists to be more closely related to humans than to guinea pigs. Among evolutionary taxonomists there is considerable controversy as to where guinea pigs fit in. Nevertheless, on the basis of their analysis Nishikimi and his colleagues claim the GULO pseudogene in guinea pig formed about 20 Ma irrespective of which phylogenetic tree is to be believed.

We see now the difficulty. Over half of the supposedly random mutations in the primate and guinea pig pseudogenes are in fact identical! Although the number of mutations found is small, when they did occur the same nucleotide resulted, and then these putative mutations tended not to change afterwards. We need not continue critiquing the evolutionary interpretations, because we will see shortly that a fatal assumption was made, repeated in multiple subsequent papers by other evolutionists.

The first creation-based interpretation was wrong

To our knowledge, no evolutionist has so far raised questions about the datasets used in the various papers nor interpretations thereof. Since at least 3 lines of reasoning, pointed out above, all seemed consistent with the neo-Darwinian framework, there seemed no reason to do so. This led to an obvious non-evolutionary interpretation. Since obtaining so many identical mutations on both alleles of the pseudogene by chance made no sense, then we surely have here some extreme hot-spots accompanied by biased mutations. This would also help explain the shared deletion found in the Hominidae samples at nt position 97 (table 2).

Physico-chemical features in the vicinity of position 97 may well be responsible for facile deletions. Should this occur in very low populations, such as during or shortly after the Flood, this genetic bottleneck would permit fixing throughout the populations.

Where everyone went wrong

In claiming multiple mutational hotspots in exon X, we implicitly assumed that the sequences had long ago been identical and subsequently mutated. This might make sense from an evolutionary point of view, and is an example where this theory has led researchers astray. The model envisions speciation with formation of new lineages and new morphological traits, although sequence comparisons often suggest chimerical mixtures of genes, with various ones resembling different ancestral relationships.

But the creation scientist has no reason to assume initially identical gene sequences across disparate organisms. We should not even assume that a single male and female individual of a specific 'kind' surviving the Flood shared 100% identical gene sequences. This is an important issue, since it seems very likely that multiple genome variants were created among the same unicellular 'species', and these have exchanged genes and portions of genes over thousands of years. The same or similar nucleotide or amino acid sequences could have been deliberately created across independent species for functional reasons, since multiple codes, and not only the one coding for amino acid sequences, are superimposed. A large number of requirements at the gene, mRNA and protein level would in all likelihood often be optimally satisfied, according to environmental details, by different sequences. Bioinformatic tools, such as alignment algorithms and especially tree-drawing tools, automatically guide the researcher into thinking similar sequences mutated from a common ancestral version. We believe a correct interpretation of the sequence data should focus on the functional purposes of various sequences of amino acids and nucleotides, independent of evolutionary speculations.

In the case of the dataset we have provided, if the initial state for the Hominidae and guinea pig could reasonably have been different, can we be sure that a large number of hotspot mutations have indeed occurred?

It seemed prudent to examine the exon X sequences from as many other organisms as we could. Blast searches[7]

Table 1. Aligned nucleotides for hominidae exon X of GULO pseudogene, rat and intact GULO consensus sequences. Identical nucleotides not shown.

	1	10	16	19	22	28	31	34	37	39	40	47	48	50	58	59	61	72	79	81	91	92	94	100	106	109	111	121	133	157
Orang:	A	C	G	C	G	G	G	G	C	T	G	G	C	C	G	G	G	A	G	G	C	A	C	G	C	C	T	C	G	C
Human:	A	C	G	C	G	G	G	G	C	T	G	G	C	C	G	G	G	A	G	G	C	A	C	G	C	C	T	C	G	C
Chimp:	A	C	G	C	G	G	G	G	C	T	G	G	C	C	G	G	G	A	G	G	C	A	C	G	C	C	T	C	G	C
Macaque:	A	C	G	C	G	G	G	G	C	T	G	G	C	C	G	G	G	A	G	G	C	A	C	G	C	C	T	C	G	C
Guinea Pig:	A	C	G	C	G	G	G	G	C	T	G	G	C	C	G	G	G	A	G	G	C	A	C	G	C	C	T	C	G	C
Consensus:	A	C	G	C	G	G	G	G	C	T	G	G	C	C	G	G	G	**A**	G	A	C	A	C/T	G	C	C	T	C	G	C
Rat:	**G**	C	G	C	**A**	G	**A**	G	C	T	G	G	C	C	**A**	G	G	**A**	**A**	A	C	A	C	**C**	C	**T**	T	C	G	**T**

0 cases where mutations are actually involved
9 cases where biased mutations had seemed to be occurring (bold)

were performed leading to a more complete dataset (table 2). Sequences were aligned using ClustalX[8] and phylogenetic trees were generated with programs *dnapars* (figure 2) and *dnaml* (figure 3) which are part of the *phylip* package.[9] The trees created with these alternative algorithms differed somewhat, but general pattens could be discerned. We can interpret figure 2 and figure 3 as simply reflecting how similar different GULO genes are. Evolutionary theory would predict a common gene ancestor several hundred million years ago for the chicken and all other organisms in table 2. The extant seven organisms possessing an intact GULO gene would have arrived through different lineages from a common starting point involving the same amount of time. To a first approximation, these should be roughly equidistant from a central point for all these organisms, point *p1* in figure 2 and figure 3.

We observed several interesting features in the computer generated trees, which we interpret using evolutionary reasoning:

1. Rat vs mouse: these supposedly share a recent common ancestor. Since their divergence, the rat genome seems to have mutated abnormally fast.
2. Pig vs chicken: the number of mutations should have been comparable, but chicken exon X seems to have mutated four times as fast.
3. Cow vs pig: the pattern of mutation implies a common ancestor a short time ago.
4. Hominidae vs point P1: the Hominidae seem to have evolved from the overall common ancestor, P1, and not rat. The evidence fails to support a Hominidae/rat common ancestor.
5. A large number of mutations seem to have occurred between point P1 and a common ancestor for human (*Hu*), chimpanzee (*Ch*), orangutan (*Or*) and macaque (*Ma*), at point P2.
6. The tree suggests a possible clustering into three categories of GULO gene: bird, Hominidae and non-Hominidae mammals.

Since the sequence studied is rather small, 164 nts only, statistical abnormalities could occur, but our observations may suggest theories for further investigation. Points 1–4 above make no sense in evolutionary terms. Thinking as evolutionists on point 5, we would ignore the tree structure implied by the data and assume that a common Hominidae ancestor, near point P2, had evolved from a rat-like ancestor, although the data here does not support this claim. Based on an evolutionary timescale, we are forced to conclude that between the time implied by points P1 and P2, the rate of mutation of exon X was about an order of magnitude faster than experienced by organisms with a presently intact GULO gene.

Observations 1 and 4 motivate a more careful examination of the raw data. We discovered that the rat Exon X sequence is not representative of intact GULO sequence. Specifically, eight times[10] (table 1) a different nucleotide was found in the rat exon X *which was identical to all the exon X of intact GULOs*! Comparing rat and mouse, which supposedly share a relatively close ancestor (table 2),[11] revealed three nt differences, although the nts found in mouse exon X were identical with those of all other organisms in the dataset, and a fourth nt (position 22) showed only one exception. Therefore, our evolutionist friends erred in comparing Hominidae sequences with that of rat exon X. In an evolutionary interpretive framework, the pseudogenes should be compared to the intact GULO genes, excluding the one from rat.

In building the consensus sequence (table 1) the data from guinea pig was taken into account. Evolutionary phylogenetic theory places a divergence of rat and guinea pig at either > 80 Ma,[12–15] or, according to others, about 60 Ma ago.[16–19] In either case, GULO is claimed to have been destroyed by a mutation only about 20 Ma ago.[2,20] Sometimes more than one nt was found in intact GULO genes at that location and the sequence from guinea pig favoured the majority, decreasing uncertainty as to whether the correct choice for the consensus nt was made.

