Single Nucleotide Polymorphisms Reveal Genetic Structuring of the Carpathian Newt and Provide Evidence of Interspecific Gene Flow in the Nuclear Genome

Piotr Zielinski1*, Katarzyna Dudek1, Michal Tadeusz Stuglik1, Marcin Liana2, Wieslaw Babik1

1 Institute of Environmental Sciences, Jagiellonian University, Krakow, Poland, 2 Department of Comparative Anatomy, Institute of Zoology, Jagiellonian University, Krakow, Poland

Abstract

Genetic variation within species is commonly structured in a hierarchical manner which may result from superimposition of processes acting at different spatial and temporal scales. In organisms of limited dispersal ability, signatures of past subdivision are detectable for a long time. Studies of contemporary genetic structure in such taxa inform about the history of isolation, range changes and local admixture resulting from geographically restricted hybridization with related species. Here we use a set of 139 transcriptome-derived, unlinked nuclear single nucleotide polymorphisms (SNP) to assess the genetic structure of the Carpathian newt (Lissotriton montandoni, Lm) and introgression from its congener, the smooth newt (L. vulgaris, Lv). Two substantially differentiated groups of Lm populations likely originated from separate refugia, both located in the Eastern Carpathians. The colonization of the present range in north-western and south-western directions was accompanied by a modest loss of variation; admixture between the two groups has occurred in the middle of the Eastern Carpathians. Local, apparently recent introgression of Lv alleles into several Lm populations was detected, demonstrating increased power for admixture detection in comparison to a previous study based on a limited number of microsatellite markers. The level of introgression was higher in Lm populations classified as admixed than in syntopic populations. We discuss the possible causes and propose further tests to distinguish between alternatives. Several outlier loci were identified in tests of interspecific differentiation, suggesting genomic heterogeneity of gene flow between species.

Citation: Zielinski P, Dudek K, Stuglik MT, Liana M, Babik W (2014) Single Nucleotide Polymorphisms Reveal Genetic Structuring of the Carpathian Newt and Provide Evidence of Interspecific Gene Flow in the Nuclear Genome. PLoS ONE 9(5): e97431. doi:10.1371/journal.pone.0097431

Introduction

Most species are genetically structured, and genetic structure is often observed at multiple spatial scales [1,2]. Genetic structure is the result of a complex interplay of drift, gene flow, and natural selection acting on standing variation and new mutations [3–5]. The relative importance of these evolutionary forces is contingent on biological features of the organisms [6], and has also been affected by large-scale historical events, such as the Pleistocene climatic oscillations [2,7]. Identification of factors responsible for the observed spatial structuring of genetic diversity is a major goal of population genetics [8]. The quantification and understanding of genetic structure within species are of fundamental importance for inferential studies of population history, population ecology and biodiversity conservation [3,9,10]. Analyses of genetic structure are also essential for several aspects of the study of adaptation [5,11–18].

Genetic variation within species is commonly structured in a hierarchical manner which may result from superimposition of processes acting at different spatial and temporal scales. For example the impact of major climatic oscillations is clearly visible in the patterns of genetic differentiation observed currently in temperate and boreal species [2,19]. This is believed to reflect mainly secondary contact and partial admixture of populations derived from separate refugia with a contribution of processes related to the expansion itself, such as allele surfing [4,7]. Within these major geographic groups, populations are differentiated due to limited dispersal producing isolation by distance [20,21].

In species with limited dispersal capabilities, signatures of past subdivision are detectable for a long time [2,22]. This may be due to a combination of limited dispersal per se and differential adaptation in refugia superimposed on contemporary ecological gradients [23]. Admixture may also be delayed or prevented by the accumulation of intrinsic incompatibilities between populations [24,25] and poor dispersers appear to speciate on smaller geographic scales [26]. Thus studying contemporary genetic structure in taxa characterized by limited dispersal is likely to
provide ample information about historical and demographic events. Amphibians and in particular salamanders are ideal for such inferences [27,28]. Another advantage of such taxa is that they retain historical information about spatial variation of genetic exchange with related, incompletely reproductively isolated species ([29] and references therein).

Detecting, quantifying and interpreting genetic structure requires appropriate tools. Single nucleotide polymorphisms (SNPs) are powerful markers well suited for assessing genetic structure [30,31]. They are amendable to high throughput, cost-effective and reliable genotyping through array-based [32] and genotyping by sequencing [33] approaches. If SNP discovery and genotyping are performed separately, the researcher has control over the location and other characteristics of the SNPs selected for genotyping. For instance, if polymorphism data from transcriptome sequencing are available, SNPs located in known protein coding genes can be selected for genotyping, providing information about variation in functionally important regions. On the other hand, a discover-then-genotype approach introduces ascertainment bias, which distorts the picture of variation obtained from a larger sample [34]. However, initial discovery of SNPs in a smaller sample may be desirable for some applications, such as detection and quantification of population structure [35,36]. This is because the discovery process is biased towards more variable SNPs, thus increasing the per-marker information content, especially if SNP discovery is performed in a random or geographically diverse sample [37]. A distinct advantage of SNPs over microsatellites is that orders of magnitude more locations in the genome, can be easily interrogated. Thus SNPs offer a truly genome-wide perspective, essential if the biological processes of interest affect portions of the genome differentially [38–40].

