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], Vodzak ME[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

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

The psychrophilic (cold-growing) fungus *P. destructans*, (Lorch et al., 2011; Minnis and Lindner, 2013; Warnecke et al., 2008) acts as an opportunistic pathogen of bats, invading the skin tissues of hibernating hosts (Cryan et al., 2010; Meteyer et al., 2012). Susceptible species, such as *Myotis lucifugus* and *M. septentrionalis*, have shown population declines

greater than 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).

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

Here, we utilise high-throughput whole-genome sequencing on the WNS-affected species *M. lucifugus*, allowing us to look at entire genomes via single nucleotide polymorphisms (SNPs).

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

2. Materials and methods

2.1. Ethics Statement

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

2.2. Sample collection and DNA extraction

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

of the greater number of polymorphic sites one would expect to arise from a more diverse collection of individuals (see Section 2.3.). Tissue samples were collected using either 3.0 mm or 5.0 mm biopsy punches (Integra Milltex, Plainsboro, NJ) and stored in 95% ethanol until extraction. We extracted DNA using QIAmp DNA Mini Kits (Qiagen, Hilden Germany) and stored DNA at -80°C until sequencing.

2.3. Sequencing methods

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

2.4. Bioinformatics methods

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

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

Average heterozygosity within the pooled samples was calculated using npstats (Ferretti et al., 2013), which controls for number of individuals within each sequencing pool and sequencing depth, thus allowing robust comparisons between pooled samples. One million

SNPs were randomly picked and permuted, with 1000 sets of 1000 SNPs used to calculate the observed heterozygosity within each pooled set of individuals and, from these, we produced the median and 95% confidence intervals on heterozygosity.

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

From the comparison of the NY and PA pooled samples (i.e. the comparison that produced the highest F_{ST} values), we visually identified across the whole genome windows with median F_{ST} values greater than 5 standard deviations from the mean, as well as those with F_{ST} values greater than 0.05, indicating at least 'moderate' genetic differentiation (Wright, 1978). The numbers of such complete windows within gene models were counted for each gene to identify genes potentially under selective pressure.

3. Results

3.1. Pre-WNS vs Post-WNS comparisons

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

3.2. Comparisons within geographical locations

Due to our sampling and sequencing setup, we were also able to separately look at individuals of the same population from two discrete geographical locations within their range. In our case a divide was made between bats sampled in New York and Pennsylvania based on geographical distance between the sampling sites. Although population genetic analysis based on a considerably lower number of genetic markers assigns *M. lucifugus* on the eastern coast of North America to the same population (Vonhof et al., 2015), the species shows high degrees of philopatry (Davis and Hitchcock, 1965; Norquay et al., 2013), warranting the examination of these arbitrarily assigned pools of individuals outside the mean dispersal range of the species using more accurate, whole-genome sequencing based population genetic analyses.

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

3.3. Post-WNS comparisons between geographically isolated sample sets

When comparing the Post-WNS individuals sampled from PA with the Post-WNS individuals from NY, the peak on scaffold GL429835 is no longer visible, but a far more

pronounced spike in F_{ST} values can be clearly seen in scaffold GL429776 (Fig. 4a), indicating relatively large allele frequency differences between the sample sets in this region of the genome. This is also seen, to a lesser extent, when comparing all samples from NY with all samples from PA (Fig. 4b). However, this significant peak in this comparison of all samples collected does not affect average fixation indices across the genome relative to the other comparison of all samples, Pre-WNS vs Post-WNS (Table 2). This peak could be attributed to mapping error, but our coverage depths show an average depth of 66.33x across the whole genome, whereas the focal region was sequenced at an average depth of 66.49x. This means that there has been no collapsing of duplicated regions, nor poor coverage of the region. Also, viewing the data on the NCBI genome viewer tool, with RepeatMasker histogram enabled, suggested no immediately obvious increase in repetitive sequence in the region of interest.

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

To clarify why we saw no fixation at this site when comparing Pre-WNS and Post-WNS populations (Fig. 3a), nor when we compared Pre-WNS samples from both geographical origins (Figs. 3b and 3c), we calculated for the relevant subpopulations the proportion of reference alleles called at each site included in the F_{ST} window analyses (Fig. 5). This revealed a small decrease over time in the proportion of the reference alleles seen at this locus in the PA subpopulation (Fig. 5a), commensurate with an increase in proportion of

reference alleles in the NY sample set (Fig. 5b). As such, existing differences between the NY and PA sample sets were insufficient to cause a relative rise in fixation indices prior to WNS spreading through the population (Fig 5c). Conversely, Post-WNS, allele proportions at this site in both sample sets had not changed sufficiently to suggest selection *within* the sample set but demonstrated clear differences *between* the NY and PA sample sets (Fig. 5d).

