Genetic diversity and evolution in *Lactuca* L. (Asteraceae)

from phylogeny to molecular breeding

Zhen Wei

Thesis committee

Promotor

Prof. Dr M.E. Schranz

Professor of Biosystematics

Wageningen University

Other members

Prof. Dr P.C. Struik, Wageningen University

Dr N. Kilian, Free University of Berlin, Germany

Dr R. van Treuren, Wageningen University

Dr M.J.W. Jeuken, Wageningen University

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences.

Genetic diversity and evolution in *Lactuca* L. (Asteraceae)

from phylogeny to molecular breeding

Zhen Wei

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Monday 25 January 2016
at 1.30 p.m. in the Aula.

Zhen Wei

Genetic diversity and evolution in *Lactuca* L. (Asteraceae)

- from phylogeny to molecular breeding,

210 pages.

PhD thesis, Wageningen University, Wageningen, NL (2016)

With references, with summary in Dutch and English

ISBN 978-94-6257-614-8

Contents

*Manuscript published

Chapter 1

General introduction

Zhen Wei

Biosystematics Group, Wageningen University and Research Centre, 6708 PB Wageningen, The Netherlands

Chapter 1

Lettuce (*Lactuca sativa* L.) is an economically important vegetable crop. The total value of world lettuce sales in 2007 was US $1,934,983,000. The world production of lettuce and the related chicory (*Cichorium intybus*) has been increasing yearly and the total yield reached 23,733,803 metric tons in 2009 (U.S.D.A. 2011). The edible parts of domesticated lettuce include the leaves and/or stems and are usually consumed as fresh or cooked products. Due to its economic importance, most current studies of the genus *Lactuca* have focused on lettuce cultivars and the species that can be easily crossed to domesticated lettuce (Koopman et al. 1998; Zohary 1991), including studies of phylogeny and trait breeding. However, the broader phylogenetic relationships of domesticated lettuce and wild lettuce species remain unclear, including the taxonomic boundary of the genus *Lactuca* L. itself. In lettuce breeding, quantitative trait loci (QTLs) related to biotic (Christopoulou et al. 2015; Jeuken et al. 2008; Simko et al. 2013; Simko et al. 2009) and abiotic stresses have been identified (Hartman et al. 2014; Jenni et al. 2013a; Uwimana et al. 2012a), providing the possibilities to improve the tolerance of lettuce to different stresses.

In this thesis, I will provide the most extensive phylogenetic reconstruction of *Lactuca* domesticated and wild species, based on chloroplast genes, genome and nuclear DNA (Internal Transcribe Spacer, ITS) sequences. A QTL analysis of the responses to salinity in a recombinant inbred line population, derived from a cross between cultivated lettuce (*L. sativa* 'Salinas') and wild lettuce (*L. serriola*), will also be presented and the potential candidate gene associated with salinity stress will be tested. Therefore, in this introduction, I will provide an overview of lettuce cultivars and uses, its hypothesized domestication history, the taxonomic position of the genus *Lactuca*, the current status of lettuce molecular breeding and mechanisms of salinity tolerance in plants, especially the High-affinity K^+ Transporter (*HKT*) gene family.

Overview of domesticated lettuce

According to different leaf shape and size, degree of rosette and head formation and less so on colour, stem type and other traits, lettuce cultivars have been classified into seven types: Butterhead, Crisphead, Cos, Cutting, Stalk, Latin, and Oilseed (De Vries and Van Raamsdonk 1994; Křístková et al. 2008; Ryder 1999; Vries 1997). Butterhead lettuce is a head type lettuce with soft and tender broad leaves, originating from Europe. Crisphead lettuce with a large firm head has two subtypes: the iceberg subtype, larger in weight and volume with a dense head, the Batavia subtype, smaller with a less dense head (Ryder 1999; Vries 1997). Cos lettuce has an erect, elongated or loaf-shaped head, with a predominant midrib running almost to the apex (Lindqvist 1960b). The leaf colour ranges from yellowish to dark green (Lebeda et al. 2007; Ryder 1999). Cutting lettuce normally doesn't form a head or have an enclosure stage. The leaves vary in leaf margin (entire, frilled), shapes (broad, elongated, lobed, curled), sizes, texture (crisp, soft), and colours (red, green; dark, light) (Křístková et al. 2008; Lebeda et al. 2007; Ryder 1999). The stalk lettuce has a thick elongated stalk and narrow leaves. The stalk is tender and consumed raw in Egypt and cooked in China (Lindqvist

1960b). There are two stalk lettuce types: one type is represented by the Chinese cultivars with light grey leaves and the leaves can be as broad as Cos type. The other type has long lanceolate leaves with pointed apex (Lindqvist 1960b). Latin, or grass é, lettuce forms a loose head and has thick leathery leaves of dark green colour. It is of European origin, but also grown around the Mediterranean including North Africa, in South America and on small areas in the U.S. (Křístková et al. 2008; Rodenburg 1960). Oilseed lettuce is a primitive type of *L. sativa* with larger seeds than those of other lettuces. The seeds are crushed to produce oil for cooking. This group has not yet been fully domesticated (Boukema 1990; Křístková et al. 2008; Lebeda et al. 2007; Ryder 1999; Ryder 1986).

Different regions and countries produce different types of lettuce cultivars. Butterhead and Crisphead lettuces are overwhelmingly popular in the United States, the United Kingdom, France, the Netherlands, Belgium, Germany and other European countries (Ryder 1999; Vries 1997). Cos lettuce came from the Greek island Cos (Kos) (Helm 1954) and remains popular in the Mediterranean Basin, northern Africa, southwest Asia, and southern Europe (Lebeda et al. 2007; Ryder 1999). The Greeks and Romans cultivated cutting lettuce (Křístková et al. 2008). More recently, Cos and Cutting lettuces have been increasingly used in the U.S. and in other countries (Lebeda et al. 2007; Vries 1997). Stalk lettuce is found in Egypt and Middle Eastern countries, and is also very common in China and India (Ryder 1999; Vries 1997). The ancient practice of making oil from Oilseed lettuce has continued to the present time in Egypt (Ryder 1999).

In ancient Egypt, lettuce was considered as an aphrodisiac and played an important role in the yearly festival of Min, God of fertility and procreation. A long-leafed lettuce type was depicted on walls of Egyptian tombs (Harlan 1986). Lindqvist (1960b) referred some primitive forms of *L. sativa* in Egypt and considered them as in a semi-wild state, rather than cultivated. Whitaker (1969) concluded that the ancestors of cultivated lettuce is indigenous to the eastern Mediterranean Basin, probably Egypt. Zeven & De Wet (1982) mentioned part of the European-Siberian region (the Middle East) as the primary centre of origin of *L. sativa*. Rulkens (1987) presumed cultivated lettuce originated in the Kurdistan-Mesopotamia area instead of in Egypt. Boukema et al. (1990) indicated that the domestication of lettuce happened in South-West Asia in the region between Egypt and Iran. De Vries (1997) deemed that the cultivated lettuces originated in South-West Asia, from the area around the Euphrates and Tigris rivers. In his point of view, there were two main reasons: 1. the highest number of related wild species can be found between the Euphrates and Tigris rivers, whereas only one related wild species - *L. serriola* is found in the Nile Valley. 2. Cereal-growing cultures in Kurdistan-Mesopotamia were known long before the first known Egyptian-grown cereals (Rulkens 1987), indicating a more ancient origin of agriculture.

Taxonomic position of *Lactuca*

The Mediterranean basin, South-Western Asia and Africa comprise centres of diversity of wild *Lactuca* species and can be considered as hot-spots for lettuce conservation (Lebeda et al. 2008; Lebeda et al. 2004; Lebeda et al. 2009). *Lactuca* species are distributed in temperate and warm regions of the northern hemisphere (Europe, Asia, Indonesia, North and Central America, Africa) (Feráková and Májovský 1977; Lebeda et al. 2004). Most of them are xerophytes except for some scandent, liana-like endemic species in the central African mountains (Stebbins 1937).

Taxonomic and phylogenetic analyses place the genus *Lactuca* in the subtribe Lactucinae, tribe Lactuceae (Cichorieae), subfamily Cichorioideae of the family Asteraceae (Compositae) (Judd et al. 2007; Kadereit and Jeffrey 2007). This genus was established by Linné in 1753 (Linné 1753). However, since it was first proposed, the circumscription and delimitation of *Lactuca* has remained obscure, mainly due to the complex and variable morphological characters of species in the genus. No agreement has been come to on the number and the boundary of *Lactuca* species. Stebbins (1937) described 8 scandent *Lactuca* species in Africa. Jeffrey (1966) modified Stebbins' treatment of the scandent species and elaborated a total of 33 African *Lactuca* species. Lebeda et al. (2004) reviewed all the published literature and stated that there are about 100 wild *Lactuca* species with two centres of diversity, Asia (51 species) and Africa (43 species). In China, Shih (1987, 1988a, b, 1991, 1997) established several new genera (*Notoseris* Shih, *Chaetoseris* Shih, *Stenoseris* Shih, *Pterocypsela* Shih) and revised the genera of *Paraprenanthes* Chang, *Mulgedium* Cass. and *Lagedium* Soják, by segregating species from *Lactuca*. Shih and Kilian (2011) considered there to be about 50 to 70 *Lactuca* species in total. Wang et al. (2013) transferred some species in *Pterocypsela*, *Lagedium*, *Mulgedium*, and *Steptorhamphus* Bunge back to *Lactuca*.

However, all the authors mentioned above dealt mostly with only regional (Asian, European, African) *Lactuca* species and the genus has never been revised in its entirety. So far, phylogenetic relationships within the genus *Lactuca* are primarily domesticated lettuce gene pool centred, due to the economic importance of lettuce. According to Harlan and Wet (1971), the primary gene pool (GP-1) of cultivated plants contains biological species (including spontaneous and cultivated races), which can cross with each other easily. Hybrids are generally fertile with good chromosome pairing, normal gene segregation, and easy gene transfer (Harlan and Wet 1971). The secondary gene pool (GP-2) includes biological species that will cross with crops. Gene transfer is possible but with barriers. Hybrids tend to be sterile with poor chromosomes pairing or not at all and may be too weak to mature. The tertiary gene pool (GP-3) comprises of species that have difficulty to cross with cultivated plants. Hybrids tend to be anomalous, lethal or completely sterile. Gene transfer is either not possible or extreme or radical methods need to be used to (Harlan and Wet 1971).

Daniel Zohary (1991) established the first lettuce gene pool concept, including *L. sativa*, *L. serriola* L., *L. aculeata* Boiss. & Ky., *L. scarioloides* Boiss., *L. azerbaijanica* Rech., *L. georgica* Grossh., *L. dregeana* DC. and *L. altaica* Fisch. & C.A. Mey as lettuce GP-1 and *L.*

saligna L. as GP-2. Koopman et al. (1998) inferred the phylogenetic relationships among *Lactuca* and related genera based on ITS-1 DNA sequences and AFLP (amplified fragment length polymorphism) fingerprints, indicating *L. sativa*, *L. serriola*, *L. aculeata*, *L. dregeana* DC., *L. altaica* Fischer & C.A.Meyer, as GP1, *L. virosa* L. and *L. saligna* as GP-2 and *L. quercina* L., *L. viminea* Presl & C.Presl, *L. sibirica* Benth. ex Maxim., and *L. tatarica* (L.) C.A. Meyer as GP-3. Later on, *L. sativa*, *L. serriola*, *L. dregeana*, and *L. altaica* were postulated conspecific (Koopman et al. 2001).

The genus *Lactuca* can be classified into seven sections: *Lactuca* subsection *Lactuca* and *Cyanicae* DC., *Phaenixopus* (Cass.) Bentham, *Mulgedium* (Cass.) C.B. Clarke, *Lactucopsis* (Schultz Bip. Ex Vis. Et Pančić) Rouy, *Tuberosae* Boiss., *Micranthae* Boiss., *Sororiae* Franchet, and two (African and North American) geographic groups (Lebeda et al. 2004; Lebeda et al. 2009; Lebeda et al. 2007). All the known *Lactuca* species in the subsection *Lactuca* are belonging to the GP-1.

The closest species to the lettuce cultivar is *L. serriola*, which can be easily crossed with *L. sativa*. The hybrids of these two species are self-fertile (Hartman et al. 2012a; Thompson et al. 1941). Although most studies have shown that *L. sativa* has a polyphyletic origin, *L. serriola* is the only widely known and accepted progenitor of crop lettuce (de 1996; Kesseli et al. 1991; Kesseli and Michelmore 1986; Lindqvist 1960b; M. Hill et al. 1996; Whitaker 1969; Yang et al. 2007; Zohary 1983). Additionally, *L. virosa* has been introgressed into some Crisphead lettuce cultivars ('Salinas', 'Vanguard', 'Vanguard 75', 'Vanmax') for its robust root system and decreased leaf drop (Mikel 2007). Somatic chromosome studies have shown that *L. serriola* and *L. sativa* have almost identical chromosome morphology. *L. saligna* differs slightly from them, but *L. virosa* is quite distinct from the other three species (Koopman et al. 1993; Lindqvist 1960a; Matoba et al. 2007; Mej ás 1993). *L. saliga* can cross with *L. serriola* and *L. sativa* and the F1 hybrids were shown to be partially fertile or self-fertile (Jeuken and Lindhout 2002; Jeuken et al. 2001; Thompson et al. 1941; Zohary 1991). Though the cross between *L. virosa* and *L. sativa* often fails, it is still possible to obtain the self-sterile hybrid and a very low percentage of the pollens are viable (Thompson et al. 1941; Whitaker and Thompson 1941; Zohary 1991). *L. viminea* is partly fertile with *L. virosa* (Groenwold 1983). *L. tatarica* could be somatically hybridized with *L. sativa* (Chupeau et al. 1994; Maisonneuve et al. 1995).Another domesticated species in *Lactuca* is *L. indica* L. (Indian lettuce), which is native to China and has been cultivated for its succulent leaves (Kadereit and Jeffrey 2007; Yamaguchi 1983). *L. indica* can somatically hybridize with *L. sativa*, generating viable callus (Mizutani et al. 1989).

Current status of lettuce molecular breeding

Since the development of DNA markers in the 1980s, molecular breeding has become a common practice (Rafalski and Tingey 1993). A number of DNA markers, including AFLP, RAPD (random amplified polymorphic), RFLP (restriction fragment length polymorphism),

SSR (simple sequence repeats or microsatellites), SCAR (sequence characterized amplified region), SNP (single nucleotide polymorphism) and so on, have been developed to construct genetic maps for crop improvement (Collard et al. 2005). Quantitative trait loci (QTL) regions within genomes, containing genes associated with a particular quantitative trait (e.g. plant height), can be identified using DNA markers and genetic maps (Collard et al. 2005). Once the QTLs related to agronomical important traits and their tightly linked DNA markers have been validated, the DNA markers can be used as molecular tools for marker-assisted selection (MAS) in plant breeding (Collard and Mackill 2008; Ribaut and Hoisington 1998).

In lettuce, AFLP, RAPD, RFLP, SSR (microsatellites) markers (Dziechciarková et al. 2004; Kesseli et al. 1991; Koopman 2005; Koopman et al. 2001; M. Hill et al. 1996; van de Wiel et al. 1998; van de Wiel et al. 1999; Witsenboer et al. 1997; Yang et al. 2007) and biochemical markers (isozymes) (Cole et al. 1991; Doležalová et al. 2003; Tanaka 2003) have been used to study the relationships between lettuce cultivars and wild lettuces. Several lettuce genetic maps based on RFLP, RAPD, AFLP, SSR, and EST markers have been developed as well. Landry et al. (1987) constructed a linkage map of lettuce using 41 RFLP loci, 5 downy mildew resistance genes, 4 isozyme loci and 3 morphological markers. Kesseli (1994) developed a genetic linkage map of *L. sativa* from RFLP and RAPD Markers. Jeuken et al. (2001) constructed an integrated interspecific AFLP map of lettuce, derived from two *L. sativa* x *L. saligna* F_2 populations. Truco et al. (2007) integrated seven linkage maps of lettuce into a high-density one comprising of 2,744 DNA markers. Later, an ultra-high-density genetic map of lettuce using 213 $F_{7:8}$ recombinant inbred lines (RILs) derived from a cross between *L. sativa* 'Salinas' and *L. serriola* was generated, including 12,842 unigenes (13,943 markers) (Truco et al. 2013).

The molecular breeding of lettuce mainly concerns three issues: beneficial characters for crops (root architecture, seed germination, leaf size, leaf weight, shelf life, storage, fitness etc.), biotic and abiotic stresses. Genomic regions associated with root architecture and deep soil water exploitation were determined in a $F_{2:3}$ population of *L. sativa* 'Salinas' x *L. serriola* using AFLP markers (Johnson et al. 2000). QTLs associated with leaf area, leaf thickness measured from specific leaf area, leaf dry and fresh weight, epidermal cell area, epidermal cell number etc., have also been detected in a RIL population derived from a cross between *L. sativa* 'Salinas' x *L. serriola* (Zhang et al. 2007). The identification of QTLs related to seed longevity under controlled deterioration and conventional storage conditions were performed using F_8 RILs from a cross between *L. sativa* 'Salinas' x *L. serriola* (Schwember and Bradford 2010b). QTLs related to domestication traits (germination time, rosette leaf length, plant height, number of stem leaves etc.) in Crisphead lettuce were revealed by a RIL population from a cross between *L. sativa* 'Salinas' and *L. serriola* f. *serriola* (UC96US23), grown in greenhouse condition (Hartman et al. 2012a). The same RIL population was also used to perform QTL analyses on fitness related traits (germination rate, biomass, days to first flower, seed output etc.) in field environments to evaluate the impact of domestication genes

(Hartman et al. 2012b). QTL analyses of fitness related traits were measured in two different crop-wild hybrids of lettuce (back-cross and RIL populations) and the fitness distribution of early- and late-generation hybrids were estimated as well (Hartman et al. 2013). Atkinson et al. (2013) constructed an intra-specific linkage map of a RIL population derived from lettuce cultivar 'Saladin' and 'Iceberg' and identified QTLs for postharvest discolouration traits for MAS.

Compared to crop improvement and abiotic stress, the resistance to biotic stress in lettuce has been more extensively studied. Resistance genes against downy mildew, corky root, lettuce mosaic, lettuce dieback, *Verticillium* wilt, turnip mosaic, root downy mildew, powdery mildew, big-vein, *Fusarium* wilt, and anthracnose have been mapped on the genetic map of lettuce and assays for MAS are developed or under development (Simko 2013). Among all the lettuce diseases, downy mildew is probably the most frequent and devastating one and can infect lettuce at any developmental stage (Lebeda et al. 2013; Simko et al. 2013; van Treuren et al. 2011). The resistance genes to downy mildew are either single dominant genes (*Dm*) or resistance factors (*R*), or multiple genes with minor effects (Beharav et al. 2013; Bonnier et al. 1992; Bonnier et al. 1994; Jeuken and Lindhout 2002; Jeuken et al. 2008; Kesseli et al. 1994; Kuang et al. 2006; Kuang et al. 2008; Kuang et al. 2004; Lebeda and Reinink 1994; Maisonneuve et al. 1994; Meyers et al. 1998; Paran et al. 1991; Paran and Michelmore 1993; Simko et al. 2013; Zhang et al. 2009). Corky root is a disease in lettuce caused by the bacterium *S. suberifaciens*, previously known as *Rhizomonas suberifaciens* (Van Bruggen et al. 1989; Van Bruggen and Jochimsen 1992; Yabuuchi et al. 1999). A recessive allele at a single locus (*cor*) for the resistance to corky root has been identified (Brown and Michelmore 1988) and located using a $F_{2:3}$ population of *L. sativa* based on RFLP and SNP markers (Moreno-Vazquez et al. 2003). A large number of lettuce cultivars, *L. serriola* and *L. virosa* lines have also been screened for resistance to corky root in lettuce, however, none of the resistant lines had the two DNA markers closely linked to the *cor* locus published earlier (Moreno-Vazquez et al. 2003; Mou and Bull 2004). Some lettuce breeding lines and cultivars have been evaluated for resistance to corky root and lettuce mosaic virus (Beiquan Mou 2007). *Lettuce mosaic virus* (LMV) is the major agent of lettuce mosaic disease and can be transmissible by aphid vectors (Candresse et al. 2006; Simko 2013; Soleimani et al. 2011). Recessive (Nicaise 2003; Ryder 2002; Ryder 1970a, b) and dominant resistance genes to LMV have been identified in lettuce (Candresse et al. 2006; Pink et al. 1992a; Pink et al. 1992b; Revers et al. 1997). In addition, resistant genes to lettuce dieback (Grube et al. 2005; Simko et al. 2010; Simko et al. 2009), Verticillium wilt (Hayes et al. 2011), turnip mosaic (Montesclaros et al. 1997; Robbins et al. 1994), root downy mildew (Kesseli et al. 1993; Vandemark et al. 1991), powdery mildew (Simko et al. 2014), big-vein (Hayes and Ryder 2007; Hayes et al. 2008), Fusarium wilt (Aruga et al. 2012), and anthracnose (McHale et al. 2009) have been identified and located on lettuce genomes as well.

The genetic variation studies for abiotic stress focus on tolerance to temperature, drought, salinity and nutrient deficiency. QTLs related to seed germination thermo-tolerance were identified in an F_8 RIL population from *L. sativa* 'Salinas' x *L. serriola* (Argyris et al. 2005). One of the identified QTLs (*Htg6.1*) was found to be co-located with *LsNCED4*, a gene encoding an enzyme in the ABA biosynthetic pathway (Argyris et al. 2008). One major QTL associated with seed priming (controlled hydration followed by drying) effect was detected. The expression of genes encoding regulated enzymes in the gibberellin and ethylene biosynthetic pathways was enhanced by this priming effect and suppressed by imbibition at elevated temperatures (Schwember and Bradford 2010a). Uwimana et al. (2012b) used two backcross populations (BC_1 and BC_2, a F_1 hybrid plant resulting from a cross between *L. serriola* and *L. sativa* 'Dynamite' were hand-pollinated with pollen from the *L. serriola* parental line) to identify QTLs associated with drought, salinity and nutrient deficiency. This was done to mimic possible natural introgression in the wild from lettuce cultivars into its wild relative. The contribution of domesticated lettuce to the vigour of crop-wild hybrids under the same abiotic stress conditions, using 98 $F_{2:3}$ families from a cross between *L. serriola* and *L. sativa* 'Dynamite', was also measured (Uwimana et al. 2012a). Genomic regions containing candidate genes related to heat stress-induced physiological disorders and maturity traits have been identified in Crisphead lettuce (Jenni et al. 2013b). Abiotic QTLs under drought, low nutrients, salt and aboveground competition stresses were identified in greenhouse and field environments. The results implicated that the introgression risk of stress tolerance (trans-) genes under field conditions could not easily be predicted by genomic background selection patterns inferred from controlled conditions in greenhouse (Hartman et al. 2014).

Mechanism of salinity tolerance in plants

Mechanisms of salinity tolerance in plants include three main types: osmotic tolerance enabling efficient water potential maintenance within the plant tissues, Na^+ or Cl^- exclusion preventing damage to photosynthetic tissues, and tissue tolerance to accumulated Na^+ or Cl^- (mainly into vacuoles) (Munns and Tester 2008; Roy et al. 2014). The osmotic tolerance immediately happens after plants contact external salinity and plants reduce cell expansion in root tips and young leaves, leading to stomatal closure (Mano and Takeda 1997; Munns and Tester 2008). Although ROS waves (Jiang et al. 2012; Mittler et al. 2011; Suzuki et al. 2012), Ca^{2+} waves (Roy et al. 2014) or long distance electrical signal (Maischak et al. 2010) have been indicated to involve in this osmotic phase, little knowledge about this phase was known (**Figure 1**).

Two gene families have been considered to play a critical role in Na^+ accumulation or exclusion in plants, including the salt overly sensitive (SOS) pathway (Huertas et al. 2012; Jarvis et al. 2014; Katschnig et al. 2015; Qiu et al. 2002) and the high affinity potassium transporter (*HKT*) gene family (Ali et al. 2012; Davenport et al. 2007; Hauser and Horie 2010; Horie et al. 2009; Platten et al. 2013; Rus et al. 2006; Rus et al. 2004) (**Figure 1**). Of the two

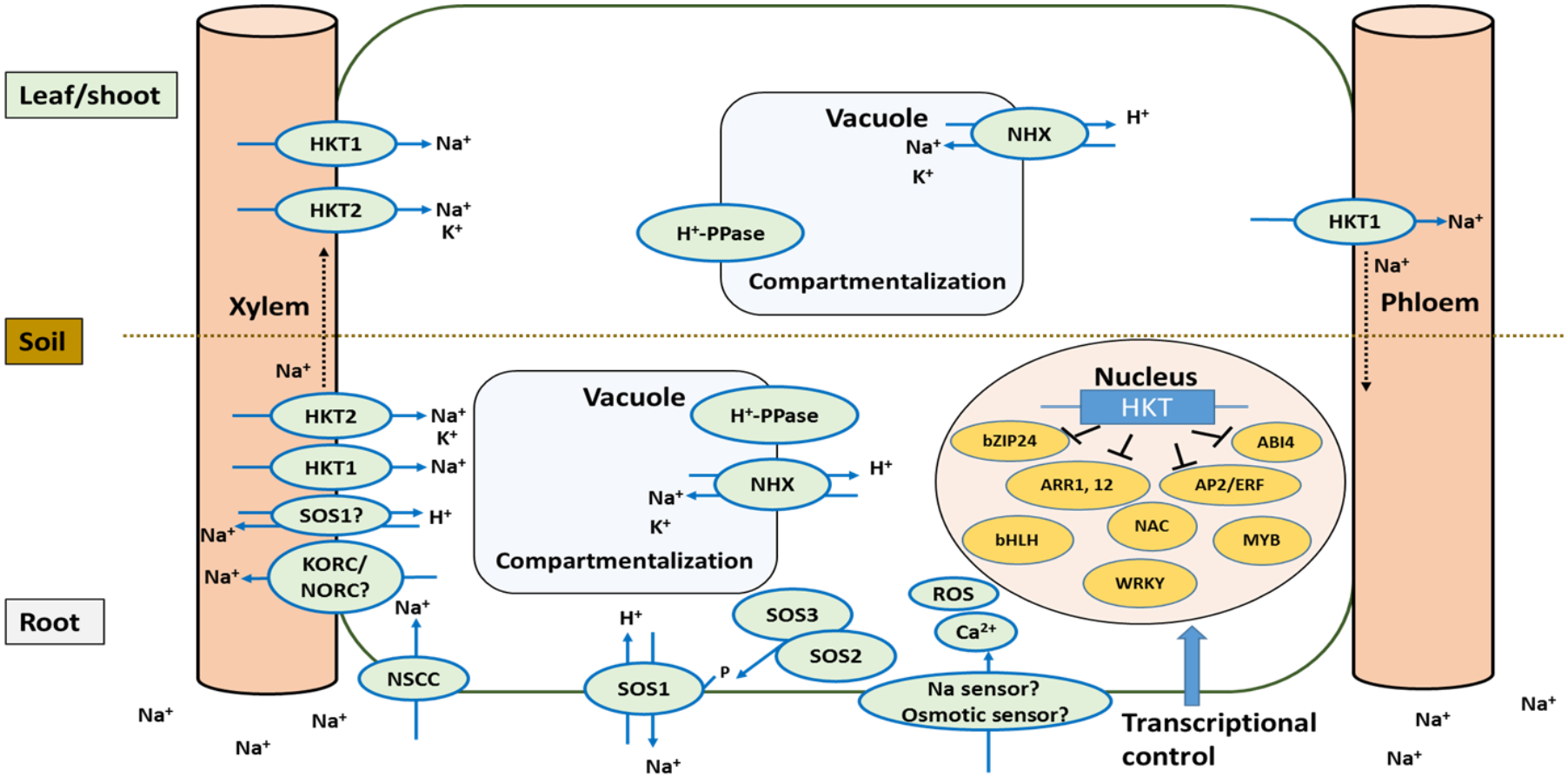


Figure 1 Overview of Na^+ transport mechanisms and important components of responses to external salinity in plants (Barragán et al. 2012; Berthomieu et al. 2003; Cui et al. 2013; de Boer and Wegner 1997; Deinlein et al. 2014; Guo et al. 2011; Hamamoto et al. 2015; Olsen et al. 2005; Roy et al. 2014; Su et al. 2015; Wegner and De Boer 1997). Abbreviations: NSCCs, non-selective cation channels; ROS, reactive oxygen species; SOS, SALT OVERLY SENSITIVE; HKT, High-affinity K^+ Transporter; NHX, Na^+/H^+ exchanger; H^+-PPase, proton-translocating vacuolar inorganic pyrophosphatases; AP2/ERF, APETALA2/ETHYLENE RESPONSE FACTOR; bZIP, basic leucine zipper; ARR, Arabidopsis response regulator; ABI, ABA-INSENSITIVE; bHLH, basic helix-loop-helix.

gene families involving in sodium accumulation, *HKT1* has been frequently identified as the most likely candidate for QTLs associated with salt tolerance and/or Na^+ exclusion in mutant and mapping populations (Ahmadi et al. 2011; Ren et al. 2005; Rus et al. 2006) and found to improve the salinity tolerance of plants (James et al. 2012; James et al. 2006; Munns et al. 2012). Tissue tolerance has been shown to be successfully improved to different extent by vacuolar Na^+/H^+ antiporters (NHX) (Barragan et al. 2012; Barragán et al. 2012; Rodríguez-Rosales et al. 2009), vacuolar H^+ pyrophosphatases (Pasapula et al. 2011), proteins involved in the synthesis of compatible solutes (e.g. proline) (Vendruscolo et al. 2007) and enzymes responsible for the detoxification of reactive oxygen species (Begara-Morales et al. 2014) (**Figure 1**).

Outline of this thesis

The circumscription and limitation of *Lactuca* have been undergoing controversy for centuries, since the genus was established by Linné (Linné 1753). Until recently, the evolutionary relationships within the *Lactuca* genus have still focused on the species in lettuce gene pool. The phylogeny and revision of the entire *Lactuca* genus has never been undertaken. In particular, Africa is one of the most diverse centres of *Lactuca* species, but the African wild species have never been analysed using molecular phylogenetic approaches. The molecular phylogeny using the most extensive sampling (mostly herbarium) of *Lactuca* species will be reconstructed, based on single chloroplast genes and whole chloroplast genomes (Chapter 2 and 3).

Due to the drying methods (high temperature, alchohol etc.), herbarium DNA usually degrades into small fragments and good quality of herbarium DNA has always been difficult to obtain by Sanger sequencing. Recently, complete/partial chloroplast genomes from herbarium samples had been shown to be able to obtain using NGS by Staats et al. (2013). Also, chloroplast phylogenomics had been reported to resolve deep phylogenetic relationships at tribe and species levels (Ma et al. 2014; Nikiforova et al. 2013). However, before I use chloroplast phylogenomics to figure out the phylogentic relationships with *Lactuca*, I need to first perform the phylogenetic analysis of the relationships between species from *Lactuca* and other subfamilies and elucidate the monophyly of *Lactuca*, due to the limited information about these two questions (Chapter 2). Then I can use complete/partial cp genome sequences to resolve deep-level relationships within *Lactuca* (Chapter 3).

In addition, the molecular breeding of *Lactuca* has centred on the domesticated lettuce and its closest relatives, *L. serriola*, *L. saligna* and *L. virosa*. The genetic diversity of different populations derived from cultivated lettuce and these relatives has been studied, especially about the beneficial characters for crops and the resistance to disease. However, the QTLs related to abiotic stress have not yet been as well studied. Previous studies have identified QTLs related to salinity (Hartman et al. 2014; Uwimana et al. 2012a; Uwimana et al. 2012b).

In this thesis, QTL analyses associated with salt tolerance will be performed and a candidate gene related to salt stress will be analysed (Chapter 4 and 5).

Therefore, I will try to address the following issues in this thesis:

In Chapter 2, I provide the most extensive molecular phylogenetic reconstruction of *Lactuca*, based on two chloroplast DNA sequences (*ndhF* and *trnL-F*). *NdhF* and *trnL-F* genes have been used to construct phylogeny in Asteraceae and *Lactuca* (Kim and Jansen 1995; Wang et al. 2013) and therefore were chosen for our phylogenetic analyses. DNA sequences from all the subfamilies of Asteraceae in Genbank and those generated from *Lactuca* herbarium samples were used to establish the affiliation and monophyly of *Lactuca* within Asteracaeae. The sampling covers nearly 40% of the total endemic African *Lactuca* species and 34% of the total *Lactuca* species, and the African endemic species were sequenced for the first time. Biogeographic, chromosomal and morphological character states were also reconstructed over the phylogenetic tree topology.

In Chapter 3, I reveal a deep level of phylogenetic relationships within *Lactuca* using chloroplast phylogenomics, and compared the phylogenetic trees based on the whole chloroplast genome sequences and nuclear Internal Transcribe Spacer (ITS). The taxa used in this study were sequenced together with 93 samples (mostly herbarium) as part of the SYNTHESYS Joint Research Activities 4 (JRA4: Plants/fungi herbarium DNA). A methodology paper of this sequencing project was published, titled as 'Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly (IOGA) pipeline' (Bakker et al. 2015).

In Chapter 4, I have identified quantitative trait loci (QTLs) associated with salt-induced changes in Root System Architecture (RSA) and ion accumulation using a recombinant inbred line population derived from a cross between cultivated lettuce (*L. sativa* 'Salinas') and wild lettuce (*L. serriola*). This study determined regions of lettuce genome contributing to salt-induced changes in RSA and ion accumulation.

In Chapter 5, a previously published QTL region (qLS7.2) from Chapter 4 was found containing one *HKT1* homolog gene of *Arabidopsis thaliana*. I used lettuce seedlings growing in a hydroponic system to test: a) if the *HKT1* gene on Chromosome 7 has different expression patterns and levels between the cultivated and wild lettuce; b) if the difference in the upstream regions (probably promoter regions) of the two *HKT1*s from the two genotypes has an effect on the expression levels and patterns; c) different ion accumulations in lettuce roots and leaves.

In the last chapter, all the results from previous chapters will be discussed and summarized and the future research of *Lactuca* phylogeny and lettuce breeding will be discussed.

References

Ahmadi N, Negrão S, Katsantonis D, Frouin J, Ploux J, Letourmy P, Droc G, Babo P, Trindade H, Bruschi G, Greco R, Oliveira MM, Piffanelli P, Courtois B (2011) Targeted association analysis identified japonica rice varieties achieving Na^+/K^+ homeostasis without the allelic make-up of the salt tolerant indica variety Nona Bokra. Theoretical and Applied Genetics 123:881-895

Ali Z, Park HC, Ali A, Oh DH, Aman R, Kropornicka A, Hong H, Choi W, Chung WS, Kim WY, Bressan RA, Bohnert HJ, Lee SY, Yun DJ (2012) TsHKT1;2, a HKT1 homolog from the extremophile Arabidopsis relative *Thellungiella salsuginea*, shows K^+ specificity in the presence of NaCl. Plant Physiol 158:1463-1474

Argyris J, Dahal P, Hayashi E, Still DW, Bradford KJ (2008) Genetic variation for lettuce seed thermoinhibition is associated with temperature-sensitive expression of abscisic Acid, gibberellin, and ethylene biosynthesis, metabolism, and response genes. Plant physiology 148:926-947

Argyris J, Truco MJ, Ochoa O, Knapp SJ, Still DW, Lenssen GM, Schut JW, Michelmore RW, Bradford KJ (2005) Quantitative trait loci associated with seed and seedling traits in *Lactuca*. Theor Appl Genet 111:1365-1376

Aruga D, Tsuchiya N, Matsumura H, Matsumoto E, Hayashida N (2012) Analysis of RAPD and AFLP markers linked to resistance to Fusarium oxysporum f. sp. lactucae race 2 in lettuce (*Lactuca sativa* L.). Euphytica 187:1-9

Atkinson LD, McHale LK, Truco MJ, Hilton HW, Lynn J, Schut JW, Michelmore RW, Hand P, Pink DA (2013) An intra-specific linkage map of lettuce (*Lactuca sativa*) and genetic analysis of postharvest discolouration traits. Theor Appl Genet 126:2737-2752

Bakker FT, Lei D, Yu J, Mohammadin S, Wei Z, van de Kerke S, Gravendeel B, Nieuwenhuis M, Staats M, Alquezar-Planas DE, Holmer R (2015) Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline. Biol J Linn Soc doi:10.1111/bij.12642

Barragán V, Leidi EO, Andrés Z, Rubio L, De Luca A, Fernández JA, Cubero B, Pardo JM (2012) Ion Exchangers NHX1 and NHX2 Mediate Active Potassium Uptake into Vacuoles to Regulate Cell Turgor and Stomatal Function in Arabidopsis. The Plant Cell 24:1127-1142

Begara-Morales JC, Sánchez-Calvo B, Chaki M, Valderrama R, Mata-Pérez C, López-Jaramillo J, Padilla MN, Carreras A, Corpas FJ, Barroso JB (2014) Dual regulation of cytosolic ascorbate peroxidase (APX) by tyrosine nitration and S-nitrosylation. Journal of experimental botany 65:527-538

Beharav A, Ochoa O, Michelmore R (2013) Resistance in natural populations of three wild *Lactuca* species from Israel to highly virulent Californian isolates of Bremia lactucae. Genetic Resources and Crop Evolution 61:603-609

Beiquan Mou RJH, and Edward J. Ryde (2007) Crisphead Lettuce Breeding Lines with Resistance to Corky Root and Lettuce Mosaic Virus. Hort Sci 42:701-703

Berthomieu P, Conéjéro G, Nublat A, Brackenbury WJ, Lambert C, Savio C, Uozumi N, Oiki S, Yamada K, Cellier F, Gosti F, Simonneau T, Essah PA, Tester M, Véry A-A, Sentenac H, Casse F (2003) Functional analysis of AtHKT1 in *Arabidopsis* shows that Na^+ recirculation by the phloem is crucial for salt tolerance. The EMBO journal 22:2004-2014

Bonnier FJM, Reinink K, Groenwold R (1992) New sources of major gene resistance in *Lactuca* to *Bremia lactucae*. Euphytica 61:203-211

Bonnier FJM, Reinink K, Groenword R (1994) Genetic analysis of *Lactuca* accessions with new major gene resistance to lettuce downy mildew Phytopathology:462-468

Boukema IW, Hazekamp, Th., van Hintum, Th.J.L. (1990) CGN Collection Reviews: The CGN Lettuce Collection. Centre for Genetic Resources, The Netherlands (CGN), Wageningen

Brown PR, Michelmore RW (1988) The genetics of corky root resistance in lettuce. Phytopathology 78:1145-1150

Candresse T, Le Gall O, Mazier M, Maisonneuve B (2006) Virus susceptibility and resistance in lettuce. In: Loebenstein G, Carr J (eds) Natural resistance mechanisms of plant to viruses. Springer, Dordrecht, The Netherlands, pp 383-397

Christopoulou M, McHale LK, Kozik A, Reyes-Chin Wo S, Wroblewski T, Michelmore RW (2015) Dissection of Two Complex Clusters of Resistance Genes in Lettuce (*Lactuca sativa*). Molecular Plant-Microbe Interactions 28:751-765

Chupeau M-C, Maisonneuve B, Bellec Y, Chupeau Y (1994) A *Lactuca* Universal Hybridizer, and its use in creation of fertile interspecific somatic hybrids. Mol Gen Genet 245:139-145

Cole RA, R.A. S, Riggall WE (1991) The use of polyacrylamide gradient gel electrophoresis to identify variation in isozymes as markers for *Lactuca* species and resistance to the lettuce root aphid Pemphigus bursarius. Euphytica 56:237-242

Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical transactions of the Royal Society of London Series B, Biological sciences 363:557-572

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169-196

Cui MH, Yoo KS, Hyoung S, Nguyen HTK, Kim YY, Kim HJ, Ok SH, Yoo SD, Shin JS (2013) An *Arabidopsis* R2R3-MYB transcription factor, AtMYB20, negatively regulates type 2C serine/threonine protein phosphatases to enhance salt tolerance. FEBS letters 587:1773-1778

Davenport RJ, Munoz-Mayor A, Jha D, Essah PA, Rus A, Tester M (2007) The Na^+ transporter AtHKT1;1 controls retrieval of Na^+ from the xylem in *Arabidopsis*. Plant, cell & environment 30:497-507

de Boer AH, Wegner LH (1997) Regulatory mechanisms of ion channels in xylem parenchyma cells. Journal of experimental botany 48:441-449

de VIM (1996) Characterization and identification of *Lactuca sativa* cultivars and wild relatives with SDS-electrophoresis (Lactuca sect. Lactuca, Compositae). Genet Res Crop Evol 43:193-202

De Vries IM, Van Raamsdonk LWD (1994) Numerical morphological analysis of Lettuce cultivars and species (Lactuca sect. Lactuca, Asteraceae). Pl Syst Evol 193:125-141

Deinlein U, Stephan AB, Horie T, Luo W, Xu G, Schroeder JI (2014) Plant salt-tolerance mechanisms. Trends in plant science 19:371-379

Doležalová I, Lebeda A, Dziechciarková M, Křístková E, ASTLEY D, van De Wiel CCM (2003) Relationships among Morphological Characters, Isozymes Polymorphism and DNA Variability – the Impact on *Lactuca* Germplasm Taxonomy. Czech J Genet Plant Breed 39:59-67

Dziechciarková M, Lebeda A, Doležalová I, Astley D (2004) Characterization of *Lactuca* spp. germplasm by proteinand polecular markers – a review. Plant soil environ 50:47-58

Fer áková V, M ájovský J (1977) The genus Lactuca L. in Europe. Bratislava: Univerzita Komenske ho

Groenwold R (1983) Onderzoek van de relatie tussen Lactuca en Bremia Lactucae. Verslag van een voorlichtingsbijeenkomst voor slaveredelaars (vervolg). Zaadbelangen 37:132

Grube R, Wintermantel W, Hand P, Aburomia R, Pink D, Ryder E (2005) Genetic analysis and mapping of resistance to lettuce dieback: A soilborne disease caused by tombusviruses. Theor Appl Genet 110:259-268

Guo R, Yu F, Gao Z, An H, Cao X, Guo X (2011) GhWRKY3, a novel cotton (*Gossypium hirsutum* L.) WRKY gene, is involved in diverse stress responses. Molecular Biology Reports 38:49-58

Hamamoto S, Horie T, Hauser F, Deinlein U, Schroeder JI, Uozumi N (2015) HKT transporters mediate salt stress resistance in plants: from structure and function to the field. Current opinion in biotechnology 32:113-120

Harlan JR (1986) Lettuce and the sycomore: sex and romance in ancient Egypt. Economic Botany 40:4-15

Harlan JR, Wet JMJd (1971) Toward a Rational Classification of Cultivated Plants. Taxon 20:509-517

Hartman Y, Hooftman DAP, Schranz EM, van Tienderen PH (2012a) QTL analysis reveals the genetic architecture of domestication traits in Crisphead lettuce. Genetic Resources and Crop Evolution 60:1487-1500

Hartman Y, Hooftman DAP, Uwimana B, Schranz ME, van de Wiel CC, Smulders MJ, Visser RG, Michelmore RW, van Tienderen PH (2014) Abiotic stress QTL in lettuce crop-wild hybrids: comparing greenhouse and field experiments. Ecology and evolution 4:2395-2409

Hartman Y, Hooftman DAP, Uwimana B, van de Wiel CC, Smulders MJ, Visser RG, van Tienderen PH (2012b) Genomic regions in crop-wild hybrids of lettuce are affected differently in different environments: implications for crop breeding. Evolutionary applications 5:629-640

Hartman Y, Uwimana B, Hooftman DAP, Schranz ME, van de Wiel CC, Smulders MJ, Visser RG, van Tienderen PH (2013) Genomic and environmental selection patterns in two distinct lettuce crop-wild hybrid crosses. Evolutionary applications 6:569-584

Hauser F, Horie T (2010) A conserved primary salt tolerance mechanism mediated by HKT transporters: a mechanism for sodium exclusion and maintenance of high K^+/Na^+ ratio in leaves during salinity stress. Plant, cell & environment 33:552-565

Hayes R, Ryder E (2007) Introgression of novel alleles for partial resistance to big vein disease from *Lactuca virosa* into cultivated lettuce. HortScience 42:35-39

Hayes RJ, McHale LK, Vallad GE, Truco MJ, Michelmore RW, Klosterman SJ, Maruthachalam K, Subbarao KV (2011) The inheritance of resistance to Verticillium wilt caused by race 1 isolates of *Verticillium dahliae* in the lettuce cultivar La Brillante. Theor Appl Genet 123:509-517

Hayes RJ, Ryder EJ, Wintermantel WM (2008) Genetic variation for big-vein symptom expression and resistance to Mirafiori lettuce big-vein virus in *Lactuca virosa* L., a wild relative of cultivated lettuce. Euphytica 164:493-500

Helm J (1954) Lactuca sativa in morphologisch-systematischer Sicht. Kulturpflanze 2:72-129

Horie T, Hauser F, Schroeder JI (2009) HKT transporter-mediated salinity resistance mechanisms in *Arabidopsis* and monocot crop plants. Trends in plant science 14:660-668

Huertas R, Olias R, Eljakaoui Z, Galvez FJ, Li J, De Morales PA, Belver A, Rodriguez-Rosales MP (2012) Overexpression of SlSOS2 (SlCIPK24) confers salt tolerance to transgenic tomato. Plant, cell & environment 35:1467-1482

James RA, Blake C, Zwart AB, Hare RA, Rathjen AJ, Munns R (2012) Impact of ancestral wheat sodium exclusion genes Nax1 and Nax2 on grain yield of durum wheat on saline soils. Functional Plant Biology 39:609-618

James RA, Davenport RJ, Munns R (2006) Physiological Characterization of Two Genes for Na^+ Exclusion in Durum Wheat, Nax1 and Nax2. Plant Physiology 142:1537-1547

Jarvis DE, Ryu CH, Beilstein MA, Schumaker KS (2014) Distinct roles for SOS1 in the convergent evolution of salt tolerance in *Eutrema salsugineum* and *Schrenkiella parvula*. Molecular biology and evolution 31:2094-2107

Jeffrey C (1966) Notes on Compositae: I. Kew Bull 18:427-486

Jenni S, Truco M, Michelmore R (2013a) Quantitative trait loci associated with tipburn, heat stress-induced physiological disorders, and maturity traits in crisphead lettuce. Theoretical and Applied Genetics 126:3065-3079

Jenni S, Truco MJ, Michelmore RW (2013b) Quantitative trait loci associated with tipburn, heat stress-induced physiological disorders, and maturity traits in crisphead lettuce. Theor Appl Genet

Jeuken M, Lindhout P (2002) *Lactuca saligna*, a non-host for lettuce downy mildew (*Bremia lactucae*), harbors a new race-specific Dm gene and three QTLs for resistance. Theor Appl Genet 105:384-391

Jeuken M, van Wijk R, Peleman J, Lindhout P (2001) An integrated interspecific AFLP map of lettuce (*Lactuca*) based on two *L. sativa* × *L. saligna* F_2 populations. Theor Appl Genet 103:638-647

Jeuken MJ, Pelgrom K, Stam P, Lindhout P (2008) Efficient QTL detection for nonhost resistance in wild lettuce: backcross inbred lines versus F_2 population. Theor Appl Genet 116:845-857

Jiang C, Belfield EJ, Mithani A, Visscher A, Ragoussis J, Mott R, Smith JA, Harberd NP (2012) ROS-mediated vascular homeostatic control of root-to-shoot soil Na delivery in *Arabidopsis*. The EMBO journal 31:4359-4370

Johnson WC, Jackson LE, Ochoa O, van Wijk R, Peleman J, St.Clair DA, Michelmore RW (2000) Lettuce, a shallow-rooted crop, and *Lactuca serriola*, its wild progenitor, differ at QTL determining root architecture and deep soil water exploitation. Theor Appl Genet 101:1066-1073

Judd WS, Campbell C, Kellogg EA, Stevens PF, Donoghue MJ (2007) Plant Systematics: A Phylogenetic Approach. Sinauer Assoc., Sunderland

Kadereit JW, Jeffrey C (2007) Flowering Plants. Eudicots: Asterales. In: Kubitzki K (ed) The families and genera of vascular plants, pp 187-189

Katschnig D, Bliek T, Rozema J, Schat H (2015) Constitutive high-level SOS1 expression and absence of HKT1;1 expression in the salt-accumulating halophyte *Salicornia dolichostachya*. Plant science : an international journal of experimental plant biology 234:144-154

Kesseli R, Ochoa O, Michelmore R (1991) Variation at RFLP loci in *Lactuca* spp. and origin of cultivated lettuce (*L. sativa*). Genome 34:430-436

Kesseli R, Witsenboer H, Stanghellini M, Vandermark G, Michelmore R (1993) Recessive resistance to Plasmopara lactucae-radicis maps by bulked segregant analysis to a cluster of dominant disease resistance genes in lettuce. Mol Plant-Microbe Interact 6:722-728

Kesseli RV, Michelmore RW (1986) Genetic variation and phylogenies detected from isozyme markers in species of *Lactuca*. J Hered 77:324-331

Kesseli RV, Paran I, Michelmore RW (1994) Analysis of a Detailed Genetic Linkage Map of *Lactuca* sativa (Lettuce) Constructed From RFLP and RAPD Markers. Genetics 136:1435-1446

Kim K-J, Jansen RK (1995) *ndhF* sequence evolution and the major clades in the sunflower family. Proceedings of the National Academy of Sciences of the United States of America 92:10379-10383

Koopman WJ (2005) Phylogenetic signal in AFLP data sets. Systematic biology 54:197-217

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. American Journal of Botany 85:1517–1530

Koopman WJM, Jong JHd, Vries IMd (1993) Chromosome banding patterns in Lettuce species (Lactuca sect. Lactuca, Compositae). Plant Systematics and Evolution 185:249-257

Koopman WJM, Zevenbergen MJ, Van den Berg RG (2001) Species Relationships in *Lactuca* s.l. (Lactuceae, Asteraceae) Inferred from AFLP Fingerprints. American Journal of Botany 10:1881-1887

Křístková E, Doležalová I, Lebeda A, Vinter V, Novotná A (2008) Description of morphological characters of lettuce (*Lactuca sativa* L.) genetic resources. Hort Sci (Prague) 35:113-129

Kuang H, Ochoa OE, Nevo E, Michelmore RW (2006) The disease resistance gene Dm3 is infrequent in natural populations of *Lactuca serriola* due to deletions and frequent gene conversions at the RGC2 locus. The Plant journal : for cell and molecular biology 47:38-48

Kuang H, van Eck HJ, Sicard D, Michelmore R, Nevo E (2008) Evolution and genetic population structure of prickly lettuce (*Lactuca serriola*) and its RGC2 resistance gene cluster. Genetics 178:1547-1558

Kuang H, Woo SS, Meyers BC, Nevo E, Michelmore RW (2004) Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. The Plant cell 16:2870-2894

Landry BS, Kesseli RV, Farrara B, Michelmore RW (1987) A Genetic Map of Lettuce (*Lactuca sativa* L.) With Restriction Fragment Length Polymorphism, Isozyme, Disease Resistance and Morphological Markers. Genetics 116:331-337

Lebeda A, Doležalová I, Ben-David R, Beharav A (2008) Eco-geographical distribution of *Lactuca saligna* natural populations in Israel. Israel Journal of Plant Sciences 56:195-206

Lebeda A, Doležalová I, Feráková V, Astley D (2004) Geographical Distribution of Wild *Lactuca* Species (Asteraceae, Lactuceae). The Botanical Review 70:328-356

Lebeda A, Doležalová I, Křístková E, Kitner M, Petrželová I, Mieslerová B, Novotná A (2009) Wild *Lactuca* germplasm for lettuce breeding: current status, gaps and challenges. Euphytica 170:15-34

Lebeda A, Křístková E, Kitner M, Mieslerová B, Jemelková M, Pink DAC (2013) Wild *Lactuca* species, their genetic diversity, resistance to diseases and pests, and exploitation in lettuce breeding. European Journal of Plant Pathology 138:597-640

Lebeda A, Reinink K (1994) Histological characterization of resistance in *Lactuca saligna* to lettuce downy mildew (*Bremia lactucae*). Physiological and Molecular Plant Pathology 44:125-139

Lebeda A, Ryder E, Grube R, Dolezalova I, Krıstkova E (2007) Lettuce (Asteraceae; Lactuca spp.) In: Singh R (ed) Genetic resources, chromosome engineering, and crop improvement series, vol 3-vegetable crops. CRC Press, Boca Raton, pp 377-472

Lindqvist K (1960a) CYTOGENETIC STUDIES IN THE *SERRIOLA* GROUP OF *LACTUCA*. Hereditas 46(1-2): 75-151

Lindqvist K (1960b) On the origin of cultivated lettuce. Hereditas 46:319-350

Linné Cv (1753) Species Plantarum, 1 edn

M. Hill, H. Witsenboer, M. Zabeau, P. Vos, R. Kesseli, Michelmore R (1996) PCR-based fingerprinting using AFLPs as a tool for studying genetic relationships in *Lactuca* spp. Theor Appl Genet 93:1202-1210

Ma PF, Zhang YX, Zeng CX, Guo ZH, Li DZ (2014) Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (poaceae). Systematic biology 63:933-950

Maischak H, Zimmermann MR, Felle HH, Boland W, Mithöfer A (2010) Alamethicin-induced electrical long distance signaling in plants. Plant Signaling & Behavior 5:988-990

Maisonneuve B, Bellec YA, P., Michelmore RW (1994) Rapid mapping of two genes for resistance to downy mildew from *Lactuca serriola* to existing clusters of resistance genes. Theor Appl Genet 89:96-104

Maisonneuve B, Chupeau MC, Bellec Y, Chupeau Y (1995) Sexual and somatic hybridization in the genus *Lactuca*. Euphytica 85 85:281-285

Mano Y, Takeda K (1997) Mapping quantitative trait loci for salt tolerance at germination and the seedling stage in barley (*Hordeum vulgare* L.). Euphytica 94:263-272

Matoba H, Mizutani T, Nagano K, Hoshi Y, Uchiyama H (2007) Chromosomal study of lettuce and its allied species (*Lactuca* spp., Asteraceae) by means of karyotype analysis and fluorescence in situ hybridization

McHale LK, Truco MJ, Kozik A, Wroblewski T, Ochoa OE, Lahre KA, Knapp SJ, Michelmore RW (2009) The genomic architecture of disease resistance in lettuce. Theor Appl Genet 118:565-580

Mejías JA (1993) Cytotaxonomic studies in the Iberian taxa of the genus *Lactuca* (Compositae). Bot Helvet 103:113-130

Meyers BC, Shen KA, Rohani P, Gaut BS, Michelmore RW (1998) Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. The Plant cell 10:1833-1846

Mikel MA (2007) Genealogy of contemporary North American lettuce. Hortscience 42:489-493

Mittler R, Vanderauwera S, Suzuki N, Miller G, Tognetti VB, Vandepoele K, Gollery M, Shulaev V, Van Breusegem F (2011) ROS signaling: the new wave? - 16:- 309

Mizutani T, Liu XJ, Tashiro Y, Miyazaki S, Shimasaki K (1989) Plant regeneration and cell fusion of protoplasts from lettuce cultivars and related wild species in Japan. Bulletin of the Faculty of Agriculture, Saga University 67:109-118

Montesclaros L, Nicol N, Ubalijoro E, Leclerc-Potvin C, Ganivet L, Lalibert ́e J, Fortin M (1997) Response to potyvirus infection and genetic mapping of resistance loci to potyvirus infection in *Lactuca*. Theor Appl Genet 94:941-946

Moreno-Vazquez S, Ochoa OE, Faber N, Chao S, Jacobs JM, Maisonneuve B, Kesseli RV, Michelmore RW (2003) SNP-based codominant markers for a recessive gene conferring resistance to corky root rot (*Rhizomonas suberifaciens*) in lettuce (Lactuca sativa). Genome 46:1059-1069

Mou B, Bull C (2004) Screening Lettuce Germplasm for New Sources of Resistance to Corky Root. J Amer Soc Hort Sci 129:712-716

Munns R, James RA, Xu B, Athman A, Conn SJ, Jordans C, Byrt CS, Hare RA, Tyerman SD, Tester M, Plett D, Gilliham M (2012) Wheat grain yield on saline soils is improved by an ancestral Na^+ transporter gene. Nature Biotechnology 30:360-364

Munns R, Tester M (2008) Mechanisms of salinity tolerance. Annual review of plant biology 59:651-681

Nicaise V (2003) The Eukaryotic Translation Initiation Factor 4E Controls Lettuce Susceptibility to the Potyvirus Lettuce mosaic virus. Plant physiology 132:1272-1282

Nikiforova SV, Cavalieri D, Velasco R, Goremykin V (2013) Phylogenetic analysis of 47 chloroplast genomes clarifies the contribution of wild species to the domesticated apple maternal line. Molecular biology and evolution 30:1751-1760

Olsen AN, Ernst HA, Leggio LL, Skriver K (2005) NAC transcription factors: structurally distinct, functionally diverse. Trends in plant science 10:79-87

Paran I, Kesseli RV, Michelmore RW (1991) Identification of restriction fragment length polymorphism and random amplified polymorphic DNA markers linked to downy mildew resistance genes in lettuce, using near-isogenic lines. Genome 34:1021-1027

Paran I, Michelmore RW (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theor Appl Genet 85:985-993

Pasapula V, Shen G, Kuppu S, Paez-Valencia J, Mendoza M, Hou P, Chen J, Qiu X, Zhu L, Zhang X, Auld D, Blumwald E, Zhang H, Gaxiola R, Payton P (2011) Expression of an *Arabidopsis* vacuolar H^+-pyrophosphatase gene (AVP1) in cotton improves drought- and salt tolerance and increases fibre yield in the field conditions. Plant biotechnology journal 9:88-99

Pink D, Lot H, Johnson R (1992a) Novel pathotypes of lettuce mosaic virus - breakdown of a durable resistance? Euphytica 63:169-174

Pink DAC, Kostova D, Walkey DGA (1992b) Differentiation of pathotypes of *Lettuce mosaic virus*. Plant Pathol 41:5-12

Platten JD, Egdane J, Ismail A (2013) Salinity tolerance, Na^+ exclusion and allele mining of HKT1;5 in *Oryza sativa* and *O. glaberrima*: many sources, many genes, one mechanism? BMC Plant Biology 13:32

Qiu Q-S, Guo Y, Dietrich MA, Schumaker KS, Zhu J-K (2002) Regulation of SOS1, a plasma membrane Na^+/H^+ exchanger in *Arabidopsis thaliana*, by SOS2 and SOS3. Proceedings of the National Academy of Sciences 99:8436-8441

Rafalski J, Tingey S (1993) Genetic diagnostics in Plant Breed: RAPDs, microsatellites and machines. Trends Genet 9:275-280

Ren Z-H, Gao J-P, Li L-G, Cai X-L, Huang W, Chao D-Y, Zhu M-Z, Wang Z-Y, Luan S, Lin H-X (2005) A rice quantitative trait locus for salt tolerance encodes a sodium transporter 37:1141-1146

Revers F, Lot H, Souche S, Le Gall O, Candresse T, Dunez J (1997) Biological and molecular variability of lettuce mosaic virus isolates. Phytopathology 87:397-403

Ribaut J-M, Hoisington D (1998) Marker-assisted selection: New tools and strategies. Trends in plant science 3:236-239

Robbins M, Witsenboer H, Michelmore R, Laliberte J, Fortin M (1994) Genetic mapping of turnip mosaic virus resistance in *Lactuca sativa*. Theor Appl Genet 89:583-589

Rodenburg CM (1960) Varieties discription. Varieties of Lettuce: an international monograph. Tjeenk Willink, Zwolle, p 260

Rodríguez-Rosales MP, Gálvez FJ, Huertas R, Aranda MN, Baghour M, Cagnac O, Venema K (2009) Plant NHX cation/proton antiporters. Plant Signaling & Behavior 4:265-276

Roy SJ, Negrao S, Tester M (2014) Salt resistant crop plants. Current opinion in biotechnology 26:115-124

Rulkens AJH (1987) DE CGN sla collectie: inventarisatie, paspoort gegevens en enkele richtlijnen voor de toekomst. CGN report: CGN-T 1987-1. CGN, Wageningen

Rus A, Baxter I, Muthukumar B, Gustin J, Lahner B, Yakubova E, Salt DE (2006) Natural variants of *AtHKT1* enhance Na^+ accumulation in two wild populations of Arabidopsis. PLoS genetics 2:e210

Rus A, Lee BH, Munoz-Mayor A, Sharkhuu A, Miura K, Zhu JK, Bressan RA, Hasegawa PM (2004) *AtHKT1* facilitates Na^+ homeostasis and K+ nutrition in planta. Plant Physiol 136:2500-2511

Ryder E (1999) Lettuce, Endive and Cichory. CABI Publishing, Wallingford

Ryder E (2002) A mild systemic reaction to lettuce mosaic virus in lettuce: Inheritance and interaction with an allele for resistance. J Am Soc Hort Sci 127:814-818

Ryder EJ (1970a) Inheritance of resistance to common lettuce mosaic. J Amer Soc Hortic Sci 95:378-379

Ryder EJ (1970b) Screening for resistance to lettuce mosaic. HortSci 5:47-48

Ryder EJ (1986) Lettuce breeding. In: Bassett M (ed) Breeding Vegetable Crops. AVI Publishing Co., Westport, pp 433-474

Schwember AR, Bradford KJ (2010a) A genetic locus and gene expression patterns associated with the priming effect on lettuce seed germination at elevated temperatures. Plant molecular biology 73:105-118

Schwember AR, Bradford KJ (2010b) Quantitative trait loci associated with longevity of lettuce seeds under conventional and controlled deterioration storage conditions. Journal of experimental botany 61:4423-4436

Shih C (1987) On the circumscription of the genus *Prenanthes* L. and *Notoseris* Shih: a new genus of Compositae from China. Acta Phytotaxon Sin 25

Shih C (1988a) Revision of *Lactuca* L. and two new genera of tribe Lactuceae(Compositae) on the mainland of Asia. Acta Phytotaxon Sin 26:382-393

Shih C (1988b) Revision of *Lactuca* L. and two new genera of tribe Lactuceae(Compositae) on the mainland of Asia. Acta Phytotaxon Sin 26:418-428

Shih C (1991) On circumscription of the genus *Cicerbita* Wall., and two new genera of Compositae from Sino-Himalayan region. Acta Phytotaxon Sin 29:394-417

Shih C (1997) Compositae (10), Cichorioideae, Lactuceae. Beijing: Science Press

Shih C, Kilian N (2011) Cichorieae [excl.*Tragopogon*, *Taraxacum*, *Hieracium* and *Pilosella*]. Science Press & St Louis: Missouri Botanical Garden, Beijing

Simko I (2013) Marker-Assisted Selection for Disease Resistance in Lettuce. In: Tuberosa RKVaR (ed) Translational Genomics for Crop Breeding: Biotic Stress. John Wiley & Sons Ltd, Chichester, UK

Simko I, Atallah AJ, Ochoa OE, Antonise R, Galeano CH, Truco MJ, Michelmore RW (2013) Identification of QTLs conferring resistance to downy mildew in legacy cultivars of lettuce. Scientific reports 3:2875

Simko I, Pechenick D, McHale L, Truco M, Ochoa O, Michelmore R, Scheffler B (2010) Development of molecular markers for marker-assisted selection of dieback disease resistance in lettuce (*Lactuca sativa*). Acta Hortic 859:401-408

Simko I, Pechenick DA, McHale LK, Truco MJ, Ochoa OE, Michelmore RW, Scheffler BE (2009) Association mapping and marker-assisted selection of the lettuce dieback resistance gene Tvr1. BMC plant biology 9:135

Simko I, Rauscher G, Sideman RG, McCreight JD, Hayes RJ (2014) Evaluation and QTL mapping of resistance to powdery mildew in lettuce. Plant Pathology 63:344-353

Soleimani P, Mosahebi G, Habibi MK (2011) Identification of some viruses causing mosaic on lettuce and characterization of Lettuce mosaic virus from Tehran Province in Iran. African Journal of Agricultural Research 6:3029-3035

Staats M, Erkens RH, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J, Richardson JE, Bakker FT (2013) Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. PloS one 8:e69189

Stebbins GL (1937) The Scandent Species of *Prenanthes* and *Lactuca* in Africa. Bulletin du Jardin botanique de l'État a Bruxelles 14:333-352

Su Y, Luo W, Lin W, Ma L, Kabir MH (2015) Model of Cation Transportation Mediated by High-Affinity Potassium Transporters (HKTs) in Higher Plants. Biological procedures online 17:1

Suzuki N, Koussevitzky S, Mittler R, Miller G (2012) ROS and redox signalling in the response of plants to abiotic stress. Plant, cell & environment 35:259-270

Tanaka TMaT (2003) Genetic analyses of isozyme in lettuce, *Lactuca sativa*, and its relatives. J Japan Soc Hort Sci 72:`122-127

Thompson RC, Whitaker TW, Kosar WF (1941) Interspecific genetic relationships in *Lactuca*. Journal of Agricultural Research 63:91-107

Truco MJ, Antonise R, Lavelle D, Ochoa O, Kozik A, Witsenboer H, Fort SB, Jeuken MJ, Kesseli RV, Lindhout P, Michelmore RW, Peleman J (2007) A high-density, integrated genetic linkage map of lettuce (*Lactuca* spp.). Theor Appl Genet 115:735-746

Truco MJ, Ashrafi H, Kozik A, van Leeuwen H, Bowers J, Reyes Chin Wo S, Stoffel K, Xu H, Hill T, Van Deynze A, Michelmore RW (2013) An Ultra High-Density, Transcript-Based, Genetic Map of Lettuce. G3 (Bethesda) 3:617–631

U.S.D.A. (2011) U.S. Lettuce Statistics.

Uwimana B, Smulders MJ, Hooftman DA, Hartman Y, van Tienderen PH, Jansen J, McHale LK, Michelmore RW, van de Wiel CC, Visser RG (2012a) Hybridization between crops and wild relatives: the contribution of cultivated lettuce to the vigour of crop-wild hybrids under drought, salinity and nutrient deficiency conditions. Theor Appl Genet 125:1097-1111

Uwimana B, Smulders MJ, Hooftman DA, Hartman Y, van Tienderen PH, Jansen J, McHale LK, Michelmore RW, Visser RG, van de Wiel CC (2012b) Crop to wild introgression in lettuce: following the fate of crop genome segments in backcross populations. BMC plant biology 12:43

Van Bruggen AHC, Brown PR, Jochimsen KN (1989) Corky Root of Lettuce Caused by strains of a Gram-Negative Bacterium from Muck Soils of Florida, New York, and Wisconsin. Applied and Environmental Microbiology 55:2635-2640

Van Bruggen AHC, Jochimsen KN (1992) First report of Rhizomonas sp. causing corky root of lettuce in Europe. Neth J P1 Path 98:45-56

van de Wiel C, Arens P, Vosman B (1998) Microsatellite fingerprinting in lettuce (*Lactuca sativa* L.) and wild relatives. Plant Cell Reports 17:837-842

van de Wiel C, Arens P, Vosman B (1999) Microsatellite retrieval in lettuce (*Lactuca sativa* L.). Genome 42:139-149

van Treuren R, van der Arend AJM, Schut JW (2011) Distribution of downy mildew (*Bremia lactucae* Regel) resistances in a genebank collection of lettuce and its wild relatives. Plant Genetic Resources 11:15-25

Vandemark G, Stanghellini M, Rasmussen S, Michelmore R (1991) Inheritance of resistance in lettuce to *Plasmopara lactucae-radicis*. Phytopathology 82:272-274

Vendruscolo ECG, Schuster I, Pileggi M, Scapim CA, Molinari HBC, Marur CJ, Vieira LGE (2007) Stress-induced synthesis of proline confers tolerance to water deficit in transgenic wheat. Journal of plant physiology 164:1367-1376

Vries IMd (1997) Origin and domestication of *Lactuca sativa* L. Genetic Resources and Crop Evolution 44: 165–174

Wang ZH, Peng H, Kilian N (2013) Molecular phylogeny of the *Lactuca* alliance (Cichorieae subtribe Lactucinae, Asteraceae) with focus on their Chinese centre of diversity detects potential events of reticulation and chloroplast capture. PloS one 8:e82692

Wegner LH, De Boer AH (1997) Properties of two outward-rectifying channels in root xylem parenchyma cells suggest a role in K^+ homeostasis and long-distance signaling. Plant Physiology 115:1707-1719

Whitaker TW (1969) Salads for Everyone: A Look at the Lettuce Plant Economic Botany 23:261-264

Whitaker TW, Thompson RC (1941) Cytological Studies in Lactuca Bulletin of the Torrey Botanical Club 68:388-394

Witsenboer H, Vogel J, Michelmore RW (1997) ldentif ication, genetic localization, and allelic diversity of selectively amplified microsatellite polymorphic loci in lettuce and wild relatives (*Lactuca* spp.). Genome 40:923-936

Yabuuchi E, Kosako Y, Naka T, Suzuki S, Yano I (1999) Proposal of *Sphingomonas suberifaciens* (van Bruggen, Jochimsen and Brown 1990) comb. nov., *Sphingomonas natatoria* (Sly 1985) comb. nov., *Sphingomonas ursincola* (Yurkov et al. 1997) comb. nov., and emendation of the genus *Sphingomonas*. Microbiol Immunol 43:339-349

Yamaguchi M (1983) World vegetables : principles, production and nutritive values. Ellis Horwood Chichester, GB

Yang T-J, Jang S-W, Kim W-B (2007) Genetic relationships of *Lactuca* spp. revealed by RAPD, inter-SSR,AFLP and PCR-RFLP analyses. J Crop Sci Biotech 29:29-34

Zeven AC, de Wet JMJ (1982) Dictionary of cultivated plants and their centres of diversity. Pudoc, Wageningen, the Netherlands

Zhang FZ, Wagstaff C, Rae AM, Sihota AK, Keevil CW, Rothwell SD, Clarkson GJ, Michelmore RW, Truco MJ, Dixon MS, Taylor G (2007) QTLs for shelf life in lettuce co-locate with those for leaf biophysical properties but not with those for leaf developmental traits. Journal of experimental botany 58:1433-1449

Zhang NW, Lindhout P, Niks RE, Jeuken MJM (2009) Genetic dissection of *Lactuca saligna* nonhost resistance to downy mildew at various lettuce developmental stages. Plant Pathology 58:923-932

Zohary D (1983) Wild genetic resources of crops in Israel. Israel Journal of Botany 32:97-127

Zohary D (1991) The wild genetic resources of cultivated lettuce (*Lactuca saliva* L . Euphytica 53:31-35

Chapter 2

Phylogenetic relationships within *Lactuca* L. (Asteraceae), including African species, based on chloroplast DNA sequence comparisons

Zhen Wei[1], Shi-Xin Zhu[2], R. G. Van den Berg[1], Freek T. Bakker[1], M. Eric Schranz[1]

1 Biosystematics Group, Wageningen University and Research Centre, 6708 PB Wageningen, The Netherlands

2 School of Life Sciences, Zhengzhou University, 450001 Zhengzhou, China

Published in *Genetic Resources and Crop Evolution* DOI: 10.1007/s10722-015-0332-5

Chapter 2

Abstract

Lettuce (*Lactuca sativa* L.) belongs to the genus *Lactuca* L. and is an important vegetable worldwide. Over the past decades, there have been many controversies about the phylogeny of *Lactuca* species due to their complex and diverse morphological characters and insufficient molecular sampling. In this study we provide the most extensive molecular phylogenetic reconstruction of *Lactuca*, including African wild species, using two chloroplast genes (*ndhF* and *trnL-F*). The sampling covers nearly 40% of the total endemic African *Lactuca* species and 34% of the total *Lactuca* species. DNA sequences from all the subfamilies of Asteraceae in Genbank and those generated from *Lactuca* herbarium samples were used to establish the affiliation of *Lactuca* within Asteracaeae. Based on the subfamily tree, we selected 33 *ndhF* sequences from 30 species and 79 *trnL-F* sequences from 48 species to infer relationships within the genus *Lactuca* using Randomized Axelerated Maximum Likelihood (RAxML) and Bayesian Inference (BI) analyses. Biogeographical, chromosomal and morphological character states were reconstructed over the Bayesian tree topology. We conclude that *Lactuca* contains two distinct phylogenetic clades - the crop clade and the *Pterocypsela* clade. Other North American, Asian and widespread species either form smaller clades or mix with the *Melanoseris* species. The newly sampled African endemic species probably should be treated as a new genus.

Key words

African *Lactuca*; *Lactuca* phylogeny; lettuce; *ndhF*; phylogenetic relationships; *trnL-F*

Introduction

Domesticated lettuce (*Lactuca sativa* L.) is a member of the genus *Lactuca* L., which is grouped in the subtribe Lactucinae, tribe Cichorieae (Lactuceae), subfamily Cichorioideae of the family Asteraceae (Compositae) (Judd et al. 2007; Kadereit and Jeffrey 2007). As one of the most important vegetables, lettuce is commercially produced worldwide, especially in Asia, North and Central America, and Europe (Lebeda et al. 2007). There are a large number of lettuce cultivars within *L. sativa*. These cultivars can be divided in seven distinct cultivar groups: Butterhead Group, Crisphead Group, Cos Group, Cutting Group, Stalk Group, Latin Group and Oilseed Group (Vries 1997). Many studies have focused on domesticated lettuce (Hartman et al. 2012; Kerbiriou et al. 2013; Uwimana et al. 2012; Zhang et al. 2009). However, there are still uncertainties about the phylogenetic relationships within *Lactuca,* mainly due to the complex and variable morphological characters of the species in the genus. Some of the controversies stem from the different circumscriptions proposed for the genus, which vary from extremely broad to very narrow concepts. Bentham (1873) included *Lactuca* species not only from the present subtribe Lactucinae, but also from the present subtribes Crepidinae and Hyoseridinae; this broad concept was maintained by Hoffmann (1890-1894). Stebbins (1937a, 1937b; 1939), Feráková and Májovský (1977) and Lebeda et al. (2004; 2007) used a moderately wide concept of *Lactuca* that comprised a total of approximately 100 species. Tuisl (1968), Shih (1988a, b), and Kadereit and Jeffrey (2007) established a narrow circumscription. In this concept, Shih and Kilian (2011) consider there to be between 50 - 70 *Lactuca* species. However, all these authors mentioned before only dealt mostly with regional *Lactuca* species and the genus has never been revised in its entirety.

Lebeda et al. (2004) provided an overview of the biogeographical distribution of wild *Lactuca* species based on the available literature data and showed that Asia (containing 51 species) and Africa (containing 43 species) are the two centres of diversity for *Lactuca* species. Lebeda et al. (2004; 2009) elaborated a classification of *Lactuca* from taxonomic and biogeographical criteria and divided the genus into seven sections (*Lactuca* (subsection *Lactuca* and *Cyanicae* DC.), *Phaenixopus* (Cass.) Bentham, *Mulgedium* (Cass.) C.B. Clarke, *Lactucopsis* (Schultz Bip. Ex Vis. Et Pančić) Rouy, *Tuberosae* Boiss., *Micranthae* Boiss., *Sororiae* Franchet) and two geographical groups (African and North American). Recently, Wang et al. (2013) constructed a DNA-based phylogenetic tree of the *Lactuca* alliance with a focus on the Chinese centre of diversity. This study fills the gap in our understanding of Asian diversity centre of *Lactuca* species and related genera, especially for the Chinese species. However, a study of the African diversity centre of *Lactuca* species is still lacking.

Despite the lack of studies focused on the entire *Lactuca* genus, there have been a number of studies focused on cultivated lettuce and closely-related wild species. These studies concentrated on aspects of interest for lettuce breeding to improve growth related to abiotic and biotic stresses using genetic resources from wild lettuce species (Hartman et al. 2014; Hartman et al. 2012; Jeuken MJ 2008; van Treuren et al. 2011). Zohary (1991) established a

concept of the 'lettuce gene pool' and Koopman et al. (1998; 2001) modified Zohary's lettuce gene pool concept and provided the first molecular phylogenetic relationships among *Lactuca* species based on nrDNA ITS-1 and AFLPs. Koopman et al. (1998) described *L. sativa*, *L. serriola* L., *L. dregeana* DC., *L. aculeata* Boiss. and *L. altaica* Fischer & C.A.Meyer as the primary gene pool, *L.virosa* L. and *L. saligna* L. as the secondary gene pool, and *L. quercina* L., *L. viminea*, *L. sibirica* Benth. ex Maxim. and *L. tatarica* (L.) C.A. Meyer as the tertiary gene pool. Apart from Koopman et al. (2001) and Wang et al. (2013), there is limited information about the molecular phylogenetic relationships within the genus *Lactuca*, especially for the African species since they were first described (Jeffrey 1966; Stebbins 1937b).

More than 4000 years ago, the Egyptians started to cultivate wild lettuce (*L. serriola*) in Africa and this species is thought to be the ancestor of modern lettuce cultivars (Harlan 1986). Lindqvist (1960) doubted that only *L. serriola* was involved in the domestication of the cultivated lettuce, but he did not specify what species might have played a role. Kesseli et al. (1991) suggested a polyphyletic origin of *L. sativa* using RFLP loci. Mikel (2007) reported that apart from *L. serriola*, the current crisphead cultivar 'Salinas' was also derived from *L. virosa* for its robust root system and decreased leaf drop. Wei et al. (2014), using a recombinant inbred line population derived from *L. sativa* 'Salinas' (crop) and *L. serriola* (wild), found that alleles from the cultivated lettuce contribute more to lateral root development than those from wild lettuce.

The aim of this present study is to provide a DNA based phylogenetic tree of *Lactuca*, and 34 % of known *Lactuca* species and 40% of the total endemic African *Lactuca* species were included in the taxon sampling. We reconstruct ancestral states for geographic areas, chromosome number and selected morphological characters over the phylogenetic trees. Novel potential genetic resources for lettuce breeding are proposed as well.

