BioOne COMPLETE

Sequencing and De Novo Transcriptome Assembly of Brachypodium sylvaticum (Poaceae)

Authors: Fox, Samuel E., Preece, Justin, Kimbrel, Jeffrey A., Marchini, Gina L., Sage, Abigail, et al.

Source: Applications in Plant Sciences, 1(3)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1200011

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Downloaded From: https://bioone.org/journals/Applications-in-Plant-Sciences on 16 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

Applications in Plant Sciences 1(3): 1200011

Genomic Resources Note

Sequencing and de novo transcriptome assembly of *Brachypodium sylvaticum* (Poaceae)[1]

Samuel E. Fox[2], Justin Preece[2], Jeffrey A. Kimbrel[3], Gina L. Marchini[4], Abigail Sage[2], Ken Youens-Clark[5], Mitchell B. Cruzan[4], and Pankaj Jaiswal[2,6]

[2]Department of Botany and Plant Pathology and Center for Genome Research and Biocomputing, 2082 Cordley Hall, Oregon State University, Corvallis, Oregon 97331 USA; [3]Joint BioEnergy Institute, Lawrence Berkeley National Laboratory, 5885 Hollis Street, Fourth Floor, Emeryville, California 94608 USA; [4]Department of Biology, Portland State University, Portland, Oregon 97201 USA; and [5]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724 USA

- *Premise of the study:* We report the de novo assembly and characterization of the transcriptomes of *Brachypodium sylvaticum* (slender false-brome) accessions from native populations of Spain and Greece, and an invasive population west of Corvallis, Oregon, USA.
- *Methods and Results:* More than 350 million sequence reads from the mRNA libraries prepared from three *B. sylvaticum* genotypes were assembled into 120,091 (Corvallis), 104,950 (Spain), and 177,682 (Greece) transcript contigs. In comparison with the *B. distachyon* Bd21 reference genome and GenBank protein sequences, we estimate >90% exome coverage for *B. sylvaticum*. The transcripts were assigned Gene Ontology and InterPro annotations. *Brachypodium sylvaticum* sequence reads aligned against the Bd21 genome revealed 394,654 single-nucleotide polymorphisms (SNPs) and >20,000 simple sequence repeat (SSR) DNA sites.
- *Conclusions:* To our knowledge, this is the first report of transcriptome sequencing of invasive plant species with a closely related sequenced reference genome. The sequences and identified SNP variant and SSR sites will provide tools for developing novel genetic markers for use in genotyping and characterization of invasive behavior of *B. sylvaticum*.

Key words: *Brachypodium sylvaticum*; comparative genomics; de novo transcriptome; invasive species; simple sequence repeat (SSR); single-nucleotide polymorphism (SNP).

Brachypodium sylvaticum (Huds.) P. Beauv. (slender false-brome; Poaceae), with an estimated genome size of 470 Mb and 17 chromosomes (Foote et al., 2004), is a perennial bunchgrass native to Europe, Asia, and North Africa and is closely related to the bioenergy feedstock model grass *B. distachyon* (L.) P. Beauv. (Wolny et al., 2011), which has a sequenced genome of 272 Mb and five chromosomes. In its native range, *B. sylvaticum* occurs in habitats ranging from forest understory to open meadows and tolerates conditions from full shade to full sun (Holten, 1980; Long, 1989; Aarrestad, 2000; Kirby and Thomas, 2000). In the United States, *B. sylvaticum* is invasive and listed as a noxious weed covering the west coast of California, Oregon, and Washington (Oregon Department of Agriculture, 2009; Washington State Department of Agriculture, 2009; Lionakis Meyer and Effenberger, 2010). It is also expanding into the eastern United States where it has been reported in Missouri and Virginia (Roy, 2010). In Oregon, *B. sylvaticum* forms thick monocultures in open forests at elevations from nearly sea level to approximately 1200 m. It threatens the endangered Oregon Willamette Valley oak savanna ecosystem by replacing the native flora and reduces habitat for rare butterflies (Kaye and Blakeley-Smith, 2006; Severns and Warren, 2008). False brome is shade tolerant (Murchie and Horton, 1997; Holmes et al., 2010), which makes it a particularly dangerous invasive threat to undisturbed habitats (Martin et al., 2009).

False brome was first introduced into Oregon via plant introduction studies in the early part of the twentieth century, and later identified and collected from the wild in 1939 (Chambers, 1966; Kaye and Blakeley-Smith, 2006). Field trials for exotic grasses were performed by the United States Department of Agriculture (USDA) in Corvallis at the Oregon State University facilities, and *B. sylvaticum* was widely planted to "improve range" throughout the western United States (Hull, 1974). Oregon State University herbarium records indicate that two separate experimental gardens were established in Eugene and Corvallis, Oregon. Genetic profiling with microsatellite markers confirm these introductions were independent and that they probably consisted of the same set of multiple accessions from the native range in Europe that had been collected by the USDA Division of Plant Introduction (Rosenthal et al., 2008). Accessions in each of these two plantings have crossed, and the invasive plants that are now spreading across Oregon forests are recombinant products of hybridization. *Brachypodium sylvaticum*

[1]Manuscript received 21 December 2012; revision accepted 7 February 2013.

The authors thank members of the Center for Genome Research and Biocomputing at Oregon State University for sequencing and computational support, Tanya Cheeke for suggestions on the manuscript, and Cathy Gresham and Fiona McCarthy for InterProScan analysis. Thanks to Sharon Wei and Doreen Ware from Gramene database, and Henry Priest and Todd Mockler from BrachyBase for help with the integration of data sets in the respective databases. Funded by laboratory startup funds provided to P.J.

[6]Author for correspondence: jaiswalp@science.oregonstate.edu

doi:10.3732/apps.1200011

 © 2013 Fox et al. Published by the Botanical Society of America. This work is licensed under a Creative Commons Attribution License (CC-BY-NC-SA).

Downloaded From: https://bioone.org/journals/Applications-in-Plant-Sciences on 16 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

has thus become increasingly common in the past 15 yr (Chambers, 1966; Kaye and Blakeley-Smith, 2006; Rosenthal et al., 2008). The introductions in Corvallis and Eugene retain unique marker signatures, but Bayesian cluster analyses indicate that similar sets of native accessions from Western Europe contributed to the hybrid genotypes that are spreading from each introduction location (Rosenthal et al., 2008).

