Running title: Adaptive evolution of *Aquilegia* species

Genetic and epigenetic mechanisms underpinning the adaptive radiation of *Aquilegia* species

Tianyuan Lu[1,2,3], Ming-Rui Li[1], Ning Ding[1], Zhen-Hui Wang[4], Li-Zhen Lan[1], Xiang Gao[5], and Lin-Feng Li[1,*]

[1]Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, School of Life Sciences, Fudan University, Shanghai 200438, China;

[2]McGill University and Genome Quebec Innovation Center, Montreal H3A 0G1, Quebec, Canada;

[3]Lady Davis Institute, SMBD JGH, Montreal H3A 1A3, Quebec, Canada;

[4]Faculty of Agronomy, Jilin Agricultural University, Changchun 130118, China;

[5]Key Laboratory of Molecular Epigenetics of the Ministry of Education, Northeast Normal University, Changchun 130024, China.

***Correspondence author:**

Lin-Feng Li (lilinfeng@fudan.edu.cn)

Summary

• Elucidating how the intrinsic factors interact with extrinsic triggers to determine species diversification is crucial to understanding the evolution and persistence of biodiversity. The genus *Aquilegia* is a model system to address the evolutionary mechanisms underpinning rapid adaptive radiation.

• We surveyed the genomes and methylomes of ten worldwide *Aquilegia* species to investigate whether specific genetic and epigenetic architectures are involved in the diversification of Asian, European and North American columbine species.

• The resulting phylogenies and population structure inferences revealed high divergence among the Asian, European and North American species. Candidate genes identified at the genetic and epigenetic levels are functionally correlated with diverse adaptation-related traits such as cell reproduction, flower development, and stress tolerance. In particular, a considerable proportion of the selection genes and their associated pathways show overlaps among the intra- and inter-lineage comparisons. Moreover, while CG-loss variations can lead to depletion of cytosine methylation level, epigenetic modification is a complementary intrinsic factor that intertwines with genetic mechanism to facilitate the diversification of *Aquilegia* species.

• Our findings suggest that specific genetic and epigenetic architectures have conferred high adaptability to the columbine species to cope with diverse external conditions, which eventually led to the rapid radiation of *Aquilegia* species.

Key words: Adaptive radiation; *Aquilegia*; Methylome; Speciation; Whole genome resequencing

Introduction

Adaptive radiation is the rapid diversification of a single ancestral species into a vast array of common descendants that inhabit different ecological niches or use a variety of resources and differ in phenotypic traits required to exploit diverse environments (Schluter, 1996; Givnish, 1997; Schluter, 2000; Givnish, 2015). Disentangling the evolutionary mechanisms underpinning adaptive radiation is fundamental to understanding the evolution and persistence of biodiversity (Simpson & Olson, 1953; Losos, 2010). Well-known animal and plant lineages that diversified through adaptive radiation include Hawaiian silversword, Caribbean anoles, Darwin's finches, and African cichlids (Carlquist & Motley, 2003; Losos & Ricklefs, 2009; Lamichhaney *et al.*, 2015; Irisarri *et al.*, 2018). However, it still remains under-investigated as to why some lineages can diversify rapidly but their close relatives or other sympatrically distributed lineages do not. In the past decades, accumulating evidence from diverse radiation lineages suggests that extrinsic environmental variables (e.g., resource availability) and intrinsic factors (e.g., lineage-specific genetic variations) can interact to determine the rate and volume of species diversification (Wagner *et al.*, 2012). Among the extrinsic triggers, ecological opportunity is considered as the primary mechanism that causes rapid adaptive radiation through the acquisition of key innovations, invasion of new environments and extinction of competitors (Simpson, 1949; Schluter, 2000). On the other hand, new species can also arise as a result of natural selection acting on intrinsic factors (e.g., advantage alleles) that ultimately generate both the phenotypic disparity and similarity among closely related species (Berner & Salzburger, 2015). An illustration of rapid adaptive radiation is the African cichlid fishes, on which the extrinsic environmental factors (e.g., ecological specialization) and extrinsic traits (e.g., adaptive introgression) have acted together to provoke the repeated adaptive radiation in geographically isolated lakes (Wagner *et al.*, 2012; Brawand *et al.*, 2014; Ford *et al.*, 2016; Irisarri *et al.*, 2018).

The genus *Aquilegia* L. (columbine) is a well-recognized model system to address the evolutionary mechanisms underlying rapid adaptive radiation (Fior *et al.*, 2013; Filiault *et al.*, 2018). It includes approximately 70 recently diversified species that are widely distributed in the temperate areas of North America and Eurasia (Munz, 1946). Phylogenetic and geographic inferences have illustrated two independent adaptive radiations of North American and European lineages from the distinct ancestral Asian species (Bastida & Herrera, 2010; Fior *et al.*, 2013). Floral diversification of the North American species is highly correlated with the pollinator specialization (Whittall & Hodges, 2007; Alcántara *et al.*, 2010; Kramer & Hodges, 2010; Sharma *et al.*, 2014). In contrast, ecological adaptation and geographic isolation are thought as the major driving forces that promoted the rapid radiation of European species

73 (Fior *et al.*, 2013; Garrido *et al.*, 2017). In Asia, shifts in pollinator and ecological habitat are both
74 proposed to be the underpinning mechanisms that resulted in the diversification of more than 20
75 morphologically distinct species (Tang *et al.*, 2007; Li *et al.*, 2014). Unlike the North American and
76 European lineages, the Asian columbines consist of four highly divergent lineages corresponding to
77 respective geographic origins and evolved independently under distinct external triggers (Tang *et al.*,
78 2007; Fior *et al.*, 2013; Li *et al.*, 2014). The independent adaptive radiations of North American and
79 European lineages from the more ancestral but independently diversified Asian columbine species
80 provide an ideal system to examine how the extrinsic and intrinsic triggers interact to promote the
81 diversification of *Aquilegia* species.

82 In this study, we surveyed the genomes and DNA methylomes of 36 accessions from ten worldwide
83 columbine species to address the underlying evolutionary mechanisms that caused the diversification of
84 Asian, European and North American columbine species. Of the Asian lineage, four phylogenetically
85 distinct species (*A*. *yabeana*, *A*. *viridiflora*, *A. oxysepala* and *A. japonica*) were selected according to their
86 geographic distributions and ecological habitats. *Aquilegia japonica* and *A. oxysepala* are sister species
87 inhabiting alpine tundra and low altitude forest niches in northeastern China, respectively (Li *et al.*, 2011;
88 Li *et al.*, 2014). Our previous studies have documented that natural selection during ecological
89 specialization together with genetic drift under geographic isolation caused the rapid evolution of
90 reproductive isolation between the two species (Li *et al.*, 2014; Li *et al.*, 2019). Here we further
91 investigate how the genetic and epigenetic mechanisms conferred adaptability to the two species to
92 cope with contrasting ecological conditions. In addition, patterns of nucleotide variation and cytosine
93 methylation were also evaluated for the *A*. *yabeana* and *A. viridiflora*. The former species shares highly
94 similar morphological traits and ecological niches with the *A. oxysepala* but is allopatrically distributed in
95 the northern China. In contrast, while the *A. viridiflora* is sympatrically distributed with *A*. *yabeana* and *A.*
96 *oxysepala* in the northern and northeastern China, it often occupies rocky and sandy ecological niches.
97 These attributes allow us to experimentally test how the genetic and epigenetic mechanisms
98 contributed to the phenotypic divergence and ecological specialization of these Asian species. In
99 addition, we also surveyed the genetic and epigenetic architectures of six North American and European
100 columbine species. We aimed to examine whether the adaptive radiations of North American and
101 European lineages are determined by extrinsic factors related to different ecological opportunities or
102 triggered by intrinsic traits utilizing the lineage-specific genomic and epigenomic architectures. Our
103 study will provide a genome-wide view of how the diverse evolutionary mechanisms facilitate the rapid
104 species diversification in the genus *Aquilegia*.

105 **Methods and Materials**

106 **Sample collection, DNA extraction and sequencing**

107 In this study, a total of 36 accessions from ten worldwide *Aquilegia* species were collected (**Table S1**).
108 Among these samples, 18 accessions were collected to represent the two Asian species, *A. japonica* and
109 *A. oxysepala*, and their putative hybrid. In addition, four accessions were collected from the other two
110 Asian species *A. yabeana* and *A. viridiflora*. Furthermore, six and eight accessions were sampled from
111 the European and North American lineages, respectively. All the 36 accessions were grown in green
112 house under the same conditions (25°C/12 hours, 16°C/12 hours). Genomic DNA was extracted from
113 fresh mature leaves using TianGen plant genomic DNA kit. Whole genome resequencing and bisulfite
114 sequencing were performed on the extracted genomic DNA using the Illumina X-ten platform (Illumina,
115 California, USA). Short-insert (350bp) DNA libraries of all accessions were constructed by NovoGene
116 (NovoGene, Tianjin, China). Genome sequences of the American species *A. coerulea* was obtained from
117 Phytozome v12.1 (Filiault *et al.*, 2018, https://phytozome.jgi.doe.gov). All data generated from the study
118 were submitted to EBI under the accession number PRJEB34182.

119

120 **Sequence assembly, population genetic structure and functional annotation**

121 Whole genome sequences of each accession were aligned against the reference genome of *A. coerulea*
122 using Burrows-Wheeler Aligner (BWA) (Li & Durbin, 2009). Raw assemblies were realigned using
123 IndelRealigner provided in the Genome Analysis Tool Kit (Mckenna *et al.*, 2010). Single nucleotide
124 polymorphisms (SNPs) and insertions/deletions (INDELs) were reported using SAMtools (Li *et al.*, 2009).
125 Only the high-quality variants (SNPs and INDELs) (read depth > 3 and mapping quality > 20) were
126 retained for subsequent population genomics analyses. Genomic context and predicted functional
127 effects of the identified variants were reported for each of the 36 samples separately. Intergenic
128 variants occurring within 5 kb up- and down-stream of transcription start site (TSS) or transcription end
129 site (TES) were defined as regulatory variations. Functional annotation of each identified variant was
130 performed using SnpEff (Cingolani *et al.*, 2012).

131 To infer the phylogenetic relationships of the ten *Aquilegia* species, neighbor-joining (NJ) trees were
132 reconstructed for each chromosome and the whole genome dataset using MEGA 7 (Kumar *et al.*, 2016).
133 Principal component analysis (PCA) was also carried out to obtain the clustering pattern of the 36
134 *Aquilegia* accessions (Zheng *et al.*, 2012). Ancestral components were estimated using ADMIXTURE
135 (Alexander & Novembre, 2009) with the assumed population number ranging from one to five. Optimal
136 population composition with the lowest cross-validation error was selected to account for ancestry. To

137 obtain the genome-wide nucleotide variation pattern, nucleotide diversity (π) and genetic
138 differentiation (F_{ST}) were calculated for each 100 kb non-overlapping sliding window using VCFtools
139 (Danecek *et al.*, 2011). Pair-wise non-synonymous (d_N)/synonymous (d_S) ratios of the ten species were
140 inferred by yn00 program in the Phylogenetic Analysis by Maximum Likelihood (PAML) package (Yang
141 2007). Candidate genes that showed the top 5% highest and lowest d_N/d_S values were supposed to
142 undergo positive and purifying selection, respectively.
143
144 **Cytosine methylation pattern and epigenetic population structure**
145 Whole genome bisulfite sequencing data were pre-processed using TrimGalore (https://www.
146 bioinformatics.babraham.ac.uk/projects/trim_galore/, accessed August 21, 2018). Trimmed paired-end
147 reads were aligned to the reference genome of *A. coerulea* using Bismark (Krueger & Andrews, 2011)
148 with a moderately stringent minimum-score function (L,0,-0.3). De-duplicated alignments of the 36
149 *Aquilegia* accessions were used to report cytosine methylation level using bismark_methylation_
150 extractor, on loci with a read depth > 3. Genomic annotations of the methylated cytosine site were
151 identified based on the reference genome using an in-house Python script. Overall CG methylation level
152 for each accession was calculated by pooling the methylation on all CG site. PCA was conducted to infer
153 the genome-wide CG methylation clustering pattern of the ten *Aquilegia* species. Differential cytosine
154 methylation was determined at both the gene and chromosome levels, respectively. At the gene level,
155 we determined differentially methylated region (DMR) for each 100 bp non-overlapping sliding window
156 using Cochran-Mantel-Haenszel (CMH) test to account for imbalanced read depth (Supplementary
157 Information). Only the genomic regions that possess Benjamini-Hochberg adjusted p value < 0.05 and
158 show inter-specific or inter-lineage methylation divergence higher than 25% are defined as significant
159 DMRs. Genes with > 20% of the genic region being DMR(s) were defined as differentially methylated
160 gene (DMG). Chromosome-level methylation pattern was measured by chromosomal methylation
161 discrepancy index (MDI) according to (O'Sullivan *et al.*, 2016). Methylation patterns of the identified
162 DMGs were visually verified before biological interpretation on Integrative Genomics Viewer
163 (Thorvaldsdottir *et al.*, 2013). In addition, we identified CG islands from the *A. coerulea* reference
164 genome using EMBOSS cpgplot with default settings (Rice *et al.*, 2000). Only the identified CG-enriched
165 genomic regions with > 200 bp were defined as CG islands. We then investigated inter-specific and inter-
166 lineage methylation divergence in and around the CG islands.
167
168 **Associations between the genetic variation and cytosine methylation**

169 To address how the genetic and epigenetic variations interact to promote species diversification, we
170 tested for associations between the identified DMGs and genes under positive selection by Chi-square
171 test. Linear regression model was adopted to measure the direct causal effect of CG-loss variation on CG
172 methylation *in situ*. To further assess whether genetic variations drive the establishment of DMG,
173 driving mutations of DMRs between the *A. japonica* and *A. oxysepala* were identified using the
174 Eigenstrat method (Supplementary Information) (Price *et al.*, 2006). Given that only a limited number of
175 DMGs were identified, differentially stringent cutoffs (5×10^{-5}, 5×10^{-8}, and 5×10^{-11}) were adopted to
176 identify the driver mutations.
177
178 **Identification of conservative clade-specific variant and functional enrichment analysis**
179 Conservative clade-specific variations (CCVs) were defined as these that were conserved across all
180 samples belonging to the same species/lineage but not present in any sample of the other
181 species/lineages. Since the biological consequences of heterozygous variants are less affirmable, only
182 the homozygous point mutations and INDELs were considered in the characterization of CCVs, including
183 frameshift, stop-gain, stop-loss, start-loss and splicing-alteration variations. Genes carrying CCVs are
184 supposed to undergo relaxed selection during the adaptation process. The foregoing genetic and
185 epigenetic analyses identified candidate genes that might be associated with the rapid diversification of
186 the *Aquilegia* species. Candidate genes identified at the genetic and epigenetic levels were employed to
187 conduct enrichment analyses using R package topGO (Alexa *et al.*, 2006). Enriched GO terms that
188 possessed a p value <0.05 and occurred at least five times in the background were considered
189 statistically significant. Structures of functional domains of targeted genes were determined based on
190 the InterPro database (https://www.ebi.ac.uk/interpro, accessed January 25, 2019). Distribution pattern
191 of the identified candidate genes and their related functional pathways was visualized using the R
192 package Jvenn (Baidou *et al*., 2014).
193
194 **Results**
195 **Population structure and nucleotide variation pattern**
196 NJ trees were reconstructed for the 36 *Aquilegia* accessions based on 689,123 homozygous SNPs. The
197 established phylogenies show that accessions of the ten species are broadly clustered as three distinct
198 lineages corresponding their geographic origins (**Figure 1a** and **Figure S1**). In brief, all the 22 accessions
199 of the four East Asian species (*A. oxysepala*, *A. japonica*, *A*. *yabeana* and *A*. *viridiflora*) cluster as a
200 monophyletic lineage, with the former two species and their hybrid forming a clade and the latter two

201 species grouping as a sister clade. In contrast, the West Asian species *A. fragrans* clusters with
202 geographically adjoining European species. Within the North American lineage, phylogenetic
203 relationships of the four species vary across the seven chromosomes. In particular, one European
204 accession of the *A. alpina* var. *alba* falls into the North American lineage at the chromosomes 1 and 3.
205 Likewise, the PCA and population structure inferences also revealed distinct genetic structure of the
206 three phylogenetic lineages (**Figure 1b** and **c**). Consistent with the above phylogenies, the *A. alpina* var.
207 *alba* accession shares the same ancestral genetic cluster with the North American lineage and the
208 putative hybrid of the *A. oxysepala* and *A. japonica* possesses a mixed genomic constitution of its
209 parental species (**Figure 1b** and **c**).
210 To examine whether specific genomic architectures are involved in the adaptive speciation in the
211 three lineages, we assessed the nucleotide diversity (π) and genetic divergence (F_{ST}) at both the intra-
212 and inter-lineage levels. At the inter-lineage level, the Asian columbine species harbored relatively
213 higher nucleotide diversity compared to the European and North American lineages (**Figure S2**). Notably,
214 the genetic divergence between the European and North American lineages is obviously higher than
215 those of the Asian-to-European and Asian-to-North American (**Figure S3**). At the intra-lineage level,
216 while both the *A. oxysepala* and *A. japonica* possess lower nucleotide diversity, apparently higher overall
217 genetic divergence (excepting the chromosome 4) is found compared to any of the inter-lineage
218 comparisons (**Figure S2** and **S3**). We also identified in total 325 high divergence genomic regions (HDGRs,
219 5% highest F_{ST}) and 241 low diversity genomic regions (LDGRs, 5% lowest π) from the inter-lineage
220 comparisons. Among these candidate genomic regions, 96 (29.5%) HDGRs and 116 (48.1%) LDGRs are
221 common between either two or three of these lineages (**Figure S4a** and **b**).
222
223 **Identification of the selection genes and highly impactful genetic variations**
224 Candidate genes associated with the adaptive divergence at the intra- and inter-lineage levels were
225 determined according to the following three strategies. First, genes that localized within the HDGR and
226 LDGR are considered to be exposed to selection in the adaptive evolution process. Between the *A.*
227 *oxysepala* and *A. japonica*, we identified seven 100-kb genomic regions that show overlaps between the
228 HDGRs and LDGRs (**Table S2**). Genes within these genomic regions are functionally associated with the
229 meiotic nuclear division, adenine methyltransferase and basic cellar activities. At the inter-lineage level,
230 candidate genes identified from the HDGRs and LDGRs play important roles in telomere maintenance
231 (e.g., telomeric single stranded DNA binding), plant growth (e.g., gibberellin receptor GID1) and flower
232 development (e.g., AGAMOUS MADS-box protein) (**Table S2**). In addition, we also find that some

233 photosynthesis related genes resided in two adjacent HDGRs (chr6:23.9-24.1Mb), including the
234 cytochrome C protein, NADH-ubiquinone oxidoreductase, photosystem I *ycf3* and photosystem II *PsbT*
235 (**Table S2**). Of significance, a majority of these candidate genes possess fixation sites at the intra- and
236 inter-lineage comparisons (**Supplementary file 1**).

