Science Controversies online

Genomics: Keystone Paper with annotations

Originally published in Science Express as 10.1126/science.1061036 on May 17, 2001
Science, Vol. 292, Issue 5523, 1903-1906, June 8, 2001

Microbial Genes in the Human Genome: Lateral Transfer or Gene Loss?

Annotation

Steven L. Salzberg,* Owen White, Jeremy Peterson, Jonathan A. Eisen

A1 The human genome was analyzed for evidence that genes had been laterally transferred into the genome from prokaryotic organisms. A2 Protein sequence comparisons of the proteomes of human, fruit fly, nematode worm, yeast, mustard weed, eukaryotic parasites, and all completed prokaryote genomes were performed, and all genes shared between human and each of the other groups of organisms were collected. A3 About 40 genes were found to be exclusively shared by humans and bacteria and are candidate examples of horizontal transfer from bacteria to vertebrates. A4 Gene loss combined with sample size effects and evolutionary rate variation provide an alternative, more biologically plausible explanation.

The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA.

* To whom correspondence should be addressed. E-mail: salzberg@tigr.org

A1 The scientific question this research project aimed to answer

A2 Information about materials and methods

A3 Short description of results

A4 A brief discussion and conclusion

A5 Studies of the evolution of species long assumed that gene flow between species is a minor contributor to genetic makeup, generally thought to only occur between closely related species. This picture changed when researchers began to study the genetics of microorganisms. Genes, including those encoding antibiotic resistance, can be exchanged between even distantly related bacterial species (horizontal or lateral gene transfer). A6 A growing body of evidence suggests that lateral gene transfer may be a much more important force in prokaryotic evolution than was previously realized (1). Lateral gene transfers involving eukaryotes have also been well documented, in most cases involving transfers from

A5 In the first paragraph the authors provide a brief introduction to the topic.

A6 After a brief description, the authors pinpoint two relevant research papers from 1999 (1) and 1998 (2).

organellar genomes into the eukaryotic nucleus (2).

A7 Analysis of the rough draft of the human genome led to the suggestion recently (3) that 223 bacterial genes have been laterally transferred into the human genome sometime during vertebrate evolution. A8 Such a possibility is of interest because it implies that bacterial infections have led to permanent transfer of genes into their hosts. A9 One possible implication is that bacteria might be manipulating the human genome for their own benefit and that this process may be continuing. A10 Such an event would require (i) that genes be transferred into the germ cell lineage, not just into any somatic cell, and (ii) that the transferred genes be stably maintained in the host cell, either by insertion into a chromosome or as extrachromosomal elements. For these genes to spread through the population, they need either to provide a selective advantage to their host or to exhibit some kind of "selfish" properties, such as the ability to duplicate and transpose.

A11 Although the possibility of lateral gene transfer has gained much support in recent years from analysis of complete genome sequences (1, 4, 5), A12 the inference of such gene transfer events is still fraught with difficulty, because of problems with methods and with the data analyzed (6, 7). A13 As in the recent study (3), we focused on detecting possible gene transfers from bacteria to vertebrates by analysis of gene distribution patterns across taxa. Those genes found in bacteria and vertebrates but not in nonvertebrates are considered possible cases of lateral transfer (putative bacteria to vertebrate transfers, or BVTs). A14 Our study differed in that it included the human proteome reported by Venter *et al*. (8) and it included proteins from parasite lineages not included in the previous study (9).

A15 We focused on analyzing complete genome sequences because the absence of a gene from a species cannot be inferred from incomplete genome sequences. Human genes for which homologs are found in completed prokaryotic genomes were identified by searching against all publicly available complete genome sequences. For our analysis of the human proteome, we used the Ensembl set, containing 31,780 proteins (3), and the Celera set, containing 26,544 proteins (8). In the Ensembl proteome, 4388 genes have BlastP matches with E-values less than 10^{-10} to a protein from a complete prokaryotic genome. Likewise, 3915 genes from the Celera proteome match at least one prokaryotic gene with the same E-value threshold (Table 1). As in (3), transfers into vertebrates were ruled out if a homolog of a gene was found in a nonvertebrate eukaryotic genome.

