5′ and 3′ region variability in the dopamine transporter gene (SLC6A3), pesticide exposure and Parkinson’s disease risk: a hypothesis-generating study

Samir N.P. Kelada1,*, Harvey Checkoway1,2, Sharon L.R. Kardia4, Christopher S. Carlson5, Paola Costa-Mallen1, David L. Eaton1, Jordan Firestone1,3, Karen M. Powers1, Phillip D. Swanson3, Gary M. Franklin1,3, W.T. Longstreth Jr2,3, Terri-Smith Weller1, Zahra Afsharinejad1 and Lucio G. Costa1,6

1Department of Environmental and Occupational Health Sciences, 2Department of Epidemiology and 3Department of Neurology, University of Washington, Seattle, WA 98105, USA, 4Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI, USA, 5Fred Hutchinson Cancer Research Center, Seattle, WA, USA and 6Department of Human Anatomy, Pharmacology and Forensic Sciences, University of Parma Medical School, Parma, Italy

Received July 28, 2006; Revised August 28, 2006; Accepted September 5, 2006 GenBank accession no. DQ307031

The dopamine transporter gene (SLC6A3) is a candidate gene for Parkinson’s disease (PD) on the basis of its critical role in dopaminergic neurotransmission. Previously, we identified 22 SNPs in the 5′ region of SLC6A3, which segregate as eight haplotypes that differ in transcriptional activity when transfected in rat dopamine-producing cells. In the present work from a case–control study size of 293 cases and 395 controls, we employed a cladistic approach to examine gene–disease association. First, we found strong evidence of balancing selection in this region, as determined by a Tajima’s D statistic of 2.97 ($P < 0.001$). Second, we found that the eight haplotypes fit into two main clades and that diplotypes of these clades were marginally associated with PD. Then, after we classified cases and controls by the number of risk alleles, accounting for the well-known 3′ region VNTR polymorphism, we found that having two or more risk alleles resulted in a modest but significant increase in PD risk [odds ratio = 1.58; 95% confidence interval (CI): 1.03–2.40]. Finally, we detected a significant interaction between occupational pesticide exposure in men and the number of risk alleles. Among pesticide-exposed subjects, the odds ratio for having two or more risk alleles was 5.66 (95% CI: 1.73–18.53). Thus, allelic variants in SLC6A3, which affect gene expression, are associated with PD in this population and may interact with occupational pesticide exposure to increase PD risk.

INTRODUCTION

Parkinson’s disease (PD) is a movement disorder characterized by bradykinesia, cogwheel rigidity, reflex impairment and rest tremor (1), which result from the loss of dopaminergic neurons in the substantia nigra (2,3). One critical determinant of the dopaminergic signaling in these neurons is the dopamine transporter (DAT), which serves to terminate the synaptic signal by transporting dopamine back into the neuron from the synaptic cleft after its release (4–6). Once back in the neuron, dopamine can either be sequestered into vesicles by the vesicular monoamine transporter (VMAT2) or is subject to enzymatic or non-enzymatic oxidation, which generates reactive oxygen species (7–9). Reactive oxygen species are known to damage cellular proteins and induce apoptotic pathways which are thought to contribute to neuronal loss in PD (3,10,11).

*To whom correspondence should be addressed at: National Human Genome Research Institute, National Institutes of Health, Building 50, Room 3510, 50 South Drive MSC 8004, Bethesda, MD 20892, USA. Tel: +1 3014022270; Fax: +1 3014809667; Email: keladas@mail.nih.gov

© The Author 2006. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
Because of the important role of DAT in dopaminergic neuron function, the DAT gene, known as SLC6A3 (chromosome 5p15.3), is a strong biological candidate gene for PD (12). The coding region of SLC6A3 has been examined thoroughly by several groups. These studies, which have examined both European American and African American populations, have shown that there are only a small number of rare non-synonymous SNPs and a few synonymous SNPs of low-to-moderate allele frequency (13–16). Results from multiple population-based studies, have shown that there are only a small number of low-to-moderate allele frequency SNPs, indicating that there is selective pressure against changes to the coding region of this gene.

