Analysis of phylogenetic relationships of *Hordeum* (Poaceae) polyploids

Dissertation

zur Erlangung des
Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der

Naturwissenschaftlichen Fakultät I – Biowissenschaften –

der Martin-Luther-Universität
Halle-Wittenberg,

vorgelegt

von Herrn Jonathan Brassac

geb. am 05.11.1986 in Lyon, Frankreich

Gutachter/in

1. Dr. Frank R. Blattner

2. Prof. Dr. Martin Röser

3. Prof. Dr. Christoph Oberprieler

Halle (Saale), verteidigt am 06. Mai 2015

CONTENTS

1. INTRODUCTION

THE GENUS *HORDEUM*

TAXONOMY

Hordeum L. belongs to the grass tribe Hordeeae. The name Hordeeae (Martinov, 1820), recently rediscovered, has priority over the widely used term Triticeae (Dumortier, 1824). The genus consists of 33 species and altogether 45 taxa are described (Bothmer *et al.*, 1995; Blattner, 2009). The key character to recognize *Hordeum* within the tribe is the possession of three single-flowered spikelets at each rachis node of the inflorescence (Fig. 1.1), the lateral ones being often sterile in the wild species (Bothmer *et al.*, 1995). The genus is distributed in arid and temperate regions of the world (Fig. 1.2) with two centers of diversity, one in Eurasia with 11 species and one in southern South America with 16 species (Bothmer *et al.*, 1995). Barley (*H. vulgare*) is the economically most important species of the genus with nearly 133 million tons produced in 2012 (FAOSTAT). Together with other species of the tribe Hordeeae such as wheat and rye, it is used for human nutrition, to feed livestock and malted for beer and whisky production. Due to its high economic importance and the potential of the wild relatives as gene pools for crop improvement (Xu & Kasha, 1992; Pickering *et al.*, 1995; Ruge *et al.*, 2003; Wendler *et al.*, 2014), relationships within the genus were intensively investigated (Blattner, 2009).

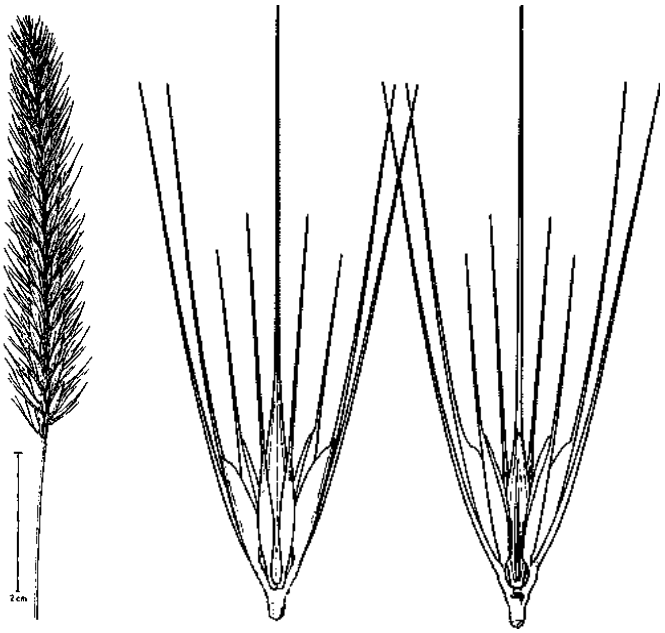


Figure 1.1 *Hordeum* spike (left; *H. brevisubulatum*) and spikelets (right; *H. marinum*). The key character for *Hordeum* determination is the possession of three single-flowered spikelets at each rachis node (modified from Bothmer et al., 1995).

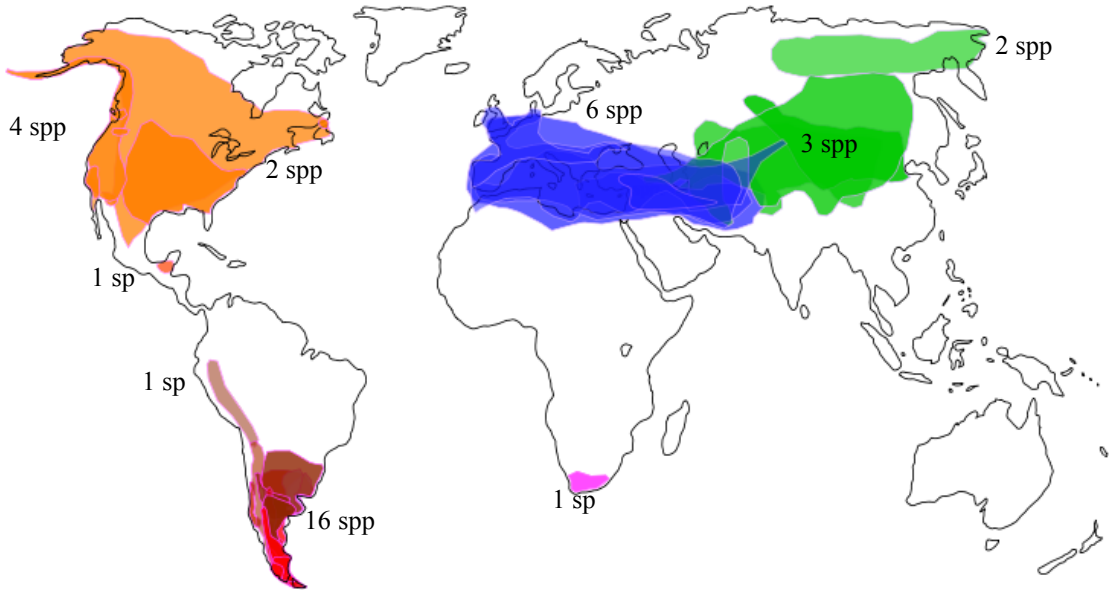


Figure 1.2 Distribution map of *Hordeum* species with number of species per area**.**

Several taxonomic studies were undertaken, including (Linnaeus, 1753) describing eight species, with five of them actually corresponding to cultivated barley. A multitude of infrageneric treatments were proposed based on the life cycle (Nevski, 1941) or the morphology of the species (Bothmer & Jacobsen, 1985; Bothmer *et al.*, 1995). In parallel, cytogenetic studies of meiotic chromosome-pairing behavior in interspecific hybrids in the tribe Hordeeae (Löve, 1984; Dewey, 1984) led to the definition of four different genomes within *Hordeum* (Bothmer *et al.*, 1986). Molecular phylogenies based on nuclear markers showed that the delimitations based on morphological characters do not represent natural groups, whereas the four different genomes are monophyletic (Komatsuda *et al.*, 1999; Blattner, 2004, 2006; Kakeda *et al.*, 2009). Blattner (2009) proposed a new infrageneric treatment consisting of two subgenera and five sections. For genome denomination I follow this last treatment, with the **H** genome occurring in *H. vulgare* and *H. bulbosum* (subgenus *Hordeum*, section *Hordeum*), **Xu** in *H. murinum* (subgenus *Hordeum*, section *Trichostachys*), **Xa** in *H. marinum* and *H. gussoneanum* (subgenus *Hordeastrum*, section *Marina*), and **I** in the remaining species (subgenus *Hordeastrum*, section *Stenostachys*). Section *Stenostachys* is further divided in series *Sibirica* and *Critesion* to account for the natural distribution and genomic differences, the latter occurring in the Americas. Arriving at a completely consistent and logical taxonomy of the genus is hampered by the high proportion of polyploid taxa partly combining genomes from different sections, such as section *Nodosa* consisting of three species with **I**/**Xa** genome combinations (*H. brachyantherum* (6x), *H. capense* and *H. secalinum*), or series (all the American polyploids except *H. depressum*).

Geographical dispersion and isolation, and polyploidization are the two major speciation mechanisms that led to the diversity observed in *Hordeum*. The genus originated in western Eurasia in the middle Miocene (Blattner, 2009) about 15.3 million years ago (Ma) when the barley and wheat lineages diverged (Marcussen *et al.*, 2014). Multiple intercontinental dispersals shaped the distribution area of *Hordeum* species starting out in Western Eurasia and reaching Eastern Asia and then North and South America through the Bering land bridge. This route of exchange between Old and New Worlds was available and free of ice through most of the Tertiary and Quaternary (Tiffney, 1985, 2008; Hewitt, 2000). During the last two million years, 10 diploid species evolved through a fast radiation in South America. Finally one species, *Hordeum capense*, colonized South Africa after the split from its sister species *H. secalinum* approximately 0.4 Ma (Blattner, 2004, 2006). Polyploidization played an important role as a speciation mechanism, as nearly half of the 45 taxa are polyploids, with 13 tetraploid ($2n = 4x = 28$) and seven hexaploid ($2n = 6x = 42$) taxa. With mostly allopolyploid species combining genomes from two or three different parental species but also autopolyploids (Bothmer *et al.*, 1995), *Hordeum* is a model of choice to study hybridization and polyploidization in grasses.

POLYPLOIDY

Polyploidy, a term introduced by Winkler (1917), is "the presence of three or more chromosome sets in an organism" (Grant, 1981). Two different concepts exist for the definition of the type of polyploidy. On the one hand is the classic cytogenetic definition where the presence of only bivalent-forming meiotic chromosomes characterizes allopolyploids while multivalent formation of homoeologous chromosomes indicates autopolyploids (Stebbins, 1947, 1950). The second definition is based on a taxonomic concept (Lewis, 1980), where polyploids formed through interspecific hybridization followed by genome duplication (allopolyploids) contrast with hybrids formed through crossing of different genotypes from within a species and/or genome duplication (autopolyploids). However, the variety of cytological behaviors from organisms with known origin led to the description of intermediary status emphasizing the fact that allopolyploids and autopolyploids are only the endpoints of a continuum (Stebbins, 1950; Grant, 1981). For example segmental allopolyploids, as taxonomic allopolyploids are often termed in the cytogenetic reference frame, are polyploids of hybrid origin possessing chromosomes forming bivalents (Ramsey & Schemske, 1998). I here use the taxonomic system of polyploid definition and explicitly refer to cytogenetic allopolyploids by indicating their genome composition.

Far from being the evolutionary dead-end or noise described by Wagner (1970), polyploidization, or whole genome duplication, is a major mechanism in plant evolution. During the past 80 years, biologists have investigated many aspects concerning polyploid organisms (reviewed in Soltis *et al.*, 2004b) such as their frequency, their mode of formation and systematic distribution. Numerous studies have tried to evaluate the proportion of polyploidy in plants. With an estimated frequency in angiosperms varying widely between 30 and 80% (Masterson, 1994), polyploidy is ubiquitous. Wood *et al.* (2009) estimated that polyploidy might be involved in up to 15% of speciation events in angiosperms, thus polyploidization "may be the single most common mechanism of sympatric speciation in plants" (Otto & Whitton, 2000). Because of its mode of formation involving hybridization and genome duplication, and therefore providing immediate reproductive isolation from their progenitors, polyploidy is considered as an instantaneous speciation mechanism (Leitch & Leitch, 2008). Polyploid genomes show high degrees of plasticity in particular due to their recurrent origins and are actually evolutionary successful, especially under environmental stress and/or fluctuations (Vanneste *et al.*, 2014). This evolutionary success might be responsible for angiosperm diversification (Soltis *et al.*, 2009) and survival during the Cretaceous-Paleogene mass extension around 66 Ma (Fawcett *et al.*, 2009; Vanneste *et al.*, 2014). Thus, probably all angiosperm lineages experienced one or several rounds of polyploidization in their history (Jaillon *et al.*, 2007; Velasco *et al.*, 2007; Tang *et al.*, 2010). Many studies (reviewed in Soltis *et al.*, 1993; Soltis & Soltis, 1999) revealed that most polyploid species originated repeatedly involving different parental genotypes, thus creating genetically diverse populations that can interbreed afterwards (Fig. 1.3). Polyploidy might

also be an important component in favoring the adaptation and range expansion of invasive species (Beest *et al.*, 2011) such as polyploid species of *Fallopia* (Bailey *et al.*, 2007), *Spartina* (Ainouche *et al.*, 2009) or *Hordeum* (Bothmer *et al.*, 1995). Theoretical studies showed that under certain conditions polyploid organisms could not only successfully establish and persist but also outcompete one or both of their progenitors (Fowler & Levin, 1984; Rodriguez, 1996). This hypothesis was confirmed, among others, by molecular data in *Cardamine* (Lihovà *et al.*, 2006) and *Hordeum murinum* (Jakob & Blattner, 2010).

A few polyploids have been extensively studied including recently formed wild species such as *Tragopogon* (Soltis *et al.*, 2004a) and *Spartina* (Ainouche *et al.*, 2004), which formed within the last 150 years, and hexaploid bread wheat that originated with the onset of agriculture (Petersen *et al.*, 2006; Zohary *et al.*, 2012; Marcussen *et al.*, 2014). Studies aiming at resolving the phylogeny of an entire genus such as *Hordeum* with about half of its species being polyploids are still rare (but see for example Triplett *et al.*, 2012; Harpke *et al.*, 2013; Mason-Gamer, 2013).

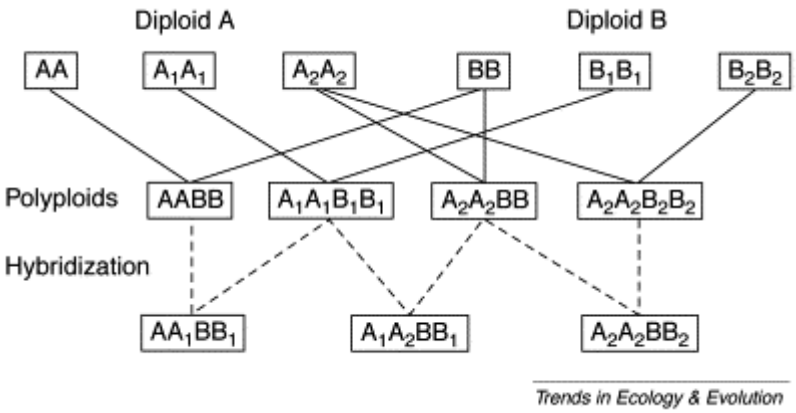


Figure 1.3 The multiple origin of polyploids hypothesis (taken from Soltis & Soltis, 1999). The genetic diversity in the parental genotypes, recurrent formation, and potential subsequent hybridization result in genetically diverse polyploid species.

Previous Work: Hordeum *and the Development of Phylogenetic Markers*

A known phylogeny is a prerequisite to many evolutionary and ecological studies such as understanding speciation mechanisms, or testing for niche shifts or conservatism after speciation (Jakob *et al.*, 2007, 2009, 2010). Especially in the case of polyploid taxa, knowing the relationships with their progenitors is necessary to understand and compare their evolution (Jakob *et al.*, 2007). Combining the genomes of at least two different species, allopolyploids are more than simple additive genomes (Doyle *et al.*, 2008). Although the genome donors of polyploids should be identifiable in karyological studies or on a phylogenetic tree including all potential progenitors, studies on polyploid taxa are generally impeded by the complex evolution of these organisms. The dynamic nature of polyploids involving recurrent formation (Brochmann *et al.*, 1992; Soltis & Soltis, 1999; Soltis *et al.*, 2004a; Cai *et al.*, 2012), gene loss or retention (Blattner, 2004; Kotseruba *et al.*, 2010; Buggs *et al.*, 2012), and homoeologous recombination (Doyle *et al.*, 2008; Salmon *et al.*, 2010; Flagel *et al.*, 2012; Weiss-Schneeweiss *et al.*, 2013) tend to blur the phylogenetic signal. The numerous analyses carried out to study relationships in *Hordeum* illustrate the development of new methods and markers to provide an accurate phylogeny of the genus. Their drawbacks and advantages will be described thereafter.

Before the advent of molecular techniques, relationships within *Hordeum* were estimated using morphological characters (*e.g.*, Booth & Richards, 1976) and cytological analyses such as chromosome pairing affinities in interspecific hybrids and karyological analyses (*e.g.*, Bothmer *et al.*, 1986, 1988) but the relationships among closely related species can be difficult to define. Although more recently, new cytological methods based on genomic and/or fluorescence *in situ* hybridization (GISH and FISH) can illustrate the genome composition of polyploid taxa in regard to their potential diploid progenitors (Taketa *et al.*, 1999, 2009; Carmona *et al.*, 2013) providing new insights into the evolution of polyploid complexes (Cuadrado *et al.*, 2013).

Doebley *et al.* (1992) used chloroplast DNA variation, one of the first molecular marker to be widely used (Savolainen & Chase, 2003), on nearly all taxa and cytotypes of the genus to provide a first picture of the relationships within *Hordeum*. But chloroplast DNA is usually maternally inherited in angiosperms (Reboud & Zeyl, 1994) and therefore not able to detect hybridization events. Moreover, chloroplast loci were learned afterwards to result in phylogenetic relationships in *Hordeum* incongruent with nuclear data due to processes such as incomplete lineage sorting, persisting polymorphisms and far-reaching haplotype extinctions (Jakob & Blattner, 2006; Petersen *et al.*, 2011). They can, however, be used to identify the direction of hybrid speciation in polyploids, *i.e.* to determine maternal parents (Doebley *et al.*, 1992; Nishikawa *et al.*, 2002).

Amplification fragment length polymorphisms (AFLP; Vos *et al.*, 1995) were also developed to study interspecific relationships in *Hordeum* (El-Rabey *et al.*, 2002; Pleines & Blattner, 2008).

Although usually used for population genetic and intraspecific studies, those anonymous dominant genome-wide markers can provide useful information among closely related species. For example, AFLP resulted in a mostly well-resolved tree for the relatively young (2 Ma; Blattner, 2006) South American clade of *Hordeum* for which nuclear and chloroplast markers arrived only at a low phylogenetic resolution (Pleines & Blattner, 2008). But none of those AFLP studies included all species of the genus as the uncertain homology between comigrating bands across species increases with the taxonomical scope (El-Rabey *et al.*, 2002). In *Hordeum*, relationships within the *H. murinum* polyploid complex were investigated using AFLP alone (El-Rabey *et al.*, 2002) or together with specific nuclear markers (Jakob & Blattner, 2010), as polyploid taxa require special care when using AFLP (Albach, 2007).

Blattner (2004) conducted a thorough phylogenetic analysis of all taxa of the genus, including mostly multiple individuals per taxon using the highly repetitive internal transcribed spacer (ITS) of the nuclear ribosomal gene array (nrDNA). ITS is a fast evolving marker bordered by highly conserved regions making it possible to develop universal primers (White *et al.*, 1990; Blattner, 1999), and, as such, a very efficient marker to study phylogenetic relationships at the infrageneric level and also within families (Baldwin *et al.*, 1995; Hsiao *et al.*, 1995, 1999). The highly repetitive nature of ITS makes it easy to amplify but is also responsible for its peculiar mode of evolution via gene conversion (Álvarez & Wendel, 2003). When no concerted evolution occurs, the presence of multiple types of ITS in a polyploid taxon traces back to its progenitors (Sang *et al.*, 1995; Blattner, 2004). However homogenization of tandem repeats (Wendel *et al.*, 1995; Blattner, 2004) or the loss of entire rDNA clusters (Kotseruba *et al.*, 2010) might result in effectively uniparental inheritance of the ITS region.

As an alternative approach low- or single-copy nuclear loci have been proposed as a source of phylogenetic information and for improving resolution and robustness in comparison to plastid and ribosomal DNA (Sang, 2002; Small *et al.*, 2004), particularly if polyploid taxa are studied (Sang *et al.*, 2004). Low-copy nuclear markers also have some disadvantages, as there are no universal PCR primers available that are applicable in all plant groups, and additional costs and lab work arise due to the necessity of cloning PCR amplicons prior to sequencing (Triplett *et al.*, 2012). Furthermore PCR might favor the formation of chimerical sequences (Cronn *et al.*, 2002; Zhou *et al.*, 2003) combining the sequences issued from different alleles. Chimerical sequences are formed during PCR when at least two almost identical sequences occur in a single PCR amplification reaction, as it is usually the case with polyploids (Cronn *et al.*, 2002). Incompletely extended amplicons can then hybridize to their homoeologous template, be extended in the next cycle and propagated during subsequent cycles (Meyerhans *et al.*, 1990; Bradley & Hillis, 1997). *In vitro* cloning and sequencing will then reveal each artifact, which would probably be masked by direct Sanger sequencing. Removing recombinant sequences from datasets is necessary, as they contribute a mixed signal within phylogenetic analyses.

They can be identified either by carefully inspecting the sequences and their placement on a phylogenetic tree (Triplett *et al.*, 2012) or by using specific programs such as RDP3 (Martin *et al.*, 2010). Single-molecule PCR (Marcussen *et al.*, 2012), using homoeologue-specific primers (Petersen & Seberg, 2004; Lihovà *et al.*, 2006) or single-strand conformation polymorphism (Rodríguez *et al.*, 2011; Cai *et al.*, 2012) are alternatives to cloning when working with polyploids but require also intensive lab work and/or *a priori* knowledge of the allele diversity at the analyzed loci. With the advent of next-generation sequencing methods, it became clear that the properties of *in silico* cloning and high-throughput sequencing would be an alternative of choice reducing artificial chimerical sequences formation and allowing naïve exploration of the allelic diversity at the analyzed loci (Griffin *et al.*, 2011; O'Neill *et al.*, 2013).

For the diploid taxa of *Hordeum*, datasets that include multiple nuclear loci converged during the last years onto similar results (Blattner, 2009; Petersen *et al.*, 2011). Thus, a good phylogenetic hypothesis seems to be accomplished for this group, although studies including several individuals per species are yet rare. For polyploid species and cytotypes of *Hordeum* phylogenetic relationships were studied mainly for small taxon groups (Petersen & Seberg, 2004; Taketa *et al.*, 2005, 2009; Kakeda *et al.*, 2009; Komatsuda *et al.*, 2009; Jakob & Blattner, 2010; Tanno *et al.*, 2010; Wang & Sun, 2011) using nuclear low-copy number loci. Therefore, a phylogenetic hypothesis for all *Hordeum* species, including all cytotypes, based on several individuals per species and single- or low-copy nuclear loci is still lacking, which severely restricts evolutionary studies in polyploidy *Hordeum* taxa.

OBJECTIVES OF THIS STUDY

The aim of this dissertation is to resolve and clarify all the relationships within the genus *Hordeum* with a particular focus on the progenitor-derivative kinships of polyploid taxa as a prolongation of the synthesis (Fig. 1.4) from Blattner (2009).

First a single low-copy number locus (*TOPO6*) was used on a large sample including mainly several individuals per species. This first analysis aimed at evaluating the nucleotide diversity of *TOPO6*, a nuclear low-copy number locus already successfully tested to resolve relationships within the *H. murinum* complex (Jakob & Blattner, 2010), for the whole genus and its capacity to identify the different species. But because a single locus might not be enough to clarify all relationships, a second analysis based on 12 nuclear low-copy number loci and one chloroplast region on a smaller sample including only up to four individuals per species was then carried out.

Overall this dissertation aims to (*i*) define parental species of allopolyploids, (*ii*) analyze the status of putative autopolyploids, (*iii*) infer single or multiple origins for polyploids, (*iv*) infer speciation events on the polyploid level, (*v*) check for indications of introgression from outside of *Hordeum*, (*vi*) compare the influence of two different types of DNA polymerases in PCR on the results of polyploid analyses, and (*vii*) compare different methods of multilocus phylogenetic inference. Ultimately a species tree will be inferred from the different gene trees. On a larger scale, this work will establish new nuclear marker regions useful in Hordeeae and eventually within the whole Poaceae family.

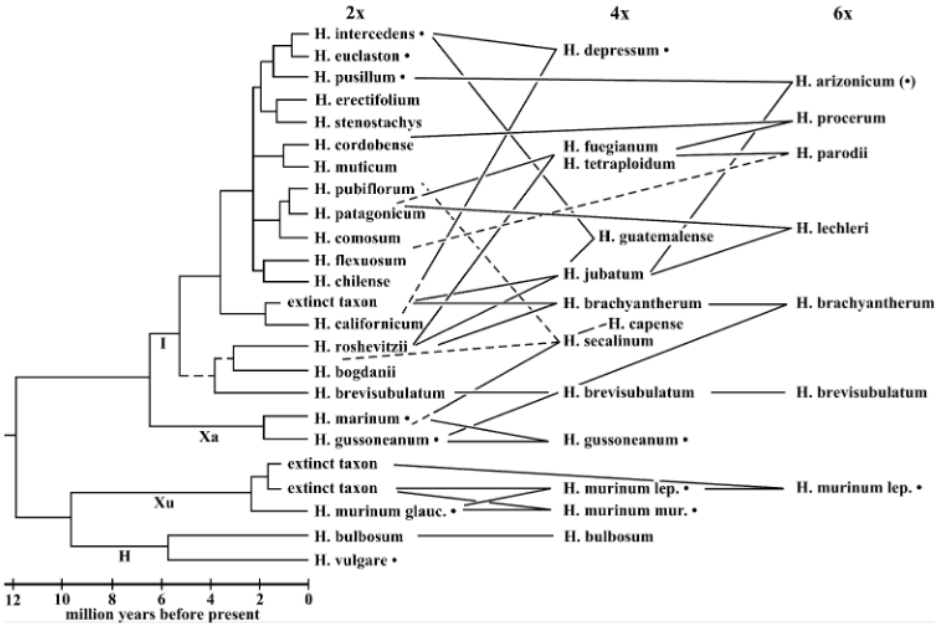


Figure 1.4 Scheme summarizing the phylogenetic relationships within *Hordeum* (taken from Blattner, 2009). Diploid taxa were drawn directly to the tree, while tetra- and hexaploids were connected by lines to their putative progenitors. Dots behind species names depict annual taxa, dashed lines topological uncertainties. Age estimations were calculated with a penalized likelihood approach (Blattner, 2006).

METHODS

This section will give a short introduction to the specific methods employed in the rest of this dissertation including sequencing, phylogenetic inference and molecular dating. In a first part, molecular methods will be described. In a second part, all the types of phylogenetic inferences that were used throughout this work will be shortly presented. Finally, the coalescent-based model, a more complex type of Bayesian Phylogenetic Inference, applied to create a dated multispecies phylogeny will be introduced. Advantages and pitfalls of all the different methods will be discussed.

IN VITRO VS IN SILICO *CLONING*

Because directly sequencing of a mixture of PCR amplicons, constituted of the different alleles within polyploid individuals, creates undecipherable chromatogram traces (but see Flot *et al.*, 2006), sequencing of polyploid taxa requires special care. One way to avoid this problem consists in selecting and isolating specific amplicons, *i.e.* cloning, prior to sequencing.

Traditional *in vitro* cloning uses the capacity of transforming bacteria, commonly *Escherichia coli* strains, with plasmids carrying the locus of interest (Sambrook *et al.*, 1989). In genetic engineering, plasmids are called vectors corresponding to small and artificial circular DNA molecules that are independent from chromosomal DNA. PCR amplicons are inserted into vectors containing genes that make cells resistant to an antibiotic. Bacteria are then transformed with the vectors and exposed to the particular antibiotic. Only bacteria effectively transformed with the vector survive. Antibiotics serve as a screen to select only the transformed bacteria, plus a reporter gene indicates if the vector present in the bacteria is recombinant, *i.e.* contains the PCR amplicon insert, or not. DNA fragments are inserted in the sequence of a reporter gene, thus disrupting its activity. Reporter genes can express an enzymatic cascade coloring the colonies, like the *lacZ-α* used in the common blue/white selection (Ullmann *et al.*, 1967), or a lethal protein, such as *eco47IR* present in the cloning vector pJET (Stankevičius *et al.*, 1995). A particular DNA fragment can then be isolated and multiplied within *E. coli* cells, before Sanger sequencing.

In vitro cloning is an easy and well established method used by many laboratories (*e.g.*, Blattner, 2004; Hoot *et al.*, 2004; Triplett *et al.*, 2012). The method is however relatively expensive and possesses other pitfalls. For example, knowing how many clones must be sequenced to discover all alleles of a certain locus is a difficult question depending on the ploidy level of the individual and the relative frequency of the different copies (Rauscher *et al.*, 2002). Between eight (for diploids) and 24 (for hexaploids) clones are usually sequenced (*cf.* Chapter 2). Moreover, *in vitro* cloning has the

tendency to reveal artifacts, such as DNA polymerase error and PCR recombinants (mentioned earlier), so that it can be hard to infer the "true" sequence for a particular locus.

Ever-larger experiments and ever-increasing needs for sequencing data led to the development of high-throughput sequencing platforms. The 454-sequencing platform (Roche) (Margulies *et al.*, 2005), commercialized since 2005, initiated the development of next-generation sequencing (NGS) platforms. Miniaturized and massively parallelized sequencing reactions are the common characteristics of NGS platforms. The 454-sequencing platform together with the HiSeq platform (Illumina) (Bentley *et al.*, 2008), another NGS automaton, revolutionized sequencing by dramatically decreasing sequencing costs (Lemmon *et al.*, 2012). This allowed very large projects to be achieved (The International Barley Genome Sequencing Consortium, 2012; The International Wheat Genome Sequencing Consortium, 2014). The high throughput of the sequencing platforms provides an oversampling ensuring the high quality of the base call (Margulies *et al.*, 2005) and the possibility to tag DNA fragments with known nucleotide patterns (Meyer *et al.*, 2007, 2008b) before pooling different individuals or even experiments, dramatically reducing the sequencing cost.

By using very finely tuned protocols, NGS platforms can isolate, amplify and sequence single DNA molecules without the help of bacterial cloning, hence cloning *in silico* (Rothberg & Leamon, 2008). With the 454-sequencing platform, single DNA molecules are ligated to 28 μm beads in 1:1 dilution conditions. Clonal amplification of the individual fragment is performed in droplets of an emulsion PCR and each bead is then deposited individually in one of the 1.6 million picoliter-sized reaction wells (Fig. 1.5; Margulies *et al.*, 2005). Finally, sequencing by synthesis is performed and bases are called in a highly parallelized fashion producing routinely one million sequences, called reads, on average 400 bases long (Margulies *et al.*, 2005) and up to 1000 bp with the newest chemistry (GS FLX Titanium XL+, see http://454.com).

The downstream analysis process consists in separating the reads by barcodes (barcode deconvolution), followed by "*de novo*" assembling or mapping the reads to a known sequence, and finally quality-check. Individual reads are aligned to one another in a so-called "contig" by using specific algorithms (Miller *et al.*, 2010) and a consensus is created, summarizing hundreds or thousands of reads to a single sequence. According to the sensitivity of the algorithm the reads produced by different alleles can either be assembled together or in separate contigs (*cf.* Chapter 3). The relatively long reads, compared to Illumina, produced by 454 simplify the assembly procedure (Li & Homer, 2010; Chin *et al.*, 2013). Besides the huge load of data inherent to next-generation sequencing platforms (Baker, 2010) and the necessity to use specific bioinformatic tools, 454 sequencing has also some specific disadvantages. For example, handling of homopolymers of length four or more is delicate and should be carefully inspected (Margulies *et al.*, 2005; Huse *et al.*, 2007). With the development of a third-generation sequencing platform, single molecule real time sequencing (Eid *et al.*, 2009; Chin *et al.*, 2013), very long reads (up to 30 kb) are available. Although still with a

reduced throughput (up to 250 Mb), the third-generation provides phased haplotype genomes allowing researchers to explore new questions (Adey *et al.*, 2013).

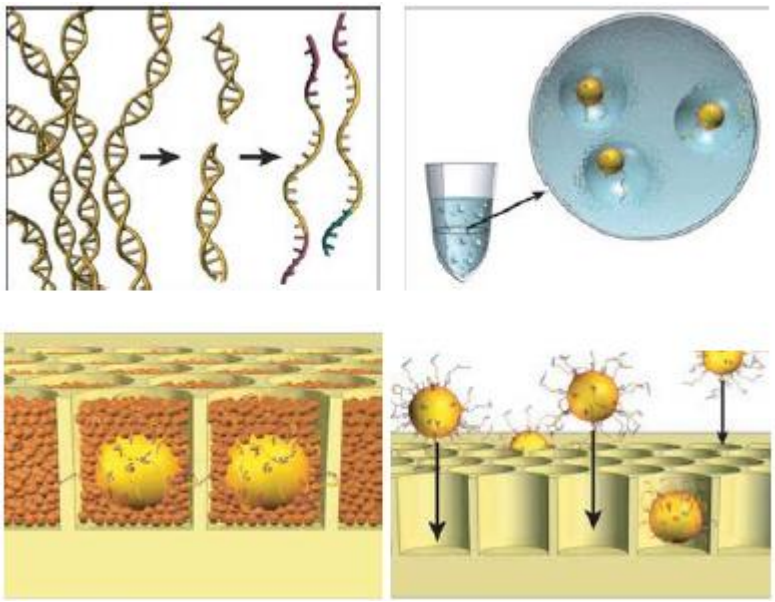


Figure 1.5 Sample preparation for 454 sequencing (taken from Margulies *et al.*, 2005). Genomic DNA is isolated, fragmented, ligated to adapters and separated into single strands (top left). Fragments are bound to beads under conditions that favor one fragment per bead, the beads are captured in the droplets of a PCR-reaction-mixture-in-oil emulsion and PCR amplification occurs within each droplet, resulting in beads each carrying ten million copies of a unique DNA template (top right). The emulsion is broken, the DNA strands are denatured, and beads carrying single-stranded DNA clones are deposited into wells of a fiber-optic slide (bottom right). Smaller beads carrying immobilized enzymes required for pyrophosphate sequencing are deposited into each well (bottom left).

Phylogenetic Inference Methods

All three types of phylogenetic inferences (*i*) distance-based method, (*ii*) maximum parsimony (MP), and (*iii*) model-based method (Maximum-Likelihood, ML and Bayesian Inference, BI) were used throughout this work. Finally, the coalescent-based model, a more complex type of BI, applied to create a dated multispecies phylogeny will be introduced. All methods have advantages and inconveniences but a pluralistic approach, that is to use different methods and compare the resulting topologies, is advocated (Doyle & Gaut, 2000; Nei & Kumar, 2000).

Distance methods

Distance methods such as neighbor joining (NJ; Saitou & Nei, 1987) estimates evolutionary distances by calculating a matrix summarizing dissimilarities between each pair of taxa. NJ is an iterative bottom-up algorithm clustering at each step the closest taxa. A single tree is finally created minimizing the sum of branch lengths at each iteration. Models of nucleotide substitution (described in section Model-based Methods) were developed to correct the distance between taxa. In contrast to other distance methods, such as UPGMA (Sokal & Michener, 1958), NJ does not assume a constant rate of evolution (*i.e.* a molecular evolutionary clock) between lineages, a hypothesis not always valid (Kumar, 2005).

NJ is an extremely fast method especially useful to have a first evaluation of the relationships between taxa. But the original information is lost during the creation of the distance matrix, making it impossible to reconstruct polymorphisms supporting the topology. Furthermore, NJ produces a single tree preventing comparisons with other topologies (Soltis & Soltis, 2003). Finally, NJ can produce different topologies according to the entry order of the sequences (Farris *et al.*, 1996). Thus a measure of the node support such as bootstrap (Felsenstein, 1985) is necessary.

Maximum Parsimony

Maximum parsimony is a method based on the philosophical idea that the simplest explanation is the best (Ockham's razor). Introduced to phylogenetic inference by Fitch (1971), this idea consists in selecting the tree supposing the least number of character changes to explain the data. In contrast to NJ, all possible topologies have to be explored to find the most parsimonious one. But depending on the data set, between a few and a hundreds equally short trees can be found. A strict consensus is then calculated summarizing only the nodes present in all shortest trees.

Because MP explicitly identifies the shared characters, called synapomorphies, supporting a clade, this method can infer ancestral traits. But when considering DNA sequences, each site can have only four different states and reverse mutations are likely to happen. Thus when mutation rates differ largely between taxa, the probability increases that they will converge, and fast evolving lineages will "wrongly" cluster together. MP is especially prone to this phenomenon called long-branch attraction (Felsenstein, 1978). However, this artifact seems to be primarily a sampling issue, when only distantly related taxa are included without intermediate ones (Bergsten, 2005).

Model-based methods

Two kinds of model-based methods to infer phylogenies were developed, maximum likelihood (ML) and Bayesian inference (BI). They both rely on the models of nucleotide (or amino acid or morphological character) substitution to reconstruct both topology and branch lengths. For DNA sequences, those models describe the frequency of each nucleotide and the probabilities of mutations changing character states between the different nucleotides. The simplest model, JC (Jukes & Cantor, 1969), assumes equal mutations and equal proportion of nucleotides. The most complex one, generalized time-reversible GTR (Tavaré, 1986), provides a different parameter for all proportions in bases and all mutations. Several hundred other intermediate models are also available (Darriba *et al.*, 2012). In addition to the models of substitution, parameters describing the distribution of variable sites in the alignment can be implemented. Two such parameters were developed, (*i*) the proportion of invariable sites (Hasegawa *et al.*, 1985) describes a fraction of the data as invariant and the rest vary at the same rate and (*ii*) the rate variation among sites is modeled by a gamma distribution (Yang, 1993, 1994). But as Box (1976) put it "all models are wrong but some are useful" and it is necessary to identify the best fit model. The likelihood of a model is susceptible to increase by adding parameters. However, too many parameters in regard to the data available can lead to calculation problems as less data are available to estimate each parameter (Hasegawa *et al.*, 1985; Posada & Crandall, 1998). For model selection, the Akaike information criterion (AIC) (Akaike, 1973, 1974) and the Bayesian information criterion (BIC) (Schwarz, 1978) are the two main methods available to identify this trade-off. Both AIC and BIC penalize the likelihood of the model by the number of its parameters. With BIC, taking into account the sample size, the penalty is higher than in AIC (Posada & Buckley, 2004), usually proposing a simpler model (Posada & Crandall, 2001).

If both ML and BI are similar they however differ in the way they handle the problem. Likelihood asks "what is the probability of the data given the model?" (Wiley & Lieberman, 2011). ML will search for the phylogeny (topology and branch length) that maximizes the likelihood of observing the data according to a specific model and topology. That is maximizing the probability that a given model and a history (tree) produced the data we are considering (Swofford *et al.*, 1996). Likelihood is calculated for a given tree and associated branch lengths, so that all potential transformations at each character need to be calculated (Wiley & Lieberman, 2011). The major disadvantage of probabilities is that they are not explanatory, *i.e.* all possible explanations have a non-zero probability but only one is correct (Siddall & Kluge, 1997). As a consequence, the search of the "tree space" can be computationally very demanding. However, recent algorithms, such as the ones implemented in RAxML (Stamatakis, 2014) which can handle very large data sets, made the search more efficient. Finally, the other inconvenience of ML is that it provides only one optimal tree.

Bayesian Inference investigates the alternative question, *i.e.* "what is the probability that the model is correct given the data?" (Wiley & Lieberman, 2011). It computes a value called the posterior

probability of trees, corresponding to the distribution of correct trees according to the data and the model. As with ML, the likelihood of the data given the model is calculated but is only a parameter of the posterior probability along with prior information explaining the data. Bayes (1763) formulated the theorem but BI inference could be applied in phylogeny only since 1996 (Li, 1996; Mau, 1996; Rannala & Yang, 1996) due to the difficulty to calculate the posterior probability of a tree as it "involves a summation over all possible trees and, for each tree, integration over all combinations of branch lengths and substitution-model parameter values" (Huelsenbeck *et al.*, 2002). Except for very simple cases, it is actually impossible to calculate it analytically. Monte Carlo sampling using Markov chains, MCMC (Metropolis *et al.*, 1953; Hastings, 1970), allows the approximation of the posterior probability distribution of trees. MCMC can be pictured as a robot (*i.e.* random walk) across a hilly landscape (*i.e.* the tree space) who wants to visit the highest peak(s) (Lewis, 2001). At each step (*i.e.* generation), the robot chooses randomly a direction and a distance (*i.e.* a set of parameters drawn from the parameters' prior distributions). The robot can always go uphill (*i.e.* better likelihood), it can go downhill but with some restrictions (*i.e.* the ratio of previous and new state compared to a randomly drawn value). The "path" chosen by the chain depends on its starting point so usually several (two to four) chains and/or analyses starting from different random points are run in parallel. The chains need to be run long enough to cover the entire space; generally a few millions of generations are necessary to arrive at a good evaluation of the landscape (*i.e.* convergence of the chains). After discarding the first values (*i.e.* burn-in), the probability distribution is approximated by periodically sampling values of the chains' stationary phase. As other methods, MCMC can be trapped in a local optimum not exploring equally high or higher peaks separated by deep valleys. The parallel Metropolis coupled MCMC algorithm (MCMCMC or MC^3; Altekar *et al.*, 2004) implemented for example in MrBayes (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003; Ronquist *et al.*, 2012b) uses chains that can explore remote tree space more easily (*i.e.* "heated" chains with lower restrictions) and can exchange their information with the main "cold" chain.

Advantages of BI include the possibility to accommodate for uncertainties, the direct estimation of support indicated by the values of posterior probability (pp) at each node and the possibility to use prior information (scientist's expertise) if available. The potential subjectivity of priors is also the main criticism of BI (but see Huelsenbeck *et al.*, 2002), as incorrectly specifying the model can lead to a wrong result, however this might be the case for every method used (Susko *et al.*, 2004). Although MC^3 is a relatively efficient way to explore the tree space (Huelsenbeck *et al.*, 2002), analysis can take up to several days sampling over tens of millions of generations to achieve convergence.

Gene trees vs. species trees

The copies of a single locus sampled across various species will always produce only a gene tree. However, the ultimate goal in phylogeny is to provide a species tree illustrating the process of speciation and the relationships between species (Nichols, 2001). It is often assumed that a gene tree is a good indicator of species relationships, especially on a large phylogenetic scale (Maddison, 1997; Degnan & Rosenberg, 2006; Heled & Drummond, 2010). However, incongruences between gene trees, and eventually with the species tree, can arise from various processes (Maddison, 1997) and can only be recognized when sequencing multiple loci as well as multiple individuals per species (Pamilo & Nei, 1988; Knowles, 2009).

Genes can be horizontally transferred between species either by vectors (viruses for example) or by hybridization. Allopolyploid organisms are by nature the result of horizontal gene transfer but, unless gene homogenization (Álvarez & Wendel, 2003) and/or diploidization (Wolfe, 2001) occurred, the incongruence is resolved by modifying the dichotomic tree into a network (Blattner, 2004). Gene duplication, by creating multiple copies that will evolve independently between lineages and eventually become differentially extinct, can subvert phylogenies due to the potential comparison of orthologous and paralogous loci (Fitch, 1970). Finally also the persistence of ancestral polymorphisms through speciation events will result in incongruences between gene and species trees (Edwards, 2009). Incomplete lineage sorting (ILS) or deep coalescence occurs when population sizes are large, and low genetic drift occurs and differential selective pressure is low in regard to speciation times (Pamilo & Nei, 1988).

Methods to accommodate incongruences between gene trees and reconstruct the species tree fall into two general classes: the supermatrix approach that consists in concatenating the different loci (de Queiroz & Gatesy, 2007) versus model-based methods. The first one is relatively straightforward and can use traditional methods of phylogenetic inference with the hope that genome-scale data would converge to a correct solution (Rokas *et al.*, 2003) while the latter explicitly integrates the processes leading to incongruent data and/or models relationships between gene trees and species trees (reviewed in Knowles, 2009). Although it is possible to use different models of sequence evolution per locus within a partitioned supermatrix, the concatenation approach assumes that all the data evolved according to a single evolutionary tree (Degnan & Rosenberg, 2009). However, loci can have different evolutionary histories for example due to genome recombinations. Therefore, the first assumption will be violated, potentially leading to overconfident support for incorrect species trees (Kubatko & Degnan, 2007; Edwards, 2009; Heled & Drummond, 2010). New methods accounting for the stochastic history of genes were developed among them the Bayesian concordance analysis (BCA; Ané *et al.*, 2007) and the multispecies coalescent (Degnan & Rosenberg, 2009; Heled & Drummond, 2010). Computed within BUCKY (Larget *et al.*, 2010), the first one provides an estimation of the proportion of the sampled loci or of the genome supporting a given topology by computing the clade's

concordance factor (CF) without making any assumption about the reason for loci discordance (Ané *et al.*, 2007; Baum, 2007; Larget *et al.*, 2010). While the second one, an extension of the coalescent theory, is assuming ILS as the only source of discrepancy between loci (Heled & Drummond, 2010).

Coalescent theory is a part of theoretical population genetics that aims at reconstructing, in a time-backward manner, the most recent common ancestor of orthologous gene copies sampled in a population (Kingman, 1982; Hudson, 1990; Hein *et al.*, 2004). The coalescent infers ancestral parameters such as population size history and divergence time directly from the sequence polymorphisms and the conflicting gene histories (Wakeley & Hey, 1997; Wall, 2003; Li & Durbin, 2011). Originally designed to model gene genealogies within populations evolving essentially under genetic drift, the theory was then extended to multispecies serving as a prior when building a species tree in a Bayesian framework (Heled & Drummond, 2010). Different species are actually considered as different related populations and for which the time for a common ancestor of a gene cannot be more recent than the splitting times of the respective species (Degnan & Rosenberg, 2009). Gene trees are constrained to a species tree and follow the stochastic coalescent process back in time from the present within each branch. Multiple individuals per species are necessary as intraspecific polymorphisms provide information on population sizes and the more samples are available, the better the estimations will be (Heled & Drummond, 2010). In *BEAST, the method used in Chapter 2, all gene trees and the species tree are simultaneously inferred to take advantage of the independent loci in estimating population sizes. Species divergence times are estimated by using interspecific polymorphisms in combination with a molecular clock (Rannala & Yang, 2003) and calibration points to obtain absolute times (Ronquist *et al.*, 2012a).

The idea of a molecular clock, *i.e.* the rate(s) at which sequences evolve, dates back for 50 years (for a review see Kumar, 2005). Thought to be a constant rate across lineages and over time, the strict molecular clock was proposed at first by Zuckerkandl & Pauling (1965). With the accumulation of molecular data it became clear that sequences evolve under many parameters violating the strict molecular clock (Gaut *et al.*, 1992). Life history traits of populations, such as size, growth and generation time (Ohta, 2002), and biochemical parameters, for example recombination and repair mechanisms (Huttley *et al.*, 2000), might be different between species or even between loci. The likelihood ratio test (LRT), a general statistic method, can be used to test for rate constancy across lineages by comparing the likelihood of a tree assuming a strict molecular clock versus no clock (Huelsenbeck & Rannala, 1997). A number of authors proposed to "relax" the molecular clock in order to accommodate for non-clock-like relationships between lineages and for calibration points not fitting with a strict molecular clock (*e.g.*, Thorne *et al.*, 1998; Drummond *et al.*, 2006; Drummond & Suchard, 2010). Assuming that different clades can have different rates but with a constant rate within each clade, local clocks (Yoder & Yang, 2000) are an option allowing all possible clock configurations, including both strict molecular clock and unconstrained models. The random local

clock (RLC) is an adaptation of the local clock method to a Bayesian framework from Yoder & Yang (2000) to sample over the large space of all 2^{2n-2} potential local clocks, where n is the number of sequences analyzed (Drummond & Suchard, 2010). However, taking into account the non-constancy of rates increases dramatically the complexity of the model potentially leading to problems with the analysis (*cf.* Chapter 3).

2. A Single Locus Analysis: Cloning

Progenitor-derivative relationships of *Hordeum* polyploids (Poaceae, Triticeae) inferred from sequences of *TOPO6*, a nuclear low-copy gene region.

The content of this chapter was published by Brassac, Jakob and Blattner in 2012 in *PLoS ONE* **7**: e33808.

