Lactuca georgica is a wild species belonging to the secondary lettuce gene pool: additional evidence, obtained by KASP genotyping

Alex Beharav (abeharav@univ.haifa.ac.il)
Institute of Evolution, University of Haifa https://orcid.org/0000-0003-1707-0606

Research Article

Keywords: Conservation, Gene bank, Genetic diversity, Genotyping platform, Single nucleotide polymorphisms (SNPs), Wild Lactuca relatives (WLRs)

Posted Date: May 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1482340/v1

Version of Record: A version of this preprint was published at Genetic Resources and Crop Evolution on December 27th, 2022. See the published version at https://doi.org/10.1007/s10722-022-01502-7.

Abstract

This work evaluated the genetic relationships between 442 single-seed descent (SSD) accessions, representing eight *Lactuca* spp., including five wild *Lactuca* relatives (WLRs) (*Lactuca georgica*, *L. altaica*, *L. saligna*, *L. serriola*, *L. aculeata*), *L. tuberosa*, *L. undulata*, and the domesticated lettuce, *L. sativa*, most of them (437) representing a core subset of the Institute of Evolution's Wild Lettuce Gene Bank (IoE's WLGB) collection. The analysis was performed by profiling 115 single-nucleotide polymorphism (SNP) markers by means of the fluorescent KASP genotyping assay. The KASP marker fragments were scored as either allele “A” or allele “B”, that were used across analyses as bi-allelic data, but included relatively a lot of U-scores, i.e., an absence of the specific sequence, that were treated as missing data. Often U-scores were specific for a certain species. Data analysis of the five WLRs showed that allele frequencies of 103 (97.2%) out of 106 differentiating loci varied significantly among the species, where 59.7% of the KASP marker diversity was between species. A neighbor-network analysis between samples belonging to the five WLRs and a single *L. sativa* cv. clearly clustered all 430 samples in accordance with their taxonomic determination. The results obtained here via multiple complementary analyses of large natural populations and individuals for germplasm variation, question assignment of *L. georgica* to the primary lettuce gene pool (LGP1). Together with our previous results obtained using TRAP markers, and hybridization experiments, we conclude that *L. georgica* is a constituent of the LGP2.

Introduction

The genus *Lactuca* L. [Compositae (Asteraceae), tribe Cichorieae, subclade Lactucinae is comprised of 123 accepted species (WFO 2022), which are primarily found in the Northern Hemisphere. The domesticated species in the genus, *Lactuca sativa* L. (lettuce), is one of the most important and widely distributed leafy vegetables around the world (Beharav 2020, and literature cited therein). Domestication has resulted in limited genetic variation, rending crop vulnerable to diseases, pests, and environmental stresses. Therefore, breeders have stimulated the use of gene banks germplasm to promote food security and sustainable agricultural production (van Treuren et al. 2013).

Development of efficient conservation strategies to maintain genetic variability of crop progenitors is key to genetic resource management of plants. To achieve this goal, it is crucial to understand the genetic structure of progenitor species at both the population and species levels (Kitner et al. 2008). Inter- and intraspecies distributions of genetic diversity should serve as driving forces for the collecting strategy. The Institute of Evolution's Wild Lettuce Gene Bank (IoE's WLGB) recently set out to characterize the population structure of wild *Lactuca sativa* relatives (WLRs) originating from natural habitats in Southwest Asia, that is the center of diversity for WLRs (Zohary 1991). Studies primarily based on our new and extensive collections of five WLRs: *L. serriola* L., *L. aculeat*a Boiss., *L. georgica* Grossh., *L. altaica* Fisch. & C.A. Mey., and *L. saligna* L. (Beharav and Hellier 2020; Beharav et al. 2020, and literature cited therein; Beharav 2021), strongly support exploitation of WLRs as rich genetic sources for lettuce improvement.

Our previous studies extensively evaluated the genetic relationships and structured diversity of most species in our WLRs collections (Kitner et al. 2008, 2015; Lebeda et al. 2012; Jemelková et al. 2015; Beharav et al. 2018a, b). According to Zohary (1991), with the except of *L. saligna*, all the species represent the primary lettuce Gene Pool (LGP1), while *L. saligna* represents the LGP2. However, the high genetic distance between *L. georgica* and *L. sativa* samples in our recent study (Beharav et al. 2018a), as well as initial crosses between *L. georgica* and *L. sativa* which displayed only partial levels of interfertility (Beharav, personal communication), question the assignment of *L. georgica* to LGP1 (Zohary 1991; Lebeda et al. 2007; Gabrielian and Zohary 2004). A similar conclusion was reached following a recently reported phytochemical study (van Treuren et al. 2018) and whole-genome resequencing (Wei et al. 2021).

The competitive allele-specific PCR (currently called Kompetitive Allele Specific PCR, or KASP™) genotyping assay utilizes a unique form of competitive allele-specific PCR combined with a novel, homogeneous, fluorescence-based reporting system for the identification and measurement of genetic variation occurring at the nucleotide level to detect single nucleotide polymorphisms (SNPs) or inserts and deletions (InDels) (He et al. 2014). The KASP technology is suitable for use on a variety of equipment platforms and provides flexibility in terms of the number of SNPs and the number of samples able to be analyzed. The KASP chemistry functions equally well in 96-, 384-, and 1,536-well microtiter plate formats (Livak et al. 1995) and has been extensively utilized in both small- and large-scale human, animal, and plant genetic laboratories. The KASP assay can distinguish two alleles of a certain locus in a co-dominant manner. If the genotype at a given SNP is homozygous, only one of the two possible fluorescent signals will be generated. But, if the genotype is heterozygous, a mixed fluorescent signal will be generated (Semagn et al. 2014). When compared to several sequence-based markers, KASP has been reported to improve cost-effectiveness and reliability (Shikari et al. 2021, and literature cited therein). In the present study, the KASP assay technique was applied to gain insights into the genetic relationship between the untapped genetic resources of *L. georgica* and samples representing four other WLRs: *L. serriola*, *L. altaica, L. aculeata*, and *L. saligna*). All five mentioned species belong to the section *Lactuca* L., subsection *Lactuca* L. (Lebeda et al. 2007). We also included: some samples of *Lactuca tuberosa* Jacq. [syn. *Steptorhamphus tuberosus* (Jacq.) Grossh.], a wild edible plant species in Jordan (Stojakowska et al. 2013, and literature cited therein); a single sample of *Lactuca undulata* Lebed. belongs to the section *Micranthae* Boiss. (Lebeda et al. 2007); and a single sample of domesticated lettuce, *L. sativa* (section *Lactuca* L., subsection *Lactuca* L.). The present study aimed to shed additional light on the position of the putative LGP1 *L. georgica* within the population structure of the main WLRs, and to discuss the feasibility of using germplasm of this WLR for crop promotion and improvement.

Materials And Methods

Plant material

A unique core subset of 437 single-seed descent (SSD), representing the regular collection of the IoE's WLGB collection, each referred to as a sample (accession), representing seven wild *Lactuca* species collected in natural habitats, were used for this study (Tables 1, 2): *L. georgica* (*N* = 43) and *L. altaica* (*N* = 12) samples were collected from six and five localities, respectively, throughout Armenia, between August 30 and September 4, 2011; *L. saligna* (*N* = 168), *L. serriola* (*N* = 40), *L. aculeata* (*N* = 162), and *L. tuberosa* (*N* = 11) samples were collected from 32, 23, 21, and 9 localities, respectively, most of them throughout Israel, between 2003 to 2014. The exceptions were: five *L. saligna* and a single *L. serriola* samples that were collected from two localities in Italy in 2004, a

single *L. serriola* and eight *L. aculeata* samples that were collected from two localities in Jordan in 1996, and ten *L. serriola* samples that were collected from five localities in Armenia in 2011; and A single *L. undulata* sample that was collected in Wadi Mujib area, Moab province, Jordan, in 2012. All those wild accessions are documented in the *Lactuca* database at the IoE (IoE'sLDB) and deposited in the seed storage facilities of the IoE's WLGB. A single sample from each of the wild species: *L. georgica, L. saligna*, *L. serriola*, and *L. aculeata*, and a single sample representing the cultivated lettuce, cv. Salinas M. were added from the collection of Rijk Zwaan B.V. (De Lier, The Netherlands) to the germplasm set of the present study. Altogether, 442 seed samples were included in the germplasm set of the present study.

