

GENETIC ANALYSIS OF COMPLEX DEMOGRAPHIC SCENARIOS: SPATIALLY EXPANDING POPULATIONS OF THE CANE TOAD, *BUFO MARINUS*

ARNAUD ESTOUP,^{1,2} MARK BEAUMONT,³ FLORENT SENNEDOT,¹ CRAIG MORITZ,^{4,5} AND JEAN-MARIE CORNUET¹

¹Centre de Biologie et de Gestion des Populations, Institut National de Recherche Agronomique, Campus International de Baillarguet CS 30 016 34988, Montferrier-sur-Lez cedex, France

²E-mail: estoup@ensam.inra.fr

³School of Animal and Microbial Sciences, University of Reading, Whiteknights, Reading RG6 6AJ, United Kingdom

⁴Museum of Vertebrate Zoology, University of California, Berkeley, California, 94720-3110

⁵Department of Zoology and Entomology, University of Queensland, Brisbane, 4072, Australia

Abstract.—Inferring the spatial expansion dynamics of invading species from molecular data is notoriously difficult due to the complexity of the processes involved. For these demographic scenarios, genetic data obtained from highly variable markers may be profitably combined with specific sampling schemes and information from other sources using a Bayesian approach. The geographic range of the introduced toad *Bufo marinus* is still expanding in eastern and northern Australia, in each case from isolates established around 1960. A large amount of demographic and historical information is available on both expansion areas. In each area, samples were collected along a transect representing populations of different ages and genotyped at 10 microsatellite loci. Five demographic models of expansion, differing in the dispersal pattern for migrants and founders and in the number of founders, were considered. Because the demographic history is complex, we used an approximate Bayesian method, based on a rejection-regression algorithm, to formally test the relative likelihoods of the five models of expansion and to infer demographic parameters. A stepwise migration-foundation model with founder events was statistically better supported than other four models in both expansion areas. Posterior distributions supported different dynamics of expansion in the studied areas. Populations in the eastern expansion area have a lower stable effective population size and have been founded by a smaller number of individuals than those in the northern expansion area. Once demographically stabilized, populations exchange a substantial number of effective migrants per generation in both expansion areas, and such exchanges are larger in northern than in eastern Australia. The effective number of migrants appears to be considerably lower than that of founders in both expansion areas. We found our inferences to be relatively robust to various assumptions on marker, demographic, and historical features. The method presented here is the only robust, model-based method available so far, which allows inferring complex population dynamics over a short time scale. It also provides the basis for investigating the interplay between population dynamics, drift, and selection in invasive species.

Key words.—Approximate Bayesian computation, demographic inferences, invasive species, microsatellites, rejection-sampling algorithm, spatially expanding populations.

Received October 10, 2003. Accepted May 17, 2004.

Invasive species and populations pose major threats to biodiversity, ecosystem integrity, agriculture, fisheries, and public health. In recent decades, the frequency of biological invasions has increased to an unprecedented level, stimulating a multitude of research projects in ecology (reviewed in Keane and Crawley 2002; Shea and Chesson 2002). By contrast, evolutionary aspects of invasions, and especially population genetic aspects, have remained relatively unexplored (reviewed in Mooney and Cleland 2001; Lee 2002). However, it is imperative to move beyond treating populations of invasive species as genetic black boxes in mitigation and management strategies (Lee 2002). Numerous demographic and evolutionary parameters are crucial to the dynamics of spatially expanding populations and to their adaptation to a novel environment: migration rate and pattern, population and propagule size or density, life cycle, mating systems, intensity and variation of selective factors over space and time, and genetic architecture of traits under selection (reviewed in Lenormand 2002). Among those parameters, migration rate and pattern, effective population size, and number of pioneer individuals (i.e., founder events) may be estimated using molecular markers. Gene flow tends to oppose the effects of local selection and thus limit adaptation and the geographic limits of expansion (e.g., Kirkpatrick and Barton 1997). However, it can also replenish the genetic variance of local populations, which is a prerequisite for evolution by natural se-

lection (e.g., Tufto 2001). Local genetic variation also largely depends on effective population size and number of pioneer individuals (e.g., Wright 1931; Grant 1998). Moreover, theoretical and empirical studies have shown that dominance and epistatic variance within populations could be converted into additive genetic variance through genetic drift, especially through temporary bottlenecks (i.e., founder events) that may occur during a colonization process (e.g., Cheverud and Routman 1996; Reznick and Ghalambor 2001).

The dynamics of spatially expanding populations typically involve a complex interaction between migration and changing (effective) population size. Analyzing molecular data in this context remains a major challenge for population genetic theory. Significant progress has been possible thanks to the development of the coalescent theory (e.g., Donnelly and Tavaré 1995; Nordborg 2001), as well as the application of a new category of genetic markers, microsatellites (reviewed in Estoup and Angers 1998). Most importantly, it appears that genetic data can be profitably combined with information from other sources such as demography and historical records, using a Bayesian approach (e.g., Tavaré et al. 1997; Wilson and Balding 1998; Pritchard et al. 1999; Estoup et al. 2001; Estoup and Clegg 2003).

Recently, advances have been made toward estimating the probability of obtaining a given gene sample configuration to make fully likelihood-based statistical inference from mo-

lecular data, rather than drawing inferences based on summary statistics (Kuhner et al. 1998; Beaumont 1999; Beerli and Felsenstein 1999; Bahlo and Griffiths 2000; Stephens and Donnelly 2000; Nielsen and Wakeley 2001). However, those approaches remain extremely challenging for populations with a complex evolutionary history, so that most applications have focused on relatively simple demographic models (e.g., Wilson and Balding 1998; Beaumont 1999; but see Beerli and Felsenstein 2001, and Wilson et al. 2003). For more complex models, inferential methods that are not fully likelihood-based (i.e., methods based on summary statistics) still appear to be the best option available (reviewed in Fu and Li 1999; Beaumont et al. 2002). Such methods have been recently applied to the analysis of nonequilibrium populations (e.g., Pritchard et al. 1999; Estoup et al. 2001; Estoup and Clegg 2003), but more complex demographic scenarios such as recent population expansions over a large geographical range and that involve the joint effects of migration and population size fluctuation have not been considered. Approximate Bayesian inferential methods (e.g., Beaumont et al. 2002) hold promise for the analysis of complex demographic scenarios, provided that a sufficient amount of information from other sources such as demography and historical records are available.

The cane toad, *Bufo marinus*, is by far the most widely successfully introduced amphibian species, and it has one of the most extensive, documented histories of introduction of any vertebrates (reviewed in Easteal et al. 1981). The species is native to the American tropics and was deliberately introduced in 1935 as a biocontrol agent in Australia, where it spread across more than 1 million km². The geographic range of cane toads in Australia is still expanding in northern Australia and to a lesser extent in eastern Australia. A large amount of demographic and historical information, especially the rate of spread and the age of populations, is available on both expansion areas (Van Beurden and Grigg 1980; Sabath et al. 1981; Easteal et al. 1985; Freeland and Martin 1985; Easteal and Floyd 1986; Seabrook 1991). Interestingly, the rate of spread is about 16 times higher in the northern expansion area than in the eastern expansion area. Hence, the recent introduction of *B. marinus* on the Australian continent provides a unique opportunity to study the expansion process of a major invading species through the use of molecular markers. Moreover, the comparison of distinct expansion areas is likely to provide useful insights into the demographic factors underlying different spatial dynamics.

In the present paper, we combined a specific sampling scheme, microsatellite data, demographic and historical information, and an adaptation of the inferential methods of Beaumont et al. (2002), to characterize the expansion processes of *B. marinus* in northern and eastern Australia. This gave useful information on the general dynamics of expansions and allowed inferences to be made about key demographic parameters (migration rate and pattern, effective population size, and number of pioneer individuals).

MATERIALS AND METHODS

Historical Information Available on Expansion Areas

Bufo marinus has a generalized anuran life history with a terrestrial, ground-dwelling adult stage and aquatic embry-

onic and larval stages (Zug and Zug 1979). Native to the American tropics, the species was introduced in 1935 in Australia, where it spread across large areas (see Fig. 1). The presently active expansion zones are located in North Territory (the northern expansion area, NEA) and New South Wales (the eastern expansion area, EEA). These two expansion areas initially started from two independent introductions near Byron Bay and Normanton (Fig. 1). The populations at Byron Bay and Normanton were established as isolates in 1960 and 1961, respectively, presumably by long-distance transportation by humans. They have remained isolated from the main core of cane toad range of expansion for about 15 years and more than 23 years for Normanton and Byron Bay, respectively (Van Beurden and Grigg 1980; Sabath et al. 1981; Easteal et al. 1985; Easteal and Floyd 1986). Immediately after its establishment, the Byron Bay population spread southward at an average speed of 1.5 km per year. At the conclusion of sampling for this study (1999), the front of expansion was at Woodburn (New South Wales), about 60 km south from Byron Bay (pers. obs.). The Normanton population spread westward at an average speed of 24 km per year. In 1999, the front of expansion was at Elsey station (North Territory), about 910 km west from Normanton (pers. obs.).

Although the speed of expansion is higher for the NEA than the EEA, the two expansion areas have several features in common. There is no indication that the spread of *B. marinus* within either area occurred by long-distance founding of new populations: if such isolates did form, it was at a spatial scale smaller than the level of resolution used for tracking expansion (Van Beurden and Grigg 1980; Easteal and Floyd 1986). Although long dispersal events may not be excluded, the pattern of diffusion is hence approximately gradual over these two areas. One possible exception concerns the NEA for which there may have been a second independent foundation in 1979 at James Creek, about 160 km west of Normanton (Freeland and Martin 1985). For both the EEA and the NEA, the spread of cane toads has been constrained by several geographical and climatic factors such as ocean, mountains, and arid zones, so that the expansion pattern is roughly unidirectional and linear (Fig. 1). Ecological conditions, although substantially different between areas with a warmer and more humid climate in the NEA, are relatively homogeneous within each area, so that the degree of permeability of the habitats by cane toad seems to be relatively homogeneous within each area, at least along the roughly linear expansion routes.