Introduction

In this chapter, results from a phylogenetic analysis of all *Hordeum* species using cloned sequences of the nuclear low-copy region *TOPO6* (Jakob & Blattner, 2010) are reported. *TOPO6* is a partial sequence of the SPO11 gene, a conserved plant homologue of the important archaean topoisomerase VI subunit B involved in inducing meiotic DNA double-strand breaks during recombination (Bergerat *et al.*, 1997; Hartung & Puchta, 2001). It consists of many variable introns (Hartung *et al.*, 2002) with exons conserved enough to design PCR primers. To determine genetic diversity of this locus within species and to be able to detect possible independent origins of polyploids we included for all taxa except one (*H. guatemalense*) an average of five individuals per taxon, representing the geographic distribution of the species. The aims of this chapter are to (*i*) define parental species of allopolyploids, (*ii*) analyze the status of putative autopolyploids, (*iii*) infer single or multiple origins for polyploids, (*iv*) infer speciation events on the polyploid level, (*v*) check for

indications of introgression from outside of *Hordeum*, and (*vi*) compare the influence of two different types of DNA polymerases in PCR on the results of polyploid analyses. We do not assume that the gene tree obtained from the analyzed locus will represent the 'true' species phylogeny of the genus. This will, however, not impair the objectives of this analysis, as the main interest is the relationships between polyploids and their diploid progenitors, which should be independent from the gross topology of the diploids in a phylogenetic tree.

MATERIALS AND METHODS

PLANT MATERIALS

We included 341 individuals representing all 33 species and most subspecies of the genus plus seven diploid Hordeeae species outside *Hordeum* as outgroups (Table 2.1). Included individuals were obtained from germplasm repositories or sampled from natural populations (Table S2.1; all figure and table numbers starting with 'S' indicate Supplementary Information) and all necessary permits were acquired to use these materials. Herbarium vouchers of the analyzed materials were deposited in the herbaria of the IPK Gatersleben (GAT) or the Museum of Natural History, Buenos Aires (BA).

Table 2.1 Taxa included in the study

Taxon	Ploidy level (N)[1]	Haploid genome	Distribution area
Hordeum* subgenus *Hordeum			
Section *Hordeum*			
H. vulgare L.			
subsp. *spontaneum* (C.Koch.) Thell.	2x (2)	**H**	SW Asia
H. bulbosum L.	2x (5), 4x (3)	**H, HH**	Mediterranean to C Asia
Section *Trichostachys* Dum.			
H. murinum L.			
subsp. *glaucum* (Steud.) Tzvel.	2x (3)	**Xu**	Mediterranean to C Asia
subsp. *murinum*	4x (3)	**XuXu**	NW Europe to Caucasus
subsp. *leporinum* (Link) Arc.	4x (4), 6x (3)	**XuXu, XuXuXu**	Mediterranean to C Asia
***Hordeum* subgenus *Hordeastrum* (Doell) Rouy**			
Section *Marina* (Nevski) Jaaska			
H. gussoneanum Parl.	2x (4), 4x (6)	**Xa, XaXa**	Mediterranean to C Asia

H. marinum Huds.	2x (3)	**Xa**	Mediterranean
Section *Stenostachys* Nevski			
Series *Sibirica* Nevski			
H. bogdanii Will.	2x (41)	**I**	C Asia
H. brevisubulatum (Trin.) Link [2]	2x (5), 4x (9), 6x (4)	**I, II, III**	C Asia
H. roshevitzii Bowden	2x (17)	**I**	C Asia
Series *Critesion* (Raf.) Blattner			
H. californicum Covas & Stebb.	2x (13)	**I**	SW California
H. chilense Roem. & Schult.	2x (22)	**I**	Chile and W Argentina
H. comosum Presl	2x (24)	**I**	S Argentina
H. cordobense Bothmer et al.	2x (22)	**I**	C Argentina
H. erectifolium Bothmer et al.	2x (1)	**I**	C Argentina
H. euclaston Steud.	2x (14)	**I**	C Argentina, Uruguay
H. flexuosum Steud.	2x (8)	**I**	E+C Argentina
H. intercendens Nevski	2x (7)	**I**	SW California, NW Mexico
H. muticum Presl	2x (10)	**I**	C to N Andes
H. patagonicum (Haum.) Covas [2]	2x (14)	**I**	S Argentina
H. pubiflorum Hook.f. [2]	2x (17)	**I**	S Argentina
H. pusillum Nutt.	2x (13)	**I**	C+E USA
H. stenostachys Godr.	2x (19)	**I**	C Argentina
H. depressum (Scribn. & Sm.) Rydb.	4x (4)	**II**	W USA
Interserial allopolyploids of series *Critesion* and *Sibirica*			
H. brachyantherum Nevski	4x (3)	**II**	W North America, Kamchatka, Newfoundland
H. fuegianum Bothmer et al.	4x (3)	**II**	S Argentina, S Chile
H. guatemalense Bothmer et al.	4x (1)	**II**	Guatemala, S Mexico
H. jubatum L.	4x (4)	**II**	NE Asia, NW+W North America, C Argentina
H. tetraploidum Covas	4x (4)	**II**	C Argentina
H. arizonicum Covas	6x (3)	**III**	SW USA
H. lechleri (Steud.) Schenk	6x (7)	**III**	C+S Argentina
H. parodii Covas	6x (4)	**III**	C Argentina
H. procerum Nevski	6x (4)	**III**	S Argentina
Section *Nodosa* (Nevski) Blattner			
H. brachyantherum Nevski	6x (2)	**IIXa**	C California

H. capense Thunb.	4x (2)	**IXa**	S Africa
H. secalinum Schreb.	4x (4)	**IXa**	Mediterranean to W Europe
Outgroup species			
Dasypyrum villosum (L) Candargy	2x (2)	**V**	
Eremopyrum triticeum (Gaertn.) Nevski	2x (1)	**FXe**	
Taeniatherum caput-medusae (L.) Nevski	2x (1)	**Ta**	
Triticum monococcum L.	2x (1)	**A^m**	
Triticum urartu Tumanian ex Gandilyan	2x (1)	**A^u**	
Secale strictum (C. Presl) C. Presl	2x (1)	**R**	
Secale vavilovii Grossh.	2x (1)	**R**	

[1] Number of individuals included per species or cytotype.

[2] Species with subspecies not further detailed here.

MOLECULAR METHODS

Genomic DNA was extracted from approximately 10 mg of silica gel-dried leaves with the DNeasy Plant Mini Kit (Qiagen) according to the protocol of the manufacturer. DNA quality and concentrations were checked on 1% agarose gels.

TOPO6 was amplified as described in Jakob & Blattner (2010) using primers Top6-15F (5'-GTG YTG TST YCA ACT GAA GTC-3') and Top6-17R (5'-CGT ACT CCA RYG CCA TTT C-3') designed to bind in exons 15 and 17 of the gene. Thus the amplification products consist of introns 15 and 16 together with exon 16, and are of lengths between 800 and 1200 base pairs (bp) in many pooid grasses (Blattner, unpublished). PCR was performed for all except 12 individuals obtained after the first sequencing results, using 1 U of a standard DNA polymerase (Qiagen Taq DNA polymerase) in 50 µl reaction volume containing approximately 10-50 ng of genomic DNA, 1x Coral Load PCR Buffer (Qiagen), 1x Q Solution (Qiagen), 1.25 mM $MgCl_2$, 0.2 mM of each dNTP, and 0.5 µM of each primer. The amplification process consisted of initial denaturation for 2 min at 95 °C, followed by 35 cycles of 30 sec at 96 °C, 1 min at 56 °C, 2.3 min at 72 °C, and a final extension of 12 min at 72 °C. To reduce PCR errors, which become visible when amplicons are cloned, amplification of *TOPO6* was performed in 27 recalcitrant polyploids and in five diploid individuals using 1 U proof-reading polymerase (Finnzymes OY, Phusion Hot Start DNA polymerase) with the same PCR conditions as before but using the supplied 1x Phusion HF Buffer. Amplification conditions were modified as suggested by the provider with higher denaturation (98 °C) and annealing temperatures (59 °C).

Amplicons were purified using Nucleofast 96 Spin Plates (Macherey-Nagel) according to the protocol of the manufacturer, eluted in 20 µl of TE buffer, and sequenced on an ABI 3730XL

automatic DNA sequencer (Applied Biosystems). For most of the diploid *Hordeum* and outgroup species amplicons were directly sequenced, while for all polyploids and eight diploids amplicons were ligated into the pJET1.2 vector (Fermentas) and transformed into DH5α *E. coli* strains. On average 12 individual colonies per individual were randomly selected for screening the insertion of a *TOPO6* fragment via PCR employing the primers pJET-F and pJET-R (Fermentas). Colonies showing products of the correct size (around 900 bp) were transferred to 200 µl LB broth medium with 0.1 mg/ml ampicillin and incubated overnight at 37 °C. A total of 952 colonies plasmids were isolated, and for each clone 1 µl was used for sequencing forward and reverse strands as described in Blattner, 2004 using primers pJET-F and pJET-R. For eight diploid individuals six to ten clones were sequenced to test if *TOPO6* behaves like a single-copy locus, while for all polyploids an average of 12 clones each were sequenced.

Data Analyses

Manual editing of sequences and multiple sequence alignments were performed with Geneious Pro v5.4 (Drummond *et al.*, 2011). Most sequences obtained by cloning of amplicons of regular *Taq* polymerase from single individuals had very similar sequences, differing in only one to ten mutations, which were not shared by more than one clone. These differences are very likely PCR errors from the *Taq* polymerase that occur during cloning or sequencing. In these cases consensus sequences for highly similar sequences were created in order to reduce the number of singletons in the alignment.

Chimerical sequences can be the result of natural recombination between alleles of orthologous or homoeologous genes and/or PCR-mediated recombination. Allopolyploid species are especially prone to the formation of chimerical sequences due to the presence of two or more homoeologous copies. Bifurcating phylogenetic trees cannot represent precisely the evolutionary histories of recombinant sequences and the presence of chimerical sequences disturbs analysis algorithms, as they combine signals from different phylogenetic groups. Automated methods included in RDP3 (Heath *et al.*, 2006; Martin *et al.*, 2010) to account for recombination events were used but the results were not conclusive due to the high number of PCR-mediated mutations present in the raw data set. Therefore, sequences were thoroughly inspected by eye to identify sequences showing combinations of polymorphic sites present in different alleles (Fig. S2.1). Recombinant sequences were excluded from the data set prior to the analysis. In cases of identical sequences derived from the same individual only one sequence was included in the data analysis.

After a preliminary analysis, out of the 341 individuals sequenced, a reduced dataset representative of all the diversity found was used for the different analyses. Thus alleles shared by more than one individual per species were included only once. The final alignment consisted of 278

sequences. This dataset and a subset containing only the sequences derived from diploid species and cytotypes consisting of 109 sequences were analyzed using parsimony and Bayesian methods. A maximum parsimony (MP) analysis was conducted in PAUP* 4b10 (Swofford, 2002) using the two-step procedure described in Blattner (2004). In an initial heuristic search with 1000 random addition sequences and TBR branch swapping the number of trees retained was restricted to five per random addition. The best trees from this search were used as starting trees in a second heuristic search utilizing TBR branch swapping, restricting the number of saved trees to 50,000. To test the statistical support of clades a bootstrap analysis with 50,000 re-samples and the fast-and-stepwise algorithm was conducted in PAUP*.

For Bayesian inference (BI), different models of sequence evolution were investigated with MrModeltest version 2.3 (Nylander, 2004). As partitioning of the data to account for intron/exon differences did not change the outcome of an initial BI analysis we inferred an overall model of sequences evolution for the entire marker region. Among the 24 models tested, the best-fit model selected by the hierarchical Likelihood Ratio Test (hLRT; Posada & Crandall, 2001) and Akaike Information Criterion (AIC; Akaike, 1973, 1974) was HKY+Γ, a transition/transversion model (Hasegawa *et al.*, 1985) with rates variation according to a gamma distribution (Yang, 1994). The analysis conducted with MrBayes version 3.1.2 (Huelsenbeck & Ronquist, 2001) consisted of two Metropolis Coupled Monte Carlo Markov Chain analyses with six chains per run for 5 million generations and sampling trees every 1000 generations. The temperature parameter was set at 0.05 to obtain a value of state swap frequency within the range of 10% to 70%. The convergence of the parameters was evaluated with the standard deviation of split frequencies (<0.01) and with the program Tracer version 1.5 (Rambaut & Drummond, 2007). The topology convergence was checked using the compare function of the online application AWTY (Nylander *et al.*, 2008), which plots posterior probabilities of clade support values for both runs against each other. The first 25% trees were discarded as burn-in and a consensus tree was computed in MrBayes 3.1.2.

To visualize species relationships, the final analysis was summarized in a schematic tree as proposed in Blattner (2004). In this scheme, the phylogeny of sequences derived from diploid species was used as a backbone, and the polyploid species were connected to the diploids according to the placement of their sequences in the complete analysis.

RESULTS

The *TOPO6* sequences obtained in this study varied in lengths between 868 and 1057 bp and were stored in the EMBL nucleotide database under accession numbers HE655746-HE656023. The alignment of 278 *TOPO6* sequences was 1275 bp long and contained 367 variable sites (281 for

diploids only), of which 248 were parsimony-informative (205 for diploids only). All analysis algorithms resulted in very similar tree topologies, thus only the BI trees are presented (Fig. 2.1 for diploids only, Fig. 2.2 for the complete dataset), while results of the MP analyses are available as supplemental Figures S2.2 and S2.3. Both analyses are summarized in a scheme of the *TOPO6*-based species and cytotype relationships within *Hordeum* (Fig. 2.3).

All analyses revealed the sequences derived from *Hordeum* species to be monophyletic with one exception (Fig. 2.1). Two cloned sequences from a single diploid *H. brevisubulatum* individual (PI229753) clustered outside the *Hordeum* clade, together with *Eremopyrum triticeum*. Sequences from the four genome groups in *Hordeum* (**H**, **I**, **Xa**, **Xu**) were mainly found monophyletic (Fig. 2.1) with few exceptions: (*i*) the *H. brevisubulatum* sequences already mentioned, (*ii*) a tetraploid individual of *H. brevisubulatum* (BG156/07) having one sequence falling outside of the **I** clade in a polytomy together with the **Xa**+**I** clade, and (*iii*) two diploid individuals of *H. murinum* (**Xu** genome, Jakob & Blattner, 2010) with two clones clustering with **H** genome *H. bulbosum* (derived from *H. murinum* PI218078) and one clone clustering with **I** genome *H. pubiflorum* (from *H. murinum* BCC2002). The *Hordeum* clade received strong support with a posterior probability (pp) of 0.99 in BI. The **H** genome sequences formed the sistergroup to the remaining species, with **Xu** genome sequences grouping as sister to the clade including sequences derived from **Xa** and **I**-genome taxa.

In the **H**-genome clade (section *Hordeum*), sequences of both cytotypes (2x and 4x) of *H. bulbosum* clustered together (Fig. 2.2) in one strongly supported clade (1.0 pp). Between one and three similar sequences were recovered per individual (Figs. 2.1 and 2.2), indicating either duplication of the *TOPO6* locus in this species or strong allelic diversity within individuals. We did, however, retrieve no sequences indicating that a taxon outside the extant diploid of *H. bulbosum* contributed to the formation of the tetraploid cytotype, which confirms the autopolyploid origin of this cytotype.

The **Xu**-genome group (section *Trichostachys*) of the *H. murinum* taxon complex consists of one clade (1.0 pp) in the analysis of diploids (*H. murinum* subsp. *glaucum*) and two clades when polyploids are included (Figs. 2.1 and 2.2). This second clade consists only of sequences derived from polyploids, indicating the existence of alleles not occurring in extant diploid individuals (Jakob & Blattner, 2010).

In the **Xa**-genome clade (section *Marina*), sequences clustered in three strongly supported groups (Fig. 2.2): (*i*) *H. marinum* sequences (1.0 pp), (*ii*) *H. gussoneanum* (2x and 4x, type B), *H. capense* (type A), *H. secalinum* (type A) and hexaploid *H. brachyantherum* (type C) with 0.98 pp, and (*iii*) *H. gussoneanum* sequences derived from tetraploids only (type A, 1.0 pp). As before, this latter group indicates alleles that do not occur in extant diploid individuals of section *Marina*.

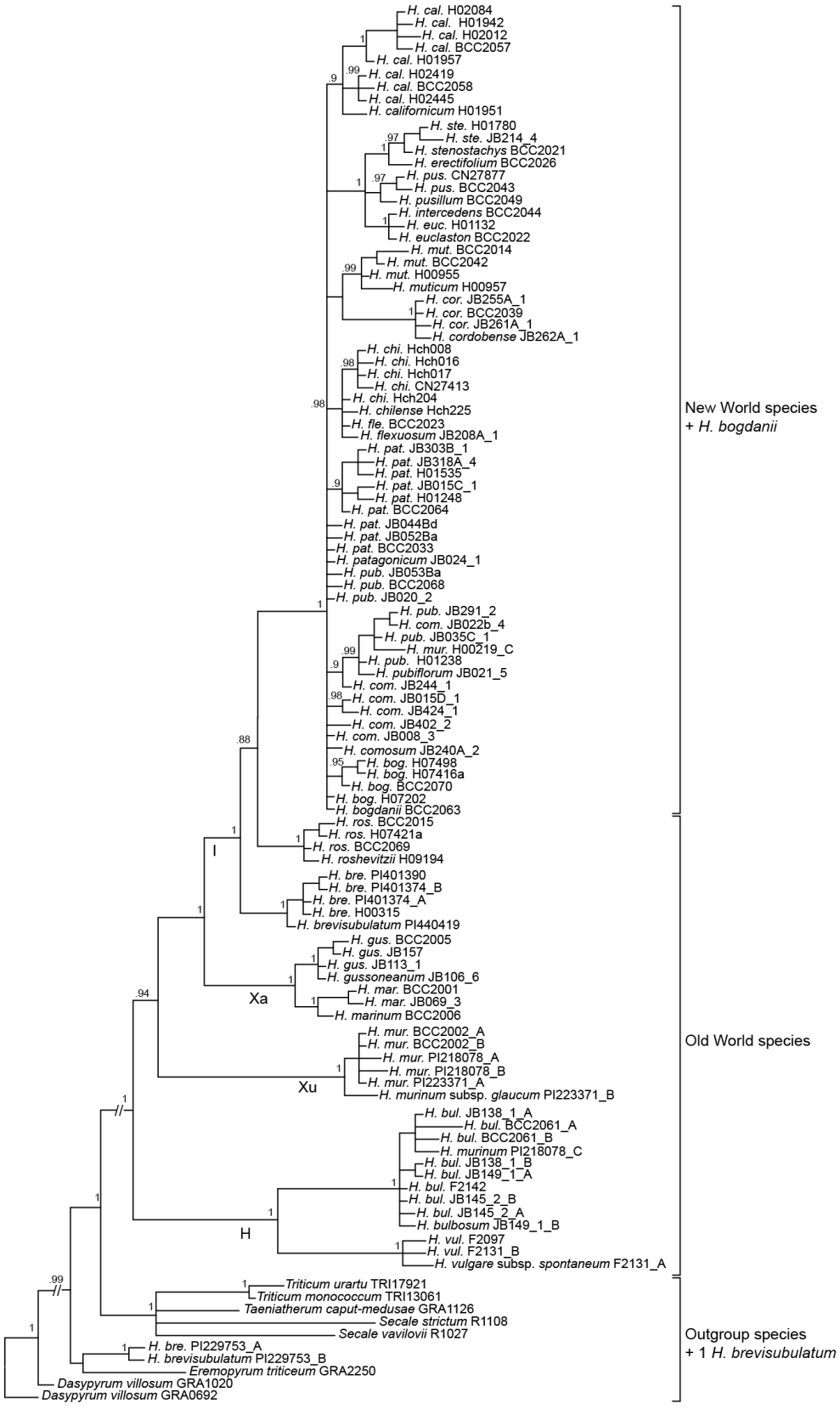


Figure 2.1 Phylogenetic tree derived from *TOPO6* sequences of the diploid *Hordeum* taxa and eight outgroup species calculated with Bayesian inference. Posterior probability values of the clades are indicated along the branches. Bold letters depict genome denominations following Blattner (2009).

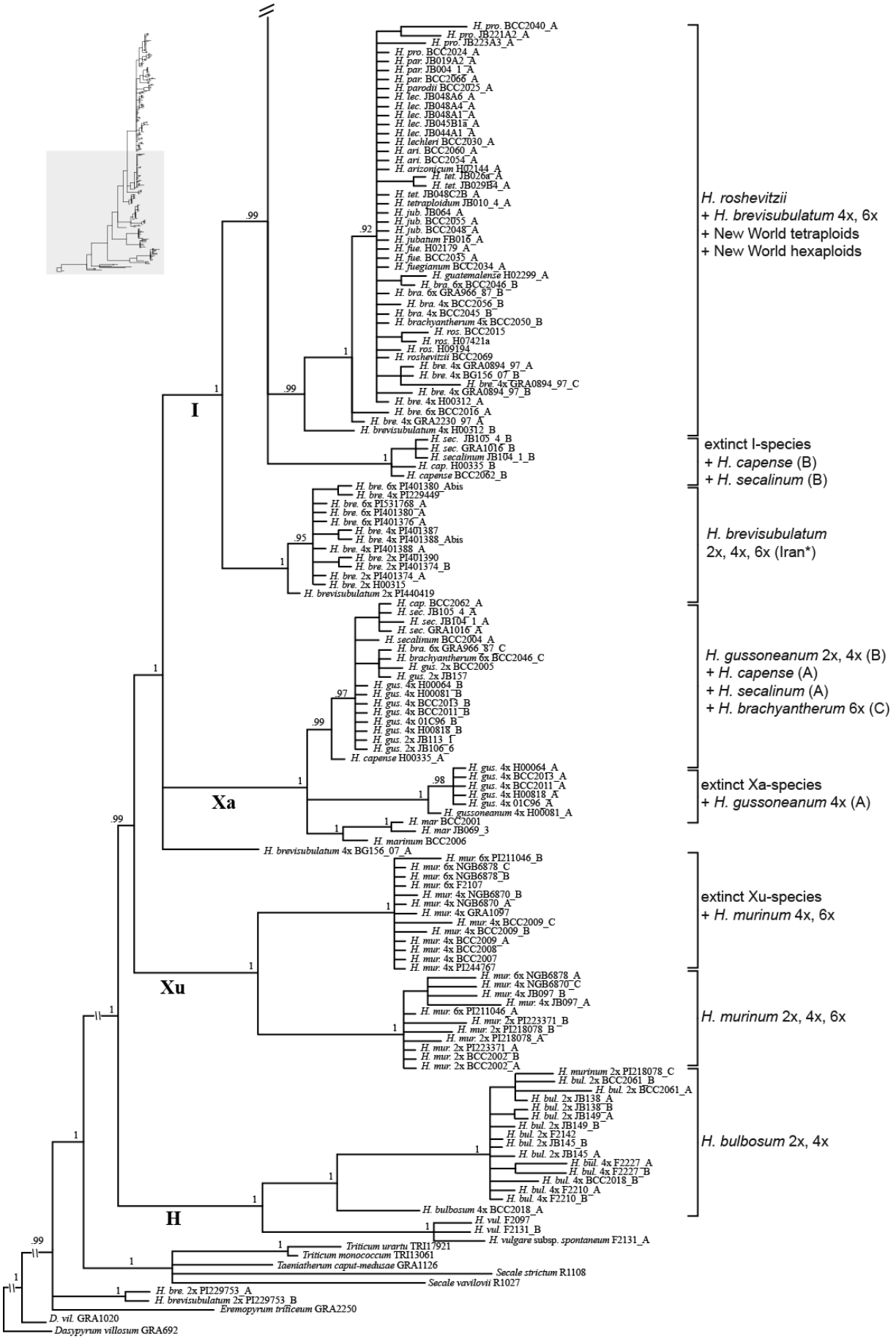
After the species name and individual number, the different copies found per individual are indicated (A-C) in case of cloned sequences.

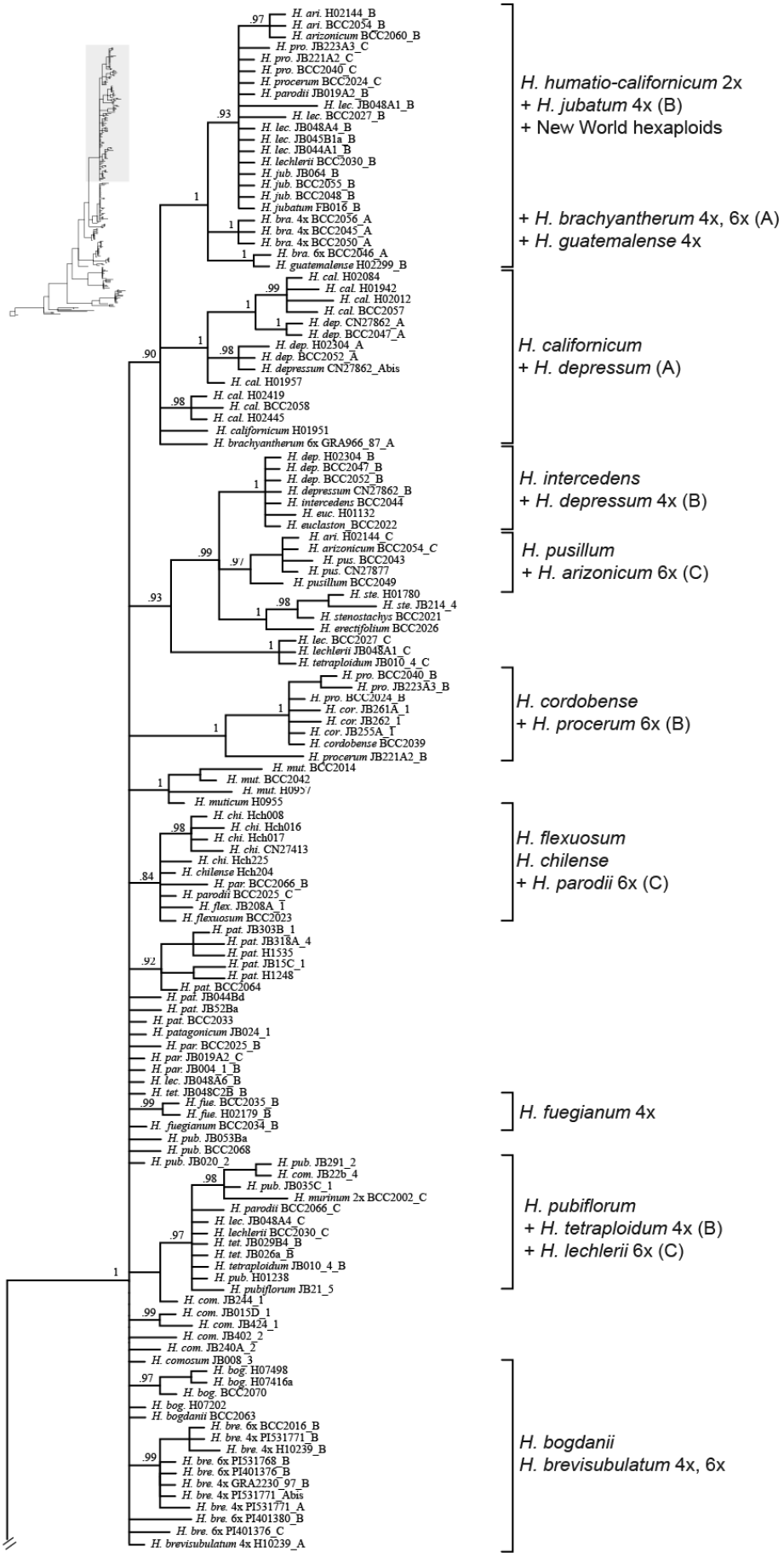
The **I-**genome group (section *Stenostachys*) consists of the Asian species *H. brevisubulatum* and *H. roshevitzii* being successive sister taxa to the large clade of sequences from Asian *H. bogdanii* together with all sequences derived from American species (Fig. 2.1). In this latter clade, sequences of single species and species groups are arranged along a large polytomy. *Hordeum bogdanii* and *H. californicum* sequences, as well as sequences of the three closely related Patagonian species *H. comosum*, *H. patagonicum*, and *H. pubiflorum*, occur directly along this polytomy or form clades of sequences derived from single species, while sequences from the other species form mixed clades.

Thus, *H. chilense* groups together with *H. flexuosum*, *H. cordobense* with *H. muticum*, and *H. euclaston* with *H. intercedens*, *H. pusillum*, *H. erectifolium*, and *H. stenostachys* (Fig. 2.1). The sequences derived from *H. bogdanii* possess a 33 bp region (alignment positions 925-957) shared by all Old World *Hordeum* species and absent in the American species (Fig. S2.4).

In the tree including polyploid-derived sequences (Fig. 2.2) in the **I-**genome group a clade (1.0 pp), consisting only of sequences from the three cytotypes (2x, 4x and 6x) of *H. brevisubulatum*, is sister to a large clade with a basal trichotomy. Its first clade (1.0 pp) contains only sequences of *H. capense* and *H. secalinum* (type B) without any sequences from an extant diploid species. The second clade (0.99 pp) consists of sequences of tetraploid and hexaploid *H. brevisubulatum* together with sequences originating from the diploid species *H. roshevitzii*, the tetraploid species *H. jubatum*, *H. brachyantherum*, *H. guatemalense*, *H. tetraploidum*, and *H. fuegianum*, and the hexaploid species *H. arizonicum*, *H. brachyantherum*, *H. lechleri*, *H. parodii*, and *H. procerum*. In the third clade, in addition to the diploid-derived sequences, Asian polyploid *H. brevisubulatum* and all American polyploids group mostly together with specific diploids into smaller subclades. Thus, *H. tetraploidum* (4x), *H. lechleri* (6x) and *H. parodii* (6x) fall in a clade with diploid *H. pubiflorum* and, augmented by *H. fuegianum* (4x), also in the group of sequences along the basal polytomy. *Hordeum parodii* (6x) groups together with diploid *H. flexuosum*, *H. procerum* (6x) with diploid *H. cordobense*, *H. arizonicum* (6x) with diploid *H. pusillum*, *H. depressum* (4x) with diploid *H. intercedens*, and *H. depressum* (4x) and *H. brachyantherum* (6x) also with diploid *H. californicum*. Finally, there is a clade consisting of sequences found in *H. brachyantherum* (4x), *H. guatemalense* (4x), *H. jubatum* (4x) and all American hexaploid species. In this latter clade no sequences derived from any extant diploid species can be found.

Next double-page:
Figure 2.2 Phylogenetic tree derived from cloned *TOPO6* sequences from diploid and polyploid *Hordeum* taxa and eight outgroup species calculated with Bayesian inference. Posterior probability values of the clades are indicated along the branches. Clades containing diploid and polyploid-derived sequences are indicated on the right. Genome denominations are given in bold type.


H. pro. BCC2040_A
H. pro. JB221A2_A
H. pro. JB223A3_A
H. pro. BCC2024_A
H. par. JB019A2_A
H. par. JB004_1_A
H. par. BCC2065_A
H. parodii BCC2025_A
H. lec. JB048A6_A
H. lec. JB048A4_A
H. lec. JB048A1_A
H. lec. JB045B1a_A
H. lec. JB044A1_A
H. lechleri BCC2030_A
H. ari. BCC2060_A
H. ari. BCC2054_A
H. arizonicum H02144_A
H. tet. JB026a_A
H. tet. JB029B4_A
H. tet. JB048C2B_A
H. tetraploidum JB010_4_A
H. jub. JB064_A
H. jub. BCC2055_A
H. jub. BCC2048_A
H. jubatum FB016_A
H. fue. H02179_A
H. fue. BCC2035_A
H. fuegianum BCC2034_A
H. guatemalense H02299_A
H. bra. 6x BCC2046_B
H. bra. 6x GRA966_87_B
H. bra. 4x BCC2056_B
H. bra. 4x BCC2045_B
H. brachyantherum 4x BCC2050_B
H. ros. BCC2015
H. ros. H07421a
H. ros. H09194
H. roshevitzii BCC2069
H. bre. 4x GRA0894_97_A
H. bre. 4x BG156_07_B
H. bre. 4x GRA0894_97_C
H. bre. 4x GRA0894_97_B
H. bre. 4x H00312_A
H. bre. 6x BCC2016_A
H. bre. 4x GRA2230_97_A
H. brevisubulatum 4x H00312_B
H. sec. JB105_4_B
H. sec. GRA1016_B
H. secalinum JB104_1_B
H. cap. H00335_B
H. capense BCC2062_B
H. bre. 6x PI401380_Abis
H. bre. 4x PI229449
H. bre. 6x PI531768_A
H. bre. 6x PI401380_A
H. bre. 6x PI401376_A
H. bre. 4x PI401387
H. bre. 4x PI401388_Abis
H. bre. 4x PI401388_A
H. bre. 2x PI401390
H. bre. 2x PI401374_B
H. bre. 2x PI401374_A
H. bre. 2x H00315
H. brevisubulatum 2x PI440419
H. cap. BCC2062_A
H. sec. JB105_4_A
H. sec. JB104_1_A
H. sec. GRA1016_A
H. secalinum BCC2004_A
H. bra. 6x GRA966_87_C
H. brachyantherum 6x BCC2046_C
H. gus. 2x BCC2005
H. gus. 2x JB157
H. gus. 4x H00064_B
H. gus. 4x H00081_B
H. gus. 4x BCC2013_B
H. gus. 4x BCC2011_B
H. gus. 4x 01C96_B
H. gus. 4x H00818_B
H. gus. 2x JB113_1
H. gus. 2x JB106_6
H. capense H00335_A
H. gus. 4x H00064_A
H. gus. 4x BCC2013_A
H. gus. 4x BCC2011_A
H. gus. 4x H00818_A
H. gus. 4x 01C96_A
H. gussoneanum 4x H00081_A
H. mar BCC2001
H. mar JB069_3
H. marinum BCC2006
H. brevisubulatum 4x BG156_07_A
H. mur. 6x PI211046_B
H. mur. 6x NGB6878_C
H. mur. 6x NGB6878_B
H. mur. 6x F2107
H. mur. 4x NGB6870_B
H. mur. 4x NGB6870_A
H. mur. 4x GRA1097
H. mur. 4x BCC2009_C
H. mur. 4x BCC2009_B
H. mur. 4x BCC2009_A
H. mur. 4x BCC2008
H. mur. 4x BCC2007
H. mur. 4x PI244767
H. mur. 6x NGB6878_A
H. mur. 4x NGB6870_C
H. mur. 4x JB097_B
H. mur. 4x JB097_A
H. mur. 6x PI211046_A
H. mur. 2x PI223371_B
H. mur. 2x PI218078_B
H. mur. 2x PI218078_A
H. mur. 2x PI223371_A
H. mur. 2x BCC2002_B
H. mur. 2x BCC2002_A
H. murinum 2x PI218078_C
H. bul. 2x BCC2061_B
H. bul. 2x BCC2061_A
H. bul. 2x JB138_A
H. bul. 2x JB138_B
H. bul. 2x JB149_A
H. bul. 2x JB149_B
H. bul. 2x F2142
H. bul. 2x JB145_B
H. bul. 2x JB145_A
H. bul. 4x F2227_A
H. bul. 4x F2227_B
H. bul. 4x BCC2018_B
H. bul. 4x F2210_A
H. bul. 4x F2210_B
H. bulbosum 4x BCC2018_A
H. vul. F2097
H. vul. F2131_B
H. vulgare subsp. spontaneum F2131_A
Triticum urartu TRI17921
Triticum monococcum TRI13061
Taeniatherum caput-medusae GRA1126
Secale strictum R1108
Secale vavilovii R1027
H. bre. 2x PI229753_A
H. brevisubulatum 2x PI229753_B
Eremopyrum triticeum GRA2250
D. vil. GRA1020
Dasypyrum villosum GRA692
I
Xa
Xu
H
H. roshevitzii
+ H. brevisubulatum 4x, 6x
+ New World tetraploids
+ New World hexaploids
extinct I-species
+ H. capense (B)
+ H. secalinum (B)
H. brevisubulatum
2x, 4x, 6x (Iran*)
H. gussoneanum 2x, 4x (B)
+ H. capense (A)
+ H. secalinum (A)
+ H. brachyantherum 6x (C)
extinct Xa-species
+ H. gussoneanum 4x (A)
extinct Xu-species
+ H. murinum 4x, 6x
H. murinum 2x, 4x, 6x
H. bulbosum 2x, 4x


H. humatio-californicum 2x
+ H. jubatum 4x (B)
+ New World hexaploids
+ H. brachyantherum 4x, 6x (A)
+ H. guatemalense 4x
H. californicum
+ H. depressum (A)
H. intercedens
+ H. depressum 4x (B)
H. pusillum
+ H. arizonicum 6x (C)
H. cordobense
+ H. procerum 6x (B)
H. flexuosum
H. chilense
+ H. parodii 6x (C)
H. fuegianum 4x
H. pubiflorum
+ H. tetraploidum 4x (B)
+ H. lechlerii 6x (C)
H. bogdanii
H. brevisubulatum 4x, 6x

For some tetraploid Iranian individuals of *H. brevisubulatum*, we obtained no indication that other taxa than diploid *H. brevisubulatum* contributed *TOPO6* sequences, suggesting an autopolyploid origin of these individuals.

The *TOPO6* alignment revealed the presence of two insertions of transposable elements (TE). Thus, a TE of variable size ranging from 78 bp (in *Secale*) to 123 bp (in *Triticum*) is present at alignment positions 968–1092 of the outgroup species. BLAST searches (Altschul *et al.*, 1990) of this element retrieved sequences featuring a *Stowaway* miniature inverted repeat transposable element (MITE). All *Hordeum* species plus *Dasypyrum villosum* and *Eremopyrum triticeum* were missing this element. In addition, the *H. bulbosum* accession BCC2018 had an allele with an insertion of 162 bp similar to a *Stowaway* MITE annotated as *Thalos* (EMBL sequence accession number AF521177.1, position 76813-76977; 92% similarity) that is located in a TA target site (alignment positions 709–872).

Recombinant Clones of TOPO6 *and Comparison of Two Polymerases*

Among the 945 clone-derived sequences obtained for this study, 365 from 19 tetra- or hexaploid individuals could be analyzed for the influence of proof-reading versus standard DNA polymerases. Separate PCRs were conducted for these individuals, with both standard and proof-reading enzymes. For this subset, *Taq* resulted in an average of 40% (SD = 0.24) of chimerical sequences, while the Phusion Hot Start DNA Polymerase resulted in only half of this proportion (22%, SD = 0.19). This difference was significant (paired Student's t-Test, $p = 0.01$) when testing for a higher proportion of chimerical sequences with the regular *Taq* than with the Phusion enzyme. Concerning only the tetraploids (10 individuals), we found a significant difference (paired Wilcoxon signed rank test, $p = 0.04$) with the same alternative hypotheses. In the case of hexaploids (nine individuals), no significant difference was found (paired Student's t-Test). This result might be due to the small amount of comparisons and/or to the presence of more true chimerical sequences present in the genome of hexaploid individuals.

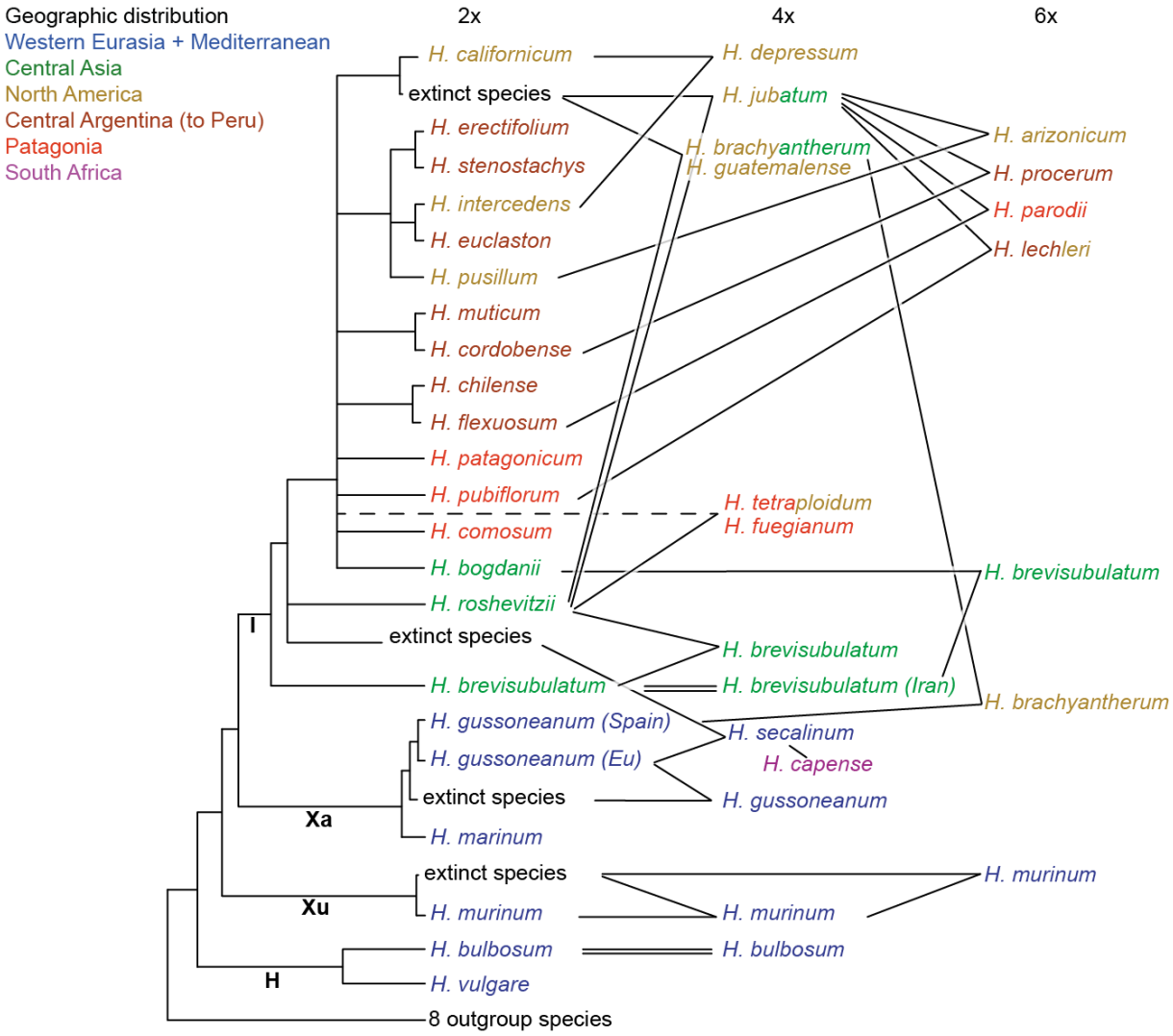


Figure 2.3 Scheme summarizing phylogenetic relationships of species and cytotypes in the genus *Hordeum* based on *TOPO6*. Diploid taxa were drawn directly at the tree, while tetra- and hexaploids were connected by lines to their inferred parental taxa. Dashed lines indicate uncertainties and double lines depict an autopolyploid origin.

DISCUSSION

Our phylogenetic analysis revealed *TOPO6* to be a single-copy locus in most diploid *Hordeum* species, as we found regularly only two different sequences per diploid individual, which we interpret as allelic variation. In a few cases we detected more than the expected maximum of two, four and six alleles in diploid, tetraploid and hexaploid individuals, respectively. However, except for three surprising cases (*H. murinum* H00219 and PI218078 and *H. bulbosum* BCC2018), these were very similar to each other so that we cannot safely discern if they stem from gene duplication or are artifacts originating during DNA amplification and sequencing. However, as using a proof-reading

DNA polymerase in PCR greatly reduced the number of sequence polymorphisms within individuals, we assume that these are essentially PCR artifacts.

The phylogenetic relationships obtained from *TOPO6* are mainly congruent with previous work (reviewed in Blattner, 2009) although they deviate from the total evidence phylogenetic tree of *Hordeum* (Blattner, 2009). This is not an unexpected result, as up to now no single marker region was able to 'correctly' resolve all species relationships among diploid *Hordeum* species (Blattner, 2009; Petersen *et al.*, 2011). The major difference (Fig. 2.1) is the non-monophyly of subgenus *Hordeum* due to *H. murinum* being sister to subgenus *Hordeastrum* instead of the **H**-genome taxa, yet with weak statistical support. Minor differences, although no incompatibilities, concern the positions of (*i*) *H. bogdanii* and (*ii*) *H. californicum*, which group in the large polytomy of the New World species instead of *H. bogdanii* being sister to *H. californicum* plus the clade of the mainly South American taxa, and (*iii*) the non-monophyly of sequences derived from the three closely related Patagonian species *H. comosum*, *H. patagonicum*, and *H. pubiflorum*. Sequences of these species also group along the large polytomy in Figure 2.1, providing no hard contradiction to monophyly of this group. As the *TOPO6* sequences derived from polyploids mostly cluster with sequences obtained from specific diploid species in phylogenetic analyses, they enable the identification of the parental species involved in polyploid formation, which is the major goal of this study.

Inference of Extinct Diploid Progenitors of Allopolyploids

Four statistically well-supported and genetically distinct clades in the phylogenetic trees contain only sequences derived from polyploid species but lack sequences of a currently existing diploid taxon: (*i*) in *H. murinum* (1.0 pp), (*ii*) in *H. gussoneanum* (1.0 pp), (*iii*) for *H. capense*/*H. secalinum* in the **I**-genome group (1.0 pp), and (*iv*) within the *H. californicum* group (1.0 pp). As we included all taxa of *Hordeum* in this study, mostly with multiple individuals representing the distribution areas of the species, we can safely interpret these sequences as the footprints of diploid species, which contributed their genome to allopolyploid taxa in the past and went extinct sometime after polyploid formation (Roelofs *et al.*, 1997). These data support similar findings from ITS and *EF-G* sequences for *H. murinum*; (Blattner, 2004; Jakob & Blattner, 2010; Tanno *et al.*, 2010) and ITS sequences for *H. californicum*-related taxa (Blattner, 2004), *EF-G* data for tetraploid *H. gussoneanum* (Komatsuda *et al.*, 2001), and are compatible with the results of analyses of the *HTL* gene (Kakeda *et al.*, 2009) and ITS (Blattner, 2004) in *H. gussoneanum*. For *H. capense*/*H. secalinum* the identification of an extinct diploid progenitor in addition to *H. gussoneanum* is new. Based on the position of the *TOPO6* sequences between the clades formed by diploid *H. brevisubulatum* and *H. roshevitzii* (Fig. 2.2), we conclude that this extinct taxon belonged to the Central Asian group of *Hordeum* species, *i.e.* series *Sibirica*. Extinct progenitors of allopolyploids have been previously inferred (Roelofs *et al.*, 1997;

Blattner, 2004; Hoot *et al.*, 2004; Ayres *et al.*, 2004; Lihovà *et al.*, 2006). However, the extent of this phenomenon is currently unclear, as the inclusion of single or few individuals as representatives of species is still common. This does not allow for the discerning of closely related species from conspecific individuals with high intraspecific genetic variation. Thus, Petersen & Seberg (2004) obtained one type of *DMC1* sequences found in *H. capense* and *H. secalinum* as sistergroup of *H. brevisubulatum* while sequences of the second type grouped in a polytomy together with the sequence of *H. gussoneanum*. Accordingly, they interpreted this topology as an indication for *H. gussoneanum* and *H. brevisubulatum* being the parents of the tetraploids. The strong differentiation between their *H. capense*/*H. secalinum* and *H. brevisubulatum* sequences in *DMC1* are, however, completely compatible with the scenario we propose here, *i.e.* that the genepools of diploid *H. brevisubulatum* and the extinct progenitor of *H. capense*/*H. secalinum* belonged to separate taxa. Particularly as neither *DMC1* nor *TOPO6* showed a comparably large differentiation for the tetraploid's sequences derived from *H. gussoneanum*, which should be the case if differentiation of the homoeologues evolved only after polyploid formation.

Due to insufficient data on such extinct species it is currently not possible to infer if extinction rates for diploids generally rise after they contribute to polyploid formation and, thus, increase competition in their habitats (Rodriguez, 1996) or if we see the normal rate of background extinction in *Hordeum*. The ratio of 3:1 of Old versus New World extinct species maintained in polyploids fits with the proposed generally higher Pleistocene extinction rates in Eurasia in comparison to the Americas inferred from the distribution of chloroplast haplotypes in *Hordeum* (Jakob & Blattner, 2006).

