Author for correspondence:
Jason Munshi-South
e-mail: jmunshisouth@fordham.edu

THE ROYAL SOCIETY
PUBLISHING

Evolutionary biology

Urbanization shapes the demographic history of a native rodent (the white-footed mouse, *Peromyscus leucopus*) in New York City

Stephen E. Harris[1], Alexander T. Xue[1,2], Diego Alvarado-Serrano[2], Joel T. Boehm[2], Tyler Joseph[2], Michael J. Hickerson[1,2] and Jason Munshi-South[3]

[1]The Graduate Center, City University of New York (CUNY), New York, NY 10016, USA
[2]The City College of New York, City University of New York, New York, NY 10031, USA
[3]Louis Calder Center-Biological Field Station, Fordham University, Armonk, NY 10504, USA

(ID) JM-S, 0000-0002-8067-4341

How urbanization shapes population genomic diversity and evolution of urban wildlife is largely unexplored. We investigated the impact of urbanization on white-footed mice, *Peromyscus leucopus,* in the New York City (NYC) metropolitan area using coalescent-based simulations to infer demographic history from the site-frequency spectrum. We assigned individuals to evolutionary clusters and then inferred recent divergence times, population size changes and migration using genome-wide single nucleotide polymorphisms genotyped in 23 populations sampled along an urban-to-rural gradient. Both prehistoric climatic events and recent urbanization impacted these populations. Our modelling indicates that post-glacial sea-level rise led to isolation of mainland and Long Island populations. These models also indicate that several urban parks represent recently isolated *P. leucopus* populations, and the estimated divergence times for these populations are consistent with the history of urbanization in NYC.

1. Introduction

Urbanization is a particularly potent driver of environmental change around the world [1]. Understanding population genomic responses of organisms to human-driven change provides important context for predicting future evolutionary responses [2]. Using genome-wide single nucleotide polymorphism (SNP) data, we investigate the effects of post-glacial environmental events and urbanization in the New York City (NYC) metropolitan area on historical demography of the white-footed mouse, *Peromyscus leucopus*. We examine the influence of climatic history over thousands of generations and also the effects of recent environmental events tens of generations in the past. This study is the first to examine the impact of urbanization on demographic history using patterns of genomic variation in wild populations.

NYC is particularly well suited for studies on urbanization because the city's recent history of geological [3], ecological [4,5] and cultural [6,7] change has been meticulously recorded. NYC also has clearly defined urban green spaces that are delimited by anthropogenic and natural barriers, and occupied by independently evolving populations of species with poor mobility through the urban matrix [8].

Natural barriers include the Hudson and East Rivers, which separate the mainland portion of the city (i.e. Bronx) from Manhattan and Long Islands. The establishment of Long Island did not begin until the retreat of the late Wisconsin glacier that covered much of present-day NYC [9]. The glacier began retreating

northward approximately 21 000 years before present (ybp) [10], and over the next few thousand years white-footed mice recolonized the region from southern refugia [11]. During this time, *P. leucopus* presumably maintained continuous populations until sea-level rise separated Long Island from mainland NY between 12 000 and 15 000 ybp [10]. Except for occasional land-clearing by Native Americans, anthropogenic barriers were not erected until after European settlement of the area around 1600 CE [4]. During early phases of urbanization in NYC (1609–1790), green spaces within the city were parade grounds, cemeteries, farms or private estates with highly manicured landscapes. In the mid-nineteenth century heavily used land plots, like present-day Prospect and Central Parks, were taken over by city officials and transformed for aesthetic purposes [12]. Private estates were also acquired by the NYC government and redesigned as vegetated parkland [13]. Remnant fauna in these parks, surrounded by a dense urban infrastructure, may have recovered from bottlenecks caused by urban fragmentation as the parks developed mature forests.

