Admixture on the northern front: population genomics of the white-footed mouse (*Peromyscus leucopus*) in Quebec and evidence of hybridization with the deer mouse (*Peromyscus maniculatus*)

Alan Garcia-Elfring

Department of Biology, Faculty of Science

McGill University

Montreal, Quebec

April 2017

A thesis submitted to McGill University

in partial fulfillment of the requirements of the degree of Master of Science

Abstract

Processes like allele surfing and genetic bottlenecks during range expansion have genetic consequences expected to result in differentiated wave-front populations with low genetic variation and potentially introgression by a local species. The northward expansion of *Peromyscus leucopus* in southern Quebec provides an opportunity to test these predictions using population genomic tools. Our results show evidence of recent and post-glacial expansion. Genome-wide variation in *P. leucopus* indicates two putative glacial lineages are separated by the St. Lawrence River, with a more recent post-glacial divergence of populations isolated by the Richelieu River. Most regressions of genetic diversity against relative expansion distance were not significant and half contradicted expectation by showing positive trends. However, in two of three transects we documented northern populations with the lowest diversity in at least one genetic measure. Consistent with bottlenecks and allele surfing during northward expansion, we document a northern-most population having low nucleotide diversity, divergent allele frequencies, the most private alleles, as well as the lowest effective population size and heterozygosity levels that indicate outcrossing. Ancestry proportions revealed putative hybrids of *P. leucopus* and *P. maniculatus*. A formal test for gene flow confirmed secondary contact, showing that a reticulate population phylogeny between *P. maniculatus* and *P. leucopus* was a better fit to the data than a bifurcating model without gene flow. Thus, we provide the first genomic evidence of gene flow between this pair of species in natural populations. Understanding the evolutionary consequences of secondary contact is an important conservation concern as climate-induced range expansions are expected to result in new hybrid zones between closely related species, with potential consequences for emerging infectious diseases.

Résumé

L’expansion géographique des espèces a des conséquences sur la structure génétique des populations au front d’expansion, qui peut être altérée par de processus tels que le surf d’alleles et les goulots d'étranglement résultant en une variabilité génétique réduite au sein des populations périphériques. L’expansion de la souris à pattes blanches (*Peromyscus leucopus*) vers le Nord dans le sud du Québec offre l'occasion de tester ces prédictions à l'aide d'outils génomiques. Nos résultats révèlent la signature d’une 'expansion récente et post-glaciaire. La structure génétique des populations de *P. leucopus* indique la présence au Québec de deux lignées séparées par le fleuve Saint-Laurent, ainsi que d’une divergence secondaire des populations de part et d’autre de la rivière Richelieu. Par ailleurs, nos données ne permettent pas de détecter une réduction de diversité génétique le long du gradient d’expansion mais au contraire une tendance positive pour certaines des lignées, ce qui va à l’encontre de la théorie. Cependant, nous documentons l’évidence de goulots d’étranglement et de surf d’allèles dans les populations les plus au nord qui possèdent une diversité nucléidique réduite, des fréquences d’allèles divergentes et d’allèles privés élevées, ainsi qu’une taille de population effective et des niveaux d’hétérozygotie qui suggèrent l’existence de croisements. Nous avons ainsi détecté l’existence d’individus hybrides de *P. leucopus*, avec une autre espèce présente au Québec, *P. maniculatus*, suite à un contact secondaire entre les deux espèces. Nos travaux documentent ainsi pour la première fois l’existence d’hybridation entre les deux espèces de *Peromyscus* en milieu naturel. Une meilleure compréhension des conséquences évolutives du contact secondaire entre espèces apparentées est un important problème de conservation, car l’une des conséquences de l’expansion des espèces suite au réchauffement climatique est de créer de nouvelles hybrides.

CONTENTS

Acknowledgements

I would first and foremost like to thank my supervisor, Dr. Virginie Millien, for providing me the opportunity to work on this research and for all her support and advice throughout the project, as well as Dr. T. Jonathan Davies. I also thank committee member Dr. Rowan D. H. Barrett for his guidance and insightful feed-back. I thank my collaborators from the Munshi-South lab, in particular Dr. Jason Munshi-South and his PhD student, Mathew Combs, for their assistance in this project and for the welcomed visit to their lab in Armonk NY. Additionally, I'm grateful for the hospitality that Dr. Munshi-South and his wife provided during my short stay in their home in New York City. I thank Dr. Brian Boyle from IBIS for the time he took answering questions during the library preparation process and assisting with the transfer of libraries to Genome Quebec. I'm likewise grateful for the time taken by Dr. Antoine Paccard to help me get acquainted with the computer cluster Guillimin, and Dr. Frederique Chain for support with Bash scripting. I thank former lab mate Dr. Sarah T. Leo for her assistance and feedback with PCR work. I thank Visiting Professor Yanping Wang for the stimulating office discussions and helpful feedback, a sentiment I extend to lab mates Julia Nordlund, Kirsten Crandall and Katherine Hebert. I also thank Professor Andrew P. Hendry for helpful feedback on the first draft of this thesis. This thesis was built on work done by past lab members and associates who took part in field work and collected samples - hard work that was essential to this research – and so I thank them. Lastly, I thank friends and family, in particular my mother, Rosalia Elfring, for the unwavering support.

Contribution of Authors

The study design of this thesis (Chapter 2) was developed with the guidance of Virginie Millien and with feed-back from committee members T. Jonathan Davies and Rowan D. H. Barrett. Rowan D. H. Barrett provided computational resources. Jason Munshi-South and Mathew Combs helped identify SNPs and analyze data. Virginie Millien funded sequencing efforts. I (A.G.) wrote the manuscript (Chapter 2 and 3) with intellectual input and revisions from all co-authors. Final thesis was approved by members of the Supervisory Committee, which include Drs.Virginie Millien, T. Jonathan Davies, and Rowan D. H. Barrett. The results and interpretations presented in this thesis have been submitted to the journal *Heredity* for review.

This work was funded with the support of an NSERC grant held by Virginie Millien and T. Jonathan Davies and through McGill's Department of Biology Research Travel Award.

CHAPTER I
LITERATURE REVIEW

Genetic consequences of range expansion

Changes in the distribution of species are thought to be an important process driving evolutionary change and diversification (Rosenzweig 1995; Price et al. 2014). The study of range expansion, in particular invasive species, was founded as a sub-discipline of ecology (i.e. invasion ecology) by Charles Elton's *The ecology of invasions by animals and plants* (1958). As an ecologist, Elton focused on the ecological factors that facilitate expansion and on species interactions, overlooking the evolutionary changes that might occur during colonization (Barrett 2015). *The genetics of colonizing species* (1965), edited by Baker and Stebbins, was the first comprehensive evolutionary-based publication discussing how the genetic composition of a colonizer may affect range expansion or change during the process (Barrett 2015).The study of range expansion has been dominated by invasive species, in large part due to their high economic toll (Pimentel et al. 2005). However, range expansions also occur, for example, following conservation efforts (Lubina and Devin 1988; Hagen et al. 2015), when pathogens pass from host to host (Biek et al. 2007), or as a result of climate warming over varying time scales, ranging from millennia (Hewit 1996, 2000, 2004) due to Milankovitch cycles, to decades with more rencent human-mediated climate warming (Parmesan and Yohe 2003; Chen et al. 2011). This chapter reviews current concepts and empirical evidence on the genetic consequences of range expansion and introduces the study system.

Loss of genetic variation

The process of expansion has both stochastic and deterministic elements that can be viewed as a spatial analog to genetic drift (Slatkin and Excoffier 2012) and natural selection (Shine et al., 2011). Range expansion is gennerally thought of as a series of successive founder events extending along an axis of expansion. A subsample of the total genetic variation in a given population is taken to a new range limit in each founder event, leading to a cline of reduced genetic variation from the historical range to the recently established range margin (Excoffier et al. 2009; Slatkin and Excoffier 2012). This prediction has been substantiated by multiple empirical studies (Estoup et al. 2004; Heckel et al. 2005; Handley et al. 2007; Besold et al. 2008;Watts et al. 2010; Garroway et al. 2011; White et al. 2013; Cahill et al. 2016). However, this model of expansion, based on invasive species, assumes the rate of expansion across space is limited by traits intrinsic to the colonizing population (e.g. reproduction rate, dispersal distance, etc.) rather than by extrinsic biotic and abiotic factors. Fewer genomic studies have investigated the genomic consequences of climate-driven range expansion (but see Pluess 2011; Swaegers et al. 2015; Monzón et al. 2016), although recent theoretical work shows that colonizing populations can maintain genetic diversity when expansion is both driven and limited by a gradual shift in optimal habitat (e.g. Nullmeier and Hallatschek 2013; Dai et al. 2014; Garnier and Lewis 2016). Colonizing populations that can avoid becoming genetically impoverished may have a higher evolutionary potential to adapt to the novel environment and become established.

Differentiation and changes in population structure

During range expansion, individuals colonizing a new area will likely carry mutations that are present at low frequency in the source population but can increase in frequency upon subsequent generations on the wave front. Because this occurs at random loci, this neutral phenomenon can be though of as a spatial analogue of random genetic drift (Slatkin and Excoffier 2012) and has been termed ‘allele surfing’ (Klopfstein et al. 2006). Although alleles commonly found segregating throughout a range margin population are more likely to ‘surf’, the surfing of rare alleles has a greater effect on population structure and differentiation (Edmonds et al. 2004; Hallatschek et al. 2007; Excoffier & Ray 2008). Allele surfing is therefore expected to result in spatial clines of allele frequencies at random loci, including loci with deleterious effects (Travis et al. 2007; Henn et al. 2016). In small founder populations, low effective population size reduces the efficacy of natural selection and the influence of stochastic processes increases, thereby partially masking deleterious alleles and facilitating their fixation on the wave-front (Peischl et al. 2013, 2015). This can lead to an enrichment of deleterious mutations with increasing expansion distance, a theoretical expectation documented empirically in humans (Henn et al. 2015; Peischl et al. 2016).

Introgressive hybridization

Another known process associated with range expansions is hybridization and introgression (reviewed in Brennan et al. 2015; Canestrelli et al 2016). The potential of interspecific gene flow as a force of evolutionary change as long been known to botanists (e.g. Anderson 1948, 1953; Heiser 1949) who also noticed an association between hybridization and abrupt anthropogenic disturbances (Wiegand 1935; Riley 1938; Anderson 1948). Edgar Anderson was among the first evolutionary biologists post-modern synthesis to write about interspecific gene-flow in nature and, with Ledyard Stebbins, gave a now classic disquisition of its role in adaptive evolution (Anderson and Stebbins 1954). A zoological counterpart was provided fifty years ago by

Lewontin and Birch (1966), in which they proposed that hybridization between species of *Drosophila* facilitated range expansion. However, their results were later attributed to phenotypic plasticity (Birch and Vogt 1969; reviewed in Arnold 2004).

By the mid-20th century hybridization was well accepted and integrated into a conceptual evolutionary framework within botany (e.g. allopolyploidy), although the subject took a backseat in zoological studies. A likely factor for the disregard among zoologists, with the exception of a few researchers (e.g. Arnold 1992; Dowling and Secor 1997), for the idea that introgression may be an important, if not relatively common source of genetic variation for adaptive evolution likely stems from Ernst Mayr's profound and lasting influence on evolutionary thinking. Mayr, an ornithologist, defined the concept of species by the property of reproductive isolation and held the opinion that animal hybrids were rare in nature and genetic dead-ends, thus evolutionarily unimportant (Mayr 1942, 1963). Furthermore, Mayr postulated that gene exchange between lineages under divergent selection could only erode built-up variation underlying local adaptation. It is the case that, relative to plants, animal hybrids were much rarer in Mayr's time, although modern methods now suggest this may have been an issue of detection rather than an evolutionary pattern.

The advent of next-generation sequencing and the development of genomic methods has brought a flood of data along with a resurgence of interest in hybridization and the evolutionary role of gene flow during primary divergence and secondary contact (Seehausen 2004; Mallet 2005; Mallet et al. 2007; Gompert et al. 2012; Gompert et al. 2014; Lowe et al. 2015; Abbott et al. 2016; Gompert & Buerkle 2016). It is now widely recognized that during speciation and secondary contact, species boundaries may be blurred and species gene-pools porous (Feder et al. 2012; reviewed in Seehausen et al. 2014).

Expansion has historically been viewed as a driver of range expansion rather than a consequence of expansion dynamics. However, recent simulation studies show that introgression, in addition to facilitating expansion into a new environment through adaptive introgression (reviewed in Lee 2002; Pfennig et al. 2016), can result from the colonization process itself. Currat et al. (2008) and Excoffier et al. (2009) showed that differential demography characterizing the colonizer and local species (low population density, low Ne; at carrying capacity, respectively) led to massive introgression from the local species into the colonizer. It is known that effective population size (Currat et al., 2008; Amorim et al., 2017) and intra-specific gene flow (Petit and Excoffier, 2009) is inversely proportional to the probability of introgression. However, despite this prediction, recent genomic studies on ongoing range expansions (e.g. White et al. 2013; Swaegers et al. 2015; Monzón et al. 2016) have tended to focus on the colonizing species and neglected how the gene-pool of the colonizer may interact with the local species (but see Mastrantonio et al. 2016). Advances in theory (e.g. Currat et al., 2008; Seehausen et al. 2014; Lindtke and Buerkle, 2015; Gompert & Buerkle, 2016) and technology (Davey et al. 2011) have revolutionized genomic studies and now provide a means to answer long standing questions regarding the frequency and consequences of hybridization between non-model species in natural populations.

Hybridization in mammals is typically regarded as rare relative to other animals, like birds and butterflies (reviewed in Mallet 2005). The paucity of mammalian hybrids has been hypothesized to be due to faster evolution of gene regulation relative to evolution of protein coding genes (Wilson et al. 1974; Prager and Wilson 1975) and a hyper-mutable Y chromosome (Short 1997) within mammals. However, recent genomic methods show that loci associated with mouse hybrid incompatibility can be found throughout the genome (e.g. Janousek et al 2012). The apparent rarity of mammalian hybridization is, at least in part, due to many related small mammal species being nocturnal and of similar phenotypes, making the visual identification of hybrids more difficult. Nonetheless, one study documented that 6% of European mammals are known to hybridize (Mallet 2005), including the house mouse (Staubach et al. 2012). There is also evidence of mammalian species hybridizing during colonization or changes in distribution, including canids during in North America (Reich et al. 1999; von Holdt et al. 2016), polar bears (Kelly et al. 2010; Cahill et al. 2015), and indeed humans, which are thought to have hybridized with at least three hominin species during expansion out of Africa (Huerta-Sánchez et al. 2014; Fu et al. 2015; Mondal et al. 2016).

Interspecific gene-flow can provide unique adaptive benefits through the introgression of pre-adapted alleles, including in animals (Hedrick 2013). One example is the Western Mediterranean mouse (*Mus spretus*), which adapted to rodenticide through the interspecific introgression of resistance genes from the house mouse, *M. musculus* (Song et al. 2011). Additionally, introgression of Denisovan DNA in the ancestors of Tibetan people is thought to have resulted in altitude adaptation (Huerta-Sanchez et al. 2014). Remarkably, a parallel seems to have occurred in the ancestors of modern Tibetan Mastifs after hybridizing with gray wolves from the Tibetan Plateau (Miao et al. 2016). In this regard, interspecific gene-flow on the wave-front may provide the pre-adapted genetic substrate needed to uncouple gene-pools from a local adaptive maximum and reach a higher global peak upon which a population may establish (Wright 1932).

Selection and range expansion

Range expansion can expose populations in newly colonized areas to novel selective pressures and shift evolutionary trade-offs optimized for core populations at equilibrium. For example, simulations of range expansion show strong selection for the evolution of higher dispersal propensity, although this can be inhibited by allee effects (Travis and Dytham 2002). Moreover, dispersal is expected to evolve to a lower density threshold such that dispersal is favoured even when a newly colonized area is still markedly below carrying capacity (Travis et al. 2009). The evolution of increased dispersal on the wave-front can be largely attributed to interbreeding between the best dispersing individuals in expanding demes, a process known as spatial sorting (Shine et al. 2011). In other words, within a generation of expansion, phenotypic effects arising from genetic variation in traits with quantitative effects on dispersal will assort individuals across space such that those with the collection of traits which confer the highest propensity for dispersal will aggregate on the wave-front ('Olympic village' effect). Consequently, by merit of dispersal abilities, mating amongst the best dispersing individuals found on the wave-font will increase the mean dispersal distance in the following generation, thereby accelerating the rate of expansion (Shine et al. 2011).

The invasive cane toad in Australia not only serves as a classic example of the potential pitfalls of biological control, but also of contemporary evolution from selection on the wave-front. Cane toad individuals on the wave have increased their dispersal distance by 5-fold in the70 years since their introduction (Phillips et al. 2006). The increased dispersal was attributed to spatial sorting due to the evolution of longer legs and jumping distance in wave-front toads. Furthermore, parasites can lag behind toads on the wave-front by 1-3 years (e.g. Phillips et al. 2010), thus colonizing populations may experience fewer immunological challenges that alter evolutionary trade-offs (e.g. Burton et al., 2010). To the degree that variation in traits involved in competition, reproduction, immunity and dispersal is genetically determined, selection is expected to produce allele frequencies clines along the axis of expansion at loci involved in such traits (e.g. White et al. 2013). However, clines from allele surfing can mimic selection and lead to false positives. Because allele surfing occurs at random loci, identifying outliers across independent transects or routes of expansion helps mitigate false positives and strengthens the case that a given outlier locus is under selection or in linkage disequilibrium with a selected locus.

Introduction to system: The colonizer and the local species

White-footed mouse (*Peromyscus leucopus*) populations in southern Quebec provide an excellent opportunity to investigate the genomic consequences of spatial expansion in a small mammal. *P. leucopus* is native to eastern North America (King 1968), and one of many species undergoing range expansion due to anthropogenic climate change (Myers et al. 2009; Chen et al. 2011; Roy-Dufresne et al. 2013). The expansion of *P. leucopus* has been estimated at 15 km per year across the Upper Michigan Peninsula (Myers et al. 2009) where it is ecologically replacing *P. maniculatus* (sensu *P. maniculatus gracilis*, Wan, 2104) and 10 km per year northward into southern Quebec (Roy-Dufresne et al. 2013). *P. leucopus* co-exists across most of its distribution with *P. maniculatus* and both are key components of small mammal communities in woodland across Quebec. Although sympatric across most of their range, *P. leucopus* is typically found in deciduous forests (Klein et al 1960; reviewed in Vessey and Vessey 2007), while *P. maniculatus* is more associated with conifer forests and reaches higher latitudes with a greater tolerance to colder winter temperatures (Pierce and Vogt 1993).