Polyploid Species of the Old World

In *H. bulbosum*, occurring in the Mediterranean and adjacent Southwest Asia, two cytotypes (2x, 4x) exist. All sequences but one (BCC2018_A) derived from this species are quite similar and group in a single strongly supported clade (1.0 pp), while the single outlier is sister to this group. For the diploid individuals we found, as expected, up to two *TOPO6* alleles and also in the tetraploid no more than two alleles were detected. Finding only one kind of sequence in a tetraploid could result from the loss of one copy from the genome, gene conversion, PCR-drift, limited clone sampling or autopolyploidy. As we included multiple individuals of the tetraploid in our study we do not expect that all would behave in the same way regarding technical shortcomings. Thus, our result supports the long-standing assumption that the tetraploid cytotype of *H. bulbosum* is of autopolyploid origin (Linde-Laursen *et al.*, 1990). The peculiar position of the second type (A) of individual BCC2018 could indicate ancient introgression and/or incomplete lineage sorting (Jakob & Blattner, 2006) or pseudogenization of one *TOPO6* copy.

In the *H. murinum* taxon group, two clades of *TOPO6* sequences were obtained. As already inferred by Jakob & Blattner (2010) and Tanno *et al.* (2010) in detailed studies of these taxa and cytotypes, all polyploids are of allopolyploid origin involving extant and extinct species from within the **Xu**-genome group.

As discussed before, in the **Xa**-genome clade sequences obtained from *H. marinum* and *H. gussoneanum* (2x, 4x) formed three distinct clusters. The two copies of the tetraploid cytotype appeared in two of these clusters, one being exclusive to this cytotype (type A), which we interpret as an indication for an extinct diploid progenitor. Although *H. gussoneanum* was also involved in the evolution of three allopolyploid taxa combining **I** and **Xa** genomes, only extant *H. gussoneanum* contributed in these cases. One group of *TOPO6* sequences of the two tetraploid sister species *H. capense* and *H. secalinum* was clustering with the sequences derived from diploid *H. gussoneanum*, pinpointing this taxon as one parental species and are thus in accord with previous analysis (Taketa *et al.*, 1999; Blattner, 2004, 2006; Petersen & Seberg, 2004). However, we found no indications for a contribution of *H. marinum* to these tetraploids, as proposed by Taketa *et al*. (2009) based on cytogenetic and by Jakob & Blattner (2006) on chloroplast data. Our data also confirm that *H. capense* and *H. secalinum* are very closely related, contrarily to Baum & Johnson (2003), but clearly separated taxa, which most probably speciated after long-distance dispersal of *H. secalinum* from Europe to South Africa, resulting in the geographically isolated *H. capense* (Blattner, 2006).

Diploid *H. gussoneanum* introduced into North America contributed to the formation of the hexaploid cytotype of *H. brachyantherum* (haploid genome composition **IIXa**) via hybridization with *H. brachyantherum*, most probably in historic times. This result confirms previous observations (Taketa *et al.*, 1999; Blattner, 2004; Jakob & Blattner, 2006; Komatsuda *et al.*, 2009). Our *TOPO6* sequences indicate that *H. gussoneanum* individuals originating from Spain (BCC2005) might have formed the Californian population (JB157) and contributed to *H. brachyantherum* (6x), occurring only in the Californian Bay area.

HORDEUM ROSHEVITZII, *A KEY SPECIES IN THE EVOLUTION OF TETRAPLOID* HORDEUM *SPECIES*

In addition to polyploid Central Asian *H. brevisubulatum* (discussed in detail below) all American tetraploid species with the exception of *H. depressum*, *viz. H. brachyantherum*, *H. fuegianum*, *H. guatemalense*, *H. jubatum*, and *H. tetraploidum*, carry one *TOPO6* type grouping with the sequences of *H. roshevitzii*, a diploid species endemic to Central Asia. The second sequence type derived from these species clustered with different diploid species of the American clade. North American *H. brachyantherum* and *H. jubatum* and Central American *H. guatemalense* thus have the *roshevitzii*-like *TOPO6* copy plus one of the extinct species identified within the *H. californicum*

clade. In contrast to results from an ITS analysis (Blattner, 2004), we found no indication for introgression of *H. intercedens* in *H. guatemalense*. In South American *H. tetraploidum* and *H. fuegianum*, in addition to the *roshevitzii*-like type, *TOPO6* copies close to that of the three Patagonian species *H. comosum*/*H. patagonicum*/*H. pubiflorum* occur. For *H. tetraploidum*, the sequences seem to hint towards a polyphyletic origin, as a group of individuals possess a different *TOPO6* copy derived from *H. pubiflorum*. From our data, we cannot discern if *H. fuegianum* evolved from *H. tetraploidum* through a speciation event on the tetraploid level or if both taxa evolved through independent allopolyploidization involving the same parental species. To resolve this we would need a higher number of informative characters in DNA sequences and much more individuals of both species included in an analysis.

Overall, *H. roshevitzii* appears as the key species in the evolution of the tetraploid species of the New World. Our data confirm findings from Blattner (2004) using ITS, where *H. roshevitzii* formed a clade with American allopolyploid species, and also cytogenetic data (FISH) that infer a contribution of *H. roshevitzii* to American allopolyploids (Taketa *et al.*, 2005). Surprisingly, Wang & Sun (2011), based on *DMC1* sequences, did not find a contribution of *H. roshevitzii* to American polyploid *Hordeum* taxa. Instead they detected in some American tetraploid and hexaploid species (*H. jubatum*, *H. fuegianum*, *H. tetraploidum*, and *H. arizonicum*) *DMC1* copies with close relationships to other Hordeeae genera (*Taeniatherum* and *Pseudoroegneria*). The use of single individuals as representatives for taxa in this study makes it impossible to infer reasons for these differences, *i.e.* if this is a general feature of the polyploids or a peculiarity of the *DMC1* locus for single individuals.

In a geographic context, the occurrence of *roshevitzii*-like sequences in North American polyploids indicates a second colonization event from Asia to North America after the initial establishment of *bogdanii*-like *Hordeum* diploids on that continent (Blattner, 2006). The *roshevitzii*-like sequences of the two South American tetraploids *H. tetraploidum* and *H. fuegianum* require either initial colonization of South America by *H. roshevitzii* and its later extinction after the formation of the tetraploids or the introduction of this sequence types through a polyploid (*H. jubatum*, see below) and introgression of the alleles via hybridization.

The Extinct californicum-like Taxon, a Key Species in the Evolution of American Polyploids

The extinct close relative of *H. californicum*, to which we refer informally as *H. humatio-californicum* (due to the fact that it is closely related to *H. californicum* but that its genome today can only be found 'buried' in polyploid taxa), contributed its genome to tetraploid *H. brachyantherum* and *H. jubatum* and all New World hexaploids, *i.e. H. arizonicum* from North America and South American *H. lechleri*, *H. parodii* and *H. procerum*. Therefore, its importance for the evolution of

Hordeum polyploids is comparable to that of *H. roshevitzii*. The American hexaploids all show essentially the *TOPO6* homoeologue pattern of tetraploid *H. jubatum*. In addition, each hexaploid has a third copy clustering with American diploid species. Thus, the formation of these hexaploids can be explained by hybridization of *H. jubatum* with (*i*) diploid *H. pusillum* resulting in *H. arizonicum*, with (*ii*) *H. cordobense* resulting in *H. procerum*, with (*iii*) *H. pubiflorum* resulting in *H. lechleri*, and with (*iv*) *H. flexuosum* or, geographically less likely, *H. chilense* resulting in *H. parodii*. While the first two combinations seem plausible regarding the partly overlapping distribution areas of the species involved, and the third also for the overall high morphological similarity of *H. lechleri* with its proposed progenitors, in *H. parodii* we found only one individual with the *H. humatio-californicum* sequence, whereas four individuals possess a *TOPO6* copy related to *H. comosum*/*H. patagonicum*/*H. pubiflorum*. In this latter case, it might also be that *H. tetraploidum* instead of *H. jubatum* contributed to hexaploid formation and later-on introgression with *H. jubatum* or one of the other hexaploids contributed to additional genomic diversity. Alternatively, also a polyphyletic origin of *H. parodii* is compatible with the data. Chloroplast sequences do not contribute to the clarification of this topic, as apart from *H. lechleri*, all South American hexaploids possess chloroplast haplotypes derived from their South American diploid progenitors, *i.e. H. parodii* chloroplast haplotypes are shared with diploid *H. patagonicum* and *H. pubiflorum* and tetraploid *H. tetraploidum* (Jakob & Blattner, 2006).

In any case, we have to assume that *H. jubatum* was present in South America for a long enough time to allow the evolution of at least three hexaploid taxa and their expansion to their extant, partly very large and mainly allopatric, distribution areas. This is in contradiction to the present assumption of a natural distribution of *H. jubatum* only in north-western North America and north-eastern Siberia (Bothmer *et al.*, 1995) and its introduction into other areas of the world as ornamental in historic times. From the data obtained from the South American polyploids, we infer the natural occurrence of *H. jubatum* in southern South America prior to European settlements and trade routes in this area. Thus, the scattered stands of this taxon in the grasslands of Central Argentina might well have originated by ancient bird-mediated long-distance dispersal of this taxon from North America (Blattner, 2006; Jakob & Blattner, 2006).

HORDEUM DEPRESSUM, AN ALL-AMERICAN TETRAPLOID

Hordeum depressum possesses a *TOPO6* copy (A) derived from *H. californicum* and a copy (B) clustering with diploid *H. intercedens* and *H. euclaston*. As it obtained no *roshevitzii*-like *TOPO6* copy, it is the only 'purely' American tetraploid species. *Hordeum intercedens* from south-western California and adjacent Mexico phylogenetically groups within South American *H. euclaston* (Pleines & Blattner, 2008). Pronounced ecological differentiation (Jakob *et al.*, 2010) together with the geographically caused *de facto* reproductive isolation (Blattner, 2006; Pleines & Blattner, 2008)

warrants the recognition as two independently evolving lineages, *i.e.* separate species. In contrast to Wang & Sun (2011), who assumed that *H. euclaston* is a parent of *H. depressum*, we propose that in the frame of the geographical co-occurrence of *H. californicum* and *H. intercedens* in southern California these species are much more likely to hybridize and thus to have contributed to the evolution of the tetraploid.

HORDEUM BREVISUBULATUM, *A COMPLEX GROUP*

The taxon complex of *H. brevisubulatum* comprises five described subspecies from a large geographic area reaching from western Turkey to north-eastern China with diploid, tetraploid and hexaploid cytotypes. Identification of the different subspecies is delicate especially for herbarium samples and materials from areas where subspecies overlap, thus we decided to consider only the ploidy level and the country of origin. The species is assumed to be of autopolyploid origin (Blattner, 2004). We were able to discern two major groups, with Iranian individuals found to be molecularly different from individuals from the rest of the species' distribution area. With one exception (H00312) the tetraploid individuals of *H. brevisubulatum* originating from Iran appeared to be autopolyploid, as the *TOPO6* sequences derived from these individuals clustered with the diploids only. Surprisingly, polyploid individuals originating from Siberia did not have a diploid *brevisubulatum*-like copy but instead one type of sequences derived from these eastern individuals clustered with *H. roshevitzii*. The other copy was either found in a peculiar position on a polytomy with the clades formed by **Xa** and **I**-genome sequences (BG156_07) or in the New World clade. In this latter clade they cluster together with sequences recovered from a tetraploid accession from Kirgizstan and a Tajik individual. For the hexaploid cytotypes, the individuals from Iran had a diploid *brevisubulatum*-like copy and the other one was clustering in the New World clade together with other *H. brevisubulatum* sequences. The hexaploid Tajik accession (BCC2016) had one sequence falling in the *H. roshevitzii* clade and one in the tetraploid/hexaploid *H. brevisubulatum* clade embedded in the New World clade. The *H. brevisubulatum* sequences present in this latter clade are similar to *H. bogdanii*, especially regarding the absence of the 33 bp long deletion characteristic for the American taxa.

To summarize our findings regarding *H. brevisubulatum*, according to the *TOPO6* phylogeny, only the Iranian *H. brevisubulatum* tetraploids are of autopolyploid origin, while in the remaining distribution area allopolyploids occur, which do not even include the diploid's *TOPO6* type. This species complex thus seems polyphyletic and/or exhibits signs of long-term interspecific hybridization with the three diploid Asian taxa *H. brevisubulatum, H. bogdanii* and *H. roshevitzii*. Moreover, *H. brevisubulatum* is the only species having one diploid individual with a *TOPO6* sequence clustering with the outgroup species *Eremopyrum*, indicating introgression from outside of *Hordeum* (Mahelka & Kopecký, 2010; Wang & Sun, 2011). What we cannot estimate is if the genetic diversity found here

represents the natural state of the species' lineages or if hybridization occurred during reproductive cycles in germplasm repositories. Obligate outbreeding of *H. brevisubulatum* (Bothmer *et al.*, 1995) might facilitate introgression in comparison to most other *Hordeum* species.

PCR Recombination Results in Chimerical Sequences

Conducting this study, we tested for the influence of two different DNA polymerases on the proportion of chimerical sequences, combining parts of different *TOPO6* types occurring in *Hordeum*. Lahr & Katz (2009) found that for genes of the major histocompatibility complex proof-reading DNA polymerases greatly reduce the proportion of recombinant sequences originating during PCR. In *Hordeum*, we found the same effect; however, it seemed to vary according to the ploidy level. Indeed, in tetraploid individuals the reduction of chimerical amplicons was much higher than in hexaploids. There are two possible explanations for this result. The first reason may be due to a lack of statistical power, as fewer comparisons were conducted (we did not increase the amount of screened colonies for hexa- in comparison to tetraploids). The second might be due to biological reasons, as recombinant copies might already be present in the genome. Cronn *et al.* (2002) also found recombinant sequences of low-copy number genes in allopolyploid cotton species, which indicates that this phenomenon is not restricted to *Hordeum*. Although low-copy nuclear genes are very promising to reconstruct species phylogenies especially in polyploids (Sang, 2002; Small *et al.*, 2004), the problem of recombinant sequences limits the usage of such markers. The origin of chimerical sequences during PCR amplification seems inherent to the type of sequences and to the use of universal primers, as they uniformly amplify all homoeologues, resulting in a mixture of amplicons. To minimize this problem one can design homoeologue-specific primers (Cronn *et al.*, 2002; Petersen & Seberg, 2004; Lihovà *et al.*, 2006). This means, however, that sequence information for the homoeologues has to be present in advance and prevents naïve exploring of allele diversity. Another solution, using single-molecule (sm) PCR (Kraytsberg *et al.*, 2008), was employed by Marcussen *et al.* (2012) to disentangle reticulate evolution in *Viola*. The smPCR seems particularly suitable for high-polyploid species and reduced sampling size. Nevertheless, from our experience in *Hordeum* we suggest that the use of a proof-reading DNA polymerase, probably together with low initial DNA concentration in PCR (Lahr & Katz, 2009), can reduce potential ambiguities regarding artificial recombinant amplicons. However, we are well aware that this cannot supersede a careful inspection of the data, particularly when higher ploidy levels are involved.

CONCLUSIONS

Using cloned sequences of *TOPO6*, a low-copy nuclear region, in a comprehensive framework including all *Hordeum* species mostly with several individuals, covering the geographic distribution of the species, we were able to infer parental relationships of *Hordeum* polyploids. The phylogenetic hypothesis presented here (Fig. 2.3) brought several new insights and supported other earlier data. Thus, it is likely that a close relative of Asian *H. bogdanii* was the starting point for the evolution of American diploid species. Diploid *H. roshevitzii* together with an extinct close relative of *H. californicum* and tetraploid *H. jubatum* were pivotal species for the evolution of the American allopolyploids. The involvement of *H. jubatum* in the formation of South American polyploids necessitates the presence of this taxon in South America well before the onset of European settlement. Thus, we propose that also Central Argentina belongs to the natural distribution area of this species, resulting in a disjunct Northern–Southern Hemisphere distribution of *H. jubatum*. We were able to analyze the status of two putative autopolyploids, confirming autopolyploidy for *H. bulbosum* and tetraploid Iranian populations of *H. brevisubulatum*, while the latter taxon shows otherwise very complex and still poorly understood allopolyploid patterns. This species complex, as well as some polyploids in South America, might result from multiple independent origins or had a long history of hybridization and introgression. The use of proof-reading DNA polymerase in PCR can reduce phylogenetic noise when analyzing polyploid sequences by cloning. More in-depth analyses have to be performed to resolve the still unclear parental relationship, for example in some cases using more loci to arrive at a higher resolution for closely related species groups (in South America) or by including much more geographically representative individuals from natural populations for species complexes (particularly in *H. brevisubulatum*).

3. A MULTILOCUS ANALYSIS: NEXT-GENERATION-SEQUENCING

SPECIES LEVEL PHYLOGENY AND POLYPLOID RELATIONSHIPS IN *HORDEUM* (POACEAE) INFERRED BY NEXT-GENERATION SEQUENCING AND *IN-SILICO* CLONING OF MULTIPLE NUCLEAR LOCI

An updated version of this chapter was published by Brassac and Blattner in 2015 in *Systematic Biology* **64**: 792-808.

INTRODUCTION

In this chapter, results from a phylogenetic analysis based on 12 single-copy nuclear loci, distributed on six of the seven barley chromosomes, and one chloroplast region are presented. Eight of the nuclear loci were newly explored for phylogenetics and are derived from rice genes (Ishikawa *et al.*, 2009), while four genes had previously been used (Petersen *et al.*, 2011; Brassac *et al.*, 2012). Initially we started out with a higher number of genes then removed loci that did not easily amplify, were not single copy, or comprised motifs that turned out to be problematic to sequence. All species were included with one to four individuals per species to additionally sample part of the intraspecific diversity. Taking advantage of next-generation sequencing (NGS) technologies including barcoding

(Meyer *et al.*, 2008b) and *in silico* cloning of multiplex-sequenced DNA fragments, we present here an extension of the method described in Griffin *et al.* (2011). In contrast to Griffin *et al.* (2011), our approach allows sequencing PCR amplicons of a large size [the longest sequence had a length of 3500 base pairs (bp)] not specifically designed to fit the read length of the employed NGS platform. Furthermore, we propose a method to disentangle reads from different genomes in a contig. Phylogenetic analyses were conducted for single loci, concatenated data from the nuclear loci, using multispecies coalescent to infer the species tree from multiple gene trees, and to evaluate incongruences between gene trees. These approaches allowed us to explore the allelic diversity simultaneously at numerous loci in multiple individuals in order to retrieve the phylogeny of the genus *Hordeum* and infer relationships between diploid and polyploid taxa and cytotypes. Also, the efficiency of multilocus phylogenetic methods in the presence of incomplete lineage sorting (ILS) and hybridization could be tested.

Materials and Methods

Plant Materials

We included 105 individuals representing all 33 species and most subspecies of the genus plus 1–2 individuals from each of five diploid Hordeeae species outside *Hordeum* plus *Bromus* as outgroups (table 3.1). Included individuals (Table S3.1) were chosen to reflect the intraspecific diversity observed in the *TOPO6* locus (Chapter 2). Herbarium vouchers of the analyzed materials were deposited in the herbaria of the IPK Gatersleben (GAT) or the Museum of Natural History, Buenos Aires (BA).

Table 3.1 Taxa included in the study

Taxon	Ploidy level (N)[1]	Haploid genome	Distribution area
Hordeum* subgenus *Hordeum			
Section *Hordeum*			
H. vulgare L.			
subsp. *spontaneum* (C.Koch.) Thell.	2x (2)	**H**	SW Asia
H. bulbosum L.	2x (1), 4x (3)	**H, HH**	Mediterranean to C Asia
Section *Trichostachys* Dum.			
H. murinum L.			
subsp. *glaucum* (Steud.) Tzvel.	2x (2)	**Xu**	Mediterranean to C Asia
subsp. *murinum*	4x (2)	**XuXu**	NW Europe to Caucasus
subsp. *leporinum* (Link) Arc.	4x (2), 6x (1)	**XuXu, XuXuXu**	Mediterranean to C Asia
***Hordeum* subgenus *Hordeastrum* (Doell) Rouy**			
Section *Marina* (Nevski) Jaaska			
H. gussoneanum Parl.	2x (2), 4x (2)	**Xa, XaXa**	Mediterranean to C Asia
H. marinum Huds.	2x (2)	**Xa**	Mediterranean
Section *Stenostachys* Nevski			
Series *Sibirica* Nevski			
H. bogdanii Will.	2x (3)	**I**	C Asia
H. brevisubulatum (Trin.) Link [2]	2x (5), 4x (4), 6x (3)	**I, II, III**	C Asia
H. roshevitzii Bowden	2x (2)	**I**	C Asia
Series *Critesion* (Raf.) Blattner			
H. californicum Covas & Stebb.	2x (3)	**I**	SW California
H. chilense Roem. & Schult.	2x (2)	**I**	Chile and W Argentina
H. comosum Presl	2x (3)	**I**	S Argentina
H. cordobense Bothmer et al.	2x (2)	**I**	C Argentina
H. erectifolium Bothmer et al.	2x (1)	**I**	C Argentina
H. euclaston Steud.	2x (3)	**I**	C Argentina, Uruguay
H. flexuosum Steud.	2x (1)	**I**	E+C Argentina
H. intercendens Nevski	2x (3)	**I**	SW California, NW Mexico
H. muticum Presl	2x (2)	**I**	C to N Andes
H. patagonicum (Haum.) Covas [2]	2x (3)	**I**	S Argentina
H. pubiflorum Hook.f. [2]	2x (2)	**I**	S Argentina
H. pusillum Nutt.	2x (2)	**I**	C+E USA

H. stenostachys Godr.	2x (2)	**I**	C Argentina
H. depressum (Scribn. & Sm.) Rydb.	4x (2)	**II**	W USA
Interserial allopolyploids of series *Critesion* and *Sibirica*			
H. brachyantherum Nevski	4x (2)	**II**	W North America, Kamchatka, Newfoundland
H. fuegianum Bothmer et al.	4x (2)	**II**	S Argentina, S Chile
H. guatemalense Bothmer et al.	4x (1)	**II**	Guatemala, S Mexico
H. jubatum L.	4x (2)	**II**	NE Asia, NW+W North America, C Argentina
H. tetraploidum Covas	4x (4)	**II**	C Argentina
H. arizonicum Covas	6x (3)	**III**	SW USA
H. lechleri (Steud.) Schenk	6x (3)	**III**	C+S Argentina
H. parodii Covas	6x (3)	**III**	C Argentina
H. procerum Nevski	6x (2)	**III**	S Argentina
Section *Nodosa* (Nevski) Blattner			
H. brachyantherum Nevski	6x (1)	**IIXa**	C California
H. capense Thunb.	4x (2)	**IXa**	S Africa
H. secalinum Schreb.	4x (2)	**IXa**	Mediterranean to W Europe
Outgroup species			
Eremopyrum triticeum (Gaertn.) Nevski	2x (1)	**FXe**	
Dasypyrum villosum (L.) Candargy	2x (2)	**V**	
Taeniatherum caput-medusae (L.) Nevski	2x (1)	**Ta**	
Triticum monococcum L.	2x (1)	**A^m**	
Secale vavilovii Grossh.	2x (1)	**R**	
Bromus tectorum	2x (1)		

[1] Number of individuals included per species or cytotype.

[2] Species with subspecies not further detailed here.

MOLECULAR METHODS

Genomic DNA was extracted from approximately 10 mg of silica gel-dried leaves with the DNeasy Plant Mini Kit (Qiagen) according to the protocol of the manufacturer. DNA quality and concentrations were checked on 1% agarose gels.

To arrive at PCR primers amplifying putative single-copy regions in *Hordeum*, the PCR-based Landmark Unique Gene (PLUG) system (Ishikawa *et al.*, 2007) was used. It consists of mapped

single-copy rice genes, which were used to detect conserved regions in wheat expressed sequence tag (EST) sequences, and map orthologous loci to the three genomes of bread wheat (Ishikawa *et al.*, 2009). The PLUG database (http://plug.dna.affrc.go.jp/) was screened for rice landmark loci with unique mapped orthologues in wheat and barley. PCR primers different from the suggested PLUG primers were designed in conserved exons with the intention to amplify regions of ~2000 bp and spanning one or several introns from all seven *Hordeum* chromosomes. The target regions thus resulted in larger amplicons in comparison to the PLUG markers, which have an average length of 951 bp (Ishikawa *et al.*, 2009). Initially 24 loci were PCR-screened with diploid representatives of the four genomes occurring in *Hordeum*. PCR products were checked on 1.4% agarose gels and numbers and sizes of the amplicons were determined. Finally, ten loci producing single fragments were chosen.

Including six already published loci, a total of two chloroplast and 14 nuclear loci were amplified by PCR (Table S3.2). PCR amplifications for all loci were performed in 30 µl reaction volume containing approximately 10-50 ng of genomic DNA, 1x Phusion HF Buffer, 0.2 mM each dNTP, 0.5 µM of each primer, and 1 U Phusion Hot Start DNA polymerase, a proof-reading enzyme (Finnzymes OY). The amplification program consisted of initial denaturation for 3 min at 98 °C, followed by 35 cycles of 30 sec at 98 °C, 1 min at 59 °C, 1 min at 72 °C, and a final extension of 10 min at 70 °C. Variations of this protocol with regards to the primer annealing temperatures for each locus are summarized in supplementary table S3.2. PCR amplification of one chloroplast region, *ndh*F, failed in more than a third of the individuals and thus *ndh*F was not included in the following steps.

Two sequencing approaches were used. (*i*) The amplicons of 96 individuals (all ploidy levels) were 454 sequenced. (*ii*) Amplicons of eight diploid individuals and one autotetraploid (*H. bulbosum* F2142) were cloned and eight clones per individual were sequenced following Brassac *et al.* (2012). The cloned amplicons were used as a control for the NGS approach and to provide reference sequences for assembling of 454 reads. All amplicons were purified using Nucleofast 96 Spin Plates (Macherey-Nagel) according to the protocol of the manufacturer and eluted in 20 µl of TE buffer.

For the NGS set, the concentrations of each amplicon were quantified using a Quant-iT Picogreen dsDNA (Molecular Probes) assay with a standard curve ranging from 1.25 pg/µL to 40 pg/µL. Fluorescence at 520 nm was measured on a Tecan Infinite M200 plate reader. The quantification of each amplicon was replicated three times and the average value was used to calculate the concentration of each sample.

Library preparation was performed as described in Meyer *et al.* (2008b) with the modifications defined below. For each individual, all amplicons were pooled in roughly equimolar ratio with a total amount of 1.5 µg of DNA in a reaction volume of 130 µl. Each pool was sheared from 700–2500 bp to a targeted size of 500–1000 bp via sonication in a microTUBE using a Covaris S220 (Covaris).

For barcoding, we used 7-nucleotides barcodes differentiated from each other by at least two substitutions on a maximum of 400 ng of DNA per individual to limit chimera formation. Tagging efficiency was verified by comparing the lengths of an untagged and a tagged fragment of a known size (200 bp) on a 2% agarose gel.

The tagged samples were quantified using a Quant-iT Picogreen dsDNA (Molecular Probes) assay as described earlier. After a sequencing trial of a diploid (*H. bogdanii* BCC2070) and a hexaploid individual (*H. arizonicum* BCC2054), we decided to standardize the DNA input according to the ploidy of the individuals in order to achieve sufficient coverage for the different alleles of the polyploids. Hence, twice the DNA amount of a diploid was used for a tetraploid and three times for a hexaploid individual.

All individually tagged sheared fragments were pooled, concentrated using MinElute PCR purification kit (Qiagen) and loaded on a SYBR Safe (Invitrogen) stained 1.5% agarose gel for size selection. The part of the gel lane comprising the fragments between 500 and 1000 bp was cut and DNA was extracted and purified using NucleoSpin Gel Clean-up (Macherey-Nagel) according to the manufacturer's protocol. DNA concentration of the pool was determined using a Qubit High Sensitivity (Invitrogen) assay. The quality of the pool was also checked using an Agilent High Sensitivity Chip assay. Dephosphorylation of the DNA pool was performed using the FastAP Thermosensitive Alkaline Phosphatase (Fermentas) and cut using the *Srf*I restriction enzyme, leaving the 5' phosphates free for the ligation of universal adapters for sequencing. To assess the efficiency of adapter ligation and quantify DNA concentration, a quantitative PCR was performed as described in (Meyer *et al.*, 2008a) using the emPCR primers.

The pool of DNA fragments from 96 individuals was sequenced in a single run on Roche's 454 sequencing platform using a PicoTiterPlate and the GS FLX Titanium XL+ chemistry in order to obtain long reads (up to 1000 bp) facilitating the identification of alleles and homoeologues. With an estimated length of 16 kbp amplicon lengths per haploid genome and two and three fold this amount for the polyploids and considering a typical output of 700 Mbp for a 454 run, we calculated that on average 270-fold coverage should be reached.

Quality Control and Haplotype Phasing

Barcode deconvolution, *i.e.* sorting fragments according to the single individuals, was performed with a custom script in Perl. The online tool TagCleaner (Schmieder *et al.*, 2010) was used to detect and trim adapters and barcodes. Geneious R6.1 (Biomatters Ltd.) was used for quality control and downstream analyses. All alignments were performed in Geneious using Mafft 6.814b (Katoh *et al.*, 2002; Katoh & Toh, 2008) with the E-INS-i algorithm (with default settings) and

manually checked. First, the sequences obtained from the diploid individuals via cloning were used to map the reads obtained from NGS with the High Sensitivity parameter (maximum mismatches at 40%) available in GENEIOUS. For loci with large insertions/deletions (indels) relative to the reference sequence, an iterative procedure was used to assemble the reads to one of the primer sequence until the second primer-binding site was retrieved. To check for potential extra copies, all the reads were remapped to the loci obtained via this initial method with High Sensitivity. Each assembly was then carefully inspected, especially regarding the coverage profile and the presence of single nucleotide polymorphisms (SNPs), trimmed and a consensus sequence was created.

Assemblies for the polyploids contained a mixture of reads belonging to the different homoeologues. To disentangle these copies, we used an *ad hoc* method consisting of *de novo* assembling of the reads mapped to a particular locus with Low Sensitivity (<10% mismatches allowed). This created a set of smaller assemblies with a more homogenous read content. The assemblies with the most reads (more than 20) were carefully checked for coverage and SNP presence, and their consensuses aligned together in order to identify the different copies. Homoeologous copies were identified as unique combinations of SNPs. Two copies were expected for the tetraploids and three for the hexaploids. Finally all the reads were mapped simultaneously, with Medium Sensitivity (20% mismatches), to confirm the different copies. To verify this method, we compared the outcome with results obtained using NEXTALLELE, a script for haplotype phasing for NGS data described in O'Neill *et al.* (2013), on a tetraploid individual (*H. depressum* BCC2047). Although the method was not designed for polyploids, the results were completely concordant due to the heterozygous-like behavior of the tetraploids. However, it appeared that NEXTALLELE was sensitive to the ratio between the reads corresponding to the two genomes resulting in ambiguities at the SNPs when this ratio departed from 50%. As expected, it could not deal with hexaploid individuals because of its limitation to two alleles. The single-copy status and the location of all nuclear loci were checked by blasting the sequences obtained for *H. vulgare* against its genome database (http://webblast.ipk-gatersleben.de/barley/viroblast.php; The International Barley Genome Sequencing Consortium, 2012).

PHYLOGENETIC ANALYSES

To infer the phylogeny of *Hordeum*, we adopted the following analysis approach consisting of nine steps. After aligning the sequences from all loci, (*i*) models of sequence evolution were determined for each locus. Gene trees were calculated for each locus with (*ii*) sequences derived from the diploid taxa by Bayesian phylogenetic inference (BI), and (*iii*) sequences from all diploid plus, consecutively, single polyploid individuals were clustered by neighbor-joining analysis to determine phylogenetic affiliation of the homoeologous gene copies found in polyploid taxa. Concatenated sequences from all loci (supermatrices) were used for BI of (*iv*) diploid and (*v*) diploid plus polyploid

taxa. (*vi*) A multispecies coalescent-based analysis was conducted to infer species trees from gene trees for the diploid individuals. (*vii*) To date nodes within the *Hordeum* phylogeny, a molecular clock approach was conducted together with the multispecies coalescent. (*viii*) A Bayesian concordance analysis was conducted on the diploid taxa only to estimate gene tree incongruences. Finally (*ix*) chloroplast *mat*K sequences were analyzed by BI to detect the maternal lineages in allopolyploids. These analysis steps are detailed below.

Model of sequence evolution

The best-fitting model of sequence evolution for each locus was identified with JMODELTEST 2.1.4 (Guindon & Gascuel, 2003; Darriba *et al.*, 2012) using default parameters. The best partitioning scheme for the concatenated nuclear loci was identified with PARTITIONFINDER 1.1.1 (Lanfear *et al.*, 2012) using the greedy algorithm. In both cases the Bayesian Information Criterion (BIC, Schwarz, 1978) was used for model choice because of its high accuracy (Darriba *et al.*, 2012) and its tendency to favor simpler models than the Akaike Information Criterion (Posada & Crandall, 2001). The preferred partitioning scheme and models of evolution are summarized in table S3.3.

Homoeologue identification

Homoeologous copies in polyploid individuals were identified from clustering of the sequences derived from polyploids with such derived from diploid taxa. For each locus and each polyploid species separately the copies from the different individuals were aligned with the sequences obtained from the diploid taxa. The neighbor-joining (NJ) method (Saitou & Nei, 1987) with HKY distances (Hasegawa *et al.*, 1985) was used to build a tree with 100 bootstrap replicates (not shown) in GENEIOUS using the Geneious Tree Builder option, as initial tests showed that NJ was able to safely discern and place copies. The closest neighbor of the polyploids' sequences were identified and coded (A, B or C) in the same way across the 12 different loci, and sequences were concatenated according to the designation as A, B, or C copies. When more copies than expected were retrieved, all sequences were included in the phylogenetic analysis per locus (diploids and polyploids) and a consensus sequence was created for the one clustering in the same clade, while the most distant copies (compared to the other individuals) were excluded from the supermatrices and the multispecies coalescent-based analysis, for example the *BLZ1* sequence Bre_2x_PI401390_B (Fig. S3.2). From the prior knowledge acquired during the analysis of *TOPO6* (Chapter 2), we included the sequences of tetraploid *H. jubatum*, which is assumed to be one of the parents of the American hexaploids, when analyzing the sequences from hexaploid species.

Bayesian phylogenetic inference

All BI analyses were performed in MRBAYES 3.2.2 (Ronquist *et al.*, 2012b) using the models and partitioning inferred by JMODELTEST and PARTITIONFINDER. Each analysis consisted of two independent analyses each running four sequentially heated chains (temperature set at 0.05) for 10 million generations and sampling a tree every 1000 generations. Convergence of the runs was assessed in TRACER 1.5 (Rambaut & Drummond, 2007) and the online application AWTY (Nylander *et al.*, 2008), which plots posterior probabilities of clade support values for both runs one against the other. The first 25% of sampled trees were discarded as burn-in and a consensus tree was computed in MRBAYES. For the supermatrix analyses with all loci concatenated, the data matrix was partitioned applying the respective model of sequence evolution for each locus/partition.

Sequences derived from nuclear loci were first analyzed for just the diploid individuals using the following approaches: (*i*) all sequences belonging to single loci were analyzed separately in MRBAYES to infer locus phylogenies (gene trees); (*ii*) the trees sampled from the MRBAYES analyses were analyzed with BUCKY 1.4.2 to obtain concordance factors for all clades. (*iii*) The sequences from all loci were concatenated (supermatrix) and analyzed in MRBAYES with partitions for all loci applying their respective models of sequence evolution. On a data set reduced to all *Hordeum* species plus *Dasypyrum villosum* and *Brachypodium distachyon*, (*iv*) the multilocus multispecies coalescent (MLMSC) was generated with *BEAST (Heled & Drummond, 2010) as part of the cross platform BEAST 1.8.1 (Drummond & Rambaut, 2007; Drummond *et al.*, 2012).

Incongruences between gene trees

Concordance among gene trees was estimated with BUCKY 1.4.2 (Ané *et al.*, 2007; Larget *et al.*, 2010) on the diploids plus *D. villosum*. Concordance factors (CF) can be interpreted as the proportion of loci supporting a specific topology (Ané *et al.*, 2007; Baum, 2007). BUCKY computes the CF for all potential topologies supported by the loci and provides a primary concordance tree "featuring relationships inferred to be true for a large proportion of genes" (Larget *et al.*, 2010). The trees sampled for each locus by the MrBayes analyses were summarized discarding 50% of each chain as burn-in. CF were estimated using the default *a priori* level of discordance ($\alpha = 1$) and 1,000,000 generations.

Dating of nodes and multispecies coalescent

Species trees and clade dating were estimated in *BEAST. The analysis was performed on a dataset corresponding to all nuclear loci except *NUC* (high amount of missing data) consisting of the diploid *Hordeum* individuals plus *D. villosum*, a species nested in a clade with *Triticum* (Escobar *et*

al., 2011), and *B. distachyon* as primary outgroup. Sequences for *B. distachyon* were retrieved by performing a BLASTN search of our loci on the *B. distachyon* genome database (http://www.brachypodium.org/). Priors on the root age (normal distribution; 44.4 Ma ± 3.53) and on the Hordeeae crown age (normal distribution; 15.32 Ma ± 0.34) were set as inferred by Marcussen *et al.* (2014). Monophyly was constrained for the *Hordeum* clade and for the Hordeeae clade. The analysis was run using the partitioning scheme and models of sequence evolution identified by PARTITIONFINDER, the Yule species tree prior, as well as the piecewise linear and constant root population model. Although rate constancy was systematically rejected for all loci based on the likelihood ratio test as proposed by Huelsenbeck & Rannala (1997), a strict clock model (uniform clock.rate; min 0, max 1.0) was used to ensure convergence and stability of the chains by using a simpler model. To speed up the analysis, we used a starting tree obtained by NJ of the supermatrix. Five independent analyses were computed for 100 million generations, sampling the states every 1000 generations. A standard burn-in of 20 million generations was discarded and all analyses were combined with LOGCOMBINER. Effective sample size (ESS) and convergence of the analyses were assessed using TRACER. A maximum clade credibility (MCC) tree was summarized with TREEANNOTATOR (part of the BEAST package).

Maximum parsimony

To see if the phylogeny obtained by BI is robust regarding different analysis algorithms, a parsimony analysis (MP) of the supermatrix of diploid taxa was conducted in PAUP* 4.0b10 (Swofford, 2002) using a heuristic search with 200 random sequence additions and TBR branch swapping, saving all shortest trees. Node support was evaluated by 500 bootstrap re-samples with the same settings but without random addition sequences.

Inference of parental progenitors of polyploids

To infer parental species of polyploids, the supermatrix of the concatenated sequences derived from each individual, including polyploids, was analyzed by BI as previously described. If homoeologues of polyploids fell within clades with diploid species these were interpreted as progenitor taxa. Clades containing polyploid-only sequences were interpreted as indication of extinct lineages (Blattner, 2004; Chapter 2). Species relationships were summarized in a schematic tree consisting of a backbone as inferred from the multispecies coalescent obtained from the diploids, and the polyploids were connected to their closest relative as revealed by the supermatrix analysis.

Inference of maternal progenitors

Chloroplast *mat*K sequences were analyzed separate from the nuclear loci, as chloroplast and nuclear phylogenies are partly incongruent in *Hordeum* (Petersen & Seberg, 2003; Jakob & Blattner, 2006; Petersen *et al.*, 2011). BI was used to determine the direction of crosses resulting in allopolyploid taxa, *i.e.* to infer the maternal progenitor, by comparing the position of chloroplast haplotypes derived from polyploids with the respective positions of parents in the phylogenetic tree derived from nuclear loci.

RESULTS

SEQUENCING AND SEQUENCE ASSEMBLY

454 sequencing of the DNA library combining barcoded PCR amplicons from one chloroplast and 12 nuclear loci (table 3.2) analyzed for 96 individuals resulted in 1,170,496 sequence reads, of which 999,492 (85%) were assembled to reference sequences, obtained from nine individuals by direct Sanger sequencing (*mat*K) or cloning and Sanger sequencing of PCR amplicons (nuclear loci) of all analyzed loci. The average read length was 472.1 (±23.8) bp. On average the number of sequence reads per individual (Table S3.1) was high but uneven. As expected, hexaploid individuals generated more reads than tetraploids and diploids (Fig. S3.1). The number of reads per locus and coverage was high for all loci but *TOPO6* and *NUC*. For these two loci only 66% and 77% of the individuals respectively received sufficient sequence reads mapping to the reference. For *TOPO6*, sequences from our previous study (Chapter 2) were used to complete the data set. Two loci, namely *TNAC1577* and *TNAC1781*, were sequenced but not further analyzed as they did not appear to be single copy in the diploid individuals and were difficult to assemble because of many homopolymer regions. Two *H. brevisubulatum* individuals (PI229753 and GRA2230/97) were excluded from the analyses because of generally bad-quality sequences.

For the 13 loci included in the study and all of the individuals, the alignment lengths per locus varied between 759 bp and 5074 bp with 210 to 1,333 variable sites (average 466) and 100 to 993 parsimony-informative characters (average 316) per locus (Table S3.3). Some species showed >1000 bp insertions in *TNAC1142* (in *Bromus tectorum* and *Triticum monococcum*) and *TNAC1463* (in *H. murinum* and *H. gussoneanum*). A BLASTN (Altschul *et al.*, 1990) search of these insertions returned only a highly similar match for the *H. murinum* element identified as a partial non-long terminal repeat retrotransposon. Concatenation of the 12 nuclear loci in a supermatrix resulted in an alignment of 24,996 bp length, while chloroplast *mat*K had a length of 2,625 bp.

Table 3.2 Loci initially explored in the study and location on barley chromosomes

Locus name	RAP2 description	Chromosome[1]	No. of reads	Coverage mean ± 1 SD
TNAC1035	Kinesin, motor region domain containing protein	1H	62,836	518 ± 340.0
TNAC1142	Similar to COP9 signalosome complex subunit 5b	2H (?)	51,409	305 ± 214.4
XYL	Xylose isomerase	2HS	87,926	547 ± 329.8
TNAC1364	Ubiquitin domain containing protein	3HL	53,286	334 ± 243.1
NUC	Nucellin	4HL	31,916	260 ± 275.4
TNAC1403	Similar to SAC domain protein 1 (FIG4-like protein AtFIG4)	4HL	151,385	490 ± 334.9
TNAC1463	Proteasome subunit beta type 2	4HS	97,610	380 ± 389.5
BLZ1	Barley leucin zipper	5HL	56,064	376 ± 236.8
TNAC1610	Peptidase S16, ATP-dependent protease La family protein	5HL	71,133	374 ± 342.0
TOPO6	Topoisomerase VI subunit B	5HL	27,416	226 ± 159.1
TNAC1577[2]	Conserved hypothetical protein	5HL (?)	Na	Na
TNAC1497	Similar to Nucleoside diphosphate kinase II, chloroplast precursor	5HS	71,439	502 ± 327.2
TNAC1740	Heat shock protein Hsp70 family protein	6HL	49,843	297 ± 241.7
TNAC1781[2]	Beta 5 subunit of 20S proteasome	7HS (?)	Na	Na
*mat*K	Maturase K	cp-LSC	179,382	426 ± 284.2
*ndh*F[2]	Subunit 6 of NADH-dehydrogenase	cp-SSC	Na	Na

[1] Chromosome locations were checked by blasting sequences obtained on the barley genome, no significant result was obtained from *TNAC1142*, locations for this locus and the other loci missing were inferred from synteny with rice and wheat.

[2] *ndh*F was excluded because of large amount of missing data and *TNAC1577* and *TNAC1781* were not single-copy loci and difficult to sequence (large homopolymer regions)

NGS sequences of the *TOPO6* locus, that were included as a control to compare the 454 results with those obtained by traditional Sanger sequencing, were longer than those obtained previously (Jakob & Blattner, 2010; Chapter 2). This was due to high confidence in the bases close to PCR primer binding sites provided by NGS compared to traditional Sanger sequencing, where bases adjacent to sequencing primers were often not reliably recovered. Overall, the similarity between sequences obtained by 454 and cloning plus Sanger sequencing was high (99.81% ± 0.67 pairwise identity).

For most of the nuclear loci, the expected two homoeologous copies for a tetraploid individual and three copies for a hexaploid individual were retrieved. Notable exceptions with partly only one of the copies observed were the tetraploids *H. capense* and *H. secalinum* (at *NUC* and *TNAC1142*) and two copies for hexaploid *H. arizonicum* (at *BLZ1*). All tetraploid and hexaploid *H. murinum* cytotypes possessed only one of the two *XYL* copies. The high sequence coverage obtained allowed us to recover very rare copies (*e.g.*, for *TNAC1610* in *H. capense* BCC2062) occurring with only about 3% of reads mapping to a locus. All sequences obtained in this study were submitted to the NCBI nucleotide database (accession numbers KM039139–KM040760).

Phylogenetic Relationships

To arrive at a species-level phylogeny of *Hordeum*, we initially analyzed sequences from all loci separately for diploid taxa and then repeated the analysis including the sequences from the polyploid taxa. The corresponding single-locus trees for all individuals are available as supplementary information (Fig. S3.2-S3.13). The gene trees varied substantially between loci. For example, *Hordeum* appeared non-monophyletic in five cases (*TNAC1142*, *TNAC1364*, *TNAC1403*, *TNAC1463* and *TNAC1497*), and the **I**-genome taxa (section *Stenostachys*) were split in two clades in six cases (*TNAC1035*, *TNAC1142*, *TNAC1403*, *TNAC1463*, *TNAC1497* and *XYL*).

Diploid taxa were analyzed separately to overcome inconsistencies due to different evolutionary histories of the analyzed loci. (*i*) A BCA allowed us to evaluate those inconsistencies and to estimate the proportion of loci supporting different topology hypotheses. *(ii)* All loci were concatenated to obtain a supermatrix that was subjected to BI in MrBayes 3.2.2 (Ronquist *et al.*, 2012b) and MP in PAUP* 4.0b10 (Swofford, 2002). Moreover, we (*iii*) checked for topological stability if loci number was further increased with a BI of a supermatrix of all nuclear loci of our study plus the nuclear loci published by Petersen *et al*, (2011) for a dataset reduced to one individual per species. Finally (*iv*) the multilocus multispecies coalescent of gene trees (MLMSC) was computed in *BEAST (Heled & Drummond, 2010) to arrive at a species tree based on the individual gene trees.

The BUCKy concordance factors obtained from the BCA computed on the single locus trees sampled by BI are relatively low confirming a general discrepancy between loci. All species, with the

exception of one, appeared monophyletic usually with a higher CF than groups of species. *Hordeum patagonicum*, a diverse species, was paraphyletic, including its sister species *H. pubiflorum* (CF of 0.06) in the primary concordance tree (Fig. S3.14), or monophyletic (CF of 0.02) in an alternative topology. The BI of the supermatrix (Fig. 3.1) produced a highly congruent tree topology with the primary concordance tree (Fig. S3.14). Within *Hordeum* all clades were highly supported except for *H. californicum*, which was found paraphyletic with the series *Critesion* species grouping within the *H. californicum* grade, and *H. patagonicum*, found paraphyletic with *H. pubiflorum* nested within the *H. patagonicum* grade. Within the South American clade, characterized by short internal branches, the nucleotide diversity was too low to resolve all sister relationships. The exclusion of *NUC*, the locus with the most missing data, from the supermatrix had no effect on the topology of the resulting tree. The consensus tree of 66 MP trees derived from the supermatrix (Fig. S3.15) was very similar to the BI tree. Exceptions included the monophyly of *H. californicum*, sister to all other species of series *Critesion*, and the paraphyly of *H. brevisubulatum*, with one individual (PI440419) clustering with low support at the basis of the two other Asian species. The inclusion of seven additional nuclear loci (Petersen *et al.*, 2011), although for a dataset with only one individual per diploid species, resulted in a highly supported topology (lowest pp 0.98) and confirmed generally the topology obtained with the supermatrix (Fig. S3.16). *Hordeum pusillum* is the only species differentially affected, clustering at the basis of the two sister species *H. intercedens* and *H. euclaston* in the 19 loci analysis.

