

Integrating GIS-based environmental data into evolutionary biology

Kenneth H. Kozak¹, Catherine H. Graham² and John J. Wiens²

¹ Bell Museum of Natural History and Department of Fisheries, Wildlife & Conservation Biology, University of Minnesota, St. Paul, MN 55108 USA

² Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245, USA

Many evolutionary processes are influenced by environmental variation over space and time, including genetic divergence among populations, speciation and evolutionary change in morphology, physiology and behaviour. Yet, evolutionary biologists have generally not taken advantage of the extensive environmental data available from geographic information systems (GIS). For example, studies of phylogeography, speciation and character evolution often ignore or use only crude proxies for environmental variation (e.g. latitude and distance between populations). Here, we describe how the integration of GIS-based environmental data, along with new spatial tools, can transform evolutionary studies and reveal new insights into the ecological causes of evolutionary patterns.

The application of GIS in evolutionary biology

Understanding many evolutionary patterns and processes requires consideration of the variation in the abiotic environment (e.g. temperature, precipitation and topography). Although most evolutionary biologists appreciate this point in theory, it is not always apparent in practice. For example, most phylogeographic studies only consider how the geographic distance between sampling localities influences genetic structure, while ignoring variation in the environment separating them (e.g. changes in topography and/or climate). Similarly, studies of speciation rarely quantify environmental factors that might cause sister species to be allopatric (geographically separated) or parapatric (geographically abutting), even though these factors can be crucial in deducing the role of ecology in speciation. Likewise, studies of character evolution rarely include quantitative data on environmental variables where the species occur, even though environmental factors can influence the evolution of many fundamental biological traits, including what an organism feeds on, when it reproduces, how many offspring it has and how large it grows.

In the past, one could hardly blame evolutionary biologists for not explicitly incorporating environmental data into many of their studies. For example, how could one graduate student or even an entire team of researchers hope to obtain detailed environmental data for 100 sampling localities of a wide-ranging species, let alone from a group of 50 species distributed across the world?

Fortunately, the availability of geographic information system (GIS) maps of environmental variables (Table 1, Box 1) now makes it relatively simple to incorporate such information into evolutionary studies. By plotting the sampling locations of organisms on GIS maps, researchers can readily obtain environmental data for hundreds of localities and dozens of species across the globe. These data can then be combined with information on the phylogeny, ecology and phenotypes of organisms to help to address many fundamental and applied evolutionary questions.

Here, we review the application of GIS-based data and tools to three of the main endeavors of modern evolutionary biology: the study of within-species genetic variation over space and time (phylogeography), the study of how new species originate (speciation) and studies of phenotypic character evolution within and among species. Many of the applications that we discuss involve ecological niche modeling (Box 2), a general methodology that uses species distributional data in combination with GIS maps of environmental variation to predict areas where the species can be expected to occur (over both space and time). Other applications involve using these datasets and associated tools, for example DIVA GIS (<http://diva-gis.org>), to obtain environmental data for each locality and then analyzing those data statistically (e.g. to test for correlations between environmental variables and spatial patterns of phenotypic divergence). Like any other approach, these GIS-based analyses must be used with appropriate caution. Nevertheless, we anticipate that incorporating these data and tools will not only transform how evolutionary biology is done in practice, but will also concomitantly change the way in which evolutionists think about many fundamental topics.

Applications to phylogeography

Phylogeography is a burgeoning subfield within evolutionary biology that addresses the geographic structure of genetic variation within species and its underlying causes [1]. It is also a field in which environmental data seem most necessary, but until recently, they have only rarely been applied. Many of the fundamental questions in phylogeography relate to how physical and climatic variation over space and time shape patterns of genetic divergence. GIS data, in combination with new analytical tools, now allow researchers to address these questions in ways that would not have been possible before.

Corresponding author: Kozak, K.H. (kozak016@umn.edu).

Table 1. Examples of some GIS-based environmental variables that can be applied to evolutionary studies

Environmental variable(s)	Minimum resolution	Website
Climate (temperature, precipitation, seasonality, solar radiation, relative humidity, vapor pressure) ^a	1 km	http://www.fao.org/sd/eidirect/climate/eisp0002.htm http://www.cru.uea.ac.uk/ http://daymet.org http://worldclim.org
Rainfall, cloud cover ^b	0.25°	http://trmm.gsfc.nasa.gov/data_dir/data.html http://edcimswww.cr.usgs.gov/pub/imswelcome/
Vegetation type, per cent tree cover ^b	1 km	http://edcdaac.usgs.gov/glcc/glcc_version1.html#Global http://glcf.umiacs.umd.edu/data/ http://edcimswww.cr.usgs.gov/pub/imswelcome/
Topography (elevation, slope, aspect, grid complexity) ^b	90 m	http://www2.jpl.nasa.gov/srtm/dataprod.htm http://edc.usgs.gov/products/elevation.html http://www.worldwildlife.org/freshwater/hydrosheds.cfm http://edc.usgs.gov/products/elevation/gtopo30/hydro/index.html
Hydrology (streams, drainage basins, flow accumulation, flow direction) ^b	1 km	http://glcf.umiacs.umd.edu/data/
Energy, productivity ^b	1 km	http://glcf.umiacs.umd.edu/data/
Landcover, land use ^b	30 m	http://edcimswww.cr.usgs.gov/pub/imswelcome/ http://glovis.usgs.gov/ http://www.landsat.org/ortho/default.html
Soils (soil type, texture, soil water capacity and pH) ^a	~10 km	http://www.fao.org/AG/agl/agll/ptsoil.stm http://www.daac.ornl.gov/SOILS/soils_collections.html

Data source: ^aspatial interpolation; ^bremote sensing.

Box 1. Obtaining environmental data for evolutionary studies

Perhaps one of the biggest impediments to incorporating GIS-based data into evolutionary studies is that researchers are simply unaware of the data that are available. These datasets include digital maps that quantify spatial variation and temporal variation in temperature, precipitation, humidity, topography, soils, vegetation, energy and productivity (Table 1). All of these can be downloaded free of charge.

