Urbanization shapes the demographic history of a city-dwelling native rodent

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], 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, Fordham University, Armonk, NY 10504 USA

**Corresponding author:* Jason Munshi-South

E-mail: jmunshisouth@fordham.edu

ABSTRACT

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 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 SNPs genotyped in 23 populations sampled along an urban-to-rural gradient. Both prehistorical climatic events and recent urbanization impacted these populations. Our modeling 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 New York City.

Keywords: ddRAD-seq, *Peromyscus leucopus*, composite likelihood, site-frequency-spectrum

INTRODUCTION

Urbanization is a particularly potent driver of environmental change around the world [1]. By understanding population genomic responses of organisms to human-driven change, we can establish baselines for examining future evolutionary responses [2]. Using genome-wide 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 look at climatic history over thousands of generations and add to this traditional approach by examining the effect 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 that 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 NYC and much of Long Island [9]. The glacier began to retreat northward ~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—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-19th century heavily used land plots, like 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.

P. 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. 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 (Fig 1) representative of urban, suburban, and rural habitats [19]. We then used *sNMF* version 0.5 [21] to examine population structure and *TreeMix* [22] to build population trees and identify likely genetic clusters of *P. leucopus*. We examined demographic models for isolated populations of white-footed mice in NYC parks to test the hypothesis that temporal patterns of population isolation were due to urbanization (Fig. 2, see supplement for full

methods). We estimated demographic parameters from the site-frequency-spectrum (SFS) using the composite-likelihood and coalescent simulation approach implemented in *fastsimcoal2* [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, and demographic bottlenecks to answer questions about the impact of urbanization and demographic inference from model based simulations. Can we distinguish recent, human-driven demographic changes from older natural events under a complex model? What natural and anthropogenic barriers drive population divergence in NYC populations? Do population sizes change in response to habitat fragmentation during urbanization? How do human-built barriers to dispersal affect gene flow between urban rodent populations?

RESULTS AND DISCUSSION

Evidence for genetic structure and admixture

Our ddRAD dataset of 14,990 SNPs from 191 individuals sampled at 23 sites [19] captured enough genetic variability to effectively 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.* 2014) supported assignment of individuals into two main groups separated by the East River and Long Island Sound (see supplemental material): 1) Mainland & Manhattan (MM) and 2) Long Island (LI) (Fig. S1). Population trees from *TreeMix* [22] supported the patterns inferred using *sNMF*. *TreeMix* also indicated that several urban parks contain recently-fragmented populations (Fig 1) with no evidence of admixture with other sites (Table S2-3, and supplement). When assigning individuals to populations for demographic model development, we compared our results to those

of a previous study that examined population structure using genome-wide loci [19]. Genetically differentiated populations included Central, Inwood, and Van Cortlandt Parks in MM, and Jamaica Bay and Fort Tilden in LI. These urban parks are all large, extensively vegetated, and surrounded by dense urban development. No rural sampling locations exhibited patterns consistent with genetically isolated populations, suggesting the parks above were isolated due to urbanization.

P. leucopus population history during recent urbanization in NYC

Inferred parameter estimates show a consistent signal of an older split between LI and MM populations in line with geologic records followed by recent divergence of NYC park populations. All models had relatively narrow confidence intervals around divergence times for MM and LI (~13,600 ybp, Fig S2) followed by strong population contraction (Fig. 2, Table 1). These divergence estimates concur with geologic records that date the separation of Long Island and the Mainland from ~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 ~800 *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

124 urban park divergence are in line with the founding of urban parks in NYC (282 ybp – present,
125 Table 1). These results indicate that isolation in urban fragments was sufficiently strong to impact
126 the evolutionary history of urban fauna.

127 We detected bottlenecks immediately after isolation of urban populations, suggesting that
128 a small remnant population within these parks at the time of the bottleneck provided most of the
129 urban genetic variation found today. Our inferred migration rates between all populations were
130 high and variable, but we estimated consistent patterns of low migration between MM and LI,
131 and asymmetrical migration of individual mice from MM into urban populations (Table 1).
132 However, given the extremely recent divergence times, these high migration rates could be due to
133 retained ancestral polymorphisms from incomplete lineage sorting or geographic structure that
134 are difficult to distinguish from admixture [25].

