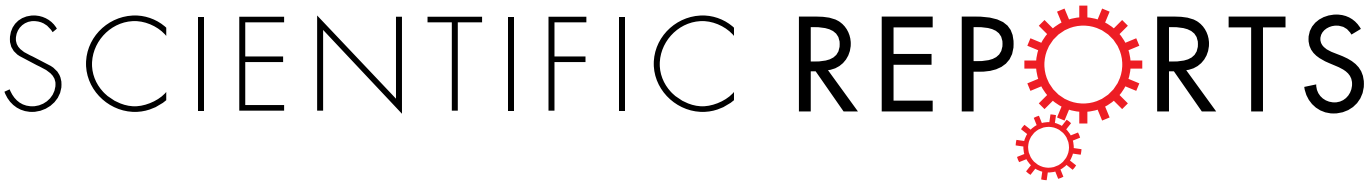


OPEN

Unraveling the evolutionary dynamics of ancient and recent polypoidization events in *Avena* (Poaceae)

Received: 14 July 2016
Accepted: 30 December 2016
Published: 03 February 2017

Qing Liu[1], Lei Lin[1,2], Xiangying Zhou[1,2], Paul M. Peterson[3] & Jun Wen[3]

Understanding the diversification of polyploid crops in the circum-Mediterranean region is a challenging issue in evolutionary biology. Sequence data of three nuclear genes and three plastid DNA fragments from 109 accessions of *Avena* L. (Poaceae) and the outgroups were used for maximum likelihood and Bayesian analyses. The evolution of cultivated oat (*Avena sativa* L.) and its close relatives was inferred to have involved ancient allotetraploidy and subsequent recent allohexaploidy events. The crown ages of two infrageneric lineages (*Avena* sect. *Ventricosa* Baum ex Romero-Zarco and *Avena* sect. *Avena*) were estimated to be in the early to middle Miocene, and the *A. sativa* lineages were dated to the late Miocene to Pliocene. These periods coincided with the mild seasonal climatic contrasts and the Mediterranean climate established in the Mediterranean Basin. Our results suggest that polyploidy, lineage divergence, and complex reticulate evolution have occurred in *Avena*, exemplifying the long-term persistence of tetraploids and the multiple origins of hexaploids related to paleoclimatic oscillations during the Miocene-Pliocene interval in the circum-Mediterranean region. This newly-resolved infrageneric phylogenetic framework represents a major step forward in understanding the origin of the cultivated oat.

Genome duplication following hybridization (allopolyploidy) is common among flowering plants, and is found in nearly a quarter of Poaceae that provide crops and fuel worldwide[1]. Phylogenetic evidence from nuclear loci has accumulated to identify allopolyploidy events because they produce characteristic double-labelled phylograms in which allopolyploids appear more than once[2]. This approach does require sufficient depth of sequencing and the identification of paralogues produced by gene duplication events[3].

The genus *Avena* L. (Poaceae) contains ca. 29 species exhibiting considerable morphological and ecological diversity in the Mediterranean Basin, Eastern Africa, Europe, Asia, and the Americas[4,5]. Based on glume shape, lemma apex, and the insertion of lemmatal awn, seven sections have been recognized for *Avena*: *Avenotrichon* (Holub) Baum, *Ventricosa* Baum ex Romero-Zarco, *Agraria* Baum, *Tenuicarpa* Baum, *Ethiopica* Baum, *Pachycarpa* Baum, and *Avena*[6]. The genus forms a polyploid series ranging from A- and C-genome diploids ($2x = 14$), AB- and A'C (DC)-genome tetraploids ($4x = 28$), to ACD-genome hexaploids ($6x = 42$)[7]. The A- and C-genome diploids are distinguished by the structural differentiation of isobrachial and heterobrachial chromosomes[8], while the B and D genomes are not found in any extant diploids[9,10].

Molecular data support a close relationship between D and A genomes[11]. Molecular and genome size analyses suggest that D-genome diploids hybridized with AC-genome tetraploids followed by chromosome doubling to form hexaploids[12,13]. Alternatively, the hexaploid D genome was inferred to originate from C-genome *A. clauda* Dur[14] rather than from *A. canariensis* Baum & Raj. & Samp[8]. Recent molecular evidence suggest that three tetraploids *A. insularis* Ladiz., *A. maroccana* Grand., and *A. murphyi* Ladiz. may contain the D genome found in hexaploid oat[11,12]. A clear molecular delineation on D-genome origins would lead to a better understanding and utilization of genetic resources in *Avena*.

[1]Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China. [2]University of Chinese Academy of Sciences, Beijing 100049, China. [3]Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013-7012, USA. Correspondence and requests for materials should be addressed to Q.L. (email: liuqing@scib.ac.cn) or J.W. (email: wenj@si.edu)

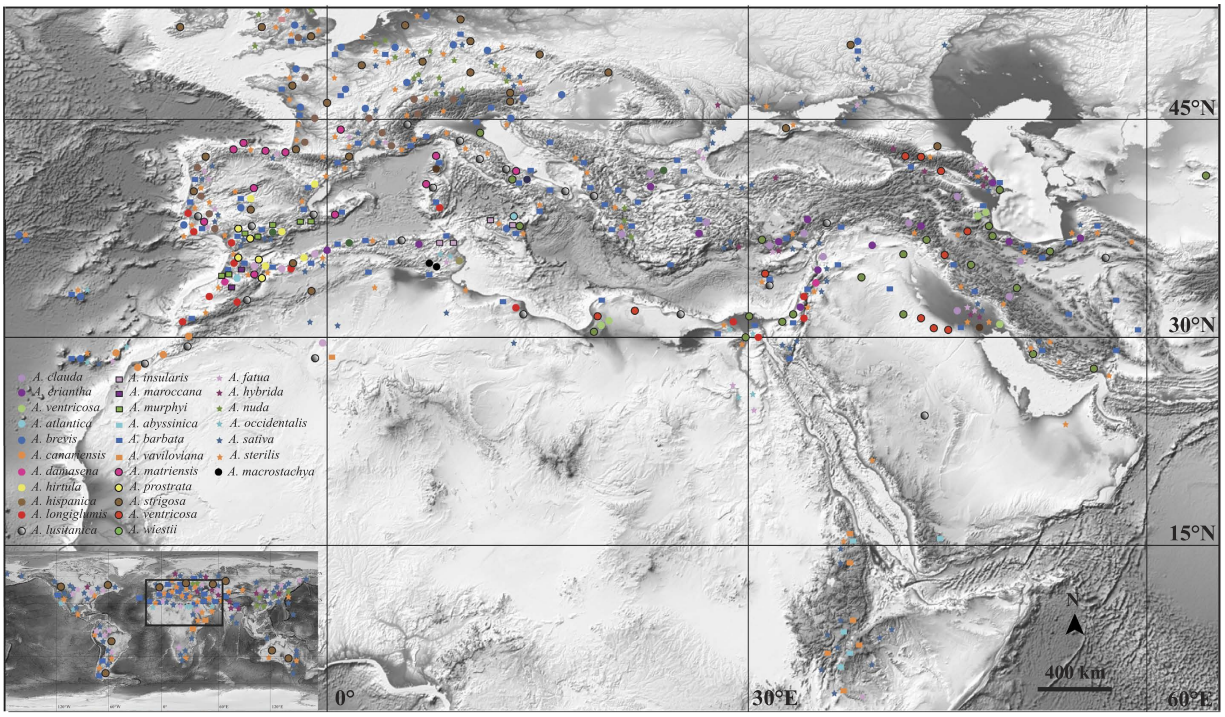


Figure 1. The diversification centre for 28 species of *Avena* (only *A. abyssinica* in eastern Africa and western Asia)[5,24,25] in the Mediterranean Basin using software Adobe Photoshop CS6. The background map was downloaded from http://www.ngdc.noaa.gov/mgg/global/global.html (Amante C, Eakins BW. ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA. Doi:10.7289/V5C8276M, March 2009).

Cultivated oat offers a model for unraveling the dynamic evolutionary process of polyploid crops in the Mediterranean Basin[15]. Initial study of repeat sequences indicated that *A. strigosa* Schreb. DNA was homologous to the A-genome sequences of the cultivated oat[16]. Some studies proposed that *A. canariensis*[9], *A. longiglumis* Dur[13]. or *A. weistii* Steud[17]. might be the A-genome progenitor. Recently, nuclear data demonstrated that the A genome evolved from multiple maternal lineages such as *A. damascena* Rajah & Baum, *A. hirtula* Lag., and *A. wiestii* Steud. rather than from one particular species[9,18]. Numerous intergenomic translocations complicate A-genome progenitor identification for the cultivated oat[8,12,16]. However, broader sampling of nuclear genes should make it possible to resolve this question.

Given chromosome structural differentiation, the C-genome origin of cultivated oat has been under intense scientific scrutiny. Eighteen chromosomes were involved in intergenomic translocations between C and A genomes of *A. sativa*[19]. Cytogenetic study indicated that none of the extant C-genome diploids could be the C-genome progenitor[20]. Plastid data supported that A'C(DC)-genome tetraploids served as the C-genome donors[21,22], whereas nuclear data proposed that the C genome originated from a C-genome diploid (*A. clauda*)[18]. Thus, the C-genome ancestry of cultivated oat remains a challenging mystery.

The Mediterranean Basin, encompassing an area between 28°–48°N and 10°–39°E, is one of the 34 global biodiversity hotspots with c. 24,000 (10% of all seed) plant species[23], and a diversification centre of *Avena* with 28 (96.55%) species (except for *A. abyssinica* Hochst.; Fig. 1)[5,24,25]. The origin of western Mediterranean dated to the Eocene (35 million years ago, mya), and the eastern Mediterranean was formed during the mid-Miocene (16 mya) by collision of Arabian and Eurasian tectonic plates, which led to the configuration of the modern Mediterranean Basin[26]. The mild seasonal contrasts were characterized by greater fluctuations in rainfall than in temperature during the early Miocene (23–16 mya), the repeated cooling events subsequently developed in the mid-Miocene (14–10 mya)[27]. Modern Mediterranean climate became established from 9–8 mya (onset of an arid climate) to 3.2–2.3 mya (onset of a seasonal climate)[28]. The mild climatic oscillation led to the extinction of tropical-subtropical floristic components (e.g., Taxodiaceae)[27] together with the harsh climatic oscillation apparently contributing to the expansion of xerophytic taxa (e.g., *Anthemis*)[28]. The establishment of Mediterranean climate was considered to have triggered the speciation of C_3 polyploid cool-season grasses, e.g., fodder ryegrasses[29].

Here we sample the majority of *Avena* species (Supplementary Table S1[30]) and present a phylogenetic analysis with divergence time estimates based on nuclear and plastid sequences (Table S2[31]). The objectives are to elucidate infrageneric phylogenetic relationships within *Avena*, clarify A-, C-, and D-genome evolutionary history for the cultivated oat, and provide a hypothesis for the early diversification history of *Avena* in the circum-Mediterranean region.

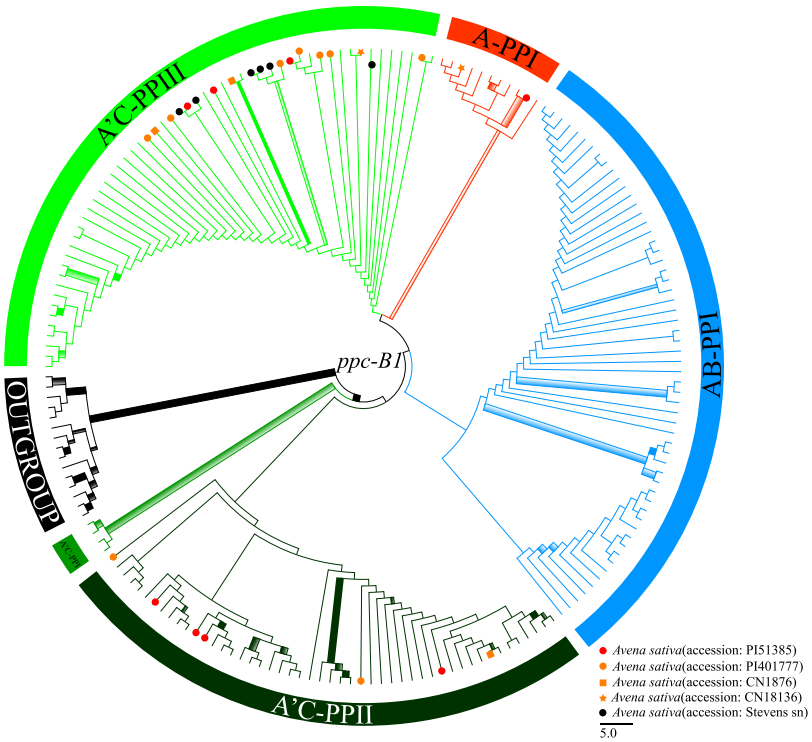


Figure 2. Maximum likelihood tree of *Avena* inferred from nuclear *ppcB1* data including three clades (A-PPI in red, A'C-PPI in green, and A'C-PPIII in light green) and two nodes (AB-PPI in blue and A'C-PPII in dark green). Branch thickness indicate maximum likelihood bootstrap support/Bayesian posterior probability (MLBS/PP): thickest solid = MLBS ≥ 90% and PP ≥ 0.90; thickest shadow = MLBS ≥ 90% or PP ≥ 0.90; thick solid = 89% ≥ MLBS ≥ 70% and 0.89 ≥ PP ≥ 0.70; thick shadow = 89% ≥ MLBS ≥ 70% or 0.89 ≥ PP ≥ 0.70; and thin solid = 69% ≥ MLBS ≥ 50% and 0.69 ≥ PP ≥ 0.50). Red, orange, and black of terminal symbols (circle, square, and star for different accessions) represent thrice, twice, and once clade/node appearance of the cultivated oat. Terminal taxon names and branch support values are shown in Figs S1–S3.

Results

***PpcB1* sequence analysis.** The aligned *ppcB1* matrix had 1017 characters, including exons 8 and 9, and intron 9; with the lengths of 783 bp, 54 bp, and 180 bp, respectively (Table S3). The *ppcB1* data provided 220 (21.63%) parsimony-informative characters. The maximum likelihood (ML) analyses and the Bayesian inference (BI) showed an identical topology for *Avena* (Fig. 2).

The monophyly of *Avena* received strong support (MLBP = 96%, PP = 1.00). Three clades and two nodes were observed in the *ppcB1* phylogram: A'C-PPI (*A. longiglumis*, A-type sequences of *A. agadiriana* and A'-type sequences of *A. maroccana* (PP = 0.98) (Supplementary Fig. S1); node A'C-PPII [*A. atlantica*, *A. damascena*, *A. longiglumis*, *A. wiestii*, and A'-type sequences of tetraploids (*A. agadiriana*, *A. insularis*, *A. maroccana*, and *A. murphyi*), and A- and A'-type sequences hexaploids (*A. fatua*, *A. hybrida*, *A. nuda*, *A. occidentalis*, *A. sativa*, and *A. sterilis*)] (Fig. S1); node AB-PPI [*A. brevis*, *A. canariensis*, *A. damascena*, *A. hirtula*, *A. hispanica*, *A. lusitanica*, *A. prostrata*, *A. strigosa*, *A. wiestii*, tetraploids (*A. abyssinica*, *A. barbata*, *A. vaviloviana* and *A. maroccana*) and hexaploids (without *A. sativa* and *A. occidentalis*)] (Fig. S2); A-PPI [*A. wiestii* and A'(D)-type sequences of hexaploids (without *A. fatua* and *A. nuda*)] (PP = 0.80) (Fig. S3); and A'C-PPIII [*A. hirtula*, C-genome diploids (*A. clauda*, *A. eriantha*, and *A. ventricosa*), and A' and C-type sequences of tetraploids (*A. insularis*, *A. maroccana*, and *A. murphyi*) and hexaploids (without *A. nuda*)] (PP = 0.54) (Fig. S3). The clade A'C-PPI was sister to a single monophyletic lineage (PP = 0.62) containing nodes AB-PPI and A'C-PPII and clades A-PPI and A'C-PPIII in *Avena* (Fig. 2).

Three [A, A'(D), and C]-types of *ppcB1* sequences were identified for one accession of *A. sativa* (*Liu 273*), consistent with its hexaploid origin. These sequences fell into three distinct groups. In clade A'C-PPII, A'-type sequences of *A. sativa* clustered with tetraploids (*A. atlantica*, *A. agadiriana*, *A. insularis*, and *A. murphyi*) and hexaploids (without *A. nuda*) in subclade A'C-PPII-A1 (MLBS = 74%, PP = 0.54), whereas A-type sequence of *A. sativa* clustered with *A. longiglumis* in subclade A'C-PPII-A2 (MLBS = 94%, PP = 1.00) (Fig. S1). C-type sequences of hexaploids grouped with *A. hirtula*, C-genome diploids and three tetraploids (*A. insularis*, *A. maroccana*, and *A. murphyi*) in clade A'C-PPIII (Fig. S3). As for clade A-PPI, A'(D)-type sequence of *A. sativa* clustered with *A. wiestii* in subclade A-PPI-D (MLBS = 70%, PP = 0.94), which was labelled as "A'(D)" due to its distinct status in *Avena* (Fig. S3).