We compared next the sequences of pseudogene exon Xs and the consensus of the functional version. At many nucleotide positions different nts were found in the case of the function GULO, but the same nt was seen for all the

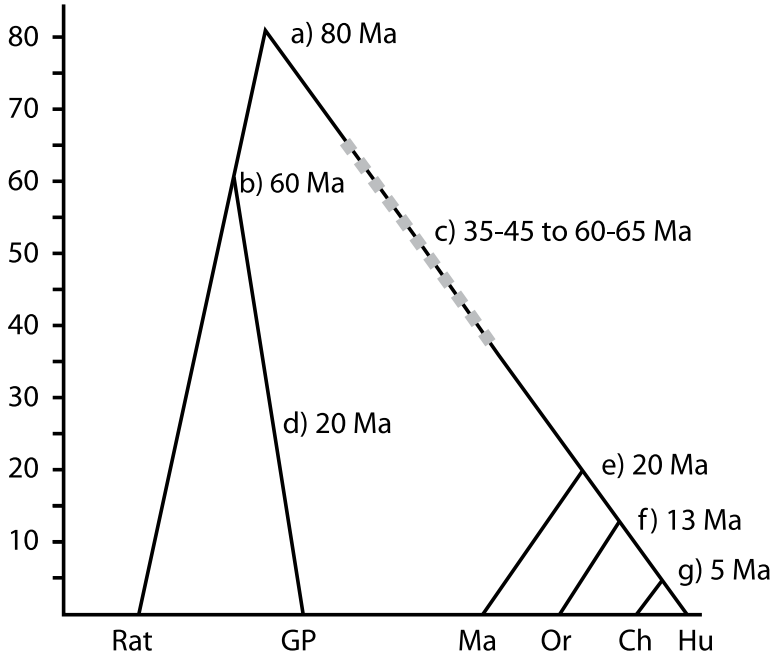

Figure 1. Evolutionist phylogenetic relationships. Some taxonomists place a common Rat/GP ancestor > 80 Ma ago. Abbreviations: GP: guinea pig; Ma: macaque; Or: orangutan; Ch: chimpanzee; Hu; human. Ma: million years

Table 2. Aligned nucleotide seq-uences of exon X from GULO genes and pseudogenes. Truman and Terborg dataset. Positions with identical nucleotides not shown.

	1	2	10	12	13	15	16	19	22	28	29	31	34	35	36	37	38	39	40	46	47	48	49	50	55	56	58	59	61	62	63	64	65	72
Orang	A	A	C	C	G	A	G	C	G	G	C	G	G	G	C	C	A	T	G	G	G	C	C	C	T	G	G	G	G	G	T	G	T	A
Human	A	A	C	C	G	A	G	C	G	G	C	G	G	G	C	C	G	T	G	G	G	C	C	C	T	G	G	G	G	G	T	G	T	A
Chimp	A	A	C	C	G	A	G	C	G	G	C	G	G	G	C	C	A	T	G	G	G	C	C	C	C	G	G	G	G	G	T	G	T	A
Macaque	A	A	C	C	A	G	G	C	G	G	A	G	G	G	C	C	A	T	G	G	G	C	C	C	T	G	G	G	G	G	T	G	T	A
GuineaPig	A	G	C	A	G	A	G	C	G	G	C	G	G	A	G	C	A	T	G	A	G	C	T	C	C	A	G	G	G	G	C	A	G	A
Mouse	G	G	C	A	G	A	G	C	G	G	C	A	G	G	C	C	A	T	G	G	G	C	C	C	C	A	G	G	G	G	T	A	G	A
Cow	A	G	C	A	A	A	G	C	G	G	C	G	G	G	C	C	A	T	G	G	G	C	G	A	C	A	G	G	A	G	T	G	G	A
Chicken	T	G	A	A	G	A	A	A	G	G	C	G	G	G	C	T	G	C	C	G	A	A	C	A	C	A	G	A	G	G	T	G	G	A
Pig	A	G	C	A	G	A	G	C	C	G	C	G	G	G	C	C	A	T	G	G	G	C	C	C	C	A	G	G	G	G	T	G	G	A
Dog	A	G	C	A	G	A	G	C	G	A	C	G	A	G	C	C	A	T	G	G	G	C	C	C	C	A	G	A	G	G	T	G	G	T
Rat	G	G	C	A	G	A	G	C	A	G	C	A	G	G	C	C	A	T	G	G	G	C	C	C	C	A	A	G	G	G	T	A	G	A

	73	75	76	79	81	83	85	91	92	94	95	96	97	98	99	100	101	103	109	111	112	114	115	118	121	127	128	130	131	133	134	135
Orang	C	C	G	G	G	G	G	C	A	C	C	A	*	G	A	G	G	T	C	T	A	T	G	C	C	C	C	G	C	G	G	A
Human	C	T	G	G	G	G	A	C	A	C	T	G	*	G	A	G	G	T	C	T	A	T	G	C	C	C	C	G	T	G	G	A
Chimp	C	T	G	G	G	C	A	C	A	C	T	G	*	G	A	G	G	T	C	T	A	T	G	C	C	C	C	G	C	G	G	A
Macaque	A	C	G	G	G	G	G	C	A	C	C	A	*	A	G	G	G	T	C	T	A	T	G	C	C	C	C	G	C	G	G	A
GuineaPig	C	C	T	G	G	G	G	C	A	C	C	G	G	G	G	G	G	C	C	T	G	T	G	C	C	C	C	G	A	G	G	A
Mouse	C	C	C	G	A	G	G	C	A	C	C	G	A	G	G	T	G	T	C	T	G	T	G	C	G	C	C	G	A	G	G	A
Cow	C	C	C	G	A	G	A	C	A	T	C	G	C	G	G	G	G	C	C	T	G	T	G	C	C	C	C	G	A	C	G	A
Chicken	C	C	T	G	A	G	G	T	G	T	C	G	A	G	C	G	G	T	C	G	G	T	G	C	C	C	C	G	C	G	G	A
Pig	C	C	C	G	A	G	G	C	A	T	C	G	G	G	C	G	G	C	C	T	G	T	G	C	C	C	C	G	A	G	G	A
Dog	C	C	T	G	A	G	C	C	A	C	C	G	C	G	G	G	G	T	C	T	G	T	G	C	C	C	C	G	A	G	G	A
Rat	C	C	C	A	A	G	G	C	A	C	C	G	A	G	G	C	G	T	T	T	G	T	G	C	C	C	C	G	A	G	G	A

	136	137	138	139	145	146	147	148	149	154	155	157	158	(GenBank)
Orang	C	A	G	C	T	C	T	G	A	C	A	C	C	>gi|4589754
Human	C	A	G	C	C	C	T	G	A	C	A	C	C	>gi|493656
Chimp	C	A	G	C	C	C	T	G	A	C	A	C	C	>gi|4589757
Macaque	C	A	G	C	C	C	T	G	G	C	A	C	C	>gi|4589758
GuineaPig	C	A	G	C	C	A	T	G	A	C	T	C	A	>gi|62899630
Mouse	C	A	G	C	C	A	T	G	A	C	A	T	A	>gi|38325769
Cow	C	A	G	C	C	A	T	G	A	C	A	C	A	>gi|77404230
Chicken	C	A	G	C	C	A	T	G	A	C	A	C	A	>gi|46425804
Pig	C	A	G	C	C	A	T	G	A	C	A	C	A	>gi|24637282
Dog	C	A	G	C	C	A	T	G	A	C	A	C	A	>gi|73993943
Rat	C	A	G	C	C	A	T	G	A	C	A	T	A	>gi|60683826

pseudogenes. We compared these with the consensus and the rat sequence (table 1). If the rat were representative of intact GULO, the Hominidae and guinea pig common ancestor would have had a similar sequence. Thus one would have to conclude there were 9 cases of biased mutations leading to the exact same nt at those pseudogene-derived positions. Until we had our full dataset, this seems like an obvious conclusion.

However, when compared to the consensus sequence of intact GULO genes, all but one (position 81) of the examples of supposed biased mutation are shown to be incorrect.