In both expansion areas, the date of first sighting was used to infer the age of the sampled populations. These dates could be obtained for six of the 10 populations sampled along the EEA (Byron Bay, Lennox Head, Ballina, Pimlico, Broadwater, and Woodburn) in Van Beurden and Grigg (1980), Easteal and Floyd (1986), Seabrook (1991), and using our own personal observations (for the expansion front). For the NEA, the first sighting dates are known for all nine sampled populations (Freeland and Martin 1985; public documents from the Darwin Office of the Conservation Commission of the North Territory; pers. obs. by G. Grigg and A. Estoup for the most recent populations). For the EEA, first sighting dates indicate that the mean rate of spread southward declined

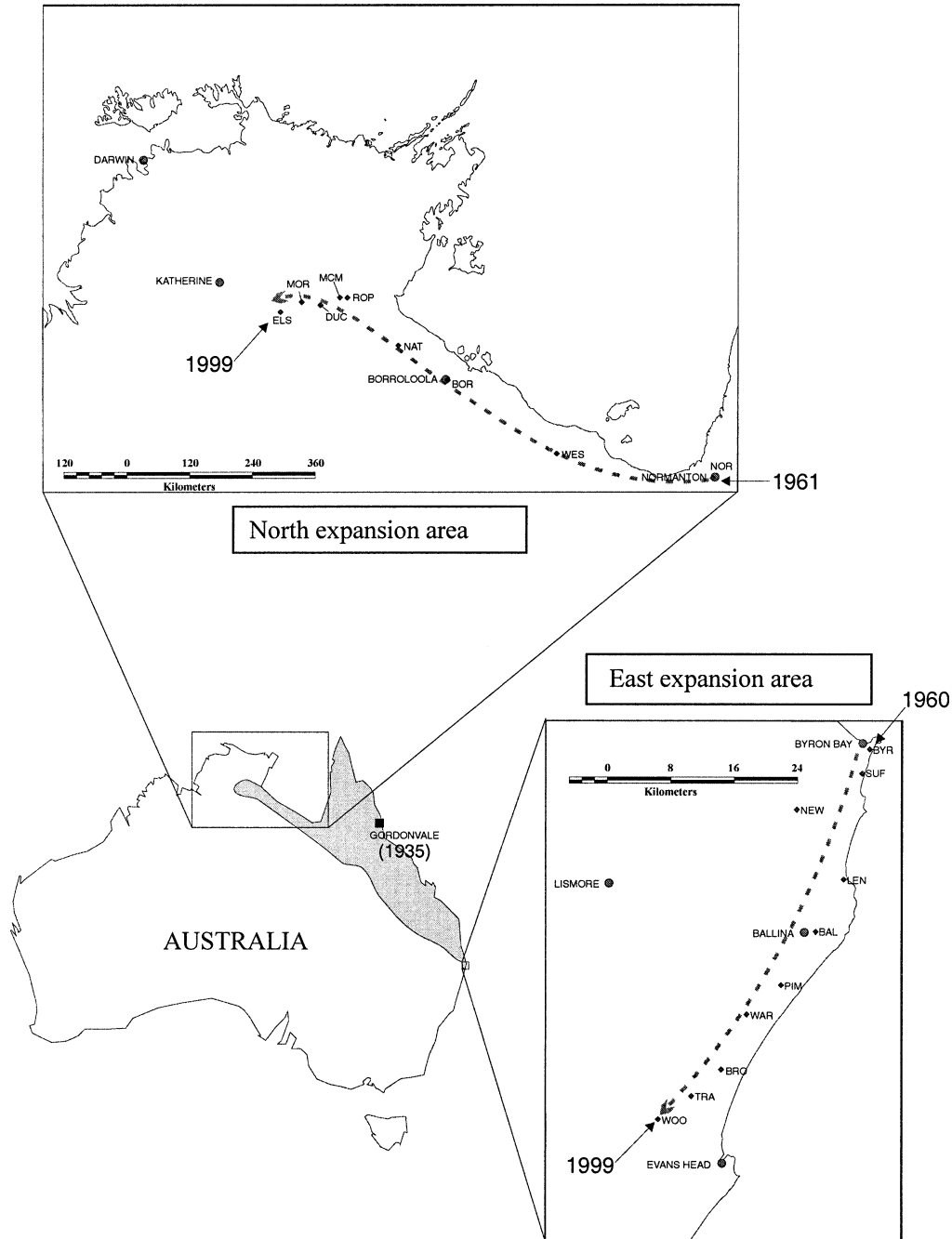


FIG. 1. Map showing the distribution range of cane toad in Australia and the location of the sampled sites in the two main active expansion areas. Gray area, distribution range in 1999; Gordonvale (1935), site and date of first introduction in Australia; gray dotted arrow, main direction of spreading; gray circles, main towns; dark gray diamonds, sampled sites; BYR, Byron Bay; SUF, Suffolk Park; NEW, Newrybar; LEN, Lennox Head; BAL, Ballina; PIM, Pimlico; WAR, Wardell; BRO, Broadwater; TRA, Tradewind; WOO, Woodburn; NOR, Normanton; WES, Westmoreland; BOR, Borroloola; NAT, Nathan River; ROP, Roper Bar; MCM, McMinn Station; DUC, Duck Pond; MOR, Moroak Station; ELS, Elsey Station.

more or less regularly from 2.75 km/year between Byron Bay and Lennox Head (Van Beurden and Grigg 1980) to 1.1 km/year between Broadwater and Woodburn (Seabrook 1991; pers. obs.). For the NEA, first sighting dates indicate that the rate of spread westward varies irregularly among years and zones (i.e., between 15 km/year for the Normanton-Westmoreland subarea and 43 km/year for the Nathan River–Rop-

er Bar subarea). The age of the four populations sampled along the EEA for which first sighting dates are not available was estimated using their exact geographic position and the rate of spread in the subarea defined by surrounding populations with known age and location. Similarly, the age, number and geographic position of the populations not sampled here but that were part of our demographic models of gradual

TABLE 1. Age and geographic distance of sampled sites and statistics summarizing the within- and between-population diversity in the two expansion areas. Age in number of generations (one generation per year), distance in kilometres from the initial introduction site; \bar{A} , mean number of alleles; \bar{H} , mean heterozygosity (gene diversity, Nei 1987); \bar{V} , mean allele size variance; sampling site order in pairwise F_{ST} matrices is the same as in the tables.

Eastern expansion area (EEA)										
	Byron Bay	Suffolk Park	Newrybar	Lennox Head	Ballina	Pimlico	Wardell	Broadwater	Tradewind	Woodburn
Age	40	38	35	32	27	22	17	11	6	1
Distance	0	5.0	13.5	22.0	29.5	37.0	43.0	50.5	56.5	62.5
\bar{A}	3.000	2.857	2.857	2.857	2.857	2.857	2.857	2.714	2.571	2.571
\bar{H}	0.510	0.562	0.576	0.500	0.548	0.462	0.505	0.433	0.548	0.452
\bar{V}	20.621	21.597	20.028	31.412	22.563	17.558	20.173	14.554	10.357	11.767
Pairwise F_{ST}										
0.0465										
0.0275	0.0066									
0.0311	0.0397	0.0187								
0.0396	0.0637	0.0267	0.0403							
0.0437	0.0196	0.0042	0.0295	0.0665						
0.0796	0.0892	0.0650	0.0663	0.1253	0.0465					
0.1258	0.0844	0.0781	0.0755	0.1315	0.0613	0.0334				
0.2242	0.1483	0.1458	0.1461	0.2223	0.1428	0.1057	0.0596			
0.1317	0.0666	0.0487	0.0869	0.1295	0.0305	0.0804	0.0665	0.0823		
Northern expansion area (NEA)										
	Normanton	Westmoreland	Borooloola	Nathan River	Roper Bar	McMinn Station	Duck Pond	Moroak Station	Elsley Station	
Age	39	19	13	8	5	4	3	2	1	
Distance	0	300	545	655	785	800	840	875	910	
\bar{A}	2.889	3.000	2.778	2.667	2.778	2.778	2.889	2.778	2.778	
\bar{H}	0.467	0.455	0.454	0.432	0.446	0.411	0.467	0.449	0.415	
\bar{V}	19.114	20.003	16.847	17.311	16.404	17.451	19.782	17.590	15.262	
Pairwise F_{ST}										
0.0315										
0.0309	0.0032									
0.0405	0.0250	0.0017								
0.0401	0.0250	0.0035	-0.0075							
0.0540	0.0245	0.0062	0.0022	-0.0021						
0.0294	0.0183	-0.0073	-0.0063	-0.0056	0.0024					
0.0461	0.0178	-0.0017	-0.0093	-0.0085	-0.0025	-0.0064				
0.0621	0.0550	0.0119	-0.0072	-0.0029	0.0012	-0.0054	-0.0042			

expansion (cf. *Demographic Models* below) were inferred from the rate of spread in the corresponding subareas.

Population Sampling

Ten and nine population samples (30 individuals per sample) were collected in February and June, 1999, along an approximately linear transect for the EEA and the NEA, respectively (Fig. 1). The samples consisted of adult toads (usually 15 males and 15 females) that were caught by hand around water holes, ponds, roads, and gardens. All population sites and samples were different from those analyzed in Estoup et al. (2001). Each transect went from the initial introduction site to the front of expansion (e.g., from Byron Bay to Woodburn for the EEA). Because of different sampling constraints, the distribution of geographical positions of sampling sites was different for the EEA and the NEA. The distances between two successive samples were very similar for the EEA while they were more heterogeneous for the NEA, with five of the nine sampling sites located in the last 125 km of the NEA (Table 1). Because, the geographic positions of sampled sites are specified in our demographic models, the different distributions of those positions in the

two expansion areas are fully accounted for in our model-based treatments. Furthermore, the sampled populations represent only a subset of the populations in the demographic models used for our treatments (cf. *Demographic Models* below).