In the Post-WNS comparison, peaks which were above our arbitrary threshold of 5 standard deviations above the mean across the whole genome, but below F_{ST} = 0.05, can also be observed in small scaffolds on the right-hand side of the F_{ST} plots. The 63 windows comprising these peaks lie in 47 scaffolds. Of these, high F_{ST} windows lie in only 6 protein-coding regions across 6 scaffolds (Table 3). Two of these encode partial gene models. Of the 4 remaining models containing high F_{ST} regions, all but 1 is described on NCBI as a "Low Quality Protein". The exception is a gene encoding the ankyrin 2 protein, within which 7 high F_{ST} windows lie.

3.4. Measures of heterozygosity within subpopulations

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

4. Discussion

Examining our data as a whole, we found low F_{ST} values between the Pre- and Post-WNS individuals of our eastern North American *M. lucifugus* population, the absence of fixation

suggesting that WNS has not decreased the population in size sufficiently for the disease to have an effect on genetic diversity. Also, we found no significant changes in within-population heterozygosity Post-WNS, despite the massive population declines caused by WNS over the past decade, and found overall high heterozygosity across the genome. These results are in accordance with a similar, low-coverage study conducted by Gignoux-Wolfsohn et al. (2018). Panmixia of populations in evolutionary time in bats to the east of the Rocky Mountains has been previously documented (Vonhof et al., 2016; Wilder et al., 2015). Gignoux-Wolfsohn et al. (2018) present evidence for a soft selective sweep, where selection acts on standing genetic variation, between Pre- and Post-WNS populations. The analysis is based on a change in minor allele frequencies due to selection, distinguished from background demography using a simulation approach; and it shows a complex picture of potentially up to 85 unlinked SNPs, each exhibiting small effects and with little or no selective signal at adjacent SNPs. In contrast, a recent case study of selection on standing genetic variation, by myxoma virus in rabbits, showed that soft selective sweeps caused concerted changes in allele frequencies at sets of linked SNPs around coding loci (Alves et al., 2019). The robust pooled sequencing approach allows us to examine for similar, larger scale changes in the genome. We acknowledge that both methods have their advantages and disadvantages, such as the increased possibility for false outliers in our pooled approach (Li et al., 2008). Hovewer we took these into account with parameters used for aligning the reads (Kofler et al., 2011), sequencing depth (Ferretti et al., 2013; Kofler and Schlötterer, 2014) pool size (Futschik and Schlötterer, 2010; Gautier, 2015) and library preparation protocol (Kofler et al., 2016).

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

Due to our sampling design, we were able to separate bats geographically within our sampling range. We made a cut-off point for a more in-depth investigation of bats sampled in the states of Pennsylvania (PA) compared to New York (NY). Although the cut-off may seem rather arbitrary, the reasoning behind this is geographic distance between sampling sites, and diving the data between the two states is appropriate for our purposes as evidenced by the results. Here, we see low F_{ST} values between the two groups of bats prior to the onset of WNS, indicating an absence of selection across the genome, with the exception of a small peak in scaffold GL429835, which lies within a gene model encoding the NIPBL cohesion loading factor. Relatively high F_{ST} values are also visible within this region when comparing the PA population before and after the spread of WNS, as well as the NY population pre- and post-WNS. However, this peak is not evident when comparing PA and NY post-WNS (Fig. 4a). This suggests that this coding sequence was under selective pressures, manifesting as differentiating allelic frequencies, in both subpopulations prior to the arrival of WNS, but that any selection was nullified by the population-level impact of WNS (or another event occurring during the same timeframe). In addition, this result confirms that although the pre-WNS bats from NY were from the initial batch of bats found dead of WNS at Howe's cave in

2006, they were genetically similar to pre-WNS bats from PA, which were sampled before the arrival of WNS. This is most likely because of the high frequency of initial mortality in the population, which indiscriminately wiped out most of the population, eroding any genetic differentiation. As for the protein NIPBL, it plays an important role in the function of the cohesin complex, which is vital for chromosome segregation, DNA damage repair, and gene expression regulation (Peters et al., 2008). As these results suggest the loss of low-level functional diversity at this locus in our subpopulations, it would be interesting to further investigate what impact this will have on the population as a whole.

The F_{ST} comparison of Post-WNS individuals sampled from PA and NY depicts considerable genetic differentiation within scaffold GL429776; differentiation which did not exist in the Pre-WNS comparison of individuals sampled from these geographical areas. This is a difference which would have gone unnoticed had we not made the comparison between bats sampled in NY and PA, or used fewer genetic markers instead of a whole-genome approach (Miller-Butterworth et al., 2014; Vonhof et al., 2015). The region's appearance post-WNS when considering all geographical locations together is not, however, reflected when the data are broken down into representatives of each geographic location. A comparison of allele frequencies of individual SNPs within this region shows that individuals sampled from PA showed an increase in deviation from the reference genome over time, whilst a greater proportion of reference alleles were seen in NY samples Post-WNS compared with Pre-WNS. The differences in proportions, growing as they were over time, were insufficient to exhibit a high fixation index Pre-WNS, but were notably different after the disease had spread through the population. One could argue, that this is due to sampling at maternal colonies and swarming sites, which has known effects on allele frequency differences

(Johnson et al., 2015). However, we sampled at a number of maternal colonies, which would dilute this effect, and furthermore, recent studies suggest philopatry within maternal colonies may not be as high as previously expected (Olivera-Hyde et al., 2019, although on *M. septentrionalis*). Therefore we assume our sampling of multiple maternal colonies and hibernation sites has removed any effects of relatedness within our sample sets.

