1   *Chloroplast Genome Sequencing and Comparative Analysis for*
2   *Five Species of the Turfgrass Fine Fescues (Festuca* **L., Poaceae***)*

3   **Yinjie Qiu**[1,*]**, Cory D. Hirsch**[2]**, Ya Yang**[3] **and Eric Watkins**[1,*]

4   [1]   Department of Horticultural Science, University of Minnesota, St. Paul, MN, 55108, USA;
5   qiuxx221@umn.edu (Y.Q.); ewatkins@umn.edu (E.W.)
6   2   Department of Plant Pathology, University of Minnesota, St. Paul, MN, 55108, USA; cdhirsch@umn.edu (C.
7   H.)
8   3   Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN, 55108, USA;
9   yayang@umn.edu (Y.Y.)
10   *   Correspondence: qiuxx221@umn.edu (Y.Q.); ewatkins@umn.edu (E.W.); Tel.: +1- 612-624-7496 (E.W.)

11   **Abstract:** Fine fescues (*Festuca* L., Poaceae) are turfgrass species that perform well in low-
12   input environments. Based on morphological characteristics, the most commonly-utilized
13   fine fescues are divided into five taxa: three are subspecies within *F. rubra* L. and the
14   remaining two are treated as species within the *F. ovina* L. complex. Morphologically, these
15   five taxa are very similar, and both identification and classification of fine fescues remain
16   challenging. In an effort to develop identification methods for fescues, we used flow
17   cytometry to estimate genome size, and sequenced the chloroplast genome of all five taxa.
18   Fine fescue chloroplast genome sizes ranged from 133,331 to 133,842 bp and contained 113
19   to 114 genes. Phylogenetic reconstruction using whole chloroplast genome sequences
20   agreed with previous work based on morphology. Comparative genomics suggested
21   unique repeat signatures for each fine fescue taxon that could potentially be used for
22   marker development for taxon identification.

23   **Keywords:** Fine fescue, chloroplast genome, phylogeny, comparative genomics
24

1. Introduction

26   With ca. 450 species, Fescues (*Festuca* L., Poaceae) is a large and diverse genus of
27 perennial grasses [1]. Fescue species are distributed mostly in temperate zones of both the
28 northern and southern hemispheres, but most commonly found in the northern hemisphere
29 [2]. Several of the fescue species have been commonly used as turfgrass. Based on both leaf
30 morphology and nuclear ITS sequences, fescue species can be divided into two groups:
31 broad-leaved fescues and fine-leaved fescues [3]. Broad-leaved fescues commonly used as
32 turfgrass include tall fescue (*F. arundinacea* Schreb.) and meadow fescue (*F. pratensis* Huds.).
33 Fine-leaved fescues are a group of cool-season grasses that include fine fescues. Five fine
34 fescue taxa: hard fescue (*F. brevipila* Tracey, 2n=6x=42), sheep fescue (*F. ovina* L., 2n=4x=28),
35 strong creeping red fescue (*F. rubra* ssp. *rubra* 2n=8x=56), slender creeping red fescue (*F.
36 rubra* ssp. *litoralis* (G. Mey.) Auquier 2n=6x=42), and Chewings fescue (*F. rubra* ssp. *fallax*
37 (Thuill.) Nyman 2n=6x=42) are commonly used as perennial turfgrasses [4]. All five taxa
38 share very fine and narrow leaves and have been used for forage, turf, and ornamental
39 purposes. They are highly tolerant to shade and drought, prefer low pH (5.5-6.5) and low
40 fertility soils [5]. Additionally, fine fescues grow well in the shade or sun, have reduced
41 mowing requirements, and do not need additional fertilizer or supplemental irrigation [4].
42   Based on morphological and cytological features, fine fescues are currently divided into
43 two groups referred to as the *F. rubra* complex (includes *F. rubra* ssp. *litoralis*, *F. rubra* ssp.
44 *rubra*, *F. rubra* ssp. *fallax*) and the *F. ovina* complex (includes *F. brevipila* and *F. ovina*) [4].
45 While it is relatively easy to identify fine fescue species into their proper complex, it is
46 challenging to identify taxa within the same complex. In the *F. rubra* complex, both ssp.

47   *litoralis* and ssp. *rubra* are rhizomatous while ssp. *fallax* is non-rhizomatous. However, the
48   separation of ssp. *litoralis* from ssp. *rubra* using rhizome length is challenging. Species
49   identification within the *F. ovina* complex relies heavily on leaf characters; however,
50   abundant morphological and ecotype diversity within *F. ovina* makes species identification
51   difficult [6]. This is further complicated by inconsistent identification methods between
52   different continents. For example, in the United States, sheep fescue is described as having
53   a bluish gray leaf color and hard fescue leaf blade color is considered green [5], while in
54   Europe, it is the opposite [7]. Beyond morphological classifications, laser flow cytometry has
55   been used to determine ploidy level of fine fescues and some other fescue species [8]. A wide
56   range of DNA contents within each complex suggests that the evolutionary history of each
57   named species is complicated, and interspecific hybridization might interfere with species
58   determination using this approach. Plant breeders have been working to improve fine
59   fescues for turf use for several decades, with germplasm improvement efforts focused on
60   disease resistance, traffic tolerance, and ability to perform well under heat stress [9].
61   Turfgrass breeders have utilized germplasm collections from old turf areas as a source of
62   germplasm [10]; however, confirming the species identity in these collections has been
63   challenging. A combination of molecular markers and flow cytometry could be a valuable
64   tool for breeders to identify fine fescue germplasm [11].

65   Due to the complex polyploidy history of fine fescues, sequencing plastid genomes
66   provides a more cost-effective tool for taxon identification than the nucleus genome because
67   it is maternally inherited, lack of heterozygosity, present in high copies and usable even in
68   partially degraded material. Previous studies have developed universal polymerase chain
69   reaction (PCR) primers to amplify non-coding polymorphic regions for DNA barcoding in
70   plants for species identification [12, 13]. However, the polymorphisms discovered from
71   these regions are often single nucleotide polymorphisms that are difficult to apply using
72   PCR screening methods. For these reasons, it would be helpful to assemble chloroplast
73   genomes and identify simple sequence repeat (SSR) polymorphism and tandem repeats.
74   Chloroplast genome sequencing has been simplified due to improved sequencing
75   technology. In turfgrass species, high throughput sequencing has been used to assemble the
76   chloroplast genomes of perennial ryegrass (*Lolium perenne* cv. Cashel) [14], tall fescue
77   (*Lolium arundinacea* cv. Schreb) [15], diploid *Festuca ovina*, *Festuca pratensis*, *Festuca altissima*
78   [16], and bermudagrass (*Cynodon dactylon*) [17]. To date, there is limited molecular biology
79   information on fine fescue species identification and their phylogenetic position among
80   other turfgrass species [16, 18]. In this study, we used flow cytometry to confirm the ploidy
81   level of five fine fescue cultivars, each representing one of the five commonly utilized fine
82   fescue taxon. We then reported the complete chloroplast genome sequences of these five
83   taxa, carried out comparative genomics and phylogenetic inference. Based on the genome
84   sequence we identified unique genome features among fine fescue taxa and predicted taxa
85   specific SSR and tandem repeat loci for molecular marker development.

86   **2. Results**

87   *2.1 Species Ploidy Level Confirmation*

88   We used flow cytometry to estimate the ploidy levels of five fine fescue taxa by
89   measuring the DNA content in each cell. DNA content was reflected by the flow cytometry
90   mean PI-A value. Overall, fine fescue taxa had mean PI-A values roughly from 110 to 180
91   (**Figure 1 and Figure S1**). *F. rubra* ssp. *rubra* cv. Navigator II (2n=8x=56) had the highest mean
92   PI-A value (181.434, %rCV 4.4%). *F. rubra* ssp. *litoralis* cv. Shoreline (2n=6x=42) and *F. rubra*

