Journal of Biogeography (J. Biogeogr.) (2013) 40, 769–777

ORIGINAL ARTICLE

Allele elimination recalculated: nested subset analyses for molecular biogeographical data

Jan Christian Habel[1]*, Werner Ulrich[2] and Thorsten Assmann[3]

[1]Department of Ecology and Ecosystem Management, Technische Universität München, D-85350 Freising-Weihenstephan, Germany, [2]Nicolaus Copernicus University, Department of Animal Ecology, PL-87100, Toruń, Poland, [3]Leuphana University Lüneburg, Institute of Ecology, D-21335, Lüneburg, Germany

*Correspondence: Jan Christian Habel, Department of Ecology and Ecosystem Management Technische Universität München Hans-Carl-von-Carlowitz-Platz 2 D-85350 Freising-Weihenstephan, Germany
E-mail: Janchristianhabel@gmx.de

ABSTRACT

Aim Post-glacial colonization of species from low-latitude refugia to high latitudes, or from lower to higher elevations, often involves repeated founder effects due to stepwise colonization. This may cause repeated population bottlenecks and the subsequent loss of alleles. Regression analyses have traditionally been used to analyse the correlation between the mean numbers of alleles and geographical distances from refugia. Here, we describe and evaluate the performance of nested subset analyses for detecting allele elimination.

Methods Genetic data sets from five butterfly and one beetle species were reanalysed using regression and nested subset analyses.

Results The data sets analysed here showed both congruent and divergent results under regression and nested subset analyses. Some data sets did not feature a significant correlation between the mean number of alleles and the colonization trajectory, but did show significant nested structure. Others showed the opposite effects. Using allele frequencies from the same data sets, we did not obtain significant patterns of nestedness.

Main conclusions Our results indicate that classical regression analyses are not always a suitable tool for analysing allele elimination, and nestedness analyses are much more meaningful. Local natural selection can alter allele frequencies, thereby erasing biogeographical patterns that have evolved as a result of the stochastic processes involved in colonization. Thus, an appropriate means of documenting allele elimination *sensu* Reinig is the joint application of nested subset and regression analyses based on presence/absence and abundance data for genetic diversity.

Keywords
Allele elimination, bottleneck, butterflies, Carabidae, colonization, founder effect, linear regression, nestedness analyses, range shifts, Rhopalocera.

INTRODUCTION

Climatic oscillations during the Pleistocene caused severe range shifts among Northern Hemisphere temperate organisms, which often survived the cold phases in southern refugia and expanded northwards during post-glacial warming. This process of post-glacial colonization of higher latitudes and elevations has been analysed intensively using morphological and phenotypic characters (Reinig, 1938; Rensch, 1954, 1958; de Lattin, 1967). The development of molecular markers has led to an increase in phylogeographical studies, which have substantially expanded our understanding of biogeographical processes (Hewitt, 1996, 2004; see review by Avise, 2009). One of the major outcomes of many studies is the detection of intraspecific differentiation into multiple genetic lineages as a product of long-term isolation in discrete refugia during glacial phases. This has been extensively documented for the Iberian, Italian and Balkan peninsulas in the Mediterranean (Comes & Kadereit, 1998; Taberlet *et al.*, 1998; Hewitt, 2004), and for North Africa (Weingartner *et al.*, 2006; Habel *et al.*, 2008, 2011). Similar biogeographical structures were found in North America for numerous plants, invertebrates and vertebrates (e.g. Pruett & Winker, 2005; Milá *et al.*, 2006, 2007; Soltis *et al.*, 2006;

© 2012 Blackwell Publishing Ltd

doi:10.1111/jbi.12054

Journal of Biogeography

Hill *et al.*, 2011 and references therein). In contrast, Armbruster *et al.* (1998) reported a lack of concordance between genetic variability and geographical isolation in pitcher-plant mosquitoes.

Colonization processes over large geographical distances are generally reflected in population genetic structures: the mean number of alleles in populations at the leading edge (e.g. northern distribution margin) is often markedly lower than in populations in the previous refugial areas (rear edge), as a consequence of the founder effects that generally accompany colonization. This phenomenon was first described by Reinig (1938) as 'allele elimination', the stochastic loss of genetic variation as a result of repeated founder effects and population bottlenecks in the course of range extension. More recently, such gradients have been detected with molecular markers (Hewitt, 1996; Merilä *et al.*, 1996). Theoretical models have shown that the extent of allele elimination depends on the type of movement (Hewitt, 1996). 'Phalanx-wise' expansion (colonization by many individuals and populations in one direction at the same time) is accompanied by little loss of genetic variation over space and time, and is typical of widespread species with high abundance and strong dispersal ability. Examples of this type of movement are shown here by the three butterfly species *Maniola jurtina* (Linnaeus, 1758) (Nymphalidae), *Melanargia galathea* (Linnaeus, 1758) (Nymphalidae), and *Polyommatus icarus* (Rottemburg, 1775) (Lycaenidae). In contrast, 'stepping-stone' movement (the migration of a single or few individuals bridging large distances) is accompanied by the loss of genetic variation due to the repeated foundation of new populations by a single or few individuals and concomitant severe population bottlenecks. Examples of this colonization type are provided here by the butterfly species *Polyommatus coridon* (Poda, 1761) (Lycaenidae) and *Coenonympha arcania* (Linnaeus, 1761) (Nymphalidae), and by the ground beetle *Carabus auronitens* Fabricius, 1792 (Carabidae).