The MLMSC species tree (Fig. 3.2) resulted in a slightly different topology providing a better resolution for the closely related American species. The main disagreement concerned the three southern Patagonian species (*H. comosum*, *H. patagonicum* and *H. pubiflorum*) found to be monophyletic and *H. chilense*/*H. flexuosum* and *H. cordobense/H. muticum* as successive sister clades to the Patagonian clade. A second disagreement occurred with the Asian clade and the relationships within this clade where MLMSC provided only a low support (0.87 and 0.66 pp, respectively) with *H. roshevitzii* outside. However, the differences are minor and do not influence the recognition of the major clades within *Hordeum*. In all analyses the recently proposed infrageneric taxonomic groups within *Hordeum* (Blattner, 2009) were found to be monophyletic, for the first time with high support values (Fig. 3.1).

Incongruences Among Loci and Between Methods

Despite the general discrepancy between loci, the multilocus methods are congruent to the exception of the two clades harboring Asian and the South American species. In the case of the former, the four methods provided four different topologies (table 3). BUCKY, with 9.6% of the loci supporting this topology, favored *H. bogdanii* as sister to the two other species. It is almost equivalent to the BI topology (9.3%) while the MLMSC topology is supported by only 6.2% of the loci. The MP

topology with *H. brevisubulatum* paraphyletic is supported by 6.9% of the loci. For the South American species, and especially the Patagonian clade, only the MLMSC recovered its monophyly while BUCKY retrieved it only for 0.3% of the loci (table 3).

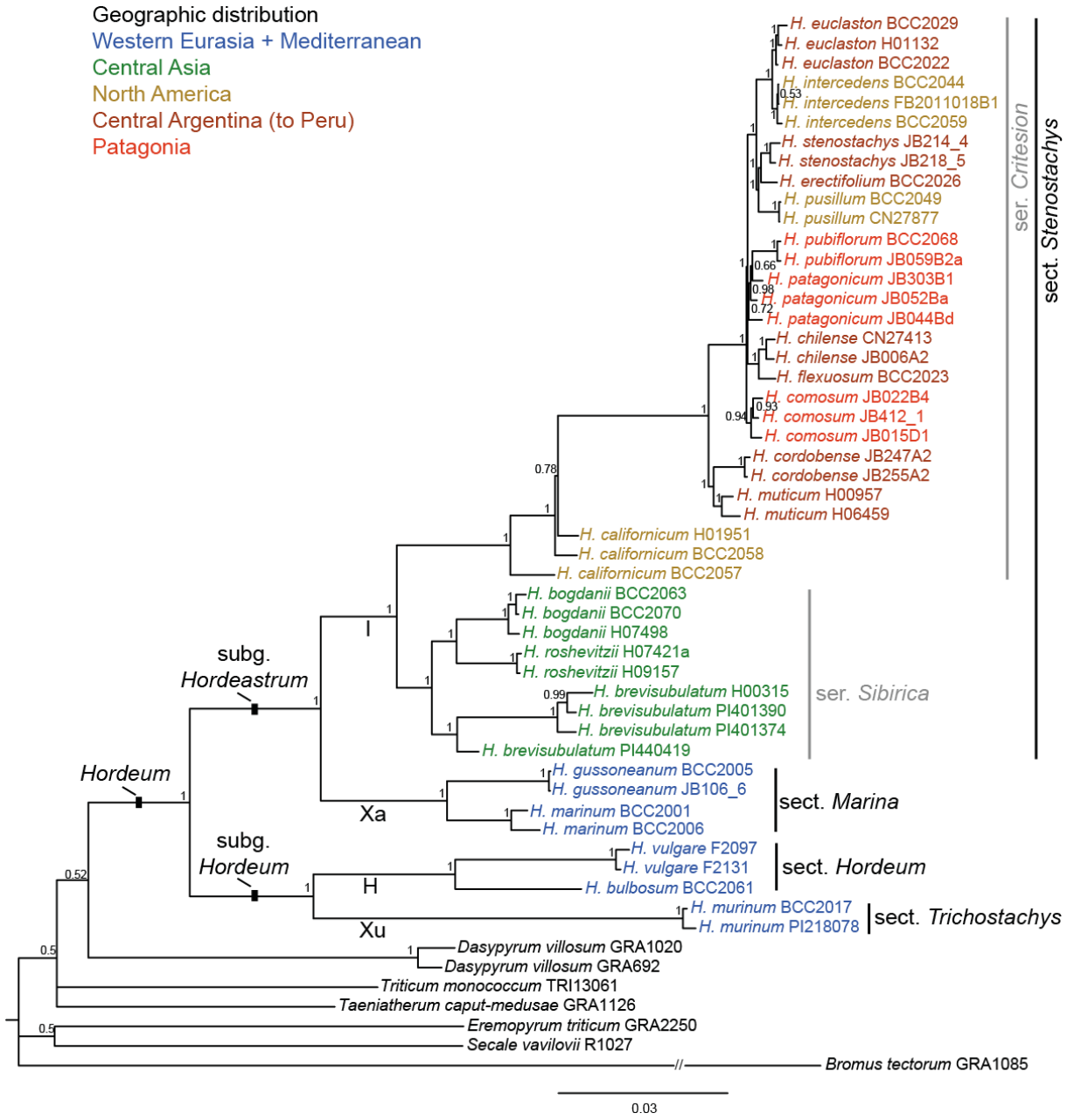


Figure 3.1 Phylogenetic tree derived from the concatenated supermatrix consisting of 12 single-copy nuclear loci of the diploid *Hordeum* taxa and six outgroup species calculated with Bayesian inference. Posterior probability values of the clades are indicated along the branches. Infrageneric treatment of the genus and genome denominations (bold letters along branches) in *Hordeum* follow Blattner (2009).

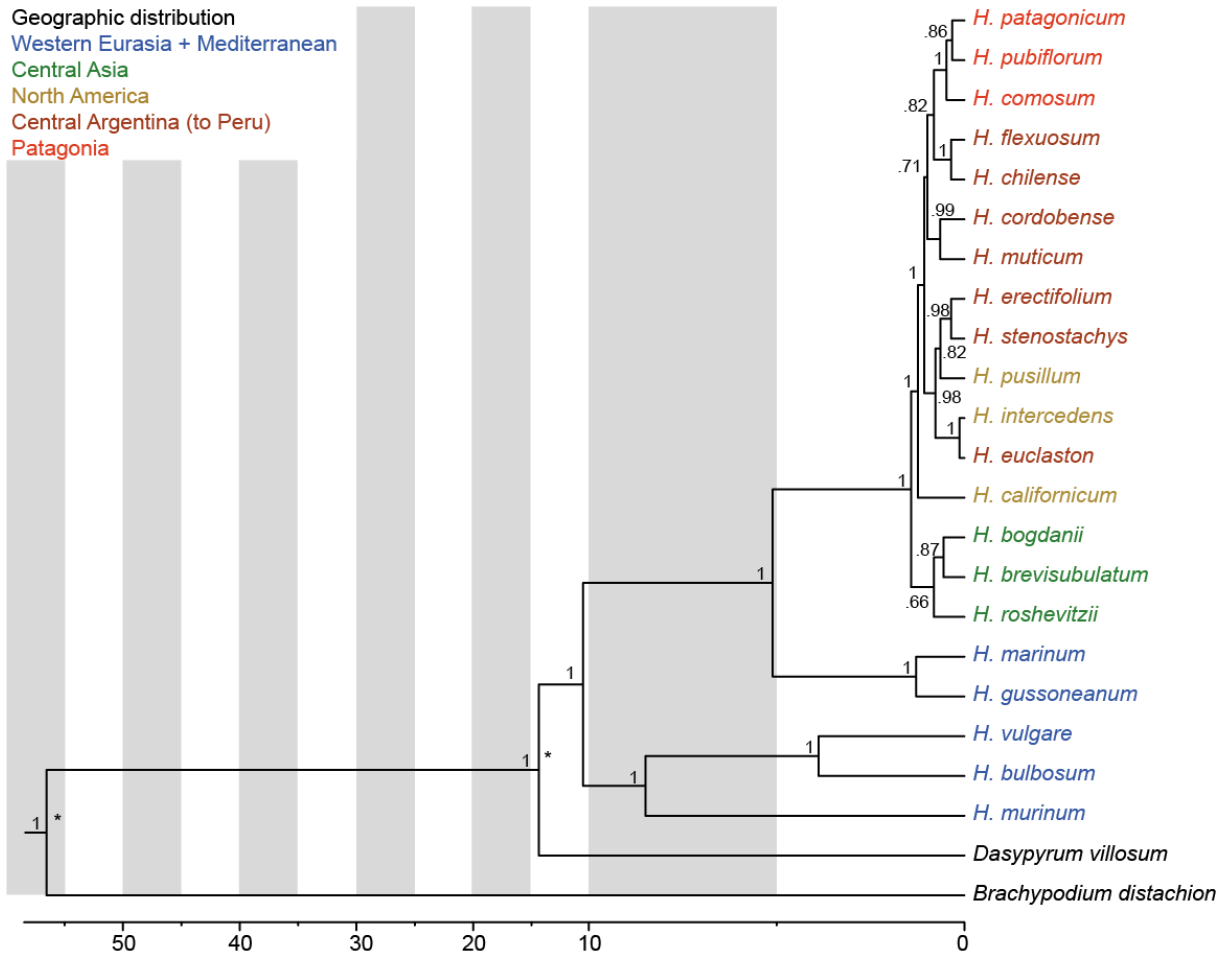


Figure 3.2 Calibrated multilocus multispecies coalescent of the diploid *Hordeum* species as calculated with *BEAST from the nuclear loci, excluding *NUC*. Posterior probability values of the clades are indicated along the branches. Ages (in million years ago, Ma) inferred from the crown clade age of *B. distachyon* + Hordeeae at 44.4 Ma and from the divergence between *Hordeum* and *Triticum* lineages at 15.32 Ma (Marcussen *et al*., 2014); calibration points are depicted by asterisks, divergence dates are reported in Table 3.3

AGES OF CLADES

Divergence times (table 3.3) estimated with a strict clock model and the secondary calibration on the most recent common ancestor between *B. distachyon* and Hordeeae and on the Hordeeae crown clade in *BEAST resulted in ages with relatively narrow 95% highest probability density intervals (HPD). A test with a random local clock (Drummond & Suchard, 2010) resulted in very similar ages but required 500 million generations per chain to obtain good mixing for all parameters (not shown).

Table 3.3 Ages of the clades in the MLMSC analysis and concordance factors (CF)

Clade (MLMSC/BI/MP)	Ages Ma (95% HPD)	Sample-wide CF (95% CI)
H. erectifolium/stenostachys (0.98/1/97)	0.35 (0.13,0.57)	0.251 (0.091,0.455)
H. erectifolium/pusillum (0.82/1/95)	0.64 (0.45,0.83)	0.184 (0.000,0.364)
H. euclaston/intercedens (1/1/100)	0.14 (0.06,0.23)	0.636 (0.455,0.727)
H. euclaston/pusillum[a]		0.090 (0.000,0.273)
H. erectifolium/intercedens (1/1/100)	0.78 (0.60,0.94)	0.476 (0.273,0.636)
H. pubiflorum/patagonicum (0.86/0.72/-)	0.33 (0.17,0.49)	0.056 (0.000,0.182)
H. pubiflorum/comosum (1/-/-)	0.48 (0.31,0.66)	0.003
H. flexuosum/chilense (1/1/100)	0.36 (0.14,0.57)	0.648 (0.545,0.727)
H. pubiflorum/chilense (0.82/-/-)	0.82 (0.62,1.03)	0.013
H. pubiflorum/cordobense (0.71/-/-)	0.99 (0.82,1.18)	0.019
H. muticum/cordobense (0.99/1/99)	0.65 (0.38,0.91)	0.278 (0.182,0.364)
H. pubiflorum/erectifolium (1/1/100)	1.08 (0.91,1.27)	0.147 (0.091,0.182)
New World I-clade taxa (0.98/1/94)	1.25 (1.08,1.45)	0.116 (0.091,0.182)
H. bogdanii/brevisubulatum (0.87/-/-)	0.57 (0.31,0.86)	0.062 (0.000,0.091)
H. roshevitzii/brevisubulatum[b]		0.096 (0.000,0.182)
H. roshevitzii/PI440419[c] (-/-/67)		0.069 (0.000,0.091)
H. bogdanii/roshevitzii[d] (-/1/100)		0.093 (0.000,0.182)
Old World I-clade taxa (0.66/1/94)	0.85 (0.46,1.28)	0.158 (0.091,0.182)
I clade (1/1/100)	1.44 (1.21,1.69)	0.187 (0.091,0.364)
Xa clade (1/1/100	1.30 (0.68,1.97)	0.813 (0.636,0.909)
I plus Xa clade (1/1/100)	5.11 (4.19,6.06)	0.237 (0.091,0.364)
H clade (1/1/100)	3.86 (2.65,5.05)	0.595 (0.455,0.727)
H plus Xu clade (1/1/100)	8.49 (7.23,9.83)	0.375 (0.273,0.545)
Hordeum (1/1/100)	10.18 (8.82,11.60)	0.45
Hordeeae[e]	14.81 (14.13,15.48)	
B. distachyon/Hordeeae[e]	56.54 (50.96,62.06)	

Notes: MLMSC, multilocus-multispecies coalescence; BI, Bayesian inference; MP, maximum parsimony; HPD, highest posterior density; CF, concordance factor; CI, credibility interval.

[a]Node present only in the 19 loci supermatrix (Fig. S3.16), [b]Node present only in the BUCKY analysis (Fig. S3.14), [c]Node present in the 12 loci supermatrix MP analysis (Fig. S3.15), [d]Node present in the 12 loci supermatrix analyses (Fig.3.1), [e]No 95% CI available and genome-wide values.

The posterior distribution for the age of the root appeared to be older than our prior (56.5 Ma, 95% HPD = 51–62). The Hordeeae crown clade age (14.8, 95% HPD = 14.1–15.5) fitted to our prior. The most recent common ancestor of *Hordeum* occurred 10.2 Ma (95% HPD = 8.8–11.6), while the split between the **H** and **Xu-**genome groups of subg. *Hordeum* happened 8.5 Ma (95% HPD = 7.2–9.8), and the divergence of **Xa** and **I-**genome lineages within subg. *Hordeastrum* started 5.1 Ma (95% HPD = 4.2–6.1). The colonization of the Americas occurred around 1.3 Ma (95% HPD = 1.1–1.5). The analysis revealed nearly contemporaneous speciation events for three couples of species in southern South America: *H. pubiflorum* and *H. patagonicum* (0.33 Ma, 95% HPD = 0.17–0.49), *H. erectifolium* and *H. stenostachys* (0.35 Ma, 95% HPD = 0.13–0.57), and *H. chilense* and *H. flexuosum* (0.36 Ma, 95% HPD = 0.14–0.57) indicating a possible common climatic and/or geographic reason.

Inference of Parental Progenitors of Polyploids

To identify the progenitors of the polyploid species a supermatrix including the sequences derived from diploid and polyploid taxa was phylogenetically analyzed with BI (Fig. S3.17). The positions of the different homoeologues of polyploids in relation to their closest relatives derived from diploids were used to infer the lineages contributing to polyploids. If sequences from a polyploid lineage grouped within different diploids, this was interpreted as an indication for allopolyploidy, while autopolyploidy was inferred if all sequences of a polyploid were in a clade with a single diploid species. Clades consisting solely of polyploid-derived sequences were interpreted as indication of extinct progenitor lineages (Blattner, 2004; Jakob & Blattner, 2010; Chapter 2). The results of this analysis were summarized in a scheme where polyploids were integrated in the modified diploid species tree (Fig. 3.3). The MLMSC topology was modified to take into account the incongruences between the different methods and to integrate the inferred extinct lineages. The polyploid relationships could mostly be identified with confidence. The wide genetic variety found in some polyploid species could partly also indicate multiple origins of such taxa. *Hordeum parodii*, a hexaploid species, as well as *H. tetraploidum*, one of its potential tetraploid progenitors together with *H. fuegianum*, appeared to be polyphyletic involving the two closely related diploid species *H. chilense* and *H. flexuosum*. The partially autopolyploid taxon *H. brevisubulatum* (Chapter 2) was treated differently. The high diversity of the copies recovered for the different individuals and the difficulty to assign the parental species/individuals across loci hampered us to create phased haplotypes. Only one tetraploid (PI401387) appeared autopolyploid with gene copies clustering essentially with the species' diploid cytotypes while sequences of other individuals clustered with species within the Asian *Hordeum* clade (*H. roshevitzii* and *H. bogdanii*). The position of many sequences (for example Bre_PI401376_C at *BLZ1*, Fig. S3.2) retrieved for the polyploid taxa indicates ongoing intergenomic recombination as already suggested by the high proportion of chimerical

sequences in cloned *TOPO6* sequences (Chapter 2). Sequencing and analysis of the individual H00312 revealed it to be probably mislabeled and it was then excluded from our conclusions (see Discussion).

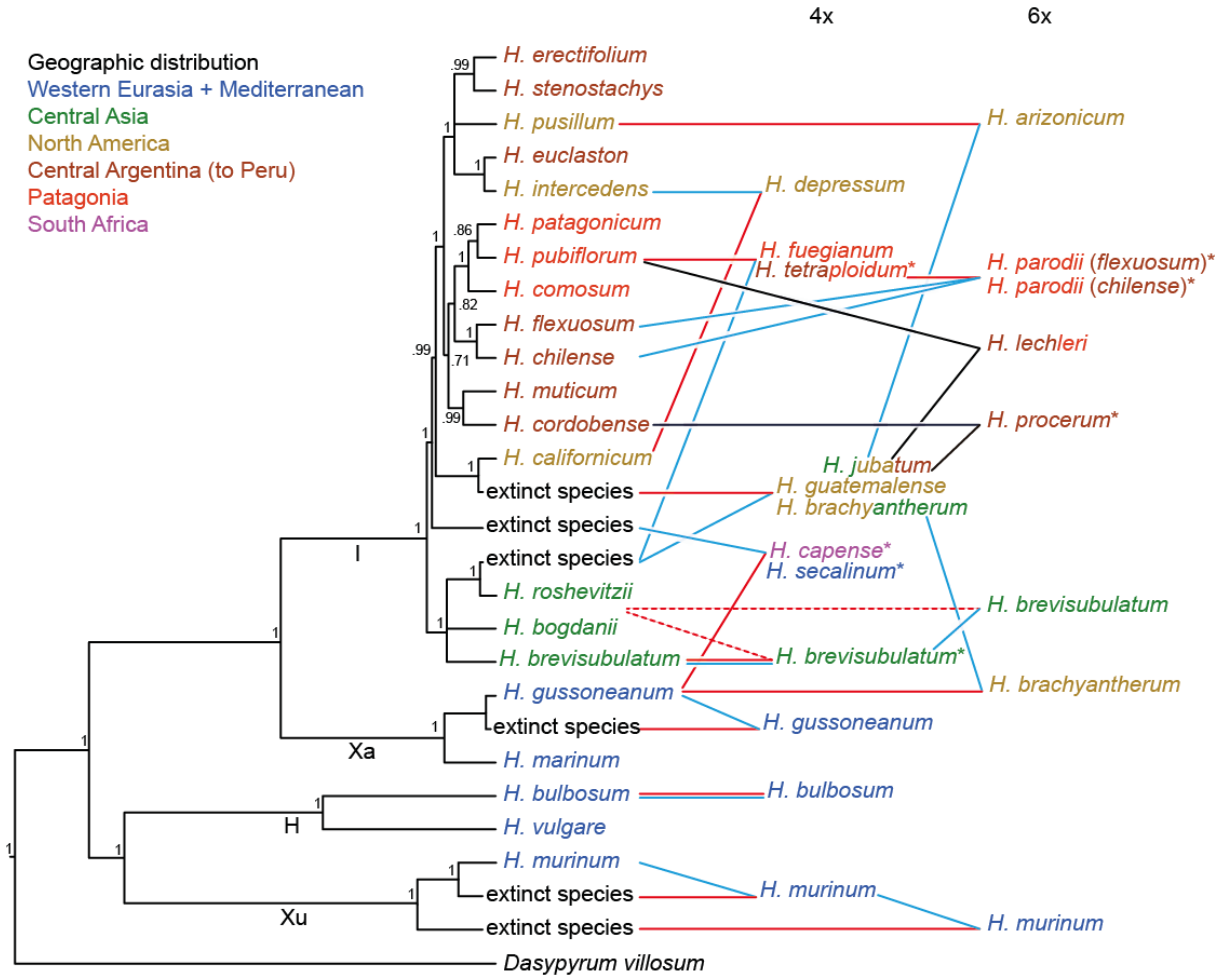


Figure 3.3 Scheme summarizing phylogenetic relationships of species and cytotypes in the genus *Hordeum* based on 12 single copy nuclear loci and one chloroplast region. Diploid taxa were drawn directly at the tree obtained from MLMSC analysis (Fig. 3.2) modified to reflect differences with BI analysis based on 12 (Fig. 3.1) and 19 loci (Fig. S3.15) and position of extinct taxa inferred from the BI tree including diploid and polyploids (Fig. S3.16). Tetra- and hexaploids were connected to their inferred progenitors. Double lines indicate autopolyploid origin, dashed lines mark uncertainties and colors refer to the direction of crosses resulting in allopolyploid taxa (blue for paternal, red for maternal parent, black for both observed) based on the chloroplast data (Fig. S3.17). Asterisks indicate incongruent relationships of polyploids recovered by chloroplast sequences (Jakob & Blattner, 2006) compared to nuclear sequences that probably indicate polyphyletic origins.

BI of the chloroplast *mat*K sequences of all individuals (Fig. S3.18) resulted in a poorly resolved phylogeny due to the small number of informative characters, and in an incongruent topology compared to the nuclear loci (see Jakob & Blattner, 2006 for a larger sampling of chloroplast haplotypes). Despite the numerous individuals grouping in the basal polytomy and in single-species clades, chloroplast data identify *H. roshevitzii* as paternal progenitor of all the polyploids it is related to. The polyploid cytotypes of *H. murinum* carry the chloroplast type of their inferred extinct progenitors. The two sister species *H. secalinum* and *H. capense* possess a Eurasian chloroplast type. The first one is closer to *H. brevisubulatum* and the latter to *H. marinum*, while their paternal progenitor *H. gussoneanum* possesses a very different type (Petersen & Seberg, 2003; Jakob & Blattner, 2006). The American hexaploid *H. arizonicum* falls in a clade with *H. pusillum*, its diploid progenitor. At least three polyploid species (*H. lechleri*, *H. procerum* and *H. tetraploidum*) were found polyphyletic regarding their chloroplast types, with haplotypes occurring in different clades.

DISCUSSION

COMBINING PCR AMPLIFICATION WITH SECOND-GENERATION SEQUENCING

In this study we took advantage of the long reads of 454 sequencing together with parallel barcoding to sequence one chloroplast and 12 nuclear single-copy loci, distributed among nearly all barley chromosomes, in 96 individuals representing all species and cytotypes of the genus *Hordeum*. In addition we cloned and sequenced these loci in eight diploids and one autotetraploid to serve as references for mapping the reads of the 454-sequenced individuals. To evaluate the capacity of our assembling strategy to recover all true haplotypes, a *de novo* approach was applied on some individuals. Although successful, this method can be very time consuming and fastidious due to the high number of contigs produced and thus was not further used.

The cloned sequences together with previously published sequences (Petersen *et al.*, 2011; Chapter 2) for some of the loci were used as a control to compare the results of traditional Sanger sequencing and NGS. Apart from the lengths of the obtained sequences, due to the high-quality base calls close to the PCR priming sites with NGS, nearly no differences occurred between both sequencing approaches. High-coverage NGS allowed us to recover rare copies that would have otherwise required sequencing of at least an order of magnitude more clones than traditionally used in phylogenetic studies. The amount of chimerical sequences was lower compared to our previous analysis (Chapter 2) due to the exclusive use of a proof-reading DNA polymerase. The obtained unambiguous sequences allowed us to safely exclude one accession (H00312) because of its peculiar placement. First described as an Iranian *H. bogdanii* for which the Giemsa-C banding pattern was analyzed (Linde-Laursen *et al.*, 1980), it was then assumed to be a tetraploid *H. brevisubulatum*

because of its genome size (unpublished result). This analysis revealed its close relationship with sequence copies of the hexaploid *H. brachyantherum*, a cytotype only known from a very small population in California (Komatsuda *et al.*, 2009). Based on our results we now believe that this individual does not represent a true *Hordeum* taxon but was instead either mislabeled or hybridized during *ex situ* propagation (Jakob *et al.*, 2014) and is therefore removed from our conclusions. Uncertainty regarding the correct resolution of mononucleotide repeats, an error specific for pyrosequencers like the 454-sequencing platform, was minor. This was likely due to the very high sequence coverage (270 fold on average) that we aimed for.

Due to the lack of dedicated bioinformatic tools that separate homoeologues (see O'Neill *et al.*, 2013; Ranwez *et al.*, 2013) we used a method based on a combination of reference-based mapping and *de novo* assembly to disentangle the sequence reads from polyploid individuals. We were able to successfully reconstruct phased haplotypes for all the individuals analyzed by NGS, resulting in the most comprehensive phylogenetic analysis of the genus *Hordeum* to date. For this approach, the long reads of the 454 platform in combination with a relatively high coverage were favorable. The recent progress in NGS, with increasing read lengths and paired-end sequencing of libraries, means that sequencing of long PCR amplicons and reconstructing phased haplotypes is no longer restricted to the long contiguous reads obtained by 454 sequencing, and therefore will become much cheaper.

Allopolyploid species possess merged genomes and thus require special care when sequencing nuclear loci. This is traditionally done by cloning of PCR amplicons and sequencing of a certain number of clones. Moreover, obtaining sufficient resolution in recently diverged species requires many characters, and longer loci might be favorable for reconstructing species trees from gene trees. The required locus length, together with sequencing of many clones per locus, makes molecular phylogenies of even a medium sized genus with many polyploid species time consuming and expensive. Here we combined traditional methods, *i.e.* amplifying long loci not necessarily designed to fit the NGS platform's size optimum, with the capacity of these new sequencing techniques to handle a mixture of sheared and barcoded PCR amplicons, extending the method proposed by Griffin *et al.* (2011). Moreover, in certain cases the high throughput of NGS allowed us to overcome what might be the result of PCR drift (Wagner *et al.*, 1994), where an allelic variant is randomly favored during the PCR, and thus to explore a potentially more complete set of allelic diversity in comparison to a cloning approach.

A major benefit of phylogenetic studies in grasses is the availability of genomic information for *Brachypodium*, rice, sorghum, barley and many other species. This makes design of PCR primers for a set of nuclear single-copy loci relatively easy. We took advantage of the rice PCR-based Landmark Unique Gene (PLUG) system (Ishikawa *et al.*, 2007) that lists a high number of potentially single-copy loci and their chromosomal position in the rice genome. With rapidly increasing genomic

information for many plant taxa (Van Bel *et al.*, 2012), it is expected that similar analyses are already or will soon be possible in taxonomic groups throughout the angiosperms.

PHYLOGENY OF HORDEUM

The phylogenetic relationships obtained from the set of 12 nuclear single-copy loci (Fig. 3.1) is in accord with the recently proposed new infrageneric treatment of *Hordeum* (Blattner, 2009). Compared to previous studies of the genus a better resolution was generally obtained. The main differences appear in the grouping of the recently diverged American species. Thus, for the first time we were able to show the monophyly of the Patagonian diploid species *H. comosum*, *H. patagonicum* and *H. pubiflorum* using nuclear data, although this relationship was already deduced from the distribution pattern of shared chloroplast haplotypes among these species (Jakob *et al.*, 2009). However, long-term large population sizes in this clade result in incomplete lineage sorting (ILS) that impedes phylogenetic analysis (Jakob & Blattner, 2006). In addition, the three Central Asian diploids could be confirmed to form a clade with high support values in the supermatrix analyses, although low support was obtained in MLMSC.

The increased resolution resulting from this study allowed us to both confirm and more precisely define the progenitor-derivative relationships previously identified (Blattner, 2004; Chapter 2). The importance of the Asian diploid *H. roshevitzii* in the evolution of the American polyploids is emphasized here, and the *mat*K analysis showed that this species was never a maternal progenitor. Jakob & Blattner (2006) did not discuss this particular result when analyzing chloroplast relationships in *Hordeum*, but the two studies are completely concordant. An extinct close relative of *H. californicum* was confirmed as a second key lineage (Chapter 2) in the evolution of most American polyploids, and chloroplast sequences indicated that this species functioned as the maternal progenitor of polyploids. Additionally, the extinct lineages contributing to polyploids within the *H. murinum* taxon complex (Jakob & Blattner, 2010; Tanno *et al.*, 2010) and tetraploid *H. gussoneanum* (Chapter 2; Carmona *et al.*, 2013) could be safely confirmed, as well as the probably extinct Central Asian paternal parent of tetraploid *H. capense* and *H. secalinum* (Chapter 2). We could identify *H. pubiflorum* as the second progenitor of the two tetraploid sister species *H. fuegianum* and *H. tetraploidum*. However, the chloroplast sequences of *H. fuegianum* clustered with those of *H. pubiflorum*, while the sequences derived from *H. tetraploidum* were placed in the basal polytomy. This suggests that *H. tetraploidum* either evolved independently or that gene flow and chloroplast capture might still be ongoing for these species (Jakob & Blattner, 2006). The recurrent formation of polyploids has been extensively demonstrated (Soltis *et al.*, 1993; Soltis & Soltis, 1999) partially explaining the high diversity and complexity of polyploid genomes. Based on our analysis, the hexaploid species *H. parodii* seems to be a good example of recurrent formation involving two diploid

species, *H. chilense* and *H. flexuosum*, and the tetraploids *H. fuegianum* or *H. tetraploidum*. The *H. brevisubulatum* species complex (di-, tetra- and hexaploid cytotypes), an obligate out-crossing taxon distributed between Iran and northeastern Siberia, shows signs of recurrent auto- and allopolyploidization involving the entire Asian genepool (including *H. roshevitzii* and *H. bogdanii*) as well as *H. californicum* or its Asian progenitor. To understand the evolution of this very diverse lineage, extensive sampling and extended population-based studies would be necessary.

Our dated phylogeny provided younger and more accurate age estimations within *Hordeum* than previously inferred (Blattner, 2006). However, within the American species group, such ages are surprisingly low taking into account their clear ecological diversification (Jakob *et al.*, 2010). The entire American clade seems to be only slightly older than one million years (1.3 Ma) for the split between *H. californicum* and the other American species. This fast speciating clade (Jakob & Blattner, 2006) seems to have been shaped at least in part by the repeated glaciations of the Pleistocene (2.6–0.01 Ma). We obtained no fewer than three nearly contemporaneous speciation events (*H. pubiflorum* and *H. patagonicum*, *H. erectifolium* and *H. stenostachys*, and *H. chilense* and *H. flexuosum*) in South America coinciding with the glacial period 0.35–0.40 Ma (Petit *et al.*, 1999) which seems to have left no major geological evidence in Patagonia (Rabassa *et al.*, 2011). The effects of ice ages on speciation and divergence are complex processes (Hewitt, 1996) but have been shown to be of major importance in shaping current biodiversity in general (Comes & Kadereit, 1998) and specifically in *Hordeum* (Jakob *et al.*, 2007, 2009, 2010).

Incongruences Among Loci and Between Methods

The use of different methods of multilocus phylogeny inference allowed us to test the performance of those methods in the presence of ILS and potential hybridization. Despite the general discrepancy between loci, only two clades appeared incongruent between methods. The Patagonian clade, characteristic of ILS with young species (0.31–0.66 Ma) and large population sizes, seems to be better dealt with by the coalescent approach. The pattern observed in the Asian clade, for which none of the four methods converged for the topology within the clade, is most probably due to far-reaching ILS and/or hybridization, especially involving the *H. brevisubulatum* species complex. The concatenation approach is known to potentially lead to overconfident support for incorrect species trees compared to the coalescent approach (*e.g.*, Kubatko & Degnan, 2007; Xi *et al.*, 2014). However, it can be used as a null hypothesis to compare to other methods explicitly modeling the biological processes that are resulting in gene tree incongruences. BUCKY's performance decreases with increasing number of taxa and/or when loci have few informative sites (Chung & Ané, 2011), nevertheless we interpret the BCA result as an indication of hybridization. The "simple" model in *BEAST excludes gene flow between species (Heled & Drummond, 2010), which in turn can

potentially disrupt the signal. Interestingly, increasing the number of loci in the supermatrix did not recover the monophyly of the Patagonian clade, confirming the superiority of the coalescent approach to summarize conflicting phylogenies. More studies analyzing the discrepancy between multilocus phylogeny inference methods such as Zwickl *et al.* (2014) are necessary to better understand their relative performance in the presence of natural processes leading to loci incongruences.

CONCLUSIONS

Our study of the phylogeny of *Hordeum* using second-generation sequencing of PCR amplicons obtained, for the first time, the species phylogeny and progenitor-derivative relationships of all di- and polyploid *Hordeum* taxa within a single analysis. We were also able to provide a time frame for the evolution of the genus. The general shift of paradigm towards multilocus analyses in phylogeny is still limited by the initial selection of enough single-copy loci. The resources available for grasses and especially Hordeeae made it possible to use a PCR-based method, which considerably reduced the data complexity. A sequence-capture approach (Lemmon *et al.*, 2012; Mascher *et al.*, 2013) might result in much more demanding data, as a substantial amount of sequences may be off-target and therefore harder to handle (own unpublished data). We chose the 454 platform for its long reads, but with generally increasing read lengths in NGS, the possibility to sequence the ends of fragments up to *c.* 700 bp length in paired-end mode on an Illumina HiSeq (even longer on a MiSeq) might define the maximum distance of two polymorphic sites to be used to correctly phase alleles or homoeologues. Thus, long conserved stretches of DNA, separating more variable parts within a sequenced locus, can prevent phasing till the very long reads of single-molecule sequencing becomes widely available. Finally, it appears necessary to analyze multilocus data using different methods to disentangle biological and methodological biases.

Most of the open questions regarding relationships among di- and polyploid taxa of the genus were solved in this phylogenetic analysis. However, the *H. brevisubulatum* polyploid complex is still puzzling, and might only be solvable by extensive population studies throughout the distribution area of the taxon. Also for the hexaploid *H. parodii* and the two closely related tetraploid species pairs *H. fuegianum*/*H. tetraploidum* and *H. capense*/*H. secalinum* some additional attention might be needed regarding their mode of evolution. This must, however, include many more individuals/populations in comparison to this study, whose aim was the overall phylogeny within *Hordeum*.

4. SYNTHESIS

Hordeum, with nearly half of its 45 taxa being polyploids, has been challenging systematists and molecular evolutionists for decades. But a known phylogeny of the genus is necessary to provide relationships between species as wells as an evolutionary time frame. Both of which are the foundations to all evolutionary and ecological studies aiming to analyze progenitor-derivative evolution and adaptation (Jakob *et al.*, 2007, 2009, 2010). During the course of this DFG-funded project, the aim was to resolve all phylogenetic relationships within this genus. Including all species and cytotypes, mostly several individuals per species, and combining single and multilocus analyses we were able to provide, for the first time, the species phylogeny and progenitor-derivative relationships of all di- and polyploid *Hordeum* taxa.

The whole project was divided in two complementary parts. The first one evaluated the intraspecific, as well as interspecific, nucleotide diversity at the single-copy nuclear locus *TOPO6* considering more than 300 individuals and nearly 1000 clones analyzed. After the first results on a single clade (Jakob & Blattner, 2010), the results confirmed this marker as a reliable tool to identify the progenitors of polyploid taxa at the scale of the whole genus. However, the resolution of a single marker is too low to infer safe relationships, especially for the recently radiated South American species. To be able to safely resolve all species affiliations, the second part of the project was based on sequencing 12 nuclear single-copy loci and one chloroplast locus on a subset of individuals from the first analysis, *i.e.* 105 individuals. Taking advantage of the second-generation sequencing method, we focused on confirming and refining the previously inferred relationships.

A Shift of Paradigm in Phylogenetics: Multilocus Analyses

With the advent of NGS, the phylogeny paradigm changed from the single locus/deep sampling strategy (*e.g.*, Nürk *et al.*, 2013) towards the multilocus or even genomic/shallow sampling strategy (Lemmon *et al.*, 2012). Though, with the very high throughput of the Illumina sequencing platforms and the possibility to barcode hundreds or thousands of individuals at the same time, studies tend towards population phylogenomics (Peterson *et al.*, 2012). For the study presented here, the single locus analysis (Chapter 2) required an intense laboratory work but only straightforward sequence analysis on the computer. However, the chimerical sequences obtained from the cloning slightly increased the computer load and pointed to the necessity to carefully inspect the sequences. The NGS analysis (Chapter 3) was relatively light in the laboratory, although PCR amplicons were generated for 16 loci, but the computer part was much more intensive. The lack of dedicated tools for halpotype phasing from NGS data and the need to run highly computing and memory-demanding analyses requires bioinformatic skills to handle the heavy data load. Finally, the general incongruence among loci and between phylogenetic inference methods necessitates the comparison of all methods taking into account what they are actually modeling.

Progenitors of Polyploid Taxa

The main objective of this work, the identification of the progenitors of polyploid species, was achieved. However, all but the allopolyploid species *H. depressum* revealed that at least one extinct lineage contributed to their formation. Although a sub-structure appeared in the two clades representing the extinct taxa related to *H. roshevitzii* and *H. californicum*, to a smaller extent, we cannot be sure if this diversity originated before or after the polyploid formation (Fig. S3.16). With not less than six extinct taxa inferred, further studies focusing on the consequences of polyploidization (Jakob *et al.*, 2007) will be hampered by the lack of direct comparisons. Accurately dating the origin of polyploid events, for example, will be impossible as only maximum ages can be estimated and will be overestimated when progenitors are extinct (Doyle & Egan, 2010; Jakob & Blattner, 2010). Several studies have inferred extinct progenitors of allopolyploid taxa (Roelofs *et al.*, 1997; Blattner, 2004; Hoot *et al.*, 2004; Lihovà *et al.*, 2006). However, we are currently unable to evaluate the extent of such phenomenon as the inclusion of all species and cytotypes with more than one representative per taxon in phylogenetic analyses is still rare.

The contribution of an extinct progenitor to a derivative taxon could actually be an artifact. Two factors will have the same effect: one at the population and one at the molecular level. We might have not sampled the lineages involved in the formation of polyploid taxa. Despite our large sampling,

including all extant species and cytotypes and with an average of eight individuals per cytotype in the first analysis, we cannot exclude this factor but its probability is low. Theoretical studies showed that a neopolyploid taxon can coexist or replace one or both of its progenitors depending, for example, on selfing rates and niche shifts (Levin, 1975; Fowler & Levin, 1984; Felber, 1991; Rodriguez, 1996). Ayres *et al.* (2004) showed that the extinction of an American native taxon followed the introduction of a European invasive congener by formation of hybrid individuals. On the molecular level, sequences might have undergone differential evolution between progenitors and descendants to the extent of behaving like a sister species. The potential evolutionary fates of duplicated loci in polyploids have been extensively described but are difficult to demonstrate (Wendel, 2000). On the one hand, processes such as pseudogenization, due to genetic redundancy and reduced selective pressure, or functional divergence result in higher divergence rates in duplicated loci. On the other hand, it has been shown that, despite the rapid genomic changes that can occur, (*e.g.*, Song *et al.*, 1995; Buggs *et al.*, 2012), loci can evolve independently and at the same rate after polyploidization (Cronn *et al.*, 1999). Likewise, the probability of observing a similar pattern across 12 loci and ten taxa in *Hordeum*, partially clustering together, is low.

Essentially in the case of *H. bulbosum*, autopolyploidy was confirmed as no evidence could be found for lineages distinct from the diploid cytotype participating in the autotetraploid. *Hordeum brevisubulatum* appeared to be a very complex taxon, requiring population level studies to understand its full extent. Distributed over a very large area, consisting of five subspecies described on three different ploidy levels (Bothmer *et al.*, 1995) and with both allo- and autotetraploid individuals, *H. brevisubulatum* represents by itself an example for the continuum between allopolyploidy and autopolyploidy (Stebbins, 1950; Grant, 1981).

Origin of Allopolyploids

The sequences obtained from particular polyploid taxa were mostly forming unique clades per homoeologue copy, eventually with other di- or polyploid species, suggesting that these taxa originated probably through a single hybridization event. Very low nucleotide diversity, as observed in the hexaploid species *H. procerum*, is an indication for a monophyletic origin. Polyploid taxa with relatively large nucleotide diversity, such as *H. tetraploidum*, or even displaying two different diploid parental species, such as *H. parodii*, are most probably of recurrent origin. Although a larger sampling than only two to four individuals per species would be necessary to confirm the single origin hypothesis, it seems that both scenarios (Soltis & Soltis, 1999) took place in *Hordeum*.

PHYLOGENETIC RELATIONSHIPS AND EVOLUTIONARY TIME FRAME

The infrageneric treatment of the genus proposed by Blattner (2009) could not be strictly retrieved with the single locus analysis (*e.g.*, *H. bogdanii* clustering with the American species; Chapter 2). However, it appeared robust to both the phylogenetic methods and the number of loci in the multilocus analysis (Chapter 3). The latest multilocus analysis in *Hordeum* based on nine nuclear and seven plastid sequence regions (Petersen *et al.*, 2011) presented a similar topology but could not recover the monophyly of the Asian clade, series *Sibirica*, neither with the nuclear regions dataset only nor with the combined datasets including also chloroplast and mitochondrial sequences. The main difference between the analysis presented in Chapter 3 and the latter seems to be the taxon and accessions sampling. Strong focus has been put on the importance of taxon sampling for the accuracy of phylogenetic inferences (*e.g.*, Zwickl & Hillis, 2002; Nürk *et al.*, 2013). Capturing the intraspecific diversity seemed informative and is definitely necessary for the inference of species trees from gene trees, as more individuals per species should increase the precision of the trees (Heled & Drummond, 2010).

To conclude, I would like to review all the evidences accumulated during this project and previous analyses in order to update the biogeographical scenario of the genus *Hordeum* formulated by Blattner (2006) including all polyploids. The multilocus analysis revealed a new time frame for the evolution of the genus at the diploid level based on the latest dating available for the tribe Hordeeae (Marcussen *et al.*, 2014). The most recent common ancestor of all species of the tribe Hordeeae probably occurred *ca.* 15 Ma (Chapter 3; Marcussen *et al.*, 2014) in Eurasia, the area described as the cradle of the tribe (Hsiao *et al.*, 1999). *Hordeum*, a basal lineage in Hordeeae (Escobar *et al.*, 2011), started to diversify *ca.* 10 Ma and dispersed to the Mediterranean area (*H. bulbosum*, *H. marinum*). The crown age of the genus was younger than previously thought (Wolfe *et al.*, 1989; Gaut, 2002; Blattner, 2004, 2006; Chalupska *et al.*, 2008). The **I**-genome lineage, occurring in Asia and Americas, separated from the Mediterranean **Xa**-genome about 5 Ma. The relationships within the genus revealed by this work necessitate multiple colonization of the New World from Siberia through Beringia. An Asian lineage first crossed Beringia *ca.* 1.3 Ma colonizing North America (*H. californicum*) and, probably through long distance dispersal or migration along the western American mountain ridges (*cf.* Blattner, 2006 and citations therein), South America. There, in less than one million years, a fast radiation occurred at the diploid level. The synchronicity of speciation events in South America (*e.g.*, *H. patagonicum-pubiflorum*, *H. flexuosum-chilense* and *H. erectifolium-stenostachys* at *ca.* 0.3 Ma) seems to indicate vicariance events caused by climatic events. Moreover, long distance dispersal back to North America also occurred twice: first for *H. pusillum* and then for *H. intercedens*.

The net rate of diversification, considering only diploids and no extinction, in America is up to 10 times higher than in Eurasia (1.21 and 0.15 species^{-my}, respectively). However, this is still modest

when compared to the most rapid angiosperm radiations (reviewed in Valente *et al.*, 2010). The rate of diversification in the Old-World is coherent with the one estimated for "Higher" Monocots (Magallon & Sanderson, 2001). A higher extinction rate during ice ages, potentially enhanced by the East-West mountain ranges in Eurasia, is assumed to explain the differential haplotype richness between both areas (Jakob & Blattner, 2006) and the high proportion of extinct lineages having contributed their genome to polyploid formation (Chapter 3).

Although no dating is available for the polyploids, we can reconstruct their potential biogeographical scenario from their sequence information. The formation of polyploid taxa in the New World involved an additional passage from Asia to America via the Bering land bridge, as all polyploids except *H. depressum* have an Asian genome. Although they are morphologically different, the three North American tetraploids, *H. jubatum* (long awned), *H. brachyantherum* and *H. guatemalense* (both short awned), have the same genome composition involving the two extinct progenitors related to *H. roshevitzii* as paternal donor and to *H. californicum* as maternal donor. The three taxa probably evolved from a vicariance event splitting a large tetraploid population and adaptation to somewhat dryer environments for *H. jubatum*. An alternative explanation would involve independent migrations from Siberia to explain the formation of each polyploid taxon and the diversity observed. The Asian progenitor involved in those tetraploid taxa also reached South America where it hybridized with *H. pubiflorum*, or a close relative, to form the two sister species *H. tetraploidum* and *H. fuegianum*. The newly formed *H. jubatum* expanded its distribution, now covering parts of Siberia, most of North America and also in Argentina. In the two latter, it hybridized with the diploid taxa to form all American hexaploids, except *H. brachyantherum* involving the tetraploid cytotype together with the European *H. gussoneanum* and *H. parodii* involving *H. tetraploidum* as paternal progenitor. Finally, the two sister species *H. secalinum* and *H. capense* probably occurred from a reciprocal cross between an extinct Asian lineage, distinct from the one that colonized the Americas, and *H. gussoneanum* before *H. capense* dispersed from Europe to South Africa.

Now that we have a robust phylogeny of the genus, open questions such as the evolution of complex and challenging taxa can be asked. For example, an in-depth study of the South American taxon *H. parodii* is necessary to investigate its taxonomy. The analysis of the two couples of sister species *H. capense*/*secalinum* (ongoing project) and *H. fuegianum*/*tetraploidum* is interesting to understand their formation, difference and evolution. Finally *H. brevisubulatum*, the very challenging Asian species, requires the analysis of hundreds of specimens and population genomics to understand how hybridization shaped this taxon.

ABSTRACT

Hordeum L. belongs to the grass tribe Hordeeae of the Poaceae family. The genus consists of 33 species and altogether 45 taxa are described. The genus is distributed in arid and temperate regions of the world. Polyploidization, after hybridization for example, played a major role in the evolution of this genus. Nearly half of the described taxa are polyploids, with both allo- and autopolyploids, as well as tetra- and hexaploids. However, the presence of multiple entire genomes within polyploid taxa complicates phylogenetic analyses, as the different copies of the loci need to be deciphered either before sequencing or after. In this thesis, I analyze the relationships within the diploid species, between the diploid and the polyploid taxa, and within the polyploid species themselves.

Relationships among all *Hordeum* species were first analyzed with the single copy nuclear locus *TOPO6* for 341 *Hordeum* individuals and eight outgroup species. PCR products were either directly sequenced or cloned. This first analysis confirmed the ability of *TOPO6* to identify most of the progenitors of polyploid taxa and provided information about the intraspecific diversity of this locus, which was used to select individuals reflecting most of the diversity for the next analysis. The next-generation sequencing platform 454 was then used to sequence simultaneously one chloroplast and 12 nuclear single-copy loci, that were first PCR-amplified, from 96 individuals representing all *Hordeum* species and cytotypes. Traditional parsimony and model-based methods, including the more recent coalescent-based species tree and molecular dating, were used to analyze relationships between the sequences derived from all loci separately as well as combined.

The results of both analyses are mostly congruent, although the multilocus analysis provided better resolution allowing the identification of all the progenitor-derivative relationships. However, at least six clades are interpreted as footprints of extinct diploid taxa, which contributed to allopolyploid evolution. Finally, a biogeographical scenario taking into account all evidences is proposed. Three key species involved in the evolution of the American polyploids of the genus were identified. (*i*) The Central Asian diploid *H. roshevitzii*, or one of its extinct relatives, provided its genome to all but one of the American tetraploids, the second originating from different American diploid species. (*ii*) All hexaploid species from the New World, except *H. parodii*, have a genome of an extinct close relative of *H. californicum* and (*iii*) possess the genome of tetraploid *H. jubatum*, each with an additional genome from different American diploids. Tetraploid *H. bulbosum* is an autopolyploid, while the assumed autopolyploid *H. brevisubulatum* (4×, 6×) was identified as allopolyploid throughout most of its distribution area.

