Hierarchical clustering complicates baraminological analysis

Matthew Cserhati and Robert Carter

Baraminology is the study of the created kinds in Genesis 1. Its main goal is to determine the boundaries between each kind and the species membership within kinds. The science of baraminology has been developing for several decades. Besides morphology-based methods, multiple molecular baraminology studies have been performed using DNA and protein sequences. All these studies attempt to use some statistical measurement parameter to determine which species are close to one another, and thus may belong to the same kind, and which ones are not. There have been difficulties in determining a universal cut-off parameter that can be easily used to separate species into different kinds. One major issue is the hierarchical structure of species relationships. This makes it extremely difficult to know where to draw the line between kinds and prevents a simple statistical determinant. This paper examines this problem in the analysis of 28 mammalian species. Several possible ways forward in determining kinds are discussed.

Baraminology is the discipline of biblical taxonomy. Its main goal is to determine the species membership of the created kinds, called baramins, mentioned in Genesis 1, and to determine the boundaries between them. Species within the same kind should show continuity with one another on a morphological and a molecular level. They should also show discontinuity with all other species outside their kind. Over 10 years ago, Peer Terborg proposed a method to just do so based on genetic fingerprints in essential (i.e. non-redundant) genes. A handful, sometimes only one, of essential genes might suffice to tell one baramin from another. Terborg coined them "indicator genes".[1] Other researchers have called them "taxonomically restricted" or "orphan" genes.[2] Indeed, modern humans, Neandertals, and Denisovans all share the same form of sialic acid in the sugar coating on their cells (N-acetylneuraminic acid, or Neu5Ac). All apes and most other mammals have a different sialic acid (N-glycolyneuraminic acid, or Neu5Gc). This is a profound difference that would, for example, prevent egg-sperm recognition, creating instant reproductive incompatibility.[3] This can be added to the list of other unique genes that separate humans (including Neandertals and Denisovans) from apes, like the FOXP2[4] gene and the structure of chromosome 2.[5] Yet, even though discontinuity can be found, the difficulty is in *quantifying* that discontinuity.

Cut-off values used in baraminology studies

Baraminology studies have traditionally attempted to use a statistical measurement with which they can separate species into their corresponding kinds. Morphology-based baraminology methods include the baraminic distance (BDIST) measurement.[6] Molecular baraminology methods include sequence alignment and identity matrix analysis, the GCM (Gene Content Method),[7] and the WGKS (Whole Genome K-mer Signature) algorithm.[8]

BDIST is a widely used morphology-based baraminology method which measures a set of characters in a group of species using a data matrix.[9] It measures 'baraminic' distances between species based on the proportion of mismatched characters over all characters. The method determines relevant characters which are present in a minimum proportion of the studied species. This relevance cut-off has been arbitrarily set between 75 to 95% in various morphological baraminology studies. BDIST also uses bootstrapping to determine which correlations between taxa are robust. The minimum bootstrap value of 90% is also arbitrary.

In molecular baraminology studies, genetic distance/similarity can be determined in several ways. One can calculate overall percentage-wise genome similarity. This can be done for shorter sequences where aligning sequences, such as mitochondrial, plastid, or bacterial genomes is relatively straightforward. Several earlier baraminology studies have focused on this.[10,11] Another method used a popular DNA searching algorithm (BLAST) to probe similarities between chimpanzee and human genomes. This revealed much greater distances between the two than earlier studies had shown,[12] further putting to rest the 'myth of 1%'.[13]

Another way is by calculating the proportion of the overlap of orthologous protein content and the total orthologous protein content between two species (orthologs are genes and their corresponding proteins that can be matched between species). This is described using the Jaccard Coefficient Value (JCV). The JCV is the defining step of the GCM, which has been used to study several groups of organisms, from Nucleocytoplasmic Large DNA Viruses (NCLDV),[14] Archaea,[15] fungi, and mammals.[16]

A third way to measure genetic similarity is to calculate the Pearson Correlation Coefficient (PCC) of the k-mer content between two species. This has been implemented in the WGKS, which has been run on algae,[17] insects,[18] and several groups of mammals, including bears, mustelids (minks and weasels), felids (cats), procyonids (racoons), and mephitids (skunks).[19]

Complicating factors

A common problem in baraminology studies is determining the cut-off value, above which two species belong to the same baramin, and below which they belong to separate baramins. Both morphological and molecular baraminology studies have failed, as of yet, to define a precise cut-off value which can be of general use for determining baraminic membership. If one chooses a cut-off score that can delimit created kinds among one group of species, this does not automatically mean the same cut-off score can be used to separate kinds using a different set of species.

This may be due to several factors. One is the type of species being studied. Prokaryotes and eukaryotes have different gene structures. Bacteria contain many passenger genes besides the core genes found in their genomes, which may lower JCV, since passenger genes are usually not shared between species. Also, genes may be pleiotropic in their effects, so a relatively small number of genes may cause large phenotypic differences (i.e. large cognitive differences between chimps and humans). This would raise the JCV cut-off in the study of such species, meaning that two genetically similar species might belong to 'obviously' separate kinds. These considerations make the conclusions of all baraminology studies purely relative, dependent on intuition more than numbers.

Complicating the picture for both molecular and morphological baraminology studies is that God could have created several baramins which show some genetic similarity, but which are different overall morphologically. For example, an evolutionary study, based on the insertion of L1 retrotransposons alleges that Chiroptera (bats) and Perissodactyla (odd-toed ungulates, such as horses) are closer to one another than horses are to cows.[20] Yet another example is the overall relative genetic similarity between humans and the great apes. It is definitely possible that God could have created different kinds of organisms this way. We simply do not know what God had in mind when He created the living world.