A7 The authors are more specific when they mention the results reported in a research paper published in 2001 (3) that is relevant to their work.

A8 By mentioning the relevance of the BVT hypothesis, the authors justify their work.

A9 Implication of the BVT hypothesis

A10 According to the authors, the following phenomena are preconditions for acceptance of the BVT hypothesis.

A11 The authors cite previous observations related to the BVT hypothesis. They cite reference 1 again and add two references from 1999 (4, 5). By doing so they make it clear that the BVT hypothesis is supported by other research results.

A12 At this point, they mention that other groups have criticized the instrumentation used in previous research that set the basis for the BVT hypothesis. Discussing divergent views in science is a valuable procedure, because it shows that the authors take a realistic view of the existing state of knowledge that is relevant to the work.

A13 Every scientific

nonvertebrate eukaryotic genome.

Table 1. Proteome sizes and number of genes shared with each of the human protein sets, with a Blast cutoff of 10^{-10}.

Organisms	Number of proteins	Number matching Ensembl proteome	Number matching Celera proteome
Human	-	31,780	26,544
Bacteria/Archaea	85,824	4,388	3,915
Yeast	9,030	7,508	7,103
C. elegans	19,400	13,770	12,660
D. melanogaster	14,080	15,324	14,302
A. thaliana	25,470	9,151	9,081
Parasites	11,606	5,146	4,756

investigation has a scientific question and the aim of research is to find a suitable answer to the question. Not every research project succeeds in answering the initial scientific question. Sometimes it just offers clues for additional investigation. Other times it offers an answer. Usually the scientific question is not presented as a regular question, with an inquisitive tone followed by a question mark. The scientific question the authors are trying to answer here is: How strong is the evidence that all 223 genes have been laterally transferred into the human genome from prokaryotic organisms? Because the question they are trying to answer was presented before by other groups as a hypothesis to explain an observation (the presence of 223 "bacteria-like" genes in the human genome) we simply call that the BVT (bacteria to vertebrate transfer) hypothesis.

A14 The authors mention a few words about the difference between their work and previous work by other groups. With that they inform that they are not simply repeating previous work but rather are using a different approach to answer the same question. In this

total # of BVTs to be independent.

A16 If the pattern of genes shared between prokaryotic and eukaryotic species is a robust measure of lateral gene transfer, A17 then we would expect that the total number of true BVTs would be independent of which and how many nonvertebrate genomes have been sampled. A18 However, as the number of nonvertebrate proteomes screened against human increased, the number of BVTs decreased (Fig. 1). A19 The two plots show comparable results for the Ensembl and Celera protein sets, and each line shows the effect with a different starting proteome. Subsequent points on the plots show averages after removing one more proteome; for example, the "fruit fly" line shows the average number of genes remaining in the BVT set after removing all *Drosophila melanogaster* genes plus one, two, three, and four additional protein sets. After removal of all genes found in complete nonvertebrate genomes, only 135 Ensembl genes and 89 Celera genes remained as possible BVTs.

Fig. 1. Genes shared by humans and

case, the information added here indicates that their methodology is more complete than that used by the other team.

A15 Here we can find a lot of information about the materials and methods they used to carry out their research.

A16 Every hypothesis formulates a prediction that should represent a logical outcome. The hypothesis is formulated as a possible answer to a scientific question and it can anticipate the results. If the prediction is confirmed, then the hypothesis should be accepted. If the prediction is not confirmed, then at least two situations are possible: the hypothesis is not valid or the assumptions underlying the hypothesis and its predictions were misleading. [more information]

A17 Prediction-logical outcome

A18 The authors first present the results of their experimental test that do not corroborate the predictions.

A19 The authors present their results in detail.

