# Genome-wide changes in genetic diversity in a population of *Myotis lucifugus* affected by white-nose syndrome

Lilley TM [1,2], Wilson IW [1], Field KA [3], Reeder DA [3], Turner G [4], Kurta A [5], Blomberg AS [6], Hoff S [7], Herzog C [7], Paterson S [1].

[1]Institute of Integrative Biology, University of Liverpool, United Kingdom. [2]Finnish Museum of Natural History, University of Helsinki, Finland. [3]Biology Department, Bucknell University, 1 Dent Drive, Lewisburg, PA 12837, USA. [4]Pennsylvania Game Commission, Harrisburg, PA, USA. [5]Department of Biology, Eastern Michigan University, Ypsilanti, MI, USA. [6]Department of Biology, University of Turku, Finland. [7]Wildlife Diversity Unit, State Department of Environmental Conservation, Albany, NY, USA.

Corresponding author: tmlill@liverpool.ac.uk

## Abstract

Novel pathogens can cause massive declines in populations, but seldom lead to extirpation of hosts. Rather, disease can act as a selective pressure on survivors, driving the evolution of resistance or tolerance. Bat white-nose syndrome (WNS) is a rapidly spreading wildlife disease in North America. The fungus causing the disease invades skin tissues of hibernating bats, resulting in disruption of hibernation behaviour, premature energy depletion, and subsequent death. We use whole-genome sequencing to investigate changes in allele frequencies within a population of *Myotis lucifugus* in eastern North America to scan for genetic resistance to WNS. Our results show low $F_{ST}$ values within the population across time, i.e. prior to WNS (Pre-WNS) compared to the population that has survived WNS (Post-WNS). However, when dividing the population with a geographical, distance-based cut-off between the states of Pennsylvania and New York, a sharp increase in values on scaffold GL429776 is evident in the Post-WNS samples. Genes present in the diverged area are associated with thermoregulation and promotion of brown fat production. Thus, although WNS may not have subjected the entire *M. lucifugus* population to selective pressure, it could allowed the rise of a selection for alleles in Pennsylvania through decreased gene flow within the population, which may contribute to survival. However, the existence of remnant sub-populations in the aftermath of WNS is likely due to multiple factors in bat life history.

Keywords: Opportunistic pathogen, white-nose syndrome, Selective pressure, Genetic resistance.

## 1. Introduction

The emergence and spread of multiple infectious wildlife diseases during recent decades has had devastating consequences for biodiversity (Daszak et al., 2000; Smith et al., 2006). Unfortunately, for many taxa, anthropogenic activities and climate change appear to be the main causes of exposure to sources of infection (Gallana et al., 2013; Tompkins et al., 2015). With shifting pathogen ranges, disease-related declines in naïve wildlife often threaten the persistence of populations. Examples include chytridiomycosis, which has wiped out populations of amphibians globally (Daszak et al., 1999; Lips et al., 2006), and avian malaria, which caused the sharp decline of island populations of birds (van Riper et al., 1986). More recently, white-nose syndrome (WNS) has been described as one of the most rapidly spreading wildlife diseases ever recorded (Blehert et al., 2009; Frick et al., 2010). Since the discovery of WNS in North America in 2006, ten species of bats have been diagnosed with the disease in 33 U.S. states and 7 Canadian provinces (www.whitenosesyndrome.org, 2019). Genetic evidence suggests that *Pseudogymnoascus destructans*, the causative agent of WNS, was introduced to North America from Europe (Campana et al., 2017; Lorch et al., 2013; Minnis and Lindner, 2013; Ren et al., 2012; Wibbelt et al., 2010), where affected species do not experience associated mortality (Harazim et al., 2018; Puechmaille et al., 2011; Warnecke et al., 2012; Wibbelt et al., 2013; Zukal et al., 2016).

When a disease enters a naïve host population, the initial wave of infection often causes epizootics that result in mass mortality, which may extirpate local host populations or even cause species extinction (Daszak et al., 1999; De Castro and Bolker, 2005; Frick et al., 2010). However, extinctions due to a disease appear to occur seldomly, and a disease can act as a

56 selective pressure on survivors, driving the evolution of tolerance or resistance, which may
57 transform a disease from being epizootic to being enzootic (Boots et al., 2009; Karlsson et al.,
58 2014; Robinson et al., 2012). This may occur through rapid changes in the distribution of
59 genetic variants associated with disease susceptibility over short timescales (Gallana et al.,
60 2013). These powerful selective pressures pathogens exert on naïve host populations (Sironi
61 et al., 2015) and genetic signatures of infection-related selective sweeps may be detectable for
62 generations (Deschamps et al., 2016; Di Gaspero et al., 2012; Groot et al., 2002). Detection of
63 selective sweeps on particular genes or gene families can confirm or exclude potential
64 mechanisms of host susceptibility or pathogen virulence (Campbell and Tishkoff, 2008).
65 Rapid, extreme bottlenecks (such as those caused by panzootic events), meanwhile, are
66 associated with a more stochastic loss of alleles, which does not necessarily indicate selection
67 (Luikart et al., 1998). Much of the research on disease-induced selection has focused on the
68 major histocompatibility complex (MHC), and indeed, diseases can drive the evolution and
69 maintenance of MHC diversity in natural populations (Davy et al., 2017a; Jeffery and
70 Bangham, 2000; Paterson et al., 1998; Spurgin and Richardson, 2010; Teacher Amber G. F. et
71 al., 2009; Zeisset and Beebee, 2014). However, factors not directly associated with interactions
72 between host and pathogen, such as environmental conditions and competition with other
73 species, can have a considerable influence on the manifestation of a disease (Scholthof, 2007).

74 The psychrophilic (cold-growing) fungus *P. destructans*, (Lorch et al., 2011; Minnis and
75 Lindner, 2013; Warnecke et al., 2008) acts as an opportunistic pathogen of bats, invading the
76 skin tissues of hibernating hosts (Cryan et al., 2010; Meteyer et al., 2012). Susceptible species,
77 such as *Myotis lucifugus* and *M. septentrionalis*, have shown population declines greater than
78 90% in affected hibernacula (Frick et al., 2015). The infection disrupts the hibernation

behaviour of susceptible species and leads to more frequent arousals from torpor, evaporative water loss, premature energy depletion, and death of susceptible individuals due to emaciation (McGuire et al., 2017; Reeder et al., 2012; Verant et al., 2014; Warnecke et al., 2013; Willis et al., 2011). Naïve infected *M. lucifugus* upregulate genes involved in immune pathways during the hibernation period (Davy et al., 2017b; Field et al., 2015, 2018; Lilley et al., 2017). These responses are weak during torpor but are robust during the intermittent arousals (Field et al., 2018; Luis and Hudson, 2006). Therefore, increased arousals may be attempts by the host to counter the pathogen, in addition to quenching thirst, grooming, expelling waste and possibly foraging (Bernard and McCracken, 2017; Willis et al., 2011), and supplementing electrolytes (Cryan et al., 2013).

White-nose syndrome is a prime example of a disease which is manifested when factors associated with the environment (i.e. temperature and humidity of the hibernaculum), the host (and the host's response to infection) and the pathogen (optimum growth range, suitable host) overlap, i.e. intersect within the "disease triangle" (Scholthof, 2007). Therefore bat species, and populations within species, are affected differently according to hibernation behaviour and prevailing micro-climate conditions in available hibernacula (Grieneisen et al., 2015; Johnson et al., 2014; Langwig et al., 2015). As such, with an environmentally persistent opportunistic pathogen such as *P. destuctans* capable of subsistence in the absence of the host, loci not associated with MHC diversity or other immune response-associated factors (Donaldson et al., 2017), such as the amount of body fat prior to hibernation (Cheng et al., 2019), may also affect survival of species and populations. This highlights the importance of examining the genome as a whole (Sparks et al., 2019).