Table 1
List of 437 samples representing seven wild *Lactuca* species from the IoE'sWLGB, used for KASP analysis

Lactuca species	Pop. no. [a]	N	Country	District	Locality	Ln	Lt	El
L. georgica	433	8	Armenia	Ararat	Geghard I	44°48′51.2"E	40°08′13.7"N	1722
	434	5	Armenia	Ararat	Geghard II	44°49′10.7"E	40°08′25.7"N	1757
	441	7	Armenia	Gegharkunik	Tsovagiugh I	44°58′15.7"E	40°37′31.1"N	1915
	443	6	Armenia	Gegharkunik	Tsovagiugh III	45°00′22.0"E	40°37′37.5"N	1920
	444	3	Armenia	Gegharkunik	Tsovagiugh IV	45°03′25.7"E	40°36′39.7"N	1915
	446	14	Armenia	Aragatsotn	Hamberd	44°16′05.3"E	40°22′28.7"N	1964
L. serriola	224	1	Jordan	Ammon province	Amman, 10 km S			
	282	1	Israel	Southern Mt. Carmel	Zikhron Ya'aqov-1	34°57′31"E	32°34′07"N	143
	320	1	Italy		Bettole Este			
	335	1	Israel	Upper Galilee	Sasa	35°23′54"E	33°01′41"N	811
	336	2	Israel	Lower Galilee	Lavi	35°27′40"E	32°47′23"N	189
	338	1	Israel	Golan Heights	Nov, Haspin	35°47′19"E	32°49′40"N	407
	339	3	Israel	Philistean Plain	Zafriyya	34°50′51"E	32°00′50"N	30
	341	2	Israel	Upper Galilee	Tarshiha, Me'ona	35°16′14"E	33°00′51"N	488
	342	1	Israel	Golan Heights	Giv'at Yo'av	35°41'22"E	32°48'08"N	333
	361	1	Israel	Golan Heights	Nahal Sa'ar	35°42′17"E	33°14′23"N	463
	365	1	Israel	Sharon Plain	Avi'el	34°58′21.5"E	32°32′22.6"N	32
	370	3	Israel	Galilee Panhandle	Metulla	35°34′34"E	33°16′43"N	506
	377	2	Israel	Lower Galilee	Kaukab Abu El Hija	35°14'56.6"E	32°50'12.1"N	375
	390	1	Israel	Sharon Plain	Binyamina	34°56′58.46"E	32°31′38.27"N	12
	395	2	Israel	Upper Galilee	Rosh-Pinna-1	35°32'26.8"E	32°58'23.7"N	378
	396	1	Israel	Upper Galilee	Zomet Rosh-Pinna	35°33′20.50"E	32°58′12.02"N	336
	430	1	Armenia	Ararat	Charentsi Arc SFS	44°38′11.9"E	40°10′29.2"N	1665
	431	1	Armenia	Kotayq	Road to Garni	44°41′03.6"E	40°07′51.7"N	1460
	438	5	Armenia	Gegharkunik	Sevan I	44°59′08.5"E	40°34′55.0"N	1910
	439	1	Armenia	Gegharkunik	Sevan II	44°57′26.8"E	40°36′07.4"N	1923
	445	1	Armenia	Aragatsotn	Ashtarak	44°25′47.1"E	40°16′19.6"N	1917
	446	1	Armenia	Aragatsotn	Hamberd	44°16′05.3"E	40°22′28.7"N	1964
	450	1	Israel	Upper Galilee	Zefat-1	35°30'53.7"E	32°57'38.1"N	760
	451	5	Israel	Upper Galilee	'En Lior	35°29′50.4"	32°58′37.3"N	732
L. aculeata	224	4	Jordan	Ammon province	Amman, 10 km S			
	232	1	Jordan	Ammon province	Amman, 10 km S			
	233	3	Jordan	Ammon province	Mafraq Junction, 35 km N Zarka			
	344	12	Israel	Golan Heights	Nov, Zomet	35°47'19"E	32°49'40"N	407
	366	4	Israel	Golan Heights	Gamla, next to Zomet	35°46'15"E	32°54'40"N	469
	367	4	Israel	Golan Heights	Hamappalim, Zomet	35°45'01"E	32°59'14"N	524
	368	5	Israel	Golan Heights	En Ziwan, close to	35°49'17"E	33°06'21"N	968
	371	6	Israel	Golan Heights	Kela' Alone	35°41'04"E	33°07'58"N	644
	372	4	Israel	Golan Heights	Hashiryon, Zomet	35°44'39"E	33°03'44"N	715

[a] Pop. no. = Population number corresponds to IoE'sLDB; N = Number of accessions; Ln = longitude; Lt = latitude; El = elevation (m a. s. l.).

Lactuca species	Pop. no. [a]	N	Country	District	Locality	Ln	Lt	El
	373	12	Israel	Upper Galilee	Elifelet	35°32'32"E	32°57'05"N	357
	374	5	Israel	Hula Plain	Mishmar-Hayarden	35°36'10"E	33°00'11"N	236
	375	8	Israel	Golan Heights	Qidmat-Zevi-1	35°41'46"E	33°01'21"N	432
	376	6	Israel	Golan Heights	Qidmat-Zevi-2	35°43'13"E	33°02'18"N	581
	380	2	Israel	Golan Heights	Nov	35°47′16.7"E	32°49'35.0"N	406
	395	10	Israel	Upper Galilee	Rosh-Pinna-1	35°32'26.8"E	32°58'23.7"N	378
	448	2	Israel	Upper Galilee	Rosh-Pinna-2	35°32'24.0"E	32°58'22.8"N	384
	449	6	Israel	Upper Galilee	Ramat Razim	35°31'18.3"E	32°57'21.9"N	677
	450	5	Israel	Upper Galilee	Zefat-1	35°30'53.7"E	32°57'38.1"N	760
	452	8	Israel	Upper Galilee	En-Zetim, Zomet	35°29'01.5"E	32°59'21.4"N	691
	453	43	Israel	Golan Heights	Haspin	35°47'34"E	32°50'31"N	420
	454	12	Israel	Golan Heights	Nov, Zomet	35°47′19"E	32°49'40"N	407
L. altaica	429	2	Armenia	Ararat	Charentsi Arc NFS	44°38′12.1"E	40°10′25.1"N	1655
	437	5	Armenia	Vayots Dzor	Shatin	45°18′22.9"E	39°50′29.6"N	1303
	438	1	Armenia	Gegharkunik	Sevan I	44°59′08.5"E	40°34′55.0"N	1910
	439	2	Armenia	Gegharkunik	Sevan II	44°57′26.8"E	40°36′07.4"N	1923
	446	2	Armenia	Aragatsotn	Hamberd	44°16′05.3"E	40°22′28.7"N	1964
L. saligna	317	3	Israel	Mt. Carmel	Nahal-Oren			
	320	3	Italy		Bettole Este			
	321	2	Italy	North	Sale			
	328	10	Israel	Philistean Plain	Zafriyya	34°50′51"E	32°00′50"N	30
	329	2	Israel	Philistean Plain	Nahalat-Yehuda	34°47′53"E	31°58′52"N	39
	330	1	Israel	Shefela	Lod	34°53′46"E	31°56′36"N	61
	331	4	Israel	Esdraelon Plain	Megiddo	35°11′05"E	32°34′21"N	123
	332	4	Israel	Lower Galilee	Kefar-Tavor	35°25′06"E	32°41′31"N	131
	333	8	Israel	Upper Galilee	Tarshiha, Me'ona	35°16′14"E	33°00′51"N	488
	334	11	Israel	Upper Galilee	Hurfeish	35°20′56"E	33°00′58"N	647
	335	10	Israel	Upper Galilee	Sasa	35°23′54"E	33°01′41"N	811
	336	8	Israel	Lower Galilee	Lavi	35°27′40"E	32°47′23"N	189
	337	6	Israel	Golan Heights	Giv'at Yo'av	35°41′22"E	32°48′08"N	336
	338	10	Israel	Golan Heights	Nov, Haspin	35°47′19"E	32°49′40"N	407
	348	7	Israel	Philistean Plain	Ramat-Gan	34°50′30"E	32°03′55"N	46
	349	6	Israel	Philistean Plain	Ashdod, Mehlaf	34°42′27"E	31°50′05"N	31
	350	5	Israel	Philistean Plain	Yad-Mordehay	34°33′26"E	31°34′59"N	33
	352	5	Israel	Philistean Plain	Tel-Aviv	34°47′55"E	32°03′54"N	30
	353	9	Israel	Lower Galilee	Tur'an	35°21′52"E	32°45′42"N	180
	354	8	Israel	Upper Galilee	Zomet Korazim	35°32′51"E	32°54′23"N	152
	355	2	Israel	Upper Galilee	Rosh-Pinna	35°32′49"E	32°58′18"N	344
	356	5	Israel	Hula Plain	Zomet Hagome	35°34′15"E	33°10′19"N	75
	357	10	Israel	Galilee Panhandle	Metulla	35°34′34"E	33°16′43"N	506
	358	1	Israel	Galilee Panhandle	Kefar-Gil'adi	35°34′38"E	33°14′20"N	293

