Article type: Research

Genus-wide comparison reveals divergence and evolution of the four sections within the genus *Claviceps* are the result of varying mechanisms driving genome evolution and host range

Stephen A. Wyka[1], Stephen J. Mondo[1,2], Miao Liu[3], Jeremy Dettman[3], Vamsi Nalam[1], Kirk D. Broders[4*]

[1] Department of Agricultural Biology, Colorado State University, Fort Collins, CO 80523, USA.

[2] U.S. Department of Energy Joint Genome Institute, Berkeley, CA 94720, USA.

[3] Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa K1A 0C6, Canada.

[4] Smithsonian Tropical Research Institute, Apartado 0843-03092 Panamá, República de Panamá.

[*] Corresponding authors: Kirk D. Broders, email: BrodersK@si.edu

Abstract:

Background:

The genus *Claviceps* has been known for centuries as an economically important fungal genera for pharmacology and agricultural research. Despite this little is known about its evolution. Only recently have researchers begun to unravel the evolutionary history of the genus, with origins in South America and classification of four distinct sections through ecological, morphological, and metabolic features (*Claviceps* sects. *Citrinae, Paspalorum*, *Pusillae*, and *Claviceps*). The first three sections are additionally characterized by narrow host range, while sect. *Claviceps* is considered evolutionarily more successful and adaptable as it has the largest host range and biogeographical distribution. However, the reasons for this success and adaptability remain unclear.

Results:

Our study elucidates these factors by sequencing and annotating 50 *Claviceps* genomes, representing 21 species, for a comprehensive comparison of genome architecture and plasticity in relation to host range potential. Our results show the trajectory from specialized one-speed genomes (*C.* sects. *Citrinae* and *Paspalorum*) to adaptive two-speed genomes (*C.* sects. *Pusillae* and *Claviceps*) through co-localization of TEs around predicted effectors and loss of RIP resulting in unconstrained tandem gene duplication coinciding with increased host range potential and speciation.

Conclusions:

Alterations of genomic architecture and plasticity can substantially influence and shape the evolutionary trajectory of fungal pathogens and their adaptability. Furthermore, our study provides a large increase in available genomic resources to propel future studies of *Claviceps* in

pharmacology and agricultural research, as well as, research into deeper understanding of the evolution of adaptable plant pathogens.

Keywords: adaptive evolution, gene family expansion, two-speed genomes, fungal plant pathogens, RIP, ergot

Background:

The genus *Claviceps* (Hypocreales: Clavicipitaceae) contains biotrophic fungal species that share a specialized ovarian-specific non-systemic parasitic lifestyle with their grass hosts (Píchová *et al.* 2018). Species of *Claviceps* are most notably known for their production of toxic alkaloids which have been used in pharmacology but are also known for their great impact on global cereal crop production and livestock farming. Despite their ecological and agriculture importance, little is known about the evolution of these important fungal species in comparison to other grass pathogens such as members of the *Puccinia* (Cantu *et al.* 2013; Kiran *et al.* 2016, 2017), *Zymoseptoria* (Grandaubert *et al.* 2015, 2019; Estep *et al.* 2015; Poppe *et al.* 2015; Testa *et al.* 2015; Wu *et al.* 2017; Stukenbrock & Dutheil 2017), or *Fusarium* (Kvas *et al.* 2009; Ma *et al.* 2010; Rep & Kistler 2010; Watanabe *et al.* 2011; Sperchneider *et al.* 2015).

The genus *Claviceps* contains 59 species divided into four *Claviceps* sections; *C.* sects. *Claviceps*, *Pusillae*, *Citrinae*, and *Paspalorum* (Píchová *et al.* 2018). These four sections are distinguished based on combinations of phylogeny, morphology, secondary metabolite production, geographic distribution, and host spectrum (Píchová *et al.* 2018). Evolutionary differences between the sections have been identified through molecular clock dating and ancestral state reconstruction. In general, *C.* sects. *Citrina*, *Paspalorum*, and *Pusillae* are the

62 oldest divergent lineages occurring at 60.5, 57.2, and 50.3 Mya, respectively (Píchová *et al.*
63 2018). It was postulated that *C.* sects. *Citrina* and *Paspalorum* originated in South America,
64 while *C.* sect. *Pusillae* experienced speciation throughout the Eocene, Oligocene, and Miocene
65 as these species encountered newly emergent PACMAD warm-season grasses (subfamilies
66 Panicoideae, Aristidoideae, Chloridoideae, Micrairoideae, Arundinoideae, Danthonioideae)
67 when an ancestral strain was transferred from South America to Africa (Píchová *et al.* 2018). In
68 contrast, the crown node of *C.* sect. *Claviceps* is estimated at 20.4 Mya and is followed by a
69 radiation of the section corresponding to a host jump from ancestral sedges (Cyperaceae) to the
70 Pooideae tribe (BEP cool-season grasses; subfamilies Bambusoideae, Ehrhartoideae, Pooideae)
71 in North America (Bouchenak-Khelladi *et al.* 2010; Píchová *et al.* 2018).

72 With such a broad host range and cosmopolitan distribution of the pre-molecular concept
73 of *C. purpurea* (*sensu lato* (s.l.)), the potential of a species complex has been suggested.
74 Multigene studies have confirmed this in *C. purpurea* s.l. with evidence pointing towards
75 adaptation to ecological niches, and division into 8 different lineages (Pažoutová *et al.* 2000,
76 2002, 2015; Douhan *et al.* 2008; Shoukouhi *et al.* 2019; Liu, M *et al. Submitted*). However, these
77 cryptic species contain overlapping host ranges, geographic distributions, and alkaloid
78 production. In addition, individual gene phylogenies showed cases of disparate results
79 (Shoukouhi *et al.* 2019; Liu, M *et al. Submitted*) and estimates of genetic differences between
80 cryptic species show relatively low average number of nucleotide substitutions per site ranging
81 from 0.006 to 0.025 (Pažoutová *et al.* 2015).

82 *Claviceps* sect. *Claviceps* shows the largest host range with species having the ability to
83 infect grasses across 13 subfamilies of BEP grasses, with some species able to infect PACMAD
84 grasses. Notable examples include *C. humidiphila* which has been reported on grasses across

four Pooidea subfamilies (BEP) and Arundineae (PACMAD), or *C. purpurea sensu stricto* (s.s.) which has been reported on up to 400 different species of Pooidea grasses (Alderman *et al.* 2004, Píchová *et al.* 2018) across six subfamilies, while retaining the ability to infect sedges (Cyperaceae) (Jungehülsing and Tudzynski 1997). In contrast *C.* sect. *Pusillae* were not as successful in adapting to multiple subfamilies of PACMAD grasses. In general, *C.* sect. *Pusillae* specialized to the subfamilies Paniceae and Andropogoneae, with few species infecting subfamilies Paspaleae, Cynodonteae, or Eragrostideae (Píchová *et al.* 2018). This is in further contrast to *C.* sects. *Citrinae* and *Paspalorum* which show specialization to subfamily Paspaleae and subfamily Cynodonteae, respectively (Píchová *et al.* 2018). While these latter two sections are monotypic, *C. paspali* (*C.* sect. *Paspalorum*) is a globally distributed pathogen but has not been reported on any other grasses (Píchová *et al.* 2018).

The reason for the evolutionary success (increased host range) of *C. purpurea* s.s. and other species in *C.* sect. *Claviceps* is still unknown. Evolution of secondary metabolites (i.e. ergot alkaloids, ergochromes, etc) is a likely mechanism as these can theoretically increase fitness by altering infection potential, stress tolerance, or antimicrobial resistance (Avalos and Carmen Limon 2015; Píchová *et al.* 2018). Ergot alkaloids have also been shown to provide enhanced host fitness through protection against herbivores (Wäli et al 2013) and other conditional mutualistic effects on host plant fitness, such as neutral to positive effects on seed set and weight (Raybould *et al.* 1998; Fisher *et al.* 2007; Wäli et al 2013). However, it is unclear if these fitness advantages are due to ergot alkaloids, fungal phytohormone production (Hinsch *et al.* 2015, 2016; Mlynarčíková 2015), or some other functional response. Furthermore, *C.* sect. *Claviceps* species share similar secondary metabolites products, albeit with some abundance variations (Píchová *et al.* 2018; Flieger *et al.* 2019), but have been reported on a narrower range

108 of hosts than *C. purpurea* s.s, indicating that secondary metabolites might not be solely
109 responsible for the divergence of this section nor the decrease in host range.
110 There are many genomic mechanisms that can govern adaptation and diversification of
111 fungal species. Recently, Moller & Stukenbrock (2017) provided an extensive review
112 summarizing many key examples of molecular factors that facilitate evolutionary processes in
113 fungal plant pathogens. In brief, fungal adaptation is mediated by changes in genome
114 architecture and plasticity such as genome size, transposable element (TE) content, localization
115 of TEs to specific genes, genome compartmentalization, and presence/absence polymorphism of
116 virulence factors. It is becoming increasingly evident that variations in these factors also give
117 rise to one-speed and two-speed genomes (Dong *et al.* 2015). However, further research in this
118 field has identified that these terms are not as clear cut and often blurred, especially as new
119 fungal genomes are analyzed (Frantzeskakis *et al.* 2019). An additional important factor in how
120 fungal genomes evolve is the presence/absence of repeat-induced point (RIP) mutation (Galagan
121 *et al.* 2003, 2004; Raffaele & Kamoun 2012; Moller & Stukenbrock 2019). RIP is known to
122 identify repeat/duplicated sequences within a genome, greater than ~400 bp with ≥80% identity,
123 and introduces C:G to T:A mutations to effectively silence these regions (Galagan *et al.* 2003,
124 2004; Urguhart *et al.* 2018). While RIP is considered a mechanism for the control and spread of
125 transposable elements, it has a global effect as it controls gene duplication on a genome-wide
126 scale (Galagan *et al.* 2003, 2004; Urguhart *et al.* 2018). Research into the genomic plasticity of
127 *Claviceps* species is very limited with Schardl *et al.* (2013) presenting the only current
128 information on the topic. They stated that RIP is present across all studied members of the
129 Claviciptaceae (i.e. *Claviceps, Epichloe, Aciculosporium*), except in *C. purpurea* s.s., and that
130 secondary metabolite clusters tended to be embedded in TE-rich regions. However, this

131 observation of co-localized TEs was more evident in *Epichloe* species than *Claviceps* species
132 (Schardl *et al.* 2013). Since their study only looked at several secondary metabolite clusters and
133 three *Claviceps* species, it questions whether this trend is consistent across the genome and
134 across multiple *Claviceps* species.

135 Píchová *et al.* (2018) has provided an excellent first account detailing the evolutionary
136 history of the *Claviceps* genus using combinations of morphology, multigene analysis, and
137 alkaloid production. Since previous genomic research of *Claviceps* have generally been restricted
138 to secondary metabolite production of three reference genomes (*C. purpurea* s.s, *C. fusiformis*,
139 and *C. paspali*) there is a large knowledge gap in the overall understandings of *Claviceps*
140 genomes and their evolution. Our research looks to further the understanding of *Claviceps*
141 evolution by characterizing the genomic plasticity and architecture in relation to adaptive host
142 potential. To this end we, sequenced and annotated 50 *Claviceps* genomes, representing 19
143 species, for a comprehensive comparison of the genus. These 50 newly annotated genomes will
144 subsequently propel further research into secondary metabolites of the genus and perhaps help
145 explain why *C. purpurea* s.s. has eluded all attempts of finding genetic resistance in crops.
146

147 **Results:**

148 *Genome Assembly and Annotation*:

149 To provide a comprehensive view of variability across *Claviceps*, we sequenced and
150 annotated 50 genomes (19 *Claviceps* spp.). Including *C. citrina* the single species of *C.* sect.
151 *Citrinae*, six species belonging to *C.* sect. *Pusillae*, and 44 genomes (12 species) belonging to *C.*
152 sect. *Claviceps*, of which 23 genomes belong to *C. purpurea* s.s. (Table 1; Additional file 1:
153 Table S1). The assemblies and annotations were of comparable quality to the reference strains

154 (Table 1). A more detailed representation of the assembly and annotation statistics can be seen in
155 Table 1 and Additional file 1: Figure S1.
156 Overall, species of *C.* sect. *Claviceps* had better assemblies and annotations than species
157 of other sections regarding contig numbers, N50's, and BUSCO completeness scores (Table 1).
158 Nearly all species of *C.* sect. *Claviceps* showed higher BUSCO scores than the references, while
159 species of *C.* sects. *Pusillae* and *Citrinae* generally showed lower scores, likely due to their
160 higher repeat content (average 40.7% ± 6.6%; Table 1). Exceptions to the low BUSCO scores
161 were *C. digitariae* and *C. maximensis* (*C.* sect. *Pusillae*), which had lower repeat content 23.3%
162 and 23.1%, respectively, than the rest of the species in *C.* sect. *Pusillae* (Table 1).
163
164 *Phylogenomics and genome fluidity:*
165 The maximum likelihood (ML) tree, based on 2,002 single copy orthologs showed
166 statistical support for four sections of *Claviceps* with a near concordant topology to the Bayesian
167 five-gene phylogeny in Píchová *et al.* (2018). Notable exceptions were the placement of *C.*
168 *paspali* (*C.* sect. *Paspalorum*) which is more closely related to *C. citrina* (*C.* sect. *Citrinae*)
169 instead of *C.* sect. *Claviceps* and *C. pusilla* was more closely related to *C. fusiformis* instead of
170 *C. maximensis* (Fig. 1). We also found that *C.* sect. *Claviceps* diverged from a common ancestor
171 with *C.* sect. *Pusillae* as opposed to *C.* sect. *Paspalorum*. Our results provide support for the
172 deeply divergent lineages of *C.* sects. *Pusillae*, *Paspalorum*, and *Citrinae* with a long divergent
173 branch resulting in *C.* sect. *Claviceps* (Fig. 1). The discrepancies between our results and
174 Píchová *et al.* (2018) could be caused by deficient taxon sampling (Nabhan and Sarkar *et al.*
175 2012) or deficient gene sampling (Posada *et al.* 2016). Píchová *et al.* (2018) did observe these
176 discrepancies (placement of *C. pusilla*) in their maximum parsimony (MP) and (placement of *C.*

177 *paspali*) in their ML topologies, suggesting that algorithm bias could also be a factor. An
178 additional Neighbor-Joining (NJ) and MP analysis were concordant with our ML analysis,
179 suggesting that *C.* sect. *Paspalorum* is more closely related to *C.* sect. *Citrina* and that *C.* sect.
180 *Claviceps* shared a common ancestor with *C.* sect. *Pusillae* (Additional file 1: Figure S2, S3). In
181 addition, our analyses of genomic evolution (this study) largely agree with the divergence of the
182 sections represented in our ML analysis (Fig. 1).

183 Within *C.* sect. *Claviceps* we observed support for the separation of recently described
184 species from *C. purpurea* s.s. (Fig.1) (Pažoutová *et al.* 2000, 2002, 2015; Douhan *et al.* 2008;
185 van der Linde *et al.* 2016; Shoukouhi *et al.* 2019; Liu, M *et al. Submitted*). Overall, *C.* sect.
186 *Claviceps* showed an abundance of short divergent branches separating the species. To further
187 elucidate trends of divergence within the genus we tested genomic fluidity using gene families
188 with and without paralogs (Kislyuk *et al.* 2011). Genomic fluidity estimates the dissimilarity
189 between genomes by using ratios of unique gene families to the sum of gene families in pairs of
190 genomes averaged over randomly chosen genome pairs from within a group on *N* genomes. For
191 example, a fluidity value of 0.05 indicates that genomes in a group have 5% unique gene families
192 and share 95% of their gene families (Kislyuk *et al.* 2011).

