American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 147B:463–466 (2008)

Haplotype Evolution of SLITRK1, a Candidate Gene for Gilles de la Tourette Syndrome

William C. Speed,[1] Brian J. O'Roak,[1] Zsanett Tárnok,[2] Csaba Barta,[3] Andrew J. Pakstis,[1] Matthew W. State,[1] and Kenneth K. Kidd[1]*

[1]*Department of Genetics, Yale University School of Medicine, New Haven, Connecticut*
[2]*Vadaskert Child and Adolescent Psychiatric Clinic, Budapest, Hungary*
[3]*Institute of Medical Chemistry, Molecular Biology and Pathobiochemistry, Semmelweis University, Budapest, Hungary*

Gilles de la Tourette syndrome (GTS) is a complex disorder with a clear genetic component but no clearly identified genes with variation of etiologic relevance. Various candidate regions and genes show some evidence of affecting risk, though clearly not all patients/families can be explained by any one of them. Resequencing one candidate gene, SLITRK1, has identified four new variants. Including them, we have typed over 2,300 normal individuals from 44 populations for 11 SNPs spanning the gene. The unusual global pattern seen is that one non-ancestral haplotype is the single most common haplotype worldwide. Other haplotypes appear to result from accumulation of mutations with no evidence of historical recombination. Although there is no evidence of selection, the haplotype frequency variation seen around the world will need to be considered in any future association studies of this locus with GTS or any other neuropsychiatric disorder.

KEY WORDS: Tourette syndrome; SLITRK1; haplotype; evolution

Please cite this article as follows: Speed WC, O'Roak BJ, Tárnok Z, Barta C, Pakstis AJ, State MW, Kidd KK. 2008. Haplotype Evolution of SLITRK1, a Candidate Gene for Gilles de la Tourette Syndrome. Am J Med Genet Part B 147B:463–466.

INTRODUCTION

Gilles de la Tourette syndrome (GTS) is a neuropsychiatric disease found in populations worldwide, although accurate estimates of regional disease prevalence have been affected by inconsistency among diagnostic criteria [Singer, 2005]. Attempts to map the disease by linkage have been unable to come up with a gene of major effect [see The Tourette Syndrome Association, 2007 for review]. However, by studying a GTS patient with a de novo chromosome 13 inversion, the gene SLITRK1 was implicated as one cause of GTS [Abelson et al., 2005]. SLITRK1 is predominantly expressed in the cerebral cortex [Aruga et al., 2003] and induces neurite outgrowth [Aruga and Mikoshiba, 2003], making it a strong candidate as a gene of major effect. More recent work [Züchner et al., 2006] has implicated SLITRK1 mutations in trichotillomania, which, like GTS, has obsessive-compulsive features.

Haplotype analysis is a powerful way to analyze the effect of putative causative SNPs in genes of interest. In previous studies of COMT [Palmatier et al., 2004], we found the haplotype associated with schizophrenia by Shifman et al. [2002] to have a limited distribution around the world with obvious implications to schizophrenia studies in other populations. Our extended COMT haplotype that included the P2 promoter SNP pointed to the relevance to schizophrenia of that promoter. In order to allow researchers to accurately dissect the haplotypes at SLITRK1, using appropriate regional frequencies, we undertook to identify as many SNPs as possible at SLITRK1 and to determine the global pattern of haplotype frequencies. While dbSNP details eight SNPs within the SLITRK1 gene, only three SNPs were confirmed with frequency information, and no detailed examination of haplotypes and their frequencies has previously been conducted.

SUBJECTS AND METHODS

Samples

For resequencing, 24 individuals affected with GTS were selected from families and sib-pair trios from Canada, Michigan, Oregon, Germany, and Hungary. Individuals were selected from divergent branches within the larger families to broaden the genetic variability of the resequencing sample. The North American samples have been studied in our lab previously [Zhang et al., 2002; Paschou et al., 2004] and the pedigrees have been described in detail elsewhere [Kurlan et al., 1986; Pauls et al., 1990; Pakstis et al., 1991]. Polymorphisms found by resequencing were studied in samples of 44 populations to calculate allele and haplotype frequencies. These populations include 10 African (Biaka, Mbuti, Yoruba, Ibo, Hausa, Chagga, Masai, Sandawe, Ethiopian Jews, and African Americans); 3 Southwest Asian (Yemenite Jews, Druze, Samaritans); 10 European (Adygei, Chuvash, Vologda Russians, Archangel Russians, Ashkenazi Jews, Finns, Hungarians, Danes, Irish, and European Americans); 2 Northwest Asian (Komi Zyriane, Khanty), 8 East Asian (Chinese from San Francisco, Taiwan Han Chinese, Hakka, Koreans, Japanese, Ami, Atayal, Cambodians), 1 Siberian (Yakut), 2 from Pacific Islands (Nasioi Melanesians, Miconesians), 4 North American (Cheyenne, Pima from Arizona, Pima from Mexico, Maya), and 4 South American (Quechua, Ticuna, Rondonia Surui, Karitiana). Sample descriptions and sample sizes can be found in ALFRED (http://alfred.med.yale.edu) starting from the UIDs provided in Table I.

