1   Title: Mapping the sex determination region in the *Salix* F$_1$ hybrid common parent population

2   confirms a ZW system in six diverse species

3

4   Authors: Dustin G. Wilkerson[1], Bircan Taskiran[1,2], Craig H. Carlson[1,3], and Lawrence B. Smart[1]

5

6   Affiliations: [1]Horticulture Section, School of Integrative Plant Sciences, Cornell University,

7   Cornell AgriTech, Geneva, NY 14456

8   Current Affiliations: [2]Biology Department, Faculty of Science, Cankiri Karatekin University,

9   Cankiri, Turkey; [3]USDA-ARS, Cereal Crops Research Unit, Edward T. Schafer Agricultural

10   Research Center, Fargo, ND 58102

11

12   Short Running Title: ZW SDR in *Salix* F$_1$ families

13

14   Key Words: QTL Mapping; *Salix*; sex determination; linkage map; *Salix purpurea*; *Salix viminalis*;

15   *Salix suchowensis*

16

17   Corresponding Author: Lawrence Smart; 630 W. North Street, 111 Surge Lab, Geneva, NY

18   14456; (315) 787-2490; lbs33@cornell.edu

19

ABSTRACT

20

21   Within the genus *Salix*, there are approximately 350 species native primarily to the

22   northern hemisphere and adapted to a wide range of habitats. This diversity can be exploited to

23   mine novel alleles conferring variation important for production as a bioenergy crop, but also to

24   identify evolutionarily important genes, such as those involved in sex determination. To

25   leverage this diversity, we created a mapping population by crossing six *Salix* species (*S.*

26   *viminalis*, *S. suchowensis*, *S. integra*, *S. koriyanagi*, *S. udensis*, and *S. alberti*) to common male

27   and female *S. purpurea* parents. Each family was genotyped via genotyping-by-sequencing and

28   assessed for kinship and population structure as well as the construction of 16 backcross

29   linkage maps to be used as a genetic resource for breeding and selection. Analyses of

30   population structure resolved both the parents and F_1 progeny to their respective phylogenetic

31   section and indicated that the *S. alberti* parent was misidentified and was most likely *S.*

32   *suchowensis*. Sex determining regions were identified on *Salix* chromosome 15 in the female-

33   informative maps for seven of the eight families indicating that these species share a common

34   female heterogametic ZW sex system. The eighth family, *S. integra* × *S. purpurea*, was entirely

35   female and had a truncated chromosome 15. Beyond sex determination, the *Salix* F_1 hybrid

36   common parent population (*Salix F₁* HCP) introduced here will be useful in characterizing

37   genetic factors underlying complex traits, aid in marker-assisted selection, and support genome

38   assemblies for this promising bioenergy crop.

39   INTRODUCTION

40   The establishment of genomic resources is an important step in developing a fully

41   realized breeding program, reinforced by modern tools for trait mapping, candidate gene

42   identification, and marker-assisted selection. *Salix* and *Populus* (poplars) comprise the majority

43   of species in the family Salicaceae, which consists of dioecious trees, shrubs, and subshrubs that

44   are highly heterozygous. Shrub willow (*Salix* spp.) are grown in northern latitudes as a

45   sustainable, high-yielding, carbon neutral, bioenergy crop that can grow on marginal land and

46   provide multiple ecosystem services (Smart et al. 2005; Stoof et al. 2014; Clifton-Brown et al.

47   2019; Fabio and Smart 2020). While shrub willow breeding in the United States has been active

48   since the 1980's, most of the nearly 350 species have yet to be tapped as a source of genetic

49   diversity (Dickmann and Kuzovkina 2014; Stanton et al. 2014).

50   Genomic resources developed for poplar were used in early genomic studies in *Salix*,

51   because their genomes are largely co-linear (Hanley et al. 2006; Berlin et al. 2010). As

52   sequencing technologies became increasingly more affordable, more *Salix* genome sequencing

53   was completed and now there are several high-quality assemblies. Genomic resources for *Salix*

54   are currently centered around a few key species. In Europe, *Salix viminalis* is an important

55   bioenergy crop species with a recently published, high quality genome assembly (Almeida et al.

56   2020). *S. viminalis* has been used in several QTL mapping studies for resistance to willow leaf

57   rust (*Melampsora larici-epitea*) (Rönnberg-Wästljung et al. 2008; Samils et al. 2011; Sulima et

58   al. 2017), drought tolerance (Rönnberg-Wästljung et al. 2005), and growth and phenology

59   (Hallingbäck et al. 2016; Hallingbäck et al. 2019). While in the United States, *Salix purpurea* is

60   the model species for bioenergy willow breeding, genetics, and genomics. The US Department

61   of Energy Joint Genome Institute has produced the highest quality annotated *Salix* reference

62   genomes assembled in the genus on male and female *S. purpurea*, available on Phytozome

63   (Zhou et al. 2018; Zhou et al. 2020) (https://phytozome-

64   next.jgi.doe.gov/info/SpurpureaFishCreek_v3_1; https://phytozome-

65   next.jgi.doe.gov/info/Spurpurea_v5_1). Using joint linkage and association mapping

66   approaches focused on *S. purpurea*, Carlson et al. (2019) identified numerous QTL for a wide

67   range of morphological, physiological, insect and disease resistance and biomass composition

68   traits. A naturalized species in North America, *S. purpurea* is a potential donor of broad

69   adaptability traits for species susceptible to pests and diseases in the Northeast United States.

70   There is value in studying the genomes of less-characterized *Salix* species for phylogenomic

71   analysis and to discover diverse sources of alleles for introgression into elite yielding cultivars.

72   Here, we introduce the *Salix* F_1 hybrid common parent population (*Salix* F_1 HCP). The

73   parents, described in Fabio et al. (2019) and Crowell et al. (2020), represent a diverse selection

74   of species from *Salix* subgenus Vetrix (Dickmann and Kuzovkina 2014). Six *Salix* species (*S.*

75   *viminalis*, *S. suchowensis*, *S. integra*, *S. koriyanagi*, *S. udensis*, and *S. alberti*) were crossed to

76   common male and female *S. purpurea* parents to form eight species hybrid families. Literature

77   describing these species ranges from the high-quality reference genomes available for *S.*

78   *purpurea* and *S. viminalis* to the scarcely studied *S. alberti* (Rosso et al. 2013). *Salix suchowensis*

79   is native to China and has been used recently to generate a chromosome scale genome

80   assembly (Wei et al. 2020). This species has been assessed for its response to drought stress (Jia

81   et al. 2020) and was one of the first *Salix* species used to map the sex determination region

82   (SDR) (Liu et al. 2013). *Salix udensis*, formally known as *S. sachalinensis*, has been described as a

83   Japanese riparian willow species that acts as a natural nest cavity for fish owls (Niiyama 2008;

4

84   Slaght et al. 2018) and is suggested to have sexually dimorphic characteristics (Ueno and Seiwa

85   2003; Ueno et al. 2006). Some genomic resources are available for the Korean *S. koriyanagi*, as

86   its chloroplast genome have been sequenced (Kim et al. 2019; Park et al. 2019) while *S. integra*

87   has been assessed for its phytoremediation potential (Cao et al. 2020; Yang et al. 2020b).

88   Developed to interrogate the genetics of several understudied *Salix* species, the *Salix* F_1 HCP is

89   an important step in the development of genomic resources in *Salix*.

90   The Salicaceae represents an interesting family for the study of the evolution of dioecy

91   and the mechanisms of sex determination. Genetic mapping of SDRs in the Salicaceae have

92   revealed considerable intra- and inter-chromosomal variability and contrasting sex

93   determination systems (Yang et al. 2021). While *P. alba* and *P. trichocarpa* are male

94   heterogametic (XY) with SDRs located on chr19, the SDR of *P. euphratica* was identified on

95   chr14 and is female heterogametic (ZW) (Paolucci et al. 2010; Geraldes et al. 2015; Yang et al.

96   2021). Conversely, the SDR of *S. purpurea* has been mapped to a large, pericentromeric region

97   on *Salix* chr15, but has been found to maintain a few orthologous regions present within the *P.*

98   *trichocarpa* chr19 SDR (Zhou et al. 2018). However, it is unclear whether dioecy evolved before

99   or after their divergence (Hou et al. 2015). The sex systems of *Salix* also lack a consensus model.

100   For instance, similar to *S. purpurea*, both *S. suchowensis* (Hou et al. 2015; Chen et al. 2016) and

101   *S. viminalis* (Pucholt et al. 2015; Pucholt et al. 2017b) have female heterogametic (ZW) systems

102   with SDRs located on *Salix* chr15, yet tree-form *S. nigra* has an XY system SDR on chr07

103   (Sanderson et al. 2021). Considering the variability in sex determination systems and SDR

104   locations already discovered within *Salix*, elucidating the mechanisms of sex determination in

105   more species could help to build a more cohesive understanding of SDR evolution. Further

106   investigation could provide evidence to implicate SDR turnovers as a factor in the considerable

107   species diversity within the genus.