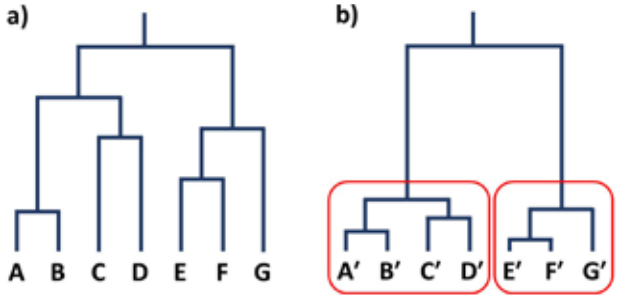

Figure 1. a) Phylogenetic tree in which baramins are not clearly distinguishable. b) Phylogenetic tree in which species can be clearly divided into two putative clusters.

When analyzing fossil data, researchers should strive to find the most complete data set possible. Incomplete fossil data sets make baraminic predictions very tentative. With more complete data, it is possible that these predictions may even change. For example, when first analyzing only craniodental characters in *Australopithecus sediba*, Wood classified this species as a member of the human holobaramin.[21] In a later analysis, including post-cranial data, he reversed his prediction, classifying *A. sediba* as an australopith.[22] One of the current authors initially placed *Homo naledi* in the human holobaramin, but then reassigned it as an australopith in a later study after including post-cranial features.[23,24] This highlights the difficulties inherent in the field. There is nothing wrong with these statistical analyses, but we need to always remember that the results are always tentative.

Gene-level similarities also occur. One example is the high sequence similarity between the oxygen-carrying protein hemoglobin in the blood of animals and the leghemoglobin protein, which fixes nitrogen in the root nodules of leguminous plants.[25] Another example is the *prestin* gene, which groups the bottlenose dolphin together with microbats.[26] Since genes with the same function and high sequential similarity are found in very different organisms, these genes can be viewed as functional design elements. But they complicate the baraminological landscape since they give the impression that very different baramins are actually similar to one another.

Another possibility is that, after the Fall, boundaries between kinds could have broken down. This does not seem to be true at this stage, but we cannot arbitrarily discount the possibility from first principles. Contrarily, genetic boundaries may have arisen which now block two species from breeding with one another, despite them belonging to the same original created kind at creation. Such mutations could involve chromosome incompatibilities, which make the hybrid offspring sterile, as in the (usual) case of the mule (63 chromosomes), which is the offspring of the horse (64 chromosomes) and the donkey (62 chromosomes). Such could also be the case with crypto-species, such as in several corals within the genus *Orbicella*, where incomplete gametic incompatibility exists, despite their belonging to the same genus, and hence most likely to the same created kind.[27,28]

Yet another thing to consider is that God could also have created multiple kinds, which seemingly belong to the same group, but are still separate from one another. For example,

God could have created multiple turtle kinds,[29] seal kinds,[30] snake kinds,[31] or squid kinds.[32]

Finally, it is quite possible that massive gene loss, duplication, rearrangement, or genetic mutation and/or scrambling could create situations where statistics are unable to correctly identify baraminological relationships. This could easily be a contributing factor to why we have struggled to come up with an objective measure of intra-baraminic differences.

Hierarchical relationships

Yet the greatest problem facing baraminology might well be the hierarchical structure of life. Even though God created organisms separate from one another, different kinds can still be placed into larger and larger groups, as in a hierarchy. For example, humans are primates, primates are mammals, mammals are tetrapods, tetrapods are vertebrates, vertebrates are deuterostomes, and deuterostomes are eukaryotes. As we ascend the hierarchy, organisms are classified less and less specifically.

Another example is the vertebrate forelimb, which follows the same basic plan in all classes of vertebrates. The bones in the forelimb have a homologue in nearly all vertebrate classes with forelimbs. This has long been cited as evidence of common descent. Departing from the traditional evolutionary view, Wagner has recently argued that homology is an artefact of the evolution of gene regulatory networks.[33] ReMine argues that common traits within a hierarchy exhibit unity because of a common Designer and not common descent.[34] Different created kinds differ from each other in the specific implementation of common design elements, such as differences in the order of limb bone development, origin of soft tissues, and differences in DNA codes.[35] Either way, similarity of design makes it harder to separate the kinds into distinct groups.

This simply means that cladistics can assemble any group of organisms into a hierarchy, regardless of whether they are truly related or not. And baraminology could be considered as just a mathematical expression of cladistics.

The biggest problem with hierarchical clustering is in deciding where to cut a cladistic/phylogenetic tree. In figure 1 we can see two extremes. In figure 1a, there is no clear clustering between species A–G. The tree looks like one big cluster. Species distances are uniformly distributed between 0 and 1. A high distance value would have to be used to get any kind of meaningful clustering. This is not optimal for baraminology studies.

In figure 1b, there is clear clustering. Here only a low distance value (corresponding to a high similarity value) is enough to distinguish between the two clusters. This kind of situation is optimal for baraminology studies. Yet, there was nothing preventing God from creating any particular pattern of similarity. Baraminology was designed to elucidate the real patterns, but it had no way of making absolute judgments about the boundaries between kinds. This became more obvious as variable levels of difference were discovered between groups of species belonging to what we thought were obviously distinct kinds.