An evolutionary re-analysis

An evolutionist could use our dataset and reinterpret earlier claims. The usual approach here would be to simply state that exon X of the rat is not useful for phylogenetic purposes, claim phylogenetic relationships in figure 1 are correct and to ignore the issue. Perhaps rat exon X mutated very rapidly for unknown reasons, although unexpected and not easily explainable in evolutionary terms.

Ignoring rat exon X

We wish to emphasize that neglecting discordant data, risks maintaining a theory *contra* available evidence. Rat populations in the world are huge, and one would be forced to argue strong positive selection must have been at play. Now, relatively few different amino acids are found at the variable residue positions in exon X of intact GULO genes (table 3). This makes it easy to create a reliable consensus sequence, which presumably indicates what the ancestral sequence was. A comparison of this consensus sequence with the rat sequence suggests that 9 putative mutations in exon X region 164 did not actually result in any different amino acids! We also observe that although rat exon X appeared to mutate very fast *following* divergence from the new mouse lineage, the modern exon X amino acid sequences of these two organisms are identical. Given the universal evolutionary belief that only non-synonymous mutations (which code for other amino acids) can be recognized by natural selection in higher organisms, we are left with no evolutionary reason as to why a lineage with many silent mutations in exon X should fix in a large rat population. We caution that only three sequences of this exon are available in GeneBank, all from *Rattus norvegicus* and all reported by Nishikimi. Although these three reported sequences are 100% identical and no one questions that this data is representative of the world rat population, geneticists have pointed to compelling evidence that accelarated mutation rates is leading to genome meltdown. We elaborate below on a book written by genetics Professor Sanford, and hope that additional rat specimens will be examined.

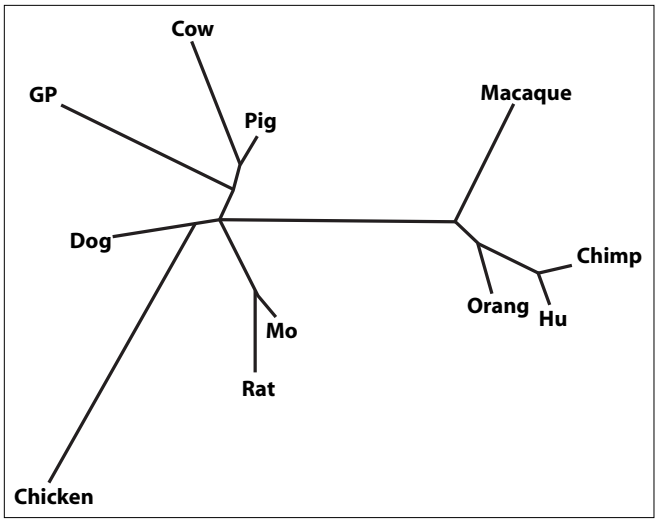

Figure 2. Degree of similarity based on exon X of GULO genes. Sequences aligned with ClustalX, tree generated with *dnapars* software from the phylip package. p_1: approximate location of consensus sequence for intact GULO exon X; p_2: approximate location of common ancestor for hominidae according to evolutionist theory. Abbreviations: Mo: mouse; GP: guinea pig; Chick: chicken; Macaq: macaque; Orang: orangutan; Hu: human; Chimp: chimpanzee

Now, if enough fairly neutral mutations occurred among rat-like creatures for tens of millions of years some could eventually fix throughout most the population. But all the rat[21] and all the mouse[22] exon X nt sequences reported in the GenBank7 were 100% identical. Since allele copies from both parents are identical, there is no evidence for millions of years of random mutations. Only limited replicate data for exon X is available through online databases, but these imply very little variability within the same major taxonomic group: humans, 100% identity;[23] pigs, 100% identity;[24] cows, 1 nt difference (A or G at position 51).[25]

Detailed examination of the sequences

Using evolutionary thinking, we would treat rat exon X as anomalous and assume that a common ancestor would have a sequence based on a consensus of the organisms with functional GULO sequences. This is reasonable, since at almost all positions all or most had the same nt (table 2). We found it very difficult to honestly play evolutionist advocate with this dataset, however. The main problem, in addition to various discordances already mentioned, is that only a few positions seem to have mutated and yet generally only two of the four nts (A, C, T, or G) were found. The patterns could not be explained by assuming common ancestors at key points.

The reader is invited to peruse nt positions 13, 38, 50, 59, 64, 76, 94, 97, 103, 131 and 132 of table 2. For example, at position 94 all organisms, including those having only pseudogenes, posses a *C* nucleotide, except for cow, chicken and pig, which all have a *T* nucleotide. Or, at position 103 eight organisms display a *T* nucleotide (leading to codon GA*T*, including those having GULO pseudogenes, whereas guinea pig, cow and pig, display a *C* (leading to codon GA*C*). Both codons translate to aspartic acid. *C* ⇨ *T* are more likely than *C* ⇨ *A* and *C* ⇨ *G* mutations, but the entire pattern of putative mutations shown in table 2 makes no evolutionary sense if assuming a common ancestor. Conceivably, mutations at some positions could be so strongly biased that most members of different taxa could soon display a particular mutation. This remains to be determined. Should this be the case, evolutionists can no longer argue, however, that various patterns reflect common ancestry. If these sets of three nts had conformed to an obvious evolutionary phylogenetic interpretation, no one would have questioned the strength of this *pro*-evolutionism evidence. The fact is that sequence databases are full of the right patterns associated with the wrong evolutionary trees.[26] Are we obliged to believe that random and yet somehow identical mutations occur again and again whenever evolutionary inconsistencies are found?

Evolutionists are also reporting[27] that phylogenetically discordant sequence patterns are increasingly being found, forcing creative new constructs such as 'homoplasy', or 'convergent evolution'. We must point out, however, that the same combination of mutations will rarely be generated by chance, and natural selection cannot have a guiding long-term goal. Since the Luria and Delbrück experiments in the 1940s,[28] most biologists consider mutations to be independent of environmental signals and also to only be a randomly generated first approximation.

Another observation based on the full dataset (table 2) does not lend itself to easy evolutionary interpretation. The mutations on exon X of the GULO pseudogene are too narrowly concentrated. Based on table 2, and using the consensus sequence, a large number of mutations seem to have occurred at the same position between time lapse *t1* = creation of the primate pseudogene and when the primate common ancestor lived. This is the only sensible way to

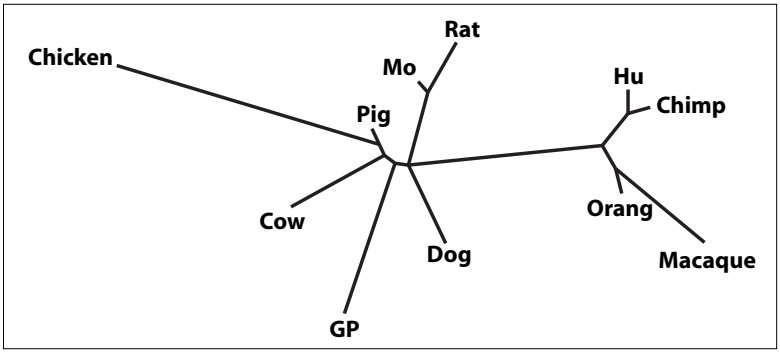

Figure 3. Degree of similarity based on exon X of GULO genes. Sequences aligned with ClustalX, tree generated with *dnaml* software from the phylip package. p_1: approximate location of consensus sequence for intact GULO exon X; p_2: approximate location of common ancestor for hominidae according to evolutionist theory. Abbreviations: Mo: mouse; GP: guinea pig; Chick: chicken; Macaq: macaque; Orang: orangutan; Hu: human; Chimp: chimpanzee