Marker Analysis

In making inferences about population demography, it is desirable to use data from many unlinked loci (e.g., Donnelly and Tavaré 1995). In this study, 10 microsatellite loci were analyzed (BM217, BM118, BM121, BM229, BM235, BM239, BM279, BM322, BM224 and BM231; Tikel et al. 2000; Estoup et al. 2001). Seven and nine of those loci were polymorphic in the EEA and the NEA, respectively. Monomorphic loci were removed from all analyses (i.e., BM217, BM218, and BM231 in the EEA and BM217 in the NEA). Previous population studies did not show evidence for statistical association between loci other than that expected under particular demographic scenarios (Leblois et al. 2000; Tikel et al. 2000). DNA extractions were performed on individual piece of toe stored in 95% ethanol following the methods of Estoup et al. (1996). Microsatellite loci were

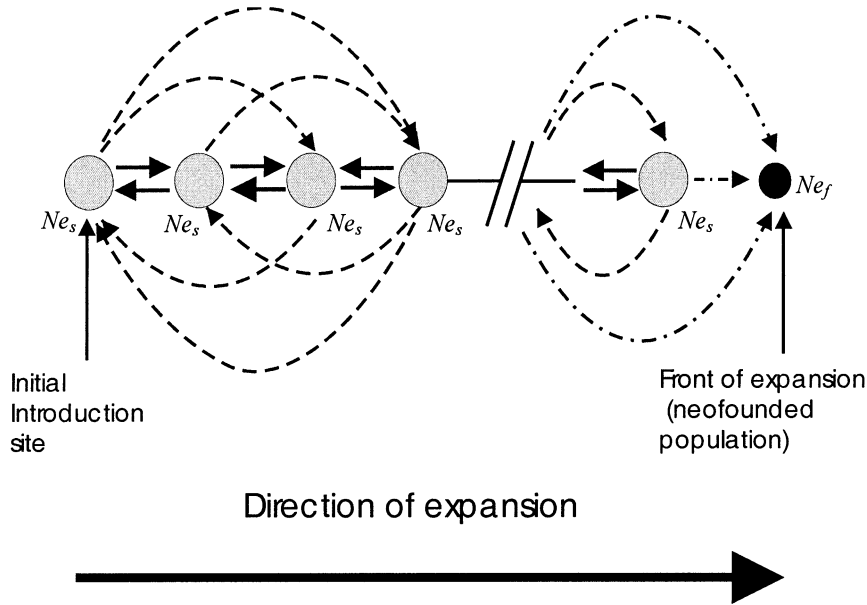


FIG. 2. Demographic models for spatial expansion of populations. N_{e_s} , stable effective population size; N_{e_f} , effective number of founders; full line arrows, migration events between adjacent populations (gray circles); dashed line arrows, migration events between nonadjacent populations; dot-dashed arrows, origin of founders in a neofounded population on the expansion front (black circle). Stepwise migration-foundation model includes only the full line arrows and the dot-dashed arrow connecting the neofounded population to the adjacent one. General isolation by distance migration-foundation model includes all arrows. Age difference between two adjacent populations is one generation. At each generation a new population is founded next to the previous one at a geographic distance that corresponds to the rate of spread in the subarea.

genotyped using fluorescent polymerase chain reaction (PCR) and an ABI sequencing machine (Applied Biosystems, Foster City, CA) as described in Tikel et al. (2000), except that new nonfluorescent-labeled primers were designed for two loci (BM235 and BM224) to allow a better separation of loci during the multilocus migration step. New primer sequences are 5'-ATGGCTTCCTCAACTGCAG-3' for BM235 and 5'-ACAATCTTTGATATTTGCTCG-3' for BM224, and give PCR fragments 47 bp and 11 bp shorter as compared to allele sizes in Tikel et al. (2000), respectively.

Demographic Models

Five different demographic models were considered to represent the expansion dynamics of *B. marinus* in both studied areas. In all models a linear series of sequentially founded populations is assumed, which is reasonable given strong geographical constraints in both areas. Also, as there is little evidence of long-distance founding (see above), at each generation a new population is assumed to be founded next to the previous one and at a geographic distance that corresponds to the rate of spread in the subarea. Each foundation involves a certain number of pioneer individuals (effective number of founders) originating from one or several previously founded populations depending on the dispersal pattern assumed in the models. *Bufo marinus* is a prolific species (7500–20,000 eggs/female; Alford et al. 1995) and a sudden population explosion was documented in most newly founded populations (Easteal 1981; pers. obs.). Thus, we assume that each newly founded population remains at low effective population sizes for a single generation only and reach larger stable effective population sizes in a single generation. We

also assume that the effective number of founders was the same for each newly founded population within each area so that this part of our model was specified by a single parameter (N_{e_f}). Similarly, the stable effective population size reached after foundation was assumed to be the same in all populations within each area and was thus specified by a single parameter too (N_{e_s}). Once at N_{e_s} , each population exchanges migrants at a rate m .

The difference between the five demographic models considered in this study relates to the dispersal pattern (i.e., the origin of dispersing individuals) for migrants and founders and to the number of founders. Those differences are summarized in Figure 2 and Table 2.

Model 1.—General model of isolation by distance for migration and foundation with founder events. During the expansion process, the number of populations exchanging migrants (s) increases over time. Each population receives a fraction of migrant genes m that have all other populations at stable population size as potential source. The relative contribution to migration of each source population depends on their geographical distance from the population receiving migrants. The probability that a migrant gene in population i originates from a given source population j is equal to:

$$p_{ij} = \frac{d_{ij}^{-k}}{\sum_{r=1, r \neq i}^s d_{ir}^{-k}}, \quad (1)$$

with $d_{\alpha\beta}$ the geographic distance between population α and β among the s populations at stable population size. The parameter k controls the intensity of the effect of geographic distance on migration. The higher the values of k , the stronger

TABLE 2. Variable and fixed demographic parameters in models. Model 1, general model of isolation by distance for migration and foundation with founder events and similar effect of geographic distance on migration and foundation; model 2, model 1 with potentially different effect of geographic distance on migration and foundation; model 3, island migration-foundation with founder events; model 4, stepwise migration-foundation with founder events; model 5, stepwise migration-foundation without founder events; prior, variable parameter with values drawn in a prior distribution given in Table 3.

Parameter	Model 1	Model 2	Model 3	Model 4	Model 5
Stable effective population size (Ne_s)	prior	prior	prior	prior	prior
Effective number of founders (Ne_f)	prior	prior	prior	prior	$Ne_s = Ne_f$
Migration rate (m)	prior	prior	prior	prior	prior
Effect of geographic distance on migration (k)	prior	prior	$k = 0$	$k = \infty$	$k = \infty$
Effect of geographic distance on founders (k_f)	$k_f = k$	prior	$k_f = k$	$k_f = k$	$k_f = k$

is the isolation by distance among populations. At each generation a population is founded next to the population founded last. The number of founders (Ne_f) is a parameter that can take low values, leading to the possibility of founder events (cf. *Prior Distributions of Demographic Parameters*). The origin of founding individuals is assessed using equation (1) and using the same k value as for migrants.

Model 2.—Differential migration and foundation. This model is identical to model 1 except that it is assumed that parameter k modulating the intensity of the effect of geographic distance may be different for migrants and founders. Hence, model 2 has one additional parameter compared to model 1, k_f , which is drawn independently from k using a prior distribution similar to that of k (cf. section *Prior Distributions of Demographic Parameters*). Such a model mimics a potentially different origin for migrants and founders. For instance, founders may preferentially originate from geographically closer populations than migrants because of the strong demographic pressure that may be associated with the sudden short-term demographic boom observed in recently founded populations (Easteal 1981; Estoup et al. 2001; pers. obs.).

Model 3.—Island migration-foundation with founder events. In this model the migrants and founders originate from any existing populations at stable population size independently of the geographic distance between populations. This model can be visualized as a particular case of model 1 with k , the parameter modulating the intensity of the effect of geographic distance on dispersal, equal to zero. Here also the number of founders is a parameter that can take low values, leading to the possibility of founder events.

Model 4.—Stepwise migration-foundation with founder events. In this model the migrants and founding individuals originate from the adjacent population(s) only. This model can be visualized as a particular case of model 1, with the value of the variable k being infinite or at least very large (e.g., $k = 10,000$). The number of founders is a parameter that can take low values, leading to the possibility of founder events.

Model 5.—Stepwise migration-foundation without founder events. This model is identical to model 4 except that the number of founders is equal to the stable effective population size (Ne_s), so that no founder event is assumed to occur during the spatial expansion process.

Estimation Procedure for Demographic Parameters

Although the Bayesian approach resolves, via integration, the theoretical problems caused by the presence of nuisance

parameters, there are severe numerical difficulties in performing the calculations of the probability distribution of the demographic (and mutational) parameters given the full genetic data (reviewed in Stephens 2003), especially when the demographic history considered is complex. To surmount these difficulties, we used an approximate Bayesian method based on summary statistics (e.g., F_{ST} and number of alleles) instead of the full data and a rejection-regression algorithm, to infer posterior distributions of variable parameters without explicit likelihood calculations (for details see Beaumont et al. 2002).

Briefly, the method of Beaumont et al. (2002) includes two steps: a rejection step and a regression adjustment and weighting step. The rejection step consists of accepting (i.e., recording) only sets of values of the variable parameters that give values of summary statistics computed from simulated datasets close to the values of summary statistics computed from the observed dataset, our target summary statistics. A Euclidian distance is computed between the entire set of normalized summary statistics and the normalized targets to measure the difference between the observed and simulated datasets. An iteration is accepted when the Euclidian distance is lower than a given threshold, whereas all summary statistics had to be in their own acceptance region to accept the iteration in former versions of the algorithm (e.g., Pritchard et al. 1999; Estoup et al. 2001; Estoup and Clegg 2003). The second step is a local linear regression adjustment that attempts to model the relationship between the parameter values and the summary statistics in the vicinity of the target summary statistics and thereby correct the accepted parameter values (for details see Beaumont et al. 2002). By using this adjustment, more points can be accepted, allowing a larger set of summary statistics to be handled and hence increasing the information extracted from the data. In our spatial expansion model, the allele frequency spectrum of the 10 or nine sampled populations is summarized by F_{ST} -values between all pairs of sampled populations (Weir and Cockerham 1984), as well as the mean number of different alleles, the mean gene diversity (Nei 1987), and the mean allele size variance computed across loci for each sampled population (see also Estoup et al. 2001; Estoup and Clegg 2003). Hence, the Euclidian distance is computed from 75 and 63 normalized summary statistics for the EEA and the NEA, respectively. Such a large number of statistics could not be handled by previous rejection-sampling algorithms without using prohibitively high acceptance thresholds.

A key feature of methods based on rejection-sampling al-

TABLE 3. Prior distributions for demographic parameters. Quantiles estimated over 100,000 values.