Peromyscus leucopus represents one of these remnant species, and we investigated the demographic history of populations occupying contemporary forest fragments in NYC and the surrounding area. *Peromyscus leucopus* are abundant across North America, have a typically short lifetime dispersal capability of approximately 100 m, prefer oak–hickory secondary forests and consume a diet of arthropods, fruits, nuts, vegetation and fungus. White-footed mice are abundant in small, fragmented urban forests [14–16] and exchange migrants only through vegetated corridors between isolated NYC parks [17]. Substantial genetic structure at microsatellite loci exists between NYC parks [8], and there is evidence of divergence and selection in genes underlying functional traits in urban populations [18].

In this study, we estimated the demographic history of *P. leucopus* in NYC to test hypotheses about population expansion and divergence in response to urbanization. We used a genome-wide SNP dataset previously generated [19] from a double-digest restriction site-associated DNA sequencing (ddRADseq) [20] protocol. Loci came from 23 white-footed mouse populations (figure 1) representative of a rural to urban gradient [19]. We used percentage impervious surface cover and human population density around sampling sites as proxies for the extent of urbanization around each site (see table 1 and fig. 1 in [19]). We then used *sNMF* v. 0.5 [21] to examine population structure, and *TreeMix* [22] to build population trees and identify likely genetic clusters of *P. leucopus*. We used data from five populations of white-footed mice in NYC parks that showed evidence of genetic isolation and had relatively high urbanization metrics to test the hypothesis that temporal patterns of population isolation resulted from urbanization (table 1). We estimated demographic parameters from the site–frequency spectrum using the composite-likelihood and coalescent simulation approach implemented in *fastsimcoal2* v. 2.5.1 [23]. *fastsimcoal2* efficiently calculates the approximate likelihood from unlinked SNP loci and accommodates complex demographic models. We used these estimates of effective population sizes, divergence times, migration and population size changes to infer the influence of urbanization on the demography of these populations. Can we distinguish recent, human-driven demographic changes from older natural events under a complex model? See the electronic supplementary material, S1, for full details on the methodology for this study.

2. Results and discussion

(a) Evidence for genetic structure and admixture

Our ddRAD dataset of 14 990 SNPs from 191 individuals sampled at 23 sites (mean of 8 ± 0.17 individuals/site) [19] captured sufficient genetic variation to estimate the post-glacial demographic history of white-footed mouse populations in the NYC metropolitan area. Before inferring demography, a sparse non-negative matrix factorization approach (*sNMF*, Frichot *et al.* [21]) supported assignment of individuals into two main groups separated by the East River and Long Island Sound: (i) mainland and Manhattan (MM) and (ii) Long Island (LI; electronic supplementary material, figure S1). Population trees from *TreeMix* [22] supported the patterns inferred using *sNMF*. *TreeMix* also indicated that several urban parks contain recently fragmented populations (figure 1b) with no evidence of admixture with other sites (electronic supplementary material, S2). When assigning individuals to populations for demographic model development, we compared our results with those of a previous study that examined population structure using genome-wide loci [19]. Genetically differentiated populations included: Central (area: 344.05 ha, 2 km buffer % impervious surface and human population size: 60.2, 351 698.8), Inwood (79.21 ha, 2 km buffer % impervious surface and human population size: 30, 121 354.2) and Van Cortlandt (433.15 ha, 2 km buffer % impervious surface and human population size: 27.7, 77 541.7) parks in MM (790 142 ha); and Jamaica Bay (263.38 ha, 2 km buffer % impervious surface and human population size: 3.2, 1438.4) and Fort Tilden (248.96 ha, 2 km buffer % impervious surface and human population size: 8.5, 2357.5) in LI (362 900 ha). These urban parks are all large, extensively vegetated and surrounded by dense urban development (figure 1a). No rural sampling locations exhibited patterns consistent with genetically isolated populations, suggesting the parks above were isolated due to urbanization.

(b) *Peromyscus leucopus* population history during recent urbanization in NYC

Inferred parameter estimates exhibit a consistent signal of an older split between LI and MM populations in line with geological records followed by recent divergence of NYC park populations. Models had tight confidence intervals around divergence times for MM and LI (approximately 13 600 ybp, electronic supplementary material, figure S2E) except for the two-population model. The two-population model had the lowest likelihood and this result may reflect the relatively poor fit of the model. Divergence was followed by a strong population contraction (table 1, electronic supplementary material, figure S3). These divergence estimates concur with geological records that date the separation of Long Island and the mainland from approximately 13 000–15 000 ybp [24].

