WILEY-BLACKWELL

Cladistics 27 (2011) 428–446

Cladistics

10.1111/j.1096-0031.2011.00347.x

When is enough, enough in phylogenetics? A case in point from *Hordeum* (Poaceae)

Gitte Petersen[a,*], Lone Aagesen[b], Ole Seberg[a] and Ida Hartvig Larsen[a]

[a]*Botanical Garden and Museum, Natural History Museum of Denmark, University of Copenhagen, Sølvgade 83 opg. S, DK-1307 Copenhagen, Denmark;* [b]*Instituto de Botánica Darwinion, Labardén 200, CC 22, San Isidro B1642HYD, Argentina*

Accepted 19 December 2010

Abstract

Direct optimization was used to reconstruct the phylogeny of the 26 diploid taxa included in the genus *Hordeum*. The total data set was composed of 16 nucleotide sequence regions from the nuclear as well as the plastid genome. The nine nuclear regions were from single-copy, protein coding genes located on six of the seven chromosome pairs in the diploid *H. vulgare* genome. The seven plastid regions comprise protein coding genes as well as intergenic regions. Studies of character congruence between data partitions showed no correlation between chromosomal location and congruence among the nuclear sequences and a level of congruence among the plastid sequences comparable with the level among the nuclear sequences. Combined analysis of all data resolved the phylogeny completely with most clades being robust and well supported. However, due to incongruence among data partitions some relationships are still and likely to remain ambiguously inferred. Rather than adding still more genes to the phylogenetic analyses, patterns of incongruence may be better explored by adding data from multiple specimens per taxon. For some species relationships the plastid data appear positively misleading, emphasizing the need for caution if plastid data are the only or dominant type of data used for phylogenetic reconstruction and subsequent re-classification.

The genus *Hordeum* L. consists of 31 diploid, tetraploid, and hexaploid species including diploid, cultivated barley, *H. vulgare* ssp. *vulgare* (Baden and Bothmer, 1994; Bothmer et al., 1995). *Hordeum* is part of the tribe Triticeae, renowned for its pronounced incongruence among gene trees, although *Hordeum* itself is always recovered as monophyletic (e.g. Kellogg et al., 1996; Mason-Gamer and Kellogg, 1996; Seberg and Petersen, 2007).

In the most recent taxonomic treatment of the whole genus, *Hordeum* is divided into four sections [*Hordeum*, *Stenostachys* Nevski, *Critesion* (Raf.) Nevski, and *Anisolepis* Nevski] based mainly on morphological characteristics (Bothmer et al., 1995). However, a recent phylogenetic analysis based on two single-copy nuclear genes, one plastid gene, and plastid restriction fragment length polymorphisms (RFLPs) showed that only *Hordeum* sect. *Hordeum* was monophyletic, whereas *H.* sect. *Stenostachys* and *H.* sect. *Critesion* had to be emended in order to be monophyletic (Petersen and Seberg, 2003). Further, *Hordeum* sect. *Anisolepis* was synonymized with *H.* sect. *Critesion* and *H.* sect. *Sibirica* (Nevski) G. Petersen & Seberg was created. In a recent review of *Hordeum* phylogenies a re-classification is proposed by Blattner (2009). Unfortunately this classification is not based on a phylogenetic analysis, but on the author's personal perception of the "true" phylogeny.

Attempts to resolve the phylogeny within *Hordeum* are complicated by the postulated, and in some cases demonstrated, allopolyploid origin of most of the polyploid species (Bothmer et al., 1995; Petersen and Seberg, 2004; Komatsuda et al., 2009). Phylogenetic analyses including the entire genus have therefore been

*Corresponding author:
E-mail address: gittep@snm.ku.dk

based on either plastid data or on nuclear internal transcribed spacer (ITS) or 5S rDNA spacer sequence data (Doebley et al., 1992; Petersen and Seberg, 1998; Nishikawa et al., 2002; Blattner, 2004; Baum et al., 2010). In a phylogenetic context, all of these sequences are, however, burdened with problems: Plastid data from allopolyploids only provide information about relationships to one parent, and as nuclear and plastid data further have been shown to provide conflicting results, plastid data alone cannot reliably be used for phylogenetic reconstruction (Petersen and Seberg, 2003). ITS may be subjected to various degrees of concerted evolution, and thus interpretation of results from inclusion of ITS data in phylogenetic analysis is ambiguous (Wendel et al., 1995). The ITS data provided by Blattner (2004) clearly demonstrate that several different ITS copies may be present even in diploid species of *Hordeum*. Like ITS, 5S rDNA gene and spacer sequences occur in tandem arrays (Sastri et al., 1992; Kellogg and Appels, 1995). Both the number of arrays and the number of sequence copies within each array vary among taxa, but there is no homogenization either within or between arrays, and thus multiple different sequence copies are expected to occur, rendering sequence retrieval as well as subsequent phylogenetic analyses highly spurious.

Several phylogenetic analyses have therefore included sequences from single-copy nuclear genes, but these are problematic to retrieve and difficult to interpret in polyploid taxa. Consequently, these phylogenies have included only diploid taxa (Komatsuda et al., 1999; Petersen and Seberg, 2003, 2009; Blattner, 2006; Sun et al., 2009; Jakob et al., 2010) or diploids plus a few, selected polyploid taxa (Petersen and Seberg, 2003, 2004; Kakeda et al., 2009; Komatsuda et al., 2009).

In general and despite their caveats, the majority of the above-mentioned phylogenetic analyses show broad agreement about relationships. However, some crucial conflicts exist between phylogenetic trees derived from plastid and nuclear sequences (see, for example, Petersen and Seberg, 2003; Blattner, 2009). Further, the supposedly closely related American species in *H*. sect. *Critesion* have remained largely unresolved due to lack of informative characters. Hence, additional data are needed both to resolve the relationships among the American species, and to explore incongruence between data from the plastid and the nuclear genomes further. In the present analysis we supplement the available data sets with a large quantity of novel data from both nuclear and plastid genes in order to answer the following questions: Are plastid genes, which obviously belong to the same linkage group, congruent? Are nuclear genes, which obviously belong to different linkage groups, congruent? Are nuclear genes located on the same chromosome (at least in *H. vulgare*) more congruent than genes located on different chromosomes? Is congruence higher among plastid genes than among nuclear genes? Are plastid and nuclear genes mutually incongruent?

To answer these questions a total of 16 data sets have been used. The nine nuclear data sets comprise gene regions from six of the seven chromosome pairs of *H. vulgare* (Bothmer et al., 1995). Five have been used previously, namely disrupted meiotic cDNA 1 (DMC1; Petersen and Seberg, 2003, 2009) located on chromosome 3 (= 3H) (V. Klimyuk, pers. commun.), elongation factor G (EF-G; Petersen and Seberg, 2003) located on chromosome 2 (= 2H) (Komatsuda et al., 1999), xylose isomerase (XYL; Petersen and Seberg, 2009) located on chromosome 2 (= 2H) (Pillen et al., 2000), nucellin (NUC; Petersen and Seberg, 2009) located on chromosome 4 (= 4H) (GenBank accession no. U87129), and barley leucin zipper 1 (BLZ1; Petersen and Seberg, 2009) located on chromosome 7 (= 5H) (Vicente-Carbajosa et al., 1998), and four are produced for this study, namely acyl carrier protein III (ACL3) located on chromosome 1 (= 7H) (Hansen and von Wettstein-Knowles, 1991), NAD(P)H-bispecific nitrate reductase (NAR7) located on chromosome 6 (= 6H) (GenBank accession no. X60173), vrn3-H3 gene similar to Arabidopsis flowering locus T (VRN3) located on chromosome 1 (= 7H) (Yan et al., 2006), and granule-bound starch synthase I (GBSSI = waxy) located on chromosome 1 (= 7H) (Genbank accession no. X07931).

Seven plastid data sets are included: six from the large, single-copy region, namely *atp*B-*rbc*L (including the entire intergenic region and *rbc*L; Nishikawa et al., 2002; Petersen and Seberg, 2003), *mat*K (Nishikawa et al., 2002), *trn*L-F (Nishikawa et al., 2002), *psb*K-I, *atp*F-H, and *trn*H-*psb*A, and one from the small, single-copy region, namely *ndh*F. The latter four regions are produced for this study, and for completeness some new sequences were added to the three former data sets.

We have chosen not to include the previously produced ITS and 5S rDNA data (Blattner, 2004; Baum et al., 2010) because of the potential problems related to paralogy and concerted evolution. Additionally, we have chosen not to include a data set composed of plastid RFLPs (Doebley et al., 1992), which belong to a different category of data, with an unknown relationship and a potential overlap with the increased amount of plastid nucleotide sequence data used here. Three recently published nuclear data sets cannot readily be combined with our data due to differences in taxon sampling (Kakeda et al., 2009; Sun et al., 2009; Jakob et al., 2010).

A large fraction of the available sequence diversity is confined to non-protein coding regions with length differences among taxa reflecting frequent occurrence of insertion/deletion events. Previous phylogenetic analyses of *Hordeum* based on nucleotide sequence data have

used either manual sequence alignments (Petersen and Seberg, 1998, 2003, 2009; Komatsuda et al., 1999; Blattner, 2004, 2006) or multiple alignments (Nishikawa et al., 2002; Kakeda et al., 2009; Sun et al., 2009; Jakob et al., 2010) constructed by ClustalX (Thompson et al., 1997) or PRANK (Löytynoja and Goldman, 2005). We prefer to use direct optimization as a means for constructing phylogenetic hypotheses (Wheeler, 1996) to ensure reproducibility and to avoid the ambiguity of manual sequence alignment and the potential conflict between parameter choice during alignment and phylogenetic reconstruction (Phillips et al., 2000).

Materials and methods

Plant material

Taxon sampling for the present study is as used previously by Petersen and Seberg (2003, 2009) and includes 26 accessions of *Hordeum*, representing all diploid species and subspecies. However, the DNA sample of the accession (H1296) of *H. pubiflorum* previously used was found to be contaminated most likely with DNA from *H. pusillum*, and in the present study all sequences were replaced with sequences from accession H1236 [Argentina, Prov. Chubut, N. Jacobsen 79–3136 (C)]. As outgroups, three accessions representing two species of *Psathyrostachys* Nevski, the most likely sister group to *Hordeum* (Seberg and Petersen, 2007), are included. Voucher information is listed elsewhere (Petersen and Seberg, 2003). In the BLZ1 data set the sequences from *H. bogdanii* and *H. marinum* ssp. *marinum* previously used by us were mixed up (Petersen and Seberg, 2009[1]), but here this error has been corrected. In a few cases sequences already available in GenBank were used instead of producing new ones. Thus, GenBank sequences were used for NAR7 (*H. vulgare*), waxy (*H. bogdanii*, *H. chilense*, *H. pusillum*, *H. stenostachys*), and *ndh*F (*H. brevisubulatum* ssp. *violaceum*, *H. bogdanii*, *H. vulgare* ssp. *vulgare*). For *mat*K, *trn*L-F, and the *atp*B-*rbc*L spacer sequences from Nishikawa et al. (2002) were used except for some produced by ourselves: from *H. brevisubulatum* ssp. *violaceum*, *H. chilense*, and *Psathyrostachys* spp. we sequenced all regions and from *H. cordobense*, *H. vulgare* ssp. *spontaneum*, *H. patagonicum* ssp. *patagonicum*, *H. patagonicum* ssp. *santacrucense*, and *H. patagonicum* ssp. *magellanicum* we only sequenced *trn*L-F. The latter five taxa were not included in the study published by Nishikawa et al. (2002), but nonetheless sequences were deposited in GenBank for *mat*K and the *atp*B-*rbc*L spacer. For *H. marinum* ssp. *gussoneanum*, Nishikawa et al. (2002) have provided sequences from several accessions; one of these (H299) is the same used by us and we included the sequences from this accession. All GenBank accession numbers are listed in Appendix Table A1, as are the authority names of all species and subspecies.