Parameter	Distribution	Range of supported values (2.5% and 97.5% quantiles)
Stable population size (N_{e_s})	uniform[50–3000]	124–2927
Number of founders (N_{e_f})	uniform[2–500]	14–487
Migration rate (m)	exp(-uniform[0–12])	8.3×10^{-6} to 0.74
Effect of geographic distance (k and k_f)	uniform[0–7]	0.18–6.83

gorithms is that they can handle complex models provided that simulation of data under the model is feasible. The coalescent model (e.g., Nordborg 2001) provides the theoretical basis required to simulate the genealogical processes underlying the patterns of shared ancestry among the genes in the samples. As the demographic history is particularly complex, we used a generation-by-generation algorithm where gene movements (i.e., migration between populations and foundation of new populations) and coalescent events are assessed at each generation going backward in time (for details see Leblois et al. 2003). The origin of the cane toad individuals initially introduced in Byron Bay and Normanton is unknown and the demographic history of the ancestral population(s) (i.e., before introduction) is certainly extremely complex and difficult to specify (Easteal 1981; Estoup et al. 2001). Thus, the coalescent trees were stopped at the time of the initial introduction, in year 1960 in the Byron Bay population for the EEA and 1961 in the Normanton population for the NEA. At this stage, the ancestral allelic states for the noncoalesced genes were simulated by random drawing in a multinomial distribution of κ allele frequencies obtained from a uniform Dirichlet distribution $D(\alpha_t)$ with $\alpha_t = (\alpha_{t,1}, \dots, \alpha_{t,i}, \dots, \alpha_{t,\kappa})$ and $\alpha_{t,i} = 1$ (e.g., Chikhi et al. 2001). The number of alleles κ and their state (i.e., allele size) are those observed over the entire dataset obtained in the EEA or the NEA. With this prior, all possible allele frequencies have equal probability at the time of the initial introduction. A model of pure drift without mutation is assumed. In practice, it means that mutations since the time of the initial introduction in Byron Bay or Normanton have negligible effect on our estimates. This assumption is reasonable because these introductions occurred only 40 generations before sampling time.

For the rejection step of the method of Beaumont et al. (2002), we set a tolerance threshold δ to be the quantile $p_\delta = 0.001$ of the empirical distribution function of the simulated Euclidian distance values. To avoid storing a large number of outputs, the normalizing factors and the critical quantile, δ , were first calculated from 100,000 iterations for each population expansion model and for each expansion area. Simulations were then run, keeping only those outputs with summary statistics within the tolerance threshold (i.e., with an Euclidian distance $< \delta$) until 10,000 sets of parameter values were accepted. These 10,000 values were obtained in around 48 h using a computer cluster of 20 standard (single processor) 1-Ghz platforms. The regression step was then processed on the 10,000 accepted values as described in Beaumont et al. (2002). Excessive summary statistics may decrease the efficiency of the method for a reasonable number of simulations (Beaumont et al. 2002). This is why demographic parameter values were adjusted on such a large number of accepted sets of values. To reduce heteroscedasticity

(i.e., inequality of variances among parameters) in the regression, all demographic parameters values were transformed on a log scale, with the exception of the migration rate (m) for which a logistic transformation was applied (i.e. $m^* = \ln[m/1 - m]$). The transformed values of the parameters were adjusted simultaneously using a multivariate regression on the 10,000 corresponding values of the 75 or 63 summary statistics. Adjusted values were then back-transformed taking the exponential for all parameters and $m = \exp(m^*)/[1 + \exp(m^*)]$ for the migration rate, to express posterior densities on a normal scale. Another personal program was used for this second step, which lasted less than 1 min on a single standard 1-Ghz platform.

For all the posterior density estimation from the adjusted sample of parameter values we used the *locfit* function (Loader 1996) implemented in the version 1.8.1 of the package R (Ihaka and Gentleman 1996; <http://cran.r-project.org>). Mean and quantile values were estimated using the *mean* and *quantile* functions in R.

Prior Distributions of Demographic Parameters

Information available from literature or personal observation on *B. marinus* populations were used to inform prior beliefs about demographic parameters (Table 3).

Regarding the stable effective population size (N_{e_s}), previous ecological, historical, and demographic data gave imprecise neighborhood size estimations of Australian populations between a few hundred to several thousand individuals (Easteal and Floyd 1986). Other estimations based on population models without migration and using enzyme and microsatellite loci supported effective population sizes of a few hundred individuals (Easteal 1985; Estoup et al. 2001). Because those values refer to different population models than those considered here, we adopted a diffuse prior distribution for N_{e_s} : a uniform distribution bounded between 50 and 3000. Personal observations suggest a considerably lower density of pioneer individuals right on the expansion front for both expansion areas. Therefore, for models including the possibility of founder events, we adopted a uniform distribution bounded between two and 500 for the number of effective founders (N_{e_f}). The founding ratio $F_R = N_{e_s}/N_{e_f}$, was also recorded, the prior on F_R being obtained by combining the priors on N_{e_s} and N_{e_f} .

Information available on dispersal parameters for the cane toad is imprecise and not necessarily adapted to the population models studied here. Using an individual-based analysis of microsatellite genotypes in a continuous population model, Leblois et al. (2000) found no evidence for strong isolation by distance in a cane toad population sampled over about 20 km around Byron Bay, suggesting that substantial

dispersal (and/or individual density) occurs at a relatively short geographic scale. Both mark-recapture and radiotracking studies have shown that cane toads rapidly move away from the location where they were captured (with individual distance per night ranging from 0 m to 1.3 km) and very seldom return (Schwartzkopf and Alford 2002). These studies, as well as the high capability for rapid colonization of large areas in Australia, suggest that migration rate and distance may be large. Since virtually any spectrum of migration rate and distance are possible, we adopted a diffuse prior distribution for the migration rate defined by $m = \exp(-\text{uniform}[0-12])$. This prior gives support to a large range of values from 8×10^{-6} to 0.74 (2.5% and 97.5% quantiles respectively, Table 3). A diffuse prior distribution was also chosen for the parameters k and/or k_f modulating the intensity of the effect of geographic distance on migration and/or foundation: a uniform distribution bounded between zero and seven. The effective number of migrants per generation $Ne_s m$ was also recorded, the prior on this parameter being obtained by combining the priors on Ne_s and m .

Cane toads reach sexual maturity at approximately 1 year and are then immediately reproductively active (Zug and Zug 1979). The death rate is high among adults, so that the assumption of a generation time of approximately 1 year seems reasonable (see also Eastale 1985; Eastale and Floyd 1986).

Robustness of Inferences

The potential effect of the prior distribution assumed for allelic frequencies at the time of the initial introduction was first tested. A Dirichlet distribution $D(\alpha_i)$ with $\alpha_i = (\alpha_{i,1}, \dots, \alpha_{i,i}, \dots, \alpha_{i,\kappa})$ and $\alpha_{i,i} = 1/\kappa$ (instead of $\alpha_{i,i} = 1$) was used to simulate allele frequencies of the κ observed allelic states. Such a distribution is also commonly used as a non-informative prior on allele frequencies in theoretical population genetics (e.g., Rannala and Mountain 1997); it tends to give more weight to low-frequency alleles than a distribution with $\alpha_{i,i} = 1$. Another simulation test was processed by assuming a Dirichlet distribution with parameter values corresponding to allele frequencies estimated by pooling data from all genotyped populations in each expansion area. In contrast to previous prior distributions, the mean for all allele frequencies is far from being equal in this case. To better investigate how much our results depend on peculiar choice of priors on m and k , we ran additional simulations using less diffuse priors for those parameters: a beta (1,20) distribution for m (mean = 0.0476; Q2.5% = 0.00126; Q50% = 0.0341; Q97.5% = 0.168) and an exponential distribution for k with mean = 3.5 (Q2.5% = 0.0891; Q50% = 2.427; Q97.5% = 12.956).

The generation time for *B. marinus* in subtropical Australian regions such as the EEA may be slightly longer than 1 year because of a colder and less humid climate than in tropical regions such as the NEA (e.g., Alford et al. 1995). Therefore, we ran an additional simulation assuming a generation time of 1.5 years instead of 1 year in the EEA. This corresponds to an expansion model with a lower number of founded populations in 39 years and a lower number of generations before reaching the initial introduction site in Byron Bay. Finally, we ran an additional simulation to assess the effect

of an initial foundation of the NEA at James Creek in 1979, a site located about 160 km West from Normanton, rather than at Normanton in 1961 (Freeland and Martin 1985). This corresponds to an expansion model starting from James Creek and excluding all populations older than 1979. In this case our model includes eight sampled populations instead of nine.

Comparative Tests of Demographic Models

We aimed to assess whether the observed microsatellite data discriminate among the five demographic models that have been considered as potentially representative of the expansion process of *B. marinus* (cf. *Demographic Models*). An attractive feature of rejection-sampling methods is that qualitative and quantitative model comparisons are feasible. We first qualitatively ranked the five models by estimating the distributions of Euclidian distances under each model for each expansion area (with $\delta = \text{infinite}$; 200,000 iterations for each model). The best model minimizes Euclidian distance values. We then quantitatively compared our models, opposing the apparently most likely model to the other four models. In rejection-sampling methods, the ratio of acceptance under two models approximates the Bayes factor, so that quantitative pairwise model comparison is feasible (Pritchard et al. 1999; Estoup et al. 2001). Half of the prior weight was placed on each model ($p_{m1} = p_{m2} = 0.5$). Then, we took δ so that $p_\delta = 0.001$ for the apparently most likely model. Using the same δ value, p_δ was estimated for each other four models. Hence, for a given δ value, we could estimate the relative posterior probabilities of the most likely model and each other four models.

RESULTS

Traditional Treatments for Analysis of within and between Population Genetic Variability

Some interesting trends can be observed when looking at mean number of alleles, heterozygosity, and allele size variance (computed over polymorphic loci) in association with the age of populations (Table 1). The populations sampled in the EEA showed a significant positive correlation between the number of alleles or allele size variance with the age of the population (Spearman rank order correlation test; $P = 0.0007$ and $P = 0.016$, respectively; $P = 0.088$ for heterozygosity). By contrast, such correlations were not significant for the populations sampled in the NEA ($P > 0.185$). These results suggest different dynamics of expansion in the two areas since a stronger reduction of genetic diversity during the expansion process in the EEA may reflect the occurrence of founder events of stronger intensity and/or a lower exchange of migrants among populations during this process.

All but three of the 45 pairwise differentiation tests (Raymond and Rousset 1995) were significant in the EEA ($p < 0.05$), whereas only 13 of the 36 tests were significant in the NEA. In agreement with this, the F_{ST} -value computed among all populations sampled in an expansion area was significantly larger in the EEA ($F_{ST} = 0.0799$) than in the NEA ($F_{ST} = 0.0132$; Wilcoxon signed-rank test, $P = 0.018$). These differences in population structure contrasts with the difference in geographic scales, the NEA being much larger than

the EEA (Fig. 1, Table 1). Again, these results suggest different dynamics of expansion, as described above. A significant correlation was found between pairwise F_{ST} -values (Table 1) and geographic distances between sampled populations in both expansion areas ($P = 0.004$ and $P = 0.003$ for the EEA and the NEA, respectively), noting that the NEA populations were sampled on a much larger spatial scale. This result suggests that isolation by distance should be taken into account when describing the expansion process in both areas.

