Euphytica (2017) 213:120
DOI 10.1007/s10681-017-1906-4

CrossMark

Genetic diversity of flax accessions originating in the Alpine region: a case study for an ex situ germplasm evaluation based on molecular marker

Eva-Maria Halbauer · Valentina Bohinec · Melanie Wittenberger · Karin Hansel-Hohl · Stephan Gaubitzer · Eva M. Sehr

Received: 5 January 2017 / Accepted: 28 April 2017
© Springer Science+Business Media Dordrecht 2017

Abstract Agro-biodiversity is currently experiencing severe genetic erosion due to mankind's unsustainable activities. Because of initiatives following the goal of the conservation of biological diversity, so far seven million crop accessions are being conserved ex situ in gene banks worldwide. Many of these accessions are landraces being rich in gene diversity, silently awaiting their proper characterisation. This is a very critical part of any long-term strategy to enhance the productivity and resilience of crops and agricultural systems and—most importantly—to ensure the preservation of our cultural and biological heritage. In this study of an ex situ germplasm evaluation we analysed 27 flax (*Linum usitatissimum* L.) accessions originating in the Alpine region, provided by five local gene banks/providers. Based on genomic microsatellite markers (gSSRs), a varying extent of accession-specific gene diversity (expected heterozygosity, H_E) was revealed ranging from 0.05 to 0.51. Admixture of individuals between accessions was uncovered, pointing towards past processes related to gene bank management activities (e.g. intentional selection, unintentional cross-pollination during regeneration) or towards the evolution of the landrace itself (e.g. same regional origin, traditional naming), highlighting the co-existence of cultural and biological diversity. Such an genetic analysis of accessions stored ex situ not only produces valuable agronomic and breeding data, but also is useful for the clarification of past processes leading to duplicates within and between collections or mislabelling, contributing to the potential for rationalisation of collections, which in turn can help ensure that the limited resources available for regeneration are used most efficiently and effectively.

Keywords Germplasm · Characterisation · Case study · Genetic diversity · Flax · *Linum usitatissimum* L.

This article is part of the Topical Collection on *Plant Breeding: the Art of Bringing Science to Life. Highlights of the 20th EUCARPIA General Congress, Zurich, Switzerland, 29 August–1 September 2016*
Edited by Roland Kölliker, Richard G. F. Visser, Achim Walter & Beat Boller

Electronic supplementary material The online version of this article (doi:10.1007/s10681-017-1906-4) contains supplementary material, which is available to authorized users.

E.-M. Halbauer · V. Bohinec · M. Wittenberger · K. Hansel-Hohl · S. Gaubitzer · E. M. Sehr (✉)
Center for Health & Bioresources, AIT Austrian Institute of Technology GmbH, Konrad-Lorenz Str. 24, 3430 Tulln, Austria
e-mail: eva-maria.sehr@ait.ac.at

Introduction

Biodiversity is currently experiencing severe genetic erosion due to mankind's unsustainable activities.

Published online: 10 May 2017

This exponential loss of diversity through the world has led to initiatives to conserve biological diversity incorporating both ex situ and in situ techniques. Particular the former, the conservation of components of biological diversity outside—ex situ—their natural habitats, is an important strategy for crops, their landraces and wild relatives (also called plant genetic resources for food and agriculture, PGRFA), since very often original habitats are under threat. Currently, approximately 7.4 million germplasm accessions are being conserved in ex situ collections worldwide (FAO 2011). Among the most popular sites is the Svalbard Global Seed Vault securing predominantly duplicate samples from gene banks around the world, illuminating the present agro-biodiversity that exists on earth. However, until a collection has been characterised and its properties become known to germplasm managers and ultimately to breeders, it has little practical use. And so far, only a fraction of the stored accessions is properly described and is ultimately introduced into breeding programs. Thus, in order to unlock the still hidden wealth of diversity that exists in gene banks, the evaluation of germplasm collections via the application of state-of-the-art genomic, phenomic and molecular technologies is a first but very critical step.

Traditionally, the genetic diversity of germplasm collections were described based on morphological traits and were further screened particularly for those traits of interest to users (agro-morphological traits). Since the 1990s, molecular genetic tools exist to characterise a set of accessions, or even an entire germplasm collection (Mondini et al. 2009), currently ranging from genotyping based on genetic markers via sequencing of reduced representation libraries (RRLs) such as RAD-seq (Hohenlohe et al. 2012), MRE-seq (Wischnitzki et al. 2016) or even RNA-seq (De Wit et al. 2012) up to sequencing and assembly of whole genomes (Cao et al. 2011). Although these next-generation sequencing (NGS) technologies have vastly reduced their cost and have been increasingly applied in several studies (Kilian and Graner 2012), the former approach—genetic marker based genotyping—is still a widely used method of elucidating the distribution of genetic variability within and among populations. In this respect, microsatellite (SSR) markers have become the marker system of choice for population genetics in many plant species due to their high reproducibility, and multiallelic and codominant nature at a single locus (Agarwal et al. 2008; Guichoux et al. 2011; Zalapa et al. 2012). Although SSRs are getting increasingly replaced by SNPs (Single Nucleotide Polymorphisms) for rather high-throughput applications in linkage mapping and broad-scale population differentiation (Kumar et al. 2012, 2015), they are still considered as affordable and useful markers for resolving fine-scale population structuring (DeFaveri et al. 2013; Hodel et al. 2016) which is necessary to get a first glimpse into the genetic makeup of a given germplasm.

For the present study we chose flax (*Linum usitatissimum* L.) as model species, a well-known predominantly self-pollinated crop of annual growth habit with a chromosome number of $n = 15$ and a genome size of $\sim$370 Mb (Ragupathy et al. 2011). The domestication of flax as one of the eight 'founder crops' dates back to 7000 BC in the Neolithic Near East initially for oil, and later, around 2000 BC, for the fibre lineages (Fu et al. 2012). In the Alpine region, flax was evidently used since the Neolithic Age (around 3000 BC) as a source of fibre; and already back then, different flax varieties seemed to be cultivated (Herbig and Maier 2011). But flax is not only a historically important crop species; it remains a versatile and worldwide expanding crop until today. Following a global trend, the demand for linseed and linseed oil is increasing again in the last years in the Alpine region as can be seen in the Austrian production of 811 tons in 2008 rising up to 1734 tons in 2016 (STATISTIK AUSTRIA 2017), presumably due to its constantly growing image as regional super food. Although this positive trend, only one variety is still registered in Austria, the 'Ötztaler Lein', whereby in the overall European Union 176 flax/linseed varieties are recognised (European Commission 2016).

The very early distribution of flax cultivation resulted in a wide range of flax landraces created jointly by the (local) environment and by the (local) people, indicating that cultural diversity goes hand in hand with biological diversity (Gorenflo et al. 2012). Part of this diversity, namely over 46,000 flax accessions comprising both, landraces and breeding lines, is stored in gene banks all over the world (Diederichsen 2007). Out of these, currently 59 accessions are secured in Austrian gene banks, whereby 16 accessions have their origin in Austria itself (National Inventory 2017); all of them awaiting their proper characterisation and further usage.

Especially breeding programmes would benefit from the integration of characterised landraces and cultivars featured with desired traits such as higher seed yield, enhanced oil production and quality, disease resistance, and resilience against drought and heat. Despite of quite some attempts of flax germplasm characterisation (e.g. Fu et al. 2002; Smýkal et al. 2011; Soto-Cerda et al. 2014; Nag et al. 2015; Choudhary et al. 2017) we are still at the very beginning of this endeavour, which is a very critical part of any long-term strategy to enhance the productivity, sustainability and resilience of crop varieties and agricultural systems, and ultimately to secure agro-biodiversity.