Our other demographic models examined whether contemporary urban populations diverged from MM or LI within the historical timeframe of urbanization in NYC. In 1609, shortly after European arrival, only 1% of the Manhattan landscape was urbanized. Over the next 400 years, humans converted 97% of natural green spaces to human use [4]. Urban populations experienced strong population bottlenecks at the time of divergence (except Jamaica Bay) and the inferred time of divergence was always within the 400 year window of European settlement (table 1). While 400 years, representative of approximately 800
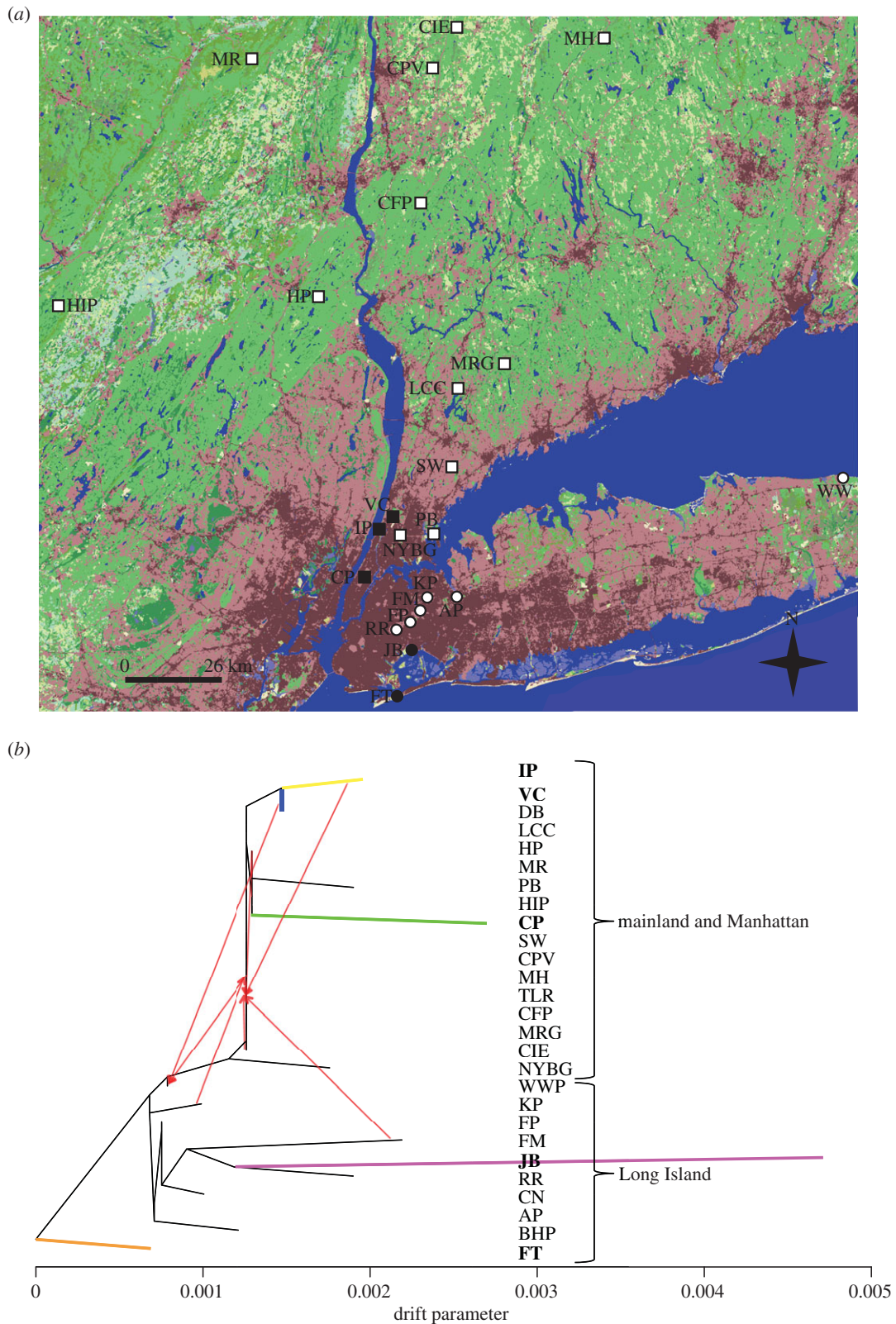
Figure 1. (*a*) Map of all sampling sites in NYC and the surrounding region. Colours correspond to the National Land Cover Database: dark red, Urban High Density Development; light red, Urban Medium to Low Density Development; greens, Forested areas; yellow, Grasslands. Squares, sampling sites from MM; circles, sites from LI; yellow shapes, sites used for urban population demographic analysis. (See electronic supplementary material S1, table S1 for full site names). (*b*) *TreeMix* population tree. Red arrows represent significant admixture using *TreeMix* and *f3* statistics. The drift parameter is plotted along the *x*-axis and represents the amount of genetic drift along the branch. Letters = sampling site codes (see electronic supplementary material, table S1 for full names; AP and CN were combined for all other analyses). Letters in bold and coloured branches correspond to urban sampling sites described in (*a*) and show urban populations with relatively high levels of divergence to non-urban populations, as evidenced by long-branch lengths.