135

136 **CONCLUSIONS**

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

145

146

METHODS

Sampling and DNA extraction

Two previous studies [8, 19] sampled individual white-footed mice between 2009 and 2013 from 23 separate localities that were used to generate the genomic data used in this study. Sites were chosen to represent a rural to urban gradient (Fig. 1). Rural sites were defined as large tracts of relatively undisturbed natural habitat, and urban sites were fragmented habitat surrounded by urban infrastructure and impervious surface. For all sampling locations, we trapped individuals over a period of 1-3 nights each. At each site, we set between one and four 7x7 m transects with 49 3" x 3" x 9" Sherman live traps, depending on the total area of each sampling site. We weighed, sexed, and took morphological measurements for all individual mice. At all sites except Central Park, Flushing Meadow, New York Botanical Garden, Brook Haven Park & Wild Wood Park, High Point Park, and Clarence Fahnestock Park, we collected tissue by taking 1 cm tail clips, placing in 80% ethanol, and storing at -20C in the laboratory. For the previously mentioned six sites, we sampled tissue from previously-collected liver samples stored in RNAlater (Ambion Inc., Austin, TX) at -80C. We extracted genomic DNA using standard extraction protocols, quantified the yield, and checked quality before genomic sequencing library preparation. See methods in (Munshi-South *et al.* 2015) for full details. All animal handling procedures were approved by the Institutional Animal Care and Use Committee at Brooklyn College, CUNY (Protocol Nos. 247 and 266) and by Fordham University's Institutional Animal Care and Use Committee (Protocol No. JMS-13-03).

RAD sequencing and SNP calling

We filtered out relatives and low-quality samples [2], retaining 191 *P. leucopus* individuals from 23 sampling sites for the genome-wide SNP dataset. Briefly, we followed standard protocols for ddRADseq presented in Peterson *et al.* (2012) [20], starting with DNA extraction using Qiagen DNEasy kits with an RNAse treatment. Next we used a combination of the enzymes, SphI-HF and MluCI to generate similarly sized DNA fragments. Using AMPure XP magnetic beads we cleaned the digested DNA then ligated barcodes and Illumina sequencing adapters to the fragments. We used a Pippin Prep for precise DNA fragment size excision from gels and then Phusion High-fidelity PCR reagents to add Illumina sequencing primers. The resulting fragments were sent to the NYU Center for Genomics and Systems Biology who sequenced the samples using three lanes of Illumina HiSeq 2000 2x100 bp paired-end sequencing. We checked initial quality of the raw reads using FastQC and subsequent primer removal, low-quality nucleotide trimming, and *de novo* SNP calling was conducted using the Stacks 1.21 pipeline. We called and filtered SNPs in Stacks using default setting except for requiring that loci occur in 22 / 23 sampling sites, and within each site, occur in at least 50% of individuals. We chose a random SNP from each RAD tag to avoid linkage between loci. Additionally, we removed individuals if they had too few reads resulting in extremely small SNP datasets or if they showed high levels of relatedness to other white-footed mice sampled. These filters resulted in 14,990 SNPs in the final dataset we used for demographic modeling.

Population structure and migration

We investigated observed patterns of genetic diversity in order to define evolutionary clusters that could be used to inform demographic modeling of *P. leucopus* populations in the NYC region. We examined population structure and evidence of migration among all 23

sampling sites. The program *TreeMix* [22] was used to build population trees and find likely migration events. *TreeMix* infers populations splitting and mixing using allele frequencies from large genomic datasets. Using a composite likelihood approach given allele frequency data, TreeMix returns the most likely population tree and admixture events given a user-specified number of admixture events. The number of admixture events tested ranged from 0 - 12 while the rest of the parameters used default settings. P-values were generated for each admixture event and comparisons made between all trees. We confirmed admixture between populations by running *f3* three-population analyses in *Treemix*. These statistics look for admixture between tested populations by identifying correlations between allele frequencies that do not fit the evolutionary history for that group of three populations. We used 500 bootstrap replicates for significance and counted migration events that were significant in *Treemix* and contained admixed populations as identified with *f3* statistics.

