EXTENDED PDF FORMAT
SPONSORED BY

BIO-RAD

http://discover.bio-rad.com



Whole-Genome Analysis of Photosynthetic Prokaryotes

Jason Raymond, *et al.*

Science **298**, 1616 (2002);
DOI: 10.1126/science.1075558

The following resources related to this article are available online at www.sciencemag.org (this information is current as of December 6, 2006):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/cgi/content/full/298/5598/1616

Supporting Online Material can be found at:
http://www.sciencemag.org/cgi/content/full/298/5598/1616/DC1

A list of selected additional articles on the Science Web sites **related to this article** can be found at:
http://www.sciencemag.org/cgi/content/full/298/5598/1616#related-content

This article **cites 22 articles**, 10 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/298/5598/1616#otherarticles

This article has been **cited by** 68 article(s) on the ISI Web of Science.

This article has been **cited by** 19 articles hosted by HighWire Press; see:
http://www.sciencemag.org/cgi/content/full/298/5598/1616#otherarticles

This article appears in the following **subject collections**:
Genetics
http://www.sciencemag.org/cgi/collection/genetics

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:
http://www.sciencemag.org/help/about/permissions.dtl

Downloaded from www.sciencemag.org on December 6, 2006

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright c 2005 by the American Association for the Advancement of Science; all rights reserved. The title SCIENCE is a registered trademark of AAAS.

eages could be surviving in some unsurveyed modern Native American breeds or local dog populations (*14*, *25*). However, genetic analysis of a diverse sample of 19 Mexican hairless dogs (xoloitzcunitle), a distinct ancient breed that has been present in Mexico for over 2000 years (*25*), only revealed mtDNA sequences previously observed in dogs of Eurasian origin (*26*). The absence of ancient North and South American dog haplotypes from a large diversity of modern breeds, including the Mexican hairless, illustrates the considerable impact that invading Europeans had on native cultures.

Our data strongly support the hypothesis that ancient American and Eurasian domestic dogs share a common origin from Old World gray wolves. This implies that the humans who colonized America 12,000 to 14,000 yr B.P. brought multiple lineages of domesticated dogs with them. The large diversity of mtDNA lineages in the dogs that colonized the New World implies that the ancestral population of dogs in Eurasia was large and well mixed at that time. Consequently, dogs, in association with humans or through trade, spread across Europe, Asia, and the New World soon after they were domesticated. Alternatively, if domestication was a more ancient event, as suggested by previous genetic results (*5*), human groups that first colonized the subarctic mammoth steppe of Siberia may have had dogs with them 26,000 to 19,000 yr B.P. (*11*). If the archaeological date of 12,000 to 14,000 yr B.P. for first domestication is accepted, the dog, as an element of culture, would have had to be transmitted across Paleolithic societies on three continents in a few thousand years or less. This would imply extensive intercultural exchange during the Paleolithic (*27*, *28*). Regardless, the common origin of New and Old World dogs demands a reconsideration of the relationship between humans and dogs in ancient societies.

References and Notes

1. S. J. Olsen, *Origins of the Domestic Dog: The Fossil Record* (Univ. of Arizona Press, Tucson, AZ, 1985).
2. D. F. Morey, *Am. Sci.* **82**, 336 (1994).
3. J. P. Scott, J. L. Fuller, *Dog Behavior: The Genetic Basis* (Univ. of Chicago Press, Chicago, IL, 1965)
4. R. K. Wayne, S. J. O'Brien, *Syst. Zool.* **36**, 457 (1987).
5. C. Vilà *et al.*, *Science* **276**, 1687 (1997).
6. B. F. Koop, M. Burbidge, A. Byun, U. Rink, S. J. Crockford, in *Dogs Through Time: An Archaeological Perspective*, S. J. Crockford, Ed. (British Archaeological Reports, Oxford, 2000), pp. 271–285. Two of the five ancient remains studied were dated to 1938 and 1940; the rest were designated as "prehistoric" by the authors.
7. D. K. Grayson, *Anthropol. Pap. Am. Mus. Nat. Hist.* **66** (1988).
8. G. Nobis, *Umschau* **19**, 610 (1979).
9. T. Dayan, *J. Archaeol. Sci.* **21**, 633 (1994).
10. J. Clutton-Brock, *A Natural History of Domesticated Mammals* (Cambridge Univ. Press, Cambridge, ed. 2, 1999).
11. T. Goebel, *Evol. Anthropol.* **8**, 208 (1999).
12. S. J. Fiedel, *J. Archaeol. Res.* **8**, 39 (2000).
13. M. Schwartz, *A History of Dogs in the Early Americas* (Yale Univ. Press, New Haven, CT, 1997).
14. R. Valadez, G. Mestre, *Historia del Xoloitzcuintle en México* (Museo Dolores Olmedo-UNAM, México, 1999).
15. Supporting material is available on *Science* Online.
16. C. Vilà *et al.*, *Mol. Ecol.* **8**, 2089 (1999).
17. The 350 modern dog sequences are from over 250 dogs corresponding to 124 haplotypes with sequences deposited in GenBank, and 100 dogs from 20 different breeds (C. Vilà, data not shown).
18. I. Barnes, P. Matheus, B. Shapiro, D. Jensen, A. Cooper, *Science* **295**, 2267 (2002).
19. J. C. Avise, *Phylogeography. The History and Formation of Species* (Harvard Univ. Press, Cambridge, MA, 2000).
20. J. Clutton-Brock, A. C. Kitchener, J. M. Lynch, *J. Zool.* **233**, 19 (1994).
21. R. Valadez, *et al. Asoci. Mexi. Med. Vet. Especi. Pequeñas Especies* **13**, 6 (2002).
22. C. Vilà, R. K. Wayne, *Conserv. Biol.* **13**, 195 (1999).
23. E. Randi, V. Lucchini, *Conserv. Genet.* **3**, 31 (2002).
24. J. H. Zarr, *Biostatistical Analysis* (Prentice Hall, Upper Saddle River, NJ, ed. 4, 1999); the calculation is made according to example 24.5.
25. R. Valadez, in *Dogs Through Time: An Archaeological Perspective*, S. J. Crockford, Ed. (British Archaeological Reports, Oxford, 2000), pp. 193–204.
26. C. Vilà, J. E. Maldonado, R. K. Wayne, *J. Hered.* **90**, 71 (1999).
27. C. Gamble, *The Palaeolithic Societies of Europe* (Cambridge Univ. Press, Cambridge, 1999).
28. A. P. Derev'anko, in *The Paleolithic of Siberia, New Discoveries and Interpretations*, A. P. Derev'anko, D. B. Shimkin, W. R. Powers, Eds. (Univ. of Illinois Press, Chicago, IL, 1998), pp. 5–13.
29. M. Stuiver, G. W. Pearson, *Radiocarbon* **35**, 1 (1993).

