Received: 9 June 2017 | Revised: 8 September 2017 | Accepted: 18 September 2017

DOI: 10.1111/mec.14380

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Inferring the geographic origin of a range expansion: Latitudinal and longitudinal coordinates inferred from genomic data in an ABC framework with the program X-ORIGIN

Qixin He[1] | Joyce R. Prado[2] | Laura Lacey Knowles[3]

[1]Department of Ecology and Evolutionary Biology, University of Chicago, Chicago, IL, USA

[2]Departamento de Ciências Biológicas, Escola Superior de Agricultura 'Luiz de Queiroz', Universidade de São Paulo, Piracicaba, Brazil

[3]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA

Correspondence
Qixin He, Department of Ecology and Evolutionary Biology, University of Chicago, Chicago IL, USA.
Email: heqixin@uchicago.edu

Abstract

Climatic or environmental change is not only driving distributional shifts in species today, but it has also caused distributions to expand and contract in the past. Inferences about the geographic locations of past populations especially regions that served as refugia (i.e., source populations) and migratory routes are a challenging endeavour. Refugial areas may be evidenced from fossil records or regions of temporal stability inferred from ecological niche models. Genomic data offer an alternative and broadly applicable source of information about the locality of refugial areas, especially relative to fossil data, which are either unavailable or incomplete for most species. Here, we present a pipeline we developed (called X-ORIGIN) for statistically inferring the geographic origin of range expansion using a spatially explicit coalescent model and an approximate Bayesian computation testing framework. In addition to assessing the probability of specific latitudinal and longitudinal coordinates of refugial or source populations, such inferences can also be made accounting for the effects of temporal and spatial environmental heterogeneity, which may impact migration routes. We demonstrate X-ORIGIN with an analysis of genomic data collected in the Collared pika that underwent postglacial expansion across Alaska, as well as present an assessment of its accuracy under a known model of expansion to validate the approach.

KEYWORDS

Ψ-statistics, approximate Bayesian computation, niche modeling, range expansion, refugia inference

1 | INTRODUCTION

Population expansions leave signatures in the distribution of population genetic variation across a landscape. This pattern of genetic variation is commonly used for making inferences about the underlying demographic processes. For example, the decreasing pattern of genetic diversity along expansion routes has been used to infer the origin of human migrations (DeGiorgio, Jakobsson, & Rosenberg, 2009; Ramachandran et al., 2005). Similarly, such genetic signatures have been applied to study postglacial expansions in other species, as well as their corresponding geographic refugia during glacial periods of the Pleistocene (reviewed in Hewitt, 2000).

However, this approach comes with an inherent issue. Specifically, genetic diversity patterns (e.g., heterozygosity, F_{ST}) can reflect not only signatures from recent distributional shifts, but also local habitat suitability or long-term geographic isolation (Austerlitz, Jung-Muller, Godelle, & Gouyon, 1997; Ray, Currat, & Excoffier, 2003). Thus, while the isolation-by-distance model applies relatively well to species that have a broad habitat, such as human beings, species with narrower niches tend to track their habitats, displaying a genetic diversity pattern of isolation by barriers or resistance (McRae

& Beier, 2007). Therefore, sole reliance on the gradients of population size/heterozygosity or the principal components without spatial models is inadequate for making accurate inferences about the ancestral source population or directions of expansion (François et al., 2010). Due to the rich, yet confounding information retained in the genetic diversity patterns, most phylogeographic studies infer the location of hypothesized refugia from the data that are independent of the genomic information (reviewed in Knowles, 2009). Ecological niche models (ENMs), for instance, could be applied to infer areas with temporal stability as suitable habitats. In addition, the associated genetic data could then be used to evaluate the hypothesis that such geographic regions would have served as refugial source population (e.g., see Carnaval, Hickerson, Haddad, Rodrigues, & Moritz, 2009; Knowles, Massatti, He, Olson, & Lanier, 2016).

Attempts to address the issue of complex historical processes shaping the current genetic patterns have witnessed the development of spatially explicit demographic models as well as spatial genetic indices. Ray, Currat, Berthier, and Excoffier (2005) systematically tested the likelihood of different geographic locations as human origins by evaluating the goodness of fit of R_{ST} values from different spatial simulations of expansions using the empirical values. Itan, Powell, Beaumont, Burger, and Thomas (2009) estimated the origin of lactase persistent mutations in Europe by fitting empirical frequencies of lactase persistent mutations to those from spatial simulations of the gene expansion along with dairy groups. These pioneer studies demonstrate the potential of using spatially explicit models for estimating migration histories. However, these models do not take temporal changes in habitat suitability into account, which limit their applicability in flora and fauna that underwent expansions largely driven by climatic oscillations.

Spatial genetic indices, on the other hand, are designed to pick up "range expansion"-specific signatures—that is, the directions of gene flow. By analysing the allele frequency clines created by consecutive founder events during the expansion of a population across a landscape, as captured by a directionality index Ψ, Peter & Slatkin (2013) demonstrated how information on the geographic origin and the direction of expansion could be extracted from genomic data through asymmetrical gene flow. That is, regression between pairwise differences of Ψ and geographic distances between populations can be used to *directly* infer the geographic origin of expansion. However, several aspects of this approach limit its utility in practice. For example, this method does not account for the heterogeneity in the underlying landscape during the inference procedure (i.e., assuming a strict isolation-by-distance model). Ψ may also be biased towards nonzero values when local population sizes differ substantially (Peter & Slatkin, 2013). Also, although it is possible to recover a signature of expansion from the magnitude of Ψ, assessing the significance of Ψ-values, and hence, the confidence of the inferred origin, is not straightforward.

Here, we present a pipeline specifically developed for making statistical inferences about the geographic origin of range expansion (called X-ORIGIN) that addresses these aforementioned shortcomings. This pipeline builds upon earlier developments in spatial demographic models (e.g., Ray, Currat, Foll, & Excoffier, 2010) and spatially explicit summary statistics (e.g., Peter & Slatkin, 2013). Specifically, with the X-ORIGIN we couple the Ψ-index (Peter & Slatkin, 2013) with a spatially explicit coalescent model for hypothesis testing in an approximate Bayesian computation (ABC; Beaumont, Zhang, & Balding, 2002) framework. Information based on current and/or historical habitat suitability can be estimated using ENMs and subsequently incorporated into the spatially explicit coalescent model (i.e., a modified application of SPLATCHE2; Ray et al., 2010). In addition, with the ABC framework, the estimation of the geographic origin of range expansion will not be sensitive to the uncertainties in the underlying demographic parameters if a wide range of priors of demographic parameters is specified in spatial simulations. Hereafter, we refer to the geographic origin of range expansion as a parameter, Ω. Together, the significance of expansion and the confidence of a particular geographic location for the ancestral source population are provided by the X-ORIGIN. As such, the pipeline couples information from a series of independent analyses (Figure 1), making X-ORIGIN a useful tool for inferring the geographic origin of ancestral sources with confidence.

It should be noted that there are general procedural parallels with the integrative distributional, demographic, and coalescent (iDDC) approach for model selection, which also involves a series of independent analyses (i.e., estimates of habitat suitability, demographic modelling, and spatially explicit coalescent; He, Edwards, & Knowles, 2013). However, the X-ORIGIN pipeline differs in that (i) it infers a novel model parameter of interest Ω (i.e., the actual latitudinal and longitudinal coordinates), and (ii) it utilizes information from spatial summary statistics, specifically, pairwise population measures of F_{ST} and the directionality index, Ψ (Peter & Slatkin, 2013). As such, X-ORIGIN is an approach that focuses on the estimation of a specific parameter of interest—Ω, whereas the iDDC is an approach for model selection among a set of biologically informed demographic hypotheses, the foci of which vary significantly among studies (e.g., Bemmels, Title, Ortego, & Knowles, 2016; Knowles & Massatti, 2017; Massatti & Knowles, 2016).

