MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 2901–2915 doi: 10.1111/j.1365-294X.2011.05151.x

Recent vicariance and the origin of the rare, edaphically specialized Sandhills lily, *Lilium pyrophilum* (Liliaceae): evidence from phylogenetic and coalescent analyses

NORMAN A. DOUGLAS,* WADE A. WALL,* QIU-YUN (JENNY) XIANG,* WILLIAM A. HOFFMANN,* THOMAS R. WENTWORTH,* JANET B. GRAY† and MATTHEW G. HOHMANN‡

**Department of Plant Biology, PO Box 7612, North Carolina State University, Raleigh, NC 27695, USA, †Directorate of Public Works, Endangered Species Branch, United States Army, Fort Bragg, NC 28310, USA, ‡US Army Corps of Engineers, Engineer Research and Development Center, PO Box 9005, Champaign, IL 618262, USA*

Abstract

Establishing the phylogenetic and demographic history of rare plants improves our understanding of mechanisms that have led to their origin and can lead to valuable insights that inform conservation decisions. The Atlantic coastal plain of eastern North America harbours many rare and endemic species, yet their evolution is poorly understood. We investigate the rare Sandhills lily (*Lilium pyrophilum*), which is endemic to seepage slopes in a restricted area of the Atlantic coastal plain of eastern North America. Using phylogenetic evidence from chloroplast, nuclear internal transcribed spacer and two low-copy nuclear genes, we establish a close relationship between *L. pyrophilum* and the widespread Turk's cap lily, *L. superbum*. Isolation-with-migration and coalescent simulation analyses suggest that (i) the divergence between these two species falls in the late Pleistocene or Holocene and almost certainly post-dates the establishment of the edaphic conditions to which *L. pyrophilum* is presently restricted, (ii) vicariance is responsible for the present range disjunction between the two species, and that subsequent gene flow has been asymmetrical and (iii) *L. pyrophilum* harbours substantial genetic diversity in spite of its present rarity. This system provides an example of the role of edaphic specialization and climate change in promoting diversification in the Atlantic coastal plain.

Keywords: coalescence, divergence, edaphic, *Lilium*, Pleistocene, rarity

Received 18 November 2010; revision received 28 April 2011; accepted 5 May 2011

Introduction

Molecular studies of rare plant taxa usually aim to quantify the level and patterns of genetic diversity in a particular species (Karron 1987; Hamrick & Godt 1990; Ellstrand & Elam 1993; Gitzendanner & Soltis 2000). Phylogeographic studies, on the other hand, often focus on widespread species and try to discern continental-scale patterns (Taberlet *et al.* 1998; Brunsfeld *et al.* 2001; Soltis *et al.* 2006). However, the tools of phylogeography, particularly coalescent-based analyses that provide information about the age and historical demography of species (Knowles 2009), have only rarely been applied to investigate the history of rare species (Raduski *et al.* 2010; Whittall *et al.* 2010).

Of the 'seven forms of rarity' (Rabinowitz 1981), the most extreme describes taxa that have a narrow geographic range, require specific habitats and maintain only small local populations. Many edaphic endemics (plants restricted to soils with unusual physical or chemical properties) belong to this category. While the textbook examples of edaphic endemic plants are restricted to serpentine, various substrates support edaphic endemics, including guano, alkali, salt, and gypsum deposits, limestone, chalk, and granite outcrops, oligotrophic bogs and deep porous sands (Ornduff 1965; Axelrod 1972; Parsons 1976; Kruckeberg &

Correspondence: Norman A. Douglas, Fax: (919) 515 3436; E-mail: norman_douglas@ncsu.edu

Rabinowitz 1985; Kruckeberg 1986; Williamson & Bazeer 1997). Many aspects of the origin of edaphic endemic species are poorly understood (Rajakaruna 2004). For instance, such species often occur in close geographic proximity to their progenitor lineages (e.g. Baldwin 2005), yet it is not usually known whether or how strongly gene flow is interrupted. While taxa displaying edaphic endemic syndromes often show reduced genetic diversity compared with their close relatives (Godt & Hamrick 1993; Baskauf *et al.* 1994; Ayres & Ryan 1999; but see Raduski *et al.* 2010), this may reflect genetic drift due to lower population sizes or the effects of selection. Strong selection imposed by edaphically challenging soils could be sufficient to foster population divergence (Nosil *et al.* 2009; Freeland *et al.* 2010). Some edaphic endemics may represent vicariant populations isolated in narrow parts of formerly wider ranges and niches of their progenitors (e.g. Crawford *et al.* 1985), which may themselves be able to grow on the unusual substrate without being restricted to it.

Edaphic specialists (especially in bog and sand habitats, Sorrie & Weakley 2001) are an important component of the endemic-rich flora of the coastal plain of eastern North America. Few coastal plain endemics have been the subject of molecular analyses. Sand dune habitats in Florida apparently served as Pleistocene refugia for the genera *Dicerandra* and *Conradina* (Edwards *et al.* 2006; Oliveira *et al.* 2007), and in general, Florida has been proposed as a major Pleistocene refugium for many taxa in eastern North America (Soltis *et al.* 2006). Yet, recent phylogeographic work indicates that some coastal plain endemic species likely persisted in northerly latitudes throughout the Pleistocene. For instance, the Atlantic coastal plain endemic Pyxie Moss, *Pyxidanthera* (Diapensiaceae), shows long-term range stasis (Wall *et al.* 2010).

The Fall-Line Sandhills of North and South Carolina (which occur at the western boundary of the coastal plain) provide one of the clearest examples of the edaphic contribution to the botanical diversity of the Atlantic coastal plain. This region is comprised of rolling hills of open, fire-maintained longleaf pine (*Pinus palustris*) savanna dissected by numerous blackwater streams and wetlands, providing a diverse matrix of habitats that support at least eight endemic plants (and numerous near-endemics, Sorrie & Weakley 2001). In the core of the Sandhills region in southern North Carolina, the uppermost deposit is the Pinehurst formation, which is characterized by loose coarse-grained sands found along ridgetops. This formation was deposited in a tidal environment (J. Nickerson, North Carolina Geological Survey, personal communication) and may date to the Eocene (Cabe *et al.* 1992). Below the Pinehurst formation (and exposed along drainages and slopes throughout the region) lies the Cretaceous Middendorf formation, which is of deltaic origin and thus has more abundant clays (Sohl & Owens 1991). At the interface between these (and similar formations in the Carolinas and southeastern Virginia) occur Sandhills seep and streamhead pocosin ecotone communities. When kept open by frequent fires encroaching from the surrounding xeric pine savannas, these wetlands can support extremely high local species richness, among the highest values ever recorded in North America (>102 species per 1/100 ha, Schafale & Weakley 1990). The age of the formations implies that endemic species have potentially had a very long time to adapt to the unusual edaphic conditions.

In this study, we consider the Sandhills lily, *Lilium pyrophilum* (Liliaceae), a striking endemic of the Sandhills in the Carolinas and southeastern Virginia. Formally described only recently (Skinner & Sorrie 2002), specimens of this species were previously identified in herbaria as any of three similar species in the region (*L. superbum*, *L. michauxii* or *L. iridollae*) that share the distinctive 'Turk's cap' morphology, in which flowers are pendent with the tepals reflexed upward. Skinner & Sorrie (2002) identified three specific plant communities (Schafale & Weakley 1990; Sorrie *et al.* 2006) that support *L. pyrophilum*, including Sandhills seep and streamhead pocosin ecotones. The third, small stream swamps are affected by frequent flooding events in addition to seepage and rarely support *L. pyrophilum* (Sorrie *et al.* 2006).

Lilium pyrophilum is a very rare species. There are fewer than 75 historical and extant locations in North and South Carolina, and Virginia (North Carolina Natural Heritage Program 2007), and between 2007 and 2009, a survey of all known populations located <500 stems across 35 populations (W. Wall, unpublished data). Approximately half of the extant populations and a quarter of the individuals occur on Fort Bragg Military Reservation in North Carolina, where prescribed and ordnance-ignited fires maintain appropriate habitat.

