Tracking the evolutionary history of the Allium ampeloprasum L. complex (section Allium) provides evidence of the contribution of North African diploids to the formation of allopolyploid horticultural groups

Thinhinan Khedim, Abdelkader Ainouche, Nabila Amirouche, Malika Ourari, Jean Keller, Malika-Lily Ainouche, Rachid Amirouche

Accepted Manuscript
"The final publication is available at link.springer.com"
Genetic Resources and Crop Evolution
Published Online: 21 April 2020
https://doi.10.1007/s10722-020-00948-x

Tracking the evolutionary history of the *Allium ampeloprasum* L. complex (section *Allium*) provides evidence of the contribution of North African diploids to the formation of allopolyploid horticultural groups

Thinhinan Khedim[1], Abdelkader Aïnouche[2], Nabila Amirouche[1], Malika Ourari[3], Jean Keller[4] Malika Lily Aïnouche[2], Rachid Amirouche[1]

[1]Université des Sciences et de la Technologie Houari Boumediene, Faculté des Sciences Biologiques, LBPO, BP n°32 El-Alia, Bab-Ezzouar, 16110 Alger, Algérie
[2]Université de Rennes 1, UMR-CNRS 6553, Ecobio, Campus de Beaulieu, Bat. 14, 35042 Rennes Cedex, France
[3]Université Abderrahmane Mira, Faculté des Sciences de la Nature et de la Vie, Route de Targa Ouzemour, 06000, Bejaia, Algérie
[4]Université de Toulouse, LRSV, CNRS, UPS, 24 chemin de Borde Rouge, Auzeville, BP42617, 31326 Castanet-Tolosan, France

Corresponding author
Rachid Amirouche
e-mail: ramirouche@usthb.dz

Abstract The economically important *Allium ampeloprasum* L. represents a polyploid complex, comprising hexa- and octoploid Great Headed Garlic horticultural cultivars (6*x*-8*x* GHG) and several traditional varieties of the tetraploid (4*x*) leeks (Leek, Bulbous leek, Kurrat and Pearl onion). Its wild representatives were indicated as rare in the Mediterranean region. This study aims to explore the diversity and origin of polyploidy in this complex, including its wild relatives *A. baeticum* and *A. guttatum* with particular focus on the poorly investigated North-African region. Natural populations were sampled in Algeria in various bioclimatic conditions, then subjected to karyological and molecular phylogenetic analyses based on nuclear rDNA *ITS* region and chloroplast *trnL−trnF* and *trnD−trnT* intergenic spacers. Comparative analyses included available Genbank accession sequences representing old-world relatives. Chromosome count surveys revealed an unexpected higher occurrence of diploid (2n = 16) than tetraploid (2n = 32) cytotypes. The phylogenetic analyses first allowed positioning the Algerian material within the *A. ampeloprasum* complex. Interestingly, all the Algerian diploid and tetraploid populations from *A. ampeloprasum* and *A. baeticum* form a distinct monophyletic group. The results provide novel and robust evidence demonstrating that the North African diploid *A. ampeloprasum* genetic pool widely contributed as a source of progenitors not only for the *A. ampeloprasum* and *A. baeticum* Algerian tetraploids, but also in the formation of the GHG and Leek cultivated allopolyploids. Therefore, the North African populations emerge as an important reservoir of new wild genetic resources of great interest for tracing the origin of crop domestication and for breeding programs of cultivated varieties.

Key words Molecular phylogeny, diploids, polyploids, *Allium ampeloprasum* complex, Algeria, wild genetic resources

Introduction

Genus *Allium* L. is the largest group of Amaryllidaceae, comprising over 800 species widely distributed in the Northern hemisphere, mainly in the Mediterranean region and Central Asia (Fritsch et al. 2010). Its remarkable taxonomic diversity is accompanied by occurrence of polyploidy (4*x*, 5*x*, 6*x*, 8*x* and 10*x*) and a high disploidy as shown by several base chromosome numbers *x* = 7, 8, 9, and 11 (Peruzzi et al. 2017). It constitutes an interesting group of species which has led to numerous molecular studies in order to elucidate infrageneric relationships and sectional classifications (Nguyen et al. 2008; Fritsch et al. 2010; Gurushidze et al. 2010; Wheeler et al. 2013;

Herden et al. 2016; Li et al. 2010, 2016; Sinitsyna et al. 2016). The origin of *Allium* crops and ornamental taxa were also investigated (Friesen et al. 1999; Gurushidze et al. 2007; Hirschegger et al. 2010; Veiskarami et al. 2019).

Presently, 15 subgenera and 56 sections are recognized (Friesen et al. 2006). The largest subgenus *Allium* comprises 280 species the majority of which being grouped in the economically important Mediterranean section *Allium*, including the genus type *A. sativum* L. (garlic) (Mathew 1996; Fritsch and Friesen 2002). Species referred to this section are undoubtedly among the most variable and taxonomically difficult to circumscribe, as it is exemplified by the noteworthy *A. ampeloprasum* L. complex. The latter represents a polyploid complex (4*x*, 6*x* and 8*x*) comprising several crops and horticultural varieties e.g. 4*x* leek, kurrat, pearl onion, bulbous leek and 6*x*-8*x* Great Headed Garlic (GHG) cultivar's which origin remains enigmatic and controversial (Bohanec et al. 2005). Several species native to the Mediterranean region (*A. polyanthum* Schultes and Schultes f., *A. pyrenaicum* Costa and Vayreda, *A. commutatum* Guss., *A. sphaerocephalon* L., *A. atroviolaceum* Boiss., *A. bourgeaui* Rechinger, *A. acutiflorum* Lois., *A. tuncelianum* (Kollm.) Ozhatay, B. Mathew and Siraneci and *A. leucanthum* C. Koch), as well as others from the Middle East (e.g., *A. truncatum* (Feinbr.) Kollmann and D. Zohary and *A. iranicum* (Wend.) Wend.), were regarded as their wild relatives (Bothmer 1970, 1974, 1982; Guern et al. 1991; Jauzein and Tison 2005). Autotetraploidy has been suggested in the leek group by cytological studies and meiotic behavior (Levan 1940; Stack and Roelofs 1996; Maragheh et al. 2018) while segmental allopolyploidy was also suggested (Khazanehdari and Jones 1997). Conversely, the genetic diversity and cytogenetic characteristics observed in polyploid genomes of some GHG accessions emphasized their allopolyploid origin (Kollmann 1972; Figliuolo et al. 2001; Hirschegger et al. 2006). Molecular phylogenetic analyses provided support to the occurrence of allopolyploidy in the GHG group, based on the detection of three phylogenetically divergent ITS ribotypes in its genome (Hirschegger et al. 2010). Nevertheless, this study could not identify the diploid progenitors of polyploids due to the lack of sampling over the natural range of the species. In fact, the scarcity, narrow geographical distribution and rarity of its diploid relatives were previously underlined (Jauzein and Tison 2005). Another molecular phylogenetic study, using two *A. ampeloprasum* samples, provided some clues suggesting a potential hybrid origin of *A. ampeloprasum* cultivars, with probably *A. iranicum* as one of their potential parents (Veiskarami et al. 2019). Interestingly, recent studies in North Africa revealed that numerous scattered diploid populations of *A. ampeloprasum* occur in Algeria and Tunisia (Khedim et al. 2010; Guenaoui et al. 2013). Accordingly, it was of interest to deepen our knowledge on these natural populations, in order to explore their relationships within the *A. ampeloprasum* complex (including the horticultural groups) and their potential as novel genetic resources for crop improvement.

In this study, chromosome numbers and phylogenetic relationships were examined within the *Allium ampeloprasum* polyploid complex and its two related taxa, *A. baeticum* Bossier and *A. guttatum* Steven. The objectives were to inventory and better understand the natural diversity of the *A. ampeloprasum* complex in the poorly explored northern Algerian regions (North Africa), in order to evaluate its evolutionary history. Here we present: (1) novel data on the geographical distribution of the wild diploid and polyploid populations of this species-group; and (2) new insights on their relationships and their involvement as progenitors of the horticultural groups, based on the *ITS* nuclear rDNA region and two chloroplast intergenic spacers *trnL−trnF* and *trnD−trnT*.

Material and Methods

Sampling and taxonomic identification

Fresh material was sampled in various bioclimatic conditions of Northern Algeria, including thirty-four wild populations belonging to *Allium ampeloprasum* and two other species, *A. baeticum* and *A. guttatum* (Supplementary Table S1). In each sampling site, 5 to 10 plants per taxon were collected and cultivated in the Experimental garden of Houari Boumediene University of Sciences and Technology (Algiers, Algeria). Type specimens were also examined from K, MPU, P and the ENSA herbaria (National High School of Agronomy, Algiers, Algeria). Plant determination was based on the specialized literature (Desfontaines 1798; Battandier and Trabut 1895, 1902; Maire 1958; Quézel and Santa 1962; De Wilde-Duyfjes 1976; Stearn 1980; Boulos 2005). The three sampled species share the diagnostic criteria of section *Allium* consisting of papyraceous or rarely fibrous bulb tunics, inflorescence possessing bracteoles, stamens mostly exserted with tri-cuspidate internal filaments (Fig. 1). Some characters such as papilla on tepals, tubercles on the margins of leaves and shape of bulbils were useful at species-level delimitation (Bothmer 1974, 1975; Jauzein and Tison 2005). We have also attempted to assign collected populations to the different varieties described by Maire (1958) for each taxon. All samples were subjected to karyological investigations in order to establish their chromosome number and ploidy level. Molecular phylogenies were based on sequences generated from our fresh materiel compared to sequences obtained from GenBank (http://www.ncbi.nlm.nih.gov/), corresponding to representatives of the *A. ampeloprasum* complex including all ploidy levels (2*x*, 4*x*, 6*x* and 8*x*). Other Mediterranean and Eurasian taxa belonging to the subgenus *Allium* were used as out-group (Supplementary Table S2).

Chromosome counts

The method followed the procedure used in Khedim et al. (2016). Briefly, chromosomes preparation was obtained from metaphase plates of root-tip cells from cultivated bulbs. Young roots (8–12 mm long) were pretreated with

0.03 % 8-hydroxyquinolein for 12 hours at 4°C and fixed in ethanol-acetic acid (3:1) for 48 hours, then hydrolyzed in 1N hydrochloric acid for 12–15 min at 60°C. Root tips were stained according to the Feulgen procedure, and then squashed in a drop of 45 % acetic acid. Metaphase plates were examined with a Zeiss Axiostar-Plus Microscope equipped with a Canon digital Camera. Chromosome counts of each population were performed from at least five exploitable metaphases plates.

Molecular phylogenetic analysis

DNA extraction, amplification and sequencing

Total genomic DNA of the 32 sampled populations was extracted from 100 mg of fresh leaf, using the Nucleospin Plant II kit (Macherey-Nagel). Nuclear and chloroplast DNA sequences were amplified. For nuclear sequences, we screened the internal transcribed spacer region (*ITS*) of ribosomal DNA including ITS_1 + *5.8S* + ITS_2 using universal ITS_1 and ITS_4 primers (Baldwin et al. 1995). Two non-coding chloroplast DNA regions corresponding to intergenic spacers (*IGS*), *trnL−trnF* and *trnD−trnT*, were amplified using universal primers: *trnL*(c) and *trnF*(f) (Taberlet et al. 1991) and *trnD*(d) and *trnT*(t) (Demesure et al. 1995). PCR was performed in a solution of 50 µL with 5 µL of 10× Taq-buffer, 5 µL of 2 mM dNTPs mix, 0.5 µM of each primer (forward and reverse), 2.5 µl of Green Go Taq DNA polymerase (Promega), and 1µl of DNA (20-50 ng/µl). PCR program was started by 3 min of DNA denaturation at 94°C, followed by 30 cycles of 1min denaturation at 94°C, 1 min at 48°C for primer annealing and 2 min of extension at 72°C for each cycle. A 7 min final extension at 72°C followed cycle 30. PCR products were purified with the EZ-10 Spin Column PCR Products Purification Kit (BioBasic Inc.).

Sequence homogeneity in polyploid genomes of *A. ampeloprasum*, *A. guttatum* and *A. baeticum* were verified by searching multiple *ITS* copies within individuals. ITS-PCR products were cloned using the pGEM-T easy Vector System Cloning Kit (Promega). The DNA of clones was isolated with Wizard plus Minipreps Purification Kit (Promega). Positive plasmids containing an insert were sequenced employing T7 and SP6 primers. 80 amplicons of *ITS* region (with average of 10 amplicons per individual) were obtained. Finally, 130 clean PCR products and amplicons were carried out using ABI system sequencer (Macrogen Inc. Company, Seoul, South Korea). All sequences were deposited under GenBank accession numbers (MG546691 − MG546821).