Typically, GIS maps of environmental variation are derived using two methods: remote sensing and spatial interpolation. Remote sensing uses aircraft or satellite-borne sensors to detect spatial variation in electromagnetic energy. This variation is then used to generate digital images of environmental variation [56]. For example, spatial variability in energy that is reflected or emitted from the Earth's surface is often used to map variation in vegetation types.

Spatial interpolation is a statistical procedure for estimating data values at unsampled sites between actual data collection locations (e.g. weather stations [57]). For example, the WorldClim dataset [58] includes spatially interpolated GIS maps of 19 climatic variables, including annual means, ranges, variability and extreme values of temperature and precipitation [59]. These maps are derived from monthly values gathered from thousands of weather stations around the world between 1950 and 2000 [58] (~50,000 locations for precipitation and ~25,000 locations for temperature). The data have been spatially interpolated at a fine resolution (~1 km²) [58], making it possible to obtain a reasonable estimate of the climatic conditions at almost any point on the surface of the earth.

Once the GIS maps of interest have been obtained, the next step is to use the geographic coordinate of each locality to plot its location on the environmental map(s) of interest. Relevant locality data can be obtained in the field (using a GPS device), from natural history museums or from the literature. Locations that lack geographic coordinates can often be georeferenced (assigned a latitude and longitude) using the verbal descriptions of the sites and online gazetteers (databases providing the latitude and longitude of towns and other landmarks, such as <http://herpnet.org/Gazetteer/GeoreferResources.htm>). Many natural history museums are also presently georeferencing specimen localities on a large scale and making the information available in a variety of online databases [60].

Environmental barriers to dispersal and gene flow

A recurring theme in phylogeographic studies is that many widely distributed species have been subdivided by historical and contemporary barriers to gene flow. GIS data now allow researchers to investigate the environmental factors that underlie the origin and maintenance of these morphologically cryptic genetic lineages (here and throughout, we refer to a lineage as a set of evolutionarily related and interbreeding populations that might or might not be formally recognized as a distinct species). For example, several recent studies have examined the relationship between the climatic and genetic similarity of populations to address whether parapatric distributions between cryptic sister lineages are associated with selection across climatic gradients. In the European bushcricket *Ephippiger ephippiger* and the western North American black salamander *Aneides flavipunctatus* divergent selection associated with climatic factors seemingly maintains the distributions of such lineages [2,3]. By contrast, the parapatric distributions of many sister lineages of plethodontid salamanders in eastern North America were neither generated nor maintained by climatic gradients [4].

Another near-ubiquitous pattern in phylogeographic studies is the decreasing genetic similarity of populations with increasing geographic distance (i.e. isolation by distance). To date, most studies have used relatively simplistic measures of geographic distance (i.e. simple linear distances). However, the distance separating populations can be a trivial factor in creating and maintaining geographic genetic structure, relative to other features of the environment that limit dispersal (e.g. climatic gradients and changes in topography). This issue is a major topic in the field of landscape genetics, which tends to focus on a smaller spatial scale than does phylogeography [5,6]. Many studies in landscape genetics have used GIS-based data to show that distances based on topography, habitat types and river drainages (for aquatic organisms) provide better predictions of genetic distances between populations than

Box 2. Ecological niche modeling

Ecological niche modeling (also referred to as species distribution modeling) generates a map of the expected distribution of a species using information on the environmental conditions where it is known to occur [61,62]. Niche modeling requires two kinds of data: georeferenced localities and GIS-based maps of the environmental variables (e.g. temperature and precipitation) that are likely to influence the suitability of the environment for that species (Figure 1). A variety of niche modeling methods exist (e.g. Bioclim [59], GARP [63] and Maxent [64]); readers should see Ref. [65] for a comparison of methods and their performance. For a general introduction to ecological niche modeling, readers can also consult the Center for Biodiversity and Conservation's online introduction to ecological niche modeling (<http://biodiversityinformatics.amnh.org/index.php?section=sdm>).

This general approach can be used to address a variety of questions at the interface of evolution and ecology. For example, a niche model for a given species can show that climatically suitable areas occur beyond its known range limits. This pattern is called overprediction. If the overprediction occurs in areas that are geographically adjacent to the known range, one interpretation is that the range limits are not actually set by climatic factors. Instead, they might be set by biotic interactions (e.g. competitors [66], or by nonclimatic factors, such as lack of time for dispersal into climatically suitable areas [67], or rivers and water bodies for terrestrial organisms). By contrast, if the niche model of a species overpredicts into geographically isolated patches of nearby habitat, this pattern suggests that its range might be more extensive than is presently known, or even that closely related, undescribed species might exist in those patches [68].

In the example illustrated in Figure 1, a combination of climatic and biotic factors seem to limit the range of *Plethodon jordani*, a

salamander that is endemic to the Great Smoky Mountains in eastern North America. The niche model for this species identifies climatically suitable locations (shown in orange, yellow and green) that extend beyond its actual geographic distribution (localities shown with blue dots). Climatically inhospitable lowland areas seemingly prevent the dispersal of *P. jordani* to suitable patches of montane habitat to the southwest (A, B), southeast (C), and northeast of its range (D). By contrast, the niche model suggests that this species should be able to disperse to climatically suitable regions (E) that border its southeastern range limit, given the climatic similarity to regions where this species occurs. These sites are occupied by the closely related species *P. metcalfei*, suggesting the possibility that ecological interactions with this species, rather than climate, set the southeastern range limit of *P. jordani* [4].

Niche modeling does have a variety of limitations [61,62,69–71]. Perhaps one of the most important is that it assumes that the modeled distribution of a species reflects its actual environmental tolerances, although this assumption is rarely tested (but see Refs [72,73]). A species might be able to tolerate physiologically a broader range of conditions than would be suggested by the extent of its predicted geographic distribution (e.g. when its range is limited by a competitor [66,70] or is not at equilibrium with the current climate as a result of dispersal limitations [67,70]). Underestimating the environmental tolerances of a species could lead to a variety of errors, such as underestimating the potential geographic distribution of a species and assuming that certain environmental factors limit its distribution that do not [70,74]. Some additional limitations and potential problems that are relevant to niche modeling but apply to GIS data in general are discussed in Box 3.

