| | |
|---|---|
| 氏　　名 | Harshita Sharma |
| 博士の専攻分野の名称 | 博士（学術） |
| 学位記号番号 | 博理工甲第 959 号 |
| 学位授与年月日 | 平成 26 年 9 月 19 日 |
| 学位授与の条件 | 学位規則第 4 条第 1 項該当 |
| 学位論文題目 | Demonstration of the feasibility of the genome distance-based universal analysis by distinct applications such as microbiome analysis and familial clustering of mice<br>（マイクロバイオーム解析やマウス家系クラスタリング等の異なる用途への応用による「ゲノム距離解析汎用法」の有用性実証研究） |
| 論文審査委員 | 委員長　教　授　松岡　浩司<br>委　員　准教授　根本　直人<br>委　員　准教授　藤野　　毅<br>委　員　准教授　石丸　雄大<br>委　員　名誉教授　西垣　功一 |

## 論文の内容の要旨

The universal method that can apply to all of the organisms by a single, simple approach is ideal also in the field of genome analysis. In this point of view, all of the technologies so far established including the NGS approach were evaluated to hold some drawbacks.

Among these, the GP method was shown in this study that it can be universal as it is applicable to any type of cells for identification and classification both widely and in depth. Through this study, we could clear off the fog that had made us less visible of the true value of the genome distance-based approach. The metaphorical expression of fog has a meaning that patchy studies so far performed had been failing to provide the whole image of the GP approach due to the missing link.

By the studies performed here which include a large number of anonymous organisms and ultimately close relationship of a species (mouse), the genome distance-based approach was proven to be effective to both of the quite different methods. Therefore, it is now evident that the GP method is effective and powerful for any type of research so far as the genome distance (difference in the genome sequence) is involved.

Independently from the main theme of examining the feasibility of the genome distance-based approaches for genome analyses, this study includes additional achievements as follows. We established the MMV approach as a device for efficient DNA amplification and further analysis of genetic information from a single cell in sub-μL for NGS-non-dependent microbiome analysis (NNMA). The MMV chip endows potential for highly parallel and multi-step processing of 1,024 different cells simultaneously, with succeeding GP-based processing for obtaining sequence specific information from each cell. First of all, different substrate coating effects were tested to check their efficiency

in the prevention of surface adsorption of biomolecules on the MMV wall: mPEG thiol coating, silicon coating and BSA coating. BSA coated MMV was found to be most suitable as it resulted in higher cell viability and lower level of cell adsorption on the MMV wall compared to other coatings. In the next step, preliminary experiments were performed to optimize PCR conditions in MMV and *E. coli* genomic DNA (from $10^4$, $10^3$, $10^2$, and single molecule/0.5 μL) were successfully PCR-amplified in BSA coated MMV. Next, oral microbes were used as a test case for NNMA. Oral sample was subjected to deflocculation, limit dilution, single cell DNA extraction and single cell random PCR in MMV followed by PCR product analysis using micro-temperature gradient gel electrophoresis (μTGGE) and computer aided normalization. As a result, some of the MMV wells generated distinct genome profiles, suggesting successful bacterial cell distribution in different MMV wells and the extraction/amplification of DNA in each well. The DNAs recovered from the μTGGE gel as commonly conserved genetic fragments (ccgfs) were further analyzed by conventional cloning/sequencing analysis, and the resulting DNA sequences were subjected to BLAST analysis. In this analysis, *Leuconostoc citreum*, *Haemophilus parainfluenzae*, *Rothia dentocariosa*, *Rothia mucilaginosa*, *Aspergillus nidulans*, *Cronobacter sakazakii*, *and Psychrobacter sp*. were tentatively identified based on sequence similarity with organisms in the NCBI database. Because all of these microbes can be assigned as stable or transient oral inhabitants, this finding strongly supports the fact that NNMA analysis was totally performed. The genome profiles obtained in this analysis were subjected to clustering analysis, in which partial number of clones was assigned to the possible species based on their own ccgf DNA sequence. The results were reproducible for two other trials, supporting the effectiveness of the NNMA method. NNMA offers the possibility of detecting, in principle, a minority of 1 in 10,000 if tens of 1,024-well MMV plates are used.