Materials and methods

Taxon sampling

Twenty-seven *Lactuca* species, including thirteen African endemic species, and four species from *Lactuca*-allied genera were sampled (**Table 1**). For the species *L. viminea* two samples representing two subspecies were included. Following the treatment of Lebeda et al. (2004), this sampling represents 34% of the total *Lactuca* species and 40% of the total endemic African species. The 32 samples come from fresh leaf, sillica-dried leaf and herbarium specimens (**Table 1**). Four of the fresh-collected materials were from Centre for Genetic Resources, the Netherlands (CGN, http://www.wageningenur.nl/en/Expertise-Services/Statutory-research-tasks/Centre-for-Genetic-Resources-the-Netherlands-1.htm).
Herbarium materials were provided by the National Herbarium of the Netherlands (WAG) and the Botanic Garden and Botanical Museum Berlin-Dahlem (B), herbarium codes

Table 1 Taxon sampling information (including herbarium specimen, silica-dried and fresh materials).

No.	Taxon name	Collection number	Deposited in[a]	Collected year	Sample type[b]	Country of origin	Note[c]
1	*Lactuca aculeata* Boiss.	Koopman, W.J.M.; CGN15692	WAG	1995	F	Turkey	
2	*L. altaica* Fischer & C.A.Meyer (*L. serriola* 2)	Koopman, W.J.M.; CGN15711	WAG	1995	F	Georgia	
3	*L. attenuata* Stebbins	Lewalle, J.; 5982	WAG	1971	H	Burundi	*
4	*L. calophylla* C.Jeffrey	Pawek, J.; 12254	WAG	1977	H	Malawi	*
5	*L. formosana* Maximowicz	Zhu, S.X.; 2011-1576	HEAC	2011	S	China	3
6	*L. glandulifera* Hook.f.	Breteler, F.J.; 111	WAG	1962	H	Cameroon	*
7	*L. imbricata* Hiern	Witte, G.F. de; 7284	WAG	1949	H	Congo	2*
8	*L. indica* L.	Zhu, S.X.; 2010-1191	HEAC	2010	S	China	3
9	*L. inermis* Forssk.	Jongkind, C.C.H.; 2635	WAG	1996	H	Ghana	
10	*L. lasiorhiza* (O.Hoffm.) C.Jeffrey	Phillips, E.; 4048	WAG	1978	H	Malawi	*
11	*L. orientalis* Boiss.	Bayer, Ch.; B 100191996	B	1989	H	Jordan	2
12	*L. paradoxa* Sch.Bip. ex A.Rich.	Friis, I. et al.; 491	WAG	1970	H	Ethiopia	*
13	*L. perennis* L.	Wieringa, J.J.; 5779	WAG	2006	S	France	
14	*L. praevia* C.D.Adams	Simons, E.L.A.N.; 855	WAG	2012	H	Guinea	1*
15	*L. raddeana* Maximowicz	Zhu, S.X.; 09-208	HEAC	2009	S	China	3
16	*L. saligna* L.	Koopman, W.J.M.; CGN15705	WAG	1991	F	Georgia	
17	*L. schulzeana* Büttner	Pauwels, L.; 5453	WAG	1976	H	Cameroon	2*
18	*L. schweinfurthii* Oliv. & Hiern	Wilde, W.J.J.O. de; 2528	WAG	1964	H	Cameroon	*
19	*L. serriola* L. 1	Jeuken, MJW; MJ19	L	2013	F	Turkey	3
20	*L. setosa* Stebbins ex C.Jeffrey	Blittersdorff, R. von; B100426945	B	2011	H	Tanzania	*
21	*L. tatarica* (L.) C.A. Meyer	Koopman, W.J.M.; 397	WAG	1996	H	Netherlands	
22	*L. tenerrima* Pourr.	Wilde, J.J.F.E. de; 3038	WAG	1961	H	Morocco	
23	*L. tinctociliata* I.M.Johnst. (*Launaea cornuta* (Hochst. ex Oliv. & Hiern) C.Jeffrey)	Masens, B.; 180	WAG	1990	H	Congo	*
24	*L. ugandensis* C.Jeffrey (*Lactuca* sp.)	Wilde, W.J.J.O. de; 2457	WAG	1964	H	Cameroon	*
25	*L. viminea* subsp. *chondrilliflora* (Boreau) Malag.	Lewalle, J.; 10014	WAG	1981	H	Morocco	
26	*L. viminea* subsp. *ramosissima* (All.) Malag.	Wieringa, J.J.; 5974	WAG	2007	H	France	1
27	*L. virosa* L.	CGN09364	L	2013	F	Iran	**

28	*L. zambeziaca* C.Jeffrey	Niangadouma, R.; 391	WAG	2004	H	Gabon	*
29	*Cicerbita alpina* Wallr.	Breteler, F.J.; 7538	WAG	1977	H	France	
30	*Notoseris triflora* (Hemsl.) C.Shih	Zhu, S.X.; 2012-1818	HEAC	2012	S	China	3
31	*Paraprenanthes diversifolia* (Vaniot) N.Kilian	Zhu, S.X.; 2012-1817	HEAC	2012	S	China	3
32	*Prenanthes purpurea* (Vaniot) N.Kilian	Wieringa, J.J.; 5375	WAG	2004	H	France	

[a] Refer to Index Herbariorum (Thiers B 2011)
[b] H-herbarium, F-fresh, S- silica-dried
[c] * African endemic species (Lebeda et al. 2004); ** seeds of the same accession can be required for free; 1 means the plastid gene sequences were obtained by Sanger sequencing; 2 indicates NGS and Sanger sequencing for this sample both failed; 3 voucher specimen are being submitted to herbarium.

Table 2 Characteristics of individual gene alignment and concatenated plastid matrix.

Data set	No. of char.[a]/No. of char.[b]	No. of parsimony inform. sites[a]/No. of inform. sites[b]
trnL-F	863/853	65(7.5%)/58(6.8%)
ndhF	2251/2250	71(3.2%)/70(3.1%)
trnL-F+ndhF	3114/3103	136(4.4%)/128(4.1%)

char. character, *inform.* informative
[a] With indel
[b] Without indel

following Thiers (2011). All necessary permissions for the described plants and specimen samplings were obtained from the respective curators, dr. ir. J.J. Wieringa (Naturalis Biodiversity Center, Leiden) and dr. Norbert Kilian (Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Berlin).

DNA extraction and purification

DNA was extracted from 10 - 30 mg of plant material using the cetyltrimethyl-ammonium-bromide (CTAB) method (Doyle and Doyle 1987), modified for herbarium specimens as in Särkinen et al. (2012) and Staats et al. (2011). The DNA extraction was then purified by Wizard DNA clean-up system (Promega Corp.) with a vacuum manifold (Promega Corp.) The quality of the DNA extractions was visualized on 1% agarose gel and measured by Qubit 2.0 Fluorometer (Invitrogen). Polymerase chain reaction and Sanger sequencing were also performed for some of the herbarium samples to check for potential degradation of DNA. PCR amplifications were performed in 10 µl reactions using MyTaq^{TM} DNA polymerase (Bioline, London, UK). Thermal cycling for PCR included 2 min. at 95 ℃, followed by 30 cycles of 30 sec. at 95 ℃, 30 sec. at 50 ℃, 1min. at 70 ℃, and ended by 5 min. at 72 ℃. The forward and reverse primer sequences of *trnL-F* were 5'-GCAATCCTGAGCCAAATCC-3' and 5'-GCTCGATGCATCATCCCGCTAAA-3', respectively. Two pairs of primers (*ndhF* 5' forward-1074 reverse and 913 forward- *ndhF* 3' reverse) were used for the amplification of *ndhF* due to the large size of the gene (Karis et al. 2001). PCR products were then purified and sequenced as described in Schneider et al. (2014).

Next Generation Sequencing and *de novo* assembly

The dataset of plastid gene sequences presented in this work was generated as part of the SYNTHESYS Joint Research Activities 4 (JRA4: Plants/fungi herbarium DNA: http://www.synthesys.info/joint-research-activities/synthesys-2-jras/jra4-plantsfungi-optimised-dna-extraction-techniques/). The *Lactuca* samples were sequenced by National High-Throughput DNA Sequencing Centre of University of Copenhagen, using the next generation sequencing Illumina HiSeq 2000 platform (http://seqcenter.ku.dk/facilities/). The protocols for DNA library preparation and PCR amplification was described in Bakker et al. (2015). Contig assembly and read clean-up were performed using standard method similar to the 'MitoBIM' approach outlined in Hahn et al. (2013) for mitochondrial genomes. This method is called the *Iterative Organelle Genome Assembly* pipeline (IOGA), aiming to assemble paired-end reads into a series of candidate assemblies and selecting the best one based on likelihood estimation (Bakker et al. 2015). The IOGA pipeline can be briefly described in the following steps: (1) Trimmomatic was used to trim low quality, adapter and other Illumina-specific sequences from individual reads (Bolger et al. 2014); (2) chloroplast genome-derived reads were filtered out of the entire read pool in Bowtie 2, by aligning the latter to a range of reference Angiosperm chloroplast genome sequences (Langmead and Salzberg 2012); (3) de novo assemblies from the trimmed, filtered and corrected chloroplast

reads, were performed in SOAPdenovo2, using k-mer values ranging from 37-97 (Luo et al. 2012); (4) 'best assemblies' were selected using the N50 criterion and then used as a new reference to find target-specific reads not selected in the first iteration; (5) step 4 was repeated until no more chloroplast genome-derived reads were found, followed by assembly of the final set of assemblies with SPAdes3.0 (Bankevich et al. 2012), under a range of different k-mer settings; (6) finally, Assembly Likelihood Estimation (Clark et al. 2013) was performed to select the best assembly (LnL score) among candidate assemblies as the final assembly. Chloroplast genes (*trnL-F* and *ndhF*) were annotated and extracted in DOGMA (Wyman et al. 2004). The IOGA script can be obtained from Github at https://github.com/holmrenser/IOGA.

Sequence alignment and phylogenetic analyses

From GenBank we obtained 218 *ndhF* gene sequences from 211 species and 301 *trnL-F* gene sequences from 250 species by Blasting *L. sativa*, *L. inermis* Forssk., *L. paradoxa* Sch.Bip. ex A.Rich. and *L. canadensis* A.Gray (**Table S1** and **Table S2**) against the NCBI nucleotide database. This sampling comprises a wide range of taxa from all the subfamilies in Asteraceae, according to the Angiosperm Phylogeny Website (http://www.mobot.org/MOBOT/research/APweb/). Together from with the *Lactuca* sequences generated in this study, we achieved 34 % taxonomic sampling for *Lactuca*. *Barnadesia caryophylla* was selected as outgroup based on the phylogenetic tree of Asteraceae in APG (http://www.mobot.org/MOBOT/research/APweb/trees/asteraceae.gif). All the DNA sequences were first automatically aligned with MAFFT (version 7, http://mafft.cbrc.jp/alignment/server/) (Katoh et al. 2002) and then manually adjusted in Mesquite 2.75 (Maddison and Maddison 2011), following the criteria used by Borsch et al. (2003), Bremer et al. (2002), Kim and Jansen (1995) and Taberlet et al. (2007). The alignments for *trn*L-F and *ndhF* genes were separately optimised by first performing Neighbour Joining in PAUP* version 4.0b10 (Swofford 2003). The following parameters were used: Outgroup: *Barnadesia caryophylla*, Dset Distance = GTR, Rates = Gamma. The vertical order of accessions in the two alignments was then adjusted according to the NJ tree in order to maintain a phylogenetic continuum and to see if local rearrangements in the alignment of nucleotides were needed. Presumably homologous indel events (gaps) were coded as additional presence/absence characters. Regions left doubts about the homology of indels or could not be aligned were treated as in Bremer et al. (2002).

Phylogenetic trees at the subfamily level were then reconstructed for *ndhF* and *trnL-F* regions separately using Randomized Axelerated Maximum Likelihood (RAxML)-HPC2 run on XSEDE (Stamatakis 2014) from the Cyber-infrastructure for Phylogenetic Research (CIPRES) Science Gateway (V. 3.3, available at http://www.phylo.org/) (Miller et al. 2010) (**Figure S1 & S2**). Simultaneously, MrBayes 3.2.2 on XSEDE from CIPRES Science Gateway was also used to perform phylogenetic analyses (Ronquist et al. 2012), using the same alignment (**Figure S3 & S4**).

In order to estimate phylogenetic relationships at the generic level, we then subsampled our subfamily level alignments based on the generated trees (**Figure S1 - S4**) and trees from Wang et al. (2013). 79 *trnL-F* and 33 *ndhF* accessions were selected to represent *Lactuca* and related genera. *Leontodon saxatilis* is the nearest sister group to *Lactuca* and related genera and therefore was chosen as the outgroup (**Figure S1 - S4**). The subsampled sequences were re-aligned using MAFFT version 7. Indels were manually coded for *trnL-F* and *ndhF* genes following the Simple Indel Coding (SIC) method (Simmons and Ochoterena 2000) in Mesquite 2.75. The selected sequences were then concatenated using SequenceMatrix-Windows 1.7.8 (Vaidya et al. 2011).

The joined alignment, containing the two plastid DNA sequences, as well as the two separate gene alignments were used for further phylogenetic analyses. For the joined alignment, the dataset was analysed in three different ways for Bayesian Inference (BI): no partition, two partitions (*trnL-F* / *ndhF*) and three partitions (*trnL-F* / codon position1 + 2 of *ndhF* / codon position 3 of *ndhF*). The parameters for BI were as follows: outgroup *Leontodon saxatilis*; lset nst = mixed, rates = gamma; unlink statefreq = (all), revmat = (all), shape = (all), pinvar = (all); prset applyto = (all), ratepr = variable; mcmcp ngen = 50000000, relburnin = yes, burninfrac = 0.25, printfreq = 1000, samplefreq = 50000 nchains = 4 temp = 0.05; Report tree = brlens. Other parameters were default settings. For the single gene alignments, the dataset of *ndhF* gene was treated in two ways for BI: no partition and two partitions (codon position1 + 2 / codon position 3) and the alignment of *trnL-F* gene was not partitioned as it is not a coding sequence.

The Markov Chain output parameter files generated by MrBayes 3.2.2 were then used in Tracer v1.6 (available at http://tree.bio.ed.ac.uk/software/tracer/) to select the best partition for constructing phylogenetic trees by selecting the marginal density centred around the highest log likelihood (LnL). The chosen partition was then subjected to RAxML analysis using default settings. TreeGraph 2 was used to add Bootstrap (BS) and Posterior Probability (PP) values on one tree (Stover and Muller 2010).

Biogeographical, chromosomal and morphological data analyses

Biogeographical distributions were inferred from The Cichorieae Portal (Hand et al. 2009+) and Lebeda et al. (2004). We used RASP (Reconstruct Ancestral State in Phylogenies) to reconstruct ancestral biogeographical areas whereby distribution areas were delineated as A(Asia), B(Europe), C(Africa) and D(North America) (Yu et al. 2015). We did not delineate more detailed distributions due to the restriction of the number of biogeographical areas in RASP. We used 1000 trees inferred from BI analyses and the condensed Bayesian tree in RASP. The Bayesian Binary MCMC (BBM) (Experimental) method and the Fixed (JC) + Gamma model were used to reconstruct the biogeographical areas. Other settings were default.

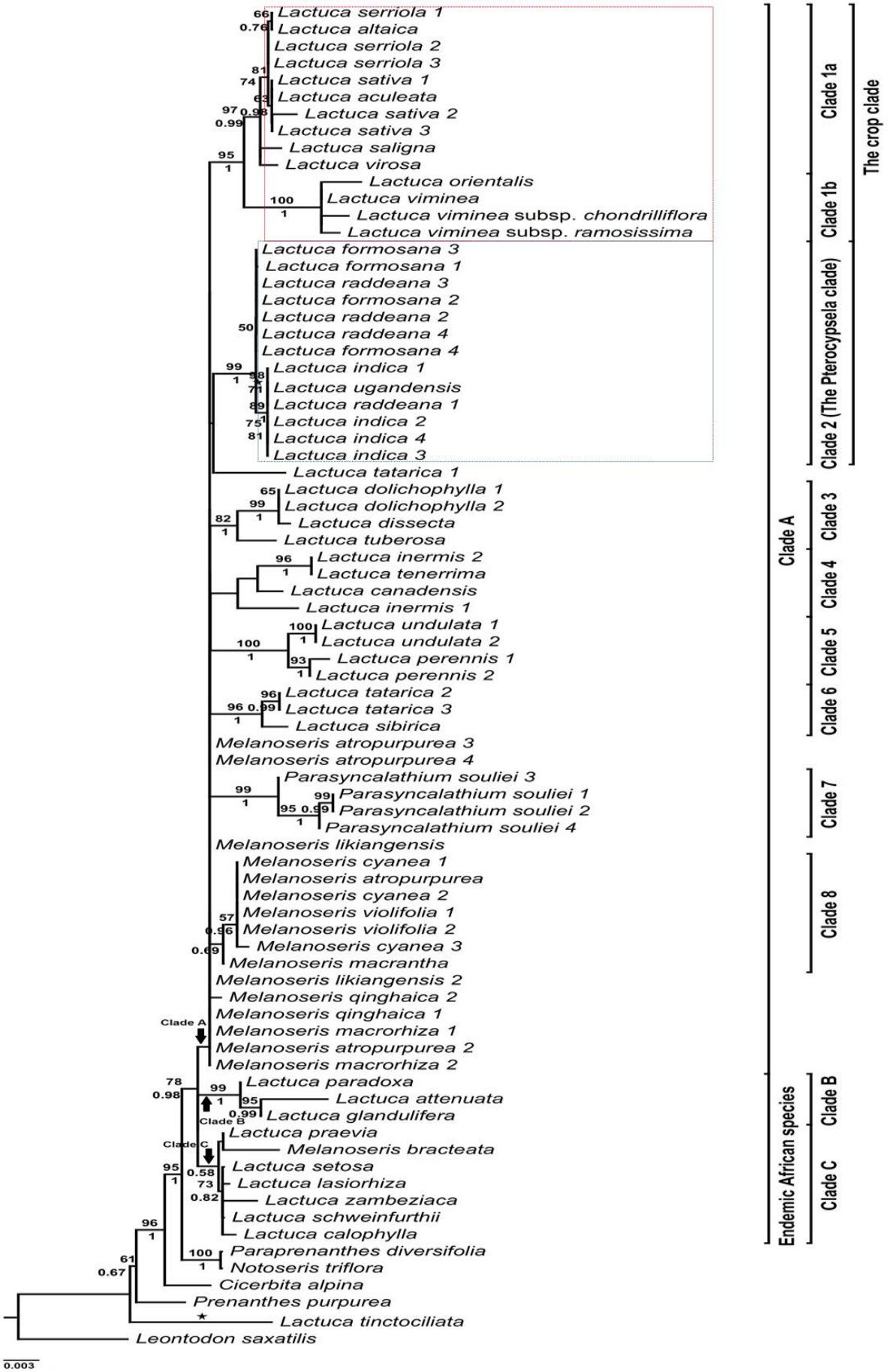
Chromosome numbers were scored according to Koopman et al. (1993), Matoba et al. (2007) and the Index to Plant Chromosome Numbers (IPCN) (Missouri Botanical Garden 2014). Selected morphological characters, such as floret number, achene winged or not and rib number were scored from The Cichorieae Portal (Hand et al. 2009+). We selected these characters because they are considered as important identification keys. Subsequently, we reconstructed the ancestral states for chromosomal and morphological characters over the same trees used for estimating the ancestral state of the biogeographical data in RASP. All the settings were the same.

Results

The *ndhF* and *trnL-F* sequences of 27 species were successfully sequenced by NGS, whereas the sequences of *L. praevia* C.D.Adams and *L. viminea* J.Presl & C.Presl subsp. *ramosissima* (All.) Malag. were failed for NGS and obtained using Sanger sequencing. In addition, the sequencing of *L. imbricata* Hiern, *L. orientalis* Boiss. and *L. schulzeana* Büttner was neither successful by NGS or Sanger. The *trnL-F* region had 863 (including indels)/853 characters in the alignment. Of the total 863/853 characters, 65(7.5%)/58(6.8%) were parsimony informative sites (**Table 2**). The alignment of *ndhF* gene contained 2251 (including indels) /2250 characters and 71(3.2%)/70(3.1%) of them were informative sites (**Table 2**). The total number of characters in the concatenated alignment was the sum of *trnL-F* and *ndhF* and 136(4.4%)/128(4.1%) of them were informative sites. The phylogenetic trees of 247 *ndhF* and 331 *trnL-F* gene sequences from different subfamilies using RAxML and BI analyses are shown in **Figure S1 - S4.** The no partition model for the concatenated dataset performed better than the partition models, as its marginal density was centred around a higher log likelihood (LnL), and therefore was chosen for further analyses. One 'best ML tree' for the concatenated sequences was inferred automatically from the RAxML analysis, which is generally congruent in topology with the BI 50% majority rule consensus tree. We present the RAxML phylogram topology combined with BS and PP values (**Figure 1**). The phylogenetic trees for single gene alignments are shown in **Figure S5 and S6**. We also reconstructed ancestral states for biogeographical, chromosomal and morphological characters over the condensed Bayesian trees of the concatenated sequences (**Figure S7 - S11**).

The phylogenetic analyses showed that *L. tinctociliata* I.M.Johnst. is outside the *Lactuca* clade and the sister group to all *Lactuca* and *Melanoseris* species, *Notoseris triflora* (Hemsl.) C.Shih, *Paraprenanthes diversifolia* (Vaniot) N.Kilian, *Cicerbita alpina* Wallr. and *Prenanthes purpurea* (Vaniot) N.Kilian (**Figure 1**, name indicated with a star). A *Lactuca* clade (BS = 78, PP = 0.98) divides into three clades, Clade A, B and C. We will describe the clades in the following sections.

Clade 1 (BS = 95, PP = 1) includes the lettuce crop and closely related wild lettuce species. It contains two subclades. Clade 1a (BS = 97, PP = 0.99) consists of the domesticated lettuce *L. sativa* and its closest relatives *L. serriola, L. altaica, L. aculeata, L. saligna* and *L. virosa*.


Lactuca serriola 1
Lactuca altaica
Lactuca serriola 2
Lactuca serriola 3
Lactuca sativa 1
Lactuca aculeata
Lactuca sativa 2
Lactuca sativa 3
Lactuca saligna
Lactuca virosa
Lactuca orientalis
Lactuca viminea
Lactuca viminea subsp. chondrilliflora
Lactuca viminea subsp. ramosissima
Lactuca formosana 3
Lactuca formosana 1
Lactuca raddeana 3
Lactuca formosana 2
Lactuca raddeana 2
Lactuca raddeana 4
Lactuca formosana 4
Lactuca indica 1
Lactuca ugandensis
Lactuca raddeana 1
Lactuca indica 2
Lactuca indica 4
Lactuca indica 3
Lactuca tatarica 1
Lactuca dolichophylla 1
Lactuca dolichophylla 2
Lactuca dissecta
Lactuca tuberosa
Lactuca inermis 2
Lactuca tenerrima
Lactuca canadensis
Lactuca inermis 1
Lactuca undulata 1
Lactuca undulata 2
Lactuca perennis 1
Lactuca perennis 2
Lactuca tatarica 2
Lactuca tatarica 3
Lactuca sibirica
Melanoseris atropurpurea 3
Melanoseris atropurpurea 4
Parasyncalathium souliei 3
Parasyncalathium souliei 1
Parasyncalathium souliei 2
Parasyncalathium souliei 4
Melanoseris likiangensis
Melanoseris cyanea 1
Melanoseris atropurpurea
Melanoseris cyanea 2
Melanoseris violifolia 1
Melanoseris violifolia 2
Melanoseris cyanea 3
Melanoseris macrantha
Melanoseris likiangensis 2
Melanoseris qinghaica 2
Melanoseris qinghaica 1
Melanoseris macrorhiza 1
Melanoseris atropurpurea 2
Melanoseris macrorhiza 2
Lactuca paradoxa
Lactuca attenuata
Lactuca glandulifera
Lactuca praevia
Melanoseris bracteata
Lactuca setosa
Lactuca lasiorhiza
Lactuca zambeziaca
Lactuca schweinfurthii
Lactuca calophylla
Paraprenanthes diversifolia
Notoseris triflora
Cicerbita alpina
Prenanthes purpurea
Lactuca tinctociliata
Leontodon saxatilis
Clade A
Clade B
Clade C
0.003
Clade 1a
Clade 1b
Clade 2 (The Pterocypsela clade)
The crop clade
Clade 3
Clade 4
Clade 5
Clade 6
Clade 7
Clade 8
Clade A
Clade B
Clade C
Endemic African species

◄ **Figure 1** RAxML phylogram ('best ML tree') of the concatenated sequences of *ndhF* gene and *trnL-F* gene used in this study; Bootstrap (BS > 50) support values are given above the branches and Posterior Probability (PP > 0.5) support values are below; the names of Chinese taxa are referred to Wang et al. (2013); star *L. tinctociliata* was mis-identified and it could be *Launaea cornuta*; *L. ugandensis* should be *Lactuca* sp.

One *L. serriola* accession is the sister group to *L. altaica* (BS = 66, PP = 0.76). *L. aculeata* and *L. sativa* are grouped together (BS = 63, PP = 0.98). *L. saligna* and *L. virosa* are the sister groups of *L. serriola*, *L. altaica*, *L. aculeata* and *L. sativa*. Clade 1b (BS = 100, PP = 1) comprises *L. orientalis*, *L. viminea* J.Presl & C.Presl, *L. viminea* J.Presl & C.Presl subsp. *chondrilliflora* (Boreau) Malag. and *L. viminea* subsp. *ramosissima*. Clade 1 (PP = 1) comprises widely spread *Lactuca* species from Asia, Europe and Africa (**Figure S7**). The species in Clade 1 have a chromosome number of eighteen (2n = 18) except *L. orientalis* (2n = 18 or 36) (**Figure S8)**. Most species in Clade 1a have a floret number between 6 -15 (20) or even more than 20 florets (**Figure S9)**. Other species in Clade 1b have less than 6 florets (**Figure S9)**. The achenes of most species in Clade 1 are not winged except *L. virosa* (**Figure S10)**. Most species in Clade 1 have a rib number between 3 and 9 (**Figure S11**).

Clade 2 (BS = 99, PP = 1) comprises of ex-*Pterocypsela* species, including *L. indica* L., *L. raddeana* Maximowicz, *L. formosana* Maximowicz and *L. ugandensis* C.Jeffrey (not ex-*Pterocypsela* species). Four *L. indica* accessions, one *L. raddeana* accession and *L. ugandensis* are in one subclade (BS = 89, PP = 1) whereas the other three *L. raddeana* accessions and four *L. formosana* accessions are in one clade (BS = 50). In addition, one *L. tatarica* accession is the sister group to Clade 2, though the BS support is very low (BS < 50). This clade contains Asian species and one African species *L. ugandensis* clade (PP = 1) (**Figure S7**). *Lactuca* species in Clade 2 have eighteen chromosomes (2n = 18) but this information for *L. ugandensis* is missing (**Figure S8)**. They usually have a floret number between 6 -15 (sometimes more than 20) (**Figure S9)**. Most species in Clade 2 (excluding *L. ugandensis*) have winged achenes (**Figure S10)** and a rib number between 1 and 7 (**Figure S11**).

Clade 3 (BS = 82, PP = 1) consists of *L. dolichophylla* Kitamura, *L. dissecta* D. Don and *L. tuberosa* Jacq. Clade 4 (lacking support) is composed of *L. tenerrima* Pourr., *L. inermis* and *L. canadensis*. *L. inermis* 1 from Ghana is the sister group of *L. tenerrima*, *L. canadensis* and *L. inermis* 2 from Togo. Clade 5 (BS = 100, PP = 1) includes *L. undulata* Ledebour and *L. perennis* L. Clade 6 (BS = 96, PP = 1) contains two *L. tatarica* accessions and *L. sibirica*. Clade 3 and 4 (PP = 1) include species from Asia and widespread species (**Figure S7**). Most species in Clade 5 and 6 are from Asia, North America or widespread species (**Figure S7**). The *Lactuca* species in Clade 3 have sixteen chromosomes (2n = 16) (**Figure S8)**. *Lactuca* species in Clade 5 and 6 have a chromosome number of eighteen (2n = 18). *L. tenerrima* and *L. inermis* in Clade 4 have sixteen chromosomes (2n = 16) while *L. canadensis* has thirty-four chromosomes (2n = 34) (**Figure S8)**. Most species in Clade 3 - 6 have a floret number usually between 6 -15 (sometimes more than 20) (**Figure S9)** and non- winged achenes (excluding *L.*

canadensis and *L. tuberosa* (**Figure S10**). Most species in Clade 3 and 4 have a rib number between 3 (1) and 7. Species in Clade 5 and 6 have 1 - 3 ribs (**Figure S11**).

Clade 7 contains four *Parasyncalathium souliei* (Franch.) J.W.Zhang, Boufford & H.Sun accessions with a good support value (BS = 99, PP = 1) (**Figure 1**). **Clade** 8 lacks support (BS<50, PP = 0.69) but may become stronger after adding more taxonomic sampling. It includes *Melanoseris cyanea* Edgew, *M. violifolia* (Decne.) N.Kilian, *M. atropurpurea* (Franch.) N.Kilian & Ze H.Wang and *M. macrantha* (C.B.Clarke) N.Kilian & J.W.Zhang. Other *Melanoseris* species, *M. atropurpurea*, *M. qinghaica* (S.W.Liu & T.N.Ho) N.Kilian & Ze H.Wang, *M. macrorhiza* (Royle) N.Kilian, *M. likiangensis* (Franch.) N.Kilian & Ze H.Wang are in a huge polytomy. *Melanoseris* and *Parasyncalathium* species are from Asia or widespread species (**Figure S7**). They have sixteen chromosomes (2n = 16) (**Figure S8**). *Melanoseris* species have a floret number between 6 -15 (sometimes more than 20) while *Parasyncalathium souliei* has a floret number less than 6 (**Figure S9**). *Melanoseris* and *Parasyncalathium* species do not have winged achenes (**Figure S10)**. The rib number of most *Melanoseris* species is unknown (**Figure S11**). *Parasyncalathium souliei* in Clade 8 has 1 - 3 ribs.

Clade B (BS = 99, PP = 1) contains three scandent African species, *L. glandulifera* Hook.f., *L. attenuata* Stebbins and their herbal sister group *L. paradoxa* (**Figure S7)**. Clade C (PP = 0.58) includes the African species *L. lasiorhiza* (O.Hoffm.) C.Jeffrey, *L. schweinfurthii* Oliv. & Hiern, *L. calophylla* C.Jeffrey, *L. zambeziaca* C.Jeffrey, *L. setosa* Stebbins ex C.Jeffrey, *L. praevia* and *Melanoseris bracteata* (Hook.f. & Thomson ex C.B.Clarke) N.Kilian. Chromosome number is only available for *L. attenuata* (2n = 32) and *L. glandulifera* (2n = 16) (**Figure S8)**. Species in Clade B and C have a floret number less than 6 (**Figure S9)** and they do not have winged achenes (**Figure S10)**. Most species in Clade B have a rib number between 3 and 7. Species in Clade C have 1 - 3 ribs (**Figure S11)**.

Discussion

Lettuce is an economically important crop and consequently most studies have mainly focused on *L. sativa* and closely related wild species (Koopman et al. 1998; Koopman et al. 1993; Koopman et al. 2001). Conversely, the entire *Lactuca* genus is poorly studied, especially for the two regions with the highest diversity, Asia (51 species) and Africa (43 species) (Lebeda et al. 2004). Recently, a publication focused on the Chinese centre of diversity, including 15 Asian *Lactuca* species (Wang et al. 2013). However, the African *Lactuca* center of diversity remains unstudied. We here present the first study focused on the phylogenetic relationships within *Lactuca* and related genera with extensive sampling of the African diversity centre, based on plastid genes. This is the first molecular phylogeny for 40% of the endemic African *Lactuca* species, especially for the scandent species since they were described and revised by Stebbins (1937b).

The mapping of biogeographical, chromosomal and morphological character states lend additional supports to the topologies of the RAxML trees. For biogeographical data, Clade B and Clade C only contain *Lactuca* species endemic to African continent, although other clades do not show distinctive pattern. The chromosome numbers (excluding the accessions with unknown chromosome number in Clade 8) supported the topology of the RAxML tree. *Lactuca* species in Clade 1, 2, 5 and 6 have a chromosome number of eighteen (2n = 18) except *L. orientalis* (2n = 18 or 36). Species in Clade 3, B, C and *Melanoseris* species have sixteen chromosomes (2n = 16). *L. tenerrima* and *L. inermis* in Clade 4 have sixteen chromosomes (2n = 16) while *L. canadensis* has thirty-four chromosomes (2n = 34). In Clade 9, *L. glandulifera* has sixteen chromosomes (2n = 16) while *L. attenuata* has thirty-two (2n = 32). The floret number also validated the topology of the RAxML tree. Most species in Clade 1a, 2 - 6 and C have a floret number usually between 6 -15 (sometimes more than 20). Other species in Clade 1b, 7, B and C have a floret number less than 6. For the state of achene, most species in the *Lactuca* clade do not have winged achenes. Only *L. virosa*, *L. canadensis*, *L. tuberosa* and species in Clade 2 (excluding *L. ugandensis*) have winged achenes. For rib number, most species in Clade 1, 4 and B have a rib number between 3 and 9. Species in Clade C, 5, 6 and Clade 8 have 1 - 3 ribs. Species in Clade 2 and 3 have a rib number between 1 and 7. The rib number of most *Melanoseris* species is unknown.

Monophyly of the subtribe Lactucinae

Our RAxML tree for concatenated sequences shows that *C. alpina*, *Faberia*, *P. purpurea* and *L. tinctociliata* should be excluded to maintain the monophyly of the subtribe Lactucinae (**Figure S1 - S4**). *L. tinctociliata* is placed outside Lactucinae and nested in Hyoseridinae (**Figure S1 - S4**). It is clustered with *Launaea sarmentosa* with a very high support (BS = 100, PP = 1) in the *trnL-F* tree and is sister group of *Sonchus oleraceus* in the *ndhF* trees (BS < 50, PP = 0.64) (**Figure S1 - S4**). This species was first published and described by I.M. Johnst in 1925 (Jeffrey 1966; Anonymous 1925). No detailed description or molecular data have been made available since then. According to I.M. Johnst, *L. tinctociliata* is very well characterized by its narrow firm purple leaf-margins which commonly bear purplish-tinged teeth and fleshy cilia, the capitula with about 12 yellow flowers, a very compressed achene, marginal, oblong-ovate or oblanceolate 5-6 mm long, thin beak > 1 mm long, about 12 ribs, bristle white pappus, 5 - 6 mm long (Anonymous 1925). From the image of the *L. tinctociliata* specimen used in this study, we can see (image available at http://medialib.naturalis.nl/file/id/WAG.1288514/format/large?width=800px&height=800px) that it has broader leaves than the type specimen (image available at http://plants.jstor.org/stable/10.5555/al.ap.specimen.gh00009514) and does not have purple leaf-margins. Although we could only compare the specimen images, the '*L. tinctociliata*' used in our study is clearly not *L. tinctociliata*. Based on our molecular data and the woody habit (typical of the species), the specimen is most likely *Launaea cornuta* (Hochst. ex Oliv. & Hiern) C.Jeffrey.

Wang et al. (2013) indicated that when *Faberia* and *P. purpurea* lineages are excluded, the subtribe Lactucinae is monophyletic. Moreover, they suggested that *C. alpina* should be disregarded while the other *Cicerbita* species are placed inside the Lactucinae. A narrow circumscription of *Prenanthes* L. was proposed making it a probably monospecific genus (Kilian and Gemeinholzer 2007; Kilian et al. 2009). Wang et al. (2013) transferred species from *Prenanthes* to *Notoseris* Shih and confirmed this narrow concept of *Prenanthes*. The BI tree of *ndhF*, including species from different subfamilies (**Figure S3**), shows that the genus *Tolpis* from the subtribe Cichoriinae is the sister group of the clade comprising *P. purpurea*, *C. alpina*, *N. triflora*, *Paraprenanthes diversifloria* and the genus *Lactuca* (PP = 0.54), but support for this pattern is lacking. The RAxML *ndhF* tree indicates *P. purpurea* is the sister group of *Tolpis* species (**Figure S1)**. In our *trnL-F* trees, *P. purpurea* is the sister group of *Ixeridium gracile*, a species from the subtribe Crepidinae (BS = 61, PP = 0.93) (**Figure S2 & S4**). Although all BS and PP values involved are low, these results would confirm the narrow concept of *Prenanthes* and indicate that *P. purpurea* probably belongs to the subtribe Cichoriinae or Crepidinae and is far away from the subtribe Lactucinae.

Our RAxML tree reveals that *Notoseris* and *Paraprenanthes* are the sister groups to *Lactuca* in the subtribe Lactucinae (**Figure 1**). When the genus *Notoseris* was first described, it comprised 12 species, with shared morphological characters such as capitula with 3-5 florets, beakless achene apices and 6 - 9 ribs on each side of achene (Shih 1987). Shih (1997) then reduced the number of species to 11. Wang et al. (2013) recently removed several species from *Notoseris* and transferred two scandent species from *Prenanthes* to *Notoseris*, based on ITS and plastid DNA sequences. *Paraprenanthes* was first proposed by C. C. Chang and formally established by Shih (1988a), who added new species and transferred some species from *Lactuca*, *Crepis* and *Mycelis* based on morphological characters, e.g. capitula with 6-23 cyanic florets, achenes with 5 main ribs and two rather similar secondary ribs in-between, and a single pappus (1988a). Shih and Kilian (2011) maintained the circumscription of *Paraprenanthes* but used a wider species concept and separated 3 species from the genus. Recently, Wang et al. (2013) revised the genus by reducing the species recognized by Shih & Kilian (2011) to 6 and adding 4 new species. Although the phylogenetic relationships among *Paraprenanthes* and *Notoseris*species remains unresolved based on *trnL-F* DNA sequence comparisons (**Figure S2 & S4**), our results indicate that *Notoseris* and *Paraprenanthes* are closely related to *Lactuca*.

Circumscription of *Lactuca* and its subgeneric classification

The phylogenetic tree for the concatenated sequences indicates that the *Lactuca* species, autochthonous to the African continent, are far away from the other *Lactuca* species. Meanwhile, the other *Lactuca* species (not endemic to Africa), *Melanoseris* and *Paracyncalathium* are nested within Clade A (lacking support) as part of the large polytomy (**Figure 1**).

The African *Lactuca* species (Clade B and C, 2n = 16, 32 or ?) The African species include *L. paradoxa*, *L. attenuata*, *L. glandulifera*, *L. lasiorhiza*, *L. schweinfurthii*, *L. calophylla*, *L. zambeziaca*, *L. setosa* and *L. praevia*. Of all of these species we present, as far as we know, the first molecular phylogeny since they were summarized and described by Jeffrey (1966). Jeffrey (1966) elaborated a total of 33 African *Lactuca* species but Lebeda et al. (2004) reported that this group contains at least 43 species and 75% of the group (31 in total) can be considered as endemic. In our sampling, only autochthonous African *Lactuca* species are included in these two clades with one exception - *M. bracteata*. The support between *L. praevia* and *M. bracteata* is very low), hence it is difficult to tell if *M. bracteata* does or does not belong to Clade C. Other species occuring in Africa but not endemic to the African continent, such as *L. inermis*, *L. tenerrima*, *L. saligna* and *L. virosa*, are distributed in other clades. This may indicate an independent evolution of the African endemic species. Based on their scandent or herbal habits, these endemic species can be divided into two groups: the scandent group and the herbal group. According to Stebbins (1937b), there were 7 scandent *Lactuca* species in Africa: *L. stipulata*, *L. elgonensis*, *L. paradoxa*, *L. attenuata*, *L. semibarbata*, *L. wildemaniana*, and *L. glandulifera*. Jeffrey (1966) combined the last two species as *L. glandulifera* and added *L. attenuatissima* to the scandent group. Our scandent samples include *L. paradoxa*, *L. attenuata* and *L. glandulifera*. These scandent species are not related to the two scandent species from *Notoseris*, which indicates two independent evolutions of the scandent habit in Lactucinae (**Figure S2 & S4**). These African species share some characters, such as capitula with less than 6 yellow florets (an exception from *L. lasiorhiza* with 10 - 14 florets) and 1 to 3 ribs on each side of achene. Chromosome number is only available for *L. attenuata* (2n = 32) and *L. glandulifera* (2n = 16) (Missouri Botanical Garden 2014). Wang et al. (2013) used the same dataset of *Melanoseris* species as in our study and showed that the genus *Melanoseris* is closely related to the genus *Lactuca*. In our results, *Melanoseris* and *Parasyncalathium* species are in Clade A and the African *Lactuca* species in Clade B and C are even further away from other *Lactuca* species in Clade A than *Melanoseris* and *Parasyncalathium* species. Our molecular, biogeographical, chromosomal and morphological data all show that the endemic African *Lactuca* species have a unique position and evolved independently. We suggest that the African species in Clade B and Clade C could be removed from *Lactuca* and treated as a new genus. However, further taxonomic, cytological and molecular studies are still needed to do an official taxonomic revision.

The Melanoseris species (Clade 7 and 8, 2n = 16 or ?) Clade 7 contains *Parasyncalathium souliei* accessions with a very high support value (BS = 99, PP = 1) (**Figure 1**). This implication is in line with Stebbins (1940) and Zhang et al. (2009; 2011). However, Wang et al. (2013) preferred to put this species in *Melanoseris* while Zhang et al. (2011) proposed that this species should be either put back in *Lactuca* or treated as a new genus. Clade 8 includes *M. cyanea*, *M. violifolia*, *M. atropurpurea* and *M. macrantha*. One *M. atropurpurea* accession is in this clade while other three *M. atropurpurea* accessions are in an

unresolved polytomy together with *M. macrorhiza*, *M. likiangensis* and *M. qinghaica*. The name *Melanoseris* was first proposed by Decaisne in 1843 for two species from the Himalayas, which are now treated as *M. lessertiana*. Edgeworth (1846) then added more Himalayan species to *Melanoseris*. Shih (1991) established two new genera from Sino-Himalayan region, *Chaetoseris* and *Stenoseris*, by transferring species from *Lactuca* and *Cicerbita*. *Chaetoseris* was distinguished from *Lactuca* and *Cicerbita* because of its achene corpus with broad and thickened lateral ribs and a pappus with an outer ring of minute hairs (Shih 1991, 1997). *Stenoseris* was established with 5 species and circumscribed by 3-5 flowered capitula and an achene with an outer ring of minute hairs (Shih 1991). Shih and Kilian (2011) revised this lineage and reused the name *Melanoseris* for the lineage based on their molecular data. They transferred species that were formerly placed in *Chaetoseris*, *Cicerbita*, *Lactuca*, *Mulgedium*, *Prenanthes* and the genus *Stenoseris* to *Melanoseris*. Furthermore, Wang et al. (2013), using nrITS1 and plastid genes, concluded that *Melanoseris* could be divided into three groups: *M. cyanea* group, *M. macrorhiza* group and *M. graciliflora* group. Although our results do not separate the *Melanoseris* lineage from *Lactuca* species, they reveal a close relationship between *Lactuca* and *Melanoseris*. Compared with previous molecular and morphological investigations, we still think *Melanoseris* and *Lactuca* are two separate but closely related genera (Shih and Kilian 2011; Wang et al. 2013).

We will now discuss the clades (1-6) that can be highlighted within *Lactuca*:

Clade 1 (The Crop Clade) (2n = 18 or 36) This clade comprises Clade 1a and 1b. Clade 1a contains the cultivated lettuce and can be referred to as *Lactuca* section *Lactuca* subsect. *Lactuca* (Lebeda et al. 2009). This clade includes *L. serriola*, *L. altaica*, *L. aculeata*, *L.virosa* and *L. saligna*. All the species in Clade 1a are interfertile or partly interfertile with *L. sativa* (Hartman et al. 2012; Thompson et al. 1941). Koopman et al. (1998) considered *L. serriola* and *L. altaica* to be conspecific based on their identical ITS-1 sequences and the results of crossing experiments. Our phylogenetic tree confirms his conclusion and also show that *L. aculeata* is closer to *L. sativa* than *L. serriola*. *L. sativa*, *L. serriola*, *L. altaica* and *L. aculeata* comprise the primary lettuce gene pool (Koopman et al. 1998). *L. virosa* and *L. saligna* are the sister groups to the species in the primary gene pool and form the secondary lettuce gene pool (Koopman et al. 1998). Crosses between *L. serriola* and *L. saligna*, and between *L. sativa* and *L. saligna* were shown to be partly fertile or self-fertile (Jeuken et al. 2001; Thompson et al. 1941; Zohary 1991). Chromosomal studies have demonstrated that *L. saligna* is potentially more closely related to *L. sativa* - *L. serriola* than *L. virosa* (Koopman et al. 1993; Matoba et al. 2007). Conversely, nrITS1 and AFLP fingerprints with moderate support indicated that *L. virosa* is closely-related to *L. sativa* - *L. serriola* (Koopman et al. 1998; Koopman et al. 2001). Although the cross between *L. virosa* and *L. sativa* often failed, it was still possible to obtain the cross and the hybrid was found to be self-sertile (Thompson et al. 1941; Whitaker and Thompson 1941; Zohary 1991). All the species in Clade 1a are widespread and share some characters, like a floret number > 6 (**Figure S7 - 11**).

Clade 1b includes *L. orientalis* and *L. viminea* and refers to section *Phaenixopus* (Lebeda et al. 2009). *L. orientalis* and *L. viminea* belonged to the genus *Scariola* but recently they were both treated as *Lactuca* species (Flann et al. 2010; Shih 1997; Shih and Kilian 2011; Wang et al. 2013). *L. orientalis* (2n = 18, 36) is a subshrub, which is very rare in *Lactuca*, all the other *Lactuca* species are herbs (Shih and Kilian 2011). It has whitish, rigid, intricately and divaricately branched stems, glaucous green leaves, solitary capitula with 4 or 5 pale yellow florets and a narrowly cylindrical involucre, and narrowly ellipsoid achenes with 5 - 7 ribs on either side (Shih and Kilian 2011). *L. viminea* subsp. *viminea*, *L. viminea* subsp. c*hondrilliflora* and *L. viminea* subsp. *ramosissima* (2n = 18) share many morphological characters although they differ from each other in certain characteristics. For example, *L. viminea* subsp. *chondrilliflora* has a beak length as long as ¼ - ½ of the achene body while *L. viminea* subsp. *viminea* and *L. viminea* subsp. *ramosissima* have a beak length equal to the achene body. Furthermore, *L. viminea* subsp. *viminea* branches only in the upper part of the stem whereas *L. viminea* subsp. *ramosissima* branches mostly in the basal part (Fer áková & M ájovský 1977). According to Koopman et al. (1998), *L. viminea* from the section *Phaenixopus* belongs to the tertiary lettuce gene pool, which also contains *L. quercina* from section *Lactucopsis*, *L. sibirica* and *L. tatarica* from section *Mulgedium*. In our phylogentic inferences, *L. quercina* was not included and *L. sibirica* and *L. tatarica* form a seperate Clade 4. Wang et al. (2013) using their nrITS1 sequences indicated a tertiary gene pool similar to Koopman's but showed that *L. sibirica* and *L. tatarica* form a well-supported seperate clade using their plastid gene sequences. Hybridization experiments showed that *L. viminea* is partly fertile with *L. virosa* (Groenwold 1983) and *L. tatarica* could be somatically hybridized with *L. sativa* (Chupeau et al. 1994; Maisonneuve et al. 1995). As the chance of generating fertile seeds from hybrids of *L. tatarica* and *L. sativa* is very low in nature (Chupeau et al. 1994; Maisonneuve et al. 1995), we consider *L. orientalis* and the three *L. viminea* subspecies as the tertiary gene pool and keep *L. sibirica* and *L. tatarica* beyond the tertiary gene pool.

The lettuce gene pool can provide rich genetic resources for improving lettuce growth, e.g. with respect to resistance to abiotic and biotic stresses. For example, *L. serriola* from the primary gene pool has been proven to possess interesting alleles for acquiring water and fertilizer in soil, increasing germination and seed longevity (Argyris et al. 2005; Johnson et al. 2000; Schwember and Bradford 2010). *L. aculeata* from the primary gene pool, *L. saligna* and *L. virosa* from the secondary gene pool, *L. viminea* from the tertiary gene pool, and *L. tatarica*, *L. biennis*, *L. canadensis*, *L. homblei*, *L. indica* and *L. perennis* beyond the lettuce gene pool all showed high resistance to downy mildew (Jeuken MJ 2008; van Treuren et al. 2011). These species may provide rich genetic resources for the crop lettuce. *L. orientalis*, belonging to the tertiary gene pool, could be a potential resource to improve the growth, development and resistance to diseases of the lettuce crop as well.

Clade 2 (The *Pterocypsela* Clade) (2n = 18 or ?) This clade comprises species mostly distributed in Asia: *L. indica* (2n = 18, although Lebeda et al. (2004) indicate it is also in

Africa based on floras), *L. raddeana* (2n = 18) and *L formosana* (2n = 18) (Hand et al. 2009+; Jeffrey 1966). The only exception is *L. ugandensis* (2n = ?) from Africa. The first three species belonged to the genus *Pterocypsela*, which was established by Shih (Shih 1988b) with type species *Pterocypsela indica* (L.) Shih. They have some shared characters, such as involucral bracts in 4 - 5 rows, capitula with 9 - 25 florets, broadly winged achenes with 1 or 3(5) prominent ribs on either side of the achene body and double pappus (Shih 1988b, 1997). Shih and Kilian (2011) transferred these three *Pterocypsela* species to *Lactuca*. Although *L. ugandensis* is grouped together with these ex-*Pterocypsela* species, it is depicted without winged achene (Jeffrey 1966; Jeffrey and Beentje 2000). This *L. ugandensis* specimen could be mis-identified. Therefore we treat it as *Lactuca* sp. Clade 2 confirms the nrITS-1 and plastid gene trees of Wang et al. (2013) and is also comparable to section *Tuberosae* (Lebeda et al. 2009; Lebeda et al. 2007). In addition, *L. indica* (Indian lettuce) has been cultivated for its edible leaves (Kadereit and Jeffrey 2007). Somatic hybridizations between *L. sativa* and *L. indica* have shown that a viable callus can be generated but it cannot produce a viable plant (Mizutani et al. 1989). Moreover, *L. indica* is resistant to downy mildew (van Treuren et al. 2011). Thus, *L. indica* could be a useful genetic resource for lettuce breeding.

Clade 3 (2n = 16) This clade is composed of *L. dolichophylla*, *L. dissecta* and *L. tuberosa* (BS = 82, PP = 1). The support value between *L. dolichophylla* and *L. dissecta* (BS = 99, PP = 1) is even higher. These three species all have a chromosome number of 16 (Shih and Kilian 2011; Vogt and Aparicio 1999). *L. dolichophylla* and *L. dissecta* have some shared characters such as capitula with 6 - 15(20) blue florets and 3 - 5 ribs on either side of the achene while *L. tuberosa* has tuberous roots and broadly winged achenes (Hand et al. 2009+; Shih and Kilian 2011). *L. dolichophylla* and *L. dissecta* are distributed in Asia, mainly in South Asia and East Asia, whereas *L. tuberosa* occurs in Asia and Europe (Geltman 2003; Hand et al. 2009+).

Clade 4 (2n = 34, 16) This clade includes *L. canadensis* (2n = 34) originating from North America, *L. tenerrima* (2n = 16) and *L. inermis* (2n = 16). *L. inermis* 1 (collected in Ghana) is the sister group to *L. canadensis*, *L. tenerrima* and *L. inermis* 2 (collected in Togo) while *L. tenerrima* and *L. inermis* 2 is close to each other (BS = 96, PP = 1) (**Figure 1**). This could be the result of mis-identification of any of the *L. inermis* accessions or not enough evidence to distinguish these species. The American *Lactuca* group includes 12 species, 7 of them are endemic with 34 chromosomes (2n = 34) and different relative DNA content (Babcock et al. 1937; Doležalová et al. 2002; Lebeda and Astley 1999). *L. tenerrima* and *L. inermis* (treated as *L. capensis* before) have been shown to cluster together due to their low DNA content while *L. canadensis* is far away from them as a result of high DNA content (Doležalová et al. 2003). The crosses between *L. canadensis* and *L. tatarica* (2n = 18), and between *L. canadensis* and *L. raddeana* (2n = 18) can generate self-sterile hybrid plants (Thompson et al. 1941). Other North American *Lactuca* species, *L. graminifolia* (2n = 34), *L. floridana* (2n = 34) and *L. spicata* (2n = 34) could be crossed with *L. indica*, *L. laciniata* (now treated as *L. indica*), *L. raddeana*, and *L. tatarica* and produce self-sterile or partly fertile hybrid plants

(Thompson et al. 1941; Wang et al. 2013). In addition, *L. canadensis*, *L. raddeana* and *L. indica* share a distinctive character, broadly winged achene, from other *Lactuca* species although their beak length are clearly different. The North American *Lactuca* species are supposed to have an amphidiploid origin and arose by subsequent crossings, doubling of chromosomes and hybrid stabilization. Their chromosome complement can be represented by the formula AABB (A = 8, B = 9) (Feráková and Májovský 1977). Our phylogenetic inferences and all these experimental hybridizations support the assumption that the North American *Lactuca* species could have a possible origin from the hybridization between *Lactuca* species with a haploid chromosome number of 8 (e.g. *L. tenerrima*) and 9 (e.g. *L. tatarica*, *L. raddeana* and *L. indica*).

Clade 5 (2n = 18) This clade comprises *L. undulata* from the section *Micranthae* and *L. perennis* from the section *Lactuca* subsect. *Cyanicae* (Lebeda et al. 2009; Lebeda et al. 2007). *L. undulata* shares characters with *L. perennis*, for example, 1 - 3 ribs per side of achene and beak as long as achene body (Feráková & Májovský 1977; Shih 1997). This close relationship between *L. undulata* and *L. perennis* is supported by Wang et al. (2013). According to Lebeda et al. (2007), species in the section *Micranthae* have a chromosome number of 16, which is not the case for *L. undulata*. Therefore, we suggest placing *L. undulata* into the section *Lactuca* subsect. *Cyanicae*.

Clade 6 (2n = 18) This clade contains *L. tatarica* and *L. sibirica* from Asia. These species are considered to belong to the section *Mulgedium* (Lebeda et al. 2009; Lebeda et al. 2007). Shih (1988b) revised the concept of genus *Mulgedium* (including *L. tatarica*) and considered *Lagedium* Soják (only including *L. sibirica*) as a monospecific genus, based on the absence of a true beaked achene and a weakly compressed achene body. But Shih's concept of *Mulgedium* and *Lagedium* is not accepted by most taxonomists. Shih and Kilian (2011) revised these two genera and transferred these species into *Lactuca*. *L. sibirica* is fully fertile with *L. tatarica*, indicating a close relationship between these two species (Koopman et al. 2001). However, another European *L. tatarica* 1 is the sister group to Clade 2 (**Figure 1**). This accession is the sister group to Clade 2 in the *ndhF* tree (**Figure S5**) and the sister group to the whole *Lactuca* clade in the *trnL-F* tree (**Figure S6**). *L. indica* in Clade 2 can be crossed with *L. tatarica*, although producing self-sterile seeds (van Treuren et al. 2011). The conflicting positions of *L. tatarica* accessions could be the consequence of hybridization. More samples and evidence are needed to solve the problem.

Conclusions

This work presents the first molecular phylogeny of *Lactuca* with representatives of African species and includes the most extensive sampling of *Lactuca* species analyzed to date. Based on the results of the phylogenetic trees, we draw the following conclusions:

1. The genus *Lactuca* contains two well-distinguished clades: the crop clade and the *Pterocypsela* clade. Other North American, Asian and widespread species either form small clades or are mixed with the *Melanoseris* species. However, we still think *Melanoseris* and *Lactuca* are two separate but closely related genera based on previous studies. The newly identified African endemic species could be treated as a new genus, though more evidence is still needed.
2. We confirm the primary and secondary lettuce gene pool and modify the tertiary gene pool concept: adding *L. orientalis* and three *L. viminea* subspecies to the tertiary gene pool while excluding *L. sibirica* and *L. tatarica*.
3. *L. indica*, *L. orientalis* and *L. viminea* could be useful genetic resources for lettuce breeding.
4. *L. undulata* should be transferred from section *Micranthae* to the section *Lactuca* subsect. *Cyanicae* based on our molecular data and its chromosome number.
5. There are at least two independent origins of the scandent habit in Lactucinae.

Although the sampling used in this study only covers 34% of the total known *Lactuca* species, we provide the most extensive molecular sampling for *Lactuca* species to date. Until now, most species in *Lactuca* have never been revised or sequenced since they were published. In the future, we will sample more species and use whole chloroplast genome data to resolve the polytomy in *Lactuca*.

Acknowledgements

Thanks to China Scholarship Council (http://en.csc.edu.cn/) for providing funding to Z.W. The contribution from SYNTHESYS Joint Research Activities 4 (JRA4: Plants/fungi herbarium DNA) were supported by European Union-funded integrated Activities grant awarded to dr. F.T.B. and dr. W.G. Berendsohn. We appreciate dr. J.J. Wieringa, dr. MJW (Marieke) Jeuken and dr. N.K. for providing plant materials for DNA extraction. Thanks to the anonymous reviewer for identifying herbarium specimen in this study.

References

Anonymous (1925) Contributions from the Gray Herbarium of Harvard University. The Gray Herbarium of Harvard University, Cambridge, Mass, U.S.A.

Argyris J, Truco MJ, Ochoa O, Knapp SJ, Still DW, Lenssen GM, Schut JW, Michelmore RW, Bradford KJ (2005) Quantitative trait loci associated with seed and seedling traits in *Lactuca*. Theor Appl Genet 111:1365-1376

Babcock EB, Stebhins GL, Jenkins JA (1937) Chromosomes and phylogeny in some genera of the Crepidinae. Cytologia, Tokyo, Fujii Jub:188-210

Bakker FT, Lei D, Yu J, Mohammadin S, Wei Z, van de Kerke S, Gravendeel B, Nieuwenhuis M, Staats M, Alquezar-Planas DE, Holmer R (2015) Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline. Biol J Linn Soc doi: 10.1111/bij.12642

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology : a journal of computational molecular cell biology 19:455-477

Bentham G (1873) Compositae. In: Bentham G, Hooker JD (eds) Genera Plantarum. Reeve & Company, London, pp 163-533

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics 30(15): 2114-2120

Borsch T, Hilu KW, Quandt D, Wilde V, Neinhuis C, Barthlott W (2003) Noncoding plastid *trnT-trnF* sequences reveal a well resolved phylogeny of basal angiosperms. J Evol Biol 16:558-576

Bremer B, Bremer K, Heidari N, Erixon P, Olmstead RG, Anderberg AA, Källersjö M, Barkhordarian E (2002) Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. Molecular Phylogenetics and Evolution 24:274-301

Chupeau M-C, Maisonneuve B, Bellec Y, Chupeau Y (1994) A Lactuca Universal Hybridizer, and its use in creation of fertile interspecific somatic hybrids. Mol Gen Genet 245:139-145

Clark S, Egan R, Frazier P, Wang Z (2013) ALE: a generic assembly likelihood evaluation framework for assessing the accuracy of genome and metagenome assemblies. Bioinformatics 29:435-443

de Vries IM (1997) Origin and domestication of *Lactuca sativa* L. Genet Resour Crop Evol 44: 165-174

Doležalová I, Lebeda A, Janeček J, Číhalíková J, Křístková E, Vránová O (2002) Variation in chromosome numbers and nuclear DNA contents in genetic resources of *Lactuca* L. species (Asteraceae). Genet Resour Crop Evol 49:383-395

Doležalová I, Lebeda A, Dziechciarková M, Křístková E, Astley D, van De Wiel CCM (2003) Relationships among morphological characters, isozymes polymorphism and DNA variability - the impact on *Lactuca* germplasm taxonomy. Czech J Genet Plant Breed 39:59-67

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phyt Bull 19:11-15

Edgeworth M (1846) Descriptions of some unpublished species of plants from North-Western India. Trans Linn Soc London 20:23-91

Ferákova V, Majovský J (1977) The genus *Lactuca* L. in Europe. Bratislava: Univerzita Komenského

Flann C, Greuter W, Hind DJN (2010) Cassini's Compositae genera: A nomenclatural and taxonomic assessment. Taxon 59:1206-1244

Geltman DV (2003) Magnoliophyta (=Angiospermae) Magnoliopsida (=Dicotyledones). In: Tzvelev NN (ed) Flora of Russia The European part and bordering regions. A.A. Balkema, a member of Swts and Zeitlinger pp 195-196

Groenwold R (1983) Onderzoek van de relatie tussen *Lactuca* en *Bremia Lactucae*. Verslag van een voorlichtingsbijeenkomst voor slaveredelaars (vervolg). Zaadbelangen 37:132

Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - a baiting and iterative mapping approach. Nucleic Acids Res 41:e129

Hand R, Kilian N, Raab-Straube Ev (2009+) International Cichorieae Network: Cichorieae Portal. http://wp6-cichorieaee-taxonomyeu/portal/. Accessed 17 February 2015

Harlan JR (1986) Lettuce and the sycomore: sex and romance in ancient Egypt. Economic Botany 40:4-15

Hartman Y, Hooftman DA, Uwimana B, Schranz ME, van de Wiel CC, Smulders MJ, Visser RG, Michelmore RW, van Tienderen PH (2014) Abiotic stress QTL in lettuce crop-wild hybrids: comparing greenhouse and field experiments. Ecology and evolution 4:2395-2409

Hartman Y, Hooftman DAP, Eric Schranz M, van Tienderen PH (2012) QTL analysis reveals the genetic architecture of domestication traits in Crisphead lettuce. Genet Resour Crop Evol 60:1487-1500

Hoffmann O (1890-1894) Compositae. In: Hoffmann O (ed) Die natürlichen Pflanzenfamilien. Engelmann, Leipzig, pp 87-387

Jeffrey C (1966) Notes on Compositae: I. Kew Bull 18:427-486

Jeffrey C, Beentje HJ (2000) Cichorieae. In: Beentje HJS, S. A. L (ed) Flora of Tropical East Africa, Compositae 1, Rotterdam, pp 63-108

Jeuken M, van Wijk R, Peleman J, Lindhout P (2001) An integrated interspecific AFLP map of lettuce (*Lactuca*) based on two *L. sativa* × *L. saligna* F_2 populations. Theor Appl Genet 103:638-647

Jeuken MJ, Pelgrom K, Stam P, Lindhout P (2008) Efficient QTL detection for nonhost resistance in wild lettuce: backcross inbred lines versus F_2 population. Theor Appl Genet 116:845-857

Johnson WC, Jackson LE, Ochoa O, van Wijk R, Peleman J, St. Clair DA, Michelmore RW (2000) Lettuce, a shallow-rooted crop, and *Lactuca serriola*, its wild progenitor, differ at QTL determining root architecture and deep soil water exploitation. Theor Appl Genet 101:1066-1073

Judd WS, Campbell C, Kellogg EA, Stevens PF, Donoghue MJ (2007) Plant Systematics: A Phylogenetic Approach. Sinauer Assoc., Sunderland

Kadereit JW, Jeffrey C (2007) Flowering Plants. Eudicots: Asterales. In: Kubitzki K (ed) The families and genera of vascular plants, pp 187-189

Karis PO, Eldenäs P, Källersjö M (2001) New Evidence for the Systematic Position of Gundelia L. with Notes on Delimitation of Arctoteae (Asteraceae). Taxon 50:105-114

Katoh K, Misawa K, Kuma k-i, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059-3066

Kerbiriou PJ, Stomph TJ, Putten PEL, Lammerts Van Bueren, ET, Struik PC (2013b) Shoot growth, root growth and resource capture under limiting water and N supply for two cultivars of lettuce (*Lactuca sativa* L.). Plant Soil 371:281–297

Kesseli R, Ochoa O, Michelmore R (1991) Variation at RFLP loci in *Lactuca* spp. and origin of cultivated lettuce (*L. sativa*). Genome 34:430-436

Kilian N, Gemeinholzer B (2007) Studies in the Compositae of the Arabian Peninsula and Socotra: 7. *Erythroseris*, a new genus and the previously unknown sister group of *Cichorium* (*Cichorieae* Subtribe *Cichoriinae*). Willdenowia 37: 283-296

Kilian N, Gemeinholzer B, Lack HW (2009) Tribe Cichorieae. In: Funk VA SA, Stuessy T and Bayer R (ed) Systematics, evolution and biogeography of the Compositae. IAPT, Vienna, pp 343-383

Kim K-J, Jansen RK (1995) *ndhF* sequence evolution and the major clades in the sunflower family. Proc Natl Acad Sci USA 92:10379-10383

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

Koopman WJM, Jong JHd, Vries IMd (1993) Chromosome banding patterns in Lettuce species (*Lactuca* sect. *Lactuca*, Compositae). Plant Syst Evol 185:249-257

Koopman WJM, Zevenbergen MJ, van den Berg RG (2001) Species relationships in *Lactuca* s.l. (Lactuceae, Asteraceae) inferred from AFLP fingerprints. Am J of Bot 10:1881-1887

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature methods 9:357-359

Lebeda A, Astley D (1999) World genetic resources of *Lactuca* spp.: Their taxonomy and biodiversity. In: Lebeda A, Kristkova E (eds) Eucarpia leafy vegetables '99. Palacky Univ., Olomouc, Czech Republic, pp 81-94

Lebeda A, Doležalová I, Feráková V, Astley D (2004) Geographical distribution of wild *Lactuca* species (Asteraceae, Lactuceae). The Botanical Review 70:328-356

Lebeda A, Doležalová I, Křístková E, Kitner M, Petrželová I, Mieslerová B, Novotná A (2009) Wild *Lactuca* germplasm for lettuce breeding: current status, gaps and challenges. Euphytica 170:15-34

Lebeda A, Ryder E, Grube R, Doležalová I, Křístková E (2007) Lettuce (Asteraceae; *Lactuca* spp.) In: Singh R (ed) Genetic resources, chromosome engineering, and crop improvement series, Vol 3 - vegetable crops. CRC Press, Boca Raton, pp 377-472

Lindqvist K (1960) On the origin of cultivated lettuce. Hereditas 46:319-350

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Zhu X, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T-W, Wang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1:18

Maddison WP, Maddison DR (2015) Mesquite: a modular system for evolutionary analysis. Version 3.02 http://mesquiteproject.org

Maisonneuve B, Chupeau MC, Bellec Y, Chupeau Y (1995) Sexual and somatic hybridization in the genus *Lactuca*. Euphytica 85 85:281-285

Matoba H, Mizutani T, Nagano K, Hoshi Y, Uchiyama H (2007) Chromosomal study of lettuce and its allied species (*Lactuca* spp., Asteraceae) by means of karyotype analysis and fluorescence in situ hybridization. Hereditas 144:235-247

Mikel MA (2007) Genealogy of contemporary North American lettuce. Hortscience 42:489-493

Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, pp 1-8

Missouri Botanical Garden (2014) Index to Plant Chromosome Numbers (IPCN). http://www.tropicos.org/Project/IPCN. Accessed 07 Oct 2014

Mizutani T, Liu XJ, Tashiro Y, Miyazaki S, Shimasaki K (1989) Plant regeneration and cell fusion of protoplasts from lettuce cultivars and related wild species in Japan. Bulletin of the Faculty of Agriculture, Saga University 67:109-118

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst biol 61:539-542

Särkinen T, Staats M, Richardson JE, Cowan RS, Bakker FT (2012) How to open the treasure chest? Optimising DNA extraction from herbarium specimens. PloS one 7:1-7

Schneider JV, Bissiengou P, Amaral Mdo C, Tahir A, Fay MF, Thines M, Sosef MS, Zizka G, Chatrou LW (2014) Phylogenetics, ancestral state reconstruction, and a new infrafamilial classification of the pantropical Ochnaceae (Medusagynaceae, Ochnaceae s.str., Quiinaceae) based on five DNA regions. Mol Phylogenet Evol 78:199-214

Schwember AR, Bradford KJ (2010) Quantitative trait loci associated with longevity of lettuce seeds under conventional and controlled deterioration storage conditions. J Exp Bot 61:4423-4436

Shih C (1987) On the circumscription of the genus *Prenanthes* L. and *Notoseris* Shih: a new genus of Compositae from China. Acta Phytotaxon Sin 25:189-203

Shih C (1988a) Revision of *Lactuca* L. and two new genera of tribe Lactuceae (Compositae) on the mainland of Asia. Acta Phytotaxon Sin 26:418-428

Shih C (1988b) Revision of *Lactuca* L. and two new genera of tribe Lactuceae (Compositae) on the mainland of Asia. Acta Phytotaxon Sin 26:382-393

Shih C (1991) On circumscription of the genus *Cicerbita* Wall., and two new genera of Compositae from Sino-Himalayan region. Acta Phytotaxon Sin 29:394-417

Shih C (1997) Compositae (10), Cichorioideae, Lactuceae. Beijing: Science Press

Shih C, Kilian N (2011) Cichorieae [excl.*Tragopogon*, *Taraxacum*, *Hieracium* and *Pilosella*]. Science Press & St Louis: Missouri Botanical Garden, Beijing, pp 233-238

Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic analyses. Syst biol 49:369-381

Staats M, Cuenca A, Richardson JE, Vrielink-van Ginkel R, Petersen G, Seberg O, Bakker FT (2011) DNA damage in plant herbarium tissue. PloS one 6:e28448

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313

Stebbins GL (1937a) Critical notes on *Lactuca* and related genera. J Bot 75:12-18

Stebbins GL (1937b) The scandent species of *Prenanthes* and *Lactuca* in Africa. Bulletin du Jardin botanique de l'État a Bruxelles 14:333-352

Stebbins GL (1939) Notes on some Indian species of *Lactuca*. Indian Forester Bot 1:237-245

Stebbins GL (1940) Studies in the Cichorieae: *Dubyaea* and *Soroseris*, endemics of the Sino-Himalayan Region. Mem Torrey Bot Club 19:1-76

Stover BC, Muller KF (2010) TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC bioinformatics 11:7

Swofford D (2003) PAUP*: Phylogenetic analysis using parsimony (* and other methods). In: Associates S (ed), 4.0b 10 edn, Sunderland Massachusetts USA

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E (2007) Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. Nucleic Acids Res 35:e14

Thiers B (2011) Index Herbariorum: A global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium. http://sweetgum.nybg.org/ih/. Accessed 17 February 2015

Thompson RC, Whitaker TW, Kosar WF (1941) Interspecific genetic relationships in *Lactuca*. Journal of Agricultural Research 63:91-107

Tuisl G (1968) Der Verwandtschaftskreis der Gattung *Lactuca* L. im iranischen Hochland und seinen Randgebieten. Ann Nat Hist Mus Wien 72:587-638

Uwimana B, Smulders MJ, Hooftman DA, Hartman Y, van Tienderen PH, Jansen J, McHale LK, Michelmore RW, Visser RG, van de Wiel CC (2012b) Crop to wild introgression in lettuce: following the fate of crop genome segments in backcross populations. BMC Plant Biol 12:43

Vaidya G, Lohman DJ, Meier R (2011) SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27:171-180

van Treuren R, van der Arend AJM, Schut JW (2011) Distribution of downy mildew (*Bremia lactucae* Regel) resistances in a genebank collection of lettuce and its wild relatives. Plant Genet Resour 11:15-25

Vogt R, Aparicio A (1999) Chromosome numbers of plants collected during Iter Mediterraneum IV in Cyprus. Bocconea 11:117-168

Wang ZH, Peng H, Kilian N (2013) Molecular phylogeny of the *Lactuca* alliance (Cichorieae subtribe Lactucinae, Asteraceae) with focus on their Chinese centre of diversity detects potential events of reticulation and chloroplast capture. PloS one 8:e82692

Wei Z, Julkowska MM, Lalo ë J-O, Hartman Y, de Boer G-J, Michelmore RW, van Tienderen PH, Testerink C, Schranz ME (2014) A mixed-model QTL analysis for salt tolerance in seedlings of crop-wild hybrids of lettuce. Mol Breed 34:1389-1400

Whitaker TW, Thompson RC (1941) Cytological Studies in *Lactuca* Bulletin of the Torrey Botanical Club 68:388-394

Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252-3255

Yu Y et al. (2015) RASP (Reconstruct Ancestral State in Phylogenies): A tool for historical biogeography. Mol Phylogenet Evol 87: 46-49.

