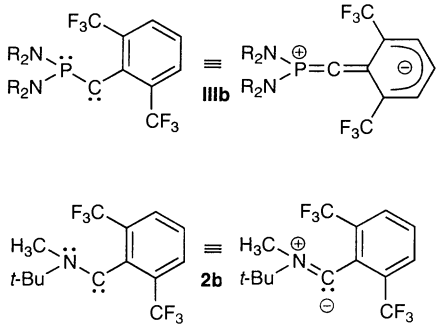
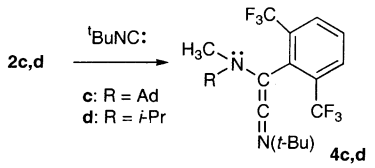


Fig. 3. Best representations for the (phosphino)-(aryl)carbene **IIIb** and the (amino)(aryl)-carbene **2b**.



Scheme 3.

2. Despite the steric protection of the carbene carbon atom, **2c,d** undergo coupling reactions with *tert*-butyl isocyanide at room temperature and afford the corresponding ketenimines **4c,d** in good yields (Scheme 3). Because this reaction, which is typical of transient singlet carbenes (*27*), is not observed for push-push carbenes **II**, we concluded that the isocyanide acts here as a Lewis base toward carbenes **2**. This result demonstrates that, in contrast with **II**, the vacant carbene orbital of **2** remains accessible.

Up to now, the number and variety of stable carbenes have been limited by the perceived necessity for two strongly interacting substituents. Despite this perceived limitation, these species have found applications (*9, 28–33*) even on a large scale. This work establishes that only a single electron-active substituent is necessary to isolate a carbene. Therefore, a broad range of these species will soon be readily available, which will open the way for new synthetic developments and applications in various fields.