To test the significance of the loss of genetic variation over geographical space, most authors have used regression analysis to compare the total mean number of alleles with latitude (specifically, the distance from refugia; Schmitt & Seitz, 2002; Schmitt *et al.*, 2002; Besold *et al.*, 2008). However, according to Reinig's hypothesis of 'allele elimination' *sensu stricto*, alleles which are lost in the wake of colonization processes cannot reappear in subsequently founded populations. Analysis of the mean number of alleles with respect to the distance from the source population does not include information from allele elimination patterns, which are a hallmark signature of a recolonization process. Thus, genetic evidence for recolonization is better detected using nested subset analyses.

Until now, nested subset analyses have been used to detect nested metacommunity structures and an ordered loss of species along environmental gradients, whereby sites that are poorer in species are subsets of species-rich sites (Patterson & Atmar, 1986; Ulrich *et al.*, 2009). A nested subset pattern is therefore connected to predictable differences in species richness and composition among sites (Ulrich *et al.*, 2009). Nested subset analysis was originally developed to describe patterns of species composition between isolated habitats such as islands and fragmented landscapes, and has become popular since Patterson & Atmar (1986) proposed that nested subset patterns reflect an orderly sequence of extinctions in these situations. More recently, nestedness analysis has become a common tool in macroecology and biogeography to study the impact of habitat size, isolation and habitat quality traits on community assembly along environmental gradients and colonization trajectories (Cutler, 1991; Atmar & Patterson, 1993; Patterson & Atmar, 2000; Ulrich *et al.*, 2009). It is important to note that this form of nested subset pattern is by definition always linked to differences in species richness (which can be studied by regression analysis), while a difference in richness is not necessarily connected to nestedness (Ulrich *et al.*, 2009) (Fig. 1). A nested subset pattern

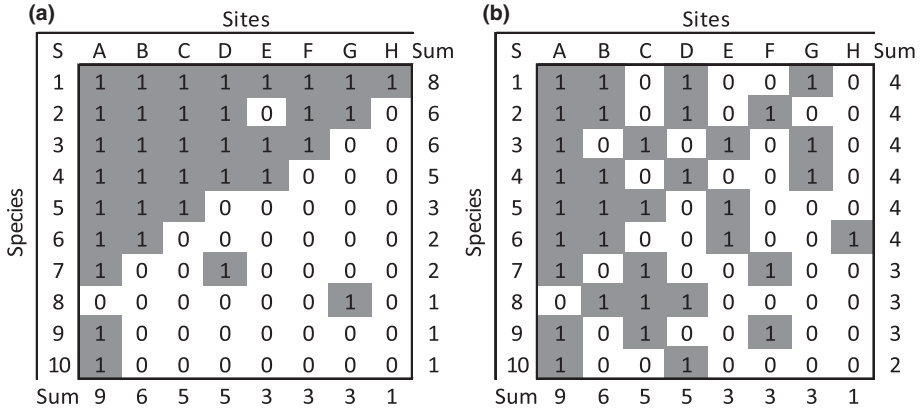


Figure 1 A hypothetical nested pattern of species occurrences (a) requires an ordered loss of species and is necessarily connected to a difference in species richness between sites. Absence of a particular species in a particular site (b) cannot be predicted from its overall number of occurrences. The matrix in panel (a) is significantly nested (NODF = 78.8; $P < 0.001$) while the matrix in panel (b) is not nested (NODF = 40.1; $P > 0.2$). Species and sites were sorted by the respective marginal (row/column) totals.

© 2012 Blackwell Publishing Ltd

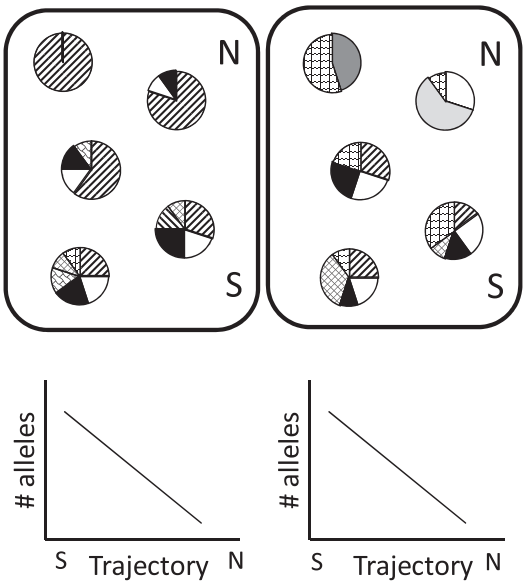
thus contains additional information about the probabilities of species' occurrences and absences in particular sites (for reviews on the relationship between richness differences and ordered species loss, see Baselga, 2010; Almeida-Neto *et al.*, 2011; Podani & Schmera, 2011).

The nestedness approach can also be adapted to analyse molecular biogeographical data sets. In population genetic studies, the presence of taxa (e.g. species) is replaced by genetic variants (e.g. alleles or haplotypes). A nested structure of gene pools should be found if the colonization occurred via the stepping-stone mode with the progressive loss of alleles. Such a scenario should yield a negative correlation between allelic variation and distance from the colonization source. Regression analysis might fail to detect the nested structure of the distribution pattern of alleles because it does not use information about the trajectory of allele loss. Thus, the identification of single alleles is more relevant than the detection of the sum of alleles using classical regression analyses (see Figs 1 and 2).