Here, we describe the approach and test the accuracy of the X-ORIGIN pipeline in inferring Ω under a known expansion history (i.e., simulated history; see Figure 2). Specifically, we model a history of expansion that involves temporal shifts in the habitat suitability of a landscape (i.e., we validate the approach by implementing a complex model which cannot be accommodated by any other currently existing programs). We also demonstrate the utility of the X-ORIGIN with an analysis of empirical data. Specifically, we analyse the SNP data set collected in the Collared pika (*Ochotona collaris*) (i.e., data from Lanier, Massatti, He, Olson, & Knowles, 2015). The impact of the glaciations is pronounced in small Alaskan mammals (Galbreath, Cook, Eddingsaas, & DeChaine, 2011; Knowles et al., 2016; Lanier et al., 2015). While previous analyses in the Collared pikas also suggested that contemporary environmental factors contribute less to genomic structure than a dynamic history involving the founding of current populations by ancestral source populations (Lanier et al.,

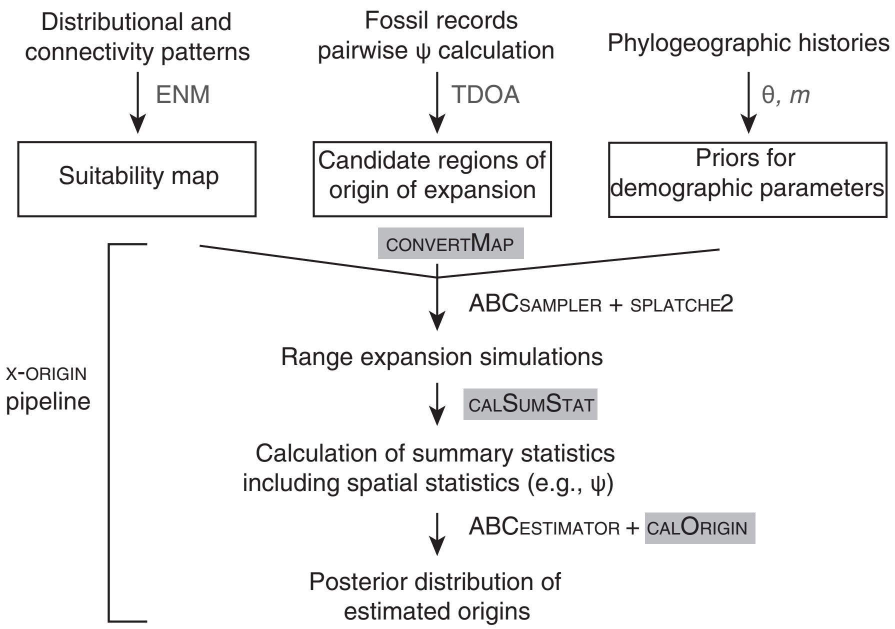
Distributional and connectivity patterns
ENM
Suitability map
Fossil records pairwise ψ calculation
TDOA
Candidate regions of origin of expansion
Phylogeographic histories
θ, m
Priors for demographic parameters
CONVERTMAP
ABCSAMPLER + SPLATCHE2
Range expansion simulations
CALSUMSTAT
Calculation of summary statistics including spatial statistics (e.g., ψ)
ABCESTIMATOR + CALORIGIN
Posterior distribution of estimated origins
X-ORIGIN pipeline

FIGURE 1 The required data inputs (shown in boxes) and workflow of the X-ORIGIN pipeline are highlighted in the schematic. Specifically, to infer the geographic location from which an expansion originates, Ω (i.e., the actual latitudinal and longitudinal coordinates of the ancestral source population), a habitat suitability map, candidate regions of Ω and priors for demographic parameters are required. To consider how habitat heterogeneity might impact the range expansion process, the habitat suitability map can be informed by spatial (as well as temporal) variation in suitability (e.g., from ENMs based on contemporary bioclimatic variables, or palaeoclimatic variables; see He et al., 2013). Otherwise, the expansion process can be modelled as a diffusion process (i.e., equal habitat suitability across space and time). Likewise, users have the option of either entering candidate regions of Ω (e.g., a region identified by the regression approach of Peter & Slatkin, 2013; as discussed in the text), or the entire map area can be evaluated during the inference procedure. The pipeline calls up different software packages for downstream generation of simulations and estimation of the expansion origin, candidate regions of Ω. Specifically, spatially explicit coalescent simulations are used to generate expected patterns of genetic variation under a demographic model the expansion process (either informed or not by spatial and temporal heterogeneity of the landscape) using a modified version of the program SPLATCHE2 (Ray et al., 2010). Summary statistics are calculated from each simulated data set using R script that are incorporated in the pipeline, which are compared with those calculated for empirical data to inform the posterior distribution of Ω using ABC. Note that all steps can be performed seamlessly in X-ORIGIN, which has a wrapper for connecting all the steps in R or python scripts. Scripts for the pipeline are shown in grey shaded boxes, while external programs called in the pipeline are shown without boxes

2015), the location of putative ancestral source populations remains unclear.

2 | METHODS

2.1 | Statistic inferences using the X-ORIGIN pipeline

The X-ORIGIN pipeline couples information from a series of independent analyses to make inferences about Ω, the geographic location of ancestral source populations, by estimating the posterior probability of Ω under an ABC framework (Figure 1). Scripts are provided in the X-ORIGIN pipeline for all the steps involved, and a detailed tutorial is provided on GitHub (see https://github.com/KnowlesLab/X-ORGIN).

Briefly, the approach employs a spatially explicit coalescent to generate expected patterns of genomic variation under a set of priors, including a prior on Ω and priors on demographic parameters of the expansion process (i.e., *k* and *m*, the local population sizes and migration rates, and an ancestral population size, N_A). That is, genomic simulations of range expansion are initiated at different random locations within the geographic range specified by the prior on Ω and for different population size and migration rate values. If there is no prior knowledge on possible geographic origins, all demes on the map used for demographic simulations will be tested. Otherwise, a prior on Ω can be based on the fossil record, or a general candidate region might be based on the regression between pairwise population differences of Ψ and geographic distances (see Peter & Slatkin, 2013).

To make inferences using X-ORIGIN that considers the effects of spatial and temporal environmental heterogeneity on the expansion process, X-ORIGIN models the impact of this environmental heterogeneity on the expansion process. Specifically, heterogeneity in habitat suitability might be derived from ecological niche models (ENMs) for the present or the past (Sindato et al., 2016; Waltari et al., 2007), or from information on known barriers (e.g., mountain ranges, glaciers and bodies of water; Boehm et al., 2013; Knowles & Massatti, 2017; Waltari & Hickerson, 2013). These suitability maps are used to inform demographic dynamics associated with the expansion process by specifying different likely migration events as a function of spatial and/or temporal environmental heterogeneity. Specifically,

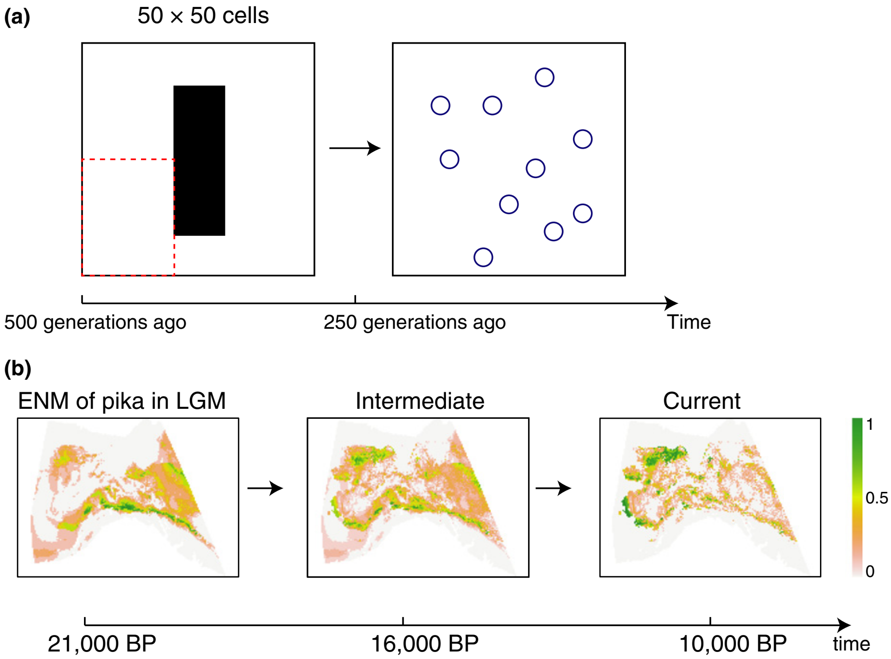

FIGURE 2 Simulated scenario used to evaluate the performance of the X-ORIGIN pipeline for inferring the geographic origin of a range expansion. In the simulated scenario, (a) expansion proceeded from the lower left corner of the map (shown as the red dotted area) across a homogeneous landscape with a centrally located geographic barrier during the first 250 generations, but not the last 250 generations (i.e., there is spatial and temporal habitat heterogeneity, where the area of the barrier has zero suitability). Due to the symmetry of the landscape, we varied the origin of expansion in the simulations within the red dotted area instead of the whole map. Circles mark populations that are sampled and for which summary statistics are calculated from multiple individuals. (b) An empirical application of X-ORIGIN in the Collared pika in which habitat suitability varied spatially and temporally across the Alaskan landscape. Ecological niche models were used to estimate habitat suitabilities for the present and past (i.e., the LGM) using climatic data (see Lanier et al., 2015 for details about ENMs). Specifically, the demographic expansion process proceeded across a temporally and spatially heterogeneous landscape, in which the habitat suitabilities from an ENM estimated for the LGM was used to inform the first 5,000 years of the simulated demographic expansion, followed by 6,000 simulated years of expansion across an intermediate surface (i.e., a map with average habitat suitability scores between those from the ENM for the present and LGM), and then 10,000 years of expansion with the habitat suitabilities from an ENM based on current climatic conditions

the habitat suitability scores for each deme determine local population sizes, thereby influencing the actual number of migrants across demes per generation. If distributional shifts are induced by climatic changes, then temporal shifts in habitat suitability can be incorporated into the demographic modelling (i.e., applying different relative weighting of suitability information from past vs. current ENMs to mirror trends of climatic change; see Brown & Knowles, 2012), given that shifts in connectivity over time can influence the expansion process, and consequently, the patterns of genetic variation across the landscape.