In describing *L. pyrophilum* (Skinner & Sorrie 2002), the authors outlined three phylogenetic hypotheses concerning the origin of the species. First, they speculated that *L. pyrophilum* may represent a peripheral isolate of the Turk's cap lily, *L. superbum*, which it most resembles morphologically (albeit with significant differences, Skinner & Sorrie 2002). *Lilium superbum* is distributed throughout much of eastern North America (Fig. 1), and in contrast to the edaphically specialized *L. pyrophilum*, it is a generalist, occurring in rich woods and oligotrophic wetlands from high elevation to sea level. Especially in northern parts of its range (e.g. the Pine Barrens of New Jersey), it can be found in saturated

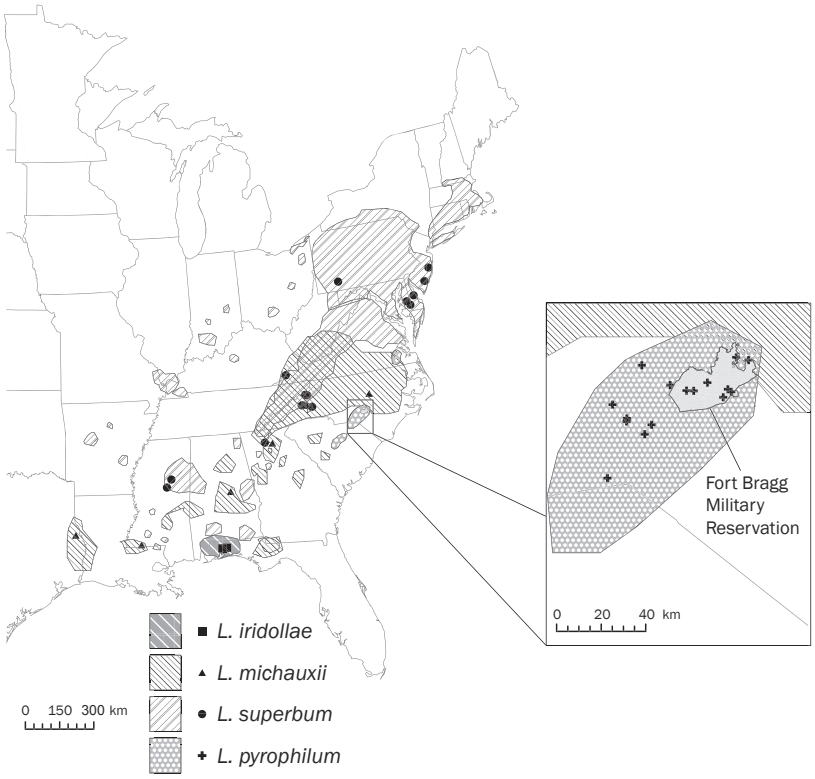

Fig. 1 Distribution of populations included in this study and geographic ranges of the four focal species.

sandy habitats not unlike those preferred by *L. pyrophilum*, but it is not restricted to them. However, it is essentially absent from the Piedmont and Atlantic coastal plain from the Carolinas southward. Thus, it is disjunct from *L. pyrophilum* by at least 150 km everywhere except in southeastern Virginia (Fig. 1) where the coastal plain narrows.

Second, they speculated that *L. pyrophilum* may represent a hybrid species, with the widespread Carolina lily (*L. michauxii*) and *L. superbum* as progenitors. Homoploid hybrid speciation has been implicated in the origin of other edaphic specialists, e.g. *Helianthus paradoxus* (Rieseberg *et al.* 1990) and Hawaiian *Scaevola* (Howarth & Baum 2005). Of the three potentially related species, *L. pyrophilum* resembles *L. michauxii* least, differing in leaf shape and producing fragrant flowers (Skinner 2002). While the range of *L. michauxii* does overlap the range of *L. pyrophilum* (Fig. 1), they occur in contrasting habitats, with *L. michauxii* favouring much drier sites. Notably, *L. michauxii* and *L. superbum* co-occur throughout much of their ranges (Fig. 1), yet natural hybrids are apparently rare (Skinner 2002).

Finally, Skinner and Sorrie suggested the possibility that *L. pyrophilum* may represent a disjunct population of the Pot-o'-gold or Panhandle lily (*L. iridollae*), a narrow endemic of wet pine savannas in northwestern Florida (where it is listed as endangered) and adjacent Alabama. This hypothesis emphasizes similar habitat requirements of the two species, but downplays consistent morphological differences (e.g. details of rhizome structure, Skinner 2002; Skinner & Sorrie 2002) and a range separation of over 700 km (Fig. 1).

In this study, we report the results of a molecular study focused on *L. pyrophilum* and its close relatives. First, we investigated the phylogeny of the eastern pendent species of *Lilium* to address whether *L. pyrophilum* represents a peripheral isolate of *L. superbum*, a hybrid between *L. superbum* and *L. michauxii*, or a disjunct population of *L. iridollae*. Second, we analysed the distribution of genetic variation within and among the taxa thought to be closely related to *L. pyrophilum* and used coalescent-based methods to explicitly evaluate the possible timing of the divergence of *L. pyrophilum*. Our results are interpreted in the context of the evolution of rare, edaphically specialized lineages in the Atlantic coastal plain.

Materials and methods

Sampling and molecular data

Samples were obtained from 50 populations spanning the geographic range of each of the four focal species (Fig. 1). We also sampled two populations of *Lilium*

canadense, another pendant species that lacks the Turk's cap morphology. Sampling information is provided in Table S1 (Supporting Information). Populations were located in the field based on documented occurrences from herbarium specimens, element occurrence records from state Natural Heritage Programs and communication with local botanists. We endeavoured to sample a similar number of populations of *L. superbum* and *L. michauxii* spanning the geographic range of each species. Our sampling of the rare *L. iridollae* was limited to two populations. In general, one individual was taken to represent each population. Genomic DNA was isolated from fresh or frozen leaves, using the CTAB method (Doyle & Doyle 1987). Nuclear ribosomal internal transcribed spacer ('ITS') sequences were obtained with primers ITS4 and ITS5a (White *et al.* 1990; Stanford *et al.* 2000). This locus was sequenced to facilitate comparison with abundant existing data available in GenBank to determine whether the species in this study form a monophyletic group. We screened eight chloroplast markers from Shaw *et al.* (2007); of these, three (the *atpI-atpH*, *psbD-trnT* and *rpl32-trnL* intergenic spacers) consistently amplified and contained variable sites. As the chloroplast behaves as a single nonrecombining locus, sequences of these three regions were concatenated, and this marker is hereafter referred to as 'CP'.

We developed single-copy nuclear markers for *Lilium*. In general, we screened EST or complete CDS sequences from *Lilium* against the *Oryza sativa* genomic sequence at GenBank using SPIDEY (Wheelan *et al.* 2001) with the 'divergent sequences' and 'use large intron sizes' options. Candidate sequences were downloaded and manually aligned in Se-Al (Rambaut 1996) using amino acid translations. Homologous sequences from GenBank were incorporated into the alignments. When we were confident of the positions of the introns in the rice genome, we then designed primers using Primer3 (Rozen & Skaletsky 2000), which were screened against DNA extracted from *L. longiflorum* and an Asiatic hybrid cultivar (which served as positive controls because nearly all of our candidate regions were based on sequences from these cultivated lilies) and the four taxa in our study. We were able to obtain single amplicons for relatively few of these regions even after extensive PCR optimization; it was often the case that primers would amplify nontarget regions or that introns would be small, invariant or missing entirely. The closely related *L. canadense* has a phenomenally large genome (1C = 47.90 pg, 46.9 Gbp; Zonneveld *et al.* 2005; Peruzzi *et al.* 2009), which may have contributed to the difficulty we encountered in obtaining single-copy nuclear sequences. However, we were able to design primers that amplified two novel regions. The first includes two introns between exons 8 and 10 of the *L. longiflorum* alkaline phytase gene, *LlAlp* ('AP', primers: AP8f, 5'-TCTCCTTGGGCTCTTTCTTG and AP10r, 5'-GAAAACCTCAAATGGGCAGAG), which is involved in phytic acid metabolism (Mehta *et al.* 2006). While GenBank contains sequences for two isoforms of this gene, our PCR experiments are consistent with these representing splice variants of a single locus. The second region corresponds to a region between exons 5 and 10 of the *AKT1*-like potassium channel *LilKT1* ('AKT', primers: AKT5f, 5'-AGAGACTCTTGATGCACT TCCTAAA and AKT10r, 5'-AAGAGAACAACA-CAACTTTCATTCC). This locus was more difficult to amplify, and we were unable to generate sequences for *L. iridollae*. Primers and PCR conditions for ITS and the chloroplast loci followed White *et al.* (1990) and Shaw *et al.* (2007). For AP and AKT, PCR contained 2.5 μL 10× PCR buffer, 1% BSA, 200 μM dNTPs, 2.5 mM $MgCl_2$, 4 μM of each primer and 0.5 U Taq DNA polymerase. Cycling conditions were 95 °C for 4 min, followed by 35 cycles of 95 °C for 30 s, 58 °C for 30 s, 72 °C for 2.5 min, and a final extension step of 72 °C for 4 min. Amplicons were cleaned with Antarctic Phosphatase and Exonuclease I (New England Biolabs, Ipswich, MA, USA). Sequencing was performed on an Applied Biosystems 3730 capillary sequencer (Foster City, CA, USA) using Big Dye chemistry. Chromatograms were edited in Sequencher 4.1.2 (Gene Codes Corporation, Ann Arbor, MI, USA). Heterozygous bases were easily identified in the chromatograms for the three nuclear regions and coded with standard IUPAC notation. Because of the low levels of divergence among our sequences, alignment was trivial and performed manually in Se-Al. The most likely haplotypic phases of AP and AKT genotype sequences were ascertained with a combination of cloning and the program PHASE 2.1 (Stephens *et al.* 2001; Stephens & Donnelly 2003) called by the 'Open/Unphase genotype' option in DnaSP v. 5 (Librado & Rozas 2009); the inferred alleles form the basis for all further analyses involving these loci. The preferred model of sequence evolution for each locus (ITS: TIM3ef + I + G; CP: K81uf + I; AP: TVM + I; AKT: TVM + I + G) was determined according to Akaike Information Criterion (AIC) in jModelTest (Posada 2008). Sampling details, genotype information and GenBank accession numbers are provided in Tables S1 and S2 (Supporting Information).

