Materials and Methods S1, Supporting information:

1. Species distribution modeling For each species of leaf beetle and host plant we used MaxEnt 3.3.3a to predict areas of suitable climate conditions for three time periods: at present, at the Last Glacial Maximum (LGM), and during the Last Interglacial (LIG). The models were performed with 10,000 randomly generated background points and, generally, with the recommended default settings (Phillips and Dudík 2008), except for the feature selection. In MaxEnt, the choice of features determines the complexity of the modeled relationship between species occurrence and their environment. We tested which feature scheme was most appropriate by examining the response curves (data not shown) for each variable and identifying the ability of the model to capture biologically meaningful trends and its transferability (preferring less complex response curves). The latter attribute is important to insure a good projection of the models for current conditions in other environments and across time. Based on the preliminary analyses of different feature selections, we showed that the use of hinge feature was more appropriate than the quadratic and linear features normally used automatically, when using small sample sizes such as those analyzed here. Hence, all subsequent models were conducted with only hinge features.

Recent studies have highlighted the following as being important for generating meaningful predictions of potentially suitable climatic conditions: (i) model complexity, that can have an important impact on model performance (Warren and Seifert 2011); (ii) multicolinearity among explanatory variables; and (iii) selecting ecological meaningful variables (e.g. Elith et al. 2011). The best variables are expected to be predictors known to play an important role for the species survival (proximal variables). Unfortunately, environmental factors influencing the viability of *Goniomena* species have not been studied in sufficient details. MaxEnt generally down-weight variables that are contributing less to the model performance. Therefore, a preliminary variable selection would normally be unnecessary; nonetheless, an appropriate feature selection is recommended (Elith et al. 2010). Furthermore, studies have shown that the variable composition counts (e.g. Rödder et al. 2009). Here, we selected a small set of bioclimatically meaningful variables by exploring the correlation among the 19 WorldClim variables, and the explanatory power of the variables. We used three different variable selection approaches, and selected: (i) six non-correlated variables that we assumed to be ecologically important, and among these (ii) those contributing most to the model after successive reduction steps, (iii) those performing best according to the AIC scores (computed in ENMtools; Warren et al. 2010).

Many of the variables were highly correlated. Our first selection aimed at identifying variables that catch maximum information without being redundant. For each species, it resulted in a selection of six variables that were not significantly correlated (i.e., pair-wise Spearman correlation score below the general rule of thumb of 0.7; Dormann et al. 2012). In the selection process, variables potentially linked to a cold-adaptation of the leaf beetles were favored, i.e proximal variables, while we excluded annual distal ones (e.g., annual mean temperature and annual precipitation). The six selected variables were all used in a model identified hereafter as the “6var-model” for each species.

From the “6var-model” variable composition we conducted two additional selection procedures aiming at identifying the variable(s) contributing the most to the model while also offering the best compromise between model performance and it’s transferability with the chosen variables. A standard method to assess the performance of predictions in discriminating suitable from unsuitable areas for a species is based on the metric of the area under the receiver operating characteristic (ROC) curve (AUC). This statistic, also provided by MaxEnt itself, is however sorely criticized (Lobo et al. 2008; Peterson et al. 2008) and is certainly not recommended when working on “presence-only” datasets (Boyce et al 2002), in particular because of the impossibility in that case to estimate the best score (upper limit) of

AUC values (Raes and ter Steege 2008; optimal suitability not equal to 1). In this context, variable selection based on the AUC scores were performed with the jackknife and cross-validation (ten-fold) options using logistic output format while also considering the general properties of variables influencing the models (eg., test Gain). Among these properties, we focused in particular on their transferability, calculated by subtracting the test GAIN at the training GAIN ($GAIN_{diff}$), as the models need to be projected in different time periods. The best transferability is evidenced by the smaller loss in gain from training to test. Therefore, from the group of six variables, we proceeded to successive steps of removing one variable at a time, excluding the less important one, until no significant loss in performance was observed. As an objective alternative statistic to the AUC, we used the Akaike Information Criterion (AIC) shown to be a good selection model criterion (Johnson and Omland 2004) conducted on raw predictions issued from MaxEnt, as proposed with the program ENMTools (Warren et al. 2010). Each of the 63 possible combinations from the six uncorrelated variables have been used to produce a prediction from which the AIC can be calculated (data not shown) and the variable combination with the best score was selected.

