This is a repository copy of *A systematic conservation strategy for crop wild relatives in the Czech Republic*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/119229/

Version: Supplemental Material

Article:
Taylor, NG orcid.org/0000-0002-8643-826X, Kell, SP, Holubec, V et al. (3 more authors) (2017) A systematic conservation strategy for crop wild relatives in the Czech Republic. Diversity and Distributions, 23 (4). pp. 448-462. ISSN 1366-9516

https://doi.org/10.1111/ddi.12539

© 2017 John Wiley & Sons Ltd. This is the peer reviewed version of the following article: Taylor, N. G., Kell, S. P., Holubec, V., Parra-Quijano, M., Chobot, K. and Maxted, N. (2017), A systematic conservation strategy for crop wild relatives in the Czech Republic. Diversity Distrib., 23: 448–462. doi:10.1111/ddi.12539, which has been published in final form at https://doi.org/10.1111/ddi.12539. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving. Uploaded in accordance with the publisher's self-archiving policy.

Reuse
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher’s website.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Supporting Information for

A systematic conservation strategy for crop wild relatives in the Czech Republic

N.G. Taylor, S.P. Kell, V. Holubec, M. Parra-Quijano, K. Chobot and N. Maxted

---

Appendix S2: Species distribution modelling (detailed methods)

Species Distribution Models (SDMs) were created in MaxEnt (version 3.3.3k; Phillips *et al*., 2004, 2006). MaxEnt is a presence-only distribution modelling approach that estimates a target probability distribution by finding the distribution of maximum entropy (i.e. closest to uniform, or most spread out) within given constraints, typically environmental variables (Phillips *et al*. 2006). MaxEnt was used because (i) it has performed well in comparisons with other SDM techniques using presence-only data (Elith *et al*., 2006; Phillips *et al*., 2006; Pearson 2010) (ii) it is simple to understand and use and (iii) it is freely available online (www.cs.princeton.edu/~schapire/maxent/).

Location Records

Taxon location records were retrieved from a variety of databases (primarily the species occurrence database of the Nature Conservation Agency of the Czech Republic; AOPK ČR, 2012). The number of records from each data source is listed in Table S2.1. Location records were processed to remove duplicates based on location and date of collection, records from before 1950 and from gardens, as described in the main text. Note that only 130,426 records were available for distribution modelling. By default, MaxEnt only retains one record of each species per 30 arc-second cell, thus removing records that are aggregated in space or from the same location but different dates.

Table S2.1 Data sources for location records of Czech priority CWR species. These are spatiotemporally independent observations of priority CWR species (i.e. different location and/or date of observation). Note that only a subset of records (130,426) were available for distribution modelling: one record per 30 arc-second cell for each species.

Data source	Description	Number of records	% of records
AOPK ČR (2012)	Species occurrence database of the Nature Conservation Agency of the Czech Republic	204,838	99.07
Holubec *et al*. (2010)	Collecting databases at the Crop Research Institute, Prague	868	0.42
MZM (2013)	Online database of Czech herbarium records	329	0.16
GBIF (2012)	Global Biodiversity Information Facility: occurrence database	86	0.04
EVIGEZ (2014)	Czech National Plant Genetic Resource Information System: gene bank accession data	603	0.29
GENESYS (2014)	Global Portal on Plant Genetic Resources: plant genetic resources accession level data	36	0.02
		206,760	100.00

Environmental Variables

Of the vast array of environmental variables available, a set of 13 were selected to be used for the final SDMs (Table S2.2). These were the variables hypothesised to be of greatest relevance to the species being modelled (Phillips *et al*., 2006), based on a combination of subjective (expert opinion, literature review) and objective (percent contribution to pilot MaxEnt models) criteria (Parra-Quijano *et al*., 2012). Notably, land cover was not included as this is likely to have changed over the six decades for which we had CWR distribution data. In contrast, plants will have had time to adapt to comparatively stable bioclimatic and geological variables (Anderson & Martinez-Meyer, 2004). Spearman rank correlations between bioclimatic variables used in the final SDMs were $< |0.54|$.

Interpolated bioclimatic data in 30 arc-second resolution were downloaded from WorldClim (www.worldclim.org/tiles.php), 30 arc-second resolution topographic data from the DIVA GIS website (www.diva-gis.org/data) and edaphic data from the European Soils Database (ESDB, 2004; Panagos, 2006). A geological map was obtained from the Department of Nutrition at the Crop Research Institute, Prague. The European Joint Research Council provided 1.5 arc-minute resolution monthly irradiation data (Šúri *et al*., 2007). A continentality index k

(expressed as a percentage) was calculated using BIO7 (annual temperature range) and equation S1 (Gorczynski, 1922), where A is the annual temperature range in °C and Θ is the latitude in radians. Continentality is a basic climatic feature and an important influence on the climate of landlocked Central and Eastern European countries like the Czech Republic. It reflects the balance between the ocean and large land masses in influencing climate (Mikolášková, 2009).

$$k = \frac{1.7(A - 12 sin\theta)}{sin\theta} = \frac{1.7A}{sin\theta} - 20.4 \qquad \text{[eqn. S1]}$$

Matching resolutions and extents of environmental variables are necessary for MaxEnt to operate, so all environmental layers were resampled across the Czech Republic at a 30 arc-second resolution. Consequently, the SDMs were only built with reference to environmental conditions within the study area.

Table S2.2 Environmental variables selected for use in Species Distribution Models. CRI – Crop Research Institute Prague; CV – coefficient of variation; JRC – Joint Research Centre of the European Commission; Max. – maximum; Min. – minimum; Temp – temperature.

