Selection in the dopamine receptor 2 gene: A candidate SNP study

Tobias Göllner, Martin Fieder

Dopamine is a major neurotransmitter in the human brain and is associated with various diseases. Schizophrenia, for example, is treated by blocking the dopamine receptors type 2. In 2009, Shaner, Miller and Mintz stated that schizophrenia was the low fitness variant of a highly variable mental trait. We therefore explore whether the dopamine receptor 2 gene (DRD2) underwent any selection processes. We acquired genotype data of the 1000 Genomes project (phase I), which contains 1093 individuals from 14 populations. We included single nucleotide polymorphisms (SNPs) with two minor allele frequencies (MAFs) in the analysis: MAF over 0.05 and over 0.01. This is equivalent to 151 SNPs (MAF > 0.05) and 246 SNPs (MAF > 0.01) for DRD2. We used two different approaches (an outlier approach and a Bayesian approach) to detect loci under selection. The combined results of both approaches yielded nine (MAF > 0.05) and three candidate SNPs (MAF > 0.01), under balancing selection. We also found weak signs for directional selection on DRD2, but in our opinion these were too weak to draw any final conclusions on directional selection in DRD2. All candidates for balancing selection are in the intronic region of the gene and only one (rs12574471) has been mentioned in the literature. Two of our candidate SNPs are located in specific regions of the gene: rs80215768 lies within a promoter flanking region and rs74751335 lies within a transcription factor binding site. We strongly encourage research on our candidate SNPs and their possible effects.
Selection in the dopamine receptor 2 gene: A candidate SNP study.

Tobias Göllner, Department of Anthropology, University of Vienna, Vienna, Austria.

Martin Fieder, Department of Anthropology, University of Vienna, Vienna, Austria.

ABSTRACT:

Dopamine is a major neurotransmitter in the human brain and is associated with various diseases. Schizophrenia, for example, is treated by blocking the dopamine receptors type 2. In 2009, Shaner, Miller and Mintz stated that schizophrenia was the low fitness variant of a highly variable mental trait. We therefore explore whether the dopamine receptor 2 gene (DRD2) underwent any selection processes. We acquired genotype data of the 1000 Genomes project (phase I), which contains 1093 individuals from 14 populations. We included single nucleotide polymorphisms (SNPs) with two minor allele frequencies (MAFs) in the analysis: MAF over 0.05 and over 0.01. This is equivalent to 151 SNPs (MAF > 0.05) and 246 SNPs (MAF > 0.01) for DRD2. We used two different approaches (an outlier approach and a Bayesian approach) to detect loci under selection. The combined results of both approaches yielded nine (MAF > 0.05) and three candidate SNPs (MAF > 0.01), under balancing selection. We also found weak signs for directional selection on DRD2, but in our opinion these were too weak to draw any final conclusions on directional selection in DRD2. All candidates for balancing selection are in the intronic region of the gene and only one (rs12574471) has been mentioned in the literature. Two of our candidate SNPs are located in specific regions of the gene: rs80215768 lies within a promoter flanking region and rs74751335 lies within a transcription factor binding site. We strongly encourage research on our candidate SNPs and their possible effects.
INTRODUCTION:

The catecholamine dopamine is a neurotransmitter in the human brain. Dopaminergic neurons can be divided into four major pathways: nigrostriatal, mesolimbic, mesocortical and tuberoinfundibular (Andén et al., 1964; Dahlström and Fuxe, 1964). These neurons play an important role in voluntary movement, feeding, reward and learning, as well as certain other functions. Outside the brain, dopamine takes on a physiological role in cardiovascular functions, hormonal regulation, renal and other functions (Snyder et al., 1970; Missale et al., 1998; Sibley, 1999; Carlsson, 2001; Iversen and Iversen, 2007). Due to this involvement in many different processes and systems, dopamine is also related to a variety of diseases. Parkinson’s disease, caused by a loss of dopaminergic innervations in the striatum, is a prominent example (Ehringer and Hornykiewicz, 1960). Additionally, the expected associations between the dopaminergic system and schizophrenia stem from the fact that various dopamine receptor 2 blockers are used as antipsychotics in treating that condition (Snyder et al., 1970; Creese et al., 1976; Seeman et al., 1976; Carlsson et al., 2001). Further relationships with dopamine dysregulation are expected in Tourette’s syndrome and attention deficit hyperactivity disorder (ADHD) (Mink, 2006; Swanson et al., 2007; Gizer et al., 2009). The strong involvement of dopamine in the reward system suggests an association with drug abuse and addiction (Hyman et al., 2006; Di Chiara and Bassareo, 2007; Koob and Volkow, 2010). Many more diseases and conditions are expected to involve dopamine dysfunctions. (As reviewed by Beaulieu & Gainetdinov, 2011)

