An oligogenic architecture underlying ecological and reproductive divergence in sympatric populations

Dušica Briševac[1], Carolina M. Peralta[1], Tobias S. Kaiser[1]*

[1] Max Planck Research Group *Biological Clocks*, Max Planck Institute for Evolutionary Biology, August-Thienemann-Strasse 2, 24306 Ploen, Germany

* kaiser@evolbio.mpg.de

Abstract

The evolutionary trajectories and genetic architectures underlying ecological divergence with gene flow are poorly understood. Sympatric timing types of the intertidal insect *Clunio marinus* (Diptera) from Roscoff (France) differ in lunar reproductive timing. One type reproduces at full moon, the other at new moon, controlled by a circalunar clock of yet unknown molecular nature. Lunar reproductive timing is a magic trait for a sympatric speciation process, as it is both ecologically relevant and entails assortative mating. Here we show that the difference in reproductive timing is controlled by at least four quantitative trait loci (QTL) on three different chromosomes. They are partly associated with complex inversions, but differentiation of the inversion haplotypes cannot explain the different phenotypes. The most differentiated locus in the entire genome, with QTL support, is the *period* locus, implying that this gene could not only be involved in circadian timing but also in lunar timing. Our data indicate that magic traits can be based on an oligogenic architecture and can be maintained by selection on several unlinked loci.

Introduction

After a long debate on whether speciation can only happen with geographic isolation (*allopatry*) or also with full range overlap (*sympatry*)[1-4], there is growing evidence for population divergence with gene flow[4,5]. The process is generally assumed to start with divergent ecological selection, but for speciation to complete, additional mechanisms for reproductive isolation must come into play[6,7]. Traits underlying ecological divergence and reproductive isolation must be genetically coupled, as otherwise they cannot withstand the homogenizing effects of gene flow and recombination[6]. One possibility to achieve such coupling are *magic traits*, i.e. traits that entail ecological divergence and reproductive isolation at the same time[8]. For example, divergent ecological adaptations can lead to differences in reproductive timing, i.e. *allochrony*[9] or *isolation by time*[10]. One example is found in the apple maggot *Rhagoletis*[11,12]. Magic traits imply that ecological divergence and reproduction are pleiotropically controlled by the same set of genes[13], usually involving only one or very few loci. As another possibility, genomic regions of low or suppressed recombination, such as chromosomal inversions, can entail a coupling of ecological and reproductive traits even if they are controlled by different sets of genes. Examples of inversions associated with ecologically relevant traits are found in *Littorina* snails[14,15] and *Heliconius* butterflies[16]. In this study, we assessed if these mechanisms play a role in sympatric population divergence in the marine midge *Clunio marinus* (Diptera: Chironomidae).

C. marinus is found in the intertidal zone of the European Atlantic coast. Larvae live at the lower levels of the intertidal, which are almost permanently submerged. For successful reproduction, the adults need these regions to be exposed by the low tide. The lowest tides predictably recur just after new moon and full moon. Therefore, *C. marinus* adults emerge

52 only during full moon or new moon low tides, reproduce immediately and die a few hours
53 later in the rising tide. This life cycle adaptation is based on a circalunar clock, i.e. an
54 endogenous time-keeping mechanism which synchronizes development and maturation with
55 lunar phase. Circalunar clocks are common in marine organisms, but their molecular basis is
56 still unknown[17]. Additionally, a circadian clock ensures that *C. marinus* adults only emerge
57 during the low tide[18]. As the pattern and amplitude of the tides differ dramatically along the
58 coastline, *C. marinus* populations from different geographic sites show various genetic
59 adaptations in circadian and circalunar timing[18-20].

60 Notably, full moon and new moon low tides are ecologically equally suitable for *C. marinus'*
61 reproduction, representing different *timing niches* that are occupied by different *timing types*
62 (for details and definitions see[21]). Some timing types of *C. marinus* use both niches (*"semi-*
63 *lunar rhythm"*; SL type), but there are also dedicated full moon (FM) or new moon (NM) timing
64 types, which occupy only one timing niche[21]. Divergence into FM and NM timing types
65 represents a magic trait, as it automatically entails assortative mating. We recently discovered
66 that in Roscoff (France) FM and NM timing types occur in sympatry, largely separated by
67 reproductive timing, but still connected by gene flow[21]. Here we present evidence for
68 polymorphic chromosomal inversions in all chromosome arms of the sympatric FM and NM
69 types. Quantitative trait loci (QTL) mapping for divergence in lunar reproductive timing
70 identifies four unlinked and additive QTL in different chromosomes and inversions. Individual
71 loci inside the inversions are more differentiated than the inversions, suggesting that the
72 inversions themselves are not directly required to maintain genetic linkage of adaptive loci
73 and that ecological divergence is kept up by permanent selection on unlinked loci on different
74 chromosomes.

Results

Genome sequencing confirms limited genetic differentiation between the sympatric FM and NM types. In order to gain insights into the loci and processes underlying sympatric divergence in reproductive timing, we sequenced 48 individual genomes for the FM and NM timing types (24 individuals each, defined by their lunar timing phenotype). The resulting set of 721,000 genetic variants (608,599 SNPs; 112,401 small indels) indicated limited genetic differentiation between the FM and NM types (Supplementary Fig. 1; weighted genome-wide F_{ST} = 0.028). Principal component analysis (PCA) and ADMIXTURE identify one individual as a *migrant in time*, which was caught at full moon among FM type individuals, but genetically clearly is of NM type (Supplementary Fig. 1A and B; blue circle and arrow). Many other individuals, particularly in the NM type, show ADMIXTURE fractions close to 0.5 or 0.25, suggesting they are F1 hybrids or backcrosses (Supplementary Fig. 1A and B; yellow circle and arrows). Finally, there are four FM individuals, which are genetically distinct along principal component 2 and in ADMIXTURE at K=4 (Supplementary Fig. 1A and B; red circle and arrows). These might constitute either a sub-lineage of the FM type or migrants from a different geographic location. Taken together, full genome resequencing confirms that there is strong gene flow between sympatric timing types.

A differentiated chromosomal inversion system on chromosome 1. Plotting genetic differentiation of the phenotypically defined FM and NM types along the genome (Supplementary Fig. 1C) revealed a region of elevated F_{ST} on the telocentric chromosome 1 (Fig. 1A). This region coincides with a block of long-range linkage-disequilibrium (LD) in the FM type (Fig. 1B; Supplementary Fig. 2). Structural variant (SV) calling based on additional

99 long-read sequencing data supports that this block represents a chromosomal inversion in
100 both timing types (Fig. 1C). Notably, the NM type shows a smaller LD block (Fig. 1B;
101 Supplementary Fig. 2), suggesting that a second structural variant occurs within the limits of
102 the detected large chromosomal inversion. While this second variant was not picked up in SV
103 calling, genetic linkage information obtained from crosses (described in detail below)
104 indicates that this is a second inversion (see double inverted marker order in Fig. 4A,B).

105 For a more detailed analysis of this inversion system, we subdivided it into three windows,
106 based on the inversion coordinates suggested by long-range LD. Two windows correspond to
107 those parts of the large inversion that do not overlap with the small inversion (roughly 2.7-
108 4.4 Mbp and 11.3-18.7 Mbp; Supplementary Table 1; Fig. 1A-C). The central window
109 corresponds to the overlap of both inversions (roughly 4.4-11.3 Mbp; Supplementary Table
110 4; Fig. 1A-C). In a Principal component analysis (PCA) on the genetic variants in these windows
111 (Supplementary Table 2), the four FM individuals that were already identified as genetically
112 distinct (Supplementary Fig. 1) always cluster separately and cannot be assessed (Fig. 1 D to
113 G, red question mark). The remaining 44 individuals are split into three genotype classes when
114 PCA is performed on the non-overlapping regions of the large inversion (Fig. 1D and F). These
115 classes correspond to individuals homozygous for the standard haplotype (SS), individuals
116 homozygous for an inverted haplotype (I_1I_1), and heterozygotes for the inversion (SI_1). The
117 haplotype with the higher observed heterozygosity in homozygotes is considered the
118 ancestral standard haplotype S (see Fig. 1H-J). Observed heterozygosity is clearly elevated in
119 individuals heterozygous for the inversion (Fig. 1H and J), suggesting there are polymorphisms
120 specific to the S or I_1 haplotypes. This is corroborated by the patterns of genetic
121 differentiation between SS and I_1I_1 homozygotes, which show F_{ST} values of up to 1
122 (Supplementary Figure 3).

123 In the region where both inversions overlap (Fig. 1E), the 44 individuals separate into six
124 genotype clusters, corresponding to three clusters of inversion haplotype homozygotes (SS,
125 I_1I_1, I_2I_2) and three clusters of heterozygotes (SI_1, SI_2, I_1I_2). Individuals that are I_1I_1 genotypes
126 in the 2.7-4.4 Mbp and 11.3-18.7 Mbp windows (Fig. 1D and F) now separate into the I_1I_1, I_1I_2
127 and I_2I_2 clusters (Fig. 1E). This indicates that the small inversion leading to haplotype I_2
128 happened in the already inverted haplotype I_1 of the large inversion (compare Fig. 1L). This is
129 backed up by genetic linkage data (Fig. 4A,B) and consistent with I_2I_2 homozygotes showing
130 even lower heterozygosity than I_1I_1 homozygotes (Fig. 1I). Again, heterozygosity is clearly
131 elevated in the inversion heterozygotes (Fig. 1I). A sliding-window PCA analysis along the
132 chromosomes (500kb windows, 100kb steps; Supplementary Figure 4) confirms these
133 patterns of segregation and the approximate genomic positions of inversion breakpoints.
134 Sliding-window PCA also indicated that there might be a third inversion, which ends at 24.8
135 Mbp (Fig. 1A and G; Supplementary Figure 4). Its start overlaps with the differentiated
136 inversion system and is therefore hard to identify, but according to observed heterozygosity
137 along the chromosome, it might lie at around 9 Mbp (Supplementary Fig. 5A). This inversion
138 is only clearly detected in three SI_3 heterozygotes (Fig. 1G and K). These three individuals are
139 found in the SS cluster in the other genomic windows, implying that the inversion leading to
140 haplotype I_3 happened in the standard haplotype S (Fig. 1L).
141 One peculiarity deserves further investigation: In the window from 2.7-4.4 Mbp some
142 individuals of the I_1I_1 genotype cluster close to the SI_1 heterozygotes. However, the pattern in
143 which these individuals segregate in the 4.4-11.3 Mbp window unequivocally indicates that
144 these must be I_1I_1 homozygotes. They also show much lower heterozygosity than the SI_1
145 heterozygotes. Their peculiar clustering may indicate a certain degree of gene conversion
146 between inversion haplotypes or a complex demographic history.

147 Taken together, our data support three inversion events, which we name In(1a), In(1b) and
148 In(1c), leading to four distinct haplotypes of chromosome 1 (S, I_1, I_2 and I_3; Fig. 1L). From the
149 PCA-derived genotypes (Fig. 1D-G), we can infer the frequency of each haplotype (Fig. 1L).
150 There is marked genetic differentiation between the sympatric timing types. The FM type
151 carries all haplotypes but is dominated by the S haplotype, while the NM type almost
152 exclusively carries the inverted I_1 and I_2 haplotypes.
153
154 **Non-differentiated inversions on chromosomes 2 and 3.** The analysis of long-range LD
155 also identifies one putative chromosomal inversion on each arm of the metacentric
156 chromosomes 2 and 3 (Fig. 2B,C; Supplementary Fig. 6). We call them In(2L), In(2R), In(3L) and
157 In(3R). In(2L) is confirmed by inverted marker order in linkage mapping (Fig. 4A,B) and In(2R)
158 is confirmed by SV calling. As above, windowed PCA along the chromosomes confirms the
159 approximate inversion breakpoints and allows to genotype the individuals for the inversion
160 haplotypes (Fig. 2D-G). Heterozygosity is clearly elevated in inversion heterozygotes (SI; Fig.
161 2H-K) and lowered in the inversion homozygotes (II; Fig. 2H-K). Notably, inversion
162 homozygotes are absent for two of these inversions (Fig. 2D and F) and very rare for the other
163 two (Fig. 2E and G), suggesting these inversions might impose some fitness constraints. These
164 four inversions are only weakly differentiated between populations (Fig. 2 L-O). Only In(2L)
165 shows up as a block of mildly elevated genetic differentiation between the strains (Fig. 2A).
166
167 **Crosses between the FM and NM types indicate that lunar reproductive timing is**
168 **heritable and controlled by four QTL.** Next, we tested for a genetic basis of lunar reproductive
169 timing by performing crossing experiments between the FM and NM types (Fig. 3). F1 hybrids
170 emerge intermediate between the FM and NM type, with a slight shift towards FM emergence

171 (Fig. 3C). In the F2 generation, the phenotype distribution is spread out, but does not fully
172 segregate into parental and F1 phenotype classes (Fig. 3D). This indicates that the difference
173 in lunar reproductive timing is controlled by more than one genetic locus. From the crossing
174 experiment, we picked a set of several F2 families that together comprised 158 individuals,
175 which all go back to a single parental pair – a FM type mother and a NM type father. Because
176 of the limited genetic differentiation between the timing types (see Supplementary Fig. 1C
177 and [21]), we sequenced the full genomes of the two parents in order to identify informative
178 genetic markers for linkage mapping and QTL mapping. We picked a set of 32 microsatellite
179 markers, 23 of which turned out to be reliably amplified and informative in the F2 families, as
180 well as four insertion-deletion polymorphisms (Supplementary Table 3).

181 Genetic linkage mapping confirmed the existence of the inversions In(1a) and In(1b), both
182 of which are supported by inverted marker order on the genetic linkage map (Fig. 4A). Marker
183 order is also inverted for In(2L), but not for the other inversions (Fig. 4A). We then performed
184 stepwise QTL identification and Multiple QTL Mapping (MQM) as implemented in R/qtl and
185 found four largely additive QTL for the difference in lunar reproductive timing (Fig. 4A and
186 Supplementary Table 4). A specific scan for epistatic effects only shows a weak interaction of
187 chromosome 1 with the QTL on chromosome 2 (Supplementary Figure 7). The four QTL model
188 explains 54% of the total variance in lunar reproductive timing and the estimated additive
189 effects of the QTL account for 6 days of the timing difference between the FM and NM type
190 (Fig. 4A, see Supplementary Table 4 for individual QTL effects and variance explained).

191 In both MQM and additional Composite Interval Mapping (CIM; Supplementary Figure 8)
192 we estimated confidence intervals for the QTL (Fig. 4A). There are two QTL on chromosome
193 1, one of which overlaps with the inversion system. While the inversions reduce
194 recombination and thus map length for the respective genomic region (Fig. 4A), there was

195 sufficient recombination to assign the QTL interval to only one end of In(1a). Based on the
196 first genetic marker outside the QTL's confidence intervals, we matched the QTL region with
197 the corresponding genomic reference sequence (Fig. 4B). The second QTL on chromosome 1
198 maps outside the inversion system and is not linked with the other QTL (Fig. 4A,B). The other
199 two QTL are found on chromosomes 2 and 3 and overlap with In(2R) and In(3L) respectively.
200 The genomic regions underlying these QTL are large and show several peaks in genetic
201 differentiation (Fig. 4B). Possibly, the QTL on chromosome 2 and 3 harbor a set of linked loci
202 influencing the lunar emergence phenotype. Taken together, there are at least four unlinked
203 loci which influence this magic trait, giving it a complex genetic architecture.
204
205 **Divergent alleles are not associated with specific inversion haplotypes.** While some
206 chromosomal inversions show up as genetically differentiated blocks, in all chromosomal
207 inversions there are loci that are more differentiated than the inversion itself (Fig. 4B). This
208 implies that the inversions are not required for protecting genetic differentiation. Instead,
209 there must be sufficient levels of recombination and gene conversion to place the divergent
210 alleles in all inversion haplotypes, so that individual loci can diverge, while the surrounding
211 inversions remain less differentiated. We corroborated this scenario for the most
212 differentiated locus in the genome (Fig. 4B, red asterisk), which is the *period* locus (Fig. 4C).
213 We identified an insertion-deletion (indel) mutation in the highly differentiated first intron of
214 the *period* gene (Fig. 4C, green arrows) and re-genotyped the 48 individuals for this indel (Fig.
215 4D; Supplementary Figure 9). This confirmed that *period* alleles indeed strongly differentiated
216 between the FM and NM types. However, we could not find a close association between the
217 *period* alleles and the inversion haplotypes of chromosome 1 (Fig. 4D). This confirms that
218 differentiation of the *period* locus is not protected by the inversions.

219

220 **Candidate loci.** The existence of individual divergence peaks within the QTL, despite the
221 overlap with chromosomal inversions, allows us to pinpoint candidate genes for controlling
222 the magic trait, i.e. phenotypic divergence in lunar reproductive timing (see Supplementary
223 Table 5 for an overview). Some loci are particularly conspicuous in that they are by far the
224 most divergent within a specific QTL. The QTL in the inversion system on chromosome 1 is
225 narrow and harbors a single strongly divergent locus, which is the *period* locus (Fig. 4B, red
226 asterisk). The *period* locus is the most differentiated locus in the entire genome (maximum
227 F_{ST} = 0.86) and *period* is a core circadian clock gene. The QTL at the end of chromosome 1,
228 outside the inversion system, also shows a single strong divergence peak (Fig. 4B, red circle).
229 This is the first intron of the *plum* gene (Supplementary Figure 10). The large intron likely
230 contains regulatory regions for both *plum* and the directly adjacent *cask* gene. On
231 chromosome arm 2L the QTL has several divergence peaks, but the strongest peak hits the
232 *stat1* gene (Fig 4B, red cross; Supplementary Figure 11). *Plum*, *cask* and *stat1* are all involved
233 in nervous system development in *Drosophila melanogaster*[22-24], suggesting that divergence
234 in lunar reproductive timing involves nervous system remodeling as well as circadian
235 timekeeping.

236 **Discussion**

237

238 We found multiple large and polymorphic chromosomal inversions in the sympatric FM
239 and NM strains, covering all chromosome arms of the *C. marinus* genome. This is consistent
240 with large non-pairing and inverted regions observed in *C. marinus* polytene chromsomes[25].
241 These inversions reduce recombination, as is observed in large LD blocks (Supplementary
242 Figures 2 and 6), and possibly lock several loci involved in divergent reproductive timing into
243 supergenes (Figure 1A and 2A). However, most of these inversions are not differentiated
244 between timing types and in all inversions, there are loci more divergent than the inversion
245 itself. We show that divergent alleles at the *period* locus are not associated with specific
246 inversion haplotypes or the standard haplotype (Figure 4D). There must be considerable
247 recombination and/or gene conversion inside the inversion, placing the *period* alleles in
248 different inversion haplotypes. This is likely also true for other divergent loci and forms the
249 basis for genetic differentiation at individual loci to exceed the inversion background. We
250 propose that natural selection on the divergent loci is more important for maintaining the
251 magic trait than the lowering of recombination rate by the inversions. Ecological divergence
252 rather drives differentiation of the inversion system on chromosome 1 than being protected
253 by it. Nevertheless, differentiation of the inversion system creates a genetic substrate for
254 further ecological and reproductive divergence.

255 At the same time, QTL mapping revealed at least four independent loci underlying lunar
256 reproductive timing. Therefore, we are neither dealing with a single pleiotropic locus that
257 affects the magic trait, nor with a single supergene that would combine genes underlying both
258 ecological divergence and reproductive isolation. Our data show that an oligogenic
259 architecture can underlie ecological and reproductive divergence in sympatric populations. In

260 addition, the four QTL we identified here for the difference in lunar reproductive timing
261 between the FM and NM types in Roscoff do not overlap with the two QTL we identified
262 previously for difference in lunar reproductive timing between Normandy (Por-1SL) and
263 Basque Coast (Jean-2NM) strains[26,27]. This indicates that within the species *C. marinus*,
264 different populations have found different genetic solutions for adapting lunar reproductive
265 timing to the local tidal regime. Thus, the genetic architecture underlying the evolution of this
266 magic trait is considerably more complex than what is generally assumed in models of
267 sympatric population divergence. Notably, the first records of *C. marinus* for Roscoff date to
268 1978 and 1989 and only mention the FM type[28,29]. The authors sampled larvae, which would
269 most likely have revealed the NM type as well, had it been present at the time. This sets a
270 very recent time frame to the establishment of sympatric timing types and the evolution of
271 the magic trait. Given the two timing niches imposed by the tides, divergent selection seems
272 to lead to rapid ecological and reproductive divergence. This may eventually lead to
273 speciation through a process of evolutionary branching[30,31].

274 Genetic differentiation between the inversion haplotypes is much higher than what could
275 evolve in such a recent time-frame (Supplementary Figure 3). We thus assume that the
276 inversions were already segregating in the ancestral populations or may have recently
277 introgressed from a different geographic site.

278 The detection of *period* as the most differentiated locus between NM and FM types,
279 overlapping with a QTL for lunar reproductive timing, suggests that this circadian clock gene
280 might also be involved in lunar time-keeping. The other prominent candidates, *plum*, *cask* and
281 *stat1* are involved in neuronal or synaptic development and plasticity. Notably, *cask* is a direct
282 interaction partner of *CaMKII* [23], which we previously found to be involved in circadian timing
283 differences in *Clunio*[26]. Hence, both QTL on chromosome 1 suggest a close connection of

284 circadian and circalunar time-keeping. This is backed by recent findings in Baltic and Arctic
285 *Clunio* strains, for which genomic analysis of the loss of lunar rhythms also implied circadian
286 GO terms as the top candidates for affecting circalunar time-keeping, followed by GO terms
287 involved in nervous system development[32]. Thus, our data not only call for a new evaluation
288 of the genetic architectures underlying magic traits, but also give hints to the genetic basis of
289 the yet enigmatic circalunar clock.

290 **Methods**

291

292 **Sampling and laboratory strains.** Laboratory strains for crossing experiments and field
293 samples for whole genome sequencing of 48 individuals were available from a previous
294 study[21]. Laboratory strains were kept in standard culture conditions[33] at 20°C and under a
295 light-dark cycle of 16:8. An artificial moonlight cycle with four nights of dim light every 30 days
296 served to synchronize reproduction in the laboratory strains.

297 **Sequencing, read mapping and genotype calling.** DNA from the 48 field caught individuals
298 were subject to whole genome sequencing in the Max Planck Sequencing Centre (Cologne)
299 on an Illumina HiSeq3000 according to standard protocols. Independent sequencing runs
300 were merged with the *cat* function. Adapters were trimmed with Trimmomatic [34] using the
301 following parameters: ILLUMINACLIP <Adapter file> :2:30:10:8:true, LEADING:20,
302 TRAILING:20, MINLEN:75. Overlapping read pairs were merged with PEAR [35] using -n 75 -c 20
303 -k and mapped to the Cluma_2.0 reference genome (available in the Open Research Data
304 Repository of the Max Planck Society under DOI 10.17617/3.42NMN2; manuscript in
305 preparation) with bwa mem [36] version 0.7.15-r1140. Mapped reads were merged into a single
306 file, filtered for -q 20 and sorted with samtools v1.9 [37]. SNPs and small indels were called using
307 GATK v3.7-0-gcfedb67 [38]. All reads in the q20 sorted file were assigned to a single new read-
308 group with 'AddOrReplaceReadGroups' script with LB=whatever PL=illumina PU=whatever
309 parameters. Genotype calling was then performed with HaplotypeCaller and parameters --
310 emitRefConfidence GVCF -stand_call_conf 30, recalibration of base qualities using GATK
311 BaseRecalibrator with '-knownSites'. Preparing recalibrated BAM files with GATK PrintReads
312 using -BQSR. Recalling of genotypes using GATK HaplotypeCaller with previously mentioned
313 parameters. Individual VCF files were combined into a single file using GATK GenotypeGVCFs.

314 **Genetic differentiation (F_{ST}).** The vcf file containing GATK-called SNPs and small indels
315 from 48 Ros-2FM and Ros-2NM males was filtered with vcftools version 0.1.14 [39] for minor
316 allele frequency of 0.05, maximum of 2 alleles, minimum quality (minQ) of 20, maximum
317 missing genotypes of 20%, which finally left 721.000 variants. Genetic differentiation between
318 the two populations (fst) was estimated using vcftools parameters --weir-fst-pop --fst-
319 window-size 1 --fst-window-step 1.

320 **Long-range linkage disequilibrium (LD)**. Linkage disequilibrium was calculated between all
321 variants along the three chromosomes in each of the populations in search of the signatures
322 of large genomic inversions. Vcf file containing GATK-called SNPs and small indels from 24
323 Ros-2FM males was filtered for minor allele frequency of 0.20, maximum of 2 alleles,
324 minimum quality (minQ) of 20, maximum missing genotypes per site of 20% leaving 344.331
325 variants. The same was done for the 24 Ros-2NM individuals resulting in 352.915 variants.
326 They were converted to plink input files and plink version 1.90 beta [40] was used to calculate
327 linkage disequilibrium between all variants with parameters --r2 --inter-chr. Large LD file was
328 then sorted into 3 files, one for each chromosome. Breakpoints of large chromosomal
329 inversions were identified by plotting r^2 and manually looking for obvious breaks in the r^2
330 scores (Supplementary Figure 2 and 6; Supplementary Table 1).

331 **Principal component analysis (PCA).** PCA was run for the entire genome, for sliding
332 windows along the chromosome ("windowed PCA"), and for windows corresponding to the
333 inversions ("local PCA"). The vcf file containing GATK-called SNPs and small indels from 48
334 Ros-2FM and Ros-2NM males was filtered with vcftools version 0.1.14 [39] for minor allele
335 frequency of 0.05, minimum quality (minQ) of 20, maximum missing genotypes per site of
336 20%, leaving 703.579 variants. Principle component analysis was calculated with plink version
337 1.90 beta [40] using flags --nonfounders --pca var-wts --chr-set 3 no-xy no-mt. Windowed PCA:

338 In order to investigate regions of the genome with unusual degree of variance, PCA was run
339 in windows along the three chromosomes. Vcf files were subdivided into vcf files containing
340 variants that belong to 500kb sliding windows with 100kb steps. They were further used to
341 calculate PCA values for each window as noted above. Local PCA: To get the genotype-
342 estimate of the identified inversions (see long-range disequilibrium section), we subdivided
343 vcf files according to the estimated inversion breakpoints (Supplementary Table 4), and
344 calculated principal components 1 and 2 as described above. Number of variants that belong
345 to each window are listed in Supplementary Table 2.

346 **Admixture.** Ancestry and relatedness of the 48 individuals of Ros-2FM and Ros-2NM
347 population were investigated with admixture version 1.3.0 [41]. Admixture was run for the
348 entire genome, for sliding windows along the chromosome ("windowed admixture"), and for
349 windows corresponding to the inversions ("local admixture"). We used vcf files previously
350 described in the PCA section as input. Admixture was calculated using k of 2 to 4. Windowed
351 admixture: In order to identify regions of the genome with ancestry different from the general
352 population, we ran admixture on vcf files containing variants that belong to 500kb sliding
353 windows with 100kb steps. Local admixture: In order to genotype the inversions (see long-
354 range disequilibrium section) we calculated admixture using vcf files containing only variants
355 from those regions of the genome (Supplementary Tables 1 and 2).

356 **Observed heterozygosity.** In order to complement the windowed and local PCA and
357 admixture analyses and further corroborate the indirect genotyping of the chromosomal
358 inversions, we calculated observed heterozygosity using vcftools version 0.1.14 [39] with --het
359 flag. Proportion of observed heterozygosity was calculated for each file by dividing the
360 number of observed heterozygotes with the number of variants. Patterns of heterozygosity
361 along the chromosomes were assessed based on the vcf files containing variants that belong

362 to 500kb sliding windows with 100kb steps (see PCA and admixture section). Finally, in order
363 to calculate observed heterozygosity of the chromosomal inversions, observed
364 heterozygosity was calculated per inversion window for each individual (Supplementary
365 Tables 1 and 2). Then observed heterozygosity was calculated and plotted for each inversion
366 genotype according to local PCA and admixture.

367 **Crosses, phenotyping and genotyping.** After synchronizing the NM and FM strains by
368 applying different moonlight regimes, single pair crosses were performed. F1 egg clutches
369 were reared individually and the emerging adults were allowed to mate freely within each F1
370 family, leading to several sets of F2 families that go back to a single pair of parents. One of
371 these sets was picked for QTL mapping. As the second peak is not under clock control, but a
372 direct response to moonlight[21], analysis was restricted to individuals emerging in the first
373 peak (n=158). DNA was extracted with a salting out method[42] and amplified with the *REPLI-g*
374 *Mini Kit* (Qiagen) according to the manufacturer's instructions. The two parents were subject
375 to whole genome sequencing with 2x150bp reads on an Illumina HiSeq2000 according to
376 standard protocols. Read mapping and genotype calling were performed as described above.
377 Microsatellite and indel genotypes were obtained by custom parsing of the vcf files. PCR
378 primers (Supplementary Table 3) for amplifying the microsatellite and indel regions were
379 designed with Primer3. Indels were PCR amplified and then run and scored on 1.5% agarose
380 gels. Microsatellites were amplified with HEX- and FAM-labeled primers and run on an ABI
381 PRISM 3100 Genetic Analyzer. Chromatograms were analyzed and scored with in R with the
382 *Fragman* package[43]. The resulting genotype matrix can be found in the R/qtl input file
383 (Supplementary File 1).

384 **Linkage & QTL mapping**. Linkage analysis and QTL mapping were performed in R/qtl
385 according to the script in Supplementary File 2. Briefly, the recombination fraction and

386 distribution of alleles were checked and the correct marker order was inferred by rippling.
387 Missing data and error load were assessed. Then the best model was obtained in a stepwise
388 selection procedure (*stepwiseqtl*). Additional interactions were checked for in a two QTL scan
389 (*scantwo*), but were negligible and not considered in the final model. Finally, a model with 4
390 additive QTLs was subject to model fitting for Multiple QTL Mapping (*fitqtl*) and QTL
391 confidence intervals were estimated (*bayesint*). Composite Interval mapping was performed
392 in QTLcartographer[44] with various selection procedures (forward, backward), exclusion
393 windows (10 cM and 20 cM) and covariates (3, 5, 10; Supplementary Figure 8). The LOD
394 significance threshold was estimated by running 1000 permutations and a p value of 0.05.
395 **PacBio Data and structural variant calling.** PacBio long read data was obtained for pools
396 of 300 to 500 individuals from laboratory strains of Ros-2NM and Ros-2FM (Roscoff, France).
397 DNA was extracted as above and sequenced with standard protocols on a PacBio Sequel II at
398 the Max Planck Sequencing Facility in Cologne, Germany. Raw long-reads of Ros-2FM and Ros-
399 2NM were mapped against the reference CLUMA 2.0 (publication in preparation) using NGM-
400 LR v.0.2.7 [45] with default settings. Alignments were sorted, filtered (q20) and indexed with
401 samtools v.1.9 [37]. Three different SV-calling tools were used per population to discover SVs.
402 Sniffles v.1.0.11 [45] was run with parameters -- min_het_af 0.1 and –genotype. SVIM v1.2.0 [46]
403 was run with default parameters (svim alignment). Finally, Delly v0.8.6 [47] was run with
404 parameters lr -y pb -q 20 --svtype (INS, insertions; INV, inversions; DUP, duplications; DEL,
405 deletions) and the output VCF files per SV type were merged with a custom script. The
406 resulting VCF files were then sorted using vcf-sort and filtered for quality "PASS" using a
407 custom bash script. SURVIVOR v1.0.6 [48] was used to filter variants for a minimum size of 300
408 bp and at least 5 reads supporting each variant (SURVIVOR filter NA 300 -1 0 5). BND variants
409 detected with Sniffles and SVIM calls were excluded with a custom bash script before using

410 SURVIVOR merge (options set to 50 1 1 0 0 300) on all VCFs produced per population. The
411 merged VCFs were then used as an input to reiterate SV calling with Sniffles using the same
412 parameters as above plus --Ivcf option and --min_support 5 (minimum number of reads
413 supporting a SV).

414 ***Period* genotyping.** An insertion-deletion (indel) mutation in the *period* locus was
415 genotyped. The fragment was PCR amplified (primers: 5'-GAATACTGAGTGTAAGACTTGGC and
416 5'-ACAACGTGACCTGTGACAAT) and run and scored on 1.5% agarose gels.

417

418

419 **Data availability**

420 Sequencing data was submitted to ENA under project number PRJEB54033. The CLUMA2.0
421 reference genome is available on the Open Research Data Repository of the Max Planck
422 Society (EDMOND) under DOI 10.17617/3.42NMN2.

423

424

425 **Materials & Correspondence**

426

427 Requests should be addressed to Tobias S. Kaiser (kaiser@evolbio.mpg.de).

428 **Author contributions**

429

430 TSK conceived, designed and supervised the study, performed crosses and QTL mapping,
431 interpreted data and wrote the manuscript. DB analyzed short-read sequencing data and
432 participated in QTL mapping. CMP analyzed long-read sequencing data. All authors read and
433 approved of the manuscript.

434

435

436 **Acknowledgements**

437

438 Tjorben Nawroth, Kerstin Schäfer, Susanne Mentz, Elke Bustorf and Sina Schirmer provided
439 laboratory assistance. Jürgen Reunert provided animal care. We thank all members of the
440 research group *Biological Clocks,* as well as Diethard Tautz for feedback and discussion. This
441 work was supported by the Max Planck Society via an independent Max Planck Research
442 Group and by an ERC Starting Grant (No 802923) awarded to TSK.

443

444

445 **Competing interests**

446

447 There are no competing interests.

References

1 Mayr, E. Ecological factors in speciation. *Evolution*, 263-288 (1947).

2 Maynard Smith, J. Sympatric speciation. *Amer Natur* **100**, 637-650 (1966).

3 Via, S. Sympatric speciation in animals: the ugly duckling grows up. *Trends in ecology & evolution* **16**, 381-390 (2001).

4 Foote, A. D. Sympatric speciation in the genomic era. *Trends in Ecology & Evolution* **33**, 85-95 (2018).

5 Richards, E. J., Servedio, M. R. & Martin, C. H. Searching for sympatric speciation in the genomic era. *BioEssays* **41**, 1900047 (2019).

6 Smadja, C. M. & Butlin, R. K. A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology* **20**, 5123-5140, doi:Doi 10.1111/J.1365-294x.2011.05350.X (2011).

7 Seehausen, O. *et al.* Genomics and the origin of species. *Nature Reviews Genetics* **15**, 176-192, doi:Doi 10.1038/Nrg3644 (2014).

8 Gavrilets, S. *Fitness Landscapes and the Origin of Species*. Vol. 41 476 (Princeton University Press, 2004).

9 Taylor, R. S. & Friesen, V. L. The role of allochrony in speciation. *Mol Ecol* **26**, 3330-3342, doi:10.1111/mec.14126 (2017).

10 Hendry, A. P. & Day, T. Population structure attributable to reproductive time: isolation by time and adaptation by time. *Molecular Ecology* **14**, 901-916 (2005).

11 Filchak, K. E., Roethele, J. B. & Feder, J. L. Natural selection and sympatric divergence in the apple maggot Rhagoletis pomonella. *Nature* **407**, 739-742, doi:10.1038/35037578 (2000).

12 Doellman, M. *et al.* Genomic differentiation during speciation-with-gene-flow: comparing geographic and host-related variation in divergent life history adaptation in Rhagoletis pomonella. *Genes* **9**, 262 (2018).

13 Servedio, M. R., Van Doorn, G. S., Kopp, M., Frame, A. M. & Nosil, P. Magic traits in speciation:'magic'but not rare? *Trends in ecology & evolution* **26**, 389-397 (2011).

14 Koch, E. L. *et al.* Genetic variation for adaptive traits is associated with polymorphic inversions in Littorina saxatilis. *Evol Lett* **5**, 196-213 (2021).

15 Faria, R. *et al.* Multiple chromosomal rearrangements in a hybrid zone between Littorina saxatilis ecotypes. *Molecular ecology* **28**, 1375-1393 (2019).

16 Joron, M. *et al.* Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**, 203-U102, doi:10.1038/nature10341 (2011).

17 Kaiser, T. S. & Neumann, J. Circalunar clocks—Old experiments for a new era. *BioEssays* **43**, 2100074 (2021).

18 Neumann, D. Genetic adaptation in emergence time of *Clunio* populations to different tidal conditions. *Helgoländer wissenschaftliche Meeresuntersuchungen* **15**, 163-171 (1967).

19 Kaiser, T. S. in *Annual, Lunar, and Tidal Clocks* (eds Hideharu Numata & Barbara Helm) Ch. 7, 121-141 (Springer Japan, 2014).

20 Kaiser, T. S., Neumann, D. & Heckel, D. G. Timing the tides: Genetic control of diurnal and lunar emergence times is correlated in the marine midge *Clunio marinus*. *BMC Genetics* **12**, 49, doi:10.1186/1471-2156-12-49 (2011).

21 Kaiser, T. S., von Haeseler, A., Tessmar-Raible, K. & Heckel, D. G. Timing strains of the marine insect Clunio marinus diverged and persist with gene flow. *Molecular Ecology* **30**, 1264-1280, doi:https://doi.org/10.1111/mec.15791 (2021).

22 Yu, X. M. *et al.* Plum, an immunoglobulin superfamily protein, regulates axon pruning by facilitating TGF-β signaling. *Neuron* **78**, 456-468 (2013).

23 Gillespie, J. M. & Hodge, J. J. CASK regulates CaMKII autophosphorylation in neuronal growth, calcium signaling, and learning. *Frontiers in molecular neuroscience* **6**, 27 (2013).

24 Ngo, K. T. *et al.* Concomitant requirement for Notch and Jak/Stat signaling during neuro-epithelial differentiation in the Drosophila optic lobe. *Developmental biology* **346**, 284-295 (2010).

25 Michailova, P. Comparative External Morphological and Karyological Characteristics of European Species of Genus Clunio HALIDAY 1855 *VII. International Symposium of Chironomidae*, 9-15 (1980).

26 Kaiser, T. S. *et al.* The genomic basis of circadian and circalunar timing adaptations in a midge. *Nature* **540**, 69-73, doi:10.1038/nature20151 (2016).

27 Kaiser, T. S. & Heckel, D. G. Genetic Architecture of Local Adaptation in Lunar and Diurnal Emergence Times of the Marine Midge *Clunio marinus* (Chironomidae, Diptera). *PLoS ONE* **7**, e32092, doi:10.1371/journal.pone.0032092 (2012).

28 Neumann, D. & Heimbach, F. in *Cyclic Phenomena in Marine Plants and Animals* (eds E. Naylor & R.G. Hartnoll) 423-433 (Pergamon Press, 1979).

29 Neumann, D. Circadian Components of Semilunar and Lunar Timing Mechanisms. *Journal of Biological Rhythms* **4**, 285-294 (1989).

30 Doebeli, M. & Dieckmann, U. Speciation along environmental gradients. *Nature* **421**, 259-264 (2003).

31 Doebeli, M. & Dieckmann, U. Evolutionary branching and sympatric speciation caused by different types of ecological interactions. *The american naturalist* **156**, S77-S101 (2000).

32 Fuhrmann, N., Prakash, C. & Kaiser, T. S. Polygenic adaptation from standing genetic variation allows rapid ecotype formation. *bioRxiv* (2021).

33 Neumann, D. Die lunare und tägliche Schlüpfperiodik der Mücke *Clunio* - Steuerung und Abstimmung auf die Gezeitenperiodik. *Zeitschrift für Vergleichende Physiologie* **53**, 1-61 (1966).

34 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).

35 Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614-620 (2014).

36 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

37 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

38 McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).

538 39 Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158
539 (2011).
540 40 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer
541 datasets. *Gigascience* **4**, s13742-13015-10047-13748 (2015).
542 41 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in
543 unrelated individuals. *Genome research* **19**, 1655-1664 (2009).
544 42 Reineke, A., Karlovsky, P. & Zebitz, C. P. W. Preparation and purification of DNA from
545 insects for AFLP analysis. *Insect Molecular Biology* **7**, 95-99 (1998).
546 43 Covarrubias-Pazaran, G., Diaz-Garcia, L., Schlautman, B., Salazar, W. & Zalapa, J.
547 Fragman: an R package for fragment analysis. *BMC genetics* **17**, 1-8 (2016).
548 44 Windows QTL Cartographer (Department of Statistics, North Carolina State University,
549 Raleigh, NC, 2006).
550 45 Sedlazeck, F. J. *et al.* Accurate detection of complex structural variations using single-
551 molecule sequencing. *Nat Methods* **15**, 461-468, doi:10.1038/s41592-018-0001-7
552 (2018).
553 46 Heller, D. & Vingron, M. SVIM: structural variant identification using mapped long
554 reads. *Bioinformatics* **35**, 2907-2915 (2019).
555 47 Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-
556 read analysis. *Bioinformatics* **28**, I333-I339, doi:10.1093/bioinformatics/bts378
557 (2012).
558 48 Jeffares, D. C. *et al.* Transient structural variations have strong effects on quantitative
559 traits and reproductive isolation in fission yeast. *Nature communications* **8**, 1-11
560 (2017).
561

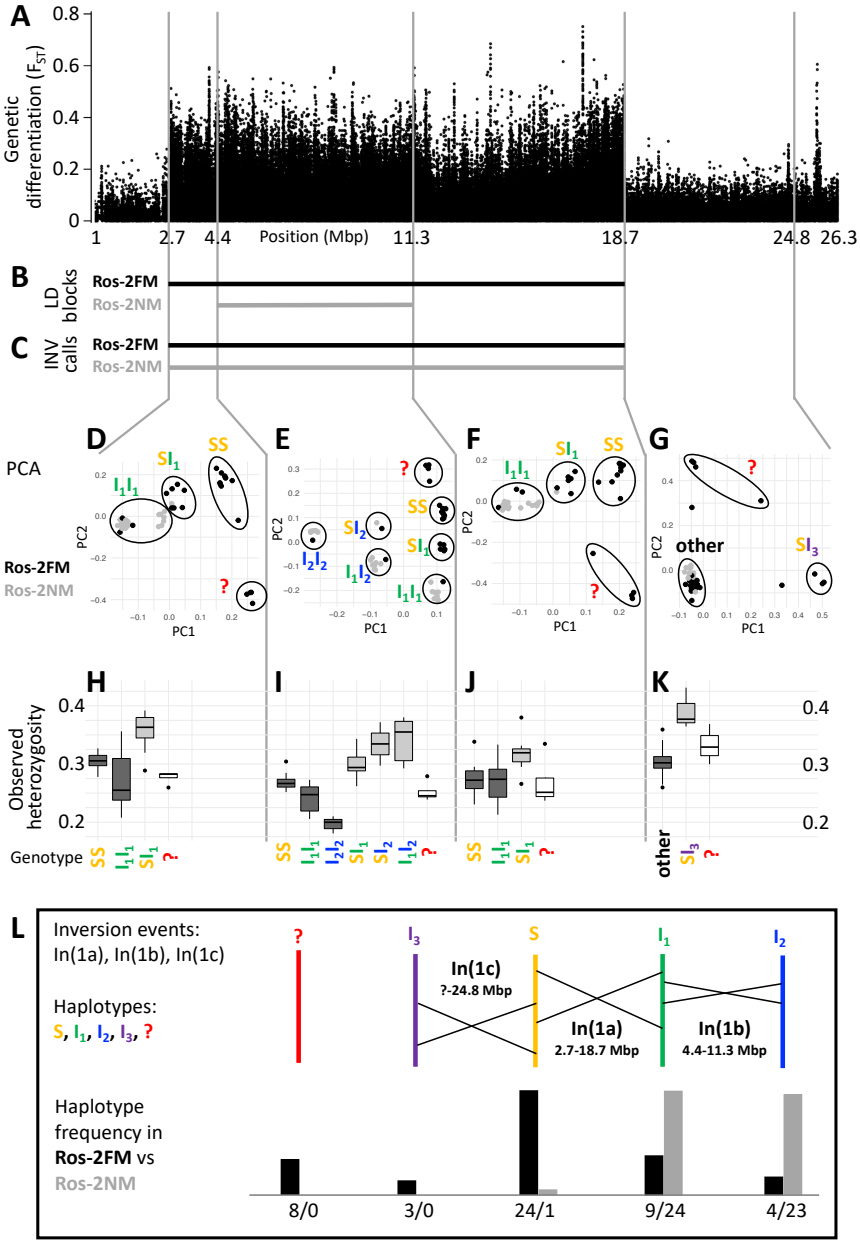


Figure 1 A complex inversion system on chromosome 1. (**A**) Chromosome 1 harbors a block of markedly elevated genetic differentiation. (**B**) This genomic block coincides with two windows of elevated long-range LD in the FM and NM strains. (**C**) Structural variant (SV) calling from long read sequencing data supports that the larger LD block is due to an inversion. (**D-G**) Principal component analysis (PCA) for chromosomal sub-windows corresponding to suggested inversions separates the individuals into clusters corresponding to standard haplotype homozygotes (SS), inversion homozygotes (I_1I_1, I_2I_2) and inversion heterozygotes (SI_1, SI_2, SI_3, I_1I_2). The four individuals which were already found distinct in whole genome analysis (red question mark) cannot be assessed. (**H-K**) Observed heterozygosity is clearly elevated in inversion heterozygotes, underpinning substantial genetic differentiation between the inversions. (**L**) A schematic overview of the sequence of inversion events and the resulting haplotypes. The frequencies of the haplotypes differ markedly between the FM and NM types.

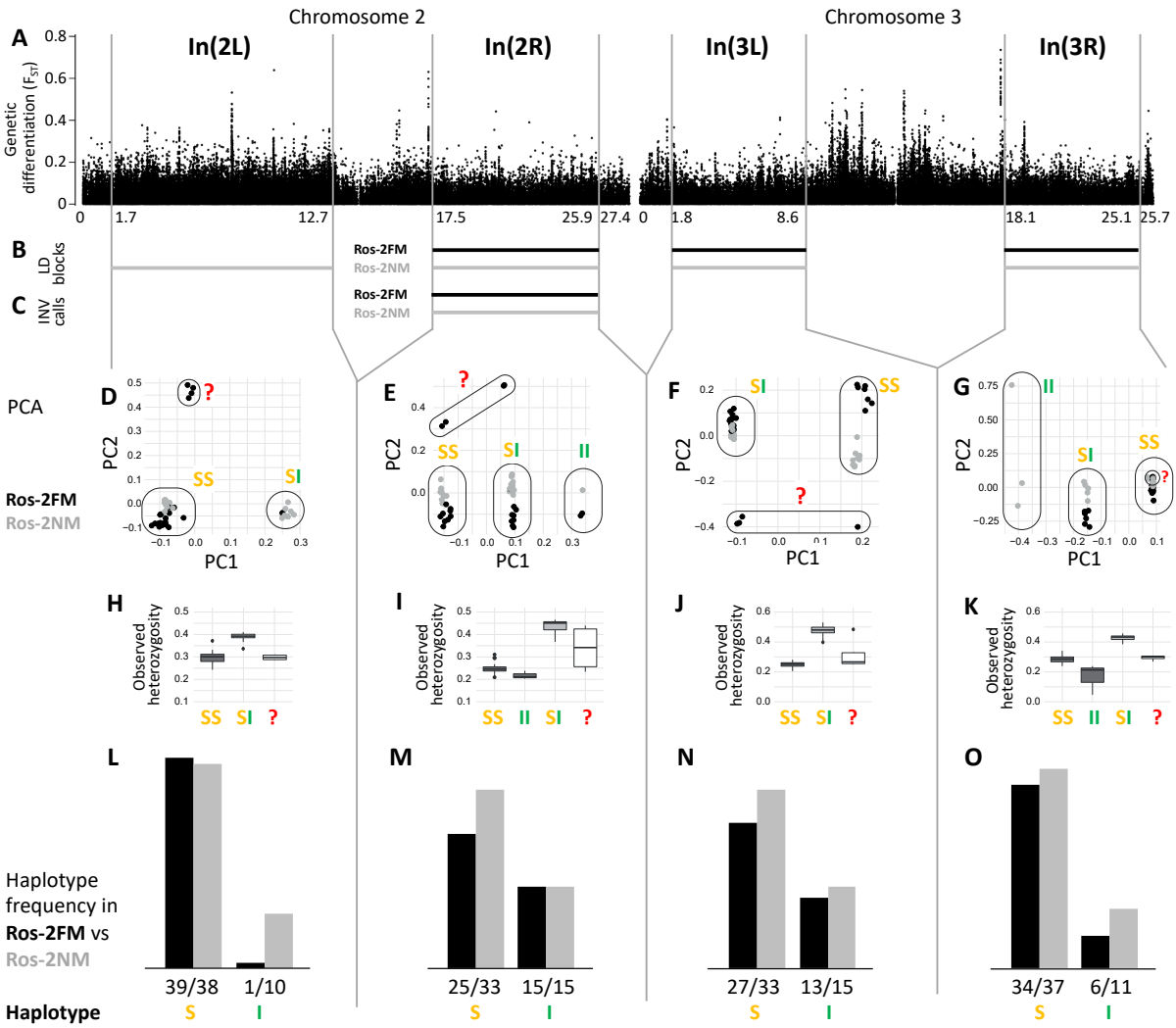
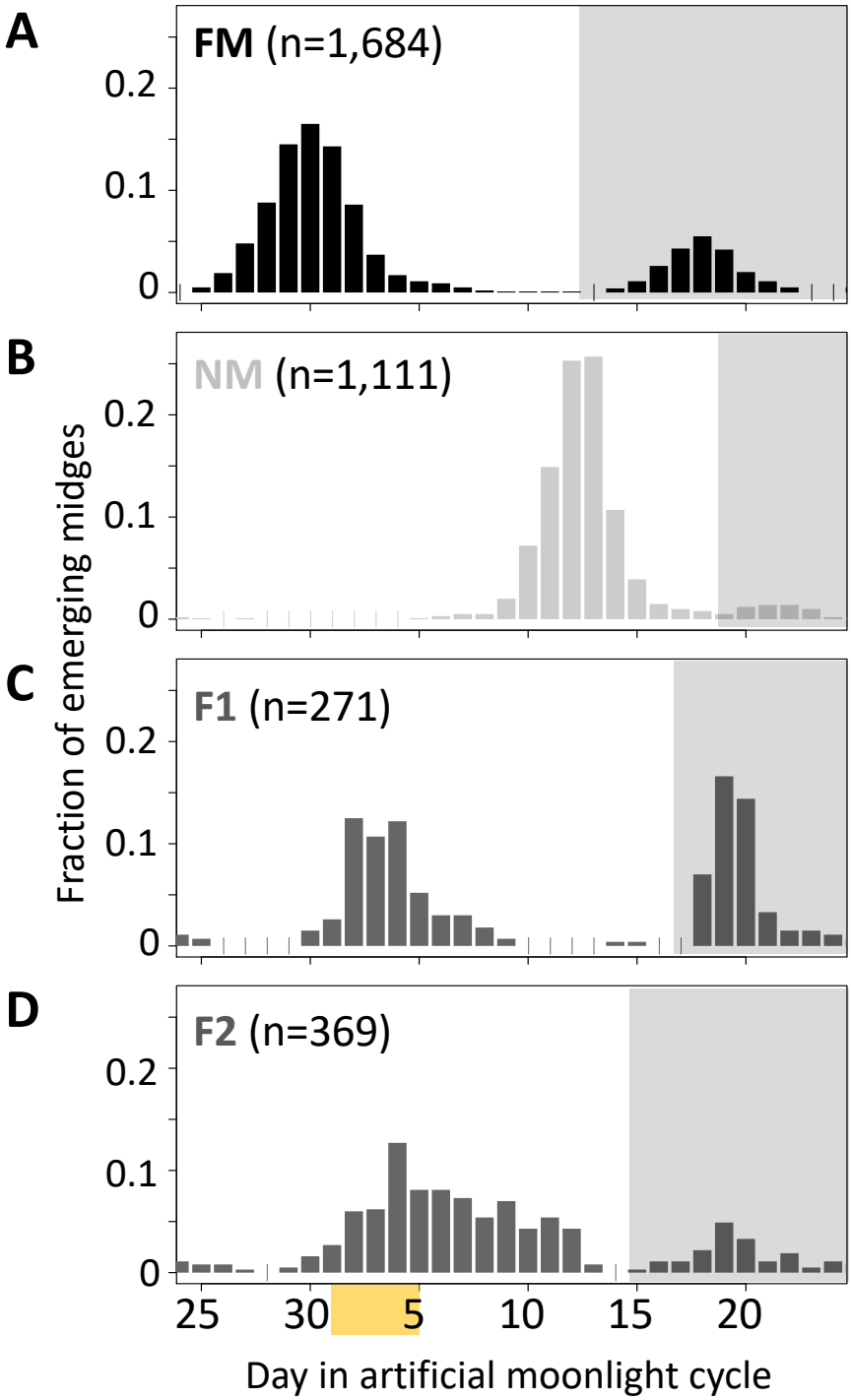


Figure 2 Inversions in chromosomes 2 and 3. (**A**) Chromosome arm 2L harbors a block of mildly elevated genetic differentiation. (**B**) Several blocks of long-range LD point to additional inversions in the FM and NM strains, one on each chromosome arm. (**C**) Structural variant (SV) calling from long read sequencing data supports the inversion on chromosome arm 2R. (**D-G**) Principal component analysis (PCA) for chromosomal sub-windows corresponding to suggested inversions separates the individuals into clusters corresponding to standard haplotype homozygotes (SS), inversion homozygotes (II) and inversion heterozygotes (SI). The four individuals which were already found distinct in whole genome analysis (red question mark) cannot be assessed. (**H-K**) Observed heterozygosity is clearly elevated in inversion heterozygotes, underpinning substantial genetic differentiation between the inversions. (**L-O**) The frequencies of the haplotypes do not differ much between the FM and NM types.



593

594 **Figure 3 Lunar emergence time is heritable.** In a cross between FM type (**A**) and NM type
595 (**B**), the F1 hybrids emerges intermediate between the parents (**C**). In the F2 (**D**), the
596 phenotypes spread out again, but do not segregate completely, suggesting that more than
597 one locus controls lunar emergence time. The peaks around day 20 (grey shading) are direct
598 responses to the artificial moonlight treatment (see [21]). They do not occur in the field and
599 were not considered in our analyses. Yellow shading indicates the days with artificial
600 moonlight treatment in the laboratory cultures.
601

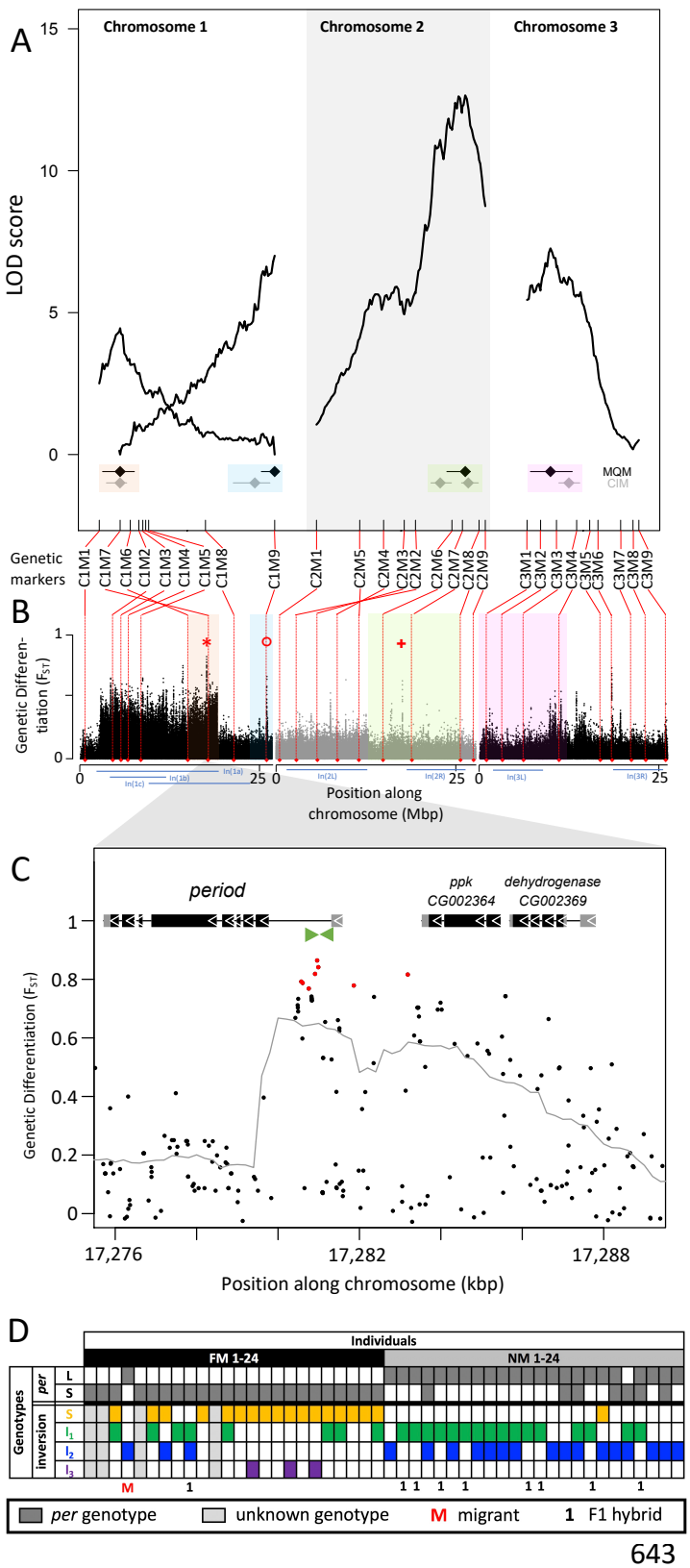
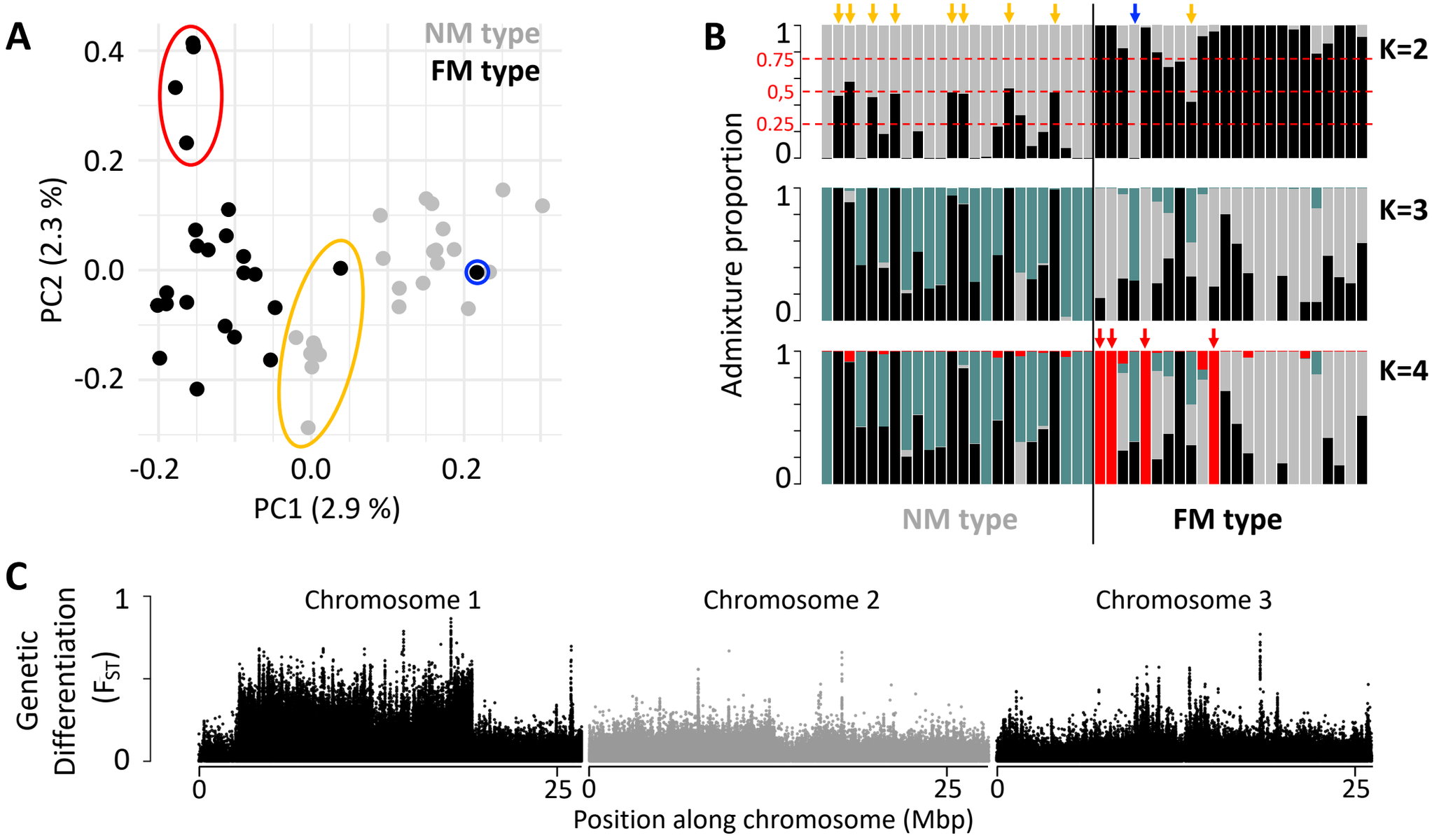


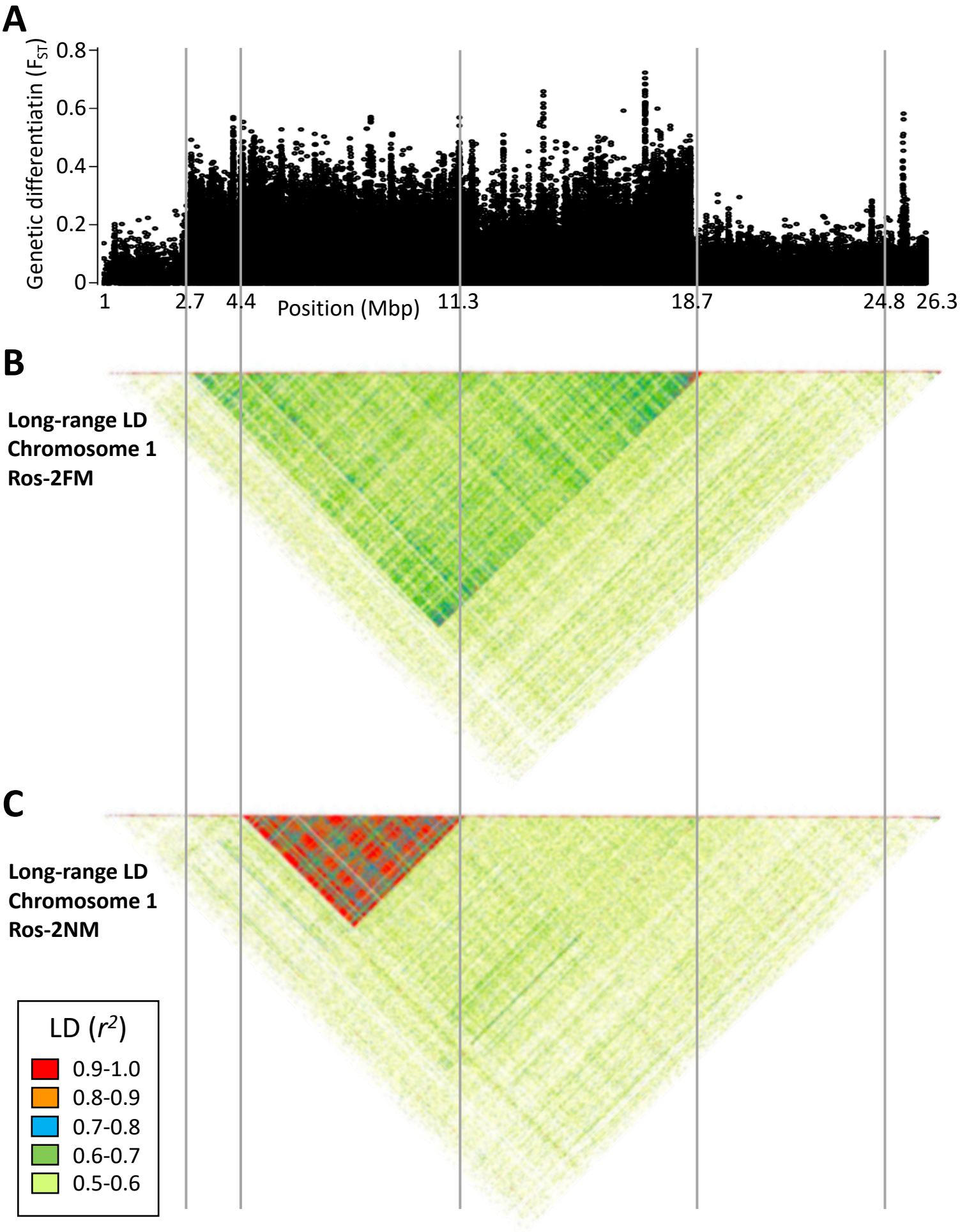
Figure 4 Quantitative trait loci (QTL) and genomic regions underlying FM vs NM emergence. (**A**) Multiple QTL Mapping (MQM) identified four significant QTL controlling lunar emergence time. Confidence intervals were determined in MQM and Composite interval mapping (CIM). The marker order supports In(1a) and the nested In(1b) on chromosome 1, as well as In(2L) on chromosome 2. (**B**) QTL intervals were colour-coded, transferred to the reference sequence and assessed for loci which are strongly differentiated (F_{ST}) between the FM and NM types. Inversions are represented by blue bars below the plot. The QTL in In(1a) overlaps with the most differentiated locus in the genome (red asterisk) – the *period* locus. Other markedly differentiated loci within the QTL intervals are the *plum/cask* locus (red circle) and the *stat1* locus (red cross). (**C**) Genetic differentiation is strongest in the first intron and the intergenic region just upstream of the *period* gene. (**D**) The *period* locus was re-genotyped based on an insertion-deletion mutation in the first intron (green arrows in (C)). The long (L) and short (S) *period* alleles are only loosely associated with inversion haplotypes.