Phylogenetic analyses

Data sets were constructed for *ITS*, *trnL−trnF* and *trnD−trnT* sequences including data from this study and from GenBank (Friesen et al. 2006; Hirschegger et al. 2010; Figliuolo and Di Stefano 2007; Guenaoui et al. 2013). The analyzed taxa are presented in Table S2, including origins, distributions, chromosome numbers, ploidy levels, references and GenBank accession numbers. Sequences representing species from section *Allium* and other sections from the subgenus *Allium* were selected, based on the tree topology obtained following phylogenetic analysis of the entire genus data sets. Statistic parameters were computed for each sequence data set (Table S3).

The original *ITS* data set included 160 sequences, among them 22 from diploids and 80 amplicons from Algerian tetraploids. *Allium umbilicatum* Boiss., and *A. brevidens* Vved. from subgenus *Allium*, *A. cepa* L. and *A. fistulosum* L. from subgenus *Cepa*, were chosen as outgroup based on previous studies (Hirschegger 2010). Phylogenetic analyses were first conducted on diploids (from Algeria and other origins) including 45 sequences, and then a global analysis was conducted on a data set including 147 sequences with all *ITS* amplicons from Algerian tetraploids (4*x*) and other polyploids (4*x*, 6*x*, 8*x*) of various origins.

Chloroplast DNA analyses were conducted separately for each region *trnL−trnF* and *trnD−trnT*, then combined with Bioedit4 software (Hall 1999). The final chloroplast data set included 56 sequences of which sequences from *A. cepa* and *A. fistulosum* were used as outgroup. Sequences were aligned with the software MUSCLE (Edgar 2004). Phylogenetic analyses based on Maximum Parsimony (MP) and Maximum Likelihood (ML) methods were conducted using Mega 10 software (Tamura et al. 2013). MP analyses were performed by heuristic searches with Subtree-Pruning-Regrafting (SPR) and 10 random addition sequence replicates. The unweighted MP analysis resulted in most parsimonious trees. Consistency index (CI) and retention index (RI) were calculated. ML analyses were performed following the procedure described in Harrison and Langdale (2006). The appropriate nucleotide substitution model of sequence evolution was determined using Mega 10, based on the lowest value of the Bayes Information Criterion (BIC). For each nuclear or plastid DNA analysis, the robustness of clades was estimated by bootstrap method (BS) with 1000 replicates (Felsenstein 1985).

Results

Chromosome numbers and ploidy level

Karyological data on the three investigated species *Allium ampeloprasum, A. baeticum* and *A. guttatum* emphasized two ploidy levels, a diploid one with 2n = 2*x* = 16 and a tetraploid 2n = 4*x* = 32 (Fig. 2; Table S2). The morphology and number of satellized chromosomes differ at infraspecific level. In all cases, satellite numbers varied between two and four per cytotype, with distal (terminal) or proximal (intercalary) position on the short arm of chromosomes (Fig. 2b-d,i,k,l). It is interesting to notice that the diploid cytotypes of *A. ampeloprasum* sampled in Algeria were scattered in the northeastern region whereas the tetraploid cytotypes occur mainly in the west-south localities. Both diploid and tetraploid cytotypes of *A. guttatum* were found in North of Algeria. With

regard to *A. baeticum*, the diploid cytotypes are widespread from the North to the South part of Central Algeria, while the tetraploid cytotype was restricted to the western area.

DNA sequences diversity

Sequences of *ITS* and cpDNA (*trnL–trnF* and *trnD–trnT*) regions were analyzed separately with the Maximum Likelihood (ML) and Parsimony analysis (MP) methods. The summary of statistical parameters for each data sets and phylogenetic analysis is presented in Supplementary Table S3.

Diversity of the ITS sequences of diploid taxa

The aligned *ITS* matrix includes 45 sequences of 11 taxa. The *ITS* region ranged from 585 bp for *A. brevidens* Vved. (AJ412721) to 614 bp for *A. sphaerocephalon* (AJ412717). The final alignment was 617 bp long with 216 variable sites, of which 146 were parsimony-informative. The unweighted MP resulted in five most parsimonious trees for 279 steps long, with a consistency index of CI = 0.7397 and a retention index RI = 0.9068. For the ML analysis, the nucleotide substitution model selected by the lowest value of BIC was Kimura 2-parameter with invariable evolution (K2+I).

Diversity of the ITS sequences of diploid and polyploid taxa

The aligned *ITS* matrix includes 147 sequences from both diploid and polyploid taxa analyzed (13 taxa). The *ITS* region ranged from 585 bp in *A. brevidens* (AJ412721) to 614 bp in *A. sphaerocephalon* L. (AJ412717). No major length polymorphisms were observed among amplicons of the Algerian tetraploids and other amplicons previously generated. The final alignment was 620 bp long with 303 variable sites, of which 182 were parsimony-informative. The unweighted MP analysis resulted in three most parsimonious trees of 491 steps long, with a CI = 0.5374 and RI = 0.8840. For the ML analysis, the nucleotide substitution model selected was Tamura 3-parameter with discrete gamma distribution (T92+G).

Chloroplast DNA data set

The *trnL−trnF* region matrix includes 57 sequences from 17 taxa. The length of sequences ranged from 545 bp in *A. ampeloprasum* Kristel (MG546697) to 558 bp in *A. cepa* (FJ628603). The final alignment was 574 bp long with 82 variable sites of which 45 were parsimony-informative. The t*rnD−trnT* matrix includes 60 sequences of the same set of taxa. The length of sequences ranged from 859 bp in *A. leucanthum* (EU626273) to 890 bp in *A. ampeloprasum* Tikjda (MG546703). The final alignment was 928 bp long with 181 variable sites of which 54 were parsimony-informative. The t*rnL−trnF* and *trnD−trnT* combined data set includes 56 sequences. The length of sequences ranged from 1306 bp in *A. leucanthum* (EU626259) to 1344 bp in *A. ampeloprasum* Tikjda (MG546703/MG546730). The final alignment was 1388 bp long with 188 variable sites of which 79 were parsimony-informative. The unweighted MP analysis resulted in one most parsimonious tree of 245 steps long with a CI = 0.6343 and RI = 0.8689. For the ML analysis, the nucleotide substitution model was Tamura 3-parameter with discrete gamma distribution (T92+G).

Phylogenetic analyses

Maximum Likelihood and Parsimony analyses generated similar tree topologies and all major relationships were well supported by the two methods. Only ML phylograms or cladograms are presented (Figs. 3, 4, 5).

Analysis of diploids

The *ITS* ML tree (Fig. 3) shows that all diploids taxa, traditionally recognized as members of section *Allium*, form monophyletic group (99 % BS) which split into two major sister clades, A and B supported by 95 % and 100 % BS respectively. Within clade A, one supported monophyletic group (86%) includes all North African (NA) diploid populations referred to *A. ampeloprasum* and *A. baeticum*, hereafter called the NA clade (Fig. 3). Moreover, two well supported subclades can be distinguished in the latter NA clade: The first (hereafter called NA1) contains all Algerian diploid populations of *A. ampeloprasum* and *A. baeticum* (99 % BS); the second one (NA2, 100 % BS) includes only two populations of *A. ampeloprasum* originating from Samaeillette and Logroa in Tunisia. Within the clade A, other minor well supported groups are represented by accessions from *A. sativum*, *A. scorodoprasum*, and by three Algerian populations of *A. guttatum* which are closely related to a sample of *A. sphaerocephalon.* The *A. guttatum* Algerian population from Yakouren (central-eastern Algeria) is not resolved and stands apart from the *A. guttatum-A. sphaerocephalon* group. The clade B (100 % BS) only includes four GenBank accessions, two referred to *A. commutatum* and two other referred to *A. ampeloprasum* originating from Sardinia.

Analysis of diploids and polyploids

In order to clarify the relationship of polyploids within the *Allium ampeloprasum* complex, novel ML and MP phylogenetic analyses were performed on a data set including all diploid and polyploid *ITS* sequences generated from this study, together with those available from GenBank to represent the diversity of section *Allium*. The resulting general ML tree is provided in Fig. S1. A reduced version is presented in Fig. 4. In the latter tree, remarkable clades are condensed: The North African subclade NA1 (including G2) and the two distinct groups

(G1 and G3), representing amplicons of undetermined origin found by Hirschegger et al. (2010) in various polyploid accessions of the *A. ampeloprasum* horticultural group (Leek group and GHG).

In this tree, all amplicons generated from Algerian tetraploid populations of *A. ampeloprasum* and *A. baeticum* are linked to the Algerian diploids of these same two taxa within subclade NA1 (91% BS). Moreover, all amplicons of the group G2 of Hirschegger et al. (2010), representing clones coming from GHG hexaploid and octoploid accessions and from the tetraploid South-European *A. polyanthum*, are clearly linked to the Algerian diploids and tetraploids (*A. ampeloprasum* and *A. baeticum*) within the NA1 subclade. The monophyletic group G3 (93% BS) of Hirschegger et al. (2010), which contains amplicons originating from octoploid GHG accessions, from the tetraploid Leek group (Leek, Kurrat, Pearl onion and Bulbous leek), *A. truncatum*, *A. pyrenaicum* and *A. iranicum*, obviously shares a common origin with the two Tunisian diploids of the NA2 subclade (51 % BS). The amplicons generated from the Algerian tetraploid population of *A. guttatum* (from Tazrout) rather share a common origin with the diploid population from Yakouren (outside the NA clade) than with the other *A. guttatum* diploid group (from Chréa, Benchicao and Cap Ténès). The remaining amplicons from GHG (6*x* and 8*x*) accessions (including those of *A. ampeloprasum* var. *babingtonii*) and from *A. pyrenaicum* (4*x*), form the monophyletic group G1 (99% BS) of Hirschegger et al. (2010) outside the NA clade.

Two other slightly divergent amplicons from *A. pyrenaicum* are related to group G1, as well as few other remaining amplicons representing the leek group (Bulbous leek and Pearl onion) and *A. iranicum*, which are poorly resolved outside the NA clade.

Initially, this phylogeny also included GenBank *ITS* sequences from European polyploid accessions (*A. atroviolaceum* (4*x*), *A. bourguoui* (4*x*), *A. dregeanum* (8*x*), *A. pseudoampeloprasum* (4*x*), *A. leucanthum* (4*x*?)), and other accessions with uncertain ploidy levels (*A. tuncelianum* (2*x*?) and *A. acutiflorum* (2*x*?)) which were all positioned outside the NA clade. However, as these sequences come from direct sequencing (i.e. not cloned), they were not included in Fig. 4 and Fig. S1 to prevent phylogenetic ambiguities, due to sequence uncertainties, which did not affect further interpretation of the tree topology. Thus, there is evidence that the undetermined sequences of the groups G2 and G3 (from Hirschegger et al. 2010) most likely derive from the wild North African *A. ampeloprasum* and *A. baeticum* gene pool (from Algeria and Tunisia). It appears that this gene pool obviously contributed, not only to the formation of the Algerian tetraploids but also to the formation of the polyploid accessions of the GHG group (6*x*, 8*x*), the leek group *sensu lato* (Leek, Kurrat, PO, BL), and of the tetraploids *A. truncatum*, *A. iranicum* and *A. pyrenaicum*.

Analysis of chloroplast DNA

The chloroplast regions *trnD−trnT* and *trnL−trnF* were first analyzed separately. Nearly the same and poorly resolved tree topology resulted from these two data sets; with however moderate bootstrap support to some clades in the *trnL−trnF* tree (not shown). The ML tree of the combined dataset presented in Fig. 5 shows two main clades, in accordance with the clades I and II identified by Hirschegger et al. (2010). The clade II (86% BS) includes *Allium* species with different ploidy levels, where *A. tuncelianum* (2*x*?) is sister (with 91% BS) to the group comprised of *A. scorodoprasum* (2*x*), *A. sativum* (2*x*), *A. leucanthum* (4*x*?), and *A. pseudoampeloprasum* (4*x*). All diploid and tetraploid Algerian samples of *A. ampeloprasum*, *A. baeticum* and *A. guttatum*, exclusively fall in the large clade I (84 % BS) together with Genbank accessions representing the remaining polyploid taxa in the section *Allium*, of which the *A. ampeloprasum* Leek and GHG groups. Although a low cpDNA sequence divergence, some moderately supported subgroups can be observed in this clade. The tetraploid *A. ampeloprasum* Leek group (63% BS) forms a subclade (63% BS) together with the tetraploids *A. iranicum*, *A. truncatum* and *A. atroviolaceum*. Another subclade with a higher support (87% BS) includes both hexa and octoploid accessions of the *A. ampeloprasum* GHG groups (85% BS), the group of the tetraploid *A. commutatum* and *A. bourgeaui* (95% BS), *A. polyanthum*, together with both 2*x* and 4*x* Algerian accessions of *A. guttatum* and one tetraploid population of *A. ampeloprasum* (Kristel, from West Algeria). Apart from the latter population, the cpDNA tree supports close maternal relationships of the *A. ampeloprasum* and *A. baeticum* tetraploids Algerian accessions with their North African diploid relatives.