In humans, five different dopamine receptors exist. They are classified into two categories based on their structure and their pharmacological and biochemical properties. The D1-class includes the dopamine receptors 1 and 5, while the D2-class consists of the dopamine receptors 2, 3 and 4 (Andersen et al., 1990; Niznik and Van Tol, 1992; Sibley and Monsma, 1992; Sokoloff et al., 1992a; Civelli et al., 1993; Vallone et al., 2000). The focus of our study is on the dopamine receptor 2 and its gene DRD2. The dopamine receptor 2 gene lies on the long arm of chromosome 11 (11q23.1). It spans from 113,280,317 to 113,346,413 for a total of 66,096 base pairs (bp) (information accessed on NCBI in the GnRH37 assembly). For the gene card, see Figure 1 in Results. DRD2 has six introns (Gingrich and Caron, 1993). Alternative splicing between intron 4 and 5 of an 87 bp exon generates two variants of the dopamine receptor 2. The difference between D2S (short) and D2L (long) is a 29-amino-acids-long chain in the third
intercellular loop of the protein (Giros et al., 1989; Monsma et al., 1989). While the short form 
(D2S) is mainly expressed at the presynapase, the long form (D2L) is expressed postsynaptically 
(Usiello et al., 2000; De Mei et al., 2009). The D2S are mainly autoreceptors, i.e. they reduce the 
expression of dopamine when activated. This leads to an important negative feedback 
mechanism (Wolf and Roth, 1990; Missale et al., 1998; Sibley, 1999). (Again, as reviewed by 
Beaulieu & Gainetdinov, 2011)

Among the many single nucleotide polymorphisms (SNPs) of DRD2, one prominent example is 
rs6277, also known as C957T. It has been associated with schizophrenia in Han Chinese in 
Taiwan (Glatt et al., 2009), in Russians (Monakhov et al. 2008) and in Bulgarians (Betcheva et 
al. 2009). Together with the -141C allele, the 957T allele is associated with the diagnosis of 
anorexia nervosa (Bergen et al., 2005). A meta-analysis showed that the Ser311Cys 
polymorphism (rs1801028) in DRD2 is a risk factor for schizophrenia. The heterozygotes 
(Ser/Cys) and the homozygotes for Cys were both at elevated risk for schizophrenia when 
compared to the Ser/Ser genotypes (Glatt and Jönsson, 2006). In a study with alcoholic patients 
and controls, the A allele of rs1076560 was more frequent in alcoholic patients (Sasabe et al., 
2007). In 2012, Mileva-Seitz et al. conducted a study with Caucasian mothers and their infants. 
They taped mother-infant behaviour and genotyped various SNPs of DRD2 and also DRD1. 
Rs1799732 and the previously mentioned rs6277 were both associated with direct vocalization of 
the mother towards the infant.

The body of literature on SNPs and their possible effects is growing rapidly. Considering the 
influences those SNPs could have on human behaviour, and bearing in mind the different 
ecological habitats of Homo sapiens, we explore if DRD2 underwent any selection processes. In 
2009 an interesting proposal by Shaner, Miller and Mintz stated that schizophrenia was the low 
fitness variant of a highly variable mental trait. Based on the connection between dopamine 
receptor 2 and schizophrenia, as stated above, we focused our analysis on DRD2.

