Published in Genetic Resources and Crop Evolution 2008
which should be used for any reference to this work

Spontaneous gene flow and population structure in wild and cultivated chicory, *Cichorium intybus* L.

L. P. Kiær · F. Felber · A. Flavell · R. Guadagnuolo · D. Guiatti · T. P. Hauser · A. M. Olivieri · I. Scotti · N. Syed · M. Vischi · C. van de Wiel · R. B. Jørgensen

Abstract Spontaneous gene flow between wild and cultivated chicory, *Cichorium intybus* L., may have implications for the genetic structure and evolution of populations and varieties. One aspect of this crop-wild gene flow is the dispersal of transgenes from genetically modified varieties, e.g. gene flow from GM chicory to natural chicory could have unwanted consequences. With the purpose to identify and quantify crop-wild gene flow in chicory, we analysed introgression in 19 wild chicory populations and 16 accessions of chicory varieties and landraces distributed across Northern, Central and Mediterranean Europe. The analysis used 281 AFLP markers and 75 SSAP markers giving a total of 356 polymorphic markers. Results from model based assignments with the program STRUCTURE indicated many incidents of recent gene flow. Gene flow was observed both between cultivars and wild populations, between landraces and wild populations, between different wild populations as well as between cultivars. Population structure visualized by distance-based clustering showed a North–South geographical structuring of the wild populations, and a general grouping of the cultivars corresponding to known origin. The results indicated, however, that the structuring between the two groups of wild and cultivated types was weak. As crop and wild recipients are genetically close and genes are transferred between the two types rather frequently, focus on mitigating crop-wild gene flow should be increased, before transgenic varieties are cultivated openly.

Keywords AFLP · GM co-existence · Gene dispersal · Introgression · Risk assessment · SSAP

L. P. Kiær · R. B. Jørgensen (✉)
Biosystems Department, Risø National Laboratory, DTU, Frederiksborgvej 399, 4000 Roskilde, Denmark
e-mail: rikke.bagger.jorgensen@risoe.dk

F. Felber
Institut de Biologie, Laboratoire de Botanique Évolutive, Université de Neuchâtel, Rue Emile-Argand 11, 2009 Neuchatel, Switzerland

A. Flavell · N. Syed
Division of Plant Sciences, University of Dundee at SCRI, Invergowrie, Dundee DD2 5DA, Scotland, UK

R. Guadagnuolo · D. Guiatti · A. M. Olivieri · M. Vischi
Dipartimento di Scienze Agrarie e Ambientali, University of Udine, Via delle Scienze 208, 33100 Udine, Italy

T. P. Hauser
Department of Ecology, University of Copenhagen, Rolighedsvej 21, 1958 Frederiksberg C, Denmark

I. Scotti
Institut National de la Recherche Agronomique, National Agricultural Research Institute, UMR ECOFOG Campus Agronomique, BP 709 Avenue de France, 97387 Kourou Cedex, France

C. van de Wiel
Plant Research International, P.O. Box 16, 6700 AA Wageningen, The Netherlands

Introduction

The wild and cultivated types of chicory, *Cichorium intybus* L., can be deliberately crossed (Kiers 2000), but apparently little is known about the spontaneous introgression between crop and wild chicory, and thus one of the potential drivers in the evolution of chicory is largely unstudied. Lately, demands for more knowledge about the spontaneous gene flow in chicory has been brought about by sanctioning of genetically modified (GM) chicory lines (C/NL/94/25; http://europa.eu.int/comm/environment/biotechnology/authorised_prod_1.htm) for breeding activities; at present the statutory environmental risk assessment of GM chicory is hampered by the poor knowledge on gene escape. Also after the GM chicory has been approved for cultivation, knowledge on gene flow is necessary to regulated co-existence of the GM cultivars that are grown in the same area as non-GM chicory. Lastly, also the practise when producing high purity chicory seed would benefit from information about intraspecific gene flow, as would strategies for conservation of chicory material for the future.

Therefore, we set out to evaluate if such spontaneous gene flow takes place in chicory, and if so to what extent. A crude approach to the investigation of introgression is to take close genetic relationships between populations and varieties as an indication of significant gene flow between them. However, if a wild population clusters with one or more cultivated populations, it is not in itself evidence of gene flow from an agricultural field to the wild habitat, but could also be due to the reciprocal process, namely the deliberate incorporation of wild material during breeding of cultivars. Likewise, the relationships between the cultivars of *C. intybus* could be influenced by breeders' deliberate introgression of traits among varieties (e.g. Ryder 1999; Kiers 2000). In other words, controlled crosses may complicate the investigation of spontaneous introgression. One possible way to overcome this is to evaluate the multilocus genotype of each individual separately against the allele distribution in the total set of individuals. Recently several methods have been presented that could perform this task, e.g. STRUCTURE (Pritchard et al. 2000).

Chicory (*Cichorium intybus* L., $2n = 18$) is a perennial herb belonging to the Asteraceae family (Kiers 2000). It is native to Europe and has been cultivated at least since Greek and Roman times (Simmonds 1976). When cultivation of chicory began is uncertain, but it was used in the Bronze Age, and around 50 AD chicory was registered by the Roman historian Plinius together with lettuce (Kiers 2000). Today numerous chicory varieties for different food and feed purposes are found. Four cultivar groups exist (Kiers 2000): (1) the root cultivars are grown for their large roots, from which inulin is extracted for industrial purposes or the roots are used for feed or coffee surrogate; (2) the Witloof cultivars, cultivated mainly in Central and North-Western Europe (e.g. Belgium and The Netherlands), are grown to produce etiolated apical buds, which are consumed as a salad or cooked; (3) the Pain de Sucre ($\sim$sugarloaf chicory) cultivars are cultivated mostly in Central and North-western Europe and are also eaten as salads; and (4) the Radicchio cultivars, which are cultivated in Southern Europe, predominantly in Northern Italy, have reddish leaves and are salad-types.

Chicory has a self-incompatibility system (Eenink 1981), which may break down in the cultivated forms as a result of selection. Insects, mainly bees, pollinate the brilliant blue flowers. The wild form of chicory has a wide distribution in Europe, North Africa and Central Asia, and has been naturalized in North America (Simmonds 1976). It is growing in well-drained habitats along roads, rivers and in disturbed places (Kiers 2000). Gene flow between cultivated and wild chicory might be likely due to the overlap in distribution areas (Van Cutsem et al. 2003), and the high rate of self-incompatibility of the species (Schoofs and de Langhe 1988). In nature, spontaneous hybridization between wild and cultivated chicory has been detected in a Danish population by Kiær et al. (2007) and Sørensen et al. (2007). In this population the crop-wild hybrids had a deviating morphology, and therefore they were easily recognized. Also, spontaneous hybridization with the closely related species, *C. endivia* L., is apparently possible (Rick 1953). Hybridization with other chicory species (e.g. *C. spinosum* L.) might also be possible (Gemeinholzer and Bachmann 2005), however, these species are considered rare along the north-south gradient, where we studied introgression.

