Adaptation to heavy-metal contaminated environments proceeds via selection on pre-existing genetic variation.

Kevin M. Wright 1*, Uffe Hellsten 2, Chenling Xu 3, Annie L. Jeong 4,5, Avinash Sreedasyam 6, Jarrod A. Chapman 2, Jeremy Schmutz 6, Graham Coop 3, Daniel S. Rokhsar 2,7, John H. Willis 4

1. Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA.
2. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA.
3. Department of Evolution and Ecology, University of California, Davis, CA, 95616, USA.
4. Department of Biology, Duke University, Durham, NC 27708, USA.
5. University Program in Genetics and Genomics, Duke University Medical Center, Durham, North Carolina, USA
6. Hudson Alpha Institute for Biotechnology, Huntsville, Alabama 35806, USA.
7. Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA

* Corresponding Author: wright@fas.harvard.edu

Across a species range, islands of stressful habitats impose similar selection pressures on isolated populations. It is as yet unclear, when populations respond to these selective pressures, the extent to which this results in convergent genetic evolution and whether convergence is due to independent mutations or shared ancestral variation. We address these questions investigating a classic example of adaptation by natural selection - the colonization of plant species to heavy metal contaminated soils. We use field-based reciprocal transplant experiments to demonstrate that mine alleles at a major copper tolerance QTL, *Tol1*, are strongly selected in the mine environment, but are neutral, or nearly so, in the off-mine environment. To identify scaffolds in genetic linkage with this locus, we assemble the genome of a mine adapted *M. guttatus* genotype and sequence near isogenic lines (NILs) homozygous for tolerant or non-tolerant alleles at *Tol1*. We identify genes with differential expression between NILs and differences in allele frequency between independent pairs of mine and off-mine populations to identify *Tol1* candidate genes. We identify a single gene, a multicopper oxidase, with large differences in expression between NILs and allele frequency between populations. Furthermore, we find patterns of genetic variation at *Tol1*, and four additional candidate adaptation loci, are consistent with selection acting upon beneficial haplotypes that predates the existence of the copper mine habitat. We estimate the age of selected *Tol1* haplotype to be at least 1700 years old and was at a frequency of 0.4-0.6% in the ancestral population when mining was initiated 150 years ago. These results suggest that adaptation to the mine habitat routinely occurs via selection on ancestral variation, rather than independent *de-novo* mutations or migration between populations.

The most spectacular demonstrations of evolution often involve independent populations repeatedly evolving similar traits in response to similar environmental pressures (Endler 1977; Joron and Mallet 1998; Schluter, 2000). Convergent evolution of similar phenotypes suggests convergence in genotype, as may occur when selection operates on the same ancestral allele in independent populations or species (Colosimo *et al.* 2005; The Heliconius Genome Consortium, 2012). Additionally, there are multiple instances of the similar phenotypes evolving via independent genetic mutations at the same locus (Tishkoff *et al.* 2007, Karasov *et al.* 2010). When evolution occurs via selection on pre-existing genetic variation, adaptation to novel environments will proceed quite quickly. Conversely, when evolution proceeds via independent genetic mutations, adaptation will be limited by the waiting time for beneficial mutations to occur and go to high frequency. In order to understand the circumstances under which either of these evolutionary scenarios is likely to occur, and make predictions about how species will respond to rapid environmental changes such as global climate change, we investigate the genomic basis of a classic example of convergent phenotypic evolution.

Investigations into the repeated colonization of stressful soil environments by plants species have a long history in the development of evolutionary biology (Kruckeberg 1951; Turner 1969; Antonovics *et al.* 1971; Macnair 1987). With the initiation of the first, and still on-going, field experiments at Park Grass, UK,

scientists demonstrated that soil nutrient levels and pH impose strong selection on the local plant communities causing some plant species to rapidly evolve tolerance, whereas other species, despite ample migration opportunities, never colonize these stressful habitats (Brenchley 1958). This result has been replicated in investigations of plants colonizing heavy metal contaminated soils (Walley *et al.* 1974, Bradshaw 1984, Symeonidis *et al.* 1985; Al-Hiyaly *et al.* 1990, 1993). Interestingly, the only species to colonize contaminated soils also have tolerant individuals in populations unexposed to soil contamination. Whereas species that did not evolved tolerance, lack tolerant individuals in unexposed populations (*summarized in* Bradshaw, 1991). Given these results, Bradshaw concluded that tolerant species possessed the necessary genetic variation in ancestral populations, and species which remain non-tolerant lack the necessary genetic variation to adapt to novel environmental pressures (1991). However these results do not eliminate the possibility that tolerance was absent in the ancestral population and spread, by migration, to uncontaminated sites following adaptation. To distinguish between these two hypotheses, we need additional evidence on the source of adaptive genetic variation in natural populations.

Here, we investigate the recent and repeated adaptation of the annual wildflower, *Mimulus guttatus*, to copper contaminated sites near Copperopolis, CA (Macnair *et al.* 1993). Contaminated sites have 10-100 fold increase in the concentration of copper in the soil compared to uncontaminated sites (Allan and Sheppard 1971; Macnair et al. 1993; S. Table 1). Copper mining in this region of the Sierra Nevada foothills started in 1861 (Aubury 1908). Mine populations are located in close geographic proximity, within 40km of each other, and *M. guttatus* is very common in uncontaminated soils adjacent to mine sites (Allan and Sheppard 1973; Macnair *et al.* 1993). Copper tolerance is nearly fixed at four mine populations (99.77%, N =2796); at intermediate frequency (56.0%; N=1121) at eleven sites, mostly located within one mile of contaminated soil, and at low frequency (2.2%; N=1056;) in 11 populations located 1-10 miles from contaminated soil (Macnair *et al.* 1993). This suggests that the copper tolerance phenotype, measured as root growth in a hydroponic solution, is under strong selection in contaminated habitats. It remains to be determined whether each mine was colonized independently and whether the low frequency of tolerance at uncontaminated sites represents ancestral variation or back migration from the mines. Previously, we identified a single large effect QTL for copper tolerance, *Tol1*, in line from the Copperopolis mine population (Wright *et al.* 2013). It is unknown whether this same locus contributes to copper tolerance in other mine adapted populations. We use field based transplant experiments, population genomic sequencing, and coalescence modeling to address three key questions: Does *Tol1* affect fitness in mine habitat, and if so, is it the primary determinate of survival in this habitat? Did convergent phenotypic evolution in independent mine populations occur via convergent genetic evolution? Did mine alleles at the candidate adaptation loci evolve from new beneficial mutations or via selection on pre-existing standing variation?

Tol1 **affects fitness in copper mine habitat**

We directly measured whether the mine and off-mine genotypes are locally adapted to their respective habitats using a reciprocal transplant experiment across multiple years (S. Figure 1). We found that plants from the mine population have greater fitness than off-mine genotypes in mine habitat, supporting the hypothesis of strong selection in this environment (Figure 1; S. Table 2-9). In the off-mine habitat, mine genotypes were also slightly favored compared to off-mine genotypes (Figure 1; S. Table 2-9). This suggests that high fitness in the mine environment does not cause a detectable trade-off in fitness in the off-mine environment. In order to measure whether the copper tolerance locus affects fitness, we conducted a reciprocal transplant experiment with introgression lines possessing alternate alleles at *Tol1*. We constructed introgression lines by reciprocally backcrossing lines with mine and off-mine alleles at the tolerance locus ($Tol1_{CC}$ or $Tol1_{MM}$) to mine (CCC_{BC}) and off-mine (MED_{BC}) genotypes (S. Figure 2). In the off-mine habitat, we found no significant effect of the *Tol1* locus in our CCC_{BC} introgression lines, (95% CI: 0.14, -0.36, p = 0.68, S. Table 6) which suggests that tolerant alleles at *Tol1* are neutral (or only weakly selected) in off-mine populations. Conversely, we found that mine alleles at *Tol1* greatly increases the probability of survival to flowering in the mine habitat (Figure 1; S. Table 2). The strength of selection (s) on *Tol1*, defined as the relative fitness of each allele, was significantly stronger in the off-mine genomic background (MED_{BC} s=0.69) compared to the mine genomic background (CCC_{BC} s=0.30). Additionally, the difference in fitness between parental lines, (s=0.90; S. Table 2) is greater than the fitness effect of *Tol1*. These two results support the hypothesis that additional loci contribute to fitness in the mine habitat. Our finding that the *Tol1* mine allele does not impose a measureable fitness cost in the off-mine habitat is consistent with either back-migration or ancestral variation as possible scenarios explaining the low frequency of copper tolerant plants in this environment. To distinguish between these two scenarios we need to measure segregating variants at *Tol1* in mine and off-mine populations.