References and Notes

1. J. B. Dumas, E. Peligot, *Ann. Chim. Phys.* **58**, 5 (1835).
2. E. Buchner, T. Curtius, *Ber. Dtsch. Chem. Ges.* **8**, 2377 (1885).
3. H. Staudinger, O. Kupfer, *Ber. Dtsch. Chem. Ges.* **45**, 501 (1912).
4. M. Jones, R. A. Moss, Eds., *Carbenes* (Wiley, New York, vols. I and II, 1973 and 1975).
5. M. Regitz, Ed., *Carbene (Carbenoide)*, vol. E19b of *Methoden der Organischen Chemie (Houben-Weyl)* (Thieme, Stuttgart, 1989).
6. U. H. Brinker, Ed., *Advances in Carbene Chemistry* (JAI Press, Greenwich, CT, vols. 1 and 2, 1994 and 1998).
7. F. Z. Dörwald, Ed., *Metal Carbenes in Organic Synthesis* (Wiley, Weinheim, Germany, 1999).
8. H. Tomioka, *Acc. Chem. Res.* **30**, 315 (1997).
9. W. A. Herrmann, C. Köcher, *Angew. Chem. Int. Ed. Engl.* **36**, 2163 (1997).
10. A. J. Arduengo III, *Acc. Chem. Res.* **32**, 913 (1999).
11. D. Bourissou, O. Guerret, F. P. Gabbaï, G. Bertrand, *Chem. Rev.* **100**, 39 (2000).
12. Y. Takahashi *et al.*, *Angew. Chem. Int. Ed. Engl.* **39**, 3478 (2000).
13. A. J. Arduengo III, R. L. Harlow, M. Kline, *J. Am. Chem. Soc.* **113**, 361 (1991).
14. R. W. Alder, C. P. Butts, A. G. Orpen, *J. Am. Chem. Soc.* **120**, 11526 (1998).
15. A. Igau, H. Grützmacher, A. Baceiredo, G. Bertrand, *J. Am. Chem. Soc.* **110**, 6463 (1988).
16. T. Kato *et al.*, *J. Am. Chem. Soc.* **122**, 998 (2000).
17. C. Buron, H. Gornitzka, V. Romanenko, G. Bertrand, *Science* **288**, 834 (2000).
18. L. Pauling, *J. Chem. Soc. Chem. Commun.* **1980**, 688 (1980).
19. Selected spectroscopic data for derivatives **2a-d**, **3a**, and **4c,d** are available at *Science* Online (*33*).
20. Amino-phenyl-carbenes have been postulated as transient species [R. A. Moss, D. P. Cox, H. Tomioka, *Tetrahedron Lett.* **25**, 1023 (1984)].
21. A similar C-H insertion reaction has been reported for the triplet (2,4,6-tri-*tert*-butylphenyl)(phenyl)carbene [K. Hirai, K. Komatsu, H. Tomioka, *Chem. Lett.* **1994**, 503 (1994)].
22. For di(amino)carbenes, the reported C-H insertion reactions initially involve the protonation of the carbene center and thus only occur with strongly polarized C-H bonds [A. J. Arduengo III *et al.*, *Helv. Chim. Acta* **82**, 2348 (1999)].
23. D. L. S. Brahms, W. P. Dailey, *Chem. Rev.* **96**, 1585 (1996).
24. H. Tomioka, K. Taketsuji, *J. Chem. Soc. Chem. Commun.* **1997**, 1745 (1997).
25. Crystal data for **2b**: Cell constants and an orientation matrix for data collection correspond to the orthorhombic space group *Pbcm*, with a = 6.0039(4) Å, b = 19.6483(13) Å, c = 12.4792(8) Å, and V (cell volume) = 1472.1(2) Å^3. A half molecule of $C_{14}H_{15}F_6N$ per asymmetric unit (number of formula units per cell = 4), giving a formula weight of 311.27 and a calculated density (D_c) of 1.404 Mg m^{-3}. The data of the structure were collected on a Bruker-AXS CCD 1000 diffractometer at a temperature of 173(2) K with graphite-monochromated Mo Kα radiation (wavelength = 0.71073 Å) by using phi- and omega-scans. We solved the structure by direct methods, using SHELXS-97 [G. M. Sheldrick, *Acta Crystallogr.* **A46**, 467 (1990)]. The linear absorption coefficient, μ, for Mo K radiation is 0.136 mm^{-1}. The structure was refined with all data on F^2 with a weighting scheme of $w^{-1} = \sigma^2(F_{obs}^2) + (g1 \cdot P)^2 + (g2 \cdot P)$ with $P = (F_{obs}^2 + 2F_{calc}^2)/3$ using *SHELXL-97* [G. M. Sheldrick, *Program for Crystal Structure Refinement* (Universität Göttingen, Göttingen, 1997)] [g is the weighting factor (g_1 = 0.051400 and g_2 = 0.755700) and F_{obs} and F_{calc} are the observed and calculated structure factors, respectively]. All non-H atoms were treated anisotropically. The H atoms were located by difference Fourier maps and refined with a riding model. The final cycle of full-matrix least square refinement was based on 11,530 measured (1472 unique) reflections and 108 variable parameters and converged with the unweighted agreement factor equal to R_1 [$I > 2\sigma(I)$ = 0.0424 (where I is the observed reflection intensity) and the weighted factor wR_2 = 0.1198 for all data. Crystallographic details may be obtained from the corresponding author or from the Cambridge Crystallographic Data Centre, by quoting the full literature citation. Crystallographic details are available at *Science* Online (*34*).
26. It has been predicted that electronegative and electropositive elements such as nitrogen and phosphorus favor the bent and linear conformation, respectively [W. W. Schoeller, *J. Chem. Soc. Chem. Commun.* **1980**, 124 (1980)].
27. A. Halleux, *Angew. Chem. Int. Ed. Engl.* **3**, 752 (1964).
28. D. S. McGuiness, K. J. Cavell, *Organometallics* **19**, 741 (2000).
29. J. Schwarz *et al.*, *Chem. Eur. J.* **6**, 1773 (2000).
30. V. P. W. Böhm *et al.*, *Angew. Chem. Int. Ed. Engl.* **39**, 1602 (2000).
31. C. W. Bielawski, R. H. Grubbs, *Angew. Chem. Int. Ed. Engl.* **39**, 2903 (2000).
32. S. C. Schürer, S. Gessler, N. Buschmann, S. Blechert, *Angew. Chem. Int. Ed. Engl.* **39**, 3898 (2000).
33. J. Louie, R. H. Grubbs, *Angew. Chem. Int. Ed. Engl.* **40**, 247 (2001).
34. For supplementary data, see *Science* Online (www.sciencemag.org/cgi/content/full/292/5523/1901/DC1).

