

様式 C-19、F-19-1、Z-19 (共通)

## 科学研究費助成事業 研究成果報告書



令和元年6月17日現在

機関番号：12702

研究種目：基盤研究(B) (一般)

研究期間：2016～2018

課題番号：16H04821

研究課題名(和文) ヒトの精神活動に関わる遺伝子の進化と集団遺伝学

研究課題名(英文) Evolution and genomic population genetics of genes related to human mental activity

研究代表者

颯田 葉子 (SATTA, Yoko)

総合研究大学院大学・先端科学研究科・教授

研究者番号：20222010

交付決定額(研究期間全体)：(直接経費) 13,300,000円

研究成果の概要(和文)：ヒトの疾病に関わる遺伝子のリスクSNPsの多くが他の霊長類で保存されている祖先型であることが知られている。ヒト特異的な社会的あるいは環境の変化によりこれらのSNPsがヒトの疾病と関わるようになったとの仮説を立て、これまで明らかになっている128の統合失調症関連SNPsをコアとして、連鎖不平衡(LD)、正の自然選択、SNPsの分岐時間等を解析している。東アジア集団で調べたところ、66SNPsで強いLDブロックを、その内15SNPsで自然選択の可能性を検出した。これら15SNPsの分岐時間は半数が最近5万年以降であった。これは考古学的遺物からヒトの脳活動が亢進したと考えられる時期と一致している。

研究成果の学術的意義や社会的意義

現生人類は30万年ほど前にアフリカ大陸で誕生後、およそ5～8万年前に全世界へとその生息域を広げた。これに伴い、洞窟の壁画、遺跡から発掘される装飾品や石器などに見られるようにヒトは文化や技術の面で大きな発展を遂げた。この文化的・技術的な発展にはヒトの脳機能の亢進がともなっていることは想像に難くない。統合失調症は認知や思考の障害であり、その発症に関わる遺伝子は、そのようなヒトの脳機能の亢進に関わると考えられる。そこで本研究では、最近、大規模関連解析で同定された統合失調症に関連する128個のSNPを対象に、ヒトの脳機能の亢進に関与したと思われる遺伝子を検出し、その適応的役割を明らかにする。

研究成果の概要(英文)：It is known that many disease-related SNPs (risk SNPs) in humans are ancestral, shared by non-human primates, implying that human-specific environmental or social changes would be a trigger of the onset of SNP-related diseases. To examine this hypothesis, we studied linkage disequilibrium (LD) surrounding target SNPs, a signal of positive selection on ancestral and derived SNPs and estimated the divergence time of a derived allele group. A recent large-scale association study has identified 128 schizophrenia-related SNPs. Of these, 66 derived SNPs in East Asia showed strong LD blocks. Our newly developed statistical test revealed significant signals of ongoing positive selection at 15 SNPs and the most recent common ancestor of a derived allele group at a half of the SNPs occurred no earlier than 50,000 years ago. This age of positively selected alleles at each target SNP coincides with the time of drastic changes in human brain activities that became evident in archaeological remains.