Targeted assembly of *Tol1* region

To accurately measure genetic variation at *Tol1*, we need to first improve the assembly of this region. *Tol1* is located in a poorly assembled pericentromeric region of linkage group 9 in the M *guttatus* reference assembly, notably characterized by large amounts repetitive DNA and low recombination rates (Wright *et. al.* 2013). We sought to map all genomic scaffolds linked to *Tol1* by resequencing plants derived from our original Near Isogenic Line (NIL) mapping population (Wright *et al.* 2013). To accomplish this, we sequenced two pools of plants each composed of 100 plants homozygous for either tolerant (NIL_$Tol1_{TT}$) or nontolerant (NIL_$Tol1_{NTNT}$) alleles at the genetic marker, *Sc84_37kb*, located 0.32 cM from *Tol1* (Wright *et al.* 2013). We reason that that scaffolds in this region should, except in the case of a rare recombination event, be fixed for different haplotypes in the NIL_$Tol1_{TT}$ and NIL_$Tol1_{NTNT}$ pools. We estimate the recombination distance between genomic scaffolds and the *Sc84_37kb* marker as the frequency of SNPs in the two pools to used to define the two pools (S. Text). We identified eight scaffolds, comprising 4.9 Mb, within a somewhat enlarged search space of 0.80 cM (S. Figure 3; S. Table 10). We also identified small portions of other scaffolds that are in tight

linkage to *Tol1* (data not shown), suggesting problems with the v1.1 reference assembly and/or translocations between the reference and mapping lines. To address these assembly issues we sequenced a Copperopolis inbred line, CCC52, to 220X coverage and constructed a *de novo* genome assembly (S. Text, S. Table 11). We repeated our mapping effort using this *de novo* assembly and identified 53 CCC52 scaffolds, totaling 3.1Mb in length, within 0.80 cM of the *Sc84_37kb* marker (S. Text, S. Table 12). While most of the scaffolds are homologous to reference assembly *Tol1* linked scaffolds, eight of them are unique to the CCC52 assembly. Using this improved assembly, we sought to identify genes that exhibit significant differences in expression between NIL_*Tol1*$_{TT}$ and NIL_*Tol1*$_{NTNT}$ lines.

Differential expression of *Tol1* link genes

We measured gene expression in the NIL_*Tol1*$_{TT}$ and NIL_*Tol1*$_{NTNT}$ lines grown in a full-strength Hoaglands solution and Hoaglands enriched with 1.25 ppm CuSO4. We extracted mRNA from root tissue, sequenced each mRNA library to 140X coverage, and aligned all reads to a *de novo* transcriptome assembly (S. Text). We find that the majority of differentially expressed genes are due to allelic variation between the NIL_*Tol1*$_{TT}$ and NIL_*Tol1*$_{NTNT}$ lines, the copper treatment had negligible effect on gene expression. We identified 277 genes that are differentially expressed (DE) (Benjamini & Hockberg adjusted p-value < 0.1 and $\log_2$ fold change > 1.0) between lines in either treatment, 37 of these genes are located on the 53 *Tol1* linked scaffolds (S. Figure 4, S. Table 13). Some *Tol1* link DE genes - including a heavy metal ion transporter, a potassium ion transporter, and a multicopper oxidase – function to maintain intracellular ion homeostasis and may underlie the copper tolerance phenotype. Next, we sought to determine whether *Tol1* linked DE genes exhibit large differences in allele frequency in mine/off-mine population comparison.

Identification of candidate adaptation loci at *Tol1*

We sought to identify loci at *Tol1* that may underlie adaptation to the copper mine environment by measuring the genetic variation within and between two mine and two offmine populations. In a classic hard selective sweep, selection drives a single new beneficial mutation to near fixation, causing linked sites to hitchhike to high frequency, leading to a local reduction in genetic variation and large differences in allele frequency between the ancestral and derived populations (Maynard Smith & Haigh 1974, Barton 2000, Berg and Coop 2015). Alternatively, selection on pre-existing adaptive variants, which reside on multiple haplotypes, is predicted to produce a much smaller change in allele frequency at surrounding sites because no single haplotype goes to fixation (Innan and Kim 2004, 2008; Prezeworski *et al.* 2005). This event is sometimes termed a soft selective sweep (Messer and Petrov 2013). By comparing independent populations that have both adapted to similar environments, we can distinguish between selection on newly derived beneficial mutations or shared ancestral variation.

We sequenced pooled DNA from 20-31 natural isolates from two mine and two off-mine populations to 34-72X genome-wide coverage (S. Text, S. Table 1, S. Figure 5).

Genome wide nucleotide diversity is similar across all populations, (π = 0.0254 – 0.0268 ; S. Table 14). Genome-wide estimates of genetic differentiation between populations, measured with F_{st} and the *G* statistic (accounts for difference in coverage among sites and populations, S. Text) are higher between mine/off-mine compared to off-mine/off-mine population comparisons (F_{st} M/OM= 0.07-0.14; F_{st} OM/OM=0.02; S. Table 15). Elevated genome-wide estimates of differentiation between mine/off-mine populations is caused by a few, large effect genomic regions (ie. S. Figure 6) whereas the majority of the genome shows little differentiation. These results suggest that the mine populations did not undergo a strong bottleneck and/or gene flow readily occurs between populations in contrasting habitats.

We identified genomic regions with elevated difference in allele frequency at *Tol1*, as well as at other sites across the genome. Differentiation at *Tol1* is not elevated across all scaffolds in tight linkage to this locus (S. Figure 7, 8), as would be expected from a single hard selective sweep (S. Text), instead we find multiple loci with nearly fixed differences in allele frequency between mine/off-mine populations surrounded by large regions of shared genetic variation. We focus on loci that exceed the 0.5% genome-wide threshold of allelic differentiation in both mine/off-mine population comparisons (allelic differentiation is measured using the *G* statistic in 500bp bins, S. Text). We identify five distinct candidate adaptation loci, harboring seven genes, within the 53 *Tol1* link CCC52 scaffolds (S. Table 16). The signal of differentiation at these loci is relatively narrow (<20kb) with no significant differentiation at physically adjacent regions. To identify which, if any, of these genes maybe a *Tol1* candidate, we compared these genes to the 37 differential expression genes from the RNA-seq experiment. We find, one gene, a multicopper oxidase (MCO) is significantly differentiated in both experiments. MCO is located on a small scaffold (*s47895 - 1.7kb*) that is absent from the reference genome. Furthermore, we find that mine populations exhibit a 6X fold enrichment of aligned reads at MCO (S. Table 17), suggesting a recent tandem duplication of this gene is responsible for the 12X increase in expression observed in copper tolerant lines (S. Table 13). This gene represents an interesting *Tol1* candidate, but additional functional studies are required to determine whether it is responsible for the copper tolerance phenotype. Building upon the results from our transplant experiment, that *Tol1* accounts for some but not all of the variation in fitness in the mine environment (Figure 1), we next sought to identify additional loci with a population genetic signature of selection.