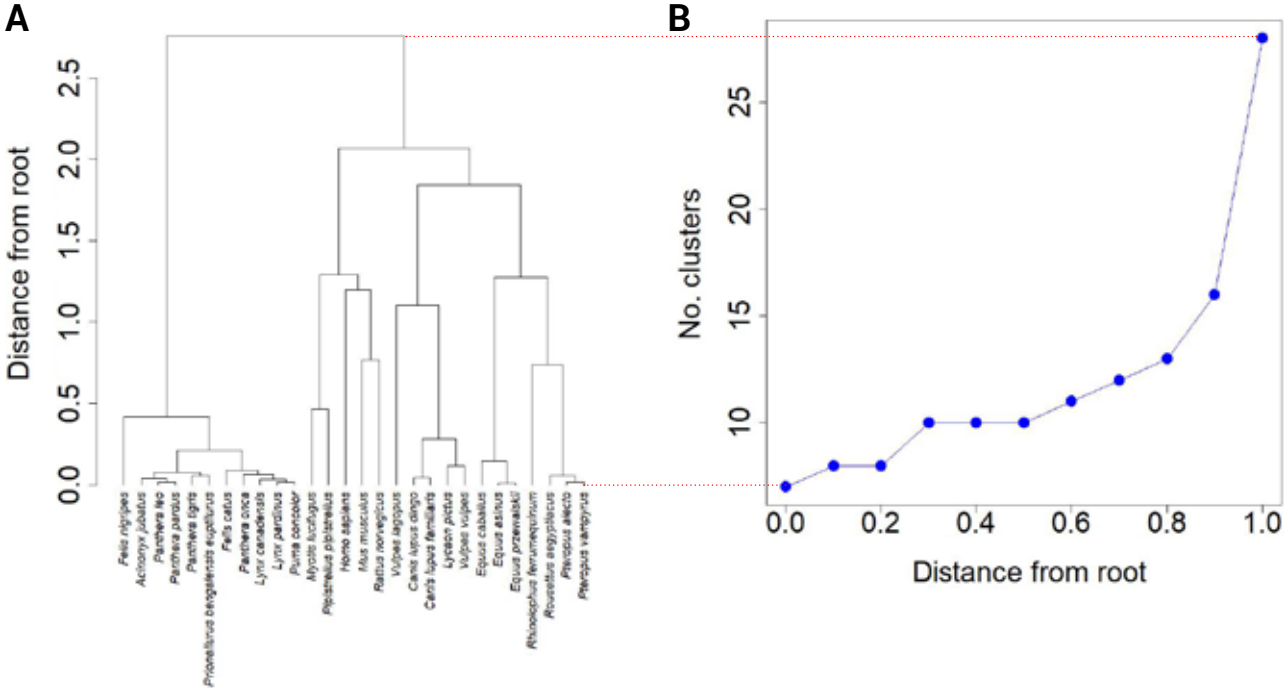

Figure 2. A) Phylogenetic tree showing relationships between the 28 species in this study. B) The number of clusters varies depending on where a cut-off ('imaginary horizontal line') is placed at a certain distance from the root (top) of the tree shown in part A.

Figure 2 shows a cladogram connecting 28 species using WGKS. Where should one draw a horizontal line to determine the boundaries between baramins? If this line is set at the top of the cladogram and slowly moved downward, the number of clusters steadily increases. This increase is shown on the right, in figure 2B. The number of clusters rises exponentially as we transition from (assumed) created kinds to post-Flood species. The inflection point might denote the optimal number of clusters, but this would be a matter of convention or convenience only, since we still don't know where the breaks between kinds are.

Hierarchical clustering problems also present themselves when selecting species for baraminology studies. Species must be chosen at the right taxonomic level. If species are chosen too broadly, different clustering algorithms could put different species together which don't belong to the same baramin.

Conversely, if species are chosen from one single baramin, the algorithms will still find clusters. This was demonstrated in a study with 25 species of dogs using BDIST. It found different groups of dogs based merely on their general size.[36] Another study based on genome-wide pairwise divergences and mitochondrial DNA analysis involving two species of ancient cave lions (*Panthera leo spelaea*), 12 historic lions (*Panthera leo melanochaita*), and six modern-day lions (*Panthera leo*) showed three separate lineages. The first lineage comprising ancient lions in Siberia and the Yukon, a second lineage in western and northern Africa and south-west Asia, and a third in central and southern Africa.[37] Nevertheless, these three lineages correspond to a monobaramin. The cat kind is known to encompass 38 species based on evidence from hybridization.[38] But if there are groups within groups, clearly determining the cut-off score is the single most critical factor in determining the boundaries between the created kinds.

The species relationships are very much like a fractal. The branching pattern on the tree of life appears similar at multiple scales. Thus, the promise of baraminology has not yet manifested itself. We do not yet have a way to make statistical determinations of group membership.

Table 1. Classification of the 28 species included in the mtDNA and WGKS analyses

Species	True cluster	mtDNA analysis	WGKS kmeans
Acinonyx jubatus	cat	*cat 2*	cat
Canis lupus dingo	dog	dog 1	dog
Canis lupus familiaris	dog	dog 1	dog
Equus asinus	horse	horse	horse
Equus caballus	horse	horse	horse
Equus przewalskii	horse	horse	horse
Felis catus	cat	cat 1	cat
Felis nigripes	cat	cat 1	cat
Homo sapiens	human	human	human
Lycaon pictus	dog	dog 1	dog
Lynx canadensis	cat	cat 1	cat
Lynx pardinus	cat	cat 1	cat
Mus musculus	mouse	mouse	mouse
Myotis lucifugus	microbat 1	*mouse*	microbat 1
Panthera leo	cat	*cat 2*	cat
Panthera once	cat	*cat 2*	cat
Panthera pardus	cat	*cat 2*	cat
Panthera tigris	cat	*cat 2*	cat
Pipistrellus pipistrellus	microbat 1	-	microbat 1
Prionailurus bengalensis	cat	cat 1	cat
Pteropus alecto	megabat	megabat	megabat
Pteropus vampyrus	megabat	megabat	megabat
Puma concolor	cat	cat 1	cat
Rattus norvegicus	mouse	mouse	mouse
Rhinolophus ferrumequinum	microbat 2	megabat	microbat 2
Rousettus aegyptiacus	megabat	megabat	megabat
Vulpes lagopus	dog	dog 2	dog 2
Vulpes vulpes	dog	dog 2	dog

Materials and Methods

Species from generally accepted created 'kinds' (which have been well studied, such as cats and dogs) were selected to determine whether they

really are members of these baramins. We used two methods to measure species similarity and dissimilarity, with a special focus on molecular baraminology methods. First, the *Kalign alignment algorithm* was applied to a set of mitochondrial sequences belonging to 27 of the selected species. The alignments of the mitochondrial DNA are needed to