研究分野：進化生理学、ゲノム集団遺伝学

キーワード：統合失調症関連遺伝子群 脳機能 人類進化 正の自然選択 祖先型・派生型SNP リスク型・非リスク型SNP 環境適応 社会構造の変化

## 様式 C-19、F-19-1、Z-19、CK-19（共通）

### 1. 研究開始当初の背景

私たちの研究グループでは、これまでヒト疾病に関する SNP でリスク型（疾病関連 SNP）が他の霊長類では固定している例（リスク型が祖先型である例）を調べてきた。そのなかで、統合失調症との関連が指摘されている *ST8SIA2* という脳内の神経細胞接着分子（NCAM）にポリシアル酸を付加する酵素遺伝子の promoter 活性について、興味ある結果を得た。*ST8SIA2* 遺伝子の promoter には、統合失調症の発症リスクに関わる 3 つの SNP があり、ヒト集団にはそれらの組み合わせからなる 4 つのハプロタイプが存在する。その中の一つ CGC タイプは、ヒトに特異的に獲得された非リスク型であり、他のタイプと比較して promoter 活性がおおよそ半分に低下している。4 つの promoter タイプの世界各地の集団での頻度を見ると、CGC タイプのみが東アジア集団（EAS）で頻度が上昇していることがわかった。この CGC タイプは、およそ 40 万年まえに誕生していることから、アフリカに起源すると考えられる。一方、この CGC タイプを持つ染色体の promoter の周辺領域およそ 40 kb の塩基配列をヒト 1000 人ゲノムのデータを用いて解析したところ、1) EAS では、promoter を含む 18 kb の領域でのヘテロ接合度が極めて低く維持されていること、2) この 18 kb の領域の外側ではヒト集団の平均的な塩基多様度を維持し 18 kb の領域を区分する明確な境界が存在していることが明らかとなった。これまでの集団遺伝学的解析により、CGC タイプの promoter 領域には正の自然選択が働き、特定の集団でその頻度が増加しつつあると考えられる。このように *ST8SIA2* 遺伝子についての研究は、リスク型の SNP ではなく、非リスク型 SNP に自然選択が働いている例を提示するとともに、自然選択の検出において 1000 人ゲノムのデータを利用した特定の SNP の頻度やその周辺の多様性の検討が有効であることを示した。

### 2. 研究の目的

本研究では、統合失調症に関わる遺伝子を網羅的に調べる。特に、最近同定された 108 個の統合失調症関連遺伝子座の SNP<sup>①</sup>を対象に、*ST8SIA2* と同様にその進化・多型を解析することで、それら SNP およびその周辺領域のハプロタイプの頻度とハプロタイプにかかる自然選択の可能性を精査する。そして、これらの多型の維持機構の解明を通して、その適応的役割を進化・集団遺伝学的観点から明らかにする。統合失調症は民族を問わず、どの集団でも 1%程度の発症率が観察されている。このような現象を説明するために、統合失調症の遺伝子が果たしたヒトの進化への適応的な役割が議論されている<sup>②③</sup>。もしそのようなことがあれば、例えば、統合失調症関連 SNP の中に、Global あるいは Local に頻度を上げている SNP がある可能性があるし、さらに、組換えとの関連で、連鎖不平衡がどの程度の強さで観察されるか、これらの関連 SNP が祖先型か、派生型かという観点での分類なども、進化学的な意味を考察する際に重要な役割を果たす。またこの SNP の頻度上昇には適応的な役割<sup>④</sup>を見いだすことも可能である。

#### <引用文献>

- ① Schizophrenia Working Group of the Psychiatric Genomics Consortium, Biological insights from 108 schizophrenia-associated genetic loci, *Nature* 511: 421-427 (2014)
- ② DF Horrobin, CN Bennett, New gene targets related to schizophrenia and other psychiatric disorders: enzymes, binding proteins and transport proteins involved in phospholipid and fatty acid metabolism, *PLEFA* 60: 431-437 (1999)
- ③ MV Olson, A Varki, SEQUENCING THE CHIMPANZEE GENOME: INSIGHTS INTO HUMAN EVOLUTION AND DISEASES, *Nature Review Genetics* 4: 20-28 (2003)
- ④ B Crespi, K Summers S Dorus, Adaptive evolution of genes underlying schizophrenia, *Proc. R. Soc. B* 274:2801-2810 (2007)
- ⑤ P Khaitovich, HE Lockstone, MT Wayland, TM Tsang, SD Jayatilaka, AJ Guo, J Zhou, M Somel, LW Harris, E Holmes, S Pääbo, S Bahn, Metabolic changes in schizophrenia and human brain evolution *Genome Biology* 9: R124 (2008),
- ⑥ K Xu, EE Schadt, KS Pollard, P Roussos, JT Dudley, Genomic and Network Patterns of Schizophrenia Genetic Variation in Human Evolutionary Accelerated Regions, *Mol. Biol. Evol.* 32: 1148-1160 (2015)