93   ssp. *fallax* cv. Treazure II (2n=6x=42) had similar mean PI-A values of 137.852, %rCV 3.7 and
94   145.864, %rCV 3.5, respectively. *F. brevipila* cv. Beacon (2n=6x=42) had a mean PI-A of 165.25,
95   %rCV 1.9, while *F. ovina* cv. Quatro (2n=4x=28) had a mean PI-A of 108.43, %rCV 2.9.
96   Standard reference *L. perenne* cv. Artic Green (2n=2x=14) had a G1 phase mean PI-A of 63.91,
97   %rCV 3.0. USDA *F. ovina* PI 230246 (2n=2x=14) had a G1 mean PI-A of 52.73 (histogram not
98   shown). The estimated genome size of USDA PI 230246 was 4.67 pg/2C. Estimated ploidy
99   level of *F. brevipila* cv. Beacon was 6.3, *F. ovina* cv. Quatro was 4.11, *F. rubra* ssp. *rubra* cv.
100   Navigator II was 6.9, *F. rubra* ssp. *litoralis* cv. Shoreline was 5.2, and *F. rubra* ssp. *fallax* cv.
101   Treazure II was 5.5 (**Table 1**). All newly estimated ploidy levels roughly correspond to
102   previously reported ploidy levels based on chromosome counts.
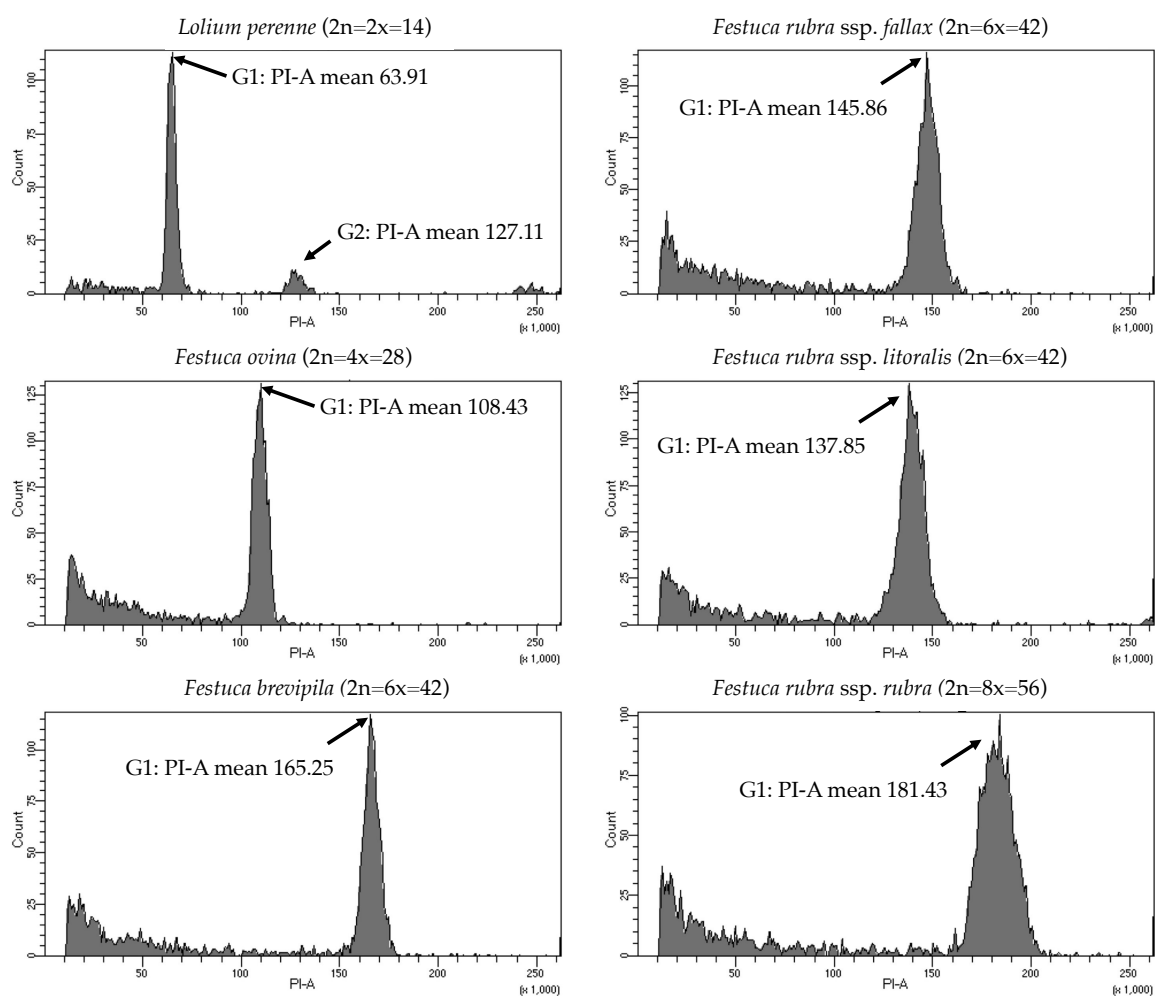
104   **Figure 1.** Flow cytometry results for the five fine fescue taxa. *Lolium* (2n=2x=14) was used as the
105   reference. Flow cytometry was able to separate *F. rubra* ssp. *rubra* from the other two subspecies in
106   the *F. rubra* complex. The mean PI-A values of *F. rubra* ssp. *fallax* and *F. rubra* ssp. *litoralis* were similar
107   (145.86 to 137.85).

108   **Table 1:** Summary of flow cytometry statistics, genome size estimation, and ploidy level estimation
109   of fine fescue species. *Lolium perenne* 2C DNA content was used to calculate fine fescue and USDA *F.*
110   *ovina* PI 230246 genome size, calculated PI 230246 DNA content was used as reference to infer fine
111   fescue ploidy level

Species name	Chromosome count	Cultivar name	Mean PI-A	%rCV *	Estimated Genome Size (pg/Nuclei)	Estimated Ploidy Level

F. brevipila	2n=6x=42	Beacon	165.3	1.9	14.6	6.3
F. ovina	2n=4x=28	Quatro	108.4	2.9	9.6	4.1
F. ovina PI 230246	2n=2x=14	NA	52.7	3.1	4.7	1.7
F. rubra ssp. rubra	2n=8x=56	Navigator II	181.4	4.4	16.1	6.9
F. rubra ssp. litoralis	2n=6x=42	Shoreline	137.9	3.7	12.2	5.2
F. rubra ssp. fallax	2n=6x=42	Treazure II	145.9	3.5	12.9	5.5
L. perenne	2n=2x=14	Artic Green	63.9	3.0	5.7	2.0

112   *** %rCV:** Quality of laser alignment. Low %rCV suggests high resolution sensitity.

113   *2.2 Plastid Genome Assembly and Annotation of Five Fescue Taxa*

114   A total of 47,843,878 reads were produced from the five fine fescue taxa. After Illumina
115   adaptor removal, we obtained 47,837,438 reads. The assembled chloroplast genomes ranged
116   from 133,331 to 133,842 bp. The large single copy (LSC) and small single copy (SSC) regions
117   were similar in size between the sequenced fine fescue accessions (78 kb and 12 kb,
118   respectively). *Festuca ovina* and *F. brevipila* in the *F. ovina* complex had exactly the same size
119   inverted repeat (IR) region (42,476 bp). In the *F. rubra* complex, *F. rubra* ssp. *rubra* and *F.*
120   *rubra* ssp. *litoralis* had the same IR size (21,235 bp). Species in the *F. rubra* complex had a
121   larger chloroplast genome size compared to species in the *F. ovina* complex. All chloroplast
122   genomes shared similar GC content (38.4%) (**Figure 2**, **Table 2**). The fine fescue chloroplast
123   genomes encoded for 113-114 genes, including 37 transfer RNAs (tRNA), 4 ribosomal RNAs
124   (rRNA), and 72 protein-coding genes (**Table 2**). Genome structures were similar among all
125   five fine fescue taxa sequenced, except that the pseudogene *accD* was annotated in all three
126   subspecies of *F. rubra*, but not in the *F. ovina* complex (**Table 3**).

127
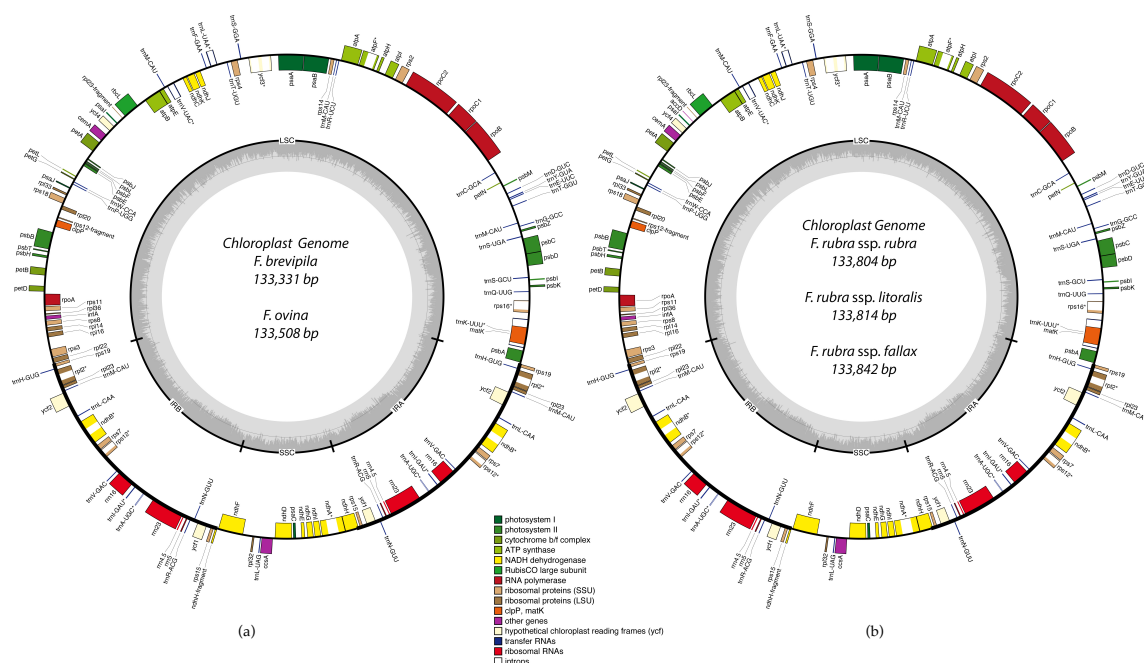
129   **Figure 2**: Whole chloroplast genome structure of *F. ovina* complex (a) and *F. rubra* complex (b). Genes
130   inside the circle are transcribed clockwise, genes outside are transcribed counter-clockwise. Genes
131   belong to different functional groups are color coded. GC content is represented by the dark gray
132   inner circle, the light gray corresponded to the AT content. IRA(B), inverted repeat A(B); LSC, large
133   single copy region; SSC, small single copy region.