[1]The error has no effect on the conclusions of that paper.

Table 1
New primers used for PCR and sequencing of three nuclear genes: ACL3, NAR7 and VRN3

Gene	Primer name	Primer sequence
ACL3	acl3_E2F	5′-cttcgcctgcagccagtgcc-3′
	acl3_E3R	5′-gtgccttcaggaacagcaagc-3′
NAR7	nar7_E1F	5′-cacgtcgacgccgagctcgccaa-3′
	nar7_E2F	5′-cttggtacaagccggagtgc-3′
	nar7_E2R	5′-ggagtaggcatatcccttcatcg-3′
	nar7_E3R	5′-ccagcaccagtacttgccgtac-3′
VRN3	vrn3_142F	5′-tgcgagctcaagccgtccatgg-3′
	vrn3_1002R	5′-ctggcagttgaagtagacggc-3′

Molecular methods

Total DNA was extracted previously from fresh leaves (Petersen and Seberg, 2003). New primers used for PCR and sequencing are listed in Table 1.

For the nuclear genes new primers were initially designed based on the *Hordeum* sequences M58754 (ACL3), X60173 (NAR7), and DQ900686 (VRN3) available in GenBank. The ACL3 primers are located in exon 2 and exon 3 (Table 1). PCR amplifications were performed at annealing temperatures of 60–68 °C using a standard polymerase (Ampliqon Taq DNA Polymerase, Ampliqon) and the buffer supplied with the kit diluted to a final $MgCl_2$ concentration of 1.5 mM. Most NAR7 sequences were amplified with primers located in exon 1 and exon 3, but some were amplified in two parts using additional primers located in exon 2 (Table 1). PCR was performed as above using annealing temperatures of 65–68 °C for the long product and 60–65 °C for the shorter fragments. VRN3 primers are located in exon 1 and exon 3 (Table 1). PCR was performed as above using annealing temperatures of 65–68 °C. For amplification of waxy we initially used the primer F-for and M-bac (Mason-Gamer et al., 1998) located in exon 9 and exon 14 in *Zea mays* L. Amplification was only successful for some taxa at annealing temperatures between 65 and 68 °C. We attempted to amplify the region in two overlapping fragments using additional primers L1-for and L2-bac, both located in exon 12 (Mason-Gamer et al., 1998) in combination with the above primers. PCR amplification with the primer combination F-for/L2-bac was successful at an annealing temperature of 65 °C, but multiple attempts to amplify the other fragment failed, and only sequences corresponding to the one amplified within the F-for/L2-bac region were subsequently used.

For PCR amplification of the plastid regions only previously published primers were used. The *trn*L-F region was amplified using primers trnLFE/trnLFF (Taberlet et al., 1991) at annealing temperatures of 55–57 °C. The *trn*H-*psb*A region was amplified using primers trnHF/psbA3F (Sang et al., 1997; Tate and Simpson, 2003) at an annealing temperature of 50 °C and with bovine serum albumin added to the reaction. The *psb*I-K and *atp*F-H regions were amplified under conditions as for *trn*H-*psb*A using the primers psbI/psbK and atpF/atpF (psbK, 5′-TTAGCCTTTGT TTGGCAAG-3′; psbI, 5′-AGAGTTTGAGAGTAAG CAT-3′; atpF, 5′-ACTCGCACACACTCCCTTTCC-3′; atpH, 5′-GCTTTTATGGAAGCTTTAACAAT-3′ kindly provided by Ki-Joong Kim, Korea University). The *atp*B-*rbc*L region was amplified using primers atpBF/rbcL17R (our primer designations; sequences from Nishikawa et al., 2002) at an annealing temperature of 55 °C. The *ndh*F gene was partially amplified using primers ndhF1318/ndhF2110R (Olmstead and Sweere, 1994) at an annealing temperature of 50 °C, and the *mat*K gene was amplified using primers matKF/matKR (our primer designations; sequences from Nishikawa et al., 2002) at an annealing temperature of 57 °C. Additional primers labelled by us matK562F, matK572R, matK1061F, and matK1093R (Nishikawa et al., 2002) were subsequently used for sequencing.

All PCR products were purified using the QIAquick PCR purification kit (QIAGEN) according to the manufacturer's instructions. For sequencing the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Mix was used with AmpliTaq DNA Polymerase, FS (PE Biosystems). The products were purified as above. DNA fragments were separated on an ABI 3110 XL (PE Biosystems) automated sequencer and manually edited using Sequencher versions 4.6–4.8.

Phylogenetic analyses

The 16 data sets were analysed using POY version 4.0 (Varón et al., 2008). Whenever applicable, the data sets were separated into blocks corresponding to the intron/exon or gene/intergene structure (Table 2). In three data sets, either previously defined (BLZ1, XYL) or newly discovered (VRN3), miniature transposable inverted-repeat elements (MITEs) belonging to the *Stowaway* family were removed from the sequences prior to the main analyses. The justification for removing the MITE sequences is that these sequences almost per definition may not need to track speciation events. Even if inserted at apparently homologous positions in the genome the elements may not be homologous (Mason-Gamer, 2007; Petersen and Seberg, 2009). Additionally, long inserts/deletions may have a strong influence on phylogenetic hypotheses in direct optimization as implemented in POY (Petersen et al., 2004; Aagesen et al., 2005). However, to test the effect of the MITEs an additional set of analyses was nonetheless performed including the MITEs.

Analyses were run in parallel using 4–10 nodes of the computer cluster Biocluster at the University of Copenhagen (130 lx24-x86 each with a 2.4 GHz processor and 40 quad-processor lx24-amd64 running at 2.4 GHz as well). The 16 data sets were run individually and in various combinations: all data, all nuclear data, all plastid data, all pair-wise combinations of nuclear data, and all pair-wise combinations of plastid data. Searches with more than two data partitions were performed using the strategy *build*(5000) *select*() *swap*(10) *select*(), while for the individual data sets and the pair-wise analyses *build*(1000) *select*() *swap*(10) *select*() were used. All searches were performed under six costs regimes: 11 [=1(0)1], 2(1)2, 3(1)2, 4(1)2, 5(1)2, 6(1)2, where, for example, 3(1)2 means gap opening cost 3, extension gap cost 1, and substitution cost 2. These latter costs are specified in POY using the command *transform* (tcm:(2,1),gap_opening:2).

As an optimality criterion allowing us to select among the most-parsimonious trees derived from analyses using different cost sets (Wheeler, 1995) we used the incongruence length difference (ILD; Mickevich and Farris, 1981; Farris et al., 1995) normalized through division by the length of the combined data cladogram (ILD $= [L_{AB} - (L_A + L_B)]/L_{AB}$, where L_{AB} is the minimum length of the combined data set and L_A and L_B are the minimum length of data sets A and B analysed individually). On the selected trees, the robustness of each clade to differential weighting is shown as sensitivity plots (''Navajo rugs'') (Wheeler, 1995).

To assess branch support, jackknife frequencies and Bremer support values were calculated using POY version 4.1.2.1 according to the program documentation (Varón et al., 2008). Jackknife analyses were carried out using the fragments as characters during character resampling in each pseudoreplicate. We chose this approach as opposed to the recommendation in the documentation because converting the dynamic characters into a static matrix used for resampling would convert long gaps into several independent characters and lead to inflation of support values. Jackknife analyses were run for several days to complete 1000 pseudoreplicates with 33% of the characters being deleted during each pseudoreplicate; *calculate_support* [jackknife:(remove:0.33, resample:1000)]. Bremer support values were calculated using a two-step procedure as described in the documentation. First, 200 random addition trees were built; *build*(200). Then each tree was swapped for a maximum of 3600 s saving all trees in a temporary file; *swap*(all,visited:''tmp.trees'', timeout:3600), and finally the commands *select*() and *report*(''my.tree'',trees) select and store the optimal trees in a separate file. In the second step, the output files

Table 2
Information on individual data sets and the chromosomal location of sequences

Data set	Chromosome	Length (bp)*	Sequence fragments used for analyses
ACL3	1 (7H)	246–638	1; only intron 2†
Waxy	1 (7H)	787–800	7; 4 exons, 3 intron
EF-G	2 (2H)	346–373	1
XYL	2 (2H)	783–817	5; 2 exons, 3 introns‡
DMC1	3 (3H)	1875–2118	12; 6 exons, 6 introns§
NUC	4 (4H)	718–795	6; 3 exons, 3 introns¶
BLZ1	7 (5H)	768–919	3; 2 exons, 1 intron**
NAR7	6 (6H)	519–612	4; 2 exons, 2 introns††
VRN3	1 (7H)	748–1001	4; 2 exons, 2 introns‡‡
*ndh*F	plastid, SSC	777	1
*mat*K	plastid, LSC	1530–1536	1
*trn*L-F	plastid, LSC	319–345	1
*atp*F-H	plastid, LSC	(312)440–459	1; only intergenic region included§§
*psb*I-K	plastid, LSC	405–412	1; only intergenic region¶¶
*trn*H-*psb*A	plastid, LSC	(230)259–277	2; only intergenic regions included***
*atp*B-*rbc*L	plastid, LSC	2210–2224	2; intergenic region and *rbc*L

Notes describe particular sequence features, excluded sequence fragments, and missing (due to sequencing problems) sequences if covering entire fragments or assumed to be longer than 100 bp. LSC, large single-copy region, SSC, small single-copy region.

*Disregarding missing leading and trailing sequence. Numbers in parentheses indicates outgroup length if it varies significantly from ingroup length.

†Exon 2 (25 bp) and exon 3 (47 bp) sequences are excluded due to limited variation.

‡Exon 6 (1 bp) and exon 9 (4 bp) without information; excluded. 5′-end of intron 6 trimmed (maximum 27 bp) because of missing data in many taxa. Intron 6 includes a *Stowaway* MITE (66–71 bp) in all species of *Hordeum* (Petersen and Seberg, 2009).

§Exon 10 (6 bp) and exon 15 (5 bp) without information; excluded. Entire exon 1 sequence (11 bp) and part of intron 1 sequence (ca. 350 bp) are lacking from *Psathyrostachys fragilis* ssp. *fragilis*. From *P. stoloniformis* ca. 170 bp are missing from intron 1.

¶Exon 3 (36 bp) excluded due to missing data in many taxa. Sequence from *H. stenostachys* is entirely lacking. Intron 3 (ca. 100 bp) sequence data are entirely lacking from *H. marinum* ssp. *gussoneanum*, *H. brachyantherum* ssp. *californicum*, and *Psathyrostachys fragilis* ssp. *villosus*. Exon 6 (55 bp) sequence data are entirely lacking from *H. vulgare* ssp. *spontaneum*, *H. flexuosum*, and *Psathyrostachys fragilis* ssp. *villosus*.

**Intron 3 includes two assumed non-homologous *Stowaway* MITEs. One MITE is present in four species of *Hordeum*, the other in 22 species of *Hordeum* (Petersen and Seberg, 2009).

††Exon 1 (7 bp) is invariable; excluded. For *H. bulbosum* sequence is only available for exon 3 and a minor fraction (7 bp) of intron 2.

‡‡Exon 1 (38 bp) is excluded. Exon 2 (62 bp) and a major part of intron 1(ca. 300 bp) is missing from *Psathyrostachys stoloniformis*. Intron 1 includes a *Stowaway* MITE (183–243 bp) in 15 species of *Hordeum*.

§§*atp*F (132 bp) invariable, *atp*H (52 bp) with one character separating the ingroup from the outgroup.

¶¶*psb*K (20 bp) includes no variation, *psb*I not sequenced.

***trn*H (30 bp) and *psb*A (53 bp) include no information. *rps*19 (282 bp) includes only characters separating the ingroup from the outgroup.

generated in the first step are use for calculating the final support values using the command *report*(''support tree.pdf'', graphsupports: bremer: ''tmp.trees'').