Secondly, we analyzed genome distance and familial relationships of members of three mouse families using GP method. Sixteen blood samples from pedigree-known mice were collected and processed to the GP analysis with three different random PCR probes. Phylogenetic trees constructed based on the $d_G$ data showed family-dependent clustering of samples, meaning that samples could be clustered with their family members. This result was reproducible with all the probes used. Moreover, sequence-specific information from commonly appearing bands in all of the samples' genome profiles referred as ccgf were collected and processed to sequencing analysis. For comparison, 18S rDNA sequence in all samples were also amplified and sequenced. 18S rDNA sequencing failed to determine mouse familial relationship while ccgf sequencing analysis provided better result and was successful to some extent. In addition a single family-multigeneration genome analysis of 31 mouse samples was also performed. The result demonstrated the possibility of genome distance-based clustering of siblings from the same parent without *a priori* knowledge of their sequence.

Furthermore, genome distance data of different species was plotted in a 2D map and in genome sequence space (GSS) to investigate properties of their genomic sequences and intra/inter-specific relations. At the same time, this approach has a merit to understand the mutual relationship of genomes, which is difficult from mere numerical data or isolated one. The genome distance-dependent two-dimensional mapping of 48 samples from phylum chordata (mammalia, aves, amphibia, reptilia) and arthropoda (insecta) is presented using principal coordinate analysis (POA) and its correlation with Euclidean distance was measured. Subsequently, a genome sequence space of these samples was constructed, where each sample is separated from the others by their corresponding genome distance.

This study has developed the potency of genome distance based universal method in microbial and animal

genotyping. The NNMA method has overcome the limitations that were associated with NGS-dependent microbiome analysis and enabled identification of members of a microbiota with ease of handling and simplicity of data interpretation. The genome distance method was further found to be capable of discriminating between two families and clustering of siblings born from the same parent. The results of this study have re-confirmed the universal nature of the genome distance-based approach and marked a path for future GP-based wide-scale exploration of microbial communities and mammalian familial relationships.

# 論文の審査結果の要旨

The applicant's dissertation was titled as: *Demonstration of the feasibility of the genome distance-based universal analysis by distinct applications such as microbiome analysis and familial clustering of mice.*（マイクロバイオーム解析やマウス家系クラスタリング等の異なる用途への応用による「ゲノム距離解析汎用法」の有用性実証研究）

