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 (xoloitzcuintle), 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
15. Supporting material is available on Science Online. 
17. The 350 modern dog sequences are from over 250 dogs corresponding to 124 haplotypes with sequenc- es deposited in GenBank, and 100 dogs from 20 different breeds (C. Vila, data not shown).
24. J. H. Zarr, Biostatistical Analysis (Prentice Hall, Upper Saddle River, NJ, ed. 4, 1999); the calculation is made according to example 2.4.5.
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. Göttherström, 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
13 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†

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 these early phototrophs are still
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 marked 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 paralogous 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-taxon 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-

<table>
<thead>
<tr>
<th>Orthologs supporting topology</th>
<th>RNA polymerase β subunit, bldA carboxylase</th>
</tr>
</thead>
</table>
| 1 Synechocystis                 | C. tepidum  
| 2 Synechocystis                 | C. tepidum, H. mobilis  
| 3 Synechocystis                 | C. aurantiacus  
| 4 Synechocystis                 | C. Mobilis  
| 5 H. mobilis                   | C. aurantiacus  
| 6 Synechocystis                 | C. aurantiacus  
| 7 Synechocystis                 | C. Mobilis  
| 8 C. aurantiacus               | H. Mobilis  
| 9 Synechocystis                 | C. aurantiacus  
| 10 Synechocystis                | H. Mobilis  
| 11 Synechocystis                | C. aurantiacus  
| 12 H. Mobilis                  | C. aurantiacus  
| 13 Synechocystis                | H. mobilis  
| 14 Synechocystis                | C. aurantiacus  
| 15 Synechocystis                | H. Mobilis  

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-taxon 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.
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 *Synechocystis* and *H. mobilis*, followed by *Synechocystis* and *Chloroflexus*. Averaged substitutions per site for *C. tepidum* and *H. mobilis*, followed by *Synechocystis* and *C. tepidum*, grouping 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 *Synechocystis* and *H. mobilis*, followed by *Synechocystis* and *Chloroflexus*. Averaged substitutions per site for *C. tepidum* and *H. mobilis*, followed by *Synechocystis* and *C. tepidum*, grouping was favored in only 8 cases.

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 non-photosynthetic 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 compart-