Results

The data sets vary considerably in length from < 300 bp to more than 2200 bp (Table 2). Some sequenced regions (mostly protein coding) were excluded from the analyses due to very limited (mostly autapomorphic) or complete lack of variation. Details are given in Table 2. The DMC1 data set is composed of two non-contiguous fragments, namely a region including intron 1 and some flanking exon sequences and a region spanning from exon 10 to exon 15. A complete, supposedly monocistronic DMC1 gene sequence is available from *Hordeum vulgare* (GenBank accession no. AF234170) and we consider it likely that the DMC1 gene is structured likewise in all *Hordeum* species. Hence, we treat the two fragments as one data set.

Tree lengths from both individual and combined analyses and ILD values under the six different cost sets are listed in Table 3. Results from pair-wise analyses of the nuclear and the plastid genes are listed in Tables 4 and 5, respectively.

Based on the ILD values the highest congruence among all data sets (excluding the MITEs) is achieved at costs 3(1)2 and the tree derived from combined analysis under this cost set is shown in Fig. 1. When MITEs are included the optimal cost set is 4(1)2. The highest congruence among the nuclear data sets is achieved at costs 5(1)2 both with and without inclusion of *Stowaway* MITEs (Figs 2 and 4), and for the

Table 3
Tree lengths from individual and combined analyses of nine nuclear and seven plastid data sets under six different sets of costs

										Nuclear combined	
Cost set	ACL3	Waxy	EF-G	XYL	DMC1	NUC	BLZ1	NAR7	VRN3	L	ILD
11	653	321	129	450	997	474	598	376	511	4886	0.0772
2(1)2	922	606	227	791	1594	767	914	594	822	7768	0.0684
3(1)2	987	635	237	834	1676	821	982	644	922	8274	0.0648
4(1)2	1045	655	246	877	1753	862	1028	689	993	8722	0.0658
5(1)2	1092	676	255	919	1830	907	1078	732	1069	9148	**0.0645**
6(1)2	1130	695	264	960	1897	951	1125	774	1127	9547	0.0654

								Plastid combined		All combined	
Cost set	*ndh*F	*mat*K	*trn*L-F	*atp*F-H	*psb*I-K	*trn*H-*psb*A	*atp*B-*rbc*L	L	ILD	L	ILD
11	55	83	87	194	57	79	88	688	**0.0654**	5645	0.0873
2(1)2	110	161	135	228	107	112	162	1089	0.0680	8983	0.0814
3(1)2	112	162	158	240	117	122	175	1163	0.0662	9577	**0.0786**
4(1)2	112	163	179	252	126	129	188	1230	0.0659	10100	0.0795
5(1)2	112	164	200	264	134	136	200	1295	0.0656	10610	0.0794
6(1)2	122	165	204	274	142	143	212	1357	0.0774	11083	0.0812

ILD values for combined analyses, with lowest values marked in bold.

plastid data sets congruence is highest at costs 11 (Fig. 3).

Discussion

In the discussion below of species relationships we will compare the present results with relationships suggested by previous phylogenetic analyses including other sources of data (e.g. Doebley et al., 1992; Blattner, 2004, 2006; Pleines and Blattner, 2008; Kakeda et al., 2009; Sun et al., 2009; Jakob et al., 2010), whereas previous analyses based on single or a few of the data partitions also included in the present study will only rarely be cited (e.g. Komatsuda et al., 1999; Nishikawa et al., 2002; Petersen and Seberg, 2003, 2009).

Even though a recent study of 5S rDNA spacer sequences (Baum et al., 2010) uses data complementing those included in the present study, we will refrain from commenting on the relationships suggested by it. In addition to the general problem of retrieving reliable sequence data from multiple-copy sequences prone to PCR recombination using a cloning approach (e.g. Cronn et al., 2002; Petersen and Seberg, 2004, 2005), we have previously (Petersen and Seberg, 2004) criticized the analytical approach based on the concept of sequence "unit classes" used by Baum et al. (2010; and papers cited therein). Prior to phylogenetic analysis, Baum et al. (2001, 2010) divide obtained sequence copies into "unit classes" based on repeated cycles of alignment manipulations, thus deciding *a priori* which sequences are homologous and which are not. To their credit, Baum et al. (2010. p. 193) cite our previous critique (Petersen and Seberg, 2004; p. 871), and apart from—understandably—expressing disagreement with our interpretation, present no evidence in favour of their point of view. On the contrary, we can easily extend our critique. The goal of reproducibility—the core of all scientific investigations—remains persistently unattainable by repeatedly manipulating the data and aligning and re-aligning them; and not publishing the alignments. Thus, "unit classes" are purely artificial, phenetic constructs and even though primary hypotheses of orthology (= homology) may be made in this (or any other) way the "unit classes" are immune to testing. Additionally, by using nearly every available standard methodology for analysing data, even UPGMA, but strangely enough not Bayesian inference, all conclusions are distorted and difficult to follow, especially as the different methodologies have very different interpretations of homology—if any at all. Such an approach is ill justified as there are "an infinite number of possible methods that could collectively yield any possible topology" (Brower, 2000, p. 148). Needless to say we have no confidence in the conclusion (Baum et al., 2010, fig. 3), which is largely at odds with all previous findings based on other types of sequences and less esoteric analyses.

Incongruence among nuclear genes

Pair-wise comparisons of the nine nuclear genes show that the EF-G data set in general is the one with the highest congruence with other data sets irrespective of cost set. Five data sets, ACL3, BLZ1, DMC1, NAR7, and VRN3, are always most congruent with EF-G, two data sets, NUC and XYL, are most congruent with EF-G under four or three cost sets, respectively, and under a single cost set, waxy, is most congruent with EF-G (Table 4). For every cost set the globally lowest ILD value

Table 4
Tree lengths (below diagonal) and ILD values (above diagonal) from pairwise analyses of nine nuclear data sets under six different cost sets

	ACL3	BLZ1	DMC1	EF-G	NAR7	NUC	VRN3	Waxy	XYL
11									
ACL3	–	0.029	0.092	**0.016**	0.029	**0.017**	0.038	0.043	0.033
BLZ1	1288	–	0.027	**0.018**	0.041	0.041	0.040	0.039	0.047
DMC1	1817	1639	–	**0.006**	0.017	0.022	0.019	**0.021**	0.028
EF-G	795	740	1133	–	**0.004**	0.024	**0.017**	0.028	0.024
NAR7	1060	1016	1397	507	–	0.022	0.019	0.031	0.026
NUC	1146	1118	1504	618	869	–	0.020	0.033	0.021
VRN3	1210	1155	1537	651	904	1005	–	0.030	**0.020**
Waxy	1018	956	1346	463	719	822	858	–	0.039
XYL	1141	1100	1488	593	848	944	981	802	–
2(1)2									
ACL3	–	0.034	0.054	**0.010**	0.036	0.032	0.051	0.052	0.043
BLZ1	1900	–	0.032	**0.018**	0.051	0.043	0.053	0.045	0.053
DMC1	2661	2591	–	**0.007**	0.024	0.023	0.029	**0.022**	0.029
EF-G	1161	1162	1833	–	**0.004**	0.021	**0.017**	0.023	**0.021**
NAR7	1573	1589	2242	824	–	**0.020**	0.022	0.036	0.034
NUC	1745	1757	2416	1015	1389	–	0.027	0.034	0.023
VRN3	1838	1834	2487	1067	1448	1633	–	0.031	0.024
Waxy	1612	1591	2249	853	1245	1422	1474	–	0.041
XYL	1790	1801	2457	1040	1434	1595	1652	1456	–
3(1)2									
ACL3	–	0.035	0.044	**0.011**	0.038	0.027	0.044	0.046	0.040
BLZ1	2041	–	0.038	**0.019**	0.051	0.037	0.059	0.050	0.057
DMC1	2786	2762	–	**0.006**	0.025	0.022	0.028	0.023	0.032
EF-G	1237	1243	1924	–	**0.001**	**0.017**	**0.016**	**0.021**	0.023
NAR7	1695	1713	2379	882	–	0.021	0.019	0.038	0.037
NUC	1858	1872	2553	1076	1496	–	0.025	0.032	**0.021**
VRN3	1997	2024	2674	1178	1597	1788	–	0.030	0.021
Waxy	1700	1703	2365	891	1329	1504	1605	–	0.039
XYL	1896	1925	2592	1096	1534	1690	1795	1529	–
4(1)2									
ACL3	–	0.037	0.034	**0.011**	0.036	0.032	0.045	0.046	0.039
BLZ1	2153	–	0.039	**0.020**	0.052	0.042	0.061	0.046	0.057
DMC1	2897	2893	–	**0.006**	0.022	0.024	0.030	**0.023**	0.030
EF-G	1305	1300	2011	–	**0.003**	**0.018**	**0.016**	0.026	0.023
NAR7	1799	1812	2498	938	–	0.018	0.019	0.037	0.038
NUC	1970	1972	2680	1128	1579	–	0.026	0.034	**0.022**
VRN3	2134	2152	2832	1259	1715	1904	–	0.029	0.023
Waxy	1782	1764	2464	925	1395	1570	1697	–	0.041
XYL	2000	2019	2712	1149	1628	1779	1913	1597	–
5(1)2									
ACL3	–	0.038	0.033	**0.011**	0.034	0.034	0.045	0.044	0.039
BLZ1	2255	–	0.038	**0.021**	0.051	0.049	0.058	0.047	0.058
DMC1	3021	3024	–	**0.006**	0.024	0.024	0.029	**0.021**	0.031
EF-G	1362	1361	2089	–	**0.006**	**0.018**	**0.016**	0.027	**0.023**
NAR7	1888	1908	2624	993	–	0.019	0.018	0.036	0.038
NUC	2070	2087	2804	1183	1671	–	0.026	0.033	0.024
VRN3	2262	2279	2985	1345	1834	2028	–	0.027	0.024
Waxy	1849	1840	2561	957	1461	1637	1794	–	0.041
XYL	2093	2119	2836	1201	1717	1870	2036	1663	–
6(1)2									
ACL3	–	0.036	0.030	**0.011**	0.035	0.035	0.050	0.045	0.039
BLZ1	2339	–	0.037	**0.021**	0.049	0.047	0.062	0.046	0.058
DMC1	3122	3138	–	**0.006**	0.025	0.023	0.032	**0.022**	0.032
EF-G	1409	1419	2175	–	**0.008**	**0.019**	**0.016**	0.027	**0.021**
NAR7	1973	1996	2739	1046	–	0.019	0.017	0.036	0.039
NUC	2156	2178	2916	1238	1759	–	0.025	0.033	0.025
VRN3	2376	2402	3124	1413	1934	2131	–	0.027	0.025
Waxy	1910	1908	2649	986	1524	1702	1872	–	0.042
XYL	2174	2212	2951	1250	1804	1959	2140	1727	–

For each data set the combination yielding the lowest ILD is marked in bold.