30. We thank R. Tedford (American Museum of Natural History, New York) and W. Isbell (Department of Anthropology, State University of New York at Binghamton) for samples. This research was supported by grants from the University of California Institute for Mexico and the United States and NSF (grant OPP-9817937). We thank C. Anderung, J. Brantingham, A. Götherström, B. VanValkenburgh, and M. Zeder for comments on the manuscript.

Supporting Online Material

www.sciencemag.org/cgi/content/full/298/5598/1613/DC1

Materials and Methods

References

5 August 2002; accepted 26 September 2002

Whole-Genome Analysis of Photosynthetic Prokaryotes

Jason Raymond,[1]* Olga Zhaxybayeva,[2]* J. Peter Gogarten,[2] Sveta Y. Gerdes,[3] Robert E. Blankenship[1]†

The process of photosynthesis has had profound global-scale effects on Earth; however, its origin and evolution remain enigmatic. Here we report a whole-genome comparison of representatives from all five groups of photosynthetic prokaryotes and show that horizontal gene transfer has been pivotal in their evolution. Excluding a small number of orthologs that show congruent phylogenies, the genomes of these organisms represent mosaics of genes with very different evolutionary histories. We have also analyzed a subset of "photosynthesis-specific" genes that were elucidated through a differential genome comparison. Our results explain incoherencies in previous data-limited phylogenetic analyses of phototrophic bacteria and indicate that the core components of photosynthesis have been subject to lateral transfer.

Photosynthesis is an essential biological process in which solar energy is transduced into other forms of energy that are available to all life. Primary production by photosynthetic organisms supports all ecosystems, with the noted exceptions of deep-sea hydrothermal vents and subsurface communities. Oxygen, one of the by-products of photosynthesis by cyanobacteria and their descendants (including algae and higher plants), transformed the Precambrian Earth and made possible the development of more complex organisms that use aerobic metabolism (*1*, *2*). Understanding the origin and evolution of the process of photosynthesis is, therefore, of considerable interest.

All available evidence suggests that (bacterio)chlorophyll-based photosynthesis arose within the bacterial domain of the tree of life and was followed by subsequent endosymbiotic transfer into eukaryotes. Accurate dates for appearance of the first photosynthetic organisms are not known. Substantial information, including biomarkers, stromatolites, and paleosols, as well as data from molecular evolution studies, indicates that oxygenic (oxygen-evolving) photosynthesis arose by 2500 million years ago (*2*–*5*). On the basis of phylogenetic analyses and the well-detailed complexity of the photosynthetic machinery, mechanistically simpler anoxygenic (non–oxygen-evolving) photosynthesis almost certainly preceded and was ancestral to oxygenic photosynthesis (*1*, *6*). Therefore the cyanobacteria, as ancient as they appear to be, were probably preceded by a diverse group of more primitive phototrophs. The supposed progeny of those early phototrophs are still

[1]Department of Chemistry and Biochemistry, Arizona State University (ASU), Tempe, AZ 85287–1604, USA. [2]Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06269–3044, USA. [3]Integrated Genomics, 2201 West Campbell Park Drive, Chicago, IL 60612, USA.