Grant sponsor: National Institute of Health; Grant numbers: GM57672, AA09379, NS056276.

*Correspondence to: Dr. Kenneth K. Kidd, Ph.D., Department of Genetics, Yale University School of Medicine, P.O. Box 208005, 333 Cedar Street, SHM I-353, New Haven, CT 06520-8005.
E-mail: kenneth.kidd@yale.edu

Received 23 March 2007; Accepted 4 September 2007
DOI 10.1002/ajmg.b.30641

TABLE I. Polymorphisms at SLITRK1

Site #	dbSNP	TaqMan ID	chr 13 position	Ancestral	Avg. HET	Fst	ALFRED UID
1	rs9531519	C___1797115_10	83,347,193	C	0.270	0.078	SI003944U
2	ss68095339*	E____218516_10	83,349,098	A	0.274	0.081	SI003902O
3	rs3164	C__26349388_10	83,349,510	T	0.352	0.086	SI004033K
4	ss68095337*	E217535V351_10	83,350,052	T	0.030	0.044	SI004030H
5	rs9602286	E_rs9602286_10	83,350,393	G	0.079	0.103	SI003901N
6	rs3737193	C__27509799_10	83,350,419	A	0.106	0.128	SI003899D
7	rs9593836	C__30133648_20	83,350,752	T	0.088	0.121	SI003900M
8	rs9546538	C___1797119_10	83,354,736	C	0.427	0.038	SI003897B
9	ss68095334*	E_5UTR2SNP1_10	83,355,043	A	0.031	0.039	SI004031I
10	ss68095332*	E_5UTR1SNP2_10	83,355,434	A	0.011	0.040	SI004032J
11	rs9546539	C___1797120_10	83,355,495	C	0.359	0.089	SI003898C

The numeric values for allele frequencies at each of the sites can be found under the site UID in ALFRED (http://alfred.med.yale.edu). Positions are based on UCSC sequence build #hg17. The average heterozygosity and F_{st} are based on the 44 populations listed in the text. The newly identified SNPs are identified with an asterisk.

Resequencing

Affected individuals were selected for two reasons. First, the hope was to find causative SNPs in the coding and/or regulatory regions of the gene. Alternatively, informative SNPs in the families could be identified for linkage studies. Secondly, if this putative disease gene has undergone negative selective pressure, the linked variation on the disease-carrying chromosome may be at lower frequency within the population, and we wished to enrich our search pool for such variation.

Long range PCR was performed using a mixture of Taq, Pfu, and Pwo enzymes with the primers PROM1F and 3′UTR7R [supplementary data, Abelson et al., 2005] on 24 affected individuals. This 7 kb PCR includes the entire promoter and 3′ untranslated region of the gene. Bidirectional sequencing was performed using the amplification primers and also internal primers 5′UTR1R, CDS3R, CDS7R, 3′UTR2R, 3′UTR3R, 3′UTR5R, 3′UTR7F [supplementary data, Abelson et al., 2005]. Automated fluorescent sequencing performed on an ABI 3730xl by W.M. Keck Foundation Biotechnology Resource Laboratory showed no allelic bias in the long-range amplification, based on heterozygote peak heights in the electropherograms.

Marker Typing

Polymorphisms discovered by resequencing were then typed on a set of globally representative populations by the TaqMan™ (Applied Biosystems) method [Livak, 1999]. We determined the ancestral states of the SNPs by using the same TaqMan assays to genotype genomic DNA for nonhuman primates—three chimpanzees (*Pan troglodytes*), three gibbons (*Hylobates*), three gorillas (*Gorilla gorilla*), three orangutans (*Pongo pygmaeus*), and three bonobos (*Pan paniscus*).

Analyses

Genotype and allele frequencies for each individual site were calculated by direct gene counting, assuming biallelic codominant inheritance. The Hardy–Weinberg test was executed by an auxiliary program, FENGEN, which also creates the input file for the program HAPLO [Hawley and Kidd, 1995] from raw data records. Maximum-likelihood estimates of haplotype frequencies were calculated from the individual multi-site phenotypes of individuals in each population using the program HAPLO [Hawley and Kidd, 1995].