### 3. 研究の方法

本研究計画は、統合失調症関連の SNPs とその周辺領域（100 kb）の多型の集団遺伝学的な解析のために（1）自然選択検出の方法の開発、および（2）開発した方法を用いてのデータ解析の二つのステップか

ら構成される。

(1) 自然選択検出の方法の開発：SFS (Site Frequency Spectrum) と連鎖不平衡 (LD : Linkage Disequilibrium) を組み合わせた  $F_c$  法とこの  $F_c$  法に改良を加え遺伝的組換えの効果を区別でき、また単一のハプロタイプに選択が働く時に観察されるハード選択浄化 (hard sweep) と複数のハプロタイプに選択が働く時に観察されるソフト選択浄化 (soft sweep) を定量的に区別できる 2D SFS (2 dimensional SFS) 法を開発し、computer simulation や既知の自然選択の可能性が示唆されている遺伝子座を用いて、その power を確認した。

(2) 開発した方法を用いての統合失調症関連 SNPs の解析：上記の  $F_c$  法を用いて、プロモータ領域の多型と統合失調症の発症との関連が指摘されている *ST8SIA2* の自然選択を検出した。また、108 遺伝子座にある 103 SNPs について、 $F_c$  法と 2D SFS 法で自然選択の痕跡を調べるとともに、2D SFS 法を用いて、ゲノム中のイントロンに存在するランダムに抽出した 100 SNPs について、ゲノム全体像を検討した。

#### 4. 研究成果

##### (1) 自然選択検出の方法の開発

①  $F_c$  法と 2D SFS 法の特性：いずれの方法も core SNP の派生型 SNP、祖先型 SNP のそれぞれと連鎖している派生型 SNP を D-group, A-group とした時、D-group 内の変異 (IAV : Intra-allelic variation) の空間的、時間的分布を記述する統計量を開発した。この時重要なのは、D-group, A-group に蓄積された変異が de novo なものなのか、あるいは組換えによって 2 次的にもたらされたものなのかを区別することである。このため、 $F_c$  法では D-group 内で頻度の高い連鎖した派生型 SNP をのぞいて、 $F_c$  統計量 (全派生型 SNP の中で、core SNP の派生型に連鎖しているものの割合) を計算する。しかしこの方法では、頻度が高いが de novo で連鎖している SNP まで解析から除外することになり、 $F_c$  統計量そのものが過小評価されることになる。そこで、2D SFS では、まず、core SNP との LD の強い領域 ( $r^2 > 0.75$ ) を決め、その領域の中で、 $F_c$  も含めた統計量を計算する。さらに、各 SNP サイトでの派生型 SNP を、D-group に  $i$  個と、A-group に  $j$  個と区別して  $\varphi(i, j)$  として表すこととした ( $i+j$  は各 SNP サイトでの派生型 SNP の数)。このようにすると  $\varphi(i, 0)$  となった SNP サイトは組換えの影響を全く受けていないサイトとして定義できる。この  $\varphi(i, 0)$  を用いて、 $G_{c0}$  (segregating sites 当たりの組換えを経験していない派生型 SNP の数) と  $L_{c0}$  (全派生型 SNP 中の組換えを経験していない派生型 SNP の割合) という新たな統計量でも IAV を定義することとした。

$F_c$  統計量については、成果を GGS 93:149-161 (2018) に、また  $F_c$  統計量を計算するプログラムを研究室の Web page (<https://sites.google.com/site/sattalab/home>) に掲載。2D SFS については、投稿中である。

② computer simulation を用いての power の検証：図 1A にハード選択浄化 (hard sweep) での、図 1B にはソフト選択浄化 (soft sweep) でのそれぞれの統計量 ( $F_c, G_{c0}, L_{c0}$ ) の power を示した。