35. We are grateful to the Centre National de la Recherche Scientifique, Rhodia, and the Deutsche Forschungsgemeinschaft for financial support of this work.

9 March 2001; accepted 20 April 2001

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

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

The human genome was analyzed for evidence that genes had been laterally transferred into the genome from prokaryotic organisms. 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. About 40 genes were found to be exclusively shared by humans and bacteria and are candidate examples of horizontal transfer from bacteria to vertebrates. Gene loss combined with sample size effects and evolutionary rate variation provide an alternative, more biologically plausible explanation.

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). 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 organellar genomes into the eukaryotic nucleus (*2*).

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. Such a possibility is of interest because it implies that bacterial infections have led to permanent transfer of genes into their hosts. One possible implication is that bacteria might be manipulating the human genome for their own benefit and that this process may be continuing. 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.

Although the possibility of lateral gene transfer has gained much support in recent years from analysis of complete genome sequences (*1*, *4*, *5*), the inference of such gene transfer events is still fraught with difficulty, because of problems with methods and with the data analyzed (*6*, *7*). 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). 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*).

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

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

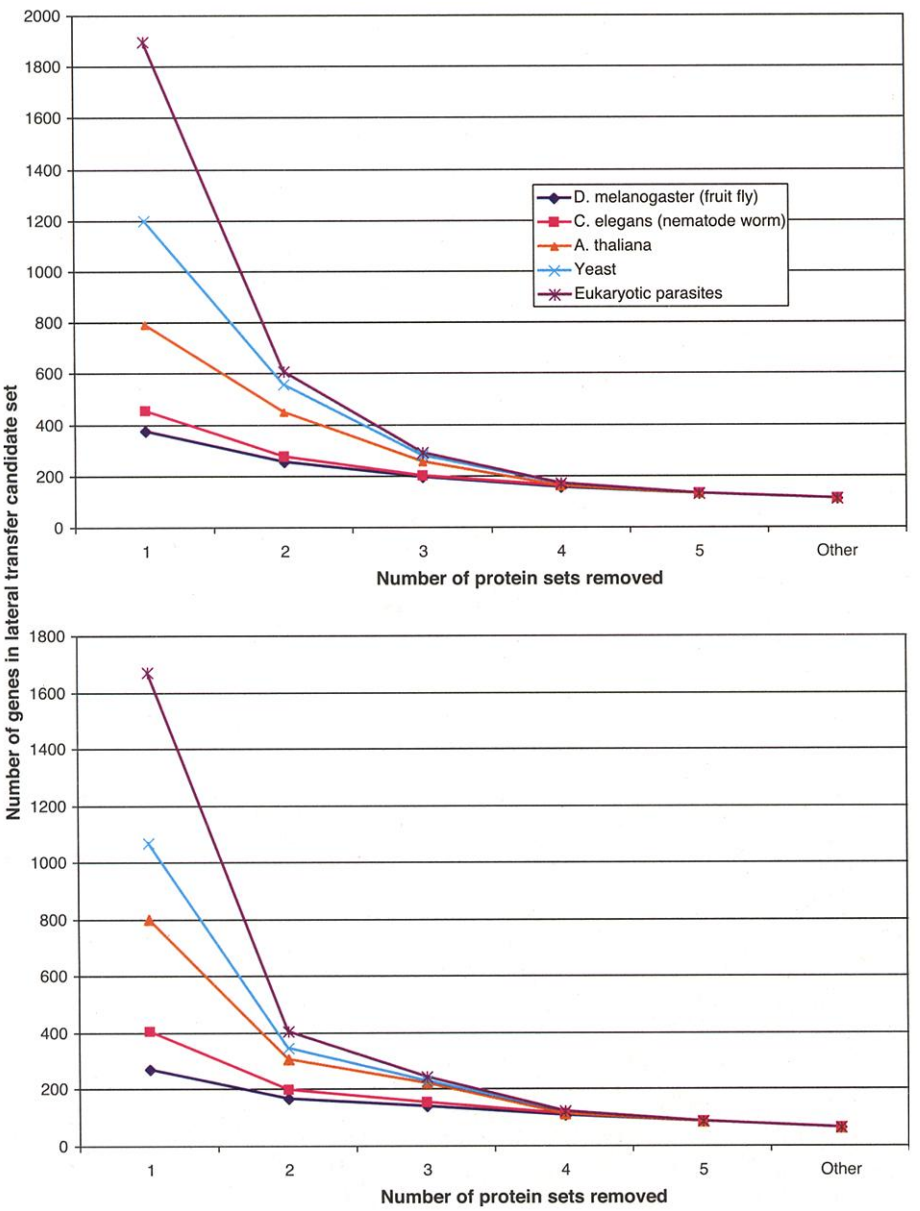


Fig. 1. Genes shared by humans and prokaryotes after removing successive proteome sets from five nonvertebrates and a collection of miscellaneous nonvertebrates ("Other"). (**Top**) Ensembl protein set. (**Bottom**) Celera protein set.

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.

If the pattern of genes shared between prokaryotic and eukaryotic species is a robust measure of lateral gene transfer, then we would expect that the total number of true BVTs would be independent of which and how many nonvertebrate genomes have been sampled. However, as the number of nonvertebrate proteomes screened against human increased, the number of BVTs decreased (Fig. 1). 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.

The downward trend of the plot in Fig. 1 suggests that the number of BVTs might decrease further if more nonvertebrate ge-

