胃がんのリスクアレルを通して探る、
日本人の遺伝的な多様性及び集団動態の解析

Study of genetic variation and demographic history of the Japanese from the viewpoint of a risk allele of gastric cancer

岩﨑理紗

博士（理学）

総合研究大学院大学

先導科学研究科

生命共生体進化学専攻

令和 2 年度

（2020）

目次

図・表リスト

第 1 章 序章

第 1 節　胃がんの疫学的特性

がんは現在の人類の主要な死因の一つである。このうち、胃がんの発症率・死亡率は人類全体で見ても高い（図 1）。更に、胃がんの発症率・死亡率は、人類集団によって偏りがあり、東アジアの人類集団では特に高いことが知られている（図 2）[1,2]。また、東アジア集団の中でも、日本人集団での胃がんの発症率は非常に高い（図 3）[1,3]。胃がんは Lauren の解剖学的分類では、intestinal type （IGC）と diffuse type（DGC）に分類される[4]。胃がんのうち、IGC は全体の 60%、DGC は全体の 30%を占めることが知られている[5]。日本では、手術に至った胃がんのうち、DGC はその 55~60%を占めており、重篤な疾患に繋がりうるものの一つとなっている[6]。IGC の胃がんの発症は、ピロリ菌の感染と特に強い関連が示唆されており、発症の頻度が高い地域はピロリ菌の感染頻度が高い地域と分布が重なっているが、ピロリ菌の除去技術の発達に伴い、発症数は徐々に減少傾向にある[2]。一方、DGC の発症地域はより一様分布に近く、日本やアメリカでの発症数は増加傾向にある[2,7,8]。

日本人集団を対象とした GWAS（Genome Wide Association Study）により、DGC の発症リスクと強い関連性（case = 925, control = 1396, $P = 9.4 \times 10^{-8}$）を持つリスク SNP（Single Nucleotide Polymorphism）である、rs2294008（T/C）が報告されている[9]。この SNP は 8 番染色体上の *PSCA* 遺伝子（Prostate Stem Cell Antigen）の開始コドンに位置している[10,11]。*PSCA* 遺伝子は、細胞増殖のパスウェイに含まれる PSCA タンパク質をコードしており、膀胱がん、すい臓がん、肝臓がん、前立腺がん、食道がん、胃がんなど、非常に多種のがんとの関連が指摘されている[11-14]。このうち、いくつかのがんについては、がんの悪化ステージの進行と発現量との関係が調べられている。例えば、前立腺がんでは、がんが進行するほど *PSCA* 遺伝子の発現量が上昇している[12]。逆に、食道がんや胃がんでは、細胞が悪性腫瘍化することによって発現が制御されており、過剰な細胞増殖を抑えている[9,14]。更に、DGC の細胞では、転写レベルで PSCA タンパク質の発現量が低下していることが、RT-PCR[9]及び免疫組織染色法[9,11]によって確認された。以上のことから、胃がんを含む一部の上皮細胞のがんでは、がん細胞の増殖を抑制する機能を持つが、前立腺がんをはじめとするがんではおそらくその逆の機能を持っており、機能的に複雑な遺伝子だと推測されている[9,11]。また、rs2294008 が、祖先型である T アレルを

コードしている場合は、派生型の C アレルをコードしている場合と比較して、翻訳後の PSCA タンパク質の長さは、9 アミノ酸長くなる[10,11]。ルシフェラーゼアッセイで C アレルと T アレルのプロモーター活性を比較したところ、ハプロタイプが T アレルを持つ場合は、発現量が低下したとの報告がある[9]。また、T アレルを持つ *PSCA* 遺伝子がコードする長いタンパク質（long PSCA）は、短いタンパク質（short PSCA）とは機能が異なり、long PSCA は細胞表面に局在するが、short PSCA は細胞内部で発現していることが、免疫染色法によって確認された[11]。以上のことから、DGC での *PSCA* 遺伝子とリスクアレルの関係は、以下のように推測されている。もし T アレルを持つ場合は、発現した long PSCA は細胞表面で細胞増殖のシグナルを受け取り、グリコシル化されることにより細胞増殖のシグナルを伝達する。一方、C アレルを持つ場合、short PSCA は細胞内部に位置しており、特に細胞増殖シグナル伝達の機能を持たないか、または分解されて T 細胞に抗原提示されることで、免疫系ががん細胞を認識し、病態の進行を止める機能がある。このため、T アレルを持つことで、過剰な細胞増殖の誘発による DGC の発達リスクが高まると考えられている[11]。アレル頻度と発症数の関係から、T アレルがヘテロ・ホモのいずれの場合も、その個体の DGC の発症リスクは高くなる[9]。このリスクアレル（T アレル）と DGC 発症との関連性は、韓国人、ウズベキスタン人及びコーカソイドでも示されている[15-17]。以上のことから、rs2294008 の T アレルは、集団を問わず、DGC の発症と強い関連性のあるリスクアレルとみなすことができる。

第 2 節　自然選択の検出方法と解析方法

正の自然選択の検出方法には様々なものが知られている。多くの検定は、遺伝的浮動によるアレル頻度の変化と比較して急速な頻度上昇や、ハプロタイプの中でアレル同士の組み合わせが連鎖不平衡（Linkage Disequilibrium）で保持されている（LD の）長さ、selective sweep による自然選択のターゲットサイトの周辺領域の多様性の低下、あるいは配列情報から作成する genealogy の歪みに着目し、中立下での多様性と比較して、自然選択のシグナルの検出を行う。本研究では、主にアレル頻度ベースの検定、あるいはハプロタイプの LD の長さに着目した検定、塩基多様度の比較を行って自然選択の検出を試みた。検定によって検出方法の特性は異なり、例えば、アレル頻度ベースの検定は、ハプロタイプベースの検定と比較して、より古いシグナルも検出することができる[18]。一方、ハプロタイプベースの検定は、染色体同士の recombination

によって、アレル同士の連鎖関係がまだ大きく崩されていない（連鎖不平衡の状態が保たれている）比較的最近に起こったシグナルを検出することができる[18,19]。そして、塩基多様度[20]は、配列内の多様性の程度を領域ごとに数値化した要約統計量である。

本研究で用いたアレル頻度ベースの検定は、F_{ST}[21,22]や、Tajima's D[23]、normalized Fay and Wu's H[24,25]、PBS[26,27]及び、二次元サイト頻度スペクトラム（2D SFS）である。F_{ST} は集団間のアレル頻度差に基づき、集団分化の程度を調べる指標である[21,22]。また、Tajima's D と normalized Fay and Wu's H はサンプル内の segregating site 数に基づく多型の統計量とペアワイズに比較して得られる統計量の差を用いて、正の自然選択や、集団動態の変化を検出する手法[24,25]である。この二つのテストを組み合わせて用いることで、多型の統計量の差の原因が自然選択か、集団動態の変化であるかを区別するためにも用いられる[28,29]。また、PBS は、ターゲット集団とその近縁集団及びアウトグループのアレル頻度差（F_{ST} 値）を集団間の距離とし、ターゲット集団で、遺伝的浮動によって生じる頻度差からは大きく外れた頻度差を持つサイトを検出する手法である[26,27]。そして、2D SFS は、自然選択のターゲットアレルを持つ配列（D グループ）と、ノンターゲットアレルを持つ配列（A グループ）で、それぞれが持つ派生型のアレルのサイト数を 2 次元マトリクスで表現し、F_c/G_{c0}/L_{c0} という要約統計量で表す手法である[30,31]。ターゲットサイトに正の自然選択が働いている場合は、中立下にある A グループの配列と比べ、D グループでは派生型のアレルを持つサイト数が（selective sweep によって）減少する。これらの統計量は特に正の自然選択が現在進行中で、selective sweep が完了していない状態の、soft/hard sweep のシグナルの検出に用いられる。F_c は、ターゲットアレルが現在の頻度に至る以前に蓄積した D グループにおける派生型アレルの個数を、A グループと比較する統計量である。また、G_{c0} は D グループ固有の派生型アレルの個数を、L_{c0} は D グループ固有の派生型アレルのあるサイト数を、それぞれターゲットアレルが中立と仮定した場合と比較する統計量である。これらの統計量が中立の場合より有意に小さい場合には、正の自然選択が働いている可能性がある。

また、自然選択が働いている領域でのハプロタイプの長さに着目した検定として、EHH（Extended Haplotype Homozygosity）[32]、nS_L（number of segregating sites by length）[33]、H12（Haplotype Homozygosity）[34]を用

いた。これらの解析は、ターゲットアレルを含む領域では、連鎖不平衡にあるブロック（LD ブロック）がより長く維持されていることに基づいた解析である。一度正の自然選択が働き始めると、自然選択の対象であるアレル及び、連鎖の強い周辺のアレルの組み合わせは維持されたまま、短期間で集団中に広がる。この連鎖不平衡状態にある配列が、相同染色体同士の組み換えにより平衡状態に至るよりも短期間で集団中に共有されるため、正の自然選択のターゲットサイトを含む LD ブロックは、平均的なブロックよりも長くなることを利用して正の自然選択を検出する。EHH は、コアとしたサイトの各アレルと連鎖しているハプロタイプのホモ接合度を非コアのサイトとの連鎖と比較し、このホモ接合度が長く維持されるかどうかによって、hard sweep のシグナルを検出する[32]。また、nS_L はコアサイト（派生型と祖先型）と連鎖している SNP によるハプロタイプの長さを比較し、hard/soft sweep のシグナルを検出する正の自然選択の検出方法の一つで、これが有意な負の値をとる場合は、派生型のアレルに対する自然選択のシグナルである可能性がある[33]。H12 は window 毎に含まれるハプロタイプの頻度に基づき、hard/soft sweep のシグナルを検出できる[34]。

最後に、塩基多様度（π）[20]は、ターゲット領域の多様性をその周辺領域、または自然選択の働いていない集団で同じ遺伝子座と比較することで、多様性の低下を評価することで、正の自然選択を検出する。

本研究では以上の手法を用いて、正の自然選択のシグナル検出を試みた。

また、配列の解析方法として、本研究では、ハプロタイプネットワークを用いた。この解析は、ハプロタイプ間の配列を比較し、突然変異を距離として表すことで、サンプルの系統関係を調べる系統解析の 1 種である。配列が似ているハプロタイプ同士は、図中でお互いに近い場所に位置する。このため、配列の系統関係がまだわからない新規のサンプル（例えば発掘された遺骨）から抽出した古代ゲノム配列を用いて、現生種との系統関係を調べたり、あるいは、現生種のどの集団と近縁なのかを調べるために利用される。また、本研究では、上記に加えて、近縁な集団間で共通して自然選択が働いているハプロタイプと、自然選択が働いていない集団も含むハプロタイプで、ネットワークのパターンにそれぞれ違いが出るかを調べるために使用した。

第 3 節　日本人の集団動態

日本に現在居住する集団の、表現型・遺伝子型の多様性を説明する集団動態

について、明治時代以降、様々な説が唱えられてきた。これらの説は「変形説」「置換説」「交雑説」の 3 つに大別され、縄文時代の人類集団と、弥生時代以降の人類集団間に認められる人骨の形態の大きな差異を説明する仮説として提唱された。変形説は、縄文時代の集団が環境変異によって形態が変化したが、縄文系の集団は、文化的な移行を伴いながら連続的に弥生時代以降の集団となり、現代の集団の祖先となったとする説である[35]。また、置換説は、縄文時代から日本に住んでいた集団が、系統的に大きく離れた大陸からの渡来系集団によって駆逐され、主要な人類集団が交代したとする説である[36]。そして、交雑説は、土着の集団である縄文系の集団と、渡来系の集団が交雑し、現代の日本人集団の祖先集団となったとする説である[37]。これらのモデルのうち、現在最も支持されているのは交雑説である。交雑説でも、どの規模でどの由来の集団がどの程度交雑したかについては様々なモデルが提唱されているが、本研究では、埴原によって提唱された「二重構造モデル」[38]に基づいて解析を行なったので、このモデルを紹介する。このモデルは、南方由来の、遺伝的に均一な縄文系人類集団が現在の日本列島に居住し、その後、大陸から稲作と鉄加工技術とともに渡ってきた渡来系弥生集団と交雑することで、現代の日本人集団が生まれたとするモデルである。また、このモデルでは、現代の日本人集団の中の多様性を、渡来系弥生集団との交雑割合の違いによるものとしている。具体的には、交雑が進んだ本土の集団を現代の本土日本人の集団とし、中央から遠く離れた北方や南方に居住しており、渡来系集団と交雑が進まなかった集団を、現代のアイヌ集団や琉球集団として説明している。現在でも、縄文系統集団の由来[39,40]や、大陸由来の集団との交雑回数や交雑の中心地[41]については、まだ議論がある。しかし、「縄文系統と、由来の異なる渡来系集団の交雑」という点や、「交雑比の違いによって日本人集団の中に多様性が生まれた」点については、形態情報からだけではなく、ゲノム情報を用いた多くの研究によって支持されている[40,42-44]。

第 4 節　本研究の目的とアプローチ

DGC のリスクアレルである rs2294008 の T アレルは、1000 人ゲノムプロジェクト（1KGP）での東京在住の日本人集団（JPT）ではその頻度が 0.50 を超える major allele である（0.63）[45]。このリスクアレルが高い頻度を示すことは、1KGP の JPT のみならず、他の報告によっても示されている[9,11,46,47]。しかし、このアレルは、遺伝的に近縁な他の東アジア集団ではほとんどで minor

allele であり、日本人集団とは大きな頻度差がある（0.25（1KGP での漢族集団（CHB））、0.26（台湾バイオバンクプロジェクトでの台湾人集団）、0.50（Bae らがゲノムワイド DNA アレイ解析を行った韓国人集団））[45,48,49]。例えば、JPT-韓国人集団間には 0.13～0.14 の頻度差があるが、もし自然選択がいずれのアレルにも働いておらず、両者の有効集団サイズが一定（例えば 10^4）である場合、共通祖先での初期頻度が 0.001 の時には、集団分岐以降、一方の集団の頻度が 0.63 に至るには平均で 16622 世代かかる。もう一方の集団が 0.50 に至るには、平均で 12254 世代かかるため、頻度差 0.13~0.14 が生じるには、平均 4368~4700 世代（1 世代を 20~30 年とすると約 10 万年）の差が必要である[50]。共通祖先での初期頻度がたとえ 0.10 であったとしても、この頻度差を生じるのにかかる時間はほとんど変わらない。このため、遺伝的に近縁な集団間でアレル頻度に大きな差が生じている場合は、集団の分岐後に、遺伝的浮動が強く影響したか、あるいは正の自然選択のように、何らかの進化的な要因が働いている可能性を示唆している。

このリスクアレルを持つことは、DGC の発症や、発症後の悪化リスクを上昇させる[9,51]ため、適応度に不利な効果をもたらすと考えられる。これまでに、このような生物学的に不利なリスクアレルについて機能解析を試みた研究や、日本人集団の DGC の発症との関連性を指摘する研究は存在するが、何故現代の日本人集団ではリスクアレルの頻度が高頻度に保たれているのか、進化という観点から論じた研究は存在しない。本研究の目的は、このようなリスクアレルが、何故現代の日本人では高頻度であるのかを、集団遺伝学的な観点から明らかにすることである。遺伝的に近縁な集団において、大きな頻度差が生じるには、中立下で遺伝的浮動の影響を強く受けた（例えば日本人集団において集団動態が大きく変化したなどの原因）か、あるいは、このアレルに対して自然選択（例えば特定のアレルに対しての正の自然選択）が働いた可能性が考えられる。そこで本研究では、具体的な手段として、以下の三つのアプローチを用いた。はじめに、頻度差が生じた原因を探るため、i）1KGP のサンプル集団である JPT 及び、遺伝的に近縁な東アジアの集団を用いて、リスクアレル及び、ノンリスクアレルに自然選択が働いている可能性について検討した。この際、検出された自然選択のシグナルに基づいて、ターゲットとなるアレルが集団内に複数あるかどうかについても検討を行い、さらに、この領域で自然選択のシグナルが世界中の他の集団に見られるかどうかや、人類集団間の系統関係を踏ま

えて selection status の変遷についても検討した。次に、JPT と遺伝的に近縁な集団間で、リスクアレルの頻度差がどのような過程で生じたのかについて、日本人の集団動態を二重構造モデルに基づき、ii) コンピュータシミュレーションによって、リスクアレルの頻度の変遷を調べた。最後に、日本人の祖先集団の一つである縄文人のゲノムのデータを用いて、iii) 塩基配列のネットワーク解析を行い、ii)で推定した、リスクアレルを含むハプロタイプの由来に関して、日本人集団成立の歴史と関連づけて議論した。

第 2 章　方法論

解析の概要

本研究は、以下の手順で進めた。まず、遺伝的に近縁な東アジア集団と、JPT で rs2294008 の T アレル頻度を比較し、F_{ST} によってゲノムワイド SNP における分化の程度を調べた。F_{ST} とは、2 集団間でのアレルの頻度差を指標として、その集団の分化の程度を調べる要約統計量であり、頻度差が大きいほど F_{ST} の値は大きくなり最大値 (1) に近づいていく。本研究では、この F_{ST} を用いて JPT とその近縁集団での rs2294008 及び周辺領域の SNP の分化程度を調べた。次に、JPT と、遺伝的に近縁な集団である CHB とで、各種の中立性検定によって、両集団での各アレルにおける正の自然選択の検出を試み、その結果を両集団で比較した。更に、JPT で T アレルの頻度が上昇した原因を調べるため、日本人の集団動態を二重構造モデルに基づきアレル頻度シミュレーションを行った。最後に、縄文系統の人類の古代ゲノムを用いて、JPT とのハプロタイプの系統関係を調べた。

使用したデータ

1000 人ゲノムプロジェクト（1KGP)、phase 3 のヒト集団のゲノムワイド SNP データを利用した[45]。本研究では、1KGP の集団の定義に従い、東アジアの subpopulation として、JPT（n=104)、CHB（n=103)、南方出身の漢族（n=105）(CHS)、中国居住のダイ族（n=93）(CDX) 及びベトナム人（n=99）(KHV) の 5 集団を使用した。さらに、アフリカの subpopulation として、メンデ族（n=85）(MSL)、ヨルバ族（n=108）(YRI)、エサン人（n=99）(ESN)、ガンビア人（n=113）(GWD)、ルヒヤ人（n=99）(LWK)、アフリカ系カリブ人（n=96）(ACB) 及びアメリカ系アフリカ人（n=61）(ASW) の 7 集団を使用した。metapopulation としては、JPT と CHB を除く東アジア（n=297）(non JPT/CHB EAS)、ヨーロッパ（n=503）(EUR)、南アジア（n=489）(SAS)、アメリカ（n=347）(AMR)、アフリカ(n=661) (AFR) の 5 集団を使用した。各 SNP の位置情報は GRCh37 に従い、各アレルの祖先型・派生型の区別は 1KGP の定義に従った[45]。また、共同研究者の Kim Hie Lim 博士（シンガポール・Nanyang Technological University）より、韓国の 1000 人ゲノムプロジェクト（1KKOR）[52]でシーケンスされた韓国人（n=151）(KOR) 及び、一部に JPT と同一個体を含む日本人（n=35）(JPN) のゲノムワイド SNP データの提供を受け、2DSFS を用いた自然選択の検討や、1KGP の東アジア集団 (EAS)

と合わせたデータセットでのネットワーク解析を行った。

F_{ST} の算出方法

Hudson の F_{ST} によって JPT と東アジアの各集団間の、ゲノムワイド SNP の遺伝距離の算出を行った[21,22]。この計算では、比較している集団間で 3 種類以上のアレルを持つ SNP を除外した。また、balancing selection などの特殊なメカニズムによって進化している遺伝子を多く含む *MHC* 領域（chr6: 25726291-33368333）は計算から除外した[53]。各 SNP から算出した F_{ST} の値は、qqman[54]を用いて作成したマンハッタンプロットによって比較した。
また、rs2294008 を含み、JPT で連鎖不平衡の状態にある 21.9 kb の領域（chr8: 143752235-143774193）の分化程度を 8 番染色体上の同じ大きさの他の領域と比較した。分化程度の評価は、21.9 kb あたりに含まれる SNP の F_{ST} を平均した。また、隣り合ったブロックが重ならないように sliding window analysis を行った。

LD ブロックの定義

rs2294008 を含み、連鎖不平衡の状態にある領域を *D'*の値によって、LD ブロックと定義した[55]。領域の定義には Haploview[56]で使用されている基準（*D'* > 0.98）[57]を使用し、マイナーアレルの頻度が 0.05 未満の SNP は LD ブロックを決定する際の対象外とした。

自然選択の検討

(1) ハプロタイプの LD に基づいた正の自然選択の検討（EHH, nS_L, H12）
rs2294008 の各ハプロタイプの LD 領域の長さや頻度に基づいて、正の自然選択の検出を 3 種類の方法（EHH, nS_L 及び H12）で試みた[32-34]。まず、EHH によって、各アレルを含むハプロタイプ同士のホモ接合度の比較を行い、自然選択の検出を試みた。rs2294008 及び、F_{ST} の順位の高い 5SNP の各アレルをコアとするハプロタイプを定義し、同じサイトの 2 種類の塩基（祖先型と派生型）をコアとするそれぞれのハプロタイプ同士でホモ接合度を計算し比較した。この際、マイナーアレルの頻度が 0.05 未満の SNP は、EHH に用いるハプロタイプを定義する際の対象外とした。コアアレルに自然選択が働いている場合、コアアレルを中心としたハプロタイプの LD の長さは中立下にあるアレル（非コアアレル）のそれより長くなることが期待される。本研究では、自然選択の対象となるアレルのハプロタイプの広がりを検出するため、合計 400 kb の範囲（143563622-143963622）で、ハプロタイプのホモ接合度の低下のパターンを

比較した。また、nS_Lによって、rs2294008 の C アレル（派生型）及び、T アレル（祖先型）のハプロタイプでの LD の長さの比をとった。同様に、周辺のF_{ST}の高い SNP の派生型及び祖先型のアレルに対しても、ハプロタイプの LD の長さの比を求め、これらを他の（中立な）サイトの派生型・祖先型の比と比較した。 ハプロタイプの中心になる SNP が中立の場合、頻度が同じ他の SNP と平均的には同程度のハプロタイプの LD 長を持つと考えられるため、同じ頻度を持つ SNP 間で比較を行った。ソフトウェアには selscan を用い[58]、8 番染色体の segregating sites からマイナーアレルの頻度が 0.01 未満の SNP やアレル数が 3 以上あるものは解析から除外し、JPT では 303,438 sites, CHB では 377,335 sites を対象として解析した。更に、H12 によって、rs2294008 を含む LD ブロックのハプロタイプ頻度を計算しこの頻度に基づいたホモ接合度を、同じ染色体上の同じ数の中立に保たれているであろう SNP を含む他の領域と比較した。比較には隣り合った window は重ならないように sliding window analysis を行った。1 window のサイズは、rs2294008 を含む LD ブロックに含まれる、125 SNP（JPT）または、124 SNP（CHB）とした。自然選択のターゲットになるサイトを含むハプロタイプは急激に頻度が上昇するため、ホモ接合度が中立で期待されるよりも高く保たれる。

(2)サイト頻度スペクトラムに基づいた自然選択の検討（Tajima's *D* 及び normalized Fay and Wu's *H*）

rs2294008 を含む LD ブロックについて、Tajima's *D* 及び、normalized Fay and Wu's *H* によってサイト頻度スペクトラムに基づく多型量を比較した[23-25]。ソフトウェアにはどちらも Dnasp v6.0[59]を用い、祖先型の不明なサイトは検討から除外した。

(3)塩基多様度（π）の比較

JPT と CHB で、rs2294008（T/C）の各アレル及び、各アレルと連鎖する配列（以下、T ハプロタイプ及び C ハプロタイプ）の塩基多様度を、Nei のπ[20]によって算出し（π_T及びπ_C）、z 検定によってこれらに有意差があるかどうか検討を行った。使用した領域は、rs2294008 を含む LD ブロックと、その上流・下流の各 100 kb（143652235-143752234 及び 143774194-143874193）で、これらのπは、Dnasp v6.0 を用いて算出した[59]。また、分散の計算には、

Takahata *et al.*の方法[60]を用いた。P-value を算出し、False Discovery Rate（FDR）を制御するため、Benjamini-Hochberg 法によって Q-value に補正し[61]、危険率 5%で有意性を確認した。この際、selective sweep の影響によるπの低下をより正確に評価するため、rs2294008 を含む LD ブロックに組み換えの確認された配列を CHB から 2 本除外したが、JPT ではそのような配列は確認されなかった。

また、C ハプロタイプに確認された 2 種類のサブハプロタイプ（後述）についても、サブハプロタイプごとにπを算出し、サブハプロタイプ間で多様性に差があるかどうか比較を行った。また、このπに基づいて、サブハプロタイプ間の分岐年代の推定も行った。mutation rate には、0.5×10^{-9}/site /year を用いた[62]。

(4)二次元サイト頻度スペクトラム（2D SFS）を用いた検定

正の自然選択を受けていると考えられるアレル（ターゲットアレル）を持つ配列（以下、D グループ）に蓄積している変異（IAV: intra-allelic variability）の量を、アレルの多様性の程度として、2D SFS（Φi,j）及び、F_c、G_{c0} 及び L_{c0} を利用して評価し、正の自然選択の検定を行った[30,31]。F_c、G_{c0} 及び L_{c0} は、中立下で作成したそれぞれの要約統計量の帰無分布と比較し、有意性を検討した。帰無分布は、中立条件下で各集団の集団動態[63]を反映させ、ms[64]で最小 1000 回のシミュレーションを行うことによって作成した[30,31]。算出した P-value は、FDR を制御するため、Benjamini-Hochberg 法によって Q-value に補正し[61]、有意性を確認した。本研究では、2 種類以上の要約統計量で Q-value が有意水準 5%未満のシグナルを検出した場合を自然選択が有効であると定義した。更に、1 統計量のみで自然選択のシグナルが検出されたマージナルなサンプルについては、統計量間の共分散の影響を除外した combined P[65]によってシグナルの強さを評価した。また、本検定に使用した領域は、ターゲットサイトである rs2294008 との連鎖が強い領域のみ（$r^2 > 0.75$）とし[31]、JPT では 143755915-143770914、CHB では 143755876-143771875 の領域を使用した。

また、C ハプロタイプを、rs2976391 （C/A） と rs2978983 （A/G）の 2 種類のアレルの組み合わせに着目して 2 種類のサブハプロタイプに区分した。各サブハプロタイプについても、2DSFS を用いて、それぞれのハプロタイプの

IAV を評価した。

また、各ターゲットアレルに対し、要約統計量の一つである t_D を用いて、D グループの IAV に基づき、D グループの配列の TMRCA（現在から共通祖先に至るまでの時間の平均値）を算出した[31]（図 4）。ここで計算に用いる IAV は、selective sweep の影響で star-like になった系統樹上に生じる、互いに独立な突然変異の数として表現される。このため、IAV は純粋に分子時計に従った量になり、自然選択による系統樹の歪みは IAV の量に特に影響を与えない。また、ターゲットアレルが自然選択下にある場合は、selective sweep によって（中立である場合と比較して）D グループ内の IAV が減少するため、その t_D は中立下にある場合よりも短くなる。この値は、特にその集団でターゲットアレルに正の自然選択が働いている場合は、その自然選択が遅くともいつまでに開始していたかを示す[31]（一方、自然選択が始まった時期の上限は、ターゲットアレルとノンターゲットアレルの分岐年代として表現される）。mutation rate には、0.5×10^{-9} /site /year を用いた[31,62]。

(5)集団特異的な自然選択の検出（PBS）

CHB に対し、PBS によって C アレルが集団特異的な正の自然選択が検出されるかどうかを検討した。使用した集団は、JPT 及び、アウトグループとして EUR を用い、ゲノムワイドな SNP のうち、いずれの集団でもマイナーアレルの頻度が 0 にならない（いずれのサンプル集団の中でも多型がある）サイトのみを使用して PBS 値を算出した[26,27]。

現代の集団でのサブハプロタイプのネットワーク解析

1KGP の EAS に加えて、KOR と JPN のゲノム配列を用いてサブハプロタイプごとのネットワーク解析を行った。まず、各サンプルから、rs2294008 を含む LD ブロック内に組み換えのある配列を除外した。次に、各 VCF ファイルで多型のあるサイトを調べ、一部のデータセットにしか多型が存在しないサイトは missing data（N）として、network（v. 5.0.1.1）[66]を用い、median-joining 法によって各サブハプロタイプのネットワークを作成した。

日本人集団の集団動態を反映した forward simulation

Wright-Fisher モデル[67]に従い、半数体でのアレル頻度シミュレーションを行って、以下に示す集団動態モデルのもとで、T アレルが中立下でも rs2294008 のように 2 集団間で高い分化を示すかどうかを調べた（図 5）。自然選択に関す

る条件は、(i) どちらのアレル（C と T）も、中立である場合及び、(ii) シミュレーション上の CHB 系統で、C アレルに対して正の自然選択がある場合の 2 つを検討した。

(1) 集団動態モデル

日本人集団の集団動態モデルの一つである「二重構造モデル」[38]を用いて、アレル頻度シミュレーションを行った。縄文系人類集団は、考古遺物の証拠から、遅くとも 16000 年前には既に日本列島に居住していたことが知られている[68]。ここに、大陸より農耕技術を持った大陸由来の人類集団（弥生系人類集団）が渡来し、約 2500 年前に弥生時代が開始した[68]。この二つの人類集団の交雑が現代の日本人集団の遺伝的特徴を形成したと考えられており、この交雑を伴う集団動態モデルは現代のゲノムデータ及び、古代ゲノムを用いた先行研究から支持されている[40,42-44]。以上のことから、二重構造モデルに則ったモデルを構築した。

縄文系人類集団（JMN）及び、大陸の人類集団の祖先集団（A_CNT）が共通祖先から分岐し、両集団は *t1* 年隔離された状態で世代交代を行う。その後、JMN と、A_CNT が r : (1-r)の割合で一代限りの交雑を経験した集団を simJPT とし、交雑を経験しなかった A_CNT 集団を simCHB とする。*t2* 年後、simJPT 及び simCHB 間の F_{ST} を計算し、rs2294008 の F_{ST} 値（観察値）以上の値を持つ場合のパラメータの組み合わせを調べた。*t2* 年前の交雑を除いて、遺伝的交流については簡便のために特にモデルに含まないものとした[40,42-44]。

本モデルでは、どちらのアレル（C と T）も中立の場合は 4 つのパラメータ（*t1*、N_{JMN}、N_{A_CNT}、r）を使用した。まず、*t1* の範囲は、縄文系人類集団と弥生系人類集団の分岐年代を考慮した。縄文系人類集団の分岐はパプアの人類集団の分岐（51,000 年前[69]）よりも遅く、現生の東アジアの人類集団の分岐よりは早い[44]と推定されている。また、先行研究による分岐年代推定の結果[40,43-44,70]や、考古学的な年代推定の結果（~16000 年前[68]）よりは集団の分岐は早くなる点を考慮し、*t1* は 17500 年前～47500 年前の範囲で、5 つ（それぞれ 875、1125、1375、1875、2375 世代に相当）から 1 つをとるものとした。なお、本研究では、現生の狩猟最終民族での女性の第一子出産年齢の平均値[71]に基づき、1 世代時間を 20 年と設定した。この数値は男性と女性のそれぞれの世代時間の平均値（28.6 年）[71]よりも短いため、一定時間（年単位）内の世代数が多くなり、遺伝的浮動によってアレル頻度の変化は大きくなる。

次に、N_{JMN} 及び N_{A_CNT} は JMN 及び A_CNT の集団サイズを表し、それぞれの子孫である集団（JPT と CHB）の推定されている有効集団サイズ[43]を上限として、N_{JMN} は 500、1000、2000、4000、8000、10000、12500、15000 から 1 つを、N_{A_CNT} は 1000、2000、4000、8000、16000、25000、30000 から 1 つをとるものとした。また、それぞれの集団の環境収容力を考慮し、N_{JMN} は常に N_{A_CNT} 以下の大きさとした。最後に、交雑時の JMN のゲノムの割合を示す r の範囲を選んだ。このパラメータは、JPT のゲノムにおける JMN の割合[40,42-44]を考慮し 0.4、0.2、0.1 から 1 つをとるものとした。一方、交雑後から最終世代に至るまでの時間（*t2*）、simJPT の集団サイズ（N_{simJPT}）、simCHB の集団サイズ（N_{simCHB}）は固定値とした。*t2* は弥生時代の開始[68]から、現在に至るまでの 2500 年（125 世代）とした。また、N_{simJPT} 及び N_{simCHB} は、本研究のモデルに最も近い Nakagome らのモデルで推定された数値を使用し、それぞれ 12,824 及び 29204 とした[43]。可変パラメータの組み合わせは総数 570 種あった。JMN と A_CNT の共通祖先での T アレルの初期頻度（f_i）として、0.1 から 0.9 までアレル頻度を 0.1 刻みで設定した。1 つの f_i につき最低 10,000 回のシミュレーションを行ったので、1 つのパラメータの組み合わせに対して最低 90,000 回の頻度シミュレーションを行ったことになる。

simCHB の系統での自然選択を含む場合は、中立下でのシミュレーションに用いた上記の 4 つの可変パラメータと 3 つの固定値に加え、正の自然選択の強さを示す値（選択係数）である s を可変パラメータとした。$2 \times N_{simCHB} \times s$ が、1、10、50、100 から 1 つをとるものとし、JMN での交雑の世代を含む 126 世代の間 simCHB の系統でアレル頻度に影響を与えるものとした。

最終的にパラメータの組み合わせの総数は、2 つのアレルが中立な場合は 570 通りとなった。また、simCHB 系統で C アレルに対して正の自然選択をモデルに加えた場合は、総数 2280 通り（570 通りのパラメータの組み合わせ ×4 通りの選択係数）のうち、KS test を通過した（詳細は後述）1640 通りの、合計 2210 通りである。パラメータの組み合わせそれぞれについて T アレルの頻度をシミュレートし、交雑後、*t2* 世代経過した simJPT と simCHB 間で F_{ST} を計算し、下記の方法で適切なパラメータの組み合わせを選択した。