193 *Claviceps* sect. *Claviceps*, which is composed of 12 different species, showed a relatively
194 small genomic fluidity (0.0619 ± 0.0019) with little variation, indicating that the pairwise gene
195 family dissimilarity between randomly sampled species is quite low. Interestingly, the amount of
196 variation between 12 different *Claviceps* species is similar to the variation between 24 *C.*
197 *purpurea* s.s. strains, however, there was a significant difference between the fluidities
198 (Additional file 1: Table S3; $P < 0.0001$). In comparison, the fluidity of *C.* sect. *Pusillae* (0.126 ±
199 0.014; $P < 0.0001$) was two times greater than the fluidity of *C.* sect. *Claviceps* and exhibited

200 greater variation, indicating greater dissimilarities in gene families between randomly sampled
201 species of *C.* sect. *Pusillae*. Lastly, at the genus level, *Claviceps* had a significantly lower fluidity
202 than other genera, i.e. *Epichloe*, *Trichoderma*, and *Fusarium* (Fig. 2; Additional file 1: Table S3;
203 $P < 0.05$). While genus-wide comparisons can be confounded by differences between genera (i.e.
204 genera age, species diversity, gene content), this could suggest an overall trend of more
205 conserved gene families between species in the genus *Claviceps* than the other genera.
206 Overall, our ML phylogeny (Fig. 1) and genome fluidity analysis (Fig. 2) indicate a large
207 evolutionary divergence separating *C.* sect. *Claviceps*. Species within *C.* sect. *Claviceps* and
208 *Pusillae* show signs of different evolutionary strategies governing their divergence. These results
209 could suggest that differences between species of *C.* sect. *Pusillae* are attributed to greater
210 amounts of amino acid substitutions and potential presence/absence polymorphism of gene
211 families (Fig. 1, 2). Due to the short branch lengths of our ML analysis (Fig. 1) and low genomic
212 fluidity (Fig. 2), genetic mutations and gene family dissimilarity do not appear to be factors
213 driving the diversification of species within *C.* sect. *Claviceps.* Our subsequent analyses of the
214 genomic architecture of all *Claviceps* species may facilitate determining factors that could be
215 associated with the evolutionary divergence of *C.* sect. *Claviceps* and those driving the cryptic
216 speciation.
217
218 *Transposable element divergences and locations:*
219 Transposable element (TE) divergence landscapes revealed an overrepresentation of LTR
220 elements in *C.* sects. *Pusillae*, *Citrinae*, and *Paspalorum*. All three sections showed a similar
221 large peak of LTRs with divergences clustered between 5–10% (Fig. 3; Additional file 1: Figure
222 S4), indicating a relatively recent expansion of TEs. The landscapes of *C.* sects. *Pusillae*,

223 *Citrinae*, and *Paspalorum* are in striking contrast to species of *C.* sect. *Claviceps* which showed
224 varying, but generally similar, abundances of LTR, DNA, LINE, SINE, and RC elements.
225 Although, unclassified TEs often constituted the majority of TE content in this section. Species
226 of *C.* sect. *Claviceps* showed broader divergence peaks that clustered between 5–30% divergence
227 (Fig. 3; Additional file 1: Figure S4). The TE landscape of *C. cyperi* showed a more striking
228 divergence peak clustering between 5–10% that more closely resembled the landscapes of
229 species in *C.* sects. *Pusillae*, *Paspalorum*, and *Citrinae*. However, the content of the TE peak in
230 *C. cyperi* largely contained DNA, LINE, and unclassified TEs as opposed to LTR's (Additional
231 file 1: Figure S4). *Claviceps cyperi* is considered an ancestral species of *C.* sect. *Claviceps*
232 (Píchová *et al.* 2018) (Fig. 1), therefore, the TE landscape observed in this species could provide
233 hints as to how TEs changed during the divergence of *C.* sect. *Claviceps* from *C.* sect. *Pusillae*.
234 This suggests that the common ancestor of these sections could have experienced a drastic
235 genomic reduction that effectively removed large amounts of repetitive content. This is
236 supported through the low genome size of *C. cyperi* (26.6 Mb), the lowest of all *C.* sect.
237 *Claviceps* species (Table 1), which reduced ~67% from the average genome size of *C.* sect.
238 *Pusillae* species (39.9 ± 6.6 Mb).
239 To identify where genes are located in relation to TEs, we calculated the average distance
240 (kbp) of each gene to the closest TE fragment. This analysis was performed for predicted
241 effectors, secreted (non-effector) genes, secondary metabolite (non-secreted) genes, and all other
242 genes. Results show that non-effector secreted genes and predicted effectors of *C.* sects.
243 *Claviceps* and *Pusillae* species were significantly closer to TEs (Fig. 4; $P < 0.05$), suggesting
244 that these genes could be located in more repeat-rich regions of the genome. In both *C.* sects.
245 *Claviceps* and *Pusillae* non-secreted secondary metabolites genes were located farther away

246 from TEs (Fig. 4; $P < 0.05$), i.e. repeat-poor regions of the genome. These trends hold true for
247 individual strains, with a notable exception of *C. pusilla* (*C.* sect. *Pusillae*) showing no
248 significant differences in the proximity of TEs to specific gene types (Additional file 1: Figure
249 S5; $P > 0.05$). Variation existed in whether particular strains had significant differences between
250 all other genes and non-effector secreted genes and non-secreted secondary metabolite genes but
251 all species in *C.* sects. *Claviceps* and *Pusillae* (aside from *C. pusilla*) had predicted effectors
252 located significantly closer to TEs (Additional file 1: Figure S5; $P < 0.05$). No significant
253 differences in the proximity of TEs to specific gene types were observed in *C.* sects. *Citrinae* and
254 *Paspalorum* (Fig. 4; $P > 0.05$), suggesting that TE's are more randomly distributed throughout
255 these genomes.

256 Overall, these results indicate a difference in evolutionary strategies between *C.* sects.
257 *Citrinae* and *Paspalorum* and *C.* sects. *Claviceps* and *Pusillae*. Given that there is an over-
258 proliferation of TEs which are randomly distributed throughout the genome, *C.* sects. *Citrinae*
259 and *Paspalorum* appear to fit the concept of a one-speed genome model (Dong *et al.* 2015;
260 Frantzeskakis *et al.* 2019). This could help explain the specialized nature of these sections (i.e.
261 lack of section lineages and restricted host range to one grass subfamily), as similar patterns of
262 large genome size, abundant TE content, and equal distribution of TEs has been observed in the
263 specialized barley pathogen *Blumeria graminis* f.sp. *hordei* (Frantzeskakis *et al.* 2018). In
264 contrast, while TE abundance differs between *C.* sects. *Claviceps* and *Pusillae*, these two
265 sections show TEs co-localized with predicted effectors, suggesting aspects of a more rapid
266 "adaptive" evolution more fitting of a two-speed model (Dong *et al.* 2015; Frantzeskakis *et al.*
267 2019). The observed switch in genome architecture between these four sections could be the
268 factor responsible for the divergence of *C.* sects. *Claviceps* and *Pusillae* from *C.* sects. *Citrinae*

269 and *Paspalorum* (Fig. 1). Additionally, this could suggest that the shift from randomly
270 distributed TEs to co-localized TEs around effectors allowed *C.* sects. *Claviceps* and *Pusillae* to
271 develop a broader host range (13 and 5 grass subfamilies, respectively) and might have
272 facilitated the increased number of speciation events in these sections (Raffaele & Kamoun
273 2012; Stuckenbrock 2013; Moller & Stukenbrock 2017).
274
275 *Genome compartmentalization*:
276 To further examine genome architecture, we analyzed intergenic regions to examine
277 evidence of genome compartmentalization by plotting the 5` and 3` intergenic lengths of all
278 genes within each genome. Results showed that all 53 *Claviceps* strains exhibited a one-
279 compartment genome (lack of large-scale compartmentalization). This provides support for the
280 one-speed model for *C.* sects. *Citrinae* and *Paspalorum* but contrasts the two-speed model
281 predicted from the TE analysis for *C.* sects. *Claviceps* and *Pusillae*. Although, these results could
282 indicate the potential evolution towards a two-compartment genome as species in *C.* sects.
283 *Claviceps* and *Pusillae* showed slightly larger intergenic regions than *C.* sects. *Citrinae* and
284 *Paspalorum* (Fig. 5; Additional file 1: Figure S6).
285 To further clarify evolutionary strategies within the four sections we determined if
286 different genes types tended to reside in different regions of the genome. Results showed that all
287 species of *C.* sect. *Claviceps* had significantly larger intergenic flanking regions for predicted
288 effector genes, indicating that they reside in more gene-sparse regions of the genome (Fig. 5;
289 Additional file 1: Figure S6). This is supported by the proximity of TEs to predicted effectors
290 (Fig. 4). Only *C. digitariae* and *C. lovelessi* (Additional file 1: Figure S6; $P < 0.01$, $P = 0.024$,
291 respectively) of *C.* sect. *Pusillae* had predicted effector genes with significantly larger intergenic

292 regions. All secreted genes for each species showed similar results and were often significant
293 (Fig. 5; Additional file 1: Figure S6). In contrast, secondary metabolite genes exhibited a
294 widespread distribution of intergenic lengths that were not significantly different than non-
295 metabolite genes in all 53 *Claviceps* strains (Fig. 5; Additional file 1: Figure S6).

296 It appears that species in *C.* sects. *Claviceps* and *Pusillae* neither align with the concept
297 of the two-speed model as they do not exhibit large-scale genomic compartmentalization, nor do
298 they fit into the one-speed model as predicted effectors are in close proximity to TEs and exist in
299 gene-sparse regions of the genome (Fig. 4, 5; Additional file 1: Figure S5). Frantzeskakis *et al.*
300 (2019) has stated that borders between these models are often blurred as many intermediate and
301 ambiguous models have been shown to exist in other fungi. RIP activity may help explain the
302 plasticity and evolutionary strategy of these gene-sparse regions rich in TEs and effectors. In the
303 presence of RIP these gene-sparse regions could harbor increased levels of local mutagenesis
304 through RIP leakage, while in the absence of RIP these regions could be hot spots for
305 duplication, deletion, and recombination (Galagan *et al.* 2003, 2004; Raffaele & Kamoun 2012;
306 Dong *et al.* 2015; Faino *et al.* 2016; Moller & Stukenbrock 2017; Frantzeskakis *et al.* 2018,
307 2019).

308

309 *RIP analysis:*

310 To test for RIP activity, we assessed the bi-directional similarity of genes against the
311 second closet BLASTp match within each isolate's own genome (Galagan *et al.* 2003; Urguhart
312 *et al.* 2018), supported by a BLASTp analysis against the *rid-1* RIP gene of *Neurospora crass*.
313 Results showed that *C.* sects. *Pusillae*, *Citrinae*, and *Paspalorum* had homologs of *rid-1* and
314 fewer genes with close identity (≥80%), potentially indicating past or current activity of RIP

315 (Fig. 6; Table 2; Additional file 2: Table S4). This was surprising given the abundance of TEs
316 within genomes of these sections (Table 1; Fig. 3; Additional file 1: Figure S4) as RIP should be
317 working to silence and inactive these TEs. While we did not directly test whether the TEs within
318 our genomes are active/inactive, the peaks in the divergence landscapes of *C.* sects. *Pusillae*,
319 *Citrinae*, and *Paspalorum* (5-10%; Fig. 3) suggest recent activity of TEs (Frantzeskakis *et al.*
320 2018). If recent TE activity is occurring within these species it could suggest that these LTR
321 elements have found a way to avoid RIP or indicate that these species harbor a less potent
322 version of RIP (Kachroo *et al.* 1994; Nakayashiki *et al.* 1999; Dayiere *et al.* 2001; Graia *et al.*
323 2001; Ikeda *et al.* 2002; Chalyet *et al.* 2003; Kito *et al.* 2003). Nonetheless, these results could
324 indicate that species within *C.* sect. *Pusillae* exhibit increased levels of local mutagenesis,
325 potentially through RIP leakage, in gene-sparse regions rich in TEs and effectors, although,
326 further studies would need to confirm this.

327 In comparison, species in *C.* sect. *Claviceps* lack *rid-1* homologs and showed larger
328 amounts of gene similarity, with some variation occurring across species and strains, suggesting
329 that RIP is inactive (Fig. 6; Table 2; Additional file 2: Table S4), supporting Schardl *et al.*
330 (2013). Gene pairs sharing a ≥80% identity to each other were often located near each other. On
331 average 27.02% ± 5.91% of the pairs were separated by 5 or less genes, while 15.95% ± 3.50%
332 of the pairs were located next to each other, indicating signs of tandem gene duplication within
333 the section (Table 2). Aside from *C. cyperi*, which showed a large reduction in the proportions of
334 highly similar tandem genes (7.77% and 5.7%, respectively), clear variations in these proportions
335 between species of *C.* sect. *Claviceps* did not exist as proportions appeared to vary more between
336 strains.

337 Overall, these results suggest that the divergence of *C.* sect. *Claviceps* from *C.* sect.
338 *Pusillae* is likely due to a loss of RIP, subsequently leading to unconstrained gene duplication
339 (Fig. 1, 6). It could be suggested that this change further increased the adaptability of *C.* sect.
340 *Claviceps* allowing its members to have a larger host range of 13 grass subfamilies versus the 5
341 subfamilies affected by *C.* sect. *Pusillae.* However, factors governing the divergence of species
342 within *C.* sect. *Claviceps* are still unclear. A BLASTp analysis against 56 important *Claviceps*
343 genes including known virulence factors, phytohormone biosynthetic genes, and secondary
344 metabolite genes did not indicate any patterns of presence/absence polymorphism which could
345 help elucidate the divergence factors within *C.* sect. *Claviceps* (Additional file 2: Table S4).
346 Similar patterns of low TE content (~7%) with TEs co-localized to tandem virulence genes
347 clusters have been reported from smut fungi *Ustilago maydis*, *Sporisorium scitamineum*, and *S.*
348 *reilianum* (Kämper *et al.* 2006; Schirawski *et al.* 2010; Dutheil *et al.* 2016). The virulence
349 clusters in these biotrophic smut fungi were shown to be associated with rapid speciation and
350 incomplete lineage sorting (Dutheil *et al.* 2016). Similar disparate lineage sorting was observed
351 in the classification of recent cryptic species in *C.* sect. *Claviceps* (Shoukouhi *et al.* 2019; Liu, M
352 *et al. Submitted*), suggesting similar evolutionary patterns could be driving the diversification of
353 *C.* sect. *Claviceps*. To better understand the types of genes with observed duplication we
354 investigated gene family expansion.
355
356 *Gene family expansion:*
357 Our OrthoFinder analysis revealed evidence of gene family expansion within *C.* sect.
358 *Claviceps* as species contained more genes per gene family than species of the other three
359 sections (Additional file 1: Figure S7). To identify the types of gene families that were showing

360 expansion we filtered our families by two criteria; 1) at least 1 species in a gene family had 2 or
361 more genes, 2) there was a significant difference in the mean number of genes per family
362 between all 44 strains in *C.* sect. *Claviceps* and the 9 species from *C.* sects. *Pusillae*, *Citrinae*,
363 and *Paspalorum* ($\alpha \leq 0.01$, Welch's test).

364 There was extensive expansion (families with observations of ≥10 genes of a strain)
365 present in many predicted effector families, families containing genes with conserved domains,
366 unclassified, and several secreted (non-effector) gene families (Fig. 7; Additional file 1: Figure
367 S8, S9). Transmembrane gene families also showed evidence of expansion with several families
368 containing species with 5-10 genes per family, while metabolite (non-secreted) families showed
369 the lowest amount of expansion (Fig. 7). Overall, *C.* sect. *Claviceps* showed expansion in a
370 greater number of gene families than *C.* sect. *Pusillae*, *Citrinae*, and *Paspalorum* in all family
371 categories except transmembrane families (Additional file 1: Figure S10). Gene families with an
372 average ≥5 genes per strain within *C.* sect. *Claviceps* contained a variety of functional proteins,
373 with generally more of the gene families containing proteins with protein/serine/tyrosine kinase
374 domains (Additional file 3: Table S5). Additional details can be obtained from Additional file 4:
375 Table S6 (ordered gene families corresponding to heatmaps; Fig. 7 and Additional file 1: Figure
376 S8, S9) and Additional file 5-6: Table S7-1,2 (gene family identification and functional
377 annotation of all proteins).

378 Within *C.* sect. *Claviceps* patterns of gene counts per gene family appeared to breakdown
379 and contain variations in the number of genes per family and some presence/absence of families
380 occurring between strains and species. Notably, *C. cyperi* (CC265) showed the lowest amount of
381 expansion, across all taxa, in comparison to other species of *C.* sect. *Claviceps*. In addition, *C.*
382 *spartinae* (CCC535), *C. capensis* (CCC1504), *C. monticola* (CCC1483), *C. pazoutovae*

383 (CCC1485), *C. occidentalis* (LM77,78,84), and *C. quebecensis* (LM458, Clav32,50) also showed
384 lower expansion (Fig. 7, Additional file 1: Figure S8, S9). However, no defined patterns were
385 observed linking the variation in gene family expansions with the literature stated host range of
386 different species within *C.* sect. *Claviceps*. While this disconnect could suggest gene family
387 expansions do not play a role in host ranges, it could also be suggested that a lack of knowledge
388 on true host ranges could be the reason. With the detection of cryptic speciation within *C.*
389 *purpurea* s.l. it puts into questions whether previously identified host ranges of *C. purpurea* s.s.
390 are correct. It is possible that host range may be more restricted in some of the new cryptic
391 species or conversely that the host range of *C. purpurea* s.s. is not as broad. This would still
392 indicate that *C.* sect. *Claviceps* has the largest host range of the four *Claviceps* sections, but it
393 could dispute the suggested 400+ grass species host range of *C. purpurea* s.s. (Alderman *et al.*
394 2004, Píchová *et al.* 2018). Extensive cross inoculation trials would be needed to more
395 confidently determine true host ranges for each of the cryptic species in *C.* sect. *Claviceps*. This
396 could then be used to re-evaluate whether there are correlations of genes or gene families to
397 defined host ranges for species of *C.* sect. *Claviceps*.
398
399 **Discussion:**
400 Our comparative study of 50 newly annotated genomes from four sections of *Claviceps*
401 has provided us with enhanced understanding of evolution in the genus through knowledge of
402 factors that could be contributing to its diversification. Our results have revealed that despite
403 having nearly identical life-strategies, these closely related species have substantially altered
404 genomic architecture and plasticity, which may drive genome adaptation. One key difference we
405 observe is a shift from characteristic one-speed genomes in narrow host-range *Claviceps* species

406 (*C.* sects. *Citrinae* and *Paspalorum*) to two-speed genomes in broader host-range lineages of *C.*
407 sects. *Pusillae* and *Claviceps* (Fig. 1).