BIOCLIMATIC VARIABLES		**Data Type**	**Resolution**	**Source**
BIO1	Annual Mean Temp. (°C)	Continuous	30 arc-second	www.worldclim.org/tiles.php
BIO3	Isothermality (BIO2/BIO7) (* 100)	Continuous	30 arc-second	www.worldclim.org/tiles.php
BIO9	Mean Temp. of Driest Quarter (°C)	Continuous	30 arc-second	www.worldclim.org/tiles.php
BIO15	Precipitation Seasonality (CV)	Continuous	30 arc-second	www.worldclim.org/tiles.php
BIO16	Precipitation of Wettest Quarter (mm)	Continuous	30 arc-second	www.worldclim.org/tiles.php

GEOPHYSICAL VARIABLES		**Data Type**	**Resolution**	**Source**
Altitude	Elevation (m above sea level)	Continuous	30 arc-second	www.diva-gis.org/Data
Slope	Slope of land surface (degrees)	Continuous	30 arc-second	www.diva-gis.org/Data
Continentality	Gorczynski's (1922) Index of Continentality	Continuous	30 arc-second	Gorczynski (1922) www.worldclim.org/tiles.php
Irradiation	Mean solar irradiation of brightest quarter (May-Jul) 2006-11 (Wh/m^2)	Continuous	1.5 arc-minute	Šúri *et al*., (2007), JRC (2012)

EDAPHIC VARIABLES		**Data Type**	**Resolution**	**Source**
geology_bedrock	Bedrock Class	Categorical	—	CRI (pers. comm.)
stu_dom	Soil type (Dominant Soil Taxonomic Unit)	Categorical	—	Panagos (2006)
parmado	Dominant parental material	Categorical	—	Panagos (2006)
bs_top	Base saturation of topsoil	Categorical	30 arc-second	Panagos (2006)

Modelling Methods

SDMs were constructed largely using the default settings in MaxEnt as these suited our purpose and have been shown to give robust and reliable results (Phillips & Dudik, 2008). The SDM for each species was based on all spatially distinct records (in separate 30 arc-second cells). SDMs were constructed at the species level, but it is important to note that, for approximately 10% of species (highlighted in bold in Appendix S1) models were based on records for a subset of infraspecific taxa that met prioritization criteria (see main text). Thus, SDMs for these species do not represent the *entire range* of the species but only the range of the prioritized subspecies or variety.

Uneven sampling effort, leading to artificial clustering of species records (as opposed to clustering purely reflecting the species ecology) can severely affect the predictions of SDMs. To control for sampling bias, we fed a bias file into MaxEnt (in the Advanced Settings). The bias file described sampling intensity across all CWR taxa, proxied by the number of records of priority CWR species, extracted from our distribution database using a circular neighbourhood method with 30 arc-second grid cells and 10km diameter (Table S2.1; Fig. S2). A bias file of this kind is used to give weight to background data in proportion to sampling effort, and has been shown to substantially improve the accuracy of model predictions (Phillips *et al*. 2009; Syfert *et al*. 2013). We acknowledge that methods to deal with sampling bias in SDMs are under active research, but have chosen a method that is suitable for our data (with varying numbers of records per species, some relatively low) and is supported by empirical research.

The ‘tenth percentile training presence’ threshold was applied to convert the SDMs into binary presence/absence maps. If the probability of a species occurring in a cell is greater than the threshold, it is assumed to be present; otherwise the species is assumed to be absent. The tenth percentile training presence describes, for each species, the presence probability above which 90% of the species locations (training data) fall. This threshold performs better than choosing arbitrary thresholds and is relatively high compared to other thresholds calculated in MaxEnt, increasing our confidence that predicted suitable habitat will actually contain the associated species.

We only retained robust SDMs. We rejected models based on based on records in fewer than 10 separate 30 arc-second cells, because models built on few data points are not reliable (van Proosdij *et al*., 2016). We also rejected SDMs with an area under the curve (AUC) < 0.7 following Pearce & Ferrier (2000), unless they performed significantly better than random (AUC > 0.5) in at least 70% of replicate crossvalidated runs. We added this final criterion to avoid penalising models with relatively low AUC, because the use of presence-only data and bias files reduces the maximum obtainable AUC (Phillips *et al*., 2006; Phillips *et al*., 2009; Syfert *et al*., 2013). For each species, we obtained 10 replicate runs using crossvalidation, in which the data were randomly split into ten groups and each group used in turn as test data. Significance was assessed using binomial tests of omission for thresholded models (Phillips *et al*., 2006).

We retained 171 SDMs which we considered robust. For these species, mean AUC across the ten replicate runs ranged from 0.547 to 0.999. For 141 species, mean test AUC exceeded 0.7. For the other 30 species, at least 70% of replicate runs performed significantly better than random. Models for 28 species were rejected because of a small number (≤ 10) of spatially distinct records, whilst models for five species (*Glyceria declinata, Papaver argemone, Prunus avium*, *Vicia sativa* and *Vicia sepium*) were rejected because fewer than 70% of their replicate runs performed significantly better than random.

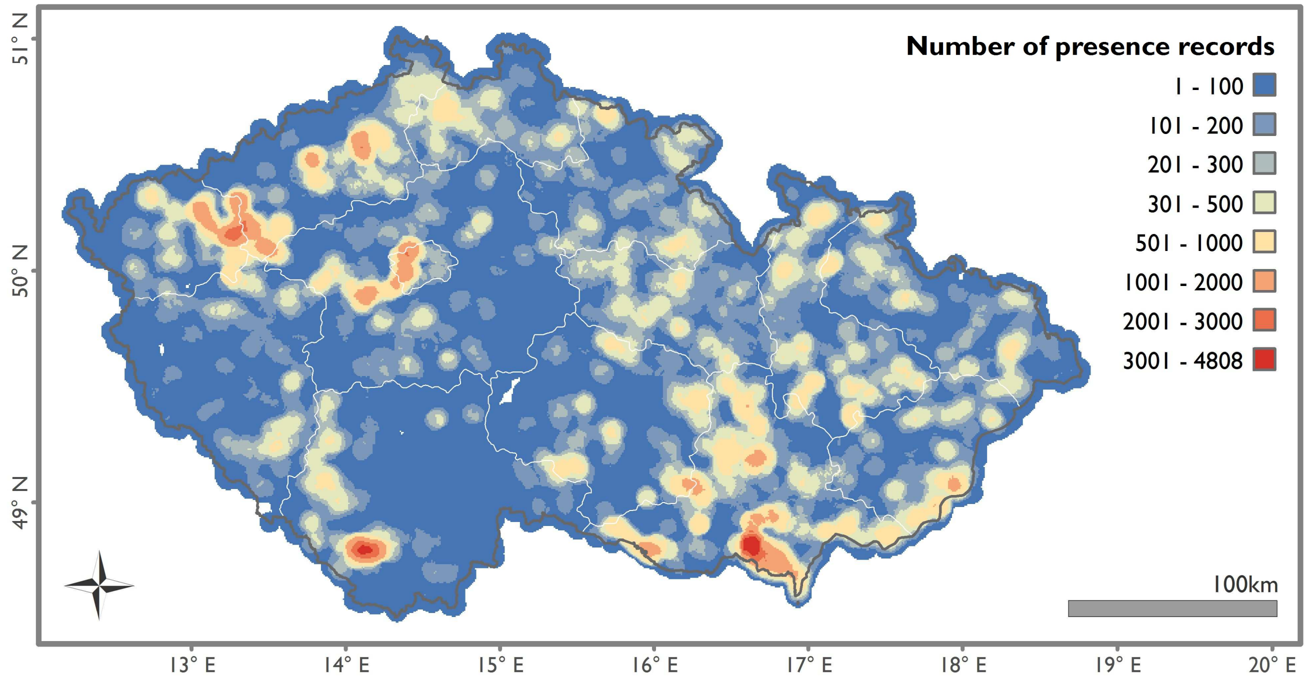


Figure S2 Number of records of 204 priority CWR species, extracted using a circular neighbourhood analysis with 30 arc-second grid cells and a 10km diameter. Total number of records: 206,760. Projection: Transverse Mercator 33N.