In this study we employ nestedness analysis for the first time as a tool for molecular biogeography, and highlight its explanatory power through examples of its use on multilocus data from six insect species that represent both phalanx and stepping-stone modes of colonization. We discuss concordant and contrasting results obtained from nestedness and regression analyses of colonization patterns.

MATERIALS AND METHODS

We reanalysed genetic data sets consisting of polymorphic allozymes from five butterfly and one beetle species with extensive geographical ranges. The molecular data include three categories of allele elimination obtained from regression analyses: marked elimination of genetic information over space, little elimination of genetic information over space, and zero allele elimination. As some of the study species have been inferred to have been isolated in multiple refugia during glaciations, and are likely to have recolonized central and northern Europe from multiple source populations, we divided the data sets into their source lineages and analysed them separately. We used the following organisms and data sets (n = number of populations): the butterflies *Polyommatus coridon* (western lineage: $n = 55$, Schmitt *et al.*, 2002; eastern lineage: $n = 18$, Schmitt & Seitz, 2002; total $n = 73$), *Coenonympha arcania* ($n = 19$; Schmitt & Besold, 2010), *Melanargia galathea* ($n = 42$; Habel *et al.*, 2011), *Maniola jurtina* (western lineage: $n = 4$; southern lineage: $n = 10$; eastern lineage: $n = 10$; total $n = 24$; Habel *et al.*, 2009), and *Polyommatus icarus* ($n = 31$; Schmitt *et al.*, 2003), as well as the ground beetle *Carabus auronitens* ($n = 33$; Assmann *et al.*, 1994). Most of the samples for local populations used in our analyses consist of 40 individuals. Populations with small sample sizes (less than 10% of the mean sampling size) were excluded from further analyses. Furthermore, alleles making up less than 5% of the total allele frequencies were excluded to prevent allele bias in the presence/absence matrix; however, for analyses based on allele frequencies we used also rare alleles with low frequency values.



S → N

Nested	A	B	C	D	E	Incidence	Abundance
1	0.25	0.30	0.60	0.80	1.00	5	2.95
2	0.20	0.25	0.15	0.10	0	4	0.7
3	0.20	0.20	0.15	0.10	0	4	0.65
4	0.15	0.15	0.10	0	0	3	0.4
5	0.10	0.10	0	0	0	2	0.2
6	0.10	0	0	0	0	1	0.1
Richness	6	5	4	3	1		
Frequency	1	1	1	1	1		

Figure 2 Proportions of six alleles at five sites from south (S) to north (N) (different alleles symbolized by different hatchings). Regression analysis yields identical results for both data sets. Organization is nested in the pattern on the left (NODF = 96, WNODF = 63, both $P < 0.001$) and not nested on the right (NODF = 63, WNODF = 23, both $P > 0.1$). The matrix below shows the nested pattern of the left distribution.

The presence and absence of alleles as well as the frequency of single alleles at multiple enzyme loci were used as input data. Allele × site presence/absence matrices were sorted according to occurrences (row totals: Fig. 2) (the input data files are given in Appendix S1 of Supporting Information). Sites were sorted using latitude as a proxy for the presumed colonization trajectory. To assess the degree of nestedness we used two metrics with different focuses: the incidence-based NODF (nestedness by overlap and decreasing fill) metric of Almeida-Neto *et al.* (2008) compares differences in occurrence across rows and sites, while the matrix temperature metric (Atmar & Patterson, 1993) is a distance metric and designed particularly for biogeographical

© 2012 Blackwell Publishing Ltd

applications (Ulrich *et al.*, 2009). Further, we assessed patterns of allele co-occurrence using the common *C*-score metric (Stone & Roberts, 1992; Ulrich & Gotelli, 2010), which is a normalized count of the total number of {{1,0}, {0,1}} submatrices within the presence/absence matrix. High *C*-score values point to a segregated pattern of allele co-occurrence within populations. To assess spatial turnover of alleles we sorted rows and columns according to the first axis of a correspondence analysis. In a matrix sorted by correspondence analysis, the position of the non-empty cells is specified by the row and column numbers. Following Gotelli & Ulrich (2012), we used the squared coefficient of correlation between the row and column numbers, r^2, as an index of how orderly the pattern of the non-empty cells along the matrix diagonal is; r^2 is thus a measure of spatial turnover of alleles.

To assess how allele elimination is linked to a decline in allele frequency, if at all, we used an abundance-based approach to nestedness (Almeida-Neto & Ulrich, 2010). For each locus with quantitative data for at least five alleles in a given species, we generated the relevant alleles × sites matrix (74 matrices in total) and assessed the degree of nestedness from the WNODF (weighed nestedness by overlap and decreasing fill) metric (Almeida-Neto & Ulrich, 2010), which extends NODF to quantitative data. WNODF measures the matrix-wide degree of ordered decline in allele frequency along a predefined environmental or diversity gradient. Of course, in frequency matrices where the total has to sum to 100%, decreases in some entries are necessarily connected to increases in other entries. In a nested situation, this increase always affects the allele with the higher frequency. However, this does not influence the performance of WNODF or the assessment of patterns (Almeida-Neto & Ulrich, 2010).