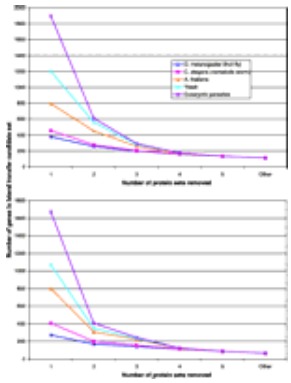

Fig. 1. Genes shared by humans and prokaryotes after removing successive proteome sets from five nonvertebrates and a collection of miscellaneous nonvertebrates ("Other"). (**Top**) Ensembl protein set. (**Bottom**) Celera protein set. [View Larger Version of this Image (18K GIF file)]

The downward trend of the plot in Fig. 1 suggests that the number of BVTs might decrease further if more nonvertebrate genomes are added to the analysis. Our analysis confirms this: Searching through all proteins in GenBank from numerous other eukaryotic nonvertebrates (labeled "Other" in Fig. 1), most of which have a relatively small number of characterized genes, identified matches to organisms such as *Suberites domuncula* (sponge), soybean, and *Aspergillus terreus*. As a result of this filtering, 21 genes were removed from the Ensembl BVTs and 21 from the Celera BVTs, leaving only 114 and 68 genes in the two sets, respectively.

A20 One explanation for the species-sampling effect shown in Fig. 1, and the reason why species distribution patterns must be interpreted with great caution, is the phenomenon of gene loss. It is likely that many genes shared by the eukaryotic common ancestor have been lost in some lineages. This seems especially likely in some of the species analyzed here, such as *Arabidopsis thaliana*, which was chosen for genome sequencing in part because of its small genome size, and *Saccharomyces cerevisiae*, for which extensive gene loss has been documented (10). A21 A simple computation illustrates the possible contribution of gene loss to the pattern. Suppose the five eukaryotic genomes analyzed all resulted from a single adaptive radiation. If this common ancestor started with 10,000 genes [see Rubin *et al.* (11) for a discussion of "core proteome" sizes] and each lineage lost 30% of its genes, then the probability that any one gene was lost from four lineages is $(0.3)^4 = 0.00081$, or 81 genes lost from all four of the nonvertebrate lineages. Of course, some genes are probably less likely to be lost than others (e.g., DNA polymerase genes). Supposing that 20% of a proteome cannot be lost, then 30% loss translates into 65 genes lost in all four lineages. It appears likely that gene loss alone could account for a large proportion of the BVT set.

A22 Another important aspect of the species-sampling effect is the phylogenetic bias in the data sets being analyzed. All of the eukaryotic complete genomes are from so-called "crown" eukaryotes: animals, plants, and fungi. In addition, three of these

A20 The authors start their discussion by formulating a hypothesis to explain one of their results. This hypothesis predicts that a known phenomenon called "gene loss" is the most suitable explanation for their findings. Gene loss is the first explanation of a final hypothesis that comprises three explanations.

A21 Here they add experimental evidence that supports the gene loss hypothesis.

A22 They indicate an additional observation--limitation of sample diversity--that is

(*Caenorhabditis elegans*, *D. melanogaster*, and *Homo sapiens*) are animals, further limiting the sample of evolutionary diversity. In contrast, the sampling of prokaryotic evolutionary diversity is much broader, containing representatives from many widely divergent bacterial and Archaeal lineages (12). A23 It seems likely that the sequencing of a broader variety of eukaryotic genomes will lead to a further reduction in the number of BVTs.

A24 The rate of nucleotide substitution varies for different genes within a genome as well as for the same gene in different species. This rate variation is due to a combination of factors, including variation in DNA replication accuracy, DNA repair, selection, recombination, genetic drift, and generation time (13). A25 Because of the effects of rate variation, sequence similarity alone is not an accurate measure of evolutionary relatedness (14, 15). Thus, Blast E-values, which are measures of sequence similarity, should not be used to measure evolutionary relatedness (15). This is particularly true in analyses of complete genomes, where it can be expected that at least some genes will be nonessential, with low selective pressure allowing more rapid mutation. A26 In the analysis used to support the claim that 223 genes have been laterally transferred into human (3), a gene was considered a BVT if the Blast score for the bacterial match was at least 10^{-9}-fold smaller than the nonvertebrate match score. From a statistical perspective, the null hypothesis should be that two genes with sufficiently high sequence similarity share a common ancestor. A27 Our analysis used the same threshold for prokaryotic and nonvertebrate matches, with a maximum E-value cutoff of 10^{-10} (i.e., the likelihood that any Blast hit was due to chance is less than 1 in 10^{10}). The use of any fixed E-value cutoff, though, will miss genes with slightly weaker similarity to nonvertebrate proteins. Because the weaker alignment scores may simply be the result of more rapid mutation in the invertebrate lineage, it is impossible to rule out common ancestry on the basis of this evidence alone. By reducing the E-value cutoff for nonvertebrate genes to 10^{-7}, we reduced the size of the Ensembl BVT set to 74 genes and the Celera BVT set to 56 genes. In addition, after comparing the 74 Ensembl BVTs to invertebrate mitochondrial genomes, we found two genes of mitochondrial origin, reducing that BVT set to 72 genes.