Initially, the logistic output estimates the probability of presence and ranges from zero to one (Phillips and Dudík 2008). It was converted to a binary prediction of potential presence (suitable climate) or absence (unsuitable climate) using two different thresholds: the maximum training sensitivity plus specificity (restrictive) threshold as recommended by Jiménez-Valverde & Lobo (2007) and the minimum presence (less restrictive) threshold. The latter focuses on a zero omission error and offers a conservative alternative to the former. These thresholds were selected to represent a strict and a less restrictive approach, respectively, and thus to overcome potential biases in the models. Notably, climatic models for the LGM generally tend to underestimate the dry conditions suggested by pollen-based estimates (Ramstein et al 2007), which could have led to potential overestimation of the species presence. On the contrary, lack of knowledge of the total current range of the beetle, might have lead us to underestimate the climatic tolerance of the species and thus its potential LGM range. In addition, the performances of CCSM and MIROC models to reconstruct LGM climate has been shown to allow reliable reconstruction of past species distributions, as evaluated using fossil records (Fuentes-Hurtado et al 2016). The ability of MaxEnt models to predict the present observed distribution of the species was evaluated using the area under the receiver operating characteristics curve (AUC; Phillips et al. 2006) and the test and training gains. These values were computed by 10-fold cross validation. However, only the initial (unconverted) probabilities of presence were used to define models for PhyloGeoSim analyses, while presence/absence data were used to visually compare SDM outputs inferred with different variable combinations and for the three time periods. This was done to avoid complications associated with choosing a threshold when designing our historical hypotheses.

2. Coalescence simulations and hypotheses testing.

2.1 Spatially explicit coalescence simulations

A grid of 60 by 80 cells was superimposed on each map of present or past species distribution inferred with MaxEnt, and the presence/absence of the species was determined for each grid cell based on the MaxEnt estimates (see below); these grids, representing estimated potential species ranges at three different time periods, were fed to the program PhyloGeoSim, to define detailed spatially explicit models of population evolution, for the purpose of running coalescence simulations. In this coalescence model, each cell is considered a distinct population (defining main geographic regions as groups of contiguous cells). At each generation, going backward in time, these populations can host coalescence events between two or more gene copies and/or exchange gene copies with neighbor populations. In a first step, preliminary forward simulations were conducted to estimate backward simulation parameters (migration rates and effective population sizes for each generation backward in time). Different parameter values were tested in the preliminary forward simulations for the following parameters: (i) the reproduction rate within a cell (in the absence of *a priori* knowledge, 2, 5 and 10), (ii) one or more matrices of maximal effective population sizes, (iii) the migration rate between adjacent cells on the grid, and finally, (iv) the number of generations in the past at which these matrices occurred. To define the maximum effective population size *Ne* of each cell, we considered the presence probability values given in the MaxEnt estimates. Cells whose probability values were in the range 0.00-0.25, 0.25-0.5, 0.5-0.75 or 0.75-1, were respectively set to 0, 0.5, 1.0 or 2.0*N_e. For *G. intermedia*, two different generic effective population sizes N_e were tested taking values of 2,500 and 25,000 for the COI mitochondrial locus and associated values four times larger for the nuclear loci, 10,000 and 100,000. These values of N_e were suggested by estimates inferred with the software IMa2 (Hey and Nielsen 2004, 2007) on the same DNA sequence data. We also tested the two following forward migration rates: 0.0001 and 0.00001 migrants/generation (based on results from Mardulyn and Milikovitch (2005)). Therefore, each tested scenario was simulated with all four possible combinations of two parameters (forward migration rate and the effective population sizes). For each of these combinations, we performed a total of 1,000 independent backward simulations. To take the stochastic variation associated with the forward simulation into account, we reiterated a forward simulation every 100 backward simulations to re-estimate backward simulations parameters (yielding a total of 10 forward for 1,000 backward simulations). The sequence length and fixed number of mutations for each simulated locus corresponded to the values observed in our real DNA sequences alignments.

For all hypotheses simulated under a spatially explicit model of coalescence, the occurrence of a strong range contraction in the oldest geographic layer (resulting in a single and relatively small ancestral region) was compulsory to allow generating sequence variation data similar to our observed data. Indeed, failing to do so resulted in the simulations generating genealogies with very long internal branches attracting most of the simulated mutations. This type of genealogies would correspond to a dataset featuring clusters of almost identical sequences separated by a large number of mutations, which is not observed in our sequence data. Indeed, because genes can coalesce only if located in the same grid cell (i.e., geographically very close), the last coalescence events would take too long to occur if genes were free to move over a large portion of the grid. Because the alpine populations currently display a higher level of diversity than the rest of the species range (Quinzin and Mardulyn 2014) and because the region remained habitable during all three modeled periods according to SDM results, we hypothesized that the ancestral region was located in the Alps (Fig. 2 and S4, Supporting information). However, the exact location of that ancestral region has likely little impact on the current pattern of genetic diversity, as long as the range contraction allows coalescence of all remaining lineages in a relatively short amount of time (i.e., internal branches of the genealogy not attracting most mutations).