Genome-wide survey of candidate adaptation loci

We identified multiple genomic regions with elevated levels of differentiation between mine and offmine populations. As with *Tol1*, we focused on candidate adaptation loci that exceed the 0.5% genome-wide *G* statistic threshold in mine/off-mine population comparisons. Some loci are unique to one mine/off-mine population comparison (S. Table 18, 19), however, loci with the largest difference in allele frequency, 0.5% and 0.1% outliers, are much more likely to be shared between the mine/off-mine population comparisons than would be expected from random chance (S. Table 20; S. Figure 9). We rank the loci by their degree

differentiation (the quantity of physically adjacent 0.5% outlier bins) in both mine/off-mine population comparisons and find that *Tol1*, although it has a large effect on fitness, ranks as the fifth most differentiated region of the genome (Table 1). The identify of the top candidate adaptation loci in our screen are largely robust to changes in the genome-wide *G* statistic threshold (1%, 0.5%, 0.1%; S. Table 21), clustering of adjacent outlier bins (5kb, 10kb, 100kb; S. Table 22), the minimum number of SNPs per bin and measures of population differentiation (5 or 10 SNPs; *G* or F_{st}; S. Table 23). Nearly all candidate adaptation loci are restricted to single genes which encode proteins that function to maintain intracellular ion homeostasis (copper ion transporter, iron-zinc ion transporter, transmembrane ferric reductase, and potassium ion transporter;Table 1) and, in some cases, they have been implicated in heavy-metal tolerance (ie. Hanikenne *et al.* 2008 ; Grennan *et al.* 2009). The young age of mine populations and narrow interval of divergence (<30kb) suggests that beneficial variants pre-date the establishment of mine populations.

Selection on pre-existing genetic variation

The narrow interval of genetic differentiation at each candidate adaptation locus is inconsistent with expectations from a hard selective sweep. Given the intense selection on *Tol1* (s=0.69, Figure 1), a hard selective sweep would cause the mine allele to reach near fixation in ~20 generations and reduce genetic diversity by >50% for ~3Mb from the selected site (S. Text). In contrast, we observe that for *Tol1* outlier loci, heterozygosity recovers to 50% of background within a maximum of 20kb (S. Table 22). This quick recovery, despite strong selection, strongly suggests that the selected allele at *Tol1* was segregating in the ancestral population prior to the establishment of the mine populations. If this is true, we predict that each mine population would share the same haplotype at the candidate loci. In order to assess this hypothesis for the *Tol1* region, and all other candidate adaptation loci, we measured genetic differentiation between mine populations.

We find that three loci in tight linkage to *Tol1* and four additional unlinked candidate adaptation loci exhibit regions of very low genetic differentiation between mine populations (Table 2; i.e. Figure 2, S. Figure 9-15). This is inconsistent with independent sweeps of new mutations, but rather with selection on shared standing variation. Mine populations share the same *core haplotype* at these loci (extending from 1-18kb; Table 2) and outside this core, genetic differentiation between mine populations rises dramatically to levels comparable to differentiation between mine and off-mine populations (Figure 2). This peak-valley-peak pattern is additional evidence that the *core haplotype* was present in the ancestral population and recombined onto multiple genomic backgrounds prior to the establishment of the copper mines (Roesti *et al.* 2014). How old must the *core haplotypes* be to have an opportunity to recombine onto different genomic backgrounds prior to the initiation of mining in the region?

Age of core haplotypes

We estimate the age of the *core haplotypes* using the rate at which genetic diversity around the mine allele recovers to background levels (S. Text). We find that all *core haplotypes* predate the establishment of the copper mines (Table 2). Not surprisingly, the locus with the strongest signature of selection, MET, has the youngest *core haplotype* (639 generations), but even this haplotype existed in the ancestral Copperopolis population long before the establishment of the mine (150 years; Table 2). In contrast to the four unlinked *core haplotype* regions, we find three tightly linked *core haplotypes* at *Tol1*. This suggests these haplotypes were in linkage disequilibrium (LD) in the ancestral population, a hypothesis consistent with the result that each scaffold maps to <0.05 cM of the same position (S. Table 12), and LD was maintained as they rose to high frequency in each mine population. We estimate the age of the *core haplotypes* at the three *Tol1* linked genes to predate the establishment of the copper mines by hundreds-thousands of years (Table 2). The CDK and PUM haplotypes have been present in the ancestral Copperopolis population for 1.7-2.7K generations, whereas the MCO haplotype is much older (12.5K generations). Using the age of these haplotypes and the strength of selection on *Tol1*, we calculate that the CDK and PUM haplotypes were at 0.4-0.6% and the MCO at 4.5% frequency when the copper mines were first established (S. Table 24). These values are inflated for MCO because it is on a small scaffold with few informative SNPs (S. Figure 15), thus we consider the CDK and PUM estimates to better reflect the age and frequency of the selected allele at *Tol1*.

Conclusion

In this study, we investigate a classic example of rapid adaptation, colonization of heavy metal contaminated environments. We investigate the genetic control of a major copper tolerance QTL, *Tol1*. This locus has a significant effect on fitness in the mine environment (s = 0.69). We improve the assembly of the pericentromeric region harboring *Tol1*. Next we use whole genome RNA expression measurements and population genomic sequencing to identify candidate genes at this locus. We identify a single gene, a multicopper oxidase, with large differences in expression between tolerant and nontolerant lines and allele frequency between mine and off-mine populations. Interestingly, three tightly linked genes at *Tol1*, including MCO, possess *core haplotypes* that are strongly differentiated from the off-mine populations and highly similar between mine populations. This pattern suggests these *core haplotypes* predates the establishment of the mine populations. We develop a novel method to estimate the age of these haplotypes and find that they predate the establishment of the copper mines by 1000s of generations. Furthermore, we estimate the frequency of this haplotype in the ancestral population at the onset of mining (< 1%) is consistent with estimates of the frequency of tolerant plants in current populations inhabiting uncontaminated sites (Macnair *et al.* 1993). Looking throughout the genome we find that the top four candidate adaptation loci exhibit a *core haplotype* pattern similar to what we observe at *Tol1*. The function of the majority of these genes, to maintain intracellular ion homeostasis, suggests they contributed to adaptation in these copper contaminated environments. This work combining field-based estimates of selection, population genomics, and theoretical analyses provides thorough support

342 for the classic hypothesis that rapid adaptation in natural populations primarily
343 proceeds via selection on standing genetic variation (Bradshaw 1991).
344
345

Figures

Figure 1 (A) Mean probability of survive to flowering for all genotypes in mine plots in 2012 reciprocal transplant experiment (error bars = standard deviation). The probability a plant survives to flower, and the p-values testing whether there is a significant effect of genotype, is calculated from the predicted values using a logistic regression. The different genotypes are: parental mine, parental off-mine, F1 hybrid, and backcross lines (BC). Backcross lines are homozygous for Copperopolis mine allele at *Tol1* (red bars) or MED off-mine allele at *Tol1*, blue bars. **(B)** Mine plot, replicate 1. **(C)** Survival to flowering for all genotypes in offmine plots in 2012 reciprocal transplant experiment. Due to low germination, we have no data for the BC MED with off-mine allele at *Tol1*. **(D)** Off-mine plot, replicate 1.

Figure 2 Genetic differentiation at top ranked candidate adaptation locus on *Scaffold8*. (**A**) Genetic differentiation measured with G statistic for individual SNPs . Arrows depict annotated genes from *M. guttatus* genome V1.1. – Black arrows are genes within *core haplotype* **(B)** ratio: $\pi_{Mine}/\ \pi_{Off\text{-}mine}$ measured in 500bp windows. Orange dots/lines: Mine1 vs. Off-mine1. Blue dots/lines: Mine2 vs. Off-mine2. Green dots in **(A)**: Mine1 vs. Mine2. Green line in **(B)** ratio of pairwise π: $\pi_{Mine1+Mine2}/\ \pi_{Mine\text{-}Off\text{-}mine}$. Grey dashed line indicates genomic intervals with no SNP data.