2.2 | Programs called up in the X-ORIGIN pipeline

In the X-ORIGIN pipeline, demographic and spatially explicit coalescent simulations are performed in SPLATCHE2 (Ray et al., 2010) in conjunction with a customized script in the X-ORIGIN pipeline to allow for temporally changing landscapes. Local demographic parameters (i.e., *k* and *m*) are informed from habitat suitability by scaling these parameters proportionally to the habitat suitability values of local demes (Figure 1), which might be temporally dynamic (i.e., the habitat suitability for a particular location may change in each generation based on shifting climatic conditions; see Brown & Knowles, 2012). Each generation *m* proportion of the population migrates out of the local deme; migration occurs to the adjacent four cells (north, south, west, east). After the exchange of individuals, local demes grow logistically with a rate *r* and are regulated by the local carrying capacity (which are also rescaled as a function of the habitat suitability of a deme); *r* can be set to a specific value (e.g., He et al., 2013), and as we do here ($r = 1$), or it can also be estimated as a parameter. For each time-forward simulation (i.e., a spatially explicit map of per generation local population sizes and migration events), a series of corresponding time-backward coalescent genetic simulation are run, with a separate coalescent simulation generated for each independent locus in the study. The ancestry of an allele will trace back from the present into ancestral source populations, where the pattern of gene lineage coalescence across the landscape and the timing of coalescence is defined by the time-forward local demographic simulations (i.e., the per generation *k* and *m* parameter values). SNP mutation models are then used to simulate patterns of genomic variation in SPLATCHE2, where the state of each SNP is generated across the independent coalescent simulations.

To generate patterns of genomic variation to compare with the empirical data, the simulated data sets are constructed by sampling the same populations (in geographic space), the same number of individuals, and the same number of SNPs as the empirical scenario. Summary statistics are calculated for both the empirical and simulated data sets. These include the spatial summary Ψ statistics calculated within, between, across all populations, as well as pairwise population F_{ST} values; ARLEQUIN 3.5 (Excoffier & Lischer, 2010) is used to calculate F_{ST}. Note that other nonspatial statistics often used in ABC analyses were also considered (e.g., *K*, the number of haplotypes, and *H*, observed heterozygosity). These additional summary statistics are not used in the analyses presented here because of the lack information they contained under the expansion scenarios (see Fig. S1); however, a user could employ them in X-ORIGIN if they determine they are relevant to the expansion history under study.

The empirical summary statistics are compared to those from the simulated data using approximate Bayesian computation (ABC), as implemented with ABCESTIMATOR in ABCTOOLBOX (Wegmann, Leuenberger, Neuenschwander, & Excoffier, 2010). Rather than conducting ABC analyses directly on the summary statistics, principal components (PCs) are extracted from all predictor variables to remove the effects of interactions between summary statistics, as well as to reduce "the curse of dimensionality" (i.e., when too many statistics are included, the distance between the simulated and empirical values systematically increases, reducing the accuracy of parameter estimates and making it more difficult to distinguish among models) (Wegmann & Excoffier, 2010; Wegmann, Leuenberger, & Excoffier, 2009).

Five thousand simulations (0.5%) whose transformed summary statistics are closest to those calculated from the empirical genomic data are retained for estimating the model parameters (i.e., Ω, the geographic locations of the ancestral source populations, and the demographic parameters *k*, *m*, and N_A). To jointly estimate the likelihood of a specific deme as the origin Ω (i.e., a specific longitude and latitude), the kernel densities of Ω across the retained simulations were estimated and used as the likelihood. This provides a nonparametric way of smoothing and estimating the likelihood of the origin based on the limited retained simulations (i.e., from the 0.5%, or five thousand retained simulations).

To check whether the inferred model is capable of generating the observed data, the likelihood of the empirical data given the model is compared with the likelihoods of the retained simulations. The fraction of simulations that have a smaller likelihood than the empirical data is expressed as a *p*-value, with small *p*-values indicating that a model is highly unlikely (Wegmann et al., 2010). Likewise, we conduct standard evaluations of the quality of the inferences from ABC (e.g., bias in parameter estimates; described below).

2.3 | Performance of the X-ORIGIN pipeline

We tested the pipeline on a simulated scenario (Figure 2a) to evaluate the performance of the approach for inferring the geographic location of the source population, Ω, under a temporally changing landscape. Specifically, simulations were conducted on a 50 × 50 deme landscape with a centrally located geographic barrier that was present in the past but not the present and expansion proceeded from the lower left deme (Figure 2a). Simulations were run for 500 generations, in which the barrier persisted for 250 generations. At the end of the simulations, 10 diploid individuals were sampled from 10 demes from across the distributional map. A range of migration rate (10^{-3}, 10^{-2}), ancestral population size (10^{-3}, 10^{-4}) and carrying capacity values (10^{-3}, 10^{-4}) per deme were simulated to check whether the inferred origin is sensitive to particular details of the demographic expansion process.

The accuracy of X-ORIGIN was evaluated by measuring the geographic distance between the actual and inferred geographic location of the source population (i.e., differences in the actual and inferred latitudinal and longitudinal coordinates). In addition to evaluating the accuracy of the estimated Ω under the model in which expansion proceeded from the upper left deme (Figure 2), we also tested whether the accuracy of Ω varied depending upon the geographic origin of the expansion. Specifically, we investigated the performance of the model by inspecting the average error of the inferred Ω of 10 pseudo-observed data sets (i.e., PODs from the simulations) in which the geographic origin of the expansion differed. Specifically, Ω was systematically varied so that each deme across the entire map served as the source of expansion.

In addition, the accuracy of X-ORIGIN pipeline is compared with Peter and Slatkin (2013)'s original "time difference of arrival location estimation" (TDOA) approach as well as a modified TDOA approach, which incorporates spatial heterogeneity in migration patterns (Olave, He, & Knowles, unpublished data). Specifically, we calculated the distance between the actual geographic origin with the one estimated from the TDOA approaches. The TDOA approach identifies the origin of the expansion by locating the deme that explains the highest proportion of variation in the correlation of pairwise Ψ differences and the pairwise differences of geographic distances of the populations to the potential origin. The modified TDOA approach correlates pairwise Ψ differences with pairwise resistance differences (McRae & Nürnberger, 2006) in which heterogeneous landscape is considered (Olave et al., unpublished data), whereas the original TDOA (Peter & Slatkin, 2013) assumes migration occurs on a homogeneous landscape (i.e., according to a random diffusion model). We

TABLE 1 Prior ranges for demographic and genetic parameters used in the demographic simulations of Collared pika

Parameters	Description	Prior ranges	Distribution
m	Migration rate between demes	($10^{-3.6}$, 10^{-2})	Log-uniform
N_{ans}	Ancestral population size before expansion	(36,880, 508,318)	Uniform
K	Carry capacity per deme	($10^{3.3}$, $10^{4.6}$)	Log-uniform
Lat	Latitude range of origin	(1,073,893, 1,850,478)	Uniform
Long	Longitude range of origin	(616,487, 899,496)	Uniform

conducted a cursory examination of the robustness of X-ORIGIN to model misspecification as well.

2.4 | Demonstration of X-ORIGIN with application to Alaskan Collared pika

In addition to details about the ABC analyses, here we briefly describe the empirical genomic data we analysed with X-ORIGIN, given that all data used here are from previous publications and are referenced below. Specifically, we analyse a genomic data set collected in the Alaskan Collared pika (for details on library construction and rigorous quality filtering see Lanier et al., 2015). Maps of environmental heterogeneity used in the X-ORIGIN analyses to infer Ω, the geographic location of the ancestral source population for the Collared pika, were generated from ENMs for the present and the last glacial maximum, LGM (see details in Knowles et al., 2016).