The two species discussed here, *P. leucopus* and *P. maniculatus*, belong to sister clades which, according to the fossil record, diverged approximately 500,000 years ago (Hibbard 1968). Both species have a more closely related sister species with which they can hybridize and produce fertile offspring under, *P. polionotus* and *P. gossypinus,* respectively (Dice 1937; 1940; McCarley 1954; Maddock & Dawson 1974). Mating experiments suggest some pre-zygotic isolation between *P. gossypynus* and *P. leucopus* depending on whether individuals come from sympatric or allopatric natural populations (McCarley 1964; Bradshaw 1968), a pattern indicative of reinforcement. After unsuccessful attempts to experimentally hybridize these species (e.g. Dice 1933) and with no documented observations of hybrids in nature, Dice (1968) concluded that complete reproductive isolation between the *P. maniculatus* and *P. leucopus* had likely occurred. However, visual observation of hybrids in nature is hampered by their similar phenotype and inconspicuous coats. Decades later, through hormonally-induced ovulation and artificial insemination, Dawson et al. (1972) and Maddock and Dawson (1974) produced live hybrids, although they were not deemed viable; essentially short lived genetic dead-ends. Subsequently, a tacit consensus that *P. maniculatus* and *P. leucopus* were reproductively isolated became established. Challenging this notion, however, Leo and Millien (2016) recently used microsatellite markers and found putative hybrids of both species in natural populations.

There has been an implicit assumption in evolutionary biology that the alleles of loci associated with intrinsic isolation or species boundaries are differentially fixed between species across their entire geographic range. As a consequence, hybrids between a given species-pair have been treated as a monolithic group in which all hybrids more or less experience the same fitness cost (i.e. degree of outbreeding depression) regardless of parental genotype. However, it is a fact that every fixed difference, including those affecting hybrid fitness, arose as a mutation that underwent a polymorphic stage before becoming fixed in a population.

Assumptions of species-wide fixed incompatibilities have likely been made for model simplicity (Cutter 2012) and because of the expectation that variants affecting hybrid fitness are likely under divergent selection and likely to be quickly fixed over evolutionary time scales. However, loci under divergent selection that affect hybrid fitness can remain polymorphic under scenarios of polygenic and or balancing selection (reviewed in Cutter 2012). Furthermore, cryptic genetic variation, that is loci affecting hybrid fitness but neutral within a species, may take longer to fix if it is not lost to random drift. There is accumulating empirical evidence for the view that reproductive isolation in recently diverged species can vary between and within populations, being conditional on individual genotype (e.g. Kozlowska et al. 2012; Mandeville et al. 2015; Araripe et al., 2016) with variable reproductive isolation now documented in a range of diverse taxa including mammals, arthropods, and nematodes (Cutter 2012).

Responses to climate change such as range expansion are affecting co-evolved life cycle and transmission dynamics between symbionts and their hosts, with far reaching effects on human, crop, and wildlife diseases (reviewed by Altizer *et al.*, 2013). Both *P. leucopus* and *P. maniculatus* are competent reservoir hosts for a number of infectious diseases. The deer mouse has been associated with severe Hantavirus outbreaks (MacNeil et al. 2011), while *P. leucopus* has a notorious association with *Borrelia burgdorferi* (LoGiudice et al. 2003), the bacterial agent of Lyme disease, an emerging infectious disease in Quebec (Ogden et al. 2008). In this context, climate-induced range expansion of *P. leucopus* into Quebec is likely to put more human populations at risk for Lyme disease.

Research objectives

For this thesis, we investigate the population genomics of the *P. leucopus* range expansion in southern Quebec using genomic methods based on RAD sequencing (reviewed in Andrews et al. 2016). To infer the post-glacial demographic history of *P. leucopus* in Quebec, we (i) analyzed population genomic structure. To look for the signature of recent expansion, we explore predictions regarding differentiation and loss of variation on the wave-front (i.e. consequences of allele surfing and genetic bottlenecks) and look for northward clines of (ii) decreasing genetic variation and (iii) increasing differentiation. Furthermore, to assess predictions of introgression during expansion and previous conclusions regarding reproductive isolation between *P. leucopus* and *P. maniculatus* (Dice 1933; Dice 1968; Dawson et al. 1972; Maddock & Dawson 1974) we look for the sign of recent admixture between these species. We use this species-pair as a case study to test the hypothesis that sympatric species, typically reproductively isolated under equilibrium conditions, are more likely to hybridize during colonization. To this end, we (ii) analyze population genomic structure and (iii) test for gene flow.

CHAPTER II

Admixture on the northern front: population genomics of the white-footed mouse (*Peromyscus leucopus*) in Quebec and evidence of hybridization with the deer mouse (*Peromyscus maniculatus*)

A. Garcia-Elfring[1, 2], R. D. H. Barrett[1], M. Combs[3], T. J. Davies[2], J. Munshi-South[3], V. Millien[1]

[1] Redpath Museum, McGill University, 859 Sherbrooke Street West, Montreal, QC, Canada, H3A OC4

[2] Department of Biology, McGill University, 1205, Docteur Penfield, Montreal, QC, Canada, H3A 1B1

[3] Louis Calder Center, Biological Field Station, Fordham University, Armonk, NY 10504, USA

*The results and interpretations presented in this thesis have been submitted to the journal *Heredity* for review.

MATERIALS AND METHODS

Study system and DNA extraction

We used *P. leucopus* samples from 13 forested sites (n=229) in southern Quebec during the summer of 2013 and 2014. To explore the possibility of interspecific admixture, we also genotyped *P. maniculatus* from one forested site (‘Pm’, n=11). Sites were grouped into three transects, the northern (N), south-eastern (SE) and south-western transects (SW), separated by the St. Lawrence and Richelieu Rivers (Figure 1). Studies of *P. leucopus* using microsatellites and mitochondrial markers have indicated that the St. Lawrence and Richelieu Rivers inhibit gene-flow, with differentiated populations on either side (Rogic *et al.*, 2013; Ledevin and Millien 2013; Fiset *et al.*, 2015), while agricultural fields are relatively ineffective barriers (Marrotte *et al.*, 2014). We thus suggest that our three transects represent distinct expansion routes that may provide independent genetic patterns of range expansion, taking the northern sites in each transect (e.g. N4, SW5/SW4, SE4) as the approximate range margin. A search of capture records on Vertnet reveals that a putative *P. leucopus* individual has been caught further north of our northern-most sample site. However, this individual was a juvenile and visually identified, making it especially difficult to differentiate from *P. maniculatus*. In fact, according to the measurements taken by the collector (e.g. ear, tail), this individual more closely resembles *P. maniculatus* (Lindquist *et al.*, 2003). We extracted DNA from liver and muscle tissue using a 3-day phenol-chloroform protocol (Sambrook *et al.*, 1989). Species were identified via PCR amplification of a mitochondrial COIII sequence, as described in Rogic *et al.*, (2013).

RAD-seq library

Extracted DNA was sent to the Institut de biologie intégrative et des systèmes (IBIS) at Université Laval for individual-based library preparation using a modified version of the original genotyping-by-sequencing (GBS) protocol (Elshire *et al.,* 2011). DNA from each individual was digested with the rare cutting (1 cut /4096 bp) restriction enzyme PstI (CTGCAG) to seed markers, followed by MspI (CCGG), a frequent cutter (1 cut /256 bp) used to define read size. Next, a barcoded adapter with a complimentary end to the PstI site and a common adapter with a complimentary end to the MspI site were ligated to the DNA fragments. Following ligation, sets of 48 samples were pooled and PCR was applied to the barcoded DNA fragments using primers that hybridize to the adapters. Each library composed of 48 samples had 100 bp paired-end reads sequenced on a single Illumina HiSeq 2000 lane at the Genome Quebec Innovation Center.

SNP genotyping and genetic diversity measures

We used the Stacks pipeline (Catchen *et al.,* 2013a) to process raw data and identify single nucleotide polymorphisms (SNPs). Raw reads were de-multiplexed and quality filtered using the *process_radtags* script of Stacks 1.37. We followed the recommendation of recent studies (Nam *et al.,* 2016; Shafer *et al.,* 2016) and aligned to a reference genome prior to SNP calling. We aligned all processed reads to the scaffold-level assemblage of the *P. maniculatus bairdii* reference genome (NCBI assembly accession: GCF_000500345.1) with Bowtie2 V 2.2.8 (Langmead and Salzberg, 2012) using local alignment. We removed one *P. maniculatus* and six *P. leucopus* individuals that had relatively few reads sequenced. This left 223 *P. leucopus* and 10 *P. maniculatus* samples (Table 1). Scripts were executed using the *ref_map.pl* wrapper in Stacks (v. 1.44) to generate a catalogue of RAD loci with a minimum depth of coverage of 4 reads (-m

4). The *populations* script was then used to obtain two SNP datasets for population genomic analysis, one consisting only of the 13 *P. leucopus* populations (n=223, '*P. leucopus*-only' dataset hereafter), and a second dataset that included the 10 *P. maniculatus* individuals (n=233, 'complete' dataset hereafter). Genome-wide markers were filtered so that each SNP had a minimum allele frequency (MAF) of at least 2.5% (--min_maf 0.025) and a heterozygosity of 0.50 or less to avoid paralogs (--max_obs_het 0.50). In addition, each locus had to be found in every population, that is 13 populations for the *P. leucopus*-only dataset (-p 13) and 14 populations in the complete dataset (-p 14), as well as in at least sixty percent of individuals in a given population (-r 0.60).

The following population genetic summary statistics were then calculated for each dataset: nucleotide diversity (π), homozygosity (Hom_{obs}, Hom_{exp}) and heterozygosity (Het_{obs}, Het_{exp}), the inbreeding coefficient (F_{IS}), private alleles, and differentiation as given by F_{ST}. Nucleotide diversity is related to expected heterozygosity and is an overall measure of genetic variation (Catchen *et al.,* 2013b). F_{IS} measures the degree of inbreeding due to non-random mating, or the change in observed homozygosity relative to the expected value. F_{ST} represents the amount of inbreeding due to random mating in a finite population and is used as a measure of population subdivision and genetic drift (Crow and Kimura, 1970; Summarized in Ewen, 2012). Thus, F_{IS} measures changes in genotype frequencies while F_{ST} measures changes in allele frequencies. We applied an F_{ST} correction, using Fisher's exact test to asses whether allele frequencies at a given locus are statistically different from zero. Loci with F_{ST} estimates that fail to reach statistical significance according to the p-value (α=0.05) are set to zero (--fst_correction p_value). To estimate contemporary effective population size, ancestry proportions, and construct a population tree, the complete dataset was thinned to the first SNP of each RAD locus to reduce linkage disequilibrium (--write_single_snp). The ordination methods used (DAPC and PCA) do not make assumptions of independence and thus all SNPs identified in each RAD locus were used in those analyses.

To examine patterns of genetic diversity, we followed White *et al.*, (2013) and regressed measures of heterozygosity and nucleotide diversity onto ‘relative expansion distance’. For sites on the northern transect which are aligned in a north-eastern direction (Figure 1), we use distance to the southern-most site (DSS). The St. Lawrence River partitions the northern transect along a north-eastern axis and thus we hypothesize that northward range expansion in this transect is likely occurring in a north-eastern direction. For this transect we therefore used the Euclidian distances (km) of sites N2, N3 and N4 to site N1 (Table 1). South of the St. Lawrence River, the Richelieu River is aligned in a north-south direction with no major barrier to direct northward expansion. For the south-western and south-eastern transects, we therefore used the distance to the southern-most latitude (DSL), or the Euclidian distance to the latitude of the southern-most site (i.e. SW1 and SE1). We regressed genetic diversity measures for each transect separately.

To evaluate the extent of drift, we used the two-dimensional isolation by distance model of Rousset (1997), regressing pair-wise F_{ST} estimates (F_{ST}/1- F_{ST}) of 10 populations and their southern-most site as a function of the natural log of relative expansion distance. The inverse of the slope of the regression provides an estimate of gene flow via the parameter of ‘neighbourhood size’ or $4\pi D\sigma^2$, where D is the effective density of individuals and $\sigma^2$is the mean squared parent-offspring distance (Slatkin and Barton, 1989; Rousset, 1997). Lastly, contemporary effective population size (Ne) was estimated using a linkage disequilibrium method (Waples and Do, 2008) implemented in the software NeEstimator V2 (Do et al., 2014).

This method was performed on file that was not imputed. However, NeEstimator corrects for missing data by calculating a fixed-inverse variance-weighted harmonic mean (Peel et al., 2013).

Imputation of missing genotypes

Next generation sequencing, especially the RAD-seq protocol, produces a patchy genotype matrix with a considerable amount of missing information. Imputation of missing data has been shown to aid allele frequency estimates and improve the power of genomic studies (Li *et al.*, 2009). For this reason, we used the software LinkImpute (Money *et al.,* 2015) to estimate missing genotypes in PLINK files (Purcell *et al.,* 2007) for all data analyses (except estimation of Ne). LinkImpute uses a *k*-nearest neighbour genotype imputation method and was designed for RAD-seq data from non-model organisms.

Genomic structure and admixture

To summarize overall genetic variation among individuals, we performed a principal component analysis (PCA) on both datasets (with and without *P. maniculatus* samples) using the package *LEA* (Frichot & Francois, 2015) in R. To look for major lineages within *P. leucopus*, we further analyzed *P. leucopus* population structure by transforming genomic variation to PCs and applying a discriminant analysis (Jombart *et al.*, 2010) in the R package *adegenet* (Jombart, 2008). DAPC describes genomic clusters using synthetic variables and focuses on the genetic variation observed between groups, while minimizing within-group variance. Genomic clusters (K) were identified using a *k*-means algorithm and the number of clusters evaluated with the Bayesian information criterion (BIC; Figure S1a). DAPC was performed on the two K values with the lowest BIC (K=2 and K=3). We used alpha-score optimization to examine the trade-offs between over-fitting and power to discriminate. This procedure identifies the PCs (Supplementary Figure S1) with the highest alpha-scores using spline interpolation. The PC with the highest (‘optimal’) alpha-score provides an approximation and with other integers of similar alpha-scores a range from which it is adequate to choose the number of PCs to retain (Jombart and Collins, 2015). We performed this procedure on 110, 10, and six PCs (Supplementary Figure S2), which showed that the first six PCs have very similar high alpha scores. We thus kept the first six PCs (12.1% of variance) for DAPC presented here. Individual ancestry proportions of *P. leucopus* and *P. maniculatus* were estimated using a model-based method implemented in ADMIXTURE using default settings (Alexander *et al.*, 2009; Alexander & Lange, 2011). ADMIXTURE uses cross-validation to evaluate models with K ancestral source populations and maximum-likelihood algorithms to estimate ancestry proportions. We define a hybrid as any individual with visible hetero-specific ancestry.

Population splits and test of bifurcating model

We used the software TreeMix 1.13 (Pickrell and Pritchard, 2012) to build a maximum likelihood (ML) tree and analyze the demographic histories of our *P. leucopus* populations. To construct a bifurcating tree, this method uses a Gaussian approximation of genetic drift and the covariance in allele frequencies between population pairs. We used PLINK 1.9 (Purcell and Chang, 2015) to stratify allele frequencies and then the plink2treemix.py script (Pickrell and Pritchard, 2012) to convert the stratified file (.frq) to TreeMix format. The ML tree was rooted using *P. maniculatus*. The amount of genetic drift estimated to have occurred in a lineage is proportional to the horizontal branch length on a given branch.

Unlike the clustering methods described above, TreeMix explicitly tests for the presence of gene flow by identifying population pairs that poorly fit a bifurcating evolutionary history, and then models gene flow events between these populations to increase the fit. Migration events are added in order of their statistical significance, such that the first gene flow event added to the tree is the one that most increases the likelihood of the model. The topology of the tree is estimated after each migration event. The direction of gene flow shown is based on asymmetries in the covariance matrix of allele frequencies, as shown on Figure 1 of Pickrell and Pritchard (2012). We modelled the tree with zero to three gene flow events. Below we present the results of the first two ML trees, a tree based on a strict bifurcating model and a tree with one gene flow event. Residual plots indicate the fit of the model.

RESULTS

Data

Sequencing our 240 samples generated 1,117,579,057 raw paired-end reads. After removing seven individuals with few reads sequenced and filtering for population genomic analyses, we obtained 38,144 bi-allelic SNPs for the *P. leucopus*-only dataset and 33,919 SNPs for the complete dataset. Thinning the complete dataset to the first SNP per RAD locus yielded 12,507 SNPs for analysis of ancestry proportions and construction of a ML tree.

P. leucopus population genetic summary statistics

Summary of population genetic statistics are presented on Table 1. The mean observed heterozygosity (Het_{obs}) in *P. leucopus* is 0.145 and mean inbreeding coefficient (F_{IS}) is 0.074 (SD = 0.056). We expected a northern decrease in genetic diversity in each transect. Counter to expectations, we found that most regressions were non-significant and half showed contradicted our hypothesis, showing a positive trend (Figure 2). However, northern populations in the northern and in south-eastern transects show relatively low genetic polymorphism in at least one measure, including a significant linear decrease in nucleotide diversity north of the St. Lawrence River (p-value = 0.002; adjusted R^2: 0.993; n = 4). In the south-western transect, northern populations from sites SW4 and SW5 show higher levels of diversity than sites further south, which contradicts our expectation under a northward range expansion. Interestingly, the northern-most population, N4, shows exceptionally high levels of observed heterozygosity (Het_{obs} = 0.169) and low levels of inbreeding (F_{IS} = 0.026) relative to other sites in the northern transect. As a comparison, the population on Montreal Island (N2) has an almost 8-fold higher F_{IS} (F_{IS}=0.204) than the northern-most site. The high variation seen in the northern-most site also stands in contrast to the relatively low genetic diversity observed in the nearest site to the south, N3 (Het_{obs}=0.118; F_{IS} = 0.157). Private alleles ranged from zero (in eight of 13 sites) to 14 (in the northern-most site, N4).

We used NeEstimator to approximate the contemporary effective population size for each population (Supplementary Table S1). Consistent with our hypothesis of northward expansion, the northern-most site also has the lowest effective population size (Ne = 14). This is consistent with our result that indicates outcrossing in this site given effective population size is inversely related to the probability of introgression. However, the expectation is contradicted south of the St Lawrence River, with northern populations such as SW5 and SE4 showing a higher effective population sizes than SW1 and SE1, respectively. Concordant with post-glacial expansion, we found a significant reduction in observed heterozygosity relative to the expected levels under Hardy-Weinberg equilibrium (HWE) in two of three *P. leucopus* transects (Table 1): the northern and south-western transects (one-tailed paired t-test; northern transect: p-value = 0.031; south-western transect: p-value = 0.017; south-eastern transect: p-value = 0.104). However, it is worth noting that apparent reductions in overall heterozygosity can also occur as a result of sub-population structure (‘Wahlund effect’).

To further look for a history of recent and post-glacial range expansion, we analyzed patterns of allele frequency divergence among populations. Overall, there was relatively low levels of differentiation among *P. leucopus* populations, with a mean F_{ST} of 0.037 (95% confidence interval [CI]: 0.033 – 0.041). 78 pair-wise F_{ST} estimates (Supplementary Table S2) between and within transects show relatively high divergence of *P. leucopus* south and north of the St. Lawrence River (Figure 5). Population divergence within transects is also higher along the northern transect, consistent with populations having experienced greater amounts of drift.

Populations south of the St. Lawrence River are less differentiated from each other than from populations north of this river, a pattern which supports our expectation of distinct post-glacial routes separated by the St. Lawrence River.