Non-coding regions of the gene, however, are more diverse. A well-known variable number of tandem repeat (VNTR) polymorphism exists in the 3′ untranslated region (UTR), consisting of 3–13 repeats, with alleles of 9- and 10-repeat being most common among several populations (6,17). Studies of this polymorphism and PD have revealed some associations, primarily with the rare 11-repeat allele (18–20), though conflicting data have also been reported (21,22). Recently, Lin et al. (23) reported that the common 10-repeat allele conferred reduced PD risk compared with the 9- and 11-repeat alleles.

The 5′ region variation has been a topic of recent interest as well. We expanded prior work by other groups (24–26) and identified 22 SNPs in the region spanning ∼5000 bp 5′ of exon 1 (transcriptional start site) through the start of exon 2 (where the translational start site lies). These 22 SNPs segregate as eight haplotypes, six of which are common (>5% frequency). The six common haplotypes differ in transcriptional activity, as determined by a reporter gene assay conducted in dopamine-producing rat PC12 cells (27). Most importantly, we observed that the two most common haplotypes (two and six) differed in activity by ~40%. Despite observed differences in transcriptional activity among 5′ region haplotypes, we did not detect associations between haplotypes and PD in a case–control study of 261 cases and 376 controls (27). More recently, Drgon et al. (28) also examined this region and again found a large haplotype block with two main haplotypes. They reported that the two main haplotypes categories differ in transcriptional activity, as in vivo human ventral striatal DAT levels and post-mortem tissue DAT levels differed by 5′ region haplotype.

Like many other chronic diseases, PD is thought to be multifactorial (3,29), with gene–environment interactions playing a key role. Among potential environmental risk factors for PD, pesticides have been examined extensively. Recent studies have shown that rotenone (30) and a combined paraquat–maneb (31) exposure can induce PD-like pathology and motor signs in rodents. Other studies have shown that organochlorine pesticides can alter dopaminergic neurotransmission in treated animals by increasing DAT expression (32,33). Epidemiological studies of pesticides and PD in humans have also been provocative. Several case–control studies have reported positive associations of PD with pesticide exposure (34–39), and a meta-analysis of 19 case–control studies reported a significant overall association [odds ratio (OR) = 1.9; 95% confidence interval (CI): 1.49–2.53] (40). Cohort studies conducted in the USA (41,42) and France (43), in which exposure was assessed prospectively, have also reported significant findings. So far, two case–control studies have reported genetic–environment interactions with pesticides and CYP2D6 (44) and GSTZ1 (45).

In the present work, using an expanded case–control study population, we employed a cladistic approach to understand the relationships among haplotypes and the association of the major 5′ region clades with PD. We then modeled PD as a function of the number of risk alleles, taking the 3′ region VNTR into account. Finally, we tested for interactions between occupational pesticide exposure and the number of SLC6A3 risk alleles.

RESULTS

We first examined sequence diversity in the 5′ region of SLC6A3 spanning back ~5000 bp of exon 1 through the start of exon 2 (+2106) with the sequence data from 24 subjects (27). With the sequence data obtained and using other GenBank sequences, a reference sequence (GenBank accession no. DQ307031) was assembled. In the ~7.5 kb of 5′ region sequence examined, 22 SNPs were found, most with high allele frequency. Average heterozygosity (θe = 1.27 × 10−2) in this region was considerably higher than nucleotide diversity (θs = 6.65 × 10−4), resulting in a large Tajima’s D statistic of 2.97 (P < 0.001). The excess of high-frequency polymorphism in this region was confirmed in a multiethnic set of 90 samples from the Polymorphism Discovery Resource (PDR90) sequenced by the Environmental Genome Project (http://egp.gs.washington.edu/). This result indicates population contraction, balancing selection or population admixture. Admixture seems unlikely, as results were consistent in our European American population and in the highly admixed PDR90, and population contraction would be inconsistent with data from the SeattleSNPs Program for Genomic Applications, where Tajima’s D across more than 5 mb in the European population is 0.34. Thus, our results strongly suggest long-term balancing selection acting on this region.

The 22 SNPs in this region form eight haplotypes. We constructed a phylogenetic tree to visualize the relations among haplotypes (Fig. 1). From this figure, two major clades can be seen: the top branch containing haplotypes one through four, designated clade A, and the bottom branch containing haplotypes five through eight, designated clade B. The relatively low nucleotide distance between the two major clades provides further evidence for balancing selection occurring in this region (46).