2.4.1 | Genomic data set

We analysed RADseq data for eight populations; note, we excluded the Pika Camp (Wrangell-St. Elias Mtns; GIS coordinates 61.2170, −138.2670) from our analyses because previous analyses indicate that it was founded from a separate ancestral refugial source (Lanier et al., 2015). Of the 23,493 RADseq loci with at least one biallelic SNP across populations, we analysed 6,816 loci with one SNP retained per RADseq loci in 50 individuals (i.e., 6–8 individuals per population, with the exception of Jawbone Lake, where $n = 2$); loci in <50% of the samples or were not present in more than one individual per population were excluded. Note that this is an expanded data set relative to those previously published (i.e., Lanier et al., 2015; Knowles et al., 2016) because we recovered more genetic information using ddRAD aligned to a reference genome for *Ochotona princeps* (American pika; ID: 771).

The directionality index Ψ requires information on the ancestral vs. derived states of SNPs because the statistic is calculated by counting the difference in derived allelic frequencies between pairs of populations (see Eq. 1 in Peter & Slatkin, 2013). Ancestral states of independent biallelic SNPs were determined by aligning the sequences with *Ochotona princeps* (American pika; ID: 771; https://www.ncbi.nlm.nih.gov/genome).

2.4.2 | Prior on Ω, the geographic locations of the origin of expansion

The TDOA approach was conducted to select candidate regions of origin to inform the prior on Ω (as opposed to considering the entire state of Alaska). Specifically, for each potential geographic location as the site of the ancestral source population (i.e., each deme from the distributional map), linear regression was performed between pairwise Ψ differences and the pairwise differences of geographic distances of the populations to the potential origin. The linear regression was repeated for each of the different potential geographic origins, and the geographic locations with R^2-values larger than 0.5 were used to specify the prior on the geographic location of the ancestral source population (regression analyses were conducted using modified scripts from Peter & Slatkin, 2013; which we provide on KnowlesLab/Github). This generated a target area of approximately 442,300 km^2 (i.e., 1,302 demes, with a size of 18.4 × 18.4 km^2 for each deme; Table 1) to analyse in detail regarding the posterior probability of Ω, the geographic location of the ancestral source population for the set of eight Collared pika populations collected across its range (see Lanier et al., 2015 for details).

2.4.3 | Estimates of habitat heterogeneity across space and time

Maps of environmental heterogeneity for the Collared pika were generated from ENMs (see details in Knowles et al., 2016). Briefly, inferences about differences in habitat suitability across space were made for the present and the LGM from ENMs based on bioclimatic data for the present and palaeoclimatic data from 21 kya. The models were tested over combinations of regularization parameters from 0.25 to 3 in intervals of 0.25 and the Linear, Quadratic, Hinge, Product and Threshold features using SDMTOOLBOX (Brown, 2014). Each model parameter class was replicated 25 times using cross-validation.

In addition, temporal shifts in habitat suitability were represented using differences in the relative weighting of habitat suitabilities estimated for the present and LGM across time to reflect climatic trends in the region over the past 21,000 years (Brown & Knowles, 2012). Specifically, the current ENM suitability map was used to represent the present to 5,000 years ago, an intermediate suitability map (i.e., an average suitability between the current and LGM ENMs) for the time period 5,000–11,000 years ago, and the LGM ENM suitability map for 11,000–21,000 years ago.

2.4.4 | ABC analyses

Data sets were simulated for 2,100 generations (based on a scaling factor of 10 to reduce the computational requirements; see He et al., 2013) to represent the range expansion from last glacial maximum. Priors for the local carrying capacity (k), ancestral population size (N_{ans}) and migration rates (m) were set according to Lanier et al. (2015) (see Table 1). Note that a geographic grid of 18.4 × 18.4 km^2 corresponded to a single deme and expansion was modelled across the Alaskan landscape (i.e., over approximately 2,197,850 km^2).

As with tests of the general performance of X-ORIGIN, we compared the estimates of Ω, the geographic location of the ancestral source of expansion, with results from (i) the TDOA method, where heterogeneity in the present landscape is not incorporated (i.e., the geographic distances separating populations were represented as pairwise Euclidean distances), (ii) the modified TDOA method, where resistance distances based on heterogeneity in the current habitat suitability is used, and (iii) X-ORIGIN, where temporal shifts in the heterogeneity of the landscape over time are accounted for. To evaluate the accuracy of estimates of Ω, five thousand pseudo-observations were generated from the prior distributions of the parameters. If the estimated parameters are unbiased, posterior quantiles of the parameters from the pseudo data sets should be uniformly distributed (Cook, Gelman, & Rubin, 2006; Wegmann et al., 2010). The posterior quantiles of true parameters for each pseudo run were calculated based on the posterior distribution of the regression adjusted 5,000 simulations closest to the pseudo-observed data sets.

3 | RESULTS

3.1 | Performance of the X-ORIGIN pipeline

For the example history considered here, which involved a central barrier that was present in the past, but not the present (i.e., there is both spatial and temporal heterogeneity in habitat suitabilities) X-ORIGIN gives more accurate inferences of Ω, the geographic location of the source population of the expansion, than the TDOA approach. In fact, the performance of X-ORIGIN was quite good, estimating the most likely origin within 1–4 demes of the actual origin (mean *p*-value = .67) from different starting positions across the map (and hence, differences in when and where the expansion process interacted with the geographic barrier), except for the lower left grid of the geographic area (Figure 3a c; see Fig. S2 for detailed examples of variation in inferences across PODs for different locations of origins).

In contrast, the majority of analyses with the TDOA approach give inferred locations that differ markedly from the actual area where the expansion originated, irrespective of where on the map the expansion originates (Figure 3b). The performance of the TDOA approach was especially poor (i.e., large discrepancies between the inferred and actual geographic origin of expansion) when the ancestral source area was near the barrier (Figure 3b). This variation in accuracy highlights the importance of explicitly modelling the temporal heterogeneity of landscapes (also see Wegmann, Currat, & Excoffier, 2006), as it strongly distorts the Ψ signatures, especially if the heterogeneity is present in the early stage of the expansion.

3.2 | Inferred geographic origin of expansion in the Alaskan Collared Pika

For the set of Collared pika populations studied here, the highest likelihood (marginal density: 1.82 × 10^{-8}; *p*-value: .996) for the location of the expansion origin, Ω, is the Mackenzie Mountains in Yukon Territory, Canada (Figure 4). This inference is based on the retained 5,000 simulations whose summary statistics were to those of empirical data. The geographic origin of expansion (i.e., the latitudinal and longitudinal coordinates) was estimated using a two-dimensional kernel density of the retained simulations, implemented using the kde2d function in the MASS package of R (Venables & Ripley, 2002).

The geographic origin of expansion inferred using X-ORIGIN differed from the TDOA results (Figure 4). Moreover, neither the

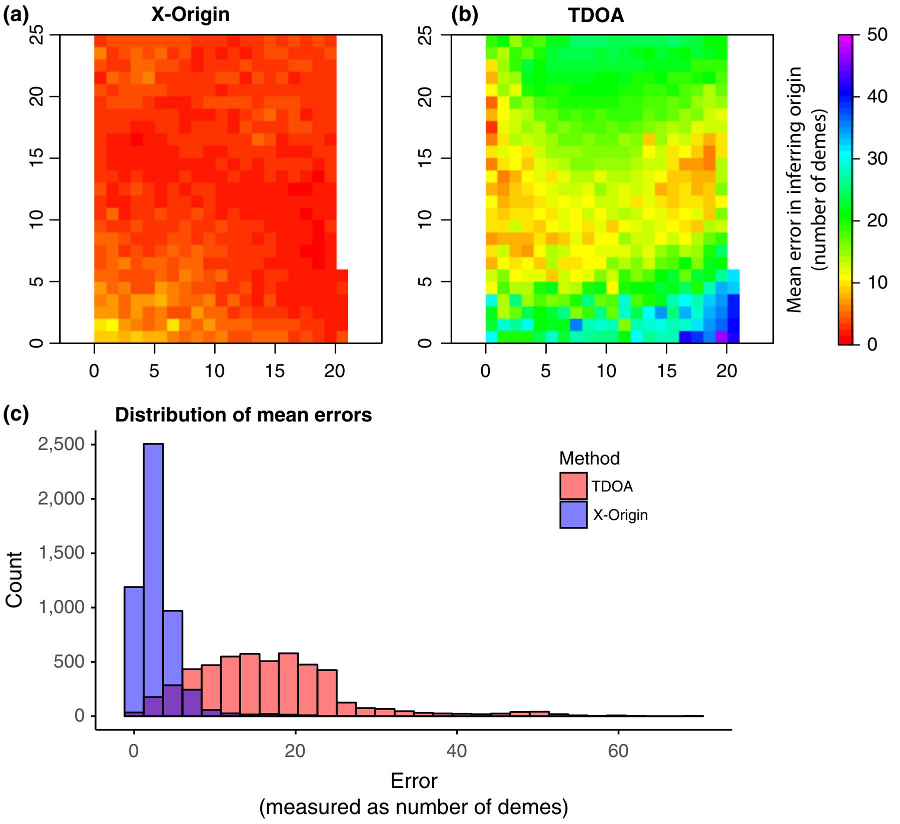

FIGURE 3 Distribution of the mean errors in the estimated Ω across the map (i.e., for different geographic locations for the origin of expansion; the red dotted area in Fig. 2a) under the simulated scenario using (a) X-ORIGIN vs. (b) the TDOA approach. Colour of each deme shows the accuracy of origin estimation if the expansion starts from that particular deme, which is measured by the distance between its inferred origin, Ω, and the actual origin, averaged across 10 simulations. Also shown are the histograms of accuracy across all 5,000 instances (c) from X-ORIGIN vs. the TDOA approach. Distances are in the units of the number of demes from the actual origin