By analysing 23 accessions of Austrian and four accessions of Swiss origin using genomic microsatellite (gSSR) markers (Cloutier et al. 2012), our aim was (1) to characterise the general genetic structure of the local Alpine flax germplasm, (2) to identify accessions with high genetic diversity, and (3) to identify individual diversity patterns of each gene bank. Such an analysis of accessions stored ex situ based on molecular tools produces valuable data of genetic diversity within and between accessions, which sets the baseline for the search for desired traits being hidden in accessions of high diversity. Furthermore, such a study is also useful for the identification of duplicates within and between collections, contributing to the potential for rationalisation of collections, which in turn can help ensure that the limited resources available for regeneration are used most efficiently and effectively.

Materials and methods

Plant material and DNA extraction

Seeds of 22 *Linum usitatissimum* L. accessions with their origin in Austria were retrieved via the National Inventory (2015) from the gene bank of the Austrian Agency for Health and Food Safety (AGES), the Tyrolian gene bank, and the gene bank of the Arche Noah association. Another Austrian accession was provided by the private farm Gut Neuhof. Furthermore, four Swiss accessions were provided by the gene bank Agroscope Changins. In total, summing up to 27 fibre and oil flax accessions (Table 1). Seeds were germinated in a petri dish and the aerial parts of the seedlings were harvested and kept frozen until further processing. Per accession, 16 seedlings were harvested with the exceptions BVAL-903452 (Barbara) and the harvest of 2013 from Gut Neuhof, where only nine and ten seedlings, respectively, could be obtained due to low germination success.

The extraction of genomic DNA was performed following the protocol described in van der Beek et al. (1992) with minor modifications for high-throughput handling using robotics. The extracted genomic DNA is deposited at the Repository Centre at the AIT Austrian Institute of Technology and is available upon request (Stierschneider et al. 2016).

SSR marker testing

Using a polymorphism information content (PIC) value over 0.75 as selection criterion, the following 20 genomic microsatellite markers (gSSRs) were chosen from literature (Cloutier et al. 2012) with the position on the according linkage group—if available (Cloutier et al. 2012)—in brackets: Lu2105 (LG8) Lu2157 (LG10), Lu2161 (LG3), Lu2183 (LG1), Lu2194 (LG3), Lu2292 (LG5), Lu2457 (LG2), Lu2509 (LG5), Lu2578 (LG8), Lu2589 (LG1), Lu2633 (LG3), Lu2778 (LG12), Lu2810 (LG7), Lu2825 (LG7), Lu2832 (LG7), Lu2853 (LG1), Lu2950 (N/A), Lu3097 (LG9), Lu3157 (LG8), Lu3180 (LG7). For each marker, an initial test-PCR was performed on four chosen samples in a total volume of 25 µl consisting of 5 µl of 5× HOT FIREPol Blend Master Mix (Solis BioDyne), 0.5 µl of 5 µM primer forward, 0.5 µl of 5 µM primer reverse, 1 µl DNA (undiluted), and ddH_2O. The conditions of the PCR amplification were as follows: 95 °C (15 min), followed by 30 cycles including 92 °C (60 s), 52–56 °C (60 s), 72 °C (60 s), ending in 72 °C (10 min) with a final halt at 10 °C. The PCR products were electrophoresed on 1.5% agarose gel. The following 11 markers which produced a clear single band were chosen for further analysis: Lu2105, Lu2157, Lu2509, Lu2589, Lu2633, Lu2810, Lu2825, Lu2853, Lu3097, Lu3157, and Lu3180.

SSR marker application and fragment analysis

The chosen 11 SSR markers were applied on the total sample set whereby the primer constellation in the PCR mix was changed to 0.4 µl of 5 µM primer forward, 0.8 µl of 5 µM primer reverse, and by adding 1 µl of 4 µM FAM-labelled M13 primer (Schuelke 2000).

Table 1 List of accessions

Accession number	Accession name	Biological state	Fibre/oil	Country of origin	Location	Entry	Year of reproduction	Gene bank	Form	Donor/ breeder	Corolla colour
BVAL-901519		Landrace		AT		2003	2009	AUT001		DNr. 213	Blue
BVAL-903252	Öllein	Landrace	Oil	AT	Maria Luggau	1987	2009	AUT001	Aestatis	DNr. 149	Blue
BVAL-903253	Faserlein	Landrace	Fibre (selection of BVAL-903252)	AT	Maria Luggau	1987	2009	AUT001	Aestatis	DNr. 149	White
BVAL-903350	McGregor	Breeding line	Oil	AT		2001	2009	AUT001	Aestatis	J. A. Turner	Violet
BVAL-903364	Belinka	Breeding line	Fibre	AT		2002	2009	AUT001	Aestatis	NLD014	Blue
BVAL-903426	Nike	Breeding line	Fibre	AT		2006	2009	AUT001	Aestatis	POL026	Violet
BVAL-903449	Laura	Breeding line	Fibre	AT		2012	1997	AUT001	Aestatis	Innoseeds	White
BVAL-903450	Hungarian gold	Breeding line	Oil	AT		2012	1996	AUT001	Aestatis	HUN019	Blue
BVAL-903451	Sandra	Breeding line	Oil	AT		2012	1994	AUT001	Aestatis	HUN019	Blue
BVAL-903452	Barbara	Breeding line	Oil	AT		2012	1996	AUT001	Aestatis	HUN019	Blue
LE006	LIN1020/78	Unknown		AT		1994	2012	AUT046		DEU146	
LE014	Ötztaler Lein	Landrace	Fibre	AT	Tirol	1994	2010	AUT046	var. elatum-multicaule	DEU146	Blue, partly white
LE026	Kolm	Unknown	Oil	AT		1990	2008	AUT046		Kolm (Collector)	
LE027	Lein 4	Unknown		AT		1990	2009	AUT046		Kolm (Collector)	
LE028	Lein 3	Unknown	Fibre	AT		1990	2007	AUT046		Kolm (Collector)	
LE036	Isegrim	Landrace	Selection (sister accession of LE037)	AT	Oberlienz	2001	2010	AUT046			White

Table 1 continued

Accession number	Accession name	Biological state	Fibre/oil	Country of origin	Location	Entry	Year of reproduction	Gene bank	Form	Donor/ breeder	Corolla colour
LE037	Marbod	Landrace	Selection (sister accession of LE036)	AT	Oberlienz	2001	2010	AUT046			Blue
LU001		Landrace		CH		2012	2009	CHE001	var. intermedium f. erecta	RUS204 (K-7747)	Blue
LU002		Landrace		CH		2012	2013	CHE001	var. intermedium f. erecta	RUS204 (K-5735)	Blue
LU003		Landrace		CH		2012	2013	CHE001	var. elongatum f. erecta	RUS204 (K-5954)	Blue
LU004	Weisser Faserlein	Landrace	Fibre	CH	Brüningen Bern	2014	2013	CHE001		DIV-1881 (CHE063)	Blue
Neuhof	cf. Ötztaler Lein			AT			2013	Gut Neuhof		Gut Neuhof	
RINN-822004	Ötztaler Lein	Landrace	Fibre	AT	Ötztal		2012	AUT005	Aestatis		Blue, partly white
RINN-822006	Umhausen	Landrace	Fibre	AT	Umhausen		2011	AUT005	Aestatis		
RINN-822007	ÖHV Lehn Längenfeld	Landrace	Fibre	AT	Lehn bei Längenfeld		2012	AUT005	Aestatis		
RINN-822008	Umhausen Christoph Schmidt	Landrace	Fibre	AT	Umhausen		2012	AUT005	Aestatis		
RINN-824005	Ötztaler x Lussatia	Breeding line	Fibre	AT			2011	AUT005	Aestatis		

AT: Austria; CH: Switzerland; AUT001: AGES Austrian Agency for Health and Food Safety: Dpt. Plant Genetic Resources; AUT005: Tyrolean gene bank, Office of the Tyrolean Regional Government; AUT046: Arche Noah association; CHE001: Agroscope Changins

The resulting PCR products were quality checked on a 1.5% agarose gel and were diluted and mixed with Hi-Di Formamide and GeneScan 350 ROX dye Size Standard according to the manufacturers protocols (Life Technologies). The size of the fragments was resolved based on capillary electrophoresis using the ABI 3110 XL Genetic Analyzer. Allele calling was performed using GeneMapper® Software 5 (Applied Biosystems). Non-amplified loci were scored as missing data. The few occurring polyploid loci were scored as well as missing data.