To optimize our search for evidence of past and recent introgression in chicory using a large dataset of dominant genetic markers, we explored genetic

relationships between populations and cultivars using distance-based trees, and we used a model-based clustering technique that provided estimates for the relative proportions of ancestry of each individual in the sample. In our analysis of crop-wild gene flow, different scenarios of introgression could be possible: (1) If spontaneous introgression of crop genes into a wild population is a recent event, then crop genes would only be found in some individuals of the wild population. Should introgression be very extensive, the wild recipient and the donor cultivar might link closely in the distance-based trees. (2) If crop-wild introgression dates back in time, cultivar genes would be more evenly distributed among all wild individuals of a recipient population. The wild populations that have been most exposed to previous crop introgression, could have close relations to the donor cultivar in trees based on genetic distance. (3) In the cases where breeders have broadened the genetic base of their cultivars by crossing with wild material, we would expect to see an even distribution of wild genetic material in all the individuals of a given variety, and most likely distantly related varieties would represent different amounts and representation of wild genes. In the distance-based trees this might be reflected by relatively close links between the cultivated variety and its wild donor population. (4) The spontaneous introgression from wild to cultivated types would leave different footprints, as the introgression would be sporadic within a given variety—at least if we are dealing with recent introgression. In dendrograms based on genetic differences, the relations between cultivars and between wild types and cultivars would probably not be affected by this sporadic transfer of wild material.

In the present study of the genetic structure of cultivated populations, we might expect to see one or the other of the following two patterns, depending on the breeding history of chicory: (5) If domestication of chicory took place several times, we expect to find several well separated groups of chicory cultivars in the distance-based trees, as well as when we force the formation of groups by the model-based clustering. (6) If domestication was a single event the model-based clustering should produce a group of wild chicory populations and a group of cultivars, when forced to form two groups.

Through a comprehensive molecular analysis of wild populations, landraces and cultivars representing all of the described cultivar groups of chicory, we set out to evaluate, whether the scenarios described above could be confirmed.

Materials and methods

Plant material

Fresh leaves were sampled from 19 wild chicory populations distributed across Northern, Central and Mediterranean Europe (Table 1, Fig. 1). Two of the wild populations derived from gene bank seeds (see Table 1), the remaining accessions were collected in this project. Additionally, we obtained seeds of 12 chicory varieties/lines from seed companies and gene bank accessions, and we collected seeds of 4 Radicchio landraces from small-scale farms in Northern Italy, where many farmers maintain their own private gene pools (Table 2). Fresh leaves were obtained from seed populations through germination in soil-trays under identical greenhouse conditions and random harvest of the seedlings. Numbers of seedlings analyzed are given in Tables 1 and 2.

Genetic marker methods

The sampled individuals were genotyped with two sets of AFLP markers and one set of SSAP markers, each set applying four combinations of selective primers (see Table 3 and below). Each marker set was applied in a separate laboratory (Table 3), but all analyses were based on the same set of DNA extractions to ensure identical starting material. A few populations were genotyped in two laboratories only.

DNA extraction

DNA was extracted following the CTAB-based procedure of Doyle and Doyle (1987) with minor modifications: active carbon and mercaptoethanol was added to the CTAB buffer before extraction (5 mg and 3 µl/ml, respectively), samples were treated with RNAse before (and not after) isopropanol precipitation, pellet isolation and resuspension following precipitation was omitted, the procedure

Table 1 The studied wild populations of chicory, including population abbreviation, country of origin, sample size and expected heterozygosity, Hj (with Std. Err.)

Country	Population	Abbreviations	*N*	Hj
Denmark	Borup	BRP	15	0.22273 (0.01209)
	Dråby	DBY	15	0.24736 (0.00956)
	Eggeslev	EGG	14	0.28087 (0.00933)
	Nyborg	NYG	15	0.32856 (0.00859)
	Odden	ODD	15	0.28022 (0.00981)
	Roskilde	ROS	13	0.29611 (0.01233)
	Svenstrup	SVP	15	0.29195 (0.00931)
France	Villedubert[a]	VILL	15	0.32039 (0.00947)
Germany	Friedrichshafen	FHN	14	0.33131 (0.00883)
	Koblenz	KOB	15	0.3474 (0.00928)
	Owingen	OWN	13	0.3084 (0.00915)
	Tübingen	TBN	15	0.34798 (0.00865)
	Weil der Stadt	WST	15	0.30091 (0.00916)
Italy	Amaro	AMR	10	0.34678 (0.01482)
	Mossano	MOS	15	0.33057 (0.01557)
	San Dorligo della Valle	SAN	5	0.35112 (0.0152)
	Subit[a]	SBT	15	0.2382 (0.00957)
	Udine	UDN	14	0.35476 (0.01447)
Switzerland	Fribourg	FRIB	10	0.31029 (0.01618)
Average				0.30715

[a] Gene bank accessions obtained from the Institut für Pflanzengenetik und Kulturpflanzenforschung, Gatersleben, Germany (IPK)

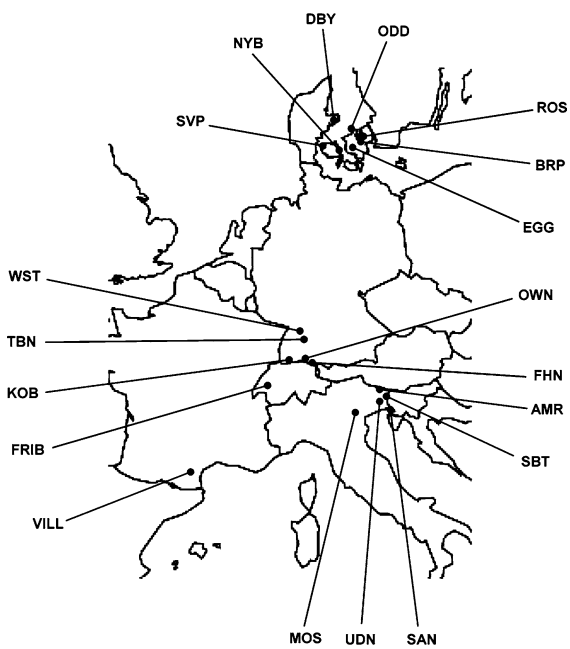

Fig. 1 Distribution of wild populations across Europe (see Table 1 for population name legend)

was halted following first resuspension in TE, and the samples were stored at 5°C for at least 24 h before storage at −20°C. The samples analyzed with SSAP markers were taken through another step of purification using the Qiagen DNeasy kit on the first DNA extractions.

AFLP profiling

The AFLP procedure used with primer set I was performed according to Vos et al. (1995) with the following modifications: digestion of DNA and ligation of adaptors for the AFLP was done with modifications described by Johannessen et al. (2002), and template DNA was amplified twice with modifications described by Shim and Jørgensen (2000). The amplified DNA samples were mixed with 98% formamide loading buffer, heated at 96°C for 5 min, and quickly cooled on ice. The products for marker set I were loaded for separation of DNA fragments on a 5% denaturing polyacrylamide long-range gel on a LI-COR® 4200 series Automated DNA Sequencer.