Despite the economic and environmental impact of *B. sylvaticum*, there remains a lack of adequate genomic resources for the study of the invasive populations. To thoroughly investigate genetic differences between invasive and native populations, a reference transcriptome specific for *B. sylvaticum* is needed to precisely align, map, and interrogate gene sequences. The major aim of this study was to assemble, annotate, and characterize a high-quality reference transcriptome that will enable researchers to assess gene expression levels, conduct comparative analyses, and identify putative single-nucleotide polymorphism (SNP) and simple sequence repeat (SSR) sequence sites in the genomes of *B. sylvaticum* populations for developing genetic markers to be used in future genotyping, identification, and genetic tracking studies.

Over the past several years, next-generation sequencing (NGS) has emerged as a low-cost, large-scale, fast, and accurate approach for de novo transcriptome sequencing (reviewed in Ward et al., 2012). Moreover, the tremendous depth of coverage generated by the Illumina sequencing platform in particular enables marker and gene discovery, comparative genomics, and gene expression analysis in nonmodel organisms (Wang et al., 2010; Huang et al., 2012; Nicolai et al., 2012; Varshney et al., 2012; Wang et al., 2012; Zhao et al., 2012). We conducted RNA-Seq transcriptome assemblies on *B. sylvaticum* plants originating west of Corvallis, Oregon (hereafter referred to as Corvallis, or Brasy-Cor), Spain (Brasy-Esp), and Greece (Brasy-Gre). We generated >36 Gb of *B. sylvaticum* transcriptome and assembled the sequences into 120,091, 104,095, and 177,095 transcript contigs with an average length of 1652, 1728, and 1566 bp for samples from Corvallis, Spain, and Greece, respectively. The cDNA sequence files are provided in fasta format for each population in Appendices S1 (Brasy-Cor), S2 (Brasy-Esp), and S3 (Brasy-Gre). We estimate that these transcriptomes represent >90% of the *B. sylvaticum* gene space. Furthermore, we identified SNP and SSR sequences that could be used to design genetic markers for use in future population studies. Along with the substantial genomic resources available in a close congener (*B. distachyon*; Mur et al., 2011) providing reference and comparative material, this transcriptome assembly will be useful on a broad scale as a greatly needed resource for ecologists and geneticists conducting research on native and invasive populations of *B. sylvaticum*, and on the role of climate change and adaptation toward successful invasiveness.

METHODS AND RESULTS

***Transcriptome sequencing and de novo assembly*—**Populations from Spain and Greece were selected to investigate correlations between the Oregon, USA samples and European progenitors used in the USDA field trials. Our samples were drawn from populations in Corvallis, Oregon (population OR-C1; Rosenthal et al., 2008; GPS coordinates: 44°39′35″N, 124°45′41″W), Avila, Spain (population SAV, USDA accession PI 318962; GPS coordinates: 40°39′27″N, 5°18′38″W), and Thessaloniki, Greece (population GRE, USDA accession PI 206546; GPS coordinates: 40°37′48″N, 22°57′36″E). The OR-C1 plants were selected from field-collected seed, and the GRE and SAV seed samples were collected and maintained by the USDA Plant Germplasm division in Pullman, Washington (USA). All plants were grown in a common greenhouse garden at the Portland State University campus in Portland, Oregon, under 12 hours light at 25°C and 12 hours dark at 15°C. At 60 wk, leaf tissue was collected from two individuals per population and ground in liquid nitrogen. Total cellular RNA was extracted using a modified protocol described elsewhere (Fox et al., 2009). In brief, RNA was extracted using RNA Plant Reagent (Life Technologies, Grand Island, New York, USA) and treated with RNase-free Turbo DNase (Life Technologies). Concentration, integrity, and extent of contamination by ribosomal RNA were assessed using an ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and Bioanalyzer 2100 (Agilent Technologies, Santa Clara, California, USA). Samples were prepared using the TruSeq RNA Sample Preparation Kit (v2) and sequenced on the Illumina HiSeq 2000 instrument (Illumina, San Diego, California, USA). We generated >358 million 101-bp paired-end reads with a fragment size of ~325 bp. This represents an overall total of 36.19 Gb of 101-bp paired-end *B. sylvaticum* transcriptome sequence. The reads from each sample were indexed for use in preliminary expression analyses and represent between 47 and 79 million reads per sample (Table 1). A total of 12.1 Gb paired-end sequences were used to assemble the Corvallis reference transcriptome, and 10.9 and 12.8 Gb were used for the Greece and Spain references, respectively.

The raw Illumina reads were processed for quality and parsed for index sequences and pairs using custom Perl scripts prior to assembly. The metrics used to assess transcriptome assembly quality included the overall number (coverage) of contigs, the average length of contigs, and the diversity of contigs (the estimated number of discrete loci assembled). We compared de novo assemblies using Velvet/Oases and Trinity algorithms, two published software programs built specifically to assemble de novo transcriptomes from short-read sequence data (Grabherr et al., 2011; Schulz et al., 2012). After evaluating the performance of the Velvet/Oases and Trinity algorithms, we conducted our final assembly using Velvet/Oases (Appendix S4). Our assembly generated 120,091, 104,095, and 177,682 contigs for the Corvallis, Spain, and Greece assemblies (Table 1; Appendices S1, S2, and S3). In all three *B. sylvaticum* assemblies, ~96% of assembled contigs were longer than 250 bp, with the longest contig for each of the three assemblies being greater than 15,000 bp. The average contig length was 1652, 1728, and 1566 bp for samples from Corvallis, Spain, and Greece, respectively, while the average length of the reference *B. distachyon* Bd21 gene models (v1.2) is 1086 bp. The median for each assembly, while lower, was near the average, indicating that the assemblies were not composed of an over-representative set of small contigs. Furthermore, the frequency distribution of contig scaffold sizes shows the majority of the lengths near the median and the overall frequency distribution similar to that observed in *B. distachyon* (Fig. 1A). However, when we compared the number of assembled contigs with the average length, we observed many more contigs in our de novo assemblies than the number of gene models in *B. distachyon*, and the average contig lengths in *B. sylvaticum* were greater than that of *B. distachyon* as well (Fig. 1B). The greater number of contigs in the *B. sylvaticum* assemblies could result from several factors, including gene splicing isoforms,

Table 1. Statistics of data sets from the sequencing and de novo transcriptome assemblies of the *Brachypodium sylvaticum* samples from Corvallis (Brasy-Cor), Spain (Brasy-Esp), and Greece (Brasy-Gre) and their comparison with the transcriptome of *B. distachyon* Bd21 sequenced genome.