134 **Table 2**: Characteristics of fine fescue chloroplast genomes.

	F. brevipila cv. Beacon	F. ovina cv. Quatro	F. rubra ssp. rubra cv. Navigator II	F. rubra ssp. litoralis cv. Shoreline	F. rubra ssp. fallax cv. Treazure II
Total Genome Size (bp)	133,331	133,508	133,804	133,814	133,842
Large Single Copy (bp)	78,462	78,632	78,888	78,909	78,882
Small Single Copy (bp)	12,393	12,400	12,446	12,435	12,452
Inverted Repeat (bp)	42,476	42,476	42,470	42,470	42,508
Ratio of LSC (%)	58.85	58.9	58.96	58.97	58.94
Ratio of SSC (%)	9.29	9.29	9.3	9.29	9.3
Ratio of IRs (%)	31.86	31.82	31.74	31.74	31.76
GC content (%)	38.4	38.4	38.4	38.4	38.4

135 **Table 3.** Fine fescue chloroplast genomes gene content by gene category.

Category	Group of Gene	Name of gene			
Self-replication (58/77)	Ribosomal RNA genes (4/8)	rrn4.5 [a]	rrn5 [a]	rrn16 [a]	rrn23 [a]
	Transfer RNA genes (27/38)	trnA-UGC *[a]	trnC-GCA	trnD-GUC	trnE-UUC
		trnF-GAA	trnG-GCC	trnH-GUG [a]	trnI-GAU *[a]
		trnK-UUU*	trnL-CAA [a]	trnL-UAA*	trnL-UAG
		trnM-CAU [c]	trnN-GUU [a]	trnP-UGG	trnQ-UUG
		trnR-ACG [a]	trnR-UCU	trnS-GCU	trnS-GGA
		trnS-UGA	trnT-GGU	trnT-UGU	trnV-GAC[a]
		trnV-UAC*	trnW-CCA	trnY-GUA	
	Small subunit of ribosome (12/16)	rps2	rps3	rps4	rps7 [a]
		rps8	rps11	rps12 *[ab]	rps14
		rps15 [a]	rps16*	rps18	rps19 [a]
	Large subunit of ribosome (9/11)	rpl2*[a]	rpl14	rpl16	rpl20
		rpl22	rpl23[a]	rpl32	rpl33
		rpl36			
	RNA polymerase subunits (4)	rpoA	rpoB	rpoC1	rpoC2
Photosynthesis (45/46)	Subunits of Photosystem I (6)	psaA	psaB	psaC	psaI
		psaJ	ycf3**		

	Subunits of Photosystem II (15)	*psbA*	*psbB*	*psbC*	*psbD*
		psbE	*psbF*	*psbH*	*psbI*
		psbJ	*psbK*	*psbL*	*psbM*
		psbN	*psbT*	*psbZ*	
	Subunits of cytochrome (6)	*petA*	*petB*	*petD*	*petG*
		petL	*petN*		
	Subunits of ATP synthase (6)	*atpA*	*atpB*	*atpE*	*atpF**
		atpH	*atpI*		
	Large subunit of Rubisco (1)	*rbcL*			
	Subunits of NADH Dehydrogenase (11/12)	*ndhA**	*ndhB*a*	*ndhC*	*ndhD*
		ndhE	*ndhF*	*ndhG*	*ndhH*
		ndhI	*ndhJ*	*ndhK*	
	Translational initiation factor (1)	*infA*			
	Maturase (1)	*matK*			
	Envelope membrane protein (1)	*cemA*			
Other genes (5)	C-type cytochrome (1)	*cssA*			
	Protease (1)	*clp*			
	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta	*accD$*			
Unknown function (5)	Conserved open reading frames (3/5)	*ycf1a*	*ycf2a*	*ycf4*	

136 ᵃ Two gene copies in IRs; ᵇ Gene divided into two independent transcription units; ᶜ Gene that has five copies; *
137 One intron-containing genes; ** Two intron-containing genes. $ Gene annotated in *F. rubra* spp. only. Fine fescue
138 species chloroplast genomes share high structure similarity and gene content. Acetyl-coenzyme A carboxylase
139 carboxyl transferase subunit beta pseudogene is annotated in *F. rubra* ssp.

140 *2.3 Chloroplast Genome IR Expansion and Contraction*

141 Contraction and expansion of the IR regions resulted in the size variation of chloroplast
142 genomes. We examined the four junctions in the chloroplast genomes, LSC/IRa, LSC/IRb,

143 SSC/IRa, and SSC/IRb of the fine fescue and the model turfgrass species *L. perenne*. Although
144 the chloroplast genome of fine fescue species was highly similar, some structural variations
145 were still found in the IR/LSC and IR/SSC boundary. Similar to *L. perenne*, fine fescue species
146 chloroplast genes *rpl22-rps19, rps19-psbA* were located in the junction of IR and LSC; *rps15-*
147 *ndhF* and *ndhH-rps15* were located in the junction of IR/SSC. In the *F. ovina* complex, the
148 *rps19* gene was 37 bp into the LSC/IRb boundary while in the *F. rubra* complex and *L. perenne*,
149 the *rps19* gene was 36 bp into the LSC/IRb boundary (**Figure 3**). The *rsp15* gene was 308 bp
150 from the IRb/SSC boundary in *F. ovina* complex, 307 bp in *F. rubra* complex, and 302 bp in *L.*
151 *perenne*. Both the *ndhH* and the pseudogene fragment of the *ndhH* (*ψndhH*) genes panned the
152 junction of the IR/SSC. The *ψndhH* gene crossed the IRb/SSC boundary with 10 bp into SSC
153 in *F. brevipila* and *F. ovina*, 21 bp in *F. rubra* ssp. *rubra*, 4 bp in *F. rubra* ssp. *litoralis* and *F. rubra*
154 ssp. *fallax*, and 7 bp in *L. perenne*. The *psbA* gene was 87 bp apart from the IRa/LSC boundary
155 into the LSC in *L. perenne* and *F. ovina* complex species but 83 bp in the *F. rubra* complex.
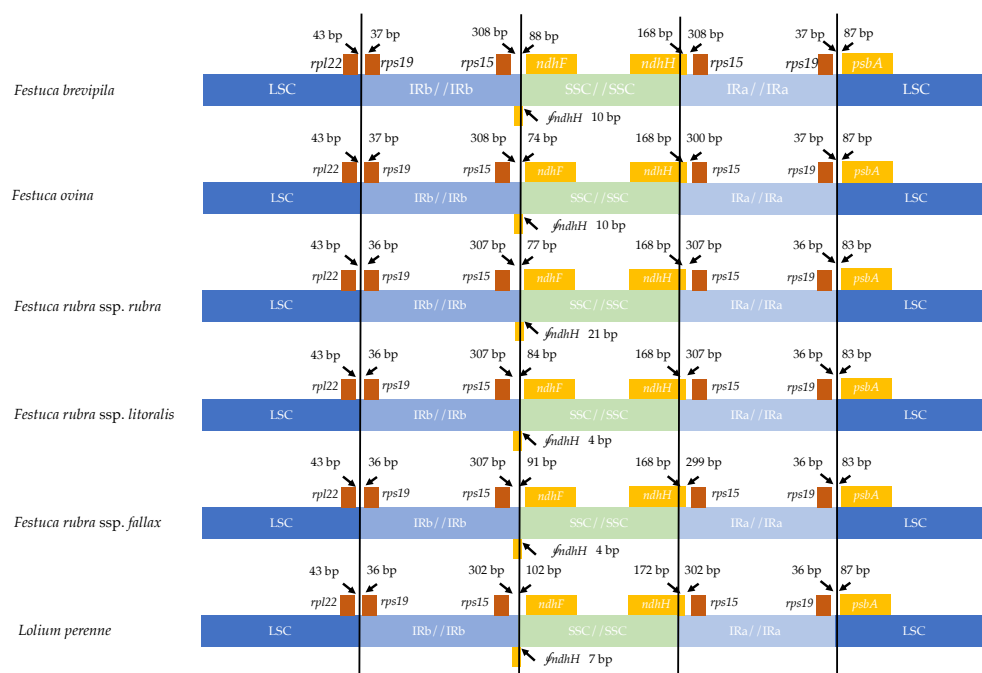
156

157 **Figure 3.** Comparison for border positions of LSC, SSC and IR regions among five fine fescues and *L.*
158 *perenne*. Genes are denoted by boxes, and the gap between the genes and the boundaries are indicated
159 by the number of bases unless the gene coincides with the boundary. Extensions of genes are also
160 indicated above the boxes.