Discussion

Novel diploid and tetraploid North-African wild resources for the *A. ampeloprasum* complex and section *Allium*

Polyploidy has already been considered as a major mechanism of evolution in the genus *Allium* (Mathew 1996). As demonstrated recently by Han et al. (2019), the diversification within genus *Allium* can be determined by intraspecific polyploid frequencies through climatic and habitats shifts. In the Mediterranean region, the *A. ampeloprasum* group displays a stable chromosome base number $x = 8$ (Hirschegger et al. 2006; Peruzzi et al. 2017). In this region, polyploids have been widely distributed with high prevalence of tetraploids, pentaploids and hexaploids, whereas diploids have been sporadically encountered in the eastern areas (Kollmann 1972). In fact, the Eastern Mediterranean region accumulates the whole set of the ploidy levels (3*x*, 4*x*, 5*x*, 6*x* and 7*x*) centered around the Aegean area (Bothmer 1970, 1975; Karavokyrou and Tzanoudakis 1991). Throughout the Western

Europe, 4*x*, 5*x* and 6*x* cytotypes have been reported in the Iberian Peninsula (Pastor 1982) and only hexaploids were observed in Holm Island and Ile D'Yeu Island (Stearn 1980; Guern et al. 1991; Jauzein and Tison 2005).

The natural populations from northern Algeria referred to as *A. ampeloprasum* (this study) exhibited an unexpected high frequency of diploids (2*x*) and tetraploids (4*x*). Among the two other related taxa *A. guttatum* and A. *baeticum*, only 2*x* and 4*x* cytotypes were found. Populations of *A. guttatum* were almost all diploids, except one tetraploid encountered in the locality of Tazrout. For this species, diploids were reported only in Eastern Mediterranean region (Tzanoudakis and Vosa 1988), as it was for *A. ampeloprasum* diploids. *A. guttatum* polyploids (4*x*, 5*x* and 6*x*) were reported in southeastern Europe (Stearn 1980). Interestingly for *A. baeticum*, while only tetraploid number was reported in Iberian material (Pastor 1982), our study allowed detection of diploids for the first time among the Algerian populations. Therefore, these new reports of diploid and tetraploid populations in North Africa represent a substantial enlargement of natural resources available to investigate origins, evolutionary relationships and wild relatives of subspontaneous and cultivated polyploids within *A. ampeloprasum* complex and section *Allium*.

North-African wild resources provide new phylogenetic insights into section *Allium*

In order to situate the Algerian populations within the section *Allium*, and in turn, to fill the gap of our knowledge on species diversity in the North African region, we used a previous phylogenetic estimate of section *Allium* as a reference framework (Hirschegger et al. 2010). We reconstructed novel phylogenies based on *ITS*, *trnL−trnF* and *trnD−trnT* data sets, including a wide range of sequences available in GenBank (Friesen et al. 2006; Hirschegger et al. 2010; Figliuolo and Di Stefano 2007; Guenaoui et al. 2013) and those generated from this study. Numerous former studies demonstrated that *ITS* sequence variation is not only helpful for phylogenetic inference at infrageneric and specific levels in a wide range of taxa, but also that it might be a useful marker for the detection of past cases of reticulation and allopolyploid speciation (Wendel et al. 1995; Aïnouche and Bayer 1997; Wendel 2000). Intragenomic diversity of *ITS* sequences may reflect evolutionary divergences among paralogous copies, but also may results from different parental *ITS* ribotypes acquired following hybridization events. These multiple *ITS* loci are generally subjected to a more or less rapid homogenization through concerted evolution (Álvarez and Wendel 2003; Kovarik et al. 2005). Depending on their age, their life traits (life cycle, mode of reproduction), their intrinsic genomic features and evolutionary rate, the recently formed hybrids and/or allopolyploids may show an incomplete sequence homogenization and may retain parental *ITS* ribotypes (Suárez-Santiago et al. 2007; Liu et al. 2008; Soltis et al. 2008; Logacheva et al. 2010; Poczai and Hyvönen 2010). The intra-individual polymorphism of *ITS* loci was previously reported in several *Allium* species (Dubouzet and Shinoda 1998; Figliuolo and Di Stefano 2007; Gurushidze et al. 2007, 2008). Hirschegger et al. (2010) revealed a diversity of ribotypes at both inter- and intragenomic levels within section *Allium* and were able to suggest or confirm the auto- or allopolyploid origin of the horticultural groups and their close relatives. However, the detection of unidentified groups of ribotypes (G1, G2, G3) from Hirschegger et al. (2010), and their limited sampling of diploids, did not allow them to accurately identify their parental origins. The results generated from this study, based on a wider sampling and cloning to detect intragenomic sequence heterogeneity, allowed phylogenetic circumscription of the novel wild North African *Allium* resources, and shed new light on diversity and relationships at different ploidy levels within the *A. ampeloprasum* complex and section *Allium*. The results hereafter discussed are summarized in the phylogenetic schema presented in supplementary Fig. S2.

New insights on the main basic diploid evolutionary lineages in section *Allium*

In accordance with Hirschegger et al. (2010), the main basic lineages of *ITS* sequences derived from known extant diploid taxa included in this study were well circumscribed in the *ITS* phylogeny, although their inter-relationships remain widely unresolved within section *Allium*. These subgroups correspond to *A. ampeloprasum*, *A. sphaerocephalon*, *A. scorodoprasum*, *A. sativum* and *A. commutatum*. Ribotypes representing subgenomes of the allotetraploids *A. pyrenaicum* and *A. iranicum* are also positioned in the trees (Fig. 4; Fig. S2). Our results demonstrated that all the North-African (NA) diploid populations of *A. ampeloprasum*, as well as those of the previously uninvestigated *A. baeticum* and *A. guttatum*, are clearly positioned in section *Allium*. Within the North African genomic pool (NA), we detected two divergent diploid subgenomic pools, NA_1 which is widely distributed from western Algeria to eastern Tunisia (including accessions referred to as *A. ampeloprasum* and *A. baeticum*) and NA_2 which seems restricted to South Tunisia (Fig. 4; Fig. S2). The *A. guttatum* diploid populations were clearly distinguished from the NA clade, and splitted into two unresolved lines roughly corresponding to a western-eastern divergence of the Algerian populations; the western ones (Chréa, Benchicao and Cap Ténès) being unambiguously related to the representative of the diploid species *A. sphaerocephalon* included in this study (AJ412717 from Turkey). Although a low sequence divergence (Fig. 5, clade I), the cpDNA data also lend support to the *ITS* distinction between the NA and the *A. guttatum* populations, which suggests that they derive from close but divergent maternal origins. Additionally, it is interesting to notice from these results that accessions referred to as *A. ampeloprasum* (Figliuolo and Di Stefano 2007) in Sardinia (Italy) are out of the *A. ampeloprasum* diploid lineages and are rather part of the *A. commutatum* lineage, which is strongly supported from both nuclear and chloroplast data. These new insights, on the main basic ribotype lines revealed in section *Allium* from the North-African wild genetic resources, provide crucial complementary information to illuminate the general evolutionary framework of section *Allium* (Fig. S2).

New insights on the origin of the horticultural groups and uncultivated polyploids in section *Allium*

All tetraploid accessions referred to as *A. ampeloprasum* and *A. baeticum* in North Algeria clearly derive from the NA_1 genomic pool, as supported by both *ITS* and cpDNA data, suggesting their autotetraploid formation, which will need to be fully verified with other nuclear markers. The other three tetraploid Algerian accessions referred to as *A. guttatum* are unambiguously related to the eastern *A. guttatum* line (represented by an accession from Yakouren). Otherwise, one of the major results highlighted by this study is that our *ITS* data illuminate the origin of the previously unidentified groups of amplicons derived from several polyploids in section *Allium* (by Hirschegger et al. 2010). As summarized in Fig. S2, it has been shown that: (i) all G2 amplicons of Hirschegger et al. (2010), representing clones coming from GHG 6*x* and 8*x* accessions (including one clone of *A. ampeloprasum* var. *babingtonii*) and from the tetraploid *A. polyanthum*, clearly derive from the NA_1 genomic pool; (ii) that G3 amplicons, representing clones from GHG 8*x* accessions, and from the tetraploids of the Leek group (Leek, Kurrat, Pearl onion and Bulbous leek), *A. truncatum*, *A. pyrenaicum* and *A. iranicum*, share a common origin with the NA_2 Tunisian diploid genomic pool; and (iii) that G1 amplicons, originating from both 6*x* and 8*x* GHG accessions (including those of *A. ampeloprasum* var. *babingtonii*) clearly stand out of the NA clade and are closely related to other amplicons representing one subgenome of the tetraploid *A. pyrenaicum*, which are close relatives to *A. atroviolaceum*, *A. leucanthum*, *A. pseudoampeloprasum* (Hirschegger et al. 2010). All together, these results provide a more comprehensive overview of the diversity within the *A. ampeloprasum* complex and section *Allium* and allow making assumptions on the evolutionary history of the main horticultural groups and uncultivated polyploids, hereafter discussed.

***Allium porrum* L.** (4*x* Leek group) ─This group (including all its forms: leek, kurrat, bulbous leek and pearl onion) was extensively studied for its economic interest. Traditionally, these cultivars have been assigned to *A. ampeloprasum* as varieties or subspecies (*A. ampeloprasum* subsp. *porrum* (L.) Hayek) (Maire 1958; Stearn 1980; Guern et al. 1991; Khazanehdari and Jones 1997). Also, the leek group was referred to as the species *A. porrum* (Levan 1940; Gray et al. 1987; Stack and Roelofs 1996), which found support from various genetic data (Bohanec et al. 2005). Regardless of their taxonomic treatment, both *ITS* rDNA and cpDNA phylogenies (Hirschegger et al., 2010) also support the leeks as a homogeneous evolutionary entity. Previous studies based on meiotic chromosome behavior already suggested an autotetraploid origin of leeks (Levan 1940; Stack and Roelofs 1996), while others supported segmental allopolyploidy based on karyotype details (Khazanehdari and Jones 1997). As demonstrated by the *ITS*-based phylogenies (Hirschegger et al. 2010; Veiskarami et al.2019; and this study Fig. 4), the leeks are close relative to *A. iranicum*. Together, the latter appear to share a common origin of at least part of their ribotypes with the East Mediterranean tetraploid *A. truncatum*. Interestingly, our results revealed that few diploid populations from South Tunisia (NA_2 group) are not members of the wide NA_1 clade, but are rather closer to *A. truncatum*, and hence close to the group including one subgenome of "*A. iranicum* and the Leeks". The derived phylogenetic placement of these populations within *A. truncatum* (Fig. 4 and supplementary Fig. S2), raises the question of their taxonomic status, and it can be speculated that they could represent recently introduced diploid genotypes from the Middle East (the native region of *A. truncatum*) to South Tunisia. However, elucidating these questions will need further more accurate investigations.

***Allium ampeloprasum* var. *holmense* (Mill.) Asch. and Graebn. (6*x*-8*x* Great Headed Garlic group)**

Previous reports suggested the separation of the *A. ampeloprasum* var. *holmense* from the tetraploid cultivars of the leek group (Kik et al. 1997; Ariga et al. 2002; Bohanec et al. 2005), while *A. ampeloprasum* var. *babingtonii* was confirmed as an isoclonal form of GHG (Treu et al., 2001). According to the phylogeny based on cpDNA data (Fig. 5), the GHG accessions (regardless of their ploidy level) appear to share the same maternal progenitor, which most likely derived (together with those of the diploids *A. guttatum*, *A. commutatum* and *A. bourgeaui*), from *A. polyanthum* within the *ampeloprasum* complex. As shown above, the leek group clearly showed a distinct maternal origin from that of the GHG groups.

