氏　　名	Deepti Diwan
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 978 号
学位授与年月日	平成 27 年 3 月 24 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	Study on the demonstration of the systematic genome mutation within a single body based on genome distance method （ゲノム距離法によって単一個体内での系統的ゲノム変異の存在を立証する研究）
論文審査委員	委員長　教　授　松岡　浩司 委　員　准教授　根本　直人 委　員　准教授　藤野　　毅 委　員　准教授　幡野　　健 委　員　名誉教授　西垣　功一

論文の内容の要旨

Since mutations are mainly caused during the process of DNA replication, repair, and movable elements (such as transposon) activity, measuring the degree of mutation frequency is crucially important for understanding the activities of life including these phenomena. In this sense, various approaches have been made to elucidate this issue. The most promising even was the advent of the next generation sequencing (NGS) around 10 years ago to solve this problem. With all its high powerfulness, the current stage of NGS-based approach is not reaching the level to be able to detect the subtle mutation rate sufficiently though it has already disclosed the differences in the genome sequence between cancer cells and genetic variation in somatic tissues. A major bottleneck in the implementation and capitalization of the NGS approach resides in such attributes as less universal applications, low cost-effectiveness, complicated data processing which is still left for future endeavors. In this situation, genome sequence differences between normal cells within a single organism have not yet been uncovered. That is, whether each cell within an individual organism possesses a systematically different genome sequence or not is not clear. Therefore, to challenge such an issue, we have taken an advantage of the well-established approach, the Genome Profiling (GP) method.

By the study performed here, which includes the measure of genomic distance between two cells within an individual organism 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 evident that the GP method can be applied to quite different purposes; genome difference analysis within a body or between bodies.

In this study, we have first discovered that there is a systematic (from bottom to top branches) variation in the genome sequence among cells of leaves in trees (Yoshino cherry and Japanese beech trees) by employing Genome Profiling (GP)

method. In other words, we detected considerable systematic variation in genome sequences among cells in individual woody plants. The degree of genome sequence difference (genomic distance) varied *systematically* from the bottom to the top of the plant, such that the greatest divergence was observed between leaf genomes from uppermost branches and the remainder of the tree. This systematic variation was observed within both Japanese beech (*Fagus crenata*) and Yoshino cherry (*Prunus* × *yedoensis*) trees. As measured by GP, the genome distance between two cells within an individual organism was non-negligible, and was correlated with physical distance (i.e., branch-to-branch distance). This phenomenon was concluded to be the result of accumulation of mutations occurring in each cell division, implying that the degree of divergence is proportional to the number of generations separating the two cells. Since this phenomenon was detected by the GP method that is estimated to detect the mutation of less than 10-5/base/replication, a large number of accumulated mutations must exist between distantly located cells in the tree.

Secondly, the feasibility of genome distance-based accumulation of mutations is demonstrated by analyzing genomic variations in a leaf from *Arabidopsis thaliana*, it was further revealed that the genome sequence varies *systematically* (between neighboring strips of a leaf) along the physical position. Based on these observations and results, we could establish the individuality of the genome of each cell belonging to a single body. Intriguingly, the genomes of all leaf portions are not completely identical but, rather, are different from each other depending on their positions along the growth direction. The results obtained here suggest that there is a systematic difference in the genome distance; i.e. positional difference which is by no means random, following the same context formerly reported (i.e., woody plants), considering that the DNA replication should be strictly controlled to be an ultimately low mutation rate. In addition, we also analyzed the DNA polymerase behavior against methylated DNA, demonstrating that the DNA polymerase may be affected, though slightly, in its function by DNA methylation depending on the degree of methylation of genomic DNA. In general, the methylation has an effect of protecting the site from the restriction enzyme cleavages and thus resuscitates random PCR products which might have disappeared due to the cleavage of their template regions. As a result, the methylation effect was determined to be negligible in our experimental system and the difference in the genomic DNA should be reduced to the DNA sequence (GATC) alteration. Besides, this study opened a way to the evaluation of genomic DNA methylation by introducing a random PCR-based novel approach which neither depends on the high-throughput screening nor DNA microarray.

Furthermore, as my supplementary study, we analyzed genome distance and familial relationships of members of three mouse families using GP method and the phylogenetic trees constructed based on the genome distance data showed family-dependent clustering, this result was reproducible with all the probes used. Thus, this study demonstrated the possibility of genome distance-based clustering of siblings from the same parent without *a priori* knowledge of their sequence.

These studies have developed the potency of genome distance based method universally in comparative genome studies. Until now, only sequencing-dependent genome-wide analysis has been done with requiring conserved DNA elements to be analyzed. The genome distance based method which was; capable of discriminating between two families and clustering of siblings born from the same parent without *a priori* knowledge of their genome sequences and was; successful in revealing the systemic difference of genomic DNAs between cells in a single body was shown to be not only easy, reproducible but also widely applicable to any type of genome analysis. These properties have not been

realized by the other methodologies.

Finally, the essence of genome distance itself was considered and the reason of its potency was unveiled to come from its well-balanced random sampling of DNA fragments from the original genomic DNA which leads to reproducible and ready sampling and yet providing a necessary and sufficient amount of information to compare two genomes. The genome distance based GP method was indicated to have advantages in readiness, reproducibility, and universality over the conventional approaches mainly based on genetic distance since the latter cannot be universally applicable to all of organisms in addition to its methodological burden. On the other hand, the current success in discrimination between highly closely-related cells was shown to be a consequence of relatively higher rate of spontaneous mutations than in prior expected.