***GBSSI* sequence analysis.** The aligned *GBSSI* matrix had 1352 characters, including exons 9, 10, 11, 12, 13, and 14, and introns 8, 9, 10, 11, 12, 13, and 14, with the lengths of 53 bp, 81 bp, 194 bp, 88 bp, 129 bp, 22 bp,

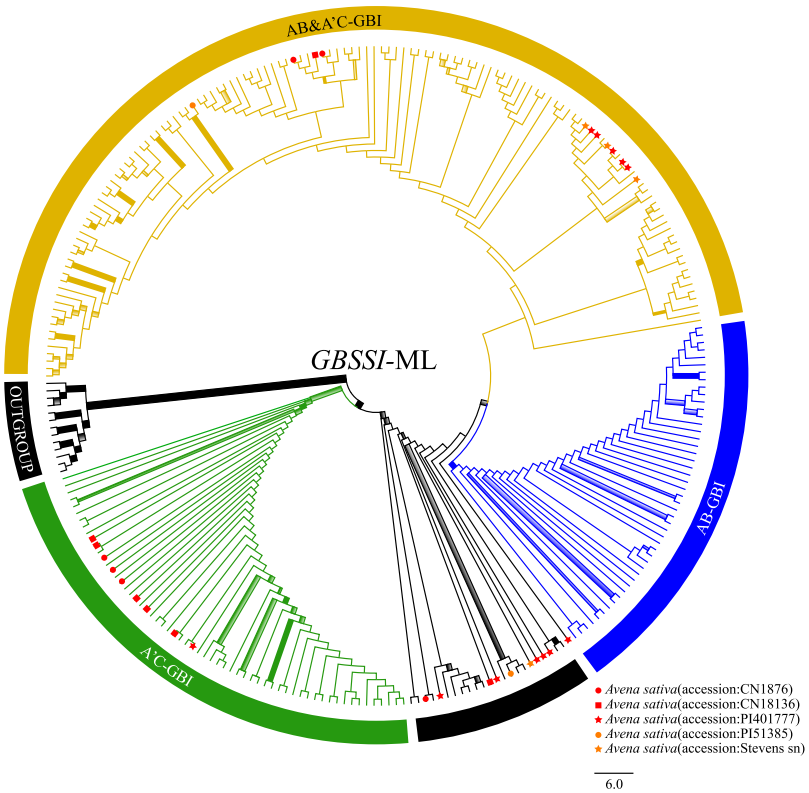


Figure 3. Maximum likelihood tree of *Avena* inferred from nuclear *GBSSI* data including three clades (AB-GBI in blue, A'C-GBI in green, and AB&A'C-GBI in brown) plus eight polyploids in unmarked black. Explanation of branch thickness and colorful terminal symbols refer to Fig. 2. Terminal taxon names and branch support values are shown in Figs S4–S6.

47 bp, 148 bp, 153 bp, 127 bp, 154 bp, 152 bp, and 4 bp, respectively (Table S3). The *GBSSI* data provided 434 (32.1%) parsimony-informative characters. The ML and BI analyses generated different topologies for *Avena* (Figs 3 and 4).

The monophyly of *Avena* received strong support (MLBS = 94%, PP = 1.00) (Figs 3 and 4). In the ML analysis, three clades plus eight polyploids [A'(D)-type sequences of tetraploids *A. insularis* and *A. maroccana*, and hexaploids (without *A. nuda*)] were observed in the *GBSSI* tree: A'C-GBI [C-genome diploids, C-type sequences of tetraploids (*A. insularis*, *A. maroccana*, and *A. murphyi*) and hexaploids (without *A. nuda*)] (MLBS = 66%, PP = 1.00) (Fig. S4); AB-GBI [*A. atlantica*, *A. hirtula*, *A. longiglumis*, *A. wiestii*, A-type sequences of tetraploids (*A. abyssinica*, *A. barbata*, and *A. vaviloviana*) and hexaploid *A. fatua*] (MLBS = 100%, PP = 0.93) (Fig. S5); and AB&A'C-GBI [*A. atlantica*, *A. brevis*, *A. canariensis*, *A. damascena*, *A. hirtula*, *A. hispanica*, *A. longiglumis*, *A. lusitanica*, *A. strigosa*, *A. wiestii*, A-type sequences of tetraploids (*A. abyssinica*, *A. agadiriana*, *A. barbata*, and *A. vaviloviana*), A'-type sequences of *A. maroccana* and *A. murphyi* and hexaploids] (Fig. S6). The clade A'C-GBI was sister to a single lineage (PP = 0.98) containing clades AB-GBI and AB&A'C-GBI in *Avena* (Fig. 3).

In BI analyses, four clades plus eight polyploids [A'(D)-type sequences of tetraploids *A. insularis* and *A. maroccana*, and hexaploids (without *A. nuda*)] were observed in the *GBSS1* tree: A'C-GBI (C-type sequences of clade AC-GBI members in ML analysis) (MLBS = 66%; PP = 1.00) (Fig. S7); A'C-GBII [*A. brevis*, *A. canariensis*, *A. hirtula*, *A. hispanica*, *A. longiglumis*, *A. lusitanica*, *A. strigosa*, *A. wiestii*, A-type sequences of *A. agadiriana* and hexaploids (*A. hybrida*, *A. nuda*, and *A. sativa*), and A'-type sequences of *A. maroccana* and *A. murphyi*)] (PP = 0.50) (Fig. S8); AB-GBI [A-type sequences of clade AB-GBI members in ML analysis] (MLBS = 100%; PP = 0.94) (Fig. S9); and AB-GBII [*A. atlantica*, *A. damascena*, *A. hirtula*, *A. longiglumis*, *A. wiestii*, A-type sequences of three AB-genome tetraploids (*A. abyssinica*, *A. barbata*, and *A. vaviloviana*) and hexaploids (without *A. nuda* and *A. sterilis*)] (PP = 0.66) (Fig. S10). Clades A'C-GBII plus AB-GBII included the same members as clade AB&AC-GBI (without *A. sterilis*). Clade AB-GBI was sister to clade AB-GBII, and this group (PP = 0.66) plus clade A'C-GBII (PP = 0.50) was assigned to a single monophyletic lineage (PP = 0.98), which was sister to clade A'C-GBI. This large clade received strong support (MLBS = 94%, PP = 1.00) in *Avena* (Fig. 4).

Three [A, A'(D), and C]-types of *GBSSI* sequences were identified in four accessions of *A. sativa* (*Liu 272*, *310*, *311*, and *348*), consistent with its hexaploid origin. These sequences fell into three distinct groups. In clade A'C-GBI, C-type sequences of *A. sativa* clustered with C-genome diploids, C-type sequences of tetraploids (*A. insularis*, *A. maroccana*, and *A. murphyi*) and hexaploids (without *A. nuda*) (MLBS = 78%, PP = 1.00) (Figs S4 and S7). A-type sequences of *A. sativa* were inserted into clade AB&A'C-GBI and clade AB-GBII (Figs S6 and S10), respectively. However, A'(D)-type sequences of *A. sativa* were embedded within a single lineage containing clades AB-GBI and AB&A'C-GBI in the ML analysis (Fig. S6), and containing clades A'C-GBII, AB-GBI, and AB-GBII in BI analysis (Fig. S10).

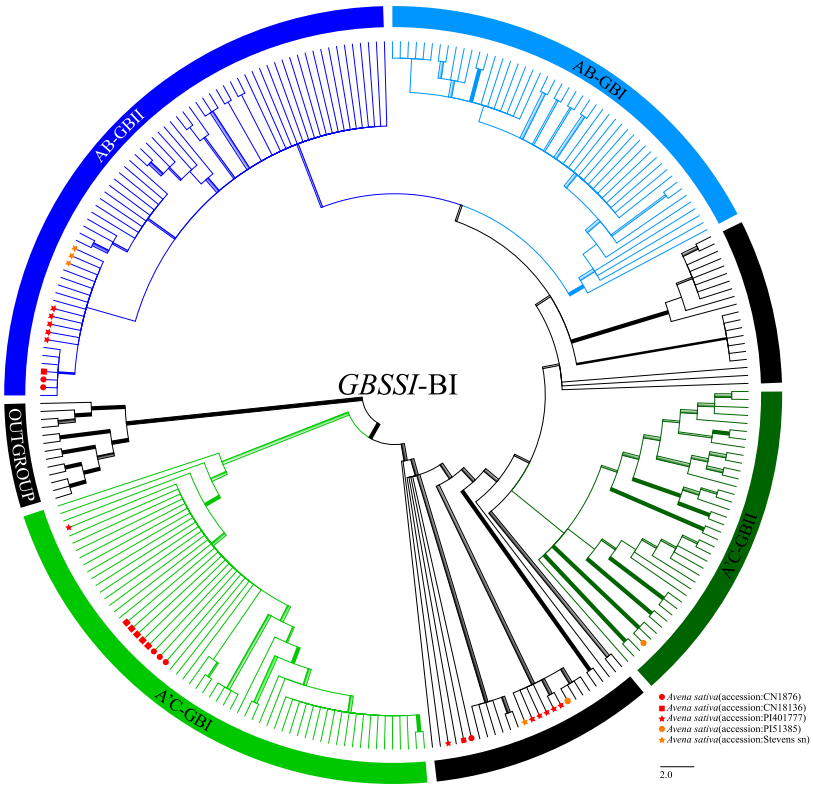


Figure 4. Bayesian inference phylogram of *Avena* inferred from nuclear *GBSSI* data including four clades (AB-GBI in light blue, AB-GBII in blue, A'C-GBI in green, and A'C-GBII in green) plus eight polyploids in unmarked black. Explanation of branch thickness and colorful terminal symbols refer to Fig. 2. Taxon names and branch support value are shown in Figs S7–S10.

***Gpa1* sequence analysis.** The aligned *gpa1* matrix had 1034 characters, including exons 10, 11, 12, introns 10, 11, and 12; with the lengths of 22 bp, 94 bp, 60 bp, 681 bp, 92 bp, and 85 bp, of which 137 (13.25%) were parsimony-informative. ML and BI analyses had an identical topology for *Avena* (Fig. 5).

The monophyly of *Avena* received strong support (MLBP = 100%, PP = 1.00). Seven clades were observed for the *gpa1* tree (Fig. S11): C-GPI (C-genome diploids) (MLBS = 100%, PP = 1.00); A'C-GPI [C-type sequences of tetraploids (*A. insularis*, *A. maroccana*, and *A. murphyi*) and five hexaploids (without *A. nuda*)] (MLBS = 88%, PP = 1.00); A'C-GPII [A-type sequences of *A. agadiriana*, and A'-type sequences of *A. insularis* and *A. murphyi*) and A'(D)-type sequences of four hexaploids (*A. fatua*, *A. occidentalis*, *A. sativa*, and *A. sterilis*)] (MLBS = 65%, PP = 0.96); A-GPI (*A. canariensis* and A-type sequence of *A. hybrida*) (MLBS = 96%, PP = 1.00); AB-GPI [diploids (*A. atlantica*, *A. damascena*, *A. hirtula*, and *A. wiestii*), A-type sequences of tetraploids (*A. abyssinica*, *A. barbata*, *A. vaviloviana*) and A'-type sequences of *A. maroccana*)] (MLBS = 73%, PP = 1.00); A'C-GPIII (*A. hirtula* and A'-type sequences of *A. maroccana*) (MLBS = 87%, PP = 1.00); and AB-GPII [*A. atlantica*, *A. brevis*, *A. damascena*, *A. hirtula*, *A. hispanica*, *A. longiglumis*, *A. lusitanica*, *A. strigosa*, and *A. wiestii*, A-type sequences of four AB-genome tetraploids and A'(D)-type sequences of *A. insularis* and hexaploids] (MLBS = 52%, PP = 0.93). Clades A-GPII, A-GPI, AB-GPI, A'C-GPIII, and A'C-GPII formed one monophyletic lineage (MLBS = 99%, PP = 1.00), and this lineage in turn was sister to clade A'C-GPI with strong support (MLBS = 92%, PP = 0.99), then the large group was sister to clade C-GPI with strong support (MLBS = 100%, PP = 1.00) (Fig. S11).

Two [A'(D)- and C-] types of *gpa1* sequences were identified in a single accession of *A. sativa* (*Liu 310*). These sequences fell into two distinct groups, with A'(D)-type sequence of *A. sativa* nested within clade AB-GPII, and C-type sequences of *A. sativa* nested within clade A'C-GPI (Fig. S11).

Divergence times. The combined plastid data of 104 accessions comprised 2819 characters, of which 232 (8.23%) were parsimony-informative. The BEAST analysis generated a well-supported tree, which was identical to the topologies obtained from ML and BI analyses of *Avena* (Fig. 6). Two clades were recognized in the plastid phylogram: C-NRR (C-genome diploids *A. clauda*, *A. eriantha*, and *A. ventricosa*; MLBS = 99%, PP = 1.00); A'C-NRR [*A. brevis*, *A. canariensis*, A-type sequences of *A. barbata* and *A. agadiriana*, and A'(D)-type sequences of tetraploids (*A. insularis*, *A. maroccana*, and *A. murphyi*) and hexaploids] + AB-NRR [*A. atlantica*, *A. damascena*, *A. hirtula*, *A. longiglumis*, *A. lusitanica*, *A. prostrata*, *A. strigosa*, *A. wiestii*, A-type sequences of AB-genome tetraploids and A'-type sequence of *A. maroccana*) and hexaploids (without *A. sativa* and *A. occidentalis*)]. Clade

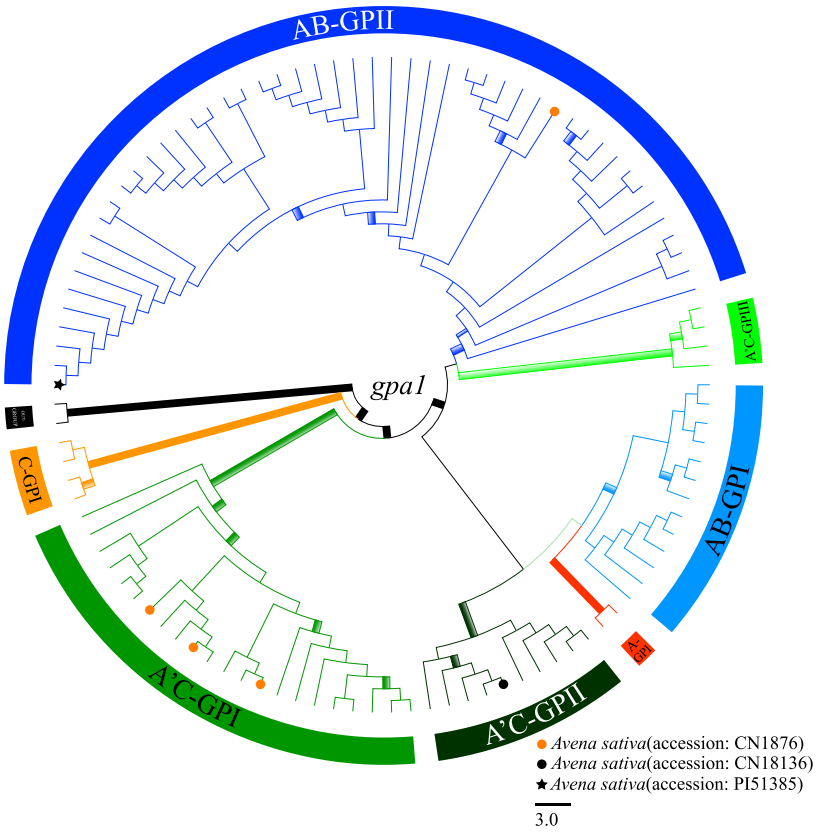


Figure 5. Maximum likelihood tree of *Avena* inferred from nuclear *gpa1* data including seven clades (A-GPI in red, C-GPI in brown, AB-GPI in light blue, AB-GPII in blue, A'C-GPI in green, A'C-GPII in dark green, and A'C-GPIII in light green). Explanation of branch thickness and colorful terminal symbols refer to Fig. 2. Taxon names and branch support value are shown in Fig. S11.

A'C-NRR + AB-NRR (MLBS = 97%, PP = 0.96) was sister to clade C-NRR in *Avena* (Fig. 6). Here we discuss divergence times for the lineages of interest as shown in Table S4.

The uncorrelated-rate relaxed molecular clock suggests that the crown age of *Avena* was 20.04 [95% highest posterior density (HPD) 3.56–35.06] mya (node 1). This was also the stem ages for clades C-NRR and A'C-NRR + AB-NRR, whose crown ages were 10.71 (HPD: 1.62–20.25) and 14.54 (HPD: 2.68–25.02) mya, respectively (nodes 2 and 3). The crown age of clade A'C-NRR + AB-NRR was also the divergence time for nodes A'C-NRR and AB-NRR (nodes 4 and 8). The crown ages of the *A. sativa* lineages were 2.43, 2.46, and 2.97 mya (nodes 5, 6, and 7), respectively (Fig. 6).

Discussion

Infrageneric phylogeny and allopolyploidy events in *Avena*. Two strongly supported infrageneric lineages within *Avena* were identified by the plastid data: the C-genome diploid lineage (*Avena* sect. *Ventricosa*) containing *A. clauda*, *A. eriantha*, and *A. ventricosa* in clade C-NRR; and the A-genome diploid-polyploid lineage (*Avena* sect. *Avena*) containing other congeneric species in clade AB-NRR + A'C-NRR (Fig. 6). Members of C-genome diploid lineage were distributed from the south Mediterranean to the Irano-Turanian region[5,6], and they were easily distinguished based on their unequal glumes[15], fusiform caryopses with striate sculpturing[32], and heterobrachial chromosomes with heterochromatin blocks along long-arm terminals[8]. Morphological, cytogenetic, and phylogenetic evidence supported recognizing this lineage as a distinct section, *Avena* sect. *Ventricosa*, which was embedded within clades A'C-PPIII (Fig. S3) and A'C-GBI based on nuclear data (Figs S4 and S7). In the *ppcB1* and *GBSSI* trees, *Avena* sect. *Ventricosa* shared a high degree of genetic similarity with C-type homoeologues of polyploids. Consequently, the ancestor of *Avena* sect. *Ventricosa* was probably the C genome donor for A'C(DC)-genome tetraploids and hexaploids.