Table 2 The studied cultivars and landraces of chicory, including population abbreviations, cultivar type, sample size and expected heterozygosity, Hj, with standard errors (following Lynch and Milligan 1994)

Type	Cultivar	Abbreviations	*N*	Hj
Radicchio	Treviso 1	TREV1	15	0.28577 (0.01031)
	Treviso 2[a]	TREV2	15	0.28667 (0.0096)
	Verona	VERO	15	0.31762 (0.00898)
	Castelfranco	CAST	15	0.29565 (0.00971)
	Chioggia 1	CHIO1	15	0.32358 (0.00955)
	Chioggia 2[a]	CHIO2	15	0.27177 (0.00987)
	Chioggia F1[a]	CHIOH	15	0.31682 (0.01037)
	Catalogna[a]	CALO	14	0.22952 (0.01021)
	Frastagliata			
Landraces	Treviso	LR-TREV	14	0.29017 (0.0105)
	Verona	LR-VERO	14	0.28146 (0.01153)
	Castelfranco	LR-CAST	14	0.25803 (0.0106)
	Chioggia	LR-CHIO	14	0.30702 (0.0107)
Pain de Sucre	Pain de Sucre	SUCR	15	0.22965 (0.01013)
Witloof	Brüsseler Witloof[b]	BRÜS	12	0.29165 (0.00947)
Root	Cassel	CASL	15	0.34087 (0.00855)
	Ventiva[b]	VENT	15	0.26172 (0.01113)
Average				0.28675

[a] Cultivars not used in the three population subsets (DK, I and DECHF)

[b] Gene bank accessions obtained from the Institut für Pflanzengenetik und Kulturpflanzenforschung, Gatersleben, Germany (IPKG)

The AFLP procedure used with primer set II was performed according to Vos et al. (1995), modified by Bonin et al. (2004), and the products for marker set II were run on a denaturing 5% polyacrylamide gel using an automated sequencer (ABI 377; Applied Biosystems).

SSAP marker analysis

The SSAP profiling protocol was as described in Syed and Flavell (2007), using a novel primer Chi-46 (ATGGAGGCTGGAAATCAACAC), identified from the long terminal repeat (LTR) of a chicory Ty1-*copia* group LTR retrotransposon by the method of Syed et al. (2005). For pre-amplification *Pst*I and *Mse*I were used, and for selective amplification, the *Mse*I adapter primer, with two selective bases and the $^{\gamma 33}$P-labelled Chi-46 primer with one selective base were used. Four combinations of selective primer nucleotides were used (Table 3). SSAP reaction products were separated by electrophoresis on 6% denaturing polyacrylamide gels. The fixed, dried gels

Table 3 Selective nucleotides applied for the AFLP primers (*Eco*RI, *Mse*I, and *Pst*I) of marker sets I (Risø National Laboratory, Denmark) and II (University of Neuchâtel, Switzerland) and for the SSAP primers of marker set III (Universities of Dundee and Udine)

Marker set	Primer	Primer
Set I	E-ACC	M-CAA
	E-ACG	M-CAA
	E-CAC	M-AGG
	E-CAG	M-AGG
Set II	E-AGA	P-CAG
	E-AGT	P-CAG
	E-ATC	P-CAG
	E-ATG	P-CTA
Set III	Chi46[a]-C	M-AC
	Chi46[a]-T	M-AG
	Chi46[a]-C	M-CG
	Chi46[a]-C	M-CT

[a] Selective bases added to the 3′ ends of primers are shown

M = *Mse*I adapter primer

E = *Eco*RI adapter primer

P = *Pst*I adapter primer

were exposed to an autoradiography film for 5 days and then developed.

Data analysis

Marker scoring

Clear and reproducible markers were identified (approx. 100–1000 bp.) and scored manually. Markers were discarded if their frequency among all individuals fell outside the range 5–95%, resulting in a complete set of 356 polymorphic markers derived from data set I (90 markers), II (191 markers), and III (75 markers). All analyses were performed on the combined data set except for a few populations that were only genotyped with two of the data sets.

Genetic variation

The expected heterozygosity was calculated for each population based on estimates of their true allele frequencies, assuming Hardy–Weinberg proportions (Zhivotovsky 1999). The distribution of genetic variation within and among populations was assessed with analyses of molecular variance (AMOVA), using "RFLP bands" as haplotype input for the software ARLEQUIN (Excoffier et al. 1992). Three population subsets, each consisting of wild and cultivated populations, were selected and analysed separately in order to differentiate between three existing cultivation scenarios in various regions of European: low cultivation intensity in Denmark (DK), the common cultivation of Witloof, root chicory, and Pain de Sucre in Central and North-western Europe (DECHF), and the intensive Radicchio cultivation and seed propagation in Northern Italy (IT). The DK population set included all collected wild populations from Denmark, as well as the cultivars BRUS, VENT, CASL, TREV1, VERO, and CHIO1. The DECHF population set included all available wild populations from Germany (DE), France (F) and Switzerland (CH), as well as SUCR, BRUS, VENT, CASL, TREV1, VERO, and CHIO1. The IT population set included all wild populations from Italy, as well as the cultivars TREV1, VERO, CAST and CHIO1, and the four Radicchio landraces. Only a fraction of all the cultivars were represented in each subset; but within each set we included cultivars of several cultivar-groups and especially those cultivars that might be likely gene donors during introgression in that area.

Distance-based clustering

An unrooted phylogenetic tree was obtained for each population subset based on Neighbour-Joining (Saitou and Nei 1987) and pair-wise distances between populations as proposed by Reynolds et al. (1983). We assumed that the outcrossing nature of chicory and the network of interbreeding among chicory cultivars (Kiers 2000) would result in rather close genetic relationships among populations. To incorporate this into the analysis, we produced reticulograms (Legendre and Makarenkov 2002) to infer the population clusters from the distance data. The algorithm implemented in the software T-REX (Makarenkov and Legendre 2004), iteratively infers alternative edges (reticulations) between the nodes of an existing tree. It continues until the loss of information that arises when multivariate distance data are collapsed into a tree structure is minimized. However, the algorithm is purely arithmetic and in this sense has no biological criteria. The appropriateness of the produced reticulations should therefore be evaluated (Makarenkov and Legendre 2004, manual). We decided to remove all reticulations that were longer than the longest branch length of their respective neighbour-joining trees.

For each subset, a thousand replicate distance matrices were obtained by bootstrap using AFLPsurv (Vekemans 2001), and consensus trees were generated in PHYLIP (Felsenstein 1993), providing statistical support for the obtained tree-structures.

Model-based clustering

We applied the model-based clustering algorithm implemented in the software STRUCTURE (Pritchard et al. 2000). The algorithm establishes a number of synthetic clusters with distinct allele frequencies, based on a user-specified number of allowed clusters (K), and returns a posterior probability at the specified K. The algorithm allows that individuals can have membership in multiple clusters, with membership coefficients summing to 1 across clusters. In this way it was possible to identify

clustering groups of populations, admixed populations, and individuals with mixed ancestry in a single analysis. In accordance with the software manual, the data were coded for input as diploid individuals with missing data in one copy of all loci.