643

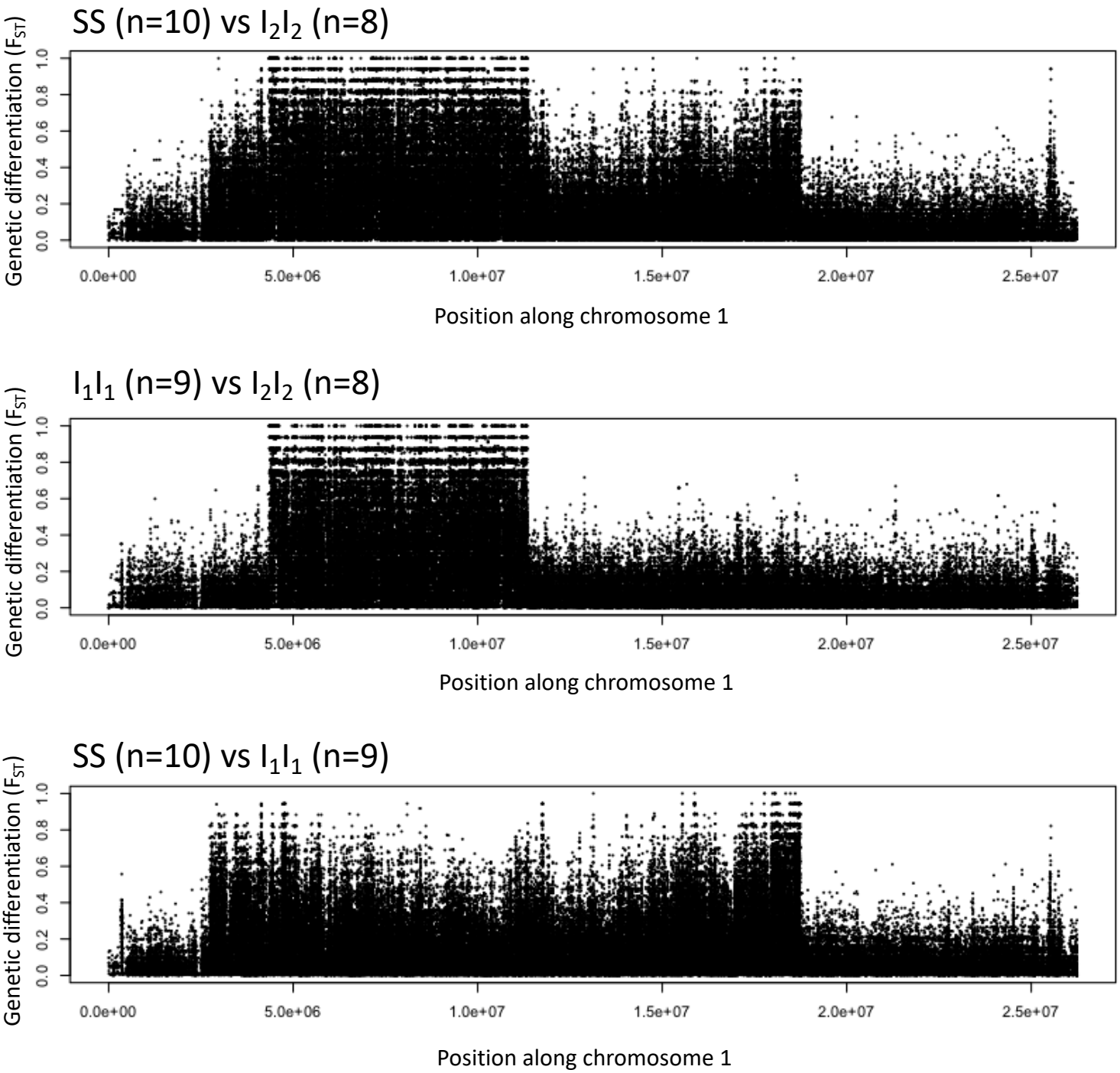


Supplementary Figure 1 Population structure of the FM and NM types in Roscoff based on 721,000 genetic variants.

(**A, B**) Principal component analysis (PCA; **A**) and admixture analysis (**B**) identify one migrant in time (blue; pure NM genotype caught at full moon), many potential F1 hybrids (yellow) and four individuals in the FM strain that appear genetically distict from all other samples (red). (**C**) Global genetic differentiation is limited, but there is a block of strong differentiation on chromosome 1.

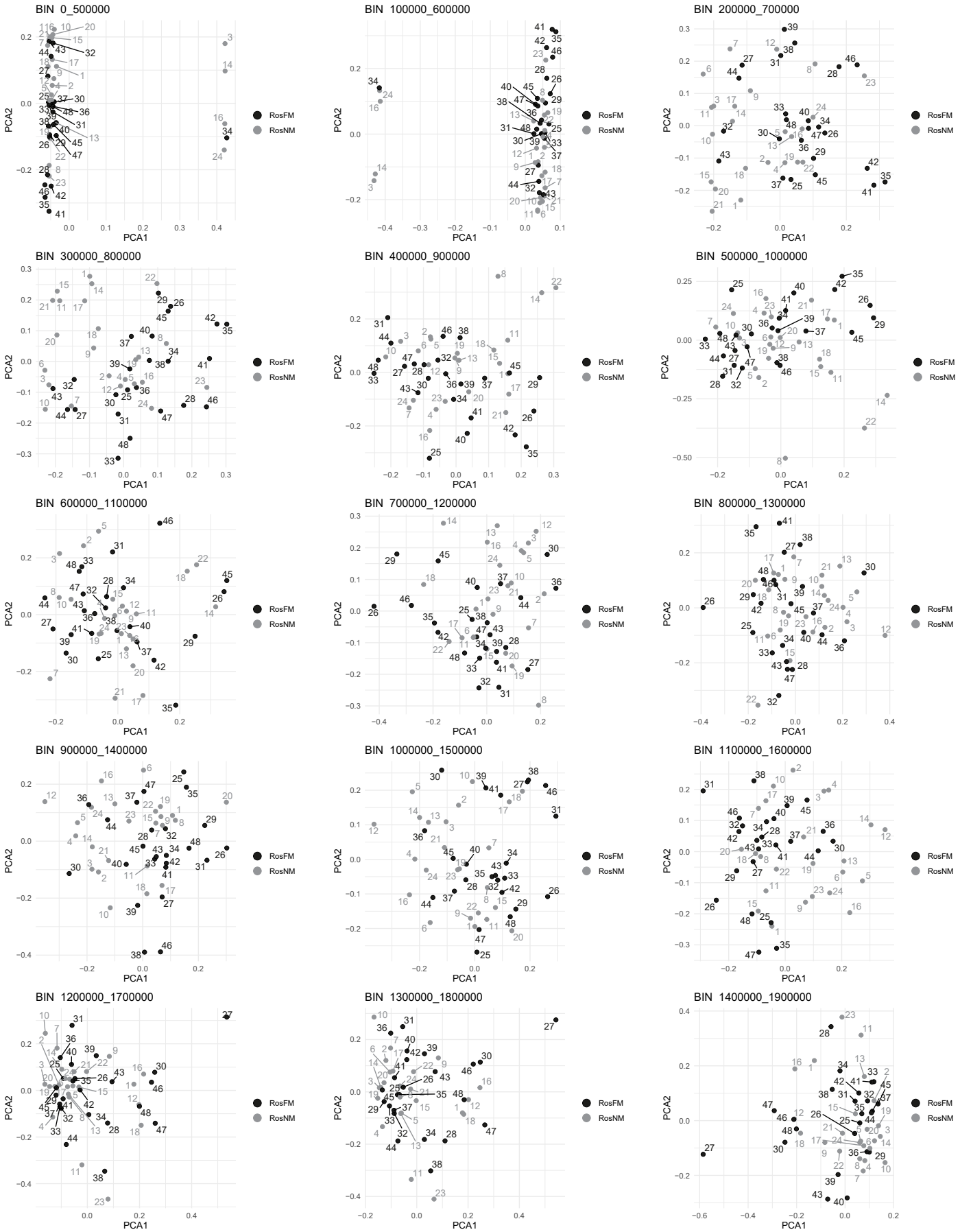


Supplementary Figure 2 Long range linkage disequilibrium along chromosome 1. Blocks of elevated genetic differentiation (**A**) correspond to blocks of elevated long-range LD in the FM type (**B**) and the NM type (**C**). Elevated long-range LD suggets that the inversion is polymorphic in the respective population.

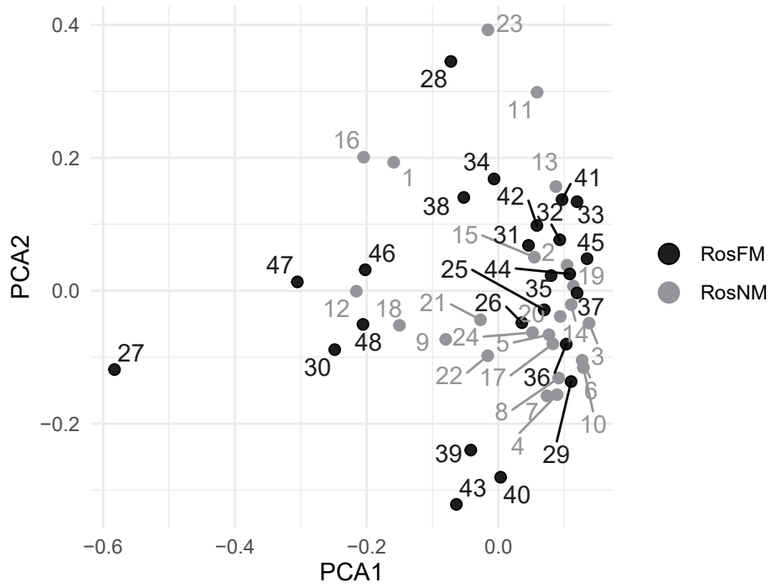


Supplementary Figure 3 Genetic differentiation between homozygotes of the standard haplotype (SS) and inversion haplotypes (I_1I_1, I_2I_2).

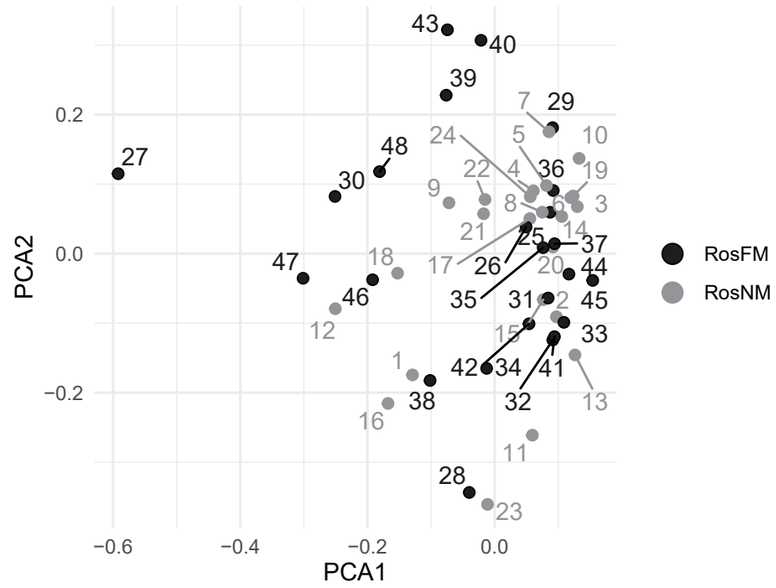
Supplementary Figure 4 Windowed PCA along chromosome 1. (500kb sliding windows with 100kb steps).