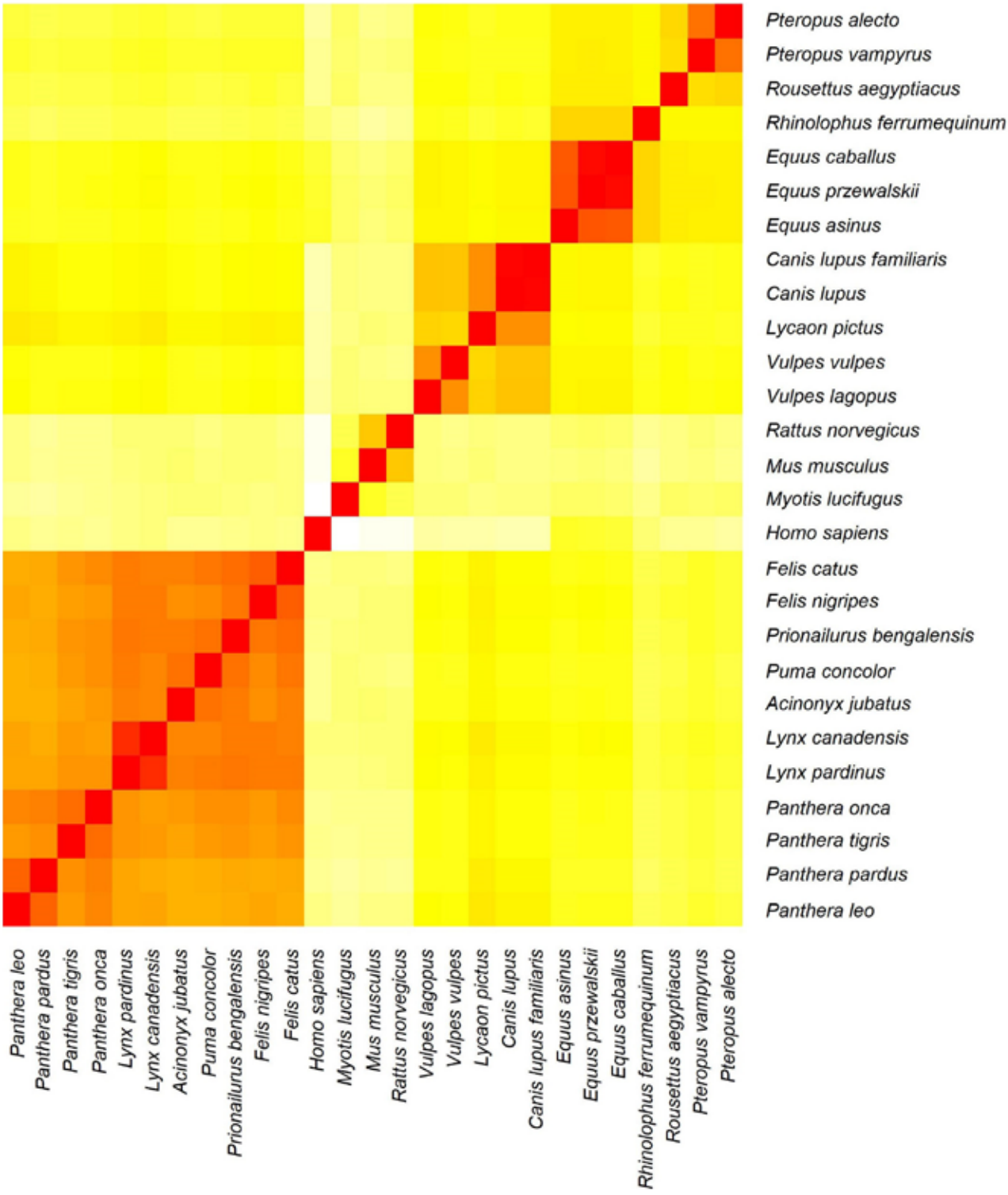

Figure 3. Heatmap of 27 species based on mtDNA sequence similarity. Higher species similarities between a given species pair correspond to redder colours. Lower similarity values correspond to yellow colours.

measure pairwise sequence similarity between all species. Second, WGKS was used to analyze the whole genomes of these baramins. These include the human, cat, dog, mouse, horse, and several putative bat kinds. The list of species, their genome sequence, and mitochondrial genomes are listed in Supplementary file part 1.

Mitochondrial genomes and whole genome sequences were downloaded from NCBI at ncbi.nlm.nih.gov/genome. The *Kalign2* software program was used to align mitochondrial genomes on the EBI website at ebi.ac.uk/Tools/msa/kalign.[39] Pairwise whole genome sequence similarity was calculated and visualized in a heatmap in R, version 3.6.0. using the heatmap command under the ward.D2 clustering method. Previously established protocol was followed for *WGKS*, and pairwise Pearson Correlation Coefficient (PCC) values were visualized in a similar way.[8] Supplementary Files for this study are available at creation.com/hierarchical-clustering-baraminology-analysis. The *kmeans* clustering algorithm was also applied, using the '*kmeans*' command to the PCC matrix with eight predicted clusters.

Results

Analysis of mitochondrial genomes

The identity matrix, which represents the percentage similarity between species, for the mitochondrial alignment is depicted as a heatmap in figure 3. The pairwise genome similarity values are available in Supplementary file part 2. The Hopkins clustering measurement value is 0.775, which means that the sequence similarity matrix has a reasonably good clustering quality. Seven groups are visible, these are the cat, dog, human, mouse, horse, and putative microbat and megabat kinds.[40] These groups are listed in column 2 of table 1. Using the *kmeans* clustering algorithm, the 27 species grouped together into eight clusters. The classification results are presented in column 3 of table 1.