(2) KS test 及び、F_{ST} 値による適切なパラメータの組み合わせの選択

本シミュレーションからは、同一のアレルが JMN と A_CNT の両方で固定した

場合や、同一のアレルが simJPT と simCHB の両方で固定した場合は有効なシミュレーションからは除外した。また、F_{ST} が負の場合は、F_{ST} をゼロとして扱い、有効なシミュレーションとして数に加えた。また、上記のパラメータの組み合わせによっては、F_{ST} の分布が実際の F_{ST} の分布とは大きく異なるものが生じうるため、以下の手法でそのようなものを除外し適切なものを抽出した。まず、シミュレーション後、90,000 個の F_{ST} からヒストグラムを作成した。この際、f_i の割合によって F_{ST} 値の重み付けを行った。fi の割合は以下の手順で決定した。まず、f_i の割合は、「SNP の突然変異率は大きく変わらないため、JMN と A_CNT の共通祖先での fi の割合と、現在の派生型のアレルの割合は大きく変わらない」と仮定した。次に、1KGP の JPT の 8 番染色体で、派生型が判明している 570,129 SNP から、派生型のアレル頻度を推定した。これらを 0.1 刻みの 9 つの bin に分けた（各 bin の割合は、0.5589、0.0984、0.0734、 0.0623、0.0494、0.0436、0.0360、0.0342 及び 0.0438）。次に、現代の JPT と CHB 間の F_{ST} で作られたヒストグラム分布と、570 種類のシミュレーションによって作成したヒストグラム分布を比較した。SNP のほとんどは中立であり、自然選択下にある SNP は全体の中ではわずかであるため、ゲノムワイドな SNP から作成される F_{ST} の分布は、完全に中立下で行ったアレル頻度シミュレーションによる F_{ST} の分布と近くなると考えられる。そこで、現実の F_{ST} の分布とシミュレーションによる F_{ST} の分布を scipy（python package, v 0.19.1）の Kolmogorov-Smirnov 2 sample test（以下、KS test）によって比較し、F_{ST} のヒストグラム分布と有意に異なるヒストグラム分布を作るパラメータの組み合わせを、p-value < 0.05 を基準に除外した。この結果残った 410 種に対して、「合計幾つの組み合わせで F_{ST} の最高値が rs2294008 の F_{ST} の観察値（0.2547）を超える SNP が（1 回以上）生じるか」を調べた。更に、この条件を満たす組み合わせに対して、(i)「$F_{ST} > 0.2547$」かつ（ii)「simJPT の方が simCHB より T アレルの頻度が高く」、かつ（iii)「simJPT の T アレルの頻度が 0.62（JPT での rs2294008 の頻度－2σ、$\sigma = 0.3 \times 10^{-2}$）を超える」3 条件を同時に満たす SNP が何回出現するかを調べた。頻度の標準偏差（σ）は、二項分布を用いて求めた。正の自然選択が simCHB の系統に起こることを仮定したシミュレーションの場合は、総数 2280 通りのパラメータの組み合わせのうち、中立下のパラメータの組み合わせで KS test をクリアした 410 通りのパラメータの組み合わせに対し 4 通りの選択係数を仮定した合計 1640 通りのパラメータに対して、上

記の 3 条件を同時に満たす SNP の出現する回数を調べた。

更に、韓国人集団を大陸由来の渡来系集団の子孫とみなした場合でも、CHB と同様に、rs2294008 と頻度や F_{ST} 値が似ている SNP が出現するかどうかを調べた。さらにこの時の、JMN の交雑前の最終世代の T アレル頻度の分布を調べた。韓国人集団の current/recent effective population size を推定した先行研究では、それぞれ 9457、20165 と推定された[72]。この値は、Nakagome *et al.* の CHB の有効集団サイズの推定値（29204）[43]とは大きく変わらないため、simCHB の集団をシミュレーション上の韓国人集団として扱った。rs2294008 と頻度や F_{ST} 値が似た SNP の探索には、(i)「$F_{ST} > 0.02342$」かつ (ii)「simJPT の方が simCHB より T アレルの頻度が高く」、かつ（iii）「simJPT の T アレルの頻度が 0.62（JPT での rs2294008 の頻度－2σ）を超え、0.63（JPT での rs2294008 の頻度＋2σ）未満」の 3 条件を同時に満たすものが何度出現するかを調べ、r との関係を調べた。(i)の F_{ST} 値は、KOR と JPT の rs2294008 のアレル頻度・染色体の本数より算出した。

古代ゲノムを用いたハプロタイプネットワーク解析

現代の JPT 及び CHB に見られる T ハプロタイプと、大陸由来の人類集団と交雑する前の縄文系統の集団のハプロタイプの関係を調べるため、渡来系集団と交雑する前の縄文系統である、2720~2418 年前の愛知県の伊川津の縄文系統のサンプル（IK002）[73]と、3960~3550 年前の礼文島の船泊の縄文系統の 2 個体のサンプル（FUN5 及び FUN23）[40]を用いて、ハプロタイプネットワークによる解析を試みた。まず、現生人類（JPT と CHB）の 21.9kbLD ブロック部分の配列を用いて、現生人類のハプロタイプの種数と頻度を確認した。ハプロタイプネットワークの作成には、network（v. 5.0.1.1）[66]を用い、median-joining 法を使用した。次に、チンパンジーゲノム（Pan tro 3.0）から、ヒトの 21.9 kb の LD ブロックをクエリとして nucleotide blast によって相同な配列を同定した。その結果、チンパンジーの 8 番染色体の 145391694-145412716 はヒトの 21.9kb の LD ブロック配列とオーソログであることが確認できたのでこれをアウトグループの配列として使用した。最後に、上記のサンプルと、下記でクオリティコントロールを行った古代ゲノムで、ハプロタイプネットワークを作成した。この時、network の segregating サイト数制限のため、ハプロタイプの定義には、missing data を含むサイトを除外する complete deletion 法を用いた。

古代ゲノムのクオリティコントロール

さらに、古代ゲノムの配列データのうち、信頼性の高いサイトのみを残して以下の基準でフィルタリングを行った。IK002 は、まず GATK（v. 3.8）内の、HaplotypeCaller の gVCF モードによってクオリティの低いサイト（mapping quality < 20 または base quality < 10）を除外し、rs2294008 のアレル情報が残るように、depth > 2 かつ genotype quality > 1 となるようなサイトのみを残した。更に、サンプルの平均カバレッジが 8 未満の場合の処理[73,74]として、バイアレリックサイトではよりカバレッジの高いアレルを選択して、1 倍体としての配列を再現した。FUN5 及び FUN23 は、HaplotypeCaller の gVCF モード及び、VQSR 値によってクオリティの低いサイトを除外し、1KGP に存在するサイト（ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/accessible_genome_masks/20141020.strict_mask.whole_genome.bed）のみを残した。更に、平均カバレッジが 8 を超える FUN23 は genotype quality > 30 かつ depth > 30 のサイトを残し、平均カバレッジが 8 未満の FUN5 は rs2294008 のアレル情報が残るように、genotype quality > 20 となるサイトを残した。FUN23 はサンプルとしてクオリティが高かったため、2 倍体として扱い、FUN5 は IK002 と同様の処理で 1 倍体として扱った。最後に、シーケンスエラーを除外するため、infinite site モデルに基づいて、古代ゲノムで各サンプルだけに限られる変異を持つサイトを除外した。

第 3 章　結果

第 1 節　F_{ST} によるアレル頻度の評価

rs2294008（chr8: 143,761,931）の T アレルは、遺伝的に近縁な集団である様々な東アジアの集団の中でも特に JPT で高く、他の集団と比較しても大きな頻度差がある（表 1）。そこで、JPT と 1KGP で最も遺伝的に近縁な CHB 間で、14,653,076 個のゲノムワイド SNP に対し、rs2294008 の分化の程度を F_{ST} を用いて調べた（図 6）。その結果、rs2294008 の示す F_{ST} 値の順位は、ゲノムワイドな SNP の中でも 32 位と高く、JPT と CHB の間で非常に大きな頻度差があることが分かった（F_{ST} = 0.2547（表 1・図 6・図 7））。ゲノムワイド SNP のうち、F_{ST} 値の第 1 位を示すのは、rs2294008 の近傍にある rs2976394（F_{ST} = 0.2669, chr8 : 143763622）であった（表 2）。そこで、rs2294008 と連鎖不平衡の状態にある SNP も同様に大きな F_{ST} 値を示すかどうか、rs2294008 の作る LD ブロック（chr8: 143752235-143774193、D' > 0.98）に含まれる SNP を調べた（図 8）。この LD ブロックは JPT では 125 SNP で構成されていたが、F_{ST} 値の 50 位までのランキングの中にこの LD ブロック中の SNP がいくつ含まれるかを調べたところ、rs2294008 を含む 49 SNP が含まれていることがわかった（表 2）。50 位までのうち、唯一 LD ブロックの外に位置していた rs2717562（2 位 ; chr8: 143776668）も、rs2294008 の下流およそ 15 kb に位置しており、特に rs2294008 とは完全連鎖の関係にあることがわかった。rs2294008 を含む LD ブロックの長さは 21.9 kb であったため、この長さを元に、8 番染色体を 21.9 kb のブロックに分け、ブロックごとの F_{ST} 値を調べたところ、rs2294008 を含む領域は最も分化が進んでいた（F_{ST} = 0.0975; 6510 block 中 1 位、図 9）。以上のことから、rs2294008 及びその近傍には、非常に大きな頻度差を持つ SNP が集中しており、進化的になんらかのプロセスによってこの遺伝的な分化がもたらされた可能性が示唆された。

第 2 節　中立性検定による正の自然選択の検討

遺伝的に近縁な集団間での、rs2294008 及び近傍の SNP での分化が、正の自然選択によるものであるかどうか、以下の二つの可能性を検討した。(i) C アレルに CHB では正の自然選択が働いている。(ii) T アレルに JPT では正の自然選択が働いている。このいずれか、または両方である可能性について、各種の中立性検定を用いて検討した。

本研究で使用した中立性検定は、その性質によって以下の二種類に大別され

る。(1) EHH、nS_L 及び H12 の属するハプロタイプの LD の長さやその頻度に着目した、ハプロタイプベースの検定と、(2) Tajima's D、normalized Fay and Wu's H、PBS の属するサイト頻度スペクトラムベースの検定である。(1) の中立性検定の結果、rs2294008 をコアとした EHH では、どちらの集団の C 及び T アレルいずれのアレルにも、有意な正の自然選択のシグナルは見られなかった（図 10（a））。同様に、rs2294008 と同じ LD ブロックに所属し、F_{ST} 順位が 2 位を除く 6 位までの 5SNP（F_{ST} = 0.2608~0.2669）のいずれのアレルでも正の自然選択のシグナルは JPT・CHB のいずれの集団にも検出されなかった（図 11）。更に、nS_L では、rs2294008 には、hard sweep や soft sweep のシグナルは検出されなかった（nS_L = 0.4063、図 10（b））。ただし、CHB で、21.9kb の LD ブロックに属する SNP のうち、biallelic な上位 10 SNP の nS_L の値を調べたところ 10 SNP のうち 6 SNP（rs1045574、rs2976396、rs1045605、rs2976398、rs2920292、rs2572910）では負の値を示し、その P 値はマージナルな値（nS_L = -1.978 ~ -2.106（0.01 < P < 0.05））であった。このことは、一部の派生型のサイトに対して弱い selective sweep のシグナルが検出されたことを示す。しかしながら、このような SNP を含む LD ブロック単位で H12 によってハプロタイプ頻度を調べたところ、いずれの集団においても、特に有意なシグナルは検出されなかった（JPT : H12 = 0.118（P > 0.05）、CHB : H12 = 0.126（P > 0.05）、図 10（c））。

（2）のサイト頻度スペクトラムベースの検定を行った。用いた領域は、上の解析と同様の 21.9 kb の LD ブロックである。Tajima's D の検定の結果、いずれの集団からも、有意な正の自然選択のシグナルは検出されなかった（JPT : Tajima's D = 2.031（0.10 > P > 0.05）、CHB : Tajima's D = 0.745（P > 0.10））。一方で、normalized Fay and Wu's H の検定からは、CHB に高頻度のアレルが蓄積されていることを示す弱いシグナルが検出された（JPT : normalized Fay and Wu's H = -0.319（P > 0.10）、 CHB : normalized Fay and Wu's H = -2.174（P < 0.05））。normalized Fay and Wu's H で検出できる高頻度のアレルの蓄積のシグナルは、正の自然選択、または集団動態の影響（最近起こった集団のボトルネックや集団構造の存在）を示唆することが報告されている[28,29]が、一方で Tajima's D の値は有意ではなく、そのいずれの可能性も強く示唆していなかった。また、PBS によって局所的な正の自然選択の可能性を検討したところ、JPT における T アレルの自然選択と、CHB における C アレルの自然選択

の、いずれも強く示唆する値とはならなかった（JPT：PBS value = 0.0592（9,051,837 SNP 中 23,975 位）、CHB： PBS value = 0.0685（9,051,837 SNP 中 11,542 位）、図 12）。以上をまとめると、いずれの自然選択の検定からも、rs2294008 における JPT または CHB での明白な正の自然選択のシグナルは検出できなかった。

第 3 節　2D SFS による正の自然選択の検討

上記の中立性検定の結果を踏まえ、本研究では、新しく開発した 2D SFS を利用した検定（F_c、G_{c0}、L_{c0}）によって、rs2294008 を含む領域の selective sweep シグナルの検出を試みた [30,31]。この検定は、ターゲットアレルと連鎖しているハプロタイプに蓄積した変異（IAV: intra-allelic variation）が、ノンターゲットアレルと連鎖しているものと比較して低下していることに着目し、自然選択のシグナルを検出する。この方法は特に selective sweep が完了していないターゲットサイトでの自然選択を検出することに優れている。この特徴を生かし、JPT 及び CHB で、rs2294008 の各アレルに対して自然選択のシグナルの検出を試みた。この方法では、ターゲットサイトと連鎖の強い領域（$r^2 > 0.75$）を用いる。そのため、解析に用いた領域は、LD ブロックの内側で、JPT は 15 kb、CHB は 16 kb となった。

この結果、CHB では、C アレルをターゲットと仮定した検定で全ての要約統計量（F_c、G_{c0}、L_{c0}）が有意な値を示し、C アレルをターゲットとした自然選択の強いシグナルを検出した（表 3）。また、T アレルをターゲットと仮定した要約統計量より小さい値を示し、C アレルと連鎖しているハプロタイプの IAV が T アレルのそれと比較して低いことを示唆した。一方で、JPT では、いずれのアレルでも有意な値を示した要約統計量は存在しなかったが、CHB と同様、C アレルと連鎖しているハプロタイプの IAV が、T アレルのそれよりも低いことが分かった。

次に、CHB に検出された C アレルでの正の自然選択が、hard sweep であるのか、それとも soft sweep であるかを調べた。hard sweep 及び soft sweep の定義は Satta *et al.*[31]に従い、soft sweep の基準を、「C アレルを持つ配列内（D グループ内）に、自然選択のターゲットになるハプロタイプが複数あるかどうか」とした。この基準を満たすかどうか調べるために、G^*_{c0} 及び$\gamma^*(10)$、i_{max} を求めた。この統計量は、D グループに特有の派生型アレルの個数や、派生型アレルを持つサイトの個数を表しており、ターゲットとなるハプロタイプが単独

である場合（hard sweep の場合）は小さく、ハプロタイプが複数ある場合（soft sweep の場合）は値が大きくなる[31]。これらの値は検出したシグナルが hard sweep であることを示唆した[31]が、一方で、D グループでの派生型アレルの個数の最大値を示す、i^*_{max} の値は D グループの本数（155）に対して 75 と、非常に大きな値を示した。この値が、非常に小さい場合は D グループ内にはハプロタイプは一つであるとみなすことができるが、この値が中間的である場合は、複数のハプロタイプ（サブハプロタイプ）が存在する可能性を示唆する。そこで、D グループ内で 75 本の派生型アレルを持つサイトを精査したところ、そのようなサイトは 2 種類（rs2976391（C/A）及び rs2978983（A/G））見つかった（図 13)。このサイトが潜在的な自然選択のターゲットサイトである可能性を考慮し、この 2 つのサイトの連鎖関係を調べた。2 本の組み換え体を除けば、rs2976391 の A アレルと rs2978983 の G アレル、rs2976391 の C アレルと rs2978983 の A アレルは完全に連鎖しており、D グループはこの 2 つのサイトによって、2 サイトとも派生型アレルを持つ 74 本と、2 サイトとも祖先型アレルを持つ 79 本の 2 種類のサブハプロタイプに分類されることが分かった。そこで、前者を A-G サブハプロタイプ、後者を C-A サブハプロタイプと名付け、このサブハプロタイプそれぞれを対象として、自然選択のシグナルの検出を行った。その結果、CHB ではどちらのサブハプロタイプでも、正の自然選択のシグナルが検出された（表 3)。また、この 2 つのサブハプロタイプの IAV は同程度であった。更に、この二つのサブハプロタイプに selective sweep の影響がみられるかどうかを各サブハプロタイプの塩基多様度（π）によって確認したところ、どちらのサブハプロタイプも同様にサブハプロタイプ間の多様度より低く、同程度の塩基多様度を示しており（$\pi_{A\text{-}G} = 0.4 \times 10^{-4}$、$\pi_{C\text{-}A} = 0.4 \times 10^{-4}$、$\pi_{A\text{-}G\ vs\ C\text{-}A} = 2.4 \times 10^{-4}$)、2D SFS の示唆する結果と矛盾しなかった。このことは、CHB の各サブハプロタイプで観察されたシグナルは hard sweep のものであり、C アレルで検出されたシグナルはこの両者を合わせた soft sweep のものであったことを示す。

次に、このサブハプロタイプが JPT にも観察されるかどうかを調べた。その結果、多型的な 2 サイトが確認され、2 サイト間の連鎖関係も JPT でも維持されており、JPT の D グループも CHB と同様の 2 サブハプロタイプで構成されていることが分かった。そこで、このサブハプロタイプに対して 2D SFS による正の自然選択の検討を行ったところ、C-A サブハプロタイプに対してのみ自

然選択のシグナルが検出された。一方、A-G サブハプロタイプには自然選択のシグナルは確認できなかった（表 3）。G^{*}_{c0} 及び$\gamma^{*}(10)$、i_{max} 、i^{*}_{max} の値は、C-A サブハプロタイプに働いた自然選択は hard sweep であることを示しており、A-G サブハプロタイプの塩基多様度と比較すると C-A サブハプロタイプの多様性の方が低かった（$\pi_{A\text{-}G} = 0.5 \times 10^{-4}$、$\pi_{C\text{-}A} = 0.2 \times 10^{-4}$、$\pi_{A\text{-}G\ vs\ C\text{-}A} = 2.7 \times 10^{-4}$）。以上の結果から、C-A サブハプロタイプの hard sweep のシグナルが、中立下にある A-G サブハプロタイプの多様性の高さによって隠されたために、JPT で C アレルには自然選択のシグナルが検出できなかった可能性がある。一方で、CHB では C アレルを持つ配列のうちの 2 サブハプロタイプの両方に自然選択があったため、C アレルをターゲットと仮定した解析で自然選択のシグナルが検出されたと考えられる。

以上をまとめると、JPT と CHB では、C アレルを持つ配列は、C-A と A-G の２つのサブハプロタイプで構成されていた。C-A サブハプロタイプにはどちらの集団でも正の自然選択が働いていた。一方、両集団は遺伝的に近縁であるにも関わらず、CHB では A-G サブハプロタイプは C-A サブハプロタイプと同程度の選択が働いていたが、JPT では A-G サブハプロタイプには自然選択のシグナルは検出されなかった。両集団のサブハプロタイプは異なる自然選択のモード下にあり、かつ、異なる選択圧下にあることが示唆された。

次に、これらのサブハプロタイプに掛かる自然選択の開始した時期について調べた。soft sweep の場合、複数のサブハプロタイプの分岐年代は、自然選択の開始時期の上限と解釈できる[31]。CHB の $\pi_{A\text{-}G\ vs\ C\text{-}A}$ の値より、この２つのサブハプロタイプの分岐年代は、240,000 年前と推定された。このことは、CHB での 2 つのサブハプロタイプに対する自然選択は、最も早い場合は 240,000 年前には始まっていたことを意味する。また、個々のサブハプロタイプに対する自然選択の開始時期の下限（t_D）は、個々のサブハプロタイプにおける TMRCA で算出される[31]。CHB では、A-G サブハプロタイプは遅くとも 27,027±14,333 年前、C-A サブハプロタイプは遅くとも 30,063±12,358 年前には自然選択が働いていたことが分かった。一方、JPT での C-A サブハプロタイプに対する自然選択は、遅くとも 10,811±8,058 年前には始まっていたことが分かった。

以上の結果と、集団の歴史を考慮すると、CHB では、過去のいずれかの時点（240,000 年前～27,027 年前）で二つのサブハプロタイプに自然選択が働きはじめたために、両者と連鎖する C アレルの頻度が上昇しており、現在もこの自

然選択は働き続けている。しかし、JPT では C-A サブハプロタイプと比べて、A-G サブハプロタイプに働く選択圧が十分強くなくなっていた（hardening）ため、C アレルをターゲットと仮定した解析では自然選択のシグナルが検出されなかった。その結果として、CHB では C アレルの頻度が高くなったが、JPT では C アレル頻度は高くならず、これが C アレルの大きな頻度差をもたらした一つの要因と考えられる。

1KGP の subpopulation のうち、最も近縁な集団同士である JPT と CHB であってもサブハプロタイプ間の selection status には違いが生じていた（表 4（a））。更に、EAS の subpopulation での rs2294008 の C アレル頻度は、CHB と同様にいずれも高頻度を示していた（表 1）。そこで、EAS のその他の subpopulation 間で、これらのサブハプロタイプの selection status に違いが見られるかどうか、2DSFS を用いて検討を行った（表 4（b））。JPT と CHB で保存されていた 2 つのサブハプロタイプは EAS の各 subpopulation でも同様に保存されており、C-A サブハプロタイプに対する自然選択のシグナルはどの subpopulation からも共通に検出された。一方、A-G サブハプロタイプに対する自然選択は、CDX 及び KHV ではシグナルが検出されず、CHS では L_{c0} にのみシグナルが検出された（Q-value=0.04）。このシグナルはマージナルなものであるため、統計量間の共分散の影響を除外した combined P で評価を行なったところ、CHS は正の自然選択の弱いシグナルが検出された（P-value=0.04）。更に、本土日本人集団と最も近縁な集団として韓国人集団が知られている[75,76]ため、共同研究者より提供いただいた韓国人集団の配列[52]でもシグナルの有無を調べた。その結果、他のアジア人集団と対照的に、KOR では自然選択のシグナルは A-G 及び C-A サブハプロタイプのいずれでも検出できなかった（表 4（b））。東アジア集団はサブハプロタイプの selection status に多様性が見られることから、この selection status が各サブハプロタイプ内部のハプロタイプの多様性に影響している可能性を考慮して、C-A 及び A-G サブハプロタイプ、T ハプロタイプそれぞれのハプロタイプ数を調べ（表 5（a）~（c））、ハプロタイプネットワークの形を比較した（図 14~図 16）。この結果、サブハプロタイプの種類は、C-A が最も少なく（34 種類）、A-G（50 種類）、T ハプロタイプ（83 種類）の順に多くなった。3 ハプロタイプのネットワークのパターンは全く異なっており、A-G サブハプロタイプは、中央の主要なハプロタイプ（H_1）以外にも数の多いハプロタイプが複数観察され（H_3、H_8 他）、ハプロタイプによっては

subpopulation の割合にも偏りがあった（H_6、H_9）（図 14）。一方、C-A サブハプロタイプの主要なハプロタイプは H_1 のみで、H_1 から派生したシングルトンを持つハプロタイプが多く、これらのハプロタイプが H_1 の周りに放射状に位置していた（図 15）。これは、1 つのハプロタイプが自然選択の対象となり、急激に集団内で頻度が上昇する過程で新規突然変異が生じ、新しいハプロタイプが出現してくる、hard sweep 下にあるハプロタイプが示すネットワークの典型的なパターンであると考えられる。T ハプロタイプは主要なハプロタイプが 3 つに分かれ（H_1、H_2、H_4）、それぞれからハプロタイプが放射状に派生していた（図 16）。ハプロタイプの種類が A-G では C-A ハプロタイプに対して多く、主要なハプロタイプが複数存在していることは、A-G サブハプロタイプの多様性の高さを示しており、ほぼ全ての集団に自然選択が働いている C-A サブハプロタイプとは対照的に、東アジア集団内での A-G サブハプロタイプの selection status に多様性があることを反映していると考えられる。また、A-G サブハプロタイプのネットワークの形状は、T ハプロタイプのような「主要なハプロタイプが複数存在するがお互いの距離が遠く、個々の主要ハプロタイプが派生するハプロタイプを持っている」パターンよりは、「主要なハプロタイプは単一で、そこから全てのハプロタイプが派生する」C-A サブハプロタイプのものと似ている。このことは、現在の C-A サブハプロタイプのように、かつて東アジア集団に共通して自然選択が働いていたが、A-G サブハプロタイプではどこかのタイミングで選択圧が変化したことを示唆していると考えられる。一方で、個々の subpopulation の現在の selection status の違いはネットワークのパターンに現れていない。例えば、自然選択の働いていない subpopulation のみで構成された（シングルトンではない）大きなクラスターが出現する、あるいは、複数の subpopulation に共有されるハプロタイプでも、subpopulation の構成比が特に大きく偏ったクラスターが出現するという例は観察されなかった。以上のことから、ネットワークの形状が主に反映しているのは、東アジア集団の共通祖先時点での selection status であり、現生集団に分岐した後の selection status の違いを区別できるほど解像度は高くない手法である可能性がある。以上をまとめると、東アジアのほとんどの集団では C-A サブハプロタイプに自然選択が働いているが、A-G サブハプロタイプの selection status は、環境が近くとも集団によってかなりばらつき、一部でしか働いていないことがわかった。

東アジア集団の「C-A サブハプロタイプに自然選択が働いている一方、A-G サブハプロタイプはそうではない」傾向は、他の人類集団にも見られるかどうか、1KGP の各 metapopulation を用いて調べた。この結果、SAS では多くの東アジア集団と同様に C-A サブハプロタイプのみに自然選択のシグナルが見られたが、EUR 及び AMR では、C-A 及び A-G サブハプロタイプに対していずれのシグナルも見られなかった（表 4（c））。更に面白いことに、AFR では、アジア集団とは逆に、C-A サブハプロタイプに自然選択が働かない一方、A-G サブハプロタイプが自然選択の対象になっていた。AFR は A-G サブハプロタイプの頻度（0.416）が C-A サブハプロタイプ（0.209）よりも高く、大きな頻度差が生じていた。これは両サブハプロタイプの頻度差がほぼない EAS のパターンとは異なることから、AFR と EAS では、サブハプロタイプ間に異なる選択圧が働いている可能性が示唆された。non JPT/CHB EAS では、A-G 及び C-A サブハプロタイプの両方で自然選択のシグナルが検出されたが、EAS の subpopulation レベルで見ると、KHV や CDX では、A-G サブハプロタイプに自然選択が働いていない（表 4（b））。このように、metapopulation の結果が全ての subpopulation での selection status を反映するとは限らないため、AFR を subpopulation に分けて更に解析を行った。この結果、やはり AFR では全ての subpopulation で A-G サブハプロタイプに自然選択が働いているが、ESN を除く全ての subpopulation で C-A サブハプロタイプで自然選択が働いていないことがわかった（表 4（d））。ESN では唯一 L_{c0} に自然選択のシグナルが出ていることから（Q-value=0.02）、combined P で評価を行なったところ、弱い自然選択のシグナルを示した（P-value=0.04）。一方で、ESN の C-A サブハプロタイプの t_D の値（61,920±30,193 年）は、自然選択下にあるよりは、中立の場合に近い、比較的大きな値を示した。L_{c0} は F_c や G_{c0} よりも長いタイムラインで有意な値を示すことがあり、過去に働いていた自然選択のシグナルに反応する[31]ため、ESN での C-A サブハプロタイプのシグナルは、過去の自然選択のものであった可能性がある。以上をまとめると、rs2294008 を含む領域はアジア・ヨーロッパ・アフリカでそれぞれ別の selection status を持っており、それぞれの集団で異なるハプロタイプが自然選択のターゲットになっていた。また、その selection status も subpopulation によって異なり、東アジアでは A-G サブハプロタイプの selection status に大きなばらつきが見られた。

また、自然選択が働いている集団を対象に、t_D を推定した（図 17）。この結

果、遅くとも、A-G サブハプロタイプは 38,596～18,889 年前、C-A サブハプロタイプは 30,063～10,811 年前に自然選択が働き始めていたと考えられる。両サブハプロタイプの自然選択の開始時期は重複しており、人類集団間でも大きな差は見られなかった。

第 4 節　T アレルの中立下でのアレル頻度の推移

JPT では、C アレルと連鎖した C-A サブハプロタイプに自然選択が働いているにも関わらず、T アレルは依然として頻度が高い。この原因が日本人集団特異的な集団動態によるものである可能性を考慮し、集団動態だけで JPT の T アレルの頻度が現在の高さに至り、頻度差が生じるのかを検討するため、T アレルのアレル頻度シミュレーションを行った。条件は、T アレルと C アレルが JPT・CHB のいずれの集団でも自然選択が働いていない場合と、C アレルが CHB 系統のみで自然選択を受けている場合の二種類について検討した。なお、本研究では、自然選択の開始時期は CHB 系統に入ってからとし、JPT の自然選択の停止時点や、両集団の分岐時点での自然選択の有無については考慮しない保守的な条件でも、rs2294008 で観測した大きな頻度差を持つ高頻度のアレルが生まれるかどうかを検証した。

日本人集団に特有の集団動態として、「二重構造モデル（dual structure model）」が知られている[38]。これを反映したモデルを作成し（図 5）、4 パラメータの組み合わせ（t1、N_{JMN}、N_{A_CNT} 及び r）で合計 570 通りのシミュレーションを行った。シミュレーションに用いたのは、日本人集団（simJPT）と中国人集団（simCHB）で、両集団間の F_{ST} を様々なパラメータの組み合わせで計算し、高 F_{ST} かつ T アレルが JPT で高頻度を示す結果の回数を調べた。その結果、中立下では、高い F_{ST} を持つパラメータの組み合わせは 92 通り出現したが（図 18）、最終的に、rs2294008 のように高い F_{ST} を持ち、simJPT で高い T アレル頻度を持つ組み合わせは 3 通りのうち、3 回出現した（図 19）。この 3 回のアレル頻度の推移を調べたところ、(i) 交雑直前の JMN での T アレルの頻度が非常に高く（0.94~1.0）、(ii) T アレル頻度は A_CNT と交雑した世代で約半分に低下することが分かった。これらの条件が現実の JPT の観察値に近いアレル頻度（0.6）を生み出している。更に中立下にある T アレルの頻度が交雑以降遺伝的浮動の影響で大きく変化するには、t = 2500 年（125 世代）という時間は短すぎることを示唆している。言い換えると、JPT-CHB 間で大きなアレル頻度差が生じ、かつ、JPT で T アレルが高頻度であるためには、縄文系統では T ア

レルが高頻度であることが必要であると考えられる。このモデルでは、交雑前の両祖先集団間で、T アレルの頻度差が大きければ大きいほど F_{ST} が高くなることがわかった（表 6（a））。このことは、JMN 系統で T アレルが高頻度であれば、交雑後の simJPT で T アレルが高頻度になることを裏付けている。

次に、正の自然選択が simCHB の系統だけで C アレルに働いている場合のシミュレーションを行った。完全中立下で KS test を通過した 410 通りのパラメータの組み合わせは、JPT の集団動態を反映していると考えられる。これらに対して選択圧の強さを 4 種類に変えた、合計 1640 通りのパラメータで中立の場合と同じく、高い F_{ST} を持ち、かつ simJPT で高頻度の T アレルが出現するパラメータの組み合わせを調べた。実行している組み合わせが中立条件の 4 倍なので、高い F_{ST} を出す組み合わせや、rs2294008 と似た条件下にある SNP を持つ組み合わせは約 4 倍程度あると予想した。その結果、高 F_{ST} が 1 つでも出る SNP が観察された組み合わせは 363 通り確認されたが、rs2294008 と似た条件を持つ SNP が出てくる組み合わせは 24 通り、35 回と、4 倍より多く観察された。更に、JMN での固定に近い T アレルの高頻度化と、交雑後のアレル頻度の低下は、自然選択を実装したシミュレーションでも観察された（図 20）。また、それぞれの祖先集団間に大きな頻度差があるほど、渡来系集団との交雑後の集団と渡来系集団の子孫の間でも頻度差が大きくなり、F_{ST} が高くなるという特徴は、simCHB に C アレルに対する自然選択が働いている条件下でも確認された（表 6（b-1）～（b-4））。このことから、渡来系弥生集団の祖先集団の自然選択の有無に関わらず、交雑前の縄文系の集団では T アレルの頻度が高かったことが示唆された。

以上の解析では、先行研究で日本人集団と漢族集団の有効集団サイズ等のパラメータが推定されている[43]ため、この数値に基づいて、simCHB を渡来系集団の子孫として扱った。この結果、rs2294008 と似た条件下にある SNP が出現するパラメータの組み合わせは、いずれも r は 0.4 のみであった。この値は、古代ゲノムデータから直接算出された、「多くても 2 割程度」という推定値よりもかなり高い[40,44]。そこで渡来系集団の子孫を、より日本人集団と遺伝的に近い韓国人集団[75,76]とし、中立・自然選択の両方の条件で、rs2294008 と似た条件下にある SNP が出現する回数を調べ、r ごとに比較した（図 21・図 22）。この結果、中立条件下では 29291 回、C アレルに自然選択が働いている状態では 224431 回、rs2294008 に似た条件下にある SNP が出現していた。このこと