The universal method that can apply to all of the organisms by a single, simple approach is ideal in the field of genome analysis. In this point of view, all of the technologies so far established including the NGS approach were evaluated to hold some drawbacks. Among these, the GP method was shown in this study that it can be universal as it is applicable to any type of cells for identification and classification both widely and in depth. Through this study, we could clear off the fog that had made us less visible of the true value of the genome distance-based approach. The metaphorical expression of fog has a meaning that patchy studies so far performed had been failing to provide the whole image of the GP approach due to the missing link. By the studies performed here which include a large number of anonymous organisms and ultimately close relationship of a species (mouse), the genome distance-based approach was proven to be effective to both of the quite different methods. Therefore, it is now evident that the GP method is effective and powerful for any type of research so far as the genome distance (difference in the genome sequence) is involved. Independently from the main theme of examining the feasibility of the genome distance-based approaches for genome analyses, this study includes additional achievements as follows. We established the MMV approach as a device for efficient DNA amplification and further analysis of genetic information from a single cell in sub-μL for *NGS-nondependent microbiome analysis* (NNMA). The MMV chip endows potential for highly parallel and multi-step processing of 1,024 different cells simultaneously, with succeeding GP-based processing for obtaining sequence specific information from each cell. First of all, different substrate coating effects were tested to check their efficiency in the prevention of surface adsorption of biomolecules on the MMV wall: e.g., mPEG thiol coating, silicon coating and BSA coating. BSA coated MMV was found to be most suitable as it resulted in higher cell viability and lower level of cell adsorption on the MMV wall compared to other coatings. In the next step, preliminary experiments had been performed to optimize PCR conditions in MMV. After those preliminary experiments, NNMA was performed using oral microbes as a test case. Oral sample was subjected to deflocculation, limit dilution, single cell DNA extraction and single cell random PCR in MMV followed by PCR product analysis using micro-temperature gradient gel electrophoresis (μTGGE) and computer aided normalization. As a result, some of the MMV wells generated distinct genome profiles, suggesting successful bacterial cell distribution in different MMV wells and the extraction/amplification of DNA in each well. The DNAs recovered from the μTGGE gel as commonly conserved genetic fragments (ccgfs) were further analyzed by conventional cloning/sequencing analysis, and the resulting DNA sequences were subjected to BLAST analysis. In this analysis, *Leuconostoc citreum*, *Haemophilus parainfluenzae*, *Rothia dentocariosa*, *Rothia mucilaginosa*, *Aspergillus nidulans*, *Cronobacter sakazakii*, and *Psychrobacter sp*. were tentatively identified based on sequence similarity with organisms in the NCBI database. Because all of these microbes coould be assigned as stable or transient oral inhabitants, this finding strongly supported the fact that NNMA analysis was totally performed in success. The genome profiles obtained in this analysis were subjected to clustering analysis, in which partial number of clones was assigned to the possible species based on their own ccgf DNA sequence. The results were reproducible for two other trials, supporting the effectiveness of the NNMA method. The NNMA method has overcome the limitations that were associated with NGS-dependent microbiome analysis and enabled identification of members of a microbiota with ease of handling and

simplicity of data interpretation.

Secondly, the applicant analyzed genome distance and familial relationships of members of three mouse families using the GP method. Sixteen blood samples from pedigree-known mice were collected and processed to the GP analysis with three different random PCR probes. Phylogenetic trees constructed based on the $d_G$ data showed family-dependent clustering of samples, meaning that samples could be clustered with their family members. This result was reproducible with all the probes used. Moreover, sequence-specific information from commonly appearing bands in all of the samples' genome profiles referred as ccgf were collected and processed to sequencing analysis. For comparison, 18S rDNA sequence in all samples were also amplified and sequenced. 18S rDNA sequencing failed to determine mouse familial relationship while ccgf sequencing analysis provided better result and was successful to some extent. In addition a single family-multi-generation genome analysis of 31 mouse samples was also performed. The genome distance method was further found to be capable of discriminating between two families and clustering of siblings born from the same parent.The result demonstrated the possibility of genome distance-based clustering of siblings from the same parent without *a priori* knowledge of their sequence.

Furthermore, genome distance data of different species was plotted in a 2D map and in genome sequence space (GSS) to investigate properties of their genomic sequences and intra/inter-specific relations. At the same time, this approach has a merit to understand the mutual relationship of genomes, which is difficult from mere numerical data or isolated one. The genome distance-dependent two-dimensional mapping of 48 samples from phylum chordata (mammalia, aves, amphibia, reptilia) and arthropoda (insecta) is presented using principal coordinate analysis (POA) and its correlation with Euclidean distance was measured. Subsequently, a genome sequence space of these samples was constructed, where each sample is separated from the others by their corresponding genome distance.

This study has developed the potency of genome distance based universal method in microbial and animal genotyping. The results of this study have re-confirmed the universal nature of the genome distance-based approach and marked a path for future GP-based wide-scale exploration of microbial communities and mammalian familial relationships.

Therefore, the applicant has first succeeded in uniformly demonstrating the usability of $d_G$ (genome distance) based approach for universal identification and clustering of organisms experimentally and theoretically.