MOLECULAR ECOLOGY

Molecular Ecology (2012) doi: 10.1111/mec.12065

Retracing the routes of introduction of invasive species: the case of the *Sirex noctilio* woodwasp

E. BOISSIN,* B. HURLEY,† M. J. WINGFIELD,* R. VASAITIS,‡ J. STENLID,‡ C. DAVIS,§ P. DE GROOT,§[1] R. AHUMADA,¶ A. CARNEGIE,** A. GOLDARAZENA,†† P. KLASMER,‡‡ B. WERMELINGER§§ and B. SLIPPERS*

**Department of Genetics, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria 0002, South Africa, †Department of Zoology and Entomology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria 0002, South Africa, ‡Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, PO Box 7026, SE-75007, Uppsala, Sweden, §Great Lakes Forestry Centre, Canadian Forest Service, 1219 Queen St. East, Sault Ste Marie, Ontario, Canada, P6A 2E5, ¶Bioforest S.A., Casilla 70-C, Concepción, Chile, **NSW Department of Primary Industries, Forest Science Centre, PO Box 100, Beecroft, NSW 2119, Australia, ††Departamento de Protección, Neiker, Instituto Vasco de Investigación y Desarrollo Agrario, Vegetal E-01080 Apdo 46, Vitoria-Gasteiz, Spain, ‡‡Laboratorio de Ecología de Insectos, Campo Forestal Gral. San Martin, INTA EEA, Bariloche, El Bolsón, Rio Negro, Argentina, §§Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf 8903, Switzerland*

Abstract

Understanding the evolutionary histories of invasive species is critical to adopt appropriate management strategies, but this process can be exceedingly complex to unravel. As illustrated in this study of the worldwide invasion of the woodwasp *Sirex noctilio*, population genetic analyses using coalescent-based scenario testing together with Bayesian clustering and historical records provide opportunities to address this problem. The pest spread from its native Eurasian range to the Southern Hemisphere in the 1900s and recently to Northern America, where it poses economic and potentially ecological threats to planted and native *Pinus* spp. To investigate the origins and pathways of invasion, samples from five continents were analysed using microsatellite and sequence data. The results of clustering analysis and scenario testing suggest that the invasion history is much more complex than previously believed, with most of the populations being admixtures resulting from independent introductions from Europe and subsequent spread among the invaded areas. Clustering analyses revealed two major source gene pools, one of which the scenario testing suggests is an as yet unsampled source. Results also shed light on the microevolutionary processes occurring during introductions, and showed that only few specimens gave rise to some of the populations. Analyses of microsatellites using clustering and scenario testing considered against historical data drastically altered our understanding of the invasion history of *S. noctilio* and will have important implications for the strategies employed to fight its spread. This study illustrates the value of combining clustering and ABC methods in a comprehensive framework to dissect the complex patterns of spread of global invaders.

Keywords: approximate Bayesian computation, coalescence, effective number of founders, founder events, scenario testing, unsampled population

Received 14 April 2012; revision received 4 August 2012; accepted 17 August 2012

Introduction

Globalization has resulted in substantially increased invasion rates during the course of the past 50 years (Hulme 2009; Richardson 2010). This is of great concern

Correspondence: Emilie Boissin, Fax: +27 12 362 5327; E-mail: eboissin@gmail.com

[1]Deceased.

because invasive species represent a major threat to biodiversity (Simberloff 2000; Clavero & Garcia-Berthou 2005), and they can also result in dramatically negative economic impacts (Everett 2000; Pimentel *et al.* 2001). Retracing the routes of invasions and determining the source populations of invaded areas define the quality of management strategies either in the source region, along the pathways of invasion or at the points of entry of invaded areas (Hulme 2009).

Identification of source populations and pathways can be achieved either by direct or indirect methods (Estoup & Guillemaud 2010; Guillemaud *et al.* 2010). Direct methods are based on historical data such as interceptions at ports and dates of first discovery in the invaded areas. Indirect methods are based on molecular analyses of populations from both source and invaded areas (Estoup & Guillemaud 2010; Lawson Handley *et al.* 2011). Where direct methods are used, it remains difficult to infer the routes of invasion accurately because the information is invariably incomplete, and even where interception records exist, an introduction will not necessarily result in establishment and a successful invasion. Indirect molecular methods have traditionally compared the genetic diversity and relatedness among populations to retrace the origin and spread of a species (Guillemaud *et al.* 2010). Invaded areas are expected to show lower genetic diversities than source areas due to founder events and subsequent population bottlenecks that can occur during the invasion process (Dlugosch & Parker 2008). However, this assumption is not always appropriate because multiple introductions can result in introduced populations being genetically more diverse than the source populations (Kolbe *et al.* 2004; Roman & Darling 2007). Furthermore, traditional genetic methods do not take into account the stochasticity of the demographic and genetic processes involved (Lawson Handley *et al.* 2011) and they do not reflect the existence of putative unsampled populations that could be the true source of invasive populations.

Recently developed methods using Approximate Bayesian Computation (ABC, Beaumont *et al.* 2002) and coalescent theory make it possible to test complex demographic scenarios and to estimate the likelihood of each scenario (Beaumont 2010; Bertorelle *et al.* 2010; Csillery *et al.* 2010). These methods also allow for the incorporation of historical data such as the dates of first observation in an invaded area. It is consequently possible to hypothesize several scenarios with putative routes of introduction, source populations and genetic make-up of invasive populations and to statistically test for the most likely scenario (Estoup & Guillemaud 2010; Guillemaud *et al.* 2010). Such ABC scenario testing has recently made it possible to determine the routes of introduction of a number of invasive species, revealing in some cases multiple introductions (Miller *et al.* 2005), the presence of unsampled populations in the genetic make-up of populations in invaded areas (Zepeda-Paulo *et al.* 2010) and bridgehead effects, where an invasive population is the source for further invaded areas (Lombaert *et al.* 2010; Ascunce *et al.* 2011). Indirect methods can also reveal the microevolutionary processes involved in a successful introduction. For instance, a recent study of the introduced *Bombus* populations in New Zealand (Lye *et al.* 2011) showed that as few as two specimens of *B. subterraneus* constituted the founder population for a successfully established invasion. Such genetic inference can also enable studies that seek to understand the effect of bottlenecks on population genetic diversity and the role of these processes on invasion success (Facon *et al.* 2011b).

Sirex noctilio Fabricus (Hymenoptera, Siricidae) is an invasive woodwasp species originating from Europe, northern Asia and northern Africa (Spradbery & Kirk 1978) that has spread around the world during the course of the last century. *Sirex* populations remain at low density in the native range and the insect is, therefore, considered as a secondary pest (Hall 1968; Wermelinger & Thomsen 2012). This is in contrast to the invaded areas, where the insect populations are large and they have resulted in serious and damaging outbreaks (Hurley *et al.* 2007; Corley & Villacide 2012). Such invasive population outbreaks have often been explained by the enemy release hypothesis (Keane & Crawley 2002; Hurley *et al.* 2007). The newly invaded areas are unlikely to host the natural enemies (parasites and predators) of the pest, and the population dynamics are freed from the usual natural regulation in the native range. *S. noctilio* has a symbiotic relationship with the fungus, *Amylostereum areolatum* (Slippers *et al.* 2003) that decomposes the wood and provides food for the wasp larval stage, which can result in further wood loss.

The global spread of *S. noctilio* began in the Southern Hemisphere, where it was first recorded in New Zealand around 1900 (Miller & Clark 1935). Subsequently, the wasp spread to Tasmania (1951) and soon after that to mainland Australia in 1961 (Carnegie *et al.* 2005). In the 1980s, *S. noctilio* was recorded in South America: Uruguay (1980, Rebuffo 1988; Maderni 1998), Argentina (1985, Klasmer *et al.* 1998; Klasmer & Botto 2012), Brazil (1988, Iede *et al.* 1988) and in Chile (2001, SAG 2001; Beèche *et al.* 2012). In South Africa, the pest was first detected in 1994 in the Western Cape (Tribe 1995) and it then spread slowly eastwards, through the Eastern Cape (2002) and more recently to Kwazulu-Natal, Mpumalanga and Limpopo (Hurley *et al.* 2007; Hurley unpublished). The North American continent was last to be invaded, with a first detection of *S. noctilio* in

New York State in 2004 (Hoebeke *et al.* 2005) and subsequently in Canada in 2005 (Bergeron *et al.* 2008).

Apart from the first records of *S. noctilio* in newly invaded countries, additional information has been gained through studies on the population structure of the symbiotic fungus *A. areolatum*. Slippers *et al.* (2002) used RFLP analyses of the nuclear intergenic spacer region (IGS) of the rDNA operon to show that the spread of the wasp within the Southern Hemisphere was probably confined to the Southern Hemisphere countries, with no additional introductions from Europe. Analyses of vegetative compatibility groups showed similar results and in particular that isolates from South America and South Africa shared a common origin (Slippers *et al.* 2001). Overall, introductions appear to have occurred serially from one Southern Hemisphere country to another (from Oceania to South America and from there to South Africa) rather than via distinct independent introductions from Europe. More recently, Nielsen *et al.* (2009) used the IGS rDNA of *A. areolatum* from Canada to show that the Canadian genotypes were more similar to those from Europe than to the Southern Hemisphere genotypes, suggesting that Europe was the source of the North American invasion. However, Bergeron *et al.* (2011) showed that one of the *A. areolatum* multilocus genotypes retrieved from the Great Lakes area (Canada) was shared with Chile and South Africa, whereas the other was from an unknown source. This suggests introductions from the Southern Hemisphere to North America and the putative existence of unsampled populations participating in the global spread of the pest.

The historical records of first observations together with the conclusions derived from the genetic analyses of the *S. noctilio–A. areolatum* complex make it possible to establish hypotheses regarding the putative routes of introduction that can be tested. The aim of this study was thus to test these hypothesized routes of introduction for the worldwide spread of *S. noctilio*. In particular, we aimed to identify the source populations of the global invasion and the number of founder specimens involved in the establishment of the invasive populations. To achieve these goals, 14 recently developed microsatellites markers (Santana *et al.* 2009) and sequences of the mitochondrial Cytochrome Oxidase subunit I gene (COI) were used to assess population genetic diversity and relatedness. Apart from traditional analyses, the microsatellite data set was also analysed in a coalescent framework using approximate Bayesian computation that allows statistical testing of distinct scenarios (Cornuet *et al.* 2008, 2010). The results revealed a much more complex scenario than previously believed, illustrating the importance and power of the combined approaches used in this study to better understand and manage the increasing threat of invasive organisms in general, and pests in particular, worldwide.