We also compared the SNPs and haplotypes analyzed here with those recently reported by Drgon et al. (28), who scanned ~18 kb of the 5′ region in 12 European American subjects. As shown in the Supplementary Material, Table S1, for the ~7.5 kb of 5′ region sequence that was examined in both studies, most SNPs were detected in both studies, with the exceptions being SNPs with low allele frequency. Similarly, there was considerable overlap in the haplotypes inferred from the genotype data, as shown in the Supplementary Material, Table S2. Of note, the two main clades identified here, A and B, correspond to the ‘TA’ and ‘CG’ haplotype groups, respectively, as reported by Drgon et al. (28).
Differences in the frequencies of the common haplotypes and clades are probably due in part to the relatively small sample size of the two study populations examined.

We then tested for associations between the major 5' region clades and PD. Since our last analysis (27), we have accrued 32 new cases and 21 more controls. All subjects from the prior analysis were included in this study. Demographic characteristics of cases and controls are shown in the Supplementary Material, Table S3. The distributions of 5' region diplotypes shown in Table 1, were not significantly different between cases and controls but did indicate a higher frequency of clade A in cases. Relative risk estimates (ORs), adjusted for age, gender and smoking status (ever/never) for the AB and AA diplotypes relative to the reference group (diplotype BB) were 1.14 (95% CI: 0.81–1.61) and 1.40 (95% CI: 0.90–2.20), respectively. Nested clade analysis did not reveal further branchpoints in the phylogenetic tree associated with PD but may have been underpowered. The distributions of common 3' region VNTR (9- or 10-repeat) genotypes (Table 1) were not significantly different between cases and controls but indicated a higher frequency of the 9-repeat allele in cases. Neither age nor gender appeared to modify this association.

We next examined the distribution of haplotypes comprising both the 5' (clade A or B) and 3' (VNTR 9- or 10-repeat allele) regions, as shown in Table 2. The B-10 haplotype frequency was higher in controls, and the A-9 haplotype was slightly higher in cases. Overall, the distribution of combined 5'-3' region haplotypes was not significantly different between cases and controls ($\chi^2 = 3.73$ (3 d.f.), $P = 0.29$).

Subsequently, we categorized cases and controls by the number of 5' and 3' risk alleles, where the 5' A clade and 3' VNTR 9-repeat alleles were considered the risk alleles. The best fit model referred to subjects as having zero, one or two or more (‘2+’) risk alleles. Using this scale, the $\chi^2$ value for the difference in the distribution between cases and controls was 4.47 (2 d.f.), with a $P$-value of 0.11. The age- and gender-adjusted ORs for subjects having 1 or 2+ risk alleles were 1.16 (95% CI: 0.75–1.78) and 1.58 (95% CI: 1.03–2.40), respectively. The association appears to be stronger in older subjects, as the ORs for 1 or 2+ risk alleles among subjects $\geq 60$ years of age were 1.24 (95% CI: 0.75–2.07) and 1.68 (95% CI: 1.03–2.73), respectively, compared with 0.92 (95% CI: 0.39–2.17) and 1.33 (95% CI: 0.55–3.17) among subjects $< 60$ years of age. Additionally, the ORs for having 2+ risk alleles was somewhat stronger in women versus men [OR = 1.84 (95% CI: 0.91–3.71) in women versus 1.46 (95% CI: 0.86–2.47) in men].

Finally, interactions with occupational pesticide exposure were assessed using stratified analysis and logistic regression models. Male subjects were classified as exposed if they