nomes 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.

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*). 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.

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 (*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*). It seems likely that the sequencing of a broader variety of eukaryotic genomes will lead to a further reduction in the number of BVTs.

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*). 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. 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. 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.

If a gene was transferred from a prokaryotic lineage into the vertebrate lineage, this likely occurred within the past 400 to 500 million years, after most of the major prokaryotic phyla were established. 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. 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. 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*). The phylogenetic analysis reveals that the vertebrate genes do not branch within any particular prokaryotic lineage. 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.

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. First, not all of these genomes are complete. Second, the annotation of the completed portions of some eukaryotic genomes is still in progress, and the state of the art in eukaryotic gene finding is imperfect. 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. 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.

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

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

eliminated some genes (including possible contaminants), collapsed multiple genes into one, and otherwise improved the data. We screened the 70 BVTs against the newer proteome and found that 23 genes had been eliminated, reducing the BVT set to 47 genes. If the original 135 Ensembl BVTs are screened against the newer release, this set is reduced to 89 genes. There were also 89 genes in the initial Celera BVT set.

Comparing the 47 Ensembl BVTs against the 56 Celera BVTs yields some interesting final reductions in the data set. 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.

After careful reexamination of the human proteome, we find 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*). The evidence presented here provides several plausible biological explanations for the presence of these genes in the human genome. 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. In cases where equally if not more plausible mechanisms exist, extraordinary events such as horizontal gene transfer do not provide the best explanation. 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.