Increased genetic drift due to allele surfing during range expansion is expected to result in wave-front populations with high F_{ST}. Consistent with this expectation, the northern-most site N4 shows high F_{ST} (N4: $F_{ST} = 0.059$, CI: 0.050 – 0.068) relative to the average. The isolation by distance model shows a significant linear relationship between differentiation relative to southern-most sites and distance (Figure 6; p-value = 0.006; Adjusted $R^2 = 0.58$). The slope of the regression is 0.0089, thus a rough approximation of the neighbourhood size (i.e. neighbourhood size or $4\pi D\sigma 2$) is 112 individuals. A substantially greater divergence is expected across species than between populations, and summary statistics from the complete dataset shows F_{ST} estimates between *P. leucopus* and *P. maniculatus* are approximately an order of magnitude greater (mean $F_{ST} = 0.384$; SD = 0.018) than the mean F_{ST} observed between *P. leucopus* populations in southern Quebec.

Population genomic structure

Overall genetic variation in *P. leucopus* was summarized with PCA which shows population structure is geographically correlated (Figure 3). Lineages on either side of the St. Lawrence River are separated by PC1 (8.1% of variance) while individual populations tend to differentiate across PC2 (2.6% of variance). A gradient in population structure is observed within some populations, including SE4, SW4, N3 and especially N4. These populations cluster more widely across PC space relative more southern populations, such as SE1, N1, SE2, and N2, where individuals tend to cluster more closely together. However, the northern site SW5 did not display this gradient pattern while it was apparent in more southern sites, like SW1 and SW2.

Analyzing between-group differences with DAPC supports two main lineages (K=2) north and south of the St. Lawrence River (Figure 4a), with more subtle variation separating populations east and west of the Richelieu River (Figure 4b). These results provide support our hypothesis that the Richelieu River and the preceding southern waterways split a glacial lineage expanding northward from an eastern refugium. We also evidence of migration across major these rivers. Both PCA and DAPC show one individual from the south-western transect (SW1) with genetic variation more typical of that found in the northern transect. Similarly, a few individuals from sites SE3 and SE4 have genotypes more similar to those found in the south-western transect.

We used ADMIXTURE to analyze the ancestry of *P. leucopus* from southern Quebec and look for evidence of hybridization with *P. maniculatus*. After cross validating 12 potential ancestral populations, our results indicate up to nine relatively well-supported source populations contributing to ancestry (Supplementary Figure S2b). Ancestry proportions under the particularly well supported model of K=4 shows *P. leucopus* population genomic structure associated with extrinsic barriers to dispersal (rivers), as well as admixture in SW5 and SW4 from different transects (Figure 7), which is consistent with the relatively high nucleotide diversity and heterozygosity observed in these two sites. Supporting ordination, the most ancestral *P. leucopus* clusters (K=3) consist of individuals on either side of the St. Lawrence River, while the Richelieu River separates more subtle substructure which is consistent with a more recent divergence. We find additional sub-structure in the northern-most site, N4 and admixture in N3 (K=6).

An admixture pattern consistent with range expansion or admixture (Falush *et al.*, 2016) is observed in the two northern sites, in particular at K values six through nine. The second northern-most population, N3, is admixed with ancestry common to the south (e.g. N1) and to the north (e.g. N4). Falush *et al.*, (2016) showed through simulations that this type of pattern would be consistent with gene flow, in this case from N1 and N4 to N3. However, such patterning of population genomic structure is also consistent with a bottleneck, such as during expansion of the N4 lineage subsequent to the split from N3 (Excoffier and Ray, 2008; Falush *et al.*, 2016), or through introgression in N4 from an unsampled ('ghost') population (Currat *et al.*, 2008; Falush *et al.*, 2016). Indeed the relatively high observed levels of heterozygosity and low inbreeding (F_{IS}) in N4 suggest the latter. Further sampling of *P. leucopus* and *P. maniculatus* north of the St. Lawrence River is warranted.

Ancestry proportions also show evidence of ancestral gene flow across the St. Lawrence River (e.g. K=3). Consistent with older admixture, the shared ancestry proportions are well homogenized within the populations. More recent admixture can be inferred when admixture proportions vary within a population because recombination has not had time to evenly distribute genomic blocks across chromosomes. We find hierarchical structure in SW5 and SE1 (K=6) likely due to isolation. Site SW5 is situated in a park surrounded by the City of Longueuil, a city which has experienced a 58-fold increase in census human population during the most recently documented 143 year period (approx. 286 mice generations), increasing from 3,977 in 1871 to 231,409 in 2014 (Statistique Canada, 2014). Site SE1 is surrounded, other than to the north-east, by Lake Champlain and the Richelieu River. Finer population genetic structure is apparent in SW1 (K=9).

Admixture with *P. maniculatus*

Notably, our ancestry results show evidence of hybridization with *P. maniculatus*. A predicted consequence of expanding populations with low effective population size is potential introgression from closely related local species. ADMIXTURE results at K=2 shows the presence of one putative *P. leucopus* hybrid in site SE3 with 15.5% of its ancestry inferred as coming from *P. maniculatus* (Figure 7). Two additional individuals have 1.6% and 1% inferred *P. maniculatus* ancestry, a percentage at least three orders of magnitude greater than in non-admixed individuals (0.001%). Given that most individuals in this site showed little to no interspecific ancestry, we suggest this result may be explained by a recent hybridization. Specifically, these ancestry proportions are consistent with ancestry expected in F3 (12.5%) and F6 (1.6%) hybrids backcrossed with *P. leucopus*. Our results also show that two out of the 10 *P. maniculatus* sampled share 18.8% and 7.5% ancestry proportions with *P. leucopus*. Further support for hybridization comes from PCA (McVean, 2009) on the complete dataset, which shows the two species are separated by PC1 (22% of cumulative variance) and the putative hybrids with more than 10% hetero-specific ancestry as intermediate genotypes compared other individuals (Supplementary Figure S4).

Bifurcating population ML tree

We used TreeMix to visualize population relationships and infer demographic history of *P. leucopus*. Matching our analyses above, the ML tree supports an older divergence of mice on either side of the St. Lawrence River and a more recent split caused by the Richelieu River (Figure 8a). Interestingly, however, the ML tree places populations N2 and SW5 as basal lineages. The ML tree indicates that the inbred population on Montreal Island (N2) is from an earlier split sharing common ancestor with a lineage that later diverged to expand north (N3 and N4), while another remained at a more southern latitude (N1). Consistent with drift during expansion, the northern-most site shows substantial accumulated drift since splitting from N3, as indicated by the relatively long branch length. The ML tree indicates the population partially isolated by the City of Longueuil (SW5) is a basal *P. leucopus* lineage historically present in that area. Surprisingly, however, the ML tree indicates SW5 diverged from other *P. leucopus* in Quebec prior to the split of populations on either side of the Richelieu River. This would suggest that other south-western populations, like SW3, are more closely related to mice in the south-eastern transect than they are to mice in SW5, a hypothesis not supported by DAPC or the ancestry proportions. Indeed, allele frequencies in SW5 are significantly less diverged (as indicated by F_{ST}) from other south-western populations (mean pair-wise $F_{ST} = 0.017$) than they are relative to south-eastern populations (mean pair-wise $F_{ST} = 0.026$; one-tailed paired t-test: p-value = 0.016). A possible explanation for the placement of SW5 as a unique lineage south of the St. Lawrence River is the complex evolutionary history of mice in this site, which show ancestral admixture from both transects (K=4; Figure 7), as well as built-up sub-structure likely due to more recent isolation (K=9).

Test for introgression

Methods that calculate ancestry proportions, like ADMIXTURE, can give misleading inferences of population structure (e.g. optimal K, Puechmaille, 2016) when sampling is incomplete or uneven (Falush *et al*., 2016; Wang, 2016). We therefore used TreeMix to formally test the bifurcating model of genealogical evolution. A residual fit of the bifurcating tree without gene flow (Figure 8b) shows that this model does not fit the evolutionary history of the *Peromyscus* populations in this thesis, instead supporting a reticulate event between *P. leucopus* from SE3 and out-group *P. maniculatus* (Figure 8c). Allowing for two more migration events supports past gene flow from N1 to SW1 (Supplementary Figure S5a, b), and from SE1/SE2 to *P. maniculatus* (Supplementary Figure S5c, d),although adding three migration events may be over-fitting as the topology of the ML tree is altered under this model.

Table 1 Site and Genetic diversity summary of P. leucopus individuals sampled from 13 sites on one northern and two southern transects. Site 14 consisted of only P. maniculatus. Genetic diversity measures include nucleotide diversity (Pi), observed heterozygosity (Hetobs) and homozygosity (Homobs), expected heterozygosity (Hetexp) and homozygosity (Homexp), the inbreeding coefficient (FIS) and number of private alleles. In the northern transect, distance to the southern-most site (DSS) is included. For sites south of the St. Lawrence River, the distance to the latitude of the southern-most site (DSL) in the SW and SE transect is shown.

Site	Transect	N	2013	2014	Latitude	Longitude	DSS (km)	π	Het_{obs}	Hom_{obs}	Het_{exp}	Hom_{exp}	F_{IS}	Private
N1	N	19		19	45.33	-74.4	0	0.184	0.141	0.859	0.177	0.823	0.132	5
N2	N	19		19	45.43	-73.94	37.56	0.182	0.118	0.882	0.176	0.824	0.204	9
N3	N	11		11	46.07	-73.28	120.05	0.178	0.118	0.882	0.167	0.833	0.157	0
N4	N	20	20		46.3	-73.09	148.25	0.175	0.169	0.831	0.171	0.830	0.026	14

Site	Transect	N	2013	2014	Latitude	Longitude	DSL (km)	π	Het_{obs}	Hom_{obs}	Het_{exp}	Hom_{exp}	F_{IS}	Private
SW1	SW	18		18	45.04	-74.46	0	0.154	0.135	0.865	0.149	0.851	0.059	0
SW2	SW	17		17	45.22	-73.93	19.94	0.145	0.112	0.888	0.139	0.861	0.101	0
SW3	SW	19	19		45.25	-73.47	23.55	0.167	0.159	0.841	0.163	0.838	0.030	0
SW4	SW	12		12	45.54	-73.32	55.5	0.165	0.137	0.863	0.157	0.843	0.081	0
SW5	SW	20		20	45.55	-73.47	56.83	0.172	0.162	0.838	0.168	0.833	0.033	2

Site	Transect	N	2013	2014	Latitude	Longitude	DSL (km)	π	Het_{obs}	Hom_{obs}	Het_{exp}	Hom_{exp}	F_{IS}	Private
SE1	SE	20	20		45.06	-73.29	0	0.171	0.159	0.841	0.166	0.834	0.037	6
SE2	SE	8	8		45.42	-73.07	40.35	0.173	0.166	0.835	0.162	0.838	0.021	0
SE3	SE	20		20	45.66	-72.75	66.39	0.169	0.160	0.840	0.164	0.836	0.031	0
SE4	SE	20	13	7	45.87	-72.56	89.53	0.167	0.153	0.848	0.163	0.837	0.047	0
Pm	SE	10		10	45.74	-72.41								

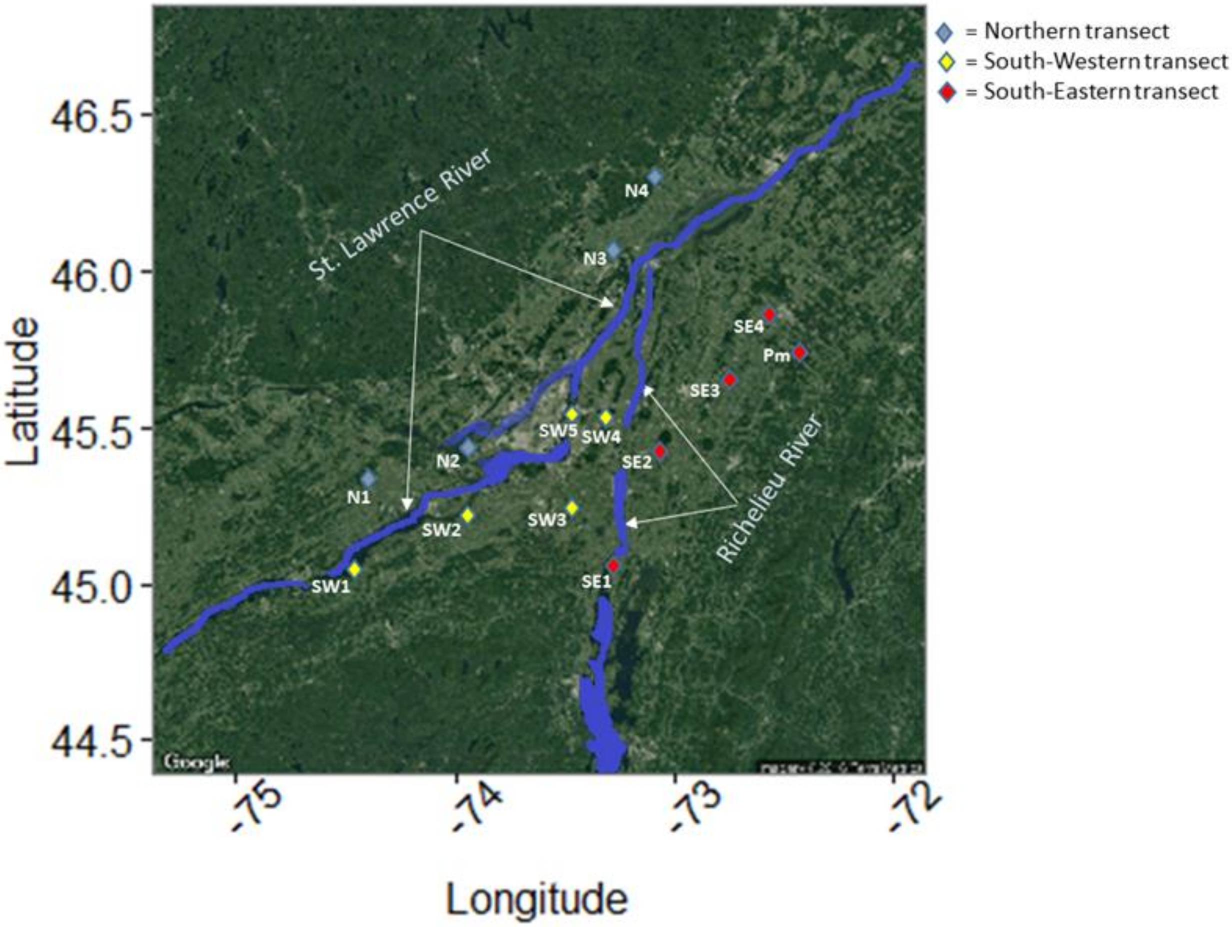


Figure 1 Study area and the geographic location of 13 *P. leucopus* sites and one *P. maniculatus* (Pm) site genotyped in southern Quebec. Major barriers to gene flow (rivers) are labeled.

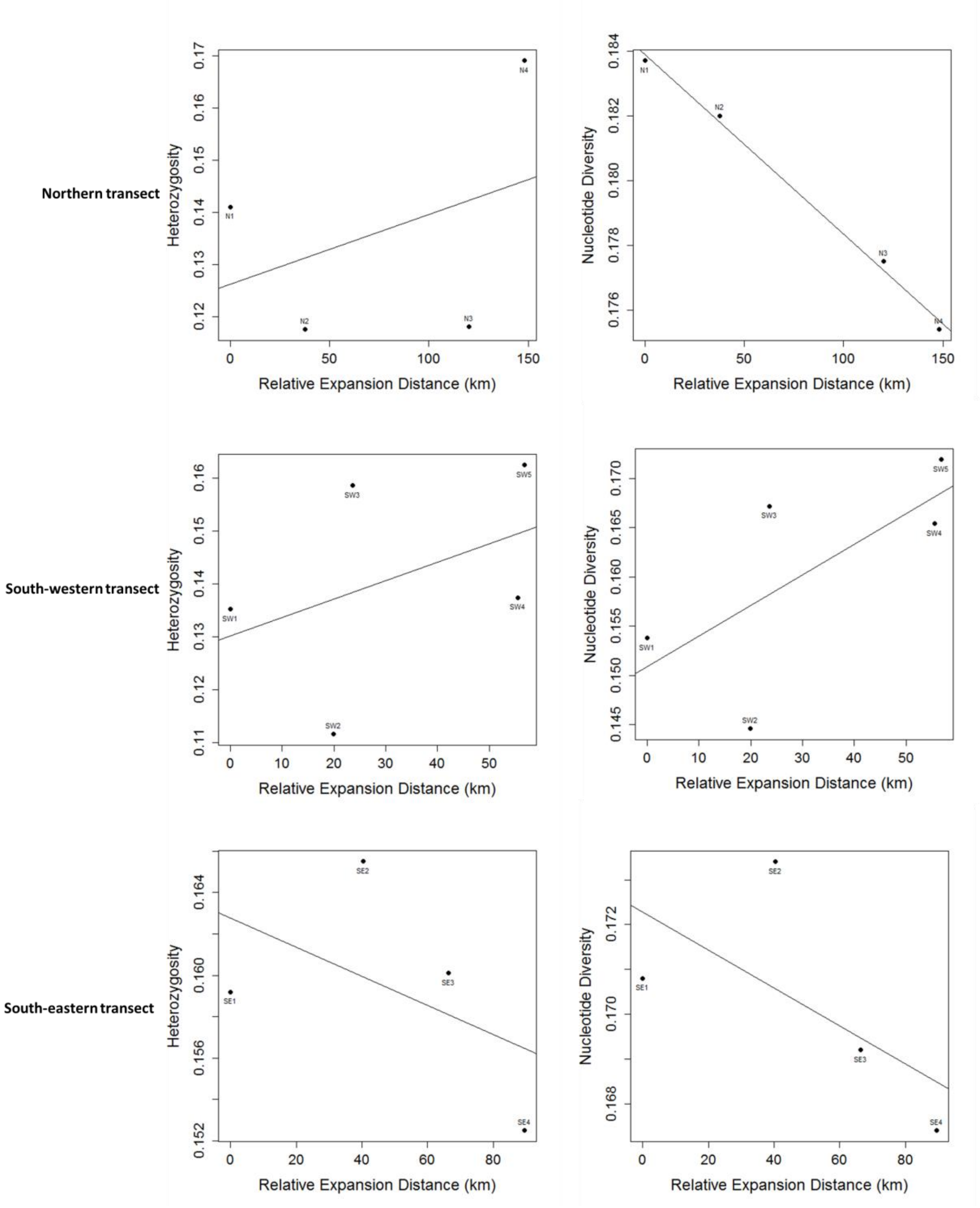


Figure 2 Regressions of *P. leucopus* genetic diversity against relative expansion distance. Plots show heterozygosity in the northern (a), south-western (c), and south-eastern (e) transects, as well as nucleotide diversity. Only one regression was significant, showing a linear decrease in nucleotide diversity in populations from the northern transect (p-value = 0.002; adjusted R^2: 0.993).