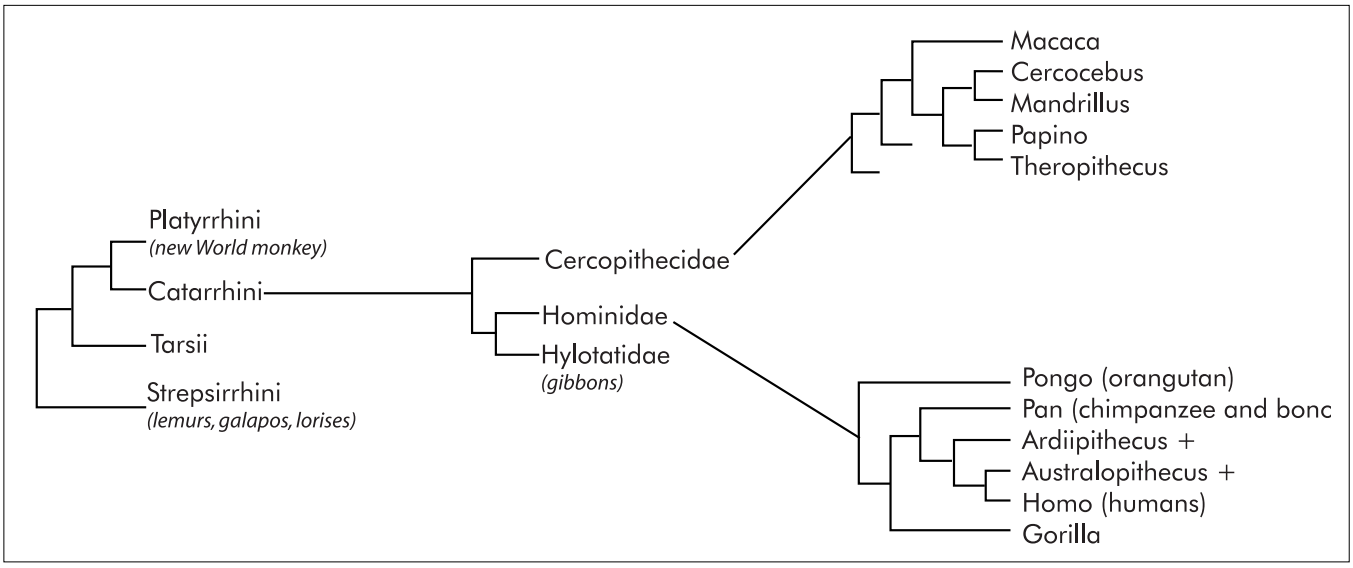

Figure 4. Evolutionist phylogenetic relationships for primates. (From Arnason *et al.*, [37] p. 8155)

explain the large number of times the same nts are found in the dataset for all primates which differ from the consensus (see positions 2, 12, 56, 65, etc.). (Don't overlook that in table 2, only data is shown for nt positions which are not identical for all organisms in the dataset. Identical nts are not informative for the purposes of our analysis).

It seems that time t1 is too short for neutral mutations to produce this effect. We need to recall some facts. Rats create vast amounts of vitamin c because they apparently need to do so.[29] Why a damaged genome would be favoured is not apparent. Based on Drake's studies,[30] a rodent or mammal should undergo a mutation rate of about $2x10^{-10}$ nts/generation. For this 164 nt exon, assuming an average generation time from one to 30 Ma (others claim 41 Ma) implies that only one or two mutations would be generated per creature. It seems that both parental alleles are identical at these positions in extant Hominidae in spite of the absence of positive selection. This suggests strong inbreeding between siblings, with all the genetic disadvantages this would generate. Even then, most new mutations would be lost by genetic drift.

The deletion of nt position 97

A very intriguing observation is the reading frame shift deletion at nt position 97 in the GULO pseudogene of Hominidae (table 2). This also seems to imply a common ancestor, since once it occurred repair by subsequent point mutations would be unlikely. We need to establish that the same deletion is present in exon X of gorillas, an evolutionary phylogenetic requirement (figure 4). It has been reported[4] that GULO is non-active in New World monkeys (spider monkey and squirrel monkey). If the deletion is also found there, then it needs to be present also in gibbons (figure 4). It would also be sensible to examine several geographically widely separated Hominidae samples. Some comments need to be made here:

For the six species in our dataset with intact GULO genes, three different nts were found at this position (97). This looks like a mutational hotspot. Apropos hotspots, the data summarized in table 2 indicates strongly that at the time point mutations supposedly occurred, they all did so at the same location in a manner that cannot explainable by common descent.[31] For example, the same mutation is implied at position 13 for cow and macaque, although the latter is a pseudogene. At position 38, the same mutation is implied for human and chicken. In virtually all cases when mutations are assumed, we find only two different nts were present.[32]

Exon VI of the guinea pig pseudogene also shows several deletions[3] within a very short sequence (data not included here). Data for the corresponding exons of other organisms with pseudogenes is not available, but nevertheless, this demonstrates that deletional hotpots is a feasible notion.

Evolutionary theory did not predict this deletion. We wish to emphasize again the huge potential for pattern coincidences, some of which may be accommodated *postfacto* into any of many evolutionary scenarios. We shall be happy below to also propose *post-facto* possibilities from a creation scientist perspective.

A new analysis based on creation science

In earlier work, one of us analysed[33] the amino acid sequence variability of ubiquitin chains. Since evolutionists assume that most mutations are randomly generated and rarely offer a selective advantage, there seemed no reason to search for special patterns which could explain the differences in terms of informative signals. But creation scientists realize that this hinders a deeper scientific search for hidden design principles which may be embedded in genomes. Deprecating terms such as 'junk DNA', in our opinion, has not only hindered research but is based on a

Table 3. Aligned amino acid sequences of exon X from GULO genes and pseudogenes. Truman and Terborg dataset. Positions with identical amino acids not shown.

	1	4	5	10	12	13	16	17	19	20	21	22	24	25	27	28	31	32	33	35	36	37	41	44	47	49	50	52	53
Orangutan (a)	K	T	E	L	A	M	A	H	E	V	V	S	Y	P	G	V	T	H/Q	E	D	V	L	C	Q	R	L	N	N	L
Macaque (b)	K	T	G	M	A	M	A	H	E	V	V	S	END	P	G	V	T	H/Q	R	D	I	I	C	Q	C	L	D	N	L
Human (c)	K	T	E	L	A	V	A	H	E	V	V	S	Y	L	G	V	T	C/W	E	D	I	L	C	W	R	L	N	N	L
Chimpanzee (d)	K	T	E	L	A	M	A	H	E	V	V	S	Y	L	G	L	T	C/W	E	D	I	L	C	R	R	L	N	N	L
GuineaPig (e)	E	K	E	L	S	M	A	H	K	V	A	A	Y	P	G	V	T	R	G	D	I	L	S	R	C	M	N	C	I
Rat (f)	E	K	E	L	A	M	A	H	K	V	V	A	Y	P	E	V	T	R	G	D	I	L	C	R	C	M	N	I	M
Mouse (g)	E	K	E	L	A	M	A	H	K	V	V	A	Y	P	E	V	T	R	G	D	I	L	C	R	C	M	N	I	M
Pig (h)	E	K	E	L	A	M	A	H	K	V	V	A	Y	P	E	V	T	R	A	D	I	L	C	R	C	M	N	I	M
Dog (i)	E	K	E	L	A	M	A	H	K	M	V	A	F	P	E	V	T	R	G	D	I	L	C	R	C	M	N	I	M
Cow (j)	E	K	E	L	A	M	A	N	K	V	V	A	Y	P	E	V	T	R	G	D	I	L	C	R	C	M	N	I	M
Chicken (k)	E	K	E	L	A	A	N	N	K	M	V	A	Y	P	E	V	A	R	A	E	I	W	C	R	C	M	N	I	M

(a) Pongo_pygmaeus(gi|4589754)
(b) Macaca_fascicularis(gi|4589758)
(c) Homo_sapiens(gi|493656)
(d) Pan_troglodytes(gi|458975
(e) Cavia_porcellus(gi|6C/W899630)
(f) Rattus_norvegicus(gi|606838C/W6)
(g) Mus_musculus(gi|383C/W5769)
(h) Sus_scrofa(gi|C/W4637C/W8C/W)
(i) Canis_familiaris(gi|73993943)
(j) Bos_taurus(gi|77404C/W3)
(k) Gallus_gallus(gi|464C/W5804)

fundamentally flawed paradigm. One of us also discovered[33] that a unique, almost perfectly invariable three amino acid pattern, was a characteristic of the protein ubiquitin in all animals, plants and fungi. This was interpreted as reflecting alternative designs of this protein. The cellular effect of these minor differences has not been elucidated yet. Discoveries such as these warn us to make sure we are not accepting assumptions from our materialist colleagues which are not required by our fundamentally different model.