Here we investigate the genetic structure of populations of the Carpathian newt (Lissotriton montandoni, Lm), a species which has apparently survived the glacial period in the Carpathians [41], an important refugial area [42,43]. Two processes appear to have profoundly affected this species and shaped the genetic structure currently observed. The first includes climatic oscillations during the Pleistocene, likely responsible for the observed regional-scale genetic structuring. The second involves hybridization with and introgression from its widely distributed congener, the smooth newt (L. vulgaris, Lv). Previous studies [29,41,44,45] demonstrated that Lv is genetically differentiated across its range in both mitochondrial and nuclear (microsatellites) genome. Patterns of genetic differentiation and species distribution modeling performed by Zieďinski et al. [29] suggest several glacial refugia in the Carpathians. While multiple, spatially and temporally distinct introgression events from Lv resulted in complete mtDNA replacement in Lm, very little recent interspecific nuclear gene flow was suggested by microsatellite markers [29]. However, interspecific gene flow in some parts of the nuclear genome has been extensive, as evidenced by data from the Major Histocompatibility Complex genes [45].

A set of transcriptome-derived SNPs and extensive sampling are used herein to address the following issues. First, we compare genetic structure inferred from the genome-wide sample of SNP markers with that estimated previously [29] from a much smaller number of microsatellites. Specifically we wanted to determine the number of genetic clusters (which may correspond to glacial refugia) supported by SNP markers, delineate their distribution and estimate the extent of admixture between them. To this end we use a comprehensive, uniform sampling, including previously undersampled Ukrainian Carpathians where admixture between genetic clusters was expected. Second, we test whether introgression from Lv is detectable in the nuclear genome of Lm with an increased number of markers, and if so, whether introgression varies geographically. Populations in which both species co-occur were also sampled across the range to estimate the admixture in syntopy. Third, we apply outlier analysis to test heterogeneity of gene flow within and between species and identify genes departing from the genomic average; such genes may be involved in population or species-specific adaptations.

Material and Methods

Sample collection

Altogether we analyzed 473 individuals from 40 populations: 25 populations of Lm (298 individuals), 7 syntopic populations in which both species co-occur (83 individuals) and 8 populations of Lv (92 individuals) [Fig. 1; Table 1]. Sampling sites were selected to cover the Lm range uniformly and to reflect Lv diversity in the surrounding areas. The average per population sample size, 12, might be considered low for some of the population genetic analyses, however we decided to rather maximize the number of markers as it was shown that this might be beneficial for robust landscape genetic inferences [46]. Throughout the text, we use terms population or locality interchangeably to refer to a particular breeding site consisting of one or more closely located water bodies. Adult newts were sampled by dip-netting during breeding season. Animals were released after tailtips were collected. Tissue samples were stored in 95% ethanol until DNA extraction. DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega).

Ethics statement

Animal samples were collected under permits: DOPozgiz-4200/II-78/3702/10/JRO provided by the Polish General Director for Environmental Protection, 03.04.12 No. 67 provided by the National Academy of Sciences of Ukraine and 3256/9.07.2010 provided by the Romanian Commission for Protection of Natural Monuments. Samples were collected with institutional animal ethics approval (number 101/2009), issued by the First Local Ethical Committee on Animal Testing at the Jagiellonian University in Krakow. Tissue samples were collected according to the requirements of these institutions; adult newts were captured by dip netting and tissue samples from tail tips were taken under anesthesia. Immediately after recovery from anesthesia, newts were released at the collection site. The sampling locations were not privately owned or protected in any way.

SNP discovery, assay development, and genotyping

SNPs were identified in liver transcriptomes of six Lm individuals sampled to encompass the genetic diversity of the species (Fig. 1). Transcriptomes were sequenced using Illumina technology and de novo assembled with Trinity [47]. Details of transcriptome sequencing and assembly will be provided elsewhere (Stuglik et al. in prep). We used a custom bioinformatic pipeline [48] to construct transcriptome-based gene models (TGM) from the Trinity output. Reads were mapped to this reference transcriptome with Bowtie2 [49] and SNPs were called with SAMTools 0.1.18. [50]. Next, we used blast searches against Xenopus tropicalis transcripts to identify TGM representing protein coding genes. To be included in the design of the genotyping assay, the SNP had to fulfill the following criteria: i) occur in a TGM which produced an unambiguous hit to a single Xenopus gene and to not exhibit high similarity to other TGMs in the newt reference transcriptome; this criterion was applied to minimize the incidence of false “SNPs” derived from paralogous regions; ii) have a minimum sequencing depth of 15 x and minimum
genotype quality of 30 phred; iii) be located at least 60 but not more than 1000 bp from the exon boundary; the latter filter was used because last exons of many genes are long and consist mostly of 3’ untranslated regions (UTR) which are poorly conserved between species; thus the length of such exons in the newt could not be reliably determined and particularly long last “exons” may be artifacts of misassembly [51]. Filtering was performed with a custom Python script. A total of 251 SNPs and their flanking sequences were scored with Illumina Assay Design Tool (ADT) and the Illumina VeraCode GoldenGate Assay was designed for 192 best scoring SNPs. GoldenGate provides codominant genotype data for polymorphic positions with two segregating variants [32]. Genotyping, primary visualization, quality assessment and filtering were performed with Illumina GenomeStudio Data Analysis Software. All loci with cluster separation score and gen train score lower than 0.2 and 0.7, respectively, were excluded from further analysis. We also excluded loci with minor allele frequency (MAF) <1% and less than 90% genotyped individuals.