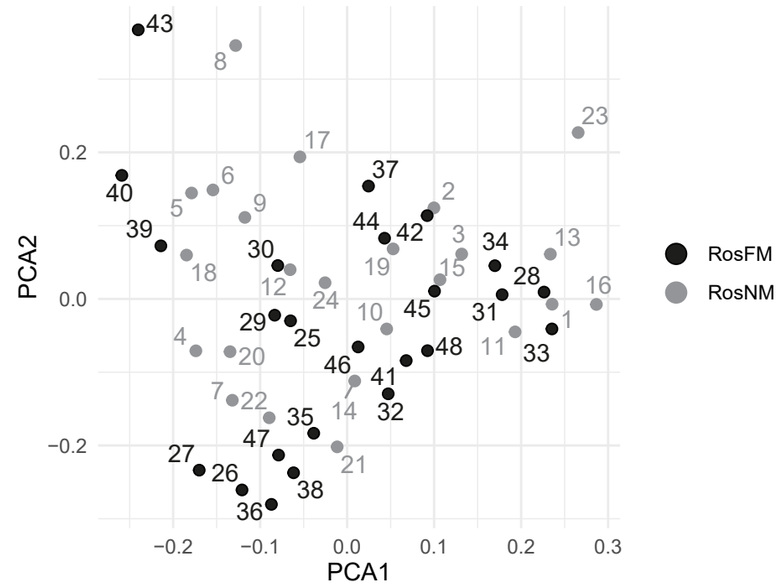
BIN 1500000_2000000



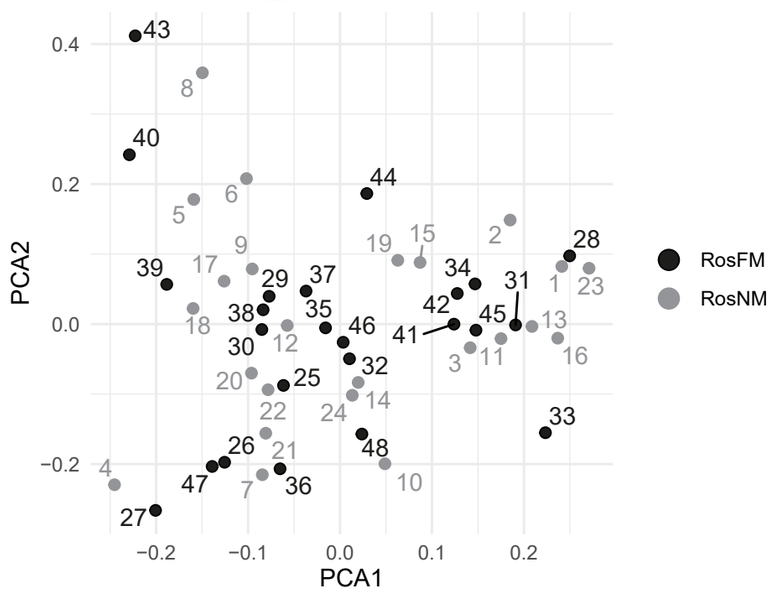
BIN 1600000_2100000



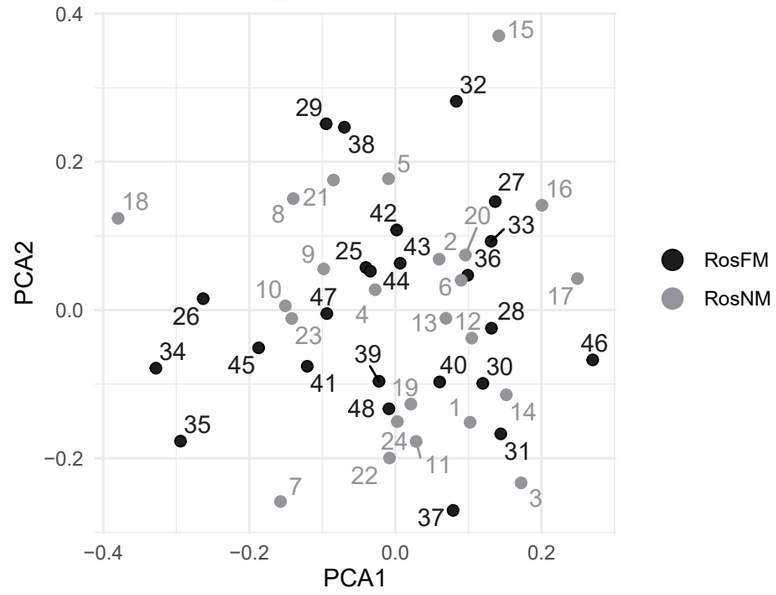
BIN 1700000_2200000



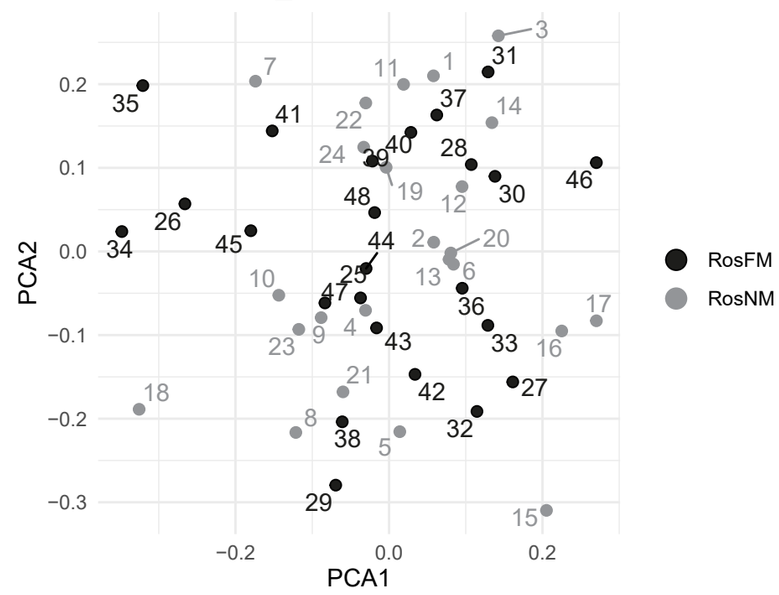
BIN 1800000_2300000



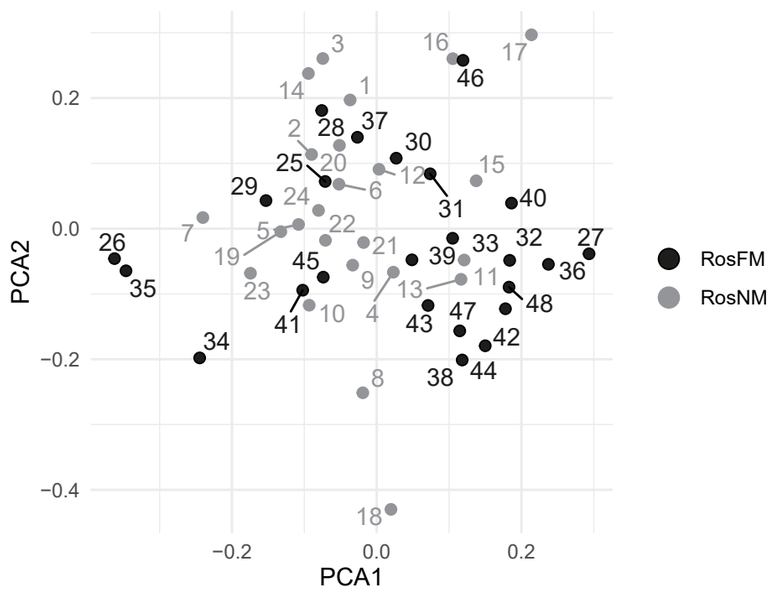
BIN 1900000_2400000



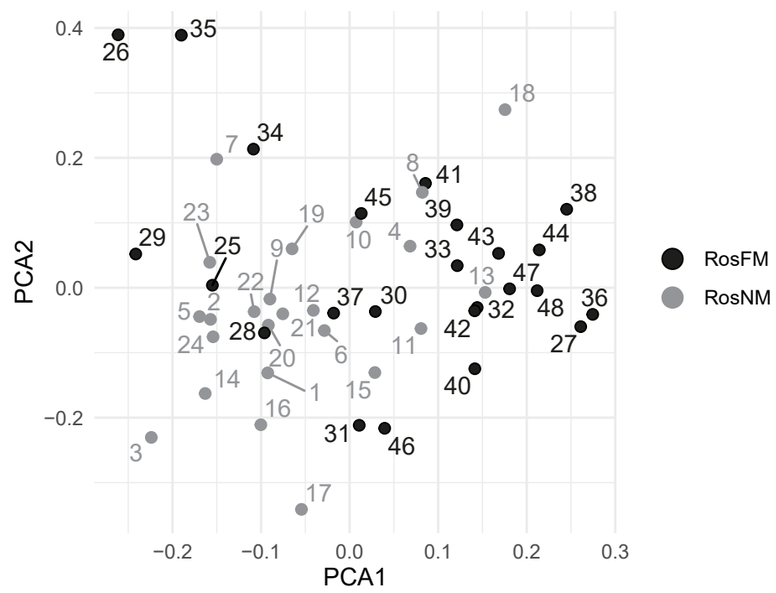
BIN 2000000_2500000



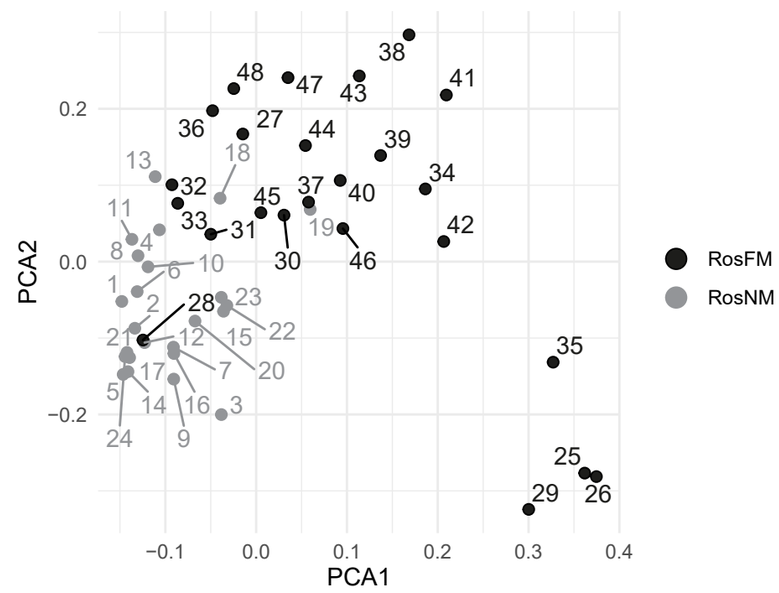
BIN 2100000_2600000



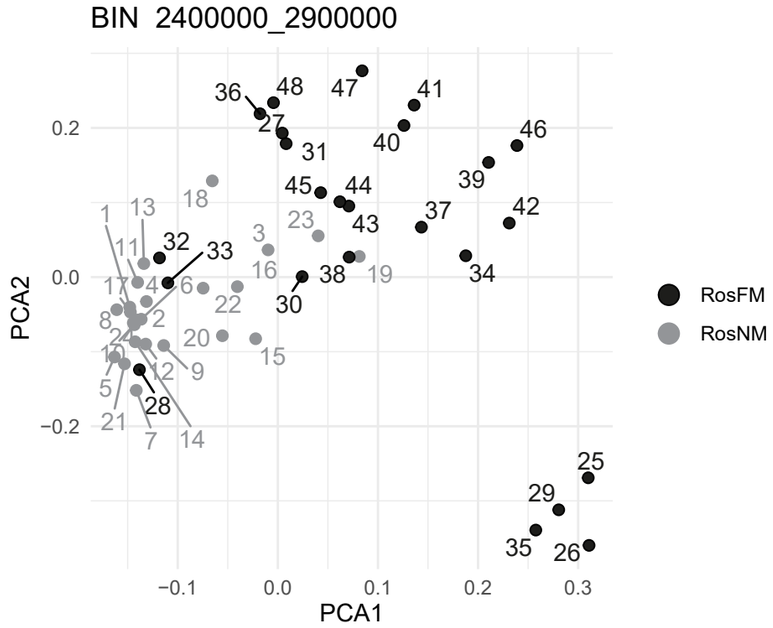
BIN 2200000_2700000



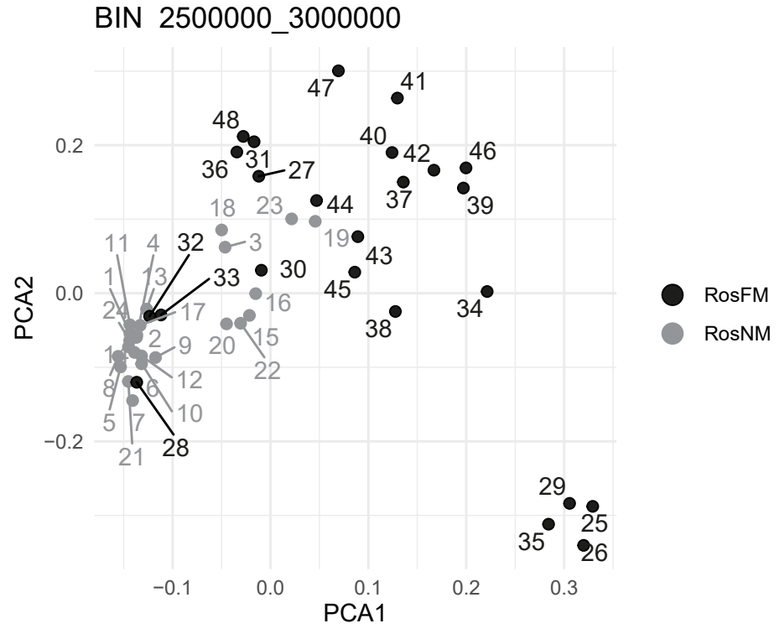
BIN 2300000_2800000



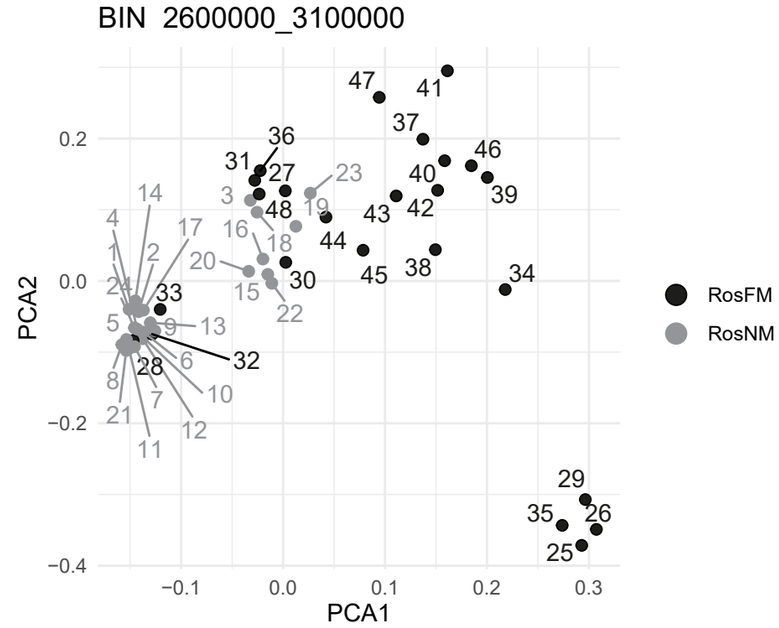
BIN 2400000_2900000



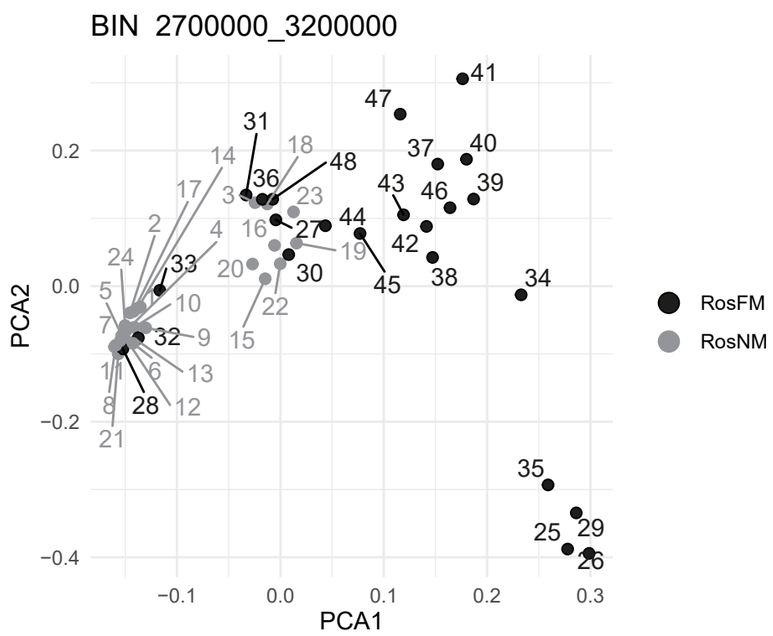
BIN 2500000_3000000



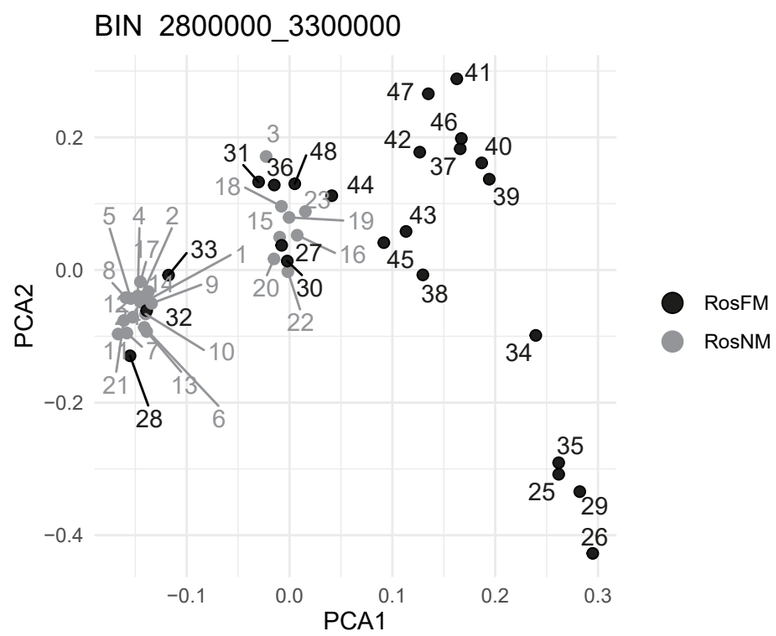
BIN 2600000_3100000



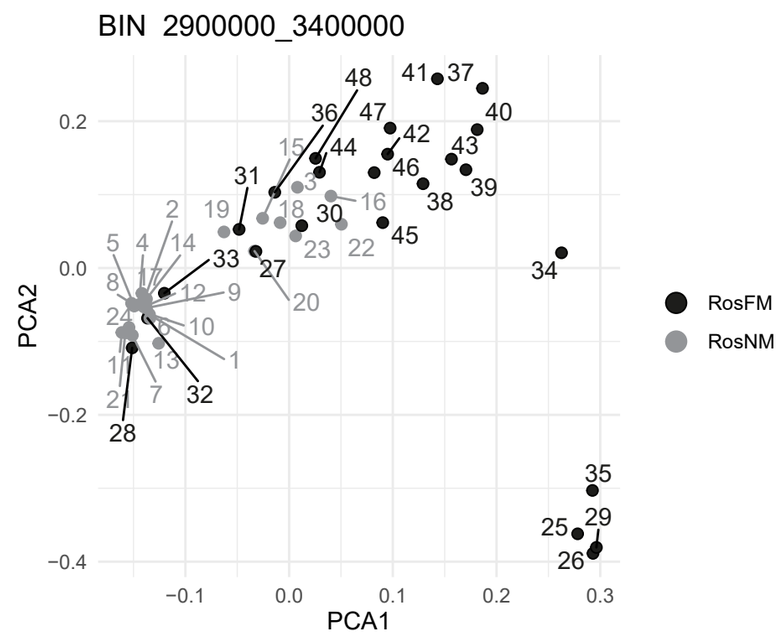
BIN 2700000_3200000



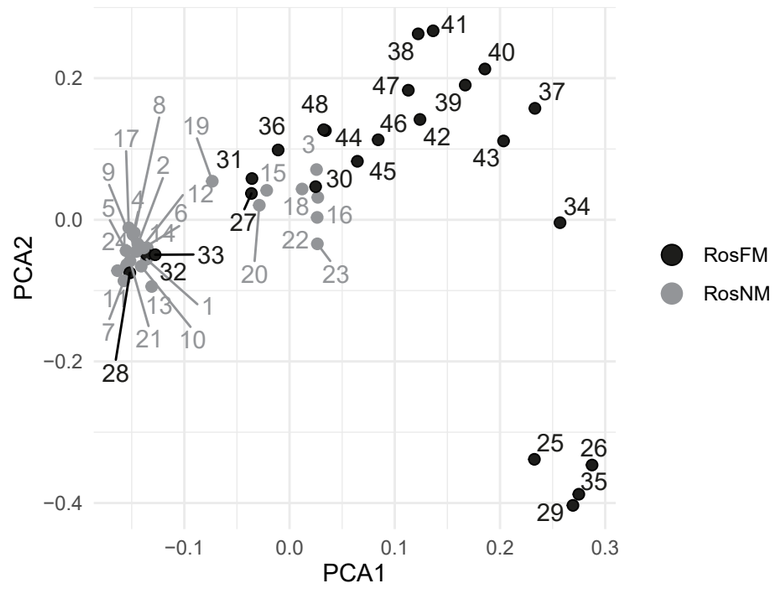
BIN 2800000_3300000



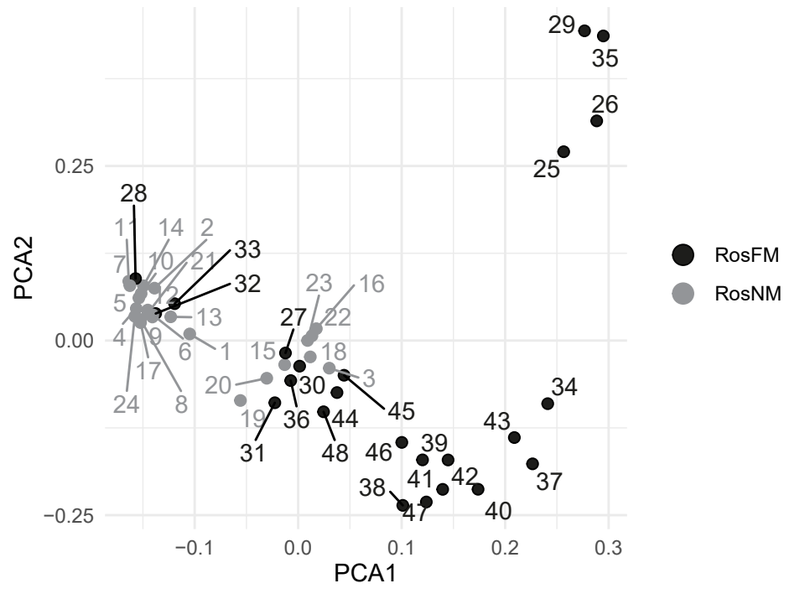
BIN 2900000_3400000



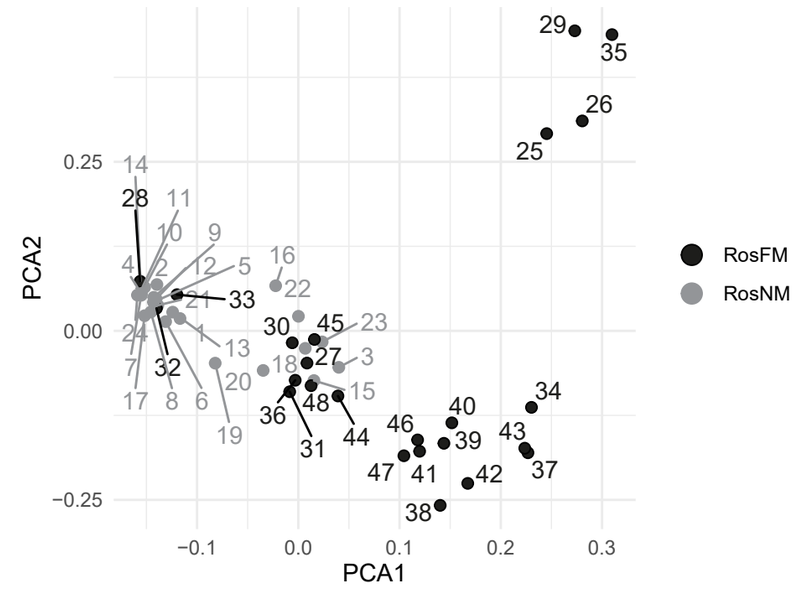
BIN 3000000_3500000



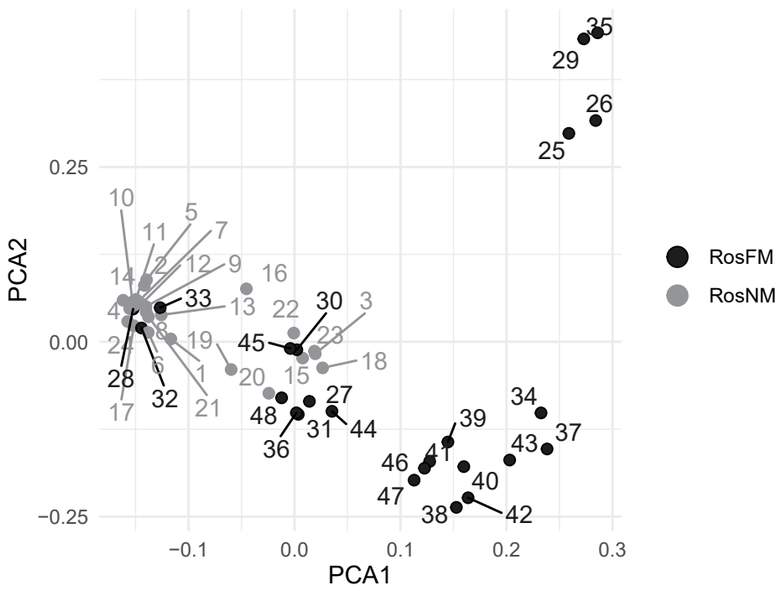
BIN 3100000_3600000



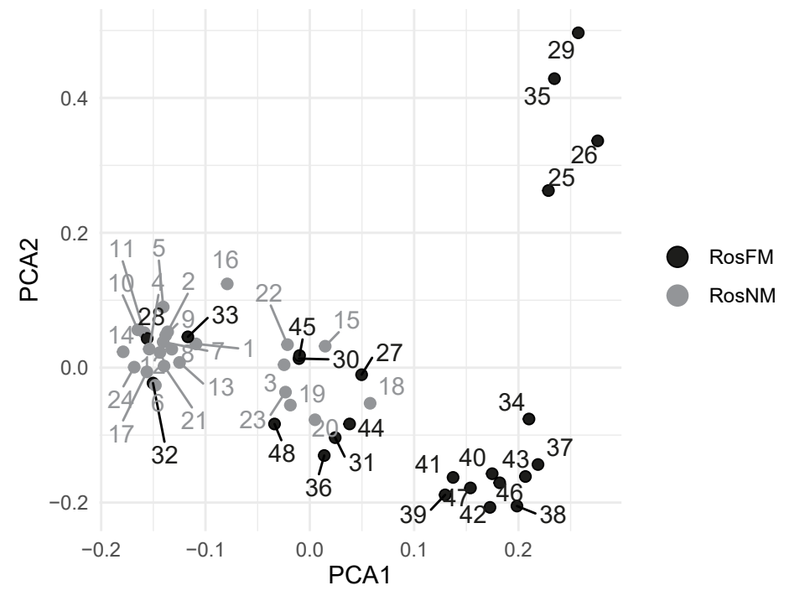
BIN 3200000_3700000



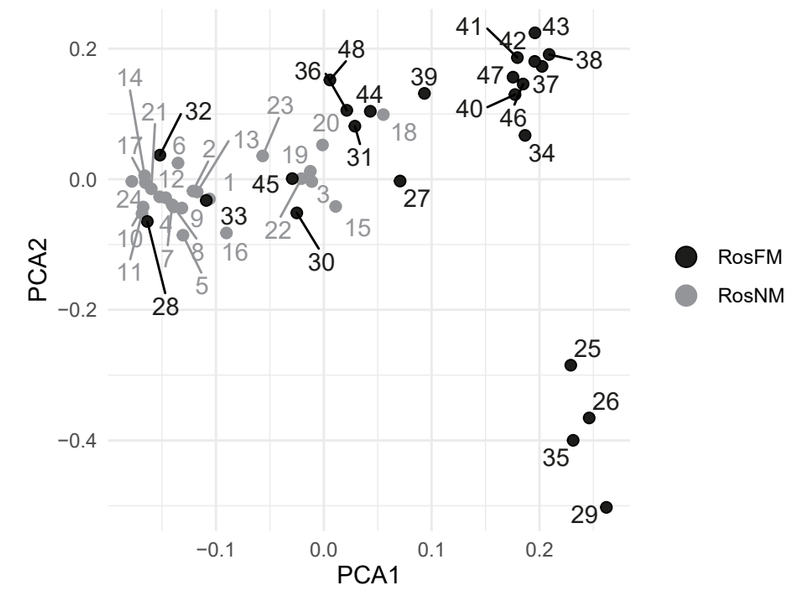
BIN 3300000_3800000



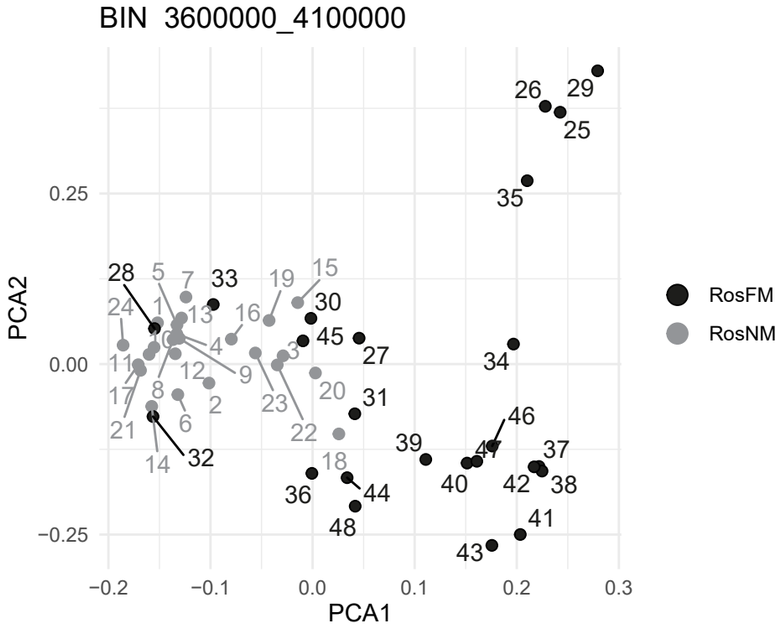
BIN 3400000_3900000



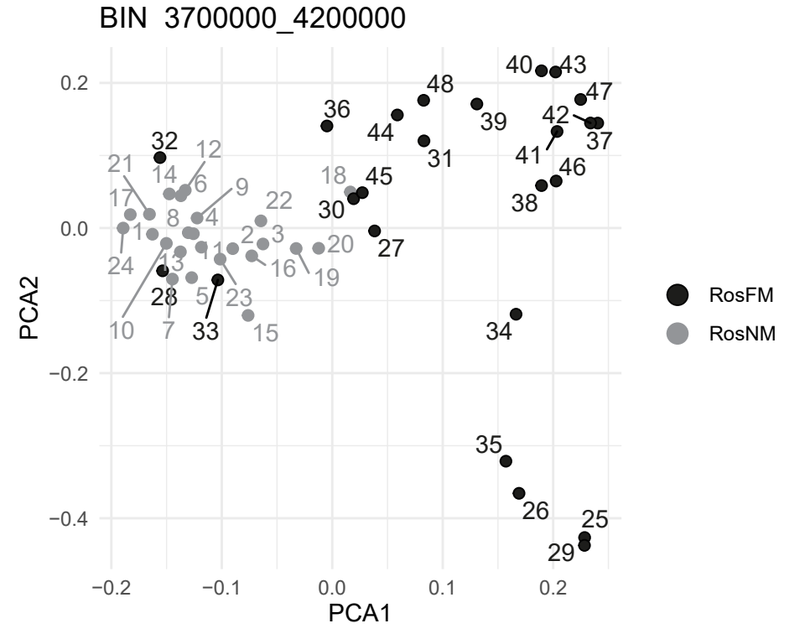
BIN 3500000_4000000



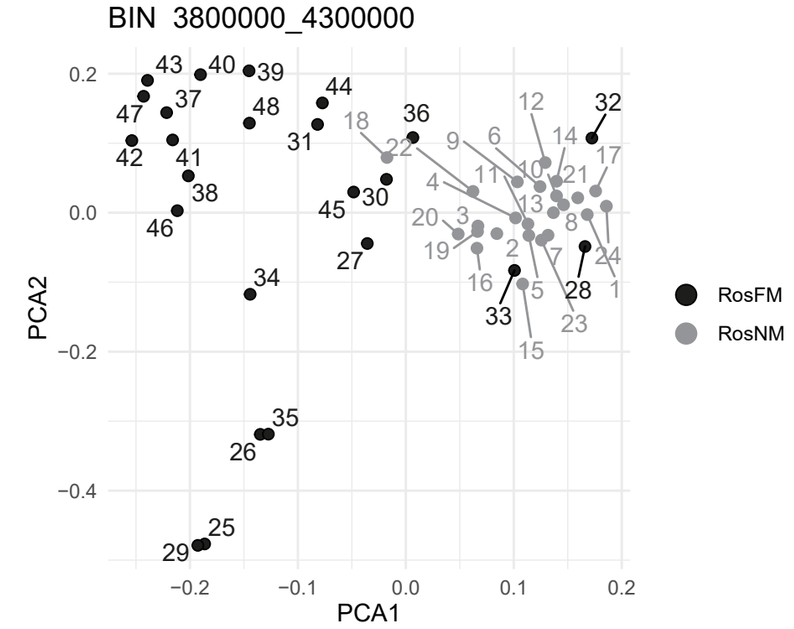
BIN 3600000_4100000



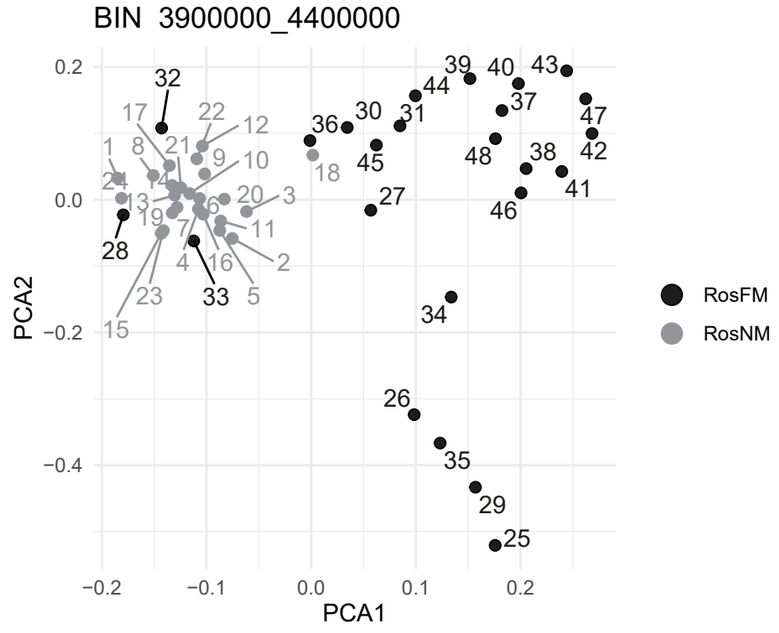
BIN 3700000_4200000



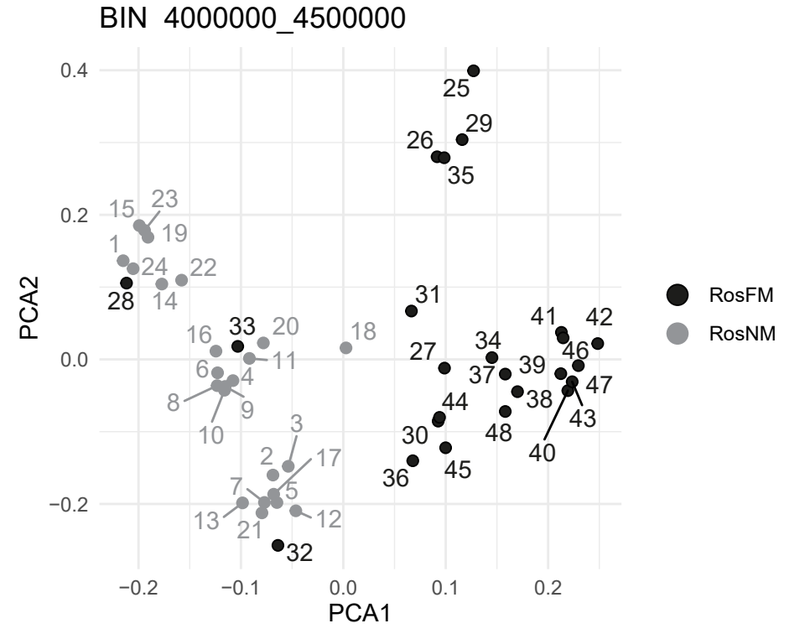
BIN 3800000_4300000



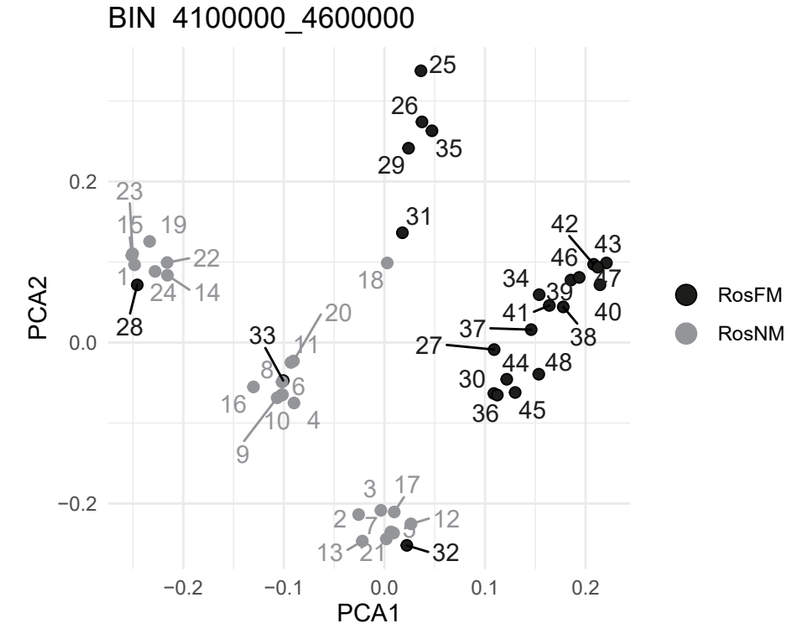
BIN 3900000_4400000



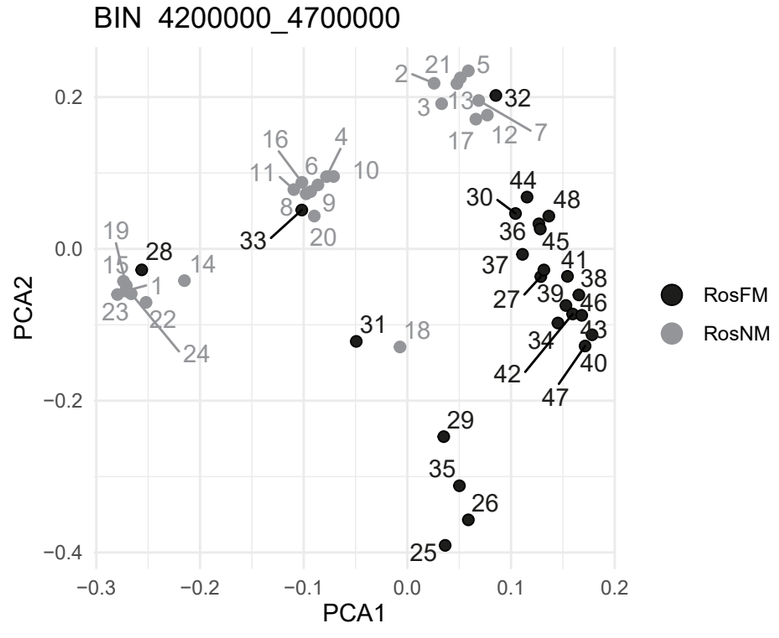
BIN 4000000_4500000



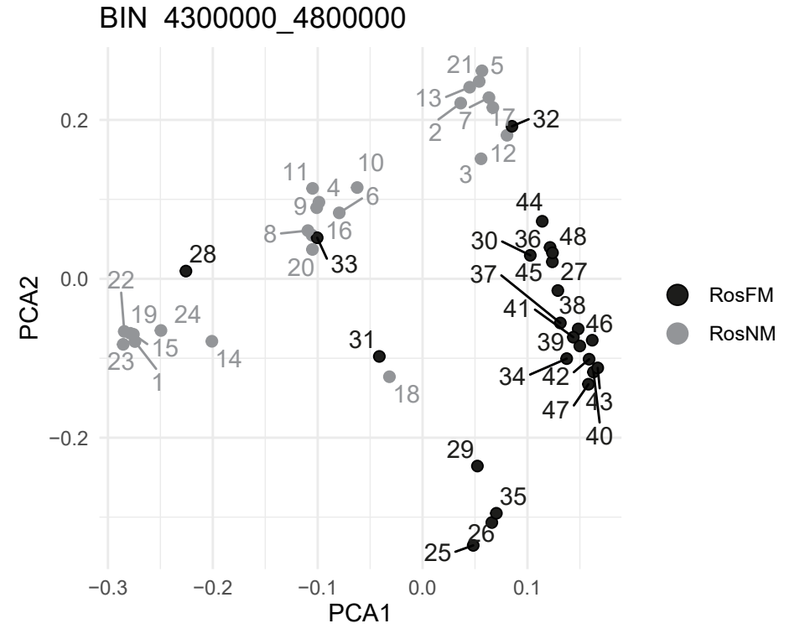
BIN 4100000_4600000



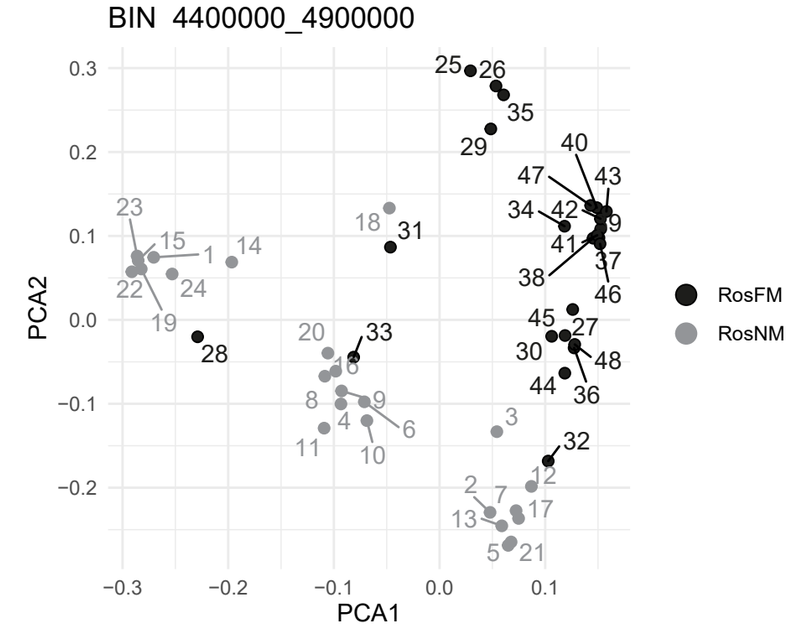
BIN 4200000_4700000



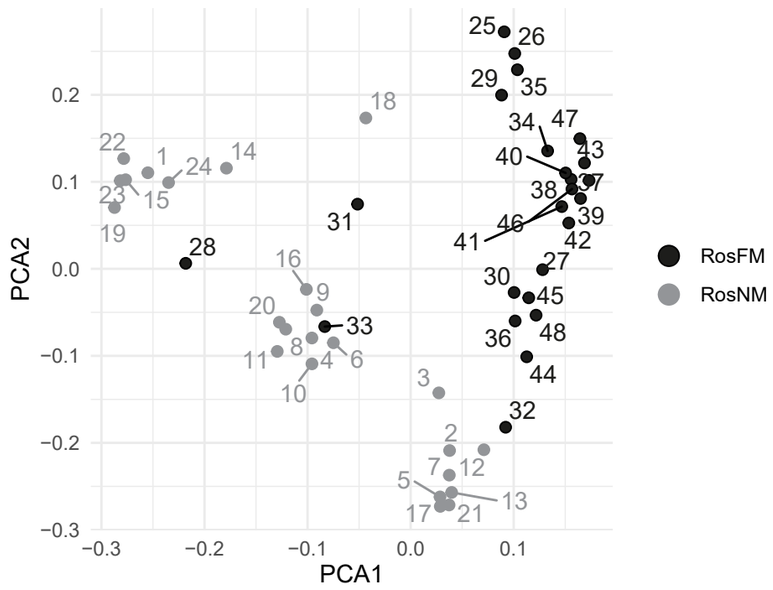
BIN 4300000_4800000



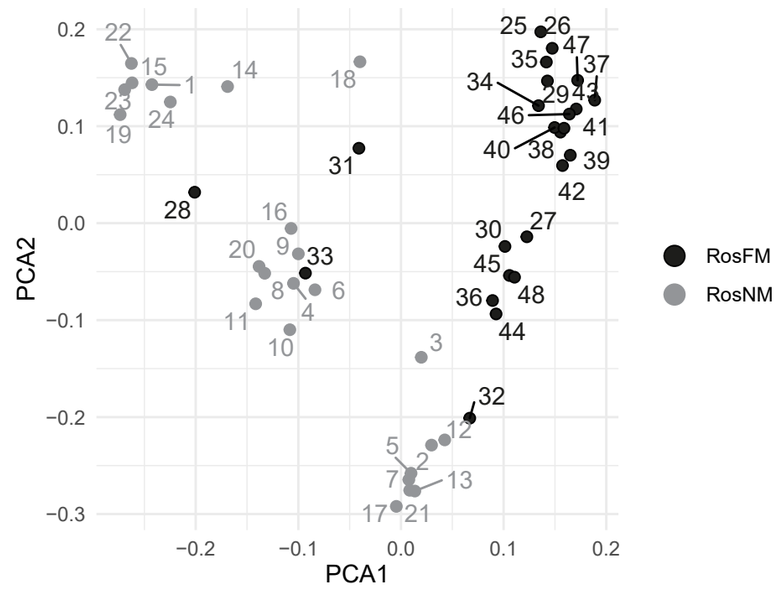
BIN 4400000_4900000



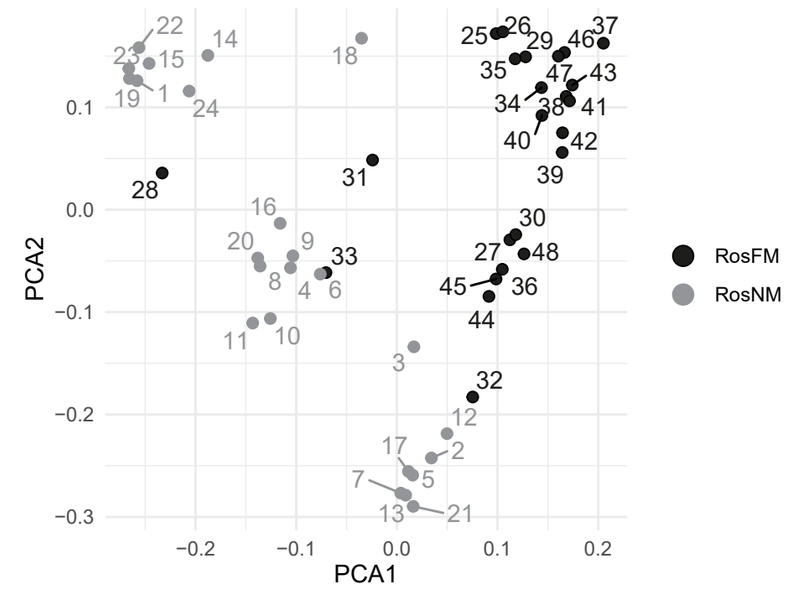
BIN 4500000_5000000



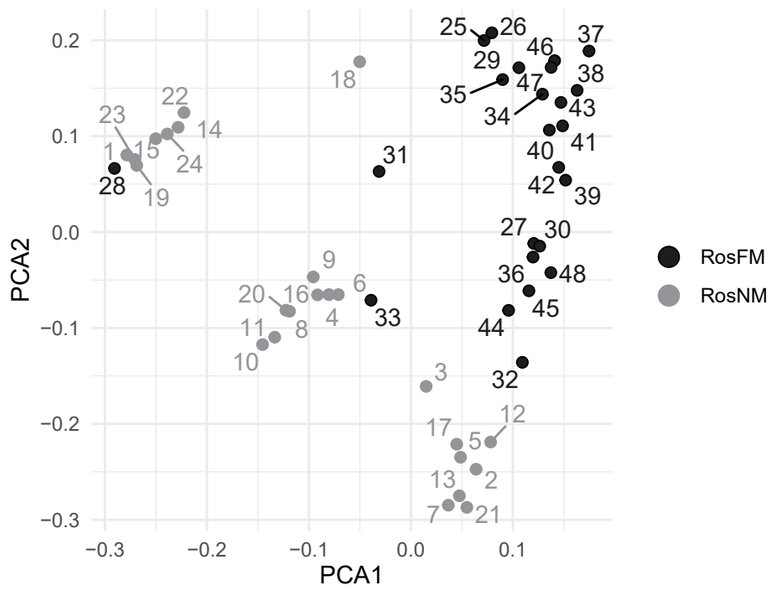
BIN 4600000_5100000



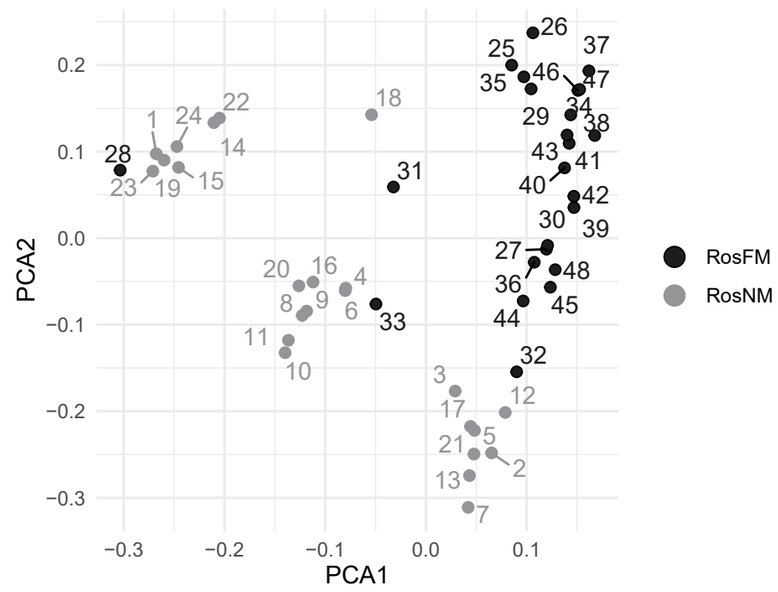
BIN 4700000_5200000



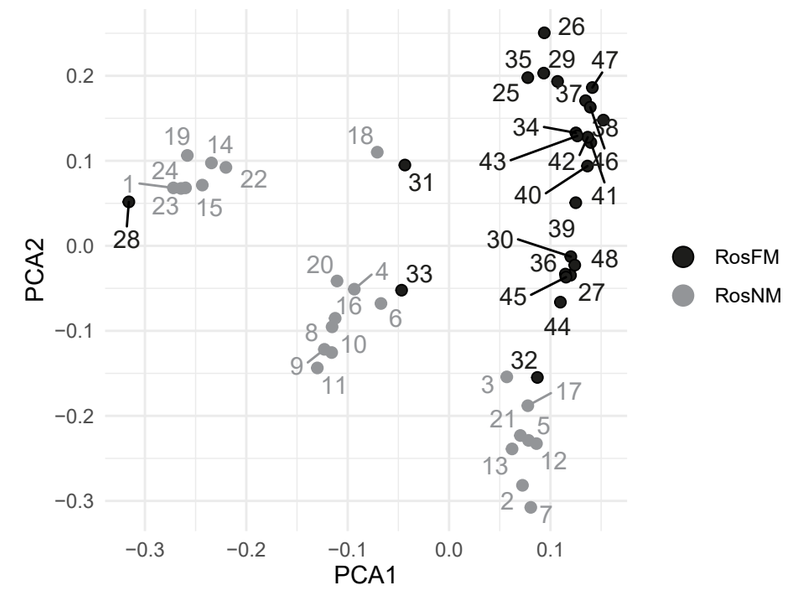
BIN 4800000_5300000



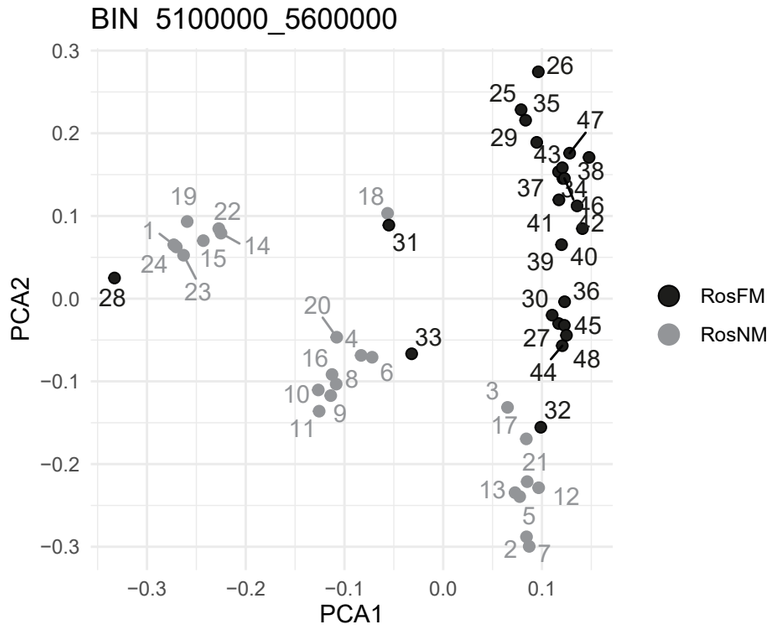
BIN 4900000_5400000



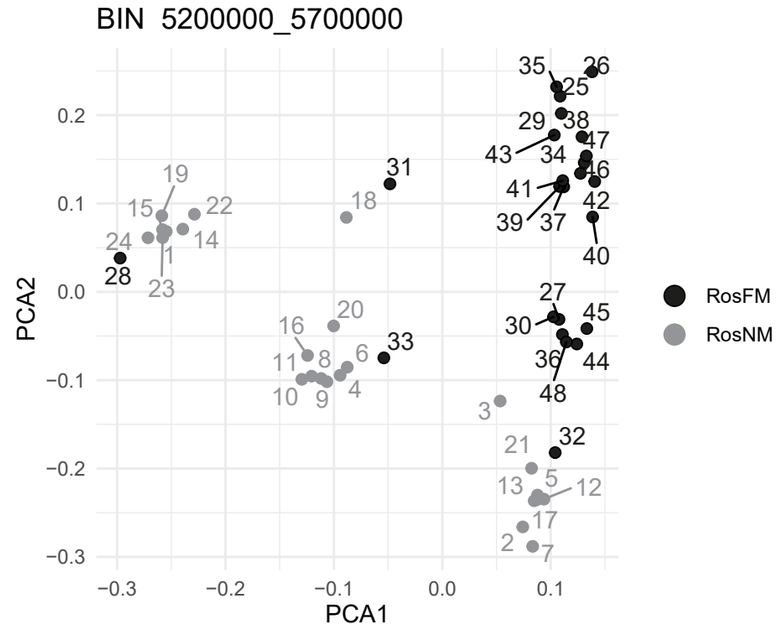
BIN 5000000_5500000



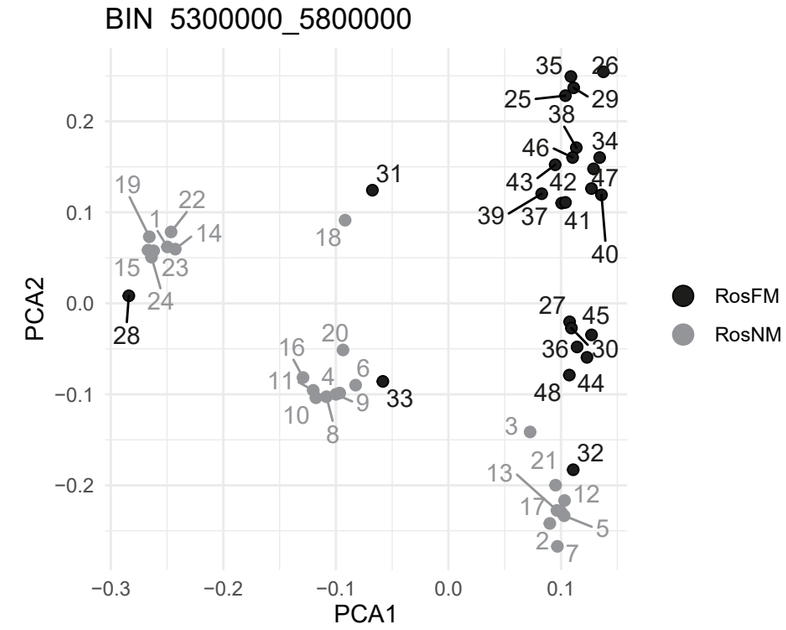
BIN 5100000_5600000



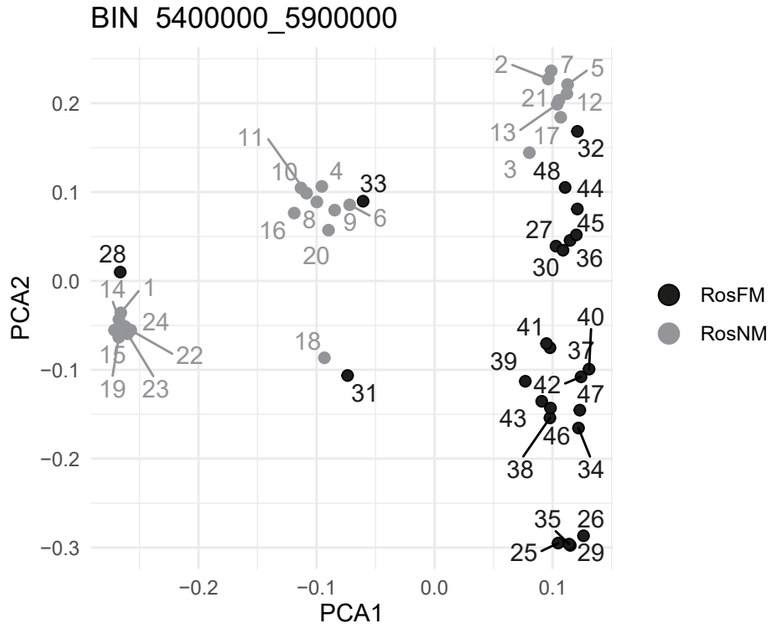
BIN 5200000_5700000



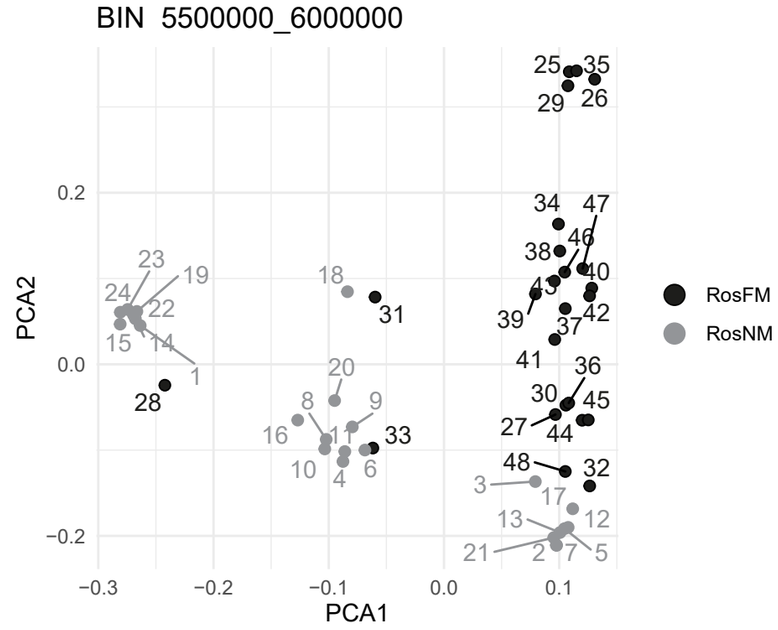
BIN 5300000_5800000



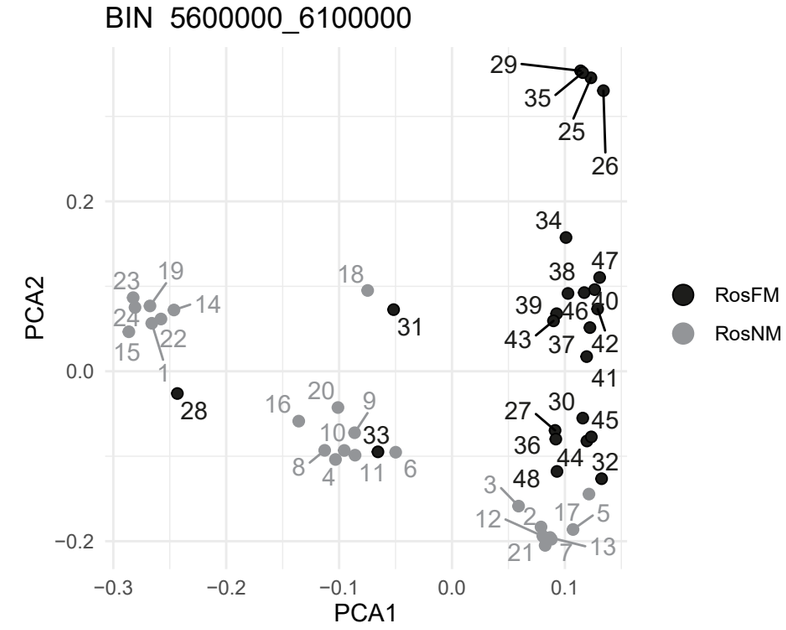
BIN 5400000_5900000



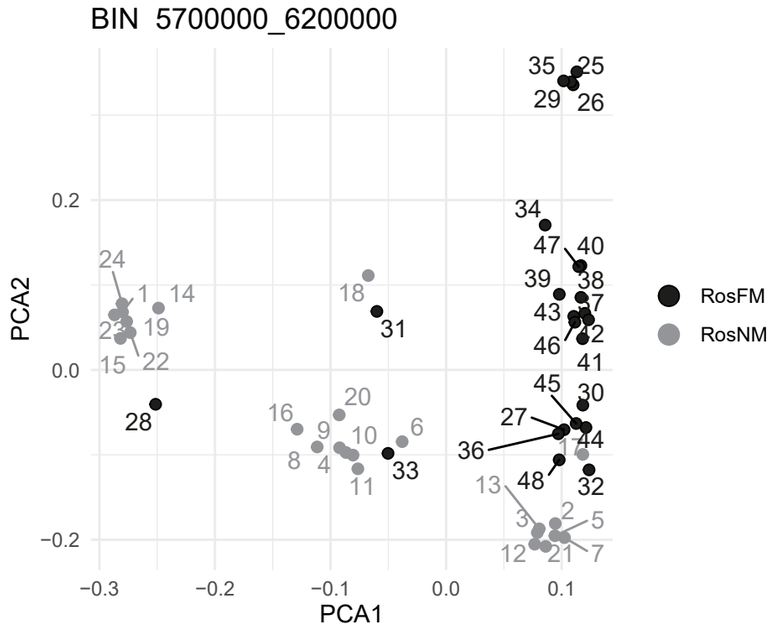
BIN 5500000_6000000



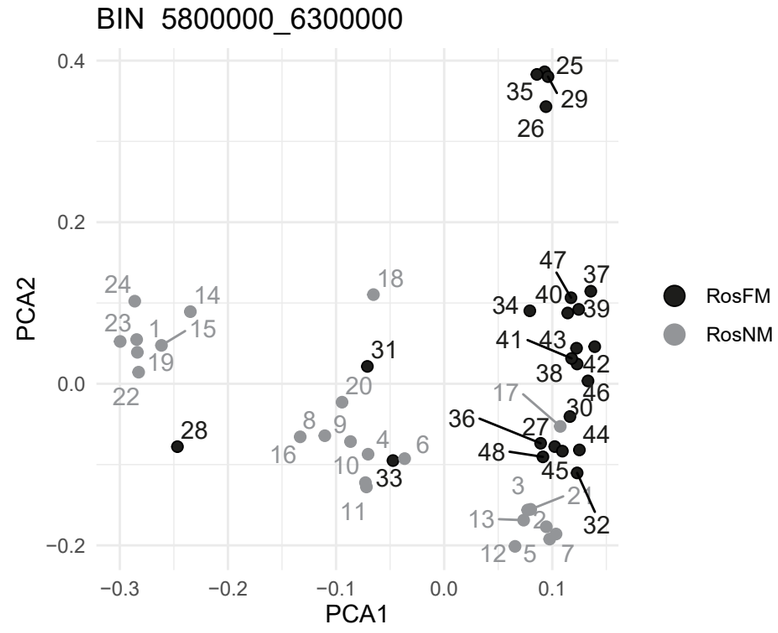
BIN 5600000_6100000



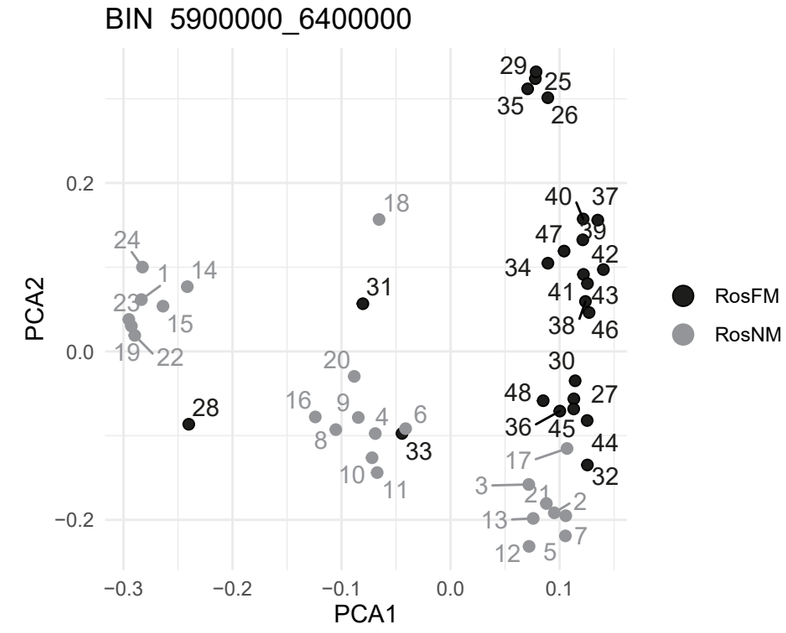
BIN 5700000_6200000



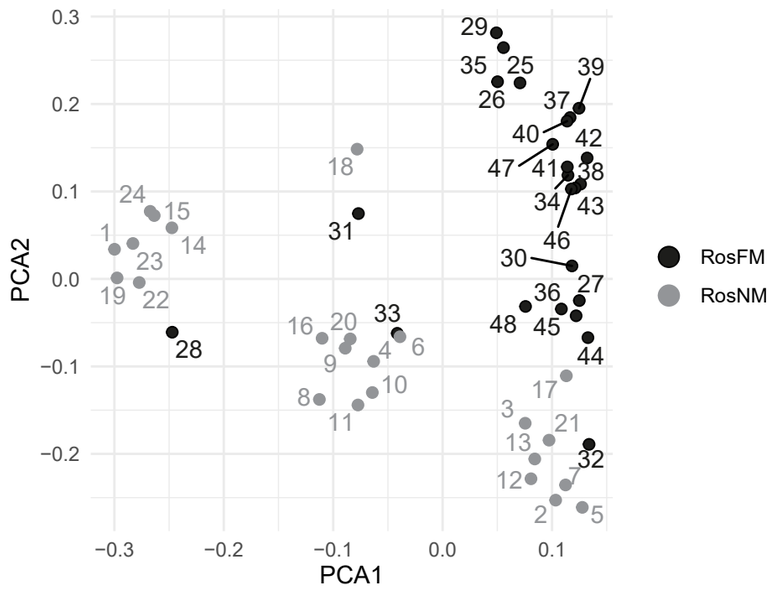
BIN 5800000_6300000



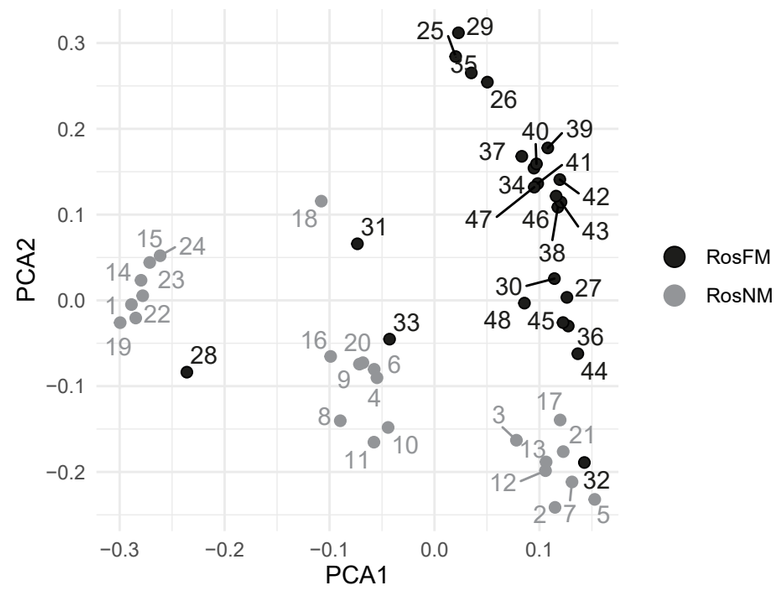
BIN 5900000_6400000



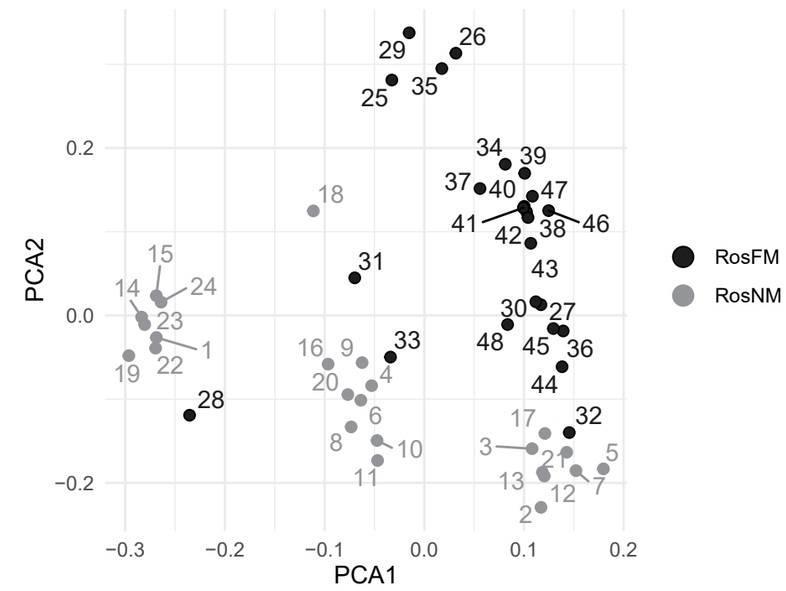
BIN 6000000_6500000



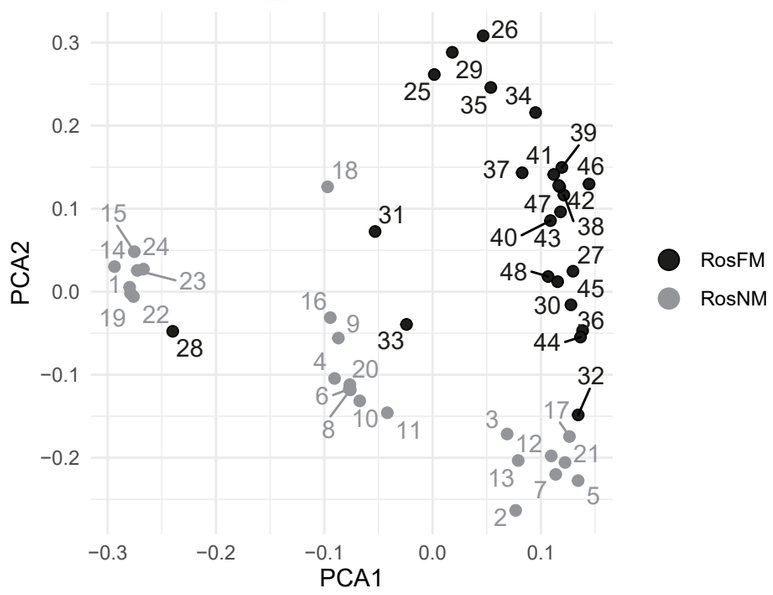
BIN 6100000_6600000



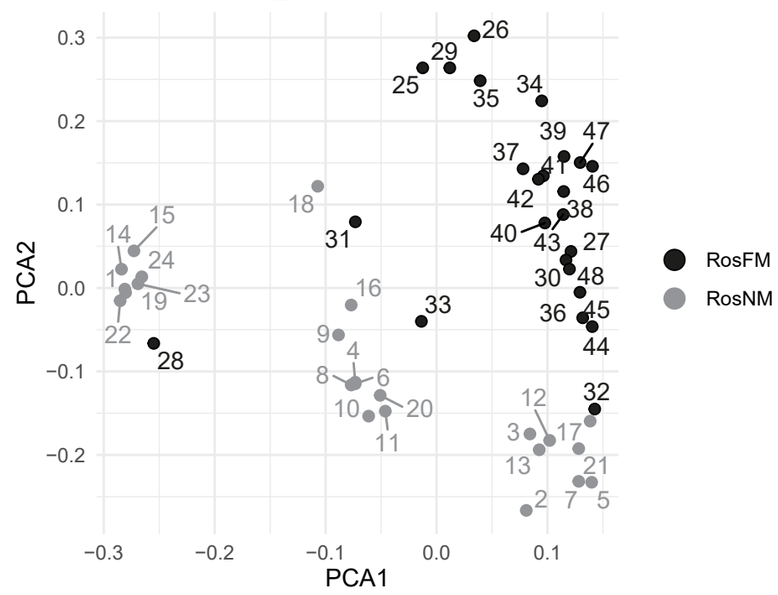
BIN 6200000_6700000



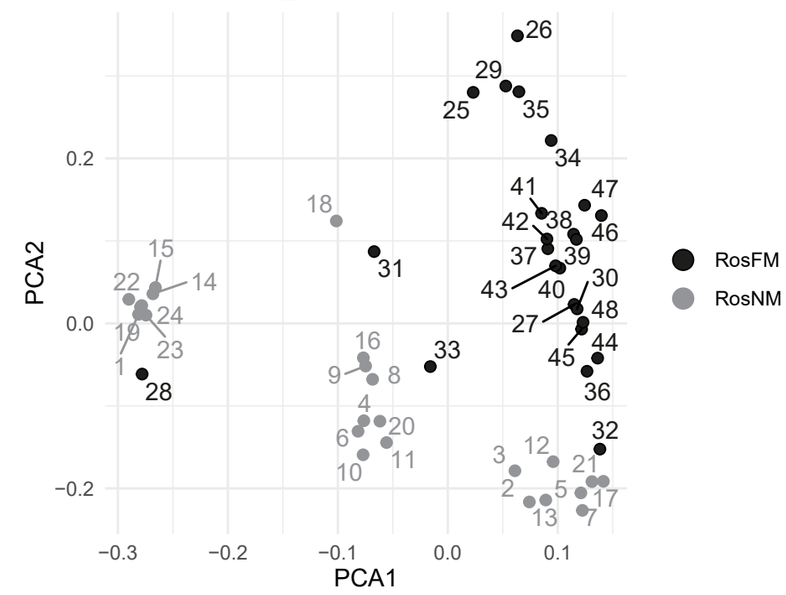
BIN 6300000_6800000



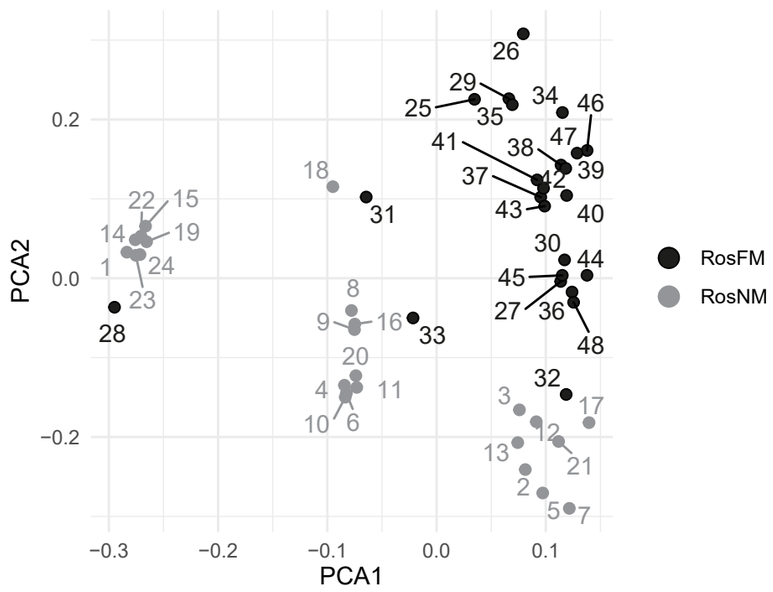
BIN 6400000_6900000



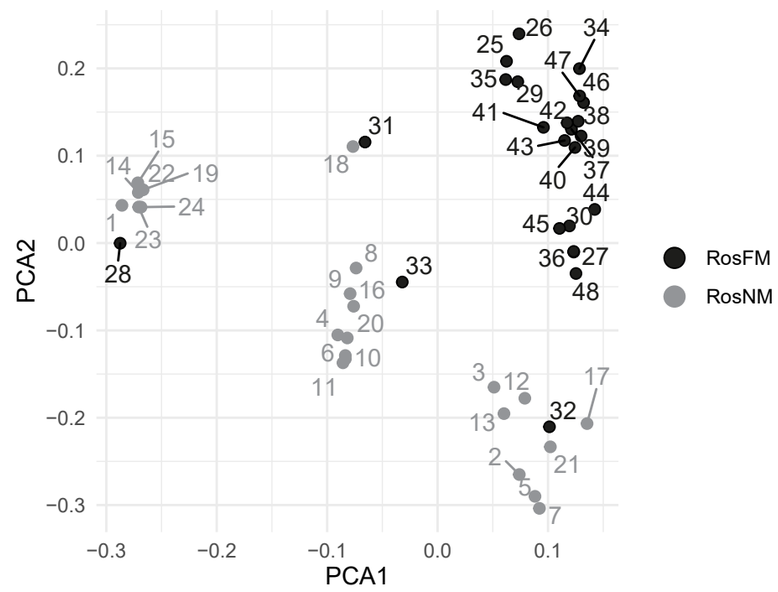
BIN 6500000_7000000



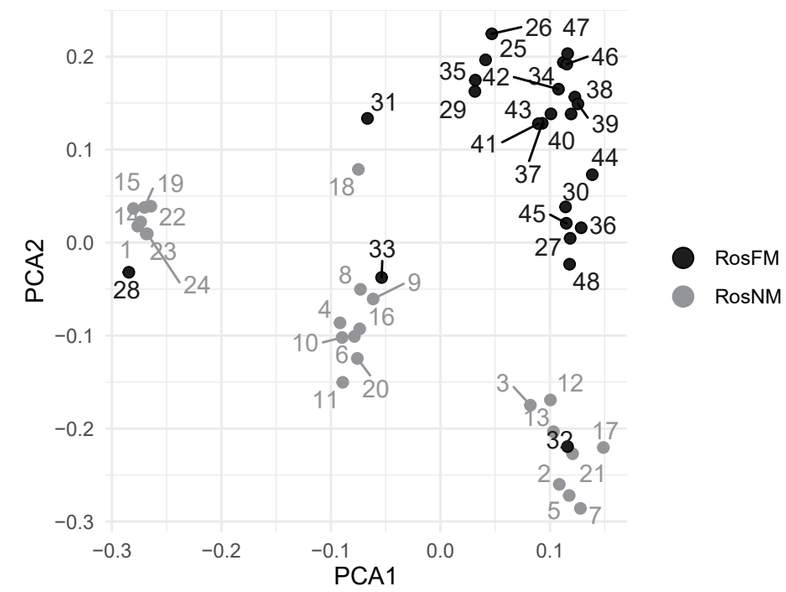
BIN 6600000_7100000



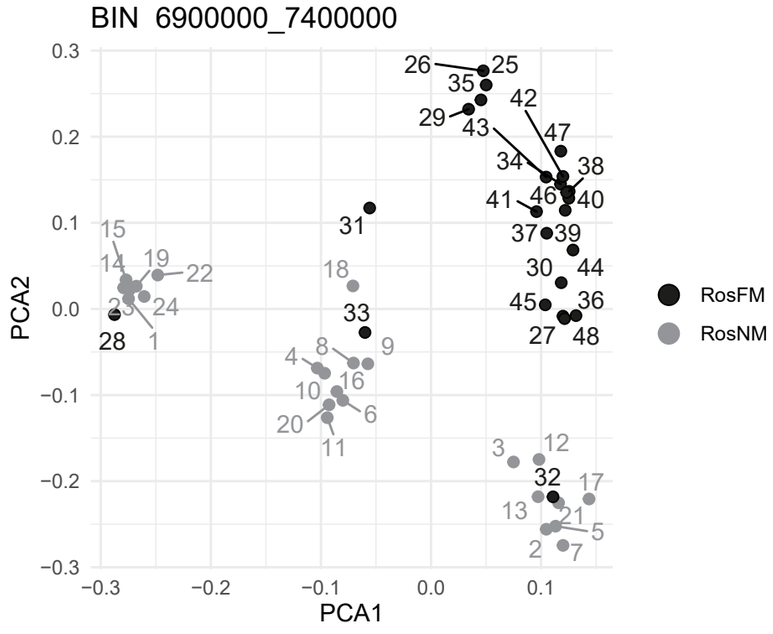
BIN 6700000_7200000



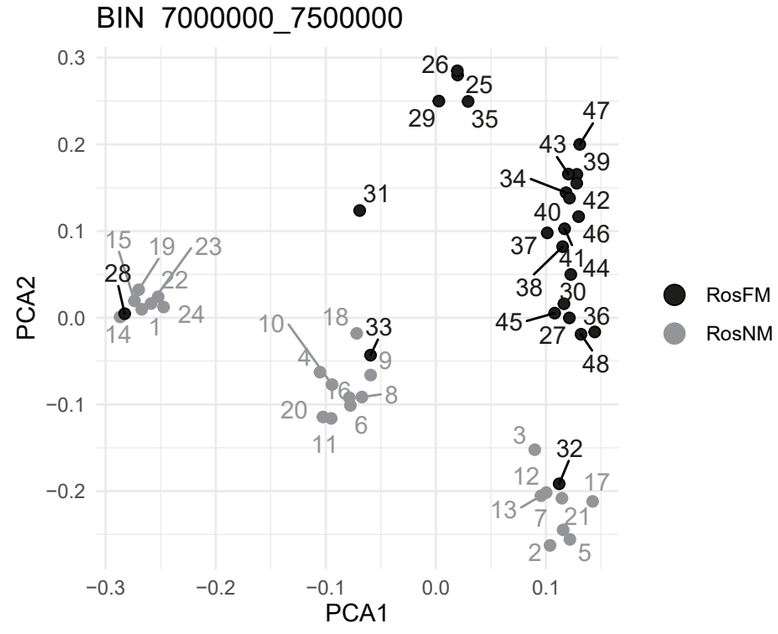
BIN 6800000_7300000



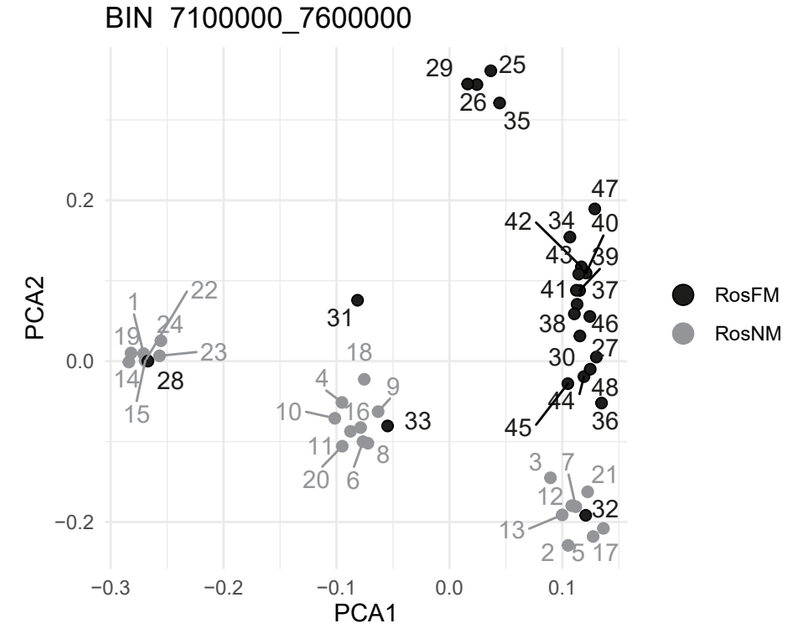
BIN 6900000_7400000



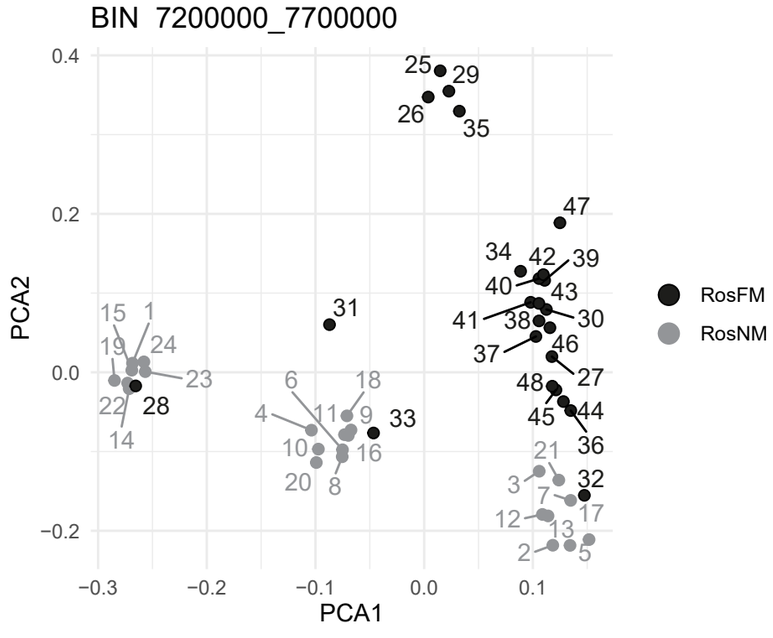
BIN 7000000_7500000



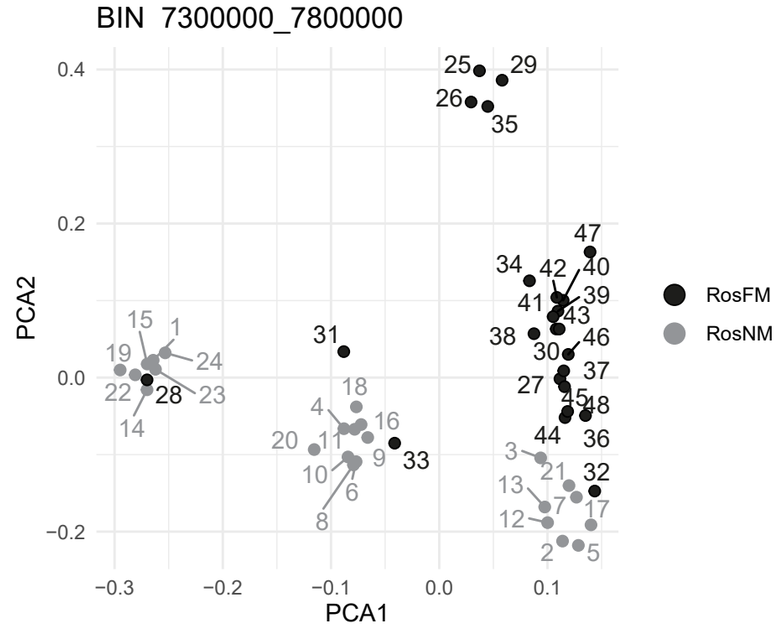
BIN 7100000_7600000



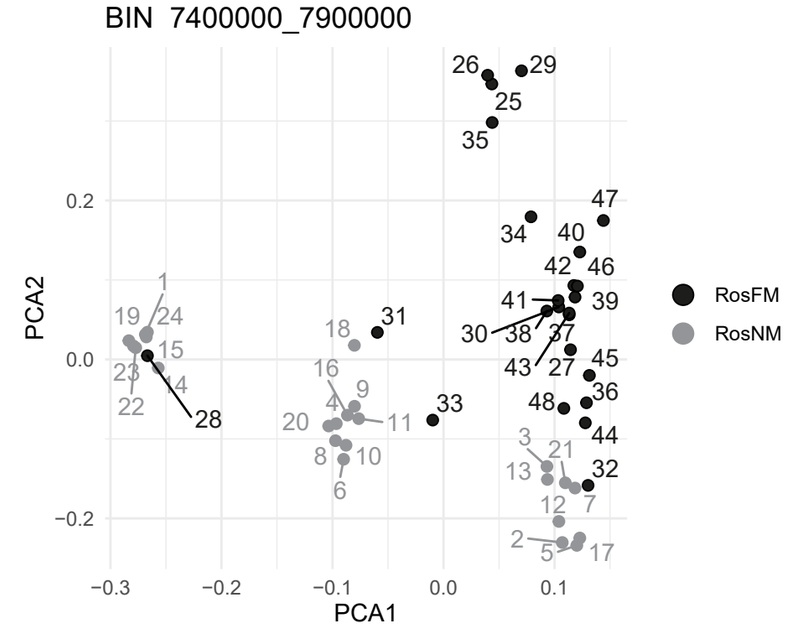
BIN 7200000_7700000



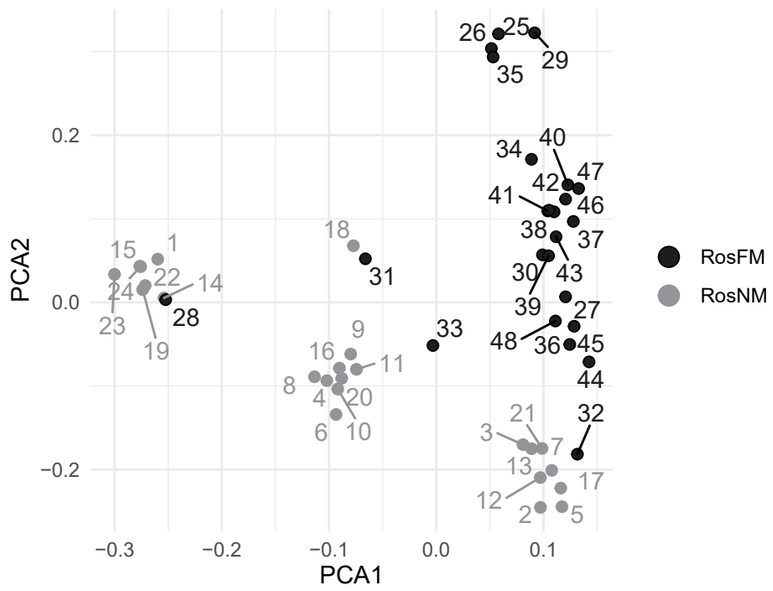
BIN 7300000_7800000



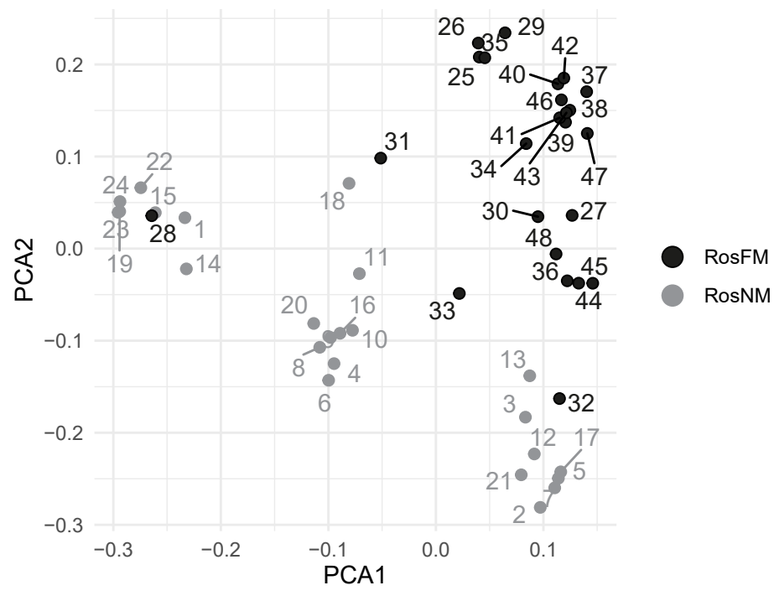
BIN 7400000_7900000



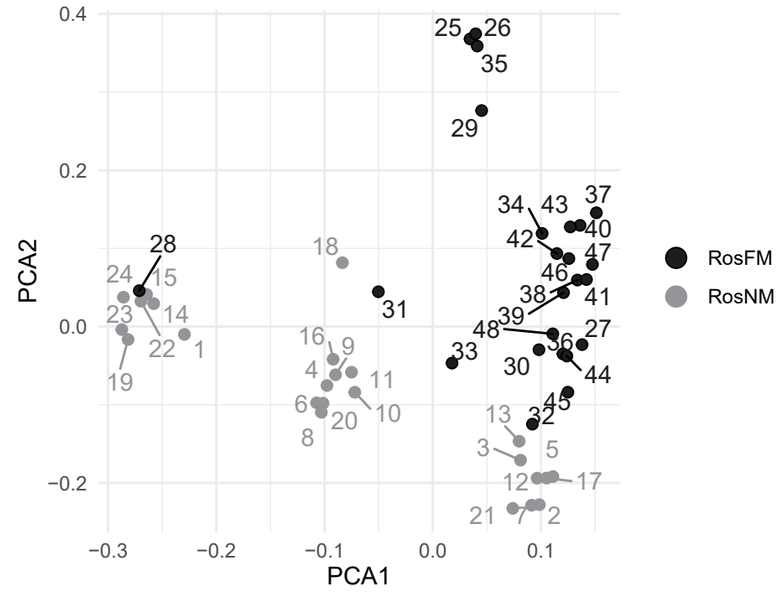
BIN 7500000_8000000



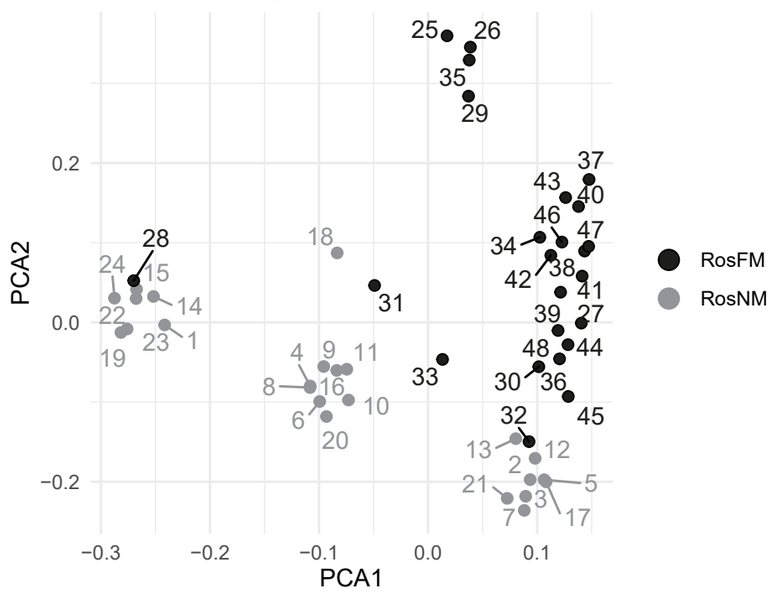
BIN 7600000_8100000



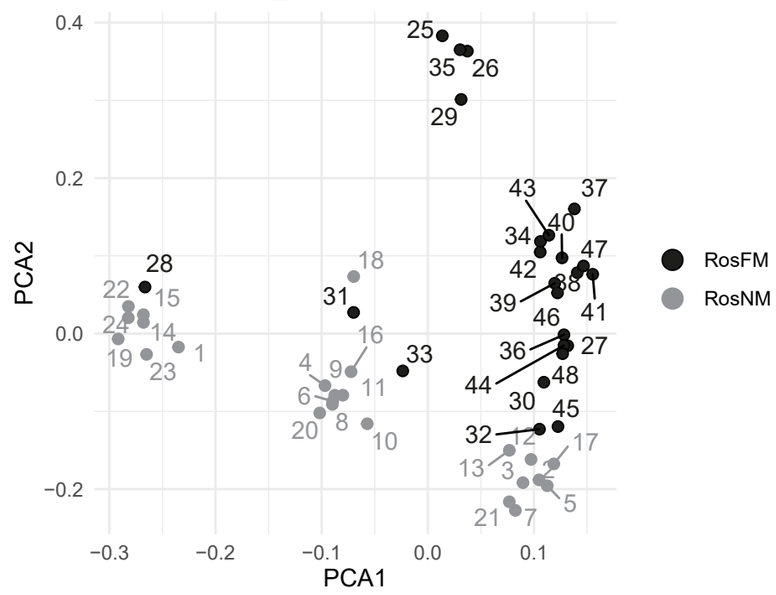
BIN 7700000_8200000



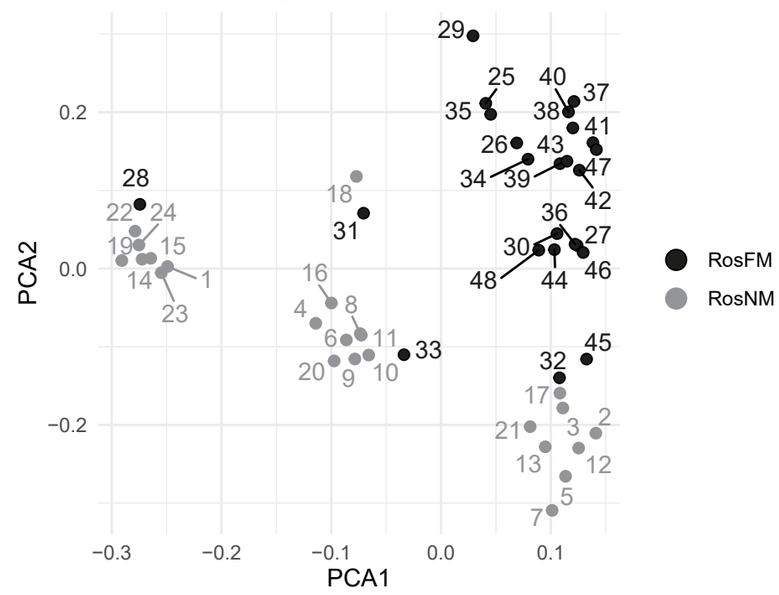
BIN 7800000_8300000



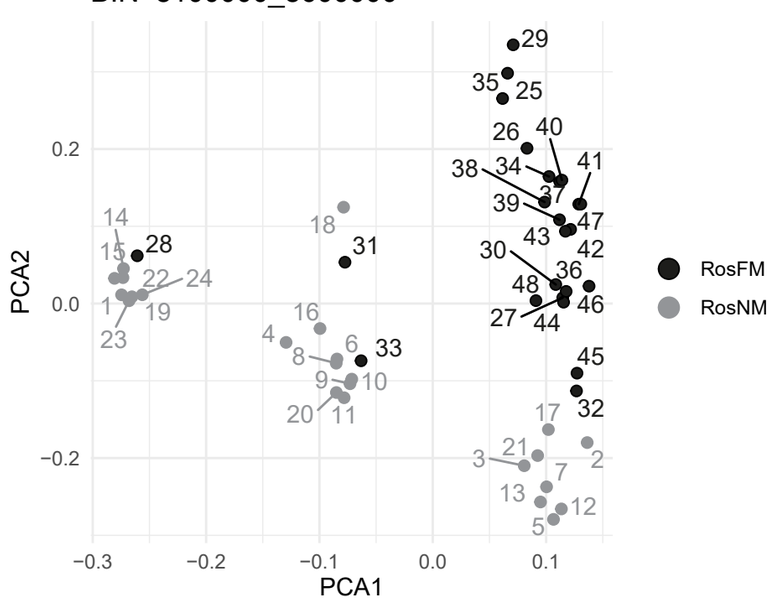
BIN 7900000_8400000



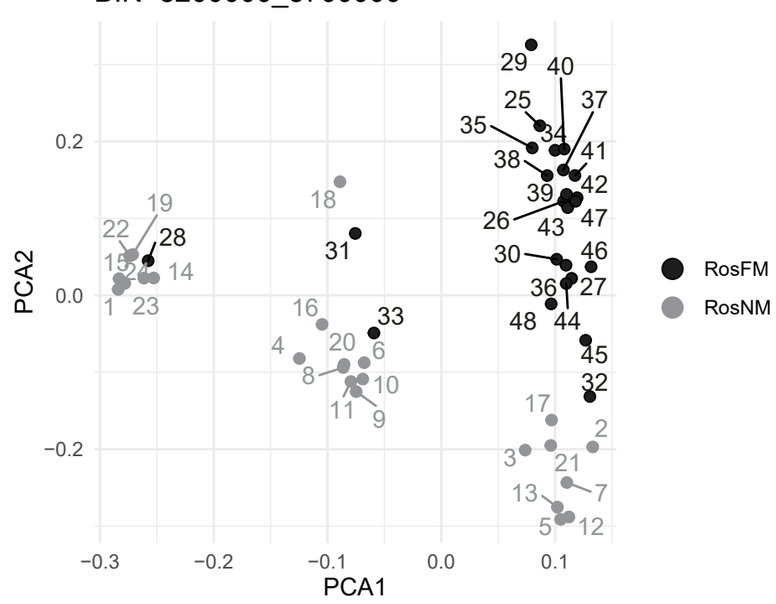
BIN 8000000_8500000



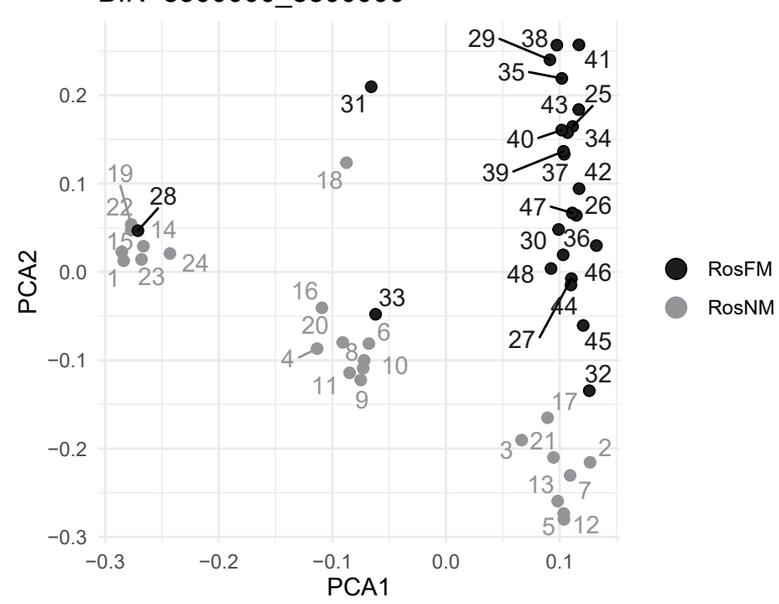
BIN 8100000_8600000



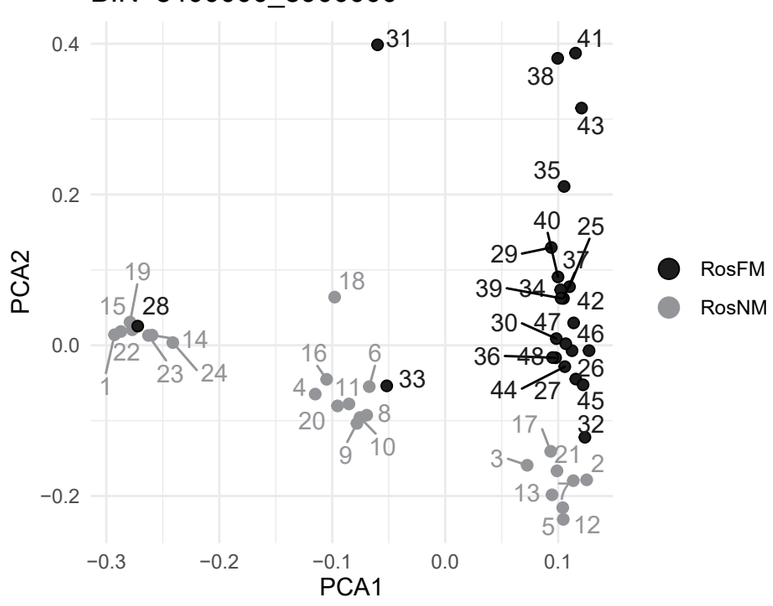
BIN 8200000_8700000



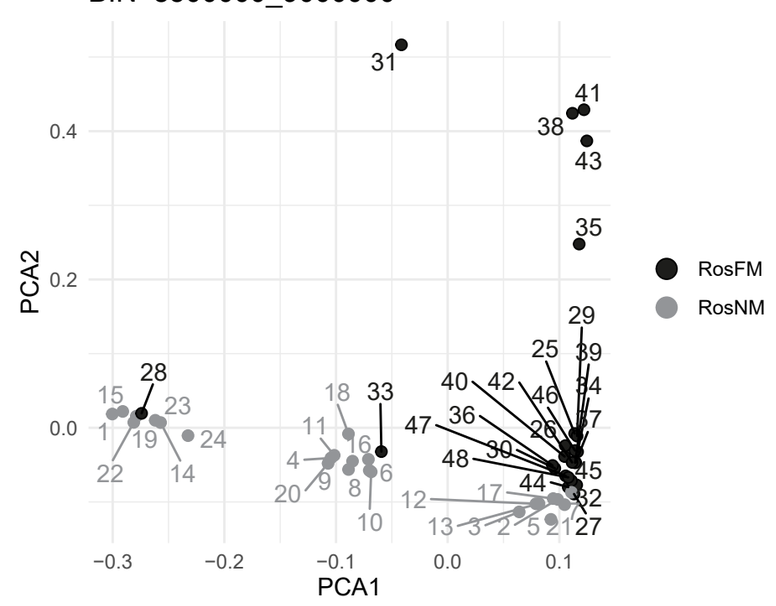
BIN 8300000_8800000



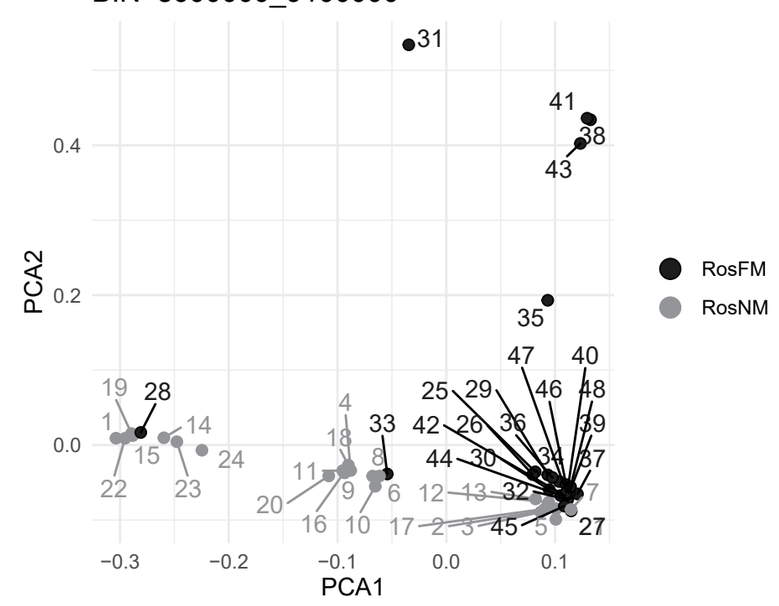
BIN 8400000_8900000



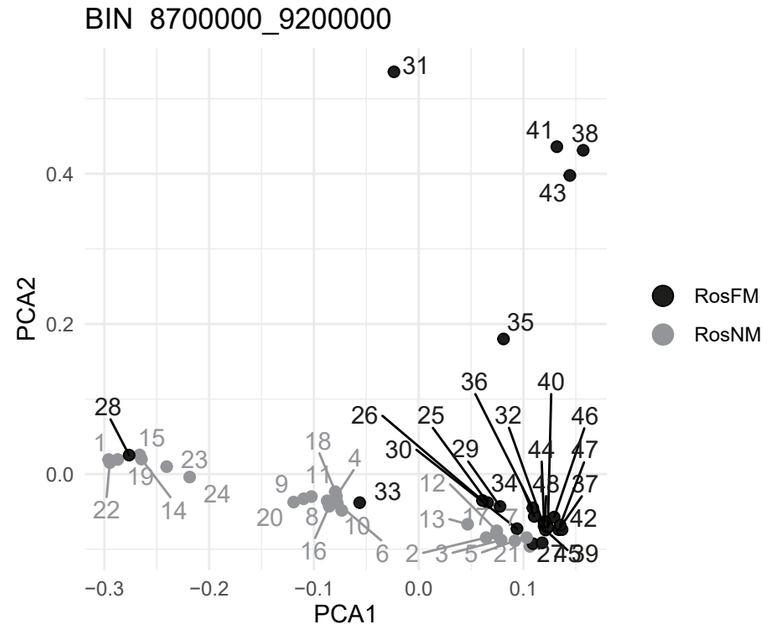
BIN 8500000_9000000



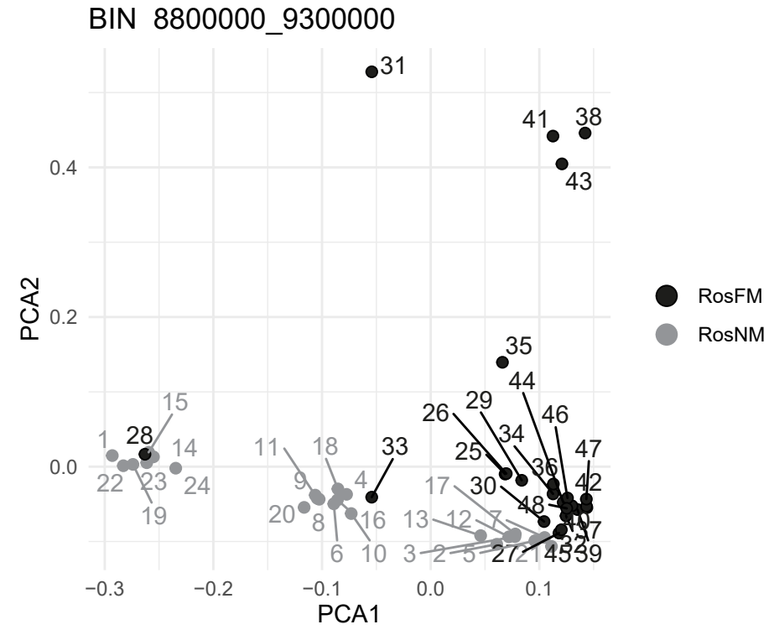
BIN 8600000_9100000



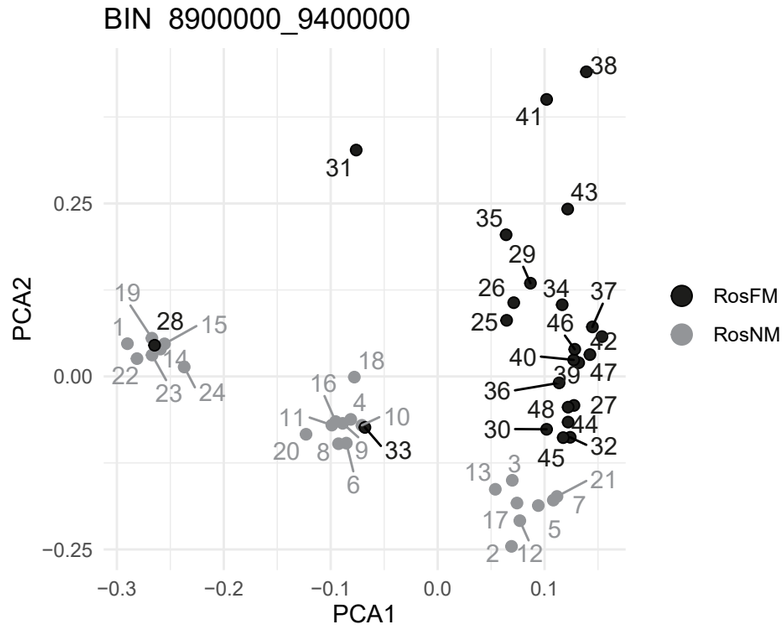
BIN 8700000_9200000



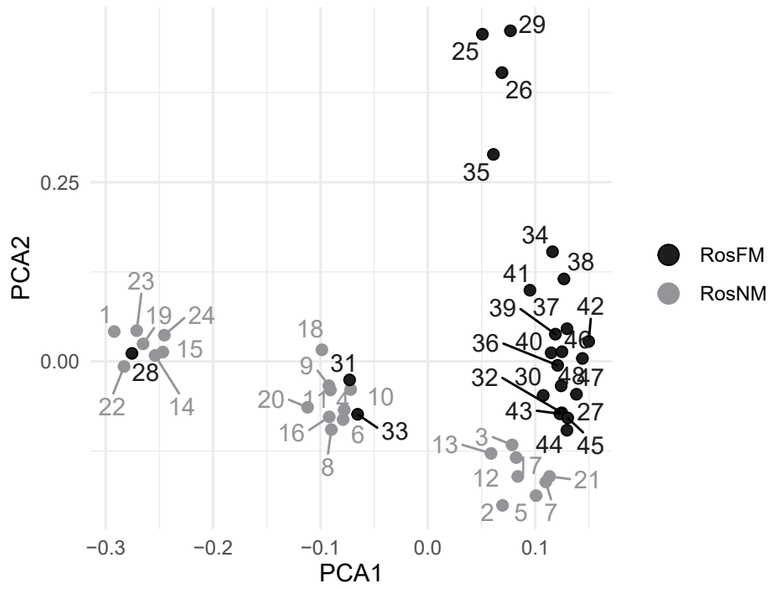
BIN 8800000_9300000



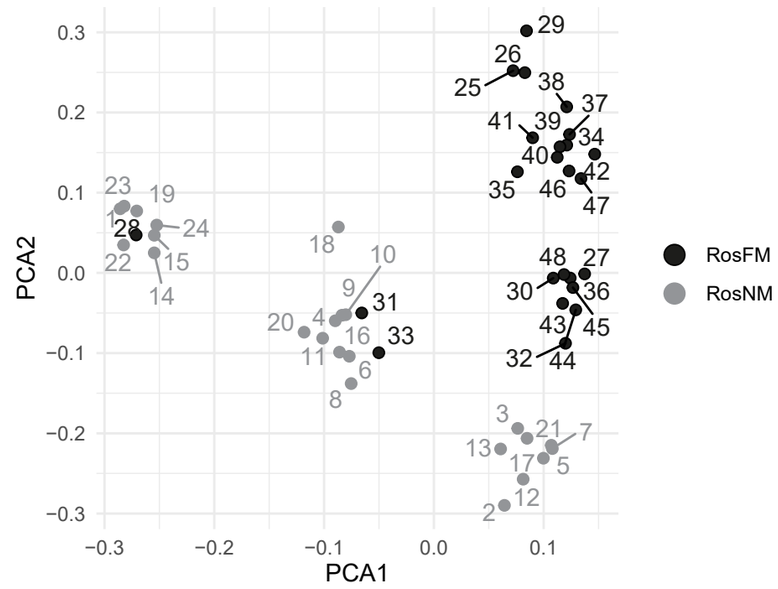
BIN 8900000_9400000



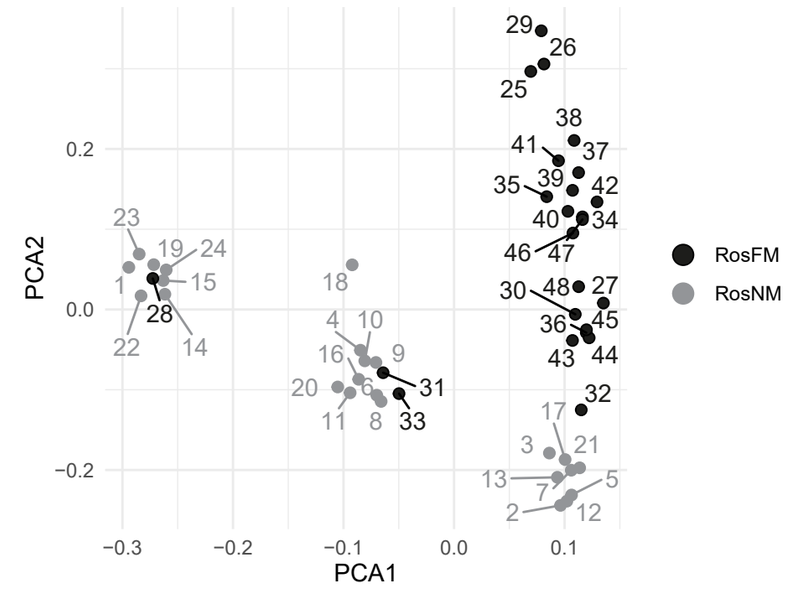
BIN 9000000_9500000



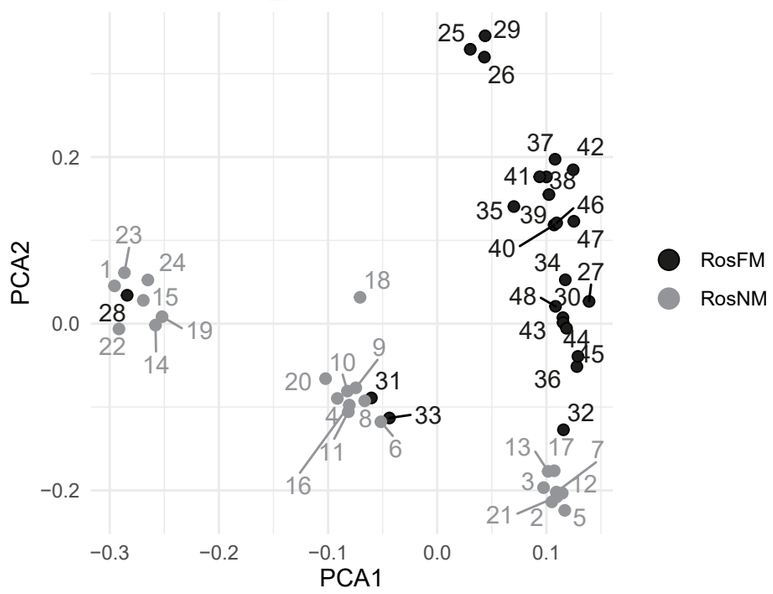
BIN 9100000_9600000



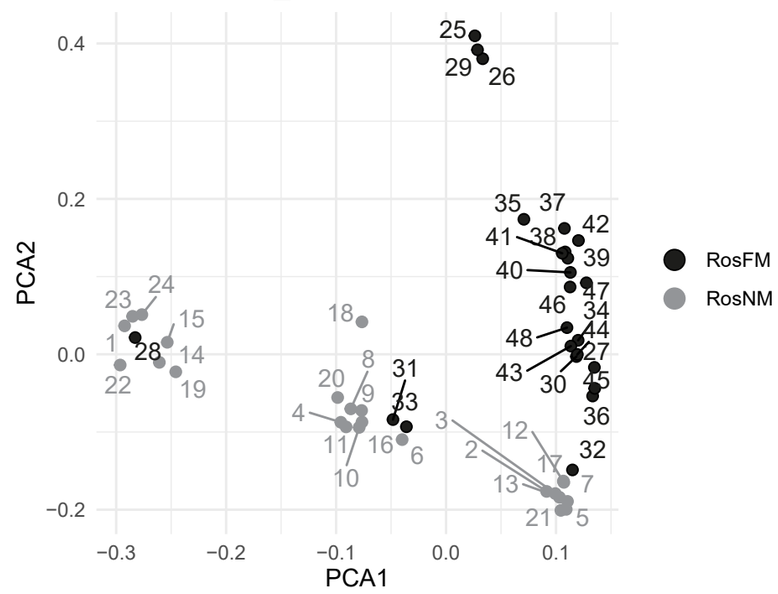
BIN 9200000_9700000



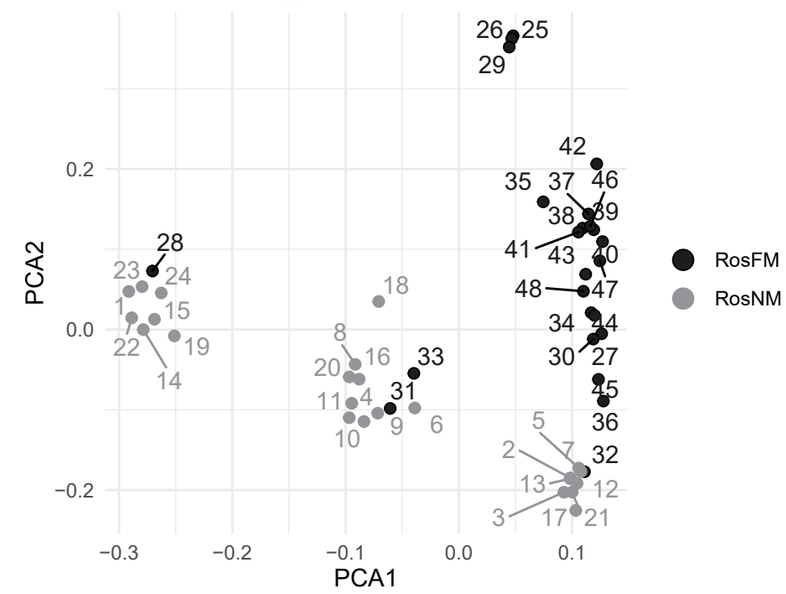
BIN 9300000_9800000



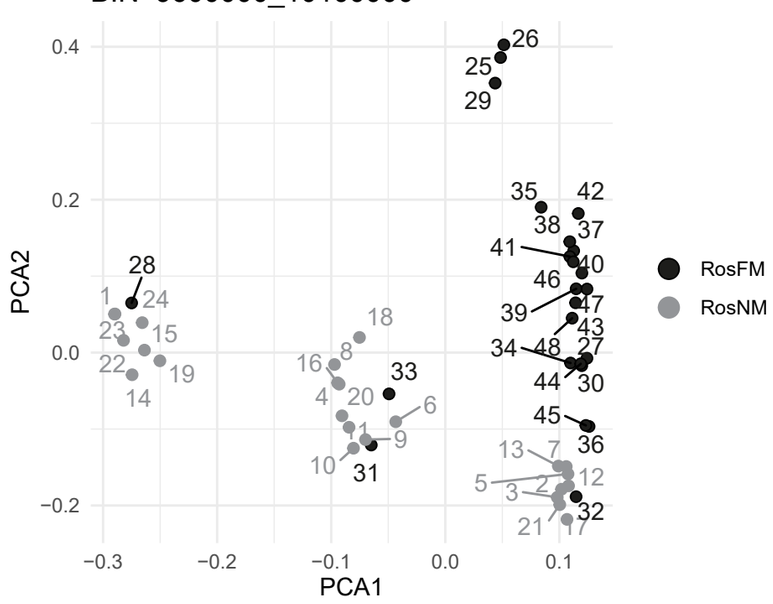
BIN 9400000_9900000



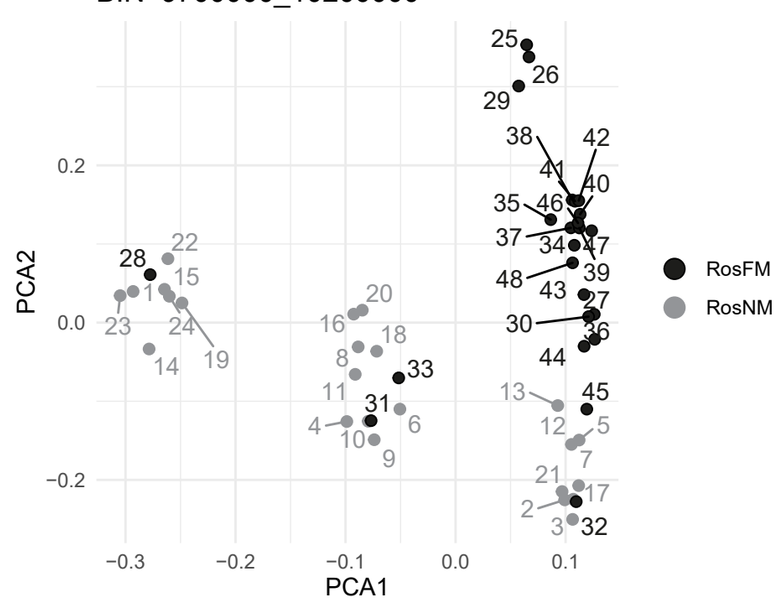
BIN 9500000_10000000



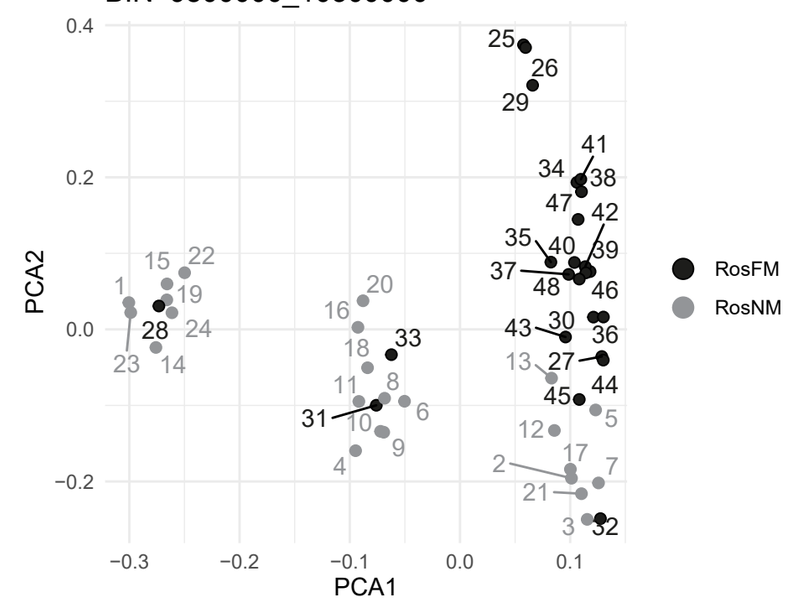
BIN 9600000_10100000



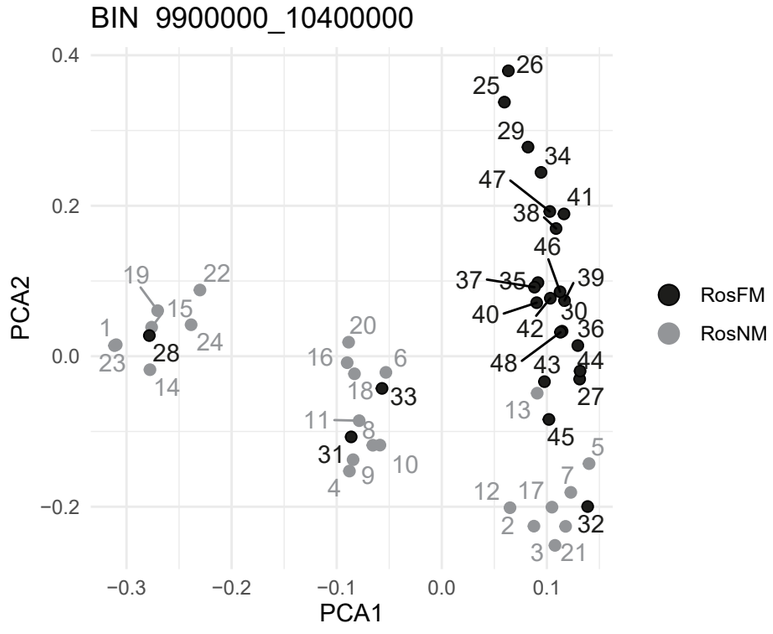
BIN 9700000_10200000



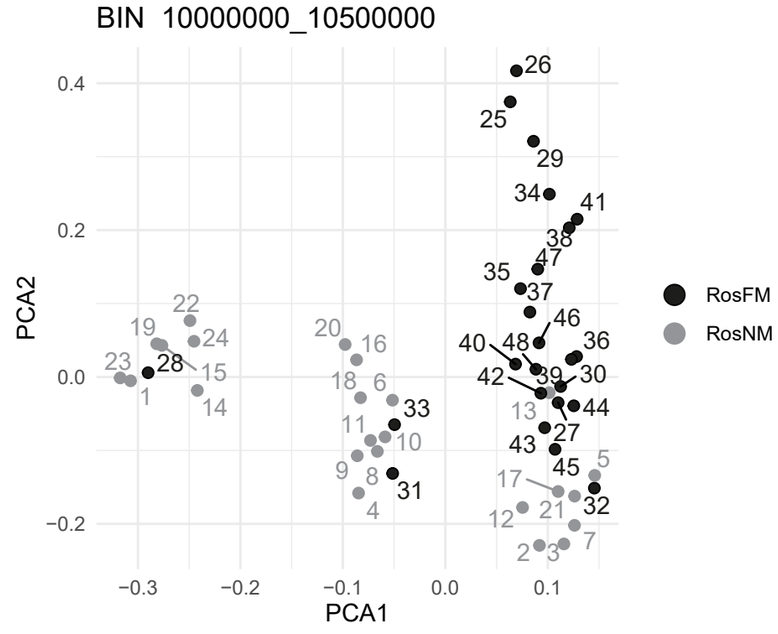
BIN 9800000_10300000



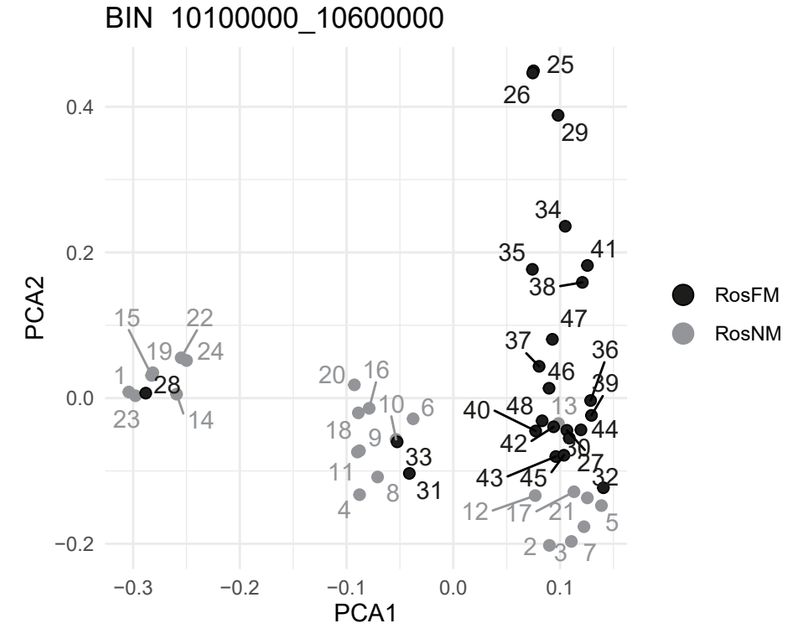
BIN 9900000_10400000



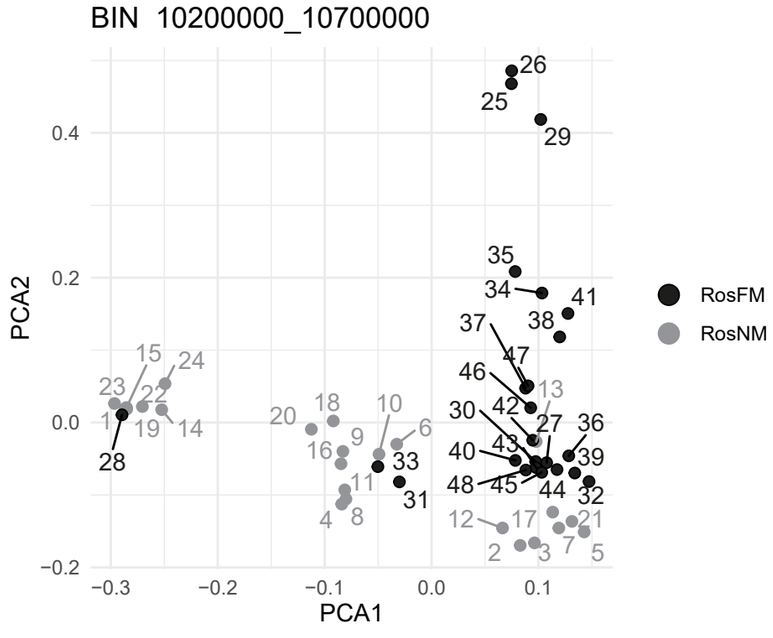
BIN 10000000_10500000



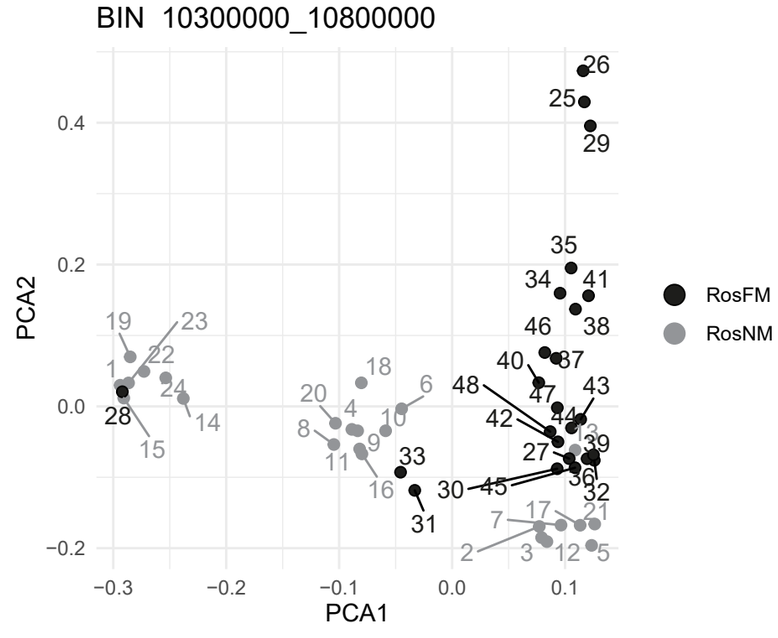
BIN 10100000_10600000



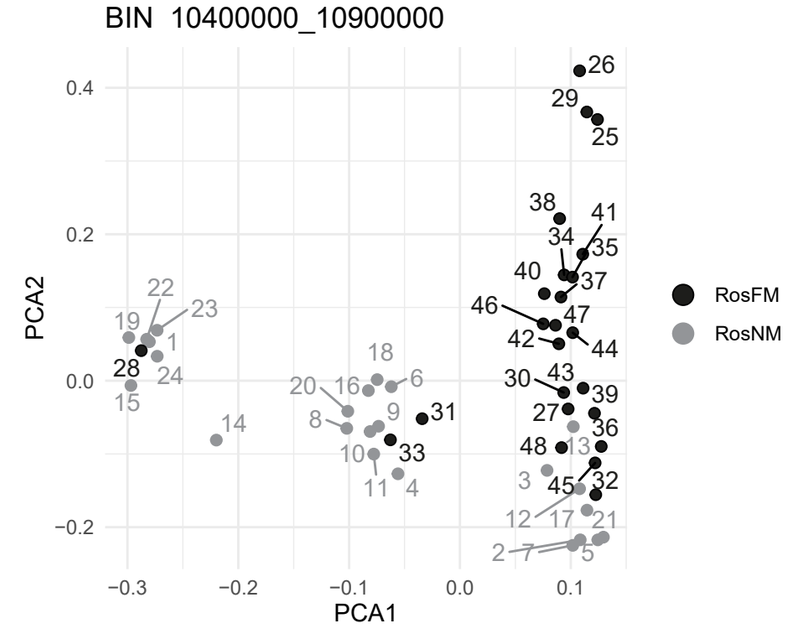
BIN 10200000_10700000



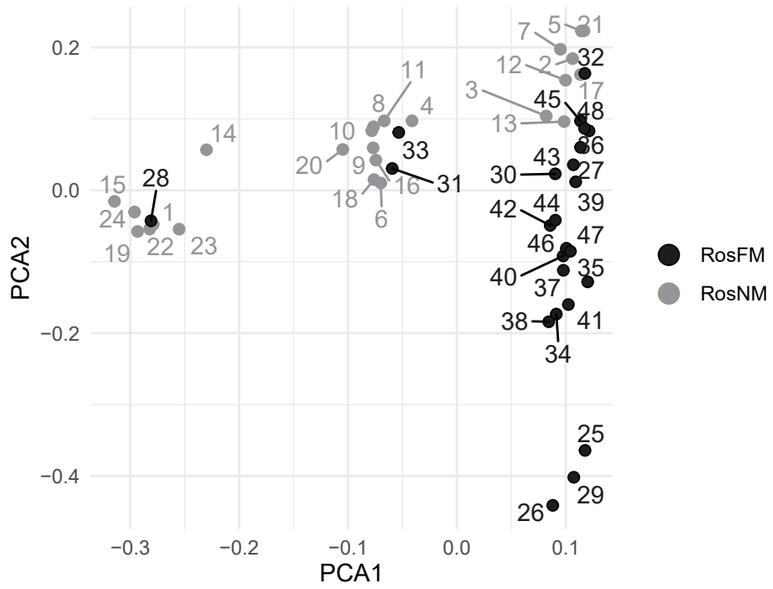
BIN 10300000_10800000



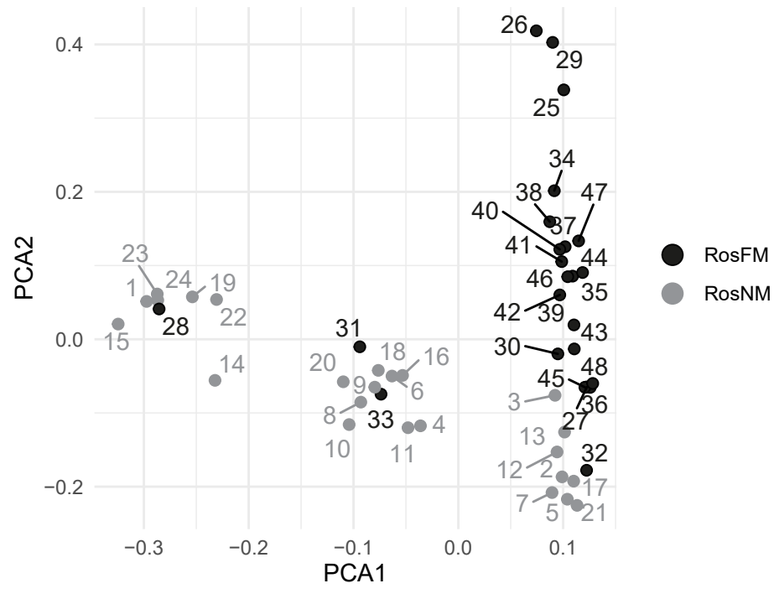
BIN 10400000_10900000



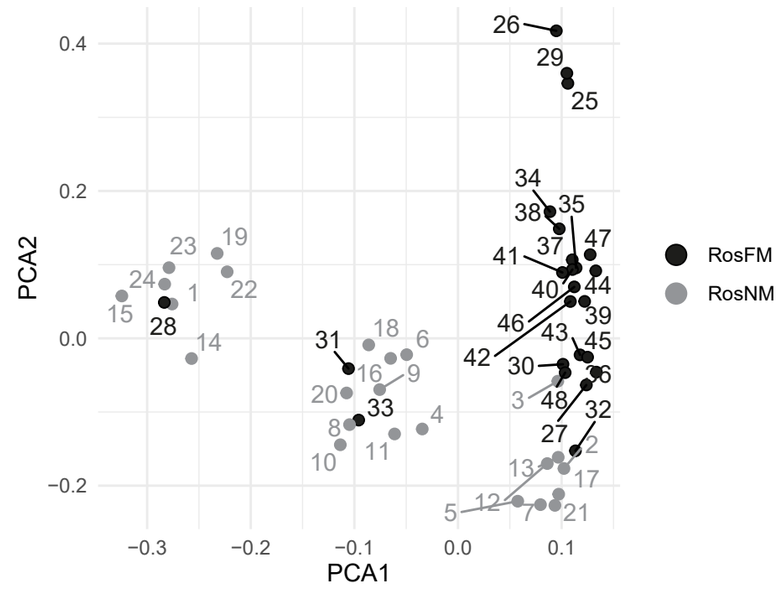
BIN 10500000_11000000



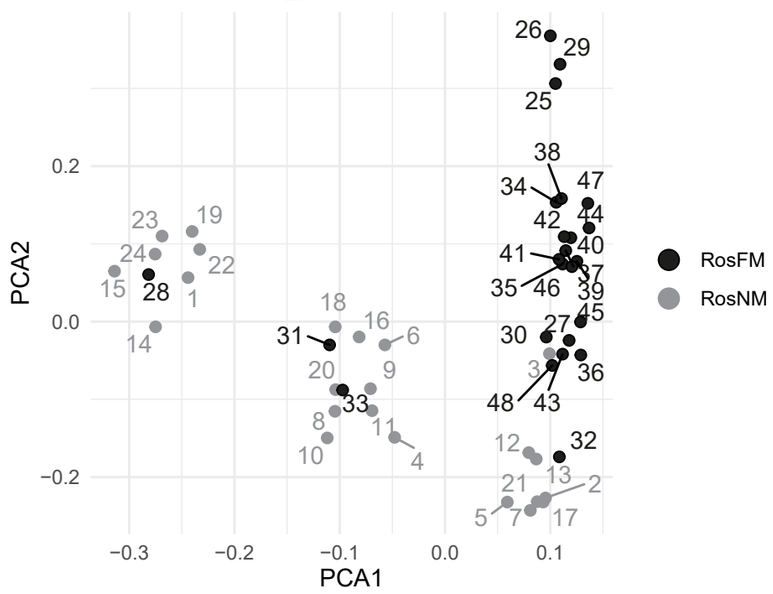
BIN 10600000_11100000



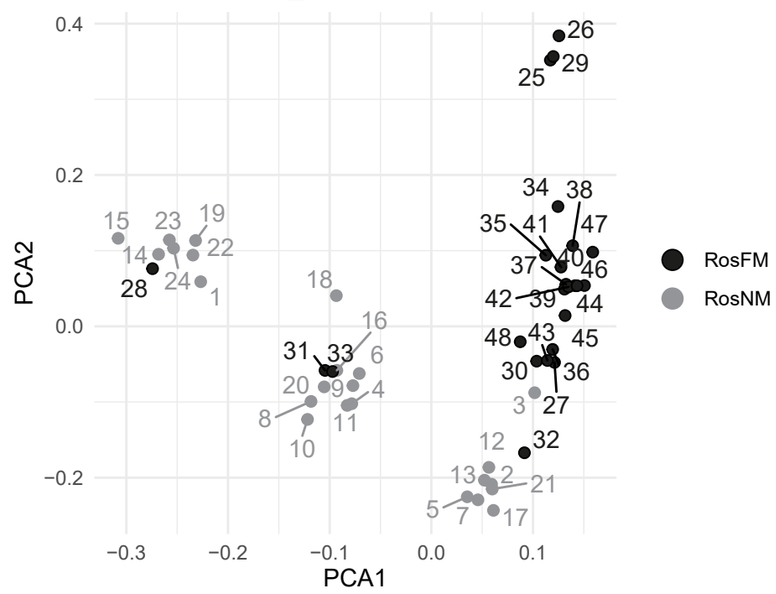
BIN 10700000_11200000



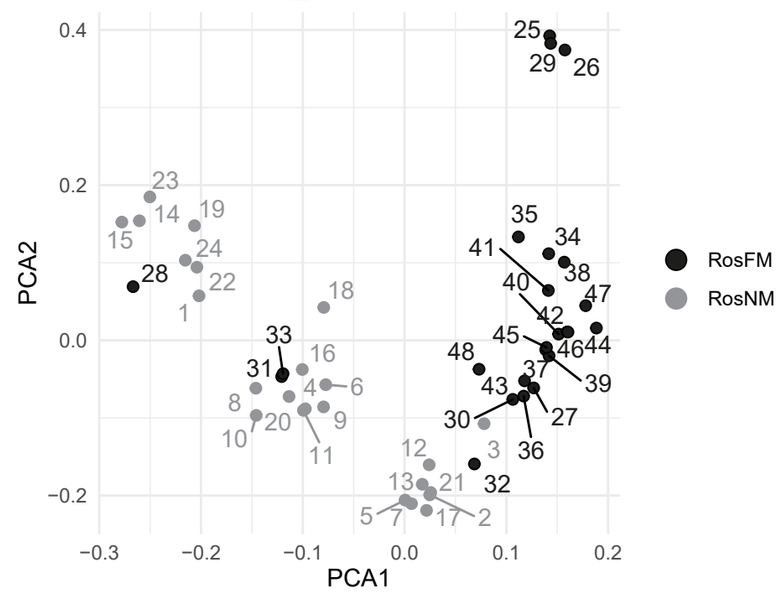
BIN 10800000_11300000



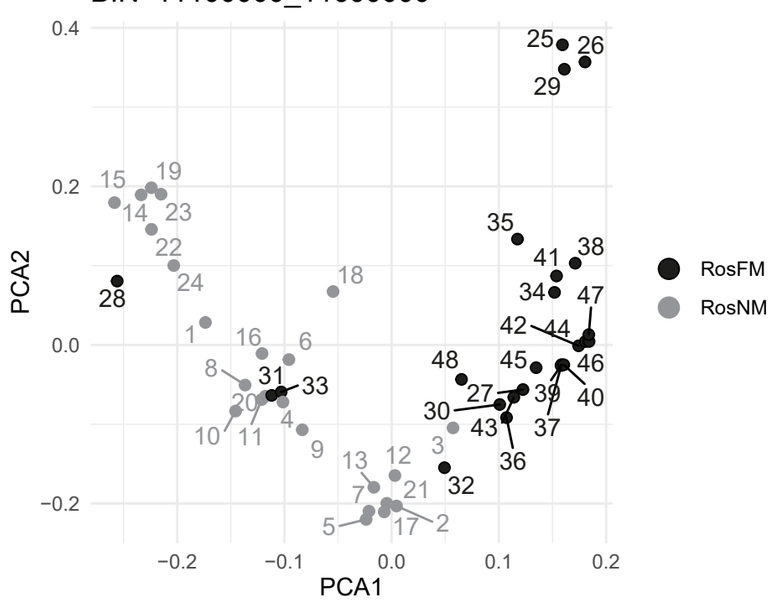
BIN 10900000_11400000



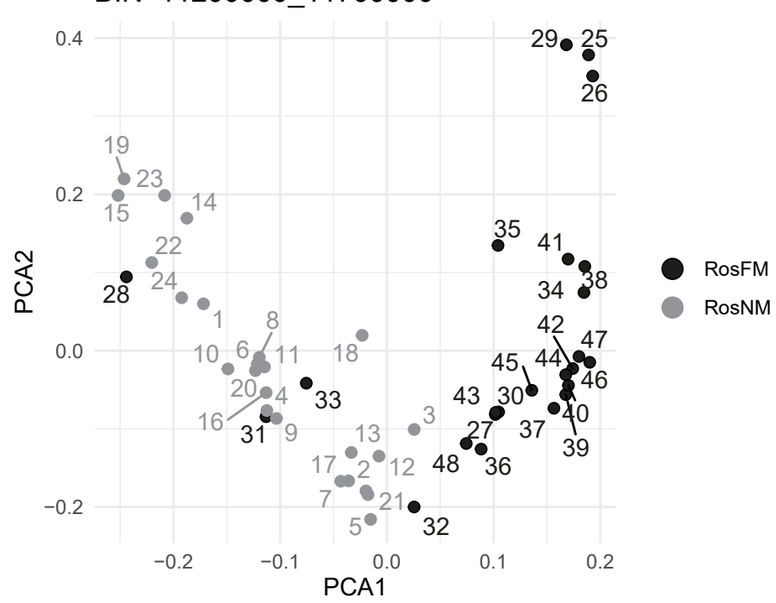
BIN 11000000_11500000



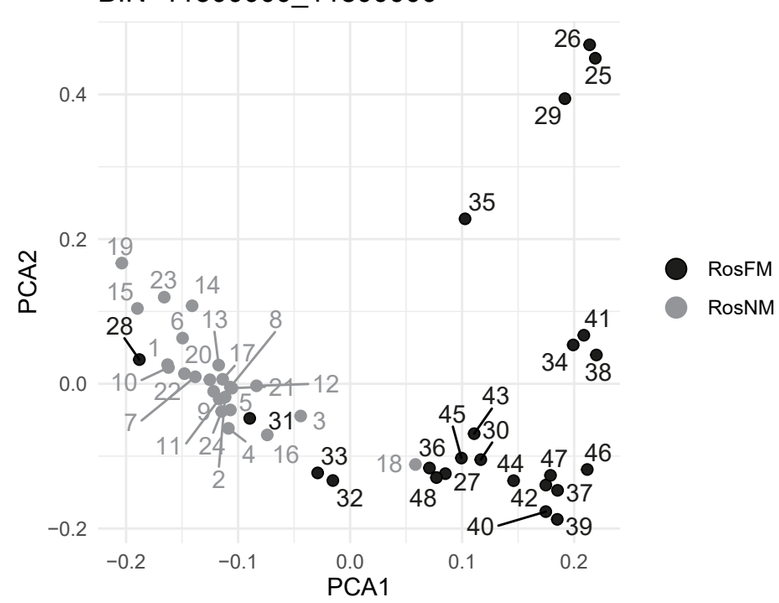
BIN 11100000_11600000



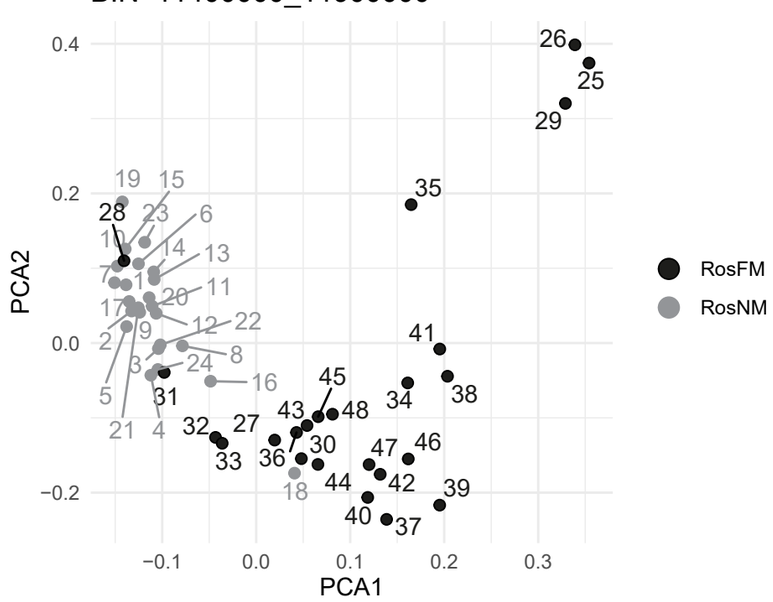
BIN 11200000_11700000



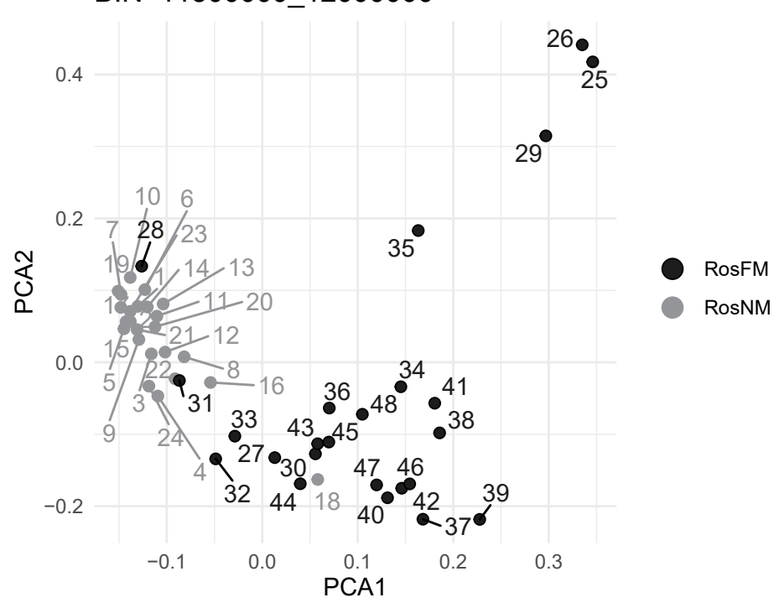
BIN 11300000_11800000



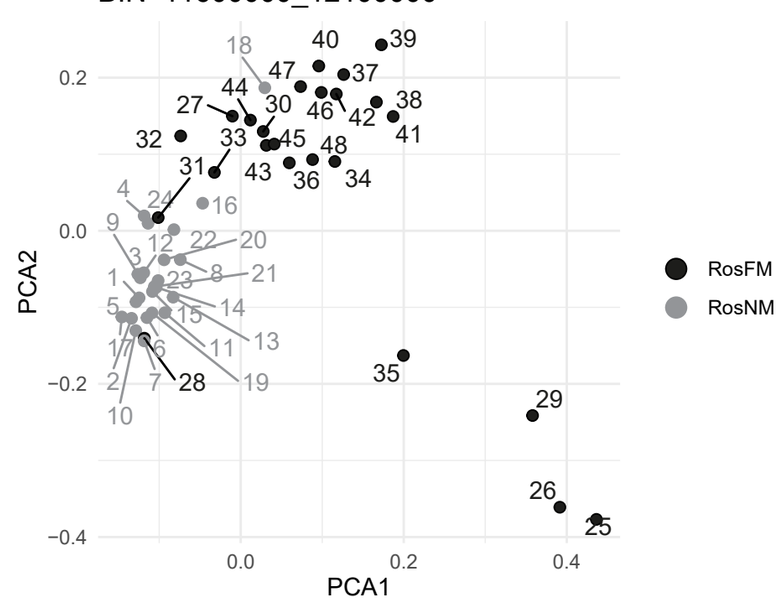
BIN 11400000_11900000



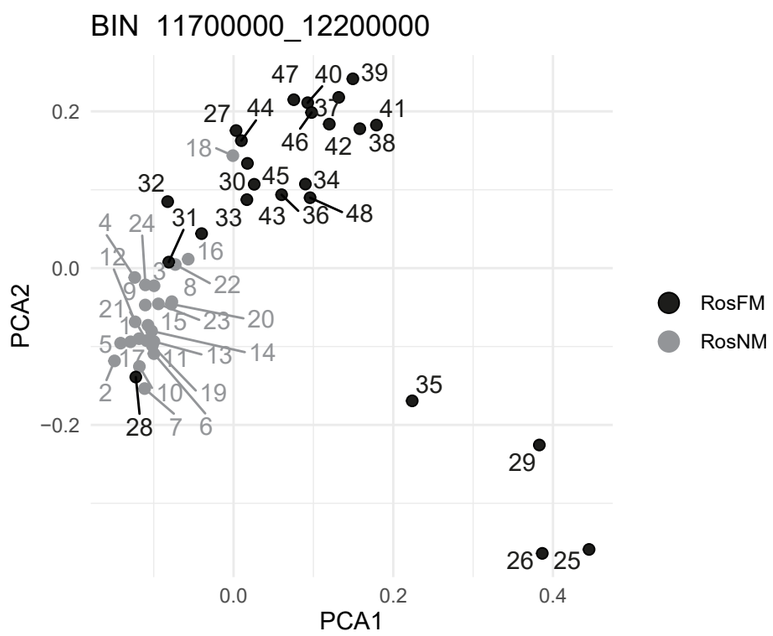