Phylogenetic analyses and descriptive population genetics

For the ITS analysis, 44 new sequences were aligned with 49 from GenBank to create a matrix of 93 sequences. Included were the four species in this study, plus 37 other taxa including the pendent eastern North

American species, *L. michiganense*, *L. canadense* and *L. grayi*, and eight others from *Lilium* section *Pseudolirium*, the monophyletic group of North American species (Nishikawa *et al.* 1999) to which all taxa in this study belong. Unweighted parsimony analysis for the ITS locus was accomplished using PAUP* 4.0b10 (Swofford 2002) using 100 random-addition sequence replicates with TBR branch swapping; owing to overall low sequence divergence, parsimony bootstrapping was conducted with 10^6 'fast' stepwise addition sequences (Soltis & Soltis 2003). Maximum-likelihood (ML) analysis for this locus was conducted in GARLI v. 1.0 (Zwickl 2006). Likelihood bootstrap values were obtained with 1000 replicate searches. The statistical parsimony haplotype network was computed for complete sequences of the three chloroplast regions, *atpI-atpH*, *psbD-trnT* and *rpl32-trnL* (38 sequences), using TCS (Clement *et al.* 2000). The nuclear loci (AP: 82 haplotypes; AKT: 62 haplotypes) have a more complicated evolutionary history than chloroplast sequences; thus, network analyses for the two were conducted using the geodesically pruned quasi-median network algorithm (Ayling & Brown 2008) as implemented in SplitsTree4 (Huson & Bryant 2006), which produces pruned networks that connect all sequences (including multistate characters) by at least one shortest path. ML trees (not shown) were inferred for these sequences as well; they were poorly resolved and showed few supported nodes. However, neither nuclear locus showed phylogenetic evidence of paralogy. For *L. michauxii*, *L. superbum* and *L. pyrophilum*, Arlequin v. 3.5 (Excoffier & Lischer 2010) was used to estimate haplotype richness, number of segregating sites, nucleotide diversity π (Nei 1987) and Watterson's (1975) population mutation parameter θ, for the chloroplast and single-copy nuclear loci.

Testing divergence between L. michauxii, L. pyrophilum *and* L. superbum

As our data include a single individual per 'population', we treated species as the main hierarchical level for the purposes of these analyses. Pairwise F_{ST} (Weir & Cockerham 1984) and the exact test of population differentiation (Raymond & Rousset 1995; Goudet *et al.* 1996) between *L. michauxii*, *L. superbum* and *L. pyrophilum* were calculated in Arlequin v. 3.5 (Excoffier & Lischer 2010), with individuals and species used as the hierarchical groupings. Significance was assessed with 10^3 permutations (F_{ST}) or 2×10^6 Markov chain steps (exact test).

The nature of the divergence between *L. superbum* and *L. pyrophilum* was further investigated using the isolation-with-migration model (Nielsen & Wakeley 2001), implemented in IMa2 (Hey & Nielsen 2007). The full model in the two-population case includes six parameters (divergence time, θ for the ancestral and two descendent populations and migration rates between the descendent populations). This model assumes no recombination within loci and free recombination between loci and that markers are selectively neutral. Thus, several recombination detection methods available in the program RDP3 (beta 40; Martin *et al.* 2005) were used to search for recombinant alleles. As selection or demographic changes can cause departures from neutral expectations, DnaSP v. 5 (Librado & Rozas 2009) was used to perform three different tests of neutrality: Tajima's D (Tajima 1989), Fay and Wu's H (Fay & Wu 2000) and R_2 (Ramos-Onsins & Rozas 2002). Critical values for these statistics were obtained using 10^5 coalescent simulations. The chloroplast data set showed no evidence of recombination; the AP and AKT data sets were filtered with IMgc Online (Woerner *et al.* 2007) to create data sets that were free of detectable recombination and infinite sites violations. Maximum priors for the IMa2 analysis were based on recommended starting values given in the program documentation and refined after preliminary exploratory runs. Priors ultimately selected were population mutation rates (for *L. pyrophilum*, *L. superbum* and ancestral population) θ_0, θ_1 and $\theta_2 = 47$, splitting time parameter $t = 3$ and population migration rate m_1 and $m_2 = 10$. Mutation rate priors (CP: 1.5×10^{-9}, AP & AKT: 6.03×10^{-9}) were specified based on values given by Gaut (1998). Seventy geometrically heated chains (using the heating parameters ha = 0.98, hb = 0.50) were run for 750 000 generations beyond a 150 000 generation burn-in and trees were sampled every 75 generations. This process was repeated 10 times using different random number seeds.

Because results from each replicate were similar, 10^5 trees were concatenated into a single run in load-trees mode and the 'test nested models' option was activated. This option evaluates the likelihood of 24 models simpler than the full isolation-with-migration model by constraining parameters (other than divergence time) and rejecting those that are significantly worse than the full model based on a likelihood ratio test. We also compared models using an information-theoretic method (Carstens *et al.* 2009), which allows the relative performance of nested and non-nested models to be compared using AIC. Compared with a hypothesis-testing approach, which simply identifies models that are rejected as significantly worse than the full model, the information-theoretic approach provides model weights that allow the relative performance of each of a given set of models, including the full model, to be compared directly with others given the data (Burnham &

Anderson 2002). We used the full model posterior probability and the 2(log-likelihood ratio) values, which IMa2 estimates for each model under the assumption that the model's posterior probability is proportional to its likelihood, to calculate the AIC for the full model and each nested model. Subsequently, Akaike weights and evidence ratios were calculated (Burnham & Anderson 2002; Carstens *et al.* 2009).

Conversion of the IMa2 parameter estimates from coalescent to demographic units was accomplished assuming a generation time of 20 years. This is arbitrary but conservative, based on what little is known about the natural history of these species. Germination and establishment is slow, taking two seasons, and plants need 7 years to reach flowering size. Year-to-year survivorship is relatively high (>0.95, Wade Wall, unpublished data). Using the equation $T = \alpha + [s/(1 - s)]$, where T = generation time, α = age of first reproduction and s = adult survivorship (Lande *et al.* 2003), we obtain a value of 26 years. Although estimates of survivorship could be too high, the Lande equation does not account for the fact that older plants are typically larger and more fecund than younger ones. In either case, our generation time should be considered a minimum estimate.

Because isolation is implicit in the isolation-with-migration framework, we tested this assumption with a series of coalescent simulations. Briefly, we estimated N_e for each locus using BEAST (Drummond & Rambaut 2007). Because only *L. pyrophilum* and *L. superbum* sequences were included, simpler ML models were utilized (CP: HKY, AP: TnN + I + G, AKT: K81uf + I). We then used Mesquite v. 2.73 (Maddison & Maddison 2010) to simulate 1000 data sets under each of several simple divergence models (using estimated substitution models for each locus). We treated each species as a population such that *L. superbum* had a N_e 3× that of *L. pyrophilum* (the total N_e corresponding to the value from BEAST). The two populations coalesced at times corresponding to 2.58 Ma (earliest Pleistocene), 126 ka (upper Pleistocene) or 18 ka (last glacial maximum). We then conducted parsimony searches using PAUP* 4.10b (Swofford 2002) on each simulated data set saving 1000 consensus trees. Slatkin and Maddison's s (i.e. the number of parsimony steps implied by a given topology treating source population as a character, Slatkin & Maddison 1989) was computed for each tree to create a null distribution for each locus and divergence time. This was compared with the value of s for the empirical data. When minimum empirical values for s were higher than 95% of the simulated values, we rejected the scenario. To evaluate the effect of the level of migration inferred by IMa2, we duplicated these analyses, but allowing migration. Because Mesquite only allows symmetrical migration, we specified a rate of 9.8×10^{-6} migrants per individual per generation, which corresponds to the estimated value of the parameter under the 'equal migration rate' nested model in IMa2. Finally, following Gugger *et al.* (2010), we evaluated the no-divergence scenario by simulating 1000 data sets per locus under a single population scenario. The resulting parsimony consensus trees were contained within the two-population model described previously, and the null distributions of s were calculated. In this case, the scenario was rejected if the maximum empirical values of s were lower than 95% of the simulated values. As coalescent parameter estimates based on single loci are highly sensitive to stochastic error (Edwards & Beerli 2000), these simulations were conducted for both the upper and lower 90% HPD estimates of N_e from BEAST.