Table 5
Tree lengths (below diagonal) and ILD values (above diagonal) from pairwise analyses of seven plastid data sets under six different cost sets

	*atp*B-*rbc*L	*atp*F-H	*mat*K	*ndh*F	*psb*I-K	*trn*H-*psb*A	*trn*L-F
11							
*atp*B-*rbc*L	–	**0.024**	0.039	0.040	0.040	0.051	0.033
*atp*F-H	289	–	0.028	0.012	**0.020**	**0.014**	**0**
*mat*K	178	285	–	**0.007**	0.060	0.069	0.045
*ndh*F	149	252	139	–	0.043	0.063	0.034
*psb*I-K	151	256	149	117	–	0.042	0.040
*trn*H-*psb*A	176	277	174	143	142	–	0.046
*trn*L-F	181	281	178	147	150	174	–
2(1)2							
*atp*B-*rbc*L	–	**0.020**	0.036	0.032	0.032	0.035	0.033
*atp*F-H	398	–	0.030	0.017	0.026	**0**	**0.003**
*mat*K	335	401	–	**0.007**	0.066	0.055	0.048
*ndh*F	281	344	273	–	0.040	0.043	0.028
*psb*I-K	278	344	287	226	–	**0.022**	0.047
*trn*H-*psb*A	284	340	289	232	224	–	0.031
*trn*L-F	307	364	311	252	254	255	–
3(1)2							
*atp*B-*rbc*L	–	**0.019**	0.037	0.037	0.036	0.042	0.035
*atp*F-H	423	–	0.029	0.022	0.022	**0**	**0**
*mat*K	350	414	–	**0.007**	0.057	0.053	0.048
*ndh*F	298	360	276	–	0.038	0.041	0.025
*psb*I-K	303	365	296	238	–	**0.017**	0.042
*trn*H-*psb*A	310	362	300	244	243	–	0.031
*trn*L-F	345	398	336	277	287	289	–
4(1)2							
*atp*B-*rbc*L	–	**0.018**	0.041	0.042	0.031	0.048	0.039
*atp*F-H	448	–	0.028	0.022	0.021	**0**	**0**
*mat*K	366	427	–	**0.007**	0.052	0.052	0.045
*ndh*F	313	372	277	–	0.033	0.040	0.027
*psb*I-K	324	386	305	246	–	**0.015**	0.032
*trn*H-*psb*A	333	381	308	251	259	–	0.034
*trn*L-F	382	431	358	299	315	319	–
5(1)2							
*atp*B-*rbc*L	–	**0.017**	0.045	0.046	0.032	0.051	0.040
*atp*F-H	472	–	0.027	0.021	0.020	**0**	**0**
*mat*K	381	440	–	**0.007**	0.051	0.051	0.040
*ndh*F	327	384	278	–	0.031	0.039	0.022
*psb*I-K	345	406	314	254	–	**0.015**	0.026
*trn*H-*psb*A	355	400	316	258	274	–	0.029
*trn*L-F	418	464	379	319	343	346	–
6(1)2							
*atp*B-*rbc*L	–	**0.016**	0.048	0.050	0.033	0.058	0.082
*atp*F-H	494	–	0.027	0.020	0.019	**0**	**0.032**
*mat*K	396	451	–	**0.007**	0.050	0.049	0.078
*ndh*F	341	394	279	–	0.031	0.038	0.068
*psb*I-K	366	424	323	262	–	**0.014**	0.067
*trn*H-*psb*A	377	417	324	265	289	–	0.070
*trn*L-F	453	494	400	339	371	373	–

For each data set the combination yielding the lowest ILD is marked in bold.

is also obtained in a comparison involving the EF-G data set (Table 4). The most likely explanation is the low phylogenetic information content of the EF-G data set, reflected by the relatively short tree length obtained in analyses of this data set alone (Table 3). A data set completely without variation is congruent with any other data set, but congruence is then trivial.

In general, genes located on the same chromosomes do not appear to be more congruent than genes on different chromosomes (Table 4). For XYL and EF-G, both located on chromosome 2, the EF-G data set is always more congruent with several other data sets than XYL, and for XYL three out of six cost sets show higher congruence with other genes than EF-G (Table 4). For ACL3, VRN3, and waxy, all located on chromosome 1, VRN3 and waxy are more congruent with each other than any of them are with ACL3, but the level of congruence is not higher than in comparisons with other genes. It is to be expected that linked characters are more congruent than unlinked characters if

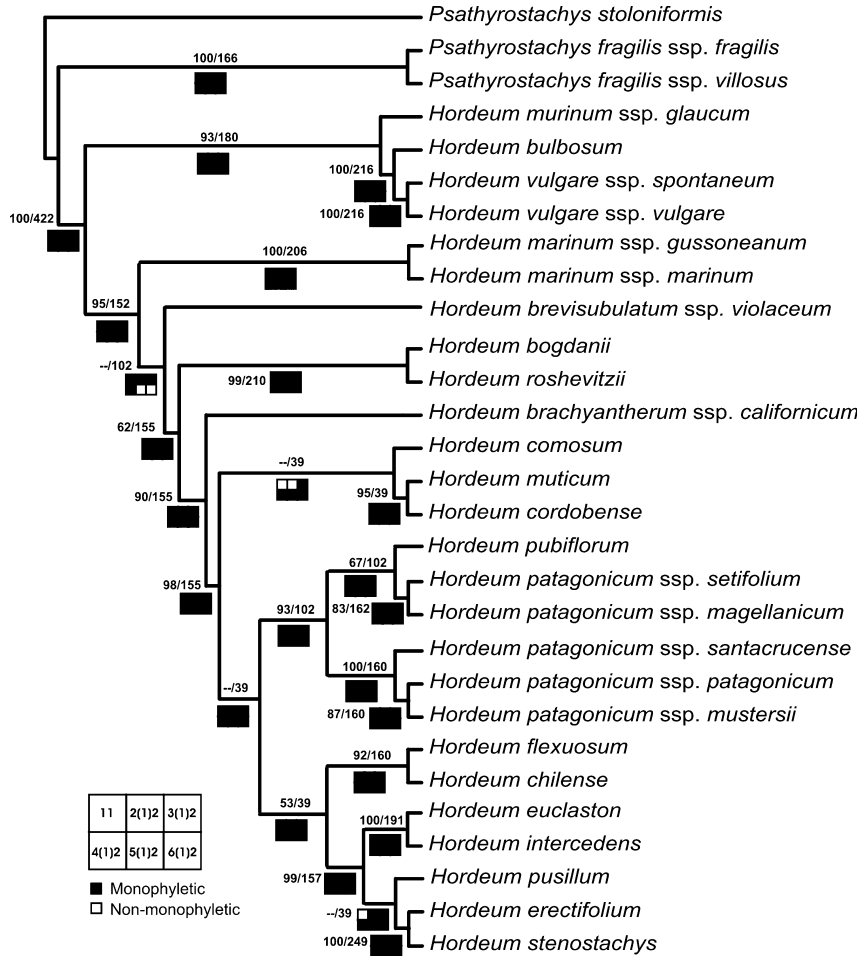

Fig. 1. Phylogenetic tree of the diploid taxa of *Hordeum* based on combined analysis of nine nuclear and seven plastid sequence regions. Four *Stowaway* MITEs have been excluded from the nuclear sequences. Robustness of clades to differential weighting is shown as sensitivity plots (''Navajo rugs''). Numbers are jackknife/Bremer support values.

hybridization or introgression has been instrumental during evolution of species (Linder and Rieseberg, 2004). Although location on the same chromosome is certainly not equal to being closely linked, loci on the same chromosome are surely more closely linked than loci on different chromosomes. Hence, the present data at least do not lend support to any hypotheses about hybridization or introgression as a possible explanation of the incongruence, leaving incomplete lineage sorting as a more likely explanation (Wendel and Doyle, 1998). The present data also suggest that the nuclear genes evolved independently from each other and that the phylogenetic inference using nuclear genes is unbiased by chromosomal location.

The individual nuclear gene trees [under optimal costs, 5(1)2] are moderately to well resolved, having from 11 (EF-G) to 21 (ACL3) out of 25 possible nodes, respectively. However, the only consistently resolved nodes include the two subspecies of *Hordeum vulgare* and the two subspecies of *H. marinum*.

Incongruence among plastid genes

Pair-wise comparisons of the seven plastid sequence regions show that irrespective of cost set the *atp*F-H data set generally exhibits the highest congruence with other data sets, and that the *atp*F-H data set is completely congruent with both the *trn*L-F and the *trn*H-*psb*A data sets under four and five cost sets, respectively (Table 5). In this case, congruence is not influenced by the low character information content of the *atp*F-H region; on the contrary, it yields the longest single plastid gene trees (Table 3). However, a substantial part of the tree length is attributable to length difference (minimum 128 bp) between the ingroup and the outgroup, and thus the amount of character information within *Hordeum* is generally not higher than for the other regions.

Location in the small, single-copy region of the plastids (*ndh*F) versus location in the large, single-copy region (all other regions) has no influence on congruence

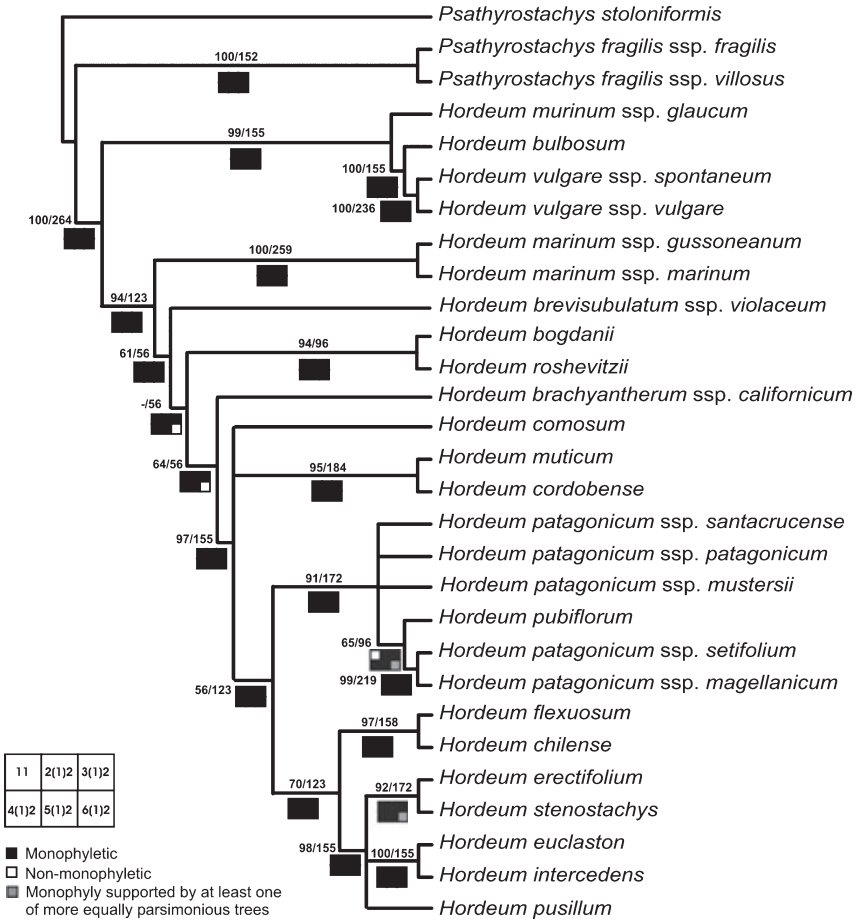

Fig. 2. Phylogenetic tree of the diploid taxa of *Hordeum* based on combined analysis of nine nuclear sequence regions. Four *Stowaway* MITEs have been excluded from the nuclear sequences. Robustness of clades to differential weighting is shown as sensitivity plots (''Navajo rugs''). Numbers are jackknife/Bremer support values.

and neither does sequence proximity (see Table 5): the *atp*B-*rbc*L region and the *trn*L-F region are located just some 6 kb apart and the *psb*I-K, *mat*K, and *trn*H-*psb*A regions are all located within a ca. 8-kb region, whereas these two sequence regions are separated by some 40 kb in the plastid genome of *H. vulgare* (GenBank NC_008590).

The two data sets exclusively containing protein coding sequence, *ndh*F and *mat*K, also have a very high level of congruence (Table 5). Thus, there is a trend towards congruence being correlated with the protein coding versus non-protein coding nature of the sequences in the data sets. A possible explanation lies in the different evolutionary patterns of these two sequence categories, as length mutations are much rarer in protein coding sequences.

The individual gene trees [under optimal costs, 1(0)1] are mostly very unresolved including only from six (*atp*F-H, *psb*I-K) to 13 (*atp*B-*rbc*L) of the possible 25 resolved nodes. Except for monophyly of the genus *Hordeum* and the two subspecies of *H. vulgare*, no other clades are consistently recovered.

Plastid versus nuclear genes

The level of congruence among the plastid sequence regions is largely comparable with the level found among the nuclear regions (Tables 4 and 5), despite the plastid regions belonging to the same linkage group, whereas the nuclear regions are located on six different chromosomes.