161 *2.4 Whole Chloroplast Genome Comparison and Repetitive Element Identification*

162 Genome-wide comparison among five fine fescue taxa showed high sequence similarity
163 with most variations located in intergenic regions (**Figure 4**). To develop markers for species
164 screening, we predicted a total of 217 SSR markers for fine fescue taxa sequenced (*F. brevipila*
165 39; *F. ovina* 45; *F. rubra* ssp. *rubra* 45; *F. rubra* ssp. *litoralis* 46; *F. rubra* ssp. *fallax* 42) that
166 included 17 different repeat types for the fine fescue species (**Figure 5a, Table S1**). The most
167 frequent repeat type was A/T repeats, followed by AT/AT. The pentamer AAATT/AATTT
168 repeat was only presented in the rhizomatous *F. rubra* ssp. *litoralis* and *F. rubra* ssp. *rubra*,
169 while ACCAT/ATGGT was only found in *F. ovina* complex species *F. brevipila* and *F. ovina*.
170 Similar to SSR loci prediction, we also predicted long repeats for the fine fescue species and
171 identified a total of 171 repeated elements ranging in size from 20 to 51 bp (**Figure 5b, Table**
172 **S2**). Complementary (C) matches were only identified in *F. brevipila* and *F. ovina*. *F. rubra*

173   species had more palindromic (P) and reverse (R) matches. Number of forward (F) matches
174   were similar between five taxa.
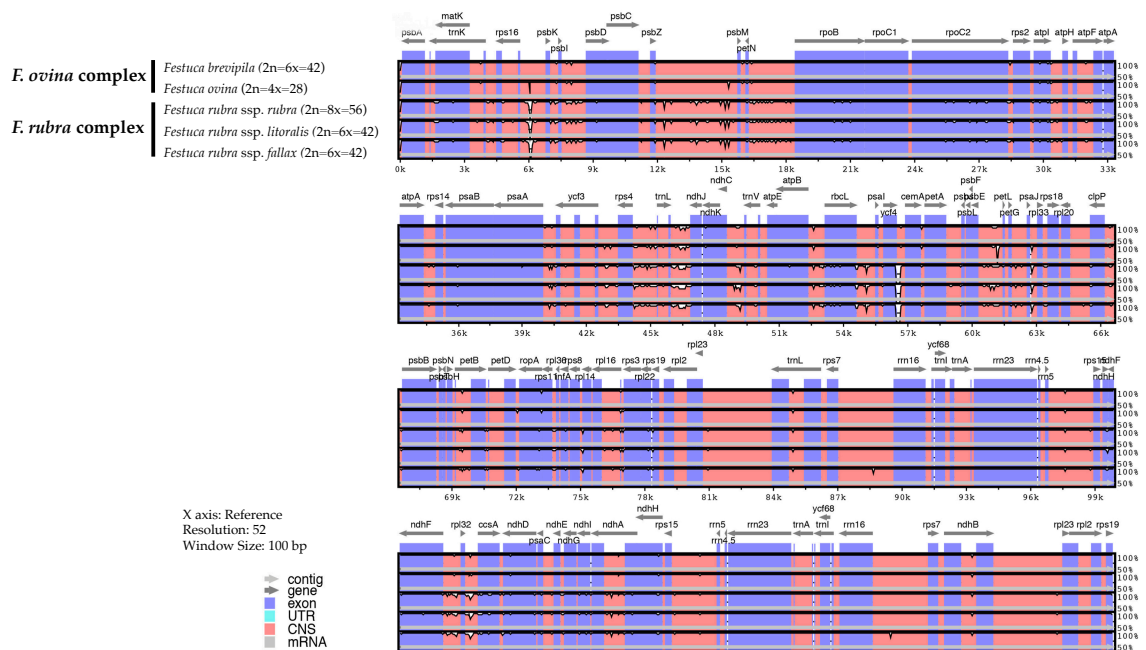
175
176   **Figure 4**. Sequence identity plot of fine fescues chloroplast genome sequences with *F. ovina* (2x) as the
177   reference using mVISTA. A cut-off of 70% identify was used for the plots, and the percent of identity
178   varies from 50% to 100% as noted on the y-axis. Most of the sequence variation between fine fescues
179   were in intergenic regions. Species in the *F. ovina* complex, *F. brevipila* and *F. ovina* showed high
180   sequence similarity. Similarly, subspecies within *F. rubra* complex also showed high sequence
181   similarity.
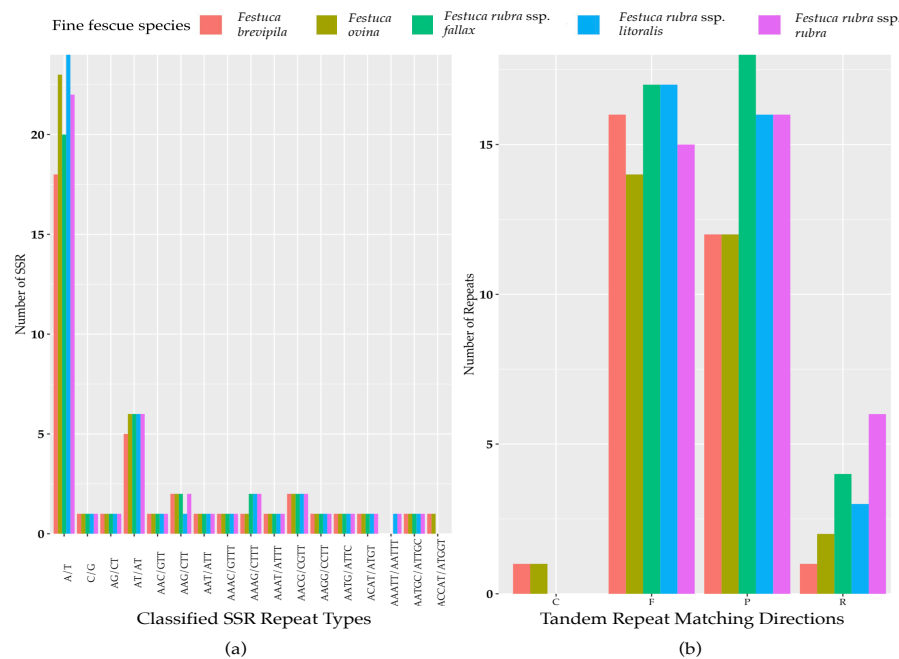
182

183   **Figure 5**. (a) SSR repeat type and numbers in the five fine fescue taxa sequenced. Single nucleotide
184   repeat type has the highest frequency. No hexanucleotide repeats were identified in the fine fescue
185   chloroplast genomes sequenced. One penta-nucleotide repeat type (AAAT/AATTT) is unique to *F*.

186   *rubra* ssp. *rubra* and *F. rubra* ssp. *litoralis*; One penta-nucleotide repeat type (ACCAT/ATGGT) is
187   unique to *F. brevipila* and *F. ovina* (b) Long repeats sequences in fine fescue chloroplast genomes.
188   Complement (C) match was only identified in the *F. ovina* complex; Reverse (R) match has the most
189   number variation in fine fescues.

