apple *M. domestica* (Cornille *et al.* 2012) and is considered to be endangered (Jacques *et al.* 2009). Efforts to decipher the genetic structure and demographic history of this tree species are thus timely (i) to mitigate the effects of climate change and prevent negative effects of habitat fragmentation on this endangered species and (ii) to increase the genetic resources available for apple breeding programs, through the identification of genes conferring resistance to pathogens or tolerance of diverse abiotic stress conditions. The genetic structure of the European crabapple has been studied only in limited areas (Coart *et al.* 2003; Larsen *et al.* 2006) and appears to be weak at fine scale, suggesting a high dispersal capacity. No study has yet investigated the phylogeography of this emblematic wild European species across its entire distribution range. We therefore used population genetic analyses and ENM to investigate the glacial refugia and postglacial recolonization history of *M. sylvestris*. For genetic analyses, we used a comprehensive set of *M. sylvestris* individuals sampled throughout Europe and 26 microsatellite markers. We addressed the following questions: (i) Did wild apples survive the last glacial period in a single refuge area in Europe or at multiple areas? Did all relict populations contribute equally to the postglacial recolonization of Europe by wild apples? (ii) Can we detect the genetic consequences of successive founder events during postglacial colonization, that is, does genetic diversity decline with increasing distance from refugia? (iii) Can we detect genetic patterns of isolation-by-distance (IBD) and obtain information about dispersal capacities? (iv) Did admixture occur between recolonizing populations during expansion of the postglacial range of this species? (v) Can we reconstruct habitats that were suitable for European crabapple in the past by ENM methods?

Materials and methods

Sampling

We have previously demonstrated the existence of gene flow from the cultivated apple *M. domestica* to the European wild apple (Cornille *et al.* 2012). We therefore initially carried out a STRUCTURE 2.3.3 (Pritchard *et al.* 2000) analysis, including 40 reference *M. domestica* cultivars previously identified as displaying no introgression from the European crabapple (i.e. with membership coefficients >0.9 to the *M. domestica* gene pool), and the entire *Malus sylvestris* data set comprising 837 individuals. We retained only the 381 individuals assigned with a value of >0.9 to the *M. sylvestris* gene pool for further analyses. Leaf material was collected at 37 sites (including 25 sites with at least four individuals) across Europe (Table S1 and Fig. S1, Supporting information), covering most of the geographical distribution of the European crabapple except Spain and Sweden (see Euforgen map of the European crabapple http://www.euforgen.org/distribution_maps.html). As the European crabapple is a scattered species (i.e. 1 individual/ha), we defined site as a group of close individuals (<20 km); the geographical coordinates for each site were defined as the average of geographical coordinates over all individuals from that site.

DNA extraction and microsatellite genotyping

DNA was extracted with the NucleoSpin® plant DNA extraction kit II (Macherey & Nagel, Düren, Germany). PCR amplification was performed with the Multiplex PCR kit (Qiagen). We used 26 microsatellites, distributed over the 17 chromosomes, typed in 10 different multiplex reactions as previously described (Patocchi *et al.* 2009; Cornille *et al.* 2012). We retained only multilocus genotypes with fewer than 25% missing data. We checked the suitability of the markers for population genetic analysis with ARLEQUIN (Excoffier & Lischer 2010). None of the 26 microsatellite markers deviated significantly from the neutral equilibrium model, as shown by the nonsignificant *P*-values obtained in Ewens–Watterson tests, and no pair of markers was in significant linkage disequilibrium (Raymond & Rousset 1995; Rousset 2008). The markers used may therefore be considered to be unlinked and to be evolving in a quasi-neutral manner.

Descriptive statistics

We tested for the occurrence of null alleles with MICROCHECKER 2.2.3 (Van Oosterhout *et al.* 2004). Allelic richness and private allele richness were calculated with ADZE (Szpiech *et al.* 2008), for sites (i.e. geographical locations) and populations (i.e. clusters inferred by TESS analyses including hybrids up to 0.55 membership coefficient in the given cluster), using sample sizes of $N = 12$ (six individuals × two chromosomes) and $N = 200$ (100 individuals × two chromosomes), corresponding to the smallest number of observations for sites and populations, respectively. Heterozygosity, Weir and Cockerham *F*-statistics and Hardy–Weinberg equilibrium were assessed with GENEPOP 4.0 (Raymond & Rousset 1995; Rousset 2008). Only sampling sites with at least four successfully genotyped specimens were included in site-specific computations (25 sites in total comprising $N = 344$ individuals) except for ADZE analysis in which 22 sites were included (i.e. geographical locations with at least four individuals and within which at least two individuals were successfully genotyped for each marker). The population-specific