The evolution of genetic variability at the LRRK2 locus

Matthew Farrer  

m.farrer@ufl.edu  

University of Florida  
https://orcid.org/0000-0003-1159-5321  

Dylan Guenther  
University of Florida  

Jordan Follett  
University of Florida  

Samia Ben Sassi  
Mongi Ben Hamida National Institute of Neurology  

Rim Amouri  
Mongi Ben Hamida National Institute of Neurology  

Faycel Hentati  
Mongi Ben Hamida National Institute of Neurology  

Brief Communication

**Keywords:** LRRK2, Parkinson's Disease, Evolution

**Posted Date:** March 13th, 2024

**DOI:** https://doi.org/10.21203/rs.3.rs-4001026/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License

**Additional Declarations:** There is a conflict of interest MJF reports US patents associated with LRRK2 mutations and mouse models (8409809, 8455243), and methods of treating neurodegenerative disease (20110092565).
Abstract

Leucine-rich repeat kinase 2 (LRRK2) c.6055G>A (p.G2019S) is a frequent cause of Parkinson's disease (PD) accounting for >30% of Tunisian Arab-Berber patients. LRRK2 is widely expressed in the immune system and its kinase activity confers a survival advantage from infection in animal models. Here we assess haplotype variability in cis and in trans of the LRRK2 c.6055G>A mutation, define the age of the pathogenic allele, explore its relationship to age of disease onset (AOO), and provide evidence for its positive selection.

Introduction

Globally, genetic variability within LRRK2 confers the high genotype and population attributable risk for PD, and the frequency of many variants appear to be population specific\(^1\). All pathogenic LRRK2 mutations elevate its kinase activity, whereas p.G2019S (rs34637584_A) directly breaks the hinge of the 'activation segment' keeping the enzyme constitutively active\(^2,3\). In Arab Berbers, rs34637584_A has a background frequency of 0.9%, it accounts for >30% of sporadic patients and 40% of those with a family history of PD\(^4\). Despite the gene's identification through linkage as a dominant Mendelian disorder\(^5\), the penetrance of rs34637584_A is incomplete\(^6\); while subtle prodromal signs may be missed, including hyposmia, REM sleep behavior disorder and orthostasis, the expressivity of LRRK2 parkinsonism is as variable as idiopathic PD\(^7\). Conceivably, penetrance and expressivity are a function of other genetic and environmental modifiers. Notably, polymorphisms within the LRRK2 locus are associated with PD susceptibility\(^1\) and AOO in progressive supranuclear palsy\(^8\), and these may influence LRRK2 expression, protein interactions, and kinase activation\(^9\). Penetrance may also be influenced by LRRK2's role in intracellular innate immunity and pathogen response. In animal models LRRK2 p.G2019S promotes survival in infections disease\(^10,11\). Variability in the LRRK2 locus is also associated with inflammatory bowel disease\(^12\), pediatric immune disorders\(^13\), and type-1 response in leprosy\(^14\). Here we investigate the effects of genetic variability in cis and in trans of rs34637584_A in a sample from the Tunisian Arab-Berber population. We assess the relationship with AOO and the genetic evidence for ancestral LRRK2 haplotype selection.