This region within the scaffold containing the peak contains six annotated genes with very different functions. Two, POC1B and CEP290, are involved in cilia development and maintenance. The ATP2B1 gene encodes a protein which belongs to the family of P-type primary ion transport ATPases, whereas TMTC3 is involved in the control of endoplasmic reticulum stress response. MGAT4C is a glycosyltransferase that participates in the transfer of N-acetylglucosamine (GlcNAc) to the core mannose residues of N-linked glycans. Finally, KITLG produces a protein involved in mast cell development, migration and function, and melanogenesis. KITLG has been documented as a target of evolution in recent studies and in humans it has been linked with thermoregulation (Yang et al., 2018), which is a critical component of bat life history (Studier and O'Farrell, 1972). In fact, upregulation of KITLG at low temperature helps promote the production of brown fat for heat generation (Huang et al., 2014). Bats utilise brown adipose tissue in arousals to normothermia from different degrees of torpor (Thomas et al., 1990). There is also evidence of parallel evolution of pigmentation in sticklebacks and humans linked to KITLG (Miller et al., 2007), as well as its role in determining resistance/susceptibility to swine respiratory diseases in Erhulian pigs (Huang et al., 2017). KITLG, by influencing mast cell development, could also play a role in either protective or pathological immune responses to *P. destructans*, which may involve IgE-mediated recognition of secreted fungal proteins. As such, temporal changes we found in

allele frequencies at this locus could, for instance, be linked to differences in hibernation site temperatures, and/or thermoregulatory needs in bats sampled at different locations (Humphries et al., 2002; Thomas et al., 1990). Spatial patterns, as seen here between the arbitrarily pooled samples from PA and NY, could be due to local adaptation, in which strong selective sweeps may be linked to gene variants favoured in local interactions (Hansen et al., 2012; Kyle et al., 2014; Rico et al., 2015; Schoville et al., 2012). Whilst we lack data concerning migration patterns of the bats studied here, as well as the environmental conditions within their microfugia, future work would benefit from these data in a bid to understand whether bats repeatedly frequent habitats with similar conditions and whether, indeed, there is a significant difference in conditions across these Eastern states that explains the rapid rise in allelic differentiation, and thus F_{ST} values, at this locus.

The increase in F_{ST} values in this scaffold GL429776 could, alternatively, be due to the massive population decline caused by WNS (Langwig et al., 2017). Although not supported by existing data, the decline could potentially have resulted in a decrease in connectivity and gene flow between remnant local populations allowing for a local adaptation to arise (Bohonak, 1999; Lilley, Anttila, & Ruokolainen, 2018). Genetically resolved examples of local adaptations are rare (Hoekstra et al., 2006) and, in our focal system, pinpointing the causality leading to local adaptation is especially difficult due to the cryptic life history of bats. Determining how these variants are spread or localised among populations is essential to understanding and managing the emergence of new selective pressures (Eizaguirre et al., 2012; Kyle et al., 2014). We cannot fully exclude the possibility of WNS selecting for alleles in our scaffold of interest though. It is likely that there is concurrent selection occurring across multiple populations, and that selection is not acting to the same degree or even on the

same loci in different geographic areas affected by WNS (Messer and Petrov, 2013). This also applies to interpreting the results on a smaller scale: the samples had not been collected from exactly the same locations before and after WNS, which may contribute to our results. Individuals sampled during the winter, at their hibernacula, may not inhabit the vicinity of these locations during the summer, although dispersal distances are often short (Davis & Hitchcock, 1965; but see Norquay et al., 2013). Understanding how local separate sub-populations cope with pathogen pressure would benefit from an RNAseq-approach identifying responses at transcript-level (Lilley et al., 2019).