likely to be relevant to the results found. This is the second explanation of their hypothesis.

A23 The authors predict future results, suggesting that newly available data will likely contribute additional confirmation of their current hypothesis.

A24 They introduce a different aspect (rate of nucleotide substitution) to the issue that corroborates their results and disputes the BVT hypothesis. This is the third explanation in their hypothesis.

A25 Apart from their results and based on the phenomenon of "rate variation of nucleotide substitution," the authors explain the reasons that led them to reject the BVT hypothesis. These reasons are based on the methodology previously used to develop the BVT hypothesis.

A26 The authors comment and criticize the methodology used by the group supporting the BVT hypothesis.

A27 The authors draw a parallel between the method used to analyze the data in the paper mentioned above and their method. They explain why their method of analysis is more suitable.

A28 If a gene was transferred from a prokaryotic lineage into the vertebrate lineage, A29 this likely occurred within the past 400 to 500 million years, after most of the major prokaryotic phyla were established. A30 Therefore, any transferred gene should be more closely related to its donor lineage than to any other prokaryotic lineage, which would be detectable in phylogenetic trees. A31 For example, phylogenetic trees built from genes that have been transferred from mitochondrial or plastid genomes to eukaryotic nuclei (16-18) indicate that the transferred genes branch with α-proteobacteria and cyanobacteria, respectively. A32 We generated phylogenetic trees for genes from the BVT sets for which sufficient numbers of related genes were available and found that most did not show patterns consistent with bacterial to vertebrate gene transfer. One such example is shown in Fig. 2, which shows a phylogenetic tree of three human hyaluronan synthase paralogs, all from the BVT set reported in (3). A33 The phylogenetic analysis reveals that the vertebrate genes do not branch within any particular prokaryotic lineage. A34 Instead, the placement of groups in the tree is consistent with normal vertical inheritance*; the absence of the gene from nonvertebrate lineages may be due either to gene loss or rate variation.

A28 Still working on rejecting the BVT hypothesis as the main explanation for the presence of all 223 bacteria-like genes in the human genome, the authors develop a conditional argument. The conditional argument functions as a logical model to test a hypothesis. The conditional argument comprises two promises that can be expressed in the form "If p...then q," where p is the antecedent and q is the conclusion. The conditional argument includes an additional fact or observation. For example, "if p...taken r in consideration, then q." Here is the first promise of their conditional argument (p).

A29 Here is the "taken this in account" (r)

A30 Prediction of the hypothesis. Then...(q)

A31 The authors added a previous observation that corroborates the stated conclusion of the conditional argument. This observation validates the prediction, showing that the statement to design the conditional argument was not misleading. Likewise, if an observation does not