The topologies of the combined nuclear gene tree (Fig. 2) and the combined plastid gene tree (Fig. 3) differ in several areas, but there are also areas of agreement. A sister-group relationship between *H. vulgare* and *H. bulbosum* is strongly supported both by nuclear and plastid data (Figs 2 and 3) as well as by all previous analyses (Doebley et al., 1992; Blattner, 2004, 2006; Kakeda et al., 2009; Sun et al., 2009; Jakob et al., 2010). Nuclear and plastid data also agree on a

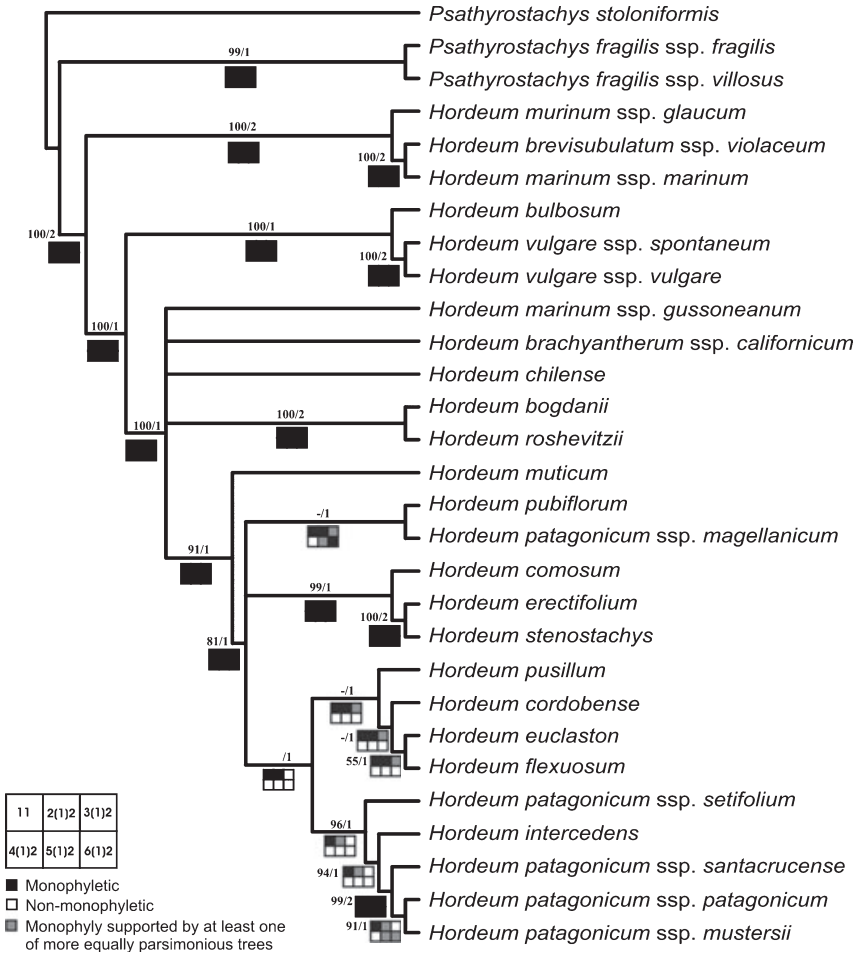

Fig. 3. Phylogenetic tree of the diploid taxa of *Hordeum* based on combined analysis of seven plastid sequence regions. Robustness of clades to differential weighting is shown as sensitivity plots (“Navajo rugs”). Numbers are jackknife/Bremer support values.

sister-group relationship between *H. bogdanii* and *H. roshevitzii* (Figs 2 and 3) and this relationship seems robustly recovered by the majority of other data (Doebley et al., 1992; Blattner, 2004; Kakeda et al., 2009). Only one more species pair (*H. erectifolium* + *H. stenostachys*) is recovered by both nuclear and plastid data (Figs 2 and 3).

One of the major inconsistencies between the trees derived from nuclear and plastid data concerns the basal split within *Hordeum*. Plastid data strongly support a split between a group consisting of *H. murinum*, *H. brevisubulatum*, and one subspecies of *H. marinum* and a group including all other taxa (Fig. 3). Nuclear data equally strongly support a group consisting of *H. vulgare*, *H. bulbosum* and *H. murinum* as the sister to the remaining species (Fig. 2). Below we discuss some of the suggested species relationships in further detail, and in general it appears as if nuclear data more closely reflect relationships implied by other often non-discrete data (e.g. morphology, chromosome pairing data).

Hordeum murinum

Whereas *H. murinum* is firmly placed as sister group to *H. vulgare* and *H. bulbosum*, both by the combined analysis and by analysis of the nuclear data only, the plastid data equally firmly place *H. murinum* in a clade with *H. marinum* ssp. *marinum* and *H. brevisubulatum*. Previous analyses of nuclear sequence data, the multicopy array ITS and three single-copy genes [thioredoxin-like gene (HTL), RNA polymerase II (RPB2), topoisomerase 6 (Topo6)], also placed *H. murinum* as sister to *H. vulgare* and *H. bulbosum* (Blattner, 2004; Kakeda et al., 2009; Sun et al., 2009; Jakob et al., 2010). These three species together comprise *H.* sect. *Hordeum* and share the presence of long, clasping auricles—one of the very few well-defined morphological traits in *Hordeum*. Thus, the nuclear sequences track the supposed, traditionally defined species phylogeny more precisely than the plastid sequences and incomplete lineage sorting may be the most likely explanation.

Hordeum brevisubulatum

The exact position of the Asiatic species *H. brevisubulatum* is perhaps the most controversial problem within *Hordeum* phylogenetics. The combined analysis places *H. brevisubulatum* as sister to a clade including two other Asiatic species, *H. bogdanii* and *H. roshevitzii*, plus the clade including all the American species (Fig. 1). Using a taxonomically reduced data set this position was also recovered in an analysis of nuclear HTL sequence data (Kakeda et al., 2009). However, the position is neither robust to cost changes nor to inclusion of MITE sequences (see below). A gap opening cost of 5 places *H. brevisubulatum* as sister to the American species, *H. roshevitzii*, *H. bogdanii*, plus *H. marinum*, and a gap opening cost of 6 places it as sister to *H. marinum* (trees not shown). The nuclear data analysed alone (Fig. 2) agree with the combined analysis, and only at the highest gap opening cost (6) does *H. brevisubulatum* change position to a sister-group relationship with *H. brachyantherum* (tree not shown). Other nuclear data analyses have placed *H. brevisubulatum* in a clade with *H. bogdanii* and *H. roshevitzii* (ITS; Blattner, 2004), with *H. bogdanii* alone (RPB2; Sun et al., 2009), or as sister to the American species (DMC1 + Topo6; Jakob et al., 2010). These relationships are not found by any of our analyses, including the nine individual nuclear gene analyses under the optimal costs set 5(1)2 (not shown). In contrast to the nuclear data, the plastid sequences consistently place *H. brevisubulatum* as sister to *H. marinum* ssp. *marinum*.

With the position of *H. brevisubulatum* as sister to *H. bogdanii*, *H. roshevitzii*, and the American species (Fig. 1) the four genomic groups proposed by Bothmer and Jacobsen proposed in a series of papers based on genome analysis all become monophyletic (e.g. Bothmer and Jacobsen, 1991). Bothmer et al. suggested that the diploid species of *Hordeum* could be divided into the I genome group (*H. vulgare* and *H. bulbosum*), the Xu genome group (*H. murinum*), the Xa genome group (*H. marinum*), and the H genome group (the remaining diploid species) (genome designations according to Wang et al., 1996). Blattner (2004, 2006) using ITS sequence data alone or in combination with other nuclear sequences also found that the four groups were monophyletic, as did Kakeda et al. (2009) using nuclear HTL sequence data.

We have previously placed *H. brevisubulatum* in *H.* sect. *Stenostachys* together with *H. marinum* and the two allotetraploid species *H. capense* Thunb. and *H. secalinum* Schreb.; both apparently of hybrid origin involving *H. marinum* and *H. brevisubulatum* (Petersen and Seberg, 2003, 2004). Given the results from the present analyses of both the combined data set and the nuclear data, only *H.* sect. *Stenostachys* is not monophyletic. Blattner (2009) suggests that *H. brevisubulatum* should be included in what he recognizes as *H.* series *Sibirica* Nevski, which otherwise only includes *H. roshevitzii* and *H. bogdanii*. However, circumscribed in this manner *H.* series *Sibirica* is not monophyletic either. Monophyly at the sectional level can only be retained by merging *H.* sect. *Stenostachys*, *H.* sect. *Sibirica*, and *H.* sect. *Critesion* (the American species) sensu Petersen and Seberg (2003), which will result in only two sections of *Hordeum* being recognized corresponding to the subgenera *Hordeum* and *Hordeastrum* (Doell) Rouy sensu Blattner (2009).

In it questionable whether addition of more sequence data, nuclear as well plastid, will increase the robustness of the phylogenetic position of *H. brevisubulatum*. Collection of sequence data from additional specimens including polyploids may be more rewarding.

Hordeum marinum

Another striking difference between the plastid and nuclear trees concerns the position of the two subspecies of *H. marinum*. The nuclear genes strongly support monophyly of the two taxa under all costs (Fig. 2), and so do other nuclear sequence data, namely HTL, RPB2, Topo6 + DMC1, and ITS (Blattner, 2004; Kakeda et al., 2009; Sun et al., 2009; Jakob et al., 2010). In contrast, they are widely separated by the plastid sequences, which place *H. marinum* ssp. *gussoneanum* in an unresolved clade with the American species plus *H. bogdanii* and *H. roshevitzii* (Fig. 3). At gap opening costs of 4 or more the clade is resolved in a manner which places *H. marinum* ssp. *gussoneanum* as the sister to the American species. Due to the generally accepted taxonomic status of the two taxa as conspecific and due to their pronounced morphological similarity (e.g. Bothmer et al., 1995) we consider the phylogenetic relationship suggested by the nuclear data most convincing, and we have previously explained the deviant plastid result as lineage sorting and will not repeat the arguments here (Petersen and Seberg, 2003).

The American species

A robust and well-supported clade including all the American species (= *H.* sect. *Critesion* sensu Petersen and Seberg, 2003) is recovered by the combined analysis (Fig. 1). The clade is also recovered by the nuclear data, although with less support and slightly less robustness (Fig. 2). An analysis of ITS data from the diploid species by Blattner (2004) also resulted in a monophyletic group of American species. Under optimal costs, 11, the plastid data alone do not recover this clade, although it is not contradicted either (Fig. 3). At gap opening costs of 4 or higher the clade is recovered. In our previous analysis based on considerably less data, a monophyletic group of American species was indeed

recovered, but the resolution within the clade was poor (Petersen and Seberg, 2003). The present combined analysis resolves the relationships within the clade almost completely, allowing a more detailed discussion of species relationships.

H. brachyantherum is robustly located as sister to the remaining species within the clade both by the combined analysis and by the nuclear data on their own (Figs 1 and 2). The plastid gene tree does not contradict this relationship (Fig. 3), which is recovered at gap opening costs of 4 or higher. A sister-group relationship between North American, diploid *H. brachyantherum*, and the remaining American taxa have been recovered previously based on analyses of various nuclear data sets (Blattner, 2006; Jakob et al., 2010).