Genetic data analysis

Each marker was evaluated for the number of alleles (N_A), the observed heterozygosity (H_O), and the expected heterozygosity [H_E; gene diversity (D)] based on the entire dataset. To avoid allele frequencies bias due to sibship (Anderson and Dunham 2008), clonality within the dataset was determined in silico by measuring the number of 100% multilocus matches by GenAlEx. Repeated matching multilocus genotypes were removed from the dataset for subsequent population genetic and structure analysis. Population structure of the reduced dataset was examined using the Bayesian model-based approach implemented in Structure 2.3.4 (Pritchard et al. 2000). The number of clusters (K) evaluated ranged from 1 to 30. The analysis was performed using five replicate runs per K value, a burn-in period length of 10,000, and a run length of 50,000. The admixture model (AD) was used to determine the correlated cluster, because some accessions were derived from breeding programs and shared pedigrees. The information about the assignment of each individual to the accession was provided via the locprior setting. Two different approaches were used to detect the most likely K value: the first is based on the rate of change of lnP(D) for each K between 1 and 30 (Pritchard et al. 2000) and the second, proposed by Evanno et al. (2005), is based on the second order rate of change of the likelihood function with respect to K (ΔK) calculated by the R package pophelper (Francis 2016). To determine the associations between the accessions, statistical parameters [H_E, H_O, fixation index (F)] were calculated and an analysis of molecular variance (AMOVA) using 1000 permutations was performed. All calculations were done using Arlequin 3.5 (Excoffier et al. 2005; Excoffier and Lischer 2010) and/or GenAlEx 6.502 (Peakall and Smouse 2012). Based on pairwise genetic distance matrices and pairwise population matrix of Nei unbiased genetic distance values calculated in GenAlEx, trees based on the Neighbor-Joining (NJ) method (Saitou and Nei 1987) using MEGA 6 (Tamura et al. 2013) were created to visualize genetic diversity and the evolutionary history among individuals and accessions.

Results and discussion

Heterozygosity on the locus level

Heterozygosity and polymorphism were calculated for each locus separately (Table 2). Considering the entire

Table 2 Heterozygosity and polymorphism per locus based on the entire dataset calculated using GenAlEx without population grouping

	Lu2105	Lu2157	Lu2509	Lu2589	Lu2633	Lu2810	Lu2825	Lu2853	Lu3097	Lu3157	Lu3180	Mean
LG	LG8	LG10	LG5	LG1	LG3	LG7	LG7	LG1	LG9	LG8	LG7	
N	409	419	415	385	414	73	94	403	79	411	377	316.27
N_A	17	7	8	16	14	4	2	9	4	15	15	10.09
N_E	5.46	3.86	2.28	7.18	5.22	3.95	4.94	4.05	2.38	1.99	1.61	3.90
H_O	0.06	0.06	0.03	0.76	0.07	0.25	0.03	0.10	0.05	0.94	0.04	0.22
H_E	0.82	0.74	0.56	0.86	0.81	0.75	0.80	0.75	0.58	0.50	0.38	0.69
F	0.93	0.92	0.95	0.12	0.91	0.66	0.96	0.87	0.91	−0.88	0.90	0.17
PIC[a]	0.77	0.78	0.79	0.85	0.84	0.81	0.84	0.77	0.78	0.84	0.80	0.81

Linkage group (LG), number of individuals (N), different alleles per locus (N_A), number of effective alleles per locus (N_E), expected (H_E) and observed heterozygosity (H_O), and fixation index (F). Except for N and N_A, the values for each locus are depicted as mean values over all populations

[a] Polymorphism information content (PIC) values taken from Cloutier et al. (2012)

dataset of 419 individuals, a total of 111 alleles were detected at 11 SSR loci, whereby per locus, the number of alleles ranged from two (at locus Lu2825) to 17 (at Lu2105) with an average of 10.09 alleles per locus. This value exceeds the range reported in previous studies (2.72–5.32) (cf. Soto-Cerda et al. 2013) and might be due to the initial selection of loci with high values for the polymorphism information content (PIC). The expected heterozygosity (H_E) values were in the range of 0.38–0.86 with a mean value of 0.69, whereby the observed heterozygosity (H_O) values varied between 0.03 and 0.94 with a rather low mean value of 0.22. At seven loci, the H_E values were above 0.7. Taken from Cloutier et al. (2012), the PIC value for each locus varied between 0.77 and 0.85 with an average of 0.81. Loci with high H_E and PIC values are considered as highly polymorphic and are therefore a good indicator for consistent and accurate population studies, also, when small sample numbers are analysed (Hale et al. 2012).

Genetic structure of the germplasm

When using certain software tools such as the program Structure (Evanno et al. 2005) to infer the presence of population structure in a given sample set, it has been discussed, that sibship, consanguinity or clonality influence allele frequencies and lead to false family-induced structures (Anderson and Dunham 2008; Goldberg and Waits 2010; Peterman et al. 2016). Therefore, a reduced dataset without repeated matching multilocus genotypes was generated resulting in 241 individuals out of initially 419. Despite DNA was extracted originally from 16 individuals per accession (except for BVAL-903452 and Gut Neuhof, where only nine and ten seedlings germinated, respectively), after the removal of repeated matching multilocus genotypes, on average 8.89 individuals remained per accession. However, according to Hale et al. (2012), a good sample size for SSR-based population genetic studies would be 25–30 individuals per population. Although the definition of who and what a population is can be controversial (Krieger 2012), in genetics, *per definitionem*, a population reflects a group of organisms within which breeding occurs. Based on that fact, although or better because of artificial management and isolation, each accession is treated as a single population in all subsequent analyses. However, with these low sample sizes per accession, an effect on the accuracy of allele frequency estimation and thus also on the thereupon calculated population genetic values cannot be ruled out.

Germplasm structure was analysed using the reduced dataset (n = 241) by applying the model-based approach in the Structure software (Pritchard et al. 2000; Anderson and Dunham 2008). According to Evanno et al. (2005) based on ΔK, the germplasm was partitioned into three clusters (Supplementary Fig. S1). Despite these model-based estimations of population structure, a certain bias e.g. due to unbalanced sample sizes (Kalinowski 2011) but especially the biological meaning must be taken into consideration when choosing the optimal value for K. Since the herein analysed accessions had their origin in Austria and Switzerland, we assumed a rather narrow genetic ancestry and thus, partitioned the dataset into the calculated three clusters. Cluster 1 (dark blue) was present in 3 gene banks (AGES, Arche Noah, and Agroscope) encompassing seven accessions, cluster 2 (light blue) appeared in all gene banks encompassing 20 accessions, and cluster 3 (red) was present in four gene banks (AGES, Arche Noah, Agroscope, Gut Neuhof) encompassing seven accessions (Fig. 1). In other words, the gene banks of AGES, Arche Noah and Agroscope were characterised by all three clusters, whereby in the latter gene bank only one individual of the accession LU001 was assigned to cluster 3 (red, Fig. 1). The Tyrolean gene bank is the only one in which all accessions, even all individuals, were characterised by one cluster (cluster 2, light blue, Fig. 1), indicating a uniform genetic background. Over the entire germplasm, within most of the accessions genetic uniformity could be observed (Fig. 1), however, seven accessions were composed of individuals with different ancestry patterns (see Table 3), highlighting the diversity within these accessions. Finally, some individuals of the accessions BVAL-903252 (n = 5) and BVAL-903253 (n = 3) showed admixture from the three clusters, and can thus not be assigned accordingly to one cluster.