Scenarios P1, 2 and 3 correspond to alternative hypotheses suggested by SDMs inferred under the LGM MIROC model (Fig. 1 and Figs. S2 and S4, Supporting information). They differ from each other by the following two characteristics: (i) whether an isolated northern population survived during the last glaciation and its LGM (Scenarios P2 and 3), and/or (ii) a connecting corridor exists between Northern and Southern regions (Scenarios P1 and 2) at the end of the last glaciations. Scenarios P4, 5 and 6 correspond to P1, 2 and 3, respectively, but are inferred with a LGM geographic distribution projected on the CCSM climatic model instead of the MIROC model (Fig. 1 and Figs. S2, S4, Supporting information). Scenarios P1-3 (MIROC) include a larger continuous distribution across regions during the last glaciation, except for the Pyrenees which are more isolated, while scenarios based on projections of CCSM climatic models, P4 to 6, indicate a more restricted range and a more distant northern population. For all six scenarios, the initial layer with one single Alpine population is constrained to occur 131,000 generations ago (one generation/year), corresponding to the time of our distribution estimate for the previous interglacial. Scenarios P7 to P12 correspond to scenarios similar to those of P1 to P6, but implementing the ancestral Alpine population 110,000 generations ago; which corresponds to the beginning of the last glacial episode (Fig. S4, Supplementary information).

For each scenario we compared the corresponding simulated sequences to the observed sequences. Several summary statistics were used for this comparison. To identify the most appropriate summary statistics, we compared the distribution of all statistics computed with SPADS 1.0 (Dellicour & Mardulyn 2014) for each dataset simulated under each scenario in PhyloGeoSim (Dellicour et al. 2014). Based on these comparisons, we selected four summary statistics, two measuring population differentiation and two measuring genetic diversity, as the most promising to discriminate among all tested scenarios: (i) global Φ_{ST} (Excoffier et al. 1992) calculated over all populations, (ii) nucleotide diversity (Nei and Li 1979) estimated within the Alps, (iii) relative nucleotide diversity (Mardulyn et al. 2009) estimated within Scandinavia, and (iv) pairwise Φ_{ST} (Excoffier et al. 1992) between the Carpathian and the Scandinavian populations. From this comparison, we derived a global *p-value* for each scenario, using the method described in Dellicour et al. (2014).

2.2 Approximate Bayesian Computation with ABCtoolbox.

ABC methods rely on simulating a large number of data sets under a specific coalescence model, but selecting values for the variable parameters of the model at random (according to a specified prior distribution) for each simulation, followed by a selection step aiming at isolating the simulations that generated data most similar to the observed data, and a regression step (Beaumont et al. 2002). The selection is based on the computation of summary statistics chosen to best summarize the information contained in the data. Only simulations for which the statistics calculated for the simulated data were similar enough to the observed data are retained, and are used to estimate a posterior distribution of model parameters. ABC analyses were conducted with the program ABCtoolbox (Wegmann et al. 2010), using SimCoal 2.0 (Laval and Excoffier 2004) to simulate sequence data under each tested scenario and arlsumstat (Excoffier and Lischer 2010) for the computation of summary statistics. In a first phase, we adopted a Monte Carlo Markov Chain (MCMC) sampling approach, for which a lower number of simulations is necessary, to identify and reject models associated with acceptance rates too low or posterior marginal densities near or equal to zero. During this first phase, we identified the best topology for the two divergence events between the three main colonized regions by *G. intermedia*. In a second phase, standard ABC sampling analyses were performed for the remaining scenarios that had not been rejected during the first phase (scenarios described in Table S4, supporting information).

For all MCMC-sampling analyses, we proceeded with a set of 100,000 simulations. We computed the distance between observed and simulated data based on linear combinations of summary statistics, via Partial Least Squares (PLS), that explain best the variance in the model parameter space (Wegmann and Excoffier 2010). To define the PLS, we used the R script provided with ABCtoolbox on a set of 10,000 simulations conducted with a standard sampling. After MCMC sampling, accepted simulations (a variable subset from the 100,000 simulations) were subsequently subjected to an extra regression adjustment step. Because the regression step is more robust to a large number of statistics (compared to the sampling step) and because PLS are model specific, we did not use the PLS transformed statistics for the regression step. The model specificity of the PLS makes their use inappropriate for model choice. We therefore used these analyses only to identify bad models, i.e. those with a posterior probability equal or close to zero.

