FIGURE 7.26. Phylogenetic profile analysis (grouping genes by their distribution patterns across species) helps predict their functions. This method analyzes complete genome sequences to identify genes that are found in similar sets of species. First, complete genomes are searched for the presence of a gene that is found in a particular (reference) genome. Second, a binary profile is created of the gene's presence and absence across the species. These two steps are repeated for all the genes in the reference genome. Third, a clustering method is used to identify genes with similar profiles (and thus similar distribution patterns across species). This figure shows one major cluster based on an analysis using the genome of the bacterium *Carboxydothermus hydrogenoformans* as the reference genome. In the *center* of the figure, profiles are shown for a subset of genes from this species: Genes are in *rows*, species are in *columns*, *red boxes* mean a gene is present in a species, and *black* means it is absent. Gene names are shown on the *right*; species names are not shown. The tree at the *left* represents the cluster diagram showing the similarity of the profiles for these genes. The columns of *red* in the central portion of the figure indicate that those species possess all or nearly all of the genes shown. Most of the genes listed, whose functions are known, have some role in sporulation (see the gene descriptions on the *right*), because species that sporulate use basically the same set of genes, and those that do not sporulate lack these genes (as indicated by the *mostly black spots* in the *central* figure). In turn, it can be inferred that genes found in the cluster, but with no known function, probably have some as yet unknown role in the sporulation process.

7.26, reprinted from Wu M. et al., *PLoS Genetics* **1:** e65, © 2005 Public Library of Science

Evolution © 2007 Cold Spring Harbor Laboratory Press