BIN 11500000_12000000



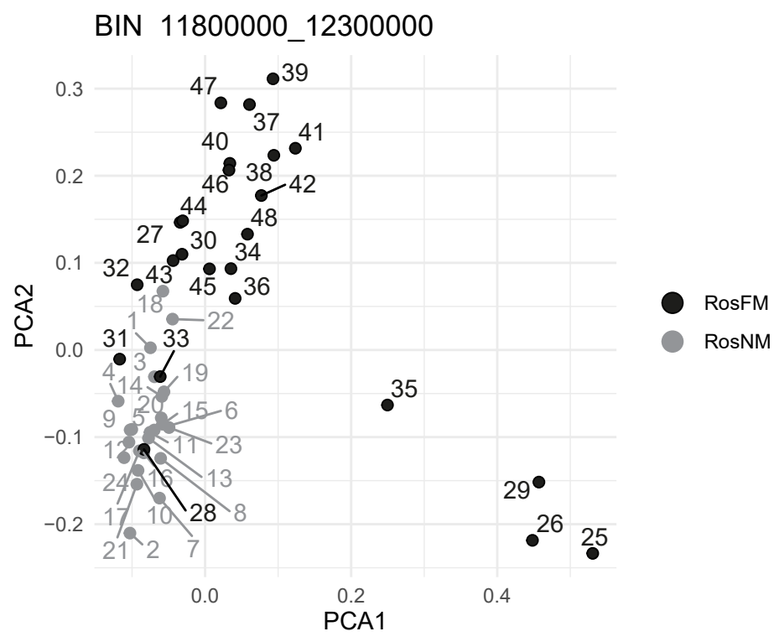
BIN 11600000_12100000



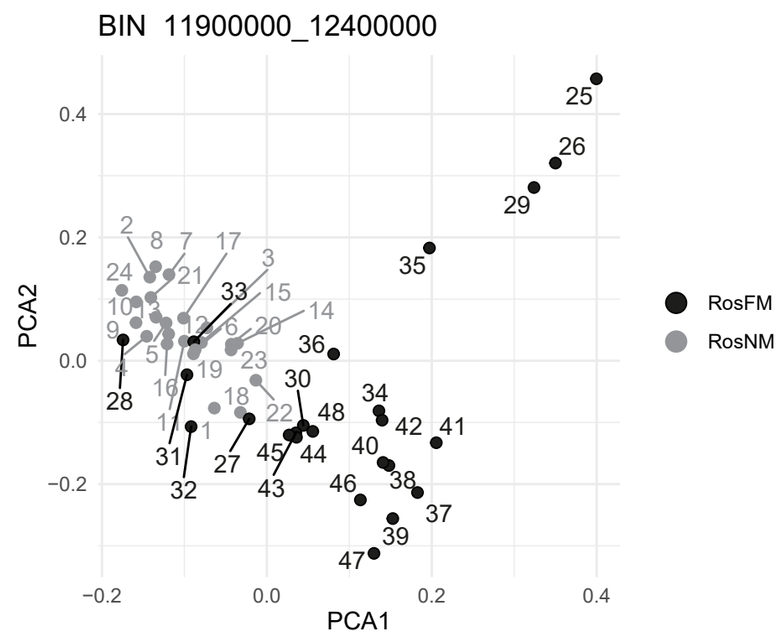
BIN 11700000_12200000



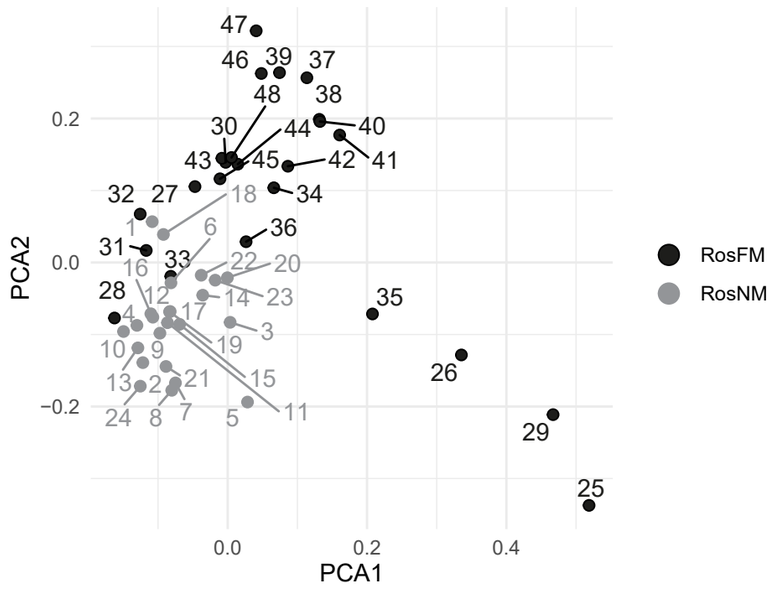
BIN 11800000_12300000



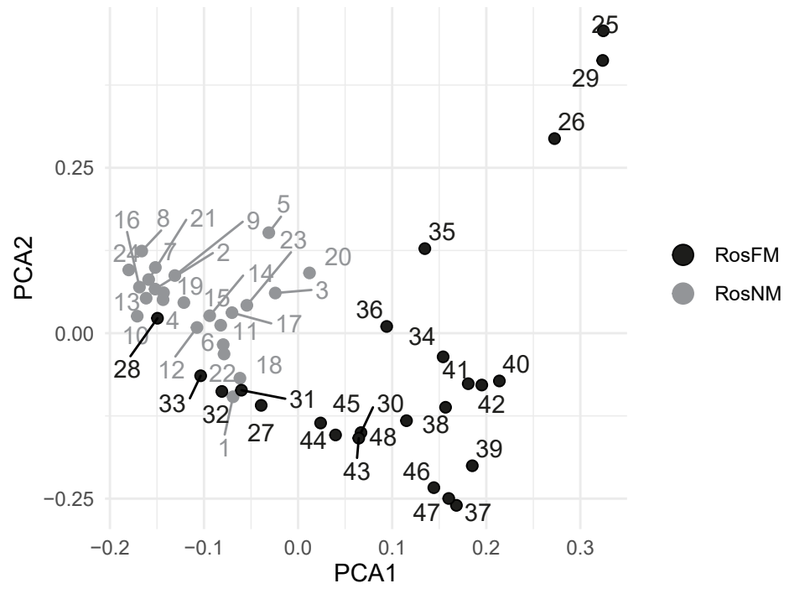
BIN 11900000_12400000



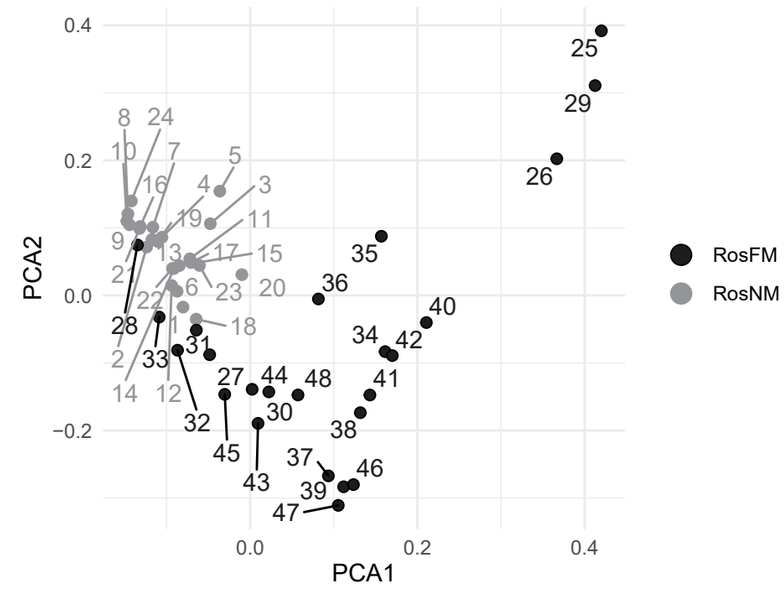
BIN 12000000_12500000



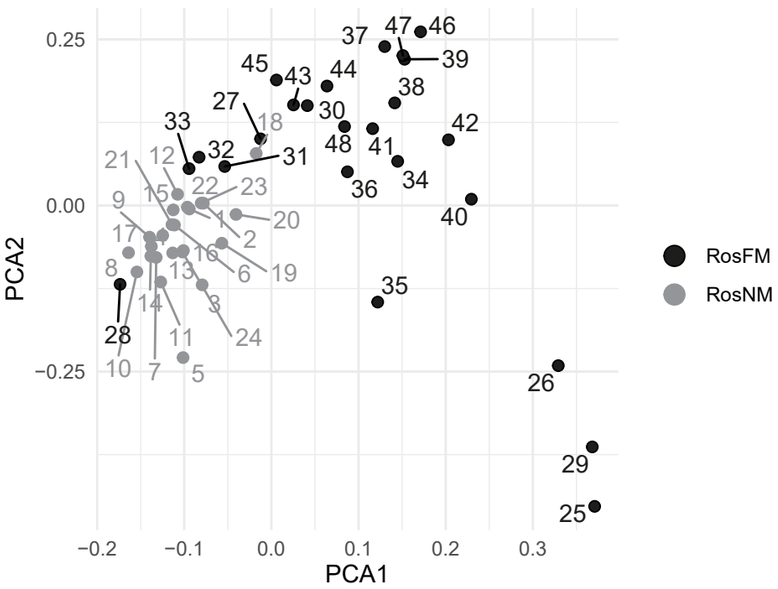
BIN 12100000_12600000



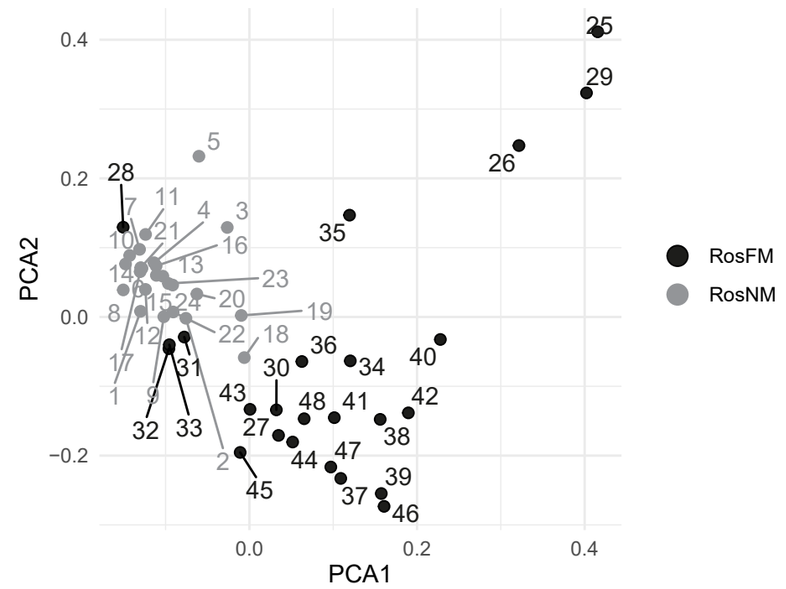
BIN 12200000_12700000



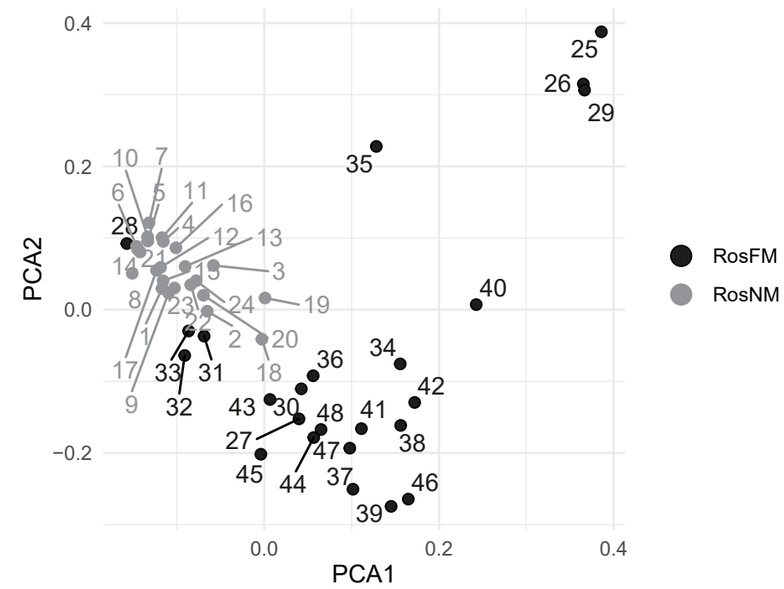
BIN 12300000_12800000



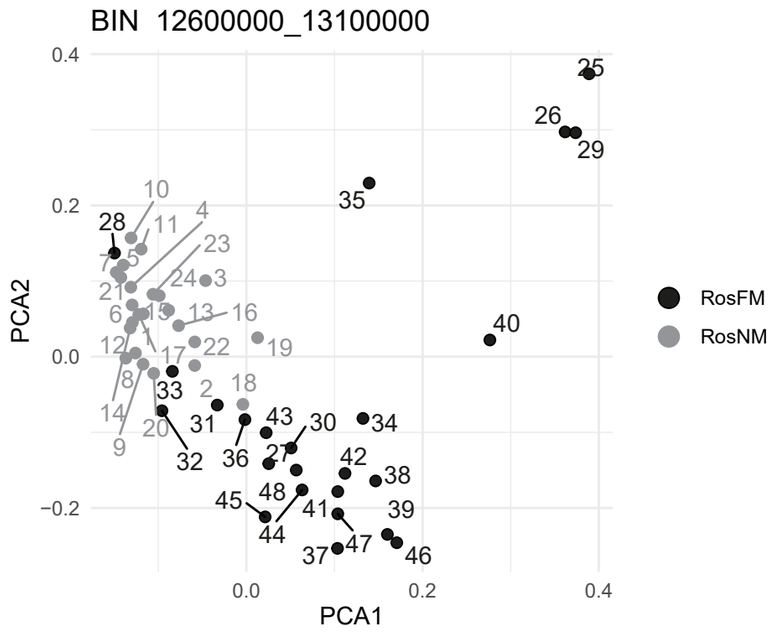
BIN 12400000_12900000



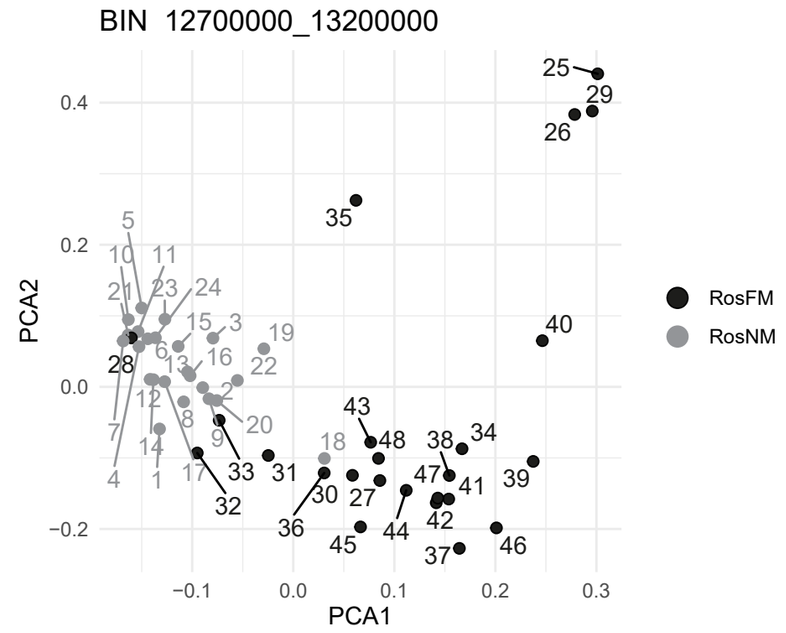
BIN 12500000_13000000



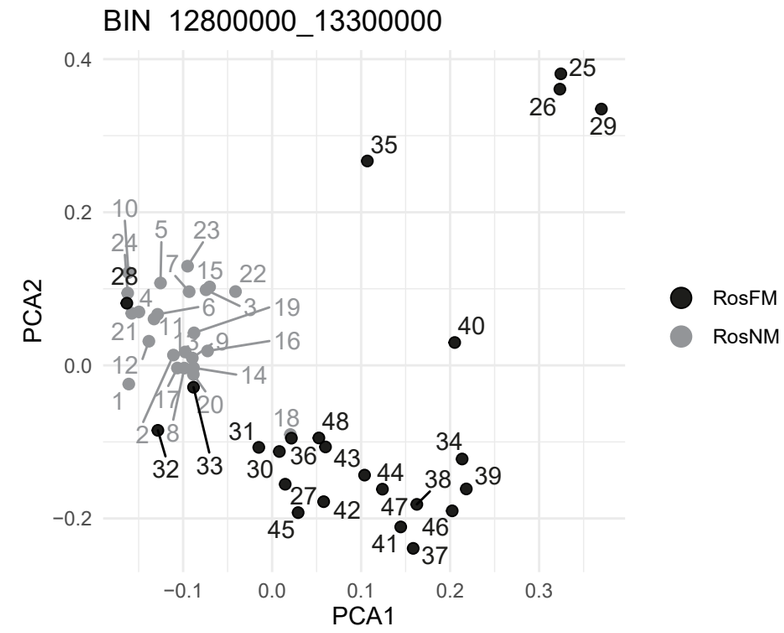
BIN 12600000_13100000



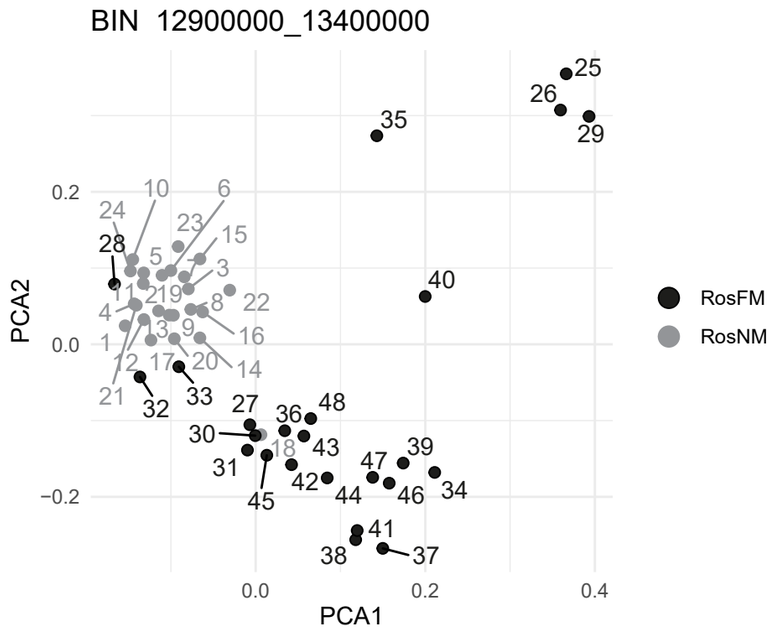
BIN 12700000_13200000



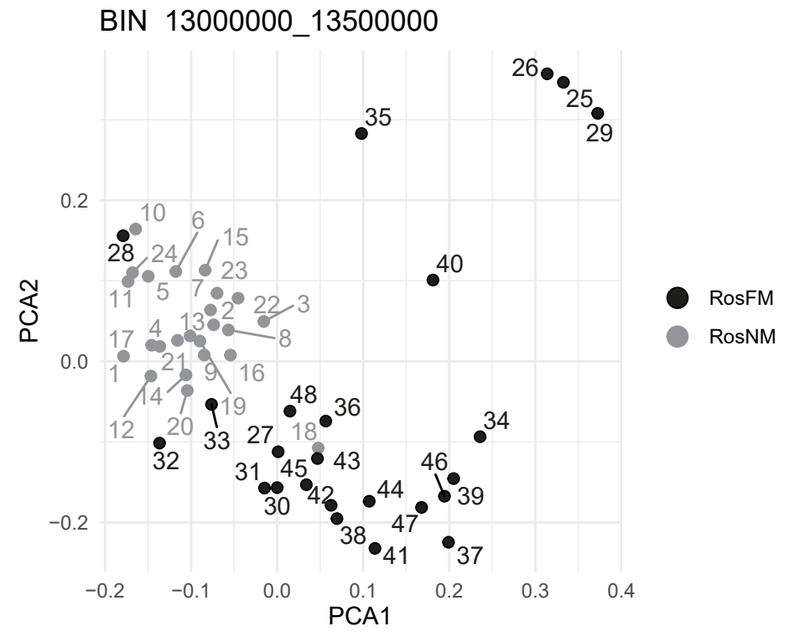
BIN 12800000_13300000



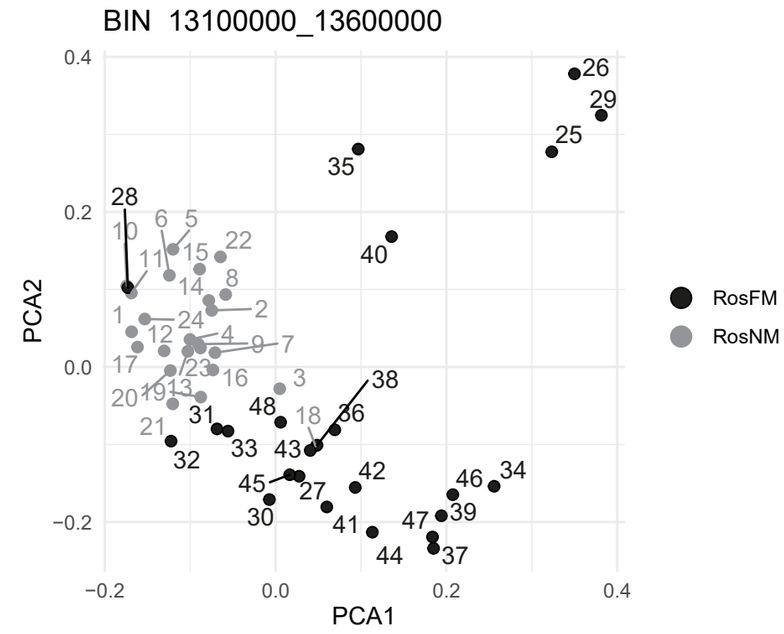
BIN 12900000_13400000



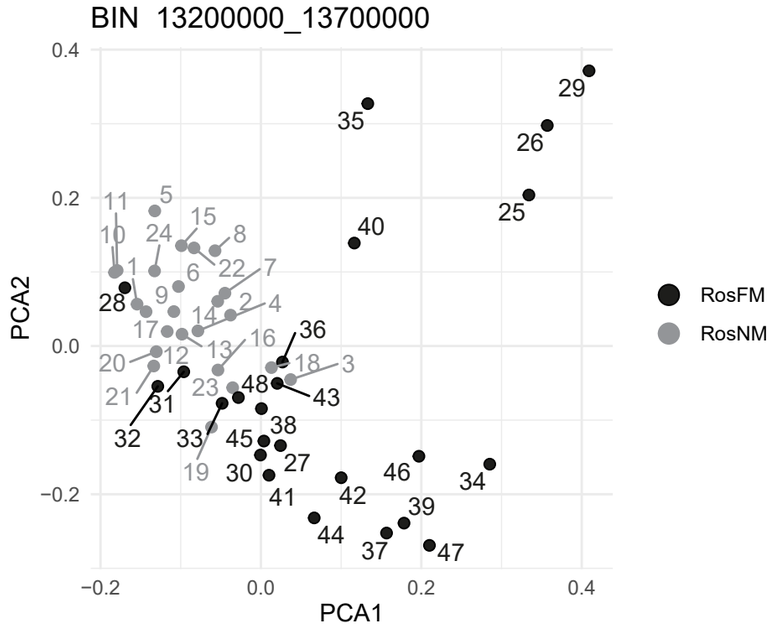
BIN 13000000_13500000



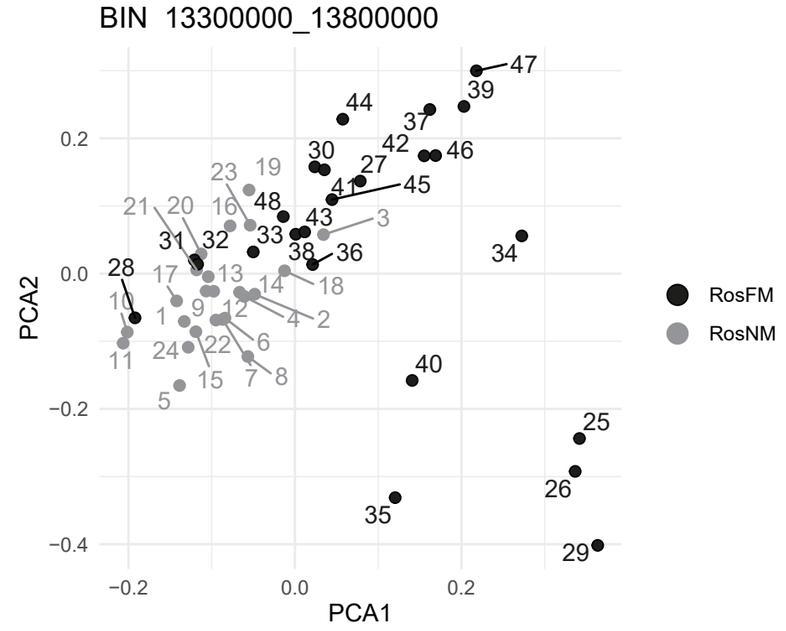
BIN 13100000_13600000



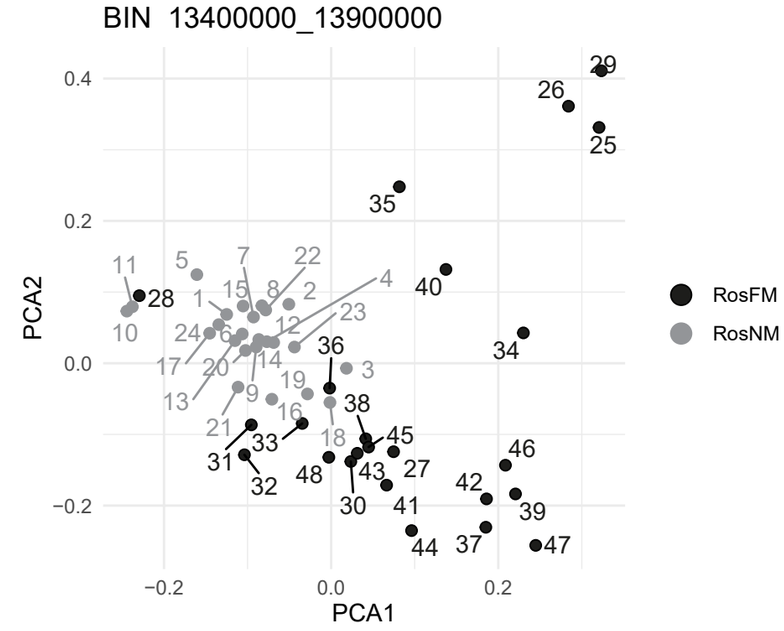
BIN 13200000_13700000



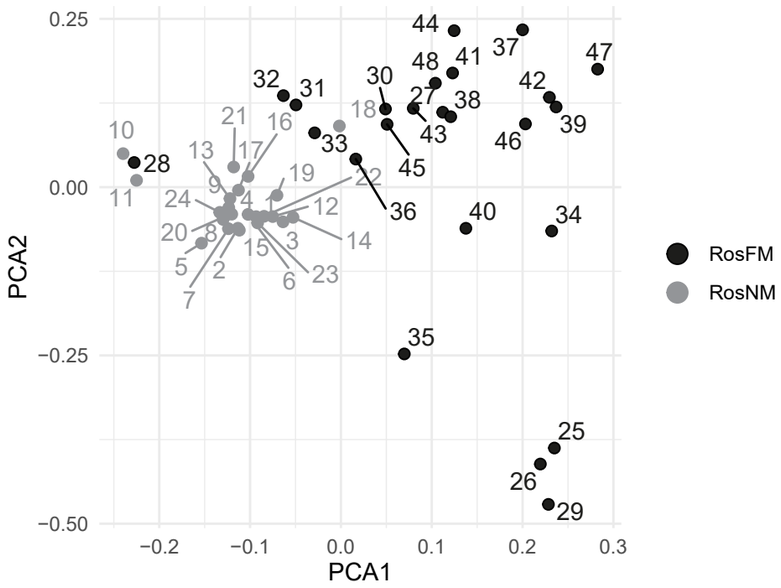
BIN 13300000_13800000



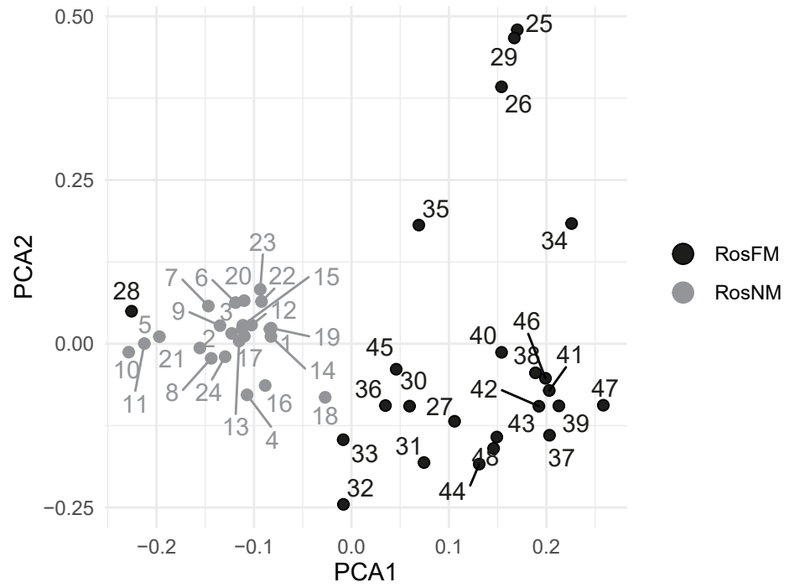
BIN 13400000_13900000



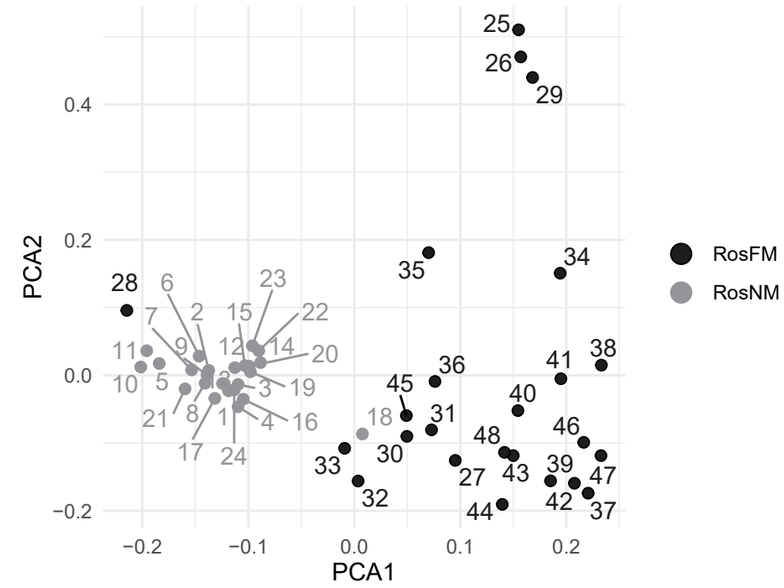
BIN 13500000_14000000



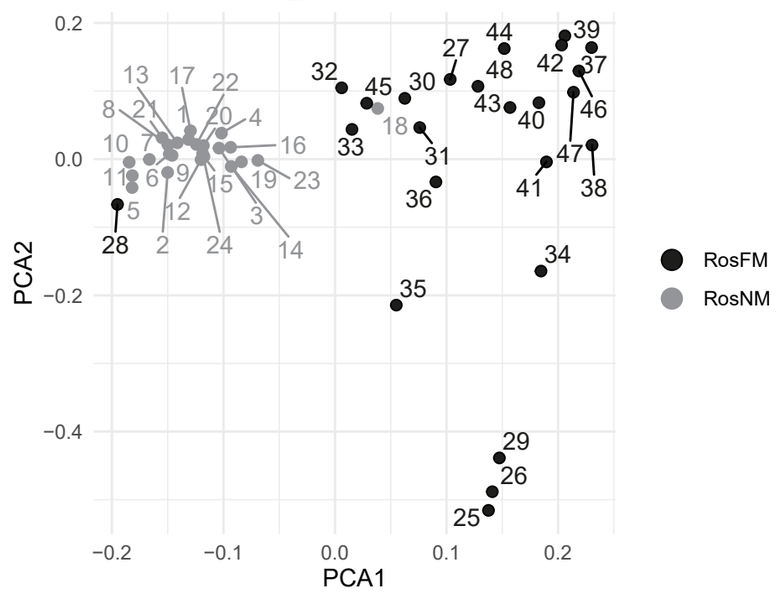
BIN 13600000_14100000



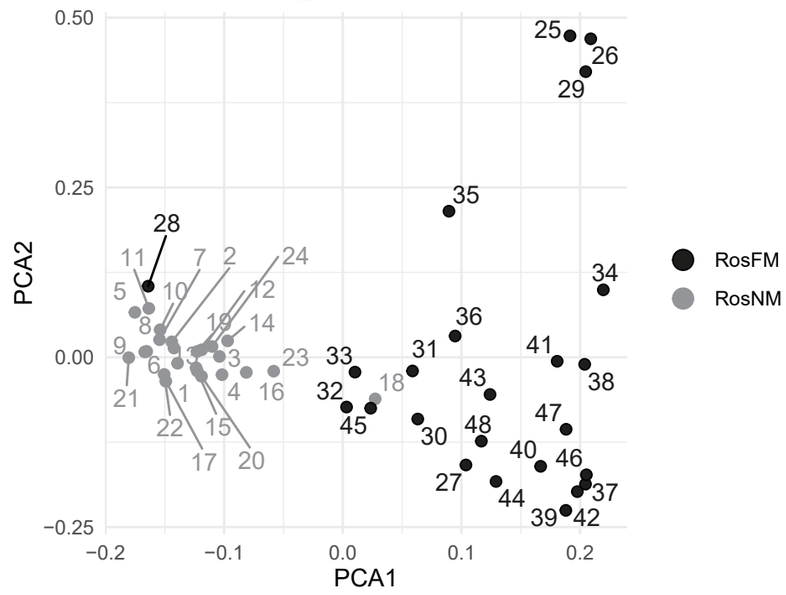
BIN 13700000_14200000



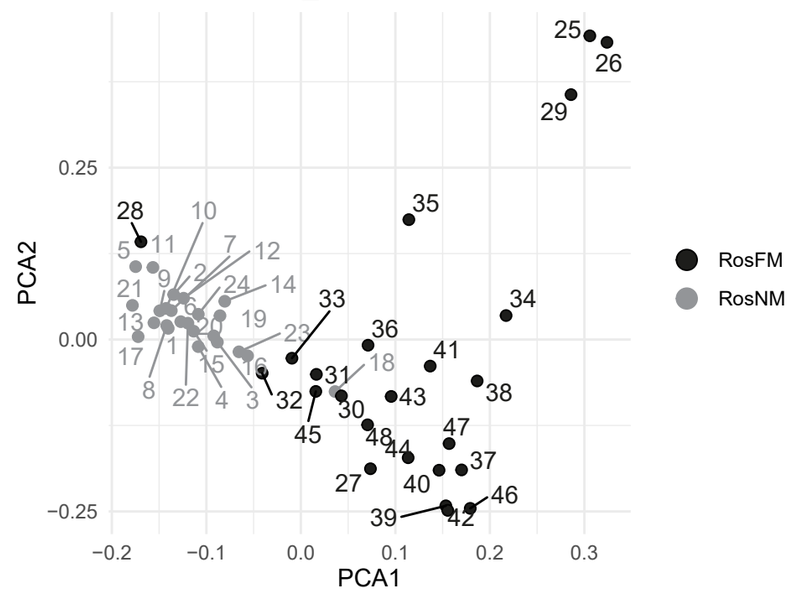
BIN 13800000_14300000



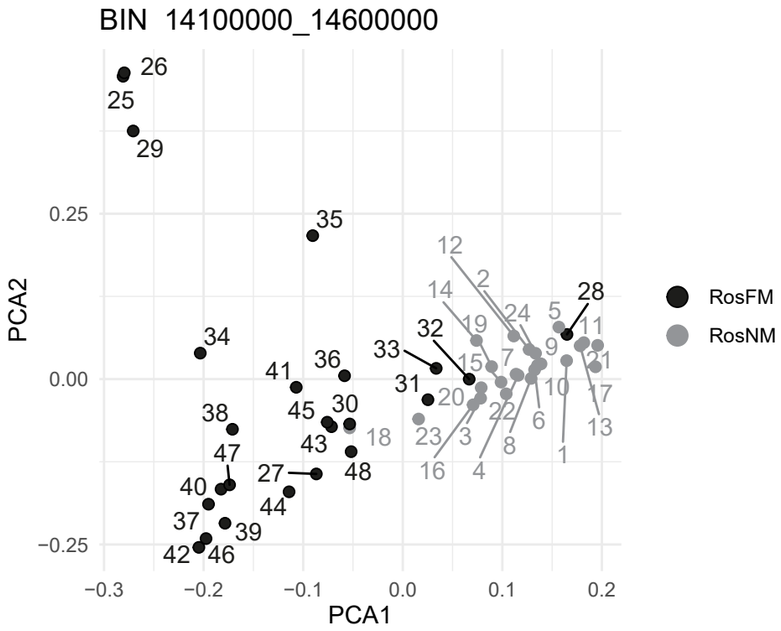
BIN 13900000_14400000



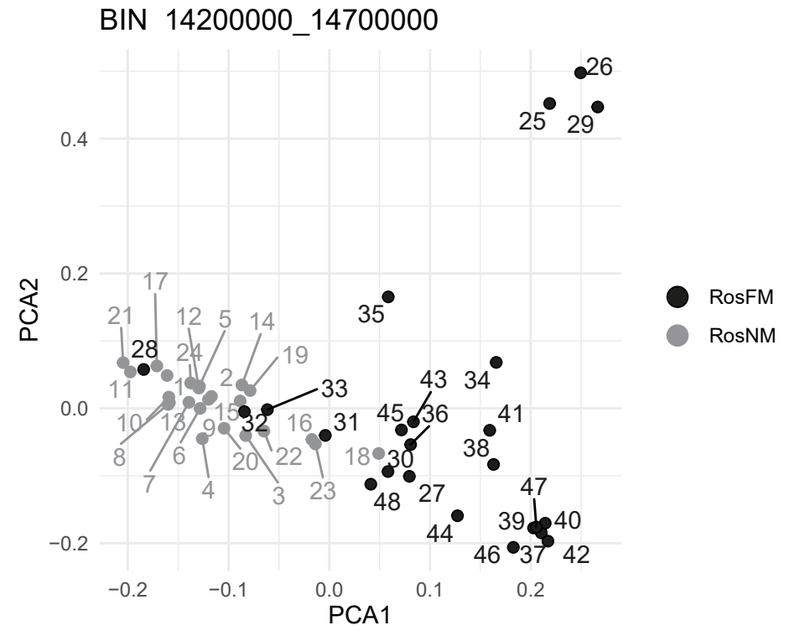
BIN 14000000_14500000



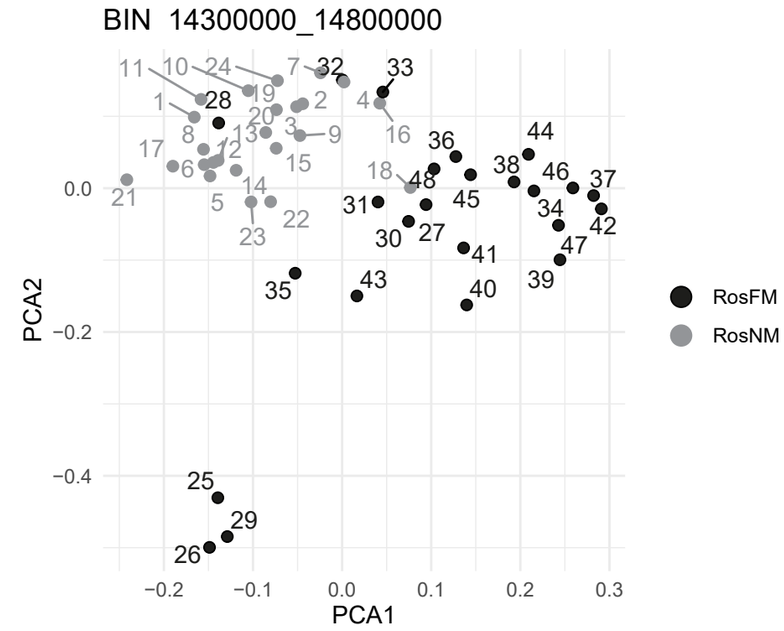
BIN 14100000_14600000



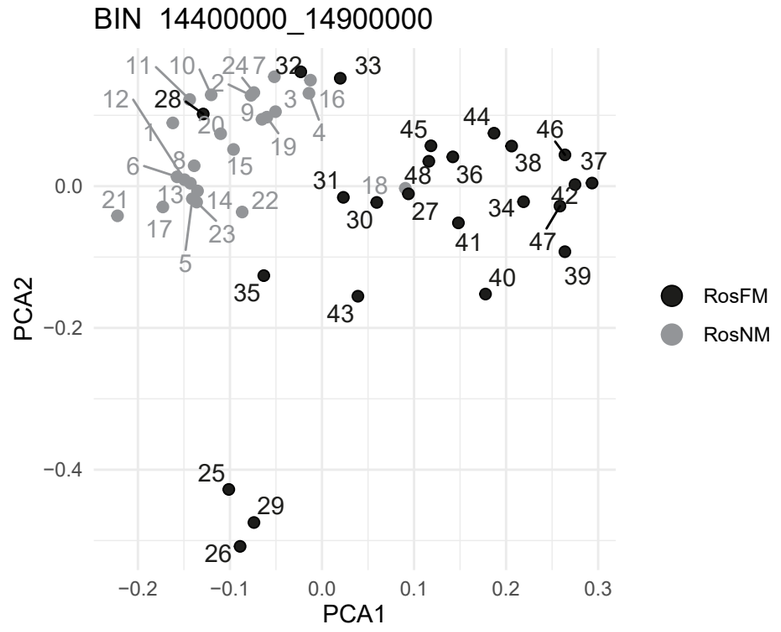
BIN 14200000_14700000



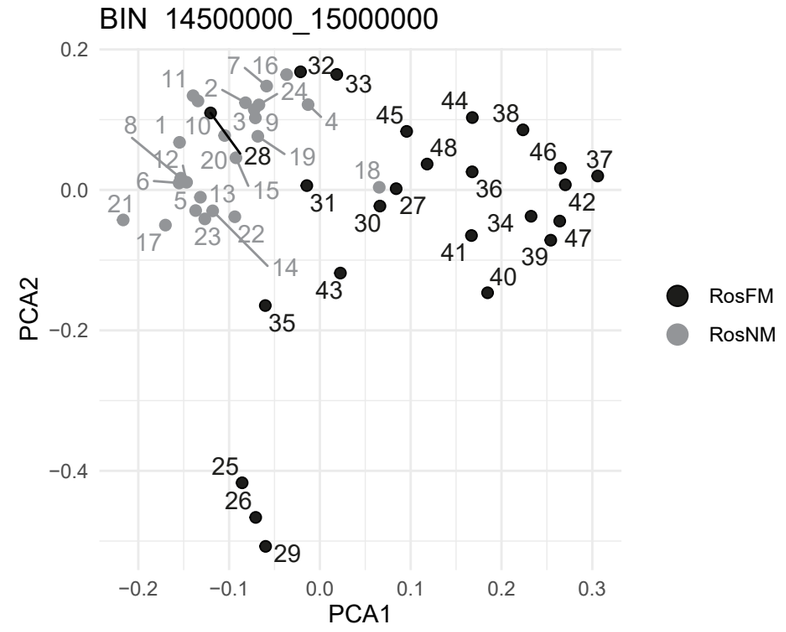
BIN 14300000_14800000



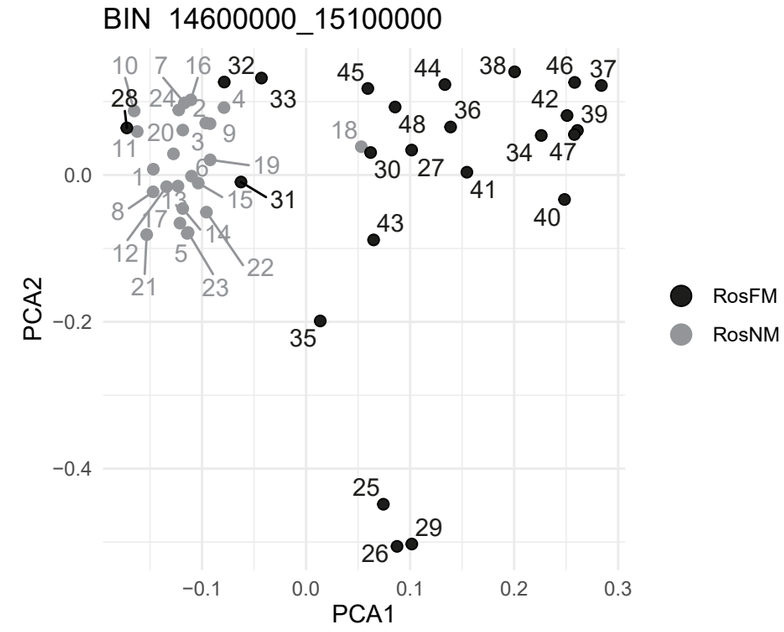
BIN 14400000_14900000



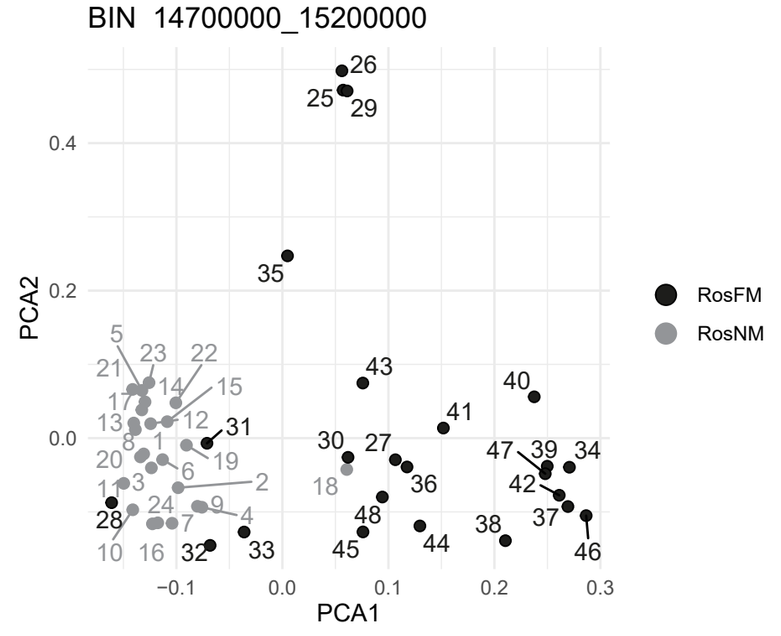
BIN 14500000_15000000



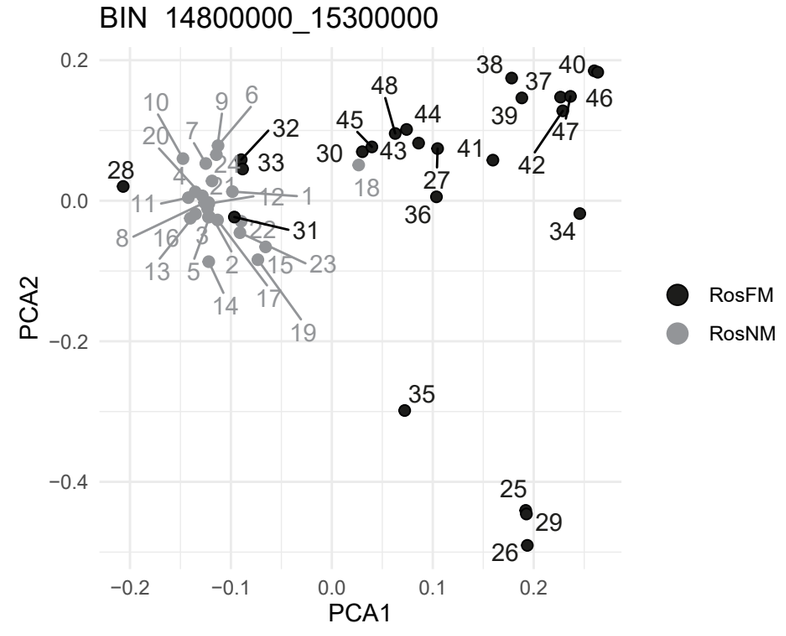
BIN 14600000_15100000



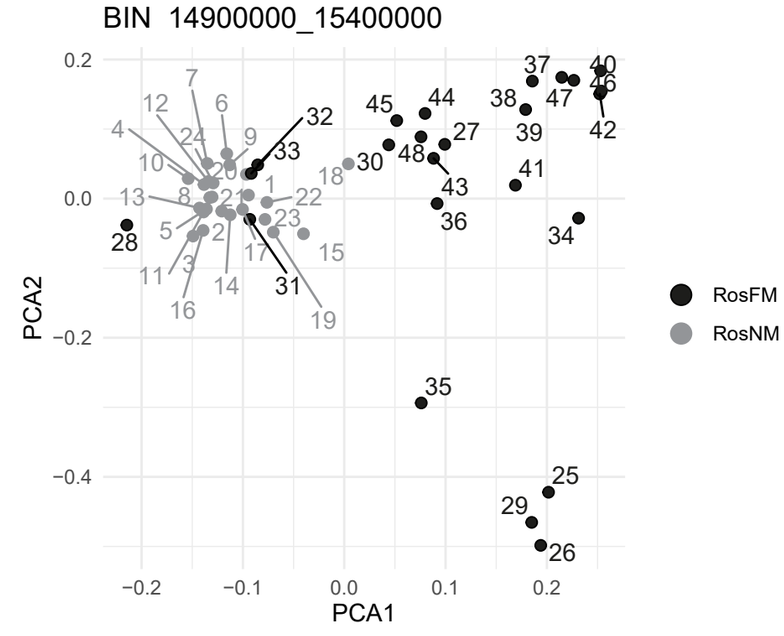
BIN 14700000_15200000



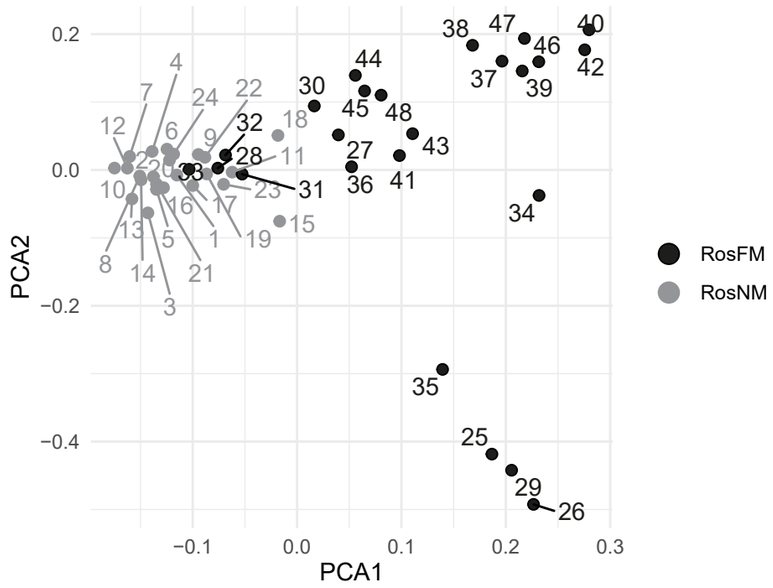
BIN 14800000_15300000



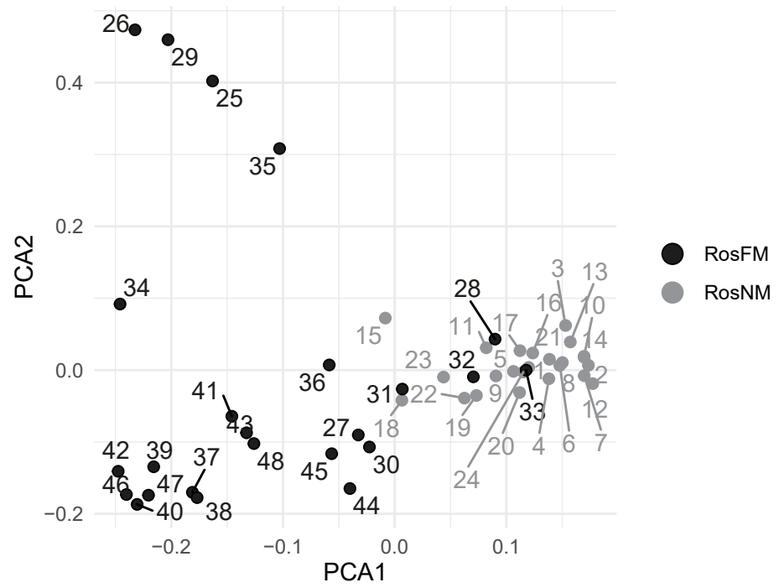
BIN 14900000_15400000



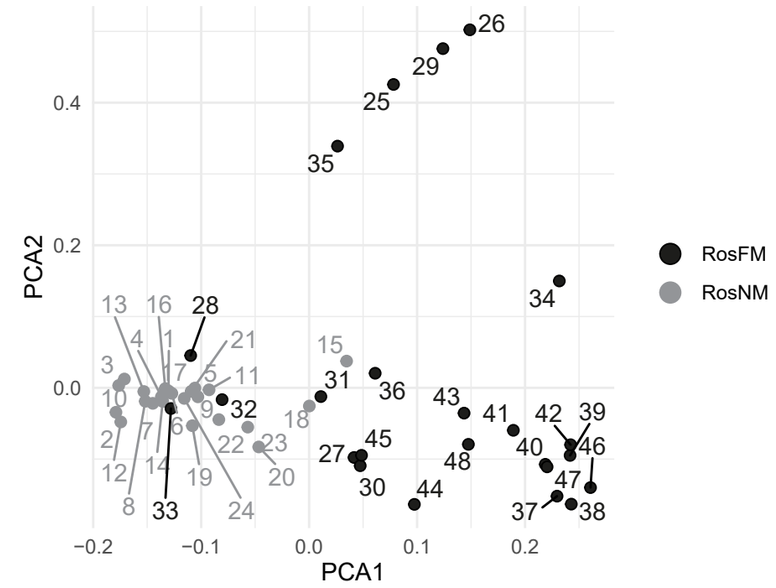
BIN 15000000_15500000



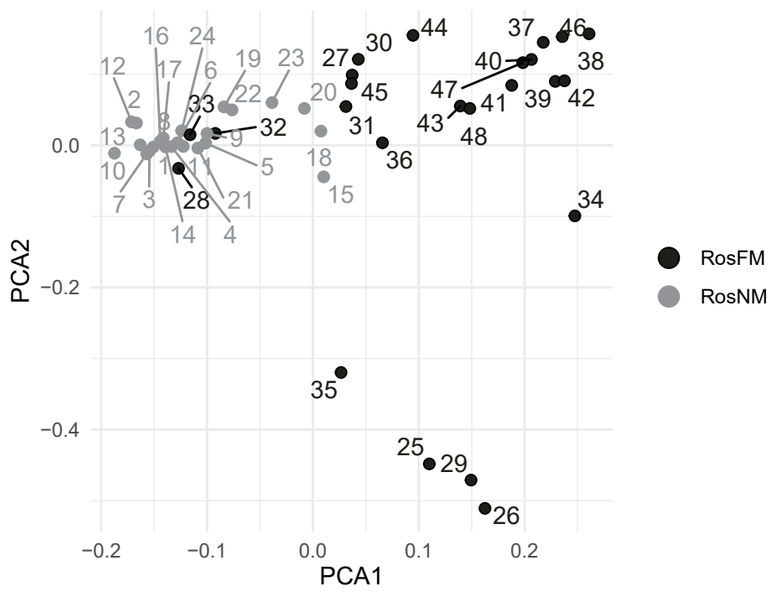
BIN 15100000_15600000



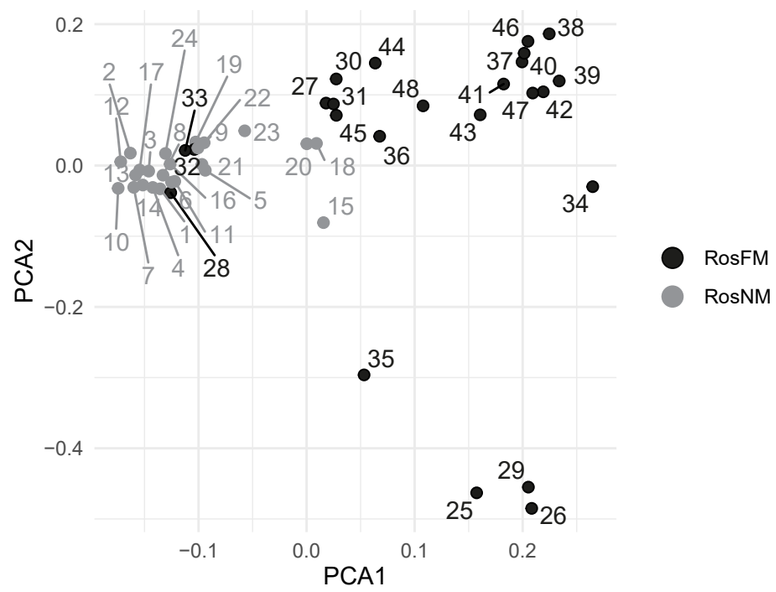
BIN 15200000_15700000



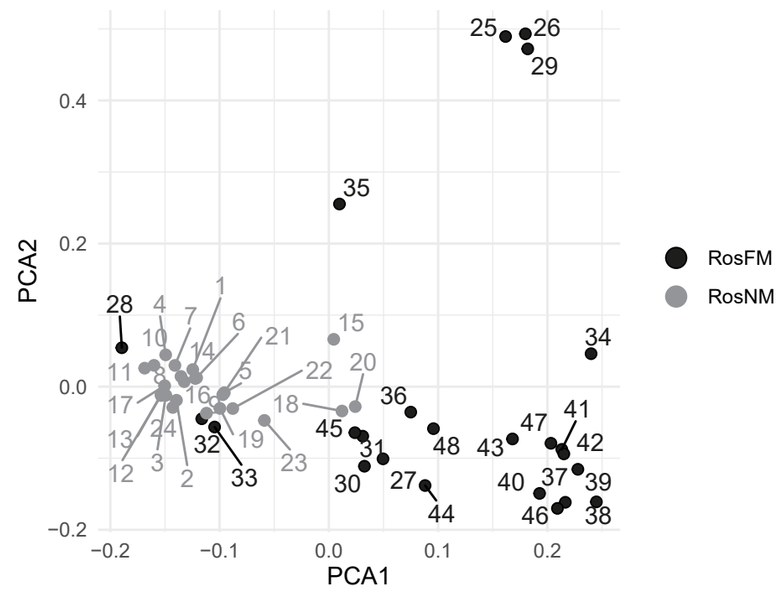
BIN 15300000_15800000



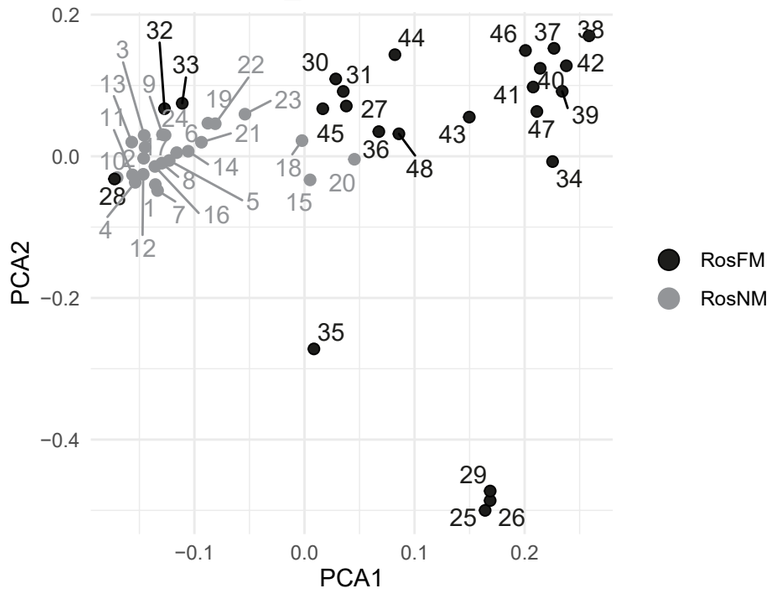
BIN 15400000_15900000



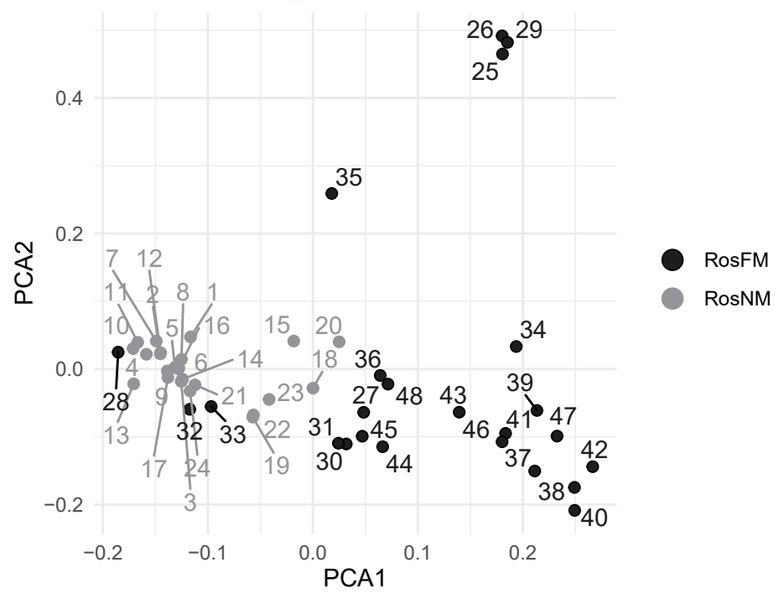
BIN 15500000_16000000



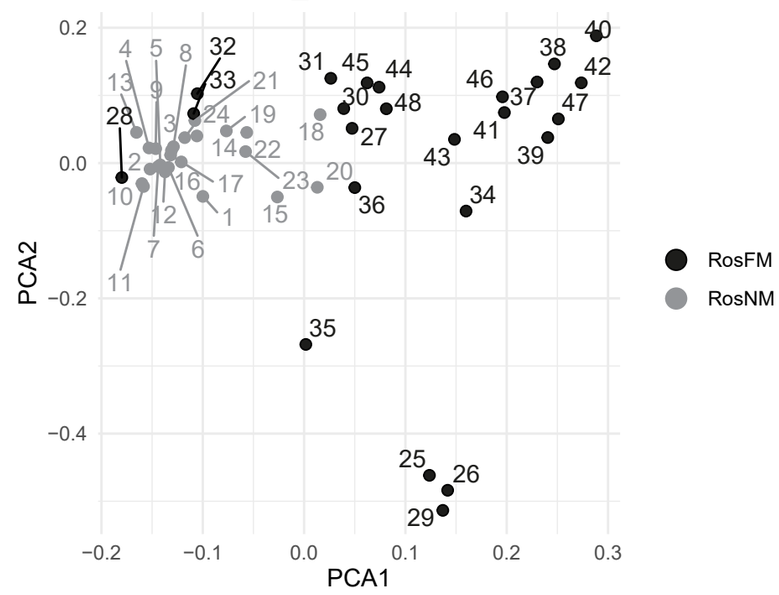
BIN 15600000_16100000



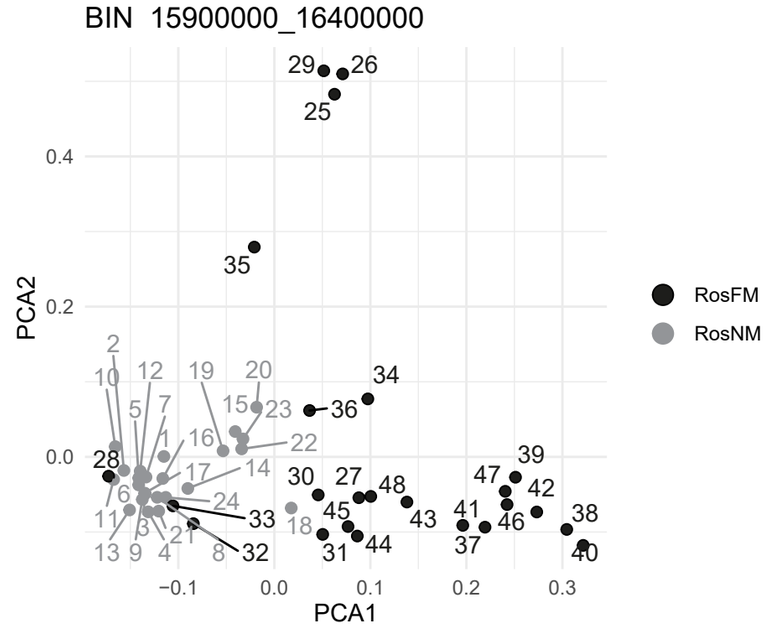
BIN 15700000_16200000



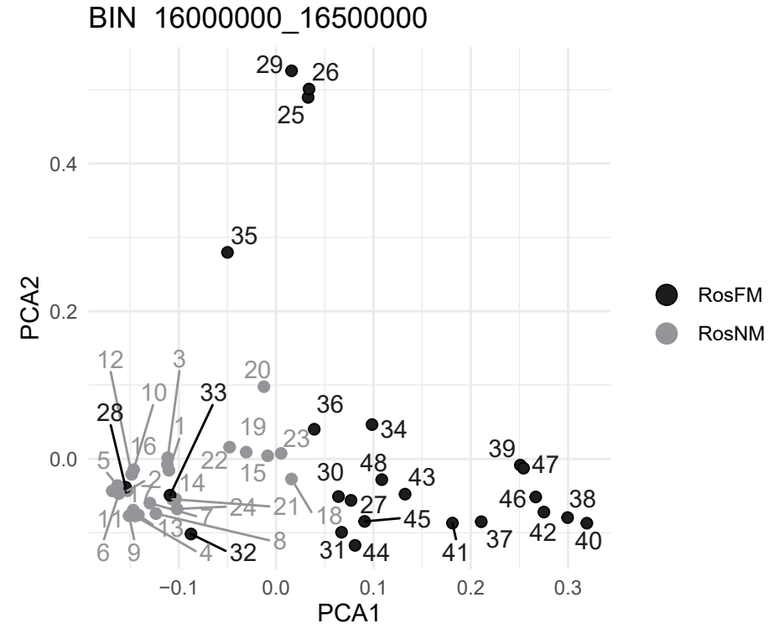
BIN 15800000_16300000



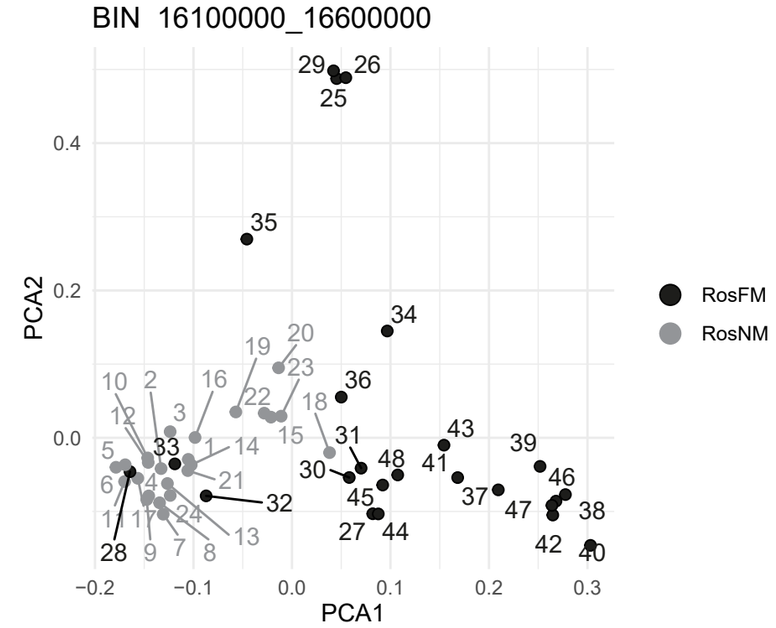
BIN 15900000_16400000



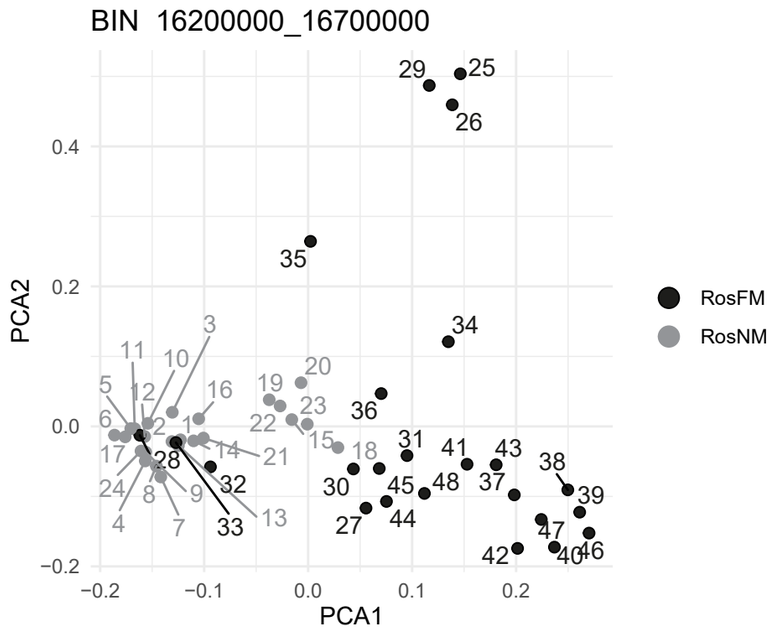
BIN 16000000_16500000



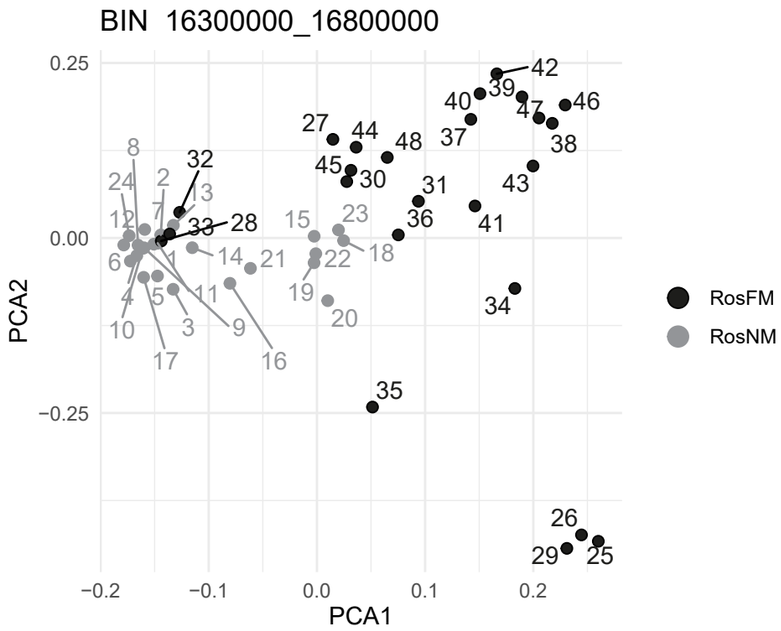
BIN 16100000_16600000



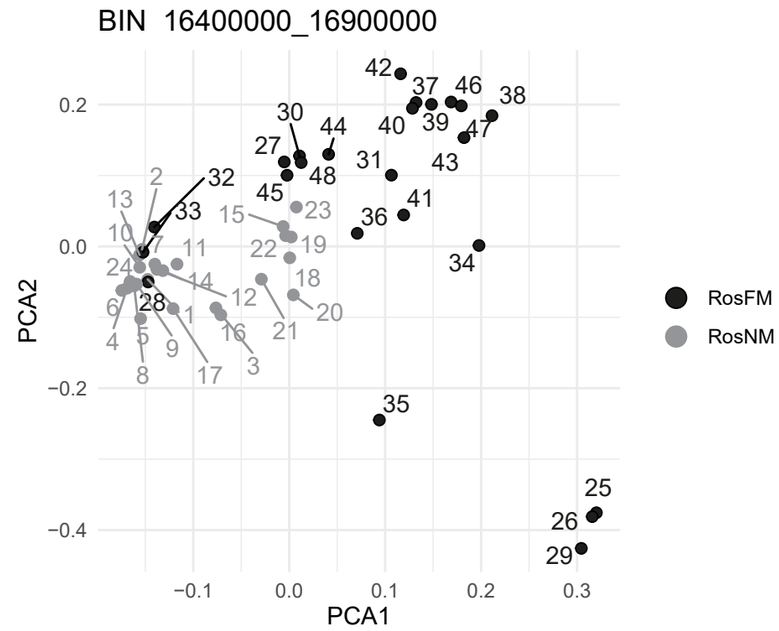
BIN 16200000_16700000



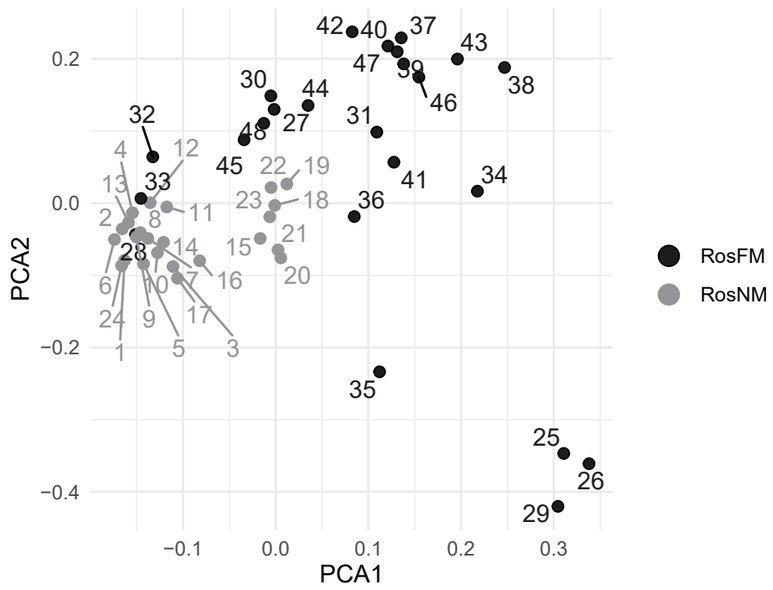
BIN 16300000_16800000



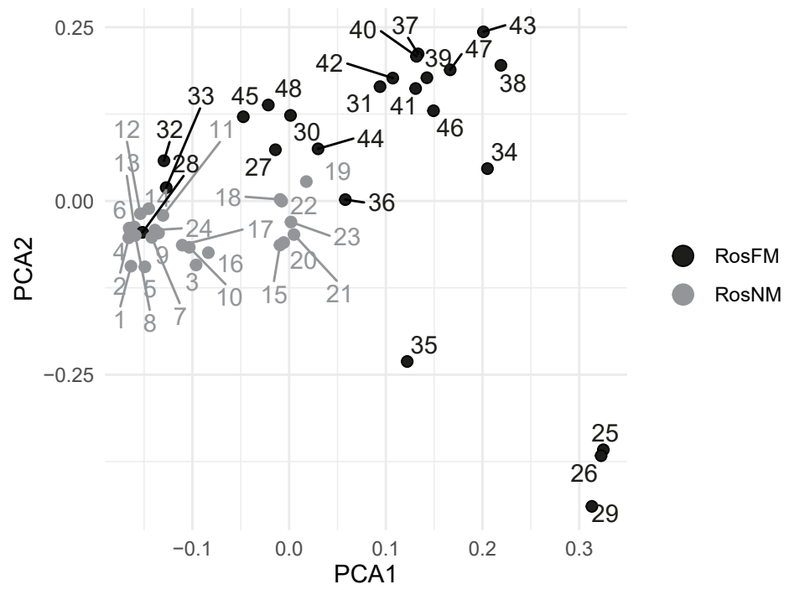
BIN 16400000_16900000



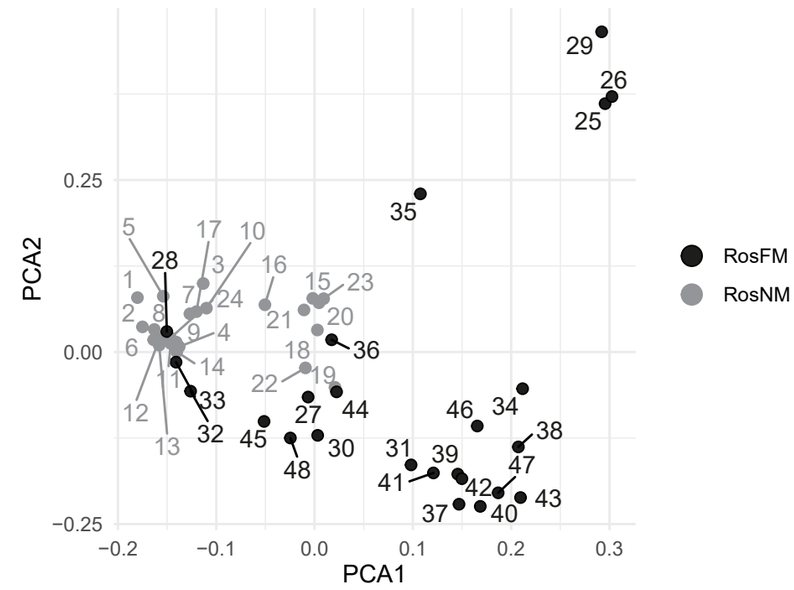
BIN 16500000_17000000



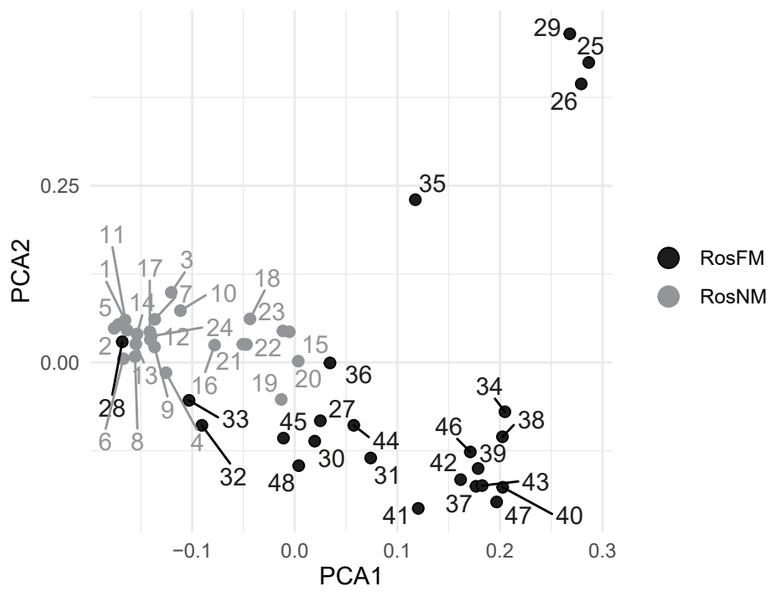
BIN 16600000_17100000



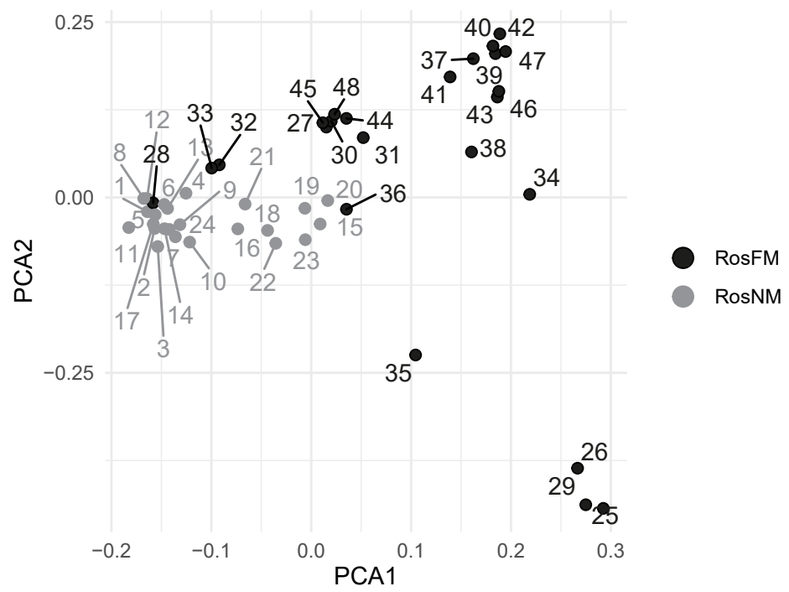
BIN 16700000_17200000



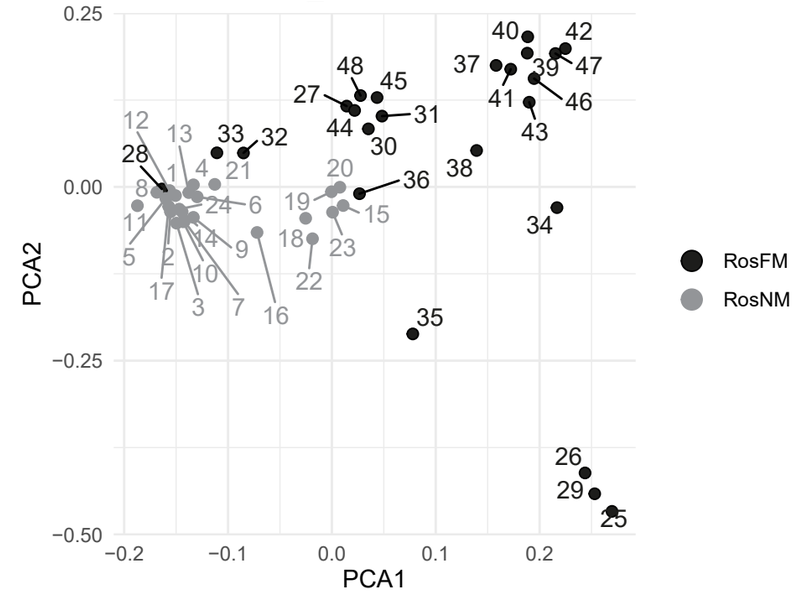
BIN 16800000_17300000



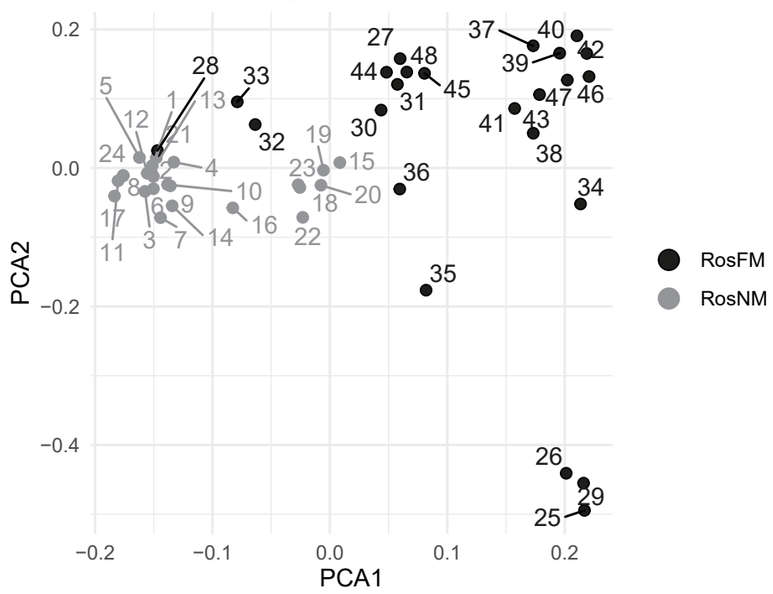
BIN 16900000_17400000



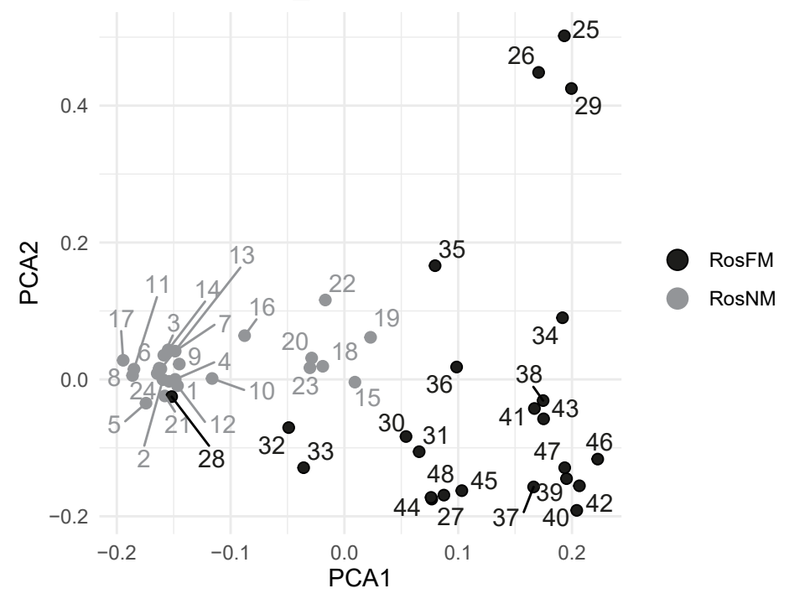
BIN 17000000_17500000



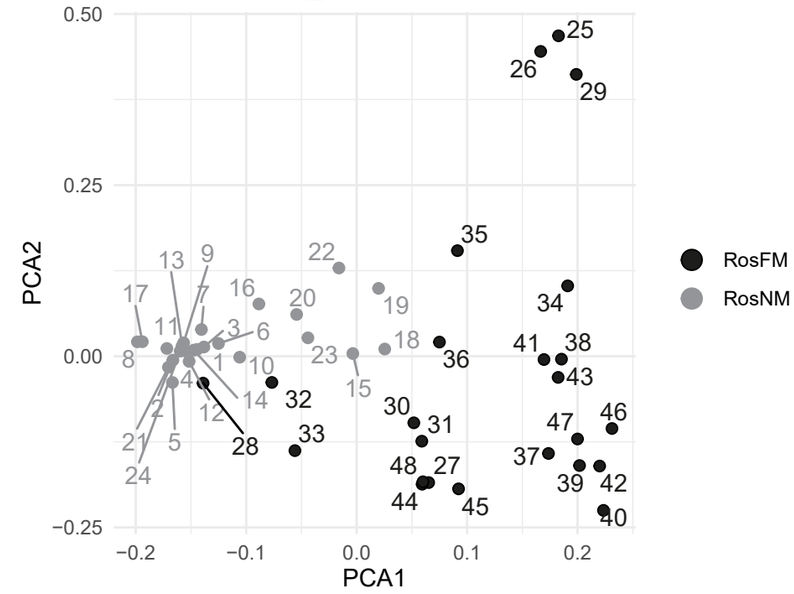
BIN 17100000_17600000



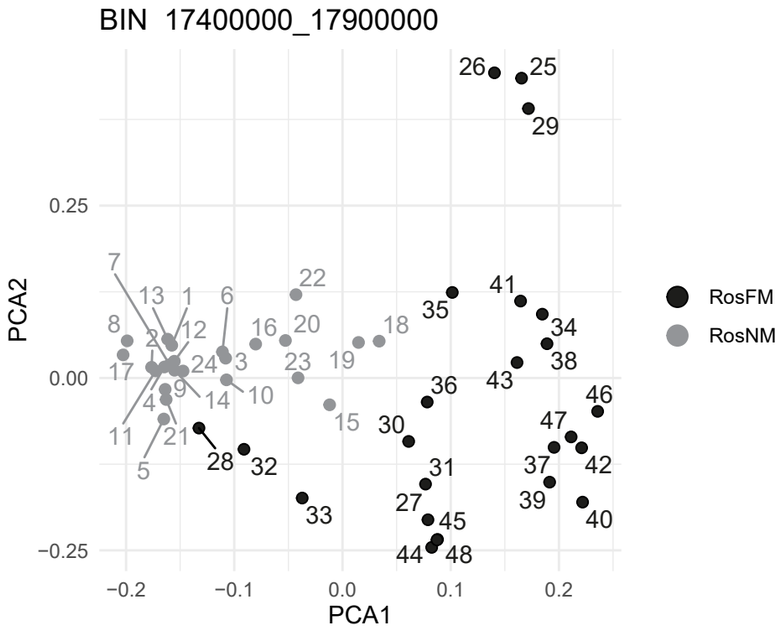
BIN 17200000_17700000



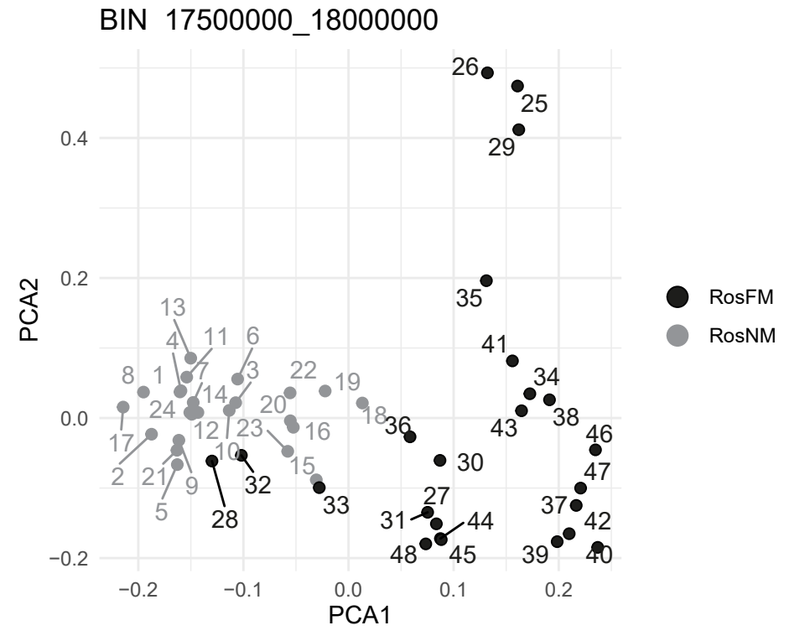
BIN 17300000_17800000



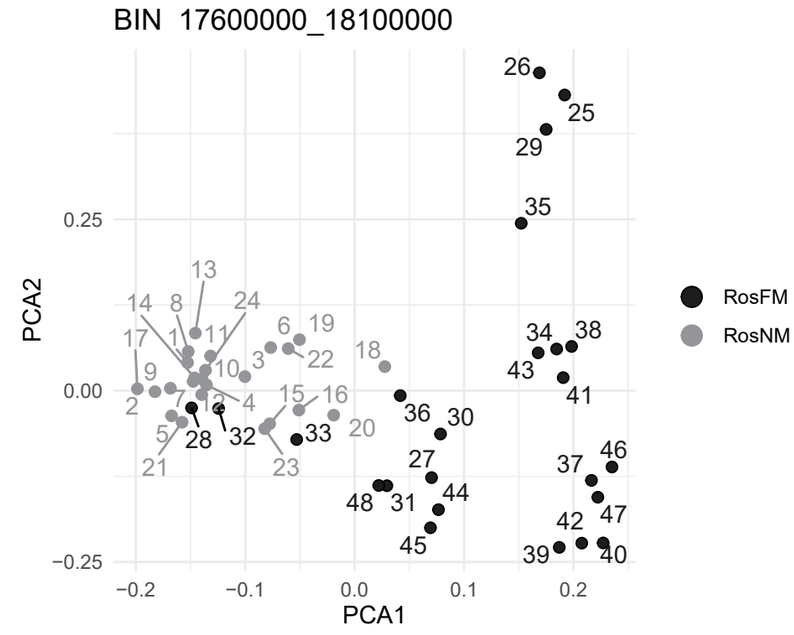
BIN 17400000_17900000



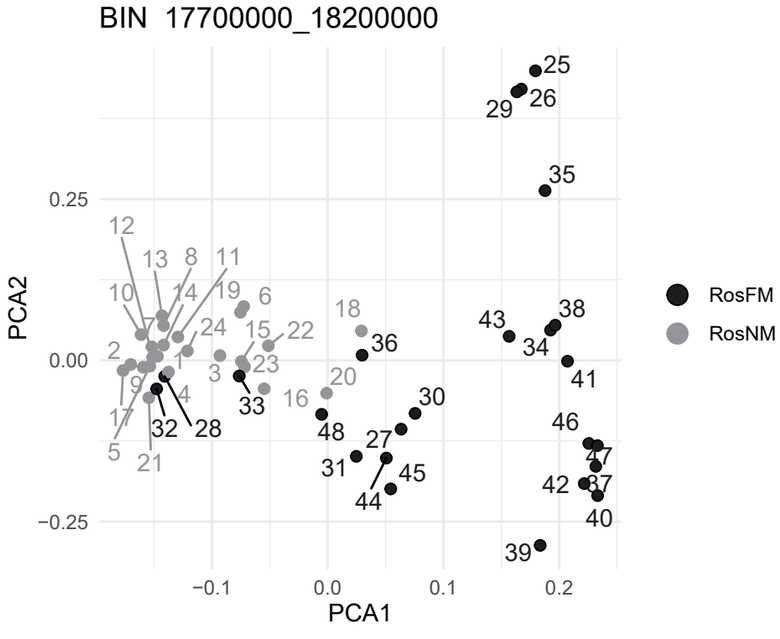
BIN 17500000_18000000



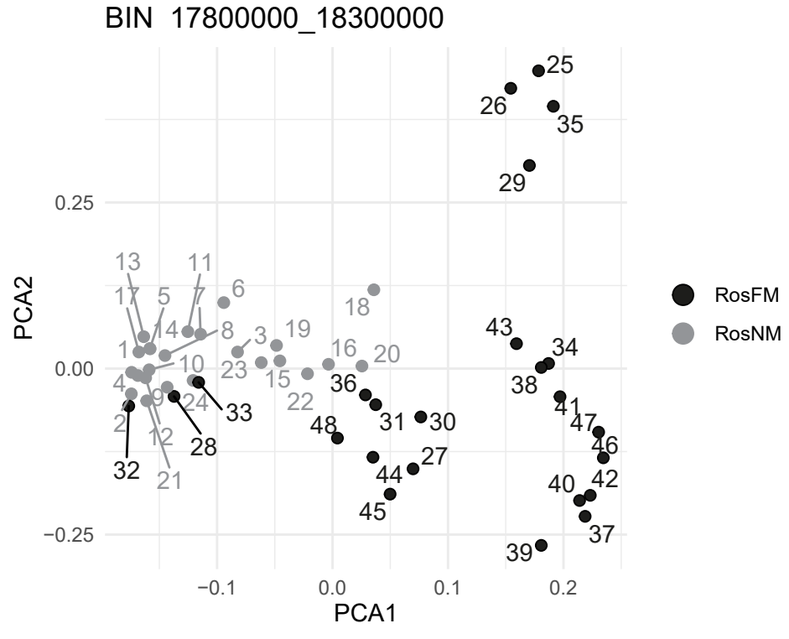
BIN 17600000_18100000



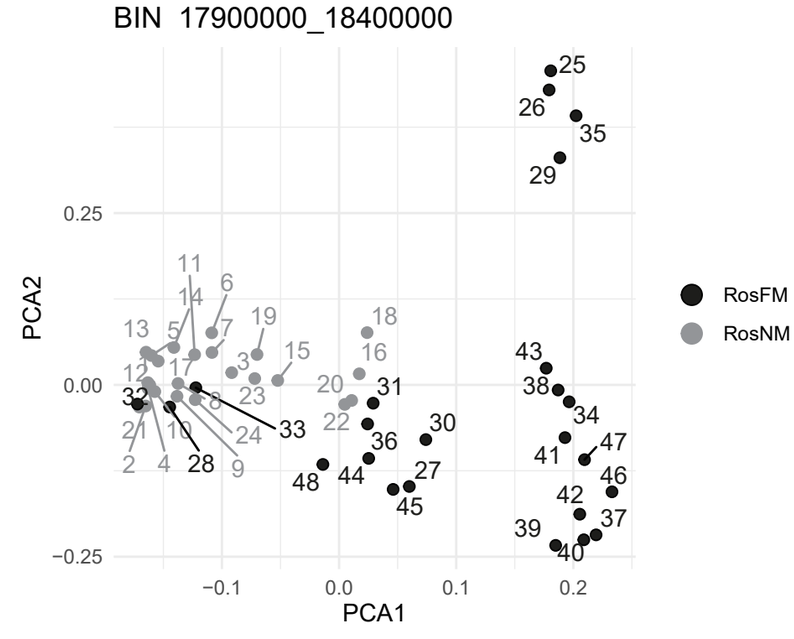
BIN 17700000_18200000



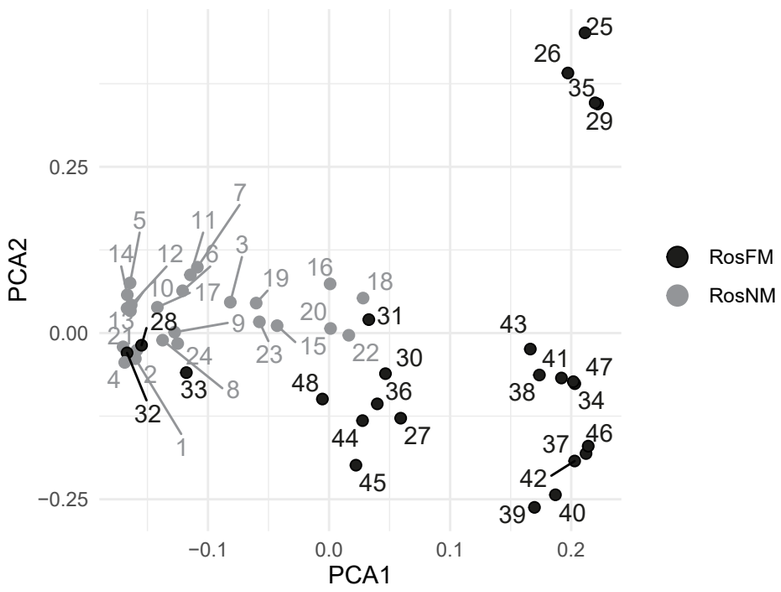
BIN 17800000_18300000



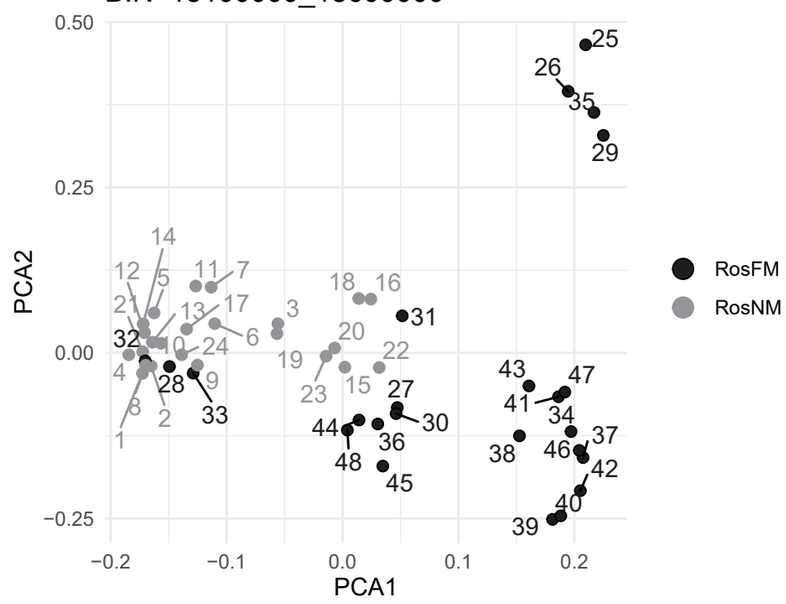
BIN 17900000_18400000



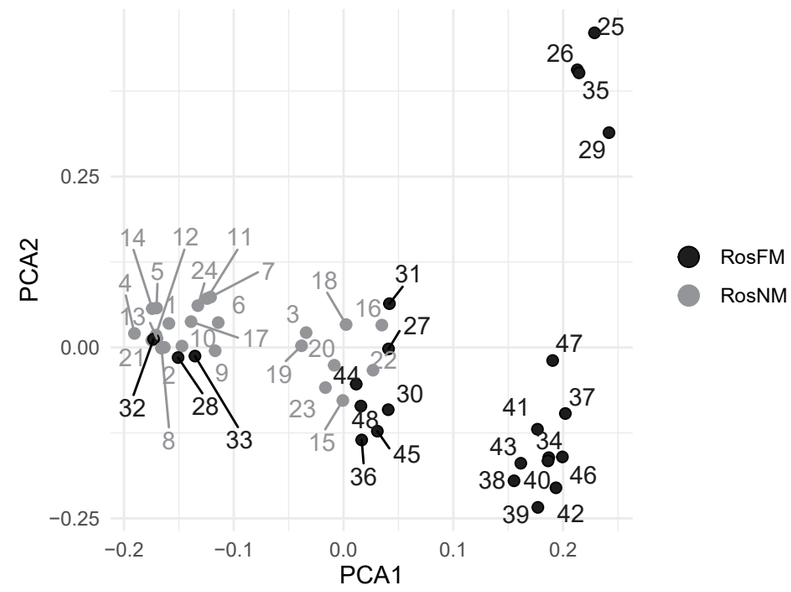
BIN 18000000_18500000



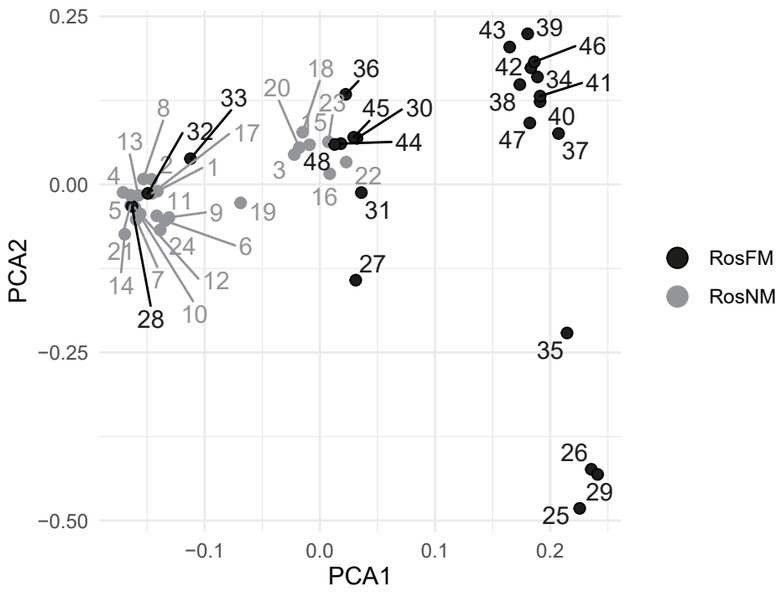
BIN 18100000_18600000



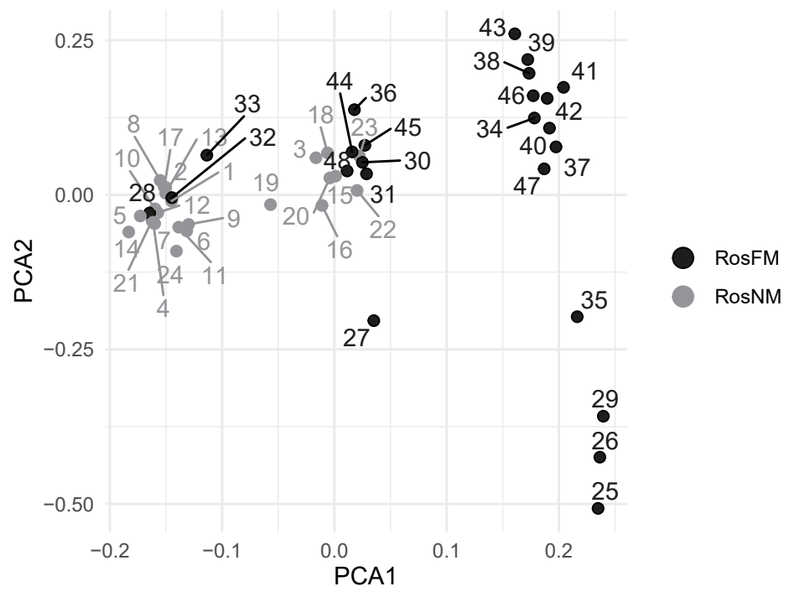
BIN 18200000_18700000



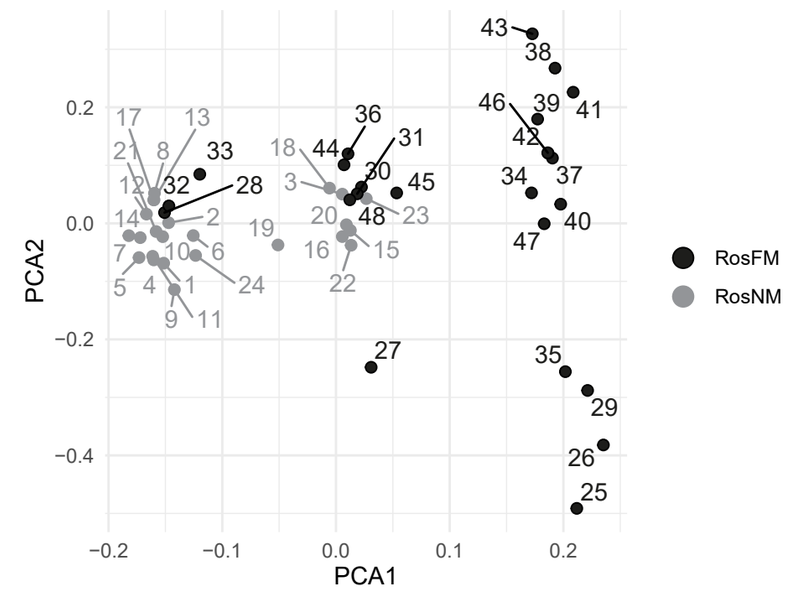
BIN 18300000_18800000



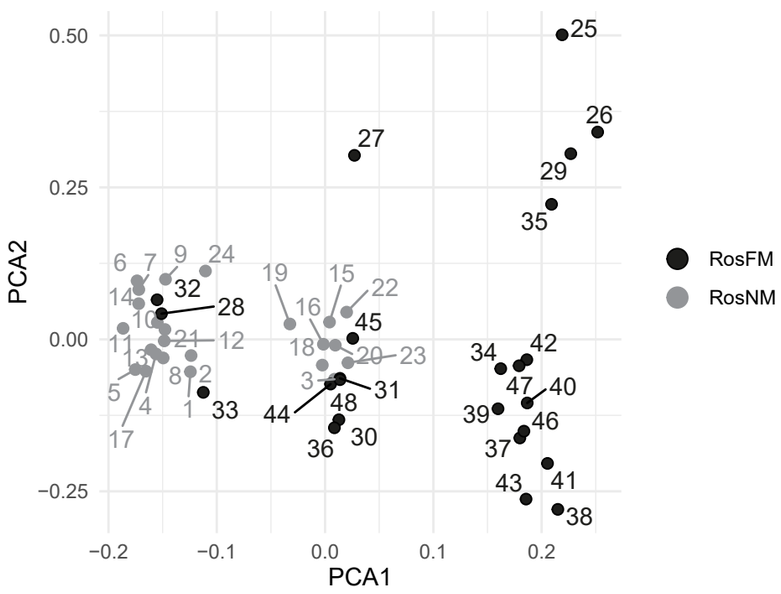
BIN 18400000_18900000



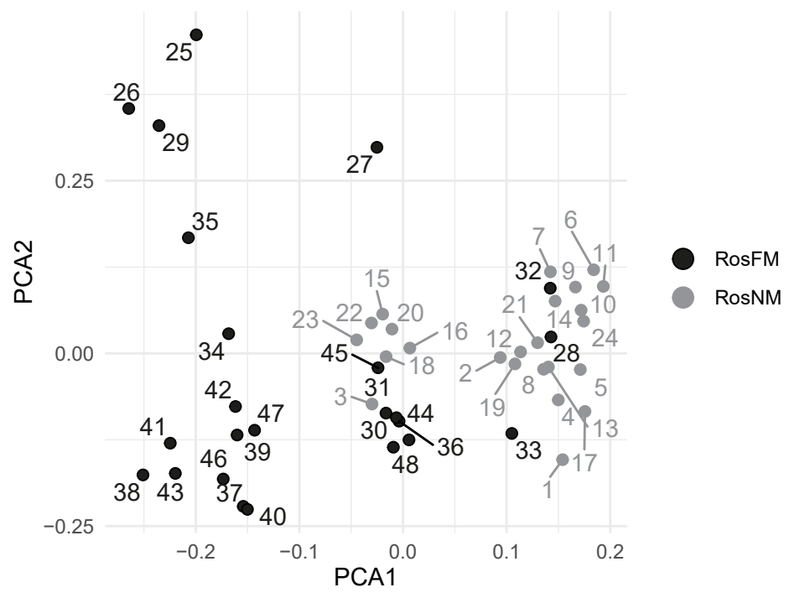
BIN 18500000_19000000



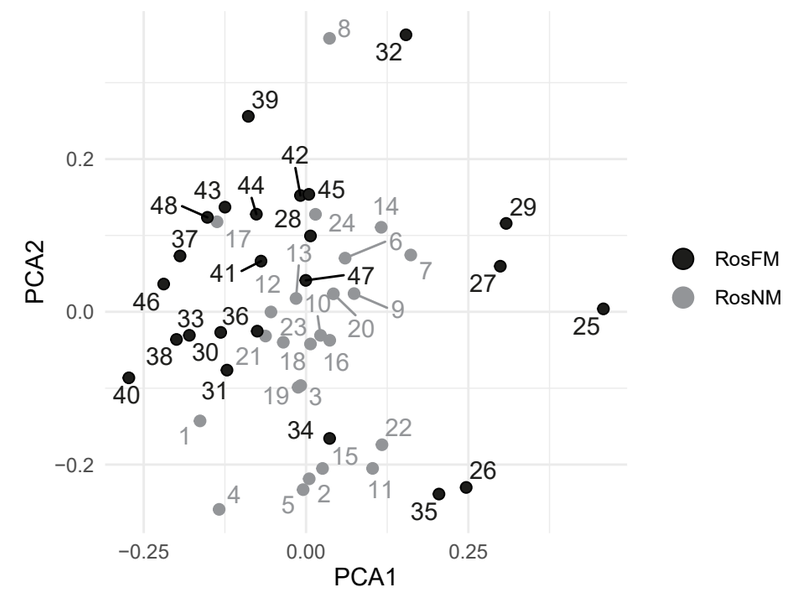
BIN 18600000_19100000



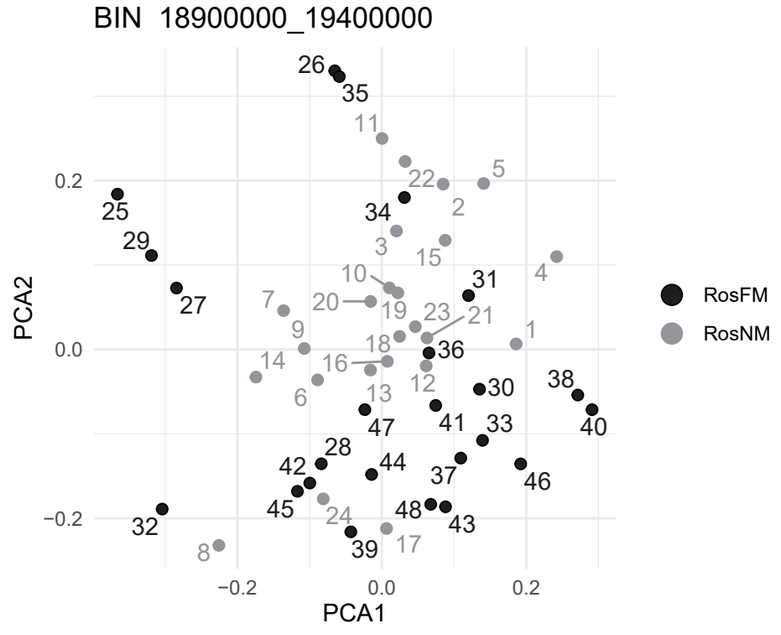
BIN 18700000_19200000



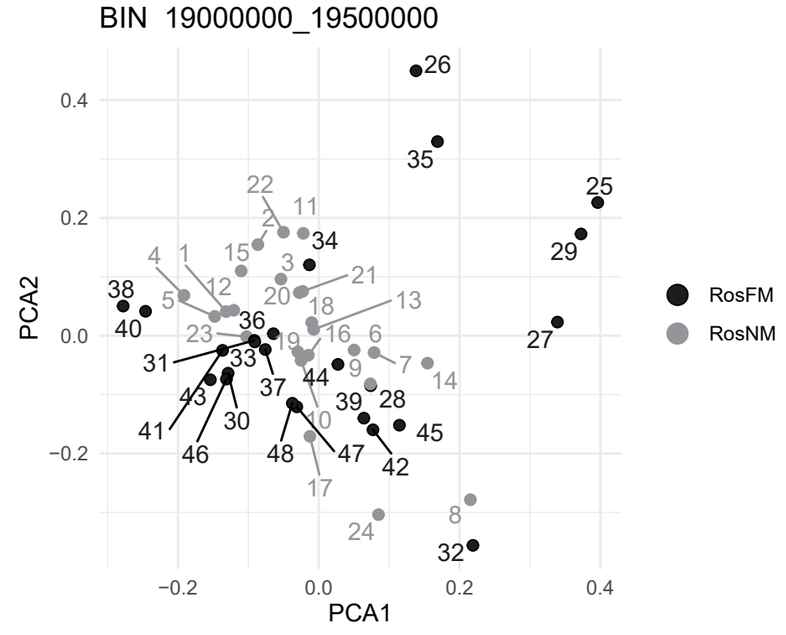
BIN 18800000_19300000



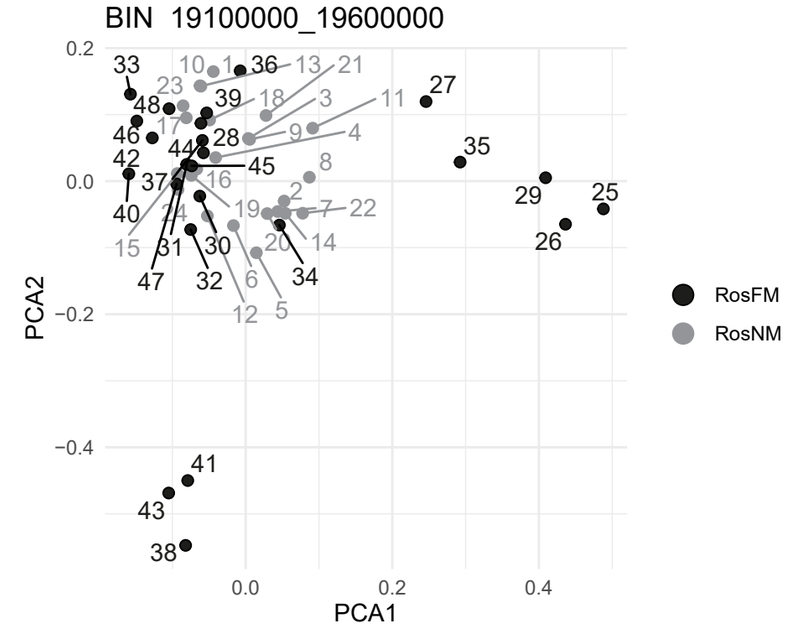
BIN 18900000_19400000



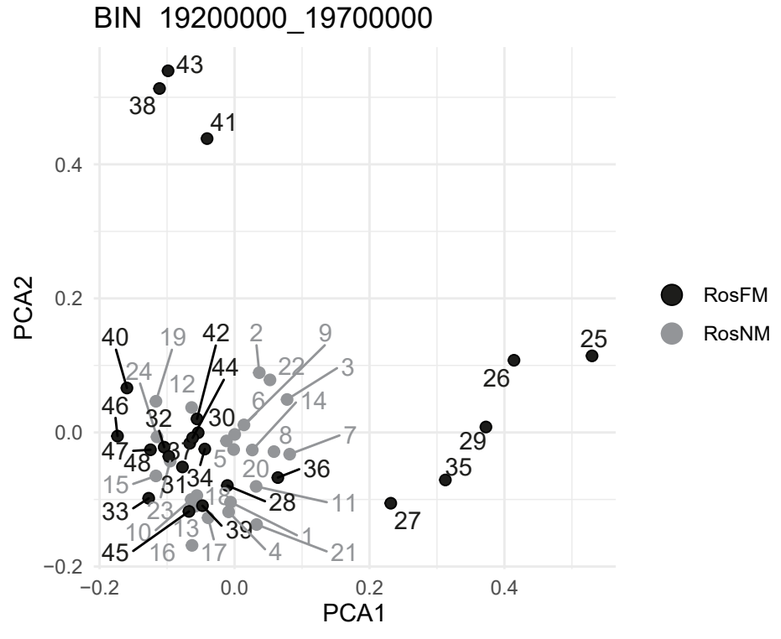
BIN 19000000_19500000



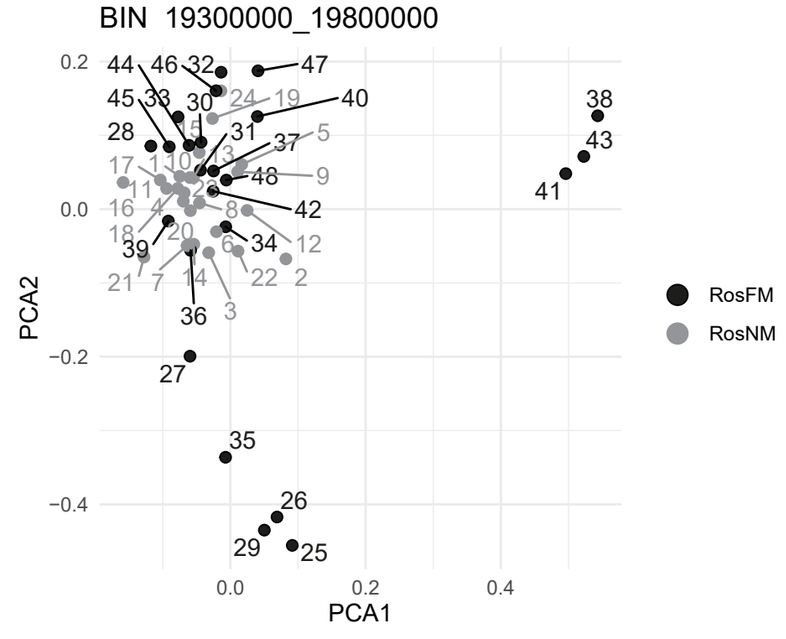
BIN 19100000_19600000



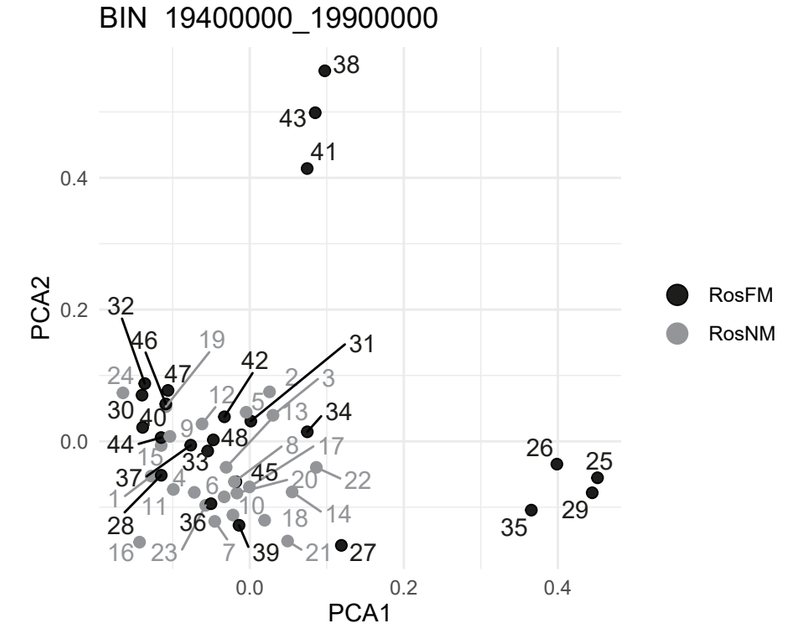
BIN 19200000_19700000



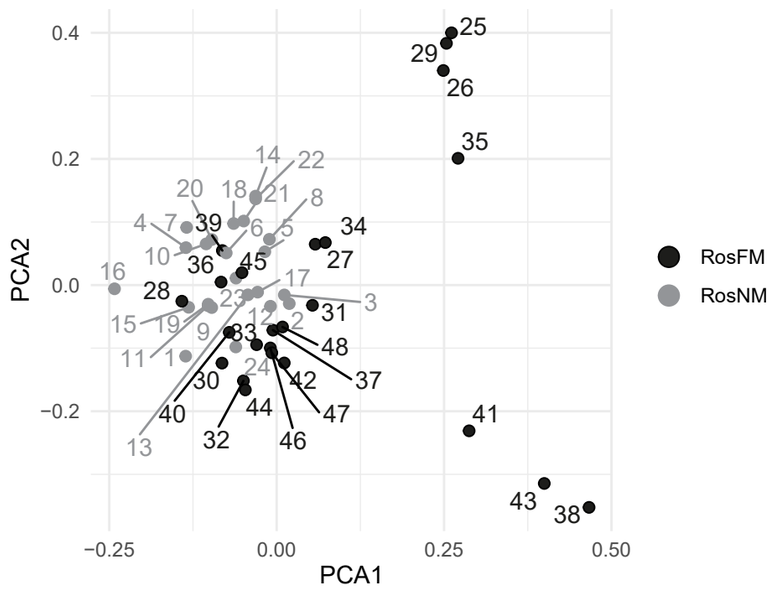
BIN 19300000_19800000



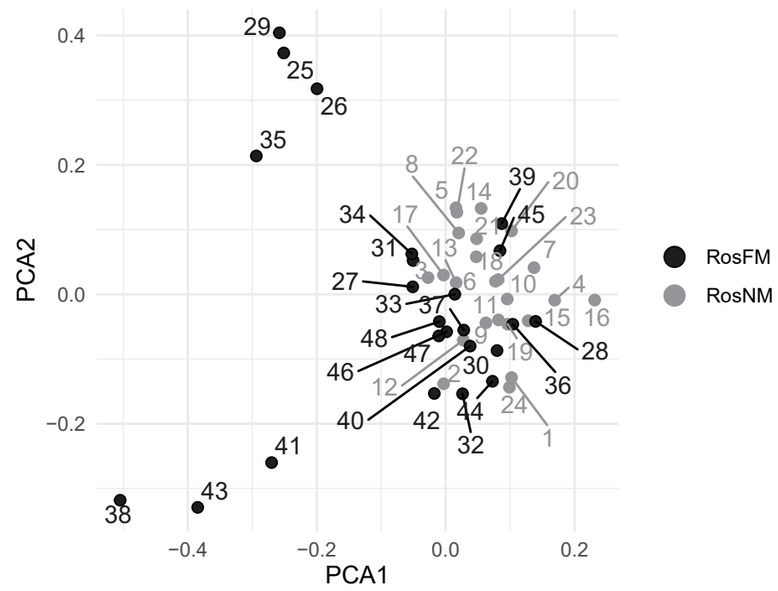
BIN 19400000_19900000



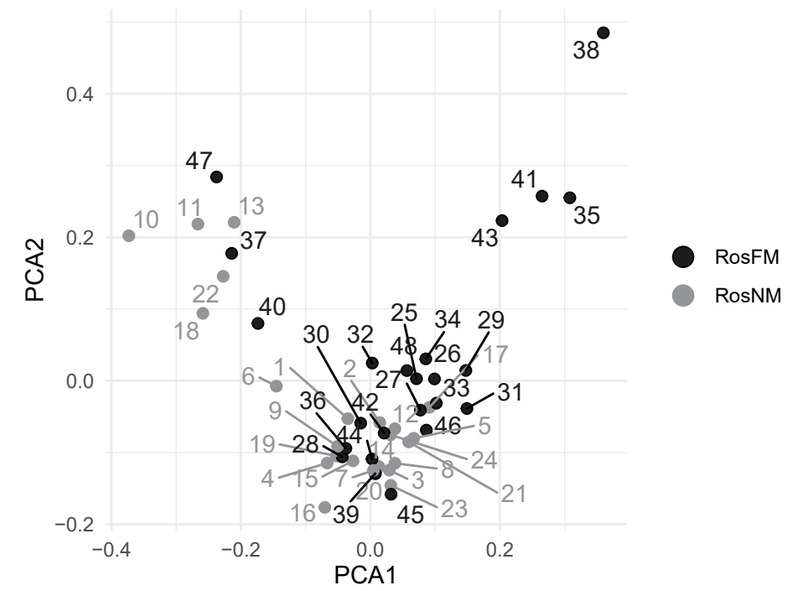
BIN 19500000_20000000



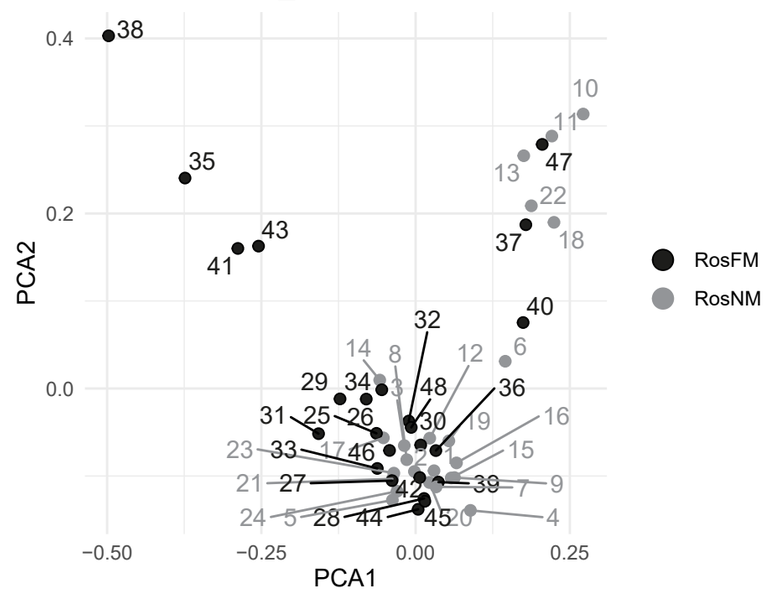
BIN 19600000_20100000



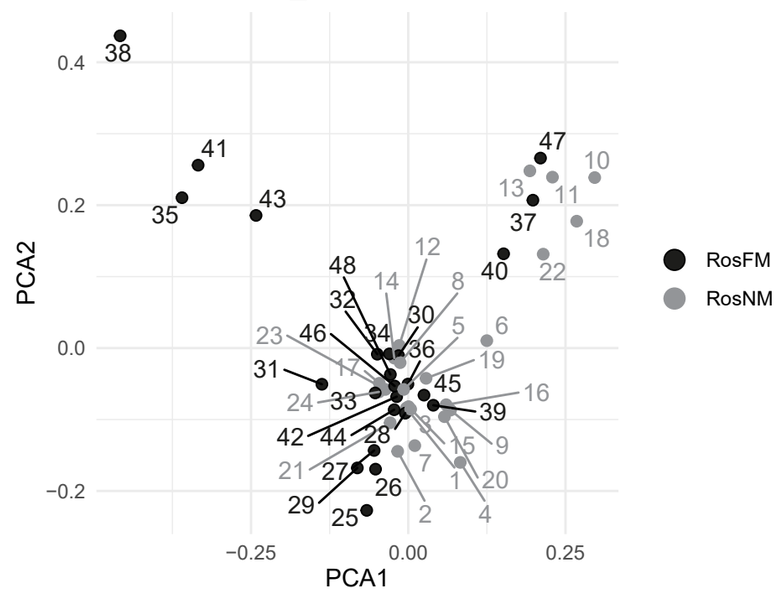
BIN 19700000_20200000



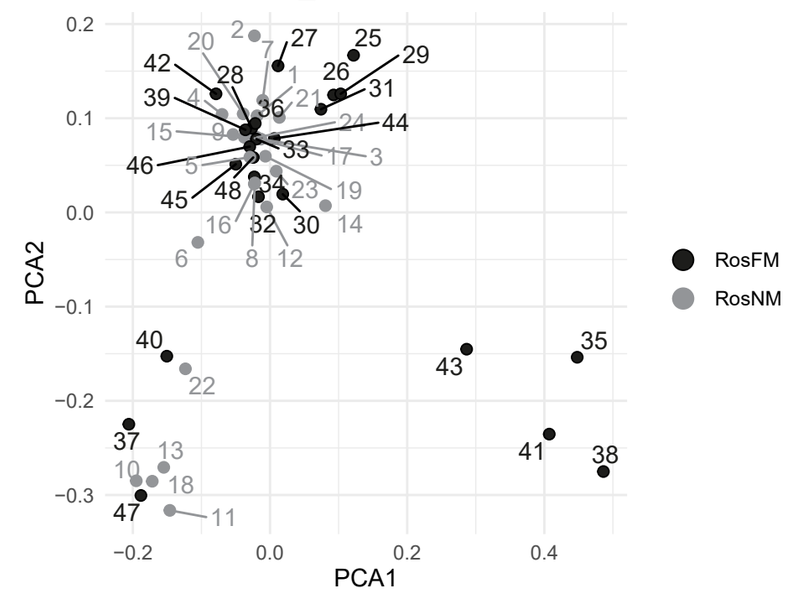
BIN 19800000_20300000



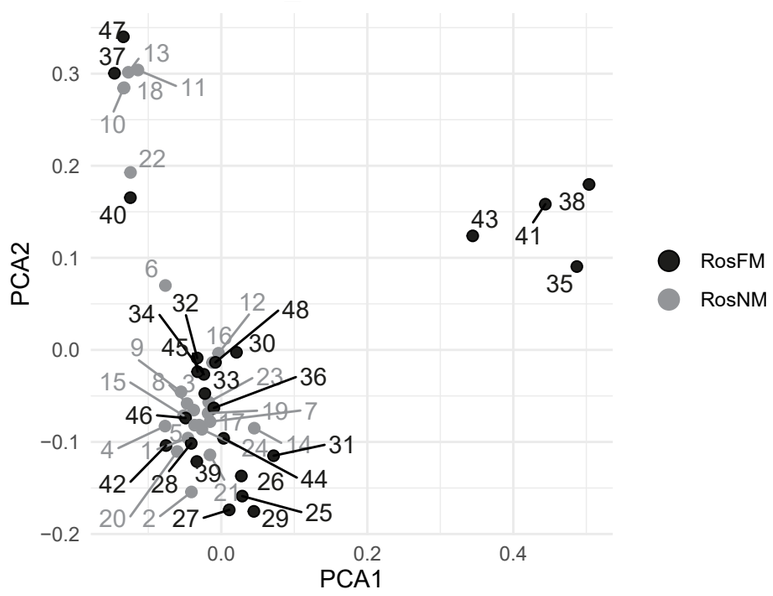
BIN 19900000_20400000



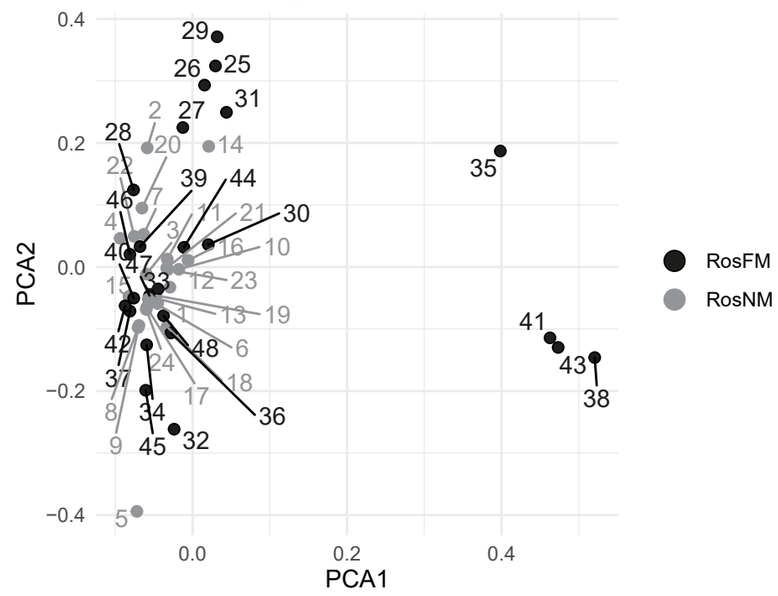
BIN 20000000_20500000



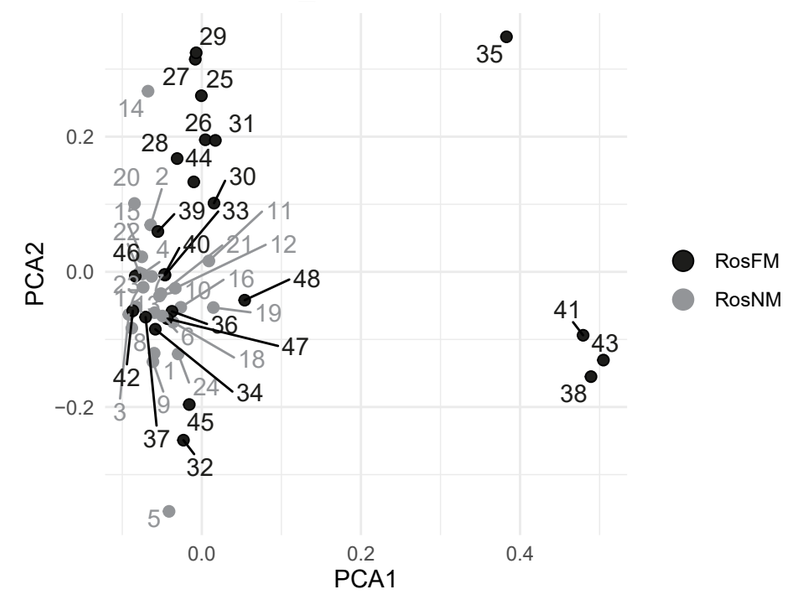
BIN 20100000_20600000



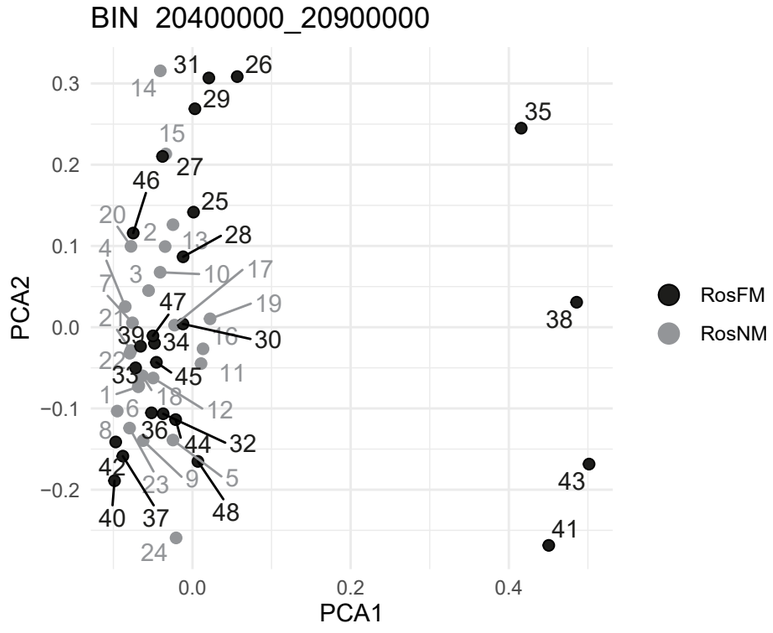
BIN 20200000_20700000



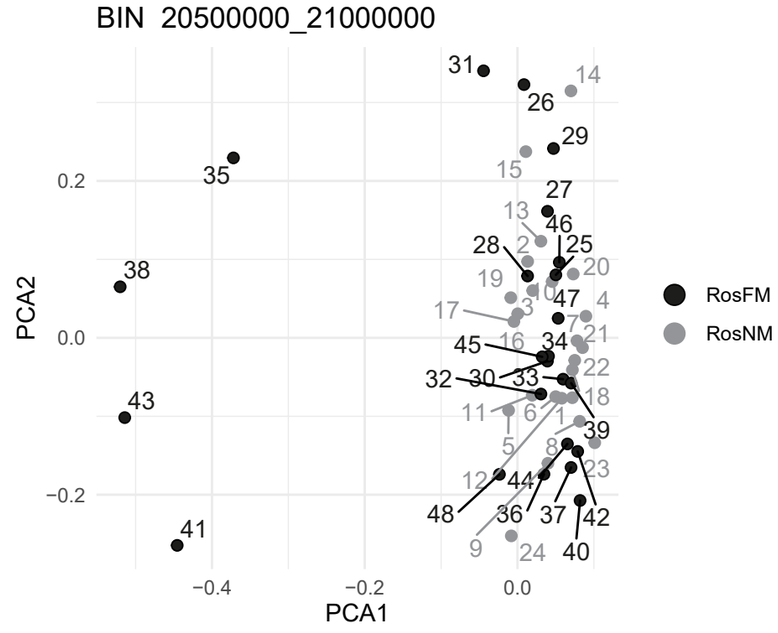
BIN 20300000_20800000



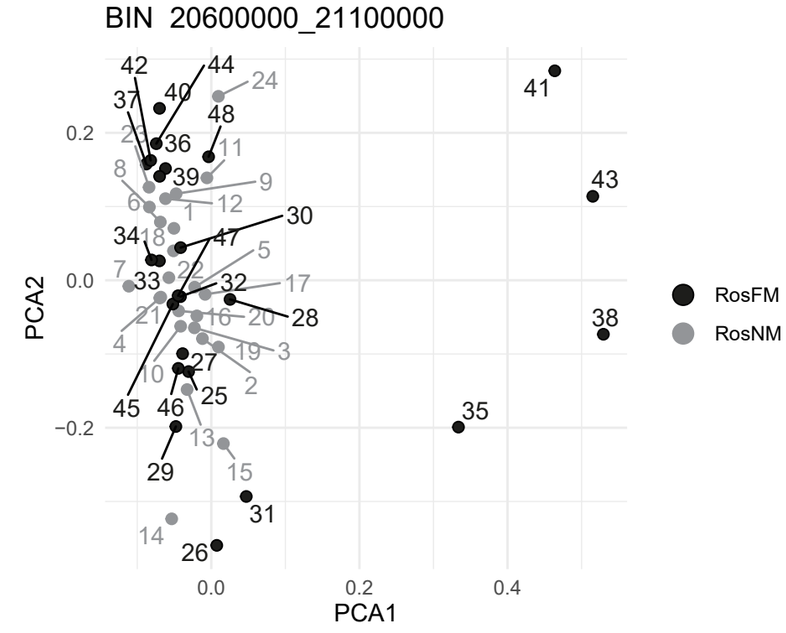
BIN 20400000_20900000



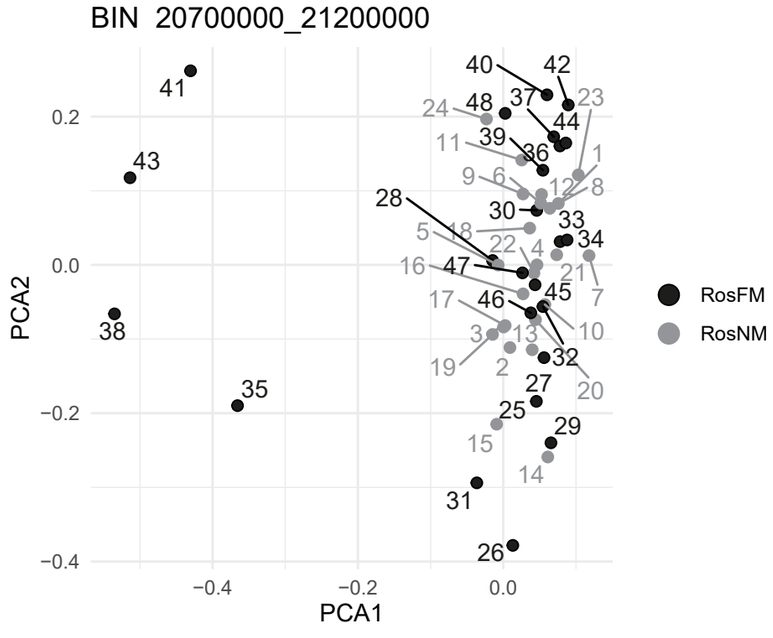
BIN 20500000_21000000



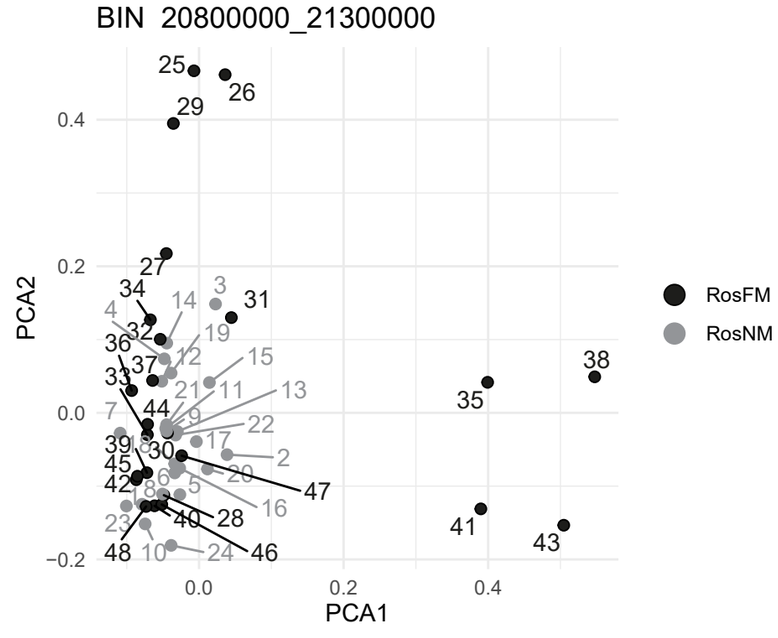
BIN 20600000_21100000



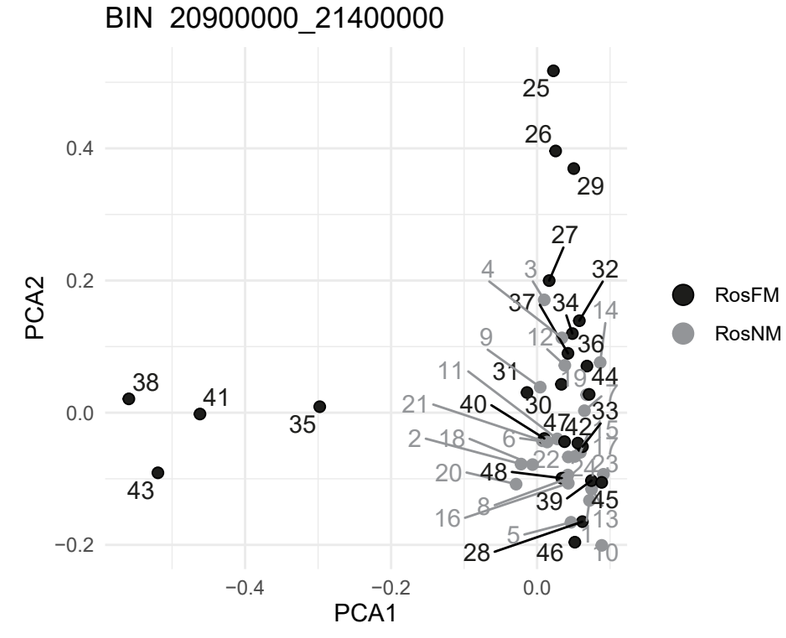
BIN 20700000_21200000



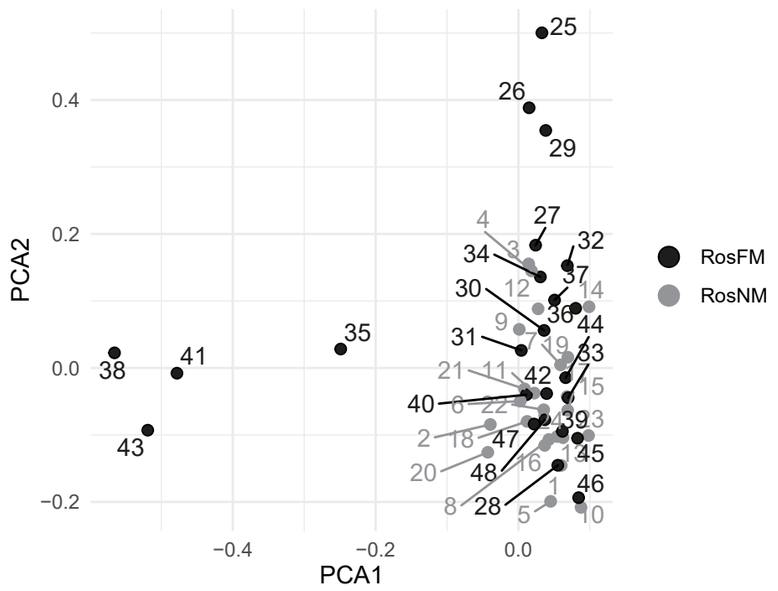
BIN 20800000_21300000



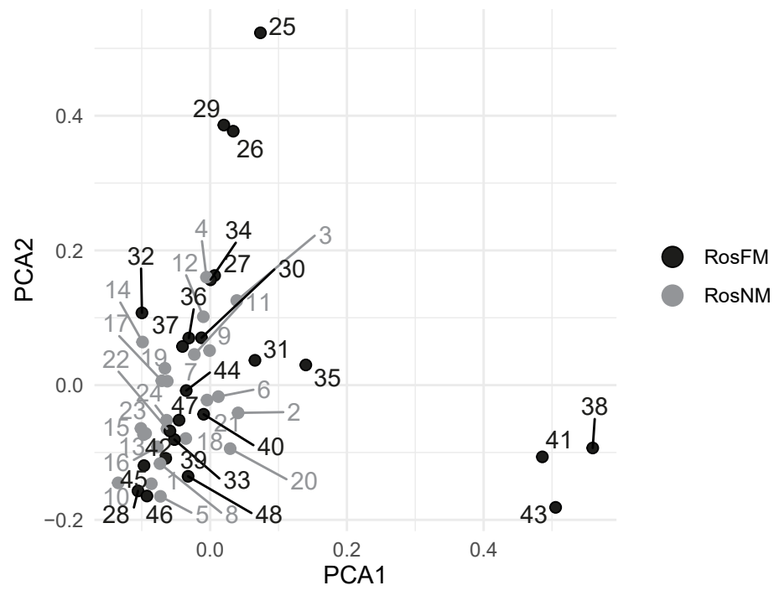
BIN 20900000_21400000



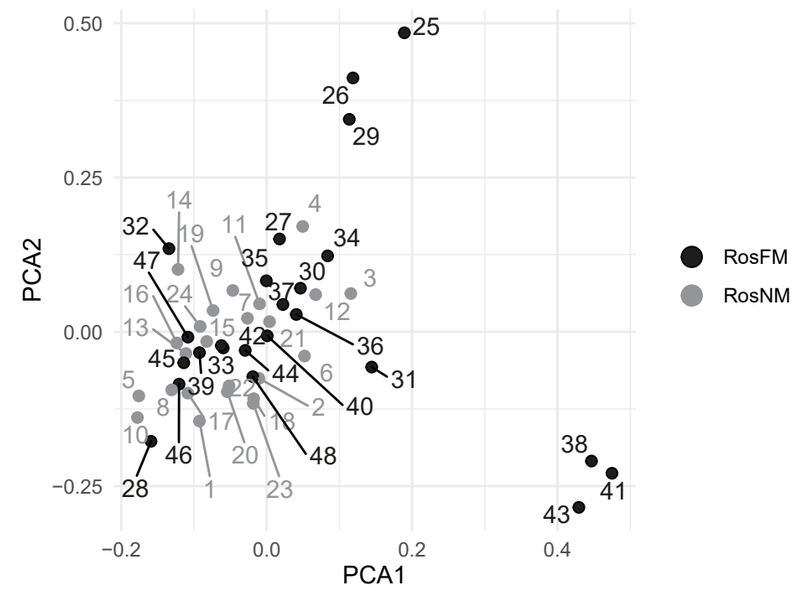
BIN 21000000_21500000



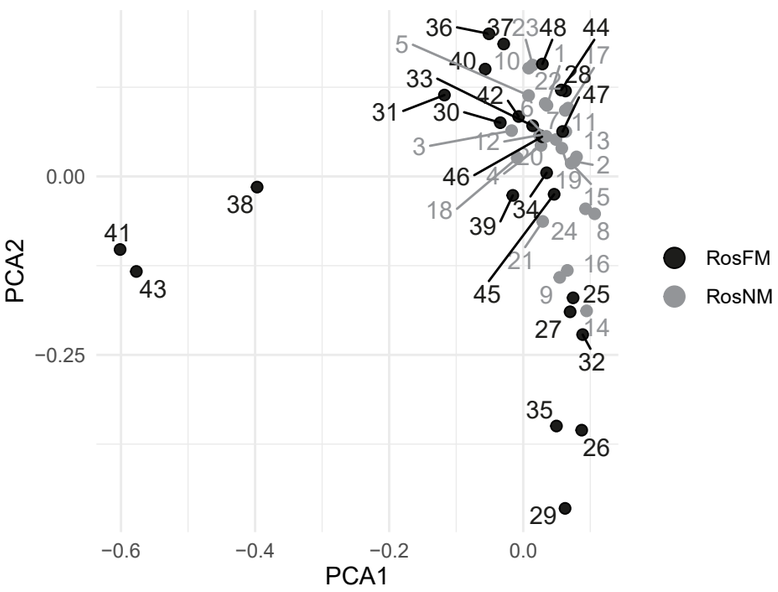
BIN 21100000_21600000



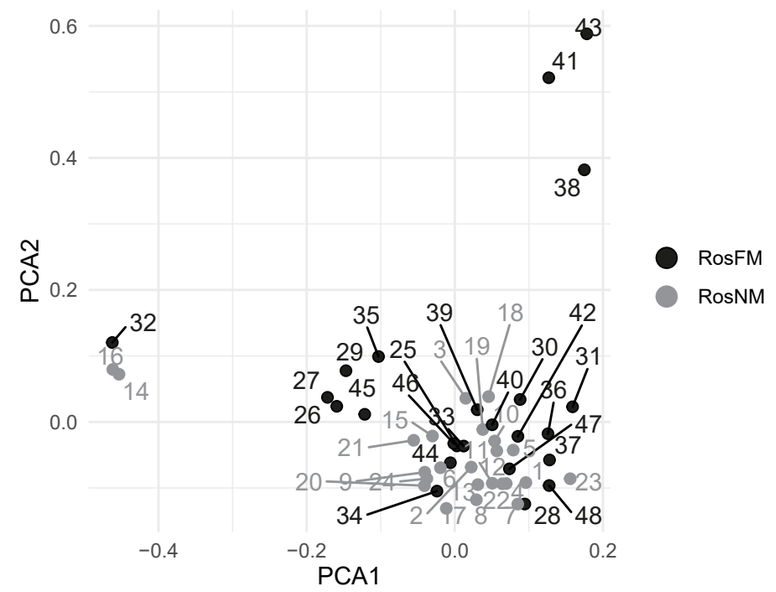
BIN 21200000_21700000