We also used sNMF version 0.5 to examine population structure sNMF explores patterns of genetic structure by assigning individual ancestry coefficients using sparse non-negative matrix factorization. sNMF does not make any model assumptions like requiring populations to be in Hardy-Weinberg and linkage equilibrium, as opposed to other likelihood models like STRUCTURE. For the number of putative ancestral populations tested, we chose a range from K = 1 to K = 11 using default parameters, with 10 replicate runs for each value of K. We ran sNMF on the full 14,990 SNP dataset (≤ 50% of SNPs missing per population) and on a more conservative dataset with only ≤ 15% of SNPs missing per population. sNMF imputes missing genotypes by resampling from the empirical frequency at each locus, and using fewer missing data ensured any population structure patterns found were not due to incorrectly imputed

212 genotypes (Fig. S1). To infer the most likely number of ancestral populations, each model run
213 generates a cross-entropy estimation based on ancestry assignment error when using masked
214 genotypes. The model with the smallest cross-entropy score implies it is the best prediction of the
215 true number of K ancestral populations.

216 **Demographic inference from genome-wide site frequency spectra**

217 In order to reduce model complexity for demographic inference, we attempted to group
218 individuals into the minimum number of populations representing unique evolutionary clusters.
219 Global analyses in TreeMix and sNMF showed the highest support for two populations split by
220 the East River, and hierarchical analyses using discrimant analysis of principal components
221 showed support for isolated urban populations. Collectively, results suggested a minimum of
222 seven putative populations captured most of the genetic variation between populations (Mainland
223 & Manhattan: MM, Long Island: LI, Central Park: CP, Van Cortlandt Park: VC, Inwood Hill
224 Park: IP, Jamaica Bay: JB, Fort Tilden: FT, Fig. 1). Along with hierarchical population structure
225 results, we chose several of the urban populations to include in demographic modeling based on
226 the size of the park, the relative isolation of the park due to urbanization, and the population
227 density of white-footed mice in the park. We generated the multi-population site frequency
228 spectrum (MSFS) for subsets of populations to test specific demographic history scenarios. We
229 used custom scripts (see supplemental materials) and the *dadi.Spectrum.from_data_dict*
230 command implemented in *dadi* to generate the MSFS. When we created the SNP dataset, we
231 required a SNP to occur in ≥ 50% of individuals from each population, so the MSFS was down-
232 projected to 50% to ensure the same number of individuals for all loci. Once the MSFSs were
233 generated, we used the software program *fastsimcoal2* [23] for demographic inference.

234 *Fastsimcoal2* (fsc2) uses a composite multinomial likelihood approach to infer demographic
235 histories from the site frequency spectrum generated from genomic scale SNP datasets. The
236 expected SFS under user defined demographic scenarios is obtained using coalescent simulations.

237 We tested demographic histories under a scenario of population isolation with migration
238 (IM model). This included six hierarchical IM models and we compared inferred parameters
239 between models (Fig. 2). All models were tested using the same dataset. There was one two-
240 population IM model (seven free parameters) to test older divergence patterns between MM and
241 LI suggested from the geologic record. The remaining five models were three-population IM
242 models (15 free parameters each) testing for recent urban population divergence. We chose to
243 run separate models investigating each urban population separately in order to avoid plausible
244 inconsistencies from over-parameterization. For these remaining models we considered an
245 ancestral population that split at time T_{div1} and then an urban population that split more recently at
246 time T_{div2}. For T_{div1} we included a range of divergence times based on the LGM of the Wisconsin
247 glacier, ~18,000 ybp. For T_{div2} we considered divergence times incorporating the timeframe of
248 urbanization in NYC, ~400 ybp. We allowed for migration between all populations, and tested
249 occurrences of population bottlenecks when urban isolation was incorporated into the model (See
250 appendix 3.3 for demographic models and parameter ranges). During likelihood calculation, a
251 conditional maximization algorithm (ECM) is used to maximize the likelihood of each parameter
252 while keeping the others stabilized. This ECM procedure runs through 40 cycles where each
253 composite-likelihood was calculated using 100,000 coalescent simulations. While increasing the
254 number of simulations can increase precision, accuracy does not significantly increase past
255 100,000 simulations [9]. Additionally, in order to avoid likelihood estimates that oversample

parameter values at local maxima across the composite likelihood surface, we ran 50 replicates with each starting from different initial conditions. We chose the replicate with the highest estimated maximum likelihood score for each model. Using parametric bootstrapping, we generated confidence intervals for the most likely inferred demographic parameters generated. The SFS was simulated with the parameter values from the highest likelihood model and then new parameter values re-estimated from the simulated SFS. We ran 100 parametric bootstraps. In order to find consistent signals of divergence which could be attributed to urbanization, we compared parameter values and overlapping confidence intervals between models.