*These authors contributed equally to this paper.
†To whom correspondence should be addressed. E-mail: blankenship@asu.edu

Downloaded from www.sciencemag.org on December 6, 2006

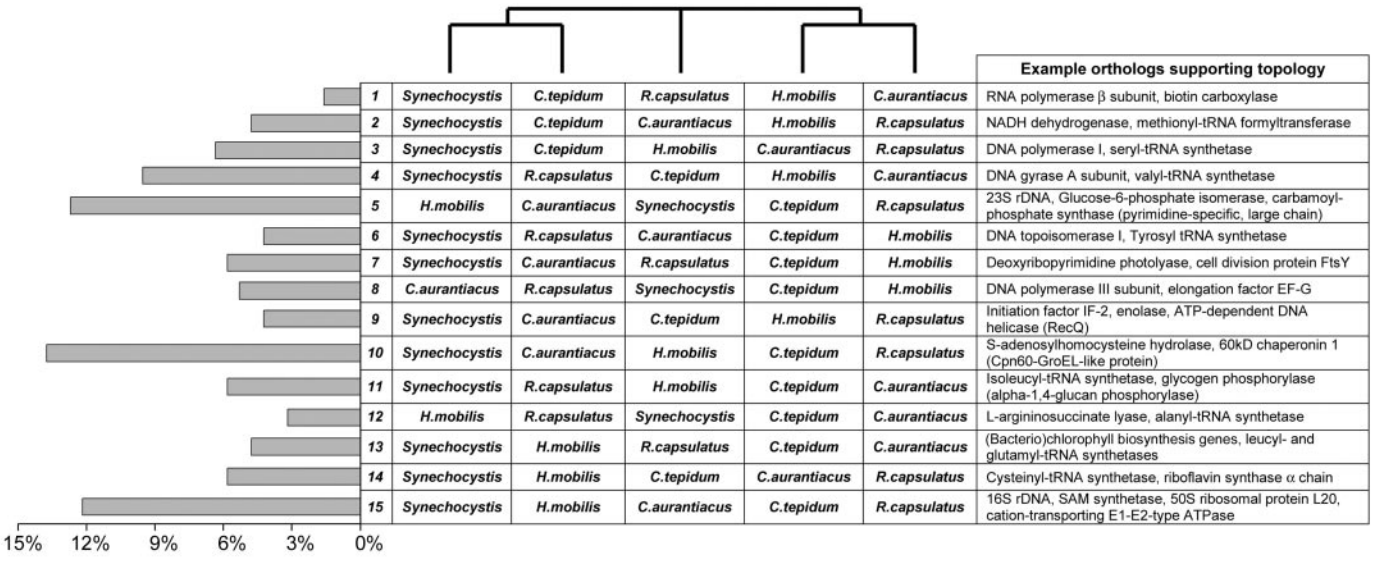
found throughout diverse ecosystems and may provide key evidence toward unraveling the early origins of photosynthesis.

There are five known bacterial phyla with photosynthetic members. These phyla are widely distributed within the bacterial domain and include the cyanobacteria (the only oxygenic group), proteobacteria (purple bacteria), green sulfur bacteria, green filamentous bacteria, and the Gram-positive heliobacteria. With respect to traditional ribosomal-based phylogenies, the distribution of photosynthesis is markedly paraphyletic (*7*, *8*). There have been a number of different hypotheses proposed to resolve the disparate phylogenetic distribution of these organisms (*6*, *9*–*11*). However, in the absence of conclusive data, none of these proposals has won unanimous acceptance. On the basis of genomic comparisons presented here, we propose that horizontal gene flow has played a major role in the evolution of bacterial phototrophs and that many of the essential components of photosynthesis have been among these horizontally transferred genes.

A crucial early step of any sequence-based analysis is the selection of genes for phylogenetic comparison, which should minimize the inclusion of potentially error-causing paralogs or nonhomologous genes. Here this was done by carrying out whole-genome BLAST comparisons of all proteins for every possible pairing of organisms that make up the sample. Putative orthologs were required to have BLAST scores with expectation values for chance similarity below a preset threshold. Sets of orthologous sequences were then compiled from genes that are reciprocal best BLAST hits across all of the genomes compared, therefore, given a set of orthologs from each of the five genomes, each individual ortholog returns all of the other four as a top-scoring BLAST hit when searching that particular genome (*12*). These computationally intensive procedures aim to avoid the erroneous results that can arise from comparing paralogous or nonhomologous genes [for methodology, application, and further discussion, see (*13*–*15*)]. Even with these rigorous ortholog selection requirements, we were able to perform phylogenetic analyses on nearly 200 sets of orthologous genes, providing a previously unattainable look into the early evolution of photosynthetic organisms.

With the use of the above methods, we found a total of 188 orthologs common to the genomes of *Synechocystis* sp. PCC6803 (cyanobacteria), *Chloroflexus aurantiacus* (green filamentous bacteria), *Chlorobium tepidum* (green sulfur bacteria), *Rhodobacter capsulatus* (proteobacteria), and *Heliobacillus mobilis* (heliobacteria). These genes encompass a broad range of functions, including housekeeping genes involved in protein synthesis, DNA replication and transcription, and manufacture of structural components of the cell, as well as the genetic components of various metabolic or biosynthetic pathways common to all the organisms. We individually evaluated each set of orthologs using maximum likelihood to determine which of the 15 possible five-taxa tree topologies provided the best fit to the observed sequence data. Posterior probabilities were calculated from log likelihood values with the use of an approach developed by Strimmer and von Haesler (*16*). Figure 1 shows all 15 possible topologies as well as the percentage of the 188 sets of protein-coding genes for which the given topology was the most probable. Also shown in Fig. 1 are example functional annotations, some of which are frequent choices for phylogenetic inference, listed by their corresponding topology those genes supported. The most unexpected result from this analysis is the distinct lack of unanimous support for a single topology. Plurality support is seen for the three trees (5, 10, and 15) that group together *Synechocystis* sp., *C. aurantiacus*, and *H. mobilis* separate from a distinct *R. capsulatus* and *C. tepidum* cluster. The data suggest that even strongly supported phylogenies and highly conserved genes from these organisms often show very different evolutionary histories.