References

Anderson, R. & Martinez-Meyer, E. (2004) Modeling species' geographic distributions for preliminary conservation assessments: an implementation with the spiny pocket mice (*Heteromys*) of Ecuador. *Biological Conservation*, **116**, 167–179.

AOPK ČR (2012) *Nálezová database (Species occurrence database)*. Agentura Ochrany Přírody a Krajiny České Republiky (Nature Conservation Agency of the Czech Republic). Available with permission at: http://portal.nature.cz (last accessed 23 October 2012).

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

ESDB (2004) The European Soil Database distribution version 2.0, European Commission and the European Soil Bureau Network, CD-ROM, EUR 19945 EN 2004.

EVIGEZ (2014) *Evidence genetických zdrojù rostlin v ČR (Czech National Plant Genetic Resource Information System).* Available at: http://genbank.vurv.cz/genetic/resources/asp2/default_c.htm (last accessed 28 October 2014).

GBIF (2012) *Global Biodiversity Information Facility: biodiversity occurrence data published by Biologiezentrum Linz, Bundesamt fuer Naturschutz/Netzwerk Phytodiversitaet Deutschland*; *Hatikka Observation Data Gateway; IPK Genebank*; *Natural History Museum, Vienna - Herbarium WU*; *Naturhistorisches Museum Mainz, Botanical Collection; The Erysiphales Collection at the Botanische Staatssammlung München; The Fungal Collection at the Senckenberg Museum für Naturkunde Görlitz* & *Tiroler Landesmuseum Ferdinandeum.* Available at: http://gbif.org/occurrence (last accessed 29 October 2012).

GENESYS (2014) *Global Portal on Plant Genetic Resources*: *plant genetic resources accession level data*. Available at: http://www.genesys-pgr.org (last accessed 28 October 2014).

Gorczynski, L. (1922) The calculation of the degree of continentality. *Monthly Weather Review*, **50**, 370.

Holubec, V., Hauptvogel, P., Paprštein, F., Podyma, W., Ševčíková, M. & Vymyslický, T. (2010) Results of projects on collecting, mapping, monitoring, and conserving of plant genetic resources 1990 – 2008. *Czech Journal of Genetics and Plant Breeding*, **46**, S2–8.

JRC (2012) *Photovoltaic geographical information system*. European Joint Research Council. Available at: http://re.jrc.ec.europa.eu/pvgis (accessed 11 September 2012).

Mikolášková, K. (2009) A regression evaluation of thermal continentality. *Geografie – Sborník České Geografické Společnosti*, **4**, 350–362.

MZM (2013) *Database of the herbarium records in the Czech Republic*. Moravské Zemské Museum (Moravian Museum). Available at: http://puvodni.mzm.cz/Botanika/CS/uvod.html (last accessed 10 October 2013).

Panagos, P. (2006) The European Soil Database. *GEO:connexion*, **5**, 32–33.

Parra-Quijano, M., Iriondo, J.M., Frese, L. & Torres E. (2012) Spatial and ecogeographic approaches for selecting genetic reserves in Europe. *Agrobiodiversity Conservation: Securing the Diversity of Crop Wild Relatives and Landraces* (ed. by N. Maxted), pp. 20–28. CAB International, Wallingford, UK.

Pearce, J. & Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225–245.

Pearson, R.G. (2010) Species' distribution modeling for conservation educators and practicioners. *Lessons in Conservation*, **3**, 54-89.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Phillips, S.J. & Dudik, M. (2008) Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.

Phillips S.J., Dudik M., Elith J., Graham C.H., Lehmann A., Leathwick J. & Ferrier S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181–197.

Phillips, S.J., Dudik, M. & Schapire, R.E. (2004) A maximum entropy approach to species distribution modeling. *Proceedings of the 21st International Conference on Machine Learning* (ed. by C. Brodley), pp. 655–662. ACM Press, New York, NY, USA.

Šúri, M., Huld, T.A., Dunlop, E.D. & Ossenbrink, H.A. (2007) Potential of solar electricity generation in the European Union member states and candidate countries. *Solar Energy*, **81**, 1295–1305.

Syfert M.M., Smith M.J. & Coomes D.A. (2013) The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PLoS ONE*, **8**, e55158.

van Proosdij, A.S.J., Sosef, M.S.M., Wieringa, J.J. & Raes, N. (2016) Minimum required number of specimen records to develop accurate species distribution models. *Ecography*, **39**, 542–552.

Appendix S3: Generalized ecogeographic land characterization for the Czech Republic

An ecogeographic land characterization (ELC) map (Figure S3) was constructed according to Parra-Quijano *et al*. (2011). Values of edaphic, geophysical and bioclimatic variables (Table S3.1) were extracted for each cell on a grid of 30 arc-second resolution, with extent covering the Czech Republic. Two-step cluster analysis of environmental variables in SPSS (IBM Corp., 2012) was used to group cells with similar bioclimatic, geophysical and edaphic properties. A Bayesian Information Criterion automatically defined the number of clusters (two bioclimatic, two geophysical and six edaphic) which, in various combinations, define 22 ecogeographic categories (Table S3.2). These categories are represented as ecogeographic zones (EGZs) on an ELC map (Figure S3). The ELC map was re-sampled to match the resolution and extent of data in the UTM 33N projection by taking the EGZ from the mid-point of every cell in the UTM projection.