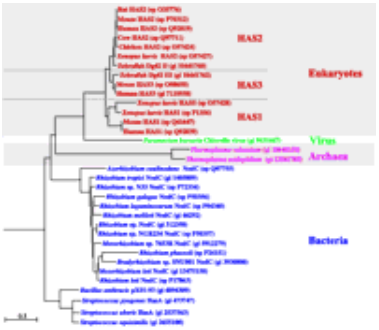

Fig. 2. Phylogenetic tree of homologs of three human hyaluronan synthase (HAS) proteins that were proposed as lateral transfers from bacteria to vertebrates (3). Homologs of the human HAS genes were identified with iterative Blastp searches of a low-redundancy protein database and aligned with clustalW. More distantly related proteins were used as outgroups to root the tree. The tree was generated from the alignment (variable regions and gaps excluded) with the neighbor-joining algorithm implemented by Phylip (25) with a PAM-based distance matrix. Species names, major evolutionary groupings, gene names if available, and sequence IDs (gi for Genpept and sp for Swissprot) are indicated in the tree. Scale bar corresponds to estimated evolutionary distance units. The presence of multiple HAS genes in different vertebrate species is likely due to duplication in vertebrates.
[View Larger Version of this Image (45K GIF file)]

A35 The absence of a gene from the annotation for fruit fly, nematode, or any other organism is not proof that the gene is missing from that organism's genome. A36 First, not all of these genomes are complete. A37 Second, the annotation of the completed portions of some eukaryotic genomes is still in progress, A38 and the state of the art in eukaryotic gene finding is imperfect. A39 To check for genes missing from the annotation, we

corroborate the prediction, then the conditional argument should be reformulated.

A32 After validating the logical outcome of their hypothesis, they perform an experimental test to investigate whether the predictions can be confirmed.

A33 The predictions for the conditional argument cannot be confirmed.

A34 When the predictions cannot be confirmed, the authors rule out the BVT hypothesis and offer an alternative hypothesis to fill the blank ("gene loss or rate variation"). Note that the authors reject the BVT hypothesis because their results reject the predictions of a conditional argument they formulated with an analogy to the following observation: "For example, phylogenetic trees built from genes that have been transferred from mitochondrial or plastid genomes to eukaryotic nuclei (16-18) indicate that the transferred genes branch with proteobacteria and cyanobacteria, respectively".

A35 The authors add three reasons--observations--to justify nonacceptance of the BVT hypothesis.

A36 First reason

A37 Second reason

imperfect. A39 To check for genes missing from the annotation, we used TBlastN to search the human proteins from the initial BVT sets against the nucleotide sequences of the genomes of complete Eukaryotes. A40 This analysis resulted in two matches between Ensembl BVTs and *A. thaliana* and three matches to *Caenorhabditis elegans*, all with E-values of 10^{-32} or lower. Three of these five genes had already been removed in the steps that reduced the set to 72 BVTs; removal of the other two left 70 Ensembl BVTs.

A41 The Ensembl proteome set has been further curated, and numerous genes have been removed from the 31,780 used for the analysis in (3). The October release (version 8.0), containing 29,304 genes, has eliminated some genes (including possible contaminants), collapsed multiple genes into one, and otherwise improved the data. A42 We screened the 70 BVTs against the newer proteome and A43 found that 23 genes had been eliminated, reducing the BVT set to 47 genes. A44 If the original 135 Ensembl BVTs are screened against the newer release, A45 this set is reduced to 89 genes. A46 There were also 89 genes in the initial Celera BVT set.

A47 Comparing the 47 Ensembl BVTs against the 56 Celera BVTs yields some interesting final reductions in the data set. A48 Both sets contain genes not included in the other set; more interesting, though, are the genes shared between the two sets. In most cases, the sequences do not match exactly, and the differences in the gene models sometimes yield further matches to nonvertebrate genes. Of the 56 Celera BVTs, 10 genes match an Ensembl protein that in turn matches one or more nonvertebrates; six of these match all four of the complete nonvertebrate genomes. This reduces the Celera BVT set to 46 genes. Of the 47 Ensembl BVTs, five genes match Celera proteins that in turn match nonvertebrates, and one short (115 amino acid) protein falls on an 825-base pair unmapped contig, which appears to be a contaminant. This reduces the Ensembl BVT set to 41 genes.

A49 After careful reexamination of the human proteome, we find

A38 Third reason

A39 They perform a series of experiments to test their three-piece hypothesis (gene loss, rate variation of nucleotide substitution, and limitation of sample diversity).

A40 The experimental results confirm their hypothesis.