Zhang J-W, Nie Z-L, Sun H (2009) Cytological study on the genus *Syncalathium* (Asteraceae-Lactuceae), an endemic taxon to alpine scree of the Sino-Himalayas. Journal of Systematics and Evolution 47:226-230

Zhang J-W, Nie Z, Wen J, Sun H (2011) Molecular phylogeny and biogeography of three closely related genera, *Soroseris, Stebbinsia*, and *Syncalathium* (Asteraceae, Cichorieae), endemic to the Tibetan Plateau, SW China. Taxon 60:15-26

Zhang NW, Lindhout P, Niks RE, Jeuken MJW (2009) Genetic dissection of Lactuca saligna nonhost resistance to downy mildew at various lettuce developmental stages. Plant Pathol 58:923–932

Zohary D (1991) The wild genetic resources of cultivated lettuce (*Lactuca sativa* L.). Euphytica 53:31-35

Supplementary tables and figures

Table S 1 Information of GenBank accessions (*ndhF* gene)

Number	Accession number	Specimen name
1	EU385129.1	*Adenocaulon bicolor*
2	EU385130.1	*Ainsliaea apiculata*
3	AF384690.3	*Alepidocline annua*
4	FM208905.1	*Amblyocarpum inuloides*
5	AF218338.1	*Andryala integrifolia*
6	AF063070.1	*Anisopappus smutsii*
7	EU385132.1	*Aphyllocladus spartioides*
8	AB530944.1	*Arctium lappa*
9	DQ444742.1	*Arctotheca calendula*
10	DQ444743.1	*Arctotis acaulis*
11	DQ444744.1	*Arctotis arctotoides*
12	DQ444745.1	*Arctotis aspera*
13	DQ444746.1	*Arctotis breviscapa*
14	DQ444747.1	*Arctotis dregei*
15	EU385133.1	*Arctotis hirsuta*
16	DQ444748.1	*Arctotis perfoliata*
17	DQ444749.1	*Arctotis sp.*
18	L39425.1	*Arctotis stochadifolia*
19	DQ444751.1	*Arctotis venusta*
20	AF384802.3	*Arnica dealbata*
21	AF384695.3	*Arnica mollis*
22	L39455.1	*Athroisma gracile*
23	EU385134.1	*Atractylis cancellata*
24	L39413.1	*Atractylodes japonica*
25	L39394.1	*Barnadesia caryophylla*
26	DQ444738.1	*Berkheya carduoides*
27	EU385136.1	*Berkheya purpurea*
28	L39456.1	*Blepharispermum zanqebaricum*
29	FM208909.1	*Blumea densiflora*
30	KC589928.1	*Brachylaena discolor*
31	EU385138.1	*Brachylaena elliptica*
32	AY780818.1	*Caesulia axillaris*
33	L39439.1	*Calendula officinalis*
34	AY780819.1	*Calostephane marlothiana*
35	L39412.1	*Carlina vulgaris*
36	FM208917.1	*Carpesium cernuum*
37	L39417.1	*Carthamus tinctorius*
38	AF218349.1	*Catananche caerulea*
39	JQ922543.1	*Cavea tanguensis*
40	EU385140.1	*Centaurea melitensis*
41	HE862371.1	*Centipeda nidiformis*

42	EU385141.1	*Centratherum punctuatum*
43	L39419.1	*Cephanonoplus segetum*
44	AF384700.3	*Chaetymenia peduncularis*
45	EU385144.1	*Chimantaea humilis*
46	L39375.1	*Chromolaena sp*
47	EU334465.1	*Chrysanthemum coronarium*
48	FM208926.1	*Chrysophthalmum montanum*
49	GU817844.1	*Cichorium intybus*
50	GU817845.1	*Cirsium discolor*
51	L39418.1	*Cirsium texanum*
52	AF233823.1	*Cnicothamnus lorentzii*
53	EU385147.1	*Cnicothamnus lorentzii 2*
54	AF384723.3	*Coespeletia timotensis*
55	AY780852.1	*Coleocoma centaurea*
56	L39451.1	*Conyza sp*
57	EU385148.1	*Corymbium glabrum*
58	JF754840.1	*Cousinia microcarpa*
59	AY780821.1	*Cratystylis conocephala*
60	AF218345.1	*Crepis biennis*
61	AF218339.1	*Crepis pyrenaica*
62	AB530927.1	*Crepis rubra*
63	AF218348.1	*Crepis tectorum*
64	DQ444739.1	*Cuspidia cernua*
65	DQ444752.1	*Cymbonotus lawsonianus*
66	AB530947.1	*Cynara scolymus*
67	L39392.1	*Dasyphyllum argenteum*
68	EU385152.1	*Dicoma capensis*
69	AF303923.1	*Didelta carnosa*
70	AY780822.1	*Dielitzia tysonii*
71	EU385154.1	*Dinoseris salicifolia*
72	FJ813488.1	*Dipterocome pusilla*
73	AY780823.1	*Dittrichia viscosa*
74	AY780858.1	*Doellia bovei*
75	AY466429.1	*Dolomiaea tibetica*
76	AJ276493.1	*Doronicum columnae*
77	FM208928.1	*Duhaldea cuspidata*
78	DQ444753.1	*Dymondia margaretae*
79	AF384713.3	*Dyscritothamnus mirandae*
80	L39411.1	*Echinops exaltatus*
81	EU385158.1	*Echinops ritro*
82	AY780824.1	*Epaltes cunninghamii*
83	EU385159.1	*Eremanthus erythropappus*
84	L39424.1	*Eremothamnus marlothianus*
85	AF153645.1	*Eriocephalus africanus*
86	EF089566.1	*Feddea cubensis*
87	L39459.1	*Fitchia sp.*

88	DQ444740.1	*Gazania krebsiana*
89	L39423.1	*Gazania splendens*
90	AF063074.1	*Geigeria ornativa*
91	EU385166.1	*Gochnatia hiriartiana*
92	EU385165.1	*Gochnatia hypoleuca*
93	AF233808.1	*Gochnatia hypoleuca 2*
94	L39397.1	*Gochnatia paucifolia*
95	EU385168.1	*Gorteria diffusa*
96	AF384728.3	*Greenmaniella resinosa*
97	EU385169.1	*Gundelia tournefortii*
98	L39429.1	*Gutenbergia polytrichotoma*
99	AY780825.1	*Gymnarrhena micrantha*
100	DQ444754.1	*Haplocarpha lanata*
101	DQ444755.1	*Haplocarpha lyrata*
102	DQ444756.1	*Haplocarpha nervosa*
103	DQ444757.1	*Haplocarpha rueppellii*
104	L39426.1	*Haplocarpha scaposa*
105	DQ444758.1	*Haplocarpha scaposa 2*
106	DQ444759.1	*Haplocarpha schimperi*
107	EU385171.1	*Hecastocleis shockleyi*
108	AB530934.1	*Helianthus annuus*
109	AB254899.1	*Hemistepta lyrata*
110	EU385172.1	*Hesperomannia arbuscula*
111	AF092584.1	*Hesperomannia lydgatei*
112	EU385173.1	*Heterolepis aliena*
113	AF218351.1	*Hieracium longipilum*
114	DQ444741.1	*Hirpicium echinus*
115	EU385174.1	*Hoplophyllum spinosum*
116	EU385176.1	*Hyaloseris rubicunda*
117	AF218344.1	*Hyoseris radiata*
118	AF218333.1	*Hypochaeris uniflora*
119	AF384736.3	*Idiopappus quitensis*
120	FM208960.1	*Inula robynsi*
121	AY780853.1	*Iphionopsis rotundifolia*
122	AY780844.1	*Jasonia tuberosa*
123	AY780854.1	*Karelinia caspia*
124	AF384739.3	*Kingianthus paradoxus*
125	GU817880.1	*Lactuca canadensis*
126	L39389.1	*Lactuca sativa*
127	AB530948.1	*Lactuca sativa 2*
128	AY780861.1	*Laggera decurrens*
129	AF218330.1	*Leontodon saxatilis*
130	L39421.1	*Liabum glabrum*
131	EU385183.1	*Lycoseris crocata*
132	EU385184.1	*Macledium zeyheri*
133	AF218335.1	*Malacothrix saxatilis*

134	AF384750.3	*Milleria quinqueflora*
135	AY780863.1	*Mollera angolensis*
136	L39420.1	*Munnozia gigantea*
137	AF233834.1	*Mutisia hieronymi*
138	AF233836.1	*Mutisia ledifolia*
139	EU385185.1	*Mutisia retrorsa*
140	EU385188.1	*Oldenburgia grandis*
141	AY780828.1	*Ondetia linearis*
142	EF155707.1	*Orbivestus cinerascens*
143	AF384756.3	*Oteiza scandens*
144	EU385192.1	*Pachylaena atriplicifolia*
145	EF155729.1	*Parapolydora fastigiata*
146	EU385193.1	*Pasaccardoa grantii*
147	AY780830.1	*Pegolettia oxydonta*
148	AY780832.1	*Pentanema glanduligerum*
149	AB288543.1	*Pertya glabrescens*
150	EU385195.1	*Pertya scandens*
151	AF218336.1	*Pilosella aurantiaca*
152	L39431.1	*Piptocarpha axillaris*
153	EU385196.1	*Platycarpha carlinoides*
154	EU385197.1	*Plazia daphnoides*
155	JX091732.1	*Pleiotaxis welwitschii*
156	EU385198.1	*Pluchea carolinensis*
157	EF155709.1	*Polydora fastigiata*
158	EU385199.1	*Proustia cuneifolia*
159	AY780859.1	*Pseudoconyza viscosa*
160	AF384772.3	*Psilostrophe gnaphalodes*
161	AF063078.1	*Pterocaulon sphacelatum*
162	FM209026.1	*Pulicaria undulata*
163	AF384773.3	*Raillardella argentea*
164	AF218350.1	*Reichardia tingitana*
165	AF218331.1	*Rhagadiolus stellatus*
166	AY780847.1	*Rhanterium epapposum*
167	AY226799.1	*Rhodogeron coronopifolius*
168	EU385202.1	*Richterago amplexifolia*
169	EU385127.1	*Richterago angustifolia*
170	AY780835.1	*Sachsia polycephala*
171	AY466427.1	*Saussurea erecta*
172	AF303926.1	*Scolymus hispanicus*
173	AF218332.1	*Scolymus hispanicus 2*
174	EU385204.1	*Scolymus maculatus*
175	KC589998.1	*Siebera pungens*
176	AF384781.3	*Silphium perfoliatum*
177	EU385181.1	*Sinclairia palmeri*
178	L39422.1	*Sinclairia pringlei*
179	EF155710.1	*Sipolesia languinosa*

180	AF384782.3	*Smallanthus microcephalus*
181	EU385206.1	*Sonchus oleraceus*
182	AF384785.3	*Squamopappus skutchii*
183	AY780869.1	*Stenachaenium campestre*
184	EU385207.1	*Stenopadus talaumifolius*
185	EU385208.1	*Stifftia chrysantha*
186	L39430.1	*Stokesia laevis*
187	EU385210.1	*Stomatochaeta condensata*
188	AF063082.1	*Streptoglossa liatroides*
189	EF155686.1	*Strobocalyx arboreum*
190	L39415.1	*Synurus deltoides*
191	KC590004.1	*Syreitschikovia spinulosa*
192	AF218346.1	*Taraxacum officinale (Leontodon vulgars)*
193	L39409.1	*Tarchonanthus camphoratus*
194	EU385212.1	*Tarchonanthus camphoratus 2*
195	AY780838.1	*Telekia speciosa*
196	EF155671.1	*Tephrothamnus paradoxa*
197	AF384789.3	*Tetrachyron orizabaensis*
198	AF218327.1	*Tolpis azorica*
199	AF218326.1	*Tolpis barbata*
200	AF218342.1	*Tolpis capensis*
201	AF218329.1	*Tolpis coronopifolia*
202	AF218328.1	*Tolpis farinulosa*
203	AF218337.1	*Tolpis staticifolia*
204	AF218325.1	*Tolpis virgata*
205	L39391.1	*Tragopogon porrifolius*
206	AF384794.3	*Tridax balbisioides*
207	KC590006.1	*Tugarinovia mogolica*
208	KC590007.1	*Tyrimnus leucographus*
209	AF218352.1	*Urospermum dalechampii*
210	AF384796.3	*Varilla mexicana*
211	GU817937.1	*Vernonia gigantea*
212	L39427.1	*Vernonia mespilifolia* subsp. *tomentosa*
213	EU385216.1	*Warionia saharae*
214	AY702088.1	*Warionia saharae 2*
215	EU385217.1	*Wunderlichia mirabilis*
216	KC590009.1	*Xeranthemum annuum*
217	EU385218.1	*Youngia japonica*
218	AY780850.1	*Zoutpansbergia caerulea*

Table S 2 Information of GenBank accessions (*trnL-F* gene)

Number	Accession number	Species name
1	EU385035.1	*Adenocaulon bicolor*
2	EU243153.1	*Ainsliaea apiculata*
3	EU243154.1	*Ainsliaea macrocephala*
4	FM997836.1	*Allagopappus canariensis*
5	DQ444809.1	*Arctotis acaulis*
6	DQ444810.1	*Arctotis arctotoides*
7	EU846486.1	*Arctotis argentea*
8	EU846488.1	*Arctotis aspera var. scabra*
9	EU846492.1	*Arctotis campanulata isolate*
10	EU846476.1	*Arctotis debensis*
11	EU846475.1	*Arctotis debensis 2*
12	EU846494.1	*Arctotis decurrens*
13	EU846506.1	*Arctotis incisa*
14	EU846479.1	*Arctotis microcephala*
15	EU846478.1	*Arctotis microcephala 2*
16	DQ444814.1	*Arctotis perfoliata*
17	EU846517.1	*Arctotis revoluta*
18	EU846482.1	*Arctotis scapiformis*
19	EU846519.1	*Arctotis semipapposa*
20	EU846525.1	*Arctotis sp.*
21	DQ444815.1	*Arctotis sp. 2*
22	EU846528.1	*Arctotis sp. 3*
23	JX083840.1	*Atractylodes carlinoides*
24	EF028336.1	*Atractylodes koreana*
25	AY504768.1	*Barnadesia caryophylla*
26	EU527264.1	*Berkheya angolesis*
27	EU527261.1	*Berkheya annectens*
28	EU527263.1	*Berkheya cardopatifolia*
29	AY504791.1	*Berkheya carlinopsis*
30	EU527257.1	*Berkheya cirsiifolia*
31	EU527254.1	*Berkheya echinacea*
32	EU527264.1	*Berkheya eriobasis*
33	EU527249.1	*Berkheya fruticosa*
34	EU527259.1	*Berkheya pannosa*
35	EU385043.1	*Berkheya purpurea*
36	EU527256.1	*Berkheya rhapontica*
37	EU195605.1	*Blumea aromatica*
38	EU243162.1	*Brachylaena elliptica*
39	EF211059.1	*Buphthalmum salicifolium*
40	EU195603.1	*Caesulia axillaris*
41	EF211060.1	*Carpesium divaricatum*
42	HM002862.1	*Carthamus arborescens*
43	KF486119.1	*Cephalorrhynchus macrorhizus (Melanoseris macrorhiza)*

44	KF486120.1	*Cephalorrhynchus macrorhizus 2(Melanoseris macrorhiza2)*
45	GU109295.1	*Chaetoseris cyanea (Melanoseris cyanea)*
46	GU109298.1	*Chaetoseris cyanea 2(Melanoseris cyanea 2)*
47	KF486144.1	*Chaetoseris grandiflora (Melanoseris atropurpurea)*
48	HQ436124.1	*Chaetoseris grandiflora 2 (Melanoseris atropurpurea 2)*
49	HQ436125.1	*Chaetoseris grandiflora 3(Melanoseris atropurpurea 3)*
50	KF486125.1	*Chaetoseris hastata(Melanoseris cyanea 3)*
51	KF486135.1	*Chaetoseris hispida(Melanoseris cyanea 4)*
52	KF486143.1	*Chaetoseris likiangensis (Melanoseris likiangensis 2)*
53	KF486129.1	*Chaetoseris lutea (Melanoseris cyanea 5)*
54	KF486130.1	*Chaetoseris lutea 2(Melanoseris cyanea 6)*
55	GU109296.1	*Chaetoseris lyriformis (Melanoseris cyanea 7)*
56	KF486132.1	*Chaetoseris lyriformis 2 (Melanoseris cyanea 8)*
57	KF486133.1	*Chaetoseris lyriformis 3 (Melanoseris cyanea 9)*
58	KF486121.1	*Chaetoseris macrantha (Melanoseris macrantha)*
59	KF486064.1	*Chaetoseris roborowskii(Cicerbita roborowskii)*
60	KF486065.1	*Chaetoseris roborowskii 2(Cicerbita roborowskii2)*
61	HQ436126.1	*Chaetoseris roborowskii 3(Cicerbita roborowskii3)*
62	KF486131.1	*Chaetoseris sichuanensis (Melanoseris cyanea 10)*
63	KF486137.1	*Chaetoseris sichuanensis 2 (Melanoseris cyanea 11)*
64	KF486148.1	*Chaetoseris sp.*
65	KF486145.1	*Chaetoseris taliensis (Melanoseris atropurpurea 4)*
66	KF486146.1	*Chaetoseris taliensis 2 (Melanoseris atropurpurea 5)*
67	KF486126.1	*Chaetoseris yunnanensis (Melanoseris yunnanesis)*
68	FM997842.1	*Chrysophtalmum gueneri*
69	KF486060.1	*Cicerbita azurea*
70	KF486061.1	*Cicerbita azurea 2*
71	KF486066.1	*Cicerbita oligolepis*
72	KF486067.1	*Cicerbita oligolepis 2*
73	KF486068.1	*Cicerbita oligolepis 3*
74	KF486122.1	*Cicerbita sikkimensis (Melanoseris violifolia)*
75	KF486123.1	*Cicerbita sikkimensis 2 (Melanoseris violifolia 2)*
76	GU817987.1	*Cichorium intybus*
77	EU385055.1	*Corymbium glabrum*
78	AB598610.1	*Crepidiastrum platyphyllum*
79	AF528396.1	*Crepis aurea*
80	AF528397.1	*Crepis viscidula*
81	EU527269.1	*Cullumia aculeata*
82	AY504795.1	*Cullumia bisulca*
83	EU527266.1	*Cullumia decurrens*
84	EU527265.1	*Cullumia patula*
85	AY504796.1	*Cullumia rigida*
86	EU527270.1	*Cuspidia cernua*
87	AY504797.1	*Cuspidia cernua 2*
88	EU846473.1	*Cymbonotus lawsonianus*
89	DQ444818.1	*Cymbonotus lawsonianus 2*

90	DQ889654.1	*Cymbonotus maidenii*
91	DQ889653.1	*Cymbonotus maidenii 2*
92	DQ889655.1	*Cymbonotus preissianus*
93	AY504767.1	*Dasyphyllum reticulatum*
94	EU527272.1	*Didelta carnosa*
95	AY504798.1	*Didelta carnosa 2*
96	EF530305.1	*Dolichlasium lagascae*
97	KF196085.1	*Dubyaea glaucescens*
98	KF196083.1	*Dubyaea hispida*
99	KF196082.1	*Dubyaea hispida 2*
100	HQ436127.1	*Dubyaea hispida 3*
101	EF211068.1	*Duhaldea cappa*
102	DQ444819.1	*Dymondia margaretae*
103	GU817998.1	*Endocellion sibiricum*
104	GU817999.1	*Erato polymnioides*
105	JN837197.1	*Eremothamnus marlothianus*
106	AB217695.1	*Eupatorium cannabinum*
107	GU818001.1	*Eupatorium serotinum*
108	KF196098.1	*Faberia cavaleriei*
109	KF196100.1	*Faberia cavaleriei 2*
110	KF196099.1	*Faberia faberi*
111	KF196101.1	*Faberia faberi 3*
112	KF196102.1	*Faberia nanchuanensis*
113	KF196104.1	*Faberia nanchuanensis 2*
114	KF196103.1	*Faberia nanchuanensis 3*
115	KF196105.1	*Faberia sinensis*
116	KF196106.1	*Faberia thibetica*
117	AY504769.1	*Gerbera crocea*
118	EU385071.1	*Gochnatia hypoleuca*
119	AY504773.1	*Gundelia tournefortii*
120	EU385076.1	*Gymarrhena micrantha*
121	EU729338.1	*Gypothamnium pinifolium*
122	DQ444820.1	*Haplocarpha lanata*
123	DQ444821.1	*Haplocarpha lyrata*
124	DQ889659.1	*Haplocarpha nervosa*
125	DQ889656.1	*Haplocarpha rueppellii*
126	DQ444824.1	*Haplocarpha scaposa*
127	DQ889660.1	*Haplocarpha scaposa 2*
128	AY504790.1	*Haplocarpha scaposa 3*
129	DQ444825.1	*Haplocarpha schimperi*
130	EU385077.1	*Hecastocleis shockleyi*
131	GU818008.1	*Helianthus tuberosus*
132	EU385079.1	*Heterolepis aliena*
133	AY504782.1	*Heterolepis aliena 2*
134	AF528399.1	*Hieracium glaucum*
135	AF528400.1	*Hieracium murorum*
136	AY504806.1	*Hirpicium echinus*

137	GU818013.1	*Homogyne alpina*
138	JN837196.1	*Hoplophyllum spinosum*
139	EU385080.1	*Hoplophyllum spinosum 2*
140	AY504784.1	*Hoplophyllum spinosum 3*
141	AF528401.1	*Hyoseris radiata*
142	AF528361.1	*Hypochaeris achyrophorus*
143	AY504774.1	*Hypochaeris glabra*
144	AF528374.1	*Hypochaeris maculata*
145	AF528382.1	*Hypochaeris robertia*
146	FM997851.1	*Inula paniculata*
147	FM997852.1	*Inula peacockiana*
148	FM997854.1	*Inula shirensis*
149	KF196069.1	*Ixeridium gracile*
150	GU818025.1	*Lactuca canadensis*
151	KF486161.1	*Lactuca dissecta*
152	GU109297.1	*Lactuca dolichophylla*
153	KF486162.1	*Lactuca dolichophylla 2*
154	GU109286.1	*Lactuca indica 2*
155	GU109288.1	*Lactuca indica 3*
156	KF486164.1	*Lactuca inermis 2*
157	KF486074.1	*Lactuca parishii*
158	KF486158.1	*Lactuca perennis 2*
159	AP007232.1	*Lactuca sativa*
160	GU109303.1	*Lactuca sativa 2*
161	AY504775.1	*Lactuca sativa 3*
162	GU109302.1	*Lactuca serriola*
163	KF486175.1	*Lactuca serriola 2*
164	KF486173.1	*Lactuca sibirica*
165	GU109301.1	*Lactuca tatarica 3*
166	KF486174.1	*Lactuca tatarica 2*
167	KF486159.1	*Lactuca undulata*
168	KF486160.1	*Lactuca undulata 2*
169	KF486048.1	*Launaea sarmentosa*
170	AF528391.1	*Leontodon autumnalis*
171	DQ449600.1	*Leontodon boryi*
172	DQ449602.1	*Leontodon carpetanus*
173	DQ449610.1	*Leontodon cichoraceus*
174	AF528392.1	*Leontodon crispus*
175	DQ449605.1	*Leontodon helveticus*
176	AF528393.1	*Leontodon hispidus*
177	DQ449604.1	*Leontodon longirostris*
178	DQ449608.1	*Leontodon maroccanus*
179	DQ449609.1	*Leontodon nevadensis*
180	AF528394.1	*Leontodon saxatilis (outgroup)*
181	DQ449599.1	*Leontodon saxatilis 2*
182	DQ449611.1	*Leontodon tingitanus*
183	KF486049.1	*Leontodon tuberosus*

184	EF155873.1	*Lepidaploa borinquensis*
185	EF155874.1	*Lepidaploa canescens*
186	EF155875.1	*Lepidaploa tortuos*
187	KF196097.1	*Melanoseris beesiana (cyanea 12)*
188	KF486128.1	*Melanoseris cyanea13*
189	KF196088.1	*Melanoseris likiangensis (Lactuca likiangensis)*
190	KF196086.1	*Melanoseris sp.*
191	KF196089.1	*Melanoseris taliensis (Melanoseris atropurpurea6)*
192	JN837207.1	*Moquinia racemosa*
193	KF486118.1	*Mulgedium bracteatum(Melanoseris bracteata)*
194	KF486117.1	*Mulgedium lessertianum (Melanoseris qinghaica2)*
195	KF486124.1	*Mulgedium qinghaicum (Melanoseris qinghaica)*
196	KF486075.1	*Mulgedium umbrosum*
197	KF486076.1	*Mulgedium umbrosum 2*
198	JN837245.1	*Munnozia pinnatipartita*
199	KF486108.1	*Notoseris henryi*
200	KF486109.1	*Notoseris henryi*
201	KF196093.1	*Notoseris macilenta*
202	KF196092.1	*Notoseris macilenta 2*
203	KF486092.1	*Notoseris melanantha*
204	KF486093.1	*Notoseris melanantha 2*
205	KF486094.1	*Notoseris melanantha 3*
206	KF196091.1	*Notoseris porphyrolepis*
207	KF486110.1	*Notoseris psilolepis*
208	KF486111.1	*Notoseris psilolepis 2*
209	KF486112.1	*Notoseris rhombiformis*
210	KF486113.1	*Notoseris rhombiformis 2*
211	KF486114.1	*Notoseris rhombiformis 3*
212	KF486104.1	*Notoseris triflora 2*
213	KF486095.1	*Notoseris wilsonii*
214	KF486096.1	*Notoseris wilsonii 2*
215	KF486097.1	*Notoseris wilsonii 3*
216	EU385094.1	*Oldenburgia grandis*
217	AB598611.1	*Paraixeris denticulata*
218	KF486085.1	*Paraprenanthes diversifolia*
219	KF486080.1	*Paraprenanthes glandulosissim*
220	KF486087.1	*Paraprenanthes gracilipes*
221	KF486082.1	*Paraprenanthes hastata*
222	KF486086.1	*Paraprenanthes heptantha*
223	KF486088.1	*Paraprenanthes longiloba*
224	KF486077.1	*Paraprenanthes luchunensis*
225	KF486078.1	*Paraprenanthes multiformis*
226	KF486083.1	*Paraprenanthes pilipes*
227	KF486084.1	*Paraprenanthes pilipes 2*
228	KF486079.1	*Paraprenanthes polypodifolia*
229	KF486081.1	*Paraprenanthes prenanthoides*
230	KF486089.1	*Paraprenanthes sagittiformis*

231	KF486090.1	*Paraprenanthes sagittiformis 2*
232	KF196096.1	*Paraprenanthes sororia*
233	GU109299.1	*Paraprenanthes yunnanensis (Melanoseris yunnanesis 2)*
234	KF486091.1	*Paraprenanthes yunnanensis 2(Melanoseris yunnanesis 3)*
235	KF486115.1	*Parasyncalathium souliei*
236	KF486116.1	*Parasyncalathium souliei 2*
237	GU818045.1	*Perityle emoryi*
238	EU385101.1	*Pertya scandens*
239	DQ449612.1	*Picris echioides*
240	GU818051.1	*Polymnia canadensis*
241	KF486098.1	*Prenanthes scandens*
242	KF486099.1	*Prenanthes scandens x Prenanthes yakoensis*
243	GU109292.1	*Prenanthes sp.*
244	KF486100.1	*Prenanthes yakoensis*
245	KF486101.1	*Prenanthes yakoensis 2*
246	GU109287.1	*Pterocypsela elata (Lactuca raddeana 3)*
247	KF486170.1	*Pterocypsela elata 2 (Lactuca raddeana 4)*
248	GU109300.1	*Pterocypsela formosana (Lactuca formosana 2)*
249	KF486167.1	*Pterocypsela formosana 2 (Lactuca formosana 2)*
250	GU109291.1	*Pterocypsela laciniata (Lactuca indica 4)*
251	KF486169.1	*Pterocypsela raddeana (Lactuca raddeana 2)*
252	GU109290.1	*Pterocypsela sonchus (Lactuca formosana 3)*
253	KF486168.1	*Pterocypsela sonchus 2 (Lactuca formosana 4)*
254	FM997884.1	*Rhanterium suaveolens*
255	EU385108.1	*Richterago amplexifolia*
256	EU385033.1	*Richterago angustifolia*
257	AY328109.1	*Saussurea erubescens*
258	AJ606153.1	*Saussurea erubescens 2*
259	AY328096.1	*Saussurea hookeri*
260	AY328133.1	*Saussurea iodostegia*
261	AJ606143.1	*Saussurea muliensis*
262	AY328105.1	*Saussurea tangutica*
263	AJ606145.1	*Saussurea velutina*
264	KF486171.1	*Scariola orientalis (Lactuca orientalis)*
265	KF486172.1	*Scariola viminea (Lactuca viminea)*
266	KF196067.1	*Sonchus oleraceus*
267	KF196068.1	*Sonchus oleraceus 2*
268	KF196066.1	*Sonchus sp.*
269	KF486062.1	*Stenoseris auriculiformis*
270	KF486063.1	*Stenoseris auriculiformis 2*
271	KF486155.1	*Stenoseris graciliflora*
272	KF486157.1	*Stenoseris graciliflora 2*
273	KF486070.1	*Stenoseris leptantha*
274	KF486071.1	*Stenoseris leptantha 2*
275	KF486156.1	*Stenoseris taliensis*
276	KF486149.1	*Stenoseris tenuis*
277	KF486150.1	*Stenoseris tenuis 2*

278	KF486151.1	*Stenoseris tenuis 3*
279	KF486152.1	*Stenoseris tenuis 4*
280	KF486072.1	*Stenoseris triflora*
281	KF486073.1	*Stenoseris triflora 2*
282	KF486163.1	*Steptorhamphus tuberosus (Lactuca tuberosus)*
283	JF920297.1	*Stifftia fruticosa*
284	JF920298.1	*Stifftia parviflora*
285	EU385116.1	*Stomatachaeta condensata*
286	HQ436141.1	*Syncalathium souliei (Parasyncalathium souliei3)*
287	HQ436140.1	*Syncalathium souliei 2 (Parasyncalathium souliei 4)*
288	EF211066.1	*Telekia speciosa*
289	GU818101.1	*Urostemon kirkii*
290	EF155891.1	*Vernonanthura patens*
291	EF155894.1	*Vernonia altissima*
292	EF155915.1	*Vernonia brasiliana*
293	EF155914.1	*Vernonia profuga*
294	EF155918.1	*Vernonia subplumosa*
295	EF155919.1	*Vernonia texana*
296	EU385122.1	*Wunderlichia mirabilis*
297	KF196072.1	*Youngia denticulata*
298	AB598609.1	*Youngia erythrocarpa*
299	KF196075.1	*Youngia heterophylla*
300	EU385123.1	*Youngia japonica*
301	KF196076.1	*Youngia paleacea*
302	AB598607.1	*Youngia pseudosenecio*

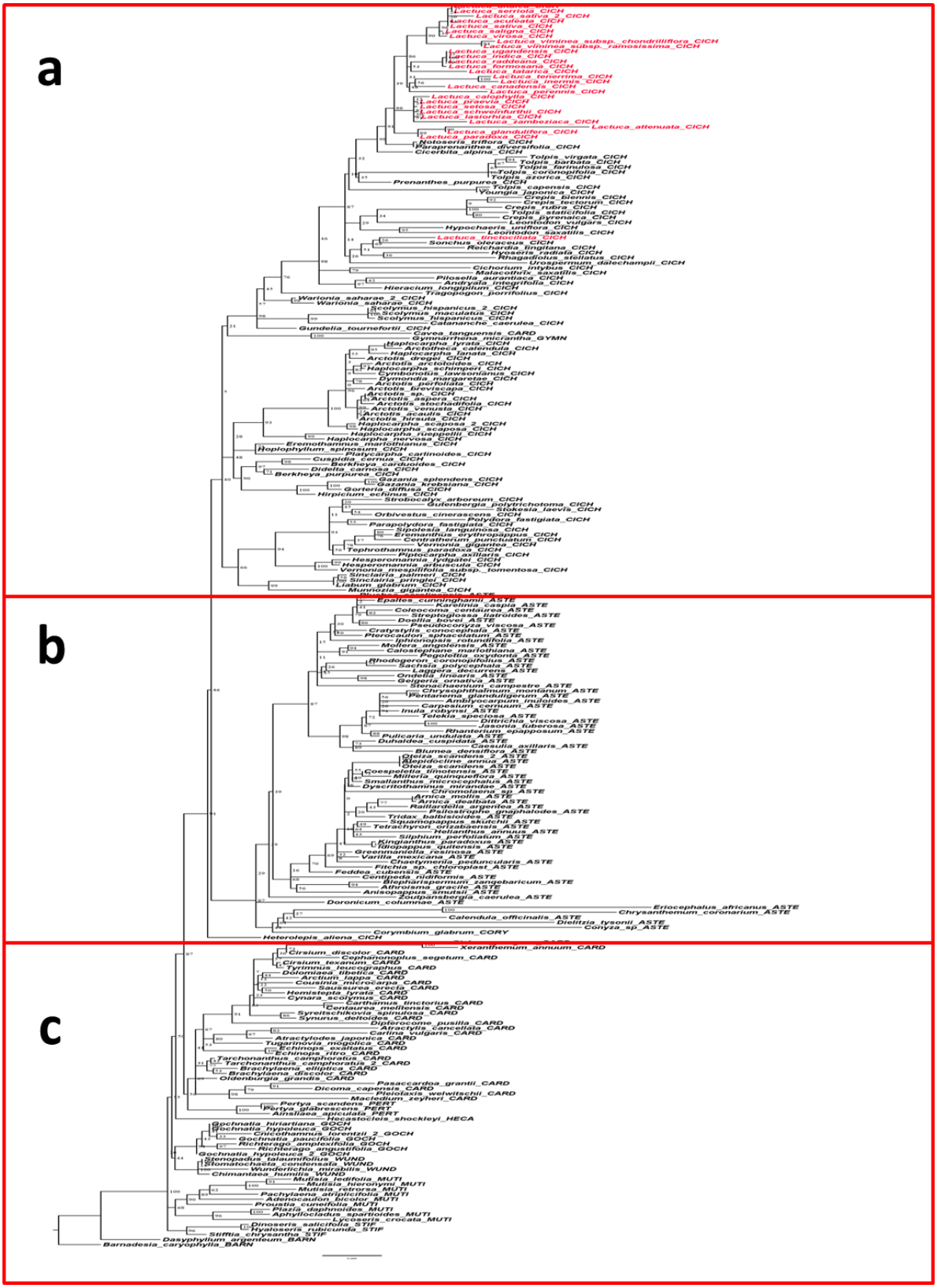


Figure S1 RAxML phylogram (best tree) of 247 accessions from different sub-families of Asteraceae, based on *ndhF* gene; *Lactuca* species are shown in red colour; bootstrap (BS) supporting values are given; *L. tinctociliata* was mis-identified and it could be *Launaea cornuta*; *L. ugandensis* should be *Lactuca* sp.

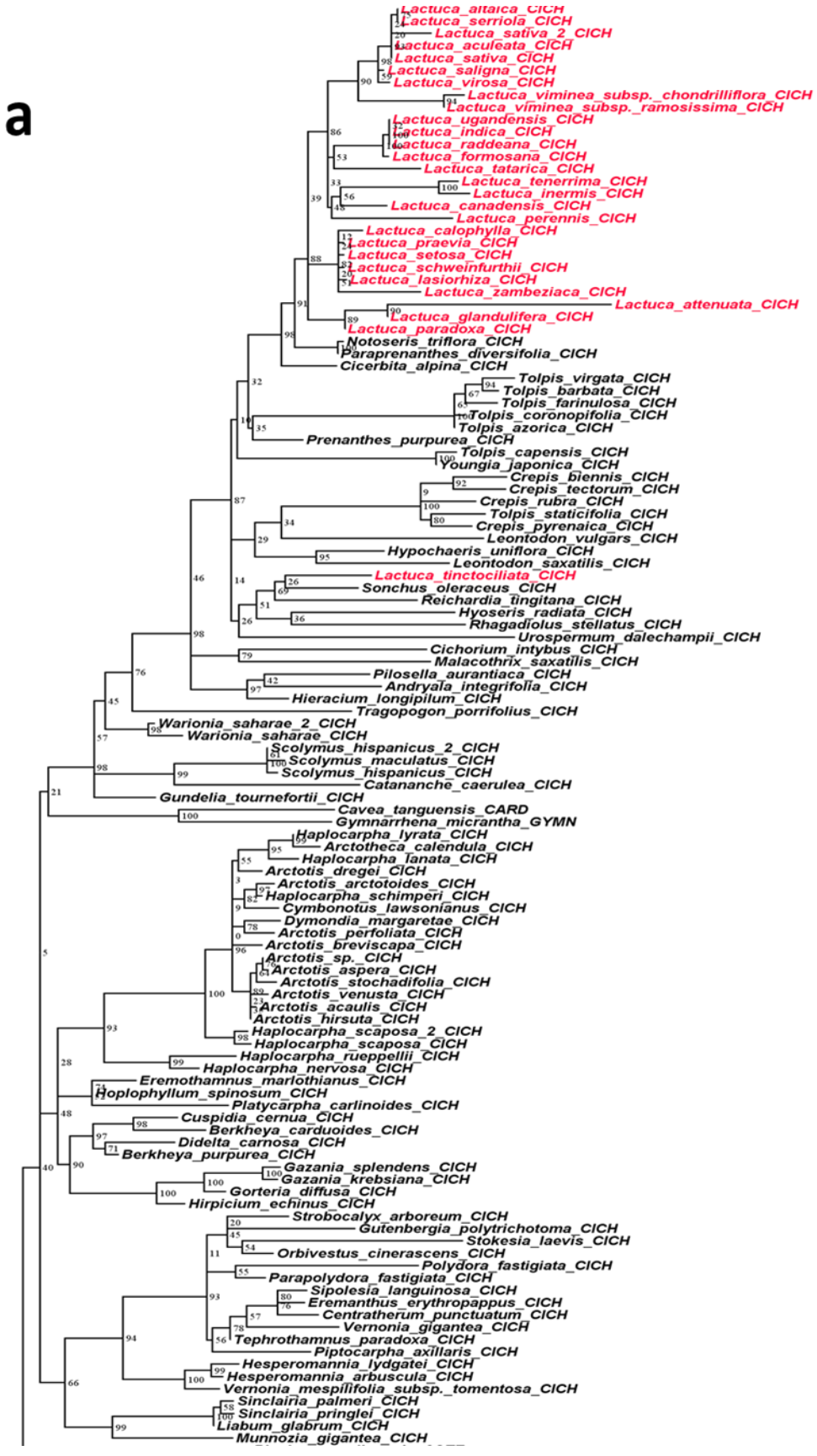

a
Lactuca_altaica_CICH
Lactuca_serriola_CICH
Lactuca_sativa_2_CICH
Lactuca_aculeata_CICH
Lactuca_sativa_CICH
Lactuca_saligna_CICH
Lactuca_virosa_CICH
Lactuca_viminea_subsp._chondrilliflora_CICH
Lactuca_viminea_subsp._ramosissima_CICH
Lactuca_ugandensis_CICH
Lactuca_indica_CICH
Lactuca_raddeana_CICH
Lactuca_formosana_CICH
Lactuca_tatarica_CICH
Lactuca_tenerrima_CICH
Lactuca_inermis_CICH
Lactuca_canadensis_CICH
Lactuca_perennis_CICH
Lactuca_calophylla_CICH
Lactuca_praevia_CICH
Lactuca_setosa_CICH
Lactuca_schweinfurthii_CICH
Lactuca_lasiorhiza_CICH
Lactuca_zambeziaca_CICH
Lactuca_attenuata_CICH
Lactuca_glandulifera_CICH
Lactuca_paradoxa_CICH
Notoseris_triflora_CICH
Paraprenanthes_diversifolia_CICH
Cicerbita_alpina_CICH
Tolpis_virgata_CICH
Tolpis_barbata_CICH
Tolpis_farinulosa_CICH
Tolpis_coronopifolia_CICH
Tolpis_azorica_CICH
Prenanthes_purpurea_CICH
Tolpis_capensis_CICH
Youngia_japonica_CICH
Crepis_biennis_CICH
Crepis_tectorum_CICH
Crepis_rubra_CICH
Tolpis_staticifolia_CICH
Crepis_pyrenaica_CICH
Leontodon_vulgaris_CICH
Hypochaeris_uniflora_CICH
Leontodon_saxatilis_CICH
Lactuca_tinctociliata_CICH
Sonchus_oleraceus_CICH
Reichardia_tingitana_CICH
Hyoseris_radiata_CICH
Rhagadiolus_stellatus_CICH
Urospermum_dalechampii_CICH
Cichorium_intybus_CICH
Malacothrix_saxatilis_CICH
Pilosella_aurantiaca_CICH
Andryala_integrifolia_CICH
Hieracium_longipilum_CICH
Tragopogon_porrifolius_CICH
Warionia_saharae_2_CICH
Warionia_saharae_CICH
Scolymus_hispanicus_2_CICH
Scolymus_maculatus_CICH
Scolymus_hispanicus_CICH
Catananche_caerulea_CICH
Gundelia_tournefortii_CICH
Cavea_tanguensis_CARD
Gymnarrhena_micrantha_GYMN
Haplocarpha_lyrata_CICH
Arctotheca_calendula_CICH
Haplocarpha_lanata_CICH
Arctotis_dregei_CICH
Arctotis_arctotoides_CICH
Haplocarpha_schimperi_CICH
Cymbonotus_lawsonianus_CICH
Dymondia_margaretae_CICH
Arctotis_perfoliata_CICH
Arctotis_breviscapa_CICH
Arctotis_sp._CICH
Arctotis_aspera_CICH
Arctotis_stochadifolia_CICH
Arctotis_venusta_CICH
Arctotis_acaulis_CICH
Arctotis_hirsuta_CICH
Haplocarpha_scaposa_2_CICH
Haplocarpha_scaposa_CICH
Haplocarpha_rueppellii_CICH
Haplocarpha_nervosa_CICH
Eremothamnus_marlothianus_CICH
Hoplophyllum_spinosum_CICH
Platycarpha_carlinoides_CICH
Cuspidia_cernua_CICH
Berkheya_carduoides_CICH
Didelta_carnosa_CICH
Berkheya_purpurea_CICH
Gazania_splendens_CICH
Gazania_krebsiana_CICH
Gorteria_diffusa_CICH
Hirpicium_echinus_CICH
Strobocalyx_arboreum_CICH
Gutenbergia_polytrichotoma_CICH
Stokesia_laevis_CICH
Orbivestus_cinerascens_CICH
Polydora_fastigiata_CICH
Parapolydora_fastigiata_CICH
Sipolesia_languinosa_CICH
Eremanthus_erythropappus_CICH
Centratherum_punctuatum_CICH
Vernonia_gigantea_CICH
Tephrothamnus_paradoxa_CICH
Piptocarpha_axillaris_CICH
Hesperomannia_lydgatei_CICH
Hesperomannia_arbuscula_CICH
Vernonia_mespilifolia_subsp._tomentosa_CICH
Sinclairia_palmeri_CICH
Sinclairia_pringlei_CICH
Liabum_glabrum_CICH
Munnozia_gigantea_CICH

Figure S1a

b

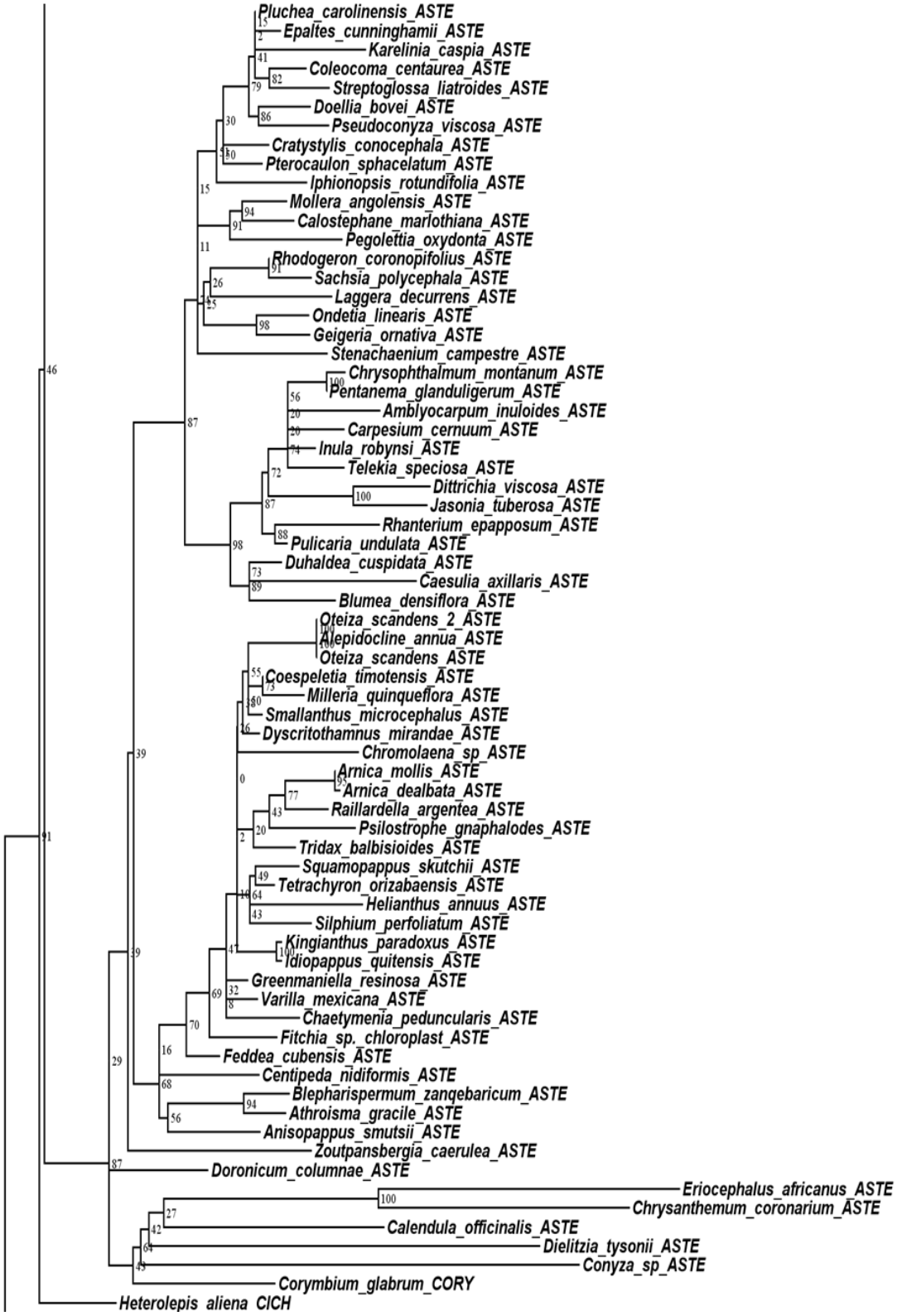

Pluchea_carolinensis_ASTE
Epaltes_cunninghamii_ASTE
Karelinia_caspia_ASTE
Coleocoma_centaurea_ASTE
Streptoglossa_liatroides_ASTE
Doellia_bovei_ASTE
Pseudoconyza_viscosa_ASTE
Cratystylis_conocephala_ASTE
Pterocaulon_sphacelatum_ASTE
Iphionopsis_rotundifolia_ASTE
Mollera_angolensis_ASTE
Calostephane_marlothiana_ASTE
Pegolettia_oxydonta_ASTE
Rhodogeron_coronopifolius_ASTE
Sachsia_polycephala_ASTE
Laggera_decurrens_ASTE
Ondetia_linearis_ASTE
Geigeria_ornativa_ASTE
Stenachaenium_campestre_ASTE
Chrysophthalmum_montanum_ASTE
Pentanema_glanduligerum_ASTE
Amblyocarpum_inuloides_ASTE
Carpesium_cernuum_ASTE
Inula_robynsi_ASTE
Telekia_speciosa_ASTE
Dittrichia_viscosa_ASTE
Jasonia_tuberosa_ASTE
Rhanterium_epapposum_ASTE
Pulicaria_undulata_ASTE
Duhaldea_cuspidata_ASTE
Caesulia_axillaris_ASTE
Blumea_densiflora_ASTE
Oteiza_scandens_2_ASTE
Alepidocline_annua_ASTE
Oteiza_scandens_ASTE
Coespeletia_timotensis_ASTE
Milleria_quinqueflora_ASTE
Smallanthus_microcephalus_ASTE
Dyscritothamnus_mirandae_ASTE
Chromolaena_sp_ASTE
Arnica_mollis_ASTE
Arnica_dealbata_ASTE
Raillardella_argentea_ASTE
Psilostrophe_gnaphalodes_ASTE
Tridax_balbisioides_ASTE
Squamopappus_skutchii_ASTE
Tetrachyron_orizabaensis_ASTE
Helianthus_annuus_ASTE
Silphium_perfoliatum_ASTE
Kingianthus_paradoxus_ASTE
Idiopappus_quitensis_ASTE
Greenmaniella_resinosa_ASTE
Varilla_mexicana_ASTE
Chaetymenia_peduncularis_ASTE
Fitchia_sp._chloroplast_ASTE
Feddea_cubensis_ASTE
Centipeda_nidiformis_ASTE
Blepharispermum_zanqebaricum_ASTE
Athroisma_gracile_ASTE
Anisopappus_smutsii_ASTE
Zoutpansbergia_caerulea_ASTE
Doronicum_columnae_ASTE
Eriocephalus_africanus_ASTE
Chrysanthemum_coronarium_ASTE
Calendula_officinalis_ASTE
Dielitzia_tysonii_ASTE
Conyza_sp_ASTE
Corymbium_glabrum_CORY
Heterolepis_aliena_CICH

Figure S1b

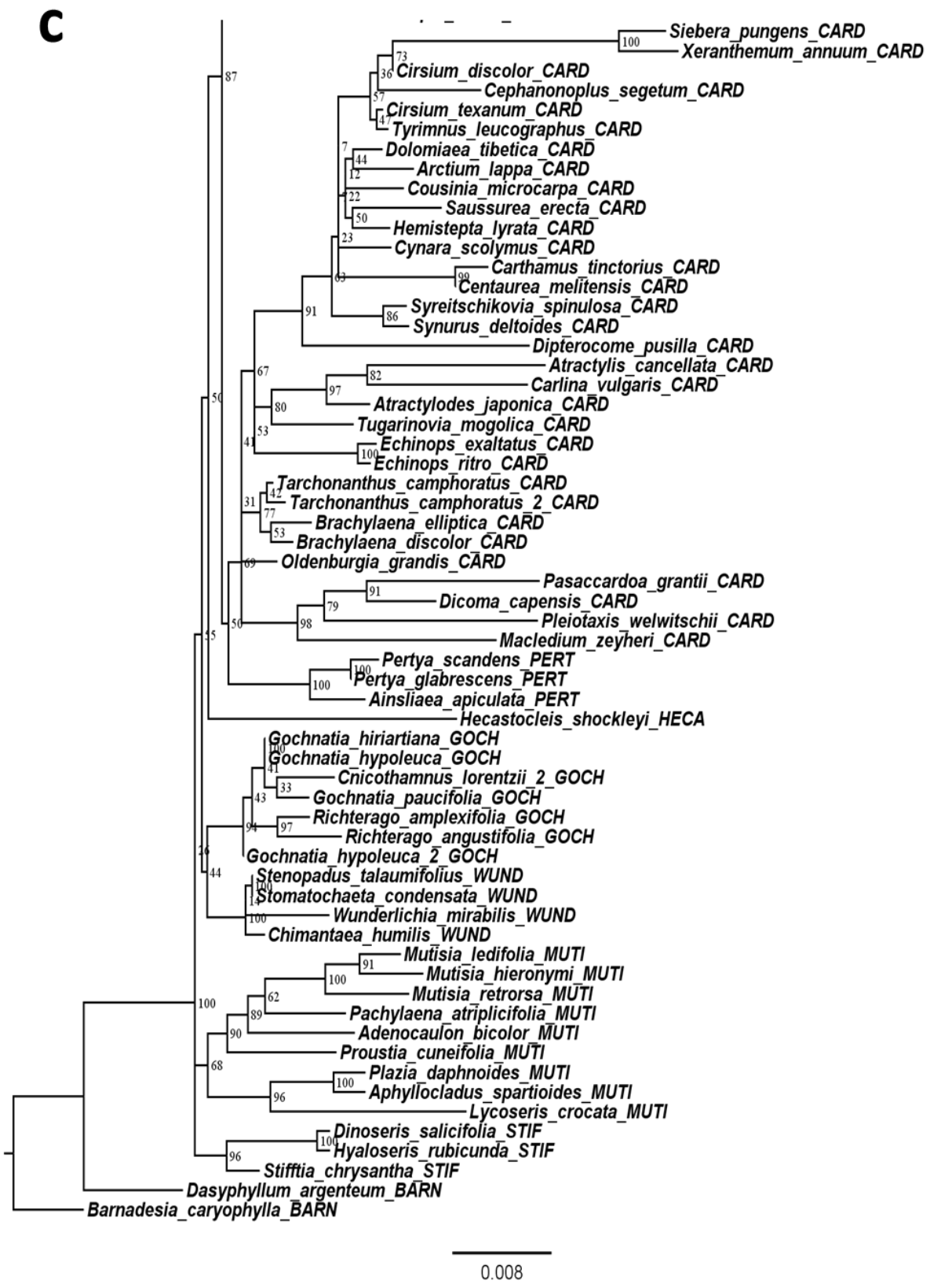

C
Siebera_pungens_CARD
Xeranthemum_annuum_CARD
Cirsium_discolor_CARD
Cephanonoplus_segetum_CARD
Cirsium_texanum_CARD
Tyrimnus_leucographus_CARD
Dolomiaea_tibetica_CARD
Arctium_lappa_CARD
Cousinia_microcarpa_CARD
Saussurea_erecta_CARD
Hemistepta_lyrata_CARD
Cynara_scolymus_CARD
Carthamus_tinctorius_CARD
Centaurea_melitensis_CARD
Syreitschikovia_spinulosa_CARD
Synurus_deltoides_CARD
Dipterocome_pusilla_CARD
Atractylis_cancellata_CARD
Carlina_vulgaris_CARD
Atractylodes_japonica_CARD
Tugarinovia_mogolica_CARD
Echinops_exaltatus_CARD
Echinops_ritro_CARD
Tarchonanthus_camphoratus_CARD
Tarchonanthus_camphoratus_2_CARD
Brachylaena_elliptica_CARD
Brachylaena_discolor_CARD
Oldenburgia_grandis_CARD
Pasaccardoa_grantii_CARD
Dicoma_capensis_CARD
Pleiotaxis_welwitschii_CARD
Macledium_zeyheri_CARD
Pertya_scandens_PERT
Pertya_glabrescens_PERT
Ainsliaea_apiculata_PERT
Hecastocleis_shockleyi_HECA
Gochnatia_hiriartiana_GOCH
Gochnatia_hypoleuca_GOCH
Cnicothamnus_lorentzii_2_GOCH
Gochnatia_paucifolia_GOCH
Richterago_amplexifolia_GOCH
Richterago_angustifolia_GOCH
Gochnatia_hypoleuca_2_GOCH
Stenopadus_talaumifolius_WUND
Stomatochaeta_condensata_WUND
Wunderlichia_mirabilis_WUND
Chimantaea_humilis_WUND
Mutisia_ledifolia_MUTI
Mutisia_hieronymi_MUTI
Mutisia_retrorsa_MUTI
Pachylaena_atriplicifolia_MUTI
Adenocaulon_bicolor_MUTI
Proustia_cuneifolia_MUTI
Plazia_daphnoides_MUTI
Aphyllocladus_spartioides_MUTI
Lycoseris_crocata_MUTI
Dinoseris_salicifolia_STIF
Hyaloseris_rubicunda_STIF
Stifftia_chrysantha_STIF
Dasyphyllum_argenteum_BARN
Barnadesia_caryophylla_BARN
0.008

Figure S1c

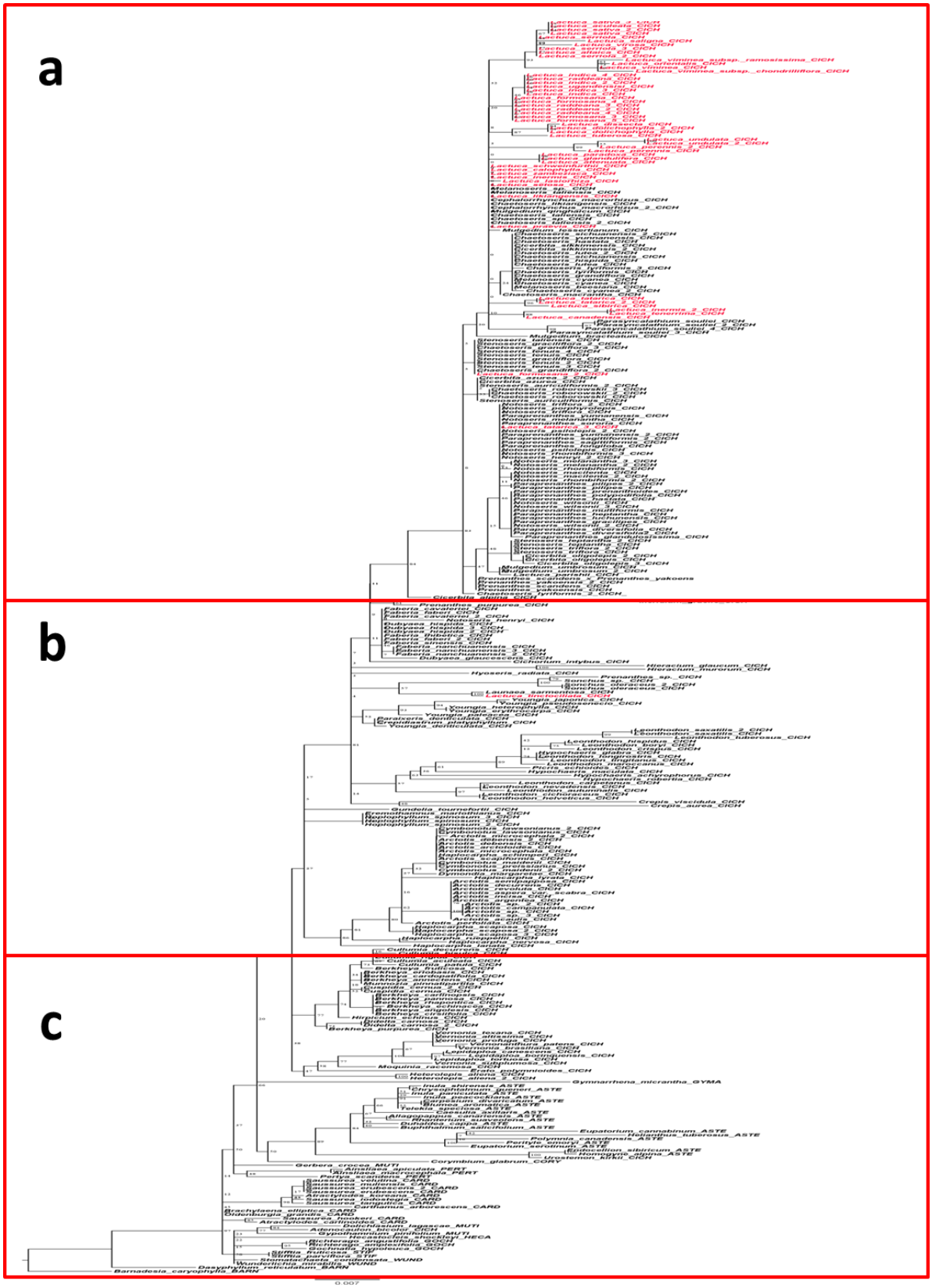


Figure S2 RAxML phylogram (best tree) of 331 accessions from different sub-families of Asteraceae, based on *trnL-F* gene; *Lactuca* species are shown in red colour; bootstrap (BS) supporting values are given; *L. tinctociliata* was mis-identified and it could be *Launaea cornuta*; *L. ugandensis* should be *Lactuca* sp.

a

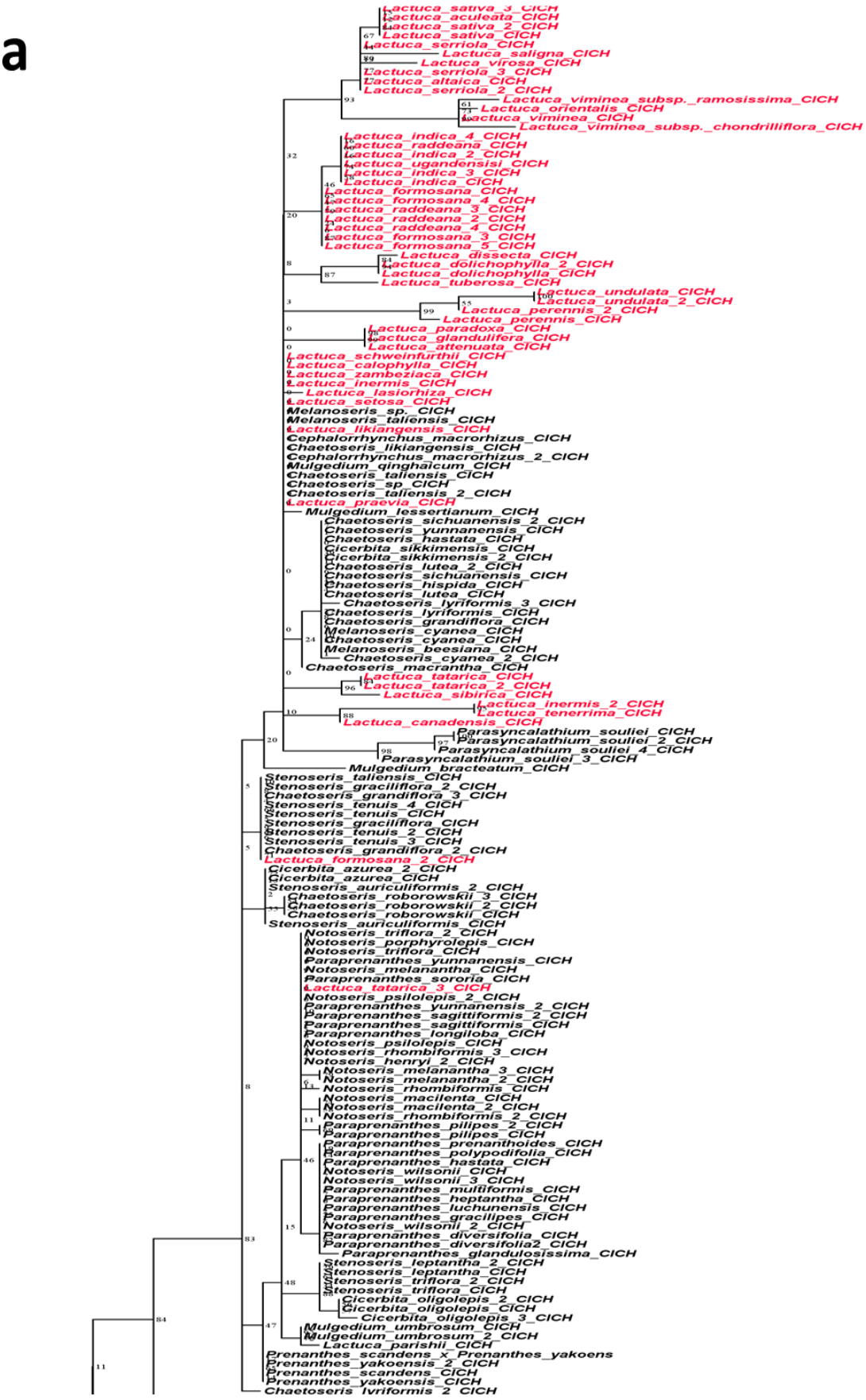

Lactuca_sativa_3_CICH
Lactuca_aculeata_CICH
Lactuca_sativa_2_CICH
Lactuca_sativa_CICH
Lactuca_serriola_CICH
Lactuca_saligna_CICH
Lactuca_virosa_CICH
Lactuca_serriola_3_CICH
Lactuca_altaica_CICH
Lactuca_serriola_2_CICH
Lactuca_viminea_subsp._ramosissima_CICH
Lactuca_orientalis_CICH
Lactuca_viminea_CICH
Lactuca_viminea_subsp._chondrilliflora_CICH
Lactuca_indica_4_CICH
Lactuca_raddeana_CICH
Lactuca_indica_2_CICH
Lactuca_ugandensisi_CICH
Lactuca_indica_3_CICH
Lactuca_indica_CICH
Lactuca_formosana_CICH
Lactuca_formosana_4_CICH
Lactuca_raddeana_3_CICH
Lactuca_raddeana_2_CICH
Lactuca_raddeana_4_CICH
Lactuca_formosana_3_CICH
Lactuca_formosana_5_CICH
Lactuca_dissecta_CICH
Lactuca_dolichophylla_2_CICH
Lactuca_dolichophylla_CICH
Lactuca_tuberosa_CICH
Lactuca_undulata_CICH
Lactuca_undulata_2_CICH
Lactuca_perennis_2_CICH
Lactuca_perennis_CICH
Lactuca_paradoxa_CICH
Lactuca_glandulifera_CICH
Lactuca_attenuata_CICH
Lactuca_schweinfurthii_CICH
Lactuca_calophylla_CICH
Lactuca_zambeziaca_CICH
Lactuca_inermis_CICH
Lactuca_lasiorhiza_CICH
Lactuca_setosa_CICH
Melanoseris_sp._CICH
Melanoseris_taliensis_CICH
Lactuca_likiangensis_CICH
Cephalorrhynchus_macrorhizus_CICH
Chaetoseris_likiangensis_CICH
Cephalorrhynchus_macrorhizus_2_CICH
Mulgedium_qinghaicum_CICH
Chaetoseris_taliensis_CICH
Chaetoseris_sp._CICH
Chaetoseris_taliensis_2_CICH
Lactuca_praevia_CICH
Mulgedium_lessertianum_CICH
Chaetoseris_sichuanensis_2_CICH
Chaetoseris_yunnanensis_CICH
Chaetoseris_hastata_CICH
Cicerbita_sikkimensis_CICH
Cicerbita_sikkimensis_2_CICH
Chaetoseris_lutea_2_CICH
Chaetoseris_sichuanensis_CICH
Chaetoseris_hispida_CICH
Chaetoseris_lutea_CICH
Chaetoseris_lyriformis_3_CICH
Chaetoseris_lyriformis_CICH
Chaetoseris_grandiflora_CICH
Melanoseris_cyanea_CICH
Chaetoseris_cyanea_CICH
Melanoseris_beesiana_CICH
Chaetoseris_cyanea_2_CICH
Chaetoseris_macrantha_CICH
Lactuca_tatarica_CICH
Lactuca_tatarica_2_CICH
Lactuca_sibirica_CICH
Lactuca_inermis_2_CICH
Lactuca_tenerrima_CICH
Lactuca_canadensis_CICH
Parasyncalathium_souliei_CICH
Parasyncalathium_souliei_2_CICH
Parasyncalathium_souliei_4_CICH
Parasyncalathium_souliei_3_CICH
Mulgedium_bracteatum_CICH
Stenoseris_taliensis_CICH
Stenoseris_graciliflora_2_CICH
Chaetoseris_grandiflora_3_CICH
Stenoseris_tenuis_4_CICH
Stenoseris_tenuis_CICH
Stenoseris_graciliflora_CICH
Stenoseris_tenuis_2_CICH
Stenoseris_tenuis_3_CICH
Chaetoseris_grandiflora_2_CICH
Lactuca_formosana_2_CICH
Cicerbita_azurea_2_CICH
Cicerbita_azurea_CICH
Stenoseris_auriculiformis_2_CICH
Chaetoseris_roborowskii_3_CICH
Chaetoseris_roborowskii_2_CICH
Chaetoseris_roborowskii_CICH
Stenoseris_auriculiformis_CICH
Notoseris_triflora_2_CICH
Notoseris_porphyrolepis_CICH
Notoseris_triflora_CICH
Paraprenanthes_yunnanensis_CICH
Notoseris_melanantha_CICH
Paraprenanthes_sororia_CICH
Lactuca_tatarica_3_CICH
Notoseris_psilolepis_2_CICH
Paraprenanthes_yunnanensis_2_CICH
Paraprenanthes_sagittiformis_2_CICH
Paraprenanthes_sagittiformis_CICH
Paraprenanthes_longiloba_CICH
Notoseris_psilolepis_CICH
Notoseris_rhombiformis_3_CICH
Notoseris_henryi_2_CICH
Notoseris_melanantha_3_CICH
Notoseris_melanantha_2_CICH
Notoseris_rhombiformis_CICH
Notoseris_macilenta_CICH
Notoseris_macilenta_2_CICH
Notoseris_rhombiformis_2_CICH
Paraprenanthes_pilipes_2_CICH
Paraprenanthes_pilipes_CICH
Paraprenanthes_prenanthoides_CICH
Paraprenanthes_polypodifolia_CICH
Paraprenanthes_hastata_CICH
Notoseris_wilsonii_CICH
Notoseris_wilsonii_3_CICH
Paraprenanthes_multiformis_CICH
Paraprenanthes_heptantha_CICH
Paraprenanthes_luchunensis_CICH
Paraprenanthes_gracilipes_CICH
Notoseris_wilsonii_2_CICH
Paraprenanthes_diversifolia_CICH
Paraprenanthes_diversifolia2_CICH
Paraprenanthes_glandulosissima_CICH
Stenoseris_leptantha_2_CICH
Stenoseris_leptantha_CICH
Stenoseris_triflora_2_CICH
Stenoseris_triflora_CICH
Cicerbita_oligolepis_2_CICH
Cicerbita_oligolepis_CICH
Cicerbita_oligolepis_3_CICH
Mulgedium_umbrosum_CICH
Mulgedium_umbrosum_2_CICH
Lactuca_parishii_CICH
Prenanthes_scandens_x_Prenanthes_yakoens
Prenanthes_yakoensis_2_CICH
Prenanthes_scandens_CICH
Prenanthes_yakoensis_CICH
Chaetoseris_lyriformis_2_CICH

Figure S2a

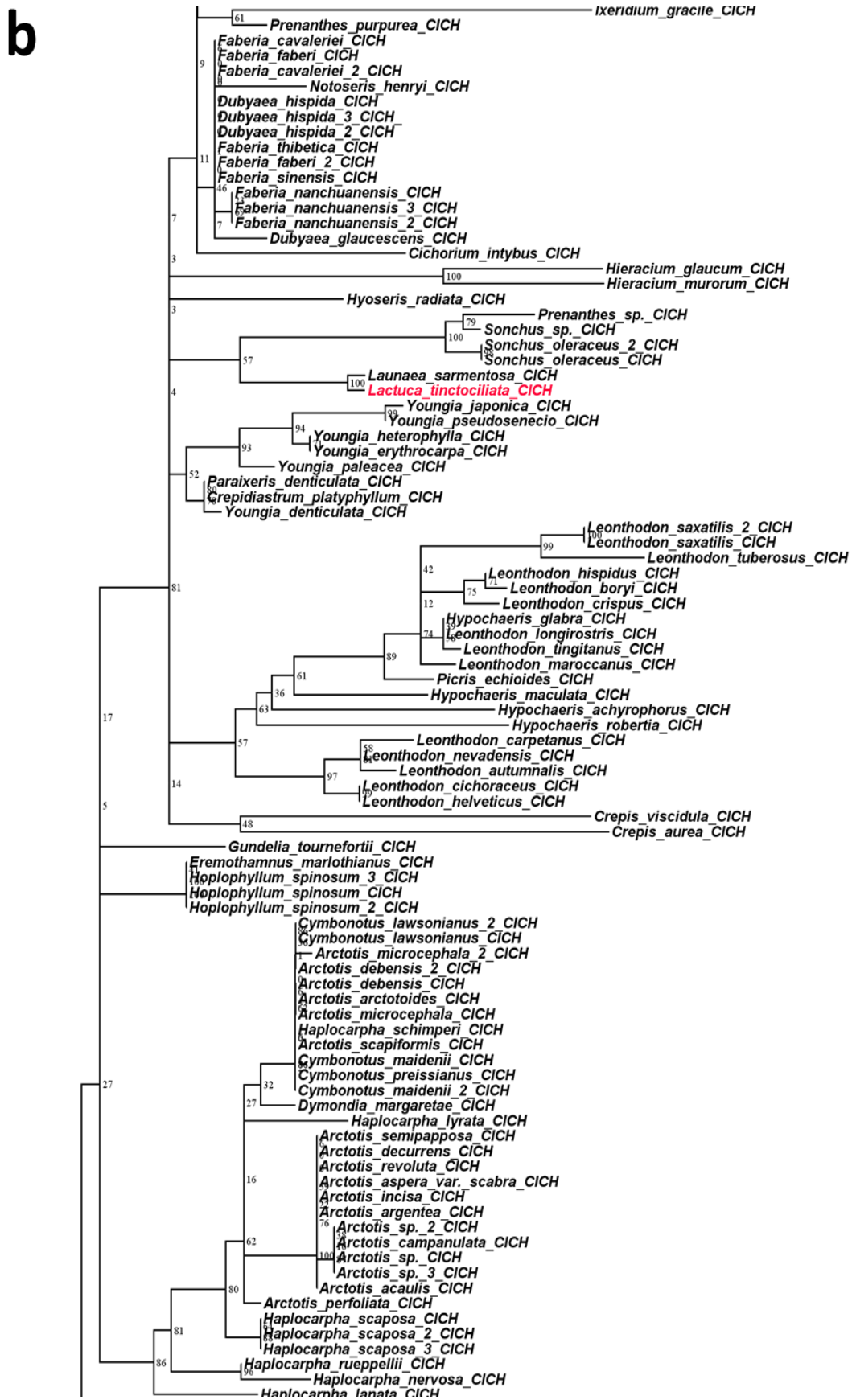

b
Ixeridium_gracile_CICH
Prenanthes_purpurea_CICH
Faberia_cavaleriei_CICH
Faberia_faberi_CICH
Faberia_cavaleriei_2_CICH
Notoseris_henryi_CICH
Dubyaea_hispida_CICH
Dubyaea_hispida_3_CICH
Dubyaea_hispida_2_CICH
Faberia_thibetica_CICH
Faberia_faberi_2_CICH
Faberia_sinensis_CICH
Faberia_nanchuanensis_CICH
Faberia_nanchuanensis_3_CICH
Faberia_nanchuanensis_2_CICH
Dubyaea_glaucescens_CICH
Cichorium_intybus_CICH
Hieracium_glaucum_CICH
Hieracium_murorum_CICH
Hyoseris_radiata_CICH
Prenanthes_sp._CICH
Sonchus_sp._CICH
Sonchus_oleraceus_2_CICH
Sonchus_oleraceus_CICH
Launaea_sarmentosa_CICH
Lactuca_tinctociliata_CICH
Youngia_japonica_CICH
Youngia_pseudosenecio_CICH
Youngia_heterophylla_CICH
Youngia_erythrocarpa_CICH
Youngia_paleacea_CICH
Paraixeris_denticulata_CICH
Crepidiastrum_platyphyllum_CICH
Youngia_denticulata_CICH
Leonthodon_saxatilis_2_CICH
Leonthodon_saxatilis_CICH
Leonthodon_tuberosus_CICH
Leonthodon_hispidus_CICH
Leonthodon_boryi_CICH
Leonthodon_crispus_CICH
Hypochaeris_glabra_CICH
Leonthodon_longirostris_CICH
Leonthodon_tingitanus_CICH
Leonthodon_maroccanus_CICH
Picris_echioides_CICH
Hypochaeris_maculata_CICH
Hypochaeris_achyrophorus_CICH
Hypochaeris_robertia_CICH
Leonthodon_carpetanus_CICH
Leonthodon_nevadensis_CICH
Leonthodon_autumnalis_CICH
Leonthodon_cichoraceus_CICH
Leonthodon_helveticus_CICH
Crepis_viscidula_CICH
Crepis_aurea_CICH
Gundelia_tournefortii_CICH
Eremothamnus_marlothianus_CICH
Hoplophyllum_spinosum_3_CICH
Hoplophyllum_spinosum_CICH
Hoplophyllum_spinosum_2_CICH
Cymbonotus_lawsonianus_2_CICH
Cymbonotus_lawsonianus_CICH
Arctotis_microcephala_2_CICH
Arctotis_debensis_2_CICH
Arctotis_debensis_CICH
Arctotis_arctotoides_CICH
Arctotis_microcephala_CICH
Haplocarpha_schimperi_CICH
Arctotis_scapiformis_CICH
Cymbonotus_maidenii_CICH
Cymbonotus_preissianus_CICH
Cymbonotus_maidenii_2_CICH
Dymondia_margaretae_CICH
Haplocarpha_lyrata_CICH
Arctotis_semipapposa_CICH
Arctotis_decurrens_CICH
Arctotis_revoluta_CICH
Arctotis_aspera_var._scabra_CICH
Arctotis_incisa_CICH
Arctotis_argentea_CICH
Arctotis_sp._2_CICH
Arctotis_campanulata_CICH
Arctotis_sp._CICH
Arctotis_sp._3_CICH
Arctotis_acaulis_CICH
Arctotis_perfoliata_CICH
Haplocarpha_scaposa_CICH
Haplocarpha_scaposa_2_CICH
Haplocarpha_scaposa_3_CICH
Haplocarpha_rueppellii_CICH
Haplocarpha_nervosa_CICH
Haplocarpha_lanata_CICH

Figure S2b

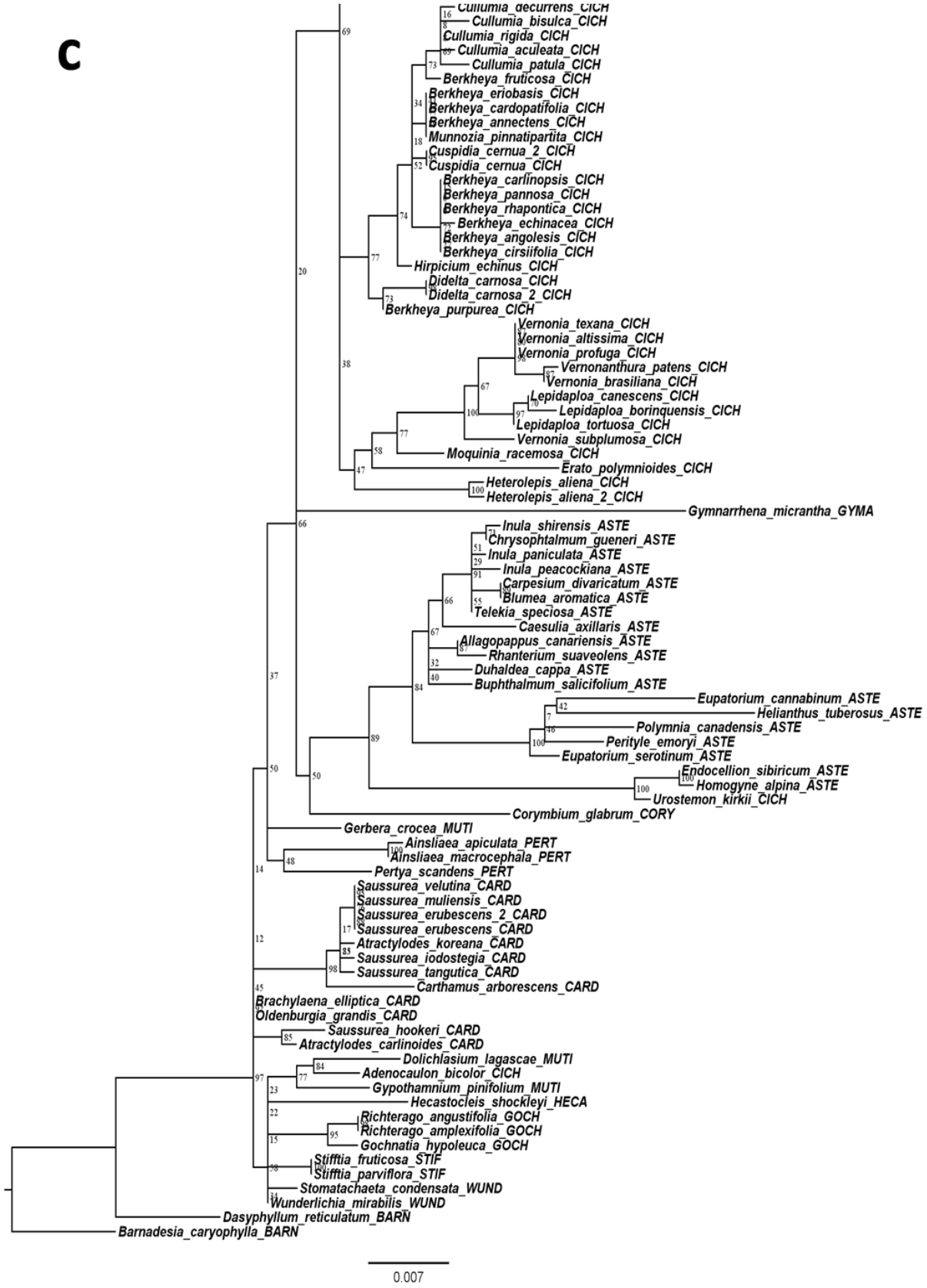

C
Cullumia_decurrens_CICH
Cullumia_bisulca_CICH
Cullumia_rigida_CICH
Cullumia_aculeata_CICH
Cullumia_patula_CICH
Berkheya_fruticosa_CICH
Berkheya_eriobasis_CICH
Berkheya_cardopatifolia_CICH
Berkheya_annectens_CICH
Munnozia_pinnatipartita_CICH
Cuspidia_cernua_2_CICH
Cuspidia_cernua_CICH
Berkheya_carlinopsis_CICH
Berkheya_pannosa_CICH
Berkheya_rhapontica_CICH
Berkheya_echinacea_CICH
Berkheya_angolesis_CICH
Berkheya_cirsiifolia_CICH
Hirpicium_echinus_CICH
Didelta_carnosa_CICH
Didelta_carnosa_2_CICH
Berkheya_purpurea_CICH
Vernonia_texana_CICH
Vernonia_altissima_CICH
Vernonia_profuga_CICH
Vernonanthura_patens_CICH
Vernonia_brasiliana_CICH
Lepidaploa_canescens_CICH
Lepidaploa_borinquensis_CICH
Lepidaploa_tortuosa_CICH
Vernonia_subplumosa_CICH
Moquinia_racemosa_CICH
Erato_polymnioides_CICH
Heterolepis_aliena_CICH
Heterolepis_aliena_2_CICH
Gymnarrhena_micrantha_GYMA
Inula_shirensis_ASTE
Chrysophtalmum_gueneri_ASTE
Inula_paniculata_ASTE
Inula_peacockiana_ASTE
Carpesium_divaricatum_ASTE
Blumea_aromatica_ASTE
Telekia_speciosa_ASTE
Caesulia_axillaris_ASTE
Allagopappus_canariensis_ASTE
Rhanterium_suaveolens_ASTE
Duhaldea_cappa_ASTE
Buphthalmum_salicifolium_ASTE
Eupatorium_cannabinum_ASTE
Helianthus_tuberosus_ASTE
Polymnia_canadensis_ASTE
Perityle_emoryi_ASTE
Eupatorium_serotinum_ASTE
Endocellion_sibiricum_ASTE
Homogyne_alpina_ASTE
Urostemon_kirkii_CICH
Corymbium_glabrum_CORY
Gerbera_crocea_MUTI
Ainsliaea_apiculata_PERT
Ainsliaea_macrocephala_PERT
Pertya_scandens_PERT
Saussurea_velutina_CARD
Saussurea_muliensis_CARD
Saussurea_erubescens_2_CARD
Saussurea_erubescens_CARD
Atractylodes_koreana_CARD
Saussurea_iodostegia_CARD
Saussurea_tangutica_CARD
Carthamus_arborescens_CARD
Brachylaena_elliptica_CARD
Oldenburgia_grandis_CARD
Saussurea_hookeri_CARD
Atractylodes_carlinoides_CARD
Dolichlasium_lagascae_MUTI
Adenocaulon_bicolor_CICH
Gypothamnium_pinifolium_MUTI
Hecastocleis_shockleyi_HECA
Richterago_angustifolia_GOCH
Richterago_amplexifolia_GOCH
Gochnatia_hypoleuca_GOCH
Stifftia_fruticosa_STIF
Stifftia_parviflora_STIF
Stomatachaeta_condensata_WUND
Wunderlichia_mirabilis_WUND
Dasyphyllum_reticulatum_BARN
Barnadesia_caryophylla_BARN
0.007

Figure S2c

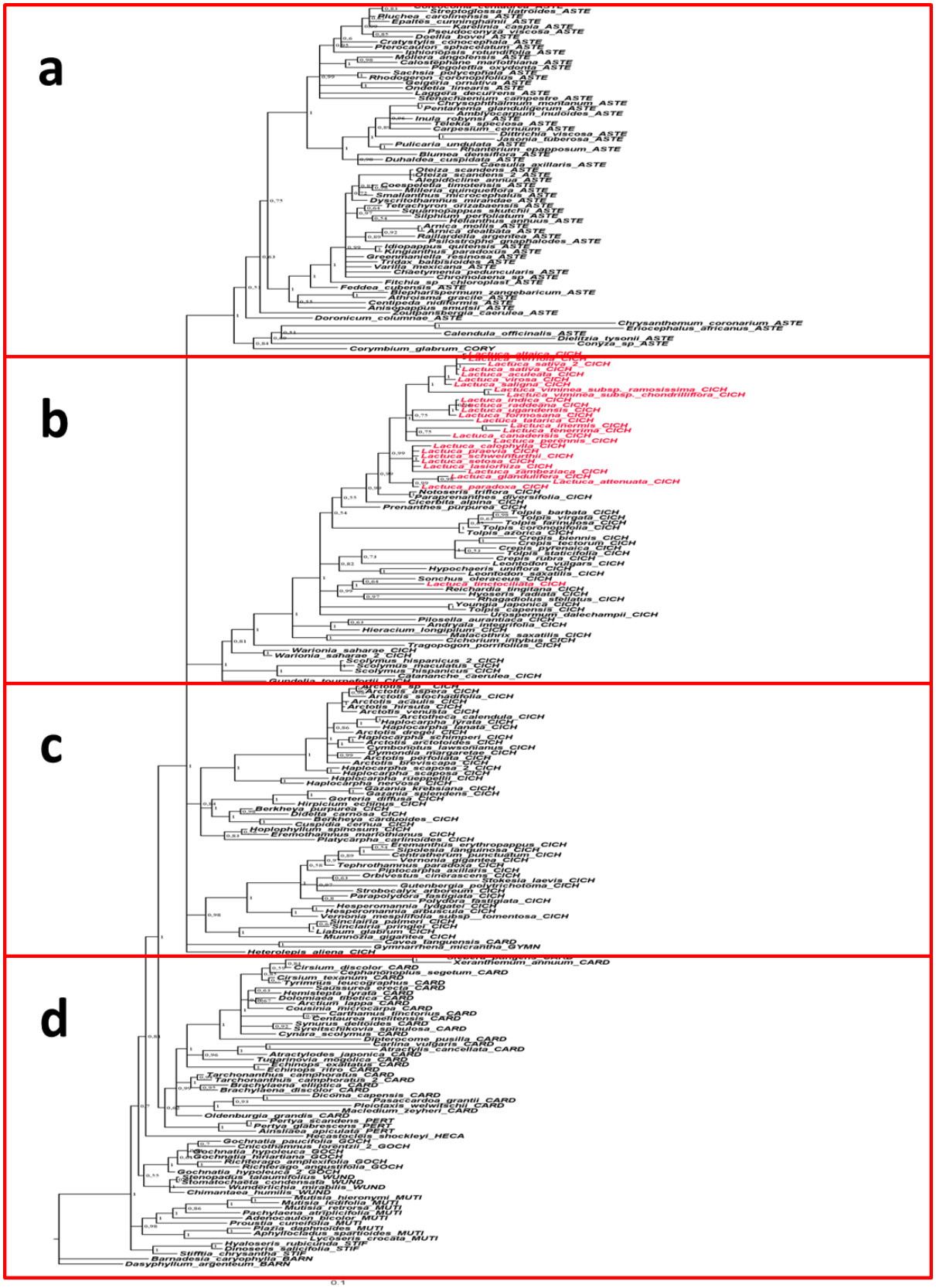


Figure S3 Bayesian phylogram (50% majority rule consensus tree) of 247 accessions from different sub-families of Asteraceae, based on *ndhF*; *Lactuca* species are shown in red colour; posterior probability (PP) supporting values are given; *L. tinctociliata* was mis-identified and it could be *Launaea cornuta*; *L. ugandensis* should be *Lactuca* sp.