190   *2.5 SNP and InDel Distribution in the Coding Sequence of Five Fine Fescue Species*

191   To identify single nucleotide polymorphisms (SNPs) we used the diploid *F. ovina* gff3
192   file (JX871940.1) as the template gene model to count SNPs distribution. We found more
193   SNPs were located within intergenic regions in the *F. ovina* complex, while in the *F. rubra*
194   complex, SNPs were located evenly between gene coding and intergenic regions. Most
195   InDels were located in intergenic regions of the fine fescue species (**Table 3**). Between *F.*
196   *ovina* and the *F. rubra* complex, *the ropC2* gene had the most SNPs (4 vs 31). *rbcL* gene also
197   has a high level of variation (1 vs 14.3). *rpoB, ccsA,* NADH dehydrogenase subunit genes
198   (*ndhH, ndhF, ndhA*), and ATPase subunit genes (*atpA, atpB, aptF*) also showed variation
199   between *F. ovina* and *F. rubra* complexes. Less SNP and InDel variation were found within
200   each complex (**Table 4, Table S3 and S4**).

201   **Table 4.** Distribution of SNPs and InDels for the five fine fescue taxa sequenced in this study.

Species	F. brevipila	F. ovina	F. rubra ssp. rubra	F. rubra ssp. litoralis	F. rubra ssp. fallax
Total number of SNPs	98	134	638	615	624
SNPs in the coding region	35	52	306	301	300
SNPs in intergenic region	63	82	332	314	324
Percentage of SNPs in the intergenic region	64.29	61.19	52.04	51.06	51.92
Total number of InDels	112	102	149	156	149
InDels in the coding region	22	17	27	26	27
InDels in intergenic region	90	85	122	130	122
Percentage of InDels in the intergenic region	80.36	83.33	81.88	83.33	81.88

202   *2.6 Nucleotide Diversity and Mutation Hotspot Identification*

203   A sliding window analysis successfully detected highly variable regions in the fine
204   fescue chloroplast genomes (**Figure 6, Table S5**). The average nucleotide diversity (Pi)
205   among fine fescue species was relatively low (0.00318). The IR region showed lower
206   variability than the LSC and SSC region. There were several divergent loci having a Pi value
207   greater than 0.01 (*psbK-psbI, trnfM-trnE, trnC-rpoB, psbH-petB, trnL-trnF, trnS-rps4, aptB-rbcL-*
208   *psaI,* and *rpl32-trnL*), but mostly within intergenic regions. The *rbcL-psaI* region contained a
209   highly variable *accD-like* region in some species, we looked at the structural variation of 10
210   species in the *Festuca - Lolium* complex. We found species in broad-leaved fescue and *F. rubra*
211   complex had similar structure, while *F. ovina* (2x, 4x) and *F. brevipila* had a 276 bp deletion
212   in the *rbcL-psaI* intergenic region (**Figure 7**).
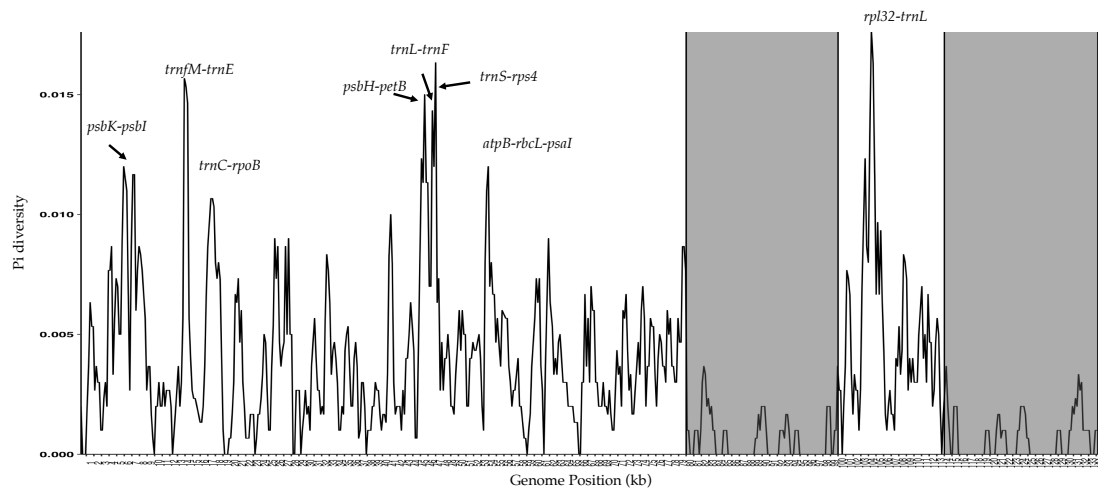
213

214 **Figure 6**. Sliding window analysis of fine fescue whole chloroplast genomes. Window size: 600 bp,
215 step size: 200 bp. X-axis: the position of the midpoint of a window (kb). Y-axis: nucleotide diversity
216 of each window. Inverted repeat regions are highlighted in grey. *rpl32-trnL* region has the most
217 nucleotide diversity followed by *psbH- petB-trnL-trnF-trnS-rps4* region.

218
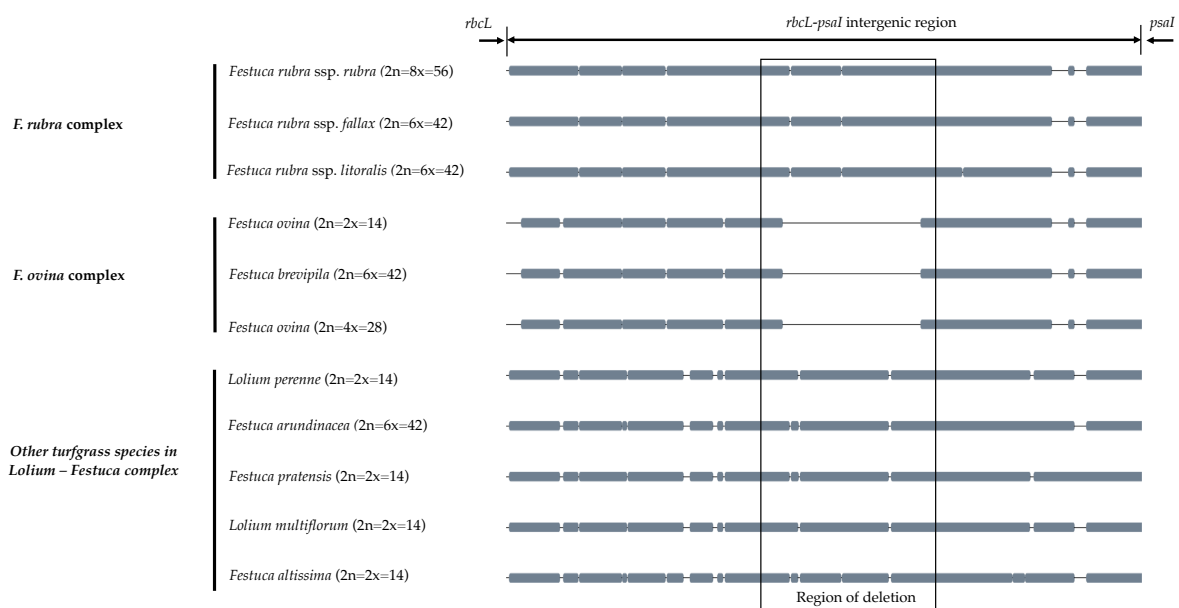
219

220 **Figure 7**. Alignment of *rbcL-psaI* intergenic sequence alignment shows the pseudogene *accD* is missing
221 in both *F. ovina* (2x, 4x) and *F. brevipila* but present in the *Festuca rubra* complex and other species
222 examined in this study. Species were ordered by complexes.

223 *2.7 Phylogenetic Reconstruction of Fine Fescue Species*

224   We reconstructed the phylogenetic relationships of species within the *Festuca - Lolium*
225 complex using the chloroplast genomes sequenced in our study and eight publicly available
226 complete chloroplast genomes including six species within the *Festuca-Lolium* complex
227 (**Figure 8**). The dataset included 125,824 aligned characters, of which 3,923 were parsimony-
228 informative and 91.11% characters are constant. The five fine fescue taxa were split into two
229 clades ([ML]BS=100). In the *F. ovina* complex, two *F. ovina* accessions included in the

230 phylogenetic analysis, a diploid one from GenBank, and a tetraploid one newly sequenced
231 in this study are paraphyletic to *F. brevipila* ([ML]BS=100). All three subspecies of *F. rubra*
232 formed a strongly supported clade ([ML]BS=100). Together they are sisters to the *F. ovina*
233 complex ([ML]BS=100).

234

235 **Figure 8.** Maximum likelihood (ML) phylogram of the *Festuca - Lolium* complex with 1,000 bootstrap
236 replicates. Fine fescues were grouped into previous named complexes (*F. ovina* and *F. rubra*), sister to
237 broad leaved fescues in the *Festuca – Lolium* complex.

