

氏 名 西山 久美子

学位(専攻分野) 博士(理学)

学位記番号 総研大甲第 2207 号

学位授与の日付 2020 年 9 月 28 日

学位授与の要件 先導科学研究科 生命共生体進化学専攻
学位規則第6条第1項該当

学位論文題目 The investigation of the signature of selection on genes
associated with dyslexia of Chinese characters

論文審査委員 主 査 教授 颯田 葉子

講師 五條堀 淳

准教授 大田 竜也

准教授 中山 一大

東京大学 大学院新領域創成科学研究科

(様式3)

博士論文の要旨

氏名 西山 久美子

論文題目 The investigation of the signature of selection on genes associated with dyslexia of Chinese characters

漢字ディスレクシア関連遺伝子に対する自然選択の痕跡について

Dyslexia, or reading disability, is found to have a genetic basis, and several related genes have been reported. Because writing systems were developed just a few thousand years ago and used by only a limited number of people before modern times, dyslexic people would have lived without the reading difficulties/disadvantages that are present in modern society. Based on this viewpoint, alleles related to reading ability are expected to be under neutral evolution. Otherwise, if natural selection has acted on such alleles, its target should be traits other than reading ability itself. The question in the study is whether natural selection has acted on the alleles of SNPs that were reported to be risk/non-risk for reading ability. In this thesis, I focused on 15 SNPs that were found to be associated with dyslexia of Chinese characters in previous studies (“core SNPs”, hereafter), and applied two types of summary statistics (*nSL* and 2D SFS-based statistics) to SNP data of East Asian populations, in order to examine whether there is any sign of selective sweep. Because core SNPs themselves are not necessarily causal, and instead, the causal site may be a site that is tightly linked to a reported SNP, I also checked such linked SNPs in this study, considering that they also could be selection targets. The findings of my study are shown in chapter 3.

In my study, first, I conducted a brief study of principal component analysis (PCA) to confirm genetic background of the study populations. Second, as a neutrality test based on extended haplotype homozygosity, I applied *nSL* to the core SNPs. *nSL* did not detect any signatures of positive selection for any of the core SNPs. Third, as a neutrality test based on site frequency spectrum (SFS), I conducted 2D SFS-based statistics, which would be main analyses of my study.

In the first section of 2D SFS-based statistics, I carried out screening of core SNPs using F_c from 2D SFS-based statistics. This examined whether a high linkage disequilibrium (LD) region containing a core SNP (core region) could be under selective sweep. Before applying F_c , I checked the global derived allele count of every core SNP and its linked SNPs, considering that if a linked SNP showed a similar number of derived alleles in EAS but it showed a different number from the core SNP when looking at the global population, the level of polymorphism should be different between the core SNP and such a linked SNP due to the difference in age. I found that three core SNPs had some linked SNPs with global derived allele counts smaller than that

of their core SNPs. I applied F_c statistic to one linked SNP in a younger classes (“younger SNP”) in each of these three cases, as well as core SNPs. Then, Two SNPs remained after this screening ($p < 0.1$): rs17031962 on *GNPTAB* and rs3789228 (younger SNP for rs1091047) on *DCDC2*.

In the second section of 2D SFS-based statistics, I analyzed the two core regions that passed the screening, in order to search for the target site of natural selection. In this analysis, I compared the level of polymorphism around each of the candidate SNPs (core SNP and its linked SNPs) in the core region, using G_{co} from 2D SFS-based statistics. The first case is the core region of rs17031962 on *GNPTAB*. In addition to *GNPTAB*, this region also contains genes of *CHPT1* (partial) and *SYCP3*. I searched for the target site and inferred that the target site could be a SNP (rs3751248), because the level of polymorphism around the SNP was significantly low, and it is located in an open chromatin region. The second case is the core region of rs3789228 on *DCDC2*. This SNP was distinguished as the younger SNP to the core SNP of rs1091047, based on global derived allele count. An SNP (rs12055879) is located in both enhancer region and CTCF binding site and showed significantly low level of polymorphism. Thus, I inferred that the target site could be this SNP. After 2D SFS-based statistics, I investigated the phylogenetic relationship of haplotypes in global populations about the two cases.

In the general discussion of chapter 5, I considered the results of my study. For most of the core SNPs, both *nSL* and 2D SFS did not detect any signatures of selective sweep, then neutrality was not rejected. Because most people were not engaged in reading and writing until recently, the genetic variations that my study focused on were unlikely to be maintained by natural selection, which is consistent with my results. Nevertheless, 2D SFS-based statistics suggested that two core regions could be under selective sweep. In both regions, I found candidate target sites which may have an effect on expression regulation and have been selected. However, which genes these SNPs affect remain unknown. Thus, my results suggest a possibility of genetic hitchhiking, whereby alleles of the reported SNPs may have increased in frequency together with the selected target, which could have functions for other genes and traits apart from reading ability. This scenario is also consisted with unlikelihood that there has been selective pressure on reading ability itself. These results on *GNPTAB* and *DCDC2* were not supported by results of *nSL*. I considered several reasons for this discrepancy in this chapter, and further analyses would be need to confirm the results.

The focus of my study was polymorphisms that were reported to be associated with risk/non-risk for dyslexia. Now, dyslexia is considered as one of neurodevelopmental disorders. Through my thesis, I tried to consider the notion of disorder, and what are “risk/non-risk alleles”. Modern society has introduced public education and demands universal literacy. Primarily, the environment of the modern society likely determines which allele is “risk” or “non-risk” for reading ability.