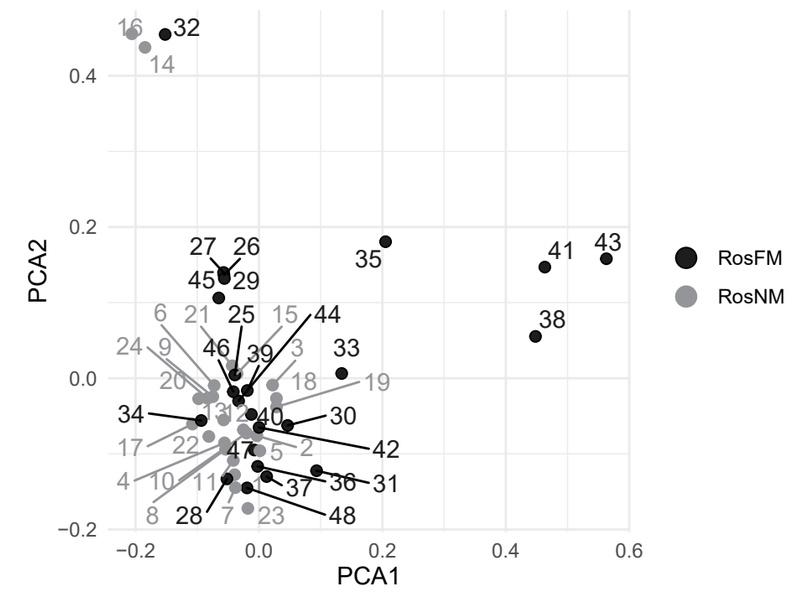
BIN 21300000_21800000



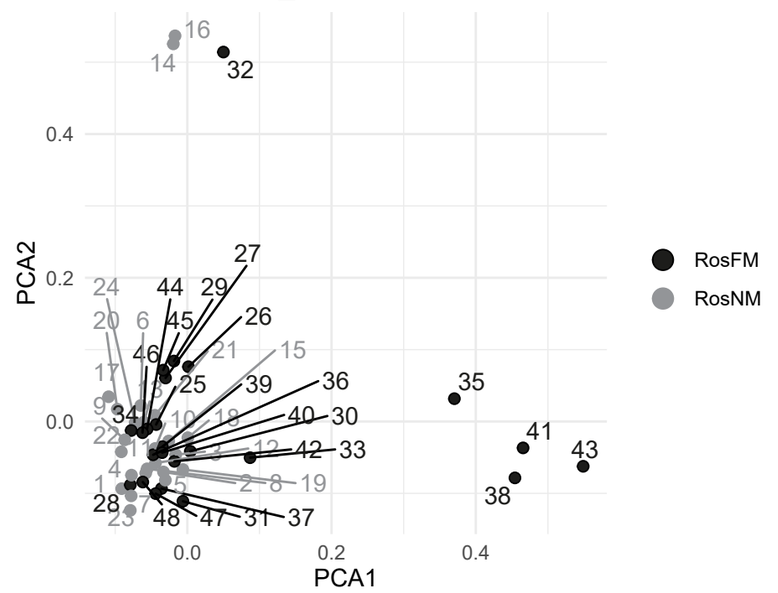
BIN 21400000_21900000



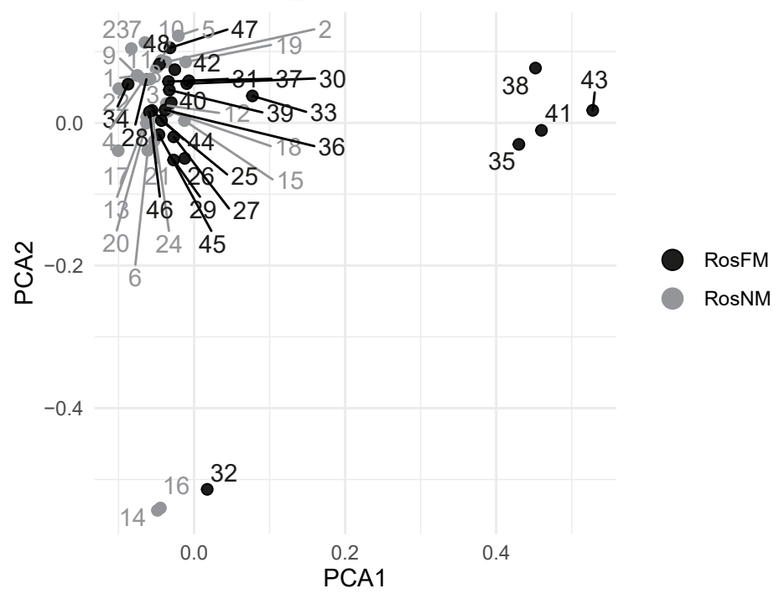
BIN 21500000_22000000



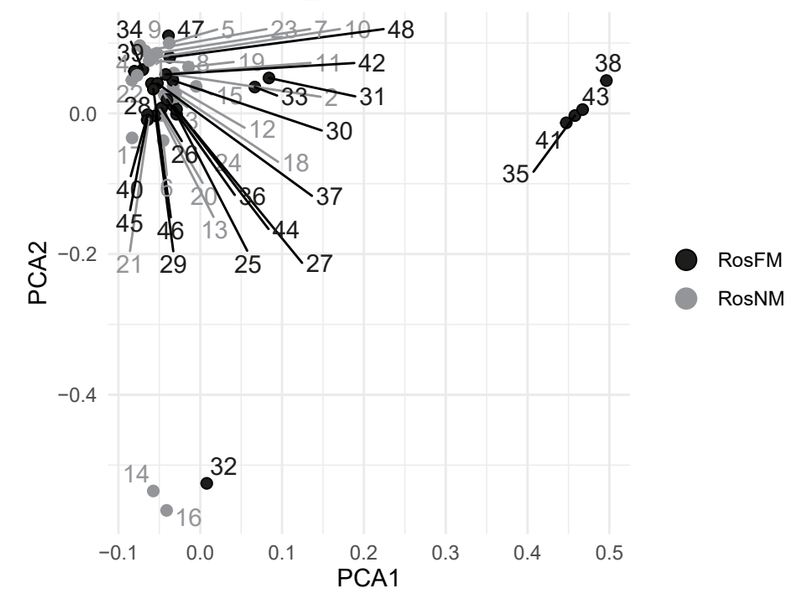
BIN 21600000_22100000



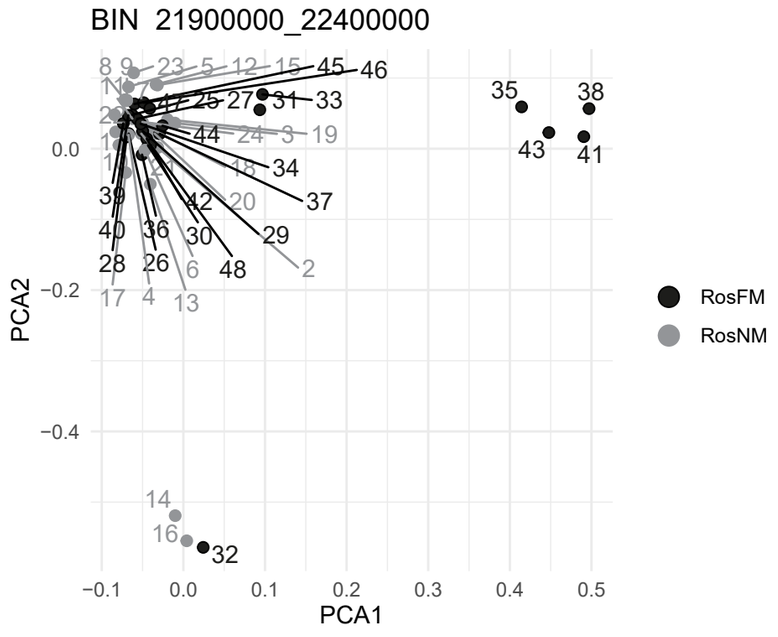
BIN 21700000_22200000



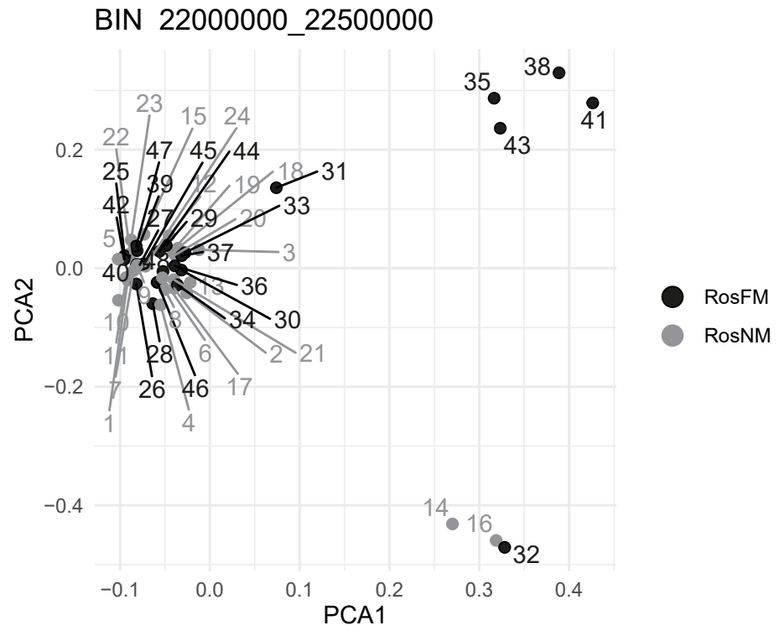
BIN 21800000_22300000



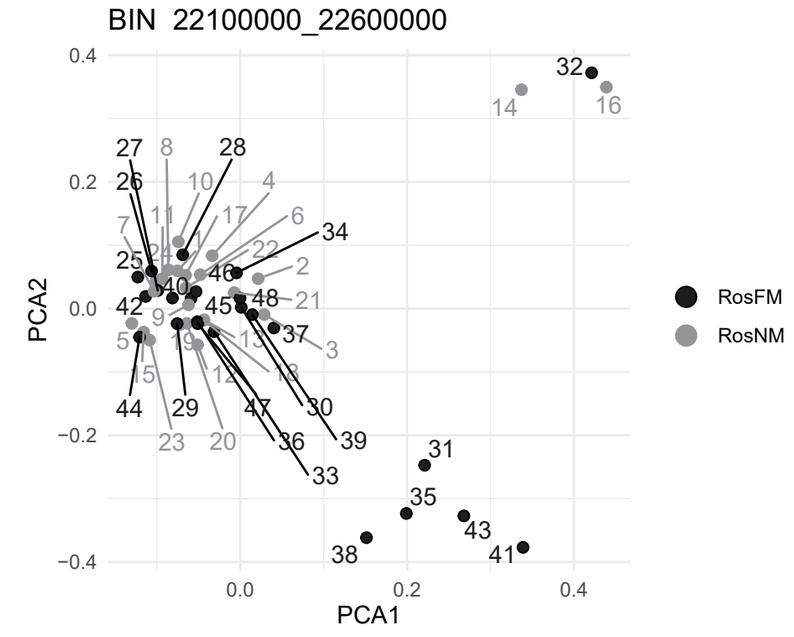
BIN 21900000_22400000



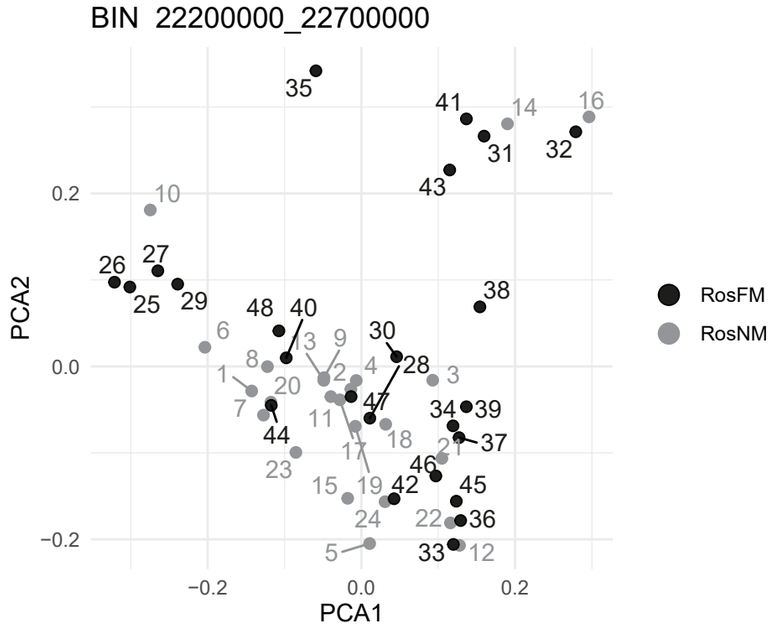
BIN 22000000_22500000



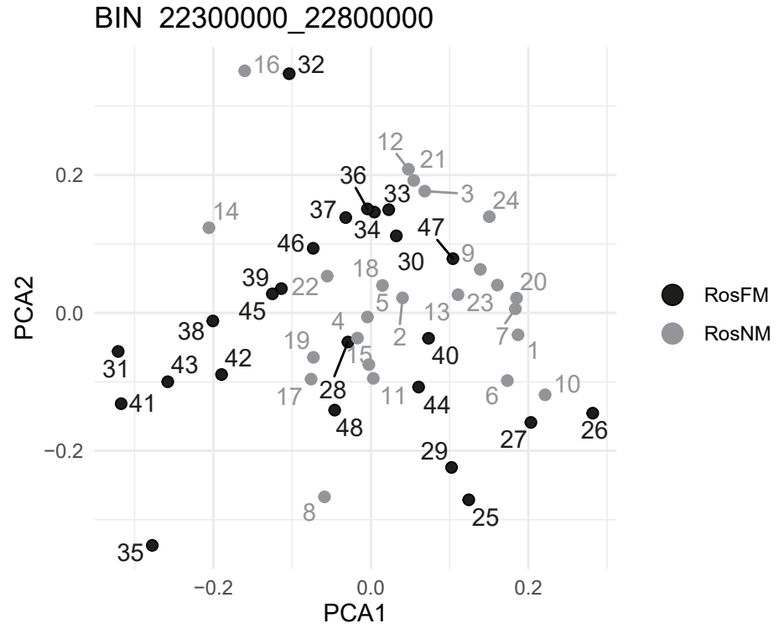
BIN 22100000_22600000



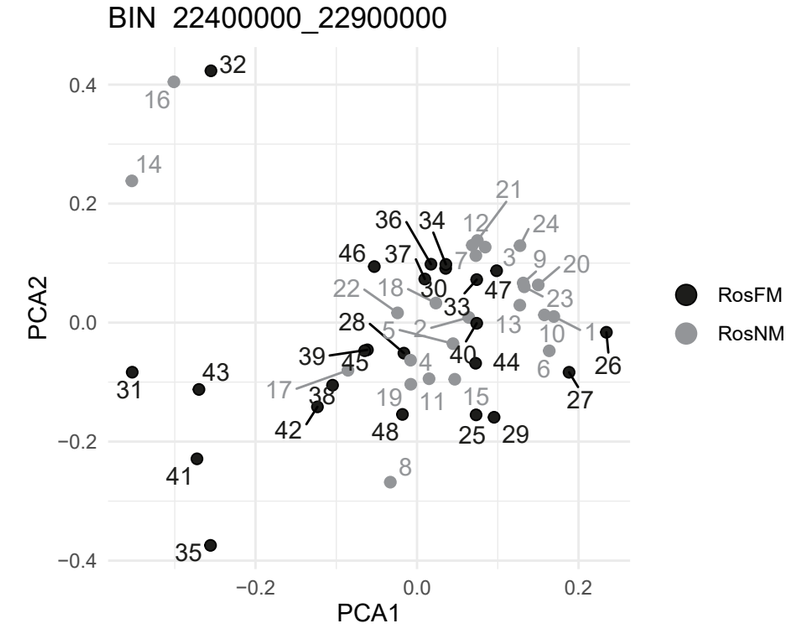
BIN 22200000_22700000



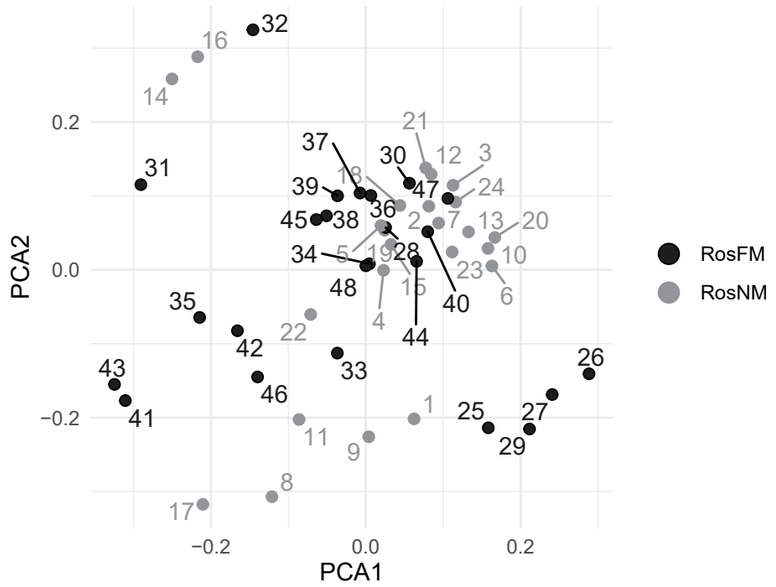
BIN 22300000_22800000



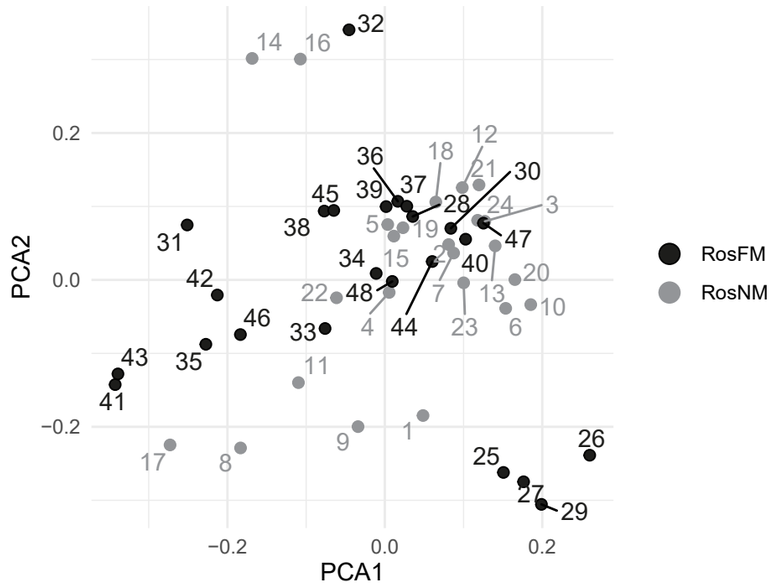
BIN 22400000_22900000



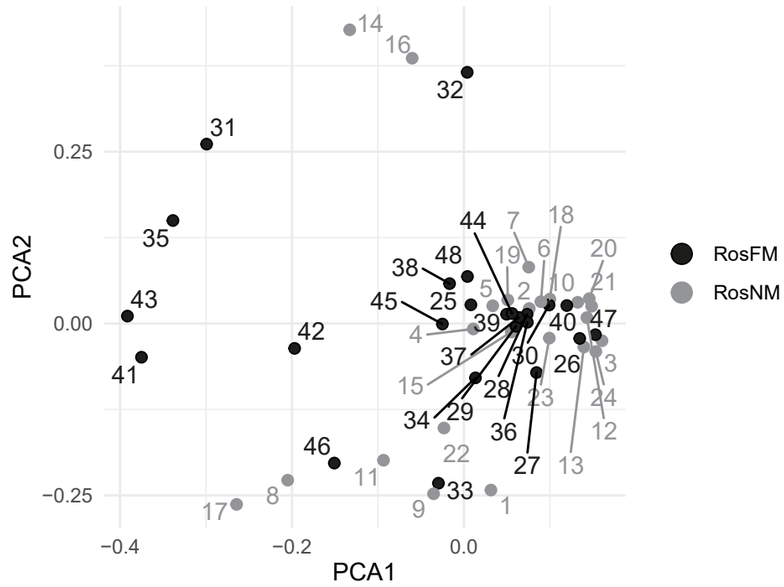
BIN 22500000_23000000



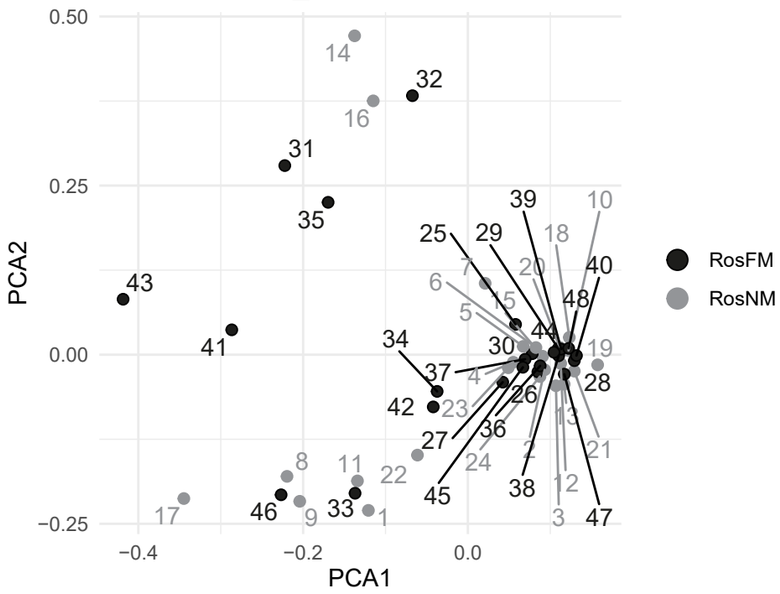
BIN 22600000_23100000



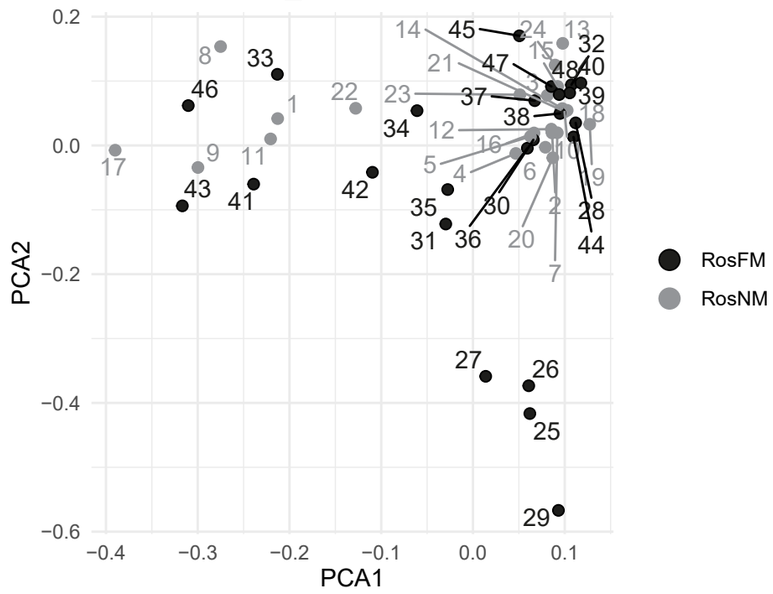
BIN 22700000_23200000



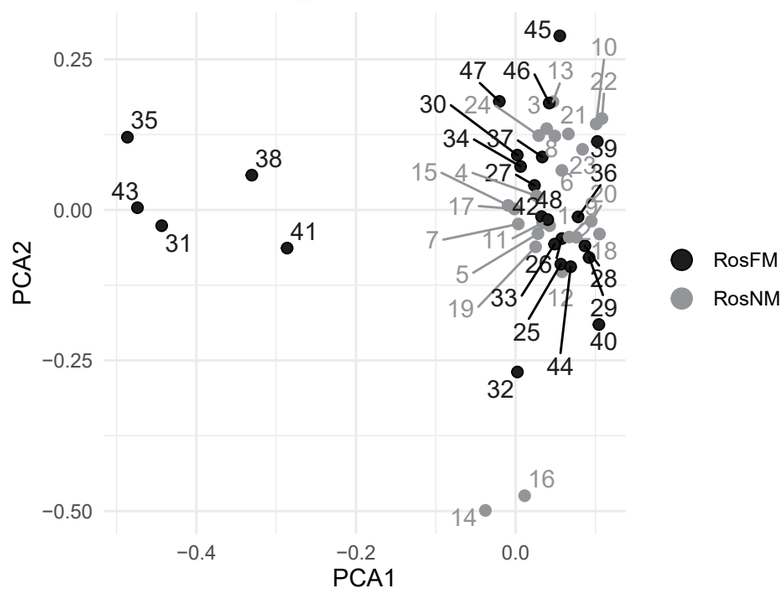
BIN 22800000_23300000



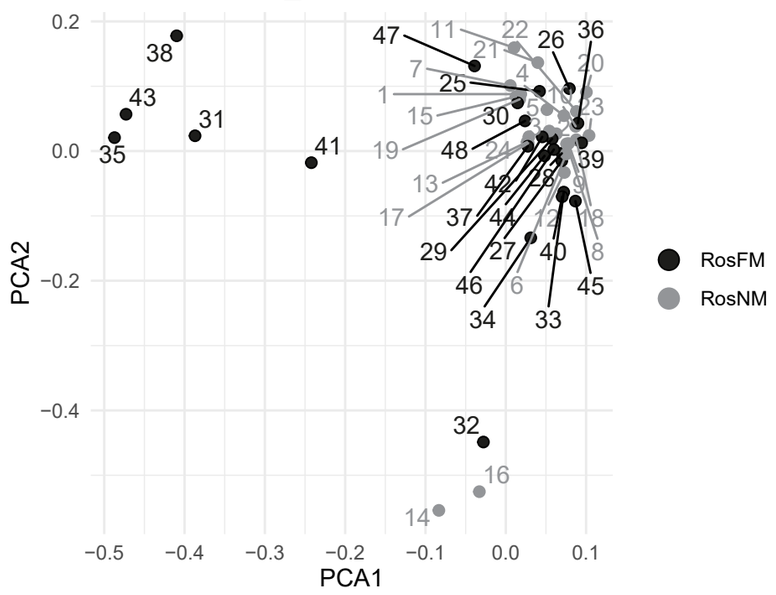
BIN 22900000_23400000



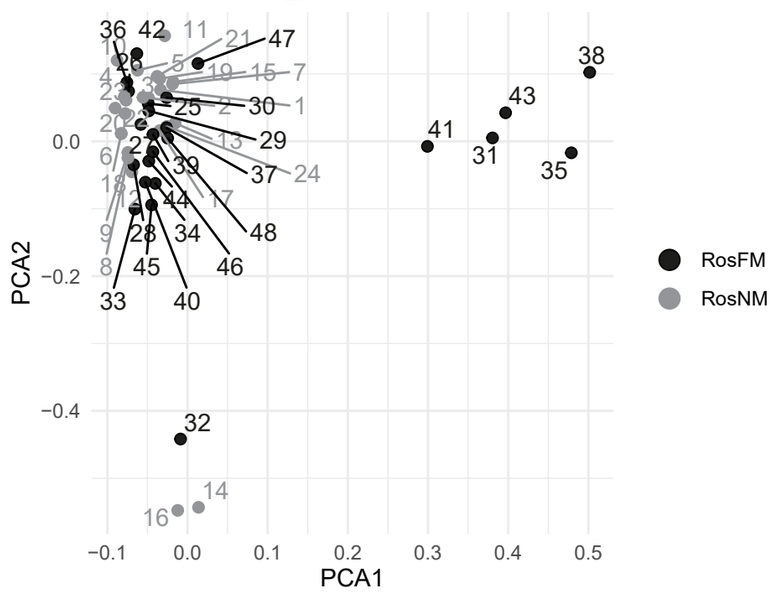
BIN 23000000_23500000



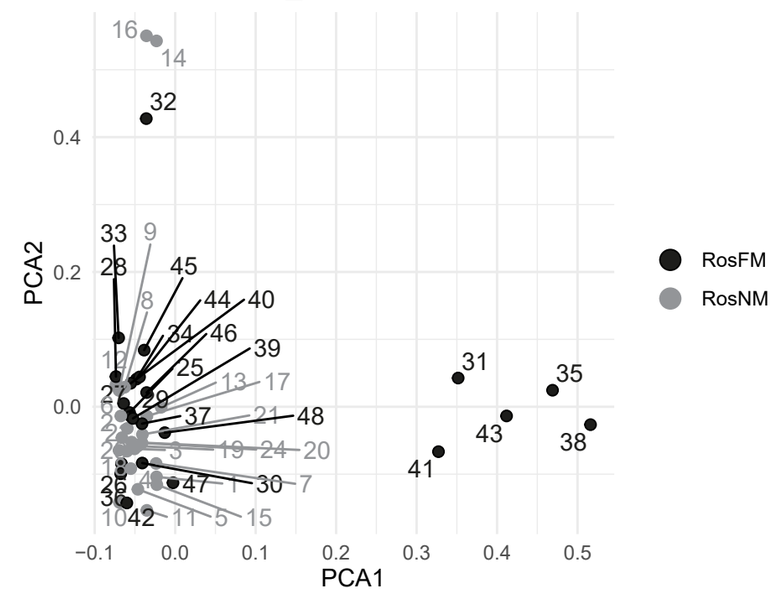
BIN 23100000_23600000



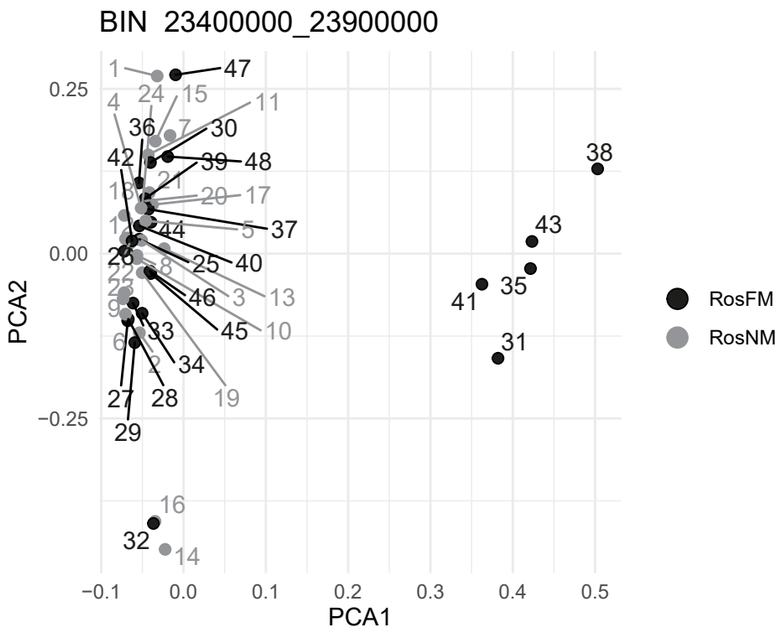
BIN 23200000_23700000



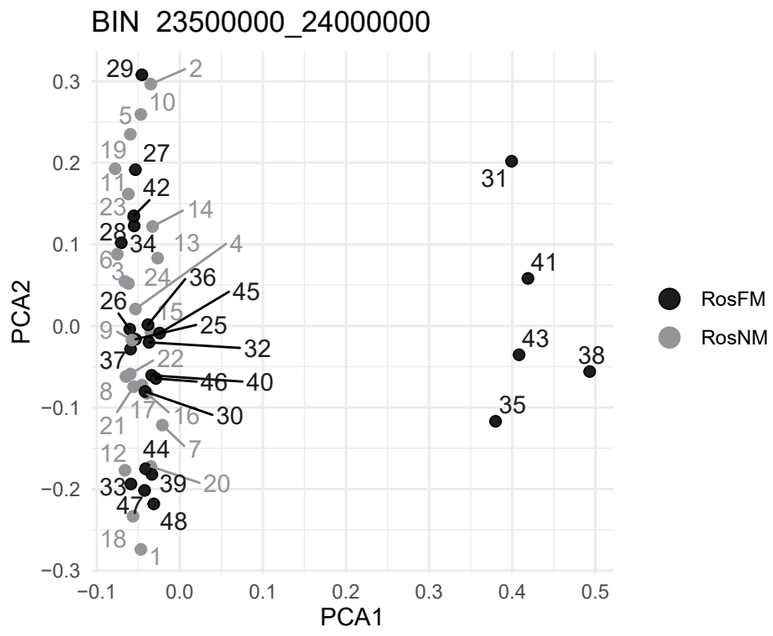
BIN 23300000_23800000



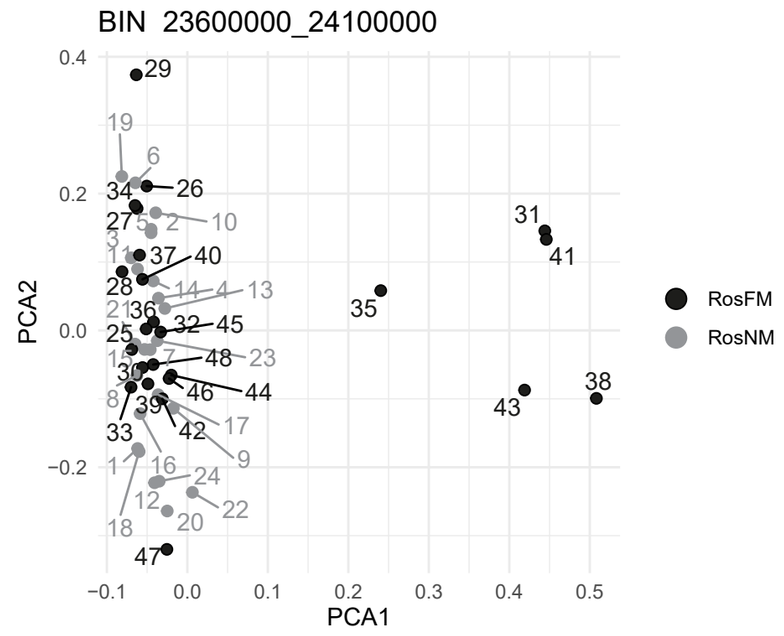
BIN 23400000_23900000



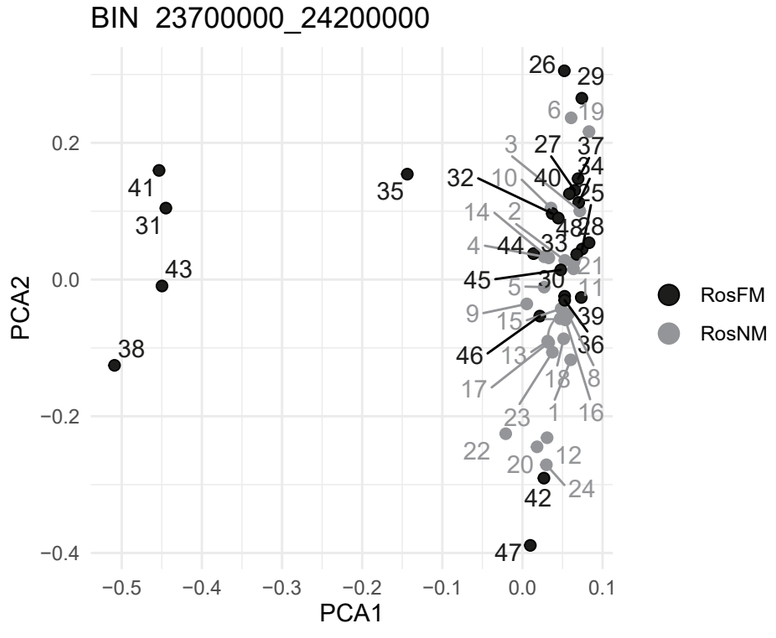
BIN 23500000_24000000



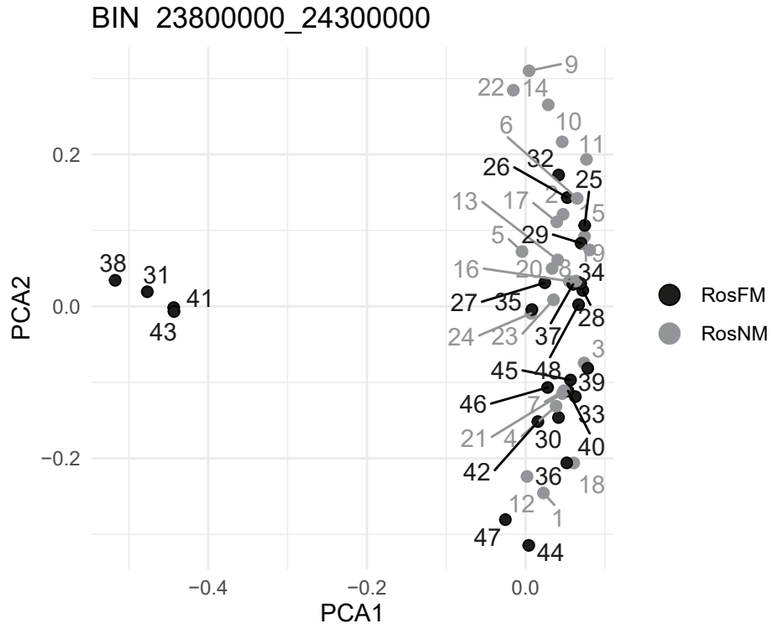
BIN 23600000_24100000



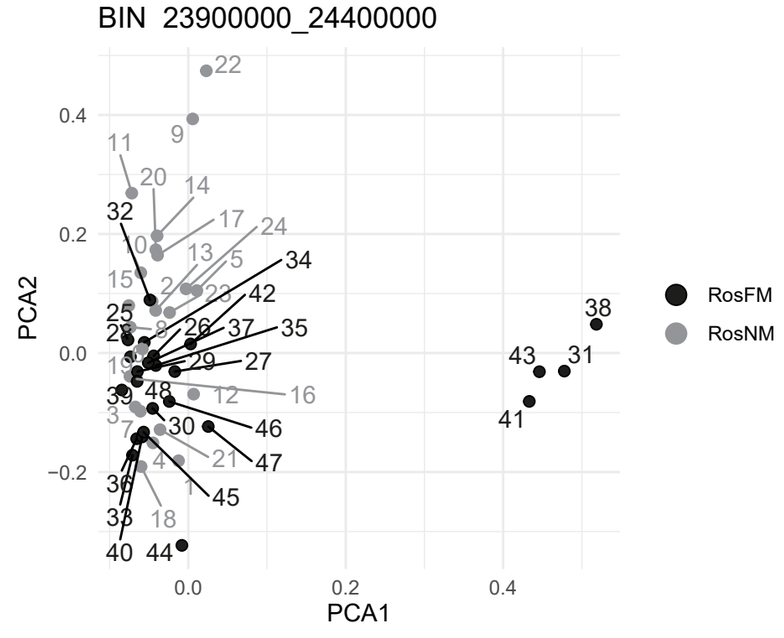
BIN 23700000_24200000



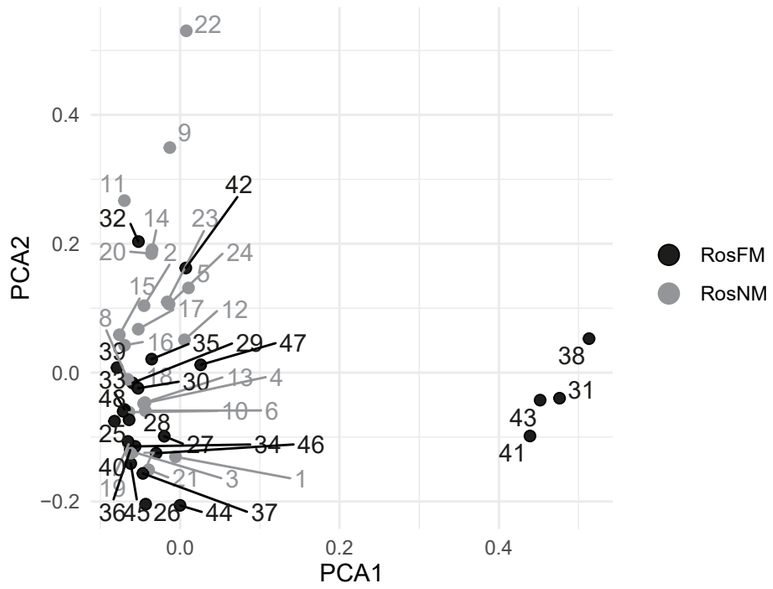
BIN 23800000_24300000



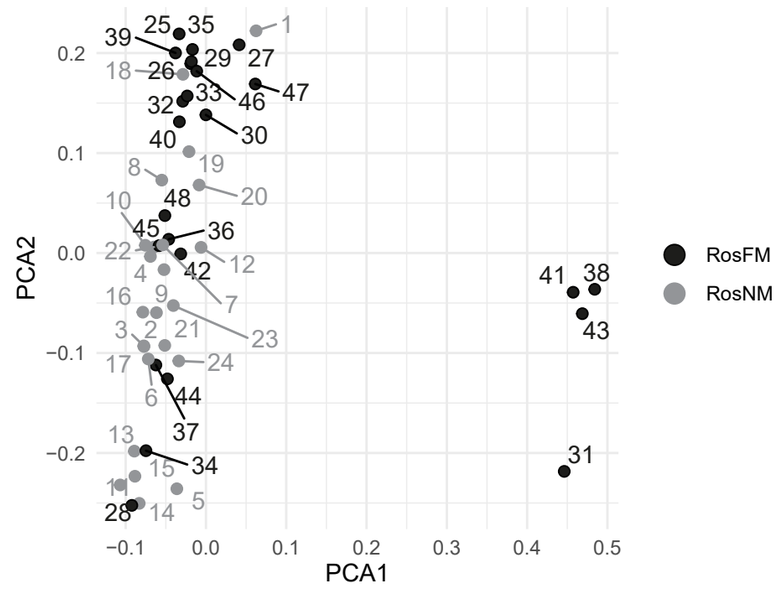
BIN 23900000_24400000



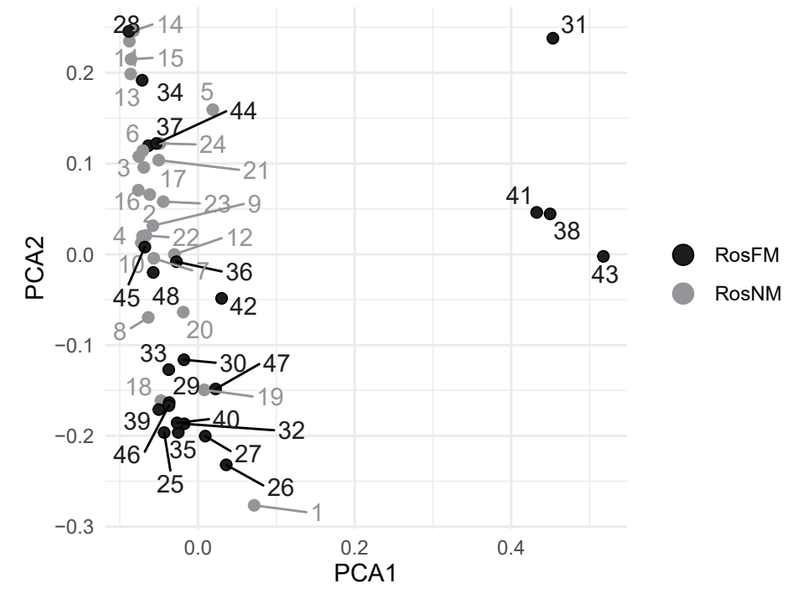
BIN 24000000_24500000



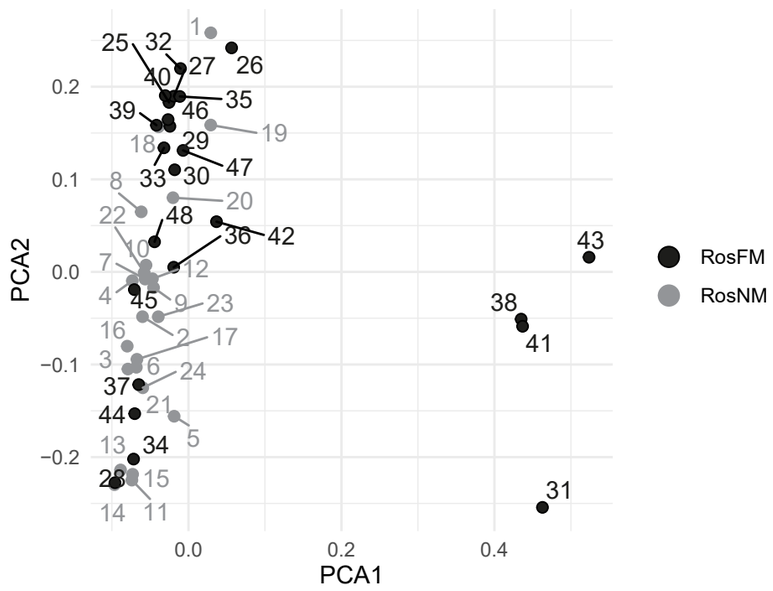
BIN 24100000_24600000



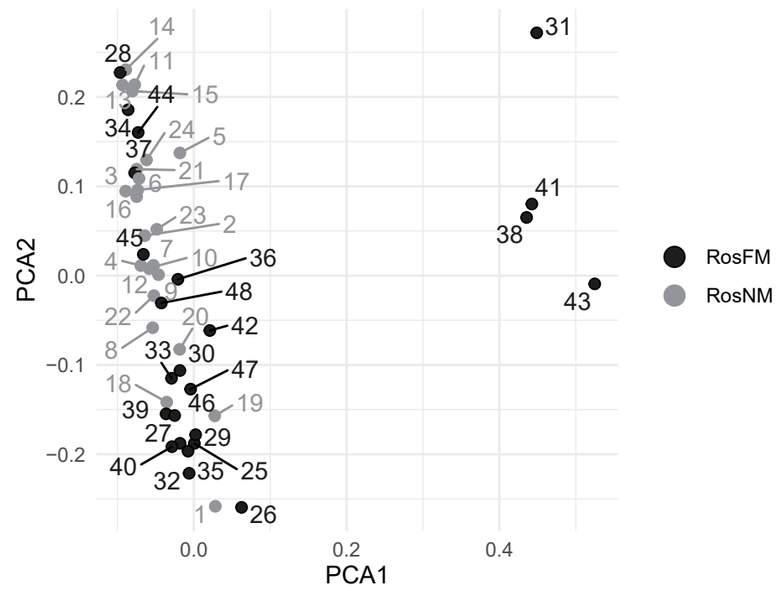
BIN 24200000_24700000



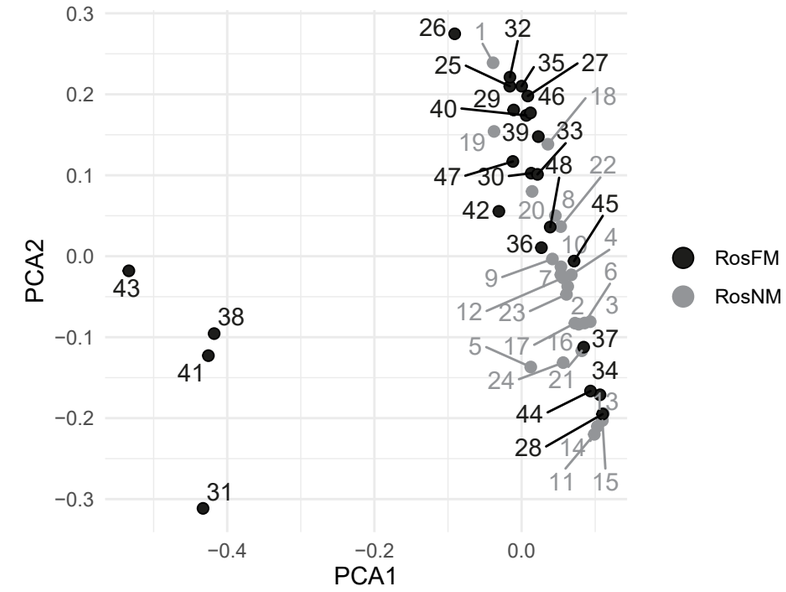
BIN 24300000_24800000



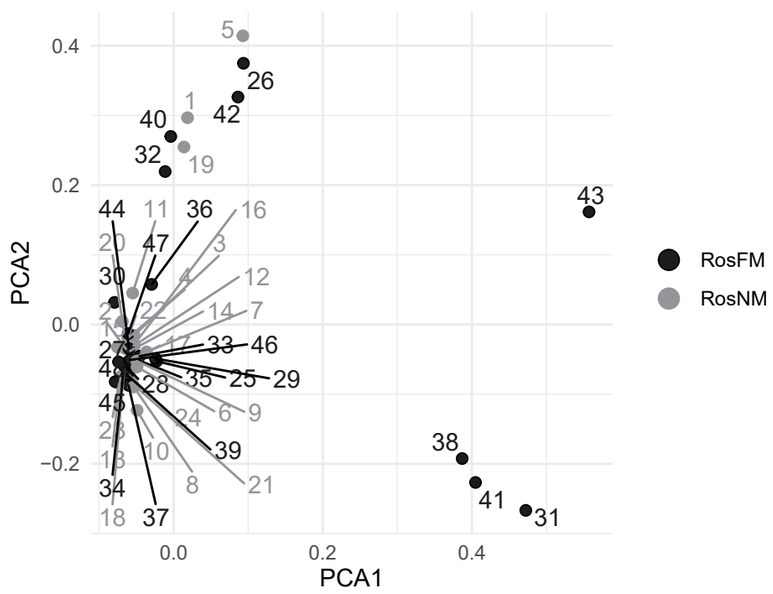
BIN 24400000_24900000



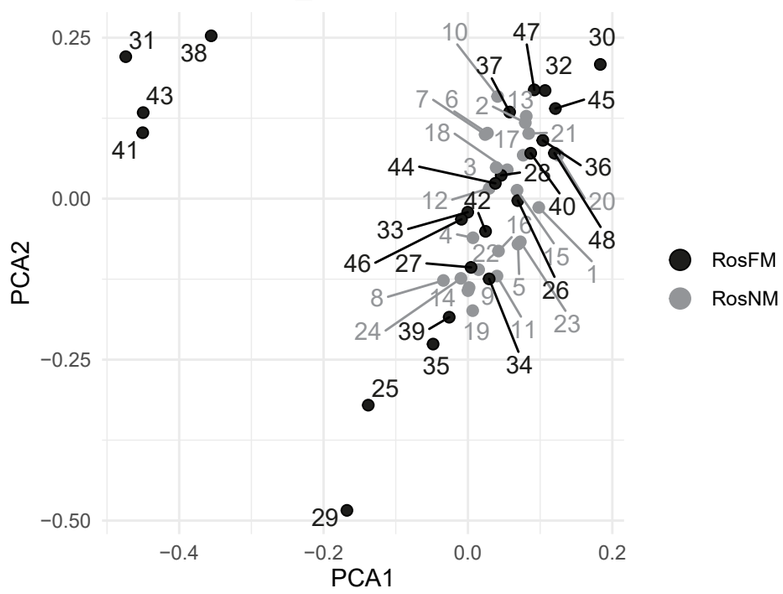
BIN 24500000_25000000



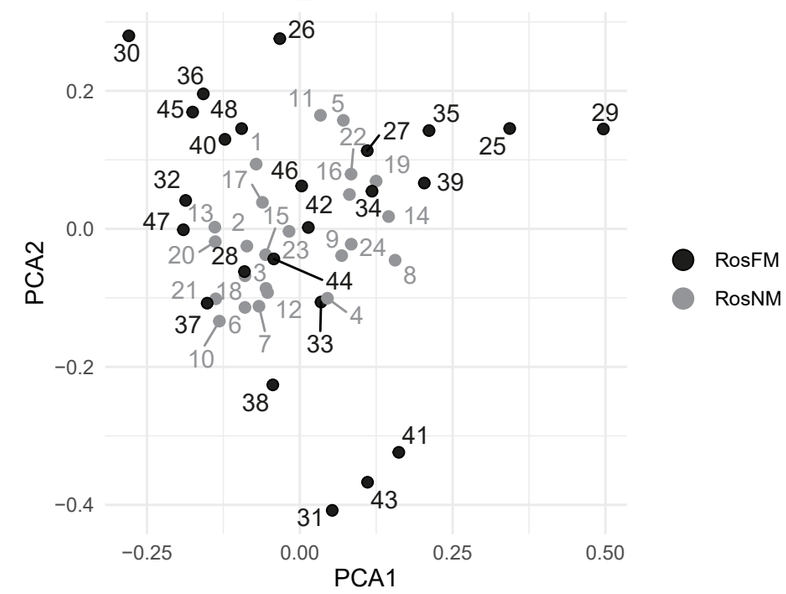
BIN 24600000_25100000



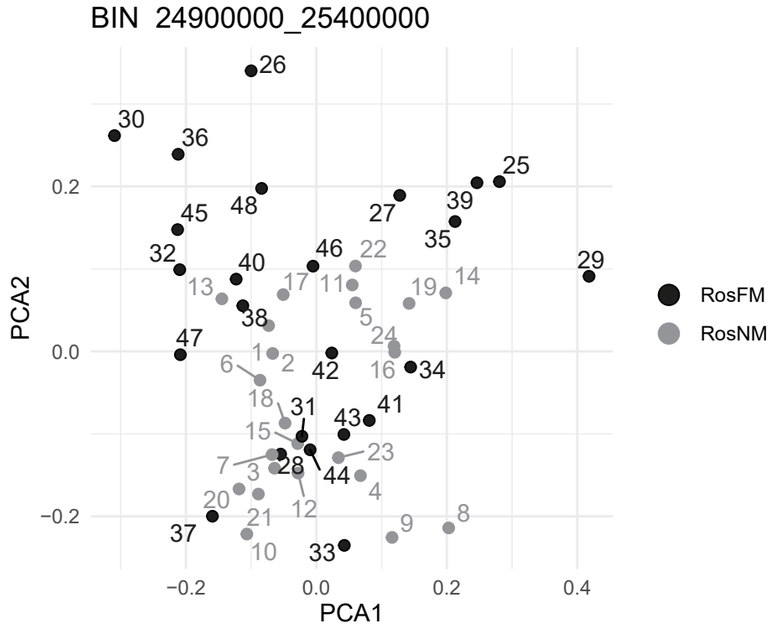
BIN 24700000_25200000



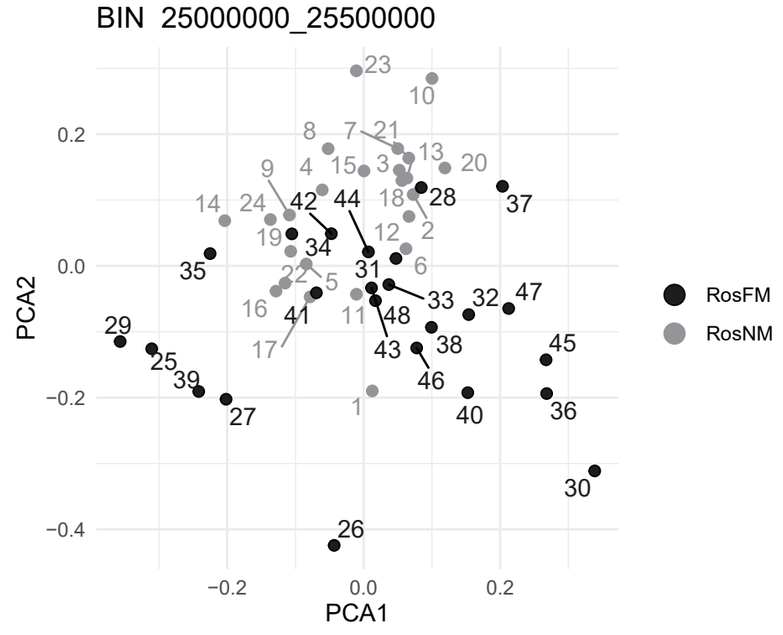
BIN 24800000_25300000



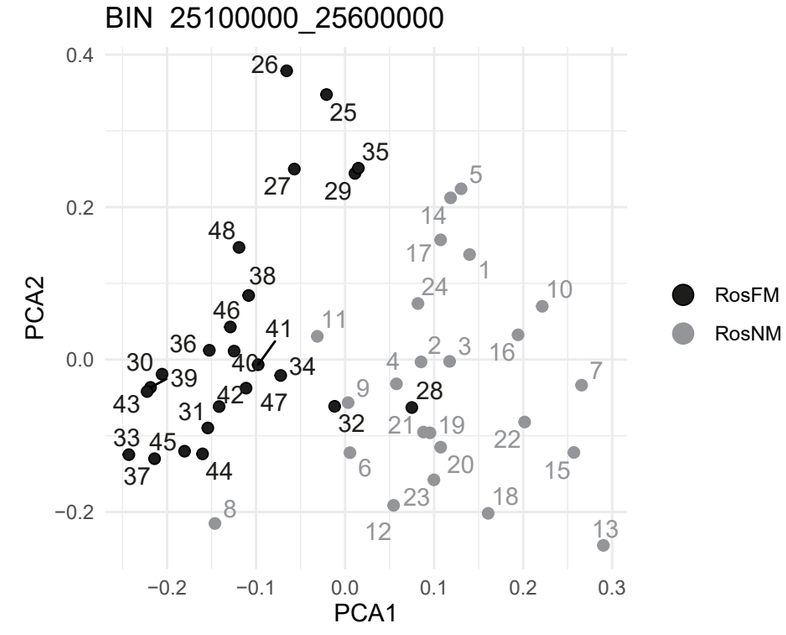
BIN 24900000_25400000



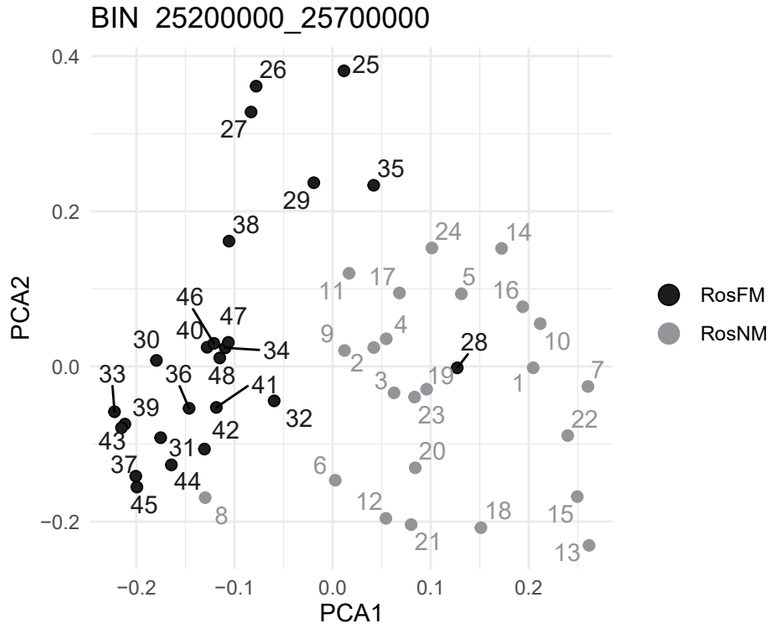
BIN 25000000_25500000



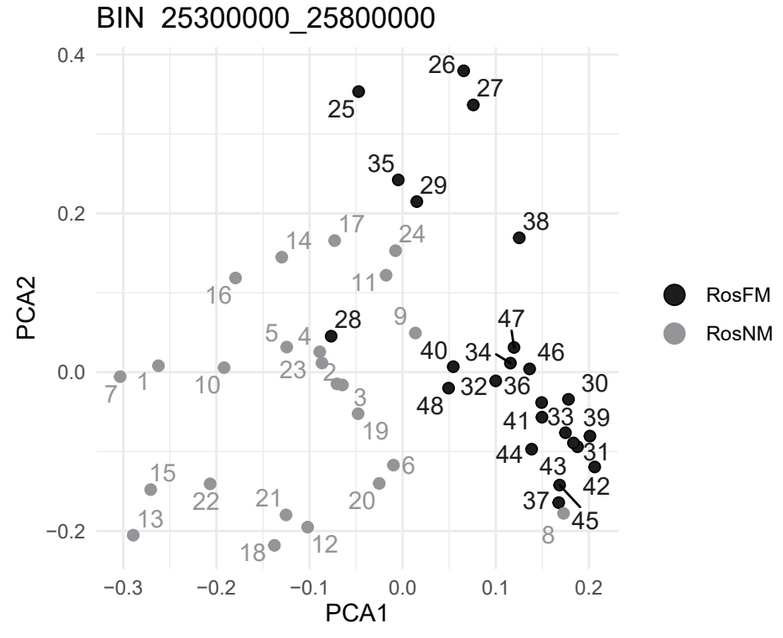
BIN 25100000_25600000



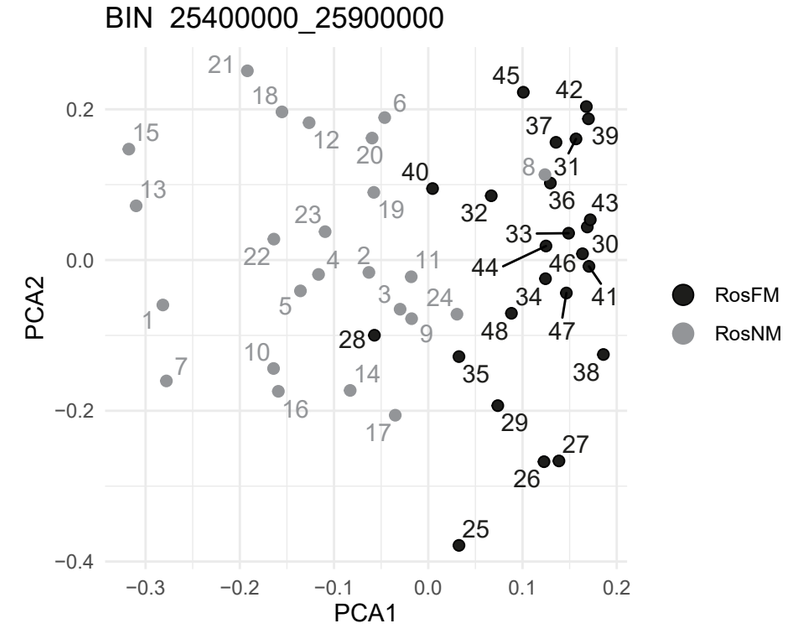
BIN 25200000_25700000



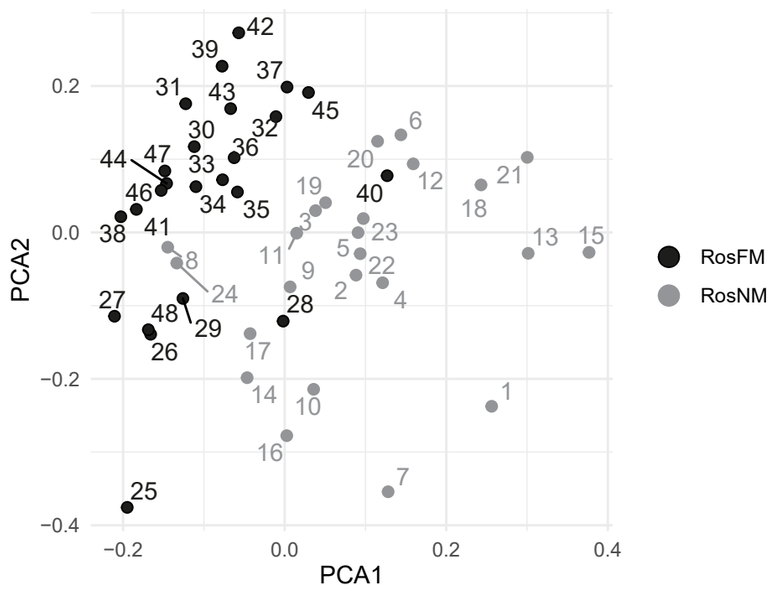
BIN 25300000_25800000



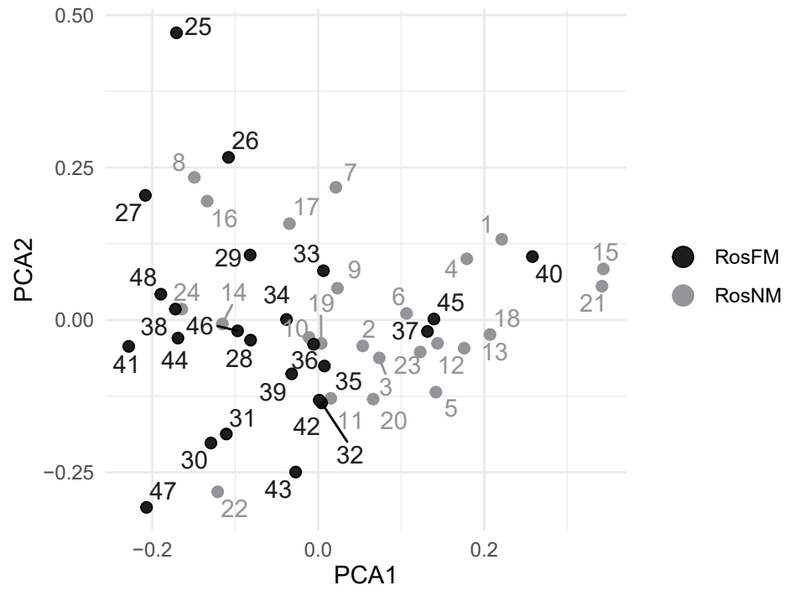
BIN 25400000_25900000



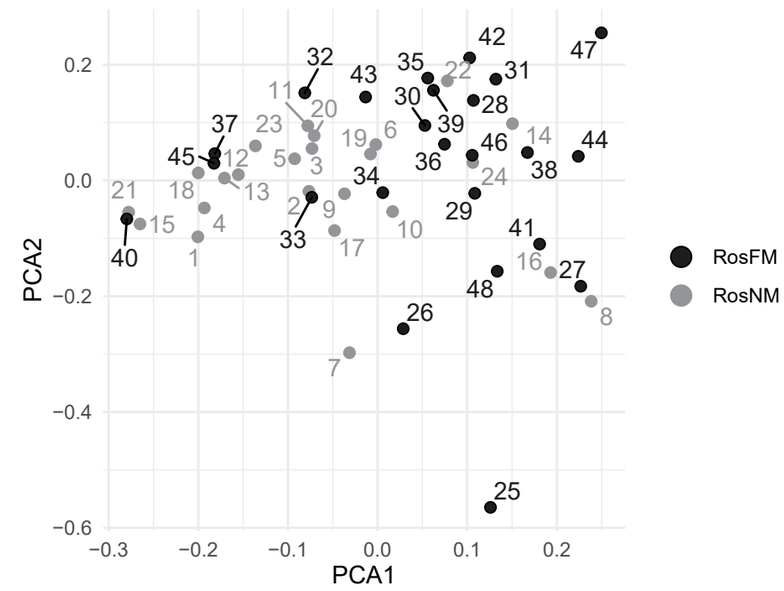
BIN 25500000_26000000



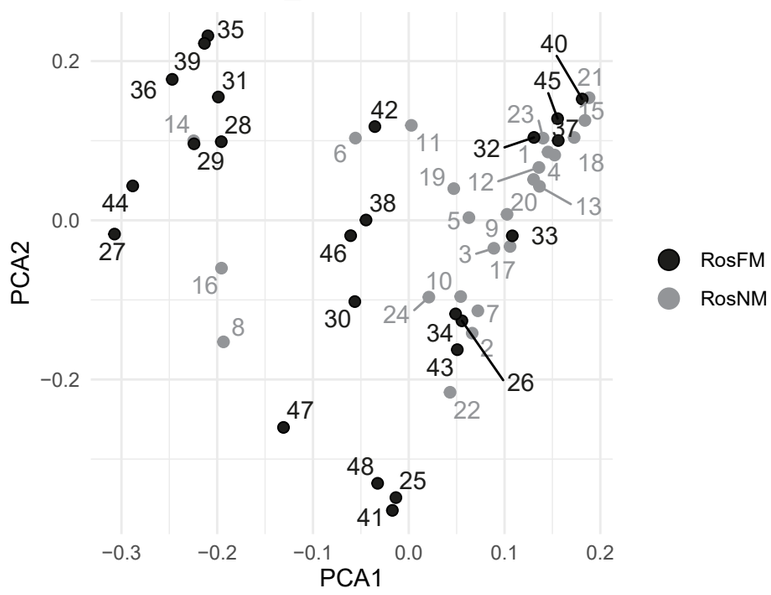
BIN 25600000_26100000



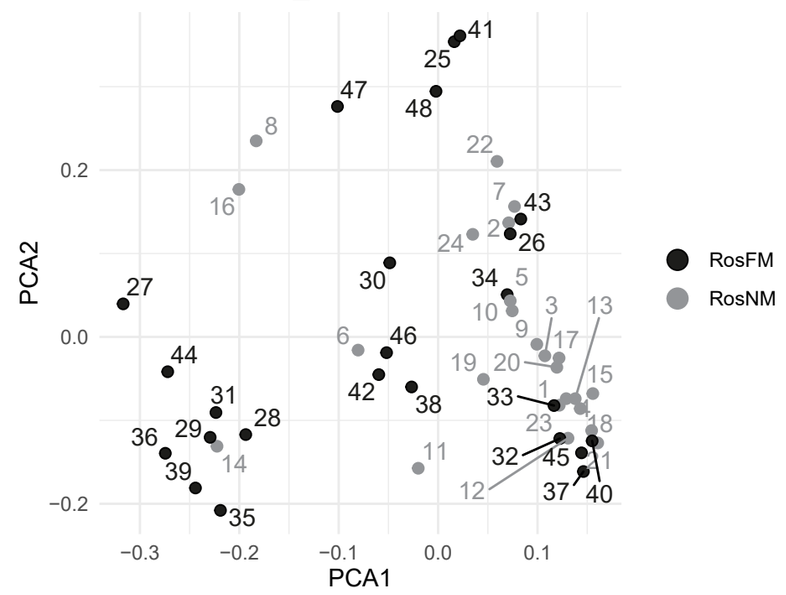
BIN 25700000_26200000



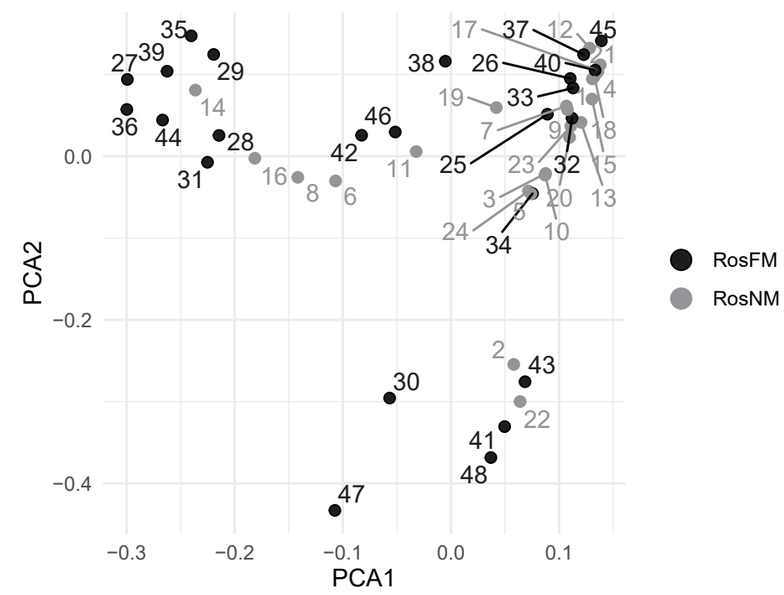
BIN 25800000_26300000



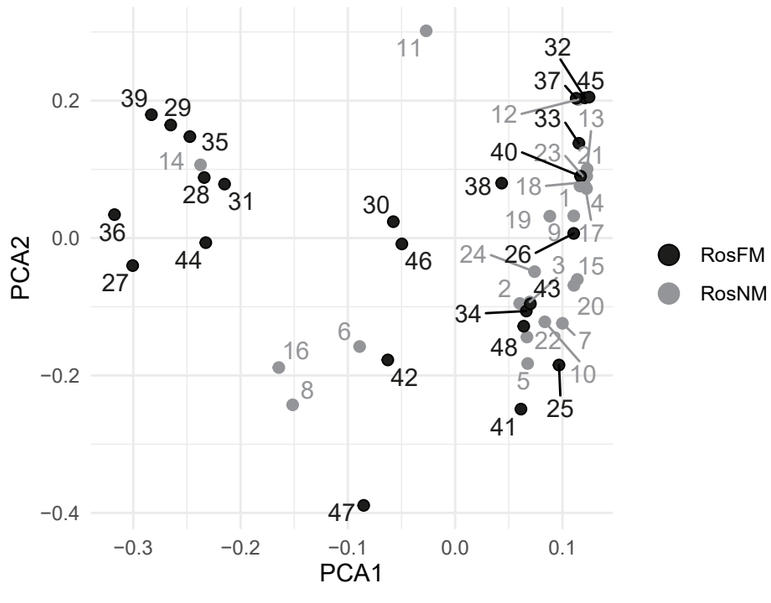
BIN 25900000_26400000



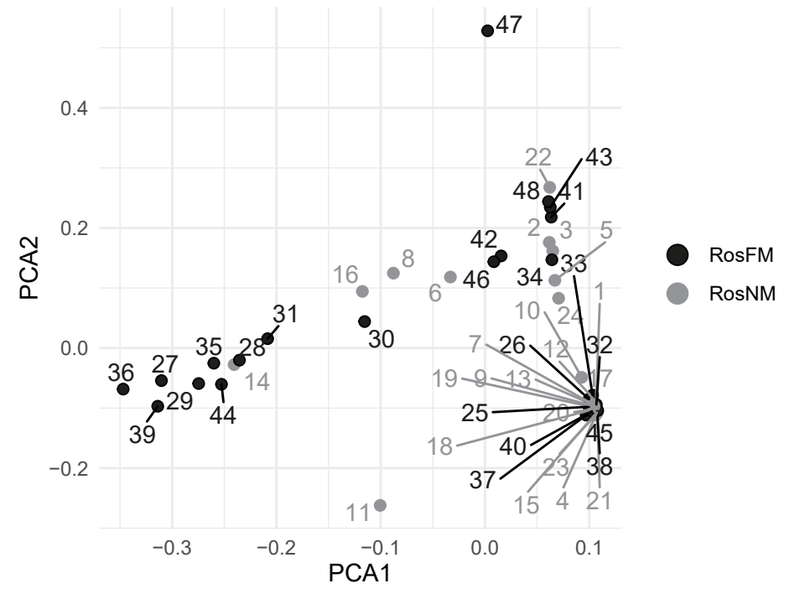
BIN 26000000_26500000

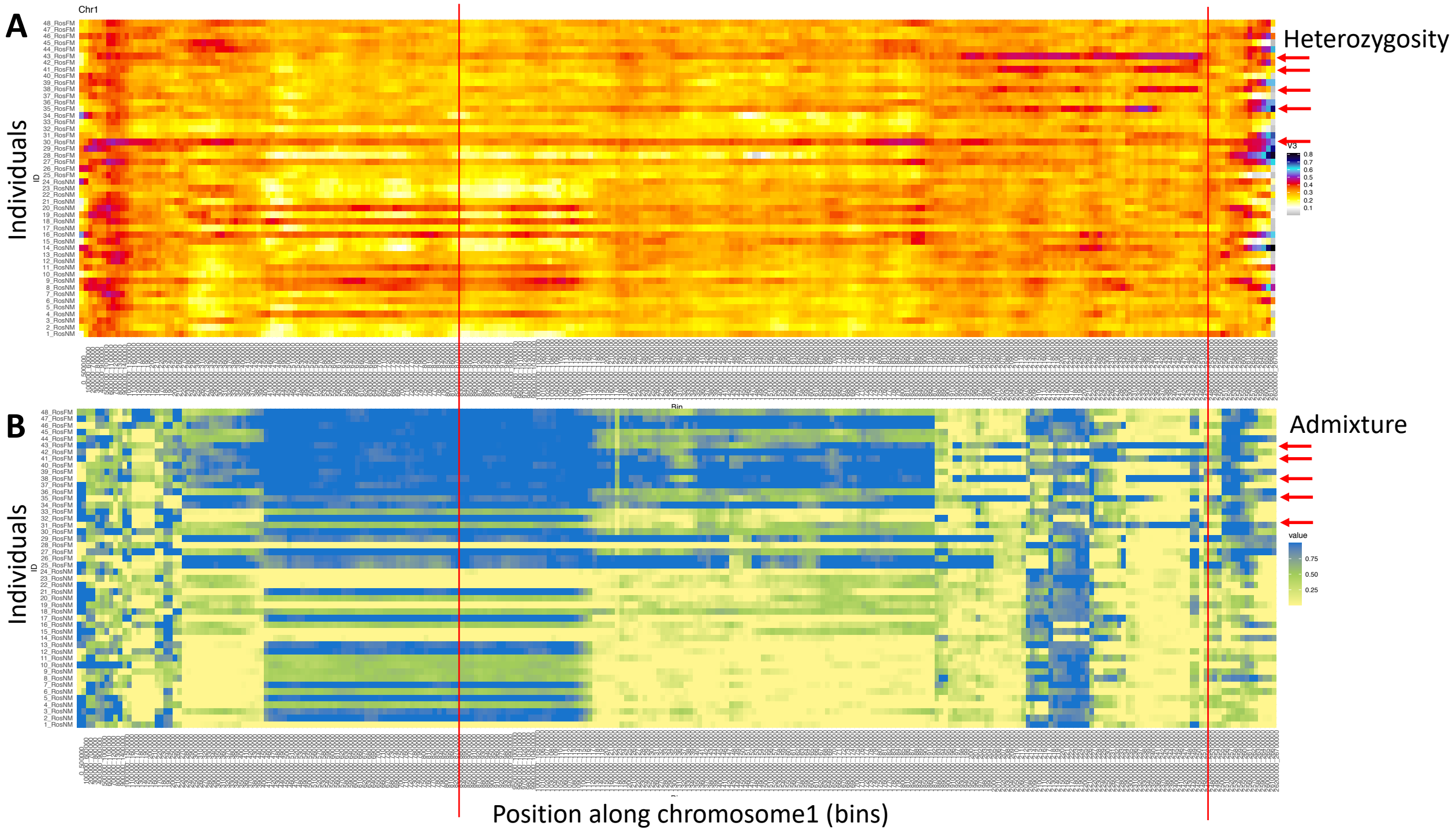


BIN 26100000_26600000

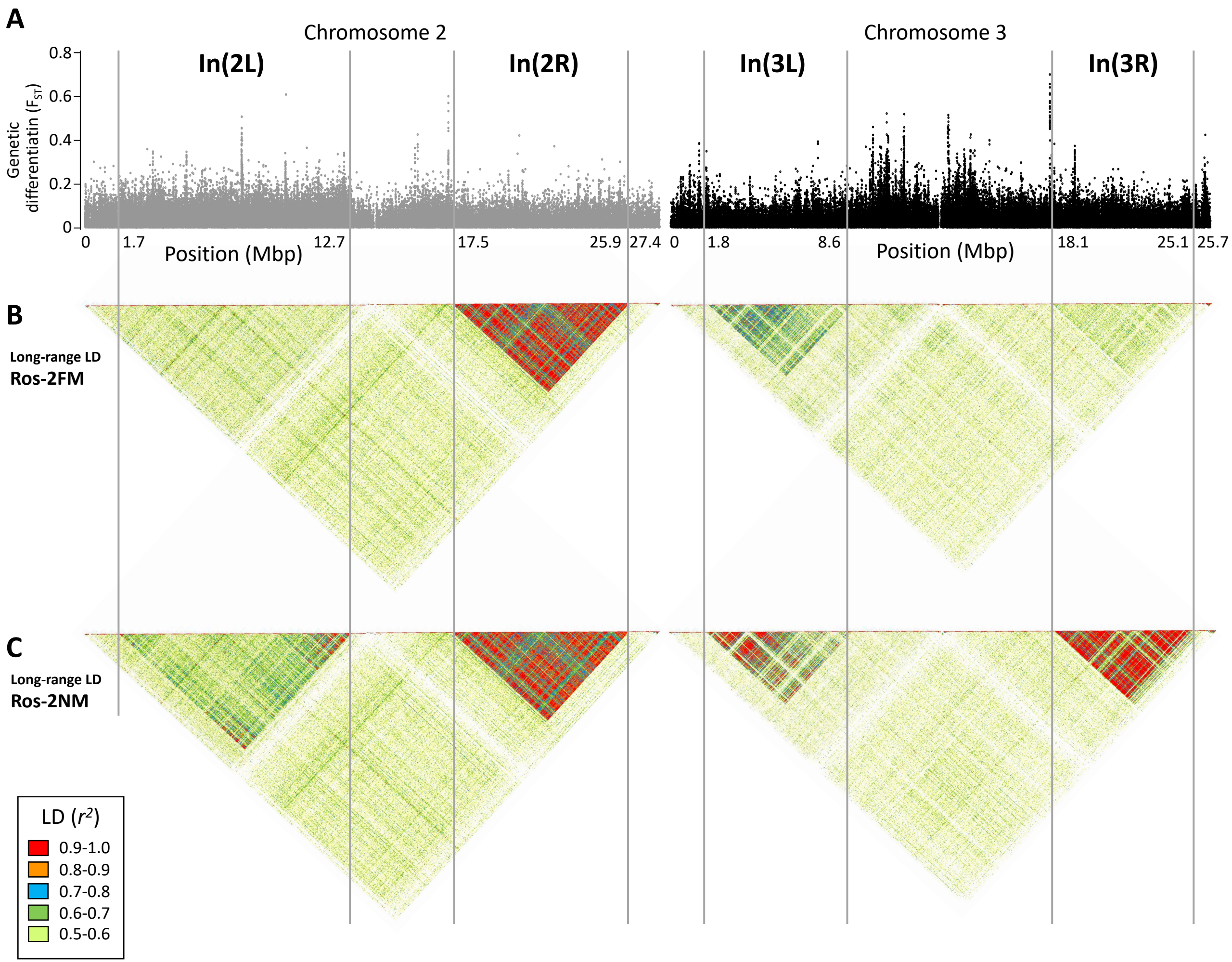


BIN 26200000_26700000





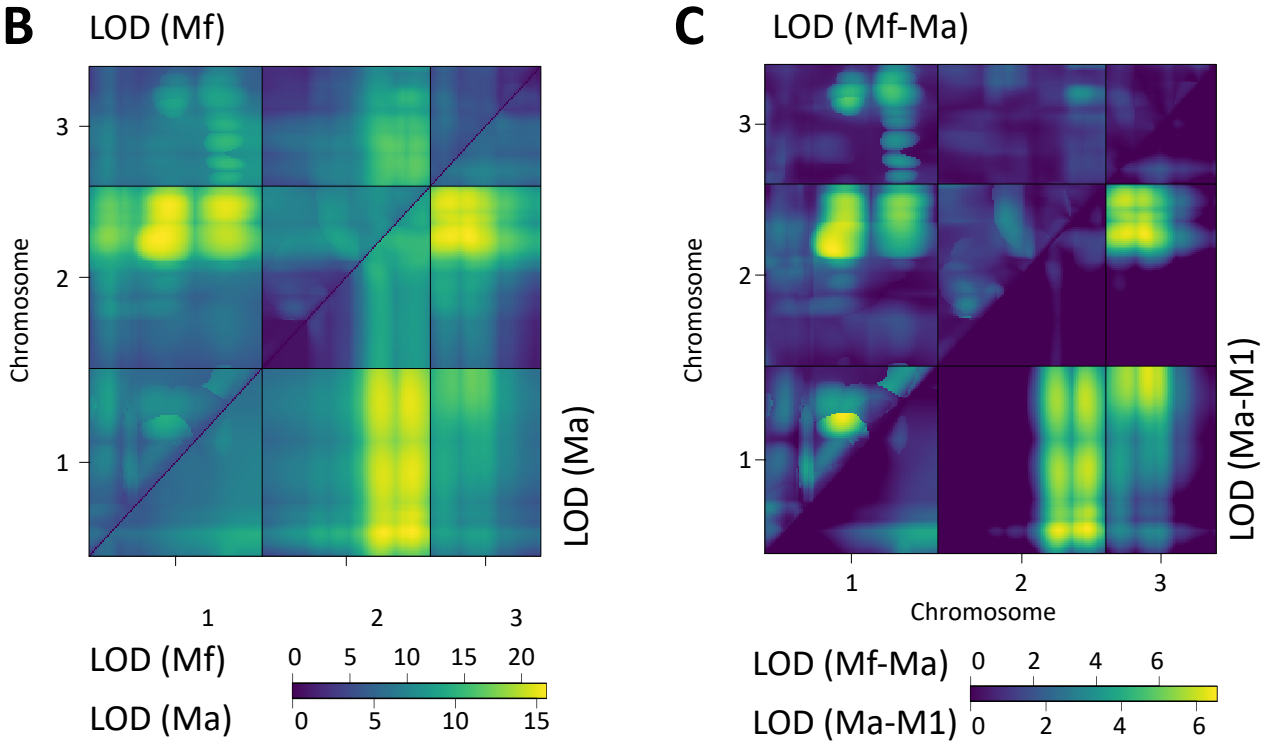
Supplementary Figure 5 Windowed analysis for observed heterozygosity (A) and Admixture (B) along chromosome 1. (**A**) A number of individuals (red arrows) show elevated heterozygosity beyond the end of In(1a). In three of them (individuals 38, 41 and 43) the ends of that region coincide (red lines). These individuals are considered SI_3 heterozygotes, carrying In(1c). In the other two individuals the region is shorter or longer, possibly pointing to yet other inversions. (**B**) Admixture scores from windowed analysis support the same individuals to be genetically distinct in the right arm of In(1c).



Supplementary Figure 6 Long range linkage disequilibrium along chromosomes 2 and 3. The block of slightly elevated genetic differentiation in chromosome arm 2L (**A**) correspond to a block of elevated long-range LD in the NM type (**C**), but not the full moon type (**B**). Elevated long-range LD suggets that the inversion is polymorphic in the repective population, congruent with the genotyping results presented in Figure 2 (the higher the allele frequency of the inverted haplotype, the stronger is LD).

A Tested models for each QTL pair

Mf = full model	*additive + epistatic effects*
Ma = additive model	*only additive effects*
M1 = one QTL model	*only one of the two QTL exists*



D

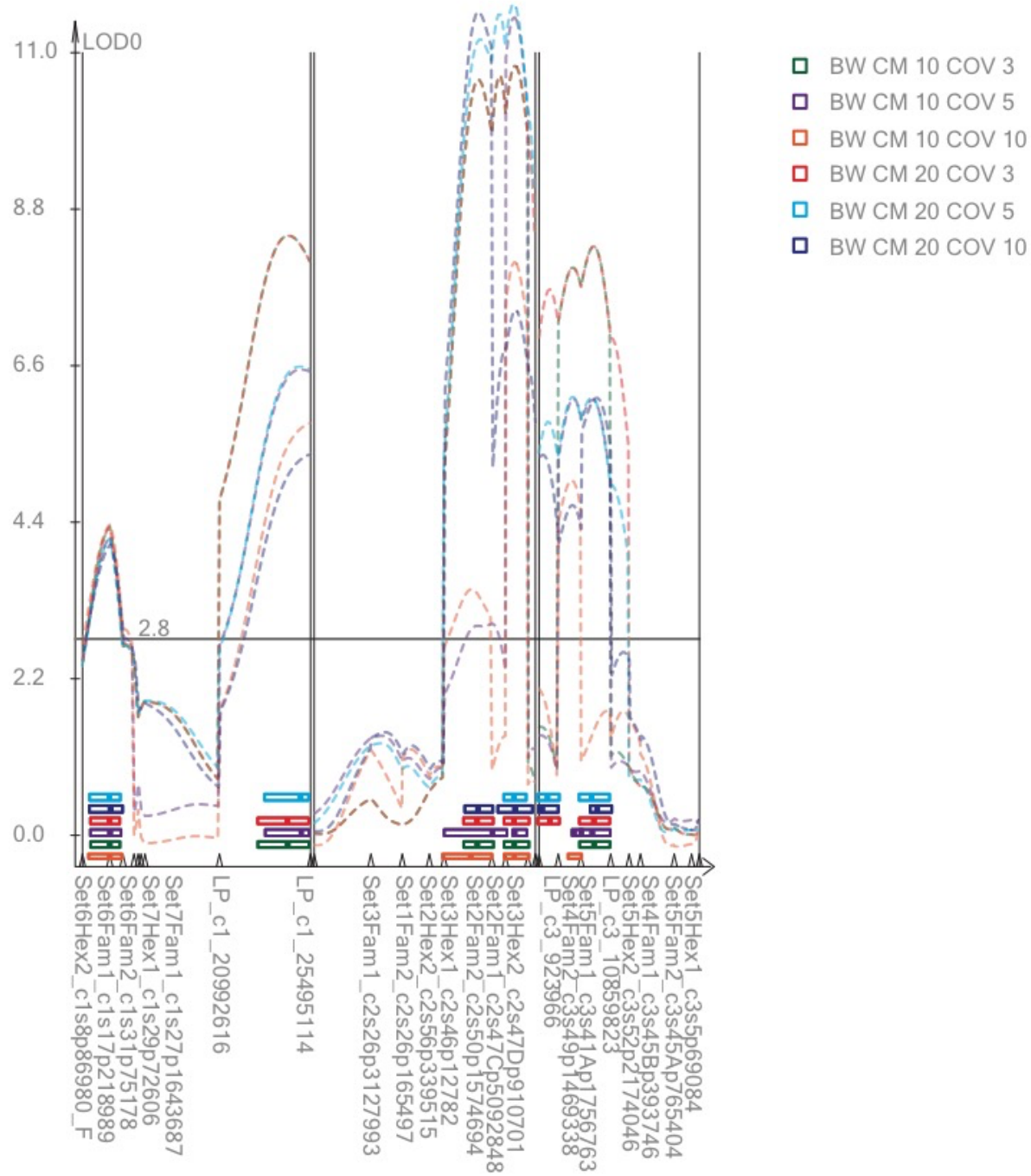
Interacting QTL			Tests for epistatic interactions					
Chr	Pos 1 (cM)	Pos 2 (cM)	LOD(Mf)	p(Mf)	LOD(Mf-M1)	p(Mf-M1)	LOD(Mf-Ma)	p(Mf-Ma)
1:1	46	78	**13.9**	**0**	**8.12**	**0.016**	4.25	0.460
1:2	42	72	**22.1**	**0**	**13.09**	**0.000**	**6.56**	**0.018**
1:3	84	12	**14.7**	**0**	**9.01**	**0.004**	2.78	0.925
2:3	93	12	**16.6**	**0**	**7.61**	**0.033**	1.16	1.000

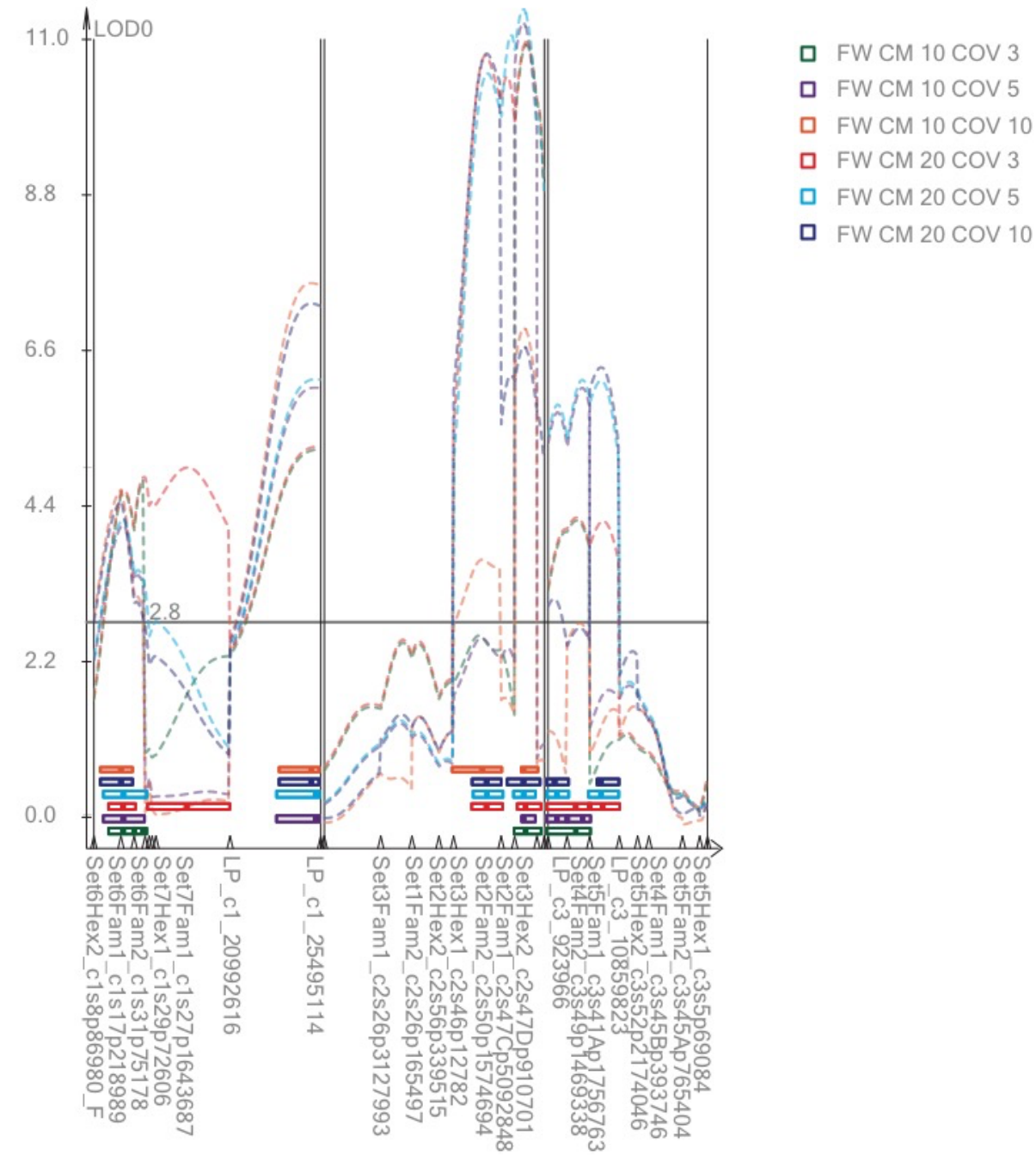
E

Interacting QTL			Tests for additivity			
Chr	Pos 1 (cM)	Pos 2 (cM)	LOD(Ma)	p(Ma)	LOD(Ma-M1)	p(Ma-M1)
1:1	12	100	**9.62**	**0**	**3.87**	**0.003**
1:2	13	90	**15.57**	**0**	**6.53**	**0.000**
1:3	101	25	**11.97**	**0**	**6.23**	**0.000**
2:3	74	23	**15.49**	**0**	**6.45**	**0.000**

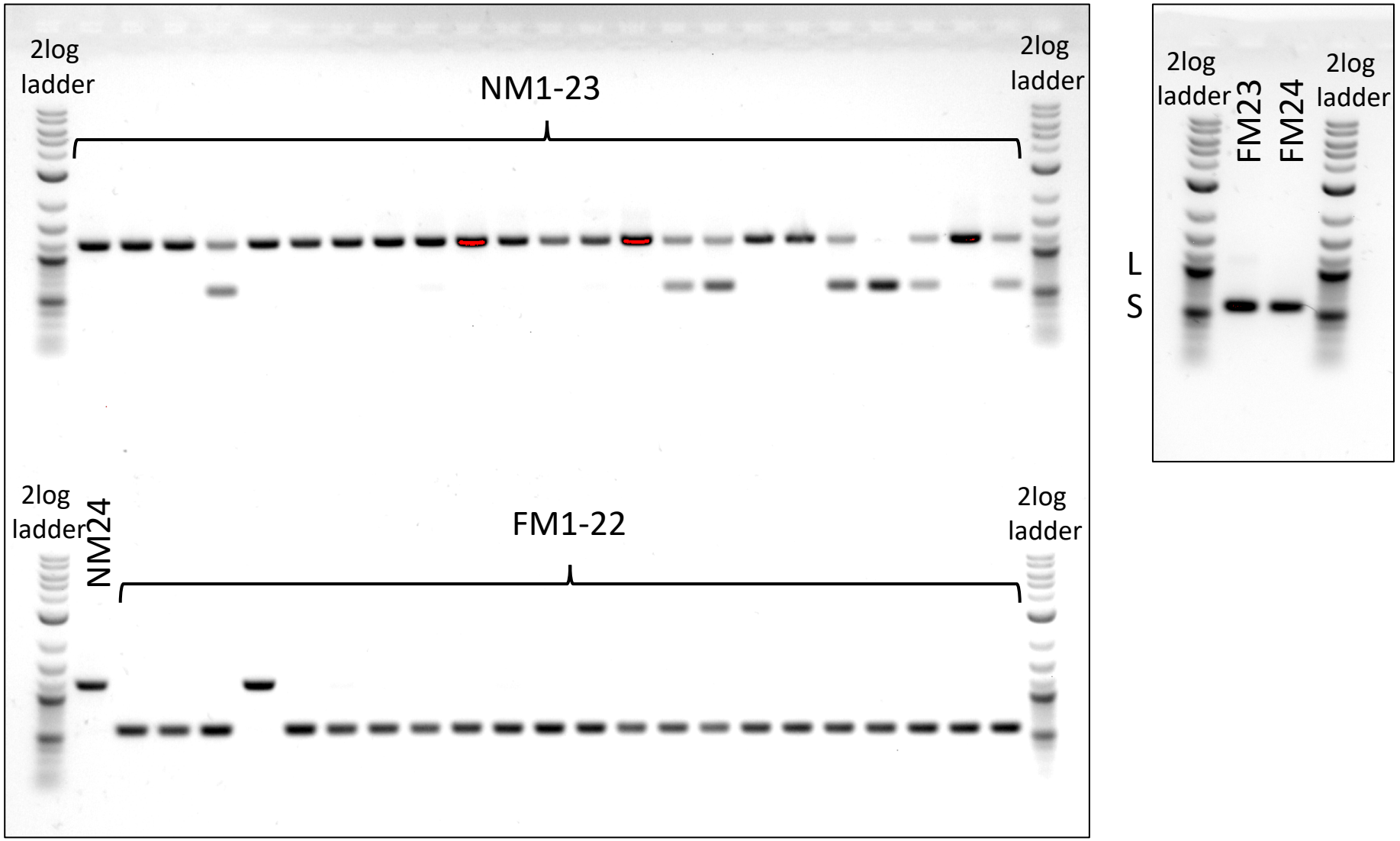
Supplementary Figure 7 Scan for interacting QTL

All chromosmes were scanned for interacting QTLs with the *scantwo* function of R/qtl. (**A**) For each possible QTL combination three models are tested, allowing for either additve and epistatic effects (full model; Mf), only additive effects (Ma) or assuming that one of the two QTL does not exist (M1). (**B**) The LOD score for the full model (Mf) and the additive-only model (Ma) are similar across the genome. (**C**) Congruently, there LOD scores are much lower for the difference between Mf and Ma (upper triangle). The additive model Ma does much better in explaining the data than assuming only one QTL exists (Ma-M1; lower triangle). (**D**,**E**) Statistical tests for the detected QTL combinations. Only in one case the full model is significantly better than the additive-only model (see p(Mf-Ma)), suggesting that epistatic effects are limited to an interaction between the middle of chromosome 1 and the end of chromosome 2 (D). The additive model is significantly better than assuming there is only one QTL (E).