Orthologs from each data set were further stratified by their putative functional assignments on the basis of cluster of orthologous groups (COG) categories (*12*, *14*, *17*) (fig. S1, table S4). It might have been expected that, for example, genes functioning in information processing would as a subset show preference for a single topology (*18*). However, the results indicate that even at this level of grouping-by-function no unanimous support for a particular topology is seen. Additionally, because branch length information is necessarily disregarded when segregating orthologs by most likely topology, we reexamined branch lengths for every tree constructed and tabulated distances determined by maximum likelihood analysis of the individual sets of orthologous genes. This step incorporated another level of stringency into the overall analysis, because potentially error-causing cases in which one or more orthologs displayed anomalously long branch lengths could be recognized and eliminated. We ob-



						Example orthologs supporting topology
1	Synechocystis	C.tepidum	R.capsulatus	H.mobilis	C.aurantiacus	RNA polymerase β subunit, biotin carboxylase
2	Synechocystis	C.tepidum	C.aurantiacus	H.mobilis	R.capsulatus	NADH dehydrogenase, methionyl-tRNA formyltransferase
3	Synechocystis	C.tepidum	H.mobilis	C.aurantiacus	R.capsulatus	DNA polymerase I, seryl-tRNA synthetase
4	Synechocystis	R.capsulatus	C.tepidum	H.mobilis	C.aurantiacus	DNA gyrase A subunit, valyl-tRNA synthetase
5	H.mobilis	C.aurantiacus	Synechocystis	C.tepidum	R.capsulatus	23S rDNA, Glucose-6-phosphate isomerase, carbamoyl-phosphate synthase (pyrimidine-specific, large chain)
6	Synechocystis	R.capsulatus	C.aurantiacus	C.tepidum	H.mobilis	DNA topoisomerase I, Tyrosyl tRNA synthetase
7	Synechocystis	C.aurantiacus	R.capsulatus	C.tepidum	H.mobilis	Deoxyribopyrimidine photolyase, cell division protein FtsY
8	C.aurantiacus	R.capsulatus	Synechocystis	C.tepidum	H.mobilis	DNA polymerase III subunit, elongation factor EF-G
9	Synechocystis	C.aurantiacus	C.tepidum	H.mobilis	R.capsulatus	Initiation factor IF-2, enolase, ATP-dependent DNA helicase (RecQ)
10	Synechocystis	C.aurantiacus	H.mobilis	C.tepidum	R.capsulatus	S-adenosylhomocysteine hydrolase, 60kD chaperonin 1 (Cpn60-GroEL-like protein)
11	Synechocystis	R.capsulatus	H.mobilis	C.tepidum	C.aurantiacus	Isoleucyl-tRNA synthetase, glycogen phosphorylase (alpha-1,4-glucan phosphorylase)
12	H.mobilis	R.capsulatus	Synechocystis	C.tepidum	C.aurantiacus	L-argininosuccinate lyase, alanyl-tRNA synthetase
13	Synechocystis	H.mobilis	R.capsulatus	C.tepidum	C.aurantiacus	(Bacterio)chlorophyll biosynthesis genes, leucyl- and glutamyl-tRNA synthetases
14	Synechocystis	H.mobilis	C.tepidum	C.aurantiacus	R.capsulatus	Cysteinyl-tRNA synthetase, riboflavin synthase α chain
15	Synechocystis	H.mobilis	C.aurantiacus	C.tepidum	R.capsulatus	16S rDNA, SAM synthetase, 50S ribosomal protein L20, cation-transporting E1-E2-type ATPase

Fig. 1. Distribution of orthologs among the 15 possible unrooted trees. The tree at top gives branching order for the photosynthetic organisms listed in the center grid for each of the 15 possible five-taxa trees. Bars show the percentage of 188 sets of orthologs that chose a particular tree topology as most likely. Examples of genes supporting each topology, based on *Synechocystis* annotations, are shown at right and include 16S and 23S trees constructed from ribosomal DNA sequences from these genomes.

Downloaded from www.sciencemag.org on December 6, 2006

served a positive correlation between overall number of substitutions per site and posterior probability score for the most likely tree, indicating that genes that are less diverged are more likely to map to an explicit topology (*19*). The shortest between-taxa distances were recovered from each 5 × 5 pairwise distance matrix generated during phylogenetic reconstruction. In 117 cases, the shortest between-taxa distance favored clustering one of the three possible pairings of *H. mobilis*, *Synechocystis*, and *Chloroflexus*, whereas the *C. tepidum–R. capsulatus* cluster was favored in only 8 cases. Overall averaged estimates of substitutions per site corroborate these findings, with the lowest number of substitutions per site between *Chloroflexus* and *H. mobilis*, followed by *Synechocystis* and *H. mobilis*. Averaged substitutions per site for the *C. tepidum–R. capsulatus* grouping were second highest overall. These results imply an overall close relationship between *H. mobilis*, *Synechocystis*, and *Chloroflexus* (though the relationship between the latter two is not as strong, on average) and reveal that the *C. tepidum–R. capsulatus* grouping that is frequently observed when unrooted topologies are considered becomes less relevant when estimated distances between these two organisms are taken into account.