The stability of the model behind STRUCTURE is known to be hampered in large systems with complex structure (Rosenberg et al. 2001). The analysis was therefore applied to each of the three population subsets. We used the admixture model option and allowed allele frequencies to differ between populations (all priors as pre-set). For each subset, we ran the algorithm at values of K ranging from 2 to the total number of populations in the subset, using 100,000 iterations for burn-in and 100,000 iterations for parameter estimation. This was repeated 5 times at each K for each subset, and the ranges of K with the highest model probabilities were identified. Then longer runs were performed, using 100,000 iterations for burn-in and 800,000 iterations for parameter estimation. To study the collapse of populations with decreasing K, long runs were performed within the identified ranges plus two steps down.When among-population variation is low (i.e. most variation is within pops), independent STRUCTURE runs are known to be less consistent (Pritchard et al. 2000; Rosenberg et al. 2002). To assess this, the longer runs were replicated 10 times for each population subset, and the median probability of all runs at each value of K was calculated in order to identify the K with the highest model probability. Which K-value is the most probable has been disputed by Evanno et al. (2005), but these authors had natural migration models as starting point for their discussion, and such models probably apply poorly to the interaction between wild and cultivated chicory, where deliberate intercrossing can be expected. We have chosen the most probable K-value as appointed by STRUCTURE.

Results

Genetic variation

In general, populations exhibited equal amounts of genetic variation (Tables 1, 2) and there were no significant pair-wise differences in expected heterozygosity in any of the population sub-groupings used for the AMOVA (Table 4). In all of the sub-groupings, most of the variation was found among individuals within populations and only smaller proportions of the total genetic variation were found among the population groups (Table 4). The negative variance component observed in the AMOVA of the IT population subset indicates absence of genetic structure and should be regarded as zero.

Table 4 Analysis of molecular variance (AMOVA) among the complete set of populations, wild populations, cultivars, and geographic population sub-groups, respectively

Population set	Variation among individuals within populations (%)	Variation among populations within groups (%)	Variation among groups of populations (%)
Total[a]	67.87	30.20	1.93
Wild[b]	75.30	18.82	5.88
Cultivated[c]	60.08	26.83	13.09
DK[d]	63.42	30.29	6.29
IT[d]	94.36	34.80	−29.16
DECHF[d]	68.08	26.22	5.70

The populations were grouped as follows: [a] wild populations versus cultivars and landraces; [b] Danish wild populations versus German/French/Swiss wild populations versus Italian wild populations; [c] landraces versus Radicchio cultivars versus Pain de Sucre versus Witloof and Root cultivars; [d] wild populations versus cultivars included in each population subset, respectively

Distance-based population structure

The observed genetic relationships between wild populations reflected the geographical distribution of wild populations with some distinct exceptions (Figs. 1, 2). Exceptions were the two Danish populations ROS and BRP, which seemed to be genetically more closely related to the Italian and German populations than to the other Danish populations, and BRP, which was relatively isolated from the rest of the wild populations. The same was true for an Italian population, SBT, which was clearly closely related to the German populations, but somewhat isolated at the end of a long branch. Finally, the Swiss population FRIB was significantly more closely related to most of the Italian populations than to the German populations. The French population, VILL, was most closely related to the main group of the Danish populations. All of the groupings were supported by high bootstrap values. Valid

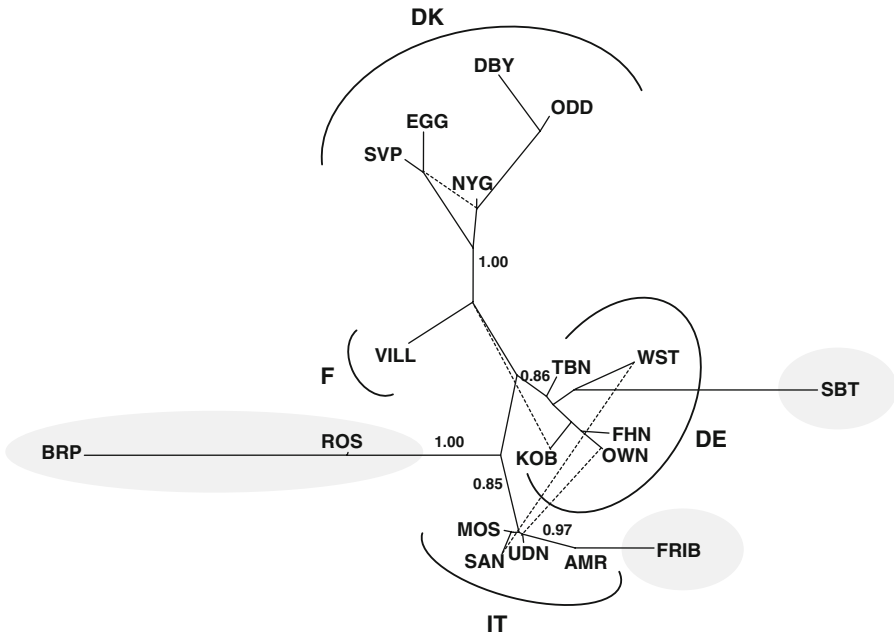

Fig. 2 Neighbour-Joining based dendrogram of all the wild populations. Reticulations (see text) are shown as stipulated lines. Bootstrap support values (1000 replications) for essential edges is shown alongside the given edge. Grey area marks four populations that diverge from the main part of wild populations

reticulations were found between two of the DE populations and Italian populations, and between the German population, KOB, and the French population VILL.

The same populations (ROS, BRP, SBT, FRIB) that deviated in the dendrogram of all wild populations, also deviated in the tree subsets of populations from DK, DECHF and IT (Figs. 3–5). Overall, the known groups of chicory cultivars and the groups of wild populations formed rather well separated clusters. In the DK dendrogram (Fig. 3), most of the wild populations formed a distinct cluster, except for ROS and BRP found on a different branch and most closely related to the Witloof cultivar BRUS. No valid reticulations were found in the DK data set. In the IT dendrogram (Fig. 4), all wild populations clustered tightly, except SBT. In the IT subset all landraces were more closely related to the wild Italian populations than to the Radicchio cultivars. Two valid reticulations were found that suggested stronger relationships between the landraces and the Radicchio cultivars than indicated by tree topology only. In the DECHF dendrogram (Fig. 5), all of the wild German populations formed a distinct cluster. The Swiss population (FRIB) was isolated from all populations. The French population (VILL) seemed to be more closely related to the root- and Witloof-

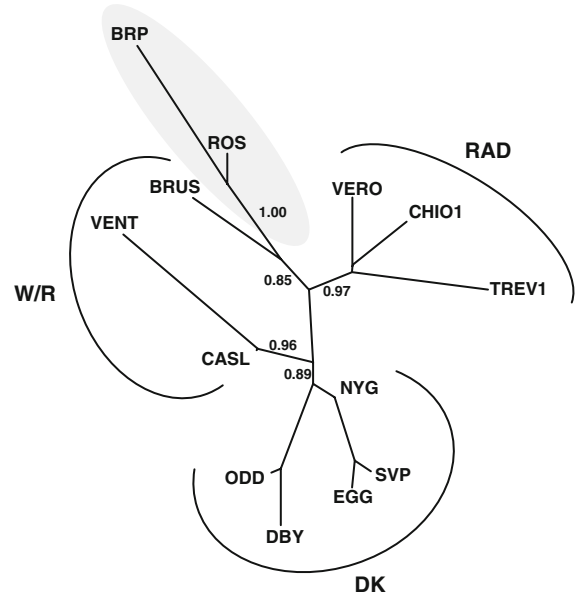

Fig. 3 Relationship between the 7 wild Danish populations (DK), 3 Root and Witloof cultivars (W/R), and 3 Radicchio cultivars (RAD), shown as a Neighbour-Joining based dendrogram. Bootstrap support values (1000 replications) for essential edges is shown alongside the given edge. Grey area marks two wild populations that diverge from the main part of wild Danish populations

type cultivars than to the other wild populations. A single valid reticulation was found between two wild German populations FHN and OWN.