To test for statistical significance, we used a null model approach (Gotelli & Ulrich, 2012) and compared the observed metrics with the distribution of metrics obtained from 200 randomized matrices. Because there was no *a priori* constraint on the occurrences of alleles across the sample sites, apart from selection and bottleneck effects, we used two null models that controlled only for the total number of alleles across the sample sites but did not fix the total numbers of incidences for particular alleles or sites. For the presence/absence matrices, we used a null model that randomized allele incidences within the matrix (the *ee* null model; Gotelli, 2000). For the frequency matrices, we applied the *aa* null model of Ulrich & Gotelli (2010), which is based on total counts and assigns incidences to matrix cells in proportion to the observed row and column totals until the observed total number of incidences in the matrix is reached. In order to transform the observed frequencies into counts, we divided the frequency of each allele by the frequency of the least common allele. All analyses were performed with the programs NODF (Almeida-Neto & Ulrich, 2010) and TURNOVER (Ulrich & Gotelli, 2012).

RESULTS

Allele distributions showed a strong spatially nested pattern across sample sites for all data sets (Table 1), independent of the nestedness metric used. Degrees of nestedness were similar when ordered according to richness and colonization trajectory. Consequently, all matrices were significantly aggregated as inferred from the low *C*-scores in comparison to the null expectation. There was no significant degree of allele turnover across sites (Table 2).

The number of alleles decreased significantly along the colonization trajectory in *Carabus auronitens*, *Coenonympha arcania* and *Polyommatus coridon* (Fig. 3; $P_{(r < 0)} < 0.01$). In *Maniola jurtina*, this decrease was only visible in the eastern lineage ($P_{(r < 0)} < 0.01$) while in the western and southern populations, as well as in *Melanargia galathea,* allele variation was stable across the colonization trajectory ($P_{(r < 0)} > 0.05$). The allele number/trajectory plots of *Carabus auronitens* and the southern lineage of *Maniola jurtina* are not homogenous

Table 1 Raw scores of the matrix temperature metric (T) and NODF metric and expected scores (*ee* null model: mean ± standard deviation) of our nine data sets where sites were sorted according to incidence or colonization trajectory. Significant scores ($P < 0.01$) are shown in bold type.

	Sorted by			Sorted by			Sorted by		
Metric	Incidence	Colonization trajectory	Expected	Incidence	Colonization trajectory	Expected	Incidence	Colonization trajectory	Expected
	Carabus auronitens			*Coenonympha arcania*			*Melanargia galathea*		
T	**6.39**	**10.6**	55.76 ± 3.59	**10.53**	**14.67**	68.55 ± 2.43	**20.24**	**25.43**	74.98 ± 1.62
NODF	**61.06**	**51.58**	25.97 ± 1.17	**63.69**	**62.76**	44.06 ± 1.12	**56.56**	**48.85**	42.40 ± 0.61
	Maniola jurtina East			*Maniola jurtina* South			*Maniola jurtina* West		
T	**6.92**	**21.3**	36.72 ± 4.23	**15.43**	**20.35**	59.55 ± 3.23	**13.15**	**26.73**	59.75 ± 3.57
NODF	**35.46**	**33.69**	47.45 ± 2.54	**60.33**	**57.65**	51.55 ± 1.63	**49**	**46.49**	51.35 ± 1.59
	Polyommatus coridon West			*Polyommatus coridon* East			*Polyommatus icarus*		
T	**11.1**	**12.29**	77.74 ± 1.13	**15.46**	**15.92**	68.79 ± 1.95	**17.17**	**20.58**	75.35 ± 1.59
NODF	**55.7**	**53.13**	34.67 ± 0.34	**54.57**	**54.37**	39.76 ± 0.80	**57.41**	**55.85**	42.27 ± 0.64

© 2012 Blackwell Publishing Ltd

Table 2 Spatial turnover among study sites (*C*-score and r^2) and respective expected scores (the *ee* null model: mean ± standard deviation) of our nine data sets. Significant scores ($P < 0.01$) are shown in bold type.

Metric	Observed	Expected	Observed	Expected	Observed	Expected
	Carabus auronitens		*Coenonympha arcania*		*Melanargia galathea*	
C-score	**0.01**	0.07 ± 0.001	**0.01**	0.12 ± 0.001	**0.03**	0.12 ± 0.01
r^2	**0.12**	0.29 ± 0.03	**0.08**	0.13 ± 0.01	**0.04**	0.09 ± 0.01
	Maniola jurtina East		*Maniola jurtina* South		*Maniola jurtina* West	
C-score	**0.02**	0.13 ± 0.01	**0.02**	0.13 ± 0.005	**0.02**	0.12 ± 0.005
r^2	**0.04**	0.27 ± 0.03	**0.05**	0.14 ± 0.01	**0.04**	0.15 ± 0.01
	Polyommatus coridon West		*Polyommatus coridon* East		*Polyommatus icarus*	
C-score	**0.01**	0.1 ± 0.005	**0.02**	0.11 ± 0.001	**0.03**	0.12 ± 0.01
r^2	**0.01**	0.08 ± 0.003	**0.06**	0.14 ± 0.01	**0.05**	0.09 ± 0.005