108   To establish the *Salix* F_1 HCP as a genetic resource, this study sought to: (1) describe the

109   genetic relationships between the parent species, (2) develop female and male parent

110   informative linkage maps for each family, and (3) define and compare their respective SDRs.

<div align="center">MATERIALS AND METHODS</div>

Population Development

113   Females of *S. viminalis*, *S. integra*, *S. alberti*, and *S. suchowensis* were crossed with male

114   *S. purpurea* clone ID 94001, while males of *S. viminalis*, *S. udensis*, *S. koriyanagi*, and *S.*

115   *suchowensis* were crossed with female *S. purpurea* clone ID 94006 creating eight F_1 species

116   hybrid families (Fig. 1). The common *S. purpurea* parents, 94006 and 94001, were chosen based

117   upon their differential resistance to willow leaf rust (Crowell et al. 2020), adaptation to the

118   Northeastern U.S., and availability of high-quality reference genomes (Zhou et al. 2020).

119   Crosses were made by forcing floral catkins of the parent genotypes from dormant

120   shoots in a greenhouse. All crosses were made in isolation to prevent pollen contamination.

121   When anthers began to dehisce, male catkins were excised and placed in falcon tubes for pollen

122   extraction using toluene, as described in Kopp et al. (2002), then stored in 2 mL microcentrifuge

123   tubes at –20°C until female catkins were receptive. Seedlings were established in a standard

124   peat-based potting mix in a greenhouse, then transplanted to nursery beds near Cornell

125   AgriTech (Geneva, NY). Winter dormant cuttings from all parents and progeny were collected

126   from one-year old stems and hand-planted in the field in a randomized complete block design

127 with three plants per plot and four replicate blocks. Field trials were established with 1.83 m

128 spacing between rows and 40.6 cm between plants within rows.

DNA Extraction and Genotyping-by-Sequencing

130 Shoot tips for DNA extraction were collected from plants in nursery beds and stored in

131 desiccant. Dried shoot tips were ground to a fine powder with a Geno/Grinder (SPEX

132 SamplePrep, Metuchen, NJ, USA) prior to genomic DNA extraction using the DNeasy Plant Mini

133 Kit (QIAGEN Inc., Valencia, CA, USA). After checking the DNA quality using gel electrophoresis,

134 DNA quantity was estimated using the Invitrogen Qubit dsDNA Broad Range Assay kit on a

135 Qubit Fluorometer (ThermoFisher Scientific, Waltham, MA). Genomic DNA was submitted to

136 the University of Wisconsin Biotechnology Center (Madison, WI) for 96-plex Genotyping-by-

137 Sequencing (GBS) library preparation using the *ApeK*I restriction enzyme and sequenced (1 ×

138 100 bp) on the Illumina HiSeq 2500 (Illumina, Inc., San Diego, CA, USA) platform.

Variant Discovery and Imputation

140 Initial variant discovery and filtering was performed using the TASSEL GBS v2 Discovery

141 Pipeline (Bradbury et al. 2007). Sequence reads were trimmed to 64 bp and aligned using BWA

142 mem (Li and Durbin 2009) under default parameters to the *S. purpurea* v5.1 reference genome

143 (Zhou et al. 2020). As this genome includes both chr15Z and 15W, chr15Z was excluded to

144 reduce mapping errors within the pseudoautosomal regions surrounding the SDR in *S.*

145 *purpurea*. This process was repeated once for all eight families together and then for each of

146 the eight F_1 families separately to identify population-wide and family-specific SNPs. By calling

147 variants on all eight families and then on each family separately, variants were based on inter-

7

148   and intra- familial genetic differences, respectively. The resulting VCF files contained 684,412

149   SNPs for the full analysis and ranged from 174,762 to 266,797 SNPs depending on the family.

150   On the eight F_1, SNPs with > 70% missing data and minor allele frequency < 0.01 and F_1

151   individuals were dropped if they had > 80% missing data or determined to be outliers based on

152   principal component analysis. Missing genotype calls and low read depth are common in GBS

153   (Elshire et al. 2011), therefore imputation was performed separately on each family to validate

154   calls in haplotype blocks. Using LinkImputeR (Money et al. 2017), genotypes called with a read

155   depth < 5 were set to missing before filtering again for missingness > 70%, resulting in 95,281 to

156   145,944 imputed SNPs with accuracies ranging from 84.3 to 93.1%.

157   **Population Structure**

158   SNPs were filtered to retain markers and individuals with ≤ 20% missing data and SNPs

159   with a minor allele frequency > 0.01, which resulted in 55,398 SNPs. Analysis of principal

160   components and kinship using default parameters were performed in Tassel 5 and visualized in

161   R (R Core Team 2020). Multiple runs of the parents and 10 randomly selected F_1 progeny were

162   analyzed using fastSTRUCTURE (Raj et al. 2014). Only a subset of the F_1 were used in this

163   analysis in order to manage file size and computation requirements. Multiple analyses of

164   'structure.py' were completed (K = 3 - 10). Using 'chooseK.py', K = 6 represented the model

165   complexity that maximized the marginal likelihood and best explained the data structure,

166   suggesting six separate populations.

167   **Linkage Map Construction and Analysis**

8

168   Unless otherwise indicated, linkage map construction and analysis were performed

169   using custom R code, available on Github (link in Data Availability). Multiple runs of each parent

170   were used to form consensus genotypes for each SNP. In the absence of a clear consensus, the

171   genotype was set to missing. If both parents were set to missing, the SNP was removed from

172   the analysis. If only one parent was set to missing, its genotype was inferred based on the

173   genotype of the known parent and the segregation of the F_1. These parental consensus

174   genotypes were used to identify the female informative (AB × AA) and male informative (AA ×

175   AB) markers used to generate backcross linkage maps for each parent using a combination of

176   R/qtl (Broman et al. 2003) and ASMap (Taylor and Butler 2017). Co-located markers and those

177   exhibiting extreme segregation distortion were removed using ASMap function 'pullCross'.

178   Linkage groups were then created using 'mstmap' with default parameters except for

179   objective.fun = "ML" and bychr = FALSE. The p-value for determining linkage groups varied

180   between families, ranging from 1e-6 to 1e-12, depending on the demarcation of the 19

181   expected linkage groups. A custom R function was then used to perform simple error correction

182   to reduce the number of double crossovers and deflate map distances before reforming linkage

183   groups. Briefly, this function relies on the marker order determined after formation of linkage

184   groups, identifying double crossovers of ≤ two SNPs, and correcting them. Map quality was

185   checked using two strategies. The 'heatMap' function in ASMap, which plots LOD linkage

186   between markers on the upper triangle and estimated recombination frequency on the bottom,

187   reveals markers that are problematic or out of phase. Then by comparing the physical position

188   (Mb) based on alignment to the *S. purpurea* reference genome and the genetic distance (cM)

189   within each linkage group shows issues with marker order or potential chromosomal

190   rearrangements.

191   To delimit the SDR in each family, the sex of all F_1 individuals was recorded by inspecting

192   flowering catkins over two growing seasons in the field. Sex ratio bias was tested using a chi-

193   square test for a 1:1 sex ratio. Using functions in R/qtl, genotype probabilities were calculated

194   using 'calc.genoprob' (step = 0, error.prob = 0.01, map.function = "kosambi", stepwidth =

195   "fixed"). Next, QTL mapping for individual sex was performed using 'scanone': model =

196   "binary", method = "em". Genome-wide significance thresholds were determined based on the

197   results of 'scanone.perm' (n.perm = 1000). QTL positions were refined using 'refineqtl' (method

198   = "hk", model = "binary"), then the 1.5 LOD support intervals were calculated using 'lodint' with

199   default parameters.

200   RESULTS

Population Structure

202   A combination of PCA, hierarchical clustering, and fastSTRUCTURE were used to

203   describe the population structure of the *Salix* F_1 HCP. A PCA of the parents revealed three

204   distinct clusters, formed by two PCs accounting for 36.8% and 22.5% of the total genetic

205   variation (Fig. 2A). The two common *S. purpurea* parents, 94006 and 94001, formed a single

206   group and were separated from the other parents by PC2. The *S. viminalis* parents, 'Jorr' and

207   07-MBG-5027, formed a cluster with *S. udensis* 04-BN-051, which was differentiated from the

208   remaining species by PC1. Including the F_1 progeny from each family into the analysis, each PC

209   accounted for 26% and 10.1% of the total variation (Fig. 2B). PC2 split *S. udensis* from the two *S.*

                                                                                        10

210   *viminalis* parents and separated *S. koriyanagi* from *S. suchowensis*, *S. integra*, and *S. alberti*. As

211   expected, the F_1 individuals were intermediate between the common parent and family specific

212   species. The F_1 progeny derived from the female parents *S. suchowensis* P295, *S. alberti* P294,

213   and *S. integra* P336 (all crossed with male *S. purpurea* 94001), co-localized.