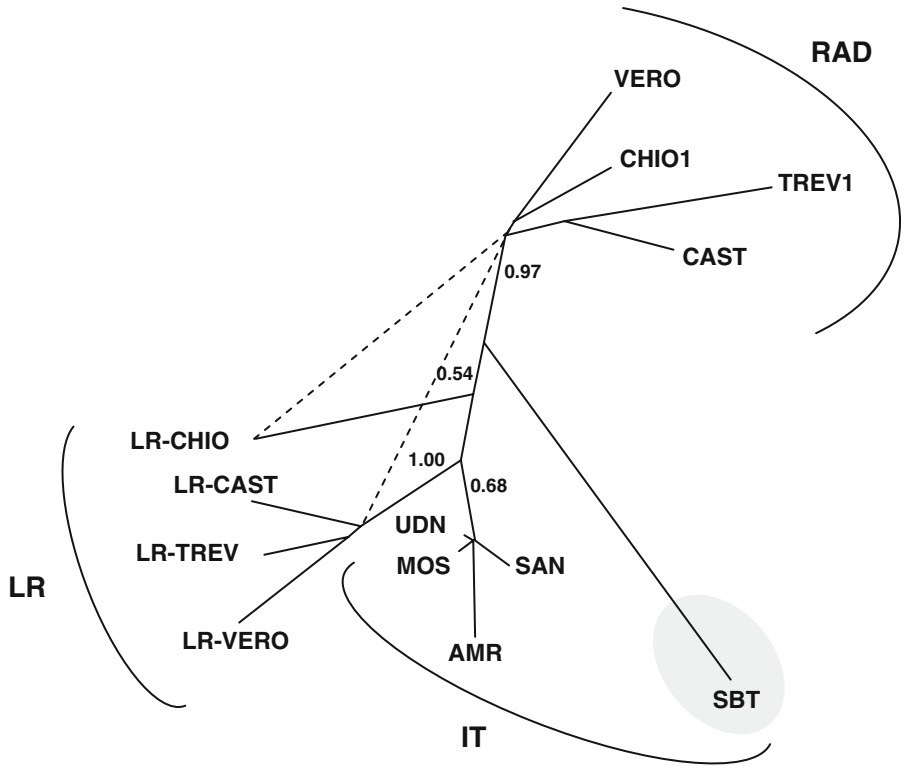

Fig. 4 Neighbour-Joining based dendrogram of the 5 wild Italian populations (IT), 4 Radicchio landraces (LR), and 4 Radicchio cultivars (RAD). Reticulations (see text) are shown as stipulated lines. Bootstrap support values (1000 replications) for essential edges is shown alongside the given edge. Grey area indicates a population that diverges from the rest of the wild populations

Introgression and model-based population structure

The majority of the individuals within single populations reassigned nearly completely to their own population (>95%), but some individuals assigned considerably to more than one population. Individual assignments in each population subset are depicted in Fig 6. We have chosen to show two *K*-values per subset—the most probable *K*-value as well as another *K*-value—and as can be seen, these was large consistency among results for different *K*-values.

The most probable model for the DK population set (Fig. 6) was found at $K = 13$ ($\sim$13 groups, ln likelihood $= -23598$). In a number of DK populations at a wide range of *K*-values including the most probable, some individuals assigned to more than one population, indicating that their genotypes were a mixture. This was especially the case for the wild population ROS, in which individuals showed widely varying degrees of admixture with BRP and another, undefined cluster. Consistently (over *K*-values) five individuals from the other wild populations shared approximately one eighth of their genes with the Witloof cultivar, BRUS.

Some of the wild Danish populations were linked in distinct clusters at all values of *K*, i.e. ODD/DBY and EGG/SVP, respectively (Fig. 6). The individuals of another wild population, NYG, were completely divided among these two distinct clusters at all values of *K*. The individuals of BRP all consistently assigned to a single cluster at all values of *K*. Population structure began to collapse as *K* decreased, but only for cultivar populations.

The most probable model for the IT population set (Fig. 6) was found at $K = 10$ (ln likelihood $= -17401$). Over all *K*-values some individuals in the IT population set showed considerable genetic admixture with other clusters than the one of their own population. For example one individual from the

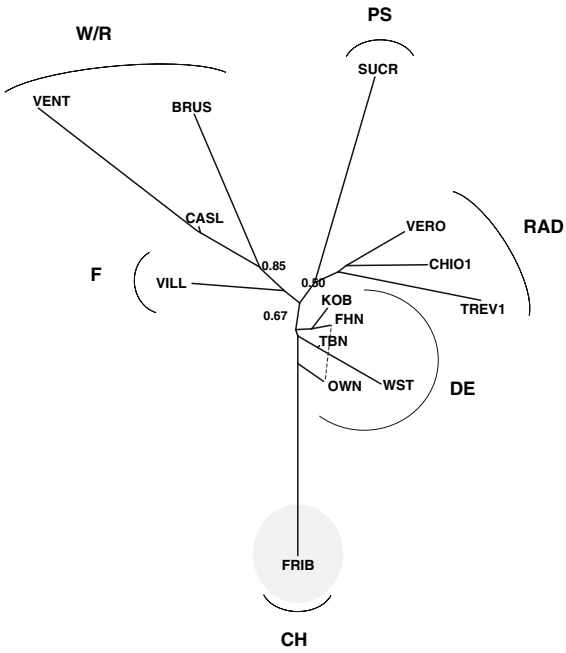

Fig. 5 Neighbour-Joining based dendrogram of the five wild German populations (DE), the wild French population (F), the wild Swiss population (CH), 3 Witloof and Root cultivars (W/R), the Pan de Sucre cultivar (PS), and 3 Radicchio cultivars (RAD). A single reticulation (see text) is shown as a stipulated line. Bootstrap support values (1000 replications) for essential edges is shown alongside the given edge. Grey area marks a wild population that diverges from the rest of the wild populations

wild UDN consistently shared a fourth of its genotype with the Radicchio cultivar TREV1. In MOS, one individual consistently shared approximately an eighth of its genes with CAST and TREV1, and another individual shared half of its ancestry with the SBT individuals. Several other wild individuals showed less pronounced signatures of introgression from cultivars. One individual from LR-VERO consistently shared half of its ancestry with the four clustering wild populations. The same was true for an individual from LR-CHIO, in which two other individuals also shared a fourth of their ancestry with wild individuals.