Population genetics analyses

Allele frequencies for each locus, tests of Hardy–Weinberg equilibrium and tests of linkage disequilibrium (LD) were calculated in GENEPOP 4.1.2 [52]; the type I error was controlled using the false discovery rate (FDR) approach implemented in QVALUE 1.0 [53,54]. Expected heterozygosity ($H_e$), was calculated in the R package adegenet [55]. Allelic richness ($AR$) was calculated in FSTAT 2.9.3.2 [56]. We interpolated geographic gradients in $H_e$ and $AR$ using inverse distance weighting (IDW) in ArcGIS v 10.0 (ESRI, Redlands, CA, USA). Pairwise $F_{ST}$ values between populations and their significance were computed in Arlequin 3.5 [57]. Multidimensional scaling (MDS) was used for visualization of the $F_{ST}$ matrix. Principal component analysis (PCA) was performed in R using adegenet. The significance of correlations between genetic and geographical distances was calculated using the Mantel test implemented in IBDWS [58]. A population tree was constructed in POPTREE2 [59] from the pairwise $F_{ST}$ [60] matrix using the neighbor-joining method. The number of genetic clusters was determined and assignment of individuals to clusters was performed using the Bayesian approach implemented in STRUCTURE 2.3.3 [61–63]. We ran Structure on two separate datasets: $Lm$, which included only morphologically pure $Lm$ populations and $Lm\&Lv$ comprising $Lm$, $Lv$ and syntopic populations. We ran 10 analyses for each $K$ 1-15 for $Lm$ and 10 replicate runs for each $K$ 1–20 for $Lm\&Lv$ dataset. In each case, the admixture model was applied and the runs consisted of 250 000 MCMC burnin steps followed by 1 000 000 post-burnin iterations. We performed inferences under the model of correlated allele frequencies for $Lm$, whereas the uncorrelated model was used for $Lm\&Lv$ dataset, because $Lm$ and $Lv$ populations were expected to be more divergent on average. To determine the most likely number of genetic clusters supported by the data, we calculated $D_K$, a measure of second order rate of change in the likelihood of data [64], using the online software Structure Harvester [65]. Analysis of molecular variance (AMOVA) in Arlequin was used to partition SNP variation into hierarchical levels. Two groupings of populations were used: i) suggested by Structure; ii) supported by the methods based on genetic distances between populations. Significance levels for variance components were estimated using 10 000 permutations.

To identify markers departing significantly from the genome-wide average of differentiation among populations a scan for $F_{ST}$ outliers was performed. In order to minimize the number of false positives the outlier detection was performed under a hierarchical island model [66] in Arlequin. We performed separate scans for differentiation within $Lm$ and for interspecific differentiation. In
each case 50 000 coalescent simulations with 2 groups of 100 demes were performed to obtain the null distribution of $F$-statistics. We selected candidate loci based on $F_{ST}$ or $F_{CT}$ values falling into the 1% upper and lower quantile as suggested by Excoffier et al. [67]. Genes containing outliers were annotated by similarity blastx search against the nr protein database.

**Results**

**Variation**

Genotyping was successful for 139 out of 192 markers (72%) and these were used for population genetic analyses (Table S1).
Genetic structure and diversity in *Lissotriton montandoni*

Genetic differentiation between *Lm* populations varied from negligible (\(F_{ST} < 0.001\)) to strong (\(F_{ST} = 0.408\)) between populations 1 and 24 at the opposite limits of the species distribution (Table S3). The MDS plot of pairwise \(F_{ST}\) revealed two major, genetically differentiated groups of populations with distinct geographic distributions in the northern and southern part of the species range (Fig. 2a). Pairwise \(F_{ST}\) values within groups were similar (averages of 0.100 and 0.098 in the northern and southern group, respectively) and overlapped only slightly with the distribution of pairwise \(F_{ST}\) between populations from different groups (mean 0.271; randomization test \(P < 0.0001\); Fig. S2). Within the northern group two populations appeared distinct from the rest. Notably both are isolated from the continuous part of the species range (Fig. 1). The westernmost locality 1 in the Sudetes Mountains is separated from the main range by the Moravian Gate and locality 6 in the Ukrainian Podolian Upland by the Dniester river. The two groups of populations around the Carpathian Basin is also pronounced, as substantial genetic distance separates the single analyzed population of *Lm* (Fig. 3), MDS (Fig. 5a) and PCA plot (PC3, Fig. 5c). Structuring those in the Carpathian Basin is visible in the population tree (Fig. S1b).