A clade consisting of the perennial South American species *H. comosum*, *H. muticum*, and *H. cordobense* is not robust to lower gap opening costs, lacks jackknife support (Fig. 1), and it is strongly contradicted by the plastid data, which robustly and with high jackknife support place *H. comosum* as the sister to two other South American species, *H. erectifolium* and *H. stenostachys* (Fig. 3). In contrast, traditional classification based mostly on morphology has placed *H. comosum* in a group together with *H. pubiflorum* and some polyploid species (Baden and Bothmer, 1994). A recent neighbour-joining analysis of amplified fragment length polymorphism (AFLP) data from the diploid American *Hordeum* species recovered a *H. muticum* – *H. cordobense* clade, but excluding *H. comosum* (Pleines and Blattner, 2008). Parsimony analysis of the same data left the relationships of the three species unresolved (Pleines and Blattner, 2008), thus not contradicting that they might form a clade. However, AFLPs, just like RFLPs, are burdened with a number of problems when used to reconstruct phylogeny, e.g. difficulties in homology assignment and an asymmetric probability of gaining and losing bands (Althoff et al., 2007; Graves, 2009), thus considerably limiting the value of hypotheses based on this type of data. A combined analysis of DMC1 and Topo6 sequence data using Bayesian inference confirmed a sister-group relationship between *H. cordobense* and *H. muticum*, but placed *H. comosum* as the sister to *H. patagonicum* and *H. pubiflorum* (Jakob et al., 2010). Thus, the sister-group relationship between *H. muticum* and *H. cordobense* seems firmly established by nuclear data, whereas the relationships of *H. comosum* must still be considered to be uncertain.

In none of the previous analyses including more specimens of the perennial South American *H. patagonicum* have they been recovered as monophyletic (Petersen and Seberg, 2003; Blattner, 2004, 2006; Sun et al., 2009). In the present combined analysis all five subspecies are included in a very robust and strongly supported clade, but which also includes another perennial South American species, *H. pubiflorum* (Fig. 1). The nuclear

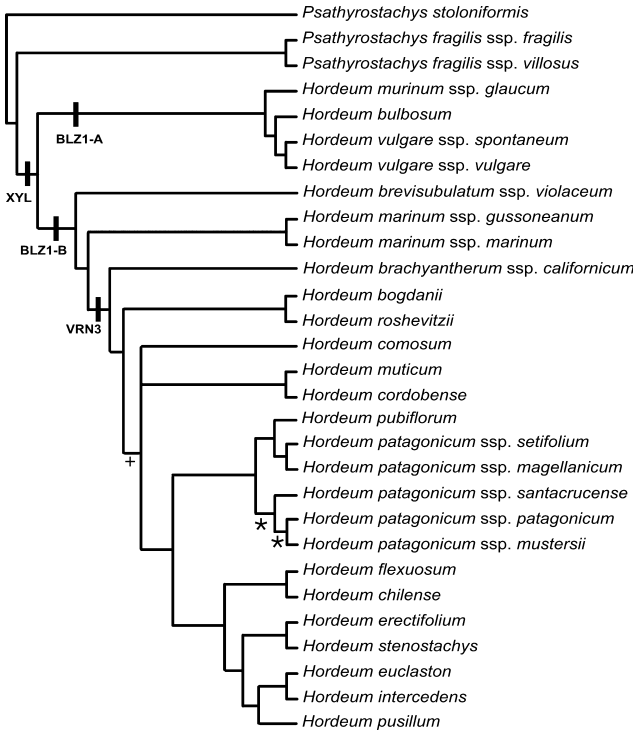

Fig. 4. Phylogenetic tree of the diploid taxa of *Hordeum* based on combined analysis of seven plastid and nine nuclear sequence regions, including MITE sequences. Insertion of MITEs is indicated as black bars on the tree. Branches that are collapsed in the consensus tree derived from combined analysis including only the nuclear regions (+ MITEs) are marked with an asterisk. The trichotomy marked with a plus is resolved in the combined nuclear gene tree, placing *H. comosum* as sister to *H. muticum* + *H. cordobense*.

data provide exactly the same result (Fig. 2), whereas the plastid data with little robustness and support recover four of the subspecies in a clade together with *H. intercedens* and *H. patagonicum* ssp. *magellanicum* in another clade with *H. pubiflorum* (Fig. 3). A close relationship between *H. pubiflorum* and at least some specimens of *H. patagonicum* have been found previously (Blattner, 2004, 2006; Pleines and Blattner, 2008; Sun et al., 2009) and some *trn*L-F haplotypes are shared between specimens of two species (Jakob and Blattner, 2006; Jakob et al., 2009). Phylogenetic analysis including additional specimens of *H. pubiflorum* in particular is needed to elucidate whether the two species should be considered conspecific.

Among the remaining seven diploid American species, three species pairs are robustly recovered and strongly supported by both the combined analysis (Fig. 1) and the nuclear data (Fig. 2). The sister-group relationships between the perennial South American species *H. flexuosum/H. chilense* and *H. erectifolium/H. stenostachys* were previously also recovered by analyses of AFLPs (Pleines and Blattner, 2008), although the latter relationship was not recovered in a parsimony analysis,

only by neighbour joining. The former relationship was also recovered by combined analysis of Topo6 and DMC1 sequence data (Jakob et al., 2010). The sister-group relationship between the annual species *H. intercedens* from California/Baja California and *H. euclaston* from South America was also recovered by Pleines and Blattner (2008) and Jakob et al. (2010), and the two species have a pronounced morphological similarity (Bothmer et al., 1995). Based also on morphological similarity, *H. pusillum*, an annual species widely distributed in the USA (but not found in California), is expected to be closely related to *H. intercedens* and *H. euclaston* (Bothmer et al., 1995), and both Pleines and Blattner (2008) and Jakob et al. (2010) did recover *H. pusillum* as the sister to the two other species. However, our combined analysis suggests a sister-group relationship between *H. pusillum* and *H. erectifolium/H. stenostachys*, although without any jackknife support and not at costs 11, where *H. pusillum* is sister to *H. intercedens/H. euclaston*. Analyses of the nuclear data alone place *H. pusillum* unresolved with respect to the two species pairs *H. erectifolium/H. stenostachys* and *H. intercedens/H. euclaston*.

Sequence diversity of the plastid *trn*L-F region has been studied in a very large number of specimens of the diploid *Hordeum* species from South America (Jakob and Blattner, 2006; Jakob et al., 2009). A remarkable intraspecific diversity was found and some *trn*L-F haplotypes were shared between specimens of different species. In particular, *H. patagonicum*, *H. pubiflorum*, and *H. comosom* share haplotypes, but also *H. flexuosum* and *H. chilense* had at least one haplotype shared with one or more other species. Whether this high level of intraspecific diversity, so far described only from the short *trn*L-F sequences, can be extended to the entire plastid genome remains unknown. Thus, additional plastid sequence data for multiple specimens of the South American *Hordeum* species are highly desirable. The available data may yet again suggest either incomplete lineage sorting or extensive hybridization among these species. Additional nuclear sequences for more specimens would be needed, too, as they would show whether the intraspecific plastid diversity is a characteristic that can be extended to the nuclear genome as well, and if so whether shared nuclear haplotypes occur. If so, the present phylogenetic hypothesis may be biased by the selection of a single specimen to represent each taxon.

Transposable elements

As indicated above, *Stowaway* MITEs found in three of the nuclear genes were removed prior to the analyses. The elements in XYL and BLZ1 have been described previously (Petersen and Seberg, 2009), but the elements found in VRN3 were new. The characterization of these elements as *Stowaway* elements is based on sequence similarity to other elements combined with diagnostic sequence characteristics; for example, all *Stowaway* elements share a particular sequence motif located as terminal inverted repeats and have the ability to fold into a hairpin-like structure (Bureau and Wessler, 1994; Petersen and Seberg, 2009). *Stowaway* elements are supposed to have insertion site preference, and identical but non-homologous elements may occur at exactly the same position in different taxa following two or more independent insertion events (Mason-Gamer, 2007; Petersen and Seberg, 2009). We have postulated this to be the case for the elements in the BLZ1 gene, where *H. vulgare*, *H. bulbosum*, and *H. murinum* share one element type (BLZ1-A), and all other *Hordeum* species share another, non-homologous element type (BLZ1-B) (Petersen and Seberg, 2009). As our primary homology statements of the *Stowaway* elements rest on sequence similarity and characteristics we acknowledge that we could be mistaken, and thus conducted another set of phylogenetic analyses including the elements. The aims of these analyses are to investigate possible changes to the phylogenetic hypotheses by inclusion of the element sequences, and to investigate, by inspection of the implied alignments under the optimal cost set, whether the putative *Stowaway* element integrity is maintained, i.e. whether the elements are aligned as expected based on their structure, or if POY analyses violate the assumption of primary homology.

Optimal congruence for the combined analysis of all data (+ MITEs) is achieved at the costs set 4(1)2. The increased optimal gap opening cost compared with the analysis without MITEs [optimal at 3(1)2] is probably an effect of the increased number on major length differences among the sequences. The optimal tree (Fig. 4) differs from the tree excluding MITEs (Fig. 1) both in level of resolution and in branching order. The assumed four independent insertion events of *Stowaway* elements (one for each of XYL and VRN3 and two for BLZ1) are, however, equally likely on both trees where all element types are confined to individual monophyletic groups (Figs 1 and 4). The exact reasons for the change in branching order cannot be determined, but it is possible that the single, significantly shorter (by approximately 40 bp) VRN3 *Stowaway* sequence in *H. brachyantherum* compared with the elements in most other species tends to "attract it" to the species completely lacking the element. This type of "long-gap" attraction is caused by shared lack of longer strings of bases, when the individual positions in the gaps are treated as characters (Petersen et al., 2004). There are no major length differences between the *Stowaway* elements of *H. brevisubulatum* and *H. marinum*, and thus the change in branching order merely illustrates the instability of this part of the tree.

Considering the incongruence between the nuclear and plastid data (see below) it may be relevant to

explore the effect of MITEs on the alignments and trees derived from analysis of nuclear data. For such analysis optimality is achieved at the same costs set as without MITEs and with all data (+ MITEs): 5(1)2. The alignments of the relevant intron regions are almost completely identical to the alignments discussed above and the minor differences are irrelevant to the element insertion history. The nuclear gene tree (Fig. 4) is congruent with the combined (+ MITEs) tree, and deviates only by its resolution.

The implied alignment of the BLZ1 sequences [from the combined analysis under the optimal costs set 4(1)2] is potentially complicated by our previous assumption of the presence of two non-homologous, but partly similar, elements (BLZ1-A, BLZ1-B; Fig. 4) in two non-overlapping groups of taxa, but POY does produce an implied alignment where the two assumed non-homologous elements are not aligned, thus confirming non-homology (alignment not shown).

However, when it comes to element integrity, POY is not always able to produce alignments consistent with our primary homology assumptions. For BLZ1, sequence fragments of *Psathyrostachys* (lacking *Stowaway* elements) are aligned with element sequence, and for *Hordeum* species only possessing the shorter BLZ1-B element, sequence assumed to flank the element is aligned with element sequence from the longer of the two elements (BLZ1-A). The implied alignment of XYL sequences also show several fragments of *Psathyrostachys* (lacking *Stowaway* elements) sequence aligned with *Hordeum* element sequence (alignment not shown). However, the VRN3 sequences are aligned in agreement with our expectations except for a single nucleotide (alignment not shown).

Thus, the implied alignments of the nuclear intron regions, assumed to include *Stowaway* MITEs, show that optimization alignment as implemented in POY is not always capable of producing alignments consistent with our assumptions of element integrity. One obvious reason may be that these assumptions are invalid, but if they are accepted as valid, which is supported by other lines of evidence (e.g. Bureau and Wessler, 1994), a correct alignment should obviously not align element sequence with sequence from species that show no sign of having an element. Accepting that some of the implied alignments produced by POY are wrong justifies exclusion of *Stowaway* sequences from the principal analyses.