The allopolyploid nature of the octo- and hexaploid GHG cultivars was demonstrated in previous studies (Hirschegger et al. 2006) and was well illustrated by the intragenomic heterogeneity of ribotypes found by Hirschegger et al. (2010) in the 8*x*-GHG accessions (containing G1, G2 and G3 ribotypes) and the 6*x*-GHG ones (with G1, G2 ribotypes) (Fig. 4). According to our phylogenetic identification of the origin of the enigmatic ribotypes observed in the GHG genomes, this study provides evidence that the North African "*ampeloprasum-baeticum*" genomic pool unambiguously contributed to the formation of GHG horticultural groups and to their close polyploid relatives in section *Allium*. The results demonstrated that both 6*x*- and 8*x*-GHG genomes (including *A. ampeloprasum* var. *babingtonii*) share very similar ribotypes (G2) that are poorly divergent from and closely related to those of most diploid and tetraploid Algerian and Tunisian accessions of the *A. ampeloprasum* and *A. baeticum* genomic pool and to those of the tetraploid *A. polyanthum* (within the NA_1 clade; Fig. 4). With respect to its plastid and ribotype phylogenetic affinities, *A. polyanthum* appears as the most likely maternal progenitor line of the 6*x*- and 8*x*-GHG groups. As already shown by Hirschegger et al. (2010), the latter groups also exhibited ribotypes (G1 amplicons) which indicates that they share close relationships (out of the NA clade) with *A. pyrenaicum* and the genetic pool formed by *A. atroviolaceum*, *A. pseudoampeloprasum* and *A. leucanthum*. Also, there is evidence from the data of Hirschegger et al. (2010) and this study that the 8*x*-GHG separately (but not the 6*x*-GHG) inherited from their progenitors a third line of ribotypes (G3) which are strongly related to those of the

leek cultivars and *A. iranicum*, and next to *A. truncatum*. Therefore, although a distinct maternal origin has been underlined between the leek and the GHG cultivars (see above), it is obvious from the nuclear data that the 8*x* GHG share a common recent origin with the leeks (Leek, Kurrat, Pearl onion and Bulbous leek).

Taxonomical remarks on related wild species in North Africa

Allium ampeloprasum L. (2n = 16; 2n = 32) ─The wild representatives of this taxon were indicated as rare in the Mediterranean region (Bothmer 1970; Jauzein and Tison 2005). Maire (1958) had reported in North Africa, six endemic varieties within the subsp. *eu-ampeloprasum* Hayek. The Algerian populations studied here correspond to four of these varieties and are illustrated in figure 1 (Fig. 1a,b: cf. var. *typicum* Regel; Fig. 1,c,d): cf. var. *duriaeanum* (Gay) Batt.; Fig. 1e,f: cf. var. *tortifolium* Batt. and Fig. 1g,h: cf. var. *getulum* Batt.).

In this study, both *ITS* and cpDNA grouped all the *A. ampeloprasum* populations in a well-supported monophyletic assemblage, but provided no clues of delimitation at the varietal level. Despite the morphological, ecological and karyotypical diversity of the tetraploid populations, no divergent *ITS* copies in amplicons were observed. Therefore, the tetraploid populations are not distinguished from the diploids and do not appear as an independent taxonomic unit within this specific complex.

Allium baeticum Boissier (2n = 16; 2n = 32) ─This species previously considered as restricted to Morocco and Iberian Peninsula (Valdès et al. 2002), represents new report for the Algerian flora. Maire (1958) has described in the Rif and Moroccan Atlas two varieties: var. *laeve* Maire and Weiller and var. *papillosum* (Lindberg) Maire and Weiller. According to cpDNA phylogeny, this species is closely related to the Iberian endemic *A. pyrenaicum*. In *ITS* phylogeny both of *A. baeticum* diploids and amplicons of tetraploids were positioned in the same clade (Fig. 4 and supplementary Fig. S1). The Algerian samples of *A. baeticum* referred to as var. *laeve*, diploids as well as tetraploids, show no significant morphological differentiation (Fig. 1i-l). The tetraploid Iberian *A. baeticum* could have derived from North African diploids following polyploidization events.

Allium guttatum Steven (2n = 16; 2n = 32) ─In the Algerian floras this species was described under the specific epithet *A. margaritaceum* Smith (Maire 1958; Quézel and Santa 1962). The numerous varieties described by Maire (1958) attest to high degree of polymorphism of this species, as shown in figure 1 (Fig. 1m,n: cf. var. *battandieri* Maire and Weiller and Fig. 1o,p: cf. var. *typicum* Regel). In Europe, three subspecies were reported, *A. guttatum* subsp. *sardum* Stearn and subsp. *dalmaticum* (Kerner ex Janchen) Stearn are endemic to Sardinia and Croatia, respectively (Stearn 1980), and the Euro-Mediterranean *A. guttatum* subsp. *guttatum* Steven, to which could belong the Algerian populations. The *ITS* phylogeny indicated in the first instance that *A. guttatum* is closely related to another Mediterranean taxon *A. sphaerocephalon*. Although this relationship is well supported, it should be reassessed in a context of a broader sampling of this taxon, as any European population of *A. guttatum* has been sequenced to date. Moreover, it is interesting to note that the Algerian populations of *A. guttatum* share a common maternal origin with *A. commutatum* and could represent a putative maternal progenitor of *A. ampeloprasum* GHG cultivars.

Biogeographical consideration and polyploidy

Most wild and cultivated taxa of the *A. ampeloprasum* polyploid complex, particularly those of high ploidy levels (6*x*, 8*x*), are closely related to the North African diploids *A. ampeloprasum* and *A. baeticum*. Multiple patterns could be hypothesized to retrace the evolutionary history which led to the formation of the *A. ampeloprasum* horticultural groups. Anyway, the involvement of North African diploid genomes is here clearly demonstrated. Phytogeographical investigations have identified a major regional hotspot of biodiversity in northern Algeria and Tunisia, which is centered in the coastal and mountainous habitats particularly from Kabylies and Kroumirie (Véla and Benhouhou 2007). The high species richness and endemism suggest that this region may hold several refugia subsequently to the Pleistocene glaciations (Médail and Diadema 2009).

Among the fifty-two refugia recognized in the Mediterranean region, eight are located in North Africa (Médail and Diadema 2009) forming discontinuous areas that could have favored an active speciation process and playing a key role in maintaining biodiversity. That is emphasized by the occurrence of diploids of *A. ampeloprasum* in Algeria while their rarity in the rest of Mediterranean region has been signaled (Bothmer 1970; Kollmann 1972). The Quaternary glacial and interglacial oscillations in the circum-Mediterranean region could have broadened the distribution of diploids of *A. ampeloprasum* complex, initially confined in refugial areas, which probably generating the actual polyploid lineages as it was demonstrated in various secondary-contact speciation model (Taberlet et al. 1998; Casazza et al. 2012; Alix et al. 2017). According to our results (Fig. 4; Fig. S1), the North African samples are separated in a first subclade of native populations from North Algeria and Tunisia (NA_1) and in a second one from South Tunisia (NA_2). The North African clade (NA) as a whole is monophyletic (BS = 89%) and this separation implies that two distinct biogeographic patterns led to the diversification within the *A. ampeloprasum* complex. The present phylogenetic analysis including diploids indicates that Northern Algeria and Tunisia likely represent the center of diversity of this clade (NA), where dispersal events have led to lineages expansion across southwestern Europe and through the Middle East. Hence, current western Mediterranean tetraploids (*A. baeticum*, *A. pyrenaicum* and *A. polyanthum*) could have originated from North African diploids having undergone paleopolyploidization in the western areas of its range, before rising towards the Iberian

Peninsula. Other lineages could have borrowed the eastern pathway generating other tetraploids (*A. truncatum* and *A. iranicum*). In contrast, some regional endemics, such as *A. pyrenaicum* within the Northern African clade, suggest an enigmatic pattern of evolution regarding the closest relations between endemics and domesticated species. Within *A. guttatum*, the diploids seem to have an extensive distribution, from northeastern Africa to eastern Mediterranean region. This is probably a taxon which diverged earlier from the *A. ampeloprasum* complex.

Conclusion

This study represents the first karyological and molecular investigation on the North African representatives of the *A. ampeloprasum* polyploid complex including *A. baeticum* and *A. guttatum*. Novel diploid (2*x*) and tetraploid (4*x*) Algerian populations constitute a substantial enlargement of natural resources to investigate evolutionary relationships. The phylogenetic analyses emphasize two divergent diploid subgenomic pools within the North African group (NA), the NA_1 which is widely distributed from Algeria to Tunisia, including accessions referred to as *A. ampeloprasum* and *A. baeticum*, and the NA_2 restricted to Tunisia including accessions referred to as *A. ampeloprasum*. The hexa and octoploid GHG cultivars and the tetraploid *A. polyanthum*, clearly share subgenomes deriving from the NA_1 genomic pool. The tetraploid Leek group, *A. pyrenaicum* and *A. iranicum*, share a common origin with the tetraploid *A. truncatum* and the Tunisian diploids from the NA_2 genomic pool. All the Algerian diploids and tetraploids share a common ancestor, suggesting that the tetraploids likely arose from autopolyploidization events. The *A. guttatum* diploid and tetraploid populations were clearly distinguished from the NA group. The phylogeny of cpDNA highlighted the importance of gene flow and the continuum between North African and South European taxa. Data on the Algerian wild gene pool of *A. ampeloprasum* have led to extend understanding of the diversity within this polyploid complex and to make assumptions on the evolutionary history of the horticultural groups and uncultivated tetraploids. Future investigations using a much wider and appropriate sampling and involving genome wide based markers will be helpful to deepen our understanding of the diversification in this relevant group.

Acknowledgements

This work was funded by the international Project CMEP-Tassili Hubert Curien 08 MDU 724, "*Polyploidy, Genome Evolution and Biodiversity*", involving the Laboratory of Biology and Physiology of Organisms, University of Sciences and Technology Houari Boumediene (USTHB, Algiers, Algeria) and the UMR-CNRS 6553 Ecobio, University of Rennes 1 (France). Also, this work found support from the CNEPRU project F00220100043 and USTHB-internship grants.

Conflicts of interest

The authors declare no conflicts of interests.

References

Aïnouche ML, Bayer RJ (1997) On the origins of the tetraploid *Bromus* species (section *Bromus*, Poaceae): Insights from the internal transcribed spacer sequences of nuclear ribosomal DNA. Genome 40:730–743. https://doi.org/10.1139/g97-796

Alix K, Gérard PR, Schwarzacher T, Heslop-Harrison JSP (2017) Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. Annals of Botany 120:183–194. https://doi.org/10.1093/aob/mcx079

Álvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. Molecular Phylogenetic Evolution 29:417–434. https://doi.org/10.1016/S1055-7903(03)00208-2

Ariga T, Kumagai H, Yoshikawa M, Kawakami H, Seki T, Sakurai H et al (2002) Garlic-like but odorless plant *Allium ampeloprasum* 'Mushuu-ninniku'. Journal of the Japanese Society for Horticultural Science 71:362–369

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on Angiosperm phylogeny. Annals of the Missouri Botanical Garden 82(2):247–277. https://doi.org/10.2307/2399880

Battandier J, Trabut L. 1895. Flore de l'Algérie et Catalogue des plantes du Maroc, Monocotylédones. Adolphe Jourdan (Ed), Alger, pp 56–62

Battandier J, Trabut L (1902) Flore analytique et synoptique de l'Algérie et de la Tunisie. Vve Giralt, Libraire-Editeur, Alger, pp 331–332

Bohanec B, Jakše M, Sešek P, Havey MJ (2005) Genetic characterization of an unknown Chinese bulbous leek-like accession and its relationship to similar *Allium* species. HortScience 40:1690–1694. https://doi.org/10.21273/HORTSCI.40.6.1690

Bothmer R von (1970) Cytological studies in *Allium* I. Chromosome numbers and morphology in Sect. *Allium* from Greece. Botaniska Notiser 123:519–551

Bothmer R von (1974) Studies in the Aegean Flora XXI. Biosystematics studies in the *Allium ampeloprasum* complex. Opera Botanica (Lund) 34:1–104

Bothmer R von (1975) The *Allium ampeloprasum* complex on Crete. Mitteilungen der Botanischen Staatssammlung Munchen 12:267–288

Boulos L (2005) Flora of Egypt. Al Ahrara publishing, Cairo. Egypt, pp 63–83

Casazza G, Granato L, Minuto L, Conti L (2012) Polyploid evolution and Pleistocene glacial cycles: A case study from the alpine primrose *Primula marginata* (Primulaceae). BMC Evol Biol 12:56. https://doi.org/10.1186/1471-2148-12-56

De Wilde-Duyfjes BEE (1976) A revision of the genus *Allium* L. (Liliaceae) in Africa. Meded Landbouwhogeschool Wageningen 76:1–237

Demesure B, Sodzi N, Petit RJ (1995) A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol Ecol 4:129–131. https://doi.org/10.1111/j.1365-294X.1995.tb00201.x

Desfontaines R (1798) Flora Atlantica, sive, Historia plantarum, quae in Atlante, agro tunetano et algeriensi crescunt. Desgranges (ed). Paris, pp 285–291. http://dx.doi.org/10.5962/bhl.title.323

Dubouzet JG, Shinoda K (1999) Relationships among Old and New World Alliums according to ITS DNA sequence analysis. Theor Appl Genet 98:422–433. https://doi:10.1007/s001220051088

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. http://dx.doi.org/10.1093/nar/gkh340

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x

Figliuolo G, Candido V, Logozzo G, Miccolis V, Spagnoletti Zeuli PL (2001) Genetic evaluation of cultivated garlic germplasm (*Allium sativum* L. and *A. ampeloprasum* L.). Euphytica 121:325–334. https://doi.org/10.1023/A:1012069532157