Four of the wild Italian populations (AMR, MOS, SAN, and UDN) formed a distinct cluster in all runs at values of K, whereas like in the reticulogram, SBT constituted an exclusive cluster at all values of K. The best run at $K = 6$ showed that most of the wild populations clustered with the three landraces LR-TREV, LR-VERO and LR-CAST. These three landraces LR-TREV, LR-VERO and LR-CAST were the first cultivars to start cluster with each other (at all $K \leq 7$).

The most probable model for the DECHF population (Fig. 6) set was found at $K = 13$ (ln likelihood $= -25611$). In the DECHF population set at all values of K, three individuals from the German wild population TBN shared between half and one forth of their genotype with the Witloof cultivar BRUS, one individual from FHN shared one forth of its markers with BRUS, and one individual from VILL shared half of its ancestry with BRUS. The wild populations generally displayed minor traces of admixture with the Pan de Sucre cultivar SUCR and the Radicchio cultivar VERO.

Some of the wild populations in the DECHF set were more or less linked in distinct clusters at all values of K. Overall, the populations KOB, FHN, and OWN assigned to the same cluster, and the individuals from TBN shared (at all values of K) a varying degree of their ancestry with both this cluster and another cluster encompassing also WST. The wild French population formed a unique cluster at all values of K. The wild Swiss population FRIB formed a unique cluster but started collapsing with the wild German populations KOB, FHN, and OWN at $K < 10$.

In all three subsets, DK, IT and DECHF, also the cultivars displayed signs of admixture, both with other cultivars and with wild populations. In a few cases admixture patterns were very alike among individuals of the same cultivar—here admixture was seen with other cultivars. However, intra-populational differences in admixture patterns were the general pattern. At lower K-values the cultivars formed groups according to expectations, i.e. root cultivars, Witloof, Pan de Sucre and Radicchio.

Discussion

The detection of introgression by means of genetic markers depends largely on the time scale of gene flow, that is, the level and duration of gene flow as well as the time since the last gene flow event. Limited and ancient gene flow is obviously harder to detect than extensive and recent gene flow. In order to detect past gene flow, its extent needs to have been substantial, so that introgression has led to a stable,

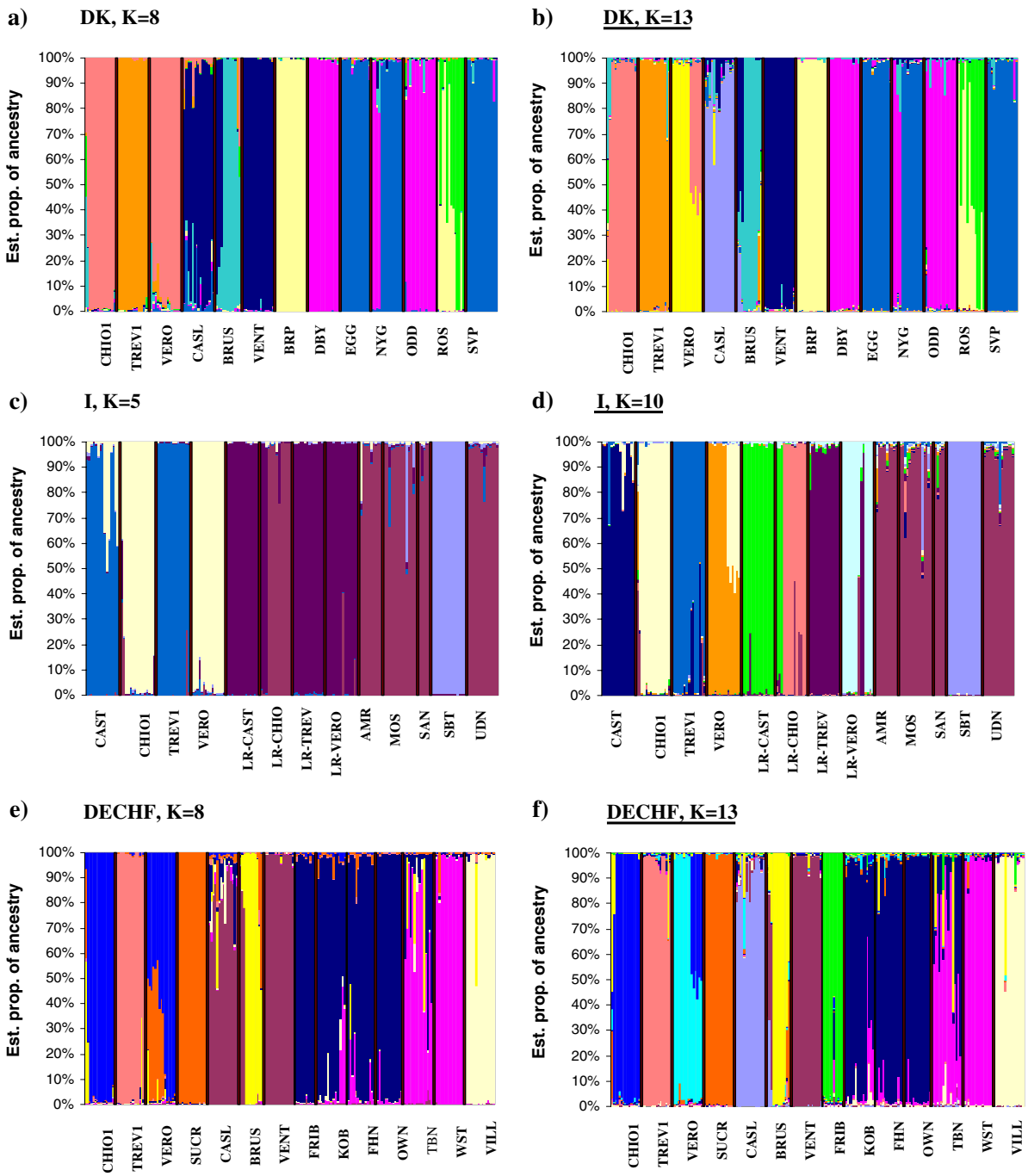

Fig. 6 Model-based clustering of individuals from the populations of the DK (**a**–**b**), I (**c**–**d**) and DECHF (**e**–**f**) subsets for different values of *K*. The runs (out of 10) with the highest model probability at each *K* are shown. Populations are divided by black vertical lines and individuals are each represented by a vertical bar. The height of different colors in each individual designates its membership coefficient to the different clusters, summing to 1 across clusters. Most probable *K*-value of each subset is underlined

mixed gene pool. In order to detect single, time-limited gene flow events, they must be recent, and exposable from the genetic composition of single individuals. As markers strictly characteristic of either the wild or the cultivated gene pool in chicory are hard to find (Van Cutsem et al. 2003; Kiers 2000), an assignment tool like STRUCTURE is extremely convenient for identifying introgressed plants in chicory.