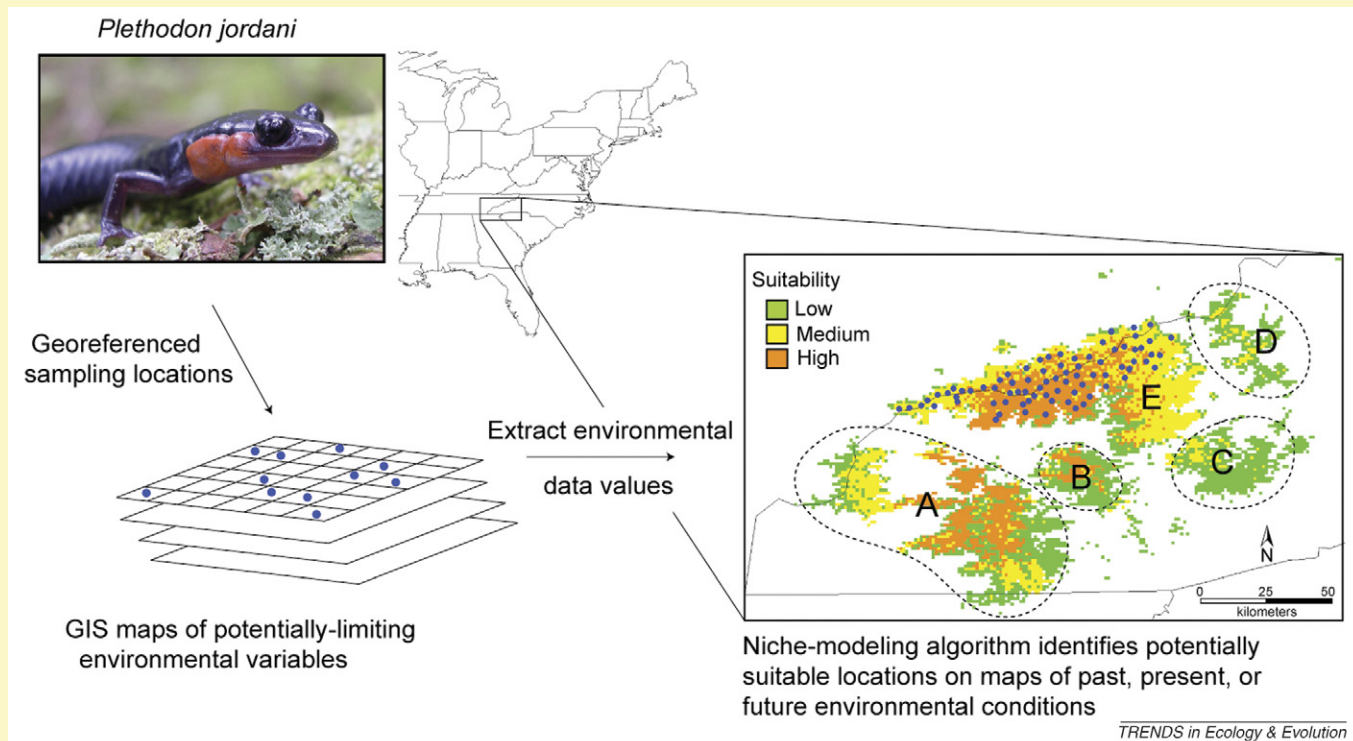


Figure 1. General example outlining the procedure for modeling the predicted geographic distribution of a species using ecological niche modeling.

do simple geographic distances (e.g., see Refs [7–10]). Surprisingly, these approaches have generally not been used in phylogeographic studies (but see Ref. [11]), but the availability of global coverages for many environmental variables (Box 1) and existing spatial tools [12–14] now make it feasible for phylogeographers to explore a large

variety of potential environmental correlates of phylogeographic structure at continental scales.

Impact of glacial refugia on phylogeographic structure
Assessing the importance of glacial cycles (and the associated expansions and contractions of habitats) on patterns

of genetic variation is a major theme in phylogeography. In the past, phylogeographic studies relied almost entirely on genetic data to make inferences about the locations of glacial refugia and their effects on organisms. Now, GIS-based methods make it possible to model the spatial patterns of climatic suitability for organisms based on their current distributions and to project these conditions onto mapped estimates of historical climates [15,16]. This development allows researchers to visualize and test hypotheses based on the potential distribution of organisms over time.

A key ingredient in these analyses is a GIS map of estimated climatic variation at a relevant point in the past. Previously, these estimates were based on palynological data, which can provide information on both paleoclimates and the distributions of favorable habitats for species (e.g. see Ref. [15]). However, palynological data are lacking for many geographic regions. Now, global estimates of paleoclimatic variation are becoming available through the Paleoclimatic Modeling Intercomparison Project (PMIP: <http://www.ncdc.noaa.gov/paleo/model.html>). Thus, it has become possible to test hypotheses regarding the refugial dynamics of many organisms and regions around the world (e.g. see Refs [17–19]). For example, Knowles and colleagues [20] recently used paleoclimatic data to model the distribution of suitable habitats of high montane populations of the North American grasshopper *Melanoplus marshalli* at the Last Glacial Maximum (LGM). They found that the paleodistribution models corroborated the genetic data, which had suggested the separation of this species into subpopulations corresponding to Pleistocene refugia.

Another major focus of phylogeographic research involves comparing patterns of genetic variation between species with partially or fully overlapping geographic distributions (comparative phylogeography). GIS-based paleoclimatic maps and models of the past distribution of species now allow researchers to test whether species with congruent phylogeographic patterns responded in a similar fashion to past climatic changes. By modeling the paleodistributions of species endemic to the Pacific Northwest mesic ecosystem, Carstens and Richards [21] showed that congruent phylogeographic patterns in three amphibian species reflect concerted responses of these lineages to historical changes in the distribution of mesic climatic conditions.

