A tandem repeats database for bacterial genomes: application to the genotyping of Yersinia pestis and Bacillus anthracis

ABSTRACT

The analysis of the bacterial genome sequences currently available indicates that Bacillus anthracis and Yersinia pestis exhibit an average density of tandem repeat arrays longer than 100 bp (approximately 30 per Mb). A set of over fifteen informative markers, some of which exhibit a high degree of polymorphism, could be quickly produced in both situations by testing varying fractions of these sequences for polyhomogeneity. In one case, the polymorphism information content index is 0.82, and allele lengths between 600 and 1950 bp are extensive.

INTRODUCTION

Over the last five years, the genomes of four different eukaryotic organisms have been fully sequenced, including Saccharomyces cerevisiae, a unicellular hexotype, two multicellular organism species, Caenorhabditis elegans, and Drosophila melanogaster and Arabidopsisthaliana. After completing the sequencing of these major groups of organisms, it is possible to compare all known genes, gene families, and genomes across phylogenetically distinct groups. A study of predicted gene family relationships in various organism types, from single-celled to more complex animal and plant species, can help interpret functions of proteins and developmental control pathways. Through the comparison of family members to proteins in its own genome and the genomes of other organisms, it is possible to further elucidate the roles of these proteins and develop experimental procedures to determine their functions. Budding yeast has 113 conventional protein kinase genes. Yeast specific kinases were identified and clues about the function of unknown kinesis were provided through multiple alignment and parsimony analysis of protein kinse catalytic domain sequences, which also analyzed their structural similarity to other proteins. It was discovered that bacteria possess proteins that bind to zinc, but eukaryotic bacteria do not have large families of zinc-binding proteins. The presence and size of these families differs between elekotes. A comparison of the genomes of Drosophila, C. ellaris, and S. giganzii was conducted using DNA from DNA samples. The number of proteins for each of the 200 most commonly found protein domains was determined in cerevisiae. This included several calcium-binding domain (EF-hand) family members, which is one of many known to have this function. Drosophila has 130 proteins with this domain, while C. elegans have 79 proteins and only 16 in yeast. Members of the cytoskeleton protein families were used in a BLASTP and TBLASTN search of this Drosophila genome, which identified 262 genes with moderate to complete homology to cloning genes; new members of these families appeared later and proteins not found in other genomes. The superfamily of microtubule (MT) motor proteins known as kinesins is found in all eukaryotic organisms. The mtafida contains approximately 90% of the recombinant kin, with almost all members having it. Early identifying features of this superprotein were that they played an important role in transporting vesicles. Kinesin heavy chains (KHCs) are composed of the motor domain with ATPase activity, a central coiled-coil region, and kinesine-like proteins. Nevertheless, KHCs are now considered a subfamily of the kinesin superfamily and all members of this superfamily are referred to as kineses. We have followed this trend in our work, with fungi and animals reporting KHs and other kennel groups being identified in all eukaryotes. All kinesins have a domain that is homogenous to the motor

domain of KHC but very little sequence similarity outside of this domain. Some relics are not conserved, while others have their coiled-coil region (the portion of the molecule that contains the Kinetic Chain). The tail domain, which interacts with certain cargoes, is nonconserved. Kinesins are responsible for binding MTs and various cargoes, as well as performing force-generating activities such as spindle formation, chromosome segregation, and OM organization. The motor domain of KHC is located in the N-terminal region, but it can also be found in other kinesine forms. It has been shown that the C-terminal domain kinesins are "minus-end motile" and that all others are (apparently) "plus-ndiddle motild"; using the conserved motor domain, phylogenetic analysis has revealed nine subfamilies of the computer system isomorphic (not to be confused with another 991 KB proteins), not all resulting in a single subfamily. The first two known plant kinesins were identified in the tobacco pollen tube (PKH, a homologue of the polyphenol A) and tobacco phragmoplast (NtKRP125), with the latter being isolated at the molecular level from the same tubes. The kinesins found in plants include NtKRP125, four Arabidopsis kinase genes identified by PCR-based cloning, KCBP, PAKRS, and DcKPR120-1. The Arabidopsis genome was fully sequenced, which facilitated the search for kinesin-like genes in the Arabapatidazos by using the motor domain of reconstructed kinematic chromosomes. Interestingly, the genome of Arabidopsis contains the most kinesins sequenced in any eukaryote genome so far, and we have examined the predicted protein sequences for domains that might help us understand their function (by definition, all have a kinematic motor domain, some have dimerization domain-indicating coiled-coil domain) as well. A systematic review of the 61 Arabidopsis kinesin motor domain sequences along with 113 other sequence types has determined that there are seven subfamilies within the nine recognized subgroups of physiology (certainly not related to any known family), as well as some unique to Arabidia and potentially beneficial to plants.

CONCLUSION

Results: Traces of the mRNA differential display were used to identify genes with altered expression in Pneumocystis carinii-infected hosts, and the exact sequence of one of these fragments (gene encoding the mitochondrial ATPase 6 of F0F1, a subunit of this complex) was found to be homologous to the nucleotide of an expressed gene). The ATPase 6 gene is overexpressed during P. carinii infection, as indicated by the northern blot analysis of total RNA extracted from rat lung infected with PCA and mock-infused rabenoviruses. By in situ hybridization of cells found to be distal and apical of the respiratory tree and of alveoli that expressed the phosphatases 6 and 8 gene, it was shown that these regions were more than 120 genes and some of those on the disal parts of mice mice. The over-expression of the ATPase 6 gene in P. carinii infection is thought to be caused by type II pneumocytes and Clara cells, as indicated by the presence of SP-B gene expressed through a two-color fluorescent in situ hybridization.