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

Despite not seeing an overarching genetic signal associated with resistance in *M. lucifugus*, remnant populations appear to be surviving year after year and even increasing in size notwithstanding infection loads during the hibernation period (Langwig et al., 2017; Lilley et al., 2016). The fact that the intensity of infection with *P. destructans* appears to be significantly lower for remnant populations compared to colonies in the midst of the epidemic phase and massive declines may suggest the development of resistance in these populations (Bernard et al., 2017; Langwig et al., 2017; but see Lilley et al., 2019). However,

the role of environmental interactions during the hibernation period, when the bat hosts are most vulnerable to the pathogen, have been largely neglected, although the effect of humidity and temperature on the growth of the fungal pathogen and survival of the host have been separately studied (Grieneisen et al., 2015; Marroquin et al., 2017; Verant et al., 2012). Modelling also suggests that environmental factors, including the cave microbiome, have an impact on the proliferation and infectivity of *P. destructans* (Hayman, Pulliam, Marshall, Cryan, & Webb, 2016; Lilley et al., 2018). This is supported by the discovery of microbes in hibernacula environments and bats that are able to retard the growth of the fungus (Micalizzi et al., 2017; Zhang et al., 2014, 2015). In light of our results and considering bat life history as a whole, adaptation and evolutionary rescue may not be a fast track for recovery of bat populations affected by WNS (Maslo and Fefferman, 2015), although ultimately these will contribute to long term survival of populations (Lilley et al., 2019). At the present, it is more likely the variability in hibernation site environmental conditions and strong selection for microbial taxa that inhibit *P. destructans* could explain why populations are recovering from the zoonosis (Cheng et al., 2016; Lemieux-Labonté, Simard, Willis, & Lapointe, 2017; Lilley et al., 2018).

It is still unclear how frequently genetic adaptation occurs in natural populations and under what circumstances it is promoted (Schoville et al., 2012). Many studies have recorded phenotypic changes in natural populations and attributed them to host-pathogen interactions (Frank et al., 2014; Kilpatrick, 2006; Langwig et al., 2017; Råberg et al., 2009). However, it is difficult to determine conclusively whether changes in phenotype are the result of selection in the genome or a result of phenotypic plasticity (Paterson et al., 2010; Routtu and Ebert, 2015). Also, the assumed benefit of change (i.e. adaptation) is often not tested

experimentally, which would allow for inference of causality with the pathogen through exclusion of other potential drivers.

To conclude, our results suggest WNS has not yet decidedly subjected populations of *M. lucifugus* to selective pressures leading to genetic adaptation in remnant populations. No population -wide signs of selection in comparisons of genomes of Pre-WNS and ten years Post-WNS populations were observed (but see Donaldson et al., 2017; Gignoux-Wolfsohn et al., 2018). This is something that is not unexpected with a species in which the effective population size most likely numbered in tens of millions of individuals prior to the onset of WNS. The disease has not caused a true population bottle-neck, as exemplified by our measures of genetic diversity not varying between pre- and post-WNS samples. Our results indicate that the existence and recent growth of remnant populations of *M. lucifugus* are more likely attributable to other factors, such as microbiome adaptation and hibernation site selection among others, rather than genetic adaption. However, we found increased variability in a specific area of the genome in a comparison of Post-WNS samples from our two different study states, relative to Pre-WNS samples. We suggest this is due to weakened connectivity between bats at different locations allowing for local adaptations to enrich in the absence of gene flow. Although genes in the high F_{ST} region of the genome were identified, and genes such as KITLG provide interesting avenues of research, even with regards to WNS, further investigation into the processes in bat life history to which these genes are related are required to expound upon the existing gene annotations.

Acknowledgements

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

Author contributions

TML and SP designed the research. TML, DMR, KAF, AK, GGR, MEV and CH collected the samples. TML prepared samples for sequencing. TML, IWW and SP analysed the data. TML, IWW and SP wrote the first draft of the manuscript, and all authors contributed to the final 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) Oley, County Berks, PA (n=12); 11) Frankstown, County Blair, PA (n=12); 12) Williams Preserve Mine, NY (n=16); 13) Hoopers Mine, NY (n=24); 14) Delmar, Albany County, NY (n=36); and 15) Willsboro, Essex County, NY (n=21).

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

Fig 3. Fixation indices (F_{ST}) across the *M. lucifugus* genome between geographically separated individuals of the study population a) comparing individuals from PA to individuals from NY before the arrival of WNS; b) comparing individuals from PA before

and after the arrival of WNS; and c) comparing individuals from NY before and after the arrival of WNS. Solid red line indicates cut-off for F_{ST} values of 5 standard deviations from the mean across the whole genome; dotted red line indicates cut-off of F_{ST}=0.05.

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

Fig 5. Proportions of reference alleles called at individual SNP sites for the region of scaffold GL429776 demonstrating high F_{ST} values when comparing a) individuals from Pennsylvania before and after the arrival of WNS; b) individuals from NY before and after the arrival of WNS; c) individuals from PA to individuals from NY before the arrival of WNS; and d) individuals from PA to individuals from NY after the arrival of WNS.