237 The above assessment mainly relied on genome-wide scanning for 100-kb non-overlapping sliding
238 window, we next employed site-based approach to identify highly impactful CCVs (e.g., frameshift
239 mutations, stop mutation and splicing alteration) from both the intra- and inter-lineage comparisons.
240 Our results reveal that a considerable proportion (17.9-40.5%) of the CCVs are identified at the gene
241 body region (**Table S3**). We then examined the impacts of these identified CCVs on the gene function.
242 Between the *A. oxysepala* and *A. japonica*, the CCV-carrying genes are enriched in several vital biological
243 pathways related to cell reproduction, including telomere maintenance, DNA repair, and DNA helicase
244 activity (**Figure 2** and **Table 1**). For example, two candidate genes (*Aqcoe6G160300* and
245 *Aqcoe7G062500*) coding *Xklp2* (*TPX2*) possess species-specific stop-gain variations, which might largely
246 influence spindle assembly during the mitotic process (Zhang & Dawe, 2011; Aguirre-Portoles *et al.*,
247 2012). At the inter-lineage level, the CCVs-harboring genes are also functionally involved in the mitotic
248 chromosome condensation, DNA ligase activity and aminopeptidase activity (**Figure 2** and **Table 1**). Two
249 CCV-containing genes (*Aqcoe2G276600* and *Aqcoe1G273400*) encoding DNA mismatch repair proteins
250 *MutS*/*MSH* and *MutS2* (Fukui *et al.*, 2008) carry one Asian-specific-to-American frameshift variant.

251 In addition, we also performed genome-wide scanning to identify selection genes from the intra-
252 and inter-lineage comparisons. Our analyses characterized a total of 4,347 high and 3,566 low d_N/d_S
253 ratio genes, with 1,773 (40.5%) and 1,584 (44.4%) of the candidate genes sharing between at least two
254 of the six comparisons (**Figure S5**). Functional analysis of the positive selection genes revealed significant
255 enrichment in various important pathways, including cell reproduction (e.g., telomere maintenance),
256 stress tolerance (e.g., defense to fungus and bacterium), and plant development (e.g., response to auxin
257 and cell growth) (**Figure S6**). Likewise, genes under purifying selection are enriched in diverse
258 functionally important categories related to cell reproduction (e.g., DNA recombination and mitotic cell
259 cycle), photosynthesis (e.g., photosystem I reaction center) and basic cellular activities (**Figure S6**).

260

261 **CG methylation patterns and differentially methylated genes**

262 In parallel with the above genomic analyses, we also investigated CG methylation pattern of the
263 representative columbine species. Overall, the North American accessions (median CG methylation
264 level: 91.99%) show significantly higher CG methylation than the European (90.49%, Wilcoxon rank sum

265 test FDR = 4.1×10^{-3}) and Asian (90.67%, Wilcoxon rank sum test FDR = 2.0×10^{-4}) accessions (**Figure 3a**
266 and **Table S1**). On the contrary, the difference in CG methylation is not significant between the
267 European and Asian lineages (FDR = 0.17). To justify pooling samples from the same species/lineage in
268 downstream analyses, we performed PCA-based clustering to examine CG-cytosine methylation
269 similarity of all the columbine accessions. The resulting overall methylation pattern highly resembled the
270 above genomic inferences, with the European and American species falling into two independent groups
271 while the four Asian species forming three separated clusters (**Figure S7**). Thus, we assessed the CG
272 methylation pattern for the European and North American lineages as well as the three Asian species (*A.*
273 *japonica*, *A. oxysepala* and *A. viridiflora*) separately. Consistent with aforementioned genomic features,
274 heterogeneous pattern of the CG methylation is also observed among the seven chromosomes, with the
275 chromosome 4 demonstrating obviously higher overall CG methylation divergence compared to the
276 other six chromosomes (**Figure 3b**). Then, we further evaluated how CG methylation is deposited at the
277 genic, putative *cis*-regulatory and CG island regions, respectively. At the genic and regulatory regions, all
278 three lineages share similar modification pattern with apparent depletion of CG methylation around the
279 TSS and TES (**Figure 3c**). However, the American lineage exhibits hyper-methylation (more than 10%)
280 around the center of CG islands and a more dramatic decrease in the CG island shores compared to the
281 European and Asian species (**Figure 3d**).
282 To examine the biological impacts of differential CG methylation, DMRs and DMGs were identified
283 from both the intra- and inter-lineage comparisons, respectively (**Tables S4** and **S5**). At the intra-lineage
284 level, 3,622 DMRs corresponding to 2,899 DMGs were identified between the *A. japonica* and *A.*
285 *oxysepala*. Functional enrichment of these DMGs indicate that the two species may have different
286 activities in photosynthesis-related pathways, including photosystem I, photosynthesis and chloroplast
287 pathways (**Figure 2**). For example, two photosynthesis-related genes, *PsaA*/*PsaB* and *CemA*, showed
288 significantly differential methylation between the two species at the genic regions (**Figure S8a** and **b**). At
289 the inter-lineage level, apparently more DMGs were identified between the North American and
290 European species (6,087 genes) compared to those of between the two lineages and Asian species
291 (3,308-5,003 genes) (**Table S4** and **S5**). DMGs characterized from the inter-lineage comparisons are
292 mainly involved in the plant growth (e.g., response to auxin) and defense, response to biotic stimulus
293 and wounding (**Figure 2**). It should be noted that while some of the identified DMGs share by at least
294 two of the six comparisons (**Figure S9**), candidate genes identified from the genetic and epigenetic levels
295 show obviously complementary patterns in the functional enrichment analyses (**Figure 2**).

296 It should be noted that the above genetic and epigenetic assessments identified similar enrichment
297 pathways from the intra- and inter-lineage comparisons, particularly those related to cell reproduction
298 and basic cellular activities (**Figure 2**, **Figure S6** and **Table S2**). We then examined how the candidate
299 genes (CCV-carrying gene, high and low d_N/d_S value gene, and DMG) distribute among the above distinct
300 strategies. While a considerable proportion of the candidate genes share among the intra- and inter-
301 lineage comparisons for each of the genetic and epigenetic assessments (**Figure S4** and **S5**), the four
302 types of candidate gene show a clearly separated distribution pattern in the Venn diagrams (**Figure S10**).
303 Likewise, the GO enrichment analyses also revealed complementary patterns among the four types of
304 candidate genes (**Figure 2** and **Figure S6**).
305
306 **Dependence of the epigenetic variability on genetic variations**
307 Both genetic variations and differential CG methylation seemed to have crucial and multifaceted
308 influences on the adaptation of the ten *Aquilegia* species. We wondered whether differential epigenetic
309 modifications are dependent on genetic variability. Among the 588,659 CG sites examined, 224,222
310 (38.09%) carried CG-loss variation(s). We then illustrated epigenetic variability for the mutation-carrying
311 and non-variant CG sites, respectively. As shown in the Figure 4, different levels of genetic-epigenetic
312 association are observed in the two types of CG sites. For example, the variation-carrying CG site
313 conveys information that highly reassembles their genetic background, with the overall methylation
314 pattern showing highly conservativity at the intra-specific level but exhibiting obvious divergence across
315 the ten columbine species (**Figure 4a**). In contrast, CG methylation divergence at the non-variant sites
316 varies with stronger variability at both the intra- and inter-specific levels (**Figure 4b**). By examining the
317 correlation of genetic variability and cytosine methylation, we find that CG methylation divergence at
318 variation-carrying CG-site is largely attributable to the CG-loss variations (**Figure 4c**). In particular, 75% of
319 the CG-loss variations occurring at the most highly variable CG-methylated dinucleotides can explain at
320 least 75% of the total epigenetic variability *per se*. Nevertheless, there is still a considerable proportion
321 of epigenetic variability that can not be sufficiently explained by the variant-CG site (**Figure 4d**).
322 Methylation at non-variant CG site could still reflect population structure, such as the neutralized
323 methylation levels in the hybrid possibly arising from heterozygosity, though with increased noise and
324 slightly compromised accuracy (**Figure 4b**). For instance, CG methylation pattern in the hybrid is
325 neutralized possibly due to heterozygosity. In addition, one *A. japonica* accession (A. japonica9, **Table**
326 **S1**) inhabiting low altitude environmental niche shows differential methylation at some CG site
327 compared to the other con-specific accessions (**Figure 4a** and **b**).

328 It is arguable that non-variant CG site possesses species-specific methylation pattern that is either
329 associated with adjacent CG-loss variations or some other non-CG-loss variations. Therefore, we have
330 attempted to identify driving mutations for each of the 1,229 DMRs between the *A. japonica* and *A.*
331 *oxysepala* (**Supplementary Information**). Our results reveal that only 568 out of the 1,229 (46.2%) DMRs
332 are significantly associated with point mutation inside or around a 500 bp upstream/downstream
333 window under the least stringent *p* value threshold (5×10^{-5}), indicating that the epigenetic changes are
334 only partially dependent on or consequences of *cis*-genetic mutations (**Figure 4e**). Notable, we observed
335 significant associations between differential CG methylation and positive selection at some specific
336 genes. In most inter-lineage comparisons, DMGs are significantly more prone to be under positive and
337 purifying selection than non-DMGs (**Table 2**), suggesting that epigenetic modifications could probably
338 assist selection pressures in shaping genotypes. In contrast, DMGs are significantly less prone to be
339 under purifying selection (**Table 2**).
340
341 **Discussion**
342 **Genetic-driven mechanisms underpinning the diversification of *Aquilegia* species**
343 Elucidating the extrinsic and intrinsic factors underpinning species diversification is crucial to
344 understanding the evolution and persistence of biodiversity (Simpson & Olson, 1953; Schluter, 2000;
345 Losos, 2010). The genus *Aquilegia* provides an ideal system to address how the diverse evolutionary
346 mechanisms promoted rapid adaptive radiation (Fior *et al.*, 2013; Filiault *et al.*, 2018). Yet, various
347 external conditions related to ecological opportunities, such as shifts in pollinator and habitat, are
348 supposed to facilitate the evolution of reproductive isolation (Kramer & Hodges, 2010; Li *et al.*, 2014),
349 genetic basis underlying the rapid diversification of *Aquilegia* species remained largely unclear. In
350 particular, it is still unclear whether the adaptive radiations of North American and European lineages
351 are caused by distinct extrinsic triggers or determined by lineage-specific genetic and epigenetic
352 architectures. In this study, we surveyed the genomes of ten worldwide columbine species to address
353 genetic basis underpinning the rapid species diversification. Broadly consistent with previously inferred
354 phylogenies (Bastida & Herrera, 2010; Fior *et al.*, 2013; Li *et al.*, 2014; Li *et al.*, 2019), the ten columbine
355 species formed three highly divergent lineages corresponding to their geographic origins, with the Asian
356 species being separated into three distinct clades but the phylogenetic relationships of European and
357 North American species varying obviously among the seven chromosomes. These genomic features
358 allow us to address whether specific genomic architectures contributed to the repeated adaptive
359 speciation in the genus *Aquilegia*.

360 It has been proposed that if intrinsic factor is the potential determinant promoted adaptive species,
361 one would expect to identify specific genetic architectures in the diversified lineages (Brawand *et al.*,
362 2014; Machado-Schiaffino *et al.*, 2014; Lamichhaney *et al.*, 2015). In Darwin's finches, for example,
363 polyphyletic topology is observed as a general pattern in the 14 morphologically distinct species.
364 Phenotypic diversity of the beak shape is mainly determined by natural selection acting on the *ALX1*
365 gene during the ecological specialization process (Lamichhaney *et al.*, 2015). A similar phenomenon is
366 observed in the East African cichlid fish wherein the radiating lineages are more dynamic in terms of
367 gene content and transcriptomic landscape compared to their non-radiating relatives (Brawand *et al.*,
368 2014; Machado-Schiaffino *et al.*, 2014). In *Aquilegia*, our previous studies have demonstrated that
369 natural selection and genetic drift together resulted in the rapid evolution of reproductive isolation
370 between the *A. japonica* and *A. oxysepala* (Li *et al.*, 2014; Li *et al.*, 2019). Here we further revealed that
371 candidate genes involved in the adaptive speciation are functionally enriched in categories related to
372 cell reproduction (e.g., telomere maintenance), stress tolerance (e.g., response to wounding) and basic
373 cellular activities. More importantly, a considerable proportion of the candidate genes (e.g., CCVs and
374 d_N/d_S) and enrichment pathways share at the intra- and inter-specific comparisons, particularly those of
375 related to cell reproduction, stress tolerance, and plant growth. Given that these functionally important
376 pathways are potentially correlated with adaptive traits, we propose that specific genetic architectures
377 are the intrinsic factor conferred adaptability to the *Aquilegia* species to respond to diverse external
378 conditions.
379 It should be noted that while both the North American and European columbines originate from
380 ancestral Asian species, shifts in pollinator and ecological habitat are supposed to drive the adaptive
381 radiation of the two lineages, respectively (Bastida & Herrera, 2010; Fior *et al.*, 2013; Filiault *et al.*,
382 2018). One hypothesis is that the ancestral Asian species carry lineage-specific genetic architectures,
383 which offered the North American and European species adaptabilities to cope with pollinator and
384 ecological specializations. With this reasoning, lineage-specific genetic variants identified in the North
385 American and European columbines should be functionally correlated with flower development and
386 ecological specialization. Alternatively, independent adaptive radiations of the North American and
387 European lineages are mainly caused by different ecological opportunities in respective geographic
388 regions. Phenotypic diversity of the North American and European species is potentially determined by
389 natural selection acting on similar genetic architectures. In our study, a considerable of the candidate
390 genes and their enrichment pathways share between the two lineages and Asian species, suggesting the
391 possibility that similar genetic architectures contributed to the adaptive radiations of the North

392 American and European columbine species. Nevertheless, we find that majority of the functionally
393 important pathways show complementary enrichment pattern in the inter-lineage comparisons, clearly
394 indicating that lineage-specific differences play crucial roles in the adaptive radiations of the North
395 American and European species. More importantly, a part of the identified lineage-specific genetic
396 differences (e.g., CCVs and d_N/d_S) are associated with adaptive divergence of the three lineages. For
397 example, four tandem duplicates of the AGAMOUS MADS-box gene exhibit high genetic divergence
398 among the North American, European and Asian lineages, particularly that three copies harbor multiple
399 lineage-specific non-synonymous substitutions. The MADS-box AGAMOUS gene plays critical roles in the
400 determination of reproductive floral oranges (e.g., carpel and stamen), floral meristem determinacy and
401 flowering transition (Kramer *et al.*, 2004; Tapia-Lopez *et al.*, 2008; Dreni & Kater, 2014). It indicates that
402 the four AGAMOUS genes are possibly associated with the floral development and reproductive
403 isolation of the columbine species. Taken together, our findings indicate that both the lineage-specific
404 and shared genetic architectures are the intrinsic factor interacts with extrinsic trigger to determine the
405 adaptive radiations of the North American and European species.
406
407 **Evolutionary potential of cytosine methylation in the adaptation of *Aquilegia* species**
408 The role of epigenetic modification in the long-term evolutionary process has long been debated
409 (Bossdorf *et al.*, 2008; Diez *et al.*, 2014; Verhoeven *et al.*, 2016). It has been noted that epigenetic
410 variations are usually under the genetic control and vary rapidly as a result of environmental induction
411 and stochastic epimutation (Richards *et al.*, 2010; Richards *et al.*, 2017). Nevertheless, it has been
412 recently recognized that some epigenetic variants can persist over generations and be highly correlated
413 with adaptive phenotypes (Verhoeven *et al.*, 2016). As illustrated in Arabidopsis, changes in cytosine
414 methylation can produce meiotically stable epialleles, which could eventually lead to phenotypic
415 diversity in the absence of genetic variations (Becker *et al.*, 2011; Schmitz *et al.*, 2011; Cortijo *et al.*,
416 2014). In this study, we aimed to assess whether the epigenetic modification also contributed the
417 adaptive speciation of the *Aquilegia* species. Consistent with the genomic features detailed above,
418 clearly high divergence of cytosine methylation was observed among the Asian, European and North
419 American lineages, particularly that differential cytosine methylation is not only found among the seven
420 chromosomes but also evident in the gene body and CG island regions. These observations indicate that
421 cytosine methylation may have contributed to the adaptability of *Aquilegia* species. Between the *A.*
422 *japonica* and *A. oxysepala*, for example, candidate genes identified at the genetic level are functionally
423 related to cell production, stress tolerance and basic cellular activities. At the cytosine methylation level,

424 the identified DMGs show significant enrichment in categories related to diverse important phenotypic
425 traits, such as photosynthesis, plant growth and development. Given that adaptive speciation of the two
426 species is mainly caused by ecological specialization (Li *et al.*, 2014; Li *et al.*, 2019), we then propose that
427 epigenetic modification is also a complementary mechanism facilitated phenotypic divergence between
428 the two species. A similar phenomenon was also observed from the inter-lineage comparison where
429 differences in cytosine methylation are associated with adaptive phenotypes such as the plant growth
430 and stress tolerance. These findings indicate that both the genetic and epigenetic architectures interact
431 to respond to diverse extrinsic conditions.