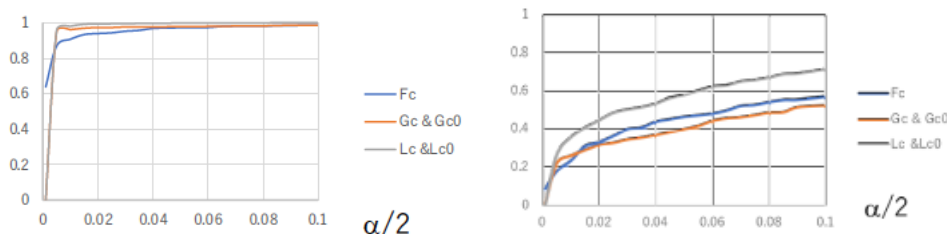


図 1 : A(左) hard sweep, B (右) soft sweep での統計量のパワー (ROC カーブ)。

横軸は false-positive rate ( $\alpha$ ) 縦軸はパワー ( $1-\beta$ )、組換えはなし。選択標的 SNP の頻度はともに 0.6。B では、頻度が 0.1 に達した時点で選択が働き始める。自然選択の強度  $2N_s = 500$ 、突然変異パラメーター  $\theta = 4N\mu = 20$ 、サンプルサイズ  $n = 1000$ 。N は集団のサイズ、 $\nu$  は領域あたりの世代あたりの突然変異率である。帰無仮説は  $\theta = 20$ ,  $n = 1000$  で完全中立とする。

Hard sweep の時には、どの統計量も非常に良いパフォーマンスを示すが、Soft sweep では、組換えがないにも関わらず、power が明らかに低下する。これは、Soft sweep では、選択の標的となる複数のハプロタイプにすでに変異が存在しているために、IAV がかさ上げされた状態になっているためと考えられる。

③ 既知の自然選択の可能性が示唆されている遺伝子座：解析に用いたのは、*LCT*(lactase)、*EDAR*(ectodysplasin A receptor)、*ASPM*(abnormal spindle-like microcephaly)、*MCPHI*(microcephalin-1)、

*ADAMI7*(a disintegrin and metalloproteinase domain17)、*NRG1*(neuregulin-1)の6遺伝子座である。 $F_c, G_{c0}, L_{c0}$ の他に、 $t_D$ (D-group)のTMRCA(Time to the Most Recent Common Ancestor)、 $F_c, G_{c0}, L_{c0}$ のP値を総合的に判断するためのFisher's combined P、その他の自然選択を検出する手法としてnSLで得られたP値を表1に示してある。

これまで様々な方法で自然選択が検出されていた、*LCT*と*EDAR*については、ここで示した新たな方法でもハード選択浄化が検出された。また自然選択に関して論争のある(nSL法ではどちらも否定される)*ASPM*と*MCPHI*については、*MCPHI*でのsoft sweepを検出した。また精神活動に関連する*ADAMI7*と*NRG1*については、*ADAMI7*でsoft sweepを検出したが*NRG1*ではcombined P値が小さいにも関わらず未検出とした。これは*NRG1*ではLDが崩れるのが早く、大きな領域を確保できず解析の用いることができた segregating sitesも13個と少数であったことや、 $F_c, G_{c0}, L_{c0}$ に正の相関があることからcombined P値を保守的に考慮したことによる。

表1 6遺伝子座における自然選択を検出する統計量

遺伝子座 <sup>*1</sup>	<i>LCT</i> (H)	<i>EDAR</i> (H)	<i>ASPM</i> (NS)	<i>MCPHI</i> (S)	<i>ADAMI7</i> (S)	<i>NRG1</i> (NS)
標的 SNP ID	rs4988235	rs3827760	rs41310927	rs930557	rs2709591	rs3924999
$m^{*2}$	511	880	413	770	871	765
LD領域 (kb)	100	65	165	32	26	3
$F_c$ (P値)	0.0054 (0.001)	0.257 (0.013)	0.123 (0.232)	0.103 (0.008)	0.229 (0.02)	0.070 (0.135)
$G_{c0}$ (P値)	1.3 (0.001)	2.3 (0.003)	5.0 (0.521)	4.0 (0.053)	5.1 (0.047)	1.9 (0.132)
$L_{c0}$ (P値)	0.0015 (0.001)	0.0091 (0.001)	0.013 (0.051)	0.013 (0.001)	0.023 (0.003)	0.010 (0.072)
Combined P	< 0.00001	< 0.00001	0.117	0.00005	0.00027	0.0383
$t_D$ (kya)	3.3 ± 0.48	19 ± 2.5	41 ± 7.9	63 ± 13	49 ± 20	17 ± 7.1
nSL (P値)	4.369 (<0.0001)	3.362 (0.003)	-0.081 (0.418)	0.628 (0.263)	2.347 (0.01)	1.895 (0.029)