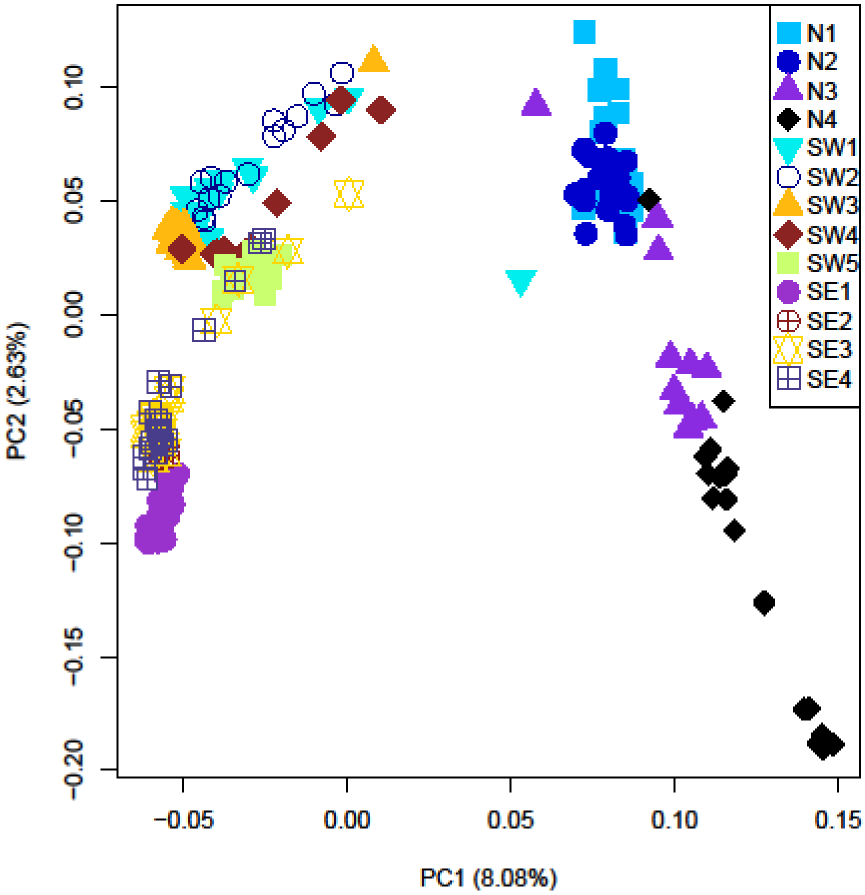


Figure 3 PCA on genome-wide variation (38,144 SNPs) of 13 *P. leucopus* populations (n=223) from southern Quebec.

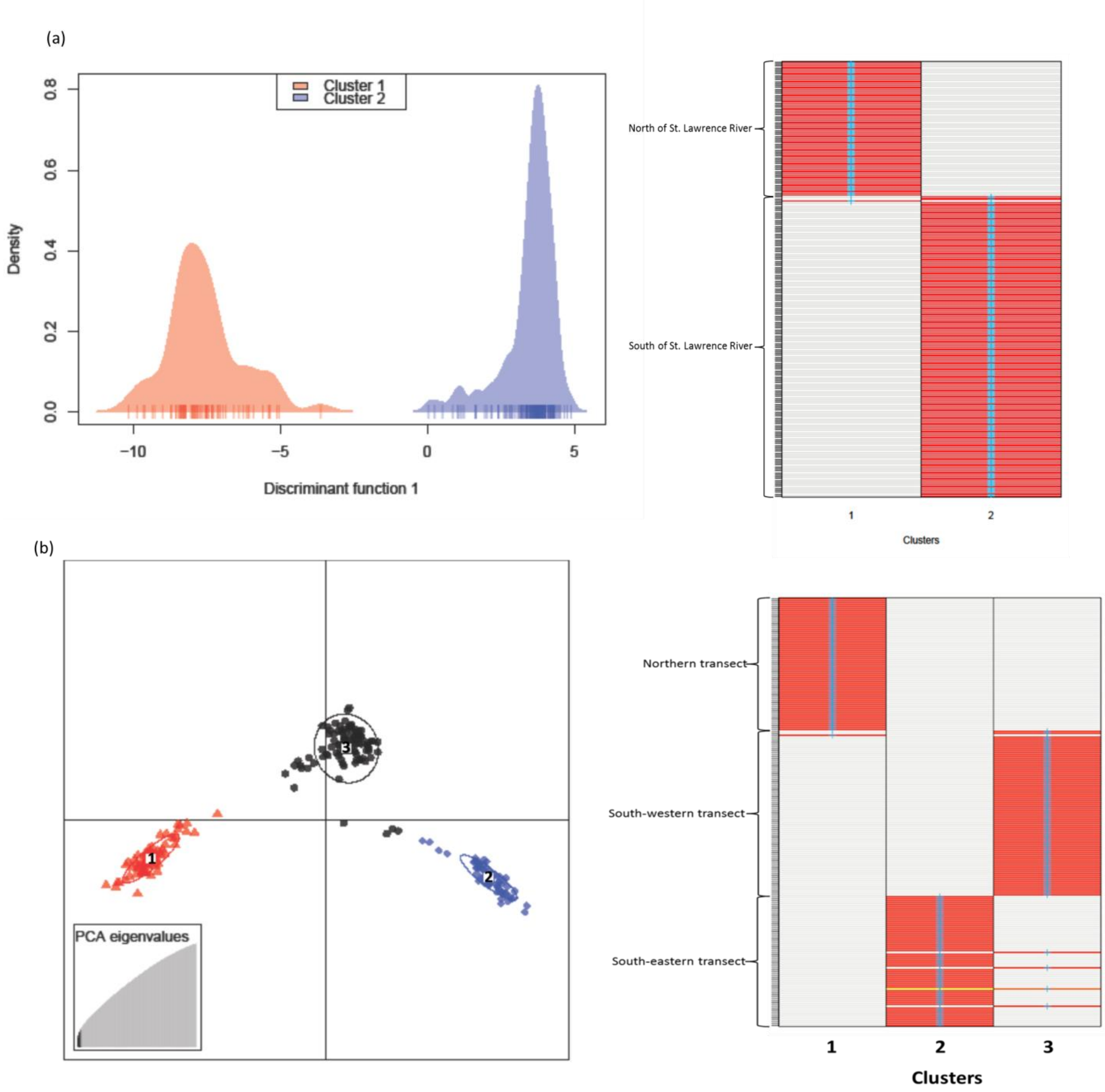


Figure 4 Population genomic structure inferred by DAPC (38,144 SNPs) on 6 retained PC eigenvalues (inset plot: 12.1% of variance) shows (a) two main *P. leucopus* lineages (K=2) separated by the St. Lawrence River, while (b) the Richelieu River divides more subtle population genomic structure. The outset plots show each individual as a row and transects delimited by arrows, colour (red) denotes which cluster each individual is assigned to. K=2 and K=3 were the best supported number of lineages, as shown by supplementary Figure S2a.

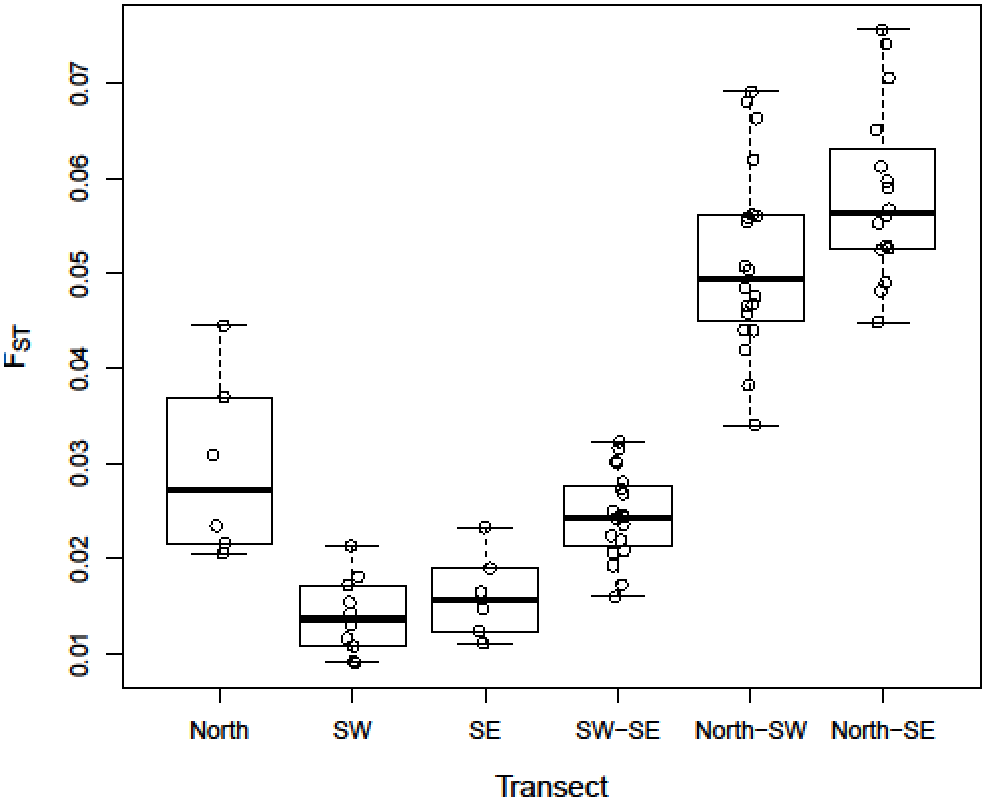


Figure 5 Summary of 78 pair-wise F_{ST} estimates of populations within and between transects (raw data is presented on supplementary Table S2).

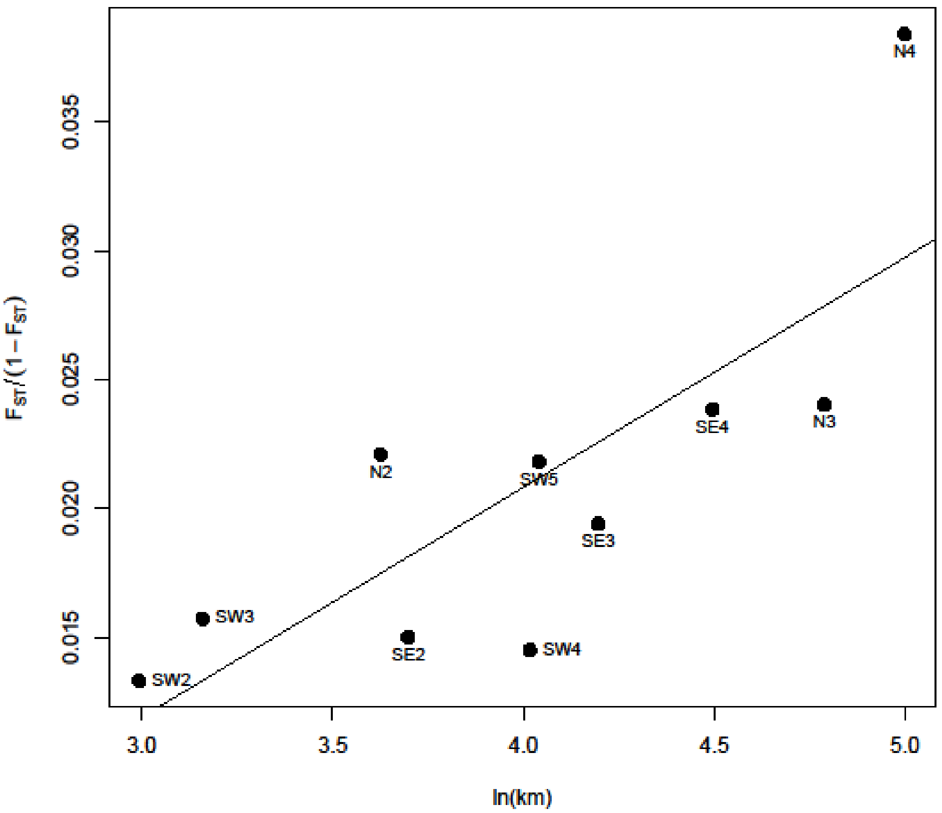
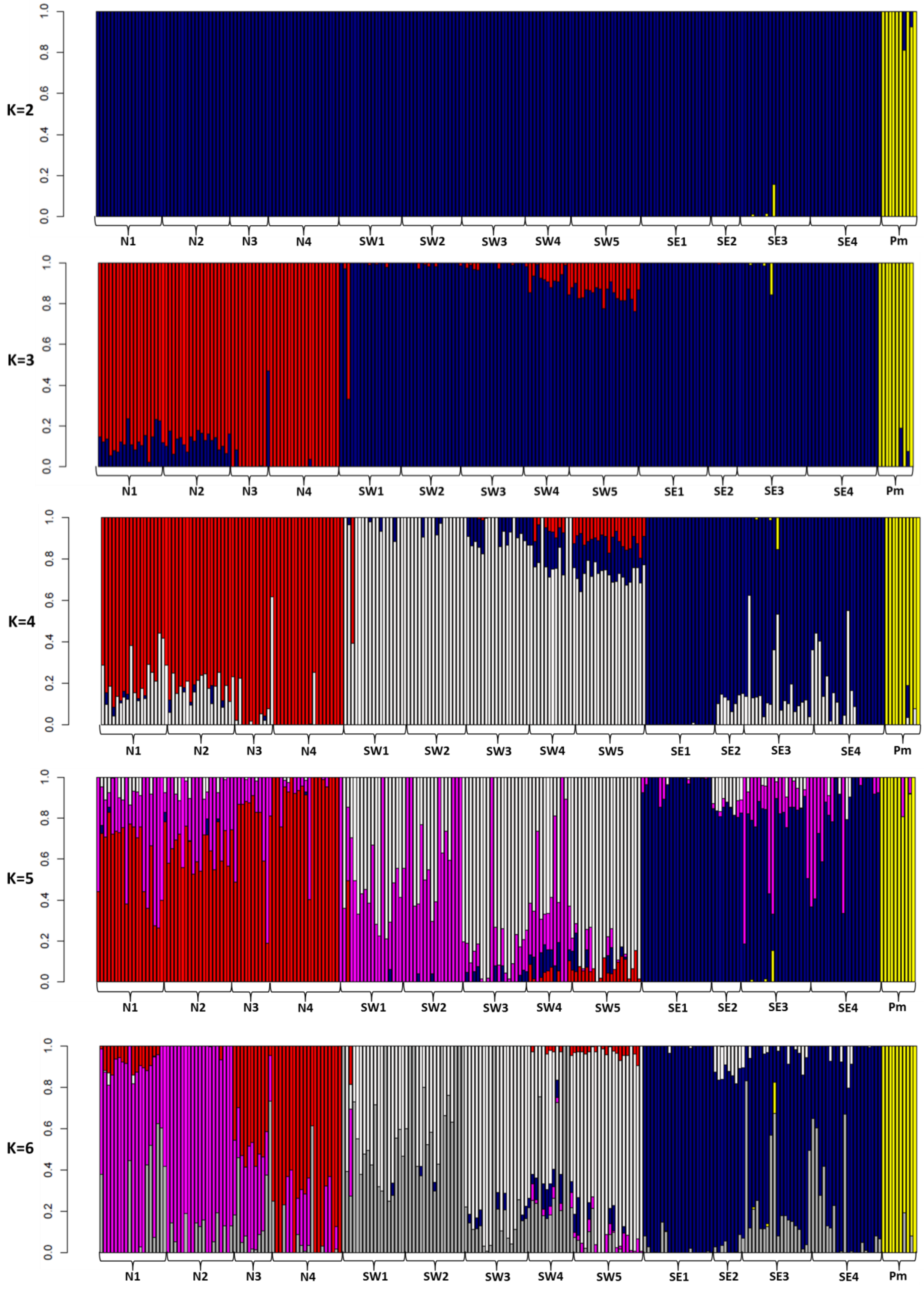


Figure 6 Isolation by distance model (Rousset, 1997) of *P. leucopus*.. Plotted is differentiation (F_{ST}/ (1-F_{ST})) between 10 populations and their southern-most site against the natural log of the distance. The regression is y = 0.0089x – 0.0149.


K=2
0.0
0.2
0.4
0.6
0.8
1.0
N1
N2
N3
N4
SW1
SW2
SW3
SW4
SW5
SE1
SE2
SE3
SE4
Pm
K=3
K=4
K=5
K=6

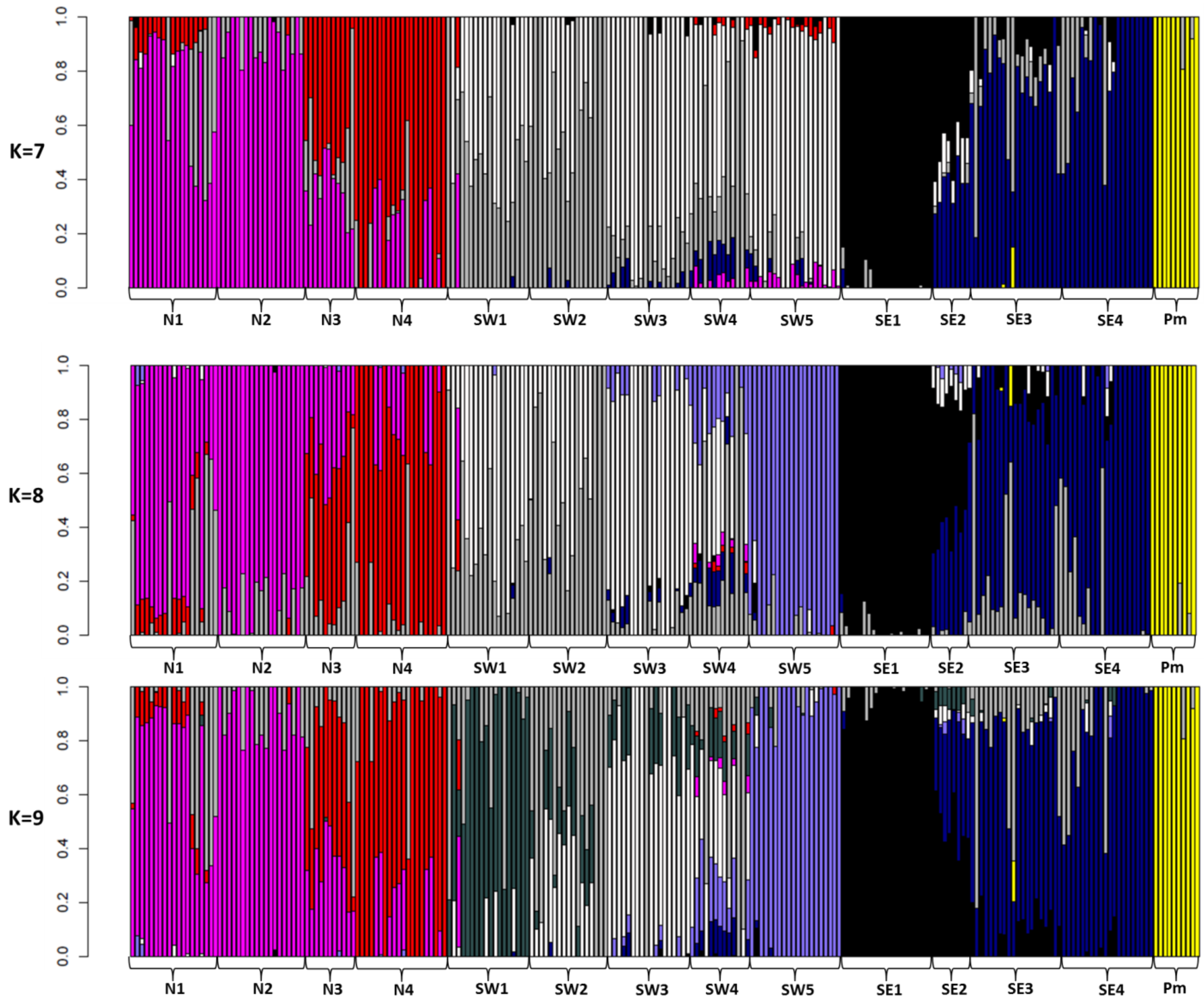
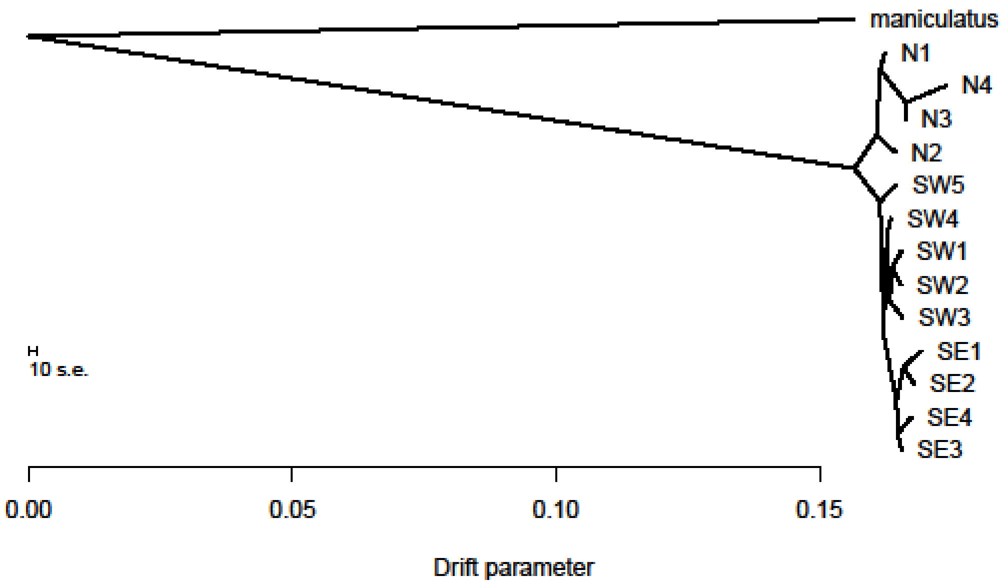