Germplasm structure analysis revealed already accessions with presumably higher diversity (e.g. BVAL-903252 or LU001, see Table 3) and illustrated the fact, that all accessions over all gene banks shared a similar genetic background highlighting their close relationship most likely due to their similar origin in the Alpine region.

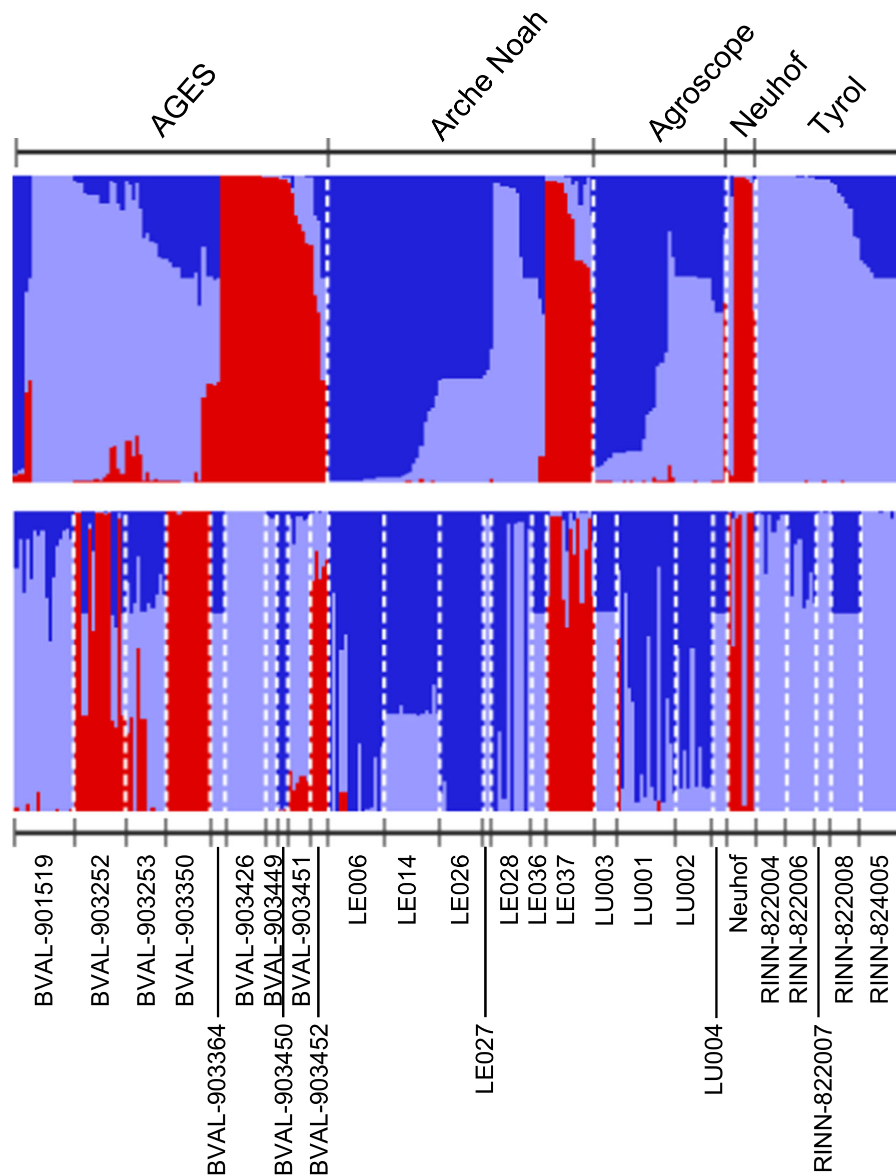

Fig. 1 Genetic structure of the germplasm grouped in gene banks (n = 5) sorted by cluster (*upper plot*), and grouped into the respective accessions (n = 27, *lower plot*) under the assumption of K = 3

Genetic diversity of the overall germplasm

Diversity parameters were calculated based on the reduced dataset (n = 241). Per accession, the calculated mean values for the expected heterozygosity (H_E, or gene diversity, D) varied from 0.05 to 0.51 (grand mean = 0.24), and the H_O (observed heterozygosity) values were below the H_E ranging from 0.09 to 0.27 (grand mean = 0.16). An overview of heterozygosity and polymorphism for each accession is given in Table 3. This is in line with previous studies, where a generally narrow genetic base of flax germplasm accessions was described (Fu et al. 2002; Smýkal et al. 2011; Soto-Cerda et al. 2012). In natural populations low H_E values can occur due to genetic isolation, historical population bottlenecks, founder effects and inbreeding. However, in breeding lines and also landraces, which is the case for most of the accessions stored in gene banks, the effect of direct selection in the domestication and breeding process is discussed to

Table 3 Heterozygosity and polymorphism per accession on the reduced dataset without repeated matching multilocus genotypes calculated using GenAlEx

Accession	N	N_M	N_A	N_E	H_O	H_E	F	Cluster (K = 3)
BVAL-901519	16	11.36	2.55	2.00	0.19	0.41	0.56	C2
BVAL-903252	14	9.73	3.73	2.48	0.19	0.51	0.62	C1[a], C2[a], C3
BVAL-903253	11	7.64	2.55	1.69	0.20	0.39	0.57	C2, C3
BVAL-903350	12	8.45	1.27	1.12	0.18	0.20	0.11	C3
BVAL-903364	4	2.91	1.09	0.98	0.23	0.14	−0.52	C2
BVAL-903426	11	7.45	1.18	1.00	0.13	0.16	0.41	C2
BVAL-903449	3	2.18	0.91	0.91	0.12	0.09	−0.33	C2
BVAL-903450	3	2.18	1.00	0.96	0.09	0.13	0.33	C1
BVAL-903451	6	3.73	1.00	0.92	0.14	0.11	−0.11	C2
BVAL-903452	4	3.45	1.00	0.96	0.14	0.12	−0.11	C3
LE006	15	9.91	2.55	1.86	0.09	0.38	0.81	C1, C2
LE014	15	14.27	2.55	1.87	0.18	0.38	0.54	C1
LE026	12	11.27	1.73	1.33	0.09	0.21	0.64	C1
LE027	2	1.91	1.09	1.09	0.09	0.05	−1.00	C2
LE028	11	10.73	2.27	1.74	0.19	0.39	0.45	C1, C2
LE036	4	3.82	1.18	1.15	0.14	0.08	−0.67	C2
LE037	13	10.18	2.55	1.62	0.27	0.33	0.17	C3
LU001	16	6.82	1.82	1.36	0.12	0.27	0.61	C1, C2, C3
LU002	10	4.09	1.18	1.01	0.19	0.15	−0.11	C1, C2[a]
LU003	6	10.27	3.55	2.20	0.21	0.47	0.55	C2
LU004	4	2.82	1.00	0.91	0.14	0.09	−0.51	C2
Neuhof	8	5.82	2.36	1.61	0.19	0.39	0.52	C2, C3
RINN-822004	8	5.64	1.91	1.31	0.19	0.24	0.40	C2
RINN-822006	8	5.82	2.00	1.46	0.23	0.34	0.38	C2
RINN-822007	4	2.91	1.18	1.03	0.11	0.16	0.38	C2
RINN-822008	8	5.55	1.00	0.93	0.14	0.11	−0.17	C2
RINN-824005	12	7.45	1.55	0.92	0.13	0.19	0.32	C2
Mean values	8.89	6.61	1.77	1.35	0.16	0.24	0.35	