3. Discussion

239 In this study, we used flow cytometry to determine the ploidy level of five fine fescue
240 cultivars, assembled the chloroplast genomes for each, and identified structural variation
241 and mutation hotspots. We also identified candidate SSR loci for marker development to
242 facilitate fine fescue species identification. Additionally, we reconstructed the phylogenetic
243 relationships of the *Festuca-Lolium* complex using plastid genome information generated in
244 this study along with other publicly-available plastid genomes.
245 Flow cytometry was able to separate *F. brevipila* cv. Beacon, *F. ovina* cv. Quatro and *F.*
246 *rubra* ssp. *rubra* cv. Navigator II based on the mean PI-A value. We noticed that the average
247 mean PI-A of the diploid *L. perenne* (63.91) was higher than the mean PI-A of diploid *F. ovina*
248 (52.73), suggesting that *F. ovina* species has smaller genome size than *L. perenne*. The ploidy
249 estimation in the *F. ovina* complex are fairly consistent while the estimations of genome sizes
250 in the *F. rubra* complex are smaller than we expected, even though these two complexes are
251 closely related. Indeed, a similar finding was reported by Huff et al [8] that *F. brevipila* has a
252 larger genome size than *F. rubra* ssp. *litoralis* and *F. rubra* ssp. *fallax*, both of which have the
253 same ploidy level as *F. brevipila*. The range of variation in DNA content within each complex
254 suggest a complicated evolutionary history in addition to polyploidization [8].
255 While most crop plants are highly distinctive from their close relatives, *Festuca* is a
256 species-rich genus that contains species with highly similar morphology and different
257 ploidy level. Consequently, it is difficult for researchers to interpret species identity. In our
258 case, it is most difficult to distinguish between *F. rubra* ssp. *litoralis* cv. Shoreline and *F. rubra*

259 ssp. *fallax* cv. Treazure II as they had similar PI-A values based on flow cytometry. Thus, we
260 need different approaches to identify them. The presence and absence of rhizome formation
261 could be taken into consideration; for example, *F. rubra* ssp. *fallax* cv. Treazure II is a bunch
262 type turfgrass, while *F. rubra* ssp. *litoralis* cv. Shoreline forms short and slender rhizomes
263 [20]. This method may not be effective, however, because rhizome formation can be greatly
264 affected by environmental conditions [21, 22].

265 To further develop molecular tools to facilitate species identification, we carried out
266 chloroplast genome sequencing. We assembled the complete chloroplast genomes of five
267 low-input turfgrass fine fescues using Illumina sequencing. Overall, the chloroplast
268 genomes had high sequence and structure similarity among all five fine fescue taxa
269 sequenced, especially within each complex. All five chloroplast genomes share similar gene
270 content except for the three species in the *F. rubra* complex that have a pseudogene Acetyl-
271 coenzyme A carboxylase carboxyl transferase subunit (*accD*). The *accD* pseudogene is either
272 partially or completely absent in some monocots. Instead, a nuclear-encoded ACC enzyme
273 has been found to replace the plastic *accD* gene function in some angiosperm linage [23].
274 Indeed, even though the *accD* pseudogene is missing in the *F. brevipila* chloroplast genome,
275 the gene transcript was identified in a transcriptome sequencing dataset (unpublished data),
276 suggesting that this gene has been translocated to nucleus genome. Previous studies have
277 shown that broad-leaf fescues, *L. perenne, O. sativa,* and *H. vulgare* all had the pseudogene
278 *accD* gene, while it was absent in diploid *F. ovina, Z. mays, S. bicolor, T. aestivum,* and *B.
279 distachyon* [16]. Broad-leaf and fine-leaf fescues diverged around 9 Mya ago [24], which
280 raises an interesting question about the mechanisms of the relocation of *accD* among closely
281 related taxa in the *Festuca-Lolium* complex and even within fine fescue species.

282 In plants, chloroplast genomes are generally considered "single copy" and lack of
283 recombination due to maternal inheritance [25]. For this reason, chloroplast genomes are
284 convenient for developing genetic markers for distinguishing species and subspecies. We
285 have identified a number of repeat signatures that are unique to a single species or species
286 complex in fine fescue. For example, complement match is only identified in *F. ovina*
287 complex, and *F. rubra* complex has more reversed matches. We also identified two SSR
288 repeats unique to each of the two complexes. The first one consists of AAATT/AATTT repeat
289 units is unique to *F. rubra* ssp. *litoralis* and *F. rubra* ssp. *rubra*, and the second one consists of
290 ACCAT/ATGGT repeat units is unique to *F. brevipila* and *F. ovina*. In cases like the
291 identification of hexaploids *F. brevipila, F. rubra* ssp. *fallax,* and *Festuca rubra* ssp. *litoralis*, it
292 is critical to have these diagnostic repeats given all three taxa share similar PI-A values from
293 flow cytometry. Taxon-specific tandem repeat could also aid the SSR repeats for species
294 identification. Nucleotide diversity analysis suggested that several variable genome regions
295 exist among the five fine fescue taxa sequenced in this study. These variable regions
296 included previously known highly variable chloroplast regions such as *trnL-trnF* and *rpl32-
297 trnL* [13, 26]. These regions have undergone rapid nucleotide substitution and are
298 potentially informative molecular markers for characterization of fine fescue species.

299 Phylogeny inferred from DNA sequence is valuable for understanding the evolutionary
300 context of a species. The phylogenetic relationship of fine fescue using whole plastid
301 genome sequences agrees with previous classification based on genome size estimation and
302 morphology [8, 18]. The *F. ovina* complex includes *F. ovina* and *F. brevipila* and the *F. rubra*
303 complex includes *F. rubra* ssp. *rubra, F. rubra* ssp. *litoralis* and *F. rubra* ssp. *fallax*, with the two
304 rhizomatous subspecies (ssp. *rubra* and ssp. *literalis*) being sister to each other. Within the
305 *Festuca − Lolium* complex, fine fescues are monophyletic and together sister to a clade
306 consists of broad-leaved fescues and *Lolium*. In our analysis, *F. brevipila* (6x) is nested within

307 *F. ovina* and sister to the diploid *F. ovina*. It is likely that *F. brevipila* arose from the
308 hybridization between *F. ovina* (2x) and *F. ovina* (4x). Further research using nuclear loci is
309 needed to confirm this hypothesis.

310   The diversity of fine fescue provides valuable genetic diversity for breeding and
311 cultivar development. Breeding fine fescue cultivars for better disease resistance, heat
312 tolerance, and traffic tolerance could be achieved through screening wild accessions and by
313 introgressing desired alleles into elite cultivars. Some work has been done using *Festuca*
314 accessions in the USDA Germplasm Resources Information Network (GRIN)
315 (https://www.ars-grin.gov) to breed for improved forage production in fescue species [27].
316 To date, there are 229 *F. ovina* and 486 *F. rubra* accessions in the USDA GRIN. To integrate
317 this germplasm into breeding programs, plant breeders and other researchers need to
318 confirm the ploidy level using flow cytometry and further identify them using molecular
319 markers. Resources developed in this study could provide the tools to screen the germplasm
320 accessions and refine the species identification so breeders can efficiently use these materials
321 for breeding and genetics improvement of fine fescue species.

322 **4. Materials and Methods**

323 *Plant Material*

324   Seeds from the fine fescue cultivars seeds were obtained from the 2014 National
325 Turfgrass Evaluation Program (www.ntep.org, USA) and planted in the Plant Growth
326 Facility at the University of Minnesota, St. Paul campus under 16 hours daylight (25 ℃) and
327 8 hours dark (16 ℃) with weekly fertilization. Single genotypes of *F. brevipila* cv. Beacon, *F.*
328 *rubra* ssp. *litoralis* cv. Shoreline, *F. rubra* ssp. *rubra* cv. Navigator II, *F. rubra* ssp. *fallax* cv.
329 Treazure II, and *F. ovina* cv. Quatro were selected and used for chloroplast genome
330 sequencing.

331 *Flow Cytometry*

332   To determine the ploidy level of the cultivars used for sequencing and compare them
333 to previous work (2n=4x=28: *F. ovina*; 2n=6x=42: *F. rubra* ssp. *litoralis*, *F. rubra* ssp. *fallax*, and
334 *F. brevipila*; 2n=8x=56: *F. rubra* ssp. *rubra*), flow cytometry was carried out using *Lolium*
335 *perenne* cv. Artic Green (2n=2x=14) as the reference. Samples were prepared using CyStain
336 PI Absolute P (Sysmex, product number 05-5022). Briefly, to prepare the staining solution
337 for each sample, 12 μL propidium iodide (PI) was added to 12 mL of Cystain UV Precise P
338 staining buffer with 6 μL RNase A. To prepare plant tissue, a total size of 0.5 cm x 0.5 cm
339 leaf sample of the selected fine fescue was excised into small pieces using a razor blade in
340 500 μL CyStain UV Precise P extraction buffer and passed through a 50 μm size filter
341 (Sysmex, product number 04-004-2327). The staining solution was added to the flow-
342 through to stain nuclei in each sample. Samples were stored on ice before loading the flow
343 cytometer. Flow cytometry was carried out using the BD LSRII H4760 (LSRII) instrument
344 with PI laser detector using 480V with 2,000 events at the University of Minnesota Flow
345 Cytometry Resource (UCRF). Data was visualized and analyzed on BD FACSDiva 8.0.1
346 software. To estimate the genome size, *L. perenne* DNA (5.66 pg/2C) was used as standard
347 [28], USDA PI 230246 (2n=2x=14) was used as diploid fine fescue relative (unpublished data).
348 To infer fine fescues ploidy, estimation was done using equations (1) and (2) [29].