Supplemental Figure 1 Local adaptation of parental genotypes in the mine habitat. A. Experimental plot 1 at the copper mine, soil Cu is 142ppm. B. Experimental off-mine plot 1, soil Cu is 12.9ppm. C. Copperopolis parental line flowering on May12, 2012. D. Probability of survival to flowering of parental genotypes in two years of reciprocal transplant experiments. Mine genotypes orange=2007; red=2012. Off-mine genotypes: purple=2007; blue=2012. E. Map of experimental plots at Napoleon Mine, 15 km from Copperopolis, CA. Orange shaded region is approximate outline of mine. Orange diamonds are two mine plots, and blue triangles are two off-mine plots.

Supplemental Figure 2 Generation of outbred BC lines of known genotype at *Tol1* locus. Backcrossing to Copperopolis parental lines, CCC, proceeded for four generations and backcross to Off-mine parental line, MED, proceeded for one generation. Afterwards heterozygous *Tol1* plants were identified with *Tol1* linked markers and selfed to produce plants homozygous for alternate alleles at *Tol1*. Independent backcross lines were then intercrossed to produce outbred plants of known genotype at *Tol1* locus.

Supplemental Figure 3 Assembly of *Tol1* region.
Plot of the recombination distance between a focal gene and the *Sc84_37kb* marker (indicated with blue arrow). Recombination distance estimated using all informative SNPs within focal gene (S. Text). *Tol1* linkage threshold (recombination distance of 0.008, or 0.8cM) is denoted with blue line. Black arrows at top indicate approximate location of v1.1 scaffolds.

Supplemental Figure 4 Differentially expressed transcripts in tolerant and non-tolerant lines.
Plot of log2 fold changes over mean expression. Blue horizontal lines represent fold change in expression of 1.0. Red dots indicate transcripts that have adjusted p-value less than 0.1. Triangles indicate transcripts that exceed 10 or -10 log2 fold change in expression. **A.** Difference in expression between NIL_*Tol1*$_{TT}$ and NIL_*Tol1*$_{NTNT}$ lines in Hoaglands + Cu treatment. We identified 93 and 107 transcripts down- and up-regulated respectively at an FDR of 0.1. **B.** Hoaglands treatment. We identified 119 and 105 transcripts to be down- and up-regulated respectively at an FDR of 0.1.

Supplemental Figure 5 Map of Copperopolis, Calaverous Cty, CA, USA.
Red circles are mine populations and blue circles denote off-mine populations. Population key: CCC – Copperopolis mine; MCN - McNulty mine; OM1 –Off-Mine1/ O'Byrnes Ferry Rd.; and OM2 Off-mine2/ Stage Coach Rd.

Supplemental Figure 6 Scaffold wide plots of allelic divergence between Copperopolis and Off-mine 1.
Plot of rank *G* statistic between Copperopolis and Off-mine 1 for scaffold 8. Each dot is 500bp bin - minimum 10 SNPs per bin- sliding window step size of 100bp. Black line is 0.5% genome wide threshold used to identify outlier bins.

Supplemental Figure 7 Genetic Differentiation at *Tol1* Linked Reference Genome Scaffolds
Order of scaffolds (reference genome v1.1) is same for S. Figure 3. Scaffold wide plots of allelic divergence, genome wide rank *G* statistic, between (A) Copperopolis / Off-mine 1, orange dots and (B) McNulty Mine / Off-mine 2, blue dots. Each dot is 500bp bin - minimum 10 SNPs per bin- sliding window step size of 100bp. Black line is 0.5% genome wide threshold used to identify outlier bins.

Supplemental Figure 8 Genetic Differentiation at *Tol1* Linked CCC52 Genome Scaffolds.
Population comparisons same as presented in S. Figure 7. Scaffolds are in numerical order from 1-53 according to name ID in S Table 12.

Supplemental Figure 9 Plot of allelic divergence between Copperopolis-Off-mine1 and McNulty-Off-mine2.
Candidate adaptation loci exceed 0.1% (dark blue), 0.5% (blue), 1.0% (light blue) rank G statistic threshold in both mine/off-mine comparisons. Allelic divergence measured as average G score for SNPs within 500bp window, each dot is a single 500bp bin with data for both population comparisons.

Supplemental Figure 10 Allelic divergence for candidate region: rank 2, scaffolds 148 and 198. These two scaffolds are adjacent in M. guttatus genome V. 2.0. (**A**) Genetic differentiation measured with G statistic for individual SNPs . Arrows depict

437 annotated genes from *M. guttatus* genome V1.1. – Black arrows are genes within
438 *core haplotype* **(B)** ratio: $\pi_{Mine}/\ \pi_{Off\text{-}mine}$ measured in 500bp windows. Orange
439 dots/lines: Mine1 vs. Off-mine1. Blue dots/lines: Mine2 vs. Off-mine2. Green dots in
440 **(A)**: Mine1 vs. Mine2. Green line in **(B)** ratio of pairwise π: $\pi_{Mine1+Mine2}/\ \pi_{Mine\text{-}Off\text{-}mine}$.
441 Grey dashed line indicates genomic intervals with no SNP data.
442
443 **Supplemental Figure 11** Allelic divergence for candidate region: rank 3, scaffold 4.
444 Figure details are provided in legend of S Figure 9.
445
446 **Supplemental Figure 12** Allelic divergence for candidate region: rank 4, scaffold 1.
447 Figure details are provided in legend of S Figure 9.
448
449 **Supplemental Figure 13** Allelic divergence for *Tol1* link candidate locus CDK,
450 scaffold 7467, CCC52 genome. Figure details are provided in legend of S Figure 9.
451
452 **Supplemental Figure 14** Allelic divergence for *Tol1* link candidate locus PUM,
453 scaffold 10201, CCC52 genome. Figure details are provided in legend of S Figure 9.
454
455 **Supplemental Figure 15** Allelic divergence for *Tol1* link candidate locus MCO,
456 scaffold 47895, CCC52 genome. Figure details are provided in legend of S Figure 9.
457
458
459
460
461

Tables

Table 1. Top ranked candidate adaptation loci

Rank	Scaffold	Length (kb)	N - Genes	N - Bins Exceed 0.5%	N - SNPs Fixed Diff.	Locus ID	Annotation
1	8	32.9	10	97	70\65	MET	Cytochrome P450; Methytransferase; Protease Inhibitor
2	148/ 198	24.1	2	39	0\32	FER	Transmembrane Ferric Reductase
3	4	7.7	2	25	0\0	ZNC	Zinc/Iron Ion Transporter
4	1	8.1	1	19	0\5	COP	Copper Ion Transporter
5A	84	1.8	1	11	0\0	MND	**Tol1 Link** - Mandelate Racemase
5B	84	5.8	1	8	3\0	CDK	**Tol1 Link** - Cyclin Dependent Kinase
6	44	2.5	1	9	0\0	SHD	Protein Binding, SH3 domain
7	80	6.2	1	6	0\0	DUF	Unknown Function
8	115	0.9	1	5	0\0	KIT	Potassium Ion Transporter
9	47	2.1	0	5	0\0	-	Intergenic
10	51	0.8	1	4	0\0	HSP	Heat Shock Protein, HSP90.

Candidate adaptation loci are composed on 500bp bins that exceed 0.5% rank G statistic threshold in both mine/off-mine population comparisons. Outlier bins are grouped together using the 10kb clustering algorithm. Loci are ranked by the number of outlier bins at a locus. The physical length of a candidate locus is defined by the region of overlap between both mine-off-mine population comparisons. The FER locus spans two adjacent scaffolds: 148 and 198. The number of SNPs with fixed differences is presented for CCC vs. Off-mine1 \ MCN vs. Off-mine2 comparisons. Gene annotation from *M. guttatus* genome v1.1, phytozome.net. Annotation of loci that span multiple genes is restricted to genes with strongest signal of differentiation.