\*1カッコの中は観察値から判断した selective sweep の mode : Hはhard sweep, Sはsoft sweep, NSはsweepが検出されない。\*2 $m$ はサンプル中の派生型アレル数

## (2) 開発した方法を用いての統合失調症関連遺伝子の解析

① *ST8SIA2*の自然選択の検出：*ST8SIA2*遺伝子のプロモーター領域の3つのSNPのハプロタイプのうちCGCタイプは他のハプロタイプと比べて、プロモータ活性が半減している。このハプロタイプは、特に東アジアで頻度が35%と高く、統合失調症の非リスクタイプであることがわかっている。このCGCを標的SNPとして、 $F_c, G_{c0}, L_{c0}$ をEASで計算すると、完全中立、デモグラフィックなモデルを導入した中立のいずれの帰無仮説も0.1%の有意性で棄却された。CGCハプロタイプ内には、5種類のサブタイプがわかっているが、これらのTMRCAはおおよそ1万2千年になる。この時期は、最終氷期の終了時期、そしてアフリカを出て北回りルートをとった集団と南回りルートの集団がユーラシア大陸の東の端に到達した時期に当たる。CGCという統合失調症の非リスクハプロタイプが東アジアで頻度を増していったのは、こうした、集団の変化に呼応したものかもしれない。

② 2D SFS 統計量のゲノム全体像：ゲノムからランダムに選んだ100個のintron SNPsに対して、 $F_c, G_{c0}, L_{c0}, Tajima's D$ を計算し、また、デモグラフィックなモデルを導入した中立の帰無仮説に対して、それぞれP値を求めた。この解析の際に、標的SNPと $r^2 > 0.75$ となるように領域を決めたが、実際には、 $r^2 > 0.75$ となる領域が1kb未満となるもの、つまり選択浄化作用の痕跡が見えないSNPsが42個もあったため新たなSNPsを追加して総計100SNPsとした。図2には、こうして求めたP値の分布が示してある。この分布の帰無仮説は一様分布なので、カイ二乗検定と、累積分布を作りKolmogorov-Smirnov(KS)検定を行ったところ、 $F_c$ を除いて、いずれかあるいは両方の検定で、帰無仮説は棄却され、 $P < 0.1$ のクラスが過剰であることが示された。さらに、FDR(false discovery rate)10%で多重検定の補正をすると、 $F_c, G_{c0}, L_{c0}, D$ それぞれで、0, 5, 8, 8個のSNPsが帰無仮説を棄却した。つまり $G_{c0}, L_{c0}, D$ の中の多くとも1個のSNPは、誤って帰無仮説を棄却していると言える。残りの4~7SNPsは中立ではない可能性