To reduce false-positives, we used two selection detection algorithms to explore DRD2. This 
exploratory (“hypothesis-free”) approach is desugned to find candidate SNPs that were under 
selection. The data basis of our analysis is the 1000 Genomes Project samples.
MATERIAL AND METHODS:

We acquired data from the 1000 Genomes Project (phase I) through SPSmart engine v5.1.1 (http://spsmart.cesga.es/engines.php; Amigo et al., 2008), using the search term “DRD2”. We included all single nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) greater than 0.05 (N= 151 SNPs) to include only the more frequently occurring SNPs. To verify our results also on the basis of a higher number of SNPs (which occur less frequently), we conducted the same analysis also based on a MAF > 0.01 sample (N= 246 SNPs; data presented in the supplementary material). The structure of the DRD2 gene (113,280,317 – 113,346,413 in the GnRH37.p13 primary assembly; gene card) is shown in the results section (Figure 1). In total we included the following populations in our analysis:

The data were converted by hand into the CONVERT format. All further format conversions were performed by PGD Spider 2.0.5.2 (Lischer and Excoffier, 2012).

Two different programs were used to detect selection. Both use F_{ST} approaches to detect outliers. The program LOSITAN calculates FDIST, which uses F_{ST} and the expected heterozygosity. It assumes an island model of migration with neutral markers. An expected distribution of Wright’s inbreeding coefficient is calculated and then outliers are identified. A neutral mean F_{ST} was computed by the program before the 50,000 simulations were performed. The infinite alleles model was used. To avoid false positive detection we set the significance level to p < 0.01 (P(Simulation F_{ST} < sample F_{ST})). (Antao et al., 2008)

BayeScan is a Bayesian statistics program. Basically it calculates two simulations for every locus: one in which it assumes the locus is under selection and the other one in which this assumption is dropped. It splits the F_{ST} coefficient into two parts. The alpha value is a locus-specific component shared by all populations. The beta value is a population-specific component shared by all loci. This is achieved via logistic regression and provides insight into selection. The alpha value serves as an indicator for selection. Significant positive values of alpha indicate directional selection, whereas significant negative values indicate balancing selection. The posterior probabilities are estimated using a reversible-jump Markov Chain Monte Carlo (MCMC) approach. The posterior probabilities are gained by counting how many times alpha is included in the model. Before computing the Markov chains, we calculate 20 pilot runs with 5000 iterations each. The initial burn-in is set to 50,000 steps and the chains are run with 5000
iterations and a thinning interval of 10. The program output consists of a posterior probability, the logarithm (base 10) of the posterior odds and a q value. These three values are all for the model with selection. Furthermore, the alpha value is reported along with an $F_{ST}$ coefficient average of all population per locus. In BayeScan the threshold of a posterior P of > 0.99 and a log10(PO) of 2 or higher is used. This threshold is labelled as “Decisive” by BayeScan (see the program manual at [http://cmpg.unibe.ch/software/BayeScan/files/BayeScan2.1_manual.pdf](http://cmpg.unibe.ch/software/BayeScan/files/BayeScan2.1_manual.pdf)). (Foll and Gaggiotti, 2008)

To compute linkage disequilibrium (LD) of the SNPs, we used the R “genetics package” ([http://cran.r-project.org/web/packages/genetics/genetics.pdf](http://cran.r-project.org/web/packages/genetics/genetics.pdf); Warnes et al., 2013). Mueller (2004) atates that $D'$ is particularly useful to assess the probability for historical recombination in a given population and $r^2$ is useful in the context of association studies. We therefore primarily calculate $D''$, but we also calculated $r^2$, which is presented in the supplementary material.

In most populations one or more SNPs had to be excluded to successfully run the computation. The population IBS was excluded entirely from this computation. IBS is a very small population ($n = 14$), and 30 SNPs caused the computation to fail. For a detailed view on all excluded SNPs see Table S1 in the supplementary material.

We accessed information on the gene via NCBI ([http://www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)) and on the specific SNPs via Ensembl ([http://www.ensembl.org/](http://www.ensembl.org/)).
RESULTS:

The combined results of LOSITAN and BayeScan yielded nine candidate SNPs under balancing selection (MAF > 0.05):

For a detailed view on the results of LOSITAN and BayeScan for all SNPs, see Table S2 in the supplementary material. Figure 1 provides a gene view of DRD2 with labels for the candidate SNPs.

The same calculations based on the MAF 0.01 sample revealed only 2 SNPs (rs60599314, rs79549222) under balancing selection by both LOSITAN and Bayescan (Figure S1, Table S3).