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

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

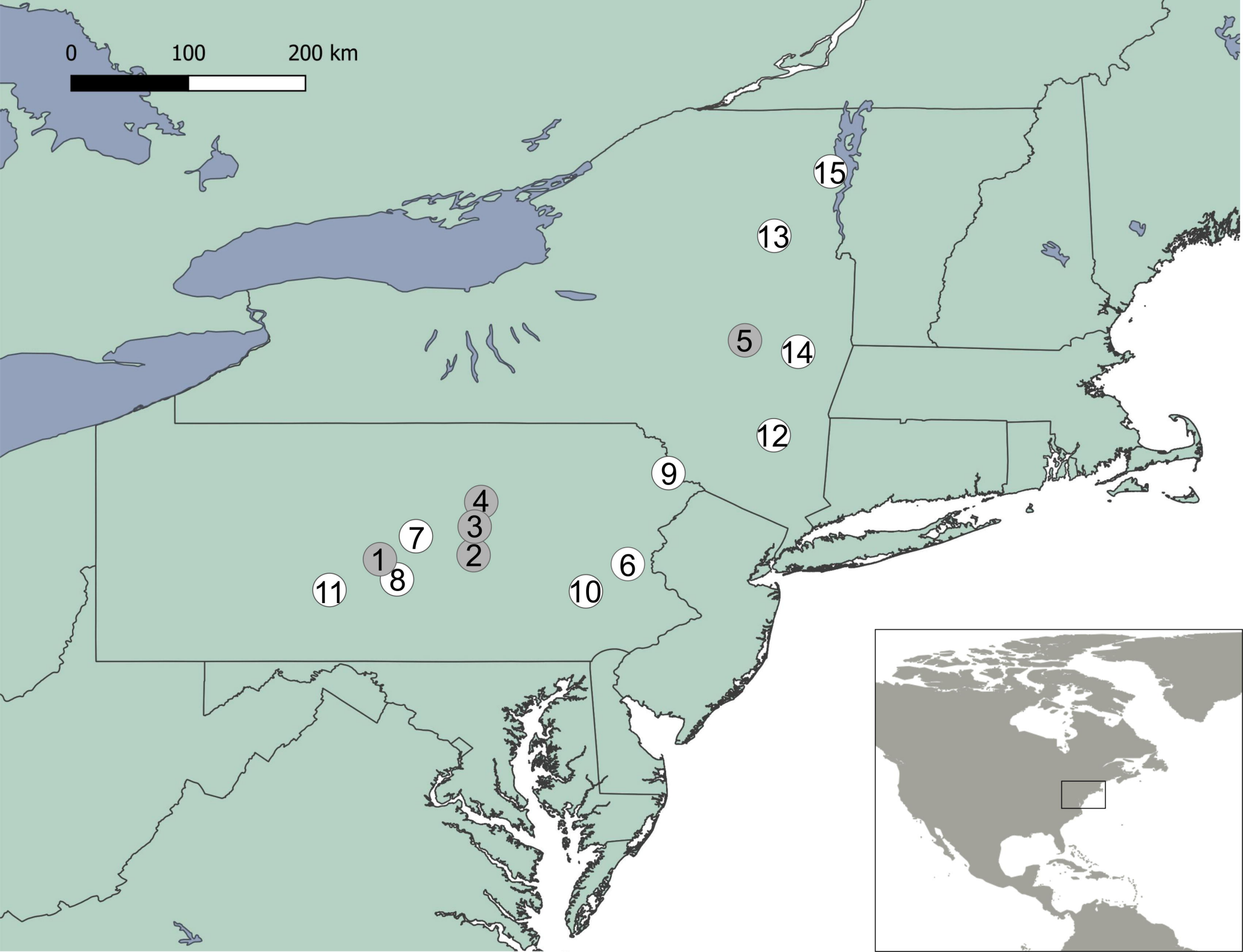
Table 4.

