Global adaptive evolution involved in neuroticism and educational behaviors through the spread of anatomically modern humans

Toshiyuki Hayakawa[1*], Naoko T. Fujito[2], Yuka Moriya[3], Masahiro Terahara[3], Kosuke M. Teshima[4], Naoyuki Takahata[3], Yoko Satta[3]

[1]Faculty of Arts and Science, Kyushu University, Fukuoka, Japan

[2]School of Advanced Sciences, SOKENDAI (The Graduate University for Advanced Studies), Hayama, Kanagawa, Japan

[3]Graduate School of Systems Life Sciences, Kyushu University, Fukuoka, Japan

[4]Department of Biology, Kyushu University, Fukuoka, Japan

* Address correspondence to

Toshiyuki Hayakawa

Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan

Phone: +81-92-802-5827; E-mail: thayakawa@artsci.kyushu-u.ac.jp

Abstract

During the spread of anatomically modern humans, cumulative cultural evolution occurred globally. It is anticipated that this evolution was accompanied by global changes in mental/behavioural phenotypes. Through the establishment of a new score summarizing allelic function at the population level (population functional score: PFS), global distributions of PFSs (PFS-26s, based on 26 different populations) for schizophrenia-associated genetic loci showing selective increases of non-risk allelic types were obtained as yardsticks, and PFS-26s similar to them were detected in alleles associated with different phenotypes. We found that alleles contributing to the suppression of neuroticism and increase in educational attainment showed dominant trends of becoming similar to the yardstick PFS-26s in their PFS-26s. This indicates that neuroticism suppression and development of educational behaviours occurred through the global spread of anatomically modern humans. Taking into account the status of psychosocial stress as an environmental factor shared among schizophrenia, neuroticism and educational attainment, it is considered that an evolutionary driver of such phenotypic changes is global adaptation to psychosocial stress, which has contributed to cumulative cultural evolution by facilitating social interactions.

Introduction

Anatomically modern humans (AMHs; *Homo sapiens*) emerged in Africa and expanded their habitat to almost every corner of the globe (1). During this spread, AMHs generated greater technological sophistication and larger bodies of adaptive knowledge to survive efficiently in new environments (2). In this cumulative cultural evolution, adaptation of mental/behavioural phenotypes would have occurred, forming a global evolutionary trend based on the distribution of associated alleles. One such trend would be adaptive evolution of the *ST8SIA2* gene (3, 4), which is related to brain/neural function (5). Selective increase of its promoter types (haplotypes of three promoter SNPs) annotated as schizophrenia non-risk types, which show lower promoter activity than the risk types, has induced a trend toward lower promoter activity at the population level (population promoter activity: PPA) through the out-of-Africa migration of AMHs. This can be interpreted as an adaptation to psychosocial stress (3, 4), which has arisen from tension during adaptation to surrounding social environments, because such stress is the sole environmental risk factor for schizophrenia dealt with by brain functions (6, 7) (Supplementary Fig. 1). On this basis, we here applied evolutionary trends of schizophrenia-associated genetic loci showing selective increases of non-risk allelic types as yardsticks to examine similar trends in allelic types classified by their effect on different phenotypes, and detected the dominance of such trends in the evolution of alleles uniquely associated with particular mental phenotypes (Supplementary Fig. 2).

Results

Population functional score (PFS) of the *ST8SIA2* gene

To apply the functional trend of the *ST8SIA2* gene as one of the yardsticks in the analysis using alleles for which no functional information was available, we first developed a score (population functional score: PFS) reflecting allelic function at the population level by focusing on the functional difference (*D*) between alleles, and then examined the correlation in PFS distributions of 26 populations spread globally (PFS-26s), encompassing Africa (AFR), Europe (EUR), South Asia (SAS), East/Southeast Asia (EAS) and the Americas

(AMR), from the 1000 Genomes Project (8), making experimental estimation of *D* unnecessary (see Materials and Methods section; Supplementary Figs. 2 and 3). The PFS-26 of the *ST8SIA2* gene (PFS-$26_{ST8SIA2}$) was obtained by determining the difference in promoter activity from the ancestral risk type showing the highest activity (Fig. 1A). PPA and PFS are negatively corelated, and PFS-$26_{ST8SIA2}$ shows a correlation coefficient (ρ) of –0.99 with the PPAs of the 26 populations, indicating the precise translation of PPA into PFS. Since PFS-$26_{ST8SIA2}$ mainly reflects the selective increase of non-risk types in non-AFR (3, 4) (see Fig. 1A), we used this as a yardstick to examine selection in non-AFR.

The yardstick PFS-26 to examine selection in AFR

To obtain a yardstick PFS-26 reflecting the selective increase of alleles in AFR, we next examined positive selection on schizophrenia-associated SNPs (9) (SCZ1 in Table 1 and Supplementary Table 1). We detected signals of positive selection on non-risk alleles of three SNPs of genes related to brain/neural function meeting the prioritization criteria [rs11693094 (F_c = 0.024, P = 0.003) in the *ZNF804A* gene, rs9922678 (F_c = 0.109, $P = 4.3 \times 10^{-4}$) in the *GRIN2A* gene and rs1501357 (F_c = 0.028, P = 0.003) in the *HCN1* gene; ref. 10]. The durations of operation of this positive selection were estimated to be 17,143, 25,237 and 33,621 years for rs11693094, rs9922678 and rs1501357, respectively. These findings indicate that the detected positive selection can be interpreted as an adaptation to psychosocial stress during the spread of AMHs. A signal of positive selection on the non-risk allele of rs1501357 was detected in AFR. We therefore selected its PFS-26 (PFS-$26_{rs1501357}$) as the yardstick to examine positive selection in AFR (Fig. 1B). As for the non-risk alleles of rs11693094 and rs9922678, signals were detected in EUR and EAS, respectively.

