Divergence in the Face of Gene Flow: The Case of Two Newts (Amphibia: Salamandridae)

Krystyna Nadachowska and Wiesław Babik†‡*

*Jagiellonian University, Institute of Environmental Sciences, Kraków, Poland; †Department of Community Ecology, Helmholtz Centre for Environmental Research—UFZ, Halle/Saale, Germany; and ‡Institute of Systematics and Evolution of Animals, Polish Academy of Sciences, Kraków, Poland

Understanding the process of divergence requires the quantitative characterization of patterns of gene flow between diverging taxa. New and powerful coalescent-based methods give insight into these processes in unprecedented details by enabling the reconstruction of the temporal distribution of past gene flow. Here, we use sequence variation at eight nuclear markers and mitochondrial DNA (mtDNA) in multiple populations to study diversity, divergence, and gene flow between two subspecies of a salamander, the smooth newt (*Lissotriton vulgaris kosswigi* and *Lissotriton vulgaris vulgaris*) in Turkey. The ranges of both subspecies encompass mainly the areas of this important glacial refugial area. Populations in refugia where species have been present for a long time and differentiated in situ should better preserve the record of past gene flow than young populations in postglacial expansion areas. Sequence diversity in both subspecies was substantial (nuclear π_{sil} = 0.69% and 1.31%). We detected long-term demographic stability in these refugial populations with large effective population sizes (N_e) of the order of 1.5–3 $\times$ 10^5 individuals. Gene trees and the isolation with migration (IM) analysis complemented by tests of nested IM models showed that despite deep, pre-Pleistocene divergence of the studied newts, asymmetric introgression from *vulgaris* to *kosswigi* has occurred, with signatures of recent gene flow in mtDNA and an anonymous nuclear marker, and evidence for more ancient introgression in nuclear introns. The distribution of migration times raises the intriguing possibility that even the initial divergence may have occurred in the face of gene flow.

Introduction

One of the long-standing questions in evolutionary biology is how populations diverge over space and time and form distinct species (Coyne and Orr 2004). At one extreme, there are situations in which descendent populations differentiated in full isolation (approximated by the isolation model), at the other, a constant gene flow could have occurred between diverging populations (the migration model) (Hey 2006). It is increasingly appreciated, however, that these two extremes are linked via a continuum of intermediate situations where gene flow may be restricted temporally, spatially, and may occur in some parts of the genome but not in others (Wu and Ting 2004; Mallet 2007; Yatabe et al. 2007). Therefore, an understanding of the process of divergence, and ultimately speciation, requires the quantitative characterization of patterns of gene flow, particularly the temporal distribution of migration events (Won and Hey 2005; Niemiller et al. 2008; Strasburg and Rieseberg 2008). Recently developed models of divergence with gene flow or isolation with migration (Nielsen and Wakeley 2001; Hey and Nielsen 2004; Hey 2005; Hey and Nielsen 2007) provide extremely powerful tools for the study of divergence processes and provide quantitative assessments. These methods stem from the observation that under different divergence scenarios the expected patterns of differentiation observed in the genomes of diverging populations or taxa differ, making it possible to test hypotheses on patterns of gene flow (Nielsen and Wakeley 2001; Hey and Nielsen 2004).

However, the histories of various parts of the genome may be dramatically discordant. First, stochasticity inherent to the coalescent process produces large variance in the depth of genealogies, even for neutrally evolving loci. This historical legacy, through the process of incomplete lineage sorting, may result in discordance between gene trees and population/species trees even in the absence of gene flow (Edwards and Beerli 2000; Hudson and Turelli 2003). Second, parts of the genome can experience distinct levels of gene flow depending on the strength of selection counteracting genetic exchange in particular genomic regions (Wu and Ting 2004; Baack and Rieseberg 2007). Distinguishing between lineage sorting and introgression, although possible, may be difficult in principle and in practice and requires genealogical information from multiple unlinked loci. Coalescent-based multilocus approaches enable the simultaneous estimation of multiple parameters of interest including divergence time, rates of historical gene flow, and effective population sizes explicitly considering genealogical information and taking into account uncertainty related to the genealogy estimation (Becquet and Przeworski 2007; Hey and Nielsen 2007; Kuhner and Smith 2007).

The application of the multilocus approach in the study of divergence, speciation, and historical demography in closely related species or differentiated populations has been heavily biased toward a privileged handful of species such as *Drosophila* (Machado et al. 2002; Hey and Nielsen 2004) or humans and other primates (Hey 2005; Won and Hey 2005; Thalmann et al. 2007). So far, the divergence population genetics of nonmodel species has received little attention, even among vertebrates (e.g., Dolman and Moritz 2006; Niemiller et al. 2008).

In the present study, we investigated the divergence population genetics of a salamander, the smooth newt (*Lissotriton vulgaris*), in Turkey. We emphasized the quantitative characterization of historical and contemporary gene flow between two subspecies in Turkey: *Lissotriton vulgaris vulgaris* and *Lissotriton vulgaris kosswigi* (further *vulgaris* and *kosswigi*, respectively), which differ strikingly in male epigamic traits. Both subspecies occur parapatrically in western Turkey, one of the putative glacial refugia of this

Key words: *Lissotriton vulgaris*, divergence, gene flow, nucleotide variation, isolation with migration.

E-mail: wieslaw.babik@uj.edu.pl.

Mol. Biol. Evol. 26(4):829–841. 2009
doi:10.1093/molbev/msp004
Advance Access publication January 8, 2009

Table 1
Markers Used in the Study

Marker	Abbreviation	Length (bp)	Category	Primers (5′->3′)
Anonymous noncoding genomic DNA fragment	*Tva4*	667	Anonymous	F: ACAGTGCAAATGCGTACAATTC; R: AGCAAGGATCTGCTCAAGAAAC
Calreticulin, fourth intron	*Calr*	552	Intron	F: GGTGAGTGTCTGGGACTATGC; R: TTTTGCAAGCCAATCTTGGTA
Chemokine receptor 4 gene	*Cxcr4*	467	Coding	F: GGACTATGACCGACAAGTACMG; R: GGTGATGTAGTAGGGCARCCA
Beta fibrinogen, seventh intron	*Fib*	423	Intron	F: GCAAAGAATGAGAGCATTGGC; R: GACATTGAAATTTAGCAAGCACA
Growth hormone, fourth intron	*Gh*	708	Intron	F: TCTCATCAAGGTGAGTTTGAACA; R: CCTTCTTGTGTCAGAGGTGCTAT
Sodium–calcium exchanger 1	*Ncx1*	598	Coding	F: AGCGATATAGAGCTGGGAAGC; R: TGCTGTCTGGGAGTTGACTTT
Orthodenticle homolog 2, second intron	*Otx2*	586	Intron	F: CGCTATCCGGACATCTTTATGA; R: GGCTCGCCAGTAATGTTGTAA
Platelet derived growth factor receptor chain alpha, 11th intron	*Pdgfr*	747	Intron	F: TGCAGCTGCCATATGACTCTA; R: TACGCTGTTCCTTCAACCACT
NADH dehydrogenase subunit 2 gene	*Nd2*	1,224	Coding	F: TCGAACCTACCCTGAGGAGAT; R: TCTGGGTTGCATTCAGAAGA

species. The distribution of *kosswigi* is restricted to the southern coast of the Black Sea near the Bosphorus, whereas *vulgaris* has a broader distribution in Europe including both formerly glaciated and unglaciated regions and is genetically highly differentiated across its range (W. Babik et al. 2005). However, *vulgaris* in Turkey (also referred to as separate subspecies *Lissotriton vulgaris schmidtlerorum*, Schmidtler and Franzen 2004) has been shown to comprise a distinct and ancient mitochondrial DNA (mtDNA) clade indicative of a long and idiosyncratic history, substantiating the treatment of Turkish populations as a separate group. This mtDNA clade extends further to Bulgaria (W. Babik, unpublished data), effectively preventing contact of *kosswigi* with any other lineage of *L. vulgaris*. MtDNA sequence data suggest pre-Pleistocene divergence of *kosswigi* and *vulgaris* (Babik et al. 2005).

The limited ranges of both subspecies encompass mainly the areas of the postulated glacial refugia. Areas in which species were constantly present during long periods of unfavorable climatic conditions during the Pleistocene are of special interest for analyses of population differentiation and speciation because they harbor particularly diverse and in many cases divergent populations (Hewitt 2004; Martinez-Solano et al. 2006; Gómez and Lunt 2007). Therefore, it is less likely that dramatic range and demographic changes caused by extinction–recolonization cycles driven by the Pleistocene climatic oscillations would have erased the record of processes that have occurred since the initial divergence. In particular, diverging populations in these areas should better reflect historical gene flow, because introgressed variants should have had more time to spread over broad areas (Hofreiter et al. 2004). The apparent long-term occurrence of both subspecies at the southern fringe of the *L. vulgaris* distribution indicates that they could have differentiated in situ and leads to the testable hypothesis that their differentiation had indeed occurred in the face of gene flow.

We characterized sequence variation at nine loci and applied coalescent-based methods to investigate the patterns of differentiation in various parts of the genome of Turkish populations of the smooth newt. We were particularly interested in testing the symmetry of gene flow and temporal distribution of migration as such data enable us to address the hypothesis that differentiation of these newts has occurred in the face of gene flow and provide a quantitative picture of their divergence. We also describe multilocus nucleotide diversity within and between subspecies, assess genetic structure of populations and reconstruct historical demography.