Taken together, these studies demonstrate how GIS data and paleodistribution models can provide an independent test of the predictions of genetic models. Nonetheless, it is important to bear in mind that the accuracy of these spatial predictions depends on whether the climatic tolerances of species are conserved over the timescale for which their geographic distributions are being modeled and that the paleoclimate maps accurately portray spatial variation in climate. Recent applications of niche modeling to contemporary and fossil locality data found that both plants and animals do show conservatism in their climatic niche characteristics over the past 20 000 years [22,23]. However, the sensitivity of species distribution models to different paleoclimatic reconstructions has been poorly explored and is an area in need of future research.

Applications to speciation research

Understanding the origin of species is one of the dominant research agendas in evolutionary biology [24]. Incorporating GIS-based tools and data has the potential to have a major impact on several persistent questions in speciation research, including: (i) what role does ecology play in speciation? (ii) what is the most common geographic mode of speciation? and (iii) what is the long-term fate of hybridizing lineages?

What role does ecology play in speciation?

GIS-based models of the potential geographic distribution of sister species are providing fresh insights into the role of ecological factors in driving allopatric speciation [4,25], a question that has largely been ignored until recently [26]. Typically, allopatry is explained by invoking dramatic geographic barriers to dispersal (e.g. rivers, oceans and deserts). However, recent GIS-based analyses suggest that allopatry can sometimes be associated with specialization to relatively subtle differences in ecological conditions, in which this specialization is maintained over evolutionary timescales (niche conservatism) [25] and limits dispersal between populations. For example, GIS-based climatic data suggest that allopatric speciation in many montane salamanders in eastern North America is related to their inability to tolerate conditions in lowland habitats [4]. Yet, both highlands and lowlands in this region consist predominantly of mesic, temperate deciduous woodlands. Thus, allopatric speciation in these species is seemingly associated with their specialization to climatic conditions at higher elevations, even though the high and low elevation habitats are generally quite similar.

Although recent studies have demonstrated that niche conservatism can play an important role in promoting speciation by limiting dispersal between populations, adaptation to different climatic conditions in allopatric or parapatric populations might also play an important role in speciation by driving phenotypic divergence and accelerating the evolution of reproductive isolation [27,28]. GIS-based models have identified pairs of sister species that seem to have originated in response to divergent selection across environmental gradients [3,29,30]. However, it is important to caution that inferring a role for adaptive divergence in speciation also requires data on phenotypic traits relevant to reproductive isolation and not just climatic data. To this end, GIS-based models seem to be a promising tool for identifying sister species that inhabit divergent climatic regimes. These species pairs can then be examined for adaptive divergence in phenotypic traits that contribute to reproductive isolation.

What is the most common geographic mode of speciation?

The typical approach for addressing this question involves comparing the current geographic overlap of sister species inferred from a phylogeny and evaluating whether their ranges are allopatric (separate), parapatric (abutting) or sympatric (overlapping) [31]. However, some authors have questioned the utility of this approach given the possibility that ranges can shift extensively after speciation and thereby obscure the original patterns of overlap [32].