Figure 2. Genetic differentiation within *L. montandoni*. (a) Non-metric two-dimensional scaling of the pairwise \(F_{ST}\) matrix; orange – populations from the northern group; blue – populations from the southern group; (b) Principal Component Analysis (PCA) performed on individual genotypes; in parentheses percentage of variance explained by principal components; orange – individuals from the northern group; blue – individuals from the southern group.

doi:10.1371/journal.pone.0097431.g002

Differentiation within *Lissotriton vulgaris*

Genetic structure among *Lv* populations around the Carpathians is stronger and presumably deeper than that within *Lm*. A deep split between populations outside the Carpathian belt and those in the Carpathian Basin is visible in the population tree (Fig. 3), MDS (Fig. 5a) and PCA plot (PC3, Fig. 5c). Structuring within the Carpathian Basin is also pronounced, as substantial genetic distance separates the single analysed population of *L. vulgaris ampelensis* (locality 38) from two populations of the nominal subspecies *L. vulgaris vulgaris* (Fig. 3, 5a, 5c).
Genetic differentiation and gene flow between species

Strong differentiation between Lm and Lv was detected by the population-based analyses (Table S3, Fig. 3, 5a). AMOVA revealed that 43.5% variation was distributed between species and 13.3% between populations within species (Table 2). Three of seven syntopic populations occupied intermediate positions in the MDS plot (Fig. 5a). However, an overwhelming majority of newts in syntopic populations fell within the range of variation of either one or the other species; only a handful of substantially admixed individuals were analyzed together (Fig. 6, S6). Structure confirmed that in syntopic populations admixture is limited, genotypes of two parental species dominate and significantly admixed individuals are rare. Structure also provided an important insight which was not visible in PCA results: a clearly detectable (>3%) admixture of Lm genes in four Lm populations. Three of these were in the southern part of the Romanian Eastern Carpathians (localities 19, 20, 25) and one was the isolated locality 1 in the Sudetes Mountains; the average admixture in these populations was 8.5%. No admixture of Lm genes was detected in Lv populations. The comparison of the average proportion of admixture on both genetic backgrounds in syntopic populations demonstrated that the mean admixture was very low, ca. 2% and that the proportion of admixture did not differ (U-test, P = 0.24) between Lm and Lm backgrounds.

Detection of outliers

The scan for outliers performed in Lm (locality 10 excluded, see above) revealed 14 FST outliers (10.0%) at the significance level of 0.01: 9 loci (9, 17, 25, 28, 34, 45, 57, 73, 137) showed an excess of differentiation among populations (candidates for local adaptations) and 5 (4, 6, 22, 84, 93) loci were less differentiated than expected under neutrality (candidates for balancing selection) (Fig. 7a). Three FCT outliers (45, 73, 75) were identified as candidates for diversifying selection between the northern and southern Lm groups and two (72, 89) as candidates for balancing selection (Fig. 7b).

Screening for outlier loci between Lm and Lv was performed excluding syntopic populations (26–32). A total of nine (6.5%) FCT outlier loci were identified: eight (15, 72, 79, 112, 116, 117, 128, 129) were more differentiated than expected under the neutral model and one (126) was less differentiated (Fig. 7c). Only one of the interspecific outliers was a nonsynonymous polymorphism. Locus 72 located within gene SRSF1, involved in splicing regulation, was classified as a candidate for balancing selection between the northern and southern Lm groups and at the same time as a candidate for divergent selection between species.

Discussion

Isolation in glacial refugia and limited dispersal determine the genetic structure of Lm

Two clearly differentiated genetic units were identified in Lm by SNP data: the northern group in the Western Carpathians and the western part of the Eastern Carpathians, and the southern group across the rest of the species range. Admixture between them occurs around the Romanian-Ukrainian border. Zielinski et al. [29] identified three units in microsatellite data: our southern group combines their eastern and southern units. Can the discrepancy between the two studies be reconciled? Below we argue that SNPs reflect the history and differentiation better than microsatellites and offer several explanations of the apparent discrepancy between the datasets. We consider the alternative explanation that SNP data fail to detect true differentiation unlikely on several grounds. First, an overwhelming majority of pairwise FST values calculated from SNPs were significant, demonstrating substantial power to detect differentiation. Second, remarkably strong isolation by distance is observed in Lm and the apparent break between the eastern and southern groups of Zielinski et al. [29] coincides with a gap in their sampling. This gap has been filled by the present study. Thus, isolation by distance and non-uniform sampling may have resulted in delineation of the apparently distinct unit [68,69]. Third, the population tree based on SNPs shows a remarkable pattern consistent with colonization from two refugia, but it is difficult to explain under the assumption of expansion from three refugia. While in both groups relationships between some populations are

Figure 3. Relationships among populations. A neighbor-joining tree was constructed from the matrix of pairwise FST; syntopic populations were excluded. Robustness of relationships was tested with 1000 bootstrap replicates. Red – L. montandoni: orange – northern, blue – southern group; green – L. vulgaris: yellow – populations in the Carpathian Basin, violet – populations outside the Carpathian Basin. doi:10.1371/journal.pone.0097431.g003
poorly resolved, in each group several populations are related in a nested fashion, with progressively longer branches at the deeper nesting levels. Nesting involves populations distributed from the center to the periphery of the species range – to the west in the Western Carpathians and to the south-west in the Romanian Carpathians. We hypothesize that the populations with poorly resolved relationships are those inhabiting the refugial areas and sharing most variation retained there. Populations related to each other in a nested fashion would be those which colonized the present range through serial events [4,70]. Taking the evidence together, we propose that the location of the refugium for the northern group was in the Eastern Carpathians close to the Polish-Ukrainian border, and that the refugium for the southern group was in the central part of the Eastern Carpathians in Romania. Species distribution models for the LGM reported by Zie ln´ski et al. [29] are broadly consistent with the proposed location of