The test for mutation-drift equilibrium of Cornuet and Luikart (1996) showed a significant excess of heterozygosity and allele size variance as compared to mutation-drift equilibrium expectations under a generalized stepwise mutation (GSM) model with $\sigma^2 = 0.36$ (Estoup et al. 2001) in all sampled populations and in both expansion areas (Wilcoxon signed-rank test, $P < 0.05$). These deviations from mutation-drift equilibrium are in agreement with demographic events such as bottlenecks associated with population foundation or the mixing of individuals from differentiated populations (Cornuet and Luikart 1996; Estoup et al. 2001). However, those events did not necessarily occur during the expansion process in the studied areas; rather they may have (also) occurred earlier during the complex introduction history of cane toad in the Caribbean and Pacific islands (including Australia in 1935). Similar deviations from mutation-drift equilibrium were indeed observed by Estoup et al. (2001) in the five initial introduction sites in Australia. Moreover, deviations from mutation-drift equilibrium were not necessarily stronger in the most recently founded populations than in the oldest populations of the studied expansion areas. Finally, excess of heterozygosity may also reflect deviation from the closed population model (i.e., model without migration among populations) assumed in the tests (Pope et al. 2000). Although Australian populations of *B. marinus* are spatially expanding, it is worth noting that no signal of population expansion (deficit of heterozygosity and/or allele size variance), that could have reflected a substantial increase of population size, has been detected in any population.

Comparative Tests of Demographic Models

Previous results give a rough indication of the demographic features of the expansion process of the two studied areas. It remains to be assessed formally whether the observed microsatellite data discriminate among the five demographic models described above. For both expansion areas, qualitative ranking of the five models, based on distributions of Euclidian distances, favors model 4 (i.e., the stepwise migration-foundation model with founder events) over other models (Fig. 3). We then quantitatively compared our models by opposing model 4 to models 1, 2, 3, and 5 for each expansion area. Posterior probabilities indicate that the statistical support is much larger for model 4 (posterior probabilities ≥ 0.96) in all four pairwise comparisons and for both expansion areas (Table 4). Similar results were obtained when using different prior distributions for m and k (posterior probabilities ≥ 0.995 for model 4; results not shown).

Inferences from Posterior Distributions

Given that model 4 is judged superior to other four models, posterior distributions of parameters were inferred for this

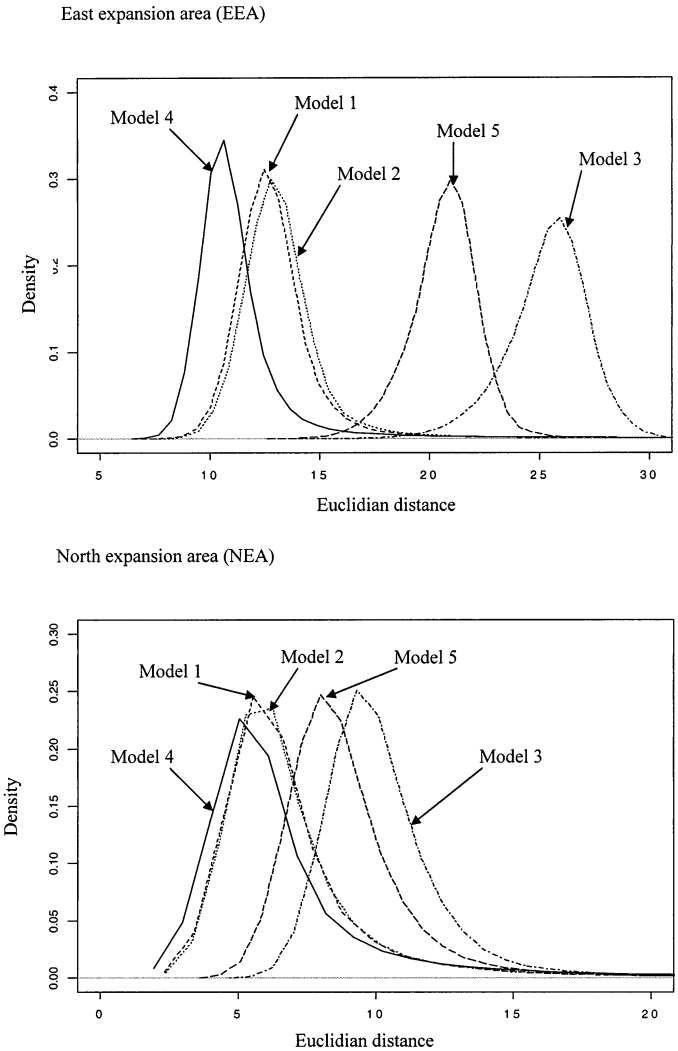


FIG. 3. Density curves of Euclidian distances under different expansion models. Model 1, general model of isolation by distance for migration and foundation with founder events and similar effect of geographic distance on migration and foundation; model 2, model 1 with potentially different effect of geographic distance on migration and foundation; model 3, island migration-foundation with founder events; model 4, stepwise migration-foundation with founder events; model 5, stepwise migration-foundation without founder events. Distributions are based on 200,000 values.

model only. The mean and standard deviation values of the mean, mode, 5%, 50%, and 95% quantiles of the posterior distributions were computed over 10 independent runs of 10,000 accepted values for each demographic parameter of the model. The posterior median (50% quantile) of each parameter was regarded as a point estimate (e.g., Wilson and Balding 1998). Table 5 shows that the variation of posterior distributions between independent runs are relatively small for all parameters of interest and for both the EEA and the NEA. This indicates that a single run of 10,000 accepted values provides a satisfactory estimate of posterior distributions.

The prior and posterior density curves for a single run of 10,000 accepted values are shown in Figure 4 before and after the regression adjustment and weighting step. Although

TABLE 4. Tests of demographic models. Model 1, general model of isolation by distance for migration and foundation with founder events and similar effect of geographic distance on migration and foundation; model 2, model 1 with potentially different effect of geographic distance on migration and foundation; model 3, island migration-foundation with founder events; model 4, stepwise migration-foundation with founder events; model 5, stepwise migration-foundation without founder events. Prior weight for each model was 0.5. EEA, eastern expansion area; NEA, northern expansion area.

Demographic models opposed	Posterior probabilities	
	EEA	NEA
Model 4 vs. model 1	0.99	0.97
Model 4 vs. model 2	0.01	0.03
Model 4 vs. model 3	0.98	0.96
Model 4 vs. model 5	0.02	0.04
Model 4 vs. model 3	$\cong 1$	$\cong 1$
Model 4 vs. model 5	$< 5 \times 10^{-3}$	$< 5 \times 10^{-3}$

processed on a large number of summary statistics, the regression adjustment and weighting step often have an important correction effect on posterior distributions. The correction increases support for higher (e.g., Ne_f in the EEA) or lower (e.g., Ne_s in the EEA) parameter values and its intensity varies among demographic parameters and expansion areas.

Figure 4 and Table 5 show that the posterior density curves of the stable effective population size (Ne_s) and the number of founders (Ne_f) differ noticeably from the priors. This means that the genetic data contain a substantial amount of information for Ne_s and Ne_f . For the EEA the posterior supports Ne_s values of a few hundreds individuals (point estimate of 693; Table 5). For the NEA, the posterior shows that populations in this area have larger Ne_s values than for the EEA (point estimate of 1698). The contrast between posterior distributions for the EEA and the NEA is even stronger for the number of effective founders. Posterior distributions support relatively low Ne_f values for the EEA (point estimate of 74) and much higher Ne_f values for the NEA (point estimate of 310). This indicates that founder events were more intense in the EEA than the NEA. Founding ratio were relatively low in both expansion areas, although almost two times higher in the EEA (point estimate of 9.2) than in the NEA (point estimate of 5.4).

The differences between the posterior and prior density curves for migration parameters, migration rate (m) and number of migrants per generation ($Ne_s m$), were less marked than for other demographic parameters (Fig. 4, Table 5). This means that our treatments extracted less information for m and $Ne_s m$ than for Ne_s and Ne_f . However, some interesting trends arise from posterior distributions. First, the posterior distributions selected toward much larger values for m and $Ne_s m$ in the NEA (modal values of 0.163 and 144 for m and $Ne_s m$, respectively) than in the EEA (modal values of 0.006 and 4.1 for m and $Ne_s m$, respectively). These values indicate

TABLE 5. Mean and standard deviation of mean, mode, and quantile values of the posteriors for the demographic parameters under the stepwise migration-foundation model with founder events. Mean and standard deviation values were computed over 10 independent runs of 10,000 accepted values. Standard deviation values are given in parentheses. Ne_s , stable effective population size; Ne_f , effective number of founders; F_R , founding ratio; m , migration rate; $Ne_s m$, effective number of migrants per generation. Values for priors were estimated from 100,000 values. EEA, eastern expansion area; NEA, northern expansion area; IR, irrelevant (cf. rectangular distribution).