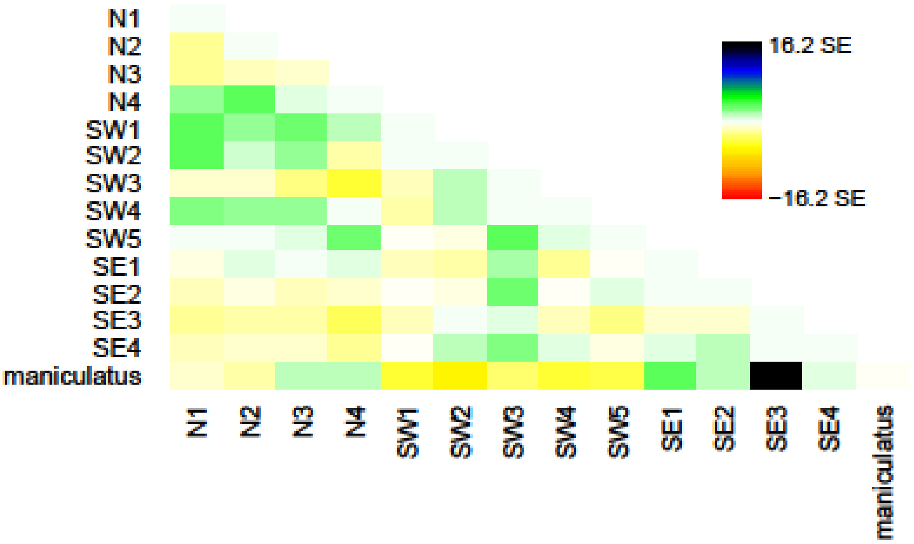
Figure 7 Ancestry proportions (12,507 SNPs) for 13 *P. leucopus* sites and one *P. maniculatus* site (Pm) for nine relatively well supported source populations. Models with four (K=4) and six (K=6) ancestral source populations had marginally best cross-validation scores (as shown on Supplementary Figure S2b).

(a)



maniculatus
N1
N4
N3
N2
SW5
SW4
SW1
SW2
SW3
SE1
SE2
SE4
SE3
10 s.e.
0.00
0.05
0.10
0.15
Drift parameter

(b)


N1
N2
N3
N4
SW1
SW2
SW3
SW4
SW5
SE1
SE2
SE3
SE4
maniculatus
16.2 SE
−16.2 SE

(c)

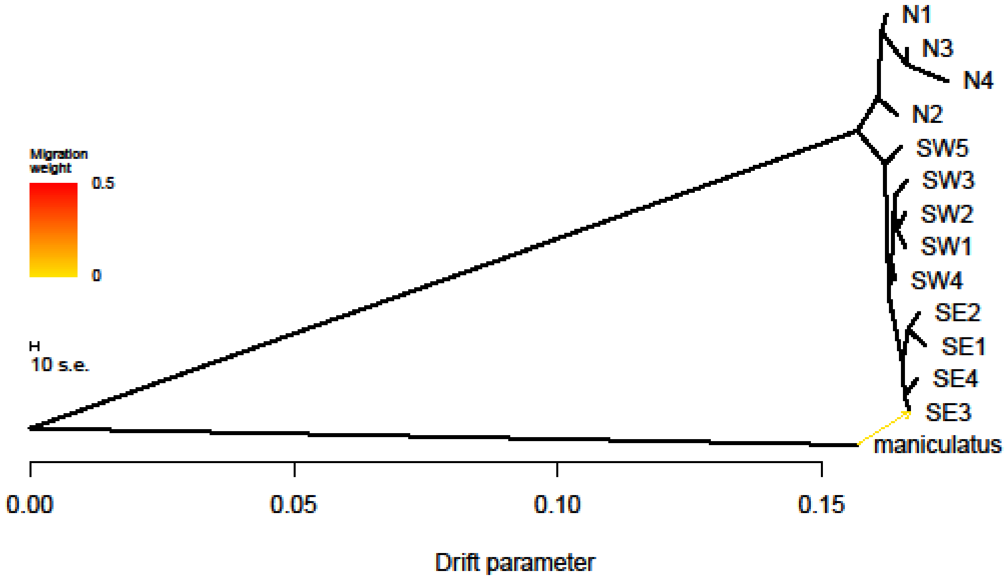


Figure 8 Population relationships inferred with TreeMix using 12,507 SNPs (a) under a bifurcating model without gene flow. Horizontal branch length represents amount of evolutionary change according to the drift parameter (related to F_{ST}). (b) The residual fit of model (a), a scenario without gene flow after divergence. The scale bar represents 10X the average standard error of the entries in the sample covariance matrix (Pickrell and Pritchard, 2012). Residuals are qualified using the colour palette. Positive residuals (green/blue/black) represent population pairs that are more closely related than presented on the tree, and are thus candidates for admixture. Residuals below zero (yellow/orange/red) represent population pairs that are less closely related to each other than shown on the tree. (c) Population tree with one migration edge, showing gene flow between *P. maniculatus* and *P. leucopus* from SE3. Direction of gene flow is shown by the arrow. The colour of the arrow denotes the weight of migration (proportional to the amount of genetic ancestry contributed by the immigrant.

CHAPTER III

DISCUSSION

We used genomic tools to investigate the history of range expansion of *P. leucopus* in southern Quebec. We predicted that allele surfing and genetic bottlenecks during recent northward expansion would lead to northern populations with reduced genetic diversity, divergent allele frequencies, and potentially admixture. There was little evidence of range expansion south of the St. Lawrence River. However, consistent with these processes, population genetic summary statistics showed the northern-most population on the northern transect with reduced nucleotide diversity, divergent allele frequencies (as indicated by F_{ST}), a high number of private alleles, and heterozygosity levels that suggest admixture with an unsampled population. We observed that population genomic structure correlates with geography. We showed that two main lineages of *P. leucopus* are present north and south of the St. Lawrence River, with additional substructure differentiating southern populations east and west of the Richelieu River. We argue that these results likely reflect the evolutionary history of isolation during post-glacial expansion of two glacial lineages. We also documented gene flow between *P. maniculatus* and *P. leucopus*, undermining the notion of complete reproductive isolation between these two species. We end by discussing the relevance of our results in an evolutionary, conservation, and public health context.

Spatial genomic gradients in *P. leucopus*

We evaluated population genetic diversity of *P. leucopus* across southern Quebec and found clines consistent with northward range expansion. Two of three transects showed northern populations with the least genetic diversity in at least one measure. The northern transect showed a significant linear decrease, although the overall difference in diversity was relatively small considering the distance between the southern-most and northern-most populations (~150 km). Contrasting our expectations, a positive trend was seen in the south-western transect due to higher levels of diversity in admixed northern locations. Our isolation by distance model estimated gene flow (112 effective migrants) along a roughly north to south direction. However, this estimate is based on a number of likely unrealistic mathematical assumptions (Whitlock and Mccauley, 1999) and thus should be viewed only as a rough approximation and interpreted with caution.

Various factors likely limited our power to detect latitudinal gradients, such as inter-site differences. For example, larger forest patches contain more individuals and are thus likely to harbour greater genetic diversity. The presence of partial barriers to dispersal between sites (e.g. roadways) in the south-western transect (Rogic et al., 2013; Marrotte et al., 2014) may further blur any relationship between genetic diversity and expansion distance at the spatial scale and sampling density used here. Wave-front dynamics may also play a role, as simulations have shown that climate-driven range expansions maintain genetic diversity on the expanding range margin (Nullmeier and Hallatschek, 2013; Dai et al., 2014), although diversity may still decrease under models of rapid climate change (Garnier and Lewis, 2016).

Further limiting our ability to detect consistent spatial gradients in genetic diversity, our results indicate that some populations are partially isolated and therefore unlikely to have undergone recent expansion. For example, the relatively high levels of inbreeding in mice on Montreal Island (N2) suggest this population has been isolated for some time. It is also likely that the demographic response of *P. leucopus* to climate change, in particular in the south-western transect, is an increase in abundance rather than northern extension. For example, although *P. maniculatus* was the more commonly found *Peromyscus* species as recently as 40 years ago, a historical presence of *P. leucopus* at low abundance (Grant, 1976) may have contributed genetic variation during more recent migration from the south.

P. leucopus is one of many North American mammals (Lessa et al., 2003) that expanded north during interglacial events of the Pleistocene. Mitochondrial DNA data indicates that at the start of the current interglacial period two *P. leucopus* lineages expanded north from eastern and western refugia approximately 17,000 and 15,000 bp, respectively (Rowe et al., 2006). The genetic legacy of the older and more geographically extensive post-glacial expansion may be reflected as a reduction of observed heterozygosity relative to Hardy-Weinberg expectation in *P. leucopus* populations sampled in southern Quebec, a region covered by the Laurentide Ice Sheet 20,000 years ago. However, it is worth noting ('Wahlund effect') that reduced heterozygosity in a population at HWE can also arise due to allele frequency differences among sub-populations. Our samples could be described as consisting of three meta-population lineages separated by geographic barriers. The role of the Wahlund effect should be minimal in our results since our comparison of observed and expected heterozygosity was based on estimates from individual sub-populations rather than from meta-populations.

Population genomic structure in *P. leucopus*

Geographic isolation is generally thought to be a common precursor for generating genetic diversity and ultimately new species (Sobel et al., 2009). We found that genomic variation in *P. leucopus,* as in human populations (Novembre et al., 2008), correlates with geography. We identified two putative glacial lineages of *P. leucopus* north and south of St. Lawrence River. These glacial lineages likely began diverging during the Late Pleistocene in western and eastern refugia, respectively (Rowe et al., 2006; Fiset et al., 2015). More subtle genetic correlations can differentiate *P. leucopus* on either side of the south-north oriented Richelieu River, which is consistent with previous studies (Rogic et al., 2013). This supports a scenario in which ancestral populations from the eastern glacial lineage were separated during northward post-glacial expansion. Populations expanding from an eastern refugium would have been separated by what was then the ancient Champlain Sea left by the retreating ice sheet, what eventually became Lake George and Lake Champlain in the states of New York and Vermont, respectively, and the Richelieu River in Quebec. *P. leucopus* populations from the western refugium, on the other hand, colonized the Great Lakes area during post-glacial expansion (Rowe et al., 2006). It is likely that the northern range limit of *P. leucopus* was relatively stable prior to more recent anthropogenic disturbances of climate and habitat. Ongoing climate change is thought to be responsible for the rapid expansion of *P. leucopus* across the Upper Michigan Peninsula (Myers et al., 2009), where it is ecologically replacing *P. maniculatus* (Wan, 2014), a process which can be facilitated by hybridization (Rhymer and Simberloff, 1996). Although we lack reference populations south of the international border, it is likely that populations north of the St. Lawrence River are an extension of the expansion detected across the Upper Michigan Peninsula and are descendant of a western glacial lineage. The higher levels of drift (F_{ST}) and inbreeding (F_{IS}) observed north of the St. Lawrence River than south of it suggests the north shore may represent a more recently established colonization route.

Populations on an expanding range margin are expected to become differentiated due to the spatial accumulation of genetic drift (allele surfing) and selection experienced during expansion. During spatial expansion into a new territory, low frequency variants can become established where they, in subsequent generations, can increase to high frequencies and lead to 'genetic revolutions' of wave-front population structure (Exoffier and Ray, 2008). Consistent with processes associated with expansion including genetic bottlenecks, allele surfing, and potentially introgression, we showed that the northern-most population had relatively low nucleotide diversity, the most differentiated allele frequencies (as indicated by F_{ST} and the drift parameter), the highest number of private alleles, the lowest effective population size and, as predicted, heterozygosity estimates that suggest outcrossing with an unsampled population. Outcrossing with an unsampled population can also explain the high number of private alleles observed in this site. Similarly, ancestry proportions showed a pattern of population structure consistent with a northward bottleneck or admixture with an unsampled population (Falush et al., 2016).

We found evidence that some populations are at least partially isolated. We found *P. leucopus* inhabiting Montreal Island with higher levels of inbreeding relative to populations on the mainland. We also found population structure in a *P. leucopus* population in an area (City of Longueuil) which has had a 58-fold increase in census human population during the most recently documented 143 year period, approximately 286 generations of *P. leucopus* (Statistique Canada, 2014). Urbanisation can have major effect on population genetic structure through isolation and likely imposes novel selective pressures. Munshi-South et al. (2016) used human population size and percent impervious surface cover in New York City to study the effects urbanization had on genomic variation of urban *P. leucopus*. They found that these proxies best explained the genome-wide variation in urban *P. leucopus* and Harris et al. (2013) identified candidate genes potentially under selection in an urban environment, many of which were associated immune system processes. The isolated *P. leucopus* populations identified in this thesis thus provide an opportunity to study different evolutionary processes, such as adaptation in response to a recent massive human expansion, and the effects of inbreeding on fitness in an island population.

Population splits

Our results of the demographic history of *P. leucopus* in southern Quebec, as shown by population splits, support a scenario of post glacial and ongoing expansion. Our analysis of population splits showed putative glacial lineages as clades separated by the St. Lawrence River, and a post-glacial divergence of populations on either side of the Richelieu River. Our results also suggest that *P. leucopus* was historically present, one lineage became isolated on Montreal Island, splitting off from a lineage remained on the mainland (N1) and one that expanded north (N3 and N4). Indeed our results show that the northern-more site has accumulated substantial genetic drift. Our results showed the population surrounded by recent urbanisation (SW5) has been historically present and had a complex evolutionary of ancient admixture and isolation. The placement of this population as basal relative to all others south of the St. Lawrence River may due to this complex evolutionary history of ancient admixture and recent isolation that has occurred in this lineage and likely violates the bifurcating model.

Secondary contact between *P. leucopus* and *P. maniculatus*

One of the more neglected consequences of range expansion is the potential for local introgression. We hypothesized that northern expansion of *P. leucopus* populations into territory historically occupied by *P. maniculatus* would lead to introgression. We used genome-wide data to test this prediction, and our results showed the presence of a few putative recombinant hybrids. However, these putative hybrids were identified using clustering methods that are sensitive to sampling design and can produce results that are easily misinterpreted (Schwartz and McKelvey, 2008; Novembre and Stephens, 2008; Falush et al., 2016; Puechmaille, 2016; Wang, 2016). Such clustering methods also do not directly test for admixture between divergent gene pools. To overcome this, we used TreeMix to explicitly test for gene flow and reject the bifurcating model of genealogical evolution, confirming secondary contact between *P. leucopus* and *P. maniculatus*.

Only five individuals showed hybrid ancestry, which is likely due to ecological differences between the two species. *P. leucopus* was sampled mostly from small woodlots scattered across an agricultural matrix where, as a generalist, *P. leucopus* thrives and our capture records indicate *P. maniculatus* is mostly absent. In contrast, there is evidence of more introgression occurring in larger and more continuous forests where *P. maniculatus* and *P. leucopus* can be locally found (Leo and Millien, 2016). In particular, Leo and Millien (2016) sampled both species where *P. maniculatus* was more abundant and found putative *P. leucopus* hybrids with ancestry proportions composed mainly of *P. maniculatus* ancestry (e.g. sample 1968 and 2023), a result consistent with predictions (Currat et al., 2008). However, the results of Leo and Millien (2016) come with the aforementioned caveat of clustering methods as well as less resolution provided by microsatellites. Taking a genomic approach and sampling broadly in sites where both species are found, including north of the St. Lawrence River, and testing for gene flow will help expose the frequency and extent of hybridization between these two species of *Peromyscus* in Quebec.

We note that although *P. maniculatus* was represented by only 10 individuals, we believe our inference of interspecific admixture is robust. For example, a mammal population genomic study showed that sufficient genetic variation could be sampled from a population with as few as

10 individuals to construct a complete evolutionary history (Trask et al., 2011). The putative hybrids found by Leo and Millien (2016) were also sampled from a more evenly distributed dataset of 69 *P. leucopus* and 84 *P. maniculatus*. Furthermore, the accuracy of hybrid identification increases with the number of markers and level of divergence, reaching 90% accuracy with fewer than 50 microsatellite markers in populations one third as diverged as the species in this thesis (Vaha and Primmer 2006). Importantly, Vaha and Primmer (2006) showed that unsampled source populations have negligible effects on hybrid identification, and McVean (2009) demonstrated that the positions of genotypes on PC space relative to non-admixed individuals can be used identify hybrids (e.g. Supplementary Figure S4), even when missing source populations.