Material and methods

Sampling

A total of 477 *Sirex noctilio* specimens were sampled from five continents both in the source (Eurasia) and the invaded regions (details of samples are given in Table 1 and Fig. 1). The native geographical distribution of *S. noctilio* is presented on Fig. 1. Sample sizes were low for European countries due to the low densities of populations in these countries and the resulting difficulty to obtain specimens. In contrast, sample sizes for invasive populations displaying outbreaks were substantially higher. In the native area, no samples from Northern Africa and almost no sample from Asia could be obtained. The native population is therefore mostly represented by European samples. A number of specimens were from museums or private collections and were dry (as opposed to other specimens preserved in ethanol). The dried samples included 10 German specimens, four French specimens, two Russian/Mongolian samples and five samples from Finland. Specimens are deposited in the Forestry and Agricultural Biotechnology Institute (University of Pretoria, South Africa) Siricid collection.

Molecular analyses

DNA was extracted from pieces of muscle from the thorax/abdomen junction or in the case of museum specimens, from a single leg. The PrepGEM™ insect DNA extraction kit (ZyGEM, New Zealand) was used and the manufacturer's protocol was followed.

Mitochondrial gene. A portion of the 5′ end of the Cytochrome Oxydase subunit I (COI) mitochondrial gene was amplified using COI-hymF (5′-CGG CGC CAT TTA TAA TTG GAG GWT TTG GWA A-3′) and COI-hymR (5′-GCA CCG ACT GCT CCT ATA GAT AAA ACA TAR TGR AA-3′). These primers were specifically designed for this study following the approach described in Hoareau & Boissin (2010) and based on the available mitochondrial genomes of Hymenoptera retrieved from the GenBank database (see Table S1, Supporting information). This set of primers should, therefore, be useful for any species belonging to Hymenoptera. These primers frame a fragment of about 900 bp of the COI gene. Additionally, because the long COI fragment could not be amplified for the museum specimens, internal primers specific to *S. noctilio* were designed to amplify shorter (300 bp) COI

Table 1 Sampling details of the *Sirex noctilio* specimens analysed in this study

Continent	Country	Locality and/or Coordinates	Sampling year	*N*	*N* female	*N* male
North America	Canada	South East Ontario (Great Lakes area)	2010	49 (18)	49	–
	USA	New York State	2006	5 (1)	5	–
		sub-TOTAL		54	54	–
Europe/Asia	Switzerland	Valais (cf. Wermelinger *et al.* 2008)	2001–2004	101 (14)	71	30
	Germany	Berlin, Stuttgart	1993	10 (3)	6	4
	France	Anglet, Villes-sur-Auzon, Bosmie-l'Aiguille, Meolans-Revel	1995–2004	4 (2)	4	–
	Greece	Vitina, Peloponissos	2004	2	2	–
	Spain	Muxica (Biscay Province, Basque Country)	2004–2011	5 (2)	3	2
	Czech Republic	Trebic 49°05′43″N/15°57′53″E	2007	3	3	–
	Finland	Helsinki	1921–1961	5	5	–
	Russia &Mongolia	Moscow/NE Ulanbaatar	–	5	5	–
		sub-TOTAL		135	99	36
South America	Argentina	Patagonia, Inta 41°07′80″S/71°16′10″W	2004	82 (14)	72	10
	Chile	La Union 40°17′0″S/73°05′0″W	2006	30 (22)	18	12
	Uruguay	Tacuarembó 31°42′51″S/55°58′57″W	2006	11 (1)	7	4
		sub-TOTAL		123	97	26
Oceania	New Zealand	North Island 38°36′11.81″S/176°34′39.18″E	2009	25	25	–
	Australia	New South Wales (Tumut, Tumba, Bombala)	2004	40 (15)	20	20
		sub-TOTAL		65	45	20
Africa	South Africa	Ngome (Natal)	2007	12	12	–
		Midlands (Natal)	2005–2006	44 (13)	12	32
		Western Cape	2004	23 (6)	14	9
		Eastern Cape	2005–2006	21 (2)	3	18
		sub-TOTAL		100	41	59
		TOTAL		477	336	141

N = number of samples analysed with the microsatellite markers. Numbers in parentheses are the number of samples also analysed at the mitochondrial Cytochrome Oxydase subunit I (COI) gene.

fragments from dried specimens (300F: 5′-CCTGT TCT TGCTGGAGCAATCAC-3′; 300R: 5′-GTGATT GCTC CAGC AAGAACAGG-3′; 600R: 5′-CGCGGCCAACAGTAA ATA-TATGATGAGCTC-3′). Previously obtained sequence data (Slippers, unpublished) were also added to this data set.

PCR amplifications were performed in a 15 μL volume containing: 1.5 mM of MgCl2, 0.2 mM of each dNTP, 1X final concentration of buffer, 0.5 μM of each primer, 0.25 unit of FastStart Taq DNA Polymerase (Roche®), 1.5 μL of DNA template and sterile water up to 15 μL. The PCR cycling parameters were as follows: 4 min of denaturation at 95 °C followed by 40 cycles of (i) denaturation at 94 °C for 45 sec, (ii) annealing at 50 °C for 60 sec and (iii) extension at 72 °C for 80 sec; completed with a final extension step at 72 °C for 4 min. PCR products were run on 1.5% agarose gel to check for correct amplification and cleaned using the Sephadex® protocol (Sigma-Aldrich). The sequencing PCR was performed in 10 μL containing 1 μL of the PCR product, 0.7 μL of Big Dye® Terminator v3.1 (Applied Biosystems), 2.5 μL of sequencing buffer (provided with Big Dye®), 1 μL of the primer and 4.8 μL of Sabax® sterilized water (Adcock Ingram). Cycling parameters were as follows: 15 sec at 96 °C, 15 sec at 59 °C and 4 min at 60 °C, repeated 25 times. The resulting products were once again cleaned using the Sephadex® protocol (Sigma-Aldrich) and sent for sequencing to a partner laboratory (DNA sequencing facility, Faculty of Natural and Agricultural Sciences, University of Pretoria).

Nuclear markers. A total of 14 recently developed microsatellite markers (Santana *et al.* 2009) were used in this study. PCRs were carried out in 11 μL using 4.5 μL of QIAGEN Multiplex PCR master mix, 2.3 μL of RNase free water (provided with the QIAGEN master mix), 0.20 μL of each primer (at a concentration of 0.5 mM) and 1 μL of DNA template. Two panels of eight and six primer pairs, respectively, were run for each specimen. Cycling parameters were as follows: 15 sec at 95 °C, 60 sec at 94 °C and 90 sec at 60 °C, repeated 50 times. PCR products were run on 1.5% agarose gels and prepare for GeneScan (9 μL of a solution containing formamide and 1:1000 Liz500 size standard (Applied Biosystems) together with 1 μL of PCR product). PCR products were run on an ABI3100 DNA analyser to determine DNA sizes (DNA

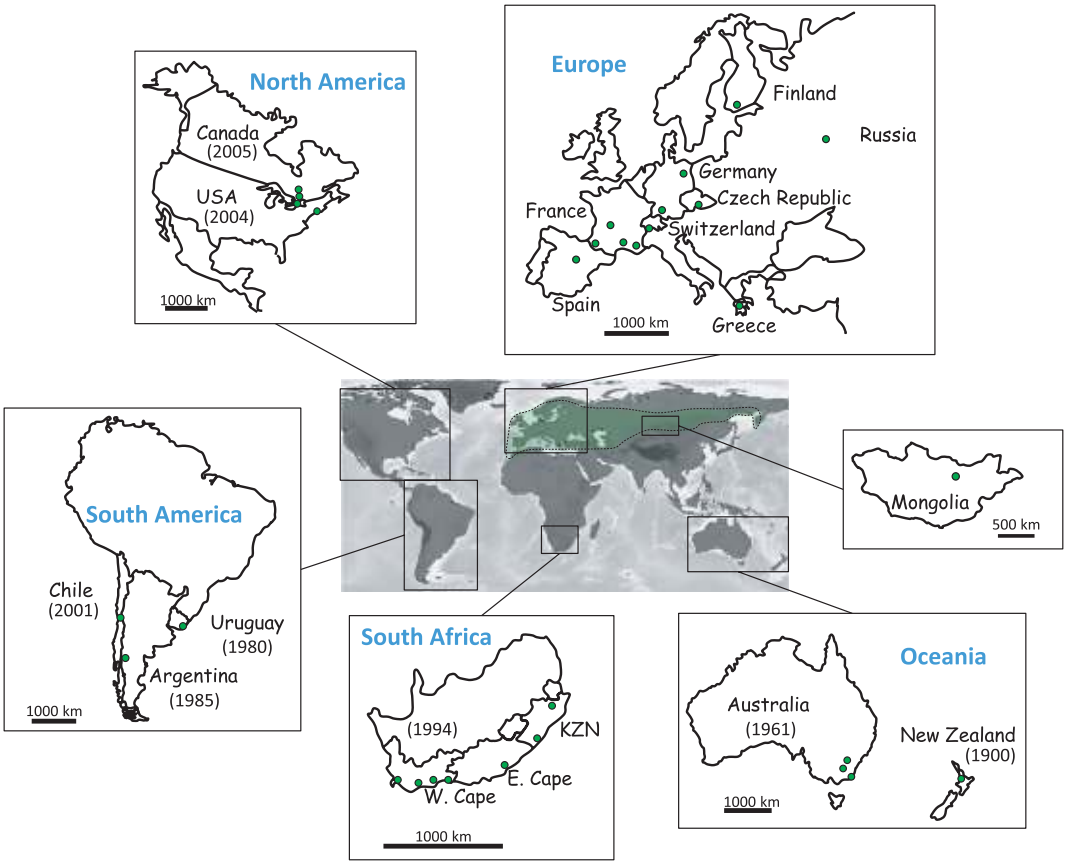

Fig. 1 Map of collection sites of *Sirex noctilio* samples analysed in this study. The native area of *S. noctilio* is indicated by the outlined shadow area of the central map and is based on Carnegie *et al.* 2006. Dates are indicative of the first detection of the pest in each country.

sequencing facility, University of Pretoria). GeneMarker® v1.90 (SoftGenetics LLC®) was used to score alleles and genotype specimens.