Three SNPs (rs6277, rs12800853, rs11608109) that do not reach significance in the MAF > 0.05 sample (Table S3a) do reach significance in the MAF > 0.01 sample (Table S3b), for directional selection. They barely reach significance (P < 0.01) based on the MAF > 0.01 sample in LOSITAN (Figure S1, Table S3b). However if we applied more stringent detection prerequisites (“force mean $F_{ST}$” and “neutral mean $F_{ST}$”; increasing computational load, but also increasing convergence and lowering the bias in $F_{ST}$ estimation) in LOSITAN, none of these three SNPs reaches significance (Figure S2). The results for balancing selection in BayeScan remained nearly unchanged in the MAF > 0.01 sample, with the exception of rs12574471, which did not reach significance (Table S3).

All nine SNPs detected based on MAF > 0.05 are intron variants (Figure 1). Only rs12574471 (3) is mentioned in the literature because it is near a supposed recombination hotspot (Glatt et al, 2009). rs80215768 (4) lies within a promoter flanking region; rs74751335 (7) lies within a transcription factor binding site. Nonetheless, we found no known associations for those two SNPs.

The $F_{ST}$ values of these nine loci indicate an overall low genetic differentiation, as well as a low differentiation between populations (Table 2). This is in accordance with balancing selection acting on the gene. The differences in $F_{ST}$ values stem from different algorithms used by the programs.
The Linkage Disequilibrium measurements $D'$ and $r^2$ were used. The heat maps for all nine populations are shown in the supplementary material (Figure S3 – S15 for $D'$ and Figure S16 – S28 for $r^2$). The relative position of the marked SNPs change because different populations had different SNPs excluded (see Table S1 for the list).
DISCUSSION:

We found nine SNPs to be candidates for balancing selection based on the sample MAF > 0.05; of those, two had been also detected under balancing selection based on the MAF > 0.01 sample. We found no SNPs based on the MAF > 0.05 sample and the MAF > 0.01 sample, under directional selection, that are detected by both algorithms (on P < 0.001 in LOSITAN) and if more stringent detection criteria were applied in LOSITAN. We therefore conclude that, if directional selection has been acting on DRD2, then the signs are rather weak, i.e. too weak to make definitive conclusions.

Checking all nine SNPs under balancing selection based on the MAF > 0.05 sample with Ensembl reveals that they are all intronic region variants. For rs60599314 (1) and rs79549222 (2) that are found by LOSITAN and BayeScan on both the MAF > 0.05 sample and the MAF > 0.01 sample, no particular additional information is known. We therefore suggest that these two SNPs may provide interesting candidates for future functional studies.

Rs80215768 (4) lies within a promoter flanking region and rs74751335 (7) lies within a transcription factor binding site (TFBS) (both SNPs detected based on the MAF > 0.05 sample). Many studies are available on the possible effects of mutations in such regions (Hayashi, Watanabe and Kawajiri, 1991; and In et al., 1997; or for a more general review on the topic, Jaenisch & Bird, 2003). Nonetheless, the SNPs show low Fst values, which is congruent with the finding of balancing selection. Sewall Wright’s guidelines for interpreting Fst values suggest little genetic differentiation in our populations (as cited by Jobling, 2013; Chapter 5, Box 5.2).

As silent mutations in DRD2 are known to alter the mRNA stability and even the synthesis of the receptor itself (Duan et al., 2003), we call for exploring the possible effects of these SNPs.

Additionally, the levels of the linkage disequilibrium measurement D’ are typical for the respective populations: African populations show a dispersed pattern and no clear LD blocks (Figure S3, S11, and S15). While the LD blocks are visible in American populations (Figure S7, S12 and S13), they are not as clear as in Asian (Figure S5, S6 and S10) or European populations (Figure S4, S8, S9 and S14). Looking up the SNPs that are in high D’ (> 0.8) with our nine candidate loci revealed no new information as to the nature of those loci. We also examined the measurement r² for all populations (again, excluded SNPs are listed in Table S1), which revealed no new insight.
The finding of balancing selection suggests that in our sample the minor alleles bear some fitness
disadvantage. Note that some individuals are homozygous for the minor allele (0.8 – 2.3% per
SNP, over all populations). Fitness is altered if survival or reproduction of an organism is
affected. This raises the possibility of a connection between our candidate SNPs and diseases or
malfunctions of dopamine receptor 2. In the list of diseases associated with dopamine (see
Introduction) the most striking example is schizophrenia because dopamine receptor 2 blockers
can successfully treat patients.