	Raw sequences		Assembled contigs			
Transcriptomes	No. of reads	Gb	Total no.	Longest sequence (bp)	Average length (bp)	Median length (bp)
Brasy-Cor	120,443,086	12.16	120,091	16,713	1616	1380
Brasy-Esp	109,942,266	11.1	104,950	21,443	1696	1456
Brasy-Gre	127,927,820	12.92	177,682	15,289	1566	1317
B. distachyon (Bd21)	—	—	31,029	14,577	1280	1086

Downloaded From: https://bioone.org/journals/Applications-in-Plant-Sciences on 16 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

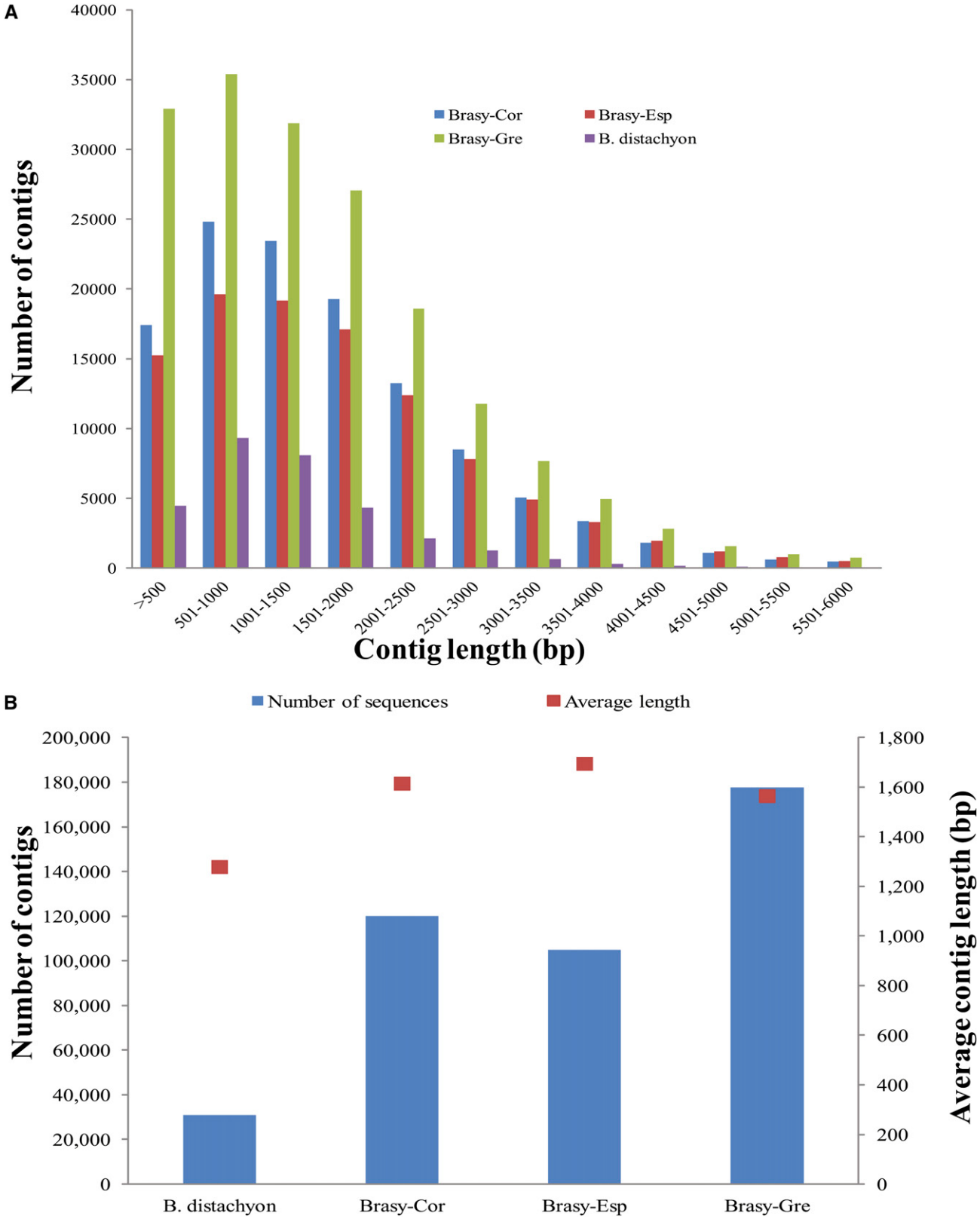

Fig. 1. Sequence comparisons. (A) Histogram of frequency distribution of contig lengths of *Brachypodium sylvaticum* samples from Corvallis (Brasy-Cor), Spain (Brasy-Esp), and Greece (Brasy-Gre) de novo transcriptome assemblies compared with *B. distachyon* cDNA lengths (X-axis has been truncated at 6 kb). The frequency distribution of the *B. sylvaticum* transcriptomes closely mirrors that of the *B. distachyon* transcriptome. (B) Comparisons of number of contigs and the average length of *B. sylvaticum* assembled contigs and *B. distachyon* transcripts. Although the overall number of contigs is far greater than the number of gene loci in *B. distachyon*, the average length observed in *B. sylvaticum* transcriptomes is larger than *B. distachyon*.

Downloaded From: https://bioone.org/journals/Applications-in-Plant-Sciences on 16 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

TABLE 2. BLAST comparisons of *Brachypodium sylvaticum* transcriptomes against various databases. Shown are the number and percentage of contigs from *B. sylvaticum* samples from Corvallis (Brasy-Cor), Spain (Brasy-Esp), and Greece (Brasy-Gre) that hit a gene from the respective database. BLASTx comparisons were made to the GenBank nonredundant peptide database (nr), while BLASTn nucleotide comparisons were made against the *B. distachyon*, *Oryza sativa*, and *Sorghum bicolor* cDNA databases (*E*-value threshold cutoff of 1e-10^{-5}).