Data analyses

Mitochondrial data. Sequences were aligned in Bioedit (Hall 1999) and manually checked. Haplotype and nucleotide diversity indices along with their standard deviations were calculated in DnaSP v5.10 (Librado & Rozas 2009). To investigate the relationships between populations, haplotype networks were reconstructed using Network v.4.6 (www.fluxus-engineering.com). As the COI fragments amplified for museum specimens and previously obtained sequences were short, two distinct haplotype networks were built. One encompassed the specimens for which 818 bp could be recovered and a second included all the specimens. For the latter network, the longer COI sequences were cut to match the shorter fragments (444 bp).

Microsatellites data. Because of the haplo-diploid life cycle of the species, some analyses were only run on the female data set, when the software could not handle haploid data (e.g. Genetix). However, other software could handle simultaneously the diploid females and the haploid males (Structure and DIYABC) and the whole data set was therefore used.

Genetic diversity. The genetic diversity of each sample was estimated by calculating the unbiased expected heterozygosity (H_E, Nei 1978) and the total number of alleles (A) in GenClone (Arnaud-Haond & Belkhir 2007) and the mean number of alleles (mean A, for females only) in Genetix v.4.05.2 (Belkhir *et al.* 2004). Additionally, in GenClone, a rarefaction method allows for comparison of the number of alleles, the number of multilocus genotypes and heterozygosity among samples with different sample sizes (A_s, MLG_s and H_{Es} respectively). The three parameters A_s, MLG_s and H_{Es} were compared for a standard size of 41 females corresponding to the sample with the lowest number of females (South Africa) and for a standard size of 20 males corresponding to the smallest male sample (Oceania).

Genetic relatedness. The genetic relatedness of populations was assessed using two methods: Factorial Corre-

spondence Analysis (FCA) available in GENETIX and Bayesian clustering using STRUCTURE v.2.3.3 (Pritchard *et al.* 2000). The FCA is a multivariate analysis based on distances between multilocus genotypes of individuals and does not make any assumption on the underlying population genetic model. STRUCTURE, on the other hand, tries to minimize the Hardy–Weinberg Disequilibrium and Linkage Disequilibrium within clusters to find the most likely number of clusters in a data set. The FCA was conducted on females only because the software accepts only diploid data. This analysis was carried out on the genotypic data using both individuals and populations as centres of inertia. STRUCTURE was used to estimate the number of populations (K) most likely present in the samples. To control for any potential biases related to the haplo-diploid life cycle of the species, several analyses were conducted. These were (i) on the whole data set (considering males as diploid and coding the second allele of males as missing as advised by the program developer, 477 specimens); (ii) on the female data set alone (336 specimens); and (iii) on the male data set alone (coded as haploid; 141 specimens). After the first batch of runs and following the uppermost level of clustering (see Results), the whole data set was further investigated by running additional analyses on each cluster separately to search for any more subtle sub-structuring. After several attempts and comparisons of likelihood, the parameters were set assuming admixture and a correlated allele frequency model. Sampling locations were not used as priors but the results were similar when using locations as priors (data not shown). The burn-in length was set to 100 000 and the simulations to 100 000 repetitions. Each run was iterated 15 times. To determine the most probable partition of the data, *K* was determined using the Δ*K* approach (Evanno *et al.* 2005) as implemented in the recently developed CORRSIEVE software (Campana *et al.* 2011).

Scenario testing. The software DIYABC (Cornuet *et al.* 2008, 2010) was used to discriminate between distinct possible introduction routes of the woodwasp *S. noctilio* across the world and from the Eurasian source. Because as many as 120 scenarios (5! = 5*4*3*2) can be tested with five populations, this number was narrowed down using both the known historical dates of the first record in invaded countries and hypotheses emerging from previous genetic studies of the symbiotic fungus *A. areolatum* (see Introduction). Additionally, because of the clustering analyses revealing two distinct gene pools (see Results), the scenario testing was applied on the two clusters separately.

Regarding cluster 1, a total of 11 distinct scenarios were considered (Figs 2A and S1, Supporting information). Scenarios 1 and 2 involve four independent introductions from Europe. However, in scenario 2, all the populations are admixed between Europe and another source, whereas in scenario 1, there is no admixture. Scenarios 3, 4 and 5 each involve three independent introductions from Europe. Scenarios 6, 7 and 8 involve two independent introductions from Europe. Scenarios 9, 10 and 11 involve a single introduction from Europe followed by subsequent serial introductions among countries in the invaded range. These scenarios differ in the number of admixed populations and the sequence of introduction. Regarding cluster 2, eight distinct scenarios were tested (Figs 2B and S1, Supporting information) with a special emphasis on the Switzerland sample due to its differentiation from the rest of Europe in the Structure plot. In particular, an unsampled population was added in scenarios 17, 18 and 19 to test for an unknown source of the second gene pool.

For both sets of analyses, prior distributions were uniform and defined as follows: $10 < N < 10000$; $1 < N_f < 100$; $5 < db < 25$; $1 < t_1 < t_2 < t_3 < t_4 < 100$; '*N*' being the effective population size, 'N_f' being the effective number of founder specimens, 'db' being the bottleneck duration in generations and '*t*' the time in generations. Priors were deliberately defined broadly because no prior information exists. Default values of prior distributions of the different parameters for the mutation model of the microsatellites were used. The mutation model was therefore the Generalized Stepwise Mutation model (Estoup *et al.* 2002) with a uniform prior distribution for the mean mutation rate ($1E^{-4}$ to $1E^{-3}$) and a uniform prior distribution for the parameter of the geometric distribution ($1E^{-1}$ to $3E^{-1}$). The 'one sample summary statistics' used were the mean number of alleles, the mean genic diversity and the mean size variance. The 'two sample summary statistics' used were the same as the 'one sample summary statistics' together with F_{ST}. For each scenario, between 600 000 (for scenarios of cluster 1) and 1 000 000 (for scenarios of cluster 2) simulated data sets were created. Prior-scenario combinations were evaluated by performing a PCA as implemented in the software. Posterior probabilities of scenarios were compared using a logistic regression as implemented in the software and using the 1% closest simulated data sets. Results were similar when using more or less closest data sets (e.g. 0.1%, 2%, 5%, data not shown) or when running a pre-processing step (Linear Discriminant Analysis) on the summary statistics before computing the logistic regression (Estoup *et al.* 2012; data not shown). Estimations of parameters were also computed. The performance of parameter estimations was assessed by computing the relative bias and the relative root mean square error (available in the software), using the mode as point estimate and 500 test data sets.

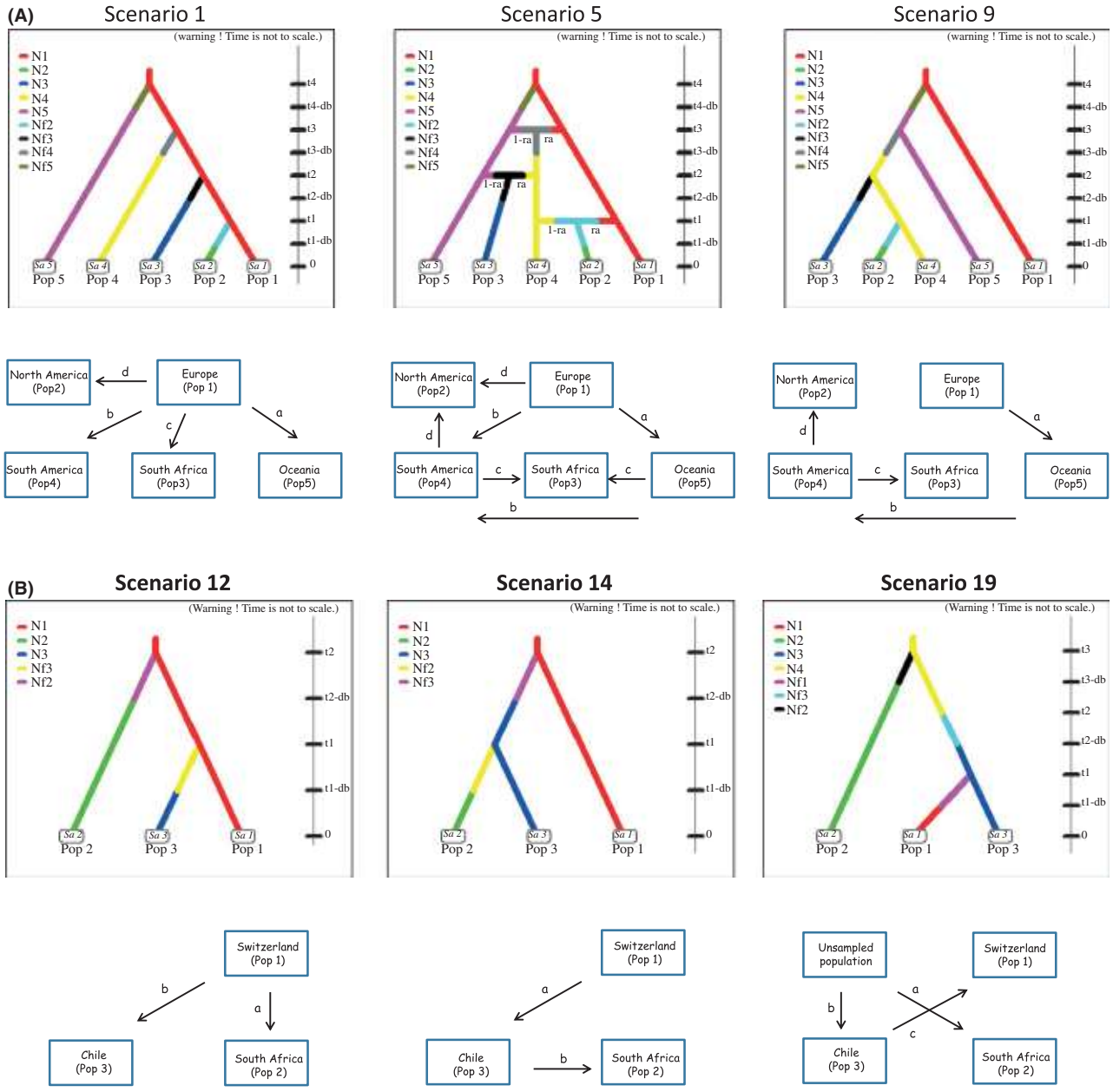

Fig. 2 Schematization of the routes of introduction and genealogy of scenarios under a coalescent framework such as implemented in the software DIYABC. (A) Example of three contrasting scenarios for cluster 1. N1–N5: effective population sizes for populations 1–5. Nf2–Nf5: effective number of founder specimens for invasive populations 2–5; (B) Example of three contrasting scenarios for cluster 2. N1–N3: effective population sizes for populations 1–3. Nf1 to Nf3: effective number of founder specimens for invasive populations 1–3. The letters a, b, c and d indicate the sequence of the invasion ('a' first, followed by 'b', etc.).