408 The ancestral species of the genus, *C. citrina* (*C.* sect. *Citrinae*) and *C. paspali* (*C.* sect.
409 *Paspalorum*), are characterized by an over-proliferation of TEs which are more evenly dispersed
410 throughout the genome (Fig. 4). Coupled with a lack of large-scale genome
411 compartmentalization (Fig. 5), these two species can be considered to fit the concept of a one-
412 speed genome which are often considered to be less adaptable and potentially more prone to
413 being purged from the biota (Dong *et al.* 2015; Frantzeskakis *et al.* 2019). Rapid adaptive
414 evolution has been noted in one-speed genomes, primarily through copy-number variation and/or
415 heterozygosity of effector loci, such as that observed in *Blumeria graminis* f.sp. *hordei* (Dong *et*
416 *al.* 2015; Frantzeskakis *et al.* 2018, 2019). We found that gene duplication is not occurring in
417 these species, likely due to the presence of RIP. Heterozygosity of effector loci or aspects of RIP
418 leakage could be the mechanisms for evolution in *C. citrina* and *C. paspali*, however, further
419 research is needed to clarify this. It should be noted that since their divergence 60.5 and 57.2
420 Mya (Píchová *et al.* 2018) it appears that these sections have remained monotypic. It cannot be
421 ruled out that lack of additional lineages within these sections could be due to limited records of
422 *Claviceps* species in South America, where the genus is thought to have originated (Píchová *et*
423 *al.* 2018). Further research into South American populations of *Claviceps* will provide significant
424 insight into the evolution of these two sections.

425 Similarly, *C.* sect. *Pusillae* exhibits an over-proliferation of TEs. However, as this section
426 diverged, from *C.* sects. *Citrinae* and *Paspalorum*, the genomic architecture evolved such that
427 TEs began to reside closer to effector genes (Fig. 4). This proximity of TEs to effectors persisted
428 in *C.* sect. *Pusillae* species (except *C. pusilla*; Additional file 1: Figure S5) and *C.* sect.

429 *Claviceps* species allowing movement of predicted effectors to more gene-sparse regions of the
430 genome (Fig. 5, Additional file 1: Figure S6). This could indicate the evolutionary beginnings of
431 a more adaptable two-speed genome potentially caused by contact with new hosts and different
432 climates as ancestral lineages migrated out of South America (Píchová *et al.* 2018), or that the
433 evolution of a two-speed genome provided an advantage in adapting to new hosts or
434 environments. While further research would be needed to clarify this, both *C.* sects. *Pusillae* and
435 *Claviceps* have larger host ranges (5 subfamilies and 13 subfamilies, respectively) and increased
436 levels of speciation (Píchová *et al.* 2018). Píchová *et al.* (2018) postulated that *C.* sect. *Pusillae*
437 was transferred to Africa (ca 50.3 Mya), while *C.* sect. *Claviceps* originated in North America
438 (ca. 20.7 Mya). It is likely that the common ancestor shared between these sections (Fig. 1) had
439 strains that were transferred to Africa, likely due to insect vectors via transatlantic long-distance
440 dispersal (LDD) (Píchová *et al.* 2018). The strains that remained, in South America, likely
441 persisted but appeared to have lacked speciation for roughly 30 Mya (Píchová *et al.* 2018),
442 despite having evolved a more adaptable two-speed genome (Fig. 4, 5). Limited sampling
443 records could again be a factor contributing to this lack of speciation during this 30 My period,
444 but it could also be suggested that the ancestral species of *C.* sects. *Claviceps* did not diverge due
445 to a lack of diversification of host species (Píchová *et al.* 2018). It is well known that *Claviceps*
446 species are fungal parasites that share a rather unique relationship with their hosts (strict ovarian
447 parasites). The co-evolution between these parasites and host plants appears to be primarily
448 driven by the evolution and diversification of the host species (Píchová *et al.* 2018). This can be
449 inferred from divergence time estimates which show that the crown node of *C.* sect. *Pusillae*
450 aligns with the crown node of PACMAD grasses (ca. 45 Mya) (Bouchenak-Khelladi *et al.* 2010,
451 Píchová *et al.* 2018), suggesting that these two organisms radiated in tandem after ancestral

452 strains of *C.* sect. *Pusillae* were transferred to Africa. Similarly, the estimated crown node of *C.*
453 sect. *Claviceps* corresponds with the origin of the core Pooidea, which occurred in North
454 America (ca. 33-26 Mya) (Bouchenak-Khelladi *et al.* 2010; Sandve & Fjellheim 2010).

455 Our analyses suggest that the divergence of *C.* sect. *Claviceps*, from *C.* sect. *Pusillae*, is
456 associated with a loss of RIP and interestingly, a loss of repeat content (Fig. 3, 6). This loss led
457 to unconstrained tandem gene duplication within the section (Fig. 6-7; Table 2; Additional file 1:
458 Figure S7-S10) allowing for rapid speciation, as has been observed in several smut fungi
459 (Kämper *et al.* 2006; Schirawski *et al.* 2010; Dutheil *et al.* 2016). This is evident in *C. cyperi*, an
460 ancestral species of *C.* sect. *Claviceps*, which showed the least amount of gene family expansion
461 and tandem duplication (Fig. 7; Table 2; Additional file 1: Figure S8, S9), suggesting that gene
462 duplication could be contributing to the divergence of new species. This is evident from
463 increases in genome size, gene count, and closely related gene pairs (≥80% identity) of other *C.*
464 sect. *Claviceps* species (Table 1, 2). Within *C.* sect. *Claviceps* unconstrained gene duplication is
465 likely facilitated by recombination events during annual sexual reproduction (Esser & Tudzynski
466 1978). Future studies on recombination will be critical to our understanding of the mechanisms
467 driving gene duplication within the section.

468 These findings, along with a long divergent branch from our phylogenetic analyses and
469 the estimation of *C.* sect. *Claviceps* crown node (ca. 20.7 Mya) (Píchová *et al.* 2018) could
470 suggest that a sudden event sparked the adaptive radiation within this section (Fig. 1). The host
471 jump to BEP grasses could have caused the rapid speciation and genomic changes of *C.* sect.
472 *Claviceps*, similar to the suggested tandem radiation of *C.* sect. *Pusillae* with the PACMAD
473 grasses in Africa (stated previously). However, unknown factors might be responsible for the
474 drastic genomic reductions (i.e. loss of RIP and repeat content) observed in *C.* sect. *Claviceps*, as

475 no such changes were observed in *C.* sect. *Pusillae*. The radiation of the core Pooidea
476 (Aveneae/Poeae, Triticeae; BEP) occurred after a global super-cooling period (ca. 33-26 Mya) in
477 North America. During this period, Pooidea experienced a stress response gene family expansion
478 which enabled adaptation and diversification to cooler, more open, habitats (Sandve & Fjellheim
479 2010). As gene family expansion was observed in *C.* sect. *Claviceps* (the only section to infect
480 BEP grasses) it suggests that the same environmental factors that caused the radiation of
481 Pooideae could have similarly affected *C.* sect. *Claviceps* (Kondrashov 2012). Interestingly, one
482 of the gene families significantly expanded in *C.* sect. *Claviceps* (OG0000016) contains proteins
483 associated with a cold-adapted (Alias *et al.* 2014) serine peptidase S8 subtilase (MER0047718;
484 S08.139) (Additional file 3: Table S5). The crown node of *C.* sect. *Claviceps* does pre-date the
485 core Pooidea radiation event by around 10-5 Mya, but the 95% highest posterior density
486 determined in Píchová *et al.* (2018) could indicate both radiation events occurred at similar
487 times.
488 Further examination of *Claviceps* species in South and Central America needs to be
489 conducted to better elucidate the evolution and dispersal of the genus (Píchová *et al.* 2018).
490 Efforts should focus on the elusive *C. junci*, a pathogen of Juncaceae (rushes), which is thought
491 to reside in *C.* sect. *Claviceps* based on morphological and geographic characteristics (Langdon
492 1952; Píchová *et al.* 2018). This species, and potentially others, will provide further insight into
493 the early evolution of *C.* sect. *Claviceps* and could bridge the current gap between the
494 environmental factors that sparked the radiation of the core Pooidea and *C.* sect. *Claviceps*.
495
496 **Methods:**
497 *Sample acquisition*:

498 Field collected samples (Clav) were surfaced sterilized, allowed to grow as mycelia, and
499 individual conidia transferred to make single spore cultures. Thirteen cultures were provided by
500 Dr. Miroslav Kolařík from the Culture Collection of Clavicipitaceae (CCC) at Institute of
501 Microbiology, Academy of Sciences of the Czech Republic. Raw Illumina reads for samples
502 (LM28, LM582, LM78, LM81, LM458, LM218, LM454, LM576, and LM583) were
503 downloaded from NCBI's SRA database. Raw Illumina reads from an additional 21 LM samples
504 were generated by Dr. Liu's lab (AAFC), sequencing protocol of these 21 samples followed
505 Wingfield *et al.* (2018). Summarized information can be found in Additional file 1: Table S1.
506
507 *Preparation of genomic DNA*:
508 Cultures grown on cellophane PDA plates were used for genomic DNA extraction from
509 lyophilized mycelium following a modified CTAB method (Doyle and Doyle 1987; Wingfield *et*
510 *al.* 2018) without using the RNase Cocktail™ Enzyme Mix, only RNASE A was used. DNA
511 contamination was checked by running samples on a 1% agarose gel and a NanoDrop One[c]
512 (Thermo Fishcer Scientific). Twenty samples (7 Clav and 13 CCC) were sent to BGI-Hong Kong
513 HGS Lab for Illumina sequencing on a HiSeq™ 4000.
514
515 *Genome assembly:*
516 Preliminary data showed that raw reads of LM458 were contaminated with bacterial
517 DNA but showed strong species similarity to Clav32 and Clav50. To filter out the bacterial DNA
518 sequences, reads of LM458 were mapped against the assembled Clav32 and Clav50 genomes
519 using BBSplit v38.41 (Bushnell 2014). All forward and reverse reads mapped to each of the
520 genomes were concatenated, respectively. Both sets were then interleaved to remove duplicates

521 and used for further analysis. Reads for all 50 samples were checked for quality with FastQC
522 v0.11.5 (Andrews 2010) and trimmed with Trimmomatic v0.36 (Bolger *et al.* 2014) using the
523 commands (SLIDINGWINDOW:4:20 MINLEN:36 HEADCROP:10) to remove poor quality
524 data, only paired end reads were used. To better standardize the comparative analysis all 50
525 sample were subject to *de novo* genome assembly with Shovill v0.9.0
526 (https://github.com/tseemann/shovill) using SPAdes v3.11.1 (Nurk *et al.* 2013) with a minimum
527 contig length of 1000bp.

528 The reference genomes of *C. purpurea* strain 20.1 (SAMEA2272775), *C. fusiformis* PRL
529 1980 (SAMN02981339), and *C. paspali* RRC 1481 (SAMN02981342) were downloaded from
530 NCBI. Proteins for *C. fusiformis* and *C. paspali* were not available on NCBI so they were
531 extracted from GFF3 files provided by Dr. Chris Schardl and Dr. Neil Moore, University of
532 Kentucky, corresponding to the 2013 annotations (Schardl *et al.* 2013) available at
533 http://www.endophyte.uky.edu. Reference genomes were standardized for comparative analysis
534 with our 50 annotated genomes, by implementing a protein length cutoff of 50 amino acids and
535 removal of alternatively spliced proteins in *C. fusiformis* and *C. paspali*, only the longest spliced
536 protein for each locus remained.

537

538 *Transposable elements:*

539 Transposable elements (TE) fragments were identified following procedures for
540 establishment of *de novo* comprehensive repeat libraries set forth in Berriman *et al.* (2018)
541 through a combined use of RepeaModeler v1.0.8 (Smit & Hubley 2015), TransposonPSI (Hass
542 2010), LTR_finder v1.07 (Xu & Wang 2007), LTR_harvest v1.5.10 (Ellinghaus *et al.* 2008),
543 LTR_digest v1.5.10 (Steinbiss *et al.* 2009), Usearch v11.0.667 (Edgar 2010), and

544 RepeatClassifier v1.0.8 (Smit & Hubley 2015) with the addition of all curated fungal TEs from
545 RepBase (Bao *et al.* 2015). RepeatMasker v4.0.7 (Smit *et al.* 2015) was then used to identify TE
546 regions and soft mask the genomes. These steps were automated through construction of a
547 custom script, TransposableELMT (https://github.com/PlantDr430/TransposableELMT).
548 Divergence landscapes for TEs in all 53 *Claviceps* genomes were generated using a
549 custom script
550 (https://github.com/PlantDr430/CSU_scripts/blob/master/TE_divergence_landscape.py) and the
551 RepeatMasker output results. The RepeatMasker results were also used with the respective GFF3
552 file from each genome to calculate the average distance (kbp) of each gene to the closest TE
553 fragment on the 5' and 3' flanking side. Values were calculated for predicted effectors, non-
554 effector secreted genes, non-secreted metabolite genes, and all other genes using a custom script
555 (https://github.com/PlantDr430/CSU_scripts/blob/master/TE_closeness.py).
556
557 *Genome annotation:*
558 AUGUSTUS v3.2.2 (Mario *et al.* 2008) was used to create pre-trained parameters files
559 using the reference *C. purpurea* strain 20.1, available EST data from NCBI, and wild-type
560 RNAseq data (SRR4428945) created in Oeser *et al.* (2017). RNA-seq data was subject to quality
561 check and trimming as above. All three datasets were also used to train parameter files for the *ab*
562 *initio* gene model prediction software's GeneID v1.4.4 (Blanco *et al.* 2007) and CodingQuarry
563 v2.0 (Testa *et al.* 2015). GeneID training followed protocols available at
564 http://genome.crg.es/software/geneid/training.html. For CodingQuarry training, RNA transcripts
565 were created *de novo* using Trinity v2.8.4 (Grabherr *et al.* 2011) on default settings and EST

566 coordinates were found by mapping the EST data to the reference genome using Minimap2 v2.1
567 (Li 2018).
568 Gene models for the 50 genomes were then predicted with GeneID and CodingQuarry
569 using the trained *C. purpruea* parameter files. CodingQuarry prediction was also supplemented
570 with transcript evidence by mapping the available EST and RNA-seq *C. purpurea* data to each
571 genome using Minimap2. BUSCO v3 (Waterhouse *et al.* 2018) was run on all 50 genomes using
572 the AUGUSTUS *C. purpurea* pre-trained parameter files as the reference organism and the
573 Sordariomyceta database. The resulting predicted proteins for each sample were used as training
574 models for *ab initio* gene prediction using SNAP (Korf 2004) and GlimmerHMM v3.0.1
575 (Majoros *et al.* 2004). Lastly, GeMoMa v1.5.3 (Keilwagen *et al.* 2016) was used for *ab initio*
576 gene prediction using the soft-masked genomes and the *C. purpruea* 20.1 reference files.
577 Funannotate v1.6.0 (Palmer & Stajich 2019) was then used as the primary software for
578 genome annotation. Funannotate additionally uses AUGUSTUS and GeneMark-ES (Ter-
579 Hovhannisyan *et al.* 2008) for *ab initio* gene model prediction, Exonerate for transcript and
580 protein evidence alignment, and EVidenceModeler (Hass *et al.* 2008) for a final weighted
581 consensus. All *C. purpurea* EST and RNAseq data were used as transcript evidence and the
582 Uniport Swiss-Prot database and proteins from several closely related species (*C. purpurea* strain
583 20.1, *C. fusiformis* PRL1980, *C. paspali* RRC1481, *Fusarium oxysporum f.sp. lycopersici* 4287,
584 *Pochonia chlamydosporia* 170, *Ustilago maydis* 521, and *Epichloe festucae* F1) were used as
585 protein evidence. The AUGUSTUS pre-trained *C. purpurea* files were used as BUSCO seed
586 species along with the Sordariomyceta database and all five *ab initio* predictions were passed
587 through the --other_gff flag with weights of 1. The following flags were also used in Funannotate
588 "predict": --repeats2evm, --optimize_augustus, --soft_mask 1000, --min_protlen 50. BUSCO

589 was used to evaluate annotation completeness using the Dikarya and Sordariomyceta databases
590 (odb9) with --prot on default settings.
591
592 *Functional annotation:*
593 Functional analysis was performed using Funannotate "annotate". The following analyses
594 were also performed on the three reference *Claviceps* genomes. Secondary metabolite clusters
595 were predicted using antiSMASH v5 (Blin *et al.* 2019) with all features turned on. Functional
596 domain annotations were conducted using eggNOG-mapper v5 (Huerta-Cepas *et al.* 2016, 2019)
597 on default settings and InterProScan v5 (Jones *et al.* 2014) with the --goterms flag. Phobius
598 v1.01 (Käll 2007) was used to assist in prediction of secreted proteins. In addition to these
599 analyses Funannotate also performed domain annotations through an HMMer search against the
600 Pfam-A database and dbCAN CAZYmes database, a BLASTp search against the MEROPS
601 protease database, and secreted protein predictions with SignalP v4.1 (Nielsen 2017).
602 For downstream analysis, proteins were classified as secreted proteins if they had signal
603 peptides detected by both Phobius and SignalP and did not possess a transmembrane domain as
604 predicted by Phobius and an additional analysis of TMHMM v2.0 (Krogh *et al.* 2001). Effector
605 proteins were identified by using EffectorP v2.0 (Sperschneider *et al.* 2018), with default
606 settings, on the set of secreted proteins for each genome. Transmembrane proteins were
607 identified if both Phobius and TMHMM detected transmembrane domains. Secondary metabolite
608 proteins were identified if they resided within metabolite clusters predicted by antiSMASH.
609 Proteins were classified as having conserved protein domains if they contained any Pfam or IPR
610 domains.
611

612 *Gene family identification and classification*:

613 OrthoFinder v2.3.3 (Emms *et al.* 2019) was used to infer groups of orthologous gene
614 families (orthogroups). To more accurately place closely related genes into orthogroups an
615 additional 78 fungal genomes (Table 2) with emphasis on plant associated fungi of the order
616 Hypocreales were added. To standardize, all 78 additional genomes were subject to a protein
617 length cutoff of 50 amino acids and genomes downloaded from http://www.endophyte.uky.edu
618 had alternatively spliced proteins removed. OrthoFinder was run using Diamond v0.9.25.126
619 (Buchfunk *et al.* 2015) on default settings. For downstream analysis, orthogroups pertaining to
620 the 53 *Claviceps* genomes were classified as secreted, predicted effectors, transmembrane,
621 metabolite, and conserved domain clusters if ≥50% of the *Claviceps* strains present in a given
622 orthogroup had at least 1 protein classified as such.