Table 1 Genetic diversity and results of neutrality tests

	Lilium michauxii			*Lilium pyrophilum*			*Lilium superbum*		
Species locus	CP	AKT	AP	CP	AKT	AP	CP	AKT	AP
Individuals (haplotypes)	8 (8)	5 (10)	7 (14)	15 (15)	13 (26)	18 (36)	13 (13)	12 (24)	15 (30)
Aligned length (bp)	2361	1428	453	2360	1428	453	2361	1428	453
Segregating sites	7	10	13	7	24	8	9	30	18
Observed haplotypes	5	7	9	4	16	9	7	17	12
Nucleotide diversity π	0.0010	0.0033	0.0098	0.0008	0.0024	0.0016	0.0009	0.0040	0.0053
Watterson's theta θ	0.0011	0.0025	0.0090	0.0009	0.0044	0.0043	0.0012	0.0061	0.0100
Tajima's D	−0.4150	0.0487	0.3349	−0.4468	−1.7637*	−1.8536**	−1.0835	−1.2142	−1.6319*
Fay and Wu's H	1.7857	0.8000	2.2418	1.3429	−8.8862*	−2.8794*	−1.9615	−4.8333	0.6437
R_2	0.1577	0.2091	0.1597	0.1301	0.0625**	0.0495***	0.1105*	0.0828	0.0692*

Sampling represents the number of individuals and the number of haplotypes (for phased nuclear loci). Significance of neutrality tests was assessed with 10^5 coalescent simulations in DnaSP v. 5.1 (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$).

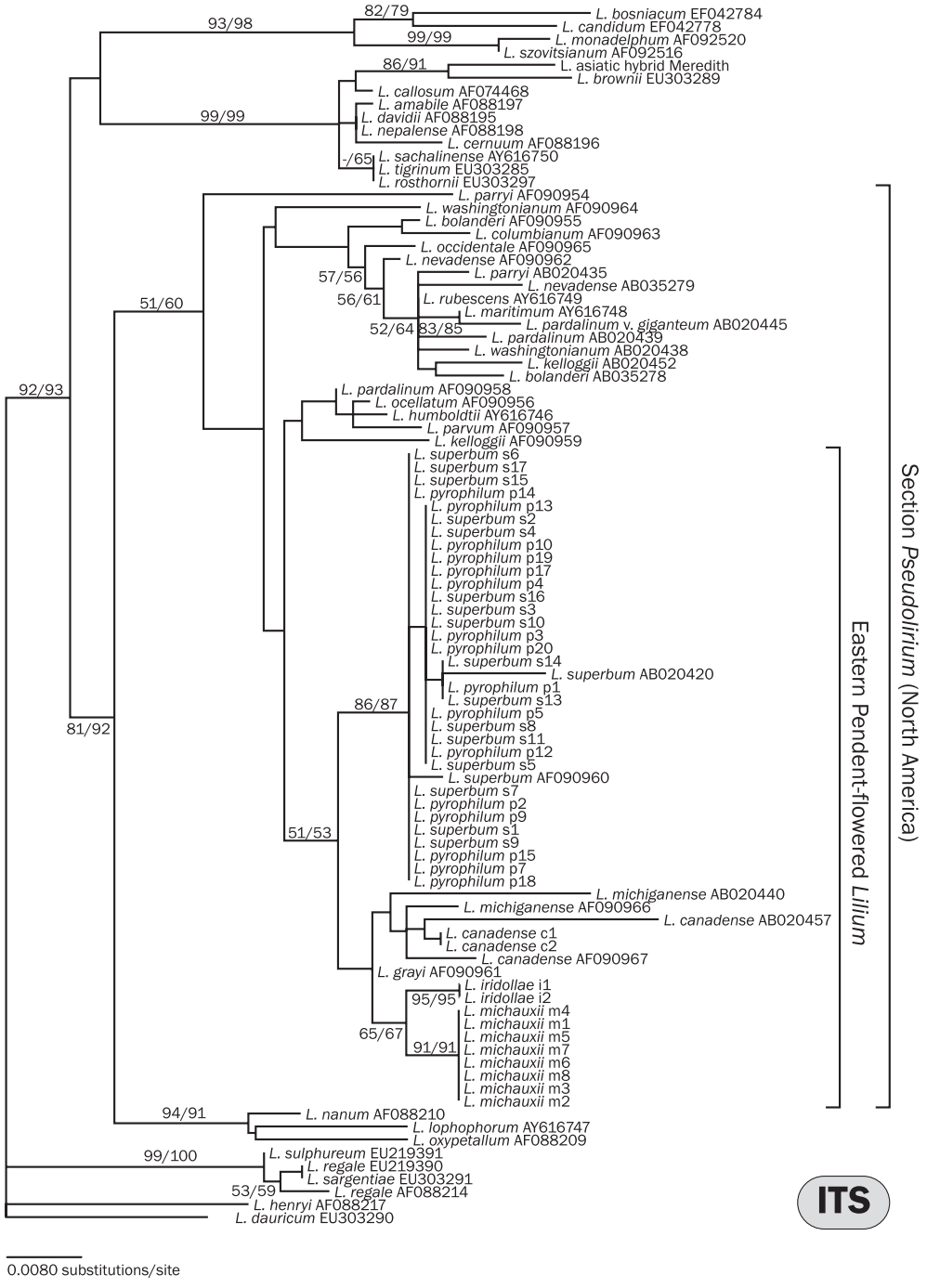

Fig. 2 Maximum-likelihood (ML) Phylogram of internal transcribed spacer sequences. Support values are ML bootstrap/Bayesian posterior probability.

Results

Phylogenetic analyses

In the analysis of ITS data, overall support is quite weak at the level of intra- and interspecific relationships, with no significant (≥70%) bootstrap support for the monophyly of the North American section *Pseudolirium* or the eastern pendent-flowered species (Fig. 2). However, there is a relatively high level of support for the branch uniting two accessions of *Lilium iridollae*, for that uniting the eight samples of *L. michauxii*, and, finally, for the branch leading to the 32 samples of *L. pyrophilum* and *L. superbum*. Little divergence is evident among the

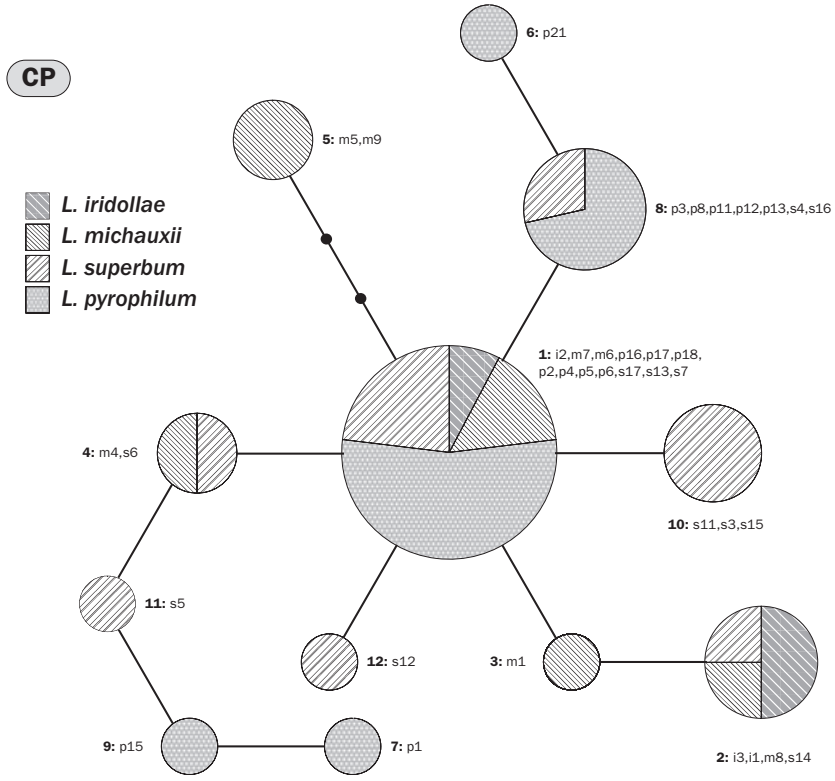

Fig. 3 Chloroplast haplotype network. Statistical parsimony network for CP haplotypes. Chart area reflects the frequency of the haplotype; each slice reflects the frequency at which each haplotype was found in each species. Haplotype numbers (bold) and sample abbreviations correspond to those in Tables S1 and S2 (Supporting Information). Edges represent mutations, black dots unsampled haplotypes.

accessions of each species (with the exception of the GenBank sequences for *L. superbum*, *L. canadense* and *L. michiganense*). The statistical parsimony network (Fig. 3) computed for the chloroplast data revealed a common haplotype (1) that was found in all four species, plus 11 less common types. Overall, four of the six non-singleton haplotypes occur in multiple species. Quasi-median networks produced for the AKT and AP loci (Fig. 4) showed that, while AP haplotype 8 is one mutational step from the nearest *L. michauxii* haplotype (m4a), most *L. michauxii* (and *L. iridollae* in AP) haplotypes are separated from a cloud of *L. pyrophilum* and *L. superbum* haplotypes, which are thoroughly intermixed and frequently shared. No haplotypes were shared between *L. pyrophilum* and *L. michauxii*.