Albeit we aim to avoid overhasty conclusions regarding directional selection on DRD2, the three
SNPs detected by BayeScan under positive selection – and that are found by LOSITAN just
below significance – could be of interest. This is particularly the case for rs6277, with its known
phenotypic associations: Rs6277 has been associated with schizophrenia in Han Chinese in
Taiwan (Glatt et al., 2009), in Russians (Monakhov et al. 2008) and in Bulgarians (Betcheva et
al. 2009).

Nonetheless, rs6277 was not identified among the 108 schizophrenia-associated loci that have
recently been published based on 36,989 cases and 113,075 controls by the Schizophrenia
Working Group of the Psychiatry Genomics Consortium (Schizophrenia Working Group of the
Psychiatry Genomics Consortium, 2014). Nevertheless, DRD2 is one of the genes that was
confirmed as being associated with schizophrenia by the Working Group. Thus, even weak signs
of directional selection on DRD2 might be important to i) identify potential new disease-related
phenotypical associations and ii) to speculate on what the “selective force” could have been
bringing mutations on DRD2 towards fixations and iii) what potential condition-related
consequences selection on DRD2 could have. The question is whether these conditions affect
fitness. Accordingly, Bassett et al. (1996) showed that reproductive fitness is reduced in groups
of familial schizophrenia, which suggests a selection process. Puzzlingly enough, they also found
some evidence for an increased fitness of a small subsample of sisters. Shaner, Miller and Mintz
in 2009 proposed that schizophrenia is the low-fitness trait of a highly variable mental trait. They
argue that the persistence of the illness at about 1% globally is too high for new mutations. Thus,
mainly balancing selection would fit this hypothesis very well, and our candidate SNPs under
balancing selection could be viable indicators for this.
223 DRD2 is clearly associated with schizophrenia. Schizophrenia, however, is a “polygenic
224 condition” including genetic loci over the whole human genome (The Schizophrenia Working
225 Group of the Psychiatry Genomics Consortium, 2014). Accordingly, the importance of DRD2
226 should not be over-estimated. Moreover, the method of selection detection does not allow direct
227 inferences about a phenotype (e.g. schizophrenia). Assumptions about the phenotype can only be
228 made based on our genotype data, but our overall results can serve as a valuable precursor to
229 future studies on the subject.

230 To untangle the possible effects of our SNPs, we propose a study in which our candidate SNPs
231 are investigated in schizophrenic and non-schizophrenic persons. A simple comparison of the
232 SNPs and the different haplotypes between the two groups should efficiently help assess our
233 findings. If this proposed study finds differences in those two groups, then the mechanisms of
234 those SNPs and their possible haplotypes must be investigated.
CONCLUSION:

We found nine candidates for balancing selection on *DRD2* but only a weak signs for directional selection. Interestingly, rs6277, a SNP known to be associated with schizophrenia, is among those SNPs for which we found weak evidence for directional selection. Some of the SNPs under balancing selection are potentially associated with various diseases. These SNPs could be important as biomarkers due to their very low F_{ST} values: the genetic differentiation of one population compared with the whole sample is very small. While all candidate SNPs may be worth exploring, we definitely recommend using rs60599314 and rs79549222 (as these were detected under balancing selection based on a MAF > 0.05 and a MAF > 0.01 sample). We also recommend rs80215768 and rs74751335 found under balancing selection for further studies on *DRD2* because the former within a promoter flanking region and the latter lies in a transcription factor binding site.
ACKNOWLEDGEMENTS:

The authors would like to thank the R Core Team for the statistical computing environment, the 1000 Genomes Project, as well as the creators of LOSITAN, BAYESCAN and PGD Spider. The University of Vienna for the “Investitionsprojekt IP 547011“. Additional thanks go to Bernard Wallner, Philipp Gewessler, Matthias Hirschmanner and Michael Stachowitsch. Furthermore we thank the reviewers for their valuable comments that helped to improve the manuscript.
REFERENCES:

Amigo, J., Salas, A., Phillips, C. & Carracedo, Á. (2008). SPSmart: adapting population based SNP genotype databases for fast and comprehensive web access. *BMC bioinformatics*, 9(1), 428.