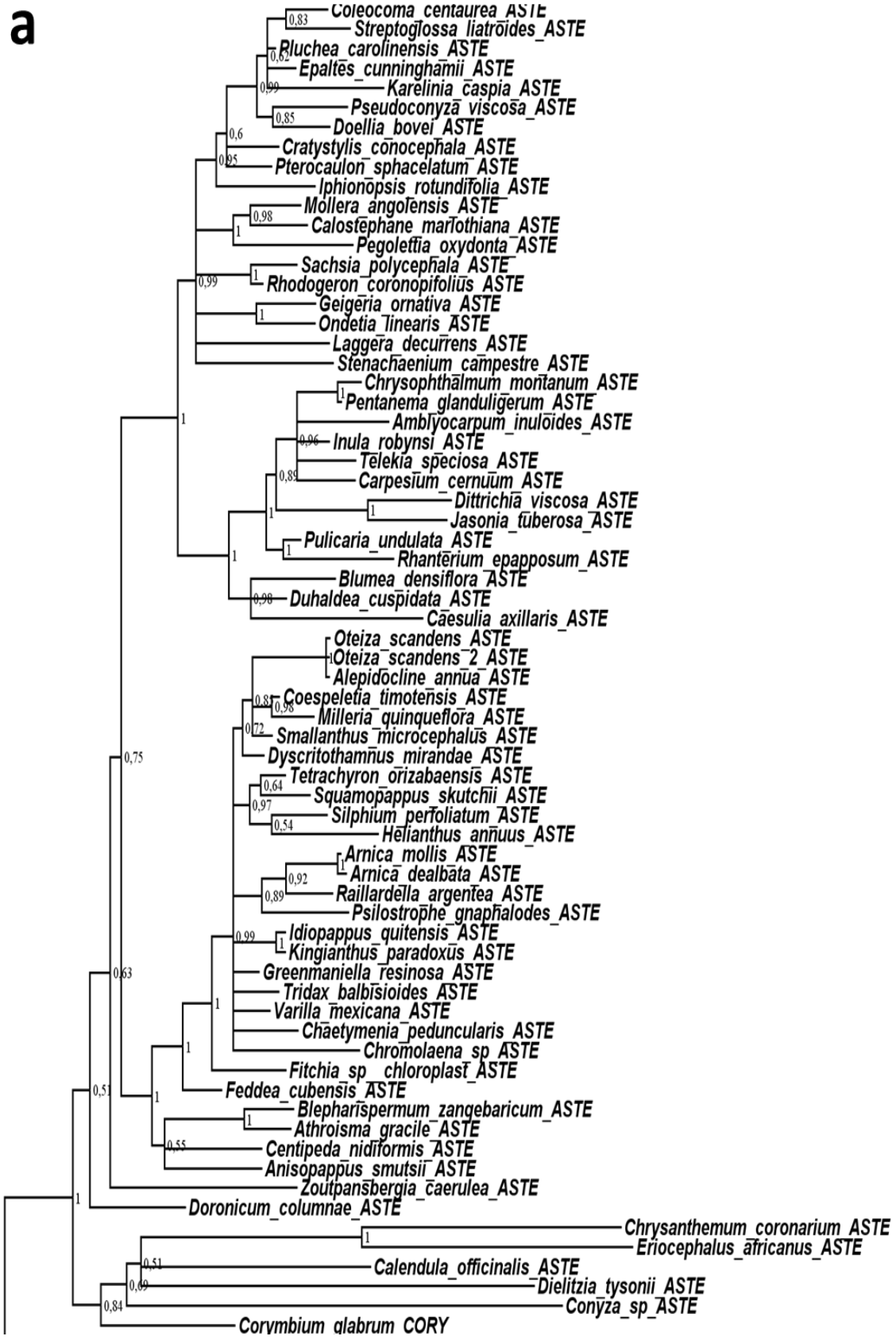

a
Coleocoma_centaurea_ASTE
Streptoglossa_liatroides_ASTE
Pluchea_carolinensis_ASTE
Epaltes_cunninghamii_ASTE
Karelinia_caspia_ASTE
Pseudoconyza_viscosa_ASTE
Doellia_bovei_ASTE
Cratystylis_conocephala_ASTE
Pterocaulon_sphacelatum_ASTE
Iphionopsis_rotundifolia_ASTE
Mollera_angolensis_ASTE
Calostephane_marlothiana_ASTE
Pegolettia_oxydonta_ASTE
Sachsia_polycephala_ASTE
Rhodogeron_coronopifolius_ASTE
Geigeria_ornativa_ASTE
Ondetia_linearis_ASTE
Laggera_decurrens_ASTE
Stenachaenium_campestre_ASTE
Chrysophthalmum_montanum_ASTE
Pentanema_glanduligerum_ASTE
Amblyocarpum_inuloides_ASTE
Inula_robynsi_ASTE
Telekia_speciosa_ASTE
Carpesium_cernuum_ASTE
Dittrichia_viscosa_ASTE
Jasonia_tuberosa_ASTE
Pulicaria_undulata_ASTE
Rhanterium_epapposum_ASTE
Blumea_densiflora_ASTE
Duhaldea_cuspidata_ASTE
Caesulia_axillaris_ASTE
Oteiza_scandens_ASTE
Oteiza_scandens_2_ASTE
Alepidocline_annua_ASTE
Coespeletia_timotensis_ASTE
Milleria_quinqueflora_ASTE
Smallanthus_microcephalus_ASTE
Dyscritothamnus_mirandae_ASTE
Tetrachyron_orizabaensis_ASTE
Squamopappus_skutchii_ASTE
Silphium_perfoliatum_ASTE
Helianthus_annuus_ASTE
Arnica_mollis_ASTE
Arnica_dealbata_ASTE
Raillardella_argentea_ASTE
Psilostrophe_gnaphalodes_ASTE
Idiopappus_quitensis_ASTE
Kingianthus_paradoxus_ASTE
Greenmaniella_resinosa_ASTE
Tridax_balbisioides_ASTE
Varilla_mexicana_ASTE
Chaetymenia_peduncularis_ASTE
Chromolaena_sp_ASTE
Fitchia_sp__chloroplast_ASTE
Feddea_cubensis_ASTE
Blepharispermum_zangebaricum_ASTE
Athroisma_gracile_ASTE
Centipeda_nidiformis_ASTE
Anisopappus_smutsii_ASTE
Zoutpansbergia_caerulea_ASTE
Doronicum_columnae_ASTE
Chrysanthemum_coronarium_ASTE
Eriocephalus_africanus_ASTE
Calendula_officinalis_ASTE
Dielitzia_tysonii_ASTE
Conyza_sp_ASTE
Corymbium_glabrum_CORY

Figure S3a

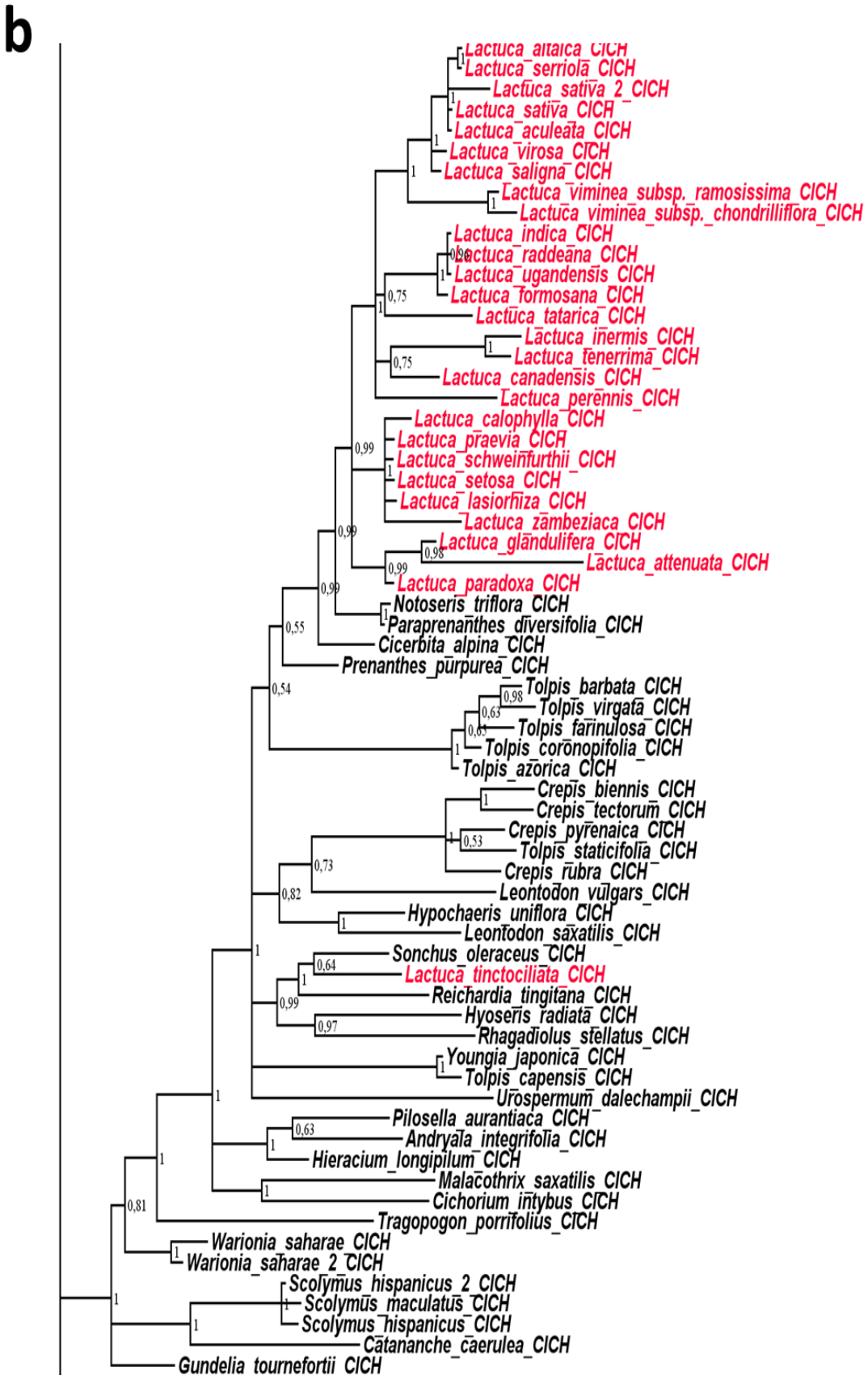

b
Lactuca_altaica_ClCH
Lactuca_serriola_ClCH
Lactuca_sativa_2_ClCH
Lactuca_sativa_ClCH
Lactuca_aculeata_ClCH
Lactuca_virosa_ClCH
Lactuca_saligna_ClCH
Lactuca_viminea_subsp._ramosissima_ClCH
Lactuca_viminea_subsp._chondrilliflora_ClCH
Lactuca_indica_ClCH
Lactuca_raddeana_ClCH
Lactuca_ugandensis_ClCH
Lactuca_formosana_ClCH
Lactuca_tatarica_ClCH
Lactuca_inermis_ClCH
Lactuca_tenerrima_ClCH
Lactuca_canadensis_ClCH
Lactuca_perennis_ClCH
Lactuca_calophylla_ClCH
Lactuca_praevia_ClCH
Lactuca_schweinfurthii_ClCH
Lactuca_setosa_ClCH
Lactuca_lasiorhiza_ClCH
Lactuca_zambeziaca_ClCH
Lactuca_glandulifera_ClCH
Lactuca_attenuata_ClCH
Lactuca_paradoxa_ClCH
Notoseris_triflora_ClCH
Paraprenanthes_diversifolia_ClCH
Cicerbita_alpina_ClCH
Prenanthes_purpurea_ClCH
Tolpis_barbata_ClCH
Tolpis_virgata_ClCH
Tolpis_farinulosa_ClCH
Tolpis_coronopifolia_ClCH
Tolpis_azorica_ClCH
Crepis_biennis_ClCH
Crepis_tectorum_ClCH
Crepis_pyrenaica_ClCH
Tolpis_staticifolia_ClCH
Crepis_rubra_ClCH
Leontodon_vulgars_ClCH
Hypochaeris_uniflora_ClCH
Leontodon_saxatilis_ClCH
Sonchus_oleraceus_ClCH
Lactuca_tinctociliata_ClCH
Reichardia_tingitana_ClCH
Hyoseris_radiata_ClCH
Rhagadiolus_stellatus_ClCH
Youngia_japonica_ClCH
Tolpis_capensis_ClCH
Urospermum_dalechampii_ClCH
Pilosella_aurantiaca_ClCH
Andryala_integrifolia_ClCH
Hieracium_longipilum_ClCH
Malacothrix_saxatilis_ClCH
Cichorium_intybus_ClCH
Tragopogon_porrifolius_ClCH
Warionia_saharae_ClCH
Warionia_saharae_2_ClCH
Scolymus_hispanicus_2_ClCH
Scolymus_maculatus_ClCH
Scolymus_hispanicus_ClCH
Catananche_caerulea_ClCH
Gundelia_tournefortii_ClCH

Figure S3b

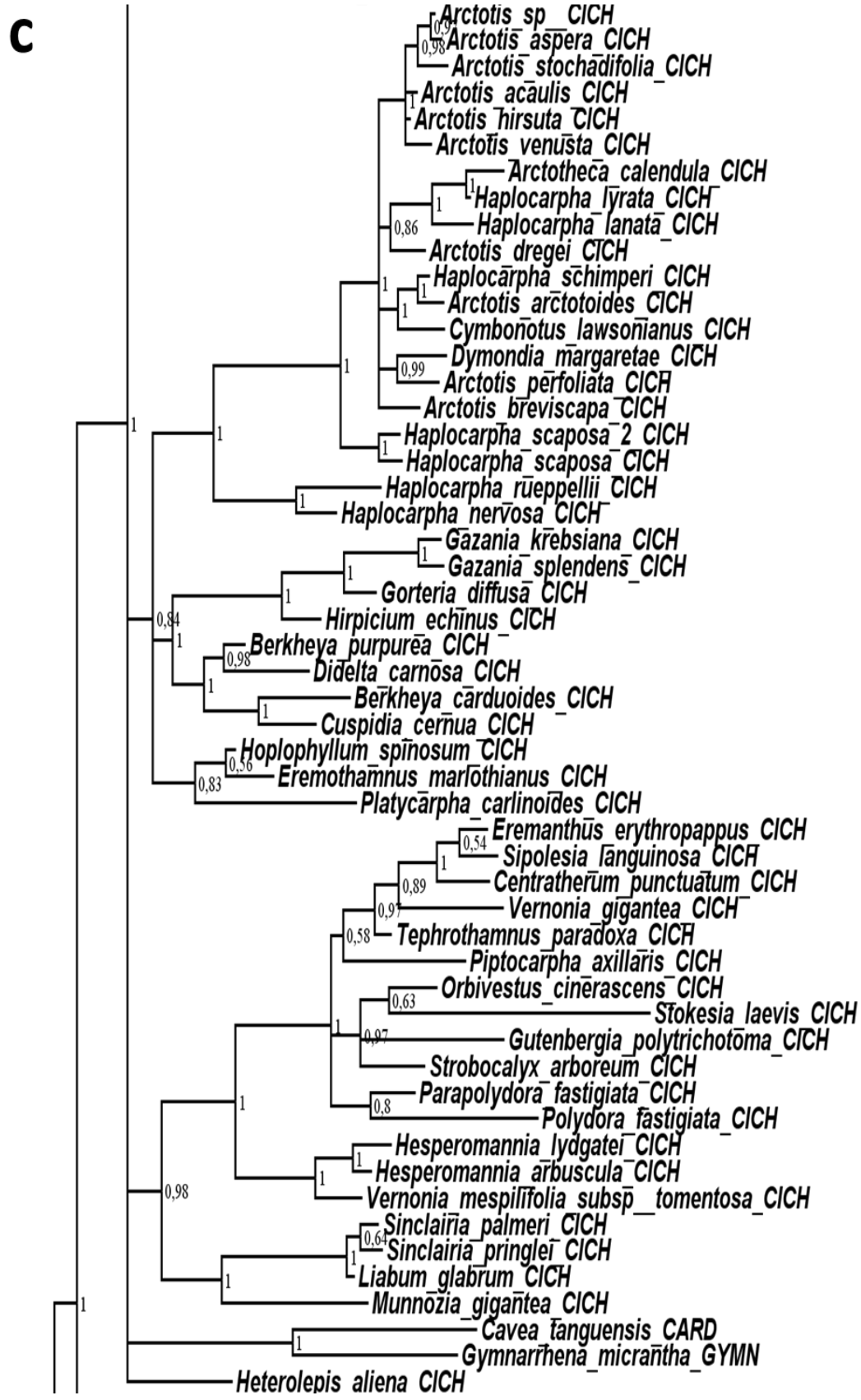

C
Arctotis_sp__ClCH
Arctotis_aspera_ClCH
Arctotis_stochadifolia_ClCH
Arctotis_acaulis_ClCH
Arctotis_hirsuta_ClCH
Arctotis_venusta_ClCH
Arctotheca_calendula_ClCH
Haplocarpha_lyrata_ClCH
Haplocarpha_lanata_ClCH
Arctotis_dregei_ClCH
Haplocarpha_schimperi_ClCH
Arctotis_arctotoides_ClCH
Cymbonotus_lawsonianus_ClCH
Dymondia_margaretae_ClCH
Arctotis_perfoliata_ClCH
Arctotis_breviscapa_ClCH
Haplocarpha_scaposa_2_ClCH
Haplocarpha_scaposa_ClCH
Haplocarpha_rueppellii_ClCH
Haplocarpha_nervosa_ClCH
Gazania_krebsiana_ClCH
Gazania_splendens_ClCH
Gorteria_diffusa_ClCH
Hirpicium_echinus_ClCH
Berkheya_purpurea_ClCH
Didelta_carnosa_ClCH
Berkheya_carduoides_ClCH
Cuspidia_cernua_ClCH
Hoplophyllum_spinosum_ClCH
Eremothamnus_marlothianus_ClCH
Platycarpha_carlinoides_ClCH
Eremanthus_erythropappus_ClCH
Sipolesia_languinosa_ClCH
Centratherum_punctuatum_ClCH
Vernonia_gigantea_ClCH
Tephrothamnus_paradoxa_ClCH
Piptocarpha_axillaris_ClCH
Orbivestus_cinerascens_ClCH
Stokesia_laevis_ClCH
Gutenbergia_polytrichotoma_ClCH
Strobocalyx_arboreum_ClCH
Parapolydora_fastigiata_ClCH
Polydora_fastigiata_ClCH
Hesperomannia_lydgatei_ClCH
Hesperomannia_arbuscula_ClCH
Vernonia_mespilifolia_subsp__tomentosa_ClCH
Sinclairia_palmeri_ClCH
Sinclairia_pringlei_ClCH
Liabum_glabrum_ClCH
Munnozia_gigantea_ClCH
Cavea_tanguensis_CARD
Gymnarrhena_micrantha_GYMN
Heterolepis_aliena_ClCH

Figure S3c

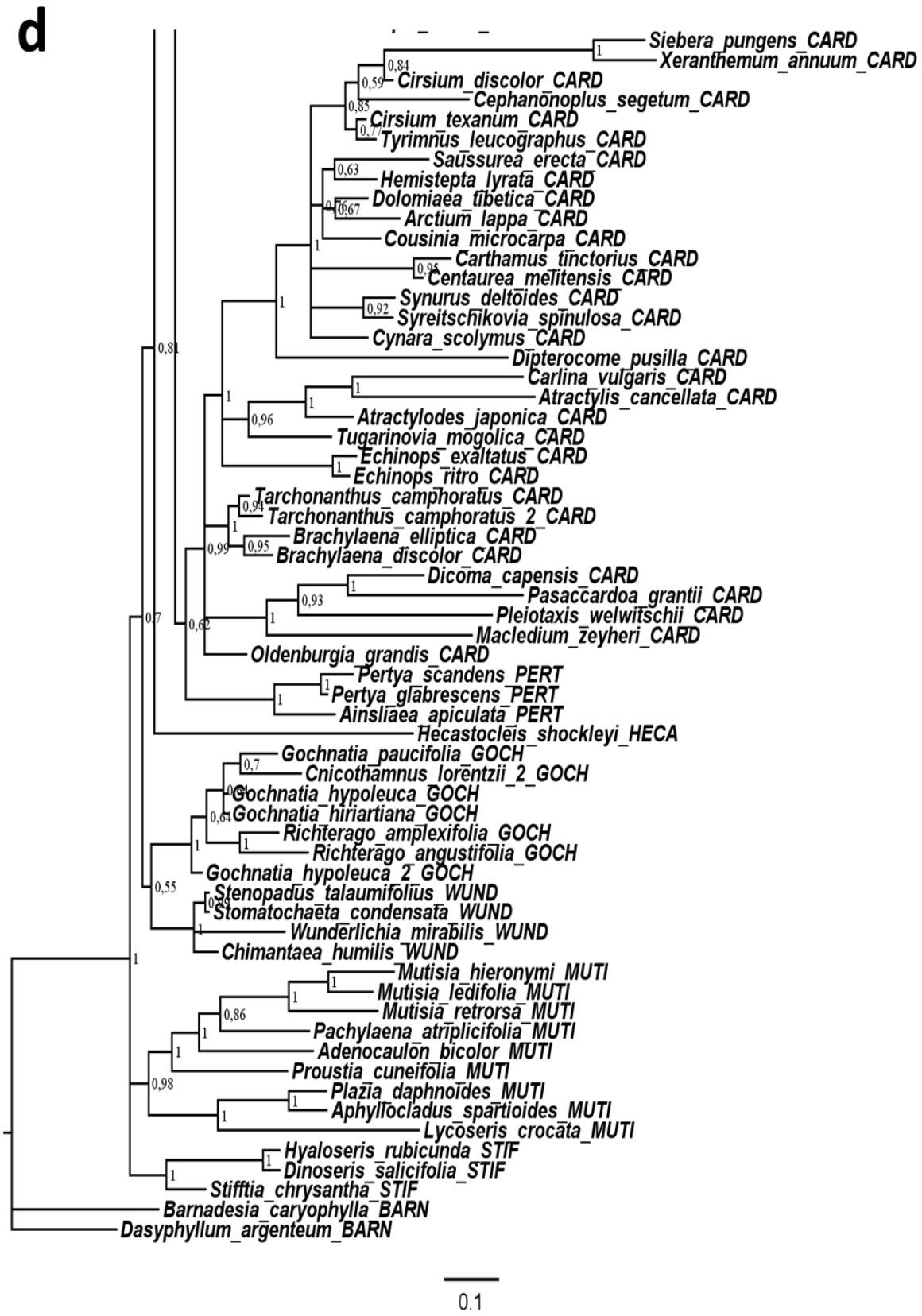

d
Siebera_pungens_CARD
Xeranthemum_annuum_CARD
Cirsium_discolor_CARD
Cephanonoplus_segetum_CARD
Cirsium_texanum_CARD
Tyrimnus_leucographus_CARD
Saussurea_erecta_CARD
Hemistepta_lyrata_CARD
Dolomiaea_tibetica_CARD
Arctium_lappa_CARD
Cousinia_microcarpa_CARD
Carthamus_tinctorius_CARD
Centaurea_melitensis_CARD
Synurus_deltoides_CARD
Syreitschikovia_spinulosa_CARD
Cynara_scolymus_CARD
Dipterocome_pusilla_CARD
Carlina_vulgaris_CARD
Atractylis_cancellata_CARD
Atractylodes_japonica_CARD
Tugarinovia_mongolica_CARD
Echinops_exaltatus_CARD
Echinops_ritro_CARD
Tarchonanthus_camphoratus_CARD
Tarchonanthus_camphoratus_2_CARD
Brachylaena_elliptica_CARD
Brachylaena_discolor_CARD
Dicoma_capensis_CARD
Pasaccardoa_grantii_CARD
Pleiotaxis_welwitschii_CARD
Macledium_zeyheri_CARD
Oldenburgia_grandis_CARD
Pertya_scandens_PERT
Pertya_glabrescens_PERT
Ainsliaea_apiculata_PERT
Hecastocleis_shockleyi_HECA
Gochnatia_paucifolia_GOCH
Cnicothamnus_lorentzii_2_GOCH
Gochnatia_hypoleuca_GOCH
Gochnatia_hiriartiana_GOCH
Richterago_amplexifolia_GOCH
Richterago_angustifolia_GOCH
Gochnatia_hypoleuca_2_GOCH
Stenopadus_talaumifolius_WUND
Stomatochaeta_condensata_WUND
Wunderlichia_mirabilis_WUND
Chimantaea_humilis_WUND
Mutisia_hieronymi_MUTI
Mutisia_ledifolia_MUTI
Mutisia_retrorsa_MUTI
Pachylaena_atriplicifolia_MUTI
Adenocaulon_bicolor_MUTI
Proustia_cuneifolia_MUTI
Plazia_daphnoides_MUTI
Aphyllocladus_spartioides_MUTI
Lycoseris_crocata_MUTI
Hyaloseris_rubicunda_STIF
Dinoseris_salicifolia_STIF
Stifftia_chrysantha_STIF
Barnadesia_caryophylla_BARN
Dasyphyllum_argenteum_BARN
0.1

Figure S3d

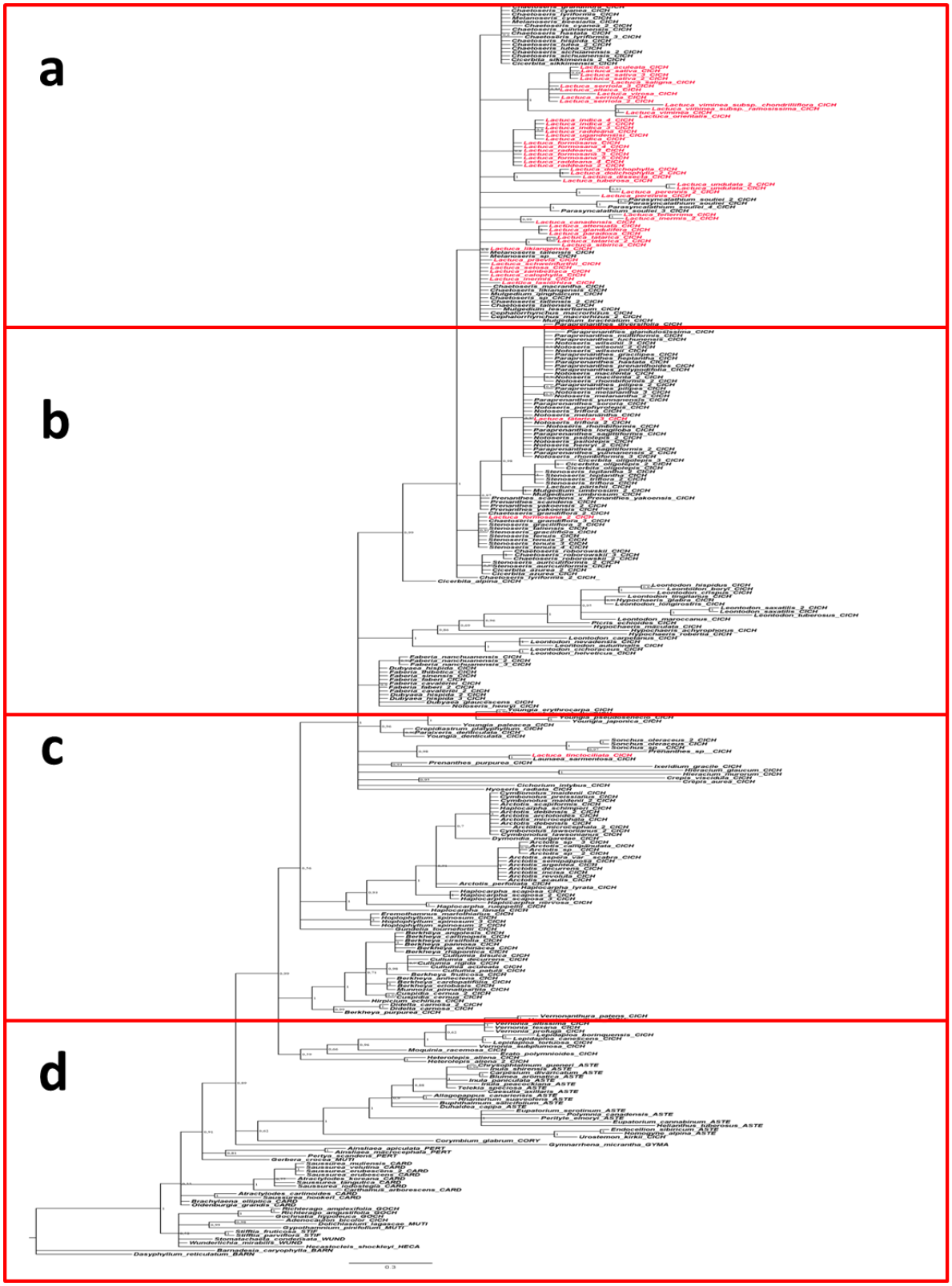


Figure S4 Bayesian phylogram (50% majority rule consensus tree) of 331 accessions from different sub-families of Asteraceae, based on *trnL-F*; *Lactuca* species are shown in red colour; posterior probability (PP) supporting values are given; *L. tinctociliata* was mis-identified and it could be *Launaea cornuta*; *L. ugandensis* should be *Lactuca* sp.

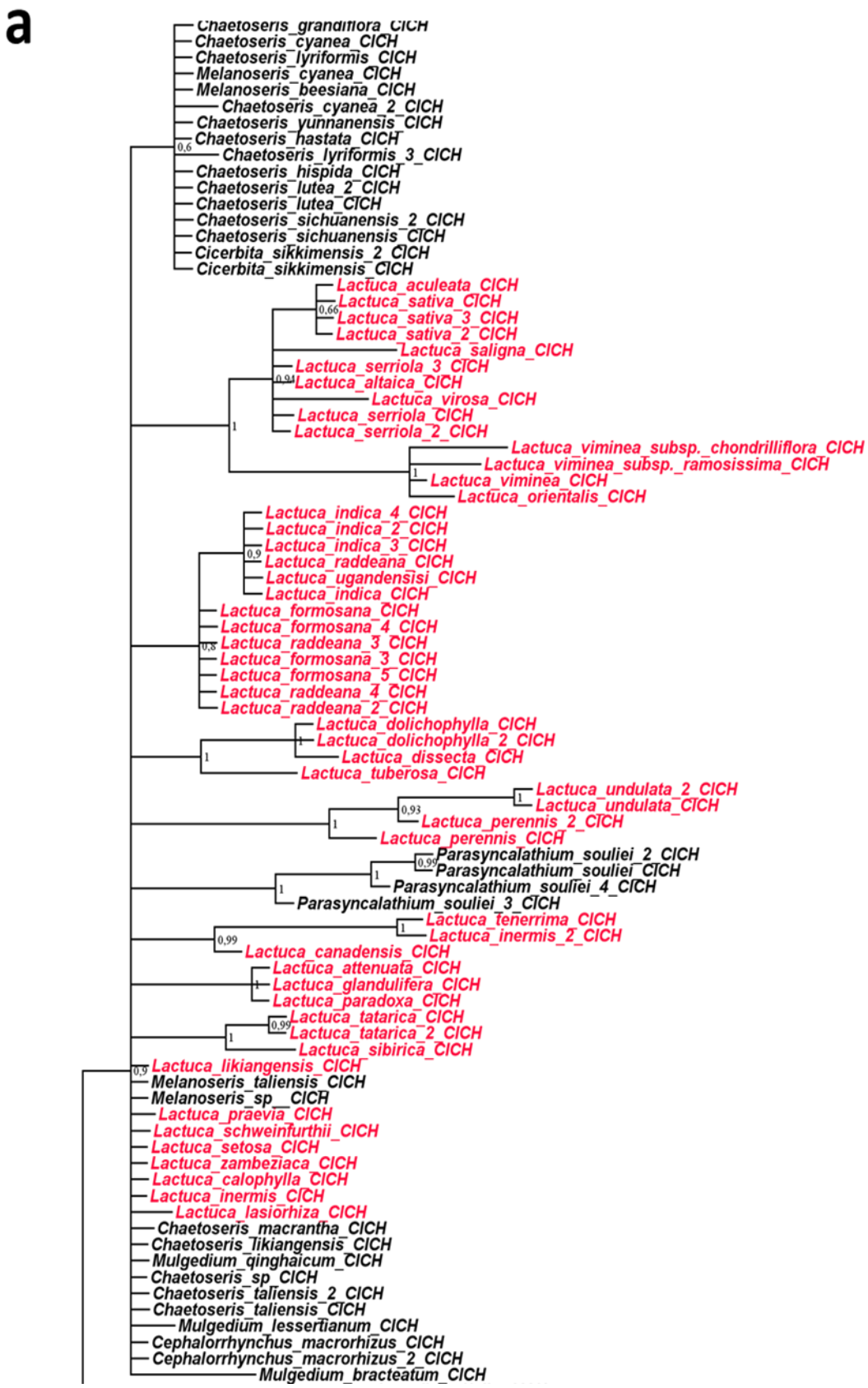

a
Chaetoseris_grandiflora_CICH
Chaetoseris_cyanea_CICH
Chaetoseris_lyriformis_CICH
Melanoseris_cyanea_CICH
Melanoseris_beesiana_CICH
Chaetoseris_cyanea_2_CICH
Chaetoseris_yunnanensis_CICH
Chaetoseris_hastata_CICH
Chaetoseris_lyriformis_3_CICH
Chaetoseris_hispida_CICH
Chaetoseris_lutea_2_CICH
Chaetoseris_lutea_CICH
Chaetoseris_sichuanensis_2_CICH
Chaetoseris_sichuanensis_CICH
Cicerbita_sikkimensis_2_CICH
Cicerbita_sikkimensis_CICH
Lactuca_aculeata_CICH
Lactuca_sativa_CICH
Lactuca_sativa_3_CICH
Lactuca_sativa_2_CICH
Lactuca_saligna_CICH
Lactuca_serriola_3_CICH
Lactuca_altaica_CICH
Lactuca_virosa_CICH
Lactuca_serriola_CICH
Lactuca_serriola_2_CICH
Lactuca_viminea_subsp._chondrilliflora_CICH
Lactuca_viminea_subsp._ramosissima_CICH
Lactuca_viminea_CICH
Lactuca_orientalis_CICH
Lactuca_indica_4_CICH
Lactuca_indica_2_CICH
Lactuca_indica_3_CICH
Lactuca_raddeana_CICH
Lactuca_ugandensisi_CICH
Lactuca_indica_CICH
Lactuca_formosana_CICH
Lactuca_formosana_4_CICH
Lactuca_raddeana_3_CICH
Lactuca_formosana_3_CICH
Lactuca_formosana_5_CICH
Lactuca_raddeana_4_CICH
Lactuca_raddeana_2_CICH
Lactuca_dolichophylla_CICH
Lactuca_dolichophylla_2_CICH
Lactuca_dissecta_CICH
Lactuca_tuberosa_CICH
Lactuca_undulata_2_CICH
Lactuca_undulata_CICH
Lactuca_perennis_2_CICH
Lactuca_perennis_CICH
Parasyncalathium_souliei_2_CICH
Parasyncalathium_souliei_CICH
Parasyncalathium_souliei_4_CICH
Parasyncalathium_souliei_3_CICH
Lactuca_tenerrima_CICH
Lactuca_inermis_2_CICH
Lactuca_canadensis_CICH
Lactuca_attenuata_CICH
Lactuca_glandulifera_CICH
Lactuca_paradoxa_CICH
Lactuca_tatarica_CICH
Lactuca_tatarica_2_CICH
Lactuca_sibirica_CICH
Lactuca_likiangensis_CICH
Melanoseris_taliensis_CICH
Melanoseris_sp__CICH
Lactuca_praevia_CICH
Lactuca_schweinfurthii_CICH
Lactuca_setosa_CICH
Lactuca_zambeziaca_CICH
Lactuca_calophylla_CICH
Lactuca_inermis_CICH
Lactuca_lasiorhiza_CICH
Chaetoseris_macrantha_CICH
Chaetoseris_likiangensis_CICH
Mulgedium_qinghaicum_CICH
Chaetoseris_sp_CICH
Chaetoseris_taliensis_2_CICH
Chaetoseris_taliensis_CICH
Mulgedium_lessertianum_CICH
Cephalorrhynchus_macrorhizus_CICH
Cephalorrhynchus_macrorhizus_2_CICH
Mulgedium_bracteatum_CICH

Figure S4a


b
Paraprenanthes_diversifolia_CICH
Paraprenanthes_diversifolia2_CICH
Paraprenanthes_glandulosissima_CICH
Paraprenanthes_multiformis_CICH
Paraprenanthes_luchunensis_CICH
Notoseris_wilsonii_3_CICH
Notoseris_wilsonii_2_CICH
Notoseris_wilsonii_CICH
Paraprenanthes_gracilipes_CICH
Paraprenanthes_heptantha_CICH
Paraprenanthes_hastata_CICH
Paraprenanthes_prenanthoides_CICH
Paraprenanthes_polypodifolia_CICH
Notoseris_macilenta_CICH
Notoseris_macilenta_2_CICH
Notoseris_rhombiformis_2_CICH
Paraprenanthes_pilipes_2_CICH
Paraprenanthes_pilipes_CICH
Notoseris_melanantha_3_CICH
Notoseris_melanantha_2_CICH
Paraprenanthes_yunnanensis_CICH
Paraprenanthes_sororia_CICH
Notoseris_porphyrolepis_CICH
Notoseris_triflora_CICH
Notoseris_melanantha_CICH
Lactuca_tatarica_3_CICH
Notoseris_triflora_2_CICH
Notoseris_rhombiformis_CICH
Paraprenanthes_longiloba_CICH
Paraprenanthes_sagittiformis_CICH
Notoseris_psilolepis_2_CICH
Notoseris_psilolepis_CICH
Notoseris_henryi_2_CICH
Paraprenanthes_sagittiformis_2_CICH
Paraprenanthes_yunnanensis_2_CICH
Notoseris_rhombiformis_3_CICH
Cicerbita_oligolepis_3_CICH
Cicerbita_oligolepis_2_CICH
Cicerbita_oligolepis_CICH
Stenoseris_leptantha_2_CICH
Stenoseris_leptantha_CICH
Stenoseris_triflora_2_CICH
Stenoseris_triflora_CICH
Lactuca_parishii_CICH
Mulgedium_umbrosum_2_CICH
Mulgedium_umbrosum_CICH
Prenanthes_scandens_x_Prenanthes_yakoensis_CICH
Prenanthes_scandens_CICH
Prenanthes_yakoensis_2_CICH
Prenanthes_yakoensis_CICH
Chaetoseris_grandiflora_2_CICH
Lactuca_formosana_2_CICH
Chaetoseris_grandiflora_3_CICH
Stenoseris_graciliflora_2_CICH
Stenoseris_taliensis_CICH
Stenoseris_graciliflora_CICH
Stenoseris_tenuis_CICH
Stenoseris_tenuis_2_CICH
Stenoseris_tenuis_3_CICH
Stenoseris_tenuis_4_CICH
Chaetoseris_roborowskii_CICH
Chaetoseris_roborowskii_3_CICH
Chaetoseris_roborowskii_2_CICH
Stenoseris_auriculiformis_2_CICH
Stenoseris_auriculiformis_CICH
Cicerbita_azurea_2_CICH
Cicerbita_azurea_CICH
Chaetoseris_lyriformis_2_CICH
Cicerbita_alpina_CICH
Leontodon_hispidus_CICH
Leontodon_boryi_CICH
Leontodon_crispus_CICH
Leontodon_tingitanus_CICH
Hypochaeris_glabra_CICH
Leontodon_longirostris_CICH
Leontodon_saxatilis_2_CICH
Leontodon_saxatilis_CICH
Leontodon_tuberosus_CICH
Leontodon_maroccanus_CICH
Picris_echioides_CICH
Hypochaeris_maculata_CICH
Hypochaeris_achyrophorus_CICH
Hypochaeris_robertia_CICH
Leontodon_carpetanus_CICH
Leontodon_nevadensis_CICH
Leontodon_autumnalis_CICH
Leontodon_cichoraceus_CICH
Leontodon_helveticus_CICH
Faberia_nanchuanensis_CICH
Faberia_nanchuanensis_2_CICH
Faberia_nanchuanensis_3_CICH
Dubyaea_hispida_CICH
Faberia_thibetica_CICH
Faberia_sinensis_CICH
Faberia_faberi_CICH
Faberia_cavaleriei_CICH
Faberia_faberi_2_CICH
Faberia_cavaleriei_2_CICH
Dubyaea_hispida_2_CICH
Dubyaea_hispida_3_CICH
Dubyaea_glaucescens_CICH
Notoseris_henryi_CICH

Figure S4b

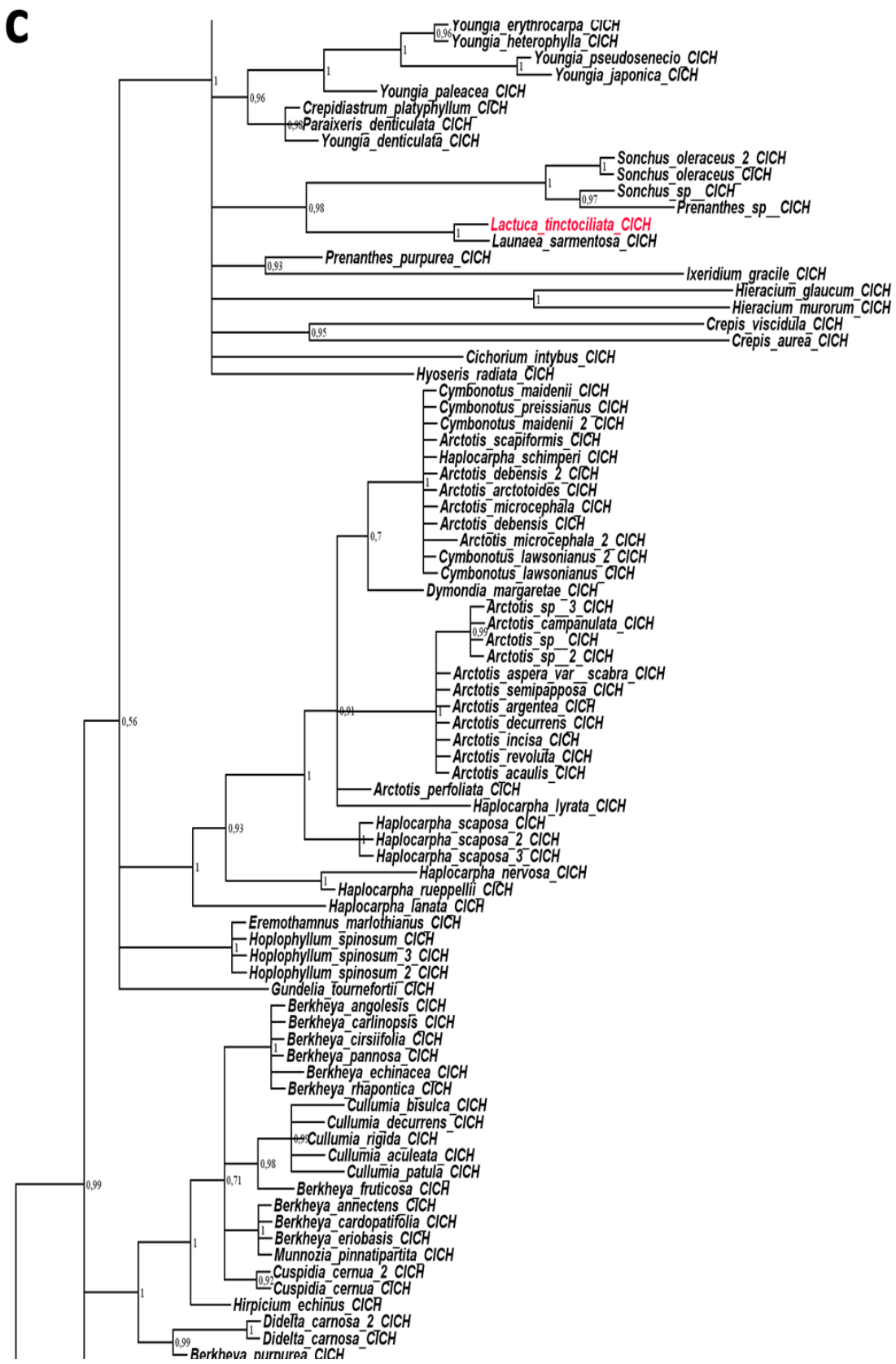

C
Youngia_erythrocarpa_CICH
Youngia_heterophylla_CICH
Youngia_pseudosenecio_CICH
Youngia_japonica_CICH
Youngia_paleacea_CICH
Crepidiastrum_platyphyllum_CICH
Paraixeris_denticulata_CICH
Youngia_denticulata_CICH
Sonchus_oleraceus_2_CICH
Sonchus_oleraceus_CICH
Sonchus_sp__CICH
Prenanthes_sp__CICH
Lactuca_tinctociliata_CICH
Launaea_sarmentosa_CICH
Prenanthes_purpurea_CICH
Ixeridium_gracile_CICH
Hieracium_glaucum_CICH
Hieracium_murorum_CICH
Crepis_viscidula_CICH
Crepis_aurea_CICH
Cichorium_intybus_CICH
Hyoseris_radiata_CICH
Cymbonotus_maidenii_CICH
Cymbonotus_preissianus_CICH
Cymbonotus_maidenii_2_CICH
Arctotis_scapiformis_CICH
Haplocarpha_schimperi_CICH
Arctotis_debensis_2_CICH
Arctotis_arctotoides_CICH
Arctotis_microcephala_CICH
Arctotis_debensis_CICH
Arctotis_microcephala_2_CICH
Cymbonotus_lawsonianus_2_CICH
Cymbonotus_lawsonianus_CICH
Dymondia_margaretae_CICH
Arctotis_sp__3_CICH
Arctotis_campanulata_CICH
Arctotis_sp__CICH
Arctotis_sp__2_CICH
Arctotis_aspera_var__scabra_CICH
Arctotis_semipapposa_CICH
Arctotis_argentea_CICH
Arctotis_decurrens_CICH
Arctotis_incisa_CICH
Arctotis_revoluta_CICH
Arctotis_acaulis_CICH
Arctotis_perfoliata_CICH
Haplocarpha_lyrata_CICH
Haplocarpha_scaposa_CICH
Haplocarpha_scaposa_2_CICH
Haplocarpha_scaposa_3_CICH
Haplocarpha_nervosa_CICH
Haplocarpha_rueppellii_CICH
Haplocarpha_lanata_CICH
Eremothamnus_marlothianus_CICH
Hoplophyllum_spinosum_CICH
Hoplophyllum_spinosum_3_CICH
Hoplophyllum_spinosum_2_CICH
Gundelia_tournefortii_CICH
Berkheya_angolensis_CICH
Berkheya_carlinopsis_CICH
Berkheya_cirsiifolia_CICH
Berkheya_pannosa_CICH
Berkheya_echinacea_CICH
Berkheya_rhapontica_CICH
Cullumia_bisulca_CICH
Cullumia_decurrens_CICH
Cullumia_rigida_CICH
Cullumia_aculeata_CICH
Cullumia_patula_CICH
Berkheya_fruticosa_CICH
Berkheya_annectens_CICH
Berkheya_cardopatifolia_CICH
Berkheya_eriobasis_CICH
Munnozia_pinnatipartita_CICH
Cuspidia_cernua_2_CICH
Cuspidia_cernua_CICH
Hirpicium_echinus_CICH
Didelta_carnosa_2_CICH
Didelta_carnosa_CICH
Berkheya_purpurea_CICH

Figure S4c


d
Vernonanthura_patens_CICH
Vernonia_brasiliana_CICH
Vernonia_altissima_CICH
Vernonia_texana_CICH
Vernonia_profuga_CICH
Lepidaploa_borinquensis_CICH
Lepidaploa_canescens_CICH
Lepidaploa_tortuosa_CICH
Vernonia_subplumosa_CICH
Moquinia_racemosa_CICH
Erato_polymnioides_CICH
Heterolepis_aliena_CICH
Heterolepis_aliena_2_CICH
Chrysophtalmum_gueneri_ASTE
Inula_shirensis_ASTE
Carpesium_divaricatum_ASTE
Blumea_aromatica_ASTE
Inula_paniculata_ASTE
Inula_peacockiana_ASTE
Telekia_speciosa_ASTE
Caesulia_axillaris_ASTE
Allagopappus_canariensis_ASTE
Rhanterium_suaveolens_ASTE
Buphthalmum_salicifolium_ASTE
Duhaldea_cappa_ASTE
Eupatorium_serotinum_ASTE
Polymnia_canadensis_ASTE
Perityle_emoryi_ASTE
Eupatorium_cannabinum_ASTE
Helianthus_tuberosus_ASTE
Endocellion_sibiricum_ASTE
Homogyne_alpina_ASTE
Urostemon_kirkii_CICH
Corymbium_glabrum_CORY
Gymnarrhena_micrantha_GYMA
Ainsliaea_apiculata_PERT
Ainsliaea_macrocephala_PERT
Pertya_scandens_PERT
Gerbera_crocea_MUTI
Saussurea_muliensis_CARD
Saussurea_velutina_CARD
Saussurea_erubescens_2_CARD
Saussurea_erubescens_CARD
Atractylodes_koreana_CARD
Saussurea_tangutica_CARD
Saussurea_iodostegia_CARD
Carthamus_arborescens_CARD
Atractylodes_carlinoides_CARD
Saussurea_hookeri_CARD
Brachylaena_elliptica_CARD
Oldenburgia_grandis_CARD
Richterago_amplexifolia_GOCH
Richterago_angustifolia_GOCH
Gochnatia_hypoleuca_GOCH
Adenocaulon_bicolor_CICH
Dolichlasium_lagascae_MUTI
Gypothamnium_pinifolium_MUTI
Stifftia_fruticosa_STIF
Stifftia_parviflora_STIF
Stomatachaeta_condensata_WUND
Wunderlichia_mirabilis_WUND
Hecastocleis_shockleyi_HECA
Barnadesia_caryophylla_BARN
Dasyphyllum_reticulatum_BARN
0.3

Figure S4d

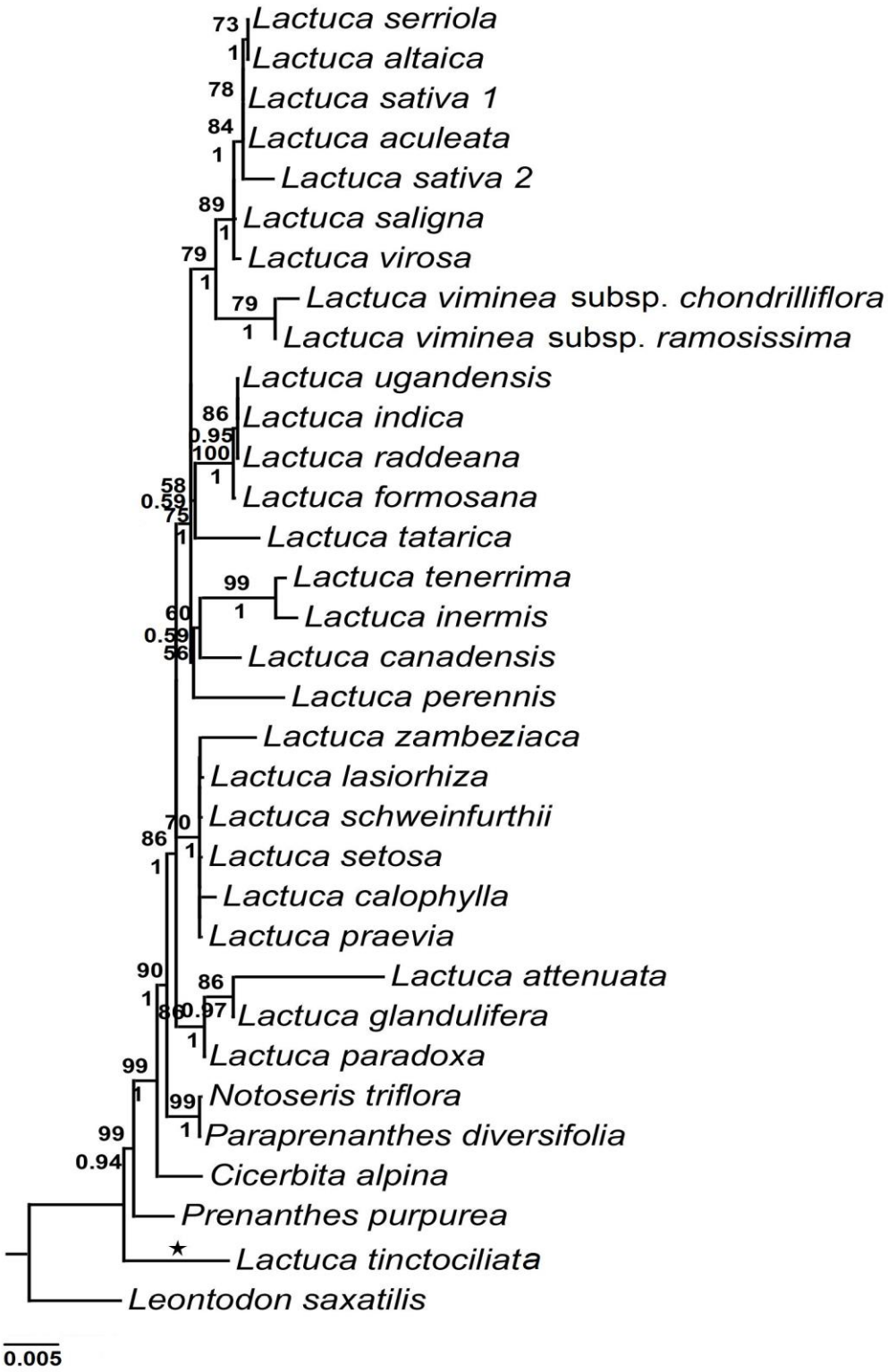


Figure S5 RAxML phylogram (best tree) of the genus *Lactuca* and related genera with a focus on African species, based on *ndhF* gene sequences; bootstrap (BS > 50) supporting values are given above the branches and posterior probability (PP > 0.5) supporting values are demonstrated below the branches; the names of Chinese taxa are referred to Wang et al. (2013); star *L. tinctociliata* was mis-identified and it could be *Launaea cornuta*; *L. ugandensis* should be *Lactuca* sp.

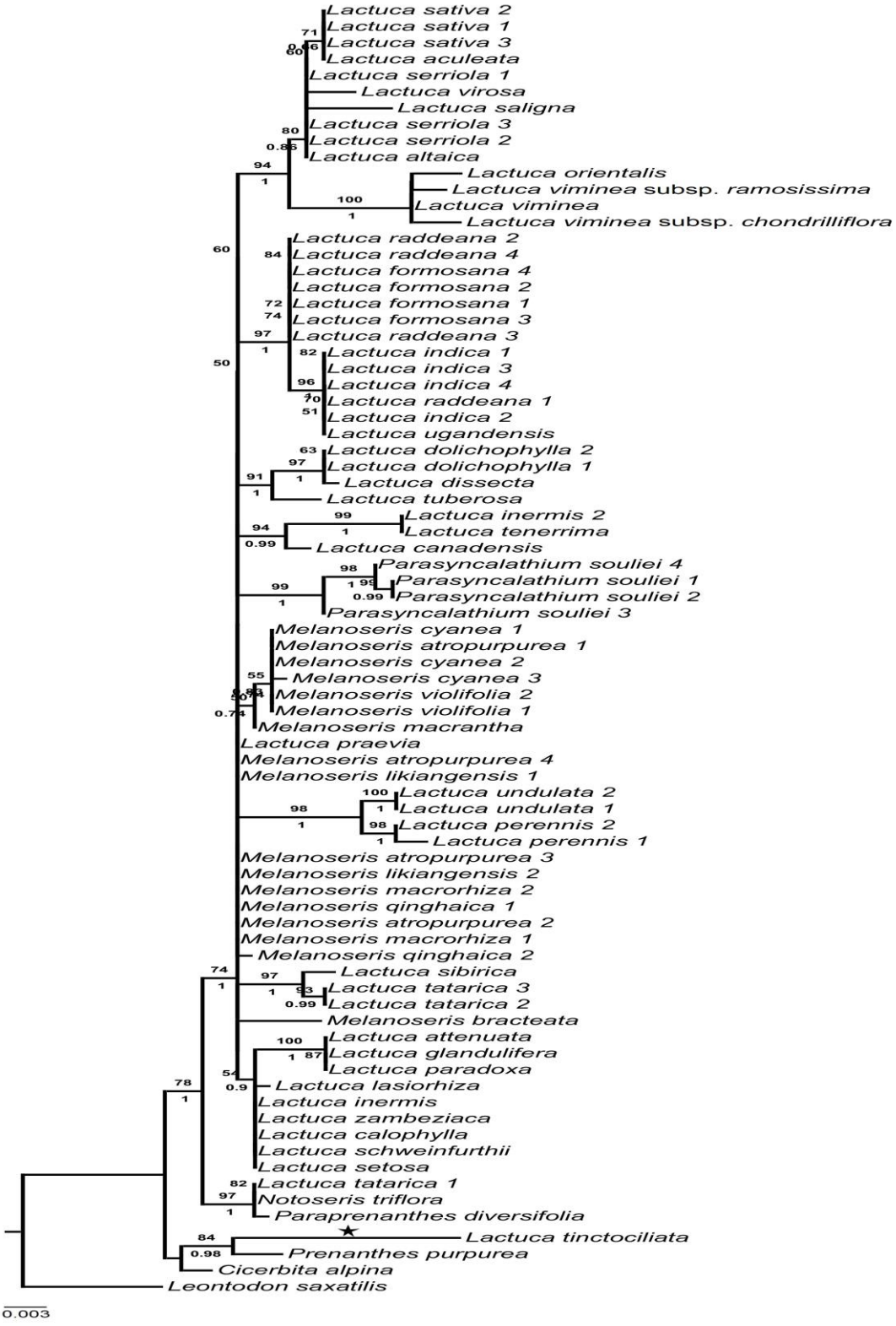


Figure S6 RAxML phylogram (best tree) of the genus *Lactuca* and related genera with a focus on African species, based on *trnL-F* gene sequences; bootstrap (BS > 50) supporting values are given above the branches and posterior probability (PP > 0.5) supporting values are demonstrated below the branches; the names of Chinese taxa are referred to Wang et al. (2013); star *L. tinctociliata* was mis-identified and it could be *Launaea cornuta*; *L. ugandensis* should be *Lactuca* sp.

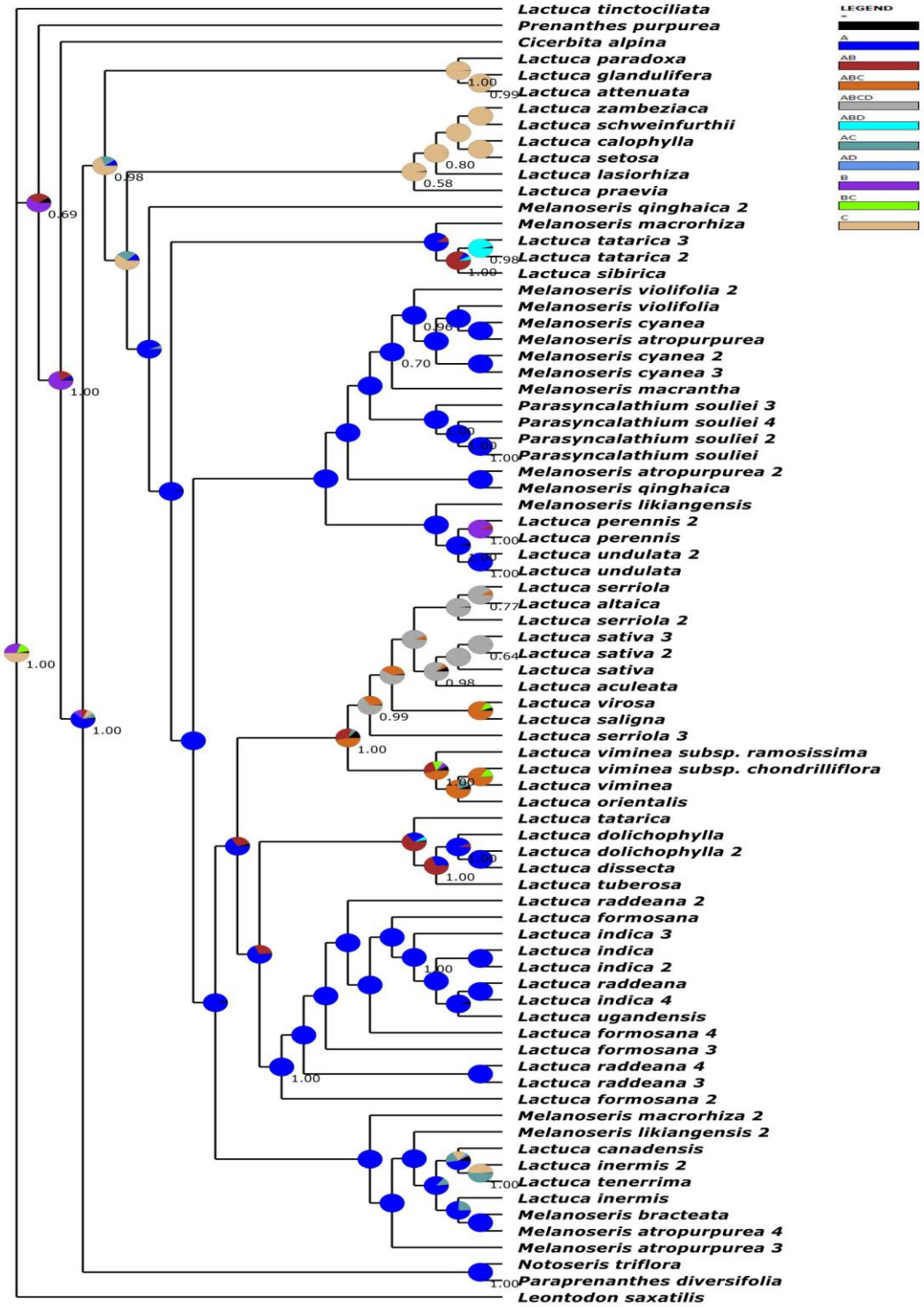


Figure S7 Ancestral stat biogeographic distribution based on Baysian consensus tree topology of the concatenated sequences of *ndhF* gene and *trnL-F* gene; pie charts on nodes show stat posterior probabilities; A Asia, B Europe, C Africa, D North America; *L. tinctociliata* was mis-identified and it could be *Launaea cornuta*; *L. ugandensis* should be *Lactuca* sp.

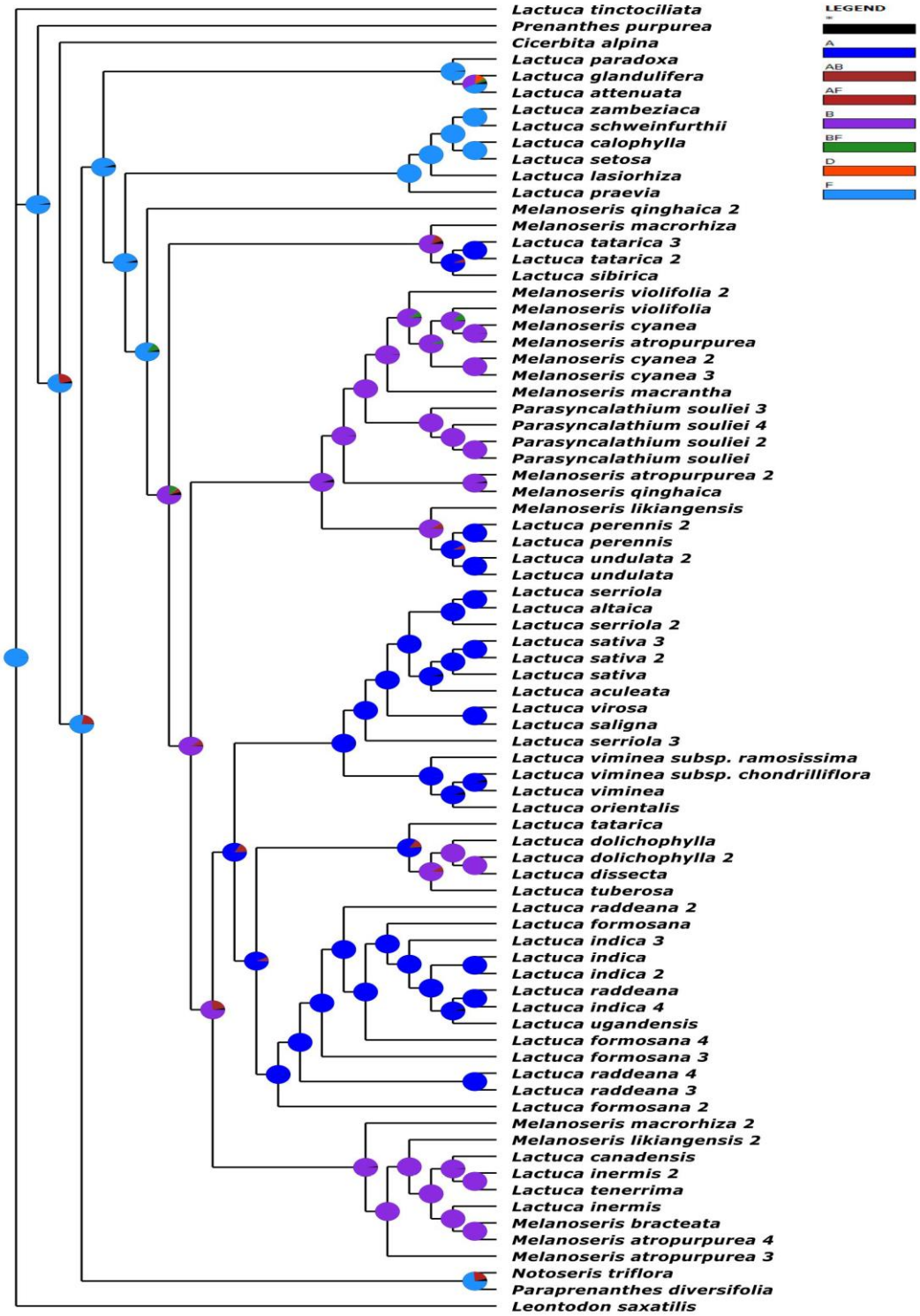


Figure S8 Ancestral stat chromosome number based on Baysian consensus tree topology of the concatenated sequences of *ndhF* gene and *trnL-F* gene; A 2n = 18, B 2n = 16, C 2n = 34, D 2n = 32, E 2n = 36, F missing data; *L. tinctociliata* was mis-identified and it could be *Launaea cornuta*; *L. ugandensis* should be *Lactuca* sp.

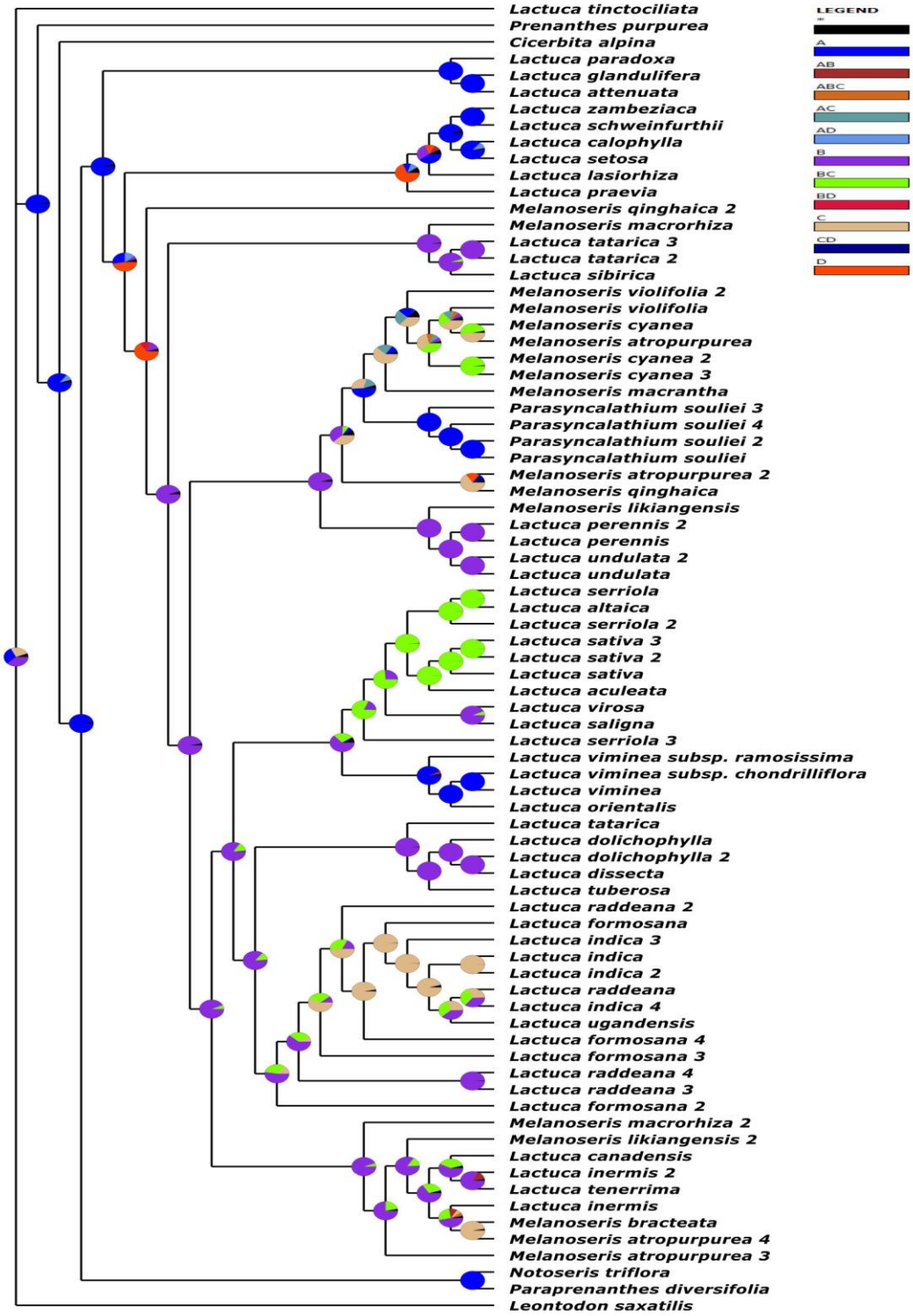


Figure S9 Ancestral stat floret number per capitula based on Baysian consensus tree topology of the concatenated sequences of *ndhF* gene and *trnL-F* gene; A < 6, B 6 - 15 (20), C > 20, D missing data; *L. tinctociliata* was mis-identified and it could be *Launaea cornuta*; *L. ugandensis* should be *Lactuca* sp.

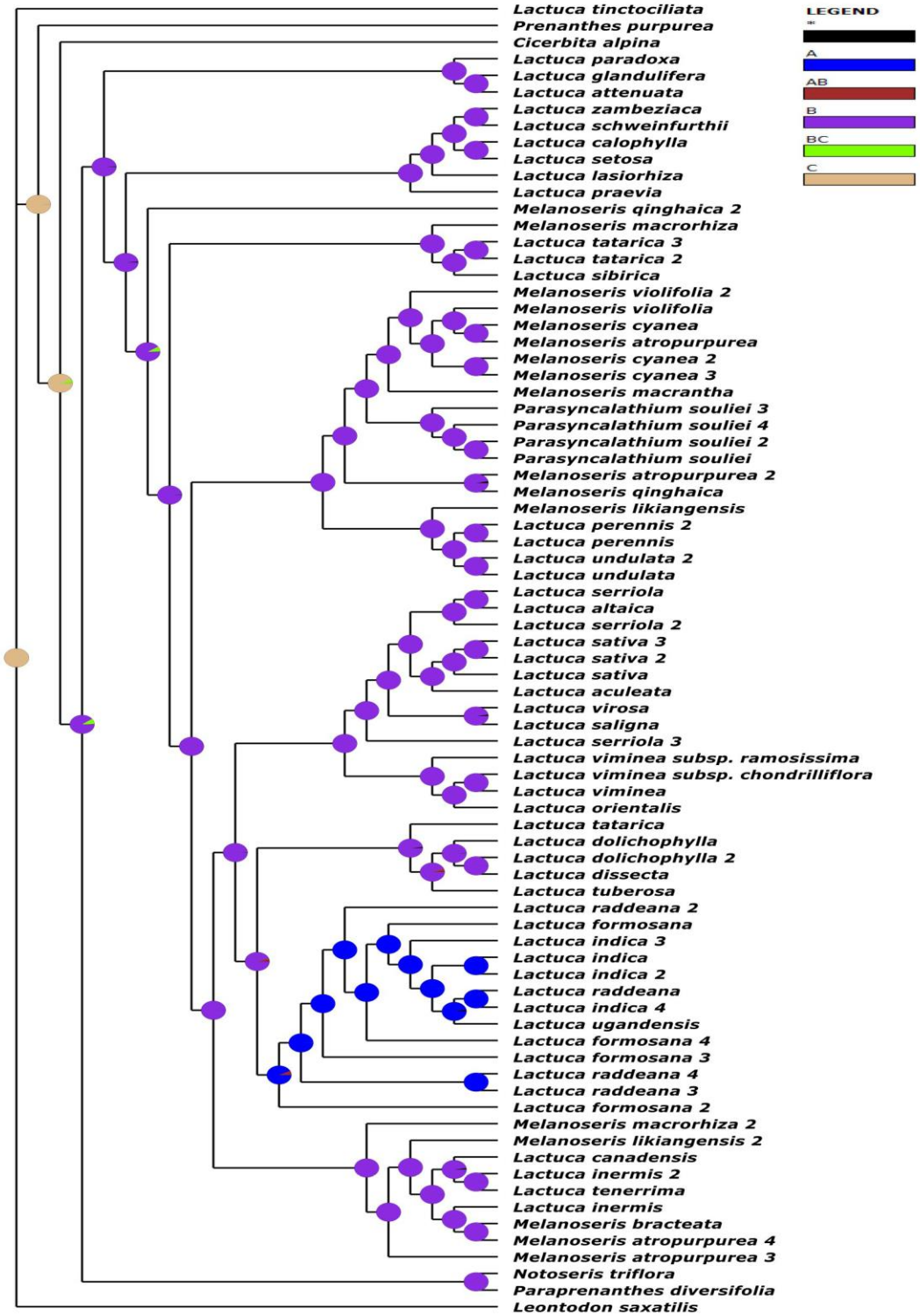


Figure S10 Ancestral stat winged achene or not based on Baysian consensus tree topology of the concatenated sequences of *ndhF* gene and *trnL-F* gene; A achene winged, B achene not winged, C missing data; *L. tinctociliata* was mis-identified and it could be *Launaea cornuta*; *L. ugandensis* should be *Lactuca* sp.