Genetic diversity

Haplotype richness h, segregating sites S, nucleotide diversity π and Watterson's θ are given in Table 1. Nucleotide diversity is relatively low, with values between 0.0008 and 0.00978 substitutions per site, and average values for AP and AKT are nearly five times the value for the chloroplast data set.

Tests of neutrality

Departures from neutrality were detected in the nuclear data sets in *L. pyrophilum* and *L. superbum*, where there were significant negative estimates of Tajima's D and R_2. Fay and Wu's H is significant in *L. pyrophilum* only. Tajima's D is sensitive to both demographic expansion and selection, and R_2 is designed to detect population expansion (Ramos-Onsins & Rozas 2002). While Fay and Wu's H is most sensitive to recent positive selection, it may be sensitive to particular demographic conditions involving structured populations (Fay & Wu 2000). We believe these loci are unlikely to be under positive selection, because there is no obvious reason two loci should deviate from neutrality more strongly in *L. pyrophilum* than in the other two taxa. The chloroplast data also show some demographic expansion in *L. superbum* (weakly significant R_2) without a significantly negative D. Thus, while we cannot eliminate the possibility of some background selection in the nuclear data sets (which does not violate the assumptions of IMa2), it is more likely that demographic factors explain the significant values for these statistics.

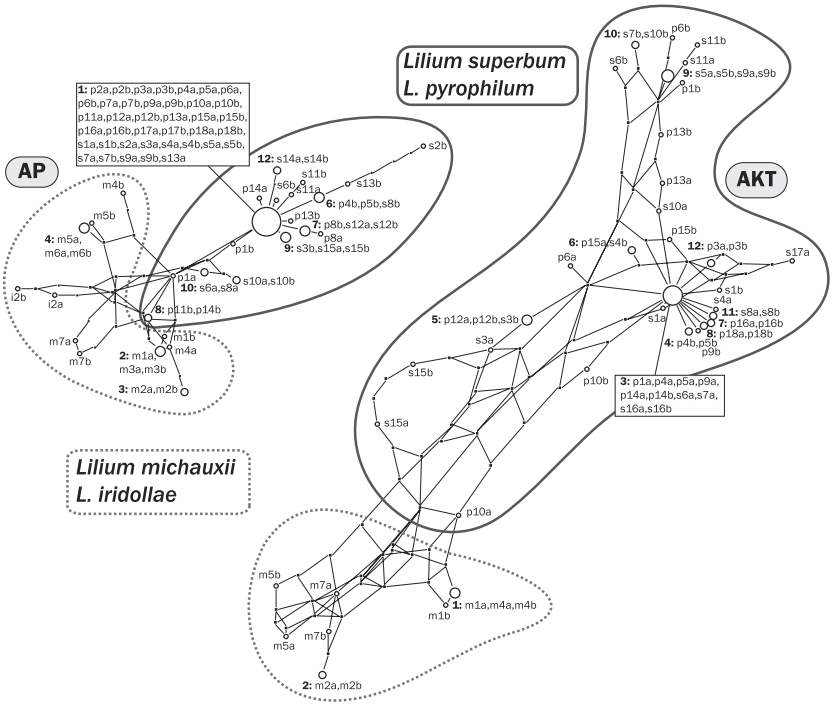

Fig. 4 Quasi-median joining networks for the nuclear loci AP and AKT. Network representations of the relationships between nuclear haplotypes (bold numbers and sample abbreviations correspond to Tables S1 and S2, in Supporting Information). In quasi-median-joining networks, each haplotype is connected to the others by at least one shortest path. Mutational steps are indicated by edges, and black dots represent potential unsampled haplotypes.

Differentiation of L. michauxii

Pairwise F_{ST} values (Table 2) revealed that *L. michauxii* was significantly divergent from *L. pyrophilum* and *L. superbum* for the AKT and AP data sets, whereas differentiation between *L. pyrophilum* and *L. superbum* was minimal and only significant in the AKT data set. No significant differentiation was detected among any of the three species for the CP data set. Conversely, all pairwise exact differentiation tests (Raymond & Rousset 1995) were significant for the two nuclear loci; for the cpDNA, a significant result was only obtained between *L. pyrophilum* and *L. michauxii*.

Divergence between L. pyrophilum *and* L. superbum

Under the isolation-with-migration model, estimates of the mutation parameter theta (θ) were *L. pyrophilum*: 3.736; *L. superbum*: 10.79; and ancestral population: 1.292, corresponding to effective population sizes (95% highest posterior density interval, abbreviated '95% HPD') of 11 400 (2800–29 700), 32 900 (12 800–86 900) and 3900 (0–14 400), respectively (Fig. 5a). The splitting time between *L. pyrophilum* and *L. superbum* was estimated as 0.7725 coalescent units, with the 95% HPD being 0.3435–2.405 (Fig. 5b). This estimate corresponds to a divergence time of 188 ka (95% HPD 84–586 ka) with the assumed mutation rates and generation time. The posterior distribution of splitting time did not reach zero (nor did it for much higher prior values in preliminary runs), so 95% HPD intervals should be interpreted with caution. The coalescent migration rate m from *L. superbum* into *L. pyrophilum* was highest at zero, while the converse was 1.915. Thus, population migration rates (2 NM, Hey & Nielsen 2004) are asymmetrical and quite high from *L. pyrophilum* into *L. superbum* (2 NM = 9.98, Fig. 5c). The model selection procedure (Table 3) preferred a model that holds the two species' population sizes equal and the migration rate from *L. superbum* to *L. pyrophilum* at zero (model weight $w = 0.32$). The next best model ($w = 0.22$) also fixed the *L. superbum* → *L. pyrophilum* migration rate at zero but allowed the population sizes to vary. The full model ($w = 0.19$) had the next highest weight, and the next three models differed in that they fixed the population sizes as above (model 4), held migration rates equal (model 5) and held the *L. pyrophilum* → *L. superbum* migration rate at zero (model 6). The six best models are assigned 95.6% of the total weight. The remaining 19 models had some combination of zero migration, and one or both of the population sizes

Table 2 Pairwise *Fst* and exact test of population differentiation

	Lilium michauxii	*Lilium pyrophilum*	*Lilium superbum*
L. michauxii		0.109/0.393***/0.625***	0.046/0.328***/0.567***
L. pyrophilum	*/***/***		0.007/0.021/0.057*
L. superbum	–/***/**	–/**/*	

Loci: CP/AP/AKT. Above diagonal, pairwise F_{ST}; below diagonal, exact test of differentiation (Goudet *et al*. 1996; Raymond & Rousset 1995). Significance assessed in Arlequin by either 10^3 permutations (F_{ST}) or 2×10^6 Markov chain steps (exact test); *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

equal to the ancestral population size. For the sake of comparison, likelihood ratio tests comparing each nested model to the full model rejected 20 of 24 nested models. The four that were not rejected, combined with the full model, represent 94.2% of the cumulative model weight from the information-theoretic analysis. Coalescent simulations under both the earliest Pleistocene (129 000 generations, 2.58 Ma) and upper Pleistocene (6300 generations, 126 ka) divergence scenarios were rejected (Table 4). However, divergence during the last glacial maximum (900 generations, 18 ka) was not rejected, and neither was the single population scenario under either the highest or lowest credible estimates for N_e. Inclusion of migration in these simulations did not qualitatively change the results.

Discussion

Three hypotheses

Our results do not favour two of the three hypotheses concerning the relationships of *Lilium pyrophilum* advanced by Skinner & Sorrie (2002). First, it is unlikely that *L. pyrophilum* simply represents a disjunct population of *L. iridollae*: the ITS phylogeny unambiguously allies *L. pyrophilum* with *L. superbum*, whereas *L. iridollae* is most closely related to *L. michauxii*. That *L. pyrophilum* and *L. iridollae* are independent only heightens the conservation concern of each of these rare species.