ZUSAMMENFASSUNG

Hordeum L. (Hordeeae, Poaceae) umfasst 33 Arten und insgesamt 45 Taxa. Die Gattung ist in trockenen und gemäßigten Regionen der Welt verbreitet. Polyploidisierung, zum Beispiel nach erfolgter Hybridisierung, spielte eine wichtige Rolle bei der Entwicklung dieser Gattung. Ungefähr die Hälfte der Taxa sind polyploid (sowohl allo- und autopolyploid, und tetra- und hexaploid). Jedoch erschwert das Vorhandensein mehrerer ganzer Genome innerhalb polyploiden Taxa phylogenetische Analysen. Die verschiedenen Kopien der Loci müssen entweder vor oder nach der Sequenzierung entschlüsselt werden. In dieser Arbeit untersuche ich die Beziehungen innerhalb der diploiden Arten, zwischen den diploiden und polyploiden, und innerhalb der polyploiden Arten selbst.

Die Beziehungen zwischen allen *Hordeum* Arten wurden zuerst mit dem nuklearen *single-copy* Locus *TOPO6* für 341 *Hordeum* Individuen und acht Außengruppen-Arten analysiert. PCR-Produkte wurden direkt sequenziert oder kloniert. Diese erste Analyse bestätigt das Vermögen des Locus, die meisten Vorfahren polyploider Taxa bestimmen zu können. Anhand der erhaltenen Informationen über die innerartliche Vielfalt dieses Locus, wurden 96 Individuen für die nächste Analyse ausgewählt, welche alle *Hordeum* Arten und Zytotypen repräsentieren und die größte genetische Vielfalt widerspiegelten. Die Next-Generation-Sequencing Plattform 454 wurde dann verwendet, um gleichzeitig einen Chloroplasten- und 12 nuklearen *single-copy* Loci, die zuvor mittels PCR amplifiziert wurden, für diese 96 Individuen zu sequenzieren. Traditionelle Parsimonie und model-basierte Methoden, inklusive der moderneren *coalescent-based species-tree* und molekularen Datierungsmethoden, wurden benutzt, um der Beziehungen zwischen den Sequenzen aller Loci separat sowie kombiniert zu analysieren.

Die Ergebnisse beider Analysen sind größtenteils kongruent, obwohl die multilocus Analyse eine höhere Aulösung lieferte, welche die Bestimmung aller „Vorfahren-Nachkommen“ Beziehungen erlaubte. Jedoch zeigen mindestens sechs *clades* Spuren ausgestorbener diploider Taxa, die zur Evolution allopolyploider Arten beigetragen haben. Schließlich wird ein biogeographisches Szenario unter Beachtung aller Beweise vorgeschlagen. Es wurden drei Schlüsselarten identifiziert, die an der Entstehung der amerikanischen Polyploiden beteiligt waren: (*i*) Die asiatische diploide Art *H. roshevitzii*, oder einer ihrer ausgestorben Verwandten, trug ihr nukleares Genom, aber niemals ihr Chloroplastengenom, zu allen amerikanischen Polyploiden (mit Ausnahme von *H. depressum*) bei. (*ii*) Alle hexaploiden Arten aus der Neuen Welt, außer *H. parodii*, weisen ein Genom eines ausgestorbenen nahen Verwandten von *H. californicum* auf und (*iii*) besitzen das Genom der tetraploiden Art *H. jubatum*, zusammen mit jeweils einem zusätzlichen Genom von verschiedenen amerikanischen diploiden Arten. Die tetraploide Art *H. bulbosum* ist autopolyploid, während der vermutete Autopolyploid *H. brevisubulatum* (4x, 6x) sowohl Allo- als auch Autopolyploid vorkommt.

REFERENCES

Adey A, Burton JN, Kitzman JO, Hiatt JB, Lewis AP, Martin BK, Qiu R, Lee C, Shendure J. **2013**. The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. *Nature* **500**: 207–211.

Ainouche ML, Baumel A, Salmon A, Yannic G. **2004**. Hybridization, polyploidy and speciation in *Spartina* (Poaceae). *New Phytologist* **161**: 165–172.

Ainouche ML, Fortune PM, Salmon A, Parisod C, Grandbastien M-A, Fukunaga K, Ricou M, Misset M-T. **2009**. Hybridization, polyploidy and invasion: lessons from *Spartina* (Poaceae). *Biological Invasions* **11**: 1159–1173.

Akaike H. **1973**. Information theory and an extension of the maximum likelihood principle. *Second international symposium on information theory* **1**: 267–281.

Akaike H. **1974**. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**: 716– 723.

Albach DC. **2007**. Amplified fragment length polymorphisms and sequence data in the phylogenetic analysis of polyploids: multiple origins of *Veronica cymbalaria* (Plantaginaceae). *New Phytologist* **176**: 481–498.

Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. **2004**. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* **20**: 407–415.

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

Álvarez I, Wendel JF. **2003**. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* **29**: 417–434.

Ané C, Larget B, Baum DA, Smith SD, Rokas A. **2007**. Bayesian estimation of concordance among gene trees. *Molecular Biology and Evolution* **24**: 412–426.

Ayres DR, Zaremba K, Strong DR. **2004**. Extinction of a common native species by hybridization with an invasive congener. *Weed Technology* **18**: 1288–1291.

Bailey JP, Bímová K, Mandák B. **2007**. The potential role of polyploidy and hybridisation in the further evolution of the highly invasive *Fallopia* taxa in Europe. *Ecological Research* **22**: 920–928.

Baker M. **2010**. Next-generation sequencing: adjusting to data overload. *Nature Methods* **7**: 495–499.

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**: 247.

Baum DA. **2007**. Concordance trees, concordance factors, and the exploration of reticulate genealogy. *Taxon* **56**: 417–426.

Baum BR, Johnson DA. **2003**. The South African *Hordeum capense* is more closely related to some American *Hordeum* species than to the European *Hordeum secalinum*: a perspective based on the 5S DNA units (Triticeae: Poaceae). *Canadian Journal of Botany* **81**: 1–11.

Bayes T. **1763**. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London* **53**: 370–418.

Beest M te, Roux JJL, Richardson DM, Brysting AK, Suda J, Kubešová M, Pyšek P. **2011**. The more the better? The role of polyploidy in facilitating plant invasions. *Annals of Botany* **109**: 19–45.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, *et al*. **2008**. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**: 53–59.

Bergerat A, de Massy B, Gadelle D, Varoutas P-C, Nicolas A, Forterre P. **1997**. An atypical topoisomerase II from archaea with implications for meiotic recombination. *Nature* **386**: 414–417.

Bergsten J. **2005**. A review of long-branch attraction. *Cladistics* **21**: 163–193.

Blattner FR. **1999**. Direct amplification of the entire ITS region from poorly preserved plant material using recombinant PCR. *BioTechniques* **27**: 1180–1186.

Blattner FR. **2004**. Phylogenetic analysis of *Hordeum* (Poaceae) as inferred by nuclear rDNA ITS sequences. *Molecular Phylogenetics and Evolution* **33**: 289–299.

Blattner FR. **2006**. Multiple intercontinental dispersals shaped the distribution area of *Hordeum* (Poaceae). *New Phytologist* **169**: 603–614.

Blattner FR. **2009**. Progress in phylogenetic analysis and a new infrageneric classification of the barley genus *Hordeum* (Poaceae: Triticeae). *Breeding Science* **59**: 471–480.

Booth TA, Richards AJ. **1976**. Studies in the *Hordeum murinum* aggregate: 1. Morphology. *Botanical Journal of the Linnean Society* **72**: 149–159.

Bothmer R von, Flink J, Landström T. **1986**. Meiosis in interspecific *Hordeum* hybrids. I. Diploid combinations. *Canadian Journal of Genetics and Cytology* **28**: 525–535.

Bothmer RV, Flink J, Landström T. **1988**. Meiosis in interspecific *Hordeum* hybrids: III tetraploid (2x × 6x) hybrids. *Hereditas* **108**: 141–148.

Bothmer R von, Jacobsen N. **1985**. Origin, taxonomy, and related species. In: Rasmusson DC, ed. Agronomy Monographs 26. Barley. Madison, USA: American Society of Agronomy, 19–56.

Bothmer R von, Jacobsen N, Baden C, Jørgensen R, Linde-Laursen I. **1995**. *An ecogeographical study of the genus* Hordeum. Rome, Italy: IPGRI.

Box GEP. **1976**. Science and statistics. *Journal of the American Statistical Association* **71**: 791–799.

Bradley RD, Hillis DM. **1997**. Recombinant DNA sequences generated by PCR amplification. *Molecular Biology and Evolution* **14**: 592–593.

Brassac J, Jakob SS, Blattner FR. **2012**. Progenitor-derivative relationships of *Hordeum* polyploids (Poaceae, Triticeae) inferred from sequences of *TOPO6*, a nuclear low-copy gene region. *PLoS ONE* **7**: e33808.

Brochmann C, Soltis PS, Soltis DE. **1992**. Recurrent formation and polyphyly of Nordic polyploids in *Draba* (Brassicaceae). *American Journal of Botany* **79**: 673–688.

Buggs RJA, Renny-Byfield S, Chester M, Jordon-Thaden IE, Viccini LF, Chamala S, Leitch AR, Schnable PS, Barbazuk WB, Soltis PS, *et al*. 2012. Next-generation sequencing and genome evolution in allopolyploids. *American Journal of Botany* **99**: 372–382.

Cai D, Rodríguez F, Teng Y, Ané C, Bonierbale M, Mueller LA, Spooner DM. **2012**. Single copy nuclear gene analysis of polyploidy in wild potatoes (*Solanum* section *Petota*). *BMC Evolutionary Biology* **12**: 70.

Carmona A, Friero E, Bustos A de, Jouve N, Cuadrado A. **2013**. The evolutionary history of sea barley (*Hordeum marinum*) revealed by comparative physical mapping of repetitive DNA. *Annals of Botany* **112**: 1845–1855.

Chalupska D, Lee HY, Faris JD, Evrard A, Chalhoub B, Haselkorn R, Gornicki P. **2008**. *Acc* homoeoloci and the evolution of wheat genomes. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 9691–9696.

Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, *et al*. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods* **10**: 563–569.

Chung Y, Ané C. **2011**. Comparing two Bayesian methods for gene tree/species tree reconstruction: simulations with incomplete lineage sorting and horizontal gene transfer. *Systematic Biology* **60**: 261–275.

Comes HP, Kadereit JW. **1998**. The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science* **3**: 432–438.

Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF. **2002**. PCR-mediated recombination in amplification products derived from polyploid cotton. *Theoretical and Applied Genetics* **104**: 482–489.

Cronn RC, Small RL, Wendel JF. **1999**. Duplicated genes evolve independently after polyploid formation in cotton. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 14406–14411.

Cuadrado Á, Carmona A, Jouve N. **2013**. Chromosomal characterization of the three subgenomes in the polyploids of *Hordeum murinum* L.: new insight into the evolution of this complex. *PLoS ONE* **8**: e81385.

Darriba D, Taboada GL, Doallo R, Posada D. **2012**. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**: 772–772.

Degnan JH, Rosenberg NA. **2006**. Discordance of species trees with their most likely gene trees. *PLoS Genetic* **2**: e68.

Degnan JH, Rosenberg NA. **2009**. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution* **24**: 332–340.

Dewey DR. **1984**. The genomic system of classification as a guide to intergeneric hybridization with the perennial Triticeae. In: Gustafson JP, ed. Stadler Genetics Symposia Series. *Gene manipulation in plant improvement*. Springer US, 209–279.

Doebley J, Bothmer R von, Larson S. **1992**. Chloroplast DNA variation and the phylogeny of *Hordeum* (Poaceae). *American Journal of Botany* **79**: 576–584.

Doyle JJ, Egan AN. **2010**. Dating the origins of polyploidy events. *New Phytologist* **186**: 73–85.

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. **2008**. Evolutionary genetics of genome merger and doubling in plants. *Annual Review of Genetics* **42**: 443–461.

Doyle JJ, Gaut BS. **2000**. Evolution of genes and taxa: a primer. In: Doyle JJ, Gaut BS, eds. Plant Molecular Evolution. Springer Netherlands, 1–23.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, *et al.*. 2011. Geneious v5.4, Available from http://www.geneious.com/.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. **2006**. Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**: e88.

Drummond AJ, Rambaut A. **2007**. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**: 214.

Drummond AJ, Suchard MA. **2010**. Bayesian random local clocks, or one rate to rule them all. *BMC Biology* **8**: 114.

Drummond AJ, Suchard MA, Xie D, Rambaut A. **2012**. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**: 1969–1973.

Edwards SV. **2009**. Is a new and general theory of molecular systematics emerging? *Evolution* **63**: 1–19.

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, *et al*. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* **323**: 133–138.

El-Rabey HA, Badr A, Schäfer-Pregl R, Martin W, Salamini F. **2002**. Speciation and species separation in *Hordeum* L. (Poaceae) evolved by discontinuous molecular markers. *Plant Biology* **4**: 567–575.

Escobar J, Scornavacca C, Cenci A, Guilhaumon C, Santoni S, Douzery E, Ranwez V, Glémin S, David J. **2011**. Multigenic phylogeny and analysis of tree incongruences in Triticeae (Poaceae). *BMC Evolutionary Biology* **11**: 181.

Farris JS, Albert VA, Källersjö M, Lipscomb D, Kluge AG. **1996**. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* **12**: 99–124.

Fawcett JA, Maere S, Van de Peer Y. **2009**. Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 5737 –5742.

Felber F. **1991**. Establishment of a tetraploid cytotype in a diploid population: effect of relative fitness of the cytotypes. *Journal of Evolutionary Biology* **4**: 195–207.

Felsenstein J. **1978**. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* **27**: 401–410.

Felsenstein J. **1985**. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.

Fitch WM. **1970**. Distinguishing homologous from analogous proteins. *Systematic Zoology* **19**: 99–113.

Fitch WM. **1971**. Toward defining the course of evolution: minimum change for a specific tree topology. *Systematic Zoology* **20**: 406–416.

Flagel LE, Wendel JF, Udall JA. **2012**. Duplicate gene evolution, homoeologous recombination, and transcriptome characterization in allopolyploid cotton. *BMC Genomics* **13**: 302.

Flot J-F, Tillier A, Samadi S, Tillier S. **2006**. Phase determination from direct sequencing of length-variable DNA regions. *Molecular Ecology Notes* **6**: 627–630.

Fowler NL, Levin DA. **1984**. Ecological constraints on the establishment of a novel polyploid in competition with its diploid progenitor. *American Naturalist* **124**: 703–711.

Gaut BS. **2002**. Evolutionary dynamics of grass genomes. *New Phytologist* **154**: 15–28.

Gaut BS, Muse SV, Clark WD, Clegg MT. **1992**. Relative rates of nucleotide substitution at the *rbcL* locus of monocotyledonous plants. *Journal of Molecular Evolution* **35**: 292–303.

Grant V. **1981**. *Plant speciation*. New York: Columbia University Press.

Griffin P, Robin C, Hoffmann A. **2011**. A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to *Poa* grasses. *BMC Biology* **9**: 19.

Guindon S, Gascuel O. **2003**. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**: 696–704.

Harpke D, Meng S, Rutten T, Kerndorff H, Blattner FR. **2013**. Phylogeny of *Crocus* (Iridaceae) based on one chloroplast and two nuclear loci: ancient hybridization and chromosome number evolution. *Molecular Phylogenetics and Evolution* **66**: 617–627.

Hartung F, Blattner FR, Puchta H. **2002**. Intron gain and loss in the evolution of the conserved eukaryotic recombination machinery. *Nucleic Acids Research* **30**: 5175 –5181.

Hartung F, Puchta H. **2001**. Molecular characterization of homologues of both subunits A (SPO11) and B of the archaebacterial topoisomerase 6 in plants. *Gene* **271**: 81–86.

Hasegawa M, Kishino H, Yano T. **1985**. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**: 160–174.

Hastings WK. **1970**. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**: 97–109.

Heath L, Van Der Walt E, Varsani A, Martin DP. **2006**. Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *Journal of virology* **80**: 11827.

Hein J, Schierup M, Wiuf C. **2004**. *Gene genealogies, variation and evolution: a primer in coalescent theory*. Oxford University Press.

Heled J, Drummond AJ. **2010**. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27**: 570–580.

Hewitt GM. **1996**. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**: 247–276.

Hewitt G. **2000**. The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907–913.

Hoot SB, Napier NS, Taylor WC. **2004**. Revealing unknown or extinct lineages within *Isoëtes* (Isoëtaceae) using DNA sequences from hybrids. *American Journal of Botany* **91**: 899 –904.

Hsiao C, Chatterton NJ, Asay KH, Jensen KB. **1995**. Molecular phylogeny of the Pooideae (Poaceae) based on nuclear rDNA (ITS) sequences. *Theoretical and Applied Genetics* **90**: 389–398.

Hsiao C, Jacobs SWL, Chatterton NJ, Asay KH. **1999**. A molecular phylogeny of the grass family (Poaceae) based on the sequences of nuclear ribosomal DNA (ITS). *Australian Systematic Botany* **11**: 667–688.

Hudson RR. **1990**. Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* **7**: 1–44.

Huelsenbeck JP, Larget B, Miller RE, Ronquist F. **2002**. Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic Biology* **51**: 673–688.

Huelsenbeck JP, Rannala B. **1997**. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* **276**: 227–232.

Huelsenbeck JP, Ronquist F. **2001**. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755.

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. **2007**. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology* **8**: R143.

Huttley GA, Jakobsen IB, Wilson SR, Easteal S. **2000**. How important is DNA replication for mutagenesis? *Molecular Biology and Evolution* **17**: 929–937.

Ishikawa G, Nakamura T, Ashida T, Saito M, Nasuda S, Endo TR, Wu J, Matsumoto T. **2009**. Localization of anchor loci representing five hundred annotated rice genes to wheat chromosomes using PLUG markers. *Theoretical and Applied Genetics* **118**: 499–514.

Ishikawa G, Yonemaru J, Saito M, Nakamura T. **2007**. PCR-based landmark unique gene (PLUG) markers effectively assign homoeologous wheat genes to A, B and D genomes. *BMC Genomics* **8**: 135.

Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C. **2007**. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**: 463–467.

Jakob SS, Blattner FR. **2006**. A chloroplast genealogy of *Hordeum* (Poaceae): long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. *Molecular Biology and Evolution* **23**: 1602–1612.

Jakob SS, Blattner FR. **2010**. Two extinct diploid progenitors were involved in allopolyploid formation in the *Hordeum murinum* (Poaceae: Triticeae) taxon complex. *Molecular Phylogenetics and Evolution* **55**: 650–659.

Jakob SS, Heibl C, Rödder D, Blattner FR. **2010**. Population demography influences climatic niche evolution: evidence from diploid American *Hordeum* species (Poaceae). *Molecular Ecology* **19**: 1423–1438.

Jakob SS, Ihlow A, Blattner FR. **2007**. Combined ecological niche modelling and molecular phylogeography revealed the evolutionary history of *Hordeum marinum* (Poaceae) — niche differentiation, loss of genetic diversity, and speciation in Mediterranean Quaternary refugia. *Molecular Ecology* **16**: 1713–1727.

Jakob SS, Martinez-Meyer E, Blattner FR. **2009**. Phylogeographic analyses and paleodistribution modeling indicate Pleistocene in situ survival of *Hordeum* species (Poaceae) in Southern Patagonia without genetic or spatial restriction. *Molecular Biology and Evolution* **26**: 907–923.

Jakob SS, Rödder D, Engler JO, Shaaf S, Özkan H, Blattner FR, Kilian B. **2014**. Evolutionary history of wild barley (*Hordeum vulgare* subsp. *spontaneum*) analyzed using multilocus sequence data and paleodistribution modeling. *Genome Biology and Evolution* **6**: 685–702.

Jukes T, Cantor C. **1969**. Evolution of protein molecules. In: Munro M, ed. Mammalian Protein Metabolism. Academic Press, 21–132.

Kakeda K, Taketa S, Komatsuda T. **2009**. Molecular phylogeny of the genus *Hordeum* using thioredoxin-like gene sequences. *Breeding Science* **59**: 595–601.

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

Katoh K, Toh H. **2008**. Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics* **9**: 286–298.

Kingman JFC. **1982**. The coalescent. *Stochastic Processes and their Applications* **13**: 235–248.

Knowles LL. **2009**. Estimating species trees: methods of phylogenetic analysis when there is incongruence across genes. *Systematic Biology* **58**: 463–467.

Komatsuda T, Salomon B, Bothmer R von. **2009**. Evolutionary process of *Hordeum brachyantherum* 6x and related tetraploid species revealed by nuclear DNA sequences. *Breeding Science* **59**: 611–616.

Komatsuda T, Salomon B, Bryngelsson T, Bothmer R von. **2001**. Phylogenetic analysis of *Hordeum marinum* Huds. based on nucleotide sequences linked to the *vrs1* locus. *Plant Systematics and Evolution* **227**: 137–144.

Komatsuda T, Tanno K, Salomon B, Bryngelsson T, Bothmer R von. **1999**. Phylogeny in the genus *Hordeum* based on nucleotide sequences closely linked to the *vrs1* locus (row number of spikelets). *Genome* **42**: 973–981.

Kotseruba V, Pistrick K, Blattner FR, Kumke K, Weiss O, Rutten T, Fuchs J, Endo T, Nasuda S, Ghukasyan A, *et al*. 2010. The evolution of the hexaploid grass *Zingeria kochii* (Mez) Tzvel. (2n= 12) was accompanied by complex hybridization and uniparental loss of ribosomal DNA. *Molecular Phylogenetics and Evolution* **56**: 146–155.

Kraytsberg Y, Nicholas A, Caro P, Khrapko K. **2008**. Single molecule PCR in mtDNA mutational analysis: genuine mutations vs. damage bypass-derived artifacts. *Methods (San Diego, Calif.)* **46**: 269–273.

Kubatko LS, Degnan JH. **2007**. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology* **56**: 17–24.

Kumar S. **2005**. Molecular clocks: four decades of evolution. *Nature Reviews Genetics* **6**: 654–662.

Lahr D, Katz L. **2009**. Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. *Biotechniques* **47**: 857–863.

Lanfear R, Calcott B, Ho SYW, Guindon S. **2012**. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29**: 1695–1701.

Larget BR, Kotha SK, Dewey CN, Ané C. **2010**. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics* **26**: 2910–2911.

Leitch AR, Leitch IJ. **2008**. Genomic plasticity and the diversity of polyploid plants. *Science* **320**: 481–483.

Lemmon AR, Emme SA, Lemmon EM. **2012**. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology* **61**: 727–744.

Levin DA. **1975**. Minority cytotype exclusion in local plant populations. *Taxon* **24**: 35.

Lewis WH. **1980**. Polyploidy in species populations. In: Lewis WH, ed. Basic Life Sciences. Polyploidy. Springer US, 103–144.

Lewis PO. **2001**. Phylogenetic systematics turns over a new leaf. *Trends in Ecology and Evolution* **16**: 30–37.

Li S. **1996**. Phylogenetic tree construction using Markov chain Monte Carlo. Ph.D. Dissertation, Ohio State University, Columbus.

Li H, Durbin R. **2011**. Inference of human population history from individual whole-genome sequences. *Nature* **475**: 493–496.

Li H, Homer N. **2010**. A survey of sequence alignment algorithms for next-generation sequencing. *Briefings in Bioinformatics* **11**: 473–483.

Lihovà J, Shimizu KK, Marhold K. **2006**. Allopolyploid origin of *Cardamine asarifolia* (Brassicaceae): incongruence between plastid and nuclear ribosomal DNA sequences solved by a single-copy nuclear gene. *Molecular Phylogenetics and Evolution* **39**: 759–786.

Linde-Laursen I, Bothmer R von, Jacobsen N. **1980**. Giemsa C-banding in Asiatic taxa of *Hordeum* section *Stenostachys* with notes on chromosome morphology. *Hereditas* **93**: 235–254.

Linde-Laursen I, Bothmer R von, Jacobsen N. **1990**. Giemsa C-banded karyotypes of diploid and tetraploid *Hordeum bulbosum* (Poaceae). *Plant Systematics and Evolution* **172**: 141–150.

Linnaeus C von. **1753**. *Species plantarum*. Stockholm.

Löve A. **1984**. Conspectus of the Triticeae. *Feddes Repertorium* **95**: 425–521.

Maddison WP. **1997**. Gene trees in species trees. *Systematic Biology* **46**: 523–536.

Magallon S, Sanderson MJ. **2001**. Absolute diversification rates in angiosperm clades. *Evolution* **55**: 1762–1780.

Mahelka V, Kopecký D. **2010**. Gene capture from across the grass family in the allohexaploid *Elymus repens* (L.) Gould (Poaceae, Triticeae) as evidenced by ITS, GBSSI, and molecular cytogenetics. *Molecular Biology and Evolution* **27**: 1370 –1390.

Marcussen T, Jakobsen KS, Danihelka J, Ballard HE, Blaxland K, Brysting AK, Oxelman B. **2012**. Inferring species networks from gene trees in high-polyploid North American and Hawaiian violets (*Viola*, Violaceae). *Systematic Biology* **61**: 107–126.

Marcussen T, Sandve SR, Heier L, Spannagl M, Jakobsen KS, Pfeifer M, The International Wheat Genome Sequencing Consortium, Wulff BBH, Steuernagel B, Mayer KFX, *et al.*. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345**: 1250092–1–1250092–4.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, *et al*. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376–380.

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. **2010**. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**: 2462–2463.

Mascher M, Richmond TA, Gerhardt DJ, Himmelbach A, Clissold L, Sampath D, Ayling S, Steuernagel B, Pfeifer M, D'Ascenzo M, *et al*. 2013. Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *The Plant Journal* **76**: 494–505.

Mason-Gamer RJ. **2013**. Phylogeny of a genomically diverse group of *Elymus* (Poaceae) allopolyploids reveals multiple levels of reticulation. *PLoS ONE* **8**: e78449.

Masterson J. **1994**. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* **264**: 421–424.

Mau B. **1996**. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. Ph.D. Dissertation, University Wisconsin, Madison.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. **1953**. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21**: 1087–1092.

Meyer M, Briggs AW, Maricic T, Höber B, Höffner B, Krause J, Weihmann A, Pääbo S, Hofreiter M. **2008a**. From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Research* **36**: e5.

Meyer M, Stenzel U, Hofreiter M. **2008b**. Parallel tagged sequencing on the 454 platform. *Nature Protocols* **3**: 267–278.

Meyer M, Stenzel U, Myles S, Prüfer K, Hofreiter M. **2007**. Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Research* **35**: e97.

Meyerhans A, Vartanian J-P, Wain-Hobson S. **1990**. DNA recombination during PCR. *Nucleic Acids Research* **18**: 1687–1691.

Miller JR, Koren S, Sutton G. **2010**. Assembly algorithms for next-generation sequencing data. *Genomics* **95**: 315–327.

Nei M, Kumar S. **2000**. *Molecular evolution and phylogenetics*. Oxford University Press.

Nevski SA. **1941**. Beiträge zur Kenntniss der wildwachsenden Gersten in Zusammenhang mit der Frage über den Ursprung von *Hordeum vulgare* L. and *Hordeum distichum* L. (Versuch einer Monographie der Gattung *Hordeum*). *Trudy Bot. Inst. Akad. Nauk, SSSR*: 64–255.

Nichols R. **2001**. Gene trees and species trees are not the same. *Trends in Ecology and Evolution* **16**: 358–364.

Nishikawa T, Salomon B, Komatsuda T, Bothmer R von, Kadowaki K. **2002**. Molecular phylogeny of the genus *Hordeum* using three chloroplast DNA sequences. *Genome* **45**: 1157–1166.

Nürk NM, Madriñán S, Carine MA, Chase MW, Blattner FR. **2013**. Molecular phylogenetics and morphological evolution of St. John's wort (Hypericum; Hypericaceae). *Molecular Phylogenetics and Evolution* **66**: 1–16.

Nylander J. **2004**. MrModeltest v2.3 program distributed by the author. *Evolutionary Biology Centre, Uppsala University*.

Nylander JAA, Wilgenbusch JC, Warren DL, Swofford DL. **2008**. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics* **24**: 581–583.

Ohta T. **2002**. Near-neutrality in evolution of genes and gene regulation. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 16134–16137.

Olmstead RG, Sweere JA. **1994**. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. *Systematic Biology* **43**: 467–481.

O'Neill EM, Schwartz R, Bullock CT, Williams JS, Shaffer HB, Aguilar-Miguel X, Parra-Olea G, Weisrock DW. **2013**. Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. *Molecular Ecology* **22**: 111–129.

Otto SP, Whitton J. **2000**. Polyploid incidence and evolution. *Annual Review of Genetics* **34**: 401–437.

Pamilo P, Nei M. **1988**. Relationships between gene trees and species trees. *Molecular Biology and Evolution* **5**: 568–583.

Petersen G, Aagesen L, Seberg O, Larsen IH. **2011**. When is enough, enough in phylogenetics? A case in point from *Hordeum* (Poaceae). *Cladistics* **27**: 428–446.

Petersen G, Seberg O. **2003**. Phylogenetic analyses of the diploid species of *Hordeum* (Poaceae) and a revised classification of the genus. *Systematic Botany* **28**: 293–306.

Petersen G, Seberg O. **2004**. On the origin of the tetraploid species *Hordeum capense* and *H. secalinum* (Poaceae). *Systematic Botany* **29**: 862–873.

Petersen G, Seberg O. **2009**. *Stowaway* MITEs in *Hordeum* (Poaceae): evolutionary history, ancestral elements and classification. *Cladistics* **25**: 198–208.

Petersen G, Seberg O, Yde M, Berthelsen K. **2006**. Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). *Molecular Phylogenetics and Evolution* **39**: 70–82.

Petit JR, Jouzel J, Raynaud D, Barkov NI, Barnola J-M, Basile I, Bender M, Chappellaz J, Davis M, Delaygue G, *et al*. 1999. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**: 429–436.

Pickering RA, Hill AM, Michel M, Timmerman-Vaughan GM. 1995. The transfer of a powdery mildew resistance gene from *Hordeum bulbosum* L. to barley (*H. vulgare* L.) chromosome 2 (2I). *Theoretical and Applied Genetics* **91:** 1288–1292.

Pillen K, Binder A, Kreuzkam B, Ramsay L, Waugh R, Förster J, Léon J. **2000**. Mapping new EMBL-derived barley microsatellites and their use in differentiating German barley cultivars. *TAG Theoretical and Applied Genetics* **101**: 652–660.

Pleines T, Blattner FR. **2008**. Phylogeographic implications of an AFLP phylogeny of the American diploid *Hordeum* species (Poaceae: Triticeae). *Taxon* **57**: 875–881.

Posada D, Buckley TR. **2004**. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology* **53**: 793–808.

Posada D, Crandall KA. **1998**. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.

Posada D, Crandall KA. **2001**. Selecting the best-fit model of nucleotide substitution. *Systematic Biology* **50**: 580–601.

De Queiroz A, Gatesy J. **2007**. The supermatrix approach to systematics. *Trends in Ecology and Evolution* **22**: 34–41.

Rabassa J, Coronato A, Martínez O. **2011**. Late Cenozoic glaciations in Patagonia and Tierra del Fuego: an updated review. *Biological Journal of the Linnean Society* **103**: 316–335.

Rambaut A, Drummond A. **2007**. Tracer v1.5. Available from http://beast.bio.ed.ac.uk/Tracer.

Ramsey J, Schemske DW. **1998**. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* **29**: 467–501.

Rannala B, Yang Z. **1996**. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *Journal of Molecular Evolution* **43**: 304–311.

Rannala B, Yang Z. **2003**. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**: 1645–1656.

Ranwez V, Holtz Y, Sarah G, Ardisson M, Santoni S, Glémin S, Tavaud-Pirra M, David J. **2013**. Disentangling homeologous contigs in allo-tetraploid assembly: application to durum wheat. *BMC Bioinformatics* **14**: S15.

Rauscher JT, Doyle JJ, Brown AHD. **2002**. Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the *Glycine tomentella* (Leguminosae) polyploid complex. *Molecular Ecology* **11**: 2691–2702.

Reboud X, Zeyl C. **1994**. Organelle inheritance in plants. *Heredity* **72**: 132–140.

Rodriguez DJ. **1996**. A model for the establishment of polyploidy in plants. *The American Naturalist* **147**: 33–46.

Rodríguez F, Cai D, Teng Y, Spooner D. **2011**. Asymmetric single-strand conformation polymorphism: an accurate and cost-effective method to amplify and sequence allelic variants. *American Journal of Botany* **98**: 1061–1067.

Roelofs D, Van Velzen J, Kuperus P, Bachmann K. **1997**. Molecular evidence for an extinct parent of the tetraploid species *Microseris acuminata* and *M. campestris* (Asteraceae, Lactuceae). *Molecular Ecology* **6**: 641–649.

Rokas A, Williams BL, King N, Carroll SB. **2003**. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**: 798–804.

Ronquist F, Huelsenbeck JP. **2003**. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572 –1574.

Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP. **2012a**. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology* **61**: 973–999.

Ronquist F, Teslenko M, Mark P van der, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. **2012b**. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**: 539–542.

Rothberg JM, Leamon JH. **2008**. The development and impact of 454 sequencing. *Nature Biotechnology* **26**: 1117–1124.

Ruge B, Linz A, Pickering R, Proeseler G, Greif P, Wehling P. **2003**. Mapping of *Rym14 Hb*, a gene introgressed from *Hordeum bulbosum* and conferring resistance to BaMMV and BaYMV in barley. *Theoretical and Applied Genetics* **107**: 965–971.

Saitou N, Nei M. **1987**. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406–425.

Salmon A, Flagel L, Ying B, Udall JA, Wendel JF. **2010**. Homoeologous nonreciprocal recombination in polyploid cotton. *New Phytologist* **186**: 123–134.

Sambrook J, Fritsch EF, Maniatis T. **1989**. *Molecular cloning: a laboratory manual*. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press.

Sang T. **2002**. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* **37**: 121–147.

Sang T, Crawford DJ, Stuessy TF. **1995**. Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: implications for biogeography and concerted evolution. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 6813–6817.

Sang T, Pan J, Zhang D, Ferguson D, Wang C, Pan K-Y, Hong D-Y. **2004**. Origins of polyploids: an example from peonies (*Paeonia*) and a model for angiosperms. *Biological Journal of the Linnean Society* **82**: 561–571.

Savolainen V, Chase MW. **2003**. A decade of progress in plant molecular phylogenetics. *Trends in Genetics* **19**: 717–724.

Schmieder R, Lim YW, Rohwer F, Edwards R. **2010**. TagCleaner: identification and removal of tag sequences from genomic and metagenomic datasets. *BMC Bioinformatics* **11**: 341.

Schwarz G. **1978**. Estimating the dimension of a model. *The Annals of Statistics* **6**: 461–464.

Siddall ME, Kluge AG. **1997**. Probabilism and phylogenetic inference. *Cladistics* **13**: 313–336.

Small RL, Cronn RC, Wendel JF. **2004**. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* **17**: 145–170.

Sokal RR, Michener CD. **1958**. A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin* **38**: 1409–1438.

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, dePamphilis CW, Wall PK, Soltis PS. **2009**. Polyploidy and angiosperm diversification. *American Journal of Botany* **96**: 336.

Soltis DE, Soltis PS. **1999**. Polyploidy: recurrent formation and genome evolution. *Trends in Ecology and Evolution* **14**: 348–352.

Soltis DE, Soltis PS. **2003**. The role of phylogenetics in comparative genetics. *Plant Physiology* **132**: 1790–1800.

Soltis DE, Soltis PS, Pires JC, Kovarik A, Tate JA, Mavrodiev E. **2004a**. Recent and recurrent polyploidy in *Tragopogon* (Asteraceae): cytogenetic, genomic and genetic comparisons. *Biological Journal of the Linnean Society* **82**: 485–501.

Soltis DE, Soltis PS, Rieseberg LH. **1993**. Molecular data and the dynamic nature of polyploidy. *Critical Reviews in Plant Sciences* **12**: 243–273.

Soltis DE, Soltis PS, Tate JA. **2004b**. Advances in the study of polyploidy since Plant Speciation. *New Phytologist* **161**: 173–191.

Song K, Lu P, Tang K, Osborn TC. **1995**. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 7719–7723.

Stamatakis A. **2014**. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*: btu033.

Stankevičius K, Povilionis P, Lubys A, Menkevičius S, Janulaitis A. **1995**. Cloning and characterization of the unusual restriction-modification system comprising two restriction endonucleases and one methyltransferase. *Gene* **157**: 49–53.

Stebbins GL. **1947**. Types of polyploids: their classification and significance. *Advances in Genetics* **1**: 403–429.

Stebbins GL. **1950**. *Variation and evolution in plants*. Columbia University Press.

Susko E, Inagaki Y, Roger AJ. **2004**. On inconsistency of the neighbor-joining, least squares, and minimum evolution estimation when substitution processes are incorrectly modeled. *Molecular Biology and Evolution* **21**: 1629–1642.

Swofford DL. **2002**. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4.0 b10.

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. **1996**. Phylogenetic inference. In: Hillis DM, Moritz C, Mable BK, eds. Molecular systematics (2nd ed.). Sinauer Associates, Inc., 407–514.

Taketa S, Ando H, Takeda K, Bothmer R von. **1999**. Detection of *Hordeum marinum* genome in three polyploid *Hordeum* species and cytotypes by genomic in situ hybridization. *Hereditas* **130**: 185–188.

Taketa S, Ando H, Takeda K, Ichii M, Bothmer R von. **2005**. Ancestry of American polyploid *Hordeum* species with the I genome inferred from 5S and 18S-25S rDNA. *Annals of Botany* **96**: 23–33.

Taketa S, Nakauchi Y, Bothmer R von. **2009**. Phylogeny of two tetraploid *Hordeum* species, *H. secalinum* and *H. capense* inferred from physical mapping of 5S and 18S-25S rDNA. *Breeding Science* **59**: 589–594.

Tang H, Bowers JE, Wang X, Paterson AH. 2010. Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proceedings of the National Academy of Sciences* **107:** 472–477.

Tanno K, Bothmer R von, Yamane K, Takeda K, Komatsuda T. **2010**. Analysis of DNA sequence polymorphism at the *cMWG699* locus reveals phylogenetic relationships and allopolyploidy within *Hordeum murinum* subspecies. *Hereditas* **147**: 34–42.

Tavaré S. **1986**. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences* **17**: 57–86.

The International Barley Genome Sequencing Consortium. **2012**. A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**: 711–716.

The International Wheat Genome Sequencing Consortium. **2014**. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345**: 1251788.

Thorne JL, Kishino H, Painter IS. **1998**. Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution* **15**: 1647–1657.

Tiffney BH. **1985**. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. *Journal of the Arnold Arboretum* **66**: 243–273.

Tiffney BH. **2008**. Phylogeography, fossils, and Northern Hemisphere biogeography: the role of physiological uniformitarianism. *Annals of the Missouri Botanical Garden* **95**: 135–143.

Triplett JK, Wang Y, Zhong J, Kellogg EA. **2012**. Five nuclear loci resolve the polyploid history of switchgrass (*Panicum virgatum* L.) and relatives. *PLoS ONE* **7**: e38702.

Ullmann A, Jacob F, Monod J. **1967**. Characterization by *in vitro* complementation of a peptide corresponding to an operator-proximal segment of the β-galactosidase structural gene of *Escherichia coli*. *Journal of Molecular Biology* **24**: 339–343.

Valente LM, Savolainen V, Vargas P. **2010**. Unparalleled rates of species diversification in Europe. *Proceedings of the Royal Society B: Biological Sciences* **277**: 1489–1496.

Van Bel M, Proost S, Wischnitzki E, Movahedi S, Scheerlinck C, Van de Peer Y, Vandepoele K. **2012**. Dissecting plant genomes with the PLAZA comparative genomics platform. *Plant Physiology* **158**: 590–600.

Vanneste K, Maere S, Van de Peer Y. **2014**. Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**: 20130353.

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M. **2007**. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE* **2**: e1326.

Vicente-Carbajosa J, Oñate L, Lara P, Diaz I, Carbonero P. **1998**. Barley *BLZ1*: a bZIP transcriptional activator that interacts with endosperm-specific gene promoters. *The Plant Journal* **13**: 629–640.

Vos P, Hogers R, Bleeker M, Reijans M, Lee T van de, Hornes M, Friters A, Pot J, Paleman J, Kuiper M, *et al*. **1995**. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**: 4407–4414.

Wagner A, Blackstone N, Cartwright P, Dick M, Misof B, Snow P, Wagner GP, Bartels J, Murtha M, Pendleton J. **1994**. Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. *Systematic Biology* **43**: 250–261.

Wagner WHJ. **1970**. Biosystematics and evolutionary noise. *Taxon* **19**: 146–151.

Wakeley J, Hey J. **1997**. Estimating ancestral population parameters. *Genetics* **145**: 847–855.

Wall JD. **2003**. Estimating ancestral population sizes and divergence times. *Genetics* **163**: 395–404.

Wang H, Sun G. **2011**. Molecular phylogeny and reticulate origins of several American polyploid *Hordeum* species. *Botany* **89**: 405–415.

Weiss-Schneeweiss H, Emadzade K, Jang T-S, Schneeweiss GM. **2013**. Evolutionary consequences, constraints and potential of polyploidy in plants. *Cytogenetic and Genome Research* **140**: 137–150.

Wendel JF. **2000**. Genome evolution in polyploids. In: Doyle JJ, Gaut BS, eds. Plant Molecular Evolution. Springer Netherlands, 225–249.

Wendel JF, Schnabel A, Seelanan T. **1995**. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Sciences of the United States of America* **92**: 280–284.

Wendler N, Mascher M, Nöh C, Himmelbach A, Scholz U, Ruge-Wehling B, Stein N. **2014**. Unlocking the secondary gene-pool of barley with next-generation sequencing. *Plant Biotechnology Journal*: 1122–1131.

White T, Bruns T, Lee S, Taylor J. **1990**. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis M, Gelfand D, Shinsky J, White T, eds. PCR Protocols: a Guide to Methods and Applications. New York, NY, USA: Academic Press, 315–322.

Wiley EO, Lieberman BS. **2011**. *Phylogenetics: theory and practice of phylogenetic systematics*. Hoboken, NJ, USA: John Wiley & Sons.

Winkler H. **1917**. Über die experimentelle Erzeugung von Pflanzen mit abweichenden Chromosomenzahlen. *Zeitschrift für Induktive Abstammungs- und Vererbungslehre* **17:** 270–272.

Wolfe KH. **2001**. Yesterday's polyploids and the mystery of diploidization. *Nature Reviews Genetics* **2**: 333–341.

Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH. **1989**. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 6201–6205.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. **2009**. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 13875–13879.

Xi Z, Liu L, Rest JS, Davis CC. **2014**. Coalescent versus concatenation methods and the placement of *Amborella* as sister to water lilies. *Systematic Biology* **63**: 919–932.

Xu J, Kasha KJ. **1992**. Transfer of a dominant gene for powdery mildew resistance and DNA from *Hordeum bulbosum* into cultivated barley (*H. vulgare*). *Theoretical and Applied Genetics* **84**: 771–777.

Yang Z. **1993**. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Molecular Biology and Evolution* **10**: 1396.

Yang Z. **1994**. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution* **39**: 306–314.

Yoder AD, Yang Z. **2000**. Estimation of primate speciation dates using local molecular clocks. *Molecular Biology and Evolution* **17**: 1081–1090.

Zhou L, Yang C, Xiao L. **2003**. PCR-mediated recombination between *Cryptosporidium* spp. of lizards and snakes. *Journal of Eukaryotic Microbiology* **50**: 563–565.

Zohary D, Hopf M, Weiss E. **2012**. *Domestication of plants in the Old World: the origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean basin*. New York, NY, USA: Oxford University Press.

Zuckerkandl E, Pauling L. **1965**. Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel H, eds. Evolving Genes and Proteins. Academic Press, 97–165.

Zwickl DJ, Hillis DM. **2002**. Increased taxon sampling greatly reduces phylogenetic error. *Systematic Biology* **51**: 588–598.

Zwickl DJ, Stein JC, Wing RA, Ware D, Sanderson MJ. **2014**. Disentangling methodological and biological sources of gene tree discordance on *Oryza* (Poaceae) chromosome 3. *Systematic Biology* **63**: 645–659.

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor Dr. Frank Blattner for giving me the chance to work on this subject, for his support, his confidence, his enthusiasm and his patience. He was always of a great help and often suggesting ways to go on with my work.

I am also thankful to Prof. Dr. Martin Röser at the Martin-Luther-Universität Halle-Wittenberg for agreeing to supervise my thesis.

I would like to thank all the members of the ETX working group for the relaxed and interesting working atmosphere. My friends and colleagues Nadine Bernhardt, Katja Herrmann, Dr. Petra Sarhanova, Dr. Dörte Harpke and all the others who were visiting during those nearly five years. But I should not forget the people who were there in the working group when I started and the IPK community: Dr. Sabine Jakob, who introduced me to the work with *Hordeum* but unfortunately left too soon, Dr. Nicolai Nürk, for interesting and challenging discussions and also for introducing me to the PhD Student Board, and Dr. Solomon Benor, always very quiet and kind but so interesting. I am also especially grateful to Dr. Christina Baier, who took care of me "socially". I would like also to thank the precious help in the lab from Petra Oswald, Birgit Kraenzlin and Brunhilde Wedemeier, as well as in the green house where was my plants taken care of by Christina Koch and Ina Faustman.

I am also indebted to Nadine Bernhardt, Katja Herrmann and Dr. Hart Poskar without whom this thesis would not have looked like this.

My gratitude goes also to the PhD Student Board of the IPK for the organization of workshops, study trips, conferences and other events.

I am grateful to all the people who made my time at IPK enjoyable between the evenings and parties at the Club and the various travels. I am thinking to Nadine Bernhardt, Katja Herrmann, Dr. Petra Sarhanova, Marco Pellino, Jana Ebersbach, Martin Mau, Tiina Liiving, Dr. Hart Poskar, Celia Maria Municio, Dr Mian Abdur Rehman Arif, Dr. Takayoshi Ishii, Quddoos ul Haq Muqaddasi and probably many others…

Finally, I am grateful to my parents, my brothers and all my family for their support and encouragements during all my studies.

ABBREVIATIONS

AFLP	=	amplified fragment length polymorphism
AIC	=	Akaike information criterion
BCA	=	Bayesian concordance analysis
BI	=	Bayesian inference
BIC	=	Bayesian information criterion
BLAST	=	basic local alignment search tool
bp	=	basepair
CF	=	concordance factor
cf.	=	*confer*, see
CI	=	credibility interval
DNA	=	deoxyribonucleic acid
e.g.	=	*exempli gratia*, for example
EMBL	=	European Molecular Biology Laboratory
emPCR	=	emulsion-PCR
ESS	=	effective sample size
et al.	=	*et alii*, and others
FISH	=	fluorescent *in situ* hybridization
GISH	=	genomic *in situ* hybridization
GTR	=	generalised time reversible
HKY	=	Hasegawa-Kishino-Yano
hLRT	=	hierarchical likelihood ratio test
HPD	=	highest posterior density
i.e.	=	*id est*, that is
ILS	=	incomplete lineage sorting
ITS	=	internal transcribed spacer
JC	=	Jukes-Cantor
kb, Mb	=	kilobase, megabase
Ma	=	million years ago
MCC	=	maximum clade credibility
MCMC	=	Monte Carlo Markov chain
MCMCMC or MC^3	=	Metropolis coupled Monte Carlo Markov chain
MITE	=	miniature inverted-repeat transposable element
ML	=	maximum likelihood
MLMSC	=	multilocus multispecies coalescent
MP	=	maximum parsimony
NCBI	=	National Center for Biotechnology Information
NGS	=	next-generation-sequencing
NJ	=	neighbor joining
nrDNA	=	nuclear ribosomal DNA
PCR	=	polymerase chain reaction
PLUG	=	PCR landmark unique gene
pp	=	posterior probability
SD	=	standard deviation

smPCR	=	single molecule PCR
SNP	=	single nucleotide polymorphism
sp, spp	=	species
TBR	=	tree bisection-reconnection
UPGMA	=	unweighted pair group method with arithmetic mean
viz.	=	*videlicet*, namely
vs	=	*versus*, against

FIGURES

TABLES

SUPPLEMENTARY INFORMATION

Table S2.1 Detailed information for all individuals analyzed in this study.