Subsequently, we set out to identify genes that play an essential role in phototrophy and whose evolution might be tightly linked to the advent and development of photosynthesis. The biochemical machinery comprising the cogwheels of photosynthesis has been continually refined over billions of years since the emergence of the first bacterial phototrophs. In some notable cases, genes within this process have originated from nonphotosynthetic genes that were incorporated by various genetic processes, including gene recruitment, gene duplication and fusion, and possibly motif shuffling (*6*, *9*). In other cases, gene origins have been masked by eons of evolution at the primary sequence level, so some homologs are detected only in other photosynthetic organisms. These so-called "photosynthesis-specific" (PS-specific) genes emerge as an obvious focus of interest in attempting to understand the evolution of photosynthesis; however, it remains unclear how extensive the set of PS-specific genes is. Therefore, we have constructed a simple method for finding members of this group.

Finding PS-specific genes can be approximated by finding all genes shared within the subset of photosynthetic organisms and then subtracting from this set those genes found in nonphotosynthetic organisms (*12*). In principle, this method for identification of pathway-specific genes can be applied to other groups of organisms whose genomes have been sequenced, giving a differential compar-

Table 1. Putative function and pathway or functional category of PS-specific and PS-related genes, and number of genomes each gene is found in (tables S1 to S4 and fig. S1). Main PS includes the five photosynthetic lineages compared in the text, other PS includes six additional phototrophic bacteria, and non-PS includes 50 nonphotosynthetic organisms. Question marks indicate unidentified functional categories.

Putative function	Main PS	Other PS	Non-PS	Pathway/functional category	GenBank accession
		PS-specific			
Mg-protoporphyrin-O-methyltransferase BchM	5	6	0	(Bacterio)chlorophyll biosynthesis	slr0525
Protochlorophyllide reductase BchB subunit	5	6	0	(Bacterio)chlorophyll biosynthesis	slr0772
Protochlorophyllide reductase BchN subunit	5	6	0	(Bacterio)chlorophyll biosynthesis	slr0750
Protochlorophyllide reductase BchL subunit	5	6	0	(Bacterio)chlorophyll biosynthesis	slr0749
Mg chelatase subunit BchH	5	6	0	(Bacterio)chlorophyll biosynthesis	slr1055
		PS-related			
Short chain alcohol dehydrogenase family	5	1	3	Oxidoreductase	slr1208
Phytoene desaturase	5	6	4	Carotenoid biosynthesis	slr1293
Mg chelatase subunit BchI	5	6	6	(Bacterio)chlorophyll biosynthesis	slr1030
Putative restriction endonuclease	5	6	6	Nucleic acid metabolism	sll1193
CbiM protein	5	2	6	Cobalamin/B_{12} biosynthesis	sll0383
CobN protein	5	6	7	Cobalamin/B_{12} biosynthesis	slr1211
Mg-protoporphyrin IX oxidative cyclase BchE	5	1	7	(Bacterio)chlorophyll biosynthesis	slr0905
Bacteriochlorophyll synthase BchG	5	6	8	(Bacterio)chlorophyll biosynthesis	sll1091
Hypothetical protein	5	3	8	?	sll1039
5′-methylthioadenosine phosphorylase	5	6	10	?	sll0135
Precorrin-8X methylmutase	5	6	10	Cobalamin/B_{12} biosynthesis	slr1467
Membrane protein	5	4	10	?	sll0932
Pyruvate:flavodoxin oxidoreductase	5	4	10	Oxidoreductase	sll0741
Hydrogenase expression/formation protein HypB	5	4	11	Urease/hydrogenase associated	sll1079
Precorrin-2 C20-methyltransferase	5	5	11	Cobalamin/B_{12} biosynthesis	slr1879
Cobalamin kinase/Cobalamin Pi guanylyltransferase	5	6	13	Cobalamin/B_{12} biosynthesis	slr0216
Precorrin-4 C11-methyltransferase	5	6	13	Cobalamin/B_{12} biosynthesis	slr0969
Precorrin-6Y C5,15-methyltransferase	5	5	13	Cobalamin/B_{12} biosynthesis	sll0099
Hydrogenase maturation protein HypF	5	3	13	Urease/hydrogenase associated	sll0322
Hypothetical protein	5	6	14	?	slr0427
Exopolyphosphatase	5	4	14	Phosphorus compounds	sll1546
Possible photoperiodic protein	4	6	0	?	sll1874
Chlorophyllide reductase 35.5-kD chain BchX	4	1	0	(Bacterio)chlorophyll biosynthesis	NP_662309
L-asparaginase II	4	5	2	?	sll0585
Ribulose bisphosphate carboxylase large chain	4	6	3	Carbon fixation	slr0009
Putative thiol-disulfide isomerase/thioredoxin	4	2	4	?	NP_662608
Hypothetical cytosolic protein	4	6	7	?	slr0351
SufE protein probably involved in FeS center assembly	4	6	8	?	sll0585
Phosphoenolpyruvate synthase/Pyruvate phosphate dikinase	4	1	9	Carbon fixation	NP_662565
Putative nucleotide-binding protein	4	1	10	?	NP_661693
Membrane lipoprotein precursor	4	0	13	?	NP_661927

Downloaded from www.sciencemag.org on December 6, 2006

ison between organisms that share a pathway and those that are missing it. Although there are obvious cases where this method will result in false negatives due to organism-specific photosynthetic proteins, even this first-order approach gives some interesting insights.