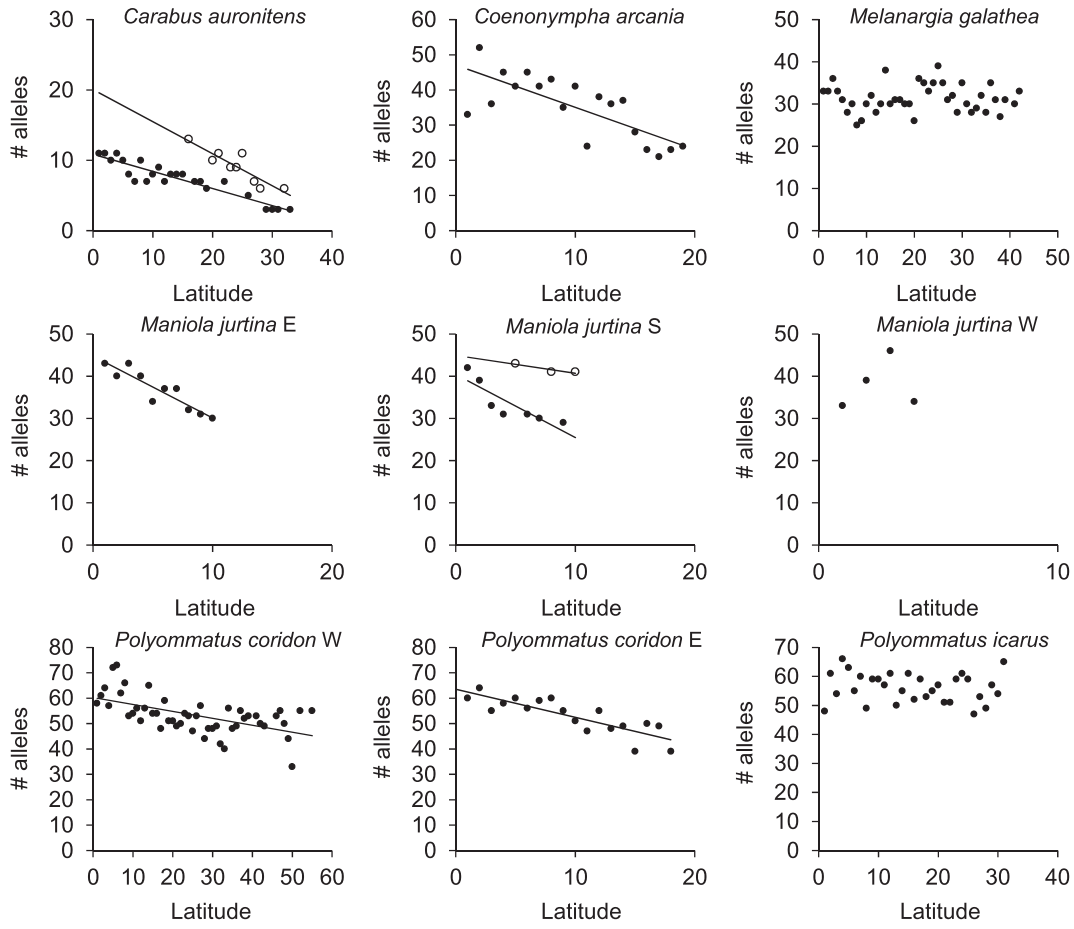


Figure 3 Colonization trajectories (latitudinal gradient) of one ground beetle (*Carabus auronitens*) and five butterfly species partly coincide with the reduced number of alleles across sample sites. In *C. auronitens* and the southern line of *Maniola jurtina*, two apparently different colonization trajectories are indicated by filled and open circles. Regression lines shown (except for *M. jurtina* West) are statistically significant at $P < 0.01$.

(Fig. 3) and are suggestive of two different colonization pathways.

Our quantitative approach to test for nestedness confirmed the nestedness results based on presence/absence (Table 3, Appendix S2). The incidence-based NODF metric used with the *aa* null model identified 33 allozymes (45%) as having a nested allele structure when sorted according to colonization trajectory ($P < 0.05$, Table 3). Sorting according to incidence and abundance resulted in only 25 (32%) significant scores (Table 3). We found no clear trends towards ordered frequency declines (Table 3, Appendix S2). For 13 alleles (18%), we found a significant nested pattern,

© 2012 Blackwell Publishing Ltd

Table 3 Summary of the results presented in Appendix S1. WNODF and NODF scores (total of 74 single comparisons) are shown above or below expected scores and above or below lower and upper two-sided 95% confidence limits of the *aa* null model. Null model distributions were obtained from 200 randomized matrices.

	Ordered according to colonization trajectory		Ordered according to incidence and frequency	
Numbers of	WNODF	NODF	WNODF	NODF
Observed score < expected score	39	16	51	28
Observed score > expected score	35	58	23	46
Observed score < lower confidence limit	25	3	32	14
Observed score > upper confidence limit	13	33	8	25

while 25 alleles (34%) appeared to be significantly non-nested, indicating an irregular pattern of allele frequency across sites.

One-way ANOVA yielded WNODF scores for single loci that differed significantly between species (sites ordered by colonization trajectory: $F = 7.2$, $P < 0.001$; ordered by incidence/abundance: $F = 6.6$, $P < 0.001$; Appendix S2). Scattered frequency distributions with scores less than 10 characterized *Carabus auronitens*, *Melanargia galathea* and *Polyommatus icarus* (Appendix S1), while the eastern and southern groups of *Maniola jurtina* and the eastern group of *Polyommatus coridon* had average WNODF scores above 15. In turn, the ANOVA did not indicate significant differences between species ($P > 0.1$ for both orders) when referring to incidences (NODF, Table 3).