は、大陸由来の渡来系集団の子孫を KOR と仮定する場合は、CHB と仮定する場合（中立条件下では 3 回、自然選択条件下では 35 回）よりも rs2294008 と似た条件下にある SNP が非常に出現しやすくなったことを示している。このうち、支持された組み合わせが最も多くなったのは、r=0.2 の場合（中立条件下では 29291 回中 15697 回、自然選択の働く条件下では 224431 回中 116601 回）であった。r の値で rs2294008 と似た条件下にある SNP が出現する回数を比較すると、r=0.4 の場合が最も少なかった（中立下では 29291 回中 10425 回、自然選択の働く条件下では 224431 回中 65599 回）が、この時、JMN での最終世代でのアレル頻度の平均値は、r=0.2 や 0.1 の時と比較して、0.125~0.161 ほど低いことがわかった。つまり、JPT-KOR 間で rs2294008 のような頻度差が生じるためには、r の値が小さいほど JMN の最終世代では T アレルが高頻度である必要がある。また、交雑相手である A_CNT の最終世代では、r の値が小さいほど T アレル頻度の平均値が上昇していた（図 21・図 22）。このことは、渡来系集団の子孫に韓国人集団を想定した場合、rs2294008 と似た条件の SNP が生じるためには、交雑後の集団での縄文系統のゲノムの割合が必ずしも高い必要はなく、交雑直前の縄文系統での T アレル頻度が高いこと及び、渡来系集団での T アレル頻度との関係が重要な要因であることを示している。

以上をまとめると、縄文系人類集団での T アレルの頻度は、大陸由来の渡来系弥生人集団よりも高く、JPT の T アレルは、主に縄文系人類集団で高頻度で維持されてきた T アレルから派生したものであること、JPT 及び縄文の系統では、重篤な疾患リスクと関連したアレルであっても高頻度で維持されてきたことが、F_{ST} で観測される大きな頻度差に繋がった一因と考えられる。

第 5 節　JPT と縄文人のハプロタイプの関係

前節のシミュレーションの結果より、JPT に見られる T アレルは主に縄文系人類集団に由来することが示唆された。rs2294008 は近傍のサイトと連鎖関係にあるため、現代の JPT の T アレルが縄文系統から派生しているならば、1SNP 単独で派生するのではなく、ハプロタイプ（配列）単位で派生していると考えられる。そこで、上記の仮説を確認するため、縄文系人類集団から抽出した古代ゲノムの塩基配列データを利用し、ハプロタイプ間の関係をハプロタイプネットワーク解析で調べた。解析に使用したのは愛知県の伊川津で発掘された 2720~2418 年前の 1 個体のサンプル（IK002）[73]及び、北海道の礼文島で発掘された 3550~3960 年前の 2 個体のサンプル（FUN23 及び FUN5）[40]のゲ

ノム配列である。JPT の LD ブロックに相当する領域（21,959 ベース）に対し、信頼性のあるサイトのみを残すフィルタリングを行い、IK002 は 11,319bp、FUN23 は 21,930bp、FUN5 は 21,795bp を用いて解析を行った。

まず、現生の集団に何種類のハプロタイプがあるのか確認した。JPT と CHB に 21.9 kb の LD ブロックにある領域のハプロタイプは、合計 93 ハプロタイプが観察された（表 7）。このうち、41 ハプロタイプが JPT 特異的なハプロタイプ、37 ハプロタイプが CHB 特異的なハプロタイプで、15 ハプロタイプが両者に共通していた。これらが rs2294008 のどちらのアレルを持つのか調べたところ、JPT 特異的な 41 ハプロタイプのうち、87.8%（36 ハプロタイプ）は T アレルを持っていたが（以下、T ハプロタイプ）、CHB 特異的な 37 ハプロタイプでは T ハプロタイプは 27.0%（10 ハプロタイプ）に留まった。JPT に特異的なハプロタイプのうち、T ハプロタイプの種類が CHB よりも多かったことは、T ハプロタイプが JPT で多様で、集団内で長く維持されてきたことを示唆している。一方、C ハプロタイプの種類は、JPT 特異的なハプロタイプではかなり少ない（5 ハプロタイプ、12.2%）のに対し、CHB 特異的なハプロタイプは全体の半数を超え（27 ハプロタイプ、73.3%）、多様化している。この結果は、CHB では C ハプロタイプは JPT と比べて高頻度であるため、新規突然変異が生じるチャンスが多く、JPT と比べて新しいハプロタイプが出現しやすかったことを示唆している。JPT-CHB 間で、C アレルに対する自然選択のモードや選択圧に違いが見られたが、ハプロタイプの数の違いは、このことを反映している可能性がある。

上記のハプロタイプの偏りを踏まえて、現代のサンプルにチンパンジーの相同配列と、船泊サンプル（FUN23・FUN5）[40]、伊川津サンプル（IK002）[73]を加えて、ネットワーク解析を行いハプロタイプ間の関係を調べた（図 23）。この結果、ネットワークは T ハプロタイプと C ハプロタイプで大きく二つに分かれた。また、全ての縄文系統のサンプルは、T アレルを持っていることが確認でき、全てが T ハプロタイプに属していた。FUN23 及び IK002 は JPT 特異的なハプロタイプと最も近縁な関係にあり、FUN23 は JPT 特異的なハプロタイプで 61 番、IK002 は現代のハプロタイプでやはり JPT 特異的なハプロタイプグループ 30 番、31 番、87 番と最も近いことが分かった。一方、FUN5 は FUN23 に最も近縁であったが、同時に JPT・CHB のどちらにも存在する主要なハプロタイプグループである 19 番、42 番、43 番、84 番のどれとも近くに位置した。

FUN23 及び FUN5 番はミトコンドリア DNA の解析から、同一の母系統に属していないことが報告されている[40]。このため、船泊の 2 個体は血縁関係が遠いことが示唆されるが、近縁な個体同士であったとしても、船泊集団の中には T ハプロタイプの多様性があったことが示された。FUN5 および IK002 はサンプルの coverage が低いため、ハプロタイプの系統関係に対する信頼性は FUN23 よりも下がる。しかしながら、縄文系人類集団のサンプルは JPT の T ハプロタイプにクラスターしていた。以上のことから、縄文系統人類集団の持っていた T ハプロタイプの多様性は高く、また、JPT の T ハプロタイプと近縁な関係にあることがわかった。しかし、遺伝的な近縁性のみでは、JPT の持つ T ハプロタイプの由来が縄文系人類集団由来であるとは言えない。縄文系統人類集団は東アジアの集団の中でも、分岐が早い集団であることが示唆されている[39,40,44]が、もし JPT の T ハプロタイプの一部が縄文系統人類集団から派生しているなら、縄文系人類集団のハプロタイプの多様性は、JPT のハプロタイプの多様性を包含する関係になるはずである。そこで、T ハプロタイプの由来についてさらに詳しく調べるため、同じサンプルを用いてハプロタイプ同士の系統関係を調べた（図 24)。その結果、FUN23 の二本の染色体は T ハプロタイプを持つ JPT/CHB の最も外群にクラスターしたが、IK002 はその内側の JPT と共にクラスターし、FUN5 は JPT/CHB と共にクラスターした。このことは、縄文系人類集団のハプロタイプの多様性が JPT のハプロタイプの多様性を内包することを示しており、アレル頻度シミュレーションから得られた、「JPT の T アレル（及びその近傍の配列）は縄文系人類集団由来である」という仮説を支持した。以上をまとめると、JPT の祖先集団である縄文系人類集団が高頻度でもっていた T ハプロタイプを受け継いだ結果、JPT では T ハプロタイプが高頻度で観察されることが示唆された。

第 4 章 総合討論

第 1 節　研究のまとめと、各集団の正の自然選択の性質の評価

本研究では、diffuse type の胃がん（DGC）のリスクアレルである rs2294008 の T アレル[9]が、なぜ JPT で高頻度に至ったのか、JPT と CHB を中心として、集団遺伝学の観点から解釈を試みた。T アレルの頻度は、遺伝的に近縁な東アジア集団の中で、JPT で最も高い（表 1）。F_{ST} によって、遺伝的に近縁な CHB 集団との間でも非常に大きな頻度差があることがわかった。しかし、F_{ST} が高い値を示していても、2D SFS を除く多くの中立性検定では、T アレルまたは C アレルを対象とした正の自然選択のシグナルを JPT と CHB のいずれの集団でも検出することが出来なかった（図 10・図 11）。一方、2D SFS による解析（F_c、G_{c0}、L_{c0}）では、CHB で C アレルに働いている正の自然選択のシグナルが検出された（表 3）。2D SFS は、ターゲットサイトを含むハプロタイプの IAV をノンターゲットサイトのそれと比較することで、selective sweep が完了していない（現在も自然選択が働いている）場合の、正の自然選択のシグナルを検出する[31]。この特徴が本研究では効果的だったと考えられる。さらに、2D SFS 解析を行った 15 kb の外側には連鎖関係の弱い領域が確認でき（図 25）、また、F_{ST} の順位が第 2 位である rs2717562 は rs2294008 の属する LD ブロック（21.9kb）の範囲外に位置している。このことは、染色体組み換えのホットスポットがターゲットサイトの近傍にあることを示している。このように組み換えが起こって配列の連鎖関係が崩れた場合には、ハプロタイプの LD の長さを用いた（最近の自然選択の検出に優れた）中立性検定では、検出が困難である可能性が指摘される。例えば、nS_L による検定では、F_{ST} の高い SNP10 個に対し、6 個のみがシグナルを示しており、そのシグナルの強さもマージナルなものであった（図 10（b））。実際に、nS_L を用いた先行研究では、比較的最近（1 万年未満）の自然選択の検討に使われている例が多かった[19,30,33]。一方で、A-G 及び C-A サブハプロタイプの CHB での自然選択の開始時期の下限を考慮すると、本研究で検出した自然選択の開始時期は古く、いずれも 3 万年より古いと推定される（表 4（a））。これに加えて、近傍に組み換えのホットスポットサイトが存在していた（図 8）ことが、長い LD を持ったハプロタイプの維持を難しくし、ハプロタイプベースの検定によるシグナル検出を困難にしていた[18]と考えられる。

CHB には C アレルに自然選択が検出された一方、JPT では C と T のどちら

のアレルにも自然選択のシグナルは検出されなかった。このことを踏まえ、πによって 21.9 kb の LD ブロック（chr8: 143752235-143774193）と上流・下流 100 kb の領域で、両集団の配列の多様性を評価した（表 8）。この結果、JPT-CHB 間で C ハプロタイプの多様性は大きく変わらなかった（JPT: $\pi_C = 1.5 \times 10^{-4}$、CHB：$\pi_C = 1.6 \times 10^{-4}$、$P = 0.398$）。また、$\pi_C$の値は、どちらの集団でも LD ブロックの多様性が、近傍の多様性よりも有意に低下していた（全て $P < 0.01$）。一方で、JPT では LD ブロックのπ_Tは上流 100 kb と比較して多様性が有意に低下していた（$P < 0.01$）が、下流とは差がなく（$P > 0.025$）、CHB では LD ブロックのπ_Tは上流 100 kb と比較して僅かに低下し（$0.01 < P < 0.025$）、下流 100 kb で有意な差は見られなかった（$P > 0.025$）（JPT：$\pi_T = 6.4 \times 10^{-4}$、CHB：$\pi_T = 8.4 \times 10^{-4}$）。以上の結果は、$\pi_C$が示した多様性の低下が領域特異的な突然変異率の変化によるものではなく、両集団に起こった C ハプロタイプに対する共通の自然選択によるものであることを示すと考えられる。C アレルを 2 サブハプロタイプに分類した結果、JPT と CHB では選択圧や自然選択のモードが異なることが分かった（表 4（a））。

C アレルをターゲットサイトした場合、CHB には soft sweep の痕跡が見られた一方で、JPT では自然選択のシグナルが検出されなかった。この両集団の解析結果の違いは、正の自然選択のモードの違い、あるいは、自然選択の要因そのものの違いを示唆している。CHB では、2 つのサブハプロタイプに対する正の自然選択は、遅くとも 27,000 年前（A-G サブハプロタイプ）または 30,000 年前（C-A サブハプロタイプ）には開始していたことが分かった。この時期は、縄文系人類集団と、東アジアの人類集団の分岐年代（15,000 年前~38,000 年前）[40,43,44]とほぼ一致する。対照的に、JPT では、正の自然選択のシグナルは C-A サブハプロタイプにしか見られず（表 4（a））、C-A サブハプロタイプに対する JPT での自然選択の開始時期は CHB での開始時期とも重なっている。以上の事は、JPT（または縄文系人類集団）と CHB の共通祖先では、両方のサブハプロタイプに自然選択が働いていたことを示しており、JPT の祖先集団である縄文系人類集団が分岐した時点でも、CHB 系統では、両方のサブハプロタイプに自然選択が働いていた可能性がある。しかしながら、JPT の系統では、A-G サブハプロタイプに対する自然選択はいずれかの時点でリラックスまたは完全に停止しており、自然選択のターゲットが 2 サブハプロタイプから 1 サブハプロタイプへと変化することによって、soft sweep から hard sweep へのモードの

変化（hardening）が起こった。以上の事から、JPT と CHB に見られる遺伝的適応は、一部に同じサブハプロタイプを自然選択のターゲットにしていながら、選択圧が異なる別の遺伝的適応である可能性がある。PBS はサイト毎のアレル頻度に基づいて、自然選択が働いていると仮定した集団の、集団特異的な遺伝的適応を検出するために用いられる[26]が、本研究の場合、JPT と CHB では自然選択のターゲットの一部が共通していたため、特定の集団だけで系統樹の枝の長さが歪むような頻度差は生じず、自然選択のシグナルとして検出されなかったと考えられる。サブハプロタイプを定義する 2 つの SNP のうち、rs2976391（C/A）は、*PSCA* 遺伝子のイントロン部分に位置しているほか、遺伝子領域が重なっている *JRK* 遺伝子（Jrk herlix-turn-helix protein）の上にも位置している（図 26、[10]）。この派生型の A アレルはプロモーターの活性や転写活性の変化に関連することが報告されている[10]。一方、rs2978983（A/G）は現時点で、Ensemble でも明らかな機能の報告はなされていなかった。C ハプロタイプは、一様に rs2294008 の C アレルを持った配列であっても、おそらく後述するように他の SNP（rs2976391 及び rs2978983）との組み合わせによって 機能的にはお互いに異なっている可能性がある。ただし、*PSCA* 遺伝子は、ヒトでは胃がんだけではなく、様々ながんの発症に関わっていることが報告されている[12-14]が、この（がんの発症を抑える）機能が選択圧として働いているかはわからない。なぜなら多くのがんの発症年齢は生殖可能な年齢よりも高いため、このがん発症の抑制機能が直接的に適応度に影響しない可能性がある。また、がんが人類の主要な死因となったのは最近のことであり、この遺伝子の長い時間の進化の選択圧として考えられるかどうかは疑問である。そのため、*PSCA* 遺伝子の他の機能に対する選択圧についても検討する必要がある。ただし、*PSCA* 遺伝子は、現在がん関連以外の機能についてはヒト以外でも解析が進んでいない。一方で、*JRK* 遺伝子は、その配列から DNA 結合タンパク質としての機能があると予測されており[77]、この機能は、アレルごとの遺伝子発現の効率の違いに関連している可能性がある。また、この遺伝子は一部のてんかん症状や、精神疾患との関連も指摘されている[78-80]。このため、これらの機能に関係した選択圧である可能性も考えられる。また、近傍の別の遺伝子に自然選択が働き、ヒッチハイキング効果によってこの領域のアレル頻度が上昇した可能性についても検討する必要がある。

いずれにしても、JPT と CHB のように、遺伝的に近縁な集団同士であって

も集団毎に同じハプロタイプ（例えば A-G サブハプロタイプ）に異なる選択圧が生じていることは、興味深い結果だと考えられる。JPT では C-A サブハプロタイプのみ、CHB では A-G 及び C-A サブハプロタイプの両方に自然選択が働き、selection status に違いが生じていたことは、たとえ正の自然選択のターゲットが一部共通していても、両集団の違いが単一の選択圧では説明できないことを示唆している。

第 2 節　rs2294008 に対する正の自然選択を通して見る、人類集団の遺伝的な多様性

C-A サブハプロタイプに自然選択が働いているにも関わらず、JPT において C アレル単位での自然選択が検出できなかった原因には、以下の 2 つの可能性が考えられる。まず、C-A サブハプロタイプは T ハプロタイプと多くの祖先型アレルを共有しているため、C-A サブハプロタイプにユニークな派生型アレルの蓄積が見えにくかったことが挙げられる。C アレル単位での自然選択のシグナルが検出できないにも関わらず、C-A サブハプロタイプ単位であればシグナルが検出されるという点は、一見矛盾して見える。しかし、C-A サブハプロタイプを特徴付ける、rs2976391（C/A）及び rs2978983（A/G）の 2 つの祖先型のアレルは、T ハプロタイプにも共有されている（図 13）。これらのサイトの祖先型のアレルを個別に自然選択のターゲットと仮定して検定を行なった場合には、特にシグナルは検出されなかった。rs2294008、rs2976391、rs2978983 の各アレルは、アレル頻度や配列の関係から、rs2294008 で新規突然変異として C アレルが出現した後、C アレルを持つ配列だけに、それぞれ rs2976391 と rs2978983 で新規突然変異が起こり、C-A サブハプロタイプと A-G サブハプロタイプが分岐したと考えられる。2D SFS はターゲットアレルと連鎖している新規突然変異の多様性に着目して自然選択を検出する手法であるが、この方法では、C-A サブハプロタイプと A-G サブハプロタイプが分岐した後に生じた突然変異のみを T ハプロタイプのみに生じた突然変異と区別して IAV を評価することが困難である。このため、rs2976391 の C アレルや、rs2978983 の A アレルをターゲットとした 2D SFS では、正の自然選択のシグナルが検出されなかった。以上のことから、rs2294008 の C アレル単独ではなく、rs2294008 の C アレルに加えて C-A サブハプロタイプに連鎖する派生型のアレルの組み合わせ、つまり、C-C-A の配列が生物学的に重要な機能と関連していると推測される。また、2 つ目の可能性として、JPT では、C-A サブハプロタイプの頻度は A-G

サブハプロタイプと比較してほとんど差がなかったことが挙げられる。このことが、C アレル単位の検定では C-A サブハプロタイプでのユニークな派生型アレルのシグナルを覆い隠しており、その結果、JPT では C アレル単位での自然選択のシグナルを検出できなかったと考えられる。

A-G 及び C-A サブハプロタイプは AFR を含む他の人類集団でも保存されていた。このことは、2 つのサブハプロタイプの分岐年代が、人類の出アフリカよりはるか以前である、240,000 年前であることと矛盾しない結果である。2 つのサブハプロタイプに対して、2D SFS を用いて、JPT と CHB を除いた他の人類集団での正の自然選択の検討を行った（表 4（c））。この結果、A-G サブハプロタイプは EAS の一部及び AFR の全 subpopulation で自然選択のターゲットとなっていた。また、C-A サブハプロタイプに対する自然選択は、韓国集団を除く東アジア集団（EAS)、及び SAS で観察された。また、KOR と EUR は、どちらのサブハプロタイプに対する自然選択のシグナルも検出されなかった。C-A サブハプロタイプに属する配列は全て、rs2976391 の C 及び、rs2978983 の A アレルを持っているが、この 2 アレルは T ハプロタイプに属する配列とも共有されており、類人猿を外群に用いた解析からも、この 2 アレルは双方とも祖先型のアレルであることがわかっている。従って、C-A サブハプロタイプは A-G サブハプロタイプよりも古くから存在することが期待される。通常、集団内に長く存在している（より古い）サブハプロタイプの方が、配列に突然変異が生じるチャンスが多いため、IAV は大きくなることが予想される。しかし、C-A サブハプロタイプは、非 AFR の集団（EAS の subpopulation 及び SAS）では A-G サブハプロタイプよりも IAV が小さかった（表 4（c））。このことは、C-A サブハプロタイプの選択圧が、非 AFR の集団では A-G サブハプロタイプよりも強い可能性を示唆している。もし互いにサブハプロタイプが中立ならば、祖先型のアレルを持ち、より集団の中で長く維持されてきた C-A サブハプロタイプの方が、突然変異が生じる機会が多かったため、A-G サブハプロタイプより多様性が高く、従ってハプロタイプの種類は多くなるはずである。しかし、実際には C-A サブハプロタイプの種数は A-G サブハプロタイプより少なく（表 5・6)、そのネットワークの形状も、C-A サブハプロタイプの方が hard sweep 下にあるハプロタイプの典型的なパターンに近かった（図 14・図 15)。そのため、2D SFS の示す結果は、東アジア集団の両ハプロタイプでのネットワーク解析の結果と矛盾しない。以上のことから、現生の多くの東アジア集団では、C-A サ

ブハプロタイプは共通の自然選択のターゲットであり、A-G サブハプロタイプよりも選択係数が大きい可能性が高い。一方、A-G サブハプロタイプは AFR でのみ、C-A サブハプロタイプよりも強いシグナルを示し、A-G サブハプロタイプのアレル頻度（0.32~0.54）も C-A サブハプロタイプ（0.15~0.29）よりは高いが、IAV が示すハプロタイプの多様性は、A-G サブハプロタイプの方が小さい。このことから、アフリカ集団は、アジア集団とは逆に、A-G サブハプロタイプに対する選択圧の方が強いと考えられる。このため、アジア集団・アフリカ集団・ヨーロッパ集団の間では、2 サブハプロタイプに対してそれぞれ進化的に異なるメカニズムが働き、現在の頻度に至った可能性がある。

以上のことと、metapopulation 間の系統関係を考慮すると、ヒト集団の歴史においては、同じ遺伝子座で正の自然選択のターゲットが変化したり、あるいは歴史が異なる集団ごとに独立にターゲット化してきたことがわかる（図 27）。例えばアジア集団の共通祖先では両サブハプロタイプに対する自然選択が働いていたのに対し、CDX/KHV/JPT では A-G サブハプロタイプに対する自然選択だけが検出されなくなり、ターゲットの変化が観察される。同様に、KOR ではいずれのハプロタイプに対しても自然選択のシグナルが検出されなくなり、ターゲットが消失している。また、アフリカ集団と東アジアの一部の集団（CHB/CHS）で、A-G サブハプロタイプに対して独立に自然選択が働くようになった。しかしアジア集団では A-G 及び C-A サブハプロタイプに対する soft sweep であるのに対して、アフリカ集団では A-G のみの hard sweep である。歴史の異なる複数集団で、同じ遺伝子座に自然選択が働いている報告は多くある[81]。例えば、乳糖耐性のように、同じ遺伝子座にそれぞれの集団で独立に出現した新規突然変異に自然選択が働くようになった場合[82]が挙げられる。しかし、本研究で検出した自然選択の働き方はこれとは異なり、集団内にもともとあったハプロタイプに対して働き、かつ、各集団で sweep のパターンもターゲットも異なる点は興味深い。また、アジア・ヨーロッパ・アフリカのそれぞれの集団で、各ハプロタイプの比が異なることがわかる（図 27）。他の metapopulation と比較して、特に東アジア集団では、自然選択の開始時期の下限と、各 subpopulation の分岐年代を考慮すると、短期間でかなり頻繁に自然選択のターゲットが切り替わっていることも興味深い点である。このことは、現在正の自然選択が働いている集団であっても、必ずしも祖先集団から自然選択の状態をそのまま引き継いだとは限らないことも示唆している。各人類集団

のサブハプロタイプに対する自然選択の開始時期を調べたところ、C-A サブハプロタイプは遅くとも約 30,000 年～11,000 年前、A-G サブハプロタイプは遅くとも約 39,000 年～19,000 年前に自然選択が始まっていたことが分かった（図 17）。遺伝的に最も多様なアフリカ集団を含む世界中の人類集団で、自然選択の下限値が狭い範囲に集中している点は特記すべきである。現生人類は、出アフリカ以降、遅くとも 47,000 年前[70]には人類集団が分集団化しており、この時期には各集団内で既に環境適応や遺伝的分化が始まっていた。人類の拡散に伴う生息環境の変化に対する遺伝的適応の一種として、同じ遺伝子座で、集団間で正の自然選択のターゲットが変化し、また、異なる集団で独立にターゲットとなっていたことは関連がある可能性がある。

第 3 節　東アジアの T アレルの歴史

rs2294008 の T アレルは JPT-CHB 間で非常に高い分化を示し（$F_{ST} = 0.2547$）、JPT でも C アレルを含むハプロタイプの一部に自然選択があることが分かったが、T アレルは高頻度を示した。このようなサイトが出現する条件をアレル頻度シミュレーションによって調べたところ、JPT の祖先集団の 1 つである、縄文系人類集団では大陸由来の渡来系弥生人集団と交雑する直前、T アレルの頻度が非常に高かったことが示唆された。この縄文系人類集団では非常に T アレルの頻度が高かったという仮説は、現在の琉球の人類集団およびアイヌ集団での T アレル頻度からも支持された。琉球の人類集団及びアイヌの人類集団は、本土日本人よりも縄文系統の遺伝的割合が高いことが知られているが[44]、彼らの T アレル頻度または T アレルと非常に強く連鎖しているアレルの頻度は本土日本人である JPT のものよりも高い（琉球 : 0.701（rs2294008）[76]、アイヌ : 0.975（rs2976396）[83]）。JPT での T アレル及び、その近傍の配列が縄文系人類集団から派生したことを確認するため、伊川津の縄文人（IK002）及び、船泊の縄文人（FUN5 及び FUN23）のゲノムデータを利用し、ハプロタイプネットワーク解析を行った。この結果、2 個体の縄文系統のサンプル（IK002 及び FUN23）は JPT に特異的な T ハプロタイプと最も近縁であることがわかり、FUN5 は FUN23 のハプロタイプ及び、JPT・CHB に共にみられる高頻度のハプロタイプの双方と近縁であることが分かった。以上の解析は、縄文系人類集団は高頻度の T ハプロタイプを集団内に維持しており、これが現在の JPT に受け継がれた可能性を支持した。

以上の解析結果を踏まえて、JPT での C ハプロタイプ、T ハプロタイプの辿

った歴史を以下のように再構築した。（縄文系統の祖先集団を含む）東アジアの祖先集団には、C-A サブハプロタイプ及び A-G サブハプロタイプが存在し、これらに対する正の自然選択が働いていた。15000 年～38000 年前[40,43,44]の時期に、縄文系人類集団が他の東アジアの祖先集団から分岐後、日本列島に移住する。縄文系統では、この分岐の後のいずれかの時期に A-G サブハプロタイプに対する自然選択がリラックスまたは完全に停止し、T ハプロタイプの頻度（rs2294008 の T アレルの頻度）が縄文系人類集団内で上昇した。韓国人集団の祖先集団ではいずれかのタイミングで両サブハプロタイプに対する自然選択が停止したが、ほとんどの東アジア集団では C ハプロタイプへの自然選択が働き続け、C アレルの頻度が上昇を続けた（図 27）。また、F_{ST} に基づいた推定値によると、3000 年～3600 年前[75]、渡来系弥生人の祖先集団が東アジアの祖先集団より分岐し、渡来民として 2500～3000 年前[68,84]に日本列島に移住し、縄文系人類集団と交雑した。縄文系人類集団では T ハプロタイプの頻度が非常に高かったが、交雑によって T ハプロタイプの頻度が交雑直前の頻度の 6 割程度に低下した。現生の多くの東アジアの系統では C-A あるいは両方（C-A 及び A-G）のサブハプロタイプへの自然選択が働き、T アレルの頻度が低下したが、JPT には C-A サブハプロタイプのみに対する自然選択が働き、また、縄文系統から派生した高頻度の T アレルによって、他の東アジア集団と比べて T アレルが高頻度を示したと考えられる。

以上をまとめ、JPT で近縁集団と比較して胃がんのリスクアレルに大きな頻度差が認められる原因は、以下の二点にあると考える。

（I）JPT の祖先集団の一つである、縄文系人類集団と、CHB の共通祖先では、胃がんのノンリスクアレルを含む C-A 及び A-G サブハプロタイプに対する自然選択が働いていた。しかし、JPT の系統では、C-A サブハプロタイプに対する自然選択が働く一方、選択圧の変化によって、いずれかの時点で A-G サブハプロタイプに対する自然選択がリラックスまたは完全に停止した。これらの複合的な要因によって、JPT では C アレルが低頻度である。

（II）祖先集団である縄文系人類集団では T アレルの頻度が非常に高かった。このため、T アレルの頻度が低い渡来系弥生人集団と交雑を経ても、現在 JPT では T アレルの頻度が他の集団と比較して高頻度に保たれている。

第 5 章 本研究の意義

本研究の新規性及び意義は、以下の 4 点であると考える。

1）私が開発を手伝った 2D SFS[31]を用いて、既存の中立性検定でははっきり自然選択のシグナルが出なかった領域に対して、自然選択のシグナルの検出例を示せた点である。本手法は、より古い自然選択のシグナルを評価し、あるいはサブハプロタイプの検出で行ったような自然選択のターゲットサイトの絞り込みが出来ると示せたことは集団遺伝学の分野において有意義であると考える。

2）日本人で罹患率が高い胃がんと強く関連するリスクアレル（rs2294008 の T アレル）頻度の高さに対して、集団遺伝学的な方向から説明が出来た点である。これまで、rs2294008 については、diffuse type の胃がんの発症との関連性や、生物学的な機能の評価が報告されてきたが、集団の歴史および正の自然選択の存在という観点から、このリスクアレルの頻度差について説明を試みた研究は、本研究が初めてである。特に、研究対象である東アジアの人類集団の中で、JPT で A-G サブハプロタイプへのシグナルが KOR と同様に失われたことを示せた点、遺伝的に近縁な集団間での動的な選択圧の変化を示したという点で、新規性のある知見を提供したと考えられる。

3)「胃がんのリスクアレルを高頻度で持つ」という、縄文系人類集団の遺伝的・生理的な特徴を明らかにできた点である。目の色や髪の毛の性質[40]などを除いて、縄文系人類集団の生理的な特徴と関わる遺伝因子については、高品質な古代ゲノムのサンプルが限られることもあって、いまだ分かっていないことが多い。また、考古学的なアプローチは、発掘されたサンプルに対する解像度の高い情報が手に入る一方で、現存せず、もう直接データが取れない集団に対しては、生理的な特徴に知見を拡張することは困難である。特に、胃がんは骨ではなく組織に起こる病変であるため、このような生理的特徴は評価可能な考古遺物として残る確率は低いと考えられる。本研究では、1KGP の JPT を用いて、現代の日本人集団で知られている胃がんに関連する遺伝因子が縄文系人類集団からどのように受け継がれてきたかを明らかにすることができた。縄文系統での胃がんの発症率や死亡率について直接的に評価することは困難であるが、縄文系統の集団が、その遺伝子プールに、現生人類にとっての胃がんのリスクアレルを高頻度で持っていたと示せたことは、日本人集団の自然人類学的知見の発展に貢献する成果の一つである。

4）rs2294008 を一例として、一遺伝子座において自然選択のターゲットが変化

し、あるいは選択圧の変化によって再びターゲットになり得る、動的で複雑なプロセスを示せた点である。これは、本研究の新規性の一つであると考える。正の自然選択の検出や、ターゲットアレルが現生集団にどのように受け継がれたのかを調べた研究はこれまでにも多く存在するが、異なる集団間で、同じターゲットアレルに自然選択が働いている場合は、最節約的に考えて共通の要因が driving force であったと考える。また、ある一種の特定の集団に働いている自然選択に焦点を当てる場合、現生個体に現在働いている（または最後に働いていた）自然選択が研究のメインターゲットとなる。しかし、本研究の成果は、機能的に重要な遺伝子座で、同じアレルが同一系統内で短期間に、複数回自然選択のターゲットが変化し、selection status が変わった例（東アジア集団の共通祖先で働いていた A-G サブハプロタイプに対する自然選択が、独立に JPT・CDX/KHV の系統で失われ、C-A サブハプロタイプのみにターゲットが変化）や、集団分岐後に独立に同じアレルに自然選択が働く例（アフリカ集団と CHB/CHS で A-G サブハプロタイプに自然選択が働いている）を示した。この点で、ヒトが進化の過程で環境の変化に有機的に応答しており、ヒトの“適応的”な在り方や、あるいは選択圧が単一であるとする前提について、疑問を投げかけるものだと考える。また、現生人類が地球上に拡散した短期間に、生息環境に遺伝的に適応する過程で、選択圧の変化や遺伝的に遠い集団間での独立な自然選択のターゲット化を通して獲得した機能的な多様性創生機構の発見は、人類進化を理解する上で非常に面白い知見を提供することができたと考える。

第 6 章　更なる研究の発展の可能性

第 1 節　本研究の未解決事項

rs2294008 及び周辺の生物学的な機能は複雑であり、T アレルと C アレルの適応度を正確に評価することは難しい。本研究では T アレルを DGC のリスクアレル、C アレルを DGC のノンリスクアレルとして紹介したが、一方で、C アレルは十二指腸潰瘍のリスクアレルとの関連性も指摘されており、GWAS からは、日本人を対象にした場合[11]と、コーカソイドの集団を対象とした場合[17]の報告がある。T アレル・C アレル間にもそれぞれのリスクと関連した生物学的な機能の違いがあり、更に、C アレルを持つサブハプロタイプ間にも機能の違いがある。これらに対する選択係数の大きさを正確に評価することは今後の課題である。

また、本研究では、1KGP に登録されている本土日本人に働いている自然選択に話を限定したが、本土日本人よりも縄文系統のゲノムの割合が高いとされる、アイヌや琉球の集団[44]での、C ハプロタイプに対する自然選択の有無や、配列情報については調べることが出来ていない。これらの集団については、JPT と同様の環境適応によって C-A サブハプロタイプに自然選択が働いているのかどうかを調べることは、この遺伝子座の選択圧について調べる上で有意義であると考える。また、現在 JPT に働いている C-A サブハプロタイプに対する自然選択は、縄文系統が東アジア集団と分岐した際に働いていた selection status をそのまま受け継いだのか、それとも遺伝的に最も近い集団である KOR と同様に一度全ての自然選択がリラックス/停止した後、JPT の系統では再度独立に働き始めたのかということや、A-G サブハプロタイプの自然選択がリラックスしたタイミングについて、近縁集団を調べることでより解像度の高い情報を提供することができると期待する。東アジア集団での T ハプロタイプ、C ハプロタイプの経験した歴史を更に詳しく調べるためにも、これらの集団や、各サブハプロタイプが selection status を変化させたと考えられる時期の、アジア集団での古代ゲノムの配列情報は、将来的にも調べる価値があると考える。また、東アジア以外の集団についても、アフリカの人類集団では非アフリカ集団と比較し、A-G サブハプロタイプの自然選択のシグナルが強いという独特の特徴を持っていることが分かったが、1KGP の AFR のデータは西アフリカの集団に限られているため、南アフリカや、本研究では焦点を当てられなかったオセアニアのデータ[85]なども含め、将来的にサブハプロタイプに対する自然選択の有無の検討

を行いたい。

また、各集団の A-G 及び C-A サブハプロタイプに対する選択圧についてもさらに検討が必要である。rs2294008 以外の胃がんのリスクアレルの有無や、その機能についてはまだ議論があるが、本文中でサブハプロタイプを定義するために用いたサイトの一つである、rs2976391（C/A）の A アレルは、A-G サブハプロタイプに含まれていながら、このアレルも胃がんのリスクアレルである可能性が指摘されている[86]ことが分かった。JPT をはじめとする東アジアのいくつかの集団では、rs2294008 の C アレルを持つ配列全てに正の自然選択が働くのではなく、「C アレルを持った上で、rs2976391 の A アレルは持っていない」C-A サブハプロタイプのみが自然選択の対象である。以上の事を考慮すると、JPT では、rs2294008 の T アレルや、rs2976391 の A アレルのような胃がんのリスクアレルを出来るだけ持たないサブハプロタイプに対する自然選択、または、C-A ハプロタイプだけが持っている機能が適応度を上げるために有利に働いている可能性がある（図 13）。一方、CHB や CHS では、C-A 及び A-G サブハプロタイプの両方に自然選択が働いている。これらの集団では、rs2976391 の A アレルによって上昇する胃がんのリスクは、A-G サブハプロタイプが持つ何らかの生理的な特徴とのトレードオフとして自然選択を受けた可能性があり、たとえ rs2294008 の C アレルを含むサブハプロタイプが自然選択下にあったとしても、JPT と同じ選択圧で説明することは困難である。本研究では、近縁集団間でも選択圧が異なることの指摘に留まり、直接的な選択圧の特定は出来なかったが、将来的には、近縁集団間の選択圧について調べてみたい。

第 2 節　本研究で得た知見の発展

本研究では、rs2294008 の頻度差を評価するため、JPT-CHB 間の F_{ST} によってゲノムワイドな SNP サイトの頻度差の比較を行った。この際、本研究では 50 位までを記載した（表 2）が、これより下位のサイトにも JPT で自然選択が働いている可能性が十分考えられる（図 6）。rs2294008 と連鎖関係にない、F_{ST} の高い SNP に対して EHH や PBS、2D SFS などで自然選択の検討を始めており、今後も JPT で自然選択が働いている領域が見つかる可能性がある（図 12・28）。

参考文献

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *C. A. Cancer. J. Clin.* **2018**, *68*, 394-424. doi:10.3322/caac.21492.