Materials and Methods

Sampling

We collected tail tips of adult or larval newts from 20 populations of *vulgaris* (213 individuals) and from 6 populations of *kosswigi* (52 individuals) spanning the entire ranges of both subspecies in Turkey (fig.1, supplementary table S1, Supplementary Material online).

Laboratory Analyses

One mtDNA fragment and eight nuclear loci were amplified and sequenced (table 1). Details of laboratory procedures are given in online Supplementary Materials. Haplotypes were inferred using Clark's (1990) method where haplotypes identified in homozygous individuals are subtracted from heterozygotes to reveal the sequence of the other allele. In cases where no reliable inference was possible and for polymerase chain reaction (PCR) products heterozygous with respect to indels, specific primers were designed to amplify and sequence individual alleles. A few particularly difficult samples were cloned. We sequenced from 4 to 10 clones per PCR product.

Nucleotide Diversity and Phylogenetic Analyses

Standard population genetic analyses of the sequence polymorphism were performed using DnaSP (Rozas et al. 2003) and MEGA4 (Tamura et al. 2007). Indels were

excluded from the analyses. We calculated haplotype diversity (h), nucleotide diversity (π), and Watterson's θ for both subspecies as well as for each population. For coding genes, we considered both total and silent variation. Pairwise comparisons between sequences from different populations were used to compute $\pi_{between}$.

DT-ModSel (Minin et al. 2003) was used to select models of sequence evolution for individual loci. Using these models, we constructed maximum likelihood trees with PHYML 2.4.4 (Guindon and Gascuel 2003) and assessed their robustness with 1,000 bootstrap (BS) replicates. Bayesian phylogenetic trees were constructed with MrBayes 3.1 (Ronquist and Huelsenbeck 2003). Two independent runs of four Metropolis Coupled Monte Carlo Markov Chains each (three of them "heated," temperature = 0.20) were run for 1.1×10^6 generations and sampled every 1,000 generations. The first 100 trees were discarded as burn-in, resulting in 2,000 sampled trees. To calculate the posterior probability (PP) of each bipartition, the majority-rule consensus tree was computed from the 2,000 sampled trees.

Testing Assumptions of the IM Model

Isolation with migration models of divergence (IM and IMa) assume that a single ancestral population split into two descendent populations t generations ago and since then the populations may have, or may have not, been subject to gene exchange (Hey and Nielsen 2004). The method uses Markov chain Monte Carlo (MCMC) sampling of gene genealogies to estimate PP distributions of six parameters: mutation-scaled effective population sizes of the ancestral (Θ_A) and the descendant populations (Θ_1 and Θ_2), rates of migration in either direction (m_1 and m_2), and the time of divergence (t).

The models were derived under several assumptions (Hey and Nielsen 2004, 2007). Because almost every empirical data set is likely to violate some of them, testing assumptions of a model is always desirable and when violations are detected, their possible impact on the results should be discussed. In our opinion, the assumption of the lack of population structure on the inferences under IM models requires some comments.

A general tool for studying subdivided populations is the structured coalescent (Notohara 1990; Wilkinson-Herbots 1998), which needs to be applied when migration rates fall below the reciprocal of the deme size (Wakeley and Aliacar 2001; Charlesworth et al. 2003). The generality of the structured coalescent comes at the cost of its high complexity making it analytically intractable in all but the simplest cases and necessitates the use of custom simulations (e.g., Knowles and Carstens 2007), conditioned on many often unknown parameters (Charlesworth et al. 2003). These difficulties prompted interest in investigating the conditions in which standard coalescent approximation may be applied to study cases with nonnegligible population structure (Charlesworth et al. 2003). In a series of papers, Wakeley and colleagues (Wakeley 1999, 2000, 2001, 2004; Wakeley and Aliacar 2001) have shown that the standard coalescent emerges in a broad class of metapopulation models provided that the number of demes is high. This is because the genealogy of the sample can be divided into a short scattering phase (when no mutations occur) at the end of which each ancestral lineage finds itself in a different deme, and a much longer collecting phase, which dominates the history of the sample and which is described by the standard coalescent process, because ancestral lineages become exchangeable. However, the rate of this process is determined by the rescaled, metapopulation effective population size (N_e) that depends on several factors, particularly the rate of among-deme migration. The effects of these parameters cannot be separated (Wakeley 2001). The complex relationship between N_e and population subdivision (Whitlock and Barton 1997; Wakeley 2000; Charlesworth et al. 2003) makes researchers reluctant to interpret "distorted" N_e estimates (Niemiller et al. 2008). However, if we consider the evolutionary history of a metapopulation or, even more generally, of a group of populations connected by appreciable migration, precisely this composite N_e is of interest because it determines the rate of neutral processes (Charlesworth et al. 2003).

In the following paragraphs, we describe the procedures used to test assumptions of the IMa model.

Lack of Substantial Population Structure

Sequence variation was partitioned into between subspecies, among populations within subspecies, and within-population components with the analysis of molecular variance (AMOVA) in Arlequin (Excoffier at al. 2005). Population differentiation was also investigated by estimating K_{ST}* (Hudson et al. 1992) and the nearest neighbor statistic S_{nn} (Hudson 2000) using Hudson's programs (http://home.uchicago.edu/~rhudson1/); pairwise statistics were calculated for *vulgaris* and *kosswigi* as well as for each pair of sampled populations. The significance of the AMOVA components and values of K_{ST}* and S_{nn} were tested with 10,000 permutations.

No Selection

Tajima's (1989) and Fu and Li's (1993) D tests were performed with DnaSP for both subspecies and for each population separately. We also performed multilocus Tajima D tests at the subspecies and population level using the HKA program (http://lifesci.rutgers.edu/~heylab/heylabsoftware.htm#HKA). McDonald–Kreitman (M–K; McDonald and Kreitman 1991) and heterogeneity tests (Hahn et al. 2002) were used to search for departures from neutrality in coding loci. The latter is based on the assumption that selection and demography have different effects on synonymous and nonsynonymous polymorphisms and is a simple modification of standard D tests. To examine whether sequence polymorphism within and sequence divergence between subspecies are correlated, a multilocus Hudson-Kreitman-Aguadé (HKA) test (Hudson et al. 1987) was conducted in the HKA program using all sequences from both subspecies. It should be noted that purifying selection does not pose a problem for the IM model—because effectively it only decreases the overall mutation rate. On the other hand, directional and balancing selection invalidate inference under the IM model.

No Recombination within Loci

The minimum number of recombination events (R_{min}) was estimated based on the four-gamete criterion, whereas the population recombination parameter $\rho = 4N_e r$, where r is the per-locus recombination, was computed with the composite likelihood method (Hudson 2001). Both statistics are implemented in LDhat 2.0 (McVean et al. 2002). In cases when recombination within loci is detected, it is common practice to extract maximally informative blocks of nonrecombining sequences for individual markers and use them in the analysis; we used the IMgc program for this purpose (Woerner et al. 2007).

Demographic Stability

An additional assumption of the IMa model is the demographic stability of diverging populations (Hey and Nielsen 2007). In order to test for signals of demographic growth/shrinkage, we estimated the exponential growth rate (g) with the Bayesian version of Lamarc (Kuhner and Smith 2007). Separate analyses for *vulgaris* and *kosswigi* were performed on recombination-filtered data (see above) and on several partial data sets. The subsets were distinguished to check whether different sampling scheme and population structure influence the estimates. The first data subset included two sequences randomly sampled per population per locus. To create the second data subset, we excluded mtDNA and randomly sampled two sequences per population per nuclear locus. In the third data subset, the most distinct samples from *kosswigi* (Alemdar, pop 24) and *vulgaris* (Keşan, pop 20) were treated as separate populations. We ran long chains of 1–2 million steps sampled every 10th step after discarding the first 10,000 steps as burn- in for each data subset. We ran the program multiple times with different seed numbers to check for consistency of estimates.

IMa Analysis

The estimation of demographic parameters, migration rates as well as number and times of migration events was carried out with the IMa program (Hey and Nielsen 2007). In contrast to previous versions (IM, Hey and Nielsen 2004), demographic parameters are not included in the MCMC and are integrated out analytically; thus, IMa has better mixing properties.

IMa analyses were performed on partial data sets. It is reasonable to perform coalescent analyses on smaller data subsets because increasing sample sizes does not usually improve parameter estimates while considerably increasing computational burden. To evaluate the effect of different sampling schemes and various levels of population structure on parameter estimates, we analyzed several subsets of the data. First, we randomly sampled two sequences per locus from "each locality" creating species-wide samples. Second, we randomly sampled 15 sequences from "each subspecies." Third, we sampled just one sequence per locus from each locality. Each sampling procedure was repeated at least twice. We also performed analysis with only nuclear loci (species wide samples). Analyses assuming no migration between populations were performed on three data sets: 1) with all markers with two sequences sampled per locus from each locality, 2) with nuclear loci, and 3) with species wide samples excluding markers having signatures of recent introgression (mtDNA and *Tva4*). Additional IMa analyses were performed to assess the possible effect of nonnegligible genetic structure within both subspecies on estimates of the model parameters. The following four groups were distinguished on the basis of pairwise K_{ST}*: 1) all *vulgaris* populations except Keşan, 2) all *kosswigi* populations except Alemdar, 3) Keşan (pop 20), and 4) Alemdar (pop 24).