214   Kinship analysis grouped the two *S. viminalis* parents with *S. udensis*, the two *S.*

215   *purpurea* parents together, and *S. koriyanagi* with *S. integra* (Fig. 2C). *Salix suchowensis* and the

216   *S. viminalis* and *S. udensis* group were the least similar while *S. purpurea* grouped centrally

217   between them, similar to PC1 (Fig. 2A). Within *S. suchowensis*, the female P295 was more

218   closely related to the female *S. alberti* P294 than it was to P63, a male. Admixture analysis

219   included the parents and a subset of the F_1 individuals of each family. The eight groups of F_1

220   individuals were comprised of roughly half the genetic background of *S. purpurea* and half the

221   other species parent, as expected (Fig. 2D). *Salix viminalis*, *S. integra*, *S. koriyanagi*, *S. udensis*

222   and *S. purpurea* formed distinct clusters, while *S. alberti* P294 grouped together with *S.*

223   *suchowensis*.

Linkage Map Construction and Analysis

225   Since there is considerable divergence between parents and pedigrees, variant

226   discovery and marker filtration were performed for each family separately. Using consensus

227   genotypes derived from multiple sequencing runs of the parents, markers were split into

228   female (AB × AA) and male (AA × AB) informative backcross markers for linkage map

229   construction, which resulted in 16 linkage maps (Fig. 3). Each linkage map consisted of 19

230   linkage groups with total map lengths ranging from 3939.9 – 6957.3 cM containing between

231   2035 and 3852 total markers (Table S1). Recombination frequency and genetic to physical

232   distance plots generated for each linkage map revealed that marker order and phase within

233   each linkage group were reasonably linear (Fig. S1). Sex phenotypes were used for QTL mapping

234   of the SDR. Six of the eight families displayed significant sex ratio bias towards females based

235   on a simple chi-square test (p < 0.05), with female to male ratios ranging from 1.4 to 1.6 (Table

236   1). Neither of the *S. viminalis* families displayed sex ratio bias, while the *S. integra* × *S. purpurea*

237   family was entirely female. Single QTL for sex were identified on chr15 within seven of the eight

238   maternal maps, excluding *S. integra* of 13X-426, explaining between 60.7 – 74.1% of the total

239   phenotypic variation (Table 2). LOD scores for the female and male linkage maps are in Table S2

240   and Table S3, respectively. Genome-wide LOD significance thresholds ranged from 3.39 – 3.69,

241   depending on the map. On the four maternal *S. purpurea* maps, the QTL as determined by the

242   permuted significance thresholds accounted for 10.3 – 12.03 Mb of the roughly 15.5 Mb chr15

243   due to suppressed recombination (Figure 4). QTL refinement and 1.5 LOD support intervals

244   narrowed this down to account for 1.17 – 7.31 Mb with peak markers located at 2.91, 2.92,

245   7.33, and 8.65 Mb when *S. suchowensis*, *S. viminalis*, *S. koriyanagi*, and *S. udensis* were the

246   paternal parent, respectively (Table 2; Figure 4). Of the three remaining maternal maps, QTL

247   determined by LOD significance accounted for 12.3, 11.45, and 13.55 Mb of chr15 on the *S.*

248   *alberti*, *S. suchowensis*, and *S. viminalis* maps, respectively, reflecting suppressed

249   recombination. Upon refinement, these ranges were narrowed to account for 1.59, 6.91, and

250   6.41 Mb of the chromosome, with peaks centered at 9.73, 2.57, and 2.92 Mb.

251   DISCUSSION

252   *Salix* is a very diverse genus, consisting of more than 350 species. As the genomic tools

253   available to the shrub willow research community increase in both number and quality, so does

254   the ability to characterize and deploy *Salix's* innate diversity in improving high yielding cultivars.

255   That diversity extends to variation between species even in traits as evolutionarily important as

256   sex determination (Yang et al. 2021). By generating mapping populations that include

257   characterized species crossed with those less studied, we will increase the number of *Salix*

258   species available for trait introgression. We developed the *Salix* F_1 HCP as a resource to

259   characterize the interactions between alleles from different, but related species, and to map

260   variation in important traits. Using GBS, we analyzed the population structure among the eight

261   families, generated linkage maps of each of the parents using backcross markers and mapped

262   the SDR in seven of the eight families using phenotypes collected from repeated field surveys.

263   PCA and hierarchical clustering predominately resolved the population by section with the F_1

264   clustering between the parents as expected. In each analysis, *S. alberti* P294 was found to very

265   closely related to the *S. suchowensis* parents. Given these results and the limited publicly

266   available information about *S. alberti*, P294 is likely *S. suchowensis* and will be described as

267   such.

268   Among the 16 linkage maps produced, all QTL for sex were detected on chr15 in seven

269   of the eight families and only in the maternal maps. In a recent study mapping sex in *S. triandra*

270   using backcross markers, Li et al. (2020) were also only able to detect QTL within the maternal

271   map, which is indicative of a ZW sex determination system on chr15. While this had been

272   known for *S. purpurea* (Zhou et al. 2018), *S. viminalis* (Pucholt et al. 2017b), and *S. suchowensis*

273   (Chen et al. 2016), this is the first study to report that both *S. koriyanagi* and *S. udensis* have ZW

13

274   SDR on chr15. Although the location of the SDR has also not been reported in *S. integra*, the *S.*

275   *integra* × *S. purpurea* family was entirely female and therefore was excluded in the linkage

276   analysis. The chr15 map from *S. integra* P336 is considerably smaller than the other families,

277   aligning to only the 14.2 – 15.5 Mb region of the *S. purpurea* reference genome, while maps of

278   from of the other families had near complete coverage. The region *S. integra* P336 chr15 that

279   did have segregating markers was outside the SDR intervals of the other female maps and is

280   likely in the pseudoautosomal region of the sex chromosome.

281   Six of the eight families were female biased - a feature prevalent in *Salix*. Of those

282   studied here, *S. purpurea* (Gouker et al. 2020), *S. viminalis* (Alström-Rapaport et al. 1997), *S.*

283   *suchowensis* (Yang et al. 2020a), *S. udensis* (Ueno and Seiwa 2003), and *S. integra* (Tozawa et al.

284   2009) have documented cases of sex ratio bias, yet this is the first time it has been reported in

285   *S. koriyanagi*. The genetic basis of sex ratio bias could be a result of secondary sex dimorphisms,

286   such as higher mortality rates in males, increased herbivory and pathogen resistance in

287   females, or the presence of a sex distorter locus (Pucholt et al. 2017a).

288   The mapping of the SDR in this study resulted in marker associations that represented a

289   majority of chr15 prior to refinement. Of the four families with *S. purpurea* as the common

290   female parent, refined QTL shared similar ranges of the SDR described by Zhou et al. (2020) (6.8

291   Mb in length starting at 2.3 Mb) with the exception of the *S. purpurea* × *S. viminalis* family (1.17

292   Mb). Carlson et al. (2019) mapped the SDR to 4.5 – 11.4 Mb on chr15 using a *S. purpurea* F_2

293   population. This study localized the SDR to the same pericentromeric region of chr15 in *S.*

294   *purpurea* using F_1 hybrid families.

295   The most recent delimitation of the SDR in *S. viminalis* spanned roughly 3.4 Mb (approx.

296   2.3 – 5.7 Mb) of chr15 (Almeida et al. 2020). Our mapping of the SDR overlaps this region by

297   2.75 Mb, including the position of our peak marker, even though we aligned markers to a

298   different reference genome. Almeida et al. (2020) aligned the SDR map to chr15 of both *S.*

299   *purpurea* v.1.0 and *S. viminalis* reference genomes and found overall synteny between species,

300   yet with several structural rearrangements. This contrasts with what is seen in our results,

301   where the chr15 of 07-MBG-5027 showed only minor rearrangements, likely due to alignment

302   to the high-quality *S. purpurea* v5.1 reference genome.

303   Based on annotations from the *Populus trichocarpa* genome, differential gene

304   expression analysis in *S. suchowensis* between male and female plants led to predictions that

305   the SDR was originally on chr14 (Liu et al. 2013). However, later work repositioned the SDR to

306   the centromeric region of chr15 when based on *Salix* alignment (Chen et al. 2016). Our study

307   defined the physical distance of the SDR at 2.44 – 9.35 Mb in P295 and 8.63 – 10.22 Mb in P294

308   with an overlap of 0.72 Mb. These two linkage maps show the greatest amount of

309   rearrangement on chr15 when aligned to the *S. purpurea* reference. It is fair to conclude that

310   alignment to a future *S. suchowensis* reference genome would aid in improving the mapping

311   resolution in this region.