	Brasy-Cor (120,091)		Brasy-Esp (104,950)		Brasy-Gre (177,682)	
Database compared	No. of hits	% hits	No. of hits	% hits	No. of hits	% hits
GenBank peptide (nr)	96,140	80.1	86,791	82.7	149,178	84.0
B. distachyon v1.2	103,752	86.4	93,291	88.9	160,309	90.2
O. sativa (japonica) vMSU6	96,375	80.3	86,577	82.5	148,617	83.6
S. bicolor v1.4	93,751	78.1	84,023	80.1	144,884	81.5

gene paralogs, and mis-assemblies. As discussed below, the vast majority of assembled contigs are well-annotated, indicating that the inflated number of contigs are not due to mis-assemblies but more likely due to transcript isoforms and/or paralogous loci.

Homology-based transcriptome annotation—We examined the number of homologous loci identified through BLAST to approximate the diversity and coverage of the assembled gene loci (Mount, 2007). We annotated genes based on BLAST similarity (*E*-value threshold cutoff of 1e-10^{-5}) to sequences available in the GenBank protein database and directly to transcripts identified in the sequenced monocot genomes. We used our transcript scaffolds as queries in BLASTx to search against the National Center for Biotechnology Information (NCBI) nonredundant protein database (Table 2), which resulted in >80% correlation between the contigs of each *B. sylvaticum* assembly and GenBank peptide sequences.

Similar trends were observed when we compared the *B. sylvaticum* nucleotide sequences directly against other monocot nucleotide sequences using BLASTn (*E*-value threshold cutoff of 1e-10^{-5} and percent identity >90%). The *B. distachyon* Bd21 (v1.2), *Oryza sativa* (Japonica, MSU6), and *Sorghum bicolor* (v1.4) transcriptome databases were obtained from Gramene BioMart (Spooner et al., 2012). In the Corvallis assembly, 103,752 (86.4%) contigs hit *B. distachyon* genes, 96,375 (80.3%) contigs hit rice genes, and 93,751 (78.1%) contigs hit *S. bicolor* genes, with similar results for the Spain and Greece assemblies (Table 2). We further examined the number of homologous loci identified through direct comparisons between *B. sylvaticum* and *B. distachyon* (Fig. 2). We found that the Corvallis assembly uncovered 28,791 (92.8%) of the

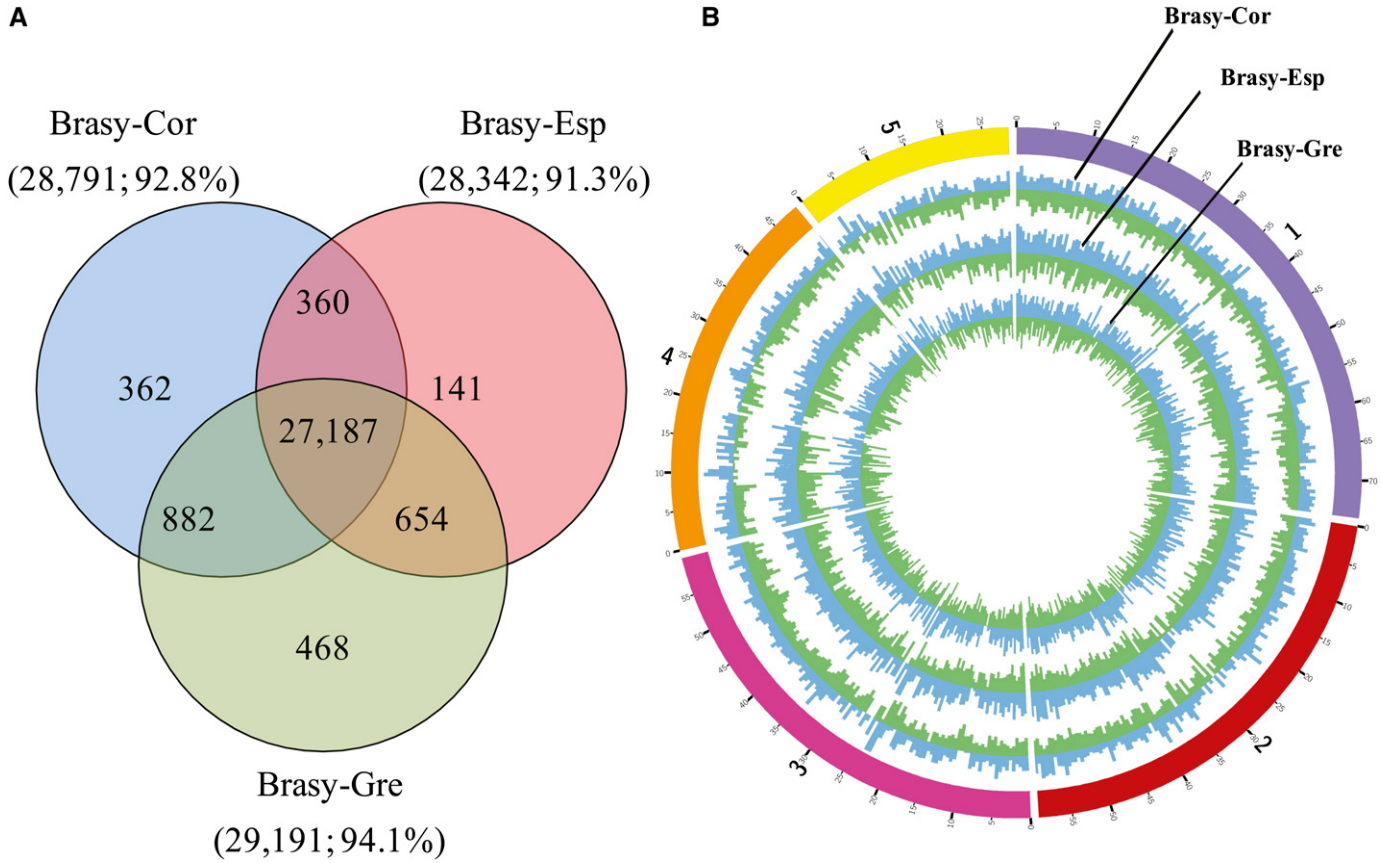

Fig. 2. Sequence comparisons between contigs from *Brachypodium sylvaticum* samples from Corvallis (Brasy-Cor), Spain (Brasy-Esp), and Greece (Brasy-Gre) and *B. distachyon* transcripts and genome. (A) *B. distachyon* genes hit by *B. sylvaticum* contigs. Greater than 87% of *B. distachyon* loci were hit by all three *B. sylvaticum* transcriptomes, and more than 96% were hit by a minimum of one contig from at least one *B. sylvaticum* transcriptome. (B) RPKM values averaged over 0.5 megabase intervals across the *B. distachyon* genome. This image shows the uniform coverage of *B. sylvaticum* reads aligned to the five *B. distachyon* chromosomes. Blue histograms indicate RPKM values averaged along the positive strand while green histograms indicate RPKM values of minus strand alignments.