432 We noted that some DMGs and enriched pathways are commonly found from the genetic and
433 epigenetic comparisons, especially for those that related to plant growth and stress tolerance. We then
434 asked whether the epigenetic mechanism works independently or largely under the control of genetic
435 variability. Intuitively, genetic variations are important drivers of epigenetic variability. Many studies
436 based on human and mouse have shown genetic variations can not only directly control CG methylation
437 patterns (Drong *et al.*, 2013; Smith *et al.*, 2014), but also manipulate long-range CG methylation at
438 specific sites to further influence phenotypes, where CG methylation serves as a mediator (Fisher *et al.*,
439 2018; Li *et al.*, 2018). However, precedent studies on explaining, in a time-span of evolutionary history,
440 whether CG methylation functions independent of genetic variations are few to our knowledge. By
441 analyzing the associations between genetic and epigenetic variability, we find that while the CG-loss
442 variation can directly lead to the depletion in CG methylation level, there still exist a considerable of
443 DMRs not attributable to *cis*-variation. Though potential non-linear additivity and possible existence of
444 distal driver mutations beyond the 1kb window of DMRs may introduce biases, our findings at least
445 suggest that the epigenetic variability also possesses autonomy to some extent. Furthermore, since
446 gene body CG methylation in plants generally stabilizes gene expression and is positively correlated with
447 gene expression level (Takuno & Gaut, 2013; Bewick *et al.*, 2016; Bewick & Schmitz, 2017; Zilberman,
448 2017), differential methylation in our study is indicative of the amount of eventual gene products. Based
449 on these attributes together with the significant association between differential methylation (DMGs)
450 and positive selection (d_N/d_S), we propose that the intrinsic epigenetic variability and the extrinsic
451 selection pressures are intertwined, where differential methylation could assist selection pressures.
452

453 **Conclusions**

454 Elucidating how the intrinsic and extrinsic mechanisms promoted species diversification is critical to
455 understanding the evolution and persistence of biodiversity. Here, we took the advantage of the model

system *Aquilegia* to address genetic and epigenetic bases underlying the rapid species radiation. Our results demonstrated that intrinsic factors related to specific genetic and epigenetic architectures conferred adaptability to the *Aquilegia* species in response to diverse external conditions, and that the epigenetic modification is a complementary mechanism that assists the selection pressure to shape the phenotypic diversity. Our study provides a genome-wide view of how the high adaptability of *Aquilegia* species have evolved in the North American, European and Asian columbine species.

Author Contributions

L.F.L. conceived this project. T.L. and M.R.L. developed statistical analysis pipeline. T.L., M.R.L., N.D., Z.H.W., L.Z.L., and X.G. carried out experiments and analyzed the data. T.L., M.R.L., N.D., Z.H.W., L.Z.L., X.G., and L.F.L. interpreted the data and participated in discussion and wrote the manuscript.

Acknowledgments

We thank Aköz Gökçe for her constructive comments on our manuscript. This work was financially supported by National Natural Science Foundation of China (31670382), Shanghai Pujiang Program (19PJ1401500), Start-up funding at Fudan University (JIH1322105) and the Department of Science and Technology of Jilin Province (20190201299JC).

References

Aguirre-Portoles C, Bird AW, Hyman A, Canamero M, de Castro IP, Malunnbres M. 2012. Tpx2 controls spindle integrity, genome stability, and tumor development. *Cancer Research* **72**(6): 1518-1528.

Alcántara JM, Bastida JM, Rey PJ. 2010. Linking divergent selection on vegetative traits to environmental variation and phenotypic diversification in the Iberian columbines (*Aquilegia*). *Journal of Evolutionary Biology* **23**(6): 1218-1233.

Alexa A, Rahnenfuhrer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**(13): 1600-1607.

Alexander DH, Novembre JK. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**(9): 1655-1664.

Bastida JM, Herrera CM. 2010. Extended phylogeny of *Aquilegia*: the biogeographical and ecological patterns of two simultaneous but contrasting radiations. *Plant Systematics and Evolution* **284**(3-4): 171-185.

Dardou P, Mariette J, Escudie F, Djemiel C, Klopp C. 2014. Jvenn: an interactive venn diagram viewer. *BMC Bioinformatics* **51**(1):293.

Becker C, Hagmann J, Jonas M, Daniel K, Oliver S, Karsten B, Detlef W. 2011. Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* **480**(7376): 245.

Berner D, Salzburger W. 2015. The genomics of organismal diversification illuminated by adaptive radiations. *Trends in Genetics* **31**(9): 491-499.

Bewick AJ, Ji L, Niederhuth CE, Willing EM, Hofmeister BT, Shi X, Wang L, Lu Z, Rohr NA, Hartwig B. 2016. On the origin and evolutionary consequences of gene body DNA methylation. *Proceedings of the National Academy of Sciences of the United States of America* **113**(32): 9111-9116.

Bewick AJ, Schmitz RJ. 2017. Gene body DNA methylation in plants. *Current Opinion in Plant Biology* **36**: 103-110.

Bossdorf O, Richards CL, Massimo P. 2008. Epigenetics for ecologists. *Ecology Letters* **11**(2): 106-115.

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan SH, Simakov O, Ng AY, Lim ZW, Bezault E, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**(7518): 375-381.

Carlquist, Motley TJ. 2003. Tarweeds and Silverswords, Evolution of the Madiinae (*Asteraceae*). *Economic Botany* **58**(1): 123-124.

Cingolani P, Platts A, Le LW, Coon M, Nguyen T, Luan W, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**(2): 80-92.

Cortijo S, Wardenaar R, Colome-Tatche M, Gilly A, Etcheverry M, Labadie K, Caillieux E, Hospital F, Aury JM, Wincker P, et al. 2014. Mapping the epigenetic basis of complex traits. *Science* **343**(6175): 1145-1148.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**(15): 2156-2158.

Diez CM, Roessler K, Gaut BS. 2014. Epigenetics and plant genome evolution. *Current Opinion in Plant Biology* **18**(1): 1-8.

Dreni L, Kater MM. 2014. MADS reloaded: evolution of the AGAMOUS subfamily genes. *New Phytologist* **201**(3): 717-732.

Drong AW, George N, Hedman AK, Eshwar M, Elin G, Small KS, So-Youn S, Bell JT, Fredrik K, Nicole S. 2013. The presence of methylation quantitative trait loci indicates a direct genetic influence on the level of DNA methylation in adipose tissue. *Plos One* **8**(2): e55923.

Filiault DL, Ballerini ES, Mandakova T, Akoz G, Derieg NJ, Schmutz J, Jenkins J, Grimwood J, Shu S, Hayes RD, et al. 2018. The *Aquilegia* genome provides insight into adaptive radiation and reveals an extraordinarily polymorphic chromosome with a unique history. *Elife* **7**.

Fior S, Mingai L, Bengt O, Roberto V, Hodges SA, Lino O, Claudio V. 2013. Spatiotemporal reconstruction of the *Aquilegia* rapid radiation through next-generation sequencing of rapidly evolving cpDNA regions. *New Phytologist* **198**(2): 579-592.

Fisher VA, Wang L, Deng X, Sarnowski C, Cupples LA, Liu CT. 2018. Do changes in DNA methylation mediate or interact with SNP variation? A pharmacoepigenetic analysis. *BMC Genetics* **19**(Suppl 1): 70.

Ford AG, Rüber L, Newton J, Dasmahapatra KK, Balarin JD, Bruun K, Day JJ. 2016. Niche divergence facilitated by fine-scale ecological partitioning in a recent cichlid fish adaptive radiation. *Evolution* **70**(12): 2718-2735.

Fukui K, Nakagawa N, Kitamura Y, Nishida Y, Masui R, Kuramitsu S. 2008. Crystal structure of MutS2 endonuclease domain and the mechanism of homologous recombination suppression. *Journal of Biological Chemistry* **283**(48): 33417-33427.

Garrido JL, Alcántara JM, Rey PJ, Medrano M, Guitián J, Castellanos MC, Bastida JM, Jaime R, Herrera CM. 2017. Geographic genetic structure of Iberian columbines (gen. *Aquilegia*). *Plant Systematics and Evolution* **303**(9): 1145-1160.

Givnish TJ. 1997. Adaptive radiation and molecular systematics: aims and conceptual issues. In: Givnish TJ, Systma KJ, eds. . *Molecular Evolution and Adaptive Radiation.*: 1-54.

Givnish TJ. 2015. Adaptive radiation versus 'radiation' and 'explosive diversification': why conceptual distinctions are fundamental to understanding evolution. *New Phytologist* **207**(2): 297-303.

541 **Irisarri I, Singh P, Koblmüller S, Torresdowdall J, Henning F, Franchini P, Fischer C, Lemmon AR,**
542 **Lemmon EM, Thallinger GG. 2018.** Phylogenomics uncovers early hybridization and adaptive
543 loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nature Communications* **9**(1): 3159-
544 3171.
545 **Kramer EM, Hodges SA. 2010.** *Aquilegia* as a model system for the evolution and ecology of petals.
546 *Philosophical Transactions Biological Sciences* **365**(1539): 477-490.
547 **Kramer EM, Jaramillo MA, Di Stilio VS. 2004.** Patterns of gene duplication and functional evolution
548 during the diversification of the *AGAMOUS* subfamily of MADS box genes in angiosperms.
549 *Genetics* **166**(2): 1011-1023.
550 **Krueger F, Andrews SR. 2011.** Bismark: a flexible aligner and methylation caller for Bisulfite-Seq
551 applications. *Bioinformatics* **27**(11): 1571-1572.
552 **Kumar S, Stecher G, Tamura K. 2016.** MEGA7: Molecular evolutionary genetics analysis version 7.0 for
553 bigger datasets. *Molecular Biology and Evolution* **33**(7): 1870.
554 **Lamichhaney S, Jonas B, Markus S?Llman A, Khurram M, Manfred G, Alvaro MB, Marta P, Carl-Johan**
555 **R, Chao W, Neda Z. 2015.** Evolution of Darwin's finches and their beaks revealed by genome
556 sequencing. *Nature* **518**(7539): 371-375.
557 **Li H, Durbin R. 2009.** *Fast and accurate short read alignment with Burrows-Wheeler transform*.
558 **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc GPD.**
559 **2009.** The sequence alignment/map format and SAMtools. *Bioinformatics* **25**(16): 2078-2079.
560 **Li L, Wang C, Lu T, Lin S, Hu YQ. 2018.** Indirect effect inference and application to GAW20 data. *BMC*
561 *Genetics* **19**(Suppl 1): 67.
562 **Li LF, Pang D, Liao QL, Xiao HX. 2011.** Genomic and EST microsatellite markers for *Aquilegia* flabellata
563 and cross-amplification in *A. oxysepala* (Ranunculaceae). *American Journal of Botany* **98**(8): 213-
564 215.
565 **Li LF, Wang HY, Pang D, Liu Y, Liu B, Xiao HX. 2014.** Phenotypic and genetic evidence for ecological
566 speciation of *Aquilegia japonica* and A. *oxysepala*. *New Phytologist* **204**(4): 1028-1040.
567 **Li MR, Wang HY, Ding N, Lu T, Huang YC, Xiao HX, Liu B, Li LF. 2019.** Rapid divergence followed by
568 adaptation to contrasting ecological niches of two closely related columbine species *Aquilegia*
569 *japonica* and A. *oxysepala*. *Genome Bioligy and Evolution* **11**(3): 919-930.
570 **Losos JB. 2010.** Adaptive radiation, ecological opportunity, and evolutionary determinism. American
571 Society of Naturalists E. O. Wilson award address. *American Naturalist* **175**(6): 623-639.
572 **Losos JB, Ricklefs RE. 2009.** Adaptation and diversification on islands. *Nature* **457**(7231): 830-836.
573 **Machado-Schiaffino G, Henning F, Meyer A. 2014.** Species-specific differences in adaptive phenotypic
574 plasticity in an ecologically relevant trophic trait: hypertrophic lips in Midas cichlid fishes.
575 *Evolution* **68**(7): 2086-2091.
576 **Mckenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D,**
577 **Gabriel S, Daly M. 2010.** The Genome Analysis Toolkit: A map reduce framework for analyzing
578 next-generation DNA sequencing data. *Genome Research* **20**(9): 1297-1303.
579 **Munz. 1946.** Aquilegia: the cultivated and wild columbines. *Gentes Herbarum* **7**: 1-150.
580 **O''Sullivan DE, Johnson KC, Skinner L, Koestler DC, Christensen BC. 2016.** Epigenetic and genetic
581 burden measures are associated with tumor characteristics in invasive breast carcinoma.
582 *Epigenetics* **11**(5): 344-353.
583 **Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, David R. 2006.** Principal components
584 analysis corrects for stratification in genome-wide association studies.
585 **Rice P, Longden I, Bleasby A. 2000.** EMBOSS: The European molecular biology open software suite.
586 *Trends in Genetics* **16**(6): 276-277.

587 **Richards CL, Alonso C, Becker C, Bossdorf O, Bucher E, Colome-Tatche M, Durka W, Engelhardt J,**
588 **Gaspar B, Gogol-Doring A, et al. 2017.** Ecological plant epigenetics: Evidence from model and
589 non-model species, and the way forward. *Ecology Letters* **20**(12): 1576-1590.
590 **Richards CL, Oliver B, Verhoeven KJF. 2010.** Understanding natural epigenetic variation. *New*
591 *Phytologist* **187**(3): 562-564.
592 **Schluter D. 1996.** Ecological causes of adaptive radiation. *American Naturalist* **148**: S40-S64.
593 **Schluter D. 2000.** The Ecology of Adaptive Radiation. *Heredity* **86**(6): 749-750.
594 **Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Ondrej L, Schork NJ, Ecker JR. 2011.**
595 Transgenerational epigenetic instability is a source of novel methylation variants. *Science*
596 **334**(6054): 369-373.
597 **Sharma B, Yant L, Hodges SA, Kramer EM. 2014.** Understanding the development and evolution of
598 novel floral form in *Aquilegia*. *Current Opinion in Plant Biology* **17**(1): 22-27.
599 **Simpson EH. 1949.** Measurement of Diversity. *Nature* **163**(4148): 688-688.
600 **Simpson GG, Olson EC. 1953.** The major features of evolution. *Quarterly Review of Biology* **8**(1): 87-88.
601 **Smith AK, Kilaru V, Kocak M, Almli LM, Mercer KB, Ressler KJ, Tylavsky FA, Conneely KN. 2014.**
602 Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry,
603 developmental stage, and tissue type. *Bmc Genomics* **15**(1): 145.
604 **Takuno S, Gaut BS. 2013.** Gene body methylation is conserved between plant orthologs and is of
605 evolutionary consequence. *Proceedings of the National Academy of Sciences of the United*
606 *States of America* **110**(5): 1797-1802.
607 **Tang LL, Yu Q, Sun JF, Huang SQ. 2007.** Floral traits and isolation of three sympatric *Aquilegia* species in
608 the Qinling Mountains, China. *Plant Systematics and Evolution* **267**(1/4): 121-128.
609 **Tapia-Lopez R, Garcia-Ponce B, Dubrovsky JG, Garay-Arroyo A, Perez-Ruiz RV, Kim SH, Acevedo F,**
610 **Pelaz S, Alvarez-Buylla ER. 2008.** An *AGAMOUS*-related MADS-box gene, XAL1 (*AGL12*),
611 regulates root meristem cell proliferation and flowering transition in *Arabidopsis*. *Plant*
612 *Physiology* **146**(3): 1182-1192.
613 **Thorvaldsdottir H, Robinson JT, Mesirov JP. 2013.** Integrative Genomics Viewer (IGV): high-
614 performance genomics data visualization and exploration. *Brief Bioinformatics* **14**(2): 178-192.
615 **Verhoeven KJ, Vonholdt BM, Sork VL. 2016.** Epigenetics in ecology and evolution: what we know
616 and what we need to know. *Molecular Ecology* **25**(8): 1631-1638.
617 **Wagner CE, Harmon LJ, Seehausen O. 2012.** Ecological opportunity and sexual selection together
618 predict adaptive radiation. *Nature* **487**(7407): 366-U124.
619 **Whittall JB, Hodges SA. 2007.** Pollinator shifts drive increasingly long nectar spurs in columbine flowers.
620 *Nature* **447**(7145): 706-709.
621 **Yang Z. 2007.** PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*
622 **24**(8): 1586-1591.
623 **Zhang H, Dawe RK. 2011.** Mechanisms of plant spindle formation. *Chromosome Research* **19**(3): 335-
624 344.
625 **Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012.** A high-performance computing
626 toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**(24):
627 3326-3328.
628 **Zilberman D. 2017.** An evolutionary case for functional gene body methylation in plants and animals.
629 *Genome Biology* **18**(1): 87.
630

631

632 **Figure legend**

633 **Figure 1**. **Phylogenetic relationship and population structure of the ten worldwide *Aquilegia* species**.
634 (a) Phylogenetic tree of the 36 accessions constructed by neighbor joining algorithm based on whole-
635 genome SNPs. (b) PCA reveals genetic similarity within each of the three lineages and genetic disparity
636 between lineages based on 15,988 LD-pruned SNPs. Curves for each lineage denote 99% probability
637 ellipses estimated from distribution of the first two principal components. (c) Population admixture of
638 the 36 *Aquilegia* accessions. Most accessions exhibit almost pure ancestry composition matching the
639 presupposed evolutionary history of the corresponding species.

640

641 **Figure 2**. **Functional enrichment of genes harboring highly impactful CCVs and DMGs.** CCV-containing
642 genes specific to either of the two lineages/species being compared were merged to construct a target
643 gene set. Ratio denotes proportion of CCV-containing genes or DMGs in the corresponding gene set of
644 interest. Absence of dot indicates no significant enrichment.