349

350 $\text{Sample 2C DNA Content} = \text{Standard 2C DNA Content (pg DNA)} \times \frac{\text{(Sample G1 Peak Mean)}}{\text{(Standard G1 Peak Mean)}}$ (1)

351

352 $$\text{Sample Ploidy} = \frac{2n \times \text{Sample pg/Nucleus}}{\text{Diploid Relative pg/Nucleus}} \qquad (2)$$

353 *DNA Extraction and Sequencing*

354 To extract DNA for chloroplast genome sequencing, 1 g of fresh leaves were collected
355 from each genotype and DNA was extracted using the Wizard Genomic DNA Purification
356 Kit (Promega, USA) following manufacturer instructions. DNA quality was examined on
357 0.8% agarose gel and quantified via PicoGreen (Thermo Fisher, Catalog number: P11496).
358 Sequencing libraries were constructed by NovoGene, Inc. (Davis, CA) using Nextera XT
359 DNA Library Preparation Kit (Illumina) and sequenced in 150 bp paired-end mode, using
360 the HiSeq X Ten platform (Illumina Inc., San Diego, CA, USA) with an average of 10 million
361 reads per sample. All reads used in this study were deposited in the NCBI Sequence Read
362 Archive (SRA) under BioProject PRJNA512126.

363

364 *Chloroplast Genome Assembly and Annotation*

365 Raw reads were trimmed of Illumina adaptor sequences using Trimmomatic (v. 0.32)
366 [30]. Chloroplast genomes were *de novo* assembled using NovoPlasty v. 2.0 [31]. Briefly, *rbcL*
367 gene sequence from diploid *F. ovina* (NCBI accession number: JX871940) was extracted and
368 used as the seed to initiate the assembly. NovoPlasty assembler configuration was set as
369 follows: *k-mer* size = 39; insert size = auto; insert range = 1.8; and insert range strict 1.3. Reads
370 with quality score above 25 were used to complete the guided assembly using *F. ovina* (NCBI
371 accession number: JX871940) as the reference. Assembled plastid genomes for each taxon
372 were manually corrected by inspecting the alignments of reads used in the assembly. The
373 assembled chloroplast genomes were deposited under BioProject PRJNA512126.

374 The assembled chloroplast genomes were annotated using the GeSeq pipeline [32] and
375 corrected using DOGMA online interface (https://dogma.ccbb.utexas.edu) [33]. BLAT
376 protein, tRNA, rRNA, and DNA search identity threshold was set at 80% in the GeSeq
377 pipeline using the default reference database with the generate codon-based alignments
378 option turned on. tRNAs were also predicted via tRNAscan-SE v2.0 and ARAGORN v 1.2.38
379 using the bacterial/plant chloroplast genetic code [34, 35]. The final annotation was
380 manually inspected and corrected using results from both pipelines. The circular chloroplast
381 map was drawn by the OrganellarGenomeDRAW tool (OGDRAW) [36].

382 *Nucleotide Polymorphism of Fine Fescue Species*

383 To identify genes with the most single nucleotide polymorphism, quality trimmed
384 sequencing reads of the five fine fescues were mapped to the diploid *Festuca ovina*
385 chloroplast genome (NCBI accession number: JX871940) using BWA v.0.7.17 [37]. SNPs and
386 short indels were identified using bcftools v.1.9 with the setting "mpileup -Ou" and called
387 via bcftools using the -mv function [38]. Raw SNPs were filtered using bcftools filter -s
388 option to filter out SNPs with low quality (Phred score cutoff 20, coverage cutoff 20). The
389 subsequent number of SNPs per gene and InDel number per gene was calculated using a
390 custom perl script SNP_vcf_from_gene_gff.pl (https://github.com/qiuxx221/fine-fescue-).

391 To identify simple sequence repeat (SSR) markers for plant identification,
392 MIcroSAtellite identification tool (MISA v 1.0) was used with a threshold of 10, 5, 4, 3, 3, 3
393 repeat units for mono-, di-, tri-, tetra-, penta-, and hexanucleotide SSRs, respectively [39].
394 The identification of repetitive sequences and structure of whole chloroplast genome was
395 done via PEPuter program online server (https://bibiserv.cebitec.uni-bielefeld.de/reputer)

396 [40]. Program configuration was set with minimal repeat size set as 20 bp and with sequence
397 identify above 90%. Data was visualized using ggplot2 in R (v 3.5.3). Finally, the sliding
398 window analysis was performed using DnaSP (v 5) with a window size of 600 bp, step size
399 200 bp to detected highly variable regions in the fine fescue chloroplast genome [41].

400 *Comparative Chloroplast Genomics Analysis*

401 To compare fine fescue species chloroplast genome sequence variations, the five
402 complete chloroplast genomes were aligned and visualized using mVISTA, an online suite
403 of computation tools with LAGAN mode [42, 43]. The diploid *Festuca ovina* (NCBI accession
404 number: JX871940) chloroplast genome and annotation were used as the template for the
405 alignment.

406 *Phylogenetic Analysis of Fine Fescues and Related Festuca species*

407 To construct the phylogenetic tree of the fine fescues using the whole chloroplast
408 genome sequence, chloroplast genomes of 8 species were downloaded from GenBank. Of
409 the 8 downloaded genomes, perennial ryegrass (*Lolium perenne*, AM777385), Italian ryegrass
410 (*Lolium multiflorum*, JX871942), diploid *Festuca ovina* (JX871940), tall fescue (*Festuca
411 arundiancea*, FJ466687), meadow fescue (*Festuca pratensis*, JX871941), and wood fescue
412 (*Festuca altissima*, JX871939) were within the *Festuca- Lolium* complex. Turfgrass species
413 outside of *Festuca-Lolium* complex including creeping bentgrass (*Agrostis stolonifera* L.,
414 EF115543) and *Cynodon dactylon* (KY024482.1) were used as an outgroup. All chloroplast
415 genomes were aligned using the MAFFT program (v 7) [44]; alignments were inspected and
416 manually adjusted. Maximum likelihood (ML) analyses was performed using the RAxML
417 program (v 8.2.12) under GTR+G model with 1,000 bootstrap [45]. The phylogenetic tree was
418 visualized using FigTree (v 1.4.3) (https://github.com/rambaut/figtree) [46].

419 **5. Conclusions**

420 Five newly-sequenced complete chloroplast genomes of fine fescue taxa were reported
421 in this study. Chloroplast genome structure and gene contents are both conserved, with the
422 presence and absence of *accD* pseudogene being the biggest structural variation between the
423 *F. ovina* and the *F. rubra* complexes. We identified SSR repeats and long sequence repeats of
424 fine fescues and discovered several unique repeats for marker development. The
425 phylogenetic constructions of fine fescue species in the *Festuca - Lolium* complex suggested
426 a robust and consistent relationship compared to the previous identification using flow
427 cytometry. This information provided a reference for future fine fescue taxa identification.

428 **Supplementary Materials**:
429 1. **Figure S1**: Flow cytometry nuclei population distribution of *L. perenne,* fine fescues, and diploid USDA
430   PI accession. G1 populations were gated in red, G2 population was only gated in *L. perenne* in green.
431 2. **Table S1**: SSR loci types and number distributions of fine fescue species predicted using MISA
432   program.
433 3. **Table S2**: Tandem repeat loci and repeat types predicted using PEPuter program in fine fescue species.
434 4. **Table S3**: SNPs number per gene distribution for fine fescue species. *rpoC2* gene has the most SNPs
435   (31) in *F. rubra* complex comparing to *F. ovina* species.
436 5. **Table S4**: InDel number per gene distribution for fine fescue species. *ndhA* gene had the most InDels
437   *F. rubra* species. *atpI* gene had the most InDesl in *F. ovina* species.
438 6. **Table S5**: Sliding window analysis using DnaSP at window size of 600 bp, step size 200 bp to detect
439   highly variable regions in the fine fescue chloroplast genome

440

Author Contributions: Y. Q. performed the experiments, analyzed the data, and wrote the manuscript; C. H. helped analyze data, wrote perl scripts; Y. Y. helped with phylogenetic analysis; E. W. secured funding for this project, supervised this research, provided suggestions, and comments. All authors contributed to the revision of the manuscript and approved the final version.