Our data for the same enzyme loci across different species allowed us to address the question of whether the pattern of allele loss was repeated at the same loci in different species. Two-way ANOVA indicated similarities in nestedness between species. Within-allele variability in the degree of nestedness was significantly lower than the between-allele variability (sorting by colonization trajectory: $F = 2.25$, $P = 0.01$; sorting by incidence/abundance: $F = 3.58$, $P < 0.001$). These differences did not depend on species identity (both orders: $P > 0.1$). The allozyme loci *G6PDH*, *ME* and *PEP* deviated from the previously described pattern; their WNODF scores differed by more than 10 units (Appendix S2).

DISCUSSION

In this study, we compared the results of allele elimination in the wake of colonization processes obtained from regression and nestedness analyses. In terms of their performance under the two analytical procedures, the data sets could potentially be assigned to three categories: (1) data sets with a negative correlation between allele number and distance to the refugial area showing congruence with the results obtained from the nestedness analyses; (2) data sets without a significant correlation, but with a significant degree of nestedness; and (3) data sets with a significant correlation, but no significant nestedness signal. Eight species/data sets fell into category 1, four into category 2 and none into category 3. In general, nestedness analysis yielded higher effects than regression analysis, and thus constitutes a sensitive tool for detecting allele loss during colonization (see Table 1).

Allele number or pattern of alleles?

Most of our data sets showed a significant correlation between latitude and allele number when using linear regression analysis: the highest correlation was found for *Coenonympha arcania* ($r = 0.82$, $P < 0.01$) (Besold *et al.*, 2008) and for the eastern lineage of *Polyommatus coridon* ($r = 0.88$, $P < 0.01$) (Schmitt & Seitz, 2002). The analyses for *Carabus auronitens* revealed significant correlation between allele numbers and latitude with low r^2 values (ranging from 0.143 to 0.261 depending on the populations considered; Reimann *et al.*, 2002). Only a marginal and insignificant reduction in the mean number of alleles in relation to latitude was detected in *Polyommatus icarus* (Schmitt *et al.*, 2002). For the two butterflies *Melanargia galathea* and *Maniola jurtina*, no significant loss of alleles was found (Habel *et al.*, 2010, 2011). For *M. jurtina*, the latter trend was only detectable for the eastern lineage, and not for the western one.

In contrast, when using nestedness analyses, we detected an ordered loss of alleles for all species, highlighting the sensitivity of our approach. That the reduction of incidence is not necessarily linked to strong nestedness is well known in biogeography (Ulrich & Gotelli, 2010) and in studies of bipartite ecological networks (Bascompte *et al.*, 2003). Here we showed a similar pattern in the spatial distribution of alleles.

Stochasticity versus natural selection

In a second calculation, we replaced the presence/absence matrix of alleles with frequency data for each population and enzyme locus, to investigate the additional information content of this more comprehensive data. Contrary to the presence/absence situation, no nested structure was detectable for most of the data sets. These contrasting results and the lack of an ordered structure of allele frequencies point to local natural selection pressures. Colonization may be accompanied by founder effects and therefore may expose the population to stochastic processes (Clegg *et al.*, 2002). Species representing a 'stepwise' process are often strongly

© 2012 Blackwell Publishing Ltd

affected by repeated founder effects and accompanying population bottlenecks. Differences in environmental conditions may reflect divergent selection pressures. Allozymes can be under strong natural selection, as has already been shown for many butterfly species (Stuber *et al.*, 1980; Neva *et al.*, 1986; Watt, 1994; Johannesson *et al.*, 1995; Eanes, 1999; Karl *et al.*, 2008, 2009). The absence of random genetic drift and the presence of selection on genes (e.g. allozymes) can generate a directional shift in allele frequencies, which works against stochastic processes during colonization (MacArthur & Wilson, 1963; Hubbell, 2001). Nevertheless, we assume that allele frequencies contain additional useful information when using non-coding (neutral) markers unaffected by local selection pressures.

CONCLUSIONS

Our data clearly suggest that the best way to detect 'allele elimination' *sensu stricto* is to analyse the pattern of allele occurrence (presence/absence) over space. While regression analyses did not detect any significant loss of alleles for the two species representing the phalanx colonization mode, we detected a significant nestedness structure. Furthermore, the degrees of significance differed markedly depending on the method used: nestedness analysis seems to be much more sensitive than classical regression calculations. Our results highlight that nested subset analysis is superior to regression analysis in studies of the genetic consequences of colonization processes. Our analyses suggest that data on the presence or absence of alleles are adequate for conducting a nested subset analysis. An important advantage to analysing allele presence/absence patterns (instead of frequencies) is that they are likely to be less sensitive to local selection pressures, and hence may reflect the signature of the colonization process more reliably. Joint analysis of occurrence and frequency data could be a tool for identifying differences in local selection pressures along colonization trajectories.

ACKNOWLEDGEMENTS

J.C.H. thanks the German Academic Exchange Service (DAAD) for funding. We thank Friedrich Weber (University of Münster) and the late Alfred Seitz (University of Mainz, Germany) for numerous stimulating discussions about the impact of post-glacial colonization processes on the genetic make-up of animal populations. We thank Claudia Drees (Hamburg, Germany) and Thomas Schmitt (Trier, Germany) for providing data sets for analyses. We thank Martin Husemann (Baylor University, Waco, TX) and two anonymous referees for improving the scientific and linguistic value of this manuscript significantly.

REFERENCES

Almeida-Neto, M. & Ulrich, W. (2010) A straightforward computational approach for measuring nestedness using quantitative matrices. *Environmental Modelling and Software*, **26**, 173–178.