Downloaded From: https://bioone.org/journals/Applications-in-Plant-Sciences on 16 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

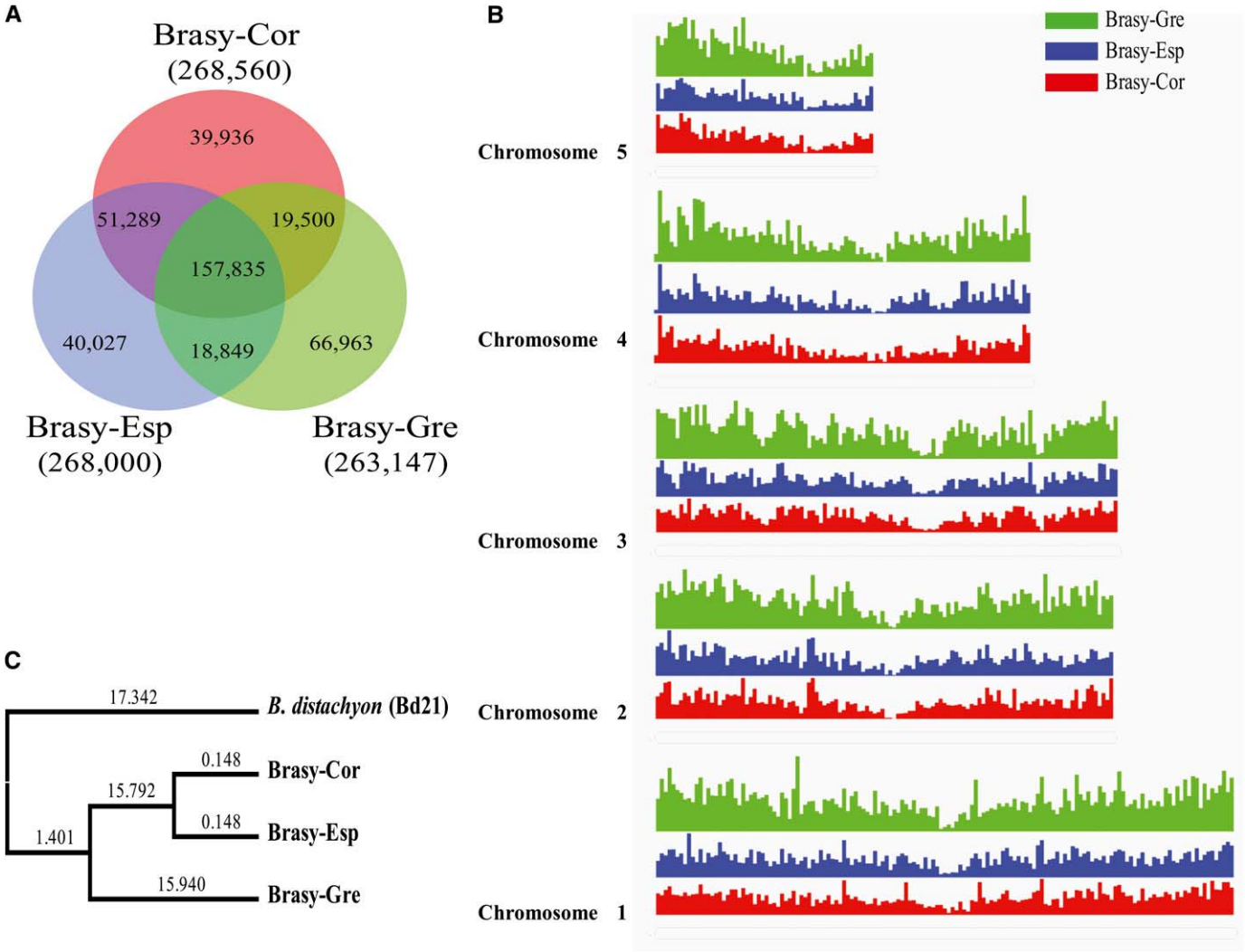

Fig. 3. Comparison of SNPs identified by mapping sequence reads from *Brachypodium sylvaticum* genotypes from Corvallis (Brasy-Cor), Spain (Brasy-Esp), and Greece (Brasy-Gre) to the *B. distachyon* Bd21 reference genome. (A) Venn diagram showing the overlap of SNPs within the three *B. sylvaticum* genotypes. The numbers inside the Venn diagram exclude the 628 common sites exhibiting variation among the *B. sylvaticum* genotypes. We identified more than 157,835 nucleotide variants common among the three *B. sylvaticum* genotypes when compared to *B. distachyon*. Brasy-Gre had the greatest number of SNPs identified. (B) Genotype-specific SNPs were mapped to show the SNP density and chromosome-wide distribution against the *B. distachyon* genome. (C) Phylogenetic construction of a maximum likelihood tree based on the 628 common sites exhibiting variation among the *B. sylvaticum* genotypes indicates the proposed evolutionary relationship among the *B. sylvaticum* genotypes and the *B. distachyon* Bd21 reference genome.

31,029 *B. distachyon* v1.2 transcripts, while the Spain and Greece assemblies hit 91.3% and 94.1% of the *B. distachyon* transcripts, respectively (Fig. 2A). When we compared the *B. distachyon* genes hit by all three *B. sylvaticum* assemblies, we found that ~96.8% (30,054) of the *B. distachyon* genes were hit by a contig from at least one of the three *B. sylvaticum* transcriptomes. In addition, 27,187 (87.6%) *B. distachyon* genes were commonly hit by all three *B. sylvaticum* transcriptome assemblies.