P. leucopus generations assuming a generation time of 0.5 years, is relatively recent, detailed demographic inference over very recent time scales is possible with adequately large genomic datasets [23]. Additionally, many point estimates for urban park divergence are in line with the founding of urban parks in NYC (282 ybp–present, table 1). These results indicate that isolation in urban fragments was sufficiently strong to impact the evolutionary history of urban fauna.

Table 1. Inferred demographic parameters with 95% confidence values from parametric bootstrapping for the three main *fastsimcoal2* model varieties. (see electronic supplementary material, table S2 for remaining models). Ne, effective population size. Time of divergence is in generations. Migration is reported as the coalescent migration rate, *m*, the proportion of individuals that move from one population to another per generation.

parameters	LI_MM site(s) (X)	LI_MM point estimate (95% CI)	LI_MM_CP site(s) (X)	LI_MM_CP point estimate (95% CI)	LI_FT_MM site(s) (X)	LI_FT_MM point estimate (95% CI)
ancestral Ne	—	68 083 (53 492, 111 400)	—	81 275 (36 559, 2 310 630)	—	61 507 (68 608, 284 971)
Long Island Ne	(LI)	1057 (798, 2203)	(LI)	12 075 (7286, 81 320)	(LI)	6515 (8325, 71 939)
mainland and Manhattan Ne	(MM)	54 886 (50 557, 105 507)	(MM)	25 140 (13 268, 106 624)	(MM)	12 584 (12 915, 77 670)
local park Ne	—	—	Central Park	9896 (75, 55 103)	Fort Tilden	6043 (3876, 56 562)
time of divergence (LI_MM)	(LI_MM)	27 198 (6110, 34 227)	(LI_MM)	29 440 (24 111, 29 637)	(LI_MM)	29 354 (12 662, 29 600)
time of divergence (Urban_LI or MM)	—	—	(CP_MM)	746 (512, 13 682)	(FT_LI)	423 (638, 6298)
ancestral resize factor	(LI_MM)	1.22 (0.94, 1.140)	(LI_MM)	3.2 (1.8, 4.6)	(LI_MM)	4.9 (3.2, 7.1)
urban pop. resize factor	—	—	(CP)	9.2×10^{-6} (5.2×10^{-8}, 2.5×10^{-1})	(FT)	8.2×10^{-8} (2.2×10^{-7}, 1.6×10^{-3})
time of urban pop. resize	—	—	(CP)	737 (512, 13 700)	(FT)	317 (638, 6298)
Mig_(X)_to_MM	(LI)	9.1×10^{-4} (4.4×10^{-4}, 1.2×10^{-3})	(LI)	1.8×10^{-6} (5.4×10^{-7}, 2.7×10^{-5})	(LI)	3.9×10^{-6} (7.9×10^{-7}, 8.1×10^{-5})
Mig_(X)_to_(LI)	(MM)	1.5×10^{-5} (1.1×10^{-6}, 3.1×10^{-5})	(MM)	1.0×10^{-6} (7.2×10^{-7}, 7.5×10^{-5})	(MM)	1.9×10^{-6} (5.3×10^{-7}, 9.5×10^{-4})
Mig_(X)_to_LI	—	—	(CP)	2.7×10^{-3} (3.1×10^{-4}, 6.1×10^{-2})	(FT)	2.3×10^{-3} (1.4×10^{-6}, 1.6×10^{-3})
Mig_(X)_to_MM	—	—	(CP)	5.1×10^{-4} (4.0×10^{-6}, 1.5×10^{-1})	(FT)	1.1×10^{-3} (1.1×10^{-4}, 8.9×10^{-4})
Mig_LI_to_(X)	—	—	(CP)	2.9×10^{-3} (9.3×10^{-6}, 6.3×10^{-4})	(FT)	1.0×10^{-2} (8.7×10^{-4}, 6.9×10^{-3})
Mig_MM_to_(X)	—	—	(CP)	5.6×10^{-3} (1.7×10^{-4}, 5.9×10^{-3})	(FT)	3.2×10^{-3} (1.3×10^{-5}, 2.1×10^{-3})