101 The initial impacts of WNS on six species of hibernating bats in the northeastern and
102 midwestern USA vary from almost complete extirpation to arrested population growth (Turner
103 et al., 2011). After the initial decline caused by WNS in *M. lucifugus*, reports have surfaced
104 describing stabilisation at smaller sizes or even increases in numbers of individuals (Dobony
105 and Johnson, 2018; Frick et al., 2017; Langwig et al., 2015; Maslo et al., 2015). Models
106 parameterised with long-term data on fungal loads, infection intensity and counts of *M.*
107 *lucifugus* at hibernacula have suggested development of either tolerance or resistance in these
108 persisting populations (Frick et al., 2017), which is supported by reports on infected individuals
109 not arousing from torpor as frequently as during the acute phase of the zoonosis (Lilley et al.,
110 2016). Since WNS has inflicted massive population declines in some *M. lucifugus*, there is a
111 possibility that selection could occur for alleles conferring resistance or tolerance from the
112 standing genetic variation, which would confer to resistance or tolerance could occur. Indeed,
113 Maslo and Feffermann (2015) suggested the occurrence of evolutionary rescue occurring and
114 Donaldson et al. (2017) described changes in an immunome across *M. lucifugus* populations
115 obtained through a sequence capture method, which may be attributed to selection by WNS. A
116 study suggests the initially affected *M. lucifugus* population is beginning to show signs of
117 resistance to the pathogen (Langwig et al., 2017). However, the possible selective pressures
118 WNS has exerted on the population, which may not even be related to immune responses (e.g.
119 Cheng et al., 2019), have only recently been described at the whole-genome level (Gignoux-
120 Wolfsohn et al., 2018), however high-sequencing depth studies are lacking.

121 Here, we utilise high-throughput whole-genome sequencing on the WNS-affected species *M.*
122 *lucifugus*, allowing us to look at entire genomes via single nucleotide polymorphisms (SNPs).
123 We compare genetic patterns ($F_{ST}$ and heterozygosity) before and after the spread of WNS in

124 the eastern North American population to gain insight into whether the disease is causing
125 selection of major loci in surviving bats. We hypothesise that, due to the massive population
126 declines caused by WNS and the fact that the affected population is now beginning to stabilise
127 or slightly increase in size (Langwig et al., 2017), a soft selective sweep based on standing
128 genetic variation has resulted in differentiation in one or many regions of the genome (Messer
129 and Petrov, 2013). Adaptation could therefore be in the process of driving the evolutionary
130 rescue of a population, in which individuals resistant or tolerant to the disease survived a
131 bottleneck and passed these traits to their offspring (Carlson et al., 2014). Alternatively,
132 survival may be due to factors not related to selection for WNS-resistant alleles, such as
133 microbiome adaptation (Ange-Stark et al., 2019; Lemieux-Labonté et al., 2017) or variation
134 across hibernation site microhabitats. This study provides a foundation for future work on
135 genome-wide effects of WNS in *M. lucifugus* and other affected species of bats.

## 136 2. Materials and methods

### 137 2.1. Ethics Statement

138 This study was carried out on bats from non-endangered species in strict accordance with
139 recommendations in the Guide for the Care and Use of Laboratory Animals of the National
140 Institutes of Health. All methods were approved by the Institutional Animal Care and Use
141 Committee at Bucknell University (protocol DMR-016). The bats were collected under
142 Pennsylvania Game Commission Special Use Permit #33085, and New York State Department
143 of Environmental Conservation Permit #427.

144 ## 2.2. Sample collection and DNA extraction

145 We conducted whole-genome sequencing on a population of *M. lucifugus* prior to, and up to
146 10 years after, the onset of WNS. White-nose syndrome in bats was first observed in upstate
147 New York, at Howe's Cave during the winter of 2006-2007 (Blehert et al., 2009). Our historic
148 samples were obtained from individuals consisting of the eastern population of *M. lucifugus*
149 (Miller-Butterworth et al., 2014; Vonhof et al., 2015) from New York (NY) (from the first
150 batch of dead bats found at Howe's Cave in 2006) and from bats caught in central Pennsylvania
151 (PA) in 2005-2006 (Fig. 1., Table 1.,Supp. Table 1.). The bats sampled from Howe's Cave,
152 New York were affected by WNS, but because they were the first bats to be documented with
153 the disease in North America, and moreover, because of the massive declines caused by the
154 disease, we believe this batch of samples is representative of the population genetic structure
155 of *M. lucifugus* prior to WNS. Bats in PA became affected by the disease in roughly 2008–
156 2009 and had not been affected by WNS at the time of sampling. We sampled wing tissue from
157 frozen specimens of known origin (Supp. Table 1.) that were collected and stored by the New
158 York State Department of Environmental Conservation (Delmar, NY) and D. Reeder to obtain
159 Pre-WNS samples. The remnants of the eastern North American population in NY and PA
160 were sampled again in 2016, and genetic diversity was compared before and after WNS (Fig.1.,
161 Table 1., Supp. Table 1.). Post-WNS bats were captured using mist nets and harp traps during
162 2015 -2016 from a number of maternal colonies, with the exception of individuals from
163 Williams Preserve Mine, NY and Woodward, Center County, PA, which were sampled from
164 hibernacula. In addition to samples from PA and NY, we collected gDNA from two individuals
165 from Upper-Peninsula Michigan (UPMI) for use in polymorphism detection. We did this in
166 order to take advantage of the greater number of polymorphic sites one would expect to arise

167 from a more diverse collection of individuals (see Section 2.3.). Tissue samples were collected
168 using either 3.0 mm or 5.0 mm biopsy punches (Integra Milltex, Plainsboro, NJ) and stored in
169 95% ethanol until extraction. We extracted DNA using QIAmp DNA Mini Kits (Qiagen,
170 Hilden Germany) and stored DNA at -80°C until sequencing.

171 ## 2.3. Sequencing methods

172 Sequencing libraries were made for two separate sets of data: **(1)** Six individuals from PA (N
173 = 2), NY (N = 2) and UPMI (N= 2) were sequenced separately in order to initially identify a
174 set of polymorphic sites within the *M. lucifugus* genome within the focal population (See Supp.
175 Table 1. for information on individuals). Individual sequencing increases the confidence of
176 SNP discovery, relative to pooled sequencing, since individuals exhibit discrete variation in
177 the number of alleles they carry at each potential SNP site; **(2)** Equimolar pools of DNA from
178 multiple individuals from within our north-eastern US population (NY and PA) at each of the
179 two timepoints (pre- and post-WNS) to give four pooled sequencing libraries (See Table 1. and
180 Supp. Table 1. for information on individuals). TruSeq Nano libraries with a 350 bp insert size
181 were prepared from all samples and run on a HiSeq4000 platform to generate 2x 150 bp reads
182 providing ~15 Gbp of sequence data (~7.5x coverage) for each of the six individuals **(1)** and
183 ~90 Gbp of sequence data (~45x coverage) for each of the four pools **(2)**. Adapter sequences
184 were removed with Cutadapt v1.2.1 and trimmed with Sickle 1.200 with a minimum quality
185 score of 20.

186 ## 2.4. Bioinformatics methods

187 First, reads from the six individually sequenced individuals (see 2.3.) were mapped onto the
188 *M. lucifugus* reference genome (https://www.broadinstitute.org/brown-bat/little-brown-bat-
189 genome-project) using Bowtie2 v2.2.5 (Langmead and Salzberg, 2012), with the '—sensitive-
190 local' preset. Reads were sorted using SAMtools v1.1 (Li et al., 2009) and combined by read
191 group (with duplicates removed using a 'lenient' validation stringency setting) using Picard-
192 tools v2.5.0 (http://broadinstitute.github.io/picard). SNPs were called across all six individuals
193 with the Genome Analysis Toolkit (GATK) v3.4 HaplotypeCaller (McKenna et al., 2010)
194 using default parameters. Within this SNP set, high-quality bivalent SNPs were selected that
195 met selection criteria across the six individuals (Quality Score > 100; AF<1; 30>=DP>=100)
196 and were input to the SAMtools mpileup function as a list of sites at which pileup was to be
197 generated. After SNP calling, these six individuals were discarded from further analyses.

198 Reads from the four pooled samples were mapped to the *M. lucifugus* reference genome as
199 above and SAMtools mpileup was used to count alleles at high-quality SNP sites identified
200 above from the six sequenced individuals (i.e. reads matching either the reference or alternate
201 allele for bivalent loci). Sites with coverage in the bottom 10% or top 10% quantiles summed
202 across the pooled samples were excluded (since these could reflect inaccuracies in the reference
203 genome), leaving about 13.5 million sites.