A41 The authors mention a fact that supports one of the reasons (reason 3) used to justify the logical basis of their hypothesis of gene loss.

A42 The authors perform some additional experimental tests that they expect will provide them with enough data to rule out BVT as the main explanation for the presence of all 223 bacteria-like genes in our genome and confirm their hypothesis of gene loss.

A43 Experimental result

A44 Experimental test

A45 Experimental result

A46 Fact

A47 Experimental test

A48 Experimental result

A49 This final result rejects

only 46 genes in the Celera protein set, and 41 in the Ensembl set, that comprise candidates for possible lateral transfer between bacteria and human (19). A50 The evidence presented here provides several plausible biological explanations for the presence of these genes in the human genome. A51 The argument for lateral gene transfer (3) is essentially a statistical one, necessarily so because of the inherent impossibility of observing events that may have occurred in the distant past. As with all statistical arguments, great care needs to be exercised to confirm assumptions and explore alternative hypotheses. A52 In cases where equally if not more plausible mechanisms exist, extraordinary events such as horizontal gene transfer do not provide the best explanation. A53 The more probable explanation for the existence of genes shared by humans and prokaryotes, but missing in nonvertebrates, is a combination of evolutionary rate variation, the small sample of nonvertebrate genomes, and gene loss in the nonvertebrate lineages.

that all 223 bacteria-like genes found in the human genome were laterally transferred from bacteria to humans. It is interesting to note that, although the authors have shown that 182 of the initial 223 genes do not correspond to BVT genes, 41 genes remain that are candidates for possible lateral transfer between bacteria and humans. At this point it becomes apparent that, instead just one hypothesis, the authors are actually testing 223 hypotheses because each bacteria-like gene presents in the human genome may or may not be the result of BVT. They conclude that only 41 genes can indeed be the result of BVT.

A50 In fact, the authors cannot rule out the BVT hypothesis completely because they could not demonstrate that all 223 genes were not the result of BVT. They demonstrated that 182 of 223 bacteria-like genes present in the human genome are NOT the consequence of BVT. For those 182 genes that are not the result of BVT, the authors formulate an alternative hypothesis.

A51 The authors criticize the methodology used to build the BVT hypothesis.

REFERENCES AND NOTES

1. W. F. Doolittle, *Science* **284**, 2124 (1999) [Abstract/Full Text].
2. W. Martin, *et al*., *Nature* **393**, 162 (1998) [CrossRef][ISI].

2. W. Martin, *et al.*, *Nature* **393**, 162 (1998) [CrossRef][ISI].
3. International Human Genome Sequencing Consortium, *Nature* **409**, 860 (2001).
4. K. E. Nelson, *et al*., *Nature* **399**, 323 (1999) [ISI][Medline].
5. W. F. Doolittle, *Trends Cell Biol*. **9**, M5 (1999) [CrossRef][ISI][Medline].
6. J. A. Eisen, *Curr. Opin. Genet. Dev*. **10**, 606 (2000) [CrossRef][ISI][Medline].
7. J. P. Gogarten, R. D. Murphey, L. Olendzenski, *Biol. Bull*. **196**, 359 (1999) [ISI][Medline].
8. J. C. Venter, *et al*., *Science* **291**, 1304 (2001) [Abstract/Full Text].
9. The complete sets of 31,780 and 26,544 proteins from the Ensembl and Celera human genome sets (www.ensembl.org/IPI and www.celera.com), which were the basis for the analyses of the human genome (3, 8), were used for all human sequence comparisons. The complete proteomes of yeast (*S. cerevisiae*) (20), nematode worm (*C. elegans*) (21), mustard weed (*A. thaliana*) (22), and fruit fly (*D. melanogaster*) (23) were collected, as was a set of all available protein sequences from the ongoing projects to sequence several eukaryotic parasites (*Plasmodium falciparum*, *Plasmodium yoelii*, *Trypanosoma brucei*, and *Theileria parva*, including preliminary genes annotated on unfinished sequences, available at www.tigr.org). The merged set of proteins from all completed prokaryotic genomes comprises 85,824 proteins (www.tigr.org/CMR). The human proteomes were searched against all proteins from all of these data sets with BlastP (24). All matches were collected, and those hits with a BLAST E-value of 10^{-10} or less were used for the initial analysis. Hits with larger E-values were collected and used for subsequent analyses. After searching all human genes against the complete prokaryotic sets, the resulting 4388 matches (for Ensembl) and 3915 matches (for Celera) formed the set of shared human-prokaryotic genes. Similarly, the genes shared by humans and each of the other four organisms or groups of organisms were collected. These databases were then compared with one another to determine the genes common to humans and prokaryotes but not found in fruit fly, worm, yeast, parasites, mustard weed, or any combination of those organisms' proteomes.
10. E. L. Braun, A. L. Halpern, M. A. Nelson, D. O. Natvig, *Genome Res*. **10**, 416 (2000) [Abstract/Full Text].
11. G. M. Rubin, *et al*., *Science* **287**, 2204 (2000) [Abstract/Full Text].
12. K. E. Nelson, I. T. Paulsen, J. F. Heidelberg, C. M. Fraser,