In order to obtain reliable estimates of population parameters, we designed IMa runs differing in heating schemes and number of chains (up to 200). We used a geometric heating scheme and a burn-in of 100,000 steps and run MCMC under the HKY model of sequence evolution. Theta scalars and maximum values for time of population splitting and migration rates were chosen experimentally by monitoring ASCII curves. The total number of genealogies saved during each run varied between 10,000 (runs of 1 million steps) and 100,000 (runs of 10 million steps). We checked the mixing properties of MCMC by monitoring effective sample size (ESS) values of parameter t, trend-line plots of the parameter, and swapping rates between chains. Long, well mixing runs were repeated at least twice with different random seed numbers. If these independent runs generated similar posterior distributions, we considered analyses to have converged on a stationary distribution.

Estimating Mutation Rates

Because both splitting time and effective population size parameters are given in mutational units, estimates of the absolute mutation rates for individual loci are needed to convert these into years and number of individuals. To this end, we estimated mutation rates as follows: Divergence times between the smooth newt and other newt species were derived by applying two calibrations. First, we used three fossil-based calibration points given by Steinfartz et al. (2007) to date the *Nd2*-based phylogeny of newts: C2: *Pleurodeles–Tylototriton* split at 44 million years ago (Ma), C3: *Notophthalmus–Taricha* split constrained as not younger than 22 Ma, and C4: *Triturus cristatus–Triturus marmoratus* split at 24 Ma. Note that in our phylogeny, the order of species is (((*vulgaris*, *italicus*), *helveticus*), *boscai*) and not (((*vulgaris*, *italicus*), *boscai*), *helveticus*) as reported by Steinfartz et al. (2007). We applied the semiparametric penalized likelihood rate smoothing method (Sanderson 2002) as implemented in r8s software with smoothing parameter $S = 40$, chosen by the crossvalidation procedure, and obtained divergence times of *Triturus cristatus* at 43, *Lissotriton boscai* at 35, *Lissotriton helveticus* at 29, and *Lissotriton italicus* at 22 Ma. In the second calibration of the molecular clock, we used the rate of 1.28% sequence divergence per My as estimated for the mtDNA fragment including the *Nd2* gene in salamanders by Weisrock et al. (2001). With this calibration, we obtained considerably younger divergence times: 22 Ma for *T. cristatus*, 18 for *L. boscai*, 16 for *L. helveticus*, and 13 Ma for *L. italicus*. These two sets of estimates differ

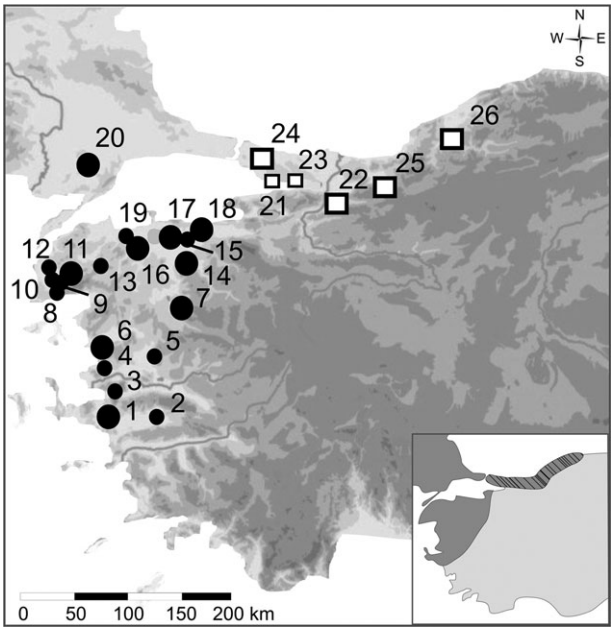

FIG. 1.—Sampling localities of *Lissotriton vulgaris vulgaris* (circles; 1–20) and *Lissotriton vulgaris kosswigi* (squares; 21–26). Larger symbols indicate populations scored for nuclear loci and mtDNA, smaller symbols—samples scored only for mtDNA. Insets show location of the study area and distributions of *vulgaris* (gray) and *kosswigi* (hatched). Note: the population 20 is Keşan and 24 is Alemdar.

by a factor of two, and it is reasonable to assume that they represent a range of plausible hypotheses on the divergence times. Thus, we use them to derive estimates of "slow" and "fast" mutation rates, respectively. To convert times of divergence into mutation rates per year, we divided mean net sequence divergence for individual loci between *L. vulgaris* and the outgroup by twice the divergence time. In order to obtain the mutation rate per generation, the per year rates were multiplied by a generation time. We used the generation time of 4 years, based on the synthesis of the literature data (Bell 1977; Marnell 1998; Cogălniceanu and Miaud 2003; Schmidtler and Franzen 2004). Sequences from the following outgroup species were used to calculate the mutation rate: *T. cristatus*: *Cxcr4*, *Gh*, *Ncx1*, *Fib*, and *Pdgfr*; *L. helveticus*: *Tva4*; and *L. italicus*: *Otx2*. Because no outgroup sequence for *Calr* was available, we assumed that it is equal to the average rate for the remaining four nuclear introns. We used the geometric mean of per-locus mutation rates to convert estimates in mutational units into real time and population sizes.

Results

MtDNA Variation, Divergence, and Structure

MtDNA sequences of 1,224 bp obtained for 205 *vulgaris* from 20 populations and 42 *kosswigi* from six populations represented 46 unique haplotypes (GenBank Accession nos: FJ588928–73). Phylogenetic analyses revealed two distinct clades with 100% BS/PP support, differing by 4.16% of net sequence divergence (Tamura–Nei distance). However, mtDNA sequences of individual newts did not always correspond with their subspecific assignment based on morphology. The larger "*vulgaris*" clade (40 haplotypes in total) included all haplotypes found in *vulgaris* populations and also the haplotypes from three western *kosswigi* populations (pops: 22, 23, and 24) (fig. 1). The "proper" *kosswigi* clade comprised six haplotypes from three populations in the central and eastern part of its range. The overall haplotype diversity (h) was 0.91 ± 0.01 for the *vulgaris* and 0.67 ± 0.08 for the *kosswigi* clade. Nucleotide diversities were $\pi = 0.80 \pm 0.03\%$ and $0.23 \pm 0.03\%$, respectively, and 1.72% and 0.48%, when considering only silent variation. Data for individual populations are given in supplementary table S2, Supplementary Material online. Tajima's and Fu and Li's D tests were not significant in either clade. Weak purifying selection was indicated by heterogeneity ($P = 0.04$) and M–K tests ($P = 0.012$). Geographic structuring of mtDNA was substantial; AMOVA performed on the entire data set divided into subspecies based on morphology indicated that most (52.3%) of variation was distributed among subspecies and among populations within subspecies (42.5%) (table 3). Also, K_{ST}* values within *vulgaris* and *kosswigi* were high: 0.85 and 0.91, respectively. Accordingly, many haplotypes were found in a single population only, although we also observed a few widespread haplotypes. There was a significant correlation between pairwise K_{ST}* and geographic distance in *vulgaris* ($r = 0.34$, $P < 0.001$, Mantel test). Statistical analysis was not possible for the *kosswigi* mtDNA clade due to only three populations sampled.

Nuclear Multilocus Variation

Eight nuclear loci were sequenced for 73 individuals from 13 populations (fig. 1). Sequenced fragments ranged from 423 to 747 bp (table 1; GenBank Accession nos for *Tva4*: FJ588119–47, for *Calr* FJ588052–79, for *Cxcr4* FJ588080–91, for *Fib* FJ588030–51, for *Gh* FJ588912–27, for *Ncx1* FJ588974–90, for *Otx2* FJ588991–9006, and for *Pdgfr* FJ588092–118). Overall nuclear nucleotide diversity (π) in *vulgaris* was 0.58% (silent $\pi = 0.69\%$) and in *kosswigi* 1.19% (silent $\pi = 1.31$, table 2). Silent π for protein-coding nuclear genes was 3–4.5 higher than overall π (table 1). Data for all populations are summarized in supplementary table S2, Supplementary Material online. Population nucleotide diversities were usually lower than overall estimates: for *kosswigi* they constituted 54–88% of the overall diversity, in *vulgaris* 64–117% (excluding the European Turkey Keşan population, where nucleotide diversity was extremely low, $\pi = 0.14\%$, only 25% of the overall value, supplementary table S2, Supplementary Material online). As could be expected, $\pi_{between}$ was higher than overall π in each subspecies, although both values were more similar in *vulgaris* (mean $\pi/\pi_{between}$ 0.97, range 0.93–1.00) than in *kosswigi* (mean 0.90, range: 0.79–0.99).

No haplotypes shared between subspecies were detected in any marker except *Tva4* (fig. 2). In addition to *Tva4*, which exhibited 27 shared polymorphisms, *Fib*, *Ncx1*, and *Otx2* also showed shared polymorphisms (one each). Both maximum likelihood and Bayesian trees revealed clearly distinct clades in each subspecies at four loci: *Cxcr4*, *Gh*, *Ncx1*, and *Otx2* (fig. 2).