[a] Pop. no. = Population number corresponds to IoE'sLDB; N = Number of accessions; Ln = longitude; Lt = latitude; El = elevation (m a. s. l.).

Lactuca species	Pop. no. [a]	N	Country	District	Locality	Ln	Lt	El
	359	3	Israel	Hula Plain	Zomet Tel-Dan	35°39′36"E	33°14′29"N	193
	360	1	Israel	Upper Galilee	Rhajar	35°39′30"E	33°15′40"N	259
	361	2	Israel	Golan Heights	Nahal Sa'ar	35°42′17"E	33°14′23"N	463
	369	5	Israel	Upper Galilee	Metulla	35°34'33.7"E	33°16'37.4"N	485
	378	10	Israel	Coastal Galilee	Judeida-Makr			
	379	1	Israel	Sharon Plain	Binyamina	34°56′49.3"E	32°31′41.9"N	11
	395	3	Israel	Upper Galilee	Rosh-Pinna-1	35°32'26.8"E	32°58'23.7"N	378
	449	3	Israel	Upper Galilee	Ramat Razim	35°31'18.3"E	32°57'21.9"N	677
L. tuberosa	381	4	Israel	Mt. Carmel	Ramat HaNadiv	34°57'04"E	32°33'25"N	121
	384	1	Israel	Esdraelon Plain	Qiryat-Tiv'on	35°06'57"E	32°42'50"N	110
	387	1	Israel	Coastal Galilee	Tel Yas'ur	35°10'11"E	32°54'08"N	36
	394	1	Israel	Lower Galilee	Shekhanya	35°14'42"E	32°51'07"N	416
	398	1	Israel	Sharon Plain	Netania	34°52'28"E	32°18'21"N	30
	399	1	Israel	Esdraelon Plain	Afula-'Illit	35°18'41"E	32°37'57"N	116
	403	1	Israel	Philistean Plain	Bene-Beraq	34°49'52"E	32°05'24"N	55
	412	1	Israel	Mt. Carmel	Nahal Kelah	35°00'54"E	32°44'17"N	347
L. undulata	512	1	Jordan	Moab province	Wadi Mujib area			

[a] Pop. no. = Population number corresponds to IoE'sLDB; N = Number of accessions; Ln = longitude; Lt = latitude; El = elevation (m a. s. l.).

Details of identification, propagation, cultivation, characterization, and re-determination of most sampled accessions are described in Beharav et al. (2018b) and Beharav (2021). Specific morphological observations supported the species identity of the *L. georgica* (Beharav et al. 2018a), *L. altaica* (Beharav et al. 2020), *L. saligna* (Beharav et al. 2008), *L. serriola* (Beharav et al. 2018b), *L. aculeata* (Beharav et al. 2010, 2018b), and *L. tuberosa* (Stojakowska et al. 2013).

DNA extraction, KASP Markers, and Data scoring

Leaf samples took place by pooling four leaf discs (diameter 5.5 mm) from true leaves of a single plant per sampled accession. These leaf samples were freeze dried and sent to LGC genomics (Hoddesdon, UK). DNA-extraction was performed by use of an sbeadex purification kit (Biosearch Technologies https://www.biosearchtech.com/products/extraction-and-purification-reagents/dna-purification-kits/sbeadex-kits). The KASP-assays were run by LGC, based on SNPs indicated by Rijk Zwaan. Genome sequences data were used from a set of *L. sativa* and *L. serriola* samples (BioSample for BioProject (Select 478460) - BioSample - NCBI (nih.gov; BioSample Links for BioProject (Select 510128) - BioSample - NCBI (nih.gov) to identify SNPs markers by 5,000 different KASP-assays, that were tested on a small number of accessions of the involved species in the present study. Then, 115 informative assays, i.e., assays that distinguish between and within species, based on the sequence variation in the *L. sativa* and *L. serriola* set, were chosen for this study. Such a marker is genotyping was based on graphs with fluorescence scores for both alleles (Semagn et al. 2014). The KASP marker fragments were scored either allele "A" or allele "B", so they scored mainly homozygous, while a very low percentage of heterozygous (H-score) was observed, as expected for predominantly self-pollinating species, such as many *Lactuca* species (Kitner et al. 2015, and literature cited therein). Relatively a lot of U-scores were observed, which can be explained by absence of the sequence to support both KASP probes in the specific assay accession. Often U-scores were specific for a certain species.

Data analysis

A detailed list of wild *Lactuca* species and the number of samples used for the various analyses is presented in Table 2. Both A-scores and B-scores of the original KASP dataset were used across analyses as bi-allelic data, while U-scores were treated as missing data.

Table 2

List of wild *Lactuca* species and number of samples (N) used for various analyses of results of 115 KASP bi-allelic SNP genotyping assays

	- List of IoE's WLGB samples (Table 1)	Genetic diversity (GD) or Basic information (BI) (Table 3)	G^2 tests and *Fst* analysis (Table 4)	Pairwise dissimilarity (Table 5), Neighbor-Network (Fig. 1) and Structure analysis (Fig. 2)
Species	N			
L. georgica	43	44 (GD)	44	44
L. serriola	40	41 (GD)	41	41
*L. aculeat*a	162	163 (GD)	163	163
L. altaica	12	12 (GD)	12	12
L. saligna	168	169 (GD)	169	169
L. tuberosa	11	11 (BI)		
L. undulata	1	1 (BI)		
L. sativa		1 (BI)		1
Total N	437	442	429	430

Genetic diversity of the five WLRs - *L. georgica* (*N* = 44), *L. serriola* (*N* = 41), *L. aculeata* (*N* = 163), *L. saligna* (*N* = 169), and *L. altaica* (*N* = 12), i.e., total of 429 sampled accessions - was analyzed at the species levels using Popgene ver 1.32 (Yeh et al. 1999), under a co-dominant mode of inheritance in a diploid organism. The number of unique alleles, percentage of polymorphic loci (polymorphism (*P*) at 1% level), observed heterozygosity (*Ho*), gene diversity (*He*; Nei 1973) that is equal to the expected heterozygosity under random mating (Hardy–Weinberg equilibrium), and unbiased gene diversity index (*UHe*; Nei 1978), were estimated. The average random mating or out-crossing rate (*t*) was calculated from the polymorphic loci at each species by the formula: $t = (1 - Fe)/ (1 + Fe)$ (Crow and Kimura, 1970), where *Fe* equals: 1 - *Ho*/ *He* (i.e., *Fe* represent the equilibrium inbreeding coefficient under partial selfing). Likelihood ratio (G^2) test for homogeneity of allele frequencies across species and F-Statistics (*Fst*; Hartl and Clark 1997) were estimated for the polymorphic loci.