Applicability of index SNPs and criterion in correlation analysis of PFS-26s

SNPs associated with phenotypes were obtained from GWASs (see Table 1 and Supplementary Table 1 for abbreviations of the GWAS SNP sets), and therefore index SNPs were used instead of causal SNPs to obtain PFS-26s. Nearly complete linkage between index and causal SNPs can be assumed only in populations targeted in the GWAS analyses. Thus, the PFS-26s would not be completely identical to those from causal SNPs. To assess the

extent of this discrepancy, we focused on SNP pairs (concordant SNP pairs) showing tight linkage between the same allele type of a certain phenotype in EUR because the European population or meta-population containing Europeans was mainly targeted to identify index SNPs used in this study (Supplementary Fig. 4). The size of the associated region is typically < 200 kb in length (11). In such < 200-kb regions, tight linkage (r^2 = 0.83–1) was found in 64 concordant SNP pairs identified independently by different GWASs for the same phenotype (neuroticism, intelligence, schizophrenia, and type 2 diabetes; see Materials and Methods section). By assuming these tight-linked SNPs as index and causal SNPs, we finally examined the correlation between PFS-26s in SNP pairs. The ρ values were mainly distributed in the range from 0.9 to 1 (40/64; Supplementary Fig. 5A). The PFS-26 proportion in a sliding window of ρ = 0.05 became consistently at most ~6% after a window shift to covering PFS-26s showing ρ < 0.9 (Supplementary Fig. 5B). This indicates that, if PFS-26 from a causal SNP is completely identical to yardstick PFS-26s (i.e., ρ = 1), the PFS-26 from the index SNP usually shows ρ > 0.9 with the yardstick PFS-26. Thus, ρ > 0.9 was considered to be a reasonable criterion to assess high correlation with the yardstick PFS-26s.

The PFS-26s of non-risk alleles of rs11693094 and rs9922678, showing positive selection in non-AFR (see above), exhibit ρ > 0.9 with PFS-26$_{ST8SIA2}$ [ρ = 0.92 (rs11693094) and 0.92 (rs9922678); see Supplementary Fig. 6]. Meanwhile, PFS-26$_{rs1501357}$ shows ρ = 0.68 with PFS-26$_{ST8SIA2}$, below the threshold of 0.9. These findings indicate that PFS-26$_{ST8SIA2}$ reflects the PFS distribution resulting from selective allelic increase in non-AFR, and are consistent with the assertion that PFS-26$_{ST8SIA2}$ does not well document the selective increase of alleles in AFR. The criterion of ρ > 0.9 with the yardstick PFS-26s is therefore applicable to detect genetic loci showing a PFS distribution matching that of the yardstick PFS-26s. Furthermore, even if index SNPs are not guaranteed to be causal SNPs, this also suggests that index SNPs can be used instead of causal SNPs in our analysis.

Appearance of PFS-26s showing high correlation with the yardstick PFS-26s

The observed rate of PFS-26s exhibiting ρ > 0.9 with the yardstick PFS-26s (PFS-26$_{ST8SIA2}$-like PFS-26s and PFS-26$_{rs1501357}$-like PFS-26s; yardstick-like PFS-26s) was examined in each allelic type (Table 1). The rate in non-risk alleles/negative-effect alleles was compared with

that in risk alleles/positive-effect alleles in each GWAS SNP set. By using PFS-26$_{ST8SIA2}$, Bonferroni-corrected significant differences ($\alpha = 0.01/13$; $P < 7.7 \times 10^{-4}$) were found in NEU1, NEU2 and EDU (see Table 1), and the rates in negative-effect alleles of NEU1 and NEU2 and positive-effect alleles of EDU were higher than those in the opposite-effect alleles [$P = 3.7 \times 10^{-5}$ (NEU1), $P = 7.2 \times 10^{-5}$ (NEU2) and $P < 7.7 \times 10^{-4}$ (EDU); binomial test] (Table 1). Meanwhile, the use of PFS-26$_{rs1501357}$ represented that Bonferroni-corrected significance was found in EDU and DEP, and the rates in negative-effect alleles of NEU1, positive-effect alleles of EDU and non-risk alleles of DEP were higher than those in the opposite-effect alleles [$P < 7.7 \times 10^{-4}$ (NEU1 and EDU) and $P = 6.5 \times 10^{-5}$ (DEP); binomial test] (Table 1). The observed significant appearance rate of the yardstick-like PFS-26s represents convergent evolution of PFS-26s through the global spread of AMHs (Supplementary Fig. 2).

Correlation between the yardstick PFS-26 and PFS-26s characteristic of phenotypes

To examine the involvement of the yardstick PFS-26s in PFS-26s unique to each phenotype, we obtained PFS-26s characteristic of each allelic type showing a significant appearance rate of the yardstick-like PFS-26s (see Supplementary Fig. 7). By using the median of ρ values in pairwise comparison among PFS-26s as a threshold to identify a trend of similarity, characteristic PFS-26s were identified as those of which the number of PFS-26s showing ρ exceeding the median value was significantly larger with the same allelic types than with the opposite-effect types ($\alpha = 0.05$/number of alleles). Such characteristic PFS-26s can be considered to reflect the allele distribution established uniquely in a certain phenotype, and were obtained in NEU1, NEU2 and EDU (Fig. 2 and Supplementary Fig. 8). Interestingly, the distribution of ρ values with the yardstick PFS-26s was remarkably restricted to a positive correlation in characteristic PFS-26s of negative-effect alleles of NEU1 [100% (8/8) in PFS-26$_{ST8SIA2}$ and PFS-26$_{rs1501357}$] and NEU2 [100% (3/3) in PFS-26$_{ST8SIA2}$ and PFS-26$_{rs1501357}$], and the positive-effect alleles of EDU [96% (2136/2223) in PFS-26$_{ST8SIA2}$; 98% (2180/2223) in PFS-26$_{rs1501357}$] (Fig. 2A–C and Supplementary Fig. 8).