Table 2
Summary of Nucleotide Variation

Subspecies	Locus	N[a]	L[b]	h[c]	π[d]	$\pi_{between}$[e]	θ[f]	π(sil)[g]	θ(sil)[h]	D[i]	D^*[j]	R_{min}[k]
vulgaris	*Tva4*	110	667	0.88	0.0223	0.0223	0.0132	0.0223	0.0132	2.138*	0.443	3
	Calr	108	552	0.75	0.0023	0.0024	0.0053	0.0023	0.0053	−1.551	−1.166	0
	Cxcr4	108	467/113	0.76	0.0027	0.0029	0.0025	0.0080	0.0084	−0.089	0.304	0
	Fib	108	423	0.75	0.0140	0.0143	0.0144	0.0140	0.0144	0.038	−0.664	4
	Gh	108	708	0.63	0.0015	0.0015	0.0042	0.0015	0.0042	−1.784*	−2.266	0
	Ncx1	108	598/127	0.69	0.0032	0.0034	0.0026	0.0142	0.0120	0.292	0.554	0
	Otx2	108	587	0.19	0.0005	0.0005	0.0027	0.0005	0.0027	−1.938	−2.010	1
	Pdgfr	108	747	0.66	0.0014	0.0014	0.0047	0.0014	0.0047	−1.959	−3.915**	1
	Average			**0.66**	**0.0058**	**0.0059**	**0.0060**	**0.0069**	**0.0072**	**−0.662**	**−1.235**	
	Nd2	226	1224	0.91	0.0080	0.0086	0.0094	0.0172	0.0173	−0.445	−0.316	
kosswigi	*Tva4*	38	667	0.84	0.0251	0.0271	0.0260	0.0251	0.0132	1.137	1.604**	3
	Calr	38	552	0.91	0.0125	0.0140	0.0089	0.0125	0.0260	1.439	0.979	1
	Cxcr4	38	467/113	0.64	0.0021	0.0025	0.0010	0.0089	0.0042	2.113*	0.777	0
	Fib	38	423	0.79	0.0148	0.0155	0.0092	0.0148	0.0010	2.075*	0.772	0
	Gh	38	708	0.67	0.0027	0.0032	0.0020	0.0027	0.0092	0.863	0.380	0
	Ncx1	38	598/126	0.49	0.0013	0.0013	0.0032	0.0048	0.0113	−1.722	−0.030	0
	Otx2	38	587	0.73	0.0025	0.0032	0.0016	0.0025	0.0032	1.288	1.032	0
	Pdgfr	38	747	0.86	0.0222	0.0236	0.0136	0.0222	0.0016	2.402*	1.456*	1
	Average			**0.77**	**0.0119**	**0.0130**	**0.0093**	**0.0131**	**0.0090**	**1.487**	**0.996**	
	Nd2	21	1,224	0.67	0.0023	0.0304	0.0021	0.0048	0.0038	0.385	−0.687	

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.
[a] Total number of sequences.
[b] Sequence length (bp), in the case of two nuclear coding genes, both total length and number of silent sites are given.
[c] Haplotype diversity.
[d] Nucleotide diversity per site.
[e] Mean number of pairwise differences computed taking into account only pairwise differences between sequences from different populations.
[f] Watterson's θ per site.
[g] Nucleotide diversity per silent site.
[h] Watterson's θ per silent site.
[i] Tajima's D (Tajima 1989).
[j] Fu and Li's D^* (Fu and Li 1993).
[k] Minimum number of recombination events.

Testing Assumptions of the IMa Model

Population Structure

Hierarchical AMOVA revealed that the majority of variation for each locus was distributed between subspecies, with the lowest percentage of variation among populations within groups, with the exception of *Tva4* in which the highest variation was found at the population level (table 3). The geographic structuring of nuclear variation within subspecies was notably lower than in the case of mtDNA and also differed between subspecies. Mean pairwise K_{ST}* averaged across loci was 0.24 in *kosswigi* and 0.14 in *vulgaris* (table 4). Interestingly, when the Keşan population was excluded from comparisons, K_{ST}* within *vulgaris* dropped even further to 0.09, indicating that geographic structuring of nuclear sequence variation in the Anatolian *vulgaris* is not strong. Hudson's *snn* analysis showed the same pattern (data not shown). A pattern of isolation by distance was apparent in *vulgaris* ($r = 0.52$, $P < 0.01$, Mantel test), the correlation between pairwise K_{ST}* and geographic distance in *kosswigi* was also very strong ($r = 0.93$), but the availability of only three populations precluded statistical testing.

Although we detected nonnegligible population structure, the robustness of the coalescent that holds in a variety of metapopulation models and concordant results obtained from the analyses of partial data sets (see below) justify the use of the IMa method.

Selection

The multilocus Tajima's D test was not significant in *vulgaris* ($D = -0.5$, $P > 0.1$, 10,000 coalescent simulations), whereas it was positive and highly significant in *kosswigi* ($D = 1.23$, $P < 0.0001$). In both subspecies, variance among loci was higher than expected from the standard neutral model (in *vulgaris* $P < 0.01$, in *kosswigi* $P < 0.05$). None of the M–K or heterogeneity tests for coding nuclear loci were significant. The multilocus HKA test using all sampled sequences from both subspecies was significant ($\chi^2 = 29.17$, $P < 0.01$). Two *kosswigi* loci (*Otx2* and *Gh*) showed low within-subspecies polymorphism and high between-subspecies divergence. The test was not statistically significant after exclusion of these loci. Altogether, we did not find unequivocal evidence for selection in our data set.

Recombination

Recombination in our data set was generally low, detected only at some loci and more often in *vulgaris* than in *kosswigi* (table 2), the difference being particularly pronounced in the composite ML recombination estimate (data not shown). The possible influence of recombination on the results of IMa and Lamarc analyses was eliminated by using maximally informative blocks of nonrecombining sequences.

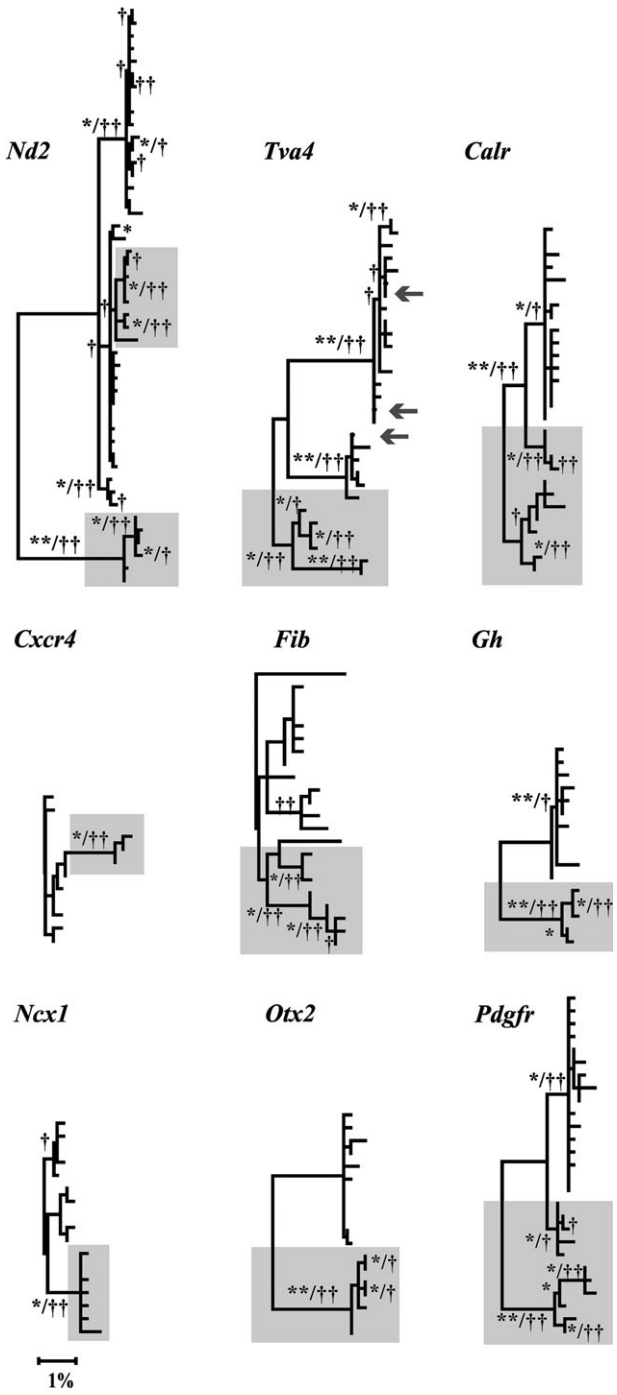

FIG. 2.—Phylogenetic trees of markers used in the study: BS values above 70% and posterior probabilities (PP) above 95% are indicated by asterisks and crosses, respectively, and 100% BS and PP with double symbols. *Kosswigi* alleles are shown on gray background or indicated with gray arrows. Trees (except *Calr*) were rooted with sequences from other newt species (data not shown, see text).

Demographic Stability

The Bayesian Lamarc growth parameter estimates were consistent between runs. The exponential growth rate estimates for *vulgaris* and *kosswigi* indicated demographic stability in the recent history of both subspecies (supplementary table S3, Supplementary Material online). No significant growth was detected in any of the partial analyses, except for *vulgaris* in the four-population analysis (treating the most distinct *kosswigi* Alemdar and *vulgaris* Keşan as separate populations); g for *vulgaris* in this analysis was estimated at 250.0 (95% CI 29.8–844.6; confidence interval, CI); however, the 99% CI included zero (−30.5 to 924.8). The analyses without mtDNA also showed no significant changes in population size. This further supports the inference of demographic stability, validating the assumption of the IMa model.