Our results challenge the idea of complete isolation between *P. leucopus* and *P. maniculatus*, supporting an emerging view that reproductive isolation can vary depending on individual genotype and demographic context (Chunco, 2014; Senerchia et al., 2016; Kozlowska et al., 2012; Gompert and Buerkle, 2016; Mandeville et al., 2015; Araripe et al., 2016). The presence of hybrids on the range margin also supports the idea that pre-mating barriers (e.g. Doty, 1972, 1973) between sympatric species may be altered during climate-driven range shifts; in essence a displacement of co-evolved genotypes driven by anthropogenic climate change (Crispo et al., 2011; Chunco, 2014).

Hybridization is expected to lead to costs in fitness due to intrinsic and extrinsic selective pressures. However, some of the fitness costs associated with hybridization during range expansion may be mitigated by evolutionary phenomena associated with expansion or small populations, including softened selection (Peischl et al., 2013), fixation of compensatory adaptive alleles (Poon and Otto, 2000), or a decrease in the rate of small-effect deleterious mutations (LaBar and Adami, 2016). Selection against, and purging of polymorphic incompatibility loci (Cutter, 2012) may also result in a rebound of hybrid fitness (Araripe et al., 2016). These processes may contribute to the conversion of a climate-driven wave-front of a single species to a moving hybrid zone (e.g. Chunco, 2014; Taylor et al., 2014, 2015).

`

Climate-induced expansions and conservation

There is ongoing debate among conservationists regarding the potential advantages (Hamilton and Miller, 2016) as well as the negative consequences (Kovach et al., 2016) of human-mediated hybridization as a means of assisting populations adapt to climate change. On the one hand, introgression can provide the genetic variation needed to evolutionary rescue a population declining to environmental stress (Gonzalez et al., 2013). However, introgression can also have negative impacts due to outbreeding depression and swamping of local genetic diversity, facilitating local extinction and homogenization of biodiversity (Kovach et al., 2016). The negative eco-evolutionary consequences may be exacerbated during range expansion if spatial sorting of hybrids disseminates mal-adapted alleles (Lowe et al., 2015; Canestrelli et al., 2016). While this debate is ongoing, natural experiments of hybridization and gene flow are likely taking place as species shift their range in response to ongoing climate change.

Range expansion and infectious diseases

Ongoing climate change is affecting co-evolved life cycle and transmission dynamics between symbionts and their hosts, altering disease systems in humans, crops, and wildlife (reviewed by Altizer et al., 2013). In addition to the expansion of *P. leucopus*, the key reservoir host of Lyme disease, the tick vector this disease (*Ixodes scapularis*) is also undergoing a northward expansion

(Leighton et al., 2012; Khatchikian et al., 2015), as well as the tick vector (*Amblyomma americanum*) of human ehrlichiosis (Monzón et al., 2016), putting more human populations at risk of infection. Mixing of interspecific genomes is likely to alter co-evolved host-symbiont relationships which may, as seen with fungi (Stukenbrock, 2016), facilitate the emergence of novel pathogens or more virulent genotypes. The introgression between *P. maniculatus* and *P. leucopus* could thus have consequences relevant to evolutionary epidemiology (Grenfell et al., 2004) and the emergence of Lyme disease in Quebec.

It is known, for example, that different strains of the causative agent of Lyme disease, *Borrelia burgdorferi*, vary in their ability to infect *P. leucopus* depending on the genetic background of the host (Hanincova et al., 2008). In addition, different *B. burgdorferi* lineages are thought differ in their pathogenic properties in humans (Hannicova et al., 2013). Although a few studies have shown that *P. maniculatus* can be a competent host of *B. burgdorferi* (Rand et al., 1993; Peavey and Lane, 1995), comprehensive studies investigating competence in sympatric *P. maniculatus* and *P. leucopus* populations are lacking (Barbour, 2016), and the competence of *P. maniculatus* and *P. leucopus* hybrids is unknown

If *P. maniculatus* is a less competent host than *P. leucopus* in sympatric populations, or if it is associated with a less virulent strain, introgression of *P. maniculatus* alleles with quantitative effects on competence may affect the propensity with which *P. leucopus* acquires, carries, and transmits a given *B. burgdorferi* genotype, and thus influence Lyme disease risk in humans. *P. maniculatus* may therefore ‘buffer’ the risk of Lyme disease in areas where *P. leucopus* is colonizing. Further work on the competence of sympatric *P. leucopus* and *P. maniculatus*, as well as that of their hybrids, will help in understanding the eco-evolutionary consequences of secondary contact between these species and the potential effect on emerging diseases.

CONCLUSISION

This thesis showed that genetic variation in *P. leucopus* from southern Quebec is associated with geography and shaped by post-glacial and recent range expansion. Our results of genome-wide population structure and genetic diversity indicate Quebec was colonized by two putative glacial lineages, one of which was further isolated during post-glacial expansion by the Richelieu River. Consistent with the predicted consequences of range expansion (Currat et al., 2008; Excoffier and Ray, 2008; Excoffier et al., 2009), we found a northern-most population showing low nucleotide diversity, divergent allele frequencies, the highest number of private alleles, the lowest effective population size and heterozygosity levels that suggest introgression from an unsampled local population. Analysis of ancestry identified putative *P. maniculatus* and *P. leucopus* hybrids. Testing the bifurcating model genealogical evolution confirmed a reticulate phylogeny and past gene flow between *P. maniculatus* and *P. leucopus* in Quebec. This result adds to the rapidly increasing list of natural hybrids discovered between species-pairs previously thought to be isolated, supporting the idea that reproductive isolation between closely related species can vary based on genotype and demographic conditions. More generally, this thesis supports the view that species boundaries not only change over time but also vary across space.

REFERENCES

Abbott RJ, Barton NH, & Good JM (2016) Genomics of hybridization and its evolutionary consequences. *Molecular ecology*, **25**: 2325-2332.

Alexander DLange K (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, **12**: 246-252

Alexander DH, Novembre J, Lange, K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, **19:** 1655-1664.

Altizer S, Ostfeld RS, Johnson PT, Kutz S, Harvell CD (2013). Climate change and infectious diseases: from evidence to a predictive framework. *Science*, **341**: 514-519.

Amorim CE, Hofer T, Ray N, Foll M, Ruiz-Linares A, Excoffier L (2017). Long-distance dispersal suppresses introgression of local alleles during range expansions. *Heredity*, **118**: 135-142.

Anderson E (1948) Hybridization of the habitat. *Evolution*, **2**: 1-9.

Anderson E (1953) Introgressive hybridization. *Biological Reviews*, **28**: 280-307.

Anderson E, Stebbins Jr GL (1954) Hybridization as an evolutionary stimulus. *Evolution*, **8:** 378-388.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe, PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**: 81-92.

Araripe LO, Tao Y, Lemos B (2016). Interspecific Y chromosome variation is sufficient to rescue hybrid male sterility and is influenced by the grandparental origin of the chromosomes. *Heredity*, **116**: 516-522.

Arnold ML (1992) Natural hybridization as an evolutionary process. *Annual review of Ecology and Systematics*, **23**: 237-261.

Arnold ML (2004) Transfer and origin of adaptations through natural hybridization: were Anderson and Stebbins right? *The Plant Cell*, **16**: 562-70.

Baker HG, & Stebbins GL (1965) Genetics of colonizing species, proceedings. In *International Union of Biological Sciences Symposia on General Biology 1964: Asilomar, Calif.* Academic Press.

Barbour AG (2016) Infection resistance and tolerance in Peromyscus spp., natural reservoirs of microbes that are virulent for humans. In Seminars in Cell & Developmental Biology. Academic Press.

Barrett S (2015) Foundations of invasion genetics: the Baker and Stebbins legacy. *Molecular Ecology*, **24**: 1927-1941.

Besold J, Schmitt T, Tammaru T, Cassel-Lundhagen A (2008) Strong genetic impoverishment from the centre of distribution in southern Europe to peripheral Baltic and isolated Scandinavian populations of the pearly heath butterfly. *Journal of Biogeography*, **35**: 2090-2101.

Biek R, Henderson J, Waller L, Rupprecht C, Real L (2007) A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. *Proceedings of the National Academy of Sciences*, **104:** 7993-7998.

Birch LC and Vogt WG (1970) Plasticity of taxonomic characters of the Queensland fruit flies *Dacus tryoni* and *Dacus neohumeralis* (*Tephritidae*). *Evolution*, **24**: 320-343.

Bradshaw WN (1968) Progeny from experimental mating tests with mice of the *Peromyscus leucopus* group. *Journal of mammalogy*, **49**: 475-480.

Brennan, A. C., Woodward, G., Seehausen, O., Muñoz-Fuentes, V., Moritz, C., Guelmami, A. et al. (2015) Hybridization due to changing species distributions: adding problems or solutions to conservation of biodiversity during global change? *Evolutionary Ecology Research*, **16**: 475-491.

Burton O, Phillips B, Travis J (2010) Trade-offs and the evolution of life-histories during range expansion. *Ecology Letters*, **13:** 1210-1220.

Cahill AE, Levinton JS (2016) Genetic differentiation and reduced genetic diversity at the northern range edge of two species with different dispersal modes. *Molecular ecology*, **25**: 515-526.

Cahill J, Stirling I, Kistler L et al. (2015) Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Molecular Ecology*, **24**: 1205-1217.

Canestrelli D, Porretta D, Lowe W et al. (2016) The Tangled Evolutionary Legacies of Range Expansion and Hybridization. *Trends in Ecology & Evolution,* **31**: 677-688.

Catchen J, Hohenlohe P, Bassham S, Amores A, Cresko W (2013a) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**: 3124-3140.

Catchen J, Bassham S, Wilson T, Currey M, O'Brien C, Yeates Q, Cresko WA (2013b). The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Molecular Ecology*, **22**: 2864-2883.

Chen I, Hill J, Ohlemuller R, Roy D, Thomas C (2011) Rapid Range Shifts of Species Associated with High Levels of Climate Warming. *Science*, **333**: 1024-1026.

Chunco AJ (2014) Hybridization in a warmer world. *Ecology and evolution*, **4:** 2019-31.

Crispo E, Moore JS, Lee-Yaw JA, Gray SM, Haller BC (2011). Broken barriers: Human-induced changes to gene flow and introgression in animals. *BioEssays*, **33**: 508-518.

Crow JF and Kimura M (1970). An introduction to population genetics theory. *Blackburn Press*

Currat M, Ruedi M, Petit RJ, Excoffier L (2008). The hidden side of invasions: massive introgression by local genes. *Evolution*, **62**: 1908-1920.

Cutter A (2012) The polymorphic prelude to Bateson–Dobzhansky–Muller incompatibilities. *Trends in Ecology & Evolution*, **27:** 209-218.

Dai Q, Zhan X, Lu B, Fu J, Wang Q, Qi D (2014) Spatial Genetic Structure Patterns of Phenotype-Limited and Boundary-Limited Expanding Populations: A Simulation Study. *PloS one*, **9:** e85778.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics,* **12**: 499-510.

Dawson WD, Mintz J, Maddock MB Lewin AR (1972) On hybridization between deermice (*Peromyscus maniculatus*) and woodmice (*P. leucopus*). ASB Bulletin 19, 64 (Abstract).

Dice LR (1933) Fertility relationships between some of the species and subspecies of mice in the genus Peromyscus. *Journal of Mammalogy*, **14**: 298-305.

Dice LR (1937) *Fertility relations in the Peromyscus leucopus group of mice*. Laboratory, University of Michigan.

Dice LR (1940) Speciation in *Peromyscus*. *American Naturalist*, **74:** 289-98.

Dice LR (1968) Speciation. *Biology of Peromyscus*, pg. 75-97.

Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources,* **14**: 209-214.

Doty RL (1972). Odor preferences of female *Peromyscus maniculatus bairdi* for male mouse odors of *P*. m. *bairdi* and *P*. *leucopus noveboracensis* as a function of estrous state. *Journal of Comp Physio Psych*, **81**: 191-197

Doty RL (1973). Reactions of deer mice (*Peromyscus maniculatus*) and white-footed mice (*Peromyscus leucopus*) to homospecific and heterospecific urine odors. *Journal of Comp Physio Psych*, **84**: 296-303

Dowling TE and Secor CL (1997). The role of hybridization and introgression in the diversification of animals. *Annual review of Ecology and Systematics*, **1**: 593-619.

Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004) Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences of the United States of America*, **101**: 975-979.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell S E (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*, **6**: e19379.

Elton CS (1958) The ecology of invasions by plants and animals. *Methuen, London*, *18*.

Estoup A, Beaumont M, Sennedot F, Moritz Cornuet J (2004) Genetic analysis of complex demographic scenarios: Spatially expanding populations of the cane toad, *Bufo marinus*. *Evolution*, **58**: 2021-2036.

Ewen JG (2012). Reintroduction biology: integrating science and management. *John Wiley & Sons*

Excoffier L and Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution*, **23**: 347-351.

Excoffier, L, Foll, M and Petit, RJ (2009) Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, **40**: 481-501.

Falush D L, van Dorp DL (2016) A tutorial on how (not) to over-interpret STRUCTURE/ADMIXTURE bar plots. *bioRxiv* doi:10.1101/066431

Feder J, Egan S, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**: 342-350.

Fiset J, Tessier N, Millien V, Lapointe F (2015) Phylogeographic Structure of the White-Footed Mouse and the Deer Mouse, Two Lyme Disease Reservoir Hosts in Québec. *PLOS ONE*, **10**: e0144112.

Frichot E, François O (2015) LEA: An R package for landscape and ecological association studies. *Methods Ecol Evol*, **6:** 925-929.

Fu Q, Hajdinjak M, Moldovan O, Constantin S, Mallick S, Skoglund P et al. (2015) An early modern human from Romania with a recent Neanderthal ancestor. *Nature*, **524**: 216-219.

Garnier J, Lewis MA (2016) Expansion under climate change: the genetic consequences. *Bulletin of mathematical biology*, **78:** 2165-85.

Garroway C, Bowman J, Holloway G, Malcom J, Wilson P (2011) The genetic signature of rapid range expansion by flying squirrels in response to contemporary climate warming. *Global Change Biology*, **17**: 1760-1769.

Gompert Z, Buerkle C (2016) What, if anything, are hybrids: enduring truths and challenges associated with population structure and gene flow. *Evol Appl*, **9**: 909-923.

Gompert Z, Parchman TL, Buerkle CA (2012) Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **36**: 439-50.

Gompert Z, Lucas L, Buerkle C et al. (2014) Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology*, **23**: 4555-4573.

Gonzalez A, Ronce O, Ferriere R, Hochberg ME (2013). Evolutionary rescue: an emerging focus at the intersection between ecology and evolution. *Proc R Soc B,* **368:** 20120404

Grant PR (1976) An 11-year study of small mammal populations at Mont St. Hilaire, Quebec. *Canadian Journal of Zoology*, **54**: 2156-2173.

Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, Holmes EC (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, **303**: 327-32.

Hagen SB, Kopatz A, Aspi J, Kojola I, Eiken HG (2015) Evidence of rapid change in genetic structure and diversity during range expansion in a recovering large terrestrial carnivore. *Proceedings of the Royal Society of London B: Biological Sciences,* **282**: 20150092.

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.

Hallatschek O, Hersen P, Ramanathan S, Nelson D (2007) Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*, **104:** 19926-19930.

Hamilton JA, Miller JM (2016) Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conservation Biology*, **30:** 33-41.

Handley L, Manica A, Goudet J, Balloux F (2007) Going the distance: human population genetics in a clinal world. *Trends in Genetics*, **23**: 432-439.

Hanincova K, Ogden N, Diuk-Wasser M et al. (2008) Fitness Variation of *Borrelia burgdorferi* Sensu Stricto Strains in Mice. *Applied and Environmental Microbiology*, **74:** 153-157.

Hanincova K, Mukherjee P, Ogden N et al. (2013) Multilocus Sequence Typing of *Borrelia burgdorferi* Suggests Existence of Lineages with Differential Pathogenic Properties in Humans. *PLoS ONE*, **8**: e73066.

Heckel G, Burri R, Fink S, Desmet J, Excoffier L (2005) Genetic structure and colonization processes in European populations of the common vole, *Microtus arvalis*. *Evolution*, **59**: 2231-2242.

Hedrick P (2013) Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, **22**: 4606-4618.

Heiser CB (1949) Natural hybridization with particular reference to introgression. *The Botanical Review*, **15**: 645-87.

Henn B, Botigué L, Peischl S et al. (2015) Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. *Proceedings of the National Academy of Sciences*, **113**: E440-E449.

Hewitt G (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**: 247-276.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**: 907-13.

Hewitt G (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359:** 183-195.

Hibbard CW (1968) Paleontology In: King JA, editor. Biology of *Peromyscus* (Rodentia), Special Publication no. 2, *American Society of Mammalogists,* pp. 6-26

Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter B, Vinckenbosch N et al. (2014) Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature,* **512**: 194-197.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics,* **24**: 1403-1405

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics,* **11**,**94:** 1-15

Jombart T, Collins C (2015) A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0. 0.

Janousek V, Wang L, Luzynski K, Dufkova P, Vyskocillova MM, Nachman MW, Munclinger P, Macholan M, Pialek J and Tucker PK (2012) Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between *Mus musculus musculus* and *M. m. domesticus*. *Molecular Ecology*, **21**: 3032-3047.

Kelly B, Whiteley A, Tallmon D (2010) The Arctic melting pot. Nature, **468**: 891-891.

Khatchikian CE, Prusinski MA, Stone M, Backenson PB, Wang IN, Foley E, Seifert SN, Levy MZ, Brisson D (2015). Recent and rapid population growth and range expansion of the Lyme disease tick vector, *Ixodes scapularis,* in North America. *Evolution,* **69**:1678-1689.

King JA (1968) Biology of *Peromyscus* (Rodentia). *American Society of Mammalogists*

Klein HG (1960) Ecological relationships of *Peromyscus leucopus noveboracensis* and *P. maniculatus gracili*s in central New York. *Ecological Monographs*, **30:** 387-407.

Klopfstein S (2005) The Fate of Mutations Surfing on the Wave of a Range Expansion. *Molecular Biology and Evolution*, **23:** 482-490.

Kovach RP, Luikart G, Lowe WH, Boyer MC, Muhlfeld CC (2016) Risk and efficacy of human-enabled interspecific hybridization for climate-change adaptation: response to Hamilton and Miller (2016). *Conservation biology,* **30:** 428-30.

Kozlowska J, Ahmad A, Jahesh E, Cutter A (2011) Genetic variation for postzygotic reproductive isolation between *Caenorhabditis briggsae* and *Caenorhabditis* SP. 9. *Evolution*, **66**: 1180-1195.

LaBar T, Adami C (2016). Evolution of drift robustness in small populations of digital organisms. *bioRxiv*. doi: https://doi.org/10.1101/071894

Langmead B and Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**: 357-359.

Lee C (2002) Evolutionary genetics of invasive species. *Trends in Ecology & Evolution,* **17**: 386-391.

Ledevin R and Millien V (2013) Congruent morphological and genetic differentiation as a signature of range expansion in a fragmented landscape. *Ecol Evol*, **3**: 4172-4182.

Leighton PA, Koffi JK, Pelcat Y, Lindsay LR, Ogden NH (2012). Predicting the speed of tick invasion: an empirical model of range expansion for the Lyme disease vector *Ixodes scapularis* in Canada. *Journal of Applied Ecology*. **49**:457-464.