A52 One additional approach the authors used to rule out the BVT hypothesis is based on OckhamÕs razor, also spelled "OccamÕs razor," and called the "law of economy" or "law of parsimony." This is a principle stated by William of Ockham, that says "entities are not to be multiplied beyond necessity." In simple terms, it is a criterion for deciding among scientific theories or explanations. The criterion is based on the belief that the simplest explanation of a scientific phenomenon and the one that requires the fewest leaps of logic is the most likely to be true. They use this approach to support their hypothesis against BVT because their results alone could not rule out that BVT is responsible for the presence of all 223 bacteria-like genes in the human genome. [more information]

A53 The authors conclude by presenting a new hypothesis. This hypothesis is based on "gene loss, limitation of sample diversity, and variation of the rate of nucleotide substitution" to explain the presence of bacteria-like genes in the human genome.

Nature Biotechnol. **18**, 1049 (2000) [ISI].
13. W. H. Li, *Molecular Evolution* (Sinauer Associates, Sunderland, MA, 1997).
14. S. F. Altschul, *J. Mol. Evol*. **36**, 290 (1993) [ISI][Medline].
15. J. A. Eisen, *Genome Res*. **8**, 163 (1998) [Full Text].
16. J. D. Palmer, *et al*., *Proc. Natl. Acad. Sci. U.S.A*. **97**, 6960 (2000) [Abstract/Full Text].
17. S. G. Andersson, *et al*., *Nature* **396**, 133 (1998) [CrossRef][ISI][Medline].
18. X. Lin, *et al*., *Nature* **402**, 761 (1999) [ISI][Medline].
19. These gene sets are available as supplementary information at *Science* Online at www.sciencemag.org/cgi/content/full/1061036/DC1.
20. A. Goffeau, *et al*., *Science* **274**, 546 (1996) [Abstract/Full Text].
21. The *C. elegans* Sequencing Consortium, *Science* **282**, 2012 (1998).
22. The *Arabidopsis* Genome Initiative, *Nature* **408**, 796 (2000).
23. M. D. Adams, *et al*., *Science* **287**, 2185 (2000) [Abstract/Full Text].
24. W. Gish and D. J. States, *Nature Genet*. **3**, 266 (1993) [ISI][Medline].
25. J. Felsenstein, *Cladistics* **5**, 164 (1989) .
26. This work was funded in part by grants from NIH (R01 LM06845 to S.L.S.) and NSF (IIS-9902923 to S.L.S. and KDI-9980088 to S.L.S. and J.A.E.).

26 March 2001; accepted 4 May 2001
Published online 17 May 2001; 10.1126/science.1061036
Include this information when citing this paper.

Volume 292, Number 5523, Issue of 8 Jun 2001, pp. 1903-1906.

Lateral or horizontal gene transfer/BVT glossary
Transfer of genes from one evolutionary organism to another. The term distinguishes from vertical gene transfer that represents the parental transfer of genetic information to the progeny. BVT is one form of Lateral Gene Transfer used specifically for gene transfer between bacteria and vertebrates.