Dyslexia should basically be a consequence of neutral variation. Even in the case where selection may have acted, the selected trait should be different from reading ability itself. While I should recognize ethical issues by bringing the perspective of evolutionary studies to human society, I hope that my study could provide an example for thinking about some dimensions of human society from the viewpoint of human evolutionary history.

博士論文審査結果

氏名 西山 久美子

論文題目 The investigation of the signature of selection on genes associated with dyslexia of Chinese characters

本論文は、4章から構成されており、それぞれの概要は下記の通りである。

第1章は、研究の背景と本研究の目的を明示している。本論文では、「識字障害 dyslexia」に関連する 15 遺伝子座とその近辺にあるリスク/非リスク SNP(Single Nucleotide Polymorphism: 一塩基多型)に対して自然選択が働いてきたかどうかを、検討することを目的としている。「識字」能力は最近(近代教育制度が普及する)まで多くの人には関わりのないことであったため、「識字」能力が自然選択の標的にはなり得ないとする従来の説がある一方、dyslexia 関連遺伝子への自然選択を検出した論文もあることを示した。特に dyslexia 関連の遺伝子座で自然選択が検出された場合、「識字能力」がそのような自然選択の対象としてありえないことを受けて、この自然選択を説明する二つのシナリオ(neuronal recycling hypothesis と pleiotropy(遺伝子機能の多面性))の可能性がまず考えられることを示した。さらに、この章の最後に、「障害」を人類進化の立場からどう捉えるかについても考察したいと述べている。

第2章は、1章で提示した問題の解決のためにどのような方法を用いたかが述べている。主に二つの要約統計量(nSL と 2D SFS)を用いての解析であること、そして PCA(主成分分析)や系統樹解析を行うことも述べている。用いたサンプルは、1000 人ゲノムプロジェクトのデータベースと、韓国人集団のデータで、データの特徴を見極め、解析により使い分けている。

第3章は、15 遺伝子座とその近辺にあるリスク/非リスク SNP(core SNP) での自然選択の痕跡検出の解析結果である。nSL と 2D SFS を用いた結果、nSL では 15 遺伝子座全てで自然選択の痕跡が検出されず、また 2D SFS では 2つの遺伝子座とその近辺にある SNP を除いて自然選択の痕跡は検出されなかった。このことから「識字障害」に関連する遺伝子座とその近辺の多型の大半は自然選択に対して中立であると結論づけた。一方、自然選択の痕跡が検出された二つの遺伝子座とその近隣領域(*GNPTAB* と *DCDC2*)で、「core SNP」と強く連鎖している他の複数の SNP の中から真の自然選択の標的となった SNP を検出する解析を行った。これは、リスク/非リスクと示唆されてはいるものの、これまでに報告された core SNP がタグ SNP である可能性を否定できないためである。この解析では、2D SFS に基づく要約統計量のみを用いた解析が行われた。真の自然選択の標的を検出する解析の結果、*GNPTAB* では core SNP から 20kb 程度上流の SNP が、*DCDC2* で

は 15kb 程度上流の SNP が、解析した領域 (*GNPTAB*:137kb, *DCDC2*:43kb) の中では、*P* 値で比較すると自然選択の標的である可能性が高いことを示した。さらにこの真の標的と考えられる SNP の生物学的意味を調べたところ、どちらも遺伝子発現に関連するらしいことがわかった。しかし、どの遺伝子の発現を調整しているのかについての情報はまだ得られていない。また、真の標的であろう SNP を持つハプロタイプの共通祖先へ至る時間(TMRCA: Time of Most Recent Common Ancestor) を推定したところ、*GNPTAB* では、およそ 15,000 年、また *DCDC2* では 22,000 年となった。

第 4 章は主に解析結果についての考察を行なっている。解析結果については、*dyslexia* 関連遺伝子座の 15 の core SNP について、2 つの遺伝子座とその近辺にある SNP を除いて、自然選択の痕跡は検出されず、「識字障害」に関連する遺伝子座とその近辺の多型は自然選択に対して中立であると結論づけた。この結論は、「識字」能力が自然選択の標的には、なり得ないとする従来の説と一致するものであった。また、この結論は、自然選択の痕跡を検出した 2SNP の推定 TMRCA がおよそ 1 万年前から 2 万年前に遡ることからも支持された。その上で、自然選択の痕跡が検出された 2 つの遺伝子座とその近辺にある SNP (真の標的 SNP) について、特にこれらの SNP が遺伝子発現の制御に関連する可能性を考察している。また、2D SFS と nSL の二つの方法間の結果に「くい違い」があったことに関して、*recombination* の可能性も含めて検討した。また、*dyslexia* 関連の二つの遺伝子座とその近傍領域にある SNP に自然選択の痕跡が検出されたことについては、第 1 章で紹介した二つの説明に加えて、第 3 の説として「*hitchhiking*」の可能性を論じた。「結論」では、これまで論じてきた結果と考察をまとめるとともに、「障害」を人類進化の立場から捉え直すことの重要性を論じている。

以上のような学位論文の内容は、①「識字障害」に関連した変異について、自然選択の可能性を 1000 人ゲノムのサンプルを使って統計学的に検証した点、②「識字障害」と自然選択の関連について従来の説明に加えて「*hitchhiking*」という説明を提示したこと、さらに付け加えるならば③「識字障害」も含めた「障害」という概念を人類進化の立場からとらえようと試みたことの以上 3 点が特徴的である。この論文で示したこと、あるいは指摘したことは今後既存の分野に新たな考え方をもたらすことにつながりうる。

以上のことから、本論文は総合研究大学院大学の博士学位の授与にふさわしいと判断される。