Further, when we extended our comparisons with other sequenced grass genomes, we found that the Corvallis contigs hit 44,538 (~67%) of 66,338 rice transcripts and 24,999 (~84%) of 29,448 sorghum transcripts with similar results for the other two *B. sylvaticum* assemblies. Overall, ~90% of all the *B. sylvaticum* contigs were assigned a homology-based annotation. While the properties of any transcriptome are uniquely associated with the spatial, temporal, and environmental factors present at the precise time of tissue sampling, these results indicate that we have sequenced the great majority of *B. sylvaticum* gene loci and have assembled three quality reference transcriptomes for *B. sylvaticum.*

Functional characterization of the transcriptome— All the *B. sylvaticum* contigs were translated into peptides by querying the longest predicted open reading frame (ORF) using the ORFPredictor tool (Min et al., 2005) and were functionally characterized using InterProScan version 4.8 (Quevillon et al., 2005; Hunter et al., 2012). Nearly half of the translated ORFs were assigned InterPro and Gene Ontology (GO) annotations (Appendix S5), which is consistent with other published annotated genomes. The functional annotations were enriched by performing Blast2GO analysis by adopting a stringent BLASTx search (*E*-value ≤1e-20 and percent identity ≥90%) against the NCBI GenBank nonredundant protein database (Conesa and Gotz, 2008), and the resulting best hits with GO annotations were used to project GO assignments to *B. sylvaticum* contigs (Gotz et al., 2008; Barrell et al., 2009). Enrichment resulted in assigning GO annotations to 69,628 Corvallis, 61,015 Spain, and 107,700 Greece contig assemblies (Appendix S6).

Gene expression analysis— We conducted relative gene expression analyses to assess the utility of these reference transcriptomes for future differential gene expression studies. We mapped *B. sylvaticum* Illumina reads to *B. distachyon* transcripts and calculated the reads per kilobase per million reads (RPKM) values. We then mapped these RPKM values to *B. distachyon* loci on all five chromosomes and graphically depict the RPKM values averaged over

Downloaded From: https://bioone.org/journals/Applications-in-Plant-Sciences on 16 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

0.5 megabase intervals (Fig. 2B). When we compared the RPKM values among the three *B. sylvaticum* plants, we found very high Pearson's correlation coefficients between gene expression data sets (Appendix S7). Although the three *B. sylvaticum* plants do not exhibit significant expression differences over our mapped intervals, we do observe a uniform distribution of contigs mapping to the *B. distachyon* genome (Fig. 2B; *B. distachyon* v1.0 genome sequence from http://mips.helmholtz-muenchen.de/plant/brachypodium/). To further demonstrate the utility of pairing the *B. sylvaticum* transcriptomes and their close congener to investigate questions regarding gene expression, we used the BrachyCyc pathway tool (http://pathway.gramene.org/gramene/brachycyc.shtml), which contains biochemical pathways consisting of over 7000 *B. distachyon* genes coding for enzymes. Using the BrachyCyc pathway tool, we mapped RPKM values to *B. distachyon* metabolic pathways to examine gene expression profiles of homologous genes (Appendix S7). These results demonstrate the potential of the *B. sylvaticum* transcriptomes for use in future studies investigating differential gene expression and metabolomics, as well as for making comparisons with *B. distachyon*.

Genetic variation—To quantify the number of SNP sites across the three transcriptomes, we mapped the reads from each *B. sylvaticum* sample to the *B. distachyon* genome v1.0 using Bowtie version 0.12.8 (Langmead et al., 2009). We then used custom Perl scripts to identify nucleotide differences in positions with at least eight aligned reads and 75% of those aligned reads confirming the SNP (Kimbrel, unpublished). Using these criteria, we identified 394,654 putative SNPs among the three *B. sylvaticum* genotypes (Fig. 3A; Appendix S8). Of these, 157,835 SNPs were in sites common to all three genotypes. Although the total number of SNPs was similar among the genotypes, we observed more SNPs unique to the Greece sample (66,963) when compared to Corvallis (39,936) and Spain (40,027). To address the biological relevance of these SNPs and their potential role to be studied in the future for relevance to invasive phenotypes, we predicted the potential effects of the variants and identified a diverse set of consequences (McLaren et al., 2010). Notably, we identified more than 230,000 downstream variants, more than 92,000 missense variants, and 234 stop codons introduced (Tables 3, 4). We observed only slight variation when we mapped the SNP densities of each *B. sylvaticum* genotype onto *B. distachyon* chromosomes. Generally, the *B. sylvaticum* SNP density mirrors the *B. distachyon* gene density and centromeric region (Fig. 3B). We also constructed a maximum parsimony tree based on concatenated variant sites where at least five reads from each *B. sylvaticum* transcriptome aligned with Bowtie (Fig. 3C). Of the 157,835 SNPs, only 628 loci were polymorphic within at least one of the three *B. sylvaticum* samples. These 628 positions were concatenated to generate the maximum likelihood tree depicting relative relationships among the three *B. sylvaticum* genotypes. While this SNP analysis shows the utility of the *B. sylvaticum* transcriptomes for genotyping studies, much work needs to be done to fully elucidate the relationships of the various native and invasive populations.

We mined the assembled *B. sylvaticum* contigs for SSRs using Perl code from the Simple Sequence Repeat Identification Tool (SSRIT; Temnykh et al., 2001; http://www.gramene.org/db/markers/ssrtool), looking for di-, tri-, tetra-, penta-, and hexanucleotide SSRs with a minimum of nine, six, six, five, and five repeat units, respectively (Table 5; Appendix S9). In total, we identified 23,535 SSRs in Corvallis, 20,303 in Spain, and 32,847 in Greece transcriptome sequences (Table 5). These SSRs were identified in 18,281 contigs from Corvallis, 15,975 from Spain, and 25,567 from Greece. To test whether SSRs showing polymorphism in our computational analysis can be used to develop potential genetic markers, we conducted a test PCR amplification of a sample SSR site from the three sequenced genotypes as well as additional *B. sylvaticum* plants from Oregon, USA, and Europe (Appendix S10).

Brachypodium sylvaticum resources and data download—All sequence, annotations, and data files are available from the project website (http://jaiswallab.cgrb.oregonstate.edu/genomics/brasy). SNPs and contig alignments to *B. distachyon* are also available from the *B. distachyon* Bd21 genome browsers available from BrachyBase (http://www.brachypodium.org/) and Gramene (http://www.gramene.org/Brachypodium_distachyon/) databases. The raw sequence data are available from the NCBI Sequence Read Archive (SRA; accession SRA062855).