645

646 **Figure 3**. **Patterns of cytosine methylation for the ten worldwide *Aquilegia* species.** (a) Genome-wide
647 CG methylation level distributions of 36 accessions. (b) MDI illustrates chromosome-level CG
648 methylation similarity. *Aquilegia viridiflora* was used as the reference. (c) CG methylation profiling in
649 genic region across the four *Aquilegia* groups. Each row represents one genic region starting at 5kb
650 upstream of its TSS and terminating at 5kb downstream of its TES, sorted by mean methylation level of
651 all analyzed CG sites. Length of gene body was normalized. (d) CG methylation profiling in and around
652 CG island.

653

654 **Figure 4**. **Association between CG-loss variations and epigenetic variability**. (a) Top 3,000 most variable
655 CG sites containing CG-loss variations. (b) top 3,000 most variable non-variant CG sites across 36
656 accessions show clade-specific methylation patterns. (c) Linear regression demonstrates that CG-loss
657 variations explain a large proportion of CG methylation variation. Composition of each category is
658 summarized in (d) with regard to whether each CG site contains a CG-loss variation. Epigenetic
659 variability was determined by standard deviation in methylation β value across all 36 accessions. CG
660 sites with top 10,000, 10,001-50,000 and 50,001-150,000 largest standard deviation were ordinally
661 labelled as possessing “very high”, “high” and “moderate” variability respectively. The rest CG sites were
662 labelled as possessing “low” variability. (e) Association test shows most DMRs are independent of *cis*-
663 acting SNPs.

664

665 **Supplementary file**

666 **Figure S1**. Per-chromosome phylogenetic trees reconstructed using neighbor-joining method.
667 Polymorphisms detected on each chromosome were retrieved separately to infer the phylogeny.

668 **Figure S2**. Distribution of nucleotide diversity (π) at the whole-genome level and the per-chromosome
669 level. Nucleotide diversity was estimated for each lineage pooling corresponding species, as well as for
670 *A. japonica* and *A. oxysepala*.

671 **Figure S3**. Distribution of inter-lineage and inter-specific genetic divergence (Fst) at the whole-genome
672 level and the per-chromosome level.

673 **Figure S4**. (a) Overlapping of HGDRs in three inter-lineage comparisons. 148 regions with 5% highest
674 genetic divergence were considered HGDRs in each comparison. 73 regions were shared by two
675 comparisons and 23 regions showed high genetic divergence in all inter-lineage comparisons. (b)
676 Overlapping of LDNRs in three lineages. 148 regions with 5% lowest nucleotide diversity were
677 considered LDNRs in each lineage. 29 regions had low genetic variability in two lineages and 57 regions
678 had low genetic variability in all three lineages.

679 **Figure S5**. Overlapping of genes harboring CCVs (a), or indicating strong positive (b) and purifying (c)
680 selection at inter-lineage/species comparisons. Genes having a high d_N/d_S (among the highest 5% or a
681 5% highest d_N when $d_S = 0$) were considered to be under strong positive selection. Genes having a low
682 d_N/d_S (among the lowest 5% or a 5% highest d_S when $d_N = 0$) were considered to be under strong
683 purifying selection.

684 **Figure S6**. Functional enrichment of genes indicating strong positive selection or purifying selection.
685 Functional enrichment analysis was performed for each inter-lineage/species comparison.

686 **Figure S7**. PCA illustrates three distinct clusters corresponding to the three lineages. Asian species
687 further demonstrated higher inter-specific divergence than the American and the European species. PCA
688 was performed based on 588,659 loci with sufficiently high sequencing quality.

689 **Figure S8**. Illustration of differential methylation in two photosynthesis genes. CG methylation pattern
690 of two genes, Aqcoe7G230600 photosystem I *PsaA*/*PsaB* (a) and Aqcoe7G231300 *CemA* (b) in *A.*
691 *japonica* and *A. oxysepala* throughout the gene body region. Red bars indicate methylation level (0-100)
692 at CG sites. Genomic coordinates on the chromosome 7 are annotated.

693 **Figure S9**. Overlapping of DMGs identified in inter-lineage/species comparisons. A considerable
694 proportion (51.3%) of these DMGs were shared by two or more inter-lineage/species comparisons.

695 **Figure S10**. Venn analyses of the candidate genes under relaxed selection (CCV-carrying), positive
696 selection (high d_N/d_S value), purifying selection (low d_N/d_S value) and differential methylated (DMG).

697 Each subpanel indicates the comparison between the *A. japonica* and *A. oxysepala* (a), *A. japonica* and
698 North American (b), *A. japonica* and European (c), *A. oxysepala* and North American (d), *A. oxysepala*
699 and European (e), North American and European (f).
700

701 **Table 1. Information of the high-impact conservative clade-specific variants (CCVs) in the cell**
702 **reproduction related genes.**

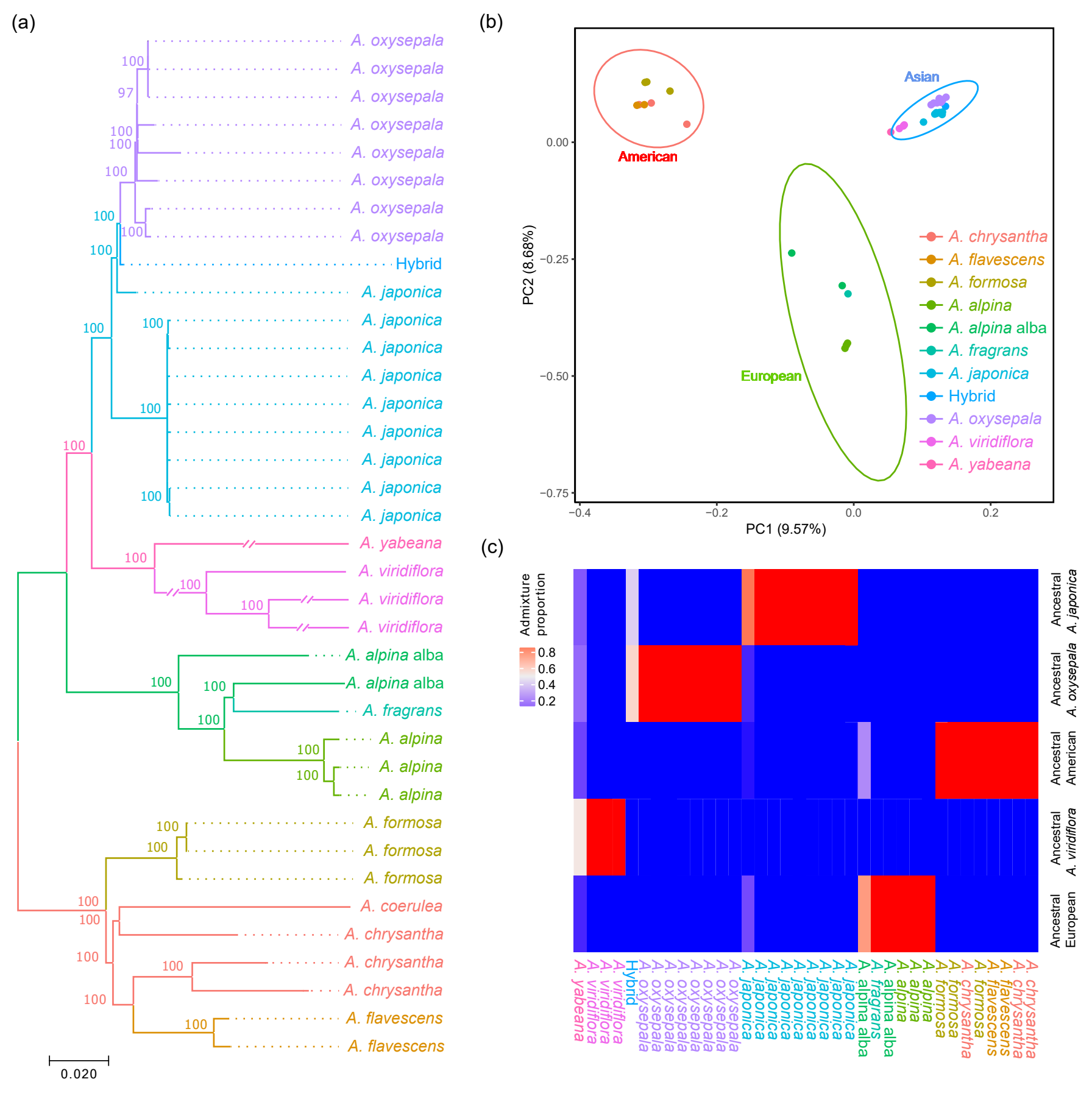
Gene	Variant-carrying	Reference	Chromosome	Position	Reference allele	Variant	Annotation	Gene function
Aqcoe1G273400	Asian	American	Chr1	18994915	GAA	GAAA	frameshift	DNA mismatch repair protein *MutS2*
Aqcoe2G151500	European	American	Chr2	15305837	A	G	splicing	PIF1-like helicase
	European	American		15307442	A	C	stop gain	
	European	American		15309865	AATATATAT	AATATATATAT	frameshift	
	European	Asian		15307442	A	C	stop gain	
	European	Asian		15309865	AATATATAT	AATATATATAT	frameshift	
	A. oxysepala	*A. japonica*		15305837	A	G	splicing	
	A. oxysepala	*A. japonica*		15309267	AT	A	frameshift	
Aqcoe2G177700	European	American	Chr2	21794397	TATGCACCAAAGGTATCACGATGC	TATGC	frameshift	PIF1-like helicase
	European	American		21794979	TT	TTGT	frameshift	
	European	Asian		21794397	TATGCACCAAAGGTATCACGATGC	TATGC	frameshift	
	A. oxysepala	*A. japonica*		21795089	CA	C	frameshift	
Aqcoe6G208600	European	American	Chr6	15364081	A	ATCTCTTCG	frameshift	PIF1-like helicase
	European	Asian		15364081	A	ATCTCTTCG	frameshift	
	A. japonica	*A. oxysepala*		15364330	TAA	TA	frameshift	
Aqcoe6G253800	European	American	Chr6	22789898	C	T	stop gain	DNA helicase
	European	American		22790012	G	A	splicing	
	European	Asian		22789898	C	T	stop gain	
	A. japonica	*A. oxysepala*		22790012	G	A	splicing	
Aqcoe2G276600	Asian	American	Chr2	33314422	AGGGGG	AGGGGGG	frameshift	DNA mismatch repair protein *Msh6*
Aqcoe6G160300	*A. japonica*	*A. oxysepala*	Chr6	9414625	G	A	stop gain	*TPX2*
Aqcoe7G062500	*A. oxysepala*	*A. japonica*	Chr7	3789055	G	A	stop gain	cell cycle regulated microtubule associated protein

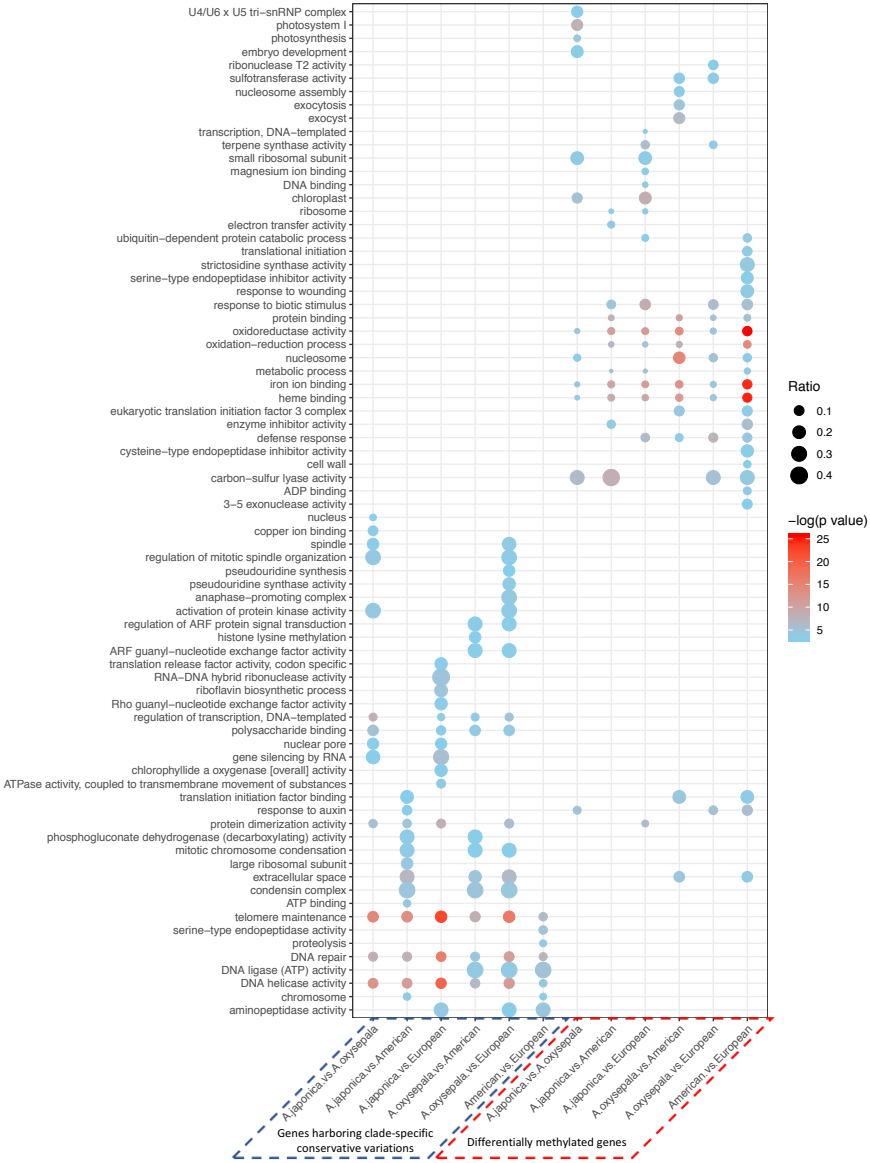
703
704
705

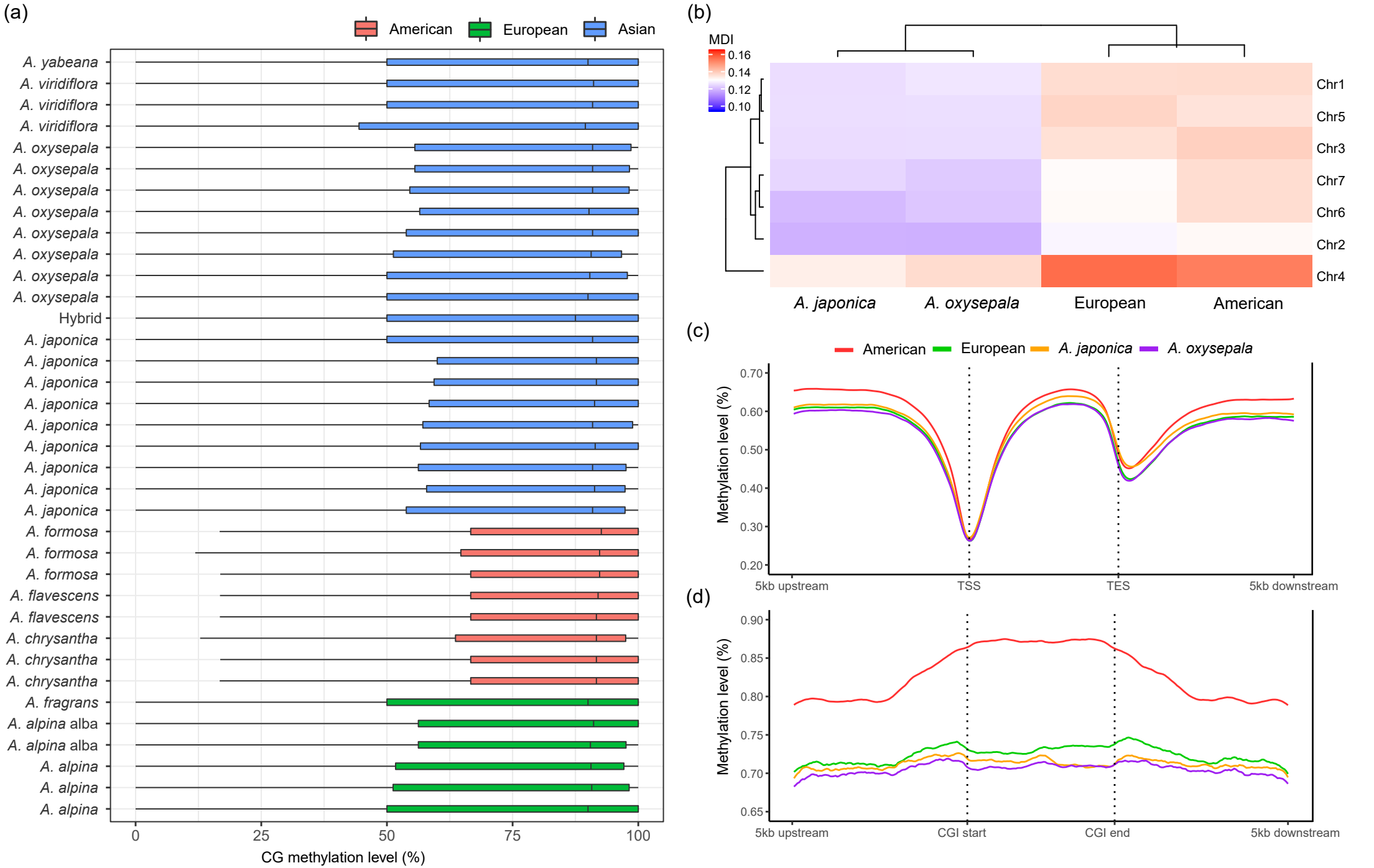
706 **Table 2. Correlation test between the differential methylation and d_N/d_S value for the overall genes.**
707