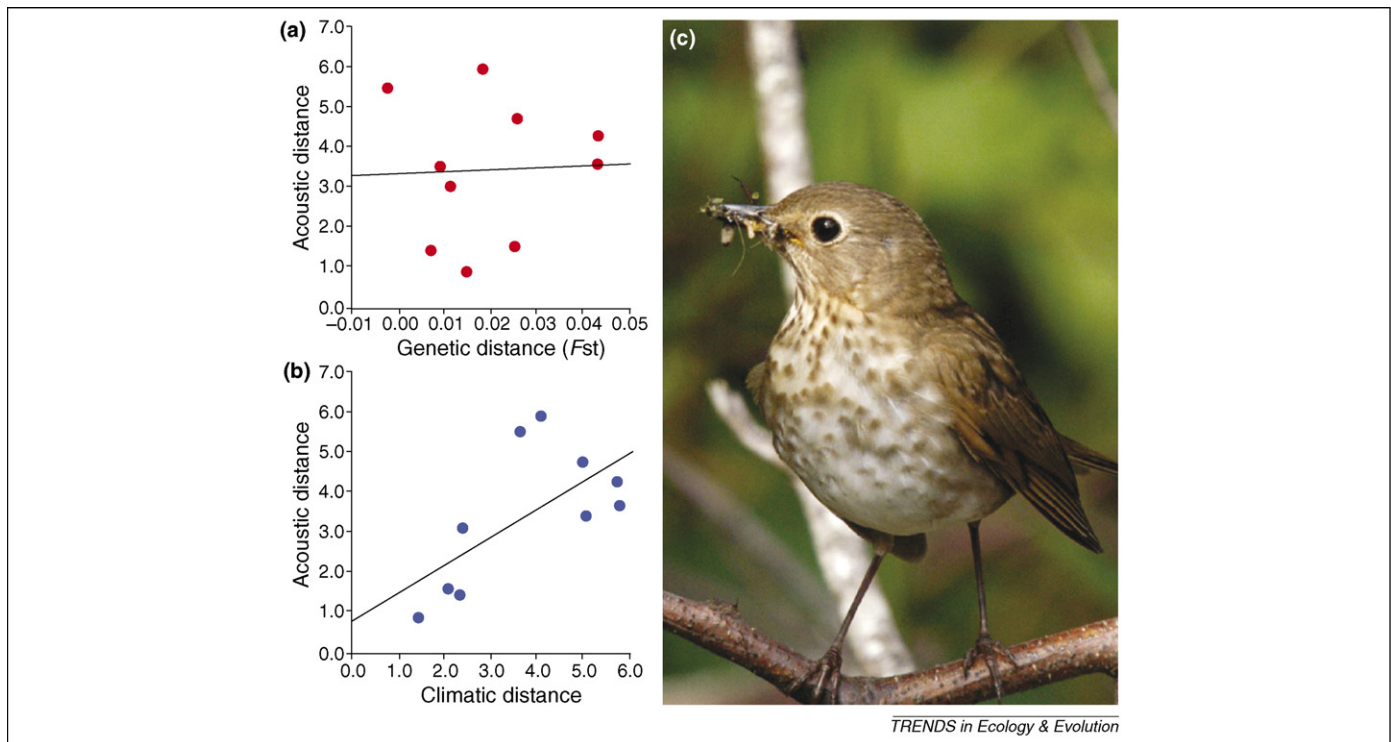


Figure 1. Combining (a) genetic and (b) GIS-based climatic data to understand geographic patterns of song evolution in (c) Swainson's thrush *Catharus ustulatus*. Ruegg *et al.* [47] assessed the association between the song divergence (acoustic distance) of populations and their genetic and climatic divergence. No association was found between the acoustic and genetic divergence of populations. By contrast, populations that inhabit different climates tend to show greater acoustic divergence than those that inhabit similar climates. Thus, natural selection rather than genetic drift seemingly drives patterns of song evolution at the population level.

GIS-based analyses of current and past climates can be used to gain insights into the stability of these overlap patterns by modeling the distribution of areas of suitable (or unsuitable) habitat over time. For example, if these analyses show that a region of unsuitable climatic conditions currently separating the ranges of two allopatric species has persisted over long timescales (e.g. as long as the estimated age of the species pair), it can be inferred with much greater confidence that allopatric speciation is the geographic mode.

Perhaps the most challenging geographic mode to demonstrate is parapatric speciation. When parapatric distributions of sister species are observed, it is difficult to determine whether this occurs because of primary contact (species evolved *in situ*, parapatrically) or secondary contact (species evolved in allopatry and extended their ranges to abut more recently) [24]. However, if species range limits are determined largely by climate and their climatic tolerances remain similar over time, it should be possible to distinguish primary and secondary contact by projecting the potential ranges of sister species backwards in time, to determine how long they are likely to have been in contact. For example, if the present-day region of contact was previously unsuitable for both species over most of the time period in which they are likely to have speciated, primary contact (and hence, parapatric speciation) seems unlikely.

Recent work on birds demonstrates the utility of this general approach for assessing historical range dynamics of sibling species. Ruegg and colleagues [33] modeled the distribution of suitable climates for two parapatric sister lineages of Swainson's thrush *Catharus ustulatus* based on

the current distribution of these lineages. They then projected those conditions onto the map of climates for the Last Glacial Maximum and inferred that the two lineages came into secondary contact after allopatric isolation in climatically distinct Pleistocene refugia.

Of course, modeling the geographic distributions of sister species backwards in time also has limitations and must be done with some caution. The accuracy of paleodistribution models of sister species will ultimately depend on the rate at which climatic niche similarity decays over time. Although recent studies have shown considerable climatic niche conservatism across large phylogenies [25,34–37], this pattern is not found in all taxa [38–40]. Thus, it is important to determine whether the climatic niches of species are generally conserved over the timescale under consideration before using paleodistribution models to infer geographic modes of speciation. A variety of spatial, statistical and phylogenetic software packages provide the analytical tools to test this hypothesis rigorously (e.g. tests of phylogenetic versus climatic distance [41,42] and null models of climatic niche overlap [43]).

What is the long-term fate of hybridizing lineages?

GIS data can even provide new insights into whether species in hybrid zones will remain separate or will fuse to form a single species. Using ecological niche modeling, Swenson [44] found that in North America, hybrid zones between eastern and western species of birds are geographically clustered and that this clustering is associated with a sharp temperature gradient corresponding to the Great-Plains–Rocky-Mountain ecotone. This pattern

suggests that these eastern and western species have adapted to different climatic conditions and that their range borders are being maintained by selection against hybrids (for a related example, see Ref. [45]). Thus, it seems unlikely that any of these species pairs will fuse to form a single species, because such a process would be impeded by the adaptive differences of the two species (see also Ref. [46]).

Applications to character evolution within and between species

The environmental conditions that a species experiences can strongly influence the evolution of many aspects of its phenotype. GIS-based environmental data can greatly facilitate analyses of correlations between evolutionary changes in species traits and the environments that those species experience. This approach can be used to study both macroevolutionary changes among species and the evolution of geographic variation within species.

For example, Ruegg and colleagues [47] used GIS-derived climatic data to examine song evolution among populations of Swainson's thrush *Catharus ustulatus* and its relationship to genetic and climatic differentiation. By quantifying geographic variation in temperature and precipitation across the range of this species, they found that the acoustic divergence of populations was not correlated with their genetic divergence (as expected under a model of genetic drift) but rather with geographic differences in climate (Figure 1). This study demonstrates the potential of GIS-based data to help to distinguish among alternative causes of intraspecific phenotypic evolution (i.e. genetic drift versus natural selection). However, it also demonstrates that a positive association between a trait and an environmental variable does not necessarily imply causation. The authors found that song evolution among populations does not appear to be related directly to climate but rather to the effects of climate in determining forest type (i.e. rainforest versus coniferous forest). Each type of forest creates a different selective environment for optimal sound transmission, because the rainforest has a dense understory and the coniferous forest has a sparse understory. Thus, the song types seem to be more directly influenced by forest type than by climate alone.

Beyond correlating climatic data with patterns of phenotypic variation, there are other applications of GIS data to studies of character evolution that might be less obvious. For example, GIS-derived climatic data can potentially be used to identify misleading phylogenetic reconstructions of character evolution [48]. In marsupial frogs (Hemiphractidae: *Cryptobatrachus*, *Flectonotus*, *Gastrotheca*, *Hemiphractus*, *Stefania*), different methods of ancestral reconstruction lead to two very different interpretations of life history evolution. Under one interpretation, direct development is ancestral for the family, and the larval (tadpole) stage, remarkably, appears to have re-evolved. This reconstruction is supported by simple parsimony and by a likelihood model with an equal rate for gains and losses of direct development across the phylogeny (one-rate model). Under the alternative reconstruction, there are multiple origins of direct development, and there is no re-evolution of the tadpole stage. This second scenario is

supported by maximum likelihood reconstruction when allowing for different rates of gain and loss of direct development (two-rate model). The two-rate model has a better statistical fit to the data than does the one-rate model and seems to make more sense intuitively (i.e. it does not require the re-appearance of the lost tadpole stage). However, when ancestral climatic regimes are reconstructed on the phylogeny, the two-rate model implies that the tadpole stage was maintained for millions of years under climatic conditions in which no marsupial frog species with a normal tadpole stage occurs today [48]. Thus, the GIS-based climatic data actually suggest that this second scenario is less plausible, and the surprising hypothesis that the tadpole stage re-evolved seems more likely. This example illustrates an unexpected way in which GIS data might be used to inform analyses of character evolution.