Species	Ploidy	Accession Number	Country	Data Source	*TOPO6* haplotype
Hordeum arizonicum Covas	6x	BCC2054 (H02313)	USA	Barley Core Collection	*H. arizonicum* BCC2054_A, B, C
		BCC2060 (H03253)	USA	Barley Core Collection	*H. arizonicum* BCC2060_A, B
		H02144 (NGB90553)	Mexico	Nordic Genetic Resource Centre	*H. arizonicum* H02144_A, B, C
Hordeum bogdanii Wilensky	2x	BCC2070 (H07804)	China	Barley Core Collection	*H. bogdanii* BCC2070
		H07065 (NGB06797)	China	Nordic Genetic Resource Centre	
		H07421b (NGB07634)	China	Nordic Genetic Resource Centre	
		H07429 (NGB06893)	China	Nordic Genetic Resource Centre	
		H07436 (NGB90003)	China	Nordic Genetic Resource Centre	
		H07461 (NGB07281)	China	Nordic Genetic Resource Centre	
		H07465 (NGB90048)	China	Nordic Genetic Resource Centre	
		H07476 (NGB08522)	China	Nordic Genetic Resource Centre	
		H07485b (NGB07282)	China	Nordic Genetic Resource Centre	
		H07569 (NGB08525)	China	Nordic Genetic Resource Centre	
		H07752 (NGB90004)	China	Nordic Genetic Resource Centre	
		H07831 (NGB90006)	China	Nordic Genetic Resource Centre	
		H07859 (NGB90007)	China	Nordic Genetic Resource Centre	
		H08704 (NGB90232)	China	Nordic Genetic Resource Centre	
		H08774 (NGB90049)	China	Nordic Genetic Resource Centre	
		H09215 (NGB90548)	China	Nordic Genetic Resource Centre	
		H03084 (NGB90002)	China	Nordic Genetic Resource Centre	
		PI499499	China	GRIN USDA Aberdeen, USA	
		PI499500	China	GRIN USDA Aberdeen, USA	
		PI531760	China	GRIN USDA Aberdeen, USA	
		BCC2063 (H04014)	Pakistan	Barley Core Collection	*H. bogdanii* BCC2063
		GRA0969	Pakistan	Genbank GTL	
		H00240 (NGB08519)	Afghanistan	Nordic Genetic Resource Centre	
		H00295 (NGB06440)	Pakistan	Nordic Genetic Resource Centre	
		H07067 (NGB06798)	China	Nordic Genetic Resource Centre	
		H07411 (NGB90229)	China	Nordic Genetic Resource Centre	
		H08769 (NGB90337)	China	Nordic Genetic Resource Centre	
		H09213 (NGB90547)	China	Nordic Genetic Resource Centre	
		PI499501	China	GRIN USDA Aberdeen, USA	
		PI499646	China	GRIN USDA Aberdeen, USA	
		H08700 (NGB90231)	China	Nordic Genetic Resource Centre	
		H00287 (NGB6439)	Pakistan	Nordic Genetic Resource Centre	
		H07498 (NGB08523)	China	Nordic Genetic Resource Centre	*H. bogdanii* H07498
		H07521 (NGB90047)	China	Nordic Genetic Resource Centre	
		H07557 (NGB08524)	China	Nordic Genetic Resource Centre	
		H07585 (NGB07283)	China	Nordic Genetic Resource Centre	
		H07727 (NGB90230)	China	Nordic Genetic Resource Centre	
		PI499498	China	GRIN USDA Aberdeen, USA	
		PI531761	China	GRIN USDA Aberdeen, USA	
		H07202 (NGB90385)	China	Nordic Genetic Resource Centre	*H. bogdanii* H07202
		H07416a (NGB08586)	China	Nordic Genetic Resource Centre	*H. bogdanii* H07416a
Hordeum brachyantherum	4x	BCC2045 (H01958)	USA	Barley Core Collection	*H. brachyantherum* 4x BCC2045_A, B

Nevski subsp. *brachyantherum*		BCC2050 (H02138)	USA	Barley Core Collection	*H. brachyantherum* 4x BCC2050_A, B
		BCC2056 (H02360)	Canada	Barley Core Collection	*H. brachyantherum* 4x BCC2056_A, B
	6x	BCC2046 (H02001, GRA0968)	USA	Barley Core Collection	*H. brachyantherum* 6x BCC2046_A, B, C
		GRA0966/87	USA	Genbank GTL	*H. brachyantherum* 6x GRA0966_87_A, B, C
Hordeum brevisubulatum (Trin.) Link	2x	PI229753	Iran	GRIN USDA Aberdeen, USA	*H. brevisubulatum* 2x PI229753_A, B
		PI401374	Iran	GRIN USDA Aberdeen, USA	*H. brevisubulatum* 2x PI401374_A, B
		PI401390	Iran	GRIN USDA Aberdeen, USA	*H. brevisubulatum* 2x PI401390
		H00315	Iran	Nordic Genetic Resource Centre	*H. brevisubulatum* 2x H00315
		PI440419	Russia	GRIN USDA Aberdeen, USA	*H. brevisubulatum* 2x PI440419
	4x	H10239 (NGB90440)	Tajikistan	Nordic Genetic Resource Centre	*H. brevisubulatum* 4x H10239_A, B
		BG156/07	Russia	Barley Core Collection	*H. brevisubulatum* 4x BG156_07_A, B
		PI401387	Iran	GRIN USDA Aberdeen, USA	*H. brevisubulatum* 4x PI401387
		PI401388	Iran	GRIN USDA Aberdeen, USA	*H. brevisubulatum* 4x PI401388_A, Abis
		PI229449	Iran	GRIN USDA Aberdeen, USA	*H. brevisubulatum* 4x PI229449
		PI531771	Kyrgyzstan	GRIN USDA Aberdeen, USA	*H. brevisubulatum* 4x PI531771_A, Abis, B
		GRA2230/97	Russia	Genbank GTL	*H. brevisubulatum* 4x GRA2230_97_A, B
		GRA0894/97	Russia	Genbank GTL	*H. brevisubulatum* 4x GRA0894_97_A, B, C
		H00312 (NGB90046)	Iran	Nordic Genetic Resource Centre	*H. brevisubulatum* 4x H00312_A, B
		BCC2016 (H10210)	Tajikistan	Barley Core Collection	*H. brevisubulatum* 6x BCC2016_A, B
	6x	PI401376	Iran	GRIN USDA Aberdeen, USA	*H. brevisubulatum* 6x PI401376_A, B, C
		PI401380	Iran	GRIN USDA Aberdeen, USA	*H. brevisubulatum* 6x PI401380_A, Abis, B
		PI531768	Tajikistan	GRIN USDA Aberdeen, USA	*H. brevisubulatum* 6x PI531768_A, B
Hordeum bulbosum L.	2x	BCC2061 (H03878)	Italy	Barley Core Collection	*H. bulbosum* 2x BCC2061_A, B
		JB138_1	Italy	Sabine Jakob, Frank Blattner	*H. bulbosum* 2x JB138_1_A, B
		JB145_2	Italy	Sabine Jakob, Frank Blattner	*H. bulbosum* 2x JB145_2_A, B
		JB149_1	Italy	Sabine Jakob, Frank Blattner	*H. bulbosum* 2x JB149_1_A, B
		F2142	Uzbekistan	Reinhard Fritsch	*H. bulbosum* 2x F2142
	4x	F2210	Iran	Reinhard Fritsch	*H. bulbosum* 4x F2210_A, B
		F2227	Iran	Reinhard Fritsch	*H. bulbosum* 4x F2227_A, B
		BCC2018 (H10298)	Tajikistan	Barley Core Collection	*H. bulbosum* 4x BCC2018_A, B
Hordeum californicum Covas & Stebbins	2x	H02419 (NGB90225)	USA	Nordic Genetic Resource Centre	*H. californicum* H02419
		H02428 (NGB90227)	USA	Nordic Genetic Resource Centre	
		H02445 (NGB90228)	USA	Nordic Genetic Resource Centre	
		H02084 (NGB90009)	USA	Nordic Genetic Resource Centre	*H. californicum* H02084
		H02408 (NGB90223)	USA	Nordic Genetic Resource Centre	
		H01942 (NGB06800)	USA	Nordic Genetic Resource Centre	*H. californicum* H01942
		H02423 (NGB90226)	USA	Nordic Genetic Resource Centre	
		H02012 (NGB06462)	USA	Nordic Genetic Resource Centre	*H. californicum* H02012
		H02414 (NGB90224)	USA	Nordic Genetic Resource Centre	
		H01957 (NGB06802)	USA	Nordic Genetic Resource Centre	*H. californicum* H01957
		BCC2057 (H02401)	USA	Barley Core Collection	*H. californicum* BCC2057
		BCC2058 (H02428)	USA	Barley Core Collection	*H. californicum* BCC2058
		H01951 (NGB06801)	USA	Nordic Genetic Resource Centre	*H. californicum* H01951
Hordeum capense Thunb.	4x	BCC2062 (H03923)	Republic South Africa	Barley Core Collection	*H. capense* BCC2062_A, B
		H00335 (NGB90596)	Lesotho	Nordic Genetic Resource Centre	*H. capense* H00335_A, B
Hordeum chilense Roemer & Schultes	2x	Hch204	Chile	Pilar Hernandez	*H. chilense* Hch204
		Hch207	Chile	Pilar Hernandez	
		Hch210	Chile	Pilar Hernandez	
		Hch212	Chile	Pilar Hernandez	
		Hch290	Chile	Pilar Hernandez	
		Hch298	Chile	Pilar Hernandez	

		Hch308	Chile	Pilar Hernandez	
		JB006A_2	Argentina	Sabine Jakob, Frank Blattner	
		JB006A_4	Argentina	Sabine Jakob, Frank Blattner	
		JB006C_12	Argentina	Sabine Jakob, Frank Blattner	
		JB006C_13	Argentina	Sabine Jakob, Frank Blattner	
		Hch017	Chile	Pilar Hernandez	*H. chilense* Hch017
		Hch068	Chile	Pilar Hernandez	
		Hch245	Chile	Pilar Hernandez	
		Hch250	Chile	Pilar Hernandez	
		Hch261	Chile	Pilar Hernandez	
		Hch300	Chile	Pilar Hernandez	
		CN27413	Argentina	AGR, Saskatoon, Canada	*H. chilense* CN27413
		Hch251	Chile	Pilar Hernandez	
		Hch016	Chile	Pilar Hernandez	*H. chilense* Hch016
		Hch225	Chile	Pilar Hernandez	*H. chilense* Hch225
		Hch008	Chile	Pilar Hernandez	*H. chilense* Hch008
Hordeum comosum Presl	2x	JB008_3	Argentina	Sabine Jakob, Frank Blattner	*H. comosum* JB008_3
		JB307_1	Argentina	Sabine Jakob, Frank Blattner	
		JB316A_1	Argentina	Sabine Jakob, Frank Blattner	
		JB316A_3	Argentina	Sabine Jakob, Frank Blattner	
		JB412_1	Argentina	Sabine Jakob, Frank Blattner	
		JB413_1	Argentina	Sabine Jakob, Frank Blattner	
		JB415_2	Argentina	Sabine Jakob, Frank Blattner	
		JB418_1	Argentina	Sabine Jakob, Frank Blattner	
		JB421B_4	Argentina	Sabine Jakob, Frank Blattner	
		JB468_1	Argentina	Sabine Jakob, Frank Blattner	
		JB015D_1	Argentina	Sabine Jakob, Frank Blattner	*H. comosum* JB015D_1
		JB053C_3	Argentina	Sabine Jakob, Frank Blattner	
		JB236A_2	Argentina	Sabine Jakob, Frank Blattner	
		JB281A_1	Argentina	Sabine Jakob, Frank Blattner	
		JB410_1	Argentina	Sabine Jakob, Frank Blattner	
		JB447B_2	Argentina	Sabine Jakob, Frank Blattner	
		JB448_3	Argentina	Sabine Jakob, Frank Blattner	
		JB467_3	Argentina	Sabine Jakob, Frank Blattner	
		JB402_2	Argentina	Sabine Jakob, Frank Blattner	*H. comosum* JB402_2
		JB402_5	Argentina	Sabine Jakob, Frank Blattner	
		JB022B_4	Argentina	Sabine Jakob, Frank Blattner	*H. comosum* JB022B_4
		JB240A_2	Argentina	Sabine Jakob, Frank Blattner	*H. comosum* JB240A_2
		JB244_1	Argentina	Sabine Jakob, Frank Blattner	*H. comosum* JB244_1
		JB424_1	Argentina	Sabine Jakob, Frank Blattner	*H. comosum* JB424_1
Hordeum cordobense Bothmer, Jacobsen & Nicora	2x	BCC2039 (H01702, GRA0974)	Argentina	Barley Core Collection	*H. cordobense* BCC2039
		BCC2067 (H06429)	Argentina	Barley Core Collection	
		JB247A_1	Argentina	Sabine Jakob, Frank Blattner	
		JB247A_3	Argentina	Sabine Jakob, Frank Blattner	
		JB249_1	Argentina	Sabine Jakob, Frank Blattner	
		JB249_2	Argentina	Sabine Jakob, Frank Blattner	
		JB249_3	Argentina	Sabine Jakob, Frank Blattner	
		JB253A_1	Argentina	Sabine Jakob, Frank Blattner	
		JB253A_2	Argentina	Sabine Jakob, Frank Blattner	
		JB253A_3	Argentina	Sabine Jakob, Frank Blattner	
		JB262_3	Argentina	Sabine Jakob, Frank Blattner	
		JB263_1	Argentina	Sabine Jakob, Frank Blattner	

		JB263_2	Argentina	Sabine Jakob, Frank Blattner	
		JB265_1	Argentina	Sabine Jakob, Frank Blattner	
		JB255A_1	Argentina	Sabine Jakob, Frank Blattner	*H. cordobense* JB225A_1
		JB255A_2	Argentina	Sabine Jakob, Frank Blattner	
		JB255A_3	Argentina	Sabine Jakob, Frank Blattner	
		JB257A_1	Argentina	Sabine Jakob, Frank Blattner	
		JB261A_1	Argentina	Sabine Jakob, Frank Blattner	*H. cordobense* JB261A_1
		JB263_3	Argentina	Sabine Jakob, Frank Blattner	
		JB262_1	Argentina	Sabine Jakob, Frank Blattner	*H. cordobense* JB262_1
		JB262_2	Argentina	Sabine Jakob, Frank Blattner	
Hordeum depressum (Scribn. & J. G. Sm.) Rydb.	4x	BCC2047 (H02006)	USA	Barley Core Collection	*H. depressum* BCC2047_A, B
		BCC2052 (H02306)	USA	Barley Core Collection	*H. depressum* BCC2052_A, B
		CN27862	USA	AGR, Saskatoon, Canada	*H. depressum* CN27862_A, Abis, B
		H02304 (NGB06810)	USA	Nordic Genetic Resource Centre	*H. depressum* H02304_A, B
Hordeum erectifolium Bothmer, Jacobsen & Jørgensen	2x	BCC2026 (H01150)	Argentina	Barley Core Collection	*H. erectifolium* BCC2026
Hordeum euclaston Steud.	2x	BCC2022 (H01115)	Argentina	Barley Core Collection	*H. euclaston* BCC2022
		BCC2029 (H01263)	Argentina	Barley Core Collection	
		CN27340	Argentina	AGR, Saskatoon, Canada	
		CN27343	Argentina	AGR, Saskatoon, Canada	
		CN27359	Argentina	AGR, Saskatoon, Canada	
		CN27369	Argentina	AGR, Saskatoon, Canada	
		H01103 (NGB08534)	Argentina	Nordic Genetic Resource Centre	
		H01107 (NGB06465)	Argentina	Nordic Genetic Resource Centre	
		H02148 (NGB90233)	Uruguay	Nordic Genetic Resource Centre	
		H06045 (NGB90342)	Argentina	Nordic Genetic Resource Centre	
		JB225A	Argentina	Sabine Jakob, Frank Blattner	
		JB228A_1	Argentina	Sabine Jakob, Frank Blattner	
		JB228A_6	Argentina	Sabine Jakob, Frank Blattner	
		H01132 (NGB07289)	Argentina	Nordic Genetic Resource Centre	*H. euclaston* H01132
H. flexuosum Steud.		BCC2023 (H01133, NGB06470)	Argentina	Nordic Genetic Resource Centre	*H. flexuosum* BCC2023
		CN27346	Argentina	AGR, Saskatoon, Canada	
		CN27358	Argentina	AGR, Saskatoon, Canada	
		JB208A_2	Argentina	Sabine Jakob, Frank Blattner	
		JB208A_3	Argentina	Sabine Jakob, Frank Blattner	
		JB208A_1	Argentina	Sabine Jakob, Frank Blattner	*H. flexuosum* JB208A_1
		CN27348	Argentina	AGR, Saskatoon, Canada	
		CN27399	Argentina	AGR, Saskatoon, Canada	
Hordeum fuegianum Bothmer, Jacobsen & Jørgensen	4x	BCC2034 (H01371)	Argentina	Barley Core Collection	*H. fuegianum* BCC2034_A, B
		BCC2035 (H01422)	Argentina	Barley Core Collection	*H. fuegianum* BCC2035_A, B
		H02179 (NGB90017)	Chile	Nordic Genetic Resource Centre	*H. fuegianum* H02179_A, B
Hordeum guatemalense Bothmer, Jacobsen & Jørgensen	4x	H02299 (NGB90554)	Guatemala	Nordic Genetic Resource Centre	*H. guatemalense* H02299_A, B
Hordeum intercedens Nevski	2x	BCC2044 (H01940, GRA979)	USA	Barley Core Collection	*H. intercedens* BCC2044
		BCC2059 (H03252)	Mexico	Barley Core Collection	
		CN28638	Mexico	AGR, Saskatoon, Canada	
		CN28639	Mexico	AGR, Saskatoon, Canada	
		CN28640	Mexico	AGR, Saskatoon, Canada	
		CN28643	Mexico	AGR, Saskatoon, Canada	
		CN28644	Mexico	AGR, Saskatoon, Canada	
Hordeum	4x	BCC2048 (H02018)	Mexico	Barley Core Collection	*H. jubatum* BCC2048_A, B

jubatum L.		BCC2055 (H02324)	USA	Barley Core Collection	*H. jubatum* BCC2055_A, B
		JB064	Germany	Sabine Jakob, Frank Blattner	*H. jubatum* JB064_A, B
		FB016	Argentina	Frank Blattner	*H. jubatum* FB016_A, B
Hordeum lechleri (Steud.) Schenck	6x	BCC2027 (H01185)	Argentina	Barley Core Collection	*H. lechleri* BCC2027_B, C
		BCC2030 (H01310)	Argentina	Barley Core Collection	*H. lechleri* BCC2030_A, B, C
		JB044A_1	Argentina	Sabine Jakob, Frank Blattner	*H. lechleri* JB044A1_A, B
		JB048A_4	Argentina	Sabine Jakob, Frank Blattner	*H. lechleri* JB048A4_A, B, C
		JB045B_1a	Argentina	Sabine Jakob, Frank Blattner	*H. lechleri* JB045B1a_A, B
		JB048A_6	Argentina	Sabine Jakob, Frank Blattner	*H. lechleri* JB048A6_A, B
		JB048A_1	Argentina	Sabine Jakob, Frank Blattner	*H. lechleri* JB048A1_A, B, C
Hordeum gussoneanum Parl.	2x	BCC2005 (H00539, NGB7294)	Spain	Barley Core Collection	*H. gussoneanum* 2x BCC2005
		JB106_6	France	Sabine Jakob, Frank Blattner	*H. gussoneanum* 2x JB106_6
		JB113_1	Italy	Sabine Jakob, Frank Blattner	*H. gussoneanum* 2x JB113_1
		JB157	USA	Sabine Jakob, Frank Blattner	*H. gussoneanum* 2x JB157
	4x	H00064 (NGB06507)	Tajikistan	Nordic Genetic Resource Centre	*H. gussoneanum* 4x H00064_A, B
		H00818 (NGB90241)	Iran	Nordic Genetic Resource Centre	*H. gussoneanum* 4x H00818_A, B
		BCC2011 (H00821, NGB6521)	Turkey	Barley Core Collection	*H. gussoneanum* 4x BCC2011_A, B
		BCC2013 (H00824, NGB6522)	Iran	Barley Core Collection	*H. gussoneanum* 4x BCC2013_A, B
		01C0509096_2	Georgia	Res. Inst. Crop Production Prague, Czech Republik	*H. gussoneanum* 4x 01C96_A, B
		H00081 (NGB06509)	Afghanistan	Nordic Genetic Resource Centre	*H. gussoneanum* 4x H00081_A, B
Hordeum marinum Huds. subsp. *marinum*	2x	BCC2006 (H00546, NGB90345)	Spain	Barley Core Collection	*H. marinum* BCC2006
		BCC2001 (H00090, NGB6821)	Greece	Barley Core Collection	*H. marinum* BCC2001
		JB069_3	France	Sabine Jakob, Frank Blattner	*H. marinum* JB069_3
Hordeum murinum L. subsp. *glaucum* (Steud.) Tzvelev	2x	PI218078	Pakistan	GRIN USDA Aberdeen, USA	*H. murinum* 2x PI218078_A, B, C
		BCC2002 (H00219)	Tunisia	Barley Core Collection	*H. murinum* 2x BCC2002_A, B, C
		PI223371 (CN064124)	Iran	GRIN USDA Aberdeen, USA	*H. murinum* 2x PI223371_A, B
Hordeum murinum L. subsp. *murinum*	4x	JB097	Germany	Sabine Jakob, Frank Blattner	*H. murinum* 4x JB097_A, B
		BCC2009 (H00721)	Denmark	Barley Core Collection	*H. murinum* 4x BCC2009_A, B, C
		H00217 (NGB06870)	Germany	Nordic Genetic Resource Centre	*H. murinum* 4x H00217_A, B, C
Hordeum murinum L. subsp. *leporinum* (Link) Arcang.	4x	BCC2007 (H00561)	Spain	Barley Core Collection	*H. murinum* 4x BCC2007
		BCC2008 (H00591)	Greece	Barley Core Collection	*H. murinum* 4x BCC2008
		PI244767	Iran	GRIN USDA Aberdeen, USA	*H. murinum* 4x PI244767
		GRA1097	Bulgaria	Genbank GTL	*H. murinum* 4x GRA1097
	6x	F2107	Uzbekistan	Reinhard Fritsch	*H. murinum* 6x F2107
		H00812 (NGB06878)	Turkey	Nordic Genetic Resource Centre	*H. murinum* 6x H00812_A, B, C
		PI211046	Afghanistan	GRIN USDA Aberdeen, USA	*H. murinum* 6x PI211046_A, B
Hordeum muticum J. Presl	2x	BCC2014 (H00958)	Bolivia	Barley Core Collection	*H. muticum* BCC2014
		H01837 (NGB07303)	Argentina	Nordic Genetic Resource Centre	
		H06457a (NGB90357)	Argentina	Nordic Genetic Resource Centre	
		H06459 (NGB90358)	Argentina	Nordic Genetic Resource Centre	
		H06468 (NGB90359)	Argentina	Nordic Genetic Resource Centre	
		H06470 (NGB90360)	Argentina	Nordic Genetic Resource Centre	
		BCC2042 (H01784, GRA982)	Argentina	Barley Core Collection	*H. muticum* BCC2042
		H06446 (NGB90060)	Argentina	Nordic Genetic Resource Centre	
		H00955 (NGB15689)	Bolivia	Nordic Genetic Resource Centre	*H. muticum* H00955
		H00957 (NGB08566)	Bolivia	Nordic Genetic Resource Centre	*H. muticum* H00957
Hordeum parodii Covas	6x	BCC2025 (H01146)	Argentina	Barley Core Collection	*H. parodii* BCC2025_A, B, C
		BCC2066 (H06328)	Argentina	Barley Core Collection	*H. parodii* BCC2066_A, B, C
		JB004_1	Argentina	Sabine Jakob, Frank Blattner	*H. parodii* JB004_1_A, B
		JB019A_2	Argentina	Sabine Jakob, Frank Blattner	*H. parodii* JB019A2_A, B, C
Hordeum patagonicum (Haumann) Covas	2x	JB044Bd	Argentina	Sabine Jakob, Frank Blattner	*H. patagonicum* JB044Bd
		JB319A_1	Argentina	Sabine Jakob, Frank Blattner	
		JB024_1	Argentina	Sabine Jakob, Frank Blattner	*H. patagonicum* JB024_1

		JB306C_1	Argentina	Sabine Jakob, Frank Blattner	
		BCC2064 (H06051)	Argentina	Barley Core Collection	*H. patagonicum* BCC2064
		JB287C_3	Argentina	Sabine Jakob, Frank Blattner	
		JB303B_1	Argentina	Sabine Jakob, Frank Blattner	*H. patagonicum* JB303B_1
		JB303B_4	Argentina	Sabine Jakob, Frank Blattner	
		H01248	Argentina	Nordic Genetic Resource Centre	*H. patagonicum* H01248
		H01535	Argentina	Nordic Genetic Resource Centre	*H. patagonicum* H01535
		JB015C_1	Argentina	Sabine Jakob, Frank Blattner	*H. patagonicum* JB015C_1
		JB318A_4	Argentina	Sabine Jakob, Frank Blattner	*H. patagonicum* JB138A_4
		JB052Ba	Argentina	Sabine Jakob, Frank Blattner	*H. patagonicum* JB052Ba
		BCC2033 (H01358)	Argentina	Barley Core Collection	*H. patagonicum* BCC2033
Hordeum procerum Nevski	6x	BCC2024 (H01136)	Argentina	Barley Core Collection	*H. procerum* BCC2024_A, B, C
		BCC2040 (H01781)	Argentina	Barley Core Collection	*H. procerum* BCC2040_A, B, C
		JB221A_2	Argentina	Sabine Jakob, Frank Blattner	*H. procerum* JB221A2_A, B, C
		JB223A_3	Argentina	Sabine Jakob, Frank Blattner	*H. procerum* JB223A3_A, B, C
Hordeum pubiflorum Hook.	2x	H01238 (NGB08538)	Argentina	Nordic Genetic Resource Centre	*H. publiflorum* H01238
		JB027B_5	Argentina	Sabine Jakob, Frank Blattner	
		JB059B_2a	Argentina	Sabine Jakob, Frank Blattner	
		JB460B_3	Argentina	Sabine Jakob, Frank Blattner	
		JB463B_2	Argentina	Sabine Jakob, Frank Blattner	
		JB463B_4	Argentina	Sabine Jakob, Frank Blattner	
		JB020_2	Argentina	Sabine Jakob, Frank Blattner	*H. publiflorum* JB020_2
		JB033Ba	Argentina	Sabine Jakob, Frank Blattner	
		JB323_4	Argentina	Sabine Jakob, Frank Blattner	
		JB447A_1	Argentina	Sabine Jakob, Frank Blattner	
		JB035C_1	Argentina	Sabine Jakob, Frank Blattner	*H. publiflorum* JB035C_1
		JB312A_1	Argentina	Sabine Jakob, Frank Blattner	
		JB053Ba	Argentina	Sabine Jakob, Frank Blattner	*H. publiflorum* JB035Ba
		JB053Bd	Argentina	Sabine Jakob, Frank Blattner	
		BCC2068 (H06687)	Bolivia	Barley Core Collection	*H. publiflorum* BCC2068
		JB021_5	Argentina	Sabine Jakob, Frank Blattner	*H. publiflorum* JB021_5
		JB291_2	Argentina	Sabine Jakob, Frank Blattner	*H. publiflorum* JB291_2
Hordeum pusillum Nutt.	2x	CN27877	USA	AGR, Saskatoon, Canada	*H. pusillum* CN27877
		Clho15663	USA	GRIN USDA Aberdeen, USA	
		CN27888	USA	AGR, Saskatoon, Canada	
		CN28655	USA	AGR, Saskatoon, Canada	
		CN32799	USA	AGR, Saskatoon, Canada	
		CN32803	USA	AGR, Saskatoon, Canada	
		CN028654	USA	AGR, Saskatoon, Canada	
		BCC2043 (H01906, GRA1176)	USA	Barley Core Collection	*H. pusillum* BCC2043
		CN27810	USA	AGR, Saskatoon, Canada	
		CN27814	USA	AGR, Saskatoon, Canada	
		CN27885	USA	AGR, Saskatoon, Canada	
		CN27886	USA	AGR, Saskatoon, Canada	
		BCC2049 (H02038)	USA	Barley Core Collection	*H. pusillum* BCC2049
Hordeum roshevitzii Bowden	2x	BCC2069 (H07202)	China	Barley Core Collection	*H. roshevitzii* BCC2069
		H07039 (NGB06891)	China	Nordic Genetic Resource Centre	
		H07046 (NGB06892)	China	Nordic Genetic Resource Centre	
		H07437 (NGB08587)	China	Nordic Genetic Resource Centre	
		H07754 (NGB90149)	China	Nordic Genetic Resource Centre	
		H08787 (NGB90068)	China	Nordic Genetic Resource Centre	
		H09152 (NGB90388)	China	Nordic Genetic Resource Centre	
		H09154 (NGB90389)	China	Nordic Genetic Resource Centre	

		H09157 (NGB90390)	China	Nordic Genetic Resource Centre	
		M4341	Mongolia	Karsten Wesche	
		M4352	Mongolia	Karsten Wesche	
		H07421a (NGB90386)	China	Nordic Genetic Resource Centre	*H. roshevitzii* H07421a
		H07879 (NGB90387)	China	Nordic Genetic Resource Centre	
		H10070 (NGB90391)	Russia	Nordic Genetic Resource Centre	
		H07883 (NGB90150)	China	Nordic Genetic Resource Centre	
		H09194 (NGB90621)	China	Nordic Genetic Resource Centre	*H. roshevitzii* H09194
		BCC2015 (H10070)	Russia	Barley Core Collection	*H. roshevitzii* BCC2015
Hordeum secalinum Schreb.	4x	BCC2004 (H00296)	Spain	Barley Core Collection	*H. secalinum* BCC2004_A
		GRA1016	France	Genbank GTL	*H. secalinum* GRA1016_A, B
		JB104_1	Germany	Sabine Jakob, Frank Blattner	*H. secalinum* JB104_1_A, B
		JB105_4	Germany	Sabine Jakob, Frank Blattner	*H. secalinum* JB105_4_A, B
Hordeum stenostachys Godr.	2x	H01780 (NGB90151)	Argentina	Nordic Genetic Resource Centre	*H. stenostachys* H01780
		H06431 (NGB90070)	Argentina	Nordic Genetic Resource Centre	
		H06484 (NGB90072)	Argentina	Nordic Genetic Resource Centre	
		JB216_2	Argentina	Sabine Jakob, Frank Blattner	
		JB217B_2	Argentina	Sabine Jakob, Frank Blattner	
		JB218_5	Argentina	Sabine Jakob, Frank Blattner	
		JB219B_2	Argentina	Sabine Jakob, Frank Blattner	
		JB241_1	Argentina	Sabine Jakob, Frank Blattner	
		JB254_1	Argentina	Sabine Jakob, Frank Blattner	
		JB254_5	Argentina	Sabine Jakob, Frank Blattner	
		JB258_1	Argentina	Sabine Jakob, Frank Blattner	
		JB258_2	Argentina	Sabine Jakob, Frank Blattner	
		JB267_1	Argentina	Sabine Jakob, Frank Blattner	
		BCC2021 (H01108)	Argentina	Barley Core Collection	*H. stenostachys* BCC2021
		CN27345	Argentina	AGR, Saskatoon, Canada	
		JB199A_1	Argentina	Sabine Jakob, Frank Blattner	
		JB211_2	Argentina	Sabine Jakob, Frank Blattner	
		JB213_5	Argentina	Sabine Jakob, Frank Blattner	
		JB214_4	Argentina	Sabine Jakob, Frank Blattner	*H. stenostachys* JB214_4
Hordeum tetraploidum Covas	4x	JB048C_2b	Argentina	Sabine Jakob, Frank Blattner	*H. tetraploidum* JB048C_2b_A, B
		JB010_4	Argentina	Sabine Jakob, Frank Blattner	*H. tetraploidum* JB010_4_A, B, C
		JB026a	Argentina	Sabine Jakob, Frank Blattner	*H. tetraploidum* JB026a_A, B
		JB029B_4	Argentina	Sabine Jakob, Frank Blattner	*H. tetraploidum* JB029B_4_A, B
Hordeum vulgare L. subsp. *spontaneum*	2x	F2097	Uzbekistan	Reinhard Fritsch	*H. vulgare* subsp. *spontaneum* F2097
		F2131	Uzbekistan	Reinhard Fritsch	*H. vulgare* subsp. *spontaneum* F2131_A, B
Dasypyrum villosum (L.) P. Candargy	2x	GRA1020	Italy	Genbank GTL	*Dasypyrum villosum* GRA1020
	2x	GRA 692	Mongolia	Genbank GTL	*Dasypyrum villosum* GRA0692
Eremopyrum triticeum (Gaertn.) Nevski	2x	GRA2250	Kazakhastan	Genbank GTL	*Eremopyrum triticeum* GRA2250
Secale strictum (C.Presl) C.Presl subsp. *kuprijanovii* (Grossh.) K.Hammer	2x	R 1108	Kazakhastan	Genbank GTL	*Secale strictum* R1108
Secale vavilovii Grossh.	2x	R1027	Italy	Genbank GTL	*Secale vavilovii* R1027
Taeniatherum caput-medusae (L.) Nevski	2x	GRA1126	Tajikistan	Genbank GTL	*Taeniatherum caput-medusae* GRA1126
Triticum monococcum L. var. *vulgare* Körn.	2x	TRI13061	Turkey	Genbank GTL	*Triticum monococcum* TRI13061
Triticum urartu Tumanian ex Gandilyan var. *spontaneoalbum* Tumanian	2x	TRI17921 (PI427516)	Turkey	USDA	*Triticum urartu* TRI17921

Figure S2.1 Examples of chimeric *TOPO6* sequences together with non-recombinant sequences derived from a tetraploid individual of *Hordeum tetraploidum* (JB048C2B). The striped alignment shows only polymorphic sites of all 16 clone-derived sequences (JB048C2Bb-q) and the two consensus sequences used in the analysis (*H. tetraploidum* JB048C2B_A and B). Sequences b–h were obtained using standard DNA polymerase and sequences i–q using proof-reading DNA polymerase in PCR. Sequences c, e, f and h show mosaic patterns indicated below the sequences as "A type" or "B type", referring to the sequences used in the analysis. For this individual no chimerical sequences were retrieved with the proof-reading DNA polymerase against four obtained via regular DNA polymerase. Also the amount of singleton SNPs (PCR errors) found in the sequences differs between both polymerases.

Figure S2.2 Strict consensus tree of 50,000 most parsimonious trees (L = 374 steps, CI = 0.84, RI = 0.96) from an analysis of *TOPO6* sequences derived from diploid *Hordeum* taxa. Numbers along branches depict bootstrap values (%) of major clades of the tree derived from a "fast-and-stepwise" of 100,000 bootstrap re-samples. After the species name and individual number, the different copies found per individual are indicated (A-C) in case of cloned sequences.

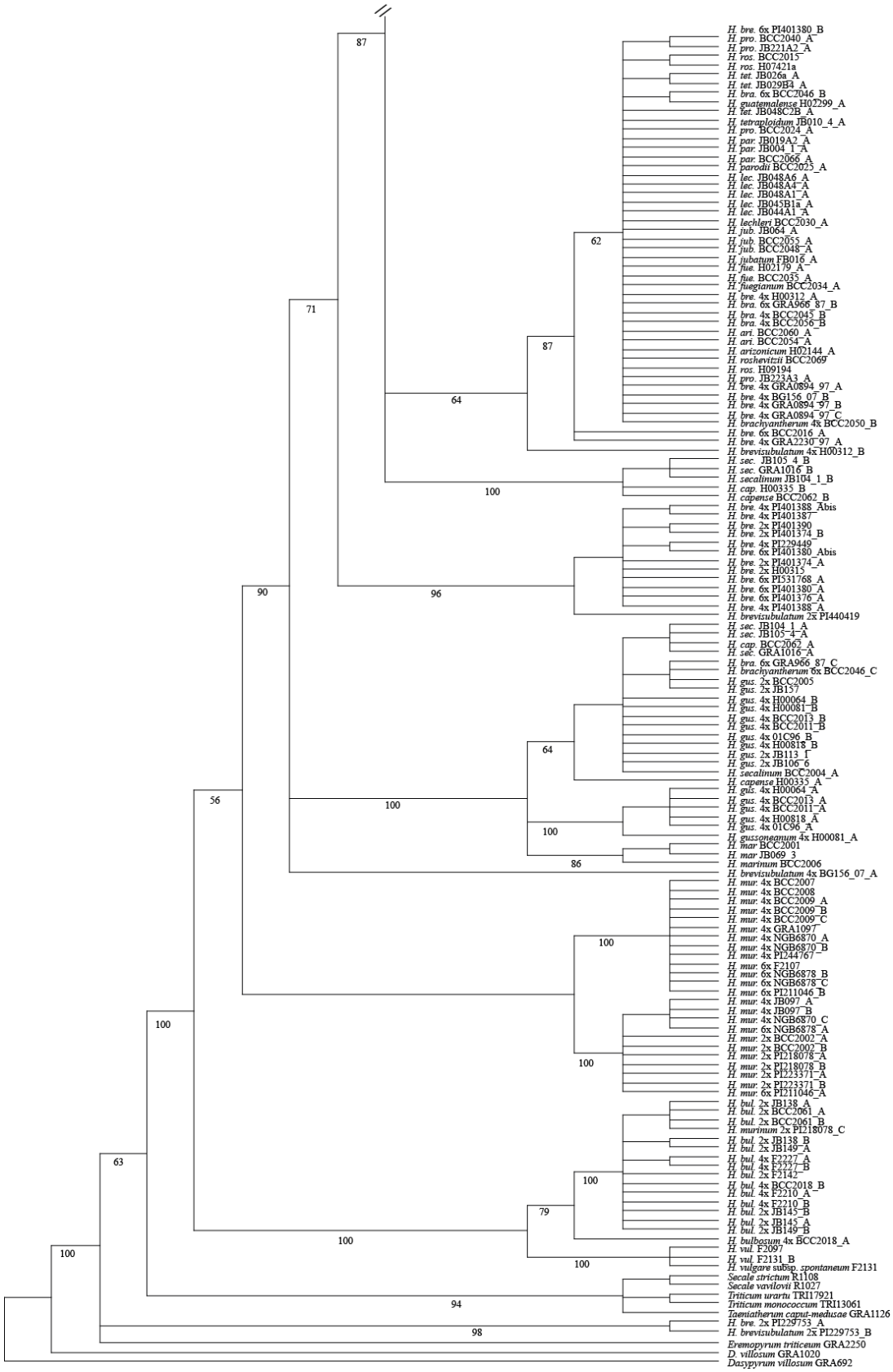


Figure S2.3 Strict consensus tree of 50,000 most parsimonious trees (L=536 steps, CI=0.80, RI=0.97) from an analysis of *TOPO6* sequences derived from di- and polyploid *Hordeum* taxa and eight outgroup species. Numbers along branches depict bootstrap values (%) of major clades of the tree derived from a 'fast-and-stepwise' analysis of 50,000 bootstrap re-samples. *Psathyrostachys juncea* was defined as outgroup taxon in the analysis

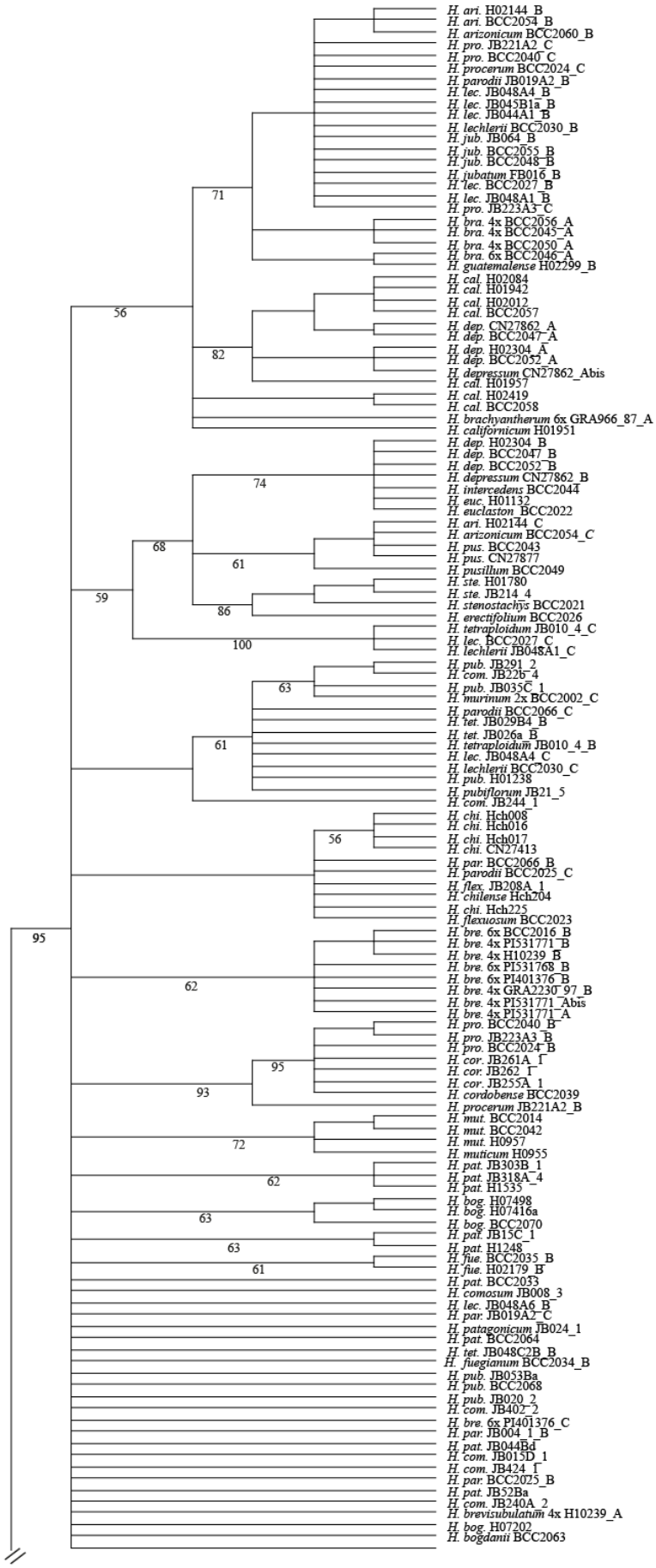


Figure S2.3 *continued*

Figure S2.4 Part of the *TOPO6* alignment showing a 33 bp deletion (alignment positions 925-957) characteristic for sequences derived from diploid New World *Hordeum* species. Sequence deletion occurred at a five basepair direct repeat (TACAC) flanking the deleted region (arrows). The absence of the deletion in *H. bogdanii* together with its presence in all American diploid species indicates that not *H. bogdanii* itself but a close relative of this species was the initial starting point for the colonization of the Americas by Asian *Hordeum*.