Results for rs34637584_A were generated by direct genotyping and added to high density arraydata. As there are no Arab-Berber reference genomes within public databases we selected thirteen rs34637584_A heterozygotes with extreme AOO phenotypes (7 with young onset PD (mean AOO=34.6 SD=7.02 (22-42) years), and 6 elderly but clinically asymptomatic individuals (mean age=78.7 SD=7.0 (69-89) years) for whole genome sequencing (WGS). Chromosome 12 SNP imputation and haplotype phasing was done using a European genome reference with and without Tunisian WGS. Nevertheless, results were similar regardless of the imputation reference used and yielded 16,997 SNPs on chromosome 12 (average minor allele frequency (MAF) = 0.25 ± 0.13 SD, range 0.50 - 0.017).
Data from LRRK2 p.G2019S heterozygotes and homozygotes were compared to define allelic variability \textit{in cis} for the longest, most parsimonious, allele for the majority of samples. This spanned a genomic distance of 396Kb, from rs878010 to rs73110066, and included a total of 69 markers in addition to the pathogenic \textit{LRRK2} c.6055A variant (rs34637584_A at 12:40340400 (GRCh38)) (Supplementary Table 1). The 396Kb \textit{cis} haplotype included complete genotyping data for all samples (n=145) and was identical in all but one unaffected control with rs2404840_G>A, which may be due to recombination. SNP frequencies in the most parsimonious \textit{LRRK2} haplotype, versus allele frequencies in unrelated control participants without rs34637584_A, enabled the age of the mutation to be calculated at approximately 40 (95% CI 28-52) generations. Assuming 30 years per generation, the rs34637584_A ancestral allele in this sample originated approximately 1,200 ± 360 years ago. Within the same dataset we observe 81 alternate \textit{LRRK2} haplotypes \textit{in trans} (unique haplotypes defined as having ≥ 1/69 difference in marker alleles). A variable length Markov chain Monte Carlo method\textsuperscript{15}, implemented in Beagle3.3, was used to identify the shortest haplotype \textit{in trans} most associated with AOO, but none were observed that reach significance after correction for multiple testing. Additionally, a maximum likelihood method was used to resolve haplotype relationships as a phylogenetic tree. This identifies 3 major clades from a central unrooted node and can be partitioned by five major SNPs (rs2638245, rs10878199, rs2638271, rs2708438, and rs1388587) that span the 40.1 – 40.3 Mb interval. Nevertheless, no clade association with AOO was apparent (z(2)=0.40, p=0.69) (Supplemental Fig. 1).

Lastly, we investigated whether the background frequency for the highly conserved rs34637584_A haplotype might be driven by recent positive selection. Integrated haplotype scores (iHS) summarize the evidence for the entirety of chromosome 12, as illustrated for affected heterozygotes and homozygotes (AG+AA), and wild type (GG) affected and unaffected individuals. A cluster of higher iHS’s (>2.5) demarks an interval between 39.8 and 41.0 Mb (Fig. 1). The distribution of iHS scores for the entirety of chromosome 12 iHS values (minus the \textit{LRRK2} locus) was bootstrapped, and suggests this cluster is highly significant compared to scores for the rest of the chromosome (p=4.50 E\textsuperscript{-18}). Curiously iHS scores within the \textit{LRRK2} locus from 40.2 – 41.0Mb were also significant for rs34637584_G wild type alleles (GG\textsubscript{all}=422, p=2.95 E\textsuperscript{-4} to 1.61 E\textsuperscript{6}.Table 2). As several inflammatory disorders are associated with the \textit{LRRK2} locus, we removed any individual with these disease-associated SNP alleles, namely rs11175593_T\textsuperscript{12}, rs4768236_C\textsuperscript{16}, and/or rs17466626_G\textsuperscript{13} (GG\textsubscript{nim} = 208) and the \textit{LRRK2} signal was ablated (Table 2). Despite the reduction in sample size, mean iHS scores and their distributions were comparable in sub-groups with and without inflammatory markers (0.79 ± 0.59SD vs 0.80 ± 0.59SD) (Supplementary Fig. 2). Overall, these results are consistent for positive evolutionary selection for the \textit{LRRK2} region not just for the rs34637584_A allele.

In conclusion, this study supports and extends prior studies suggesting LRRK2 p.G2019S heterozygotes are descendants of a common ancestral founder who originated at least 40 (95% CI 28-52) generations ago (Supplementary Table 1). This result is within the confidence interval of prior estimates (Supplementary Table 1 & 2). The \textit{LRRK2} locus includes SNPs that nominate genome wide associations to several inflammatory disorders (Crohn's disease [rs11175593_T]\textsuperscript{12}; Inflammatory bowel disease...
[rs4768236_C] pediatric immune diseases [rs1746626_G], and platelet count [rs52989481_G]. However, in our data, those alleles are not in linkage disequilibrium with rs34637584_A (the LRRK2 c.6055A haplotype). Rather, those alleles are captured on haplotypes in trans. Whether these variants confer a functional change on LRRK2 expression or activity has yet to be demonstrated.