Table 2. Results of the Analysis of Molecular variance (AMOVA).

| Source of variation                      | d.f. | Sum of squares | Percentage of variation explained | p       |
|-----------------------------------------|------|----------------|-----------------------------------|---------|
| Two groups within L. montandoni         |      |                |                                   |         |
| Among groups                            | 1    | 1260.054       | 19.52                             | <0.0001 |
| Among populations within groups         | 22   | 1330.68        | 8.16                              | <0.0001 |
| Within populations                      | 548  | 8988.36        | 72.32                             | <0.0001 |
| Two species                             |      |                |                                   |         |
| Among groups                            | 1    | 4019.136       | 43.47                             | <0.0001 |
| Among populations within groups         | 31   | 3545.323       | 13.31                             | <0.0001 |
| Within populations                      | 747  | 10316.708      | 43.22                             | <0.0001 |

AMOVAs were performed for: i) two groups within L. montandoni; ii) two species (excluding syntopic populations).

doi:10.1371/journal.pone.0097431.t002

Figure 4. Genetic structure of L. montandoni inferred by Structure for K=2 groups. For each population pie charts show the fraction of the genes from the northern (orange) and southern (blue) groups.
doi:10.1371/journal.pone.0097431.g004
refugia. The role of the Carpathians as a major refugium for European biota has recently been well documented [42,43,71,72]. Multiple species show genetic differentiation between the Western, Eastern and Southern Carpathians, pointing to the presence of several refugia (reviewed in [29,72]). So far, a refugium in the western part of the Eastern Carpathians has to our knowledge not been proposed.

Expansion from refugia is commonly accompanied by loss of variation [4,73]. Reduction of genetic variation along the postulated expansion routes is visible in our data, but the signal is not strong, and may be distorted locally by introgression from Lv (see below). Hence expansion was apparently not accompanied by severe bottlenecks and thus only a minor fraction of variation has been lost. The strongest reduction in genetic variation occurred in locality 6 in the Podolian Upland isolated from the main portion of the range. This population is a remnant of a geographically remote group of Lm populations [74] which has been hypothesized to be isolated from the main part of the range for several thousand years [75].

Salamanders often exhibit low individual mobility and strong philopatry [27,76]. Genetic differentiation between salamander populations appears to reflect these features although the geographic scale of subdivision differs among species [77,78], which may be related to habitat characteristics [79,80] and to life-history traits [81]. In continuous habitats limited dispersal abilities are likely to generate isolation by distance patterns with a gradient of genetic differentiation among sites, on which larger-scale, hierarchical differentiation reflecting geographic or environmental barriers may be superimposed [80,82]. A comparison of our results with those of a study [83] on fine-scale genetic differentiation in L. vulgaris graecus suggests that a combination of isolation by distance, probably due to limited dispersal, and spatial clustering due to historical fragmentation and/or landscape barriers occurs in Lissotriton newts at both micro- and macroscales.

Local introgression of Lv alleles into the Lm nuclear genome is detectable with SNP markers

A major finding of the present study is substantial introgression of Lv nuclear alleles into some Lm populations. This is contrary to the findings of Zielinski et al. [29] who detected very little recent nuclear introgression in either direction. One likely explanation for the difference between the studies is the number of markers employed [84]. While we analyzed 139 unlinked SNPs, inference about introgression in the previous study was based on only 10 microsatellites. The observed discrepancy does not result from differences in sampling because three of four admixed populations were analysed in both studies. As SNP markers were discovered in a sample of Lm individuals, our study did not use diagnostic markers. This could be considered a weakness if viewed from the perspective of classical studies of hybrid zones which usually employed a limited number of diagnostic markers. However, because of the widespread genomic heterogeneity of interspecific gene flow [38,40,85,86], such diagnostic markers may constitute a highly nonrandom sample of the genome, enriched in genomic regions strongly differentiated between species. In our opinion randomly selected polymorphisms are better suited for an unbiased assessment of introgression. We acknowledge that ideally
both species should be included in the discovery panel; this would however limit the number of polymorphic loci useful for the assessment of genetic structure within \textit{Lm}. The current study demonstrates two peculiar features of \textit{Lm} x \textit{Lv} hybridization. First, appreciable (>3\%) introgression was detectable only locally, in four of 25 sampled \textit{Lm} localities. In these populations most individuals were introgressed and the average admixture of \textit{Lv} genes was 8.5\%. Second, in the introgressed \textit{Lm} populations, admixture was stronger than in seven syntopic localities, where it was barely detectable. Thus current syntopy, even if it leads to occasional hybridization, as shown by a single putative F1 hybrid, does not necessarily cause introgression. This is somewhat surprising because a study of a \textit{Lm}/\textit{Lv} hybrid zone at microscale detected strong assortative mating but also found that syntopy was almost universally accompanied by some admixture \cite{44}. As the four admixed \textit{Lm} populations testify, nuclear introgression of \textit{Lv} alleles into \textit{Lm} populations extends beyond syntopy, but does not permeate into the core of the \textit{Lm} range.