623

624 *Phylogeny and genome fluidity*:

625 Phylogenetic relationship of all 53 *Claviceps* genomes, with *Fusarium graminearum*, *F.*
626 *verticillioides*, *Epichloe festucae*, and *E. typhina* as outgroups, was derived from 2,002 single-
627 copy orthologs obtained from OrthoFinder. This resulted in a dataset of 114,114 amino acids
628 sequences which were aligned using MAFFT v7.429 (Katoh & Standely 2013) on default
629 settings. Uninformative sites were removed using Gblocks v0.91 (Castresana 2000) on default
630 settings. Due to the large scale of the alignment maximum likelihood reconstruction was
631 performed using FastTree v2.1.11 (Price *et al.* 2010) using the WAG model of amino acid
632 substitution with the -gamma and -slow flag, and 1000 bootstraps. MEGA X (Sudhir *et al.* 2018)
633 was used for neighbor-joining (NJ) reconstruction using the JTT model of amino acid
634 substitution with gamma distribution and maximum parsimony (MP) reconstruction using the

635 tree bisection reconstruction (TBR) algorithm with 100 repeated searches. Nodal support for
636 both NJ and MP reconstructions were assessed with 1000 bootstraps.
637 Genomic fluidity was used to assess gene family dissimilarity within the *Claviceps* genus
638 (Kislyuk *et al.* 2011). Datasets containing gene families from representative members of *C.* sect.
639 *Pusillae*, *C.* sect. *Claviceps*, *Clavieps* genus, and all *C. purpurea* strains were extracted from our
640 OrthoFinder results. Additional species- and genus-wide gene family datasets from the additional
641 78 fungal genomes were extracted for comparative purposes. All section- and genus-wide
642 datasets contained 1 representative strain from each species to reduce phylogenetic bias. Each
643 extracted dataset was used to calculate the genomic fluidity following procedures described in
644 Kislyuk *et al.* (2011) using a custom script
645 (https://github.com/PlantDr430/CSU_scripts/blob/master/pangenome_fluidity.py). The result
646 files for each dataset were then used for figure creation and two-sample two-sided z-test statistics
647 (Kislyuk *et al.* 2011) using a custom script
648 (https://github.com/PlantDr430/CSU_scripts/blob/master/combine_fluidity.py).
649
650 *Gene compartmentalization:*
651 A custom script
652 (https://github.com/PlantDr430/CSU_scripts/blob/master/genome_speed_hexbins.py) was used
653 to calculate and plot the intergenic regions of each gene, overlay gene classifications, display
654 distribution of genes, and statistically test whether the overlaid genes had similar intergenic
655 lengths compared to all other genes in the genomes. Statistics were performed by resampling 100
656 random overlaid genes and 100 other genes 1,000 times for both the 5' and 3' flanking distances.
657 Mann-Whitney U test was used to test for significance on all 2,000 subsets corrected with

658 Benjamini/Hochberg. Corrected p-values were averaged per flanking side and then together to
659 get a final p-value. Genes that appeared on a contig alone were excluded. In addition, genes that
660 were located at the start of each contig were plotted along the x-axis, while genes located at the
661 end of each contig were plotted along the y-axis.
662
663 *RIP and BLAST analyses*:
664 For all 53 genomes a self BLASTp v2.9.0+ search was conducted to identify best hit
665 orthologs within each genome with a cutoff e-value of 10^{-5} and removal of self-hits. This process
666 was automated, using a custom script
667 (https://github.com/PlantDr430/CSU_scripts/blob/master/RIP_blast_analysis.py). For genes with
668 an identity of ≥80% we determined if the pair was located next to each other and/or separated by
669 5 or less genes. Fifty-six important *Claviceps* genes (Additional file 2: Table S4) including the
670 *rid-1* homolog (Freitag *et al.* 2002) were used in a BLASTp analysis to identify the number of
671 genes present that passed an e-value cutoff of 10^{-5}, 50% coverage, and 35% identity. Genes that
672 appeared as best hits for multiple query genes were only recorded once for their overall best
673 match.
674
675 *Statistical programs and plotting:*
676 Statistics and figures were generated using Python3 modules SciPy v1.3.1, statsmodel
677 v0.11.0, and Matplotlib v3.1.1. Heatmaps were generated using ComplexHeatmap v2.2.0 in R
678 (Gu 2016).
679
680 **Declarations:**

Acknowledgements:

We thank Dr. Miroslav Kolařík for providing *Claviceps* isolates from the Culture Collection of Clavicipitaceae at Institute of Microbiology, Academy of Sciences of the Czech Republic (CCC samples); Dr. Parivash Shoukouhi, Dr. Jim Menzies, and Dr. Zlatko Popovic for collection, isolation, maintenance, and DNA extraction of LM samples; Dr. Chris Schardl and Dr. Neil Moore, University of Kentucky, for providing the 2013 GFF3 files for *C. paspali* and *C. fusiformis*; Dr. Joshua Weitz and the Franklin Graybill Statistical Laboratory at Colorado State University for their help in data analysis of genomic fluidity; Molecular Technologies Laboratory (MTL) at the Ottawa Research & Development Centre, Agriculture and Agri-Food Canada, especially Kasia Dadej for technical assistance. For genomes downloaded from JGI these sequence data were produced by the US Department of Energy Joint Genome Institute https://www.jgi.doe.gov/ in collaboration with the user community.

Funding:

This work is supported by the Agriculture and Food Research Initiative (AFRI) National Institute of Food and Agriculture (NIFA) Fellowships Grant Program: Predoctoral Fellowships grant no. 2019-67011-29502/project accession no. 1019134 from the United States Department of Agriculture (USDA), and by the American Malting Barley Association grant no. 17037621. Whole-genome sequencing of LM samples was supported, in part, by funding provided to Dr. Jeremy Dettman from Agriculture and Agri-Food Canada's Biological Collections Data Mobilization Initiative (BioMob, Work Package 2, project J-001564).

704 **Availability of data and materials:**

705 Datasets and scripts are available on Dryad: Wyka, Stephen *et al.* (2020), Genus-wide
706 comparison reveals divergence and evolution of the four sections within the genus Claviceps are
707 the result of varying mechanisms driving genome evolution and host range, Dryad,
708 Dataset, https://doi.org/10.5061/dryad.18931zcsk. (Submitted upon publication)

709 Genomes and Illumina raw reads were deposited to NCBI under the BioProject
710 PRJNA528707 (Additional file 1: Table S1). Scripts are maintained within the GitHub repository
711 of the primary author's, https://github.com/PlantDr430/CSU_scripts. TransposableELMT can be
712 found at Zenodo DOI: 105281/zenodo3469661. All phylogenetic trees were made available at
713 TreeBase (ID:XXXX).

714

715 **Author Contributions:**

716 The project was conceived and designed by S.A.W., S.J.M., and K.B.; S.A.W. performed
717 the research, annotations, bioinformatic workflows, and analyzed the data with technical
718 troubleshooting from S.J.M.; M.L. and J.R.D initiated whole-genome sequencing of LM
719 samples; M.L., V.N., and K.B. provided management, research advice, and editorial
720 contributions; S.A.W. wrote the paper with contributions from all other authors.

721

722 **Ethics approval and consent to participate:**

723 Not applicable

724

725 **Consent for publication:**

726 Not applicable

727

728 **Competing interests:**

729 The authors declare that they have no competing interests.

730

731 **References:**

732 Alderman SC, Halse RR, White JF (2004) A reevaluation of the host range and geographical
733 distribution of *Claviceps* species in the United States. *Plant Disease* **88,** 63–81.

734 Alias N, Ahmad Mazian M, Salleh AB, Basri M, Rahman RN (2014) Molecular cloning and
735 optimization for high level expression of cold-adapted serine protease from antarctic yeast
736 *Glaciozyma antarctica* PI12. *Enzyme Research* **2014,** 197938.

737 Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available
738 online at: http://wwwbioinformaticsbabrahamacuk/projects/fastqc.

739 Avalos J, Carmen Limon M (2015) Biological roles of fungal carotenoids. *Current Genetics* **61,**
740 309–324.

741 Bao W, Kojima KK, Kohany O (2015) Repbase Update, a database of repetitive elements in
742 eukaryotic genomes. *Mobile DNA*, **6,** 11

743 Berriman M, Coghlan A, Tsai IJ (2018) Creation of a comprehensive repeat library for newly
744 sequenced parasitic worm genome. *Protocol Exchange* DOI: 101038/protex2018054.

745 Blanco E, Parra G, Guigo R (2007) Using geneid to identify genes. *Current Protocols in*
746 *Bioinformatics* **18**, 431-4328.

747 Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T (2019)
748 antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids*
749 *Research* **47,** W81-87.

750 Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence
751 data. *Bioinformatics* **30,** 2114-20.
752 Bormann J, Tudzynski P (2009) Deletion of Mid1, a putative stretch-activated calcium channel
753 in *Claviceps purpurea*, affects vegetative growth, cell wall synthesis and virulence.
754 *Microbiology* **155,** 3922-3933.
755 Bouchenak-Khelladi Y, Verboom GA, Savolainen V, Hodkinson TR (2010) Biogeography of the
756 grasses (Poaceae): a phylogenetic approach to reveal evolutionary history in geographical
757 space and geological time. *Botanical Journal of the Linnean Society* **162,** 543–557.
758 Buchfunk B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND.
759 *Nature Methods* **12,** 59-60.
760 Bushnell, B (2014) BBMap: A fast, accurate, splice-aware aligner. Available at:
761 https://sourceforgenet/projects/bbmap/.
762 Cantu D, Segovia V, MacLean D, Bayles R, Chen X, Kamoun S, Dubcovsky J, Saunders DGO,
763 Uauy C (2013) Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia*
764 *striiformis* f sp *tritici*reveal polymorphic and haustorial expressed secreted proteins as
765 candidate effectors. *BMC Genomics* **14,** 270.
766 Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in
767 phylogenetic analysis. *Molecular Biology and Evolution* **17,** 540-552.
768 Chalvet F, Grimaldi C, Kaper F, Langin T, Dabousii MJ (2003) Hop, an active Mutator-like
769 element in the genome of the fungus *Fusarium oxysporum. Molecular Biology and Evolution*
770 **20,** 1362–1375.

771 Daviere JM, Langin T, Daboussi MJ (2001) Potential role of transposable elements in the rapid
772 reorganization of the Fusarium oxysporum genome. *Fungal Genetics and Biology* **34,** 177–
773 192.
774 Dong S, Raffaele S, Kamoun S (2015) The two-speed genomes of filamentous pathogens: waltz
775 with plants. *Current Opinion in Genetics and Development* **35,** 57–65.
776 Dopstadt J, Neubauer L, Tudzynski P, Humpf H-U (2016) The epipolythiodiketopiperazine gene
777 cluster in *Claviceps purpurea*: dysfunctional cytochrome p450 enzyme prevents formation of
778 the previously unknown clapurines. *PLoS ONE* **11,** e0158945.
779 Douhan GW, Smith ME, Huyrn KL, Westbrook A, Beerli P, Fisher AJ (2008) Multigene
780 analysis suggests ecological speciation in the fungal pathogen *Claviceps purpurea*.
781 *Molecular Ecology* **17,** 2276-2286.
782 Doyle JJ, Doyle JL (1987) A rapid DNA isolation precedure for small quantities of fresh leaf
783 tissue. *Phytochemical Bulletin* **19,** 11-15.
784 Dutheil JY, Mannhaupt G, Schweizer G, Sieber CMK, Munsterkotter M, Guldener U,
785 Schirawski J, Kahmann R (2016) A tale of genome compartmentalization: the evolution of
786 virulence clusters in smut fungi. *Genome Biology and Evolution* **8,** 681–704.
787 Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*
788 **26,** 2460-2461.
789 Ellinghaus D, Kurtz S, Willhoeft U (2008) LTRharvest, a efficient and flexible software for de
790 novo detection of LTR retrotransposons. *BMC Bioinformatics* **9,** 18.
791 Emms DM, Kelly S (2018) OrthoFinder: phylogenetic orthology inference for comparative
792 genomics. *bioRxiv* 466201 DOI: 101101/466201.

793 Esser K, Tudzynski P (1978) Genetics of the ergot fungus *Claviceps purpurea*. *Theoretical and*
794 *Applied Genetics* **53,** 145-149.
795 Estep LK, Torriani SFF, Zala M, Anderson NP, Flowers MD, McDonald BA, Mundt CC,
796 Brunner PC (2015) Emergence and early evolution of fungicide resistance in North American
797 populations of *Zymoseptoria tritici*. *Plant Pathology* **64,** 961-971.
798 Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GCM, Wittenberg AHJ, Thomma
799 BPHJ (2016) Transposons passively and actively contribute to evolution of the two-speed
800 genome of a fungal pathogen. *Genome Research* **26,** 1091–1100.
801 Fisher AJ, DiTomaso JM, Gordon TR, Aegerter BJ, Ayres DR (2007) Salt marsh *Claviceps*
802 *purpurea* in native and invaded Spartina marshes in Northern California. *Plant Disease* **91,**
803 380–386.
804 Flieger M, Stodulkova E, Wyka SA, *et al.* (2019) Ergochromes: heretofore neglected side of
805 ergot toxicity. *Toxins* **11,** 439.
806 Frantzeskakis L, Kracher B, Kusch S, *et al.* (2018) Signatures of host specialization and a recent
807 transposable element burst in the dynamic one-speed genome of the fungal barley powdery
808 mildew pathogen. *BMC Genomics* **19,** 381.
809 Frantzeskakis L, Kusch S, Panstruga R (2019) The need for speed: compartmentalized genome
810 evolution in filamentous phytopathogens Molecular Plant Pathology **20(1):** 3-7
811 Freitag M, Williams RL, Kothe GO, Selker EU (2002) A cytosine methyltransferase homologue
812 is essential for repeat_induced point mutation in *Neurospora crassa*. *PNAS* **99,** 8802-8807.
813 Galagan JE Selker EU (2004) RIP: the evolutionary cost of genome defense. *TRENDS in*
814 *Genetics* **20,** 417-423.

815 Galagan JE, Calvo SE, Borkovich KA *et al*. (2003) The genome sequence of the filamentous
816 fungus *Neurospora crassa*. *Nature* **422,** 859-868.
817 Giesbert S, Schurg T, Scheele S, Tudzynski P (2008) The NADPH oxidase Cpnox1 is required
818 for full pathogenicity of the ergot fungus *Claviceps purpurea*. *Molecular Plant Pathology* **9,**
819 317-327.
820 Grabherr MG, Haas BJ, Yassour M, *et al.* (2011) Full-length transcriptome assembly from RNA-
821 seq data without a reference genome. *Nature Biotechnology* **29,** 644-52.
822 Graia F, Lespinet O, Rimbault B, Deguard-Chablat M, Coppin E, Picard M (2001) Genome
823 quality control: RIP (repeat-induced point mutation) comes to *Podospora*. *Molecular*
824 *Microbiology* **40,** 586–595.
825 Grandaubert J, Bhattacharyya A, Stukenbrock EH (2015) RNA-seq-based gene annotation and
826 comparatrive genomics of four fungal grass pathogens in the genus *Zymoseptoria* identify
827 novel orphan genes and species-specific invasions of transposable elements. *Genes,*
828 *Genomics, Genetics* **5,** 1323-1333.
829 Grandaubert J, Dutheil JY, Stukenbrock EH (2019) The genomic determinants of adaptive
830 evolution in a fungal pathogen. *Evolution Letters* **3,** 299-312.
831 Gu Z, Eils R, Schlesner M (2016) Complex heatmaps reveal patterns and correlations in
832 multidimensional genomic data. *Bioinformatics* **32,** 2847-9.
833 Hass B (2010) TransposonPSI. Available online at: http://transposonpsisourceforgenet.
834 Hass BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR
835 (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the
836 program to assemble spliced alignments. *Genome Biology* **9,** R7.