Analysis of whole genome sequences using WGKS

Next, we applied WGKS and tested for clustering performance. The Hopkins clustering value is 0.896, which indicates that the matrix is very good for clustering. The results are shown in figure 4. Statistical measures, such as the minimum, mean, maximum PCC value, standard deviation and p-value are provided in table 2.

Kmeans clustering was also applied to the PCC matrix, and nine clusters were predicted (see table 1, column 4). Supplementary file part 3 contains the PCC matrix as well as the clusters for each of the species in the WGKS analysis. Interestingly, *Homo sapiens* clusters separately from all other mammals, showing that it is indeed a unique species and forms its own kind. Its mean PCC with all other species is 0.226 (± 0.002 SD), which is very low compared to all other kinds, again telling us humans are separate from all other species (Genesis 1:27).

The cat kind forms a distinct group, discontinuous from all other species. One species, *Felis nigripes* (black-footed cat) stands out from the other cats. Based on the mitochondrial DNA, the Y chromosome, the cytochrome b gene, and the 12S rRNA, some (evolutionary) researchers think that *F. nigripes* diverged early from all other cats.[41,42] This species is found in South Africa, Botswana, Namibia, and Zimbabwe, far from the initial post-Flood dispersal site of Ararat. This species has a mean PCC of 0.903 (± 0.012 SD) with all other cat species from the cat kind, whereas the other cats have a mean PCC of 0.979 (± 0.012 SD) among themselves.

In the dog kind, the Arctic fox (*Vulpes lagopus*) stands out from the four other dog species, with a mean PCC of 0.672 (± 0.014 SD) compared with all other dog species, which have a mean PCC of 0.843 (± 0.149 SD). Why the Arctic fox stands out is unclear, but as an arctic species it is highly adapted to a cold climate, demonstrated by a faster metabolism, specialized circulation, foot pads, and smaller/shorter extremities.[43]

Mice and rats have a rather low PCC value of 0.463. However, a previous molecular baraminology study also grouped these two species into the same kind.[16]

Table 2. Statistical results for the WGKS algorithm on 28 mammal species

cluster	species	min	mean	max	stdev	p-value
cats	11	0.879	0.965	0.998	0.032	6.52E-162
dogs	4	0.939	0.957	0.985	0.020	1.76E-25
horses	3	0.924	0.950	0.995	0.039	9.34E-05
murids	2	0.463	0.463	0.463	NA	1.3E-27
vespertilionids	2	0.673	0.673	0.673	NA	3.26E-39
megabats	3	0.978	0.983	0.991	0.007	2.32E-40

Further, bats form three distinct groups, two of which are represented by the suborder *Microchiroptera* (microbats), and one by the suborder *Megachiroptera* (megabats). It is unknown if they represent different kinds if separation and diversification occurred after the Flood among the bats. As with other animal species, God could have created multiple

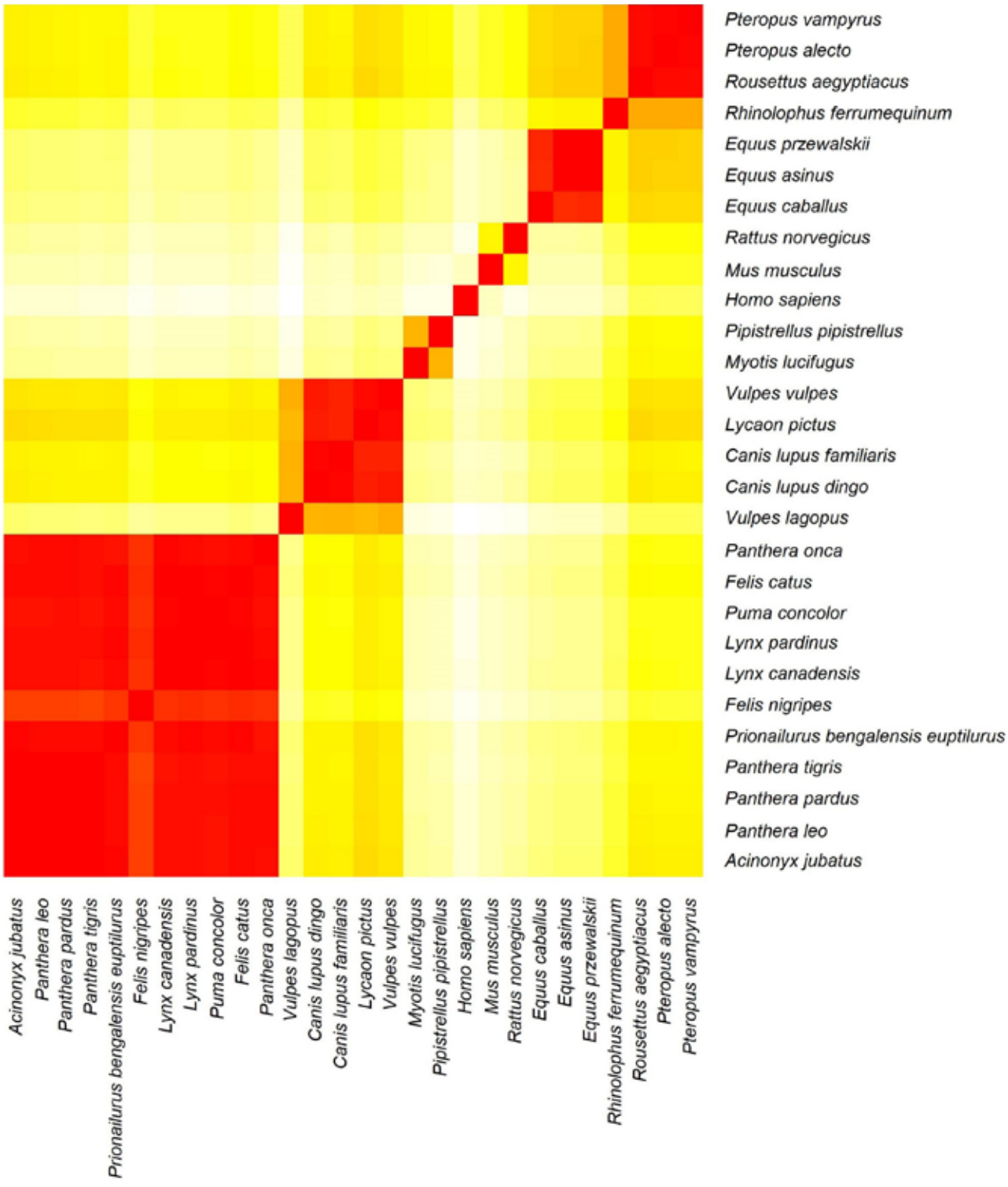

Figure 4. Heatmap of 29 species based on WGKS similarity. Higher species similarities between a given species pair correspond to redder colours. Lower similarity values correspond to yellow colours.