To visualize the genetic relationships among the analyzed samples, a Neighbor-Network based on the uncorrected p-distance was constructed in SplitsTree 4 (Huson and Bryant, 2006). Neighbor-Network is a variant of Neighbor-Joining (NJ) tree, which constructs phylogenetic networks instead of phylogenetic trees (Bryant and Moulton 2004). A Nexus input file was exported from GenAlEx ver. 6.5 (Peakall and Smouse 2006) into the SplitsTree. The Nexus file included a binary data set, based on A-scores and B-scores of 430 individual samples (details in Table 2): the 429 samples of the five WLRs and the single *L. sativa* sample, but without those of *L. tuberosa* and *L. undulata*, that contained a highly frequency of U-scores. In our case, uncorrected p-distance represents the proportion (*p*) of compared loci carrying different scored allele. It was obtained by dividing the number of loci with different alleles by the total number of loci compared. In fact, it was equal to 1 - Simple matching similarity coefficient for binary data set, seems to us the most adequate for determining similarity coefficient in predominantly autogamous species. The reliability and robustness of the network were tested by bootstrap analysis with 1000 replicates.

The Bayesian approach in STRUCTURE 2.3.4 software (Pritchard et al. 2009) was also implemented as a model-based clustering method for inferring population structure using genotype data consisting of unlinked markers (Pritchard et al. 2000). The method enabled delineation of the optimal number of K clusters ("populations") of individuals based on their genotypes at multiple loci (Evanno et al. 2005). Notably, the main modeling assumptions of STRUCTURE are Hardy-Weinberg equilibrium within populations and complete linkage equilibrium between loci within populations (Prichard et al. 2000). However, this was not the case in our dataset, where scores were mostly homozygous, as expected for predominantly self-pollinating *Lactuca* species. Nevertheless, the 'admixture model' was used, with correlated allele frequencies of the co-dominant KASP data of the same 430 individual samples that were exposed to the Neighbor-Network construction (details in Table 2). The general 'admixture model' that estimates admixture proportions for each individual was preferred allowing one to identify admixed individuals represented by a proportional mixture of two or more signals characteristic of the various clusters (Pritchard et al. 2000). The K was set 2 to 8 with ten independent runs for each K using 1,000,000 Markov Chain Monte Carlo (MCMC) iterations following the period of 100,000 burn-in iterations. STRUCTURE results were averaged using CLUMPP version 1.1.2 (Jakobsson & Rosenberg 2007) and plotted with the aid of DISTRUCT version 1.1 (Rosenberg 2004).

Results

KASP genetic diversity

A summary of the genetic diversity for each of the five WLRs, as defined based on the 115 KASP loci, is presented in Table 3. Of the 115 KASP markers, 113 (98.3%), 113 (98.3%), and 108 (93.9%) resulted in fragments that were scored as either allele "A" or allele "B", at least in a single sample out of the sampled *L. serriola*, *L. altaica*, and *L. aculeata* LGP1 accessions, respectively. Only 97 (84.3%) and 92 (80.0%) out of the 115 KASP markers resulted in fragments that were scored as either allele "A" or allele "B", at least in a single sample out of all sampled *L. saligna* (LGP2) and *L. georgica*, respectively, i.e., 18 (15.7%) and 23 (20.0%) markers all analysed *L. saligna* and *L. georgica* samples, respectively, resulted in U-scores. U-scores were obtained for 14.0%, 14.1%, and 15.1% of the

L. serriola, *L. aculeata, and L. altaica* data points, respectively, while higher percentages (38.2% and 46.7%) of all the data points for *L. saligna* (LGP2) and *L. georgica* samples, respectively, resulted in U-scores.

Table 3

Summary of genetic diversity of *Lactuca* spp., based on 115 KASP bi-allelic SNP genotyping assays

Species	Sample size [a] Nt Nm	Total SNP Present [b]	U-score (%) [c]	Unique alleles	Polymorphism (P) [d]	Observed heterozygosity (*Ho*) [e]	Gene diversity [e] *He* [f] *UHe* [g]	Out-crossing rate (*t*) [e]
L. georgica	44 29.5	92	46.7	1	20/92 = 21.7	0.012	0.214 0.225	0.030
L. serriola	41 36	113	14.0	1	81/113 = 71.7	0.004	0.363 0.369	0.006
L. aculeata	163 149	108	14.1	1	51/108 = 47.2	0.004	0.330 0.331	0.006
L. altaica	12 10.5	113	15.1	0	62/113 = 54.9	0.017	0.368 0.389	0.023
L. saligna	169 124	97	38.2	2	40/97 = 41.2	0.030	0.266 0.283	0.060
L. tuberosa	11 7	50	72.5	*	2/50 = 4.0	*	*	*
L. undulata	1 0.17	20	82.6	*	**	*	*	*
L. sativa	1 0.88	101	12.2	*	**	*	*	*

[a] Nt = Total sample size in the experiment; Nm = Mean (from all markers where fragments were being scored (present), either allele "A" or allele "B", at least in a single sample) of real sample size

[b] Number of KASP markers where fragments were being scored (present), either allele "A" or allele "B", at least in a single sample

[c] Percentage of data points resulted with U-scores

[d] Percentage of polymorphic loci (number of polymorphic/number with real data) at the 1% level

[e] Mean for whole polymorphic loci

[f] Gene diversity, equivalent to the expected heterozygosity under panmixia (Nei 1973)

[g] Unbiased gene diversity, equivalent to the unbiased expected heterozygosity under panmixia (Nei 1978)

* Not computed, due to very low sample size

** Only single sample size

Two unique alleles were obtained for *L. saligna*, while a single unique allele was obtained for each of three species: *L. georgica*, *L. serriola*, and *L. aculeata*. The polymorphism level (*P*) appeared to be highly significantly ($\chi^2 = 54.6$, $df = 4$, $p < 0.0001$) differed between the five WLRs. The highest polymorphism level was obtained for *L. serriola* (*P* = 71.7%), with the remaining species descending in the order from *L. altaica* (*P* = 54.9%), *L. aculeata* (*P* = 47.2%), *L. saligna* (*P* = 41.2%), while the lower level was obtained for *L. georgica* (*P* = 21.7%). Average gene diversity (*He*) at the species level was 0.308, ranging from 0.214 for *L. georgica* to 0.266 for *L. saligna*, 0.330 for *L. aculeata* and 0.363 for *L. serriola*, while a higher level of diversity was obtained for *L. altaica* (*He* = 0.368). Clearly, slightly higher, but same order unbiased gene diversity values (*UHe*) were obtained.

Scores were mostly homozygous, as expected for predominantly self-pollinating *Lactuca* species. As a result, very low observed heterozygosity (*Ho*) was obtained for all five WLRs, ranging from 0.004 for *L. serriola* and *L. aculeata* to 0.030 for *L. saligna*. Thus, very low out-crossing rate (*t*) values (mean: 0.025 (2.5%)) were calculated for the five species,, ranging from 0.006 (0.6%) for *L. serriola* and *L. aculeata* to 0.060 (6%) for *L. saligna* (Table 3).

Basic information was obtained for the three remaining species included in this study: *L. tuberosa* and *L. undulata*, two wild species that are genetically far from the domesticated lettuce, *L.sativa* (Table 3). Only 50 (43.5%) out of the 115 KASP markers resulted with fragments that were being scored (present), either allele "A" or allele "B", at least in a single sample out of 12 sampled accessions of *L. tuberosa*, i.e., 65 (56.5%) markers resulted with U-scores in all 12 samples. 72.3% out of the whole data points resulted with U-scores for the *L. tuberosa* samples. A single sample that represented *L. undulata* resulted with fragments that were being scored only for 20 (17.4%) out of the 115 markers, i.e., 95 (82.6%) markers resulted with U-scores in this sample. A single sample that represented *L. sativa* resulted with fragments being scored for 101 (87.8%) out of the 115 markers, i.e., only 14 (12.2%) markers resulted with U-scores in this sample.