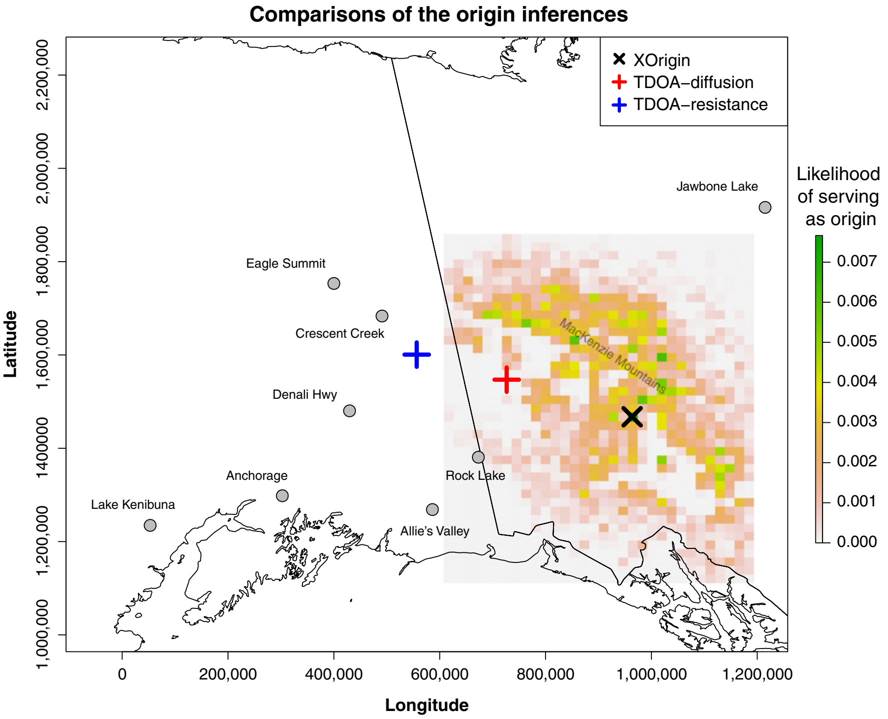

FIGURE 4 Estimates of the origin of expansion, Ω, inferred in the Collared pika using X-ORIGIN compared with the TDOA approach. The deme with the highest likelihood inferred from X-ORIGIN is marked with a black "X," whereas the location of origins estimated using TDOA methods are marked with crosses. The heat map shows differences in the probability density estimates of different demes across the map being the origin of expansion, as estimated in X-ORIGIN, with the greener shades representing higher probabilities; the shaded square area represents the prior area for Ω, a region identified by the regression approach of Peter and Slatkin (2013). Each deme in the map has equal relative size (i.e., the map is projected using the North American Datum—NAD83—readjustment of the global positioning system that accounts for the earth's curvature) and population localities of sequenced individuals are marked by grey circles

inferred area based on the pairwise Ψ matrix on a homogeneous landscape (TDOA-diffusion) nor the one based on a resistance map of the current landscape suitabilities (TDOA-resistance) are in areas with high likelihoods. That is, simulated genetic data sets where expansion proceeded from the inferred areas under the TDOA approaches do not correspond to the observed genetic data (i.e., there is a mismatch between the empirical summary statistic and those calculated from the simulations).

Based on the distances between actual vs. inferred origin for each of the different method, X-ORIGIN outperformed TDOA, although the accuracy of inferred Ω-values varied depending upon the geographic origin of the expansion (Figure 5). We also note that the accuracy was generally lower for the heterogeneous landscape inferred for pikas relative to the landscape used to validate the X-ORIGIN package (Figure 5 vs. Figure 3). In particular, populations that originated from the southeast region exhibited the lowest accuracy (i.e., the greatest difference between the inferred and actual value of Ω). This is most likely due to the lack of samples from that area, and consequently, little information of the direction of asymmetrical gene flow expected under an expansion model (see Peter & Slatkin, 2013). Nevertheless, comparison of the accuracy of inferences between X-ORIGIN and TDOA approaches indicates those from X-ORIGIN are more accurate for an expansion originating from the Mackenzie Mountain range. Specifically, analysing simulated data of expansions from the Mackenzie Mountain range (i.e., the PODs from the ABC simulations), the TDOA approaches give estimates that are generally displaced by 15–30 demes from the actual origin of expansion (i.e., a discrepancy of 750–1,500 km), and curiously, these were more inaccurate than inferences with a southwest geographic origin of expansion (Figure 5), despite sampling of populations in that region (see discussion below).

4 | DISCUSSION

Patterns of genetic variation in individuals sampled in the present harbour rich information about past movements of species. In contrast to those from nonspatial models of population demography (e.g., changes in population size or admixture proportions; see Hey, 2005; Theis, Ronco, Indermaur, Salzburger, & Egger, 2014), recent developments have focused on inferences from spatially explicit approaches. Specifically, departure from equilibrium status of population movements under a diffusion model, "isolation by distance," caused either by range expansion/contraction history, long distance admixture or habitat heterogeneity is tested through different approaches. One general approach is to quantify discrepancies between spatial genetic patterns and the expectations from geographic distances. For example, discrepancies between population's positions on a genetic PCA map can be visualized against a map of their geographic distribution using Procrustes analyses to examine where on a landscape patterns of genetic variation depart from isolation by distance (Knowles et al., 2016; Wang, Zöllner, & Rosenberg, 2012), or a "geogenetic map" can be used to infer potential long-range admixture among populations (Bradburd, Ralph, & Coop, 2016). Similarly, disruptions to past movement might be inferred by relating the effective migration rates to expected genetic dissimilarities for an interpolated geographic map of barriers or corridors among populations (see Petkova, Novembre, & Stephens, 2016).

Instead of quantifying discrepancies from isolation by distance, our approach directly models expected patterns of genetic variation using spatial genetic indices and makes inferences about historical movements—specifically, the geographic origin of expansion, Ω—under an ABC framework, while incorporating temporal shifts in habitat suitability over time. This is not the first approach for directly evaluating genetic variation under models of historical movement. For example, the spatial genetic indices applied here were developed to directly infer historical movements based on shifts in the genetic summary statistics across a landscape (Peter & Slatkin, 2013), and spatial-autocorrelation of genetic covariance information has been applied to distinguish among spatially explicit demographic scenarios (Alvarado-Serrano & Hickerson, 2016; Bertorelle & Barbujani, 1995; Coop, Witonsky, Rienzo, & Pritchard, 2010). However, our approach infers and evaluates the parameter Ω—the actual latitudinal and longitudinal coordinates for the origin of an expansion—that is not based on the assumption of a diffusion model and provides statistical rigorousness and flexible applications for inferences about historical expansion scenarios. First, we can evaluate the likelihood of different geographic locations as the origin of a population expansion, accounting for both spatial and temporal heterogeneity in habitat suitability of the landscape. Second, with the freely available X-ORIGIN pipeline we developed, users can validate any inference by determining whether the inferred model is capable of generating data that generally corresponds to the empirical data, which is equally important as estimating the most likely model for the origin of expansion (i.e., the most likely location for the origin of expansion may nonetheless be a poor fit to the observed data). Such attributes are not currently implemented in other methods for inference about expansion histories (e.g., compare with Ray et al., 2005).

Below, we discuss how these attributes make X-ORIGIN not only a practical tool, but as our analyses demonstrate, also one whose performance is better than not accommodating such dynamic histories. Likewise, we highlight how this pipeline can easily be adapted for a more general inference approach beyond inferring the origin of expansions, especially with the development of new spatial indices. However, we also note the difference in performance of X-ORIGIN between a simple demographic history (i.e., the one used to validate the approach) and the one with more extreme habitat heterogeneity, and caution users of the importance for validating the accuracy of the inference, which can be implemented in the X-ORIGIN pipeline. We apply this practice when interpreting the results from the X-ORIGIN analysis of the Collared pikas, as well as discuss aspects of the data that might contribute to uncertainty in the inferred origin of expansion, and the importance of corroborative evidence not based on the genetic data itself.