Second, the hypothesis that the species originated as a hybrid between *L. michauxii* and *L. superbum* is not supported by network analyses (Fig. 4). If *L. pyrophilum* represented a recent hybrid, single-copy nuclear loci should be related to both parental species. Instead, most *L. pyrophilum* and *L. superbum* haplotypes are closely related to each other (and many are shared), while they show less similarity to *L. michauxii*. The phylogenetic analysis of ITS sequences placed the *L. pyrophilum* samples with *L. superbum* sequences only, to the exclusion of the *L. michauxii* sequences.

Lilium pyrophilum appears to be a peripheral isolate of *L. superbum*. Our results indicate that the overall magnitude of divergence between the two lily species is very low and that the origin of *L. pyrophilum* is likely to have been very recent, i.e. during the latter Pleistocene or Holocene. Our estimated divergence date from the IMa2 analysis of 188 ka (Fig. 5b) would fall within the Illinoian glacial period. The minimum credible divergence time of 84 ka would seem to indicate that *L. pyrophilum* is in fact isolated from *L. superbum*. In spite of low F_{ST} values (Table 2), zero probability is assigned to the most recent divergence times in this analysis. The results of the simulation

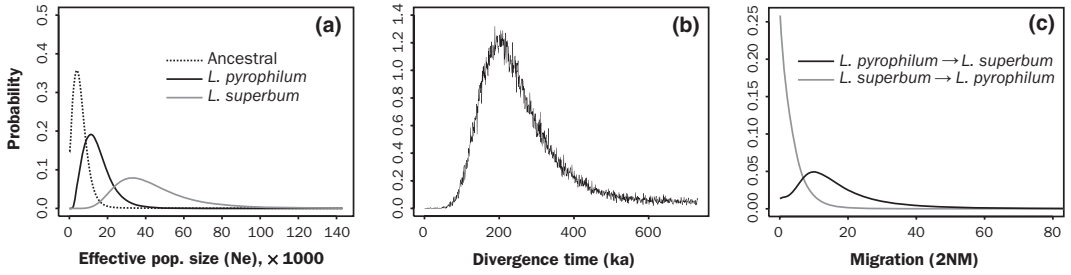

Fig. 5 Posterior probability distributions for IMa2 model parameters under the full model. (a) Effective population size for both species and the ancestral population. Both descendent taxa are inferred to have larger effective population size in this analysis. Estimated values for N_e are *Lilium pyrophilum*, 11 400 (95% HPD 2800–29 700); *L. superbum*, 32 900 (12 800–86 900); and ancestral, 3900 (0–14 400). (b) Divergence time. No probability is found for divergence times near zero; however, the distribution fails to reach zero at the upper end. The peak corresponds to a value of 188 (84–586) ka. (c) Migration rate. Highest probability for migration from *L. superbum* into *L. pyrophilum* is zero; there is, however, a higher probability of migration in the opposite direction (2 NM = 9.98).

Table 3 IMa2 analysis of nested models

Model description	log(*P*)	Terms	AIC	ΔAIC	Model weight	Cum. weight	d.f.	2LLR	*P*-value, LRT
θ (*pyrophilum*) = θ (*superbum*), *m* zero from *superbum* to *pyrophilum*	−4.442	3	14.884	0	0.301	0.301	2	2.986	0.2247
m zero from *superbum* to *pyrophilum*	−3.825	4	15.65	0.766	0.2052	0.5062	1	1.752	0.1856
Full IM model	−2.949	5	15.898	1.014	0.1813	0.6875	—	—	—
θ (*pyrophilum*) = θ (*superbum*)	−3.972	4	15.944	1.06	0.1772	0.8647	1	2.045	0.1527
Symmetrical migration	−4.803	4	17.606	2.722	0.0772	0.9419	1	3.707	0.0542
m zero from *pyrophilum* to *superbum*	−6.29	4	20.58	5.696	0.0174	0.9593	1	6.681	0.0097
θ (*pyrophilum*) = θ (ancestral), *m* zero from *superbum* to *pyrophilum*	−7.985	3	21.97	7.086	0.0087	0.968	2	10.07	0.0065
θ (*pyrophilum*) = θ (ancestral), *m* zero from *pyrophilum* to *superbum*	−8.116	3	22.232	7.348	0.0076	0.9757	2	10.33	0.0057
θ (*pyrophilum*) = θ (*superbum*), symmetrical migration	−8.408	3	22.816	7.932	0.0057	0.9814	2	10.92	0.0043
θ (*pyrophilum*) = θ (ancestral), symmetrical migration	−8.424	3	22.848	7.964	0.0056	0.987	2	10.95	0.0042
All θ equal, *m* zero from *superbum* to *pyrophilum*	−9.858	2	23.716	8.832	0.0036	0.9906	3	13.82	0.0032
θ (*pyrophilum*) = θ (ancestral)	−7.899	4	23.798	8.914	0.0035	0.9941	1	9.9	0.0017
θ (*superbum*) = θ (ancestral), *m* zero from *superbum* to *pyrophilum*	−9.192	3	24.384	9.5	0.0026	0.9967	2	12.49	0.0019
All θ equal	−9.858	3	25.716	10.832	0.0013	0.9981	2	13.82	0.001
θ (*superbum*) = θ (ancestral)	−9.192	4	26.384	11.5	0.001	0.999	1	12.49	0.0004
θ (*pyrophilum*) = θ (*superbum*), *m* zero from *pyrophilum* to *superbum*	−10.63	3	27.26	12.376	0.0006	0.9997	2	15.36	0.0005
θ (*superbum*) = θ (ancestral), symmetrical migration	−12.1	3	30.2	15.316	0.0001	0.9998	2	18.3	0.0001
All θ equal, symmetrical migration	−13.4	2	30.8	15.916	0.0001	0.9999	3	20.9	0.0001
θ (*superbum*) = θ (ancestral), zero migration	−14.26	2	32.52	17.636	0	0.9999	3	22.63	0
Zero migration	−13.35	3	32.7	17.816	0	1	2	20.8	0
θ (*superbum*) = θ (ancestral), *m* zero from *pyrophilum* to *superbum*	−14.26	3	34.52	19.636	0	1	2	22.63	0
All θ equal, *m* zero from *pyrophilum* to *superbum*	−18.52	2	41.04	26.156	0	1	3	31.13	0
θ (*pyrophilum*) = θ (*superbum*), zero migration	−24.86	2	53.72	38.836	0	1	3	43.83	0
θ (*pyrophilum*) = θ (ancestral), zero migration	−29.23	2	62.46	47.576	0	1	3	52.57	0
All θ equal, zero migration	−30.93	1	63.86	48.976	0	1	4	55.97	0

Models include the full IM model and 24 simpler nested models for the two-population case. Information-theoretic statistics, based on log(*P*), follow Burnham & Anderson (2002) and have been sorted by model weight. Models not rejected under traditional-likelihood ratio tests (LRT) are included in the 95% confidence set of models selected by AIC.

analysis lead us to interpret the IMa2 results with caution, however, because they reject divergence >6300 generations (126 ka) ago for each locus and fail to reject the scenarios with divergence at 900 generations (18 ka) and with no divergence (Table 4). The models tested in this approach, however, were simplified with respect to the full IMa2 model and treat each locus separately rather than jointly. Regardless of whether the IMa2 results or the coalescent simulation results are preferred, the isolation between the two taxa is not ancient. Mid- to late Pleistocene divergence times have been found in surprisingly few studies of plants (e.g. Strasburg & Rieseberg 2008; Bittkau & Comes 2009; Cooper *et al.* 2010).

Our results provide insight into the demographic patterns that have affected the two species. Deviations from neutral expectation indicate population expansion in both taxa (e.g. the average value for Tajima's *D* across three loci: *L. pyrophilum* = −1.35, *L. superbum* = −1.31, Table 1). This result is corroborated by the IMa2 analysis, which demonstrates modern effective population sizes higher than the ancestral, with the widespread *L. superbum* having a larger value (N_e 2.7 times that of the endemic *L. pyrophilum*, Fig. 5a). It is worth noting that the effective population size of *L. pyrophilum* (11 000 individuals) is surprisingly high considering the very small range of the species; in fact, our estimate of N_e is well in excess of the current census population size estimated by a recent inventory. Two factors may explain this discrepancy. First, our estimated generation time may be too low, which would cause us to overestimate effective population size (and underestimate divergence time). Second, agriculture, timber harvesting and fire suppression have

Table 4 Results of coalescent simulation study

	Marker		
Simulation model	AKT	AP	CP
Divergence time (in generations) without migration			
129 000	**0.000**	**0.000**	**0.000**
6300	**0.023**	**0.008**	**0.036**
900	0.992	0.379	0.394
Divergence time (in generations) with migration			
129 000	**0.001**	**0.001**	**0.006**
6300	**0.028**	**0.013**	**0.034**
900	0.839	0.438	0.422
No divergence			
High N_e	0.122	0.148	0.546
Low N_e	0.065	0.181	0.235

P-value for each model was obtained by comparison of either minimum (divergence) or maximum (no divergence) empirical *s* value (Slatkin & Maddison 1989) with simulated distributions of *s* under coalescent scenarios to test whether observed data were consistent with divergence times discussed in text. Simulations were based on assumed 20-year generation time.

dramatically transformed much of the landscape of the Sandhills over the past few hundred years, which may well have extirpated many populations. As these plants are long-lived outcrossers, too few generations may have elapsed for the impact of the current bottleneck to be fully reflected in the estimated N_e (Lande & Barrowclough 1987). Although our results suggest that the existing population has apparently been greatly reduced recently, much of the original genetic diversity remains and could be conserved, minimizing the impact of the present-day population bottleneck.