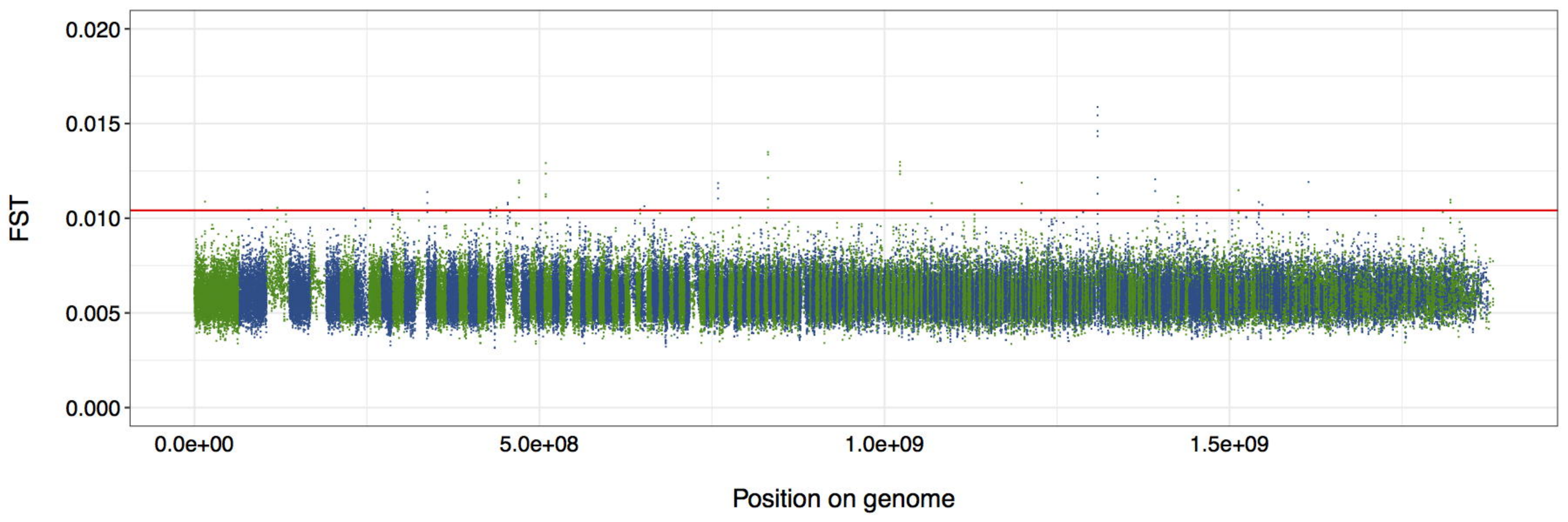
Gene ID	F_{ST}	No of windows	No of windows	Start pos.	End pos.	Details

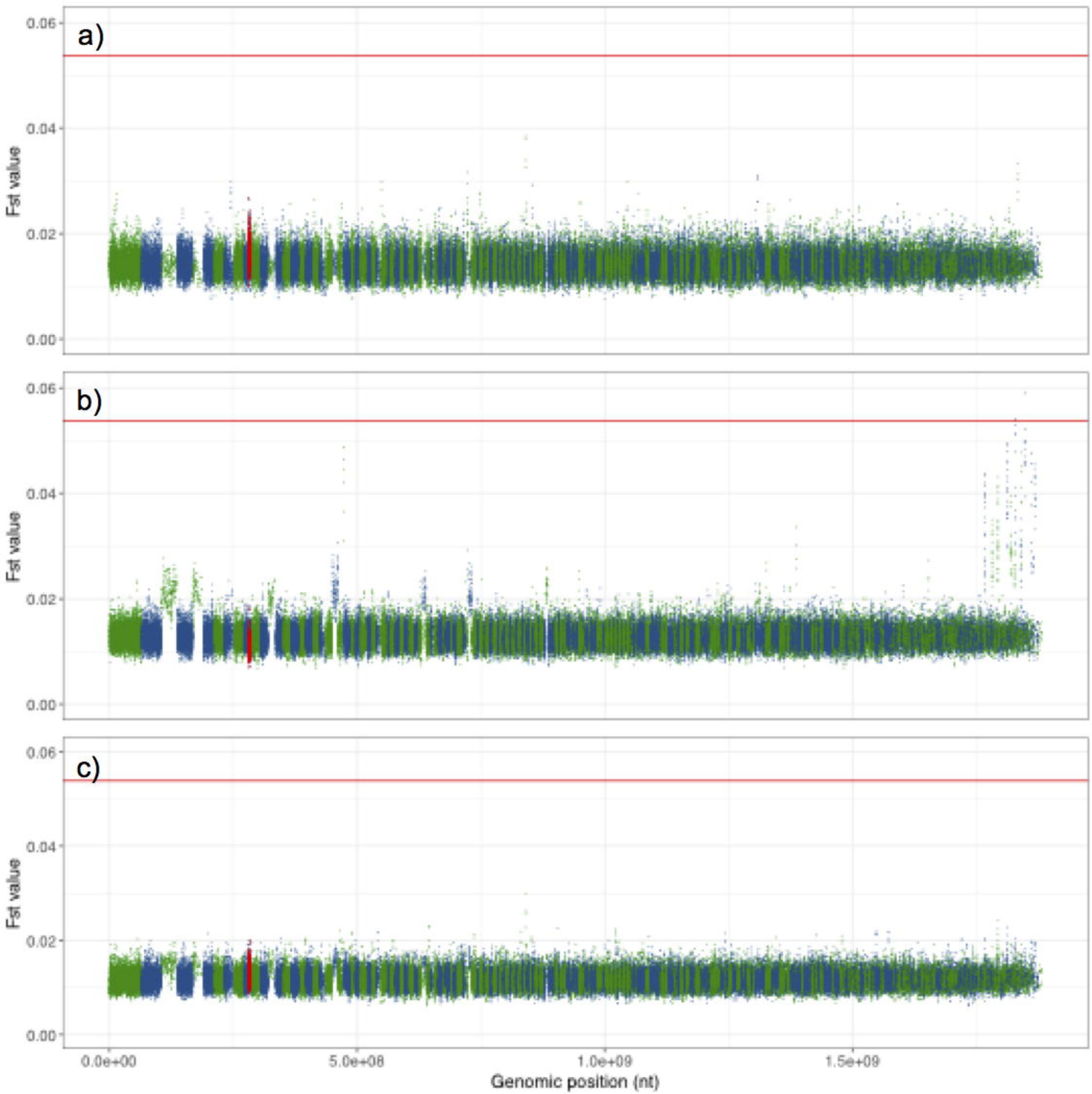
			(per 10kb)			
MGAT4C	0.0612	78	1.08	8999210	9724150	Glycosyltransferase that participates in the transfer of N-acetylglucosamine (GlcNAc) to the core mannose residues of N-linked glycans
CEP290	0.0679	12	1.01	10709418	10827912	Involved in early and late steps in cilia formation
ATP2B1	0.0826	12	0.89	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