---

**Table 1.** 5' and 3' region diplotypes among PD cases and controls

<table>
<thead>
<tr>
<th>Region</th>
<th>Cases (%)</th>
<th>Controls (%)</th>
<th>$\chi^2$</th>
<th>OR (95% CI)$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' Clade</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diplotype</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BB</td>
<td>98 (33.4)</td>
<td>145 (36.7)</td>
<td>1.0 (ref.)</td>
<td></td>
</tr>
<tr>
<td>BA</td>
<td>137 (46.8)</td>
<td>188 (47.6)</td>
<td>2.15, $P = 0.34$</td>
<td>1.14 (0.81–1.61)</td>
</tr>
<tr>
<td>AA</td>
<td>58 (19.8)</td>
<td>62 (15.7)</td>
<td>1.40 (0.90–2.20)</td>
<td></td>
</tr>
<tr>
<td>3' VNTR Genotype</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genotype</td>
<td>10/10</td>
<td>9/10</td>
<td>9/9</td>
<td>10/10</td>
</tr>
<tr>
<td>10/10</td>
<td>148 (50.5)</td>
<td>224 (56.7)</td>
<td>2.66, $P = 0.27$</td>
<td>1.31 (0.95–1.82)</td>
</tr>
<tr>
<td>9/10</td>
<td>121 (41.3)</td>
<td>141 (35.7)</td>
<td>1.0 (ref.)</td>
<td></td>
</tr>
<tr>
<td>9/9</td>
<td>24 (8.2)</td>
<td>30 (7.6)</td>
<td>1.33 (0.74–2.39)</td>
<td></td>
</tr>
</tbody>
</table>

$^a$Adjusted for age, gender and smoking status (ever/never).

---

**Table 2.** Combined 5' and 3' region haplotype frequencies among PD cases and controls

<table>
<thead>
<tr>
<th>5'-3' haplotype</th>
<th>Cases Frequency (%)</th>
<th>Controls Frequency (%)</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-10</td>
<td>244 (41.6)</td>
<td>368 (46.6)</td>
<td>3.73, $P = 0.29$</td>
</tr>
<tr>
<td>B-9</td>
<td>89 (15.2)</td>
<td>110 (13.9)</td>
<td></td>
</tr>
<tr>
<td>A-10</td>
<td>173 (29.5)</td>
<td>221 (28.0)</td>
<td></td>
</tr>
<tr>
<td>A-9</td>
<td>80 (13.7)</td>
<td>91 (11.5)</td>
<td></td>
</tr>
</tbody>
</table>

---

Figure 1. A UPGMA model of haplotype phylogeny. The x-axis represents the number of differences between haplotypes, with the bar showing a difference of two nucleotides (nts). Haplotypes were resolved using all 22 SNPs in the region, but only the tagSNPs used to differentiate them in the case–control study are shown (positions relative to reference sequence DQ307031). Italicized haplotypes (5 and 7) were not separately tagged because of low frequency.
Our results indicate that nucleotide diversity in the 5′ region of the DAT gene is quite high and that balancing selection is occurring on this region. This stands in strong contrast to purifying (or negative) selection occurring on the coding region of the gene (14). Linkage disequilibrium in the 5′ region may span further 5′ but declines somewhere 3′ of exon 2. There is another haplotype block in the distal 3′ region encompassing the VNTR, creating a total of at least three major haplotype blocks in the gene (47). The balancing selection that was detected presumably acts either on the 5′ region of SLC6A3 or on sequence further 5′, which contains other genes. Additional data are needed to elucidate which regions are associated with the selective effects. We were able to gain support from examination of SLC6A3 genotypes from another population which again showed the high intra-population variability indicative of balancing selection.

These results suggest that the two main 5′ region clades have differential effects on gene expression, which may have served to maintain them in the population, an idea which was put forth by Greenwood and Kelsoe (25) and which is supported by the very high Tajima’s D test statistic and the branch lengths of the phylogenetic tree we found here. Two in vitro studies and one in vivo imaging study provide further evidence that the two common 5′ region haplotypes differ in transcriptional activity (27,28,48). The 3′ region VNTR polymorphism has repeatedly been associated with attention-deficit hyperactivity disorder (49), and some have hypothesized that hyperactivity may have conferred a selective advantage in prior times and environments where resources were scarce (the ‘environmental mismatch’ hypothesis) (50), thus providing a rationale for selective pressure on the DAT gene. Related findings for the dopamine receptor D4 gene (DRD4) have also been reported (51).

Results from our case–control study indicate weak individual associations between the 5′ and 3′ regions with PD. However, when we examined the association between PD and SLC6A3 using the combined number of risk alleles as the predictor, the findings were stronger and significant. The genotype classification system used here produced results which suggest that disease risk is a function of a threshold number of SLC6A3 allelic variants (two) beyond which disease risk rises, although not very sharply.