204 Average heterozygosity within the pooled samples was calculated using npstats (Ferretti et al.,
205 2013), which controls for number of individuals within each sequencing pool and sequencing
206 depth, thus allowing robust comparisons between pooled samples. One million SNPs were
207 randomly picked and permuted, with 1000 sets of 1000 SNPs used to calculate the observed

208 heterozygosity within each pooled set of individuals and, from these, we produced the median
209 and 95% confidence intervals on heterozygosity.

210 Using allele counts at each locus to estimate allele frequencies, $F_{ST}$ was calculated without
211 weighting for sampling effort ($F_{ST} = \left(\bar{p}(1-\bar{p}) - \overline{p(1-p)}\right)/\bar{p}(1-\bar{p})$, where $p$ is the allele
212 frequency and either averaged between populations ($\bar{p}$) or the value $p(1-p)$ averaged across
213 populations). $F_{ST}$ was separately calculated between the Pre-WNS and Post-WNS samples and
214 between subsections thereof, specifically PA pre- vs post-WNS, NY pre- vs post-WNS, PA
215 pre- vs NY pre-WNS, PA post- vs NY post-WNS, and all PA vs all NY). Selection is expected
216 to elevate $F_{ST}$ at specific genomic regions relative to background levels due to sampling noise
217 or genetic drift, which are expected to act equally across the genome. To identify such regions
218 across the genome, a moving window was applied to calculate median pairwise $F_{ST}$ values for
219 groups of 500 consecutive SNPs at a time to even out sampling effects. Moving windows
220 shifted 100 SNPs at a time, thus providing the best balance between fine detail and practicality.
221 Allele frequency estimates were weighted by the number of individuals in each pool. This was
222 done using a custom R script (see Data availability section), run using R v3.4.1 (R
223 Development Core Team, 2011). Results were confirmed using Popoolation2 with a 10 kb
224 sliding window and with defined pool sizes. This gave the same qualitative results, but higher
225 levels of background noise (Kofler et al., 2011). Additionally, Popoolation2 was used to
226 perform Fishers Exact Tests on individual SNPs in order to test whether any very highly
227 significant SNPs were missed by the moving window approach.

228 From the comparison of the NY and PA pooled samples (i.e. the comparison that produced the
229 highest $F_{ST}$ values), we visually identified across the whole genome windows with median $F_{ST}$

230 values greater than 5 standard deviations from the mean, as well as those with $F_{ST}$ values greater
231 than 0.05, indicating at least 'moderate' genetic differentiation (Wright, 1978). The numbers
232 of such complete windows within gene models were counted for each gene to identify genes
233 potentially under selective pressure.

## 234 3. Results

### 235 3.1. Pre-WNS vs Post-WNS comparisons

236 To detect regions of the genome that may be under selection, we plotted $F_{ST}$ across the *M.*
237 *lucifugus* genome, calculated from allele counts within moving windows covering 500 SNPs
238 and shifting 100 SNPs at a time. Comparing Pre- and Post-WNS samples revealed an overall
239 low level of population differentiation due to genetic structure ($F_{ST}$ median of 0.0059; Table
240 2). Whilst a small number of windows have $F_{ST}$ values slightly beyond our arbitrary cut-off of
241 5 standard deviations above the mean, none have $F_{ST}$ values exceeding 0.05, suggesting that
242 there is not even moderate selection in any region of the genome in Post-WNS samples
243 compared to Pre-WNS samples (Fig. 2, Table 2)(Wright, 1978). Similarly, using Fisher exact
244 tests from Popoolation2 at a threshold of $p < 10^{-6}$, no individual SNPs were found to be
245 consistently associated with differences Pre- versus Post-WNS among the four sets of samples.

### 246 3.2. Comparisons within geographical locations

247 Due to our sampling and sequencing setup, we were also able to separately look at individuals
248 of the same population from two discrete geographical locations within their range. In our case
249 a divide was made between bats sampled in New York and Pennsylvania based on
250 geographical distance between the sampling sites. Although population genetic analysis based

251 on a considerably lower number of genetic markers assigns *M. lucifugus* on the eastern coast
252 of North America to the same population (Vonhof et al., 2015), the species shows high degrees
253 of philopatry (Davis and Hitchcock, 1965; Norquay et al., 2013), warranting the examination
254 of these arbitrarily assigned pools of individuals outside the mean dispersal range of the species
255 using more accurate, whole-genome sequencing based population genetic analyses.

256 A number of minor differences ($F_{ST}$ > 5 stdevs above mean, but $F_{ST}$ < 0.05) between these sets
257 of samples are evident prior to arrival of WNS, as well as a peak in scaffold GL429835 with
258 an $F_{ST}$ value exceeding 0.05 (Fig. 3a). The high peak consists of 4 windows the midpoints of
259 which span an 81,887 bp region within a coding sequence consisting of 47 exons, which
260 encodes the NIPBL cohesion loading factor (NCBI Gene ID: 102430106). The highest peak in
261 a comparison of PA populations Pre- and Post-WNS (Fig. 3b) lies in a similar position within
262 this coding region, as does a peak in a similar comparison of NY samples (Fig. 3c), though
263 neither exceed the $F_{ST}$ value cut-off of 0.05.

264 ### 3.3. Post-WNS comparisons between geographically isolated sample sets

265 When comparing the Post-WNS individuals sampled from PA with the Post-WNS individuals
266 from NY, the peak on scaffold GL429835 is no longer visible, but a far more pronounced spike
267 in $F_{ST}$ values can be clearly seen in scaffold GL429776 (Fig. 4a), indicating relatively large
268 allele frequency differences between the sample sets in this region of the genome. This is also
269 seen, to a lesser extent, when comparing all samples from NY with all samples from PA (Fig.
270 4b). However, this significant peak in this comparison of all samples collected does not affect
271 average fixation indices across the genome relative to the other comparison of all samples, Pre-
272 WNS vs Post-WNS (Table 2). This peak could be attributed to mapping error, but our coverage

273 depths show an average depth of 66.33x across the whole genome, whereas the focal region
274 was sequenced at an average depth of 66.49x. This means that there has been no collapsing of
275 duplicated regions, nor poor coverage of the region. Also, viewing the data on the NCBI
276 genome viewer tool, with RepeatMasker histogram enabled, suggested no immediately
277 obvious increase in repetitive sequence in the region of interest.

278 We examined scaffold GL429776 more closely to compare the $F_{ST}$ values between individuals
279 from the arbitrarily selected sample sets from NY Post-WNS and PA Post-WNS (Fig. 4a).
280 Genes within scaffold GL429776 containing windows with fixation indices > 0.05 between the
281 NY Post-WNS and PA Post-WNS samples, are presented in Table 4 and Fig. 4a.

282 To clarify why we saw no fixation at this site when comparing Pre-WNS and Post-WNS
283 populations (Fig. 3a), nor when we compared Pre-WNS samples from both geographical
284 origins (Figs. 3b and 3c), we calculated for the relevant subpopulations the proportion of
285 reference alleles called at each site included in the $F_{ST}$ window analyses (Fig. 5). This revealed
286 a small decrease over time in the proportion of the reference alleles seen at this locus in the PA
287 subpopulation (Fig. 5a), commensurate with an increase in proportion of reference alleles in
288 the NY sample set (Fig. 5b). As such, existing differences between the NY and PA sample sets
289 were insufficient to cause a relative rise in fixation indices prior to WNS spreading through the
290 population (Fig 5c). Conversely, Post-WNS, allele proportions at this site in both sample sets
291 had not changed sufficiently to suggest selection *within* the sample set but demonstrated clear
292 differences *between* the NY and PA sample sets (Fig. 5d).

293 In the Post-WNS comparison, peaks which were above our arbitrary threshold of 5 standard
294 deviations above the mean across the whole genome, but below $F_{ST}$ = 0.05, can also be

295 observed in small scaffolds on the right-hand side of the $F_{ST}$ plots. The 63 windows comprising
296 these peaks lie in 47 scaffolds. Of these, high $F_{ST}$ windows lie in only 6 protein-coding regions
297 across 6 scaffolds (Table 3). Two of these encode partial gene models. Of the 4 remaining
298 models containing high $F_{ST}$ regions, all but 1 is described on NCBI as a "Low Quality Protein".
299 The exception is a gene encoding the ankyrin 2 protein, within which 7 high $F_{ST}$ windows lie.

