Appendix 2: Field case study details

Functional traits - We used the functional trait database ANDROSACE (AlpiNe Database ResOurce for Species And eCosystems fEatures, Thuiller et al. unpublished). The database includes trait information for Alpine plants from individual projects (Albert et al. 2010a,b) and freely available databases such as LEDA (Knevel et al. 2003), BioFlor (Kühn et al. 2004), Ecoflora (Fitter & Peat 1994) and CATMINAT (Julve 1998).

Suitability index - The species distribution model was built using the Random Forest algorithm (Breiman 2001) for each species independently in order to estimate species abiotic niches, and thereby their probability of occurrence. To ensure a broad estimate of species niches, the full plant occurrences database for the French Alps was used together with climatic (temperature of the coldest month of the year, relative summer wetness and winter precipitation) and topographic variables (topographic wetness index and topographic position, i.e. topographic convexity or concavity). The climatic variables were originally extracted from the AURELHY (Analysis Using the RElief for HYdrometeorology) dataset (Benichou & Le Breton 1987), downscaled to a 100m resolution, while the topographic variables came from a 50m resolution Digital Elevation Model. Based on the species distribution models, we extracted the probability of presence (suitability indices) for the 373 plant species of the community data in each of the 95 communities.

Phylogeny - A genus-level phylogeny of alpine plants was built using the workflow proposed in Roquet et al. (accepted) with DNA sequences downloaded from Genbank: three conserved chloroplastic regions (rbcL, matK and ndhF) and 8 regions for certain families or orders (atpB, ITS, psbA-trnH, rpl16, rps4, rps4-trnS, rps16, trnL-F) which were aligned separately by taxonomic clustering. All sequences were aligned with 3 methods (MUSCLE, Edgar 2004; MAFFT, Katoh et al. 2005; Kalign, Lassmann and Sonnhammer 2005) and checked by eye. The best alignment for each region was selected and depurated with TrimAl (Capella-Gutierrez et al. 2009). Phylogenetic

inference by maximum-likelihood (ML) was conducted with RAxML (Stamatakis et al. 2006) applying a supertree constraint at the family-level based on Davies et al. (2004) and Moore et al. (2010). In order to obtain a phylogenetic tree with branch lengths equal to absolute evolutionary time, we dated the best ML tree with penalized-likelihood using r8s (Sanderson 2003) and 25 fossils constraints (Table S4). Finally, the tips of the phylogenetic tree were resolved with polytomies to obtain a species-level phylogeny.

References

Albert, C.H., Thuiller, W., Yoccoz, N. G., Douzet, R., Aubert, S., & Lavorel, S. 2010a. A multi-trait approach reveals the structure and the relative importance of intra-vs. interspecific variability in plant traits. *Functional Ecology* 24: 1192–1201.

Albert, C.H., Thuiller, W., Yoccoz, N. G., Soudant, A., Boucher, F., Saccone, P., & Lavorel, S. 2010b. Intraspecific functional variability: extent, structure and sources of variation. *Journal of Ecology* 98: 604–613.

Benichou, P. & Le Breton, O. 1987. Prise en compte de la topographie pour la cartographie des champs pluviométriques statistiques. *La Météorologie* 19.

Breiman, L. 2001. Random Forests. *Machine Learning* 45: 5–32.

Capella-Gutierrez, S., Silla-Martinez, J.M., Gabaldon, T. 2009 trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972-1973.

Davies, T.J., Barraclough, T.G., Chase, M.W., Soltis, P.S., Soltis, D.E., Savolainen, V. 2004. Darwin's abominable mystery: Insights from a supertree of the angiosperms. *Proceedings of the National Academy of Sciences* 101: 1904-1909.

Edgar, R.C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113

Fitter, A. & Peat, H. 1994. The ecological flora database. *Journal of Ecology* 82: 415–425.

Julve, P. 1998, ff. Baseflor. Index botanique, écologique et chorologique de la flore de France. URL: http://perso.wanadoo.fr/philippe.julve/catminat.htm.

Katoh K., Kuma K., Toh H. & Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic acids research* 33: 511-518.

Knevel, I., Bekker, R., Bakker, J. & Kleyer, M. 2003. Life-history traits of the Northwest European flora: The LEDA database. *Journal of Vegetation Science* 14: 611–614.

Kühn, I., Durka, W. & Klotz, S. 2004. BiolFlor — a new plant-trait database as a tool for plant invasion ecology. *Diversity and Distributions* 10: 363–365.

Lassmann T. & Sonnhammer E.L. 2005 Kalign--an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics* 6: 298.

Moore M.J., Soltis P.S., Bell C.D., Burleigh J.G. & Soltis D.E. 2010 Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proceedings of the National Academy of Science 107: 4623-4628.

Roquet, C., Lavergne, S. & Thuiller, W. Building megaphylogenies for macroecology:taking up the challenge. *Ecography.* Accepted. doi: 10.1111/j.1600-0587.2012.07773.x

Sanderson, M.J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19(2), 301-302.

Stamatakis, A. 2006 RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688-2690.