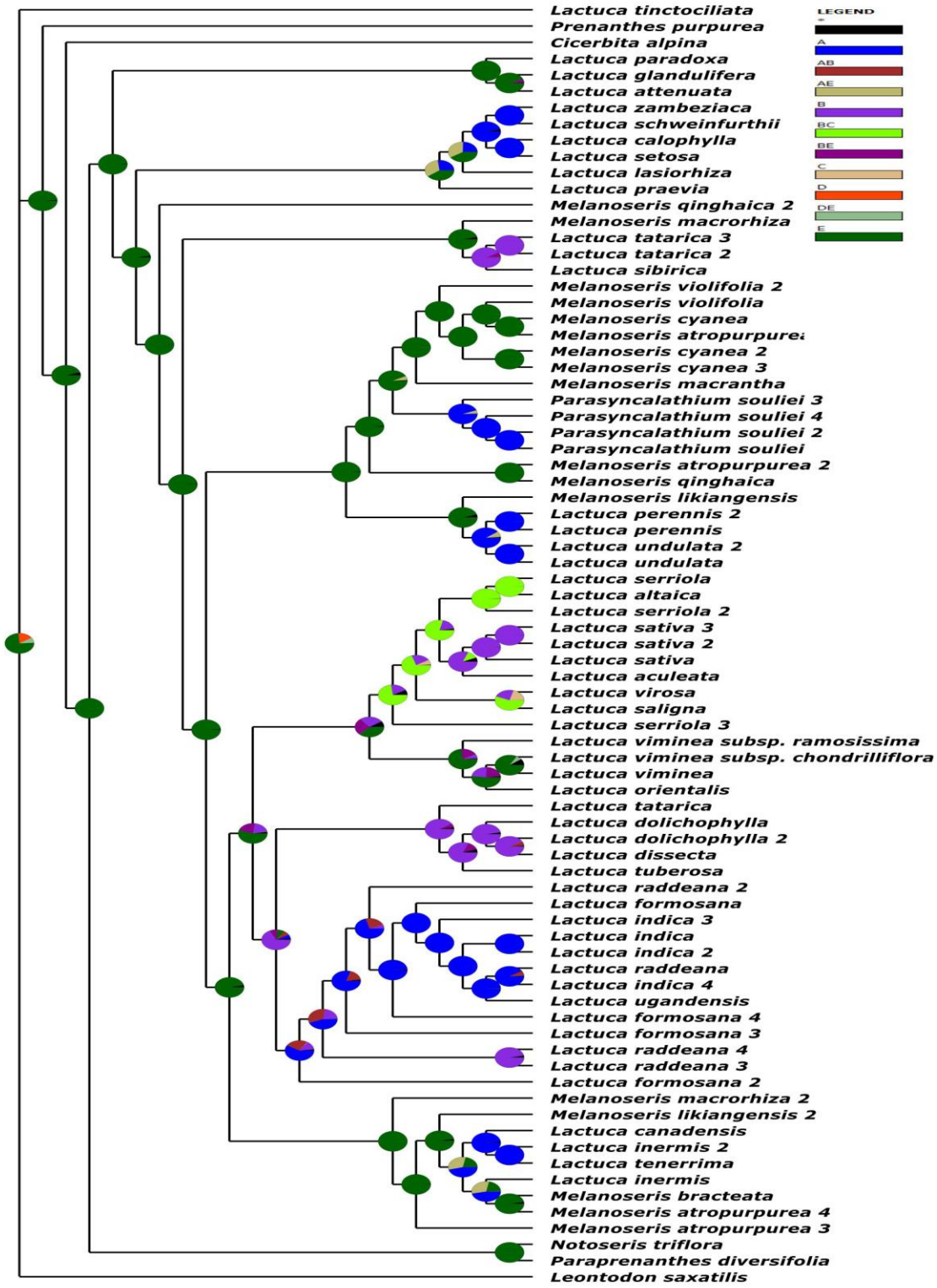


Figure S11 Ancestral stat rib number per side of achene based on Baysian consensus tree topology of the concatenated sequences of *ndhF* gene and *trnL-F* gene; A 1 - 3, B 3 - 7, C 8 - 9, D > 10, E missing data; *L. tinctociliata* was mis-identified and it could be *Launaea cornuta*; *L. ugandensis* should be *Lactuca* sp.

Figure S12 Photo of *Lactuca tinctociliata* I.M.Johnst TYPE specimen.

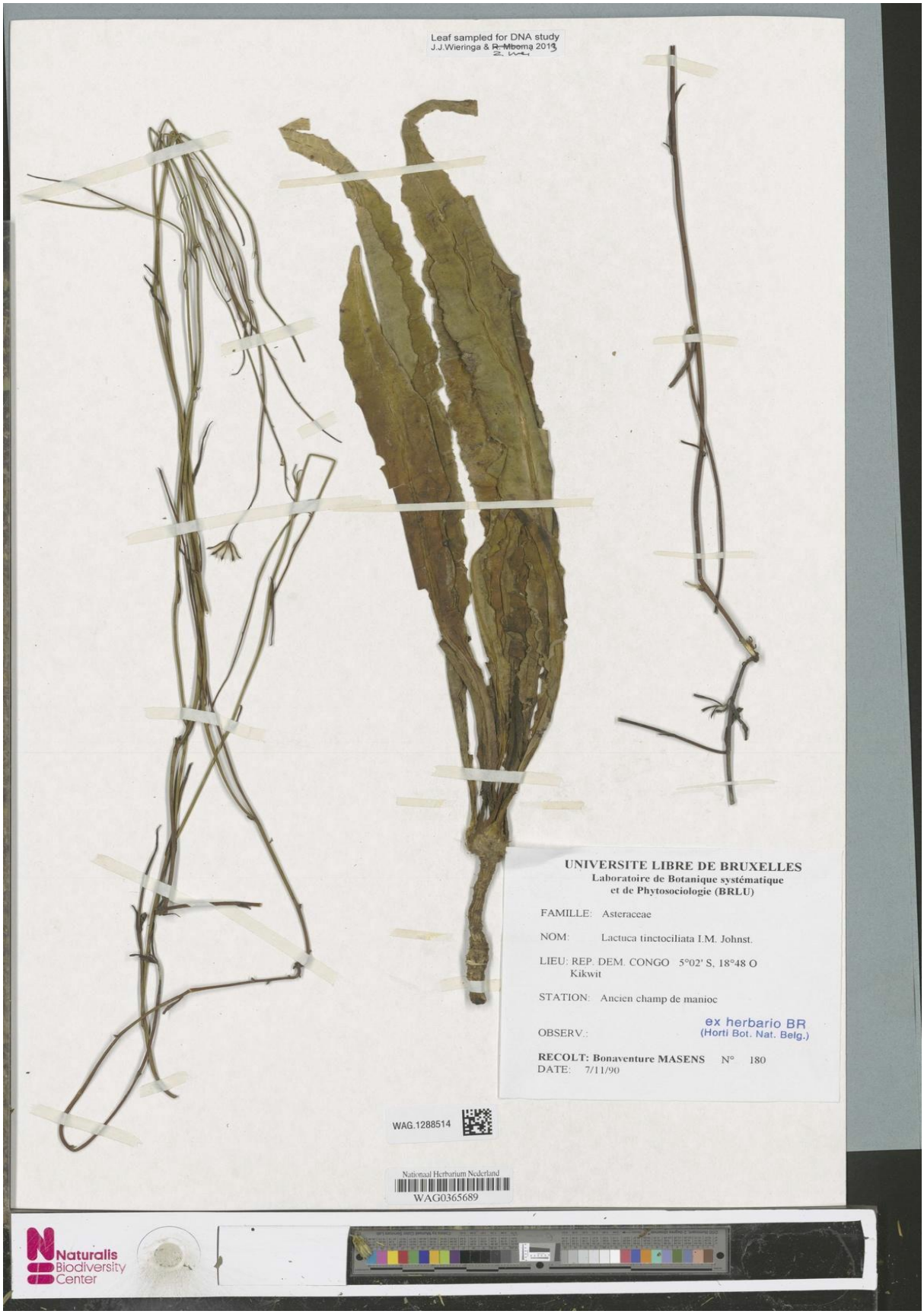


Figure S13 Photo of *Lactuca tinctociliata* I.M.Johnst WAG specimen.

Chapter 3

Phylogenetic analysis of *Lactuca* L. and closely related genera (Asteraceae), using complete chloroplast genomes and nuclear rDNA sequences

Zhen Wei, Freek T. Bakker, M. Eric Schranz

Biosystematics Group, Wageningen University and Research Centre, 6708 PB Wageningen, The Netherlands

Manuscript in preparation

Chapter 3

Abstract

Taxonomists have yet to agree on the circumscription and delimitation of *Lactuca* L., an economically important group. In this study, whole chloroplast genomes and rDNA of *Lactuca* species and four outgroups were sequenced using Illumina HiSeq technology and analysed phylogenetically. The sampling covered 36% of the total *Lactuca* species and all the important geographical groups within the genus. Complete chloroplast large single copy region (LSC), small single copy region (SSC), one inverted repeat region (IR), and the nuclear internal transcribed spacer region (ITS1+5.8S+ITS2) were successfully assembled for 31 samples. The results demonstrated that *Lactuca* as currently circumscribed is not monophyletic, unless the endemic African species (many analysed in this study) are transferred. Phylogenetic analyses based on chloroplast genome data and ITS DNA sequences indicate that there are at least four main *Lactuca* groups: the crop group, the *Pterocypsela* group, the North American group and the group containing widely distributed species (the term 'group' does not mean 'clade' here. For the widely distributed group, it is not a monophyletic clade given the current species sampling.)

.

Key words

Chloroplast phylogenomics; ITS; *Lactuca* phylogeny; lettuce; phylogenetics

Introduction

Lactuca L. belongs to the subtribe Lactucinae, tribe Cichorieae, subfamily Cichorioideae of the family Asteraceae (Kilian et al. 2009). The most well-known species in this genus is *L. sativa* L. (the cultivated lettuce), an economically important vegetable. However, the boundaries of the genus *Lactuca*, unlike other well-studied economic crops such as potato, tomato and rice (Ge et al. 2002; Ge et al. 1999; Peralta and M. 2001; Rodriguez et al. 2009), have remained controversial ever since it was established by Linné in 1753 (Linné 1753). The concepts of *Lactuca* have varied from broad to very narrow. The historically broadly defined *Lactuca*, containing species from present day Lactucinae, Crepidinae and Hyoseridinae, was established by Bentham (1873) and kept by Hoffmann (1890-1894). This very broad delimitation of *Lactuca* is nowadays not well accepted, compared to the other two concepts. The moderately wide concept of *Lactuca*, comprising a total of approximately 100 species, was proposed by Stebbins (1937a, 1937b; 1939) and revised by Feráková and Májovský (1977) and Lebeda et al. (2004; 2007). The narrow circumscription of *Lactuca*, including about 50-70 *Lactuca* species, was suggested by Tuisl (1968) and modified by Shih (1988a, b), Kadereit and Jeffrey (2007).

There are many important morphological characters for the circumscription of *Lactuca*, such as distinctly but moderately compressed and many-ribbed beaked (rarely unbeaked) achenes, pappus of many fine simple smooth or scabrid bristles without (exceptionally with) an outer ring of very short, smooth hairs, capitula with 4-30 or more florets (Lebeda et al. 2007; Shih and Kilian 2011). Among all of the characters for the delimitation of *Lactuca*, Kilian (2001) strongly emphasized three features: (1) the presence or absence of an outer row of minute pappus hairs; (2) the presence or absence of a beak; and (3) the number of flowers per capitulum (Kilian 2001). These features play important roles in distinguishing *Cicerbita*, *Cephalorhynchus*, *Steptorhamphus*, *Mycelis*, *Scariola*, *Mulgedium*, *Chaetoseris* and *Stenoseris* species from both intermediate and narrow viewpoints of *Lactuca* species (Feráková and Májovský 1977; Shih 1988a, b, 1991; Tuisl 1968).

However, Stebbins (1937a) considered the first two characters useless for generic delimitation, because they separate closely related species such as *L. tenerrima* and *L. perennis*. Jeffrey (1966) also rejected the relevance of the outer minute pappus in his treatment of the African *Lactuca*. Koopman et al. (1998; 2001) supported Stebbins' view of the useless characters based on molecular analysis of ribosomal DNA ITS-1 (internal transcribed spacer) sequences and AFLP fingerprints, though the sampling was too small to revise the delimitation of *Lactuca* and related genera. Recently, Wang et al. (2013) revised the *Lactuca* alliance with a focus on the Chinese centre of diversity based on ITS and plastid DNA sequences, including 78 species from *Lactuca*, *Cicerbita*, *Melanoseris*, *Notoseris*, *Parapernanthes*, *Faberia* and *Prenanthes*. They transferred species from *Cephalorrhynchus*, *Cicerbita*, *Mulgedium*, *Chaetoseris* and *Stenoseris* to *Melanoseris* based on molecular analyses. However, their molecular phylogeny is not in line with the important features that

Kilian (2001) emphasized for circumscription and delimitation for *Lactuca*. In addition, chromosome numbers (Mejías 1993), chromosome banding patterns (Koopman et al. 1993), and isozyme markers (Kesseli and Michelmore 1986) were used to study the relationships in *Lactuca* as well. But these studies only included a small number of *Lactuca* species closely related to the cultivated lettuce, which are clearly not enough to revise the entire genus.

The most recent study of phylogenetic relationships within *Lactuca* was by Wei et al. (Chapter 2) based on two chloroplast (cp) DNA sequences, including 32 *Lactuca* species and 16 from genera related to *Lactuca*. This study provided the first molecular phylogenetic tree for African *Lactuca* species. Nevertheless, in their phylogenetic analyses, *Melanoseris* and *Lactuca* species cannot be placed phylogenetically based on the gene sequences used and resulted in a polytomy. The African *Lactuca* species were found to be independent of other *Lactuca* species. Despite the polytomy, they suggested that some endemic African *Lactuca* species probably don't belong to *Lactuca* and more evidence is still needed to solve the polytomy in *Lactuca*.

In recent years, phylogeny inferred from genome-scale data (phylogenomics) have been proven to reveal robust and deep evolutionary relationships (Hackett et al. 2008). In particular, plastid phylogenomics of plants have achieved great success to resolve deep relationships among basal angiosperms (Jansen et al. 2007; Moore et al. 2007), monocot and dicot angiosperms (Barrett et al. 2013; Moore et al. 2010), and even at tribe level and species level (Huang et al. 2014; Ma et al. 2014; Nikiforova et al. 2013). In the present study, we aim to resolve the deeper-nodes in the *Lactuca* phylogeny by generating and analysing complete chloroplast genome data for 36% of known species, using four outgroup species.

Materials and methods

Taxon sampling, DNA extraction and purification

Twenty-seven *Lactuca* accessions representing 25 species, and 4 species from *Lactuca*-allied genera were sampled (**Table 1**). The plant materials used in this study included fresh, silica-dried and herbarium materials (**Table 1**). Information of voucher specimen can be referred to Wei et al. (Chapter 2). DNA extraction methods were modified for herbarium specimens as in Särkinen et al. (2012) and Staats et al. (2011), using a modified cetyltrimethyl-ammonium-bromide (CTAB) method (Doyle and Doyle 1987). The detailed protocol and purification method can be found in Wei et al. (Chapter 2). Additionally, four previously published cp non-coding DNA sequences (*psbA-trnH*, *5' trnL$^{(UAA)}$-trnF*, *rpl32-trnL$^{(UAG)}$* and *trnQ$^{(UUG)}$-5' rps16* spacers) from 17 *Lactuca* accessions, 8 *Melanoseris* species and *Parasyncalathium souliei*, and 54 ribosomal ITS DNA sequences of accessions from *Lactuca* and related genera were obtained from GenBank (**Table S1 & S2**). In total, the sampling covered 36 % of the total *Lactuca* wild species (Lebeda et al. 2004).

Table 1 Taxon sampling and assemble information

No.	Taxon Name	Specimen ID	Sample type	Assembled genome size (bp)	GC (%)	CDS[#] No.	tRNA[#] No.	rRNA[#] No.	ITS length	GC (%)
1	*Cicerbita alpina* Wallr.	7538	H	124,543	36.8	75	20	5	639	56
2	*Lactuca aculeata*[*] Boiss.	CGN15692	F	152,726	37.6	75	20	5	639	54.6
3	*Lactuca attenuata* Stebbins	5982	H	44,755	41.5	17	7	5	438	53.2
4	*Lactuca calophylla* C.Jeffrey	12254	H	114,450	36.9	66	13	5	640	51.4
5	*Lactuca formosana* Maximowicz	2011-1576	S	127,755	36.5	74	19	5	639	53.7
6	*Lactuca glandulifera* Hook.f.	111	H	123,944	36.9	72	19	5	640	53.6
7	*Lactuca imbricata* Hiern	7284	H	\	\	\	\	\	488	52.4
8	*Lactuca indica* L.	2010-1191	S	128,867	36.4	74	20	5	640	53.8
9	*Lactuca inermis* Forssk.	2635	H	148,704	37.8	73	20	5	626	49.4
10	*Lactuca lasiorhiza* (O.Hoffm.) C.Jeffrey	4048	H	103,764	37.9	50	17	5	641	52.7
11	*Lactuca orientalis* Boiss. 1	B 100193265	H	124,076	36.5	70	15	5	641	55.1
12	*Lactuca orientalis* Boiss. 2	B 100394477	H	129,281	36.4	73	20	5	640	54.7
13	*Lactuca paradoxa* Sch.Bip. ex A.Rich.	9405	H	127,646	36.5	73	20	5	640	52.7
14	*Lactuca perennis* L.	5779	S	127,631	36.4	72	19	5	639	52.4
15	*Lactuca praevia* C.D.Adams	855	H	124,228	36.8	67	17	5	641	52.6
16	*Lactuca raddeana* Maximowicz	09-208	S	127,660	36.5	71	20	5	640	53.6
17	*Lactuca saligna* L.	CGN15705	F	127,590	36.5	72	20	5	\	\
18	*Lactuca schweinfurthii* Oliv. & Hiern	2528	H	127,755	36.5	74	19	5	641	52.4
19	*Lactuca serriola* L. 1	CGN15711	F	127,623	36.4	74	20	5	639	54.6
20	*Lactuca serriola* L. 2	MJ19	F	152,732	37.6	75	20	5	639	54.8
21	*Lactuca setosa* Stebbins ex C.Jeffrey	B100426945	H	132,075	37	72	19	5	641	51.8
22	*Lactuca sp.*	2457	H	131,515	37	73	20	5	640	53.8
23	*Lactuca tatarica*(L.) C.A. Meyer	397	H	141,503	37.3	73	20	4	327	50.8

24	*Lactuca tenerrima* Pourr.	3038	H	109,186	37.5	58	15	5	626	50.5
25	*Lactuca viminea* subsp. *chondrilliflora*(Boreau) Malag.	10014	H	127,948	36.4	73	20	5	\	\
26	*Lactuca viminea* subsp. *ramosissima* (All.) Malag.	5974	H	65,790	42.4	1	3	0	315	53.3
27	*Lactuca virosa* L.	CGN09364	F	127,467	36.5	73	19	5	277	52
28	*Lactuca zambeziaca* C.Jeffrey	391	H	75,630	39.6	45	15	5	243	50.6
29	*Notoseris triflora* (Hemsl.) C.Shih	2012-1818	S	127,527	36.6	74	20	4	641	54.4
30	*Paraprenanthes diversifolia* (Vaniot) N.Kilian	2012-1817	S	127,246	36.6	74	20	4	640	52.7
31	*Prenanthes purpurea*(Vaniot) N.Kilian	5375	H	127,789	36.6	73	20	5	641	54.9

If there are more accessions for one species in figures, all samples in this table are indicated as accession 1.
[*]samples with two Inverted Repeat (IR), otherwise one or one complete and one partial IRs
[#] single copy; H herbarium sample, S silica-dried sample, F fresh sample

Table 2 Data characteristics of different data sets

No.	Data set	No. of taxa	No. of total sites	No. of variable/informative sites	GC (%)
1	Cp genome (LSC+SSC+IR)	31	134,523	7,112 (5.3%)	36.8
2	Cp genome (LSC+SSC+IR) + 4 cp genes	57	134,821	7,300 (5.4%)	36.7
3	ITS	83	657	309 (47.0%)	53.5
4	LSC+SSC	31	110,361	6,739 (6.1%)	35.1
5	IR	31	25,451	354 (1.4%)	43.1
6	CDS	31	62,204	1,849 (3.0%)	38.8
7	tRNA+rRNA	31	6,718	36 (0.5%)	54.2

Chloroplast genome sequencing, *de novo* assembly, annotation and alignment

The cp genome and ITS DNA sequences were generated as part of the SYNTHESYS Joint Research Activities 4 (JRA4: Plants/fungi optimised DNA Extraction Techniques: http://www.synthesys.info/joint-research-activities/synthesys-2-jras/jra4-plantsfungi-optimised-dna-extraction-techniques/). The *Lactuca* samples were sequenced by National High-Throughput DNA Sequencing Centre of University of Copenhagen (http://seqcenter.ku.dk/facilities/) and BGI Tech Solutions (HongKong) Co., Limited, using Illumina HiSeq 2000/2500 sequencing, as described in Bakker & al. (2015).

De novo assembly was performed using an Iterative Organelle Genome Assembly (IOGA) pipeline as described in Bakker et al. (2015). The IOGA pipeline included Illumina read trimming (Bolger et al. 2014), filtering (Langmead and Salzberg 2012), correcting and finally *de novo* assembly using the SOAPdenovo2 software package with k-mer values ranging from 37-97 (Luo et al. 2012). A series of candidate assemblies were generated and then the best one was selected based on an Assembly Likelihood Estimation (ALE) test (Bankevich et al. 2012; Clark et al. 2013). For a more detailed description of the IOGA pipeline and the quality of assemblies can be found in Bakker et al. (2015). As in virtually all angiosperms, there are two inverted repeats (IRs) on the cp genome of *L. sativa* (Shaw et al. 2007; Timme et al. 2007). The selected best assembled cp genomes in our study mostly show only one IR region, as a likely result of stacking of reads for the two identical IRs at the same position of the assembly (Bakker et al. 2015). The effect of the double IRs on the best assemblies was discussed in Bakker et al. (2015). In addition, one IR region contains the same information content for phylogenetic construction as the other copy (since they are identical). Thus, the use of just one IR for phylogenetic analysis is merited. The ITS regions have been used to infer the phylogenetic relationships between *Lactuca* and related genera (Koopman et al. 1998; Wang et al. 2013). ITS regions are highly repetitive sequences and the phylogeny based on ITS showed some incongruences with that based on chloroplast phylogeny (Wang et al. 2013). The ITS (ITS1+5.8S+ITS2) regions from herbarium samples were also assembled using IOGA pipeline, using a panel of *Lactuca* rDNA sequences from GenBank as references.

Cp genome assemblies were annotated against reference genome (*L. sativa*, accession number AP007232.1) in Geneious 8.1.5. (Kearse et al. 2012). Coding Sequence (CDS), rRNA and tRNA (Palmer 1985) sequences were extracted using Geneious 8.1.5 and aligned using MAFFT v7.017 (Katoh et al. 2002) plugin in Geneious. All the alignments of CDS, rRNA and tRNA were manually checked and then concatenated respectively in Geneious for further analyses. The nuclear assemblies containing the ITS regions were first aligned with previously published ITS DNA sequences from *Lactuca* species using MAFFT and then manually edited in Geneious. The optimised ITS alignments were then used in phylogenetic analyses. Variable sites and nucleotide composition were calculated in MEGA6 (Tamura et al. 2013). All the annotated genes will be available soon in GenBank.

Data sets and phylogenetic analyses

Seven data sets were used for phylogenetic analyses. **Data Set 1** included the newly generated cp genome data: large single copy region (LSC), small single copy region (SSC), and one IR. **Data Set 2** contained the cp genome data and four published non-coding cp DNA sequences from *Lactuca* and related genera (**Table 2**). **Data Set 3** included new and published rDNA ITS (ITS1 and ITS1+5.8S+ITS2) DNA sequences of species from *Lactuca* and related genera. **Data Sets 4-7** comprise subsets of the cp genome data in the following order, LSC+SSC, IR, CDS, and tRNA+rRNA. Phylogenetic analyses were performed for all the data sets, using Randomized Axelerated Maximum Likelihood (RAxML)-HPC2 run on XSEDE (Stamatakis 2014) from the Cyber-infrastructure for Phylogenetic Research (CIPRES) Science Gateway (V. 3.3, available at http://www.phylo.org/) (Miller et al. 2010). *Prenanthes purpurea*, *Notoseris triflora*, *Paraprenanthes diversifolia*, *Cicerbita alpina* were chosen as outgroups. GTR+GAMMA model was selected for bootstrapping phase (Chapter 2)(Wang et al. 2013). Phylogenetic trees were visualized and edited in TreeGraph 2 (Stover and Muller 2010).

Results

Summary of the NGS data

Twenty-seven *Lactuca* and four cp genomes from related genera were successfully assembled using the newly developed IOGA pipeline (Bakker et al. 2015). The assembled cp genome sizes (including one IR, LSC and SSC) ranged from 44,755 to 132,075 bp (**Table 1**). Only *L. aculeata* (152,726 bp) had two complete IRs whereas *L. inermis* (148,704 bp), *L.serriola2* (152,732 bp) and *L. tatarica* (141,503 bp) contained one complete and one incomplete IR regions. The GC content of cp genomes varied from 36.4% - 37.0%. However, *Lactuca* species with incomplete LSC, SSC or IR had higher GC content (e.g. *L. attenuata* 41.5%, *L. viminea* subsp. *ramosissima* 42.4% and *L. zambeziaca* 39.6%). The number of annotated CDS, tRNA and rRNA ranged from 1 to 75, 3 to 20, 0 to 5, respectively. The length of assembled rDNA ITS sequences ranged from 243 and 641 bp, including complete (partial) ITS1, complete 5.8S ribosomal RNA and complete (partial) ITS2 sequences, and the percentage of GC content was from 49.4% to 56.0% (**Table 1**). The ITS sequence of *L. imbricata* was successfully assembled but the cp genome sequence of this species failed to assemble. In contrast, the cp genomes of *L. saligna* and *L. viminea* subsp. *chondrilliflora* were successfully assembled but ITS sequence assemblies were unsuccessful.

The number of taxa, nucleotide site, variable/informative site and nucleotide composition in different data sets are presented (**Table 2**). We included tRNA and rRNA sequences as non-translated sequences. The non-coding regions of cp genmome were not extracted because the information of introns was not indicated in reference annotation. The ITS sequences contained the highest percentage of variable sites (47.0%) among all data sets. The non-

translated regions (tRNA and rRNA) had the lowest percentage of variable sites (0.5%). Other data sets show a percentage between 1.4% and 6.5%.

Phylogenetic analyses based on cp genome and four cp DNA sequences (Data Sets 1 and 2)

The RAxML tree based on 31 cp genomes (LSC+SSC+IR) (**Data Set 1**) is shown in **Figure 1**. *Notoseris triflora*, *Paraprenanthes diversifolia*, *Cicerbita alpina*, *Prenanthes purpurea*, and most endemic African *Lactuca* species were found to be closely related to *Lactuca* but not as part of the *Lactuca* clade. There are three groups within core *Lactuca*. The first group, the crop group (Bootstrap = 100), includes two clades. One clade (BS = 100) contains the domesticated lettuce (*L. sativa*), *L. aculeata*, *L. serriola*, *L saligna* and *L. virosa*. The other clade (BS = 100) comprises of *L. orientalis*, *L. viminea* subsp. *chondrilliflora* and *L. viminea* subsp. *ramosissima*. The second group, the *Pterocypsela* group (BS = 100) is formed of *L. indica*, *L. raddeana*, *L. sp.*, *L. schweinfurthii* and *L. formosana*. The third group, the widely distributed group has two clades, the first includes *L. tatarica* and group 2 contains *L. inermis*, *L. tenerrima* and *L. perennis*. Please note that the term 'group' does not mean 'clade' here. For the widely distributed group, it is not a clade given the current species sampling, and it may be divided into more groups if more taxa are included.

The African endemic '*Lactuca*' group (BS = 100) does not belong to the genus *Lactuca* and has two clades. One clade (BS = 100) includes 3 scandent species, *L. attenuata* Stebbins, *L. glandulifera* Hook.f. and *L. paradoxa* Sch.Bip. ex A.Rich. The other clade (BS = 100) comprises of *L. calophylla* C.Jeffrey, *L. setosa* Stebbins ex C.Jeffrey, *L. lasiorhiza* (O.Hoffm.) C.Jeffrey, *L. praevia* C.D.Adams and *L. zambeziaca* C.Jeffrey.

The RAxML tree (57 taxa) based on cp genome (LSC+SSC+IR) and four cp non-coding sequences (**Data Set 2**) is illustrated in **Figure 2**. This tree topology is generally consistent with the RAxML tree in **Figure 1** despite the slight differences in sampling. The crop group (BS = 100) contains one *L. viminea* J.Presl & C.Presl species and two more *L. orientalis* accessions than the crop group based on **Data Set 1**. The *Pterocypsela* group (BS = 100) includes two more *L. indica*, *L. raddeana*, and *L. formosana* accessions. The widely distributed group has four clades. Clade 1 (BS = 100) contains two *L. tatarica* accessions and *L. sibirica*. Clade 2 (BS = 100) consists of two *L. inermis* accessions, and one *L. tenerrima*. Clade 3 (BS = 100) includes two *L. undulata* Ledebour accessions, two *L. perennis* accessions. Clade 4 (BS =100) is composed of *L. dolichophylla* Kitamura, *L. dissecta* D. Don and *L. tuberosa* Jacq.

M. bracteata (Hook.f. & Thomson ex C.B.Clarke) N.Kilian is the sister group to other *Melanoseris* species, *L. dolichophylla*, *L. dissecta* and *L. tuberosa*. The *Melanoseris* group (BS = 95) contains *M. cyanea* Edgew, *M. violifolia* (Decne.) N.Kilian, *M. macrantha* (C.B.Clarke) N.Kilian & J.W.Zhang, *M. qinghaica* (S.W.Liu & T.N.Ho) N.Kilian & Ze

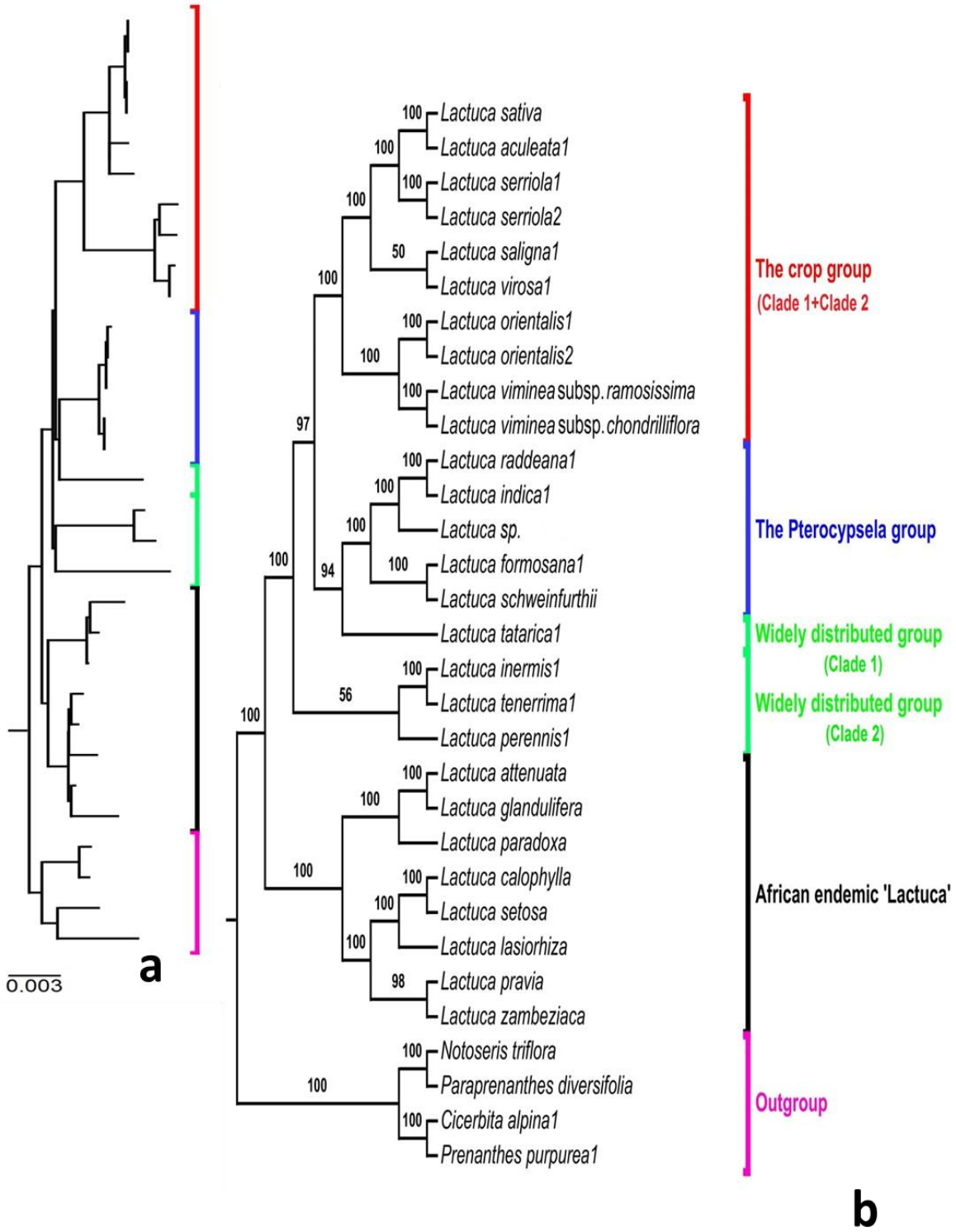


Figure 1 RAxML phylogenetic trees ('best tree') of species of *Lactuca* and related genera based on chloroplast genome (LSC+SSC+IR) DNA sequences. Bootstrap values (>50) are given above the branches. RAxML phylogram with scale bar (indicating substitutions per site) (a) and the same tree in rectangular cladogram style (b). Colours represent different groups as discussed in the main text: 'the crop group' (red), 'the *Pterocypsela* group' (dark blue), 'widely distributed group' (green), African endemic '*Lactuca*' group (black) and outgroup (pink).

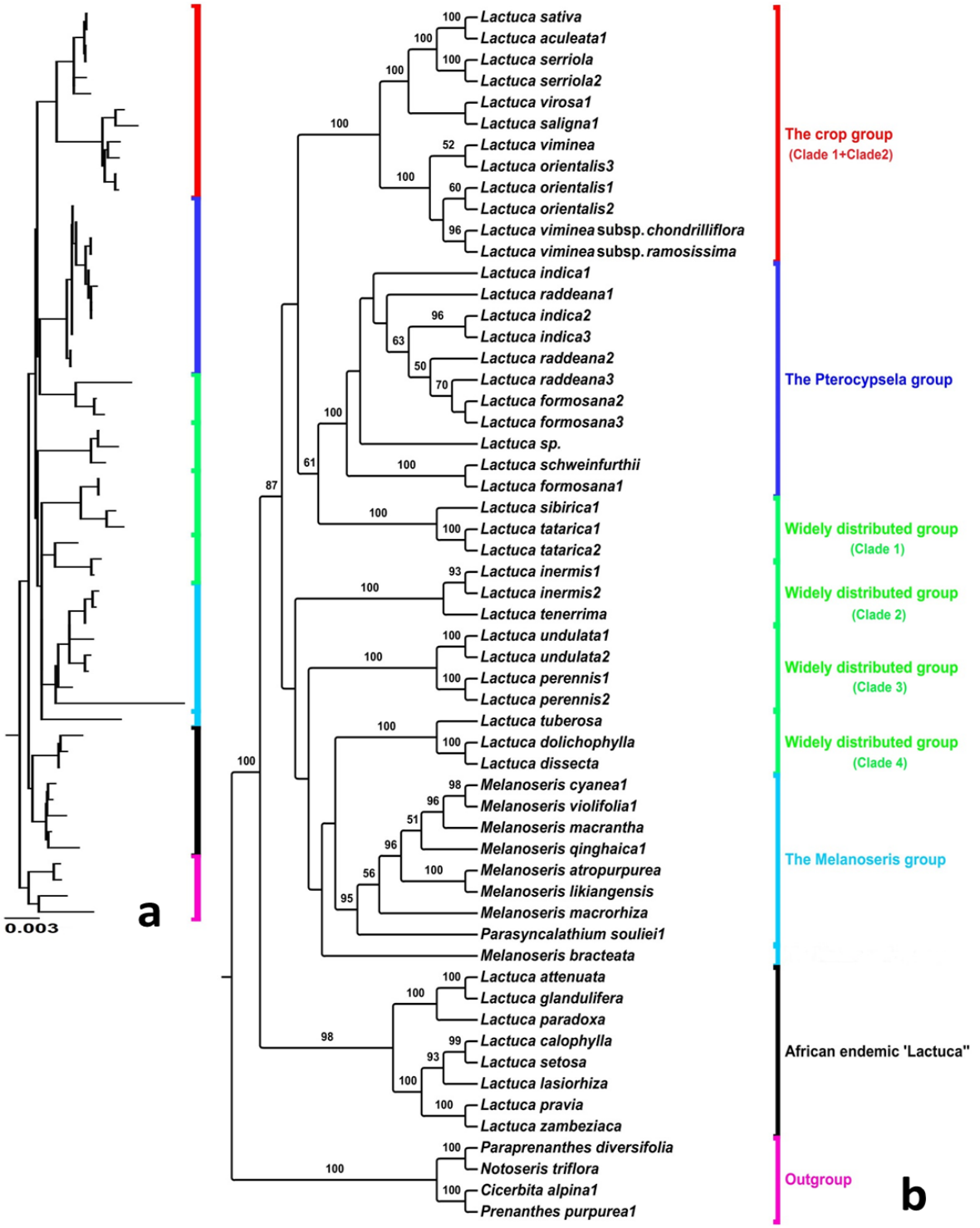


Figure 2 RAxML phylogenetic trees (‘best tree’) of species of *Lactuca* and related genera based on chloroplast genome (LSC+SSC+IR) and four chloroplast non-coding DNA (Sanger sequencing data) sequences. Bootstrap values (>50) are given above the branches. RAxML phylogram with scale bar (indicating substitutions per site) (a) and the same tree in rectangular cladogram style (b). Colours represent different groups as discussed in the main text: ‘the crop group’ (red), ‘the *Pterocypsela* group’ (dark blue), ‘widely distributed group’ (green), the *Melanoseris* group (light blue), African endemic ‘*Lactuca*’ group (black) and outgroup (pink).

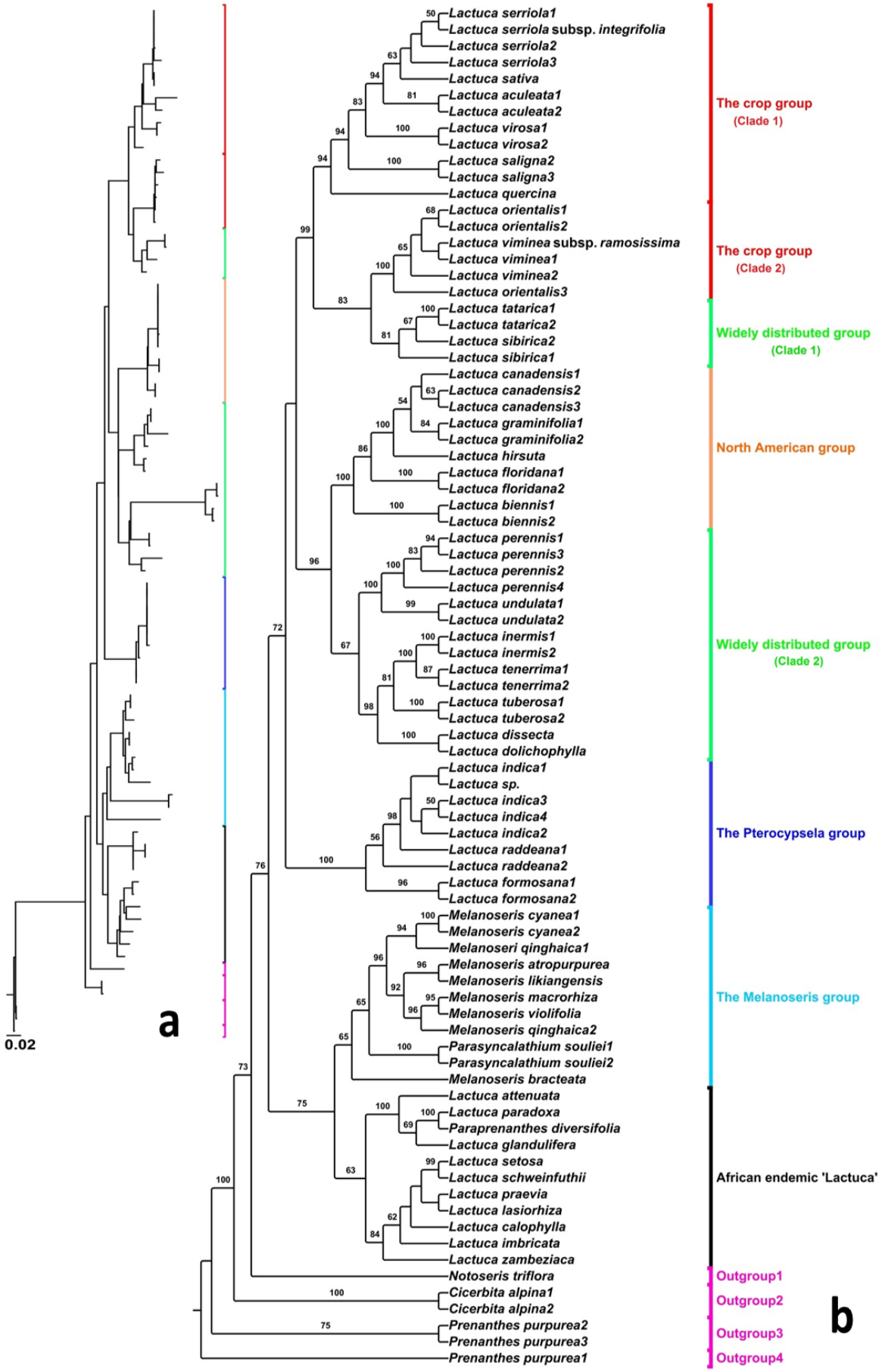
H.Wang, *M. macrorhiza* (Royle) N.Kilian, *M. likiangensis* (Franch.) N.Kilian & Ze H.Wang, *M. atropurpurea* (Franch.) N.Kilian & Ze H.Wang and *Parasyncalathium souliei* (Franch.) J.W.Zhang, Boufford & H.Sun. The African endemic '*Lactuca*' group (BS = 98) has identical clades as those in **Figure 1**, except a bit lower bootstrap value.

Phylogenetic analysis of nuclear rDNA and ITS DNA sequences (ITS1+5.8S+ITS2, Data Set 3)

Figure 3 shows the RAxML tree of species from *Lactuca* and related genera based on 29 newly generated and 54 previously published ITS1 and ITS (ITS1+5.8S+ITS2) sequences. The phylogenetic tree identifies 4 groups within *Lactuca* (the crop group, the *Pterocypsela* group, North American group, widely distributed group), and two groups close to *Lactuca* (the *Melanoseris* group and the African '*Lactuca*' group). The crop group includes two clades. Clade 1 (BS = 94) contains *L. sativa*, *L. aculeata*, *L. serriola*, *L. serriola* L. subsp. *integrifolia* (Gray) G.H.Loos, *L saligna*, *L. virosa* and *L. quercina*. Clade 2 (BS = 100) comprises of *L. orientalis*, *L. viminea* and *L. viminea* subsp. *ramosissima*. The widely distributed group has two clades. Clade 1 (BS = 81) includes *L. tatarica* accessions and *L. sibirica* while Clade 2 (BS = 67) is composed of *L. undulata*, *L. perennis*, *L. tenerrima* and *L. inermis*, *L. tuberosa*, *L. dolichophylla* and *L. dissecta*. The North American group (BS =100) is formed of *L. biennis* (Moench) Fernald, *L. canadensis* L., *L. hirsuta* Muhl. ex Nutt., *L. graminifolia* Michx., and *L. floridana* (L.) Gaertn. The *Pterocypsela* group (BS = 100) consists of *L. indica*, *L. sp.*, *L. raddeana*, and *L. formosana*. The *Melanoseris* group (BS = 65) contains seven *Melanoseris* species and *Parasyncalathium souliei*. The African endemic '*Lactuca*' group (BS = 63) includes *Paraprenanthes diversifolia*, *L. attenuata*, *L. glandulifera*, *L. paradoxa*, *L. setosa*, *L. schweinfurthii*, *L. praevia*, *L. lasiorhiza*, *L. calophylla*, *L. imbricata* and *L. zambeziaca*.

Phylogenetic analyses based on LSC+SSC, IR, CDS, tRNA and rRNA DNA sequences (Data Sets 4 - 7)

Figure 4 shows the four RAxML trees based on different data sub-sets of the cp genome data (LSC+SSC, IR, CDS, tRNA and rRNA) (**Table 2**). In general, there are five main groups in each phylogenetic tree, including the crop group, the *Pterocypsela* group, the widely distributed group, African endemic '*Lactuca*' group and the outgroups. Different colours represent the different main clades. **Figure 4a** shows the RAxML tree using LSC and SSC regions. This tree is deeply resolved with high bootstrap values (76 - 100) on the regarding nodes of six main clades, represented by different colours. **Figure 4b** demonstrates the phylogenetic trees of IR regions and the bootstrap values of the six main clades range from 57 to 100. **Figure 4c** represents the RAxML tree with concatenated CDS. The bootstrap supporting values on six main clades are between 74 and 100. **Figure 4d** illustrates the RAxML tree based on tRNA and rRNA sequences. This phylogenetic tree shows low supporting values, from 41 to 99, for only five main clades.


Lactuca serriola1
Lactuca serriola subsp. integrifolia
Lactuca serriola2
Lactuca serriola3
Lactuca sativa
Lactuca aculeata1
Lactuca aculeata2
Lactuca virosa1
Lactuca virosa2
Lactuca saligna2
Lactuca saligna3
Lactuca quercina
Lactuca orientalis1
Lactuca orientalis2
Lactuca viminea subsp. ramosissima
Lactuca viminea1
Lactuca viminea2
Lactuca orientalis3
Lactuca tatarica1
Lactuca tatarica2
Lactuca sibirica2
Lactuca sibirica1
Lactuca canadensis1
Lactuca canadensis2
Lactuca canadensis3
Lactuca graminifolia1
Lactuca graminifolia2
Lactuca hirsuta
Lactuca floridana1
Lactuca floridana2
Lactuca biennis1
Lactuca biennis2
Lactuca perennis1
Lactuca perennis3
Lactuca perennis2
Lactuca perennis4
Lactuca undulata1
Lactuca undulata2
Lactuca inermis1
Lactuca inermis2
Lactuca tenerrima1
Lactuca tenerrima2
Lactuca tuberosa1
Lactuca tuberosa2
Lactuca dissecta
Lactuca dolichophylla
Lactuca indica1
Lactuca sp.
Lactuca indica3
Lactuca indica4
Lactuca indica2
Lactuca raddeana1
Lactuca raddeana2
Lactuca formosana1
Lactuca formosana2
Melanoseris cyanea1
Melanoseris cyanea2
Melanoseri qinghaica1
Melanoseris atropurpurea
Melanoseris likiangensis
Melanoseris macrorhiza
Melanoseris violifolia
Melanoseris qinghaica2
Parasyncalathium souliei1
Parasyncalathium souliei2
Melanoseris bracteata
Lactuca attenuata
Lactuca paradoxa
Paraprenanthes diversifolia
Lactuca glandulifera
Lactuca setosa
Lactuca schweinfuthii
Lactuca praevia
Lactuca lasiorhiza
Lactuca calophylla
Lactuca imbricata
Lactuca zambeziaca
Notoseris triflora
Cicerbita alpina1
Cicerbita alpina2
Prenanthes purpurea2
Prenanthes purpurea3
Prenanthes purpurea1
0.02
a
b
The crop group
(Clade 1)
The crop group
(Clade 2)
Widely distributed group
(Clade 1)
North American group
Widely distributed group
(Clade 2)
The Pterocypsela group
The Melanoseris group
African endemic 'Lactuca'
Outgroup1
Outgroup2
Outgroup3
Outgroup4

◄ **Figure 3** RAxML phylogenetic trees ('best tree') of species of *Lactuca* and related genera based on ribosomal ITS DNA sequences. Bootstrap values (>50) are given above the branches. RAxML phylogram with scale bar (indicating substitutions per site) (a) and the same tree in rectangular cladogram style (b). Colours represent different groups as discussed in the main text: 'the crop group' (red), 'the *Pterocypsela* group' (dark blue), 'widely distributed group' (green), North American group (orange), the *Melanoseris* group (light blue), African endemic '*Lactuca*' group (black) and outgroup (pink).

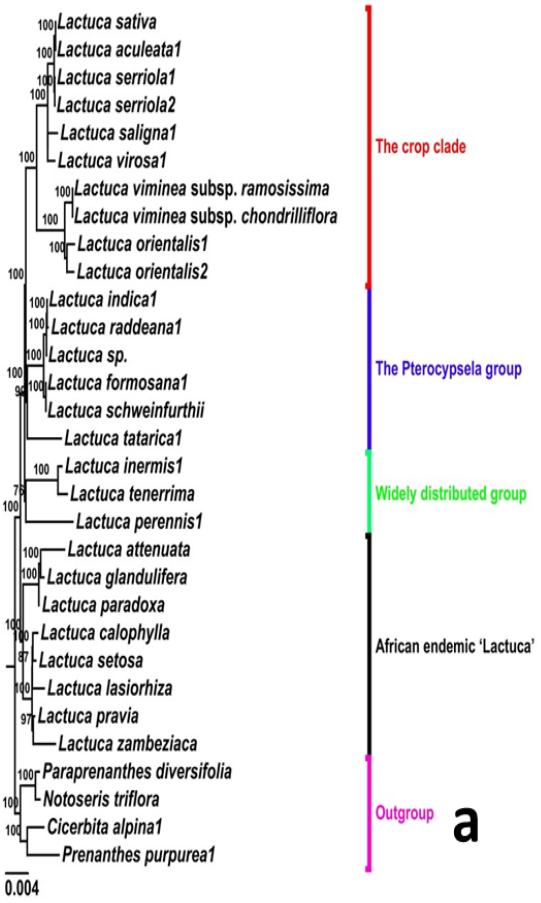
Discussion

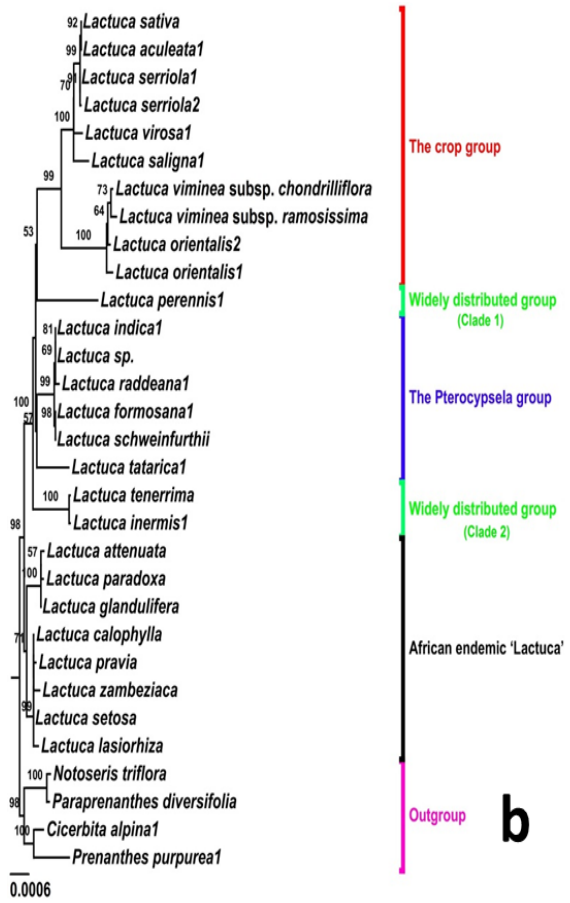
This study presents the chloroplast genome sequences of 24 *Lactuca* species and 4 outgroups, using the most extensive sampling to date. While our taxon sampling covers just 36% of the *Lactuca* species, it does sample all the important geographical groups within the genus (**Figure 1, 2 and 3**) (Lebeda et al. 2004)(Chapter 2). In addition, species from the most related genera were added as outgroups (*N. triflora*, *Paraprenanthes diversifolia*, *C. alpina* and *Prenanthes purpurea*) as well as *Melanoseris* species.

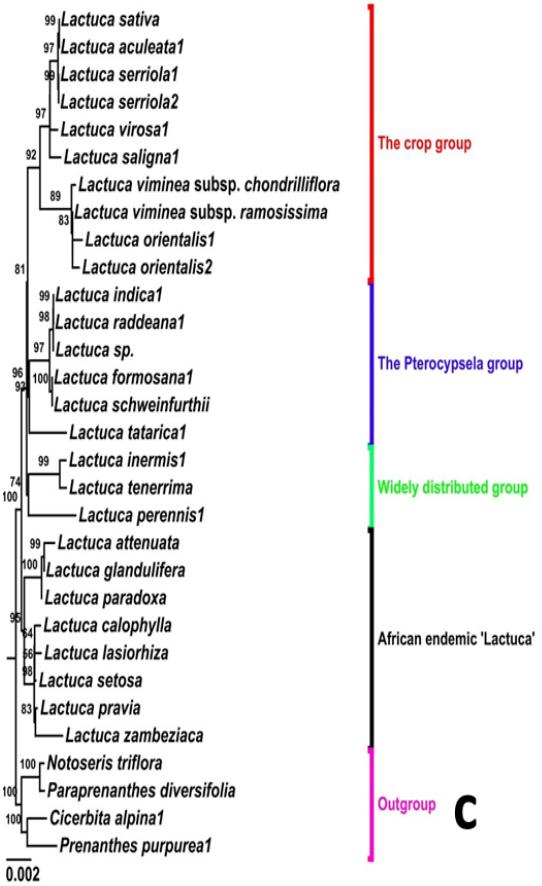
The cp genome data generated in this study mostly contain at least one complete LSC region, one complete SSC region and one complete IR with some exceptions (*L. attenuata*, *L. calophylla*, *L. lasiorhiza*, *L. tenerrima*, *L. viminea* subsp. *ramosissima*, and *L. zambeziaca*). These exceptions all are herbarium samples. Living *Lactuca* species are known to produce latex, containing guaianolide sesquiterpene lactone and lactucin, on the surface of wounded leaves and roots and the latex may affect the obtained DNA quality (Michiels et al. 2003). However, if the latex, produced as secondary metabolites and phytoalexins, would stay in herbarium samples is little known. One possible reason for the incomplete data is probably due to the damaged DNA of the herbarium specimens. The damage and degradation of DNA of herbarium samples had been shown to occur during specimen preparation, and the drying method had also been thought to strongly affect PCR success (Särkinen et al. 2012; Staats et al. 2011) instead of Illumina sequencing, as applied here. The other reason for the low output of these herbarium samples was that in order to minimise destructive sampling we used less than 30 mg of dry plant material. The small amount of sample might be too little to obtain DNA with sufficient quality with highly degraded DNA.

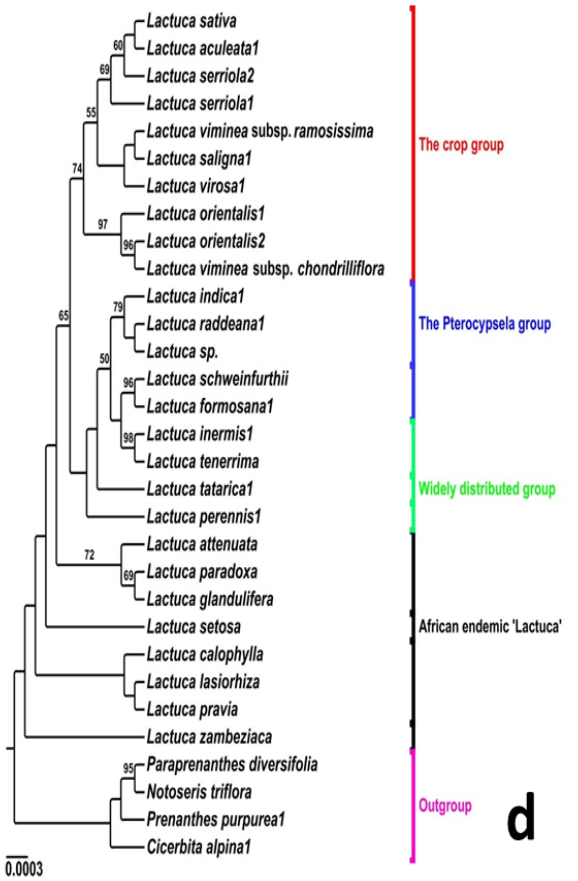
In order to incorporate more taxa in our sampling, we added previously published *Lactuca* accessions with four non-coding DNA markers. **Data Set 2** contains 57 taxa in total, including 30 new cp genomes, 1 published cp genome and 26 accessions with 4 non-coding DNA markers (**Table 2**). Compared with **Data Set 1**, the missing data of the 26 taxa in **Data Set 2** did not deduce the number of variable/informative sites, but increased 188

► **Figure 4** RAxML phylogenetic trees ('best tree') of species of *Lactuca* and related genera based on different data sets. Bootstrap values (>50) are given above the branches. Scale bars indicate substitutions per site. **a** LSC+SSC; **b** IR; **c** CDS; **d** tRNA+rRNA. Colours represent different groups as discussed in the main text: 'the crop group' (red), 'the *Pterocypsela* group' (dark blue), 'widely distributed group' (green), African endemic '*Lactuca*' group (black) and outgroup (pink).


Lactuca sativa
Lactuca aculeata1
Lactuca serriola1
Lactuca serriola2
Lactuca saligna1
Lactuca virosa1
Lactuca viminea subsp. ramosissima
Lactuca viminea subsp. chondrilliflora
Lactuca orientalis1
Lactuca orientalis2
Lactuca indica1
Lactuca raddeana1
Lactuca sp.
Lactuca formosana1
Lactuca schweinfurthii
Lactuca tatarica1
Lactuca inermis1
Lactuca tenerrima
Lactuca perennis1
Lactuca attenuata
Lactuca glandulifera
Lactuca paradoxa
Lactuca calophylla
Lactuca setosa
Lactuca lasiorhiza
Lactuca pravia
Lactuca zambeziaca
Paraprenanthes diversifolia
Notoseris triflora
Cicerbita alpina1
Prenanthes purpurea1
0.004
The crop clade
The Pterocypsela group
Widely distributed group
African endemic 'Lactuca'
Outgroup
a


Lactuca sativa
Lactuca aculeata1
Lactuca serriola1
Lactuca serriola2
Lactuca virosa1
Lactuca saligna1
Lactuca viminea subsp. chondrilliflora
Lactuca viminea subsp. ramosissima
Lactuca orientalis2
Lactuca orientalis1
Lactuca perennis1
Lactuca indica1
Lactuca sp.
Lactuca raddeana1
Lactuca formosana1
Lactuca schweinfurthii
Lactuca tatarica1
Lactuca tenerrima
Lactuca inermis1
Lactuca attenuata
Lactuca paradoxa
Lactuca glandulifera
Lactuca calophylla
Lactuca pravia
Lactuca zambeziaca
Lactuca setosa
Lactuca lasiorhiza
Notoseris triflora
Paraprenanthes diversifolia
Cicerbita alpina1
Prenanthes purpurea1
0.0006
The crop group
Widely distributed group
(Clade 1)
The Pterocypsela group
Widely distributed group
(Clade 2)
African endemic 'Lactuca'
Outgroup
b


Lactuca sativa
Lactuca aculeata1
Lactuca serriola1
Lactuca serriola2
Lactuca virosa1
Lactuca saligna1
Lactuca viminea subsp. chondrilliflora
Lactuca viminea subsp. ramosissima
Lactuca orientalis1
Lactuca orientalis2
Lactuca indica1
Lactuca raddeana1
Lactuca sp.
Lactuca formosana1
Lactuca schweinfurthii
Lactuca tatarica1
Lactuca inermis1
Lactuca tenerrima
Lactuca perennis1
Lactuca attenuata
Lactuca glandulifera
Lactuca paradoxa
Lactuca calophylla
Lactuca lasiorhiza
Lactuca setosa
Lactuca pravia
Lactuca zambeziaca
Notoseris triflora
Paraprenanthes diversifolia
Cicerbita alpina1
Prenanthes purpurea1
0.002
The crop group
The Pterocypsela group
Widely distributed group
African endemic 'Lactuca'
Outgroup
c


Lactuca sativa
Lactuca aculeata1
Lactuca serriola2
Lactuca serriola1
Lactuca viminea subsp. ramosissima
Lactuca saligna1
Lactuca virosa1
Lactuca orientalis1
Lactuca orientalis2
Lactuca viminea subsp. chondrilliflora
Lactuca indica1
Lactuca raddeana1
Lactuca sp.
Lactuca schweinfurthii
Lactuca formosana1
Lactuca inermis1
Lactuca tenerrima
Lactuca tatarica1
Lactuca perennis1
Lactuca attenuata
Lactuca paradoxa
Lactuca glandulifera
Lactuca setosa
Lactuca calophylla
Lactuca lasiorhiza
Lactuca pravia
Lactuca zambeziaca
Paraprenanthes diversifolia
Notoseris triflora
Prenanthes purpurea1
Cicerbita alpina1
0.0003
The crop group
The Pterocypsela group
Widely distributed group
African endemic 'Lactuca'
Outgroup
d

variable/informative sites. Too few complete characters had been proven to be the reason for reduced phylogenetic accuracy rather than missing data (Wiens 2003; Wiens and Morrill 2011). Thus we do not consider the missing data in **Data Set 2** will reduce the accuracy of our phylogenetic analyses. **Data set 3 is** comprised of 83 ribosomal DNA sequences, including our NGS data and published DNA sequences from GenBank. **Data Set 3** includes *L. quercina*, endemic *Lactuca* species from North America, and generally has more than one accession from one species whereas **Data Set 1** and **2** contain *L. viminea* subsp. *chondrilliflora* (not assembled successfully for ITS).

Comparison of phylogenetic trees based on different data sets

The phylogenetic trees inferred from **Data Set 1**, **Data Set 2** and **Data Set 3** are congruent in the following aspects: (1) the RAxML trees contain three main groups within *Lactuca* (the ITS tree has one more group of North American species), and two groups close to but not in *Lactuca* (*Melanoseris* group and African endemic '*Lactuca*' group) (*Melanoseris* group is not included in **Data Set 1**); (2) the main groups are comparable to each other except sampling differences, e.g. the crop group contains all the wild *Lactuca* species that are interfertile or partly interfertile with the domesticated lettuce.

However, there are still some differences among the three RAxML trees: (1) the North American *Lactuca* species were not included in the phylogenetic tree based on cp genome/four non-coding DNA sequences (**Data Set 1** and **2**), and they are the sister group of the widely distributed group, including *L. undulata*, *L. perennis*, *L. tenerrima*, *L. inermis*, *L. tuberosa*, *L. dolichophylla* and *L. dissecta*; (2) *L. tuberosa* is the sister group of *L. dolichophylla* and *L. dissecta* in the tree based on **Data Set 2** but that of *L. tenerrima* and *L. inermis* in the ITS tree (**Data Set 3**); (3) *L. schweinfurthii* is within the *Pterocypsela* group in the cp genome tree (**Data Set 1** and **2**) but placed beyond the whole *Lactuca* clade in the ITS tree (**Data Set 3**); (4) *L. tatarica* and *L. sibirica* are the sister group of the *Pterocypsela* group in both trees inferred from **Data Set 1 and 2** whereas they are the sister group of *L. orientalis*, *L. viminea*, *L. viminea* subsp. *ramosissima* in the nuclear tree (**Data Set 3**); (5) *M. bracteata* is the sister group of *L. tatarica*, *L. sibirica*, *L. tuberosa*, *L. dolichophylla*, *L. dissecta* and the species in the *Pterocypsela* group (low BS value while it is placed within the *Melanoseris* clade in the ITS tree (**Data Set 3**); (6) *Paraprenanthes diversifolia* is in the same clade with *N. triflora* and the sister group of all *Lactuca* and *Melanoseris* species in the two cp trees (**Data Set 1 and 2**) but placed in the scandent African *Lactuca* clade in the ITS tree (**Data Set 3**).

Incongruence between the cp and nuclear/ITS trees has been reported very often (Fehrer et al. 2007; Kim and Donoghue 2008; Nishimoto et al. 2003; Van Raamsdonk et al. 1997; Yu et al. 2013). Technical causes (insufficient taxon sampling, long branch attraction, sequencing errors etc.), divergent alleles among the multiple ITS copies within a nucleus, additive polymorphism and chloroplast capture after an introgression/hybridization event could be the

reasons to explain the incongruence between the cp and nuclear/ITS trees (Acosta and Premoli 2010; Fehrer et al. 2007; Fuertes Aguilar and Nieto Feliner 2003; Stegemann S 2012; Tsitrone et al. 2003; Van Raamsdonk et al. 1997; Wendel and Doyle 1998; Wolfe and Elisens 1995). In our study, taxon sampling covered all the important groups in *Lactuca* and most *Lactuca* species have more than accessions. Long branch attraction and sequencing errors were not observed in our results.

Despite the fact that the North American *Lactuca* species were not included in the cp phylogenetic trees, the ITS tree shows that species in the *Pterocypsela* group are the sister group to species from North American and widely distributed group. *L. graminifolia*, *L. floridana* and *L. spicata* could be crossed with *L. indica*, *L. laciniata* (now treated as *L. indica*), *L. raddeana*, and *L. tatarica* and produce self-sterile or partly fertile hybrid plants (Thompson et al. 1941; Wang et al. 2013). Although the North American species have an unique chromosome number (2n = 34) in *Lactuca*, they share a distinctive morphological character, flattened with somewhat thickened margins or broadly winged achene, with the *Pterocypsela* clade species (Hand et al. 2009). Our results of **Data Set 3** and the hybridization experiments confirm a close relationship between the species from the *Pterocypsela* group and North American group.

L. sibirica is closely related to *L. tatarica* and fully fertile with *L. tatarica* (Koopman et al. 2001). Meanwhile, *L. tatarica* is closely related to *L. indica* in the *Pterocypsela* group because they can be crossed with each other and produce self-sterile seeds (van Treuren et al. 2011). Consequently, *L. sibirica* and *L. tatarica* are close to species in the *Pterocypsela* group. Hybridization experiments showed that *L. tatarica* could be somatically hybridized with *L. sativa* (Chupeau et al. 1994; Maisonneuve et al. 1995), indicating a relatively far relationship between *L. tatarica* and species in the crop group. Therefore, the position of *L. sibirica* and *L. tatarica* in the cp tree is more reliable than that of in the ITS tree.

The conflicting position of *L. tuberosa* could be a putative case of chloroplast capture. *L. dolichophylla* and *L. dissecta* have some shared characters such as capitula with 6-15(20) blue florets and 3-5 ribs on either side of the achene while *L. tuberosa* has tuberous roots and broadly winged achenes (Hand et al. 2009+; Shih and Kilian 2011). The incongruent positions of *L. schweinfurthii* between the cp and ITS trees are very likely the results of reticulation and chloroplast capture. In addition, the position of *M. bracteata* in the ITS tree seems more reliable because it has higher BS values. Lastly, the position of *Paraprenanthes diversifolia* in the cp tree is a better estimate because it is more consistent with morphology (Wang et al. 2013). It should be noted that, for these species, we only sampled one accession. More accessions should be sequenced in the future to validate the hypothesis.

We constructed phylogenetic relationships within *Lactuca* and related genera in different data scales/data sets. Although the phylogenetic tree of cp genome (**Data Set 1**) has lowest taxon sampling, it has the highest BS values with only two BS values below 90, due to more

variable/informative sites than other data sets (Heath et al. 2008; Wiens 2003). Consequently, **Data Set 2**, **4** and **6** have more resolution in deep branches in the phylogenetic trees than **Data Set 3, 5** and **Data Set 7**. The phylogenetic tree based on cp genomes reveals the deepest phylogenetic relationships within *Lactuca* and the lower taxonomic sampling in the cp tree have not affected the main groups within *Lactuca*.

Implication for *Lactua* taxonomy

The results from different data sets all have demonstrated that the current *Lactuca* genus is not monophyletic. In order to maintain the monophyly of *Lactuca*, the circumscription of the current *Lactuca* genus should be revised. Specifically, we suggest that the endemic African species should be treated as a new genus since they form a monophyletic group. Other African *Lactuca* species, not considered autochthonous, should still be treated as *Lactuca* species. The African group of *Lactuca* species contains at least 43 species, and 75% of them (31 in total) should be considered as endemic (Lebeda et al. 2004). The autochthonous species are mostly distributed in central and southern Africa, and some are reported from eastern and western tropical Africa. All the African species, recorded from northern Africa, are not endemic but widely distributed in different continents (Jeffrey 1966; Lebeda et al. 2004; Stebbins 1937b). Wei et al. (Chapter 2) reported that the endemic species of African *Lactuca* group could probably be treated as a new genus based on two cp DNA sequences, though the resolution was not very good. Now we confirm this assumption with high supporting BS portion (BS = 100) on nodes of phylogenetic trees in **Figure 1 and 2** and moderate supporting BS value (BS = 76) in **Figure 3**. Nevertheless, limited information about the whole African group is available, no matter morphological or molecular characters. More molecular and morphological studies on the type specimens of the entire African group are still needed to revise the boundary of *Lactuca*.

Although the taxon sampling in this study was only 36%, and therefore may have affected our inferred tree topoplogies, it covered all the important geographical groups. It can be inferred from our phylogenetic analyses that there are at least four main groups in *Lactuca*. The first on is the crop group, including *L. sativa*, *L. aculeata*, *L. serriola*, *L. serriola* subsp. *integrifolia*, *L saligna*, *L. virosa*, *L. quercina*, *L. orientalis*, *L. viminea*, *L. viminea* subsp. *chondrilliflora* and *L. viminea* subsp. *ramosissima*. The second one is the *Pterocypsela* group (Chapter 2), containing *L. indica*, *L. raddeana*, *L. sp.* and *L. formosana*. The third group is the North American group, comprising of autochthonous *Lactuca* species from North American, *L. biennis*, *L. canadensis*, *L. hirsuta*, *L. graminifolia*, and *L. floridana*. The last group is composed of widely distributed *Lactuca* species from Europe, Asia, Africa and North America (*L. tatarica*, *L. sibirica*, *L. undulata*, *L. perennis*, *L. tenerrima*, *L. inermis*, *L. tuberosa*, *L. dolichophylla* and *L. dissecta*) and it may be divided into more groups if more taxa are included.

***Melanoseris* species**

Melanoseris species and *Parasyncalathium souliei* are closely related to *Lactuca* and ex-endemic African species. Zhang et al. (2009; 2011) suggested that this species should be either put back in *Lactuca* or treated as a new genus. However, Wang et al. (2013) proposed that this species should be placed in *Melanoseris* and our results are in line with them.

Conclusions

In this study, we provide the first cp genome sequences of wild *Lactuca* species and strong support for deep nodes in our phylogenetic trees between species from *Lactuca* and related genera based on cp genome/non-coding and nuclear rDNA and ITS sequences. This study includes all the important geographical groups within *Lactuca* and elucidates the following key issues about the *Lactuca* species used in this study:

1. The *Lactuca* species, native to the African continent, should be excluded from the genus *Lactuca* and treated as a new genus;
2. There are at least four main groups within the genus *Lactuca*: the crop group, the *Pterocypsela* group, the North American group and the group containing widely distributed species;
3. The cp genome DNA sequences can resolve deep phylogenetic relationships on species level in *Lactuca*.

Acknowledgements

Thanks to China Scholarship Council (http://en.csc.edu.cn/) for providing funding to Z.W. The funding for this study came from SYNTHESYS Joint Research Activities 4 (JRA4: Plants/fungi optimised DNA Extraction Techniques) supported by European Union-funded integrated Activities grant awarded to dr. F.T.B. and Prof. Dr. Walter G. Berendsohn. We appreciate dr. J.J. Wieringa, dr. M.J.W. (Marieke) Jeuken and dr. N.K. for providing samples for DNA extraction.