Figliuolo G, Mang S (2010) Characterization and discrimination of Mediterranean bulb-producing garlic. In: Pacurar M, Krejci G (eds) Garlic Consumption and Health. Nova Science Publishers, Inc. New York, pp 181–197

Friesen N, Fritsch R, Blattner FR (2006) Phylogeny and new intrageneric classification of *Allium* (Alliaceae) based on nuclear ribosomal DNA its sequences. Aliso 22:372–395. https://doi.org/10.5642/aliso.20062201.31

Friesen N, Pollner S, Bachmann K, Blattner RF (1999) RAPDs and non-coding chloroplast DNA reveal a single origin of the cultivated *Allium fistulosum* from *A. altaicum* (Alliaceae). American Journal of Botany 86: 554–562. https://doi.org/10.2307/2656817

Fritsch RM, Blattner FR, Gurushidze M (2010) New classification of *Allium* L. subg. *Melanocrommyum* (Webb and Berthel) Rouy (Alliaceae) based on molecular and morphological characters. Phyton 49:145–220

Fritsch RM, Friesen N (2002) Evolution, domestication and taxonomy. In: *Allium crop science: recent advances*. Wallingford, CAB International. UK, pp 5–27

Gray D, Ward JA (1987) A comparison of leek (*Allium porrum*) and onion (*Allium cepa*) seed development. Annales of Botany 60:181–197. https://doi.org/10.1093/oxfordjournals.aob.a087435

Guenaoui C, Mang S, Figliuolo G, Neffati M (2013) Diversity in *Allium ampeloprasum*: from small and wild to large and cultivated. Genet Resour Crop Evol 60:97–114. https://doi.org/10.1007/s10722-012-9819-5

Guern M, Le Corff L, Boscher J (1991) Caryologie comparée des *Allium* du groupe *ampeloprasum* en France. Bull Soc Bot France 138:303–313. https://doi.org/10.1080/01811797.1991.10824932

Gurushidze M, Fritsch RM, Blattner FR (2010) Species-level phylogeny of *Allium* subgenus *Melanocrommyum*: incomplete lineage sorting, hybridization and *trnF* gene duplication. Taxon 59:829–840. https://doi.org/10.1002/tax.593012

Gurushidze M, Fritsch RM, Blattner RF (2008) Phylogenetic analysis of *Allium* subg. *Melanocrommyum* infers cryptic species and demands a new sectional classification. Mol Phylogenet Evol 49:997–1007. https://doi.org/10.1016/j.ympev.2008.09.003

Gurushidze M, Mashayekhi S, Blattner FR, Friesen N, Fritsch RM (2007) Phylogenetic relationships of wild and cultivated species of *Allium* section *Cepa* inferred by nuclear rDNA ITS sequence analysis. Plant Syst Evol 269:259–269. https://doi.org/10.1007/s00606-007-0596-0

Hall AT (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95-98

Han TS, Zheng QJ, Onstein RE, Rojas-Andrès BM, Hauenschild F, Muellner-Riehl AN, Xing YW (2019) Polyploidy promotes species diversification of *Allium* through ecological shifts. New Phytol. Issue online 01 December 2019. https://doi.org/10.1111/nph.16098

Harrison CJ, Langdale JA (2006) A step by step guide to phylogeny reconstruction. The Plant Journal 45(4):561–572. https://doi.org/10.1111/j.1365-313X.2005.02611.x

Herden T, Hanelt P, Friesen N (2016) Phylogeny of *Allium* L. subgenus *Anguinum* (G. Don. ex W.D.J. Koch) N. Friesen (Amaryllidaceae). Mol Phylogenet Evol 95:79–93. https://doi.org/10.1016/j.ympev.2015.11.004

Hirschegger P, Galmarini C, Bohanec B (2006) Characterization of a novel form of fertile great headed garlic (*Allium sp.*). Plant Breeding 125:635–637. https://doi.org/10.1111/j.1439-0523.2006.01279.x

Hirschegger P, Jakše J, Trontelj P, Bohanec B (2010) Origins of *Allium ampeloprasum* horticultural groups and a molecular phylogeny of the section *Allium* (*Allium*: Alliaceae). Mol Phylogenet Evol 54:488–497. https://doi.org/10.1016/j.ympev.2009.08.030

Jauzein P, Tison J (2005) Le complexe d'*Allium ampeloprasum* L. Lejeunia 178:457–4184

Karavokyrou E, Tzanoudakis D (1991) The genus *Allium* in Greece: II. A cytogeographical study of the E Aegean species. Botanical Chronic 10: 777–784

Khazanehdari KA, Jones GH (1997) The causes and consequences of meiotic irregularity in the leek (*Allium ampeloprasum* spp. *porrum*); implications for fertility, quality and uniformity. Euphytica 93:313–319. https://doi.org/10.1023/A:1002914808150

Khedim T, Amirouche N, Amirouche R (2010) Caractérisation cytotaxonomique du genre *Allium* (Amaryllidaceae) en Algérie. Ann Inst Nat Agron (Alger) 31:17–59

Khedim T, Amirouche N, Amirouche R (2013) Biosystematic and Plants Genetic Resources: Example from the genus *Allium* (Amaryllidaceae) of the Algerian flora. Acta Hortic 997:19–23. doi.org/10.17660/ActaHortic.2013.997.1

Khedim T, Amirouche N, Amirouche R (2016) Morphological and cytotaxonomic data of *Allium trichocnemis* and *A. seirotrichum* (Amaryllidaceae) endemic to northern Algeria compared with *A. cupanii* group. Phytotaxa 243:247–259. http://dx.doi.org/10.11646/phytotaxa.243.3.3

Kik C, Samoylov AM, Verbeek WHJ, Van Raamsdonk LWD (1997) Mitochondrial DNA variation and crossability of leek (*Allium porrum*) and its wild relatives from the *Allium ampeloprasum* complex. Theor Appl Genet 94:465–471. https://doi.org/10.1007/s001220050438

Kollmann F (1972) *Allium ampeloprasum* -A polyploid complex II. Meiosis and relationships between the ploidy types. Caryologia 2:295–312. https://doi.org/10.1080/00087114.1972.10796484

Kovarik A, Pires JC, Leitch AR, Lim KY, Sherwood AM, Matyasek R et al (2005) Rapid concerted evolution of nuclear ribosomal DNA in two Tragopogon allopolyploids of recent and recurrent origin. Genetics 169:931–944. https://doi.org/10.1534/genetics.104.032839

Levan A (1940) Meiosis in *Allium porrum*, a tetraploid species with chiasma localization. Hereditas 26:454–462. https://doi.org/10.1111/j.1601-5223.1940.tb03248.x

Li QQ, Zhou SD, He XJ, Yu Y, Zhang YC, Wei XQ (2010) Phylogeny and biogeography of *Allium* (Amaryllidaceae: Allieae) based on nuclear ribosomal internal transcribed spacer and chloroplast rps16 sequences, focusing on the inclusion of species endemic to China. Annals of Botany 106:709–733. https://doi.org/10.1093/aob/mcq177

Li QQ, Zhou SD, Huang DQ, He XJ, Wei XQ (2016) Molecular phylogeny, divergence time estimates and historical biogeography within one of the world's largest monocot genera. *AoB PLANTS* 8 (1):1–17. https://doi.org/10.1093/aobpla/plw041

Liu ZP, Chen ZY, Pan J, Li XF, Su M, Wang LJ et al (2008) Phylogenetic relationships in *Leymus* (Poaceae: Triticeae) revealed by the nuclear ribosomal internal transcribed spacer and chloroplast *trnL-F* sequences. Mol Phylogenet Evol 46:278–289. https://doi.org/10.1016/j.ympev.2007.10.009

Logacheva MD, Valiejo-Roman CM, Degtjareva GV, Stratton JM, Downie SR, Samigullin TH et al (2010) A comparison of nrDNA ITS and ETS loci for phylogenetic inference in the Umbelliferae: an example from tribe Tordylieae. Mol Phylogenet Evol 57:471–6. https://doi.org/10.1016/j.ympev.2010.06.001

Maire R (1958) Flore de l'Afrique du Nord. Vol V, Paul Lechevalier Edition, Paris, pp 244–303

Maragheh FP, Janus D, Senderowicz M, Haliloglu K, Kolano B (2018) Karyotype analysis of eight cultivated *Allium* species. Journal of Applied Genetics 60:1–11. https://doi.org/10.1007/s13353-018-0474-1

Mathew B (1996) A review of *Allium* section *Allium*. Royal Botanical Garden, Kew, UK

Médail F, Diadema K (2009) Glacial refugia influence plant diversity patterns in the Mediterranean Basin. Journal of Biogeography 36:1333–1345. https://doi.org/10.1111/j.1365-2699.2008.02051.x

Nguyen N, Dricall H, Speacht C (2008) A molecular phylogeny of the wild onion (*Allium,* Alliaceae*)* with a focus on the western North American center of diversity. Mol Phylogenet Evol 10:1–16. https://doi.org/10.1016/j.ympev.2007.12.006

Pastor J (1982) Karyology of *Allium* species from Iberian Peninsula. Phyton 22(2):171–200

Peruzzi L, Carta A, Altinordu F (2017) Chromosome diversity and evolution in *Allium* (Amaryllidaceae, Allioideae). Plant Biosystems 151:212–220. https://doi.org/10.1080/11263 504.2016.1149123

Poczai P, Hyvönen J (2010) Nuclear ribosomal spacer regions in plant phylogenetics: problems and prospects. Molecular Biology Reports 37(4):1897–912. https://doi.org/10.1007/s11033-009-9630-3

Quézel P, Santa S (1962) Nouvelle flore de l'Algérie et des régions désertiques et méridionales. Tome 1, Edition du CNRS, Paris, pp 209–213

Sinitsyna TA, Herden T, Friesen N (2016) Dated phylogeny and biogeography of the Eurasian *Allium* section *Rhizirideum* (Amaryllidaceae). Plant Syst Evol 302:1311–1328. https://doi:10.1007/s00606-016-1333-3

Soltis DE, Mavrodiev EV, Doyle JJ, Rauscher J, Soltis PS (2008) ITS and ETS sequence data and phylogeny reconstruction in allopolyploids and hybrids. Systematic Botany 33:7–20. https://doi.org/10.1600/036364408783887401

Stack SM, Roelofs D (1996) Localized chiasmata and meiotic nodules in the tetraploid onion *Allium porrum*. Genome 39: 770–783. https://doi.org/10.1139/g96-097

Stearn WT (1980) *Allium* L. In: Tutin et al (eds), Flora Europaea, Vol 5. Cambridge University Press, pp 49–69

Stewart PH (1975) Un nouveau climagramme pour l'Algérie et son application au barrage vert. Bull Soc Hist Nat Afrique du Nord 65:239–252

Suárez-Santiago VN, Salinas MJ, Garcia-Jacas N, Soltis PS, Soltis DE, Blanca G (2007) Reticulate evolution in the *Acrolophus* subgroup (*Centaurea* L., Compositae) from the western Mediterranean: Origin and diversification of section *Willkommia* Blanca. Mol Phylogenet Evol 43:156–172. https://doi.org/10.1016/j.ympev.2006.08.006

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. Mol Ecol 7:453–464. https://doi.org/10.1046/j.1365-294x.1998.00289.x

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology 17:1105–1109. https://doi.org/10.1007/BF00037152

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA 6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol 30:2725–2729. https://doi:10.1093/molbev/mst197

Tzanoudakis D, Vosa C (1988) The cytogeographical distribution pattern of *Allium* (Alliaceae) in the Greek peninsula and Islands. Plant Syst Evol 159:193–215. https://doi.org/10.1007/BF00935972

Valdès CB, Rejdali M, Achal el kadmiri A, Jury SL, Monserat M, José M (2002) Catalogue des plantes vasculaires du nord du Maroc, incluant des clés d'identification. Vol 2, CSIC (ed). Madrid, pp 865–870

Veiskarami GH, Khodayari H, Heubl G, Weigend M, Zarre S (2019) Phylogenetic relationships of Iranian *Allium* sect. *Allium* (Amaryllidaceae, Allioideae) as inferred from nrDNA ITS, cpDNA, *rps16* and *trn*L–F sequences. Nordic Journal of Botany 37(7): e02109. https://doi.org/10.1111/njb.02109

Véla E, Benhouhou S (2007) Évaluation d'un nouveau point chaud de biodiversité végétale dans le Bassin méditerranéen (Afrique du Nord). Comptes Rendus Biologies 330:589–605

Wendel JF, Schnabel A, Seelanan T (1995) Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc Nat Acad Sciences USA 92:280–284. https://doi.org/10.2307/2366544

Wendel JF (2000) Genome evolution in polyploids. Plant Molecular Biology 42:225–249. https://doi.org/10.1023/A:1006392424384

