

DA  
3106  
2007  
416

**Protein Evolution**  
**Based on Analysis of DNA Sequences**

Division of Agriculture and Forestry  
Doctoral Degree Program in Agricultural Sciences  
University of Tsukuba

**Yuko Ohfuku**

寄贈  
大福裕子氏

03302711

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Materials and Methods</b>	<b>7</b>
2.1	Definition of orthologous genes . . . . .	7
2.2	The G+C content and the G+C content at the third codon position . . . . .	7
2.3	Division of a DNA sequence by using a Markov model . . . . .	9
2.4	Definitions of class, region and CLASS . . . . .	10
<b>3</b>	<b>Lateral genes in the genomic DNA sequence</b>	<b>13</b>
3.1	The correlation between the G+C content of the genomic DNA sequences and the G+C content at the third codon position (GC3 content) for all ORFs of the genomic DNA sequences . . . . .	13
3.2	The correlation between the G+C content of the genomic DNA sequences and the G+C content at the third position of the three types of codons found in orthologous genes . . . . .	16
3.3	Distribution of the G+C content at the third codon position of synonymous codons in the orthologous genes of <i>Pyrococcus horikoshii</i> OT3 . . . . .	19
3.4	Characteristics of the high and low GC3 content regions . . . . .	22
<b>4</b>	<b>Relationship between DNA sequence and three-dimensional structure</b>	<b>29</b>
4.1	Analysis of probability of DNA sequence . . . . .	29
4.1.1	Flavodoxin reductase . . . . .	29
4.1.2	Heat shock protein ( <i>grpE</i> ) . . . . .	32
4.1.3	Flavin oxidoreductase . . . . .	32
4.1.4	Integrase/recombinase <i>xerD</i> . . . . .	37
4.1.5	Endonuclease III . . . . .	37
4.1.6	Elongation factor Tu ( <i>tufB</i> ) . . . . .	42

4.2	Comparison between the divided regions and the domain classified in accordance with the definition of CATH . . . . .	42
<b>5</b>	<b>Structural evolution of protein</b>	<b>47</b>
5.1	Flavodoxin reductase of <i>Escherichia coli</i> . . . . .	47
5.1.1	DNA sequence, amino acid sequence, three-dimensional structure and the divided regions . . . . .	47
5.1.2	Homology search with the DNA sequences of the divided regions . . . . .	47
5.2	Heat shock protein ( <i>grpE</i> ) of <i>Escherichia coli</i> . . . . .	54
5.2.1	DNA sequence, amino acid sequence, three-dimensional structure and the divided regions . . . . .	54
5.2.2	Homology search with the DNA sequences of the divided regions . . . . .	54
5.2.3	Duplication of gene . . . . .	59
<b>6</b>	<b>Discussion</b>	<b>65</b>
	<b>Acknowledgements</b>	<b>75</b>
	<b>References</b>	<b>77</b>
	<b>Appendix</b>	<b>91</b>