Domination of the yardstick PFS-26s in characteristic PFS-26s

The degree to which positive correlations with the yardstick PFS-26s dominated in the characteristic PFS-26s can be elucidated by determining the correlation in characteristic PFS-26 between ρ values with the yardstick PFS-26s and the average ρ values with others (Fig. 2D–F; Supplementary Fig. 9). As for PFS-26$_{\mathrm{ST8SIA2}}$, high R^2 values with Bonferroni-corrected significant positive correlations ($\alpha = 0.01/3$; $P < 3.3 \times 10^{-3}$) were found in characteristic PFS-26s of NEU1 and EDU (Fig. 2D and E). Higher ρ values with PFS-26$_{\mathrm{ST8SIA2}}$ were associated with higher average ρ values with others in negative-effect alleles of NEU1 ($R^2 = 0.84$, $P = 1.5 \times 10^{-3}$) and positive-effect alleles of EDU ($R^2 = 0.43$, $P = 2.6 \times 10^{-275}$) (Fig. 2D and E). Meanwhile, for PFS-26$_{\mathrm{rs1501357}}$, a high R^2 value with Bonferroni-corrected significant positive correlation was found in EDU (Fig. 2F). Higher ρ values with PFS-26$_{\mathrm{rs1501357}}$ were associated with higher average ρ values with others in positive-effect alleles of EDU ($R^2 = 0.53$, $P < 3.3 \times 10^{-3}$) (Fig. 2F). Taken together with the significant appearance rate of the yardstick-like PFS-26s (Table 1) and an exclusive restriction toward positive ρ values with the yardstick PFS-26s (Fig. 2A–C), the observed high R^2 values with positive correlations with the yardstick PFS-26s in the characteristic PFS-26s of the negative-effect alleles of NEU1 and the positive-effect alleles of EDU (Fig. 2D–F) indicate that a trend of becoming similar to the yardstick PFS-26s dominates for the PFS-26s unique to these allele groups.

Unique involvement of the yardstick PFS-26s in neuroticism and educational attainment PFS-26s

To further confirm the uniqueness of the involvement of the yardstick PFS-26s in NEU1 and EDU, we applied the PFS-26s of alleles already known as selective targets (rs4988235 in the *MCM6* gene, rs1800414 in the *OCA2* gene and rs3827760 in the *EDAR* gene; ref. 12–14) as yardsticks (control-yardstick PFS-26s). The control-yardstick PFS-26s showed no dominant trend of becoming similar to the characteristic PFS-26s of NEU1 and EDU (Supplementary Table 2 and Supplementary Fig. 10).

Phenotype evolution driven by adaptation to psychosocial stress

The dominant trend of becoming similar to the yardstick PFS-26s was detected only in the

PFS-26s of neuroticism negative-effect alleles and educational attainment positive-effect alleles (Table 1 and Fig. 2). Taken together with the genetic correlation between lower neuroticism and increased educational attainment (15), such a dominant trend indicates that single global adaptive evolution has driven neuroticism suppression and increased behaviours currently having positive effects on educational attainment. Neuroticism is a personality trait associated with sensitivity to psychosocial stress (16). Meanwhile, educational attainment is a behavioural phenotype and is primarily determined by environmental factors. Because progress in learning via communication achieved by overcoming psychosocial stress has a positive effect on educational attainment, psychosocial stress is considered to be such an environmental factor. The tolerance to psychosocial stress is thus a shared function between neuroticism negative-effect alleles, educational attainment positive-effect alleles and schizophrenia non-risk alleles (see also Supplementary Fig. 1). It is therefore considered that the dominant trend of becoming similar to the yardstick PFS-26s in the PFS-26s of neuroticism negative-effect alleles and educational attainment positive-effect alleles means a polygenic trend toward higher frequencies of types conferring tolerance to psychosocial stress through the global spread of AMHs (i.e., global polygenic adaptation to psychosocial stress).

Discussion

The yardstick PFS-26s are just two patterns of PFS distribution resulting from global adaptation to psychosocial stress. Because other patterns would be present but are currently unknown, the yardstick PFS-26s should not be misinterpreted as reflecting the difference in adaptation to psychosocial stress between populations.

In the PFS-26s of NEU1, the dominant trend of becoming similar to the yardstick PFS-26s was detected with the PFS-$26_{ST8SIA2}$, but not with the PFS-$26_{rs1501357}$. This might mean that NEU1 has more genetic loci participating in the adaptation to psychosocial stress in non-AFR than that in AFR.

NEU1 showed the dominant trend of becoming similar to PFS-$26_{ST8SIA2}$, but NEU2 did not (Fig. 2A and Supplementary Fig. 11A). This may represent the inclusion of more

genetic loci participating in the adaptation to psychosocial stress in NEU1 than in NEU2. The difference of items used in phenotype assessment between NEU1 and NEU2 (17, 18) might have influenced the number of genetic loci identified as contributing to the adaptation to psychosocial stress.

Negative-effect/non-risk alleles of NEU2 and DEP showed only significant appearance rate of the yardstick-like PFS-26s (Table 1). Because sensitivity to psychosocial stress characterizes neuroticism and plays an important role in the onset of depression, this finding may represent only partial inclusion of the selective targets participating in adaptation to psychosocial stress in NEU2 and DEP.

Although psychosocial stress is one of the environmental risk factors associated with the onset of schizophrenia and depression (6, 7, 19, 20), we did not detect any dominant trend of becoming similar to the yardstick PFS-26s in SCZ1, SCZ2 or DEP. This may have been caused by the inclusion of their genetic loci associated with other environmental risk factors because the limited number of genetic loci associated with sensitivity to psychosocial stress would lessen the detection power in our analysis. Importantly, this indicates that the yardstick-like PFS-26s did not result from negative selection against schizophrenia and depression.

In addition to NEU1, NEU2 and EDU, characteristic PFS-26s were identified in INT, SCZ2, HEI, T2D and CHR (Supplementary Fig. 11). In the characteristic PFS-26s of the non-risk alleles of SCZ2 and the positive-effect alleles of HEI, a predominant restriction toward positive ρ values [100% (26/26) in SCZ2; 96% (1036/1077) in HEI] and high R^2 values with positive correlations ($R^2 = 0.85$, $P = 1.8 \times 10^{-11}$ in SCZ2; $R^2 = 0.33$, $P = 7.3 \times 10^{-96}$ in HEI) were found with PFS-26$_{rs1501357}$ (Supplementary Fig. 11H and I). Because of the lack of a dense appearance of the yardstick-like PFS-26s in SCZ2 and HEI (Table 1), this might be interpreted as an unknown evolutionary process resulting in PFS-26s similar to PFS-26$_{rs1501357}$, not by adaptation to psychosocial stress.

Meeting the criterion of $\rho > 0.9$ with the yardstick PFS-26s alone is insufficient to conclude that certain PFS-26s are involved in adaptation to psychosocial stress. To elucidate such involvement in each yardstick-like PFS-26 of NEU1, NEU2 and EDU, examination by detecting positive selection would be needed, but this is beyond the scope of this study.