An assignment of the accessions to clusters is based on K = 3

Number of individuals per accession of the reduced dataset (N). Mean numbers over all loci of individuals (N_M), of different alleles per locus (N_A), of number of effective alleles per locus (N_E), of expected (H_E), of observed heterozygosity (H_O), and of the fixation index (F)

[a] Indication of admixture from three clusters

be the main cause for low heterozygosities (Gepts 2003; Flint-Garcia 2013). Which can be directly transferred to flax, since its domestication as one of the eight 'founder crops' happened more than 8000 years ago in the Neolithic Near East; initially for oil, and later, around 3000 years ago, for the fibre lineages (Fu et al. 2012). Despite their difference in domestication time, their apparent difference in nucleotide diversity of 24 resequenced genomic regions (fibre flax showed a lower diversity than oil flax; Fu et al. 2012), and their difference in gene diversity at gSSR loci (fibre flax: 0.30, oil flax: 0.36; Choudhary et al. 2017), when comparing these two lineages in our dataset, gene diversity did not differ with a mean value of 0.61 for the accessions assigned to the fibre lineage and 0.62 for the oil lineage accessions (Table 4). This indicates that in both lineages, at least for the herein analysed flax germplasm of the Alpine region, a rather high level of genetic variation at the analysed loci has been

Table 4 Heterozygosity and polymorphism of the oil versus fibre lineages based on the reduced dataset without repeated matching multilocus genotypes calculated using GenAlEx

Flax lineage	N	N_M	N_A	N_E	H_O	H_E
Fibre	99	75.36	6.45	3.13	0.25	0.61
Oil	52	38.82	6.00	3.62	0.20	0.62
Unassigned	90	64.18	7.45	4.48	0.26	0.71
Mean values	80.33	59.45	6.64	3.74	0.24	0.65

Number of individuals per lineage of the reduced dataset (N). Mean numbers over all loci of individuals (N_M), of different alleles per locus (N_A), of number of effective alleles per locus (N_E), of expected (H_E), of observed heterozygosity (H_O), and of the fixation index (F)

Table 5 Analysis of molecular variance using the reduced dataset without repeated matching multilocus genotypes calculated by Arlequin

Source of variation	d.f.	Sum of squares	Variance	% total	P
Without grouping					
Among accessions	26	487.01	1.008	53.06	<0.0001
Within accessions	455	405.86	0.892	46.94	<0.0001
Total	481	892.87	1.900		
With grouping (n = 5)					
Among groups	4	115.86	0.104	5.41	0.0088
Among accessions	22	371.15	0.928	48.24	<0.0001
Within accessions	455	405.86	0.892	46.35	<0.0001
Total	481	892.87	1.924		

d.f.: degrees of freedom; P: probability of having a more extreme variance component than the observed values by chance alone

preserved over time. In contrast to managed populations of flax, genetic diversity of the proposed progenitor species of cultivated flax, pale flax (*Linum bienne* Mill.), is described to be on a higher level (Soto-Cerda et al. 2014), supporting the universality of loss of diversity in cultivated crops relative to their wild ancestors (Tanksley and McCouch 1997). Our results further support the central dogma [although probably only a hypothesis given the discrepancies between various studies described in Fu (2015)] that wild relatives but also landraces show higher genetic diversity than breeding lines of a given crop species: the landraces (comprising 14 accessions) were characterised with a mean value for gene diversity (H_E) of 0.28 and the breeding lines (eight accessions) with 0.14.

Genetic diversity on the level of accessions

From all 27 accessions, the most diverse accessions with H_E values equal or greater than the value for the positive standard deviation (mean H_E 0.24 + standard deviation 0.13 = 0.38) were three accessions from the AGES gene bank (BVAL-903253, BVAL-901519, and BVAL-903252), three from the Arche Noah gene bank (LE006, LE014, and LE028), one from Agroscope (LU001) and the single one from Gut Neuhof. The least diverse accessions with values equal or lower than the value for the negative standard deviation ($H_E \leq 0.11$) were BVAL-903449 (AGES), LE027 and LE036 (Arche Noah), and LU004 (Agroscope).

An AMOVA analysis based on the reduced dataset without grouping the accessions according to their gene bank association was done using the software Arlequin to evaluate the diversity components within and among the distinct accessions. The majority of the variance occurring among the accessions accounted for 53.06% of the total variation, whereby 46.94% of the variation was attributed to differences within the accessions (Table 5), which is in line with previous studies, where a higher among than within population

Table 6 Heterozygosity and polymorphism per gene bank on the reduced dataset without repeated matching multilocus genotypes calculated using GenAlEx

Gene bank/provider	N	N_M	N_A	N_E	H_O	H_E	F	Cluster (K = 3)
AGES	85	59.09	6.91	3.42	0.17	0.56	0.71	C1-3
Arche Noah	72	62.09	6.45	3.48	0.18	0.65	0.69	C1-3
Agroscope	36	24.00	4.73	2.72	0.17	0.51	0.68	C1-3
Gut Neuhof	8	5.82	2.36	1.61	0.19	0.39	0.52	C2, C3
Tyrolian gene bank	40	27.36	2.91	1.70	0.19	0.40	0.57	C2
Mean values	48.20	35.67	4.67	2.59	0.18	0.50	0.64	

An assignment of the clusters to the gene bank is based on the assumption of K = 3

Number of individuals per gene bank of the reduced dataset (N). Mean numbers over all loci of individuals (N_M), of different alleles per locus (N_A), of number of effective alleles per locus (N_E), of expected (H_E), of observed heterozygosity (H_O), and of the fixation index (F)

differentiation was detected, e.g. 62 versus 38% in Habibollahi et al. (2015).

The evaluation of gene diversity present in a germplasm is of special interest, since accessions harbouring high genetic diversity can be identified which may serve as a valuable resource to support plant breeders with genetic material to extend genetic variability, as a basis to create new crop varieties. These assessments not only provide guidance for better germplasm utilization for genetic improvement, but also facilitate efforts in germplasm conservation.

Migration rate and gene flow

The relative measure of migration between the accessions (Nm) was 0.13, measured as the mean value of per locus values of Nm over all populations by GenAlEx, which falls in the range of previously described gene flow values of self-pollinated plant species (Govindaraju 1989). Flax is described to be a predominantly self-pollinating species, but gene flow due to cross-pollination in the range of 1–5% has been described to occur when plants are grown in close proximity (Jhala et al. 2011). Therefore, an isolation distance of <0.8 km has been proposed for flax by Mader and Hopwood (2013) in the context of organic seed production. However, gene flow is also described to occur to a certain extent in germplasm collections (de Vicente 2005). Because of seeds are frequently regenerated to keep their viability and to replenish seed stocks of ex situ collections, gene flow may occur as the result of cross-pollination, as well as through physical mixing of seed lots. Therefore, in order to ensure the safeguard of the genetic integrity and the genetic diversity of a given germplasm collection, knowledge of pollination strategies, a proper regeneration process, as well as careful gene bank management and care are of utmost importance!

Genetic setup within individual gene banks

After grouping the accessions according to their provider/gene bank (n = 5, cf. Table 1), the degree of genetic diversity (H_E, D) within a specific gene bank ranged from 0.39 for Gut Neuhof up to 0.65 for the Arche Noah gene bank, with a grand mean over all gene banks of 0.50 (Table 6). It is not surprising, that Gut Neuhof is characterised by a limited genetic diversity since only one accession (with eight individuals in the reduced dataset) was analysed. On the other hand, in the gene banks of Arche Noah and AGES high diversity was present, with seven accessions (72 individuals) and ten accessions (85 individuals) analysed, respectively. According to the AMOVA, only 5.41% of the total variation was attributed to differences among the five gene banks. The highest proportion of variation was seen among the accessions with 48.24%, followed by the variation within the accessions with 46.35% (Table 5), supporting the results of the structure analysis.