Avena sect. *Avena* was proposed for the A-genome diploid-polyploid lineage including nodes with low support in the plastid tree (Fig. 6). Chromosome rearrangement had occurred since the divergence of *Avena* sect. *Avena* progenitors, leading to the divergence of A-genome constitution[4,10,30], which could be divided into two groups in the section. The first group, As-genome diploids (*A. brevis*, *A. hispanica*, *A. strigosa*, *A. atlantica*, *A. hirtula*, *A. lusitanica*, and *A. wiestii*), *A. damascena* (Ad-genome), and *A. longiglumis* (Al-genome) clustered with three A'C(DC)-genome tetraploids (*A. insularis*, *A. maroccana*, and *A. murphyi*) in node A'C-PPII (Fig. S1); and the second group, As-genome diploids (*Avena brevis*, *A. hispanica*, *A. strigosa*, *A. atlantica*, *A. hirtula*, *A. lusitanica*, and *A. wiestii*), *A. canariensis* (Ac-genome), *A. damascena* (Ad-genome), and *A. prostrata* (Ap-genome)

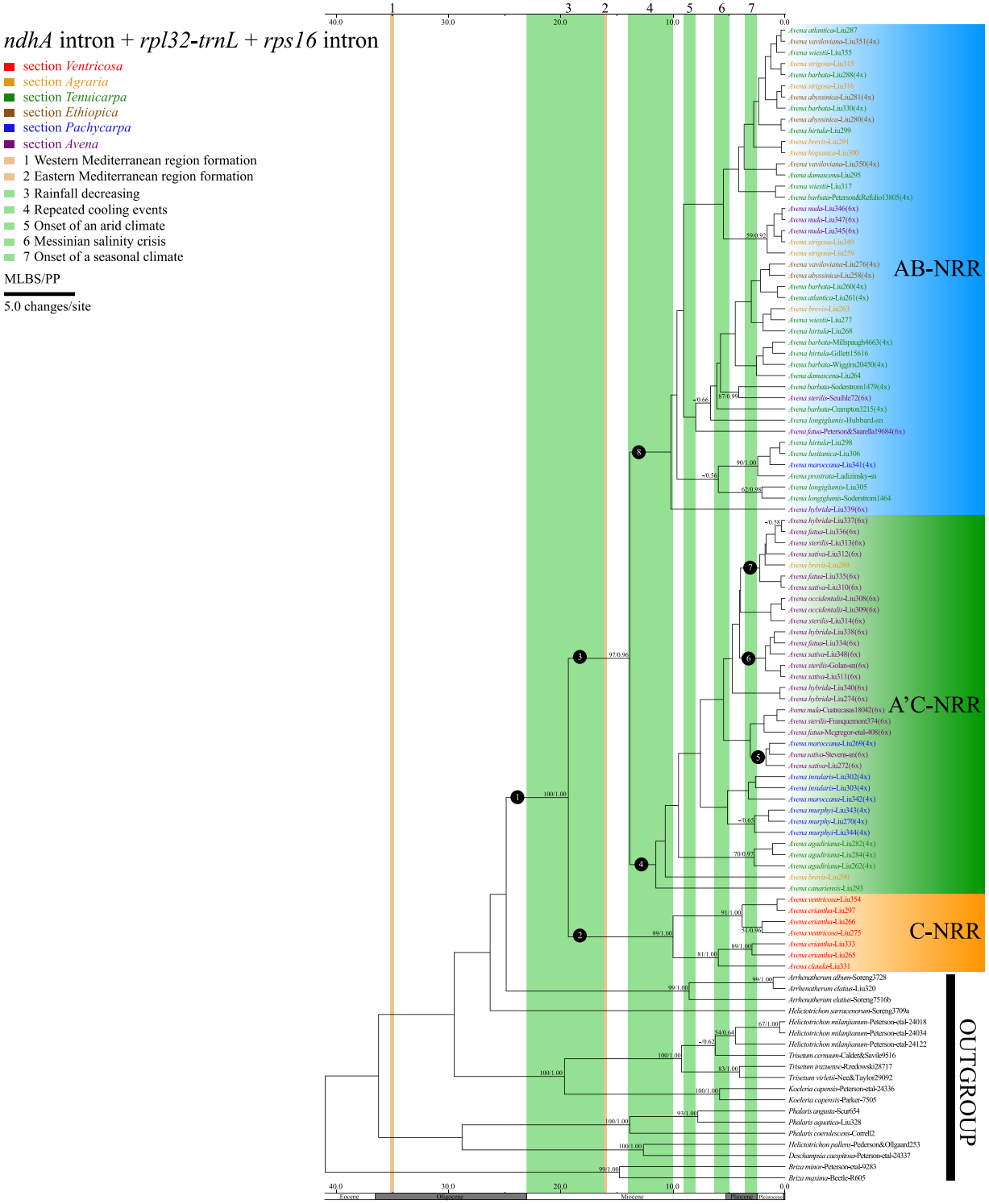


Figure 6. Chronogram of *Avena* and its close relatives based on plastid (*ndhA* intron, *rpl32-trnL*, and *rps16* intron) data including two clades (C-NRR and A'C-NRR + AB-NRR) inferred from BEAST. Numbers above the branches are MLBS/PP. Taxon labels are in the format: *Avena vaviloviana*-Liu351 (4x), where *Avena vaviloviana* indicates that the sequence belongs to the species; Liu351 indicates voucher; (4x) indicates that the species is tetraploid. Coloured taxon labels correspond to sections. Node number indicates the lineages of interest (Table S4).

clustered with AB-genome tetraploids in node AB-PPI (Fig. S2), and As-genome diploids (*A. atlantica*, *A. hirtula*, and *A. wiestii*), *A. damascena* (Ad-genome), and *A. longiglumis* (Al-genome) clustered with AB-genome tetraploids in clade AB-GBII (Fig. S10). Phylogenetic relationships among the As-, Ad-, and Al-genome diploids

and A'C-genome tetraploids, together with those of As-, Ac-, Ad-, Al-, and Ap-genome diploids and AB-genome tetraploids indicated that the close relatives of the A'C- and AB-genome tetraploids might be found within different A-genome groups based on *ppcB1* data. Therefore, we hypothesize that AB- and A'C(DC)-genome tetraploids originated from different A-genome diploid ancestors (Table S5[13,16,18,21,22]). Whole genome sequencing data including repetitive DNA might be able to detect the A (A')-genome constitution in *Avena* tetraploids[12].

Three secondary gene pool members, *A. insularis*, *A. maroccana*, and *A. murphyi* are native to the northwest Africa and adjacent environs (i.e., *A. insularis* in Sicily and Tunisia, *A. maroccana* in Moroccana, and *A. murphyi* in southern Spain and northern Morocco)[5,21]. They formed a clade A'C-GPI together with hexaploids in the *gpa1* tree (Fig. S11). In view of the chromosome pairing capacity between A'C(DC)-genome tetraploids and hexaploids[21] and sequence-based diversity data[9], the A' and C genomes in the three tetraploids matched closest with D and C genomes in cultivated oat[10]. Since the As-, Ad-, and Al-genome diploids were involved in the A'C(DC)-genome tetraploid formation, it cannot be excluded that A'C(DC)-genome tetraploids originated from the ancient allotetraploidy events owing to the isolated phylogenetic positions of *A. maroccana* in clade A'C-PPI (Fig. S1), those of *A. insularis* inserted within a monophyletic lineage of the *GBSSI* tree (Fig. S8), and that of *A. murphyi* in clade A'C-GPI (Fig. S11). If this was the case, one would expect three or more ancient A-genome diploids to have participated in the origin of A'C(DC)-genome tetraploid. The three tetraploids have been reported as AC-genome-derived based on anonymous genotyping-by-sequencing (GBS) markers[9], while the A'C(DC) designation of the tetraploids is fully compatible with our results together with another analysis based on GBS markers located on hexaploid chromosomes[10].

Within the As-genome diploids, *Avena hispanica* was isolated from the closely related *A. hirtula* and *A. lusitanica* in the clade A'C-NRR + AB-NRR of plastid tree (Fig. 6). However, *A. lusitanica* (group 5) showed specific genetic divergence from *A. hirtula* and *A. hispanica* (group 3) in high-density GBS analysis[10]. Based on the length of lemma biaristulate tips (5–12 mm[6]) and the genome size (9.08 ± 0.11[12]), *A. hirtula* can be easily differentiated from the two As-genome diploids, that have a similar genome size to the smallest Ad-genome diploid *A. damascene*[12]. The incongruencies among morphological characters and genetic differences make the identification of the As-genome species challenging. *Avena lusitanica* and *A. hispanica* might represent ecotypes of *A. hirtula* found in the circum-Mediterranean, western Asia, and Europe[5,33].

Allohexaploid origin of *Avena sativa*. Two distinct steps were inferred for the formation of the cultivated oat. The first step includes the ancient allotetraploidy events involving the hybridization between the ancient A'(or diverged A)- and C-genome diploid ancestors to form A'C (now called DC)-genome tetraploids. The second step includes subsequent recent allohexaploidy events involving hybridization between DC-genome tetraploids and the more recent A-genome diploid progenitors to form the extant ACD-genome hexaploids[18]. The close relationship between the genetically homogeneous *Avena* sect. *Ventricosa* and the C-copy sequences of A'C-genome tetraploids plus hexaploids was a novel discovery which suggested their C-genome donor to be the ancestor of *Avena* sect. *Ventricosa*. This was consistent with the hypothesis that the paleotetraploidy events pre-dated and potentially triggered divergence of the extant A'C(DC)-genome tetraploids in narrow ranges of the Mediterranean Basin[9]. In the *gpa1* tree, A'C(DC)-genome tetraploids together with hexaploids comprised the clade A'C-GPI (Fig. S11). Therefore, the nuclear data provided robust evidence for the designated D and C genomes in cultivated oat, matching closest with A'(D)- and C-genome in *A. insularis*, *A. maroccana*, and *A. murphyi*, and the A-genome designation matches better with the extant A-genome diploids in *Avena*.

The close relationships among three A-genome diploids and cultivated oat were observed in the *ppcB1* tree, i.e., *A. atlantica*, *A. longiglumis*, and *A. wiestii* were embedded within the A'C-PPII-A1, A'C-PPII-A2, and A-PPI subclades (Figs S1 and S3). The IGS-RFLP dendrogram suggested that *A. atlantica* should be placed within the cluster containing polyploids rather than within the *A. strigosa* cluster[13], showing that *A. atlantica* has genetic dissimilarities with *A. strigosa*[34]. *Avena longiglumis* formed a strongly supported subclade A'C-PPII-A2 with *A. sativa* (Fig. S1). In addition, two new *ppcB1* clones of *A. wiestii* (*Rawi 11581*, US!) located in node A'C-PPII and clade A-PPI (Figs S1 and S3), together with two reported *FL intron2* clones (*CIav 9053*)[18] indicate that the coexistance of diploid and tetraploid forms for *A. wiestii* is certainly different from other As-genome diploids. Although the different genome forms of *A. wiestii* were close in genome size[12], the intraspecific differences between *A wiestii* deserves further investigation. Two plausible explanations can be proposed for the ploidy level of allelic variation found in *A. wiestii*. First, the three diploids may have arisen by allopolyploidy and subsequent unequal diploidization led to heterozygotes. Second, introgression may have brought about very subtle morphological and genetic changes in *Avena* (Fig. 2), because stabilized introgressant species were more common than cases of dispersed introgression involving extensive gene flow among distinct species[35]. The two explanations are not mutually exclusive, such as *Leucaena*[36] sharing unequal diploidization and introgression processes. The paternally inherited genome of an allopolyploid is usually more prone to genetic change than the maternally derived genome according to the nuclear cytoplasmic interaction hypothesis[37]. In support of this hypothesis, it has been proposed that *A. atlantica*, *A. longiglumis*, or *A. wiestii* might carry the diverged A-genomes because considerable allelic variation was detected in the *ppcB1* and *FL int2* data[18].

Based on *ppcB1* and cytogenetic data, a close phylogenetic relationships between the A and D genomes substantiates the multiple origins for cultivated oat[19]. However, the integrated theory for the long-term evolutionary impact of recurrent polyploidy was unclear for hexaploid divergence in *Avena*. Based on the level of genetic variation, it is logical to postulate that recurrent polyploidy from genetically distinct diploid progenitors would introduce genetic variation into hexaploids. Nuclear data have demonstrated that recurrent polyploidy can lead to hexaploids being reproductively isolated to varying degrees. Six hexaploids were found within A'C-PPIII, AB&A'C-GBI, AB-GPII clades, while some hexaploids were dispersed within other clades in the nuclear gene trees. For example, *Avena nuda* is morphologically distinct with falling caryopses, but it was independently inserted within node AB-PPI and clades AB-GBII and A'C-GBII, demonstrating varying degrees of interfertility

when compared with *A. sativa*[15]. Clearly six hexaploids cannot be regarded as a single species designated as *A. sativa*[38], especially for wild hexaploids—*A. fatua*, *A. sterilis*, *A. hybrida*, and *A. occidentalis*, each adapted to respective microenvironments in the circum-Mediterranean region[33]. *Avena* provided a great model for studying polyploidy, especially concerning the evolutionary and genetic processes associated with extensive intergenomic translocations[30] and northward diffusion into cooler areas[33] over a time scale of c. 20 mya (Fig. 6). Future studies of *Avena* need to investigate the unique and conserved genomic signatures using phylogenomics[39,40].

Paleoclimatic hypothesis for the lineage divergence of *Avena*. It has been proposed that the Miocene-Pliocene interval was a key period in the origin of Mediterranean temperate plants and involved two major climatic oscillations[41]. The former comprised mild seasonal climatic contrasts that resulted from rainfall decreasing and repeated cooling events during the early to middle Miocene; and the latter was characterized by a high seasonal Mediterranean climate resulting from the onset of aridity and seasonality during the late Miocene to Pliocene[27]. During these mild climatic contrasts, shifts in vegetation from subtropical forest to annual grasslands occurred in the Mediterranean Basin[29]. The resultant habitat heterogeneity may have had lasting impact on the genetic and phenotypic divergence of major lineages in *Avena*[27]. Major lineages in *Avena* are distinguished by ecological differentiation: *Avena* sect. *Ventricosa* is distributed in calcareous rocky plateaus or mountain grassland habitats; and *Avena* sect. *Avena* is distributed in carbonate sands or semi-desert habitats in the circum-Mediterranean region. The crown ages of these two lineages are estimated at 14.54 (HPD: 2.68–25.02) and 10.71 (HPD: 1.62–20.25) mya, respectively (Fig. 6). These periods coincide with mild seasonal climatic contrasts that occurred during the early to middle Miocene. It appears a temporal relationship exists between the mild seasonal climatic contrasts and the divergence of major lineages in *Avena*.

Cultivated oat may have arisen multiple times in response to selection pressure such as geographic isolation. The long-term aridity of the Mediterranean Basin summer became more severe along a south-eastern to north-western gradient during the late Miocene to Pliocene[27], leading to the domination of open habitats by C_3-pooid grasses[42]. The increased colonization capacity of cultivated oat may be strongly linked to hybridization between diploid and tetraploid progenitors followed by chromosome duplication. Recurrent polyploidization events in the *Avena sativa* lineages (nodes 5, 6, and 7) seem to correlate with highly seasonal climatic oscillation. Geographic isolation might have contributed to genetic differentiation in the progenitor-derivative species pair, the presumed D(or A')-genome progenitors having disjunct distributions in the Mediterranean region (e.g., *A. atlantica* was endemic to Morocco, *A. wiestii* was endemic to the east Mediterranean, east Saharo-Arabian, and Irano-Turanian, and *A. longiglumis* was endemic to the west-south-east Mediterranean and Saharo-Arabian)[5]. The once extensive distribution of the narrow-endemic A'C(DC)-genome tetraploids underwent contraction. Hybridization might have been a key element in the successful spread of cosmopolitan cultivated oat as a result of incorporation of locally adapted genes from different progenitor genomes. If this was the case, then the initial hybridization must have pre-dated the formation of modern Mediterranean region[26], which isolated *A. wiestii* (eastern-most) from *A. atlantica* (western-most). Therefore, the independent hexaploidy events of cultivated oat were modulated by harsh climatic oscillation, thus *A. sativa* was able to adapt to new habitats.

Avena represents a remarkable model to study because its history of polyploidy, lineage divergence, and complex reticulate evolution. The complex evolution of cultivated oat and its close relatives involved paleotetraploidy events between the ancient A(or A')- and C-genome diploid ancestors and subsequent recent allohexaploidy events between A'C(DC)-genome tetraploids and the more recent A-genome diploid progenitors. The pattern of recurrent polyploidizations in *Avena* and their temporal relationships with paleoclimatic oscillations is unparalleled among polyploid crops occurring in the circum-Mediterranean region[4,43].

Methods

Taxon sampling and data collection. Eighty-nine accessions of 27 species were sampled to represent the morphological diversity and geographic range of six sections in *Avena*[5], together with outgroups comprising 20 accessions of 16 species from seven allied genera (Supplementary Table S1[30]) based on the recent phylogeny and classification of Poaceae[44]. Leaf material was obtained from seedlings and herbarium specimens.

Three low-copy nuclear genes, *phosphoenolpyruvate carboxylase B1* (*ppcB1*), *granule-bound starch synthase I* (*GBSSI*) and *G protein alpha subunit 1* (*gpa1*), were used. The *ppcB1* gene encodes *PEPC* enzyme for the oxaloacetate replenishment of the tricarboxylic acid cycle in C_3 plants[45], the *GBSSI* gene encodes *GBSSI* enzyme for the amylose synthesis in plants[46], and the *gpa1* gene encodes a G-protein α subunit for signal transduction in flowering plants[47]. These loci have previously been used for accurate phylogenetic assessments in Poaceae[2,47,48]. Based on genome-wide studies on cereal crops, the three loci appear to be on different chromosomes[4,48,49], thus each of nuclear markers can provide an independent phylogenetic estimate.

Genomic DNA was extracted following Liu *et al*[31]. and 864 new sequences were generated for nuclear (*ppcB1*, *GBSSI*, and *gpa1*) and plastid (*ndhA* intron, *rpl32-trnL*, and *rps16* intron) fragments, which were amplified using designed or published primers and protocols listed in Table S2[31]. Amplified products were purified using polyethylene glycol (PEG) precipitation protocols and sequenced using an ABI PRISM 3730XL DNA Analyzer (Applied Biosystems, Foster City, CA, USA). For accessions that unsuccessfully underwent direct sequencing, the purified PCR products were cloned into pCR4-TOPO vectors and transformed into *Escherichia coli* TOP10 competent cells following the protocol of TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA, USA). The resulting sequences were edited using Sequencher v.5.2.3 (Gene Codes Corp., Ann Arbor, MI, USA) and aligned with MUSCLE v.3.8.31[50], followed by manual adjustment in SE-AL v.2.0a11 (http://tree.bio.ed.ac.uk/software/seal/). All sequences were deposited in GenBank (KT452936–453223, KT723464–724040).

Phylogenetic analyses. Phylogenetic analyses were performed using maximum likelihood[51] and Bayesian inference[52]. Nucleotide substitution models were selected based on the Akaike Information Criterion determined

by Modeltest v.3.7[53]. ML and bootstrap analyses (MLBS) were performed using the best-fit model (Table S3) for 1,000 bootstrap replicates in GARLI v.0.96[54], with runs set for unlimited generations, and automatic termination following 10,000 generations without significant topological change (lnL increase of 0.01). The output file containing the best trees for bootstrap reweighted data was then read into PAUP* v.4.0b10[55] where the majority-rule consensus tree was constructed to calculate MLBS.