Andén, N. E., Carlsson, A., Dahlström, A., Fuxe, K., Hillarp, N. Å., & Larsson, K. (1964). Demonstration and mapping out of nigro-neostriatal dopamine neurons. *Life sciences*, 3(6), 523-530.

Andersen, P. H., Gingrich, J. A., Bates, M. D., Deary, A., Falardeau, P., Senogles, S. E., & Caron, M. G. (1990). Dopamine receptor subtypes: beyond the D 1/D 2 classification. *Trends in pharmacological sciences*, 11(6), 231-236.

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC bioinformatics*, 9(1), 323.

Bassett, A. S., Bury, A., Hodgkinson, K. A., & Honer, W. G. (1996). Reproductive fitness in familial schizophrenia. *Schizophrenia research*, 21(3), 151-160.

Bergen, A. W., Yeager, M., Welch, R. A., Haque, K., Ganjei, J. K., van den Bree, M. B., ... & Kaye, W. H. (2005). Association of multiple DRD2 polymorphisms with anorexia nervosa. *Neuropsychopharmacology*, 30(9), 1703-1710.

Carlsson, A. (2001). A paradigm shift in brain research. *Science*, 294(5544), 1021-1024.

Carlsson, A., Waters, N., Holm-Waters, S., Tedroff, J., Nilsson, M., & Carlsson, M. L. (2001). Interactions between monoamines, glutamate, and GABA in schizophrenia: new evidence. *Annual review of pharmacology and toxicology*, 41(1), 237-260.

Civelli, O., Bunzow, J. R., & Grandy, D. K. (1993). Molecular diversity of the dopamine receptors. *Annual review of pharmacology and toxicology*, 33(1), 281-307.

Creese, I., Burt, D. R., & Snyder, S. H. (1976). Dopamine receptor binding predicts clinical and pharmacological potencies of antischizophrenic drugs. *Science*, 192(4238), 481-483.
Dahlström, A., & Fuxe, K. (1964). Evidence for the existence of monoamine-containing neurons in the central nervous system. I. Demonstration of monoamines in the cell bodies of brain stem neurons. *Acta Physiologica Scandinavica. Supplementum, SUPPL-232.*

De Mei, C., Ramos, M., Iitaka, C., & Borrelli, E. (2009). Getting specialized: presynaptic and postsynaptic dopamine D2 receptors. *Current opinion in pharmacology, 9*(1), 53-58.

Di Chiara, G., & Bassareo, V. (2007). Reward system and addiction: what dopamine does and doesn’t do. *Current opinion in pharmacology, 7*(1), 69-76.

Duan, J., Wainwright, M. S., Comeron, J. M., Saitou, N., Sanders, A. R., Gelernter, J., & Gejman, P. V. (2003). Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Human molecular genetics, 12*(3), 205-216.

Ehringer, H., & Hornykiewicz, O. (1960). Distribution of noradrenaline and dopamine (3-hydroxytyramine) in the human brain and their behavior in diseases of the extrapyramidal system. *Klinische Wochenschrift, 38,* 1236.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics, 180*(2), 977-993.

Gingrich, J. A., & Caron, M. G. (1993). Recent advances in the molecular biology of dopamine receptors. *Annual review of neuroscience, 16*(1), 299-321.

Giros, B., Sokoloff, P., Martres, M. P., Riou, J. F., Emorine, L. J., & Schwartz, J. C. (1989). Alternative splicing directs the expression of two D2 dopamine receptor isoforms. *Nature, 342,* 923 – 926.

Gizer, I. R., Ficks, C., & Waldman, I. D. (2009). Candidate gene studies of ADHD: a meta-analytic review. *Human genetics, 126*(1), 51-90.

Hayashi, S. I., Watanabe, J., & Kawajiri, K. (1991). Genetic polymorphisms in the 5′-flanking region change transcriptional regulation of the human cytochrome P450IIE1 gene. *Journal of biochemistry, 110*(4), 559-565.