Table 2 –Age of core haplotypes at candidate adaptation loci.

Locus ID	Core Haplotype			Coalescence Time				Annotation
	Scaffold	Length (kb)	Num. Genes	CCC	MCN	CCC/ OM1	MCN/ OM2	
MET	8	19.6	4	639	1191	1901	5699	Cytochrome P450; Methytransferase; Protease Inhibitor; Unknown Func.
FER	148/198	7.06	1	1533	2022	2285	1725	Transmembrane Ferric Reductase
ZNC	4	2.4	1	17191	960	1622	1537	Zinc/Iron Ion Transporter
COP	1	1.27	1	5062	3119	3956	3931	Copper Ion Transporter
***Tol1* Linked Genes**								
CDK	7467*	1.26	1	1768	7840	19121	15778	Cyclin Dependent Kinase
PUM	10201*	8.51	1	2789	10162	19264	23900	Pumilio-family, RNA binding
MCO	47895*	1.06	1	12543	90664	340560	56920	Multicopper oxidase

The length and age of the core haplotypes are estimated by modeling the rate at which $\pi_{M1\text{-}M2}$ (pairwise π between mine populations) increases with distance from the selected site using the program, MSSEL, see supplemental text for additional details. The length of the core haplotype is estimated using equation 1 in S. Text. Coalesence time in presented in generations for within population (CCC; MCN) and between population (CCC/OM1; MCN/OM2) analyses. *Scaffold ID corresponds to CCC52 de novo assembly.

494
495 **References**
496 Al-Hiyaly, S. A. K., McNeilly, T., & Bradshaw, A. D. (1990). The Effect of Zinc
497 Contamination from Electricity Pylons. Contrasting Patterns of Evolution in Five
498 Grass Species. New Phytologist, 114(2), 183–190.
499
500 Al-Hiyaly, S. A. K., McNeilly, T., Bradshaw, A. D., & Mortimer, A. M. (1993). The effect
501 of zinc contamination from electricity pylons. Genetic constraints on selection for
502 zinc tolerance. Heredity, 70, 22-32.
503
504 Allen, W. R., & Sheppard, P. M. (1971). Copper Tolerance in Some Californian
505 Populations of the Monkey Flower, *Mimulus guttatus*. Proc. R. Soc. B., 177(1047),
506 177–196.
507
508 Antonovics, J., Bradshaw, A. D., & Turner, R. G. (1971). Heavy metal tolerance in
509 plants. Adv. Ecol. Res., 7:1–85.
510
511 Aubury, L. E. (1905). The copper resources of California (No. 23). WW Shannon,
512 Superintendent State Print.
513
514 Barton, N. H. (2000). Genetic hitchhiking. Phil. Tran. R. Soc. B., 355(1403), 1553-
515 1562.
516
517 Berg, J. J., & Coop, G. (2015). A Coalescent Model for a Sweep of a Unique Standing
518 Variant. Genetics. 201.
519
520 Bradshaw, A. D. (1984). The importance of evolutionary ideas in ecology and vice
521 versa. Evolutionary ecology, 1-25.
522
523 Bradshaw, A. D. (1991). The Croonian Lecture, 1991. Genostasis and the limits to
524 evolution. Phil. Tran. R. Soc. B., 333(1267), 289–305.
525
526 Brenchley, W. E. (1958). The Park Grass Plots at Rothamsted 1856-1949. Rep.
527 Rothamsted Exp. Stn.
528
529 Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood,
530 J., et al. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of
531 Ectodysplasin alleles. Science, 307(5717), 1928–1933.
532
533 Endler, J. A. (1977). Geographic Variation, Speciation and Clines (p. 268). Princeton
534 University Press.
535
536 Grennan, A. K. (2009). Identification of genes involved in metal transport in plants.
537 Plant Physiology, 149(4), 1623–4.
538

539 Hanikenne, M., Talke, I. N., Haydon, M. J., Lanz, C., Nolte, A., Motte, P., et. al. (2008).
540 Evolution of metal hyperaccumulation required cis-regulatory changes and
541 triplication of HMA4. Nature, 453(7193), 391–395.
542
543 Heliconius Genome Consortium. (2012). Butterfly genome reveals promiscuous
544 exchange of mimicry adaptations among species. Nature, 487(7405), 94-98.
545
546 Innan, H., & Kim, Y. (2004). Pattern of polymorphism after strong artificial selection
547 in a domestication event. Proc. Nat. Acad. Sci.101(29), 10667-10672.
548
549 Innan, H., & Kim, Y. (2008). Detecting local adaptation using the joint sampling of
550 polymorphism data in the parental and derived populations. Genetics, 179(3), 1713-
551 1720.
552
553 Joron, M., & Mallet, J. L. (1998). Diversity in mimicry: paradox or paradigm? Trends
554 Eco. & Evol, 13(11), 461-466.
555
556 Karasov, T., Messer, P. W., & Petrov, D. A. (2010). Evidence that Adaptation in
557 *Drosophila* Is Not Limited by Mutation at Single Sites. PLoS Genetics, 6(6),
558 e1000924.
559
560 Kruckeberg, Arthur R. (1951) Intraspecific variability in the response of certain
561 native plant species to serpentine soil. Amer. Jour. Botany (1951): 408-419
562
563 Macnair, M. R. (1987). Heavy metal tolerance in plants: A model evolutionary
564 system. Trends Eco. & Evol, 2(12), 354–359.
565
566 Macnair, M. R., Smith, S. E., & Cumbes, Q. J. (1993). Heritability and distribution of
567 variation in degree of copper tolerance in Mimulus guttatus at Copperopolis,
568 California. Heredity, 71(5), 445–455.
569
570 Maynard-Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene.
571 Genetics Research, 23(01), 23–35.
572
573 Messer, P. W., & Petrov, D. A. (2013). Population genomics of rapid adaptation by
574 soft selective sweeps. Trends Eco. & Evol, 28(11), 659–69.
575
576 Prezeworski, M., Coop, G., & Wall, J. D. (2005). The signature of positive selection on
577 standing genetic variation. Evolution, 59(11), 2312-2323.
578
579 Roesti, M., Gavrilets, S., Hendry, A. P., Salzburger, W., & Berner, D. (2014). The
580 genomic signature of parallel adaptation from shared genetic variation. Molecular
581 Ecology, 23(16), 3944-3956.
582
583 Schluter, D. (2000). The Ecology of Adaptive Radiation. Oxford, England: Oxford
584 University Press.

Symeonidis, L., McNeilly, T., & Bradshaw, A. D. (1985). Differential tolerance of three cultivars of Agrostis capillaries l. to cadmium, copper, lead, nickel and zinc. New Phytologist, 101(2), 309–315.