Wheeler EJ, Mashayekhi S, Mcneal DW, Columbus JT, Pires JC (2013) Molecular systematics of *Allium* subgenus *Amerallium* (Amaryllidaceae) in North America. American Journal of Botany 100:701–711. https://doi.org/10.3732/ajb.1200641

Titles of the Figures

Fig. 1 Morphological traits, inflorescence and bulb with outer tunics of some sampled populations. *A. ampeloprasum* (**a-d**) *diploids*, (**e-h**) tetraploids, *A. baeticum* (**i-j**) diploids, (**k-l**) tetraploids, *A. guttatum* (**m-p**) diploids. Correspondence with varieties from Maire (1958). (**a-b)** *A. ampeloprasum* cf. var. *typicum* Regel, (**c-d**) cf. var. *duriaeanum* (Gay) Batt., (**e-f)** cf. var. *tortifolium* Batt., (**g-h)** cf. var. *getulum* Batt., (**i-l**) *A. baeticum* cf. var. *laeve* Maire and Weiller, (**m-n**) *A. guttatum* cf. var. *battandieri* Maire and Weiller, (**o-p**), cf. var. *typicum* Regel

Fig. 2 Somatic metaphases of some sampled populations. Diploids (2n = 2*x* = 16): (**a)** *A. ampeloprasum* Bouzegza, (**b**) Chréa, (**c**), Tinekachine, (**d**), Bouhciène, (**e**) *A. baeticum* Redjredj, (**f**) *A. guttatum* Chréa. Tetraploids (2n = 4*x* = 32), (**g**) *A. ampeloprasum* Bouzeghaia, (**h**) El Mesrane, (**i**) Tizi Ouchir, (**j**) Matmata, (**k**) *A. baeticum* Cap Ténès, (**l**) *A. guttatum* Tazrout. Abbreviations: Sp, proximal satellite; St, terminal satellite. Scale bar = 10 µm

Fig. 3 Phylogenetic tree resulting from Maximum Likelihood analysis based on *ITS* sequences of diploids *A. ampeloprasum* and related taxa within section *Allium*. Numbers by nodes represent maximum Likelihood bootstrap supports > 50% (1000 replicates). Solid symbols are used for diploids

Fig. 4 Condensed phylogenetic tree resulting from a Maximum Likelihood analysis based on the *ITS* sequences of diploids, polyploids *A. ampeloprasum* complex and related taxa within section *Allium* (right). Focus on the NA group (lift). Numbers by nodes represent maximum Likelihood bootstrap supports > 50% (1000 replicates). Solid symbols are used for diploids and empty ones for polyploids. The scale bar indicates a branch length of 0.001 substitutions per site. Clades of amplicons (G1, G2 and G3) from Hirshengger et al. (2010) are shown. **G1**, *A. ampeloprasum* var. *holmense* GHG (6*x*, 8*x*) + *A. ampeloprasum* var. *babingtonii* + *A. pyrenaicum* (4*x*). **G2**, *A. ampeloprasum* var. *holmense* GHG (6*x*, 8*x*) + *A. ampeloprasum* var. *babingtonii* + *A. polyanthum* (4*x*). **G3**, *A. ampeloprasum* var. *holmense* GHG (8*x*) + *A. ampeloprasum* Leek group (Leek, Pearl onion, Bulbous leek and Kurrat) (4*x*) + *A. pyrenaicum* (4*x*) + *A. truncatum* (4*x*) + *A. iranicum* (4*x*)

Fig. 5 Phylogenetic tree resulting from a Maximum Likelihood analysis based on combined plastid DNA *trn L−trnF* and *trnD−trnT* regions of diploids, polyploids *A. ampeloprasum* complex and related taxa within section *Allium*. Numbers by nodes represent Maximum Likelihood bootstrap supports > 50% (1000 replicates). Solid symbols are used for diploids and empty ones for polyploids

Fig. 1

Fig. 2

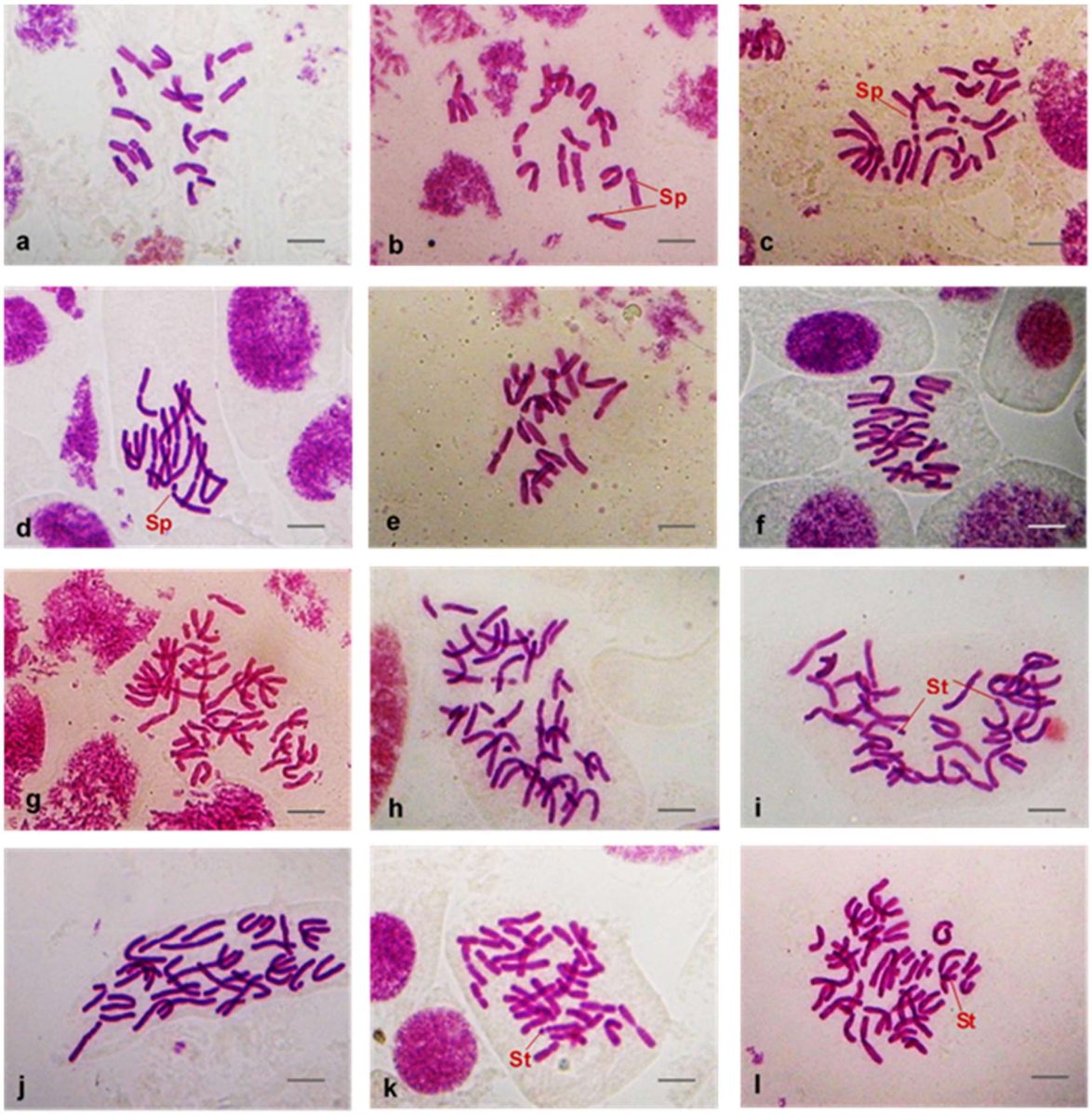


Fig. 3

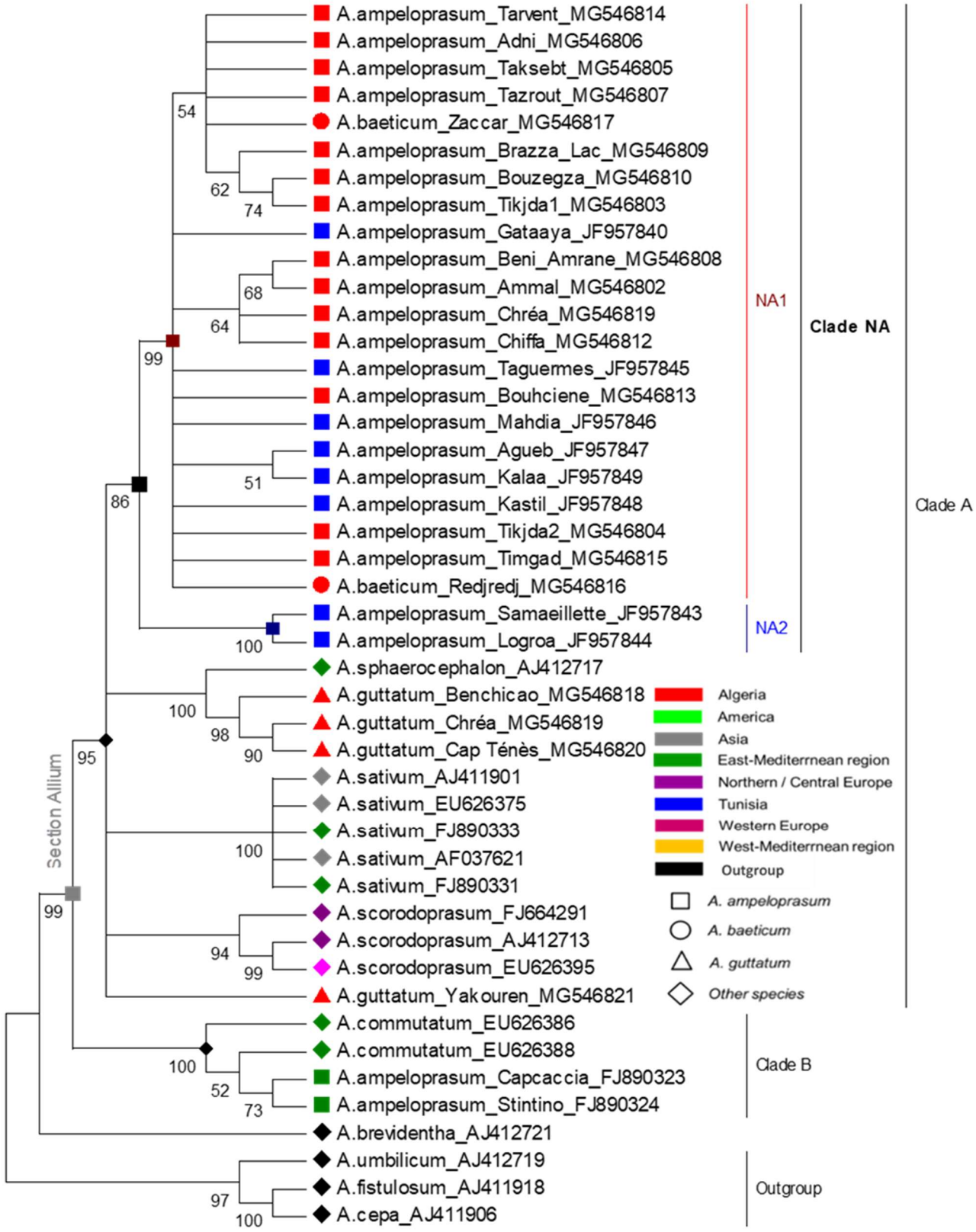

A.ampeloprasum_Tarvent_MG546814
A.ampeloprasum_Adni_MG546806
A.ampeloprasum_Taksebt_MG546805
A.ampeloprasum_Tazrout_MG546807
A.baeticum_Zaccar_MG546817
A.ampeloprasum_Brazza_Lac_MG546809
A.ampeloprasum_Bouzegza_MG546810
A.ampeloprasum_Tikjda1_MG546803
A.ampeloprasum_Gataaya_JF957840
A.ampeloprasum_Beni_Amrane_MG546808
A.ampeloprasum_Ammal_MG546802
A.ampeloprasum_Chréa_MG546819
A.ampeloprasum_Chiffa_MG546812
A.ampeloprasum_Taguermes_JF957845
A.ampeloprasum_Bouhciene_MG546813
A.ampeloprasum_Mahdia_JF957846
A.ampeloprasum_Agueb_JF957847
A.ampeloprasum_Kalaa_JF957849
A.ampeloprasum_Kastil_JF957848
A.ampeloprasum_Tikjda2_MG546804
A.ampeloprasum_Timgad_MG546815
A.baeticum_Redjredj_MG546816
A.ampeloprasum_Samaeillette_JF957843
A.ampeloprasum_Logroa_JF957844
A.sphaerocephalon_AJ412717
A.guttatum_Benchicao_MG546818
A.guttatum_Chréa_MG546819
A.guttatum_Cap Ténès_MG546820
A.sativum_AJ411901
A.sativum_EU626375
A.sativum_FJ890333
A.sativum_AF037621
A.sativum_FJ890331
A.scorodoprasum_FJ664291
A.scorodoprasum_AJ412713
A.scorodoprasum_EU626395
A.guttatum_Yakouren_MG546821
A.commutatum_EU626386
A.commutatum_EU626388
A.ampeloprasum_Capcaccia_FJ890323
A.ampeloprasum_Stintino_FJ890324
A.brevidentha_AJ412721
A.umbilicum_AJ412719
A.fistulosum_AJ411918
A.cepa_AJ411906
54
62
74
68
64
99
51
86
100
100
98
90
95
100
99
94
99
100
52
73
97
100
NA1
NA2
Clade NA
Clade A
Clade B
Outgroup
Section Allium
Algeria
America
Asia
East-Mediterrnean region
Northern / Central Europe
Tunisia
Western Europe
West-Mediterrnean region
Outgroup
A. ampeloprasum
A. baeticum
A. guttatum
Other species