Our results from the model-based clustering using STRUCTURE indicated frequent, recent gene flow between cultivated and wild chicory and between cultivars. Introgression between wild and cultivated plants was found in all three population subsets, but possibly this was most pronounced in the Italian set. For example one individual from the wild population UND seemed to be a BC_1 plant from hybridization with TREV1 as ¼ of the genome apparently derived from this cultivar (introgression scenario 1, see "Introduction"). It is certainly not surprising that introgression seemed to be most pronounced in the Italian set, as the cultivation of chicory is much more common in this region of collection compared to the other two collection areas in Germany and Denmark. Besides, the Italian wild populations and landraces were sampled in the area of chicory seed production. Fields with flowering cultivars for seed propagation provide ample opportunities of hybridization with other chicory fields and wild chicory in the neighbourhood. In the distance-based tree, the wild IT populations clustered with the locally propagated landraces, which may also be explained by pollen flow during propagation. The landraces are propagated at small scale at individual farms, and possibly isolation measures during this seed production are rather inefficient, and thus gene flow could be more pronounced compared to when seed production of conventional varieties take place. The rather frequent introgression detected between different Italian cultivars could be a product of both manipulated intercrossing during breeding or spontaneous crossings during seed propagation. Unfortunately, details on breeding history of the cultivars are hard to obtain, as information resides safely and inaccessibly with the breeders. However, as the introgression between cultivars was not evenly distributed among all individuals of a variety, spontaneous gene flow is the most likely source of gene transfer.

Also in the Danish subset some individuals had experienced introgression, one example being the wild Danish population ROS. As mentioned above some of the individuals seemed to have introgressed with a cluster not included in the present analysis. The ROS population was collected along a road where wild individuals, an unidentified chicory cultivar and morphologically hybrid-like individuals between the two were observed. The cultivar was apparently sown there as embellishment just after construction of the road; sowing exotic seed mixtures to restore road margins has been a popular practise. A comprehensive genetic and morphological analysis of wild, hybrid- and cultivar-individuals from the ROS population showed that introgression between the cultivar and wild chicory plants had taken place, and that hybrids were quite competitive (Kiær et al. 2007; Sørensen et al. 2007). The result from these detailed studies of the ROS population thus confirms the findings in the present study that also point to ROS as an introgressed population (introgression scenario 1). However, Kiær et al. (2007) and Sørensen et al. (2007) only studied this one Danish population, where the long lasting co-existence of crop and wild plants at the road margin may have provided optimal conditions for hybridization. Therefore, their results may not reflect the general status of introgression in chicory.

Several wild plants from other wild Danish populations had traces of introgression from the Witloof cultivar BRUS (introgression scenario 1). That this cultivar in particular contributed to introgression in DK is perhaps not so surprising, as BRUS (or related cultivar types) is the cultivar most often grown Denmark. There is a small production, and besides BRUS can be found in laymen's gardens. The production is a two step process. Large roots are produced in the field and then transferred to growth containers to germinate in darkness. Probably not all plants are transferred to dark-production; some are left in the field, where they may flower and introgress with the wild chicory in the area. It is also possible that cultivated chicory may disperse and establish in natural ecosystems, giving rise to gene flow with wild types. This is supported by finding of plants apparently belonging to the cultivated gene pool in ruderal places in Belgium (Van Cutsem et al. 2003).

In the STRUCTURE based modelling, we found replicated STRUCTURE runs to be inconsistent to some extent. This is most likely due to a low among-population variation relative to the within-population variation, as has been shown before (Pritchard et al. 2000; Rosenberg et al. 2002). Another intriguing result from the STRUCTURE based modelling should also be mentioned: A small part of individuals of two of the accessions NYG (DK-set, Fig. 6) and LR-CHIO (I-set, Fig. 6) divided among distinct clusters at almost all values of *K*. When the individuals of single populations are allocated so clearly to other populations, it is most likely that sampling errors or erroneous labelling of samples can have been the reason, as gene flow would have to be extremely massive to produce these results. However, seed dispersal could generate such results. The NYG individuals that did not cluster with their own population were assigned to DBY and ODD. All three populations were collected along roads close to major ferry-harbours that connect different Danish regions, so seed dispersal with cars or lorries from the other two populations to NYG is a possibility. The majority of the NYG individuals formed a cluster together with EGG, the population to which it was geographically closest. Whether seed dispersal could be at play in the LR-CHIO population is unknown.

We started out expecting different genetic patterns corresponding to different introgression scenarios. Corresponding to introgression scenario 1 of the introduction, we found that crop genes were only present in some individuals of a wild population, suggesting recent spontaneous introgression. For some of the wild populations introgression seemed to be so frequent that it created relatively tight links to cultivated populations. This was seen for the DK population ROS that was closer related to BRUS than to other wild DK populations. In the STRUCTURE assignments ROS also showed extensive admixture with another (cultivated?) population that was not included in the analysis. In no cases could we find evidence of deliberate inter-crossing with wild material in the breeding of cultivars, as the introgression seemed to be sporadic within a given variety (corresponding to introgression scenario 4).

Within the two gene pools of wild and cultivated chicory, levels of genetic variation were almost identical (AMOVA analysis). A similar trend was found when comparing wild and cultivated chicory (root types) from Belgium (Van Cutsem et al. 2003), so apparently breeding of chicory cultivars has had little effect on the general genetic variation.

Taken as a whole, the genetic relationships observed in the dendrograms were similar to the groupings of populations applied in the analysis of molecular variance (AMOVA) as wild populations, cultivars and landraces formed identifiable clusters, even though the different groups of chicory were not much differentiated from each other. This was also supported by the model-based clustering, as a group of cultivars and a group of wild populations were not found, when the program was forced to generate two group (data not shown). The results were clearly in favour of a domestication providing different cultivar-groups. All in all the genetic patterns that we found indicate that domestication of chicory could have taken place several times (introgression scenario 5).

Based on morphological characters (hairiness of phyllaries) Gemeinholzer and Bachmann (2005) were unable to find intraspecific structuring in *C. intybus*. They also carried out a molecular analysis of *C. intybus* comprising ITS, AFLP and microsatellite markers. This analysis focussed on interspecific differences between a number of chicory species, and a population was only represented by a single individual. Therefore it is perhaps not surprising that intraspecific differentiation in *C. intybus* was not revealed. In contrast to these results, the present analysis showed a grouping of most of the wild populations in agreement with their geographical European North–South distribution, although the differences in genetic variation between these subgroupings of Danish, German/French/Swiss and Italian accessions (AMOVA) were small. However, in the trees of population subsets, some of the wild populations diverged substantially from the rest of wild populations in that area. The three populations that diverged were the two Danish populations BRP and ROS and the Italian population SBT. In accordance with their close geographical location, the two Danish populations formed their own cluster in the model-based analysis, and some individuals from ROS seemed to have introgressed with another cluster, not included in the present analysis. SBT consistently formed its own cluster; according to its morphology it was also somewhat diverging from the other wild Italian chicories, so it might have

originated from naturalized plants derived from a cultivar. That the three populations BRP, ROS and SBT had relatively tight links to cultivars in the trees could be due to previous or recent (but frequent) introgression. The wild populations VILL and FRIB were geographically isolated from the other wild populations and their somewhat isolated position in dendrograms and STRUCTURE-outputs are therefore not surprising. More than 60% of the genetic variation was always found within the wild or cultivated populations, fulfilling the expectations for an outbreeding species. In the Belgian chicory study on root chicory and wild populations (Van Cutsem et al. 2003), the largest part of genetic variation was likewise found within populations or varieties, and wild and cultivated groups were vaguely separated. A similar trend for varieties and wild species of other obligate outcrossing species as such as *Festuca pratensis* Huds. and *Brassica rapa* L. has been reported by Fjellheim and Rognli (2005) and Andersen et al. (2008).