Henrich (2) argued that, once individuals evolve to learn from one other with sufficient accuracy, social groups of individuals develop what might be called *collective brains*, which drive cumulative cultural evolution. It is likely that overcoming psychosocial stress in the spread of AMHs, that is, global adaptation to psychosocial stress, has made these collective brains bigger by increasing the degree of social interaction and population size, and accelerated cumulative cultural evolution (4). This study proposes that the current genetic basis of neuroticism and educational attainment has been established through such evolution (Supplementary Fig. 12).

In conclusion, a dominant trend of PFS-26s becoming similar to the yardstick PFS-26s was identified in the evolution of neuroticism negative-effect alleles and educational attainment positive-effect alleles. This indicates a global trend toward neuroticism suppression and development of educational behaviours through the spread of AMHs. Global adaptation to psychosocial stress is considered as a driver of this trend. It is therefore suggested that our cooperativeness has improved globally.

Materials and Methods

SNPs associated with phenotypes

Single-nucleotide polymorphisms (SNPs) associated with different phenotypes were obtained from genome-wide association studies (GWASs) reported to date (9, 10, 17, 18, 21–29) (see Supplementary Table 1). SNPs identified by genome-wide significance ($P < 5.0 \times 10^{-8}$) were used in this study. Because GWASs not using conditional and joint multiple-SNP (COJO) analysis (30) were included in this study, we used SNP sets identified without COJO analysis, with the exceptions of BMI and HEI for which only SNP sets identified by COJO analysis were available. On the basis of the 1000 Genomes Project (hereafter abbreviated as 1KGP) database (8), we eliminated SNPs for which allelic frequencies are not available, SNPs for which multiple derived alleles are indicated, SNPs for which only derived alleles are found in AMHs, SNPs located on the X chromosome (see below) and indels. In this study, we mainly used GWASs that identified more than or close to 100 associated genetic loci to maintain statistical power for our analyses (see Supplementary Table 1).

Phenotypes and GWASs

Psychosocial stress acts as an environmental risk factor in the onset of schizophrenia and depressive disorder (6, 7, 19, 20). Neuroticism also acts as an environmental risk factor in the onset of conditions such as schizophrenia and depression (17), and more neurotic individuals are vulnerable to social stress. We also previously pointed out a possible role of adaptation to psychosocial stress in learning via communication (3, 4). Therefore, we selected GWASs of neuroticism (NEU1 and NEU2; ref. 17 and 18), educational attainment (EDU; ref. 21), intelligence (INT; ref. 22), schizophrenia (SCZ1 and SCZ2; ref. 9 and 10) and depressive disorder (DEP; ref. 23) for inclusion in this study. In addition to educational attainment, we choose chronotype as an additional behavioural phenotype, and included a GWAS on it (CHR; ref. 24). We also focused on GWASs of physiological phenotypes [body mass index (BMI), BMI; height, HEI; vitamin D level, VID; type 2 diabetes, T2D; migraine, MIG] (ref. 25–29) to examine whether the adaptation to psychosocial stress has made physiological contributions. In addition to these GWASs, we included GWASs identifying far fewer than 100 associated genetic loci (for neuroticism, intelligence, schizophrenia and type 2 diabetes; ref. 31–38) in the case of linkage analysis (see main text).

DNA sequence data

Frequencies of each allele of SNPs were retrieved from phase 3 of the 1KGP database (8), covering 2,504 individuals (5,008 sequences) of 26 different global populations. By following the 1KGP database (8), these populations were classified into five meta-populations: Africa (AFR), Europe (EUR), South Asia (SAS), East and Southeast Asia (EAS), and the Americas (AMR).

Population functional score (PFS)

Three SNPs (rs3759916, rs3759915 and rs3759914) in the promoter region of the *ST8SIA2* gene, which encodes a sialyltransferase synthesizing a linear homopolymer of sialic acid (polysialic acid) in the brain (5), are associated with schizophrenia risk (39–41). Selective increase of non-risk haplotypes of these three promoter SNPs (TCT and CGC types) has

occurred during the spread of AMHs (3, 4). These two promoter types show lower promoter activities than the risk types, such as TGT and CGT types (3), resulting in a trend toward lower population promoter activity (PPA) in the spread of AMHs (4). This trend toward lower PPA in the *ST8SIA2* gene can be interpreted as an adaptation to psychosocial stress, and has been established in populations worldwide through the spread of AMHs (4). Thus, the examination of local adaptation is not sufficient to identify similarity with this global functional trend in SNPs associated with phenotypes. Furthermore, because PPAs are estimated by using both frequencies and promoter activities of promoter types, PPA cannot be estimated for GWAS SNPs, the functions of which are usually unknown. By focusing on the difference in allelic function, we therefore established a score that makes it possible to examine the similarity of any SNPs with the global trend of population function of the *ST8SIA2* gene, even if allelic functions of those SNPs are unknown. SNPs identified from GWAS analysis by their association with phenotypes are biallelic. By setting D as the difference in allelic function (e.g., promoter activity) in allele A1 from allele A2 at a biallelic SNP and assuming additive allelic effects, the functional difference is given by $2D$, D, and 0 for the A1/A1 homozygote, A1/A2 heterozygote and A2/A2 homozygote, respectively. The functional difference at the population level (population function score: PFS) is then obtained by

$$PFS = \frac{2Dn_{11} + Dn_{12}}{2(n_{11} + n_{12} + n_{22})} = \frac{2n_{11} + n_{12}}{2(n_{11} + n_{12} + n_{22})} D$$

where n_{11}, n_{12} and n_{22} are the numbers of A1/A1 homozygotes, A1/A2 heterozygotes and A2/A2 homozygotes in the population, respectively. This means that PFS can also be defined simply by

$$PFS = f_{A1} D$$

where f_{A1} is A1 allele frequency $[(2n_{11} + n_{12})/2(n_{11} + n_{12} + n_{22})]$.