RESULTS

Four new and seven previously reported SNPs were identified by resequencing at the SLITRK1 locus (Table I). All 2,326 individuals in the 44 population samples were typed for these 11 SNPs. The SNPs spanned 8.3 kb. SNPs 8-11 are in the 1.8 kb upstream of the SLITRK1 initiation codon; SNPs 1–7 are in the 4.3 kb downstream from the termination codon. None was found in the single exon. No marker showed any significant deviation from Hardy–Weinberg in any of the populations. Twelve haplotypes were found with frequency greater than 5% in at least one population; all but haplotypes #7, #10, and #11 (cf. Fig. 2) were directly observed, that is, occurred in individuals with a genotype of zero or only one heterozygous SNP. Frequencies of the seven most frequent haplotypes are graphed in Figure 1. These 12 haplotypes account for 94.6–100% of the chromosomes in all populations. The ancestral haplotype (Fig. 2), inferred from genotyping non-human primates, was not seen in any population. The globally most common haplotype differs from the ancestral haplotype by two single nucleotide polymorphisms (rs3164 and rs9546538) and is the most common haplotype in every population, ranging in frequency from 38.9% to 68.0%. The haplotype and individual site frequencies for these populations are available online at http://alfred.med.yale.edu/alfred/recordinfo.asp?condition=loci.locus_uid='lo001745r.

All of the common haplotypes can be explained by accumulation of variation on the ancestral haplotype with no recombination within the 8.3 kb region (Fig. 2). Most of the directly observed haplotypes can be ordered in pairs differing by one derived nucleotide change; in one case two changes are required (asterisk (*) in Fig. 2). Collectively, these unambiguously generate two chains, each starting with a separate single derived nucleotide change from the ancestral haplotype: ancestor to #8, #8 to both #1 & #9, #9 to #4, #4 to #5; and separately ancestor to #6, #6 to #3, #6 with two steps to #2, #2 to #11. Any other evolutionary arrangement of these haplotypes would require more mutations, including back mutations. The inferred haplotype #7 differs by only one derived nucleotide substitution from haplotype #6. While not directly observed, haplotype #7 has estimated frequencies of up to 13% in over a dozen populations. The tree in Figure 2 requires two instances of recurrent "mutation": rs3737193 mutating from A to G to produce haplotypes #3 and #12; ss68095332 mutating from A to C to produce haplotypes 10 and 11. In both cases one of the two haplotypes is seen in several populations and definitely present (haplotypes #8 and #11) and the other is inferred to be present in only one isolated population but at frequencies unlikely to be errors of inference (#12 at 6.5% in Nasioi; #10 at 8.5% in Mbuti). Thus, rare double crossovers ("gene conversions") or recurrent mutations could have occurred and drifted to detectable frequencies in those populations. These rare haplotypes fit into the tree differing from a common definitely present haplotype by only the recurrent "mutation"; any other location would require additional mutations.