In terms of genotype–expression phenotype correlations, the 5′ region B clade appears to be associated with higher striatal DAT expression levels than the A clade, as determined in an in vivo human imaging study (28). The A and B clades described in this report correspond to the TA and CG haplotypes, respectively, as reported by Drgon et al. (28); it is also interesting to note that the results from the in vivo study contrast with results from in vitro studies conducted in rat PC12 cells (27) and human neuroblastoma cells (48). Conflicting data regarding the 3′ VNTR–gene expression relation have been reported (25,52–59), perhaps in part because of a lack of consideration of the 5′ region; the weight of evidence, though, is in favor of the hypothesis that the 10-repeat allele of the VNTR yields higher expression than the 9-repeat allele. Thus, at the current time and given the results of our study, the data suggests that lower DAT expression (i.e. the 5′ A clade and/or the 3′ VNTR 9-repeat allele) is associated with higher PD risk. The mechanism by which lower DAT expression increases PD risk remains to be determined. Possible mechanisms include altered dopamine disposition due to lower re-uptake by DAT and/or quantitatively different interactions with α-synuclein (60,61). It should also be noted that other regions of SLC6A3 appear to modulate transcriptional activity as well (e.g. intron 8) (47,62), and hence should be considered in future studies.

Our findings also imply that the common disease–common variant hypothesis (63) applies to PD, at least to some extent. Additionally, the finding of an interaction between SLC6A3 and occupational pesticide exposure further supports the idea that PD is a multifactorial disease caused by interactive effects between genes and environment. The interaction also provides additional support for the hypothesis that pesticide exposures are associated with PD risk among a genetically

---

Table 3. Interactions with occupational pesticide exposure among men

<table>
<thead>
<tr>
<th>Number of risk alleles</th>
<th>Exposed Cases (%)</th>
<th>Controls (%)</th>
<th>OR (95% CI)</th>
<th>Not exposed Cases (%)</th>
<th>Controls (%)</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7 (14.9)</td>
<td>17 (30.9)</td>
<td>1.0 (ref)</td>
<td>23 (17.6)</td>
<td>37 (20.1)</td>
<td>1.0 (ref)</td>
</tr>
<tr>
<td>1</td>
<td>14 (29.8)</td>
<td>23 (41.8)</td>
<td>1.63 (0.52–5.15)</td>
<td>49 (37.4)</td>
<td>60 (32.6)</td>
<td>1.21 (0.62–2.36)</td>
</tr>
<tr>
<td>2+</td>
<td>26 (55.3)</td>
<td>15 (27.3)</td>
<td>5.66 (1.73–18.53)</td>
<td>59 (45.0)</td>
<td>87 (47.3)</td>
<td>1.17 (0.62–2.23)</td>
</tr>
</tbody>
</table>

χ² 8.73, P = 0.01

0.85, P = 0.65

---

*Adjusted for age (<60, ≥60), education (quintiles) and smoking status (ever/never).
susceptible subgroup. Most current hypotheses about pesticides and PD state that pesticides may perturb mitochondrial function (e.g. through complex I inhibition) and increase the already high level of oxidative stress in dopaminergic neurons (64,65). Either subsequent to or independent of the oxidative stress, pesticides also create cellular conditions that favor the formation of α-synuclein fibrils (66) and impair proteosomal function (65). These phenomena are not mutually exclusive, so pathways of oxidative stress and proteosome dysfunction may interact, exacerbating each other’s effects (10).

However, we did not find that risk of PD among pesticide-exposed subjects was associated with alleles that confer high expression, as one would expect on the basis of whole animal and cell culture models (67). Those studies have shown that higher DAT expression confers susceptibility to neurotoxicants that are substrates for DAT, most notably MPP⁰. Organochlorine and pyrethroid pesticides can increase DAT expression in vivo and hence alter dopaminergic signaling and the effect of MPP⁰ exposure (33,68,69). In contrast, our results suggest that the deleterious effects of occupational pesticide exposure among men in this study and DAT low expression risk alleles are caused by independent cellular mechanisms. The results also suggest that both risk factors are required for disease risk to increase. Further toxicological and epidemiological studies which differentiate between specific pesticide classes with known toxic mechanisms are needed to better understand this; additionally, more information on which pesticides are substrates for DAT would also be helpful. Though DAT is often considered a gateway for neurotoxicants (67), to date, no pesticides [including paraquat and rotenone (70)] have conclusively been shown to enter dopaminergic neurons through it, and thus it appears that some pesticides exert their toxic effects to dopaminergic neurons by a mechanism that is not dependent on entering the neuron through DAT (70).