837 Hinsch J, Galuszka P, Tudzynski P (2016) Functional characterization of the first filamentous
838 fungal tRNA-isopentenyltransferase and its role in the virulence of *Claviceps purpurea*. *New*
839 *Phytologist* **211,** 980-992.
840 Hinsch J, Vrabka J, Oeser B, Novak O, Galuszka P, Tudzynski P (2015) *De novo* biosynthesis of
841 cytokinins in the biotrophic fungus *Claviceps purpurea*. *Environmental Microbiology* **17,**
842 2935-2951.
843 Huerta-Cepas J, Szklarczyk D, Jensen LJ, von Mering C, Bork P (2016) Fast genome-wide
844 functional annotation through orthology assignment by eggnog-mapper. *Molecular Biology*
845 *and Evolution* **34,** 2115-2122.
846 Huerta-Cepas J, Szklarczyk D, Heller D, *et al.* (2019) eggnog 5.0: a hierarchical, functionally
847 and phylogenetically annotated orthology resource based on 5090 organisms and 2502
848 viruses. *Nucleic Acids Research* **47,** D309-314
849 Ikeda K, Nakayashiki H, Tosa Y, Mayama S (2002) Repeat-induced point mutation (RIP) in
850 *Magnaporthe grisea*: implications for its sexual cycle in the natural field context. *Molecular*
851 *Microbiology* **45,** 1355–1364.
852 Jones P, Binns D, Chang H-Y, *et al.* (2014) InterProScan 5: genome-scale protein function
853 classification. *Bioinformatics* **30,** 1236-1240.
854 Jungehülsing U, Tudzynski P (1997) Analysis of genetic diversity in *Claviceps purpurea* by
855 RAPD markers. *Mycological Research* **101,** 1–6.
856 Kachroo P, Leong SA, Chatto BB (1994) Pot2, an inverted repeat transposon from the rice blast
857 fungus *Magnaporthe grisea*. *Molecular and General Genetics* **245,** 339–348.
858 Käll L, Krogh A, Sonnhammer EL (2007) Advantages of combined transmembrane topology and
859 signal peptide prediction--the Phobius web server. *Nucleic Acids Research* **35,** W429-32.

860 Kämper J, Kahmann R, Bolker M *et al.* (2006) Insights from the genome of the biotrophic fungal
861 plant pathogen *Ustilago maydis*. *Nature* **444,** 97–101.
862 Katoh K, Standley, DM (2013) MAFFT multiple sequence alignment software version 7:
863 improvements in performance and usability. *Molecular Biology and Evolution* **30,** 772-780.
864 Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F (2016) Using intron
865 position conservation for homology-based gene prediction. *Nucleic Acids Research* **44,** e89.
866 Kiran K, Rawal HC, Dubey H, *et al.* (2016) Draft genome of the wheat rust pathogen (*Puccinia*
867 *triticina*) unravels genome-wide structural variations during evolution. *Genome Biology and*
868 *Evolution* **8,** 2702-2721.
869 Kiran K, Rawal HC, Dubey H, Jaswal R, Bhardwaj SC, Prasad P, Pal D, Devanna BN, Sharma
870 TR (2017) Dissection of genomic features and variations of three pathotypes of Puccinia
871 striiformis through whole genome sequencing. *Scientific Reports* **7,** 42419.
872 Kislyuk AO, Haegeman B, Bergman NH, Weitz JS (2011) Genomic fluidity: an integrative view
873 of gene diversity within microbial populations. *BMC Genomics* **12,** 32.
874 Kito H, Takahashi Y, Sato J, Fukiya S, Sone T, Tomita F (2003) Occan, a novel transposon in
875 the Fot1 family, is ubiquitously found in several Magnaporthe grisea isolates. *Current*
876 *Genetics* **42,** 322–331.
877 Kondrashov FA (2012) Gene duplication as a mechanism of genomic adaptation to a changing
878 environment. *Proceedings of the Royal Society B* **279,** 5048-5057.
879 Korf I (2004) Gene finding in novel Genomes. *BMC Bioinformatics* **5,** 59.
880 Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein
881 topology with a hidden markov model: application to complete genomes. *Journal of*
882 *Molecular Biology* **305,** 567-580.

883 Kvas M, Marasas WFO, Wingfield BD, Wingfield MJ, Steenkamp ET (2009) Diversity and
884 evolution of *Fusarium* species in the *Gibberella fujikuroi* complex. *Fungal Diversity* **34,** 1-
885 21.
886 Langdon RFN (1952) Studies on ergot (PhD thesis). Queensland University, Brisbane, Australia
887 Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34,** 3094-
888 3100.
889 Ma L-J, van der Does HC, Borkovich KA, *et al*. 2010 Comparative genomics reveals mobile
890 pathogenicity chromosomes in *Fusarium*. *Nature* **464,** 367-373.
891 Majoros WH, Pertea M, Salzberg SL (2004) TigrScan and GlimmerHMM: two open-source ab
892 initio eukaryotic gene-finders. *Bioinformatics* **20,** 2878-2879.
893 Mario S, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped
894 cDNA alignments to improve *de novo* gene finding. *Bioinformatics* **24,** 637–644.
895 Mey G, Oeser B, Lebrun MH, Tudzynski P (2001) The biotrophic, non-appressorium-forming
896 grass pathogen *Claviceps purpurea* needs a *Fus3/Pmk1* homologous mitogen-activated
897 protein kinase for colonization of rye ovarian tissue. *Molecular Plant-Microbe Interactions*
898 **15,** 303-312.
899 Mey G, Held K, Scheffer J, Tenberge KB, Tudzynski P (2002) CPMK2, an SLT2-homologous
900 mitogen-activated protein (MAP) kinase, is essential for pathogenesis of *Claviceps purpurea*
901 on rye: evidence for a second conserved pathogenesis-related MAP kinase cascade in
902 phytopathogenic fungi. *Molecular Microbiology* **46,** 305-318.
903 Mlynarčíková Eva (2015) Delece genu CpiaaH u fytopatogenní houby *Claviceps purpurea*
904 Olomouc, bakalářská práce (Bc) Univerzita Palackeho V Olomouci Přírodovědecká fakulta.
905 Available from: https://theses.cz/id/chk7nx (Bachelor's Thesis).

906 Moller M, Stukenbrock EH (2017) Evolution and genome architecture in fungal plant pathogens.
907 *Nature Reviews Microbiology* **15,** 756-771.
908 Nahan AR and Sarkar IN (2012) The impact of taxon sampling on phylogenetic inference: a
909 review of two decades of controversy. *Brief Bioinformatics* **13,** 122-134.
910 Nakayashiki H, Nishimoto N, Ikeda K, Tosa Y, Mayama S (1999) Degenerate MAGGY
911 elements in a subgroup of *Pyricularia grisea*: a possible example of successful capture of a
912 genetic invader by a fungal genome. *Molecular and General Genetics* **261,** 958–966.
913 Neubauer L, Dopstadt J, Humpf H-U, Tudzynski P (2016) Identification and characterization of
914 the ergochrome gene cluster in the plant pathogenic fungus *Claviceps purpurea*. *Fungal*
915 *Biology and Biotechnology* **3,** 2.
916 Nielsen H (2017) Predicting secretory proteins with SignalP In: *Protein function prediction* (eds.
917 Kihara D). *Methods in Molecular Biology* **1611** Humana Press, New York, NY.
918 Nurk S, Bankevich A, Antipov D, *et al.* (2013) Assembling genomes and mini-metagenomes
919 from highly chimeric reads. *Research in Computational Molecular Biology RECOMB 2013*
920 *Lecture Notes in Computer Science*, vol 7821 Springer, Berlin, Heidelberg.
921 Oeser B, Heidrich PM, Muller U, Tudzynski P, Tenberge KB (2002) Polygalacturonase is a
922 pathogenicity factor in the *Claviceps purpurea*/rye interaction. *Fungal Genetics and Biology*
923 **36,** 176-186.
924 Oeser B, Kind S, Schurack S, Schmutzer T, Tudzynski P, Hinsch J (2017) Cross-talk of the
925 biotrophic pathogen *Claviceps purpurea* and its host *Secale cereal*. *BMC Genomics* **18,** 273.
926 Palmer J, Stajich J (2019) nextgenusfs/funannotate: funannotate v1.6.0 (Version 1.6.0) Zenodo
927 DOI: 105281/zenodo3354704.

928 Pažoutová S, Olšovská J, Linka M, Kolínská R, Flieger M (2000) Chemoraces and habitat
929 specialization of *Claviceps purpurea* populations. *Applied and Environmental Microbiology*
930 **66,** 5419–5425.
931 Pažoutová S, Pešicová K, Chudíčková M, Šrůtka P, Kolařík M 2015 Delimitation of cryptic
932 species inside *Claviceps purpurea*. *Fungal Biology* **119,** 7–26.
933 Pažoutová S, Raybould AF, Honzátko A, Kolínská R (2002) Specialised population of *Claviceps*
934 *purpurea* from salt marsh *Spartina* species. *Mycological Research* **106,** 210–214.
935 Pichova K, Pazoutova S, Kostovcik M, Chudickova M, Stodulokva E, Novak P, Flieger M, van
936 der Linde E, Kolarik M (2018) Evolutionat histroy of ergot with a new infrageneric
937 classificiation (Hypocreales: Clavicipitaceae: *Claviceps*). *Molecular Phylogenetics and*
938 *Evolution* **123,** 73-87.
939 Poppe S, Dorcheimer L, Happel P, Stukenbrock EH (2015) Rapidly evolving genes are key
940 players in host specialization and virulence of the fungal wheat pathogen *Zymoseptoria tritici*
941 (*Mycosphaerella graminicola*). *PLoS Pathogens* **11,** e1005055.
942 Posada D (2016) Phylogenomics for systemic biology. *Systematic Biology* **65,** 353-6.
943 Price MN, Dehal PS, Arkin AP (2010) FastTree 2 – approximately maximum-likelihood trees for
944 large alignments. *PLoS ONE* **5,** e9490.
945 Raffaele S, Kamoun S (2012) Genome evolution in filamentous plant pathogens: why bigger can
946 be better. *Nature Reviews Microbiology* **10,** 417–430.
947 Raybould AF, Gray AJ, Clarke RT (1998) The long-term epidemic of *Claviceps purpurea* on
948 *Spartina anglica* in Poole Harbour: pattern of infection, effects on seed production and the
949 role of *Fusarium heterosporum*. *New Phytologist* **138,** 497–505.

950 Rep M, Kistler HC (2010) The genomic organization of plant pathogenicity in *Fusarium* species.
951 *Current Opinion in Plant Biology* **13,** 420-426.
952 Rolke Y, Tudzynski P (2008) The small GTPase Rac and the p21-activated kinase Cla4 in
953 *Claviceps purpurea*: interaction and impact on polarity, development, and pathogenicity.
954 *Molecular Microbiology* **68,** 405-423.
955 Sandve SR, Fjellheim S (2010) Did gene family expansions during the Eocene- Oligocene
956 boundary climate cooling play a role in Pooideae adaptation to cool climates? *Molecular*
957 *Ecology* **19,** 2075–2088.
958 Schardl CL, Young CA, Hesse U, *et al*. (2013) Plant-symbiotic fungi as chemical engineers:
959 multi-genome analysis of the Clavicipitaceae reveals dynamics of alkaloid loci. *PLoS*
960 *Genetics* **9,** e1003323.
961 Scheffer J, Chen C, Heidrich P, Dickman MB, Tudzynski P (2005a) A CDC42 homologue in
962 *Claviceps purpurea* is involved in vegetative differentiation and is essential for
963 pathogenicity. *Eukaryotic Cell* **4,** 1228-1238.
964 Scheffer J, Ziv C, Yarden O, Tudzynski P (2005b) The COT1 homologue CPCOT1 regulates
965 polar growth and branching and is essential for pathogenicity in *Claviceps purpurea*. *Fungal*
966 *Genetics and Biology* **42,** 107-118.
967 Schirawski J, Mannhaupt G, Munch K *et al.* (2010) Pathogenicity determinants in smut fungi
968 revealed by genome comparison. *Science* **330,** 1546–1548.
969 Shoukouhi P, Hicks C, Menzies JG, Popovic Z, Chen W, Seifert KA, Assabgui R, Liu M (2019)
970 Phylogeny of Canadian ergot fungi and a detection assay by real-time polymerase chain
971 reaction Mycologia **111,** 1-13.

972 Smit AFA, Hubley R (2015) RepeatModeler Open-10. Available online at:
973 http://wwwrepeatmaskerorg.
974 Smit AFA, Hubley R, Green P (2015) RepeatMasker Open-40. Available online at:
975 http://wwwrepeatmaskerorg.
976 Sperchneider J, Gardiner DM, Thatcher LF, Lyons R, Singh KB, Manners JM, and Taylor JM
977 (2015) Genome-wide analysis in three *Fusarium* pathogens identifies rapidly evolving
978 chromosomes and genes associated with pathogenicity. *Genome Biology and Evolution* **7,**
979 1613-1627.
980 Sperschneider J, Dobbs PN, Gardiner DM, Singh KB, Taylor JM (2018) Improved prediction of
981 fungal effector proteins from secretomes with EffectorP 2.0. *Molecular Plant Pathology* **19,**
982 2094-2110.
983 Steinbiss S, Willhoeft U, Gremme G, Kurtz S (2009) Fine-grained annotation and classification
984 of de novo predicted LTR retrotransposons. *Nucleic Acids Research* **37,** 7002–7013.
985 Stukenbrock EH (2013) Evolution, selection and isolation: a genomic view of speciation in
986 fungal plant pathogens. *New Phytologist* **199,** 895–907.
987 Stukenbrock EH, Dutheil JY (2017) Comparison of finescale recombination maps in fungal plant
988 pathogens reveals dynamic recombination landscapes and intragenic hotspots. *Genetics* **208,**
989 1209-1229.
990 Sudhir K, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary
991 Genetics Analysis across computing platforms. *Molecular Biology and Evolution* **35,** 1547-
992 1549.

993 Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M (2008) Gene prediction in
994 novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome*
995 *Research* **18,** 1979-1990.
996 Testa A, Hane JK, Ellwood SR, Oliver RP (2015) CodingQuarry: highly accurate hidden markov
997 model gene prediction in fungal genomes using RNA-seq transcripts. *BMC Genomics* **16,**
998 170.
999 Testa A, Oliver R, Hane J (2015) Overview of genomic and bioinformatic resources for
1000 *Zymoseptoria tritici*. *Fungal Genetics and Biology* **79,** 13-16.
1001 Urguhart AS, Mondo SJ, Makela MR *et al.* (2018) Genomic and genetic insights into a
1002 cosmopolitan fungus, *Paecilomyces variotii* (Eurotiales). *Frontiers in Microbiology* **9,** 3058.
1003 Van der Linde EJ, Pešicová K, Pažoutová S, Stodůlková E, Flieger M, Kolařík M (2016) Ergot
1004 species of the *Claviceps purpurea* group from South Africa. *Fungal Biology* **120,** 917-930.
1005 Wäli PP, Wäli PR, Saikkonen K, Tuomi J (2013) Is the pathogenic ergot fungus a conditional
1006 defensive mutualist for its host grass? *PLoS ONE* **8,** e69249.
1007 Watanabe M, Yonezawa T, Lee K, Kumagai S, Sugita-Konishi Y, Goto K, Hara-Kudo Y
1008 (2011) Molecular phylogeny of the higher and lower taxonomy of the *Fusarium* genus and
1009 differences in the evolutionary histories of multiple genes. *BMC Evol Biol* **11,** 322.
1010 Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva
1011 EV, Zdobnov EM (2018) BUSCO applications from quality assessments to gene prediction
1012 and phylogenomics. *Molecular Biology and Evolution* **35,** 543-548.
1013 Wingfield BD, Miao L, Nguyen HDT, *et al.* (2018) Nine draft genome sequences of *Claviceps*
1014 *purpurea* s.lat., including *C. arundinis*, *C. humidiphila*, and *C.* cf. *spartinae*,
1015 pseudomolecules for the pitch canker pathogen *Fusarium circinatum*, draft genome of

1016 *Davidsoniella eucalypti*, *Grosmannia galeiformis*, *Quambalaria eucalypti*, and
1017 *Teratosphaeria destructans*. *IMA Fungus* **9,** 401-418.
1018 Wu B, Macielog AI, Hao W (2017) Origin and spread of spliceosomal introns: insights from the
1019 fungal clade *Zymoseptoria*. *Genome Biology and Evolution* **9,** 2658-2667.
1020 Xu Z, Wang H (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR
1021 retrotransposons. *Nucleic Acids Research* **35,** W265–268.
1022

1023 **Figure Captions:**

1024

1025 **Figure 1:** Maximum likelihood phylogenetic reconstruction of the *Claviceps* genus using amino
1026 acid sequences of 2,002 single copy orthologs with 1000 bootstrap replicates. Pink dots at
1027 branches represent bootstrap values ≥95. Arrows and descriptions indicate changes in genomic
1028 architecture between *Claviceps* sections that will be described in this study.