References

Acosta MC, Premoli AC (2010) Evidence of chloroplast capture in South American Nothofagus (subgenus Nothofagus, Nothofagaceae). Mol Phylogenet Evol 54:235-242

Bakker FT, Lei D, Yu J, Mohammadin S, Wei Z, van de Kerke S, Gravendeel B, Nieuwenhuis M, Staats M, Alquezar-Planas DE, Holmer R (2015) Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline. Biol J Linn Soc doi:10.1111/bij.12642

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of computational biology : a journal of computational molecular cell biology 19:455-477

Barrett CF, Davis JI, Leebens-Mack J, Conran JG, Stevenson DW (2013) Plastid genomes and deep relationships among the commelinid monocot angiosperms. Cladistics 29:65–87

Bentham G (1873) Compositae. In: Bentham G, Hooker J (eds) Genera Plantarum. Reeve & Company, London

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics

Chupeau M-C, Maisonneuve B, Bellec Y, Chupeau Y (1994) A *Lactuca* Universal Hybridizer, and its use in creation of fertile interspecific somatic hybrids. Mol Gen Genet 245:139-145

Clark S, Egan R, Frazier P, Wang Z (2013) ALE: a generic assembly likelihood evaluation framework for assessing the accuracy of genome and metagenome assemblies. Bioinformatics 29:435-443

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phyt Bull 19:11-15

Fehrer J, Gemeinholzer B, Chrtek J, Jr., Brautigam S (2007) Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in Pilosella hawkweeds (*Hieracium*, Cichorieae, Asteraceae). Mol Phylogenet Evol 42:347-361

Fer âkov á V, M âjovsk ý J (1977) The genus *Lactuca* L. in Europe. Bratislava: Univerzita Komenske ho

Fuertes Aguilar J, Nieto Feliner G (2003) Additive polymorphisms and reticulation in an ITS phylogeny of thrifts (*Armeria*, Plumbaginaceae). Molecular Phylogenetics and Evolution 28:430-447

Ge S, Li A, Lu B-R, Shou-Zhou; Z, Hong D-Y (2002) A Phylogeny of the rice tribe Oryzeae (Poaceae) based on MatK sequence data. American Journal of Botany 89:1967-1972

Ge S, Sang T, Lu B-R, Hong D-Y (1999) Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proceedings of the National Academy of Sciences of the United States of America 96:14400-14405

Hackett SJ, Kimball RT, Reddy S, Bowie RCK, Braun EL, Braun MJ, Chojnowski JL, Cox WA, Han K-L, Harshman J, Huddleston CJ, Marks BD, Miglia KJ, Moore WS, Sheldon FH, Steadman DW, Witt CC, Yuri T (2008) A Phylogenomic Study of Birds Reveals Their Evolutionary History. Science 320:1763-1768

Hand R, Kilian N, Raab-Straube Ev (2009+) International Cichorieae Network: Cichorieae Portal. Available: http://wp6-cichorieaee-taxonomyeu/portal/ Accessed 30 July 2014

Heath TA, Hedtke SM, Hillis DM (2008) Taxon sampling and the accuracy of phylogenetic analyses. Journal of Systematics and Evolution 46:239-257

Hoffmann O (1890-1894) Compositae. In: Hoffmann O (ed) Die natürlichen Pflanzenfamilien. Engelmann, Leipzig, pp 87-387

Huang H, Shi C, Liu Y, Mao S-Y, Gao L-Z (2014) Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC evolutionary biology 14:151

Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Muller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, McNeal JR, Kuehl JV, Boore JL (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proceedings of the National Academy of Sciences of the United States of America 104:19369-19374

Jeffrey C (1966) Notes on Compositae: I. Kew Bull 18:427-486

Kadereit JW, Jeffrey C (2007) Flowering Plants. Eudicots: Asterales. In: Kubitzki K (ed) The families and genera of vascular plants, pp 187-189

Katoh K, Misawa K, Kuma k-i, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30:3059-3066

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647-1649

Kesseli RV, Michelmore RW (1986) Genetic variation and phylogenies detected from isozyme markers in species of *Lactuca*. J Hered 77:324-331

Kilian N (2001) *Lactuca* stebbinsii (Lactuceae, Compositae), a puzzling new species from Angola. Willdenowia 31:71-78

Kilian N, Gemeinholzer B, Lack HW (2009) Tribe Cichorieae. In: Funk VA SA, Stuessy T and Bayer R (ed) Systematics, evolution and biogeography of the Compositae, Vienna: IAPT, pp pp. 343-383

Kim ST, Donoghue MJ (2008) Incongruence between cpDNA and nrITS trees indicates extensive hybridization within Eupersicaria (Polygonaceae). Am J Bot 95:1122-1135

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. American Journal of Botany 85:1517-1530

Koopman WJM, Jong JHd, Vries IMd (1993) Chromosome banding patterns in Lettuce species (Lactuca sect. Lactuca, Compositae). Plant Systematics and Evolution 185:249-257

Koopman WJM, Zevenbergen MJ, Van den Berg RG (2001) Species Relationships in *Lactuca* s.l. (Lactuceae, Asteraceae) Inferred from AFLP Fingerprints. American Journal of Botany 10:1881-1887

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature methods 9:357-359

Lebeda A, Doležalová I, Feráková V, Astley D (2004) Geographical Distribution of Wild *Lactuca* Species (Asteraceae, Lactuceae). The Botanical Review 70:328-356

Lebeda A, Ryder E, Grube R, Dolezalova I, Krıstkova E (2007) Lettuce (Asteraceae; *Lactuca* spp.) In: Singh R (ed) Genetic resources, chromosome engineering, and crop improvement series, vol 3-vegetable crops. CRC Press, Boca Raton, pp 377-472

Linné Cv (1753) Species Plantarum, 1 edn

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Zhu X, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T-W, Wang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1:18

Ma PF, Zhang YX, Zeng CX, Guo ZH, Li DZ (2014) Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (poaceae). Systematic biology 63:933-950

Maisonneuve B, Chupeau MC, Bellec Y, Chupeau Y (1995) Sexual and somatic hybridization in the genus *Lactuca*. Euphytica 85 85:281-285

Mejías JA (1993) Cytotaxonomic studies in the Iberian taxa of the genus *Lactuca* (Compositae). Bot Helvet 103:113-130

Michiels A, Van den Ende W, Tucker M, Van Riet L, Van Laere A (2003) Extraction of high-quality genomic DNA from latex-containing plants. Analytical Biochemistry 315:85-89

Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, pp 1-8

Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proceedings of the National Academy of Sciences of the United States of America 104:19363-19368

Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE (2010) Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proceedings of the National Academy of Sciences of the United States of America 107:4623-4628

Nikiforova SV, Cavalieri D, Velasco R, Goremykin V (2013) Phylogenetic analysis of 47 chloroplast genomes clarifies the contribution of wild species to the domesticated apple maternal line. Molecular biology and evolution 30:1751-1760

Nishimoto Y, Ohnishi O, Hasegawa M (2003) Topological incongruence between nuclear and chloroplast DNA trees suggesting hybridization in the urophyllum group of the genus *Fagopyrum* (Polygonaceae). Genes Genet Syst:139-153

Palmer JD (1985) Comparative organization of chloroplast genomes. Annu Rev Genet 19:325-354

Peralta IE, M. SD (2001) Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes (*Solanum* L. section *Lycopersicon* [Mill.] Wettst. subsection *Lycopersicon*). American Journal of Botany 88:1888-1902

Rodriguez F, Wu F, Ane C, Tanksley S, Spooner DM (2009) Do potatoes and tomatoes have a single evolutionary history, and what proportion of the genome supports this history? BMC evolutionary biology 9:191

Särkinen T, Staats M, Richardson JE, Cowan RS, Bakker FT (2012) How to open the treasure chest? Optimising DNA extraction from herbarium specimens. PloS one 7:1-7

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. American Journal of Botany 94:275-288

Shih C (1988a) Revision of *Lactuca* L. and two new genera of tribe Lactuceae(Compositae) on the mainland of Asia. Acta Phytotaxon Sin 26:382-393

Shih C (1988b) Revision of *Lactuca* L. and two new genera of tribe Lactuceae(Compositae) on the mainland of Asia. Acta Phytotaxon Sin 26:418-428

Shih C (1991) On circumscription of the genus *Cicerbita* Wall., and two new genera of Compositae from Sino-Himalayan region. Acta Phytotaxon Sin 29:394-417

Shih C, Kilian N (2011) Cichorieae [excl.*Tragopogon*, *Taraxacum*, *Hieracium* and *Pilosella*]. Science Press & St Louis: Missouri Botanical Garden, Beijing

Staats M, Cuenca A, Richardson JE, Vrielink-van Ginkel R, Petersen G, Seberg O, Bakker FT (2011) DNA damage in plant herbarium tissue. PloS one 6:e28448

Stamatakis A (2014) RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics

Stebbins GL (1937a) Critical notes on *Lactuca* and related genera. J Bot 75:12-18

Stebbins GL (1937b) The Scandent Species of *Prenanthes* and *Lactuca* in Africa. Bulletin du Jardin botanique de l'État a Bruxelles 14:333-352

Stebbins GL (1939) Notes on some Indian species of *Lactuca*. Indian Forester Bot 1:237-245

Stegemann S KM, Greiner S, Bock R (2012) Horizontal transfer of chloroplast genomes between plant species. Proc Natl Acad Sci USA 109:2434-2438

Stover BC, Muller KF (2010) TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC bioinformatics 11:7

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular biology and evolution 30:2725-2729

Thompson RC, Whitaker TW, Kosar WF (1941) Interspecific genetic relationships in *Lactuca*. Journal of Agricultural Research 63:91-107

Timme RE, Kuehl JV, Boore JL, Jansen RK (2007) A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. American Journal of Botany 94:302-312

Tsitrone A, Kirkpatrick M, Levin D (2003) A model for chloroplast capture. Evolution 57:1776–1782

Tuisl G (1968) Der Verwandtschaftskreis der Gattung *Lactuca* L. im iranischen Hochland und seinen Randgebieten. Ann Nat Hist Mus Wien 72:587-638

Van Raamsdonk LWD, Smiech MP, Sandbrink JM (1997) Introgression explains incongruence between nuclear and chloroplast DNA-based phylogenies in *Allium* section *Cepa*. Botanical Journal of the Linnean Society 123

van Treuren R, van der Arend AJM, Schut JW (2011) Distribution of downy mildew (*Bremia lactucae* Regel) resistances in a genebank collection of lettuce and its wild relatives. Plant Genetic Resources 11:15-25

Wang ZH, Peng H, Kilian N (2013) Molecular phylogeny of the *Lactuca* alliance (Cichorieae subtribe Lactucinae, Asteraceae) with focus on their Chinese centre of diversity detects potential events of reticulation and chloroplast capture. PloS one 8:e82692

Wendel J, Doyle J (1998) Phylogenetic incongruence: window into genome history and molecular evolution. In: Soltis DE SP, Doyle JJ (ed) Molecular systematics of plants II: DNA sequencing. Kluwer Academic Publishers, Boston, pp 265-296

Wiens JJ (2003) Missing Data, Incomplete Taxa, and Phylogenetic Accuracy. Systematic biology 52:528-538

Wiens JJ, Morrill MC (2011) Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. Systematic biology 60:719-731

Wolfe AD, Elisens WJ (1995) Evidence of chloroplast capture and pollen-mediated gene flow in *Penstemon* sect. *Peltanthera* (Scrophulariaceae). Systematic Botany 20:395-412

Yu WB, Huang PH, Li DZ, Wang H (2013) Incongruence between nuclear and chloroplast DNA phylogenies in *Pedicularis* section *Cyathophora* (Orobanchaceae). PloS one 8:e74828

Zhang J-W, Nie Z-L, Sun H (2009) Cytological study on the genus *Syncalathium* (Asteraceae-Lactuceae), an endemic taxon to alpine scree of the Sino-Himalayas. Journal of Systematics and Evolution 47:226-230

Zhang J-W, Nie Z, Wen J, Sun H (2011) Molecular phylogeny and biogeography of three closely related genera, *Soroseris*, *Stebbinsia*, and *Syncalathium* (Asteraceae, Cichorieae), endemic to the Tibetan Plateau, SW China. Taxon 60:15-26

Supplementary tables

Table S1 Information of the 4 non-coding chloroplast regions (Sanger sequencing data) from *Melenoseris* and *Lactuca* species

Species name	trnQ(UUG)-5'rps16	5'trnL(UAA)-trnF	rpl32-trnL(UAG)	psbA-trnH
Melanoseris atropurpurea (Franch.) N.Kilian & Ze H.Wang	KF486272.1	KF486144.1	KF486016.1	KF485888.1
Melanoseris bracteata (Hook.f. & Thomson ex C.B.Clarke) N.Kilian	KF485862.1	KF486118.1	KF485990.1	KF486246.1
Melanoseris cyanea Edgew	KF486256.1	KF486128.1	KF486000.1	KF485872.1
Melanoseris likiangensis (Franch.) N.Kilian & Ze H.Wang	KF486271.1	KF486143.1	KF486015.1	KF485887.1
Melanoseris macrantha (C.B.Clarke) N.Kilian & J.W.Zhang	KF486249.1	KF486121.1	KF485993.1	KF485865.1
Melanoseris macrorhiza (Royle) N.Kilian	KF486247.1	KF486119.1	KF485991.1	KF485863.1
Melanoseris qinghaica (S.W.Liu & T.N.Ho) N.Kilian & Ze H.Wang	KF486252.1	KF486124.1	KF485996.1	KF485868.1
Melanoseris violifolia (Decne.) N.Kilian	KF486250.1	KF486122.1	KF485994.1	KF485866.1
Parasyncalathium souliei (Franch.) J.W.Zhang, Boufford & H.Sun	KF486243.1	KF486115.1	KF485987.1	KF485859.1
Lactuca undulata	KF486287.1	KF486159.1	KF486031.1	KF485903.1
Lactuca undulata2	KF486288.1	KF486160.1	KF486032.1	KF485904.1
Lactuca dissecta	KF486289.1	KF486161.1	KF486033.1	KF485905.1
Lactuca dolichophylla	KF486290.1	KF486162.1	KF486034.1	KF485906.1
Lactuca tuberosa	KF486291.1	KF486163.1	KF486035.1	KF485907.1
Lactuca inermis2	KF486292.1	KF486164.1	KF486036.1	KF485908.1
Lactuca indica2	KF486293.1	KF486165.1	KF486037.1	KF485909.1
Lactuca indica3	KF486294.1	KF486166.1	KF486038.1	KF485910.1
Lactuca formosana2	KF486295.1	KF486167.1	KF486039.1	KF485911.1
Lactuca formosana3	KF486296.1	KF486168.1	KF486040.1	KF485912.1
Lactuca raddeana2	KF486297.1	KF486169.1	KF486041.1	KF485913.1
Lactuca raddeana3	KF486298.1	KF486170.1	KF486042.1	KF485914.1
Lactuca orientalis3	KF486299.1	KF486171.1	KF486043.1	KF485915.1
Lactuca sibirica	KF486301.1	KF486173.1	KF486045.1	KF485917.1
Lactuca tatarica2	KF486302.1	KF486174.1	KF486046.1	KF485918.1
Lactuca viminea	KF486300.1	KF486172.1	KF486044.1	KF485916.1
Lactuca perennis2	KF486286.1	KF486158.1	KF486030.1	KF485902.1

Table S2 Information of the rDNA ITS regions from *Melenoseris* and *Lactuca* species

Species name	Accession number
Cicerbita alpina 2	AJ228651.1
Lactuca aculeata 2	AJ228612.1
Lactuca biennis 1	HQ161959.1
Lactuca biennis 2	KP828828.1
Lactuca canadensis 1	HQ161956.1
Lactuca canadensis 2	GU818575.1
Lactuca canadensis 3	KP828829.1
Lactuca dissecta	KF485649.1
Lactuca dolichophylla	KF485650.1
Lactuca floridana 1	HQ161957.1
Lactuca floridana 2	KP828827.1
Lactuca formosana 2	KF485655.1
Lactuca graminifolia 1	HQ161958.1
Lactuca graminifolia 2	KP828830.1
Lactuca hirsuta	HQ172901.1
Lactuca indica 2	AJ228634.1
Lactuca indica 3	AY862579.1
Lactuca indica 4	KF485653.1
Lactuca inermis 2	KF485652.1
Lactuca orientalis 3	KF485659.1
Lactuca perennis 2	L48143.1
Lactuca perennis 3	AJ228636.1
Lactuca perennis 4	AJ633334.1
Lactuca quercina	AJ228623.1
Lactuca raddeana 2	KF485657.1
Lactuca saligna 1	AJ228618.1
Lactuca saligna 2	HQ161960.1
Lactuca sativa 2	L13957.1
Lactuca serriola 3	AJ633331.1
Lactuca serriola subsp. *integrifolia*	AB742457.1
Lactuca sibirica 1	AJ228624.1
Lactuca sibirica 2	KF485660.1
Lactuca tatarica 2	AJ228629.1
Lactuca tenerrima 2	AJ228642.1
Lactuca tuberosa 1	AJ228645.1
Lactuca tuberosa 2	KF485651.1
Lactuca undulata 1	KF485647.1
Lactuca undulata 2	KF485648.1
Lactuca viminea 1	AJ228627.1
Lactuca viminea 2	AJ633333.1
Lactuca virosa 2	AJ228613.1
Melanoseris atropurpurea	KF485633.1
Melanoseris bracteata	KF485607.1

Melanoseris cyanea 1	KF485617.1
Melanoseris cyanea 2	KF485615.1
Melanoseris likiangensis	KF485632.1
Melanoseris macrorhiza	KF485608.1
Melanoseris qinghaica 1	KF485606.1
Melanoseris qinghaica 2	KF485613.1
Melanoseris violifolia	KF485611.1
Parasyncalathium souliei 1	KF485604.1
Parasyncalathium souliei 2	KF485605.1
Prenanthes purpurea 2	AJ228655.1
Prenanthes purpurea 3	KF485548.1

Chapter 4

A mixed model QTL analysis for salt tolerance in seedlings of crop-wild hybrids of lettuce

Zhen Wei[1], Magdalena M. Julkowska[2], Jacques-Olivier Lalo ë[3], Yorike Hartman[3], Gert-Jan de Boer[4], Richard W. Michelmore[5], Peter H. van Tienderen[3], Christa Testerink[2], M. Eric Schranz[1]

1 Biosystematics Group, Wageningen University, 6708 PB Wageningen, The Netherlands

2 Swammerdam Institute for Life Sciences, University of Amsterdam, 1090 GE Amsterdam, The Netherlands

3 Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, 1090 GE Amsterdam, The Netherlands

4 EnzaZaden, Enkhuizen, The Netherlands

5 The Genome Center& Department of Plant Sciences, University of California, Davis, CA 95616-8816, USA

Published in *Molecular Breeding* 34(3): 1389-1400

Chapter 4

Abstract

Cultivated lettuce is more sensitive to salinity stress than its wild progenitor species potentially due to differences in root architecture and/or differential uptake and accumulation of sodium. We have identified quantitative trait loci (QTLs) associated with salt-induced changes in Root System Architecture (RSA) and ion accumulation using a recombinant inbred line population derived from a cross between cultivated lettuce (*Lactuca sativa* 'Salinas') and wild lettuce (*L. serriola*). Components of RSA were quantified by replicated measurements of seedling growth on vertical agar plates containing different concentrations of NaCl in a controlled growth chamber environment. Accumulation of sodium and potassium ions was measured in replicates of greenhouse-grown plants watered with 100 mM NaCl water. A total of fourteen QTLs were identified using multi-trait linkage analysis, including three major QTLs associated with general root development, root growth in salt stress condition, and ion accumulation. The three major QTLs, qRC9.1, qRS2.1 and qLS7.2, were linked with markers E35/M59-F-425, LE9050 andLE1053 respectively. This study provides regions of lettuce genome contributing to salt-induced changes in Root System Architecture (RSA) and ion accumulation. Future fine-mapping of major QTLs will identify candidate genes underlying salt stress tolerance in cultivated lettuce.

Key words: lettuce; salt stress; root system architecture; crop-wild hybrids; QTLs

Introduction

In the face of increasing salinization of agricultural regions and global climate change, improving salt tolerance of crops could contribute to food production and the sustainability of agricultural systems (Flowers 2004; Munns et al. 2006). Quantitative Trait Loci (QTLs) for salt tolerance have been reported in a wide-range of crops; for example, wheat (Díaz De León et al. 2011), rice (Thomson et al. 2010;Wang et al. 2012a; Wang et al. 2010), soybean (Lee et al. 2004; Tuyen et al. 2010), barley (Nguyen et al. 2012) and tomato (Foolad et al. 2001). For many of these species, candidate genes have been identified allowing for the selection of superior alleles to increase salt tolerance (Ren et al. 2005). Superior alleles can often be identified in the close-relatives of the domesticated species and then be introgressed into the crop species (Shahbaz and Ashraf 2013).

Plant adaptions to salinity-stress fall into three distinct types of biological processes: osmotic stress tolerance enabling efficient water potential maintenance within the plant tissues; Na^+ or Cl^- exclusion preventing damage to photosynthetic tissues; and tissue tolerance ensuring minimal ion toxicity in cytosol (Munns and Tester 2008). Phenotypic traits conventionally used in QTL mapping for salt stress tolerance correspond to the osmotic stress acclimation or responses specific to salinity stress (Munns 2010). Traits associated with growth, like root and leaf elongation (Mano and Takeda 1997), are considered to alter in response to the osmotic effects of salt stress. Changes in biomass production on the other hand, including fresh and dry weight (Wang et al. 2012b) and yield (Genc et al. 2010), are likely to be affected by osmotic as well as ionic components of salt stress (Munns 2010). The ion content in shoots and roots are definitive traits measured as a result of salt-specific effects (Munns 2010). Germination and survival rate can also be used to detect QTLs for salt tolerance (DeRose-Wilson and Gaut 2011; Lin et al. 2004; Zhou et al. 2011). In most studies, not only responses of phenotypic traits to osmotic effects were measured, but also traits related to salt-specific effects (Genc et al. 2013; Genc et al. 2010; Nguyen et al. 2012; Uwimana et al. 2012a; Vallejo et al. 2010; Wang et al. 2012b).

Salt stress is widely documented to inhibit root growth. But this inhibition is not equal for primary and lateral roots since the primary root is less sensitive to salt-induced growth inhibition (Geng et al. 2013). Changes in Lateral Root (LR) emergence under mild salinity stress were observed in *Arabidopsis*(Zolla et al. 2010), yet the directionallity of this response remains ellusive (Galvan-Ampudia and Testerink 2011). A number of root traits were studied for QTL mapping, but the main focus remained on traits primarily correlated with primary root traits, such as primary/total root length, primary/total root weight and primary root diameter (Sharma et al. 2011; Vaughn and Masson 2011; Zhu et al. 2005). These root traits can be referred to as Root system architecture (RSA), which indicates the spatial configuration of the root system in the soil and plays an important role in plant productivity (Lynch 1995). Cultivated lettuce with a larger RSA may perform better in harsh conditions

(Kerbiriou et al. 2013a). Changes in RSA and water capture can directly affect and enhance plant growth rate and biomass accumulation in maize (Hammer et al. 2009). Changes in RSA are also 'feed-forward' mechanisms to maintain resource capture under limiting water and Nitrogen supply for two cultivars of lettuce (*Lactuca sativa* L.) (Kerbiriou et al. 2013b). Expanding quantification of RSA to include measurements of LR Length and calculations of the density per primary root and branching zone were recently advocated to increase our understanding of RSA regulation(De Smet et al. 2012; Dubrovsky and Forde 2012). Therefore, RSA should be an informative character to measure when salt tolerance is considered in plants. In recent years, two-dimensional and three-dimensional gel-based imaging platforms have been used to study RSA. These platforms are advantageous since they make RSA visible and avoid root damage (Fang et al. 2009; French et al. 2009; Iyer-Pascuzzi et al. 2010). Multiple RSA traits can be studied in great detail using several software packages such as EZ-Rhizo and GiA Roots that have been developed to extract and quantify root traits from images (Armengaud et al. 2009; Galkovskyi et al. 2012).

The primary gene pool of cultivated lettuce includes not only *L. sativa*, but also the species *L. serriola*, *L. altaica* and *L. aculeata*. These species can all be easily crossed to *L. sativa* (Koopman et al. 1998; Koopman et al. 2001). The three species along with *L. saligna* and *L. virosa* are closely related to cultivated lettuce and considered important resources for lettuce breeding (Lebeda et al. 2009; Schwember and Bradford 2010a; van Treuren et al. 2011; Zhang et al. 2009). QTL studies of lettuce have been associated with resistance to pathogens and pests, for instance, downy mildew (Jeuken and Lindhout 2002; Jeuken et al. 2008; Zhang et al. 2009). QTLs for a number of beneficial traits, such as 107 QTLs for shelf life (Zhang et al. 2007), 13 QTLs for RSA and deep soil water exploitation (Johnson et al. 2000), 17 QTLs for seed and seedling traits related to germination (Argyris et al. 2005), and 76 QTLs for domestication traits (Hartman et al. 2012b), have also been reported. A previous QTL study of lettuce grown in salt stress conditions, focused on plant vigour-related traits and salt content in shoot tissue (Uwimana et al. 2012a; Uwimana et al. 2012b). QTLs underlying RSA responses of lettuce seedlings in salt stress have not yet been reported. Additionally, it has been demonstrated that *L. sativa* has a shallower root system and higher plasticity of its roots in the surface soil zone than *L. serriola* while the latter is more drought tolerant (Gallardo et al. 1996; Jackson 1995). Therefore an interspecific cross between these two species is appropriate for mapping QTLs associated with salt tolerance.

In this study, we used a 2-D imaging platform to analyze RSA of lettuce seedlings of a recombinant inbred line population in agar plates as well as measured the salt content in leaves of seedlings grown in a greenhouse to identify QTLs. Most QTL detections are performed by analysis of designed segregating populations derived from two inbred parental lines, where absence of selection, mutation and genetic drift is assumed. This assumption leads to unclear QTL locations and an unrealistically high number of marker-trait associations when kinship and coancestry information is ignored (Malosetti et al. 2011).Thus we chose a

mixed model QTL analysis to detect QTLs (multi-trait linkage analysis and single trait linkage analysis in single environment) and minimize false QTLs, instead of Simple Interval Mapping (SIM) and Composite Interval Mapping (CIM) (Malosetti et al. 2011; van Eeuwijk et al. 2010). Our aim was to find parental contributions to the traits of interest and to identify candidate genomic regions to improve the salt tolerance of lettuce.

Materials and methods

Plant materials

We used a Recombinant Inbred Line (RIL) population derived from a cross between cultivated lettuce (*L. sativa* 'Salinas') ('Salinas' is the name of the growing region in California for which it was bred rather than being named because it is salt-tolerant) and wild lettuce (*L. serriola*) (UC96US23), self-pollinated for nine generations (Argyris et al. 2005; Johnson et al. 2000; Truco et al. 2007; Truco et al. 2013).The two parents of the RIL population differ in RSA and the efficiency of exploitation of soil water and nutrients. Cultivated lettuce has a shallow root system and higher plasticity of its roots than wild lettuce in the surface soil zone. In contrast, wild lettuce has a deeper, more productive root system and extract more soil water from depth than crop lettuce (Gallardo et al. 1996; Jackson 1995). This is a core mapping population that is being used for diverse QTL analyses and for which there is an ultra-high density genetic map (Truco et al. 2013). For example, it has been used in previous studies of QTL mapping for domestication traits (Hartman et al. 2012b), shelf life (Zhang et al. 2007), root architecture and deep soil water exploitation (Johnson et al. 2000) as well as seed and seedling traits related to germination (Argyris et al. 2005). Fifty-nine highly informative RILs were selected by MapPop 1.0, from a set of 356 F7:8 RILs, to maximize the number of recombinants and reduce the time and expenses (Vision et al. 2000).

Trait measurement

Trait measurements of RSA of lettuce seedlings were made in the climate chamber. Seeds were first germinated in vertical square agar plates (70 ° angle) in growth chamber, under 16-hour light and 8-hour dark, 22 ℃ and 70% humidity. Themedium composition was ½ Murashi-Skoog (MS), 0.5% sucrose and 0.1% MES (pH 5.8 KOH). Four-day-old seedlings were then transferred to non-stress and salt plates, containing 0 mM, 75 mM and 150 mMNaCl respectively. Digital images of all plates were taken with a scanner (EPSON perfection V700) on day 0 (transferring day) and day 8. Four individual plants were selected randomly for data extraction. The RSA image data were quantified using the EZ-Rhizo software package 1.0 (Armengaud et al. 2009). The RSA traits included Main Root Length (MRL), Main Root Angle (MRA), the length of Branched Zone (BZ), Lateral Root Density per cm in Branched Zone (LRD/BZ), total Lateral Root Length (LRL) and Lateral Root Number (LRN).

Ion concentrations in lettuce leaves were measured in greenhouse grown plants. The RILs were first germinated in petri dishes containing filter paper and demineralized water in a climate chamber, under 16-hour light and 8-hour dark, at 20 ℃ and 10 ℃ respectively. One-week-old seedlings were then transplanted into pots containing vermiculite moistened with water containing 1g/L fertilizer (POKON) in the greenhouse, under 16-hour light and 8-hour dark, at 20 ℃ and 16 ℃ respectively. After two weeks, three-week-old seedlings were watered with 100 mM NaCl water, twice a day for three days in Week 4 and once a day for two days in Week 5. Four plants were chosen randomly for ion concentration measurement. Three punches from two different leaves of each plant were taken in Week 6 and then washed in purified water under light in a shaker for 30 min. After, they were transferred to wells containing 3 mL of 0.01% silwet-L77 solution, vacuum-infiltrated and incubated for 1 hour under light in a shaker (100 rpm). The conductivity of the samples was determined with the Horiba Twin Cond. Subsequently, the samples were cooked in a microwave to break up the cells so that total electrolyte content could be measured in the same way. Na^+ content (NAC) and K^+ content (KC) were then measured. The ratio between Na^+ content and K^+ content (NA/K) was calculated and used for testing QTLs as well.

Quantitative trait loci analysis

Mean phenotypic data of the RILs from the different environments were used to detect QTLs. The genetic map and marker data of the RILs used in our QTL analysis were generated as part of The Compositae Genome Project and are available from http://compgenomics.ucdavis.edu(Truco et al. 2007). The genetic map was composed of 1,513 predominantly AFLP and EST markers distributed over the nine lettuce chromosomes (http://cgpdb.ucdavis.edu/GeneticMapViewer/display/; map version: RIL_MAR_2007_ratio) (Truco et al. 2007).

QTL analysis was performed by GenStat 15th version and is available from http://www.vsni.co.uk/software/genstat(Payne et al. 2012). Multi-trait linkage analysis (single environment) and single trait linkage analysis (single environment) were chosen to detect QTLs. A significant QTL effect ($\alpha = 0.05$) at particular genome positions is associated with a low P value (rejection of null hypothesis of no QTL), which is graphically shown on a –log10 scale to resemble the typical LOD profile plot (Pastina et al. 2012). Genetic predictors estimated from marker information were calculated with a step size of 5cM. Two main steps were then taken: (1) genome wide QTL scan, testing first one QTL at a time (Simple Interval Mapping, SIM) followed by Composite Interval Mapping (CIM), cofactors were included at positions where there was evidence for QTLs; (2) fitting a multi-QTL model after backward selection from the set of candidate QTLs found significant in an earlier genome scan to estimate QTL locations and effects (Griffiths et al. 2010; Huang et al. 2010; Malosetti et al. 2007; Pastina et al. 2012). Multi-trait linkage analysis differs from single trait analysis in that phenotypes of all traits are simultaneously used to test for a QTL showing an effect on at least one of the traits (implying a pleiotropic model if the QTL has an effect on more than one trait).

The additive effect, standard error, high value allele, the Percentage of Explained Variance (PVE) (given by per trait at each QTL position in Multi-trait Linkage Analysis) and positions of the QTLs were estimated and used to determine what traits were affected by the specific QTL. Genomic locations of QTLs were displayed by MapChart4.0 (Voorrips 2002).

A QTL can be described as a 'major' or 'minor' QTL, based on the percentage of the phenotypic variation explained by a QTL (Collard et al. 2005). QTLs in this study were classified by the PVE values (maximum PVE value among all the traits at each position for multi-trait linkage analysis) and defined as major QTL (PVE>25%), intermediate QTL (PVE between 10% and 25%), and minor QTL (PVE<10%) (Burke et al. 2002).

Results

Phenotypic distribution

The phenotypic RSA traits varied substantially among the 59 RILs in the control, 75 mM NaCl and 150 mM NaCl conditions (**Figure 1**).In the control condition, the continuous distribution for all the traits implied that RSA traits were quantitatively inherited in nature (**Figure S1a**). The crop parent - *L. sativa* 'Salinas' showed lower values than the wild parent - *L. serriola* in all the trait distributions except in LRD/BZ (**Figure S1a**). In the 75 mM NaCl condition, the wild parent had higher values than the crop parent in all trait distributions. More precisely, *L. serriola* was about twice the values of *L. sativa* 'Salinas' for BZ, LRL, LRN and MRL traits (**Figure S1b**). In the 150 mM NaCl condition, the phenotypic distribution showed a continuous pattern for all RSA traits except MRL (**Figure S1c**). More than 90% of the 59 RILs had a MRL value between 0 to 2 cm. One line even got a mean MRL value of -0.01 cm, which might be caused by measure error.The phenotypic ion accumulation traits in the 100 mM NaCl condition showed no bimodal distribution. The frequency of KC indicated a normal distribution while that of NAC and NA/K decreased with increasing ion concentration. *L. serriola* had slightly higher KC than *L. sativa* 'Salinas' (**Figure S1d**). In contrast, the crop parent was about triple the amount of the wild parent in NAC and NA/K.

Multi-trait and single trait linkage analysis

We identified a total of fourteen QTLs for nine traits related to changes induced by salt in RSA and ion accumulation in control, 75 mM and 100 mM NaCl conditions, using multi-trait linkage analysis (**Table 1; Figure 2**). The QTLs were distributed over seven of the nine linkage groups (none were found on LG4 and LG6). Most QTLs were found in a single environment except qRC9.1 and qRS9.2, which were overlapping and found in control and salt environments, respectively (Overlapping QTLs from different experiments are shown in tables and figures but we acknowledge that they may represent a single locus) (**Table 1**). Three major QTLs (PVE > 25%) and three minor QTLs (PVE < 10%) were detected. Other QTLs were intermediate QTLs (PVE between 10% and 25%). No QTLs were detected in the 150 mM NaCl root growth condition (a very high salt concentration for *Lactuca* spp.). One

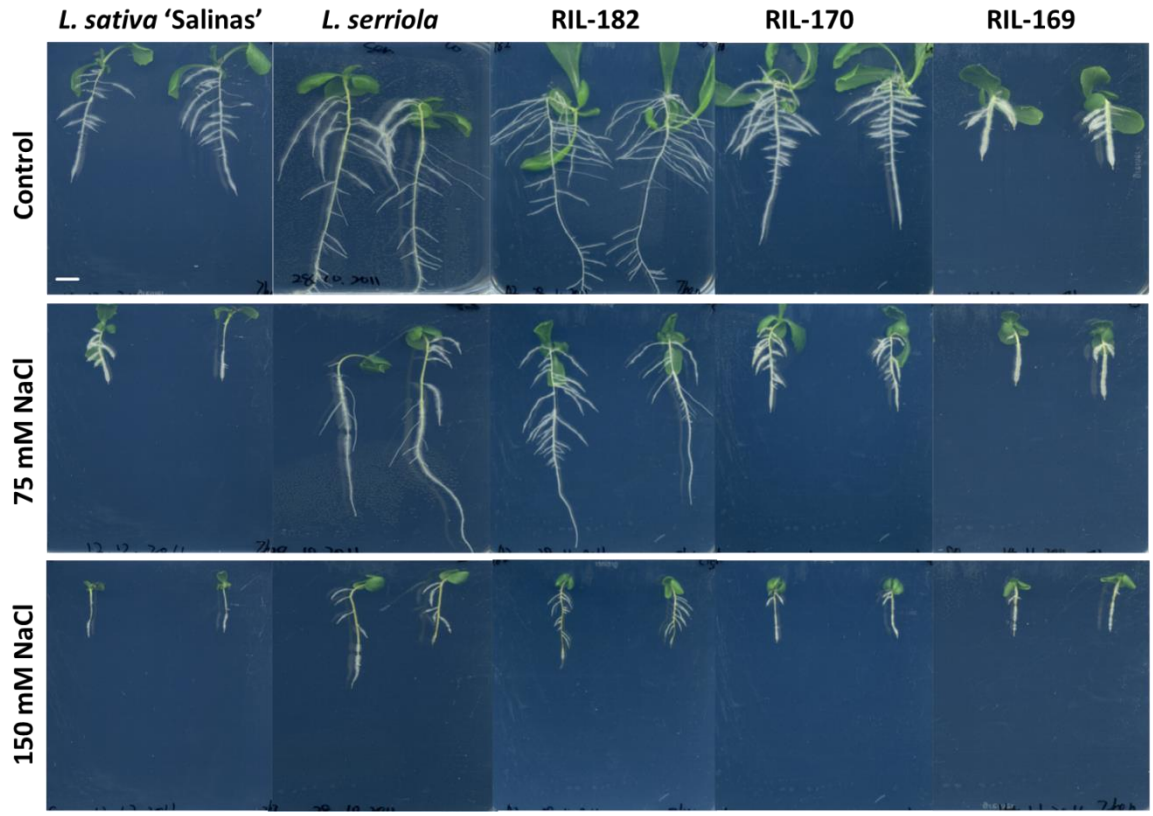


Figure 1 Twelve-day-old lettuce seedlings of parental lines and three recombinant inbred lines in agar plates in the non-salt, 75 mM NaCl and 150 mM NaCl conditions respectively; the selected RILs display large variations in RSA; White line indicates 1 cm

major QTL for RSA, qRS2.1, on LG2 was identified in the 75 mM salt environment while another major QTL for RSA, qRC9.1, on LG9 was detected in the non-salt condition and overlapped with qRS9.2 in the 75 mM salt condition (**Table 1**). The third major QTL on LG7 for ion content of leaves, qLS7.2, was found in the salt environment (**Table 1**).

In single trait linkage analysis, we detected eight QTLs associated with responses in LRL, LRN and ion contents in control, 75 mM, 100 mM and 150 mM NaCl conditions (**Table 2**). All single-trait QTLs overlapped with QTLs detected by multi-trait linkage analysis (**Table 2**). Five of single-trait QTLs were major QTLs and overlapping with the three major QTLs (qRC9.1, qRS9.2 and qLS7.2) and one intermediate QTL (qRC7.1) found by multi-trait linkage analysis. Additionally, all the single-trait QTLs related to ion accumulation in leaves were at the same position as major QTL qLS7.2, found by multi-trait linkage analysis.

Correlation analysis in multi-trait linkage analysis

In the multi-trait linkage analysis, all RSA traits were simultaneously used to test for a QTL showing an effect on at least one of these RSA traits. Therefore, all the traits showed different contributions to the same QTLs in the results. In the control condition, LRL, LRN, BZ and MRL had significant positive correlations with each other (**Table S1**). In the 75 mM NaCl condition, MRA had negative correlations with MRL and LRD/BZ (<0.05) (**Table S2**). Apart

Table 1 QTLs detected by multi-trait linkage analysis in a *L. sativa* 'Salinas' × *L. serriola* recombinant inbred line population

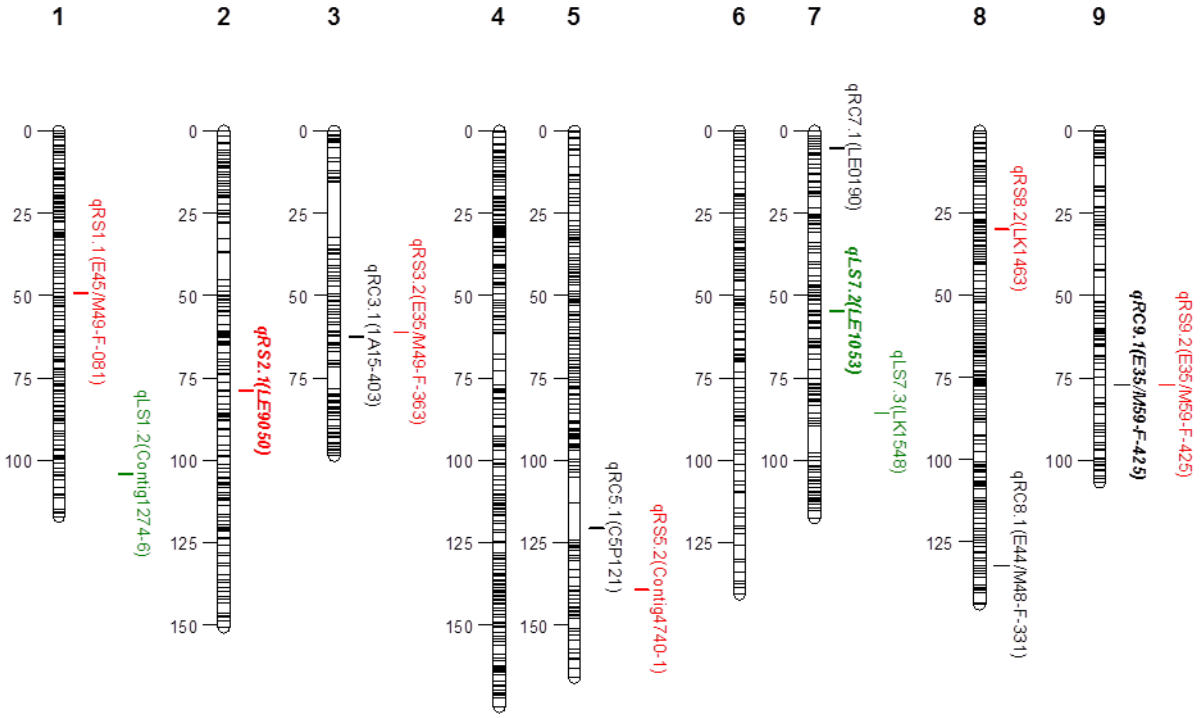
QTL name	E (mM NaCl)	Tissue	Nearest marker	Marker name	LG	Pos. (cM)	-log10(p)
qRC3.1	0	Root	421	1A15-403	3	62.47	3.4
qRC5.1	0	Root	892	C5P121[#]	5	120.60	4.1
qRC7.1	0	Root	1078	LE0190	7	5.26	2.6
qRC8.1	0	Root	1384	E44/M48-F-331	8	132.16	2.3
qRC9.1	0	Root	1491	E35/M59-F-425[b,c]	9	77.24	8.0
qRS1.1	75	Root	89	E45/M49-F-081	1	49.15	4.3
qRS2.1	75	Root	268	LE9050[c]	2	78.84	6.7
qRS3.2	75	Root	417	E35/M49-F-363	3	61.20	4.3
qRS5.2	75	Root	907	Contig4740-1	5	139.32	13.6
qRS8.2	75	Root	1242	LK1463	8	30.01	4.2
qRS9.2	75	Root	1491	E35/M59-F-425[b]	9	77.24	7.5
qLS1.2	100	Leaf	168	Contig1274-6	1	104.19	5.3
qLS7.2	100	Leaf	1132	LE1053[c]	7	54.64	46.1
qLS7.3	100	Leaf	1170	LK1548	7	85.70	4.3

qRC QTLs for RSA traits in control condition; *qRS* QTLs for RSA traits in 75 mM NaCl condition; *qLS* QTLs for ion accumulation in 100 mM NaCl condition, *E* environment, *Pos.* position, *LG* linkage group

[a] QTL position instead of a marker's name

[b] QTL has been detected in two environments

[c] Major QTL



◄ **Figure 2** Genomic locations and adjacent markers of quantitative trait loci for root and ion concentrations in different conditions; the colours indicate different environments: *black* non-salt for roots, *red* 75 mM NaCl for roots and *green* 100 mM NaCl for ion content in leaves; *bold italic* letters signify major QTLs; C5P121 is a QTL position instead of a marker's name.

Table 2 QTLs detected by single trait linkage analysis in a *L. sativa* 'Salinas' × *L. serriola* recombinant inbred line population

Trait name	*E*	Marker name	Overlaps	Pos. (cM)	LUB	-Log10(*p*)	AE	SE	PVE (%)
LRL	0	E35/M59-F-425	qRC9.1	77.24	0-106.7	3.1	-7.54	2.12	17.7
LRN	0	E35/M59-F-425	qRC9.1	77.24	0-106.7	2.8	-4.09	1.23	17.4
LRL	75	E35/M59-F-425	qRS9.2	77.24	57.5-97.0	5.1	-2.95	0.60	31.4[a]
LRN	75	E35/M59-F-425	qRS9.2	77.24	55.4-99.1	4.7	-2.78	0.59	29.4[a]
LRL	150	1A21-233	qRC7.1	3.00	0-39.1	3.5	-0.60	0.16	22.4
K	100	LE1053	qLS7.2	54.64	45.3-64.0	9.3	1.63	0.22	52.3[a]
NA	100	LE1053	qLS7.2	54.64	47.9-61.4	13.8	-3.68	0.36	68.6[a]
NA/K	100	LE1053	qLS7.2	54.64	47.9-61.4	13.9	-0.48	0.05	68.7[a]

E environment (mM NaCl), *Pos.* position, *LUB* lower-upper bounds, *AE* additive effect, *SE* standard error, *PVE* the percentage of explained variances; negative AE means effect is from *L. sativa* 'Salinas', and positive AE means effect is from *L. serriola*

[a] Major QTLs; significance level alpha=0.05

from that, all other traits had positive correlations with each other.Accumulation of KC ions was found to be negatively correlated with NAC and the NA/K ratio in 100 mM NaCl (**Table S3**).

Direction of additive effect

In multi-trait linkage analysis, LRL and LRN had high negative additive effect, increased by alleles from the crop parent, especially at qRC5.1 (-7.67 and -4.26) and qRC9.1 (-8.59 and -4.98) (**Figure 3a**) in the control condition. However, these two traits also had high positive additive effect from alleles of the wild parent at aRC3.1, qRC7.1 and qRC8.1. In contrast, the other four traits had low additive effect at all QTL positions whether the effect was positive or negative (**Figure 3a**). In the 75 mM NaCl treatment, LRL and LRN contributed relatively higher additive effect, especially at the two major QTL positions, 2.70 and 2.01 at qRS2.1, -2.28 and -2.13 at qRC9.2. In the 100 mM NaCl condition, the high value allele for KC at all QTL positions was from the wild parent (**Figure 3b**). Conversely, the high value alleles for NAC and NA/K came from the cultivated parent (**Figure 3b**). The additive effect of KC, NA/K and NAC at the major QTL position, qLS7.2, were 1.58, 0.44 and 3.39 respectively (**Figure 3b**).

In single trait linkage analysis, the crop parent increased the additive effect of QTLs for RSA, NAC and NA/K traits. The wild parent increased the additive effect of QTLs for KC.

These results confirmed the direction of effect detected by multi-trait linkage analysis (**Table 2**; **Table S4**).

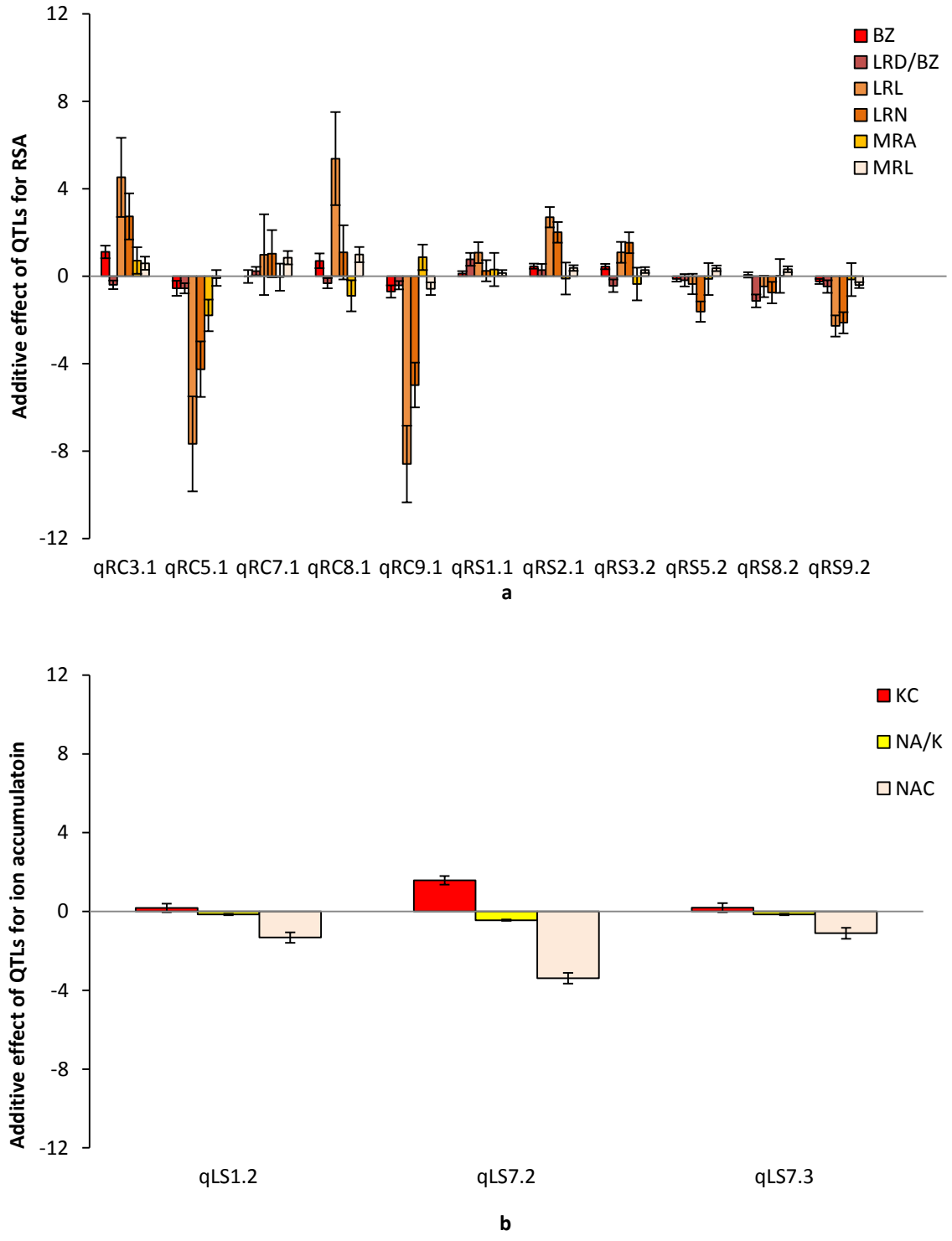


Figure 3 Additive effect at each QTL location: **a** QTLs for RSA in the control and 75 mM NaCl condition; **b** QTLs for ion accumulation in the 100 mM NaCl condition; positive effect means that the alleles from the wild parent *L. serriola* increase the trait values, negative effect means that the alleles from the crop parent *L. sativa* 'Salinas' increase the trait values; error bars represent stand error (n=4)

Major QTLs under control and salt stress

In multi-trait linkage analysis, the major QTL qRS2.1 on LG2 at position 78.84 cM was only found in the salt-related environment (**Table S4**). LRL explained the highest PVE (26.3%) among all the traits at this locus. In addition, alleles from the wild species, *L. serriola*, contributed positive effect to most traits at this locus, while alleles from *L. sativa* 'Salinas' had low effect to MRA. The PVEs of the other major root QTL qRC9.1 in control condition were higher than those of qRS9.2 in salt environment. LR traits were the main contributors to qRC9.1 since PVE of LRN and LRL was 25.7% and 23.0% respectively (**Table S4**). All the high value alleles at this locus were from the crop parent - *L. sativa* 'Salinas', except MRA in the non-salt condition. The PVE values of qLS7.2 were significantly different from other QTLs, 49.2% for KC, 57.9% for NA/K and NAC (**Table S4**).

In single trait analysis, the –log10(p) values of the five major QTLs varied from 4.7 to 13.9. The PVE values of QTLs for LRL and LRN in the 75 mM NaCl condition were 31.4% and 29.4%. The PVE values of QTLs for KC, NAC and NA/K were 52.3%, 68.6% and 68.7% respectively.

Discussion

Our study revealed genomic regions associated with RSA and ion content in leaves of lettuce seedlings in response to salinity. In multi-trait linkage analysis, we detected a total of 14 QTLs for 9 traits, 11 QTLs for RSA traits and 3 QTLs for ion accumulation in leaves. However, no QTLs were found in the 150 mM NaCl condition, probably due to the lack of variation between the RILs in this very high-salt condition. In single trait linkage analysis, we found eight QTLs for LRL, LRN and ion accumulation in all conditions. As all of them were overlapping with QTLs found in multi-trait linkage analysis, we mainly focus on the QTLs detected in multi-trait linkage analysis in further discussion.

We discovered three major QTLs using multi-trait linkage analysis (single environment). Two of the three major QTLs, qRS2.1 and qLS7.2, were identified in the 75 mM NaCl (agar plate grown seedlings) and 100 mM NaCl (greenhouse grown plants) conditions respectively. So we consider them as salt-specific QTLs or potential 'adaptive' QTLs (Collins et al. 2008), meaning they are detected only in specific environmental conditions or increased in expression with the level of an environmental factor. QRC9.1 was found in the non-stress condition and overlapped with qRS9.2 in the 75 mM NaCl condition, which implies that it is related to general root growth. Therefore we consider qRC9.1 to be a stable QTL, which is consistently detected across multiple environments (Collins et al. 2008). It should be pointed out that two closely located QTLs on LG3, qRC3.1 and qRS3.2, were found across salt and non-salt conditions. They could potentially be a stable QTL as well as qRC9.1 and qRS9.2.

Several QTLs found in our study are consistent with results from previous studies. The major QTL qRS2.1 was at nearly the identical position as QTLs reported for two traits related to root water acquisition of lettuce in field **(**Johnson et al. 2000**)**. Four QTLs in our study

overlapped with QTLs detected for longevity of lettuce seeds using the same population (Schwember and Bradford 2010b). QLS7.2 was also in the LOD interval of two QTLs for Na^+ and K^+ in a F_2 population from *L. sativa* and *L. serriola* (Uwimana et al. 2012a). Three QTLs were co-located with those found in QTL mapping for domestication and fitness related traits of lettuce using the same RILs, including major QTL qRC9.1 (Hartman et al. 2012a; Hartman et al. 2012b).

RSA in response to salt stress

In multi-trait linkage analysis, MRL had relatively evenly distributed effect values in control and salt conditions, while LRL and LRN contributed much more to the additive effects in the control condition than in the salt condition. In single trait linkage analysis, only QTLs for LRL and LRN were detected both in control and salt conditions. These results imply that lateral roots may be more sensitive to moderate salt stress than primary roots. This observation is similar to that made in a study of lateral root development under low salt stress in *Arabidopsis*, which reported that primary root elongation was not sensitive to low and moderate osmotic stress while lateral root development was very sensitive to low osmotic and ionic stress (Zhao et al. 2011).

Direction of effect at major QTL locations

In multi-trait linkage analysis, the allelic effect at one QTL position was generally increased by either crop allele or wild allele. However, this was not always the case. The allelic effect at one QTL position was sometimes from different parents according to the phenotypic traits used to test that QTL. For example, at some QTL positions, such as qRS2.1, MRA showed opposite direction of additive effect from the other five traits due to a negative correlation between them.

The crop alleles from *L. sativa* 'Salinas' increased the effect at qRC9.1 (qRS9.2), especially for the total LRL and LRN, which indicates that this major QTL is related to general root development. The wild allele from *L. serriola* increased the effect at qRS2.1, also for LRL and LRN, implying that this QTL might be correlated with LR development in salt stress condition. *L. sativa* 'Salinas' produced more lateral roots, a greater total of root length and more external links than *L. serriola* in the top soil zone (0 to 5 cm) (Jackson 1995). In contrast, dried soil in the top zone (0 to 20 cm) had no effect on biomass production in *L. serriola* but reduced final shoot production in *L. sativa* 'Salinas' (Gallardo et al. 1996). The two major QTLs for RSA identified in this study confirm the importance of these regions in RSA development and the effect of the allele were in the directions as expected. However, it had been reported that the crop allele increased the effect of two QTL regions associated with deep soil water exploitation (Johnson et al. 2000) and those regions, regarding to g H_2O per m^3 and percentage total H_2O at 25–50 cm, were at the same position as qRS2.1. As our study focused on salt induced changes in RSA of young seedlings, rather than root biomass

distribution over specific soil depths in older plants, we believe our results do not conflict with the previous data and confirm the major QTL region on LG2 for RSA.

The third major QTL associated with ion accumulation, qLS7.2, showed different directions of allelic effects between KC, NAC and NA/K. Negative correlation between KC and other two ion accumulation traits has been observed. Additionally, the wild allele contributed to the additive effect of KC whereas the crop allele increased the additive effect of NAC and NA/K. These results are similar to a study in rice, where the allele from the salt-tolerant parent increased KC while the allele from salt-susceptible parent increased NAC in shoot (Lin et al. 2004). QTLs for vigour of crop-wild hybrids of lettuce under drought, salinity and nutrient deficiency conditions also indicated that the wild allele increased additive effect for KC while the crop allele increased NAC in salt condition (Uwimana et al. 2012a).

This work is the first QTL analysis of salt tolerance in lettuce seedlings, using 2D-imaging gel system for RSA characters and salt content in leaves. Our study suggests candidate genomic regions for improving salt tolerance of cultivated lettuce and determining the changes in RSA of lettuce in response to salinity. In the future, we will fine-map the major QTL regions and validate them by backcrossing to generate isogenic lines.

Acknowledgments

Thanks to China Scholarship Council (http://en.csc.edu.cn/) for providing funding to ZW. The contributions from the Compositae Genome Project were supported by the National Science Foundation Plant Genome Program Grant # DBI0421630 to RWM. MM and CT were funded by the STW Learning from Nature program grant 10987 and NWO ALW grants 820.02.017 and 846.11.002 awarded to CT.

References

Argyris J, Truco MJ, Ochoa O, Knapp SJ, Still DW, Lenssen GM, Schut JW, Michelmore RW, Bradford KJ (2005) Quantitative trait loci associated with seed and seedling traits in *Lactuca*. Theor Appl Genet 111:1365-1376

Armengaud P, Zambaux K, Hills A, Sulpice R, Pattison RJ, Blatt MR, Amtmann A (2009) EZ-Rhizo: integrated software for the fast and accurate measurement of root system architecture. Plant J 57:945-956

Burke JM, Tang SX, Knapp SJ, Rieseberg LH (2002) Genetic analysis of sunflower domestication. Genetics 161:1257–1267

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169-196

Collins NC, Tardieu F, Tuberosa R (2008) Quantitative trait loci and crop performance under abiotic stress: where do we stand? Plant physiol 147:469-486

De Smet I, White PJ, Bengough AG, Dupuy L, Parizot B, Casimiro I, Heidstra R, Laskowski M, Lepetit M, Hochholdinger F, Draye X, Zhang H, Broadley MR, Peret B, Hammond JP, Fukaki H, Mooney S, Lynch JP, Nacry P, Schurr U, Laplaze L, Benfey P, Beeckman T, Bennett M (2012) Analyzing lateral root development: how to move forward. Plant Cell 24:15-20

DeRose-Wilson L, Gaut BS (2011) Mapping salinity tolerance during *Arabidopsis thaliana* germination and seedling growth. PLoS ONE 6:e22832

Díaz De León JL, Escoppinichi R, Geraldo N, Castellanos T, Mujeeb-Kazi A, Röder MS (2011) Quantitative trait loci associated with salinity tolerance in field grown bread wheat. Euphytica 181:371-383

Dubrovsky JG, Forde BG (2012) Quantitative analysis of lateral root development: pitfalls and how to avoid them. Plant Cell 24:4-14

Fang S, Yan X, Liao H (2009) 3D reconstruction and dynamic modeling of root architecture in situ and its application to crop phosphorus research. Plant J 60:1096-1108

Flowers TJ (2004) Improving crop salt tolerance. J Exp Bot 55:307-319

Foolad MR, Zhang LP, Lin GY (2001) Identification and validation of QTLs for salt tolerance during vegetative growth in tomato by selective genotyping. Genome 44:444-454

French A, Ubeda-Tomas S, Holman TJ, Bennett MJ, Pridmore T (2009) High-throughput quantification of root growth using a novel image-analysis tool. Plant physiol 150:1784-1795

Galkovskyi T, Mileyko Y, Bucksch A, Moore B, Symonova O, Price CA, Topp CN, Pascuzzi ASL, Zurek PR, Fang S, Harer J, Benfey PN, Weitz JS (2012) GiA Roots: software for the high throughput analysis of plant root system architecture. BMC Plant Biol 12:116

Gallardo M, Jackson LE, Thmopson RB (1996) Shoot and root physiological responses to localized zones of soil moisture in cultivated and wild lettuce (*Lactuca* spp.). Plant Cell Environ 19:1169-1178

Galvan-Ampudia CS, Testerink C (2011) Salt stress signals shape the plant root. Curr Opin Plant Biol14:296-302

Genc Y, Oldach K, Gogel B, Wallwork H, McDonald GK, Smith AB (2013) Quantitative trait loci for agronomic and physiological traits for a bread wheat population grown in environments with a range of salinity levels. Mol Breeding 32:39-59

Genc Y, Oldach K, Verbyla AP, Lott G, Hassan M, Tester M, Wallwork H, McDonald GK (2010) Sodium exclusion QTL associated with improved seedling growth in bread wheat under salinity stress.Theor Appl Genet 121:877-894

Geng Y, Wu R, Wee CW, Xie F, Wei X, Chan PM, Tham C, Duan L, Dinneny JR (2013) A spatio-temporal understanding of growth regulation during the salt stress response in *Arabidopsis*. Plant Cell 25:2132-2154

Griffiths S, Simmonds J, Leverington M, Wang Y, Fish L, Sayers L, Alibert L, Orford S, Wingen L, Snape J (2010) Meta-QTL analysis of the genetic control of crop height in elite European winter wheat germplasm. Mol Breeding 29:159-171

Hammer GL, Dong Z, McLean G, Doherty A, Messina C, Schussler J, Zinselmeier C, Paszkiewicz S, Cooper M (2009) Can changes in canopy and/or root system architecture explain historical maize yield trends in the U.S. corn belt? Crop Sci 49:299

Hartman Y, Hooftman DA, Uwimana B, van de Wiel CC, Smulders MJ, Visser RG, van Tienderen PH (2012a) Genomic regions in crop-wild hybrids of lettuce are affected differently in different environments: implications for crop breeding. Evol Appl 5:629-640

Hartman Y, Hooftman DAP, Schranz ME, van Tienderen PH (2012b) QTL analysis reveals the genetic architecture of domestication traits in Crisphead lettuce. Genet Resour Crop Ev 60:1487-1500

Huang X, Paulo M, Boer M, Effgen S, Keizer P, Koornneef M, van Eeuwijk FA (2010) Analysis of natural allelic variation in *Arabidopsis* using a multiparent recombinant inbred line population. P NATL ACAD SCI USA 108:4488-4493

Iyer-Pascuzzi AS, Symonova O, Mileyko Y, Hao Y, Belcher H, Harer J, Weitz JS, Benfey PN (2010) Imaging and analysis platform for automatic phenotyping and trait ranking of plant root systems. Plant physiol 152:1148-1157

Jackson LE (1995) Root architecture in cultivated and wild lettuce (*Lactuca* spp.). Plant Cell Environ 18:885-894

Jannink JL (2005) Selective phenotyping to accurately map quantitative trait loci. Crop Sci 45:901-908

Jeuken MJ, Lindhout P (2002) *Lactuca saligna*, a non-host for lettuce downy mildew (*Bremia lactucae*), harbors a new race-specific *Dm* gene and three QTLs for resistance. Theor Appl Genet 105:384-391

Jeuken MJ, Pelgrom K, Stam P, Lindhout P (2008) Efficient QTL detection for nonhost resistance in wild lettuce: backcross inbred lines versus F_2 population. Theor Appl Genet 116:845-857

Johnson WC, Jackson LE, Ochoa O, van Wijk R, Peleman J, St.Clair DA, Michelmore RW (2000) Lettuce, a shallow-rooted crop, and *Lactuca serriola*, its wild progenitor, differ at QTL determining root architecture and deep soil water exploitation. Theor Appl Genet 101:1066-1073

Kerbiriou PJ, Stomph TJ, Lammerts van Bueren ET, Struik PC (2013a) Influence of transplant size on the above- and below-ground performance of four contrasting field-grown lettuce cultivars. Front Plant Sci 4:379

Kerbiriou PJ, Stomph TJ, Putten PEL, Lammerts Van Bueren ET, Struik PC (2013b) Shoot growth, root growth and resource capture under limiting water and N supply for two cultivars of lettuce (*Lactuca sativa* L.). Plant Soil 371:281-297

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 squences. Am J Bot 85:1517-1530

Koopman WJM, Zevenbergen MJ, van den Berg RG (2001) Species relationships in *Lactuca* S.L. (Lactuceae, Asteraceae) inferred from AFLP fingerprints. Am J Bot 88:1881-1887

Lebeda A, Doležalová I, Křístková E, Kitner M, Petrželová I, Mieslerová B, Novotná A (2009) Wild *Lactuca* germplasm for lettuce breeding: current status, gaps and challenges. Euphytica 170:15-34

Lee GJ, Boerma HR, Villagarcia MR, Zhou X, Carter Jr TE, Li Z, Gibbs MO (2004) A major QTL conditioning salt tolerance in S-100 soybean and descendent cultivars. Theor Appl Genet 109:1610-1619

Lin HX, Zhu MZ, Yano M, Gao JP, Liang ZW, Su WA, Hu XH, Ren ZH, Chao DY (2004) QTLs for Na^+ and K^+ uptake of the shoots and roots controlling rice salt tolerance. Theor Appl Genet 108:253-260

Lynch J (1995) Root architecture and plant productivity. Plant physiol 109:7-13

Malosetti M, van Eeuwijk FA, Boer MP, Casas AM, Elia M, Moralejo M, Bhat PR, Ramsay L, Molina-Cano JL (2011) Gene and QTL detection in a three-way barley cross under selection by a mixed model with kinship information using SNPs. Theor Appl Genet 122:1605-1616

Malosetti M, Ribaut JM, Vargas M, Crossa J, Eeuwijk FA (2007) A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). Euphytica 161:241-257

Mano Y, Takeda K (1997) Mapping quantitative trait loci for salt tolerance at germination and the seedling stage in barley (*Hordeum vulgare* L.). Euphytica 94:263–272

Munns R (2010) Approaches to identifying genes for salinity tolerance and the importance of timescale. Methods Mol Biol 639:25-38

Munns R, James RA, Läuchli A (2006) Approaches to increasing the salt tolerance of wheat and other cereals. J Exp Bot 57:1025-1043

Munns R, Tester M (2008) Mechanisms of salinity tolerance. Annu Rev Plant Biol 59:651-681

Nguyen VL, Ribot SA, Dolstra O, Niks RE, Visser RGF, van der Linden CG (2012) Identification of quantitative trait loci for ion homeostasis and salt tolerance in barley (*Hordeum vulgare* L.). Mol Breeding 31:137-152

Pastina MM, Malosetti M, Gazaffi R, Mollinari M, Margarido GR, Oliveira KM, Pinto LR, Souza AP, van Eeuwijk FA, Garcia AA (2012) A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. Theor Appl Genet 124:835-849

Payne R, Murray D, Harding S, Baird D, Soutar D (2012) Introduction to GenStat for Windows, 15th edn. VSN International, Hemel Hempstead. www.vsni.co.uk/software/genstat?ref=genstat.com

Ren ZH, Gao JP, Li LG, Cai XL, Huang W, Chao DY, Zhu MZ, Wang ZY, Luan S, Lin HX (2005) A rice quantitative trait locus for salt tolerance encodes a sodium transporter. Nat genet 37:1141-1146

Schwember AR, Bradford KJ (2010a) A genetic locus and gene expression patterns associated with the priming effect on lettuce seed germination at elevated temperatures. Plant Mol Biol 73:105-118

Schwember AR, Bradford KJ (2010b) Quantitative trait loci associated with longevity of lettuce seeds under conventional and controlled deterioration storage conditions. J Exp Bot 61:4423-4436

Shahbaz M, Ashraf M (2013) Improving salinity tolerance in cereals. Crit Rev Plant Sci 32:237-249

Sharma S, Xu S, Ehdaie B, Hoops A, Close TJ, Lukaszewski AJ, Waines JG (2011) Dissection of QTL effects for root traits using a chromosome arm-specific mapping population in bread wheat. Theor Appl Genet 122:759-769

Thomson MJ, Ocampo M, Egdane J, Rahman MA, Sajise AG, Adorada DL, Tumimbang-Raiz E, Blumwald E, Seraj ZI, Singh RK, Gregorio GB, Ismail AM (2010) Characterizing the saltol quantitative trait locus for salinity tolerance in rice. Rice 3:148-160

Truco MJ, Antonise R, Lavelle D, Ochoa O, Kozik A, Witsenboer H, Fort SB, Jeuken MJ, Kesseli RV, Lindhout P, Michelmore RW, Peleman J (2007) A high-density, integrated genetic linkage map of lettuce (*Lactuca* spp.). Theor Appl Genet 115:735-746

Truco MJ, Ashrafi H, Kozik A, van Leeuwen H, Bowers J, Reyes Chin Wo S, Stoffel K, Xu H, Hill T, van Deynze A, Michelmore RW (2013) An ultra high-density, transcript-based, genetic map of lettuce. G3 (Bethesda)3:617-631

Tuyen DD, Lal SK, Xu DH (2010) Identification of a major QTL allele from wild soybean (*Glycine soja* Sieb. & Zucc.) for increasing alkaline salt tolerance in soybean. Theor Appl Genet 121:229-236

Uwimana B, Smulders MJ, Hooftman DA, Hartman Y, van Tienderen PH, Jansen J, McHale LK, Michelmore RW, van de Wiel CC, Visser RG (2012a) Hybridization between crops and wild relatives: the contribution of cultivated lettuce to the vigour of crop-wild hybrids under drought, salinity and nutrient deficiency conditions. Theor Appl Genet 125:1097-1111

Uwimana B, Smulders MJ, Hooftman DA, Hartman Y, van Tienderen PH, Jansen J, McHale LK, Michelmore RW, Visser RG, van de Wiel CC (2012b) Crop to wild introgression in lettuce: following the fate of crop genome segments in backcross populations. BMC Plant Biol 12:43

Vallejo AJ, Yanovsky MJ, Botto JF (2010) Germination variation in *Arabidopsis thaliana* accessions under moderate osmotic and salt stresses. Ann Bot-London 106:833-842

van Eeuwijk FA, Bink MC, Chenu K, Chapman SC (2010) Detection and use of QTL for complex traits in multiple environments. Curr Opin Plant Biol 13:193-205

van Treuren R, van der Arend AJM, Schut JW (2011) Distribution of downy mildew (*Bremia lactucae* Regel) resistances in a genebank collection of lettuce and its wild relatives. Plant Genetic Resources 11:15-25

Vaughn LM, Masson PH (2011) A QTL study for regions contributing to *Arabidopsis thaliana*root skewing on tilted surfaces. G3 (Bethesda) 1:105-115

Vision TJ, Brown DG, Shmoys DB, Durrett RT, Tanksley SD (2000) Selective mapping: astrategy for optimizing the construction of high-density linkage maps. Genetics 155:407-420

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77-78

Wang Z, Chen Z, Cheng J, Lai Y, Wang J, Bao Y, Huang J, Zhang H (2012a) QTL analysis of Na^+ and K^+ concentrations in roots and shoots under different levels of NaCl stress in rice (*Oryza sativa* L.). PLoS ONE 7:e51202

Wang Z, Cheng J, Chen Z, Huang J, Bao Y, Wang J, Zhang H (2012b) Identification of QTLs with main, epistatic and QTL x environment interaction effects for salt tolerance in rice seedlings under different salinity conditions. Theor Appl Genet 125:807-815

Wang Z, Wang J, Bao Y, Wu Y, Zhang H (2010) Quantitative trait loci controlling rice seed germination under salt stress. Euphytica 178:297-307

Zhang FZ, Wagstaff C, Rae AM, Sihota AK, Keevil CW, Rothwell SD, Clarkson GJ, Michelmore RW, Truco MJ, Dixon MS, Taylor G (2007) QTLs for shelf life in lettuce co-locate with those for leaf biophysical properties but not with those for leaf developmental traits. J Exp Bot 58:1433-1449

Zhang NW, Lindhout P, E. NR, Jeuken MJW (2009) Genetic dissection of *Lactuca saligna* nonhost resistance to downymildew at various lettuce developmental stages. Plant Pathol 58:923-932

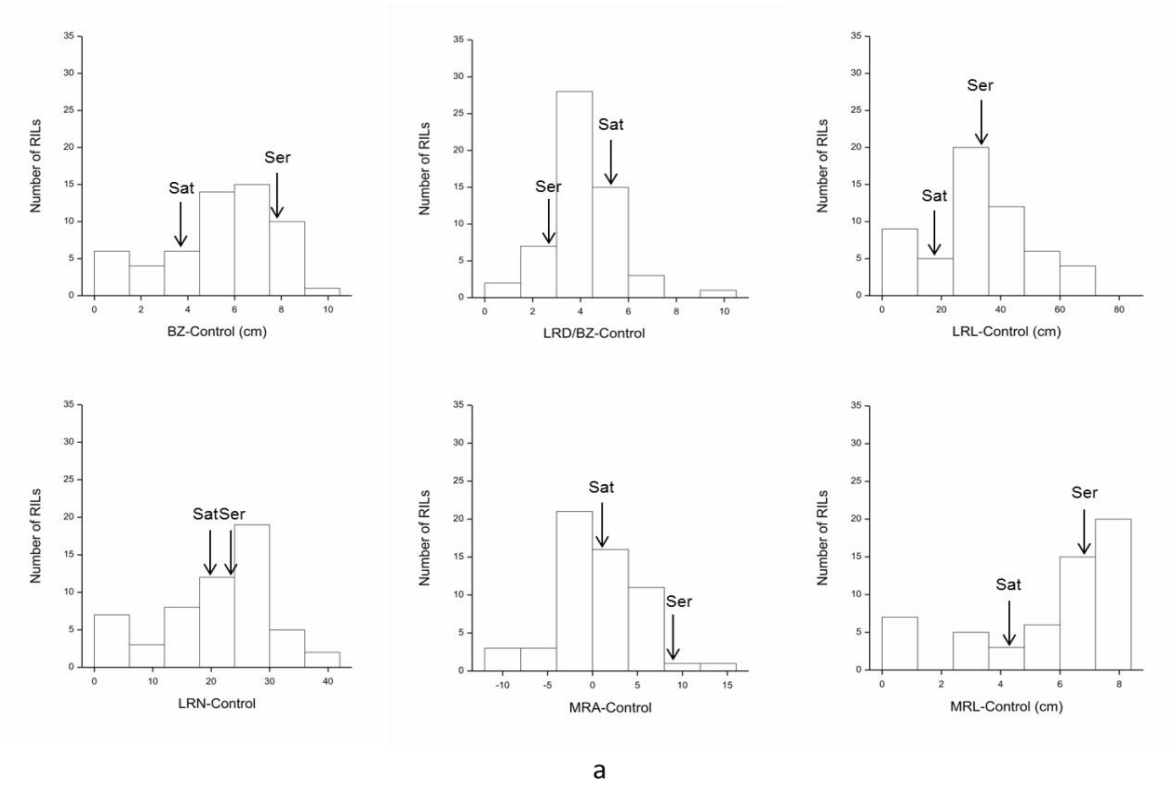
Zhao Y, Wang T, Zhang W, Li X (2011) SOS3 mediates lateral root development under low salt stress through regulation of auxin redistribution and maxima in *Arabidopsis*. New phytol 189:1122-1134

Zhou G, Johnson P, Ryan PR, Delhaize E, Zhou M (2011) Quantitative trait loci for salinity tolerance in barley (*Hordeum vulgare* L.). Mol Breeding 29:427-436

Zhu J, Kaeppler SM, Lynch JP (2005) Mapping of QTL controlling root hair length in maize (*Zea mays* L.) under phosphorus deficiency. Plant Soil 270:299-310

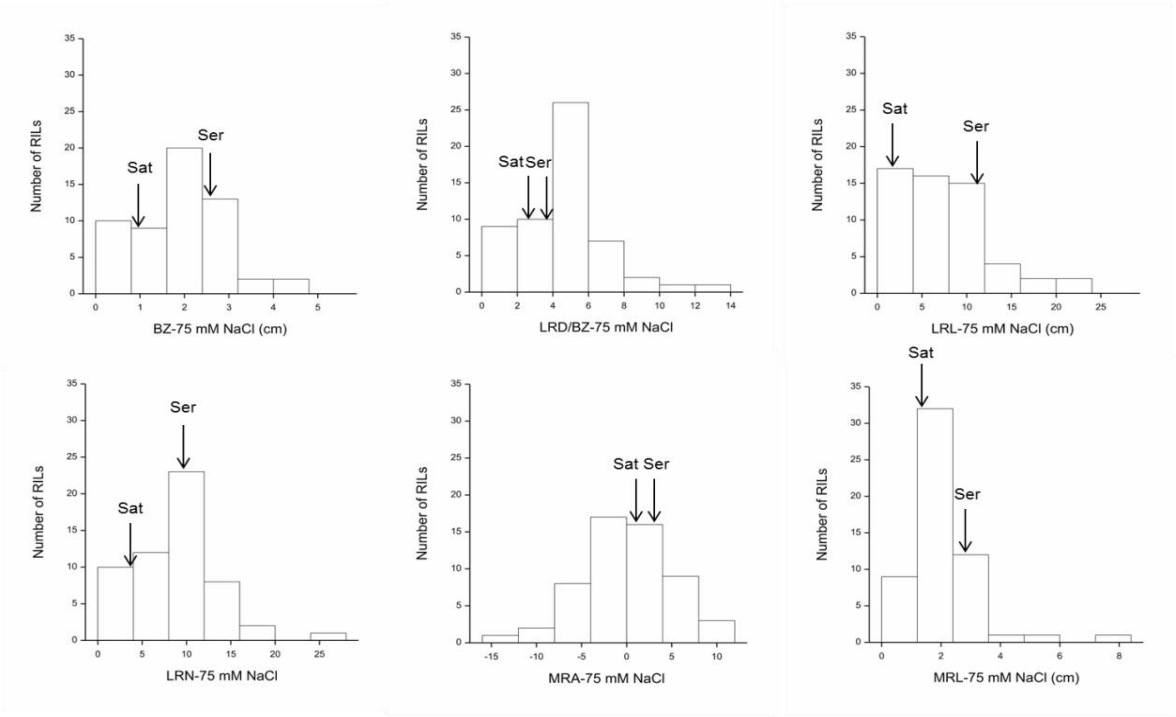
Zolla G, Heimer YM, Barak S (2010) Mild salinity stimulates a stress-induced morphogenic response in *Arabidopsis thaliana* roots. J Exp Bot 61:211-224

Supplementary figures and tables



a

Figure S1a phenotypic distribution of RSA traits in non-stress condition



b

Figure S1b phenotypic distribution of RSA traits in the 75 mM NaCl condition

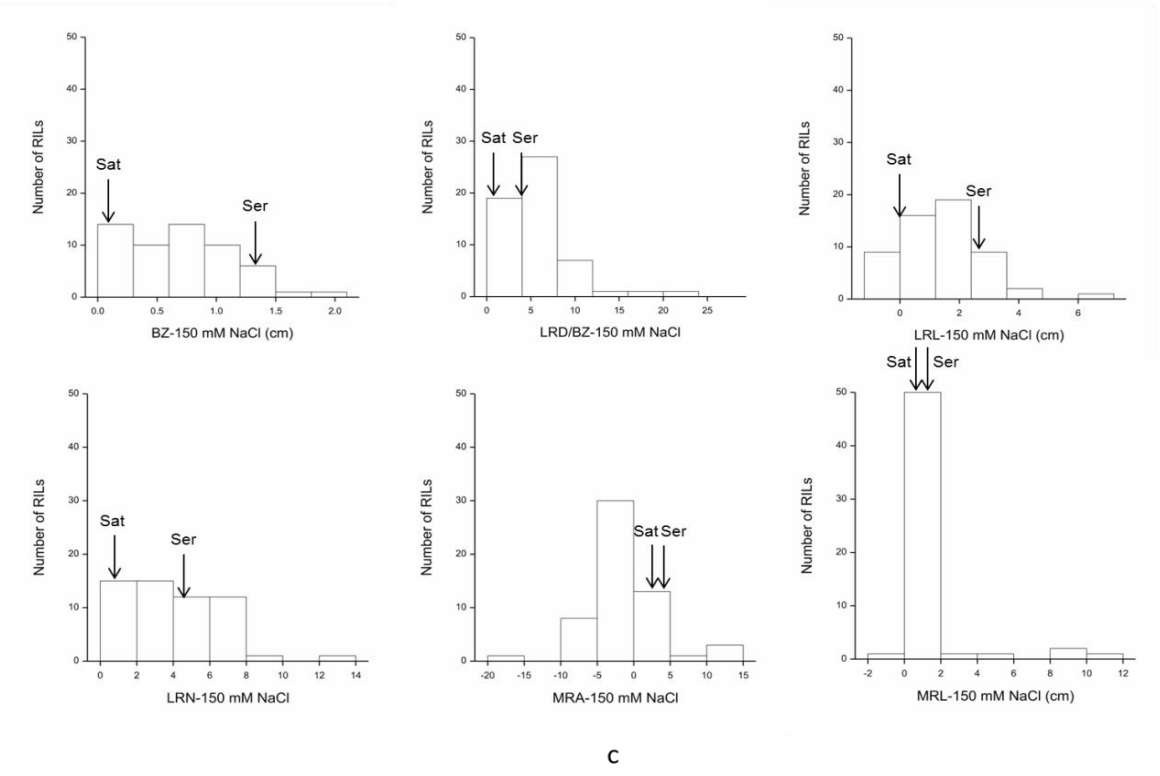


Figure S1c phenotypic distribution of RSA traits in the 150 mM NaCl condition

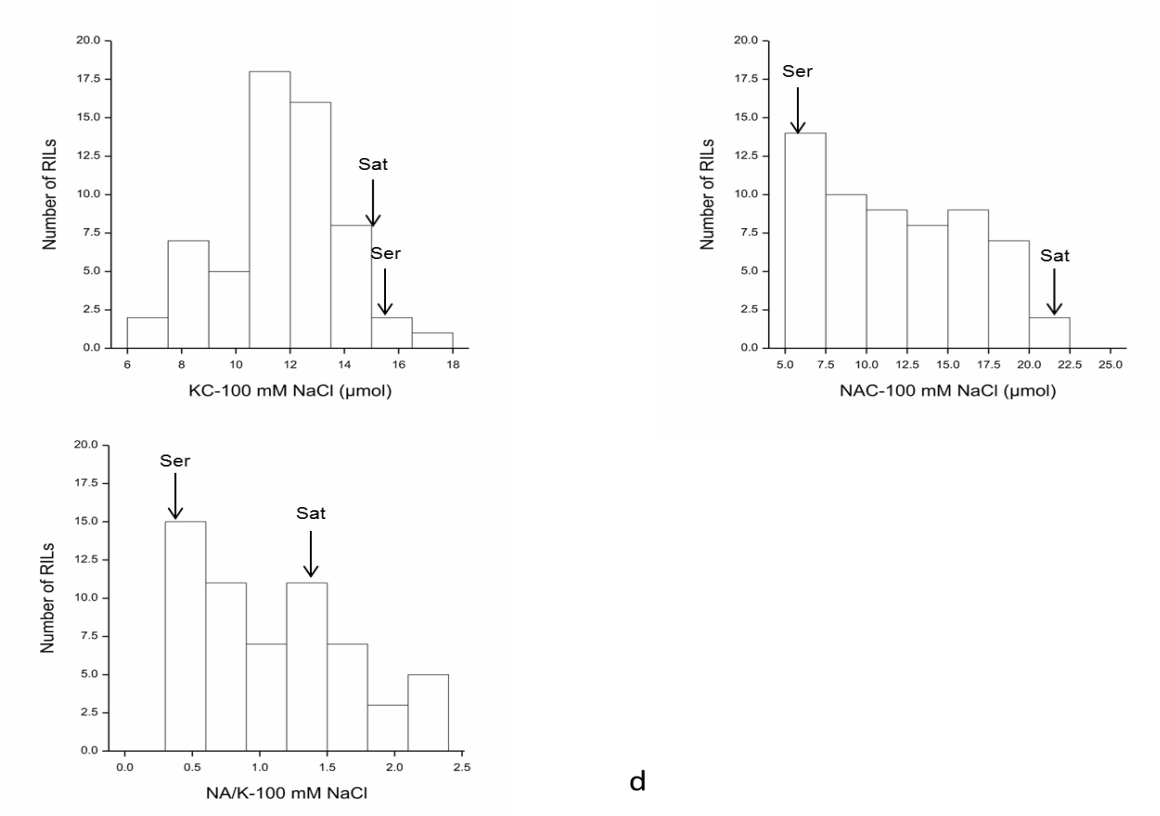


Figure S1d phenotypic distribution of ion accumulation traits in the 100 mM NaCl condition; the arrows indicate the mean values for parental lines - *L. serriola* and *L. sativa* 'Salinas'

Table S1 Correlation matrix (Pearson (n)) in the control condition:

Variables	MRL	LRD/BZ	LRN	MRA	BZ	LRL
MRL	1*	-0.003	0.525*	-0.056	0.648*	0.543*
LRD/BZ	-0.003	1*	0.416*	-0.075	0.048	0.226
LRN	0.525*	0.416*	1*	-0.009	0.846*	0.884*
MRA	-0.056	-0.075	-0.009	1*	0.047	0.044
BZ	0.648*	0.048	0.846*	0.047	1*	0.846*
LRL	0.543*	0.226	0.884*	0.044	0.846*	1*

* indicates values different from 0 with a significance level alpha=0.05

Table S2 Correlation matrix (Pearson (n)) in the 75 mM NaCl condition:

Variables	MRL	LRD/BZ	LRN	MRA	BZ	LRL
MRL	1*	0.102	0.503*	-0.022	0.676*	0.547*
LRD/BZ	0.102	1*	0.496*	-0.037	0.305*	0.417*
LRN	0.503*	0.496*	1*	0.077	0.870*	0.888*
MRA	-0.022	-0.037	0.077	1*	0.035	0.190
BZ	0.676*	0.305*	0.870*	0.035	1*	0.792*
LRL	0.547*	0.417*	0.888*	0.190	0.792*	1*

* indicates values different from 0 with a significance level alpha=0.05

Table S3 Correlation matrix (Pearson (n)) in the 100 mM NaCl condition:

Variables	NAC	KC	NA/K
NAC	1*	-0.602*	0.924*
KC	-0.602*	1*	-0.822*
NA/K	0.924*	-0.822*	1*

* indicates values different from 0 with a significance level alpha=0.05

Table S4 Information of major QTLs found in multi-trait linkage analysis

QTL	Environment (mM NaCl)	Trait	High value allele	AE	SE	P	PVE (%)
QRS2.1	75	BZ	WILD	0.46	0.12	-	18.0
QRS2.1	75	LRD/BZ	WILD	0.29	0.28	0.31	1.4
QRS2.1	75	LRL	WILD	2.71	0.47	-	26.3[a]
QRS2.1	75	LRN	WILD	2.01	0.47	-	15.4
QRS2.1	75	MRA	CROP	-0.11	0.73	0.89	0
QRS2.1	75	MRL	WILD	0.38	0.13	-	9.9
QRC9.1	0	BZ	CROP	-0.70	0.28	0.01	7.8
QRC9.1	0	LRD/BZ	CROP	-0.41	0.19	0.03	7.8
QRC9.1	0	LRL	CROP	-8.59	1.76	-	23.0
QRC9.1	0	LRN	CROP	-4.99	1.03	-	25.7[a]
QRC9.1	0	MRA	WILD	0.87	0.59	0.14	3.8
QRC9.1	0	MRL	CROP	-0.57	0.29	0.05	5.9
QLS7.2	100	KC	WILD	1.58	0.22	-	49.2[a]
QLS7.2	100	NA/K	CROP	-0.44	0.04	-	57.9[a]
QLS7.2	100	NAC	CROP	-3.39	0.27	-	57.9[a]

[a] Major QTL; significance level alpha=0.05
- P value is smaller than 0.0005
L. sativa 'Salinas' is the crop parent, *L. serriola* is the wild parent
AE additive effect, *SE* standard error, *PVE* the percentage of explained variances

Chapter 5

Allele-specific expression of *Lactuca HKT1;1* genes correlates with divergence of promoter regions and whole-plant Na^+/K^+ homeostasis

Zhen Wei[1], R. W. Michelmore[2], David E. Salt[3], M. Eric Schranz[1]

1 Biosystematics Group, Wageningen University and Research Centre, 6708 PB Wageningen, The Netherlands

2 Genome Center and Department of Plant Sciences, University of California-Davis, CA 95616, U.S.A.

3 Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen AB24, 3UU, United Kingdom

Manuscript in preparation

Chapter 5

Abstract

A previous QTL analysis of salt stress in lettuce, using seedlings from a recombinant inbred (RIL) line population derived from the cultivated (*Lactuca sativa* 'Salinas') and wild (*L. serriola*) lettuce, found one major QTL region (qLS7.2) contributing to Na^+/K^+ homeostasis in leaves. Here we report the identification and characterization of allelic variation of a *Lactuca* high-affinity K^+ transporter 1;1 (*HKT1;1*) homolog located at the maximum LOD value of qLS7.2. We constructed a phylogenetic analysis of *Lactuca HKT1*-like protein sequences with other published *HKT* protein sequences and identified transmembrane and pore segments of lettuce *HKT1;1 alleles*, according to the four-MPM structural model proposed for *AtHKT1;1*. The 5' upstream promoter regions (approximately 2kb) of both genotypes were investigated for cis-acting regulatory elements as well. The concentration of Na^+ and K^+ and the relative gene expression of the two lettuce *HKT1;1* alleles were quantified over a time-course for both shoots and roots using plants grown hydroponically. We found 37 and 21 negative cis-regulatory elements, specific to *AtHKT1;1* expression in roots, for *LseHKT1;1* (*L. serriola*) and *LsaHKT1;1* (*L. sativa* 'Salinas'), respectively. This result was consistent with the low expression of the lettuce *HKT1* alleles in roots and high expression in shoots, showing a time-dependent pattern. Significant allelic differences were identified in *Lactuca HKT1;1* expression in early stage (0-24 hours) shoots with higher expression of *LsaHKT1;1* and in late stage (2-6 days) roots with higher expression of *LseHKT1;1*. *L. sativa* 'Salinas' has higher *HKT1;1* expression and was more tolerant than *L. serriola* within 24 hours, although afterwards no significant differences of Na^+/K^+ ratios or *HKT1;1* expression were observed.