BI analyses were conducted in MrBayes v.3.2.1[56] using the best-fit model for each nuclear and the combined plastid loci (Table S3). The Bayesian Markov Chain Monte Carlo (MCMC) algorithm was run for 30 million generations with four incremental chains starting from random trees and sampling one out of every 1,000 generations. Convergence between runs and the choice of an appropriate burn-in value were assessed by comparing the traces using Tracer v.1.5 (http://tree.bio.ed.ac.uk/software/tracer). All resulting trees were then combined with LogCombiner v.1.6.1 (http://beast.bio.ed.ac.uk/) with 25% burn-ins. The remaining trees (c. 45,000) were used to calculate the Bayesian posterior probabilities (PP) for internal nodes. Data sets and phylogenetic trees are available at TreeBase (http://treebase.org, study no. TB2: S18544) (Reviewer access URL: http://purl.org/phylo/treebase/phylows/study/TB2:S18544. Figures 1–6 (Supplementary Figs S1–S11) were prepared using Photoshop CS6 v.13.0 (Adobe, San Jose, CA, USA).

Divergence time estimation. The molecular dating analyses employed plastid markers a strict molecular clock model was rejected at a significance level of 0.01 (LR = 963.1856, d.f. = 102, $P < 0.01$) based on a likelihood ratio test[51]. A Bayesian relaxed clock model was implemented in BEAST v.1.8.2[56] to estimate divergence times in *Avena*. Three plastid markers were partitioned using BEAUTI v.1.8.2 (within BEAST) with the best fit model determined by Modeltest v.3.7 (Table S3). The stipoid-Pooideae lineage including *Avena* plus outgroups was dated to be 49.71 mya based on eight phytolith fossils, and thus the crown age of *Avena* plus outgroups was set at 49.71 mya since fossil surveys provide no evidence of an earlier date for the origin of the stipoid-Pooideae lineage during the late Eocene[57].

A Yule tree prior, linked uncorrelated lognormal relaxed clock model, and default operators were defined in the BEAST xml input file. After optimal operator adjustment as suggested by the output diagnostics from preliminary BEAST runs, two independent MCMC runs were performed for 30 million generations, each run sampling every 1,000 generations with 25% burn-ins. All parameters had a potential scale reduction factor that was close to one, indicating that the posterior distribution had been adequately sampled. A 50% majority rule consensus from the retained posterior trees (c. 45,000) of three runs was obtained using TreeAnnotator v.1.8.2 (within BEAST) with a PP limit of 0.5 and mean lineage heights. The convergence between two runs was checked using the "cumulative" and "compare" functions in AWTY[58].

References

1. Meyer, R. S., Duval, A. E. & Jensen, H. R. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* **196,** 29–48 (2012).
2. Triplett, J. K., Clark, L. G., Fisher, A. E. & Wen, J. Independent allopolyploidization events preceded speciation in the temperate and tropical woody bamboos. *New Phytol.* **204,** 66–73 (2014).
3. Petersen, G., Seberg, O., Yde, M. & Berthelsen, K. Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Mol. Phylogenet. Evol.* **39,** 70–82 (2006).
4. Loskutov, I. G. & Rines, H. W. *Avena* L. In *Wild crop relatives: genomic and breeding resources*, *vol.* 1. *Cereals* (ed. Kole, C.) 109–184 (Springer, Heidelberg, 2011).
5. Lin, L. & Liu, Q. Geographical distribution of *Avena* L. (Poaceae). *J. Trop. Subtrop. Bot.* **23,** 111–122 (2015).
6. Baum, B. R. *Oats: wild and cultivated, a monograph of the genus Avena L. (Poaceae)* 1–463 (Minister of Supply and Services, Ottawa, 1977).
7. Katsiotis, A., Loukas, M. & Heslop-Harrison, J. S. Repetitive DNA, genome and species relationships in *Avena* and *Arrhenatherum* (Poaceae). *Ann. Bot.* **86,** 1135–1142 (2000).
8. Linares, C., Ferrer, E. & Fominaya, A. Discrimination of the closely related A and D genomes of the hexaploid oat *Avena sativa* L. *Proc. Natl. Acad. Sci. USA* **95,** 12450–12455 (1998).
9. Chew, P. *et al.* A study on the genetic relationships of *Avena* taxa and the origins of hexaploid oat. *Theor. Appl. Genet.* **129,** 1405–1415 (2016).
10. Yan, H. H. *et al.* High-density marker profiling confirms ancestral genomes of *Avena* species and identifies D-genome chromosomes of hexaploid oat. *Theor. Appl. Genet.* **129,** 2133–2149 (2016).
11. Peng, Y. Y. *et al.* Phylogenetic investigation of *Avena* diploid species and the maternal genome donor of *Avena* polyploids. *Taxon* **59,** 1472–1482 (2010).
12. Yan H. H. *et al.* Genome size variation in the genus *Avena*. *Genome* **59,** 209–220 (2016).
13. Nikoloudakis, N., Skaracis, G. & Katsiotis, A. Evolutionary insights inferred by molecular analysis of the ITS1-5.8S-ITS2 and IGS *Avena* sp. sequences. *Mol. Phylogenet. Evol.* **46,** 102–115 (2008).
14. Peng, Y. Y., Wei, Y. M., Baum, B. R. & Zheng, Y. L. Molecular diversity of the 5S rRNA gene and genomic relationships in the genus *Avena* (Poaceae: Aveneae). *Genome* **51,** 137–154 (2008).
15. Ladizinsky, G. *Studies in oat evolution. A man's life with Avena*, 1–96 (Springer, Heidelberg, 2012).
16. Linares, C., Gónzalez, J., Ferrer, E. & Fominaya, A. The use of double FISH to physically map the positions of 5S rDNA genes in relation to the chromosomal location of 18S–5.8S–26S rDNA and a C genome specific DNA sequence in the genus *Avena*. *Genome* **39,** 535–542 (1996).
17. Fu, Y. B. & Williams, D. J. AFLP variation in 25 *Avena* species. *Theor. Appl. Genet.* **117,** 333–342 (2008).
18. Peng, Y. Y. *et al.* Phylogenetic inferences in *Avena* based on analysis of *FL intron2* sequences. *Theor. Appl. Genet.* **121,** 985–1000 (2010).
19. Chen, Q. F. & Armstrong, K. Genomic *in situ* hybridization in *Avena sativa*. *Genome* **37,** 607–612 (1994).
20. Katsiotis, A., Hagidimitriou, M. & Heslop-Harrison, J. S. Chromosomal and genomic organization of *Ty1-copia*-like retrotransposon sequences in the genus *Avena*. *Genome* **39,** 410–417 (1996).
21. Ladizinsky, G. A new species of *Avena* from Sicily, possibly the tetraploid progenitor of hexaploid oats. *Genet. Resour. Crop. Evol.* **45,** 263–269 (1998).
22. Ladizinsky, G. Domestication via hybridization of the wild tetraploid oats *Avena magna* and *A. murphyi*. *Theor. Appl. Genet.* **91,** 639–646 (1995).

23. Mittermeier, R. A., Robles Gil, P., Hoffmann, M. & de Fonseca, G. A. B. *Hotspots revisited: Earth's biologically richest and most endangered ecoregions*, 1–389 (CEMEX, Mexico City, 2004).
24. Clayton, W. D., Vorontsova, M. S., Harman, K. T. & Williamson, H. *GrassBase-The Online World Grass Flora*. (2006 onwards) Available at: http://www.kew.org/data/grasses-db.html. (Accessed: 8th November 2006).
25. Germeier, C., Maggioni, L., Katsiotis, A. & Lipman, E. Report of a working group on *Avena*. Sixth Meeting, jointly held with the Final Meeting of project AGRI GEN RES 061 on "*Avena* Genetic Resources for Quality in Human Consumption" (AVEQ), Bucharest, Romania. Bioversity International, Rome, Italy (2011).
26. Garfunkel, Z. Origin of the eastern Mediterranean basin: a reevaluation. *Tectonophysics* **391,** 11–34 (2004).
27. Thompson, J. D. *Plant evolution in the Mediterranean*, 1–293 (Oxford University Press Inc., New York, 2005).
28. Lo Presti, R. M. & Oberprieler, C. Evolutionary history, biogeography and eco-climatological differentiation of the genus *Anthemis* L. (Compositae, Anthemideae) in the circum-Mediterranean area. *J. Biogeogr.* **36,** 1313–1332 (2009).
29. Inda, L. A., Sanmartín, I., Buerki, S. & Catalán, P. Mediterranean origin and Miocene–Holocene Old World diversification of meadow fescues and ryegrasses (*Festuca* subgenus *Schedonorus* and *Lolium*). *J. Biogeogr.* **41,** 600–614 (2014).
30. Nikoloudakis, N. & Katsiotis, A. Comparative molecular and cytogenetic methods can clarify meiotic incongruities in *Avena* allopolyploid hybrids. *Caryologia* **68,** 84–91 (2015).
31. Liu, Q., Liu, H., Wen, J. & Peterson, P. M. Infrageneric phylogeny and temporal divergence of *Sorghum* (Andropogoneae, Poaceae) based on low-copy nuclear and plastid sequences. *PLoS One* **9,** e104933 (2014).
32. Lin, L., Zeng, F. Y. & Liu, Q. Caryopsis micromorphological characteristics of *Avena* (Poaceae) and its taxonomic significances. *J. Trop. Subtrop. Bot.* **24,** 1–13 (2016).
33. Allard, R. W. *Principles of plant breedings* (*2nd ed.*), 1–254 (John Willey & Sons, Inc., New York, 1999).
34. Leggett, J. M. Interspecific hybrids involving the recently described diploid taxon *Avena atlantica*. *Genome* **29,** 361–364 (1987).
35. Heiser, C. B. Introgression re-examined. *Botanical Review* **39,** 347–366 (1973).
36. Govindarajulu, R., Hughes, C. E., Alexander, P. J. & Bailey, C. D. The complex evolutionary dynamics of ancient and recent polyploidy in *Leucaena* (Leguminosae; Mimosoideae). *Am. J. Bot.* **98,** 2064–2076 (2011).
37. Gill, B. S. Nucleocytoplasmic interaction (NCI) hypothesis of genome evolution and speciation in polyploid plants In *Kihara memorial international symposium on cytoplasmic engineering in wheat* (eds Sasakuma, T. & Kinoshita, T.) 48–53 (Kihara Memorial Foundation, Yokohama, 1991).
38. Ladizinsky, G. & Zohary, D. Notes on species delimitation, species relationships and polyploidy in *Avena*. *Euphytica* **20,** 380–395 (1971).
39. Wen, J., Liu, J. Q., Ge, S., Xiang, Q. Y. & Zimmer, E. A. Phylogenomic approaches to deciphering the tree of life. *J Syst. Evol.* **53,** 369–370 (2015).
40. Zimmer, E. A. & Wen, J. Using nuclear gene data for plant phylogenetics: progress and prospects II. Next-gen approaches. *J Syst. Evol.* **53,** 371–379 (2015).
41. Bacon, C. D., Baker, W. J. & Simmons, M. P. Miocene dispersal drives island radiations in the palm tribe Trachycarpeae (Arecaceae). *Syst. Biol.* **61,** 426–442 (2012).
42. Estep, M. C. *et al.* Allopolyploidy, diversification, and the Miocene grassland expansion. *Proc. Natl. Acad. Sci. USA* **111,** 15149–15154 (2014).
43. Blondel, J. The 'design' of Mediterranean landscapes: a millennial story of humans and ecological systems during the historic period. *Human Ecol.* **34,** 713–729 (2006).
44. Soreng, R. J. *et al.* A worldwide phylogenetic classification of the Poaceae (Gramineae). *J. Syst. Evol.* **53,** 117–137 (2015).
45. Matsuoka, M. & Hata, S. Comparative studies of phosphoenolpyruvate carbonxylase from C_3 and C_4 plants. *Plant Physiol.* **85,** 947–951 (1987).
46. Mason-Gamer, R. J., Weil, C. F. & Kellogg, E. A. Granule-bound starch synthase: structure, function, and phylogenetic utility. *Mol. Biol. Evol.* **15,** 1658–1673 (1998).
47. Ma, H. GTP-binding proteins in plants: new members of an old family. *Plant Mol. Biol.* **26,** 1611–1636 (1994).
48. Christin, P. A. *et al.* Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. *Curr. Biol.* **18,** 37–43 (2008).
49. Zhu, Q. H. & Ge, S. Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. *New Phytol.* **167,** 249–265 (2005).
50. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
51. Felsenstein, J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17,** 368–376 (1981).
52. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61,** 539–542 (2012).
53. Posada, D. & Crandall, K. A. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14,** 817–818 (1998).
54. Zwickl, D. J. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD thesis, 1–135 (University of Texas at Austin, Austin, 2006).
55. Swofford, D. L. *PAUP*. Phylogenetic analysis using parsimony (* and other methods), version 4.0b10.* (Sinauer Associates, Sunderland, 2003).
56. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29,** 1969–1973 (2012).
57. Prasad, V. *et al.* Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. *Nat. Commun.* **2,** 480 (2011).
58. Nylander, J. A. A., Wilgenbusch, J. C., Warren, D. L. & Swofford, D. L. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics* **24,** 581–583 (2008).

Acknowledgements

The research was supported by National Natural Science Foundation of China to Q.L.(31270275), Special Basic Research Foundation of Ministry of Science and Technology of the People's Republic of China to Q.L. (2013FY112100), Chinese Academy of Sciences President's International Fellowship Initiative to Q.L. (2016VBA010), China Scholarship Council Awards to Q.L. (201604910096), Undergraduate Innovation Training Program of Chinese Academy of Sciences (27), and the Endowment Grant Program of the Smithsonian Institution to J.W. The authors thank CN-Saskatchewan, ILRI-Addis Ababa, and USDA-Beltsville Germplasm System for seeds.

Author Contributions

Q.L, P.M.P, and J.W. conceived the research. Q.L., L.L., and X.Y.Z. conducted experiments, analyzed the data, and wrote the manuscript. P.M.P. and J.W. checked the final manuscript. All authors approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Liu, Q. *et al.* Unraveling the evolutionary dynamics of ancient and recent polypoidization events in Avena (Poaceae). *Sci. Rep.* **7**, 41944; doi: 10.1038/srep41944 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information

Unraveling the evolutionary dynamics of ancient and recent polyoidization events in *Avena* (Poaceae)

Qing Liu[1], Lei Lin[1,2], Xiangying Zhou[1,2], Paul M. Peterson[3], and Jun Wen[3]

[1]Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China;

[2]University of Chinese Academy of Sciences, Beijing 100049, China;

[3]Department of Botany, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013-7012, USA.

Correspondence and requests for materials should be addressed to Q.L. (liuqing@scib.ac.cn) or J.W. (wenj@si.edu)

Table of Contents

Supplemental Figure S9. Bayesian inference phylogeny of clade AB-GBI of *Avena* inferred from nuclear *GBSSI* data (Figure 4).

Supplemental Figure S10. Bayesian inference phylogeny of clade AB-GBII of *Avena* inferred from nuclear *GBSSI* data (Figure 4).

Supplemental Figure S11. Maximum likelihood phylogeny of *Avena* inferred from nuclear *gpa1* data (Figure 5).

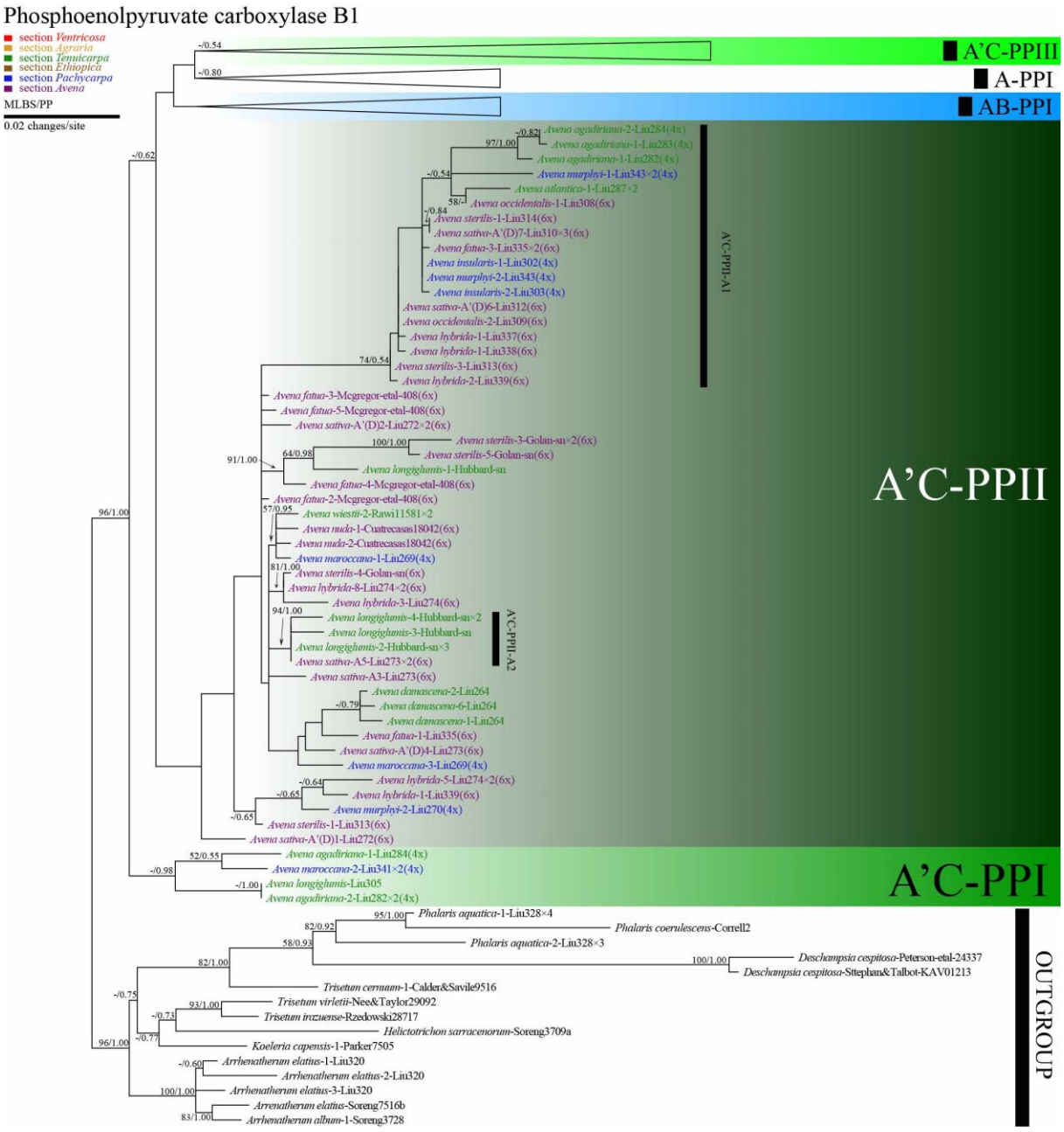
Supplemental Table S1. Taxa included in this study