Isolation with Migration Model

The marginal posterior density curves and maximum likelihood estimates (MLE, locations of the peaks) for all parameters are given in figure 3 and table 5. Supporting online information contains estimates from additional data sets (supplementary table S4, Supplementary Material online). For the subspecies-wide data set, estimated theta values were similar for both subspecies corresponding to a long-term N_e of approximately 3.16×10^5 (2.30×10^5–4.27×10^5) and 2.98×10^5 (2.00×10^5–4.33×10^5) individuals using the slow mutation rate and 1.68×10^5 (1.23×10^5–2.27×10^5) and 1.58×10^5 (1.06×10^5–2.30×10^5) individuals when using the fast rate. For the ancestral population size and time of divergence, flat curves of the marginal posterior densities without distinct peaks indicate that these parameters could not reliably be estimated from our data. The estimate for migration from *kosswigi* to *vulgaris* was effectively zero, whereas migration rate was substantially higher in the opposite direction: 0.115 (0.045–0.215) corresponding to 0.178 (0.070–0.332) migrants per generation. Also, the likelihood ratio tests, applied to evaluate the fit of simpler divergence scenarios nested within the full IM model, strongly rejected the model assuming no gene flow from *vulgaris* to *kosswigi*, confirming the presence of historical asymmetric gene flow (supplementary table S5, Supplementary Material online).

The numbers and times of migration events were recorded for each locus. First the analysis was run with all likelihood functions set to one to obtain the prior distribution of migration times. The prior distribution had a peak near zero and did not correspond to the shape of marginal posterior density curves of migration times. Consistent with the findings from migration rates, all inferred migration events were from *vulgaris* to *kosswigi*: two events for five loci (*Nd2*, *Tva4*, *Calr*, *Fib*, and *Pdgfr*) and one migration event at four remaining loci (*Cxcr4*, *Gh*, *Ncx1*, and *Otx2*) However, the signal for migration in loci with one migration event was very weak (fig. 4). The marginal posterior densities curves summed across loci are clearly bimodal (fig. 4). The time of the most recent migration was estimated as approximately 0.20 Ma (fast mutation rate) and 0.38 Ma (slow mutation rate). The "older" peak was much wider and corresponded to 1.1 to >2 Ma (fast mutation rate) and 2.1 to >4 Ma (slow mutation rate). Separate curves for each locus indicated recent migration for *Nd2* and *TvA4* (0.20 [0.38] and 0.5 [0.94] Ma for fast and slow mutation rates, respectively). The migration times for *Nd2* (second peak) was

Table 3
AMOVA—Percentages of Variation Accounted for by between Subspecies, among Populations within Subspecies, and within Population Levels

	Source of Variation (%)		
Locus	Among Subspecies	Among Populations	Within Populations
Tva4	30.9**	15.3***	53.8***
Calr	68.1**	10.1***	21.8***
Cxcr4	86.4**	6.5***	7.0***
Fib	45.7**	13.7***	40.6***
Gh	94.8**	1.6***	3.6***
Ncx1	78.1**	8.9***	13.0***
Otx2	96.7***	1.2***	2.1***
Pdgfr	68.5**	6.9***	24.6***
Nd2	52.3***	42.5***	5.1***

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

estimated to be 1.03 (1.94) Ma, for *TvA4* and *Calr* 2.10 (3.90) Ma, for *Fib* 3.53 (6.63) Ma, and for *Pdgfr* 4.46 (8.38) Ma.

Estimates of parameters obtained from other data sets were similar (summarized in supplementary table S5, Supplementary Material online). Pairwise comparisons among four groups distinguished on the basis of pairwise K_{ST}* indicate that N_e for *vulgaris* without Keşan was very close to the whole subspecies estimates. A different pattern was observed in *kosswigi* in which the estimates of population sizes were almost the same for Alemdar and the remaining *kosswigi* populations combined. Detailed results can be found in the Supporting Information (supplementary table S5, Supplementary Material online).

Discussion

Ancient Divergence and Prolonged Gene Flow?

Analyses of the full IM model failed to provide reliable estimates of the time of divergence. Therefore, two approaches were used to roughly estimate the time of divergence. First, loci with no evidence of "recent" gene flow (all except mtDNA and *Tva4*) were subjected to an IMa analysis with migrations switched off. Second, the difference between the mean sequence divergence between species and $\pi_{between}$ divided by twice the mutation rate was used as an unbiased estimate of the time of divergence (Wakeley 2000). Both methods gave comparable estimates: With the fast mutation rate, the IM estimate was approximately 4.5 Ma, and distance method gave approximately 5 Ma. With slow mutation rate, the IM estimate was approximately 8.6 and distance 9.6 Ma. Estimates based on the mtDNA clock are somewhat lower: 3.4 and 6.1 for fast and slow, respectively.

Neither of the approaches corrects for effects of introgression; thus, both possibly underestimate the time of divergence. Nevertheless, even youngest estimates are remarkably old, placing divergence of these newts deep in the Pliocene, and an even more ancient, Miocene differentiation cannot be ruled out. Perhaps, speculation on the role of environmental factors in initiating divergence is premature, but we note that the younger divergence times agree well with the Messinian salinity crisis approximately 5.9–5.3 Ma (Krijgsman et al. 1999), which profoundly affected the circumediterranean biota.

With an estimated time of divergence of the order of 4–$5N_e$ generations, some parts of the genome are expected to have not reached reciprocal monophyly (Edwards and Beerli 2000; Hudson and Turelli 2003), but paraphyly or polyphyly may be a signature of gene flow as well. Two markers, mtDNA and *Tva4*, are particularly suggestive in this respect. Each subspecies is characterized by a distinct, highly divergent mtDNA clade, the only exception being the westernmost *kosswigi* populations close to the Bosphorus exhibiting haplotypes belonging to the *vulgaris* clade. Their limited geographic distribution close to the margin of *kosswigi* range is a clear signature of introgression, as previously suggested on morphological grounds (Freytag 1957). However, the haplotypes are not identical to any of the *vulgaris* samples: Either mtDNA gene flow, although recent, is not ongoing or *vulgaris* populations that constituted the source of introgressed haplotypes have not been sampled. The Bosphorus has remained submerged for the last 5.3–7 ky (Kerey et al. 2004), and more recently, the vast Istanbul agglomeration may have formed a barrier to the dispersal of newts.

On the other hand, the subspecies share several identical haplotypes, present in the center of the *kosswigi* range, in the anonymous marker *Tva4*. The retention of full ancestral haplotypes since the time of initial divergence is extremely unlikely, particularly with the high *Tva4* mutation rate. The IM analysis suggests gene flow, however not very recent, also in the case of other markers. All available evidence converges on the asymmetric gene flow from *vulgaris* to *kosswigi*. Asymmetric introgression appears to be relatively common, as evidenced by studies using IM (Won and Hey 2005; Niemiller et al. 2008; Slotte et al. 2008). Analysis of nested models (Hey and Nielsen 2007) is unambiguous in rejecting the scenario of no gene

Table 4
Average Pairwise K_{ST}* Based on Nuclear Loci Only

vulgaris	1	6	7	11	14	16	17	18	*kosswigi*	20	21
6	0.05								21	0.33	
7	0.06	0.08							25	0.26	0.12
11	0.11	0.06	0.08								
14	0.15	0.08	0.10	0.11							
16	0.10	0.06	0.11	0.03	0.07						
17	0.10	0.10	0.09	0.08	0.02	0.05					
18	0.18	0.19	0.12	0.15	0.04	0.11	0.04				
20	0.36	0.40	0.36	0.27	0.32	0.31	0.26	0.30			

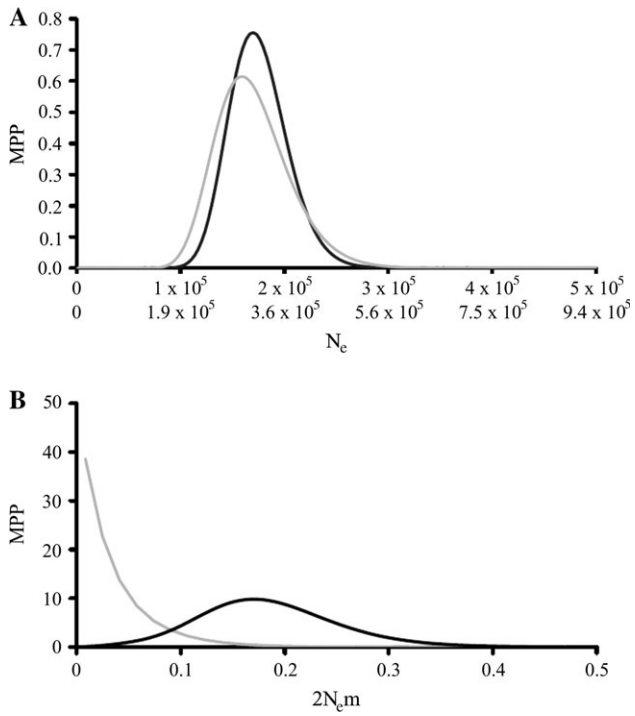

FIG. 3.—Marginal posterior probability (MPP) distributions of the IM model parameters for one of the data sets created by random sampling of two sequences from each population per locus: (*A*) effective population sizes (N_e) estimates (for fast and slow mutation rate) for *vulgaris* (black curve) and *kosswigi* (gray curve), (*B*) Population migration rates ($2N_em$) from *kosswigi* to *vulgaris* (gray curve) and from *vulgaris* to *kosswigi* (black curve).

flow from *vulgaris* to *kosswigi* providing statistical support to the inference of asymmetric gene flow.