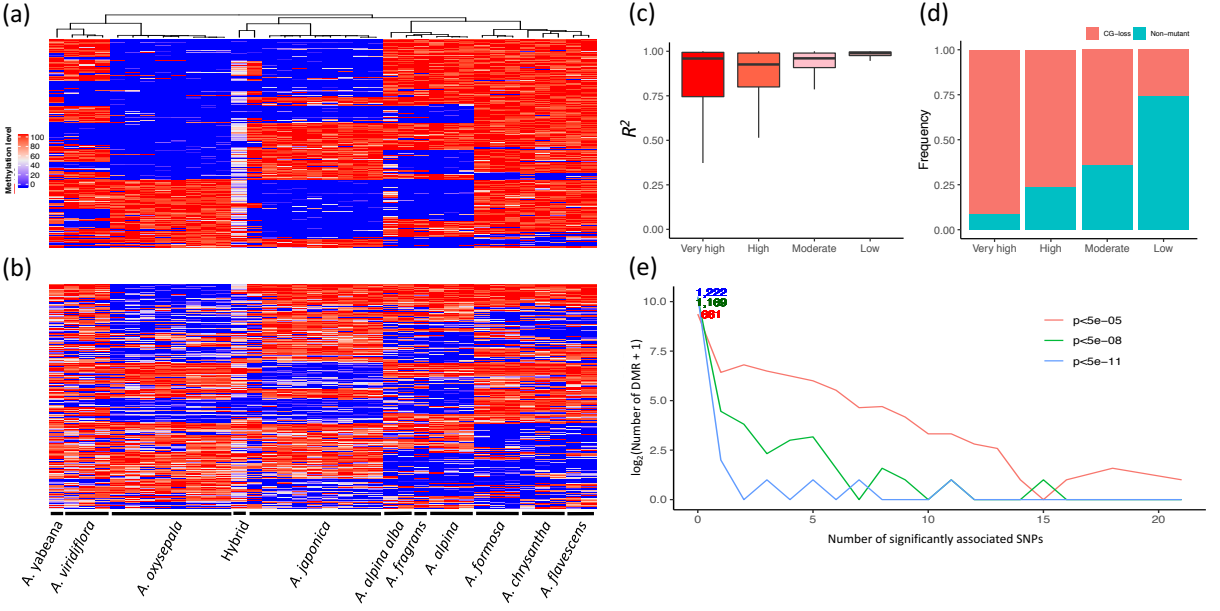
Type of selection	Differential methylation	Jap-Oxy*	Jap-Ame	Jap-Eur	Oxy-Ame	Oxy-Eur	Ame-Eur
Positive selection	DMG	7.2%	7.3%	11.9%	6.7%	8.4%	8.9%
	non-DMG	4.4%	5.0%	5.6%	4.7%	5.4%	5.4%
	p value	0.11	7.3e-02	3.9e-05	6.7e-02	1.8e-02	2.8e-04
Purifying selection	DMG	3.1%	1.8%	2.4%	2.3%	2.0%	1.9%
	non-DMG	4.3%	4.3%	4.7%	4.9%	5.1%	4.0%
	p value	0.53	3.2e-02	9.1e-02	1.3e-02	8.4e-03	1.0e-02

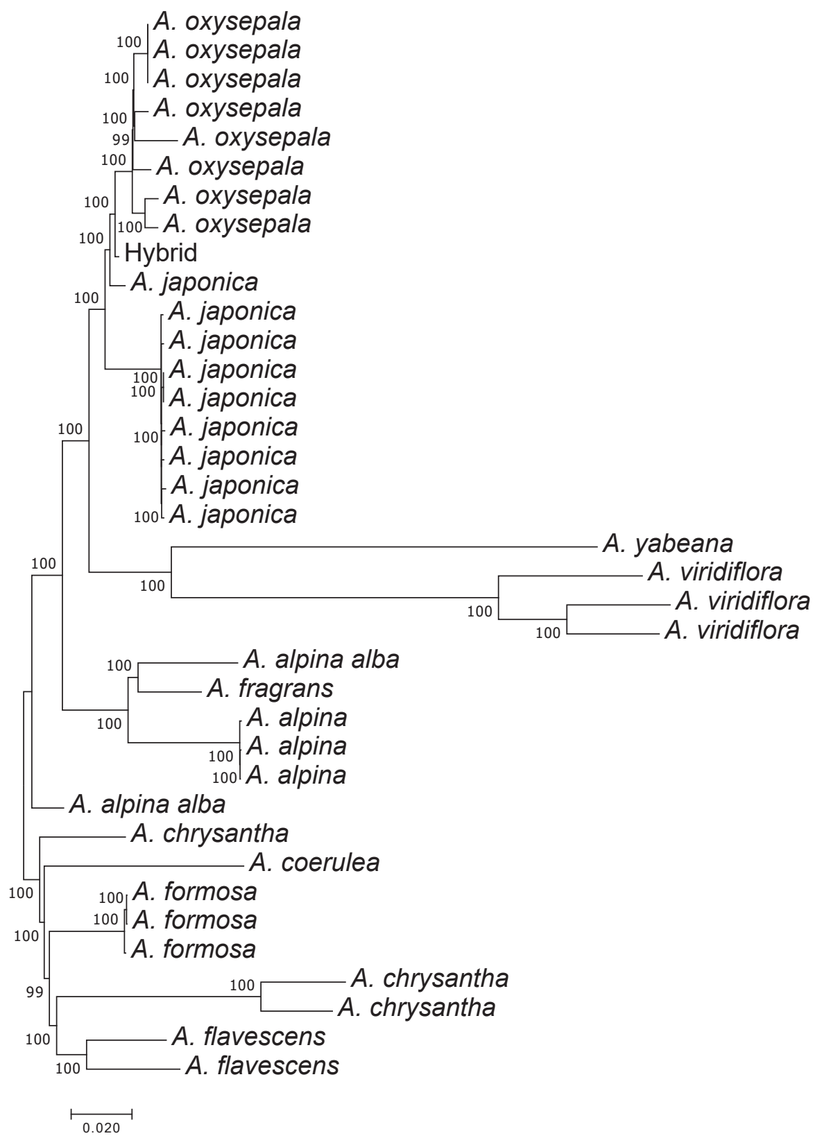
708 *, Jap, *A. japonica*; Oxy, *A. oxysepala*; Ame, American species; Eur, European species. Each percentage
709 represents the proportion of genes belonging to either DMGs or non-DMGs compared between the two
710 corresponding clades that are under corresponding or higher strength of positive selection. The *p* values
711 were obtained from Chi-square tests.

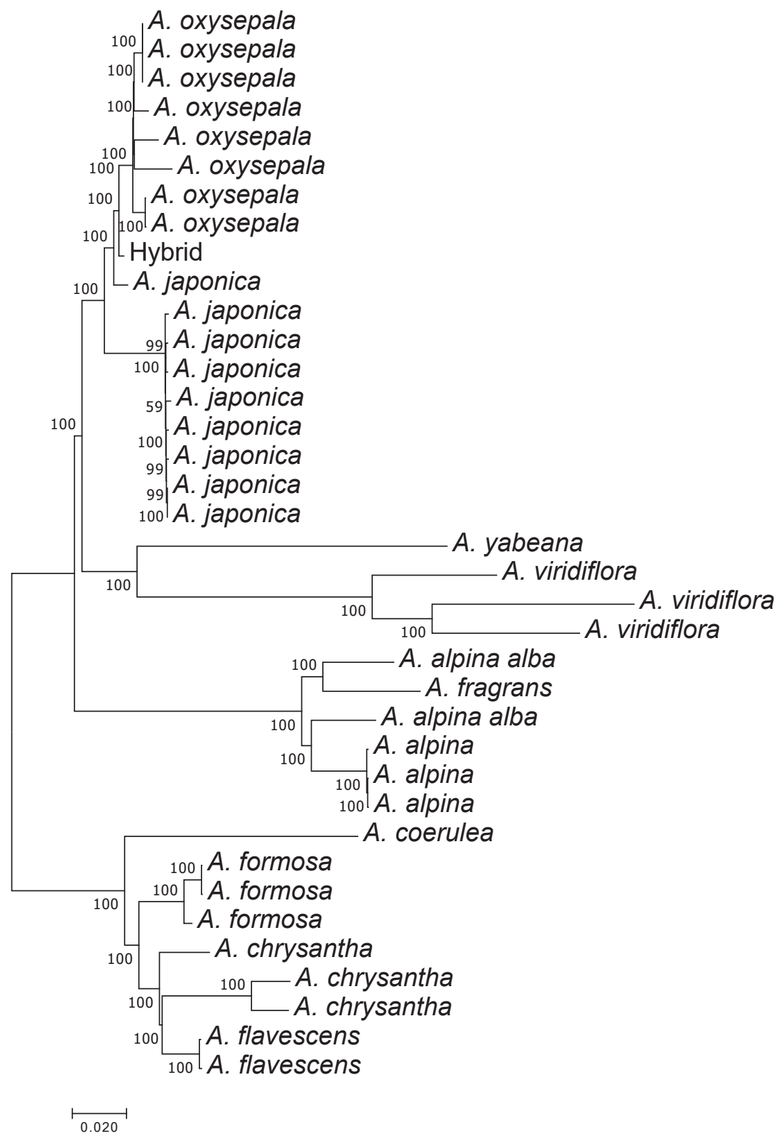

(a)
A. oxysepala
A. oxysepala
A. oxysepala
A. oxysepala
A. oxysepala
A. oxysepala
A. oxysepala
A. oxysepala
Hybrid
A. japonica
A. japonica
A. japonica
A. japonica
A. japonica
A. japonica
A. japonica
A. japonica
A. yabeana
A. viridiflora
A. viridiflora
A. viridiflora
A. alpina alba
A. alpina alba
A. fragrans
A. alpina
A. alpina
A. alpina
A. formosa
A. formosa
A. formosa
A. coerulea
A. chrysantha
A. chrysantha
A. chrysantha
A. flavescens
A. flavescens
0.020
(b)
Asian
American
European
PC2 (8.68%)
PC1 (9.57%)
0.00
−0.25
−0.50
−0.75
−0.4
−0.2
0.0
0.2
A. chrysantha
A. flavescens
A. formosa
A. alpina
A. alpina alba
A. fragrans
A. japonica
Hybrid
A. oxysepala
A. viridiflora
A. yabeana
(c)
Admixture proportion
0.8
0.6
0.4
0.2
Ancestral A. japonica
Ancestral A. oxysepala
Ancestral American
Ancestral A. viridiflora
Ancestral European

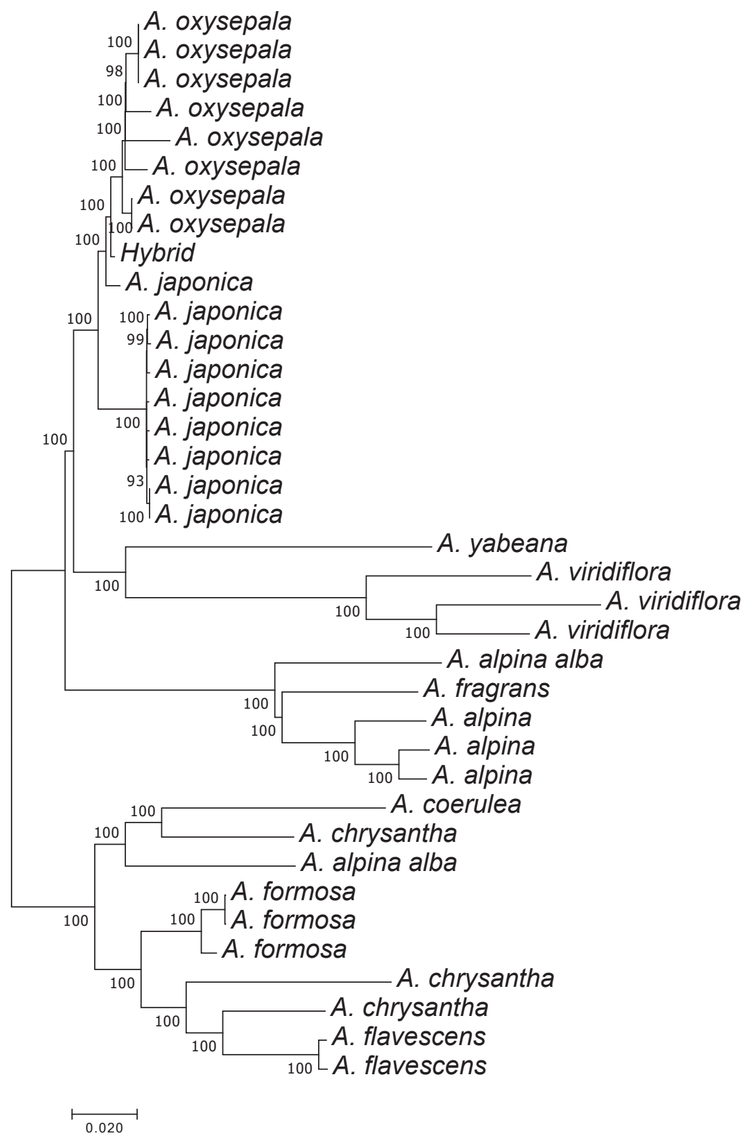

U4/U6 x U5 tri−snRNP complex
photosystem I
photosynthesis
embryo development
ribonuclease T2 activity
sulfotransferase activity
nucleosome assembly
exocytosis
exocyst
transcription, DNA−templated
terpene synthase activity
small ribosomal subunit
magnesium ion binding
DNA binding
chloroplast
ribosome
electron transfer activity
ubiquitin−dependent protein catabolic process
translational initiation
strictosidine synthase activity
serine−type endopeptidase inhibitor activity
response to wounding
response to biotic stimulus
protein binding
oxidoreductase activity
oxidation−reduction process
nucleosome
metabolic process
iron ion binding
heme binding
eukaryotic translation initiation factor 3 complex
enzyme inhibitor activity
defense response
cysteine−type endopeptidase inhibitor activity
cell wall
carbon−sulfur lyase activity
ADP binding
3−5 exonuclease activity
nucleus
copper ion binding
spindle
regulation of mitotic spindle organization
pseudouridine synthesis
pseudouridine synthase activity
anaphase−promoting complex
activation of protein kinase activity
regulation of ARF protein signal transduction
histone lysine methylation
ARF guanyl−nucleotide exchange factor activity
translation release factor activity, codon specific
RNA−DNA hybrid ribonuclease activity
riboflavin biosynthetic process
Rho guanyl−nucleotide exchange factor activity
regulation of transcription, DNA−templated
polysaccharide binding
nuclear pore
gene silencing by RNA
chlorophyllide a oxygenase [overall] activity
ATPase activity, coupled to transmembrane movement of substances
translation initiation factor binding
response to auxin
protein dimerization activity
phosphogluconate dehydrogenase (decarboxylating) activity
mitotic chromosome condensation
large ribosomal subunit
extracellular space
condensin complex
ATP binding
telomere maintenance
serine−type endopeptidase activity
proteolysis
DNA repair
DNA ligase (ATP) activity
DNA helicase activity
chromosome
aminopeptidase activity
A.japonica.vs.A.oxysepala
A.japonica.vs.American
A.japonica.vs.European
A.oxysepala.vs.American
A.oxysepala.vs.European
American.vs.European
A.japonica.vs.A.oxysepala
A.japonica.vs.American
A.japonica.vs.European
A.oxysepala.vs.American
A.oxysepala.vs.European
American.vs.European
Genes harboring clade-specific conservative variations
Differentially methylated genes
Ratio
0.1
0.2
0.3
0.4
−log(p value)
25
20
15
10
5

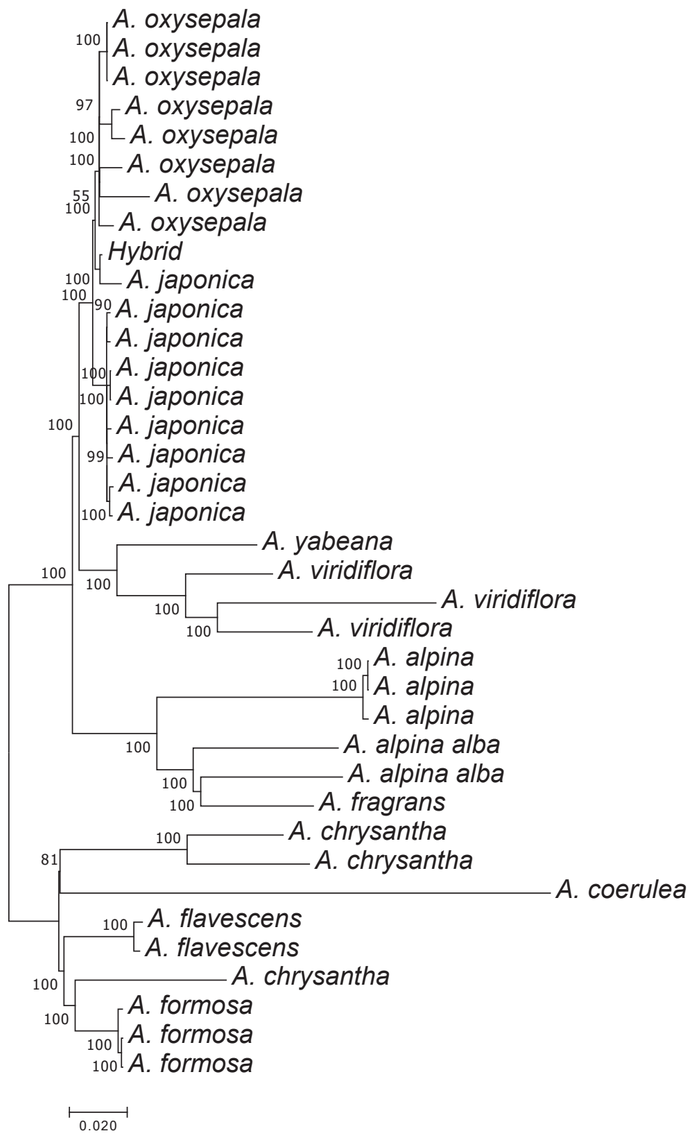

(a)
American
European
Asian
A. yabeana
A. viridiflora
A. viridiflora
A. viridiflora
A. oxysepala
A. oxysepala
A. oxysepala
A. oxysepala
A. oxysepala
A. oxysepala
A. oxysepala
A. oxysepala
Hybrid
A. japonica
A. japonica
A. japonica
A. japonica
A. japonica
A. japonica
A. japonica
A. japonica
A. japonica
A. formosa
A. formosa
A. formosa
A. flavescens
A. flavescens
A. chrysantha
A. chrysantha
A. chrysantha
A. fragrans
A. alpina alba
A. alpina alba
A. alpina
A. alpina
A. alpina
0
25
50
75
100
CG methylation level (%)
(b)
MDI
0.16
0.14
0.12
0.10
Chr1
Chr5
Chr3
Chr7
Chr6
Chr2
Chr4
A. japonica
A. oxysepala
European
American
(c)
American
European
A. japonica
A. oxysepala
Methylation level (%)
0.70
0.60
0.50
0.40
0.30
0.20
5kb upstream
TSS
TES
5kb downstream
(d)
Methylation level (%)
0.90
0.85
0.80
0.75
0.70
0.65
5kb upstream
CGI start
CGI end
5kb downstream