Hyman, S. E., Malenka, R. C., & Nestler, E. J. (2006). Neural mechanisms of addiction: the role of reward-related learning and memory. *Annual review of neuroscience, 29,* 565-598.
In, K. H., Asano, K., Beier, D., Grobholz, J., Finn, P. W., Silverman, E. K., ... & Drazen, J. M. (1997). Naturally occurring mutations in the human 5-lipoxygenase gene promoter that modify transcription factor binding and reporter gene transcription. *Journal of clinical investigation*, 99(5), 1130.

Iversen, S. D., & Iversen, L. L. (2007). Dopamine: 50 years in perspective. *Trends in neurosciences*, 30(5), 188-193.

Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics*, 33, 245-254.

Jobling, M., Hollox, E., Hurles, M., Kivisild, T. & Tyler-Smith, C. (2013). *Human evolutionary genetics*. New York and London: Garland Science.

Koob, G. F., & Volkow, N. D. (2010). *Neurocircuitry of addiction. neuropsychopharmacology*, 35(1), 217-238.

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28(2), 298-299.

Mink, J. W. (2006). Neurobiology of basal ganglia and Tourette syndrome: basal ganglia circuits and thalamocortical outputs. *Advances in neurology*. New York: Raven Press, 99, 89.

Missale, C., Nash, S. R., Robinson, S. W., Jaber, M., & Caron, M. G. (1998). Dopamine receptors: from structure to function. *Physiological reviews*, 78(1), 189-225.

Monsma, F. J., McVittie, L. D., Gerfen, C. R., Mahan, L. C., & Sibley, D. R. (1989). Multiple D2 dopamine receptors produced by alternative RNA splicing. *Nature*, 342, 926 – 929.

Mueller, J. C. (2004). Linkage disequilibrium for different scales and applications. *Briefings in bioinformatics*, 5(4), 355-364.

Niznik, H. B., & Van Tol, H. H. (1992). Dopamine receptor genes: new tools for molecular psychiatry. *Journal of psychiatry and neuroscience*, 17(4), 158.

Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511(7510), 421-427.
Seeman, P., Lee, T., Chau-Wong, M., & Wong, K. (1976). Antipsychotic drug doses and neuroleptic/dopamine receptors. *Nature*, 261(5562), 717-719.

Shaner, A., Miller, G., & Mintz, J. (2004). Schizophrenia as one extreme of a sexually selected fitness indicator. *Schizophrenia research*, 70(1), 101-109.

Sibley, D. R. (1999). New insights into dopaminergic receptor function using antisense and genetically altered animals 1. *Annual review of pharmacology and toxicology*, 39(1), 313-341.

Sibley, D. R., & Monsma, F. J. (1992). Molecular biology of dopamine receptors. *Trends in pharmacological sciences*, 13, 61-69.

Snyder, S. H., Taylor, K. M., Coyle, J. T., & Meyerhoff, J. L. (1970). The role of brain dopamine in behavioral regulation and the actions of psychotropic drugs. *American journal of psychiatry*, 127(2), 199-207.

Sokoloff, P., Andrieux, M., Besançon, R., Pilon, C., Martres, M. P., Giros, B., & Schwartz, J. C. (1992). Pharmacology of human dopamine D 3 receptor expressed in a mammalian cell line: comparison with D 2 receptor. *European journal of pharmacology: molecular pharmacology*, 225(4), 331-337.

Swanson, J. M., Kinsbourne, M., Nigg, J., Lanphear, B., Stefanatos, G. A., Volkow, N., ... & Wadhwa, P. D. (2007). Etiologic subtypes of attention-deficit/hyperactivity disorder: brain imaging, molecular genetic and environmental factors and the dopamine hypothesis. *Neuropsychology review*, 17(1), 39-59.

Usiello, A., Baik, J. H., Rougé-Pont, F., Picetti, R., Dierich, A., LeMeur, M., ... & Borrelli, E. (2000). Distinct functions of the two isoforms of dopamine D2 receptors. *Nature*, 408(6809), 199-203.

Vallone, D., Picetti, R., & Borrelli, E. (2000). Structure and function of dopamine receptors. *Neuroscience & biobehavioral reviews*, 24(1), 125-132.