312   Suppressed recombination is a hallmark of chromosomes containing an SDR. Comparing

313   the map of chr15 from each family, the SDR extends across a region of sparce marker density,

314   approximately 3 to 9 Mb and is flanked by regions with greater marker density. As described

315   above, this centromeric region with suppressed recombination is often associated with the SDR

316   in *Salix*. In all seven families with mapped SDR, the peak marker was located within this region

317   although its position varied. In the two *S. viminalis* families and two of the three *S. suchowensis*

318   families, the peak marker was located within a 0.38 Mb region (2.57 – 2.95 Mb), while the

319   remaining three families were less consistent and located proximal to the centromere. The

320   generation of additional reference genomes for use in mapping will add context to these results

321   and further refine the structure of the SDR among variable species.

322   This study described the population structure among the eight families within the *Salix*

323   F_1 HCP, constructed linkage maps for each parent, and mapped the SDR to the maternal chr15

324   in seven of the eight families. The introduction of the *Salix* F_1 HCP provides the opportunity to

325   map QTL for phenotypic traits beyond sex determination, while the linkage maps could be used

326   to anchor and scaffold contigs in the generation of new reference genomes for each of the

327   parents. While all species have a ZW sex determination system with an SDR that maps to chr15,

328   these genetic resources provide a foundation for further characterization of the mechanism of

329   sex determination and mapping of other key traits in these related species.

330   **Data Availability**

331   The GBS data used in the population structure analysis and the eight files used for

332   linkage map construction each family prior to linkage map construction are available in hapmap

333   format through figshare at https://figshare.com at "GBS Data from Wilkerson et al". R code

334   used to format and create the linkage maps is available on the Willowpedia Github site located

335   at https://github.com/Willowpedia/Wilkerson_etal_SalixLinkageMaps. Table S1 contains

336   statistics on each of the 16 linkage maps created in this study, including marker count and cM

337   length for each linkage group and the total markers and cM length for each map. Figure S1 is a

16

338 pdf slide show that, one map per slide, shows the marker cM by physical position for each

339 linkage group and a heatmap visualizing recombination frequency and linkage. Tables S2 and S3

340 contain the cM distance and sex QTL LOD scores of markers within each female and male map,

341 respectively.

Conflicts of Interest

350 The authors have no conflicts of interest to report.

Figures

352 Figure 1: Pedigrees of the *Salix* F_1 HCP. There were four full-sib families with 94001 as the
353 paternal parent and four full-sib families with 94006 as the maternal parent. Reciprocal crosses
354 were made with male and female *S. viminalis* and *S. suchowensis* while *S. integra*, *S. alberti*, *S.*
355 *udensis*, and *S. koriyanagi* were crossed only once. P294 was received with the identification of
356 *S. alberti*.

357 Figure 2: Results of PC and fastSTRUCTURE analysis of the *Salix* F_1 HCP. A: PCA of the parents; B:
358 PCA of the F_1 and the parents; C: Hierarchical clustering of the parents; D: Distruct plot using
359 fastSTRUCTURE results. In panels A, B, and C: purple - *S. purpurea*, red - *S. viminalis*, green - *S.*
360 *suchowensis*, blue - *S. udensis*, teal – *S. integra*, grey – *S. koriyanagi*, orange - *S. alberti*. Colors in
361 panel D are based on fastSRUCTURE results, affecting only *S. alberti*. *Received as *S. alberti*.

362   Figure 3: Linkage maps for each of the parents within the *Salix* F$_1$ HCP. Female maps (first and

363   third rows) were constructed using female informative markers (AB × AA), while male maps

364   (second and forth rows) were constructed using male informative markers (AA × AB).

365   Figure 4: Genetic (cM) and Physical (Mb) distances for the maternal parents in the *Salix* F$_1$ HCP.

366   Each parent has a set of two maps, genetic on the top and physical on the bottom with lines

367   connecting them indicating a marker's relative position in both maps. On the genetic maps,

368   each marker is represented by a grey vertical line while the 1.5 LOD support interval of the QTL

369   is shown with a purple box. On the physical map, grey vertical lines are markers not associated

370   with the QTL, yellow shows markers whose LOD score was above the permuted significance

371   threshold, purple are markers within the 1.5 LOD support interval, and red indicates the cM and

372   Mb position of the peak marker. P336's linkage map cM distance was artificially increased for

373   figure clarity.* Received as *S. alberti.*

374

375

376 **Tables**

Table 1: Sex phenotype statistics for the eight families in the *Salix* F_1 HCP.

Family	Maternal Species	Paternal Species	N	Females	Males	F:M Ratio	P-Value
13X-358	*S. purpurea*	*S. udensis*	150	91	59	1.6:1	0.01[*]
10X-400	*S. purpurea*	*S. suchowensis*	88	53	33	1.6:1	0.03[*]
11X-407	*S. purpurea*	*S. viminalis*	100	60	40	1.5:1	0.05[NS]
13X-438	*S. purpurea*	*S. koriyanagi*	150	93	57	1.6:1	< 0.001[*]
12X-421	*S. viminalis*	*S. purpurea*	100	53	47	1.1:1	0.55[NS]
13X-426	*S. integra*	*S. purpurea*	150	150	0	1:0	< 0.001[*]
13X-440	*S. suchowensis*	*S. purpurea*	150	91	58	1.6:1	0.01[*]
13X-443	*S. alberti*[‡]	*S. purpurea*	150	87	62	1.4:1	0.04[*]

[*] Chi-square test for sex ratio bias (p < 0.05); [NS] Not significant; [‡] Received as *S. alberti*.

377

Table 2: Sex QTL associated with the SDR within the maternal linkage maps.

Family	Maternal Species	Chr	Peak (cM)	Peak LOD	Peak (Mb)	Low-LSI (Mb)	High-LSI (Mb)	PVE (%)
13X-358	*S. purpurea*	15	85.3	29.3	8.65	4.51	9.22	73.7
10X-400	*S. purpurea*	15	81.1	16.1	2.91	2.61	9.92	61.8
11X-407	*S. purpurea*	15	114.2	18.6	2.92	2.88	4.05	60.7
13X-438	*S. purpurea*	15	70.1	35.1	7.33	2.63	8.29	71.1
12X-421	*S. viminalis*	15	86.0	21.1	2.92	2.95	9.36	66.0
13X-440	*S. suchowensis*	15	78.4	26.7	2.57	2.44	9.35	74.1
13X-443	*S. alberti*[‡]	15	145.8	35.6	9.73	8.63	10.22	70.3

LSI: 1.5 LOD Support Interval; PVE: Percent Variation Explained; cM: centiMorgan position within the linkage map; Mb: Physical position based on alignment to *S. purpurea* reference genome. [‡]Received as *S. alberti*.

378

379

380   REFERENCES

381   Almeida P., Proux-Wera E., Churcher A., Soler L., Dainat J. et al. 2020. Genome assembly of the basket

382   willow, *Salix viminalis*, reveals earliest stages of sex chromosome expansion. BMC Biol. 18(1): 78.

383   Alström-Rapaport C., Lascoux M., Gullberg U. 1997. Sex determination and sex ratio in the dioecious

384   shrub *Salix viminalis* L. Theor Appl Genet. 94(3-4): 493-497.

385   Berlin S., Lagercrantz U., von Arnold S., Ost T., Ronnberg-Wastljung A. C. 2010. High-density linkage

386   mapping and evolution of paralogs and orthologs in *Salix* and *Populus*. BMC Genomics. 11: 129.

387   Bradbury P. J., Zhang Z., Kroon D. E., Casstevens T. M., Ramdoss Y. et al. 2007. TASSEL: software for

388   association mapping of complex traits in diverse samples. Bioinformatics. 23(19): 2633-2635.

389   Broman K. W., Wu H., Sen Ś., Churchill G. A. 2003. R/qtl: QTL mapping in experimental crosses.

390   Bioinformatics. 19(7): 889-890.

391   Cao Y., Ma C., Chen H., Zhang J., White J. C. et al. 2020. Xylem-based long-distance transport and phloem

392   remobilization of copper in *Salix integra* Thunb. J Hazard Mater. 392: 122428.

393   Carlson C. H., Gouker F. E., Crowell C. R., Evans L., DiFazio S. P. et al. 2019. Joint linkage and association

394   mapping of complex traits in shrub willow (*Salix purpurea* L.). Ann Bot. 124(4): 701-716.

395   Chen Y., Wang T., Fang L., Li X., Yin T. 2016. Confirmation of single-locus sex determination and female

396   heterogamety in willow based on linkage analysis. PLoS One. 11(2): e0147671.

397   Clifton-Brown J., Harfouche A., Casler M. D., Dylan Jones H., Macalpine W. J. et al. 2019. Breeding

398   progress and preparedness for mass-scale deployment of perennial lignocellulosic biomass crops

399   switchgrass, miscanthus, willow and poplar. GCB Bioenergy. 11(1): 118-151.