Prospects for applied evolutionary research

So far, we have reviewed how GIS are providing new insights into evolutionary biology and provided examples from three areas of basic research. However, the merging of data and approaches from geographic-information science and evolutionary biology should also catalyze further integration of evolutionary perspectives into applied research.

Box 3. Limitations of GIS data in evolutionary studies

GIS data and new spatial analytical tools can provide many new insights into how environmental variation in space and time drives evolutionary patterns and processes. Nonetheless, some caution is needed at all stages of these analyses, from obtaining the basic data to the statistical interpretation of the results.

First, researchers must consider the extent to which data resolution and quality might impact their results [75]. For example, climatic data layers derived from spatial interpolation can have greater error in regions with extreme topographic heterogeneity and where the density of weather stations is sparse [57]. Likewise, maps that are derived from the interpolation of points that are evenly spaced will have greater accuracy than those generated from points that are clustered [2,75]. Furthermore, when assessing the past distribution of suitable climates for a species, it is important that researchers consider the extent to which their conclusions might be impacted by variation among alternative paleoclimatic models and different methods for modeling the potential geographic distribution of species [76–78].

Many environmental variables are tightly correlated making some of them effectively redundant. Thus, using all the variables in a database might violate the assumptions of many statistical tests, potentially leading to spurious results. Researchers can use two straightforward approaches to avoid this potential pitfall. The first approach is to test the correlations among all the variables for the localities of interest and to select a subset of variables that seem least correlated and most relevant for the question at hand. Alternatively, a data-reduction method, such as principal-components analysis can be used to generate linear combinations of the original variables that are orthogonal (i.e. independent) of each other (e.g., see Refs [3,4,79]).

It is also important to consider the extent to which available sampling locations might represent a biased representation of the true spatial distribution of a species. For example, collecting efforts for some species might be clustered near towns and cities and along roadways. Such nonrandom sampling could lead to biases in the estimated environmental distribution of a species. Thus, it might be necessary to subsample available locations to remove or to reduce such potential biases [71].

With the availability of global GIS data on environmental variation, researchers are now beginning to address the general question of whether introduced species have evolved to become more invasive. For example, it is possible to construct a GIS-based model of the distribution of an invasive species based on its native distributions and to test the ability of the model to predict the extent of its range in the area to which it has been introduced [49]. If the native distribution fails to predict the extent of the climatic distribution of the species in its introduced range, the evolution of increased climatic tolerances (relative to their native range) might be involved in their geographic expansion from the location(s) where it was first introduced [50]. This general approach is also shedding light on the extent to which adaptation to novel environmental conditions might facilitate the geographic expansion of domesticated species beyond the ranges of their wild progenitors [51]. Furthermore, new spatial analyses for testing whether natural selection at the molecular level is associated with environmental variation [52] show much promise for identifying candidate genes that might promote the geographic spread of invasive and domesticated species.

GIS-data could also be used in an evolutionary framework to address how species are likely to respond to global climate change (e.g. by shifting ranges, adapting or becoming extinct). Recent studies have used GIS data and methods to estimate how species will respond to climate change, but assuming no evolution in their climatic tolerances [53–55]. Alternatively, combining GIS-based climatic data, phylogenies and comparative methods for estimating rates of evolutionary change could be used to determine whether the rates at which species have changed in their climatic niches in the past can keep pace with estimated rates of human-induced climate change.

Conclusions

Given the increasing ease with which environmental data can be obtained and analyzed for many populations, species, clades and regions across the world, we envision that GIS technology will become increasingly integrated into evolutionary biology. By allowing evolutionary biologists to obtain data on environmental variation that would have previously been impossible or impractical to collect, there seems little doubt that GIS will continue to invigorate existing research areas and open the door to new research programs focused on both fundamental and applied research questions in evolutionary biology.

Acknowledgements

Our research was supported by NSF grants DBI-0434728 (K.H.K.), DEB-0416152 (C.H.G.) and EF 0334923 (J.J.W.), and a NASA new investigator award to C.H.G. For comments on the manuscript, we thank S. B. Baines, D. Moen and other members of the Wiens laboratory group, A. Guisan, and four anonymous reviewers. We thank M. Chatfield and M. Yip for kindly providing the photographs of *Plethodon jordani* and *Catharus ustulatus* that appear in Box 1 and Figure 1, respectively.

