

氏名	Aamer, Waleed Hussein Omer
学位(専攻分野)	博士(理学)
学位記番号	総研大甲第 1939 号
学位授与の日付	平成29年3月24日
学位授与の要件	生命科学研究科 遺伝学専攻 学位規則第6条第1項該当
学位論文題目	Genetic Structure of East African Populations Based on HLA Class I Genes
論文審査委員	主査 教授 斎藤 成也 教授 北野 潤 教授 平田 たつみ 教授 城石 俊彦 教授 徳永 勝士 東京大学

論文の要旨

Summary (Abstract) of doctoral thesis contents

East Africa is characterized by high levels of genetic, linguistic and cultural diversity. Studies have shown that, in Africans, there is a strong correlation between genetic and linguistic families. Focusing on the Sudanic region of east Africa, a recent study based on genome-wide SNP data found two substructures in the Sudanese population. The two substructures were related to Afro-Asiatic and Nilo-Saharan/Niger-Congo linguistic families. Furthermore, the Afro-Asiatic family contains several groups with different genetic background which need to be elucidated. Here, I studied the genetic diversity of HLA class I genes in eight ethnic groups from east Africa. The samples I used represent three countries in East Africa; Sudan, South Sudan and Ethiopia. The samples from Sudan have wide geographical distribution and all samples belonged to three known African linguistic families; Afro-Asiatic, Nilo-Saharan, and Niger-Congo. I first sequenced the samples using next generation sequencing method on Miseq platform. Then I analyzed the sequence data and identified HLA-A, HLA-B and HLA-C genotypes. While identifying the genotypes, I found four mutations not previously seen in their sequence background, which can be considered as new allele. I confirmed all of the mutations by Sanger sequencing. The genotypes data were then used to calculate the allele frequency and determine the allele frequency distribution of each group. To find out how diverse the study populations are, I calculated heterozygosity and F_{st} values which indicated high level of heterozygosity and population differentiation.

To examine the population structure I performed principal component analysis (PCA) using the identified alleles frequencies. I found most of the variances were explained by the first and the second principal components (PC). The PCA identified three clusters that correlated with linguistic families of the study groups. The first cluster was defined by alleles common to the Nilo-Saharan groups. The second cluster was defined by alleles that are frequent in Afro-Asiatic groups and third cluster was defined by alleles of the Sudanese Arab groups. Interestingly, the division in cluster 3 between Arab and the other groups in the Afro-Asiatic family was not previously seen in SNP data, suggesting there is further substructure within this family.

For the further analysis, I focused on cluster 3 and tried to find whether the identified alleles are on the same chromosome, I estimated haplotype frequencies and compared it to the previous clustering patterns. I found alleles of the most frequent haplotypes are also found in the same cluster and those alleles are informative for differentiation between groups because they have similar pattern of allele frequency distribution. Having identified the informative alleles in cluster 3, I checked whether they are tightly linked by estimating pairwise linkage disequilibrium (LD) between alleles at

(別紙様式 2)
(Separate Form 2)

the three HLA loci. I observed that alleles of HLA-B and HLA-C are in perfect LD most of the times. On the hand, alleles of HLA-A showed relatively lower LD scores; this is expected as LD decays with increase in physical distance along the chromosome.

Furthermore, I decided to trace the ancestry of cluster 3 alleles so I searched for the Arab frequent haplotypes in other population in the Allele Frequency Net Database (AFND), which is the largest database for global distribution of HLA alleles.

Interestingly, the most frequent 2-loci haplotypes (HLA-B and HLA-C) among Arab groups were not found in All Sub-Saharan African population and found mainly in Asia, Europe and North America. More interestingly, I found the most frequent B-C haplotype among Sudanese Arabs (B*51:01-C*15:02) has 4.7% frequency in the Saudi Arabian population.

Finally, finding cluster 3 alleles exclusively outside Africa strongly suggested that the Sudanese Arab groups have experienced gene flow from non-African sources. To further understand the structure of the Sudanese groups in the context of other populations, I integrated data from AFND including populations from Sub-Saharan Africa, Middle East and Europe. PCA on the integrated data sets confirmed the previously seen clustering patterns in our own data. Intriguingly, the last PCA showed Arab groups from Sudan were the closest to all non-African populations, particularly populations from the Middle East like the Saudi one.

In conclusion, this study identified four new HLA alleles and established a map of HLA class I allele and haplotype frequencies in Sudan. Furthermore, a new substructure within Afro-Asiatic groups was found, which separates Sudanese Arabs from non-Arabs. The new substructure seems to be affected by gene flow from the Middle East.

(別紙様式 3)
(Separate Form 3)

博士論文審査結果の要旨

Summary of the results of the doctoral thesis screening

ヒトゲノムには、第 6 番染色体の上に「H L A (Human Leukocyte Antigen ; ヒト白血球抗原)」と呼ばれる、主要組織適合性複合体 (MHC) 遺伝子群が存在するが、これらはクラス I と II にわかれ、それぞれのクラスに複数のきわめて遺伝的変異に富む遺伝子座が存在する。出願者は、自分の出身国であるスーダン内に居住する複数人類集団の集団構造を、H L A クラス I の遺伝子データを用いて、以下の研究を行なった。

本研究に用いられた D N A サンプルは、スーダン医科学技術大学が収集した、8 集団、合計 329 名のものである。これらの集団が話す言語は、アフリカの 3 言語族 (アフロ- アジア、ナイロ- サハラ、ニジェール- コンゴ) にわたっており、また出願者自身が一部集団の試料収集に参加している。次世代シーケンサーを用いて D N A 配列の決定がおこなわれ、配列解析は出願者がすべて担当した。H L A 遺伝子は遺伝子重複によって生じた類似配列が多く、それぞれの座位の塩基配列を決定するのは簡単ではないが、出願者の属する研究室は H L A 遺伝子座を詳細に研究しており、そこで蓄積されているノウハウを用いることができ、クラス I の A, B, C 3 座位について対立遺伝子の確定とハプロタイプの推定をおこなった。これらのデータを主成分分析法で解析した結果、3 クラスタが観察され、それらはほぼ言語分類に対応していた。同じサンプルをゲノム規模 S N P で解析した結果とすこし異なり、スーダン在住のアラブ人が異なるクラスタを形成した。この結果は、遺伝的多型性のきわめて高い H L A 遺伝子のデータが微細な集団構造を推定するのに役立つことを示している。

なお、博士論文には盛り込むことができなかったが、出願者は H L A クラス II 遺伝子の解析も終了しており、審査会ではクラス I とクラス II のデータを合わせた結果も紹介したが、クラス I だけのデータの解析結果と比較して、大きな変化はなかった。

この研究は、現代人のなかでもきわめて多様性が高いことが知られているアフリカ諸集団の中で、研究がまだ十分には進んでいないスーダンの人類集団における集団構造に、あたらしい知見を加えるものである。審査員全員で審査した結果、本大学院における学位授与の水準を満たす論文であると判断した。