In order to resolve the relationships among the individuals within a gene bank, NJ trees based on genetic distance values of the unreduced datasets were generated. This visualisation provides a deeper insight into the extent of genetic diversity of the accessions as well as putative admixture events between accessions. Focusing on the analysed ten accessions from the AGES gene bank (Supplementary Fig. S2, left),

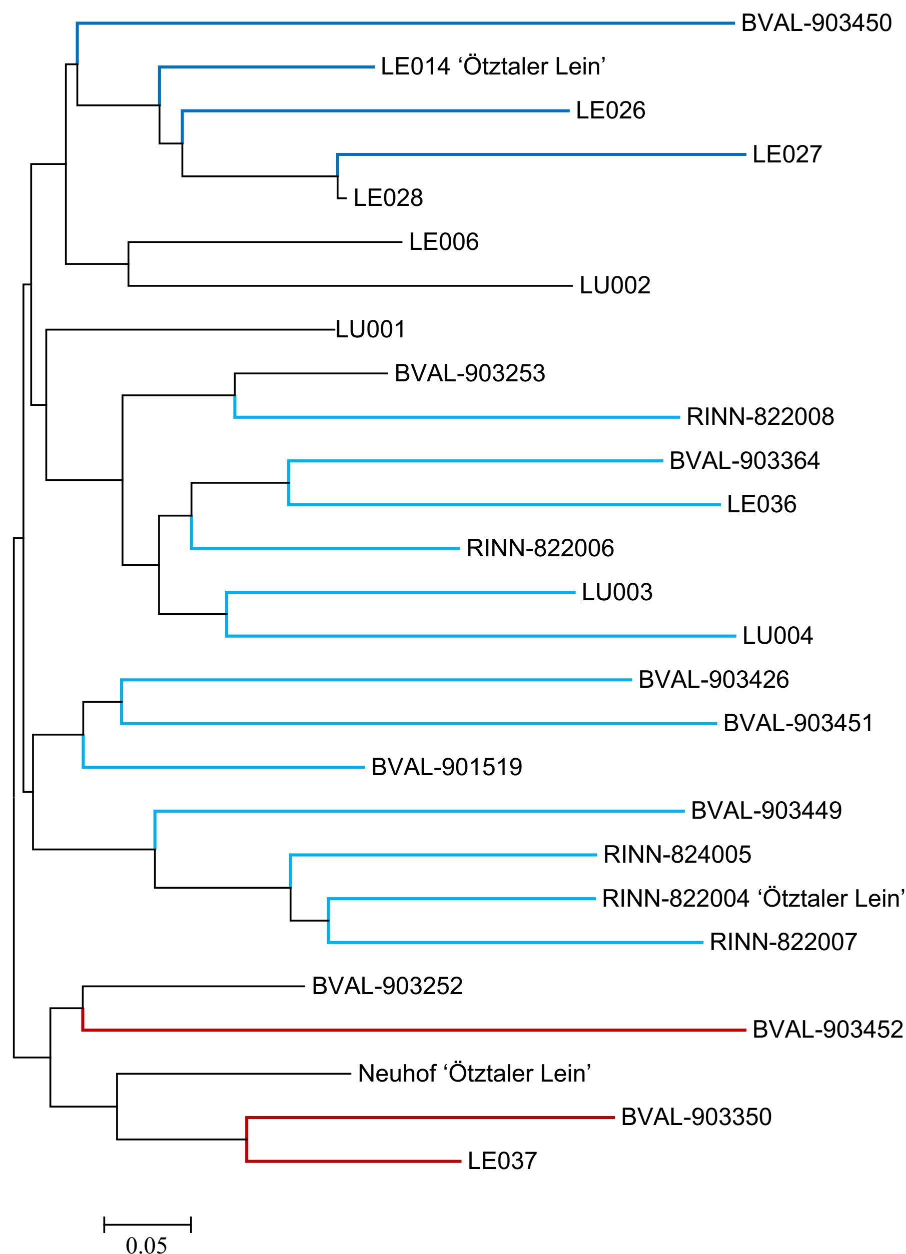

Fig. 2 The Neighbor-Joining method was used to infer the evolutionary history of the accessions based on the pairwise population matrix of Nei unbiased genetic distance values of the unreduced dataset. The optimal tree with the sum of branch length = 6.47590332 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The branch of those accessions where all individuals were assigned to one of the three clusters is marked in the respective colour [*dark blue* (cluster 1), *light blue* (cluster 2) and *red* (cluster 3)], branches of admixed accessions are kept uncoloured

accessions with higher diversity (distinct branching pattern, e.g. BVAL-901519) were clearly distinguishable from those with lower diversity (collapsed branches, e.g. BVAL-903449). In addition, prominent admixture of the accession BVAL-903252 with BVAL-903350 and with some individuals of BVAL-

903253 was revealed. The accession BVAL-903350 (McGregor) originated from a flax line traded in Austria, thus, it is striking that such a severe admixture with a landrace was recognisable. A possible explanation could be the very unlikely but still possible physical mixture of seeds or cross-pollination during the regeneration process. On the other hand, that BVAL-903252 admixed with BVAL-903253 can be explained by the fact, that the latter accession was selected from the first one during the regeneration process by the gene bank manager. Furthermore, one individual of BVAL-903426 appeared within BVAL-901519, which could be also due to any anthropogenic influence.

Similarly, high genetic diversity of accessions of the Arche Noah gene bank (Supplementary Fig. S2, right) was reflected by a deeper branching pattern (e.g. LE006, LE037) in comparison to accessions with low diversity (e.g. LE027). Here as well, admixture between two accessions, LE026 and LE028, could be visually identified. Interestingly, those accessions are stemming from the same donor/sustainer, who, presumably, regenerated the material on the same field. Moreover, one individual of LE037 appeared between LE014 and LE026, which could be due to the fact, that those three accessions were regenerated by the gene bank manager on the same field without any isolation distance.

Within the Tyrolean gene bank, the only breeding line (RINN-824005) appeared to be the most diverse accession in comparison to the locally collected landraces (Supplementary Fig. S3, left). That the two accessions RINN-822006 and RINN-822008 were originally collected in the same location (Umhausen, Tyrol), and are thus closely related, can be seen by their proximity in the NJ tree. Admixture of individuals of RINN-822007 and RINN-822006 into the accession RINN-822004 could be explained by the fact, that all these accessions were originally collected as landraces in the same valley, the Ötztal in Tyrol, and are thus sharing a similar genetic background.

Finally, based on the NJ tree, the same conclusions regarding the extent of genetic diversity can be drawn for the four accessions of the Agroscope gene bank (Supplementary Fig. S3, right). Here as well, admixture of one individual of LU003 into LU002, and two individuals of LU001 between LU002 and LU003 were detected.

Duplication, mislabelling, misidentification—or the legacy of cultural diversity?