References

- 1 Avise, J.C. (2000) *Phylogeography: The History and Formation of Species*, Harvard University Press
- 2 Ritchie, M.G. *et al.* (2001) Mitochondrial DNA variation and GIS analysis confirm a secondary origin of geographical variation in the bushcricket *Ephippiger ephippiger* (Orthoptera: Tettigoniidae), and resurrect two subspecies. *Mol. Ecol.* 10, 603–611
- 3 Rissler, L.J. and Apodaca, J.J. Adding more ecology into species delimitation: ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). *Syst. Biol.* 56, 924–942
- 4 Kozak, K.H. and Wiens, J.J. (2006) Does niche conservatism promote speciation? A case study in North American salamanders. *Evolution Int. J. Org. Evolution* 60, 2604–2621
- 5 Manel, S. *et al.* (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol. Evol.* 18, 189–197
- 6 Storfer, A. *et al.* (2007) Putting the 'landscape' in landscape genetics. *Heredity* 98, 128–142
- 7 Funk, W.C. *et al.* (2005) Population structure of Columbia spotted frogs (*Rana luteiventris*) is strongly affected by the landscape. *Mol. Ecol.* 14, 483–496
- 8 Spear, S.F. *et al.* (2005) Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). *Mol. Ecol.* 14, 2553–2564
- 9 Vignieri, S.N. (2005) Streams over mountains: influence of riparian connectivity on gene flow in the Pacific jumping mouse (*Zapus trinotatus*). *Mol. Ecol.* 14, 1925–1937
- 10 Pieltney, S.B. *et al.* (1998) Local genetic structure in red grouse (*Lagopus lagopus scoticus*): evidence from microsatellite DNA markers. *Mol. Ecol.* 7, 1645–1654
- 11 Kozak, K.H. *et al.* (2006) Gene lineages and eastern North American palaeodrainage basins: phylogeography and speciation in salamanders of the *Eurycea bislineata* species complex. *Mol. Ecol.* 15, 191–207
- 12 Guillot, G. *et al.* (2005) Geneland: a computer package for landscape genetics. *Mol. Ecol. Notes* 5, 712–715
- 13 Ray, N. (2005) PATHMATRIX: a geographical information system tool to compute effective distances among samples. *Mol. Ecol. Notes* 5, 177–180
- 14 Drielsma, M. *et al.* (2007) The spatial links tool: automated mapping of habitat linkages in variegated landscapes. *Ecol. Model.* 200, 403–411
- 15 Hugall, A. *et al.* (2002) Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). *Proc. Natl. Acad. Sci. USA* 99, 6112–6117
- 16 Peterson, A.T. *et al.* (2004) Reconstructing the Pleistocene geography of *Aphelocoma jays*. *Divers. Distrib.* 10, 237–246
- 17 Bonaccorso, E. *et al.* (2006) Pleistocene fragmentation of Amazon species' ranges. *Divers. Distrib.* 12, 157–164
- 18 Graham, C.H. *et al.* (2006) Habitat history improves prediction of biodiversity in a rainforest fauna. *Proc. Natl. Acad. Sci. USA* 103, 632–636
- 19 Waltari, E. *et al.* (2007) Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *PLoS ONE* 2, e563
- 20 Knowles, L.L. *et al.* (2007) Coupling genetic and ecological-niche models to examine how past population distributions contribute to divergence. *Curr. Biol.* 17, 940–946
- 21 Carstens, B.C. and Richards, C.L. (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution Int. J. Org. Evolution* 6, 1439–1454
- 22 Martínez-Meyer, E. *et al.* (2004) Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinction and climate change projections for biodiversity. *Glob. Ecol. Biogeogr.* 13, 1013–1044
- 23 Martínez-Meyer, E. and Peterson, A.T. (2006) Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-Recent transition. *J. Biogeogr.* 33, 1779–1789
- 24 Coyne, J.A. and Orr, H.A. (2004) *Speciation*, Sinauer Associates
- 25 Peterson, A.T. *et al.* (1999) Conservatism of ecological niches in evolutionary time. *Science* 285, 1265–1267
- 26 Wiens, J.J. (2004) Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. *Evolution Int. J. Org. Evolution* 58, 193–197
- 27 Gavrillets, S. (2003) Perspective: Models of speciation: what have we learned in 40 years? *Evolution Int. J. Org. Evolution* 57, 2197–2215
- 28 Funk, D.J. *et al.* (2006) Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proc. Natl. Acad. Sci. USA* 103, 3209–3213