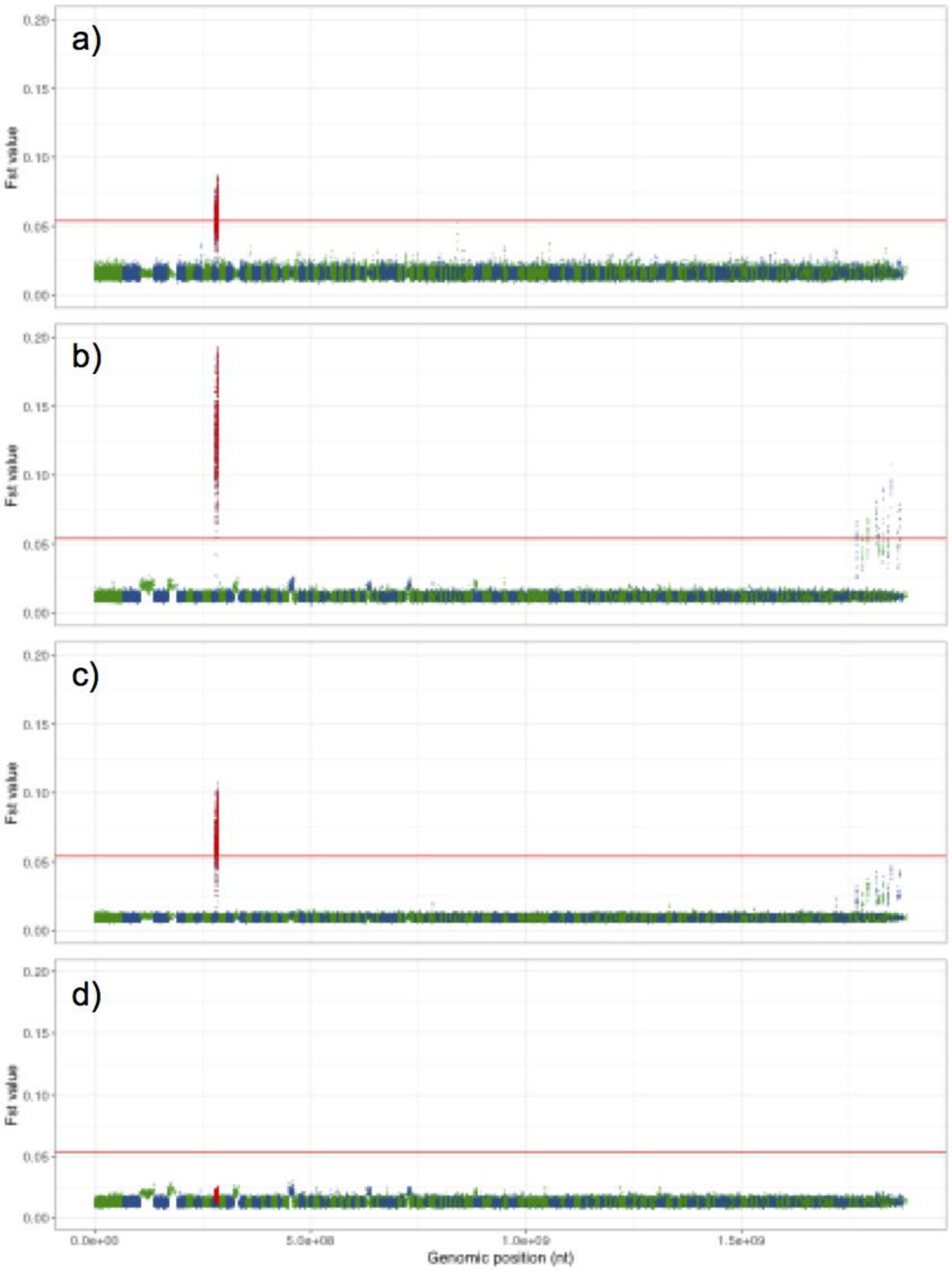
						development, migration and function, and in melanogenesis.
POC1B	0.0686	7	0.74	12044678	12139811	Plays an important role in centriole assembly and/or stability and ciliogenesis
TMTC3	0.0872	6	1.00	10827719	10887502	Involved in the positive regulation of proteasomal protein degradation in the endoplasmic 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