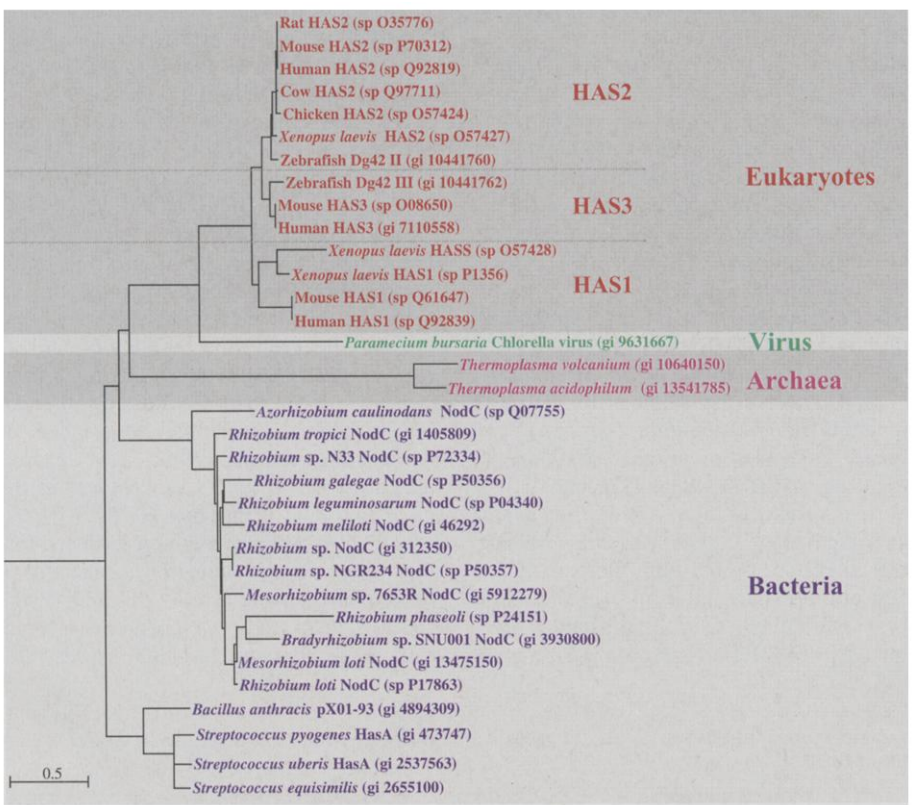


Fig. 2. Phylogenetic tree of homologs of three human hyaluronan synthase (HAS) proteins that were proposed as lateral transfers from bacteria to vertebrates (*3*). Homologs of the human HAS genes were identified with iterative Blastp searches of a low-redundancy protein database and aligned with clustalW. 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.

References and Notes

1. W. F. Doolittle, *Science* **284**, 2124 (1999).
2. W. Martin *et al.*, *Nature* **393**, 162 (1998).
3. International Human Genome Sequencing Consortium, *Nature* **409**, 860 (2001).
4. K. E. Nelson *et al.*, *Nature* **399**, 323 (1999).
5. W. F. Doolittle, *Trends Cell Biol.* **9**, M5 (1999).
6. J. A. Eisen, *Curr. Opin. Genet. Dev.* **10**, 606 (2000).
7. J. P. Gogarten, R. D. Murphey, L. Olendzenski, *Biol. Bull.* **196**, 359 (1999).
8. J. C. Venter *et al.*, *Science* **291**, 1304 (2001).
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).
11. G. M. Rubin *et al.*, *Science* **287**, 2204 (2000).
12. K. E. Nelson, I. T. Paulsen, J. F. Heidelberg, C. M. Fraser, *Nature Biotechnol.* **18**, 1049 (2000).
13. W. H. Li, *Molecular Evolution* (Sinauer Associates, Sunderland, MA, 1997).
14. S. F. Altschul, *J. Mol. Evol.* **36**, 290 (1993).
15. J. A. Eisen, *Genome Res.* **8**, 163 (1998).
16. J. D. Palmer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6960 (2000).
17. S. G. Andersson *et al.*, *Nature* **396**, 133 (1998).
18. X. Lin *et al.*, *Nature* **402**, 761 (1999).
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).
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).
24. W. Gish, D. J. States, *Nature Genet.* **3**, 266 (1993).
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.