Table S3.1 Individuals included in the study and number of reads per locus

Species	Accession Number	Ploidy	*BLZ1*	*MATK*	*NUC*	*TNAC 1035*	*TNAC 1142*	*TNAC 1364*	*TNAC 1403*	*TNAC 1463*	*TNAC 1497*	*TNAC 1610*	*TNAC 1740*	*TOPO6*	*XYL*	Total
454 sequenced individuals																
	BCC2054	6x	566	354	429	954	819	304	620	416	882	151	247	367	712	6821
H. arizonicum	H02144	6x	559	1605	518	362	767	28	2110	1405	828	369	Na	696	781	10028
	BCC2060	6x	988	5026	1091	519	1080	632	2840	2346	1967	2125	949	1052	1475	22090
	BCC2063	2x	271	1422	139	201	268	245	785	504	275	18	183	416	438	5165
H. bogdanii	BCC2070	2x	396	337	321	427	579	853	618	474	983	263	239	281	952	6723
	H07498	2x	255	1549	221	226	420	239	1670	714	608	333	265	368	549	7417
	BCC2050	4x	613	2376	337	192	606	157	1502	788	569	423	Na	408	707	8678
H. brachyantherum subsp. *brachyantherum*	BCC2056	4x	37	805	824	1058	1242	704	3479	3041	2527	623	101	157	949	15547
	BCC2046	6x	288	1144	174	262	304	204	1226	497	447	683	227	844	781	7081
	PI229753	2x	Na	Na	Na	Na	Na	Na	Na	Na	Na	Na	Na	Na	Na	Na
	PI401374	2x	72	266	47	59	50	117	181	179	180	144	37	52	122	1506
	PI401387	4x	85	367	79	33	Na	47	343	346	225	96	Na	92	108	1821
	H00312	4x	139	322	87	58	122	94	361	317	227	138	81	127	187	2260
H. brevisubulatum	GRA2230/97	4x	Na	Na	Na	Na	Na	Na	Na	Na	Na	Na	Na	Na	Na	Na
	GRA0894/97	4x	520	1299	64	285	227	241	1222	658	477	328	257	569	463	6610
	PI401376	6x	758	1369	234	179	242	232	702	1034	422	483	528	544	441	7168
	PI531768	6x	1043	1884	493	72	504	379	Na	2020	881	386	209	72	1225	9168
	BCC2016	6x	894	3985	773	545	1533	564	85	2757	1240	638	1148	886	1606	16654
	BCC2061	2x	204	1183	Na	251	199	223	681	411	380	27	333	196	508	4596
H. bulbosum	BCC2018	4x	392	1389	564	543	1175	688	1611	791	846	361	752	Na	959	10071
	F2227	4x	442	1833	458	488	505	629	1284	1544	909	345	1680	427	1183	11727
	BCC2057	2x	378	1681	64	269	346	381	902	820	398	333	468	516	537	7093
H. californicum	H01951	2x	555	1468	Na	283	336	263	1934	554	758	691	463	307	926	8538
	BCC2058	2x	287	1600	33	344	333	342	1445	824	71	537	498	344	765	7423
H. capense	BCC2062	4x	1308	2748	Na	1626	770	615	1566	2135	758	961	1029	910	2380	16806
	H00335	4x	720	Na	22	513	366	350	1234	1048	802	168	601	Na	1009	6833
H. chilense	JB006A2	2x	1052	3782	460	1009	1851	453	1384	2094	372	767	920	902	1714	16760
	CN27413	2x	537	1131	39	897	528	740	1019	580	320	319	561	Na	285	6956
	JB015D1	2x	798	1960	309	478	317	535	895	2117	433	333	Na	338	670	9183
H. comosum	JB412_1	2x	458	1662	447	1029	306	323	713	1059	278	335	Na	379	508	7497
	JB022B4	2x	234	2118	149	468	174	506	1491	Na	574	310	Na	237	1276	7537
H. cordobense	JB247A1	2x	269	663	135	186	333	162	1065	388	449	Na	196	226	346	5056
	JB255A2	2x	481	882	213	258	390	194	912	355	295	610	326	69	830	5815
H. depressum	BCC2047	4x	1155	2906	384	505	767	389	1826	820	712	545	690	333	1357	12389
	BCC2052	4x	927	3940	491	726	477	503	2015	690	980	1596	930	488	1266	15029
H. erectifolium	BCC2026	2x	870	1445	60	996	1002	1093	2748	758	861	1940	858	70	596	13297
H. euclaston	BCC2029	2x	592	2141	Na	585	727	277	1891	761	1180	1894	535	Na	834	11417
H. flexuosum	BCC2023	2x	635	1405	319	959	988	579	1037	974	1457	513	1019	744	939	11568
	BCC2035	4x	1680	5571	167	771	1247	669	1820	1800	1046	747	949	Na	1690	18157
H. fuegianum	H02179	4x	54	59	Na	12	41	53	Na	83	150	Na	28	Na	142	622
	H01422	4x	905	4760	300	290	760	240	1566	899	1134	552	203	147	1033	12789
H. guatemalense	H02299	4x	434	2632	571	1357	1275	462	2374	1970	591	762	801	792	1070	15091
H. jubatum	JB064	4x	606	1948	162	792	479	362	1558	972	467	412	609	Na	645	9012
	FB016	4x	722	1993	327	551	363	744	1412	1248	409	667	667	448	607	10158
H. lechleri	BCC2027	6x	1146	2886	407	968	642	1424	1699	1427	684	2763	Na	Na	1643	15689
	JB044A1	6x	304	1463	91	806	686	1352	1839	350	828	1625	1379	Na	1121	11844

	JB048A1	6x	659	Na	Na	678	911	820	1198	350	586	1342	767	Na	1042	8353
H. marinum subsp. *gussoneanum*	JB106_6	2x	486	562	Na	240	337	764	470	215	584	160	375	28	761	4982
	BCC2005	2x	279	730	Na	516	60	404	183	165	259	159	520	Na	630	3905
	BCC2013	4x	334	953	Na	735	254	921	994	491	236	232	197	Na	316	5663
	H00081	4x	1136	2221	17	686	280	811	2387	620	662	333	878	24	1936	11991
H. marinum subsp. *marinum*	BCC2001	2x	233	737	Na	373	166	148	685	434	324	146	309	21	Na	3576
	BCC2006	2x	396	1666	Na	507	336	378	729	470	388	120	555	551	341	6437
H. murinum subsp. *glaucum*	PI218078	2x	236	760	Na	574	172	360	823	51	844	480	321	Na	914	5535
	BCC2017	2x	329	1334	Na	931	404	440	1041	106	371	267	343	Na	408	5974
H. murinum subsp. *murinum*	BCC2009	4x	692	2889	533	821	1053	847	Na	336	1240	411	117	Na	981	9920
	H00217	4x	972	2327	944	1234	537	948	2926	408	864	483	642	Na	1603	13888
H. murinum subsp. *leporinum*	BCC2007	4x	437	1847	752	684	325	528	3550	887	804	1638	1002	Na	596	13050
	BCC2008	4x	321	2296	531	401	452	683	1447	774	721	766	21	Na	802	9215
	BCC2010	6x	979	1173	889	165	736	1183	3577	Na	1452	1054	486	Na	1902	13596
H. muticum	H06459	2x	557	1178	149	726	580	471	904	428	1136	1158	301	Na	1489	9077
	H00957	2x	1139	4446	Na	1122	940	1008	3709	2552	941	2671	540	Na	2965	22033
H. parodii	BCC2025	6x	1669	3997	940	2974	1101	2651	6514	7061	3276	2033	2331	Na	2720	37267
	BCC2066	6x	1144	4403	834	1775	1609	1108	4954	5251	1393	2126	1670	156	2359	28782
	JB004_1	6x	1202	8420	1599	1926	1948	1312	5492	4404	1792	3590	990	267	2588	35530
H. patagonicum	JB044Bd	2x	838	3662	432	836	664	1141	1775	2134	841	543	149	1098	823	14936
	JB303B1	2x	717	2905	354	868	500	649	2190	1611	673	723	589	432	712	12923
	JB052Ba	2x	689	2730	469	888	579	421	3171	1275	1043	976	255	348	1070	13914
H. procerum	BCC2024	6x	1949	3099	514	1503	1311	2383	2997	Na	1591	2512	2491	Na	3069	23419
	JB221A2	6x	1191	1574	1091	1131	697	700	1299	747	1386	1109	965	798	1925	14613
H. pubiflorum	JB059B2a	2x	337	2071	322	436	408	296	1185	1219	726	563	625	494	966	9648
	BCC2068	2x	484	2647	198	717	527	392	2010	1233	446	1170	207	406	1341	11778
H. pusillum	CN27877	2x	263	1539	215	289	434	407	687	246	577	909	337	468	596	6967
	BCC2049	2x	697	1702	638	649	363	482	1232	1523	733	391	490	612	682	10194
H. roshevitzii	H09157	2x	600	1740	370	963	496	284	1580	625	575	302	Na	697	966	9198
	H07421a	2x	467	1411	440	834	232	416	1535	1027	512	446	218	907	1025	9470
H. secalinum	BCC2004	4x	589	1855	Na	1357	403	483	1166	Na	910	964	838	Na	627	9192
	JB104_1	4x	972	2837	699	817	645	603	1994	1230	1652	841	959	1072	1385	15706
H. stenostachys	JB218_5	2x	546	1365	Na	1089	603	558	1237	272	937	1217	512	451	714	9501
	JB214_4	2x	412	1147	230	450	310	396	1403	1085	303	703	270	366	573	7648
H. tetraploidum	JB048C2b	4x	641	2511	374	896	666	699	1805	2667	908	1153	532	878	853	14583
	JB010_4	4x	1015	3953	555	1191	791	767	1826	2788	2016	1058	1048	688	1419	19115
	JB026a	4x	2165	7275	3367	2109	1756	1466	6619	4430	1971	3828	2549	1697	2464	41696
	JB029B4	4x	501	3871	Na	1258	728	719	3235	315	656	933	448	998	1371	15033
H. vulgare subsp. *spontaneum*	F2097	2x	702	2345	653	393	Na	1052	825	Na	291	Na	1116	Na	567	7944
	F2131	2x	961	1122	1400	574	Na	358	933	614	376	66	445	134	368	7351
Eremopyrum triticeum	GRA2250	2x	709	2164	472	113	336	708	1937	1361	1068	692	469	105	543	10677
Dasypyrum villosum	GRA692	2x	Na	612	Na	261	136	133	794	865	233	275	150	Na	211	3670
	GRA1020	2x	26	1309	Na	714	367	296	1209	700	1105	397	213	Na	614	6950
Secale vavilovii	R1027	2x	560	2078	Na	476	168	898	2291	2592	2026	893	413	957	807	14159
Taeniatherum caput-medusae	GRA1126	2x	432	969	652	864	227	570	1545	309	861	1288	115	237	643	8712
Triticum monococcum	TRI13061	2x	458	1067	56	494	207	767	2559	626	195	1160	415	12	679	8695
Bromus tectorum	GRA1085	2x	639	1986	Na	395	408	1031	Na	14	869	378	630	Na	423	6773
Cloned individuals																
H. brevisubulatum	H00315	2x														

	PI401390	2x
	PI440419	2x
H. bulbosum	F2142	4x
H. euclaston	BCC2022	2x
	H01132	2x
H. intercedens	BCC2044	2x
	BCC2059	2x

Table S3.2 Primers used in the PCR amplification of each locus. F: forward; R: reverse

Locus	Primer name	Primer Sequence	Annealing temperature	Reference
TNAC1035	TNAC1035F	TGCACTGGGATCTAACCTAAA	56°C	This study
	TNAC1035R	TCCAGTGATCATTTGAAGATTCC		
TNAC1142	TNAC1142F	GCCTACGAGTACATGGTCGAG	57°C	This study
	TNAC1142R	CAGCATCCATAACCAGGATGT		
TNAC1364	TNAC1364F	CGTCAGGCTCAGGGTGTC	58°C	This study
	TNAC1364R	AAAGAGCCTCTGTCTCTCAGG		
TNAC1463	TNAC1463F	CGTCTTTATCAAACCCTGCAA	63°C	This study
	TNAC1463R	GTTCACCGAGTTCATCCAGAA		
TNAC1403	TNAC1403F	CCTCCTCCATTGCGAGATAAC	63°C	This study
	TNAC1403R	GTAGTAACGCTGAAGGGTTCG		
TNAC1610	TNAC1610F	CTGTTGGCAGTTGCAACAAAT	56°C	This study
	TNAC1610R	TCCATAGCAAAGGCCCTAAAT		
TNAC1577	TNAC1577F	CGCTCCATTTCCAGATCATAA	56°C	This study
	TNAC1577R	GCAGCATATAAATGGCCAAAC		
TNAC1497	TNAC1497F	ATCAAACCTGACGGTGTTCAG	58°C	This study
	TNAC1497R	CATGCAGACTACAGGTCCAGA		
TNAC1740	TNAC1740F	CGGAAGTGCTCGATTGTATCT	55/56°C	This study
	TNAC1740R	GCGGGTTTCTTCTCAACCTT		
TNAC1781	TNAC1781F	AACTGGCAATCAGCAGCAC	55/56°C	This study
	TNAC1781R	CACCACGCTCTCTTTCATCTT		
TOPO6	Top6-15F	GTGYTGTSTYCAACTGAAGTC	58/59°C	Jakob & Blattner, 2010
	Top6-17R	CGTACTCCARYGCCATTTC		
XYL	xyl6F	CCTGATGGCAAAACACTCGC	59°C	Pillen *et al.*, 2000
	xyl9R	CAGCATAAGCATACACCTTGACCTC		
NUC	nuc3F	CTCGCCACCGACATCATTTC	59/60°C	Petersen & Seberg, 2009
	nuc6R	GCTTCTTCCCCTTCCAGCATAGAG		
BLZ1	blz3F	GTCCACATCTGGTACAAAATGCTG	59°C	Vicente-Carbajosa *et al.*, 1998
	blz4R	TGTGCTTCGAGTTCATTCAGGTG		
*mat*K	matK-F	AACCCGGAACTAGTCGGATG	60°C	Nishikawa *et al.*, 2002
	matK-R	CTCAATGGTAGAGTACTCGG		
*ndh*F	ndhF-F1	ATGGAACAKACATATSAATATGC	60°C	(Olmstead & Sweere, 1994)
	ndhF2110-R	CCCCCTAYATATTTGATACCTTCTCC		

Table S3.3 Information on individual data sets and setting used in the analyses

Locus	Alignment length	Parsimony-informative sites	Variable sites	Model of evolution [a]	Clock for *BEAST
TNAC1035	714 (759)	123 (184)	147 (250)	HKY+G (1)	Strict
TNAC1142	983 (2055)	238 (387)	310 (524)	HKY+G (2)	Strict
XYL	849 (1285)	138 (307)	199 (448)	HKY+G (1)	Strict
TNAC1364	1055 (1599)	260 (409)	293 (584)	HKY+G (2)	Strict
NUC	847 (910)	99 (176)	128 (268)	K80+G	[b]
TNAC1403	1993 (3443)	268 (367)	327 (562)	HKY+G (3)	Strict
TNAC1463	4651 (5074)	519 (993)	838 (1333)	HKY+G (4)	Strict
BLZ1	1363 (1500)	119 (211)	249 (412)	HKY+G (1)	Strict
TNAC1610	1492 (1742)	104 (236)	183 (379)	HKY+G (3)	Strict
TOPO6	890 (1269)	129 (211)	164 (303)	HKY+G (1)	Strict
TNAC1497	805 (4038)	142 (225)	165 (325)	HKY+G (1)	Strict
TNAC1740	1383 (1511)	207 (309)	254 (465)	HKY+G (1)	Strict
*mat*K	2609 (2625)	95 (100)	203 (210)	HKY+G	[b]
Supermatrix[c]	25382 (24996)	2852 (3633)	4915 (5469)	HKY+G & GTR+G	[b]

Values in brackets correspond to alignments including polyploids
[b]Data set not included in the *BEAST analysis
[c]Data set consisting of all loci except *mat*K and divided in two partitions, one consisting of *TNAC1463* with HKY+G and the rest of the loci with GTR+G

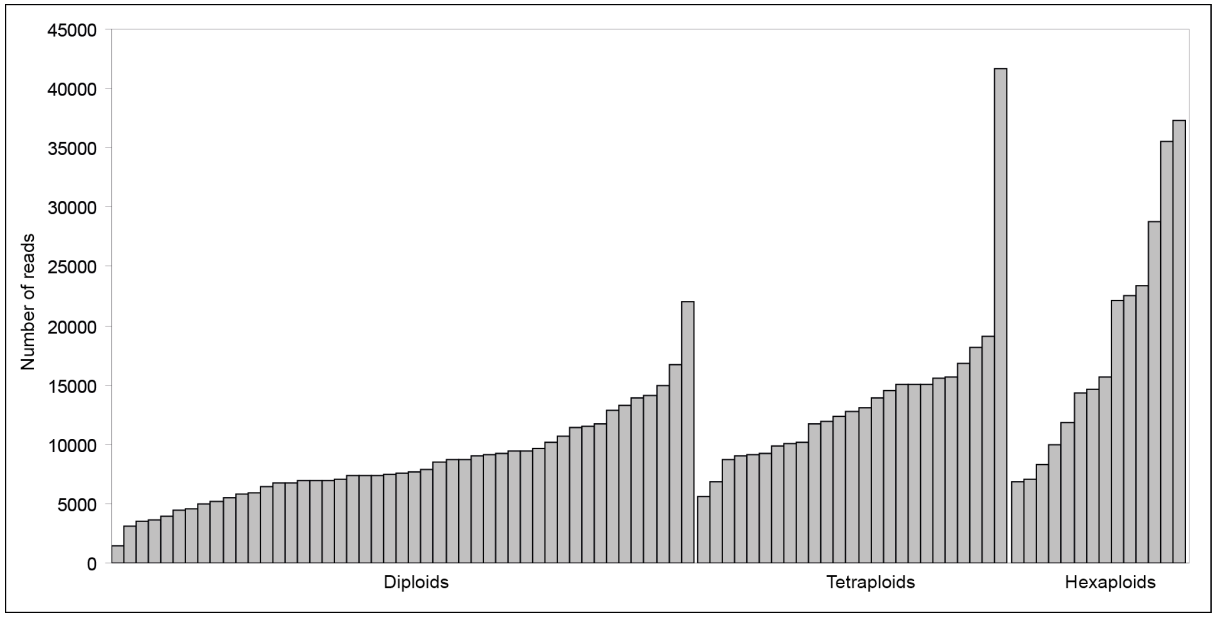


Figure S3.1 Distribution of 454 sequence reads obtained for the 92 individuals included in the final analyses. Individuals were grouped by ploidy level and ranked according to the number of reads, from smallest to largest.

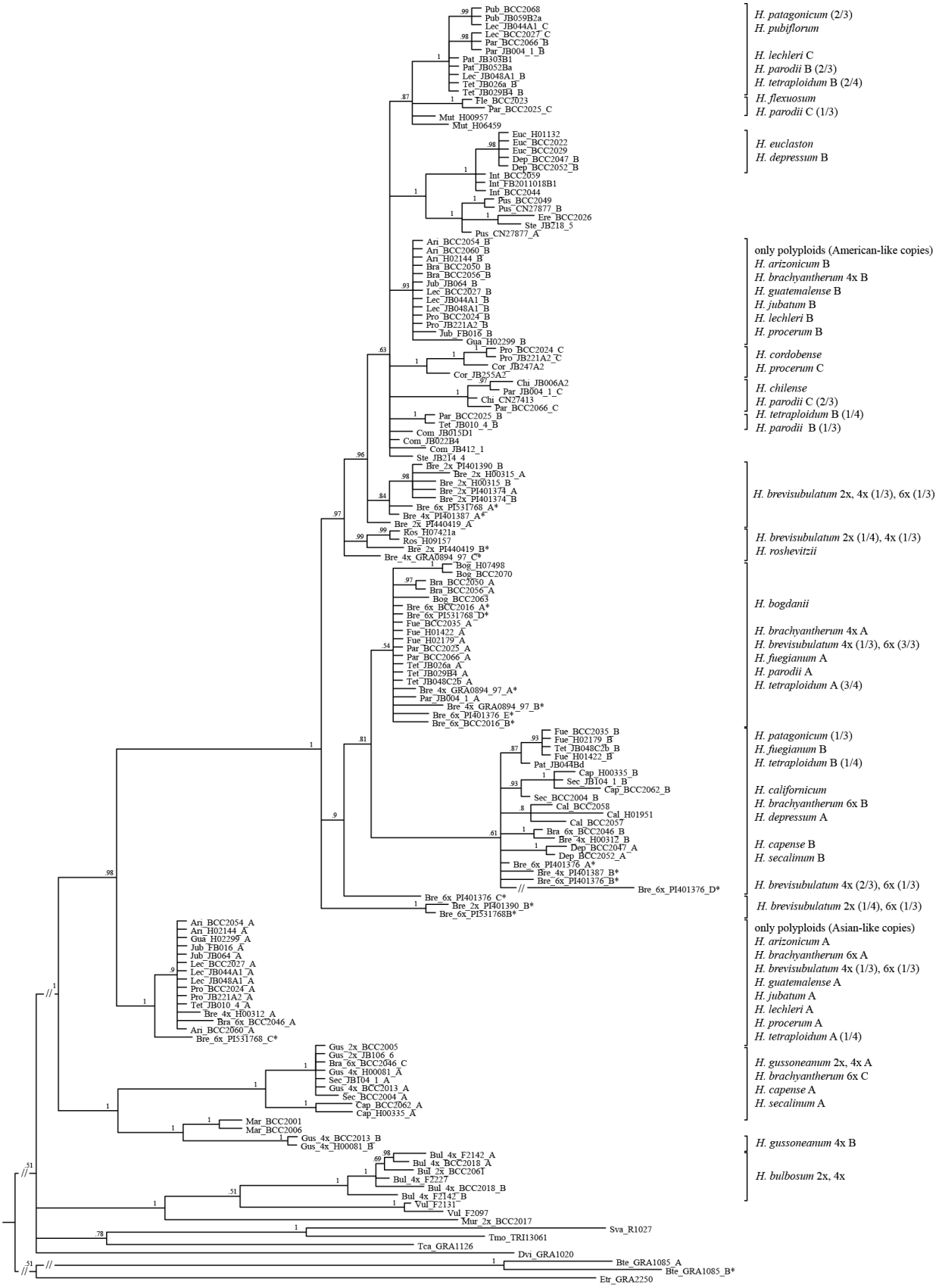


Figure S3.2 Phylogenetic tree inferred with BI for all individuals from *BLZ1* sequences. Values at nodes indicate posterior probabilities. Taxa names are abbreviated by the three first letters of the species epithet for *Hordeum* species and first letter of the genus name and two first letters of the species epithet for the outgroup species. Asterisks depict sequences excluded from the supermatrix and MLMSC analyses.

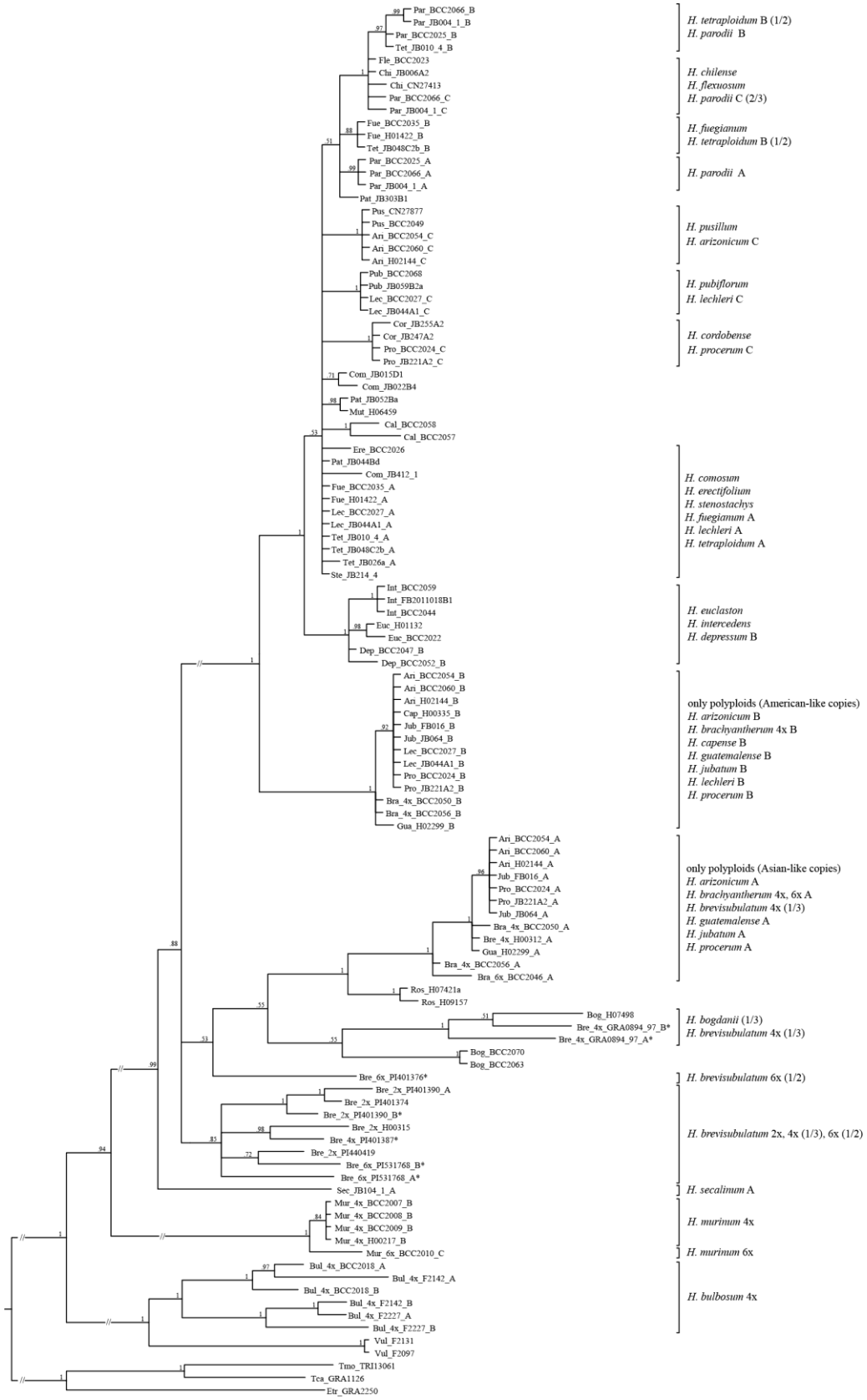


Figure S3.3 Phylogenetic tree inferred with BI for all individuals from *NUC* sequences. Values at nodes indicate posterior probabilities. Taxa names are abbreviated by the three first letters of the species epithet for *Hordeum* species and first letter of the genus name and two first letters of the species epithet for the outgroup species. Asterisks depict sequences excluded from the supermatrix and MLMSC analyses.

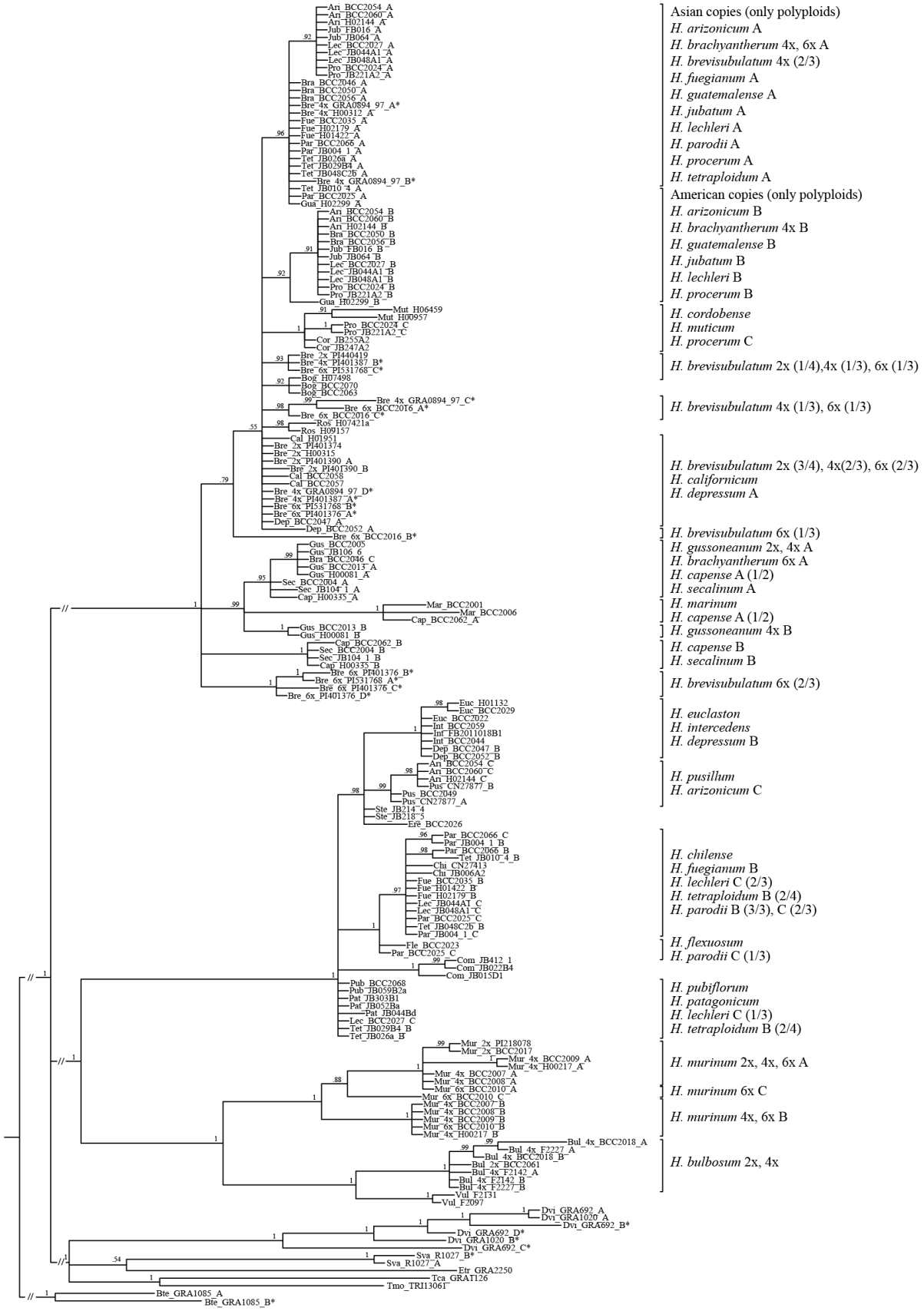


Figure S3.4 Phylogenetic tree inferred with BI for all individuals from *TNAC1035* sequences. Values at nodes indicate posterior probabilities. Taxa names are abbreviated by the three first letters of the species epithet for *Hordeum* species and first letter of the genus name and two first letters of the species epithet for the outgroup species. Asterisks depict sequences excluded from the supermatrix and MLMSC analyses.

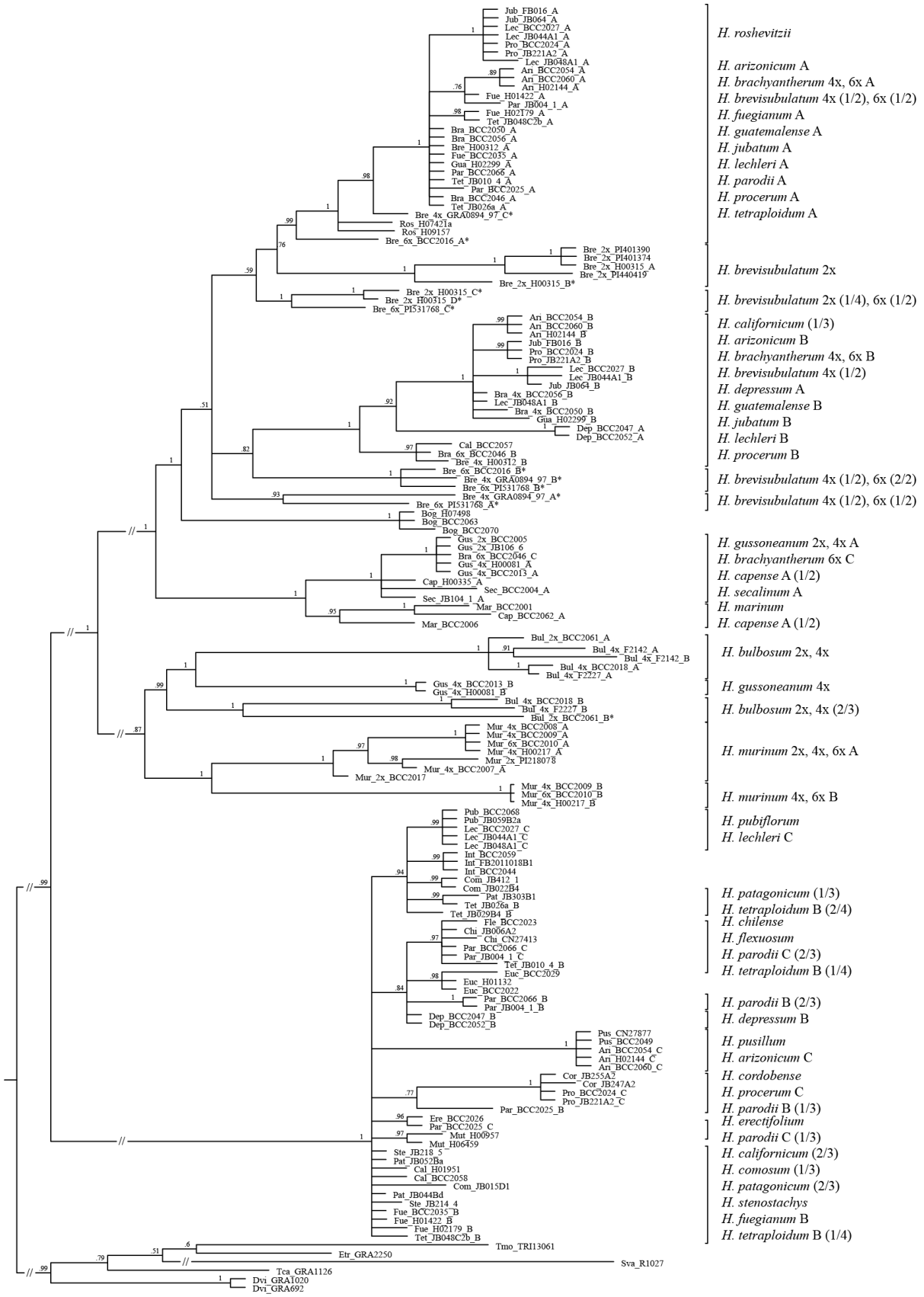


Figure S3.5 Phylogenetic tree inferred with BI for all individuals from *TNAC1142* sequences. Values at nodes indicate posterior probabilities. Taxa names are abbreviated by the three first letters of the species epithet for *Hordeum* species and first letter of the genus name and two first letters of the species epithet for the outgroup species. Asterisks depict sequences excluded from the supermatrix and MLMSC analyses.

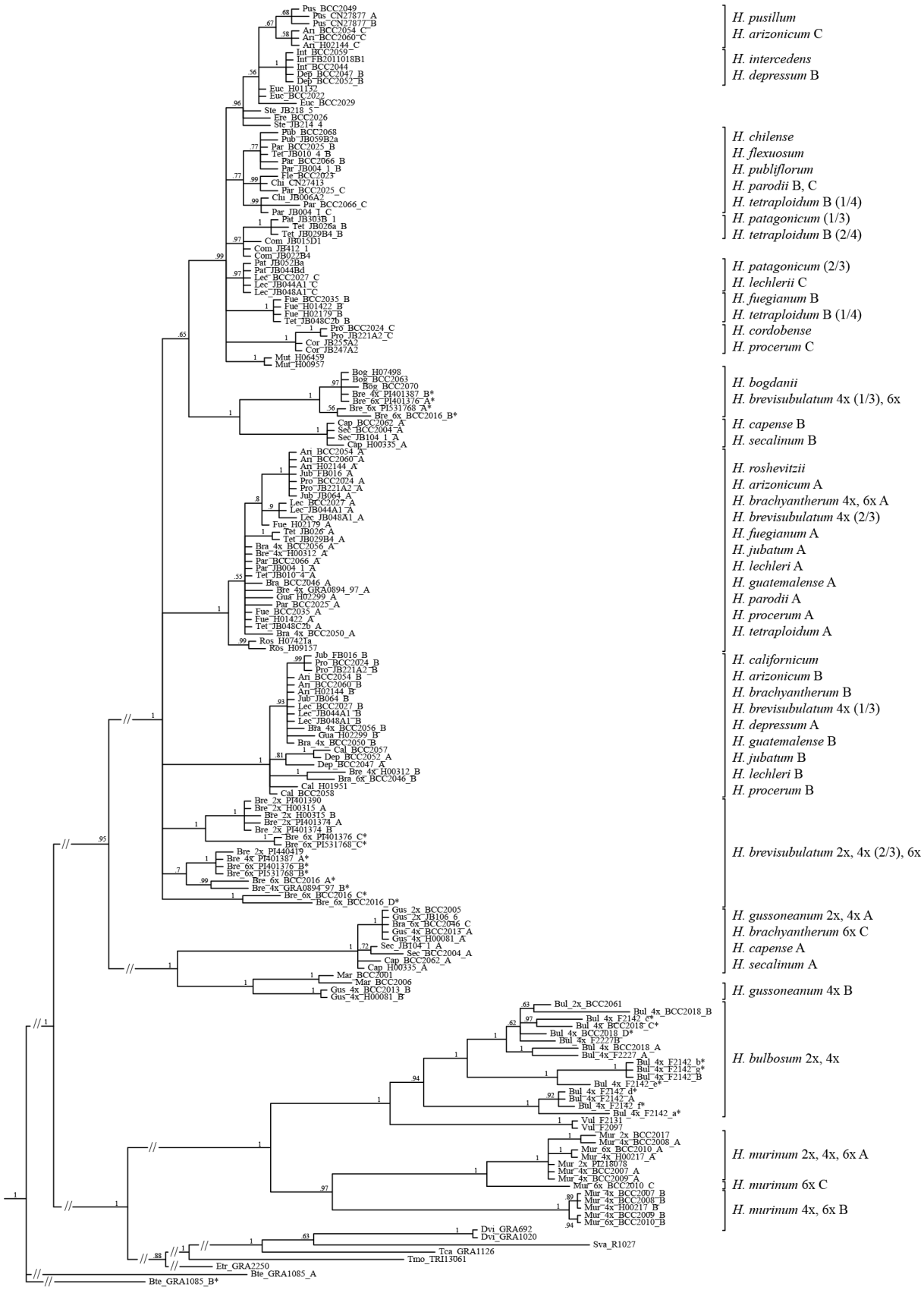


Figure S3.6 Phylogenetic tree inferred with BI for all individuals from *TNAC1364* sequences. Values at nodes indicate posterior probabilities. Taxa names are abbreviated by the three first letters of the species epithet for *Hordeum* species and first letter of the genus name and two first letters of the species epithet for the outgroup species. Asterisks depict sequences excluded from the supermatrix and MLMSC analyses.

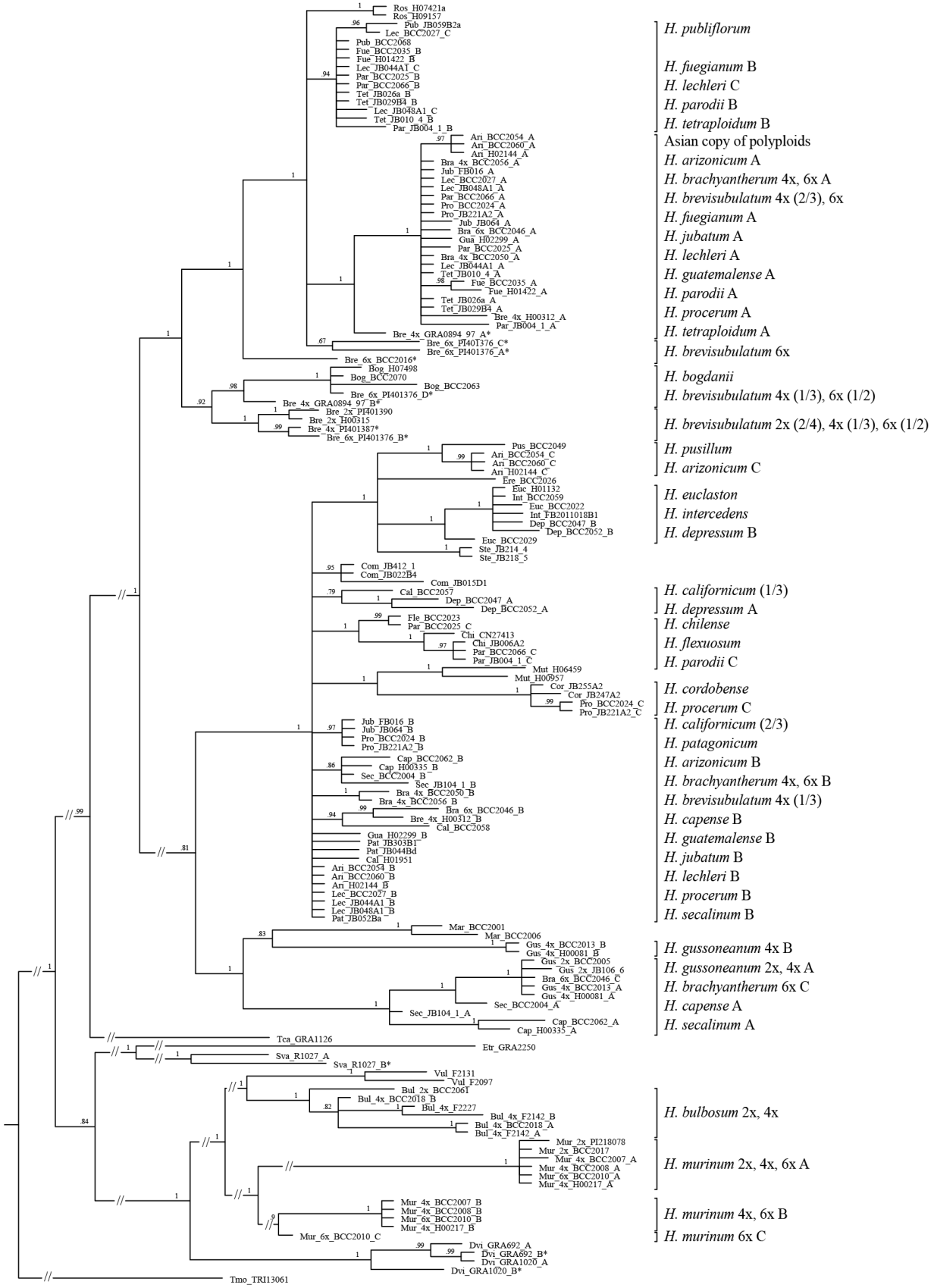


Figure S3.7 Phylogenetic tree inferred with BI for all individuals from *TNAC1403* sequences. Values at nodes indicate posterior probabilities. Taxa names are abbreviated by the three first letters of the species epithet for *Hordeum* species and first letter of the genus name and two first letters of the species epithet for the outgroup species. Asterisks depict sequences excluded from the supermatrix and MLMSC analyses.

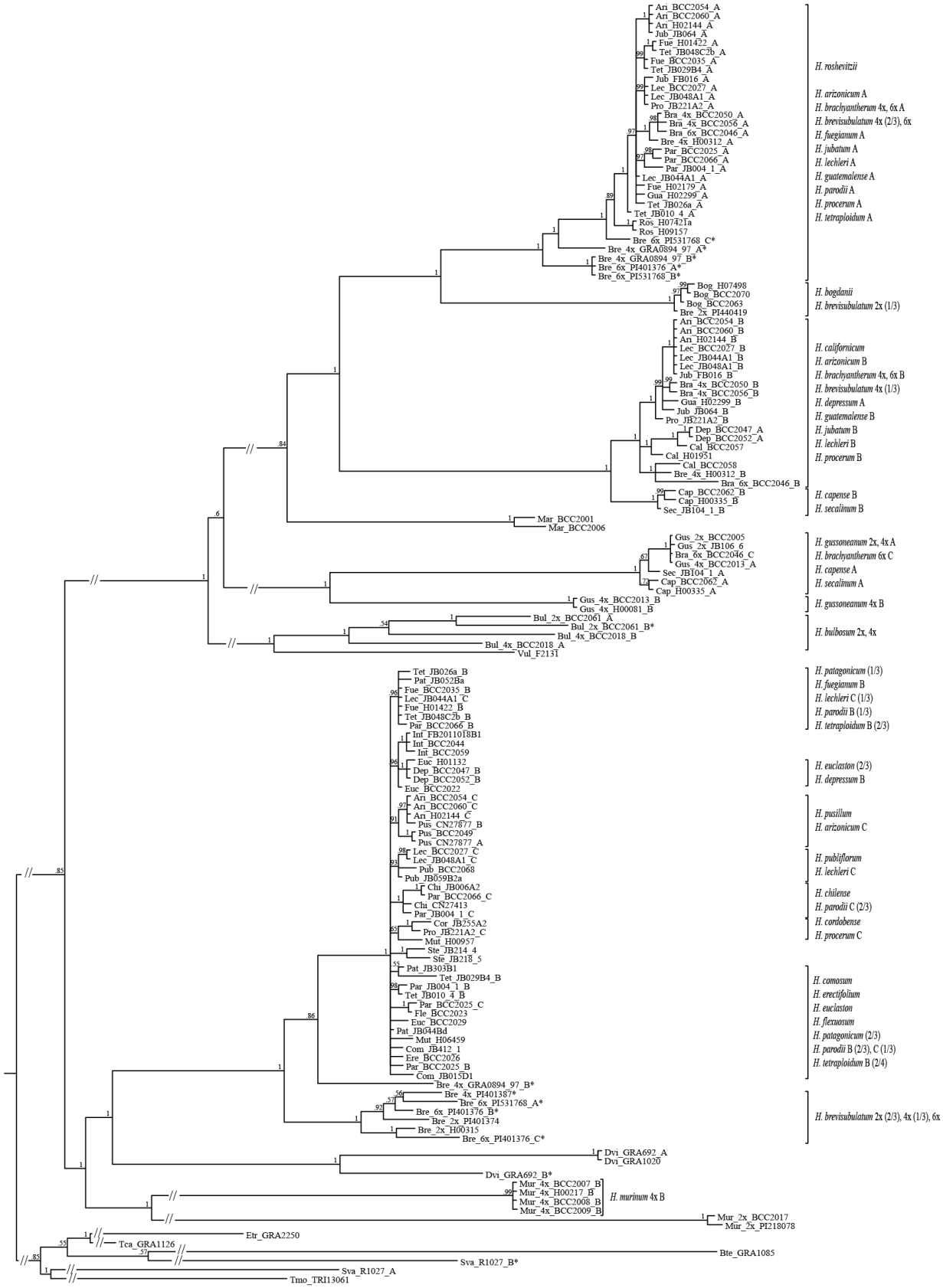


Figure S3.8 Phylogenetic tree inferred with BI for all individuals from *TNAC1463* sequences. Values at nodes indicate posterior probabilities. Taxa names are abbreviated by the three first letters of the species epithet for *Hordeum* species and first letter of the genus name and two first letters of the species epithet for the outgroup species. Asterisks depict sequences excluded from the supermatrix and MLMSC analyses.

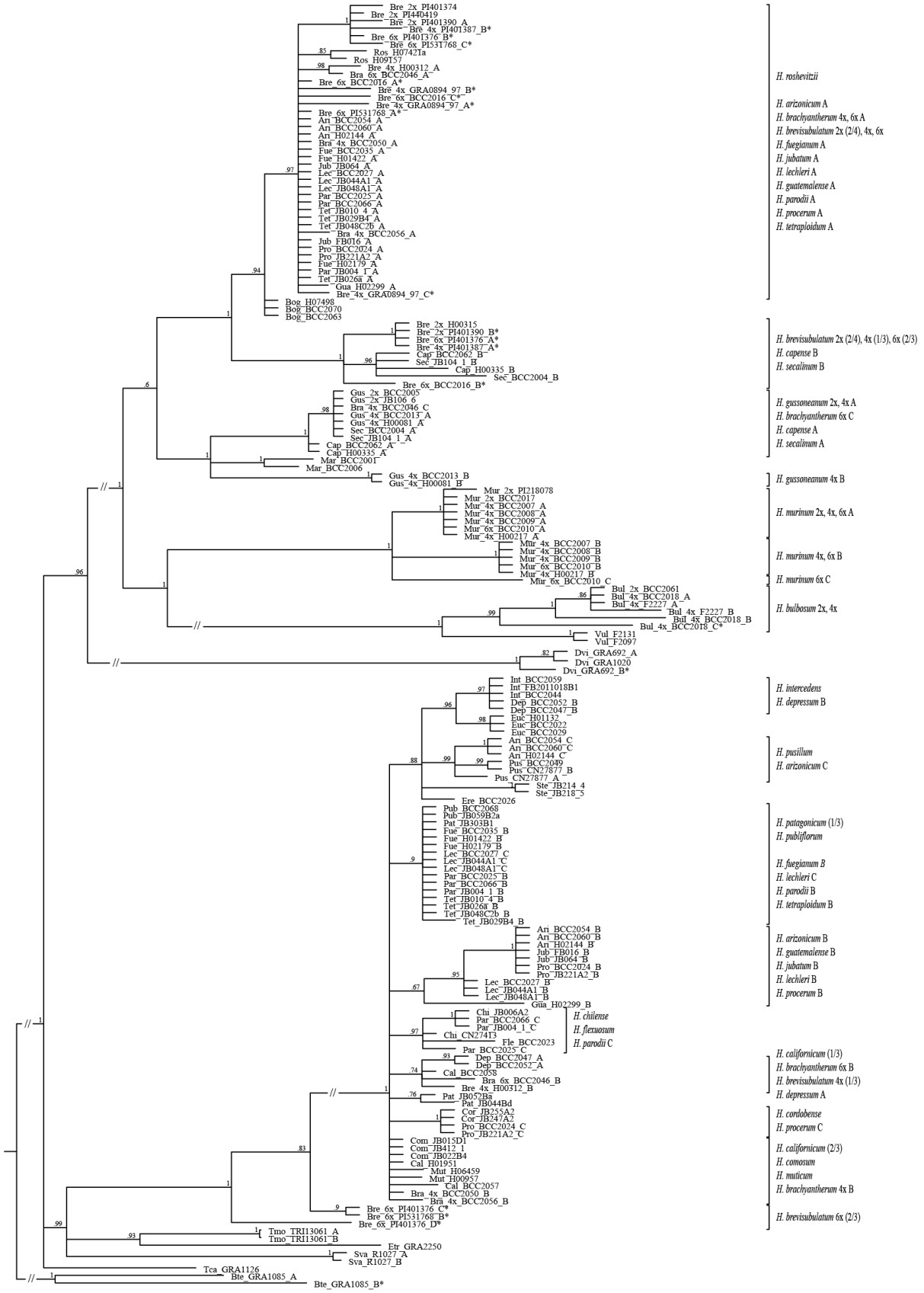


Figure S3.9 Phylogenetic tree inferred with BI for all individuals from *TNAC1497* sequences. Values at nodes indicate posterior probabilities. Taxa names are abbreviated by the three first letters of the species epithet for *Hordeum* species and first letter of the genus name and two first letters of the species epithet for the outgroup species. Asterisks depict sequences excluded from the supermatrix and MLMSC analyses.