The distribution of migration times is bimodal, supporting both ancient and recent gene flow. The controversial possibility that initial differentiation could have occurred in the face of gene flow cannot be rejected, whereas another burst of gene exchange has occurred rather recently and might be ongoing. The most recent migration times are found in mtDNA and anonymous marker *TvA4*. The migration times for other loci with a strong signal of gene flow support the hypothesis that the two subspecies may have been exchanging genes even since their initial divergence. However, additional markers sampled throughout the genome would be required to test if the bimodal distribution of migration times indeed reflects two distinct periods when gene flow occurred or is the consequence of the limited number of markers employed. For example, 454 sequencing of amplified fragment length polymorphism (AFLP) fragments could provide a wealth of homologous short genomic sequences randomly distributed over the genome, which could resolve the case quantitatively. Recent studies show that continuous gene flow between diverging populations may occur over many genomic region; this phenomenon appears to be much more common than previously thought and indicates that gene flow has to be considered as an important factor in the process of speciation, particularly driven by ecological factors (Hey et al. 2004; Won and Hey 2005; Niemiller et al. 2008; Strasburg and Rieseberg 2008).

It may seem surprising that populations so divergent still retain the ability to exchange genes. However, development of reproductive isolation can be a prolonged process. Some parts of the genome, located in proximity to genes initially involved in reproductive isolation, begin to diverge earlier than others (Wu and Ting 2004). These "genomic islands of speciation" may constitute hotspots for genome divergence while genes throughout the rest of the genome are exchanged freely between diverging populations (Turner et al. 2005; Yatabe et al. 2007). As the divergence of genomes progresses, more and more regions become differentiated, and neutral introgression becomes difficult as negatively selected genes form a considerable barrier to gene flow even for neutral, linked fragments. However, as long as the ability to hybridize is retained and hybrids are at least partially fertile, potential for gene flow in certain genomic regions is retained (Mallet 2005), even for tens of millions of generations (e.g., Kronforst 2008). Hybridization between "full" species of newts is frequent even in situations in which taxa differ profoundly in male epigamic traits crucial in mate recognition (Babik et al. 2003). Judging from the generally high level of divergence between subspecies, Turkish populations of the smooth newt have already accumulated many differences, and it is possible that the gene flow is opposed by selection in many parts of their genome.

Assumptions of IM Models and the Studied System

Selection might have influenced the evolution of several of the loci sampled, as suggested by some tests; however, these may rather reflect population structure, as suggested by consistently positive Tajima's *D* across loci in *kosswigi* or lack of introgression at individual loci manifesting itself in lower than expected polymorphism relative to divergence *Otx2* and *Gh* in *kosswigi*.

The deep divergence observed at the majority of the studied loci, together with clear morphological differences, indicates that each subspecies forms a distinct evolutionary unit. Compared with the extent of differentiation between *kosswigi* and *vulgaris*, genetic structuring within subspecies is of a relatively recent origin, and thus the use of separation of timescales approximation (Wakeley 1999, 2004) seems justified.

Although the details of population dynamics are unknown, it is reasonable to assume that each subspecies in Turkey approximates a metapopulation, whereas there is no deep and persistent subdivision in either subspecies This notion is supported by the relatively small spatial scale (maximum distance between *vulgaris* populations ca. 300 km), lack of obvious barriers to dispersal, generally sufficient habitat connectivity, and the lack of deep divergence among haplotypes across the range, even in the case of the more structured mtDNA. Local breeding populations inhabit numerous, usually transient ponds. Many amphibians, particularly salamanders, form metapopulations (Marsh and Trenham 2001).

Because separation of timescales approximation appears to hold and we are interested in the estimation of historical gene flow between subspecies, subspecies-wide growth rates and historical effective population sizes, we therefore used methods based on the standard coalescent

Table 5
Estimates of Parameters from IMa Analysis

Data set[a]	θ_1[b]	θ_2[c]	θ_A[d]	m_1[e]	m_2[f]	t[g]	N_{1slow}[h]	N_{2slow}[i]	N_{1fast}[j]	N_{2fast}[k]	$2N_1m_1$[l]	$2N_2m_2$[m]
Species wide sampling	3.28	3.09	10.66	0.000	0.115	44.84	3.16×10^5	2.98×10^5	1.68×10^5	1.58×10^5	0.000	0.178
HPD90Lo	2.39	2.07	0.16	0.000	0.045	14.83[n]	2.30×10^5	2.00×10^5	1.23×10^5	1.06×10^5	0.000	0.070
HPD90Hi	4.42	4.49	280.71	0.075	0.215	44.98[n]	4.27×10^5	4.33×10^5	2.27×10^5	2.30×10^5	0.123	0.332
Nuclear	2.23	2.39	7.18	0.000	0.115	37.51	2.71×10^5	2.92×10^5	1.44×10^5	1.55×10^5	0.000	0.178
HPD90Lo	1.49	1.55	0.14	0.000	0.015	12.44[n]	1.82×10^5	1.89×10^5	9.64×10^4	1.00×10^5	0.000	0.023
HPD90Hi	3.13	3.63	249.46	0.115	0.225	44.98[n]	3.81×10^5	4.43×10^5	2.02×10^5	2.35×10^5	0.189	0.348
Without introgressed markers	3.64	6.40	0.18			4.19	4.69×10^5	8.24×10^5	2.47×10^5	4.33×10^5		
HPD90Lo	1.52	2.23	0.18			1.58	1.96×10^5	2.87×10^5	1.03×10^5	1.51×10^5		
HPD90Hi	7.89	20.98	92.16			6.80	1.01×10^6	2.70×10^6	5.34×10^5	1.42×10^6		

NOTE.—all estimates include per gene mutation rate u, which is equal to the geometric mean of the mutation rates of all the loci.

HPD90Lo—the lower bound of the estimated 90% highest posterior density (HPD) interval.

HPD90Hi—the upper bound of the estimated 90% highest density interval.

[a] "Specieswide sampling" data set consists of two randomly sampled sequences per locus per locality; "nuclear" data set excluded mtDNA locus and "without introgressed markers" excluded *Tva4* and mtDNA and was run assuming no migration between subspecies.

[b] Theta for v*ulgaris*.

[c] Theta for *kosswigi*.

[d] Theta for ancestral population.

[e] Migration rate from *kosswigi* to *vulgaris*.

[f] Migration rate from *vulgaris* to *kosswigi*.

[g] Time since ancestral population splitting in mutational units ($t \times u$).

[h] Effective population size for *vulgaris* slow mutation rate, $u = 2.59 \times 10^{-6}$/locus/gen for all markers; 2.05×10^{-6} —for nuclear markers; 1.94×10^{-6}— for "without introgressed markers" data set.

[i] Effective population size for *kosswigi* slow mutation rate.

[j] Effective population size for *vulgaris* fast mutation rate, $u = 4.87 \times 10^{-6}$/locus/gen for all markers; 3.87×10^{-6} —for nuclear markers; 3.69×10^{-6}– for without introgressed markers data set.

[k] Effective population size for *kosswigi* fast mutation rate.

[l] Population migration rate for *kosswigi*, $M = 2Nm = (4Nu \times m/u)/2$.

[m] Population migration rate for *vulgaris*.

[n] Indicates that HPD interval did not appear to be contiguous and the estimates are not reliable.

process (e.g., Lamarc, IMa). At the same time, we tried to assess the robustness of the methods by comparing the results obtained from subsets of the data, which could be to some degree affected by population structure. These comparisons suggest that our results are reliable. However, the effect of population structure may be large in some situations; for example, it may profoundly influence the patterns of site frequency distribution and linkage, thereby invalidating tests for selection based on the standard neutral model (Wakeley and Aliacar 2001; Haddrill et al. 2005; Wright and Gaut 2005).

The influence of population structure on polymorphism spectra, estimates of coalescent-based demographic parameters, and inferences of selection have received much interest in recent years (Wakeley 2004; Arunyawat et al. 2007; Moeller et al. 2007), and there is a clear and urgent need for methods extracting demographic information from sequence data, explicitly dealing with moderate population subdivision. There is also much demand for a critical evaluation of the robustness of inferences made using existing methods (e.g., IM) under moderate departures from panmixia commonly observed in natural populations.

Notwithstanding possible controversies over the interpretation of the effective population sizes derived from the IM model, estimates of the relative amount, time, and direction of migration between two forms should not be severely affected by moderate population structure (Peters et al. 2007; Niemiller et al. 2008). Therefore, we regard our inferences on the patterns of gene flow between Turkish *vulgaris* and *kosswigi* as robust.

Nucleotide Variation

Overall, silent nuclear nucleotide diversity was almost two times higher in *kosswigi* than in *vulgaris* (1.31% vs. 0.69%); the pattern was also apparent at the population level and may reflect asymmetric introgression of alleles from *vulgaris* to *kosswigi*. Nucleotide diversity in the smooth newt in Turkey is high, much higher than observed in humans (Reich et al. 2002) or other hominoids (Thalmann et al. 2007) and comparable with that seen in *Drosophila* (Aquadro et al. 2001) and other invertebrates (Lynch 2006), although lower than in many outcrossing plants (Wright and Gaut 2005; Arunyawat et al. 2007). Local populations harbor a substantial fraction of the overall diversity found in each subspecies, consistent with migration rates sufficient to distribute variation emerging via mutation over the subspecies range.