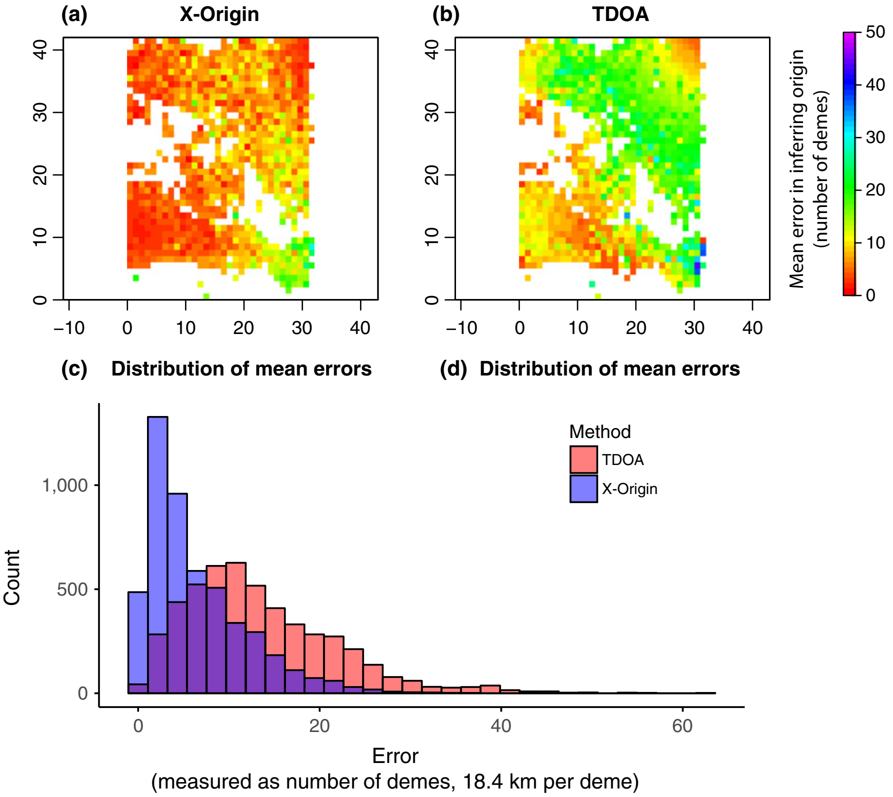

FIGURE 5 Distribution of mean errors in the estimated Ω across the map (i.e., for different geographic locations for the origin of expansion) for pseudo-observations in the Pika simulations (see Figure 2B) using (a) X-ORIGIN vs. (b) the TDOA approach. In total, 5,000 pseudo-observations are generated and colour of each deme shows the accuracy of origin estimation if the expansion starts from the particular deme, which is measured by the average distance between its inferred origins and the actual origin. White area on the map contains demes where not all populations can be colonized if the expansion starts from there. Also shown are the histograms of accuracy across all 5,000 instances (c) from X-ORIGIN vs. the TDOA approach. Distances are in the units of the number of demes from the actual origin, and each deme is 18.4 km in length

4.1 | Factors impacting the accuracy of inferences about the geographic origin of expansion

The Ψ index directly captures the overall trend of differences in frequencies of derived polymorphic alleles in populations based on the fact that expanding front of populations are experiencing serial bottlenecks. Therefore, Ψ indices are informative as long as current populations have not yet reached equilibrium. If the majority of the pairwise Ψ indices are close to zero in the system (which is not the case for pikas; Table S1), the lack of spatial gradient in the Ψ indices indicate that either there was not an expansion or a sufficient amount of time since the expansion has passed such that its genetic signature can no longer be detected by the Ψ indices (see also Peter & Slatkin, 2013). We tested a scenario in the Pika data set where there is no expansion origin to examine the performance of X-ORIGIN. Specifically, we simulated 1,000 replicate data sets in which all populations started from their sampling areas to reach equilibrium states. For these data sets, although Ψ indices deviate strongly from zero, no origin can be estimated from TDOA as no positive relationship between pairwise differences of Ψ and geographic distances among populations can be established (Table S2). Likewise, with X-ORIGIN, marginal densities of the expansion model are extremely low (on the order of 10^{-200} to 10^{-12} as compared to 10^{-8} for PODs that experienced expansion from a single origin) and *p*-values are zero (Table S2). Therefore, X-ORIGIN, like TDOA, will not give misleading results about the potential origin for expansion when no such expansion occurred.

Any inference that extracts information on the geographic distribution of genetic variation requires adequate sampling of populations as well as number of independent SNPs (i.e., at least more than 1,000 independent SNPs; Peter & Slatkin, 2013; Bradburd et al., 2016). Our results clearly show that inferences become less accurate when sampled populations are located further from the location where an expansion originated (e.g., see higher error rate at southeast corner of Figure 5a). Therefore, researchers should carefully consider the sampling design. In particular, our results (see also Peter & Slatkin, 2013; Bradburd et al., 2016) suggest that obtaining accurate inferences that utilize spatial information about the distribution of genetic variation may be dependent upon which populations are sampled, rather than whether there is sufficient power for such inferences related to the number of loci analysed. Although it is beyond the scope of this study, this general question is something that could be explored using the X-ORIGIN pipeline.

Another factor impacting the accuracy of inference relates to model misspecification. Specifically, complicated demographic scenarios such as those involving two or more geographic origins of expansion will give misleading results if not accommodated (see also Peter & Slatkin, 2013). There are a number of ways to accommodate and/or test whether an assumed expansion from a single source might be violated. For example, clustering algorithms can be run to delineate populations into different groups with potentially different expansion origins and validated by a minimum-spanning tree built from a matrix of Ψ -values (Peter & Slatkin, 2013), followed by separate inferences of Ω for each subgroup of populations. Alternatively, competing explanatory models with multiple origins vs. one expansion origin can be analysed in X-ORIGIN and compared in a model selection framework. Our results also suggest that any model, even those that might be more probable than others, should be interpreted with caution if Ω is located in areas with low confidence (based on reference to simulated data sets), or if the most likely model nevertheless has a low probability of generating data that resembles the empirical data (i.e., low *p*-value; Wegmann et al., 2010; see He et al., 2013 for details of model validation).

Despite positive aspects of X-ORIGIN related to estimating the likelihood of the expansion origin, and consequently, uncertainty surrounding this inference (e.g., the geographic area spanned by the 90% highest posterior density of Ω), as well as validating the inference using PODs (see Figure 5), one unexplored issue is how errors early in the pipeline might get amplified and generate misleading results. We did a cursory examination of how such errors might impact an inferred expansion origin. Specifically, we examined how robust the inferred origin might be to uncertainties regarding the temporal changes in habitats—in this case, the duration of a barrier, as in the scenario, we used to validate X-ORIGIN (see Figure 2). When we varied the true duration of the barrier to simulate data (i.e., simulate data with a barrier that persisted for 200–300 generations, rather than 250 of the 500 generations), we observed no difference in the accuracy of the Ω estimation (Fig. S3). This shows that the pipeline can be robust to misspecification of temporal dynamics of a historical scenario (at least for the parameter space examined here). This clearly should not be interpreted as general evidence of robustness to model misspecification. Rather we present it here to show that X-ORIGIN exhibits some robustness, but also to emphasize that all users can conduct their own investigation to robustness tailored to the specifics of their application.

There are of course other paths for errors that could impact the accuracy of inferences about Ω. For example, we use ENMs to estimate potential suitable areas to inform demographic models (see Figure 1). As a consequence, the results from X-ORIGIN could be impacted by poor ENMs (i.e., validation and best practices of ENMs should be followed). In addition, applying different transformation of habitat suitabilities into local carrying capacities can affect patterns of genetic variation (see Brown & Knowles, 2012). There are different strategies one might take to avoid biases that could result from unrealistic assumptions or errors in the upstream steps of the pipeline (Figure 1). For example, instead of using a fixed suitability score from an ENM model for each deme, suitability scores between maximum and minimum range inferred for each deme might be randomly sampled during the simulation process to generate expected patterns of genetic variation that incorporate some uncertainties in the ENM modelling. This might increase the number of simulations required for inferring Ω to get an unbiased and precise estimate under an ABC framework, given that accommodating such uncertainties may increase the variance in observed patterns of genetic variation in simulated data sets. Likewise, different transformations of habitat suitabilities into local carrying capacities (scaling habitat suitability linearly with local carrying capacity vs. a step function; Brown & Knowles, 2012) could be incorporated as alternative models to be tested (i.e., treated in a model selection framework, even when the primary interest is on inferring the origin of expansion, Ω).