がある。これらの SNPs がどのような機能をもつ遺伝子にあるのかは現在解析中である。

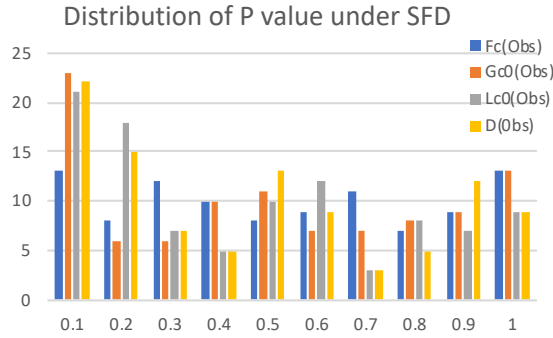


図2 デモグラフィックなモデルを導入した中立の帰無仮説に対して、P 値の分布  
縦軸は SNP の頻度、横軸は P 値を示している。

③ 統合失調症関連の 108 遺伝子座の SNP の解析：108 遺伝子座に統合失調症関連の SNPs は合計で 128 あった。その内訳は、祖先型がリスク型のもものが 60、派生型がリスク型のもものが 43、祖先型が決まらないものが 11、挿入欠失変異型 14 である。ここでは、祖先型が決まっていて、挿入欠失変異型ではない 103 SNPs について解析を進めた。また、リスク型、非リスク型に関わらず、派生型を標的 SNP として解析を進めた。まず、この標的 SNP のうち、東アジア集団で派生型の頻度が 5%以上で  $r^2 > 0.75$  となるように 1kb 以上の解析領域を決められた SNP は 66 だった。この 66 SNPs について、 $F_c$ 、 $G_{c0}$ 、 $L_{c0}$  を計算し、Fisher's combined P で評価したところ、図 3A に示すような分布となった。比較のため、同様の分布を 100 introns で描いた(図 3B)。

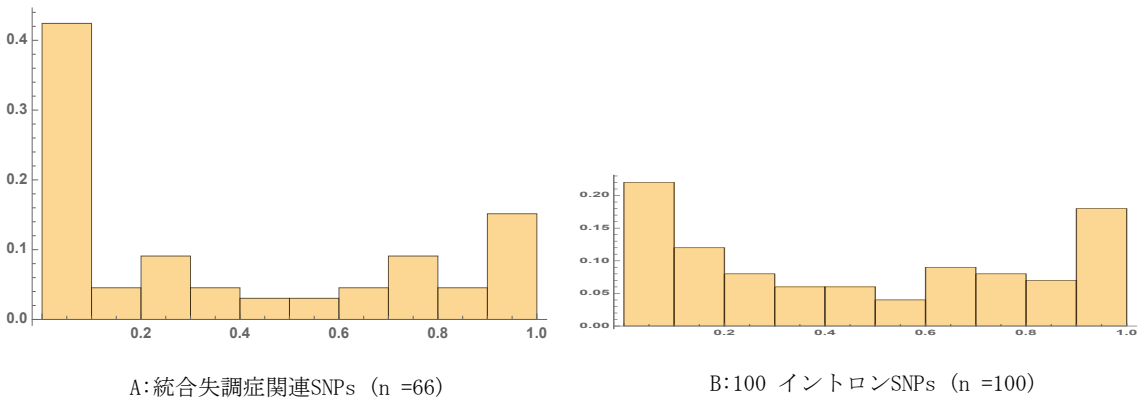


図3  $F_c$ 、 $L_{c0}$ 、 $G_{c0}$  に関する Fisher's S combined P 値のヒストグラム  
縦軸：SNP 頻度 横軸：Fisher's combined P

統合失調症関連 SNP はゲノムの全体像である 100 intron SNPs と比較して  $P < 0.1$  のクラスが過剰である。 $P < 0.1$  のクラスの内訳をみると、 $P < 0.001$  が 7 SNPs、 $0.001 < P < 0.01$  が 9 c、さらに  $0.01 < P < 0.05$  が 6 SNPs あった。さらにこれらの D-group の TMRCA ( $t_b$ ) を求めると、最も古いもので、420 kya、最も新しいもので 8 kya となり、その平均値は、

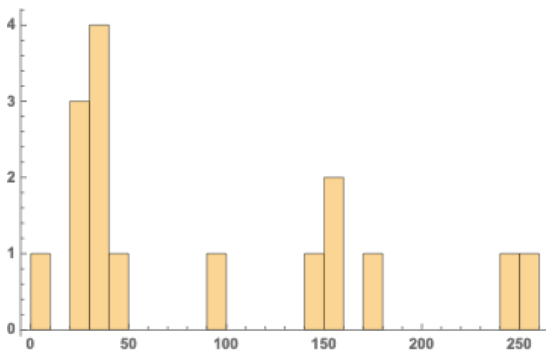


図4 Combined  $P < 0.1$  となった SNP での D-group TMRCA の分布  
縦軸は頻度、横軸は時間 (単位千年)