2. Crew, K. D., Neugut, A. I. Epidemiology of gastric cancer. *World J. Gastroenterol.* **2006**, *12*, 354–362, doi:10.3748/wjg.v12.i3.354.

3. Forman, D., Bray, F., Brewster, D. H., Gombe Mbalawa, C., Kohler, B., Piñeros, M., Steliarova-Foucher, E., Swaminathan, R., Ferlay, J. Cancer Incidence in Five Continents, Vol. X. *IARC Scientific Publication.* **2014**, *164*. Lyon: International Agency for Research on Cancer.

4. Lauren, P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathol. Microbiol. Scand.* **1965**, *64*, 31–49, doi:10.1111/apm.1965.64.1.31.

5. Ansari, S., Gantuya, B., Tuan, V. P., Yamaoka, Y. Diffuse Gastric Cancer: A Summary of Analogous Contributing Factors for Its Molecular Pathogenicity. *Int. J. Mol. Sci.* **2018**, *19*, 2424. doi:10.3390/ijms19082424

6. Yamashita, K., Sakuramoto, S., Nemoto, M., Shibata, T., Mieno, H., Katada, N., Kikuchi, S., & Watanabe, M. Trend in gastric cancer: 35 years of surgical experience in Japan. *World J. Gastroenterol.* **2011**, *17*, 3390–3397, https://doi.org/10.3748/wjg.v17.i29.3390

7. Henson, D. E., Dittus, C., Younes, M., Nugyen, H., Albores-Saavedra, J. Differential Trends in the Intestinal and Diffuse Types of Gastric Carcinoma in the United States, 1973-2000: Increase in the Signet Ring Cell Type. *Arch. Pathol. Lab. Med.* **2004**, *128*, 765–70, doi:10.1043/1543-2165(2004)128<765:DTITIA>2.0.CO;2

8. Miyahara, R., Niwa, Y., Matsuura, T., Maeda, O., Ando, T., Ohmiya, N., Itoh, A., Hirooka, Y., Goto, H. Prevalence and prognosis of gastric cancer detected by screening in a large Japanese population: data from a single institute over 30 years. *J. Gastroenterol. Hepatol.* **2007**, *22*, 1435-42, doi:10.1111/j.1440-1746.2007.04991.

9. The Study Group of Millenium Genome Project for Cancer. Genetic

variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. *Nat. Genet.* **2008**, *40*, 730–740, doi:10.1038/ng.152.
10. Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M. R., Armean, I. M., Bennett, R., Bhai, J., Billis, K., Boddu, S. *et al.*, Ensembl 2019. *Nucleic Acids Res.* **2019**, *47*, D745-D751. doi:10.1093/nar/gky1113.
11. Tanikawa, C., Urabe, Y., Matsuo, K., Kubo, M., Takahashi, A., Ito, H., Tajima, K., Kamatani, N., Nakamura, Y., Matsuda, K. A genome-wide association study identifies two susceptibility loci for duodenal ulcer in the Japanese population. *Nat. Genet.* **2012**, *44*, 430–434, doi:10.1038/ng.1109.
12. Gu, Z., Thomas, G., Yamashiro, J., Shintaku, I. P., Dorey, F., Raitano, A., Witte, O. N., Said, J. W., Loda, M., Reiter, R. E. Prostate stem cell antigen (PSCA) expression increases with high gleason score, advanced stage and bone metastasis in prostate cancer. *Oncogene.* **2000**, *19*, 1288-1296, doi:10.1038/sj.onc.120342619.
13. Wu, X., Ye, Y., Kiemeney, L. A. *et al.* Genetic variation in the prostate stem cell antigen gene PSCA confers susceptibility to urinary bladder cancer. *Nat Genet.* **2009**, *41*, 991-996, doi:10.1038/ng.421.
14. Saeki, N., Gu, J., Yoshida, T., Wu, X. Prostate stem cell antigen: a Jekyll and Hyde molecule? *Clin Cancer Res.* **2010**, *16*, 3533-8, doi: 10.1158/1078-0432.CCR-09-3169.
15. Park, B., Yang, S., Lee, J., Woo, H.D., Choi, I.J., Kim, Y.W., Ryu, K.W., Kim, Y.I., Kim, J. Genome-Wide Association of Genetic Variation in the PSCA Gene with Gastric Cancer Susceptibility in a Korean Population. *Cancer Res. Treat.* **2019**, *51*, 748-757, doi:10.4143/crt.2018.162.
16. Turdikulova, S., Dalimova, D., Abdurakhimov, A., Adilov, B., Yusupbekov, A., Djuraev, M., Abdujapparov, S., Egamberdiev, D., Mukhamedov, R. Association of rs2294008 and rs9297976 Polymorphisms in PSCA Gene with Gastric Cancer Susceptibility in Uzbekistan. *Cent. Asian J. Glob. Heal.* **2016**, *5*, 227, doi:10.5195/cajgh.2016.227.
17. García-González, M.A., Bujanda, L., Quintero, E., Santolaria, S., Benito, R., Strunk, M., Sopeña, F., Thomson, C., Pérez-Aisa, A., Nicolás-Pérez, D. *et al.*, Association of PSCA rs2294008 gene variants with poor prognosis and increased susceptibility to gastric cancer and decreased risk of duodenal

ulcer disease. *Int. J. Cancer.* **2015**, *137*, 1362–1373, doi:10.1002/ijc.29500.
18. Oleksyk, T. K., Smith, M. W., O'Brien, S. J. Genome-wide scans for footprints of natural selection. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2010**, 365, 185-205. doi:10.1098/rstb.2009.0219
19. Nakayama, K., Ohashi, J., Watanabe, K., Munkhtulga, L., Iwamoto, S. Evidence for Very Recent Positive Selection in Mongolians. *Mol Biol Evol.* **2017**, *34*, 1936-1946, doi: 10.1093/molbev/msx138.
20. Nei, M., Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA.* **1979**, *76*, 5269-5273, doi:10.1073/pnas.76.10.5269.
21. Hudson, R. R., Slatkin, M., Maddison, W. P. Estimation of levels of gene flow from DNA sequence data. *Genetics.* **1992**, *132*, 583–589.
22. Bhatia, G., Patterson, N., Sankararaman, S., Price, A. L. Estimating and interpreting FST: The impact of rare variants. *Genome Res.* **2013**, *23*, 1514–1521, doi:10.1101/gr.154831.113.
23. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* **1989**, *123*, 585–595.
24. Fay, J., Wu, C. -I. Hitchhiking Under Positive Darwinian Selection. *Genetics.* **2000**, *155*, 1405–1413, doi:10.1007/s11434-008-0202-z.
25. Zeng, K., Fu, Y. -X., Shi, S., Wu, C. -I. Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics.* **2006**, *174*, 1431-1439, doi:10.1534/genetics.106.061432.
26. Yi, X., Huerta-Sanchez, E., Jin, X., Cuo, Z. X., Pool, J. E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T. S., Zheng, H. *et al.* Sequencing of 50 human exomes reveals adaptation to high altitude. *Science.* **2010**, *329*, 75-78, doi:10.1126/science.1190371.
27. Cavalli-Sforza, L. L. Human diversity, *Proc. 12th Int. Congr. Genet.* **1969**, *3*. Tokyo. pp.405-416.
28. Wakeley, J., Alicar, N. Gene genealogies in a metapopulation. *Genetics*, **2002**, *159*, 893–905.
29. Przeworski, M. The signature of positive selection at randomly chosen loci. *Genetics.* **2002**, *160*, 1179-1189.
30. Fujito, T. N., Satta, Y., Hayakawa, T., Takahata, N. A new inference

method for detecting an ongoing selecive sweep. *Genes. Genet. Syst.* **2018,** *93*, 149-161.,doi:10.1266/ggs.18-00008.
31. Satta, Y., Zeng, W., Nishiyama, K. V., Iwasaki, R. L., Hayakawa, T., Fujito, N. T., Takahata, N. Two-dimensional site frequency spectrum for detecting, classifying and dating incomplete selective sweeps. *Genes. Genet. Syst.* **2019**, *94*, 283-300. doi:10.1266/ggs.19-00012.
32. Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V, Patterson, N. J., McDonald, G. J. *et al.*, Detecting recent positive selection in the human genome from haplotype structure. *Nature.* **2002**, *419*, 832–837, doi:10.1038/nature01140.
33. Ferrer-Admetlla, A., Liang, M., Korneliussen, T., Nielsen, R. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol. Biol. Evol.* **2014**, *31*, 1275–1291, doi:10.1093/molbev/msu077.
34. Garud, N. R., Messer, P. W., Buzbas, E. O., Petrov, D. A. Recent selective sweeps in North American Drosophila melanogaster show signatures of soft sweeps. *PLoS. Genet.* **2015**, *11*, e1005004, doi:10.1371/journal.pgen.1005004.
35. Hasebe, K. The ancient Japanese. *J. Anth. Sci. Nippon*, **1940**, *55*, 27-34. (in Japanese)
36. Howells, W. W. The Jomon people of Japan: a study by discriminant analysis of Japanese and Ainu crania. *Paper Peabody Museum Arch. Ethno Harvard Univ.* **1966**, *57*, 1-43.
37. Kiyono, K. *Kodaijinkotsu No Kenkyu Ni Motozuku Nihonjinshuron*. Iwanami Shoten, Tokyo, 1949. (in Japanese)
38. Hanihara, K. Dual Structure Model for the Population History of the Japanese. *Japan Rev.* **1991**, *2*, 1–33, doi:10.1537/ase.102.455.
39. Gakuhari, T., Nakagome, S., Rasmussen, S., Allentoft, M. E., Sato, T., Korneliussen, T., Chuinneagáin, B. N., Matsumae, H., Koganebuchi, K., Schmidt, R. et al. Ancient Jomon genome sequence analysis sheds light on migration patterns of early East Asian populations. *Commun. Biol.* **2020**, *3*, 1-10. doi:10.1038/s42003-020-01162-2.
40. Kanzawa-Kiriyama, H., Jinam, T. A., Kawai, Y., Sato, T., Hosomichi, K., Tajima, A., Adachi, N., Matsuumra, H., Kryukov, K., Saito, N., *et al.* Late

Jomon male and female genome sequences from the Funadomari site in Hokkaido, Japan. *Anthropol. Sci.* **2019**, *127*, 83-108, doi:10.1537/ase.190415.
41. Saitou, N. *History of Japanese Archiplelago people.* Tokyo, Japan: Iwanami shoten, 2015. ISBN 9784005008124 (in Japanese).
42. Jinam, T. A., Kanzawa-Kiriyama, H., Inoue, I., Tokunaga, K., Omoto, K., Saitou, N. Unique characteristics of the Ainu population in Northern Japan. *J. Hum. Genet.* **2015**, *60*, 565–571, doi:10.1038/jhg.2015.79.
43. Nakagome, S., Sato, T., Ishida, H., Hanihara, T., Yamaguchi, T., Kimura, R., Mano, S., Oota, H., Omoto, K., Tokunaga, K., *et al.* Model-based verification of hypotheses on the origin of modern Japanese revisited by Bayesian inference based on genome-wide SNP data. *Mol. Biol. Evol.* **2015**, *32*, 1533–1543, doi:10.1093/molbev/msv045.
44. Kanzawa-Kiriyama, H., Kryukov, K., Jinam, T.A., Hosomichi, K., Saso, A., Suwa, G., Ueda, S., Yoneda, M., Tajima, A., Shinoda, K. I., *et al.* A partial nuclear genome of the Jomons who lived 3000 years ago in Fukushima, Japan. *J. Hum. Genet.* **2017**, *62*, 213–221, doi:10.1038/jhg.2016.110.
45. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature.* **2015**, *526*, 68–74, doi:10.1038/nature15393.
46. Kaniwa, N., Sugiyama, E., Saito, Y., Kurose, K., Maekawa, K., Hasegawa, R., Furuya, H., Ikeda, H., Takahashi, Y., Muramatsu, M. *et al.* Specific HLA types are associated with antiepileptic drug-induced Stevens-Johnson syndrome and toxic epidermal necrolysis in Japanese subjects. *Pharmacogenomics.* **2013**, *14*, 1821-31, doi: 10.2217/pgs.13.180.
47. Tohkin, M., Kaniwa, N., Saito, Y., Sugiyama, E., Kurose, K., Nishikawa, J., Hasegawa, R., Aihara, M., Matsunaga, K., Abe, M. *et al.* A whole-genome association study of major determinants for allopurinol-related Stevens-Johnson syndrome and toxic epidermal necrolysis in Japanese patients. *Pharmacogenomics J.* **2013**, *13*, 60-9. doi: 10.1038/tpj.2011.41.
48. Chen, C., Yang, J., Chiang, C. W. K., Hsiung, C., Wu, P., Chang, L., Chu, H., Chang, J., Song, I., Yang, S. *et al.* Population structure of Han Chinese in the modern Taiwanese population based on 10,000 participants in the Taiwan Biobank project. *Hum. Mol. Genet.* **2016**, *25*, 5321–5331, doi:10.1093/hmg/ddw346.

49. Bae, J. S., Cheong, H. S., Kim, J. O., Lee, S. O., Kim, E. M., Lee, H. W., Kim, S., Kim, J. W, Cui, T., Inoue, I. *et al.* Identification of SNP markers for common CNV regions and association analysis of risk of subarachnoid aneurysmal hemorrhage in Japanese population. *Biochem. Biophys. Res. Commun.* **2008**, *373*, 593–596, doi:10.1016/j.bbrc.2008.06.083.
50. Kimura, M., Ohta, T. The age of a neutral mutant persisting in a finite population. *Genetics.* **1973**, *75*, 199-212. doi:10.1017/S0016672300014750.
51. Toyoshima, O., Tanikawa, C., Yamamoto, R., Watanabe, H., Yamashita, H., Sakitani, K., Yoshida, S., Kubo, M., Matsuo, K., Ito, H. *et al.*, Decrease in PSCA expression caused by Helicobacter pylori infection may promote progression to severe gastritis. *Oncotarget.* **2017**, *9*, 3936-3945. doi:10.18632/oncotarget.23278.
52. Jeon, S., Bhak, Y., Choi, Y., Jeon, Y., Kim, S., Jang, J., Jang, J., Blazyte, A., Kim, C., Kim, Y. et al., Korean Genome Project: 1094 Korean personal genomes with clinical information. *Sci. Adv.* **2020**, *6*, eaaz7835. doi:10.1126/sciadv.aaz7835.
53. Takahata, N., Satta, Y., Klein, Y. Polymorphism and balancing selection at major histocompatibility complex loci. *Genetics.* **1992**, *130*, 925-938, doi: 10.1007/978-3-642-51479-1_20.
54. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Software.* **2018**, 3, 731, 1-2, doi:10.21105/joss.00731.
55. Hartl, D. L., Clark, A. G. Principles of Population genetics, 4th edition. *J. Hered.* **2007**, 98, *382*, doi:10.1093/esm035.
56. Barrett, J. C., Maller, J. and Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinfomatics*. **2005**, 21, 2, 263-265, doi:10.1093/bioinfomatics/bth457.
57. Gabriel, S.B. The Structure of Haplotype Blocks in the Human Genome. *Science.* **2002**, *296*, 2225–2229, doi:10.1126/science.1069424.
58. Szpiech, Z.A., Hernandez, R.D. Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* **2014**, *31*, 2824–2827, doi:10.1093/molbev/msu211.
59. Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S.,

Librado, P., Ramos-Onsins, S. E., Sánchez-Gracia, A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302, doi:10.1093/molbev/msx248.

60. Takahata, N., Tajima, F. Sampling Errors in Phylogeny. *Mol. Biol. Evol.* **1991**, *8*, 494, doi:10.1093/oxfordjournals.molbev.a040669.

61. Benjamini, Y., Hochberg, Y. Controlling the false discovery rate: a partical and powerful approach to multiple testing. *J. Roy. Stat. Soc. Ser.* **1995**, *B57*, 289-300, doi: 10.2307/2346101.

62. Scally, A., Durbin, R. Revising the human mutation rate: implications for understanding human evolution. *Nat. Rev. Genet.* **2012.** *13*, 745-753, doi:10.1038/nrg3295.

63. Schaffner, S. F., Foo, C., Gabriel, S., Reich, D., Daly, M. J., Altshuler, D. Calibrating a coalescent simulation of human genome sequence variation. *Genome. Res.* **2005**, *15*, 1576-1583, doi:10.1101/gr.3709305.

64. Hudson, R. R. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinfomatics.* **2002**, *18*, 337-338, doi:10.1093/bioinformatics/18.2.337.

65. Brown, M. B. 400: A Method for Combining Non-Independent, One-Sided Tests of Significance. *Biometrics.* **1975**, *31*, 987-92, doi:10.2307/2529826.

66. Bandelt, H. J., Forster, P., Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* **1999**, *16*, 37-48. doi:10.1093/oxfordjournals.molbev.a026036

67. Ewens, W. J. *Mathematical Population Genetics 1*, Biomathematics vol. 9. Springer Verlag: New York, **1979**.

68. Habu, J. *Ancient Jomon of Japan*, Cambridge University Press: Cambridge, UK, **2004**, ISBN 0521776708.

69. Jacobs, G. S., Hudjashov, G., Saag, L., Kusuma, P., Darusallam, C. C., Lawson, D. J., Mondel, M., Pagani, L., Ricaut, F. -X., Stoneking, M. *et al.* Multiple Deeply Divergent Denisovan Ancestries in Papuans. *Cell.* **2019**, *177*, 1010-1021, doi:10.1016/j.cell.2019.02.035.

70. Terhorst, J.: Kamm, J. A., Song, Y, S. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat Genet.* **2017**, *49*, 303-309, doi:10.1038/ng.3748.

71. Fenner, J. N. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* **2005**, *128*, 415–423, doi:10.1002/ajpa.20188.
72. Park L. Effective population size of korean populations. *Genomics Inform.* **2014**, *12*, 208-15, doi: 10.5808/GI.2014.12.4.208.
73. McColl, H., Racimo, F., Vinner, L., Demeter, F., Gakuhari, T., Moreno-Mayar, J. V., van Driem, G., Gram Wilken, U., Selguin-Orlando, A., de la Fuente Castro, C. *et al.* The prehistoric peopling of Southeast Asia. *Science.* **2018**, *361*, 88-92, doi:10.1126/science.aat3628.
74. Sikora, M., Seguin-Orlando, A., Sousa, V. C., Albrechtsen, A., Korneliussen, T., Ko, A., Rasmussen, S., Dupanloup, I., Nigst, P. R., Bosch, M. D. et al. Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science.* **2017**, *358*, 659-662. doi:10.1126/science.aao1807.
75. Wang, Y., Lu, D., Chung, Y. –J, Xu, S. Genetic structure, divergence a1nd admixture of Han Chinese, Japanese and Korean populations. *Hereditas.* **2018**, *155*, 19. doi: 10.1186/s41065-018-0057-5.
76. Sato, T., Nakagome, S., Watanabe, C., Yamaguchi, K., Kawaguchi, A., Koganebuchi, K., Haneii, K., Yamaguchi, T., Hanihara, T., Yamamoto, K. *et al.* Genome-wide SNP analysis reveals population structure and demographic history of the ryukyu islanders in the southern part of the Japanese archipelago. *Mol. Biol. Evol.* **2014**, *31*, 2929-2940, doi:10.1093/molbev/msu230.
77. Dou, T., Gu, S., Zhou, Z., Ji, C., Zeng, L., Ye, X., Xu, J., Ying, K., Xie, Y. Mao, Y. Note: Isolation and Characterization of a Jerky and JRK/JH8 Like Gene, Tigger Transposable Element Derived 7, TIGD7. *Biochem. Genet.* **2004**, *42*, 279-285, doi: 10.1023/B:BIGI.0000034428.95802.35.
78. Pinto, D. Dissecting the Genetic Basis of Idiopathic Epilepsies [PhD thesis]. Utrecht University, 2006. ISBN 90-393-4191-5.
79. Morita, R., Miyazaki, E., Shah, P. U., Castroviejo, I. P., Delgado-Escueta, A. V., Yamakawa, K. Exclusion of the JRK/JH8 gene as a candidate for human childhood absence epilepsy mapped on 8q24. *Epilepsy Res.* **1999**, *37*, 151–158, doi: 10.1016/s0920-1211(99)00061-3.

80. Chen, T., Giri, M., Xia, Z. Y., Subedi, Y. N., Li, Y. Genetic and epigenetic mechanisms of epilepsy: a review. *Neuropsychiatr Dis Treat.* **2017**, *13*, 1841-1859, doi: 10.2147/NDT.S142032.
81. Rees, J. S., Castellano, S., Andrés, A. M. The Genomics of Human Local Adaptation. *Trends Genet.* 2020, *36*, 415-428. doi: 10.1016/j.tig.2020.03.006.
82. Anguita-Ruiz, A., Aguilera, C. M., Gil, Á. Genetics of Lactose Intolerance: An Updated Review and Online Interactive World Maps of Phenotype and Genotype Frequencies. *Nutrients.* 2020, *12*, 2689, doi: 10.3390/nu12092689.
83. Japanese Archipelago Human Population Genetics Consortium, Jinam, T. A., Nishida, N., Hirai, M., Kawamura, S., Oota, H., Umetsu, K., Kimura, R., Ohashi, J., Tajima, A. *et al.* The history of human populations in the Japanese Archipelago inferred from genome-wide SNP data with a special reference to the Ainu and the Ryukyuan populations. *J. Hum. Genet.* **2012**, *57*, 787–795, doi:10.1038/jhg.2012.114.
84. Fujio, S. *History of Yayoi period*. Tokyo, Japan: Kodansha, 2015. ISBN 9784062883306 (in Japanese).
85. Bergström, A., McCarthy, S. A., Hui, R., Almarri, M. A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J. *et al.* Insights into human genetic variation and population history from 929 diverse genomes. *Science.* **2020**, *367*, eaay5012, doi:10.1126/science.aay5012
86. Mocellin, S., Verdi, D., Pooley, K. A., Nitti, D. Genetic variation and gastric cancer risk: a field synopsis and meta-analysis. *Gut.* **2015**, *64*, 1209-1219. doi:10.1136/gutjnl-2015-309168

謝辞

はじめに、本研究を進めるにあたり、多くの先生方や関係者の方々にご協力いただきました。この場を借りて感謝を申し上げます。

まず、主任指導教員である颯田葉子教授には、本研究を始めるにあたっての機会をいただき、研究方向の相談から議論に至るまで、終始丁寧なご指導をいただきました。本テーマに出会う機会を与えていただけたこと、そして、研究を面白い方向に導くきっかけをいただけたことに感謝します。また、副指導教官の五條堀淳講師には、解析についても、あるいは研究の方向性についても毎回有意義なコメントをいただきました。本研究の共同研究の一人である、石谷孔司博士をご紹介いただいたことにも感謝いたします。更に、副指導教官の大田竜也准教授には、博士課程の研究をどのように表現するかというデザイン面や、方向性について、毎回有意義なコメントをいただきました。また、本郷一美准教授には、研究を進めるにあたって、考古学部門からみた人類に対する視点を提供していただきました。颯田葉子教授、五條堀淳講師、本郷一美准教授には、本審査の際の副査をお引き受けいただきました。また、大田竜也准教授には、主査をお引き受け頂き、博士論文の改定に当たって有意義なコメントを頂きました。高畑尚之名誉教授には、ラボミーティングを通じて、また、本研究を論文としてまとめるにあたって、有意義なコメントをいただきました。颯田研究室の皆様には、大変貴重な議論の機会をいただきました。また、研究室内外を問わず、ご指導いただいた諸先生方、総研大の関係者の皆様に深く感謝申し上げます。特に、寺井洋平助教並びに木下充代准教授には、研究生活の悩みを含め、精神的なサポートをしていただきました。本研究の一部は、総研大の教育開発センターより、助成を得て国際誌に掲載することができました。重ねて感謝申し上げます。

本研究の実施にあたり、総研大外部からも、様々な研究者の方にお世話になりました。

東京大学の太田博樹教授には、本審査の際の外部審査員をお引き受けいただき、博士論文の改定に当たって、たくさんの有意義なコメントをいただきました。また、伊川津縄文人１個体のデータをご提供いただいております。

国立科学博物館の神澤秀明博士、国立国際医療研究センターの河合洋介博士

には、船泊縄文人 2 個体の DNA データを使わせていただき、解析のご指導、ご協力をいただきました。また、産業技術総合研究所の石谷孔司博士には、伊川津縄文人 1 個体のデータを使わせていただきました。以上のお三方には研究をまとめるにあたって論文の共著者をお引き受けいただき、論文を改善にするにあたって、有意義なコメントをいただきました。

琉球大学の木村亮介准教授並びに金沢大学の佐藤丈寛博士より、Asian DNA Repository Consortium 所蔵の、アイヌ人類集団及び琉球集団のゲノムアレイデータを使わせていただきました。また、台湾の National Yang Ming University の Wen-Ya Ko 准教授には台湾集団のゲノムアレイデータを使わせていただき、国立遺伝学研究所の井上逸朗教授には韓国人のゲノムアレイデータを使わせていただきました。

また、シンガポールの Nanyang Technological University の Kim Hie Lim 助教には、151 個体の韓国人集団と、35 個体の日本人集団の貴重なゲノム SNP データをご提供いただき、共同研究を始める機会をいただきました。以上の方々にご提供いただいたデータは、研究において主張を補強する貴重な資料として活用させていただきました。

以上の方々をはじめとして、大変多くの方々にご協力いただきました。ここに感謝いたします。

最後に、どんな時も常に側で研究生活を見守ってくれていた家族に本研究の成果を捧げます。

図

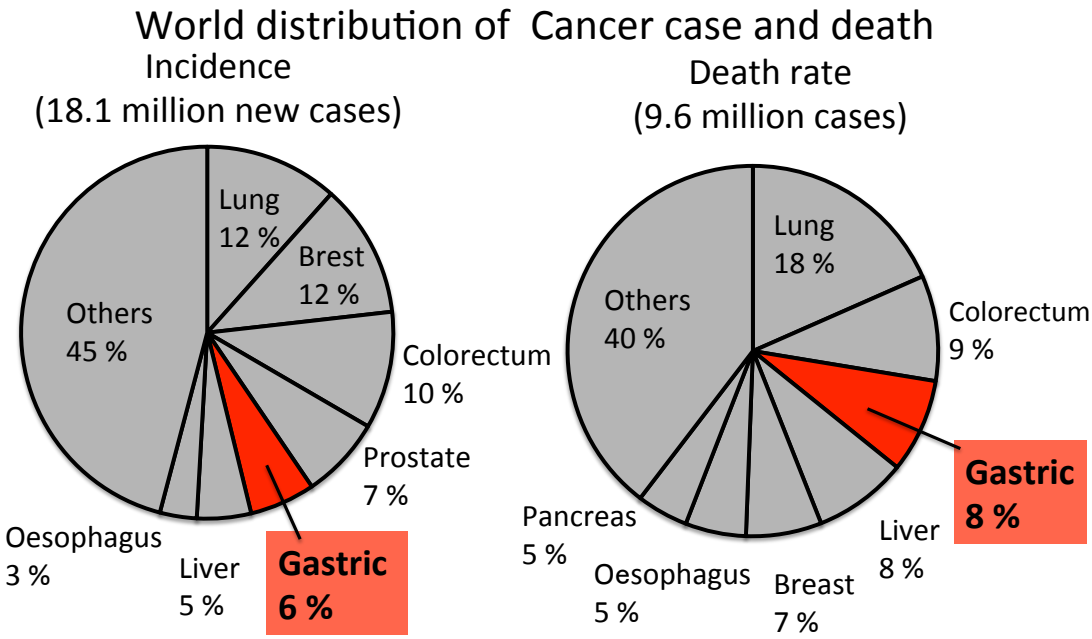

図 1. 2018 年の人類のがん発症率・死亡率の種類別の割合.
本図は GLOBOCAN[1]のデータを再編集した.

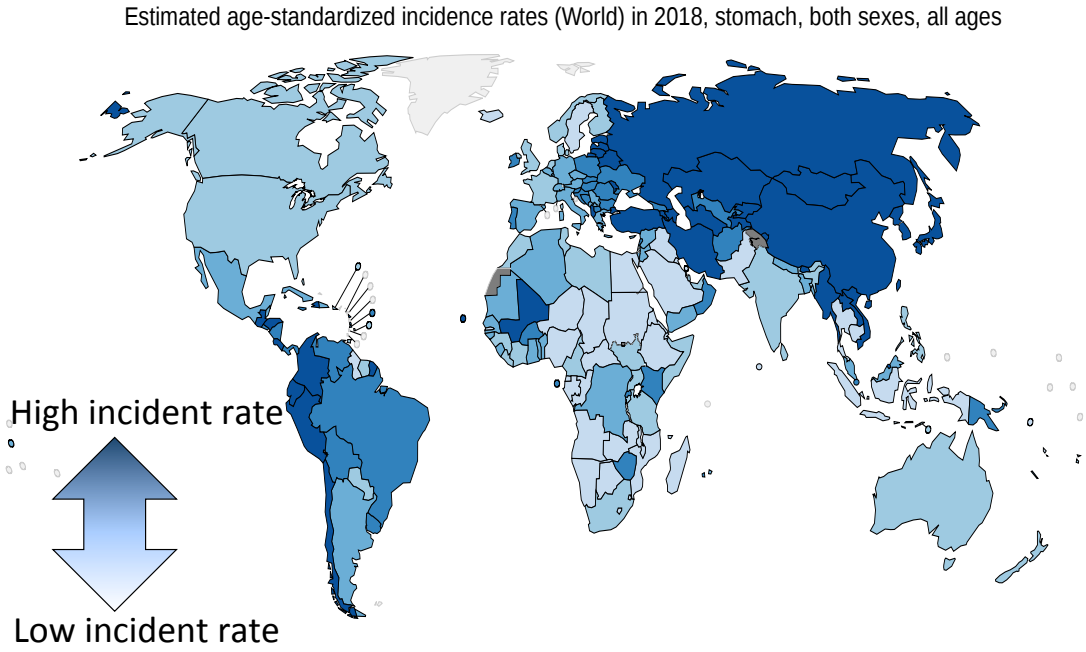

図 2. 地域別の胃がんの発症率.
胃がんの発症率の高い地域は濃い青、低い地域は薄い青で示した.
本図は GLOBOCAN[1]のデータを再編集した.

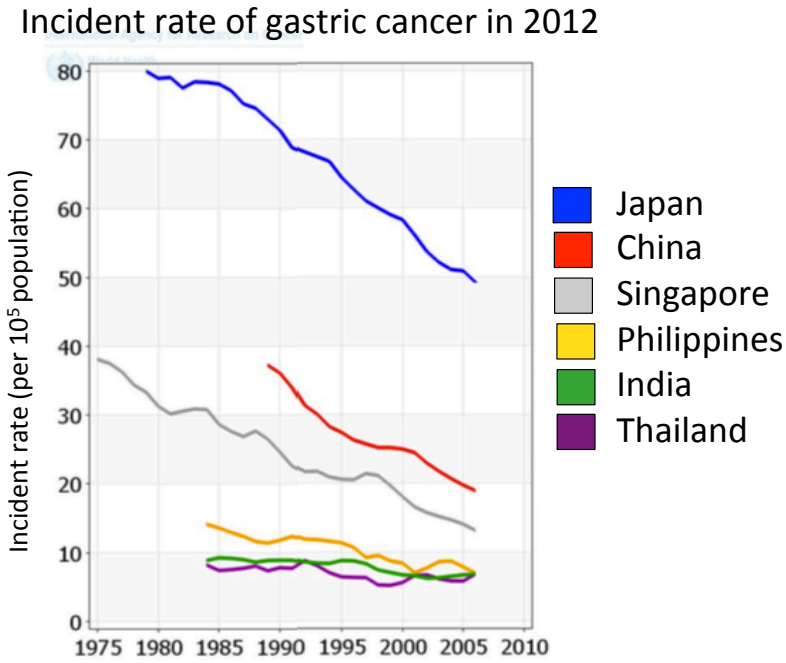

図 3. 東アジアの人類集団での男性の胃がんの発症率.
本図は WHO のデータ[3]をまとめた GLOBOCAN[1]のデータを再編集した.