Pattern of variation of KASP loci

Of the 115 KASP loci, 107 (93.0%) were differentiating among the set of 429 samples representing the five WLRs. (Table 4a). The mean percentage of total differentiating loci derived from all ten interspecies pairs comparisons was 80.5%, ranging from 71 (68.3%) out of the 104 total KASP loci representing all sampled *L. georgica* and *L. saligna* accessions, to 101 (87.8%) out of the 115 total KASP loci representing all *L. georgica* and *L. serriola* samples. Notably, the mean percentage of loci with alleles "A" and/or "B" extracted from only one species out of the total number of differentiating loci derived from the 10 interspecies pairs comparisons was 16.6%, ranging from 0 (0%) of the 80 total differentiating loci representing all samples from *L. serriola* and *L. altaica*, to 19 (26.8%) of the 71 total differentiating loci representing all samples from *L. georgica* and *L. saligna*.

Table 4

Summary of single-locus G^2 tests (a) for homogeneity of alleles ("A" and "B") frequencies, and (b) *Fst* [a] analysis of the differentiating KASP SNP assays among 429 samples of five wild *Lactuca* spp. and among samples of each pairwise species

	Number of loci						
	All five spp.	*L. georgica*– *L. serriola*	*L. georgica*– *L. aculeata*	*L. georgica*– *L. altaica*	*L. georgica*– *L. saligna*	*L. serriola*– *L. aculeata*	*L. serriola*– *L. altaica*
(a) G^2 tests							
Significance							
$p < 0.001$	93	55	53	48	35	75	20
$0.001 < p < 0.01$	0	5	3	6	3	2	5
$0.01 < p < 0.05$	0	7	1	6	1	2	11
Total $p < 0.05$	93	67	57	60	39	79	36
Allele(s) scored in one species	11 *	26	22	24	19	6	0
Sum	104	93	79	84	58	85	36
$0.05 < p$	3	8	8	12	13	11	44
Total differentiating	107	101	87	96	71	96	80
Monomorphic	8	14	24	19	33	18	33
All	115	115	111	115	104	114	113
(b) *Fst* analysis							
Fst value							
0.05 – 0.15	7	16	8	12	11	22	18
0.15 – 0.25	7	10	5	10	5	14	10
> 0.25	92	66	64	67	42	46	10
Total *Fst* > 0.05	106	92	77	89	58	82	38
< 0.05	1	9	10	7	13	14	42
Total differentiating	107	101	87	96	71	96	80
Mean – total differentiating loci	*Fst* = 0.597	*Fst* = 0.484	*Fst* = 0.584	*Fst* = 0.555	*Fst* = 0.526	*Fst* = 0.335	*Fst* = 0.099

	Number of loci			
	L. serriola– *L. saligna*	*L. aculeata*– *L. altaica*	*L. aculeata*– *L. saligna*	*L. altaica*– *L. saligna*
(a) G^2 tests				
Significance				
$p < 0.001$	64	60	63	55
$0.001 < p < 0.01$	4	2	4	6
$0.01 < p < 0.05$	9	7	2	6
Total $p < 0.05$	77	69	69	67
Allele(s) present in one species	15	7	15	16
Sum	92	76	84	83
$0.05 < p$	7	16	6	9
Total differentiating	99	92	90	92
Monomorphic	14	22	20	21
All	113	114	110	113

(b) *Fst* analysis				
Fst value				
0.05 – 0.15	29	19	18	21
0.15 – 0.25	9	9	9	10
> 0.25	48	49	50	53
Total *Fst* > 0.05	86	77	77	84
< 0.05	13	15	13	8
Total differentiating	99	92	90	92
Mean – total differentiating loci	*Fst* = 0.372	*Fst* = 0.400	*Fst* = 0.471	*Fst* = 0.470

[a] *Fst* = Gene diversity between species, relative to total gene diversity (Hartl and Clark, 1997)
* loci that were being scored for only a single allele, but was not found in the all five species

Allele frequencies

The G^2 test for homogeneity of the allele frequencies among the five major species showed 93 (96.9%) out of 96 KASP loci that resulted with fragments being scored (present) for both alleles “A” and “B” varied highly significantly ($p < 0.001$), while only three loci varied not significantly ($0.05 < p$) (Table 4a). Together with 11 loci that were being scored for only a single allele, but was not found in all five species, 104 (97.2%) out of total 107 differentiating loci varied significantly among the five species. The remaining eight loci were monomorphic for a single allele that was scored in all five species.

G^2 test between pairwise-compared species revealed that a total of 36 (45.0%) out of 80 differentiating loci varied significantly ($p < 0.05$) between *L. serriola* and *L. altaica* (for more details see Table 4a). From all other pairwise-compared species, a significantly higher total proportion of significant loci was obtained. It’s ranging from a total of 58 (81.7%) out of 71 differentiating loci (includes 19 loci where a single or both the two alleles were being scored only in one out of the two species) that varied significantly between *L. georgica* and *L. saligna*, to a total of 84 (93.3%) out of 90 differentiating loci (includes 26 loci where a single or both the two alleles were being scored only in one out of the two species) that varied significantly between *L. aculeata* and *L. saligna*. These total proportions of significant loci far exceed the 5% level expected by chance (binomial test, $p < 0.0001$) (Aiken 1955), even for the lowly proportion obtained between *L. serriola* and *L. altaica*, indicating highly significant differences in allele frequencies for all interspecific combinations.

Genetic differentiation within and among species (*Fst* analysis)

The relative degree of gene differentiation (*Fst*) among the five WLRs averaged 0.597 (ranged 0.037-1) for the total of 106 differentiating loci (Table 4b). Namely, 59.7% of the KASP diversity was between the five species, while 40.3% was within species. According Hartl and Clark (1997), *Fst* < 0.05 indicate negligible genetic differentiation, the range of 0.05 to 0.15 for *Fst* may be considered to indicate moderate differentiation, a 0.15 to 0.25 may indicate great differentiation, while *Fst* > 0.25 may indicate very great differentiation. The present data showed moderate, great, and very great differentiation between the

five species at seven (6.54%), seven (6.54%), and 92 (85.98%) out of the 107 differentiating loci, respectively, and negligible differentiation (*Fst* < 0.05) at only a single locus (0.94%).

Fst analysis of each of the ten pairwise species comparisons revealed significantly lowest average KASP diversity (9.9%) between the species combination *L. serriola* - *L. altaica* (Table 4b; Wilcoxon two-sample nonparametric test compared to each of the other combinations; $p < 0.001$). Higher average diversity (33.5%) was observed between the species combination *L. serriola* - *L. aculeata.* The highest average *Fst* values for KASP diversity was noted between *L. georgica* and its combination with the other four species: 48.4%, 52.6%, 55.5%, and 58.4% for *L. georgica* - *L. serriola, L. georgica* - *L. saligna*, *L. georgica* - *L. altaica*, and *L. georgica* - *L. aculeata*, respectively.

Species relationship among and within species

The uncorrected p-distance values for pairwise samples were lower for all intraspecies comparisons than for the interspecies comparisons (Table 5), as expected. The mean value of uncorrected p-distance derived from all intraspecies pairs was 0.240. The highest average uncorrected p-distance from the intraspecies comparisons was obtained between pairwise samples belonging to the species *L. serriola* (0.36) and *L. altaica* (0.34), with the remaining species descending in the order from *L. altaica* (0.20) to both *L. georgica* and *L. saligna* (0.15). There weren't any intraspecies comparisons for *L. sativa*, since this species was represented by only a single sample.