Results

Haplotype networks

A total of seven haplotypes were recovered from the 50 specimens analysed for the long fragment (818 bp; Fig. 3A), and seven haplotypes were recovered from the 113 specimens analysed for the shorter fragment (444 bp; Fig. 3B). Haplotype and nucleotide diversities were very low [Hd = 0.656 (SD = 0.054) and Pi = 0.00147 (SD = 0.00014), Hd = 0.657 (SD = 0.028) and Pi = 0.00183 (SD = 0.00013), for the long and short fragments respectively]. The total number of variable sites was five for the short fragment and six for the long fragment. Haplotypes did not cluster by geography or by the gene pools (Fig. 3) defined by the Structure analysis (Fig. 5A). The dominant

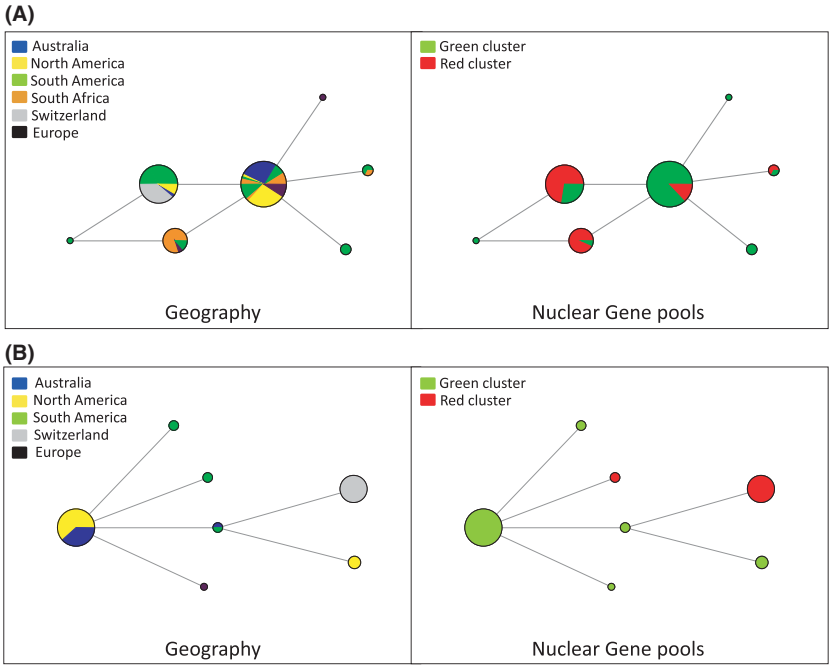

Fig. 3 Haplotype networks of the Cytochrome Oxydase subunit I (COI) mitochondrial gene. (A) for the entire fragment (818 bp, 50 specimens); (B) for the shorter fragment (444 bp, 113 specimens). Circle sizes are proportional to the number of sequences per haplotypes. Colours correspond to localities on the left graphs and clusters on the right graphs. Following the Structure results, Switzerland and the rest of Europe are considered separately.

haplotype on Fig. 3B was shared by all the countries/continents except Switzerland.

Genetic diversity

The European female sample showed the highest values of mean number of alleles (A), total number of alleles and total number of alleles for the standard size comparison (Table 2A). However, the South America female sample showed the highest multilocus genotype number and the highest heterozygosity value. The male data set provided the same results except that South America also had the highest total number of alleles (Table 2B). Samples from South Africa and Oceania always showed the lowest values of genetic diversity in both data sets, whereas those from Europe, South America and North America showed highest values. Allele frequencies for each female population are provided in Table S2 (Supporting information). The total number of alleles across populations was 94 and 26 were occurring in a single population (i.e. private alleles, see Table S2, Supporting information).

Genetic relatedness

Only females were considered in the FCA because the GENETIX software considers only diploid specimens (Fig. 4). The first axis distinguished samples from cluster 1 and cluster 2 defined by the Structure results (see below). The second axis distinguished North America from the rest of cluster 1. The third axes discriminated South Africa from the rest of cluster 2. This latter axis also helped to distinguish within cluster 1 between Oceania (Australia and New Zealand) and South America (Argentina, Uruguay and Chile), but with some overlap between them.

The uppermost level of differentiation in the Structure plot was found to be $K = 2$ clusters when considering the whole data set (Fig. 5A). Noticeably, samples from Europe were divided among the two clusters with most of the samples grouped in one cluster (cluster 1) except for Switzerland that formed a separate cluster (cluster 2). Most of the samples from Chile and from South Africa clustered with the Switzerland sample in cluster 2. For South Africa, however, most of the samples from Western Cape, the area of first introduction, belonged to cluster 1. Therefore, both samples from South Africa and South America represent a mixture of the two clusters. The rest of the samples (Oceania, North America and South America) clustered with the European samples in cluster 1.

When considering the two clusters separately and running further Bayesian clustering analyses (Fig. 5B), cluster 1 showed a further subdivision into three subgroups and cluster 2 showed a further subdivision into two subgroups. Within cluster 1, most of the samples from Europe, Argentina and Uruguay group together (dark orange cluster). Samples from South America resided in all three clusters. Samples from North America resided in a cluster on their own

Table 2 (A) Summary statistics of the 336 *Sirex noctilio* females; (B) summary statistics of the 141 *S. noctilio* males

(A) Females	*N*	Mean A	A_T	A_S	MLG_T	MLG_S	H_{ET}	H_{ES}
Europe	99	5.5	104	78.93 (0.81)	78	35.08 (0.19)	0.43	0.42 (0.003)
North America	54	3.28	59	56.14 (0.15)	51	39.27 (0.08)	0.42	0.40 (0.001)
South America	97	5	95	74.79 (0.51)	97	41 (0.00)	0.49	0.48 (0.002)
Oceania	45	2.79	57	55.6 (0.17)	37	34.29 (0.09)	0.34	0.34 (0.001)
South Africa	41	2.5	46	46	23	23	0.34	0.34

(B) Males	*N*	A_T	A_S	MLG_T	MLG_S	H_{ET}	H_{ES}
Europe	36	37	31.78 (0.30)	15	10.85 (0.14)	0.24	0.24 (0.003)
South America	26	46	42.94 (0.16)	25	19.42 (0.05)	0.41	0.40 (0.001)
Oceania	20	19	19	9	9	0.13	0.13
South Africa	59	30	26.14 (0.18)	21	9.98 (0.17)	0.18	0.18 (0.003)

N = number of specimens, A = number of alleles, MLG = multilocus genotypes, H_E = unbiased heterozygosity, the T in indices refers to the total sample size, the S in indices refers to the re-sampling procedure (considering only 41 specimens per samples for females and 20 specimens per samples for males); values between parentheses are standard deviations.

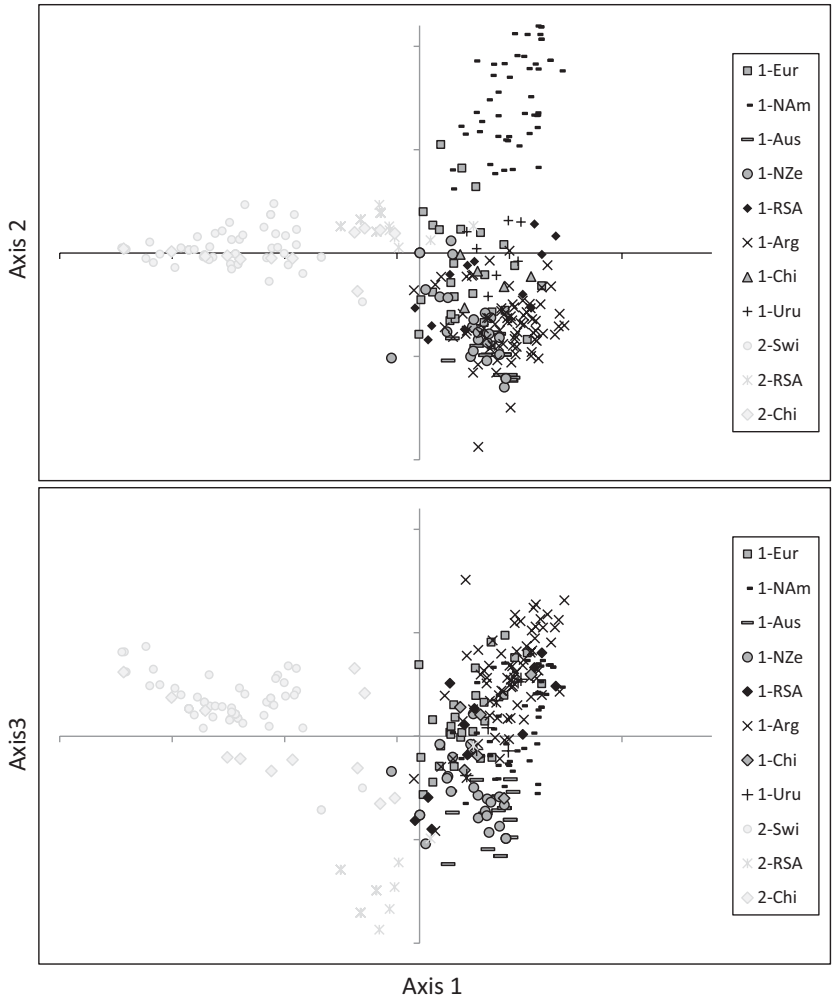

Fig. 4 Factorial Correspondence Analysis on the female data set of *Sirex noctilio* (*N* = 336). Samples are split in cluster 1 (1-followed by a 3 letter code) and cluster 2 (2-followed by a 3 letter code). 3 letter codes are as follows: Eur = Europe (without Switzerland), Swi = Switzerland, NAm = North America, Aus = Australia, NZe = New Zealand, RSA = South Africa, Arg = Argentina, Chi = Chile, Uru = Uruguay.

together with specimens from Europe (light yellow cluster). Samples from Oceania (Australia and New Zealand) clustered with South Africa and Chile (medium orange cluster). New Zealand samples were, however, more diverse than those from Australia and showed all three clusters. In cluster 2, Chile and Switzerland sam-