1029

1030 **Figure 2:** Genomic fluidity (dashed lines) for specified groups within the order Hypocreales.
1031 Species level groups contain multiple isolates of a given species, while section and genus level
1032 groups contain one strain from representative species to remove phylogenetic bias. Shaded
1033 regions represent standard error and were determined from total variance, containing both the
1034 variance due to the limited number of samples genomes and the variance due to subsampling
1035 within the sample of genomes. Letters correspond to significant difference between fluidities
1036 determined through a two-sided two-sample z-test ($P < 0.05$; Additional file 1: Table S3)
1037 (Kislyuk *et al.* 2011).

1038

1039 **Figure 3:** Transposable element (TE) fragment divergence landscapes for representative species
1040 of each *Claviceps* section; *C. purpurea* 20.1 (*C.* sect. *Claviceps*), *C. maximensis* CCC398 (*C.*
1041 sect. *Pusillae*), *C. paspali* RRC1481 (*C.* sect. *Paspalorum*), and *C. citrina* (*C.* sect. *Citrinae*).
1042 Stacked bar graphs show the non-normalized sequence length occupied in each genome (y-axis)
1043 for each TE type based on their percent divergence (x-axis) from their corresponding consensus
1044 sequence. Landscape for all remaining species can be seen in Additional file 1: Figure S4.
1045
1046 **Figure 4:** Boxplot distributions of predicted effectors, secreted (non-effectors), secondary
1047 metabolite (non-secreted) genes and other genes (i.e. genes that are not effectors, secreted, or
1048 secondary metabolite genes) in *Claviceps* sections showing the mean distance (kbp) of each gene
1049 to the closest transposable element fragment (5' and 3' flanking distances were averaged
1050 together). Kruskal Wallis (*P*-value; * < 0.05, ** < 0.01, *** < 0.001, n.s. = not significant).
1051 Pairwise comparison was performed with two-sided Mann-Whitney U-test with
1052 Benjamini/Hochberg multitest correction. Letters correspond to significant differences between
1053 gene categories within sections ($P < 0.05$). Plots for all individual species can been seen in
1054 Additional file 1: Figure S5.
1055
1056 **Figure 5:** 5' and 3' intergenic region size (y- and x-axis) of representative isolates of each of the
1057 four sections within the *Claviceps* genus; *C, purpurea* 20.1 (*C.* sect. *Claviceps*), *C. maximensis*
1058 CCC398 (*C.* sect. *Pusillae*), *C. paspali* RRC1481 (*C.* sect. *Paspalorum*), and *C. citrina* (*C.* sect.
1059 *Citrinae*). First genes of contigs are plotted along the bottom of the x-axis and genes at the end of
1060 each contig are plotted along the y-axis. Colored hexbins indicate the intergenic lengths of all
1061 genes with color-code indicating the frequency distribution (gene counts) according to the legend

1062 on the right. Overlaid markers indicate specific gene types corresponding to legends in the top
1063 right within each plot. Frequency distributions of specific gene types (corresponding legend
1064 color) and all other genes not of the specific type (black) are plotted along the x- and y-axis. For
1065 information on statistical test see Methods and for plots of all remaining species see Additional
1066 file 1: Figure S6.
1067
1068 **Figure 6:** Representative isolates of each *Claviceps* species showing the fraction of orthologs (y-
1069 axis) within each isolate (z-axis) at a given percent identity (x-axis) from self BLASTp results.
1070 Two *C. purpruea* s.s. isolates are shown to compare newly sequenced versus the reference.
1071
1072 **Figure 7:** Heatmap of gene counts in gene families (orthogroups) for all 53 *Claviceps* strains
1073 ordered based on ML tree in Fig. 1 and separated by sections. Gene families are separated based
1074 on their classification and are only represented once (i.e. metabolite families shown are those that
1075 are not already classified into the effector or secreted families) and are ordered based on
1076 hierarchical clustering, see Additional file 4: Table S6 for list of gene families corresponding to
1077 the order shown in the heatmaps. The host spectrum of each isolate is shown on the left side of
1078 the figure determined in Píchová *et al.* (2018) and Miao, L *et al.* (*Submitted*). For heatmap of
1079 conserved domains see Additional file 1: Figure S8 and for unclassified gene families see
1080 Additional file 1: Figure S9.

Table 1: Assembly and annotations statistics for the three reference *Claviceps* genomes and the 50 *Claviceps* genomes used in this study.

										BUSCO Completeness			Protein function					
Organism	**Strain**	**Section**	**Read Coverage**	**Genome size (Mb)**	**Contig (#)**	**N50**	**Genomic GC (%)**	**Masked repeats (%)**	**Gene count**	**Dikarya**	**Sordario-myceta**	**Conserved domains**	**MEROPs**	**CAZY -mes**	**2° meta-bolites**	**Trans-membrane**	**Secreted**	**Effectors**
References:																		
C. purpruea	20.1	Claviceps	--	32.1	1,442†	46,498†	51.6%	12.1%	8,703	95.30%	94.70%	6560	255	243	321	1114	547	199
C. fusiformis	PRL1980	Pusillae	--	52.3	6,930	19,980	37.3%	50.0%	9,304	96.70%	94.90%	6178	249	244	227	1315	521	152
C. paspali	RRC1481	Paspalorum	--	28.9	2,304	26,898	47.7%	21.8%	8,400	94.30%	93.30%	5943	235	230	185	1216	565	196
This study:																		
C. purpruea	Clav04	Claviceps	46x	31.8	3,288	21,051	51.7%	11.6%	8,824	95.50%	94.10%	6233	246	231	199	1145	557	221
C. purpruea	Clav26	Claviceps	59x	30.8	1,361	49,697	51.7%	10.4%	8,737	97.70%	96.50%	6221	263	229	255	1158	592	226
C. purpruea	Clav46	Claviceps	58x	30.8	1,409	49,302	51.7%	11.0%	8,597	98.00%	96.60%	6181	266	229	281	1151	560	198
C. purpruea	Clav55	Claviceps	59x	30.7	1,525	44,299	51.8%	11.1%	8,480	97.10%	95.90%	6113	260	231	253	1166	553	195
C. purpruea	LM4	Claviceps	64x	30.6	1,296	47,441	51.8%	11.3%	8,470	97.00%	95.80%	6132	249	233	230	1149	584	231
C. purpruea	LM5	Claviceps	67x	30.5	1,258	51,505	51.8%	10.3%	8,508	96.90%	95.50%	6179	257	231	287	1154	547	193
C. purpruea	LM14	Claviceps	49x	30.6	1,297	49,955	51.8%	11.3%	8,422	97.40%	95.60%	6132	259	227	273	1164	529	172
C. purpruea	LM28	Claviceps	49x	30.6	1,343	51,635	51.7%	10.8%	8,713	97.30%	96.10%	6260	254	233	279	1156	540	188
C. purpruea	LM30	Claviceps	64x	30.6	1,224	51,374	51.8%	10.7%	8,526	97.00%	95.50%	6167	262	229	268	1156	588	220
C. purpruea	LM33	Claviceps	45x	30.5	1,398	44,564	51.8%	10.5%	8,557	96.30%	95.50%	6193	263	226	264	1143	597	237
C. purpruea	LM39	Claviceps	81x	30.5	1,282	48,443	51.8%	11.4%	8,591	97.10%	96.10%	6156	265	232	256	1150	584	228
C. purpruea	LM46	Claviceps	79x	30.6	1,291	50,932	51.8%	10.9%	8,455	97.00%	95.80%	6125	261	225	250	1158	554	201
C. purpruea	LM60	Claviceps	81x	30.6	1,259	47,464	51.7%	10.6%	8,498	97.00%	95.80%	6126	256	231	290	1147	553	195
C. purpruea	LM71	Claviceps	168x	30.5	1,400	45,114	51.8%	10.9%	8,472	97.10%	95.60%	6095	240	221	248	1152	550	219
C. purpruea	LM207	Claviceps	53x	30.5	1,352	45,388	51.8%	10.5%	8,475	97.00%	95.70%	6120	259	220	251	1140	571	224
C. purpruea	LM223	Claviceps	74x	30.8	1,297	46,577	51.7%	11.8%	8,438	97.00%	95.70%	6076	234	230	250	1146	564	214
C. purpruea	LM232	Claviceps	53x	30.7	1,348	49,571	51.7%	10.7%	8,512	96.60%	95.70%	6169	260	232	284	1156	555	199
C. purpruea	LM233	Claviceps	49x	30.6	1,331	50,327	51.8%	11.2%	8,717	96.70%	95.90%	6294	264	230	266	1162	570	207
C. purpruea	LM461	Claviceps	37x	30.5	1,440	44,216	51.8%	9.7%	8,656	96.60%	95.20%	6223	237	231	246	1154	550	210
C. purpruea	LM469	Claviceps	75x	30.5	1,257	48,403	51.8%	11.4%	8,394	97.30%	96.00%	6106	255	229	273	1148	542	185
C. purpruea	LM470	Claviceps	26x	30.5	1,797	32,579	51.8%	10.2%	8,591	96.50%	95.30%	6195	256	223	210	1144	533	184
C. purpruea	LM474	Claviceps	64x	30.6	1,354	47,245	51.8%	10.7%	8,500	96.80%	95.70%	6118	242	222	290	1162	522	191
C. purpruea	LM582	Claviceps	89x	30.7	1,600	39,003	51.8%	10.8%	8,518	97.20%	95.40%	6132	257	226	229	1141	545	198
C. aff. purpruea	Clav52	Claviceps	60x	29.6	1,334	48,893	51.8%	9.7%	8,316	96.80%	96.20%	6078	252	225	250	1153	523	177
C. quebecensis‡	Clav32	Claviceps	64x	28.7	1,068	58,118	51.6%	7.7%	8,232	98.00%	96.60%	6057	244	226	266	1157	522	174
C. quebecensis‡	Clav50	Claviceps	59x	28.8	1,075	66,795	51.6%	8.5%	8,046	97.50%	96.30%	5986	248	228	260	1139	524	174
C. quebecensis‡	LM458	Claviceps	78x	28.4	1,166	45,693	51.6%	7.7%	8,055	97.10%	95.80%	6007	243	226	235	1135	508	154
C. occidentalis‡	LM77	Claviceps	58x	28.7	1,728	29,222	51.4%	7.9%	8,162	96.10%	94.70%	6020	243	222	186	1132	517	182
C. occidentalis‡	LM78	Claviceps	64x	28.8	1,689	29,608	51.4%	7.9%	8,231	95.80%	94.70%	6052	246	223	163	1133	513	174
C. occidentalis‡	LM84	Claviceps	164x	28.9	1,404	36,685	51.4%	8.0%	8,221	97.00%	95.40%	6088	244	223	189	1129	517	180
C. ripicola‡	LM218	Claviceps	146x	31.1	1,072	60,464	51.4%	11.7%	8,327	96.70%	95.70%	6090	249	228	255	1136	545	203
C. ripicola‡	LM219	Claviceps	55x	30.8	1,239	55,312	51.4%	10.8%	8,381	96.80%	95.80%	6133	255	226	270	1130	538	188
C. ripicola‡	LM220	Claviceps	91x	30.9	1,223	54,100	51.4%	10.6%	8,449	97.10%	95.90%	6168	249	225	255	1122	564	211
C. ripicola‡	LM454	Claviceps	156x	31.2	1,508	40,844	51.4%	12.0%	8,562	97.10%	96.10%	6163	261	224	261	1134	554	210
C. spartinae	CCC535	Claviceps	60x	29.3	1,456	42,688	51.4%	8.6%	8,433	97.50%	95.90%	6146	260	227	249	1153	504	163
C. arundinis	LM583	Claviceps	69x	30.6	996	70,672	51.4%	11.1%	8,235	96.80%	95.70%	6102	253	225	287	1140	526	186
C. arundinis	CCC1102	Claviceps	61x	30.3	896	91,905	51.4%	9.6%	8,486	97.70%	96.50%	6213	259	226	288	1130	548	195
C. humidiphila	LM576	Claviceps	77x	31.2	1,236	55,717	51.5%	11.3%	8,440	97.00%	95.90%	6147	253	227	282	1152	564	211
C. perihumidiphila‡	LM81	Claviceps	140x	31.2	1,003	67,487	51.5%	12.3%	8,291	97.10%	95.90%	6106	238	225	270	1124	522	176
C. cyperi	CCC1219	Claviceps	56x	26.6	1,921	27,113	51.7%	10.1%	7,673	97.70%	95.40%	5774	228	202	135	1072	392	97
C. capensis	CCC1504	Claviceps	66x	27.7	1,136	59,777	51.7%	7.7%	8,037	97.60%	95.70%	5989	245	225	250	1145	475	137
C. pazoutovae	CCC1485	Claviceps	61x	27.6	1,304	42,785	51.7%	8.3%	7,941	97.50%	96.00%	5954	232	223	225	1137	478	150
C. monticola	CCC1483	Claviceps	58x	27.8	1,144	56,619	51.6%	9.0%	7,977	98.10%	96.50%	5908	248	220	207	1118	481	148
C. pusilla	CCC602	Pusillae	52x	45.9	5,068	15,010	40.4%	45.0%	8,735	90.90%	88.30%	6276	252	227	76	1235	461	138
C. lovelessii	CCC647	Pusillae	53x	41.1	5,300	12,480	42.1%	37.6%	8,862	91.60%	88.20%	6351	248	228	110	1230	525	174
C. digitariae	CCC659	Pusillae	57x	33.4	1,773	32,638	44.8%	23.3%	8,285	95.90%	94.70%	6195	254	239	187	1213	513	158
C. maximensis	CCC398	Pusillae	58x	33.0	829	81,956	44.9%	23.1%	7,943	98.30%	96.50%	6100	244	235	226	1172	468	126
C. sorghi	CCC632	Pusillae	60x	35.6	3,660	16,225	44.4%	34.0%	8,208	89.90%	87.10%	6085	241	222	117	1139	425	123
C. africana	CCC489	Pusillae	56x	37.7	1,781	37,639	42.5%	37.0%	8,119	95.00%	91.50%	6057	241	221	165	1151	471	145
C. citrina	CCC265	Citrinae	64x	43.5	4,772	16,294	41.5%	53.8%	7,821	92.20%	88.20%	5879	224	207	120	1100	368	85

† The reference strain *C. purpurea* 20.1 was additionally assembled into 191 scaffolds with a scaffold N50 of 433,221.

‡ Newly identified species Miao, L (unpublished data, *Submitted*)

Table 2: Number of all unique gene pairs with an ≥ 80% identify and the proportion of these gene pairs that are located next to each other (separated by 0 genes) and separated by 5 or less genes (≤ 5 genes) for all 53 *Claviceps* genomes.