			Mean	Mode	Quantiles		
					5%	50%	95%
Ne_s	Prior		1526	IR	197	1526	2853
	Posterior	EEA	744 (41)	537 (41)	205 (10)	693 (39)	1442 (80)
		NEA	1685 (66)	1944 (133)	526 (23)	1698 (65)	2838 (101)
Ne_f	Prior		251	IR	27	251	475
	Posterior	EEA	78 (4)	68 (4)	48 (3)	74 (4)	118 (6)
		NEA	311 (7)	313 (9)	182 (4)	310 (7)	448 (10)
F_R	Prior		16.9	2.1	0.8	6.1	56.8
	Posterior	EEA	10.7 (0.9)	6.5 (1.0)	2.4 (0.2)	9.2 (0.8)	23.8 (2.0)
		NEA	5.9 (0.2)	4.8 (0.5)	1.6 (0.1)	5.4 (0.2)	11.8 (0.4)
m	Prior		0.084	1.7×10^{-4}	1.1×10^{-5}	0.003	0.550
	Posterior	EEA	0.014 (0.006)	0.006 (0.004)	6.0×10^{-6} (3.7×10^{-6})	8.4×10^{-4} (4.4×10^{-4})	0.064 (0.032)
		NEA	0.117 (0.015)	0.163 (0.033)	1.4×10^{-4} (3.7×10^{-6})	0.009 (0.002)	0.664 (0.063)
$Ne_s m$	Prior		127	0.266	0.01	2.96	801
	Posterior	EEA	4.7 (2.2)	4.1 (1.8)	0.005 (0.002)	0.60 (0.27)	19.9 (9.6)
		NEA	188 (47)	144 (36)	0.023 (0.005)	13.5 (2.9)	883 (223)

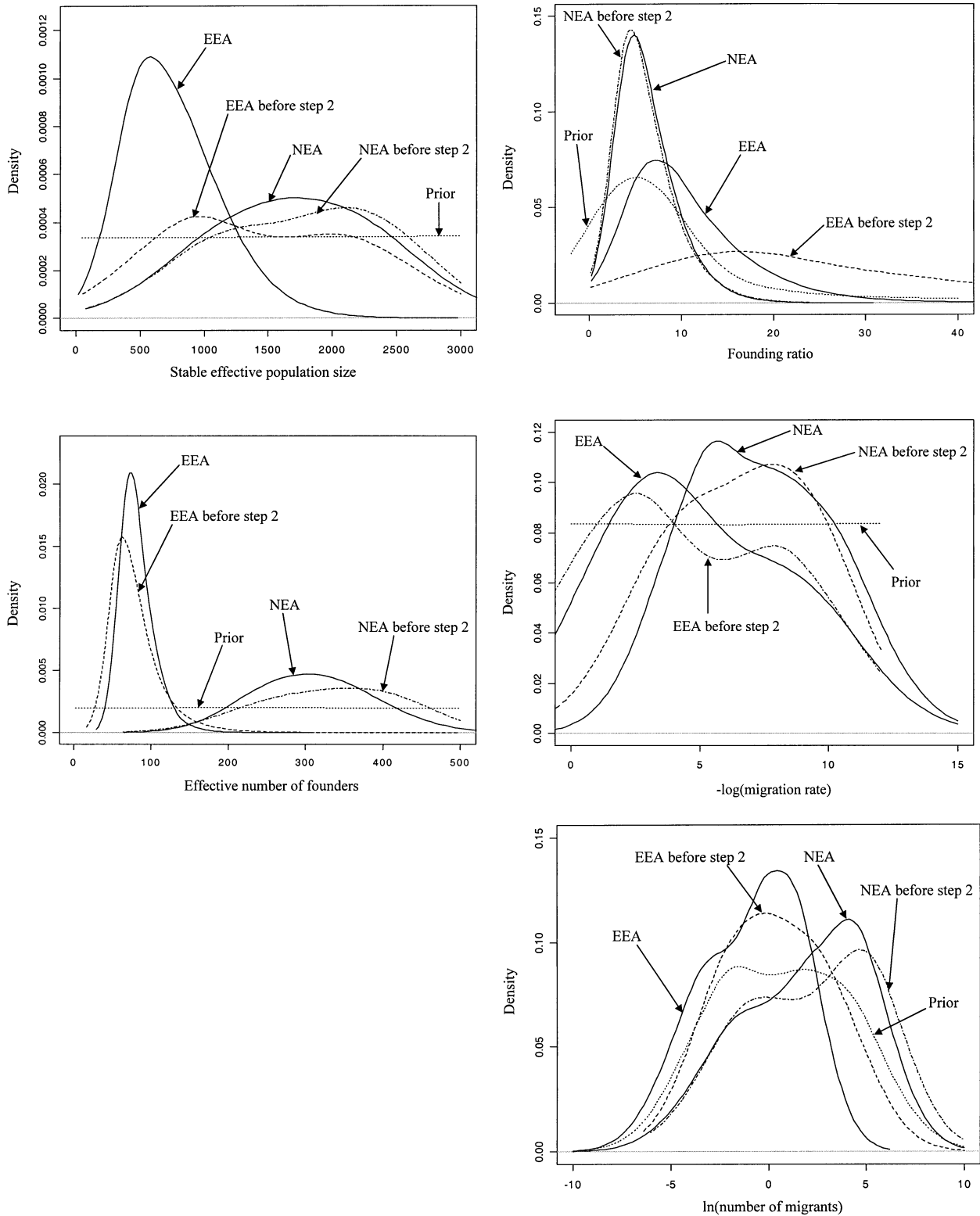


FIG. 4. Prior and posterior density curves for demographic parameters under a stepwise migration-foundation model with founder events. The short dashed lines correspond to the prior density curves. The long dashed and solid lines correspond to the posterior density curves before and after the regression adjustment and weighting step (step 2), respectively. EEA, eastern expansion area; NEA, northern expansion area. All prior and posterior densities are based on 100,000 and 10,000 values, respectively. The migration rate and the number of migrants are plotted on a logarithmic scale.

TABLE 6. Robustness of inferences. Mean, mode, and quantile values of demographic parameter posteriors were estimated under the stepwise migration-foundation model with founder events. All values were estimated from a single run of 10,000 accepted values. Ne_s , stable effective population size; Ne_f , effective number of founders; m , migration rate; $Ne_s m$, effective number of migrants per generation; EEA, eastern expansion area; NEA, northern expansion area. Superscript letters indicate (a) a Dirichlet distribution $D(\alpha_i)$ with $\alpha_{i,i} = 1/\kappa$ was used to simulate allele frequencies at the time of the initial introduction; (b) a Dirichlet distribution with parameter values corresponding to allele frequencies estimated by pooling data from all genotyped populations in each expansion area was used to simulate allele frequencies at the time of the initial introduction; (c) a Beta(1,20) distribution was assumed as prior for m ; mean, mode, 5%, 50% and 95% quantile values are then 0.048, 0.012, 2.6×10^{-3} , 0.034, and 0.138 for the prior on m and 73, 9.8, 1.96, 40.68, and 250 for the prior on $Ne_s m$; (d) Generation time is 1.5 years instead of 1 year in the EEA; (e) initial foundation of the NEA at James Creek in 1979 rather than at Normanton in 1961 (see text for details).

		Quantiles				
		Mean	Mode	5%	50%	95%
Ne_s	EEA ^a	782	696	217	728	1528
	EEA ^b	704	560	203	663	1316
	EEA ^c	435	241	115	390	918
	EEA ^d	391	247	102	349	810
	NEA ^a	1719	2054	523	1731	2893
	NEA ^b	1882	2109	539	1847	3370
	NEA ^c	1781	1958	599	1774	3011
	NEA ^e	1863	2490	555	1898	3040
	Ne_f	EEA ^a	75	74	47	72
EEA ^b		100	81	51	94	168
EEA ^c		82	73	51	79	125
EEA ^d		49	42	30	47	73
NEA ^a		322	350	193	326	439
NEA ^b		341	357	193	343	484
NEA ^c		325	317	193	323	469
NEA ^e		302	306	154	298	466
m		EEA ^a	0.009	0.005	3.0×10^{-6}	5.0×10^{-4}
	EEA ^b	0.046	0.023	1.0×10^{-5}	2.7×10^{-3}	0.272
	EEA ^c	0.024	0.026	1.0×10^{-3}	0.016	0.075
	EEA ^d	0.013	0.008	7.0×10^{-6}	7.8×10^{-4}	0.061
	NEA ^a	0.152	0.265	2.4×10^{-5}	0.017	0.772
	NEA ^b	0.156	0.296	3.4×10^{-5}	0.017	0.789
	NEA ^c	0.049	0.059	1.7×10^{-3}	0.035	0.148
	NEA ^e	0.150	0.258	1.9×10^{-5}	0.015	0.765
	$Ne_s m$	EEA ^a	2.9	3.2	0.003	0.38
EEA ^b		29	13	0.009	2.09	94.5
EEA ^c		7.9	6.5	0.732	5.55	22.7
EEA ^d		2.5	2.6	0.003	0.29	12.5
NEA ^a		310	243	0.040	24	1440
NEA ^b		418	330	0.058	25	202
NEA ^c		82	73	2.7	53	265
NEA ^e		330	308	0.034	26	1640

that adjacent populations are likely to exchange a substantial number of migrants per generation in both expansion areas and that such exchanges are larger in the NEA than in the EEA. Second, the number of effective migrants per generation exchanged between populations once demographically stabilized appears to be considerably lower than the number of effective individuals founding new populations in both expansion areas. This conclusion is less certain for the NEA because of the large 95% quantile obtained for $Ne_s m$ in this case.

Robustness of Inferences

Table 6 shows that, for the EEA, the use of a Dirichlet distribution $D(\alpha_i)$ with $\alpha_{i,i} = 1/\kappa$ gave values of drift (Ne_s and Ne_f) and migration (m and $Ne_s m$) parameters within the means \pm standard deviation values obtained using a Dirichlet distribution with $\alpha_{i,i} = 1$ (Table 5). For the NEA, the same result hold for Ne_s , while slightly and considerably larger values than corresponding means \pm standard deviation were obtained for Ne_f and migration parameters, respectively. A

Dirichlet distribution with parameter values corresponding to allele frequencies estimated by pooling data from all genotyped populations increased the support for higher values for all parameters in both expansion areas (except for Ne_s , which remained unchanged in the EEA). The assumption of a less diffuse prior on m (i.e., a beta[1,20] distribution) increased the support for lower values of Ne_s and Ne_f in the EEA and slightly increases the support for higher values of those parameters in the NEA. In agreement with the lower information content for dispersal parameters (cf. previous section), a larger effect was observed on m and $Ne_s m$. In any case, results remained in agreement with the two general conclusion that: (1) adjacent populations exchange a substantial number of effective migrants per generation in both expansion areas, and such exchanges are larger in northern than in eastern Australia; and (2) the effective number of migrants appears to be considerably lower than that of founders in both expansion areas.

The assumption of a generation time of 1.5 years instead of 1 year in the EEA increased the support for low values of

the drift parameters, while no significant effect was observed for the migration parameters (cf. mean and standard deviation values for a generation time of 1 year in Table 5). The assumption of an expansion model for the NEA starting from James Creek gave slightly lower (N_e) and considerably larger (other parameters) values than corresponding means \pm standard deviation.

Hence, although the effect of (some) prior assumptions on posterior distributions are far from being negligible, test simulations show that variation between estimated posterior distributions are generally not large, indicating that our inferences are relatively robust to assumptions on marker, demographic, and historical features, at least to those tested here. In particular, the differences initially observed in the dynamics of the two expansion areas remained or were strengthened.