DEMOGRAPHIC INFERENCE

Parameters were allowed to vary in demographic modeling using fastsimcoal2, but all six models converged on similar parameter values estimated from the observed MSFS. Parameter estimates with the highest likelihood generally fell within the upper and lower bounds generated from parametric bootstrapping (Fig. S2, Table 1). The first two-population model tested divergence time, effective population size, and migration rates between MM and LI populations (Model 1, Fig. 2). The divergence time for the MM and LI split was inferred to be 13,599 ybp and the MM effective population size (N_E) was 50x larger than the LI N_E (Table 1). Divergence times are based on a generation time of 0.5 years for *Peromyscus leucopus*. Migration was also inferred to be low (< 1 individual per generation) between MM and LI (Table 1).

The inferred demography for the more complex three-population models generally supported results from the two-population model. The first two complex models both estimated the divergence between MM and LI, but one model tested for divergence of JB and LI after the MM and LI split (Model 5, Fig. 2) while the other model tested divergence between FT and LI

after the MM and LI split (Model 6, Fig. 2). This model also tested the likelihood of a bottleneck event when FT and JB, both urban populations, diverged. We set up the other three complex models in an identical fashion, except we tested the urban populations of CP (Model 2, Fig. 2), VC (Model 3, Fig. 2), or IP (Model 4, Fig. 2) for divergence from MM after the MM and LI split. Point estimates for demographic parameters converged on similar values and generally fell within the 95% confidence limits from parametric bootstrapping (Fig. S2, Table 1). The average divergence time for MM and LI was 14,679 ybp SD = 956.19. Similar to the two-population model, the MM N_E was larger (at least 2x in each model) than the LI N_E. The individual urban populations all had N_E values 10x smaller than MM, but often similar to LI. The divergence time for the five tested urban populations, even with variation in number of generations per year, was consistent with the timeframe of urbanization (mean divergence = 233 ybp; SD = 164.5). Our demographic models proved to be rather robust in returning reasonable parameter values with consistent convergence to similar values across replicates. Although wide confidence intervals on many parameters suggest low resolution in inferring parameters values given the model and data, they are likely a consequence of the complexity of the model given the number of parameters and wide parameter ranges. The narrow confidence intervals on other parameters suggest that these inferences reliably capture important aspects of the true demographic history of white-footed mice in NYC, especially given the often biologically unrealistic parameter search space (Fig S3).

ACKNOWLEDGMENTS

This research was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R15GM099055 to JM-S, a NSF Graduate Research Fellowship to SEH, by NASA through the Dimensions of Biodiversity Program and National Science Foundation: DOB 1343578 and DEB-1253710 to M. J. H. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health.

REFERENCES

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

2. Donihue, C. M. & Lambert, M. R. 2014 Adaptive evolution in urban ecosystems. *Ambio* , 1–10. (doi:10.1007/s13280-014-0547-2)

3. Isachsen, Y. W. 2000 *Geology of New York: A Simplified Account*. New York State Museum/Geological Survey, State Education Department, University of the State of New York.

4. Sanderson, E. W. & 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, E. W. 2009 *Mannahatta: a natural history of New York City*. New York: Abrams.

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

7. Caro, R. A. 1975 *The power broker: Robert Moses and the fall of New York*. 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, R. S. & Stone, J. R. 1991 Late Quaternary stratigraphy and depositional history of the Long Island Sound basin: Connecticut and New York. *J. Coast. Res.* , 1–23.

10. Lewis, R. 1995 *Geologic History of Long Island Sound*.

11. Smith, P. W. 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:http://dx.doi.org/10.2307/1931679)

12. Rosenzweig, R. & Blackmar, E. 1992 *The Park and the People: A History of Central Park*. Cornell University Press.

13. Christen, C. A. 1988 *Three Hundred Years of Parks: A Timeline of New York City Park History*. City of New York, Parks & Recreation.