It would be useful to compare our results at the level of nucleotide variation with data from other amphibians. Unfortunately, there is very little information on multilocus nuclear sequence variation in this group. The only study to date found a mean $\pi = 0.28\%$ for eight nuclear genes from a large sample of *Ambystoma ordinarium* (Weisrock et al. 2006). Other studies, examining mainly tropical frogs, usually used only a single nuclear locus (Crawford 2003) and/or focused on phylogeographic patterns rather than on levels of nucleotide diversity (Carnaval and Bates 2007; Fouquet et al. 2007). It is worth noting however that these studies inferred very high local population sizes in tropical anurans. The mtDNA nucleotide diversity is within the

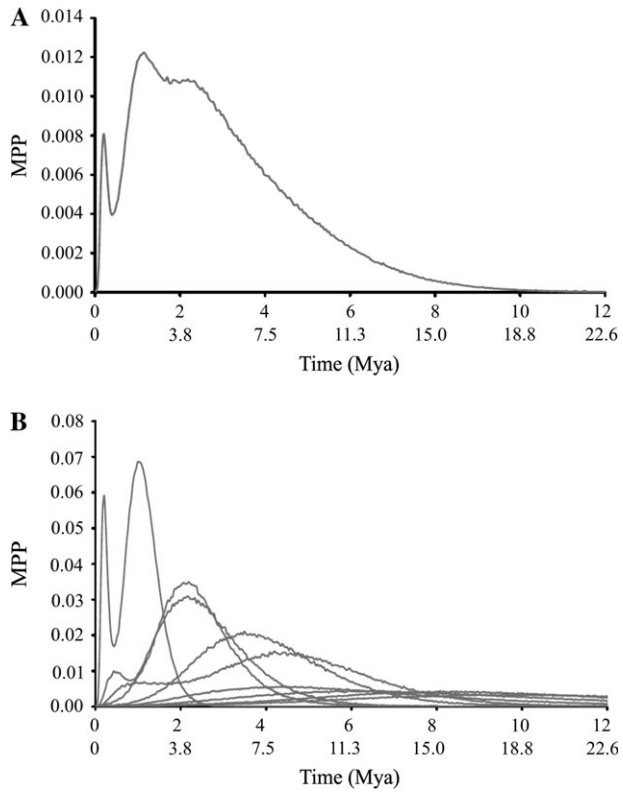

FIG. 4.—Temporal distribution of migration times. On the *x*-axis time in Ma (for fast and slow mutation rate). (*A*) Summary distribution for loci with two inferred migration events. The *y*-axis is the average of the estimates of posterior densities of migration times. (*B*) distributions for all individual loci.

ranges reported in other amphibian species (e.g., Crawford 2003; Rissler and Taylor 2003).

History of Newt Populations in the Turkish Refugium

High nucleotide variation, substantial mtDNA structuring and the apparent isolation by distance point to the long-term survival and large effective population sizes of smooth newts in western Turkey and support the conclusions of Babik et al. (2005) who located glacial refugia of the smooth newt in this area.

Numerous studies highlight the importance of Anatolia as a Pleistocene glacial refugium. High genetic diversity and basal clades have been repeatedly found there both in plants (Heuertz et al. 2006) and animals (Dubey et al. 2006), in agreement with climatic reconstructions (Hayes et al. 2005).

Populations in refugial areas often have long and complex histories (Gómez and Lunt 2006), and they frequently show high levels of genetic diversity reflecting large historical effective population sizes (Lessa et al. 2003). In contrast, populations from expansion areas often exhibit signatures of recent bottlenecks followed by strong demographic growth (Lessa et al. 2003; Hewitt 2004). The lack of significant demographic growth is well supported by the Lamarc analyses with concordant results across several data sets including different populations and markers. The negative (although not significantly different from zero) growth estimate from Lamarc together with highly positive Tajima's *D* values, both overall and in local populations, may even indicate a recent demographic contraction in *kosswigi* populations.

The apparent demographic stability and large long-term N_e are in line with the view that southern peninsulas in many cases constituted centers of endemism and biodiversity hotspots rather than sources of populations colonizing deglaciated areas (Bilton et al. 1998). In the case of the smooth newt, much of the vast postglacial range was colonized by lineages from the northern Balkans and Central Europe (Babik et al. 2005).

Conclusions

Our multilocus analysis of divergence population genetics of two subspecies of the smooth newt inhabiting Turkey has shown high nucleotide diversity implying large effective population sizes and suggested long-term demographic stability of both forms. Unidirectional gene flow from *vulgaris* to *kosswigi* has been occurring for a remarkable period of time, on the order of a million generations, possibly since their initial pre-Pleistocene divergence. The distribution of migration times does not allow to reject the hypothesis that the initial divergence may have occurred in the presence of gene flow, creating exciting prospect for future research. Our results point to the importance of divergence with gene flow as a general and apparently widespread mode of divergence.

Supplementary Material

Supplementary tables S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We are very grateful to Kurtluş Olgun who helped organized the collecting trip in Turkey and secured appropriate permits from TÜBITAK. Ben Wielstra, Gonçalo Themudo, Nazan Üzüm, and Özgür Güclü actively took part in newt collection. Pim Arntzen and Nazan Üzüm contributed additional samples. Maciej Pabijan provided valuable comments. The study would not be possible without the support of Walter Durka. The work was partially funded by the Alexander von Humboldt Foundation grant 3-Fokoop-POL/1022634 and by the Helmholtz Centre for Environmental Research—UFZ.

Literature Cited

Aquadro CF, DuMont VB, Reed FA. 2001. Genome-wide variation in the human and fruitfly: a comparison. Curr Opin Genet Dev. 11:627–634.

Arunyawat U, Stephan W, Stadler T. 2007. Using multilocus sequence data to assess population structure, natural selection,

and linkage disequilibrium in wild tomatoes. Mol Biol Evol. 24:2310–2322.
Baack EJ, Rieseberg LH. 2007. A genomic view of introgression and hybrid speciation. Curr Opin Genet Dev. 17:513–518.
Babik W, Branicki W, Crnobrnja-Isailovic J, Cogalniceanu D, Sas I, Olgun K, Poyarkov NA, Gracia-Paris M, Arntzen JW. 2005. Phylogeography of two European newt species—discordance between mtDNA and morphology. Mol Ecol. 14: 2475–2491.
Babik W, Szymura JM, Rafinski J. 2003. Nuclear markers, mitochondrial DNA and male secondary sexual traits variation in a newt hybrid zone (*Triturus vulgaris* × *T. montandoni*). Mol Ecol. 12:1913–1930.
Becquet C, Przeworski M. 2007. A new approach to estimate parameters of speciation models with application to apes. Genome Res. 17:1505–1519.
Bell G. 1977. The life of the smooth newt (*Triturus vulgaris*) after metamorphosis. Ecol Monogr. 47:279–299.
Bilton DT, Mirol PM, Mascheretti S, Fredga K, Zima J, Searle JB. 1998. Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. Proc Roy Soc Lond B. 265:1219–1226.
Cogălniceanu D, Miaud C. 2003. Population age structure and growth in four syntopic amphibian species inhabiting a large river floodplain. Can J Zool. 81:1096–1106.
Carnaval AC, Bates JM. 2007. Amphibian DNA shows marked genetic structure and tracks Pleistocene climate change in northeastern Brazil. Evolution. 61:2942–2957.
Charlesworth B, Charlesworth D, Barton NH. 2003. The effects of genetic and geographic structure on neutral variation. Ann Rev Ecol Syst. 34:99–125.
Clark AG. 1990. Inference of haplotypes from PCR-amplified samples of diploid populations. Mol Biol Evol. 7:111–122.
Coyne JA, Orr HA. 2004. Speciation. Sunderland (MA): Sinauer Associates.
Crawford AJ. 2003. Huge populations and old species of Costa Rican and Panamanian dirt frogs inferred from mitochondrial and nuclear gene sequences. Mol Ecol. 12:2525–2540.
Dolman G, Moritz C. 2006. A multilocus perspective on refugial isolation and divergence in rainforest skinks (*Carlia*). Evolution. 60:573–582.
Dubey S, Zaitsev M, Cosson JF, Abdukadier A, Vogel P. 2006. Pliocene and Pleistocene diversification and multiple refugia in a Eurasian shrew (*Crocidura suaveolens* group). Mol Phylogenet Evol. 38:635–647.
Edwards SV, Beerli P. 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution. 54:1839–1854.
Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics. Data analysis. Evol Bioinf Online. 1:47–50.
Fouquet A, Vences M, Salducci MD, Meyer A, Marty C, Blanc M, Gilles A. 2007. Revealing cryptic diversity using molecular phylogenetics and phylogeography in frogs of the *Scinax ruber* and *Rhinella margaritifera* species groups. Mol Phylogenet Evol. 43:567–582.
Freytag GE. 1957. Über *Triturus vulgaris* aus dem Gebiet von Istanbul. Zoologisches Anzeiger. 158:49–53.
Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. Genetics. 133:693–709.
Gómez A, Lunt DH. 2007. Refugia within refugia: patterns of phylogeographic concordance in Iberian Peninsula. In: Weiss S, Ferrand N, editors. Phylogeography of southern european refugia. The Netherlands: Springer. p. 155–188.
Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52:696–704.
Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P. 2005. Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. Genome Res. 15:790–799.
Hahn MW, Rausher MD, Cunningham CW. 2002. Distinguishing between selection and population expansion in an experimental lineage of bacteriophage T7. Genetics. 161:11–20.
Hayes A, Kucera M, Kallel N, Sbaffi L, Rohling EJ. 2005. Glacial Mediterranean sea surface temperatures based on planktonic foraminiferal assemblages. Quat Sci Rev. 24:999–1016.
Heuertz M, Carnevale S, Fineschi S, Sebastiani F, Hausman JF, Paule L, Vendramin GG. 2006. Chloroplast DNA phylogeography of European ashes, *Fraxinus* sp (Oleaceae): roles of hybridization and life history traits. Mol Ecol. 15: 2131–2140.
Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. Philos Trans R Soc Lond B Biol Sci. 359:183–195.
Hey J. 2005. On the number of New World founders: a population genetic portrait of the peopling of the Americas. PLos Biol. 3:965–975.
Hey J. 2006. Recent advances in assessing gene flow between diverging populations and species. Curr Opin Genet Dev. 16:592–596.
Hey J, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics. 167:747–760.
Hey J, Nielsen R. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc Natl Acad Sci USA. 104: 2785–2790.
Hey J, Won YJ, Sivasundar A, Nielsen R, Markert JA. 2004. Using nuclear haplotypes with microsatellites to study gene flow between recently separated Cichlid species. Mol Ecol. 13:909–919.
Hofreiter M, Serre D, Rohland N, Rabeder G, Nagel D, Conard N, Munzel S, Paabo S. 2004. Lack of phylogeography in European mammals before the last glaciation. Proc Natl Acad Sci USA. 101:12963–12968.
Hudson RR. 2000. A new statistic for detecting genetic differentiation. Genetics. 155:2011–2014.
Hudson RR. 2001. Two-locus sampling distributions and their application. Genetics. 159:1805–1817.
Hudson RR, Boos DD, Kaplan NL. 1992. A statistical test for detecting population subdivision. Mol Biol Evol. 9:138–151.
Hudson RR, Kreitman M, Aguadé M. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics. 116:153–159.
Hudson RR, Turelli M. 2003. Stochasticity overrules the "three-times rule": genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. Evolution. 57:182–190.
Kerey IE, Meric E, Tunoglu C, Kelling G, Brenner RL, Dogan AU. 2004. Black Sea–Marmara Sea Quaternary connections: new data from the Bosphorus, Istanbul, Turkey. Palaeogeogr Palaeocl. 204:277–295.
Knowles LL, Carstens BC. 2007. Estimating a geographically explicit model of population divergence. Evolution. 61:477–493.
Krijgsman W, Hilgen FJ, Raffi I, Sierro FJ, Wilson DS. 1999. Chronology, causes and progression of the Messinian salinity crisis. Nature. 400:652–655.
Kronforst MR. 2008. Gene flow persists millions of years after speciation in *Heliconius* butterflies. BMC Evol Biol. 8:98.
Kuhner MK, Smith LP. 2007. Comparing likelihood and Bayesian coalescent estimation of population parameters. Genetics. 175:155–165.