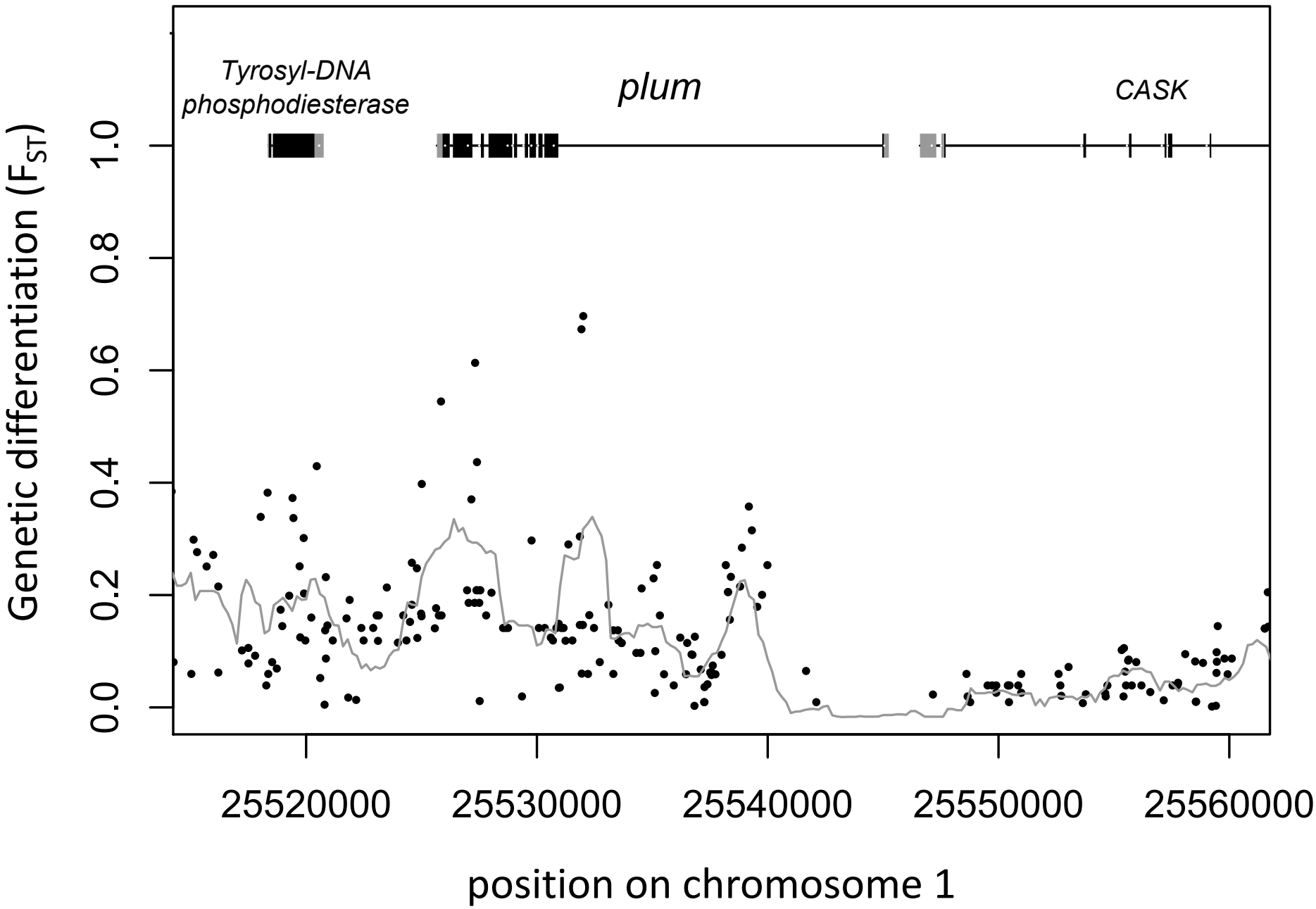




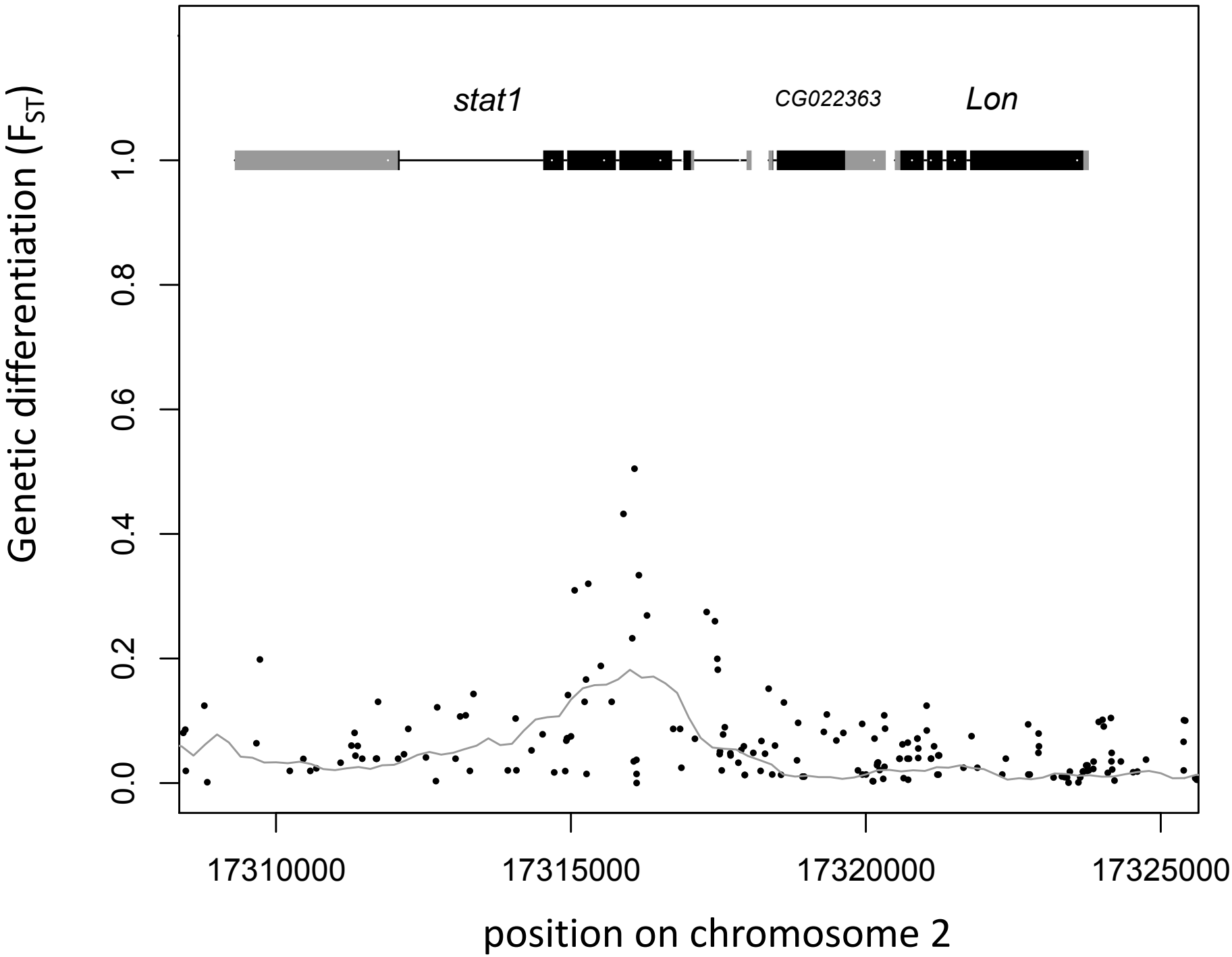
Supplementary Figure 8 LOD profiles from Composite Interval Mapping (CIM).
(**A**) Backward selection models. (**B**) Forward selection models. CM = size of the exclusion window in centimorgan; COV = number of covariates. CIM generally identifies the same four QTL as MQM, but sometimes identifies several adjacent QTL in the same genomic region.



Supplementary Figure 9 Genotyping of an insertion-deletion in the *period* locus.
The NM strain is dominated by the long allele (L), the FM strain by the short allele (S).



Supplementary Figure 10 Gene models and genetic differentiation at the *plum* locus. Dots are individual SNPs, the line is FST in 1kb sliding windows with 200 bp steps. In the gene models grey boxes are untranslated regions (UTRs), black boxes are coding sequence (CDS).



Supplementary Figure 11 Gene models and genetic differentiation at the *stat1* locus. Dots are individual SNPs, the line is FST in 1kb sliding windows with 200 bp steps. In the gene models grey boxes are untranslated regions (UTRs), black boxes are coding sequence (CDS).

Supplementary Table 1 Inversion breakpoints as estimated from long range LD.

Inversion	Chr	Left breakpoint (bp)	Right breakpoint (bp)	Size (bp)
In(1a)	1	2.741.953	18.733.448	15.991.495
In(1b)	1	4.353.782	11.342.207	6.988.425
In(2L)	2	1.675.117	12.670.058	10.944.941
In(2R)	2	17.530.360	25.926.796	8.396.436
In(3L)	3	1.806.371	8.614.086	6.807.715
In(3R)	3	18.144.782	25.076.351	6.931.569

Supplementary Table 2 Number of variants in inversion windows

	Number of variants
Chr1	261.739
In(1a), left arm	15.972
In(1b)	83.456
In(1a), right arm	92.064
In(1c), right arm	52.125
Chr2	249.948
In(2L)	114.556
In(2R)	77.001
Chr3	191.892
In(3L)	53.786
In(3R)	51.788

Name	Internal name [1]	Chromosome	Position [2]	Motif	Forward primer	Reverse primer	Size [3]	Dye	Set	Scorable [4]	Marker
MS101	MS_c1s8p86980	1	634807	AG	TCAAGACGCCAGACAAAAGA	CGCCGCTTCAAATTACTGAC	415	HEX	6	YES	C1M1
MS102	MS_c1s15p1802123	1	1467742	GAA	GGGTGCGTCACAAAAGTTTC	TGTTTGAGCTTTTCCTGTTGC	220	HEX	8	NO	-
MS103	MS_c1s29p72606	1	4419208	TC	TTGCGATTTTCTGTGGTCTG	TTGGTCTTGACTGGCTGTTG	213	HEX	7	YES	C1M2
MS104	MS_c1s59p1429135	1	5544616	TG	CGACACATATCACGGAGCAG	TCGGCAAATCTAACCATACGA	306	FAM	7	YES	C1M3
MS105	MS_c1s19p129932	1	6487528	TC	GCCCTGAAGCATTTCAAAAA	TCCTTGTAGAAAGGCGGAAA	222	HEX	6	YES	C1M4
MS106	MS_c1s27p1643687	1	8266251	AG	TTTAAGTCAATGCGGAAGAGG	GAACAAAAACCATCGCACAA	100	FAM	7	YES	C1M5
MS107	MS_c1s22Bp113007	1	8553163	TC	ACTTCGCATGTTTCCCAGAA	CCATGTCATCAACTTCCGTCT	411	HEX	7	NO	-
MS108	MS_c1s31p75178	1	14704747	AG	TGGATTGAGATGGACGTTGA	CTGTTCGAGTTCAAATGGAATG	305	FAM	6	YES	C1M6
MS109	MS_c1s17p218989	1	17462705	TG	CAAGCATGACACATCAAGCA	CACACAAACCGTATCGGTGA	124	FAM	6	YES	C1M7
MS110	MS_c1s42p265009	1	18183719	GAT	CACGACGCACCACTTAAGAA	CACAAAATCAGTCGGTGGAA	329	FAM	8	NO	-
LP1	LP_c1_20992616	1	20992616	-	TAGAGCGATCCGCGCTAATA	GATTATCTGTTCCACTTGCGATA	148/227	-	-	YES	C1M8
MS111	MS_c1s44p1006880	1	22933080	CT	CCATTTAAGGGATGCGTGAT	CAAGCGAAAGCAATTTCCAT	420	HEX	8	NO	-
LP2	LP_c1_25495114	1	25495114	-	TCAAAATGTGAAGTTTCAAAGCA	TTGGAGCATTTCTTCAATGTTT	114/156	-	-	YES	C1M9
MS112	MS_c1s61p885571	1	25862224	GA	CAAAACATCTTTTCGGGTCAC	GGGTTCTAGCAAAGCGAAGA	177	FAM	8	NO	-
Name	**Internal name [1]**	**Chromosome**	**Position [2]**	**Motif**	**Forward primer**	**Reverse primer**	**Size [3]**	**Dye**	**Set**	**Scorable [4]**	**Marker**
MS113	MS_c2s20p241395	2	502108	AG	TTGGCTATTTGGTTGGTCGT	TTGATGTGTTCATCGCACTG	244	HEX	2	YES	C2M1
MS114	MS_c2s13p558035	2	1242367	TC	GGGTCGCGTCATGATTCTTA	CCAAACCTGAAAATCCATCC	219	HEX	1	NO	-
MS115	MS_c2s46p12782	2	2800433	TC	TTTGGCGTTGTTCAAGTCTCT	TCGACTCAAGCTCCAAAGTG	215	HEX	3	YES	C2M2
MS116	MS_c2s56p339515	2	5674405	CAT	TGAATGTCGTGTTTGGGTTAAG	GCAGATCTACAATGAAAGGTAATGG	360	HEX	2	YES	C2M3
MS117	MS_c2s26p165497	2	8425973	TC	CCTTTTTGCCATCGTTCATT	TGTCGAGGGCATGAATATGA	149	FAM	1	YES	C2M4
MS118	MS_c2s26p3127993	2	11358786	AG	TGAAACCAAAAGTGGCACAA	TGGAAATCCAAAGTGGAAGAG	149	FAM	3	YES	C2M5
MS119	MS_c2s50p1574694	2	14667091	TC	TGCTTTGACGCTAGTAAACAGG	AAATGATGAACCGTGACTGTCT	161	FAM	2	YES	C2M6
MS120	MS_c2s47Cp5092848	2	18620830	CA	TGACCTTTTCAAGGCAAATTAAA	TCTCACGTTTTCGATGATGG	300	FAM	2	YES	C2M7
MS121	MS_c2s47Cp2334121	2	21305073	AG	GTTTCATCATTTCGCTGTTGC	TCCACAAGCTAAGATGCAATTC	306	FAM	3	NO	-
MS122	MS_c2s47Ap263174	2	24127353	TC	AACGATGAATGGCAATTTGA	ACCCACGCATTTTTGGTAAG	306	FAM	1	NO	-
MS123	MS_c2s47Dp910701	2	25271117	AC	AATTCGCTTTGACTGGTGCT	GCCATTGAGTGGCCAAATAC	445	HEX	3	YES	C2M8
MS124	MS_c2s47Fp635401	2	27069765	AG	GCAAATTCAAGGGAATCCTG	TTCCAGCAAAGTCTGTTTCG	461	HEX	1	YES	C2M9
Name	**Internal name [1]**	**Chromosome**	**Position [2]**	**Motif**	**Forward primer**	**Reverse primer**	**Size [3]**	**Dye**	**Set**	**Scorable [4]**	**Marker**
LP3	LP_c3_923966	3	923966	-	AAACAATGTCAATTGCCAAATA	TGAGACTGATGATAGATTATGTTCCA	126/195	-	-	YES	C3M1
MS125	MS_c3s49p1469338	3	3109346	AG	GCGATTAAATGTCGCGATAAG	CGTCAGAAGTCAAAGTGTCAGG	349	FAM	4	YES	C3M2
MS126	MS_c3s41Ap1756763	3	5995719	TAA	GTGGAAAGCTTTGGCGTAGA	TGCTTCTTGTCTTGCTCAATG	149	FAM	5	YES	C3M3
LP4	LP_c3_10859823	3	10859823	-	CGCAGCGTTTTTCAATACAA	TTTGATTTYCTGGGGTGTGA	149/228	-	-	YES	C3M4
MS127	MS_c3s48p3645480	3	12404930	CT	GCATGCAGATTCGAAAAACA	AGCGCTGGTTGAAAAAGAGA	270	HEX	4	NO	-
MS128	MS_c3s52p2174046	3	16511020	TG	CGACGACTCAGTGGTTATGC	ACTCCCACACCTTTCAATCAA	422	HEX	5	YES	C3M5
MS129	MS_c3s45Bp393746	3	18127054	AC	CGGGGCGAGTATGTTATCAA	GAATGAAGGCTGGAAAAAGTCA	151	FAM	4	YES	C3M6
MS130	MS_c3s45Ap765404	3	20771652	TG	GCACGAACGAGTTGTGAATG	GGCAGATCGAAATGTCTTTGT	301	FAM	5	YES	C3M7
MS131	MS_c3s5p69084	3	22737957	TG	TGCAATCGTACGTTGAACTCTT	TTGCAAACTGACGCAGACTC	211	HEX	5	YES	C3M8
MS132	MS_c3s3p743476	3	25467715	AC	TGGAATGGTAGCGGGTTAAA	TCGCTCGTTACAAGATGGAA	436	HEX	4	YES	C3M9

[1] Microsatellite (MS) or length polymorphism (LP); chromosme (c), scaffold (s) and position (p) on CLUMA1.0 reference assembly for MS and on CLUMA2.0 assembly for LP

[2] 5' end of forward primer in CLUMA2.0 refernce assembly

[3] expected amplicon size based on CLUM1.0 reference assembly

[4] microsatellite/length polymorphism was PCR-amplifiable and segregated in a scorable pattern in Ros-FMxNM cross

Supplementary Table 4 Final QTL model as given by the *fitqtl* function, obtained with multiple imputation, a normal phenotype model and based on 158 observations.

a) Model formula: y ~ Q1 + Q2 + Q3 + Q4

b) Full Model result

	df	SS	MS	LOD	%var	p(Chi2)	p(F)