Gene flow is inferred from *L. pyrophilum* to *L. superbum*. Models including symmetrical migration are not weighted heavily compared with models that have zero or nearly zero gene flow from *L. superbum* to *L. pyrophilum* (Table 3). Presently, the two species are disjunct. However, the plants are visited by strong-flying pollinators, such as swallowtail butterflies and hummingbirds (Skinner 2002), and the seeds are adapted for wind dispersal. Why migration would be asymmetrical is unknown, but this could be explained by pollinator behaviour, dispersal or intrinsic barriers to gene flow.

Edaphic endemism in the Sandhills

The Sandhills pre-date the Pleistocene and may be substantially older, raising the possibility that some endemic taxa may have originated in the Pliocene or earlier and maintained populations in the region continuously. How might Pleistocene climate changes have affected the distribution of *Lilium* spp. in the coastal plain and effected the isolation of *L. pyrophilum*? While periods of severe climate change may eliminate edaphic endemics that are unable to migrate to areas with a suitable climate and substrate, edaphic endemics may in fact be likely to endure climate change in their geographic ranges. As their niches are defined more by soils than climate, they are likely to remain the best competitors on restrictive soils under a wide range of conditions. In fact, the degree of edaphic restriction exhibited by a species often varies with climate: populations may be widespread in environments with low competition and edaphically restricted in more favourable climates (Brooks 1987; Harrison *et al.* 2009).

The edaphic conditions that currently support populations of *L. pyrophilum* have probably been relatively stable, because the erosional process has no doubt continually exposed the interface between permeable and impermeable soils, creating seeps. Yet, the divergence between *L. pyrophilum* and *L. superbum* is comparatively recent. Genetic diversity of *L. pyrophilum*, while lower than that of *L. superbum*, is still high, making a vicariant scenario likely. Thus, the phenotypic divergence described by Skinner & Sorrie (2002) probably occurred in the context of large populations and substantial gene flow.

The combination of long-term persistence and recent divergence of *L. pyrophilum* indicates that this species descends from locally adapted populations that were stranded in the Sandhills as *L. superbum* retreated to higher elevations. It is not clear why the intervening Piedmont region supports neither taxon; however, many groups show a similar disjunction (Braun 1955; Sorrie & Weakley 2001). This study indicates that for these lilies, at least, the disjunction coincided with Pleistocene climate oscillations; this may apply to other taxa that share similar distributions. More in-depth studies of the *L. pyrophilum/L. superbum* system, using microsatellite markers, will quantify genetic structure within *L. pyrophilum*, and gene flow within and between *L. pyrophilum* and *L. superbum*. These more detailed analyses will improve estimates of divergence time and gene flow and identify populations of high conservation priority. Better understanding of this group will provide further insight into the role of edaphic specialization, possibly brought on by climate change, in promoting diversification.

Acknowledgements

We thank Fort Bragg Military Reservation and the Endangered Species Branch for logistic support and the Construction Engineering Research Laboratory (US Army Corps of Engineers Agreement #W9132T-07-2-0019) for funding. We also thank Xiang Liu, David Thomas, Patrick Zhou, Esther Ichugo, Matt

Cleary and Jacob Hilton for assistance with laboratory work and Marshall Wilson, Tom Phillips, Mac Alford, Rob Naczi, Gary Shurette, Viola Walker, Emil Devito, Heather Sullivan, Paul Manos, John Pogacnik, James Smith, Wayne Longbottom, Misty Buchanan and others for assistance locating *Lilium* populations.

References

Axelrod DI (1972) Edaphic aridity as a factor in angiosperm evolution. *The American Naturalist*, **106**, 311–320.

Ayling SC, Brown TA (2008) Novel methodology for construction and pruning of quasi-median networks. *BMC Bioinformatics*, **9**, 115.

Ayres DR, Ryan FJ (1999) Genetic diversity and structure of the narrow endemic *Wyethia reticulata* and its congener *W. bolanderi* (Asteraceae) using RAPD and allozyme techniques. *American Journal of Botany*, **86**, 344–353.

Baldwin BG (2005) Origin of the serpentine-endemic herb *Layia discoidea* from the widespread *L. glandulosa* (Compositae). *Evolution*, **59**, 2473–2479.

Baskauf CJ, McCauley DE, Eickmeier WG (1994) Genetic analysis of a rare and a widespread species of *Echinacea* (Asteraceae). *Evolution*, **48**, 180–188.

Bittkau C, Comes HP (2009) Molecular inference of a Late Pleistocene diversification shift in *Nigella* s. lat. (Ranunculaceae) resulting from increased speciation in the Aegean archipelago. *Journal of Biogeography*, **36**, 1346–1360.

Braun EL (1955) The phytogeography of unglaciated Eastern United States and its interpretation. *The Botanical Review*, **21**, 297–375.

Brooks RR (1987) *Serpentine and its Vegetation: A Multidisciplinary Approach*. Dioscorides Press, Portland, OR.

Brunsfeld SJ, Sullivan J, Soltis DE, Soltis PS (2001) Comparative phylogeography of northwestern North America: a synthesis. In: *Integrating Ecology and Evolution in a Spatial Context* (eds Silvertown J, Antonovics J), pp. 319–340. Blackwell Science, Oxford.

Burnham KP, Anderson DR (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York.

Cabe S, Nickerson JG, Hoffman CW, Farrell K (1992) Cretaceous and Tertiary stratigraphy of the northern Sandhills area, North Carolina. In: *Geologic field guides to North Carolina and vicinity* (eds Dennison J, Stewart KG), Annual Meeting of the Southeastern Section of the Geological Society of America. Winston-Salem, NC: Geological Society of America, Boulder, CO.

Carstens BC, Stoute HN, Reid NM (2009) An information-theoretical approach to phylogeography. *Molecular Ecology*, **18**, 4270–4282.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Cooper EA, Whittall JB, Hodges SA, Nordborg M (2010) Genetic variation at nuclear loci fails to distinguish two morphologically distinct species of *Aquilegia*. *PLoS ONE*, **5**, e8655.

Crawford DJ, Ornduff R, Vasey MC (1985) Allozyme variation within and between *Lasthenia minor* and its derivative species, *L. maritima* (Asteraceae). *American Journal of Botany*, **72**, 1177–1184.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.

Drummond A, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Edwards SV, Beerli P (2000) Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, **54**, 1839–1854.

Edwards CE, Soltis DE, Soltis PS (2006) Molecular phylogeny of *Conradina* and other scrub mints (Lamiaceae) from the southeastern USA: evidence for hybridization in Pleistocene refugia? *Systematic Botany*, **31**, 193–207.

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology and Systematics*, **24**, 217–242.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405.

Freeland JR, Biss P, Conrad KF, Silvertown J (2010) Selection pressures have caused genome-wide population differentiation of *Anthoxanthum odoratum* despite the potential for high gene flow. *Journal of Evolutionary Biology*, **23**, 776–782.

Gaut BS (1998) Molecular clocks and nucleotide substitution rates in higher plants. *Evolutionary Biology*, **30**, 93–120.

Gitzendanner MA, Soltis PS (2000) Patterns of genetic variation in rare and widespread plant congeners. *American Journal of Botany*, **87**, 783–792.

Godt MJW, Hamrick JL (1993) Genetic diversity and population structure in *Tradescantia hirsuticaulis* (Commelinaceae). *American Journal of Botany*, **80**, 959–966.

Goudet J, Raymond M, de-Meeus T, Rousset F (1996) Testing differentiation in diploid populations. *Genetics*, **144**, 1933–1940.

Gugger PF, Sugita S, Cavender-Bares J (2010) Phylogeography of Douglas-fir based on mitochondrial and chloroplast DNA sequences: testing hypotheses from the fossil record. *Molecular Ecology*, **19**, 1877–1897.