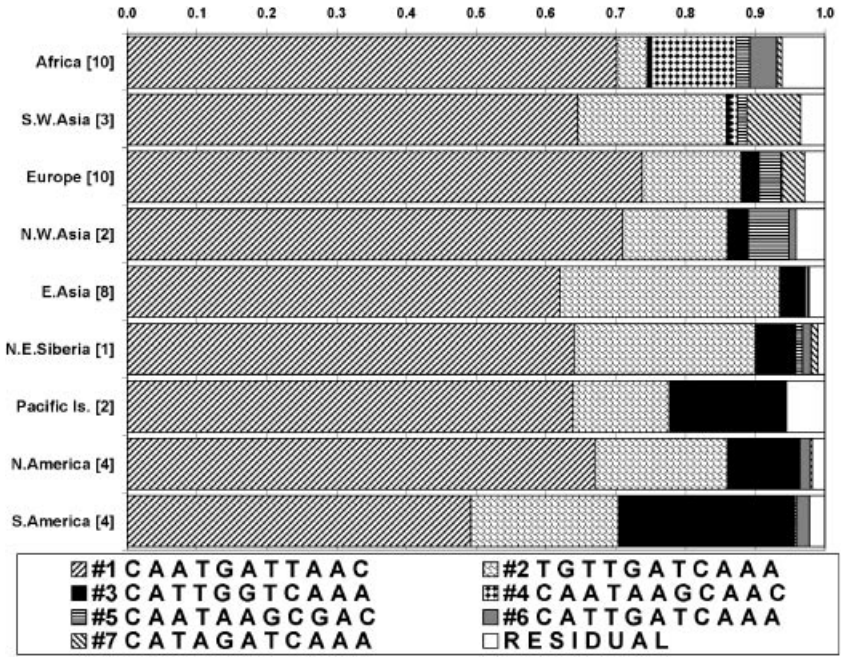

Fig. 1. SLITRK1 haplotype frequencies based on 44 populations for 11 SNPs spanning 8.3 kb region. The seven most common haplotypes are displayed as averages for the populations in each of the major geographic regions. The other haplotype frequencies with non-zero frequency estimates are combined into the residual class. Numeric values for these seven haplotypes in each of the 44 populations are available in ALFRED along with the frequencies of the five additional haplotypes that were estimated to occur with a frequency of at least 5% in at least one population (see Fig. 2) but are included in the residual of very rare haplotypes in this figure.

It is somewhat rare to find a single, non-ancestral haplotype that spans 8.3 kb at such high worldwide frequency, given that five of the SNPs have heterozygosities greater than 0.25. HapMap data for this region shows a linkage disequilibrium 'block' encompassing these SNPs in their Caucasian (CEU) sample, but haplotype diversity in their Asian (CHB + JPT) and African (YRI) samples. Based on the HapMap, we would have expected more haplotype diversity in our non-European populations. Low haplotype diversity can be one indication of a recent selective sweep or continued selective pressure at a locus. In looking at the SNP data, F_{st} values ranged from 0.038 to 0.128; there is no indication in any population of an increase in F_{st} accompanied by a decrease in heterozygosity, indications of a selective sweep. While the 8.3 kb region is too short to look for long range effects on haplotype homozygosity [by using EHH or REHH; Sabeti et al., 2002], values of haplosimilarity for each SNP in each population ranged from +7.8 to −22.0, not varying outside the normal range seen across chromosome 22 [Hanchard et al., 2006].

Combining these observations, it is unlikely that the worldwide, high frequency, non-ancestral haplotype at SLITRK1 has been the target of positive selection; instead, its frequency is likely due to drift. Among the remaining haplotypes

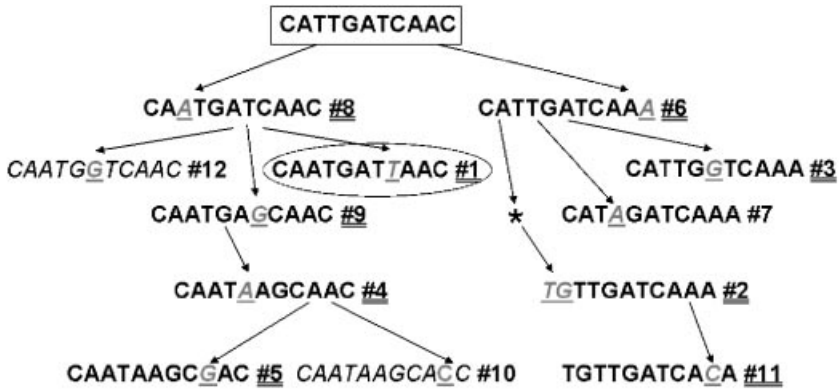

Fig. 2. Proposed evolution of SLITRK1 haplotypes. The 12 haplotypes shown are those that occur with a frequency of at least 5% in at least one population. The boxed haplotype is the ancestral state as determined by non-human primate genotypes. The globally most common haplotype is circled. Haplotypes are numbered as in Figure 1 except for numbers 8 through 12 that occur at an average frequency too low to be seen in the graphic. Double underlined rank numbers occur for haplotypes actually observed (see text). According to this schema, the underlined nucleotide is the new mutation distinguishing the new haplotype from the previous haplotype. Italicized haplotypes are those likely to involve recurrent mutation. Haplotypes 8 through 12 are seen only at low frequency in any region (<5%) and would not be visible in Figure 1. Since the order of the two mutations separating haplotypes 6 and 2 cannot be inferred from the rare inferences of the two possibilities, the asterisk represents the intermediate haplotype.

observed, considerable diversity in frequency is seen, especially between African populations and the rest of the world. Such variability should be taken into account in further studies of the association of this locus with GTS.

ELECTRONIC DATABASES CITED

ALFRED http://alfred.med.yale.edu.
dbSNP http://www.ncbi.nlm.nih.gov/snp/.
UCSC http://genome.ucsc.edu/cgi-bin/hggateway.