ELC maps typically focus on a limited number of taxa with similar evolutionary histories. However, our ELC map is generalized, attempting to represent the adaptive scenarios for all 204 priority CWR species. Parra-Quijano *et al*. (2011) found that an ELC map generated for Peninsular Spain was an appropriate representation of the adaptive scenario for eight different plant species (in a range of families, and including crops, landraces and wild taxa). Each species was not simply proportionally distributed across ecogeographic categories (the null expectation), and larger seeds (a proxy for adaptation) were associated with the overrepresented categories. Further support for the validity of a generalized map is provided by the fact that exotic plant species tend to invade hotspots of native diversity (Stohlgren *et al*., 2003), which suggests that similar environmental factors influence the growth and success of a wide range of plant species, regardless of taxonomic identity or geographical origin.

To build the generalized ELC map, we selected environmental variables hypothesized to influence both the distribution and adaptation (phenotypic and genotypic) of all focal taxa. Thus, a large number of variables were needed to cover all scenarios (Table S3.1), although we acknowledge this may introduce redundancy and irrelevance of some variables for some taxa.

References

FAO/IIASA/ISRIC/ISSCAS/JRC (2012) *Harmonised World Soil Database v1.2*. FAO, Rome, Italy and IIASA, Laxenburg, Austria.

IBM Corp. (2012) *IBM SPSS Statistics for Windows v21.0*. IBM Corp., Armonk, NY, USA.

Parra-Quijano, M., Iriondo, J.M. & Torres, E. (2011) Ecogeographical land characterization maps as a tool for assessing plant adaptation and their implications in agrobiodiversity studies. *Genetic Resources and Crop Evolution*, **59**, 205–217.

JRC (2012) *Photovoltaic geographical information system*. European Joint Research Council. Available at: http://re.jrc.ec.europa.eu/pvgis (accessed 11 September 2012).

Stohlgren, T.J., Barnett, D.T. & Kartesz, J.T. (2003) The rich get richer: patterns of plant invasions in the United States. *Frontiers in Ecology and the Environment*, **1**, 11.

Table S3.1 Ecogeographic variables used for construction of the generalized ecogeographic land characterization (ELC) map for priority CWR in the Czech Republic. FAO – Food and Agriculture Organization of the United Nations; IIASA – International Institute for Applied Systems Analysis; ISRIC – World Soil Information; ISSCAS – Institute of Soil Science, Chinese Academy of Sciences; JRC – Joint Research Centre of the European Commission; Max. – maximum; Min. – minimum; Temp – temperature.

BIOCLIMATIC VARIABLES		**Data Type**	**Resolution**	**Source**
BIO1	Annual Mean Temp. (°C)	Continuous	30 arc-second	www.worldclim.org/tiles.php
BIO5	Max. Temp. of Warmest Month (°C)	Continuous	30 arc-second	www.worldclim.org/tiles.php
BIO6	Min. Temp. of Coldest Month (°C)	Continuous	30 arc-second	www.worldclim.org/tiles.php
BIO12	Annual Precipitation (mm)	Continuous	30 arc-second	www.worldclim.org/tiles.php
BIO13	Precipitation of Wettest Month (mm)	Continuous	30 arc-second	www.worldclim.org/tiles.php
BIO14	Precipitation of Driest Month (mm)	Continuous	30 arc-second	www.worldclim.org/tiles.php

GEOPHYSICAL VARIABLES		**Data Type**	**Resolution**	**Source**
Longitude	Longitude (decimal degrees)	Continuous	—	—
Latitude	Latitude (decimal degrees)	Continuous	—	—
Altitude	Elevation (m above sea level)	Continuous	30 arc-second	www.diva-gis.org/Data
Slope	Slope of land surface (degrees)	Continuous	30 arc-second	www.diva-gis.org/Data
Irradiation	Mean solar irradiation of brightest quarter (May-Jul) 2006-11 (Wh/m^2)	Continuous	1.5 arc-minute	JRC (2012)

EDAPHIC VARIABLES		**Data Type**	**Resolution**	**Source**
TYPE	Soil Type	Categorical	—	FAO/IIASA/ISRIC/ISSCAS/JRC (2012)
DRAINAGE	Drainage class	Categorical	—	FAO/IIASA/ISRIC/ISSCAS/JRC (2012)
T_TEXTURE	Topsoil Texture	Categorical	—	FAO/IIASA/ISRIC/ISSCAS/JRC (2012)
T_OC	Topsoil Organic Carbon (% weight)	Continuous	30 arc-second	FAO/IIASA/ISRIC/ISSCAS/JRC (2012)