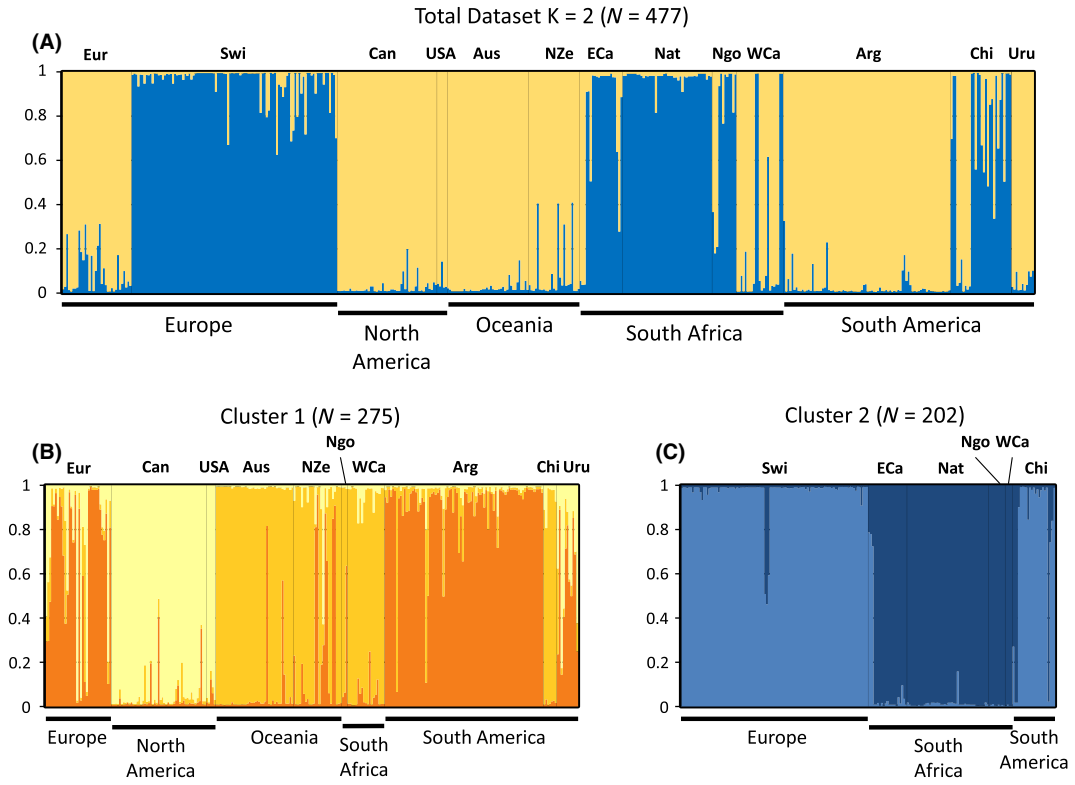

Fig. 5 Bayesian clustering analysis: (A) on the whole data set (N = 477); (B) on cluster 1 (N = 275); (C) on cluster 2 (N = 202). Each vertical column represents a specimen. Continents are indicated below the plots, sampling localities are indicated above the plots and are delimited by black vertical lines. Locality codes as in Fig. 4 with the addition of USA = United Stated of America, Can = Canada, WCa = Western Cape, ECa = Eastern Cape, Nat = Natal, Ngo = Ngome.

ples grouped together, while samples from South Africa belonged to a separate sub-group (Fig. 5C).

When considering males and females separately, the results were equivalent. The female data set had a most likely number of clusters of K = 2, showing the same clustering as the whole data set (Fig. S2A, Supporting information). For the male data set, a most likely number of clusters of K = 4 was found (Fig. S2B, Supporting information) and this also corresponds to the same partition as the whole data set with two sub-groups within each cluster. The third group of cluster 1 was absent because no male was analysed from North America.

Scenario testing

In the scenario testing analyses and when considering only cluster 1 (275 specimens, five populations), the scenario with the highest likelihood was scenario 5 (Fig. 6A). In this case, there were three independent introductions from Europe together with serial introductions within the Southern Hemisphere and to North America. North America appeared to be represented by a mixture of specimens from Europe and South America. South Africa represented an admixture from Oceania and South America. South America represented a mixture of specimens from Oceania and Europe.

When considering cluster 2 (202 specimens, three populations), the scenario testing revealed that samples from Switzerland, Chile and South Africa most likely originated via introductions from an unsampled population (Fig. 6B). Scenario 19 was the most likely, with two independent introductions into South Africa and Chile from the unknown source followed by a serial introduction from Chile into Switzerland rather than another independent introduction from the unsampled population into Switzerland (as in scenario 17).

The estimation of the effective number of founders for each population within each cluster (N_f; Fig. 6) appeared to be robust as the biases indices for each parameter are close to 0 (Table S3, Supporting information). The effective number of founders in cluster 2 varied from 3 to 74 and the ones for cluster 1 varied from 15 to 73 (Fig. 6).

The inferred routes of introduction resulting from both the clustering analyses organized by continent and the scenario testing are summarized in Figure 7.

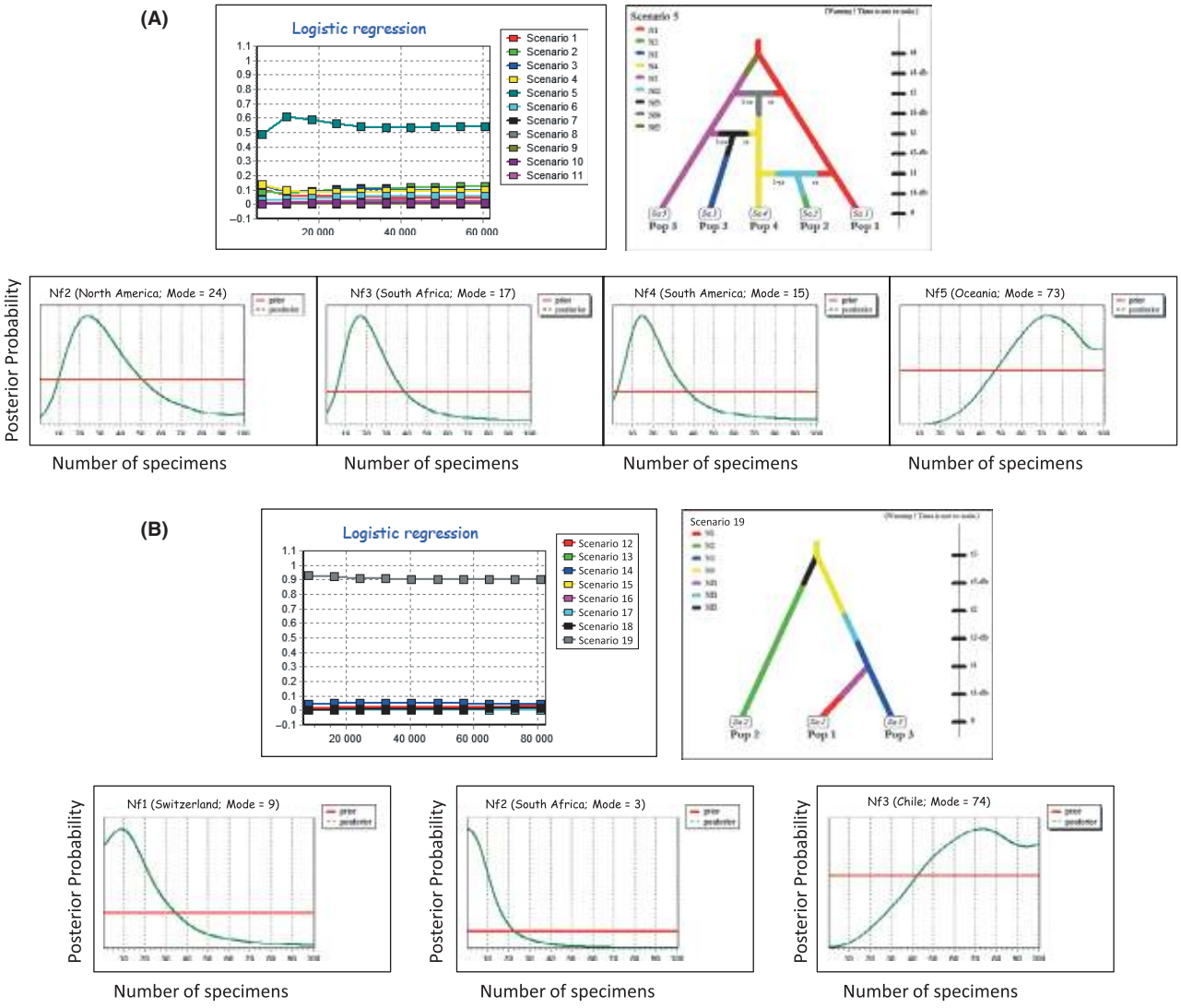

Fig. 6 Logistic regression plot showing the Posterior Probability of each scenario from the DIYABC analysis. The most likely scenario is shown as a reminder on the right side of the logistic regression plot (pop1 = Europe, pop2 = North America, pop3 = South Africa, pop4 = South America, pop5 = Oceania in panel A and pop1 = Switzerland, pop2 = South Africa and pop3 = Chile in panel B). The graphs below show the Posterior Probability (y-axis) of estimates of the effective number of founders (N_f, x-axis) for each population. (A) for cluster 1 and (B) for cluster 2.

Discussion

This study illustrates the complementarity of using traditional genetic tools and newly developed methodologies when investigating routes of invasion. This provides a robust statistical framework enabling to discriminate between different hypothesized scenarios. Analyses of a large number of samples from five continents with polymorphic microsatellites (and COI sequences) provided substantial insight into the probable routes of invasion of *S. noctilio* and they also revealed various unexpected results. The genetic make-up of most of the samples was much more complex than formerly believed and populations in most continents appeared to be the products of admixture from several sources. Particularly, two distinct source areas appear to account for the spread of the pest (Europe and an unknown source) and European pines could for the first time be considered threatened by *Sirex noctilio* invasion.