DISCUSSION

A Methodological Advance in the Analysis of Complex Demographic Scenarios

Inferences from molecular data about complex demographic scenarios such as the spatial expansion dynamics of invading species is notoriously difficult. The present study illustrates the limitations of traditional approaches by showing that traditional population genetics treatments gave useful but only qualitative indications about the demographic features of the two studied expansion areas. Although standard correlative and F_{ST} analyses suggest different dynamics of expansion in the EEA and the NEA, the differences observed are difficult to interpret as they may reflect the occurrence of founder events of different intensity and/or varying rates of migration among populations during the expansion process. Similarly, the isolation-by-distance patterns observed in both areas are difficult to interpret as they potentially are reflected by interactions between migration rate and pattern, founder events, and sampling scheme. Finally, traditional methods of analyzing nonequilibrium situations, for instance that of Cornuet and Luikart (1996), assume unrealistic population models, at least in the present situation. They are also affected by longer term demographic processes, so that it is difficult to know to what extent results obtained with such methods illustrate short time-scale expansion processes. By contrast, our original simulation-based analysis allow choosing in a formal way among a nonexhaustive set of nonequilibrium demographic models that could represent the expansion process. Moreover, such treatments allow estimation of posterior distributions of parameters of interest under complex but realistic models. The inferences made here are also independent from the complex introduction history of *B. marinus* before reaching Australia (Easteal 1981; Estoup et al. 2001), and instead focus on the demographic events that occurred during the expansion process within Australia. Our simulation-based treatments, however, present the disadvantage of being much more computationally intensive than traditional treatments, and no generic program using such methods are yet available. Because of severe numerical difficulties, a priori more powerful fully likelihood methods (references listed in the introduction) do not yet have the potential to treat evolutionary scenarios as complex as those

studied in this paper, whatever the scale of the evolutionary processes involved. By contrast, methods based on rejection-sampling algorithms do not present such numerical difficulties and have the potential to handle virtually any complex models provided that simulation of data under the model is feasible (Beaumont et al. 2002).

Our model-based reconstruction of the spatial expansion history of *B. marinus* in Australia inevitably includes approximations of the actual process, leaving room for speculation about alternative scenarios and other details that may influence patterns of differentiation. Because both expansion areas are strongly constrained geographically, a linear expansion-foundation model appears reasonable. However, a two-dimensional model with a relatively small width that may vary over space (and time) would be more realistic for the cane toad in the studied areas. Such a model is more difficult to implement than a linear one and the gain it would bring is uncertain. In any case, it would be worth developing treatments based on two-dimensional models as they should prove to be useful for many other expanding populations and species.

The models representing the continuous expansion process of cane toad were made discrete as far as possible, given the demographic and historical information available (dates of first sight and rates of spread). This led us to consider a set of connected populations separated by geographic distances computed as the distance corresponding to a single generation of spread in a given subarea. However, this may be unrealistic because it implies that no population exists between those considered in our models. Considering a higher number of populations along the expansion process would be more realistic, but this number would be arbitrary with the present demographic information available. The ultimate level of discretization would be to consider a continuous population model such as the one of Rousset (2000). However, beside the fact that the sampling scheme under this model is different from the one used here, a continuous population model may not be appropriate at the large geographic scale of this study and estimation under such model of demographic parameters in a dynamic migration-foundation system would be difficult.

New Insights on Spatial Expansion Processes

In recently introduced and hence nonequilibrium populations, the time since introduction may not have been sufficient for isolation by distance to arise (Slatkin 1993). Tsutsui and Case (2001) invoked this as hypothesis to explain the absence of isolation by distance observed for invading populations of the Argentine ants introduced approximately 100 years ago, whereas populations of the same species showed a clear pattern of isolation by distance in the country of origin. Tests for isolation by distance in more recent (60 years) cane toad expanding populations showed that it is possible for such patterns to arise quickly. Although measures of isolation by distance in equilibrium cane toad populations from the original geographic range are not available, this result suggests that it is worth looking for other explanations than the nonequilibrium hypothesis when contrasted isolation by distance pattern between populations from the original and the introduced geographical range are observed (e.g., change in social

structure in the case of the Argentine ant or, more generally, differing frequency of human-mediated dispersal).

For species strongly dependent on the presence of suitable habitats (e.g., ponds for frogs) the spread in a novel area is expected to be more or less patchy and involves a greater potential for serial bottlenecks and associated reduction of genetic variability, especially when the species is not prolific. Here we show that an ecologically flexible and prolific species characterized by a continuous spread with substantial gene flow between populations may also show such pattern, especially when ecological conditions are not optimum (e.g., in the EEA). Stable effective population size (N_{e_s}) were larger in the NEA than in the EEA. This is at least partly explained by a larger carrying capacity in the NEA due to the more tropical climatic conditions in the NEA that better fit the ecological optimum of *B. marinus* (e.g., Alford et al. 1995). Translating an effective into an actual number of individuals as traditionally estimated in many demographic or ecological studies is difficult. Census:effective population size ratios can be very high ($\sim 100:1$) in anurans, especially for prolific species such as cane toad (Scribner et al. 1997; Zeisset and Beebee 2003). Thus, our estimates of N_{e_s} are not necessarily discordant with the observation of hundreds or thousands of toads in both the EEA and the NEA (Eastel 1981; pers. obs.). High census:effective population size ratios also mean that a limited number of effective founders as low as a few dozen in the EEA may actually represent a large number of individuals (e.g., several hundreds).

If the number of individuals founding a new population is proportional to the number of individuals in the source populations, it is not surprising that a larger number of founders (N_{e_f}) was also found in the NEA. However, founding ratio estimations show that N_{e_f} is, proportionally to N_{e_s} , larger in the NEA than in the EEA, so that other factors such as ecological conditions favoring dispersal in the NEA should be considered (e.g., Schwarzkopf and Alford 2002). For instance, extensive monsoonal flooding is frequent in northern Australia and may have an important role in cane toad dispersal. Because human population density is much lower in North than in East Australia, ecological factors also provide a better explanation for the higher rate of expansion in the NEA than does the possibility of transportation by humans.

The precision of our inferences on dispersal parameters is relatively poor. They are thus of little practical value in absolute terms, but give interesting information in the context of a comparative study between two expansion areas, by indicating a larger number of migration events between adjacent populations in the NEA than in the EEA. The number of effective migrants received each generation by a stable population (N_{e_m}) also appears to be lower than the number of effective founders (N_{e_f}) in both expansion areas. One possible explanation is that disparity between census and effective population sizes may be greater in a large stable population relative to a newly founded population in which there are fewer intraspecific competitors. In this case the same number of dispersing individuals would correspond to a relatively higher effective size in the new population. Another explanation relates to the observation that recent populations of cane toad are often characterized by a strong and transitory population boom of 2–10 years (Eastel 1981; Freeland 1986;

pers. obs.). If the migration rate is constant, such transitory high densities would directly translate into a larger number of migrants (i.e., a transitory high N_{e_m}) and hence a larger number of founders colonizing a novel area from the population located on the expansion front. Beside the possibility of transitory high N_{e_m} , large densities may also represent a pressure for migration that may increase the migration rate in recently founded populations and thus the number of founding individuals colonizing a novel area (e.g., Peters 1977; Baker 1978). Although our migration-foundation models disconnect migration between demes once demographically stable from founding events, they do not take into account the possibility of transitory population booms. Estoup et al. (2001) showed that very little information about such demographic events could be obtained from genetic data. Therefore, we have chosen to consider a single and integrative stable population size that includes the possibility of transitory large effective population sizes. Other inferential methods and sampling schemes (e.g., the sampling of populations over times) should be used to characterize transitory population boom events.

What could be the consequences of the inferred demographic features of the spatial expansions of *B. marinus* in the EEA and NEA on the current and future range expansion of the species in these areas? The high level of gene flow among adjacent populations is expected to oppose the effects of local selection and thus to limit adaptation and consequently the range of expansion in nonoptimal environments (e.g., Kirkpatrick and Barton 1997; reviewed in Lenormand 2002). Ecological conditions are more homogeneous in space and closer to optimum in the NEA than in the EEA, so that local selective pressure is expected to be higher in the EEA than in the NEA. The southern limit of the EEA is geographically close to the range limit of *B. marinus* in eastern Australia as predicted by computer simulation based on bioclimatic indexes (Sutherst et al. 1995). This may explain why the rate of spread is globally much higher in the NEA than in the EEA and why this rate declines going southward in the EEA (no such effect is visible for the NEA). The disruptive role of gene flow on local adaptation and hence on geographical expansion may be a limiting factor for the spread of cane toad in the EEA, where a north–south ecological gradient is likely to exist and act as a soft range boundary. Gene flow, however, especially at intermediate rates (Gomulkiewicz et al. 1999), can replenish the genetic variance of local populations, which is a prerequisite for evolution by natural selection. Populations sampled in the EEA showed a significant reduction of the genetic variability at microsatellite loci with the age of the population. If this reflects reduced local genetic variance at quantitative traits, the restorative effect of gene flow may be an important factor in the EEA. By contrast, the NEA would rather correspond to the case of a species colonizing a region where it is well adapted so that range expansion is not much affected by the amount of genetic variation available locally for adaptation (Kirkpatrick and Barton 1997). Introducing realistic demographic scenarios within local adaptation models was found to have interesting, but complicated, consequences (reviewed in Lenormand 2002). We hope that the demographical inferences on the expansion process obtained in the present

study will stimulate further work to include these inferences in predictive models dealing with the evolution of adaptive quantitative traits relevant to the continuing spread of *B. marinus* in Australia.

ACKNOWLEDGMENTS

We thank D. Tikel and R. Leblois for assistance in toad sampling, G. Grigg for providing useful information on the expansion dynamics of cane toads in North Territory, S. Piry for installation of computer clusters and drawing Figure 1, and I. Wilson for constructive comments on the manuscript and useful discussions on Dirichlet distributions. This work was financially supported by a grant to AE from the Ministère Français de l'Ecologie et du Développement Durable on biological invasions and to CM from the Australian Research Council.