In order to be able to detect accession duplicates and possible mislabelling or misidentification across gene banks, a NJ tree based on pairwise population matrix of Nei unbiased genetic distance values of the unreduced dataset was generated (Fig. 2). In general, the structure of the tree very well reflected the three detected genetic clusters (cf. Fig. 1). A further correlation of the tree structure to the two lineages (fibre, oil) or to the five gene banks was however not detected. Furthermore, with the available information about the accessions, no duplication within and among gene banks was identified. However, on the basis of the present three accessions of the landrace called ‘Ötztaler Lein’ (LE014, RINN-822004, and the accession from Gut Neuhof) one would expect that those accessions appear next to each other in the NJ tree due to their expected genetic similarity. But the contrary was the case, leading to the discussion of the ‘correctness’ of characterisation, identification, traditional naming and further usage, regeneration and care of landraces, which was partly addressed already by Berg (2009). Traditionally, landraces are characterised based on their morphology and are often named after a distinct location. As is the case for the landrace ‘Ötztaler Lein’, which has its origin in the Ötztal valley (Tyrol, Austria); being in cultivation since centuries, currently classified as rare agricultural crop according the Austrian agro-environmental program ÖPUL (2015), and sustained as such. The obvious genetic-based differences between the three accessions of the landrace ‘Ötztaler Lein’ are puzzling and could probably be explained by the following possible scenarios: either the original accession was separated very early in time and the resulting parts have been evolved differently (possibly in different locations) since then, or that different landraces were named the same due to morphological similarity, which also appeared to happen for rice varieties (Sathya 2014), or because of the seed trading behaviour of local farmers in the past, where a mixture of varieties was the norm (Vogl-Lukasser et al. 2007) which could have resulted in different genetic backgrounds of the same landrace. All scenarios point towards the fact, that culture and tradition had a great influence in the creation, naming, maintenance and further development of landraces. It is already a fact

for quite some time that cultural diversity goes hand in hand with biological diversity (Maffi 2005; Galluzzi et al. 2010; Gorenflo et al. 2012). Especially landraces represent a subset of agro-biodiversity that has been created jointly by the (local) environment and by the (local) people, showing a tight linkage of the biological and cultural heritage. Thus, as Negri (2005) reflected already, it is of utmost importance to 'acquire a greater awareness of the importance of the agro-biodiversity, reinforce the links between rural communities, their environment and plant genetic resources, foster pride among young farmers with regards to their natural and cultural heritage' in order to be able to preserve agricultural genetic resources.

Conclusions

The results of this study highlight the importance of germplasm evaluation based on molecular markers. The analysis of the genetic structure and diversity of 27 flax accessions originating in the Alpine region kept at local gene banks and one local sustainer revealed a varying extent of gene diversity (H_E) between 0.05 and 0.51. Our results are further supporting the central dogma that landraces show higher genetic diversity than breeding lines, with mean values for gene diversity of 0.28 (landraces) versus 0.14 (breeding lines). Despite the herein and elsewhere detected narrow genetic base of flax germplasm, especially those accessions characterised by a higher gene diversity may still have the potential to serve as a valuable resource supporting plant breeders to extend the genetic variability of modern flax lines. Furthermore, our findings suggest that much of the diversity that survived through the stages of domestication has been retained ex situ in gene banks and is well represented in the local Alpine flax collection.

But these assessments not only provide guidance for better germplasm utilization for genetic improvement, but also facilitate efforts in germplasm conservation. Certain admixture was detected within and between accessions which could be partly explained by intentional gene bank management activities (e.g. selection based on morphological characteristics), but partly were due to the concerted action of several unintentional causes such as cross-pollination during regeneration processes where no isolation distance was considered or due to any anthropogenic influence in the past, highlighting a deep interrelationship between biological and cultural diversity.

Therefore, in order to ensure the safeguard of the static genetic integrity and the genetic diversity of a given ex situ germplasm collection, knowledge of past culture-based activities, but also pollination strategies, a proper regeneration process, as well as careful gene bank management, supported by the knowledge of its genetic diversity and integrity are of utmost importance in order to combat ongoing genetic erosion of existing agro-biodiversity.

Acknowledgements We thank the gene bank of the Austrian Agency for Health and Food Safety (AGES), the Tyrolian gene bank, the gene bank of the Arche Noah association, the gene bank Agroscope Changins, and the private farm Gut Neuhof, Obersiebenbrunn, Austria, for providing seeds. The project was financially supported by the AIT Austrian Institute of Technology GmbH and the FEMtech programme of the Austrian Research Promotion Agency (FFG).

References

Agarwal M, Shrivastava N, Padh H (2008) Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Rep 27:617–631. doi:10.1007/s00299-008-0507-z

Anderson EC, Dunham KK (2008) The influence of family groups on inferences made with the program structure. Mol Ecol Resour 8:1219–1229. doi:10.1111/j.1755-0998.2008.02355.x

Berg T (2009) Landraces and folk varieties: a conceptual reappraisal of terminology. Euphytica 166:423–430. doi:10.1007/s10681-008-9829-8

Cao J, Schneeberger K, Ossowski S et al (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nat Genet 43:956–963. doi:10.1038/ng.911

Choudhary SB, Sharma HK, Kumar AA et al (2017) SSR and morphological trait based population structure analysis of 130 diverse flax (*Linum usitatissimum* L.) accessions. C R Biol 340:65–75. doi:10.1016/j.crvi.2016.12.002

Cloutier S, Miranda E, Ward K et al (2012a) Simple sequence repeat marker development from bacterial artificial chromosome end sequences and expressed sequence tags of flax (*Linum usitatissimum* L.). Theor Appl Genet 125:685–694. doi:10.1007/s00122-012-1860-4

Cloutier S, Ragupathy R, Miranda E et al (2012b) Integrated consensus genetic and physical maps of flax (*Linum usitatissimum* L.). Theor Appl Genet 125:1783–1795. doi:10.1007/s00122-012-1953-0

de Vicente M (2005) Gene flow and germplasm management. In: Topical reviews in agricultural biodiversity. International Plant Genetic Resources Institute, Rome

De Wit P, Pespeni MH, Ladner JT et al (2012) The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. Mol Ecol Resour 12:1058–1067. doi:10.1111/1755-0998.12003

DeFaveri J, Viitaniemi H, Leder E, Merilä J (2013) Characterizing genic and nongenic molecular markers: comparison of microsatellites and SNPs. Mol Ecol Resour 13:377–392. doi:10.1111/1755-0998.12071

Diederichsen A (2007) Ex situ collections of cultivated flax (*Linum usitatissimum* L.) and other species of the genus *Linum* L. Genet Resour Crop Evol 54:661–678. doi:10.1007/s10722-006-9119-z

European Commission (2016) Common catalogue of varieties of agricultural plant species—35th complete edition. OJ C 478, 21.12.2016, pp 1–794

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

Excoffier L, Lischer HEL (2010) Arlequin suite ver. 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol. doi:10.1111/j.1755-0998.2010.02847.x

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47

FAO (2011) Second global plan of action for plant genetic resources for food and agriculture. FAO, Rome, p 96

Flint-Garcia SA (2013) Genetics and consequences of crop domestication. J Agric Food Chem 61:8267–8276. doi:10.1021/jf305511d

Francis RM (2016) Pophelper: an r package and web app to analyse and visualize population structure. Mol Ecol Resour. doi:10.1111/1755-0998.12509

Fu Y-B (2015) Understanding crop genetic diversity under modern plant breeding. Theor Appl Genet 128:2131–2142. doi:10.1007/s00122-015-2585-y

Fu Y-B, Diederichsen A, Richards KW, Peterson G (2002) Genetic diversity within a range of cultivars and landraces of flax (*Linum usitatissimum* L.) as revealed by RAPDs. Genet Resour Crop Evol 49:167–174. doi:10.1023/A:1014716031095

Fu Y-B, Diederichsen A, Allaby RG (2012) Locus-specific view of flax domestication history. Ecol Evol 2:139–152. doi:10.1002/ece3.57

Galluzzi G, Eyzaguirre P, Negri V (2010) Home gardens: neglected hotspots of agro-biodiversity and cultural diversity. Biodivers Conserv 19:3635–3654. doi:10.1007/s10531-010-9919-5