Table 5
Range and average of pairwise uncorrected p-distance within and between six *Lactuca* spp., based on 115 KASP bi-allelic SNP genotyping assays, scored

Species	*L. georgica*			*L. serriola*			*L. aculeata*			*L. altaica*			*L. saligna*		
	N [a]	Range	Average	N	Range	Average	N	Range	Average	N	Range	Average	N	Range	Ave
L. georgica	946	0.02–0.43	0.15												
L. serriola	1804	0.54–0.77	0.66	820	0.00-0.50	0.36									
L. aculeata	7172	0.54–0.76	0.65	6683	0.40–0.62	0.50	13203	0.00-0.40	0.20						
L. altaica	528	0.59–0.77	0.67	492	0.04–0.51	0.38	1956	0.43–0.60	0.52	66	0.02–0.43	0.34			
L. saligna	7436	0.31–0.54	0.42	6929	0.48–0.74	0.60	27547	0.47–0.72	0.60	2028	0.54–0.73	0.64	14196	0.01–0.33	0.15
L. sativa	44	0.63–0.72	0.66	41	0.46–0.60	0.53	163	0.37–0.51	0.43	12	0.48–0.58	0.52	169	0.56–0.70	0.63

[a] pairwise sample size

The mean genetic distance between all interspecies pairs was 0.561, ranging from 0.38 [comparisons between samples belonging to *L. serriola* (most likely the progenitor of *L. sativa*; (Lebeda et al. 2007; Zohary 1991) and samples belonging to *L. altaica*) to 0.67 (comparisons between samples belonging to *L. georgica* and samples belonging to *L. altaica*) (Table 5). The mean genetic distance was highest between interspecies pairs comprised of specimens belonging to *L. georgica* and samples belonging to the four LGP1 species, averaging 0.66 (range: 0.65–0.67). The mean genetic distance between the interspecies pairs comprised of samples belonging to *L. saligna* (LGP2) and samples belonging to the four LGP1 species averaged 0.618, ranging from 0.60 to 0.64. The mean distance for the eight interspecies pairs between LGP1 species averaged 0.48, ranging from 0.38 to 0.53. Notably, an average distance of 0.42 was measured between samples belonging to *L. georgica* and those belonging to *L. saligna*.

To represent the relationship across and within species, a neighbor-network cluster analysis was performed; results are presented in a dendrogram based on the uncorrected p-distance values. Four main clusters were evident (Fig. 1). The first, second and third clusters included all samples belonging to *L. aculeata*, *L. georgica*, and *L. saligna*, respectively. The fourth cluster included all samples belonging to *L. serriola* and *L. altaica*, as well as a subcluster of *L. serriola* samples. The single sample of *L. sativa* was located between the *L. aculeata* and the *L. serriola*/*L. altaica*, clusters. In conclusion, as expected, the entire set of 430 samples clustered in accordance with their species identification.

Delta K, an ad hoc statistics tool implemented in STRUCTURE Harvester (Earl and vonHoldt 2012) version 0.6.94, which is a recommended indicator of the best-fitting number of populations within a sample, was highest at K = 2, that hardly interpreted our dataset structure. In contrast, accessions classification was set at K = 5 (Fig. 2) as the best partition into clusters ("populations") and showed high resemblance to the Neighbor-Network cluster analysis output (see Fig. 1). Color discrimination between clusters in the STRUCTURE illustrated high average (from the ten independent runs) membership probability values across accessions [*L. georgica* (0.989; coloured brown), *L. serriola* (0.974; coloured yellow), *L. altaica* (0.992; coloured yellow), and *L. saligna* (0.924; coloured green], implying that an individual accession, representing one out of the four species mentioned, originated mostly from a single population. Average membership probabilities across the *L. aculeata* accessions showed that individuals from this species originated mostly from two populations (colored gray (0.673) and black (0.326)). Membership probabilities of the single *L. sativa* cv. (Salinas M) showed combination origination mostly from three populations (colored yellow (0.472), black (0.315), and gray (0.188)), i.e., a population representing the *L.serriola*/*L. altaica* samples, and the two population representing the *L. aculeata* samples.

Discussion

The KASP marker fragments in the present study were assigned an allele “A” or allele “B” score, which was then used across analyses as bi-allelic data. Relatively many U-scores were observed, i.e., the specific sequence was absent, and was treated as missing data. Often U-scores were specific for a certain species. Thus, one of our main conclusions is that KASPs assays are not the best marker system for analysis of wild *Lactuca* spp. representing different LGPs, when markers were chosen from samples representing species belonging to LGP1. Clearly, this conclusion can serve as a general conclusion for studies evaluating the genetic diversity of species from other taxa. Notwithstanding, the clear separation observed between *L. georgica* samples and those representing the other three wild LGP1 members of *Lactuca* L., subsection *Lactuca* L. (*L serriola*, *L. altaica*, and *L. aculeata*), as well as samples representing the domesticated lettuce, assessed in the present study by KASP markers and previously by TRAP markers (Beharav et al. 2018a), question the assignment of *L. georgica* to LGP1 (Zohary 1991; Gabrielian and Zohary 2004; Lebeda et al. 2007). A similar conclusion was drawn following a recent phytochemical study (van Treuren et al. 2018) and whole-genome resequencing (Wei et al. 2021). Our collections and novel results, obtained in accordance with our rule of screening for germplasm variation using a large sample of natural populations and individuals, are of substantial importance for crop breeding. The highest percentage of U-score data points observed for the *L. georgica* samples, the highest average *Fst* values of all differentiating KASP loci which were obtained between *L. georgica* in comparisons with the other four WLRs samples (*L serriola*, *L. altaica*, *L. aculeata*, and *L. saligna*), and the highest mean value of genetic distance values obtained for the interspecies pairs of samples belonging to *L. georgica* versus samples belonging to the four LGP1 species (*L. serriola*, *L. altaica*, *L. aculeata*, and *L. sativa*), strengthen our conclusion. Taken together, the present study demonstrated that a combination of multiple analyses of a single, but complex dataset, can provide a clear understanding of obtained results. Taken together with initial hybridization experiments crosses between *L. georgica* and *L. sativa* that displayed only, but partial levels of interfertility (Beharav, personal communication), we conclude that *L. georgica* is indeed a constituent of the LGP2, aligning with McGuire et al. (1993) concept of categorization of the *Lactuca* species into the various GPs and in disagreement with Wei’s et al. (2021) recent statement, that *L. georgica* should be assigned to LGP3.

Various approaches for the placement of *L. altaica* germplasm between and within the LGP1 species have been described in Beharav et el. (2020, and literature cited therein). Results obtained by various analyses of the dataset of the present study support the suggestion (e.g., Koopman et al. 1998) that *L. altaica* be considered conspecific with *L. serriola*, which later brought the species to be described as a synonym of *L. serriola* in WFO (2022). Notably, results obtained by Beharav et el. (2020) strengthened the claim that *L. altaica* shares more similarity with *L. sativa* (e.g., Shulha and Zidorn 2019), as determined by sesquiterpene lactone pattern.

Recent studies have suggested that apart from their distinct floral habit, *L. georgica* is unique in its biochemical features (Michalska et al., 2014; Beharav et al., 2015) and late bolting and flowering (Beharav and Hellier, 2020) and presents a new wild source of resistance to *Bremia lactucae* (Beharav 2021), the causal of lettuce downy mildew, which is the most harmful disease of lettuce worldwide. Thus, *L. georgica* bears potential as an attractive germplasm resource for domesticated lettuce breeding. These findings justify identification and collection of additional *L. georgica* samples from multiple geographic locations, which are restricted to the Euxinian-Hyrcanian region of southwest Asia (Caucasia, Northeast Anatolia, and North Iran) (Zohary 1991, and literature cited therein). However, many of the accessions can no longer be obtained from their original source, i,e., they may be lost in their country of origin, discarded, or entirely inaccessible (McGuire et al 1993). Natural blooming of *L. georgica* plants, a diploid ($2n = 2x = 18$) species (Zohary 1991; Gabrielian and Zohary 2004), occurs in July-August (Gabrielian and Fragman-Sapir 2008). Following our experience, the optimal dates to collect ripe seeds of *L. georgica* in natural habitats is between mid-August and mid-September. Collection and proper handling of seeds from such plants will maintain these important accessions for the benefit of the world economy and agriculture.