The evolutionary model clearly predicts a common ancestral GULO gene for all members in the dataset reported here. This is not true of the Creation model. We do not attribute miraculous fine-tuning properties to random mutations plus natural selection. But intelligence can produce designs optimised to reconcile many often contradictory performance goals. This can require differences at the gene, mRNA and protein levels. Therefore, we converted[34] the nucleotide sequences into predicted amino acid sequences to determine if there was any reason to suspect alternative categories of at least the exon X portion of the GULO gene (table 3). We suggest that the data supports this hypothesis and therefore justifies additional sequencing efforts to test such an hypothesis.

An amino acid consensus sequence based on organisms with intact GULO genes (6 very different organisms), plus guinea pig, can easily be made by visual inspection. We observe the following number of differences in exon X from the consensus (table 3):

- Rat and mouse: none
- Pig: one (position 33)
- Dog: two (positions 20, 24)
- Cow: one (position 17)
- Chicken: 8 (positions 13, 16, 17, 20, 31, 33, 35, 37)
- Guinea pig: 6 (positions 12, 21, 27, 41, 52, 53)

All of the differences found in the chicken and guinea pig exon X differ from each other. This confirms the observation12 that guinea pig genes tend to be very different from those of other rodents, *contra* evolutionary morphological expectations.

We also observe, based on table 3, that among pseudogenes in the Hominidae there are 8 candidates sharing the same amino acid differerence with the consensus sequence.

We see how creation science reasoning can provide fruitful guidance for research activities. Evolutionists would have expected a comparable degree of divergence in protein sequence from the consensus for organisms with intact GULO genes. We suggest that 'very different' members of various taxa (e.g. birds) should be examined to see whether families of exon X clusters are found. This should enable us to interpret trees such as those displayed in figure 2 and figure 3.

We would prefer to make predictions based on a thorough knowledge of design factors and not simply on visible morphologies, but we are unfortunately still far from being able to understand cell complexity at a sufficient level to allow us to do this. A creation scientist is not strictly forced to expect similar genetic features in chickens and ducks, for example. An understanding of how protein machines work and how their underlying genes are regulated, would be useful in making predictions about different gene variants. Sequence differences should reflect divergences from ancestral biblical 'kinds'.

A creation science interpretive framework

We can only present here a rough outline of the approach young-earth creation researchers may take in interpreting sequence data. Most of the biblical kinds which survived the Flood through Noah's Ark were represented initially by a single male and female (seven members of clean animals were taken in the Ark). The command to 'be fruitful, and multiply, and fill the earth'[35] and to have dominion over its living creatures seems to imply an intention by God to enrich the whole Earth with a variety of creative life forms. We assume He did not intend for most of the Flood-surviving kinds to simply go extinct a short time afterwards. One can develop detailed, computer supported, mathematical, population genetics models, which like the evolutionary ones, will have many unproven assumptions and adjustable parameters. The more constraints one can

place on such models the quicker testable predictions and research intuitions can be generated. We propose the following principles:

- The number of generations would be limited to about 4,500 years for organisms which survived by being represented in the Ark
- The male and female Ark ancestors may have possessed many different alleles and very different regulatory sequences. We see every reason to question why these alleles should have been identical at that starting point.
- Uncontested ecological niches would have been present for a large number of generations. These would have led to many fragmented, genetically isolated sub-populations for several generations. As the numbers increased, some sub-populations may have subsequently interbred.
- Predators would have reproduced much more slowly than prey, leading to relaxed Darwinian selection. Many mutations which are not deadly would be quickly fixed in the sub-populations.
- Mutation rates may have been exceedingly rapid, especially if high radiation levels were present[36] (evolutionists have also argued for a period of abnormally faster evolution/mutation).[37]
- The genomes were designed to adapt in very short time spans. Adaptation is not predicted to be a trial-and-error process involving mostly random point mutations which, over hundreds of millions of years, could leed to new genes. Genetic adaptation within the lifetime or within a single generation, would be perfectly reasonable and require 'pre-programmed' genetic informational potential.

It is interesting that in some aspects, both the evolutionary model and the Creation model may share many features. The former predicts much slow change over long periods of time, while the later rapid change in short time periods. In cases such as the dog family, both models have the potential to accommodate the number of mutational differences observed between consensus wolf and dingo aligned gene sequences. It is apparent that some of the same mathematical formalisms can be applied to both frameworks. For example, instead of a small mutation rate in a large population with a low level of positive selection, the creation scientist can use the same formalisms for a much smaller effective population and a much higher level of selection. The assumption is that the early genomes had a much higher potential for rapid change. The end results will be similar to a first approximation, but rely on different time scales and initial conditions.

But there are subtle differences which can lead to research proposals. An obvious one involves expectations about the fossil record. Rapid or slow change would lead to different evidence. For example, a factor of 10,000 longer time periods would predict more fossil samples.

A possible difference involves the amount of polymorphisms expected among the different members of a population. For example, if a mouse common ancestor lived about 30 Ma ago, then a huge number of differences would be expected between the extant genomes. Nucleotide mutations can occur on the paternal as well as the maternal allele of any gene, and cross-over during gamete formation is also possible. In the long term, we all expect greater randomness.

The Creation model can accommodate several major mutations. But how much variety is expected within the same species at this time? Sanford recently provided compelling evidence[38] that mutation rates in humans appear to be about a 1,000 times higher than commonly believed. Creation scientists need to collect mutation rate data to determine how much polymorphism would be predicted by a young-earth framework.

The deletion at position 97 revisited

We would like to determine just how wide-spread this deletion really is. If this deletion is absent in gorillas (figure 4), then evolutionists would immediately state that it is a coincidence and no longer a phylogenetic marker. Recall that evolutionary theory did not predict this specific pattern. Once discovered, an evolutionary scenario was then offered.

It would also be prudent to examine several individuals to be certain that this deletion is 100% absent in all modern descendents. Such a finding would not easily be accommodated *postfacto* in an evolutionary framework.

In a world with rapid mutations and low populations, many coincidences are bound to occur (our favoured model). Many of these mutations would fix in the small populations, and be present today. We observe in table 2 that the nt position where a deletion is observed is quite variable among all organisms in the dataset, and may very well represent a mutational hot-spot. This is one reason one needs to ensure that this deletion is totally absent in a population. It is possible that multiple deletions occurred over a period of time, especially when the populations were small, and that some of the modern lineages may still include members lacking the deletions—this scenario only applies to the Creation model.

We pointed out earlier that Exon VI of the guinea pig pseudogene also featured several deletions[3] within a very short sequence. This indicated that deletional hotspots may be present which are not often observed as they may be lethal. But if the gene is already deactivated and superfluous, such deletions may occur frequently.

Incidentally, although an indel has been used as key evidence to support a tetrapod/lungfish phylogenetic topology, one team of evolutionists recently suggested[39] that this could just be another example of homoplasy, or simple coincidence. We would not be the first to suggest that a shared deletion need not reflect common descent.