Perspectives

Based on data from 16 sequence regions from all diploid taxa of *Hordeum*, major areas of the phylogeny are becoming increasingly well corroborated. However, even this massive amount of data still leaves areas of the phylogeny ambiguously resolved. Inclusion of multiple accessions per taxon and polyploid taxa may help to answer some questions, but the fundamental problem, incongruent hypotheses of speciation invoked by plastid and nuclear data (genome tree incongruence), persists. As discussed above there are well-supported cases of incongruence, e.g. the position of *H. murinum* and one subspecies of *H. marinum*, where the phylogeny seems positively misguided by plastid data. To what extent this can be used to support other topological conflicts as being caused by genome tree conflict remains unclear. However, if this is the case, is a phylogenetic tree based entirely on nuclear data then a better hypothesis about species relationships than a tree based on all data—nuclear and plastid? In the present study, the topology of the nuclear tree and the combined tree are largely but not entirely similar, but this may merely be a consequence of a much larger amount and hence dominance of nuclear data. This discussion seems reminiscent of the fear that morphological data should be swamped by molecular data (e.g. Hillis, 1987). Adding complete plastid genome sequence data to the combined analyses may change the phylogeny drastically—at least for a while until nuclear genome sequencing becomes commonplace. To us, there is no simple answer to the question. In terms of our understanding of *Hordeum* phylogeny, we can do little else than intensively explore the behaviour of molecular data and, whenever possible, aided by other types of data, attempt to interpret incongruencies biologically. However, within the realm of coalescence theory, adherents of model-based phylogenetic methods might approach these problems in a radically different manner (e.g. Kubatko, 2009; Liu et al., 2009).

Several studies of other plant groups have demonstrated incongruence between nuclear and plastid data partitions (e.g. Doyle et al., 2003; Albach and Chase, 2004; Guo et al., 2004; Smissen et al., 2004; Weeks and Simpson, 2004; Fehrer et al., 2007; Kim and Donoghue, 2008), but none to our knowledge as firmly—in terms of the number of sequenced regions—as the present study. It is widely accepted that plastid gene trees may be phylogenetically misleading when evolutionary processes, such as hybridization or incomplete lineage sorting, have occurred during speciation (Wendel and Doyle, 1998; Wolfe and Randle, 2004). Obviously, nuclear gene trees may be misleading, too—not least ITS trees as used in the majority of the cited studies—but we expect that trees derived from combined analysis of many nuclear loci are more likely to infer the correct organismal phylogeny. Biological reasons for incongruence have for obvious reasons usually been discussed in relation to species-level phylogenies and their broader, potential consequences largely neglected in higher level phylogenetics. However, speciation processes shape not

only species phylogenies, but through time they will leave trace even in higher level phylogenies.

It might be tempting to ignore the problem, for example because of our use of direct optimization, but the congruence problem is not restricted to this methodology, as it has been found using other methods (Petersen and Seberg, 2003). The problem might even be worse if the observations of Jakob and Blattner (2006) are of more general occurrence. It is also possible to explain the problems as caused by our choice of organism, but such argument needs evidence that what we have found in *Hordeum* is indeed exceptional.

Thus, if our observations can be generalized, the wider implications for the standard use of plastid data may be alarming. Massive amounts of plastid data, as provided by recent whole plastid genome sequencing, may result in increasingly well-supported organelle trees (Cai et al., 2006; Jansen et al., 2007): trees that may or may not reflect organismal phylogeny. It might be worthwhile starting to spend more time exploring potential conflicts and signal in data from different genomic partitions instead of seeking increased support for existing hypotheses by addition of still more plastid data.

Acknowledgements

We are most grateful to Frank Blattner and Katsuyuki Kakeda for helping us sort out the unfortunate problems relating to previously mixed up DNA samples and sequences. We thank Charlotte Hansen for skillfully conducting the lab work. The work was funded by the Danish National Sciences Research Council grant no. 272-06-0436.

References

Aagesen, L., Petersen, G., Seberg, O., 2005. Sequence length variation, indel costs, and congruence in sensitivity analysis. Cladistics 21, 15–30.

Albach, D.C., Chase, M.W., 2004. Incongruence in Veroniceae (Plantaginaceae): evidence from two plastid and a nuclear ribosomal DNA region. Mol. Phylogenet. Evol. 32, 183–197.

Althoff, D.M., Gitzendanner, M.A., Segraves, K.A., 2007. The utility of amplified fragment length polymorphisms in phylogenetics: a comparison of homology within and between genomes. Syst. Biol. 56, 477–484.

Baden, C., Bothmer, R.von., 1994. A taxonomic revision of *Hordeum* sect. *Critesion*. Nord. J. Bot. 14, 117–136.

Baum, B.R., Johnson, D.A., Bailey, L.G., 2001. Defining orthologous groups among multicopy genes prior to inferring phylogeny, with special emphasis on the Triticeae (Poaceae). Hereditas 135, 123–138.

Baum, B.R., Edwards, T., Johnson, D.A., 2010. Codependence of repetitive sequence classes in genomes: phylogenetic analysis of 5S rDNA families in *Hordeum* (Triticeae: Poaceae). Genome 53, 180–202.

Blattner, F.R., 2004. Phylogenetic analysis of *Hordeum* (Poaceae) as inferred by nuclear rDNA ITS sequences. Mol. Phylogenet. Evol. 33, 289–299.

Blattner, F.R., 2006. Multiple intercontinental dispersals shaped the distribution area of *Hordeum* (Poaceae). New Phytol. 169, 603–614.

Blattner, F.R., 2009. Progress in phylogenetic analysis and a new infrageneric classification of the barley genus *Hordeum* (Poaceae: Triticeae). Breed. Sci. 59, 471–480.

Bothmer, R.von., Jacobsen, N., 1991. Interspecific hybrids within the genus *Hordeum*. In: Gupta, P.K., Tsuchiya, T. (Eds.), Chromosome Engineering in Plants: Genetics, Breeding, Evolution, Part A. Elsevier, Amsterdam, pp. 411–431.

Bothmer, R.von., Jacobsen, N., Baden, C., Jørgensen, R.B., Linde-Laursen, I., 1995. An Ecogeographical Study of *Hordeum*, 2nd edn. IPGRI, Rome.

Brower, A.V.Z., 2000. Evolution is not a necessary assumption of cladistics. Cladistics 16, 143–154.

Bureau, T.E., Wessler, S.R., 1994. *Stowaway*: a new family of inverted repeat elements associated with genes of both monocotyledonous and dicotyledonous plants. Plant Cell 6, 907–916.

Cai, Z., Penaflor, C., Kuehl, J.V., Leebens-Mack, J., Carlson, J.E., dePamphilis, C.W., Boore, J.L., Jansen, R.K., 2006. Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. BMC Evol. Biol. 6, 77.

Cronn, R., Cedroni, M., Haselkorn, T., Grover, C., Wendel, J.F., 2002. PCR-mediated recombination in amplification products derived from polyploidy cotton. Theor. Appl. Genet. 104, 482–489.

Doebley, J.R., Bothmer, R.von., Larson, S., 1992. Choloroplast DNA variation and the phylogeny of *Hordeum* (Poaceae). Am. J. Bot. 79, 576–584.

Doyle, J.J., Doyle, J.L., Rauscher, J.T., Brown, A.H.D., 2003. Diploid and polyploid reticulate evolution throughout the history of the perennial soybeans (*Glycine* subgenus *Glycine*). New Phytol. 161, 121–132.

Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1995. Testing significance of incongruence. Cladistics 10, 315–319.

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

Graves, J., 2009. Deeper AFLPs. Heredity 103, 99.

Guo, Y.-P., Ehrendorfer, F., Samuel, R., 2004. Phylogeny and systematic of *Achillea* (Asteraceae-Anthemideae) inferred from nrITS and plastid *trnL-F* DNA sequences. Taxon 53, 657–672.

Hansen, L., von Wettstein-Knowles, P., 1991. The barley genes *Acl1* and *Acl3* encoding acyl carrier proteins I and III are located on different chromosomes. Mol. Gen. Genet. 229, 467–478.

Hillis, D.M., 1987. Molecular versus morphological approaches to systematics. Annu. Rev. Ecol. Syst. 18, 23–42.

Jakob, S.S., Blattner, F.R., 2006. A chloroplast genealogy of *Hordeum* (Poaceae): long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. Mol. Biol. Evol. 23, 1602–1612.

Jakob, S.S., Martinez-Meyer, E., Blattner, F.R., 2009. Phylogeographic analyses and paleodistribution modeling indicate Pleistocene in situ survival of *Hordeum* species (Poaceae) in southern Patagonia without genetic or spatial restriction. Mol. Biol. Evol. 26, 907–923.

Jakob, S.S., Heibl, C., Rödder, D., Blattner, F.R., 2010. Population demography influences climatic niche evolution: evidence from diploid American *Hordeum* species (Poaceae). Mol. Ecol. 19, 1423–1438.

Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., dePamphilis, C.W., Leebens-Mack, J., Müller, K.F., Guisinger-Bellian, M., Haberle, R.C., Hansen, A.K., Chumley, T.W., Lee, S.-B., Peery,

R., McNeal, J.R., Kuehl, J.V., Boore, J.L., 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc. Natl Acad. Sci. USA 104, 19369–19374.

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

Kellogg, E.A., Appels, R., 1995. Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. Genetics 140, 325–343.

Kellogg, E.A., Appels, R., Mason-Gamer, R.J., 1996. When genes tell different stories: the diploid genera of the Tritceae (Gramineae). Syst. Bot. 21, 321–347.

Kim, S.-T., Donoghue, M.J., 2008. Incongruence between cpDNA and nrITS trees indicates extensive hybridization within *Eupersicaria* (Polygonaceae). Am. J. Bot. 95, 1122–1135.

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

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

Kubatko, L.S., 2009. Identifying hybridization events in the presence of coalescence via model selection. Syst. Biol. 58, 478–488.

Linder, C.R., Rieseberg, L.H., 2004. Reconstructing patterns of reticulate evolution in plants. Am. J. Bot. 91, 1700–1708.

Liu, L., Yu, L., Pearl, D.K., Edwards, S.V., 2009. Estimating species phylogenies using coalescence times among sequences. Syst. Biol. 58, 468–477.

Löytynoja, A., Goldman, N., 2005. An algorithm for progressive multiple alignment of sequences with insertions. Proc. Natl Acad. Sci. USA 102, 10557–10562.

Mason-Gamer, R.J., 2007. Multiple homoplasious insertions and deletions of a Triticeae (Poaceae) DNA transposon: a phylogenetic perspective. BMC Evol. Biol. 7, 92.

Mason-Gamer, R.J., Kellogg, E.A., 1996. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Poaceae). Syst. Biol. 45, 524–545.

Mason-Gamer, R.J., Weil, C.F., Kellogg, E.A., 1998. Granule-bound starch synthase: structure, function, and phylogenetic utility. Mol. Biol. Evol. 15, 1658–1673.

Mickevich, M.L., Farris, J.S., 1981. The implications of congruence in *Menidia*. Syst. Zool. 30, 351–370.

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

Olmstead, R.G., Sweere, J.A., 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. Syst. Biol. 43, 467–481.

Petersen, G., Seberg, O., 1998. Phylogeny and genomes in the genus *Hordeum*. In: Jaradat, A.A. (Ed.), Triticeae III. Science Publishers, Enfield, pp. 109–118.

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

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

Petersen, G., Seberg, O., 2005. Phylogenetic analysis of allopolyploid species. Czech. J. Genet. Plant Breed. 41 (special issue), 28–37.

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

Petersen, G., Seberg, O., Aagesen, L., Frederiksen, S., 2004. An empirical test of the treatment of indels during optimization alignment based on the phylogeny of the genus *Secale* (Poaceae). Mol. Phylogenet. Evol. 30, 733–742.

Phillips, A., Janies, D., Wheeler, W.C., 2000. Multiple sequence alignment in phylogenetic analysis. Mol. Phylogenet. Evol. 16, 317–330.

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. Theor. Appl. Genet. 101, 652–660.

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

Sang, T., Crawford, D.J., Stuessy, T.F., 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). Am. J. Bot. 84, 1120–1136.

Sastri, D.C., Hilu, K., Appels, R., Lagudah, E.S., Playford, J., Baum, B.R., 1992. An overview of evolution in plant 5S DNA. Plant Syst. Evol. 183, 169–181.

Seberg, O., Petersen, G., 2007. Phylogeny of Triticeae (Poaceae) based on three organelle genes, two single-copy nuclear genes, and morphology. Aliso 23, 361–371.

Smissen, R.D., Breitwieser, I., Ward, J.M., 2004. Phylogenetic implications of trans-specific chloroplast DNA sequence polymorphism in New Zealand Gnaphalieae (Asteraceae). Plant Syst. Evol. 249, 37–53.