			Separation	
Species	**Strain**	**Gene pairs (#)**	**0 genes**	**≤ 5 genes**
C. purpurea	20.1	997	11.74%	30.19%
C. purpurea	Clav04	578	8.65%	11.94%
C. purpurea	Clav26	429	17.48%	30.77%
C. purpurea	Clav46	415	18.55%	29.64%
C. purpurea	Clav55	373	16.09%	26.81%
C. purpurea	LM4	426	19.95%	34.51%
C. purpurea	LM5	412	15.78%	30.83%
C. purpurea	LM14	352	19.6%	32.95%
C. purpurea	LM28	404	14.36%	25.99%
C. purpurea	LM30	352	21.88%	38.35%
C. purpurea	LM33	395	18.23%	32.66%
C. purpurea	LM39	393	17.3%	28.24%
C. purpurea	LM46	383	15.4%	29.24%
C. purpurea	LM60	374	21.39%	34.22%
C. purpurea	LM71	332	17.77%	31.02%
C. purpurea	LM207	354	17.8%	27.68%
C. purpurea	LM223	348	21.84%	34.2%
C. purpurea	LM232	424	15.33%	26.42%
C. purpurea	LM233	673	11.59%	24.37%
C. purpurea	LM461	401	14.96%	27.93%
C. purpurea	LM469	361	20.5%	32.96%
C. purpurea	LM470	410	16.34%	27.8%
C. purpurea	LM474	319	18.81%	30.72%
C. purpurea	LM582	386	13.99%	24.09%
C. aff. purpurea	Clav52	235	20.0%	31.06%
C. capensis	CCC1504	144	13.89%	21.53%
C. pazoutovae	CCC1485	182	14.29%	20.33%
C. monticola	CCC1483	174	13.22%	21.84%
C. occidentalis	LM78	173	18.5%	26.59%
C. occidentalis	LM77	151	17.88%	28.48%
C. occidentalis	LM84	431	10.9%	18.79%
C. quebecensis	LM458	176	19.32%	26.14%
C. quebecensis	Clav32	189	14.29%	24.34%
C. quebecensis	Clav50	161	14.91%	26.09%
C. ripicola	LM218	386	16.84%	31.61%
C. ripicola	LM219	393	15.78%	28.5%
C. ripicola	LM220	412	16.02%	28.64%
C. ripicola	LM454	434	13.13%	21.43%
C. spartinae	CCC535	251	10.36%	16.33%
C. arundinis	CCC1102	431	11.6%	22.51%
C. arundinis	LM583	362	15.19%	26.24%
C. humidiphila	LM576	401	14.96%	23.44%
C. perihumidiphila	LM81	351	19.66%	33.62%
C. cyperi	CCC1219	193	5.7%	7.77%
C. pusilla	CCC602	9	0.0%	0.0%
C. fusiformis	PRL 1980	4	0.0%	0.0%
C. lovelessii	CCC647	7	0.0%	0.0%
C. digitariae	CCC659	10	0.0%	0.0%
C. maximensis	CCC398	3	0.0%	0.0%
C. sorghi	CCC632	12	0.0%	8.33%
C. africana	CCC489	8	0.0%	0.0%
C. citrina	CCC265	24	4.17%	4.17%
C. paspali	RRC 1481	1	0.0%	0.0%

Fig. 1

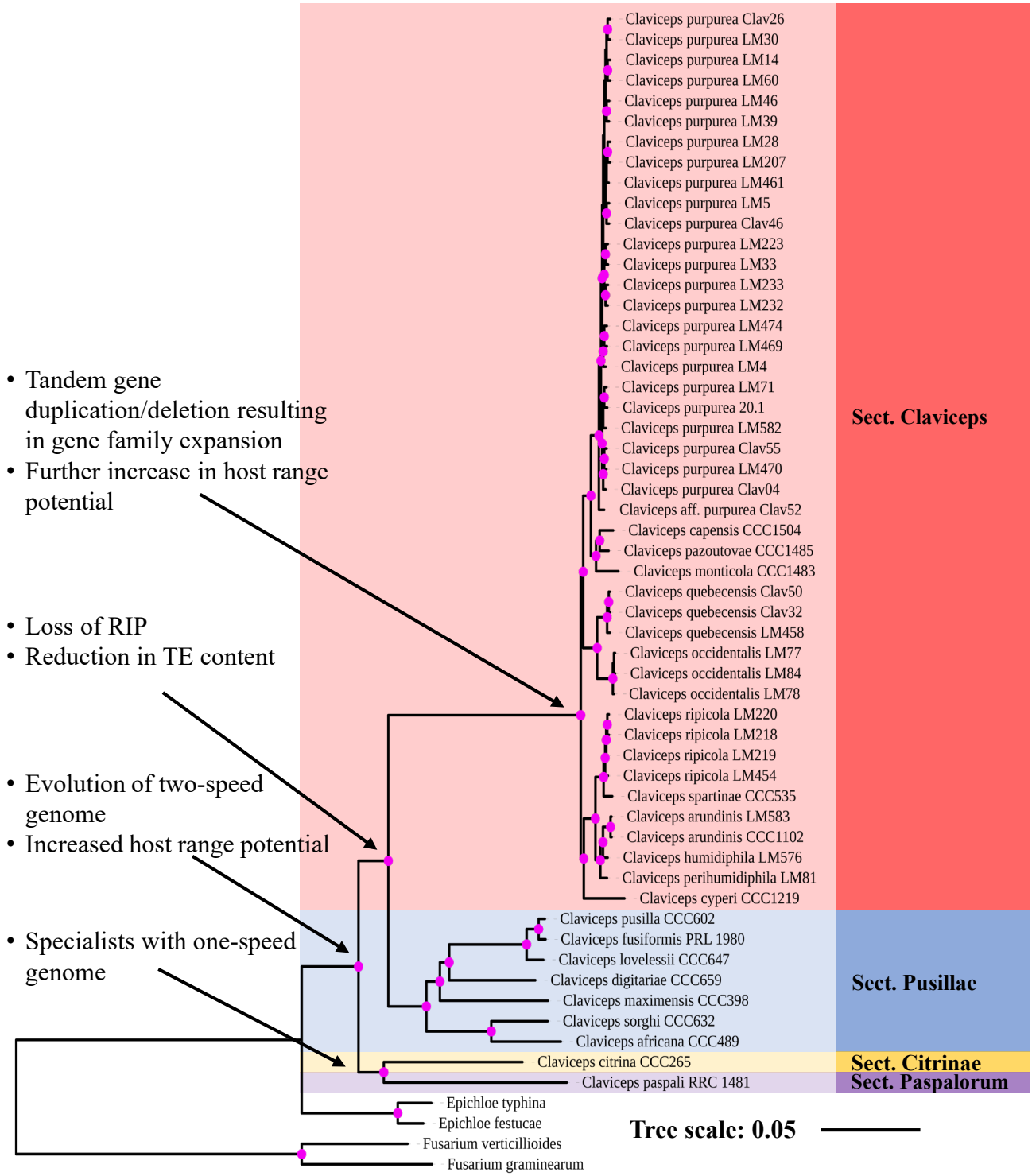

Claviceps purpurea Clav26
Claviceps purpurea LM30
Claviceps purpurea LM14
Claviceps purpurea LM60
Claviceps purpurea LM46
Claviceps purpurea LM39
Claviceps purpurea LM28
Claviceps purpurea LM207
Claviceps purpurea LM461
Claviceps purpurea LM5
Claviceps purpurea Clav46
Claviceps purpurea LM223
Claviceps purpurea LM33
Claviceps purpurea LM233
Claviceps purpurea LM232
Claviceps purpurea LM474
Claviceps purpurea LM469
Claviceps purpurea LM4
Claviceps purpurea LM71
Claviceps purpurea 20.1
Claviceps purpurea LM582
Claviceps purpurea Clav55
Claviceps purpurea LM470
Claviceps purpurea Clav04
Claviceps aff. purpurea Clav52
Claviceps capensis CCC1504
Claviceps pazoutovae CCC1485
Claviceps monticola CCC1483
Claviceps quebecensis Clav50
Claviceps quebecensis Clav32
Claviceps quebecensis LM458
Claviceps occidentalis LM77
Claviceps occidentalis LM84
Claviceps occidentalis LM78
Claviceps ripicola LM220
Claviceps ripicola LM218
Claviceps ripicola LM219
Claviceps ripicola LM454
Claviceps spartinae CCC535
Claviceps arundinis LM583
Claviceps arundinis CCC1102
Claviceps humidiphila LM576
Claviceps perihumidiphila LM81
Claviceps cyperi CCC1219
Claviceps pusilla CCC602
Claviceps fusiformis PRL 1980
Claviceps lovelessii CCC647
Claviceps digitariae CCC659
Claviceps maximensis CCC398
Claviceps sorghi CCC632
Claviceps africana CCC489
Claviceps citrina CCC265
Claviceps paspali RRC 1481
Epichloe typhina
Epichloe festucae
Fusarium verticillioides
Fusarium graminearum
Sect. Claviceps
Sect. Pusillae
Sect. Citrinae
Sect. Paspalorum
Tree scale: 0.05
• Tandem gene duplication/deletion resulting in gene family expansion
• Further increase in host range potential
• Loss of RIP
• Reduction in TE content
• Evolution of two-speed genome
• Increased host range potential
• Specialists with one-speed genome

Fig. 2

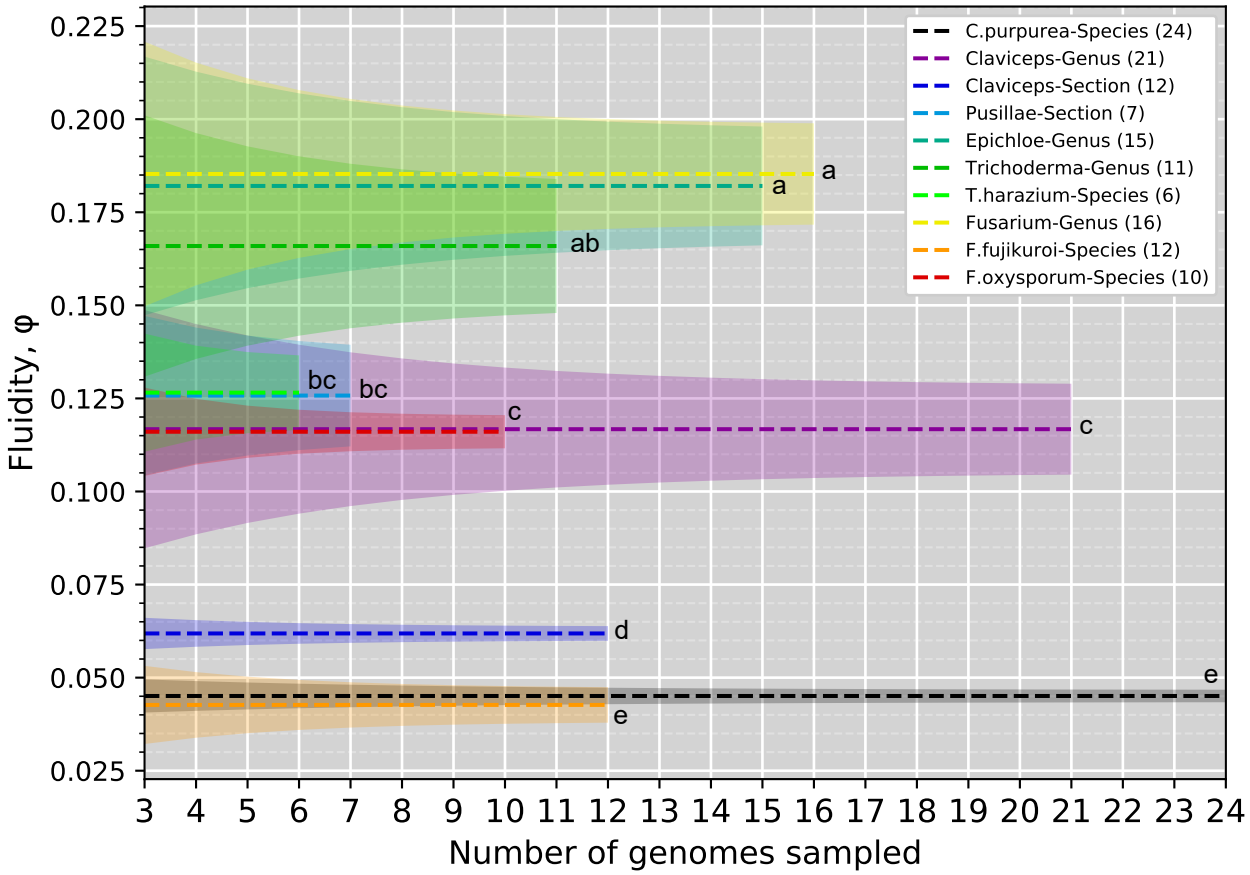

C.purpurea-Species (24)
Claviceps-Genus (21)
Claviceps-Section (12)
Pusillae-Section (7)
Epichloe-Genus (15)
Trichoderma-Genus (11)
T.harazium-Species (6)
Fusarium-Genus (16)
F.fujikuroi-Species (12)
F.oxysporum-Species (10)
Fluidity, φ
0.225
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
a
a
ab
bc
bc
c
c
d
e
e
3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Number of genomes sampled

Fig. 3

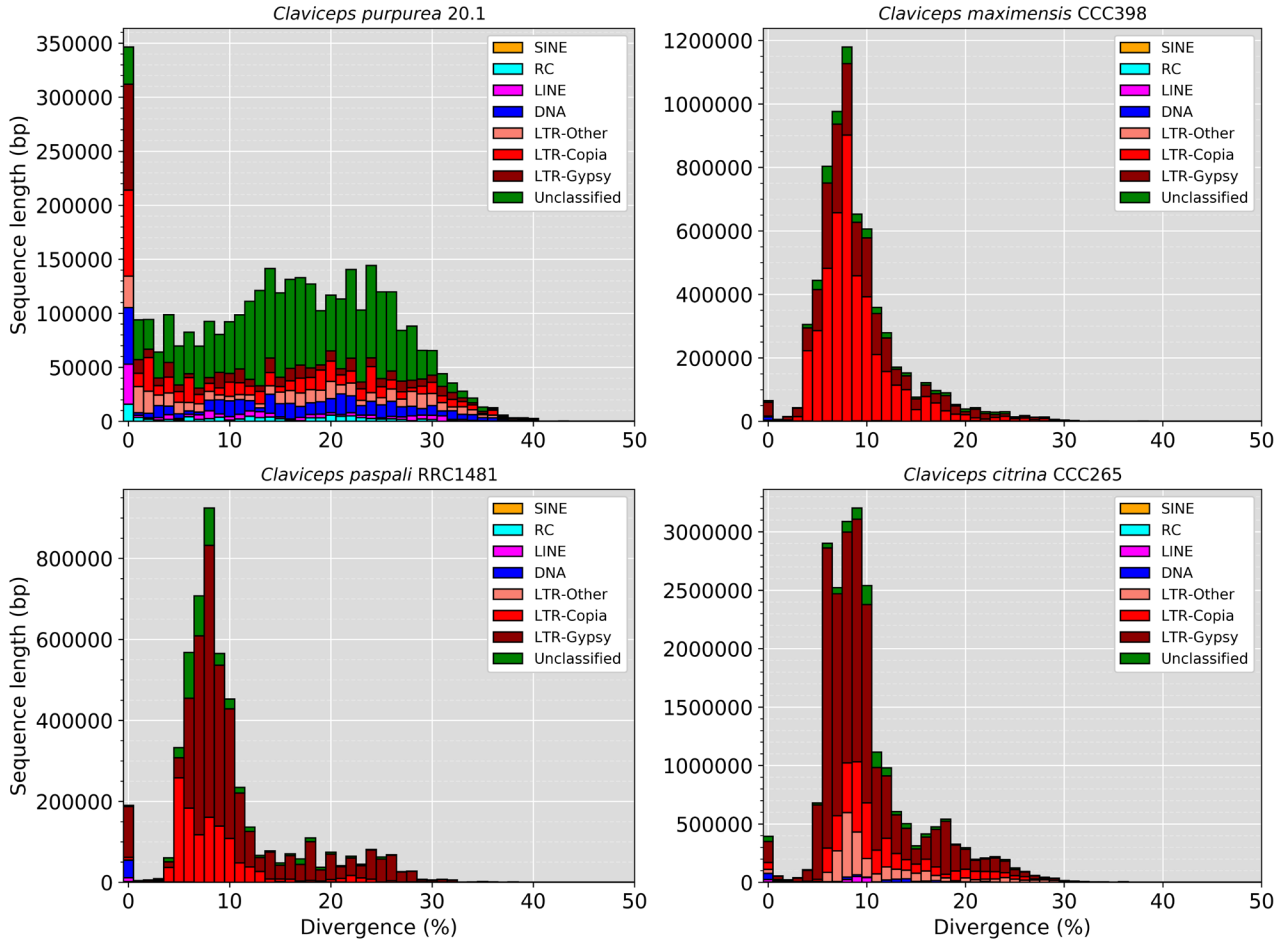

Claviceps purpurea 20.1
Claviceps maximensis CCC398
Claviceps paspali RRC1481
Claviceps citrina CCC265
SINE
RC
LINE
DNA
LTR-Other
LTR-Copia
LTR-Gypsy
Unclassified
Sequence length (bp)
Divergence (%)