95 kya、また  $P < 0.001$  のクラスの平均値は、73 Kya となった。この  $t_b$  の分布のピークが 5 万年以下になっていること(図 4) は、考古学的遺物から推定されるヒトの精神活動、特に芸術活動の亢進が見られた時期との一致は、意味のあることかもしれない。また、 $P < 0.01$  となる 15 SNPs 中 11 SNPs がイントロンに存在していた。対象とした 103 SNPs 中では 68 SNPs がイントロンにあったので特に偏りはない。今後は、 $P < 0.01$  となった SNPs について周辺領域も含め精査していく。

### (3) 今後の展望

本研究では東アジア集団の派生型 SNP を標的 SNP として解析を行ったが、今後は東アジア集団の祖先型 SNPs に働く自然選択の検出やアフリカ集団やヨーロッパ集団での同様の解析を行い、現生人類の世界拡散と統合失調症に代表される精神障害の遺伝要因の関係を探っていく計画である。

### 5. 主な発表論文等

[雑誌論文] (計 4 件)

- ① Fujito TN, Satta Y, Hayakawa T, Takahata N (2018) A new inference method for detecting an ongoing selective sweep. *Genes Gen Syst* 93:149-161. 査読有り <https://doi.org/10.1266/ggs.18-00008>
- ② Fujito TN, Satta Y, Hane M, Matsui A, Yashima K, Kitajima K, Sato C, Takahata N, Hayakawa T (2018) Positive selection on schizophrenia-associated ST8SIA2 gene in post-glacial Asia. *PLoS ONE* 13(7): e0200278. 査読有り <https://doi.org/10.1371/journal.pone.0200278>.
- ③ Satta Y, Fujito TN, Takahata N (2018) Nonequilibrium neutral theory for hitchhikers. *Mol. Biol. Evol.* 36: 1362-1365. 査読有り <https://doi.org/10.1093/molbev/msy093>.

[学会発表] (計 12 件)

- ① Satta Y: A new inference method for detecting an ongoing selection. 理論分子進化学の新機軸、遺伝研研究集会 三島、2018年11月
- ② Satta Y: Population genetics of Schizophrenia related genes. ISEGB 2018, Taipei Taiwan, October, 2018
- ③ 早川敏之: 精神疾患関連遺伝子 ST8SIA2 にみる現生人類における社会環境と遺伝子相互作用. 第20回日本進化学会、東京、2018年8月
- ④ 藤戸尚子: 現生人類での精神活動関連遺伝子の進化: シアル酸転移酵素 STX の低活性化による東アジア集団での適応. 第19回日本進化学会、長浜、2017年8月
- ⑤ Fujito TN: Adaptive evolution of mental activity-related STX gene in the out-of-Africa migration. The biology of genomes Cold Spring Harbor, April 2017
- ⑥ Satta Y: "Risk conversion" hypothesis in human evolution. ISEGB 2016, Kaohsiung, Taiwan, November 2016
- ⑦ 藤戸尚子: 出アフリカと精神疾患: シアル酸転移酵素 STX の低活性化による適応. 第88回日本遺伝学会、三島、2016年9月
- ⑧ Fujito TN: Spread of reduced activity of STX promoter in modern humans. SMBE2016. Queensland Australia, July 2016

[その他]

ホームページ等

<https://sites.google.com/site/sattalab/home> に F<sub>st</sub> 統計量を計算するプログラムを掲載した。

### 6. 研究組織

#### (1) 研究分担者

研究分担者氏名: 早川敏之

ローマ字氏名: Hayakawa Toshiyuki

所属研究機関名: 九州大学

部局名: 基幹教育院

職名: 准教授

研究者番号 (8 桁): 80418681

#### (2) 研究協力者

研究協力者氏名: 藤戸尚子

ローマ字氏名: Fujito T Naoko

研究協力者氏名: 高畑尚之

ローマ字氏名: Takahata Naoyuki

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。