Declarations

Acknowledgements

AB wishes to thank Rijk Zwaan B.V. for providing greenhouse facilities, support in assessment and cultivation of the sampled *Lactuca* accessions, as well for providing the KASP data points. Performing the Neighbor-Network and Structure analysis by Dr. Lucie Vaculná and Dr. Miloslav Kitner, Department of Botany, Faculty of Science, Palacký University in Olomouc, Olomouc-Holice, Czech Republic, are gratefully acknowledged. Financial support by the Research Authority, and calculating the uncorrected p-distance values by Mrs. Efrat Yaskil, the Statistics Consulting Unit, University of Haifa, Israel, are acknowledged. I would also like to thank Dr. Roi Ben-David, Institute of Plant Sciences, Agriculture Research Organization (ARO), Volcani Center, Bet Dagan, Israel, for his valuable comments on the first draft version of this manuscript.

Data availability

the dataset generated and analyzed during the current study is available from the author.

Conflict of interest

the author declare that he has no conflict of interest.

References

1. Aiken HH (1955) Tables of cumulative binomial probability distribution. Harvard University Press, Cambridge, MA
2. Beharav A (2021) *Lactuca georgica*, a new wild source of resistance to downy mildew: comparative study to other wild lettuce relatives. Eur J Plant Pathol 160:127–136

3. Beharav A, Hellier B (2020) Bolting and flowering response of *Lactuca georgica*, a wild lettuce relative, to low temperatures. Amer J Plant Sci 11:2139–2154
4. Beharav A, Ben-David R, Doležalová I, Lebeda A (2008) Eco-geographical distribution of *Lactuca saligna* natural populations in Israel. Isr J Plant Sci 56:195–120
5. Beharav A, Ben-David R, Doležalová I, Lebeda A (2010) Eco-geographical distribution of *Lactuca aculeata* natural populations in north-eastern Israel. Genet Resour Crop Evol 57:679–686
6. Beharav A, Hellier B, Richardson KL, Lebeda A, Kisha T (2018a) Genetic relationships and structured diversity of *Lactuca georgica* germplasm from Armenia and the Russian Federation among other members of *Lactuca* L., subsection *Lactuca* L., assessed by TRAP markers. Genet Resour Crop Evol 65:1963–1978
7. Beharav A, Khalifa S, Nevo E (2018b) Israel J. Plant Sci. New insights into the range, morphology, and natural hybridization of wild *Lactuca aculeata* in Israel. Isr J Plant Sci 65:175–185
8. Beharav A, Malarz J, Michalska K, Ben-David R, Stojakowska A (2020) Variation of sesquiterpene lactone contents in *Lactuca altaica* natural populations from Armenia. Biochem Syst Ecol 90:104030
9. Beharav A, Stojakowska A, Ben-David R, Malarz J, Michalska K, Kisiel W (2015) Variation of sesquiterpene lactone contents in *Lactuca georgica* natural populations from Armenia. Genet Resour Crop Evol 62:431–441
10. Bryant D, Moulton V (2004) Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol 21:255–265
11. Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper & Row Publishers, New York
12. Earl DA, VonHoldt BM (2012) Structure Harvestor: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361
13. 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
14. Gabrielian E, Fragman-Sapir O (2008) Flowers of the Transcaucasus and adjacent areas: including Armenia, Eastern Turkey, Southern Georgia, Azerbaijan and Northern Iran. A.R.G. Gantner Verlag, K.G
15. Gabrielian E, Zohary D (2004) Wild relatives of food crops native to Armenia and Nakhichevan. Flora Mediterranea 14:5–80. Edinburgh University Press, Palermo
16. Hartl DL, Clark AG (1997) Principles of population genetics, 3rd edn. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts
17. He C, Holme J, Anthony J (2014) SNP Genotyping: The KASP Assay. In: Fleury D, Whitford R (eds) Crop Breeding. Methods in Molecular Biology (Methods and Protocols), vol 1145. Humana Press, New York, NY, pp 75–86
18. Huson DH, Bryant D (2006) Application of Phylogenetic Networks in Evolutionary Studies. Mol Biol Evol 23(2):254–267
19. Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806
20. Jemelková M, Kitner M, Křístková E, Beharav A, Lebeda A (2015) Biodiversity of *Lactuca aculeata* germplasm assessed by SSR and AFLP markers, and resistance variation to *Bremia lactucae*. Bioc Syst Ecol 61:344–356
21. Jemelková M, Kitner M, Křístková E, Doležalová I, Lebeda A (2018) Genetic variability and distance between *Lactuca serriola* L. populations from Sweden and Slovenia assessed by SSR and AFLP markers. Acta Bot Croat 77:172–180
22. Kitner M, Lebeda A, Doležalová I, Maras M, Křístková E, Nevo E, Pavlíček T, Meglič V, Beharav A (2008) AFLP analysis of *Lactuca saligna* germplasm collections from four European and three Middle East countries. Isr J Plant Sci 56:185–193
23. Kitner M, Majesky L, Křistková E, Jemelková M, Lebeda A, Beharav A (2015) Genetic structure and diversity in natural populations of three predominantly self-pollinating wild *Lactuca* species in Israel. Genet Resour Crop Evol 62:991–1008
24. Koopman WJM, Guetta E, Van De Wiel CCM, Vosman B, Van den Berg RG (1998) Phylogenetic relationships among Lactuca (Asteraceae) species and related genera based on ITS-1 DNA sequences. Am J Bot 85:1517–1530
25. Lebeda A, Kitner M, Křístková E, Doležalová I, Beharav A (2012) Genetic polymorphism in *Lactuca aculeata* populations and occurrence of natural putative hybrids between *L. aculeata* and *L. serriola*. Bioc Syst Ecol 42:113–123
26. Lebeda A, Ryder EJ, Grube R, Doležalová I, Křístková E (2007) Lettuce (Asteraceae *Lactuca* spp.). In: Singh RJ (ed) Genetic Resources, Chromosome Engineering, and Crop Improvement. Vegetable Crops, vol 3. CRC Press, Taylor and Francis Group, Boca Raton, FL, USA, pp 377–472
27. Livak KJ, Flood SJ, Marmaro J, Giusti W, Deetz K (1995) Oligonucleotide with fl uorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. Genome Res 4:357–362
28. Michalska K, Beharav A, Kisiel W (2014) Sesquiterpene lactones from roots of *Lactuca georgica*. Phytochem Lett 10:10–12
29. McGuire PE, Ryder EJ, Michelmore RW, Clark RL, Antle R, Emery G, Hannan RM, Kesseli RV, Kurtz EA, Ochoa O, Rubatzky VE, Waycott W (1993) Genetic resources of lettuce and *Lactuca* species in California. An assessment of the USDA and UC collections and recommendations for long-term security. Report no. 12. University of California, Genet Resour Conservation Program, Davis, CA
30. Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323
31. Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590
32. Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol 6:288–295
33. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