### 300 3.4. Measures of heterozygosity within subpopulations

301 We examined heterozygosity across the genome in the subpopulations divided by time and
302 location to measure whether the epizootic reduces genetic diversity. We found the following
303 values: PA pre-WNS: Median: 0.263 (0.252 – 0.274 95% CI); PA post-WNS: 0.247 (0.239 –
304 0.255); NY pre-WNS: 0.255 (0.247 – 0.263); and NY post-WNS: 0.256 (0.247 -0.265).

## 305 4. Discussion

306 Examining our data as a whole, we found low $F_{ST}$ values between the Pre- and Post-WNS
307 individuals of our eastern North American *M. lucifugus* population, the absence of fixation
308 suggesting that WNS has not decreased the population in size sufficiently for the disease to
309 have an effect on genetic diversity. Also, we found no significant changes in within-population
310 heterozygosity Post-WNS, despite the massive population declines caused by WNS over the
311 past decade, and found overall high heterozygosity across the genome. These results are in
312 accordance with a similar, low-coverage study conducted by Gignoux-Wolfsohn et al. (2018).
313 Panmixia of populations in evolutionary time in bats to the east of the Rocky Mountains has
314 been previously documented (Vonhof et al., 2016; Wilder et al., 2015). Gignoux-Wolfsohn et
315 al. (2018) present evidence for a soft selective sweep, where selection acts on standing genetic

316 variation, between Pre- and Post-WNS populations. The analysis is based on a change in minor
317 allele frequencies due to selection, distinguished from background demography using a
318 simulation approach; and it shows a complex picture of potentially up to 85 unlinked SNPs,
319 each exhibiting small effects and with little or no selective signal at adjacent SNPs. In contrast,
320 a recent case study of selection on standing genetic variation, by myxoma virus in rabbits,
321 showed that soft selective sweeps caused concerted changes in allele frequencies at sets of
322 linked SNPs around coding loci (Alves et al., 2019). The robust pooled sequencing approach
323 allows us to examine for similar, larger scale changes in the genome. We acknowledge that
324 both methods have their advantages and disadvantages, such as the increased possibility for
325 false outliers in our pooled approach (Li et al., 2008). Hovewer we took these into account with
326 parameters used for aligning the reads (Kofler et al., 2011), sequencing depth (Ferretti et al.,
327 2013; Kofler and Schlötterer, 2014) pool size (Futschik and Schlötterer, 2010; Gautier, 2015)
328 and library preparation protocol (Kofler et al., 2016).

329 The species originally described as *M. lucifugus* in North America (Fenton and Barclay, 1980)
330 has since been divided into five non-sister species (Morales and Carstens, 2018). All of the
331 individuals sampled in our study belong to *M. lucifugus senso stricto*, with a range extending
332 from the east coast of North America to Alaska, and furthermore, to the same previously
333 assigned population (Miller-Butterworth et al., 2014). Previous nuclear genetic studies show
334 that differentiation is low, and there is no evidence for any major barriers to nuclear gene flow
335 across the range of *M. lucifugus*.

336 Due to our sampling design, we were able to separate bats geographically within our sampling
337 range. We made a cut-off point for a more in-depth investigation of bats sampled in the states

338 of Pennsylvania (PA) compared to New York (NY). Although the cut-off may seem rather
339 arbitrary, the reasoning behind this is geographic distance between sampling sites, and diving
340 the data between the two states is appropriate for our purposes as evidenced by the results.
341 Here, we see low $F_{ST}$ values between the two groups of bats prior to the onset of WNS,
342 indicating an absence of selection across the genome, with the exception of a small peak in
343 scaffold GL429835, which lies within a gene model encoding the NIPBL cohesion loading
344 factor. Relatively high $F_{ST}$ values are also visible within this region when comparing the PA
345 population before and after the spread of WNS, as well as the NY population pre- and post-
346 WNS. However, this peak is not evident when comparing PA and NY post-WNS (Fig. 4a).
347 This suggests that this coding sequence was under selective pressures, manifesting as
348 differentiating allelic frequencies, in both subpopulations prior to the arrival of WNS, but that
349 any selection was nullified by the population-level impact of WNS (or another event occurring
350 during the same timeframe). In addition, this result confirms that although the pre-WNS bats
351 from NY were from the initial batch of bats found dead of WNS at Howe's cave in 2006, they
352 were genetically similar to pre-WNS bats from PA, which were sampled before the arrival of
353 WNS. This is most likely because of the high frequency of initial mortality in the population,
354 which indiscriminately wiped out most of the population, eroding any genetic differentiation.
355 As for the protein NIPBL, it plays an important role in the function of the cohesin complex,
356 which is vital for chromosome segregation, DNA damage repair, and gene expression
357 regulation (Peters et al., 2008). As these results suggest the loss of low-level functional
358 diversity at this locus in our subpopulations, it would be interesting to further investigate what
359 impact this will have on the population as a whole.

360 The $F_{ST}$ comparison of Post-WNS individuals sampled from PA and NY depicts considerable
361 genetic differentiation within scaffold GL429776; differentiation which did not exist in the
362 Pre-WNS comparison of individuals sampled from these geographical areas. This is a
363 difference which would have gone unnoticed had we not made the comparison between bats
364 sampled in NY and PA, or used fewer genetic markers instead of a whole-genome approach
365 (Miller-Butterworth et al., 2014; Vonhof et al., 2015). The region's appearance post-WNS
366 when considering all geographical locations together is not, however, reflected when the data
367 are broken down into representatives of each geographic location. A comparison of allele
368 frequencies of individual SNPs within this region shows that individuals sampled from PA
369 showed an increase in deviation from the reference genome over time, whilst a greater
370 proportion of reference alleles were seen in NY samples Post-WNS compared with Pre-WNS.
371 The differences in proportions, growing as they were over time, were insufficient to exhibit a
372 high fixation index Pre-WNS, but were notably different after the disease had spread through
373 the population. One could argue, that this is due to sampling at maternal colonies and swarming
374 sites, which has known effects on allele frequency differences (Johnson et al., 2015). However,
375 we sampled at a number of maternal colonies, which would dilute this effect, and furthermore,
376 recent studies suggest philopatry within maternal colonies may not be as high as previously
377 expected (Olivera-Hyde et al., 2019, although on *M. septentrionalis*). Therefore we assume our
378 sampling of multiple maternal colonies and hibernation sites has removed any effects of
379 relatedness within our sample sets.

380 This region within the scaffold containing the peak contains six annotated genes with very
381 different functions. Two, POC1B and CEP290, are involved in cilia development and
382 maintenance. The ATP2B1 gene encodes a protein which belongs to the family of P-type

383 primary ion transport ATPases, whereas TMTC3 is involved in the control of endoplasmic
384 reticulum stress response. MGAT4C is a glycosyltransferase that participates in the transfer of
385 N-acetylglucosamine (GlcNAc) to the core mannose residues of N-linked glycans. Finally,
386 KITLG produces a protein involved in mast cell development, migration and function, and
387 melanogenesis. KITLG has been documented as a target of evolution in recent studies and in
388 humans it has been linked with thermoregulation (Yang et al., 2018), which is a critical
389 component of bat life history (Studier and O'Farrell, 1972). In fact, upregulation of KITLG at
390 low temperature helps promote the production of brown fat for heat generation (Huang et al.,
391 2014). Bats utilise brown adipose tissue in arousals to normothermia from different degrees of
392 torpor (Thomas et al., 1990). There is also evidence of parallel evolution of pigmentation in
393 sticklebacks and humans linked to KITLG (Miller et al., 2007), as well as its role in determining
394 resistance/susceptibility to swine respiratory diseases in Erhulian pigs (Huang et al., 2017).
395 KITLG, by influencing mast cell development, could also play a role in either protective or
396 pathological immune responses to *P. destructans*, which may involve IgE-mediated
397 recognition of secreted fungal proteins. As such, temporal changes we found in allele
398 frequencies at this locus could, for instance, be linked to differences in hibernation site
399 temperatures, and/or thermoregulatory needs in bats sampled at different locations (Humphries
400 et al., 2002; Thomas et al., 1990). Spatial patterns, as seen here between the arbitrarily pooled
401 samples from PA and NY, could be due to local adaptation, in which strong selective sweeps
402 may be linked to gene variants favoured in local interactions (Hansen et al., 2012; Kyle et al.,
403 2014; Rico et al., 2015; Schoville et al., 2012). Whilst we lack data concerning migration
404 patterns of the bats studied here, as well as the environmental conditions within their
405 microfugia, future work would benefit from these data in a bid to understand whether bats

406 repeatedly frequent habitats with similar conditions and whether, indeed, there is a significant
407 difference in conditions across these Eastern states that explains the rapid rise in allelic
408 differentiation, and thus $F_{ST}$ values, at this locus.