CONCLUSIONS

The list of noxious invasive plants identified by the Oregon Department of Agriculture (2009) includes species such as kudzu, goatgrass, knapweed, and others that exact high economic and ecological costs in regions where they have been introduced. Ecologists make distinctions among species that are introduced (able to persist), naturalized (establishes self-sustaining populations, but is not a dominant component of the vegetation), and invasive (is able to dominate habitats to the exclusion of native species). False brome meets the "invasive" criteria but has not yet become notorious because its distribution is still restricted compared to other invasive plants. Our ultimate goal is to use *B. sylvaticum* as a model for studying adaptation and invasiveness and for the general study of grasses. Therefore, we consider it a necessary first step to establish baseline resources for *B. sylvaticum* by generating de novo transcriptomes from multiple genotypes, use them to study gene expression and regulation, and identify functional nucleotide polymorphisms to develop new sets of genetic markers for future population-wide

Table 3. SNP consequence predictions. After aligning all of the 394,654 SNPs from the *Brachypodium sylvaticum* samples from Corvallis (Brasy-Cor), Spain (Brasy-Esp), and Greece (Brasy-Gre) on the *B. distachyon* (Bd21) genome v1.0, we predicted the potential effects of SNP variant loci on the Bd21 genes. We identified a diverse set of consequences for the various gene loci.

Predicted variant effect	No. of SNP sites
3′ UTR variant	46,871
3′ UTR variant, splice region variant	48
5′ UTR variant	6636
5′ UTR variant, splice region variant	34
Downstream gene variant	231,969
Initiator codon variant	38
Intron variant	36,334
Intron variant, splice region variant	13,336
Missense variant	92,532
Missense variant, splice region variant	310
Splice acceptor variant	3222
Splice donor variant	5502
Splice region variant, initiator codon variant	1
Stop gained	234
Stop gained, splice region variant	3
Stop lost	117
Stop lost, splice region variant	2
Stop retained variant	180
Synonymous variant	185,986
Synonymous variant, splice region variant	958
Upstream gene variant	174,506

Table 4. Characterization of the number of transitions and transversions that were predicted from our SNP analysis, showing a much higher prevalence of transitions over transversions as expected. The SNPs were identified in the *Brachypodium sylvaticum* samples from Corvallis (Brasy-Cor), Spain (Brasy-Esp), and Greece (Brasy-Gre).

SNP	Substitution type	Brasy-Cor	Brasy-Esp	Brasy-Gre
A → C	Transversion	14,208	13,836	14,695
A → G	Transition	43,635	42,985	43,135
A → T	Transversion	9,695	9,834	10,118
C → A	Transversion	11,998	12,241	11,459
C → G	Transversion	18,834	18,072	17,862
C → T	Transition	36,201	37,250	34,717
G → A	Transition	35,918	36,980	34,383
G → C	Transversion	18,962	18,103	17,813
G → T	Transversion	11,892	12,143	11,393
T → A	Transversion	9,647	9,970	10,039
T → C	Transition	44,121	43,302	43,790
T → G	Transversion	14,077	13,912	14,371

Downloaded From: https://bioone.org/journals/Applications-in-Plant-Sciences on 16 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

Table 5. Summary of SSR sites identified in the transcriptomes of *Brachypodium sylvaticum* samples from Corvallis (Brasy-Cor), Spain (Brasy-Esp), and Greece (Brasy-Gre). We identified di-, tri-, tetra-, penta-, and hexanucleotide SSRs with a minimum of nine, six, six, five, and five repeat units, respectively. More than 20,000 SSRs were identified in the *B. sylvaticum* transcriptomes; trinucleotide repeats were the largest class of SSRs.

Population	Total	Dimer	Trimer	Tetramer	Pentamer	Hexamer
Brasy-Cor	23,535	6011	16,346	750	183	245
Brasy-Esp	20,303	5449	14,093	513	141	107
Brasy-Gre	32,847	8194	23,280	858	276	239

screening. The results of our de novo assemblies produced a relatively large number of long, reconstructed transcripts, as demonstrated by average contig lengths. Overall, we were able to assign homology-based annotations to ~90% of *B. sylvaticum* contigs, and more than 50% of the translated sequences were functionally annotated by assigning InterPro and Gene Ontology annotations. More than 96% of *B. distachyon* Bd21 gene loci were associated with *B. sylvaticum* contigs, thereby demonstrating diversity and broad coverage in our transcriptomic data. When compared to *B. distachyon*, we discovered ~390,000 SNPs, of which ~158,000 SNPs were common to the three *B. sylvaticum* samples. Based on the SNP calls, we identified the number of SNPs with consequences to the gene and transcripts including those altering the potential intron splicing sites and translated protein sequences. These resources, when paired with well-established *B. distachyon* genomic data, will be useful in the future characterization of *B. sylvaticum* invasiveness.

LITERATURE CITED

Aarrestad, P. A. 2000. Plant communities in broad-leaved deciduous forests in Hordaland county, Western Norway. *Nordic Journal of Botany* 20: 449–466.

Barrell, D., E. Dimmer, R. P. Huntley, D. Binns, C. O'Donovan, and R. Apweiler. 2009. The GOA database in 2009: An integrated Gene Ontology Annotation resource. *Nucleic Acids Research* 37(Supplement 1): D396–D403.

Chambers, K. L. 1966. Notes on some grasses of the Pacific Coast. *Madroño* 18: 250–251.

Conesa, A., and S. Götz. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics* 2008: 619832.

Foote, T. N., S. Griffiths, S. Allouis, and G. Moore. 2004. Construction and analysis of a BAC library in the grass *Brachypodium sylvaticum*: Its use as a tool to bridge the gap between rice and wheat in elucidating gene content. *Functional and Integrative Genomics* 4(1): 26–33.

Fox, S., S. Filichkin, and T. C. Mockler. 2009. Applications of ultra-high-throughput sequencing. *In* D. A. Belostotsky [ed.], Methods in molecular biology, vol. 553: Plant systems biology, 79–108. Humana Press, Totowa, New Jersey, USA.

Götz, S., J. M. Garcia-Gomez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M. Robles, et al. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* 36(10): 3420–3435.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29(7): 644–652.

Holmes, S. E., B. A. Roy, J. P. Reed, and B. J. Johnson. 2010. Context-dependent pattern and process: The distribution and competitive dynamics of an invasive grass, *Brachypodium sylvaticum*. *Biological Invasions* 12: 2303–2318.

Holten, J. I. 1980. Distribution and ecology of *Brachypodium sylvaticum*, *Bromus benekeni* and *Festuca altissima* in central Norway. *Blyttia* 38: 137–144.