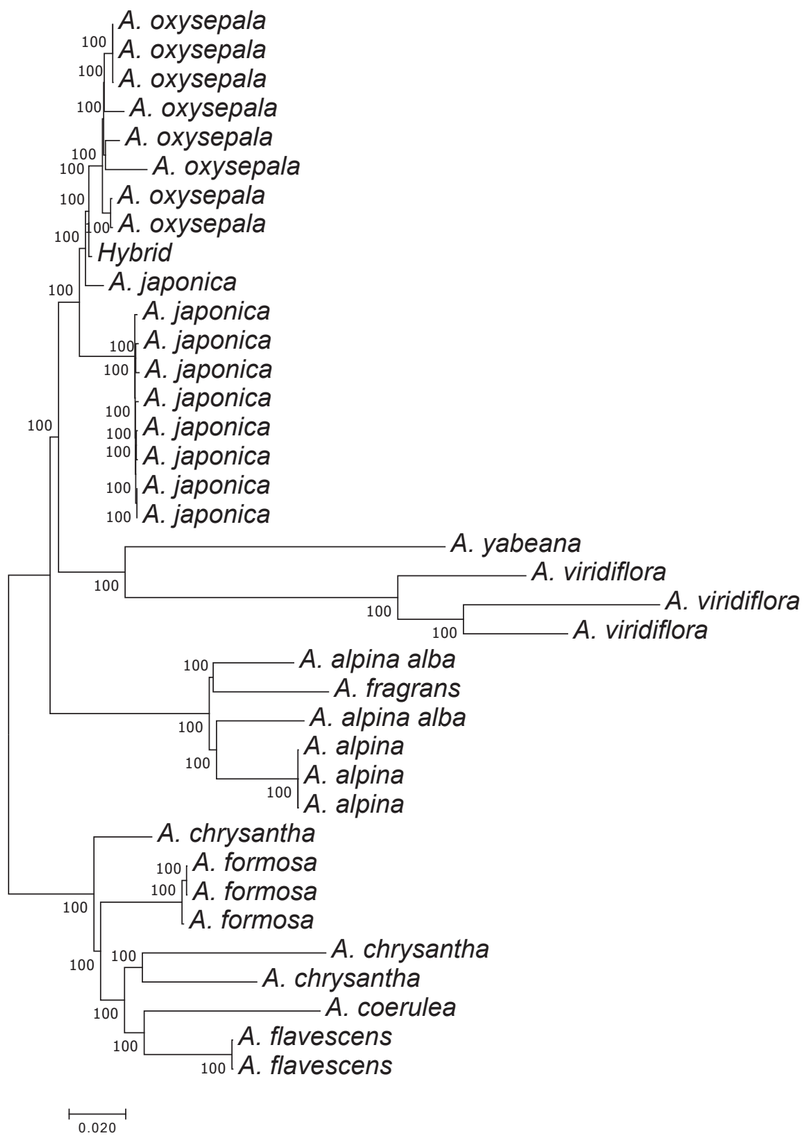

(a)
Methylation level
100
80
60
40
20
0
(b)
A. yabeana
A. viridiflora
A. oxysepala
Hybrid
A. japonica
A. alpina alba
A. fragrans
A. alpina
A. formosa
A. chrysantha
A. flavescens
(c)
R^2
1.00
0.75
0.50
0.25
0.00
Very high
High
Moderate
Low
(d)
CG-loss
Non-mutant
Frequency
1.00
0.75
0.50
0.25
0.00
Very high
High
Moderate
Low
(e)
1,222
1,169
581
p<5e−05
p<5e−08
p<5e−11
log2(Number of DMR + 1)
10.0
7.5
5.0
2.5
0.0
0
5
10
15
20
Number of significantly associated SNPs

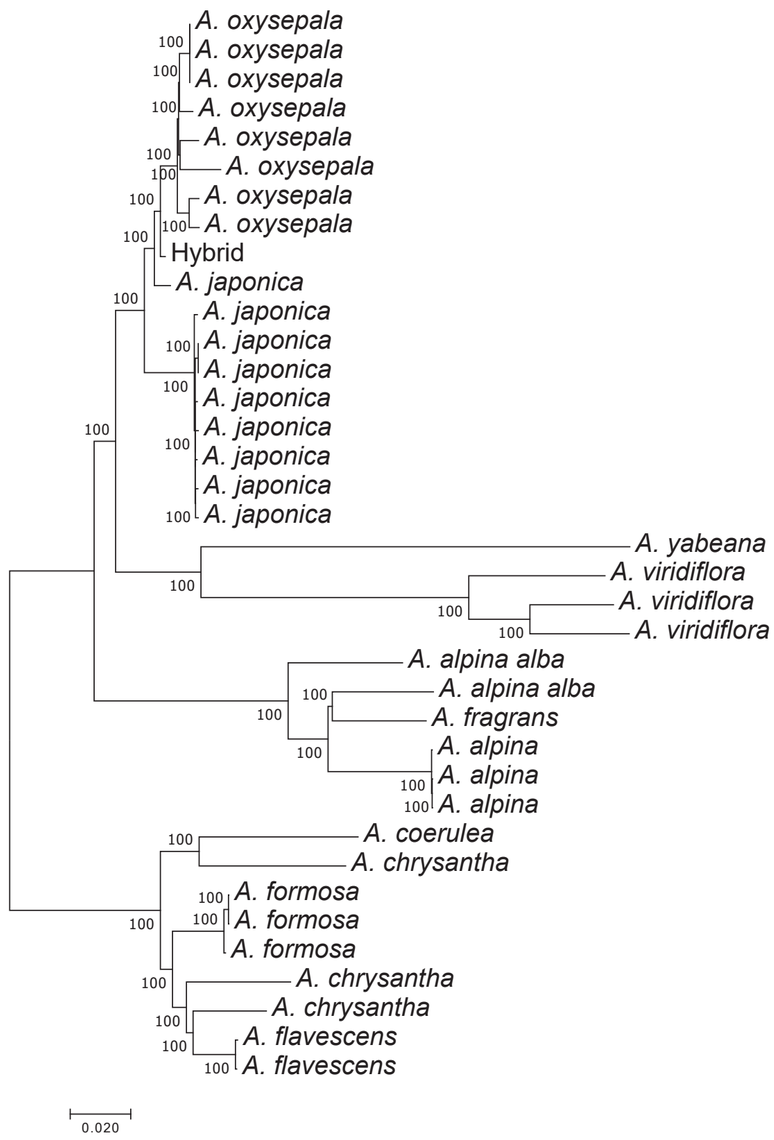

Chr1
A. oxysepala
Hybrid
A. japonica
A. yabeana
A. viridiflora
A. alpina alba
A. fragrans
A. alpina
A. coerulea
A. formosa
A. chrysantha
A. flavescens
0.020

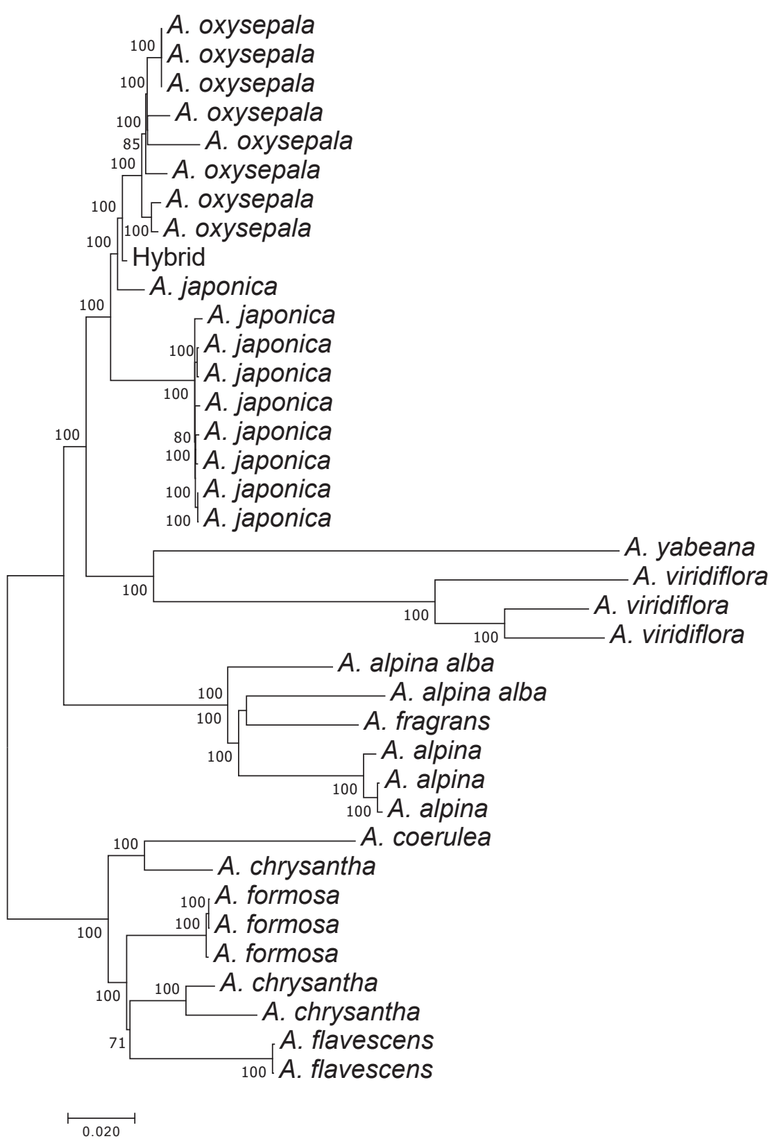

Chr2
A. oxysepala
Hybrid
A. japonica
A. yabeana
A. viridiflora
A. alpina alba
A. fragrans
A. alpina
A. coerulea
A. formosa
A. chrysantha
A. flavescens
0.020

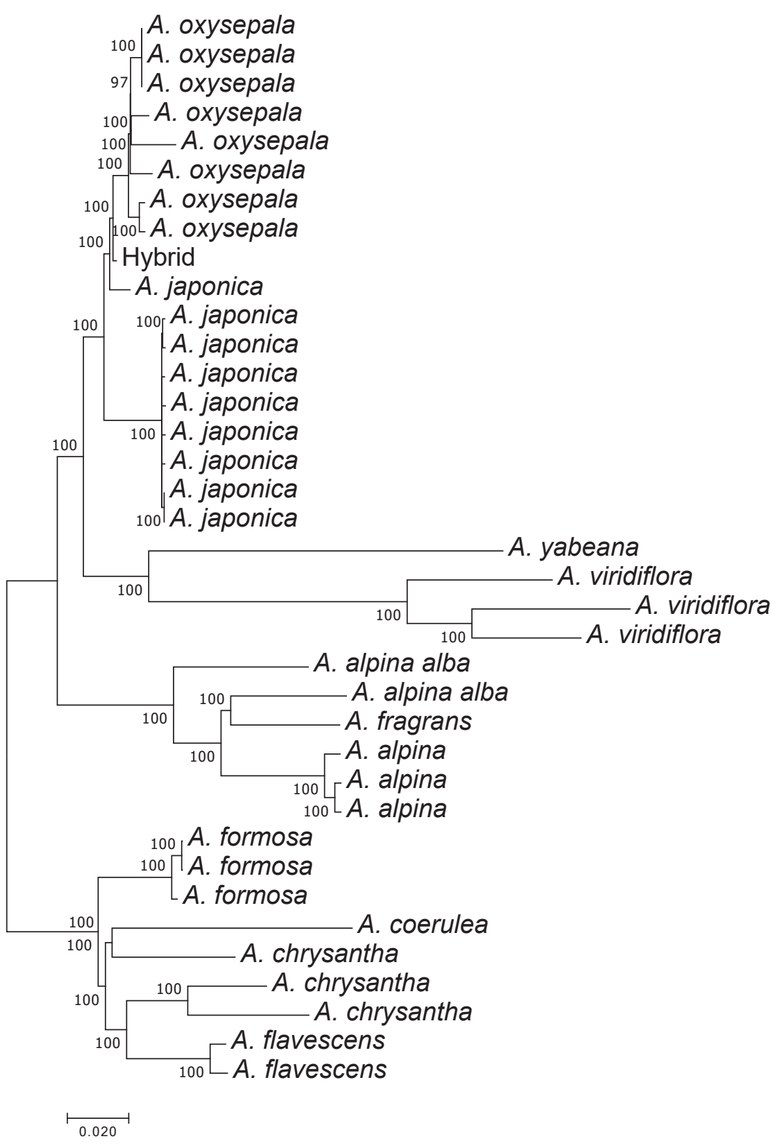

Chr3
A. oxysepala
Hybrid
A. japonica
A. yabeana
A. viridiflora
A. alpina alba
A. fragrans
A. alpina
A. coerulea
A. chrysantha
A. formosa
A. flavescens
0.020

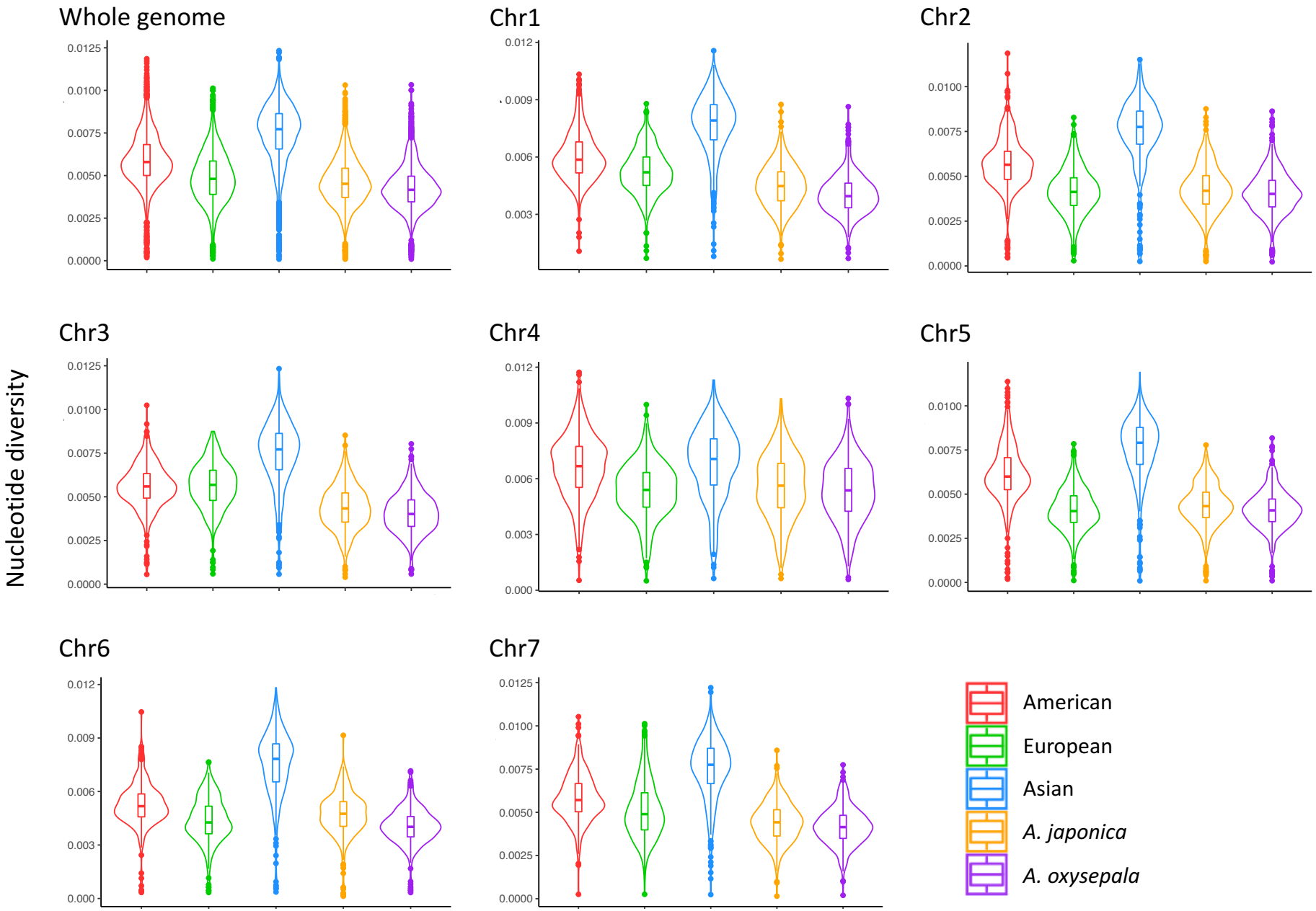

Chr4
A. oxysepala
Hybrid
A. japonica
A. yabeana
A. viridiflora
A. alpina
A. alpina alba
A. fragrans
A. chrysantha
A. coerulea
A. flavescens
A. formosa
0.020