We acknowledge the limitations of our study. An important caveat of our findings is the possibility of false-positive results owing to chance. The unadjusted P-value for the χ² distribution test of risk alleles among pesticide-exposed cases and controls, the main finding of interest, was 0.01. Correcting this estimate for the number of hypothesis tests performed (i.e. Bonferroni correction with seven tests) increases the P-value to 0.07. However, several of the tests performed were not independent. For example, the χ² distribution tests of the combined 5′–3′ haplotype frequencies among cases and controls and the number of risk alleles among cases and controls are clearly related. Similarly, the likelihood ratio test performed to assess the significance of the gene–environment interaction is not independent of the two χ² distribution tests performed in the pesticide exposure-stratified analysis. There is no currently accepted method to account for the non-independence of these tests. We suspect that true P-value lies somewhere in between the values of 0.01 and 0.07. Replication of these findings with pesticides, especially with specific pesticide classes, will be needed before confident conclusions can be reached.

It should also be noted that a recent whole-genome association study of PD did not find association with SLC6A3, although one SLC6A3 tagSNP was marginally associated in one of the two phases of that study (71). Additionally, pesticide exposure misclassification may also be a concern since exposure assessment relied on subjects’ self-reported recall of historical exposures. It is possible that the observed interactions with the broad category of occupational pesticide exposures may have masked considerably stronger effects for specific pesticides.

Clearly, larger epidemiological studies using more robust exposure assessment methodologies will be needed to thoroughly evaluate associations with specific pesticides—and gene–environment interactions. Other groups studying PD should attempt to replicate these new findings and to gain a more solid understanding of the role of DAT in PD. We estimate that a sample of size of approximately 500 cases and 500 controls would be required to replicate these results, assuming similar effect sizes to what we observed, similar allele frequencies and prevalence of exposure (25% among men) and 80% power with α = 0.05 (72). Larger sample size might also allow for examination of interactions with other genes, especially those involved in dopaminergic pathways (e.g. MAOB, COMT and VMAT2).

**MATERIALS AND METHODS**

**Case–control study**

The Institutional Review Board Committees on Human Subjects Research at the University of Washington and the Group Health Cooperative (GHC) Center for Health Studies approved this study. Newly diagnosed idiopathic PD patients (n = 293) were identified from neurology and general medical practice clinics of the GHC from the Puget Sound area in western Washington State and referred by neurology clinic at the University of Washington Medical Center (73). Diagnostic criterion for the PD cases was the presence of at least two of the four cardinal signs of PD: bradykinesia, resting tremor, cogwheel rigidity and postural reflex impairment (1), at least one of which had to be bradykinesia or resting tremor. Exclusion criteria included the use of certain medications during the 12 months preceding symptom onset, prior history of multiple cerebrovascular events or other explanations for symptoms of parkinsonism. To verify PD diagnoses, two study neurologists (P.D.S. and G.M.F.) reviewed charts for cases not referred by neurologists. A third neurologist reviewed the cases on which the first two reviewers disagreed about eligibility (W.T.L.). Descriptive statistics and demographic characteristics are shown in the Supplementary Material, Table S3.

Control subjects (n = 395) were identified from GHC enrollees without past histories of PD or other neurodegenerative disorders. Controls were frequency-matched to cases by birth decade, gender and year of enrollment in GHC. All subjects were European Americans. Study subjects were volunteers who were informed of the purpose of the study. Since our last report (27), we have accrued 32 new cases and 21 new controls. All subjects included in the prior analysis were included in this analysis as well.

During a structured interview, subjects provided data on demographics, medical and occupational history. From the interview data, subjects’ occupational pesticide exposure history was obtained, as described by Firestone et al. (74). For the purposes of this project, subjects were classified as
exposed if they reported having an occupation for 6 months or more, which involved pesticide exposure, including pest extermination, crop dusting, vegetable, crop or animal farming. Because very few women reported these exposures, analysis was restricted to men only. In this study population, residential pesticide use was not associated with PD status (74). Additionally, and shown in the Supplementary Material, Table S3, rural or farming residence history was not associated with PD.