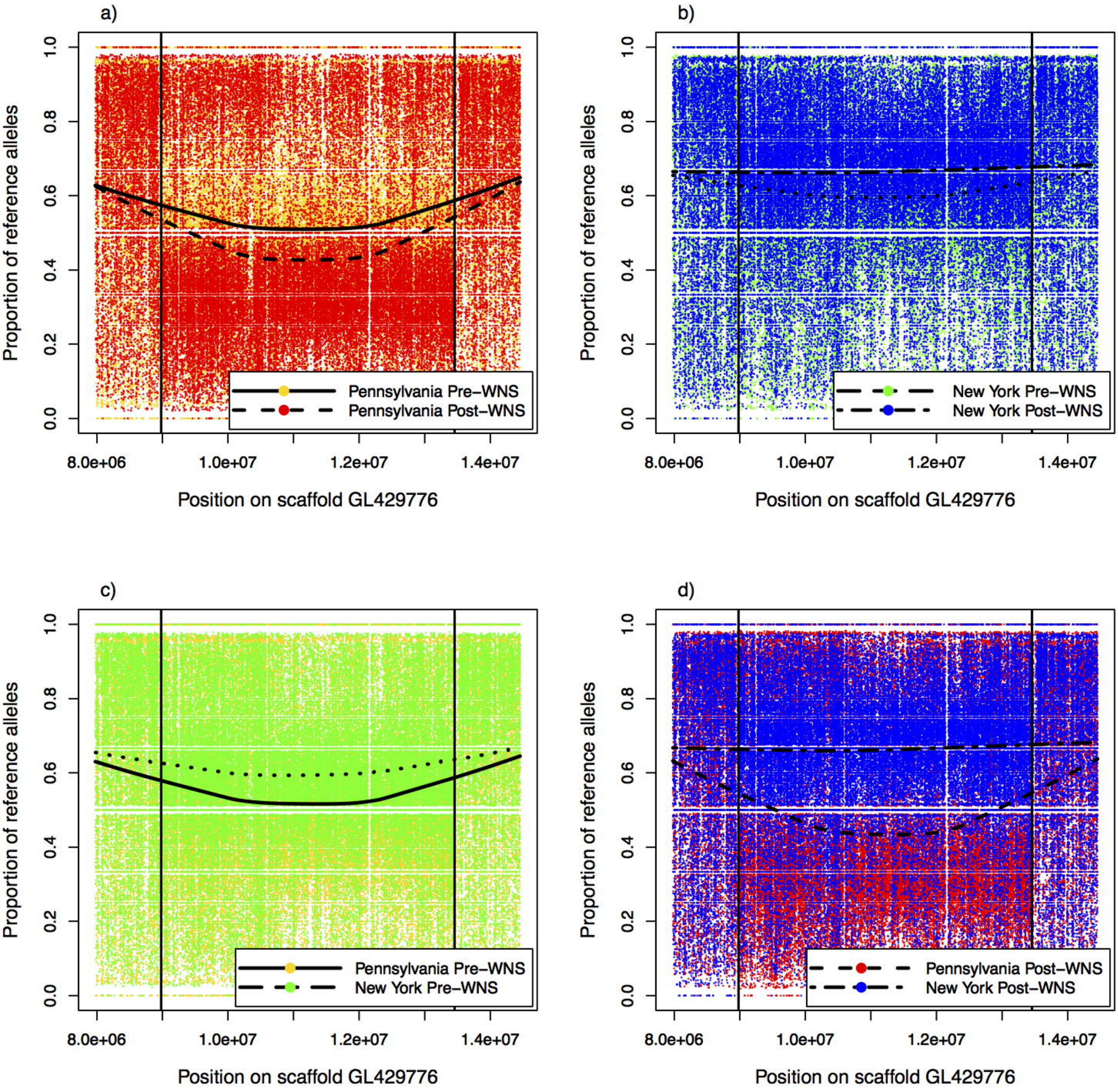
CUNH12orf29	0.0488	1	0.71	10696290	10710421	chromosome unknown C12orf29 homolog
DUSP6	0.0603	1	2.26	11976039	11980455	Inactivates MAP kinases.
GeneID:111824955	0.0459	1	1.57	12101754	12108133	uncharacterised LOC111824955


0
100
200 km
15
13
5
14
12
9
4
3
2
7
1
8
11
10
6


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