For each standard ABC sampling analysis, we ran a set of 1,000,000 simulations. Linear combinations of the statistics (PLS) were used for regression adjustment step (in order to reduce the size of the dataset). The regression was conducted on a set of 10,000 simulations. Model choice was based on posterior probabilities (marginal densities) also ensuring a good model fit to the observed data (given obs *p-value*).

To specify the range of values to explore for three different model parameters (population mutation rate ($\boldsymbol{\theta}$), and its components, effective population size (N_e) and per-generation mutation rate (μ)), we estimated those from the observed data with the program IMa2 (Hey and Nielsen 2004, 2007). Theta ($\boldsymbol{\theta}$) was estimated only for COI, the only locus for which we could define a range of possible μ values (0.0015-0.02 substitution per nucleotide and per million year), based on information from the literature (Shapiro et al. 2006; Nazari and Sperling 2007). The range of values for $\boldsymbol{\theta}$ inferred from the COI locus was used to define the range of μ and $\mathrm{N_e}$ values for the nuclear loci, assuming an effective population size four times greater than for the mitochondrial locus.

Overall, we worked with demographic model assumptions aiming at reflecting the relevant observed conflicts among different SDMs. Prior and posterior parameter distributions are presented for each model in supplementary Table S4 and coalescence models underlying the simulated scenarios are shown in Fig. 3 and supplementary Fig. S5. Tested scenarios involved different degrees of complexity (requiring at least the estimation of 16-18 parameters). The divergence times among the three regions were the only variables of the simplest models (ISOL models), which included no migration or modification of population sizes. More complex scenarios were then generated by including migration (MIGR models) or population size modifications (expansion and/or bottleneck; scenarios CCROC III and IV), or both (different scenarios were proposed with different matrices; scenarios CCSM, MIROC I-IV and CCROC I, II). In order to decrease the complexity of the migration scheme, we assumed symmetric gene flow ranging from 0.0001 to 0.001. As already mentioned earlier, the upper value is suggested by a population structure study on another *Gonioctena* species, *G. olivavea* (Mardulyn and Milikovitch 2005), assuming similar dispersal capabilities.

In the ABC analyses, the persistence of a Nordic population at the LGM was evaluated by estimating the divergence time separating this population from the remaining of the distribution (via inference of the posterior probability distribution of this model parameter). In addition, we tested whether the three major portions of the range were isolated before the Last Interglacial (c. 130,000 years BP). We also investigated if gene flow occurred among these major regions after their divergence, i.e. before their current isolation, and estimated divergence times among them. Finally, one potential factor thought to impact population ranges during the glaciations and that were not implemented in our SDM modeling is the extent of the ice sheet, which was probably an important limiting factor, even for cold-adapted species. We therefore mimic such potential impact testing for demographic events (population size expansion/contraction according to ice sheet retrieve/expansion, respectively, and climate). Overall we more specifically addressed the following questions, 1/ did the fragmentation of the range directly lead to isolated regions, 2/ did two (or three) of the current main regions remain connected with gene flow or 3/ came temporarily into contact again recently, and 4/ did the climate and ice sheet affect population's demography?

The summary statistics used estimated either differentiation between populations or genetic diversity within populations: the number of private polymorphic sites per population (Private S), the mean number of sites with segregating substitutions over population (Mean S) and the standard deviation (s.d.) interval (SD S), the total number of polymorphic sites over all populations (Tot S), the mean number of pairwise sequence differences for each population (All Pi) and over populations (Mean Pi) and the s.d. of the mean over populations (SD Pi), and finally, the mean number of differences between pairs of populations (Pairwise Pi). In the absence of selection, older stable populations (e.g., ancient refuge) are expected to display higher genetic diversity levels compared to the rest of the species range (Hewitt 2000;

Petit et al. 2003).

References

Beaumont MA, Zhang W, Balding DJ. 2002 Approximate Bayesian computation in population genetics. Genetics 162, 2025–35.

Boyce MS, Vernier PR, Nielson SE, Schmiegelow FKA. 2002 Evaluating resource selection functions. Ecol. Modell. 157, 281–300.

Dellicour S, Kastally C, Hardy OJ, Mardulyn P. 2014 Comparing phylogeographic hypotheses by simulating DNA sequences under a spatially explicit model of coalescence. Mol Biol Evol. 31, 3359-3372.

Dellicour S, Mardulyn P. 2014 SPADS 1.0: a toolbox to perform spatial analyses on DNA sequence data sets. Mol. Ecol. Res. 14, 647-651.