Fig. 4

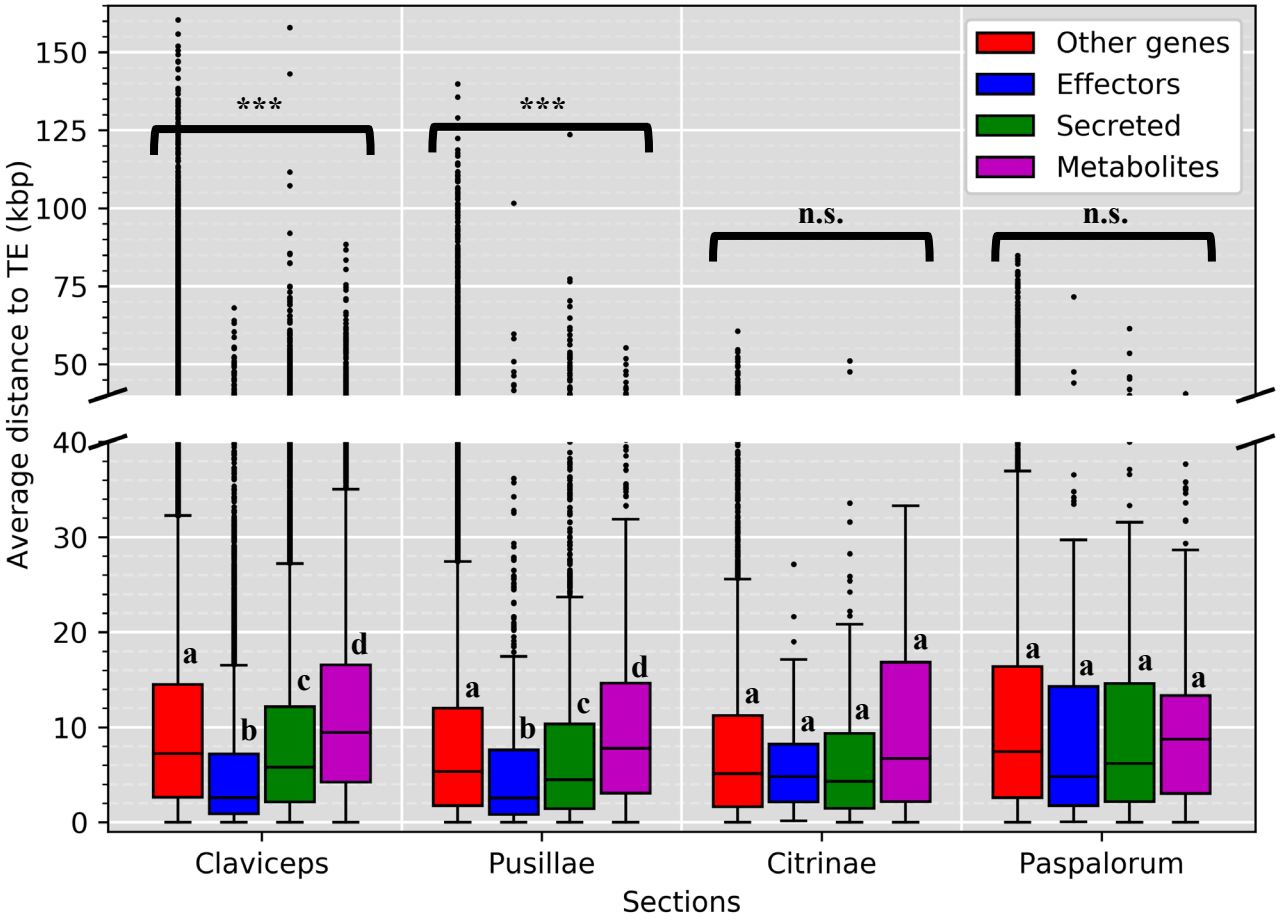

Other genes
Effectors
Secreted
Metabolites
***
***
n.s.
n.s.
Average distance to TE (kbp)
150
125
100
75
50
40
30
20
10
0
a
b
c
d
a
b
c
d
a
a
a
a
a
a
a
a
Claviceps
Pusillae
Citrinae
Paspalorum
Sections

Fig. 5

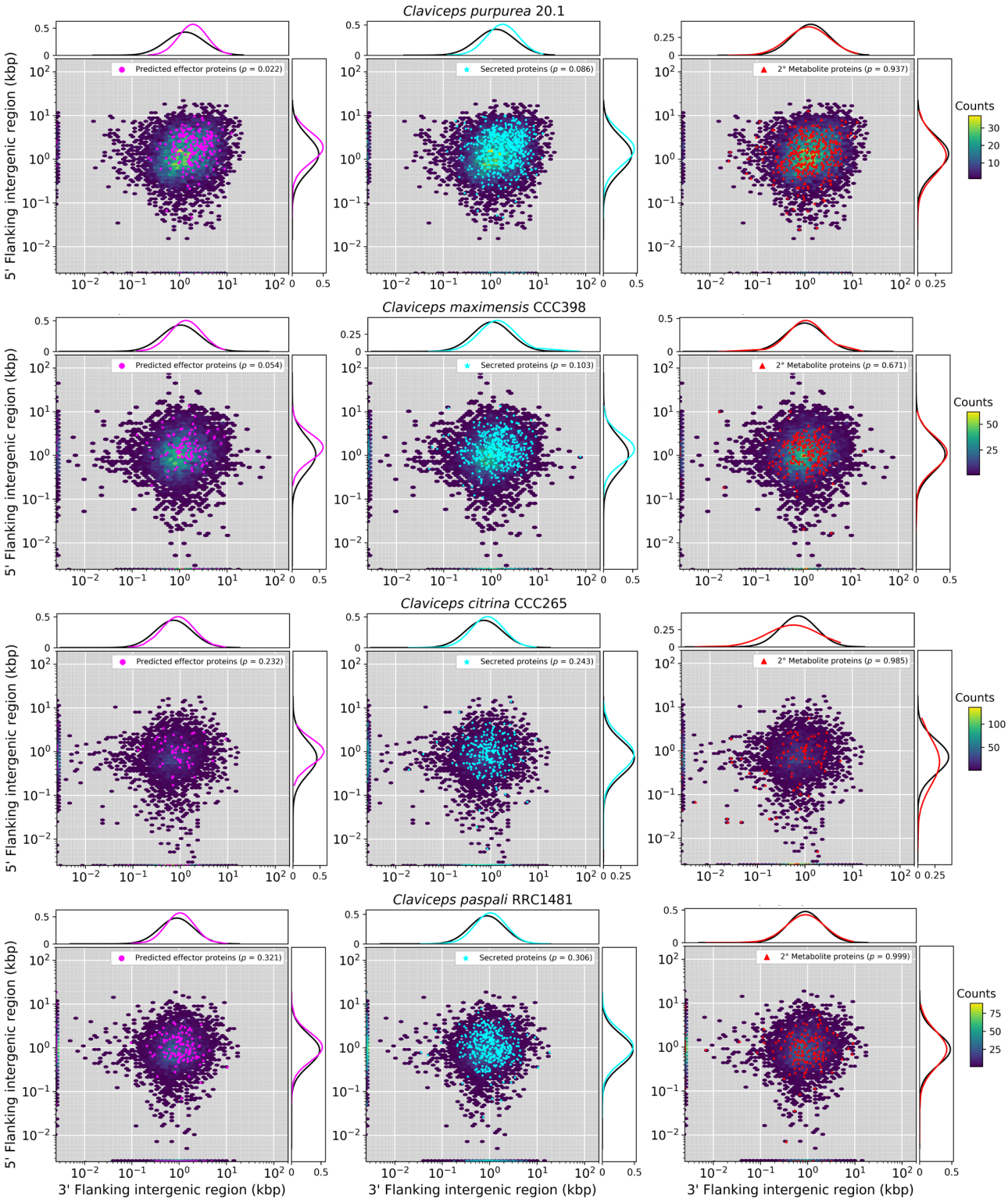

Claviceps purpurea 20.1
Predicted effector proteins (p = 0.022)
Secreted proteins (p = 0.086)
2° Metabolite proteins (p = 0.937)
Claviceps maximensis CCC398
Predicted effector proteins (p = 0.054)
Secreted proteins (p = 0.103)
2° Metabolite proteins (p = 0.671)
Claviceps citrina CCC265
Predicted effector proteins (p = 0.232)
Secreted proteins (p = 0.243)
2° Metabolite proteins (p = 0.985)
Claviceps paspali RRC1481
Predicted effector proteins (p = 0.321)
Secreted proteins (p = 0.306)
2° Metabolite proteins (p = 0.999)
Counts
5' Flanking intergenic region (kbp)
3' Flanking intergenic region (kbp)

Fig. 6

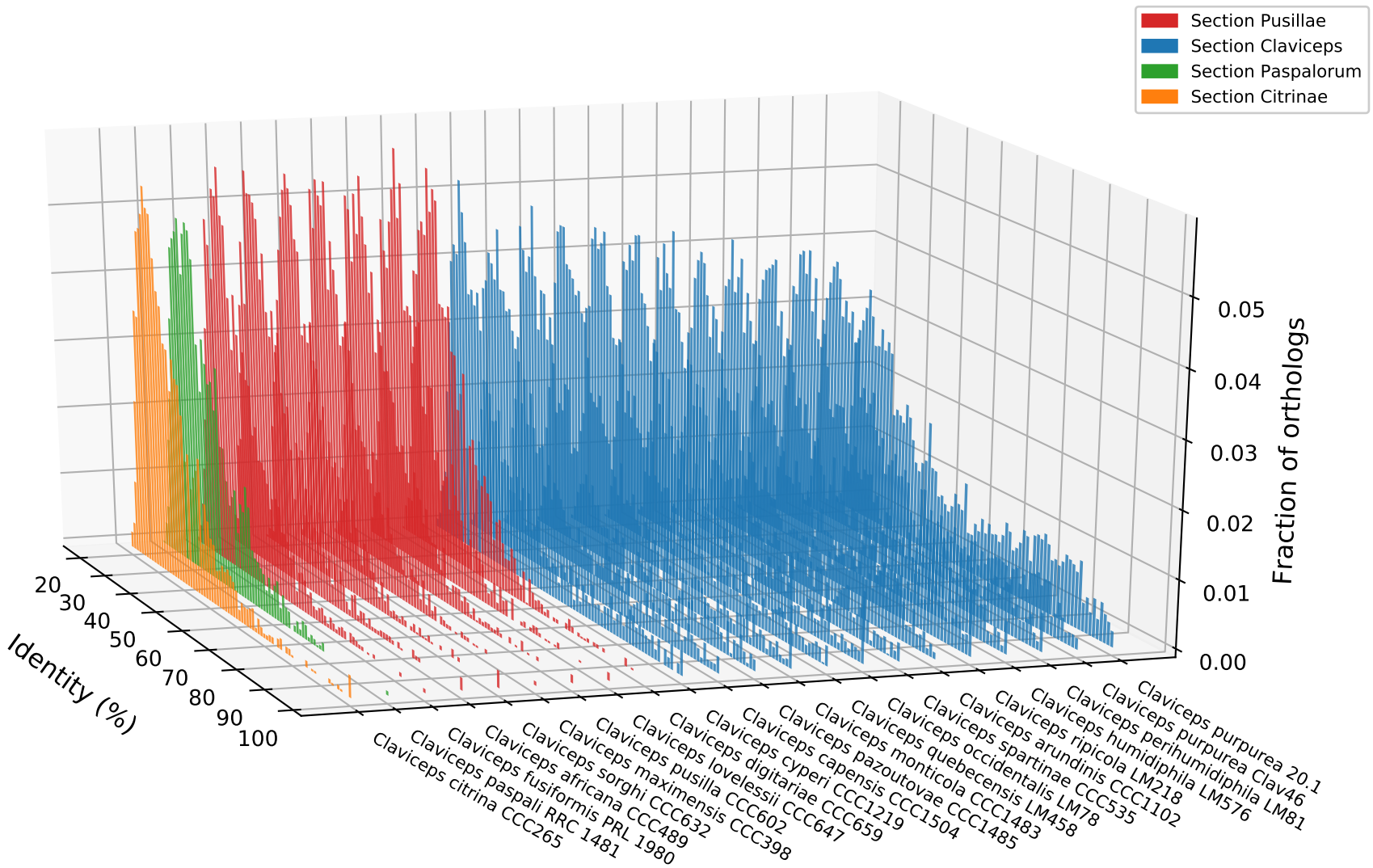
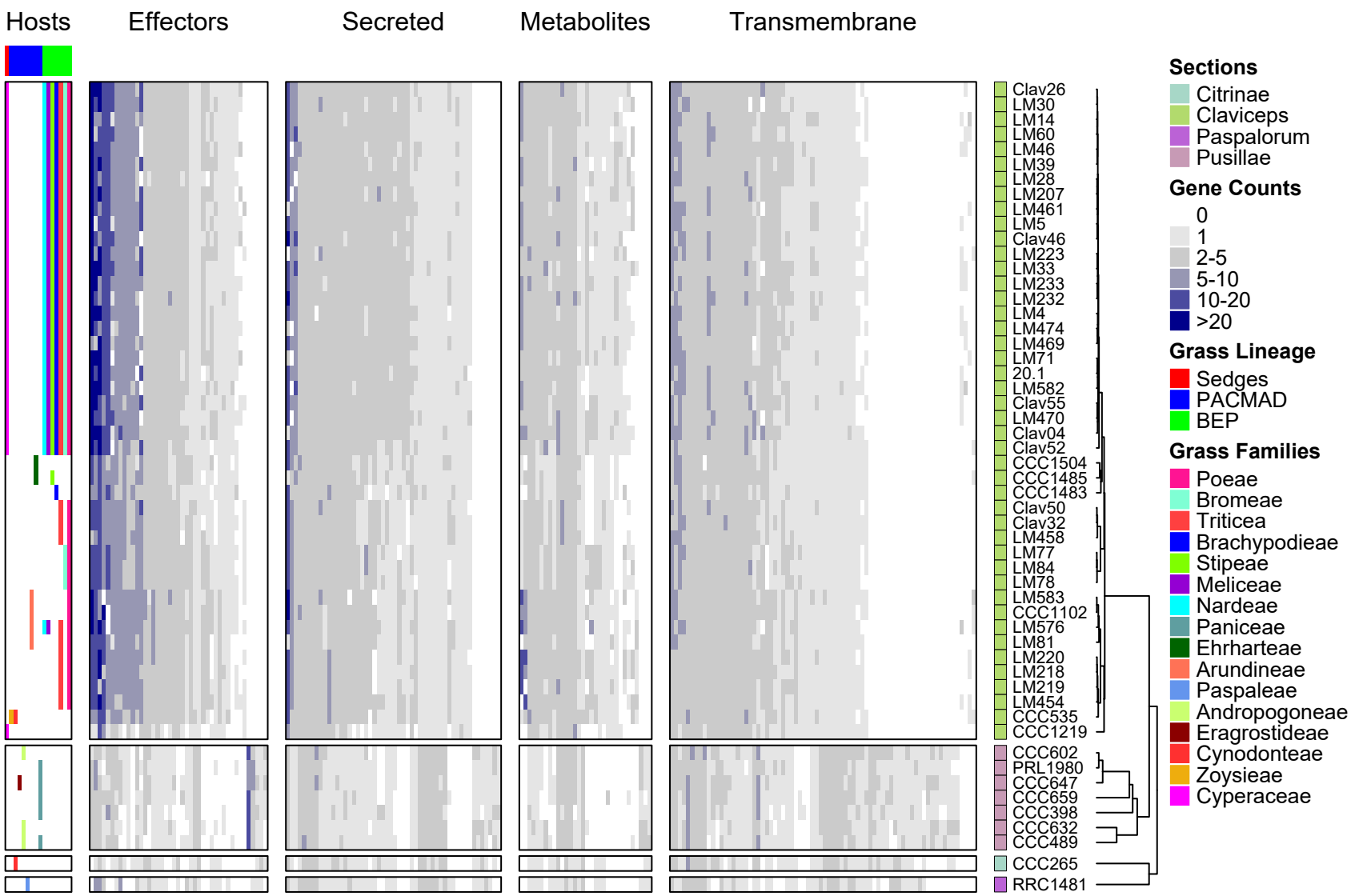

Section Pusillae
Section Claviceps
Section Paspalorum
Section Citrinae
Fraction of orthologs
0.00
0.01
0.02
0.03
0.04
0.05
Identity (%)
20
30
40
50
60
70
80
90
100
Claviceps purpurea 20.1
Claviceps purpurea Clav46
Claviceps perihumidiphila LM576
Claviceps humidiphila LM81
Claviceps ripicola LM218
Claviceps arundinis CCC1102
Claviceps spartinae CCC535
Claviceps occidentalis LM78
Claviceps quebecensis LM458
Claviceps monticola CCC1483
Claviceps pazoutovae CCC1504
Claviceps capensis CCC1485
Claviceps cyperi CCC1219
Claviceps digitariae CCC659
Claviceps lovelessii CCC647
Claviceps pusilla CCC602
Claviceps maximensis CCC398
Claviceps sorghi CCC632
Claviceps africana CCC489
Claviceps fusiformis PRL 1980
Claviceps paspali RRC 1481
Claviceps citrina CCC265

Fig. 7


Hosts
Effectors
Secreted
Metabolites
Transmembrane
Clav26
LM30
LM14
LM60
LM46
LM39
LM28
LM207
LM461
LM5
Clav46
LM223
LM33
LM233
LM232
LM4
LM474
LM469
LM71
20.1
LM582
Clav55
LM470
Clav04
Clav52
CCC1504
CCC1485
CCC1483
Clav50
Clav32
LM458
LM77
LM84
LM78
LM583
CCC1102
LM576
LM81
LM220
LM218
LM219
LM454
CCC535
CCC1219
CCC602
PRL1980
CCC647
CCC659
CCC398
CCC632
CCC489
CCC265
RRC1481
Sections
Citrinae
Claviceps
Paspalorum
Pusillae
Gene Counts
0
1
2-5
5-10
10-20
>20
Grass Lineage
Sedges
PACMAD
BEP
Grass Families
Poeae
Bromeae
Triticea
Brachypodieae
Stipeae
Meliceae
Nardeae
Paniceae
Ehrharteae
Arundineae
Paspaleae
Andropogoneae
Eragrostideae
Cynodonteae
Zoysieae
Cyperaceae