In performing this analysis on the above set of five photosynthetic genomes and a group of six taxonomically diverse, nonphotosynthetic bacteria and archaea, we found only a small set of PS-specific proteins (Fig. 2) (tables S1 to S4). Relaxing our constraints to include putative "photosynthesis-related proteins" (PS-related) — defined as missing in no more than one of the photosynthetic genomes or present in no more than one of the nonphotosynthetic genomes — notably increases the size of this set with the caveat of potentially increasing the number of false positives. Genes found in all 11 bacterial and archaeal genomes are predominantly housekeeping genes that function in nucleic and amino acid transport and metabolism as well as in translation and ribosomal structure (but not in transcription or DNA replication). PS-specific and PS-related genes function primarily in energy production (*12*). However, no single majority topology was observed in the phylogenetic trees from either of these functional subsets.

A second, more exhaustive method was then undertaken in which we compared the five photosynthetic organisms to an additional six photosynthetic and 50 nonphotosynthetic organisms from publicly available genome projects (Table 1). This comparison did not require a single key organism (such as *Synechocystis*) as with the above analysis, but rather it found homologous genes and gene families from the overlap and differences of a large set of photosynthetic and nonphotosynthetic genomes (*12*). Homologs found in this extensive analysis corroborate most of the findings from the restricted data set, and add several significant hits to the overall list and subtract some false positives. The function and topology supported by several genes at the top of these lists are congruent with recent phylogenetic analysis of pigment biosynthesis genes (*6*), though they differ from the ribosomal-based organismal phylogenies and plurality topologies in Fig. 1. These results bolster the idea that the evolution of photosynthetic genes has been disconnected from divergence and speciation in these organisms, confirming the extensive role that horizontal gene flow has played in prokaryote evolution. An additional caveat is that many genes from the PS-related set are either hypothetical or completely unknown, complicating attempts to understand the context under which many of these genes have evolved and making them candidates for further analysis. One possibility is that some elements of the photosynthetic apparatus, or factors involved in its assembly or stability, remain unknown.

Previous phylogenetic analyses of photosynthetic bacteria have necessarily used a limited subset of genes to infer relationships among these organisms, often resulting in incongruent results (*6*, *7*, *10*, *11*). New whole-genome data have allowed us to make an extensive comparison of representatives of each of the five known groups of photosynthetic bacteria and may help to reconcile multiple lines of disparate phylogenetic evidence centered on them. In line with other recent whole-genome analyses, horizontal gene transfer (HGT) appears to be an integral aspect of prokaryote evolution (*20*–*23*), and genetic components of the photosynthetic apparatus have crossed species lines nonvertically. Rather than confounding the overall picture, as is often the case in data-limited studies where HGT is apparent, in the context of whole genome comparisons HGT can further refine and resolve the history of an organism. For example, multiple lines of phylogenetic evidence, supported in part by our analysis, have placed the Gram-positive firmicutes, which include *H. mobilis*, as a sister phylum to the cyanobacteria (*8*, *15*, *24*). However, the close relationship of either of these groups with *Chloroflexus* has not previously been noted. The placement of *Chloroflexus* at the base of the bacterial radiation using 16S ribosomal RNA has been the basis for its designation as the earliest phototroph (*7*, *25*). Taking into consideration our results that indicate extensive lateral gene transfer raises the possibility that *Chloroflexus* has acquired phototrophy, perhaps largely through lateral gene transfer. This idea is bolstered by the close phylogenetic and, to a lesser degree, phenotypic relatedness of *Chloroflexus* and *Chlorobium*, evident in their highly similar pigment biosynthesis genes and light-harvesting chlorosome structures. In contrast, other components of these two bacteria, including the photosynthetic reaction centers, are markedly different; thus, other components might have been inherited vertically or through HGT from other phototrophs. These ideas suggest further tests of estimating times of divergence and lateral gene transfer for these and the other photosynthetic bacteria compared here. For all the demonstrated evolutionary complexity and antiquity of these bacteria, mapping the early events in the evolution and distribution of photosynthesis stands as a formidable but exciting challenge.