Despite our limited sample size, the cluster of iHS values around the LRRK2 locus is indicative of positive selection for LRRK2 rs34637584_A. Although Tunisia has a high frequency of consanguineous marriages, neither isolation nor genetic drift are likely to produce the distribution of values observed. Overall, the burden of evidence from our data and others suggests rs34637584_A, and the constitutive LRRK2 kinase activity it confers, offers a survival advantage to reproductive age. To date, this has enabled a >19-fold increase in the background frequency of rs34637584_A in Tunisia, in our sample, as compared to the global mean (rs34637584_A MAF_{Tunisia} = 0.0094 (7/742); MAF_{gnomADr2.1-all} = 0.0004884 (138/282542), MAF_{gnomADr2.1-African} = 0.0001202 (3/24962). Fisher's p=9.53 E^{-27}). The evolutionary forces driving positive selection are unknown but epidemiologic and experimental research on pathogens restricted by LRRK2 kinase activity may be informative. Intriguingly, the retromer is also central to innate immune responses and often corrupted by intracellular pathogens. Its core component, VPS35 p.D620N, is linked to PD and activates LRRK2 kinase. RAB32 p.S71R, recently linked and associated with PD, also causes LRRK2 kinase activation. RAB32 is central to the biogenesis and transport of melanosomes in melanocytes, and similar components are deployed in catecholamine metabolism and pigment production. RAB32 traffics mitochondrially derived itaconic acid to the pathogen-containing vacuole, to inhibit bacterial growth. It also interacts with PINK1 that instigates mitophagy, and for which loss-of-function mutations are best described in Tunisian families with parkinsonism. Hence, most Mendelian gene mutations that cause PD impinge on phagolysosome biology and intracellular innate immunity and may illustrate convergent evolution. Nevertheless, the ability to meaningfully investigate this for other linked loci that cause PD is limited by sample size. How peripheral and central immunity might influence the vulnerability of dopaminergic neurons in LRRK2 parkinsonism and idiopathic PD remains to be defined, but both cell autonomous and non-autonomous mechanisms evidently contribute.

**Methods**

*Participants and Genotyping*

A total of 434 patients and 321 controls, of which 232 were LRRK2 p.G2019S carriers were included in this study (Table 1). All were of Tunisian descent, and family relationships were sought, and pedigrees constructed when appropriate. All participants in either the discovery cohort or the replication series were aged 18 years or older at neurological assessment and provided written informed consent. Blood samples were sourced ethically and their use in research was in accordance with the terms of written informed consent. Approval for this study was obtained from the local ethics committee at the National
Institute and Ministry of Health (Tunisia), and was reviewed by GlaxoSmithKline (UK), the Institutional Review Board of Mayo Foundation (USA), the Research Ethics Board of the University of British Columbia (Canada) and the Institutional Review Board of the University of Florida.

**Declarations**

**Data availability**

All raw data is available from the investigators under a material transfer agreement, and provided its use does not conflict or compete with ongoing studies. That research must be covered under an appropriate ethics protocol.

**Acknowledgements**

We would like to thank Ekaterina Nosova, PhD and Joanne Trinh, PhD, for their help in the initial stages of data collection and analysis. The project was funded by a Lee and Lauren Fixel Chair in Parkinson's research (M.J.F). The authors thank the research participants who made this study possible and the staff the Mongi Ben Hamada National Institute of Neurology.

**Potential conflict of interest**

MJF reports US patents associated with LRRK2 mutations and mouse models (8409809, 8455243), and methods of treating neurodegenerative disease (20110092565).