Supplemental Table S2. DNA primers and PCR parameters used for amplification and sequencing

Supplemental Table S3. Statistics and evolutionary models for separate data partitions

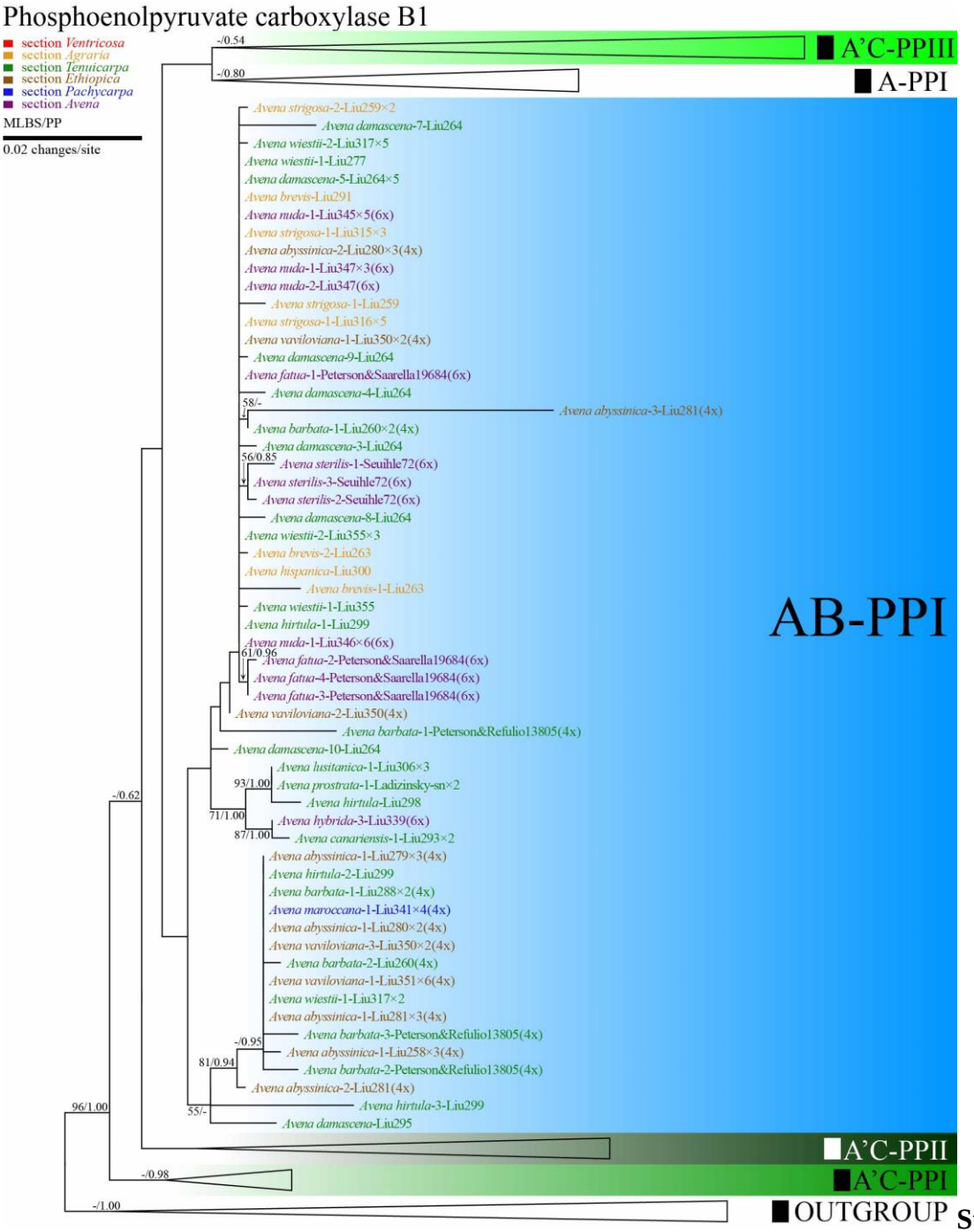
Supplemental Table S4. Posterior age distributions of major nodes in *Avena*

Supplemental Table S5. Potential paternal parents for *Avena sativa*

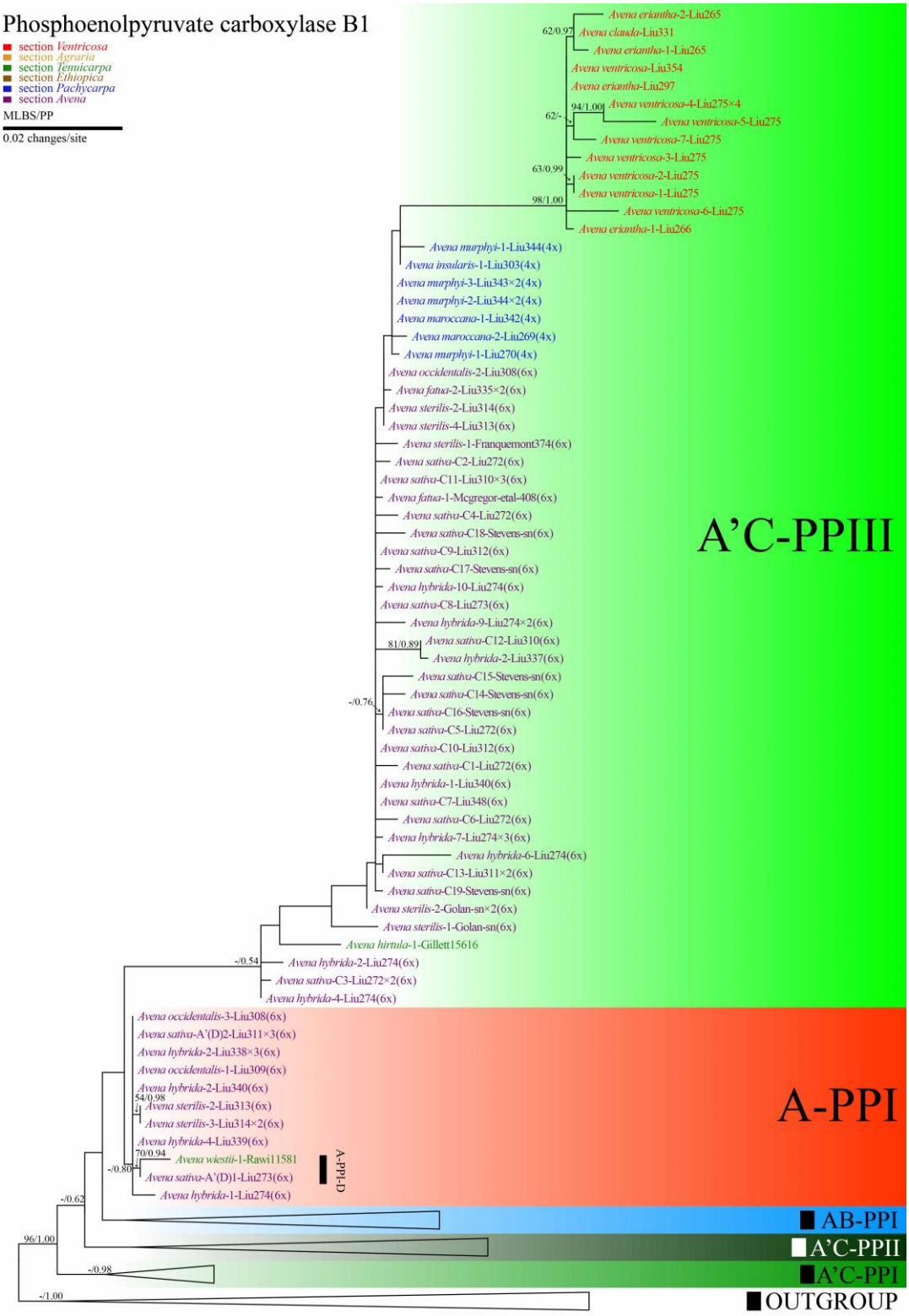


Supplementary Figure S1. Maximum likelihood phylogeny of clade A'C-PPI and node A'C-PPII of *Avena* inferred from nuclear *ppcB1* data (Figure 2). Numbers above branches are maximum likelihood bootstrap support/Bayesian posterior probability (MLBS/PP). Taxon labels are in the format: *Avena nuda*-1-Liu346×6 (6x) where *Avena nuda* indicates that the sequence belongs to the species; -1- = the first sequence cloned in Table S1 for this species; Liu346 indicates voucher; ×6 indicates we recovered 6 clones for the sequence; (6x) indicates the species is hexaploid; the absence of a mark between species name and voucher indicates that the sequence is derived from PCR-direct sequencing; the absence of a mark after voucher

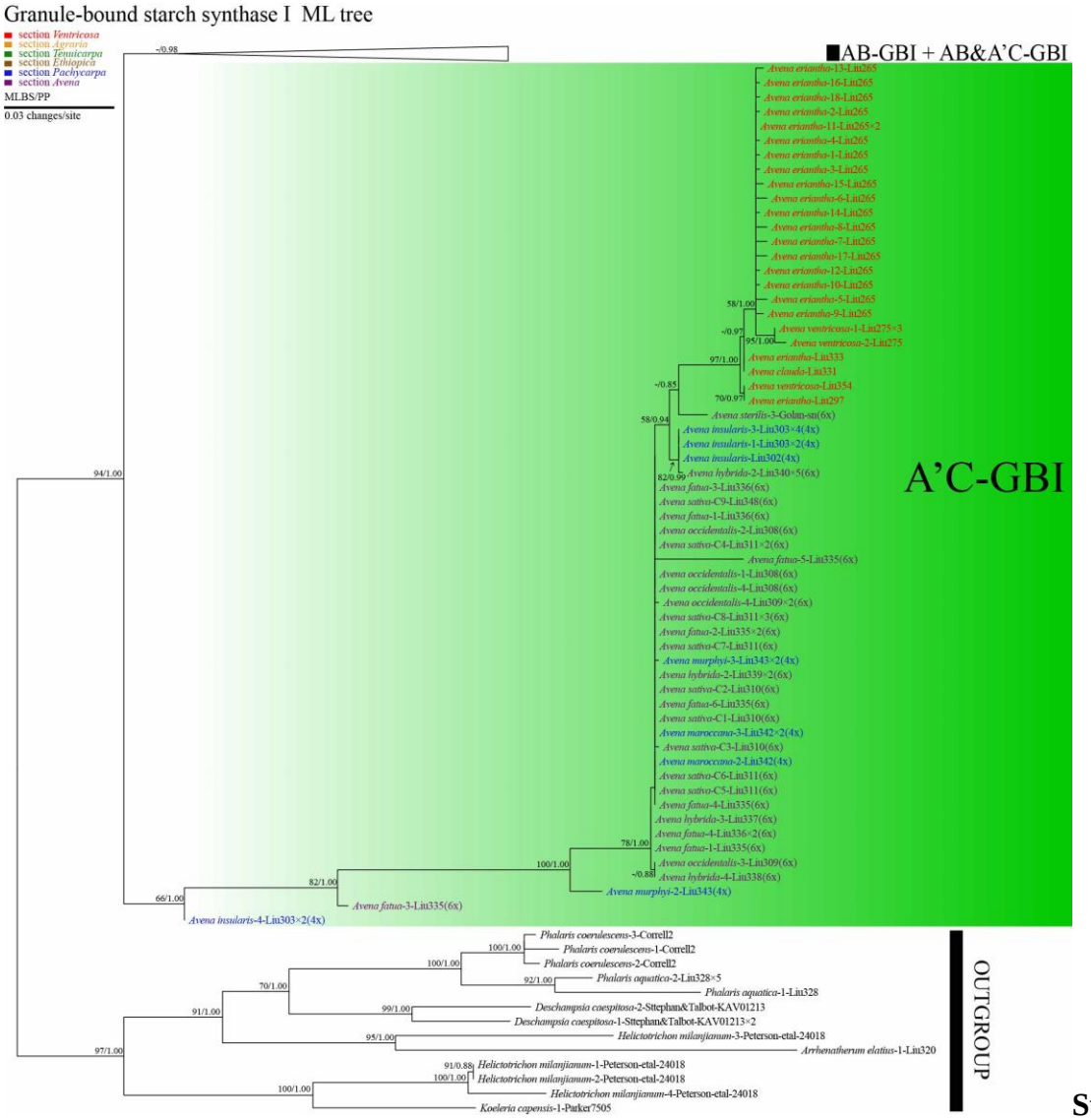
indicates that only one sequence for the diploid species was recovered. Coloured taxon labels correspond to sections listed at the top left corner of the figure.



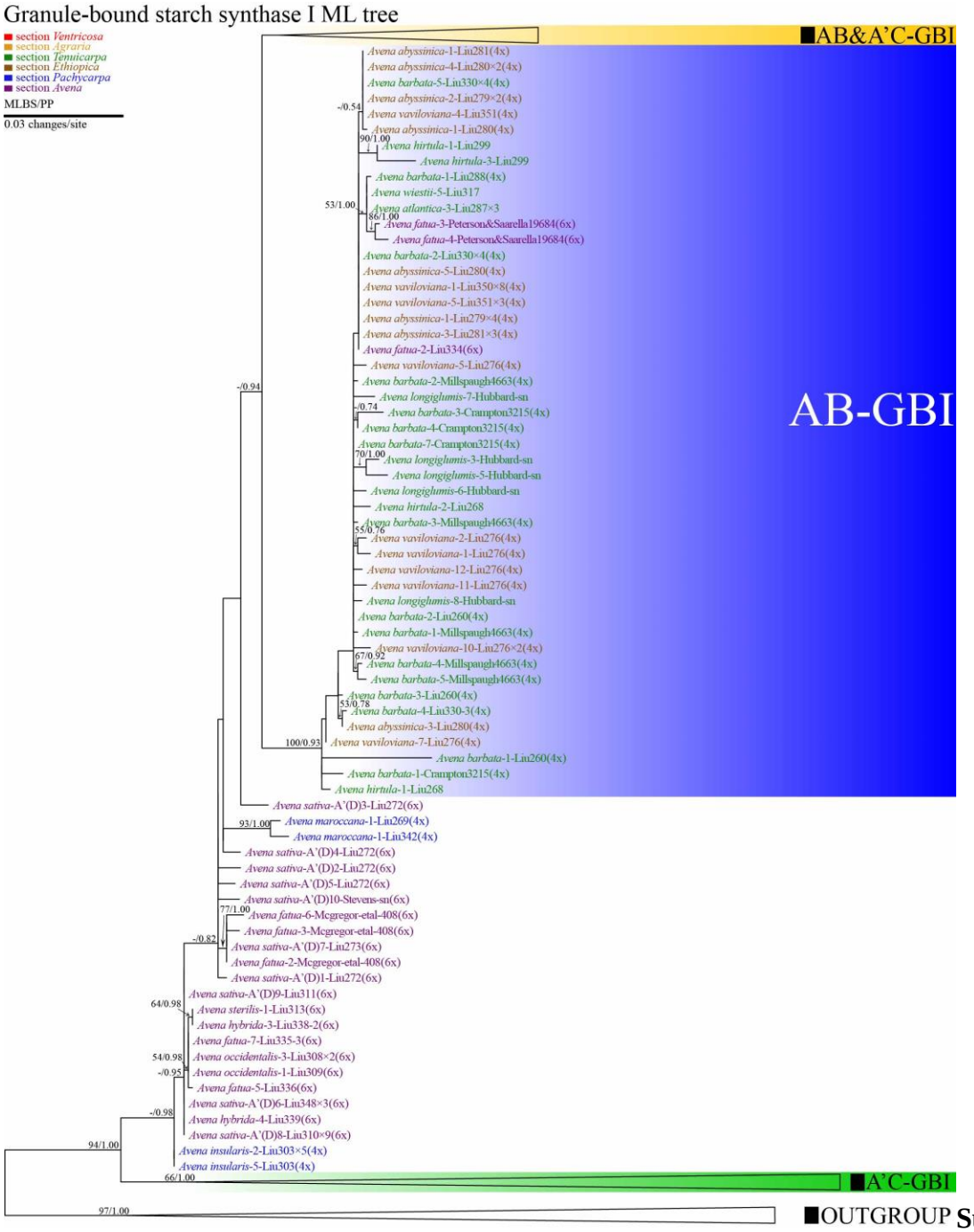
Supplementary Figure S2. Maximum likelihood phylogeny of node AB-PPI of *Avena* inferred from nuclear *ppcB1* data (Figure 2). Numbers above branches are MLBS/PP. Taxon labels are in the same format as in Figure S1. Coloured taxon labels correspond to sections listed at the top left corner of the figure.