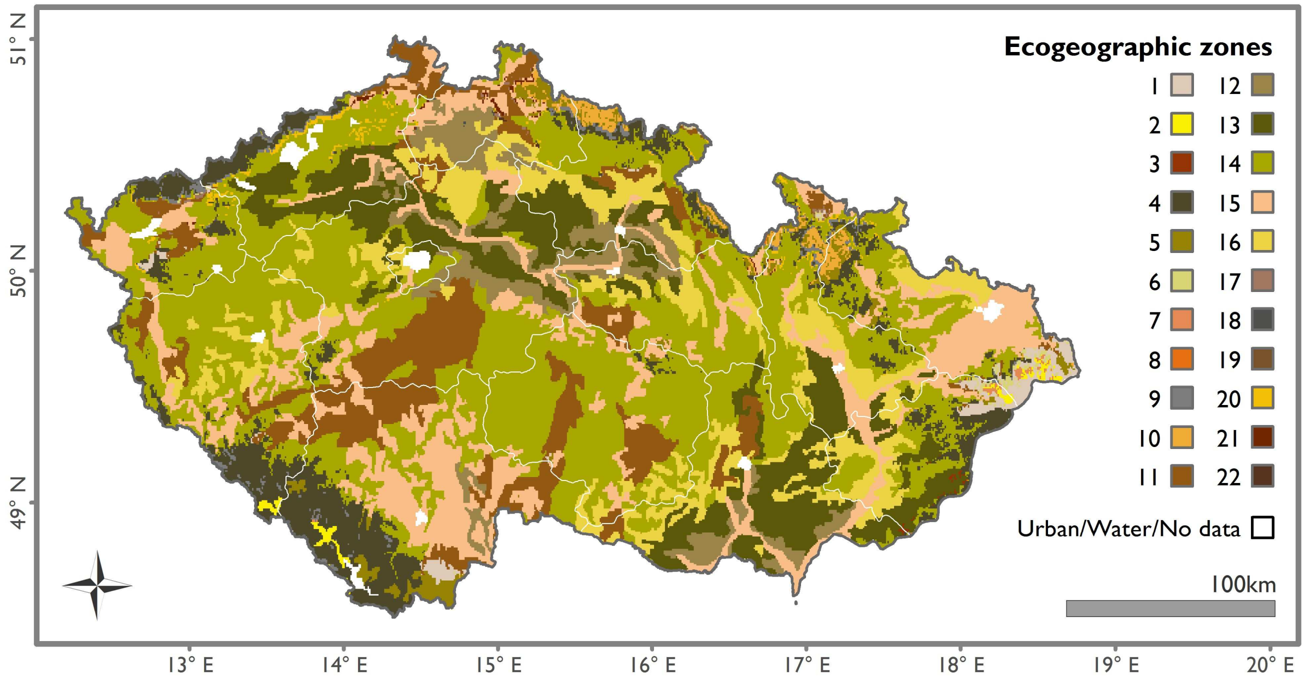


Figure S3 Generalized ecogeographic land characterization map for 204 priority CWR species in the Czech Republic, showing the 22 discrete ecogeographic zones (EGZs) identified by our analysis. Projection: Transverse Mercator 33N.

Table S3.2 Twenty-two distinct ecogeographic zones (EGZs) were defined in the Czech Republic, based on a combination of two bioclimatic (B), two geophysical (G) and six edaphic (E) clusters. For each EGZ, means (and standard deviations) of continuous variables are presented. For categorical variables, modal values are listed. Units as in Table S3.1.

EGZ	Cells	%	B	G	E	BIO1	BIO5	BIO6	BIO12	BIO13	BIO14	Elevation	Slope	Solar Radiation	Soil Type	Drainage	Texture	Organic C
1	1,791	0.98	1	1	1	6.3 (0.7)	21.4 (1.2)	-7.3 (0.7)	883.0 (68.4)	123.7 (13.8)	45.4 (4.9)	653 (120)	4.3 (2.6)	1042 (17)	Cambisols	Well	Coarse	0.7 (0.1)
2	571	0.31	1	1	2	5.4 (0.9)	20.1 (1.2)	-7.9 (0.7)	1035.7 (82.6)	131.0 (11.1)	59.5 (8.6)	866 (150)	3.3 (2.6)	1050 (9)	Cambisols	Very Poor	Fine	2.9 (2.2)
3	98	0.05	1	1	3	6.6 (0.8)	21.6 (1.4)	-7.0 (0.4)	785.2 (28.5)	108.2 (5.5)	40.7 (2.3)	595 (93)	4.6 (2.3)	1074 (30)	Cambisols	Moderately Well	Fine	1.3 (0.3)
4	13,388	7.35	1	1	4	5.8 (0.8)	20.3 (1.4)	-7.0 (0.6)	881.1 (125.9)	109.7 (11.9)	50.7 (10.8)	755 (146)	3.5 (2.2)	1035 (25)	Cambisols	Moderately Well	Medium	1.4 (0.2)
5	2,525	1.39	1	1	5	5.5 (1.1)	19.8 (2.1)	-7.3 (0.6)	844.6 (95.3)	110.5 (11.1)	46.0 (8.8)	784 (159)	3.2 (2.4)	1040 (26)	Cambisols	Poor	Medium	1.0 (0.1)
6	344	0.19	1	2	1	5.9 (0.8)	20.7 (1.3)	-7.7 (0.7)	896.5 (94.9)	127.4 (15.9)	44.7 (6.2)	703 (141)	10.3 (3.5)	1030 (15)	Cambisols	Well	Coarse	0.7 (0.2)
7	164	0.09	1	2	2	5.1 (0.7)	19.7 (1.0)	-8.4 (0.5)	1005.6 (65.9)	142.6 (8.6)	52.0 (5.3)	863 (126)	10.2 (3.2)	1040 (6)	Cambisols	Poor	Coarse	0.9 (0.5)
8	27	0.01	1	2	3	3.9 (1.5)	17.0 (2.5)	-8.2 (0.8)	857.8 (98.1)	113.6 (9.1)	44.7 (9.0)	1006 (238)	11.6 (4.2)	1027 (25)	Leptosols	Imperfectly	Fine	1.9 (0.4)
9	1,200	0.66	1	2	4	5.1 (0.7)	19.0 (1.1)	-7.3 (0.8)	852.8 (144.4)	106.0 (13.0)	47.9 (12.1)	820 (152)	10.2 (2.9)	1019 (23)	Cambisols	Moderately Well	Medium	1.4 (0.1)
10	1,450	0.80	1	2	5	4.0 (1.0)	17.0 (1.6)	-8.2 (0.6)	840.4 (84.6)	110.0 (10.1)	44.4 (8.3)	988 (179)	10.3 (3.8)	1018 (10)	Cambisols	Poor	Medium	1.0 (0.1)
11	18,345	10.08	2	1	1	7.4 (0.6)	22.3 (1.0)	-5.5 (0.6)	650.8 (61.9)	87.3 (7.1)	33.0 (5.6)	475 (93)	1.8 (1.6)	1033 (20)	Cambisols	Well	Coarse	0.7 (0.1)
12	9,320	5.12	2	1	2	8.3 (0.6)	23.5 (1.0)	-4.9 (0.5)	569.2 (47.1)	79.2 (6.3)	26.1 (3.9)	275 (79)	0.8 (1.2)	1038 (30)	Arenosols	Somewhat Excess	Coarse	2.3 (4.5)
13	20,907	11.48	2	1	3	8.4 (0.6)	23.6 (1.0)	-5.0 (0.6)	575.8 (55.1)	80.7 (7.7)	26.5 (4.0)	261 (67)	1.1 (1.4)	1048 (32)	Chernozem	Moderately Well	Fine	1.7 (0.5)
14	66,360	36.45	2	1	4	7.1 (0.6)	22.0 (1.0)	-5.8 (0.6)	657.1 (60.1)	89.2 (7.8)	33.0 (5.3)	482 (108)	2.0 (1.8)	1033 (20)	Cambisols	Moderately Well	Medium	1.4 (0.1)
15	26,589	14.61	2	1	5	7.8 (0.7)	22.9 (1.1)	-5.4 (0.6)	659.2 (70.5)	89.9 (10.0)	32.8 (5.9)	391 (131)	1.0 (1.4)	1035 (23)	Luvisols	Poor	Medium	1.0 (0.7)
16	17,648	9.69	2	1	6	8.0 (0.5)	23.2 (0.8)	-5.4 (0.5)	605.6 (48.6)	84.5 (6.8)	28.1 (3.8)	325 (72)	1.2 (1.3)	1041 (26)	Luvisols	Moderately Well	Medium	0.8 (0.1)
17	176	0.10	2	2	1	6.5 (0.5)	20.6 (0.7)	-5.9 (0.6)	666.2 (42.7)	86.9 (7.6)	34.7 (4.1)	549 (89)	10.5 (2.8)	1003 (12)	Cambisols	Well	Coarse	0.7 (0.1)
18	14	0.01	2	2	2	6.7 (0.7)	20.8 (1.0)	-5.6 (0.6)	602.0 (31.3)	78.7 (3.4)	30.1 (2.6)	505 (104)	10.1 (3.2)	1005 (3)	Podzols	Somewhat Excess	Coarse	2.2 (0.0)
19	15	0.01	2	2	3	6.5 (0.4)	20.6 (0.6)	-6.0 (0.6)	603.1 (34.9)	80.6 (5.8)	29.5 (1.8)	441 (70)	11.5 (1.7)	1007 (2)	Cambisols	Moderately Well	Fine	1.8 (0.7)
20	862	0.47	2	2	4	6.4 (0.6)	20.6 (0.9)	-5.9 (0.7)	660.6 (52.5)	87.5 (7.8)	33.9 (4.5)	513 (124)	11.0 (2.7)	1004 (14)	Cambisols	Moderately Well	Medium	1.3 (0.2)
21	226	0.12	2	2	5	6.2 (0.6)	20.2 (0.9)	-6.1 (0.7)	652.1 (38.3)	84.6 (6.3)	33.8 (2.7)	564 (117)	10.5 (3.5)	1004 (10)	Cambisols	Poor	Medium	1.0 (0.1)
22	18	0.01	2	2	6	6.7 (0.3)	20.9 (0.4)	-6.1 (0.3)	620.6 (14.0)	81.9 (2.5)	30.9 (1.0)	500 (55)	11.1 (1.2)	1011 (5)	Luvisols	Moderately Well	Medium	0.8 (0.0)
	182,038	100.00																