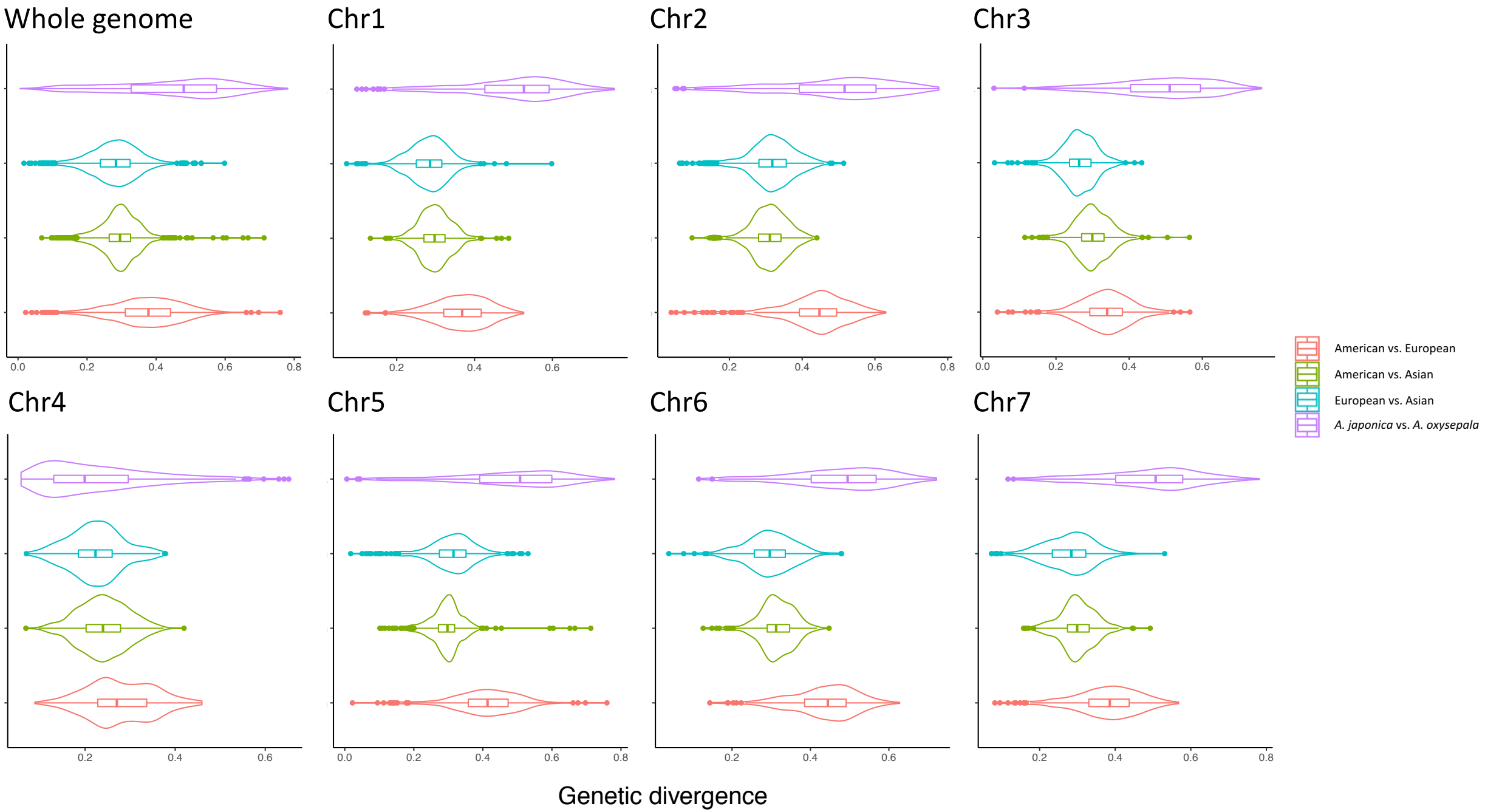

Chr5
A. oxysepala
Hybrid
A. japonica
A. yabeana
A. viridiflora
A. alpina alba
A. fragrans
A. alpina
A. chrysantha
A. formosa
A. coerulea
A. flavescens
0.020

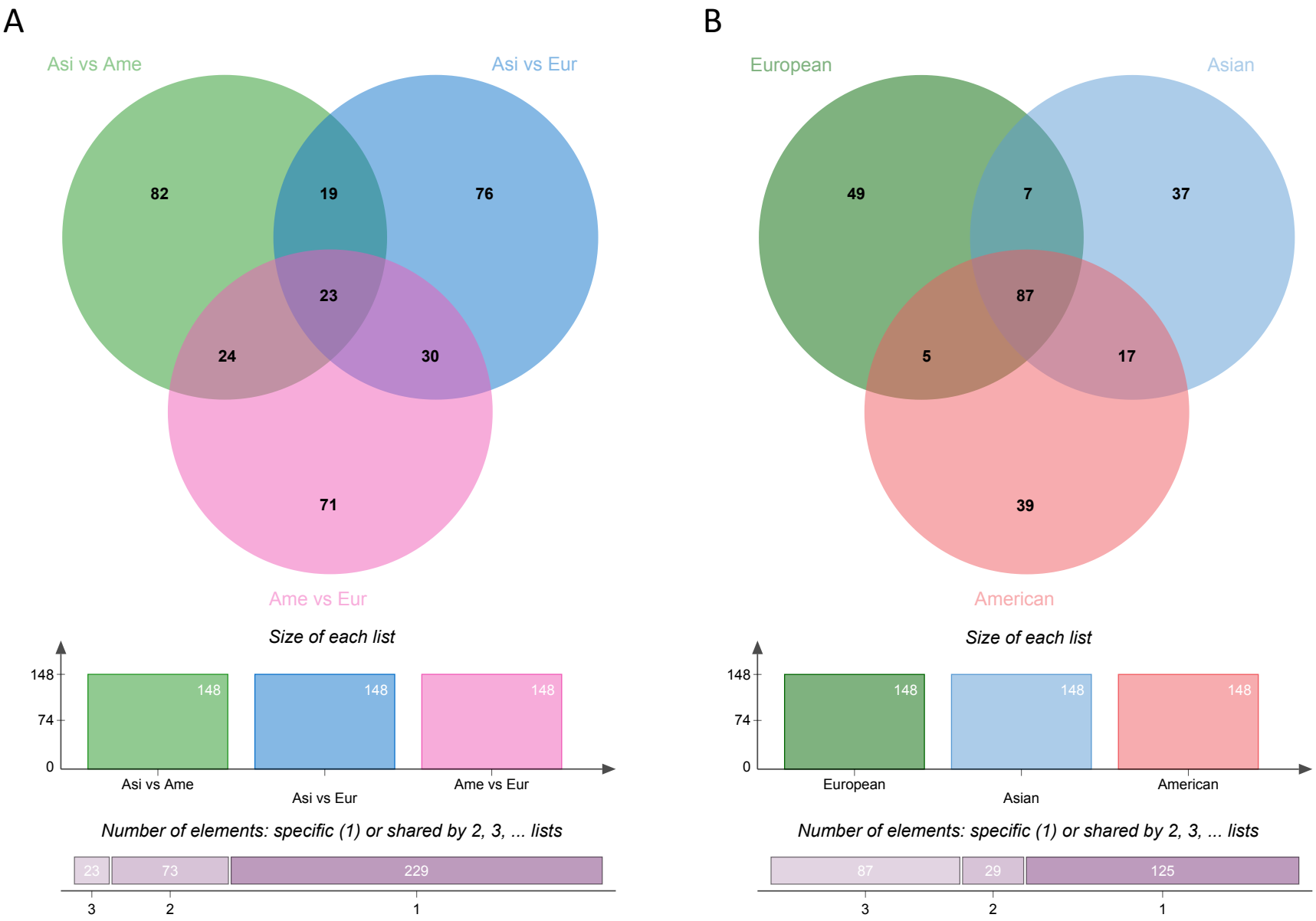

Chr6
A. oxysepala
Hybrid
A. japonica
A. yabeana
A. viridiflora
A. alpina alba
A. fragrans
A. alpina
A. coerulea
A. chrysantha
A. formosa
A. flavescens
0.020

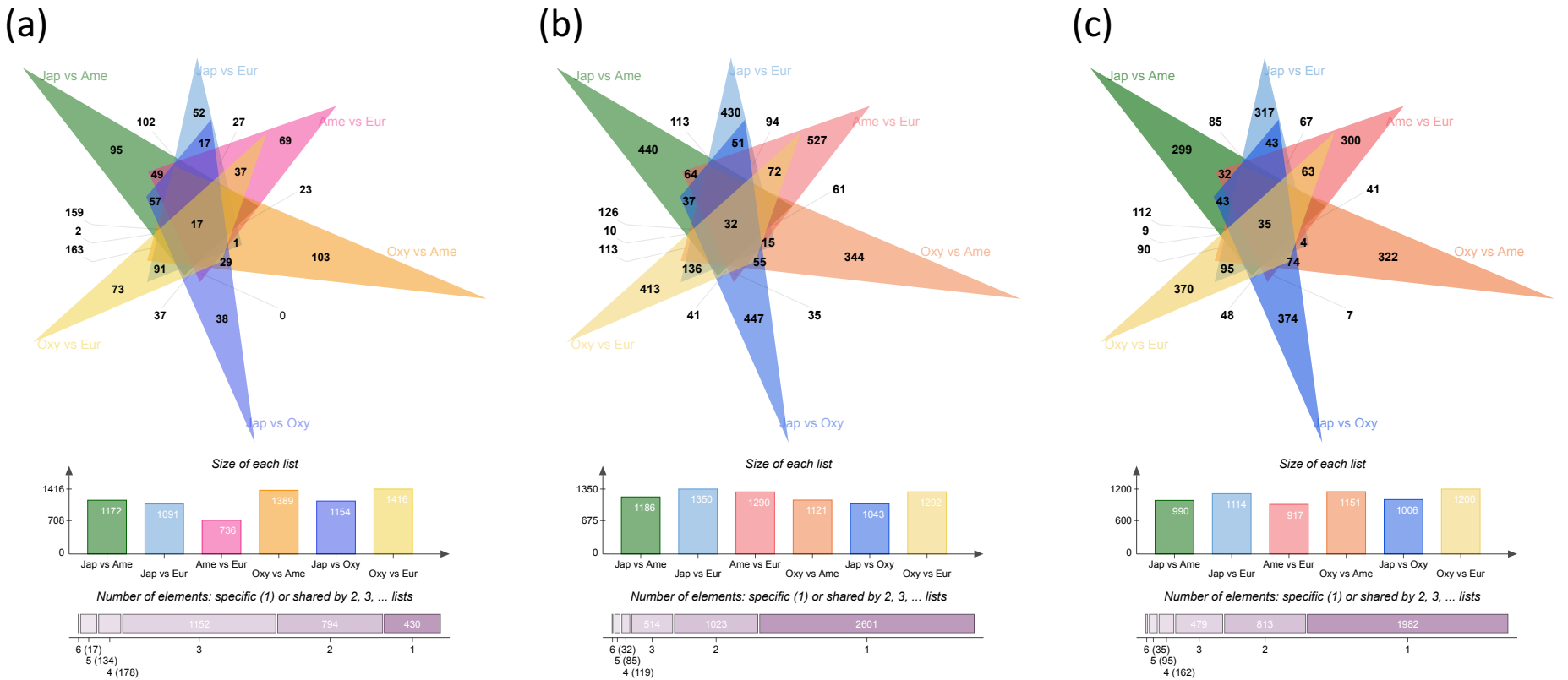

Chr7
A. oxysepala
Hybrid
A. japonica
A. yabeana
A. viridiflora
A. alpina alba
A. fragrans
A. alpina
A. coerulea
A. chrysantha
A. formosa
A. flavescens
0.020


All
A. oxysepala
Hybrid
A. japonica
A. yabeana
A. viridiflora
A. alpina alba
A. fragrans
A. alpina
A. formosa
A. coerulea
A. chrysantha
A. flavescens
0.020

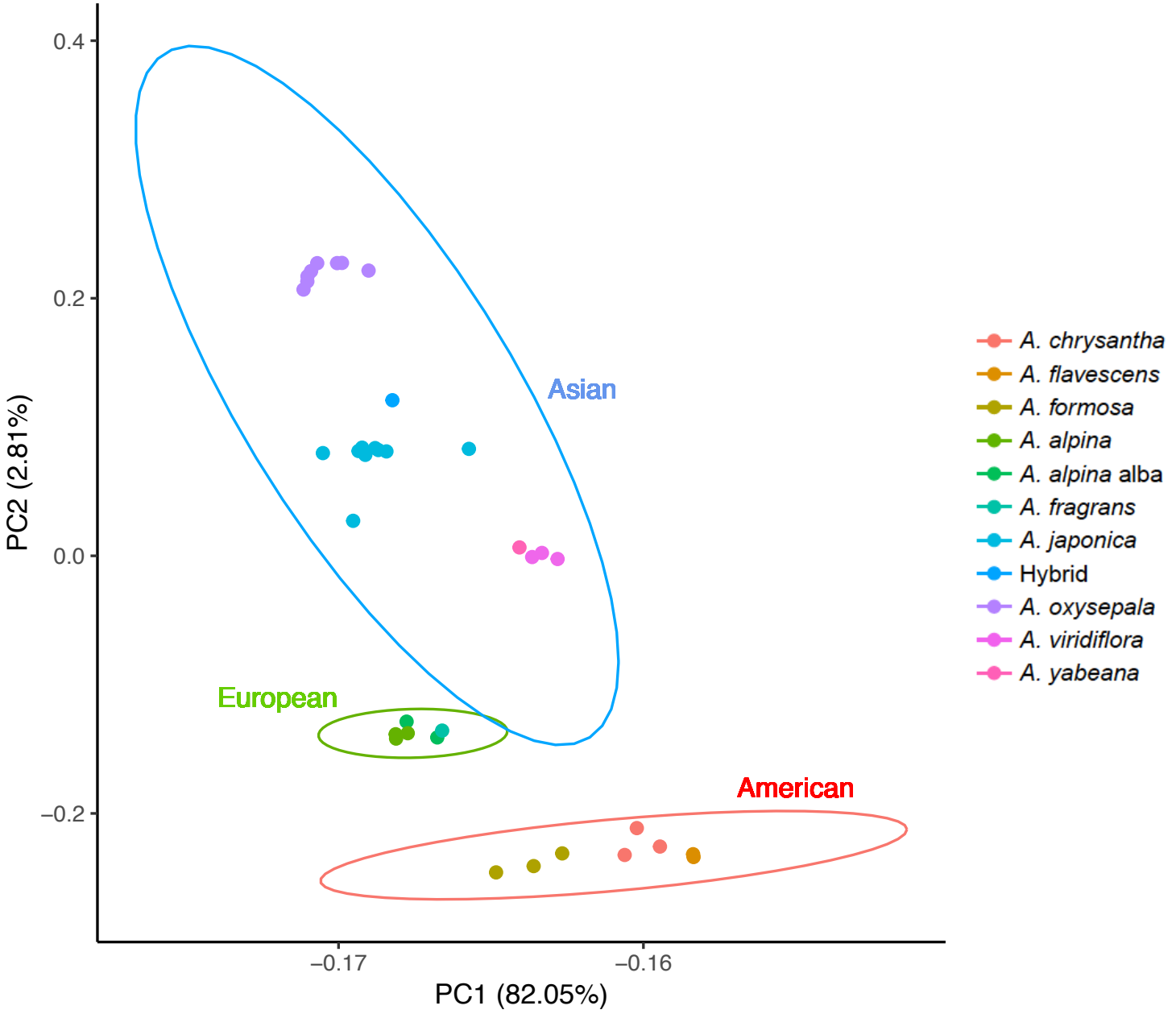

Whole genome
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Nucleotide diversity
American
European
Asian
A. japonica
A. oxysepala


Whole genome
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
0.0
0.2
0.4
0.6
0.8
American vs. European
American vs. Asian
European vs. Asian
A. japonica vs. A. oxysepala
Genetic divergence

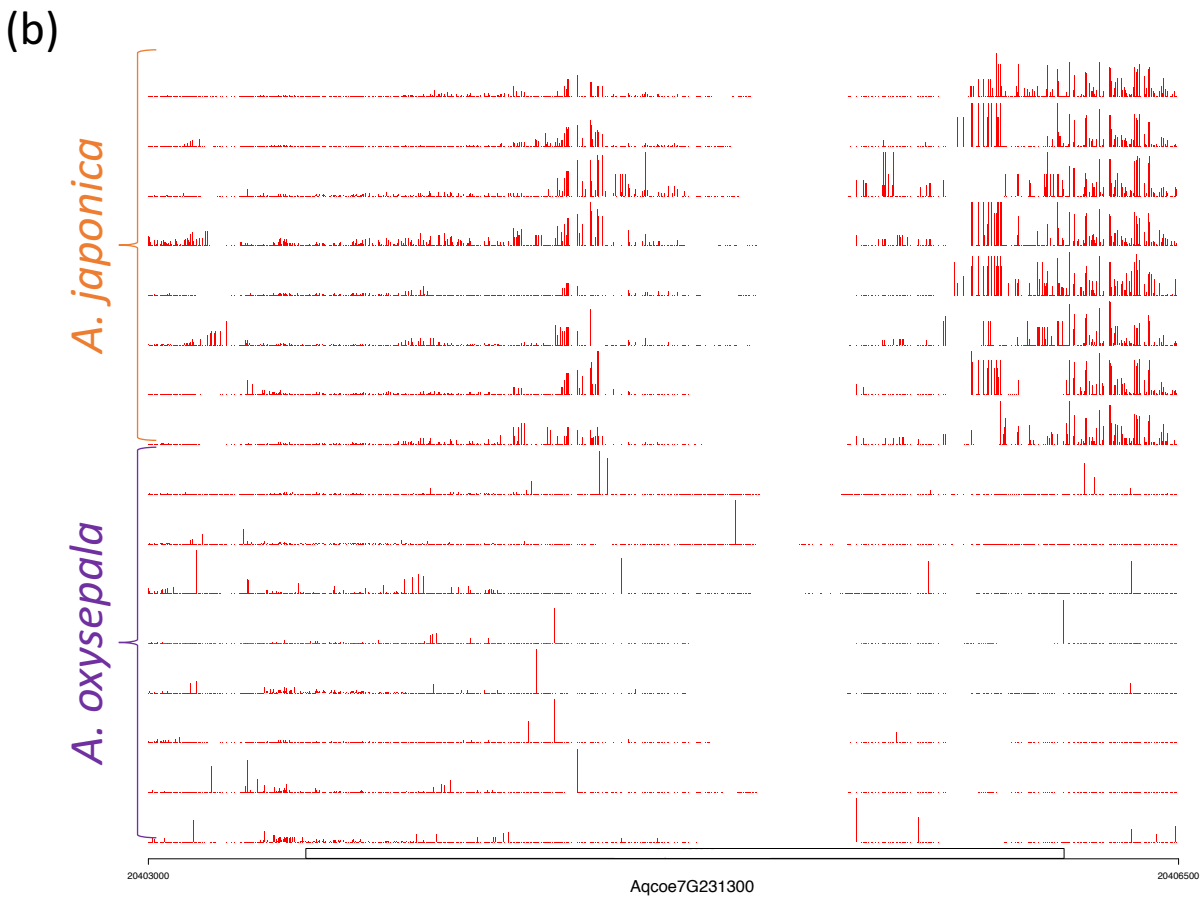

A
Asi vs Ame
Asi vs Eur
Ame vs Eur
82
19
76
23
24
30
71
Size of each list
148
74
0
148
148
148
Asi vs Ame
Asi vs Eur
Ame vs Eur
Number of elements: specific (1) or shared by 2, 3, ... lists
23
73
229
3
2
1
B
European
Asian
American
49
7
37
87
5
17
39
Size of each list
148
74
0
148
148
148
European
Asian
American
Number of elements: specific (1) or shared by 2, 3, ... lists
87
29
125
3
2
1