Although such flexibility in accounting for uncertainty or potential errors in upstream steps (Figure 1) is a strength of the X-ORIGIN package we developed, the application of X-ORIGIN (especially compared with TDOA; Peter & Slatkin, 2013) comes with much more computational expense. For example, a typical spatially explicit simulation of 2,000 generations on a 150 × 150 grid layer and the generation of 1,000 SNPs takes more than 7 min. Users are advised to calculate required computational resources before experimenting with the pipeline. This includes reducing the size of the Ω prior (e.g., by applying TDOA as a preliminary step for data inspection, as applied in the Collared pika example).

4.2 | The Mackenzie Mountain region as the most likely origin of expansion in Collared pika

As an alpine small mammal, suitable habitats for Collared pika are spatially highly heterogeneous, but also temporally heterogeneous given that Alaska was directly impacted by the glacial cycles (Figure 2b). Previous analyses have suggested a potentially complex biogeographic history involving expansion from multiple ancestral sources (Knowles et al., 2016; Lanier et al., 2015; Lanier & Olson, 2009). Limited sampling of populations inhibits analysis of data subsets to explore such models with X-ORIGIN (i.e., multiple populations are required to estimate potential sources of expansion) and therefore is beyond the scope of this manuscript. Nevertheless, it is informative to consider how our inference compares to previous characterizations for the populations analysed here.

Previous studies that made inferences about the biogeographic and demographic history of the Collared pika applied analyses that assumed equilibrium status (e.g., F_{ST}, STRUCTURE analyses, estimates of phylogenetic relationships among populations). For example, in an analysis of the relationship between F_{ST} values among populations and the geographic distance separating them (Lanier et al., 2015), the most northeastern sampled population Jawbone Lake (Figure 4) appeared to be an outlier under the expectation of isolation by distance. Based on this result, and the relative genetic distinctiveness of the Jawbone Lake population and the other two north-central populations from the Yukon-Tanana Uplands (specifically, the Eagle Summit and Crescent Creek populations), these populations were analysed separately and a distinct pattern of isolation by distance at the regional level was interpreted as possible evidence of different ancestral source populations (Lanier et al., 2015). However, our analyses here provide a compelling argument for an alternative explanation. Specifically, the genetic similarities between Jawbone Lake and the Eagle Summit and Crescent Creek populations (See Figure 5 in Lanier et al., 2015) may not reflect a refugial source that was differed from the refugial source of other sampled populations. Instead, it may reflect their proximity to the geographic origin of expansion in an ancestral species, Ω in the Mackenzie mountains (see Figure 4), and more specifically, the similar geographic distance of the populations from the source of expansion. Even though our validation tests indeed show that the degree of reliability about expansion can be considerable (e.g., differing by as much as 1,500 km from the actual expansion origin depending upon where on the landscape the expansion proceeded from; Figure 5), the mean error surrounding estimates of Ω as a function of the distance from the actual origin is quite low (i.e., less than five demes away, or 250 km) for the geographic region with the highest likelihood of Ω (Figure 4). Interestingly, Procrustes analyses in the Collared pikas, as well as other codistributed alpine mammals, suggest a stronger deviation along the longitudinal axis between genetic variation and geography (i.e., genetic similarities more centrally located than the geographic space occupied by the populations; Knowles et al., 2016). Our analysis supported this deviation as a result of an expansion history along this axis, offering an alternative interpretation to the hypothesis of a centrally located refugium.

Lastly, ENMs for the LGM are not inconsistent with our estimate (Figure 2b). However, if we consider information from the ENMs by themselves, the region of high habitat suitability encompasses a broad area that does not offer much detail about the potential location of ancestral populations. This even includes a potential northwestern source population (Figure 2b), even though former genetic (Knowles et al., 2016; Lanier et al., 2015) and fossil studies (Gunderson, Jacobsen, & Olson, 2009; Lanier & Olson, 2013) suggest the lack of support for such a putative ancestral source (e.g., in the Brooks Range). Both X-ORIGIN and TDOA analyses reinforce that despite projections from the ENM for the LGM, this region does not appear to be a likely candidate as an ancestral source of expansion.

5 | CONCLUSIONS

Our results show that failing to consider the impact of spatial and temporal heterogeneity on the expansion process can lead to much less accurate inferences (Figure 3a compared with b, and Figure 5a compared with b). Furthermore, there are also ways to minimize potential errors when inferring the origin of expansion. For example, in our simulations, we place a broad prior on parameters that are not targets of interest, but may influence estimates of Ω (e.g., ancestral population, carrying capacity; see Table 1), thereby accounting for uncertainty about the demography of the expansion process. Moreover, the summary statistics used in the inference procedure (i.e., Ψ and F_{ST} values) are not sensitive to the absolute effective population sizes, but rather the ratio of size differences between population pairs. Lastly, despite the lower accuracy of inferences for complicated scenarios, as with the analysis of the Collared pika, relative to simple expansion scenarios (Figures 3 and 5), accounting for the effects of spatial and temporal heterogeneity is generally more accurate than applying an oversimplified model if the goal is to infer the geographic location of an expansions origin (Figure 3). Therefore, we argue that the caveats and concerns associated with inferring the origin of expansion do not nullify the utility of spatially and temporally explicit models, such as those applied here in the new X-ORIGIN pipeline. In particular, we show that it is incorrect to assume that environmental heterogeneity (whether temporal or spatial) will not impact inferred origins of expansion, and that despite the caveats we highlight with X-ORIGIN, they are less problematic than many implicit assumptions made in approaches that ignore geographic and temporal constraints on population movements or population size fluctuations (see Knowles & Alvarado-Serrano, 2010). Moreover, the reliability of any inference about the origin of expansion under the more complex models implemented in the X-ORIGIN pipeline can be (and should be) rigorously explored using validation procedures.

ACKNOWLEDGEMENTS

Specimens and genomic data analysed here were collected as part of past collaborative work with Hayley Lanier and Link Olson, for which we are grateful for their helpful discussions on small Alaskan mammals, and Collared pikas in particular. We appreciate the comments from two anonymous reviewers and Benjamin Peter that helped improve earlier versions of the manuscript.

DATA ACCESSIBILITY

vcf files for Collared pika and SNP used in the analysis were deposited on Dryad for data archive (https://doi.org/10.5061/dryad.4s1gg); X-ORIGIN pipeline tutorial, scripts, example files and input files used in the study are uploaded on GitHub and released under the DOI: https://zenodo.org/badge/latestdoi/100994225.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTION

Q.H. and L.L.K. conceived and designed the study. Q.H. wrote the pipeline code. Q.H. and J.R.P. analyzed the data. Q.H., J.R.P. and L.L.K. wrote the paper.

ORCID

Qixin He http://orcid.org/0000-0003-1696-8203

REFERENCES

Alvarado-Serrano, D. F., & Hickerson, M. J. (2016). Spatially explicit summary statistics for historical population genetic inference. *Methods in Ecology and Evolution*, *7*(4), 418–427. https://doi.org/10.1111/2041-210X.12489

Austerlitz, F., Jung-Muller, B., Godelle, B., & Gouyon, P.-H. (1997). Evolution of coalescence times, genetic diversity and structure during colonization. *Theoretical Population Biology*, *51*(2), 148–164. https://doi.org/10.1006/tpbi.1997.1302

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, *162*(4), 2025–2035.

Bemmels, J. B., Title, P. O., Ortego, J., & Knowles, L. L. (2016). Tests of species-specific models reveal the importance of drought in postglacial range shifts of a Mediterranean-climate tree: Insights from integrative distributional, demographic and coalescent modelling and ABC model selection. *Molecular Ecology*, *25*(19), 4889–4906. https://doi.org/10.1111/mec.13804

Bertorelle, G., & Barbujani, G. (1995). Analysis of DNA diversity by spatial autocorrelation. *Genetics*, *140*(2), 811–819.