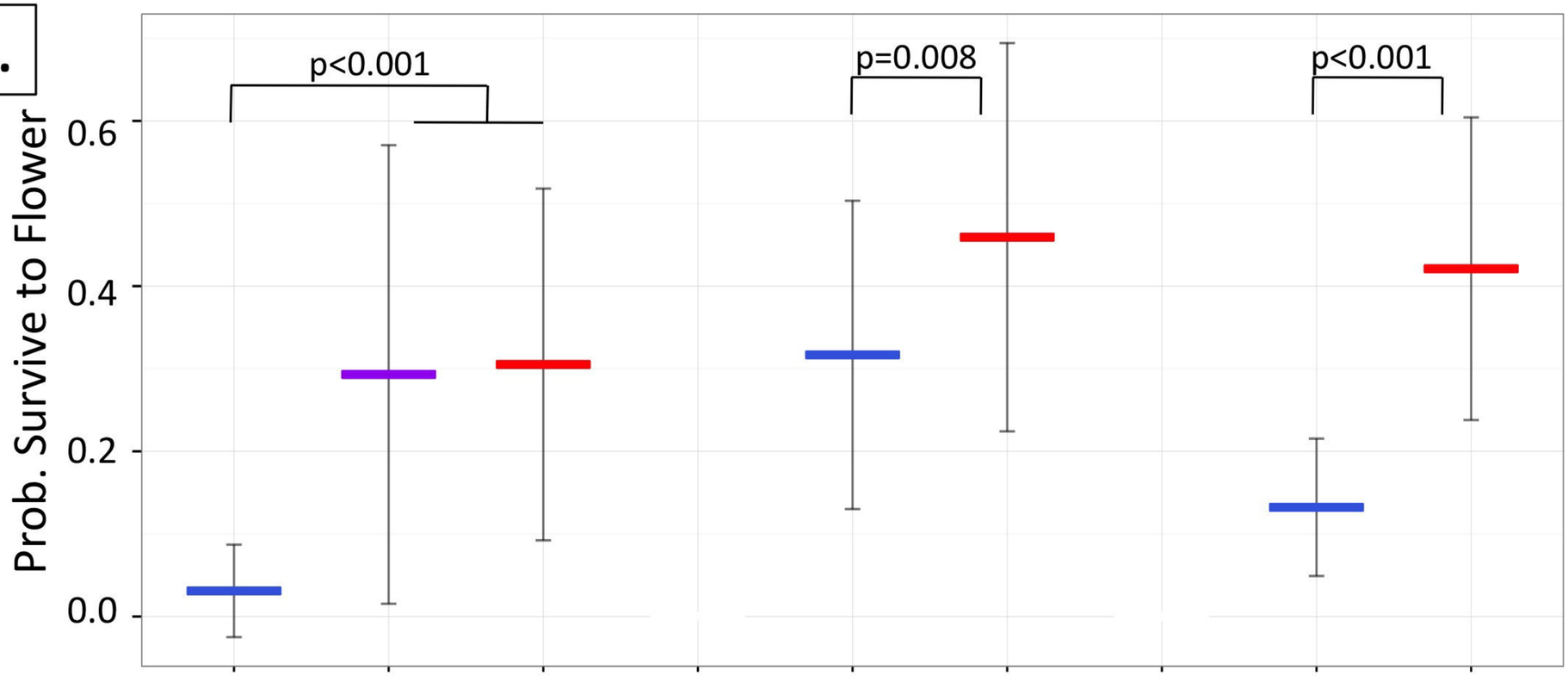
Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., … Deloukas, P. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. Nature Genetics, 39(1), 31–40.

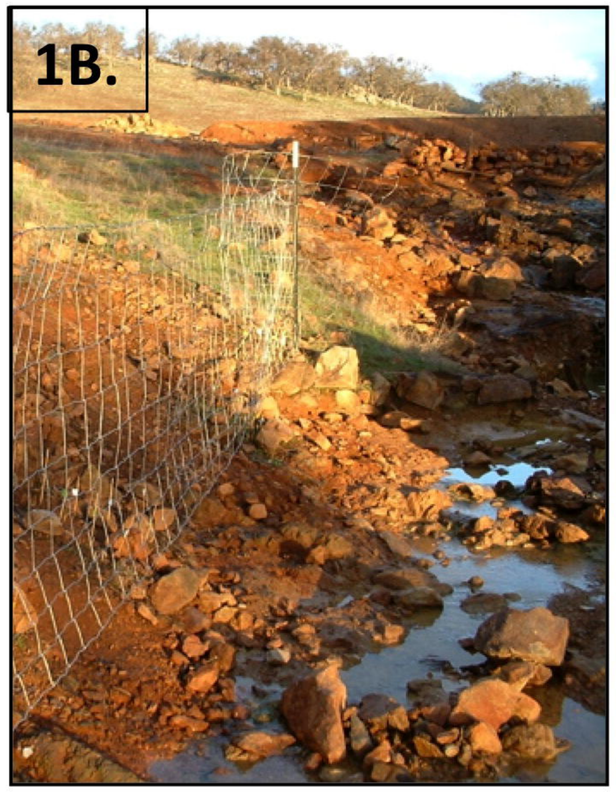
Turner, R. G. (1969). Heavy metal tolerance in plants. Vol. 9, 399-410. Blackwell Oxford.

Walley, K. A., Khan, M. S. I., & Bradshaw, A. D. (1974). The potential for evolution of heavy metal tolerance in plants. I. Heredity, 32, 309-319.

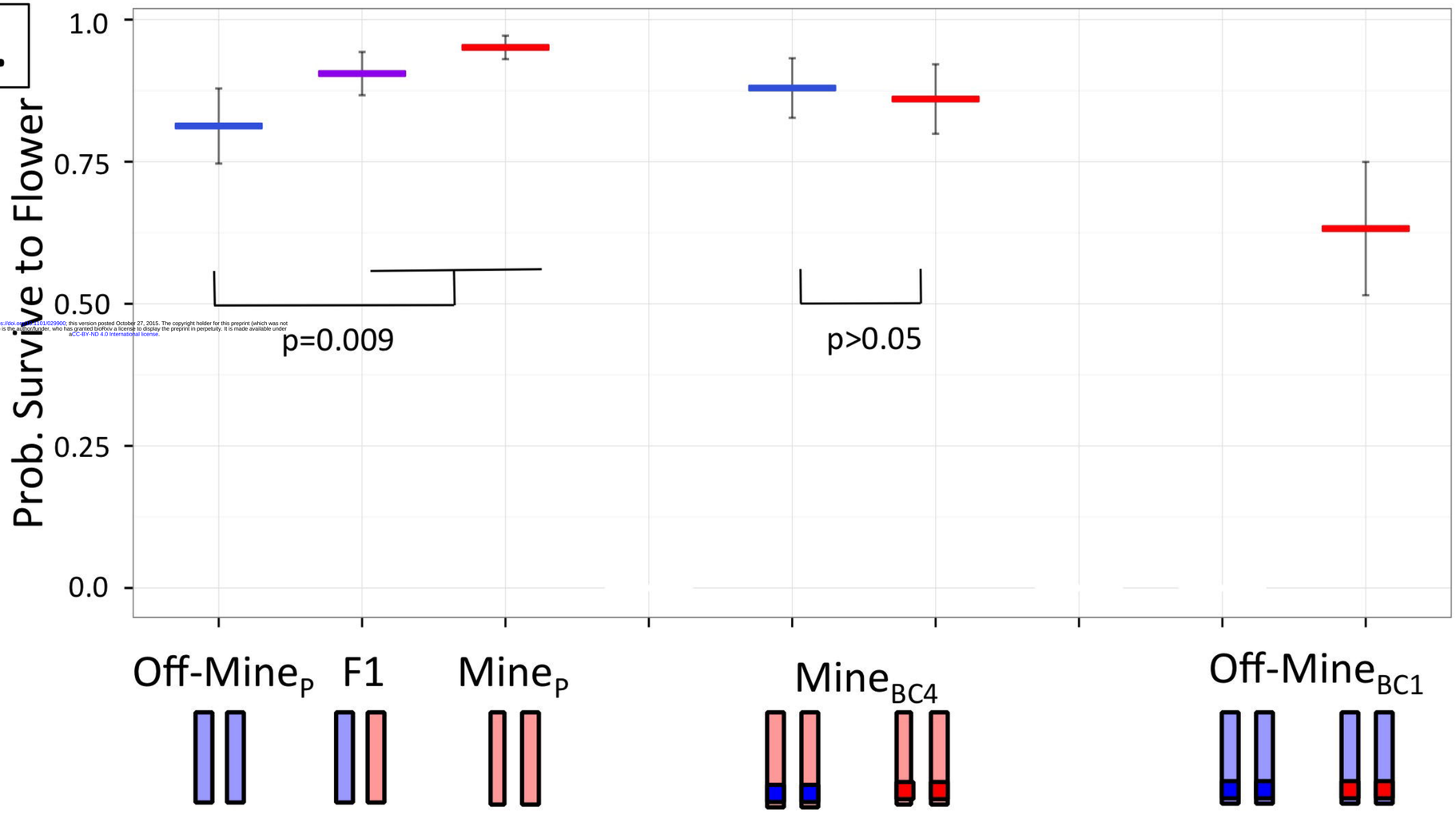
Wright, K. M., Lloyd, D., Lowry, D. B., Macnair, M. R., & Willis, J. H. (2013). Indirect evolution of hybrid lethality due to linkage with selected locus in *Mimulus guttatus*. PLoS Biol, 11(2), e1001497.