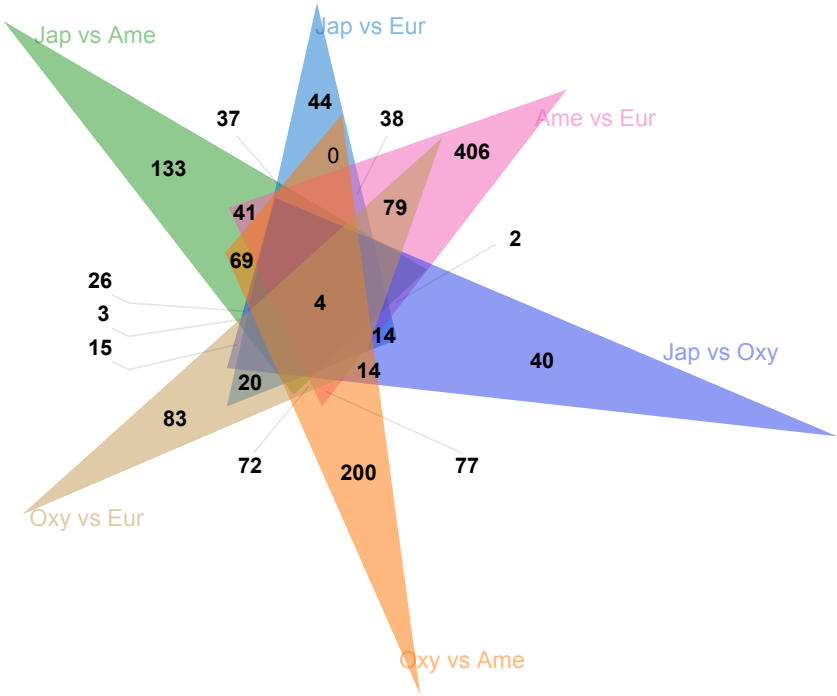

(a)
Jap vs Ame
Jap vs Eur
Ame vs Eur
Oxy vs Ame
Oxy vs Eur
Jap vs Oxy
102
52
27
95
17
69
49
37
57
23
159
17
2
1
163
103
91
29
73
37
38
0
Size of each list
1416
708
0
1172
1091
736
1389
1154
1416
Jap vs Ame
Jap vs Eur
Ame vs Eur
Oxy vs Ame
Jap vs Oxy
Oxy vs Eur
Number of elements: specific (1) or shared by 2, 3, ... lists
1152
794
430
6 (17)
5 (134)
4 (178)
3
2
1
(b)
Jap vs Ame
Jap vs Eur
Ame vs Eur
Oxy vs Ame
Oxy vs Eur
Jap vs Oxy
113
430
94
440
51
527
64
72
61
126
37
10
32
113
15
344
136
55
413
41
447
35
Size of each list
1350
675
0
1186
1350
1290
1121
1043
1292
Jap vs Ame
Jap vs Eur
Ame vs Eur
Oxy vs Ame
Jap vs Oxy
Oxy vs Eur
Number of elements: specific (1) or shared by 2, 3, ... lists
514
1023
2601
6 (32)
5 (85)
4 (119)
3
2
1
(c)
Jap vs Ame
Jap vs Eur
Ame vs Eur
Oxy vs Ame
Oxy vs Eur
Jap vs Oxy
85
317
67
299
43
300
32
63
41
112
43
9
35
90
4
322
95
74
370
48
374
7
Size of each list
1200
600
0
990
1114
917
1151
1006
1200
Jap vs Ame
Jap vs Eur
Ame vs Eur
Oxy vs Ame
Jap vs Oxy
Oxy vs Eur
Number of elements: specific (1) or shared by 2, 3, ... lists
479
813
1982
6 (35)
5 (95)
4 (162)
3
2
1

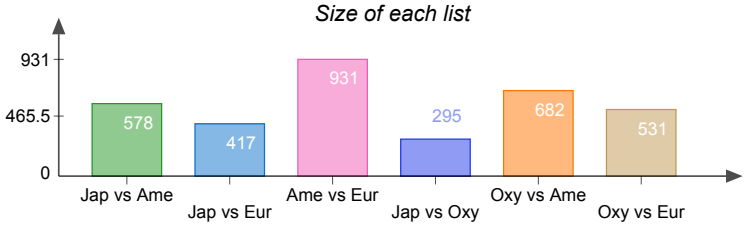
terpenoid biosynthetic process
regulation of signal transduction
protein glycosylation
peptidyl−prolyl cis−trans isomerase activity
microtubule−based process
malate dehydrogenase (decarboxylating) (NAD+) activity
COPII vesicle coat
tyrosine biosynthetic process
translation initiation factor activity
threonine−type endopeptidase activity
proteasome core complex, alpha−subunit complex
prephenate dehydrogenase (NADP+) activity
prephenate dehydrogenase (NAD+) activity
exocyst
eukaryotic translation initiation factor 3 complex
DNA recombination
ATP hydrolysis coupled proton transport
amino acid transmembrane transporter activity
amino acid transmembrane transport
thiamine pyrophosphate binding
protein serine/threonine phosphatase activity
magnesium ion transport
magnesium ion transmembrane transporter activity
intramolecular transferase activity
indole−3−glycerol−phosphate synthase activity
helicase activity
gluconeogenesis
ATP−dependent peptidase activity
transferase activity, transferring acyl groups
primary amine oxidase activity
prephenate dehydratase activity
NAD binding
mitotic cell cycle
manganese ion binding
L−phenylalanine biosynthetic process
Golgi membrane
cellular amino acid metabolic process
voltage−gated chloride channel activity
unfolded protein binding
sucrose synthase activity
SRP−dependent cotranslational protein targeting to membrane
spliceosomal complex
signal transduction
ribonucleoside binding
regulation of ARF protein signal transduction
photosystem I reaction center
phosphogluconate dehydrogenase (decarboxylating) activity
pentose−phosphate shunt
nucleolus
mRNA binding
GTP binding
fructose−bisphosphate aldolase activity
chloride transport
calcium ion binding
ATP binding
ARF guanyl−nucleotide exchange factor activity
anaphase−promoting complex
7S RNA binding
vesicle docking involved in exocytosis
trehalose biosynthetic process
structural molecule activity
structural constituent of ribosome
ribosome
Ran GTPase binding
pyruvate kinase activity
potassium ion binding
outer membrane
nucleosome
N−acetyltransferase activity
intracellular protein transport
GTPase activity
glycolytic process
four−way junction helicase activity
fatty acid biosynthetic process
biosynthetic process
acetylglucosaminyltransferase activity
6−phosphofructokinase activity
xyloglucan:xyloglucosyl transferase activity
strictosidine synthase activity
serine−type peptidase activity
protein kinase C−activating G−protein coupled receptor signaling pathway
protein dimerization activity
nuclear pore
mitochondrion organization
mitochondrial inner membrane presequence translocase complex
diacylglycerol kinase activity
cell redox homeostasis
apoplast
translation
Rho guanyl−nucleotide exchange factor activity
phosphorelay signal transduction system
mitochondrial matrix
mitochondrial inner membrane
cytokinesis
cytochrome complex assembly
chlorophyllide a oxygenase [overall] activity
cellulose microfibril organization
cell growth
anchored component of membrane
riboflavin kinase activity
riboflavin biosynthetic process
protein−disulfide reductase activity
large ribosomal subunit
glycolipid transporter activity
glycolipid transport
glycolipid binding
G−protein coupled receptor signaling pathway
enzyme inhibitor activity
cysteine−type peptidase activity
transcription, DNA−templated
serine−type carboxypeptidase activity
sequence−specific DNA binding
RNA−directed 5−3 RNA polymerase activity
ribonuclease T2 activity
prenyltransferase activity
DNA binding transcription factor activity
DNA binding
tricarboxylic acid cycle
serine−type endopeptidase inhibitor activity
response to wounding
response to stress
response to auxin
protein folding
diacylglycerol O−acyltransferase activity
zinc ion binding
ubiquitin−dependent protein catabolic process
translation release factor activity, codon specific
transcription initiation from RNA polymerase II promoter
telomere maintenance
response to oxidative stress
regulation of transcription, DNA−templated
protein repair
protein peptidyl−prolyl isomerization
protein import
peptide−methionine (S)−S−oxide reductase activity
iron−sulfur cluster assembly
intramolecular lyase activity
intracellular
glycerophosphodiester phosphodiesterase activity
defense response to fungus
defense response to bacterium
cytoplasm
chromatin binding
chitinase activity
chitin catabolic process
cell wall macromolecule catabolic process
A.japonica.vs.A.oxysepala
A.japonica.vs.American
A.japonica.vs.European
A.oxysepala.vs.American
A.oxysepala.vs.European
American.vs.European
A.japonica.vs.A.oxysepala
A.japonica.vs.American
A.japonica.vs.European
A.oxysepala.vs.American
A.oxysepala.vs.European
American.vs.European
positive selection
purifying selection
Ratio
0.2
0.4
0.6
−log(p value)
16
12
8
4

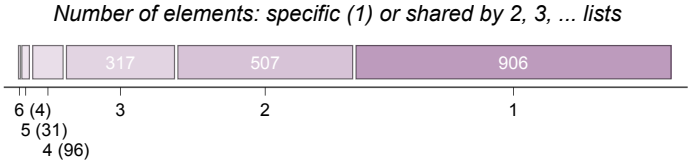

0.4
0.2
0.0
−0.2
PC2 (2.81%)
−0.17
−0.16
PC1 (82.05%)
Asian
European
American
A. chrysantha
A. flavescens
A. formosa
A. alpina
A. alpina alba
A. fragrans
A. japonica
Hybrid
A. oxysepala
A. viridiflora
A. yabeana

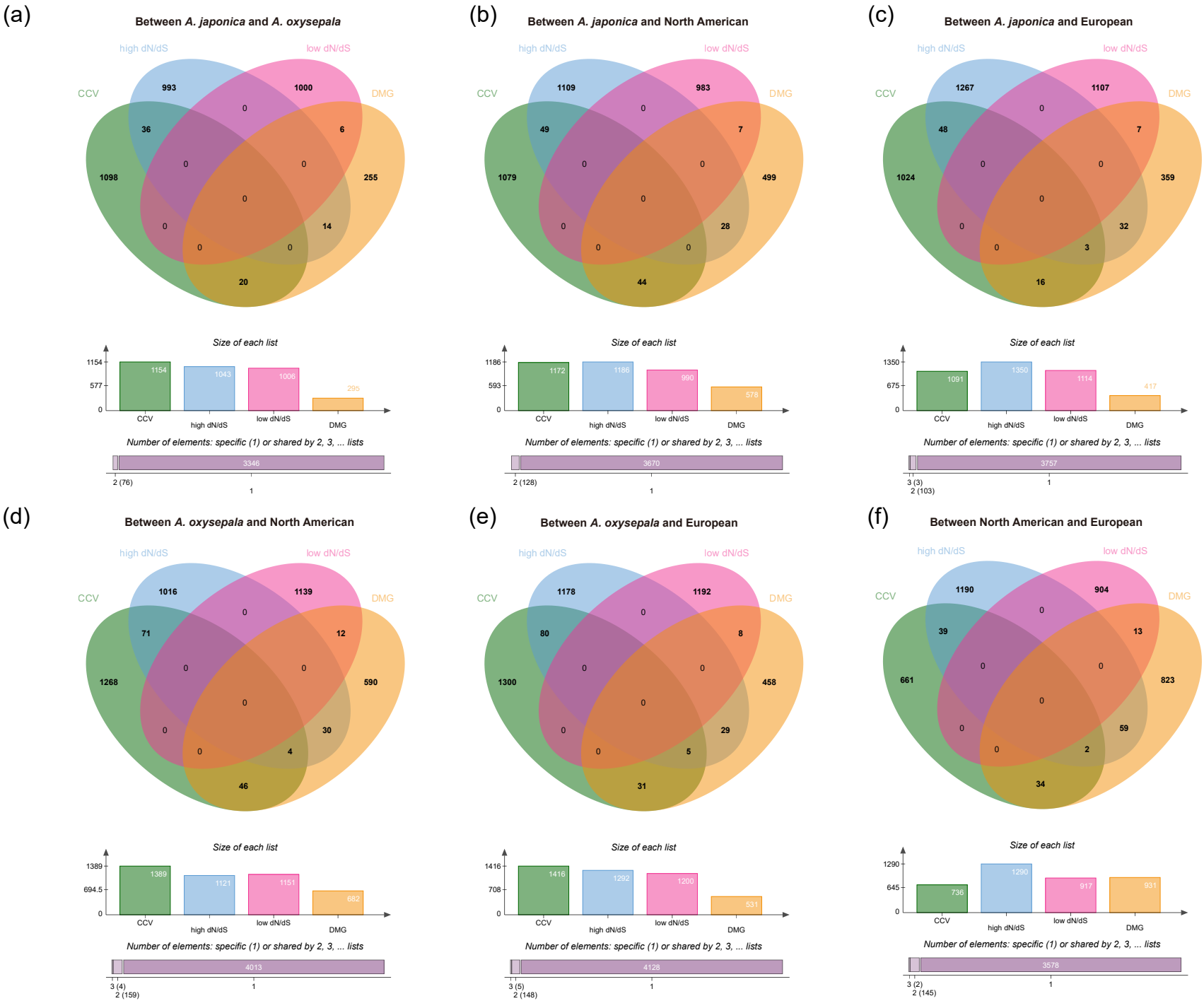
(a)
A. japonica
A. oxysepala
20392800
20394100
Aqcoe7G230600


(b)
A. japonica
A. oxysepala
20403000
20406500
Aqcoe7G231300


Jap vs Ame
Jap vs Eur
Ame vs Eur
Jap vs Oxy
Oxy vs Eur
Oxy vs Ame
133
37
44
38
0
406
41
79
2
69
26
4
3
15
14
40
14
20
83
72
200
77


Size of each list
931
465.5
0
578
417
931
295
682
531
Jap vs Ame
Jap vs Eur
Ame vs Eur
Jap vs Oxy
Oxy vs Ame
Oxy vs Eur


Number of elements: specific (1) or shared by 2, 3, ... lists
317
507
906
6 (4)
5 (31)
4 (96)
3
2
1


(a)
Between A. japonica and A. oxysepala
high dN/dS
low dN/dS
CCV
DMG
993
1000
0
36
6
1098
0
0
255
0
14
0
0
20
Size of each list
1154
577
0
1154
1043
1006
295
CCV
high dN/dS
low dN/dS
DMG
Number of elements: specific (1) or shared by 2, 3, ... lists
3346
2 (76)
1
(b)
Between A. japonica and North American
high dN/dS
low dN/dS
CCV
DMG
1109
983
0
49
7
1079
0
0
499
0
28
0
0
44
Size of each list
1186
593
0
1172
1186
990
578
CCV
high dN/dS
low dN/dS
DMG
Number of elements: specific (1) or shared by 2, 3, ... lists
3670
2 (128)
1
(c)
Between A. japonica and European
high dN/dS
low dN/dS
CCV
DMG
1267
1107
0
48
7
1024
0
0
359
0
32
0
3
16
Size of each list
1350
675
0
1091
1350
1114
417
CCV
high dN/dS
low dN/dS
DMG
Number of elements: specific (1) or shared by 2, 3, ... lists
3757
3 (3)
2 (103)
1
(d)
Between A. oxysepala and North American
high dN/dS
low dN/dS
CCV
DMG
1016
1139
0
71
12
1268
0
0
590
0
30
0
4
46
Size of each list
1389
694.5
0
1389
1121
1151
682
CCV
high dN/dS
low dN/dS
DMG
Number of elements: specific (1) or shared by 2, 3, ... lists
4013
3 (4)
2 (159)
1
(e)
Between A. oxysepala and European
high dN/dS
low dN/dS
CCV
DMG
1178
1192
0
80
8
1300
0
0
458
0
29
0
5
31
Size of each list
1416
708
0
1416
1292
1200
531
CCV
high dN/dS
low dN/dS
DMG
Number of elements: specific (1) or shared by 2, 3, ... lists
4128
3 (5)
2 (148)
1
(f)
Between North American and European
high dN/dS
low dN/dS
CCV
DMG
1190
904
0
39
13
661
0
0
823
0
59
0
2
34
Size of each list
1290
645
0
736
1290
917
931
CCV
high dN/dS
low dN/dS
DMG
Number of elements: specific (1) or shared by 2, 3, ... lists
3578
3 (2)
2 (145)
1