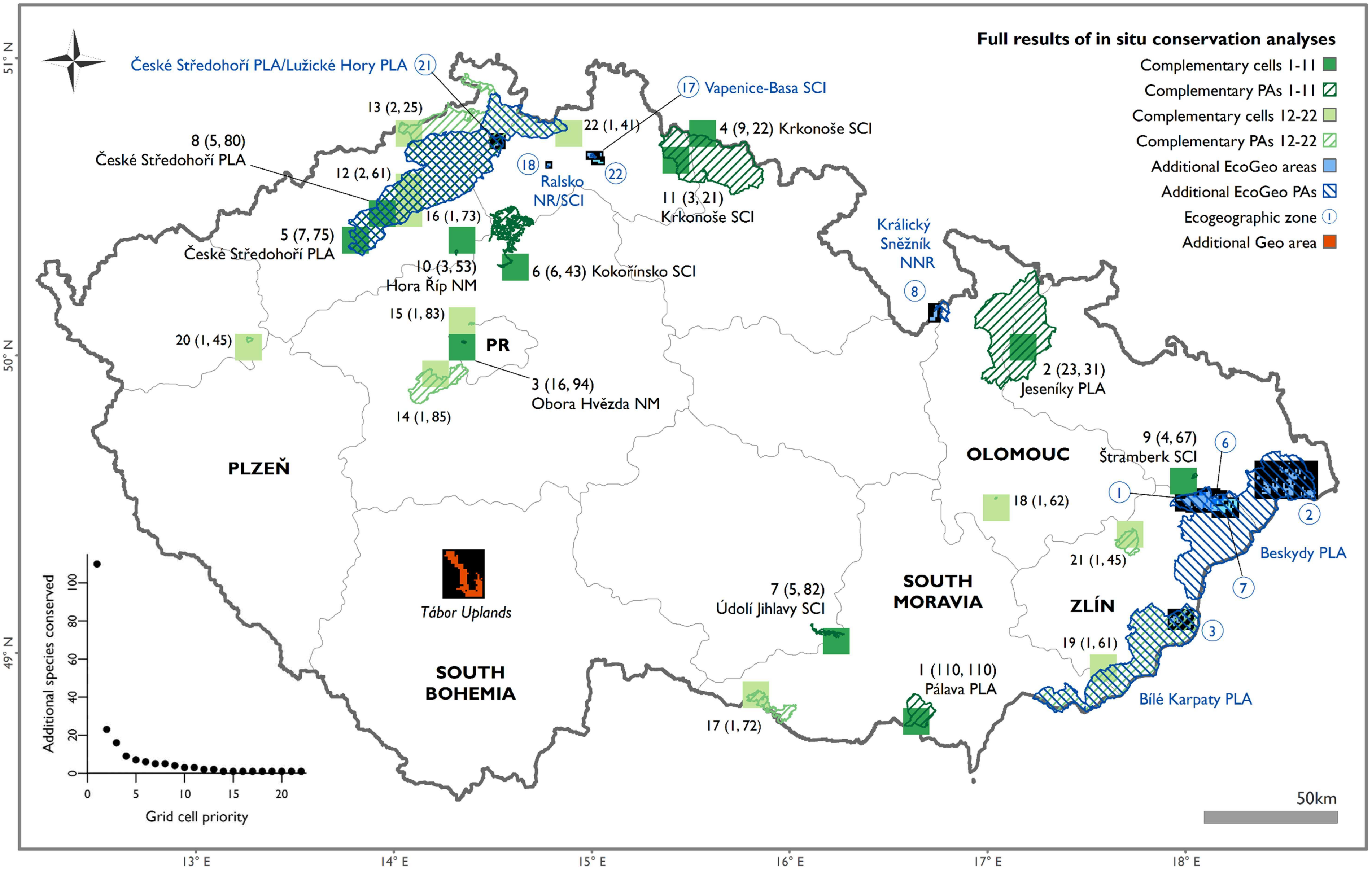
Appendix S4: Full results of in situ conservation analyses for CWR in the Czech Republic

Figure S4 Full results of in situ conservation analyses. Priority areas are (in order): the top eleven cells from complementarity analysis (dark green), the richest areas of omitted ecogeographic zones (EcoGeo areas; blue) and an additional species-rich area to fill a spatial gap (Geo area; orange). For the complementary network, numbers outside parentheses = priority rank of complementary cells; first number in parentheses = the number of additional species conserved; second number in parentheses = total number of species recorded in that cell. For the additional EcoGeo areas, the number of each ecogeographic zone is given in a blue circle. The protected area (PA) with the largest area of overlap with each priority area is also shown (hatched).