In conclusion, we started out with some expectations as to the genetic patterns that would be found for different introgression-scenarios: We found crop genes in only some of the individuals of a given introgressed wild population, and therefore we assume that this gene flow was of recent date corresponding to introgression scenario 1. Recent gene flow was common in all three areas of collection. Ellstrand et al. (1999) and Ellstrand (2001) have reviewed introgression between crop plants and their wild or weedy relatives. They found that crop-wild hybridization has been involved in the evolution of more aggressive weeds for seven of the world's 13 most important crops; they also reported that extensive introgression could increase the extinction risk for the wild taxon (Ellstrand 2001). Chicory can now be added to the series of plant species where crop and wild plants have been demonstrated to introgress spontaneously and rather frequently. So, in a scenario where GM chicory is cultivated openly in the field, flow of transgenes to wild types seems unavoidable. The consequences of this introgression will depend on the environment and the genes transferred. Gene flow can probably be controlled to some extent by co-existence measures, but wild chicory has a number of biological characteristics—obligate outcrosser, insect pollinated, perennial, small seeds, dispersal along roads, and high cross compatibility with cultivars—that might demand stringent control.

Acknowledgements The EU project ANGEL, "Analysis of gene flow from crop to wild forms in lettuce and chicory and its population-ecological consequences in the context of GM-crop biosafety", QLK3-CT-2001-01657, supported this work. F. Felber and R. Guadagnuolo were partially funded by the National Centre of Competence in Research (NCCR) Plant Survival, research programme of the Swiss National Science Foundation. We thank Anouk Béguin for help with the analyses and Dr. Guillaume Besnard for technical support.

References

Andersen NS, Poulsen G, Andersen BA, Kiær LP, D'Hertefeldt T, Wilkinson MJ, Jørgensen RB (2008) Processes affecting genetic structure and conservation—a case study of wild and cultivated *Brassica rapa.* Genet Resour Crop Evol. doi:10.1007/s10722-008-9354-6

Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. Mol Ecol 13:3261–3273

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:42309

Eenink AH (1981) Compatibility and incompatibility in witloof-chicory (*Cichorium intybus* L.). 2. The incompatibility system. Euphytica 30:77–85

Ellstrand NC (2001) When transgenes wander, should we worry? Plant Physiol 125:1543–1545

Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. Annu Rev Ecol Syst 30:539–563

Evanno G, Regnaut S, Goudet JG (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Felsenstein J (1993) PHYLIP, Phylogenetic inference package, version 3.5.7. A computer program distributed by the author. Department of genetics, University of Washington, Seattle, USA. Available from http://evolution.genetics.washington.edu/phylip.html

Fjellheim S, Rognli OA (2005) Genetic diversity within and among Nordic meadow fescue (*Festuca pratensis* Huds.) cultivars determined on the basis of AFLP markers. Crop Sci 45:2081–2086

Gemeinholzer B, Bachmann K (2005) Examining morphological and molecular diagnostic character states of *Cichorium intybus* L. (Asteraceae) and *C. spinosum* L. Pl Syst Evol 253:105–123

Johannessen MM, Mikkelsen TN, Jørgensen RB (2002) CO_2 exploitation and genetic diversity in winter varieties of oilseed rape (*Brassica napus*); varieties of tomorrow. Euphytica 128:75–86

Kiær LP, Philipp M, Jørgensen RB, Hauser TP (2007) Genealogy, morphology and fitness of spontaneous hybrids between wild and cultivated chicory (*Cichorium intybus*). Heredity 99:112–120

Kiers AM (2000) Endive, chicory and their wild relatives. A systematic and phylogenetic study of *Cichorium* (Asteraceae). Gorteria Suppl 5:1–78

Legendre P, Makarenkov V (2002) Reconstruction of biogeographic and evolutionary networks using reticulograms. Syst Biol 51:199–216

Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. Mol Ecol 3:91–99

Makarenkov V, Legendre P (2004) From a phylogenetic tree to a reticulated network. J Comput Biol 11:195–212

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Reynolds J, Weir BS, Cockerham CC (1983) Estimation for the co-ancestry coefficient: basis for a short-term genetic distance. Genetics 105:767–779

Rick CM (1953) Hybridization between chicory and endive. Proc Am Soc Hort Sci 61:459–466

Rosenberg NA, Burke T, Elo K, Feldman MW, Freidlin PJ, Groenen MAM, Hillel J, Mäki-Tanila A, Tixier-Boichard M, Vignal A, Wimmers K, Weigend S (2001) Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. Genetics 159:699–713

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kenneth KK, Lev A, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298:2381–2385

Ryder EJ (1999) Lettuce, endive and chicory. Crop production science in horticulture, vol 7. CABI Publishing, CAB International, New York

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Schoofs J, de Langhe E (1988) Chicory (*Cichorium intybus* L.). In: Bajaj YPS (ed) Biotechnology in agriculture and forestry, vol 6. Crops II. Springer-Verlag, Berlin/Heidelberg

Shim SI, Jørgensen RB (2000) Genetic structure in cultivated and wild carrots (*Daucus carota* L.) revealed by AFLP analysis. Theor Appl Genet 101:227–233

Simmonds NW (1976) Evolution of crop plants. Longman, London

Sørensen BS, Kiær LP, Jørgensen RB, Hauser TP (2007) The temporal development in a hybridizing population of wild and cultivated chicory (*Cichorium intybus* L.). Mol Ecol 16:3292–3298

Syed NH, Flavell AJ (2007) Sequence-specific amplification polymorphisms (SSAPs): a multi-locus approach for analyzing transposon insertions. Nat Protoc 1:2746–2752

Syed NH, Sureshsundar S, Wilkinson MJ, Bhau BS, Cavalcanti JJV, Flavell AJ (2005) *Ty1-copia* retrotransposon-based SSAP marker development in Cashew (*Anacardium occidentale* L). Theor Appl Genet 110:1195–1202

Van Cutsem P, du Jardin P, Boutte Beauwens T, Jacqmin S, Vekemans X (2003) Distinction between cultivated and wild chicory gene pools using AFLP markers. Theor Appl Gen 107:713–718

Vekemans X (2001) AFLPsurv V.1.0. A software for genetic diversity analysis with AFLP population data, distributed by the author, xvekema@ulb.ac.be, Université Libre de Bruxelles

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Friters A, Pot J, Paleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. Mol Ecol 8:907–913