PFS of the *ST8SIA2* gene was obtained using the actual functional difference (i.e., promoter activity). The TGT type is the ancestral type, and shows the highest promoter activity[3]. Thus, PFS of the *ST8SIA2* gene ($PFS_{ST8SIA2}$) is obtained by

$$PFS_{ST8SIA2} = f_{CGT}D_{CGT} + f_{TCT}D_{TCT} + f_{CGC}D_{CGC}$$

where D_{CGT}, D_{TCT} and D_{CGC} are differences of promoter activities from the TGT type in

the CGT, TCT and CGC types, and f_{CGT}, f_{TCT} and f_{CGC} are frequencies of the CGT, TCT and CGC types, respectively.

Importantly, D of GWAS SNP is cancelled out in the estimation of the correlation coefficient (ρ) (i.e., similarity) between the distributions of PFSs from 26 populations (PFS-26; global trend of population function). Thus, correlation analysis using PFS-26 was performed to examine the similarity of global trend of population function without experimental estimation of allelic function (see Supplementary Figs. 2 and 3).

As for SNPs on the X chromosome, the difference in PFSs among populations (i.e., PFS-26) would become inaccurate if the sex ratio differs among the populations. Because FIN in the 1KGP showed a significant difference in sex ratio from the expected one (i.e., a 1:1 male to female ratio) (P = 0.02), SNPs on the X chromosome were excluded from this study. In the GWASs used in this study, the SNPs on the X chromosome were as follows: 3 in SCZ1, 5 in SCZ2, 57 in EDU, 3 in CHR, 1 in VID and 2 in MIG. They thus constitute a very small proportion of the SNPs in each GWAS.

Rate of high correlation with yardstick PFS-26s

The observed rate of high correlation with certain PFS-26s used as a yardstick was obtained for each allele type group (e.g., risk alleles of SCZ1). A binomial test was applied to examine whether this rate became higher in certain allelic types than in opposite allelic types. Unfortunately, the standard observed rate of PFS-26s showing a high correlation with the yardstick PFS-26 is unknown in SNPs associated with phenotypes. We therefore compared the rate in negative-effect alleles/non-risk alleles with that in positive-effect alleles/risk alleles for each GWAS. Because populations have genetic kinship with each other, the correlation between PFS-26s cannot be examined under the assumption of independence among populations (see Supplementary Fig. 3). To strictly assess the high correlation between PFS-26s, a criterion considering such genetic kinship was thus needed (see main text).

Identification of characteristic PFS-26s in allelic types

Similar PFS-26s uniquely found in a certain allelic type (e.g., schizophrenia risk alleles) were

identified in pairwise comparison of PFS-26s (Supplementary Fig. 7). In this identification, we focused on the number of PFS-26s ($N_{>\mathrm{MED}}$) showing ρ exceeding the median value of all pairwise comparisons of PFS-26s. A particular type, of which $N_{>\mathrm{MED}}$ is significantly higher in comparison with the same allelic types than with the opposite-effect types ($P < 0.05$; χ^2 test), is categorized as a characteristic PFS-26. In the χ^2 test, results were not considered if $N_{>\mathrm{MED}}$ was less than 10.

Detecting positive selection acting on alleles

We applied the F_c statistic to detect positive selection on SNPs in three meta-populations: AFR, EUR and EAS (3, 42, 43). To detect all possible genetic loci, we modified a previously reported simulation method (3, 42) by including alleles showing S (number of segregating sites) of at least 100. Benjamini & Hochberg correction was performed for multiple testing. Significant discovery was then examined at a false discovery rate (FDR) of < 0.1.

Data availability

The genomes used are available from the 1000 Genomes Project (phase 3 release, https://www.internationalgenome.org/).

Acknowledgements

This research was supported by the Japanese Society for the Promotion of Science (JP16K07535, JP19K06866 and JP24K09628). We thank Edanz (https://jp.edanz.com/english-editing) for editing a draft of this manuscript.

Author contributions

T.H. conceived and supervised the study. T.H., N.T.F., and Y.M. collected sequence data. T.H., N.T.F., M.T., N.T. and Y.S. performed analysis detecting signals of positive selection. T.H. and Y.M. performed correlation and regression analyses. N.T.F., K.M.T., N.T., and Y.S. read and commented on the pre-final version of the manuscript. T.H. wrote the paper.

Competing interest declaration

The authors declare no competing interest.

Table

Table 1. Observed rate of yardstick-like PFS-26s.

	Phenotype	GWAS	PFS-26$_{5T8SIA2}$		PFS-26$_{rs1501357}$ (HCN1)	
			Allele		Allele	
Mental			Positive-effect	Negative-effect	Positive-effect	Negative-effect
	Neuroticism	NEU1	0.015 (2/131)	0.076 (10/131)	0 (0/131)	0.038 (5/131)
		NEU2	0.009 (1/109)	0.064 (7/109)	0.018 (2/109)	0.018 (2/109)
	Educational attainment	EDU	0.057 (218/3845)	0.019 (74/3845)	0.024 (94/3845)	0.006 (24/3845)
	Intelligence	INT	0.030 (7/231)	0.056 (13/231)	0.022 (5/231)	0.026 (6/231)
	Chronotype	CHR	0.039 (13/332)	0.021 (7/332)	0.039 (13/332)	0.027 (9/332)
			Risk	Non-risk	Risk	Non-risk
	Schizophrenia	SCZ1	0.038 (4/106)	0.085 (9/106)	0 (0/105)	0.038 (4/105)
		SCZ2	0.064 (17/267)	0.064 (17/267)	0.026 (7/267)	0.049 (13/267)
	Depression	DEP	0.028 (8/288)	0.052 (15/288)	0.014 (4/288)	0.049 (14/288)
Physical			Positive-effect	Negative-effect	Positive-effect	Negative-effect
	Body mass index (BMI)	BMI	0.032 (29/899)	0.052 (47/899)	0.028 (25/899)	0.020 (18/899)
	Height	HEI	0.056 (670/11872)	0.054 (640/11872)	0.025 (301/11872)	0.022 (260/11872)
	Vitamine D	VID	0.026 (2/78)	0.051 (4/78)	0.038 (3/78)	0.013 (1/78)
			Risk	Non-risk	Risk	Non-risk
	Type 2 Diabetes	T2D	0.037 (5/134)	0.037 (5/134)	0.022 (3/134)	0.030 (4/134)
	Migraine	MIG	0.052 (6/116)	0.043 (5/116)	0.026 (3/116)	0.026 (3/116)