Finally, we are not sure why other creation scientists are so quick to discard[40] the obvious possibility of divine intention. Double Nobel Prize winner L. Pauling has claimed for years[29] that humans need vast amounts of vitamin C, far more than needed to prevent scurvy, a view supported by the huge amounts known to be produced by

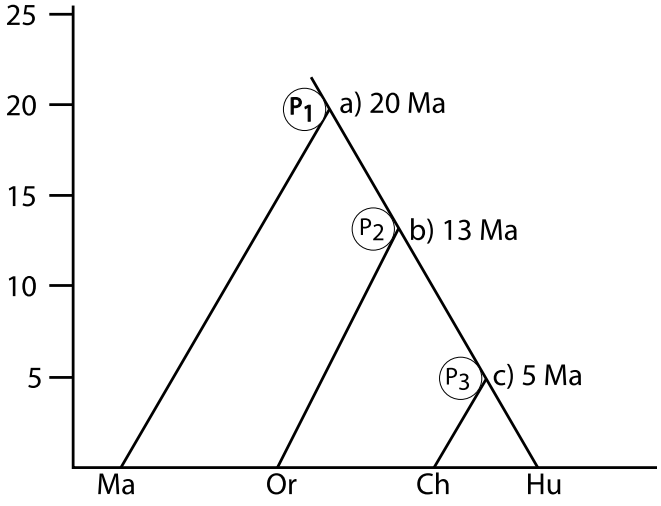

Figure 5. Evolutionist phylogenetic relationships for hominidae. P_1: presumed location of macaque, orangutan, chimpanzee and human common ancestor; P_2: presumed location of orangutan, chimpanzee and human common ancestor; P_3: presumed location of chimpanzee and human common ancestor. Abbreviations: Ma: macaque; Or: orangutan; Ch: chimpanzee; Hu: human

rats. The antioxidant properties of this molecule could originally have been designed to trap damaging free radicals. The presence of these damaging molecules may be partly responsible for the decrease in longevity of human and other primates.

It is possible that shortened generation times in small isolated populations was necessary to express the 'pre-programmed' genetic variability needed to replenish the post-Flood world. One readily thinks of examples such as polar bears, which needed special features ideally suited to their new environment. We must accept that in a strictly non materialistic world, we may not find all the answers if we always exclude divine action during key periods of world history.

Finally, in the Appendix[41] we discuss the pitfalls in interpreting sequence data using examples from the dataset presented here. Many evolutionists are persuaded the data supports in broad terms their viewpoint. We introduce some reasons as to why this is only an illusion.

Conclusions

We had originally intended to point out that identical mutations at the same location in guinea pigs, humans and various monkeys demonstrated that mutations could not have been neutral, but rather extraordinarily biased. This would reinforce the view that the same nucleotide deletion present in the human, chimpanzee, orangutan and macaque exon X pseudogene merely reflects a mutational hot spot. We are disturbed that for about 20 years authors, reviewers, publishers and readers of the GULO pseudogene literature were convinced the current neo-Darwinian theory had been established beyond question. An erroneous assumption, obvious to anyone who had enquired, has only now come to light.

When examined in detail, the full pseudogene dataset we collected does not lend itself to a reasonable neo-Darwinian interpretation. Using standard bioinformatics tools and principles, we present alternative designs for at least the exon X portion of the GULO gene. These may be plausible due to nucleotide patterns being relevant as regulatory signals or the favouring of some codons for various possible reasons. We do accept that some mutations have occurred in this exon. But these novel proposals imply that the ancestors of the organisms studied may well never have had the exact same GULO sequence.

The reasons why most Hominidae display a deletion at position 97 are not clear, but we argue that this fact should not be overrated. This position shows the characteristics of being a mutational hotspot, and during a period of high mutations and low populations many statistical coincidences can be generated.

Acknowledgments

Professor Nishikimi generously provided us with reprints of his papers which were not available to us. We hope he will collaborate with us in further researching this interesting pseudogene. Many thanks to the staff at CMI and the unknown reviewers for their helpful suggestions.

References

1. Ohta, Y. and Nishikimi, M., Random nucleotide substitutions in primate non-functional gene for L-gulono-gamma-lactone oxidase, the missing enzyme in L-ascorbic acid biosynthesis, *Biochim Biophys Acta* **1472**:408–411, 1999.
2. Nishikimi, M., Kawai, T. and Yagi, K., Guinea pigs possess a highly mutated gene for L-gulono-gamma-lactone oxidase, the key enzyme for l-ascorbic acid biosynthesis missing in this species, *The Journal of Biological Chemistry* **267**(30):21967–21972, 1992.
3. Nishikimi M., *et. al.*, Cloning and Chromosomal Mapping of the Human Non-functional Gene for L-gulono-gamma-lactone oxidase, the enzyme for l-ascorbic acid biosynthesis missing in man, *The Journal of Biological Chemistry* **269**(18):13685–13688, 1994.
4. Inai, Y., Ohta, Y. and Nishikimi, M., The whole structure of the human non-functional L-gulono-gamma-lactone oxidase gene—the gene responsible for scurvy—and the evolution of repetitive sequences thereon *J. Nutr. Sci. Vitaminol. (Tokyo)* **49**:315–319, 2003.
5. Ohta and Nishikimi, ref. 1, p. 410.
6. The intact GULO gene is believed to consist of 12 exons (Nishikimi *et al.*, ref. 3. Exon X refers to the tenth one).
7. <www.ncbi.nlm.nih.gov/BLAST/>
8. <bips.u-strasbg.fr/fr/Documentation/ClustalX/>
9. <evolution.genetics.washington.edu/phylip.html>
10. Nucleotide positions: 1, 22, 31, 58, 79, 100, 109, 157
11. Nucleotide positions: 22, 58, 79, 109
12. Li, W.-H., Hide, W.A., Zharkikh, A., Ma, D.-P. and Graur, D., The molecular taxonomy and evolution of the guinea pig, *The Journal of*

Heredity **83**(3):174–181, 1992.

13. Nishikimi *et al.*, ref. 2 p. 21972.
14. Li M., Badger, J.H., Chen, X., Kwong, S., Kearney, P. and Zhang, H., An information-based sequence distance and its application to whole mitochondrial genome phylogeny, *Bioinformatics*, **17**(2):149–154, 2001.
15. Prion molecular phylogeny, 30 May 1998, <www.mad-cow.org/moly_phylo.html>.
16. Nishikimi *et al.*, ref. 2 p. 21971.
17. Marignetti, J.A and Brosius, J., Neural BC1 RNA as an evolutionary marker: Guinea pig remains a rodent, *Proc. Natl. Acad. Sci. USA* **90**:9698–9702, 1993.
18. Robinson-Rechavi, M., Ponger, L. and Mouchiroud, D. Nuclear gene LCAT supports rodent monoophyly, *Mol. Biol. Evol*. **17**(9):1410–1412, 2000.
19. Cao, Y, Adachi, J., Taka-aki, Y. and Masami, H. Phylogenetic place of guinea pigs: no support of the rodent-polyphyly hypothesis from maximum-likelikood analyses of multiple protein sequences, *Mol. Biol. Evol.* **11**(4):593–604, 1994.
20. Nishikimi, M., Okamura, M. and Ohta, Y., Recent advances in biochemistry of L-ascorbic acid biosynthesis, *Recent Res. Biophys. Biochem*. **3**:531–545, 2003.
21. Rat GULO pseudogene sequences reported in GenBank: gi|60683826, gi|60097942, gi|204149
22. Mouse GULO pseudogene sequences reported in GenBank: gi|94397619, gi|18044344, gi|20379919, gi|20380022, gi|38325769, gi|22028124, gi|74195452, gi|31343390
23. *Homo sapiens* GenBank accession numbers: gi|32490605, gi|493656
24. Pig GenBank accession numbers: gi|24637282, gi|7381187
25. Cow GenBank accession numbers: gi|77404230, gi|5924388
26. For example see Li *et al.*, ref. 14, p. 180. Depending on phylogenetic assumptions, the lipocoring protein of guinea pig and pigeon share eight or at least 15 amino acid parallel substitutions. Facts such as these, if uniquely shared between humans and chimpanzees would be immediately published as proof for neo-Darwinian theory. The inconsistent examples draw little attention.
27. Rokas, A. and Carroll, S.B., Bushes in the Trees of Life, *PloS Biology*, **4**(11):1899–1904, 2006; <biology.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pbio.0040352>.
28. Luria S.E. and Delbrück, M., Mutations of bacteria from virus sensitive to virus resistance, *Genetics* **28**:491–511, 1943.
29. Pauling, L., *Vitamin C and the Common Cold*, Bantam Books, Inc., New York, 1971.
30. Drake, J.W., Charlesworth, B., Charlesworth, D. and Crow, J.F., Rates of spontaneous mutation, *Genetics* **148**:1667–1686, 1998.
31. Positions which could be explained as hotspot mutations include: 13, 38, 50, 55, 59, 64, 81, 85, 94, 96, 103 in table 2.
32. See positions 2, 12, 13, 38, 50, 55, 56, 59, 64, 65, 75, 81, 94, 95, 96, 99, 103, 112,132, 140, 146, 158 in table 2.
33. Truman, R., The ubiquitin protein: chance or design? *Journal of Creation* **19**(3):116–127, 2005.
34. Annex 4—Amino acids, one and three letter codes, <www.fao.org/DOCREP/004/Y2775E/y2775e0e.htm>.
35. Genesis 1:28; 9:1.
36. (a) Vardiman, L., Snelling, A.A. and Chaffin, E.D., *Radioisotopes and the Age of the Earth*, Institute for Creation Research and Creation Research Society, 2000. (b) Vardiman, L., Radioisotopes and the Age of the Earth: <www.icr.org/pdf/research/RATE_ICC_Vardiman.pdf>
37. Arnason, U. *et. al*., Mammalian mitogenomic relationships and the root of the eutherian tree, *PNAS* **99**(12):8141–8156, 2002; p. 8155: 'It has been argued that the molecular estimates suggesting early origin of eutherian orders are artifactual and caused by accelerated molecular evolution at the Cretaceous/Tertiary (K/T) boundary or immediately thereafter.'
38. Sanford, J.C., *Genetic Entropy & The Mystery of the Genome*, Ivan Press, Lima, New York, 2005.
39. Rokas and Carroll, ref. 27, p. 1901.
40. Woodmorappe, J., Are pseudogenes 'shared mistakes' between primate genomes? *Journal of Creation* **14**(3):55–71, 2000.
41. <creation.com/GULO_Appendix>