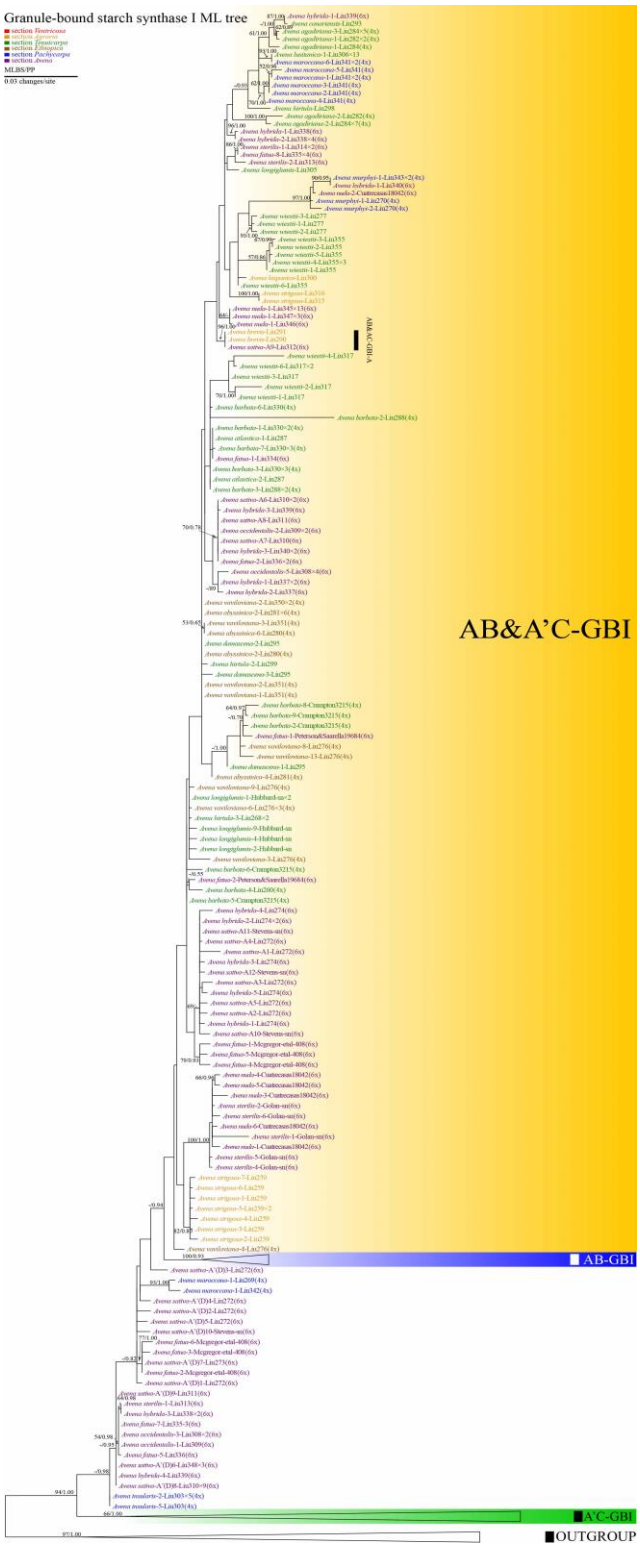
Supplementary Figure S3. Maximum likelihood phylogeny of clade A'C-PPII of *Avena* inferred from nuclear *ppcB1* data (Figure 2). Numbers above branches are MLBS/PP. Taxon labels are in the same format as in Figure S1. Coloured taxon labels correspond to sections listed at the top left corner of the figure.



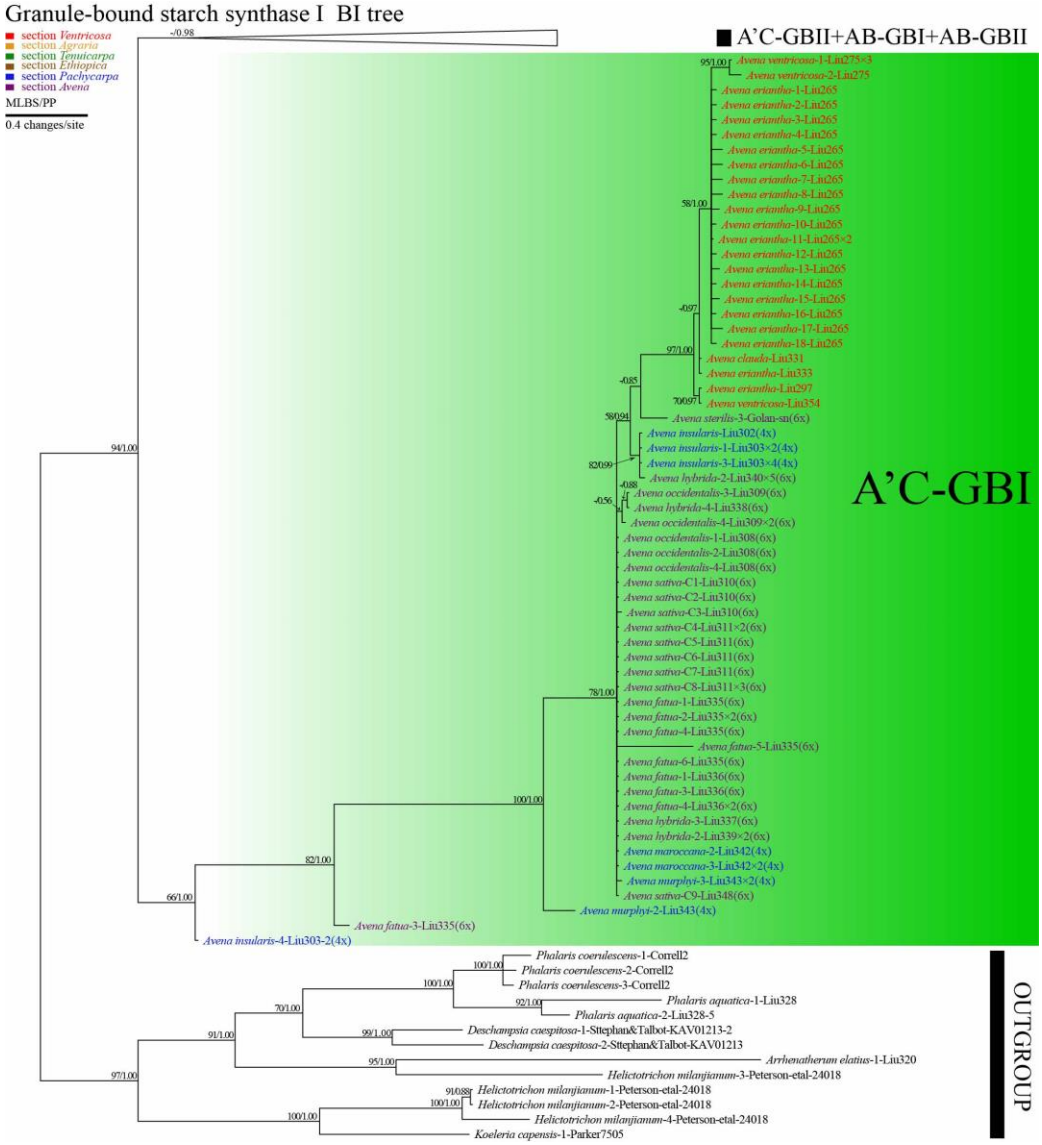
Supplementary Figure S4. Maximum likelihood phylogeny of clade A'C-GBI of *Avena* inferred from nuclear *GBSSI* data (Figure 3). Numbers above branches are MLBS/PP. Taxon labels are in the same format as in Figure S1. Coloured taxon labels correspond to sections listed at the top left corner of the figure.



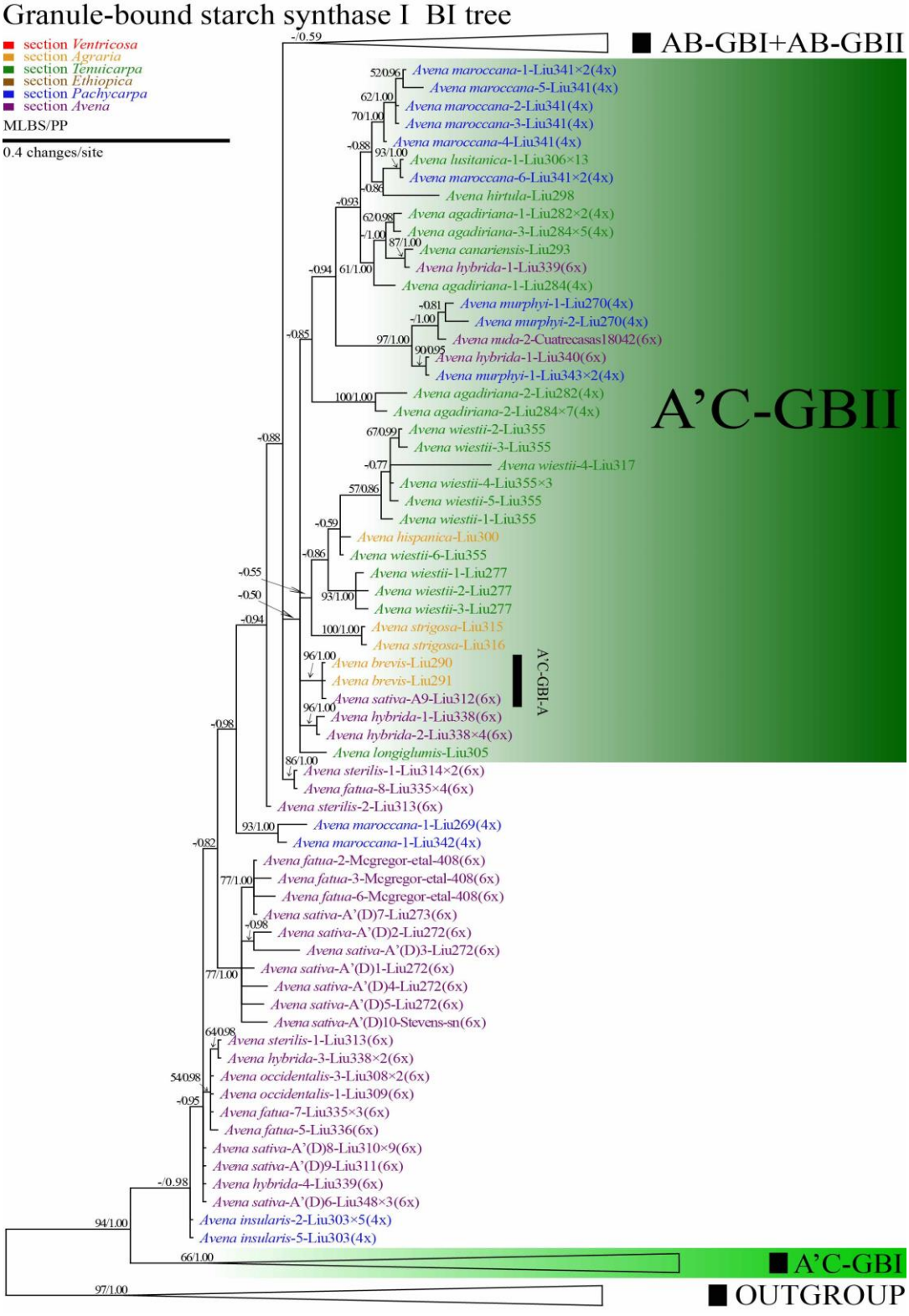
Supplementary Figure S5 Maximum likelihood phylogeny of clade AB-GBI of *Avena* inferred from nuclear *GBSSI* data (Figure 3). Numbers above branches are MLBS/PP. Taxon labels are in the same format as in Figure S1. Coloured taxon labels correspond to sections listed at the top left corner of the figure.



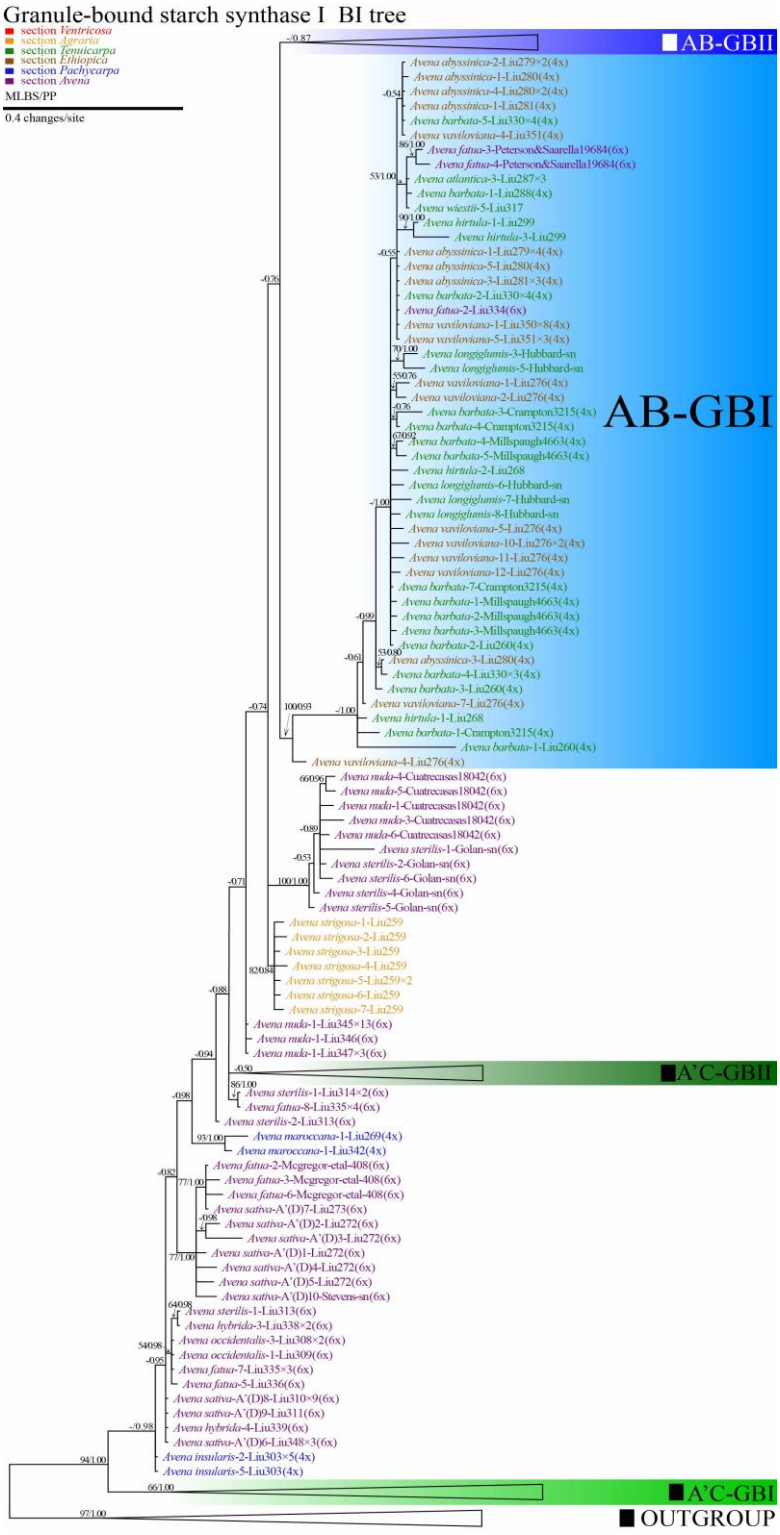
Supplementary Figure S6 Maximum likelihood phylogeny of clade AB&A'C-GBI of *Avena* inferred from nuclear *GBSSI* data (Figure 3). Numbers above branches are MLBS/PP. Taxon labels are in the same format as in Figure S1. Coloured taxon labels correspond to sections listed at the top left corner of the figure.



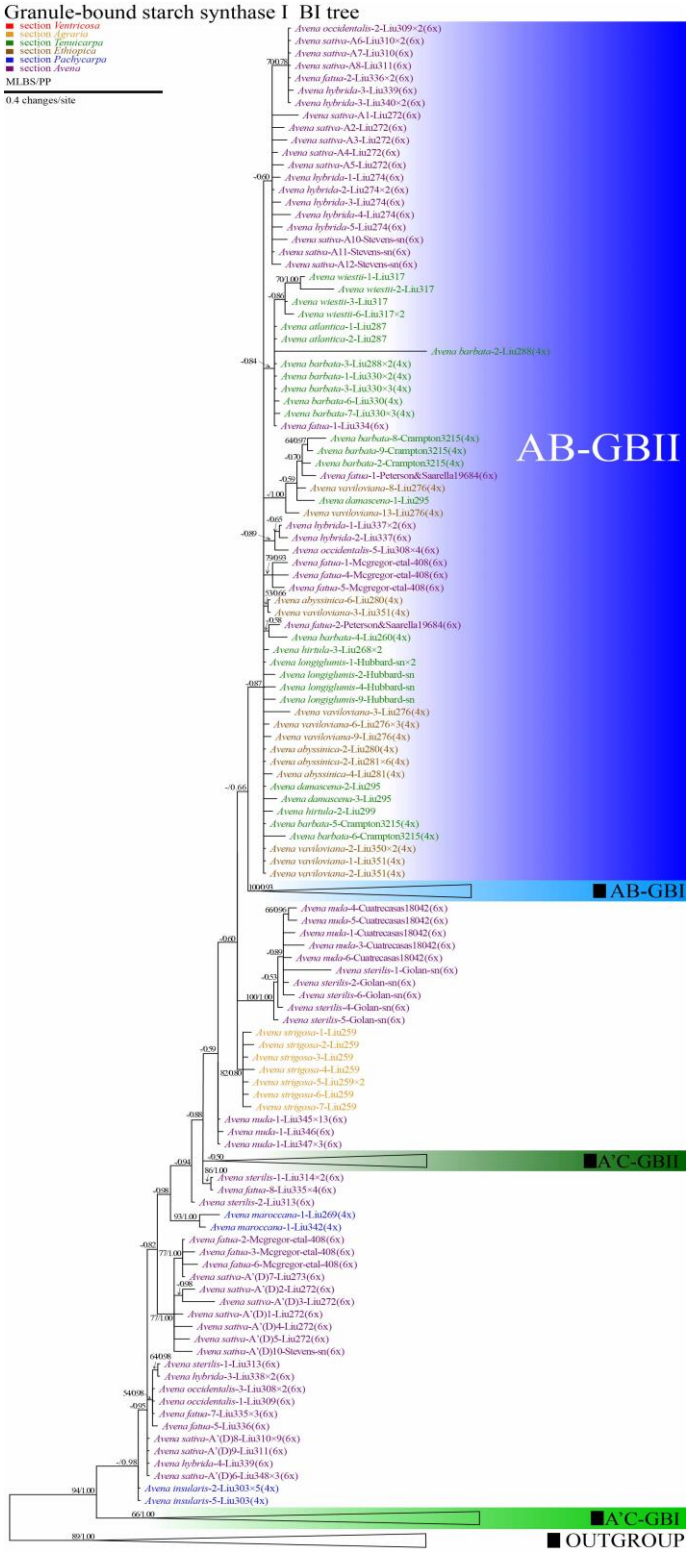
Supplementary Figure S7 Bayesian inference phylogeny of clade A'C-GBI of *Avena* inferred from nuclear *GBSSI* data (Figure 4). Numbers above branches are MLBS/PP. Taxon labels are in the same format as in Figure S1. Coloured taxon labels correspond to sections listed at the top left corner of the figure.