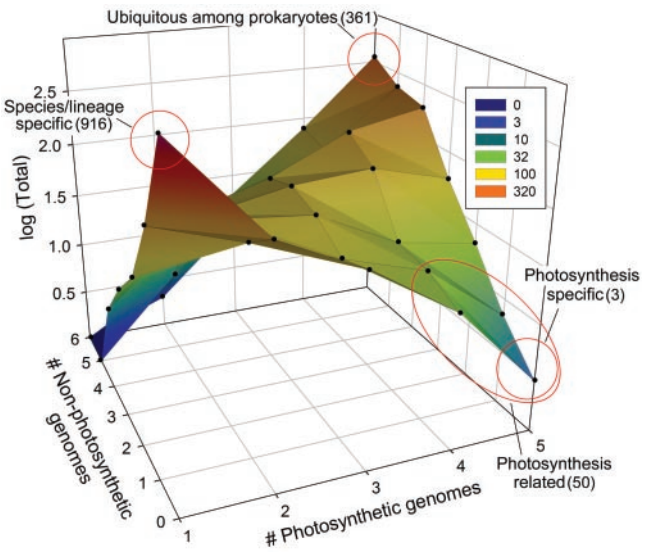


Fig. 2. Distribution of 3169 genes from *Synechocystis* by occurrence in five photosynthetic and six nonphotosynthetic bacterial and archaeal genomes, ranging from genes present in all 11 genomes to those only found in *Synechocystis*. Proposed categories are circled in red, and number of genes in each proposed category is shown in parentheses.

References and Notes

1. R. E. Blankenship, *Molecular Mechanisms of Photosynthesis* (Blackwell Science, Oxford, 2002).
2. J. Xiong, C. E. Bauer, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **53**, 503 (2002).
3. J. J. Brocks, G. A. Logan, R. Buick, R. E. Summons, *Science* **285**, 1033 (1999).
4. R. Rye, H. D. Holland, *Am. J. Sci.* **298**, 621 (1998).
5. S. B. Hedges *et al.*, *BMC Evol. Biol.* **1**, 4 (2001).
6. J. Xiong, W. M. Fischer, K. Inoue, M. Nakahara, C. E. Bauer, *Science* **289**, 1724 (2000).
7. C. R. Woese, *Microbiol. Rev.* **51**, 221 (1987).
8. B. L. Maidak *et al.*, *Nucleic Acids Res.* **29**, 173 (2001).
9. D. H. Burke, J. E. Hearst, A. Sidow, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7134 (1993).
10. R. S. Gupta, T. Mukhtar, B. Singh, *Mol. Microbiol.* **32**, 893 (1999).
11. F. Baymann, M. Brugna, U. Muhlenhoff, W. Nitschke, *Biochim. Biophys. Acta* **1507**, 291 (2001).
12. Materials and Methods are available as supporting material on *Science* Online.
13. J. P. Gogarten, L. Olendzenski, *Curr. Opin. Genet. Dev.* **9**, 630 (1999).
14. R. L. Tatusov, E. V. Koonin, D. J. Lipman, *Science* **278**, 631 (1997).
15. O. Zhaxybayeva, J. P. Gogarten, *BMC Genomics* **3**, 4 (2002).
16. K. Strimmer, A. von Haeseler, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6815 (1997).
17. R. L. Tatusov, M. Y. Galperin, D. A. Natale, E. V. Koonin, *Nucleic Acids Res.* **28**, 33 (2000).
18. R. Jain, M. C. Rivera, J. A. Lake, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3801 (1999).
19. J. Raymond, O. Zhaxybayeva, J. P. Gogarten, S. Y. Gerdes, R. E. Blankenship, unpublished data.
20. L. Aravind, R. L. Tatusov, Y. I. Wolf, D. R. Walker, E. V. Koonin, *Trends Genet.* **14**, 442 (1998).
21. H. Ochman, J. G. Lawrence, E. A. Groisman, *Nature* **405**, 299 (2000).
22. J. F. Heidelberg *et al.*, *Nature* **406**, 477 (2000).

Downloaded from www.sciencemag.org on December 6, 2006

23. E. V. Koonin, K. S. Makarova, L. Aravind, *Ann. Rev. Microbiol.* **55**, 709 (2001).
24. W. F. Vermaas, *Photosynth. Res.* **41**, 285 (1994).
25. H. Oyaizu, B. Debrunner-Vossbrinck, L. Mandelco, J. A. Studier, C. R. Woese, *Syst. Appl. Microbiol.* **9**, 47 (1987).
26. We thank L. Olendzenski for critically reading the manuscript. H. Hartman for stimulating discussions, and R. Overbeek for help with analysis using the ERGO database. The NASA Exobiology program and the NASA Astrobiology Institute at ASU are gratefully acknowledged for support. This is publication no. 559 from the ASU center for the Early Events in Photosynthesis. Sequence data for *C. tepidum* was obtained from The Institute for Genomic Research through the website at www.tigr.org. Sequencing of *C. tepidum* was accomplished with support from the DOE. The data for *C. aurantiacus* has been provided freely by the DOE Joint Genome Institute.

Supporting Online Material
www.sciencemag.org/cgi/content/full/298/5598/1616/DC1
Materials and Methods
Fig. S1
Tables S1 to S4

28 June 2002; accepted 17 September 2002

Extent of Chromatin Spreading Determined by *roX* RNA Recruitment of MSL Proteins

Yongkyu Park,[1] Richard L. Kelley,[2] Hyangyee Oh,[3] Mitzi I. Kuroda,[1,2,3]* Victoria H. Meller[4]*

The untranslated *roX1* and *roX2* RNAs are components of the *Drosophila* male-specific lethal (MSL) complex, which modifies histones to up-regulate transcription of the male X chromosome. *roX* genes are normally located on the X chromosome, and *roX* transgenes can misdirect the dosage compensation machinery to spread locally on other chromosomes. Here we define MSL protein abundance as a determinant of whether the MSL complex will spread in cis from an autosomal *roX* transgene. The number of expressed *roX* genes in a nucleus was inversely correlated with spreading from *roX* transgenes. We suggest a model in which MSL proteins assemble into active complexes by binding nascent *roX* transcripts. When MSL protein/*roX* RNA ratios are high, assembly will be efficient, and complexes may be completed while still tethered to the DNA template. We propose that this local production of MSL complexes determines the extent of spreading into flanking chromatin.