Fig.4

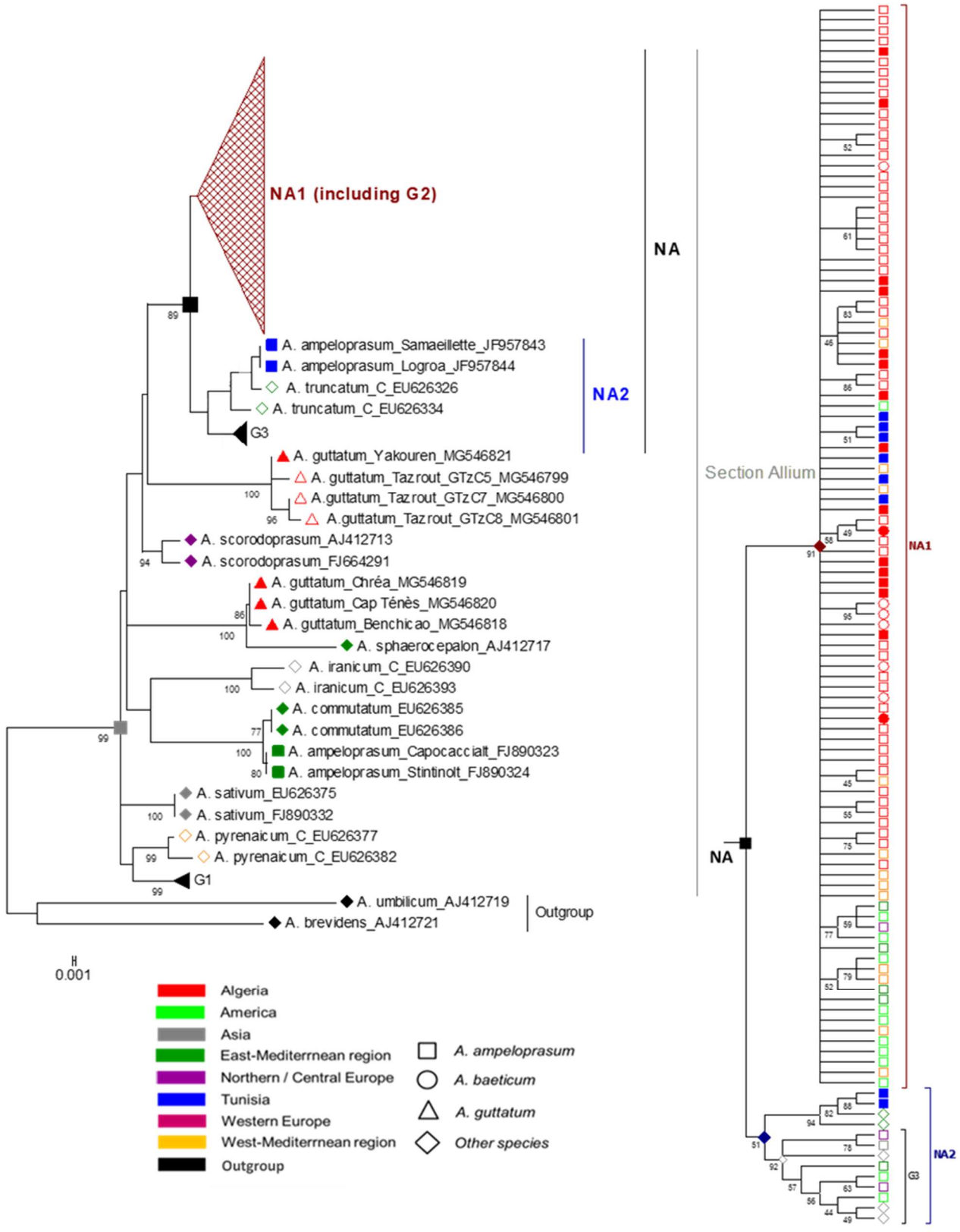

NA1 (including G2)
NA
NA2
A. ampeloprasum_Samaeillette_JF957843
A. ampeloprasum_Logroa_JF957844
A. truncatum_C_EU626326
A. truncatum_C_EU626334
G3
A. guttatum_Yakouren_MG546821
A. guttatum_Tazrout_GTzC5_MG546799
A.guttatum_Tazrout_GTzC7_MG546800
A.guttatum_Tazrout_GTzC8_MG546801
A. scorodoprasum_AJ412713
A. scorodoprasum_FJ664291
A. guttatum_Chréa_MG546819
A. guttatum_Cap Ténès_MG546820
A. guttatum_Benchicao_MG546818
A. sphaerocepalon_AJ412717
A. iranicum_C_EU626390
A. iranicum_C_EU626393
A. commutatum_EU626385
A. commutatum_EU626386
A. ampeloprasum_Capocaccialt_FJ890323
A. ampeloprasum_Stintinolt_FJ890324
A. sativum_EU626375
A. sativum_FJ890332
A. pyrenaicum_C_EU626377
A. pyrenaicum_C_EU626382
G1
A. umbilicum_AJ412719
A. brevidens_AJ412721
Outgroup
0.001
Algeria
America
Asia
East-Mediterrnean region
Northern / Central Europe
Tunisia
Western Europe
West-Mediterrnean region
Outgroup
A. ampeloprasum
A. baeticum
A. guttatum
Other species
Section Allium
NA
NA1
NA2
G3

Fig.5


Algeria
America
Asia
East-Mediterrnean region
Northern / Central Europe
Tunisia
Western Europe
West-Mediterrnean region
Outgroup
A. ampeloprasum
A. baeticum
A. guttatum
Other species
A.ampeloprasum_Tarvent_MG546700/MG546727
A.ampeloprasum_Cap Ténès_MG546707/MG546734
A.ampeloprasum_Tinekachine_MG546699/MG546726
A.ampeloprasum_Brazza_Lac_MG546693/MG546720
A.ampeloprasum_Chiffa_MG546696/MG546723
A.baeticum_Msila_MG546711/MG546738
A.baeticum_Cap Ténès_MG546713/MG546740
A.ampeloprasum_Beni_Amrane_MG546692/MG546719
A.ampeloprasum_Bouhciene_MG546698/MG546725
A.ampeloprasum_Taksebt_MG546708/MG546735
A.ampeloprasum_Adni_MG546709/MG546736
A.ampeloprasum_Tizi-Ouchir_MG546691/MG546718
A.ampeloprasum_Ouled_Farès_MG546706/MG546733
A.ampeloprasum_Tazrout_MG546710/MG546737
A.ampeloprasum_Chréa_MG546695/MG546722
A.ampeloprasum_Matmata_MG546701/MG546728
A.ampeloprasum_Tikjda1_MG546703/MG546730
A.ampeloprasum_Bouzegza_MG546694/MG546721
A.ampeloprasum_Bouzeghaia_MG546705/MG546732
A.ampeloprasum_Tikjda2_MG546704/MG546731
A.baeticum_Redjredj_MG546712/MG546739
A.pyrrenaicum_EU626257/EU626284
A.ampeloprasum_Timgad_MG546702/MG546729
A.acutiflorum_EU626256/EU626282
A.iranicum_EU626242/EU626279
A.truncatum_EU626244/EU626281
A.atroviolaceum_EU626241/EU626283
A.ampeloprasum_PO4x_EU626238/EU626272
A.ampeloprasum_BL4x_EU626243/EU626270
A.ampeloprasum_Kurrat4x_EU626240/EU626269
A.ampeloprasum_Leek4x_EU626239/EU626271
A.ampeloprasum_Kristel_MG546697/MG546724
A.polyanthum_EU626250/EU626280
A.ampeloprasum_GHG6xSpain9_EU626245/EU626267
A.ampeloprasum_GHG8xFCA_EU626247/EU626268
A.ampeloprasum_GHG8x1007_EU626251/EU626264
A.ampeloprasum_GHG6xPazin_EU626248/EU626266
A.ampeloprasum_varbab_EU626246/EU626265
A.ampeloprasum_GHG6xall1355_EU626249/EU626289
A.guttatum_Benchicao_MG546715/MG546742
A.gutatum_Ténès_MG546716/MG546743
A.guttatum_Tazrout_MG546717/MG546744
A.commutatum_EU626255/EU626287
A.commutatum_EU626253/EU626285
A.commutatum_EU626254/EU626286
A.bourgeaui_EU626252/EU626288
A.tuncelianum_EU626262/EU626278
A.scorodoprasum_FJ628600/FJ628605
A.scorodoprasum_FJ628601/FJ628606
A.scorodoprasum_EU626263/EU626274
A.leucanthum_EU626259/EU626273
A.leucanthum_EU626260/EU626275
A.pseudoampeloprasum_EU626258/EU626276
A.sativum_EU626261/EU626277
A.capa_FJ628603/FJ628608
A.fistulosum_FJ628604/FJ628609
Clade I
Clade II
Outgroup
Section Allium

Electronic Supplementary Material

ESM 1 Table S1 Geographic location, elevation and bioclimatic range of the 31 sampling sites in Algeria. Bioclimat (*) from Stewart (1975): A, Arid; H, humid; SA, semiarid; SH, subhumid. Abbreviations: AA: *A. ampeloprasum*; AB, *A. baeticum*; AG, *A. guttatum*

ESM 2 Table S2 Detailed information of the Algerian populations used for chromosome counting and DNA sequencing, and references of all analyzed taxa included from other studies

ESM 3 Table S3 Summary of phylogenetic parameters from ML, MP analysis and model test of separate and combined datasets, based on *ITS*-rDNA and cpDNA (*trnL−trnF*/*trnD−trnT*) sequences. Abbreviations: BIC, Bayes information criterion; CI, consistency index; RI, retention index

ESM 4 Fig. S1 Phylogenetic tree resulting from a Maximum Likelihood analysis based on a broad dataset of *ITS* sequences of diploid and polyploid accessions from the *A. ampeloprasum* complex and from related taxa within section *Allium*. Included are *ITS* sequences generated from this study and from GenBank. Numbers on nodes represent maximum Likelihood bootstrap supports (1000 replicates). Solid symbols are used for diploids and empty ones for polyploids

ESM 5 Fig. S2 Summary of the evolutionary relationships in the *A. ampeloprasum* complex and origins of polyploids (redrawn from *ITS* phylogenetic analyses from this study and Hirschegger et al. (2010)). Major lineages sampled among the extant diploid species in the complex are represented by black triangles in the tree. In the absence of diploid representatives (not sampled or extinct), ribotypes representing subgenomes of the allotetraploids *A. pyrenaicum* and *A. iranicum* are also indicated in grey triangles. All taxa are boxed in shaded rectangles. The lines connecting the shaded boxes (dotted, dashed or solid according to their respective ploidy level (4*x*, 6*x* and 8*x*) indicate the sharing of genomic markers (ribotypes) contributing to polyploid genomes. Abbreviations: NA_1, North African clade 1 (North Algeria and Tunisia); NA_2, North African clade 2 (South Tunisia)

Supplementary Table S1 Geographic location, elevation and bioclimatic range of the 31 sampling sites in Algeria