We detected bottlenecks immediately after isolation of urban populations, suggesting that a small remnant population within these parks at the time of the bottleneck provided most of the urban genetic variation found today. Our inferred migration rates between all populations were high and variable, but we estimated consistent patterns of low migration between MM and LI, and asymmetrical migration of individual mice from MM into urban populations (table 1). Despite asymmetrical gene flow, urban parks consistently showed a signal of some emigration to LI or MM, suggesting that urban parks contain stable, though relatively small populations. However, given the extremely recent divergence times, these high migration rates could be due to retained ancestral polymorphisms from incomplete lineage sorting or geographical structure, which are difficult to distinguish from admixture [25]. It is important to note that allelic dropout in ddRADseq data from mutations in cut sites can affect demographic analyses, but using a minimum coverage cut-off and restricting the amount of missing data can mitigate these effects (electronic supplementary material, S1).

3. Conclusion

Our results show that geography, geological events and human-driven habitat change have left a detectable genomic signature in NYC's white-footed mouse populations. Patterns of genetic variation and population structure reflect past demographic processes [26], and genome-wide SNPs generated from ddRADseq provided enough information to distinguish recent demographic events from past geological processes. Our demographic models estimated divergence times and migration patterns that are consistent with the known geological and historical record of NYC. This study is the first to use population genomic modelling to estimate the demographic impact of urbanization on wild populations.

References

1. Corlett RT. 2015 The Anthropocene concept in ecology and conservation. *Trends Ecol. Evol.* **30**, 36–41. (doi:10.1016/j.tree.2014.10.007)

2. Donihue CM, Lambert MR. 2014 Adaptive evolution in urban ecosystems. *Ambio* **44**, 194–203. (doi:10.1007/s13280-014-0547-2)

3. Isachsen YW. 2000 *Geology of New York: a simplified account*. Albany, NY: New York State Museum.

4. Sanderson EW, Brown M. 2007 Mannahatta: an ecological first look at the Manhattan landscape prior to Henry Hudson. *Northeast. Nat.* **14**, 545–570. (doi:10.1656/1092-6194(2007)14[545:MAEFLA]2.0.CO;2)

5. Sanderson EW. 2009 *Mannahatta: a natural history of New York City*. New York, NY: Abrams.

6. Burrows EG, Wallace M. 1998 *Gotham: a history of New York City to 1898*. Oxford, UK: Oxford University Press.

7. Caro RA. 1975 *The power broker: Robert Moses and the fall of New York*. New York, NY: Alfred A. Knopf. Vintage.

8. Munshi-South J, Kharchenko K. 2010 Rapid, pervasive genetic differentiation of urban white-footed mouse (*Peromyscus leucopus*) populations in New York City. *Mol. Ecol.* **19**, 4242–4254. (doi:10.1111/j.1365-294X.2010.04816.x)

9. Lewis RS, Stone JR. 1991 Late Quaternary stratigraphy and depositional history of the Long Island Sound basin: Connecticut and New York. *J. Coast. Res.* **11**, 1–23.