**References**


Tables

Table 1: Demographics of LRRK2 p.G2019S and idiopathic patients

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>214</td>
<td>321</td>
<td>220</td>
<td>12</td>
</tr>
<tr>
<td>Number of men (%)</td>
<td>105 (49.1%)</td>
<td>168 (52.5%)</td>
<td>124 (56%)</td>
<td>6 (50%)</td>
</tr>
<tr>
<td>Mean age (SD) years</td>
<td>68.1 (12.8)</td>
<td>62.4 (11.0)</td>
<td>67.6 (12.6)</td>
<td>56.7 (10.9)</td>
</tr>
<tr>
<td>Median age (IQR)</td>
<td>68 (59-76)</td>
<td>62 (53-69)</td>
<td>69 (48-90)</td>
<td>54.5 (38-72)</td>
</tr>
<tr>
<td>Mean age of onset (SD)</td>
<td>54.9 (14.5)</td>
<td>-</td>
<td>57.1 (11.6)</td>
<td>-</td>
</tr>
<tr>
<td>Median age at onset (IQR)</td>
<td>58 (46-66)</td>
<td>-</td>
<td>57 (40-74)</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2: Evidence for positive selection in LRRK2 wildtype alleles with and without Inflammation Markers
<table>
<thead>
<tr>
<th>P-Value (w/ inflammatory markers)</th>
<th>P-Value (w/out inflammatory markers)*</th>
<th>Window Bin [Start-End (Mb)]</th>
<th>LRRK2 Locus</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.16</td>
<td>0.22</td>
<td>39.6 – 39.85</td>
<td>N</td>
</tr>
<tr>
<td>5.56 E -5</td>
<td>0.011</td>
<td>39.8 – 40.05</td>
<td>Y</td>
</tr>
<tr>
<td>1.61 E -6</td>
<td>0.27</td>
<td>39.6 – 40.25</td>
<td>Y</td>
</tr>
<tr>
<td>3.64 E -4</td>
<td>0.32</td>
<td>40.2 – 40.45</td>
<td>Y</td>
</tr>
<tr>
<td>2.95 E -4*</td>
<td>0.14</td>
<td>40.4 – 40.65</td>
<td>Y</td>
</tr>
<tr>
<td>0.36</td>
<td>0.27</td>
<td>40.6 – 40.85</td>
<td>Y</td>
</tr>
<tr>
<td>0.55</td>
<td>0.08</td>
<td>40.8 – 41.05</td>
<td>Y</td>
</tr>
<tr>
<td>0.45</td>
<td>0.99</td>
<td>41.0 – 41.25</td>
<td>Y</td>
</tr>
<tr>
<td>0.0060</td>
<td>0.0058</td>
<td>41.2 – 41.45</td>
<td>N</td>
</tr>
<tr>
<td>0.34</td>
<td>0.58</td>
<td>41.4 – 41.65</td>
<td>N</td>
</tr>
<tr>
<td>0.06</td>
<td>0.15</td>
<td>41.6 – 41.85</td>
<td>N</td>
</tr>
<tr>
<td>1.28 E -4</td>
<td>5.71 E -5</td>
<td>41.8 – 42.05</td>
<td>N</td>
</tr>
<tr>
<td>0.093</td>
<td>0.12</td>
<td>42.0 – 42.25</td>
<td>N</td>
</tr>
<tr>
<td>0.26</td>
<td>0.04</td>
<td>42.2 – 42.45</td>
<td>N</td>
</tr>
<tr>
<td>0.61</td>
<td>0.53</td>
<td>42.4 – 42.65</td>
<td>N</td>
</tr>
<tr>
<td>0.51</td>
<td>0.34</td>
<td>42.6 – 44.85</td>
<td>N</td>
</tr>
</tbody>
</table>

*Inflammatory SNP alleles considered were rs11175593_T\(^{12}\), rs4768236_C\(^{16}\), and/or rs17466626_G\(^{13}\)

Figures
Figure 1

Integrated Haplotype Scores in Chromosome 12 of Tunisian Arab Berbers

Absolute value of elevated iHSs > 2.5 across chromosome 12. A cluster of elevated iHSs can be observed within the 39.8 – 41.0Mb range, which encompasses LRRK2 p.G2019S haplotypes and is consistent with positive selection within the region.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- LRRK2CisandTransSupplemental.docx