Hijmans *et al.* 2005) and records of the presence of *M. sylvestris* obtained for sampled individuals ($N = 381$). We estimated the pairwise correlation of values for the 19 bioclimatic variables by calculating Pearson's correlation coefficients and retained only the variables that were not strongly correlated (i.e. with Pearson's correlation coefficients <0.75). We tested both sets of bioclimatic variables because we had no primer assumptions about species' preferences, and we wanted to compare the respective LGM projections when taking into account autocorrelation and the overfitting of the data. We removed duplicated coordinate data points, resulting in 73 presences in total (data set S1, Supporting information) to evaluate the distribution at the LGM. These models make the assumptions that climate is one of the main factors driving species distribution and that the climatic niche of this species has remained largely unchanged in recent centuries (Text S1, Supporting information).

## Results

### *Population structure*

Summary statistics for genetic variability are shown in Tables S3 and S4 (Supporting information). For the 25 sites with at least four samples, the mean number of genotypes was 13.8 ± 8.4 (average ± standard deviation), allelic richness was 3.7 ± 0.4 (range: 2.6–4.4) and genetic diversity was 0.84 ± 0.14 (range: 0.55–0.89), on average, across markers. Heterozygote deficit, estimated over the whole data set, was highly significant ($P < 0.001$), but low ($F_{IS} = 0.03$, with a mean of 0.03 ± 0.07 per site and per marker). The mean $F_{ST}$ across loci were small ($F_{ST} = 0.10$, range: 0.008–0.280) but significant, for all pairs of sites ($P < 0.02$, Table S5, Supporting information).

The results of TESS analyses are shown in Figs 2 and 3. For $K = 2$, the analyses revealed a clear west/east partitioning. The simulations for $K = 3$ split the eastern cluster into NE and SE clusters. A similar pattern was obtained for $K = 4$, whereas for $K = 5$, a fourth cluster was identified in Bosnia-Herzegovina. STRUCTURE analyses generated congruent clustering patterns (Figs S2 and S3, Supporting information).

In TESS analyses, DIC values decreased monotonically from $K = 2$ to $K = 6$. Thus, increasing the number of clusters continually improved the fit of the model to the data. However, DIC seemed to decrease more slowly after $K = 3$ (Fig. S4, Supporting information), suggesting that further increases in $K$ provided little information. In STRUCTURE analyses, the mode of the ΔK statistic was observed at $K = 2$ ($\Delta K = 1670$, $Pr|\ln L = 2683$, Fig. S5, Supporting information), but $\Delta K$ was still high at $K = 3$ ($\Delta K$=480, $Pr|\ln L = 838$), suggesting further improvement in the fit of the model. Based on the narrow geographical distribution of the clusters inferred at $K > 3$, and the $\Delta K$ and DIC values obtained, $K = 3$ was considered the most biologically relevant clustering solution for subsequent historical inference.

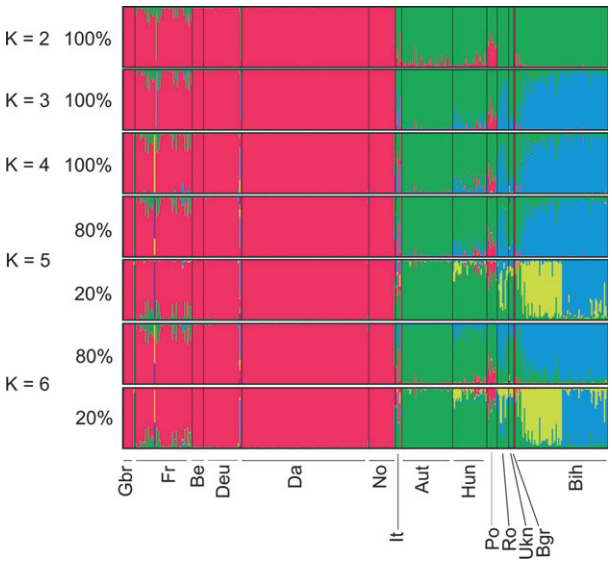


**Fig. 2** Population structure in the European crabapple *Malus sylvestris* ($N = 381$, 37 sites across Europe), inferred with the Bayesian clustering algorithm implemented in TESS. Each individual is represented by a vertical bar, partitioned into $K$ segments representing the proportions of ancestry of its genome in $K$ clusters. When several clustering solutions ('modes') were represented within replicate runs, the proportion of simulations represented by each mode is shown. For a better visualization, the 37 sites were grouped by country. Gbr, Great Britain; Fr, France; Be, Belgium; Deu, Germany; Da, Denmark; No, Norway; It, Italy; Aut, Austria; Hun, Hungary; Po, Poland; Ro, Romania; Ukn, Ukraine; Bgr, Bulgaria; Bih, Bosnia-Herzegovina.

We used the TESS membership coefficient inferred at $K = 3$ to define the three populations used in subsequent analyses. Genotypes were assigned to a given population if their membership coefficient for that population exceeded 0.55. Five genotypes could not be assigned to any population and were not included in subsequent analyses; $N = 376$ individuals were thus retained for population-specific computations. The three populations are hereafter referred to as the 'western' (W, red, $N = 213$), the 'north-eastern' (NE, green, $N = 90$) and 'south-eastern' (SE, blue, $N = 73$) populations. The W population was relatively homogeneous, with 91% of genotypes having membership coefficients >0.9 for that population (Figs 2 and 3). The NE and SE populations presented higher levels of admixture, with 31% ($N = 28$) and 26% ($N = 19$) of genotypes, respectively, having membership coefficients <0.9