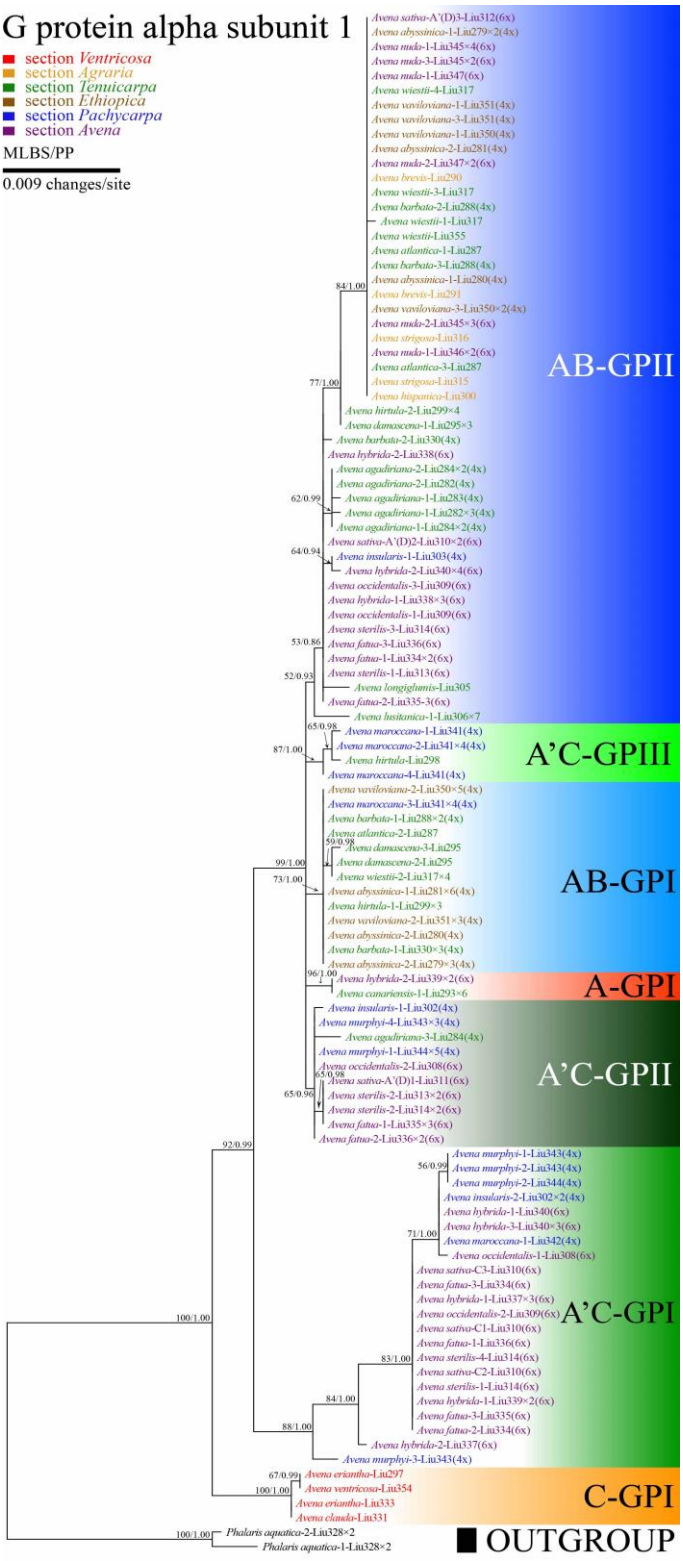
Supplementary Figure S8 Bayesian inference phylogeny of clade A'C-GBII of *Avena* inferred from nuclear *GBSSI* data (Figure 4). Numbers above branches are MLBS/PP. Taxon labels are in the same format as in Figure S1. Coloured taxon labels correspond to sections listed at the top left corner of the figure.



Supplementary Figure S9 Bayesian inference phylogeny of clade AB-GBI of *Avena* inferred from nuclear *GBSSI* data (Figure 4). Numbers above branches are MLBS/PP. Taxon labels are in the same format as in Figure S1. Coloured taxon labels correspond to sections listed at the top left corner of the figure.



Supplementary Figure S10 Bayesian inference phylogeny of clade AB-GBII of *Avena* inferred from nuclear *GBSSI* data (Figure 4). Numbers above branches are MLBS/PP. Taxon labels are in the same format as in Figure S1. Coloured taxon labels correspond to sections listed at the top left corner of the figure.



Supplementary Figure S11 Maximum likelihood phylogeny of *Avena* inferred from nuclear *gpa1* data (Figure 5). Numbers above branches are MLBS/PP. Taxon labels are in the same format as in Figure S1. Coloured taxon labels correspond to sections listed at the top left corner of the figure.

Supplementary Table S1. Taxa included in this study

Taxa	Voucher (Source)	Country of origin	GenBank accession numbers of LCN markers (*ppc-B1*; *GBSSI*; *gpa1*)	GenBank accession numbers of chloroplastid regions (*ndhA* intron; *rpl32-trnL*; *rps16* intron)
Section *Ventricosa* Baum ex Romero-Zarco				
A. clauda Dur. (2*x* = 14; C_pC_p)	*Liu 331* (*CN 19201*; IBSC)	Canada	KT723587; KT723827; KT723996	KT452977; KT453085; KT453177
A. eriantha Dur. (2*x* = 14; C_pC_p)	*Liu 265* (*CIav 9050*; IBSC)	United Kingdom	**1 KT723655**, **2 KT723656**; **1 KT723917**, **2 KT723918**, **3 KT723919**, **4 KT723920**, **5 KT723921**, **6 KT723922**, **7 KT723923**, **8 KT723924**, **9 KT723925**, **10 KT723926**, **11 KT723927**, **12 KT723928**, **13 KT723929**, **14 KT723930**, **15 KT723931**, **16 KT723932**, **17 KT723933**, **18 KT723934**; -	KT453015; KT453034; KT453223
	Liu 266 (*PI 657576*; IBSC)	Morocco	**1 KT723465**; -; -	KT452937; KT453035; KT453122
	Liu 297 (*PI 657576*; IBSC)	Morocco	KT723542; KT723753; KT723959	KT452958; KT453061; KT453151
	Liu 333 (*CIav 9050*;	United Kingdom	-; KT723828; KT723997	KT452978; KT453086; KT453178

	IBSC)			
A. ventricosa Balansa ex Coss. (2*x* = 14; CvCv)	*Liu 275* (*PI 657337*; IBSC)	Morocco	**1 KT723511**, **2 KT723512**, **3 KT723513**, **4 KT723514**, **5 KT723515**, **6 KT723516**, **7 KT723517**; **1 KT723706**, **2 KT723707**; -	KT452947; KT453044; KT453133
	Liu 354 (*PI 657337*; IBSC)	Morocco	KT723618; KT723887; KT724039	KT452998; KT453103; KT453197
Section *Agraria* Baum				
A. brevis Roth (2*x* = 14; A_sA_s)	*Liu 263* (*CIav 1783*; IBSC)	Germany	**1 KT723643**, **2 KT723644**; -; -	KT452990; KT453119; KT453221
	Liu 289 (*CN 1979*; IBSC)	Canada	-; - ; -	KT452987; KT453055; KT453145
	Liu 290 (*CN 3075*; IBSC)	Russian Federation	-; KT723747; KT723953	KT452988; KT453057; KT453147
	Liu 291 (*CIav 1783*; IBSC)	Germany	KT723539; KT723747; KT723954	KT452989; KT453058; KT453148
A. hispanica Ard. (2*x* = 14; A_sA_s)	*Liu 300* (*CN 25675*; IBSC)	Portugal	KT723549; KT723765; KT723963	KT452961; KT453065; KT453156
A. strigosa Schreb (2*x*	*Liu 259* (*PI 401794*;	United Kingdom	**1 KT723547**, **2 KT723548**; **1 KT723758**, **2**	KT452960; -; KT453155

= 14; A_sA_s)	IBSC)		**KT723759**, **3 KT723760**, **4 KT723761, 5 KT723762**, **6 KT723763**, **7 KT723764**; -	
	Liu 315 (*CN 21993*; IBSC)	Portugal	**1 KT723578**; KT723800; KT723986	KT452972; KT453077; KT453169
	Liu 316 (*CN 36500*; IBSC)	Canada	**1 KT723579**; KT723801; KT723987	KT452973; KT453078; KT453170
	Liu 349 (*PI 401794*; IBSC)	United Kingdom	-; - ; -	KT452995; -; KT453194
Section *Tenuicarpa* Baum				
A. agadiriana Baum & Fedak (4*x* = 28; AABB)	*Liu 262* (*PI 657585*; IBSC)	Morocco	-; - ; -	-; KT453118; KT453220
	Liu 282 (*CN 25823*; IBSC)	Morocco	**1 KT723531**, **2 KT723532**; **1 KT723736**, **2 KT723737**; **1 KT723941**, **2 KT723942**	KT452953; -; KT453141
	Liu 283 (*CN 25868*; IBSC)	Morocco	**1 KT723533**; -; **1 KT723943**	-; - ; -
	Liu 284 (*PI 657585*; IBSC)	Morocco	**1 KT723534**, **2 KT723535**; **1 KT723738**, **2 KT723739**, **3 KT723740**; **1 KT723944**, **2**	KT452954; KT453052; KT453142

			KT723945, **3 KT723946**	
A. atlantica Baum & Fedak (2*x* = 14; A_sA_s)	*Liu 261* (*PI 657393*; IBSC)	Morocco	-; - ; -	-; KT453112; KT453212
	Liu 287 (*PI 657393*; IBSC)	Morocco	**1 KT723536**; **1 KT723741**, **2 KT723742**, **3 KT723743**; **1 KT723947**, **2 KT723948**, **3 KT723949**	KT452955; KT453053; KT453143
A. barbata Pott ex Link (4*x* = 28; AABB)	*Liu 260* (*PI 282723*; IBSC)	Israel	**1 KT723633**, **2 KT723634**; **1 KT723902**, **2 KT723903**, **3 KT723904**, **4 KT723905**; -	KT453016; KT453110; KT453204
	Liu 288 (*CN 19357*; IBSC)	Iran	**1 KT723537**; **1 KT723744**, **2 KT723745**, **3 KT723746**; **1 KT723950**, **2 KT723951**, **3 KT723952**	KT453017; KT453054; KT453144
	Liu 330 (*PI 282723*; IBSC)	Israel	-; **1 KT723820**, **2 KT723821**, **3 KT723822**, **4 KT723823**, **5 KT723824**, **6 KT723825**, **7 KT723826**; **1 KT723994**, **2 KT723995**	KT453018; KT453084; KT453176
	Soderstrom 1479 (US)	Tunisia	-; - ; -	-; KT453064; KT453154
	Peterson & Refulio 13805 (US)	Peru	**1 KT723560**, **2 KT723561**, **3 KT723562**; -; -	-; -; KT453163
	Crampton 3215 (US)	USA	-; **1 KT723808**, **2 KT723809**, **3 KT723810**, **4 KT723811**, **5 KT723812**, **6 KT723813**, **7**	-; KT453080; KT453172

			KT723814, **8 KT723815**, **9 KT723816**; -	
	Wiggins 20450 (US)	USA	-; -; -	-; KT453083; KT453175
	Millspaugh 4663 (US)	USA	-; **1 KT723855**, **2 KT723856**, **3 KT723857**, **4 KT723858**, **5 KT723859**; -	-; KT453092; KT453184
A. canariensis Baum & Raj & Samp (2*x* = 14; AcAc)	*Liu 293* (*CN 23021*; IBSC)	Spain	**1 KT723540**; KT723749; **1 KT723955**	KT452956; KT453059; KT453149
A. damascena Rajah & Baum (2*x* = 14; A_dA_d)	*Liu 264* (*PI 657472*; IBSC)	Morocco	**1 KT723645**, **2 KT723646**, **3 KT723647**, **4 KT723648**, **5 KT723649**, **6 KT723650**, **7 KT723651**, **8 KT723652**, **9 KT723653**, **10 KT723654**; -; -	KT453014; KT453120; KT453222
	Liu 295 (*PI 657472*; IBSC)	Morocco	KT723541; **1 KT723750**, **2 KT723751**, **3 KT723752**; **1 KT723956**, **2 KT723957**, **3 KT723958**	KT452957; KT453060; KT453150
A. hirtula Lag. (2*x* = 14; AsAs)	*Liu 268* (*PI 657464*; IBSC)	Morocco	-; **1 KT723663**, **2 KT723664**, **3 KT723665**; -	KT452939; KT453037; KT453124
	Liu 298 (*CN 19738*; IBSC)	Algeria	KT723543; KT723754; KT723960	KT452959; KT453062; KT453152
	Liu 299 (*PI 657464*;	Morocco	**1 KT723544**, **2 KT723545**, **3 KT723546**; **1**	KT453021; KT453063; KT453153

	IBSC)		**KT723755**, **2 KT723756**, **3 KT723757**; **1 KT723961**, **2 KT723962**	
	Gillett 15616 (US)	Jordan	**1 KT723538**; -; -	-; KT453056; KT453146
A. longiglumis Dur. (2*x* = 14; AlAl)	*Liu 305* (*CN 21406*; IBSC)	Algeria	KT723553; KT723772; KT723967	KT452964; KT453068; KT453159
	Hubbard s.n. (US)	United Kingdom	**1 KT723471**, **2 KT723472**, **3 KT723473**, **4 KT723474**; **1 KT723666**, **2 KT723667**, **3 KT723668**, **4 KT723669**, **5 KT723670**, **6 KT723671**, **7 KT723672**, **8 KT723673**, **9 KT723674**; -	KT452940; KT453038; KT453038
	Soderstrom 1464 (US)	Tunisia	-; -; -	KT452950; KT453049; KT453138
A. lusitanica (Tab. Morais) Baum (2*x* = 14; AsAs)	*Liu 306* (*CN 26251*; IBSC)	Morocco	**1 KT723554**; **1 KT723773**; **1 KT723968**	KT452965; KT453069; KT453160
A. prostrata Ladiz. (2*x* = 14; A_pA_p)	*Ladizinsky s.n.* (K)	Spain	**1 KT723464**; -; -	KT452936; KT453026; KT453121
A. wiestii Steud. (2*x* = 14; AsAs)	*Liu 277* (*PI 53626*; IBSC)	Egypt	**1 KT723518**; **1 KT723721**, **2 KT723722**, **3 KT723723**; -	KT452949; KT453046; KT453135
	Liu 317 (*CN 19343*;	Iran	**1 KT723580**, **2 KT723581**; **1 KT723802**, **2**	KT452974; KT453079; KT453171

	IBSC)		**KT723803**, **3 KT723804**, **4 KT723805**, **5 KT723806**, **6 KT723807**; **1 KT723988**, **2 KT723989**, **3 KT723990**, **4 KT723991**	
	Liu 355 (*PI 53626*; IBSC)	Egypt	**1 KT723619**, **2 KT723620**; **1 KT723888**, **2 KT723889**, **3 KT723890**, **4 KT723891**, **5 KT723892**, **6 KT723893**; KT724040	KT452999; KT453104; KT453198
	Rawi 11581 (US)	Kuwait	**1 KT723523**, **2 KT723524**; -; -	-; -; -
Section *Ethiopica* Baum				
A. abyssinica Hochst. (4*x* = 28; AABB)	*Liu 258* (*PI 58728*; IBSC)	Ethiopia	**1 KT723505**; -; -	KT452945; KT453043; KT453131
	Liu 279 (*CN 3971*; IBSC)	Canada	**1 KT723525**; **1 KT723724**, **2 KT723725**; **1 KT723935**, **2 KT723936**	-; -; -
	Liu 280 (*CN 22051*; IBSC)	Ethiopia	**1 KT723526**, **2 KT723527**; **1 KT723726**, **2 KT723727**, **3 KT723728**, **4 KT723729**, **5 KT723730**, **6 KT723731**; **1 KT723937**, **2 KT723938**	KT452951; KT453050; KT453139
	Liu 281 (*PI 58728*; IBSC)	Ethiopia	**1 KT723528**, **2 KT723529**, **3 KT723530**; **1 KT723732**, **2 KT723733**, **3 KT723734**, **4 KT723735**; **1 KT723939**, **2 KT723940**	KT452952; KT453051; KT453140

A. vaviloviana (Malz.) Mordv (4*x* = 28; AABB)	*Liu 276* (*PI 412766*; IBSC)	Ethiopia	-; **1 KT723708**, **2 KT723709**, **3 KT723710**, **4 KT723711**, **5 KT723712**, **6 KT723713**, **7 KT723714**, **8 KT723715**, **9 KT723716**, **10 KT723717**, **11 KT723718**, **12 KT723719**, **13 KT723720**; -	KT452948; KT453045; KT453134
	Liu 350 (*CN 22004*; IBSC)	Ethiopia	**1 KT723614**, **2 KT723615**, **3 KT723616**; **1 KT723880**, **2 KT723881**; **1 KT724033**, **2 KT724034**, **3 KT724035**	KT452996; KT453101; KT453195
	Liu 351 (*PI 412766*; IBSC)	Ethiopia	**1 KT723617**; **1 KT723882**, **2 KT723883**, **3 KT723884**, **4 KT723885**, **5 KT723886**; **1 KT724036**, **2 KT724037**, **3 KT724038**	KT452997; KT453102; KT453196
Section *Pachycarpa* Baum				
A. insularis Ladiz. [4*x* = 28; AACC(DDCC)]	*Liu 302* (*CN 19178*; IBSC)	Italy	**1 KT723550**; KT723766; **1 KT723964**, **2 KT723965**	KT452962; KT453066; KT453157
	Liu 303 (*CN 108634*; IBSC)	Tunisia	**1 KT723551**, **2 KT723552**; **1 KT723767**, **2 KT723768**, **3 KT723769**, **4 KT723770**, **5 KT723771**; **1 KT723966**	KT452963; KT453067; KT453158
A. maroccana Grand. [4*x* = 28;	*Liu 269* (*CIav 8330*; IBSC)	Morocco	**1 KT723475**, **2 KT723476**, **3 KT723477**; **1 KT723675**; -	KT453019; KT453039; KT453126