1A.





1C.



2A.

Rank G Statistic

0.001

0.005

0.010

0.100

1.000

2B.

$\pi_{MINE}/\pi_{OFF\text{-}MINE}$

2.0

1.0

0.0

260

280

300 (kb)

320

340

S Figure 1

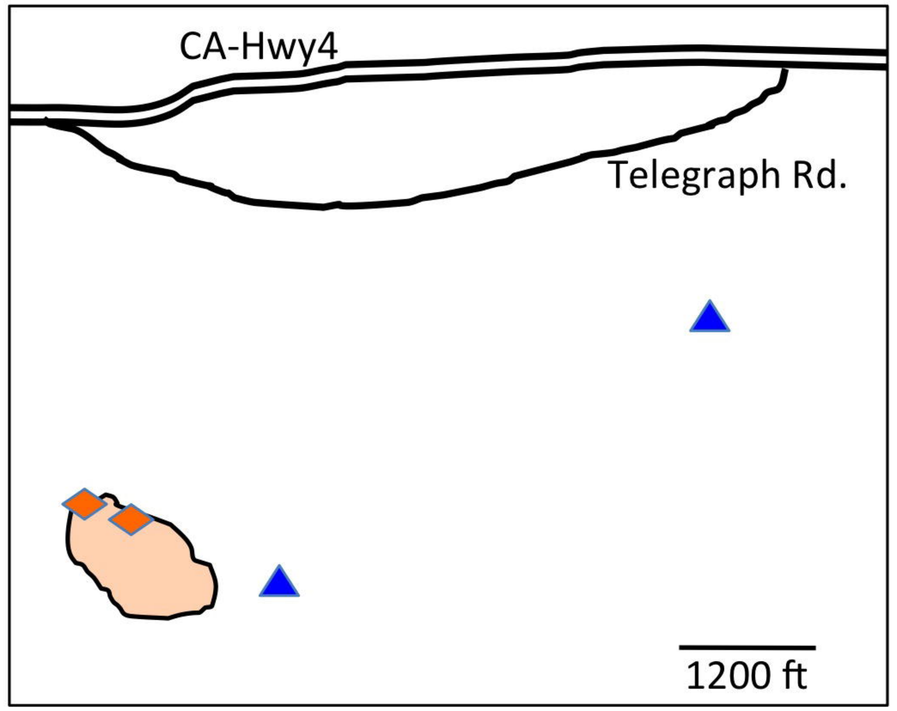
A.

B.

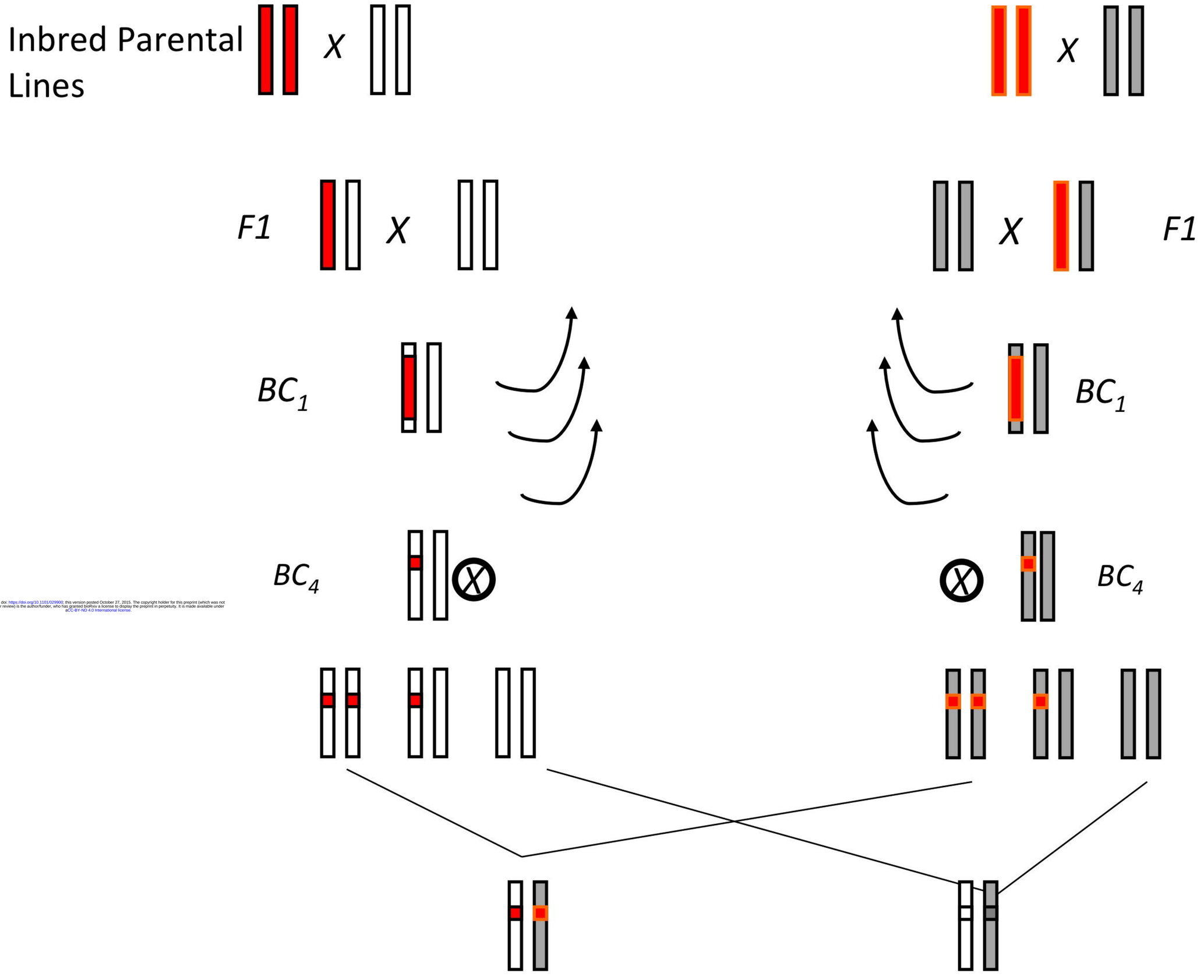
C.

D.