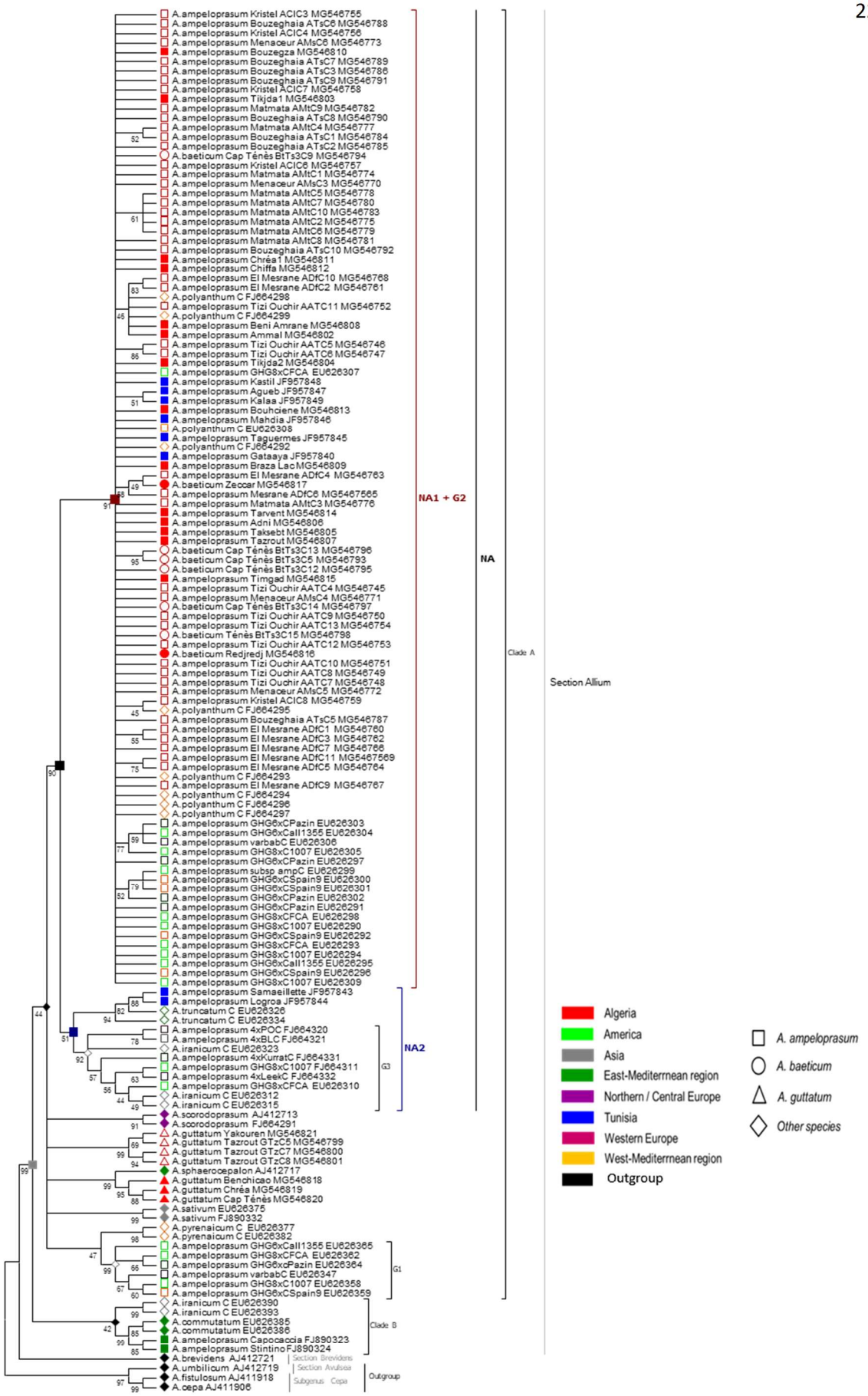
Site	Locality	Latitude	Longitude	Elevation in meters	Bioclimat [a]	Sampled species [b]
Adni	Irdjen	36°40′00″N	04°09′00″E	720	H	*AA, AG*
Ammal	Lakhdaria	36°37′10″N	03°32′43″E	650	SH	*AA*
Benchicao	Medéa	36°13′56.2″N	02°52′27.2″E	1092	H	*AG*
Beni Amrane	Thénia	36°40′00″N	03°35′30″E	294	SH	*AA*
Bouhcène	Dellys	36.8°48.8′00″N	04°33′18″E	790	SH,	*AA*
Bouzeghaia	Zeboudja	36°19′28.8″N	01°14′41.7″E	201	SA	*AA*
Bouzegza	Keddara	36°37′31″N	03°28′46″E	600	H	*AA*
Brazza Lac	Zoubiria	36°04′00″N	02°56′50″E	870	SH	*AA*
Cap Ténès	Ténès	36°30′44″N	01°18′16″E	120	SH	*AA, AB, AG*
Chiffa	Blida	36°27′22″N	02°43′06″E	300	SH	*AA*
Chréa 1	National Park of Chréa	36°25′239″N	02°53′54″E	1020	H	*AA*
Chréa 2	National Park of Chréa	36.4°23′23.6″N	02°52′66″E	1500	H	*AG*
El Mesrane	Djelfa	34°45′59.3″N	03°10′26.2″E	800	A	*AA*
Gelt Estel	Djelfa	35°08′59.0″N	03°01′59.0″E	700	A	AB
Kristel	Gdyel	35°46′38.8″N	00°30′48.8″W	49	SA	*AA*
Matmata	Oued Djemaâ	35°48′00″N	00°41′00″E	650	SA	*AA*
Menaceur	Hadjout	36°30′10.5″N	02°15′59.9″E	133	SH	*AA*
Moudjbar	Ksar El Boukhari	35°56′7.1″N	02°46′23.8″E	410	SA	*AB*
Mont Zaccar	Miliana	36°18′00″N	02°16′60″E	500	SH	*AB*
Msila	Boutlélis	35°36′52.1″N	00°53′43.3″W	343	SH	*AB*
Ouled Farès	Ténès	36°14′32.7″N	01°14′17.4″E	139	SA	*AA*
Redjredj	Medéa	36°05′10.6″N	02°57′51.1″E	1075	SH	*AB*
Taksebt	Takhoukht	36°39′00″N	04°10′00″E	450	SH	*AA*
Tarvent	Agouni Bouaklane	36.8°60′37″N	04°05.6′08″E	810	SH	*AA*
Tazrout	Semghoune	36°49′34.3″N	04°04′31″E	900	SH	*AG*
Tikjda 1	National Park of Djurdjura	36°27′43.4″N	04°09′1.8″E	1684	H	*AA*
Tikjda 2	National Park of Djurdjura	36°27′44.6″N	04°10′5.4″E	1800	H	*AA*
Timgad	Batna	35°30′21″N	06°28′00″E	1069	SA	*AA*
Tinekachine	Makouda	36°47′1.7″N	04°01′48″E	700	SH	*AA*
Tizi Ouchir	Ain Torki	36°19′60″N	02°18′09″E	884	H	*AA*
Yakouren	Akfadou	36°44′05″N	04°26′19″E	660	H	*AG*

[a] Bioclimat from Stewart (1975) A: Arid, H: humid, SA: semiarid, SH: subhumid
AA: *A. ampeloprasum,* AB: *A. baeticum,* AG: *A. guttatum*

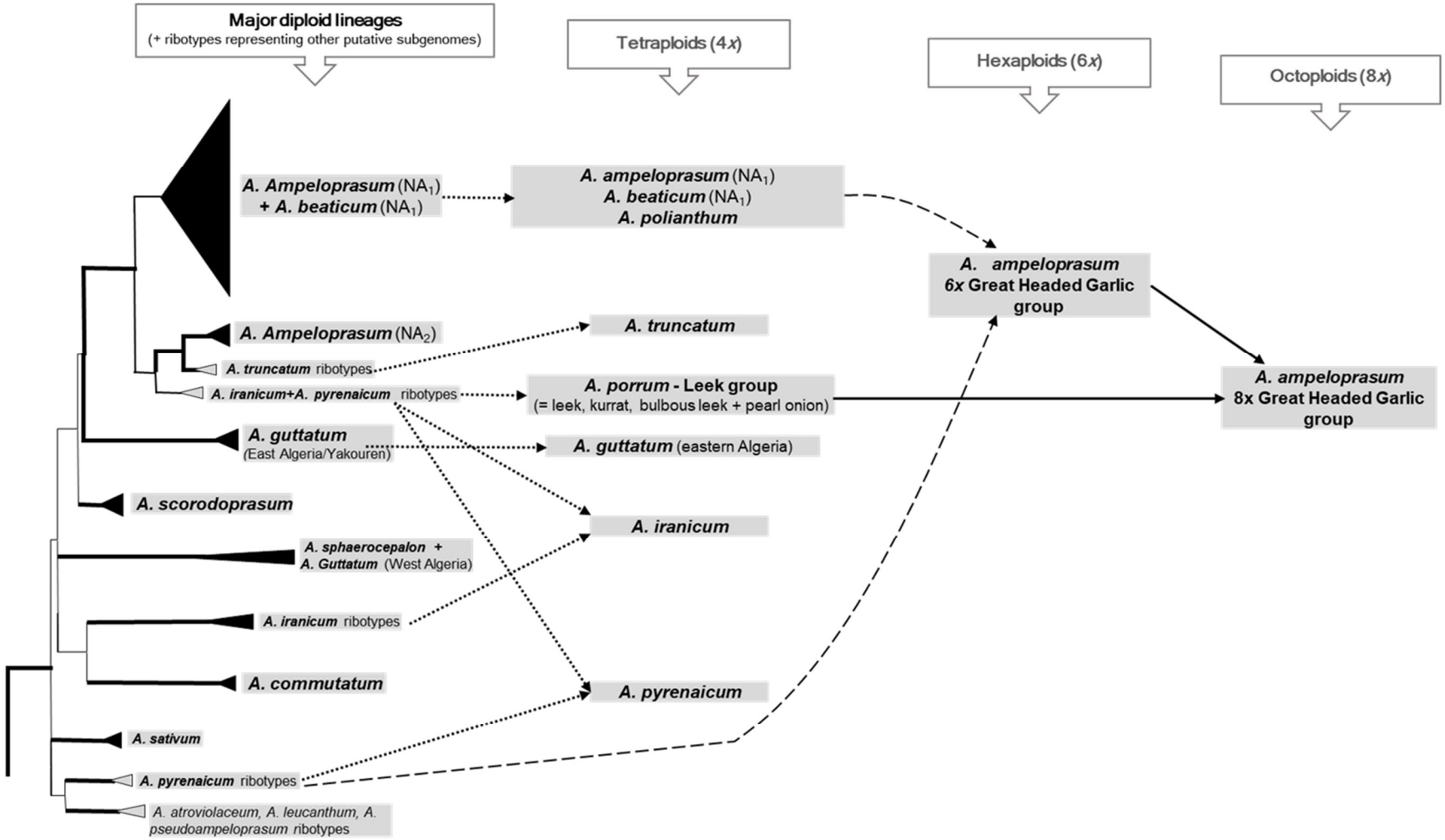
Supplementary Table S3 Summary of phylogenetic parameters from ML, MP analysis and model test of separate and combined datasets, based on *ITS*-rDNA and cpDNA *(trnL-trnF/trnD-trnT*) sequences

Parameters	*ITS*–2*x*	*ITS*–4*x*, 5*x*, 6*x*, 8*x*	*trnL–trnF*	*trnD–trnT*	Combined *trnL–trnF* + *trnD–trnT*
No. of included taxa	11	13	17	17	17
No. of included accessions	45	160	60	57	56
No. of included characters	617	620	928	574	1388
No. of variable sites	216	303	181	82	188
No. of potentially parsimony-informative sites	146	182	54	45	79
No. of trees obtained by unweighted MP	7	3	1	2	1
No. of steps for unweighted MP (tree length)	279	491	128	228	245
CI	0.7397	0.5374	0.6200	0.6138	0.6343
RI	0.9068	0.8840	0.8812	0.8368	0.8689
Nucleotide substitution model selected by lowest BIC	K2+I	T92+G	T92+G	T92+G	T92+G

BIC, Bayes information criterion; CI, consistency index; RI, retention index



Supplementary Fig. S1 Phylogenetic tree resulting from a Maximum Likelihood analysis based on a broad dataset o *ITS* sequences of diploid and polyploid accessions from the *A. ampeloprasum* complex and from related taxa withi section *Allium*. Included are *ITS* sequences generated from this study and from genbank. Numbers on nodes represen maximum Likelihood bootstrap supports (1000 replicates). Solid symbols are used for diploids and empty ones fo polyploids



Supplementary Fig. S2 Summary of the evolutionary relationships in the *A. ampeloprasum* complex and origins of polyploids (redrawn from *ITS* phylogenetic analyses from this study and Hirschegger et al. (2010)). Major lineages sampled among the extant diploid species in the complex are represented by black triangles in the tree. In the absence of diploid representatives (not sampled or extinct), ribotypes representing subgenomes of the allotetraploids *A. pyrenaicum* and *A. iranicum* are also indicated in grey triangles. All taxa are boxed in shaded rectangles. The lines connecting the shaded boxes (dotted, dashed or solid according to their respective ploidy level (4*x*, 6*x* and 8*x*) indicate the sharing of genomic markers (ribotypes) contributing to polyploid genomes. Abbreviations: NA1, North African clade 1 (North Algeria and Tunisia); NA2, North African clade 2 (South Tunisia).