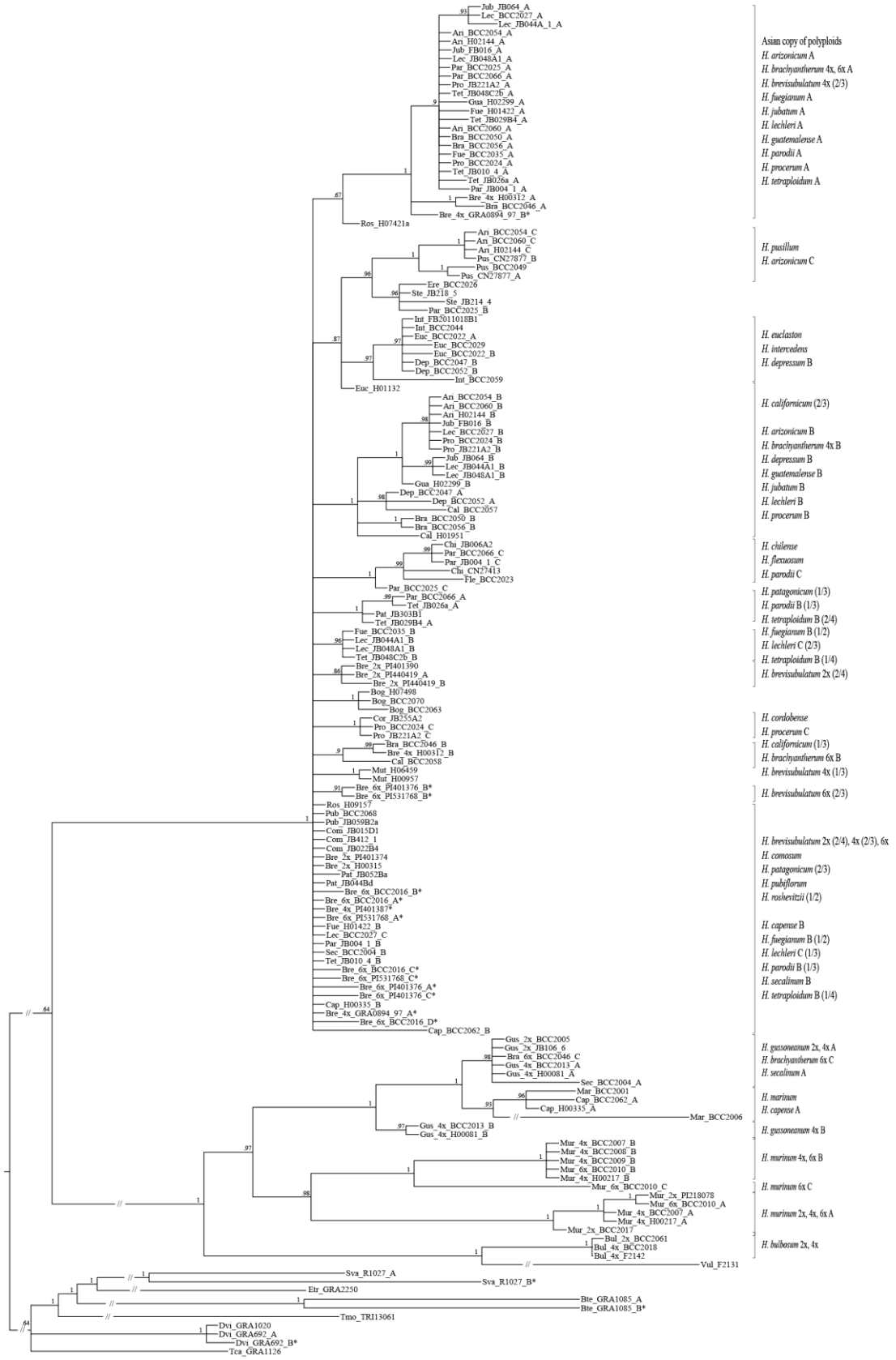


Figure S3.10 Phylogenetic tree inferred with BI for all individuals from *TNAC1610* sequences. Values at nodes indicate posterior probabilities. Taxa names are abbreviated by the three first letters of the species epithet for *Hordeum* species and first letter of the genus name and two first letters of the species epithet for the outgroup species. Asterisks depict sequences excluded from the supermatrix and MLMSC analyses.

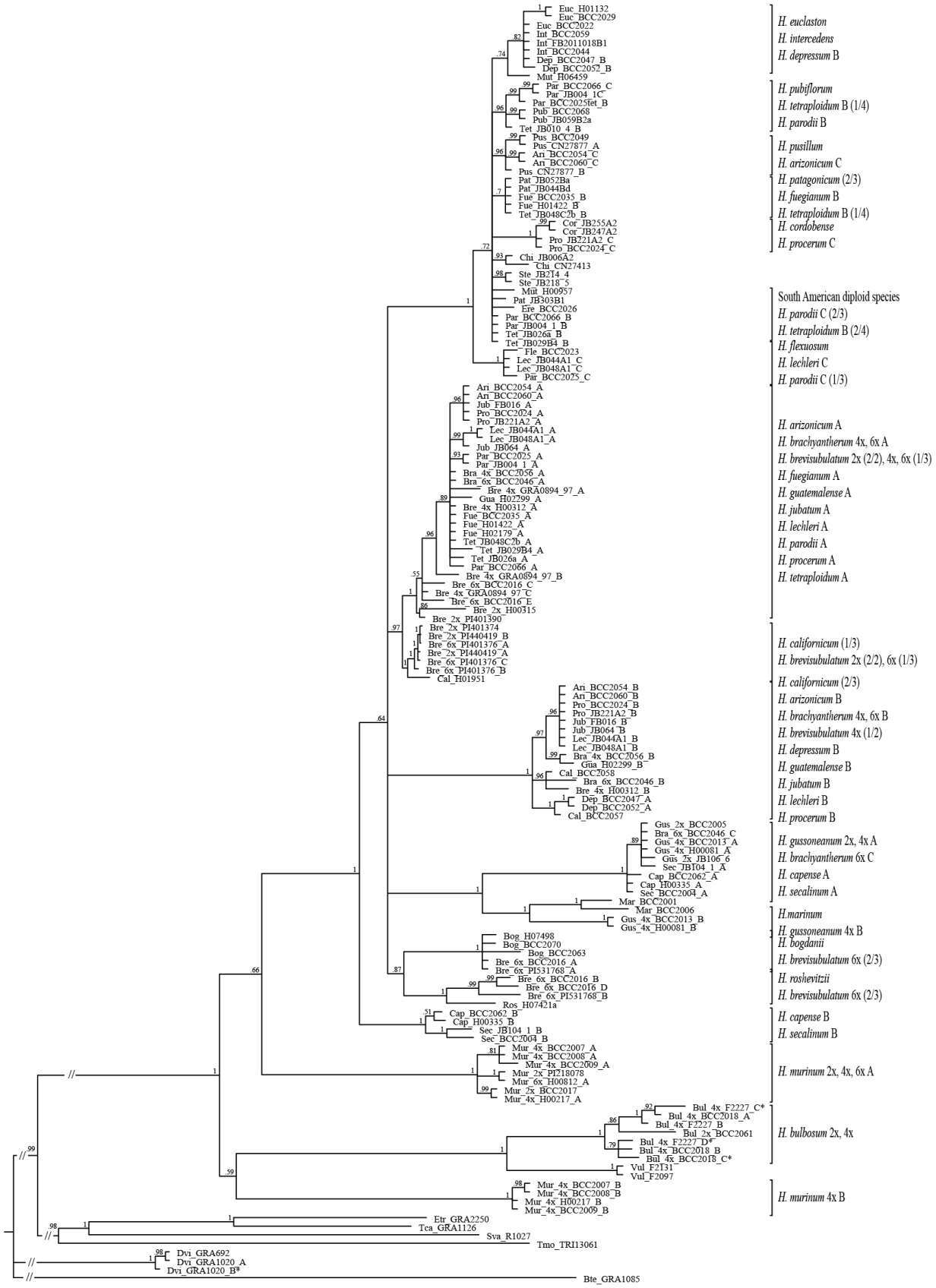


Figure S3.11 Phylogenetic tree inferred with BI for all individuals from *TNAC1740* sequences. Values at nodes indicate posterior probabilities. Taxa names are abbreviated by the three first letters of the species epithet for *Hordeum* species and first letter of the genus name and two first letters of the species epithet for the outgroup species. Asterisks depict sequences excluded from the supermatrix and MLMSC analyses.

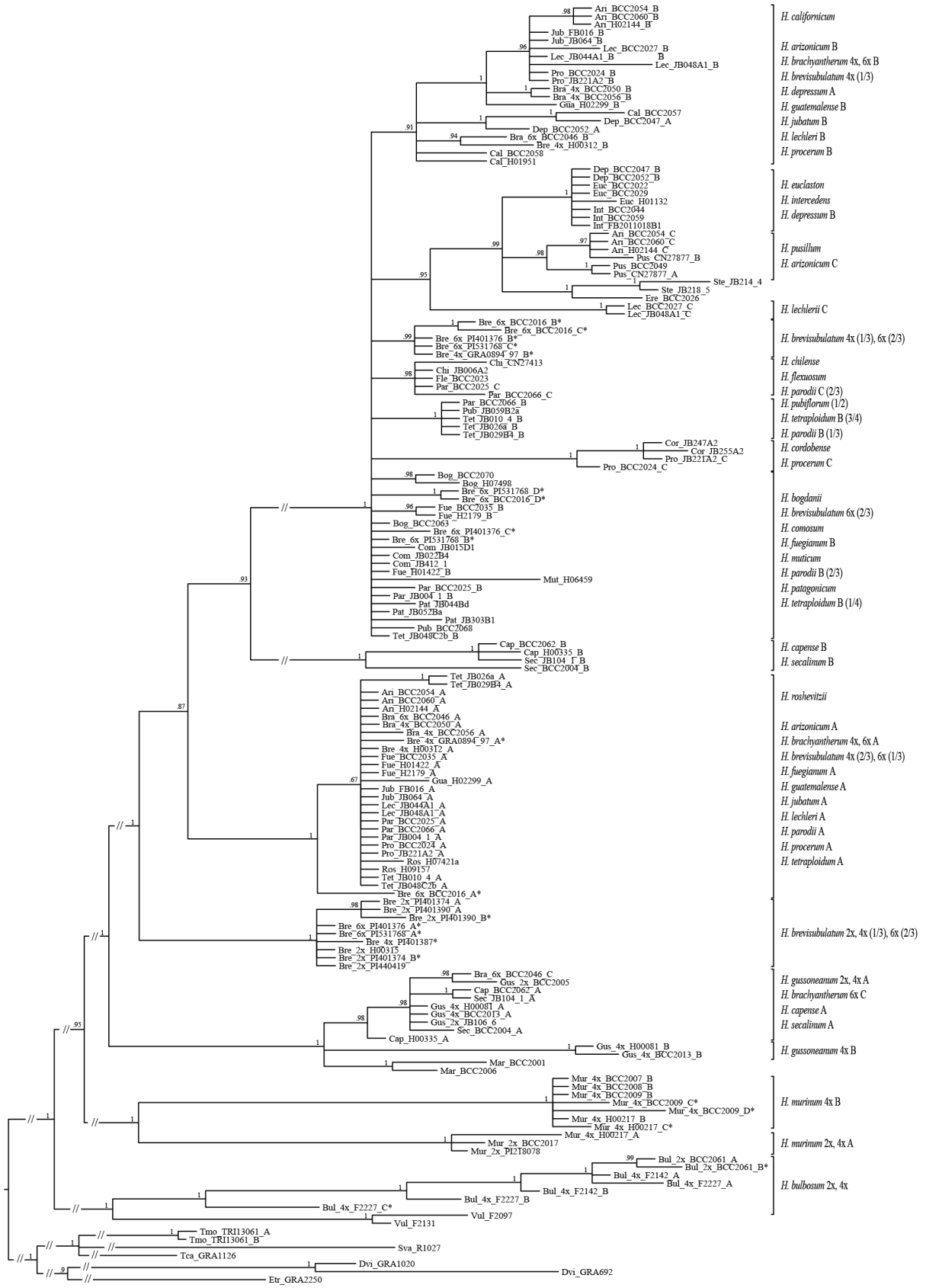


Figure S3.12 Phylogenetic tree inferred with BI for all individuals from *TOPO6* sequences. Values at nodes indicate posterior probabilities. Taxa names are abbreviated by the three first letters of the species epithet for *Hordeum* species and first letter of the genus name and two first letters of the species epithet for the outgroup species. Asterisks depict sequences excluded from the supermatrix and MLMSC analyses.

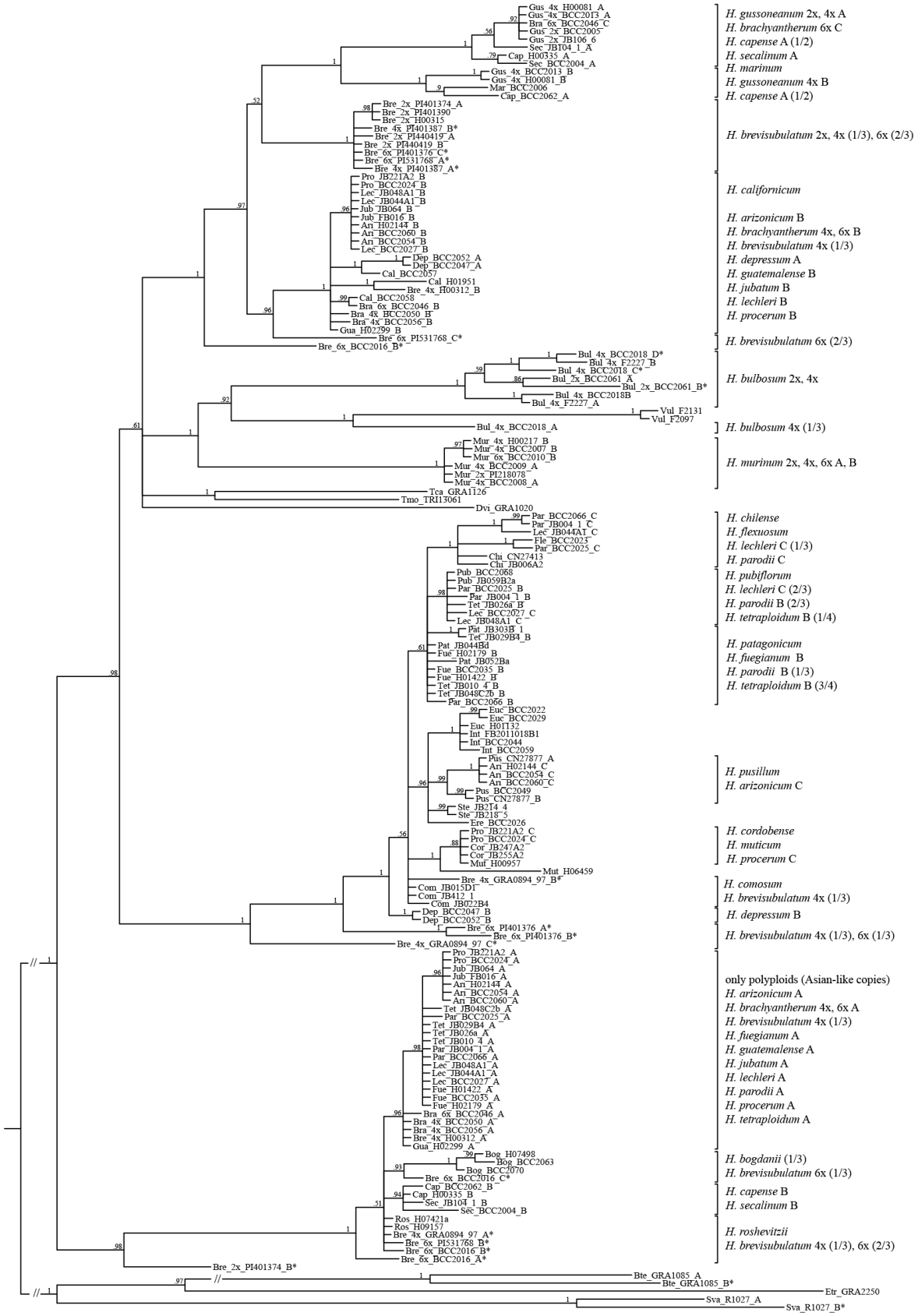


Figure S3.13 Phylogenetic tree inferred with BI for all individuals from *XYL* sequences. Values at nodes indicate posterior probabilities. Taxa names are abbreviated by the three first letters of the species epithet for *Hordeum* species and first letter of the genus name and two first letters of the species epithet for the outgroup species. Asterisks depict sequences excluded from the supermatrix and MLMSC analyses.

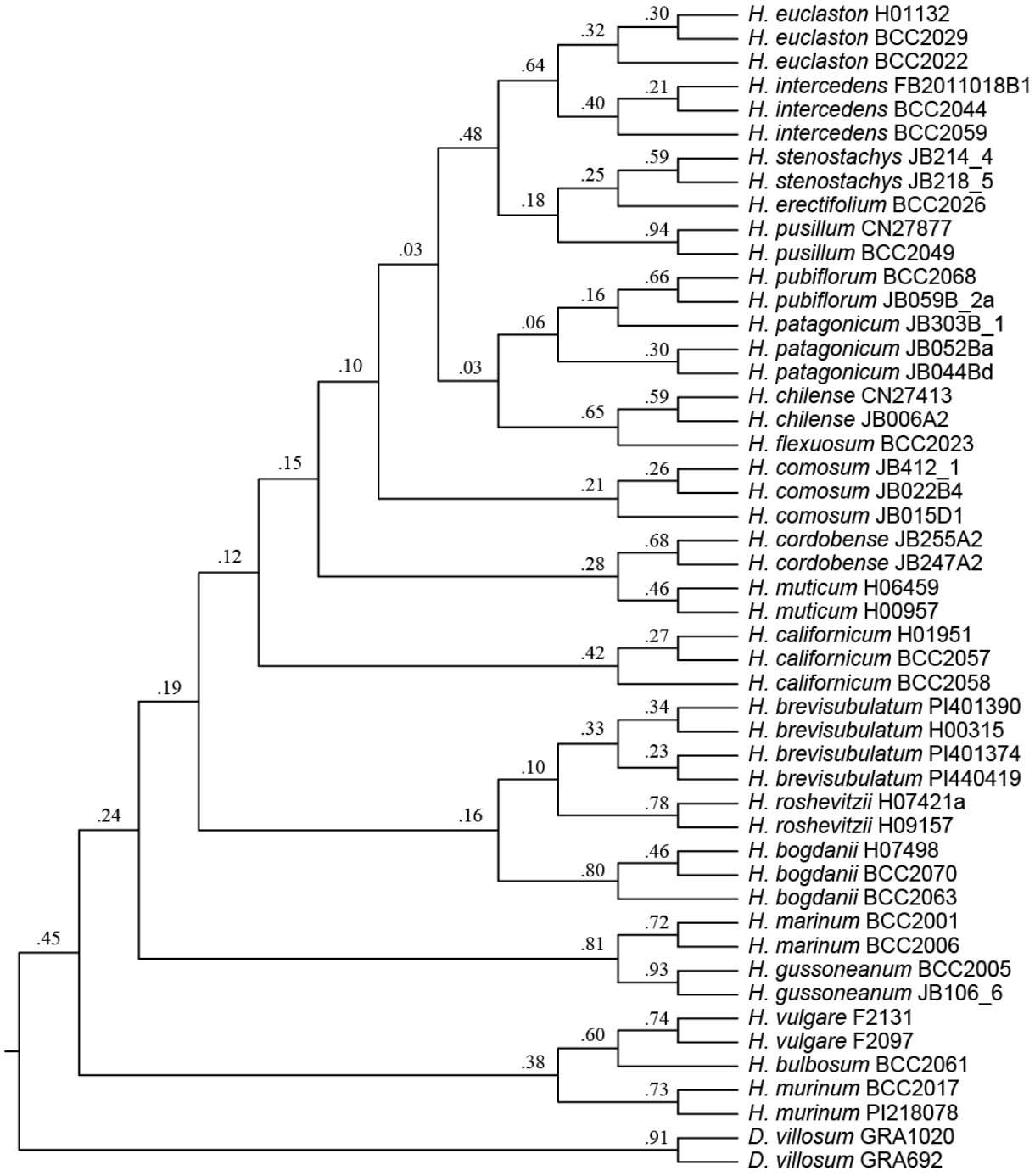


Figure S3.14 Primary Concordance Tree from BUCKY from all diploids, based on 11 loci.

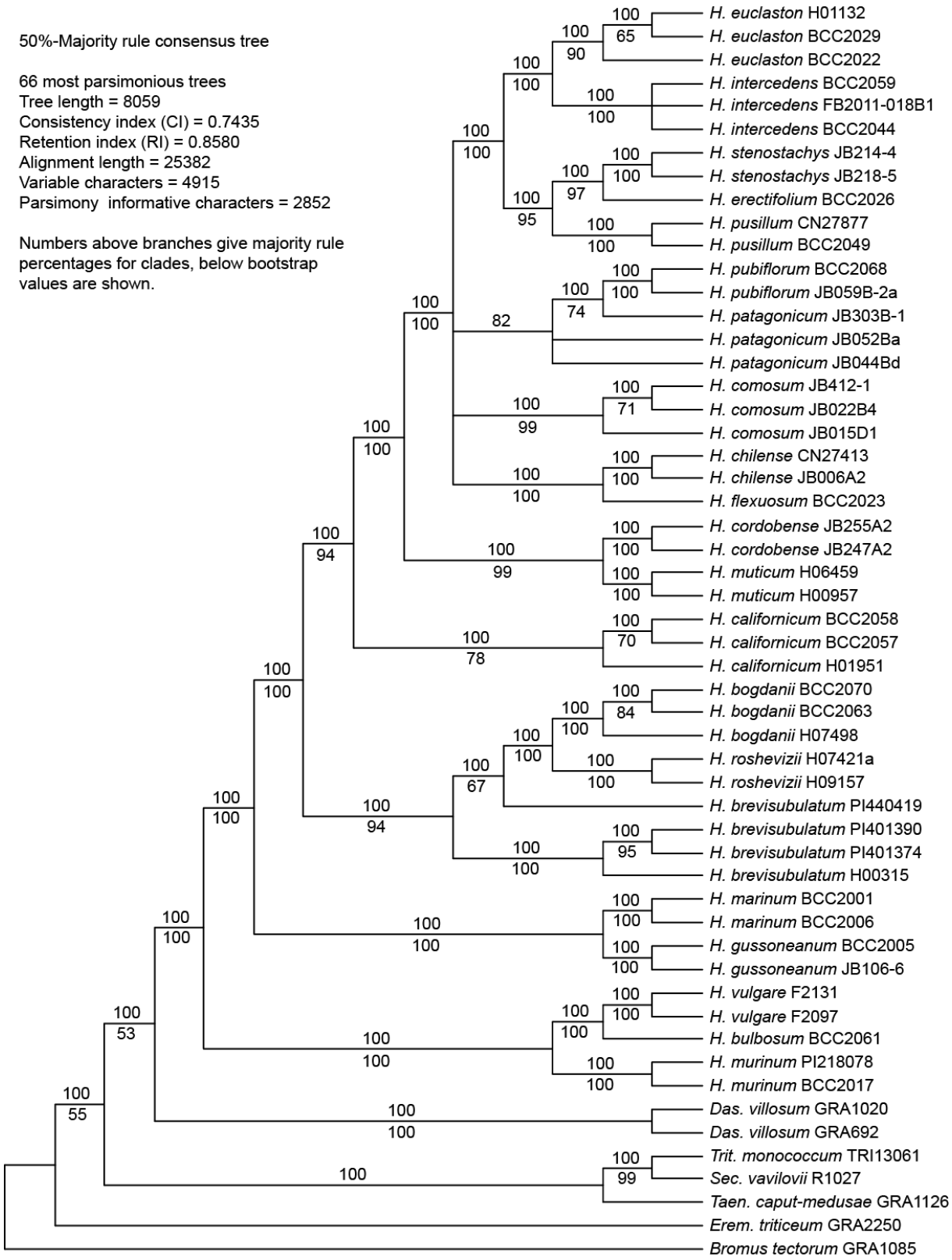


Figure S3.15 Phylogenetic tree inferred for the diploid individuals with MP from 12 single-copy nuclear loci.

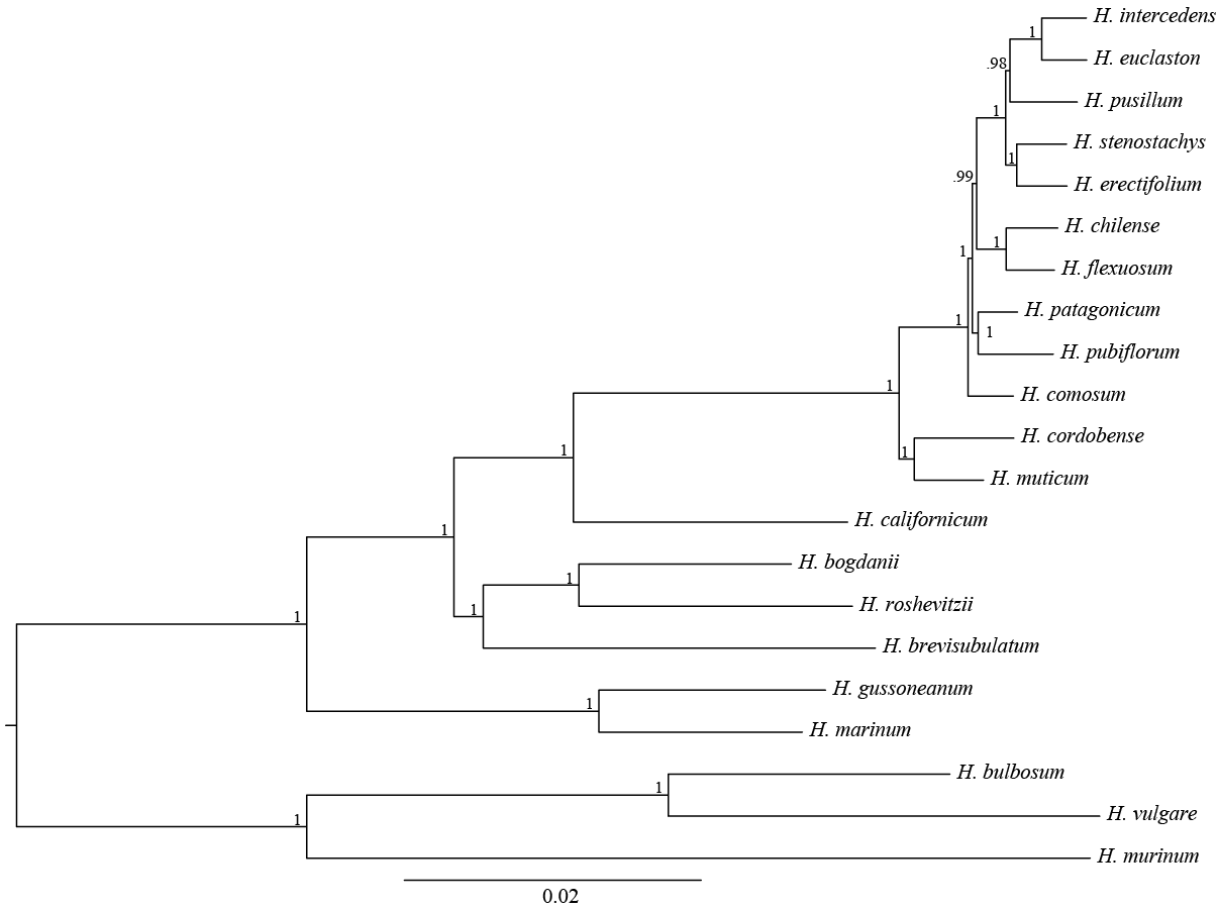
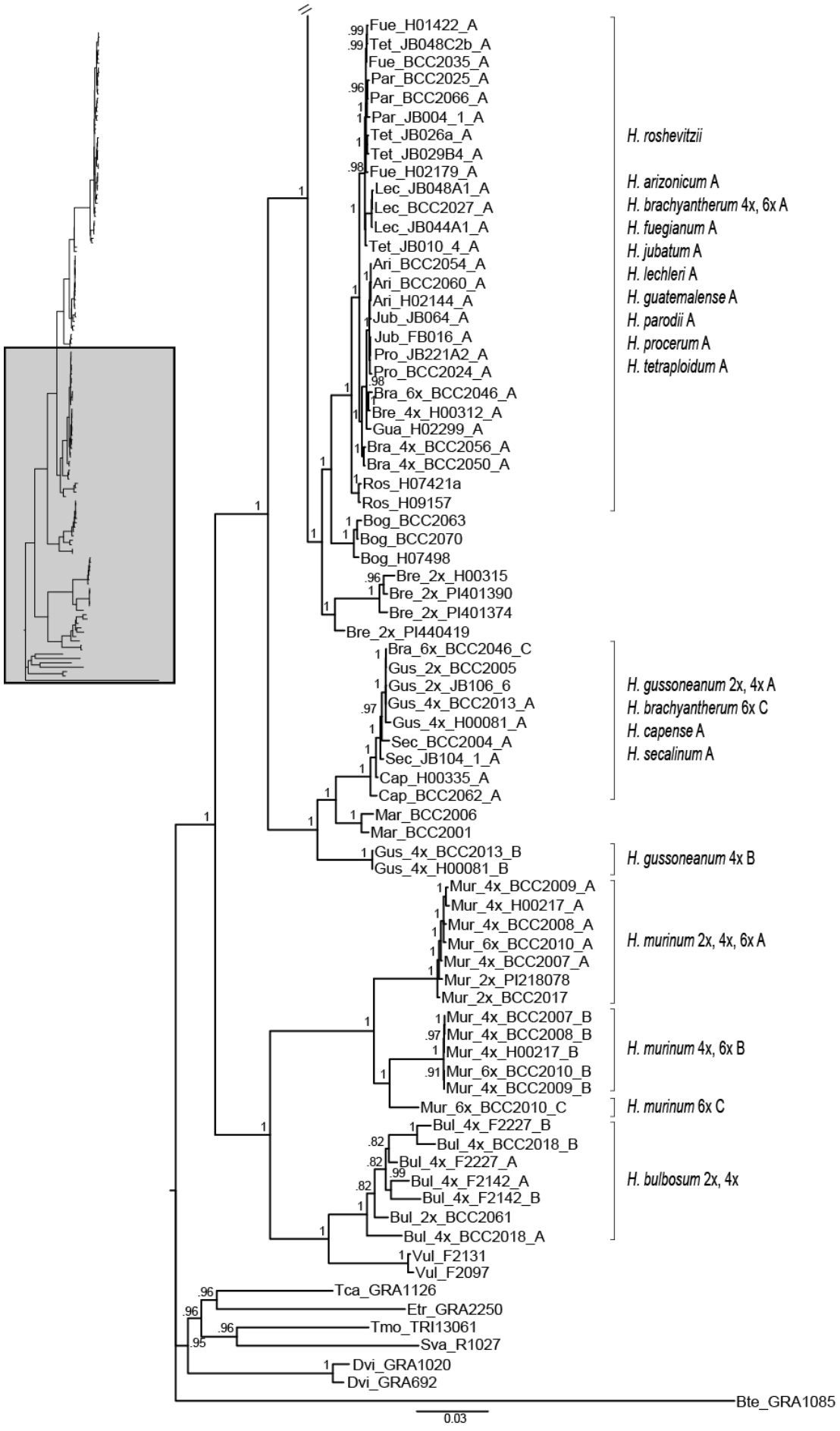
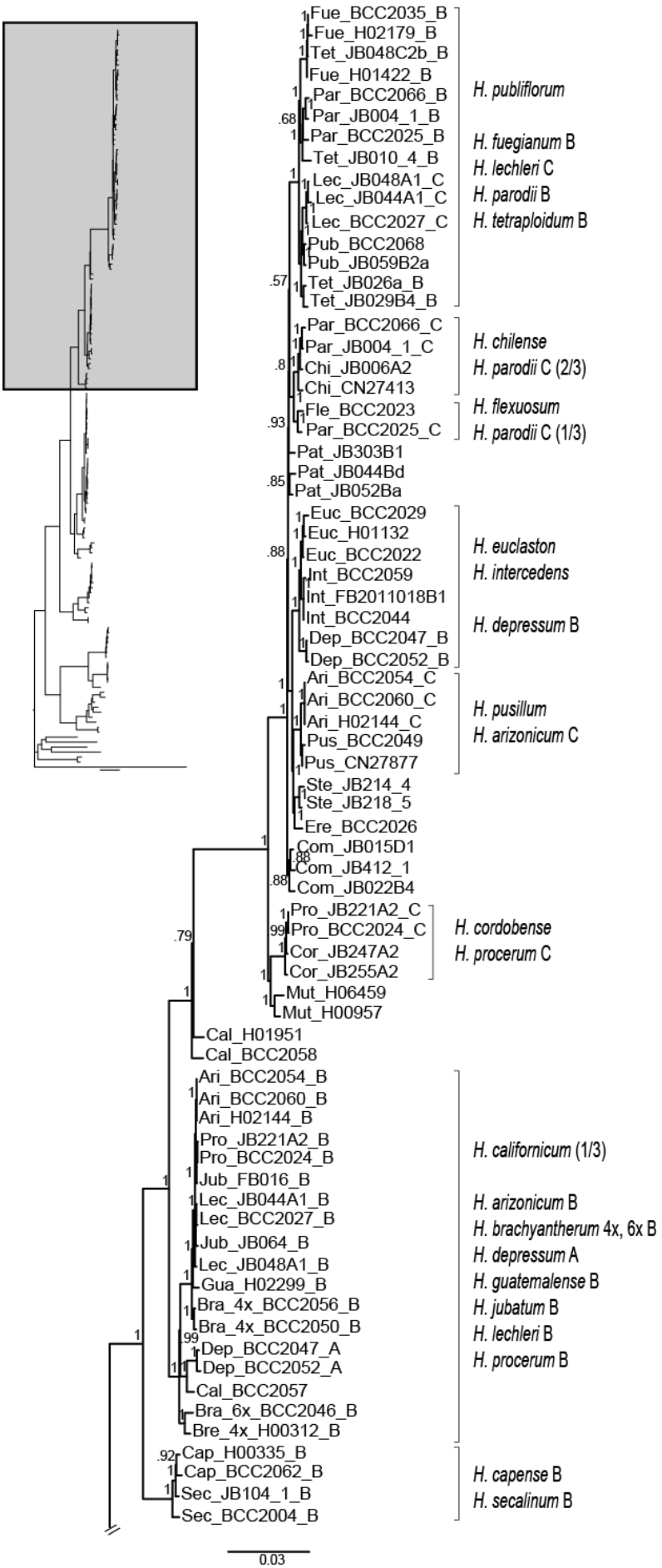


Figure S3.16 Phylogenetic tree inferred for the diploid individuals with BI from 19 single-copy nuclear loci combining the loci from this study and Petersen *et al.* (2011). Each species is represented by only one individual. Values at nodes indicate posterior probabilities.

Next double page:

Figure S3.17 Phylogenetic tree inferred with BI from all 12 nuclear single-copy sequences for all individuals. Values at nodes indicate posterior probabilities. Taxa names are abbreviated by the three first letters of the species epithet for *Hordeum* species and first letter of the genus name and two first letters of the species epithet for the outgroup species. After species name and individual number the different copies found per individual are indicated (A-C) for the polyploids.


Fue_H01422_A
Tet_JB048C2b_A
Fue_BCC2035_A
Par_BCC2025_A
Par_BCC2066_A
Par_JB004_1_A
Tet_JB026a_A
Tet_JB029B4_A
Fue_H02179_A
Lec_JB048A1_A
Lec_BCC2027_A
Lec_JB044A1_A
Tet_JB010_4_A
Ari_BCC2054_A
Ari_BCC2060_A
Ari_H02144_A
Jub_JB064_A
Jub_FB016_A
Pro_JB221A2_A
Pro_BCC2024_A
Bra_6x_BCC2046_A
Bre_4x_H00312_A
Gua_H02299_A
Bra_4x_BCC2056_A
Bra_4x_BCC2050_A
Ros_H07421a
Ros_H09157
Bog_BCC2063
Bog_BCC2070
Bog_H07498
Bre_2x_H00315
Bre_2x_PI401390
Bre_2x_PI401374
Bre_2x_PI440419
Bra_6x_BCC2046_C
Gus_2x_BCC2005
Gus_2x_JB106_6
Gus_4x_BCC2013_A
Gus_4x_H00081_A
Sec_BCC2004_A
Sec_JB104_1_A
Cap_H00335_A
Cap_BCC2062_A
Mar_BCC2006
Mar_BCC2001
Gus_4x_BCC2013_B
Gus_4x_H00081_B
Mur_4x_BCC2009_A
Mur_4x_H00217_A
Mur_4x_BCC2008_A
Mur_6x_BCC2010_A
Mur_4x_BCC2007_A
Mur_2x_PI218078
Mur_2x_BCC2017
Mur_4x_BCC2007_B
Mur_4x_BCC2008_B
Mur_4x_H00217_B
Mur_6x_BCC2010_B
Mur_4x_BCC2009_B
Mur_6x_BCC2010_C
Bul_4x_F2227_B
Bul_4x_BCC2018_B
Bul_4x_F2227_A
Bul_4x_F2142_A
Bul_4x_F2142_B
Bul_2x_BCC2061
Bul_4x_BCC2018_A
Vul_F2131
Vul_F2097
Tca_GRA1126
Etr_GRA2250
Tmo_TRI13061
Sva_R1027
Dvi_GRA1020
Dvi_GRA692
Bte_GRA1085
0.03
H. roshevitzii
H. arizonicum A
H. brachyantherum 4x, 6x A
H. fuegianum A
H. jubatum A
H. lechleri A
H. guatemalense A
H. parodii A
H. procerum A
H. tetraploidum A
H. gussoneanum 2x, 4x A
H. brachyantherum 6x C
H. capense A
H. secalinum A
H. gussoneanum 4x B
H. murinum 2x, 4x, 6x A
H. murinum 4x, 6x B
H. murinum 6x C
H. bulbosum 2x, 4x


Fue_BCC2035_B
Fue_H02179_B
Tet_JB048C2b_B
Fue_H01422_B
Par_BCC2066_B
Par_JB004_1_B
Par_BCC2025_B
Tet_JB010_4_B
Lec_JB048A1_C
Lec_JB044A1_C
Lec_BCC2027_C
Pub_BCC2068
Pub_JB059B2a
Tet_JB026a_B
Tet_JB029B4_B
Par_BCC2066_C
Par_JB004_1_C
Chi_JB006A2
Chi_CN27413
Fle_BCC2023
Par_BCC2025_C
Pat_JB303B1
Pat_JB044Bd
Pat_JB052Ba
Euc_BCC2029
Euc_H01132
Euc_BCC2022
Int_BCC2059
Int_FB2011018B1
Int_BCC2044
Dep_BCC2047_B
Dep_BCC2052_B
Ari_BCC2054_C
Ari_BCC2060_C
Ari_H02144_C
Pus_BCC2049
Pus_CN27877
Ste_JB214_4
Ste_JB218_5
Ere_BCC2026
Com_JB015D1
Com_JB412_1
Com_JB022B4
Pro_JB221A2_C
Pro_BCC2024_C
Cor_JB247A2
Cor_JB255A2
Mut_H06459
Mut_H00957
Cal_H01951
Cal_BCC2058
Ari_BCC2054_B
Ari_BCC2060_B
Ari_H02144_B
Pro_JB221A2_B
Pro_BCC2024_B
Jub_FB016_B
Lec_JB044A1_B
Lec_BCC2027_B
Jub_JB064_B
Lec_JB048A1_B
Gua_H02299_B
Bra_4x_BCC2056_B
Bra_4x_BCC2050_B
Dep_BCC2047_A
Dep_BCC2052_A
Cal_BCC2057
Bra_6x_BCC2046_B
Bre_4x_H00312_B
Cap_H00335_B
Cap_BCC2062_B
Sec_JB104_1_B
Sec_BCC2004_B
0.03
H. publiflorum
H. fuegianum B
H. lechleri C
H. parodii B
H. tetraploidum B
H. chilense
H. parodii C (2/3)
H. flexuosum
H. parodii C (1/3)
H. euclaston
H. intercedens
H. depressum B
H. pusillum
H. arizonicum C
H. cordobense
H. procerum C
H. californicum (1/3)
H. arizonicum B
H. brachyantherum 4x, 6x B
H. depressum A
H. guatemalense B
H. jubatum B
H. lechleri B
H. procerum B
H. capense B
H. secalinum B

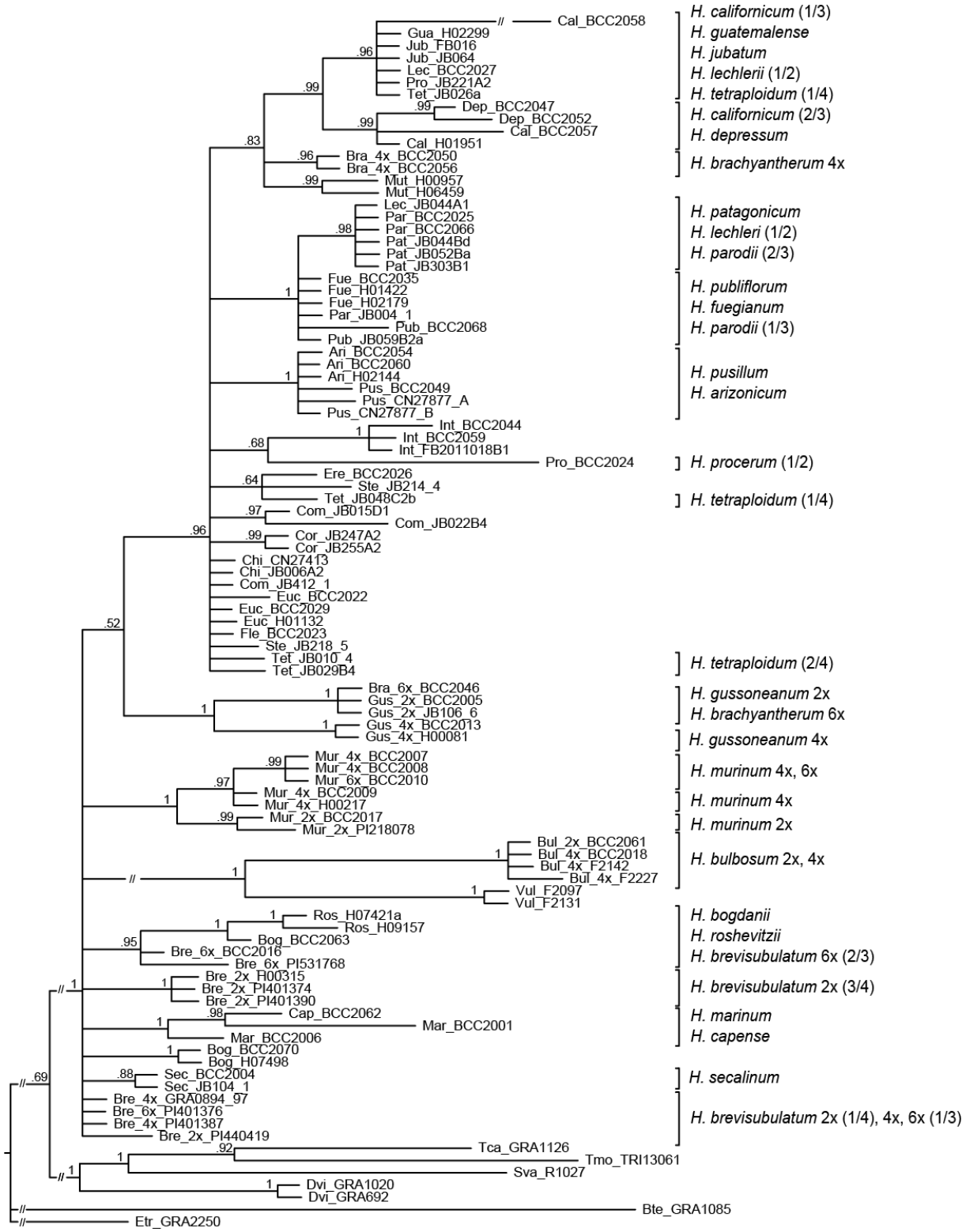


Figure S3.18 Phylogenetic tree inferred with BI for all individuals from *mat*K sequences. Values at nodes indicate posterior probabilities. Taxa names are abbreviated by the three first letters of the species epithet for *Hordeum* species and first letter of the genus name and two first letters of the species epithet for the outgroup species.

CURRICULUM VITAE & PUBLICATIONS

PERSONAL INFORMATION

Name	Jonathan Brassac
Date of birth	05th November 1986
Place of birth	Lyon, France
Nationality	French
Residency	Hans-Stubbe-Str. 22
	D-06466 Gatersleben
	Germany

STUDY AND PROFESIONAL EXPERIENCE

Since 02/2010	PhD student in the unit of Taxonomy and Evolutionary Biology at the Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben "Analysis of phylogenetic relationships of *Hordeum* (Poaceae) polyploids" (DFG grant no. BL 462/9 to FRB) in the workgroup of Dr. F.R. Blattner.
09/2007-06/2009	Study of Biodiversity, Ecology and Evolution, specialty "Diversity and Evolution of Plants and their Symbionts" at the University of Montpellier, France. Degree: Master of Science
01/2009-06/2009	Master Thesis: "The evolution of flowering gene *PHYC* during the domestication of pearl millet *Pennisetum glaucum* in West Africa" at the Institute for Research and Development (IRD), Montpellier, France.
04/2008-06/2008	Traineeship "Mating system in a population of sunflower, *Helianthus annuus* L., during neodomestication" at the National Institute of Agronomical Research (INRA), Montpellier, France.
09/2004-06/2007	Study of Biology and Ecology, specialty "Biology of Organisms and Populations" at the University of Lyon (UCBL1), France. Degree: Bachelor of Science

Extracurricular activities

Mar 2010 – Jan 2013	PhD Student Board member (workshop team, Plant Science Student Conference 2010 & 2012)
Jun 2012 – Present	Club (IPK social place) responsibility (accountant, movie and game nights)

PUBLICATIONS

Brassac J, Jakob SS, Blattner FR (2012) Progenitor-derivative relationships of *Hordeum* polyploids (Poaceae, Triticeae) inferred from sequences of *TOPO6*, a nuclear low-copy gene region. PLoS ONE 7(3): e33808. doi:10.1371/journal.pone.0033808

Brassac J, Blattner FR (2015) Species level phylogeny and polyploid relationships in *Hordeum* (Poaceae) inferred by next-generation sequencing and in-silico cloning of multiple nuclear loci. Systematic Biology 64(5): 792-808. doi: 10.1093/sysbio/syv035

PRESENTATIONS (selected)

Brassac J, Jakob SS, Blattner FR. Analysis of phylogenetic relationships of *Hordeum* (Poaceae) polyploids. *Biosystematics*, Berlin, 21.02 – 27.02.2011

Brassac J, Blattner FR. Phylogenetic analysis of *Hordeum* polyploids using next-generation sequencing. *8th Plant Science Student Conference*, Gatersleben, 04.06 – 07.06.2012

Brassac J, Blattner FR. Supermatrix or coalescent-based phylogenies? Analysis of next-generation sequenced multilocus data in *Hordeum* (Poaceae). *BioDivEvo*, Dresden, 24.03 – 27.03.2014

POSTERS (selected)

Brassac J, Jakob SS, Blattner FR. Analysis of phylogenetic relationships of *Hordeum* (Poaceae) polyploids. *7th Plant Science Student Conference*, Halle, 14.06 – 17.06.2011

Brassac J, Jakob SS, Blattner FR. Phylogenetic relationships of *Hordeum* polyploids (Poaceae, Triticeae) inferred form sequences of *TOPO6*, a nuclear low-copy gene region. *International Conference on Polyploidy, Hybridization and Biodiversity*, Průhonice, Czech Republic 07.05 – 10.05.2012

Brassac J, Herrmann K, Blattner FR. Next generation sequencing to investigate speciation in *Hordeum* (Poaceae, Triticeae) at the diploid and polyploid level. *21st International Symposium "Biodiversity and Evolutionary Biology" of the German Botanical Society (DBG)*, Mainz, 16.09 – 19.09.2012

Brassac J, Blattner FR. Phylogenetic relationships of *Hordeum* polyploids inferred from 454-sequenced single-copy genes. *55th Phylogenetic Symposium*, Oldenburg, 22.11 – 24.11.2013

EIGENSTÄNDIGKEITSERKLÄRUNG

Hiermit erkläre ich, Jonathan Brassac, dass diese Arbeit bisher weder der Naturwissenschaftlichen Fakultät der Martin-Luther-Universität Halle Wittenberg noch irgendeiner anderen wissenschaftlichen Einrichtung zum Zweck der Promotion vorgelegt wurde.

Ich erkläre, dass ich mich bisher noch nie um einen Doktorgrad beworben habe.

Ich erkläre, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe. Es wurden keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, und die den benutzten Werken wörtlich oder inhaltlich entnommen Stellen wurden als solche kenntlich gemacht.

Unterschrift

Ort / Datum: Gatersleben, den 30/10/2015