bat baramins. In this case, some use echolocation, whereas others do not.[44] Echolocation is a complex trait, and so is not expected to vary within a kind, unless it is lost.

There are two main groups of microbats. These results mirror the results of a new molecular baraminology study which places *M. lucifugus* and *P. pipistrellus* into a tentative baramin, called Vespertilionoidea, and *R. ferrumequinum* into another tentative baramin, called Rhinolophoidea (see figure 4).[45] There is but one main grouping among the megabats. Two baraminology studies put megabats into their own baramin.[45,46] *Rousettus aegyptiacus* (Egyptian fruit bat), *Pteropus alecto* (black flying fox), and *Pteropus vampyrus* (large flying fox) cluster together within a megabat monobaramin (figure 4). *Pteropus* and *Rousettus* are both genera within the subfamily Pteropodinae, family Pteropodidae (flying foxes).

In figure 4, the three horse species (*Equus asinus*, *Equus caballus*, and *Equus przewalskii*) are found next to *R. ferrumequinum*. The three horse species show a very high mean PCC value of 0.950 (± 0.039 SD). *E. asinus* and *E. caballus* have a PCC value of 0.924, *E. asinus* and *E. przewalskii* have a value of 0.995, and *E. caballus* and *E. przewalskii* have a PCC value of 0.932. Outside of their own kind, the three horse species have the highest average PCC values with the three megabat species (0.555–0.588, ± 0.013 SD), followed by the microbat *R. ferrumequinum* (0.469), see table 3. Interestingly, an evolutionary study based on L1 retrotransposon insertions concluded that *Chiroptera* (bats) and *Perissodactyla* (odd-toed ungulates, such as horses) form sister groups in the clade *Pegasoferae*, named after the mythical flying horse from Greek mythology.[47]

Nevertheless, *WGKS* accurately predicts these assumed baramins. Nine of the 27 species were misclassified by the mtDNA analysis (in bold in column 3 of table 1), corresponding to a classification efficiency of only 66.7%. In contrast, only one out of 28 species were misclassified by *WGKS* (for a clustering efficiency of 96.4%).

Discussion

This study highlights one of the main problems of baraminology: the subjective classification of species into baramins. The fact that life follows a hierarchical pattern makes it hard to delimit the created kinds. The thought that God frontloaded the created kinds with diversity-generating mechanisms[1] means that major changes to morphology, gene content, and gene sequence could have easily occurred within the creation model. Designed hemizygosity[48] could also lead to discordant gene content in the descendant species of any holobaramin. Even so, species are sometimes classified together into baramins based on intuitive considerations. While intuition may be a more subjective means of classifying organisms, it is still a part of science due to its inherent recognition of patterns. God created organisms and our goal as creation scientists is to think God's thoughts after Him. Adam, as the world's very first taxonomist was able to name the animals when God led them to him (Genesis 2:19–20), albeit with perfect, sinless mental capabilities. We may have to rely on a holistic, multi-lined approach to baraminic classification, including biblical, morphological, hybridization, and genetic data, as discussed in Ahlquist and Lightner (2019).[49]

Statistically speaking, a PCC value of > 0.7 denotes a strong correlation between two vectors. This could possibly be used as a cut-off limit, but this remains to be evaluated. Another consideration is to use good quality datasets with good clustering statistics (i.e. Hopkins measurement > 0.75). Or, the data can be pre-processed (i.e. normalization). Furthermore, apparent outliers may be excluded from the data set before clustering or noted as a small cluster. Post-clustering steps may also be taken to ensure quality results.[39] Several alternate statistical measures may indicate the presence of multiple clusters within a given data set. Testing for a Gaussian distribution in the distance/correlation measures can detect the presence of multiple clusters.[50] Different clusters may have different levels of distance values, each of which follows their own Gaussian distribution. Multiple modes (peaks of similarity) within the distribution are often important for discriminating groups. Applying principal component analysis to a non-Gaussian distribution can sometimes help to estimate the number of clusters.[51,28]

One possible way forward is to apply algorithms which automatically estimate the number of clusters within a

Table 3. PCC values between the horse species and several bat species

	Equus asinus	*Equus caballus*	*Equus przewalskii*
Pteropus alecto	0.581	0.555	0.582
Pteropus vampyrus	0.578	0.553	0.579
Rousettus aegyptiacus	0.588	0.563	0.588
Rhinolophus ferrumequinum	0.476	0.455	0.476

baraminology data set, instead of looking for a certain cut-off parameter. One such algorithm is the 'gap statistic' algorithm. This algorithm calculates the within-cluster sum of squares around the cluster mean for *k* clusters. The 'optimal' number of clusters can be thought of as the point after which the number of clusters does not decrease drastically.[52]

Another algorithmic procedure would involve estimating baramins by starting with the selection of a seed species. This seed species would represent a given baramin. Next, other species would be chosen which have the highest JCV/PCC values compared to the seed species. A mutually overlapping JCV value could be calculated for all species added to the seed species. The algorithm would stop adding species when either there is a statistically significant drop in the JCV values of the species added to the seed species.