Local differences in the extent of introgression may be explained by several mechanisms. The introgressed populations may be simply located at the tails of local hybrid zones, and would thus be sampled entirely by chance. However other potential explanations deserve consideration. Local ecological conditions may either favor introgression or delay removal of introgressed alleles by selection \cite{87}. Differences in abundance of species in a breeding locality may force the rarer species to hybridize due to scarcity of conspecific mates, but we have not observed this effect in syntopic populations. If species are genetically structured, as in our case, introgression may be easier between some genetic groups if their genomes harbor fewer incompatible alleles and thus intrinsic selection against hybrids is weaker or ecological/sexual adaptations are similar \cite{24,25}. \textit{Lv} is strongly differentiated genetically \cite{41,88,89} and various \textit{Lv} groups come into contact with \textit{Lm} populations in the Carpathian Basin, and outside the Carpathian belt. If introgression is neutral, the observed pattern may result from expansion-related phenomena \cite{4}. Under the scenario modeled by Currat et al. \cite{90}, when one species invades the range of another, neutral introgression occurs almost exclusively from the resident to the invading species. Thus, local expansion of \textit{Lm} would bring \textit{Lv} genes onto its genetic background. A comparison of the two isolated \textit{Lm} populations may be instructive in this respect. Population 1 at the western margin of the species range, probably the result of postglacial or more recent expansion, has recently introgressed \textit{Lv} mtDNA and shows clear evidence of nuclear introgression. Another isolated population (6), close to the postulated refugial area of the northern \textit{Lm} group and possibly surviving in situ for a long time, shows no trace of nuclear introgression. Scenarios related to the Currat et al. \cite{90} model were favored as the explanation of mtDNA introgression and replacement in \textit{Lm} by Zielinski et al. \cite{29}.

In addition to laboratory experiments which are difficult to perform in this system due to logistic reasons, two other kinds of

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure6.png}
\caption{Genetic differentiation and admixture between \textit{L. montandoni} and \textit{L. vulgaris} inferred by Structure for $K=2$ groups. For each population pie charts show the fraction of \textit{L. montandoni} (red) and \textit{L. vulgaris} (green) genes. doi:10.1371/journal.pone.0097431.g006}
\end{figure}
analyses would be informative with respect to the causes of the apparent differentiation in the extent of introgression. Examination of several transects through hybrid zones in the context of local environmental conditions and relative species abundance could be informative as demonstrated in multiple systems [87,91–93]. Another important way forward would be to use multilocus sequence data [94] to construct and test multipopulation models of gene flow between \( Lm \) and \( Lv \). Models distinguishing two groups within \( Lv \), inside and outside of the Carpathian basin, as well as two groups within \( Lm \) can be evaluated and hypotheses regarding the timing and extent of gene flow may be tested within an Approximate Bayesian Computations framework [95,96]. This approach would provide a longer-scale perspective on gene flow between species and its spatial and temporal variation.

Genomic heterogeneity of gene flow within and between species

Outlier loci were detected both within \( Lm \) and between \( Lm \) and \( Lv \). Such candidate loci may signal various forms of selection acting on the markers themselves or at linked sites [24,38]. Alternatively, their apparent outlier status may result from violation of the model assumptions, to which the available methods are very sensitive [13]. We do not attempt a formal functional analysis of the identified outliers but rather emphasize that the outliers detected in the \( Lm-Lv \) comparison indicate heterogeneity of interspecific gene flow in nuclear protein coding genes. Dramatic discordance in the propensity for interspecific gene flow occurs between the mitochondrial and nuclear genome ([29]; this study). Within the nuclear genome the genes of MHC class II introgress easily between the two species [45]. The present study suggests that heterogeneity of gene flow is widespread in the nuclear genome. Some genomic regions, typically linked to genes involved in intergeneric incompatibilities or underlying speciesspecific adaptations, i.e. genes which may cause reduced hybrid fitness, acquire reproductive isolation earlier than other regions [38,97,98]. The size of such regions and mechanisms responsible for maintenance of genomic differentiation have been a subject of ongoing controversy and intense recent research [85,98–101]. It is expected that the shape of the heterogeneity in gene flow will evolve over time and a comparison of the extent of heterogeneity at various stages of divergence is of great interest for the understanding of the buildup of genomic divergence as differentiation progresses [40,86]. Transcriptome data, such as those used here for the development of SNP markers, are being applied to study genomic heterogeneity of gene flow in the \( Lm-Lv \) system (Stuglik et al. in prep.).

Conclusions

Using a panel of transcriptome-derived SNP markers, our study has demonstrated that isolation in glacial refugia and limited dispersal have been the main factors determining the genetic structure of \( Lm \). Two substantially differentiated groups of \( Lm \) populations likely originated from separate refugia, both located in the Eastern Carpathians. The colonization of the present range in north-western and south-western directions was accompanied by a modest loss of variation. Local introgression of \( Lv \) alleles into several \( Lm \) populations was detected. Introgression was higher in
Lm populations classified as admixed than in syntopic populations. We discuss the possible causes of this discrepancy and propose further tests to distinguish between alternatives. Several outliers were identified in tests of interspecific differentiation, suggesting genomic heterogeneity at various stages of species divergence is of major interest. For the understanding of the buildup of differentiation across the genome and Lm/Lv is a promising study system in this respect.