### 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.

<table>
<thead>
<tr>
<th>Putative function</th>
<th>Main PS</th>
<th>Other PS</th>
<th>Non-PS</th>
<th>Pathway/functional category</th>
<th>GenBank accession</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PS-specific</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mg-protoporphyrin-O-methyltransferase BchM</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td>(Bacterio)chlorophyll biosynthesis</td>
<td>str0525</td>
</tr>
<tr>
<td>Prochlorophyllum reductase BchA subunit</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td>(Bacterio)chlorophyll biosynthesis</td>
<td>str0772</td>
</tr>
<tr>
<td>Prochlorophyllum reductase BchN subunit</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td>(Bacterio)chlorophyll biosynthesis</td>
<td>str0750</td>
</tr>
<tr>
<td>Mg-chelatase subunit BchH</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td>(Bacterio)chlorophyll biosynthesis</td>
<td>str0749</td>
</tr>
<tr>
<td><strong>PS-related</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short chain alcohol dehydrogenase family</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>Oxidoreductase</td>
<td>str1208</td>
</tr>
<tr>
<td>Phytene desaturase</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>Carotenoid biosynthesis</td>
<td>str1293</td>
</tr>
<tr>
<td>Mg chelatase subunit Bch</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>(Bacterio)chlorophyll biosynthesis</td>
<td>str1030</td>
</tr>
<tr>
<td>Putative restriction endonuclease</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>Nucleic acid metabolism</td>
<td>sll1193</td>
</tr>
<tr>
<td>CblM protein</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>Cobalamin/B_{12} biosynthesis</td>
<td>sll0383</td>
</tr>
<tr>
<td>CobN protein</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>Cobalamin/B_{12} biosynthesis</td>
<td>str1211</td>
</tr>
<tr>
<td>Mg-protoporphyrin IX oxidative cyclase BchE</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>(Bacterio)chlorophyll biosynthesis</td>
<td>str0905</td>
</tr>
<tr>
<td>Bacteriochlorophyll synthase BchG</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>Cobalamin/B_{12} biosynthesis</td>
<td>sll1091</td>
</tr>
<tr>
<td>Hypothetical protein</td>
<td>5</td>
<td>3</td>
<td>8</td>
<td>?</td>
<td>sll1039</td>
</tr>
<tr>
<td>S’-methylthioadenosine phosphorylase</td>
<td>5</td>
<td>6</td>
<td>10</td>
<td>?</td>
<td>sll1035</td>
</tr>
<tr>
<td>Precorrin-8X methyltransferase</td>
<td>5</td>
<td>6</td>
<td>10</td>
<td>?</td>
<td>sll1039</td>
</tr>
<tr>
<td>Membrane protein</td>
<td>5</td>
<td>4</td>
<td>10</td>
<td>COBAMIN/B_{12} biosynthesis</td>
<td>str1467</td>
</tr>
<tr>
<td>Pyruvate:flavodoxin oxidoreductase</td>
<td>5</td>
<td>4</td>
<td>10</td>
<td>Oxidoreductase</td>
<td>str0932</td>
</tr>
<tr>
<td>Hydrogenase expression/formation protein HypB</td>
<td>5</td>
<td>4</td>
<td>11</td>
<td>Urease/hydrogenase associated</td>
<td>sll1079</td>
</tr>
<tr>
<td>Precorrin-2 C20-methyltransferase</td>
<td>5</td>
<td>5</td>
<td>11</td>
<td>Cobalamin/B_{12} biosynthesis</td>
<td>sll1879</td>
</tr>
<tr>
<td>Cobalamin kinase/Cobalamin Pi guanylyltransferase</td>
<td>5</td>
<td>6</td>
<td>13</td>
<td>Cobalamin/B_{12} biosynthesis</td>
<td>str0216</td>
</tr>
<tr>
<td><strong>Precorrin-4 C11-methyltransferase</strong></td>
<td>5</td>
<td>6</td>
<td>13</td>
<td>?</td>
<td>str0969</td>
</tr>
<tr>
<td><strong>Precorrin-6Y C5,15-methyltransferase</strong></td>
<td>5</td>
<td>5</td>
<td>13</td>
<td>Cobalamin/B_{12} biosynthesis</td>
<td>str0099</td>
</tr>
<tr>
<td><strong>Hydrogenase maturation protein HypF</strong></td>
<td>5</td>
<td>3</td>
<td>13</td>
<td>Urease/hydrogenase associated</td>
<td>str0322</td>
</tr>
<tr>
<td>Hypothetical protein</td>
<td>5</td>
<td>6</td>
<td>14</td>
<td>?</td>
<td>str0427</td>
</tr>
<tr>
<td>Exopolyphosphatase</td>
<td>5</td>
<td>4</td>
<td>14</td>
<td>Phosphorus compounds</td>
<td>sll1546</td>
</tr>
<tr>
<td>Possible photoperiodic protein</td>
<td>4</td>
<td>6</td>
<td>0</td>
<td>?</td>
<td>sll1874</td>
</tr>
<tr>
<td>Chlorophyllide reductase 35.5-KD chain BchX</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>(Bacterio)chlorophyll biosynthesis</td>
<td>NP_662309</td>
</tr>
<tr>
<td>L-asparaginase II</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>Carbon fixation</td>
<td>str0009</td>
</tr>
<tr>
<td>Ribulose bisphosphate carboxylase large chain</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>?</td>
<td>NP_662608</td>
</tr>
<tr>
<td>Putative thiol-disulfide isomerase/thioredoxin</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>?</td>
<td>sll0935</td>
</tr>
<tr>
<td>Hypothetical cytosolic protein</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>?</td>
<td>sll0585</td>
</tr>
<tr>
<td>SulE protein probably involved in FeS center assembly</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>?</td>
<td>sll0585</td>
</tr>
<tr>
<td>Phosphoenolpyruvate synthase/Pyruvate phosphate dikinase</td>
<td>4</td>
<td>1</td>
<td>9</td>
<td>Carbon fixation</td>
<td>NP_662565</td>
</tr>
<tr>
<td>Putative nucleotide-binding protein</td>
<td>4</td>
<td>1</td>
<td>10</td>
<td>?</td>
<td>NP_661693</td>
</tr>
<tr>
<td>Membrane lipoxygenase precursor</td>
<td>4</td>
<td>0</td>
<td>13</td>
<td>?</td>
<td>NP_661927</td>
</tr>
</tbody>
</table>
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 archaean, 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—nearly increases the size of this set with the caveat of potentially increasing the number of false positives. Genes found in all 11 bacterial and archaean genomes are predominantly housekeeping genes that function in nucleic acid 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 disconected from divergence and speciation in these organisms, confirming the extensive role of 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.

References and Notes
12. Materials and Methods are available as supporting material on Science Online.
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. Meller4*

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. When MSL protein/roX RNA ratios are high, assembly will be efficient, and complexes can 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 NH2-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 analysis 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 DNA 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