409 The increase in $F_{ST}$ values in this scaffold GL429776 could, alternatively, be due to the massive
410 population decline caused by WNS (Langwig et al., 2017). Although not supported by existing
411 data, the decline could potentially have resulted in a decrease in connectivity and gene flow
412 between remnant local populations allowing for a local adaptation to arise (Bohonak, 1999;
413 Lilley, Anttila, & Ruokolainen, 2018). Genetically resolved examples of local adaptations are
414 rare (Hoekstra et al., 2006) and, in our focal system, pinpointing the causality leading to local
415 adaptation is especially difficult due to the cryptic life history of bats. Determining how these
416 variants are spread or localised among populations is essential to understanding and managing
417 the emergence of new selective pressures (Eizaguirre et al., 2012; Kyle et al., 2014). We cannot
418 fully exclude the possibility of WNS selecting for alleles in our scaffold of interest though. It
419 is likely that there is concurrent selection occurring across multiple populations, and that
420 selection is not acting to the same degree or even on the same loci in different geographic areas
421 affected by WNS (Messer and Petrov, 2013). This also applies to interpreting the results on a
422 smaller scale: the samples had not been collected from exactly the same locations before and
423 after WNS, which may contribute to our results. Individuals sampled during the winter, at their
424 hibernacula, may not inhabit the vicinity of these locations during the summer, although
425 dispersal distances are often short (Davis & Hitchcock, 1965; but see Norquay et al., 2013).
426 Understanding how local separate sub-populations cope with pathogen pressure would benefit
427 from an RNAseq-approach identifying responses at transcript-level (Lilley et al., 2019).

428 The ankyrin 2 protein encoding gene, lying within the observed small scaffolds on the right-
429 hand side of the $F_{ST}$ plots has a wide range of attributed Gene Ontology functions (The UniProt
430 Consortium, 2019), which reinforce the ubiquitous nature and far-reaching roles of ankyrin-
431 repeat-containing genes across organisms (Voronin and Kiseleva, 2007). Alternatively, the
432 peaks in this area may have arisen because only relatively few SNPs were called there given
433 the small size of the scaffolds, thus limiting our ability to generate high-quality windows, hence
434 the greater sampling noise.

435 Despite not seeing an overarching genetic signal associated with resistance in *M. lucifugus*,
436 remnant populations appear to be surviving year after year and even increasing in size
437 notwithstanding infection loads during the hibernation period (Langwig et al., 2017; Lilley et
438 al., 2016). The fact that the intensity of infection with *P. destructans* appears to be significantly
439 lower for remnant populations compared to colonies in the midst of the epidemic phase and
440 massive declines may suggest the development of resistance in these populations (Bernard et
441 al., 2017; Langwig et al., 2017; but see Lilley et al., 2019). However, the role of environmental
442 interactions during the hibernation period, when the bat hosts are most vulnerable to the
443 pathogen, have been largely neglected, although the effect of humidity and temperature on the
444 growth of the fungal pathogen and survival of the host have been separately studied (Grieneisen
445 et al., 2015; Marroquin et al., 2017; Verant et al., 2012). Modelling also suggests that
446 environmental factors, including the cave microbiome, have an impact on the proliferation and
447 infectivity of *P. destructans* (Hayman, Pulliam, Marshall, Cryan, & Webb, 2016; Lilley et al.,
448 2018). This is supported by the discovery of microbes in hibernacula environments and bats
449 that are able to retard the growth of the fungus (Micalizzi et al., 2017; Zhang et al., 2014, 2015).
450 In light of our results and considering bat life history as a whole, adaptation and evolutionary

451 rescue may not be a fast track for recovery of bat populations affected by WNS (Maslo and
452 Fefferman, 2015), although ultimately these will contribute to long term survival of populations
453 (Lilley et al., 2019). At the present, it is more likely the variability in hibernation site
454 environmental conditions and strong selection for microbial taxa that inhibit *P. destructans*
455 could explain why populations are recovering from the zoonosis (Cheng et al., 2016; Lemieux-
456 Labonté, Simard, Willis, & Lapointe, 2017; Lilley et al., 2018).

457 It is still unclear how frequently genetic adaptation occurs in natural populations and under
458 what circumstances it is promoted (Schoville et al., 2012). Many studies have recorded
459 phenotypic changes in natural populations and attributed them to host-pathogen interactions
460 (Frank et al., 2014; Kilpatrick, 2006; Langwig et al., 2017; Råberg et al., 2009). However, it is
461 difficult to determine conclusively whether changes in phenotype are the result of selection in
462 the genome or a result of phenotypic plasticity (Paterson et al., 2010; Routtu and Ebert, 2015).
463 Also, the assumed benefit of change (i.e. adaptation) is often not tested experimentally, which
464 would allow for inference of causality with the pathogen through exclusion of other potential
465 drivers.

466 To conclude, our results suggest WNS has not yet decidedly subjected populations of *M.*
467 *lucifugus* to selective pressures leading to genetic adaptation in remnant populations. No
468 population -wide signs of selection in comparisons of genomes of Pre-WNS and ten years Post-
469 WNS populations were observed (but see Donaldson et al., 2017; Gignoux-Wolfsohn et al.,
470 2018). This is something that is not unexpected with a species in which the effective population
471 size most likely numbered in tens of millions of individuals prior to the onset of WNS. The
472 disease has not caused a true population bottle-neck, as exemplified by our measures of genetic

473 diversity not varying between pre- and post-WNS samples. Our results indicate that the
474 existence and recent growth of remnant populations of *M. lucifugus* are more likely attributable
475 to other factors, such as microbiome adaptation and hibernation site selection among others,
476 rather than genetic adaption. However, we found increased variability in a specific area of the
477 genome in a comparison of Post-WNS samples from our two different study states, relative to
478 Pre-WNS samples. We suggest this is due to weakened connectivity between bats at different
479 locations allowing for local adaptations to enrich in the absence of gene flow. Although genes
480 in the high $F_{ST}$ region of the genome were identified, and genes such as KITLG provide
481 interesting avenues of research, even with regards to WNS, further investigation into the
482 processes in bat life history to which these genes are related are required to expound upon the
483 existing gene annotations.

484 ## Acknowledgements

485 We thank Svenska Kulturfonden, H2020 Marie Skłodowska-Curie Actions (TML, 706196)
486 and the Natural Environment Research Council, UK for funding the study; staff at the Centre
487 for Genomic Research, University of Liverpool, UK for DNA sequencing expertise; Jenni
488 Prokkola and Riley F. Bernard for proofreading the manuscript; and Alice Lilley, Cali Wilson,
489 Beth Rogers and Mike Scafini for assisting in sample collection.

490 ## Author contributions

491 TML and SP designed the research. TML, DMR, KAF, AK, GGR and CH collected the
492 samples. TML prepared samples for sequencing. TML, IWW and SP analysed the data. TML,
493 IWW and SP wrote the first draft of the manuscript, and all authors contributed to the final
494 version.

## Data availability statement

Raw data are available at ENA accession XXXXX and bioinformatics code at https://github.com/scottishwormboy/myoluc.

## References

**Table captions**

Table 1. Samples used in study.

Table 2. Summary of individual SNP $F_{ST}$ values across the *M. lucifugus* genome when comparing Pre-WNS with Post-WNS; and samples from PA and NY (both Pre- and Post-WNS).

Table 3. Genes within the *M. lucifugus* genome – excluding scaffold GL429776 – with windows possessing $F_{ST}$ values >= 0.05 when comparing PA and NY samples Post-WNS

Table 4. Genes within scaffold GL429776 with windows possessing $F_{ST}$ values >= 0.05 when comparing PA and NY samples Post-WNS.