Sun, G., Pourkheirandish, M., Komatsuda, T., 2009. Molecular evolution and phylogeny of the *RPB2* gene in the genus *Hordeum*. Ann. Bot. 103, 975–983.

Taberlet, P., Gielly, L., Pautou, G., Bouvet, J., 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol. 17, 1105–1109.

Tate, J.A., Simpson, B.B., 2003. Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. Syst. Bot. 28, 723–737.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Varón, A., Vinh, L.S., Bomash, I., Wheeler, W.C., 2008. POY 4.0.2870. American Museum of Natural History, New York.

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. Plant J. 13, 629–640.

Wang, R.R.-C., von Bothmer, R., Dvorak, J., Fedak, G., Linde-Laursen, I., Muramatsu, M., 1996. Genome symbols in the Triticeae (Poaceae). In: Wang, R.R.-C., Jensen, K.B., Jaussi, C. (Eds.), Proceedings of the 2nd International Triticeae Symposium. Utah State University Publication, Design and Production, Utah, pp. 29–34.

Weeks, A., Simpson, B.B., 2004. Molecular genetic evidence for interspecific hybridization among endemic hispaniolan *Bursera* (Burseraceae). Am. J. Bot. 91, 976–984.

Wendel, J.F., Doyle, J.J., 1998. Phylogenetic incongruence: window into genome history and molecular evolution. In: Soltis, D.E., Soltis, P.M., Doyle, J.J. (Eds.), Molecular Systematics of Plants II. Kluwer Academic Publishers, Boston, pp. 265–296.

Wendel, J.F., Schnabel, A., Seelanan, T., 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc. Natl Acad. Sci. USA 92, 280–284.

Wheeler, W.C., 1995. Sequence alignment, parameter sensitivity and the phylogenetic analysis of molecular data. Syst. Biol. 44, 321–331.

Wheeler, W.C., 1996. Optimization alignment: the end of multiple sequence alignment in phylogenetics. Cladistics 12, 1–9.

Wolfe, A.D., Randle, C.P., 2004. Recombination, heteroplasmy, haplotype polymorphism, and paralogy in plastid genes: implications for plant molecular systematic. Syst. Bot. 29, 1011–1020.

Yan, L., Fu, D., Li, C., Blechl, A., Tranquilli, G., Bonafede, M., Sanchez, A., Valarik, M., Yasuda, S., Dubcovsky, J., 2006. The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. Proc. Natl Acad. Sci. USA 103, 19581–19586.

Appendix

Table A1 GenBank accession numbers of *Hordeum* and *Psathyrostachys* sequences

Taxon	ACL3	BLZ1	DMC1(1)/ DMC1(2)	EFG	NAR7	NUC	VRN3	Waxy	XYL	atpB/rbcL	atpF-H	matK	ndhF	psbIK	trnH-psbA	trnLF
H. bogdanii Wilensky	HQ619393	EU862589	EU862674/ AY137412	AF135001	HQ619365	EU862616	HQ619440	EU282317	EU862645	ABO78358/ AY137455	HQ619269	AB078089	AF056181	HQ619303	EU118530	AB078438
H. brachyantherum Nevski ssp. *californicum* (Covas & Stebbins) Bothmer, N. Jacobsen & Seberg	HQ619389	EU862585	EU862670/ AF277260	AF135004	HQ619361	EU862612	HQ619436	HQ619464	EU862641	ABO78362/ AY137451	HQ619265	AB078093	DQ247887	HQ619299	HQ619325	AB078442
H. brevisubulatum (Trin.) Link ssp. *violaceum* Boiss. & Hohen.	HQ619373	EU862569	EU862654/ AY137396	AY137392	HQ619345	EU862597	HQ619420	HQ619450	EU862625	HQ619243/ AY137435	HQ619249	HQ619277	AF267691	HQ619283	HQ619312	HQ619334
H. bulbosum L.	HQ619392	EU862588	EU862673/ AY137411	AF134989	HQ619364	EU862615	HQ619439	HQ619466	EU862644	ABO78364/ AY137454	HQ619268	AB078095	HQ619415	HQ619302	EU118529	AB078444
H. chilense Roem. & Schult.	HQ619387	EU862583	EU862668/ AY137408	AF134996	HQ619359	EU862610	HQ619434	EU282318	EU862639	HQ619244/ AY137449	HQ619263	HQ619278	HQ619412	HQ619297	HQ619323	HQ610338
H. comosum C. Presl	HQ619379	EU862575	EU862660/ AY137400	AF135000	HQ619351	EU862603	HQ619426	HQ619456	EU862631	ABO78366/ AY137441	HQ619255	AB078097	HQ619404	HQ619289	HQ619317	AB078446
H. cordobense Bothmer, N. Jacobsen & Nicora	HQ619396	EU862591	EU862677/ AY137415	AF134997	HQ619367	EU862619	HQ619443	HQ619469	EU862647	ABO78367/ AY137458	HQ619272	AB078098	HQ619417	HQ619306	HQ619329	HQ619340
H. erectifolium Bothmer, N. Jacobsen & R.B. Jørg.	HQ619378	EU862574	EU862659/ AF277259	AF135005	HQ619350	EU862602	HQ619425	HQ619455	EU862630	ABO78369/ AY137440	HQ619254	AB078100	DQ247888	HQ619288	HQ619316	AB078448
H. euclaston Steud.	HQ619380	EU862576	EU862661/ AY137401	AF134993	HQ619352	EU862604	HQ619427	HQ619457	EU862632	ABO78370/ AY137442	HQ619256	AB078101	HQ619405	HQ619290	EU118531	AB078449
H. flexuosum Steud.	HQ619377	EU862573	EU862658/ AY137399	AF134994	HQ619349	EU862601	HQ619424	HQ619454	EU862629	ABO78371/ AY137439	HQ619253	AB078102	HQ619403	HQ619287	HQ619315	AB078450
H. intercedens Nevski	HQ619388	EU862584	EU862669/ AY137409	AF134992	HQ619360	EU862611	HQ619435	HQ619463	EU862640	ABO78374/ AY137450	HQ619264	AB078105	HQ619413	HQ619298	HQ619324	AB078453
H. marinum Huds. ssp. *marinum*	HQ619374	EU862570	EU862655/ AY137397	AY028799	HQ619346	EU862698	HQ619421	HQ619451	EU862626	ABO78388/ AY137436	HQ619250	AB078119	HQ619401	HQ619284	EU118533	AB078466
ssp. *gussoneanum* (Parl.) Thell.	HQ619372	EU862568	EU862653/ AF277257	AF135012	HQ619344	EU862696	HQ619419	HQ619449	EU862624	ABO78382/ AY137434	HQ619248	AB078113	DQ247889	HQ619282	HQ619311	AB078460
H. murinum L. ssp. *glaucum* (Steud.) Tzvelev	HQ619375	EU862571	EU862656/ AF277258	AF134990	HQ619347	EU862599	HQ619422	HQ619452	EU862627	ABO78389/ AY137437	HQ619251	AB078120	DQ247890	HQ619285	HQ619313	AB078467
H. muticum C. Presl	HQ619376	EU862572	EU862657/ AY137398	AF134995	HQ619348	EU862600	HQ619423	HQ619453	EU862628	ABO78393/ AY137438	HQ619252	AB078124	HQ619402	HQ619286	HQ619314	AB078471
H. patagonicum (Hauman) Covas ssp. *patagonicum*	HQ619382	EU862578	EU862663/ AY137403	AF135006	HQ619354	EU862606	HQ619429	HQ619459	EU862634	ABO78397/ AY137444	HQ619258	AB078128	HQ619407	HQ619292	EU118535	HQ619335

Appendix
(*Continued*)

Taxon	ACL3	BLZ1	DMC1(1)/ DMC1(2)	EFG	NAR7	NUC	VRN3	Waxy	XYL	atpB/rbcL	atpF-H	matK	ndhF	psbIK	trnH-psbA	trnLF
ssp. *setifolium* (Parodi & Nicora) Bothmer, Giles & N. Jacobsen	HQ619383	EU862579	EU862664/ AY137404	AF135007	HQ619355	EU862607	HQ619430	HQ619460	EU862635	ABO78399/ AY137445	HQ619259	AB078130	HQ619408	HQ619293	HQ619319	AB078474
ssp. *santacrucense* (Parodi & Nicora) Bothmer, Giles & N. Jacobsen	HQ619385	EU862581	EU862666/ AY137406	AF135008	HQ619357	EU862609	HQ619432	HQ619462	EU862637	ABO78398/ AY137447	HQ619261	AB078129	HQ619410	HQ619295	HQ619321	HQ619336
ssp. *mustersii* (Nicora) Bothmer, Giles & N. Jacobsen	HQ619384	EU862580	EU862665/ AY137405	AF135010	HQ619356	EU862608	HQ619431	HQ619461	EU862636	ABO78396/ AY137446	HQ619260	AB078127	HQ619409	HQ619294	HQ619320	AB078473
ssp. *magellanicum* (Parodi & Nicora) Bothmer, Giles & N. Jacobsen	HQ619395	EU862590	EU862676/ AY137414	AF135009	HQ619366	EU862618	HQ619442	HQ619468	EU862646	ABO78395/ AY137457	HQ619271	AB078126	HQ619416	HQ619305	HQ619328	HQ619337
H. pubiflorum Hook. f.	HQ619381	HQ619238	HQ619242	AF134999	HQ619353	HQ619239	HQ619428	HQ619458	HQ619240	ABO78401/ HQ619241	HQ619257	EU118389	HQ619406	HQ619291	HQ619318	AB078476
H. pusillum Nutt.	HQ619390	EU862586	EU862671/ AY137410	AF134991	HQ619362	EU862613	HQ619437	EU282231	EU862642	ABO78402/ AY137452	HQ619266	AB078133	HQ619414	HQ619300	HQ619326	AB078477
H. roshevitzii Bowden	HQ619397	EU862592	EU862678/ AY137416	AF135002	HQ619368	EU862620	HQ619444	HQ619470	EU862648	ABO78403/ AY137459	HQ619273	AB078134	HQ619418	HQ619307	HQ619330	AB078478
H. stenostachys Godr.	HQ619386	EU862582	EU862667/ AY137407	AF134998	HQ619358	–	HQ619433	EU282322	EU862638	ABO78405/ AY137448	HQ619262	AB078136	HQ619411	HQ619296	HQ619322	AB078480
H. vulgare L. ssp. *vulgare*	HQ619394	X80068	EU862675/ AY137413	AF134987	X60173	EU862617	HQ619441	HQ619468	X95256	ABO78407/ AY137456	HQ619270	AB078138	AF267692	HQ619304	HQ619327	AB078481
ssp. *spontaneum* (C. Koch) Thell.	HQ619391	EU862587	EU862672/ AF277262	AF134988	HQ619363	EU862614	HQ619438	HQ619465	EU862643	ABO78408/ AY137453	HQ619267	AB078139	DQ247891	HQ619301	EU118528	HQ619339
P. fragilis (Boiss.) Nevski ssp. *fragilis*	HQ619398	EU862593	EU862679/ AF277261	AY137393	HQ619369	EU862621	HQ619445	HQ619471	EU862649	HQ619245/ AY137460	HQ619274	HQ619279	DQ247896	HQ619308	HQ619331	HQ619341
ssp. *villosus* Baden	HQ619399	EU862594	EU862680/ AF277263	AY137394	HQ619370	EU862622	HQ619446	HQ619472	EU862650	HQ619246/ AY137461	HQ619275	HQ610280	DQ247897	HQ619309	HQ619332	HQ619342
P. stoloniformis Baden	HQ619400	EU862595	EU862652/ AF277264	AY137395	HQ619371	EU862623	HQ619447-8	HQ619473	EU862651	HQ619247/ AY137462	HQ619276	HQ610281	DQ247898	HQ619310	HQ619333	HQ619343