Multiple origins of threat: unexpected detection of two distinct gene pools and the existence of an unsampled source population

Clustering analyses unexpectedly revealed the presence of two distinct gene pools in the data set (herein referred to as cluster 1 and cluster 2, Fig. 5). Cluster 1 was predominant and was recovered from all populations except that from Switzerland. Cluster 2 was recovered only from

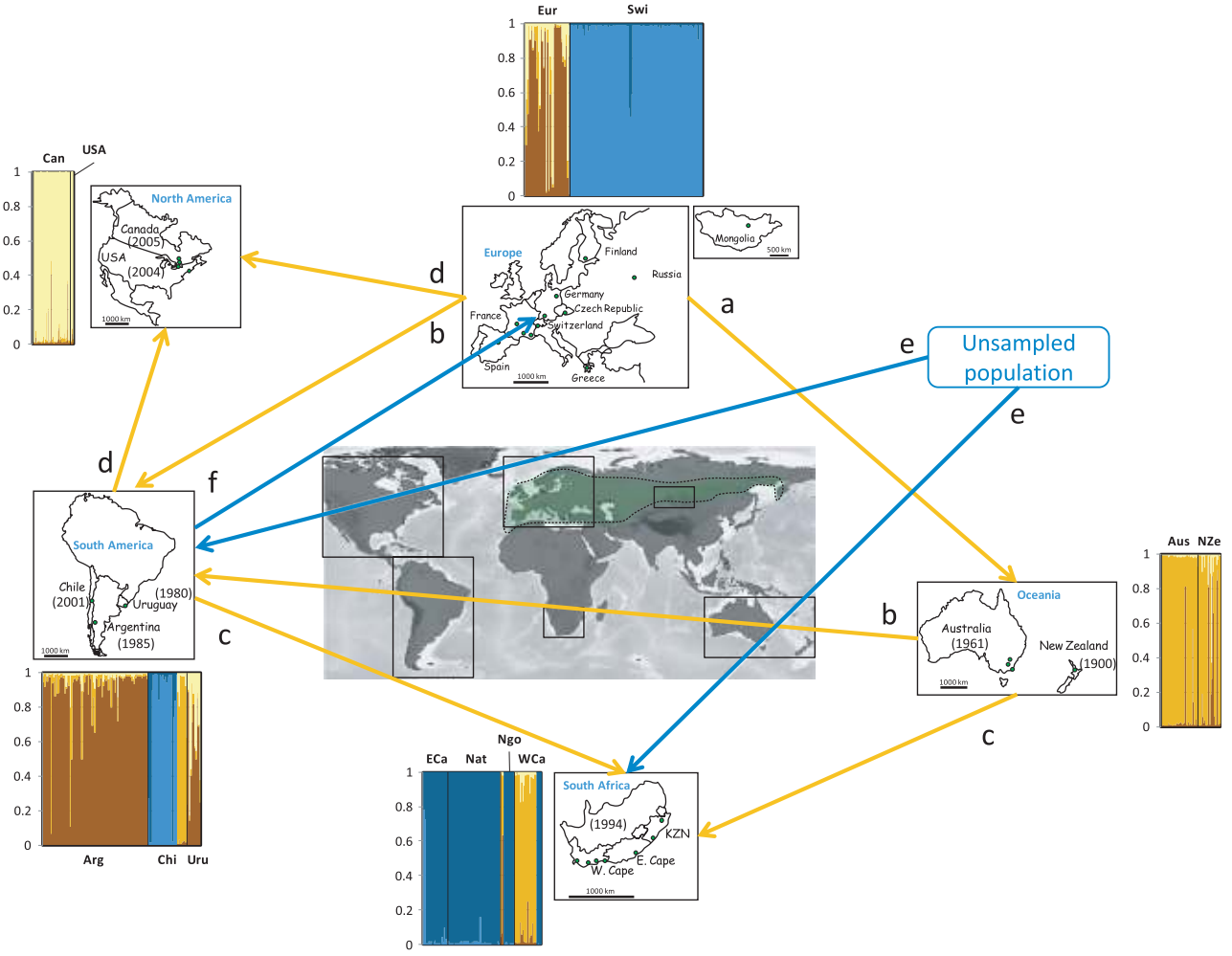

Fig. 7 Summary of the Bayesian clustering analyses per continent and schematization of the worldwide spread of *Sirex noctilio* inferred from this study. The letters a, b, c and d indicate sequence of invasion ('a' first, followed by 'b', etc.) for cluster 1 and e, f indicate sequence of invasion ('e' first, followed by 'f') for cluster 2.

Switzerland, South Africa and Chile. Further scenario testing on the second gene pool revealed that an unsampled source population is most likely to have been involved in the overall genetic make-up (scenario 19, Figs 2B and 6B) and Switzerland has been invaded by specimens from this cluster. Recognition of the Switzerland population as 'introduced' was surprising because *S. noctilio* is native in Europe. The population, however, represents a genetic make-up completely different from all the other surrounding populations sampled in Europe, which all resided in cluster 1. Furthermore, this population is exceptional as it was sampled from an area where the *S. noctilio* population increased unexpectedly during a period of drought-induced pine decline in the early 2000s (Wermelinger & Thomsen 2012). Populations in Europe do not usually result in outbreaks as their numbers are regulated by predators and parasitoids, as well as host and other factors with which they have co-evolved. Conversely, invasive populations are released from their natural enemies in their new environment and can undergo rapid and unregulated population growth (Keane & Crawley 2002). *S. noctilio* is, however, native to Europe and natural enemies must, therefore, be present in Switzerland. The population of natural enemies might, for various reasons, have not matched to the sudden outbreak of the *S. noctilio* population, as already seen in gall wasps (Schonrogge *et al.* 2006). In any case, the genetic data presented in this study support a non-native origin of the Switzerland population. This suggests that Europe should not be ignored in terms of invasion by *S. noctilio* from other areas. At the same time, this species is native and, thus, probably much more firmly embedded in a regulating environment than in areas of introduction where the species was previously absent.

The abundance of the second gene pool in most of the samples from South Africa could explain the poor results obtained with the biological control of *S. noctilio*, using the entomopathogenic nematode *Deladenus siricidicola*, in these areas (Hurley *et al.* 2007; Slippers *et al.* 2012). Indeed, the selected nematode strain used in biological control programmes across the Southern Hemisphere was selected from Australia and could, therefore, be unsuitable as our study revealed only the presence of cluster 1 in Australia. Genetic analyses revealed that these biological control populations from the Southern Hemisphere are exceptionally homozygous (Mlonyeni *et al.* 2011). Earlier studies have suggested that nematode–wasp interactions might be coevolved and specific to certain regions or populations (Bedding 1984). For instance, a strain of the nematode used in experiments could sterilize an Australian population of *S. noctilio*, but not a Belgian population of the wasp. Further laboratory experiments are required that would focus on developing a strain of *D. siricidicola* efficient against the second gene pool.

The likelihood of an unsampled source population for global invasive populations of *S. noctilio* was also recently suggested in a worldwide study on the fungal symbiont of the wasp, *A. areolatum* (Bergeron *et al.* 2011). The authors reported a multilocus genotype (MLG) that did not match with any MLG from Europe or invaded areas around the world included in that study. A potential unsampled source of the spread may be other native areas of *S. noctilio:* Asia or northern Africa. Further sampling in these regions would help to clarify the source area of this genetic cluster 2.

The worldwide routes of introduction of S. noctilio

The presence of two potential sources for the *S. noctilio* invasion complicates clarification of the pathways of introduction of the pest. Two invasion histories occurred in parallel, the one represented by cluster 2 and the other by cluster 1. Regarding cluster 2, the most likely scenario (scenario 19, Fig. 2B, Fig. 6B) involves an independent introduction to South Africa from an unsampled population and a second independent introduction into Chile, followed by a serial introduction from Chile to Switzerland. These results reflect the FCA and the clustering analysis that revealed a close relationship between populations from Switzerland and Chile (Figs 4 and 5B). This route of invasion appears to be quite recent given that Chile was invaded around 2001 (SAG 2001). In Switzerland, *S. noctilio* was favoured by a drought period in the early 2000s (Wermelinger & Thomsen 2012) and the regions of South Africa where this cluster 2 occurred (Eastern Cape and Natal) were also colonized after 2001 (Hurley *et al.* 2007).

The most likely scenario regarding invasion of cluster 1 (Fig. 6A) emerged as a complex situation involving three independent introductions from Europe and the admixture nature of three populations. South American and South African populations are admixtures from Europe and Oceania and from Oceania and South America respectively. Bayesian clustering revealed a similar situation where (i) serial introductions within the Southern Hemisphere were probable as New Zealand, Australia, South Africa and Chile belong to the same sub-cluster and (ii) independent introductions from Europe to Argentina and Uruguay were probable as they belong to the same sub-cluster. Those two analyses, therefore, confirmed the spread within the Southern Hemisphere hypothesized by Slippers *et al.* (2002). However, they also suggest multiple introductions of the invaded areas due to independent introductions from Europe. The invasive North American population was also an admixture composed both of specimens from Europe and South America. This result is similar to that of Bergeron *et al.* (2011) who found that 26% of their *A. areolatum* samples from Canada belonged to a given MLG shared with Chile and South Africa. This supports the view that some introductions from the Southern Hemisphere into the Northern Hemisphere occurred.

The results of this study show that the global pathways of introduction for *S. noctilio* are much more complex than previously assumed (Fig. 7). This is especially true regarding the fact that there have been two source areas for invasion into new areas. These results are consistent with recent genetic studies of invasive species that revealed much more complex and unexpected scenarios than suggested by historical data alone (Miller *et al.* 2005; Hoos *et al.* 2010; Lombaert *et al.* 2010).

Effects of propagule pressure on genetic diversity and invasion success

Introduction of a species into new areas triggers a loss in genetic diversity through founder events. This results from a small portion of the source's total genetic diversity being sampled (Nei *et al.* 1975; Dlugosch & Parker 2008). Furthermore, the small sizes of the newly established populations make them prone to subsequent genetic drift and bottlenecks (Dlugosch & Parker 2008; Miller *et al.* 2009). Results of this study illustrate the loss of genetic diversity through a founder event, for instance, between New Zealand and Australia, as illustrated by allele frequencies (more fixed loci for Australia) and genetic diversity (Tables S2 (Supporting information) and 1, Fig. 5). This is consistent with the knowledge that the Australian invasion originated from

New Zealand. It also probably reflects the strict and successful quarantine system in Australia, which focused on preventing *S. noctilio* introductions and spread since the 1950s (Neumann *et al.* 1987), therefore minimizing the propagule pressure on the country. In invasion biology, the role of propagule pressure is well known to influence the overall genetic diversity of a newly established population (Simberloff 2009). Propagule pressure will depend on (i) the number of founder specimens and (ii) the number of independent introductions (Simberloff 2009).