Nucleotide diversity, cladistic analysis and genotyping

We previously re-sequenced 48 chromosomes to identify common SNPs in the region spanning ~5000 bp 5′ of exon 1 (transcriptional start site) through the start of exon 2 (where the translational start site lies), a total of ~7.5 kb. We identified 22 SNPs comprising eight haplotypes (27). In the present work, we first re-examined overall nucleotide diversity by calculating diversity (θs) and average heterozygosity (θh), and then tested for possible effects of selection on the 5′ region, using Tajima’s D statistic (75). These statistics were calculated using Arlequin software (76). To understand the relationships among haplotypes and to place them into an evolutionary context, we constructed a phylogenetic tree of the haplotypes using the unweighted pair-group method, using arithmetic averages (UPGMA) method in the MEGA2 software package (77).

A haplotype-tagging approach was used to efficiently infer haplotypes of the 5′ region, allowing the distribution of common haplotypes among PD cases and controls to be examined. Six SNPs were used: −2603T>C (rs3756450), −2315G>A (rs2652510), −2299A>T (rs35032437), −2296C>T (rs2550956), I1 + 478G>T (rs11564752) and I1 + 1473G>A (rs1316830). SNP positions given are in reference to GenBank sequence DQ307031. Genotyping methods were previously re-sequenced 48 chromosomes to identify common haplotypes among PD cases and controls to be examined. Six SNPs were used: −2603T>C (rs3756450), −2315G>A (rs2652510), −2299A>T (rs35032437), −2296C>T (rs2550956), I1 + 478G>T (rs11564752) and I1 + 1473G>A (rs1316830). SNP positions given are in reference to GenBank sequence DQ307031. Genotyping methods for these SNPs were described previously (27). The VNTR was genotyped using the methods of Vandenberg et al. (78). Some subjects (n = 15) had rare VNTR alleles (containing 5, 6, 8 or 11-repeats) and were excluded from the analysis.

Despite the fact that the 5′ and 3′ regions lie within ~80 kb of each other on the same gene, the linkage disequilibrium between these two regions is quite low (25,79). Owing to the low linkage disequilibrium between these two regions, phylogenetic tree analysis could not be conducted for the entire region examined (80,81). Additionally, the association analysis can essentially be considered a two-locus problem.

Statistical analysis

Results of single-locus (5′ or 3′) analyses showed that the A clade of the 5′ region and VNTR 9-repeat allele were associated with PD in this study. Hence, in the analysis of combined 5′ and 3′ region genotypes, we categorized subjects by their number of risk alleles:

- 0: reference group, B-10/B-10; n = 140
- 1: B-9/B-10 and A-10/B-10; n = 252
- 2: A-10/B-9, B-9/B-9, A-9/B-10 and A-10/A-10; n = 216
- 3: A-9/B-9 and A-9/A-10; n = 69
- 4: A-9/A-9; n = 11.

We then used these values in a logistic regression model with indicator variables representing the different diplotype levels, adjusting for age (<60, ≥60), gender and smoking status (ever/never). Because the number of subjects with four risk alleles was small and because ORs for subjects having two or three risk alleles were nearly the same, subjects with two, three or four risk alleles were collapsed into a ‘two or more’ (called 2+) risk allele category, giving a more parsimonious model that yielded more precise estimates of risk for subjects having 1 or 2+ risk alleles. In regression models of pesticide exposure and gene–pesticide interaction, education was added to the model (in quintiles, see the Supplementary Material, Table S1) because it was associated with PD and is considered a proxy for occupation-related physical activity, which has been associated with PD risk in other studies (82,83).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

This work was supported by NIEHS grants P30ES07033, P42ES04696, R01ES10750, ES07032 and the University of Washington Magnuson Scholars Program. The authors also wish to acknowledge the contribution of Group Health Cooperative and University of Washington neurologists, Drs Robert Gotshall, Ann Hunt, Eric Kraus, Richard Mesher, Steve Pugh, Bruce Ransom, Ali Samii, Timothy Searce, Kurt Seiffert, Ken Uchino and Thurman Wheeler, for referral of PD cases. The authors also thank Drs Joan Bailey-Wilson and Jay Lubin for helpful discussions and their advice.

Conflict of Interest statement. None declared.

REFERENCES