Dormann CF, Schymanski SJ, Cabral J, Chuine I, Graham C, Hartig F, Kearney M, Morin X, Römermann C, Schröder B, et al. 2012 Correlation and process in species distribution models: bridging a dichotomy. J. Biogeogr. 39, 2119–2131.

Elith J, Kearney M, Philips S. 2010 The art of modelling range-shifting species. Methods Ecol. Evol. 1, 330-342.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ. 2011 A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17, 43–57.

Excoffier L, Smouse PE, Quattro JM. 1992 Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application. Genetics 131, 479–491.

Excoffier L, Lischer HEL. 2010 Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Res. 10, 564-567.

Fuentes-Hurtado M, Hof AR, Jansson R. 2016 Paleodistribution modeling suggests glacial refugia in Scandinavia and out-of-Tibet range expansion of the Arctic fox. Ecol. Evol. 6, 170-180.

Hey J, Nielsen R. 2004 Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167, 747–760.

Hey J, Nielsen R. 2007 Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc Natl Acad Sci USA 104, 2785–90.

Hewitt G. 2000 The genetic legacy of the Quaternary ice ages. Nature 405, 907–13.

Jiménez-Valverde A, Lobo JM. 2007 Threshold criteria for conversion of probability of species presence to either–or presence–absence. Acta Oecologica 31, 361–369.

Johnson JB, Omland KS. 2004 Model selection in ecology and evolution. Trends Ecol. Evol. 19, 101–8.

Laval G, Excoffier L. 2004 SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. Bioinformatics 20, 2485-2487.

Lobo JM, Jiménez-Valverde A, Real R. 2008 AUC: a misleading measure of the performance of predictive distribution models. Glob. Ecol. Biogeogr. 17, 145–151.

Mardulyn P, Milinkovitch MC. 2005 Inferring contemporary levels of gene flow and demographic history in a local population of the leaf beetle *Gonioctena olivacea* from mitochondrial DNA sequence variation. Mol. Ecol. 14, 1641–53.

Mardulyn P, Mikhailov YE, Pasteels JM. 2009 Testing phylogeographic hypotheses in a Euro-Siberian cold-adapted leaf beetle with coalescent simulations. Evolution 63, 2717–2729.

Nazari V, Sperling FAH. 2007 Mitochondrial DNA divergence and phylogeography in western Palaearctic Parnassiinae (Lepidoptera : Papilionidae): How many species are there ? insects Syst. Evol 38, 121–138.

Nei M, Li WH. 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76, 5269–5273.

Peterson AT, Papeş M, Soberón J. 2008 Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecol. Modell. 213, 63–72.

Petit RJ, Aguinagalde I, de Beaulieu JL, et al. 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. Science 300, 1563–5.

Phillips SJ, Anderson RP, Schapire RE. 2006 Maximum entropy modeling of species geographic distributions. Ecol. Modell. 190, 231–259.

Phillips S, Dudík M. 2008 Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31, 161–175.

Quinzin MC, Mardulyn P. 2014 Multi-locus DNA sequence variation in a complex of four leaf beetle species with parapatric distributions: Mitochondrial and nuclear introgressions reveal recent hybridization. Mol. Phylogenet. Evol. 78C, 14–24.

Ramstein G, Kageyama M, Guiot J, Wu H, Hély C, Krinner G, Brewer S. 2007 How cold was Europe at the Last Glacial Maximum? A synthesis of the progress achieved since the first PMIP model-data comparison. *Climate of the Past* 3(2), 331-339

Raes N, ter Steege H. 2008 A null-model for significance testing of presence-only species distribution models. Ecography 30, 727–736.

Rödder D, Schmidtlein S, Veith M, Lötters S. 2009 Alien invasive slider turtle in unpredicted habitat: a matter of niche shift or of predictors studied? PLoS One 4, e7843.

Shapiro LH, Strazanac JS, Roderick GK. 2006 Molecular phylogeny of Banza (*Orthoptera: Tettigoniidae*), the endemic katydids of the Hawaiian Archipelago. Mol. Phylogenet. Evol. 41, 53–63.

Warren DL, Glor RE, Turelli M. 2010 ENMTools: a toolbox for comparative studies of environmental niche models. Ecography 33, 607–611.

Warren D, Seifert S. 2011 Ecological niche modeling in Maxent : the importance of model complexity and the performance of model selection criteria. Ecol. Appl. 21, 335–342.

Wegmann D, Excoffier L. 2010 Bayesian inference of the demographic history of chimpanzees. Mol. Biol. Evol. 27, 1425–35.

Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L. 2010 ABCtoolbox: a versatile toolkit for approximate Bayesian computations. BMC Bioinformatics, 11, 116.