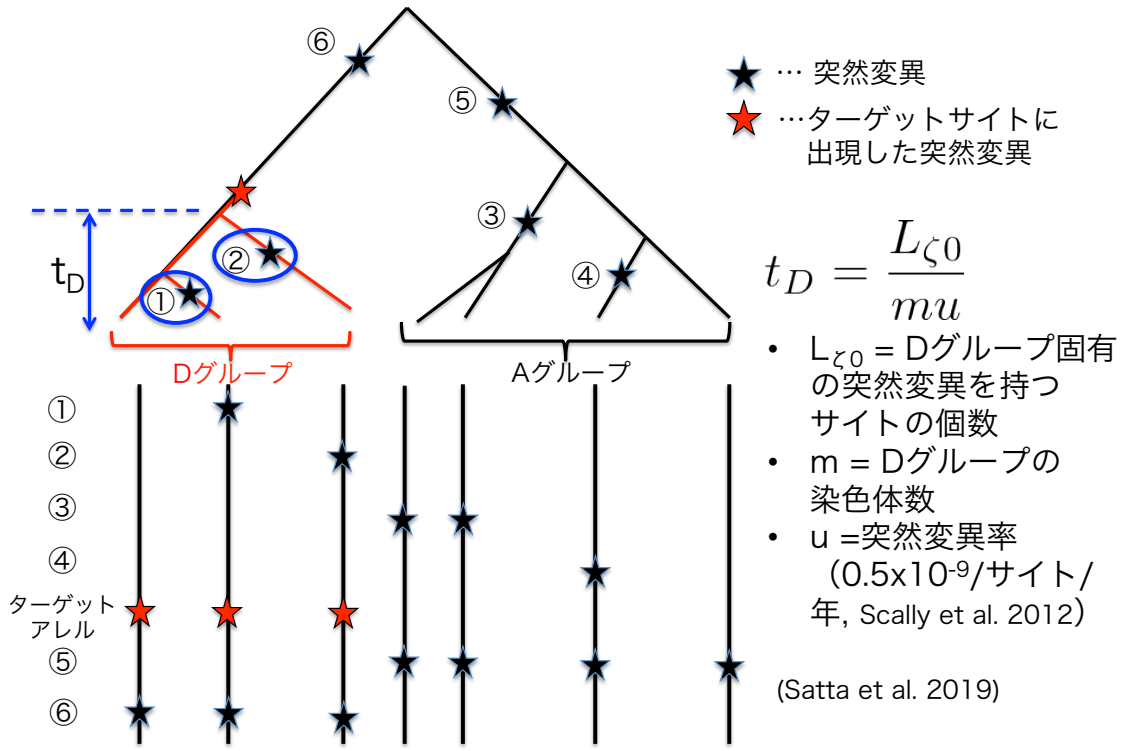

図 4. t_D による自然選択の開始時期の推定.

ターゲットアレルが正の自然選択のターゲットである場合、自然選択がいつまでに始まっていたかを D グループの IAV に基づいて推定する統計量である t_D と、系統樹上・染色体上に生じた突然変異の関係を示した．図中の赤い星は、突然変異によって生じたターゲットサイトに生じた適応度を上げるアレルを示し、黒い星は突然変異によって生じた中立なアレルを示す．図中の青丸は D グループのみに生じた突然変異を示し、D グループの IAV を示す要約統計量の一つである $L_{\zeta 0}$ によって表現される．また、図中の番号は、染色体上のサイトの位置と対応している．

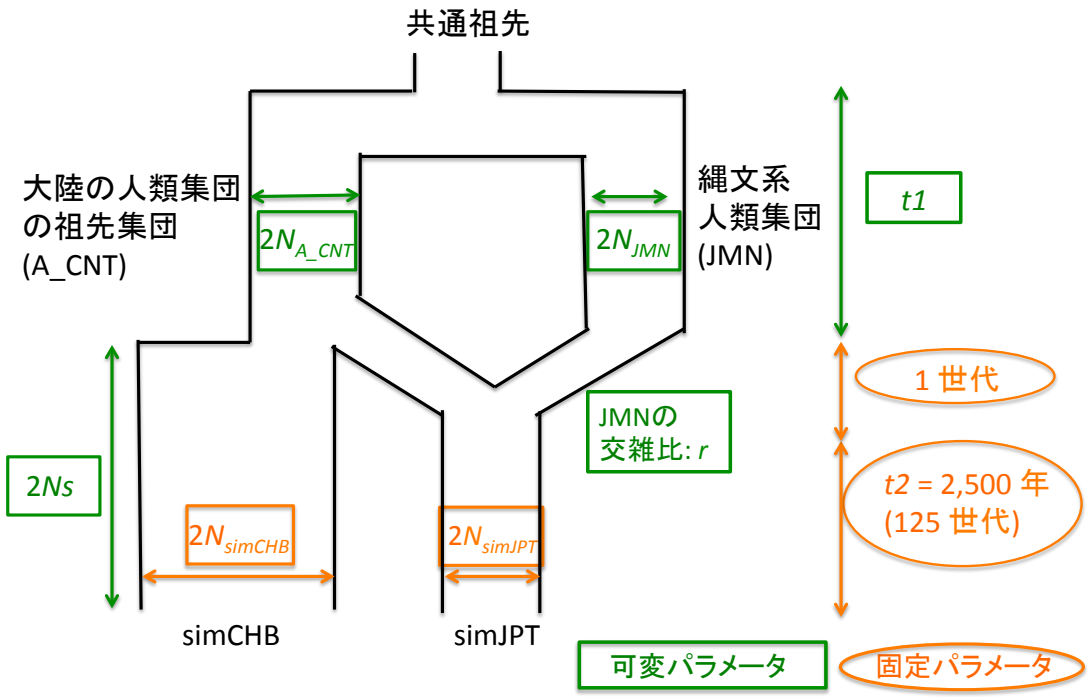

図 5. 縄文系人類集団（JMN）、大陸の人類集団の祖先集団（A_CNT）、日本人（simJPT）、漢族（simCHB）の集団動態モデル.

T と C のアレルが互いに中立な条件では、緑文字で示した 4 種のパラメータ（N_{JMN}、$N_{\mathrm{A\_CNT}}$、*t1* 及び *r*）、simCHB 系統に C アレルに自然選択が働いている条件では、中立な条件のもとでの 4 種のパラメータに加えて自然選択係数（2*Ns*）を含めた、5 種のパラメータがそれぞれ可変である．オレンジ色で示した 3 種のパラメータ（N_{simJPT}、N_{simCHB} 及び *t2*）は、このシミュレーションではいずれの条件下でも値を固定のものとした.

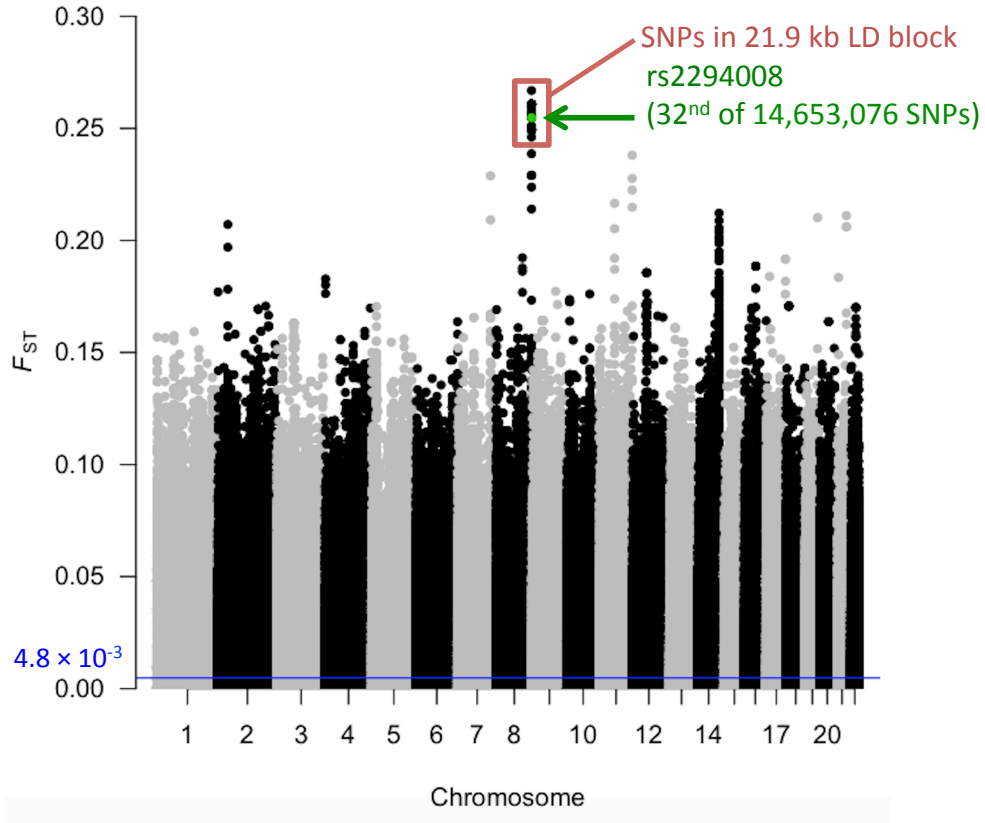

図 6. JPT-CHB 間のゲノムワイド SNP の F_{ST} 値によるマンハッタンプロット. 図中の各点は 常染色体にある 14,653,076 SNP の F_{ST} 値を示す. rs2294008 の F_{ST} 値を緑色に強調して示した. また、rs2294008 を含む LD ブロック (21.9 kb) に属する SNP を赤枠で強調して示した.これらの中で第 2 位の SNP (rs2717562) はこの LD ブロックの外側に位置する. ゲノムワイド SNP の F_{ST} 値の平均値は、青線で示した.

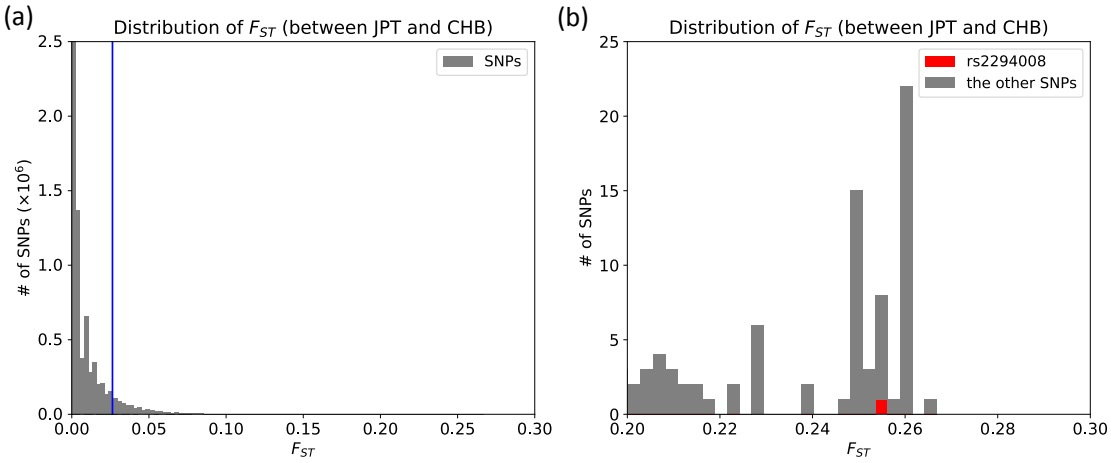

図 7. JPT-CHB 間のゲノムワイド SNP の F_{ST} 値の分布図.

（a）はゲノムワイド SNP の、JPT-CHB 間の F_{ST} 値の分布を表す. 上位 5 パーセンタイルの F_{ST} 値（$F_{ST} \geq 0.0264$）を、青線で示した.（b）は（a）の拡大図で、$0.20 \leq F_{ST} \leq 0.30$ の範囲を示す. また、rs2294008 は赤色のバーで示した.

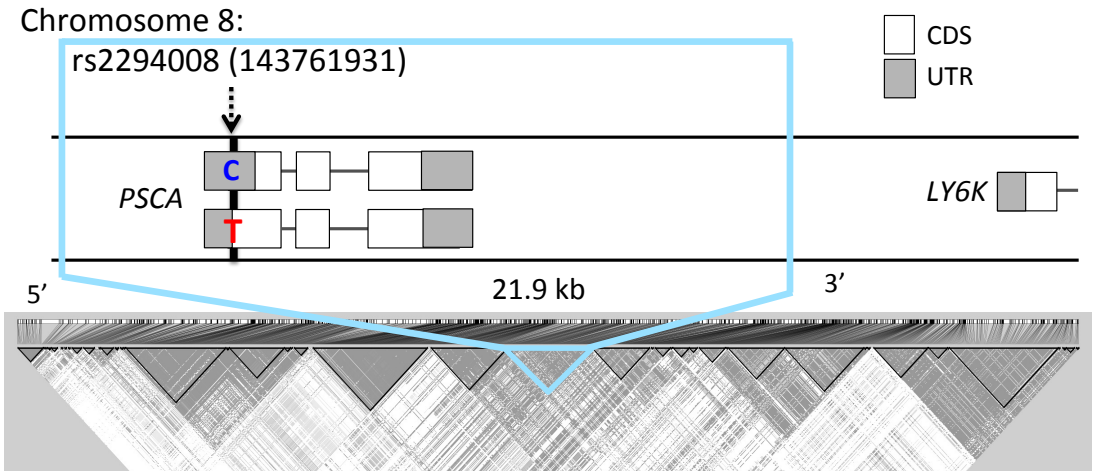

図 8. rs2294008 を含む *PSCA* 遺伝子とその近傍の遺伝子地図（上図）及び rs2294008 とその近傍の SNP との LD（下図）.

rs2294008 は *PSCA* 遺伝子のイニシエーションコドンの第二ポジションにある. C アレル（派生型）をコードする場合、T アレル（祖先型）に比べて、タンパク質に翻訳される領域（CDS）は、9 アミノ酸短くなる[11]. また、下図には、rs2294008（chr8: 143761931）及び、*PSCA* 遺伝子（12.4kb）の含まれる、21.9 kb の LD ブロックを水色の逆三角形で示した. 下図内の色の濃い部分は、アレル同士の連鎖関係が強いことを、色の薄い部分はアレル同士の連鎖関係が弱いことを示しており、逆三角形の領域は特に強く連鎖関係が保たれていることを示す.

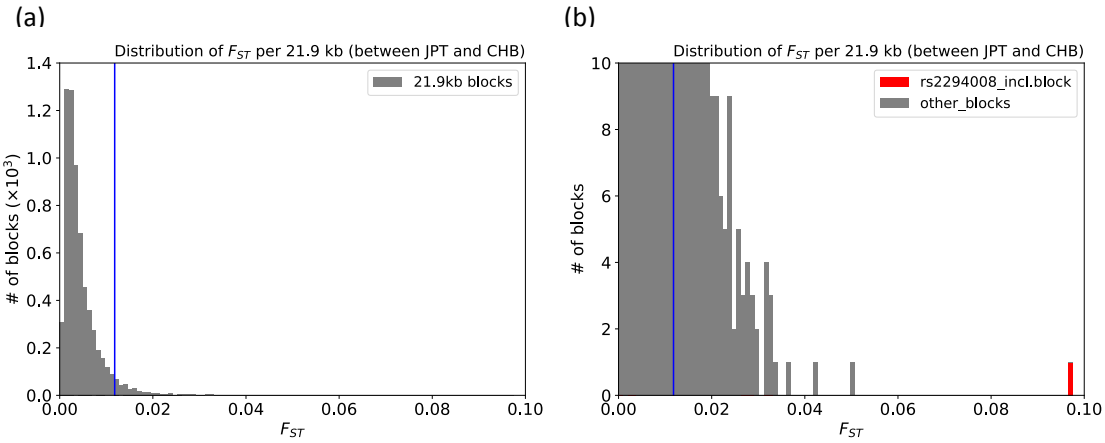

図 9. 21.9 kb ごとのブロックの F_{ST} 値の分布図.

8 番染色体を短碗の端から 21.9kb 毎のノンオーバーラップブロックに区切り、ブロック毎に含まれる SNP の F_{ST} を平均し、各ブロックの F_{ST} 値とした.

ブロックの総数は 6510 であった.

（a）全ブロックの F_{ST} の分布.（b）（a）の拡大図. rs2294008 を含むブロックは赤いバーで示し、他のブロックは灰色のバーで示した.

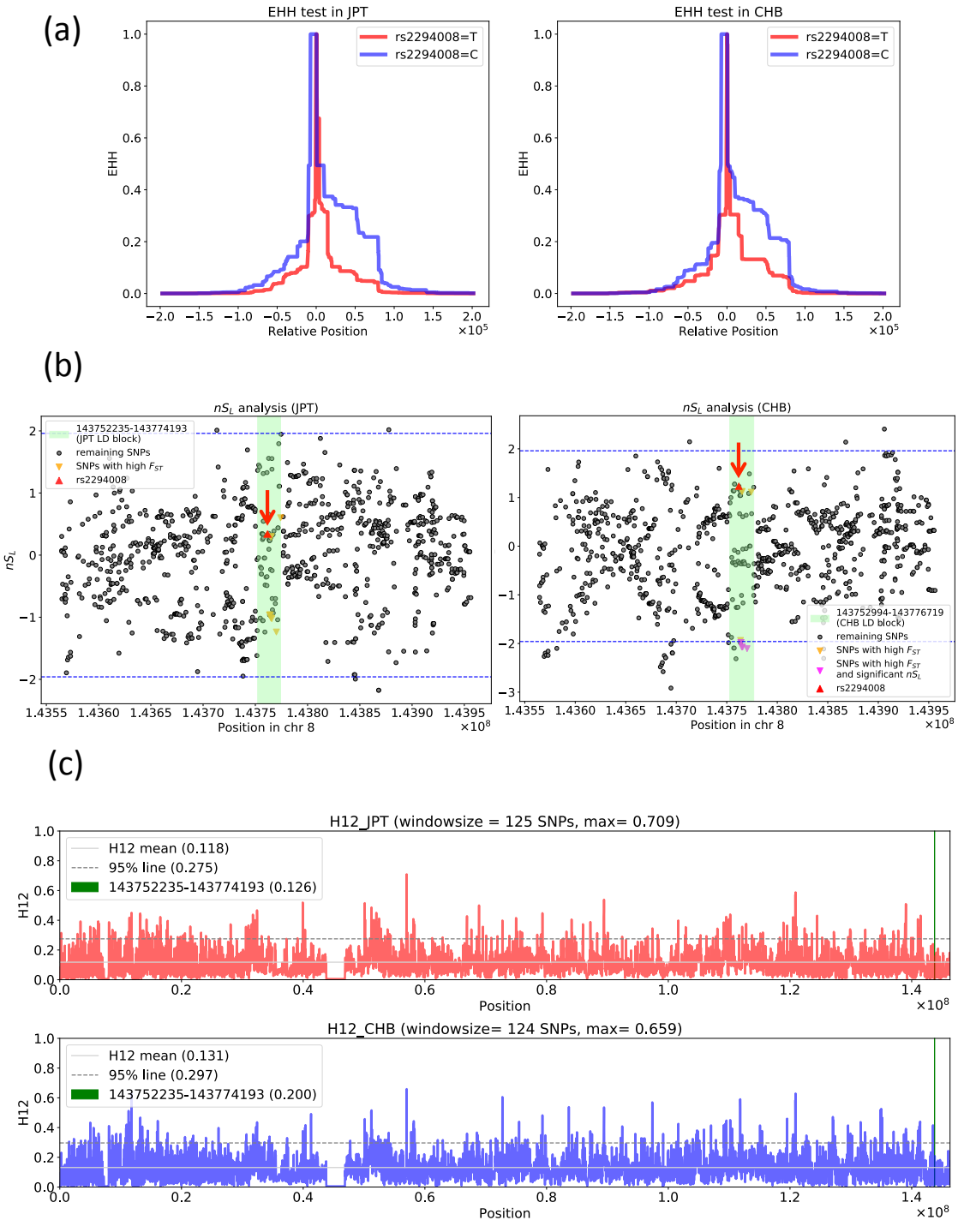

図 10. 各種の中立性検定（EHH、nS_L 及び H12）.

（a）rs2294008 をコアとして、この上流・下流の各 200 kb の EHH 値の減衰の程度を、コアとなるアレルで比較した．JPT での解析結果は左に、CHB での

解析結果は右に示した．また、コアを T アレルとする場合は赤、C アレルとする場合は青で示した．（b）rs2294008 とその近傍で、nS_L 値を解析した．rs2294008 及び、F_{ST} 値が 10 位までの SNP で、同じ LD ブロックに属する SNP をハイライトした. JPT での解析結果は左に、CHB での解析結果は右に示した．rs2294008 は赤い三角形で示し、nS_L 値が有意で、高い F_{ST} を持ち rs2294008 と連鎖関係にある SNP はピンク色の逆三角形で、高い F_{ST} を持つ SNP はオレンジの逆三角形で、そのほかの SNP は灰色の点で示した．それぞれの集団で定義した rs2294008 を含む LD ブロックの範囲は薄い緑色の帯で示した．nS_L の 95%有意水準値（-1.96~1.96）は青の点線で示した．（c）8 番染色体全体で H12 の sliding window analysis を行った．rs2294008 を含む JPT での LD ブロックは緑の帯で示した．H12 のそれぞれの集団での平均値は薄い灰色の実線で示し、全体の 95%ラインは灰色の点線で示した．

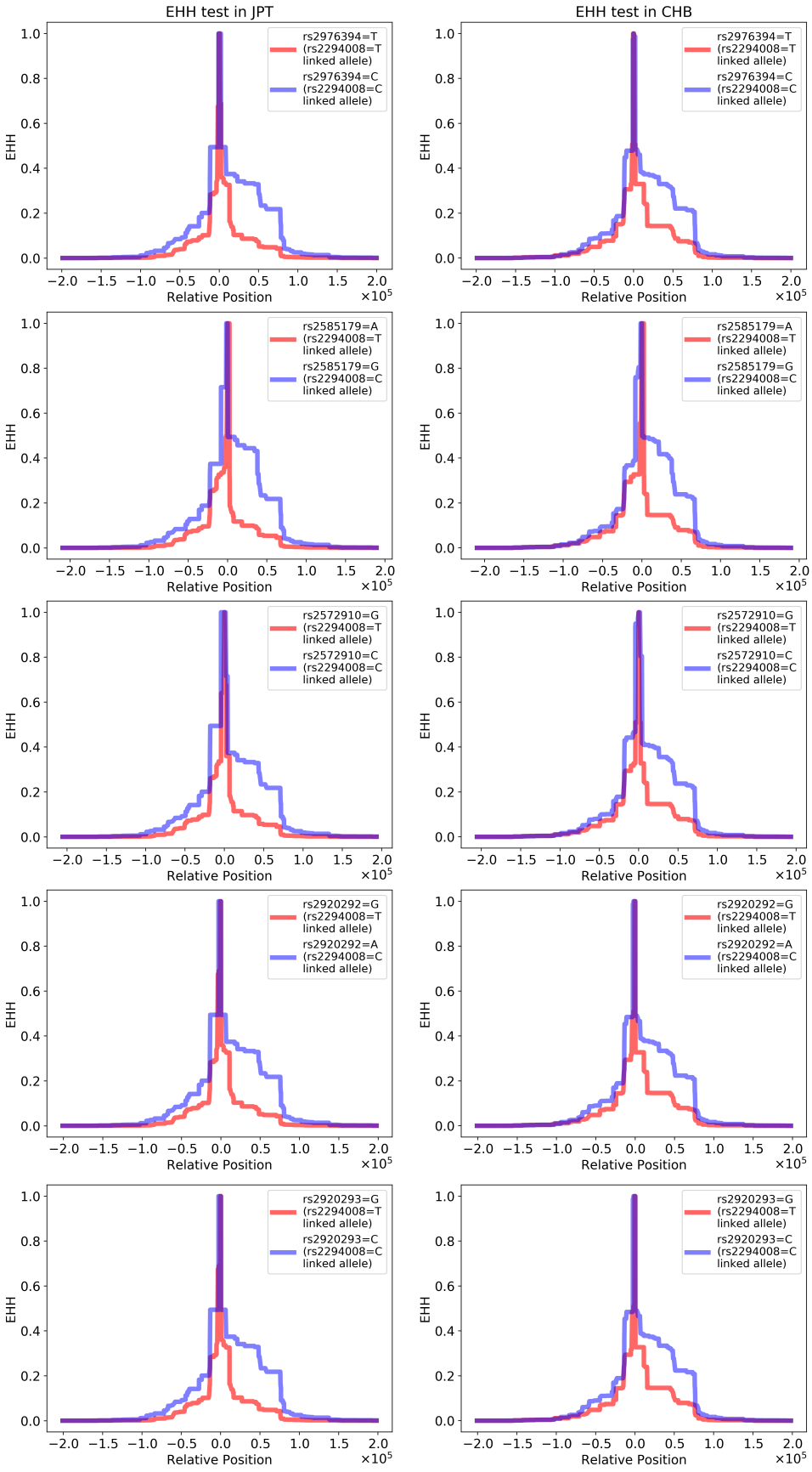
EHH test in JPT
EHH test in CHB
EHH
Relative Position
×10^5
rs2976394=T (rs2294008=T linked allele)
rs2976394=C (rs2294008=C linked allele)
rs2585179=A (rs2294008=T linked allele)
rs2585179=G (rs2294008=C linked allele)
rs2572910=G (rs2294008=T linked allele)
rs2572910=C (rs2294008=C linked allele)
rs2920292=G (rs2294008=T linked allele)
rs2920292=A (rs2294008=C linked allele)
rs2920293=G (rs2294008=T linked allele)
rs2920293=C (rs2294008=C linked allele)

図 11．rs2294008 と同じ LD ブロックに属し、高い F_{ST} を持つサイトをコアとして行った EHH 解析.
JPT での解析結果は左列に、CHB での解析結果は右列に示した．また、T アレルとリンクしているアレルをコアとする場合は赤、C アレルとリンクしている場合は青で示した.

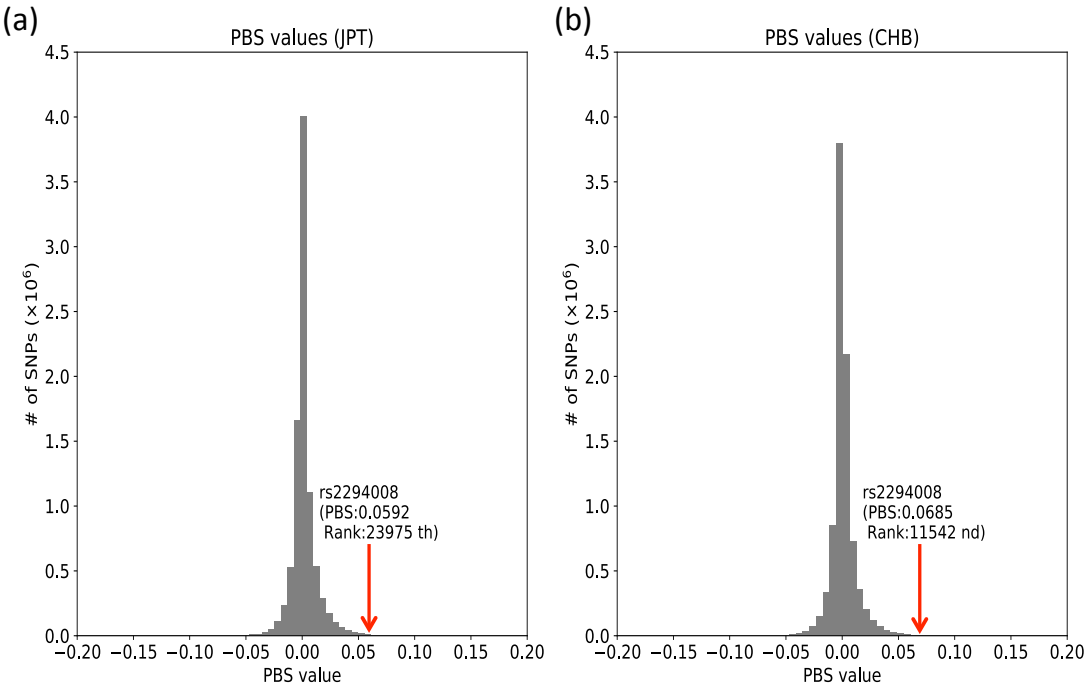

図 12. JPT 及び CHB の PBS の分布.

（a）JPT、（b）CHB. 9,051,837 SNP の PBS 値に基づく. 図中に rs2294008 の PBS 値及びランキングを示した. 赤矢印は rs2294008 を含む bin を示す.

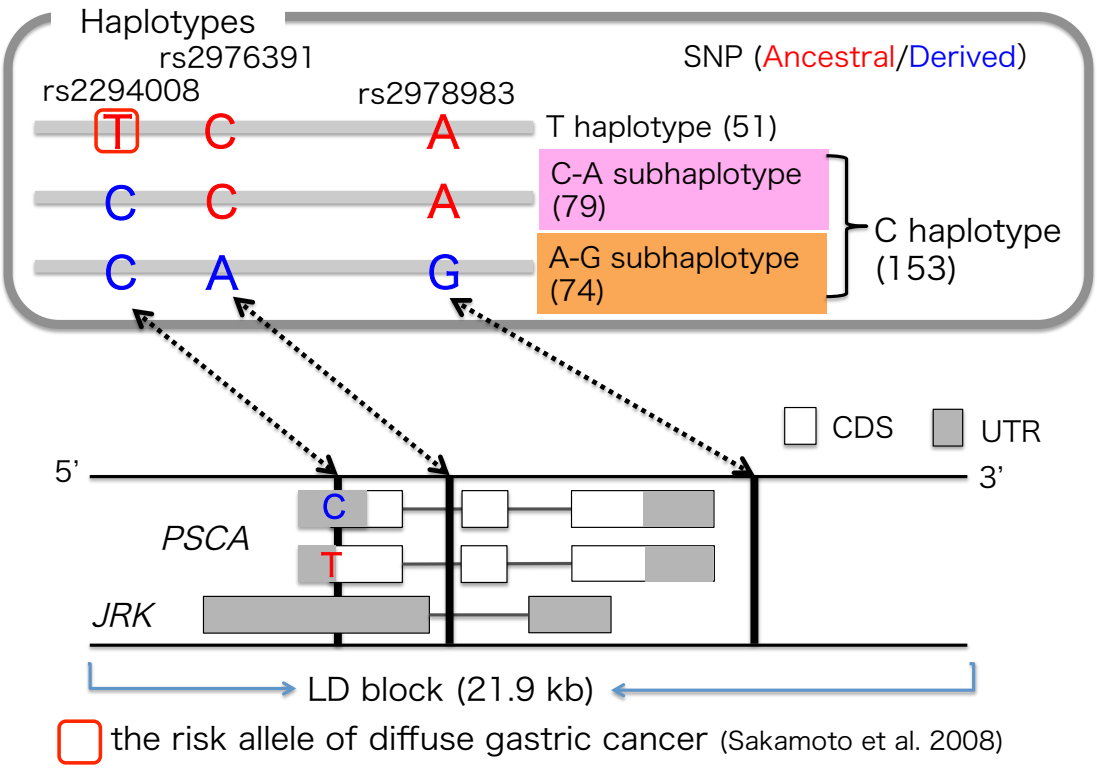

図 13. 各サブハプロタイプと SNP の関係

上図はサブハプロタイプを決定する 3SNP の染色体上の配置順と組み合わせを示す．また、ハプロタイプの後にあるカッコ内の数値は CHB での染色体本数を表す．また、下図には、3SNP の位置と *PSCA* 及び *JRK* 遺伝子との位置関係を示した．SNP 及び遺伝子の位置、タンパク質のコーディング領域（CDS）の情報は、Cunningham *et al.* [10]に従った．

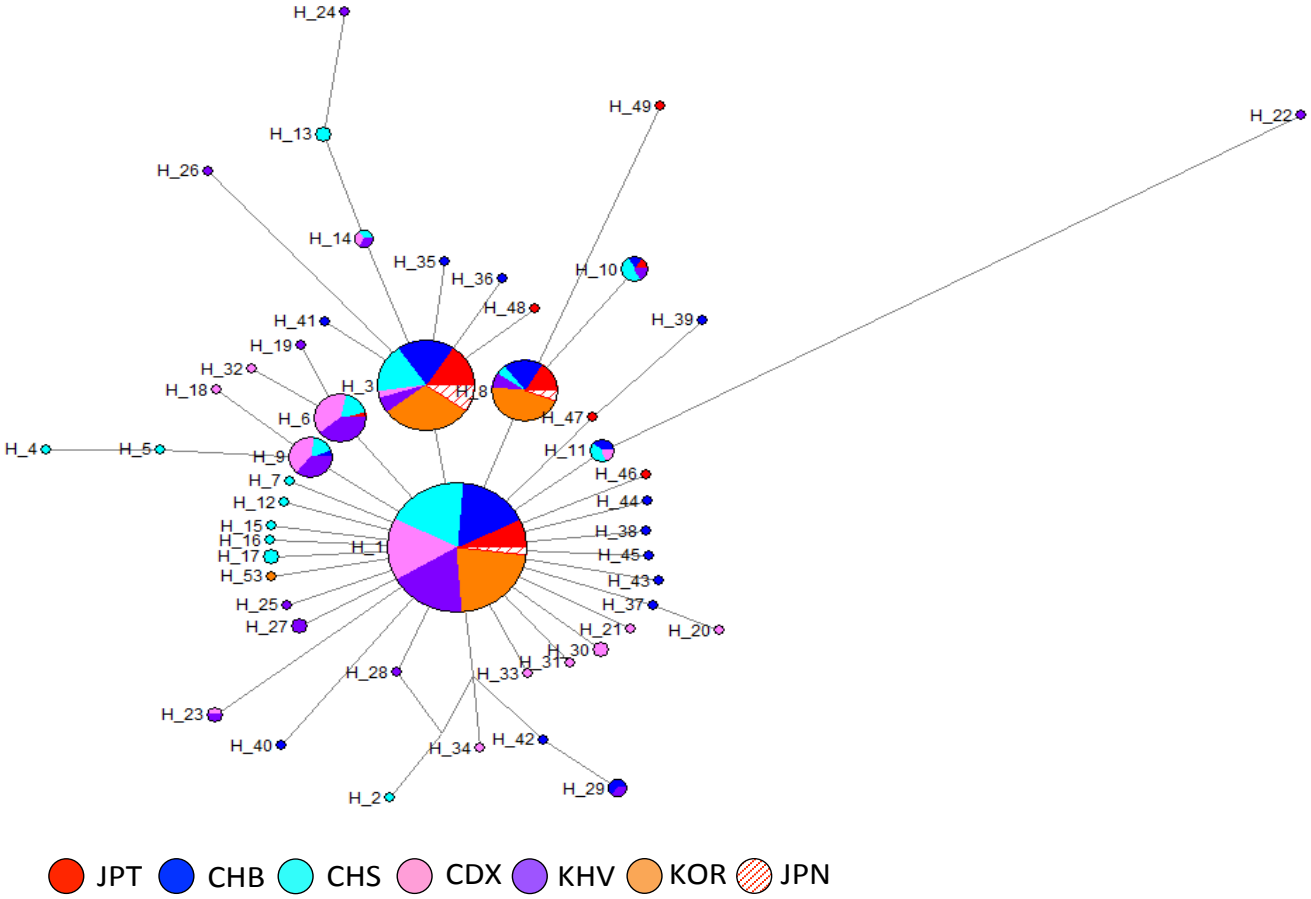

図 14. 東アジア集団の A-G サブハプロタイプのネットワーク解析.
図中のノードの円の大きさはハプロタイプ数を反映しており、円の中の色はハプロタイプの属する現代の各集団を示している. また、この図ではノード間の枝長は、ハプロタイプ間の塩基置換数を反映していない. 各ハプロタイプの名称・本数は表 5（a）の表記に従う.