Figures

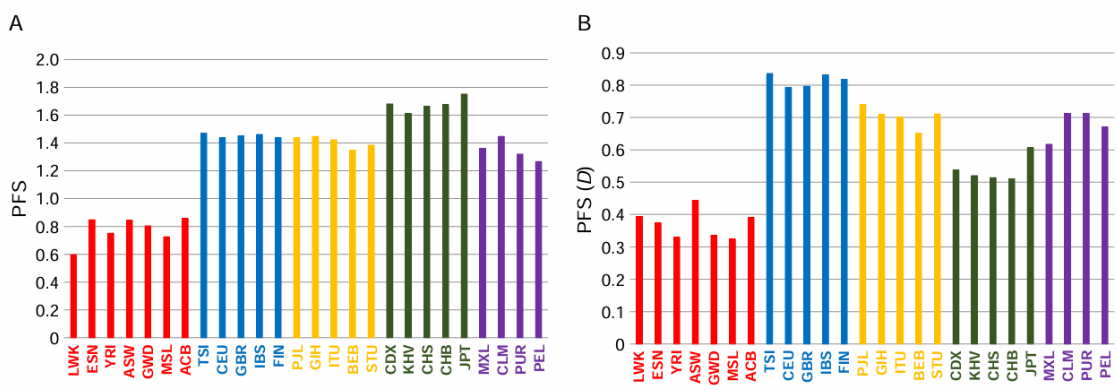

Fig. 1. (A) PFSs of the promoter types of the *ST8SIA2* gene. (B) PFSs of the schizophrenia non-risk allele of rs1501357. Populations are represented by abbreviations used in the 1000 Genomes Project (ref. 8).

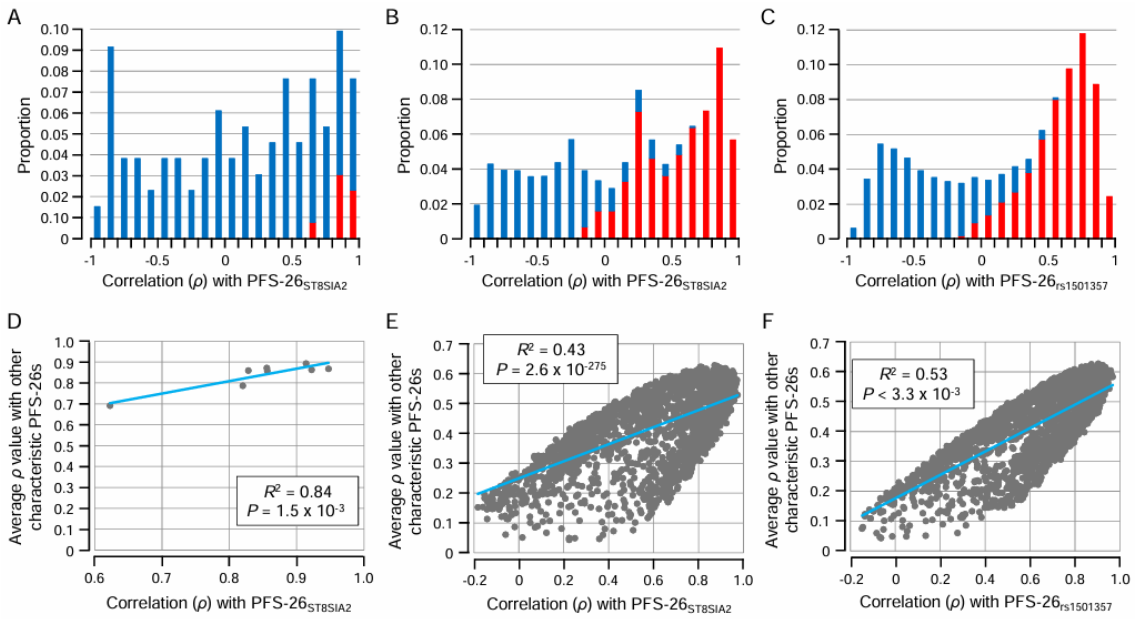

Fig. 2. Involvement of the yardstick PFS-26s in the evolution of alleles associated with neuroticism and educational attainment. (A) Distribution of correlation coefficients with PFS-$26_{ST8SIA2}$ in PFS-26s of negative-effect alleles of NEU1. (B) Distribution of correlation coefficients with PFS-$26_{ST8SIA2}$ in PFS-26s of positive-effect alleles of EDU. (C) Distribution of correlation coefficients with PFS-$26_{rs1501357}$ in PFS-26s of positive-effect alleles of EDU. Red part of each bar corresponds to the characteristic PFS-26s. (D) Correlation trend with PFS-$26_{ST8SIA2}$ in the characteristic PFS-26s of negative-effect alleles of NEU1. (E) Correlation trend with PFS-$26_{ST8SIA2}$ in the characteristic PFS-26s of positive-effect alleles of EDU. (F) Correlation trend with PFS-$26_{rs1501357}$ in the characteristic PFS-26s of positive-effect alleles of EDU.

References

1. Nielsen, R., Akey, J.M., Jakobsson, M., Pritchard, J.K., Tishkoff, S., Willerslev, E. Tracing the peopling of the world through genomics. Nature. 2017; 541(7637):302-10. https://doi.org/10.1038/nature21347 PMID: 28102248.
2. Henrich, J. The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter. Princeton, NJ: Princeton University Press; 2016.
3. Fujito, N.T., Satta, Y., Hane, M., Matsui, A., Yashima, K., Kitajima, K., Sato, C., Takahata, N., Hayakawa, T. Positive selection on schizophrenia-associated ST8SIA2 gene in post-glacial Asia. PLoS One. 2018; 13(7):e0200278. https://doi.org/10.1371/journal.pone.0200278 PMID: 30044798.
4. Hayakawa, T., Terahara, M., Fujito, N.T., Matsunaga, T., Teshima, K.M., Hane, M., Kitajima, K., Sato, C., Takahata, N., Satta, Y. Lower promoter activity of the ST8SIA2 gene has been favored in evolving human collective brains. PLoS One. 2021; 16(12): e0259897. https://doi.org/10.1371/journal.pone.0259897 PMID: 34914745.
5. Sato, C., & Hane, M. Mental disorders and an acidic glycan-from the perspective of polysialic acid (PSA/polySia) and the synthesizing enzyme, ST8SIA2. Glycoconj J. 2018; 35(4): 353-373. https://doi.org/10.1007/s10719-018-9832-9 PMID: 30058042.
6. van Os, J., Kenis, G., Rutten, B.P. The environment and schizophrenia. Nature. 2010; 468(7321):203-212. https://doi.org/10.1038/nature09563 PMID: 21068828.
7. Howes, O.D., McCutcheon, R., Owen, M.J., Murray, R.M. The Role of Genes, Stress, and Dopamine in the Development of Schizophrenia. Biol Psychiatry. 2017; 81(1):9-20. https://doi.org/10.1016/j.biopsych.2016.07.014 PMID: 27720198.
8. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015; 526(7571): 68-74. https://doi.org/10.1038/nature15393. PMID: 26432245.
9. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014; 511(7510): 421-7. https://doi.org/10.1038/nature13595 PMID: 25056061.
10. Trubetskoy, V., Pardiñas, A.F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T.B., Bryois, J., Chen, C.Y., Dennison, C.A., Hall, L.S., Lam, M., et al. Mapping genomic loci