Gepts P (2003) Crop domestication as a long-term selection experiment. In: Plant breeding reviews. Wiley, New York, pp 1–44

Goldberg CS, Waits LP (2010) Quantification and reduction of bias from sampling larvae to infer population and landscape genetic structure. Mol Ecol Resour 10:304–313. doi:10.1111/j.1755-0998.2009.02755.x

Gorenflo LJ, Romaine S, Mittermeier RA, Walker-Painemilla K (2012) Co-occurrence of linguistic and biological diversity in biodiversity hotspots and high biodiversity wilderness areas. Proc Natl Acad Sci 109:8032–8037. doi:10.1073/pnas.1117511109

Govindaraju DR (1989) Variation in gene flow levels among predominantly self-pollinated plants. J Evol Biol 2:173–181. doi:10.1046/j.1420-9101.1989.2030173.x

Guichoux E, Lagache L, Wagner S et al (2011) Current trends in microsatellite genotyping. Mol Ecol Resour 11:591–611. doi:10.1111/j.1755-0998.2011.03014.x

Habibollahi H, Noormohammadi Z, Sheidai M, Farahani F (2015) Genetic structure of cultivated flax (*Linum usitatissimum* L.) based on retrotransposon-based markers. Genetika 47:1111–1122

Hale ML, Burg TM, Steeves TE (2012) Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. PLoS ONE 7:e45170. doi:10.1371/journal.pone.0045170

Herbig C, Maier U (2011) Flax for oil or fibre? Morphometric analysis of flax seeds and new aspects of flax cultivation in Late Neolithic wetland settlements in southwest Germany. Veg Hist Archaeobot 20:527–533. doi:10.1007/s00334-011-0289-z

Hodel RGJ, Segovia-Salcedo MC, Landis JB et al (2016) The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century. Appl Plant Sci 4:1600025. doi:10.3732/apps.1600025

Hohenlohe PA, Catchen J, Cresko WA (2012) Population genomic analysis of model and nonmodel organisms using sequenced RAD tags. In: Pompanon F, Bonin A (eds) Data production and analysis in population genomics: methods and protocols. Humana Press, Totowa, pp 235–260

Jhala AJ, Bhatt H, Topinka K, Hall LM (2011) Pollen-mediated gene flow in flax (*Linum usitatissimum* L.): can genetically engineered and organic flax coexist? Heredity 106:557–566. doi:10.1038/hdy.2010.81

Kalinowski ST (2011) The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. Heredity 106:625–632

Kilian B, Graner A (2012) NGS technologies for analyzing germplasm diversity in genebanks. Brief Funct Genom. doi:10.1093/bfgp/elr046

Krieger N (2012) Who and what is a "population"? Historical debates, current controversies, and implications for understanding "population health" and rectifying health inequities. Milbank Q 90:634–681. doi:10.1111/j.1468-0009.2012.00678.x

Kumar S, You FM, Cloutier S (2012) Genome wide SNP discovery in flax through next generation sequencing of reduced representation libraries. BMC Genom 13:684. doi:10.1186/1471-2164-13-684

Kumar S, You FM, Duguid S et al (2015) QTL for fatty acid composition and yield in linseed (*Linum usitatissimum* L.). Theor Appl Genet 128:965–984. doi:10.1007/s00122-015-2483-3

Mader E, Hopwood J (2013) Pollinator management for organic seed producers. The Xerces Society, Portland, 28 pp

Maffi L (2005) Linguistic, cultural, and biological diversity. Annu Rev Anthropol 34:599–617. doi:10.1146/annurev.anthro.34.081804.120437

Mondini L, Noorani A, Pagnotta AM (2009) Assessing plant genetic diversity by molecular tools. Diversity. doi:10.3390/d1010019

Nag S, Mitra J, Karmakar PG (2015) An overview on flax (*Linum usitatissimum* L.) and its genetic diversity. Int J

Agric Environ Biotechnol 8:805–8017. doi:10.5958/2230-732X.2015.00089.3
National Inventory (2015, 2017) http://www.genebank.at. Accessed 13 May 2015, 5 Apr 2017
Negri V (2005) Agro-biodiversity conservation in Europe: ethical issues. J Agric Environ Ethics 18:3–25. doi:10.1007/s10806-004-3084-3
ÖPUL (2015) https://www.bmlfuw.gv.at/land/laendl_entwicklung/oepul/oepul2015.html. Accessed 27 Dec 2016
Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in excel. Population genetic software for teaching and research—an update. Bioinformatics 28:2537–2539. doi:10.1093/bioinformatics/bts460
Peterman W, Brocato ER, Semlitsch RD, Eggert LS (2016) Reducing bias in population and landscape genetic inferences: the effects of sampling related individuals and multiple life stages. PeerJ 4:e1813. doi:10.7717/peerj.1813
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
Ragupathy R, Rathinavelu R, Cloutier S (2011) Physical mapping and BAC-end sequence analysis provide initial insights into the flax (*Linum usitatissimum* L.) genome. BMC Genom 12:217. doi:10.1186/1471-2164-12-217
Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
Sathya A (2014) The art of naming traditional rice varieties and landraces by ancient tamils. Asian Agri-Hist 18(1):5–21
Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol 18:233–234. doi:10.1038/72708
Smýkal P, Bačová-Kerteszová N, Kalendar R et al (2011) Genetic diversity of cultivated flax (*Linum usitatissimum* L.) germplasm assessed by retrotransposon-based markers. Theor Appl Genet 122:1385–1397. doi:10.1007/s00122-011-1539-2
Soto-Cerda BJ, Maureira-Butler I, Muñoz G et al (2012) SSR-based population structure, molecular diversity and linkage disequilibrium analysis of a collection of flax (*Linum usitatissimum* L.) varying for mucilage seed-coat content. Mol Breed 30:875–888. doi:10.1007/s11032-011-9670-y
Soto-Cerda BJ, Diederichsen A, Ragupathy R, Cloutier S (2013) Genetic characterization of a core collection of flax (*Linum usitatissimum* L.) suitable for association mapping studies and evidence of divergent selection between fiber and linseed types. BMC Plant Biol 13:78. doi:10.1186/1471-2229-13-78
Soto-Cerda BJ, Diederichsen A, Duguid S et al (2014) The potential of pale flax as a source of useful genetic variation for cultivated flax revealed through molecular diversity and association analyses. Mol Breed 34:2091–2107. doi:10.1007/s11032-014-0165-5
STATISTIK AUSTRIA (2017) http://www.statistik.at. Accessed 5 Apr 2017
Stierschneider M, Gaubitzer S, Schmidt J, et al (2016) The Evoltree Repository Centre—a central access point for reference material and data of forest genetic resources. In: Evolution of trees and forest communities: ten years of the Evoltree network. PG Edition, Bordeaux, pp 15–19
Tamura K, Stecher G, Peterson D et al (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729. doi:10.1093/molbev/mst197
Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063. doi:10.1126/science.277.5329.1063
van der Beek JG, Verkerk R, Zabel P, Lindhout P (1992) Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: Cf9 (resistance to *Cladosporium fulvum*) on chromosome 1. Theor Appl Genet 84:106–112. doi:10.1007/BF00223988
Vogl-Lukasser B, Falschlunger G, Blauensteiner P, Vogl CR (2007) Erfahrungswissen über Lokalsorten traditioneller Kulturarten in Ost- und Nordtirol. Department für Nachhaltige Agrarsysteme, Universität für Bodenkultur, Wien
Wischnitzki E, Burg K, Berenyi M, Sehr EM (2016) Selecting hypomethylated genomic regions using MRE-Seq. In: Hehl R (ed) Plant synthetic promoters. Humana Press, New York City, pp 83–102
Zalapa JE, Cuevas H, Zhu H et al (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am J Bot 99:193–208. doi:10.3732/ajb.1100394