Leo SS and Millien V (2016). Microsatellite markers reveal low frequency of natural hybridization between the white-footed mouse (*Peromyscus leucopus*) and deer mouse (*Peromyscus maniculatus*) in southern Quebec, Canada. *Genome*

Lewontin R and Birch L (1966) Hybridization as a Source of Variation for Adaptation to New Environments. *Evolution,* **20**: 315-336

Li Y, Willer C, Sanna S, Abecasis G (2009). Genotype imputation. *Ann Rev Genomics and human genetics,* **10**: 387-406.

Lindtke D, Buerkle C (2015) The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. *Evolution*, **69:** 1987-2004.

LoGiudice K, Ostfeld R, Schmidt Keesing F (2003) The ecology of infectious disease: Effects of host diversity and community composition on Lyme disease risk. *Proceedings of the National Academy of Sciences,* **100**: 567-571.

Lowe W, Muhlfeld C, Allendorf F (2015) Spatial sorting promotes the spread of maladaptive hybridization. *Trends in Ecology & Evolution,* **30**: 456-462.

Lubina J, Levin S (1988) The Spread of a Reinvading Species: Range Expansion in the California Sea Otter. *The American Naturalist,* **131:** 526-543.

MacNeil A, Ksiazek TG and Rollin PE (2011) Hantavirus pulmonary syndrome, United States, 1993-2009. *Emerging infectious diseases*, **17**: 1195-1202.

Maddock MB, Dawson WD (1974) Artificial insemination of deermice (*Peromyscus maniculatus*) with sperm from other rodent species. *Development*, **3**: 621-634

Mallet J (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, **20**: 229-237.

Mallet J, Beltrán M, Neukirchen W, Linares M (2007) Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evol Biol*, **7, 28:** 1-16

Mandeville E, Parchman T, McDonald D, Buerkle C (2015) Highly variable reproductive isolation among pairs of Catostomus species. *Molecular Ecology,* **24:** 1856-1872.

Marrotte R, Gonzalez A, Millien V (2014) Landscape resistance and habitat combine to provide an optimal model of genetic structure and connectivity at the range margin of a small mammal. *Molecular Ecology,* **23:** 3983-3998.

Mastrantonio V, Porretta D, Urbanelli S, Crasta G, Nascetti G (2016) Dynamics of mtDNA introgression during species range expansion: insights from an experimental longitudinal study. *Sci. Rep.*, **6:** 30355.

Mayr E (1942) Systematics and the origin of species, from the viewpoint of a zoologist. *Harvard University Press.*

Mayr E (1963) Animal species and evolution (Vol. 797). Cambridge, Massachusetts: *Belknap Press of Harvard University Press.*

McCarley WH (1964) Ethological isolation in the cenospecies *Peromyscus leucopus*. *Evolution*, **18:** 331-332.

McCarley WH (1954) Natural hybridization in the *Peromyscus leucopus* species group of mice. *Evolution*, **8**: 314-323.

Miao B, Wang Z, Li Y (2016) Genomic Analysis Reveals Hypoxia Adaptation in the Tibetan Mastiff by Introgression of the Grey Wolf from the Tibetan Plateau. *Molecular Biology and Evolution*. doi: 10.1093/molbev/msw274

Mondal M, Casals F, Xu T, Dall'Olio GM, Pybus M, Netea MG (2016) Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. *Nature Genetics,* **48**: 1066-1070

Money D, Gardner K, Migicovsky Z, Schwaninger H, Zhong GY, Myles S (2015). LinkImpute: Fast and accurate genotype imputation for nonmodel organisms. *G3: Genes*| *Genomes*| *Genetics,* **5**: 2383-2390.

Monzón J, Atkinson E, Henn Benach J (2016) Population and Evolutionary Genomics *of Amblyomma americanum*, an Expanding Arthropod Disease Vector. *Genome Biol Evol*, **8:** 1351-1360.

Munshi-South J, Zolnik C, Harris S (2016) Population genomics of the Anthropocene: urbanization is negatively associated with genome-wide variation in white-footed mouse populations. *Evol Appl,* **9:** 546-564.

Myers P, Lundrigan BL, Hoffman SM, Haraminac AP, Seto SH (2009) Climate-induced changes in the small mammal communities of the Northern Great Lakes region. *Global Change Biology*, **15:** 1434-1454.

Nam K, Jeong H, Nam J (2016) Pseudo-Reference-Based Assembly of Vertebrate Transcriptomes. *Genes,* **7, 10**: 1-14

Novembre J, Stephens M (2008). Interpreting principal component analyses of spatial population genetic variation. *Nat Genet*, **40**: 646-649.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, Bustamante CD (2008). Genes mirror geography within Europe. *Nature*, **456**: 98-101.

Nullmeier J and Hallatschek O (2013) The coalescent in boundary-limited range expansions. *Evolution*, **67**: 1307-1320.

Ogden N, St-Onge L, Barker I et al. (2008) Risk maps for range expansion of the Lyme disease vector, *Ixodes scapularis*, in Canada now and with climate change. *International Journal of Health Geographics*, **7, 24**: 1-15

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**: 37-42.

Peavey C, Lane R (1995) Transmission of *Borrelia Burgdorferi* by Ixodes Pacificus Nymphs and Reservoir Competence of Deer Mice (*Peromyscus maniculatus*) Infected by Tick-Bite. *The Journal of Parasitology*, **81**:175-178

Peel D, Waples RS, Macbeth GM, Do C, Ovenden JR (2013). Accounting for missing data in the estimation of contemporary genetic effective population size (Ne). *Molecular ecology resources*, **13**: 243-253.

Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L (2013) On the accumulation of deleterious mutations during range expansions. *Molecular Ecology*, **22**: 5972-5982.

Peischl, S., Kirkpatrick, M., & Excoffier, L. (2015) Expansion load and the evolutionary dynamics of a species range. *The American Naturalist*, **4:** E81-E93

Peischl S, Dupanloup I, Foucal A, et al. (2016) Relaxed selection during a recent human expansion. *bioRxiv*, 064691.

Petit RJ, Excoffier L. Gene flow and species delimitation (2009). *Trends in Ecology & evolution*, **24**: 386-393.

Pfennig KS, Kelly AL, Pierce AA (2016) Hybridization as a facilitator of species range expansion. *In Proc. R. Soc. B.*, **283:** 20161329.

Phillips BL, Brown GP, Webb JK, Shine R (2006) Invasion and the evolution of speed in toads. *Nature*, **439:** 803.

Phillips B, Kelehear C, Pizzatto L et al. (2010) Parasites and pathogens lag behind their host during periods of host range advance. *Ecology*, **91:** 872-881.

Pierce S, Vogt F (1993) Winter Acclimatization in *Peromyscus maniculatus gracilis, P. leucopus noveboracensis,* and *P. l. leucopus*. *Journal of Mammalogy*, **74:** 665-677.

Pimentel D, Zuniga R, Morrison D (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological economics,* **52:** 273-288.

Pluess AR (2011) Pursuing glacier retreat: genetic structure of a rapidly expanding *Larix decidua* population. *Molecular Ecology*, **20:** 473-485.

Poon A and Otto S (2000) Compensating for our load of mutations: freesing the meltdown of small populations. *Evolution*, **54**: 1467-1479

Prager EM and Wilson AC (1975) Slow evolutionary loss of the potential for interspecific hybridization in birds: a manifestation of slow regulatory evolution. *Proceedings of the National Academy of Sciences*, **72**: 200-204.

Price T, Hooper D, Buchanan C et al. (2014) Niche filling slows the diversification of Himalayan songbirds. *Nature*, **509:** 222-225.

Puechmaille SJ (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Mol Ecol resources*, **16**: 608-627.

Purcell S and Chang C (2015). PLINK 1.9. URL https://www.cog-genomics.org/plink2

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ, Sham PC (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The Am J Hum Gene*, **81**: 559-575.

R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rand P, Lacombe E, Smith R et al. (1993) Competence of *Peromyscus maniculatus* (Rodentia: Cricetidae) as a Reservoir Host for *Borrelia burgdorferi* (Spirochaetares: Spirochaetaceae) in the Wild. *Journal of Medical Entomology*, **30:** 614-618.

Rhymer JM, Simberloff D (1996). Extinction by hybridization and introgression. *Ann Rev Ecol Syst*, **27**: 83-109.

Reich D, Wayne R, Goldstein D (1999) Genetic evidence for a recent origin by hybridization of red wolves. *Molecular Ecology*, **8**: 139-144.

Riley HP. A character analysis of colonies of Iris fulva, Iris hexagona var. giganticaerulea and natural hybrids (1938) *American Journal of Botany*, **25**: 727-38.

Rogic A, Tessier N, Legendre P, Lapointe F, 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.

Rosenzweig ML (1995) Species diversity in space and time. Cambridge University Press.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*, **145**:1219-1228.

Rowe KC, Heske EJ, Paige KN (2006) Comparative phylogeography of eastern chipmunks and white-footed mice in relation to the individualistic nature of species. *Molecular Ecology*, **15**: 4003-4020.

Roy-Dufresne E, Logan T, Simon J, Chmura G, Millien V (2013) Poleward Expansion of the White-Footed Mouse (*Peromyscus leucopus*) under Climate Change: Implications for the Spread of Lyme Disease. *PLoS ONE*, **8**: e80724.

Sambrook, J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Chapter 6 - Preparation and Analysis of Eukaryotic Genomic DNA. *Cold Spring Harbor Laboratory Press*, Cold Spring Harbor, NY

Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, **19**: 198-207.

Seehausen O, Butlin RK, Keller I et al. (2014) Genomics and the origin of species. *Nature Reviews Genetics*, **15:** 176-192.

Senerchia N, Felber F, North B, Sarr A, Guadagnuolo R, Parisod C (2016) Differential introgression and reorganization of retrotransposons in hybrid zones between wild wheats. *Molecular ecology*, **11:** 2518-2528

Shafer A, Peart CR, Tusso S, Maayan I, Brelsford A, Wheat CW, Wolf JB (2016) Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, DOI: 10.1111/2041-210X.12700

Shine R, Brown G, Phillips B (2011) An evolutionary process that assembles phenotypes through space rather than through time. *Proceedings of the National Academy of Sciences*, **108**: 5708-5711.

Short RV (1997) An introduction to mammalian interspecific hybrids. *Journal of Heredity*, **88**: 355-357.

Slatkin M and Barton NH (1989). A comparison of three indirect methods for estimating average levels of gene flow. *Evolution*, **43**: 1349-1368.

Slatkin M and Excoffier L (2012) Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift. *Genetics*, **191**: 171-181.

Sobel JM, Chen GF, Watt LR, Schemske DW (2010). The biology of speciation. *Evolution*, **64**: 295-315.

Song Y, Endepols S, Klemann N et al. (2011) Adaptive Introgression of Anticoagulant Rodent Poison Resistance by Hybridization between Old World Mice. *Current Biology*, **21:** 1296-1301.

Statistique Canada. Recensements du Canada (2014). URL: http:// www.stat.gouv.qc.ca/

Staubach F, Lorenc A, Messer P et al. (2012) Genome Patterns of Selection and Introgression of Haplotypes in Natural Populations of the House Mouse (*Mus musculus*). *PLoS Genetics*, **8**: e1002891.

Stukenbrock EH (2016) Hybridization speeds up the emergence and evolution of a new pathogen species. *Nature Genetics,* **48**:113.

Swaegers J, Mergeay J, Van Geystelen A et al. (2015) Neutral and adaptive genomic signatures of rapid poleward range expansion. *Molecular Ecology*, **24:** 6163-6176.

Taylor S, Curry R, White T, Ferretti V, Lovette I (2014) Spatiotemporally consistent genomic signatures of reproductive isolation in a moving hybrid zone. *Evolution*, **68:** 3066-3081.

Taylor SA, Larson EL, Harrison RG (2015) Hybrid zones: windows on climate change. *Trends in ecology & evolution,* **30:** 398-406.

Travis JM and Dytham C (2002) Dispersal evolution during invasions. *Evolutionary Ecology Research,* **4:** 1119-1129.

Travis J, Munkemuller T, Burton O et al. (2007) Deleterious Mutations Can Surf to High Densities on the Wave Front of an Expanding Population. *Molecular Biology and Evolution*, **24**: 2334-2343.

Travis J, Mustin K, Benton T, Dytham C (2009) Accelerating invasion rates result from the evolution of density-dependent dispersal. *Journal of Theoretical Biology*, **259**: 151-158.

Vessey S H, Vessey KB (2007) Linking behavior, life history and food supply with the population dynamics of white-footed mice (*Peromyscus leucopus*). *Integrative Zoology*, **2**: 123-130.

von Holdt B, Kays R, Pollinger J, Wayne R (2016) Admixture mapping identifies introgressed genomic regions in North American canids. *Molecular Ecology*, **25:** 2443-2453.

Wan JJ (2014). Ecological Replacement of *Peromyscus maniculatus gracilis* by *Peromyscus leucopus* in northern Michigan. *PhD dissertation*, University of Michigan.

Wang J (2016). The computer program Structure for assigning individuals to populations: easy to use but easier to misuse. *Mol Ecol Resources,* doi: 10.1111/1755-0998.12650

Waples RS and Do CH (2008). LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular ecology resources,* 8:753-756.

Watts P, Keat S, Thompson D (2010) Patterns of spatial genetic structure and diversity at the onset of a rapid range expansion: colonisation of the UK by the small red-eyed damselfly *Erythromma viridulum*. *Biological Invasions*, **12:** 3887-3903.

White T, Perkins S, Heckel G, Searle J (2013) Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. *Molecular Ecology*, **22**: 2971-2985.

Whitlock MC, Mccauley DE (1999). Indirect measures of gene flow and migration: FST≠ 1/(4Nm+ 1). *Heredity,* **82**:117-125.

Wiegand KM (1935) A taxonomist's experience with hybrids in the wild. *Science*, **81**: 161-166.

Wilson AC, Maxson LR, Sarich VM (1974). Two types of molecular evolution. Evidence from studies of interspecific hybridization. *Proceedings of the National Academy of Sciences*, **71**: 2843-2847.

Wright S (1932) The roles of mutation, inbreeding, crossbreeding, and selection in evolution. **1**: 356-366.

APPENDIX

Supplementary figures

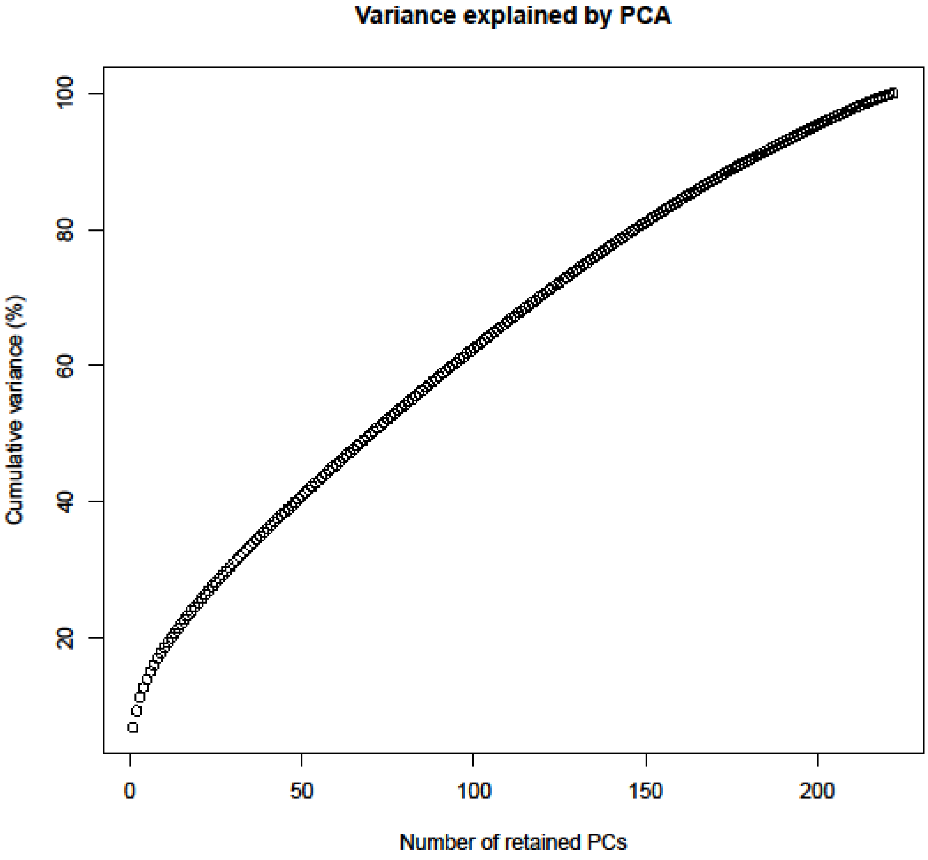


Figure S1 The cummulative genetic variance (38,144 SNPs) of *P. leucopus* explained by PCs. The first 6 PCs were used for DAPC.

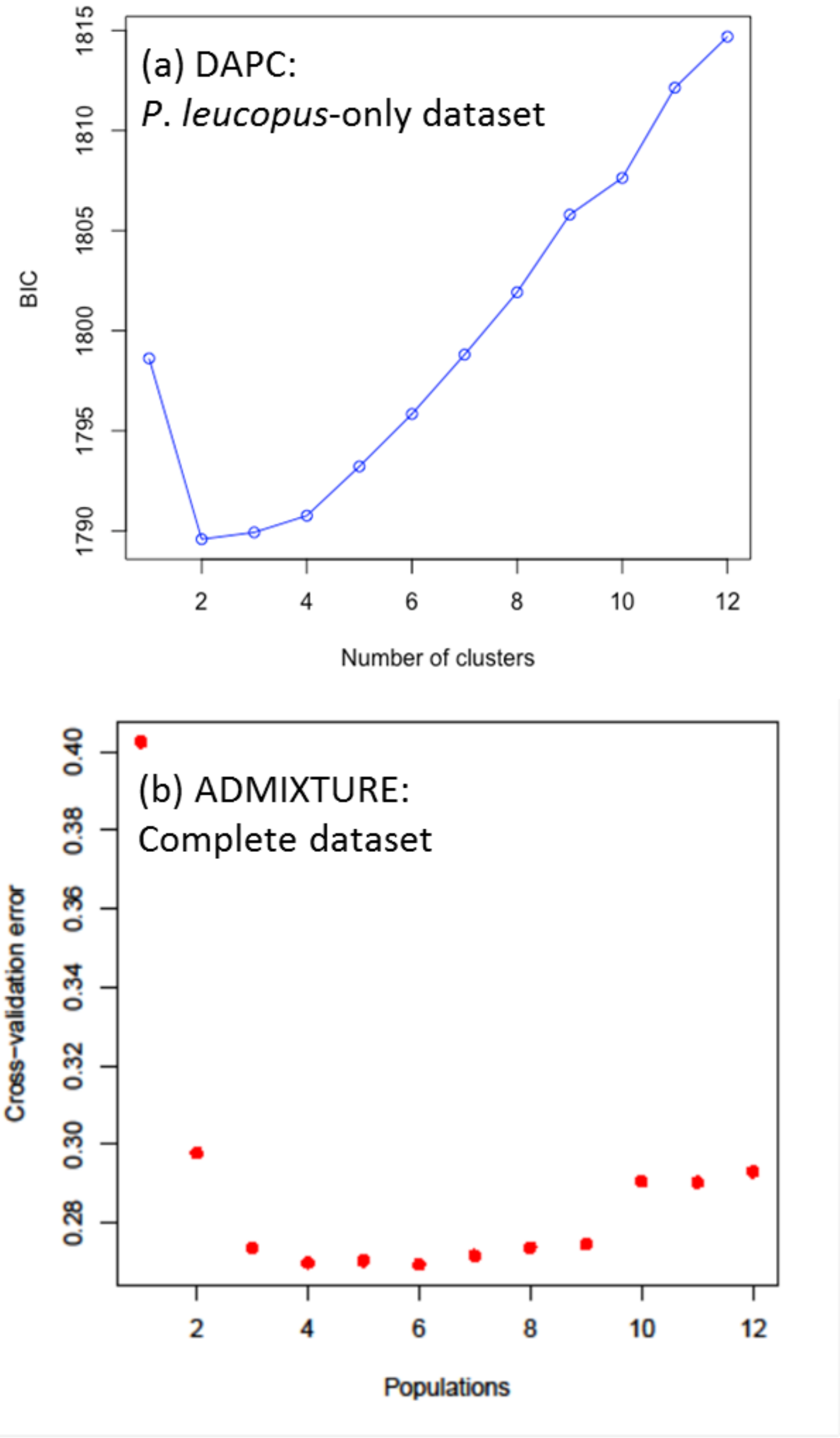


Figure S2 Estimation of genetic clusters (K) by (a) DAPC for the *P. leucopus*-only data and (b) ADMIXTURE for the complete dataset.