400   Crowell C. R., Bekauri M. M., Cala A. R., Mcmullen P., Smart L. B. et al. 2020. Differential susceptibility of

401   diverse *Salix* spp. to *Melampsora americana* and *Melampsora paradoxa*. Plant Dis. 104: 2949-

402   2957.

403   Dickmann D. I., Kuzovkina J. 2014. Poplars and willows of the world, with emphasis on silviculturally

404   important species. Poplars and willows: Trees for society and the environment. Rome, Italy:

405   FAO.

406   Elshire R. J., Glaubitz J. C., Sun Q., Poland J. A., Kawamoto K. et al. 2011. A robust, simple genotyping-by-

407   sequencing (GBS) approach for high diversity species. PLoS One. 6(5): e19379.

408   Fabio E. S., Leary C. J., Smart L. B. 2019. Tolerance of novel inter-specific shrub willow hybrids to water

409   stress. Trees. 33(4): 1015-1026.

410   Fabio E. S., Smart L. B. 2020. Genetic and environmental influences on first rotation shrub willow (*Salix*

411   spp.) bark and wood elemental composition. Bioenergy Res. 13(3): 797-809.

412   Geraldes A., Hefer C. A., Capron A., Kolosova N., Martinez-Nuñez F. et al. 2015. Recent Y chromosome

413   divergence despite ancient origin of dioecy in poplars (*Populus*). Mol Ecol. 24(13): 3243-3256.

414   Gouker F. E., Carlson C. H., Zou J., Evans L., Crowell C. R. et al. 2020. Sexual dimorphism and sex ratio

415   bias in the dioecious willow *Salix purpurea* L. bioRxiv. 2020.2004.2005.026427.

416   Hallingbäck H. R., Berlin S., Nordh N.-E., Weih M., Rönnberg-Wästljung A.-C. 2019. Genome wide

417   associations of growth, phenology, and plasticity traits in willow [*Salix viminalis* (L.)]. Front Plant

418   Sci. 10(753).

419   Hallingbäck H. R., Fogelqvist J., Powers S. J., Turrion-Gomez J., Rossiter R. et al. 2016. Association

420   mapping in *Salix viminalis* L. (Salicaceae) – identification of candidate genes associated with

421   growth and phenology. GCB Bioenergy. 8(3): 670-685.

422   Hanley S., Mallott M., Karp A. 2006. Alignment of a *Salix* linkage map to the *Populus* genomic sequence

423   reveals macrosynteny between willow and poplar genomes. Tree Genet Genomes. 3(1): 35-48.

424   Hou J., Ye N., Zhang D., Chen Y., Fang L. et al. 2015. Different autosomes evolved into sex chromosomes

425   in the sister genera of *Salix* and *Populus*. Scientific Reports. 5(1): 9076.

426   Jia H., Wang L., Li J., Sun P., Lu M. et al. 2020. Physiological and metabolic responses of *Salix*

427   *sinopurpurea* and *Salix suchowensis* to drought stress. Trees. 34(2): 563-577.

428   Kim J., Kim Y., Park J. 2019. Complete chloroplast genome sequence of the *Salix koriyanagi* Kimura ex

429   Goerz (Salicaceae). Mitochondrial DNA B: Resour. 4(1): 549-550.

430   Kopp R., Maynard C., Rocha De Niella P., Smart L., Abrahamson L. 2002. Collection and storage of pollen

431   from *Salix* using organic solvents. Am J Bot. 89: 248-252.

432   Li H., Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.

433   Bioinformatics. 25(14): 1754-1760.

434   Li W., Wu H., Li X., Chen Y., Yin T. 2020. Fine mapping of the sex locus in *Salix triandra* confirms a

435   consistent sex determination mechanism in genus *Salix*. Hortic Res. 7(1): 64.

436   Liu J., Yin T., Ye N., Chen Y., Yin T. et al. 2013. Transcriptome analysis of the differentially expressed

437   genes in the male and female shrub willows (*Salix suchowensis*). PLoS One. 8(4): e60181.

438   Money D., Migicovsky Z., Gardner K., Myles S. 2017. LinkImputeR: user-guided genotype calling and

439   imputation for non-model organisms. BMC Genomics. 18(1): 523.

440   Niiyama K. 2008. Coexistence of *Salix* species in a seasonally flooded habitat. Ecology of Riparian Forests

441   in Japan. Springer. p. 165-174.

442   Paolucci I., Gaudet M., Jorge V., Beritognolo I., Terzoli S. et al. 2010. Genetic linkage maps of *Populus*

443   *alba* L. and comparative mapping analysis of sex determination across *Populus* species. Tree

444   Genet Genomes. 6(6): 863-875.

445   Park J., Kim Y., Xi H. 2019. The complete chloroplast genome sequence of male individual of Korean

446   endemic willow, *Salix koriyanagi* Kimura ex Goerz (Salicaceae). Mitochondrial DNA B: Resour.

447   4(1): 1619-1621.

448   Pucholt P., Hallingbäck H. R., Berlin S. 2017a. Allelic incompatibility can explain female biased sex ratios

449   in dioecious plants. BMC Genomics. 18(1): 251.

450   Pucholt P., Rönnberg-Wästljung A. C., Berlin S. 2015. Single locus sex determination and female

451   heterogamety in the basket willow (*Salix viminalis* L.). Heredity. 114(6): 575-583.

452   Pucholt P., Wright A. E., Conze L. L., Mank J. E., Berlin S. 2017b. Recent sex chromosome divergence

453   despite ancient dioecy in the willow *Salix viminalis*. Mol Biol Evol. 34(8): 1991-2001.

454   R Core Team. 2020. R: A language and environment for statistical computing. Vienna, Austria: R

455   Foundation for Statistical Computing.

456   Raj A., Stephens M., Pritchard J. K. 2014. fastSTRUCTURE: variational inference of population structure in

457   large SNP data sets. Genetics. 197(2): 573-589.

458   Rönnberg-Wästljung A. C., Glynn C., Weih M. 2005. QTL analyses of drought tolerance and growth for a

459   *Salix dasyclados* x *Salix viminalis* hybrid in contrasting water regimes. Theor Appl Genet. 110(3):

460   537-549.

461   Rönnberg-Wästljung A. C., Samils B., Tsarouhas V., Gullberg U. 2008. Resistance to *Melampsora larici-*

462   *epitea* leaf rust in *Salix*: analyses of quantitative trait loci. J Appl Genet. 49: 321-331.

463   Rosso L., Facciotto G., Bergante S., Vietto L., Nervo G. 2013. Selection and testing of *Populus alba* and

464   *Salix* spp. as bioenergy feedstock: Preliminary results. Appl Energy. 102: 87-92.

465   Samils B., Rönnberg-Wästljung A.-C., Stenlid J. 2011. QTL mapping of resistance to leaf rust in *Salix*. Tree

466   Genet Genomes. 7(6): 1219-1235.

467   Sanderson B. J., Feng G., Hu N., Carlson C. H., Smart L. B. et al. 2021. Sex determination through X–Y

468   heterogamety in *Salix nigra*. Heredity. 126(4): 630-639.

469   Slaght J. C., Surmach S. G., Kisleiko A. A. 2018. Ecology and conservation of Blakiston's fish owl in Russia.

470   Biodiversity Conservation Using Umbrella Species. Springer. p. 47-70.

471   Smart L. B., Volk T., Lin J., Kopp R. F., Phillips I. S. et al. 2005. Genetic improvement of shrub willow (*Salix*

472   spp.) crops for bioenergy and environmental applications in the United States. Unasylva. 56: 51-

473   55.

474   Stanton B. J., Serapiglia M. J., Smart L. B. 2014. The domestication and conservation of *Populus* and *Salix*

475   genetic resources. Poplars and willows: trees for society and the environment. Wallingford, UK:

476   CAB International. p. 124-199.

477   Stoof C. R., Richards B. K., Woodbury P. B., Fabio E. S., Brumbach A. R. et al. 2014. Untapped potential:

478   opportunities and challenges for sustainable bioenergy production from marginal lands in the

479   Northeast USA. Bioenergy Res. 8(2): 482-501.

480   Sulima P., Przyborowski J. A., Kuszewska A., Zaluski D., Jedryczka M. et al. 2017. Identification of

481   quantitative trait loci conditioning the main biomass yield components and resistance to

482   *Melampsora* spp. in *Salix viminalis* x *Salix schwerinii* hybrids. Int J Mol Sci. 18(3).

483   Taylor J., Butler D. 2017. R Package ASMap: efficient genetic linkage map construction and diagnosis. J

484   Stat Softw. 79(6): 1-29.

485   Tozawa M., Ueno N., Seiwa K. 2009. Compensatory mechanisms for reproductive costs in the dioecious

486   tree *Salix integra*. Botany. 87(3): 315-323.