34. Pritchard JK, Wen X, Falush D (2009) Documentation for structure software: Version 2.3. University of Chicago, Chicago, USA. http://pritch.bsd.uchicago.edu/structure. html
35. Rosenberg A (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138
36. Semagn K, Babu R, Hearne S, Olsen M (2014) Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Mol Breed 33:1–14
37. Shikari AB, Najeeb S, Khan G et al (2021) KASP™ based markers reveal a population sub-structure in temperate rice (*Oryza sativa* L.) germplasm and local landraces grown in the Kashmir valley, north-western Himalayas. Genet Resour Crop Evol 68:821–834
38. Shulha O, Zidorn C (2019) Sesquiterpene lactones and their precursors as chemosystematic markers in the tribe Cichorieae of the Asteraceae revisited: an update (2008–2017). Phytochemistry 163:149–177
39. Stojakowska A, Michalska K, Malarz J, Beharav A, Kisiel W (2013) Root tubers of *Lactuca tuberosa* as a source of antioxidant phenolic compounds and new furofuran lignans. Food Chem 138:1250–1255
40. Šuštar-Vozlič J, Ugrinović K, Maras M, Křístková E, Lebeda A, Meglič V (2021) Morphological and genetic diversity of Slovene lettuce landrace 'Ljubljanska ledenka' (*Lactuca sativa* L.). Genet Resour Crop Evol 68:185–203
41. WFO (2022) : World Flora Online. http://worldfloraonline.org/. Accessed on: 14 March 2022
42. Yeh FC, Yang RC, Boyle T (1999) POPGENE version 1.32: Microsoft Windows–based freeware for population genetic analysis, quick user guide. Center for International Forestry Research, University of Alberta, Edmonton, Alberta, Canada
43. van Treuren R, van der Arend AJM, Schut JW (2013) Distribution of downy mildew (*Bremia Lactucae* Regel) resistances in a genebank collection of lettuce and its wild relatives. Plant Genet Resour: characterization and utilization 11:15–25
44. van Treuren R, van Eekelen HDLM, Wehrens R, Wehrens R, de Vos RCH (2018) Metabolite variation in the lettuce gene pool: towards healthier crop varieties and food. Metabolomics 14:146
45. Wei T, van Treuren R, Liu X et al (2021) Whole-genome resequencing of 445 *Lactuca* accessions reveals the domestication history of cultivated lettuce. Nat Genet 53:752–760
46. Weir BS (1996) Genetic Data Analysis II. Methods for discrete population genetic ata. Sinauer Associates, Inc. Publishers, Sunderland, Mass., U.S.A, p 356
47. Zohary D (1991) The wild genetic resources of cultivated letucce (*Lactuca sativa* L. Euphytica 53:31–35

Figures

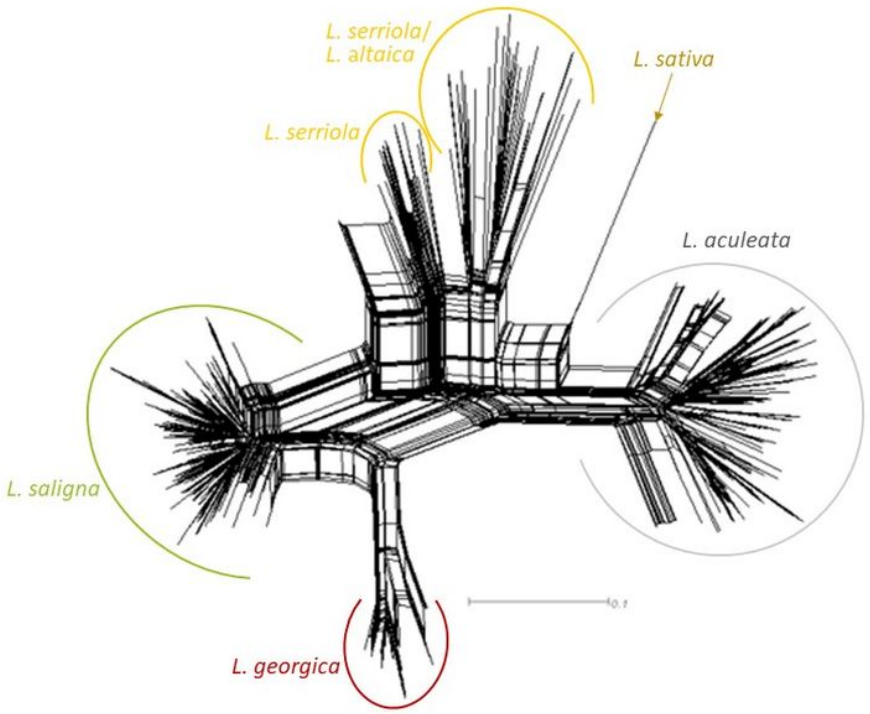

Figure 1

Neighbor-network cluster analysis based on uncorrected p-distance among 430 samples of six *Lactuca* spp., based on 115 KASP bi-allelic SNP genotyping assays. Resulting clusters are highlighted by colors that corresponds to the results of the Bayesian clustering presented in Fig. 2.

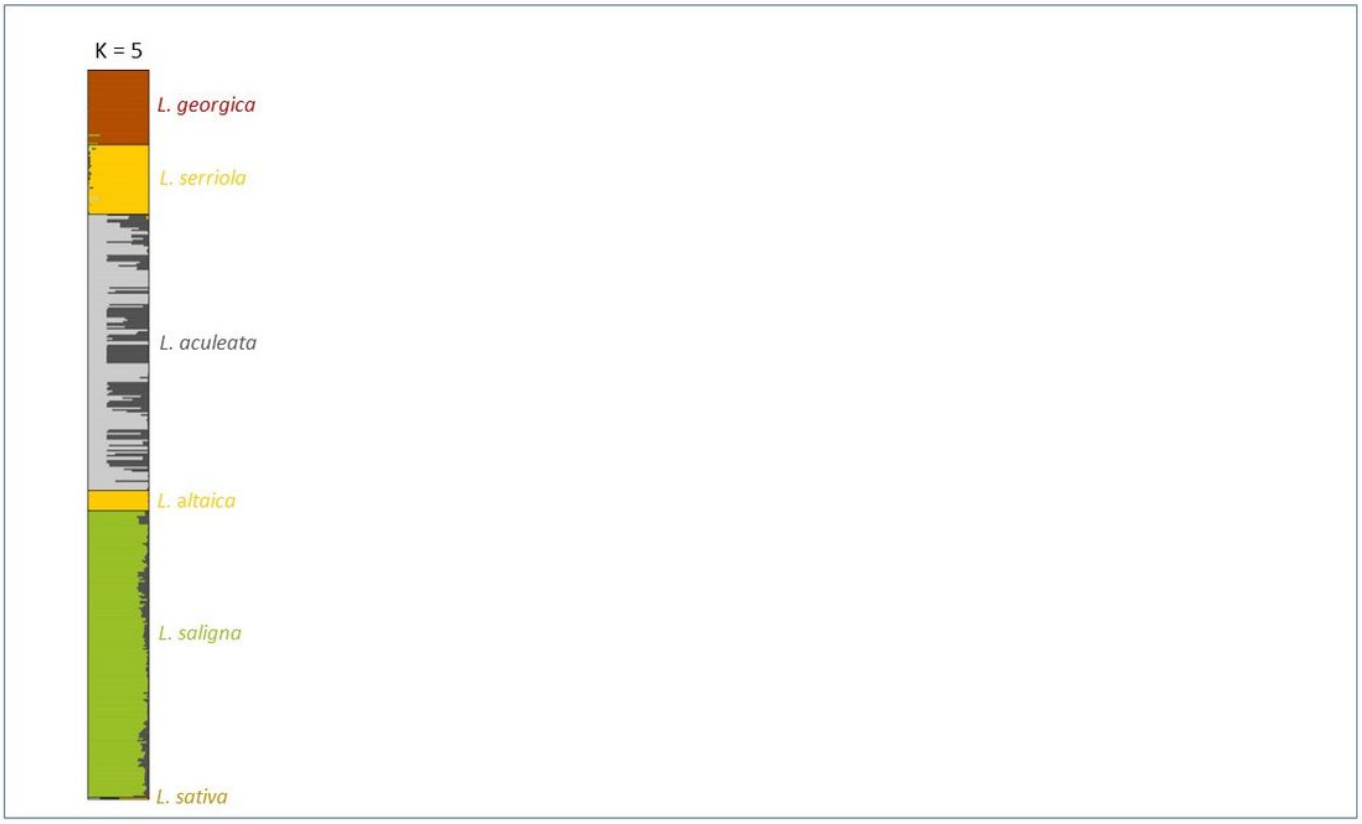

Figure 2

Species structure and membership fraction at K = 5 for 430 sampled accessions of six *Lactuca* spp., based on 'admixture model' of STRUCTURE analysis of the results of 115 bi-allelic KASP SNP genotyping assays. Each accession is represented by a horizontal line with the different colors representing distinct species.