Warnes, G., Leisch, F., Man, M., & Warnes, M. G. Package ‘genetics’ (2013).

Wolf, M. E., & Roth, R. H. (1990). Autoreceptor regulation of dopamine synthesis. *Annals of the New York academy of sciences*, 604(1), 323-343.
**Table 1** (on next page)

Populations of the 1000 Genomes Project
| Superpopulation (Code) | Population Code | Population Description | Number of Individuals |
|-----------------------|-----------------|------------------------|----------------------|
| Africa (AFR)          | ASW             | African ancestry in Southwest USA | 61                   |
|                       | LWK             | Luhya in Webuye, Kenya | 97                   |
|                       | YRI             | Yoruba in Ibadan, Nigeria | 88                   |
| Europe (EUR)          | CEU             | Utah residents with Northern and Western European ancestry | 87                   |
|                       | FIN             | Finnish from Finland | 93                   |
|                       | GBR             | British from England and Scotland | 88                   |
|                       | IBS             | Iberian populations in Spain | 14                   |
|                       | TSI             | Toscani in Italy | 98                   |
| East Asia (ASN)       | CHB             | Han Chinese in Beijing, China | 97                   |
|                       | CHS             | Han Chinese South | 100                  |
|                       | JPT             | Japanese in Tokyo, Japan | 89                   |
|                       | CLM             | Colombians from Medellin, Colombia | 66                   |
| America (AMR)         | MXL             | Mexican ancestry from Los Angeles USA | 60                   |
|                       | PUR             | Puerto Ricans from Puerto Rico | 55                   |
|                       |                 | All populations = | 1093                 |
Table 2 (on next page)

The dopamine receptor 2 gene’s nine candidate SNPs for balancing selection (MAF > 0.05)
| Locus          | Major Allele (Frequency) | Minor Allele (Frequency) | $F_{ST}$ (Lositan) | $F_{ST}$ (BayeScan) | Location          |
|---------------|--------------------------|--------------------------|--------------------|---------------------|-------------------|
| rs60599314    | C (0.871)                | T (0.129)                | 0.0110             | 0.0272              | 113,306,431 (Intronic region) |
| rs79549222    | T (0.87)                 | G (0.13)                 | 0.0106             | 0.0260              | 113,310,340 (Intronic region) |
| rs12574471    | C (0.891)                | T (0.109)                | 0.0172             | 0.0364              | 113,316,236 (Intronic region) |
| rs80215768    | G (0.925)                | A (0.075)                | 0.0304             | 0.0328              | 113,318,880 (Intronic region) |
| rs76581995    | C (0.925)                | A (0.075)                | 0.0304             | 0.0328              | 113,319,835 (Intronic region) |
| rs80014933    | T (0.923)                | C (0.077)                | 0.0304             | 0.0332              | 113,328,135 (Intronic region) |
| rs74751335    | G (0.915)                | C (0.085)                | 0.0266             | 0.0322              | 113,328,810 (Intronic region) |
| rs77264605    | A (0.915)                | G (0.085)                | 0.0266             | 0.0327              | 113,328,913 (Intronic region) |
| rs76499333    | G (0.925)                | A (0.075)                | 0.0299             | 0.0327              | 113,329,449 (Intronic region) |
Graphical output of BayeScan

X-axis: log10(q values) for all SNPs, the threshold is -2 (see Material and Methods). Y-axis: $F_{ST}$-values, where high values indicate directional selection, low values balancing selection. Red dots: the nine SNPs in the MAF > 0.05 sample that both BayeScan and LOSITAN find. For exact results see Table S2.
Location of candidate SNPs under balancing selection in *DRD2*

E1-8 are exons 1 to 8. (1): rs60599314, (2): rs79549222, (3): rs12574471, (4): rs80215768, (5): rs76581995, (6): rs80014933, (7): rs74751335, (8): rs77264605, (9): rs76499333
Detection of outlier SNPs of the DRD2 gene using FDIST (LOSITAN)

X-axis: estimated heterozygosity values. Y-axis: $F_{ST}$-values. The red area indicates positive directional selection, the grey area neutrality, and the yellow area balancing selection. Confidence intervals represent borders between “selection areas”. See Table S2 for the exact results on the SNPs.