AACC(DDCC)]				
	Liu 341 (*CN 21862*; IBSC)	Morocco	**1 KT723601**, **2 KT723602**; **1 KT723863**, **2 KT723864**, **3 KT723865**, **4 KT723866**, **5 KT723867**, **6 KT723868**; **1 KT724016**, **2 KT724017**, **3 KT724018**, **4 KT724019**	KT453022; KT453094; KT453186
	Liu 342 (*CIav 8330*; IBSC)	Morocco	**1 KT723603**; **1 KT723869**, **2 KT723870**, **3 KT723871**; **1 KT724020**	KT453023; KT453095; KT453187
A. murphyi Ladiz. [4*x* = 28; AACC(DDCC)]	*Liu 270* (*PI 657606*; IBSC)	Morocco	**1 KT723478**, **2 KT723479**; **1 KT723676**, **2 KT723677**; -	KT452941; KT453040; KT453127
	Liu 343 (*CN 21989*; IBSC)	Spain	**1 KT723604**, **2 KT723605**, **3 KT723606**; **1 KT723872**, **2 KT723873**, **3 KT723874**; **1 KT724021**, **2 KT724022**, **3 KT724023**, **4 KT724024**	KT452986; KT453096; KT453188
	Liu 344 (*PI 657606*; IBSC)	Morocco	**1 KT723607**, **2 KT723608**; -; **1 KT724025**, **2 KT724026**	KT452991; KT453097; KT453189
Section *Avena*				
A. fatua L. (6*x* = 42; AACCDD)	*Liu 334* (*CN 3214*; IBSC)	Australia	-; **1 KT723829**, **2 KT723830**; **1 KT723998**, **2 KT723999**, **3 KT724000**	KT452979; -; KT453179
	Liu 335 (*CN 3228*;	Australia	**1 KT723588**, **2 KT723589**, **3 KT723590**; **1**	KT452980; KT453087; KT453180

	IBSC)		**KT723831**, **2 KT723832**, **3 KT723833**, **4 KT723834**, **5 KT723835**, **6 KT723836**, **7 KT723837**, **8 KT723838**; **1 KT724001**, **2 KT724002**, **3 KT724003**	
	Liu 336 (*PI 544659*; IBSC)	USA	-; **1 KT723839**, **2 KT723840**, **3 KT723841**, **4 KT723842**, **5 KT723843**; **1 KT724004**, **2 KT724005**, **3 KT724006**	KT452981; KT453088; KT453181
	Mcgregor et al. 408 (US)	Mexico	**1 KT723466**, **2 KT723467**, **3 KT723468**, **4 KT723469**, **5 KT723470**; **1 KT723657**, **2 KT723658**, **3 KT723659**, **4 KT723660**, **5 KT723661**, **6 KT723662**; -	KT452938; KT453036; KT453123
	Peterson & Saarella 19684 (US)	USA	**1 KT723627**, **2 KT723628**, **3 KT723629**, **4 KT723630**; **1 KT723898**, **2 KT723899**, **3 KT723900**, **4 KT723901**; -	KT453001; KT453106; KT453200
A. hybrida Peterm. (6*x* = 42; AACCDD)	*Liu 274* (*PI 458784*; IBSC)	United Kingdom	**1 KT723495**, **2 KT723496**, **3 KT723497**, **4 KT723498**, **5 KT723499**, **6 KT723500**, **7 KT723501**, **8 KT723502**, **9 KT723503**, **10 KT723500**; **1 KT723695**, **2 KT723696**, **3 KT723697**, **4 KT723698**, **5 KT723699**; -	KT452944; -; KT453130

	Liu 337 (*CN 24885*; IBSC)	Iran	**1 KT723591**, **2 KT723592**; **1 KT723844**, **2 KT723845**, **3 KT723846**; **1 KT724007**, **2 KT724008**	KT452982; KT453089; KT453182
	Liu 338 (*CN 24926*; IBSC)	Iran	**1 KT723593**, **2 KT723594**; **1 KT723847**, **2 KT723848**, **3 KT723849**, **4 KT723850**; **1 KT724009**, **2 KT724010**	KT452983; KT453090; -
	Liu 339 (*PI 458778*; IBSC)	United Kingdom	**1 KT723595**, **2 KT723596**, **3 KT723597**, **4 KT723598**; **1 KT723851**, **2 KT723852**, **3 KT723853**, **4 KT723854**; **1 KT724011**, **2 KT724012**	KT452984; KT453091; KT453183
	Liu 340 (*PI 458784*; IBSC)	United Kingdom	**1 KT723599**, **2 KT723600**; **1 KT723860**, **2 KT723861**, **3 KT723862**; **1 KT724013**, **2 KT724014**, **3 KT724015**	KT452985; KT453093; KT453185
A. nuda L. (6*x* = 42; AACCDD)	*Liu 345* (*CN 79351*; IBSC)	Netherlands	**1 KT723609**; **1 KT723875**; **1 KT724027**, **2 KT724028**, **3 KT724029**	KT452992; KT453098; KT453190
	Liu 346 (*CN 79386*; IBSC)	Germany	**1 KT723610**; **1 KT723876**; **1 KT724030**	KT452993; -; KT453191
	Liu 347 (*CIav 9008*; IBSC)	Czechoslovakia	**1 KT723611**, **2 KT723612**; **1 KT723877**; **1 KT724031**, **2 KT724032**	KT452994; KT453099; KT453192

	Cuatrecasas 18042 (US)	Colombia	**1 KT723480**, **2 KT723481**; **1 KT723678**, **2 KT723679**, **3 KT723680**, **4 KT723681**, **5 KT723682**, **6 KT723683**; -	KT452942; KT453041; KT453128
A. occidentalis Dur. (6*x* = 42; AACCDD)	*Liu 308* (*CN 4538*; IBSC)	Spain	**1 KT723555**, **2 KT723556**, **3 KT723557**; **1 KT723774**, **2 KT723775**, **3 KT723776**, **4 KT723777**, **5 KT723778**; **1 KT723969**, **2 KT723970**	KT452966; KT453070; KT453161
	Liu 309 (*CN 21473*; IBSC)	Greece	**1 KT723558**, **2 KT723559**; **1 KT723779**, **2 KT723780**, **3 KT723781**, **4 KT723782**; **1 KT723971**, **2 KT723972**, **3 KT723973**	KT452967; KT453071; KT453162
A. sativa L. (6*x* = 42; AACCDD)	*Liu 272* (*PI 401777*; IBSC)	Poland	**A' (D)1 KT723486**, **A' (D)2 KT723489**, **C1 KT723482**, **C2 KT723483**, **C3 KT723484**, C**4 KT723485**, **C5 KT723487**, **C6 KT723488**; **A1 KT723684**, **A2 KT723686**, **A3 KT723688**, **A4 KT723689**, **A5 KT723690**, **A'(D)1 KT723685**, **A'(D)2 KT723687**, **A'(D)3 KT723691**, **A'(D)4 KT723692**, **A'(D)5 KT723693**; -	KT452943; KT453042; KT453129
	Liu 273 (*PI 51385*;	Spain	**A3 KT723490**, **A' (D)4 KT723493**, **A5**	-; -; -

IBSC)		**KT723494**, **C8 KT723491**, **A'(D)1 KT723492**; **A'(D)7 KT723694**; -	
Liu 348 (*PI 401777*; IBSC)	Poland	**C7 KT723613**; **C9 KT723879**, **A'(D)6 KT723878**; -	KT453024; KT453100; KT453193
Liu 312 (*PI 51385*; IBSC)	Spain	**A' (D)6 KT723568**, **C9 KT723569**, **C10 KT723570**; A9 KT723796; A'(D)3 KT723979	KT452969; KT453074; KT453166
Liu 310 (*CN 1876*; IBSC)	Canada	**A' (D)7 KT723565**, **C11 KT723563**, **C12 KT723564**; **A6 KT723783**, **A7 KT723788**, **C1 KT723784**, **C2 KT723785**, **C3 KT723787**, **A'(D)8 KT723786**; **A'(D)2 KT723977**, **C1 KT723974**, **C2 KT723975**, **C3 KT723976**	KT452968; KT453072; KT453164
Liu 311 (*CN 18136*; IBSC)	Canada	**C13 KT723567**, **A'(D)2 KT723566**; **A8 KT723794**, **C4 KT723790**, **C5 KT723791**, **C6 KT723792**, **C7 KT723793**, **C8 KT723795**, **A'(D)9 KT723789**; **A'(D)1 KT723978**	KT453020; KT453073; KT453165
Stevens s.n. (US)	USA	**C14 KT723621**, **C15 KT723622**, **C16**	KT453000; KT453105; KT453199

			KT723623, **C17 KT723624**, **C18 KT723625**, **C19 KT723626**; **A10 KT723895**, **A11 KT723896**, **A12 KT723897**, **A'(D)10 KT723894**; -	
A. sterilis L. (6*x* = 42; AACCDD)	*Liu 313* (*CN 3253*; IBSC)	Australia	**1 KT723571**, **2 KT723572**, **3 KT723573**, **4 KT723574**; **1 KT723797**, **2 KT723798**; **1 KT723980**, **2 KT723981**	KT453166; KT453075; KT453167
	Liu 314 (*CN 3375*; IBSC)	Australia	**1 KT723575**, **2 KT723576**, **3 KT723577**; **1 KT723799**; **1 KT723982**, **2 KT723983**, **3 KT723984**, **4 KT723985**	KT452971; KT453076; KT453168
	Golan s.n. (US)	Palaestinae	**1 KT723506**, **2 KT723507**, **3 KT723508**, **4 KT723509**, **5 KT723510**; **1 KT723700**, **2 KT723701**, **3 KT723702**, **4 KT723703**, **5 KT723704**, **6 KT723705**; -	KT452946; -; KT453132
	Seuihle 72 (US)	Jordan	**1 KT723519**, **2 KT723520**, **3 KT723521**; -; -	-; KT453047; KT453136
	Franquemont 374 (US)	Peru	**1 KT723522**; -; -	-; KT453048; KT453137
Outgroup				
Arrhenatherum album (Vahl) Clayton (2*x* =	*Soreng 3728* (US)	Spain	**1 KT723631**; -; -	KT453002; KT453107; KT453201

10,14)				
Arrhenatherum elatius (L.) P. Beauv. ex J. Presl & C. Presl (4*x* = 28)	*Liu 320* (*PI 249687*; US)	Spain	**1 KT723582**, **2 KT723583**, **3 KT723584**; **1 KT723817**; -	KT452975; KT453081; KT453173
	Soreng 7516b (US)	Greece	KT723632; -; -	-; KT453108; KT453202
Briza minor L. (2*x* = 14)	*Peterson et al. 9283* (US)	Ecuador	-; -; -	KT453003; KT453109; KT453203
Briza maxima L. (2*x* = 14)	*Beetle R605* (US)	Portugal	-; -; -	-; KT453111; KT453205
Deschampsia cespitosa (L.) P. Beauv. (2*x* = 26)	*Sttephan & Talbot KAV01213* (US)	USA	KT723635; **1 KT723906**, **2 KT723907**; -	-; -; -
	Peterson et al. 24337 (US)	Tanzania	KT723636; -; -	KT453025; KT453028; KT453206
Helictotrichon milanjianum (Rendle) C.E. Hubb. (2*x* = 26)	*Peterson et al. 24018* (US)	Tanzania	-; **1 KT723908**, **2 KT723909**, **3 KT723910**, **4 KT723911**; -	KT453005; KT453029; KT453208
	Peterson et al. 24034 (US)	Tanzania	-; -; -	KT453007; KT453031; KT453210

	Peterson et al. 24122 (US)	Tanzania	-; -; -	KT453008; -; KT453211
Helictotrichon pallens (Link) J.M. Couderc & Guédès (2*x* = 26)	*Peterson & Ollgaard 253* (US)	Denmark	-; -; -	KT453004; -; KT453207
Helictotrichon sarracenorum (Gand.) Holub. (2*x* = 14)	*Soreng 3709a* (US)	Spain	KT723637; -; -	KT453006; KT453030; KT453209
Koeleria capensis Nees (2*x* = 14)	*Peterson et al. 24336* (US)	USA	-; -; -	KT453010; KT453032; KT453214
	Parker 7505 (US)	USA	**1 KT723638**; **1 KT72 3912**; -	KT453009; KT453113; KT453213
Phalaris aquatica L. (2*x* = 28)	*Liu 328* (*PI 598922*; US)	Italy	**1 KT723585**, **2 KT723586**; **1 KT723818**, **2 KT723819**; **1 KT723992**, **2 KT723993**	KT452976; KT453082; KT453174
Phalaris angusta Nees ex Trin. (2*x* = 14)	*Scur 654* (US)	Brazil	-; -; -	KT453013; KT453117; KT453219
Phalaris coerulescens Desf. (2*x* = 14)	*Correll 2* (US)	Australia	KT723642; **1 KT723914**, **2 KT723915**, **3 KT723916**; -	-; KT453116; KT453218
Trisetum cernuum Trin. (4*x* = 28)	*Calder & Savile 9516* (US)	Canada	**1 KT723639**; -; -	-; KT453033; KT453215

Trisetum irazuense (Kuntze) Hitchc. (4*x* = 28)	*Rzedowski 28717* (US)	Mexico	KT723641; **1 KT723913**; -	KT453012; KT453115; KT453217
Trisetum virletii E. Fourn. (4*x* = 28)	*Nee & Taylor 29092* (US)	Mexico	KT723640; -; -	KT453011; KT453114; KT453216

Taxa: Chromosome numbers are based on http://mobot.mobot.org/W3T/Search/ipcn2.html; Genome assignment is based on Lin & Liu[5] and Nikoloudakis & Katsiotis[30]. Voucher (Source): CN, Plant Gene Resources at Saskatchewan, Canada; ILRI, International Livestock Research Institute at Addis Ababa, Ethiopia; PI or CIav, Germplasm Resources Information Network of United States Department of Agriculture at Beltsville, USA; IBSC, South China Botanical Garden Herbarium; K, Royal Botanic Gardens, Kew; US, United States National Herbarium; GenBank accession numbers of LCN markers (*ppc-B1*; *GBSSI*; *gpa1*) followed by sequence number (Prefix "A" or "B" or "A'(D)" indicates A- or B- or A'(D)-type homoeologue for polyploid species); interrupted line indicates unavailable sequence; Cloned sequences are labelled in bold.

Supplementary Table S2. Primers and PCR parameters used for amplification and sequencing. Chromosomal locations of nuclear genes are based on rice (*Oryza sativa* L.)

Region	**Location**	**Primers**	**Sequence (5'–3')**	**PCR parameters**	**Reference**
PpcB1	Chromosome 1	*PpcB1*-8F	AAG GCC CAG GAG GAG ATC GTG G	95 ℃/3 min; 16 × (94 ℃/20 s; 65 ℃ /40 s, -1 ℃/cycle; 72 ℃/90 s), 21 × (94 ℃/20 s; 50 ℃/40 s; 72 ℃/90 s); 72 ℃/5 min	This study
		PpcB1-9R	CAG CCG CTG CCT CAG GTA CGG GT		
GBSSI	Chromosome 6	*GBSSI* 9F	ATC GTC AAC GGC ATG GAC GTC	The same as above	This study
		GBSSI 14R	CAC GTC CTC CCA GTT CTT GGC		
gpa1	Chromosome 5	*gpa1*10F	GAG GAG RAA GTG GAT TCA TCT	The same as above	This study
		*gpa1*12R	ACC TCC TGT TTG YCA KGT GC		
ndhA intron	Plastid	*ndhA* intronF	CGC TAT TYC AAA ACC GTA CRT	95 ℃/5 min; 36 × (94 ℃/30 s; 53 ℃/60 s; 72 ℃/120 s); 72 ℃/8 min	This study
		ndhA intronR	CAA TAT CTC TAC GTG TGA TTC G		
rpl32-trnL	Plastid	*rpl32*F	CAG TTC CAA AAA AAC GTA CTT C	The same as above	Liu *et al.*, 2014[31]
		rpl32-trnL $^{(UAG)}$R	CTG CTT CCT AAG AGC AGC GT		

rps16 intron	Plastid	*rps16*F *rps16*R	TGT GGT ARA AAG CAA C AAC ATC WAT TGC AAS GAT TCG ATA	The same as above	Liu *et al.*, 2014[31]

Supplementary Table S3. Statistics and evolutionary models for separate data partitions. SL, aligned sequence length; GC, guanine and cytosine; PIC, parsimony informative characters; Ti/Tv, transition/transversion ratio; CI, consistency index excluding uninformative characters; RI, retention index

Partition	**No. of Sequences**	**SL**	**GC%**	**PIC**	**PIC/SL**	**Ti/Tv**	**CI**	**RI**	**Best-fit model**
ppcB1	193	1017	0.6226	220	0.2163	1.1290	0.2676	0.5333	GTR + I + G
GBSSI	277	1352	0.5662	434	0.3210	1.2274	0.4308	0.8864	TrN + G
gpa1	106	1034	0.3652	137	0.1321	1.5681	0.9227	0.9849	TVM + G
cpDNA	288	2819	0.3013	232	0.0823	0.9281	0.8187	0.9139	TVM + I + G

Supplementary Table S4. Posterior age distributions of major lineages in *Avena*. Lineage number in accordance with those in Figure 6; Lineage age is given by the mean age and the 95% highest posterior density (HPD) intervals in brackets; NA, not available

Lineage	Number	Stem age (Mya)	Crown age (Mya)
Avena	1	25.57 (NA)	20.04 (3.56–35.06)
C-NRR	2	20.04 (3.56–35.06)	10.71 (1.62–20.25)
A'C-NRR + AB-NRR	3	20.04 (3.56–35.06)	14.54 (2.68–25.02)
A'C-NRR	4	14.54 (2.68–25.02)	12.24 (NA)
The *A. sativa* lineage 5	5	3.82 (NA)	2.43 (NA)
The *A. satica* lineage 6	6	4.77 (NA)	2.46 (NA)
The *A. sativa* lineage 7	7	4.69 (NA)	2.97 (NA)
AB-NRR	8	14.54 (2.68–25.02)	10.88 (NA)

Supplementary Table S5. Potential paternal parents for *Avena sativa*. Bold species, strongly supported in present study (MLBS/PP); Underlined species, supported in previous studies[13,16,18,21,22]; Bold and underlined species, supported in present and previous studies

Species	2X (ML/BI)		4X (ML/BI)	
	CC genome	AA genome	AABB genome	A'A'CC genome
Avena sativa	***A. clauda***[18](100/1.00)	*A. atlantica* (84/1.00)	*A. abyssinica* (84/1.00)	*A. insularis*[21](83/1.00)
	A. eriantha (100/1.00)	***A. brevis*** (95/1.00)	*A. barbata* (84/1.00)	*A. maroccana*[22](83/1.00)
	A. ventricosa[13](100/1.00)	*A. damascena* (77/1.00)	*A. vaviloviana* (84/1.00)	***A. murphyi***[22](100/1.00)
		A. hirtula[16](77/1.00)		
		A. hispanica (84/1.00)		
		A. longiglumis[13](94/1.00)		
		A. strigosa[16](84/1.00)		
		A. wiestii[16](84/1.00)		