Complementary cells 12-22 are included for completeness, but are not priority conservation targets. Overlapping PAs are 12 – České Středohoří PLA; 13 – Labské Pískovce PLA; 14 – Český Kras PLA; 15 – Údolí Únětického Potoka NR; 16 – České Středohoří PLA; 17 – Podyjí NP; 18 – Vápenice NM; 19 – Bílé Karpaty PLA; 20 – Vladař SCI; 21 – Hostýnské Vrchy SCI; 22 – Lužické Hory PLA.

Abbreviations: NM – Nature Monument; NNR – National Nature Reserve; NP – National Park; NR – Nature Reserve; PLA – Protected Landscape Area; PR – Prague; SCI – Natura 2000 Site of Community Importance. Projection: Transverse Mercator 33N.

Inset: Addition of grid cells to the complementary network yields diminishing returns. From the twelfth cell, each additional cell only includes two priority species not already contained in the network. From the fourteenth, this drops to a single species.


Full results of in situ conservation analyses
Complementary cells 1-11 🟩
Complementary PAs 1-11 ▨
Complementary cells 12-22 🟩
Complementary PAs 12-22 ▨
Additional EcoGeo areas 🟦
Additional EcoGeo PAs ▨
Ecogeographic zone ①
Additional Geo area 🟥
České Středohoří PLA/Lužické Hory PLA ㉑
13 (2, 25)
8 (5, 80)
České Středohoří PLA
12 (2, 61)
16 (1, 73)
5 (7, 75)
České Středohoří PLA
10 (3, 53)
Hora Říp NM
6 (6, 43) Kokořínsko SCI
⑱ Ralsko NR/SCI
22 (1, 41)
㉒
⑰ Vapenice-Basa SCI
4 (9, 22) Krkonoše SCI
11 (3, 21)
Krkonoše SCI
Králický Sněžník NNR
⑧
2 (23, 31)
Jeseníky PLA
15 (1, 83)
PR
20 (1, 45)
3 (16, 94)
Obora Hvězda NM
14 (1, 85)
PLZEŇ
OLOMOUC
9 (4, 67)
Štramberk SCI
⑥
①
18 (1, 62)
②
21 (1, 45)
Beskydy PLA
⑦
Tábor Uplands
SOUTH BOHEMIA
7 (5, 82)
Údolí Jihlavy SCI
SOUTH MORAVIA
ZLÍN
③
19 (1, 61)
1 (110, 110)
Pálava PLA
Bílé Karpaty PLA
17 (1, 72)
Additional species conserved
0
20
40
60
80
100
0
5
10
15
20
Grid cell priority
50km
51° N
50° N
49° N
13° E
14° E
15° E
16° E
17° E
18° E

Appendix S5: Genetic representativeness (by proxy) of existing ex situ CWR collections

Genetic representativeness of existing ex situ collections of Czech CWR was assessed using two proxies for genetic variation: geographic and ecogeographic (GR and ER; see main text for explanation of these proxies). These two proxies were significantly correlated but imperfectly so ($\log_{10}$ GR against $\log_{10}$ ER $r = 0.655$, df = 64, $p < 0.001$), thus both were analysed. A further reason for considering both proxies is the trade-off between simplicity and realism: a geographic proxy is much simpler to implement (requiring fewer data) but the ecogeographic proxy is arguably a better representation of genetic variation in that it more explicitly considers environmental variation that is assumed to drive genetic variation.

First priorities for collection of material for ex situ conservation are the 134 species without any known accessions (see Appendix S4). Second priorities are species with existing accessions that require supplementary collection of material: (a) four species whose accessions are from unknown locations (*Festuca breviplia*, *Medicago minima*, *Poa remota* and *Vicia sativa*) and (b) 66 species whose existing accessions are likely to be poorly representative of genetic diversity (Table S6). For these species, the top 10 priorities based on each of GR and ER are highlighted in bold.

Table S5 Genetic representativeness, by proxy, of priority CWR with existing georeferenced accessions. Pr – priority for supplementary collection, based on sum of ranks; AccT – total number of accessions of wild Czech origin (with unique accession number, but including accessions not georeferenced); AccG – georeferenced accessions (including spatial duplicates); GCT – total geographic coverage of species (based on SDM) in number of cells; GCA – accession geographic coverage (circular area of 20 km diameter around accession locations) in number of cells; GR – geographic representativeness; ECT – total ecogeographic coverage (ecogeographic zones covered by SDM); ECA – accession geographic coverage (ecogeographic zones from which germplasm has been collected); ER – ecogeographic representativeness; Sum – sum of individual ranks for GR and ER.

Pr	Species	AccT	AccG	GCT	GCA	GR (%)	Rank	ECT	ECA	ER (%)	Rank	Sum
1	*Vicia sylvatica*	1	1	**103,422**	**417**	**0.40**	**2**	**22**	**1**	**4.5**	**1**	3
2	*Vicia sepium**	4	3	**166,692**	**942**	**0.57**	**7**	**22**	**1**	**4.5**	**1**	8
3	*Elymus caninus*	2	2	**140,228**	**833**	**0.59**	**8**	**22**	**1**	**4.5**	**1**	9
4	*Festuca altissima*	1	1	**85,238**	**538**	**0.63**	**10**	**22**	**1**	**4.5**	**1**	11
5	*Malva alcea*	1	1	**78,662**	**338**	**0.43**	**4**	**21**	**1**	**4.8**	**8**	12
6	*Raphanus raphanistrum*	1	1	**81,692**	**346**	**0.42**	**3**	**20**	**1**	**5.0**	**10**	13
7	*Allium ursinum*	1	1	67,882	479	0.71	13	**22**	**1**	**4.5**	**1**	14
8	*Festuca heterophylla*	1	1	**89,087**	**409**	**0.46**	**5**	**20**	**1**	**5.0**	**10**	15
8	*Lotus pedunculatus*	3	3	78,454	584	0.74	14	**22**	**1**	**4.5**	**1**	15
10	*Vicia dumetorum*	1	1	76,224	637	0.84	16	**22**	**1**	**4.5**	**1**	17
11	*Poa humilis*	2	1	**62,286**	**351**	**0.56**	**6**	19	1	5.3	13	19
12	*Festuca rubra*	6	4	**74,781**	**73**	**0.10**	**1**	22	2	9.1	21	22
13	*Poa supina*	5	4	65,951	931	1.41	22	**21**	**1**	**4.8**	**8**	30
14	*Melica ciliata*	1	1	40,719	516	1.27	20	19	1	5.3	13	33
15	*Vicia tetrasperma*	4	4	**125,484**	**760**	**0.61**	**9**	20	2	10.0	26	35
16	*Genista germanica*	2	2	97,314	648	0.67	12	21	2	9.5	24	36
17	*Humulus lupulus*	3	2	112,579	733	0.65	11	15	2	13.3	31	42
18	*Trifolium pratense*	3	3	66,709	673	1.01	19	21	2	9.5	24	43
19	*Melica picta*	1	1	30,796	258	0.84	16	9	1	11.1	28	44
20	*Dactylis polygama*	4	4	96,745	2,131	2.20	26	22	2	9.1	21	47
21	*Medicago prostrata*	1	1	17,216	218	1.27	20	9	1	11.1	28	48