487   Ueno N., Kanno H., Seiwa K. 2006. Sexual differences in shoot and leaf dynamics in the dioecious tree

488   *Salix sachalinensis*. Botany. 84(12): 1852-1859.

489   Ueno N., Seiwa K. 2003. Gender-specific shoot structure and functions in relation to habitat conditions

490   in a dioecious tree, *Salix sachalinensis*. J For Res. 8(1): 0009-0016.

491   Wei S., Yang Y., Yin T. 2020. The chromosome-scale assembly of the willow genome provides insight into

492   Salicaceae genome evolution. Hortic Res. 7(1): 45.

493   Yang G., Xu Q., Li W., Ling J., Li X. et al. 2020a. Sex-related differences in growth, herbivory, and defense

494   of two *Salix* species. Forests. 11(4): 450.

495   Yang W., Wang D., Li Y., Zhang Z., Tong S. et al. 2021. A general model to explain repeated turnovers of

496   sex determination in the Salicaceae. Mol Biol Evol. 38(3): 968-980.

497   Yang W., Wang Y., Liu D., Hussain B., Ding Z. et al. 2020b. Interactions between cadmium and zinc in

498   uptake, accumulation and bioavailability for *Salix integra* with respect to phytoremediation. Int J

499   Phytoremediation. 22(6): 628-637.

500   Zhou R., Macaya-Sanz D., Carlson C. H., Schmutz J., Jenkins J. W. et al. 2020. A willow sex chromosome

501   reveals convergent evolution of complex palindromic repeats. Genome Biol. 21(1): 38.

502   Zhou R., Macaya-Sanz D., Rodgers-Melnick E., Carlson C. H., Gouker F. E. et al. 2018. Characterization of

503   a large sex determination region in *Salix purpurea* L. (Salicaceae). Mol Genet Genomics. 293(6):

504   1437-1452.

505

506

Female *S. purpurea* 94006 (13X-358) with 2801 Markers
Female *S. purpurea* 94006 (10X-400) with 2196 Markers
Female *S. purpurea* 94006 (11X-407) with 2527 Markers
Female *S. purpurea* 94006 (13X-438) with 3503 Markers

Male *S. udensis* 04-BN-05 (13X-358) with 3355 Markers
Male *S. suchowensis* P63 (10X-400) with 2035 Markers
Male *S. viminalis* Jorr (11X-407) with 3086 Markers
Male *S. koriyanagi* 04-FF-016 (13X-438) with 3852 Markers

Female *S. integra* P336 (13X-426) with 2552 Markers
Female *S. viminalis* 07-MBG-5027 (12X-421) with 3115 Markers
Female *S. suchowensis* P295 (13X-440) with 3446 Markers
Female *S. alberti* P294 (13X-443) with 2836 Markers