Key words

Gene expression; *HKT1*; *Lactuca*; lettuce; salt stress

Introduction

Soil salinization, exacerbated by a mismatch between water demands for irrigation in food production and the amount of non-saline water, has a negative effect on crop production (Gabrijel et al. 2011). In this context, engineering salt tolerance, including marker-assisted selection and gene stacking technologies, is crucial to enhance crop production (Deinlein et al. 2014). Mechanisms of salinity tolerance in plants include three main types: osmotic tolerance, Na^+ or Cl^- exclusion, and tissue tolerance to accumulated Na^+ or Cl^- (mainly into vacuoles) (Munns and Tester 2008; Roy et al. 2014).

The osmotic stress immediately happens after plants are exposed to salt stress and plants reduce cell expansion in root tips and young leaves, leading to stomatal closure (Munns and Tester 2008). ROS waves (Jiang et al. 2012; Mittler et al. 2011; Suzuki et al. 2012), Ca^{2+} waves (Roy et al. 2014) or long distance electrical signals (Maischak et al. 2010) may be involved in this 'osmotic phase', but there is still many unknowns. Comparatively, the accumulation of Na^+ in plants is better understood. The Salt Overly Sensitive (SOS) pathway (Huertas et al. 2012; Jarvis et al. 2014; Katschnig et al. 2015; Qiu et al. 2002) and the high affinity potassium transporter (*HKT*) gene family (Ali et al. 2012; Davenport et al. 2007; Hauser and Horie 2010; Horie et al. 2009; Platten et al. 2013; Rus et al. 2006; Rus et al. 2004) have been considered to play critical roles in regulating Na^+ transport within plants. The expression levels of these genes have been frequently reported to alter accumulation of Na^+ in shoots. Vacuolar Na^+/H^+ antiporters (NHX) (Barragan et al. 2012; Barragán et al. 2012; Rodríguez-Rosales et al. 2009), vacuolar H^+ pyrophosphatases (Pasapula et al. 2011), proteins involved in the synthesis of compatible solutes (e.g. proline) (Vendruscolo et al. 2007) and enzymes responsible for the detoxification of reactive oxygen species (Begara-Morales et al. 2014) have been implied to be successful to different extent in improving plant tissue tolerance.

Of the two gene families involved in plant sodium accumulation, the *HKT1* group of the *HKT* family has frequently been listed as the best target to improve salinity tolerance in crops. *HKT1* has often been reported as the most likely candidate for quantitative trait loci (QTL) associated with salt tolerance and/or Na^+ exclusion in mutant and mapping populations (Ahmadi et al. 2011; Ren et al. 2005; Rus et al. 2006). Novel *HKT1* alleles from the diploid wheat relative, *Triticum monococcum*, were successfully incorporated into a modern durum wheat cultivar and improved the salinity tolerance by marker-assisted selection (MAS) (James et al. 2012; James et al. 2006; Munns et al. 2012). In addition, *HKT1* genes appear to increase salinity tolerance by tissue-specific (Zhang et al. 2008) and/or cell type-specific (Moller et al. 2009) Na^+ transport in *Arabidopsis*.

A previous QTL analysis of salt tolerance in lettuce discovered one major QTL region (qLS7.2) related to sodium accumulation in leaves, using seedlings from a recombinant inbred line population derived from the lettuce crop (*Lactuca sativa* 'Salinas') and the wild

species (*L. serriola*) (Wei et al. 2014). The wild allele contributed to the additive effect of K^+ concentration, while the crop allele increased the additive effect of Na^+ and Na^+/K^+. Using mapping and genome data from the Compositae Genome Project Database (CGPDB 2014) (http://compgenomics.ucdavis.edu/), the QTL qLS7.2 was located onto lettuce chromosome 7. In this work, we report and characterize an *HKT1*-like protein coding sequence (*LsaHKT1;1*) at this location. Interestingly, there is another *HKT1* copy (*LsaHKT1;2*) on chromosome 4 of *L. sativa* 'Salinas'. In this study, expression levels of *HKT1*-like genes from *L. sativa* 'Salinas' (*LsaHKT1;1*) and *L. serriola* (*LseHKT1;1*), grown in hydroponic system, were quantified at different time points in control and salt conditions. Ion accumulation in different tissues of both genotypes was measured to examine the process of induction by salinity. Cis-regulatory elements in promoter regions of both genotypes were also identified.

Materials and methods

Phylogenetic analysis of lettuce *HKT1*-like protein sequences and promoter region analysis

The scaffolds containing *HKT1;1*-like protein sequences of *L. sativa* 'Salinas' and *L. serriola* (UC96US23), and the protein sequence of *LsaHKT1;2* were obtained from the Michelmore Lab, U.C. Davis, U.S.A. (Personal communication). Scaffold 894 of *L. sativa* 'Salinas' contains the *LsaHKT1;1*-like gene gi350612817, and the scaffold 8733 of *L. serriola* contains the orthologous gene *LseHKT1;1*-like gi350612817. The protein sequences from *LsaHKT1;1* and *LseHKT1;1* (on chromosome 7) and *LsaHKT1;2* (on chromosome 4) were first aligned with 42 other known angiosperm *HKT*s (*HKT*-like) protein sequences (**Table S1**) using MAFFT (Katoh et al. 2002). Unrooted Minimum-Evolution tree was constructed using MEGA 6.06 with default settings (Tamura et al. 2013). The amino acid sequences of *HKT1*-like genes (*LsaHKT1;1* and *LseHKT1;1*) from the two lettuce species were also compared with *Arabidopsis thaliana HKT1* protein sequence (*AtHKT1;1*, accession number: NP_567354) to investigate structural differences. The positions of transmembrane and pore segments were predicted according to the four-MPM structural model (transmembrane segment, pore, transmembrane segment) proposed for *AtHKT1;1* (Durell and Guy 1999; Hamamoto et al. 2015; Kato et al. 2001). In addition, the genomic DNA sequences of *LsaHKT1;1* and *LseHKT1;1* were compared using Blast2Sequences (Altschul et al. 1990).

The upstream 5' UTR and promoter regions (approximately 2 kb) of *LsaHKT1;1* and *LseHKT1;1* alleles were investigated for cis-regulatory elements using an online database of Plant Cis-acting Regulatory DNA Elements (PLACE) (Higo et al. 1999). The presence of CpG islands was also checked using the CpG Island Searcher web tool with default settings (Takai and Jones 2002).

Experimental conditions and plant materials

Seeds of the two parental genotypes of *L. sativa* 'Salinas' and *L. serriola*, used in a previous QTL study related to salt stress (Wei et al. 2014), were put in a cold (4◦ C) dark room for one week to break seed dormancy. Seeds were then germinated for six days and grown for one week in sandy soil in the greenhouse. Plant seedlings were then transferred to hydroponic containers (20 litres) in the greenhouse filled with liquid solution (Dry Hydroponics®, cultivation system for short cycle crops) (**Figure 1**). Two-week-old seedlings were then transferred to new containers with different treatments: control (0 mM NaCl) or 75 mM NaCl. Electrical conductivity was measured before treatment using an Elmeco meter (Tasseron Sensors & Controls, Nootdorp, The Netherlands) to make sure that the environment in containers with the same treatment were homogenous. Each container included seedlings of both genotypes (5 for each genotype) and represented one biological replicate. Shoot and root materials were sampled for RNA isolation and ion concentration measurement at 0h (transfer time point), 2h (hours), 6h, 12h and 24h, 2d (days), 4d and 6d for both control and 75 mM NaCl conditions using three biological replicates per time point and treatment.

Figure 1 Example of lettuce seedlings being grown in hydroponic system in greenhouse.

Total RNA isolation and cDNA synthesis

Total RNA of *Lactuca* shoots and roots was isolated using RNeasy® Plant Mini Kit (QIAGEN, Netherlands) including RNase-Free DNase Set (QIAGEN, Netherlands). The quality and concentration of RNA isolations treated with DNase were evaluated using A_{260}/A_{280} and A_{260}/A_{230} ratios in a NanoDrop 2000 spectrophotometer (Thermo SCIENTIFIC, Netherlands). Total RNA (1µg) (DNA-free) was reverse transcribed using iScript™ cDNA Synthesis Kit (Bio-Rad Laboratories, Inc., Netherlands) in T100™ Thermal Cycler (Bio-Rad Laboratories, Inc., Netherlands), following the manufacturer's instructions (Bio-Rad Laboratories, Inc., Netherlands). Thermal cycling for cDNA synthesis was performed in 20 µl reaction, including 5 min. at 25 ℃, 30 min. at 42 ℃, ended by 5 minutes at 85 ℃.

Primer design and RT-qPCR

Specific primers for *LsaHKT1;1* and *LseHKT1;1* were designed using Primer3Plus (using default settings) (Untergasser et al. 2007). Candidate reference control genes were obtained

from previously published literature (Borowski et al. 2014; Porcel et al. 2006; Zhang et al. 2009). The information of designed and selected primers is shown in **Table 1**.

Table 1 Primer information of the reference and lettuce *HKT1;1* genes

No.	Gene	Forward primer sequence	Reverse primer sequence
1	*TUB*	5'-TAGGCGTGTGAGTGAGCAGT-3'	5'-AACCCTCGTACTCTGCCTCTT-3'
2	*40S*	5'-CAAGATTCGGTGACAGGGATG-3'	5'-CACCACCTCCAAATCCACCA-3'
3	*EIF2A*	5'-TAGGCGAGTGGAGAAGCATT-3'	5'-GTAGAAACAGCAACAGGCAAA-3'
4	*HKT1;1*	5'-ATGGAAATGTGGGGTTCTCA-3'	5'-CTTCCAGAAAACCCGTACCA-3'

Real-time PCR was performed with iQ™ SYBR® Green (Bio-Rad Laboratories, Inc., Netherlands) using CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad Laboratories, Inc.). Thermal cycling followed Borowski et al. 2014 (Borowski et al. 2014). A dilution series (10, 100, 1000 and 10000 times) of the cDNA samples in Milli-Q water were tested to identify the cDNA concentrations that produce cycle threshold values between 18 and 30. The final reaction volume was 10 µl, including 0.5 µl forward primer, 0.5 µl reverse primer, 5 µl SYBR® Green Mix and 4 µl cDNA. RT-qPCR reactions of three biological replicates (each included two technical replicates) were performed in one plate. Negative controls were included for each primer pair to avoid contaminants. The relative expression levels were calculated according to the $2^{-\Delta Ct}$ method (Julkowska 2015; Livak and Schmittgen 2001).

Measurement of Na^+ and K^+ in lettuce leaves and roots by an Inductively Coupled Plasma-mass Spectrometry (ICP-MS) and data analysis

Fresh leaf and root materials, from the same *Lactuca* plants used for RNA isolations, were dried in an oven at 88℃ overnight. The concentration of Na^+ and K^+ in the dried plant material was measured at the Ionomics Facility of School of Biological Sciences, University of Aberdeen (http://www.ionomicshub.org/home/PiiMS). The steps can be described as follows: (1) dry lettuce materials were transferred into Pyrex test tubes (16 x 100 mm); (2) after weighing the appropriate number of samples (these masses were used to calculate the rest of the sample masses (Danku et al. 2013), trace metal grade nitric acid (J. T. Baker® BAKER Instra-Analyzed™; Avantor Performance Materials; Scientific & Chemical Supplies Ltd, Aberdeen, UK) spiked with indium internal standard was added to the tubes (1.20 mL); (3) hydrogen peroxide (1.50 mL) (Primar-Trace analysis grade, 30%; Fisher Scientific, Loughborough, UK) was also added; (4) samples were left overnight to pre-digest and then digested in dry block heaters (DigiPREP MS, SCP Science; QMX Laboratories, Essex, UK) at 115˚C for 4 hours; (5) the digested samples were diluted to 11.5 mL with 18.2 MΩcm Milli-Q Direct water (Merck Millipore, Watford, UK) and aliquots transferred to 96-well deep well plates using adjustable multichannel pipette (Rainin; Anachem Ltd, Luton, UK) for analysis; (6) elemental analysis was performed with an inductively coupled plasma-mass

spectrometry (ICP-MS) (PerkinElmer NexION 300D equipped with Elemental Scientific Inc. autosampler and Apex HF sample introduction system; PerkinElmer LAS Ltd, Seer Green, UK and Elemental Scientific Inc., Omaha, NE, USA, respectively) in the standard mode; (7) twenty elements (Li, B, Na, Mg, P, S, K, Ca, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Rb, Sr, Mo, and Cd) were monitored; (8) liquid reference material composed of pooled samples of the digested materials was prepared before the beginning sample runs and was used throughout the whole ICP-MS runs; the reference material was run after every ninth sample in all ICP-MS sample sets to correct for variation between and within ICP-MS analysis runs (Danku et al. 2013).

Sample concentrations were calculated using an external calibration method within the instrument software. The calibration standards (with indium internal standard and blanks) were prepared from single element standards (Inorganic Ventures; Essex Scientific Laboratory Supplies Ltd, Essex, UK) solutions. Further data computations were made using Microsoft Excel software.

Results

Phylogenetic analysis of *Lactuca HKT1*-like protein sequences and promoter region analysis of *LsaHKT1;1* and *LseHKT1;1*

The phylogenetic analysis confirmed that the *HKT1*-like protein sequences in lettuce, *LsaHKT1;1* and *LseHKT1;1* on chromosome 7 and *LsaHKT1;2* on chromosome 4, all belonged to *HKT* Class I transporters (**Figure 2**). Further analysis was only done for *LsaHKT1;1* and *LseHKT1;1* as we focus on the *HKT1*s in the QTL (qLS7.2) region related to salinity stress. The comparison of *HKT1;1* protein sequences showed that there were only two amino acid polymorphisms between *LsaHKT1;1* (*L. sativa* 'Salinas') and *LseHKT1;1* (*L. serriola*) protein alleles and they were 48.6% identical to *AtHKT1;1*. The alignment of the amino acid sequences and the predicted transmembrane and pore segments are shown in **Figure 3**. The comparison of the genomic DNA sequences is shown in **Table 2**. The *Lactuca HKT1;1* genes contain three exons and two introns. The nucleotide DNA sequences of exons and introns of the two lettuce *HKT1;1* alleles were very similar (sequence identity between 99% and 100%). However, the 1st intron of *LsaHKT1;1* contained a large number of ambiguous based (Ns).

Table 2 Genomic sequence comparison of the lettuce *HKT1* genes

Name	1st exon	1st intron	2nd exon	2nd intron	3rd exon
LsaHKT1;1	1161	10615	231	135	324
LseHKT1;1	1161	8867	231	139	324
Coverage*	100%	84%	100%	100%	100%
Identity	99%	99%	100%	100%	99%

* compared to *LsaHKT1;1* sequence

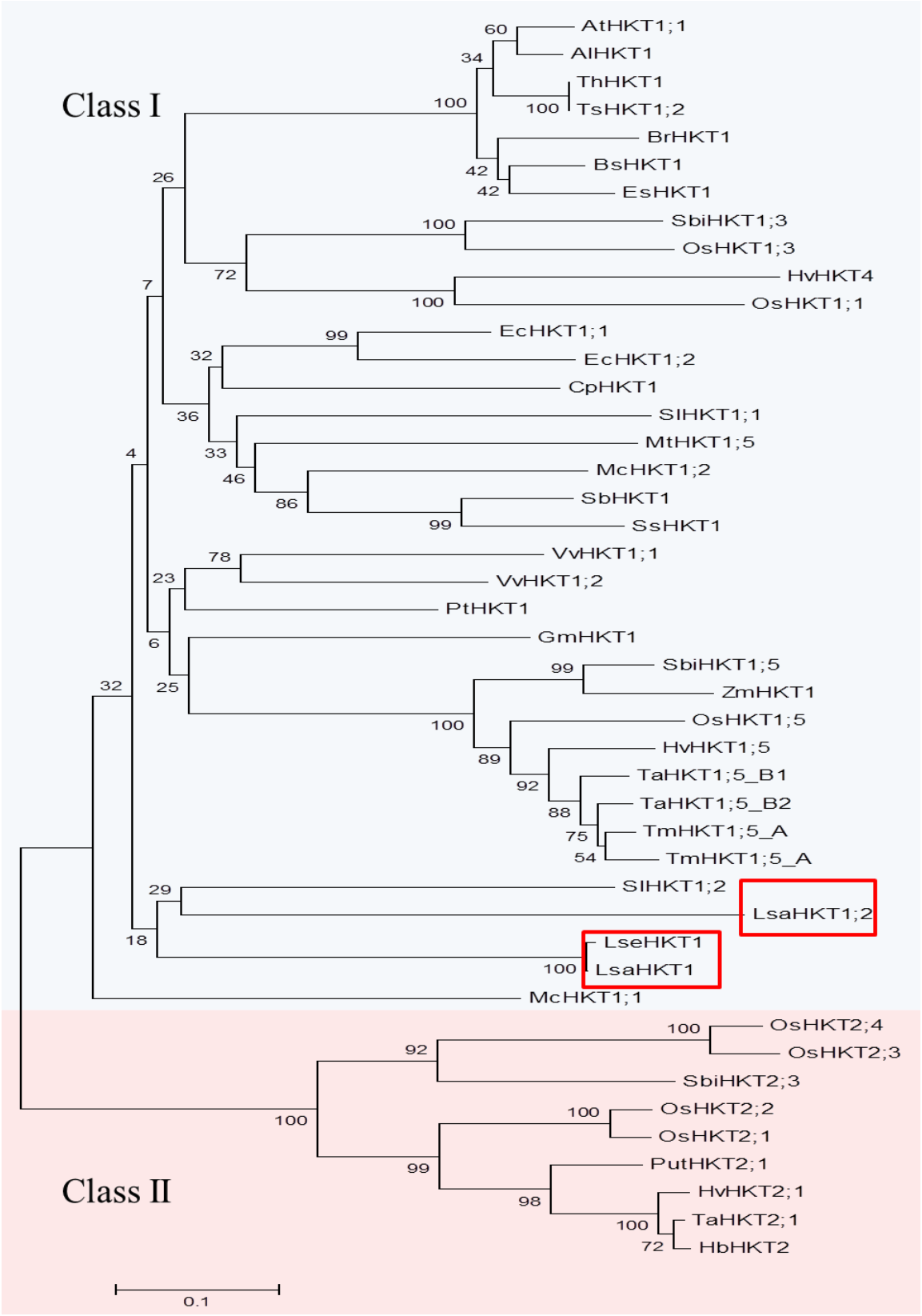


Figure 2 Phylogenetic relationships between lettuce *HKT1*-like proteins and other plant *HKT* transporters. Unrooted minimum-evolution tree was constructed with full polypeptide sequences with MEGA 6.06, using default settings. The scale bar represents a distance of 0.05 substitutions per site. Bootstrap values are shown above the branches.

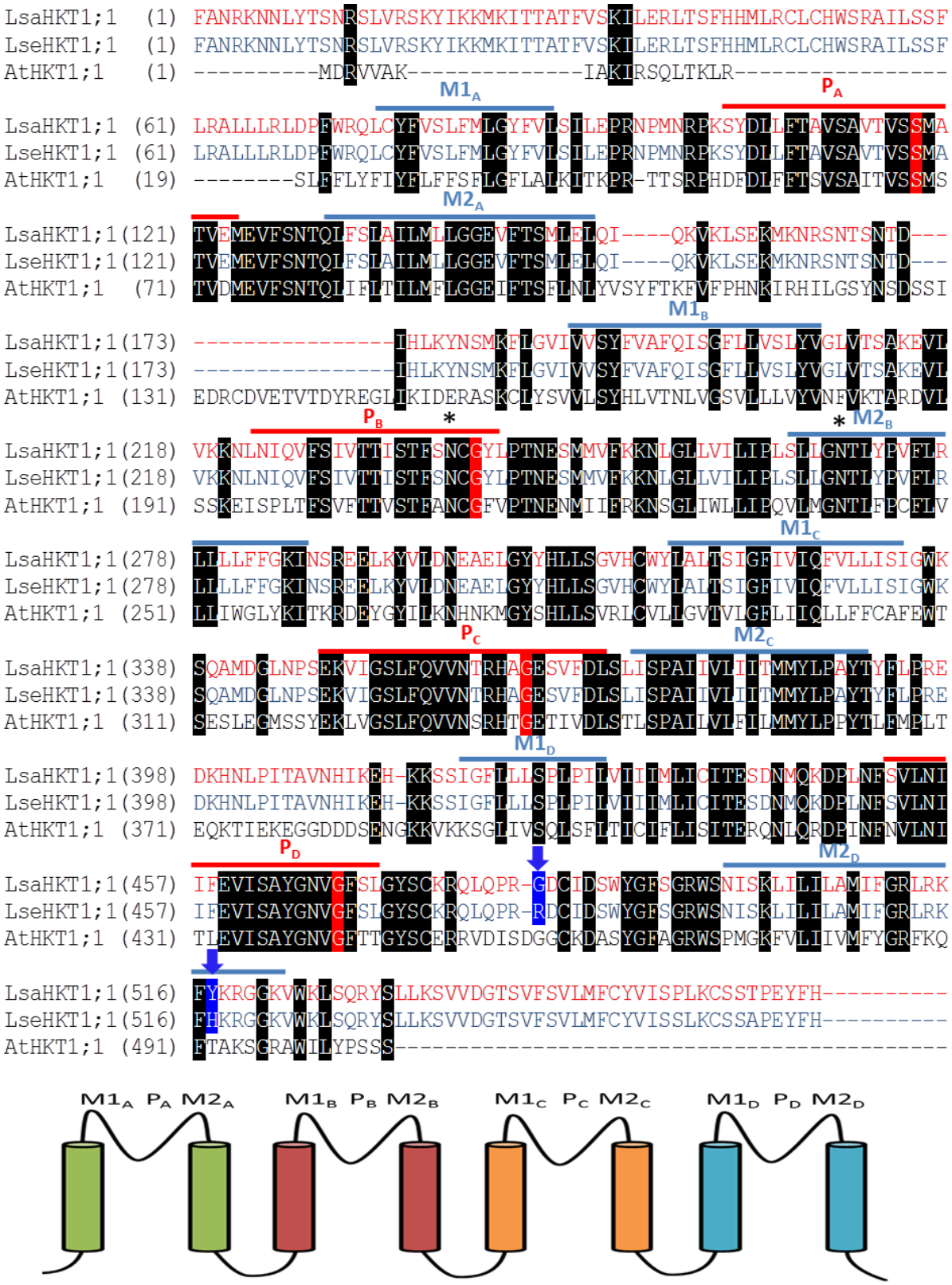


Figure 3 Alignment of *LseHKT1;1*, *LsaHKT1;1* and *AtHKT1;1* amino acid sequences. Identical residues in all sequences are highlighted in black. Residue substitutions of the two lettuce alleles are indicated in blue with arrows. Positions of transmembrane and pore segments were predicted according to the four-MPM structural model proposed for the topology of the *AtHKT1;1* protein. The conserved Gly residues in the K^+ channel selectivity filter GYG of the P-loop-like domains are highlighted in red (Mäser et al. 2002b). The

presence of Ser in the P_A-loop is conserved in Na^+ permeable *HKT* transporters (Mäser et al. 2002b; Platten et al. 2006). * defines the position of Asp residues reported to be essential for K^+ transport activity in *TsHKT1;2* (Ali et al. 2012).

The investigation of cis-acting regulatory elements in promoter regions of *LsaHKT1;1* and *LseHKT1;1* were made by analyzing the 2076 bp and 2013 bp 5' upstream sequences, respectively. No CpG islands were detected for either *HKT1;1* promoters, which would be important for potential epigenetic regulation by methylation. A total of 386 and 466 putative cis-acting elements were found in the promoter regions of *LsaHKT1;1* and *LseHKT1;1* (for two strands), respectively (**Table S2 and S3**). Most cis-acting elements found in the promoter regions of the two genes were shared elements, such as CAAT and GATA boxes (enhancer regions). The unique elements for both genes are shown in **Table 3**. Transcription factors binding sites associated with water stress (MYCATERD1), dehydration (MYCATRD22) and elevated external salinity (MYB binding site, basic leucine zipper (bZIP) (Deinlein et al. 2014; Hamamoto et al. 2015; Roy et al. 2014) were found in *LseHKT1;1*. In contrast, cis-acting elements related to drought, high-light, low temperature, and cold stresses, as well as transcription factor family genes, APETALA2/ETHYLENE RESPONSE FACTOR (AP2/ERF), MYB and bZIP (different binding sites from *LseHKT1;1*) (Deinlein et al. 2014; Hamamoto et al. 2015; Roy et al. 2014) were detected in the promoter region of *LsaHKT1;1* gene.

Time course of gene expression of *HKT1;1* alleles and Na^+ and K^+ accumulation in *Lactuca* tissues

The relative gene expressions of *LsaHKT1;1* and *LseHKT1;1* were calculated against three reference genes: *EIF2A* (Elongation initiation factor gamma subunit), *TUB* (Tubulin) and *40S* (40S ribosomal RNA) (Borowski et al. 2014) (**Table 1**). The geometric means of the expression levels calculated based on the three reference genes is depicted in **Figure 4**. The two *HKT1;1* alleles in roots both showed very low expression levels through the whole experiment. Their expression began to increase after 24 hours of salt treatment (75 mM NaCl), ranging from 0.005 to 0.06 (*LseHKT1;1*) and from 0.005 to 0.03 (*LsaHKT1;1*) respectively (**Figure 4A**). The relative expression of *LseHKT1;1* in shoots decreased (from 0.3 to 0.09) after to salinity-treatment during the first 12 hours and rose from 24 hours afterwards (from 0.09 to 0.54) (**Figure 4B**). The gene expression of *LsaHKT1;1* in lettuce shoots remained more stable (around 0.5) than that of *LseHKT1;1* and demonstrated a sharp increase (1.85) at 24h (**Figure 4B**).

The accumulation of Na^+ in roots and shoots of both *Lactuca* species showed a general increase over time in the salinity treatment (**Figure S1A and B**). *L. serriola* demonstrated a higher Na^+ concentration in roots than that in *L. sativa* 'Salinas' during the whole time course, ranging from 0.5 to 7.6 g/Kg DW and from 0.2 to 3.7 g/Kg DW respectively (**Figure S1A**). In *Lactuca* shoots, the two species showed similar sodium accumulation for the first 24 hours after salt treatment, but from 24 hours afterwards, the cultivated lettuce displayed a higher

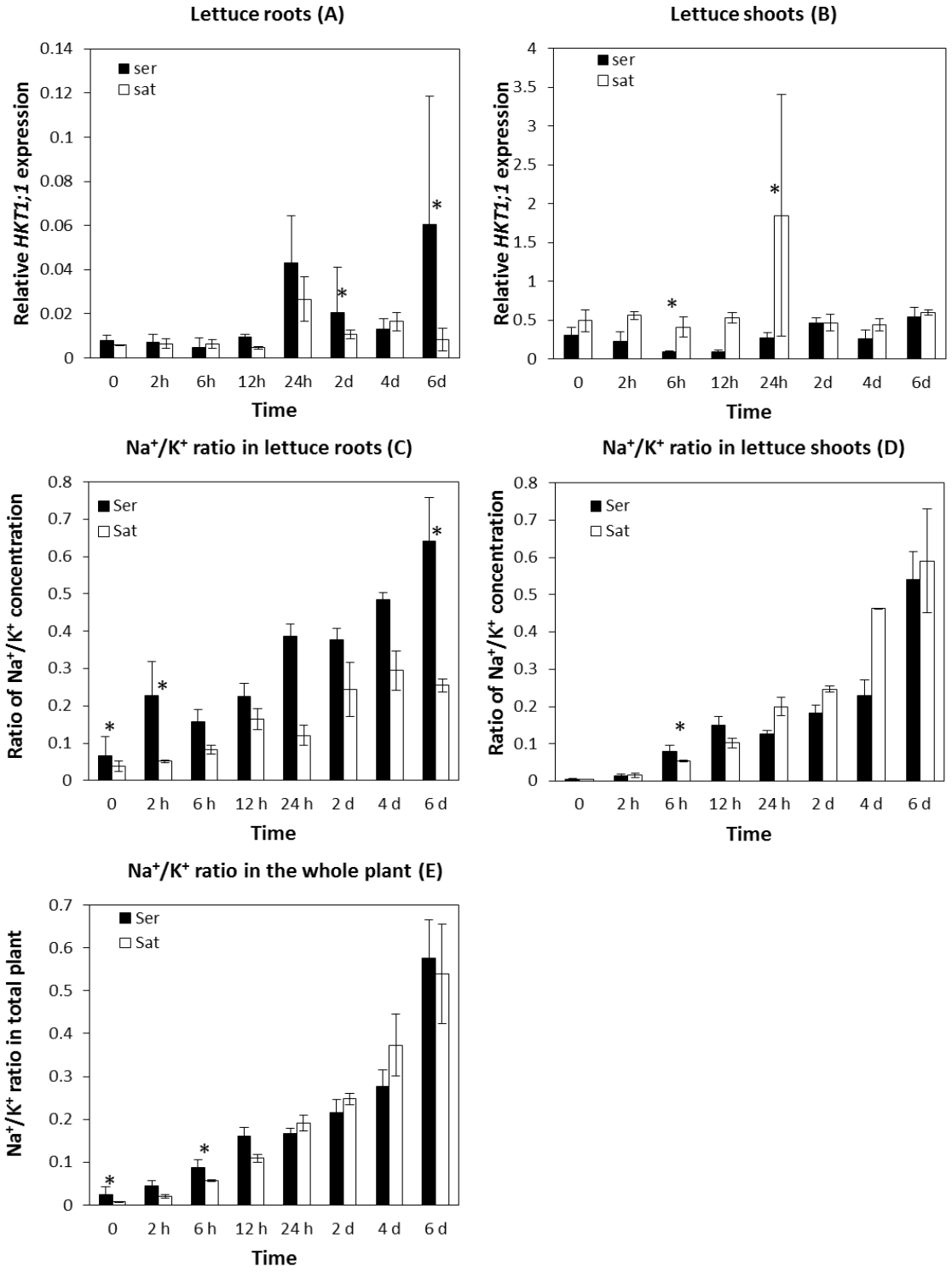
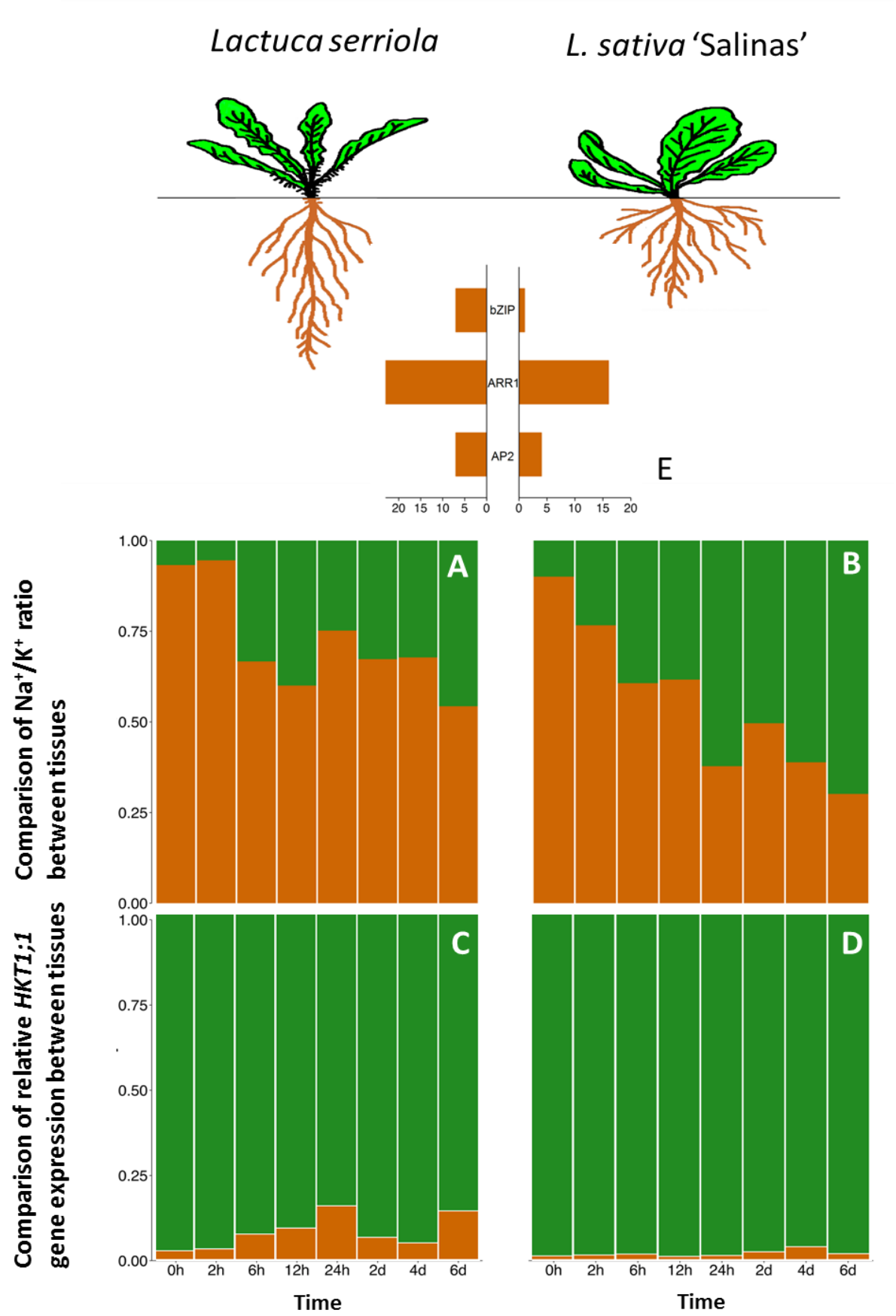


Figure 4 The relative expression of *Lactuca HKT1;1* alleles and Na^+/K^+ ratio in two lettuce species from time 0 (control condition) to 6 days (75 mM NaCl). Ser is the abbreviation of *L. serriola* and Sat is *L. sativa* 'Salinas'. (A) Relative expression of *HKT1;1* alleles in lettuce roots. (B) Relative expression of *HKT1;1* alleles in lettuce shoots. (C) Na^+/K^+ ratio in lettuce roots. (D) Na^+/K^+ ratio in lettuce shoots. (E) Na^+/K^+ ratio in the whole plant. Error bars indicate the Standard Error from three biological repeats. Student's t-test was used to test the differences between the two genotypes using IBM SPSS Statistics 22.0.


Lactuca serriola
L. sativa 'Salinas'
bZIP
ARR1
AP2
20 15 10 5 0
0 5 10 15 20
E
A
B
C
D
1.00
0.75
0.50
0.25
0.00
0h 2h 6h 12h 24h 2d 4d 6d
Time
Comparison of Na+/K+ ratio between tissues
Comparison of relative HKT1;1 gene expression between tissues

◄ **Figure 5** Summary of the expression of *HKT1* genes and Na^+/K^+ accumulation in the two lettuce genotypes through time. Left is wild lettuce, *L. serriola*. Right is cultivated lettuce, *L. sativa* 'Salinas'. The sum of Na^+/K^+ ratio and relative *HKT1;1* gene expression in roots and shoots is represented as 1. The percentages of Na^+/K^+ ratio and relative *HKT1;1* gene expression in roots (brown bar) and shoots (green bar) are shown in the figure. (A) The percentages of Na^+/K^+ ratio in roots and shoots of *L. serriola*. (B) The percentages of Na^+/K^+ ratio in roots and shoots of *L. sativa* 'Salinas'. (C) Relative *HKT1;1* gene expression in roots and shoots of *L. serriola*. (D) Relative *HKT1;1* gene expression in roots and shoots of *L. sativa* 'Salinas'. (E) The number of negative regulatory elements found in 5' promoter regions of the two genotypes.

sodium accumulation than the wild lettuce (**Figure S1B**). The K^+ accumulation in roots of *L. sativa* 'Salinas' was rising with time after exposed to external salinity, from 3.4 to 15.2 g/Kg DW whereas that of *L. serriola* did not show as strong as an increase, from 7.5 to 12.4 g/Kg DW (**Figure S1C**). The potassium contents in *Lactuca* shoots of two species both illustrated an increasing trend during the first 12 hours after salt stress and then reached a relatively stable stage from 24 hours to 8 days (**Figure S1D**). However, the concentrations of Na^+ and K^+ in the whole plants did not show significant differences between the two lettuce species, except the Na^+ contents in control condition (**Figure S1E & F**). The Na^+/K^+ ratios in roots and shoots of both species showed an increasing trend (**Figure 4C and D**). In roots, *L. serriola* showed a higher Na^+/K^+ ratio through the whole time-course while both species had similar ratios in shoots within 12 hours and afterwards *L. sativa* 'Salinas' began to show higher Na^+/K^+ ratios (**Figure 4C and D**). The Na^+/K^+ ratios in the whole plants demonstrated significances at time 0 (control condition) and 6h (75 mM NaCl treatment) (**Figure 4E**).

The tissue-specific Na^+/K^+ ratio and relative *HKT1;1* gene expression in each genotype were compared and illustrated in **Figure 5**. The composition of root Na^+/K^+ ratio in the sum of Na^+/K^+ ratio in shoot and root decreased during time series, both for *L. serriola* (**Figure 5A**) and *L. sativa* 'Salinas' (**Figure 5B**). The comparison of relative *HKT1;1* gene expression in different tissues of *L. serriola* (**Figure 5C**) and *L. sativa* 'Salinas' (**Figure 5D**) both showed much higher *HKT1;1* expression in shoot than in root. In both genotypes, the dynamically tissue-specific Na^+/K^+ ratio was generally consistent with tissue-specific *HKT1;1* gene expression at the same time points. The negative regulatory elements found in promoter regions were also shown (**Figure 5E**). A total of 23 ARR1AT boxes, 7 AP2 domains and 7 bZIP binding sites were detected in the upstream of the promoter regions of *L. serriola* whereas the numbers in *L. sativa* 'Salinas' were 16, 4 and 1, respectively (**Figure 5E**).

Discussion

The *LsaHKT1;1* (*L. sativa* 'Salinas') and *LseHKT1;1* (*L. serriola*) are alleles of a Class I *HKT* transporter

The *HKT* gene family is responsible for Na^+ distribution and Na^+/K^+ homeostasis in plants (Hamamoto et al. 2015; Hauser and Horie 2010; Rodr ǵuez-Navarro and Rubio 2006; Xue et

al. 2011) and has been reported to play a crucial role in enhancing salt tolerance in plants (Almeida et al. 2014b; Horie et al. 2009; Rus et al. 2006; Sanadhya et al. 2015; Sunarpi et al. 2005). There are two classes of *HKT* transporters in vascular plants with putatively distinct ion selectivities (Hauser and Horie 2010; Platten et al. 2006). Class I *HKT* transporters (*HKT1*) usually mediate relatively Na^+ selective transport whereas Class II *HKT* transporters (*HKT2*) are mainly responsible for Na^+/K^+ transport activity (Horie et al. 2001; Kader et al. 2006; Oomen et al. 2012; Sanadhya et al. 2015; Uozumi et al. 2000).

A previous QTL mapping study in *Lactuca* identified one major QTL region on chromosome 7 (qLS7.2) contributing to Na^+/K^+ homeostasis in lettuce leaves under salt conditions, using seedlings from a recombinant inbred line population derived from the lettuce crop (*L. sativa* 'Salinas') and wild species (*L. serriola*) (Wei et al. 2014). We have identified an *HKT1*-like locus (*HKT1;1*) near the position of the maximum significance of the LOD value for the major QTL related to Na^+ and K^+ concentrations. Another *HKT1*-like locus (*LsaHKT1;2*) was found on chromosome 4. In this study, we performed a phylogenetic analysis of *HKT1*-like protein sequences (*LsaHKT1;1*, *LseHKT1;1*, and *LsaHKT1;2*) in the two lettuce species and other published *HKT* protein sequences. The results confirmed that the two *HKT1* loci were *HKT1* homologs, and not HKT2 homologs, and showed that the two *HKT1;1* alleles in lettuce were almost identical except two residue substitutions. The two *HKT1* loci (*HKT1;1* and *HKT1;2*) are likely derived from the ancient polyploidy events occurring in the early history of the Asteraceae (Barker et al. 2008; Barreda et al. 2015).

The glycine residues in four-loop-per-subunit *HKT2* transporters have been identified to provide the potassium selectivity and *HKT1* transporters have a serine at the filter position in the P_A-loop function mainly as Na^+ transporters in plants (Corratgé-Faillie et al. 2010; Hauser and Horie 2010; Horie et al. 2009; Mäser et al. 2002b). It should be noted that *OsHKT2;1* is an exception with a serine in the P_A region and shows a robust Na^+ selectivity in yeast and *Xenopus oocytes* (Garciadeblás et al. 2003; Horie et al. 2001; Mäser et al. 2002b). The members of Class I *HKT* transporters have been reported as low affinity and specific Na^+ transporters located in the plasma membrane of parenchyma cells surrounding the xylem vessels and to upload Na^+ from xylem, preventing Na^+ accumulation in shoots (Asins et al. 2013; Ben Amar et al. 2014; Davenport et al. 2007; Jha et al. 2010; Maathuis 2014; Munns et al. 2012; Ren et al. 2005; Sunarpi et al. 2005; Uozumi et al. 2000; Xue et al. 2011). However, some exceptions have been observed for members of Class I, especially for when they were expressed in heterologous systems (Fairbairn et al. 2000; Su et al. 2003). Two *HKT1* isoforms were tested for transporter activity and ion selection in the halophytic *Thellungiella salsuginea* and one of them, *TsHKT1;2*, showed a strong K^+ transporter activity and selectivity for K^+ over Na^+ (Ali et al. 2012). The presence of two aspartic residues (D), D207 and D238, were considered as the key features of K^+ transport capacity. At these two positions, asparagine residues (N) were found in *Arabidopsis* and other known plant sequences (Ali et al. 2012).

The presence of Ser instead of Gly in the P_A-loop of the two lettuce *HKT1;1* alleles indicated a potential Na^+ selectivity for the lettuce *HKT1;1* transporters. In addition, *LsaHKT1;1* and *LseHKT1;1* both contain two asparagine residues at the two important positions for K^+ transport capacity, implying a preferential Na^+ selectivity. Functional analysis and gene expression of these two *Lactuca HKT1;1* alleles should be compared in heterologous systems to confirm this prediction in the future.

Lactuca HKT1;1 locus as candidate gene for major QTL for Na^+/K^+ homeostasis supported by differential expression of *Lactuca HKT1;1* alleles and differential ion accumulation

Salt tolerance has been indicated as a quantitative trait in plants (Asins et al. 2015; Cuartero et al. 2006; Flowers 2004; Nguyen et al. 2012; Ren et al. 2005). The *HKT1*-like coding genes have been shown to be of crucial importance to improve salt tolerance (Almeida et al. 2014a; Huang et al. 2006; James et al. 2011; James et al. 2012; James et al. 2006; Munns et al. 2012; Ren et al. 2005). Therefore, we analyzed a temporal series of gene expression of the two *Lactuca HKT1;1* alleles, *LsaHKT1;1* and *LseHKT1;1*, and Na^+ and K^+ contents in different tissues.

The results of the expression of *Lactuca HKT1;1* alleles show a complex pattern in different tissues. Although both genotypes demonstrated low expression levels in roots, *LseHKT1;1* had a higher relative expression than that of *LsaHKT1;1* (**Figure 4A**). This means a greater potential Na^+ retrieval in root xylem of *L. serriola* than that of *L. sativa* 'Salinas,' thus leading to less Na^+ accumulation in shoots of *L. serriola* than that of *L. sativa* 'Salinas'. In contrast, *LsaHKT1;1* showed a much higher expression than *LseHKT1;1* in shoots (**Figure 4B**), indicating more Na^+ recirculation in shoot xylem of *L. sativa* 'Salinas' than that of *L. serriola*. The measurements of Na^+ contents in lettuce roots and shoots were consistent with the expression patterns (**Figure S1A & B**). In the previous QTL study, the major QTL co-localizing with the *HKT1;1* genes was identified associating with Na^+ and K^+ accumulation in lettuce leaves (Wei et al. 2014), which was in line with the higher expression levels of the two lettuce *HKT1* alleles in shoots than in roots in this study. In addition, the alleles from the cultivated lettuce were found to contribute more to the major QTL region for Na^+/K^+ ratio and Na^+ accumulation in leaves (Wei et al. 2014). Indeed, the direction of the additive effects was consistent with the higher expression of *LsaHKT1;1* in shoots of *L. sativa* 'Salinas'.

The differential expression of *LsaHKT1;1* and *LseHKT1;1* might be explained by major differences in their 5' promoter sequences that could alter the potential binding of cis-acting regulatory elements. Several transcription factor family genes, e.g. *Arabidopsis* response regulator (ARR) 1 and ARR12 (Mason et al. 2010), ABA-INSENSITIVE (ABI) 4 (Shkolnik-Inbar et al. 2013), APETALA2/ETHYLENE RESPONSE FACTOR (AP2/ERF) (Kasuga et al. 1999) and basic leucine zipper (bZIP) 24 (Yang et al. 2009) have been reported to as negative regulators of *AtHKT1;1* in the roots. Other transcription factors including WRKY (Guo et al.

2011; Mondini et al. 2012), MYB (Cui et al. 2013), basic helix-loop-helix (bHLH) (Jiang et al. 2009) and NAC (Tran et al. 2004), are differentially expressed in response to external salinity. In total, 23 ARR1AT boxes, 7 AP2 domains and 7 bZIP binding sites were found in the 2 kb upstream of the promoter regions of *L. serriola* whereas the numbers in *L. sativa* 'Salinas' were 16, 4 and 1, respectively (**Figure 5E**). In addition, the unique positions of regulatory elements of *L. serriola* included 7 bZIP binding sites (potential negative regulatory elements) whereas those of *L. sativa* 'Salinas' contained only 2 AP2 domains and 1 bZIP binding sites. These negative transcription regulators of *AtHKT1;1*s in roots could be the reason for low expression of *HKT1*s in lettuce roots but not shoots. The mechanism of how these cis-acting elements working on the expression of *LsaHKT1;1* and *LseHKT1;1* remains unclear and needs further study.

The amino acid differences between *LsaHKT1;1* and *LseHKT1;1* alleles might cause phenotypic differences of the transporters. One change was in the end of $M2_D$ domain while the other was neither in the transmembrane nor pore segments (**Figure 2**). A Tyrosine (Y or Tyr) in $M2_D$ domain of *LsaHKT1;1* changes to Histidine (H or His) in *LseHKT1;1*, whereas a Threonine (T or Thr) is in *Arabidopsis*. His is positive amino acid, Tyr is hydrophilic and Thr is non-charged. Although His and Tyr have similar structures, Tyr has a phenolic OH-group. This difference in $M2_D$ domain of *Lactuca HKT1;1*s may be important for pH and ion selectivity. Further study is still needed to test the functional differences between the two alleles.

Although the accumulation of Na^+ and K^+ in the whole plants did not show significant differences between the two *Lactuca* species, the Na^+/K^+ ratios demonstrated significance at time 0 (control condition) and 6h (75 mM NaCl treatment), indicating a fast response to elevated external salinity in wild lettuce (**Figure 4E**). The Na^+/K^+ ratio is considered as an important indicator of tolerance to salinity in plants, and the more tolerant plant usually has lower Na^+/K^+ ratio (Lin et al. 2004). *LseHKT1;1* showed much lower expressions than *LsaHKT1;1* in shoots for the first 24 hours after exposed to salinity, and thus the Na^+/K^+ ratios of total plant of *L. serriola* were higher than that of *L. sativa* 'Salinas' during the same time, implying *L. sativa* 'Salinas' was more tolerant than *L. serriola*. However, from 24 hours afterwards, no significant differences of Na^+/K^+ ratios were observed between the two genotypes as well as the expression of *HKT1*s.

The high Na^+ accumulation in shoots of *L. sativa* 'Salinas' might be compensated for by other genes involved in the mechanism of salt tolerance, such as the tonoplast-localized Na^+/K^+ exchangers (such as NHX) (Barragán et al. 2012; Rodríguez-Rosales et al. 2009) and cell membrane-localized SALT OVERLY SENSITIVE (SOS) Na^+/K^+ antiporters (Huertas et al. 2012; Jarvis et al. 2014; Katschnig et al. 2015; Yamaguchi et al. 2013). Future research should include more genes related to salinity and the effect of external salinity on total growth of the plants (biomass) should also be measured.

In addition, the *HKT1* gene in *Arabidopsis* has been reported to be expressed in the root stele and leaf vasculature (Mäser et al. 2002a) and be responsible for removing Na^+ from xylem of roots to protect Na^+ accumulation in shoots (Moller et al. 2009; Sunarpi et al. 2005). Nevertheless, the expression of the two lettuce *HKT1* alleles in our study fit another alternate model (Berthomieu et al. 2003), implying *AtHKT1;1* was expressed in the phloem and its activity might contribute to the circulation of Na^+ in the whole plant. Additionally, a soil bacterium, *Bacillus subtilis* GB03, conferred salt tolerance in *A. thaliana* by concurrently down- and upregulating *HKT1* expression in roots and shoots, as observed in our study, and resulting in lower Na^+ accumulation throughout the plant compared with controls (Zhang et al. 2008).

In conclusion, the low expression of *HKT1;1*s in lettuce roots might be explained by the promoter regions containing binding sites of negative transcription factors, which have been reported to preferentially express in roots of *Arabidopsis* and reduce *AtHKT1;1* expression in roots. The *HKT1* alleles in lettuce, *LsaHKT1;1* (*L. sativa* 'Salinas') and *LseHKT1;1* (*L. serriola*), were mainly expressed in shoots, showing a time-dependent pattern. Significant differences of *HKT1;1* expression were observed in shoots in early stage (0-24 hours) and in roots in late stage (2-6 days). Functional analysis of the two lettuce *HKT1* alleles and the total growth of lettuce related to salt induction should be included in future research.

Table 3 Unique cis-acting regulatory elements in 5' UTR promoter region of the two lettuce genotypes

Site name	Loc.	Species	Orient.	Seq.	Description
PRECONSCRHSP70A	4	ser	(+)	SCGAYNRNNNNNNNNNNNNNNN HD	HSP; chlorophyl; MgProto;
WUSATAg	107	ser	(+)	TTAATGG	Target sequence of WUS in the intron of AGAMOUS gene in Arabidopsis; See Lohmann et al. Cell 105:793-803 (2003)
AMYBOX2	110	ser	(-)	TATCCAT	amylase; seed;
QELEMENTZMZM13	231	ser	(+)	AGGTCA	enhancing; ZM13; LAT52; pollen;
BOXIINTPATPB	299	ser	(-)	ATAGAA	plastid; NEP; atpB; PatpB; NCII; Box I; Box II;
TATABOX3	355	ser	(+)	TATTAAT	TATA; sporamin;
RYREPEATGMGY2	391	ser	(-)	CATGCAT	glycinin; CATGCAT; Gy2; seed;
RYREPEATLEGUMINBOX	391	ser	(-)	CATGCAY	RY repeat; legumin box; seed; storage protein;
RYREPEATBNNAPA	392	ser	(-)	CATGCA	RY repeat; RY/G box; seed; napA; napin;
DPBFCOREDCDC3	394	ser	(-)	ACACNNG	Dc3; lea class gene; embryo; ABA; DPBF-1, DPBF-2; bZIP; GIA1;,ABI5; seed;
MYCATERD1	394	ser	(+)	CATGTG	water-stress; erd;
MYCATRD22	394	ser	(-)	CACATG	Dehydration; Water stress; ABA; MYC; myc; leaf; shoot;
ASF1MOTIFCAMV	415	ser	(+)	TGACG	TGACG; root; leaf; CaMV; 35S; promoter; auxin; salicylic acid;,light; as-1; TGA1a, TGA1b; CREB; ASF1; TGA6; shoot; xenobiotic,stress; SAR; SA; Disease resistance;
HEXMOTIFTAH3H4	415	ser	(-)	ACGTCA	hexamer; HBP-1A; HBP-1B; histone H3; CaMV; 35S; NOS; HBP-1;,Leucine zipper motif; meristem; OBF1; bZIP; lip19; LIP19;
PALINDROMICCBOXGM	415	ser	(-)	TGACGTCA	C-box; bZIP; STGA1; STF; hypocotyl; TGA; SA;
PALINDROMICCBOXGM	415	ser	(+)	TGACGTCA	C-box; bZIP; STGA1; STF; hypocotyl; TGA; SA;
TGACGTVMAMY	415	ser	(+)	TGACGT	alpha-Amylase; cotyledon; seed germination; seed;
ACGTCBOX	416	ser	(-)	GACGTC	C-box; ACGT element; seed;
ACGTCBOX	416	ser	(+)	GACGTC	C-box; ACGT element; seed;

ACGTATERD1	417	ser	(-)	ACGT	ACGT; etiolation; erd;
ACGTATERD1	417	ser	(+)	ACGT	ACGT; etiolation; erd;
HEXMOTIFTAH3H4	417	ser	(+)	ACGTCA	hexamer; HBP-1A; HBP-1B; histone H3; CaMV; 35S; NOS; HBP-1;,Leucine zipper motif; meristem; OBF1; bZIP; lip19; LIP19;
TGACGTVMAMY	417	ser	(-)	TGACGT	alpha-Amylase; cotyledon; seed germination; seed;
ASF1MOTIFCAMV	418	ser	(-)	TGACG	TGACG; root; leaf; CaMV; 35S; promoter; auxin; salicylic acid;,light; as-1; TGA1a, TGA1b; CREB; ASF1; TGA6; shoot; xenobiotic,stress; SAR; SA; Disease resistance;
ACGTTBOX	433	ser	(-)	AACGTT	T-box; T box; ACGT element:
ACGTTBOX	433	ser	(+)	AACGTT	T-box; T box; ACGT element:
ACGTATERD1	434	ser	(-)	ACGT	ACGT; etiolation; erd;
ACGTATERD1	434	ser	(+)	ACGT	ACGT; etiolation; erd;
TGTCACACMCUCUMISIN	460	ser	(-)	TGTCACA	cucumisin; fruit;
AMYBOX2	467	ser	(+)	TATCCAT	amylase; seed;
DPBFCOREDCDC3	564	ser	(+)	ACACNNG	Dc3; lea class gene; embryo; ABA; DPBF-1, DPBF-2; bZIP; GIA1;,ABI5; seed;
MYBPLANT	574	ser	(+)	MACCWAMC	Myb; MYB; Myb305; AmMYB308; AmMYB330; flower; PAL; CHS; DFR;,Candi; Bz1; phenylpropanoid; lignin; leaf; shoot;
HBOXCONSENSUSPVCHS	576	ser	(+)	CCTACCNNNNNNNCT	H-box; H box; CHS; chs; light regulation; light; elicitor;,stress; transposon; wounding; leaf; shoot; Ku autoantigen;,KAP-2;
SURECOREATSULTR11	585	ser	(-)	GAGAC	sulfate uptake; sulfate transporter; ARF; -S; S;
PROXBBNNAPA	615	ser	(-)	CAAACACC	ABRE; ABA; prox B; B-box; seed; napA; napin;
BOXIINTPATPB	708	ser	(-)	ATAGAA	plastid; NEP; atpB; PatpB; NCII; Box I; Box II;
E2FCONSENSUS	725	ser	(-)	WTTSSCSS	E2F
CACGCAATGMGH3	951	ser	(+)	CACGCAAT	D1; D4; GH3; Auxin;
MARTBOX	984	ser	(-)	TTWTWTTWTT	MAR; SAR; T-box; scaffold; matrix;
MARTBOX	985	ser	(-)	TTWTWTTWTT	MAR; SAR; T-box; scaffold; matrix;

MARTBOX	986	ser	(-)	TTWTWTTWTT	MAR; SAR; T-box; scaffold; matrix;
MARTBOX	987	ser	(-)	TTWTWTTWTT	MAR; SAR; T-box; scaffold; matrix;
MARTBOX	988	ser	(-)	TTWTWTTWTT	MAR; SAR; T-box; scaffold; matrix;
MARTBOX	989	ser	(-)	TTWTWTTWTT	MAR; SAR; T-box; scaffold; matrix;
MARTBOX	990	ser	(-)	TTWTWTTWTT	MAR; SAR; T-box; scaffold; matrix;
MARTBOX	991	ser	(-)	TTWTWTTWTT	MAR; SAR; T-box; scaffold; matrix;
BOXIINTPATPB	1098	ser	(+)	ATAGAA	plastid; NEP; atpB; PatpB; NCII; Box I; Box II;
RBCSCONSENSUS	1353	ser	(+)	AATCCAA	rbcS; G box; I box; leaf; shoot;
CCA1ATLHCB1	1430	ser	(-)	AAMAATCT	CCA1; Lhcb; shoot; leaf;
RBCSCONSENSUS	1521	ser	(+)	AATCCAA	rbcS; G box; I box; leaf; shoot;
LEAFYATAG	1524	ser	(+)	CCAATGT	LEAFY; AGAMOUS;
MARARS	1630	ser	(+)	WTTTATRTTTW	MAR; SAR; ARS;
SORLREP3AT	1716	ser	(+)	TGTATATAT	phyA; phytochrome; light;
ACGTABREMOTIFA2OSEM	1754	ser	(-)	ACGTGKC	ABA; ABRE; motif A; DRE;
GADOWNAT	1754	ser	(-)	ACGTGTC	Ga; seed; germaination;
ABRELATERD1	1756	ser	(-)	ACGTG	ABRE; etiolation; erd;
ACGTATERD1	1757	ser	(-)	ACGT	ACGT; etiolation; erd;
ACGTATERD1	1757	ser	(+)	ACGT	ACGT; etiolation; erd;
HEXMOTIFTAH3H4	1757	ser	(+)	ACGTCA	hexamer; HBP-1A; HBP-1B; histone H3; CaMV; 35S; NOS; HBP-1;,Leucine zipper motif; meristem; OBF1; bZIP; lip19; LIP19;
TGACGTVMAMY	1757	ser	(-)	TGACGT	alpha-Amylase; cotyledon; seed germination; seed;
ASF1MOTIFCAMV	1758	ser	(-)	TGACG	TGACG; root; leaf; CaMV; 35S; promoter; auxin; salicylic acid;,light; as-1; TGA1a, TGA1b; CREB; ASF1; TGA6; shoot; xenobiotic,stress; SAR; SA; Disease resistance;
ABRELATERD1	1769	ser	(-)	ACGTG	ABRE; etiolation; erd;
T/GBOXATPIN2	1769	ser	(-)	AACGTG	T/G-box; JA; pin2; LAP; MYC; wounding;
ACGTATERD1	1770	ser	(-)	ACGT	ACGT; etiolation; erd;

ACGTATERD1	1770	ser	(+)	ACGT	ACGT; etiolation; erd;
QELEMENTZMZM13	1791	ser	(-)	AGGTCA	enhancing; ZM13; LAT52; pollen;
BOXIINTPATPB	1858	ser	(+)	ATAGAA	plastid; NEP; atpB; PatpB; NCII; Box I; Box II;
ACGTATERD1	1865	ser	(-)	ACGT	ACGT; etiolation; erd;
ACGTATERD1	1865	ser	(+)	ACGT	ACGT; etiolation; erd;
ACGTATERD1	1983	ser	(-)	ACGT	ACGT; etiolation; erd;
ACGTATERD1	1983	ser	(+)	ACGT	ACGT; etiolation; erd;
LTRECOREATCOR15	53	sat	(-)	CCGAC	low temperature; cold; LTRE; drought; ABA; cor15a; BN115; leaf;,shoot; phytochrome;
CEREGLUBOX2PSLEGA	72	sat	(-)	TGAAAACT	legumin; glutenin; cereal; legA; seed;
CBFHV	137	sat	(-)	RYCGAC	CBF; AP2 domain; CRT/DRE; low temperature;
DRECRTCOREAT	137	sat	(-)	RCCGAC	DRE/CRT; drought; high-light; cold; DREB; DREB1; DREB2; CBF;
LTRECOREATCOR15	137	sat	(-)	CCGAC	low temperature; cold; LTRE; drought; ABA; cor15a; BN115; leaf;,shoot; phytochrome;
SORLIP2AT	140	sat	(-)	GGGCC	phyA; phytochrome; light;
SORLIP2AT	173	sat	(-)	GGGCC	phyA; phytochrome; light;
SORLIP2AT	249	sat	(-)	GGGCC	phyA; phytochrome; light;
NRRBNEXTA	283	sat	(+)	TAGTGGAT	ext; extensin; stem; internode; petiole; root;
TATCCACHVAL21	285	sat	(-)	TATCCAC	gibberellin; GA; GARC;
TATCCAOSAMY	286	sat	(-)	TATCCA	alpha-amylase; MYB proteins; gibberellin; GA; sugar starvation;
SORLIP2AT	350	sat	(-)	GGGCC	phyA; phytochrome; light;
MYB2AT	355	sat	(-)	TAACTG	MYB; myb; SV40; enhancer; bronze; bronze-1; leaf; shoot;
MYB2CONSENSUSAT	355	sat	(-)	YAACKG	MYB; rd22BP1; ABA; leaf; seed; stress;
AUXREPSIAA4	391	sat	(+)	KGTCCCAT	Auxin; AuxRE; root; meristem;
GGTCCCATGMSAUR	391	sat	(+)	GGTCCCAT	SAUR; NDE; Auxin;
LTRE1HVBLT49	404	sat	(-)	CCGAAA	low temperature; LTRE;
SORLIP2AT	433	sat	(-)	GGGCC	phyA; phytochrome; light;

MYB2AT	438	sat	(-)	TAACTG	MYB; myb; SV40; enhancer; bronze; bronze-1; leaf; shoot;
MYB2CONSENSUSAT	438	sat	(-)	YAACKG	MYB; rd22BP1; ABA; leaf; seed; stress;
SORLIP2AT	517	sat	(-)	GGGCC	phyA; phytochrome; light;
SORLIP2AT	558	sat	(-)	GGGCC	phyA; phytochrome; light;
S1FBOXSORPS1L21	585	sat	(+)	ATGGTA	S1F; S1F box; S1F-box; S1; plastid protein; RPS1; RPL21; leaf;,negative;
CEREGLUBOX2PSLEGA	659	sat	(-)	TGAAAACT	legumin; glutenin; cereal; legA; seed;
MYB2AT	707	sat	(-)	TAACTG	MYB; myb; SV40; enhancer; bronze; bronze-1; leaf; shoot;
MYB2CONSENSUSAT	707	sat	(-)	YAACKG	MYB; rd22BP1; ABA; leaf; seed; stress;
ERELEE4	749	sat	(-)	AWTTCAAA	Ethylene; E4; GST1; senescence; ERE; fruit;
SEF1MOTIF	796	sat	(-)	ATATTTAWW	SOYBEAN; STORAGE PROTEIN; 7S; GLOBULIN; BETA-CONGLICININ; seed;
TATABOX2	796	sat	(+)	TATAAAT	TATA; legA; phaseolin;
L1BOXATPDF1	871	sat	(-)	TAAATGYA	PDF1; L1 box; L1 layer-specific expression; Shoot apical,meristem; SAM; organ primordia; cotton fiber; HDZip; homeodomain;,leucine zipper;
MYB2AT	890	sat	(-)	TAACTG	MYB; myb; SV40; enhancer; bronze; bronze-1; leaf; shoot;
MYB2CONSENSUSAT	890	sat	(-)	YAACKG	MYB; rd22BP1; ABA; leaf; seed; stress;
AACACOREOSGLUB1	907	sat	(-)	AACAAAC	glutelin; AACA; GCN4; seed; endosperm;
AMYBOX1	913	sat	(-)	TAACARA	amylase; seed;
GAREAT	913	sat	(-)	TAACAAR	GARE; GA;
MYBGAHV	913	sat	(-)	TAACAAA	myb; Myb; GAmyb; GA; gibberellin; GARC; alph-amylase; amylase;,aleurone; GARE; seed;
S1FBOXSORPS1L21	1060	sat	(+)	ATGGTA	S1F; S1F box; S1F-box; S1; plastid protein; RPS1; RPL21; leaf;,negative;
S1FSORPL21	1060	sat	(+)	ATGGTATT	S1F; S1; plastid protein; RPL21; leaf; negative;
TATABOXOSPAL	1127	sat	(-)	TATTTAA	TBP; TFIIB; pal; DNA binding and bending;
CBFHV	1362	sat	(+)	RYCGAC	CBF; AP2 domain; CRT/DRE; low temperature;
IBOX	1452	sat	(+)	GATAAG	I box; I-box; rbcS; light regulation; light; LeMYB1,

					Myb-like,protein; leaf; shoot;
IBOXCORENT	1452	sat	(+)	GATAAGR	I-box; CAM; light;
S1FBOXSORPS1L21	1473	sat	(-)	ATGGTA	S1F; S1F box; S1F-box; S1; plastid protein; RPS1; RPL21; leaf;,negative;
MYB2AT	1598	sat	(-)	TAACTG	MYB; myb; SV40; enhancer; bronze; bronze-1; leaf; shoot;
MYB2CONSENSUSAT	1598	sat	(-)	YAACKG	MYB; rd22BP1; ABA; leaf; seed; stress;
AMYBOX1	1645	sat	(-)	TAACARA	amylase; seed;
GAREAT	1645	sat	(-)	TAACAAR	GARE; GA;
MYBGAHV	1645	sat	(-)	TAACAAA	myb; Myb; GAmyb; GA; gibberellin; GARC; alph-amylase; amylase;,aleurone; GARE; seed;
PREATPRODH	1653	sat	(-)	ACTCAT	proline; ProDH; hypoosomolarity; bZIP;
MYB2AT	1671	sat	(+)	TAACTG	MYB; myb; SV40; enhancer; bronze; bronze-1; leaf; shoot;
MYB2CONSENSUSAT	1671	sat	(+)	YAACKG	MYB; rd22BP1; ABA; leaf; seed; stress;
ERELEE4	1683	sat	(-)	AWTTCAAA	Ethylene; E4; GST1; senescence; ERE; fruit;
MYB2CONSENSUSAT	1712	sat	(+)	YAACKG	MYB; rd22BP1; ABA; leaf; seed; stress;
MARABOX1	1724	sat	(+)	AATAAAYAAA	MAR; SAR; matrix; A-box; scaffold;
SEF1MOTIF	1728	sat	(-)	ATATTTAWW	SOYBEAN; STORAGE PROTEIN; 7S; GLOBULIN; BETA-CONGLICININ; seed;
SEF1MOTIF	1736	sat	(-)	ATATTTAWW	SOYBEAN; STORAGE PROTEIN; 7S; GLOBULIN; BETA-CONGLICININ; seed;
TATABOX2	1736	sat	(+)	TATAAAT	TATA; legA; phaseolin;
S1FSORPL21	1941	sat	(-)	ATGGTATT	S1F; S1; plastid protein; RPL21; leaf; negative;
S1FBOXSORPS1L21	1943	sat	(-)	ATGGTA	S1F; S1F box; S1F-box; S1; plastid protein; RPS1; RPL21; leaf;,negative;
ACGTATERD1	2046	sat	(-)	ACGT	ACGT; etiolation; erd;
ACGTATERD1	2046	sat	(+)	ACGT	ACGT; etiolation; erd;

Acknowledgements

Thanks to China Scholarship Council (http://en.csc.edu.cn/) for providing funding to Z.W. We appreciate Magdalena M. Julkowska and Christa Testerink for data analysis, Tao Zhao for promoter sequence analysis, Setareh Mohammadin for figure making and Andre Maassen for helping with the hydroponic experiment in greenhouse.