Huang, L., X. Yang, P. Sun, W. Tong, and S. Hu. 2012. The first Illumina-based de novo transcriptome sequencing and analysis of safflower flowers. *PLoS One* 7(6): e38653.

Hull, A. J. C. 1974. Species for seeding mountain rangelands in southeastern Idaho, northeastern Utah, and western Wyoming. *Journal of Range Management* 27: 150–153.

Hunter, S., P. Jones, A. Mitchell, R. Apweiler, T. K. Attwood, A. Bateman, T. Bernard, et al. 2012. InterPro in 2011: New developments in the family and domain prediction database. *Nucleic Acids Research* 40(D1): D306–D312.

Kaye, T. N., and M. Blakeley-Smith. 2006. False-brome (*Brachypodium sylvaticum*). University of Washington Press, Seattle, Washington, USA.

Kirby, K. J., and R. C. Thomas. 2000. Changes in the ground flora in Wytham Woods, southern England from 1974 to 1991—Implications for nature conservation. *Journal of Vegetation Science* 11: 871–880.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10(3): R25.

Lionakis Meyer, D. J., and J. Effenberger. 2010. California noxious weed disseminules identification manual. California Department of Food and Agriculture, Sacramento, California, USA.

Long, G. M. 1989. Morphological and physiological variation in *Brachypodium sylvaticum*. Ph.D. dissertation, College of Cardiff, University of Wales, Cardiff, United Kingdom.

Martin, P. H., C. D. Canham, and P. L. Marks. 2009. Why forests appear resistant to exotic plant invasions: Intentional introductions, stand dynamics, and the role of shade tolerance. *Frontiers in Ecology and the Environment* 7: 142–149.

McLaren, W., B. Pritchard, D. Rios, Y. Chen, P. Flicek, and F. Cunningham. 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 26(16): 2069–2070.

Min, X. J., G. Butler, R. Storms, and A. Tsang. 2005. OrfPredictor: Predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Research* 33(Supplement 2): W677–W680.

Mount, D. W. 2007. Using the Basic Local Alignment Search Tool (BLAST). *Cold Spring Harbor Protocols* 2007: doi:10.1101/pdb.top17.

Mur, L. A., J. Allainguillaume, P. Catalan, R. Hasterok, G. Jenkins, K. Lesniewska, I. Thomas, and J. Vogel. 2011. Exploiting the *Brachypodium* tool box in cereal and grass research. *New Phytologist* 191(2): 334–347.

Murchie, E. H., and P. Horton. 1997. Acclimation of photosynthesis to irradiance and spectral quality in British plant species: Chlorophyll content, photosynthetic capacity and habitat preference. *Plant Cell and Environment* 20: 438–448.

Nicolai, M., C. Pisani, J. P. Bouchet, M. Vuylsteke, and A. Palloix. 2012. Discovery of a large set of SNP and SSR genetic markers by high-throughput sequencing of pepper (*Capsicum annuum*). *Genetics and Molecular Research* 11(3): 2295–2300.

Oregon Department of Agriculture (ODA). 2009. False brome. Noxious Weed Control Program [online]. Website http://www.oregon.gov/ODA/PLANT/WEEDS/profile_falsebrome.shtml [accessed 5 December 2012].

Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, and R. Lopez. 2005. InterProScan: Protein domains identifier. *Nucleic Acids Research* 33(Supplement 2): W116–W120.

Rosenthal, D. M., A. P. Ramakrishnan, and M. B. Cruzan. 2008. Evidence for multiple sources of invasion and intraspecific hybridization in *Brachypodium sylvaticum* (Hudson) Beauv. in North America. *Molecular Ecology* 17(21): 4657–4669.

Roy, B. A. 2010. *Brachypodium sylvaticum*. Invasive species compendium [online]. Website http://www.cabi.org/isc/?compid=5&dsid=9890&loadmodule=datasheet&page=481&site=144 [accessed 5 December 2012].

Downloaded From: https://bioone.org/journals/Applications-in-Plant-Sciences on 16 Apr 2024
Terms of Use: https://bioone.org/terms-of-use

Schulz, M. H., D. R. Zerbino, M. Vingron, and E. Birney. 2012. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28(8): 1086–1092.

Severns, P. M., and A. D. Warren. 2008. Selectively eliminating and conserving exotic plants to save an endangered butterfly from local extinction. *Animal Conservation* 11: 476–483.

Spooner, W., K. Youens-Clark, D. Stains, and D. Ware. 2012. Gramene-Mart: The BioMart data portal for the Gramene project. *Database (Oxford)* 2012: doi:10.1093/database/bar056.

Temnykh, S., G. DeClerck, A. Lukashova, L. Lipovich, S. Cartinhour, and S. McCouch. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Research* 11(8): 1441–1452.

Varshney, R. K., W. Chen, Y. Li, A. K. Bharti, R. K. Saxena, J. A. Schlueter, M. T. Donoghue, et al. 2012. Draft genome sequence of pigionpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nature Biotechnology* 30(1): 83–89.

Wang, Y., X. Zeng, N. J. Iyer, D. W. Bryant, T. C. Mockler, and R. Mahalingam. 2012. Exploring the switchgrass transcriptome using second-generation sequencing technology. *PLoS ONE* 7(3): e34225.

Wang, Z., B. Fang, J. Chen, X. Zhang, Z. Luo, L. Huang, X. Chen, et al. 2010. De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*). *BMC Genomics* 11: 726.

Ward, J. A., L. Ponnala, and C. A. Weber. 2012. Strategies for transcriptome analysis in nonmodel plants. *American Journal of Botany* 99(2): 267–276.

Washington State Department of Agriculture (WSDA). 2009. Noxious weed list [online]. Website http://www.nwcb.wa.gov/nwcb_nox.htm [accessed 5 December 2012].

Wolny, E., K. Lesniewska, R. Hasterok, and T. Langdon. 2011. Compact genomes and complex evolution in the genus *Brachypodium. Chromosoma* 120(2): 199–212.

Zhao, Z., L. Tan, C. Dang, H. Zhang, Q. Wu, and L. An. 2012. Deep-sequencing transcriptome analysis of chilling tolerance mechanisms of a subnival alpine plant, *Chorispora bungeana. BMC Plant Biology* 12(1): 222.

Downloaded From: https://bioone.org/journals/Applications-in-Plant-Sciences on 16 Apr 2024
Terms of Use: https://bioone.org/terms-of-use