## Figure captions

Fig. 1. *M. lucifugus* sampling locations in Pennsylvania (PA) and New York (NY). Pre-WNS-sites in grey circles and Post-WNS-sites in white circles. 1) Aitkin Cave, Mifflin County, PA (n=3); 2) Snyder County, Penn, PA (n=19); 3) Lewisburg, Union Co, PA (n=3); 4) Northumberland County, PA (n=2); 5) Howes Cave, County Schoharie, NY (n=24); 6) Allentown, Lehigh County, PA (n=32); 7) Woodward, Centre County, PA (n=14); 8) Alexandria, Huntingdon County, PA (n=20); 9) Lackawaxen, Pike County, PA (n=4); 10)

515 Oley, County Berks, PA (n=12); 11) Frankstown, County Blair, PA (n=12); 12) Williams
516 Preserve Mine, NY (n=16); 13) Hoopers Mine, NY (n=24); 14) Delmar, Albany County, NY
517 (n=36); and 15) Willsboro, Essex County, NY (n=21).

518 Fig. 2. Fixation indices ($F_{ST}$) across the *M. lucifugus* genome, comparing the population before
519 and after the arrival of white-nose syndrome. Scaffold lengths presented in green and blue.
520 Solid red line indicates cut-off for $F_{ST}$ values of 5 standard deviations from the mean across
521 the whole genome; dotted red line indicates cut-off of $F_{ST}$=0.05.

522 Fig 3. Fixation indices ($F_{ST}$) across the *M. lucifugus* genome between geographically separated
523 individuals of the study population a) comparing individuals from PA to individuals from NY
524 before the arrival of WNS; b) comparing individuals from PA before and after the arrival of
525 WNS; and c) comparing individuals from NY before and after the arrival of WNS. Solid red
526 line indicates cut-off for $F_{ST}$ values of 5 standard deviations from the mean across the whole
527 genome; dotted red line indicates cut-off of $F_{ST}$=0.05.

528 Fig 4. Fixation indices ($F_{ST}$) across the *M. lucifugus* genome between geographically separated
529 individuals of the study population a) comparing individuals from PA to individuals from NY
530 after the arrival of WNS, with a zoomed plot of scaffold GL429776 and gene models therein
531 containing windows demonstrating $F_{ST}$ values < 0.05; and b) comparing individuals from PA
532 to individuals from NY with Pre- and Post-WNS data combined. Solid red line indicates cut-
533 off for $F_{ST}$ values of 5 standard deviations from the mean across the whole genome; dotted red
534 line indicates cut-off of $F_{ST}$=0.05.

535 Fig 5. Proportions of reference alleles called at individual SNP sites for the region of scaffold
536 GL429776 demonstrating high $F_{ST}$ values when comparing a) individuals from Pennsylvania

537 before and after the arrival of WNS; b) individuals from NY before and after the arrival of
538 WNS; c) individuals from PA to individuals from NY before the arrival of WNS; and d)
539 individuals from PA to individuals from NY after the arrival of WNS.

540 Table 1.

| Sample age | Classification | Total number of samples | Pennsylvania* | New York* |
|---|---|---|---|---|
| 2006 | Pre-WNS | 49 | 25 | 24 |
| 2007 | Pre-WNS | 3 | 3 | |
| 2015 | Post-WNS | 28 | 28 | |
| 2016 | Post-WNS | 139 | 82 | 57 |

* WNS first detected in New York in 2007 and PA in 2009

541

Table 2.

| | **Pre- vs post-WNS** | **Between site (combining pre- and post-WNS)** |
|---|---|---|
| **Median (Interquartile range)** | 0.0059 (0.0013 – 0.0171) | 0.0057 (0.0013 – 0.0165) |
| **Mean (s.d.)** | 0.0127 (0.0173) | 0.0123 (0.0170) |

Table 3.

| Scaffold | Location range (bp) | Gene ID | Protein Function | Number of overlapping high $F_{ST}$ windows |
|---|---|---|---|---|
| NW_005871715.1 | 288334 - 351644 | 102435296 | bifunctional heparan sulfate N-deacetylase/N-sulfotransferase 4-like (partial) | 4 |
| NW_005871771.1 | 26662 - 48938 | 102441697 | E3 ubiquitin-protein ligase TRIM7 | 1 |
| NW_005871823.1 | 81353 - 213965 | 102428352 | Ankyrin 2 | 7 |
| NW_005871905.1 | 175 - 139344 | 102428043 | calcium/calmodulin-dependent protein kinase type II subunit delta-like | 10 |

| | | | | |
|---|---|---|---|---|
| NW_005872026.1 | 72042 - 101269 | 102443179 | N-deacetylase and N-sulfotransferase 4 (partial) | 1 |
| NW_005872062.1 | 33266 - 86607 | 102428048 | zinc finger GRF-type containing 1 | 1 |

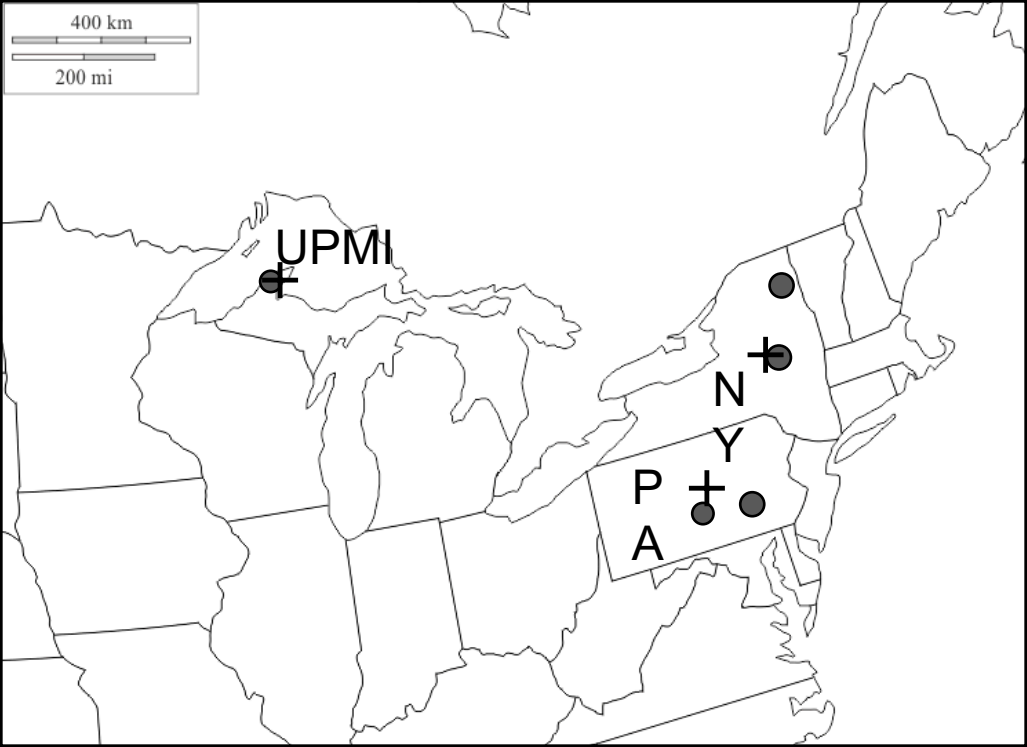
Table 4.