Supporting Information

Figure S1 Expected heterozygosity. (a) interpolated geographic gradients in L. montandoni (Lm), (b) means and standard deviations (SD) of the ln-likelihood of the probability of data for various values of K. (TIF)

Figure S2 Histograms showing the distribution of pairwise FST between populations within the northern and southern L. montandoni groups and between groups. (TIF)

Figure S3 Identification of the number of groups (K) in Structure analysis for L. montandoni. (a) Evanno et al. (2005) method; (b) means and standard deviations (SD) of the ln-likelihood of the probability of data for various values of K. (TIF)

Figure S4 Interpolated geographic gradients of allelic richness in L. montandoni. (TIF)

Figure S5 Isolation by distance in L. montandoni. Relationships between pairwise FST and log-geographic distances are presented for all populations and populations within northern and southern groups separately. (TIF)

References

1. Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science 236: 787–792.
2. Avise JC (2000) Phylogeography. The history and formation of species. Cambridge-London: Cambridge University Press.
3. Manel S, Holderegger R (2013) Ten years of landscape genetics. Trends in Ecology and Evolution 28: 614–621.
4. Escorfer L, Foil M, Petit RJ (2009) Genetic consequences of range expansions. Annual Review of Ecology, Evolution, and Systematics 40: 481–501.
5. Charlesworth B, Charlesworth D, Barton NH (2003) The Effects of Genetic and Geographic Structure on Neutral Variation. Annual Review of Ecology, Evolution, and Systematics 34: 99–125.
6. Bohonak AJ (1999) Dispersal, gene flow, and population structure. Quarterly Review of Biology 74: 21–45.
7. Hewitt GM (2011) Quaternary phylogeography: The roots of hybrid zones. Genetics 139: 617–630.
8. Foll M, Gaggiotti O (2006) Identifying the environmental factors that determine the genetic structure of populations. Genetics 174: 875–891.
9. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, et al. (2002) Genetic structure of human populations. Science 298: 2301–2305.
10. Frankham R, Ballou JD, Briscoe DA (2010) Introduction to Conservation Genetics. Cambridge-London: Cambridge University Press.
11. Beaumont MA (2005) Adaptation and speciation: what can Fst tell us? Trends in Ecology & Evolution 20: 435–440.
12. Storey JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. Molecular Ecology 14: 677–688.
13. Narum SR, Hess JE (2011) Comparison of FST outlier tests for SNP loci under selection. Molecular Ecology Resources 11: 184–194.
14. Fream M, Rainey PB, Traulsen A (2013) The effect of population structure on the rate of evolution. Proceedings of the Royal Society B: Biological Sciences 280.
15. Wakeley J (2001) The coalescent in an island model of population subdivision with variation among demes. Theoretical Population Biology 59: 153–144.
16. Ingvarsson PK (2004) Population subdivision and the Hudson-Kreitman-Aguade test: Testing for deviations from the neutral model in organelle genomes. Genetic Research 83: 31–39.
17. Nielsen R (2005) Molecular signatures of natural selection. Annual Review of Genetics 39: 197–218.
18. Stadler T, Haubold B, Merino C, Stephan W, Pfafflhuber P (2009) The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. Genetics 182: 205–216.
19. Hickerson MJ, Carstens BC, Cavinder-Bares J, Grant AL, Graham CH, et al. (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. Molecular Phylogenetics and Evolution 54: 291–301.
20. Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. Evolution 53: 1898–1914.
21. Slatkin M (1995) Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47: 264–279.
22. Landguth EL, Cushman SA, Schwartz MK, McKelvey TS, Murphy M, et al. (2010) Quantifying the lag time to detect barriers in landscape genetics. Molecular Ecology 19: 4179–4191.
23. Stewart JR (2009) The evolutionary consequence of the individualistic response to climate change. Journal of Evolutionary Biology 22: 2363–2375.
24. Bierne N, Welch J, Lloret E, Bonhomme F, David P (2011) The coupling hypothesis: Why genome scans may fail to map local adaptation genes. Molecular Ecology 20: 2044–2072.
25. Corbett-Detig RB, Zhou J, Clark AG, Hardil DL, Ayroles JF (2013) Genetic incompatibilities are widespread within species. Nature 504: 135–137.
26. Kiel Y, Timothy G, Barraclough TG (2010) Speciation has a spatial scale that depends on levels of gene flow. American Naturalist 175: 316–334.
27. Smith MA, Green DM (2005) Dispersal and the metapopulation paradigm in amphibian ecology and conservation: Are all amphibian populations metapopulations? Ecography 28: 110–128.
28. Vences M, Wake DB (2007) Speciation, species boundaries and phylogeography of amphibians. In: Heatwole H, Tyler M, editors. Amphibian Biology, Vol...
43. Tzedakis PC, Emerson BC, Hewitt GM (2013) Cryptic or mystic? Glacial tree

45. Nadachowska-Brzyska K, Zielinski P, Radwan J, Babik W (2012) Interspecific

40. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

39. Sousa V, Hey J (2013) Understanding the origin of species with genome-scale

37. Rosenblum EB, Novembre J (2007) Ascertainment bias in spatially structured

35. Paschou P, Ziv E, Burchard EG, Choudhry S, Rodriguez-Cintron W, et al. (2009) PCA-correlated SNPs for structure identification in worldwide human populations. PLoS Genetics 3: 1672–1686.

33. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29: 644–652.

58. Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service.

57. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of

56. Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F- statistics. Journal of Heredity 86: 485–486.

55. Rafinski J, Bogacka B, Babik W (2003) Comparison between single nucleotide polymorphisms and microsatellites. Journal of Heredity 106: 158–171.

54. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. PNAS USA 100: 9440–9445.

53. Storey JD, Pelig B, Przyropczynski Studentoww Uniwersytetu Jagiellońskiego.

52. Rousset F (2008) GENEPOP’007: A complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources 8: 103–106.

51. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME algorithm/Map format and SAMtools. Bioinformatics 25: 2078–2079.

50. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

48. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. Science 319: 1100–1104.

47. Sommer RS, Nadachowska A (2006) Glacial refugia of mammals in Europe: Evidence from fossil records. Mammal Review 36: 87–112.

46. Landguth EL, Fedy BC, Oyler-McAnce SJ, Garey AL, Emel SL, et al. (2012) Fine scale spatial genetic structure of two syntopic newts across a hybrid zone. BMC Evol. Biol. 12: 207–220.

45. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

44. Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. Trends in Ecology & Evolution 23: 494–498.

43. Tzedakis PC, Emerson BC, Hewitt GM (2013) Cryptic or mystic? Glacial tree

42. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

41. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

40. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

39. Sousa V, Hey J (2013) Understanding the origin of species with genome-scale

38. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. PNAS USA 100: 9440–9445.

37. Rosenblum EB, Novembre J (2007) Ascertainment bias in spatially structured

36. Slatkin M, Voelm L (1991) FST. In: Chapman & Hall.

35. Paschou P, Ziv E, Burchard EG, Choudhry S, Rodriguez-Cintron W, et al. (2009) PCA-correlated SNPs for structure identification in worldwide human populations. PLoS Genetics 3: 1672–1686.

34. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29: 644–652.

33. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29: 644–652.

32. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of

31. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME algorithm/Map format and SAMtools. Bioinformatics 25: 2078–2079.

30. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of

29. Zelinski P, Nadachowska-Brzyska K, Wielstra B, Szokat R, Covaci-Marcov SD, et al. (2013) No evidence for nuclear introgression despite complete mtDNA replacement in the Carpathian newt (Lasiotriton montandoni). Molecular Ecology 22: 1884–1903.

28. Brumfield RT, Buerki P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology and Evolution 18: 249–256.

27. Helyar SJ, Hemmer-Hansen J, Bekkivolt D, Taylor MI, Ogden R, et al. (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. Molecular Ecology Resources 11: 123–136.

26. Fan J, Chee MS, Gunderson KL (2006) Highly parallel genomic assays. Nature Reviews Genetics 7: 632–644.

25. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

24. Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. Trends in Ecology & Evolution 23: 494–498.

23. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

22. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

21. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

20. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

19. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

18. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

17. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

16. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

15. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

14. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

13. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

12. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

11. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

10. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

9. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

8. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

7. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

6. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

5. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

4. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

3. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

2. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

1. Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species

...
90. Curat M, Rueci M, Petit RJ, Excoffier L (2008) The hidden side of invasions: Massive introgression by local genes. Evolution 62: 1908–1920.

91. Dowling TE, Broughton RE, Demarais BD (1997) Significant role for historical effects in the evolution of reproductive isolation: Evidence from patterns of introgression between the cyprinid fishes, Luxilus cornutus and Luxilus chrysocephalus. Evolution 51: 1574–1583.

92. Aboim MA, Mavaře J, Bernatchez L, Coelho MM (2010) Introgressive hybridization between two Iberian endemic cyprinid fish: A comparison between two independent hybrid zones. Journal of Evolutionary Biology 23: 817–828.

93. Teeter KC, Thibodeau LM, Gompert Z, Buerkle CA, Nachman MW, et al. (2010) The variable genomic architecture of isolation between hybridizing species of house mice. Evolution 64: 472–485.

94. Zielinski P, Stuglik MT, Dudek K, Konczal M, Babik W (2014) Development, validation and high-throughput analysis of sequence markers in nonmodel species. Molecular Ecology Resources 14: 352–360.

95. Beaumont MA (2010) Approximate Bayesian computation in evolution and ecology. Annual Review of Ecology, Evolution, and Systematics 41: 379–406.

96. Nadachowska-Brzyska K, Burri R, Olason PI, Kawakami T, Smeds L, et al. (2015) Demographic divergence history of Pied Flycatcher and Collared Flycatcher Inferred from Whole-Genome Re-sequencing Data. PLoS Genetics 9.

97. Wu CI, Ting CT (2004) Genes and speciation. Nature Reviews Genetics 5: 114–122.

98. Nosil P, Feder JL (2012) Genomic divergence during speciation: Causes and consequences. Philosophical Transactions of the Royal Society B: Biological Sciences 367: 332–342.

99. Noor MAF, Bennett SM (2009) Islands of speciation or mirages in the desert Examining the role of restricted recombination in maintaining species. Heredity 103: 439–444.

100. Lawniczak MKN, Emrich SJ, Holloway AK, Regier AP, Olson M, et al. (2010) Widespread divergence between incipient Anopheles gambiae species revealed by whole genome sequences. Science 330: 312–314.

101. Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Philosophical Transactions of the Royal Society B: Biological Sciences 367: 451–460.