14. Pergams, O. R. W. & Lacy, R. C. 2007 Rapid morphological and genetic change in Chicago-area Peromyscus. *Mol. Ecol.* **17**, 450–63. (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**, 2075–88. (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, S. E., 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, C. & Harris, S. E. 2015 Population genomics of the Anthropocene: urbanization reduces the evolutionary potential of small mammal populations. *bioRxiv* (doi:http://dx.doi.org/10.1101/025007)

20. Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. 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-Sánchez, E., Sousa, V. C. & Foll, M. 2013 Robust Demographic Inference from Genomic and SNP Data. *PLoS Genet.* **9**, e1003905. (doi:10.1371/journal.pgen.1003905)

24. Lewis, R. S. & Stone, J. R. 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, L. a F. 2014 Neandertal admixture in Eurasia confirmed by maximum-likelihood analysis of three genomes. *Genetics* **196**, 1241–51. (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)

FIGURES AND TABLES

Table 1. Inferred demographic parameters with 95% confidence values from parametric bootstrapping for all fastsimcoal2 models.

	LI_MM Model 1		LI_MM_CP Model 2		LI_MM_VC Model 3		LI_MM_IP Model 4		LI_JB_MM Model 5		LI_FT_MM Model 6	
Parameters	Site(s) (X)	(Point Estimate) (95 % CI)	Site(s) (X)	(Point Estimate) (95 % CI)	Site(s) (X)	(Point Estimate) (95 % CI)	Site(s) (X)	(Point Estimate) (95 % CI)	Site(s) (X)	(Point Estimate) (95 % CI)	Site(s) (X)	(Point Estimate) (95 % CI)
Ancestral Ne	-	68083 (53492,111400)	-	81275 (36559,2310630)	-	90482 (71280,312469)	-	73627 (53734,283269)	-	114700 (85828,269490)	-	61507 (68608,284971)
Long island Ne	(LI)	1057 (798,2203)	(LI)	12075 (7286,81320)	(LI)	8991 (8180,74287)	(LI)	8723 (6639,70959)	(LI)	7354 (9849,74872)	(LI)	6515 (8325,71939)
Mainland & Manhattan Ne	(MM)	54886 (50557,105507)	(MM)	25140 (13268,106624)	(MM)	16416 (13984,93388)	(MM)	15138 (12371,80804)	(MM)	17186 (18832,76281)	(MM)	12584 (12915,77670)
Local park Ne	-	-	Central park	9896 (75,55103)	Van,Cortlandt	156 (27,36482)	Inwood Hill	6886 (32,41092)	Jamaica Bay	6279 (37,54167)	Fort Tilden	6043 (3876,56562)
Time of divergence (LI_MM)	(LI_MM)	27198 (6110,34227)	(LI_MM)	29440 (24111,29637)	(LI_MM)	29669 (17400,29663)	(LI_MM)	28662 (23600,29647)	(LI_MM)	29666 (18312,29636)	(LI_MM)	29354 (12662,29600)
Time of divergence (Urban_LI or MM)	-	-	(CP_MM)	746 (512,13682)	(VC_MM)	373 (459,8586)	(IP_MM)	462 (545,14379)	(JB_LI)	327 (667,7749)	(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)	5.5 (2.9,7.0)	(LI_MM)	4.8 (2.5,7.2)	(LI_MM)	6.7 (3.2,7.2)	(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})	(VC)	0.6 (3.8×10^{-9},0.66)	(IP)	4.8×10^{-5} (2.2×10^{-8},7.8×10^{-1})	(JB)	1.86 (1.8×10^{-8},0.83)	(FT)	8.2×10^{-8} (2.2×10^{-7},1.6×10^{-3})
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)	2.4×10^{-6} (7.0×10^{-7},1.2×10^{-5})	(LI)	2.4×10^{-6} (6.3×10^{-7},6.5×10^{-6})	(LI)	2.9×10^{-6} (7.6×10^{7},1.4×10^{-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)	1.4×10^{-6} (7.9×10^{-7},6.5×10^{-5})	(MM)	2.1×10^{-6} (7.9×10^{-7},3.7×10^{-5})	(MM)	1.3×10^{-6} (7.8×10^{-7},2.9×10^{-4})	(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})	(VC)	0.16 (6.5×10^{-4},0.39)	(IP)	5.1×10^{-3} (4.4×10^{-4},0.22)	(JB)	3.8×10^{-3} (2.5×10^{-5},1.9×10^{-1})	(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})	(VC)	0.07 (3.5×10^{-5},2.3×10^{-1})	(IP)	1.4×10^{-5} (5.6×10^{-6},1.9×10^{-1})	(JB)	1.2×10^{-3} (1.9×10^{-4},5.9×10^{-2})	(FT)	1.1×10^{-3} (1.1×10^{-4},8.9×10^{-4})
Mig_LI_to_(X)	-	-	(CP)	2.9×10^{-4} (9.3×10^{-6},6.3×10^{-4})	(VC)	1.4×10^{-3} (1.5×10^{-4},3.1×10^{-3})	(IP)	5.4×10^{-4} (8.6×10^{-5},2.7×10^{-3})	(JB)	5.3×10^{-3} (5.9×10^{-4},4.8×10^{-3})	(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})	(VC)	5.1×10^{-3} (7.5×10^{-5},3.5×10^{-3})	(IP)	5.9×10^{-3} (5.0×10^{-5},4.5×10^{-3})	(JB)	4.4×10^{-3} (1.7×10^{-5},2.4×10^{-3})	(FT)	3.2×10^{-3} (1.3×10^{-5},2.1×10^{-3})