Male *S. purpurea* 94001 (13X-426) with 3586 Markers
Male *S. purpurea* 94001 (12X-421) with 2991 Markers
Male *S. purpurea* 94001 (13X-440) with 3221 Markers
Male *S. purpurea* 94001 (13X-443) with 2825 Markers
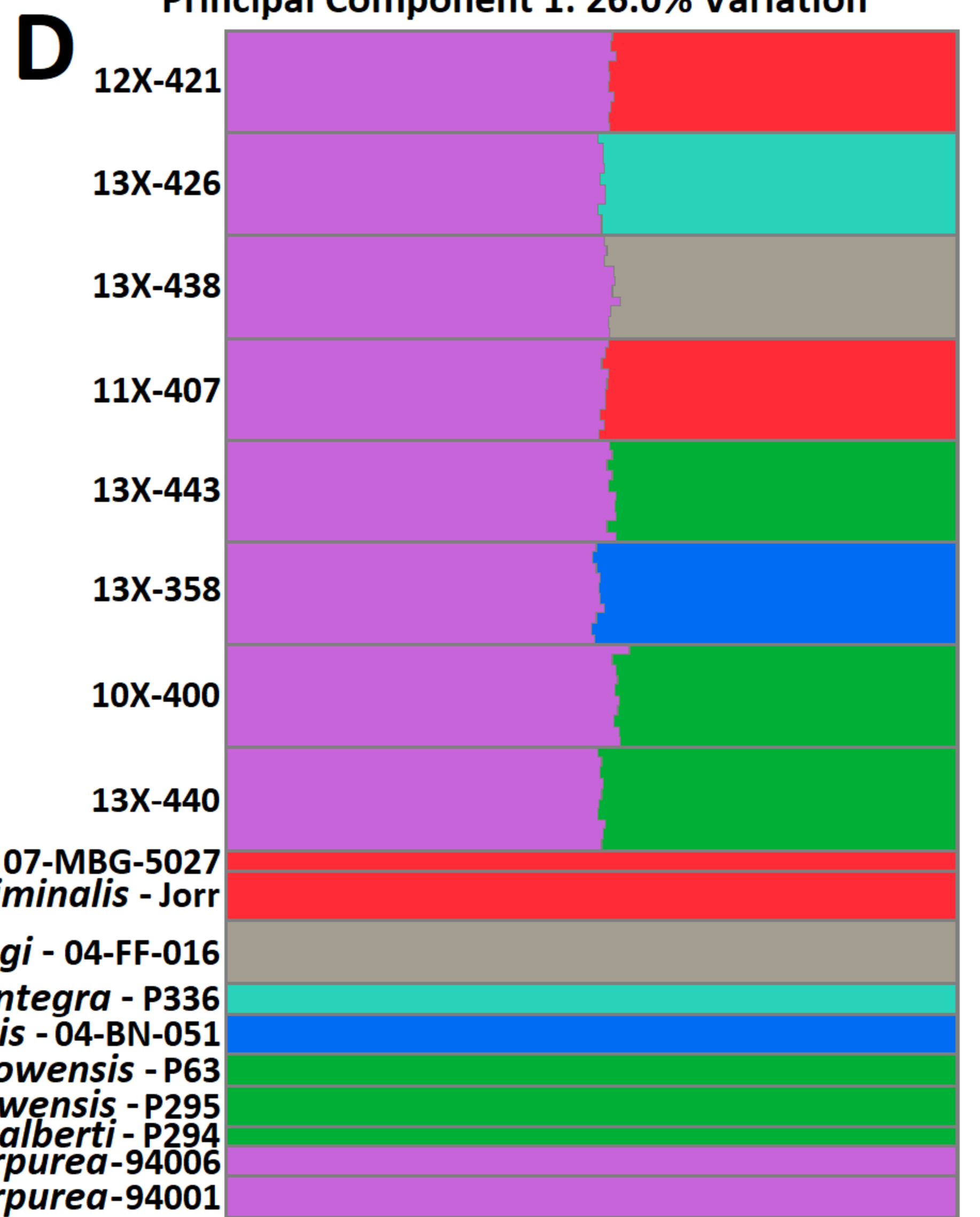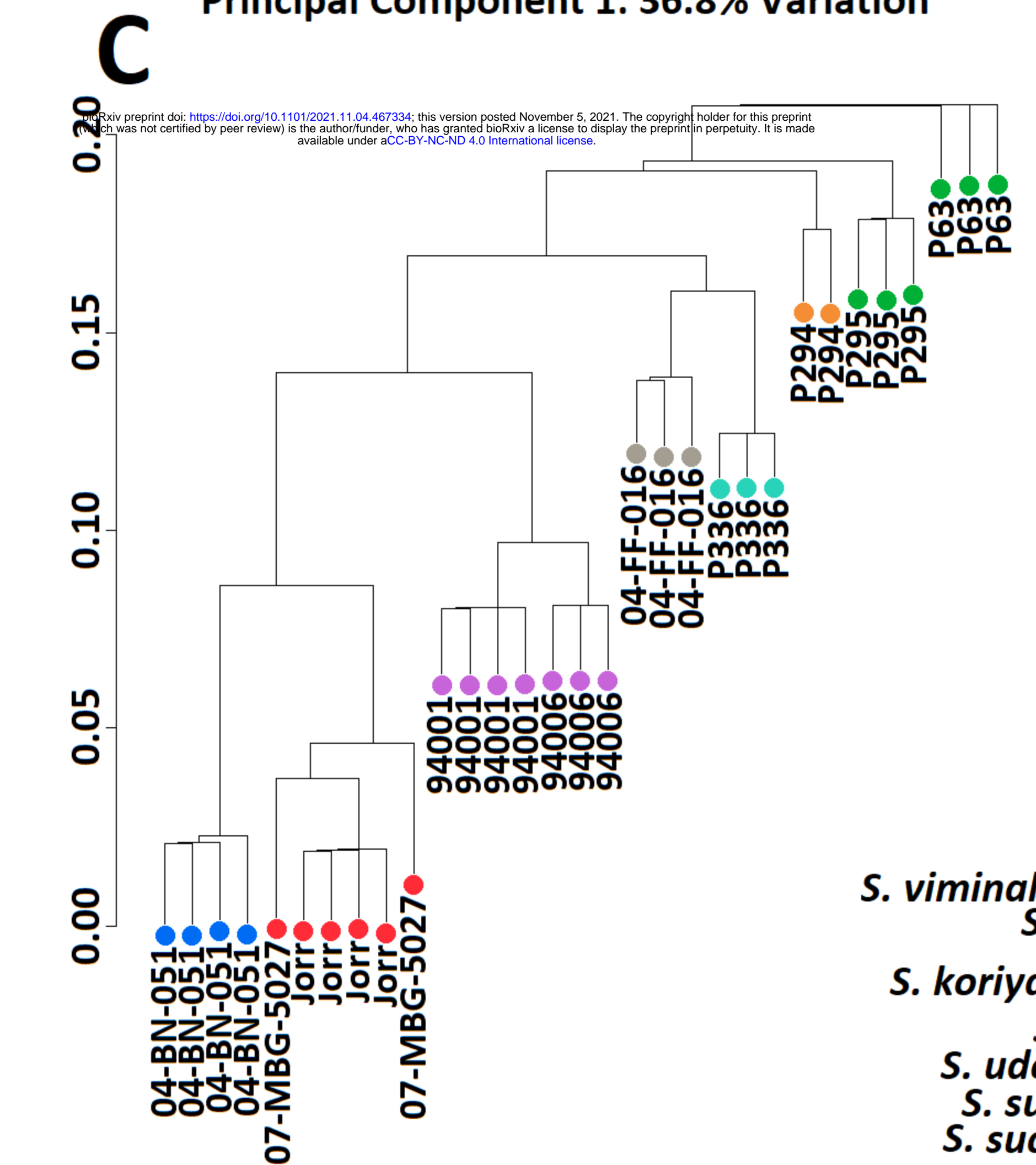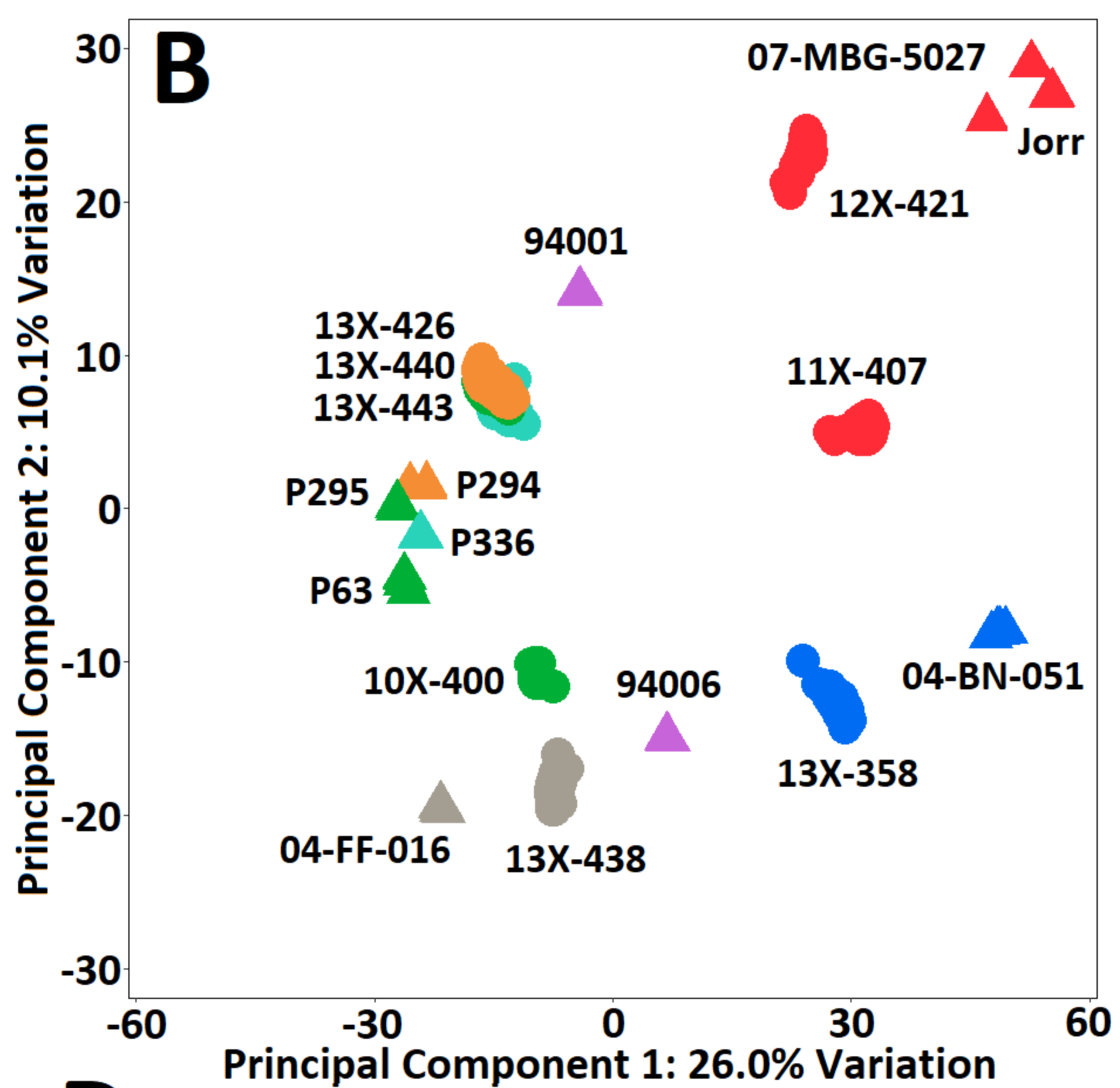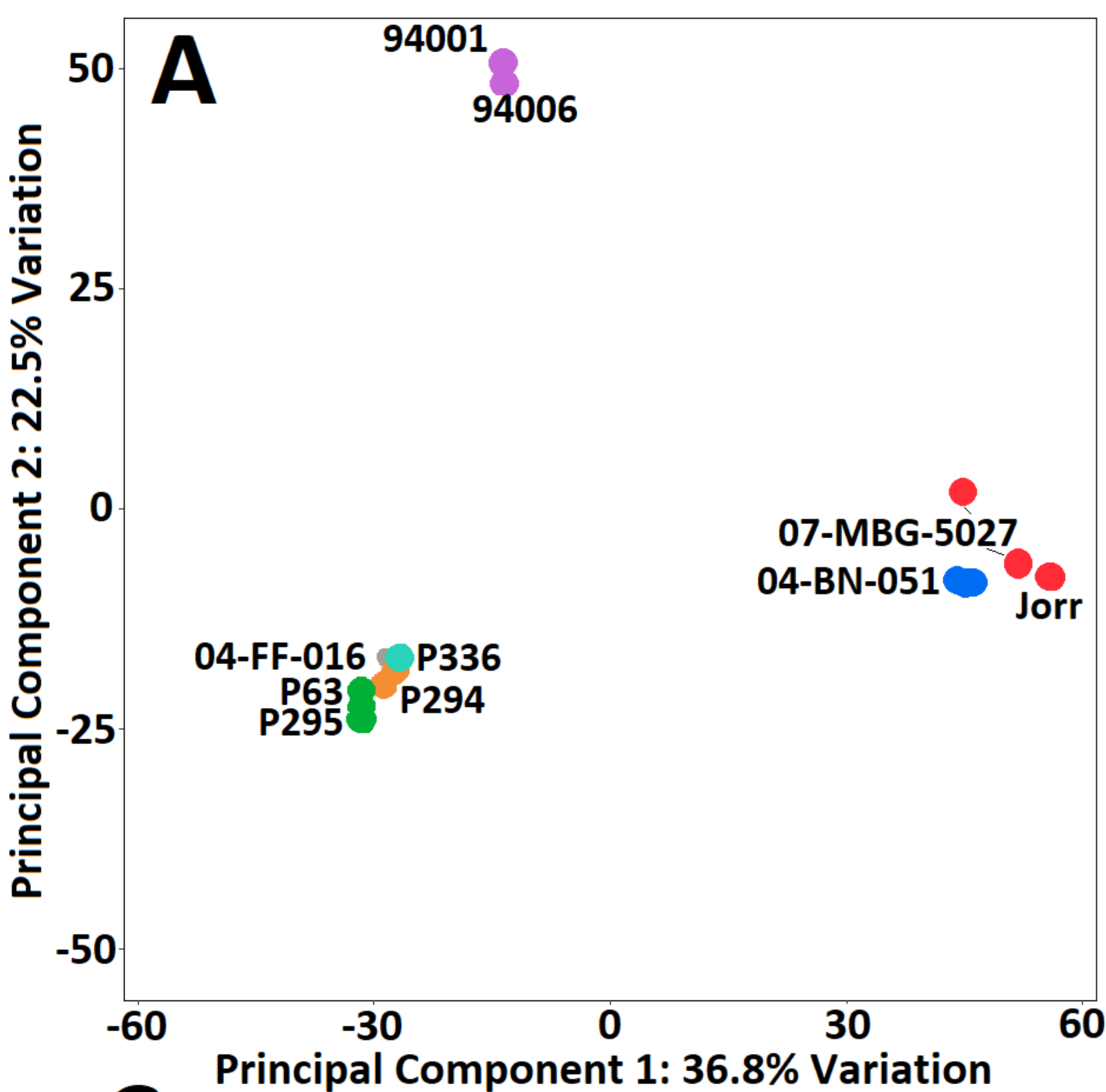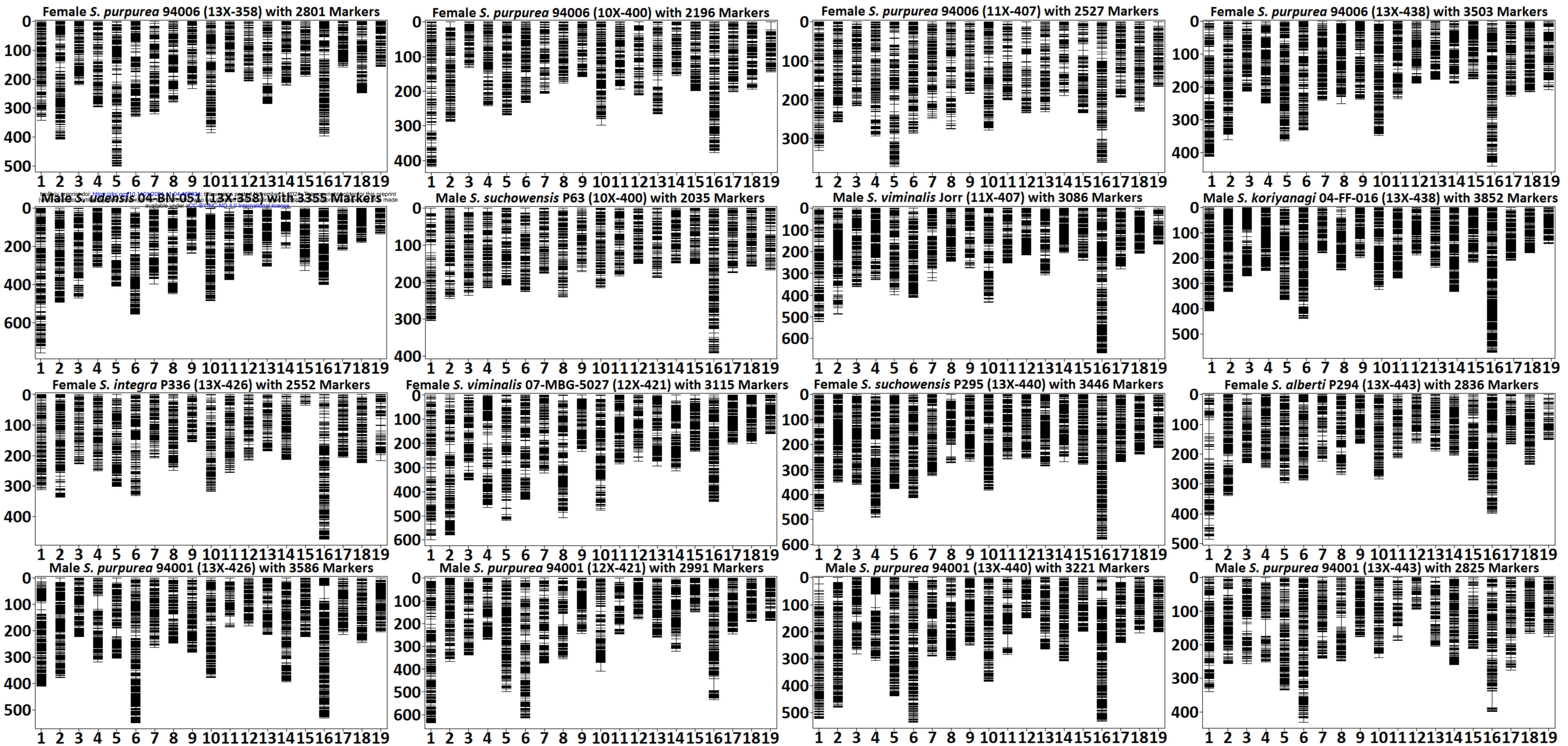
Genetic Distance (cM)