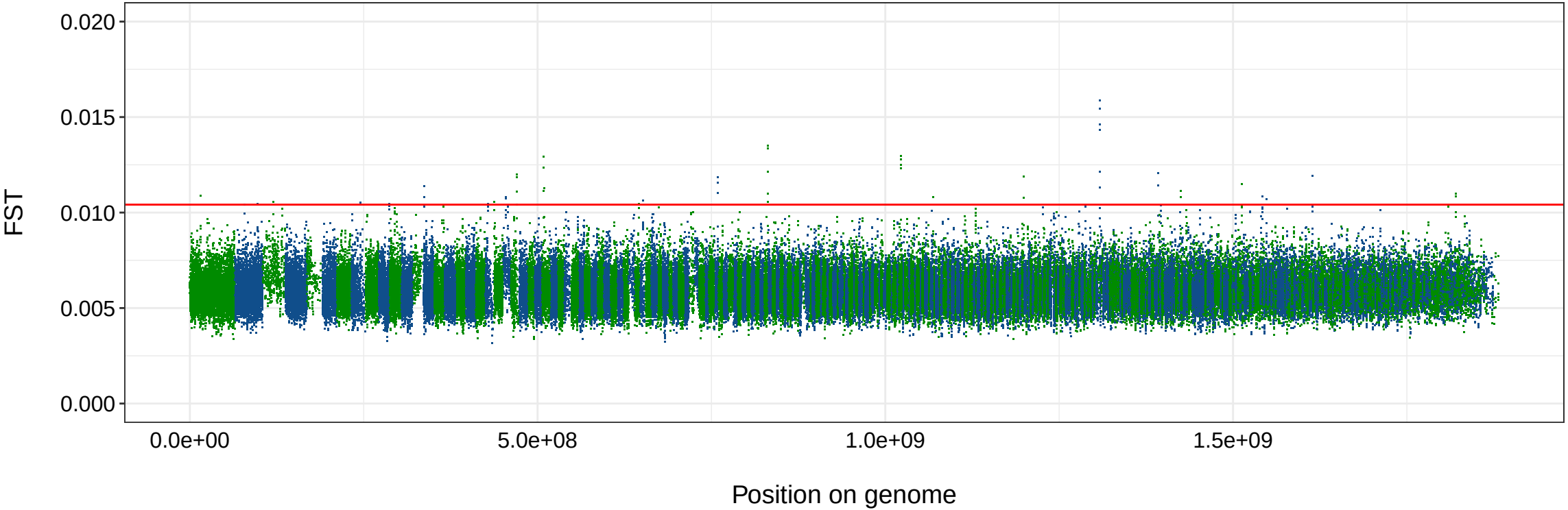
| Gene ID | $F_{ST}$ | No of windows | No of windows | Start pos. | End pos. | Details |
|---|---|---|---|---|---|---|

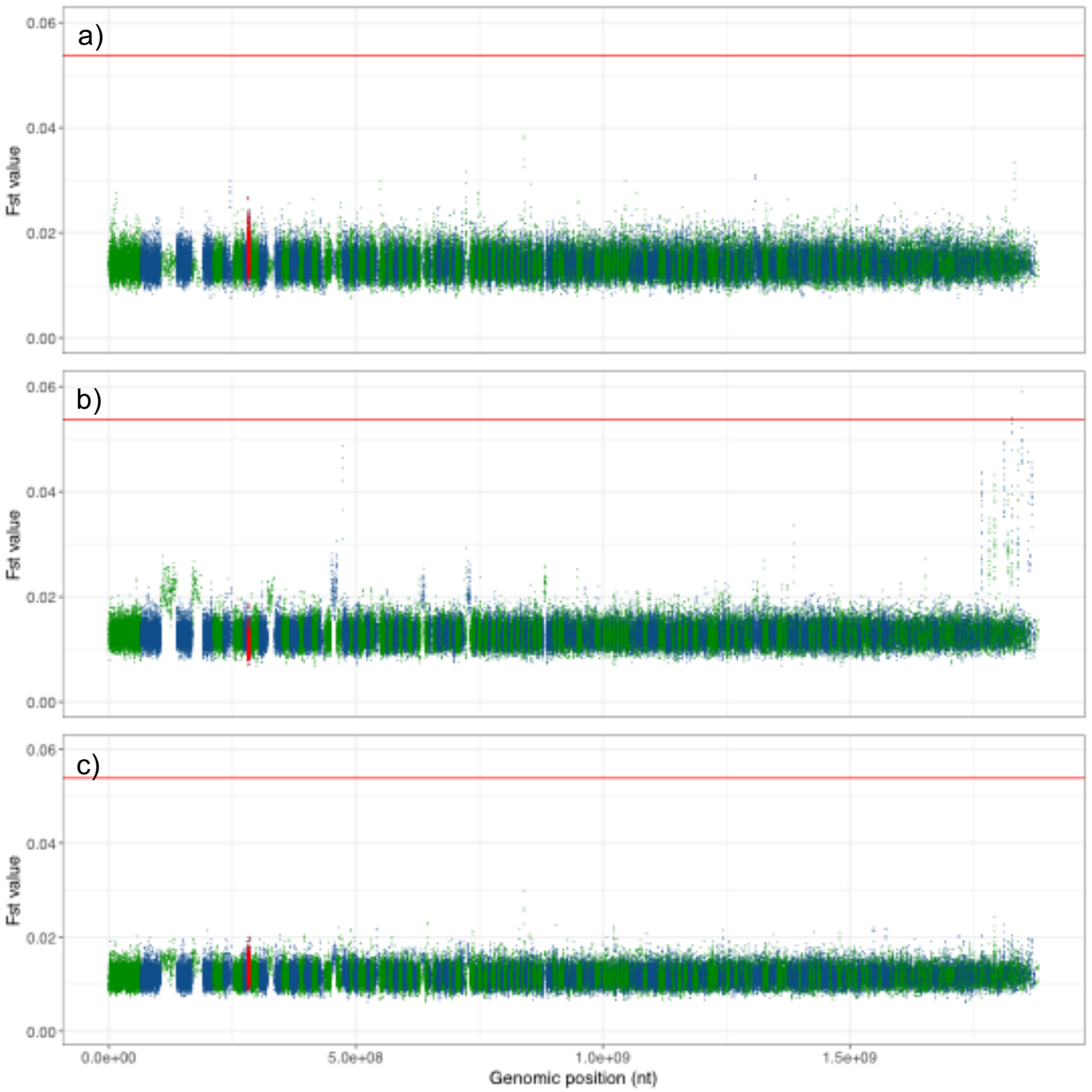
| | | | | | | |
|---|---|---|---|---|---|---|
| | | | **(per 10kb)** | | | |
| MGAT4C | 0.0612 | 78 | 1.08 | 8999210 | 9724150 | Glycosyltransferase that participates in the transfer of N-acetylglucosamine (GlcNAc) to the core mannose residues of N-linked glycans |
| CEP290 | 0.0679 | 12 | 1.01 | 10709418 | 10827912 | Involved in early and late steps in cilia formation |
| ATP2B1 | 0.0826 | 12 | 0.89 | 12228178 | 12362389 | The protein encoded by this gene belongs to the family of P-type primary ion transport ATPases characterised by the formation of an aspartyl phosphate intermediate during the reaction cycle. These enzymes remove bivalent calcium ions from eukaryotic cells against very large concentration gradients and play a critical role in intracellular calcium homeostasis. |