Model	8	797.6944	99.711800	26.71958	54.10373	0	0
Error	149	676.6853	4.541512				
Total	157	1474.3797					

c) Drop one QTL at a time ANOVA table

QTL	df	Type III SS	LOD	%var	F value	p(Chi2)	p(F)
1@12.5	2	93.71	4.450	6.356	10.32	0	6.36e-05
1@105.0	2	153.28	7.005	10.396	16.87	0	2.48e-07
2@89.0	2	301.80	12.653	20.470	33.23	0	1.17e-12
3@14.0	2	159.54	7.263	10.821	17.56	0	1.42e-07

d) Estimated effects

QTL	Additive effect			Dominance effect		
	est	SE	t	est	SE	t
1@12.5	1.00331	0.27691	3.623	0.76487	0.39691	1.927
1@105.0	1.42628	0.24778	5.756	0.01726	0.40408	0.043
2@89.0	2.10262	0.28000	7.509	0.06049	0.36010	0.168
3@14.0	1.53380	0.26992	5.682	-0.55051	0.35045	-1.571
Intercept	5.43907	0.18264	29.781			

FST	QTL	Gene ID	Putative gene	Chromosome	Position	SNPs												Indels													
						CDS: non-synonymous	CDS: synonymous	Stop gained	Stop lost	Stop retained	Start lost	Splice site	3' UTR	5' UTR	Intron	Non coding transcript	Intergenic	Intragenic	Frameshift	Inframe (conservative)	Inframe (disruptive)	Stop gained	Stop lost	Stop retained	Start lost	Start retained	Splice site	3' UTR	5' UTR	Intron	Non coding transcript	Intergenic	Intragenic
>0.6		CLUMA_CG020013	NA	chr1	4134425	.	.	.	.	.	.	.	.	.	.	.	4	.	.	.	.	.	.	.	.	.	.	.	.	2	.	2	.
>0.6		CLUMA_CG019977	(sp)_putative_Arginine-hydroxylase_NDUFAF5%2C_mitochondrial	chr1	4135600	.	.	.	.	.	.	.	.	.	.	.	4	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.
>0.6		CLUMA_CG019979	(nr)_putative_nuclear_pore_complex_protein_DDB_G0274915-like	chr1	4142501	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.6		CLUMA_CG005654	(pfam)_DUF1676_domain_containing_protein_CLUMA_CG005654	chr1	4457076	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
>0.6		CLUMA_CG005667	(pfam)_DUF1676_domain_containing_protein_CLUMA_CG005667	chr1	4457076	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
>0.6		CLUMA_CG000851	(sp)_similar_to_Dynamin-1-like_protein_CLUMA_CG000851	chr1	4714034	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
>0.6		CLUMA_CG000853	NA	chr1	4714034	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
>0.6		CLUMA_CG000841	NA	chr1	4817120	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.6		CLUMA_CG018930	(sp)_putative_5'-nucleotidase_domain-containing_protein_3	chr1	5674933	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.6		CLUMA_CG003005	(mn)_Ionotropic_receptor_60a_(IR60a)	chr1	6300137	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
>0.6		CLUMA_CG003000	(sp)_similar_to_Limbic_system-associated_membrane_protein_CLUMA_CG003000	chr1	6300137	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
>0.6		CLUMA_CG005143	(sp)_putative_Cadherin-87A	chr1	6738725	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.6		CLUMA_CG005226	(sp)_putative_Hormone_receptor_4	chr1	7036032	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	2	.	.	.
>0.6		CLUMA_CG005651	(sp)_putative_Serine_protease_3	chr1	8399023	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
>0.6		CLUMA_CG005612	(sp)_putative_Toll-like_receptor_7	chr1	8399023	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
>0.6		CLUMA_CG003833	(mn)_Ionotropic_receptor_7.x8_(IR7.x8)_pseudogene?	chr1	8526350	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.6		CLUMA_CG003841	(sp)_putative_Ras_GTPase-activating_protein_raskol	chr1	8526350	.	.	.	.	.	.	.	.	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.6		CLUMA_CG018849	NA	chr1	9209667	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	2	.	1	.	.	.
>0.6		CLUMA_CG002108	(sp)_related_to_Retinol_dehydrogenase_11	chr1	11354040	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.6		CLUMA_CG002110	(sp)_similar_to_Glutathione_S-transferase_1-1_CLUMA_CG002110	chr1	11354040	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.6		CLUMA_CG024385	(sp)_putative_Protein_muscleblind	chr1	12759391	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.6		CLUMA_CG001740	NA	chr1	12759391	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.6		CLUMA_CG002622	(sp)_putative_Coiled-coil_domain-containing_protein_lobo	chr1	13913382	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
>0.6		CLUMA_CG024509	NA	chr1	13913382	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	.
>0.6		CLUMA_CG002549	(sp)_related_to_3-hydroxyisobutyryl-CoA_hydrolase%2C_mitochondrial_CLUMA_CG002549	chr1	14033949	2	2	.	.	.	.	.	1	2	.	1	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
>0.6		CLUMA_CG022477	(sp)_related_to_3-hydroxyisobutyryl-CoA_hydrolase%2C_mitochondrial_CLUMA_CG022477	chr1	14035773	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.6	QTL 1	CLUMA_CG008386	(mn)_Tachykinin-like_receptor_at_99D_(TkR99D)	chr1	15777096	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.6	QTL 1	CLUMA_CG020610	(sp)_putative_Syntaxin-1A	chr1	16591767	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5	.	.	.
>0.6	QTL 1	CLUMA_CG020824	(mn)_period_(per)	chr1	17280426	.	.	.	.	.	.	.	7	19	.	6	.	.	.	.	.	.	.	.	.	2	14	.	3	.
>0.6	QTL 1	CLUMA_CG002364	(mn)_pickpocket_CLUMA_CG002364	chr1	17281861	.	2	.	.	.	.	1	.	1	.	9	.	.	.	.	.	.	.	.	.	.	.	.	3	.
>0.6	QTL 1	CLUMA_CG002369	(sp)_similar_to_Dehydrogenase/reductase_SDR_family_member_on_chromosome_X_CLUMA_CG002369	chr1	17285502	.	3	.	.	.	.	.	.	.	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.6	QTL 1	CLUMA_CG002460	NA	chr1	17676854	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.6	QTL 1	CLUMA_CG002472	NA	chr1	17676854	1	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.6	QTL 1	CLUMA_CG006144	(sp)_related_to_Protein_spitz_CLUMA_CG006144	chr1	17770917	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.
>0.6	QTL 1	CLUMA_CG023302	(sp)_related_to_Protein_spitz_CLUMA_CG023302	chr1	17770917	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.
>0.6	QTL 1	CLUMA_CG008012	(sp)_putative_DNA_topoisomerase_1	chr1	18087445	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4	.	.	.	.
>0.6	QTL 1	CLUMA_CG008187	(sp)_similar_to_N-acetylgalactosaminyltransferase_4_CLUMA_CG008187	chr1	18706611	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.6	QTL 2	CLUMA_CG020204	(sp)_putative_Nuclear_pore_complex_protein_Nup133	chr1	25510150	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.6	QTL 2	CLUMA_CG020208	(mn)_plum | (sp)_similar_to_Brother_of_CDO	chr1	25527322	.	.	.	.	.	.	.	.	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4		CLUMA_CG017949	(sp)_similar_to_3-oxoacyl-[acyl-carrier-protein]_reductase_FabG_CLUMA_CG017949	chr2	7452391	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	1	.