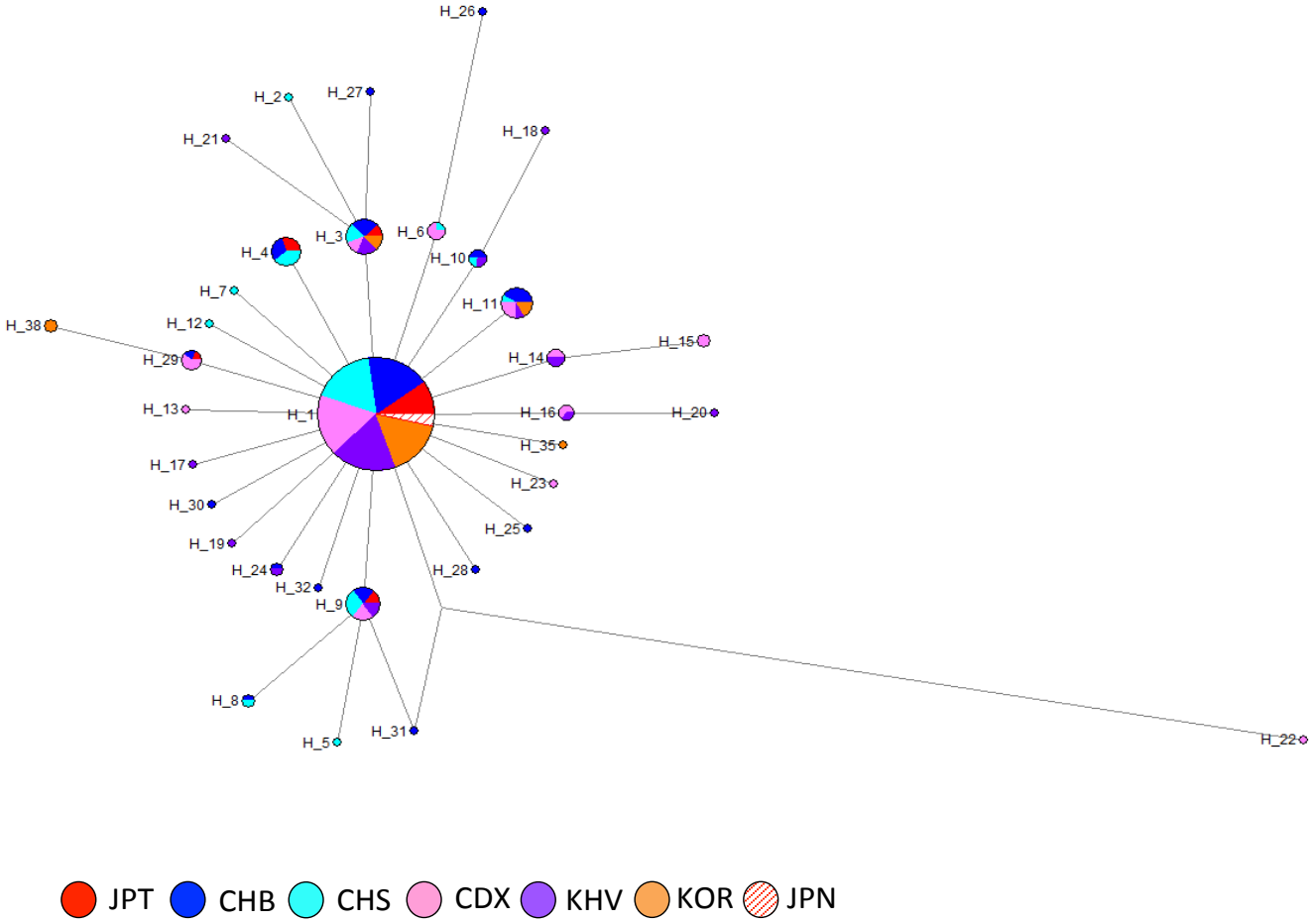

図 15. 東アジア集団の C-A サブハプロタイプのネットワーク解析.
図中のノードの円の大きさはハプロタイプ数を反映しており、円の中の色はハプロタイプの属する現代の各集団を示している. また、この図ではノード間の枝長は、ハプロタイプ間の塩基置換数を反映していない. 各ハプロタイプの名称・本数は表 5（b）の表記に従う.

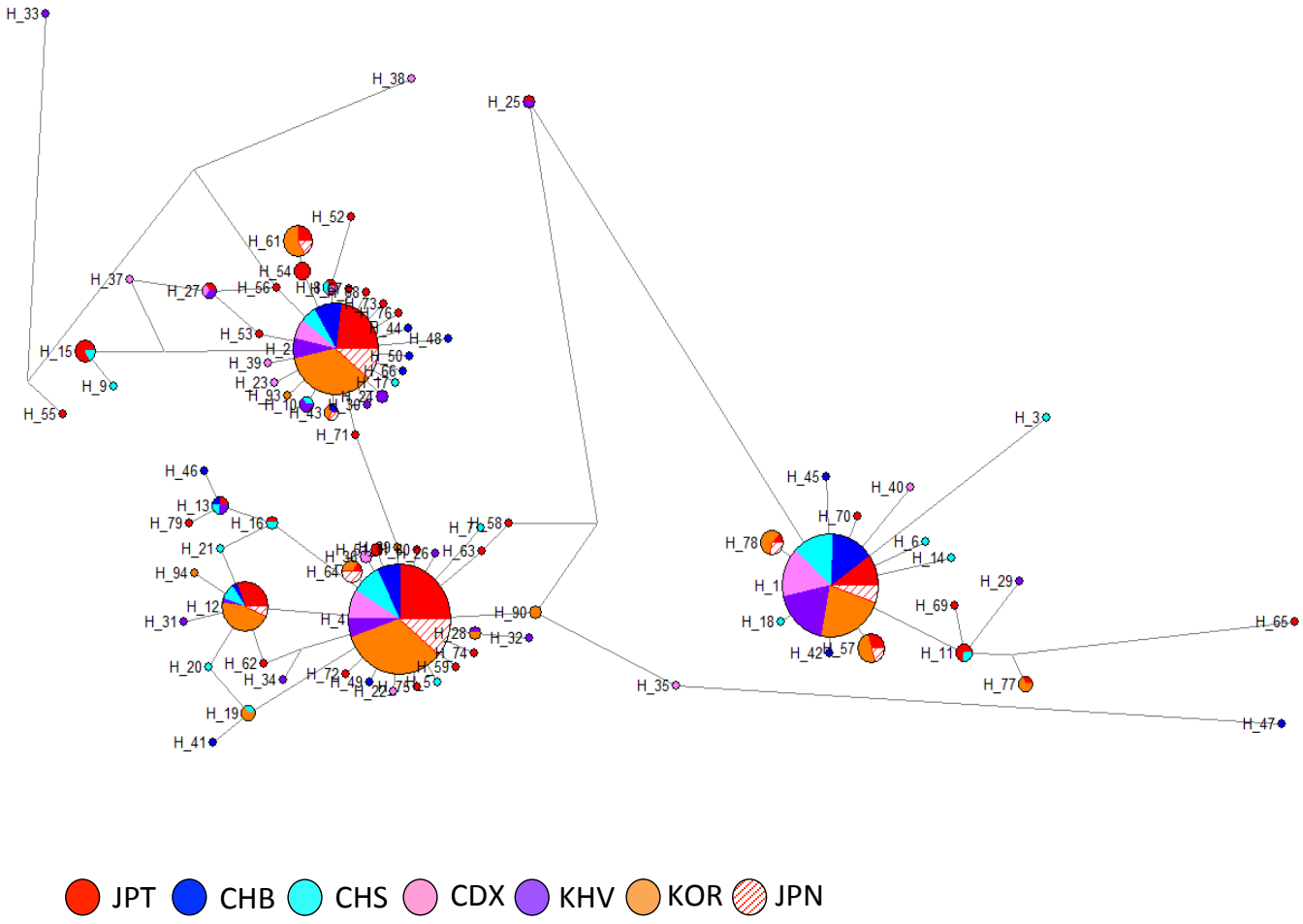

図 16. 東アジア集団の T ハプロタイプのネットワーク解析.
図中のノードの円の大きさはハプロタイプ数を反映しており、円の中の色はハプロタイプの属する現代の各集団を示している. また、この図ではノード間の枝長は、ハプロタイプ間の塩基置換数を反映していない. 各ハプロタイプの名称・本数は表 5（c）の表記に従う.

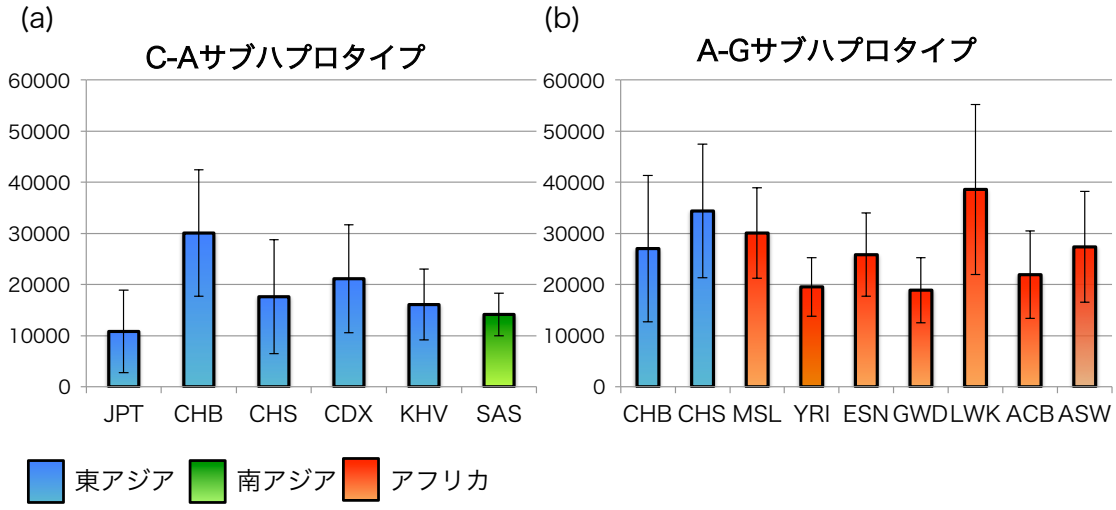

図 17. 自然選択下にある各集団の t_D 値.

各人類集団での rs2294008 の D グループでの TMRCA の平均値（t_D）及びその分散は、Satta *et al.*の手法[39]に従って算出した. 各人類集団の略称は手法を参照.

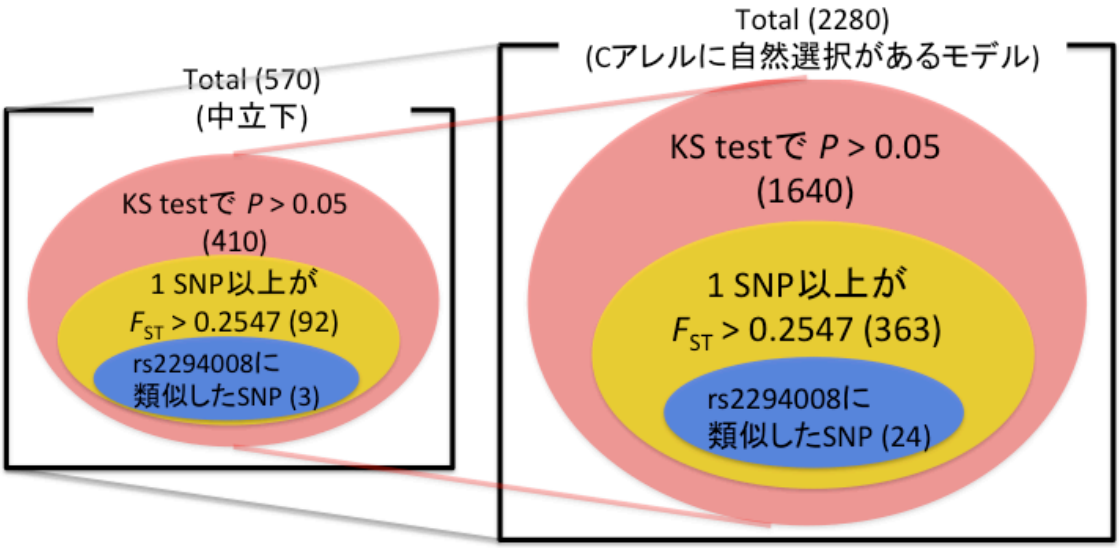

図 18. アレル頻度シミュレーションによるパラメータの組み合わせの内訳.
カッコ内の数字は、アレル頻度シミュレーションによるパラメータの組み合わせの種数を示す. 外枠のカッコは中立（左図）または自然選択（右図）がモデルに組み込まれている場合のシミュレーションの総数を示す。赤色の丸はそれぞれの条件でのシミュレーションのうち KS test を通過した総数を示している. 自然選択が働いている場合の組み合わせの数は、中立下でのパラメータの組み合わせの数（410 種）に対し、選択係数 4 種類を組み込んだ合計 1640 種となる. 黄色の丸は、それぞれの中立または自然選択下での $F_{ST} > 0.2547$ となる SNP がシミュレーションで一つ以上出たパラメータの組み合わせの総数を示す. 青色の丸は、rs2294008 と類似した条件（$F_{ST} > 0.2547$ かつ、simJPT での T アレル頻度が固定せずに 0.62 を超え、かつ、simCHB の T アレル頻度を上回る SNP が出現する）でのパラメータの組み合わせの総数を示す.

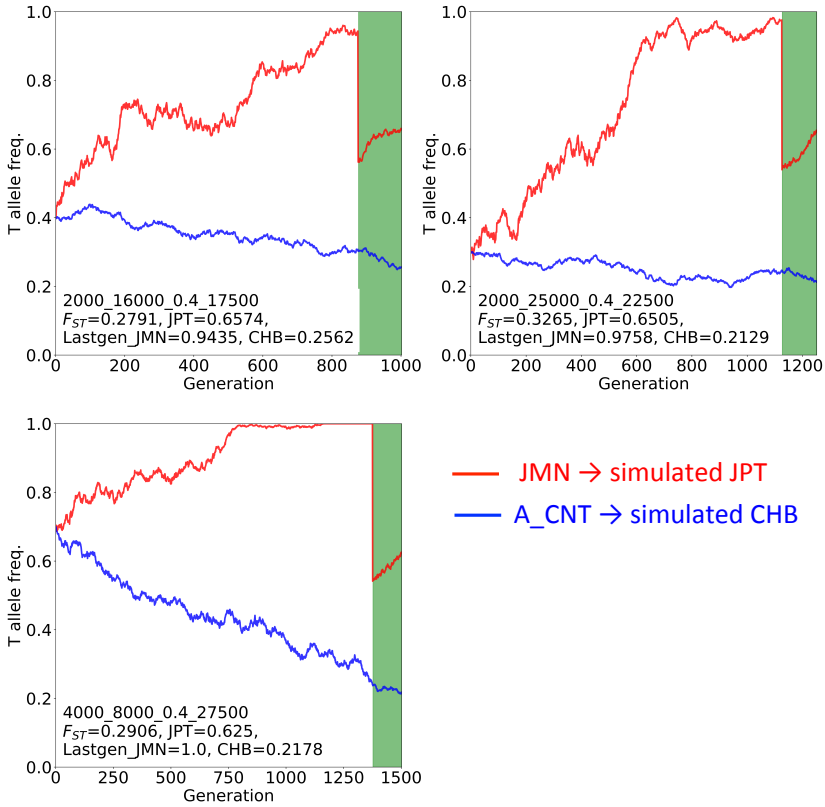

図 19. 中立下で、rs2294008 様の条件下にある 3 SNP の、T アレル頻度の軌跡. JMN 及び simJPT での T アレル頻度を赤い折れ線で示した．また、A_CNT 及び simCHB の T アレル頻度を青い折れ線で示した．薄い緑の帯は JMN と A_CNT が交雑後の期間（*t2* = 125 世代）示す．各図の左下のサブタイトルは、順に N_{JMN}、N_{A_CNT}、*r*、*t1* を示す．また、2 行目に、F_{ST} 値及び simJPT での T アレル頻度（JPT）を示した．3 行目には、JMN の交雑前の最終世代でのアレル頻度（Lastgen_JMN）、simCHB での T アレル頻度（CHB）を記載した.

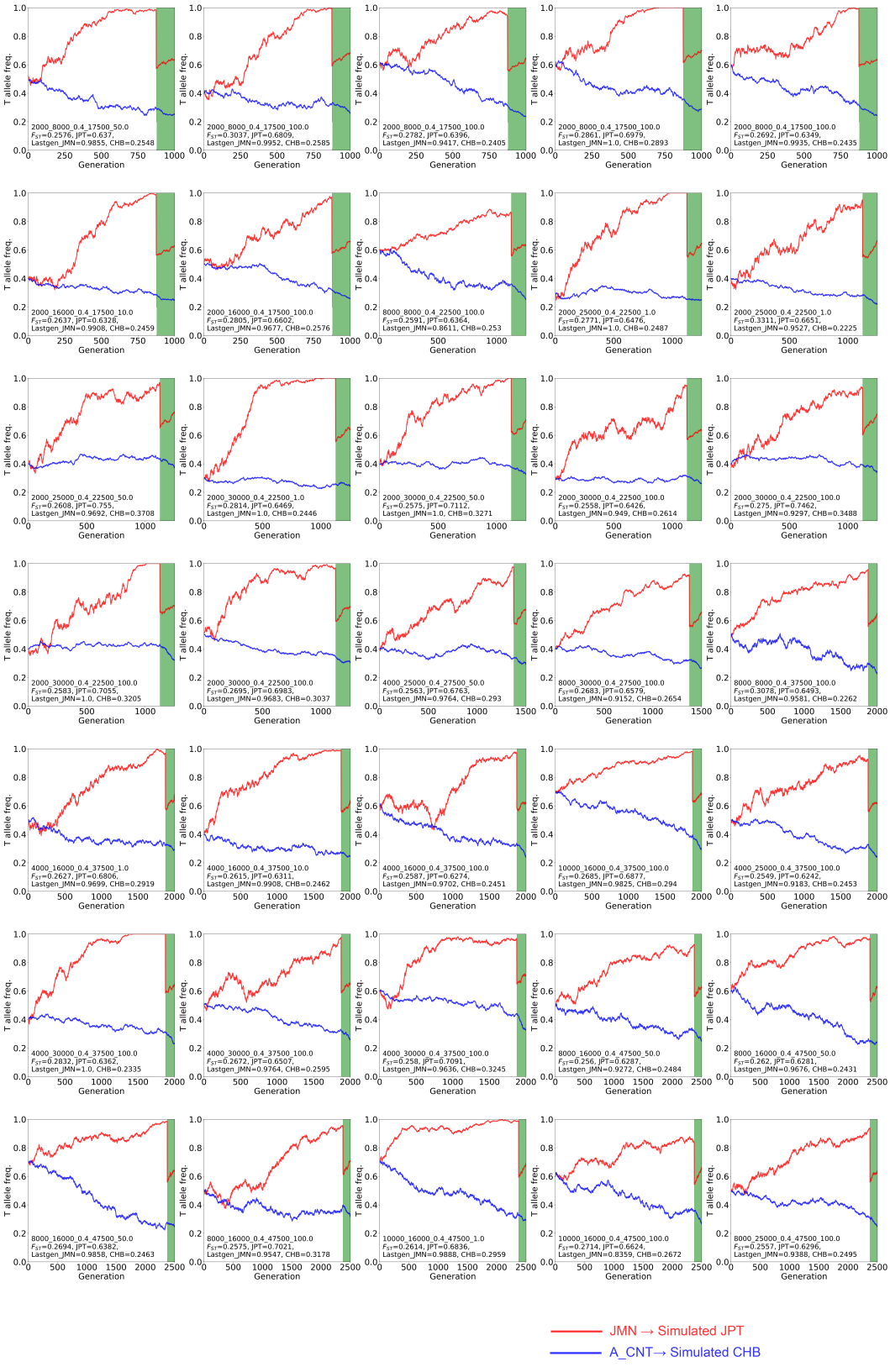
T allele freq.
Generation
2000_8000_0.4_17500_50.0
F_{ST}=0.2576, JPT=0.637,
Lastgen_JMN=0.9855, CHB=0.2548
2000_8000_0.4_17500_100.0
F_{ST}=0.3037, JPT=0.6809,
Lastgen_JMN=0.9952, CHB=0.2585
2000_8000_0.4_17500_100.0
F_{ST}=0.2782, JPT=0.6396,
Lastgen_JMN=0.9417, CHB=0.2405
2000_8000_0.4_17500_100.0
F_{ST}=0.2861, JPT=0.6979,
Lastgen_JMN=1.0, CHB=0.2893
2000_8000_0.4_17500_100.0
F_{ST}=0.2692, JPT=0.6349,
Lastgen_JMN=0.9935, CHB=0.2435
2000_16000_0.4_17500_10.0
F_{ST}=0.2637, JPT=0.6326,
Lastgen_JMN=0.9908, CHB=0.2459
2000_16000_0.4_17500_100.0
F_{ST}=0.2805, JPT=0.6602,
Lastgen_JMN=0.9677, CHB=0.2576
8000_8000_0.4_22500_100.0
F_{ST}=0.2591, JPT=0.6364,
Lastgen_JMN=0.8611, CHB=0.253
2000_25000_0.4_22500_1.0
F_{ST}=0.2771, JPT=0.6476,
Lastgen_JMN=1.0, CHB=0.2487
2000_25000_0.4_22500_1.0
F_{ST}=0.3311, JPT=0.6651,
Lastgen_JMN=0.9527, CHB=0.2225
2000_25000_0.4_22500_50.0
F_{ST}=0.2608, JPT=0.755,
Lastgen_JMN=0.9692, CHB=0.3708
2000_30000_0.4_22500_1.0
F_{ST}=0.2814, JPT=0.6469,
Lastgen_JMN=1.0, CHB=0.2446
2000_30000_0.4_22500_50.0
F_{ST}=0.2575, JPT=0.7112,
Lastgen_JMN=1.0, CHB=0.3271
2000_30000_0.4_22500_100.0
F_{ST}=0.2558, JPT=0.6426,
Lastgen_JMN=0.949, CHB=0.2614
2000_30000_0.4_22500_100.0
F_{ST}=0.275, JPT=0.7462,
Lastgen_JMN=0.9397, CHB=0.3488
2000_30000_0.4_22500_100.0
F_{ST}=0.2583, JPT=0.7055,
Lastgen_JMN=1.0, CHB=0.3205
2000_30000_0.4_22500_100.0
F_{ST}=0.2695, JPT=0.6983,
Lastgen_JMN=0.9683, CHB=0.3037
4000_25000_0.4_27500_50.0
F_{ST}=0.2563, JPT=0.6763,
Lastgen_JMN=0.9764, CHB=0.293
8000_30000_0.4_27500_100.0
F_{ST}=0.2683, JPT=0.6579,
Lastgen_JMN=0.9152, CHB=0.2654
8000_8000_0.4_37500_100.0
F_{ST}=0.3078, JPT=0.6493,
Lastgen_JMN=0.9581, CHB=0.2262
4000_16000_0.4_37500_1.0
F_{ST}=0.2627, JPT=0.6806,
Lastgen_JMN=0.9699, CHB=0.2919
4000_16000_0.4_37500_10.0
F_{ST}=0.2615, JPT=0.6311,
Lastgen_JMN=0.9908, CHB=0.2462
4000_16000_0.4_37500_100.0
F_{ST}=0.2587, JPT=0.6274,
Lastgen_JMN=0.9702, CHB=0.2451
10000_16000_0.4_37500_100.0
F_{ST}=0.2685, JPT=0.6877,
Lastgen_JMN=0.9825, CHB=0.294
4000_25000_0.4_37500_100.0
F_{ST}=0.2549, JPT=0.6242,
Lastgen_JMN=0.9183, CHB=0.2453
4000_30000_0.4_37500_100.0
F_{ST}=0.2832, JPT=0.6362,
Lastgen_JMN=1.0, CHB=0.2335
4000_30000_0.4_37500_100.0
F_{ST}=0.2672, JPT=0.6507,
Lastgen_JMN=0.9764, CHB=0.2595
4000_30000_0.4_37500_100.0
F_{ST}=0.258, JPT=0.7091,
Lastgen_JMN=0.9636, CHB=0.3245
8000_16000_0.4_47500_50.0
F_{ST}=0.256, JPT=0.6287,
Lastgen_JMN=0.9272, CHB=0.2484
8000_16000_0.4_47500_50.0
F_{ST}=0.262, JPT=0.6281,
Lastgen_JMN=0.9676, CHB=0.2431
8000_16000_0.4_47500_50.0
F_{ST}=0.2694, JPT=0.6382,
Lastgen_JMN=0.9858, CHB=0.2463
8000_16000_0.4_47500_100.0
F_{ST}=0.2575, JPT=0.7021,
Lastgen_JMN=0.9547, CHB=0.3178
10000_16000_0.4_47500_1.0
F_{ST}=0.2614, JPT=0.6836,
Lastgen_JMN=0.9888, CHB=0.2959
10000_16000_0.4_47500_100.0
F_{ST}=0.2714, JPT=0.6624,
Lastgen_JMN=0.8359, CHB=0.2672
8000_25000_0.4_47500_100.0
F_{ST}=0.2557, JPT=0.6296,
Lastgen_JMN=0.9388, CHB=0.2495
JMN → Simulated JPT
A_CNT→ Simulated CHB

図 20．C アレルに対する自然選択下で、rs2294008 様の条件下にある 35 SNP の、T アレル頻度の軌跡.

JMN 及び simJPT での T アレル頻度を赤い折れ線で示した．また、A_CNT 及び simCHB の T アレル頻度を青い折れ線で示した．薄い緑の帯は JMN と A_CNT が交雑後の期間（$t2$ = 125 世代）を示す. 各図の左下のサブタイトルは、順に N_{JMN}、$N_{\mathrm{A\_CNT}}$、r、$t1$、$2Ns$ を示す．2 行目以降の表記は、図 19 に従う.

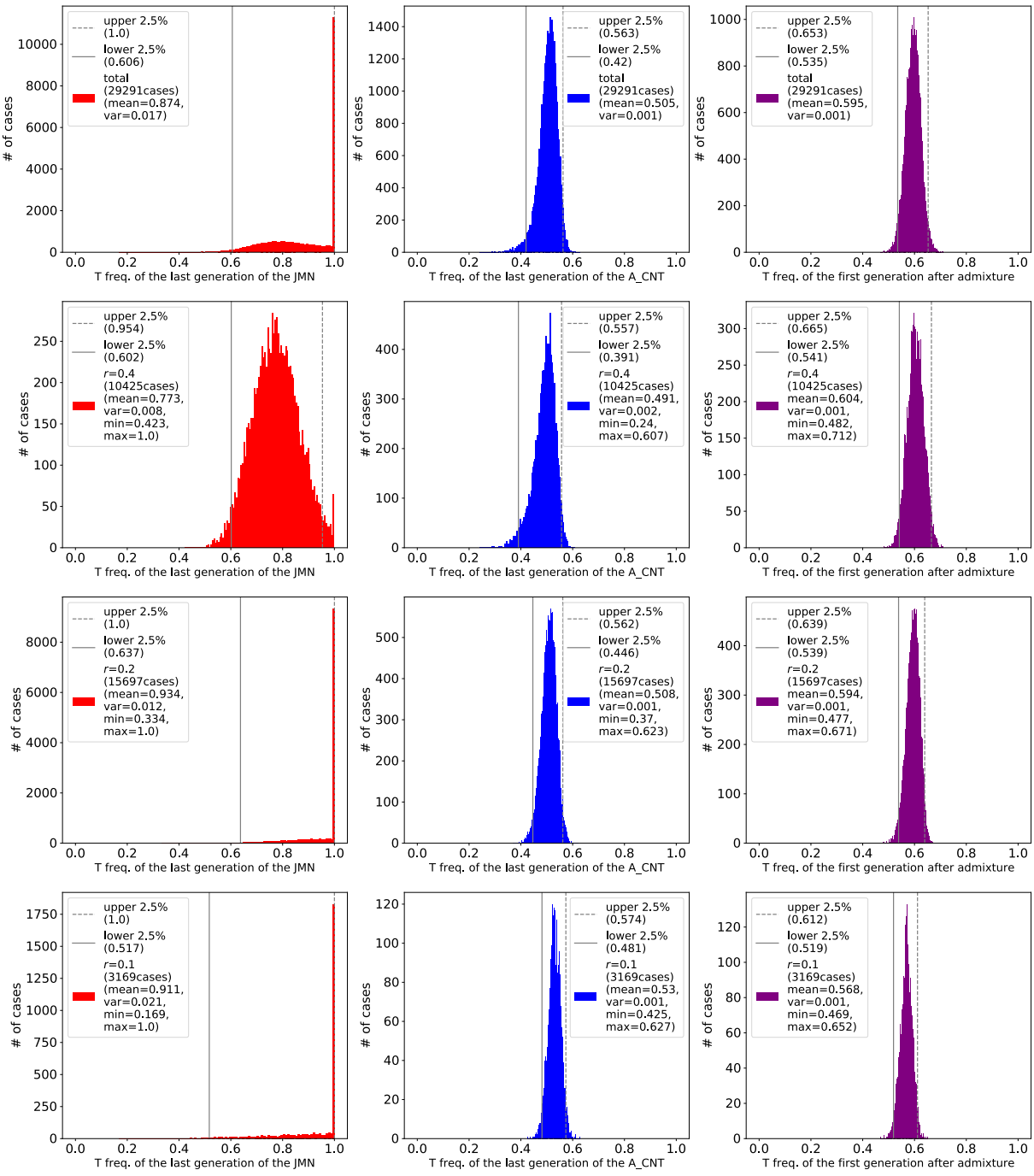

図 21．中立下で、rs2294008 様の SNP の T アレルの頻度分布．
各グラフは、rs2294008 に似た条件下にある T アレルの頻度分布を示す．左列は JMN の交雑直前の最終世代、中央列は A_CNT の最終世代、右列は交雑直後の JPT での頻度分布をそれぞれ示す．A_CNT は韓国集団を想定している．また、1 段目は r の値に関係なく頻度や F_{ST} が rs2294008 と似た値を示す条件を

満たしたものの全ての分布を示し、2 段目以降はそれぞれ r=0.4、 r=0.2、 r=0.1 での分布を示す．図中の灰色の点線及び実線は、それぞれ分布の上位 2.5%及び下位 2.5%を示す．

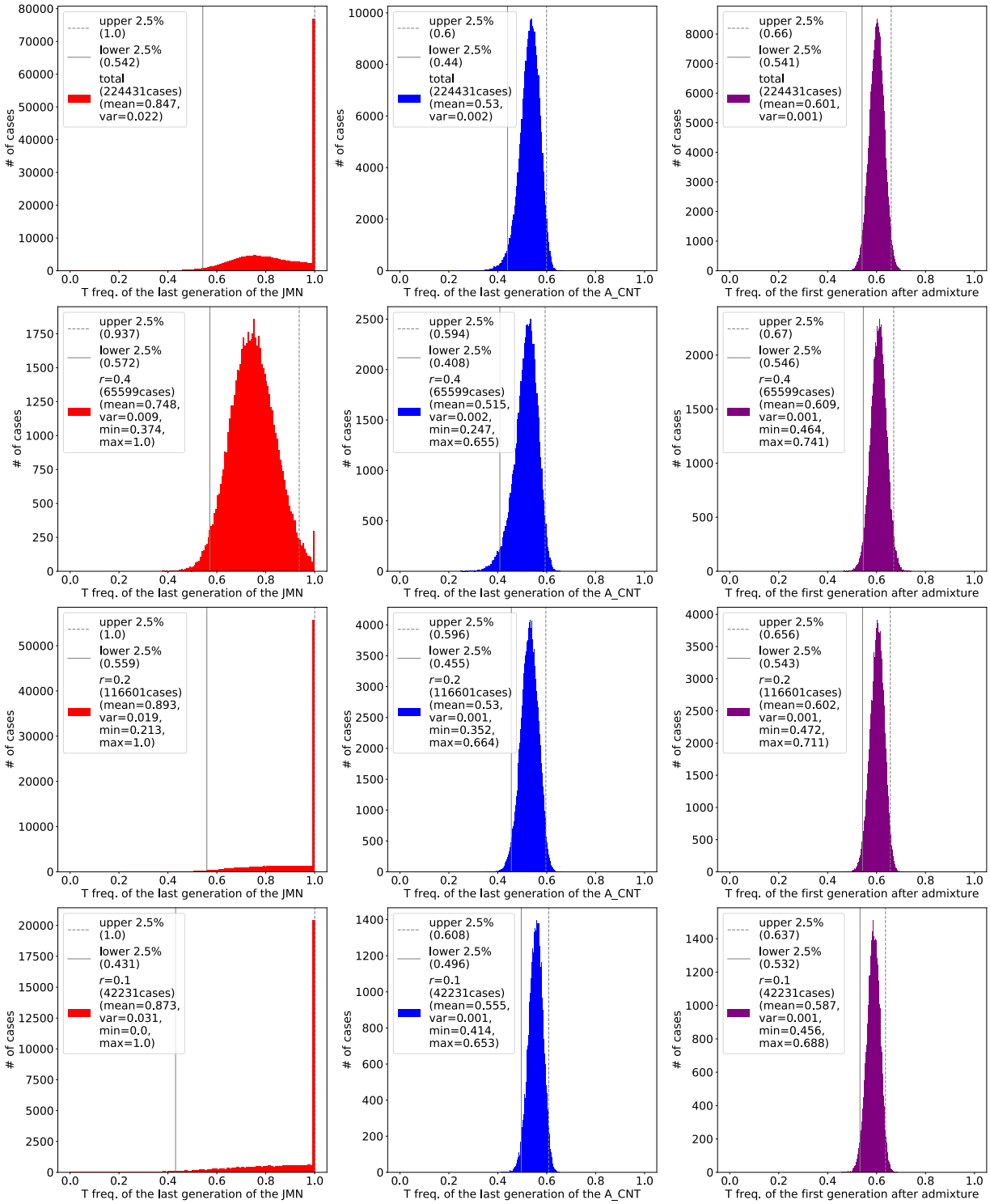

図 22. C アレルに対する自然選択下で、rs2294008 様の SNP の T アレルの頻度分布.