Conflicts of Interest: The authors declare no conflict of interest.

451
452

References

1. Clayton, W.D. and S.A. Renvoize, *Genera graminum. Grasses of the world.* Genera graminum. Grasses of the World., 1986. **13**.

2. Jenkin, T.J., *Fescue Species (Festuca L.).* In: Roemer, T. & W. Rudorf. Handbuch der Pflanzenzüchtung, 1959(4): p. 418–434.

3. Torrecilla, P. and P. Catalán, *Phylogeny of broad-leaved and fine-leaved Festuca lineages (Poaceae) based on nuclear ITS sequences.* Systematic Botany, 2002. **27**(2): p. 241-252.

4. Ruemmele, B., L. Brilman, and D. Huff, *Fine fescue germplasm diversity and vulnerability.* Crop Science, 1995. **35**(2): p. 313-316.

5. Beard, J.B., *Turfgrass: Science and Culture.* 1972.

6. Piper, C.V., *North American species of Festuca.* Vol. 10. 1906: US Government Printing Office.

7. Hubbard, C.E., *Grasses. A guide to their structure, identification, uses, and distribution in the British Isles.* Grasses. A guide to their structure, identification, uses, and distribution in the British Isles., 1968.

8. Huff, D.R. and A.J. Palazzo, *Fine fescue species determination by laser flow cytometry.* Crop Science, 1998. **38**(2): p. 445-450.

9. Casler, M.D., *Turfgrass biology, genetics, and breeding.* 2003: John Wiley & Sons.

10. Bonos, S.A. and D.R. Huff, *Cool-season grasses: Biology and breeding.* Turfgrass: Biology, use, and management, 2013(turfgrassbiolog): p. 591-660.

11. Hebert, P.D., et al., *Biological identifications through DNA barcodes.* Proceedings of the Royal Society of London. Series B: Biological Sciences, 2003. **270**(1512): p. 313-321.

12. Baldwin, B.G., et al., *The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny.* Annals of the Missouri Botanical Garden, 1995: p. 247-277.

13. Demesure, B., N. Sodzi, and R. Petit, *A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants.* Molecular Ecology, 1995. **4**(1): p. 129-134.

14. Diekmann, K., et al., *Complete chloroplast genome sequence of a major allogamous forage species, perennial ryegrass (Lolium perenne L.).* DNA Research, 2009. **16**(3): p. 165-176.

15. Cahoon, A.B., et al., *The complete chloroplast genome of tall fescue (Lolium arundinaceum; Poaceae) and comparison of whole plastomes from the family Poaceae.* American Journal of Botany, 2010. **97**(1): p. 49-58.

16. Hand, M.L., et al., *Plastome sequence determination and comparative analysis for members of the Lolium-Festuca grass species complex.* G3: Genes, Genomes, Genetics, 2013: p. g3. 112.005264.

17. Huang, Y.-Y., et al., *Complete chloroplast genome sequence of common bermudagrass (Cynodon dactylon (L.) Pers.) and comparative analysis within the family Poaceae.* PLoS One, 2017. **12**(6): p. e0179055.

18. Cheng, T., et al., *Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity.* Molecular Ecology Resources, 2016. **16**(1): p. 138-149.

19. Untergasser, A., et al., *Primer3—new capabilities and interfaces.* Nucleic Acids Research, 2012. **40**(15): p. e115-e115.

20. Meyer, W.A. and C.R. Funk, *Progress and Benefits to Humanity from Breeding Cool-Season Grasses for Turf 1.* Contributions from breeding forage and turf grasses, 1989(contributionsfr): p. 31-48.

21. Yang, M., et al., *Transcriptomic analysis of the regulation of rhizome formation in temperate and tropical lotus (Nelumbo nucifera).* Scientific Reports, 2015. **5**: p. 13059.

22. Ma, X. and B. Huang, *Gibberellin-stimulation of rhizome elongation and differential GA-responsive proteomic changes in two grass species.* Frontiers in Plant Science, 2016. **7**: p. 905.

495   23.   Rousseau-Gueutin, M., et al., *Potential functional replacement of the plastidic acetyl-CoA carboxylase*
496   *subunit (accD) gene by recent transfers to the nucleus in some angiosperm lineages.* Plant Physiology, 2013.
497   **161**(4): p. 1918-1929.

498   24.   Fjellheim, S., et al., *Phylogeographical history of the widespread meadow fescue (Festuca pratensis Huds.)*
499   *inferred from chloroplast DNA sequences.* Journal of Biogeography, 2006. **33**(8): p. 1470-1478.

500   25.   Ebert, D. and R. Peakall, *Chloroplast simple sequence repeats (cpSSRs): technical resources and*
501   *recommendations for expanding cpSSR discovery and applications to a wide array of plant species.* Molecular
502   Ecology Resources, 2009. **9**(3): p. 673-690.

503   26.   Dong, W., et al., *Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels*
504   *and for DNA barcoding.* PloS One, 2012. **7**(4): p. e35071.

505   27.   Robbins, M.D., J.E. Staub, and B.S. Bushman, *Development of fine-leaved Festuca grass populations*
506   *identifies genetic resources having improved forage production with potential for wildfire control in the western*
507   *United States.* Euphytica, 2016. **209**(2): p. 377-393.

508   28.   Arumuganathan, K., et al., *Nuclear DNA content of thirteen turfgrass species by flow cytometry.* Crop
509   Science, 1999. **39**(5): p. 1518-1521.

510   29.   Doležel, J., J. Greilhuber, and J. Suda, *Estimation of nuclear DNA content in plants using flow cytometry.*
511   Nature Protocols, 2007. **2**(9): p. 2233.

512   30.   Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.*
513   Bioinformatics, 2014. **30**(15): p. 2114-2120.

514   31.   Dierckxsens, N., P. Mardulyn, and G. Smits, *NOVOPlasty: de novo assembly of organelle genomes from*
515   *whole genome data.* Nucleic Acids Research, 2016. **45**(4): p. e18-e18.

516   32.   Tillich, M., et al., *GeSeq–versatile and accurate annotation of organelle genomes.* Nucleic Acids Research,
517   2017. **45**(W1): p. W6-W11.

518   33.   Wyman, S.K., R.K. Jansen, and J.L. Boore, *Automatic annotation of organellar genomes with DOGMA.*
519   Bioinformatics, 2004. **20**(17): p. 3252-3255.

520   34.   Lowe, T.M. and S.R. Eddy, *tRNAscan-SE: a program for improved detection of transfer RNA genes in*
521   *genomic sequence.* Nucleic Acids Research, 1997. **25**(5): p. 955-964.

522   35.   Laslett, D. and B. Canback, *ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide*
523   *sequences.* Nucleic Acids Research, 2004. **32**(1): p. 11-16.

524   36.   Lohse, M., O. Drechsel, and R. Bock, *OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation*
525   *of high-quality custom graphical maps of plastid and mitochondrial genomes.* Current Genetics, 2007. **52**(5-6):
526   p. 267-274.

527   37.   Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows–Wheeler transform.*
528   Bioinformatics, 2009. **25**(14): p. 1754-1760.

529   38.   Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.*
530   Bioinformatics, 2010. **26**(6): p. 841-842.

531   39.   Thiel, T., et al., *Exploiting EST databases for the development and characterization of gene-derived SSR-*
532   *markers in barley (Hordeum vulgare L.).* Theoretical and Applied Genetics, 2003. **106**(3): p. 411-422.

533   40.   Kurtz, S., et al., *REPuter: the manifold applications of repeat analysis on a genomic scale.* Nucleic Acids
534   Research, 2001. **29**(22): p. 4633-4642.

535   41.   Librado, P. and J. Rozas, *DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.*
536   Bioinformatics, 2009. **25**(11): p. 1451-1452.

537   42.   Brudno, M., et al., *LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic*
538   *DNA.* Genome Research, 2003. **13**(4): p. 721-731.
539   43.   Frazer, K.A., et al., *VISTA: computational tools for comparative genomics.* Nucleic Acids Research, 2004.
540   **32**(suppl_2): p. W273-W279.
541   44.   Katoh, K. and D.M. Standley, *MAFFT multiple sequence alignment software version 7: improvements in*
542   *performance and usability.* Molecular Biology and Evolution, 2013. **30**(4): p. 772-780.
543   45.   Stamatakis, A., *RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa*
544   *and mixed models.* Bioinformatics, 2006. **22**(21): p. 2688-2690.
545   46.   Rambaut, A., *FigTree v1. 4.* 2012.
546