Almeida-Neto, M., Guimarães, P., Guimarães, P.R., Jr, Loyola, R.D. & Ulrich, W. (2008) A consistent metric for nestedness analysis in ecological systems: reconciling concept and quantification. *Oikos*, **117**, 1227–1239.

Almeida-Neto, M., Frensel, D.M.B. & Ulrich, W. (2011) Rethinking the relationship between nestedness and beta diversity: a comment on Baselga (2010). *Global Ecology and Biogeography*, **21**, 772–777.

Armbruster, P., Bradshaw, W.E. & Holzapfel, C.M. (1998) Effects of postglacial range expansion on allozyme and quantitative genetic variation of the pitcher-plant mosquito, *Wyeomyia smithii*. *Evolution*, **52**, 1697–1704.

Assmann, T., Nolte, O. & Reuter, H. (1994) Postglacial colonisation of middle Europe by *Carabus auronitens* as revealed by population genetics (*Carabus*, Carabidae). *Carabid beetles: ecology and evolution* (ed. by K. Desender, M. Dufreine, M. Loreau, M.L. Luff and J.P. Maelfait), pp. 3–9. Kluwer Academic Publisher, Dordrecht.

Atmar, W. & Patterson, B.D. (1993) The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia*, **96**, 373–382.

Avise, J.C. (2009) Phylogeography: retrospect and prospect. *Journal of Biogeography*, **36**, 3–15.

Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M. (2003) The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences USA*, **100**, 9383–9387.

Baselga, A. (2010) Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, **19**, 134–143.

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

Clegg, S.M., Degnan, S.M., Kikkawa, J., Moritz, C., Estoup, A. & Owens, I.P.F. (2002) Genetic consequences of sequential founder events by an island-colonizing bird. *Proceedings of the National Academy of Sciences USA*, **99**, 8127–8132.

Comes, H.P. & Kadereit, J.W. (1998) The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, **3**, 432–438.

Cutler, A. (1991) Nested faunas and extinction in fragmented habitats. *Conservation Biology*, **5**, 496–505.

Eanes, W.F. (1999) Analysis of selection on enzyme polymorphisms. *Annual Review of Ecology and Systematics*, **30**, 301–326.

Gotelli, N.J. (2000) Null model analysis of species co-occurrence patterns. *Ecology*, **81**, 2606–2621.

Gotelli, N.J. & Ulrich, W. (2012) Statistical challenges in null model analysis. *Oikos*, **121**, 171–180.

Habel, J.C., Meyer, M., El Mousadik, A. & Schmitt, T. (2008) Africa goes Europe: the complete phylogeography of the marbled white butterfly species complex *Melanargia*

© 2012 Blackwell Publishing Ltd

galathea/*M. lachesis*. *Organisms Diversity & Evolution*, **8**, 121–129.

Habel, J.C., Dieker, P. & Schmitt, T. (2009) Biogeographical connections between the Maghreb and the Mediterraenan peninsulas of southern Europe. *Biological Journal of the Linnean Society*, **98**, 693–703.

Habel, J.C., Rödder, D., Scalercio, S., Meyer, M. & Schmitt, T. (2010) Strong genetic cohesiveness between Italy and North Africa in four butterfly species. *Biological Journal of the Linnean Society*, **99**, 818–830.

Habel, J.C., Lens, L., Rödder, D. & Schmitt, T. (2011) From Africa to Europe and back: refugia and range shifts cause high genetic differentiation in the Marbled White butterfly *Melanargia galathea*. *BMC Evolutionary Biology*, **11**, 215.

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

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

Hill, J.K., Griffiths, H.M. & Thomas, C.D. (2011) Climate change and evolutionary adaptations at species' range margins. *Annual Review of Entomology*, **56**, 143–159.

Hubbell, S.P. (2001) *The unified neutral theory of biodiversity and biogeography*. Princeton, University Press, Princeton, NJ.

Johannesson, K., Johannesson, B. & Lundgren, U. (1995) Strong natural selection causes microscale allozyme variation in a marine snail. *Proceedings of the National Academy of Sciences USA*, **92**, 2602–2606.

Karl, I., Schmitt, T. & Fischer, K. (2008) Phosphoglucose isomerase genotype affects life-history traits and cold stress resistance in a Copper butterfly. *Functional Ecology*, **22**, 887–894.

Karl, I., Schmitt, T. & Fischer, K. (2009) Genetic differentiation between alpine and lowland populations of a butterfly is related to PGI enzyme genotype. *Ecography*, **32**, 488–496.

de Lattin, G. (1967) *Grundriß der Zoogeographie*. Gustav Fischer, Jena, Germany.

MacArthur, R.H. & Wilson, E.O. (1963) An equilibrium theory of insular zoogeography. *Evolution*, **17**, 373–387.

Merilä, J., Björklund, M. & Baker, A.J. (1996) Genetic population structure and gradual northward decline of genetic variability in the greenfinch (*Carduelis chloris*). *Evolution*, **50**, 2548–2557.

Milá, B., Smith, T.B. & Wayne, R.K. (2006) Postglacial population expansion drives the evolution of long-distance migration in a songbird. *Evolution*, **60**, 2403–2409.

Milá, B., McCormack, J.E., Castañeda, G., Wayne, R.K. & Smith, T.B. (2007) Recent postglacial range expansion drives the rapid diversification of a songbird lineage in the genus *Junco*. *Proceedings of the Royal Society B: Biological Sciences*, **274**, 2653–2660.