LITERATURE CITED

- Alford, R. A., M. P. Cohen, M. R. Crossland, M. N. Hearnden, D. James, and L. Schwartzkopf. 1995. Population biology of *Bufo marinus* in North Australia. Pp 173–181 in M. Finlayson, ed. Wetland research in the wet-dry tropics of Australia. Office of the Supervising Scientist Report 101, Canberra, Australia.
- Baker, R. 1978. The evolutionary biology of animal migration. Hodder and Stoughton, London.
- Bahlo, M., and R. C. Griffiths. 2000. Inference from gene trees in a subdivided population. *Theor. Popul. Biol.* 57:79–95.
- Beaumont, M. A. 1999. Detecting population expansion and decline using microsatellites. *Genetics* 153:2013–2029.
- Beaumont, M. A., W. Zhang, and D. J. Balding. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162:2025–2035.
- Beerli, P., and J. Felsenstein. 1999. Maximum likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152:763–773.
- . 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using the coalescent approach. *PNAS* 98:4563–4568.
- Cheverud, J. M., and E. J. Routman. 1996. Epistasis as a source of increased additive genetic variance of population bottleneck. *Evolution* 47:1042–1051.
- Chikhi, L., M. Bruford, and M. A. Beaumont. 2001. Estimation of admixture proportions: a likelihood-based approach using Markov chain Monte Carlo. *Genetics* 158:1347–1362.
- Cornuet, J.-M., and G. Luikart. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144:2001–2014.
- Donnelly, P., and S. Tavaré. 1995. Coalescents and genealogical structure under neutrality. *Ann. Rev. in Genet.* 29:410–421.
- Eastale, S. 1981. The history of introductions of *Bufo marinus* (Amphibia: Anura): a natural experiment in evolution. *Biol. J. Linn. Soc.* 16:93–113.
- . 1985. The genetics of introduced populations of the giant toad *Bufo marinus*. III. Geographical patterns of variation. *Evolution* 39:1065–1075.
- Eastale, S., and R. B. Floyd. 1986. The ecological genetics of introduced populations of the giant toad, *Bufo marinus* (Amphibia: Anura): dispersal and neighborhood size. *Linn. Soc. of Lond* 27:17–45.
- Eastale, S., E. K. Van Beurden, R. Floyd, and M. Sabath. 1985. Continuing geographical spread of *Bufo marinus* in Australia: range expansion between 1974 and 1980. *J. Herpetol.* 19:185–188.
- Estoup, A., and B. Angers. 1998. Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. Pp. 55–86 in G. Carvalho, ed. *Advances in molecular ecology*. NATO Press, Amsterdam.
- Estoup, A., and S. M. Clegg. 2003. Bayesian inferences on the recent island colonization history by the bird *Zosterops lateralis lateralis*. *Mol. Ecol.* 12:657–674.
- Estoup, A., C. R. Largiadere, E. Perrot, and D. Chourrout. 1996. Rapid one-tube DNA extraction for reliable PCR detection of fish polymorphic markers and transgenes. *Mol. Mar. Biol. Biotech.* 5:295–298.
- Estoup, A., I. J. Wilson, C. Sullivan, J.-M. Cornuet, and C. Moritz. 2001. Inferring population history from microsatellite and enzyme data in serially introduced cane toads, *Bufo marinus*. *Genetics* 159:1671–1687.
- Fu, Y. X., and W. H. Li. 1999. Coalescing into the 21st century: an overview and prospects of coalescent theory. *Theor. Popul. Biol.* 56:1–10.
- Freeland, W. J. 1986. Population of the cane toad, *Bufo marinus*, in relation to time since colonization. *Aust. Wild. Res.* 13:321–330.
- Freeland, W. J., and K. C. Martin. 1985. The rate of range expansion by *Bufo marinus* in North Australia, 1980–84. *Aust. Wildl. Res.* 15:555–559.
- Grant, P. R. 1998. *Evolution on islands*. Oxford Univ. Press, Oxford, U.K.
- Gomulkiewicz, R., R. D. Holt, and M. Barfield. 1999. The effects of density dependence and immigration on local adaptation and niche evolution in a black-hole sink environment. *Theor. Popul. Biol.* 55:283–296.
- Hilborn, R., and M. Mangel. 1997. *The ecological detective: confronting models with data*. Princeton Univ. Press, Princeton, NJ.
- Ihaka, R., and R. Gentleman. 1996. R: a language for data analysis and graphics. *J. Comput. Graph. Stat.* 5:299–314.
- Keane, R. M., and J. Crawley. 2002. Exotic plant invasions and the enemy release hypothesis. *Trends Ecol. Evol.* 17:164–170.
- Kirkpatrick, M., and N. H. Barton. 1997. Evolution of species's range. *Am. Nat.* 150:1–23.
- Knowles, L. L., and W. P. Maddison. 2002. Statistical phylogeography. *Mol. Ecol.* 11:2623–2635.
- Kuhner, M. K., J. Yamato, and J. Felsenstein. 1998. Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* 149:429–434.
- Leblois, R., F. Rousset, D. Tikel, C. Moritz, and A. Estoup. 2000. Absence of evidence for isolation by distance in an expanding cane toad (*Bufo marinus*) population: an individual-based analysis of microsatellite genotypes. *Mol. Ecol.* 9:1905–1909.
- Leblois, R., A. Estoup, and F. Rousset. 2003. Influence of mutational and sampling factors on the estimation of demographic parameters in a continuous population under isolation by distance. *Mol. Biol. Evol.* 20:491–502.
- Lee, E. 2002. Evolutionary genetics of invasive species. *Trends Ecol. Evol.* 17:386–391.
- Legendre, P., and L. Legendre. 1998. *Numerical ecology*. 2nd English edition. Elsevier Science BV, Amsterdam.
- Lenormand, T. 2002. Gene flow and the limits to natural selection. *Trends Ecol. Evol.* 17:183–189.
- Loader, C. R. 1996. Local likelihood density estimation. *Ann. Stat.* 24:1602–1618.
- Mooney, H. A., and E. E. Cleland. 2001. The evolutionary impact of invasive species. *Proc. Natl. Acad. Sci. USA* 98:5446–5451.
- Nei, M. 1987. *Molecular evolutionary genetics*. Columbia Univ. Press, New York.
- Nielsen, R., and J. Wakeley. 2001. Distinguishing migration from migration: a Markov chain Monte Carlo approach. *Genetics* 158:885–896.
- Nordborg, M. 2001. Coalescent theory. Pp 179–212 in D. J. Balding, M. Bishop, and C. Cannings, eds. *Handbook of statistical genetics*. Wiley, Chichester, U.K.
- Peters, T. M. 1977. *The biology of invasions*. Technical Bulletin, University of Minnesota, Agricultural Experiment Station.
- Pope, L. C., A. Estoup, and C. Moritz. 2000. Phylogeography and population structure of an ecotonal marsupial, *Bettongia tropica*, determined using mtDNA and microsatellite. *Mol. Ecol.* 9:2041–2053.
- Pritchard, J. K., M. T. Seielstad, A. Prez-Lezaum, and M. W. Feldman. 1999. Population growth of human Y chromosomes: a

- study of Y chromosome microsatellites. *Mol. Biol. Evol.* 16: 1791–1798.
- Rannala, B., and J. L. Mountain. 1997. Detecting immigration by using multilocus genotypes. *Proc. Natl. Acad. Sci. USA* 94: 9197–9201.
- Raymond, M., and F. Rousset. 1995. GENEPOP (ver. 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* 86:248–249.
- Reznick, D. N., and C. K. Ghahambor. 2001. The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote rapid evolution. *Genetica* 112: 183–198.
- Rousset, F. 2000. Genetic differentiation between individuals. *J. Evol. Biol.* 13:58–62.
- Sabath, M. D., W. C. Boughton, and S. Easteal. 1981. Expansion of the range of the introduced toad *Bufo marinus* in Australia from 1935 to 1974. *Copeia* 1981:676–680.
- Schwarzkopf, L., and R. A. Alford. 2002. Nomadic movement in tropical toads. *Oikos* 96:492–506.
- Scribner, K. T., J. W. Arntzen, and T. Burke. 1997. Effective number of breeding adults in *Bufo bufo* estimated from age-specific variation at minisatellite loci. *Mol. Ecol.* 6:701–712.
- Seabrook, W. 1991. Range expansion of the introduced cane toad *Bufo marinus* in New South Wales. *Aust. Zool.* 27:58–62.
- Shea, K., and P. Chesson. 2002. Community ecology theory as a framework for biological invasions. *Trends Ecol. Evol.* 17: 170–176.
- Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* 47:264–279.
- Stephens, M. 2003. Inference under the coalescent. Pp. 213–238 in D. J. Balding, M. Bishop, C. Cannings, eds. *Handbook of statistical genetics*. Wiley, Chichester, U.K.
- Stephens, M., and P. Donnelly. 2000. Inference in molecular population genetics. *J. R. Stat. Soc. Lond.* 62:605–635.
- Sutherst, R. W., R. B. Floyd, and G. F. Maywald. 1995. The potential geographical distribution of the cane toad, *Bufo marinus* L. in Australia. *Conserv. Biol.* 9:294–299.
- Tavaré, S., D. J. Balding, R. C. Griffiths, and P. Donnelly. 1997. Inferring coalescence times from DNA sequence data. *Genetics* 145:505–518.
- Templeton, A. R. 1998. Nested clade analysis of phylogeographic data: testing hypotheses about gene flow and population history. *Mol. Ecol.* 7:381–397.
- Tikel, D., D. Peatkau, N. Cortinas, R. Leblois, C. Moritz, and A. Estoup. 2000. Microsatellite loci in the invasive toad species *Bufo marinus*. *Mol. Ecol.* 9:1227–1229.
- Tsutsui, N. D., and T. J. Case. 2001. Population genetics and colony structure of the Argentine ant (*Linepithema humile*) in its native and introduced ranges. *Evolution* 55:976–985.
- Tufto, J. 2001. Effects of releasing maladapted individuals: a demographic-evolutionary model. *Am. Nat.* 158:331–340.
- Van Beurden, E. K., and G. C. Grigg. 1980. An isolated and expanding population of the introduced toad *Bufo marinus* in New South Wales. *Aust. Wild. Res.* 7:305–310.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
- Wilson, I. J., and D. J. Balding. 1998. Genealogical inference from microsatellite data. *Genetics* 150:499–510.
- Wilson, I. J., M. E. Weale, and D. J. Balding. 2003. Inferences from DNA data: population histories, evolutionary processes, and forensic match probabilities. *J. R. Stat. Soc. Lond.* 166: 155–201.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16: 97–159.
- Zeisset, I., and J. C. Beebee. 2003. Population genetics of a successful invader: the marsh frog *Rana ridibunda* in Britain. *Mol. Ecol.* 12:639–646.
- Zug, G. R., and P. B. Zug. 1979. The marine toad, *Bufo marinus*: a natural history résumé of native populations. *Smiths. Contr. Zool.* 284:1–58.

Corresponding Editor: F. Bonhomme