This was done in the baraminology analysis of several halophilic species of Archaea. The results indicated that the difference in JCV between the three individual halophile species (*Halobacterium hubeiense*, *Halococcus salifodinae*, and *Halosimplex carlsbadense*) and their associated species were statistically significantly different than the JCV calculated for all other species (p-values: *H. hubeiense*: 2.4E-99; *H. salifodinae*: 1.9E-126; *H. carlsbadense*: 9E-115).[15]

Yet, this is similar to standard phylogenetic tree construction techniques, only using JCV instead of sequence data. Also, the algorithm could be tricked if it found a group of recently diverged species that, in turn, separated early from the parent baramin.

Conclusions

The developing field of molecular baraminology is an exciting, yet challenging area of study within creation science. With further study we should be able to make useful progress. More than likely, a combined approach that uses morphological and molecular statistics, including orphan genes and hemizygous states, as well as a limited amount of user inference, will be needed.

Glossary

BDIST method: a morphology-based baraminology method, based on comparing shared characteristics between a set of species.

Cladistics: a method of grouping organisms based on the proportion of characteristics that they have in common.

Cryptospecies: isolated populations of the same species which become incapable of mating with one another due to some genetic mutation.

Deuterostome: a large group of animals characterized by its embryonic development in which its first embryonic opening becomes the anus.

GCM: Gene Content Method, an algorithm, which determines baraminic membership based on the proportion of shared (orthologous) genes among species.

Holobaramin: the complete created kind, including all species pertaining to that kind.

k-*mer*: a segment of the DNA k bp long.

Monobaramin: a group of interrelated species that cluster together within a baramin.

Orthologous protein: a protein present in two species with a high degree of sequence similarity and which perform essentially the same function.

Passenger gene: genes which are not essential for an organism, and which were picked up from the environment, common in bacteria.

PCC: Pearson Correlation Coefficient, a statistical measure which tells us how similar two vectors are.

Pleiotropic gene: a gene with multiple effects in the phenotype.

Tetrapod: an animal with four limbs.

Retrotransposon: mobile genetic elements, which copy themselves into different parts of the genome, doing so through an RNA intermediate.

WGKS method: a molecular baraminology method which analyzes and compares the k-mer content of species within a study.

References

1. Terborg, P., Evidence for the design of life: part 2—Baranomes, *J. Creation* **22**(3):68–76, 2008.
2. Tomkins, J. and Bergman, J., Incomplete lineage sorting and other 'rogue' data fell the tree of life, *J. Creation* **27**(3):84–92, 2013.
3. Gibbons, A., How an ancient microbial arms race remodeled human cells, *Science* **369**(6503):491–492, 2020.
4. Terborg, P. and Truman, R., The FOXP2 gene supports Neandertals being fully human, *J. Creation* **22**(2):13–14, 2008.
5. Miga, K.H., Chromosome-specific centromere sequences provide an estimate of the ancestral Chromosome 2 fusion event in hominin genomes, *J. Heredity* **108**(1):45–52, 2017.
6. Wood, T.C. and Murray, M.J., *Understanding the Pattern of Life: Origins and organization of the species*, Broadman & Holman, Nashville, TN, 2003.
7. O'Micks, J., Baraminology classification based on gene content similarity measurement, CRSQ **54**(1):27–37, 2017.
8. Cserhati, M., A new baraminology method based on Whole Genome K-mer Signature analysis and its application to insect classification, *J. Creation* **34**(1):86–95, 2020.
9. Wood, T.C., Baraminic distance, bootstraps, and BDISTMDS, *Occasional Papers of the Baraminology Study Group* **12**:1–17, 2008.
10. Robinson, D.A., A mitochondrial DNA analysis of the Testudine apobaramin, CRSQ **33**:262–272, 1997.
11. Wood, T.C., Mitochondrial DNA analysis of three terrestrial mammal baramins (Equidae, Felidae, and Canidae) implies an accelerated mutation rate near the time of the Flood; in: Horstemeyer, M. (Ed.), *Proceedings of the Seventh International Conference on Creationism*, Creation Science Fellowship, Pittsburgh, PA, 2013.
12. Tomkins, J. and Bergman, J., Genomic monkey business—estimates of nearly identical human–chimp DNA similarity re-evaluated using omitted data, *J. Creation* **26**(1):94–100, 2012.
13. Batten, D., The myth of 1%: human and chimp DNA are very different, *Creation* **36**(1):35–37, 2014. See also 'Human and ape similarity' in Wieland, M., Convert to Creation, *Creation* **40**(3):36–39, 2018.