10. Lewis R. 1995 Geologic history of Long Island Sound. *Connecticut College Arboretum Bull.* no. 34. Hartford, CT: Connecticut Geological and Natural History Survey.

11. Smith PW. 1957 An analysis of post-Wisconsin biogeography of the Prairie Peninsula region based on distributional phenomena among terrestrial vertebrate populations. *Ecology* **38**, 205–218. (doi:10.2307/1931679)

12. Rosenzweig R, Blackmar E. 1992 *The park and the people: a history of Central Park*. Ithaca, NY: Cornell University Press.

13. Christen CA. 1988 *Three hundred years of parks: a timeline of New York City Park history*. New York, NY: City of New York, Parks & Recreation.

14. Pergams ORW, Lacy RC. 2007 Rapid morphological and genetic change in Chicago-area *Peromyscus*. *Mol. Ecol.* **17**, 450–463. (doi:10.1111/j.1365-294X.2007.03517.x)

15. Rogic A, Tessier N, Legendre P, Lapointe F-J, Millien V. 2013 Genetic structure of the white-footed mouse in the context of the emergence of Lyme disease in southern Québec. *Ecol. Evol.* **3**, 2075–2088. (doi:10.1002/ece3.620)

16. Munshi-South J, Nagy C. 2014 Urban park characteristics, genetic variation, and historical demography of white-footed mouse (*Peromyscus leucopus*) populations in New York City. *PeerJ* **2**, e310. (doi:10.7717/peerj.310)

17. Munshi-South J. 2012 Urban landscape genetics: canopy cover predicts gene flow between white-footed mouse (*Peromyscus leucopus*) populations in New York City. *Mol. Ecol.* **21**, 1360–1378. (doi:10.1111/j.1365-294X.2012.05476.x)

18. Harris SE, Munshi-South J, Obergfell C, O'Neill R. 2013 Signatures of rapid evolution in urban and rural transcriptomes of white-footed mice (*Peromyscus leucopus*) in the New York metropolitan area. *PLoS ONE* **8**, e74938. (doi:10.1371/journal.pone.0074938)

19. Munshi-South J, Zolnik CP, Harris SE. 2016 Population genomics of the Anthropocene: urbanization is negatively associated with genome-wide variation in white-footed mouse populations. *Evol. Appl.* **9**. (doi:10.1111/eva.12357)

20. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012 Double digest RADseq: an inexpensive

method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**, e37135. (doi:10.1371/journal.pone.0037135)

21. Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. 2014 Fast and efficient estimation of individual ancestry coefficients. *Genetics* **4**, 973–983. (doi:10.1534/genetics.113.160572)

22. Pickrell J, Pritchard J. 2012 Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*. **8**, e1002967. (doi:10.1371/journal.pgen.1002967)

23. Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M. 2013 Robust demographic inference from genomic and SNP data. *PLoS Genet*. **9**, e1003905. (doi:10.1371/journal.pgen.1003905)

24. Lewis RS, Stone JR. 1992 Late Quaternary stratigraphy and depositional history of the Long Island Sound Basin: Connecticut and New York. *J. Coast. Res. Spec. Issue* **11**, 1–23.

25. Lohse K, Frantz LAF. 2014 Neandertal admixture in Eurasia confirmed by maximum-likelihood analysis of three genomes. *Genetics* **196**, 1241–1251. (doi:10.1534/genetics.114.162396)

26. Li J, Li H, Jakobsson M, Li S, Sjödin P, Lascoux M. 2012 Joint analysis of demography and selection in population genetics: where do we stand and where could we go? *Mol. Ecol*. **28**, 28–44. (doi:10.1111/j.1365-294X.2011.05308.x)