SUMMARY OF Ph.D. DISSERTATION

School School of Fundamental Science and Technology	Student Identification Number	SURNAME, First name HACHIYA, Tsuyoshi

Title

Probabilistic Models for Inferring Orthology in Comparative Genome Analysis

Abstract

Dramatic increases in throughput of nucleotide sequencing machines yield exponential growth of the number of sequenced genomes. In order to mine valuable insights from the rapidly growing repository of biological sequences, inferring evolutionary relationships is one of the most fundamental prerequisites.

Evolutionary relationships among biological sequences are divided into two classes: orthology and paralogy. It is widely accepted that orthologous sequences have retained identical function to their ancestral sequences, whereas paralogous sequences have acquired new lineage-specific functions. Accordingly, identifying orthologous sequences is an effective way to infer the function of biological sequences. Exponential growth of the number of sequenced genomes arises two new challenges in the identification of orthologous sequences. The first challenge is to identify orthologous chromosomal segments. The second challenge is to identify orthologous genes encoded by closely related genomes.

In Chapter 1 of this thesis, the significance of inferring orthology was explained.

In Chapter 2, in order to address the first challenge, an algorithm named OSfinder was proposed. OSfinder identifies orthologous chromosomal segments based on probabilistic models. The algorithm enables to automatically optimize the parameter values that play a key role to compute orthologous chromosomal segments. This automation would increase the throughput of comparative genome analysis. Furthermore, our computational experiments using mammalian and bacterial genomes demonstrate that the optimization algorithm of OSfinder makes it possible to identify orthologous chromosomal segments more accurately than existing algorithms.

In Chapter 3, in order to address the second challenge, an algorithm named OASYS was proposed. Probabilistic modeling approaches provide us with a framework to integrate heterogeneous evidence in order to make inferences. By integrating the information of protein sequence homology and the information of gene order conservation based upon the framework, OASYS enables accurate identification of orthologous genes even when closely related genomes are compared.

In Chapter 4, the potential contribution of OSfinder and OASYS to systems biology and evolutionary biology was discussed.