The number of founder specimens. The ABC method allowed an estimation of the effective number of founder individuals for each population, a demographic parameter that can facilitate the understanding of the microevolutionary processes associated with an introduction. The estimated effective number of founders ranged from 3 to 74 (Fig. 6). The fact that so few founder individuals would have resulted in some of the successfully established and aggressively spreading populations might be considered surprising. However, Lye *et al.* (2011) recently showed that some populations of bumblebees in New Zealand were the result of only two founder specimens. Facon *et al.* (2011b) recently demonstrated that bottlenecks of intermediate intensity may help to purge deleterious alleles, therefore lowering the inbreeding depression of small populations. This finding helps to resolve the paradox of the success of invasive populations after they have undergone a depletion of genetic diversity, known to be associated with a decrease in the potential to adapt to new environments (Frankham *et al.* 1999) and an increase in extinction risk (Garza & Williamson 2001). Furthermore, the haplo-diploid nature of *S. noctilio* provides an additional route to purge deleterious alleles in the haploid males (Werren 1993) and will allow populations to persist with very low genetic diversities. This might at least partly explain the success of establishment and spread of *S. noctilio* worldwide.

The number of independent introductions. Most of the populations considered in this study appear to be the result of admixture events. This suggests a number of independent introductions, therefore increasing the propagule pressure. Interestingly, samples from Oceania showed the highest number of founder specimens for cluster 1 (Fig. 6A). As the area was invaded first, the probability of multiple introductions during the course of more than 100 years is higher than the one for areas invaded in the last 30 years. However, Chile in cluster 2 also displayed a high number of founder specimens (N_f = 74, Fig. 6B), whereas the introduction is much more recent (around 2001). Both the introduction of numerous specimens and multiple introductions may explain the high number of founder specimens in Chile. The populations considered in this study illustrated the possible effect of several introductions on genetic diversity. The sample from South America included the two genetic pools and the three sub-clusters of cluster 1 (Fig. 5). It, therefore, presented high values of genetic diversity comparable or superior to areas where *S. noctilio* is native (Table 2A,B). Recently, population admixture has been shown to allow heterosis effects in the harlequin ladybird, where offspring develop more quickly and are slightly bigger than their parents (Facon *et al.* 2011a). Similarly, laboratory experiments on an invasive snail showed an increased invasiveness for the admixed offspring compared with their parents (Facon *et al.* 2005, 2008). Such heterosis could, for instance, lead to failure of biological control measures. In the case of *S. noctilio*, the effects of the admixture between cluster 1 and cluster 2, in particular, should be the focus of further laboratory experiments.

Conclusions

Results of this study revealed a complex genetic make-up for a global sample of *Sirex noctilio* and they also illustrated complex pathways of invasion. The discovery of an unsampled source population will have drastic consequences on the approaches undertaken to contain the spread of the pest in various part of the world. In particular, attempts to manage *S. noctilio* populations using biological control will need to consider the existence of the two distinct gene pools revealed in this study. Furthermore, monitoring should be considered for European countries that have previously not been concerned with this pest due to its perceived native status. Future studies should seek to identify the unsampled source area for *S. noctilio* revealed in this investigation. This should diminish the number of parallel routes of introduction and minimize the potential admixture of the two sources, which would, then, prevent subsequent heterosis effects that could exacerbate the invasiveness of *S. noctilio*. A recent study predicted that *S. noctilio* could spread further in the countries where it already occurs and become established in new areas such as Central America, Africa and China (Carnegie *et al.* 2006). This threat is further emphasized in the context of climate change that could trigger extreme drought, a condition known to favour attacks of *S. noctilio* on pines in plantations and forests.

New insights into *S. noctilio* invasion history revealed in this study arose from the application of microsatellite markers combined with Bayesian clustering and scenario testing. The results highlight the fact that exploratory methods such as Bayesian clustering (and FCA) represent an important first step that makes it possible to interrogate and gain a broad view of a large data set. This view can then be used

to decrease the number of hypotheses in scenario testing. The data set complied in this study also exemplified the fact that microsatellite markers remain the marker of choice when considering recent events such as those related to human-mediated invasions. This is in comparison to the low resolution provided by the mitochondrial COI gene, which was also analysed here. The insights provided by the new methodologies, particularly those relating to the demographic history of invasive and re-introduced species and the microevolutionary processes associated with invasion, will facilitate an understanding of the processes that underpin successful establishment and invasion in new areas.

Acknowledgements

This study was supported by a University of Pretoria postdoctoral fellowship to EB. The project was funded by members of the Tree Protection Co-operative Programme (TPCP) and the THRIP initiative of the Department of Trade and Industry, South Africa. We acknowledge Prof. AV. Selikhovkin, St. Petersburg Forest Technical University for the Russian samples; T. Noblecourt, Office National des Forets for the French samples; and S. Blank, Senckenberg Deutsches Entomologisches Institut Müncheberg for the German samples. Finally, we are grateful to TB Hoareau for his critical reading of an earlier version of the manuscript and helpful discussions.

References

Arnaud-Haond S, Belkhir K (2007) GENCLONE: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. *Molecular Ecology Notes*, **7**, 15–17.

Ascunce MS, Yang CC, Oakey J *et al*. (2011) Global invasion history of the fire ant *Solenopsis invicta*. *Science*, **331**, 1066–1068.

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

Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Bedding RA (1984) Nematode parasites of Hymenoptera. In: *Plant and Insect Nematodes*, (ed. Nickle WR), pp. 755–795, Marcel Dekker Inc., New York, NY, USA.

Beèche M, Lanfranco D, Zapata M, Ruiz C (2012) Surveillance and control of the *Sirex* woodwasp: the Chilean experience. In: *The Sirex Woodwasp and its Fungal Symbiont: Research and Management of a Worldwide Invasive Pest* (eds Slippers B, de Groot P, Wingfield MJ), pp. 51–64. Springer, Dordrecht.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, Logiciel Sous Windows TM Pour la Génétique des Populations. Laboratoire Génome, Populations, Interactions. CNRS UMR 5000, Université de Montpellier II, Montpellier, France.

Bergeron MJ, Hamelin RC, Leal I, Davis C, de Groot P (2008) First report of *Amylostereum areolatum*, the fungal symbiont of *Sirex noctilio*, on *Pinus* spp. in Canada. *Plant Disease*, **92**, 1138.

Bergeron MJ, Leal I, Foord B *et al*. (2011) Putative origin of clonal lineages of *Amylostereum areolatum*, the fungal symbiont associated with *Sirex noctilio*, retrieved from *Pinus sylvestris*, in eastern Canada. *Fungal Biology*, **115**, 750–758.

Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, **19**, 2609–2625.

Campana MG, Hunt HV, Jones H, White J (2011) CorrSieve: software for summarizing and evaluating Structure output. *Molecular Ecology Resources*, **11**, 349–352.

Carnegie AJ, Eldridge RH, Waterson DG (2005) History and management of *Sirex* wood wasp in pine plantations in New South Wales, Australia. *New Zealand Journal of Forestry Science*, **35**, 3–24.

Carnegie AJ, Matsuki M, Haugen DA *et al*. (2006) Predicting the potential distribution of *Sirex noctilio* (Hymenoptera: Siricidae), a significant exotic pest of *Pinus* plantations. *Annals of Forest Science*, **63**, 119–128.

Clavero M, Garcia-Berthou E (2005) Invasive species are a leading cause of animal extinctions. *Trends in Ecology & Evolution*, **20**, 110.

Corley JC, Villacide JM (2012) Population dynamics of *Sirex noctilio*: Influence of diapause, spatial aggregation and flight potential on outbreaks and spread. In: *The Sirex Woodwasp and its Fungal Symbiont: Research and Management of a Worldwide Invasive Pest* (eds Slippers B, de Groot P, Wingfield MJ), pp. 51–64, Springer, Dordrecht.

Cornuet JM, Santos F, Beaumont MA *et al*. (2008) Inferring population history with DIYABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*, **24**, 2713–2719.

Cornuet JM, Ravigne V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, **11**, 11.

Csillery K, Blum MGB, Gaggiotti OE, Francois O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**, 410–418.

Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, **17**, 431–449.

Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology*, **19**, 4113–4130.

Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**, 1591–1604.

Estoup A, Lombaert E, Marin JM *et al*. (2012) Estimation of demo-genetic model probabilities with approximate Bayesian Computation using linear discriminant analysis on summary statistics. *Molecular Ecology Resources*, **12**, 846–855.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Everett RA (2000) Patterns and pathways of biological invasions. *Trends in Ecology & Evolution*, **15**, 177–178.

Facon B, Jarne P, Pointier JP, David P (2005) Hybridization and invasiveness in the freshwater snail *Melanoides tuberculata*: hybrid vigour is more important than increase in genetic variance. *Journal of Evolutionary Biology*, **18**, 524–535.

Facon B, Pointier JP, Jarne P, Sarda V, David P (2008) High genetic variance in life-history strategies within invasive

populations by way of multiple introductions. *Current Biology*, **18**, 363–367.

Facon B, Crespin L, Loiseau A, Lombaert E, Magro A, Estoup A (2011a) Can things get worse when an invasive species hybridizes? The harlequin ladybird Harmonia axyridis in France as a case study. *Evolutionary Applications*, **4**, 71–88.

Facon B, Hufbauer RA, Tayeh A *et al.* (2011b) Inbreeding depression Is purged in the invasive insect *Harmonia axyridis*. *Current Biology*, **21**, 424–427.

Frankham R, Lees K, Montgomery ME, England PR, Lowe EH, Briscoe DA (1999) Do population size bottlenecks reduce evolutionary potential? *Animal Conservation*, **2**, 255–260.

Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, **10**, 305–318.

Guillemaud T, Beaumont MA, Ciosi M, Cornuet JM, Estoup A (2010) Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity*, **104**, 88–99.

Hall MJ (1968) A survey of siricid attack on radiata pine in Europe. *Australian Forestry*, **32**, 5–162.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hoareau TB, Boissin E (2010) Design of phylum-specific hybrid primers for DNA barcoding: addressing the need for efficient COI amplification in the Echinodermata. *Molecular Ecology Resources*, **10**, 960–967.

Hoebeke ER, Haugen DA, Haack R (2005) *Sirex noctilio*: discovery of a Palearctic siricid wasp in New York. *Newsletter of the Michigan Entomological Society*, **50**, 24–25.

Hoos PM, Whitman Miller A, Ruiz GM, Vrijenhoek RC, Geller JB (2010) Genetic and historical evidence disagree on likely sources of the Atlantic amethyst gem clam *Gemma gemma* (Totten, 1834) in California. *Diversity and Distributions*, **16**, 582–592.