各グラフは、rs2294008 に似た条件下にある T アレルの頻度分布を示す. 左列は JMN の交雑直前の最終世代、中央列は A_CNT の最終世代、右列は交雑直後

の JPT での T アレルの頻度分布をそれぞれ示す．A_CNT は韓国集団を想定している．また、1 段目は r の値に関係なく頻度や F_{ST} が rs2294008 と似た値を示す条件を満たしたものの全ての分布を示し、2 段目以降はそれぞれ r=0.4、r=0.2、 r=0.1 での分布を示す．図中の灰色の点線及び実線は、それぞれ分布の上位 2.5%及び下位 2.5%を示す．

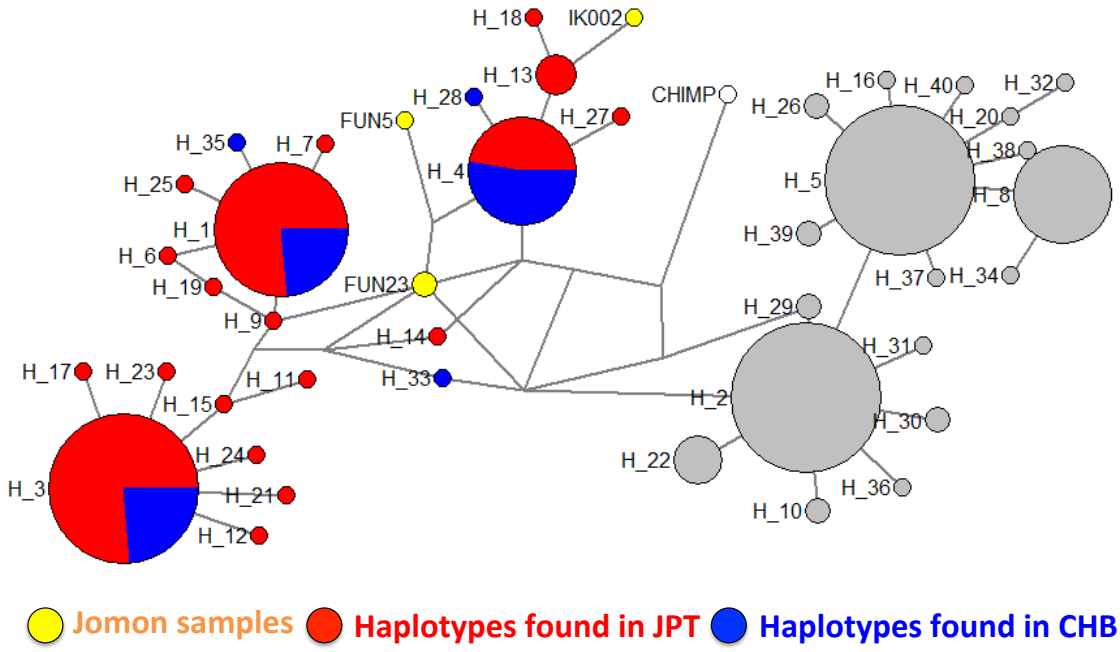

図 23. 縄文系統の 3 サンプル（IK002、FUN5、FUN23）及び、現代の JPT・CHB、チンパンジーの相同配列を用いたハプロタイプネットワーク解析.
C アレルを持つハプロタイプは灰色で示した. 縄文系統の 3 サンプル（IK002、FUN5、FUN23）は黄色の円で表し、現代の JPT で観察されるハプロタイプは赤色、現代の CHB で観察されるハプロタイプは青色で示した. また、チンパンジーのハプロタイプを白色で示した. 図中の円の大きさはハプロタイプの本数を反映している. 図中の H で始まるハプロタイプの名称は、古代ゲノムを加えた場合のハプロタイプの名称であり、現代の人類で定義したハプロタイプ名称との対応は、表 7 に従う. なお、この図では、各ノード間の枝長はハプロタイプ同士の塩基置換数を反映していない.

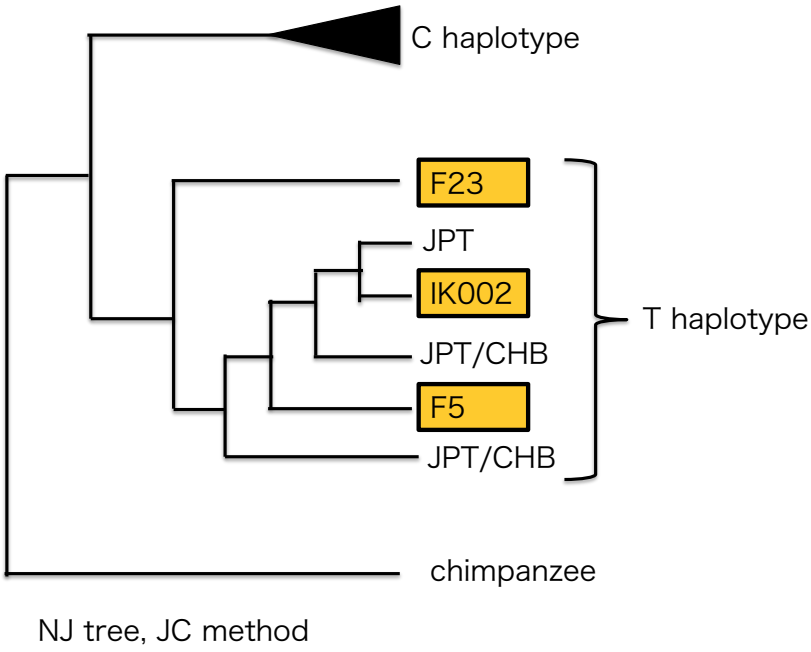

図 24. 縄文系統の 3 サンプル及び、現代の JPT・CHB、チンパンジーの相同配列を用いたハプロタイプの系統解析.
ハプロタイプ間の距離は、Jukes-Cantor モデルに従って補正を行い算出した。この距離に基づいて Neighbor-Joining 法によって系統樹を作成し、縄文系統の３サンプルと現生集団のトポロジーを決定した．なお、この図では、各ノード間の枝長はハプロタイプ同士の塩基置換数を反映していない.

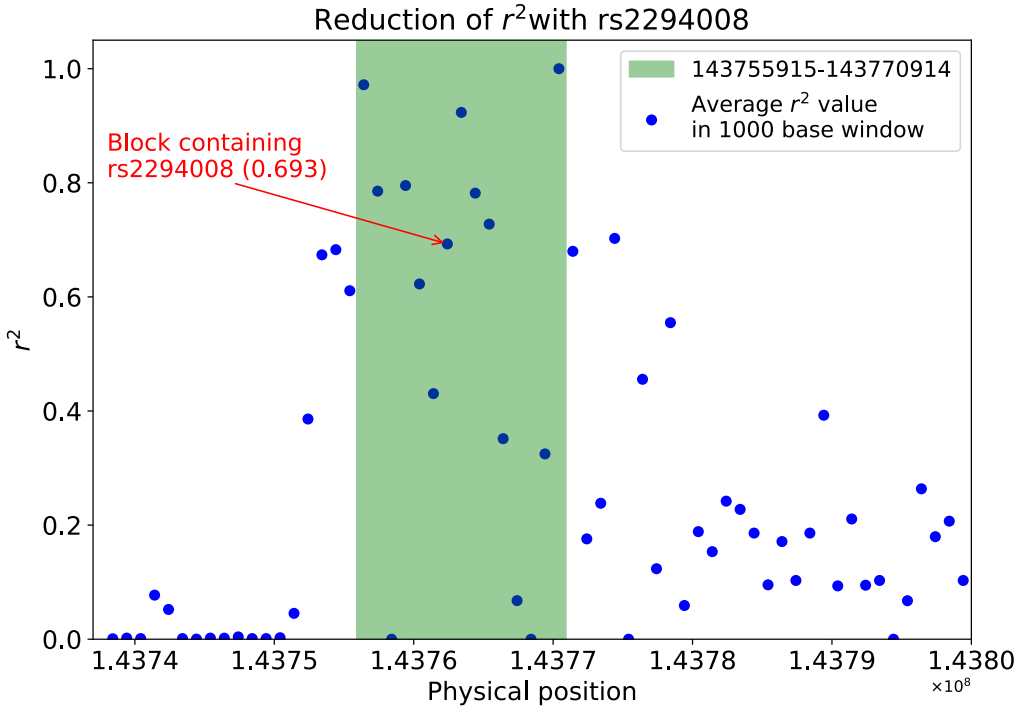

図 25. rs2294008 とその近傍 60kb の SNP 間の連鎖（r^2）の強さの減衰. rs2294008 と近傍の SNP との r^2 を 1 kb の領域毎にその平均値をとった. 各領域の左端の番地に従って青い点で示した. rs2294008 を含むブロックの r^2 の値は、図中に赤文字で示した. また、JPT で 2D SFS に用いた領域（$r^2 > 0.75$）は、図中の緑色の帯で示した.

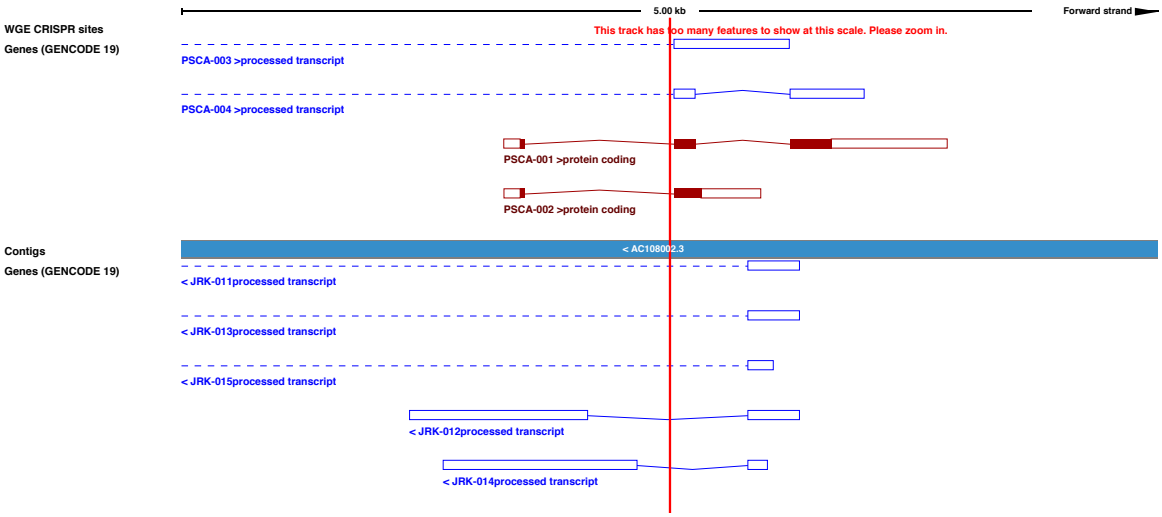

図 26. r2976391 のゲノム上の位置と、*PSCA* 遺伝子、*JRK* 遺伝子の位置関係. rs2976391 の位置は、図中の赤い縦線で示した. また、この図は Ensembl（http://grch37.ensembl.org/Homo_sapiens/Variation/Context?db=core;r=8:143762224-143763224;v=rs2976391;vdb=variation;vf=477661232、[8]）よりダウンロードした.

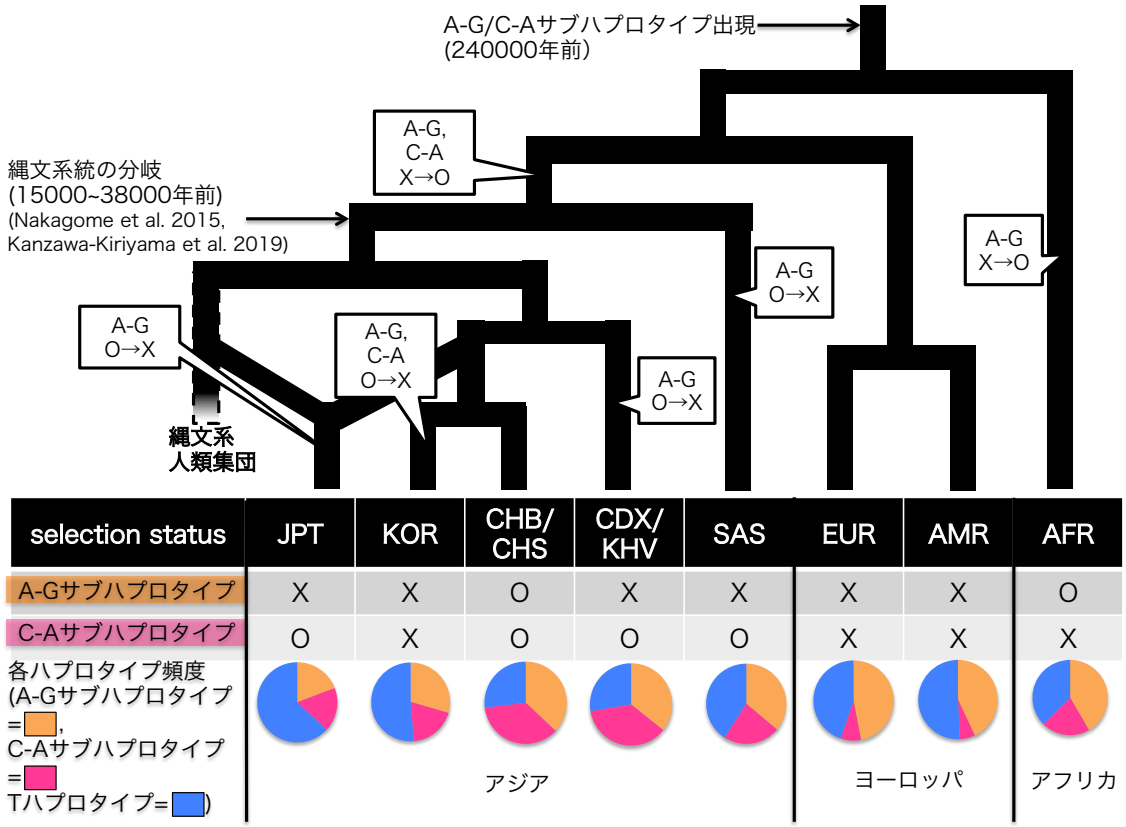

図 27. A-G 及び C-A サブハプロタイプに対する、自然選択の状態の推移.
表は現代の各集団に働いている、各サブハプロタイプに対する自然選択の状態を示す. 表中の自然選択の状態の表記は、表 3~4 に従う. 下の図は、表 4 (a)・(b) の、A-G 及び C-A サブハプロタイプ、T ハプロタイプの頻度を示す. また、上部の系統樹は各人類集団の分岐順と集団間の関係を示す. 東アジア集団の分岐順と系統関係にはまだ議論がある. そこで本研究では、まず 1KGP[44]の各集団間のペアワイズ F_{ST} 値、Kanzawa-Kiriyama *et al.*[43]の系統樹を元に、KOR を除いた subpoulation 間の系統関係を推定した. 更に、韓国人集団と最も近縁な集団を PCA[73]から推定し、KOR の系統的位置を決定した. また、縄文系統の交雑相手である、渡来系集団の子孫が具体的に現代のどの集団であるかについても、まだ議論がある. そこで、弥生時代の開始時期 (2500~3000 年前) [65,79] を渡来系弥生人集団と縄文系統の集団との交雑開始時期と仮定し、この時期を Wang *et al.* [72]の推定した韓国人集団-CHB 間の分岐年代 (47 世代前 (940~1410 年前)) と比較した. この両者の前後関係を考慮し、渡来系集団を CHB・CHS・KOR の共通祖先として系統樹を作成した. 図中の吹き出しは、同図の各サブハプロタイプの自然選択の状態の変化を示す. この自然選択の状

態変化は、最節約的に推定を行った．ただし、東アジアの人類集団の共通祖先での自然選択の状態の変化の時期については、以下の根拠に基づいて推定した．縄文系人類集団は、他の東アジア集団から、約 15,000 年～38,000 年[39,42-43]に分岐したが、この時期と、JPT 及び CHB での A-G 及び C-A サブハプロタイプに対する自然選択の開始時期の下限（約 27,000 年前及び 30,000 年前）は同時期に当たる．従って、縄文系人類集団が東アジア集団から分岐する時点では、東アジア集団の共通祖先には、A-G 及び C-A サブハプロタイプに対して自然選択が働いていたはずであり、自然選択が働いていない現代の集団（JPT、KOR、CDX、KHV）では、各集団に分岐した後に、いずれかのタイミングで自然選択がリラックス、または停止したと考えられる．

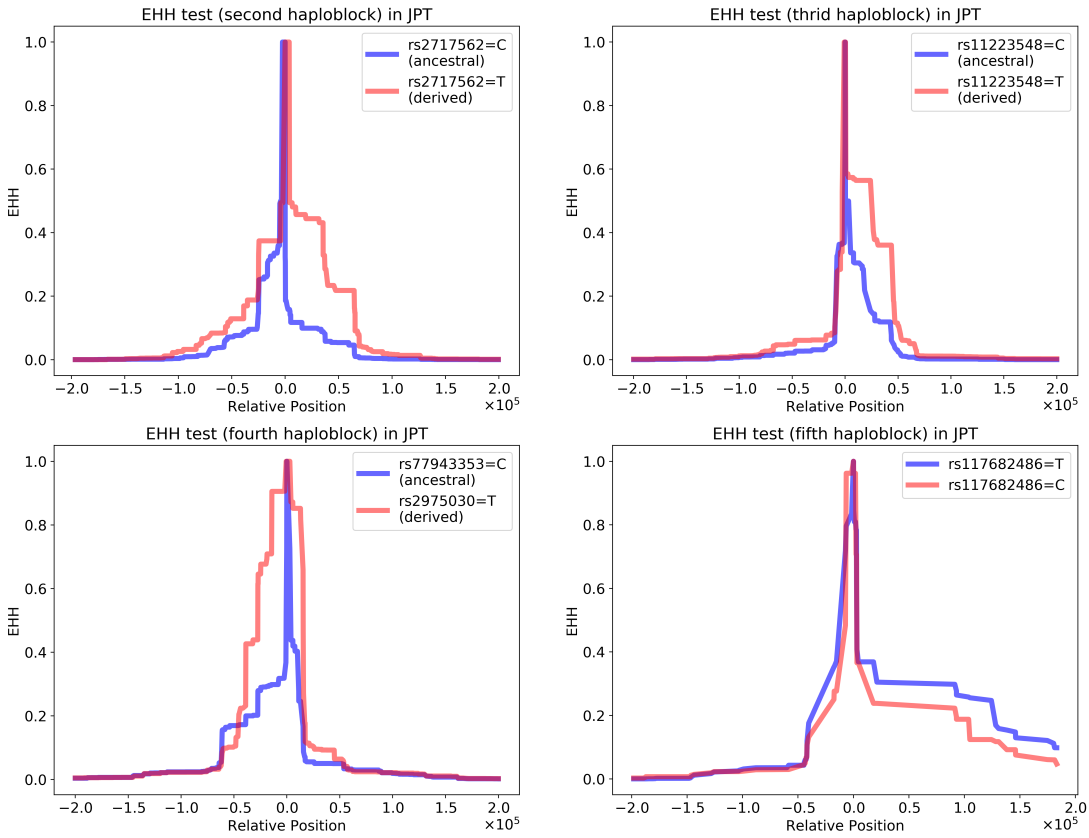

図 28. F_{ST} が高く、JPT で自然選択が働いている可能性のある SNP に対する EHH 解析.

rs2294008 及びこの SNP を含む LD ブロック外の領域に位置し、かつ、高い F_{ST} 値を持つ SNP をコアとした EHH 値の減衰の程度を、派生型と祖先型のアレルで比較した. ただし、rs117682486 では、T/C のどちらが祖先型であるかの記述が 1KGP[44]にないため、派生型と祖先型の区別は行なっていない.

表

表 1. 1KGP の東アジアの各集団の T アレル頻度と rs2294008 の F_{ST} 値.

	JPT	CHB	CHS	CDX	KHV
Tアレル頻度	0.6298	0.2476	0.2714	0.2742	0.2778
rs2294008のF_{ST}値（JPTvs）	–	0.2547	0.2259	0.2224	0.2184

*各略称は、1KGP[44]の表記に従った.

JPT…東京在住の日本人、CHB…中国の漢族、CHS…中国南部の漢族、CDX…中国のダイ族、KHV…ベトナム人を示す.

表 2. ゲノムワイド SNP の JPT と CHB 間の F_{ST} 値（1 位~50 位）とその位置情報.

順位	染色体番号	染色体上の位置	F_{ST} 値
1	8	143763622	0.266892
2	8	143776668	0.260792
3	8	143774193	0.260792
4	8	143770135	0.260792
5	8	143765699	0.260792
6	8	143765414	0.260792
7	8	143765326	0.260792
8	8	143764937	0.260792
9	8	143764879	0.260792
10	8	143764101	0.260792
11	8	143764001	0.260792
12	8	143763958	0.260792
13	8	143763757	0.260792
14	8	143763750	0.260792
15	8	143763690	0.260792
16	8	143763618	0.260792
17	8	143763490	0.260792
18	8	143763109	0.260792
19	8	143763083	0.260792
20	8	143763043	0.260792
21	8	143762932	0.260792
22	8	143762430	0.260792
23	8	143760179	0.259648
24	8	143764613	0.257827
25	8	143766059	0.256077
26	8	143761144	0.256077
27	8	143760256	0.256077
28	8	143771714	0.25537

29	8	143771712	0.25537
30	8	143763547	0.254748
31	8	143762135	0.254748
32	8	143761931	0.254748
33	8	143760444	0.250914
34	8	143759364	0.250914
35	8	143755720	0.250914
36	8	143760421	0.249349
37	8	143759137	0.249349
38	8	143757763	0.249349
39	8	143757708	0.249349
40	8	143757409	0.249349
41	8	143757286	0.249349
42	8	143757037	0.249349
43	8	143756919	0.249349
44	8	143756895	0.249349
45	8	143756892	0.249349
46	8	143756530	0.249349
47	8	143756218	0.249349
48	8	143755426	0.249349
49	8	143754728	0.249349
50	8	143752994	0.249349

表 3. JPT と CHB で、rs2294008 の T 及び C アレルをターゲットと仮定した 2D SFS.

対象集団	JPT		CHB	
使用した領域	143755915-143770914	143755915-143770914	143755876-143771875	143755876-143771875
ターゲットアレル	C	T	C	T
サンプルサイズ	208 (C = 77, T = 131)		206 (C = 155, T = 51)	
ターゲットアレル頻度	0.370	0.630	0.752	0.248
segregating site数	91	91	88	88
F_c	0.167 (0.834)	0.833 (> 0.999)	0.352×10^{-1} $(0.223 \times 10^{-2})^{**}$	0.869 (> 0.999)
G_{c0}	9.60 (0.693)	31.84 (0.975)	1.84 $(0.167 \times 10^{-2})^{**}$	25.13 (> 0.999)
L_{c0}	0.178×10^{-1} (0.708)	0.259 (> 0.999)	0.565×10^{-2} $(> 0.167 \times 10^{-2})^{**}$	0.130 (> 0.999)
G^*_{c0}	22.50	46.23	5.00	30.69
$\Upsilon^*(10)$	0.500	0.700	0.000	0.962
i^*_{max}	0	29	75	21
i_{max}	40	130	8	50
自然選択の有無§	X	X	O	X
selective sweepタイプ¶	-	-	soft	-
t_D (年)	166234	1425954	28226	1970588
t_DのSD (year)	99080	360166	9300	427512

§F_c, G_{c0} 及び L_{c0} のカッコ内の数値は Q-value を示す．Q-value < 0.01 の場合は **、0.01 ≤ Q-value < 0.05 の場合は*で示した.

¶selective sweep タイプは、Satta *et al.* [30]の基準及び、i^*_{max} の値によって soft sweep か hard sweep かを決定した.「-」は特に自然選択のシグナルが認められなかったことを示す．また、解析対象の集団が metapopulation の場合は、selection status の異なる subpopulation が混じることによって heterogeneity が生じるため、それぞれの sweep によって生じるシグナルが正しく統計量に反

映されない可能性があり、特に hard sweep か soft sweep かは判定しないものとした.

表 4. A-G または C-A サブハプロタイプをターゲットと仮定した 2D SFS.

（a）JPT と CHB

対象集団	JPT		CHB	
使用した領域	143755915-143770914	143755915-143770914	143755876-143771875	143755876-143771875
ターゲットアレル[#]	A-G	C-A	A-G	C-A
サンプルサイズ[#]	208 (A-G = 40, others = 168)	208 (C-A = 37, others = 171)	206 (A-G = 74, others = 132)	206 (C-A = 79, others = 127)
ターゲットアレル頻度	0.192	0.178	0.359	0.383
segregating site数	89	89	86	86
F_c	0.483×10^{-1} (0.296)	0.112×10^{-1} (0.354×10^{-1})*	0.156×10^{-1} (0.228×10^{-1})*	0.196×10^{-1} (0.347×10^{-1})*
G_{c0}	2.17 (0.287)	1.50 (0.180)	1.78 (0.573×10^{-1})	1.90 (0.548×10^{-1})
L_{c0}	0.241×10^{-2} (0.273)	0.555×10^{-3} (0.456×10^{-1})*	0.258×10^{-2} (0.228×10^{-1})*	0.307×10^{-2} (0.204×10^{-1})*
G^*_{c0}	8.00	2.00	8.00	4.00
$\Upsilon^*(10)$	0.000	0.000	0.000	0.000
i^*_{max}	0	0	2	1
i_{max}	8	2	8	5
自然選択の有無[§]	X	O	O	O
selective sweepタイプ[¶]	-	hard	hard	hard
t_D (年)	43333	10811	27027	30063
t_DのSD (year)	27689	8058	14333	12358

表中の、自然選択の有無の表記及び、selective sweep タイプの表記は表 3 に従う.

[#]C-A、A-G はそれぞれ C-A サブハプロタイプ、A-G サブハプロタイプを示す.

（b）EAS の subpopulation、KOR 及び non JPT/CHB EAS

対象集団	CDX		KHV		CHS		KOR		non JPT/CHB EAS	
使用した領域	143755915-143771914	143755915-143771914	143755915-143771914	143755915-143771914	143755915-143771914	143755915-143771914	143755995-143770914	143755995-143770914	143755876-143771875	143755876-143771875
ターゲットアレル[#]	A-G	C-A	A-G	C-A	A-G	C-A	A-G	C-A	A-G	C-A
サンプルサイズ[#]	186 (A-G = 64, others = 122)	186 (C-A = 71, others = 115)	198 (A-G = 73, others = 125)	198 (C-A = 70, others = 128)	210 (A-G = 80, others = 130)	210 (C-A = 71, others = 139)	302 (A-G = 89, others = 213)	302 (C-A = 58, others = 244)	594 (A-G = 217, others = 377)	594 (C-A = 212, others = 382)
ターゲットアレル頻度	0.344	0.382	0.369	0.354	0.381	0.338	0.295	0.192	0.365	0.357
segregating site数	80	80	87	87	87	87	64	64	112	112
F_c	0.899×10^{-1} (0.547)	0.146×10^{-1} (0.166×10^{-1})*	0.302×10^{-1} (0.863×10^{-1})	0.799×10^{-2} (0.699×10^{-3})**	0.230×10^{-1} (0.519×10^{-1})	0.939×10^{-2} (0.921×10^{-2})**	0.484×10^{-1} (0.285)	0.101 (0.385)	0.258×10^{-1} (0.650×10^{-1})	0.114×10^{-1} (0.675×10^{-2})**
G_{c0}	3.43 (0.313)	2.40 (0.171)	2.73 (0.186)	1.50 (0.389×10^{-1})*	2.44 (0.137)	2.00 (0.115)	17.00 (0.883)	2.33 (0.457)	3.44 (0.138)	3.10 (0.116)
L_{c0}	0.418×10^{-2} (0.399)	0.209×10^{-2} (0.142×10^{-2})**	0.508×10^{-2} (0.112)	0.152×10^{-2} (0.699×10^{-3})**	0.349×10^{-2} (0.415×10^{-1})*	0.159×10^{-2} (0.230×10^{-2})**	0.244×10^{-2} (0.901×10^{-1})	0.100×10^{-2} (0.874×10^{-1})	0.338×10^{-2} (0.231×10^{-1})*	0.169×10^{-2} ($<0.223 \times 10^{-2}$)**
G^*_{c0}	6.67	3.33	4.80	2.00	3.17	6.00	17.00	5.00	6.50	6.25
$\Upsilon^*(10)$	0.333	0.000	0.200	0.000	0.000	0.000	1.000	0.000	0.125	0.250
i^*_{max}	1	2	1	2	5	1	0	0	19	4
i_{max}	10	4	10	2	5	6	17	5	24	11
自然選択の有無[§]	X	O	X	O	△	O	X	X	△	O
combined P	-	-	-	-	O (0.04, x2=7.826, df=2.44)	-	-	-	O (0.03, x2=8.160, df=2.45)	-
selective sweepタイプ[¶]	-	hard	-	hard	-	hard	-	-		
t_D (年)	46875	21127	51370	16071	34375	17606	25605	16178	35714	18278
t_DのSD (year)	25615	10563	22062	6916	13073	11135	25605	12009	15670	8276

表中の、自然選択の有無の表記及び、selective sweep タイプの表記は表 3 に従う．また、ターゲットアレルの表記は表 4（a）に従う．

（c） metapopulation

対象集団	SAS		EUR		AMR		AFR	
使用した領域	143755864-143777863	143755864-143777863	143756771-143778770	143756771-143778770	143752915-143780914	143752915-143780914	143756253-143774252	143756253-143774252
ターゲットアレル#	A-G	C-A	A-G	C-A	A-G	C-A	A-G	C-A
サンプルサイズ#	978 (A-G = 353, others = 625)	978 (C-A = 225, others = 753)	1006 (A-G = 473, others = 533)	1006 (C-A = 82, others = 924)	694 (A-G = 299, others = 395)	694 (C-A = 43, others = 651)	1322 (A-G = 550, others = 772)	1322 (C-A = 276, others = 1046)
ターゲットアレル頻度	0.361	0.230	0.470	0.082	0.431	0.062	0.416	0.209
segregating site数	201	201	184	184	253	253	262	262
F_c	0.167 (0.452)	0.352×10^{-1} (0.136)	0.321 (0.590)	0.140×10^{-1} (0.252)	0.204 (0.521)	0.372×10^{-1} (0.462)	0.139×10^{-1} (<0.273 x 10-2)**	0.883×10^{-1} (0.320)
G_{c0}	4.34 (0.188)	1.75 (0.159×10^{-1})*	8.68 (0.265)	1.75 (0.211)	4.62 (0.159)	1.90 (0.524)	2.28 (<0.273×10^{-2})**	10.71 (0.253)
L_{c0}	0.545×10^{-2} (0.906×10^{-1})	0.107×10^{-2} (0.528×10^{-2})**	0.937×10^{-2} (0.531×10^{-1})	0.200×10^{-3} (0.501×10^{-1})	0.654×10^{-2} (0.321×10^{-1})*	0.700×10^{-3} (0.649)	0.251 x 10-2 (<0.273×10^{-2})**	0.430×10^{-2} (0.450×10^{-1})*
G^*_{c0}	9.56	3.50	15.20	2.50	9.30	3.25	4.28	24.57
$Y^*(10)$	0.250	0.000	0.450	0.000	0.240	0.000	0.060	0.710
i^*_{max}	38	3	42	1	108	2	63	5
i_{max}	49	6	472	3	35	5	10	57
自然選択の有無§	X	O	X	X	△	X	O	△
combined P	-	-	-	-	X (0.07, x2=6.302, df=2.54)	-	-	X (0.08, x2=6.243, df=2.64)
t_D (年)	45841	14141	61695	7761	43000	31561	21212	73269
t_DのSD (year)	17576	4179	22423	4294	13682	12319	4403	33280