Neva, E., Beiles, A., Kaplan, D., Golenberg, E.M., Olsvig-Whittaker, L. & Naveh, Z. (1986) Natural selection of allozyme polymorphisms: a microsite test revealing ecological genetic differentiation in wild barley. *Evolution*, **40**, 13–20.

Patterson, B.D. & Atmar, W. (1986) Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological Journal of the Linnean Society*, **28**, 65–82.

Patterson, B.D. & Atmar, W. (2000) Analyzing species composition in fragments. *Bonner Zoologische Monographien*, **46**, 9–24.

Podani, J. & Schmera, D. (2011) A new conceptual and methodological framework for exploring and explaining pattern in presence–absence data. *Oikos*, **120**, 1625–1638.

Pruett, C.L. & Winker, K. (2005) Northwestern song sparrow populations show genetic effects of sequential colonization. *Molecular Ecology*, **14**, 1421–1434.

Reimann, T., Assmann, T., Nolte, O., Reuter, H., Huber, C. & Weber, F. (2002) The paleoecology of *Carabus auronitens* Fabricius: characterization and localization of glacial refugia in southern France and reconstruction of postglacial expansion routes by means of allozyme polymorphisms. *Abhandlungen des Naturwissenschaftlichen Vereins Hamburg*, **35**, 1–151.

Reinig, W.F. (1938) *Elimination und Selektion*. Gustav Fischer, Jena.

Rensch, B. (1954) *Neuere Probleme der Abstammungslehre*. Enke, Stuttgart.

Rensch, B. (1958) Die ideale Artbeschreibung. *Uppsala Universitets Arsskrift*, **6**, 91–103.

Schmitt, T. & Besold, J. (2010) Upslope movements and large scale expansions: the taxonomy and biogeography of the *Coenonympha arcania–C. darwiniana–C. gardetta* butterfly species complex. *Zoological Journal of the Linnean Society*, **159**, 890–904.

Schmitt, T. & Seitz, A. (2002) Postglacial distribution area expansion of *Polyommatus coridon* (Lepidoptera: Lycaenidae) from its Ponto-Mediterranean glacial refugium. *Heredity*, **89**, 20–26.

Schmitt, T., Gieβl, A. & Seitz, A. (2002) Postglacial colonisation of western Central Europe by *Polyommatus coridon* (Poda 1761) (Lepidoptera: Lycaenidae): evidence from population genetics. *Heredity*, **88**, 26–34.

Schmitt, T., Giessl, A. & Seitz, A. (2003) Did *Polyommatus icarus* (Lepidoptera: Lycaenidae) have distinct glacial refugia in southern Europe? Evidence from population genetics. *Biological Journal of the Linnean Society*, **80**, 529–538.

Soltis, D.E., Morris, A.B., McLachlan, J.S., Manos, P.S. & Soltis, P.S. (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, **15**, 4261–4293.

Stone, L. & Roberts, A. (1992) Competitive exclusion or species aggregation? An aid in deciding. *Oecologia*, **91**, 419–424.

Stuber, C.W., Moll, R.H., Goodman, M.M., Schaffer, H.E. & Weir, B.S. (1980) Allozyme frequency changes associated with selection for increased grain yield in maize (*Zea mays* L.). *Genetics*, **95**, 225–236.

© 2012 Blackwell Publishing Ltd

Taberlet, P., Fumagalli, L., Wust-Saucy, A.-G. & Cosson, J.-F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Ulrich, W. & Gotelli, N.J. (2010) Null model analysis of species associations using abundance data. *Ecology*, **91**, 3384–3397.

Ulrich, W. & Gotelli, N.J. (2012) Pattern detection in null model analysis. *Oikos*, doi:10.1111/j.1600-0706.2012.20325.x.

Ulrich, W., Almeida-Neto, M. & Gotelli, N.J. (2009) A consumer's guide to nestedness analysis. *Oikos*, **118**, 3–17.

Watt, W.B. (1994) Allozymes in evolutionary genetics: self-imposed burden or extraordinary tool? *Genetics*, **136**, 11–16.

Weingartner, E., Wahlberg, N. & Nylin, S. (2006) Speciation in *Pararge* (Satyrinae: Nymphalidae) butterflies – North Africa is the source of ancestral populations of all *Pararge* species. *Systematic Entomology*, **31**, 621–632.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Alleles × sites matrices of the seven species used for the nestedness analyses.

Appendix S2 Quantitative nestedness analysis of enzyme loci.

BIOSKETCHES

Jan Christian Habel is assistant professor at the Department of Ecology and Ecosystem Management of the Technische Universität München, Germany. His special interests are molecular ecology, biogeography and conservation biology.

Werner Ulrich is head of the macroecology research group at the Nicolaus Copernicus University in Torun, Poland. His main research interests are spatial distributions of insect and plant species, patterns of species co-occurrence and nestedness, and the ecology of parasitic Hymenoptera.

Thorsten Assmann is head of the animal ecology group at the Leuphana University Lüneburg, Germany. His main research fields are conservation biology, ecology, taxonomy and biogeography.

Author contributions: J.C.H. and T.A. conceived the ideas; J.C.H. collected the data; W.U. analysed the data; and all authors contributed to data interpretation and writing the manuscript.

Editor: Judith Masters

© 2012 Blackwell Publishing Ltd