- 29 Graham, C.H. *et al.* (2004) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution Int. J. Org. Evolution* 58, 1781–1793
- 30 Kozak, K.H. and Wiens, J.J. (2007) Climatic zonation drives latitudinal variation in speciation mechanisms. *Proc. Biol. Sci.* 274, 2995–3003
- 31 Barraclough, T.G. and Vogler, A.P. (2000) Detecting the geographical pattern of speciation from species-level phylogenies. *Am. Nat.* 155, 419–434
- 32 Losos, J.B. and Glor, R.E. (2003) Phylogenetic comparative methods and the geography of speciation. *Trends Ecol. Evol.* 18, 220–227
- 33 Ruegg, K.C. *et al.* (2006) Climate change and the origin of migratory pathways in the Swainson's thrush, *Catharus ustulatus*. *J. Biogeogr.* 33, 1172–1182
- 34 Prinzing, A. *et al.* (2001) The niche of higher plants: evidence for phylogenetic niche conservatism. *Proc. Biol. Sci.* 268, 2383–2389
- 35 Ackerly, D.D. (2004) Adaptation, niche conservatism, and convergence: comparative studies of leaf evolution in the California chaparral. *Am. Nat.* 163, 654–671
- 36 Wiens, J.J. *et al.* (2006) Evolutionary and ecological causes of the latitudinal diversity gradient in hylid frogs: treefrog trees unearth the roots of high tropical diversity. *Am. Nat.* 168, 579–596
- 37 Smith, S.A. *et al.* (2005) Replicate patterns of species richness, historical biogeography, and phylogeny in Holarctic treefrogs. *Evolution Int. J. Org. Evolution* 59, 2433–2450
- 38 Rice, N.H. *et al.* (2003) Ecological niche differentiation in the *Aphelocoma* jays: a phylogenetic perspective. *Biol. J. Linn. Soc.* 80, 369–383
- 39 Knouft, J.H. *et al.* (2006) Phylogenetic analysis of the evolution of the niche in lizards of the *Anolis sagrei* group. *Ecology* 87, S29–S38
- 40 Peterson, A.T. and Holt, R.D. (2003) Niche differentiation in Mexican birds: using point occurrences to detect ecological innovation. *Ecol. Lett.* 6, 774–782
- 41 Rosenberg, M.S. (2001) Pattern analysis, spatial statistics, and geographic exegesis. Version 1.0. Department of Biology, Arizona State University. <http://www.passagesoftware.net>
- 42 Pagel, M. and Meade, A. (2007) BayesTraits. School of Biological Sciences, University of Reading. <http://evolution.reading.ac.uk/SoftwareMain.html>
- 43 Stockman, A.K. *et al.* (2008) D-NOVL: a program to simulate overlap between two niche-based distribution models. *Mol. Ecol. Notes* 8, 290–294
- 44 Swenson, N.G. (2006) GIS-based niche models reveal unifying climatic mechanisms that maintain the location of avian hybrid zones in a North American suture zone. *J. Evol. Biol.* 19, 717–725
- 45 Cicero, C. (2004) Barriers to sympatry between avian sibling species (Paridae: *Baeolophus*) in local secondary contact. *Evolution Int. J. Org. Evolution* 58, 1573–1587
- 46 Stockman, A.K. and Bond, J.E. (2007) Delimiting cohesion species: extreme population structuring and the role of ecological interchangeability. *Mol. Ecol.* 16, 3374–3392
- 47 Ruegg, K. *et al.* (2006) Divergence in mating signals correlates with ecological variation in the migratory songbird, Swainson's thrush (*Catharus ustulatus*). *Mol. Ecol.* 15, 3147–3156
- 48 Wiens, J.J. *et al.* (2007) Loss and re-evolution of complex life cycles in marsupial frogs: can ancestral reconstruction mislead? *Evolution Int. J. Org. Evolution* 61, 1886–1899
- 49 Peterson, A.T. (2003) Predicting the geography of species' invasions via ecological niche modeling. *Q. Rev. Biol.* 79, 419–433
- 50 Broennimann, O. *et al.* (2007) Evidence of climatic niche shift during biological invasion. *Ecol. Lett.* 10, 701–709
- 51 Miller, A.J. and Knouft, J.H. (2006) GIS-based characterization of the geographic distributions of wild and cultivated populations of the Mesoamerican fruit tree *Spondias purpurea*. *Am. J. Bot.* 93, 1757–1767
- 52 Joost, S. *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Mol. Ecol.* 16, 3955–3969
- 53 Peterson, A.T. *et al.* (2002) Future projections for Mexican faunas under global climate change scenarios. *Nature* 416, 626–629
- 54 Williams, S.E. *et al.* (2003) Climate change in Australian tropical rainforests: an impending environmental catastrophe. *Proc. Biol. Sci.* 270, 1887–1892
- 55 Thomas, C.D. *et al.* (2004) Extinction risk from climate change. *Nature* 427, 145–148
- 56 Turner, W. *et al.* (2003) Remote sensing for biodiversity science and conservation. *Trends Ecol. Evol.* 18, 306–314
- 57 Hutchinson, M.F. (1989) A new objective method for spatial interpolation of meteorological variables from irregular networks applied to the estimation of monthly mean solar radiation, temperature, precipitation and windrun. *CSIRO Division of Water Resources Tech. Memo* 89, 95–104
- 58 Hijmans, R.J. *et al.* (2005) Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978
- 59 Nix, H.A. (1986) A biogeographic analysis of Australian elapid snakes. In *Atlas of Elapid Snakes of Australia* (Australian Flora and Fauna Series Number 7) (Longmore, R., ed.), pp. 4–15, Australian Government Publishing Service
- 60 Graham, C.H. *et al.* (2004) New developments in museum-based informatics and applications in biodiversity analysis. *Trends Ecol. Evol.* 19, 497–503
- 61 Guisan, A. and Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecol. Model.* 135, 147–186
- 62 Guisan, A. and Thuiller, W. (2005) Predicting species distributions: offering more than simple habitat models. *Ecol. Lett.* 8, 993–1009
- 63 Stockwell, D. and Peters, D. (1999) The GARP modeling system: problems and solutions to automated spatial prediction. *Int. J. Geogr. Inf. Sci.* 13, 143–158
- 64 Phillips, S.J. *et al.* (2006) Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259
- 65 Elith, J. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151
- 66 Anderson, R.P. *et al.* (2002) Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos* 98, 3–16
- 67 Svenning, J.C. *et al.* (2006) Range filling in European trees. *J. Biogeogr.* 33, 2018–2021
- 68 Raxworthy, C.J. *et al.* (2003) Predicting distributions of known and unknown reptile species in Madagascar. *Nature* 426, 837–841
- 69 Araujo, M.B. and Guisan, A. (2006) Five (or so) challenges for species distribution modeling. *J. Biogeogr.* 33, 1677–1688
- 70 Soberon, J. and Peterson, A.T. (2005) Interpretations of models of fundamental ecological niches and species' distributional areas. *Biodiv. Inform.* 2, 1–10
- 71 Pearson, R.G. and Dawson, T.P. (2003) Predicting the impacts of climate change on the distributions of species: are bioclimate envelope models useful? *Glob. Ecol. Biogeogr.* 5, 361–371
- 72 Kearney, M. and Porter, W.P. (2004) Mapping the fundamental niche: physiology, climate, and the distribution of a nocturnal lizard. *Ecology* 85, 3119–3131
- 73 Wright, J.W. *et al.* (2006) Experimental verification of ecological niche modeling in a heterogeneous environment. *Ecology* 87, 2433–2439
- 74 Anderson, R.P. *et al.* (2002) Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Glob. Ecol. Biogeogr.* 11, 131–141
- 75 Kidd, D.M. and Ritchie, M.G. (2006) Phylogeographic information systems: putting the geography into phylogeography. *J. Biogeogr.* 33, 1851–1865
- 76 Hijmans, R.J. and Graham, C.H. (2006) The ability of climate envelopes to predict the effect of climate change on species distributions. *Glob. Change Biol.* 12, 2272–2281
- 77 Heikkinen, R.K. *et al.* (2006) Methods and uncertainties in bioclimatic envelope modeling under climate change. *Prog. Phys. Geogr.* 30, 751–777
- 78 Araujo, M.B. and New, M. (2007) Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22, 42–47
- 79 Richards, C.L. and Knowles, L.L. (2007) Tests of phenotypic and genetic concordance and their application to the conservation of Panamanian golden frogs (Anura, Bufonidae). *Mol. Ecol.* 16, 3119–3133