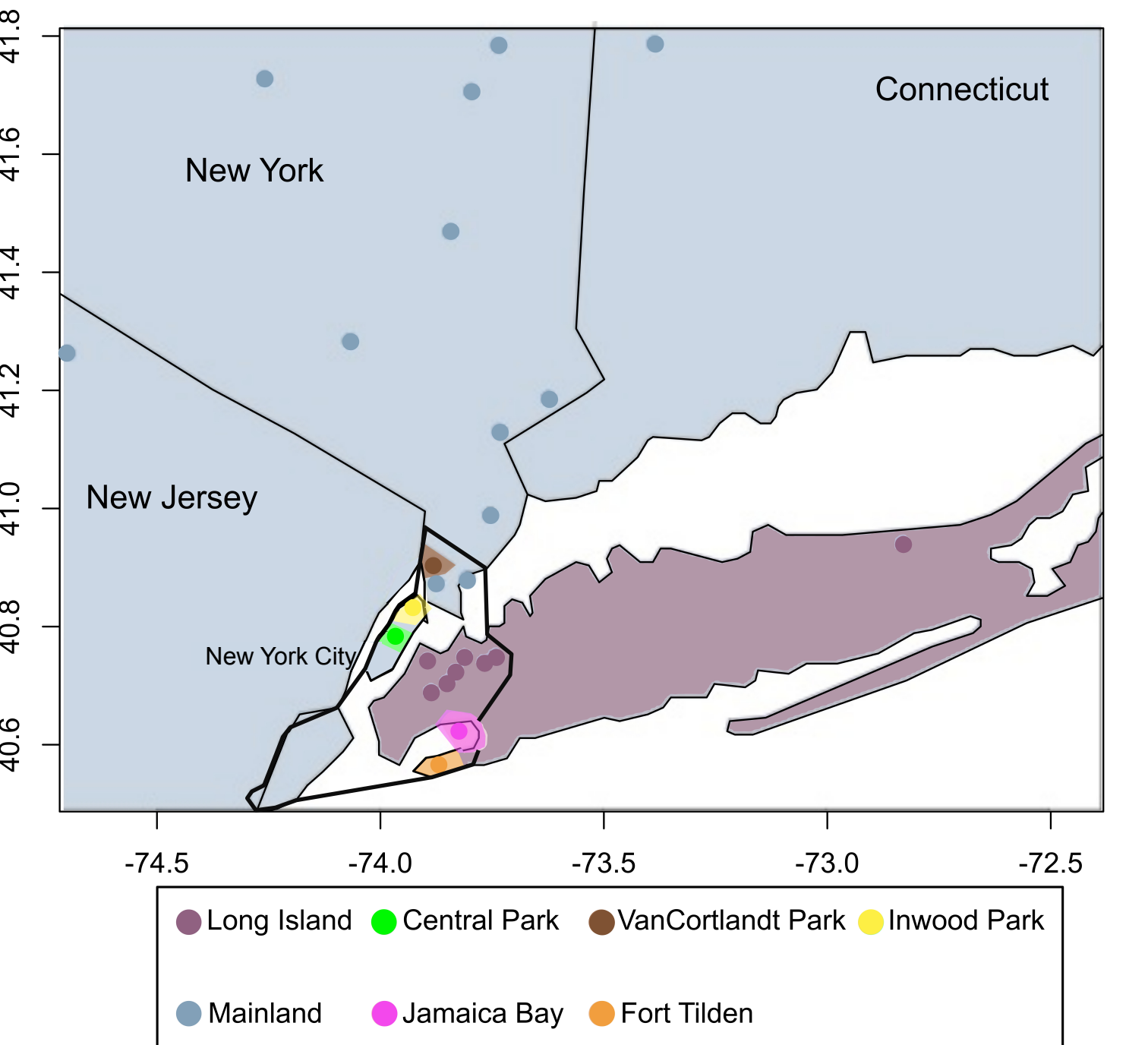
Ne = effective population size. Time of divergence is in generations. Migration is reported as the coalescent *m*, proportion of individuals that move from one population to another per generation.