References

Ahmadi N, Negrão S, Katsantonis D, Frouin J, Ploux J, Letourmy P, Droc G, Babo P, Trindade H, Bruschi G, Greco R, Oliveira MM, Piffanelli P, Courtois B (2011) Targeted association analysis identified japonica rice varieties achieving Na^+/K^+ homeostasis without the allelic make-up of the salt tolerant indica variety Nona Bokra. Theoretical and Applied Genetics 123:881-895

Ali Z, Park HC, Ali A, Oh DH, Aman R, Kropornicka A, Hong H, Choi W, Chung WS, Kim WY, Bressan RA, Bohnert HJ, Lee SY, Yun DJ (2012) TsHKT1;2, a HKT1 homolog from the extremophile Arabidopsis relative *Thellungiella salsuginea*, shows K^+ specificity in the presence of NaCl. Plant Physiol 158:1463-1474

Almeida P, de Boer GJ, de Boer AH (2014a) Differences in shoot Na^+ accumulation between two tomato species are due to differences in ion affinity of HKT1;2. Journal of plant physiology 171:438-447

Almeida PM, de Boer GJ, de Boer AH (2014b) Assessment of natural variation in the first pore domain of the tomato HKT1;2 transporter and characterization of mutated versions of SlHKT1;2 expressed in *Xenopus laevis* oocytes and via complementation of the salt sensitive athkt1;1 mutant. Frontiers in plant science 5:600

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215:403-410

Asins MJ, Raga V, Roca D, Belver A, Carbonell EA (2015) Genetic dissection of tomato rootstock effects on scion traits under moderate salinity. Theoretical and applied genetics 128:667-679

Asins MJ, Villalta I, Aly MM, Olias R, Alvarez DEMP, Huertas R, Li J, Jaime-Perez N, Haro R, Raga V, Carbonell EA, Belver A (2013) Two closely linked tomato HKT coding genes are positional candidates for the major tomato QTL involved in Na^+ /K^+ homeostasis. Plant, cell & environment 36:1171-1191

Barker MS, Kane NC, Matvienko M, Kozik A, Michelmore RW, Knapp SJ, Rieseberg LH (2008) Multiple Paleopolyploidizations during the Evolution of the Compositae Reveal Parallel Patterns of Duplicate Gene Retention after Millions of Years. Molecular biology and evolution 25:2445-2455

Barragán V, Leidi EO, Andrés Z, Rubio L, De Luca A, Fernández JA, Cubero B, Pardo JM (2012) Ion Exchangers NHX1 and NHX2 Mediate Active Potassium Uptake into Vacuoles to Regulate Cell Turgor and Stomatal Function in *Arabidopsis*. The Plant cell 24:1127-1142

Barreda VD, Palazzesi L, Tellería MC, Olivero EB, Raine JI, Forest F (2015) Early evolution of the angiosperm clade Asteraceae in the Cretaceous of Antarctica. Proceedings of the National Academy of Sciences 112:10989-10994

Begara-Morales JC, Sánchez-Calvo B, Chaki M, Valderrama R, Mata-Pérez C, López-Jaramillo J, Padilla MN, Carreras A, Corpas FJ, Barroso JB (2014) Dual regulation of cytosolic ascorbate peroxidase (APX) by tyrosine nitration and S-nitrosylation. Journal of experimental botany 65:527-538

Ben Amar S, Brini F, Sentenac H, Masmoudi K, Very AA (2014) Functional characterization in *Xenopus* oocytes of Na^+ transport systems from durum wheat reveals diversity among two HKT1;4 transporters. Journal of experimental botany 65:213-222

Berthomieu P, Conéjéro G, Nublat A, Brackenbury WJ, Lambert C, Savio C, Uozumi N, Oiki S, Yamada K, Cellier F, Gosti F, Simonneau T, Essah PA, Tester M, Véry A-A, Sentenac H, Casse F (2003) Functional analysis of AtHKT1 in Arabidopsis shows that Na^+ recirculation by the phloem is crucial for salt tolerance. The EMBO journal 22:2004-2014

Borowski JM, Galli V, Messias Rda S, Perin EC, Buss JH, dos Anjos e Silva SD, Rombaldi CV (2014) Selection of candidate reference genes for real-time PCR studies in lettuce under abiotic stresses. Planta 239:1187-1200

CGPDB (2014) Compositae genome project database. http://compgenomics.ucdavis.edu/

Corratgé-Faillie C, Jabnoune M, Zimmermann S, Véry AA, Fizames C, Sentenac H (2010) Potassium and sodium transport in non-animal cells: the Trk/Ktr/HKT transporter family. Cellular and Molecular Life Sciences 67:2511-2532

Cuartero J, Bolarín MC, Asíns MJ, Moreno V (2006) Increasing salt tolerance in the tomato. Journal of experimental botany 57:1045-1058

Cui MH, Yoo KS, Hyoung S, Nguyen HTK, Kim YY, Kim HJ, Ok SH, Yoo SD, Shin JS (2013) An Arabidopsis R2R3-MYB transcription factor, AtMYB20, negatively regulates type 2C serine/threonine protein phosphatases to enhance salt tolerance. FEBS letters 587:1773-1778

Danku J, Lahner B, Yakubova E, Salt D (2013) Large-scale plant ionomics. In: Maathuis F (ed) Methods Mol Biol. Humana Press, New York, pp 255-276

Davenport RJ, Munoz-Mayor A, Jha D, Essah PA, Rus A, Tester M (2007) The Na^+ transporter AtHKT1;1 controls retrieval of Na^+ from the xylem in *Arabidopsis*. Plant, cell & environment 30:497-507

Deinlein U, Stephan AB, Horie T, Luo W, Xu G, Schroeder JI (2014) Plant salt-tolerance mechanisms. Trends in plant science 19:371-379

Durell SR, Guy HR (1999) Structural models of the KtrB, TrkH, and Trk1,2 symporters based on the structure of the KcsA K^+ channel. Biophysical Journal 77:789-807

Fairbairn D, Liu W, Schachtman D, Gomez-Gallego S, Day S, Teasdale R (2000) Characterisation of two distinct HKT1-like potassium transporters from *Eucalyptus camaldulensis*. Plant molecular biology 43:515-525

Flowers TJ (2004) Improving crop salt tolerance. Journal of experimental botany 55:307-319

Gabrijel O, Zed R, Szilvia V (2011) Soil Salinisation and Salt Stress in Crop Production

Garciadeblás B, Senn ME, Bañuelos MA, Rodríguez-Navarro A (2003) Sodium transport and HKT transporters: the rice model. The Plant Journal 34:788-801

Guo R, Yu F, Gao Z, An H, Cao X, Guo X (2011) GhWRKY3, a novel cotton (*Gossypium hirsutum* L.) WRKY gene, is involved in diverse stress responses. Molecular Biology Reports 38:49-58

Hamamoto S, Horie T, Hauser F, Deinlein U, Schroeder JI, Uozumi N (2015) HKT transporters mediate salt stress resistance in plants: from structure and function to the field. Current opinion in biotechnology 32:113-120

Hauser F, Horie T (2010) A conserved primary salt tolerance mechanism mediated by HKT transporters: a mechanism for sodium exclusion and maintenance of high K^+/Na^+ ratio in leaves during salinity stress. Plant, cell & environment 33:552-565

Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. Nucleic acids research 27:297-300

Horie T, Hauser F, Schroeder JI (2009) HKT transporter-mediated salinity resistance mechanisms in *Arabidopsis* and monocot crop plants. Trends in plant science 14:660-668

Horie T, Yoshida K, Nakayama H, Yamada K, Oiki S, Shinmyo A (2001) Two types of HKT transporters with different properties of Na^+ and K^+ transport in *Oryza sativa*. The Plant Journal 27:129-138

Huang S, Spielmeyer W, Lagudah ES, James RA, Platten JD, Dennis ES, Munns R (2006) A Sodium Transporter (HKT7) Is a Candidate for Nax1, a Gene for Salt Tolerance in Durum Wheat. Plant Physiology 142:1718-1727

Huertas R, Olias R, Eljakaoui Z, Galvez FJ, Li J, De Morales PA, Belver A, Rodriguez-Rosales MP (2012) Overexpression of SlSOS2 (SlCIPK24) confers salt tolerance to transgenic tomato. Plant, cell & environment 35:1467-1482

James RA, Blake C, Byrt CS, Munns R (2011) Major genes for Na^+ exclusion, Nax1 and Nax2 (wheat HKT1;4 and HKT1;5), decrease Na^+ accumulation in bread wheat leaves under saline and waterlogged conditions. Journal of experimental botany 62:2939-2947

James RA, Blake C, Zwart AB, Hare RA, Rathjen AJ, Munns R (2012) Impact of ancestral wheat sodium exclusion genes Nax1 and Nax2 on grain yield of durum wheat on saline soils. Functional Plant Biology 39:609-618

James RA, Davenport RJ, Munns R (2006) Physiological Characterization of Two Genes for Na^+ Exclusion in Durum Wheat, Nax1 and Nax2. Plant Physiology 142:1537-1547

Jarvis DE, Ryu CH, Beilstein MA, Schumaker KS (2014) Distinct roles for SOS1 in the convergent evolution of salt tolerance in *Eutrema salsugineum* and *Schrenkiella parvula*. Molecular biology and evolution 31:2094-2107

Jha D, Shirley N, Tester M, Roy SJ (2010) Variation in salinity tolerance and shoot sodium accumulation in *Arabidopsis* ecotypes linked to differences in the natural expression levels of transporters involved in sodium transport. Plant, cell & environment 33:793-804

Jiang C, Belfield EJ, Mithani A, Visscher A, Ragoussis J, Mott R, Smith JA, Harberd NP (2012) ROS-mediated vascular homeostatic control of root-to-shoot soil Na delivery in *Arabidopsis*. The EMBO journal 31:4359-4370

Jiang Y, Yang B, Deyholos M (2009) Functional characterization of the *Arabidopsis* bHLH92 transcription factor in abiotic stress. Molecular Genetics and Genomics 282:503-516

Julkowska MM (2015) Tuned to survive: Salt stress induced changes in Arabidopsis.

Kader MA, Seidel T, Golldack D, Lindberg S (2006) Expressions of OsHKT1, OsHKT2, and OsVHA are differentially regulated under NaCl stress in salt-sensitive and salt-tolerant rice (*Oryza sativa* L.) cultivars. Journal of experimental botany 57:4257-4268

Kasuga M, Liu Q, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1999) Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nature Biotechnology 17:287-291

Kato Y, Sakaguchi M, Mori Y, Saito K, Nakamura T, Bakker EP, Sato Y, Goshima S, Uozumi N (2001) Evidence in support of a four transmembrane-pore-transmembrane topology model for the *Arabidopsis thaliana* Na^+/K^+ translocating AtHKT1 protein, a member of the superfamily of K^+ transporters. Proceedings of the National Academy of Sciences of the United States of America 98:6488-6493

Katoh K, Misawa K, Kuma k-i, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic acids research 30:3059-3066

Katschnig D, Bliek T, Rozema J, Schat H (2015) Constitutive high-level SOS1 expression and absence of HKT1;1 expression in the salt-accumulating halophyte *Salicornia dolichostachya*. Plant science : an international journal of experimental plant biology 234:144-154

Lin HX, Zhu MZ, Yano M, Gao JP, Liang ZW, Su WA, Hu XH, Ren ZH, Chao DY (2004) QTLs for Na^+ and K^+ uptake of the shoots and roots controlling rice salt tolerance. Theoretical and applied genetics 108:253-260

Livak KJ, Schmittgen TD (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. Methods 25:402-408

Maathuis FJ (2014) Sodium in plants: perception, signalling, and regulation of sodium fluxes. Journal of experimental botany 65:849-858

Maischak H, Zimmermann MR, Felle HH, Boland W, Mithöfer A (2010) Alamethicin-induced electrical long distance signaling in plants. Plant Signaling & Behavior 5:988-990

Mäser P, Eckelman B, Vaidyanathan R, Horie T, Fairbairn DJ, Kubo M, Yamagami M, Yamaguchi K, Nishimura M, Uozumi N, Robertson W, Sussman MR, Schroeder JI (2002a) Altered shoot/root Na^+ distribution and bifurcating salt sensitivity in *Arabidopsis* by genetic disruption of the Na^+ transporter AtHKT1. FEBS letters 531:157-161

Mäser P, Hosoo Y, Goshima S, Horie T, Eckelman B, Yamada K, Yoshida K, Bakker EP, Shinmyo A, Oiki S, Schroeder JI, Uozumi N (2002b) Glycine residues in potassium channel-like selectivity filters determine potassium selectivity in four-loop-per-subunit HKT transporters from plants. Proceedings of the National Academy of Sciences of the United States of America 99:6428-6433

Mason MG, Jha D, Salt DE, Tester M, Hill K, Kieber JJ, Eric Schaller G (2010) Type-B response regulators ARR1 and ARR12 regulate expression of AtHKT1;1 and accumulation of sodium in *Arabidopsis* shoots. The Plant Journal 64:753-763

Mittler R, Vanderauwera S, Suzuki N, Miller G, Tognetti VB, Vandepoele K, Gollery M, Shulaev V, Van Breusegem F (2011) ROS signaling: the new wave? Trends in plant science 16:300-309

Moller IS, Gilliham M, Jha D, Mayo GM, Roy SJ, Coates JC, Haseloff J, Tester M (2009) Shoot Na^+ exclusion and increased salinity tolerance engineered by cell type-specific alteration of Na^+ transport in *Arabidopsis*. The Plant cell 21:2163-2178

Mondini L, Nachit M, Porceddu E, Pagnotta MA (2012) Identification of SNP mutations in DREB1, HKT1, and WRKY1 genes involved in drought and salt stress tolerance in durum wheat (*Triticum turgidum* L. var durum). Omics : a journal of integrative biology 16:178-187

Munns R, James RA, Xu B, Athman A, Conn SJ, Jordans C, Byrt CS, Hare RA, Tyerman SD, Tester M, Plett D, Gilliham M (2012) Wheat grain yield on saline soils is improved by an ancestral Na^+ transporter gene. Nature Biotechnology 30:360-364

Munns R, Tester M (2008) Mechanisms of salinity tolerance. Annual review of plant biology 59:651-681

Nguyen VL, Ribot SA, Dolstra O, Niks RE, Visser RGF, van der Linden CG (2012) Identification of quantitative trait loci for ion homeostasis and salt tolerance in barley (*Hordeum vulgare* L.). Molecular Breeding 31:137-152

Oomen RJ, Benito B, Sentenac H, Rodriguez-Navarro A, Talon M, Very AA, Domingo C (2012) HKT2;2/1, a K^+-permeable transporter identified in a salt-tolerant rice cultivar through surveys of natural genetic polymorphism. The Plant journal : for cell and molecular biology 71:750-762

Pasapula V, Shen G, Kuppu S, Paez-Valencia J, Mendoza M, Hou P, Chen J, Qiu X, Zhu L, Zhang X, Auld D, Blumwald E, Zhang H, Gaxiola R, Payton P (2011) Expression of an *Arabidopsis* vacuolar H^+-pyrophosphatase gene (AVP1) in cotton improves drought- and salt tolerance and increases fibre yield in the field conditions. Plant biotechnology journal 9:88-99

Platten JD, Cotsaftis O, Berthomieu P, Bohnert H, Davenport RJ, Fairbairn DJ, Horie T, Leigh RA, Lin HX, Luan S, Maser P, Pantoja O, Rodriguez-Navarro A, Schachtman DP, Schroeder JI, Sentenac H, Uozumi N, Very AA, Zhu JK, Dennis ES, Tester M (2006) Nomenclature for HKT transporters, key determinants of plant salinity tolerance. Trends in plant science 11:372-374

Platten JD, Egdane J, Ismail A (2013) Salinity tolerance, Na^+ exclusion and allele mining of HKT1;5 in *Oryza sativa* and *O. glaberrima*: many sources, many genes, one mechanism? BMC Plant Biology 13:32

Porcel R, Aroca R, Azcon R, Ruiz-Lozano JM (2006) PIP aquaporin gene expression in arbuscular mycorrhizal *Glycine max* and *Lactuca sativa* plants in relation to drought stress tolerance. Plant molecular biology 60:389-404

Qiu Q-S, Guo Y, Dietrich MA, Schumaker KS, Zhu J-K (2002) Regulation of SOS1, a plasma membrane Na^+/H^+ exchanger in Arabidopsis thaliana, by SOS2 and SOS3. Proceedings of the National Academy of Sciences 99:8436-8441

Ren Z-H, Gao J-P, Li L-G, Cai X-L, Huang W, Chao D-Y, Zhu M-Z, Wang Z-Y, Luan S, Lin H-X (2005) A rice quantitative trait locus for salt tolerance encodes a sodium transporter 37:1141-1146

Rodr ́guez-Navarro A, Rubio F (2006) High-affinity potassium and sodium transport systems in plants. Journal of experimental botany 57:1149-1160

Rodr ́guez-Rosales MP, G ́alvez FJ, Huertas R, Aranda MN, Baghour M, Cagnac O, Venema K (2009) Plant NHX cation/proton antiporters. Plant Signaling & Behavior 4:265-276

Roy SJ, Negrao S, Tester M (2014) Salt resistant crop plants. Current opinion in biotechnology 26:115-124

Rus A, Baxter I, Muthukumar B, Gustin J, Lahner B, Yakubova E, Salt DE (2006) Natural variants of AtHKT1 enhance Na^+ accumulation in two wild populations of *Arabidopsis*. PLoS genetics 2:e210

Rus A, Lee BH, Munoz-Mayor A, Sharkhuu A, Miura K, Zhu JK, Bressan RA, Hasegawa PM (2004) AtHKT1 facilitates Na^+ homeostasis and K^+ nutrition in planta. Plant Physiol 136:2500-2511

Sanadhya P, Agarwal P, Khedia J, Agarwal PK (2015) A Low-Affinity K^+ Transporter AlHKT2;1 from Recretohalophyte *Aeluropus lagopoides* Confers Salt Tolerance in Yeast. Molecular biotechnology 57:489-498

Shkolnik-Inbar D, Adler G, Bar-Zvi D (2013) ABI4 downregulates expression of the sodium transporter HKT1;1 in *Arabidopsis* roots and affects salt tolerance. The Plant journal : for cell and molecular biology 73:993-1005

Su H, Balderas E, Vera-Estrella R, Golldack D, Quigley F, Zhao C, Pantoja O, Bohnert H (2003) Expression of the cation transporter McHKT1 in a halophyte. Plant molecular biology 52:967-980

Sunarpi, Horie T, Motoda J, Kubo M, Yang H, Yoda K, Horie R, Chan WY, Leung HY, Hattori K, Konomi M, Osumi M, Yamagami M, Schroeder JI, Uozumi N (2005) Enhanced salt tolerance mediated by AtHKT1 transporter-induced Na unloading from xylem vessels to xylem parenchyma cells. The Plant journal : for cell and molecular biology 44:928-938

Suzuki N, Koussevitzky S, Mittler R, Miller G (2012) ROS and redox signalling in the response of plants to abiotic stress. Plant, cell & environment 35:259-270

Takai D, Jones PA (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proceedings of the National Academy of Sciences 99:3740-3745

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular biology and evolution

Tran L-SP, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Fujita M, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2004) Isolation and Functional Analysis of Arabidopsis Stress-Inducible NAC Transcription Factors That Bind to a Drought-Responsive cis-Element in the early responsive to dehydration stress 1 Promoter. The Plant cell 16:2481-2498

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic acids research 35:W71-74

Uozumi N, Kim EJ, Rubio F, Yamaguchi T, Muto S, Tsuboi A, Bakker EP, Nakamura T, Schroeder JI (2000) The Arabidopsis HKT1 gene homolog mediates inward Na^+ currents in Xenopus laevis oocytes and Na^+ uptake in *Saccharomyces cerevisiae*. Plant Physiology 122:1249-1259

Vendruscolo ECG, Schuster I, Pileggi M, Scapim CA, Molinari HBC, Marur CJ, Vieira LGE (2007) Stress-induced synthesis of proline confers tolerance to water deficit in transgenic wheat. Journal of plant physiology 164:1367-1376

Wei Z, Julkowska MM, Laloë J-O, Hartman Y, de Boer G-J, Michelmore RW, van Tienderen PH, Testerink C, Schranz ME (2014) A mixed-model QTL analysis for salt tolerance in seedlings of crop-wild hybrids of lettuce. Molecular Breeding 34:1389-1400

Xue S, Yao X, Luo W, Jha D, Tester M, Horie T, Schroeder JI (2011) AtHKT1;1 mediates nernstian sodium channel transport properties in *Arabidopsis* root stelar cells. PloS one 6:e24725

Yamaguchi T, Hamamoto S, Uozumi N (2013) Sodium transport system in plant cells. Frontiers in plant science 4:410

Yang O, Popova OV, Süthoff U, Lüking I, Dietz K-J, Golldack D (2009) The Arabidopsis basic leucine zipper transcription factor AtbZIP24 regulates complex transcriptional networks involved in abiotic stress resistance. Gene 436:45-55

Zhang H, Kim M-S, Sun Y, Dowd SE, Shi H, Paré PW (2008) Soil Bacteria Confer Plant Salt Tolerance by Tissue-Specific Regulation of the Sodium Transporter HKT1. Molecular Plant-Microbe Interactions 21:737-744

Zhang L-Y, Zhang Y-Y, Chen R-G, Zhang J-H, Wang T-T, Li H-X, Ye Z-B (2009) Ectopic Expression of the Tomato Mi-1 Gene Confers Resistance to Root Knot Nematodes in Lettuce (*Lactuca sativa*). Plant Molecular Biology Reporter 28:204-211

Supplementary figures and tables

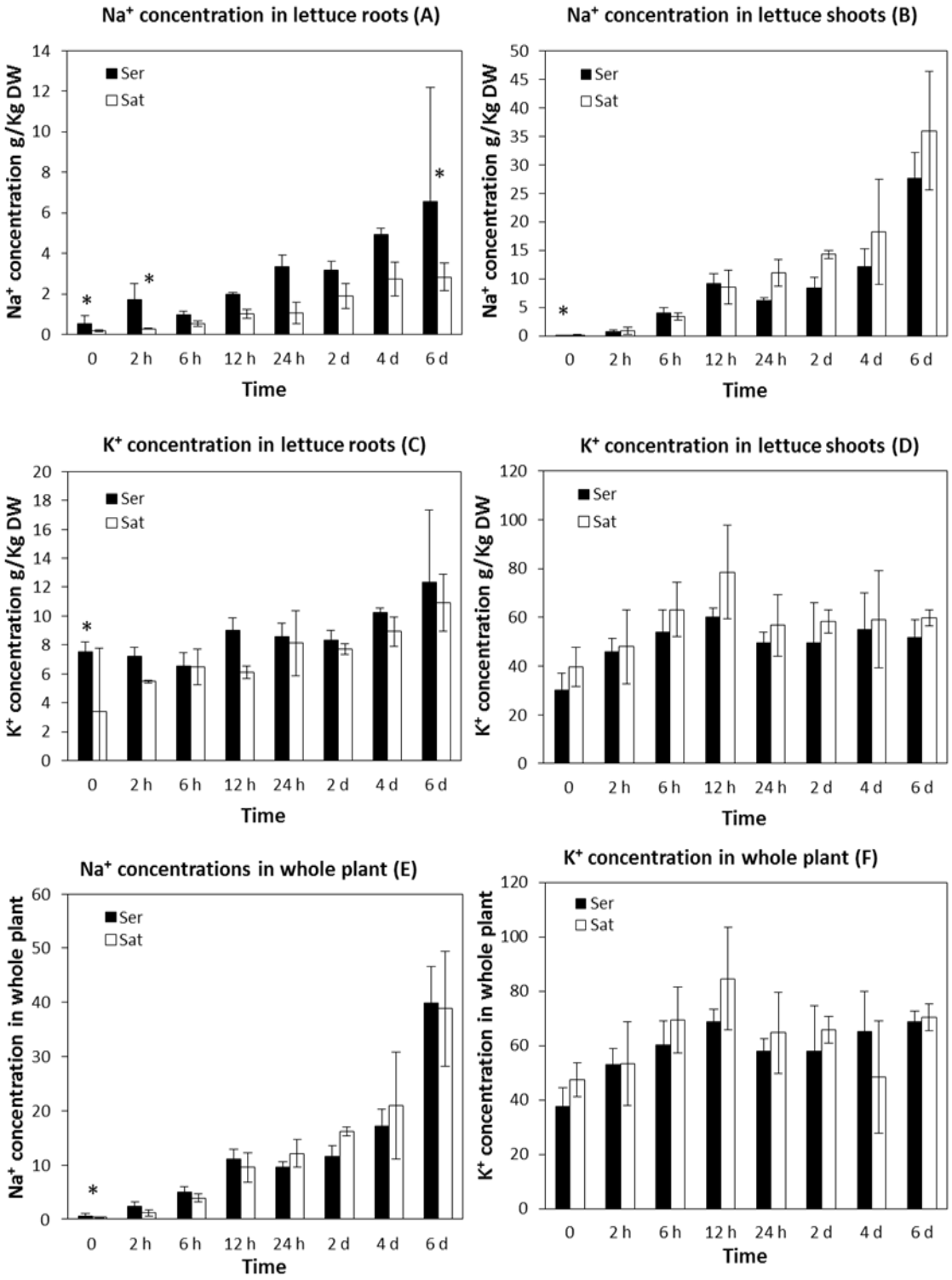


Figure S1 Sodium and potassium concentrations in lettuce roots and shoots from time 0 (control condition) to 6 days (75 mM NaCl). (A) Na^+ concentration in lettuce roots. (B) Na^+ concentration in lettuce shoots. (C) K^+ concentration in lettuce roots. (D) K^+ concentration in lettuce shoots. (E) Na^+ concentration in whole plant. (F) K^+ concentration in whole plant. Ser is *L. serriola* and Sat is *L. sativa* 'Salinas'. Error bars indicate the Standard Error from three

biological repeats. Student's t-test was used to test the differences between the two genotypes using IBM SPSS Statistics 22.0.

Table S1 Information of *HKT* transporters used in the phylogenetic analysis

Protein	Species	Accession number
AlHKT1	Arabidopsis lyrata	489874
AtHKT1;1	Arabidopsis thaliana	NP_567354
BrHKT1	Brassica rapa FPsc	Brara.B02495
BsHKT1	Boechera stricta	Bostr.25463s0287
CpHKT1	Carica papaya	evm.TU.supercontig_115.37
EcHKT1;1	Eucalyptus camaldulensis	AAF97728
EcHKT1;2	Eucalyptus camaldulensis	AAD53890
EsHKT1	Eutrema salsugineum	AFJ23835.1
GmHKT1	Glycine max	XP_003540998.1
HbHKT2	Hordeum brevisubulatum	AER42622.1
HvHKT1;5	Hordeum vulgare	ABK58096.1
HvHKT2;1	Hordeum vulgare	AEM55590.1
HvHKT4	Hordeum vulgare	AEM44690.1
McHKT1;1	Mesembryanthemum crystallinum	AF367366_1
McHKT1;2	Mesembryanthemum crystallinum	AAO73474.1
MtHKT1;5	Medicago truncatula	AES77170.1
OsHKT1;1	Oryza sativa	Q7XPF8.2
OsHKT1;3	Oryza sativa	Q6H501.1
OsHKT1;5	Oryza sativa	A2WNZ9.2
OsHKT2;1	Oryza sativa	A2YGP9.2
OsHKT2;2	Oryza sativa	Q93XI5.1
OsHKT2;3	Oryza sativa	Q8L481.1
OsHKT2;4	Oryza sativa	Q8L4K5.1
PtHKT1	Populus trichocarpa	EEF03794.1
PutHKT2;1	Puccinellia tenuiflora	ACT21087.1
SbHKT1	Salicornia bigelovii	ADG45565.1
SbiHKT1;3	Sorghum bicolor	EES04614.1
SbiHKT1;5	Sorghum bicolor	EES02856.1
SbiHKT2;3	Sorghum bicolor	EER90327.1
SlHKT1;1	Solanum lycopersicum	Solyc07g014690.2.1
SlHKT1;2	Solanum lycopersicum	Solyc07g014680.2.1
SsHKT1	Suaeda salsa	AAS20529.2
TaHKT1;5-B1	Triticum aestivum	ABG33947.1
TaHKT1;5-B2	Triticum aestivum	ABG33948.1
TaHKT1;5-D	Triticum aestivum	ABG33949.1
TaHKT2;1	Triticum aestivum	AAA52749
ThHKT1	Thellungiella halophila	BAJ34563.1
TmHKT1;5-A	Triticum monococcum	ABG33946.1
TsHKT1;2	Thellungiella salsuginea	BAJ34563
VvHKT1;1	Vitis vinifera	CAO64083
VvHKT1;2	Vitis vinifera	CAO64075

ZmHKT1	Zea mays	AEK27028.1

Table S2 A total of 386 putative cis-acting elements were found in the promoter regions of *LsaHKT1;1* (for two strands). This table can be provided if needed.

Table S3 A total of 466 putative cis-acting elements were found in the promoter regions of *LseHKT1;1* (for two strands). This table can be provided if needed.

Chapter 6

General discussion

Zhen Wei

Biosystematics Group, Wageningen University and Research Centre, 6708 PB Wageningen, The Netherlands

Cultivated lettuce belongs to the genus *Lactuca*, which has been a problematic genus in terms of its taxonomic circumscription and phylogenetic affinities for more than a century. The morphological characters of *Lactuca* species are complex and diverse, adding up the difficulty of studying phylogenetic relationships within *Lactuca*. In this thesis, I constructed the latest molecular phylogeny within *Lactuca*, based on chloroplast (single and genome) and ribosomal nuclear DNA sequences (Chapter 2 and 3). I found there are at least four phylogenetic groups within *Lactuca*, which have not been identified before. The endemic African species used in this study turned out to be more closely related to other genera and therefore should be transferred from *Lactuca* and treated as a new genus in the future (Chapter 2 and 3). Although the taxon sampling of this study covered all the important geographic groups in *Lactuca*, we only sampled one-third of all *Lactuca* species. The taxon sampling should be improved in future research.

I also undertook genetic and molecular breeding studies of cultivated lettuce. Specifically, I used a recombinant inbred line population derived from the domesticated lettuce (*L. sativa* 'Salinas') crossed to the wild species (*L. serriola* L.) to perform QTL analysis related to salt stress in lettuce seedlings. Three major QTL regions associated with responses to salinity stress in lettuce root system architecture (RSA) and leaf were discovered (Chapter 4). One *HKT*-like protein coding gene was found near the maximum LOD value of one major QTL (qLS7.2) related to sodium accumulation in lettuce leaves (Chapter 5). Structural analysis demonstrated the Na^+ selectivity of the *Lactuca HKT1*s, and expression pattern of *Lactuca HKT1*s showed they were mainly expressed in shoot and the expression changed during different time courses. Based on the conclusions of this thesis, I will start the general discussion with the implication for phylogenetics of *Lactuca*.

Implications for *Lactuca* phylogenetics and taxonomy

Key characters for the diagnosis of species within *Lactuca*

Killian (2001) considered there were 14 features important for identifying the taxonomic positions of species within the Lactucinae. He especially emphasized three features crucial for diagnosing species within *Lactuca*: (1) the presence or absence of an outer row of minute pappus hairs; (2) the presence or absence of a beak; and (3) the number of flowers per capitulum (Kilian 2001). These three important characters were also stressed by other researchers (Lebeda et al. 2007; Shih and Kilian 2011). However, in my thesis, I found that the outer row of minute pappus hairs and the presence or absence of a beak did not show clear patterns on the phylogenetic trees based on two chloroplast genes (Chapter 2). In contrast, floret number per capitulum, the presence or absence of broadly winged achene, chromosomal number, and geographic distribution are essential characters for circumscription of *Lactuca* species (Chapter 2).

Next Generation Sequencing (NGS) technology makes herbarium DNA feasible for molecular phylogenetic studies

Sequencing technology has been developing rapidly during the past four decades (Morozova and Marra 2008; Sanger and Coulson 1975; van Dijk et al. 2014). Next Generation Sequencing technology has been proven to be a useful tool for constructing molecular phylogenies of different taxonomic levels (Barrett et al. 2013; Hackett et al. 2008; Huang et al. 2014; Jansen et al. 2007; Ma et al. 2014; Moore et al. 2007; Moore et al. 2010; Nikiforova et al. 2013). Fresh plant tissues are ideal to obtain high quality DNA needed for high-throughput sequencing technology. When fresh plant materials are not available, then herbarium tissue could be an alternative for phylogenetic studies by NGS (Bakker 2015; Bakker et al. 2015). Although DNA isolated from herbarium specimens is usually degraded into small DNA fragments as a result of drying and preservation methods (Besse and Drábková 2014; Staats et al. 2011), the length of herbarium DNA is still suitable for NGS (Staats et al. 2013). Chloroplast genome DNA sequences have been successfully assembled from several historical plant herbarium samples using the Illumina HiSeq 2000 platform (Bakker et al. 2015; Staats et al. 2013). In Chapter 2, I also assembled 30 *Lactuca* chloroplast genome sequences and 29 nuclear (ribosomal) DNA sequences (ITS1+5.8S+ITS2). Twenty-four of the chloroplast genomes include one complete large single copy region, one complete small single copy region and one complete inverted repeat region. Therefore, herbarium tissues can provide enormous resources for future phylogenetic research using NGS.

Perspectives on *Lactuca* phylogeny

Asia and Africa have been considered as the two centres of diversity of wild *Lactuca* species and contain the highest number of species, 51 and 43, respectively (Lebeda et al. 2004). Surprisingly, most of the endemic African species used in my study were more closely related to non- *Lactuca* and likely should be treated as a new monophyletic genus. This result implies a possible consequence that the circumscription of *Lactuca* could be narrowed down by including more wild *Lactuca* species that are native to the African continent. If this consequence is proven to be true, then the historical viewpoint about *Lactuca*, that it has Asia and Africa as the two most diverse centres, will be changed. The elimination of the African diversity centre of *Lactuca* will make the origin of *Lactuca* (Asia) more clear.

Genome-wide association mapping for lettuce breeding

The development and application of molecular markers in breeding and the current status of lettuce molecular breeding have been discussed previously (Chapter 1). Many QTL mapping studies have been performed in lettuce, such as QTLs for shelf-life (Zhang et al. 2007), for RSA and deep soil water exploitation (Johnson et al. 2000), for seed and seedling traits related to germination (Argyris et al. 2005), for domestication traits (Hartman et al. 2013), for salt responses in RSA and leaf (Wei et al. 2014), and for resistance to pathogens and pests

(den Boer et al. 2014; Jeuken and Lindhout 2002; Jeuken et al. 2008; Simko et al. 2015; Zhang et al. 2009). However, GWAS mapping has not yet been applied to lettuce breeding.

More and more whole genome sequences of plants have been finished and released as a result of the fast developing NGS platforms, which makes GWAS mapping possible to be used in breeding (Huang and Han 2014). GWAS mapping can find all genomic regions involved in controlling complex traits of interest (Gupta et al. 2014), using single nucleotide polymorphisms (SNPs) markers (Zhao et al. 2011), simple-sequence repeat (SSR) markers (Nambeesan et al. 2015), Diversity Arrays Technology (DArT) markers (Bordes et al. 2014) or other markers (Gupta et al. 2013). The candidate genes revealed by GWAS mapping can then be validated through T-DNA mutants or genetic transformation and used for genetic modification or marker-assisted selection to develop novel varieties (Huang and Han 2014).

The genome of cultivated lettuce (*L. sativa*) and wild lettuce (*L. serriola*) have been sequenced (Truco et al. 2014), which provides the possibility of GWAS mapping (Zhao et al. 2011) for lettuce in the near future. The project of International *Lactuca* Genomics Consortium (ILGC), funded by TKI-TopSector, is aiming to re-sequence 2 new reference genomes for *L. saligna* L. and *L. virosa* L. and to assess the allelic diversity by exome sequencing. Once these *Lactuca* whole genome sequences are finished, the genetic diversity of complex traits related to biotic or abiotic stresses can be investigated in different lettuce cultivar accessions and develop new advanced crops.

Integration of phylogenetics with genomics: identifying allelic differences in lettuce at the species and/or population level

In Chapter 4, I used a QTL mapping approach to detect genetic regions associated with salt induced changes in lettuce RSA and leaf. Then I investigated the allelic differences of candidate gene (*HKT1*) from the two parental lines of the population used for QTL mapping in Chapter 5. The domesticated lettuce showed a distinctive pattern in expression level from the wild species during a time course. This raises an interesting question about whether there will be differential expression between another *Lactuca HKT1* allele, not in the QTL region on chromosome 7 but on chromosome 4 (mentioned in Chapter 5), and the two alleles studied in Chapter 5.

Multiple copies of *HKT1*s in plants have been reported not just in monocots, (Ben Amar et al. 2014; Huang et al. 2008; James et al. 2011; James et al. 2012; James et al. 2006; Platten et al. 2013), but also in dicots, (Ali et al. 2012; Asins et al. 2013). These different *HKT1* copies show discrepancy in expression, structure and/or Na^+ affinity at species level (Almeida et al. 2014a) and allelic differences between populations (Negrao et al. 2013). The research of tomato *HKT1*s can be a good example. Two closely linked *HKT1* coding genes were found in the major tomato QTL involved in Na^+/K^+ homeostasis and the complex expression pattern for the *HKT1;1* and *HKT1;2* alleles might come from the differences in their promoter

sequences (Asins et al. 2013). The tomato *HKT1;2* genes showed differences in Na^+ transport behaviour and affinity between two tomato species (Almeida et al. 2014a). However, the first pore domain of the *HKT1;2* was found to be conserved among 93 different tomato accessions (Almeida et al. 2014b).

It will be very interesting to study *HKT1* variation within and between different lettuce species and populations. Based on the phylogenetic relationships within *Lactuca* (Chapter 2 and 3), some species are close to the domesticated lettuce and could be screened for the differential *HKT1* expression: *L. aculeata* Boiss., *L. serriola*, *L saligna*, *L. virosa*, *L. quercina* L., *L. orientalis* Boiss., *L. viminea* J.Presl & C.Presl, *L. viminea* subsp. *chondrilliflora* (Boreau) Malag., *L. viminea* subsp. *ramosissima* (All.) Malag., *L. indica* L., *L. raddeana* Maximowicz, *L. formosana* Maximowicz, *L. tatarica* (L.) C.A. Meyer and *L. sibirica* Benth. ex Maxim. Novel *HKT1* alleles have been proved to improve the salt tolerance in wheat (James et al. 2012; James et al. 2006; Munns et al. 2012). Thus wild *Lactuca* species containing *HKT1* alleles with high expression and Na^+ affinity can be considered as potential genetic resources to improve the salt tolerance in cultivated lettuce. In addition, the whole genome sequences of domesticated lettuce make GWAS study of salinity stress feasible for lettuce accessions in the future. More candidate genes related to mechanisms of salt tolerance in lettuce, such as vacuolar Na^+/H^+ antiporters (NHX) and salt overly sensitive (SOS) pathway (Barragán et al. 2012; Huertas et al. 2012; Katschnig et al. 2015; Rodríguez-Rosales et al. 2009), might be revealed in future study.

Not just salinity stress

During the transition from wild to domesticated lettuce, beneficial characters like good hearting, decreased latex content, loss of spines, increased head size and bolting resistance were selected (Ryder and Whitaker 1995). The implication of my phylogenetic studies within *Lactuca* (this thesis) does not limit future work just to salt stress in lettuce, but also provide insights into novel (potential) genetic resources for research on other beneficial traits in lettuce, e.g. RSA, leaf area, flowering time, disease resistance, yield and oil proportion of seeds (for oilseed lettuce), using QTL or GWAS mapping approach. Some wild *Lactuca* species, e.g. *L. serriola*, *L. saligna*, *L. virosa*, *L. tatarica*, *L. viminea*, *L. biennis* (Moench) Fernald, *L. canadensis* L., *L. homblei* De Wild, *L. indica*, *L. perennis* L. and *L. tenerrima* Pourr. etc., have been screened for the resistance to downy mildew isolates (*Bremia lactucae* Regel) (Globerson 1980; Lebeda and Boukema 1991; Lebeda et al. 2002; Lebeda and Reinink 1994; Lebeda and Zinkernagel 2003; Ryder and Whitaker 1995; van Treuren et al. 2011; Zohary 1983). Moreover, lettuce cultivars and some wild *Lactuca* species have also been evaluated for the resistance to insect pests (e.g. leaf miners, *Liriomyza langeri* Frick), bacterium (e.g. corky root, *Sphingomonas suberifaciens*) and virus (e.g. lettuce mosaic virus) (Beiquan Mou and Liu 2004; Beiquan Mou 2007; Mou and Bull 2004; Mou and Liu 2003). The *Lactuca* species closely related to the lettuce cultivars (discussed in Chapter 2 and 3) and with resistance to pathogens and insect pests, can be considered as useful genetic resources

for disease resistance in lettuce breeding. In addition, the phylogenetic relationships of *HKT1*s can also shed some light on the evolution of *Lactuca* species. For instance, do the two loci of *HKT1* of the domesticated lettuce indicate a gene duplication event? If so, what about the *HKT1*s in other wild *Lactuca* species? How many duplication events of *HKT1*s happened during the domestication of lettuce crop or in the evolution of *Lactuca* speciation? These interesting questions can be studied using the new genome sequences of *L. sativa* 'Salinas' and *L. serriola*.

References

Ali Z, Park HC, Ali A, Oh DH, Aman R, Kropornicka A, Hong H, Choi W, Chung WS, Kim WY, Bressan RA, Bohnert HJ, Lee SY, Yun DJ (2012) TsHKT1;2, a HKT1 homolog from the extremophile Arabidopsis relative *Thellungiella salsuginea*, shows K^+ specificity in the presence of NaCl. Plant Physiol 158:1463-1474

Almeida P, de Boer GJ, de Boer AH (2014a) Differences in shoot Na^+ accumulation between two tomato species are due to differences in ion affinity of HKT1;2. Journal of plant physiology 171:438-447

Almeida PM, de Boer GJ, de Boer AH (2014b) Assessment of natural variation in the first pore domain of the tomato HKT1;2 transporter and characterization of mutated versions of SlHKT1;2 expressed in *Xenopus laevis* oocytes and via complementation of the salt sensitive athkt1;1 mutant. Frontiers in plant science 5:600

Argyris J, Truco M, Ochoa O, Knapp S, Still D, Lenssen G, Schut J, Michelmore R, Bradford K (2005) Quantitative trait loci associated with seed and seedling traits in *Lactuca*. Theoretical and Applied Genetics 111:1365-1376

Asins MJ, Villalta I, Aly MM, Olias R, Alvarez DEMP, Huertas R, Li J, Jaime-Perez N, Haro R, Raga V, Carbonell EA, Belver A (2013) Two closely linked tomato HKT coding genes are positional candidates for the major tomato QTL involved in Na^+ /K^+ homeostasis. Plant, cell & environment 36:1171-1191

Bakker FT (2015) DNA sequences from plant herbarium tissue. In: Hörandl E, Appelhans MS (eds) Next-Generation Sequencing in Plant Systematics. International Association for Plant Taxonomy (IAPT)

Bakker FT, Lei D, Yu J, Mohammadin S, Wei Z, van de Kerke S, Gravendeel B, Nieuwenhuis M, Staats M, Alquezar-Planas DE, Holmer R (2015) Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline. Biol J Linn Soc doi:10.1111/bij.12642

Barragán V, Leidi EO, Andrés Z, Rubio L, De Luca A, Fernández JA, Cubero B, Pardo JM (2012) Ion Exchangers NHX1 and NHX2 Mediate Active Potassium Uptake into Vacuoles to Regulate Cell Turgor and Stomatal Function in *Arabidopsis*. The Plant Cell 24:1127-1142

Barrett CF, Davis JI, Leebens-Mack J, Conran JG, Stevenson DW (2013) Plastid genomes and deep relationships among the commelinid monocot angiosperms. Cladistics 29:65–87

Beiquan Mou, Liu Y-B (2004) Host Plant Resistance to Leafminers in lettuce. J Amer Soc Hort Sci 129:383-388

Beiquan Mou RJH, and Edward J. Ryde (2007) Crisphead Lettuce Breeding Lines with Resistance to Corky Root and Lettuce Mosaic Virus. Hort Sci 42:701-703

Ben Amar S, Brini F, Sentenac H, Masmoudi K, Very AA (2014) Functional characterization in Xenopus oocytes of Na^+ transport systems from durum wheat reveals diversity among two HKT1;4 transporters. Journal of experimental botany 65:213-222

Besse P, Dr âbkov á L (2014) DNA Extraction from Herbarium Specimens. Molecular Plant Taxonomy. Humana Press, pp 69-84

Bordes J, Goudemand E, Duchalais L, Chevarin L, Oury F, Heumez E, Lapierre A, Perretant M, Rolland B, Beghin D, Laurent V, Le Gouis J, Storlie E, Robert O, Charmet G (2014) Genome-wide association mapping of three important traits using bread wheat elite breeding populations. Molecular Breeding 33:755-768

den Boer E, Pelgrom KB, Zhang N, Visser RF, Niks R, Jeuken MW (2014) Effects of stacked quantitative resistances to downy mildew in lettuce do not simply add up. Theoretical and Applied Genetics 127:1805-1816

Globerson D (1980) Wild lettuce as a source for improving cultivated lettuce. In: Maxon Smith JW (ed) Proceedings, Eucarpia meeting on leafy vegetables, Littlehampton, pp 86-96

Gupta PK, Kulwal PL, Jaiswal V, Theodore Friedmann JCD, Stephen FG (2014) Chapter Two - Association Mapping in Crop Plants: Opportunities and Challenges. Advances in Genetics. Academic Press, pp 109-147

Gupta PK, Varshney RK, Edwards D, Gupta P (2013) Sequence Based DNA Markers and Genotyping for Cereal Genomics and Breeding. Cereal Genomics II. Springer Netherlands, pp 57-76

Hackett SJ, Kimball RT, Reddy S, Bowie RCK, Braun EL, Braun MJ, Chojnowski JL, Cox WA, Han K-L, Harshman J, Huddleston CJ, Marks BD, Miglia KJ, Moore WS, Sheldon FH, Steadman DW, Witt CC, Yuri T (2008) A Phylogenomic Study of Birds Reveals Their Evolutionary History. Science 320:1763-1768

Hartman Y, Hooftman DP, Eric Schranz M, van Tienderen P (2013) QTL analysis reveals the genetic architecture of domestication traits in Crisphead lettuce. Genetic Resources and Crop Evolution 60:1487-1500

Huang H, Shi C, Liu Y, Mao S-Y, Gao L-Z (2014) Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC Evolutionary Biology 14:151

Huang S, Spielmeyer W, Lagudah ES, Munns R (2008) Comparative mapping of HKT genes in wheat, barley, and rice, key determinants of Na^+ transport, and salt tolerance. Journal of experimental botany 59:927-937

Huang X, Han B (2014) Natural Variations and Genome-Wide Association Studies in Crop Plants. Annual Review of Plant Biology 65:531-551

Huertas R, Olias R, Eljakaoui Z, Galvez FJ, Li J, De Morales PA, Belver A, Rodriguez-Rosales MP (2012) Overexpression of SlSOS2 (SlCIPK24) confers salt tolerance to transgenic tomato. Plant, cell & environment 35:1467-1482

James RA, Blake C, Byrt CS, Munns R (2011) Major genes for Na+ exclusion, Nax1 and Nax2 (wheat HKT1;4 and HKT1;5), decrease Na^+ accumulation in bread wheat leaves under saline and waterlogged conditions. Journal of experimental botany 62:2939-2947

James RA, Blake C, Zwart AB, Hare RA, Rathjen AJ, Munns R (2012) Impact of ancestral wheat sodium exclusion genes Nax1 and Nax2 on grain yield of durum wheat on saline soils. Functional Plant Biology 39:609-618

James RA, Davenport RJ, Munns R (2006) Physiological Characterization of Two Genes for Na^+ Exclusion in Durum Wheat, Nax1 and Nax2. Plant Physiology 142:1537-1547

Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Muller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, McNeal JR, Kuehl JV, Boore JL (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proceedings of the National Academy of Sciences of the United States of America 104:19369-19374

Jeuken M, Lindhout P (2002) Lactuca saligna, a non-host for lettuce downy mildew (*Bremia lactucae*), harbors a new race-specific Dm gene and three QTLs for resistance. Theoretical and Applied Genetics 105:384-391

Jeuken MJW, Pelgrom K, Stam P, Lindhout P (2008) Efficient QTL detection for nonhost resistance in wild lettuce: backcross inbred lines versus F_2 population. Theoretical and Applied Genetics 116:845-857

Johnson WC, Jackson LE, Ochoa O, van Wijk R, Peleman J, St. Clair DA, Michelmore RW (2000) Lettuce, a shallow-rooted crop, and Lactuca serriola, its wild progenitor, differ at QTL determining root architecture and deep soil water exploitation. Theoretical and Applied Genetics 101:1066-1073

Katschnig D, Bliek T, Rozema J, Schat H (2015) Constitutive high-level SOS1 expression and absence of HKT1;1 expression in the salt-accumulating halophyte *Salicornia dolichostachya*. Plant science : an international journal of experimental plant biology 234:144-154

Kilian N (2001) Lactuca stebbinsii (Lactuceae, Compositae), a puzzling new species from Angola. Willdenowia 31:71-78

Lebeda A, Boukema IW (1991) Further Investigation of the Specificity of Interactions Between Wild *Lactuca* spp. and *Bremia lactucae* Isolates from *Lactuca serriola*. J Phytopathology 133:57-64

Lebeda A, Doležalová I, Feráková V, Astley D (2004) Geographical Distribution of Wild *Lactuca* Species (Asteraceae, Lactuceae). The Botanical Review 70:328-356

Lebeda A, Pink D, Astley D (2002) Aspects of the interactions between wild *Lactuca* spp. and related genera and lettuce downy mildew (*Bremia lactucae*). In: SpencerPhillips P, Gisi U, Lebeda A (eds) Advances in downy mildew research. Kluwer Academic Publishers, Dordrecht, pp 85-117

Lebeda A, Reinink K (1994) Histological characterization of resistance in *Lactuca saligna* to lettuce downy mildew (*Bremia lactucae*). Physiological and Molecular Plant Pathology 44:125-139

Lebeda A, Ryder E, Grube R, Dolezalova I, Krıstkova E (2007) Lettuce (Asteraceae; Lactuca spp.) In: Singh R (ed) Genetic resources, chromosome engineering, and crop improvement series, vol 3-vegetable crops. CRC Press, Boca Raton, pp 377-472

Lebeda A, Zinkernagel V (2003) Characterization of New Highly irulent German Isolates of *Bremia lactucae* and Efficiency of Resistance in Wild *Lactuca* spp. Germplasm. J Phytopathology 151:274-282

Ma PF, Zhang YX, Zeng CX, Guo ZH, Li DZ (2014) Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (poaceae). Systematic biology 63:933-950

Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proceedings of the National Academy of Sciences of the United States of America 104:19363-19368

Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE (2010) Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proceedings of the National Academy of Sciences of the United States of America 107:4623-4628

Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. Genomics 92:255-264

Mou B, Bull C (2004) Screening Lettuce Germplasm for New Sources of Resistance to Corky Root. J Amer Soc Hort Sci 129:712-716

Mou B, Liu Y-B (2003) Leafminer Resistance in Lettuce. Hort Sci 38:570-572

Munns R, James RA, Xu B, Athman A, Conn SJ, Jordans C, Byrt CS, Hare RA, Tyerman SD, Tester M, Plett D, Gilliham M (2012) Wheat grain yield on saline soils is improved by an ancestral Na^+ transporter gene. Nature Biotechnology 30:360-364

Nambeesan S, Mandel J, Bowers J, Marek L, Ebert D, Corbi J, Rieseberg L, Knapp S, Burke J (2015) Association mapping in sunflower (*Helianthus annuus* L.) reveals independent control of apical vs. basal branching. BMC Plant Biology 15:84

Negrao S, Almadanim MC, Pires IS, Abreu IA, Maroco J, Courtois B, Gregorio GB, McNally KL, Oliveira MM (2013) New allelic variants found in key rice salt-tolerance genes: an association study. Plant biotechnology journal 11:87-100

Nikiforova SV, Cavalieri D, Velasco R, Goremykin V (2013) Phylogenetic analysis of 47 chloroplast genomes clarifies the contribution of wild species to the domesticated apple maternal line. Molecular biology and evolution 30:1751-1760

Platten JD, Egdane J, Ismail A (2013) Salinity tolerance, Na+ exclusion and allele mining of HKT1;5 in *Oryza sativa* and O. glaberrima: many sources, many genes, one mechanism? BMC Plant Biology 13:32

Rodríguez-Rosales MP, Gálvez FJ, Huertas R, Aranda MN, Baghour M, Cagnac O, Venema K (2009) Plant NHX cation/proton antiporters. Plant Signaling & Behavior 4:265-276

Ryder EJ, Whitaker TW (1995) Lettuce. In: Smartt J, Simmonds NW (eds) Evolution of crop plants. Harlow, GB: Longman London and New York

Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of Molecular Biology 94:441-448

Shih C, Kilian N (2011) Cichorieae [excl.*Tragopogon*, *Taraxacum*, *Hieracium* and *Pilosella*]. Science Press & St Louis: Missouri Botanical Garden, Beijing

Simko I, Ochoa OE, Pel MA, Tsuchida C, Font i Forcada C, Hayes RJ, Truco M-J, Antonise R, Galeano CH, Michelmore RW (2015) Resistance to Downy Mildew in Lettuce 'La Brillante' is Conferred by Dm50 Gene and Multiple QTL. Phytopathology 105:1220-1228

Staats M, Cuenca A, Richardson JE, Vrielink-van Ginkel R, Petersen G, Seberg O, Bakker FT (2011) DNA damage in plant herbarium tissue. PloS one 6:e28448

Staats M, Erkens RH, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J, Richardson JE, Bakker FT (2013) Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. PloS one 8:e69189

Truco M, Reyes-Chin-Wo S, Lavelle D, Kozik A, Froenicke L, Xu H, Michelmore RW (2014) Lettuce Genome Assembly, Structure and Annotation. International Plant and Animal Genome Conference XXII, San Diego

van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C (2014) Ten years of next-generation sequencing technology. Trends in Genetics 30:418-426

van Treuren R, van der Arend AJM, Schut JW (2011) Distribution of downy mildew (*Bremia lactucae* Regel) resistances in a genebank collection of lettuce and its wild relatives. Plant Genetic Resources 11:15-25

Wei Z, Julkowska MM, Laloë J-O, Hartman Y, de Boer G-J, Michelmore RW, van Tienderen PH, Testerink C, Schranz ME (2014) A mixed-model QTL analysis for salt tolerance in seedlings of crop-wild hybrids of lettuce. Molecular Breeding

Zhang FZ, Wagstaff C, Rae AM, Sihota AK, Keevil CW, Rothwell SD, Clarkson GJ, Michelmore RW, Truco MJ, Dixon MS, Taylor G (2007) QTLs for shelf life in lettuce co-locate with those for leaf

biophysical properties but not with those for leaf developmental traits. Journal of experimental botany 58:1433-1449

Zhang NW, Lindhout P, Niks RE, Jeuken MJW (2009) Genetic dissection of *Lactuca saligna* nonhost resistance to downy mildew at various lettuce developmental stages. Plant Pathology 58:923-932

Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa* 2:467

Zohary D (1983) Wild genetic resources of crops in Israel. Israel Journal of Botany 32:97-127

Summary

Cultivated lettuce (*Lactuca sativa* L.) is an important leafy vegetable worldwide. However, the phylogenetic relationships between domesticated lettuce and its wild relatives are still not clear. In this thesis, I focus on the phylogenetic relationships within *Lactuca* L., including an analysis of the wild *Lactuca* species that are endemic to Africa for the first time. The genetic variation of responses to salinity in a recombinant inbred line population, derived from a cross between the lettuce crop (*L. sativa* 'Salinas') and wild species (*L. serriola*), was investigated and the candidate gene in the identified QTL regions was further studied.

In Chapter 1, I introduce and discuss topics related to genetic diversity and evolution in *Lactuca*, including an overview of lettuce cultivars and uses, its hypothesized domestication history, the taxonomic position of *Lactuca*, current status of molecular breeding in lettuce and mechanisms of salinity tolerance in plants, especially the High-affinity K^+ Transporter (*HKT*) gene family.

In Chapter 2, the most extensive molecular phylogenetic analysis of *Lactuca* was constructed based on two chloroplast genes (*ndhF* and *trnL-F*), including endemic African species for the first time. This taxon sampling covers nearly 40% of the total *Lactuca* species endemic to Africa and 34% of all *Lactuca* species. DNA sequences from all the subfamilies of Asteraceae in Genbank and those generated from *Lactuca* herbarium samples were used to elucidate the monophyly of *Lactuca* and the affiliation of *Lactuca* within Asteracaeae. Based on the subfamily tree, 33 *ndhF* sequences from 30 species and 79 *trnL-F* sequences from 48 species were selected to infer phylogenetic relationships within *Lactuca* using Randomized Axelerated Maximum Likelihood (RAxML) and Bayesian Inference (BI) analyses. In addition, biogeographical, chromosomal and morphological character states were analysed based on the Bayesian tree topology. The results showed that *Lactuca* contains two distinct phylogenetic clades - the crop clade and the *Pterocypsela* clade. Other North American, Asian and widespread species either form smaller clades or mix with the *Melanoseris* species in an unresolved polytomy. The newly sampled African endemic species probably should be excluded from *Lactuca* and treated as a new genus.

In Chapter 3, twenty-seven wild *Lactuca* species and four outgroup species were sequenced using next generation sequencing (NGS) technology. The sampling covers 36% of total *Lactuca* species and all the important geographical groups in the genus. Thirty chloroplast genomes, including one complete (partial) large single copy region (LSC), one small single copy region (SSC), one inverted repeat (IR) region, and twenty-nine nuclear ribosomal DNA sequences (containing the internal transcribed spacer region) were successfully assembled and analysed. A methodology paper for which I am co-author, but is not included in this thesis, of the sequencing pipeline was published: 'Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative

Organelle Genome Assembly (IOGA) pipeline'. These NGS data helped resolve deeper nodes in the phylogeny within *Lactuca* and resolved the polytomy from Chapter 2. The results showed that there are at least four main groups within *Lactuca*: the crop group, the *Pterocypsela* group, the North American group and the group containing widely-distributed species. I also confirmed that the endemic African species should be removed and treated as a new genus.

In Chapter 4, quantitative trait loci (QTLs) related to salt-induced changes in Root System Architecture (RSA) and ion accumulation were determined using a recombinant inbred line population derived from a cross between cultivated lettuce and wild lettuce. I measured the components of RSA by replicated lettuce seedlings grown on vertical agar plates with different NaCl concentrations in a controlled growth chamber environment. I also quantified the concentration of sodium and potassium in replicates of greenhouse-grown plants watered with 100 mM NaCl. The results identified a total of fourteen QTLs using multi-trait linkage analysis, including three major QTLs associated with general root development (qRC9.1), root growth in salt stress condition (qRS2.1), and ion accumulation (qLS7.2).

In Chapter 5, one of the identified QTL regions (qLS7.2) reported in Chapter 4 was found to contain a homolog of the *HKT1* from *Arabidopsis thaliana*. I did a phylogenetic analysis of *Lactuca HKT1*-like protein sequences with other published *HKT* protein sequences and determined transmembrane and pore segments of lettuce *HKT1;1 alleles*, according to the model proposed for *AtHKT1;1*. Gene expression pattern and level of *LsaHKT1;1* (*L. sativa* 'Salinas') and *LseHKT1;1* (*L. serriola*) in root and shoot were investigated in plants growing hydroponically over a time-course. The measurements of Na^+ and K^+ contents were sampled at the same time as the samples used for gene expression test. In addition, I examined the 5' promoter regions of the two genotypes. The results showed low expression levels of both *HKT1;1* alleles in *Lactuca* root and relatively higher expression in shoot, probably due to the negative cis-regulatory elements of *HKT1* alleles found in *Lactuca* promoter regions. Significant allelic differences were found in *HKT1;1* expression in early stage (0-24 hours) shoots in and in late stage (2-6 days) roots. shoot *HKT1;1* expression/root *HKT1;1* expression was generally consistent with the ratios of Na^+/K^+ balance in the relevant tissues (shoot Na^+/K^+ divided by root Na^+/K^+).

In Chapter 6, I summarize and discuss the results from previous chapters briefly. The implications of Chapter 2 and 3 for *Lactuca* phylogenetics are discussed, including some key characters for the diagnosis of species within *Lactuca*, the use of herbarium DNA for NGS technology, and perspectives into *Lactuca* phylogeny. Future perspectives of genome-wide association mapping for lettuce breeding were also discussed. Lastly, I propose to integrate phylogenetic approaches into investigations of allelic differences in lettuce, not just associated with salinity stress but also with other stressed and beneficial characters, both within and between species.

Samenvatting

Gecultiveerde sla (*Lactuca sativa* L.) behoort tot een van de belangrijkste bladgroenten in de wereld. Desondanks is de fylogenetische relatie tussen gedomesticeerde sla en zijn wilde verwanten nog niet bekend. In mijn thesis heb ik daarom de fylogenetische verwantschap binnen het genus Lactuca L. onderzocht. Het is uniek dat ik endemische Afrikaanse Lactuca soorten aan deze fylogenie heb weten toe te voegen. Daarnaast heb ik onderzoek gedaan naar de genetische variatie die bestaat in een 'recombinant inbred line' (RIL) van Lactuca ten aanzien van de reactie op verschillende zoutconcentraties (zgn zoutrespons). Deze RIL is een kruising tussen een gecultiveerde sla (*L. Sativa* 'Salinas') en een wilde soort (*L. serriola*). De kandidaat genen verantwoordelijk voor de zoutrespons in de gevonden QTL-regio's zijn verder onderzocht.

In Hoofdstuk 1 introduceer en bediscussieer ik de onderwerpen die gerelateerd zijn aan de genetische diversiteit en evolutie van Lactuca. Hiertoe geef ik een overzicht van de sla cultivars en hoe ze gebruikt worden. Verder ga ik in op de taxonomische positie van Lactuca, en de veronderstelde domesticatie geschiedenis. Tot slot komt de huidige staat van moleculaire veredeling en mechanismen van zout tolerantie in planten, met nadruk op de 'High-affinity K^+ Transporter' (*HKT*) gen familie, aan de orde.

In Hoofdstuk 2 presenteer ik de eerste uitgebreide moleculaire fylogenie van Lactuca. De fylogenie is gebaseerd op twee chloroplast genen (*ndhF* en *trnL-F*) en bevat onder andere endemische Afrikaanse Lactuca soorten. De taxon sampling bevat ~40% van het totale aantal endemische Afrikaanse en 34% van alle Lactuca soorten. Met behulp van de DNA sequenties (verkregen van Genbank) van alle subfamilies van de Asteraceae en de DNA sequenties gegenereerd van *Lactuca* herbarium samples is de monofylie van Lactuca opgehelderd. Bovendien is daarmee de affiliatie van Lactuca in de Asteraceae aangetoond. Om de fylogenetische relatie binnen Lactuca te verduidelijken, zijn er, op basis van een subfamilie fylogenie, 33 ndhF sequenties van 30 soorten en 79 trnL-F sequenties van 48 soorten geselecteerd. De resultaten zijn geanalyseerd met behulp van Randomized Axelerated Maximum Likelihood (RAxML) en Bayesian Inference (BI) analyses. De verkregen Bayesiaanse topologie is vervolgens gebruikt om de biogeografische, chromosomale en morfologische karakterstaten te analyseren. De resultaten laten zien dat Lactuca twee duidelijke fylogenetische clades heeft: het gewas clade en de Pterocypsela clade. Andere Noord-Amerikaanse, Aziatische en globale verspreide soorten vormen kleinere clades of mengen met Melanoseris soorten in een polytomie. Ik stel voor om de opgenomen endemische Afrikaanse soorten als een nieuw zustergenus van Lactuca te behandelen.

In Hoofdstuk 3 worden 27 wilde Lactuca soorten en vier outgroup soorten gesequenced met behulp van 'next generation sequencing' (NGS) technologie. De taxonsampling bevat alle belangrijke geografische groepen van het genus en 36% van alle Lactuca soorten. Dertig

chloroplast genomen zijn succesvol geassembleerd en geanalyseerd. Hiertoe behoorden een volledige (soms gedeeltelijk) 'large single copy' (LSC) regio, een 'small single copy' (SSC) regio, een 'inverted repeat' (IR), en negentwintig nucleair ribosomale DNA sequenties (inclusief de 'internal transcribed spacers') Ik ben medeauteur van het gepubliceerde methodologische artikel over de sequentie werkwijze ('Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly (IOGA) pipeline'), maar heb besloten deze niet toe te voegen aan mijn thesis. De verkregen NGS data helpen om de diepere knopen van de fylogenie, zoals gepresenteerd in Hoofdstuk 2, op te lossen. De resultaten tonen aan dat er ten minste vier hoofdgroepen zijn in Lactuca: de gewassen, de Pterocypsela groep, de Noord-Amerikaanse groep, en de groep met de wijdverspreide soorten. Daarnaast heb ik bevestigd dat de endemische Afrikaanse groep uit Lactuca moet en als een aparte genus moet worden behandeld.

In Hoofdstuk 4 heb ik een 'recombinant inbred line' (RIL) populatie gebruikt om 'quantitative trait loci (QTLs)' te vinden die gerelateerd zijn aan zout geïnduceerde veranderingen in het wortelstelsel (Root System Architecture -RSA) en ionen accumulatie. De RIL populatie kwam voort uit een kruising tussen een wilde en gecultiveerde sla. Om de RSA componenten te meten, heb ik herhaaldelijk sla zaailingen gekweekt in verticale agar platen met verschillende zoutconcentraties (NaCl) in de gecontroleerde omgeving van een klimaatkamer. Daarnaast heb ik de concentratie van natrium en kalium ionen gekwantificeerd in replica's van planten die in de kas waren opgegroeid en gewaterd werden met 100 mM NaCl. Na een multi-trait linkage analyse vinden we veertien QTLs waaronder drie grote QTLs die geassocieerd zijn met wortel ontwikkeling (rRC9.1), wortel groei tijdens zout stress (qRS2.1) en ionen accumulatie (qLS7.2).

Hoofdstuk 5 toont de bevinding dat een van de in hoofdstuk 4 geïdentificeerde QTL regio's (qLS7.2) een homoloog van *HKT1* van Arabidopsis thaliana bevat. Ik heb een fylogenetische analyse gedaan van gepubliceerde Lactuca HKT1-like eiwit sequenties en de transmembraan en porie-segmenten van de sla *HKT1;1* allelen bepaald. Dit aan de hand van het model van *AtHKT1;1*. De gen expressie patronen en mate van *LsaHKT1;1* (*L. sativa* 'Salinas') en *LseHKT1;1* (*L. serriola*) in wortels en scheuten zijn onderzocht in planten die hydrophonisch zijn opgekweekt in een tijdsinterval. Tegelijk zijn Na^+ en K^+ gemeten. Daarnaast heb ik de 5' promotor regio van de twee genotypen onderzocht. Er was een lage expressie van allebei de *HKT1;1* allelen in Lactuca wortels en een relatief hoge expressie in de scheuten. Waarschijnlijk zijn de negatieve cis-regulators van de promotor regio's van de Lactuca *HKT1* allelen hiervoor verantwoordelijk. Allelische verschillen waren significant in de *HKT1;1* expressie in een vroeg stadium (0-24uur) scheuten en in de wortels in de late stadium (2-6 dagen). De scheut *HKT1;1* expressie/wortel *HKT1;1* expressie was over het algemeen consequent met de Na^+/K^+ ratio's in de betreffende weefsels (scheut Na^+/K^+ gedeeld door wortel Na^+/K^+).

Hoofdstuk 6 is een samenvatting van alle voorgaande hoofdstukken en bediscussieerd de resultaten. De implicaties van Hoofdstuk 2 en 3 voor de Lactuca fylogenie worden behandeld, en ter verduidelijking introduceer ik een aantal sleutelkenmerken voor het vaststellen van Lactuca soorten. Daarnaast evalueer ik het nut van het gebruik van herbarium DNA voor NGS technologie, en licht ik de toekomstige perspectieven van genoom brede associatie mappen voor sla cultivatie toe Ik stel voor om fylogenetische analyses te integreren in onderzoek naar allelische verschillen in sla, niet alleen maar voor de associatie met zout stress maar ook met andere positieve en negatieve karakters, zowel binnen als tussen soorten.

Acknowledgements

After four-year and five-month stay in the Netherlands, I finally accomplished my PhD in Wageningen UR. Before I thank everyone, I would like to thank my promoter and daily supervisor Prof. dr. ME (Eric) Schranz first. Just like what Eric always says, 'I want you to be happy', I did have a very happy PhD life. You sent me to all the PhD courses I was interested in, seminars, forums and workshops related to different topics, and even two GRC conferences. These scientific trainings broadened my horizon and gave me the chance to talk with the best scientists in the world (though I am pretty sure they did not remember me). In one word, a thousand thanks to Eric!

Then I would like to thank my dear colleagues in Biosystematics Group, especially our Amsterdam people, Erik van den Bergh (congratulations to your new baby), Setareh Mohammadin, Heo JaeYun and Johannes A. Hofberger. You all helped me a lot when I first came to The Netherlands. Thanks to all my colleagues, involved in my scientific life or not: dr. Freek T. Bakker, dr. Ronald G. van den Berg, dr. J.J. (Jan) Wieringa, dr. Lars W. Chatrou, Nynke Groendijk-Wilders, F.F.M. (Frank) Becker, Andre Maassen, dr. Roel H.M.J. Lemmens, dr. Sabrina Simon, dr. C.A. (Kitty) Vijverberg, Wilma Twigt, ZHAO Tao, Rens Holmer, Mathijs Nieuwenhuis, dr. Robin van Velzen, YU Jiaying, LEI Di, dr. T.P. (Phuong) Nguyen, Sara J. van de Kerke, Andre S.J. van Proosdij, Jennifer de Carvalho Lope, dr. Lidija Berke, dr. Laurie Grandont, E.O.D. (Deedi) Sogbohossou and C.J. (Corrie) Hanhart. Freek, thank you for critical and careful supervision in the phylogenetic part of my thesis.

I give my special thanks to dr. Christa Testerink and dr. Magdalena M. Julkowska. Thank you both for your critical suggestions and inspiration about my research. Thank you for your help and encouragement with my GRC presentation. I also feel grateful to those researchers involved in my research and provided materials for me, Prof. dr. Richard W. Michelmore, Prof. dr. David Salt, dr. M.J.W. (Marieke) Jeuken, dr. Norbert Kilian and dr. ZHU shixin.

Many thanks to those people in Amsterdam, who made my early life in Amsterdam happy: Prof. dr. Peter van Tienderen, dr. Thomas van Hengstum, dr. Melis Akman, dr. Yorike Hartman, Prof. dr. Maurice W. Sabelis (deceased) and all my best Chinese friends, LI Dan, WANG xiang, WANG zhijie, LOU zhongyu and LIU zuozhen.

Thanks to my Wageningen friends: JI xianwen, DENG yuxi, ZHANG jianhua, HUANG pingping, WANG yanli, XIAO tingting, REN guiling, CHENG xu, ZHANG yunmeng, ZHANG yanxia, WANG bo, BAI bing, ZHANG chunyue, HAN yue, HE hanzi, LIU junwei, SHI xiaoqian, SONG yin, HE jun, SONG wei, LI huchen, ZHANG yuanyuan, SUN kaile, XU xuan, ZHU feng and FENG yuan.

Many thanks to all the people I met here. You all gave me a good time.

Acknowledgements

感谢我的父母，在我远游的这段时间一直支持我，给我鼓励。在你们身体不好的时候没能陪在身边是我最大的遗憾。现在我已经完成学业，马上就能全家团聚了，我永远爱你们。

最后，感谢我的男朋友冯华。在我们相处的这七年里，我情绪低落的时候是你鼓励我，我开心的时候是你陪伴我，你包容了我的一切缺点。很幸运一路走来是你一直陪着我，不离不弃。

魏珍

2016 年 1 月于瓦赫宁根

Zhen Wei

Wageningen

January, 2016

About the author

Zhen Wei was born on 28th of June 1986, in Zhengzhou, Henan, China. She studied as an undergraduate at Zhengzhou University from 2004 to 2008. She did her graduation project, 'Phylogenetic relationships between *Notoseris* and related species', under the supervison of dr. Shixin Zhu. She obtained her Bachelor degree in biotechnology in 2008 and started her master study in the same year at Zhengzhou University. Her master project was about 'Plant communities and their relationship with composition of food items of rodents in shallow mountain area, Jiyuan, China'. In 2011, she got her master degree in botany and came to the Netherlands for PhD study. She spent her first year of PhD study at University of Amsterdam and finished her first experiment of QTL mapping related to salt stress in lettuce seedlings (Chapter 4, this thesis). Then her daily supervisor Eric M. Schranz got his professorship at Wageningen University in June 2012. She decided to transfer to Wageningen University and continue her PhD with Eric. She finished her PhD thesis, entitled 'Genetic diversity and evolution in *Lactuca* L. (Asteraceae) - from phylogeny to molecular breeding' in Biosystematics Group, Wageningen University. Results of her PhD study were presented in this thesis.

Education Statement of the Graduate School

Experimental Plant Sciences

Issued to:	**Zhen Wei**
Date:	**25 January 2016**
Group:	**Laboratory of Biosystematics**
University:	**Wageningen University & Research Centre**

1) Start-up phase	*date*
► **First presentation of your project**	
Quantitative trait loci associated with salt tolerance in lettuce	Feb 09, 2012
► **Writing or rewriting a project proposal**	
► **Writing a review or book chapter**	
Genetic diversity and evolution in *Lactuca* L. (Asteraceae) - from phylogeny to molecular breeding, submitted to Euphytica	Oct 10, 2015
► **MSc courses**	
► **Laboratory use of isotopes**	
Subtotal Start-up Phase	*7.5 credits**

2) Scientific Exposure	*date*
► **EPS PhD student days**	
EPS PhD student day 2012, University of Amsterdam	Nov 30, 2012
EPS PhD student day 2013, Leiden University	Nov 29, 2013
EPS PhD student day 2015 (Get2Gether), Soest (NL)	Jan 29-30, 2015
► **EPS theme symposia**	
EPS theme 4 symposium 'Genome Biology', Radboud University Nijmegen	Dec 07, 2012
EPS theme 4 symposium 'Genome Biology', Wageningen University	Dec 13, 2013
EPS theme 4 symposium 'Genome Biology', Wageningen University	Dec 03, 2014
EPS theme 3 symposium 'Metabolism and Adaptation', University of Amsterdam	Mar 22, 2013
EPS theme 3 symposium 'Metabolism and Adaptation', Wageningen University	Mar 11, 2014
EPS theme 3 symposium 'Metabolism and Adaptation', Utrecht University	Feb 10, 2015
► **NWO Lunteren days and other National Platforms**	
NWO-ALW meeting 'Experimental Plant Sciences', Lunteren (NL)	Apr 02-03, 2012
NWO-ALW meeting 'Experimental Plant Sciences', Lunteren (NL)	Apr 22-23, 2013
NWO-ALW meeting 'Experimental Plant Sciences', Lunteren (NL)	Apr 14-15, 2014
NWO-ALW meeting 'Experimental Plant Sciences', Lunteren (NL)	Apr 13-14, 2015
► **Seminars (series), workshops and symposia**	
Seminar 'Molecular insights into spore biology and metabolism of Phytophthora infestans, the potato blight pathogen'	May 07, 2013
Workshop 'iPlant Collaborative Wiki', Wageningen University	Mar 03-04, 2014
Seminar 'Water saving in rice: combining genetic, physiological, agronomic and modelling approaches to combat drought stress in rice crops'	May 13, 2014
Seminar 'Chromatin structure controls centromeres and secondary metabolism in filamentous fungi'	Oct 10, 2014
Seminar 'Genetics and epigenetics: a complex relationship'	Nov 09, 2014
Seminar 'Seasonal flowering in annual and perennial plants'	Jan 19, 2015
Seminar 'The evolutionary significance of gene and genome duplications'	Feb 03, 2015
Seminar 'Inferring species trees given coalescence and reticulation'	Mar 18, 2015

Seminar 'Polyploidy in wild relatives of soybean and other legumes: systematics, comparative and functional genomics, and nodulation'	May 12, 2015
Seminar 'Regulation of root morphogenesis in tomato species in the face of a changing environment'	Sep 09, 2015
► **Seminar plus**	
► **International symposia and congresses**	
Molecular mapping & marker assisted selection (Vienna, Austria)	Feb 08-11, 2012
GRC: Salt & Water Stress in Plants (Hongkong, China)	Jun 24-29, 2012
4th European Joint Retreat for PhD Students in Plant Sciences (Norwich, UK)	Aug 15-17, 2012
PhenoDays 2012 (Wageningen, The Netherlands)	Oct 12-14, 2012
5th European Joint Retreat for PhD Students in Plant Sciences (Ghent, Belgium)	Jul 23-26, 2013
GRC: Salt & Water Stress in Plants (Newry, US)	Aug 03-08, 2014
16th Young Systematists' Forum	Nov 21, 2014
► **Presentations**	
GRC: Salt & Water Stress in Plants (Poster)	Jun 26, 2012
GRC: Salt & Water Stress in Plants (Oral)	Aug 04, 2014
16th Young Systematists' Forum (Poster)	Nov 21, 2014
ALW meeting 'Experimental Plant Sciences', Lunteren (Talk)	Apr 14, 2015
► **IAB interview**	
► **Excursions**	
Subtotal Scientific Exposure	*19.0 credits**

3) In-Depth Studies	*date*
► **EPS courses or other PhD courses**	
Mixed model based QTL mapping in GenStat	May 14-16, 2012
Current Trends in Phylogenetics	Oct 22-26, 2012
Bioinformatics-A User's Approach	Mar 04-08, 2013
Introduction to R for Statistical Analysis	Jun 10-11, 2013
Basic Statistics	Jun 18,19, 24, 25, 26, 2013
Transcription Factor and Transcriptional Regulation	Dec 17-19, 2013
► **Journal club**	
► **Individual research training**	
Subtotal In-Depth Studies	*7.0 credits**

4) Personal development	*date*
► **Skill training courses**	
Practical English Plus	Sep 2012- Feb 2013
Scientific writing	Oct 17-Dec 05, 2013
Writing Grant Proposal	Apr 07-Jun 09, 2015
► **Organisation of PhD students day, course or conference**	
► **Membership of Board, Committee or PhD council**	
Subtotal Personal Development	*6.2 credits**

TOTAL NUMBER OF CREDIT POINTS*	**39.7**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits

** A credit represents a normative study load of 28 hours of study.*

The research in this thesis was performed at Biosystematics Group, Wageningen University, with financial support from China Scholarship Council (CSC) and Graduate School of Experimental Plant Sciences of the Netherlands.

Cover design: Zhen Wei and Hua Feng

Lay out: Zhen Wei

Printed by: Proefschriftmaken.nl || Uitgeverij BOXPress