Toward Chromosome-Level Comparative Analysis of Vertebrate Genomes

Katsuhiko Murakami¹
katsu@gsc.riken.go.jp
Tetsushi Yada²
yada@ims.u-tokyo.ac.jp
Hideki Noguchi¹
hide@gsc.riken.go.jp
Atsushi Toyoda¹
toyoda@gsc.riken.go.jp
Yoshiyuki Sakaki¹
sakaki@gsc.riken.go.jp
Yasushi Totoki¹
totoki@gsc.riken.go.jp
Masahira Hattori¹
hattori@gsc.riken.go.jp

¹ RIKEN Genomic Sciences Center, 1-7-22, Suehiro-cho, Tsurumi, Yokohama, Kanagawa, JAPAN
² Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan

Keywords: comparative genomics, gene finding, regulatory element

1 Introduction

Since the publication of the human draft genome sequence, steady progress around the world is made to analyze the genome. In addition, mouse genome project and other vertebrate genome projects of model organisms are rapidly producing much genomic sequences. Comparative analysis is becoming increasingly important as a tool for genome sequence analysis [3]. In this study, we focus on vertebrate genome sequences of various species, such as human, mouse and rat. The evolutionary conserved regions in human genome sequences are identified effectively by the comparison with those vertebrate genome sequences.

The main goal of this research is to identify novel genes and regulatory regions and to elucidate genome-level events in evolution from genome sequences. We have developed a comparative genome system which collects data, compares them, visualizes the results, and produces primer sets using primer3 [6] for the PCR experiments for confirmation of the exon prediction. Users can submit syntenic genomic sequences or select a publicly available syntenic region. The system allow the users to analyze and convert data easily, so that users can browse the various analysis results using PipMaker [4] (WWW server for genome comparison and java application laj), and Alfresco [2] (java application), without producing annotation files on their own.

2 Method and Results

Figure 1 gives an overview of the system. We have collected publicly available draft data to compare with human genome sequences, from Trace server [5], and other WWW sites of major sequence centers. Currently, the pre-analysis of chromosome 21 of human has been done. The analysis includes RepeatMasker, CpG island detection, BLAST against principal databases, BLAST matching syntenic sequences each other, a fast global alignment with GLASS [1], ab initio prediction, DNA motif search in regulatory regions. We have evaluated some of the procedures of genome comparison, and made a protocol for acquiring appropriate results.
3 Discussions

There are many conserved regions in the annotated non-coding regions among the alignments of human-mouse sequences. The system predicts coding regions base on the conservation of the DNA sequences. Since some of them may be true exons, we plan to check the prediction using RT-PCR.

References