表中の、自然選択の有無の表記及び、selective sweep タイプの表記は表 3 に従う．また、ターゲットアレルの表記は表 4（a）に従う．

（d）AFR の subpopulation

対象集団	GWD		YRI		LWK		ACB	
使用した領域	143756352-143774351	143756352-143774351	143756108-143771107	143756108-143771107	143756072-143771071	143756072-143771071	143756072-143771071	143756072-143771071
ターゲットアレル#	A-G	C-A	A-G	C-A	A-G	C-A	A-G	C-A
サンプルサイズ#	226 (A-G =100, others = 126)	226 (C-A = 48, others = 178)	216 (A-G = 116, others=100)	216 (C-A = 50, others = 166)	198 (A-G = 76, others = 122)	198 (A-G = 37, others = 161)	192 (A-G = 73, others = 119)	192 (A-G = 36, others = 156)
ターゲットアレル頻度	0.442	0.212	0.537	0.231	0.384	0.187	0.380	0.188
segregating site数	149	149	136	136	140	140	124	124
F_c	0.187×10^{-1} ($<0.182 \times 10^{-2}$)**	0.111 (0.591)	0.181×10^{-1} ($<0.228 \times 10^{-2}$)**	0.723×10^{-1} (0.243)	0.364×10^{-1} (0.570×10^{-2})**	0.774×10^{-1} (0.341)	0.143 x 10-1 ($<0.220 \times 10^{-2}$)**	0.719×10^{-1} (0.4424)
G_{c0}	1.42 ($<0.182 \times 10^{-2}$)**	6.57 (0.550)	1.31 ($<0.228 \times 10^{-2}$)**	4.43 (0.230)	2.44 (0.570×10^{-2})**	3.83 (0.412)	1.50 ($<0.220 \times 10^{-2}$)**	4.00 (0.306)
L_{c0}	0.236×10^{-1} ($<0.182 \times 10^{-2}$)**	0.640×10^{-2} (0.154)	0.261×10^{-2} ($<0.228 \times 10^{-2}$)**	0.476×10^{-2} (0.465×10^{-1})*	0.381 x 10-2 ($<0.228 \times 10^{-2}$)**	0.399×10^{-2} (0.929×10^{-1})	0.224 x 10-2 ($<0.220 \times 10^{-2}$)**	0.373×10^{-2} (0.944×10^{-1})
G^*_{c0}	2.67	10.75	2.00	5.00	3.60	6.67	2.33	4.00
$\gamma^*(10)$	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000
I^*_{max}	1	0	18	1	11	0	8	1
i_{max}	4	19	2	8	7	8	3	7
自然選択の有無§	O	X	O	△	O	X	O	X
combined P	-	-	-	X (0.07, x2=7.113, df=2.86)	-	-	-	-
selective sweepタイプ¶	hard	-	hard	-	hard	-	hard	-
t_D (年)	18889	106481	19540	82667	38596	82883	21918	74074
t_DのSD (year)	6383	54681	5747	35075	16644	42486	8567	36289

対象集団	ESN		ASW		MSL	
使用した領域	143757203-143774202	143757203-143774202	143756015-143771014	143756015-143771014	143752994-143781058	143752994-143781058
ターゲットアレル#	A-G	C-A	A-G	C-A	A-G	C-A
サンプルサイズ#	198 (A-G =82, others = 118)	198 (A-G =57, others = 141)	122 (A-G =39, others = 83)	122 (C-A =22, others = 100)	170 (A-G =64, others = 106)	170 (C-A =26, others = 144)
ターゲットアレル頻度	0.414	0.288	0.320	0.180	0.376	0.153
segregating site数	127	127	131	131	242	242
F_c	0.173×10^{-1} (<0.205 x 10-2)**	0.133 (0.293)	0.301×10^{-1} (0.963×10^{-2})**	0.667×10^{-1} (0.246)	0.367×10^{-1} (0.167×10^{-2})**	0.459×10^{-1} (0.247)
G_{c0}	1.50 (<0.205×10^{-2})**	3.75 (0.152)	1.14 (0.206×10^{-2})**	2.40 (0.293)	1.50 (0.251×10^{-2})**	3.18 (0.280)
L_{c0}	0.333×10^{-2} (<0.205×10^{-2})**	0.555×10^{-2} (0.179×10^{-1})*	0.229×10^{-2} (0.206×10^{-2})**	0.344×10^{-2} (0.140)	0.391×10^{-2} (0.251×10^{-2})**	0.507×10^{-2} (0.278)
G^*_{c0}	2.20	6.50	2.00	3.33	3.25	4.00
$\gamma^*(10)$	0.000	0.250	0.000	0.000	0.000	0.000
i^*_{max}	15	4	6	0	4	1
i_{max}	3	11	2	4	5	5
自然選択の有無§	O	△	O	X	O	X
combined P	-	O (0.04, x2=8.686, df=3.02)	-	-	-	-
selective sweepタイプ¶	hard	-	hard	-	hard	-
t_D (年)	25825	61920	27350	72727	30064	95931
t_DのSD (year)	8116	30193	10811	36364	8838	10899

表中の、自然選択の有無の表記及び、selective sweep タイプの表記は表 3 に従う．また、ターゲットアレルの表記は表 4（a）に従う．

ESN の C-A サブハプロタイプは、Q-value でも combined P でも弱い自然選択のシグナルが検出されているが、これは L_{c0} の値だけが非常に小さかったためだと考えられる．L_{c0} は他の要約統計量と比較して長くシグナルが出るため、過去に働いていた自然選択のシグナルを検出することがある[30]．また、近縁な他のアフリカ集団での中立な C-A サブハプロタイプの t_D と比較して大きな差が見られなかった．以上の理由から、ESN で C-A サブハプロタイプに自然選択が働いていないと判断し、selective sweep タイプについても表記しない．

表５．東アジア各集団でのサブハプロタイプの種類及びその本数.

（a）C-A サブハプロタイプ

ハプロタイプ名	JPT	CHB	CHS	CDX	KHV	KOR	日本人サンプル（1KGPJPT を一部含む）	Total
H_1	29	52	52	51	54	48	10	296
H_2	0	0	1	0	0	0	0	1
H_3	2	4	3	2	3	2	0	16
H_4	3	3	4	0	0	0	0	10
H_5	0	0	1	0	0	0	0	1
H_6	0	0	1	3	0	0	0	4
H_7	0	0	1	0	0	0	0	1
H_8	0	1	1	0	0	0	0	2
H_9	2	3	4	3	2	0	0	14
H_10	0	2	1	0	1	0	0	4
H_11	0	5	1	3	1	1	1	12
H_12	0	0	1	0	0	0	0	1
H_13	0	0	0	1	0	0	0	1
H_14	0	0	0	2	2	0	0	4
H_15	0	0	0	2	0	0	0	2
H_16	0	0	0	2	1	0	0	3
H_17	0	0	0	0	1	0	0	1
H_18	0	0	0	0	1	0	0	1
H_19	0	0	0	0	1	0	0	1
H_20	0	0	0	0	1	0	0	1
H_21	0	0	0	0	1	0	0	1
H_22	0	0	0	1	0	0	0	1
H_23	0	0	0	1	0	0	0	1
H_24	0	1	0	0	1	0	0	2
H_25	0	1	0	0	0	0	0	1
H_26	0	1	0	0	0	0	0	1
H_27	0	1	0	0	0	0	0	1
H_28	0	1	0	0	0	0	0	1
H_29	1	1	0	0	0	3	0	5
H_30	0	1	0	0	0	0	0	1
H_31	0	1	0	0	0	0	0	1
H_32	0	1	0	0	0	0	0	1
H_35	0	0	0	0	0	1	0	1
H_38	0	0	0	0	0	2	0	2
Total	37	79	71	71	70	57	11	396

表中のハプロタイプの欠番は N（missing data）を含むため、network 図では主要なハプロタイプと統合されている.

(b) A-G サブハプロタイプ

ハプロタイプ番号	JPT	CHB	CHS	CDX	KHV	KOR	日本人サンプル(1KGPJPTを一部含む)	Total
H_1	16	36	41	34	39	48	4	218
H_2	0	0	1	0	0	0	0	1
H_3	12	14	13	2	4	23	7	75
H_4	0	0	1	0	0	0	0	1
H_5	0	0	1	0	0	0	0	1
H_6	1	0	4	9	9	0	0	23
H_7	0	0	1	0	0	0	0	1
H_8	6	7	2	0	3	17	2	37
H_9	0	1	3	7	6	0	0	17
H_10	1	1	3	0	1	0	0	6
H_11	0	2	2	1	0	0	0	5
H_12	0	0	1	0	0	0	0	1
H_13	0	0	2	0	0	0	0	2
H_14	0	0	1	1	1	0	0	3
H_15	0	0	1	0	0	0	0	1
H_16	0	0	1	0	0	0	0	1
H_17	0	0	2	0	0	0	0	2
H_18	0	0	0	1	0	0	0	1
H_19	0	0	0	0	1	0	0	1
H_20	0	0	0	1	0	0	0	1
H_21	0	0	0	1	0	0	0	1
H_22	0	0	0	0	1	0	0	1
H_23	0	0	0	1	1	0	0	2
H_24	0	0	0	0	1	0	0	1
H_25	0	0	0	0	1	0	0	1
H_26	0	0	0	0	1	0	0	1
H_27	0	0	0	0	2	0	0	2
H_28	0	0	0	0	1	0	0	1
H_29	0	2	0	0	1	0	0	3
H_30	0	0	0	2	0	0	0	2
H_31	0	0	0	1	0	0	0	1
H_32	0	0	0	1	0	0	0	1
H_33	0	0	0	1	0	0	0	1
H_34	0	0	0	1	0	0	0	1
H_35	0	1	0	0	0	0	0	1

ハプロタイプ番号	JPT	CHB	CHS	CDX	KHV	KOR	日本人サンプル（1KGPJPTを一部含む）	Total
H_36	0	1	0	0	0	0	0	1
H_37	0	1	0	0	0	0	0	1
H_38	0	1	0	0	0	0	0	1
H_39	0	1	0	0	0	0	0	1
H_40	0	1	0	0	0	0	0	1
H_41	0	1	0	0	0	0	0	1
H_42	0	1	0	0	0	0	0	1
H_43	0	1	0	0	0	0	0	1
H_44	0	1	0	0	0	0	0	1
H_45	0	1	0	0	0	0	0	1
H_46	1	0	0	0	0	0	0	1
H_47	1	0	0	0	0	0	0	1
H_48	1	0	0	0	0	0	0	1
H_49	1	0	0	0	0	0	0	1
H_53	0	0	0	0	0	1	0	1
Total	40	74	80	64	73	89	13	433

(c) T ハプロタイプ

ハプロタイプ番号	JPT	CHB	CHS	CDX	KHV	KOR	日本人サンプル(1KGPJPTを一部含む)	Total
H_1	13	18	18	20	24	29	7	129
H_2	23	10	6	7	7	35	11	99
H_3	0	0	1	0	0	0	0	1
H_4	36	11	13	13	8	49	16	146
H_5	0	0	1	0	0	0	0	1
H_6	0	0	1	0	0	0	0	1
H_7	0	0	1	0	0	0	0	1
H_8	1	0	1	1	0	0	0	3
H_9	0	0	1	0	0	0	0	1
H_10	0	0	1	0	2	0	0	3
H_11	3	0	1	0	0	0	0	4
H_12	9	1	3	0	1	13	2	29
H_13	1	1	1	0	1	0	0	4
H_14	0	0	1	0	0	0	0	1
H_15	5	0	1	0	0	0	0	6
H_16	1	0	1	0	0	0	0	2
H_17	0	0	1	0	0	0	0	1
H_18	0	0	1	0	0	0	0	1
H_19	0	0	1	0	0	2	0	3
H_20	0	0	1	0	0	0	0	1
H_21	0	0	1	0	0	0	0	1
H_22	0	0	0	1	0	0	0	1
H_23	0	0	0	1	0	0	0	1
H_24	0	0	0	0	2	0	0	2
H_25	1	0	0	0	1	0	0	2
H_26	0	0	0	0	1	0	0	1
H_27	1	0	0	1	1	0	0	3
H_28	0	0	0	0	1	1	0	2
H_29	0	0	0	0	1	0	0	1
H_30	0	0	0	0	1	0	0	1
H_31	0	0	0	0	1	0	0	1
H_32	0	0	0	0	1	0	0	1
H_33	0	0	0	0	1	0	0	1
H_34	0	0	0	0	1	0	0	1
H_35	0	0	0	1	0	0	0	1
H_36	0	0	0	2	0	0	0	2
H_37	0	0	0	1	0	0	0	1
H_38	0	0	0	1	0	0	0	1
H_39	0	0	0	1	0	0	0	1
H_40	0	0	0	1	0	0	0	1

ハプロタイプ番号	JPT	CHB	CHS	CDX	KHV	KOR	日本人サンプル（1KGPJPTを一部含む）	Total
H_41	0	1	0	0	0	0	0	1
H_42	0	1	0	0	0	0	0	1
H_43	0	1	0	0	0	1	1	3
H_44	0	1	0	0	0	0	0	1
H_45	0	1	0	0	0	0	0	1
H_46	0	1	0	0	0	0	0	1
H_47	0	1	0	0	0	0	0	1
H_48	0	1	0	0	0	0	0	1
H_49	0	1	0	0	0	0	0	1
H_50	0	1	0	0	0	0	0	1
H_51	2	0	0	0	0	0	0	2
H_52	1	0	0	0	0	0	0	1
H_53	1	0	0	0	0	0	0	1
H_54	4	0	0	0	0	0	0	4
H_55	1	0	0	0	0	0	0	1
H_56	1	0	0	0	0	0	0	1
H_57	3	0	0	0	0	5	2	10
H_58	1	0	0	0	0	0	0	1
H_59	1	0	0	0	0	0	0	1
H_60	1	0	0	0	0	0	0	1
H_61	3	0	0	0	0	7	2	12
H_62	1	0	0	0	0	0	0	1
H_63	1	0	0	0	0	0	0	1
H_64	1	0	0	0	0	2	3	6
H_65	1	0	0	0	0	0	0	1
H_66	1	0	0	0	0	0	0	1
H_67	1	0	0	0	0	0	0	1
H_68	1	0	0	0	0	0	0	1
H_69	1	0	0	0	0	0	0	1
H_70	1	0	0	0	0	0	0	1
H_71	1	0	0	0	0	0	0	1
H_72	1	0	0	0	0	0	0	1
H_73	1	0	0	0	0	0	0	1
H_74	1	0	0	0	0	0	0	1
H_75	1	0	0	0	0	0	0	1
H_76	1	0	0	0	0	0	0	1
H_77	1	0	0	0	0	2	0	3
H_78	1	0	0	0	0	4	2	7
H_79	1	0	0	0	0	0	0	1
H_89	0	0	0	0	0	1	0	1
H_90	0	0	0	0	0	2	0	2
H_93	0	0	0	0	0	1	0	1
H_94	0	0	0	0	0	1	0	1
Total	131	51	57	51	55	155	46	546

表 6. 交雑直前の世代での、JMN・A_CNT の T アレル頻度と F_{ST} 値（平均値）.

（a）中立条件でのシミュレーション

		JMNの最終世代のTアレル頻度									
		0<=x<0.1	0.1<=x<0.2	0.2<=x<0.3	0.3<=x<0.4	0.4<=x<0.5	0.5<=x<0.6	0.6<=x<0.7	0.7<=x<0.8	0.8<=x<0.9	0.9<=x<0.10
A_CNTの最終世代でのTアレル頻度	0<=y<0.1	0.0046	0.0084	0.0173	0.0264	0.0350	0.0436	0.0525	0.0622	0.0733	0.0865
	0.1<=y<0.2	0.0065	0.0039	0.0060	0.0111	0.0170	0.0226	0.0279	0.0331	0.0396	0.0486
	0.2<=y<0.3	0.0099	0.0055	0.0038	0.0053	0.0093	0.0142	0.0188	0.0228	0.0270	0.0326
	0.3<=y<0.4	0.0130	0.0088	0.0051	0.0037	0.0050	0.0085	0.0129	0.0170	0.0205	0.0237
	0.4<=y<0.5	0.0157	0.0125	0.0084	0.0050	0.0037	0.0050	0.0083	0.0124	0.0162	0.0181
	0.5<=y<0.6	0.0189	0.0162	0.0125	0.0083	0.0050	0.0037	0.0050	0.0083	0.0125	0.0143
	0.6<=y<0.7	0.0237	0.0203	0.0170	0.0130	0.0085	0.0050	0.0037	0.0051	0.0088	0.0114
	0.7<=y<0.8	0.0316	0.0268	0.0227	0.0188	0.0142	0.0093	0.0053	0.0038	0.0055	0.0090
	0.8<=y<0.9	0.0474	0.0391	0.0330	0.0276	0.0227	0.0172	0.0111	0.0060	0.0039	0.0065
	0.9<=y<0.10	0.1061	0.0938	0.0806	0.0681	0.0566	0.0448	0.0331	0.0212	0.0104	0.0048

色のついたセルは、F_{ST} の 1 位~3 位である。 1 位を赤、2 位を橙、3 位を黄のセルで示す.

（b-1）simCHB に C アレルで正の自然選択が働く条件でのシミュレーション（$2Ns = 1.0$）

A_CNTの最終世代でのTアレル頻度(y) ＼ JMNの最終世代のTアレル頻度(x)	0≤x<0.1	0.1≤x<0.2	0.2≤x<0.3	0.3≤x<0.4	0.4≤x<0.5	0.5≤x<0.6	0.6≤x<0.7	0.7≤x<0.8	0.8≤x<0.9	0.9≤x<0.10
0≤y<0.1	0.0046	0.0084	0.0173	0.0263	0.0351	0.0433	0.0525	0.0625	0.0740	0.0869
0.1≤y<0.2	0.0064	0.0039	0.0060	0.0111	0.0171	0.0227	0.0279	0.0334	0.0396	0.0489
0.2≤y<0.3	0.0098	0.0054	0.0038	0.0053	0.0093	0.0142	0.0189	0.0229	0.0271	0.0328
0.3≤y<0.4	0.0128	0.0087	0.0051	0.0037	0.0051	0.0086	0.0130	0.0171	0.0204	0.0239
0.4≤y<0.5	0.0156	0.0123	0.0082	0.0049	0.0037	0.0050	0.0083	0.0125	0.0162	0.0182
0.5≤y<0.6	0.0187	0.0159	0.0122	0.0082	0.0049	0.0037	0.0050	0.0084	0.0126	0.0144
0.6≤y<0.7	0.0235	0.0201	0.0168	0.0128	0.0084	0.0050	0.0037	0.0052	0.0089	0.0115
0.7≤y<0.8	0.0317	0.0264	0.0226	0.0186	0.0140	0.0092	0.0053	0.0038	0.0055	0.0090
0.8≤y<0.9	0.0470	0.0392	0.0328	0.0275	0.0225	0.0169	0.0109	0.0059	0.0039	0.0065
0.9≤y<0.10	0.1063	0.0935	0.0810	0.0685	0.0564	0.0445	0.0329	0.0212	0.0104	0.0048

色付きのセルの意味は表 6（a）に従う.

（b-2）simCHB に C アレルで正の自然選択が働く条件でのシミュレーション（$2Ns = 10.0$）

A_CNTの最終世代でのTアレル頻度(y) ＼ JMNの最終世代のTアレル頻度(x)	0≤x<0.1	0.1≤x<0.2	0.2≤x<0.3	0.3≤x<0.4	0.4≤x<0.5	0.5≤x<0.6	0.6≤x<0.7	0.7≤x<0.8	0.8≤x<0.9	0.9≤x<0.10
0≤y<0.1	0.0045	0.0087	0.0178	0.0273	0.0359	0.0444	0.0533	0.0640	0.0753	0.0884
0.1≤y<0.2	0.0060	0.0039	0.0064	0.0119	0.0181	0.0240	0.0292	0.0346	0.0413	0.0507
0.2≤y<0.3	0.0091	0.0051	0.0038	0.0058	0.0102	0.0154	0.0202	0.0243	0.0286	0.0347
0.3≤y<0.4	0.0118	0.0080	0.0047	0.0038	0.0056	0.0095	0.0142	0.0186	0.0220	0.0256
0.4≤y<0.5	0.0144	0.0113	0.0075	0.0046	0.0038	0.0055	0.0093	0.0138	0.0176	0.0198
0.5≤y<0.6	0.0172	0.0147	0.0113	0.0074	0.0045	0.0038	0.0055	0.0094	0.0138	0.0158
0.6≤y<0.7	0.0218	0.0188	0.0155	0.0117	0.0077	0.0046	0.0038	0.0057	0.0098	0.0126
0.7≤y<0.8	0.0297	0.0250	0.0211	0.0174	0.0130	0.0085	0.0049	0.0039	0.0060	0.0099
0.8≤y<0.9	0.0452	0.0377	0.0312	0.0264	0.0215	0.0160	0.0103	0.0056	0.0040	0.0070
0.9≤y<0.10	0.1056	0.0929	0.0800	0.0678	0.0556	0.0440	0.0326	0.0208	0.0101	0.0050

色付きのセルの意味は表 6（a）に従う.

（b-3）simCHB に C アレルで正の自然選択が働く条件でのシミュレーション（$2Ns = 50.0$）

A_CNTの最終世代でのTアレル頻度(y) ＼ JMNの最終世代のTアレル頻度(x)	0≤x<0.1	0.1≤x<0.2	0.2≤x<0.3	0.3≤x<0.4	0.4≤x<0.5	0.5≤x<0.6	0.6≤x<0.7	0.7≤x<0.8	0.8≤x<0.9	0.9≤x<0.10
0≤y<0.1	0.0044	0.0102	0.0205	0.0305	0.0396	0.0482	0.0576	0.0688	0.0802	0.0938
0.1≤y<0.2	0.0047	0.0045	0.0087	0.0157	0.0230	0.0296	0.0354	0.0419	0.0490	0.0591
0.2≤y<0.3	0.0065	0.0041	0.0047	0.0086	0.0145	0.0210	0.0267	0.0317	0.0363	0.0435
0.3≤y<0.4	0.0082	0.0055	0.0040	0.0049	0.0086	0.0142	0.0203	0.0254	0.0297	0.0342
0.4≤y<0.5	0.0099	0.0077	0.0052	0.0040	0.0051	0.0087	0.0142	0.0200	0.0249	0.0277
0.5≤y<0.6	0.0120	0.0102	0.0077	0.0051	0.0040	0.0051	0.0088	0.0143	0.0201	0.0227
0.6≤y<0.7	0.0156	0.0133	0.0110	0.0081	0.0054	0.0040	0.0051	0.0089	0.0147	0.0185
0.7≤y<0.8	0.0228	0.0188	0.0156	0.0127	0.0093	0.0060	0.0041	0.0050	0.0091	0.0144
0.8≤y<0.9	0.0378	0.0309	0.0255	0.0210	0.0169	0.0123	0.0077	0.0046	0.0048	0.0099
0.9≤y<0.10	0.1025	0.0902	0.0779	0.0653	0.0533	0.0419	0.0304	0.0190	0.0093	0.0059

色付きのセルの意味は表 6（a）に従う.

（b-4）simCHB に C アレルで正の自然選択が働く条件でのシミュレーション（$2Ns = 100.0$）

A_CNTの最終世代でのTアレル頻度(y) ＼ JMNの最終世代のTアレル頻度(x)	0≤x<0.1	0.1≤x<0.2	0.2≤x<0.3	0.3≤x<0.4	0.4≤x<0.5	0.5≤x<0.6	0.6≤x<0.7	0.7≤x<0.8	0.8≤x<0.9	0.9≤x<0.10
0≤y<0.1	0.0049	0.0125	0.0241	0.0350	0.0444	0.0539	0.0637	0.0751	0.0867	0.1003
0.1≤y<0.2	0.0042	0.0063	0.0127	0.0214	0.0300	0.0373	0.0442	0.0510	0.0593	0.0699
0.2≤y<0.3	0.0050	0.0046	0.0076	0.0137	0.0215	0.0295	0.0363	0.0422	0.0479	0.0559
0.3≤y<0.4	0.0059	0.0047	0.0052	0.0086	0.0146	0.0221	0.0298	0.0360	0.0412	0.0468
0.4≤y<0.5	0.0067	0.0056	0.0048	0.0057	0.0092	0.0152	0.0227	0.0301	0.0361	0.0398
0.5≤y<0.6	0.0076	0.0068	0.0056	0.0048	0.0058	0.0094	0.0154	0.0230	0.0304	0.0339
0.6≤y<0.7	0.0099	0.0087	0.0073	0.0058	0.0048	0.0057	0.0091	0.0153	0.0233	0.0283
0.7≤y<0.8	0.0155	0.0126	0.0106	0.0087	0.0066	0.0049	0.0052	0.0084	0.0149	0.0221
0.8≤y<0.9	0.0293	0.0236	0.0192	0.0155	0.0122	0.0089	0.0059	0.0048	0.0073	0.0148
0.9≤y<0.10	0.0990	0.0866	0.0742	0.0624	0.0507	0.0393	0.0280	0.0173	0.0087	0.0077

色付きのセルの意味は表 6（a）に従う.

表 7. 現代の JPT・CHB で定義したハプロタイプと、縄文系統のサンプル、チンパンジーの相同配列を加えて定義したハプロタイプの対応・頻度表.

現生サンプルから定義したハプロタイプ名	JPT+CHBでの本数	JPT+CHBでのハプロタイプの頻度	JPTでのハプロタイプの頻度	CHBでのハプロタイプの頻度	rs2294008のアレル	ハプロタイプの所在	JPTでのヘテロ接合度	CHBでのヘテロ接合度	JPT+CHBでのヘテロ接合度	F_{ST}	現生のサンプルと古代ゲノムのサンプルから定義したハプロタイプ名
86	46	0.111	0.168	0.053	T	JPT・CHB共通	0.280	0.101	0.198	0.036	H_3
26	80	0.193	0.139	0.248	C	JPT・CHB共通	0.240	0.373	0.312	0.018	H_2
88	33	0.080	0.111	0.049	T	JPT・CHB共通	0.197	0.092	0.147	0.015	H_1
56	52	0.126	0.077	0.175	C	JPT・CHB共通	0.142	0.288	0.220	0.020	H_5
84	31	0.075	0.063	0.087	T	JPT・CHB共通	0.117	0.159	0.139	0.001	H_4
48	26	0.063	0.058	0.068	C	JPT・CHB共通	0.109	0.127	0.118	0.000	H_8
40	10	0.024	0.043	0.005	T	JPT・CHB共通	0.083	0.010	0.047	0.019	H_3
12	13	0.031	0.029	0.034	C	JPT・CHB共通	0.056	0.066	0.061	0.000	H_5
28	5	0.012	0.024	0.000	T	JPTのみ	0.047	0.000	0.024	0.017	H_1
53	4	0.010	0.019	0.000	T	JPTのみ	0.038	0.000	0.019	0.014	H_1
7	3	0.007	0.014	0.000	T	JPTのみ	0.028	0.000	0.014	0.012	H_1
16	6	0.014	0.014	0.015	C	JPT・CHB共通	0.028	0.029	0.029	0.000	H_2
31	3	0.007	0.014	0.000	T	JPTのみ	0.028	0.000	0.014	0.012	H_13
43	3	0.007	0.014	0.000	T	JPTのみ	0.028	0.000	0.014	0.012	H_4
50	2	0.005	0.010	0.000	T	JPTのみ	0.019	0.000	0.010	0.010	H_3
70	6	0.014	0.010	0.019	C	JPT・CHB共通	0.019	0.038	0.029	0.000	H_22
89	5	0.012	0.010	0.015	C	JPT・CHB共通	0.019	0.029	0.024	0.000	H_2
14	1	0.002	0.005	0.000	C	JPTのみ	0.010	0.000	0.005	0.007	H_5
15	1	0.002	0.005	0.000	C	JPTのみ	0.010	0.000	0.005	0.007	H_8
1	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_3

現生サンプルから定義したハプロタイプ名	JPT+CHBでの本数	JPT+CHBでのハプロタイプの頻度	JPTでのハプロタイプの頻度	CHBでのハプロタイプの頻度	rs2294008のアレル	ハプロタイプの所在	JPTでのヘテロ接合度	CHBでのヘテロ接合度	JPT+CHBでのヘテロ接合度	F_{ST}	現生のサンプルと古代ゲノムのサンプルから定義したハプロタイプ名
21	2	0.005	0.005	0.005	C	JPT・CHB共通	0.010	0.010	0.010	0.000	H_26
8	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_27
18	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_21
19	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_4
20	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_1
24	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_7
25	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_3
27	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_3
29	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_18
30	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_13
32	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_1
33	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_1
35	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_17
38	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_19
39	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_3
49	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_24
52	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_1
47	2	0.005	0.005	0.005	C	JPT・CHB共通	0.010	0.010	0.010	0.000	H_10
57	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_12
58	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_1

現生サンプルから定義したハプロタイプ名	JPT+CHBでの本数	JPT+CHBでのハプロタイプの頻度	JPTでのハプロタイプの頻度	CHBでのハプロタイプの頻度	rs2294008のアレル	ハプロタイプの所在	JPTでのヘテロ接合度	CHBでのヘテロ接合度	JPT+CHBでのヘテロ接合度	F_{ST}	現生のサンプルと古代ゲノムのサンプルから定義したハプロタイプ名
21	2	0.005	0.005	0.005	C	JPT・CHB共通	0.010	0.010	0.010	0.000	H_26
62	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_11
55	1	0.002	0.005	0.000	C	JPTのみ	0.010	0.000	0.005	0.007	H_20
63	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_1
65	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_25
68	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_14
69	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_6
73	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_15
83	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_1
87	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_13
91	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_23
93	1	0.002	0.005	0.000	T	JPTのみ	0.010	0.000	0.005	0.007	H_3
71	1	0.002	0.005	0.000	C	JPTのみ	0.010	0.000	0.005	0.007	H_16
67	2	0.005	0.005	0.005	T	JPT・CHB共通	0.010	0.010	0.010	0.000	H_3
77	1	0.002	0.005	0.000	C	JPTのみ	0.010	0.000	0.005	0.007	H_5
2	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_5
3	2	0.005	0.000	0.010	C	CHBのみ	0.000	0.019	0.010	0.000	H_30
5	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_5
6	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_8
10	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_2

現生サンプルから定義したハプロタイプ名	JPT+CHBでの本数	JPT+CHBでのハプロタイプの頻度	JPTでのハプロタイプの頻度	CHBでのハプロタイプの頻度	rs2294008のアレル	ハプロタイプの所在	JPTでのヘテロ接合度	CHBでのヘテロ接合度	JPT+CHBでのヘテロ接合度	F_{ST}	現生のサンプルと古代ゲノムのサンプルから定義したハプロタイプ名
13	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_2
23	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_2
34	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_34
36	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_5
37	5	0.012	0.000	0.024	C	CHBのみ	0.000	0.047	0.024	0.008	H_2
41	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_5
44	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_2
45	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_5
46	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_32
51	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_36
54	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_8
59	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_5
64	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_22
66	2	0.005	0.000	0.010	C	CHBのみ	0.000	0.019	0.010	0.000	H_39
72	2	0.005	0.000	0.010	C	CHBのみ	0.000	0.019	0.010	0.000	H_29
4	1	0.002	0.000	0.005	T	CHBのみ	0.000	0.010	0.005	−0.003	H_3
9	1	0.002	0.000	0.005	T	CHBのみ	0.000	0.010	0.005	−0.003	H_1
76	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_31
78	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_40
79	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_37

現生サンプルから定義したハプロタイプ名	JPT+CHBでの本数	JPT+CHBでのハプロタイプの頻度	JPTでのハプロタイプの頻度	CHBでのハプロタイプの頻度	rs2294008のアレル	ハプロタイプの所在	JPTでのヘテロ接合度	CHBでのヘテロ接合度	JPT+CHBでのヘテロ接合度	F_{ST}	現生のサンプルと古代ゲノムのサンプルから定義したハプロタイプ名
80	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_38
81	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_2
11	1	0.002	0.000	0.005	T	CHBのみ	0.000	0.010	0.005	−0.003	H_1
17	1	0.002	0.000	0.005	T	CHBのみ	0.000	0.010	0.005	−0.003	H_28
22	1	0.002	0.000	0.005	T	CHBのみ	0.000	0.010	0.005	−0.003	H_1
85	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_2
42	1	0.002	0.000	0.005	T	CHBのみ	0.000	0.010	0.005	−0.003	H_4
60	1	0.002	0.000	0.005	T	CHBのみ	0.000	0.010	0.005	−0.003	H_33
74	1	0.002	0.000	0.005	T	CHBのみ	0.000	0.010	0.005	−0.003	H_3
90	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_5
75	1	0.002	0.000	0.005	T	CHBのみ	0.000	0.010	0.005	−0.003	H_3
92	1	0.002	0.000	0.005	C	CHBのみ	0.000	0.010	0.005	−0.003	H_2
82	1	0.002	0.000	0.005	T	CHBのみ	0.000	0.010	0.005	−0.003	H_35

表 8. rs2294008 を含む LD ブロックと、その上流・下流各 100 kb の領域の塩基多様度（π）.

	C allele (JPT=77, CHB=155)			T allele (JPT=131, CHB=49)		
領域	上流100kb	21.9 kb LD block	下流100kb	上流100kb	21.9 kb LD block	下流100kb
JPT	1.20×10^{-3} (1.10×10^{-4})	0.15×10^{-3} (8.26×10^{-5})	0.97×10^{-3} (9.85×10^{-5})	1.33×10^{-3} (1.15×10^{-4})	0.64×10^{-3} (1.71×10^{-4})	0.95×10^{-3} (9.75×10^{-5})
CHB	1.13×10^{-3} (1.06×10^{-4})	0.16×10^{-3} (8.54×10^{-5})	0.93×10^{-3} (9.64×10^{-5})	1.38×10^{-3} (1.17×10^{-4})	0.84×10^{-3} (1.96×10^{-4})	1.01×10^{-3} (1.00×10^{-4})

カッコ内の数字は標準偏差を示す.

発表論文リスト

1. Iwasaki, R. L., Ishiya, K., Kanzawa-Kiriyama, H., Kawai, Y., Gojobori, J., Satta, Y. Evolutionary history of the risk SNP for diffuse type gastric cancer in the Japanese. *Genes.* **2020**, *11*, 775. doi:10.3390/genes11070775.

2. Satta, Y., Zeng, W., Nishiyama, K. V., Iwasaki, R. L., Hayakawa, T., Fujito, N. T., Takahata, N. Two-dimensional site frequency spectrum for detecting, classifying and dating incomplete selective sweeps. *Genes. Genet. Syst.* **2019**, *94*, 283-300. doi:10.1266/ggs.19-00012.