Pr	Species	AccT	AccG	GCT	GCA	GR (%)	Rank	ECT	ECA	ER (%)	Rank	Sum
21	*Trifolium dubium*	7	6	142,007	1,112	0.78	15	22	3	13.6	33	48
21	*Trifolium rubens*	1	1	15,397	470	3.05	32	17	1	5.9	16	48
24	*Festuca filiformis*	4	4	60,852	1,198	1.97	24	20	2	10.0	26	50
25	*Koeleria pyramidata*	5	5	109,920	925	0.84	16	19	3	15.8	35	51
26	*Lotus maritimus*	1	1	21,731	698	3.21	33	11	1	9.1	21	54
26	*Poa chaixii*	3	3	22,590	1,019	4.51	44	**20**	**1**	**5.0**	**10**	54
26	*Vicia pannonica*	6	4	33,875	1,223	3.61	37	15	1	6.7	17	54
29	*Bromus inermis*	8	8	58,706	1,164	1.98	25	15	2	13.3	31	56
30	*Vicia villosa*	6	5	48,051	1,355	2.82	30	21	3	14.3	34	64
31	*Trifolium ochroleucon*	1	1	9,242	592	6.41	52	19	1	5.3	13	65
32	*Dorycnium germanicum*	1	1	9,368	327	3.49	36	8	1	12.5	30	66
33	*Vicia angustifolia*	15	15	105,515	2,676	2.54	28	22	4	18.2	41	69
34	*Festuca valesiaca*	3	3	26,126	1,482	5.67	50	13	1	7.7	20	70
34	*Vicia tenuifolia*	5	5	65,737	2,000	3.04	31	17	3	17.6	39	70
36	*Trifolium montanum*	7	7	127,143	2,028	1.60	23	22	5	22.7	50	73
36	*Vicia hirsuta*	14	14	112,079	2,472	2.21	27	19	4	21.1	46	73
38	*Melilotus dentatus*	2	2	9,679	395	4.08	39	6	1	16.7	36	75
39	*Poa badensis*	4	3	1,540	173	11.23	60	15	1	6.7	17	77
41	*Genista tinctoria*	24	24	123,578	4,649	3.76	38	22	4	18.2	41	79
40	*Medicago falcata*	11	11	69,926	2,328	3.33	34	16	3	18.8	44	78
42	*Phleum alpinum*	1	1	4,250	502	11.81	62	14	1	7.1	19	81
42	*Trifolium campestre*	32	31	133,931	3,569	2.66	29	20	5	25.0	52	81
42	*Vicia pisiformis*	7	6	38,669	1,658	4.29	42	17	3	17.6	39	81
45	*Melilotus officinalis*	26	26	118,033	3,929	3.33	34	22	5	22.7	50	84
46	*Securigera varia*	32	29	152,080	6,424	4.22	40	20	4	20.0	45	85
47	*Dorycnium herbaceum*	2	2	7,435	534	7.18	53	11	2	18.2	41	94
47	*Lotus borbasii*	1	1	3,079	303	9.84	58	6	1	16.7	36	94
47	*Trifolium arvense*	17	17	130,280	5,507	4.23	41	19	5	26.3	53	94
50	*Genista pilosa*	1	1	5,749	672	11.69	61	6	1	16.7	36	97
50	*Trifolium alpestre*	18	16	83,130	4,181	5.03	49	18	4	22.2	48	97
52	*Trifolium fragiferum*	6	4	25,146	1,120	4.45	43		4	30.8	55	98
53	*Melilotus albus*	25	23	129,894	6,366	4.90	47	22	6	27.3	54	101
53	*Trifolium medium*	32	32	148,750	7,025	4.72	45	22	7	31.8	56	101
53	*Medicago lupulina*	34	32	139,899	6,607	4.72	45	22	7	31.8	56	101
56	*Onobrychis arenaria*	10	2	9,797	793	8.09	56	9	2	22.2	48	104
57	*Vicia cracca*	29	29	153,899	8,975	5.83	51	22	7	31.8	56	107
58	*Anthyllis vulneraria*	20	18	139,841	6,892	4.93	48	22	8	36.4	62	110
58	*Daucus carota*	19	18	9,292	1,166	12.55	64	19	4	21.1	46	110
60	*Astragalus onobrychis*	3	2	9,395	750	7.98	55	6	2	33.3	59	114
61	*Astragalus glycyphyllos*	51	47	146,068	10,764	7.37	54	20	8	40.0	63	117
62	*Lavatera thuringiaca*	2	2	9,669	983	10.17	59	6	2	33.3	59	118
62	*Trifolium aureum*	21	20	46,111	3,766	8.17	57	20	7	35.0	61	118
64	*Astragalus cicer*	18	17	28,898	3,456	11.96	63	10	5	50.0	65	128
65	*Lactuca serriola*	133	128	96,093	23,237	24.18	65	18	8	44.4	64	129
66	*Festuca supina*	3	3	356	198	55.62	66	4	2	50.0	65	131
		719	672									

* The SDM of *V. sepium* was not robust according to our criteria in Appendix S2: it did not perform significantly better than random in any replicate runs. However, we used the SDM for *V. sepium* as an estimate of total geographic coverage because it adequately reflects the wide distribution of this species (and models never performed worse than random (AUC always > 0.545).