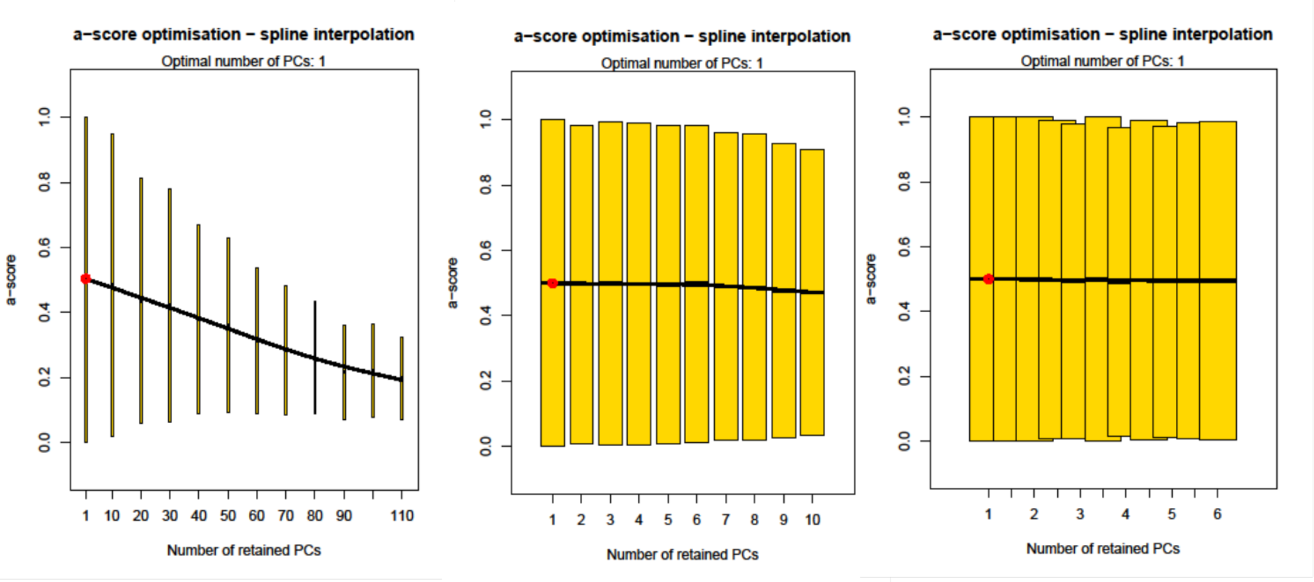


Figure S3 Alpha-score optimization to evaluate the trade-off between discrimination and over-fitting based on a DAPC. We tested (left to right) 110, 10, and 6 PCs. We kept the DAPC on 6 PCS based on the high alpha-scores across the first 6 PCs.

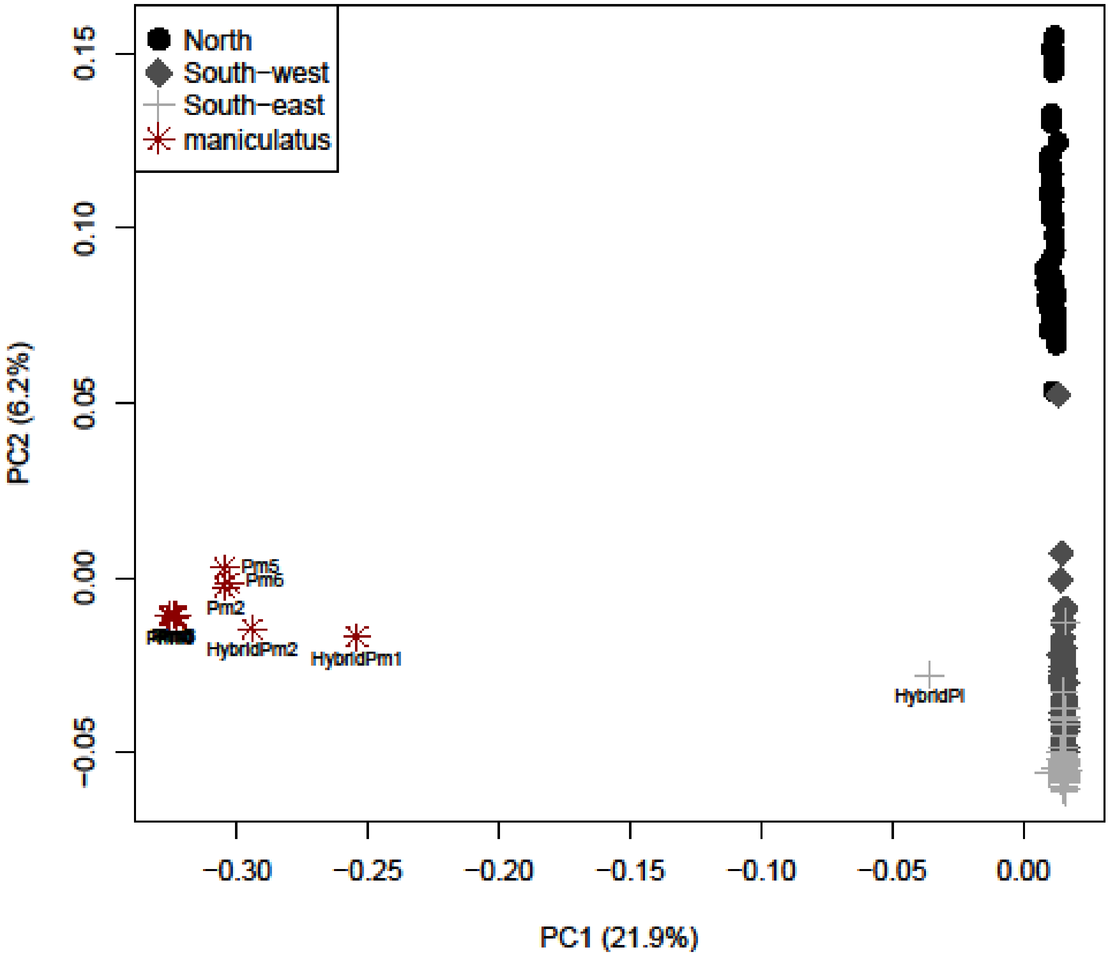
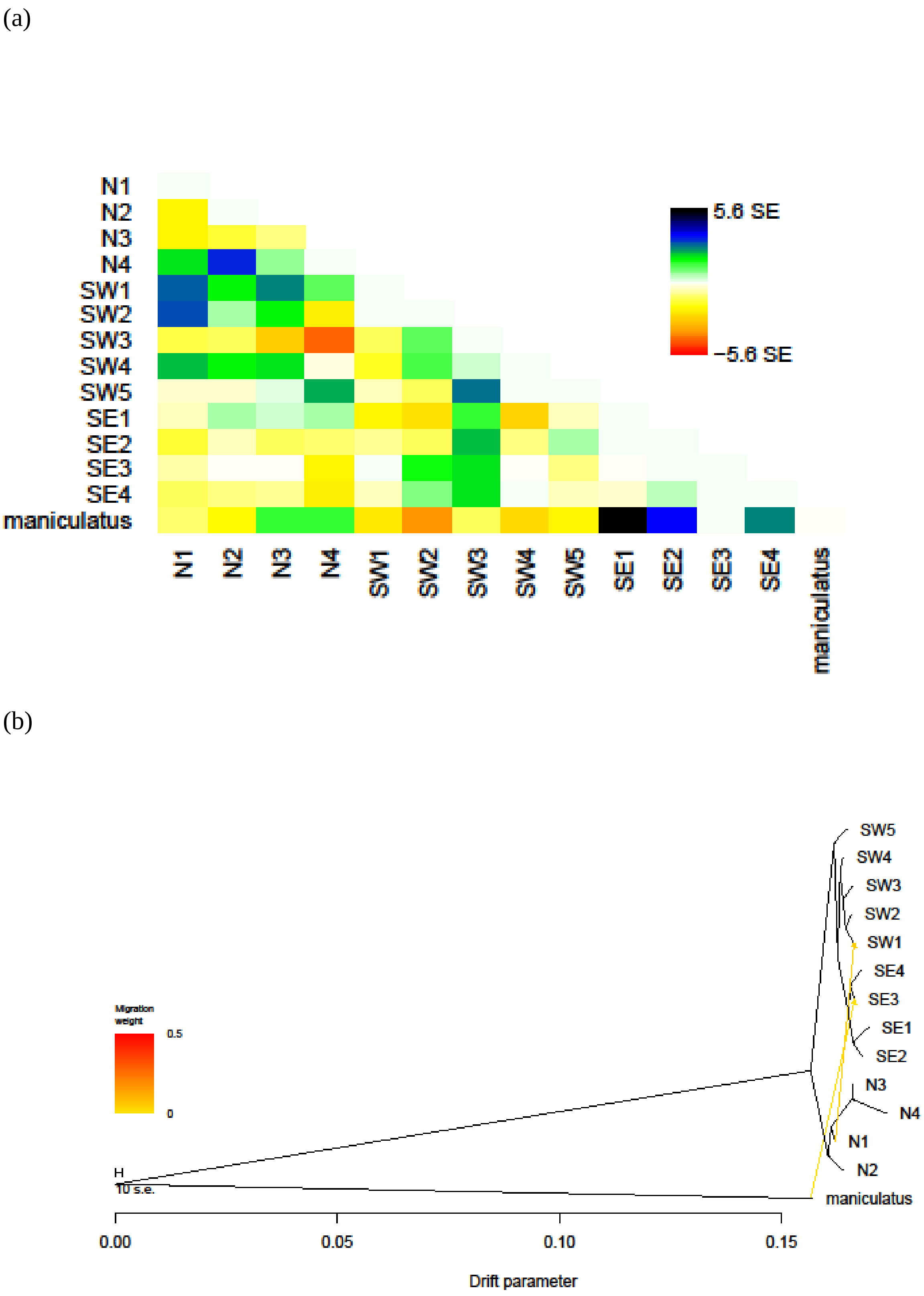
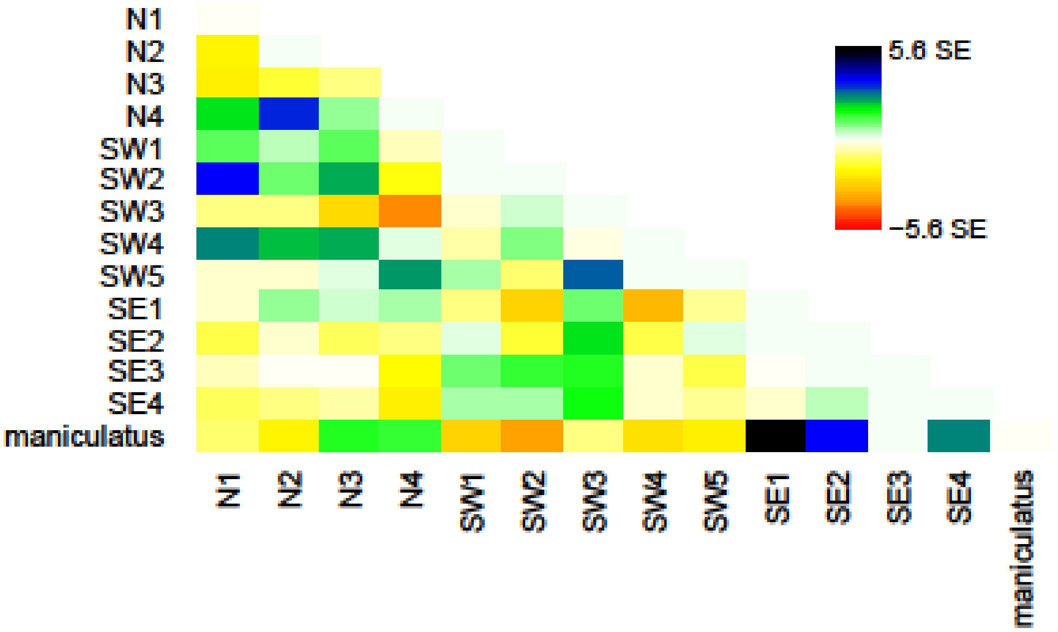


Figure S4 PCA of the complete dataset (33,919 SNPs). The *P. leucopus* hybrid with 15.5% *P. maniculatus* ancestry is labelled as 'HybridPl'. The *P. maniculatus* individuals with 18.8% and 7.5% *P. leucopus* ancestry (K=2) are labelled 'HybridPm1' and 'HybridPm2', respectively.


(a)
N1
N2
N3
N4
SW1
SW2
SW3
SW4
SW5
SE1
SE2
SE3
SE4
maniculatus
5.6 SE
−5.6 SE
(b)
Migration weight
0.5
0
SW5
SW4
SW3
SW2
SW1
SE4
SE3
SE1
SE2
N3
N4
N1
N2
maniculatus
H
10 s.e.
0.00
0.05
0.10
0.15
Drift parameter

(c)


N1
N2
N3
N4
SW1
SW2
SW3
SW4
SW5
SE1
SE2
SE3
SE4
maniculatus
5.6 SE
−5.6 SE

(d)

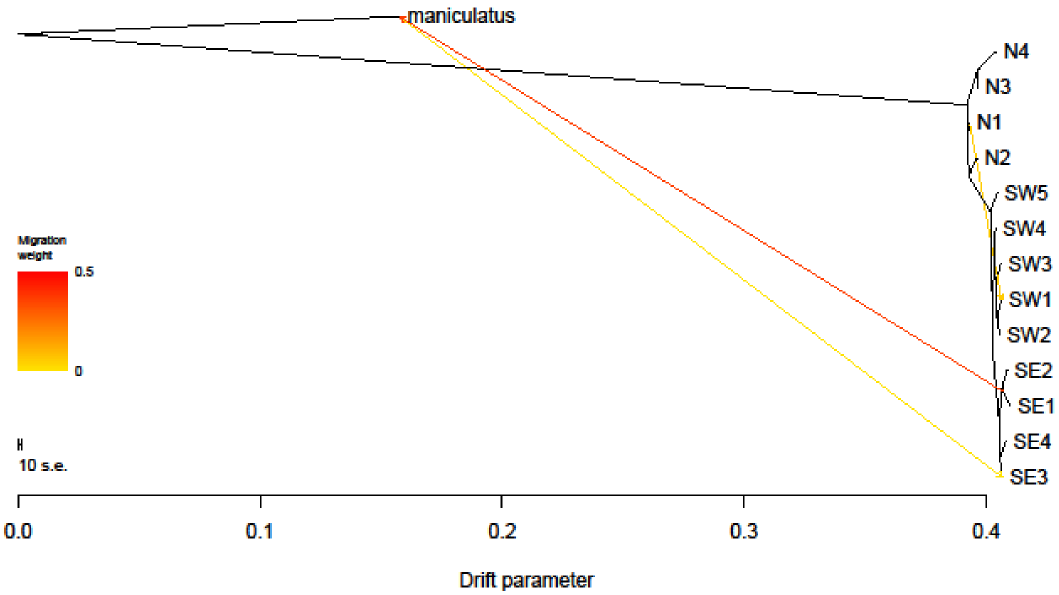


Figure S5 TreeMix results of different demographic models (12,507 SNPs). Plotted are (a) the residuals of a tree with one migration event, a (b) ML tree with two gene flow events and (c) the residuals of its fit, and a (d) tree with three gene flow events. The scale bar represents 10X the average standard error of the entries in the sample covariance matrix (Pickrell and Pritchard, 2012). Residuals are qualified using the colour palette. Positive residuals (green/blue/black) represent population pairs that are more closely related than presented on the tree, and thus candidates for admixture. Negative residuals (yellow/orange/red) represent population pairs that are less closely related than presented on the tree. Direction of gene flow is shown by the arrow. The colour of the arrow denotes the weight of migration, which is proportional to the amount of genetic ancestry contributed by the immigrant.

Supplementary tables

Table S1 Contemporary effective population size (Ne) for 13 *P. leucopus* populations assessed with NeEstimator. Estimates of infinity indicate genetic variation was explained by sampling error alone (i.e. no evidence of variation explained by a finite number of parents).

Population	**Ne $_{\text{(MAF: 5\%)}}$**	**N**	**Harmonic mean sample size**
N1	56	19	12.1
N2	77	19	13.5
N3	∞	11	7
N4	14	20	16.9
SW1	25	18	12.7
SW2	941	17	9.3
SW3	31	19	17.7
SW4	∞	12	8.3
SW5	35	20	18.8
SE1	47	20	18.7
SE2	45	8	7.7
SE3	53	20	15.9
SE4	55	20	16.1

Table S2 Pair-wise F_{ST} estimates for 13 *P. leucopus* populations.

	N1	N2	N3	N4	SW1	SW2	SW3	SW4	SW5	SE1	SE2	SE3	SE4
N1		0.022	0.023	0.037	0.047	0.044	0.047	0.034	0.042	0.055	0.045	0.049	0.053
N2			0.031	0.045	0.050	0.049	0.051	0.038	0.046	0.059	0.048	0.053	0.056
N3				0.021	0.056	0.056	0.055	0.044	0.048	0.061	0.053	0.057	0.060
N4					0.068	0.066	0.069	0.056	0.062	0.076	0.065	0.071	0.074
SW1						0.013	0.015	0.014	0.021	0.032	0.027	0.025	0.028
SW2							0.009	0.011	0.017	0.030	0.025	0.021	0.024
SW3								0.009	0.018	0.030	0.022	0.021	0.024
SW4									0.012	0.024	0.017	0.016	0.019
SW5										0.032	0.022	0.024	0.027
SE1											0.015	0.019	0.023
SE2												0.012	0.017
SE3													0.011