Royal Truman has bachelor's degrees in chemistry and in computer science from SUNY Buffalo, an M.B.A from the University of Michigan, a Ph.D. in organic chemistry from Michigan State University and post-graduate studies in bioinformatics from the universities of Heidelberg and Mannheim. He works for a large multinational in Europe.

Peer Terborg has an M.Sc. in Biology (Hons biochemistry and molecular genetics) and a Ph.D. in Medical Sciences from the University of Groningen, The Netherlands. He is currently working on the cellular and molecular aspects of pulmonary diseases, such as asthma and COPD, and is an expert on the molecular biology of signal transduction and gene expression.

Why the shared mutations in the Hominidae exon X GULO pseudogene are not evidence for common descent

Royal Truman and Peer Terborg

Appendix: The nature of bioinformatic evidence

Suppose an evolutionist has been informed that the nt *G* (Guanine) is found in all the organisms of a dataset for a particular gene, except for *Hu* (human) and *Ch* (chimpanzee), for which nt *A* (Adenine) is found at that position. Since he 'knows' the present evolutionary phylogeny for Hominidae is true, he shows figure 4 point P3 to any sceptic and presents this nucleotide information as persuasive evidence for the current model. To illustrate the trap, the reader has just been deliberately misled! From position 96 of table 2 we see that the *A* nt is unique to only *Or* (orangutan) and *Ma* (macaque), and not *Hu* and *Ch*, which makes no evolutionary sense. What seemed quite convincing a second ago, now requires a statistically improbable scenario: both neutral mutations occurred independently by chance, for only these two organisms, and this then spread throughout both populations.

Some time later you are informed that *Hu* and *Ch* uniquely share an nt *A* in a dataset, (not the same nt position as mentioned above) contrary to all other organisms which display the nt *G*. How convincing is this evidence now for a *Hu*/*Ch* common ancestor? Obviously far less, given your preceding experience. Your caution would be justified. In fact, we have deliberately misled the reader for a second time! The observation refers now to position 132 (table 2) and the imputed common ancestor is once again *Or* and *Ma* (and not for *Hu* and *Ch* as we just pretended).

There are two positions (75 and 95 see table 2) in which indeed *Hu* and *Ch* only show the same nt in the dataset.

Our evolutionist presents us next with a statistically greater challenge. We are told that *Hu*, *Ch* and *Or* all share the nt *T* (Thymine) and all the other organisms the nt *C* (Cytosine). Arguing for two such coincidences might be difficult, but for three such coincidences at exactly the same position and in accord with evolutionary thinking you are going to have difficulties. This evidence matches evolutionary theory well (figure 4 point P2), and is essentially compelling, right? Well, not really. We have chosen to mislead the reader for a third time to drive the point home. We are referring to position 55 (table 2) and the nt *T* in common refers to *Or*, *Hu* and *Ma*. Although the *Ch* is supposed to share a common ancestor with *Hu* after the *Or* line branched off, the expected nt is not found for *Ch*. The only reasonable evolutionary answer, is that precisely at that position a back-mutation occurred to the original nt *C*. But examination of table 2 implies very few mutations have occurred at all, and such coincidences demand arguing against the facts. This aspect of coincidences will be discussed further below. Incidentally, the pattern at position 140 (table 2) could indeed be interpreted in a manner the evolutionist would like: here *Or*, *Hu* and *Ch* share the nt *C*, whereas all the other organisms the nt *T*.

We see that given enough data we can easily select whatever data suits our purposes and ignore or downplay the rest.

Why are intelligent researchers being so easily misled to see evolutionary evidence in patterns of nucleotide or protein sequences? There are three principles which we hope to explain in greater detail in a future paper.

It is, hardly surprising that organisms with a *similar* Bauplan and environment will indeed share many designed genetic features. This is intuitively anticipated by those believing in design. It would be unreasonable to expect elephants and *E. coli* to possess highly similar genomes. After all, we do expect genetic information to have visible morphological outcomes! Organisms in very different taxa will on average show significantly different gene sequences. By the mathematical nature of how evolutionary trees are algorithmically programmed, in which the more similar sequences are assumed to have branched off from a common ancestor, it is inevitable that apparently reasonable evolutionary trees at this very rough degree of detail will often result.

Evolutionists and creation scientists agree that there was indeed a common ancestor for dogs, for bears, for ducks, etc. Sequence analysis at this micro-level can reveal in principle true phylogenetic relationships within the original created biblical 'kinds'. (The detailed scenarios and models do differ, however. The post-Flood environments with low population sizes would permit a large number of mutations to fix quickly, whereas evolutionists believe new information arose through a long process of random mutations plus natural selection. Most creation researchers believe organisms were endowed *ab initio* with genetic possibilities which were later expressed and that genetic information did not arise by chance.)

The evolutionary framework possesses a vast number of candidate phylogenetic markers and adjustable parameters. There are virtually no real, *a priori* predictions, uniquely limited to the evolutionists, as to what genetic data to expect.

We hope to offer a detailed analysis of point (iii) in the future. Researchers overrate the strength of evidence which

seemingly supports their theory if they can immediately map data presented to an interpretation they are very comfortable with. If two or more nts in our dataset match up in a manner consistent with a reasonable common ancestor, this explanation pops immediately into the evolutionist's mind. Both of us have spent over a decade being trained in evolution-dominated secular universities. We both can immediately offer multiple evolutionary possibilities to most data presented to us. We can also quickly offer the best evolutionary 'excuse' when the data does not meet theoretical expectations. It is our hope to show our evolutionist friends that what seems apparent is a mirage.

Note that evolutionary theory has not stated in advance which mutations in common would be expected to arise from which common ancestor. Intuitively, when wearing evolutionary glasses, we *accommodate* the data *post facto* into the theoretical framework. Therefore, if some organisms share a suitable pattern, and this is presented in a manner where the evolutionary explanation is immediately apparent, then too much significance is assigned to the finding. Particularly guilty are phylogenetic bifurcation trees (see figure 1), in which a common ancestor is directly claimed. Other data clustering methods merely indicate closer resemblance in a more neutral way, such as our figure 2 and figure 3, although almost all modern bioinformatic alignment algorithms are based on and calibrated on evolutionary assumptions.

The potential for coincidence is vast. Suppose our data implied a common ancestor for three out of four organisms in a dataset (or 4 out of six, or 5 out of eight ...). The interpretation is then 'obvious': the shared-derived character was 'obviously' present on a common ancestor, which some lineages subsequently lost.