論文の審査結果の要旨

The theme of the applicant's thesis was "Study on the demonstration of systematic genome mutation within a single body based on genome distance method". It was composed of three main chapters and was written featuring the scientifically valuable finding of systematic mutation of genome sequences among cells within a single body.

Since mutations are mainly caused during the process of DNA replication, repair, and movable elements (such as transposon) activity, measuring the degree of mutation frequency is crucially important for understanding the activities of life including these phenomena. In this sense, various approaches have been made to elucidate this issue. The most promising event was the advent of the next generation sequencing (NGS) around 10 years ago to solve this problem. With all its high powerfulness, the current stage of NGS-based approach is not reaching the level to be able to detect the subtle mutation rate sufficiently though it has already disclosed the differences in the genome sequence between cancer cells and genetic variation in somatic tissues. A major bottleneck in the implementation and capitalization of the NGS approach resides in such attributes as less universal applications, low cost-effectiveness, complicated data processing which is still left for future endeavors. In this situation, genome sequence differences between normal cells within a single organism have not yet been uncovered. That is, whether each cell within an individual organism possesses a systematically different genome sequence or not was not clear. Therefore, to challenge such an issue, the applicant and her co-workers have taken an advantage of the well-established approach, the Genome Profiling (GP) method.

By the study performed here, which includes the measure of genomic distance between two cells within an individual organism and ultimately close relationship of a species (mouse), the genome distance-based approach was proven to be effective to both of the quite different applications. Therefore, it is evident that the GP method can be applied to quite different purposes; genome difference analysis within a body or between bodies.

In this study, the applicant has first discovered that there is a systematic (from bottom to top branches) variation in the genome sequence among cells of leaves in trees (Yoshino cherry and Japanese beech trees) by employing Genome Profiling (GP) method. In other words, the applicant detected considerable systematic variation in genome sequences among cells in individual woody plants. The degree of genome sequence difference (genomic distance) varied *systematically* from the bottom to the top of the plant, such that the greatest divergence was observed between leaf genomes from uppermost branches and the remainder of the tree. This systematic variation was observed within both Japanese beech (*Fagus crenata*) and Yoshino cherry (*Prunus* × *yedoensis*) trees. As measured by GP, the genome distance between two cells within an individual organism was non-negligible, and was correlated with physical distance (i.e., branch-to-branch distance). This phenomenon was concluded to be the result of accumulation of mutations occurring in each cell division, implying that the degree of divergence is proportional to the number of generations separating the two cells. Since the GP method is estimated to detect the mutation of no less than 10^{-5}/base/replication, a large number of accumulated mutations must exist between distantly located cells in the tree.

Secondly, the feasibility of cell-to-cell distance-based accumulation of mutations was demonstrated by analyzing genomic variations in a leaf from *Arabidopsis thaliana*. It was further revealed that the genome sequence varies *systematically* (between neighboring strips of a leaf) along the physical position. Based on these observations and results, the ap-

plicant could find the difference in the genome sequence of each cell within a narrow region belonging to a single body. Intriguingly, the genomes of all leaf portions were not completely identical but, rather, different from each other depending on their positions along the growth direction. The results obtained here suggest that there is a systematic difference in the genome distance; i.e. positional difference which is by no means random, following the same conclusion above reported (i.e., woody plants). In addition, the applicant also analyzed the DNA polymerase behavior against methylated DNA, demonstrating that the DNA polymerase may be affected, though slightly, in its function by DNA methylation depending on the degree of methylation of genomic DNA. In general, the methylation has an effect of protecting the site from the restriction enzyme cleavages and thus resuscitates random PCR products which might have disappeared due to the cleavage of their template regions. As a result, the methylation effect was determined to be negligible in her experimental system and the difference in the genomic DNA should be reduced to the DNA sequence (GATC) alteration. Besides, this study opened a way to the evaluation of genomic DNA methylation by introducing a random PCR-based novel approach which neither depends on the high-throughput screening nor DNA microarray.

Furthermore, as her supplementary study, the applicant analyzed genome distance and familial relationships of members of three mouse families using GP method and the phylogenetic trees constructed based on the genome distance data showed family-dependent clustering, this result was reproducible with all the probes used. Thus, this study demonstrated the possibility of genome distance-based clustering of siblings from the same parent without *a priori* knowledge of their sequence.

These studies have developed the potency of genome distance based method universally in comparative genome studies. Until now, only sequencing-dependent genome-wide analysis has been done with requiring conserved DNA elements to be analyzed. The genome distance based method which was; capable of discriminating between two families and clustering of siblings born from the same parent without *a priori* knowledge of their genome sequences and was; successful in revealing the systemic difference of genomic DNAs between cells in a single body was shown to be not only easy, reproducible but also widely applicable to any type of genome analysis. These properties have not been realized by the other methodologies.

Finally, the essence of genome distance itself was considered and the reason of its potency was unveiled to come from its well-balanced random sampling of DNA fragments from the original genomic DNA which leads to reproducible and ready sampling and yet providing a necessary and sufficient amount of information to compare two genomes. The genome distance based GP method was shown to have advantages in readiness, reproducibility, and universality over the conventional approaches mainly based on genetic distance since the latter cannot be universally applicable to all of organisms in addition to its methodological burden. On the other hand, the current success in discrimination between highly closely-related cells was shown to be a consequence of relatively higher rate of spontaneous mutations than formerly expected.

In this thesis, the contents above mentioned were described rationally and logically, meaning that the applicant has performed academically valuable works.