Hamrick JL, Godt MJW (1990) Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding, and Genetic Resources* (eds Brown AHD, Clegg MT, Kahler AL, Weir BS), pp. 43–63. Sinauer Associates, Sunderland, MA.

Harrison S, Damschen E, Going BM (2009) Climate gradients, climate change, and special edaphic floras. *Northeastern Naturalist*, **16**, 121–130.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences*, **104**, 2785–2790.

Howarth DG, Baum DA (2005) Genealogical evidence of homoploid hybrid speciation in an adaptive radiation of

Scaevola (Goodeniaceae) in the Hawaiian Islands. *Evolution*, **59**, 948–961.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Karron JD (1987) A comparison of levels of genetic polymorphism and self-compatibility in geographically restricted and widespread plant congeners. *Evolutionary Ecology*, **1**, 47–58.

Knowles LL (2009) Statistical phylogeography. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 593–612.

Kruckeberg AR (1986) An essay: the stimulus of unusual geologies for plant speciation. *Systematic Botany*, **11**, 455–463.

Kruckeberg AR, Rabinowitz D (1985) Biological aspects of endemism in higher plants. *Annual Review of Ecology and Systematics*, **16**, 447–479.

Lande R, Barrowclough GF (1987) Effective population size, genetic variation, and their use in population management. In: *Viable Populations for Conservation* (ed. Soulé ME), pp. 87–124. Cambridge University Press, Cambridge.

Lande R, Engen S, Sæther B-E (2003) *Stochastic Population Dynamics in Ecology and Conservation*. Oxford University Press, Oxford.

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

Maddison WP, Maddison DR (2010) *Mesquite: A Modular System for Evolutionary Analysis*. http://mesquiteproject.org.

Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics*, **21**, 260–262.

Mehta BD, Jog SP, Johnson SC, Murthy PPN (2006) Lily pollen alkaline phytase is a histidine phosphatase similar to mammalian multiple inositol polyphosphate phosphatase (MINPP). *Phytochemistry*, **67**, 1874–1886.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

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

Nishikawa T, Okazaki K, Uchino T, Arakawa K, Nagamine T (1999) A molecular phylogeny of *Lilium* in the internal transcribed spacer region of Nuclear ribosomal DNA. *Journal of Molecular Evolution*, **49**, 238–249.

North Carolina Natural Heritage Program (2007) *North Carolina Natural Heritage Program Element Occurrence Database*. NC Department of Environment and Natural Resources, Raleigh, NC, 27699–1610.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Oliveira LO, Huck RB, Gitzendanner MA, Judd WS, Soltis DE, Soltis PS (2007) Molecular phylogeny, biogeography, and systematics of *Dicerandra* (Lamiaceae), a genus endemic to the southeastern United States. *American Journal of Botany*, **94**, 1017–1027.

Ornduff R (1965) Ornithocoprophilous endemism in Pacific basin angiosperms. *Ecology*, **46**, 864–867.

Parsons RF (1976) Gypsophily in plants-a review. *American Midland Naturalist*, **96**, 1–20.

Peruzzi L, Leitch IJ, Caparelli KF (2009) Chromosome diversity and evolution in Liliaceae. *Annals of Botany*, **103**, 459–475.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

Rabinowitz D (1981) Seven forms of rarity. In: *The Biological Aspects of Rare Plant Conservation* (ed. Synge H), pp. 205–217. Wiley, New York.

Raduski AR, Rieseberg LH, Strasburg JL (2010) Effective population size, gene flow, and species status in a narrow endemic sunflower, *Helianthus neglectus*, compared to its widespread sister species, *H. petiolaris*. *International Journal of Molecular Sciences*, **11**, 492–506.

Rajakaruna N (2004) The edaphic factor in the origin of plant species. *International Geology Review*, **46**, 471–478.

Rambaut A (1996) *Se-Al: Sequence Alignment Editor*. Available at http://evolve.zoo.ox.ac.uk/.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Raymond M, Rousset F (1995) An exact test for population differentiation. *Evolution*, **49**, 1280–1283.

Rieseberg LH, Carter R, Zona S (1990) Molecular tests of the hypothesized hybrid origin of two diploid *Helianthus* species (Asteraceae). *Evolution*, **44**, 1498–1511.

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology*, **132**, 365–386.

Schafale MP, Weakley AS (1990) *Classification of the Natural Communities of North Carolina, Third Approximation*. North Carolina Department of Environment, Health, and Natural Resources, Division of Parks and Recreation, Natural Heritage Program, Raleigh.

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany*, **94**, 275–288.

Skinner MW (2002) *Lilium*. In: *Flora of North America North of Mexico. Magnoliophyta: Liliidae: Liliales and Orchidales* (ed. Flora of North America Editorial Committee), pp. 172–197. Oxford University Press, Oxford.

Skinner MW, Sorrie BA (2002) Conservation and ecology of *Lilium pyrophilum*, a new species of Liliaceae from the Sandhills region of the Carolinas and Virginia, USA. *Novon*, **12**, 94–105.

Slatkin M, Maddison WP (1989) A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics*, **123**, 603–613.

Sohl NF, Owens JP (1991) Cretaceous stratigraphy of the Carolina Coastal Plain. In: *The Geology of the Carolinas: Carolina Geological Society Fiftieth Anniversary Volume* (eds Horton JW, Zullo VA, Carolina Geological Society), pp. 191–220. Univ. of Tennessee Press, Knoxville, TN.

Soltis PS, Soltis DE (2003) Applying the bootstrap in phylogeny reconstruction. *Statistical Science*, **18**, 256–267.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, **15**, 4261–4293.

Sorrie BA, Weakley AS (2001) Coastal plain vascular plant endemics: phytogeographic patterns. *Castanea*, **66**, 50–82.

Sorrie BA, Gray JB, Crutchfield PJ (2006) The vascular flora of the longleaf pine ecosystem of Fort Bragg and Weymouth Woods, North Carolina. *Castanea*, **71**, 129–161.

Stanford AM, Harden R, Parks CR (2000) Phylogeny and biogeography of *Juglans* (Juglandaceae) based on matK and ITS sequence data. *American Journal of Botany*, **87**, 872–882.

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, **73**, 1162–1169.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

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

Swofford DL (2002) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer, Sunderland, MA.

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

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

Wall WA, Douglas NA, Xiang QY, Hoffmann WA, Wentworth TR, Hohmann MG (2010) Evidence for range stasis during the latter Pleistocene for the Atlantic Coastal Plain endemic genus, *Pyxidanthera* Michaux. *Molecular Ecology*, **19**, 4302–4314.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wheelan SJ, Church DM, Ostell JM (2001) Spidey: a tool for mRNA-to-genomic alignments. *Genome Research*, **11**, 1952–1957.

White TJ, Bruns T, Lee S, Taylor J (1990) Amplifications and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR Protocols: A Guide to Methods and Applications* (eds Innis M, Gelfand D, Sninsky J, White T), pp. 315–322. Academic Press, San Diego, CA.

Whittall JB, Syring J, Parks M *et al.* (2010) Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. *Molecular Ecology*, **19**, 100–114.

Williamson PS, Bazeer SK (1997) Self-incompatibility in *Abronia macrocarpa* (Nyctaginaceae). *Southwestern Naturalist*, **42**, 409–415.

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

Zonneveld BJM, Leitch IJ, Bennett MD (2005) First nuclear DNA amounts in more than 300 angiosperms. *Annals of Botany*, **96**, 229–244.

Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. http://code.google.com/p/garli/.

N.A.D. is an evolutionary biologist who focuses on phylogenetics and phylogeography of arid-adapted and edaphic endemic plants. W.A.W. is a plant ecologist interested in the interplay between population genetics and population dynamics of rare and endemic species. Q.Y.X.'s research explores the mechanisms underlying biodiversity patterns and morphological variation, with emphasis on the Cornales. W.A.H. is interested in ecology and conservation of savanna ecosystems. T.R.W. is a vegetation scientist with research interests in vegetation/environment relationships and biological diversity. J.B.G. is a botanist working on conservation of rare plants in Coastal Plain habitats. M.G.H. uses diverse approaches to inform rare species conservation.

Data accessibility

Sample and haplotype information is found in Table S1 (Supporting Information). DNA sequences: GenBank accessions JF829316–JF829423 (Table S2, in Supporting Information). ITS data and phylogenetic tree available at http://purl.org/phylo/treebase/phylows/study/TB2:S11519.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Sampling. Haplotype numbers correspond to sample labels in Figs 3 and 4, and to names accessioned in GenBank (Table S2, in Supporting Information). For AP and AKT sequences, integers identify haplotypes recovered more than once in this study and other identifiers refer to unique haplotypes

Table S2 GenBank accession numbers. Haplotype names correspond to samples in Table S1 (Supporting Information)

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