A key mechanism for regulating eukaryotic gene expression is alteration of DNA packaging into chromatin (*1*). Modified chromatin architecture can sometimes be propagated long distances in cis from an initiation point (*2–6*), but the mechanism of such spreading is not understood. The MSL dosage compensation complex is thought to spread along the single male X chromosome in *Drosophila* (*7*). The MSL complex is composed of at least six proteins and two noncoding *roX* RNAs that paint the male X chromosome, leading to covalent modification of the NH_2-terminal tails of histones H3 and H4 and twofold hypertranscription of hundreds of linked genes (*8–10*).

The two *roX* RNAs perform redundant functions (*11*, *12*). The lethality of *roX1 roX2* double-mutant males can be rescued by expression of either *roX1* or *roX2* RNA from autosomal locations, showing that *roX* RNAs can be supplied in trans to coat the X chromosome (*12*). However, both genes synthesizing *roX* RNAs are normally located on the X chromosome, and we have suggested that this contributes to targeting dosage compensation to the correct chromosome (*7*).

In certain *msl* mutant backgrounds, the MSL complex is absent from most locations on the X chromosome, but a small subset of sites, termed chromatin entry sites, retain partial complexes (*7*, *13*). Two of these sites are the *roX* genes. When a *roX* gene is moved to an autosome, it recruits MSL complex, which occasionally spreads up to 1 megabase (Mb) into the flanking autosome in a pattern that varies considerably (Fig. 1A). This suggested that the MSL complex recognizes the X chromosome by first binding at *roX* genes (and perhaps additional sites) and then spreading in cis (*7*). The MSL proteins could recognize the *roX* genes by binding DNA, nascent RNA, or both. MSL proteins bind *roX* RNAs to form active complexes, and each *roX* gene also contains an MSL binding site (*9*, *14*).

The ectopic MSL spreading observed from autosomal *roX* transgenes was seen in only a small fraction of nuclei compared with the invariant MSL pattern in the wild-type male X chromosome (*7*, *13*). During complementation analyses of *roX1 roX2* mutants, we unexpectedly found that the genotype of the X chromosome strongly influenced ectopic MSL spreading from autosomal transgenes. We observed essentially no spreading in the presence of a wild-type X chromosome, but mutations in either *roX1* or *roX2* separately allowed modest MSL spreading from autosomal *roX* transgenes in some nuclei (Table 1; Fig. 1, B to D). In contrast, *roX1 roX2* mutants displayed extensive autosomal MSL spreading [>1 megabase pair (Mbp)] in nearly all nuclei regardless of their insertion site (Fig. 1, E to I; Fig. 2, A and B), including centric heterochromatin (Fig. 1I). In each case, MSL complexes still painted the X chromosome. Autosomal *roX* transgenes were poor sites of MSL spreading if one or both endogenous *roX* genes were functioning on the X chromosome, but the same transgenes supported efficient MSL spreading over autosomes in a *roX1 roX2* double mutant. Thus, *roX* genes appear to compete for limiting components for chromatin spreading.

We next asked if only X-linked *roX* genes could compete with autosomal MSL spreading. We found that a second autosomal *roX* transgene strongly reduced spreading from a reference *roX* transgene. For example, the MSL complex spread several megabase pairs from P{w^+GMroX2}97F (henceforth transgenics will be referred to as GMroX1-location or GMroX2-location, i.e., GMroX2-97F) in nearly all nuclei when it was the only source of *roX* RNA (Table 1; Fig. 2B). However, spreading was greatly reduced when GMroX1-67B was also present (Fig. 2C; Table 1). We tested seven pairs of *roX* transgenes and found that spreading from one site was reduced in both frequency and extent by the presence of a second *roX* gene (Table 1) (*15*). This confirms that the factors on the wild-type X chromosome responsible for competing for MSL spreading from an autosomal transgene are the endogenous *roX* genes and shows that *roX* genes are potent inhibitors of ectopic MSL spreading regardless of location.

The ability to compete with ectopic MSL spreading might reside in the *roX* RNAs or in the MSL binding sites within the *roX* genes. We constructed stocks in which MSL cis spreading from a reference GMroX2-97F transgene was challenged with two different *roX1* cDNA transgenes, both of which contain an MSL binding site. In one case, the *roX1* cDNA was transcribed from the constitutive *Hsp83* promoter (*13*). This transgene

[1]Howard Hughes Medical Institute, [2]Department of Molecular and Cellular Biology, [3]Program in Cell and Molecular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. [4]Department of Biology, Tufts University, Medford, MA 02155, USA.

*To whom correspondence should be addressed. E-mail: mkuroda@bcm.tmc.edu (M. I. K.); vmeller@tufts.edu (V.H.M.)

Downloaded from www.sciencemag.org on December 6, 2006