We can formalize this observation using decision theory. We define two statements of opinion, S1 and S2, and information fact I. Here S1 and S2 are mutually exclusive, and $p(S1) + p(S2) = 1$.

S1 : 'Evolutionary theory is the true explanation'
S2 : 'Evolutionary theory is not the true explanation'
I: 'A sequence pattern is found predicted by evolutionary theory'

Using Bayes's Rule,

$$P(S1 \mid I) = P(I \mid S1)/P(I) \times P(S1) \quad (1)$$

where $P(S1 \mid I)$ means, 'the probability we assign to statement S1 given that we have been informed about fact I'.

Given that information *I* was in fact found, the *posterior probability* $P(S1 \mid I)$, of belief statement S1, is given by the right hand side of (1). P(S1) is the *prior probability* before such data became available.

$P(I \mid S1)/P(I)$ has the potential to modify a prior belief, and cannot be less than one. Now, neo-Darwinian theory has been in a state of parameter fine-tuning for over half a century. Sequence alignment weighting matrices have been optimally calibrated[1] to provide evolutionary theory the highest consistency possible. This means that the desired date of lineage divergences, according to current theory, are typically used to calculate probabilities of conversion from one nt or amino acid into another in for example PAM matrices.[2] Frequency of events such as gene duplication and mutations are also calibrated by evolutionary assumptions. When the results don't agree well,[3] the assumed evolutionary dates can be modified.[4] Discordant genes or parts of their sequences are simply stated as providing the wrong signal.[5] All these parameter fine-tunings[6] lead to a modified model which did not result from fundamental evolutionary assumptions. Fundamental theory did not predict creation of a GULO pseudogene for guinea pig and primate lineages 20 Ma ago,[7] nor was this based on any fossil or morphological data. In fact, the morphological basis for classifying the guinea pig in the order Rodentia was proposed[4] to be irrelevant only after so many gene sequence abnormalities were discovered. Dates, parameters and interpretation are constantly readjusted to optimise internal consistency, almost totally devoid of objective constraints.

These observations imply that we may find model-optimised examples in which $P(I) > 0$ ('A sequence pattern is found predicted by evolutionary theory'). This is hardly surprising, given the rich variety of parameters available to make evolutionary scenarios fit.[8] But are these probabilities truly lower than $P(I \mid S1)$, meaning probabilities of being correct only if evolutionary theory is in fact true? Note that in the examples in which we misled the reader we cannot distinguish between $P(I \mid S1)$ and P(I). Does evolutionary theory really predict the same nt for only *Hu* and *Ch*? Sometimes we find this result. This is in accord with evolutionary theory and thus reinforces the belief the theory is true. Sometimes we don't find this result. *'"So what" thinks the evolutionist. The theory never predicted this pattern anyway.'*

The creation science theoretician also has many degrees of freedom available to create scenarios which fit the available data. Different categories of gene sequences may have been created initially. Furthermore, shortly after the Flood many species were present in very low numbers. Based on radioactivity studies[9] it is possible that mutation rates may have been very high in the past.[10] The latter two factors suggest that numerous mutations may have occurred and fixed almost immediately in the entire populations very rapidly in the past. These models also have much freedom in guessing when various speciation events may have taken place. After a large number of converging interactions (i.e. model tinkering) a fine-tuned scenario would also lead to predictions of $P(I) > 0$, where I now mean 'A sequence pattern is found predicted by creation theory'. But once again, is it truly so, that $P(I \mid S1) > P(I)$?

Statistically founded guesses in the absence of any theory will generally be far better than random guesses. One need have no opinion about the origin of life to develop strictly empirical and useful statistical models. One can

collect any cellular feature one wishes, correlate with other features, cluster as one finds appropriate, and thereby permit better predictions for an unstudied organism. *I* then becomes: 'A sequence pattern is found predicted by a statistical model'. We certainly now expect $P(I) > 0$. But is $P(I \mid S1) > P(I)$ truly due to whatever story we invent to embellish the trends extracted from empirical models? Is the story of any real value, if the predictions we make simply rely on statistical observations, properly expressed mathematically?

It is our opinion that statistical analysis can indeed be fruitful in identifying patterns which provide intuition for additional research. But to fathom the true meaning behind coding and non-coding DNA sequence patterns a much deeper understanding is needed into all the kinds of coded signals[11] and Design goals needed by various cells. Superimposed are randomising mutations ('noise') which may camouflage the original intent, and these must also be studied before sequence data is to be understood.

References

1. Read revealing statements in the Materials and Methods sections of papers in which multiple parameters have been used, calibrated in a manner to optimize a desired evolutionary outcome. Here is an example: 'The mtREV-24 model of amino acid sequence evolution and the TN-93 model of nucleotide evolution were used for distance and likelihood analyses.' 'The analyses were performed under the assumptions of ... plus four classes of variable sites'. From Arnason, U. *et. al.*, Mammalian mitogenomic relationships and the root of the eutherian tree, *PNAS* **99**(12):8141–8156, 2002.
2. Dayhoff, M.O., Eck, R.V. and Park, C.M., Atlas of protein sequence and structure, NBRF, Vol. 5, p. 75–84, 1972.
3. Arnason, U. *et. al.*, ref. 1, p. 8154: 'In comparison, the nucleotide trees differed considerably depending on the analytical approach used.'
4. Li, W.-H., Hide, W.A., Zharkikh, A., Ma, D.-P. and Graur, D., The molecular taxonomy and evolution of the guinea pig, *The Journal of Heredity* **83**(3):174–181, 1992.
5. Arnason, U. *et. al.*, ref. 1, p. 8155: 'Only the stem regions of the mt rRNA genes seem to carry a useful (albeit marginal) phylogenetic signal and inclusion of the other parts of these sequences may increase the background noise and promote the selection of the wrong signal.'
6. Arnason, U. *et. al.*, ref. 1, p. 8155: 'Finally, the pinniped results lend no support to the morphological hypothesis...underlining the problems associated with basing phylogenetic conclusions on anatomical features that may have strong adaptive values.'
7. Nishikimi, M., Kawai, T. and Yagi, K., Guinea pigs possess a highly mutated gene for L-gulono-gamma-lactone oxidase, the key enzyme for l-ascorbic acid biosynthesis missing in this species, *The Journal of Biological Chemistry* **267**(30):21967–21972, 1992.
8. For example, a large proportion of gene sequences were less similar between humans and chimpanzees than between humans and gorilla in some cases. A paper was published in which it was claimed that millions of years after a split and genetic isolation from the gorilla lineage cross-breeding re-occurred. This bizarre theory prompted headlines in Germany like 'Scientists show early humans had sex with monkeys'.

   This *ad hoc* rationalization is found in: Patterson, N., Genetic evidence for complex speciation of humans and chimpanzees, *Nature* **441**:1–6, 2006.
9. (a) Vardiman, L., Snelling, A.A. and Chaffin, E.D., *Radioisotopes and the Age of the Earth*, Institute for Creation Research and Creation Research Society, 2000. (b) Vardiman, L., Radioisotopes and the Age of the Earth: <www.icr.org/pdf/research/RATE_ICC_Vardiman.pdf>
10. Arnason, U. *et. al.* ref. 1, p. 8155: 'It has been argued that the molecular estimates suggesting early origin of eutherian orders are artifactual and caused by accelerated molecular evolution at the Cretaceous/Tertiary (K/T) boundary or immediately thereafter.'
11. Trifonov, E.N., Genetic sequences as product of compression by inclusive superposition of many codes, *Molecular Biology* **31**(4):647–654, 1997.

Royal Truman has bachelor's degrees in chemistry and in computer science from SUNY Buffalo, an M.B.A from the University of Michigan, a Ph.D. in organic chemistry from Michigan State University and post-graduate studies in bioinformatics from the universities of Heidelberg and Mannheim. He works for a large multinational in Europe.

Peer Terborg has an M.Sc. in Biology (Hons. biochemistry and molecular genetics) and a Ph.D. in Medical Sciences from the University of Groningen, The Netherlands. He is currently working on the cellular and molecular aspects of pulmonary diseases, such as asthma and COPD, and is an expert on the molecular biology of signal transduction and gene expression.