Identification of conserved C2H2 zinc-finger gene families in the Bilateria

ABSTRACT

Based on current metazoan phylogeny, the minimum complement of C2H2 zinc-finger genes in the genome of the bilaterian common ancestor is composed of these 39 groups.

INTRODUCTION

The interaction between the MS2 coat protein and its translational operator is a well-established example of RNA-protein recognition, utilizing genetic, biochemical, and structural methods. Figure 1 displays the primary and secondary structures of the recombinant rRNA hairpin that establish contacts with both subunits of each coat proteins dimer. The coat protein complex with its RNA target is highly intricate, as two unpaired adenosines are inserted into equivalent pockets on different subunits of the coat dimer (Figure 2). The interactions between A-4 and A-10 with coat proteins involve non-identical contacts with the same five amino acid residues, Val29, Thr45, Ser47, Finally, and Lys61. The use of X-ray crystallographic analysis indicates specific amino acid-nucleotide interactions, but fails to provide a clear explanation of their respective roles in RNA-binding and translational repression. In the experiments described here, we used amino acid substitutions of A-pocket amino acids in single-chain coat protein heterodimers to determine the significance of each residue's interaction with A-4 and A-10.

CONCLUSION

In order to avoid basic sequence contingency loci of limited epidemiological value, we confined our study of tandem repeats to minisatellites, which are repeat units longer than 9 base-pairs, and to make typing alleles easier with agarose gel electrophoresis. Simple sequence contingency loci are also included in the database, which is particularly useful for molecular pathogenicity studies. The tandem repeats database was used to test the use of Y. pestis and B. anthracis on two of the most genetically homogeneous human pathogenes. It is possible that a common database format will be established to identify and conduct epidemiological analyses of small-scale pathogen types that can be characterized by minisatellite typing, with additional data on tandem repeat polymorphism added to the current database to prevent duplication of work and nomenclature. The abundance and use of tandem repeats vary greatly among different bacterial species, with some having an excess of more than three tandem repetitions in their genome, such as M. tuberculosis and P. aeruginosa. The occurrence of polymorphism in tandem repeats is likely to affect protein structure rather than gene activity. In M. tuberculosis, all tandem repeateds with total length (L) above 100 bp and 9 or 15 base-pairs long units are located with ORFs. This means that a significant proportion of these tandem repetitions correspond to the PE and PPE multigene families. Tandem repeat polymorphism is strongly associated with one or more sequenced allele characteristics in the two species examined, as shown in Figure 7. In Yersinia pestis, a strong correlation exists between the number of allels observed and the homogeneity of the tandem array. In Bacillus anthracis (pictured) the strongest correlations are found with total array length and GC content. The two species do not exhibit comparable correlations, implying that the tandem repeat's polymorphism cannot be deduced from its primary sequence. Minisatellites, like those in the human genome are known for their high

polymorphism, despite having low internal homogeneity. In order to dispel any suggestion that polymorphism is linked to subclasses of alleles displaying a higher level of internal homogeneity, more comprehensive allle sequencing will be necessary. Likewise, allELE sequencing would be required to officially prove that variations in allome size are likely due to differences in the number of repeats. Of particular importance, five among the B. The anthracis markers listed (Ceb-Bams1, 3, 7, 13, and 30) are highly polymorphic with PIC values (or Nei's index) greater than 0.7. It is worth noting, however, that the allele length observed for Ceb–Bamis1 in the Ames strain is not typical according to the sequence data (Table 2). This may be due to a high mutation rate at Cec-Barms1/3 and subsequently to some sequencing error. The Ames strain is unlikely to have allele 4 (Table 3), as the expected allolicle size does not correspond to any existing allELE in the product set. The locus is moderately polymorphic, with a PIC value of 0.26 and only three alleles present in Table 2, which suggests that interpreting it may result in ambiguity among sequencing researchers. This problem could be remedied by using Ceb-Bams1 and Cec-Bean28, the same strain used for the sequencing project. The phylogenetic tree illustrated in Figure 6 tends to cluster strains with alleles of similar size at the most variable loci, despite the lack of consideration for the magnitude of allile size difference in building the distance matrix. This is comparable to observations made in H. It is suggested by influenzae and others that mutation events are primarily small-scale changes. To better understand the process of event generation, it is necessary to conduct more detailed studies using full allele sequencing.