Lessa EP, Cook JA, Patton JL. 2003. Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. Proc Natl Acad Sci USA. 100:10331–10334.

Lynch M. 2006. The origins of eukaryotic gene structure. Mol Biol Evol. 23:450–468.

Machado CA, Kliman RM, Markert JA, Hey J. 2002. Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. Mol Biol Evol. 19:472–488.

Mallet J. 2005. Hybridization as an invasion of the genome. Trends Ecol Evol. 20:229–237.

Mallet J. 2007. Hybrid speciation. Nature. 446:279–283.

Marnell F. 1998. A skeletochronological investigation of the population biology of the smooth newts *Triturus vulgaris* L. at a pond in Dublin, Ireland. Biol Environ Proc Royal Irish Acad. 1:31–36.

Marsh DM, Trenham PC. 2001. Metapopulation dynamics and amphibian conservation. Conserv Biol. 15:40–49.

Martinez-Solano I, Teixeira J, Buckley D, Garcia-Paris M. 2006. Mitochondrial DNA phylogeography of *Lissotriton boscai* (Caudata, Salamandridae): evidence for old, multiple refugia in an Iberian endemic. Mol Ecol. 15:3375–3388.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature. 351:652–654.

McVean G, Awadalla P, Fearnhead P. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics. 160:1231–1241.

Minin V, Abdo Z, Joyce P, Sullivan J. 2003. Performance-based selection of likelihood models for phylogeny estimation. Syst Biol. 52:674–683.

Moeller DA, Tenaillon MI, Tiffin P. 2007. Population structure and its effects on patterns of nucleotide polymorphism in teosinte (*Zea mays* ssp *parviglumis*). Genetics. 176:1799–1809.

Nielsen R, Wakeley J. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics. 158: 885–896.

Niemiller ML, Fitzpatrick BM, Miller BT. 2008. Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: *Gyrinophilus*) inferred from gene genealogies. Mol Ecol. 17:2258–2275.

Notohara M. 1990. The coalescent and genealogical process in structured populations. J Math Biol. 29:59–75.

Peters JL, Zhuravlev Y, Fefelov I, Logie A, Omland KE. 2007. Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*Anas* spp.). Evolution. 61:1992–2006.

Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM, Richter DJ, Lander ES, Altshuler D. 2002. Human genome sequence variation and the influence of gene history, mutation and recombination. Nat Genet. 32:135–142.

Rissler LJ, Taylor DR. 2003. The phylogenetics of Desmognathine salamander populations across the southern Appalachians. Mol Phylogenet Evol. 27:197–211.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 19:2496–2497.

Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol. 19:101–109.

Schmidtler JF, Franzen M. 2004. *Triturus vulgaris* (Linnaeus, 1758) – Teichmolch. In: Böhme W, editor. Handbuch der Reptilien und Amphibien Europas. Wiebelsheim: AULA-Verlag GmbH. p. 847–967.

Slotte T, Huirun H, Lascoux M, Ceplitis A. 2008. Polyploid speciation did not confer instant reproductive isolation in *Capsella* (Brassicaceae). Mol Biol Evol. 25:1472–1481.

Steinfartz S, Vicario S, Arntzen JW, Caccone A. 2007. A Bayesian approach on molecules and behavior: reconsidering phylogenetic and evolutionary patterns of the Salamandridae with emphasis on *Triturus* newts. J Exp Zool Mol Dev Evol. 308B:139–162.

Strasburg JL, Rieseberg LH. 2008. Molecular demographic history of the annual sunflowers *Helianthus annus* and *H. petiolaris* – large effective population sizes and rates of long-term gene flow. Evolution. 62:1936–1950.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585–595.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol. 24:1596–1599.

Thalmann OH, Fischer AH, Lankester FH, Paabo SH, Vigilant LH. 2007. The complex evolutionary history of gorillas:Insights from genomic data. Mol Biol Evol. 24:146–158.

Turner TL, Hahn MW, Nuzhdin SV. 2005. Genomic islands of speciation in *Anopheles gambiae*. PLos Biol. 3:1572–1578.

Wakeley J. 1999. Nonequilibrium migration in human history. Genetics. 153:1863–1871.

Wakeley J. 2000. The effects of subdivision on the genetic divergence of populations and species. Evolution. 54:1092–1101.

Wakeley J. 2001. The coalescent in an island model of population subdivision with variation among demes. Theoret Pop Biol. 59:133–144.

Wakeley J. 2004. Metapopulation models for historical inference. Mol Ecol. 13:865–875.

Wakeley J, Aliacar N. 2001. Gene genealogies in a metapopulation. Genetics. 159:893–905.

Weisrock DW, Macey JR, Ugurtas IH, Larson A, Papenfuss TJ. 2001. Molecular phylogenetics and historical biogeography among salamandrids of the "true" salamander clade: rapid branching of numerous highly divergent lineages in *Mertensiella luschani* associated with the rise of Anatolia. Mol Phylogenet Evol. 18:434–448.

Weisrock DW, Shaffer HB, Storz BL, Storz SR, Voss SR. 2006. Multiple nuclear gene sequences identify Phylogenetic species boundaries in rapidly radiating clade of Mexican ambystomatid salamanders. Mol Ecol. 15:2489–2503.

Whitlock MC, Barton NH. 1997. The effective size of a subdivided population. Genetics. 146:427–441.

Wilkinson-Herbots HM. 1998. Genealogy and subpopulation differentiation under various models of population structure. J Math Biol. 37:535–585.

Woerner AE, Cox MP, Hammer MF. 2007. Recombination-filtered genomic datasets by information maximization. Bioinformatics. 23:1851–1853.

Won YJ, Hey J. 2005. Divergence population genetics of chimpanzees. Mol Biol Evol. 22:297–307.

Wright SI, Gaut BS. 2005. Molecular population genetics and the search for adaptive evolution in plants. Mol Biol Evol. 22: 506–519.

Wu CI, Ting CT. 2004. Genes and speciation. Nat Rev Genet. 5: 114–122.

Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH. 2007. Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. Genetics. 175:1883–1893.

Jody Hey, Associate Editor

Accepted January 5, 2009