>0.4		CLUMA_CG017947	NA	chr2	7452391	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	1	.
>0.4		CLUMA_CG017952	(sp)_putative_Sodium/bile_acid_cotransporter_5	chr2	7461980	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4		CLUMA_CG017953	(mn)_Gustatory_receptor_43_(GR43)	chr2	7462968	1	6	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4		CLUMA_CG009535	(sp)_related_to_HIV_Tat-specific_factor_1_homolog	chr2	7464868	1	1	.	.	.	.	.	.	2	.	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4		CLUMA_CG017954	NA	chr2	7466572	.	.	.	.	.	.	.	.	.	.	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4		CLUMA_CG024356	NA	chr2	7487587	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4		CLUMA_CG003411	NA	chr2	7487587	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4		CLUMA_CG004602	(sp)_similar_to_Protein_tramtrack%2C_beta_CLUMA_CG004602	chr2	9578624	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4		CLUMA_CG004863	(sp)_putative_cAMP-specific_3'%2C5'-cyclic_phosphodiesterase	chr2	10568233	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	11	.	.
>0.4	QTL	CLUMA_CG015068	(sp)_putative_Alpha-L-iduronidase	chr2	15853287	2	1	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG015349	(sp)_putative_TBC1_domain_family_member_1	chr2	15854557	.	.	.	.	.	.	.	2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG012971	(mn)_signal_transducer_and_transcription_activator_Stat92E-like_1_(stat92EI1) | Stat1-like	chr2	17316893	.	.	.	.	.	.	.	.	9	.	1	.	.	.	.	.	.	.	.	.	.	.	3	.	.
>0.4	QTL	CLUMA_CG022363	NA	chr2	17318296	.	1	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG011298	(sp)_putative_TGF-beta_receptor_type-1	chr2	20700840	.	.	.	.	.	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG010830	(sp)_related_to_Galactosylgalactosylxylosylprotein_3-beta-glucuronosyltransferase_P_CLUMA_CG010830	chr2	22370220	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG023463	(sp)_related_to_Galactosylgalactosylxylosylprotein_3-beta-glucuronosyltransferase_P_CLUMA_CG023463	chr2	22370220	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG000081	(sp)_putative_Membralin	chr3	1365270	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG007484	NA	chr3	7027320	2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	.	.	.
>0.4	QTL	CLUMA_CG013280	(sp)_putative_Protein_ECT2	chr3	9308367	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4	.	.	.
>0.4	QTL	CLUMA_CG013356	(sp)_putative_Protein_furry	chr3	9647837	4	6	.	.	.	.	.	.	2	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG013355	NA	chr3	9647837	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG013429	(sp)_putative_Semaphorin-2A	chr3	9982618	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4	.	.	.
>0.4	QTL	CLUMA_CG013443	(sp)_putative_ARF_GTPase-activating_protein_GIT2	chr3	10042796	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG013513	(mn)_TRP_channel_transient_receptor_potential_cation_channel_subfamily_M_(Trpm)	chr3	10303595	.	24	.	.	.	.	.	.	12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG022407	(sp)_putative_Cytosolic_carboxypeptidase-like_protein_5	chr3	10362891	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG013537	NA	chr3	10377569	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG013538	NA	chr3	10377569	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG013539	(mn)_Down_syndrome_cell_adhesion_molecule_2_(Dscam2)	chr3	10380101	.	.	.	.	.	.	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG024000	(mn)_Aquaporin_big_brain_(bib)	chr3	11105680	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
>0.4	QTL	CLUMA_CG013720	(sp)_putative_Neuroligin-like_protein_glit-1	chr3	11105680	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
>0.4	QTL	CLUMA_CG013726	(sp)_similar_to_Trypsin%2C_alkaline_A_CLUMA_CG013726	chr3	11127429	1	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
>0.4	QTL	CLUMA_CG024669	(nr)_putative_protein_obstructor-E-like	chr3	11131133	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG018167	(sp)_similar_to_Trypsin_beta_CLUMA_CG018167	chr3	11131133	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG022604	(sp)_similar_to_Trypsin%2C_alkaline_A_CLUMA_CG022604	chr3	11136276	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG013727	(sp)_similar_to_Lipase_member_H_CLUMA_CG013727	chr3	11138968	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4	QTL	CLUMA_CG013728	(pfam)_DUF2236_domain_containing_protein_CLUMA_CG013728	chr3	11141251	.	.	.	.	.	.	1	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	3	.	.	.
>0.4		CLUMA_CG015788	(sp)_putative_Zinc_transporter_2	chr3	13221842	.	.	.	.	.	.	.	2	14	.	3	.	.	.	.	.	.	.	.	.	.	.	6	.	1	.
>0.4		CLUMA_CG015831	NA	chr3	13221842	.	.	.	.	.	.	.	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4		CLUMA_CG015766	(sp)_similar_to_Chaoptin	chr3	13241308	.	.	.	.	.	.	.	.	1	3	3	.	.	.	.	.	.	.	.	.	.	.	1	.	1	.
>0.4		CLUMA_CG016118	(nr)_putative_mucin-5AC_CLUMA_CG016118	chr3	13287644	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
>0.4		CLUMA_CG015894	NA	chr3	13291314	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4		CLUMA_CG024579	(sp)_putative_Inositol_monophosphatase_1	chr3	14294877	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4		CLUMA_CG015899	(mn)_ligand-dependent_nuclear_receptor_transcription_factors_ecdysone-induced_protein_78C_(Eip78c)	chr3	14313566	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	.	.	.
>0.4		CLUMA_CG016302	(nr)_related_to_LOW_QUALITY_PROTEIN:_growth_arrest_and_DNA_damage-inducible_protein_GADD45_alpl	chr3	15205513	.	2	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4		CLUMA_CG016944	(sp)_similar_to_Tryptase	chr3	15205909	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4		CLUMA_CG009336	NA	chr3	18084332	1	.	.	.	.	.	.	.	.	.	13	.	.	.	.	.	.	.	.	.	.	.	.	4	.
>0.4		CLUMA_CG009337	NA	chr3	18084693	1	.	.	.	.	.	.	3	.	.	13	.	.	.	.	.	.	.	.	.	.	.	.	4	.
>0.4		CLUMA_CG022126	NA	chr3	18088307	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4		CLUMA_CG009338	(sp)_putative_ADP-ribosylation_factor_1	chr3	18092454	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4		CLUMA_CG022849	NA	chr3	18300253	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4		CLUMA_CG009657	(sp)_related_to_Condensin_complex_subunit_2	chr3	19273849	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
>0.4		CLUMA_CG006228	(sp)_putative_Zinc_finger_MIZ_domain-containing_protein_1	chr3	25493088	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4	.	.	.