implicates genes and synaptic biology in schizophrenia. Nature. 2022; 604(7906): 502-508. https://doi.org/10.1038/s41586-022-04434-5. PMID: 35396580.

11. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491(7422): 56-65. https://doi,org/10.1038/nature11632. PMID: 23128226.
12. Tishkoff, S.A., Reed, F.A., Ranciaro, A., Voight, B.F., Babbitt, C.C., Silverman, J.S., Powell, K., Mortensen, H.M., Hirbo, J.B., Osman, M., Ibrahim, M., et al. Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet. 2007; 39(1): 31-40. https://doi.org/10.1038/ng1946. PMID: 17159977.
13. Fujimoto, A., Kimura, R., Ohashi, J., Omi, K., Yuliwulandari, R., Batubara, L., Mustofa, M.S., Samakkarn, U., Settheetham-Ishida, W., Ishida, T., Morishita, Y., et al. A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. Hum Mol Genet. 2008; 17(6): 835-43. https://doi.org/10.1093/hmg/ddm355. PMID: 18065779.
14. Sturm., R.A., and Duffy, D.L. Human pigmentation genes under environmental selection. Genome Biol. 2012; 13(9): 248. https://doi.org/10.1186/gb-2012-13-9-248. PMID: 23110848.
15. Hill, W.D., Marioni, R.E., Maghzian, O., Ritchie, S.J., Hagenaars, S.P., McIntosh, A.M., Gale, C.R., Davies, G., Deary, I.J. A combined analysis of genetically correlated traits identified 187 loci and a role for neurogenesis and myelination in intelligence. Mol Psychiatry. 2019; 24(2): 169-181. https://doi.org/10.1038/s41380-017-0001-5. PMID: 29326435.
16. Goldberg, L.R. The structure of phenotypic personality traits. Am Psychol. 1993;48(1): 26-34. https://doi.org/10.1037//0003-066x.48.1.26. PMID: 8427480.
17. Nagel, M., Jansen, P.R., Stringer, S., Watanabe, K., de Leeuw, C.A., Bryois, J., Savage, J.E., Hammerschlag, A.R., Skene, N.G., Nuñoz-Manchado, A.B.; 23andMe Research Team; White, T., et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. Nat Genet. 2018; 50(7): 920-927. https://doi.org/10.1038/s41588-018-0151-7. PMID: 29942085.
18. Luciano, M., Hagenaars, S.P., Davies, G., Hill, W.D., Clarke, T.K., Shirali, M., Harris,

S.E., Marioni, R.E., Liewald, D.C., Fawns-Ritchie, C. Adams, M.J., et al. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. Nat Genet. 2018; 50(1): 6-11. https://doi.org/10.1038/s41588-017-0013-8. PMID: 29255261.

19. Rodriguez, V., Alameda, L., Trotta, G., Spinazzola, E., Marino, P., Matheson, S.L., Laurens, K.R., Murray, R.M., Vassos, E. Environmental risk factors in bipolar disorder and psychotic depression: a systematic review and meta-analysis of prospective studies. Schizophr Bull. 2021; 47(4): 959-974. https://doi.org/10.1093/schbul/sbaa197. PMID: 33479726.
20. Slavich, G.M., & Irwin, M.R. From stress to inflammation and major depressive disorder: a social signal transduction theory of depression. Psychol Bull. 2014; 140(3): 774-815. https://doi.org/10.1037/a0035302. PMID: 24417575.
21. Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S.M., Sidorenko, J., Kweon, H., Goldman, G., Gjorgjieva, T., Jiang, Y., et al. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. Nat Genet. 2022; 54(4): 437-449. https://doi.org/10.1038/s41588-022-01016-z. PMID: 35361970.
22. Savage, J.E., Jansen, P.R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C.A., Nagel, M., Awasthi, S., Barr, P.B., Coleman, J.R.I., Grasby, K.L., et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat Genet. 2018; 50(7): 912-919. https://doi.org/10.1038/s41588-018-0152-6. PMID: 29942086.
23. Als, T.D., Kurki, M.I., Grove, J., Voloudakis, G., Therrien, K., Tasanko, E., Nielsen, T.T., Naamanka, J., Veerapen, K., Levey, D.F., Bendl, J., et al. Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses. Nat Med. 2023; 29(7): 1832-1844. https://doi.org/10.1038/s41591-023-02352-1. PMID: 37464041.
24. Jones, S.E., Lane, J.M., Wood, A.R., van Hees, V.T., Tyrrell, J., Beaumont, R.N., Jeffries, A.R., Dashti, H.S., Hillsdon, M., Ruth, K.S., Tuke, M.A., et al. Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms.

Nat Commun. 2019; 10(1): 343. https://doi.org/10.1038/s41467-018-08259-7. PMID: 30696823.

25. Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., Visscher, P.M., GIANT Consortium. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. Hum Mol Genet. 2018; 27(20): 3641-3649. https://doi.org/10.1093/hmg/ddy271. PMID: 30124842.
26. Yengo, L., Vedantam, S., Marouli, E., Sidorenko, J., Bartell, E., Sakaue, S., Graff, M., Eliasen, A.U., Jiang, Y., Raghavan, S., Miao, J., et al. A saturated map of common genetic variants associated with human height. Nature. 2022; 610(7933): 704-712. https://doi.org/10.1038/s41586-022-05275-y. PMID: 36224396.
27. Revez, J.A., Lin, T., Qiano, Z., Xue, A., Holtz, Y., Zhu, Z., Zeng, J., Wang, H., Sidorenko, J., Kemper, K.E., Vinkhuyzen, A.A.E., et al. Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration. Nat Commun. 2020; 11(1): 1647. https://doi.org/10.1038/s41467-020-15421-7. PMID: 32242144.
28. Xue, A., Wu, Y., Zhu, Z., Zhang, F., Kemper, K.E., Zheng, Z., Yengo, L., Lloyd-Jones, L.R., Sidorenko, J., Wu. Y.; eQTLGen Consortium; McRae, A.F., et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nat Commun. 2018; 9(1): 2941. https://doi.org/10.1038/s41467-018-04951-w. PMID: 30054458.
29. Hautakangas, H., Winsvold, B.S., Ruotsalainen, S.E., Bjornsdottir, G., Harder, A.V.E., Kogelman, L.J.A., Thomas, L.F., Noordam, R., Benner, C., Gormley, P., Artto, V., et al. Genome-wide analysis of 102,084 migraine cases identified 123 risk loci and subtype-specific risk alleles. Nat Genet. 2022; 54(2): 152-160. https://doi.org/10.1038/s41588-021-00990-0. PMID: 35115687.
30. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, Madden, P.A.F., Heath, A.C., Martin, N.G., Montgomery, G.W., Weedon, M.N., et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat

Genet. 2012; 44(4): 369-375. https://doi.org/10.1038/ng2213. PMID: 22426310.

31. Genetics of Personality Consortium; de Moor, M.H., et al. Meta-analysis of genome-wide association studies for neuroticism, and the polygenic association with major depressive disorder. JAMA Psychiatry. 2015; 72(7): 642-650. https://doi.org/10.1001/jamapsychiatry.2015.0554. PMID: 25993607.
32. Okbay, A., Baselmans, B.M., De Neve, J.E., Turley, P., Nivard, M.G., Fontana, M.A., Meddens, S.F., Linnér, R.K., Rietveld, C.A., Derringer, J., Gratten, J., et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nat Genet. 2016; 48(6): 624-633. https://doi.org/10.1038/ng.3552. PMID: 27089181.
33. Smith, D.J., Escott-Price, V., Davies, G., Bailey, M.E., Colodro-Conde, L., Ward, J., Vedernikov, A., Marioni, R., Cullen, B. Lyall, D., et al. Genome-wide analysis of over 106000 individuals identifies 9 neuroticism-associated loci. Mol Psychiatry. 2016; 21(6): 749-757. https://doi.org/10.1038/mp.2016.49. PMID: 27067015.
34. Lo, M.T., Hinds, D.A., Tung, J.Y., Franz, C., Fan, C.C., Wang, Y., Smeland, O.B., Schork, A., Holland, D., Kauppi, K., Sanyal, N., et al. Genome-wide analyses for personality traits identify six genomic loci and show correlation with psychiatric disorders. Nat Genet. 2017; 49(1): 152-156. https://doi,org/10.1038/ng.3736. PMID: 27918536.
35. Sniekers, S., Stringer, S., Watanabe, K., Jansen, P.R., Coleman, J.R.I., Krapohl, E., Taskesen, E., Hammerschlag, A.R., Okbay, A, Zabaneh, D., Amin, N., et al. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. Nat Genet. 2017; 49(7): 1107-1112. https://doi.org/10.1038/ng.3869. PMID: 28530673.
36. Li, Z., Chen, J., Yu, H., He, L., Xu, Y., Zhang, D., Yi, Q., Li, C., Li, X., Shen, J., Song, Z., et al. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. Nat Genet. 2017; 49(11): 1576-1583. https://doi.org/10.1038/ng.3973. PMID: 28991256.
37. Lam, M., Chen, C.Y., Li, Z., Martin, A.R., Bryois, J., Ma, X., Gaspar, H., Ikeda, M., Benyamin, B., Brown, B.C., Liu, R., et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. Nat Genet. 2019; 51(12): 1670-

1678. https://doi.org/10.1038/s41588-019-0512-x. PMID: 31740837.

38. Kooner, J.S., Saleheen, D., Sim, X., Sehmi, J., Zhang, W., Frossard, P., Been, L.F., Chia, K.S., Dimas, A.S., Hassanali, N., Jafar, T., et al. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet. 2011; 43(10): 984-989. https://doi,org/10.1038/ng.921. PMID: 21874001.
39. Arai, M., Yamada, K., Toyota, T., Obata, N., Haga, S., Yoshida, Y., Nakamura, K., Minabe, Y., Ujike, H., Sora, I. et al. Association between polymorphisms in the promoter region of the sialyltransferase 8B (SIAT8B) gene and schizophrenia. Biol Psychiatry. 2006; 59(7):652-9. https://doi.org/10.1016/j.biopsych.2005.08.016 PMID: 16229822.
40. Tao, R., Li, C., Zheng, Y., Qin, W., Zhang, J., Li, X., Xu, Y., Shi, Y.Y., Feng, G., He, L. Positive association between SIAT8B and schizophrenia in the Chinese Han population. Schizophr Res. 2007; 90(1-3):108-14. https://doi.org/10.1016/j.schres.2006.09.029 PMID: 17126533.
41. Gilabert-Juan, J., Nacher, J., Sanjuán, J., Molto, M.D. Sex-specific association of the ST8SIAII gene with schizophrenia in a Spanish population. Psychiatry Res. 2013; 210(3):1293-5. https://doi.org/10.1016/j.psychres.2013.09.001 PMID: 24070986.
42. Fujito, N.T., Satta, Y., Hayakawa, T., Takahata, N. A new inference method for detecting an ongoing selective sweep. Genes Genet Syst. 2018; 93(4): 149-161. https://doi.org/10.1266/ggs.18-00008. PMID: 30270233.
43. Satta, Y., Zheng, 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. 2020; 94(6): 283-300. https://doi.org/10.1266/ggs.19-00012. PMID: 31827009.