ACKNOWLEDGMENTS

This work was funded, in part, by National Institute of Health grants GM57672 and AA09379 to Kenneth K. Kidd and NS056276 to Matthew W. State. We thank the many colleagues who helped us assemble the population samples. Special thanks are due the many hundreds of individuals from these populations who volunteered to give blood samples for studies such as this.

REFERENCES

Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM, Mathews CA, Pauls DL, Rasin MR, Gunel M, Davis NR, Ercan-Sencicek AG, Guez DH, Spertus JA, Leckman JF, Dure LS IV, Kurlan R, Singer HS, Gilbert DL, Farhi A, Louvi A, Lifton RP, Sestan N, State MW. 2005. Sequence variants in SLITRK1 are associated with Tourette's syndrome. Science 310(5746):317–320.

Aruga J, Mikoshiba K. 2003. Identification and characterization of Slitrk, a novel neuronal transmembrane protein family controlling neurite outgrowth. MCN 24(1):117–129.

Aruga J, Yokota N, Mikoshiba K. 2003. Human SLITRK family genes: Genomic organization and expression profiling in normal brain and brain tumor tissue. Gene 315(2):87–94.

Hanchard NA, Rockett KA, Spencer C, Coop G, Pinder M, Jallow M, Kimber M, McVean G, Mott R, Kwiatkowski DP. 2006. Screening for recently selected alleles by analysis of human haplotype similarity. Am J Hum Genet 78(1):153–159.

Hawley ME, Kidd KK. 1995. HAPLO: A program using the EM algorithm to estimate the frequencies of multi-site haplotypes. J Hered 86:409–411.

Kurlan R, Behr J, Medved L, Shoulson I, Pauls D, Kidd JR, Kidd KK. 1986. Familial Tourette's syndrome: Report of a large pedigree and potential for linkage analysis. Neurology 36:772–776.

Livak KJ. 1999. Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. Genet Anal 14:143–149.

Pakstis AJ, Heutink P, Pauls DL, Kurlan R, van de Wetering BJ, Leckman JF, Sandkuyl LA, Kidd JR, Breedveld GJ, Castiglione CM, Weber J, Sparkes RS, Cohen DJ, Kidd KK, Oostra B. 1991. Progress in the search for genetic linkage with Tourette syndrome: An exclusion map covering more than 50% of the autosomal genome. Am J Hum Genet 48:281–294.

Palmatier MA, Pakstis AJ, Speed W, Paschou P, Goldman D, Odunsi A, Okonofua F, Kajuna S, Karoma N, Kungulilo S, Grigorenko E, Zhukova OV, Bonne-Tamir B, Lu RB, Parnas J, Kidd JR, DeMille MM, Kidd KK. 2004. COMT haplotypes suggest P2 promoter region relevance for schizophrenia. Mol Psychiatry 9(9):859–870.

Paschou P, Feng Y, Pakstis AJ, Speed WC, DeMille MM, Kidd JR, Jaghori B, Kurlan R, Pauls DL, Sandor P, Barr CL, Kidd KK. 2004. Indications of linkage and association of Gilles de la Tourette syndrome in two independent family samples: 17q25 is a putative susceptibility region. Am J Hum Genet 75(4):545–560.

Pauls DL, Pakstis AJ, Kurlan R, Kidd KK, Leckman JF, Cohen DJ, Kidd JR, Como P, Sparkes R. 1990. Segregation and linkage analyses of Tourette's syndrome and related disorders. J Am Acad Child Adolesc Psychiatry 29:195–203.

Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–837.

Shifman S, Bronstein M, Sternfeld M, Pisante-Shalom A, Lev-Lehman E, Weizman A, et al. 2002. A highly significant association between a COMT haplotype and schizophrenia. Am J Hum Genet 71:1296–1302.

Singer HS. 2005. Tourette's syndrome: From behaviour to biology. Lancet Neurol 4(3):149–159.

The Tourette Syndrome Association International Consortium for Genetics. 2007. Genome Scan for Tourette Disorder in Affected-Sibling-Pair and Multigenerational Families. Am J Hum Genet 80:265–272.

Zhang H, Leckman JF, Pauls DL, Tsai C-P, Kidd KK, Campos MR, Tourette Syndrome Association International Consortium for Genetics. 2002. Genomewide scan of hoarding in sib pairs in which both sibs have Gilles de la Tourette syndrome. Am J Hum Genet 70:896–904.

Züchner S, Cuccaro ML, Tran-Viet KN, Cope H, Krishnan RR, Pericak-Vance MA, Wright HH, Ashley-Koch A. 2006. SLITRK1 mutations in Trichotillomania. Mol Psychiatry 11:888–891.