0   50   100   150   200   250   300

Genetic

P294 - *S. alberti**

Physical

Genetic

P295 - *S. suchowensis*

Physical

Genetic

P336 - *S. integra*

Physical

Genetic

07-MBG-5027 - *S. viminalis*

Physical

Genetic

94006 (13X-438) - *S. purpurea*

Physical

Genetic

94006 (11X-407) - *S. purpurea*
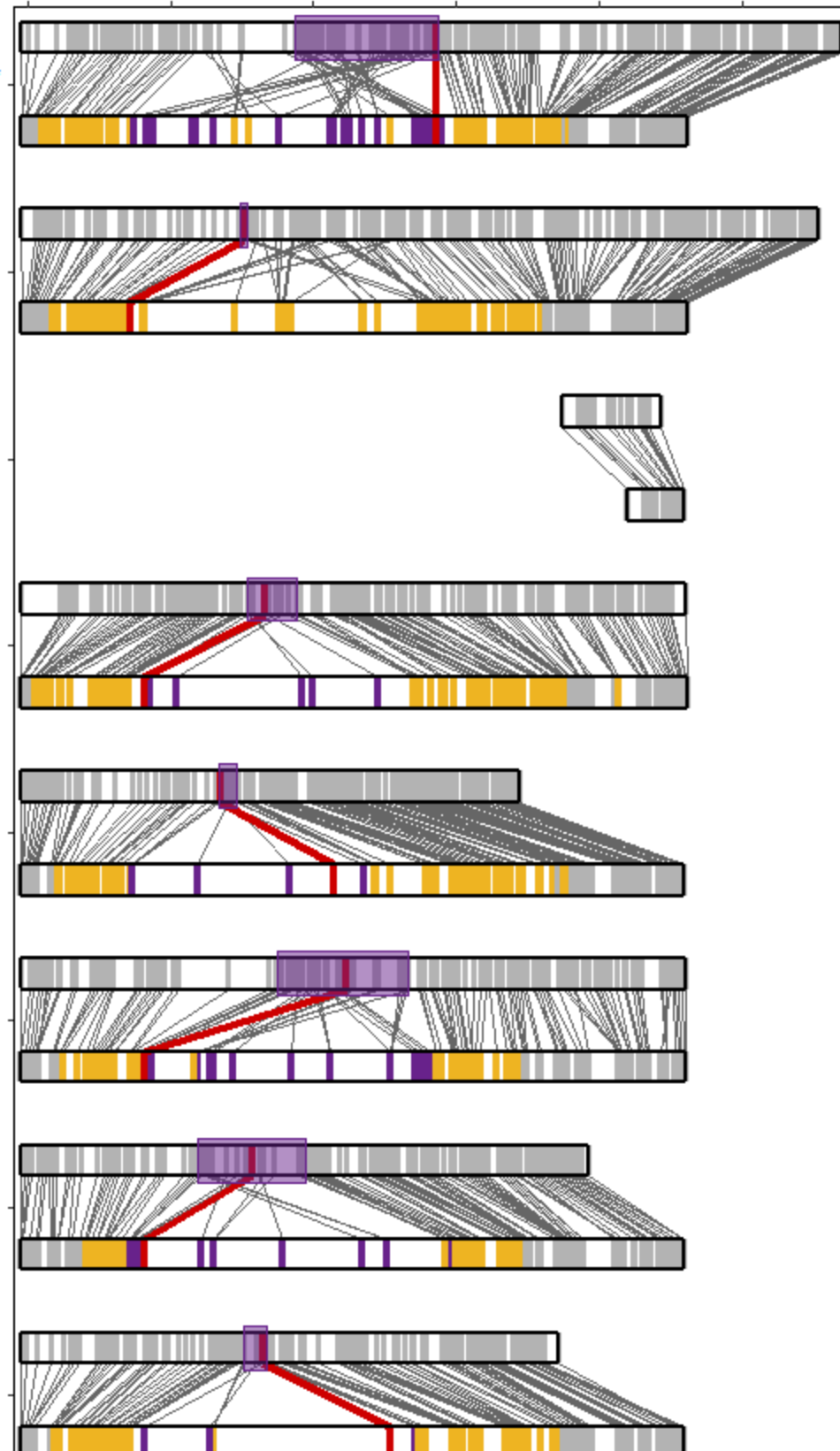
Physical

Genetic

94006 (10X-400) - *S. purpurea*

Physical

Genetic

94006 (13X-358) - *S. purpurea*

Physical

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Physical Distance (Mb)