14. O'Micks, J., Baraminic analysis of nucleocytoplasmic large DNA viruses, *J. Creation* **31**(1):73–79, 2017.
15. Yaugh, A., Baraminological analysis of a set of Archaea species based on genomic data, CRSQ **53**(2):140–154, 2016.
16. Lightner, J.K. and Cserhati, M., The uniqueness of humans is clearly demonstrated by the gene-content statistical baraminology method, CRSQ **55**(3):133–141, 2019.
17. Cserhati, M., Whole Genome K-mer Signature-based clustering of algae species, *Algal Research*, submitted, 2020.
18. Cserhati, M., A new baraminology method based on Whole Genome K-mer Signature analysis and its application to insect classification, *J. Creation* **34**(1):86–95, 2020.
19. Cserhati, M., A tail of two pandas— Whole Genome K-mer Signature Analysis of the Red Panda (*Ailurus fulgens*) and the Giant Panda (*Ailuropoda melanoleuca*), *Mol. Phylogenetics and Evolution*, submitted, 2020.
20. Nishihara, H., Hasegawa, M., and Okada N., Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions, *Proc. Natl. Acad. Sci. USA* **103**(26):9929–9934, 2006.
21. Wood, T.C., Baraminological analysis places *Homo habilis*, *Homo rudolfensis*, and *Australopithecus sediba*, in the human holobaramin, ARJ **3**:71–90, 2010.
22. Wood, T.C., *Australopithecus sediba*, statistical baraminology, and challenges to identifying the human holobaramin; in: Horstemeyer, M. (Ed.), *Proceedings of the Seventh International Conference on Creationism*, Creation Science Fellowship, Pittsburgh, PA, 2013.
23. O'Micks, J., Preliminary baraminological analysis of *Homo naledi* and its place within the human baramin, *J. Creation Theology and Science Series B: Life Sciences* **6**:31–39, 2016.
24. O'Micks, J., *Homo naledi* probably not part of the human holobaramin based on baraminic re-analysis including postcranial evidence, ARJ **9**:263–272, 2016.
25. Bogusz, D., Appleby, C.A., Landsmann, J., Dennis, E.S., Trinick, M.J., and Peacock, W.J., Functioning haemoglobin genes in non-nodulating plants, *Nature* **331**(6152):178–80, 1988.
26. Li, Y., Liu, Z., Shi, P., and Zhang J., The hearing gene Prestin unites echolocating bats and whales, *Curr. Biol.* **20**(2):R55–56, 2010.
27. Levitan, D.R. *et al.*, Mechanisms of reproductive isolation among sympatric broadcast-spawning corals of the *Montastraea annularis* species complex, *Evolution* **58**(2):308–23, 2004.
28. Manica, A. and Carter, R.W., Morphological and fluorescence analysis of the *Montastraea annularis* species complex in Florida, *Marine Biology* **137**:899–906, 2000.
29. Hennigan, T., An initial estimate toward identifying and numbering the Ark turtle and crocodile kinds, ARJ **7**:1–10, 2014.
30. Wood, T.C., *Animal and Plant Baramins*, Center for Origins Research Issues in Creation, no. 3, Wipf and Stock Publishers, Eugene, OR, pp. 24–27, 2008.
31. Cserhati, M., Do lizards and snakes form separate apobaramins? *J. Creation* **34**(1):96–101, 2020.
32. O'Micks, J., A preliminary cephalopod baraminology study based on the analysis of mitochondrial genomes and morphological characteristics, ARJ **11**:193–204, 2018.
33. Wagner, G., *Homology, Genes, and Evolutionary Innovation*, Princeton University Press, 2014.
34. ReMine, W.J., *The Biotic Message*, St Paul Science, St Paul, MN, 1993.
35. Statham, D., Problems with the evolutionary interpretation of limb design, *J. Creation* **26**(2):10, 2012.
36. Sanders, H.F. and Cserhati, M., Statistics, baramins, and interpretations: a critical evaluation of BDIST, CRSQ, submitted, 2020.
37. de Manuel, M. *et al*., The evolutionary history of extinct and living lions, PNAS **117**(20):10927–10934, 2020.
38. Pendragon, B. and Winkler, N., The family of cats—delineation of the feline basic type, *J. Creation* **25**(2):118–124, 2011.
39. Lassmann, T., Frings, O., and Sonnhammer, E.L., Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features, *Nucleic Acids Res.* **37**(3):858–865, 2009.
40. Cserhati, M., Probing the mysteries of bat baraminology, *J. Creation*, manuscript in preparation, 2020.
41. Renard, A., Lavoie, M., Pitt, J.A., and Larivière, S., *Felis nigripes* (Carnivora: Felidae), *Mammalian Species* **47**(975):78–83, 2010.
42. Pecon-Slattery, J. *et al*., Phylogenetic assessment of introns and SINEs within the Y chromosome using the cat family felidae as a species tree, *Mol. Biol. Evol.* **21**(12):2299–2309, 2004.
43. Prestrud, P., Adaptations by the Arctic Fox *(Alopex lagopus)* to the Polar Winter, *Arctic* **44**(2):132–136, 1990.
44. Parker, J. *et al*., Genome-wide signatures of convergent evolution in echolocating mammals, *Nature* **502**(7470):228–231, 2013.
45. Cserhati, M., Molecular bat baraminology predicts four bat baramins, *J. Creation*, submitted, 2020.
46. Lightner, J.K., Mammalian Ark kinds, ARJ **5**(2012):151–204, 2012.
47. Nishihara, H., Hasegawa, M., and Okada, N., Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions, *Proc. Natl. Acad. Sci. USA* **103**(26):9929–34, 2006.
48. Terborg, P., The hemizygosity hypothesis—a novel genetic paradigm for baranomes, *J. Creation* **34**(2):111–116, 2020.
49. Ahlquist, J. and Lightner, J.K., Strategies for more clearly delineating, characterizing, and inferring the natural history of baramins I, CRSQ **56**(2):97–104, 2019.
50. Kimes, P.K., Liu, Y., Hayes, D.N., and Marron, J.S., Statistical significance for hierarchical clustering, *Biometrics* **73**(3):811–821, 2017.
51. Adolfsson, A., Ackerman, M., and Brownstein, N.C., To cluster, or not to cluster: an analysis of clusterability methods, *Pattern Recognition* **88**:13–26, 2019.
52. Tibshirani, R., Walther, G., and Hastie, T., Estimating the number of clusters in a data set via the gap statistic, *J. R. Statist. Soc. B.* **63**(2):411–423, 2001.

Matthew Cserhati *has a Ph.D. in biology and a B.Sc. in software development.. He has been an active creationist for 19 years and takes a great interest in molecular biology. He has published a number of articles in* Journal of Creation*. Matthew is currently studying at Greenville Presbyterian Theological Seminary.*

Robert Carter *received his Bachelor of Science in Applied Biology from the Georgia Institute of Technology in 1992 and his Ph.D. in coral reef ecology from the University of Miami in 2003. He has studied the genetics of fluorescent proteins in corals and sea anemones and holds one patent on a particular fluorescent protein gene. His current research involves looking for genetic patterns in the human genome and the development of a biblical model of human genetic history. He works as a speaker and scientist at CMI-US.*