Hulme PE (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, **46**, 10–18.

Hurley BP, Slippers B, Wingfield MJ (2007) A comparison of control results for the alien invasive woodwasp, *Sirex noctilio*, in the southern hemisphere. *Agricultural and Forest Entomology*, **9**, 159–171.

Iede ET, Penteado SR, Bisol JL (1988) Primeiro registro de ataque de *Sirex noctilio* em *Pinus taeda* no Brasil. *EMBRAPA-CNPF, Circular tecnica*, **20**, 1–12.

Keane RM, Crawley MJ (2002) Exotic plant invasions and the enemy release hypothesis. *Trends in Ecology & Evolution*, **17**, 164–170.

Klasmer P, Botto E (2012) The Ecology and Biological Control of the Woodwasp *Sirex noctilio* in Patagonia, Argentina. In: *The Sirex Woodwasp and its Fungal Symbiont: Research and Management of a Worldwide Invasive Pest* (eds Slippers B, de Groot P, Wingfield MJ), pp. 203–215. Springer, Dordrecht.

Klasmer P, Fritz G, Corley J, Botto E (1998) Current status of research on *Sirex noctilio* F. in the Andean-Patagonian region in Argentina. In: *Conference Proceedings: Training in the control of Sirex noctilio by Use of Natural Enemies, USDA Forest Service (FHTET 98-13)* (eds Iede E, Shaitza E, Penteado S, Reardon R, Murphy T), pp. 89–90. Colombo, Brazil.

Kolbe JJ, Glor RE, Schettino LRG, Lara AC, Larson A, Losos JB (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 177–181.

Lawson Handley LJ, Estoup A, Evans DM *et al.* (2011) Ecological genetics of invasive alien species. *BioControl*, **56**, 409–428.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Lombaert E, Guillemaud T, Cornuet JM, Malausa T, Facon B, Estoup A (2010) Bridgehead effect in the worldwide invasion of the biocontrol Harlequin ladybird. *PLoS One*, **5**, 7.

Lye GC, Lepais O, Goulson D (2011) Reconstructing demographic events from population genetic data: the introduction of bumblebees to New Zealand. *Molecular Ecology*, **20**, 2888–2900.

Maderni JFP (1998) *Sirex noctilio* F.: present status in Uruguay. In: *Conference Proceedings: Training in the control of Sirex noctilio by Use of Natural Enemies* (eds Iede E, Shaitza E, Penteado S, Reardon R, Murphy T), pp. 81–82. USDA Forest Service (FHTET 98-13). Colombo, Brazil.

Miller D, Clark AF (1935) *Sirex noctilio* (Hym.) and its parasite in New Zealand. *Bulletin of Entomological Research*, **26**, 149–154.

Miller N, Estoup A, Toepfer S *et al.* (2005) Multiple transatlantic introductions of the western corn rootworm. *Science*, **310**, 992.

Miller KA, Nelson NJ, Smith HG, Moore JA (2009) How do reproductive skew and founder group size affect genetic diversity in reintroduced populations? *Molecular Ecology*, **18**, 3792–3802.

Mlonyeni XO, Wingfield BD, Wingfield MJ *et al.* (2011) Extreme homozygosity in Southern Hemisphere populations of *Deladenus siricidicola*, a biological control agent of *Sirex noctilio*. *Biological Control*, **59**, 348–353.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1–10.

Neumann FG, Morey JL, McKimm RJ (1987) The *Sirex* wasp in Victoria. Department of Conservation, Forests and Lands, Victoria, Bulletin No. 29. 41p.

Nielsen C, Williams DW, Hajek AE (2009) Putative source of the invasive *Sirex noctilio* fungal symbiont, *Amylostereum areolatum*, in the eastern United States and its association with native siricid woodwasps. *Mycological Research*, **113**, 1242–1253.

Pimentel D, McNair S, Janecka J *et al.* (2001) Economic and environmental threats of alien plant, animal, and microbe invasions. *Agriculture Ecosystems & Environment*, **84**, 1–20.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Rebuffo S (1988) La avispa de la madera *Sirex noctilio* F. en el Uruguay. Ministerio de Ganadería Agricultura y Pesca, Montevideo.

Richardson DM (2010) Fifty Years of Invasion Ecology: The Legacy of Charles Elton. Wiley Blackwell, Oxford, 456 pp.

Roman J, Darling JA (2007) Paradox lost: genetic diversity and the success of aquatic invasions. *Trends in Ecology & Evolution*, **22**, 454–464.

SAG (2001) Informativo Fitosanitario. Vigilancia Fitosanitaria. Departamento de Protección Agrícola. No 06.

Santana QC, Coetzee MPA, Steenkamp ET *et al.* (2009) Microsatellite discovery by deep sequencing of enriched genomic libraries. *BioTechniques*, **46**, 217–223.

Schonrogge K, Moriya S, Melika G *et al.* (2006) Early parasitoid recruitment in invading cynipid galls. In: *Galling Arthropods and Their Associates: Ecology and Evolution*, pp. 91–101. Springer-Verlag Tokyo, Tokyo.

Simberloff D (2000) Nonindigenous species: a global threat to biodiversity and stability. In: *Nature and Human Society: The Quest for a Sustainable World* (ed. Raven PWT), pp. 325–336. National Academy Press, Washington, District of Columbia.

Simberloff D (2009) The role of propagule pressure in biological invasions. In: *Annual Review of Ecology Evolution and Systematics*, pp. 81–102, Annual Reviews, Palo Alto.

Slippers B, Wingfield MJ, Coutinho TA, Wingfield BD (2001) Population structure and possible origin of *Amylostereum areolatum* in South Africa. *Plant Pathology*, **50**, 206–210.

Slippers B, Wingfield BD, Coutinho TA, Wingfield MJ (2002) DNA sequence and RFLP data reflect geographical spread and relationships of *Amylostereum areolatum* and its insect vectors. *Molecular Ecology*, **11**, 1845–1854.

Slippers B, Coutinho TA, Wingfield BD, Wingfield MJ (2003) A review of the genus *Amylostereum* and its association with woodwasps. *South African Journal of Science*, **99**, 70–74.

Slippers B, Hurley BP, Mlonyeni XO, de Groot P, Wingfield MJ (2012) Factors affecting the efficacy of *Deladenus siricidicola* in biological control systems. In: *The Sirex Woodwasp and its Fungal Symbiont: Research and Management of a Worldwide Invasive Pest* (eds Slippers B, de Groot P, Wingfield MJ), pp. 119–134, Springer, Dordrecht.

Spradbery JP, Kirk AA (1978) Aspects of ecology of siricid woodwasps (Hymenoptera, Siricidae) in Europe, North Africa and Turkey with special reference to biological control of *Sirex noctilio* in Australia. *Bulletin of Entomological Research*, **68**, 341–359.

Tribe GD (1995) The woodwasp *Sirex noctilio* Fabricius (Hymenoptera: Siricidae), a pest of *Pinus* species, now established in South Africa. *African Entomology*, **3**, 216–217.

Wermelinger B, Thomsen IM (2012) The woodwasp *Sirex noctilio* and its associated fungus *Amylostereum areolatum* in Europe. In: *The Sirex Woodwasp and its Fungal Symbiont: Research and Management of a Worldwide Invasive Pest* (eds Slippers B, de Groot P, Wingfield MJ), pp. 51–64, Springer, Dordrecht.

Wermelinger B, Rigling A, Schneider Mathis D, Dobbertin M (2008) Assessing the role of bark- and wood-boring insects in the decline of Scots pine (*Pinus sylvestris*) in the Swiss Rhone valley. *Ecological Entomology*, **33**, 239–249.

Werren JH (1993) The evolution of inbreeding in a haplodiploid organism. In: *The Natural History of Inbreeding and Outbreeding* (ed. Thornhill NW), pp. 42–94, University of Chicago Press, Chicago.

Zepeda-Paulo FA, Simon JC, Ramirez CC *et al.* (2010) The invasion route for an insect pest species: the tobacco aphid in the New World. *Molecular Ecology*, **19**, 4738–4752.

E.B. is interested in factors (natural or anthropogenic) shaping biodiversity patterns. B.H. interests' are in integrated management strategies for the control of forest insect pests, with a particular focus on biological control. M.J.W. research concentrates on the movement of pests and pathogens worldwide and on the bark-beetle fungus symbioses. R.V. is a forest pathologist also interested in ecology and population biology of wood-inhabiting fungi, and fungal–insect interactions. J.S. leads a research group with interests that cover various aspects of fungal biology spanning from ecology to evolution and genomics, including fungal–insect interactions. C.D. research interests' are forest health with an emphasis on forest pathology. P.G. was a forest entomologist and key collaborator, but sadly died before this study was completed. R.A. does research on the control of pathogens and pests in commercial forestry plantations. AJ.C. research interests include optimizing forest health surveillance techniques, eucalypt foliar pathogens, impact of eucalyptus/guava rust and risk mapping of key forest pests. A.G. has broad research interests in forest entomology, including systematics, biology and control. P.K. does research in Forest Entomology, mainly on ecology of forest pests, their monitoring and their biological control. B.W. is interested in the diversity of mainly bark and wood boring forest insects and their interactions with host plants, natural enemies and forest management. B.S. research focuses on the molecular ecology and evolution of fungal pathogens and insect pests of trees.

Data accessibility

COI sequences have been deposited in GenBank under the accession numbers: JX456374–JX456387.

The microsatellite data set and mtDNA alignment files have been deposited in DRYAD under the accession number: doi:10.5061/dryad.37mm8.

Supporting information

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

Table S1 List of the 13 mitochondrial genomes used to design the Hymenoptera COI primers.

Table S2 Allele frequencies of each female population, FIS estimates and occurrence of private alleles.

Table S3 Confidence indices in the demographic estimates and posterior probabilities and confidence intervals of scenarios.

Fig. S1 Distinct scenarios tested using the software DIYABC. Schematic representation of the sequence of introduction between continents and countries, their implementation in a coalescent framework using DIYABC and text explaining the differences between the scenarios.

Fig. S2 Structure plots for the female and male data sets considered separately and showing the partitioning for $K = 2$ to $K = 4$; Evanno's graphs representing the most likely number of clusters (K). They show the likelihood ($L'(K)$, on the left Y-axis) associated with each K (on the X-axis,) and the related ΔK values (on the right Y-axis).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.