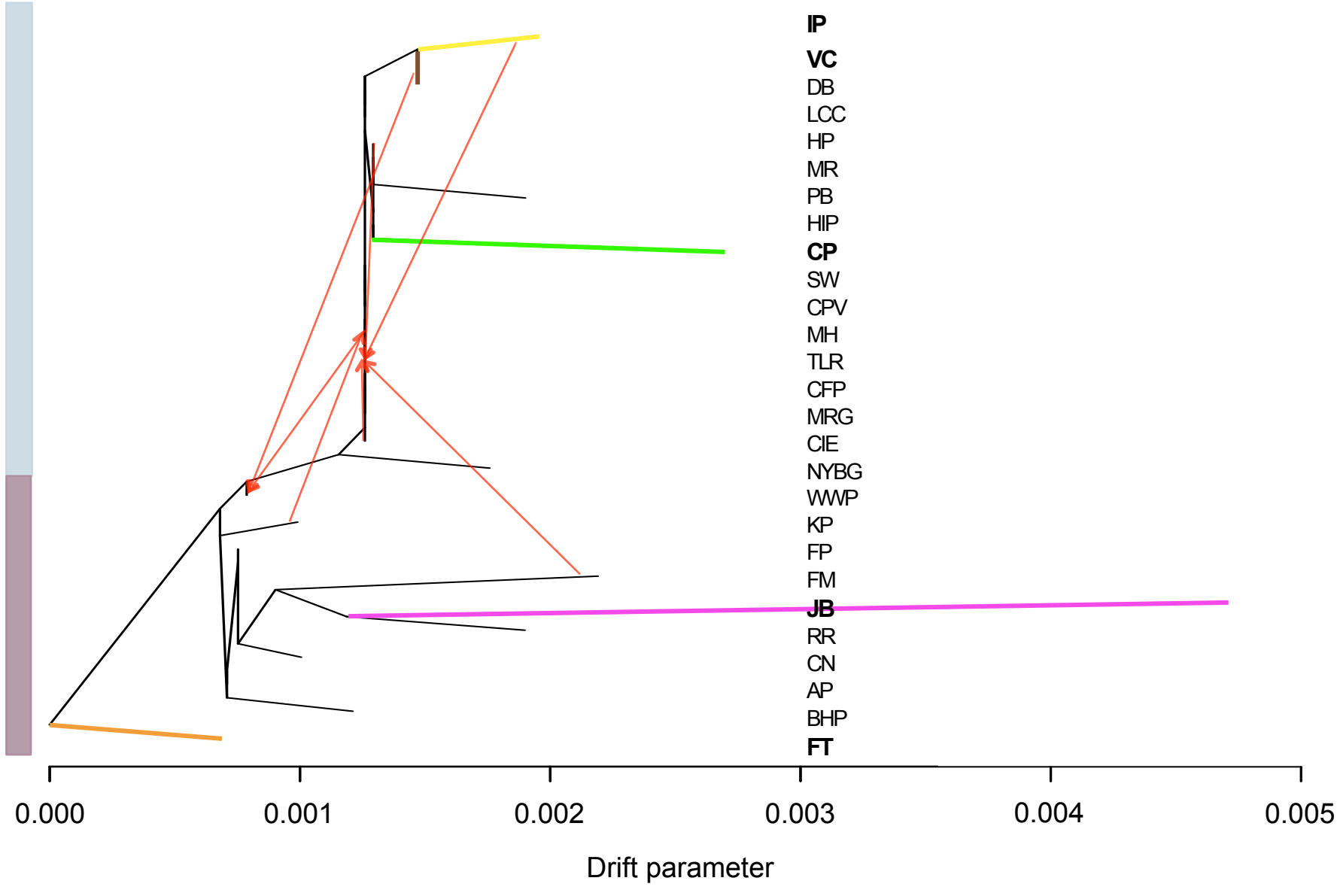
Figure 1

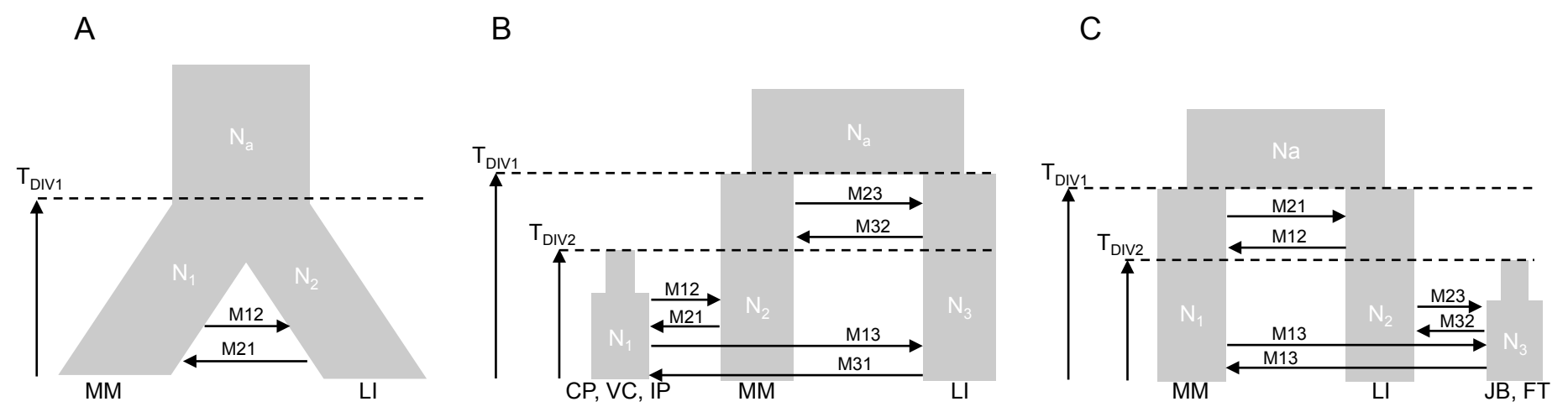
(A) Map of NYC region with Long Island (LI), Mainland (MM), Inwood Hill Park (IP), Central Park (CP), Van Cortlandt Park (VC), Fort Tilden (FT), and Jamaica Bay (JB). Points = sampling location, shaded to correspond to population assignment. Vertical axis = Latitude. Horizontal axis = Longitude. (B) *TreeMix* population tree. Red arrows represent significant admixture using *TreeMix* and *f3* statistics. Shading and bold lines correlate with colors from Fig. 1A.

Figure 2

fastsimcoal2 models. MM = Mainland & Manhattan. LI = Long Island. CP = Central Park. VC = Van Cortlandt Park. IP = Inwood Hill Park. JB = Jamaica Bay. FT = Fort Tilden. TDIV = Time of Divergence. N = Effective Population Size. M = migration rate. (A) Model 1 tests divergence between MM and LI. (B) Models 2, 3, and 4 test divergence of CP, VC, and IP from MM after the split. (C) Models 5 and 6 test for divergence of JB and FT from LI after split.


Connecticut
New York
New Jersey
New York City
41.8
41.6
41.4
41.2
41.0
40.8
40.6
-74.5
-74.0
-73.5
-73.0
-72.5
Long Island
Central Park
VanCortlandt Park
Inwood Park
Mainland
Jamaica Bay
Fort Tilden


IP
VC
DB
LCC
HP
MR
PB
HIP
CP
SW
CPV
MH
TLR
CFP
MRG
CIE
NYBG
WWP
KP
FP
FM
JB
RR
CN
AP
BHP
FT
0.000
0.001
0.002
0.003
0.004
0.005
Drift parameter


A
N_a
T_{DIV1}
N_1
N_2
M12
M21
MM
LI
B
N_a
T_{DIV1}
M23
M32
T_{DIV2}
M12
M21
N_1
N_2
N_3
M13
M31
CP, VC, IP
MM
LI
C
Na
T_{DIV1}
M21
M12
T_{DIV2}
M23
M32
N_1
N_2
N_3
M13
M13
MM
LI
JB, FT