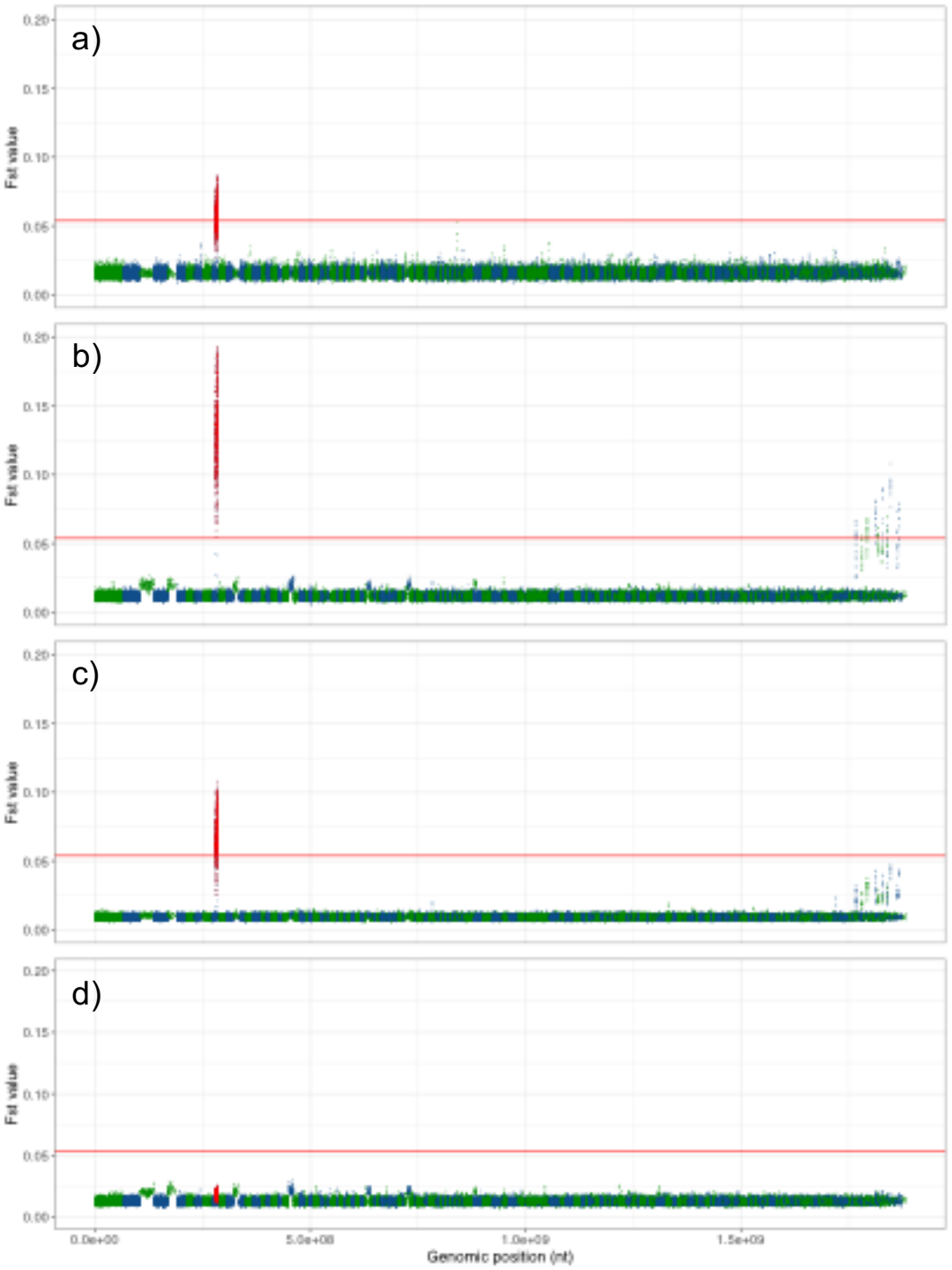
| | | | | | | |
|---|---|---|---|---|---|---|
| KITLG | 0.0706 | 10 | 1.13 | 11127130 | 11215731 | Plays an essential role in the regulation of cell survival and proliferation, hematopoiesis, stem cell maintenance, gametogenesis, mast cell development, migration and function, and in melanogenesis. |
| POC1B | 0.0686 | 7 | 0.74 | 12044678 | 12139811 | Plays an important role in centriole assembly and/or stability and ciliogenesis |
| TMTC3 | 0.0872 | 6 | 1.00 | 10827719 | 10887502 | Involved in the positive regulation of proteasomal protein degradation in the endoplasmic reticulum (ER), and the control of ER stress response. |
| GeneID:111824954 | 0.0693 | 5 | 1.14 | 11756928 | 11800690 | uncharacterised LOC111824954 |
| CUNH12orf50 | 0.0588 | 2 | 0.67 | 10663328 | 10693373 | chromosome unknown C12orf50 homolog |
| GeneID:111824948 | 0.0816 | 2 | 1.13 | 11375011 | 11392720 | uncharacterised LOC111824948 |

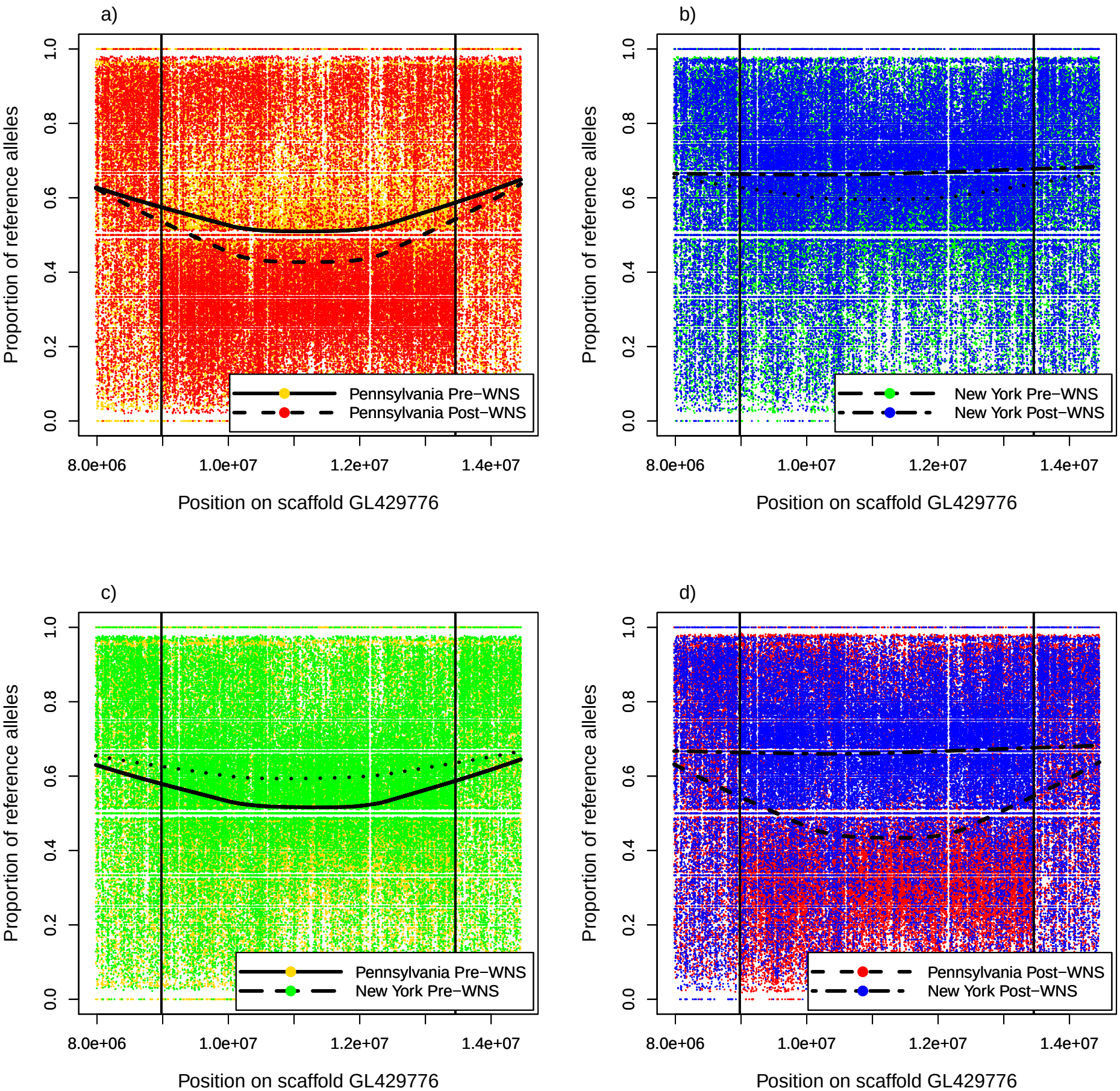
| | | | | | | |
|---|---|---|---|---|---|---|
| GeneID:111824949 | 0.0635 | 2 | 1.14 | 11802704 | 11820273 | uncharacterised LOC111824949 |
| CUNH12orf29 | 0.0488 | 1 | 0.71 | 10696290 | 10710421 | chromosome unknown C12orf29 homolog |
| DUSP6 | 0.0603 | 1 | 2.26 | 11976039 | 11980455 | Inactivates MAP kinases. |
| GeneID:111824955 | 0.0459 | 1 | 1.57 | 12101754 | 12108133 | uncharacterised LOC111824955 |


400 km
200 mi
UPMI
NY
PA


0.020
0.015
0.010
0.005
0.000
FST
0.0e+00
5.0e+08
1.0e+09
1.5e+09
Position on genome


a)
b)
c)
Fst value
0.06
0.04
0.02
0.00
0.0e+00
5.0e+08
1.0e+09
1.5e+09
Genomic position (nt)


a)
b)
c)
d)
Fst value
0.20
0.15
0.10
0.05
0.00
0.0e+00
5.0e+08
1.0e+09
1.5e+09
Genomic position (nt)


a)
b)
c)
d)
Proportion of reference alleles
0.0
0.2
0.4
0.6
0.8
1.0
8.0e+06
1.0e+07
1.2e+07
1.4e+07
Position on scaffold GL429776
Pennsylvania Pre−WNS
Pennsylvania Post−WNS
New York Pre−WNS
New York Post−WNS