Boehm, J. T., Woodall, L., Teske, P. R., Lourie, S. A., Baldwin, C., Waldman, J., & Hickerson, M. (2013). Marine dispersal and barriers drive Atlantic seahorse diversification. *Journal of Biogeography*, *40*(10), 1839–1849. https://doi.org/10.1111/jbi.12127

Bradburd, G. S., Ralph, P. L., & Coop, G. M. (2016). A spatial framework for understanding population structure and admixture. *PLOS Genetics*, *12*(1), e1005703. https://doi.org/10.1371/journal.pgen.1005703

Brown, J. L. (2014). SDMtoolbox: A python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods in Ecology and Evolution*, *5*(7), 694–700. https://doi.org/10.1111/2041-210X.12200

Brown, J. L., & Knowles, L. L. (2012). Spatially explicit models of dynamic histories: Examination of the genetic consequences of Pleistocene glaciation and recent climate change on the American Pika. *Molecular Ecology*, *21*(15), 3757–3775. https://doi.org/10.1111/j.1365-294X.2012.05640.x

Carnaval, A. C., Hickerson, M. J., Haddad, C. F. B., Rodrigues, M. T., & Moritz, C. (2009). Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, *323*(5915), 785–789. https://doi.org/10.1126/science.1166955

Cook, S. R., Gelman, A., & Rubin, D. B. (2006). Validation of software for Bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics*, *15*(3), 675–692. https://doi.org/10.1198/106186006X136976

Coop, G., Witonsky, D., Rienzo, A. D., & Pritchard, J. K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics*, *185*(4), 1411–1423. https://doi.org/10.1534/genetics.110.114819

DeGiorgio, M., Jakobsson, M., & Rosenberg, N. A. (2009). Explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. *Proceedings of the National Academy of Sciences*, *106*(38), 16057–16062. https://doi.org/10.1073/pnas.0903341106

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

François, O., Currat, M., Ray, N., Han, E., Excoffier, L., & Novembre, J. (2010). Principal component analysis under population genetic models of range expansion and admixture. *Molecular Biology and Evolution*, *27*(6), 1257–1268. https://doi.org/10.1093/molbev/msq010

Galbreath, K. E., Cook, J. A., Eddingsaas, A. A., & DeChaine, E. G. (2011). Diversity and demography in Beringia: Multilocus tests of paleodistribution models reveal the complex history of arctic ground squirrels. *Evolution*, *65*(7), 1879–1896. https://doi.org/10.1111/j.1558-5646.2011.01287.x

Gunderson, A. M., Jacobsen, B. K., & Olson, L. E. (2009). Revised distribution of the Alaska Marmot, *Marmota broweri*, and confirmation of parapatry with hoary marmots. *Journal of Mammalogy*, *90*(4), 859–869. https://doi.org/10.1644/08-MAMM-A-253.1

He, Q., Edwards, D. L., & Knowles, L. L. (2013). Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution*, *67*(12), 3386–3402. https://doi.org/10.1111/evo.12159

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*(6789), 907. https://doi.org/10.1038/35016000

Hey, J. (2005). On the number of new world founders: A population genetic portrait of the peopling of the Americas. *PLOS Biology*, *3*(6), e193. https://doi.org/10.1371/journal.pbio.0030193

Itan, Y., Powell, A., Beaumont, M. A., Burger, J., & Thomas, M. G. (2009). The origins of lactase persistence in Europe. *PLOS Computational Biology*, *5*(8), e1000491. https://doi.org/10.1371/journal.pcbi.1000491

Knowles, L. L. (2009). Statistical Phylogeography. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 593–612. https://doi.org/10.1146/annurev.ecolsys.38.091206.095702.

Knowles, L. L., & Alvarado-Serrano, D. F. (2010). Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: Insights from coupled ecological, demographic and genetic models in montane grasshoppers. *Molecular Ecology*, *19*(17), 3727–3745. https://doi.org/10.1111/j.1365-294X.2010.04702.x

Knowles, L. L., & Massatti, R. (2017). Distributional shifts – not geographic isolation – as a probable driver of montane species divergence. *Ecography*, https://doi.org/10.1111/ecog.02893

Knowles, L. L., Massatti, R., He, Q., Olson, L. E., & Lanier, H. C. (2016). Quantifying the similarity between genes and geography across Alaska's alpine small mammals. *Journal of Biogeography*, *43*(7), 1464–1476. https://doi.org/10.1111/jbi.12728

Lanier, H. C., Massatti, R., He, Q., Olson, L. E., & Knowles, L. L. (2015). Colonization from divergent ancestors: Glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). *Molecular Ecology*, *24*(14), 3688–3705. https://doi.org/10.1111/mec.13270

Lanier, H. C., & Olson, L. E. (2009). Inferring divergence times within pikas (*Ochotona* spp.) using mtDNA and relaxed molecular dating techniques. *Molecular Phylogenetics and Evolution*, *53*(1), 1–12. https://doi.org/10.1016/j.ympev.2009.05.035

Lanier, H. C., & Olson, L. E. (2013). Deep barriers, shallow divergences: Reduced phylogeographical structure in the collared pika (Mammalia: Lagomorpha: *Ochotona collaris*). *Journal of Biogeography*, *40*(3), 466–478. https://doi.org/10.1111/jbi.12035

Massatti, R., & Knowles, L. L. (2016). Contrasting support for alternative models of genomic variation based on microhabitat preference: Species-specific effects of climate change in alpine sedges. *Molecular Ecology*, *25*(16), 3974–3986. https://doi.org/10.1111/mec.13735

McRae, B. H., & Beier, P. (2007). Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences*, *104*(50), 19885–19890. https://doi.org/10.1073/pnas.0706568104

McRae, B. H., & Nürnberger, B. (2006). Isolation by resistance. *Evolution*, *60*(8), 1551–1561. https://doi.org/10.1554/05-321.1

Olave, M., He, Q., & Knowles, L. L. (unpublished data). Evidence for shared refugia based on allele frequency asymmetries of genomic data among five alpine Alaskan small mammal species.

Peter, B. M., & Slatkin, M. (2013). Detecting range expansions from genetic data. *Evolution*, *67*(11), 3274–3289. https://doi.org/10.1111/evo.12202

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100. https://doi.org/10.1038/ng.3464

Ramachandran, S., Deshpande, O., Roseman, C. C., Rosenberg, N. A., Feldman, M. W., & Cavalli-Sforza, L. L. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(44), 15942–15947. https://doi.org/10.1073/pnas.0507611102

Ray, N., Currat, M., Berthier, P., & Excoffier, L. (2005). Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations. *Genome Research*, *15*(8), 1161–1167. https://doi.org/10.1101/gr.3708505

Ray, N., Currat, M., & Excoffier, L. (2003). Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution*, *20*(1), 76–86. https://doi.org/10.1093/molbev/msg009

Ray, N., Currat, M., Foll, M., & Excoffier, L. (2010). SPLATCHE2: A spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics*, *26*(23), 2993–2994. https://doi.org/10.1093/bioinformatics/btq579

Sindato, C., Stevens, K. B., Karimuribo, E. D., Mboera, L. E. G., Paweska, J. T., & Pfeiffer, D. U. (2016). Spatial heterogeneity of habitat suitability for rift valley fever occurrence in Tanzania: An ecological niche modelling approach. *PLOS Neglected Tropical Diseases*, *10*(9), e0005002. https://doi.org/10.1371/journal.pntd.0005002

Theis, A., Ronco, F., Indermaur, A., Salzburger, W., & Egger, B. (2014). Adaptive divergence between lake and stream populations of an East African cichlid fish. *Molecular Ecology*, *23*(21), 5304–5322. https://doi.org/10.1111/mec.12939

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S (Fourth)*. New York, NY: Springer. https://doi.org/10.1007/978-0-387-21706-2

Waltari, E., & Hickerson, M. J. (2013). Late Pleistocene species distribution modelling of North Atlantic intertidal invertebrates. *Journal of Biogeography*, *40*(2), 249–260. https://doi.org/10.1111/j.1365-2699.2012.02782.x

Waltari, E., Hijmans, R. J., Peterson, A. T., Nyári, Á. S., Perkins, S. L., & Guralnick, R. P. (2007). Locating pleistocene refugia: Comparing phylogeographic and ecological niche model predictions. *PLoS ONE*, *2*(7), e563. https://doi.org/10.1371/journal.pone.0000563

Wang, C., Zöllner, S., & Rosenberg, N. A. (2012). A quantitative comparison of the similarity between genes and geography in worldwide human populations. *PLOS Genetics*, *8*(8), e1002886. https://doi.org/10.1371/journal.pgen.1002886

Wegmann, D., Currat, M., & Excoffier, L. (2006). Molecular diversity after a range expansion in heterogeneous environments. *Genetics*, *174*(4), 2009–2020. https://doi.org/10.1534/genetics.106.062851

Wegmann, D., & Excoffier, L. (2010). Bayesian inference of the demographic history of Chimpanzees. *Molecular Biology and Evolution*, *27*(6), 1425–1435. https://doi.org/10.1093/molbev/msq028

Wegmann, D., Leuenberger, C., & Excoffier, L. (2009). Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics*, *182*(4), 1207–1218. https://doi.org/10.1534/genetics.109.102509

Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABCtoolbox: A versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, *11*, 116. https://doi.org/10.1186/1471-2105-11-116

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: He Q, Prado JR, Knowles LL. Inferring the geographic origin of a range expansion: Latitudinal and longitudinal coordinates inferred from genomic data in an ABC framework with the program x-origin. *Mol Ecol*. 2017;26:6908–6920. https://doi.org/10.1111/mec.14380