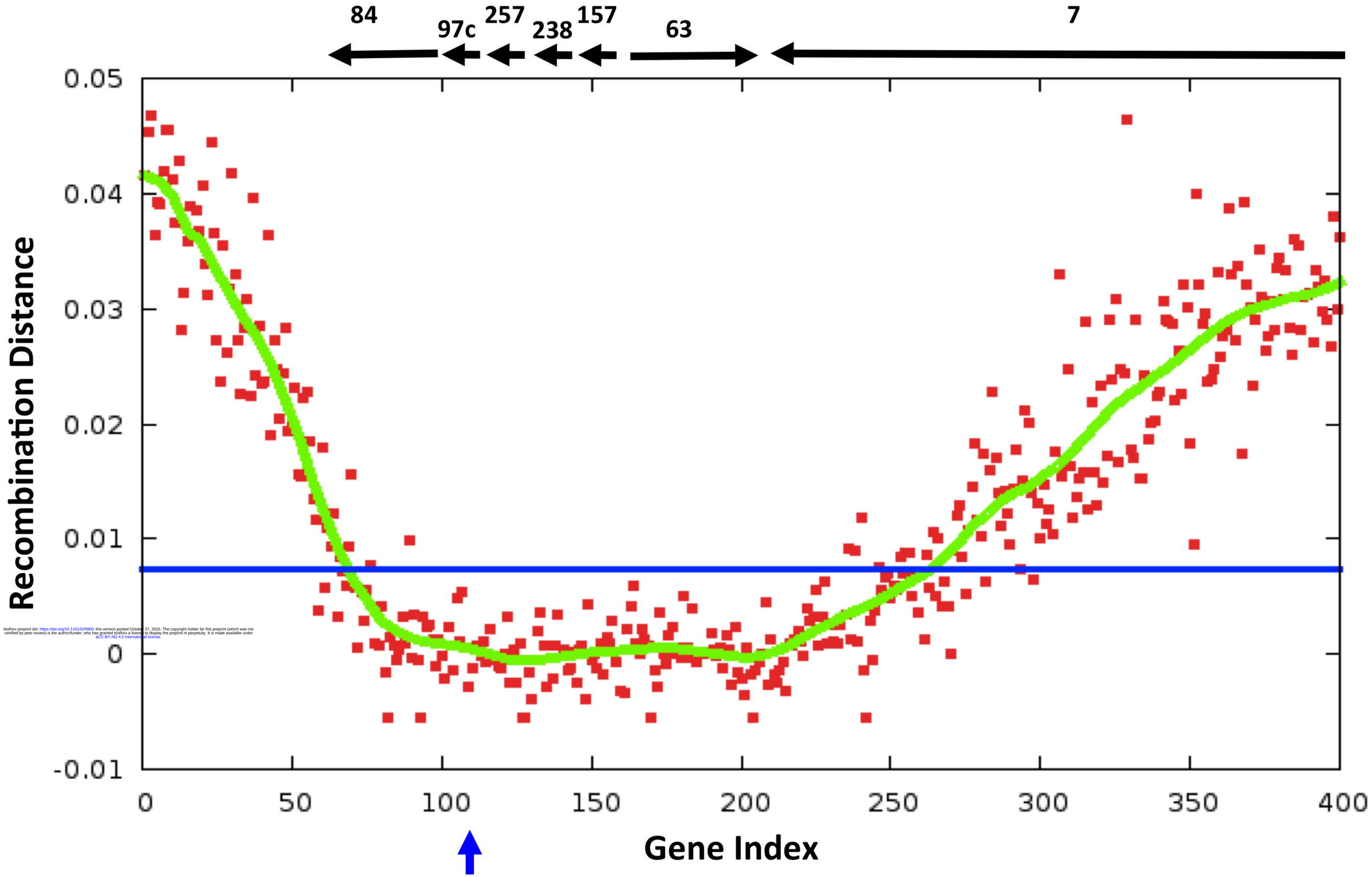
E.



S Figure 2

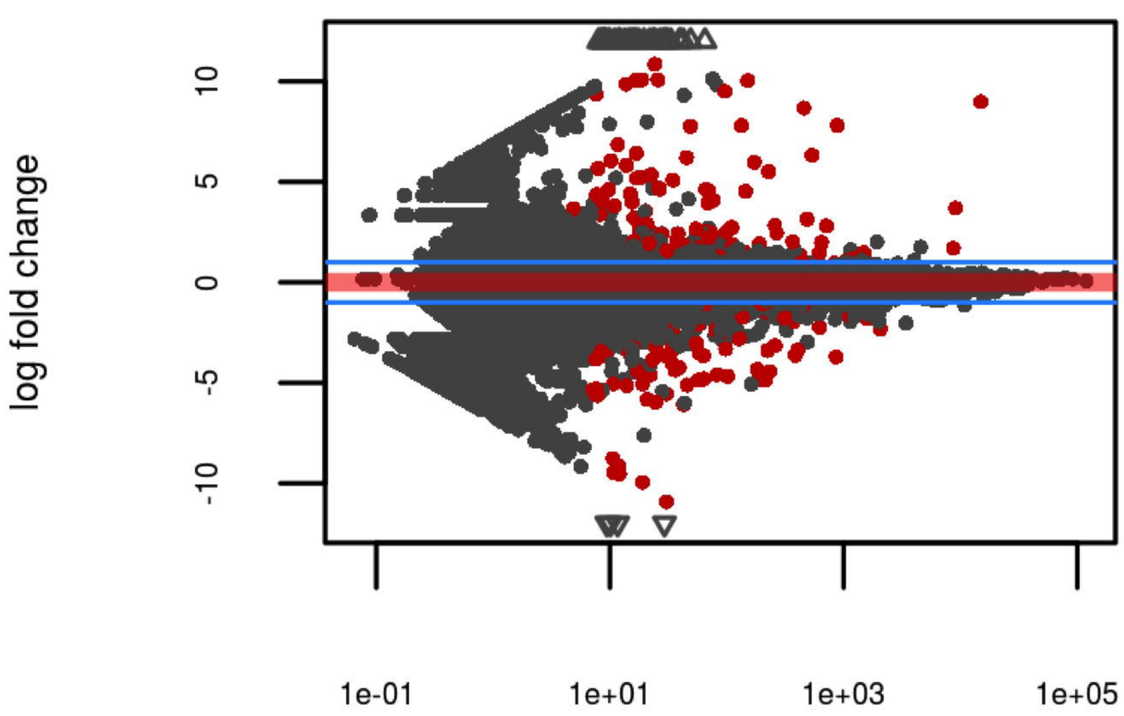


S Figure 3

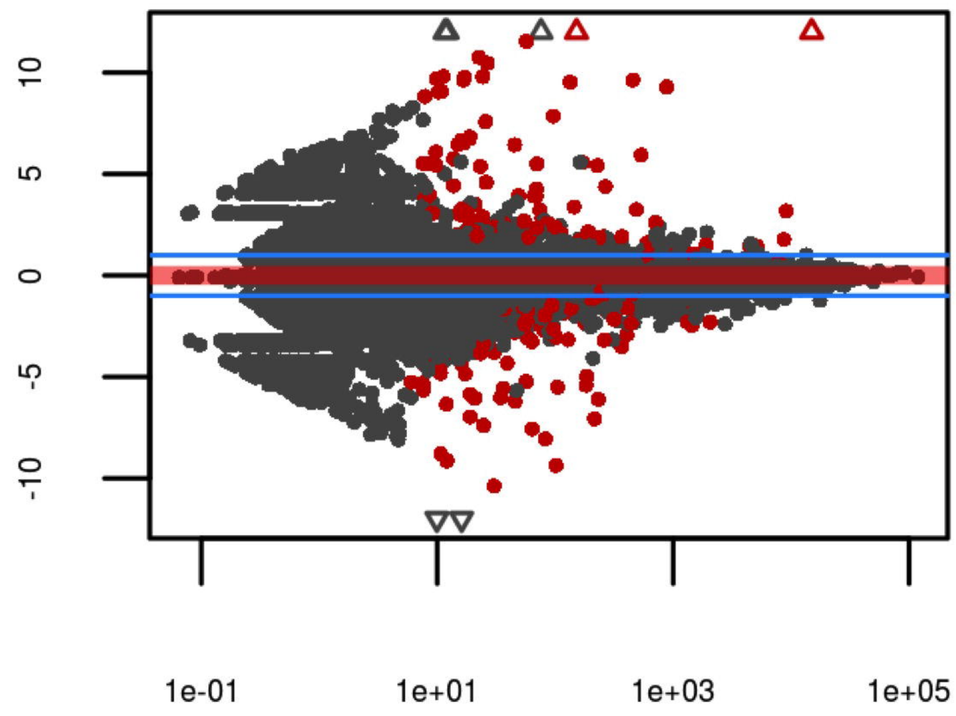


S Figure 4

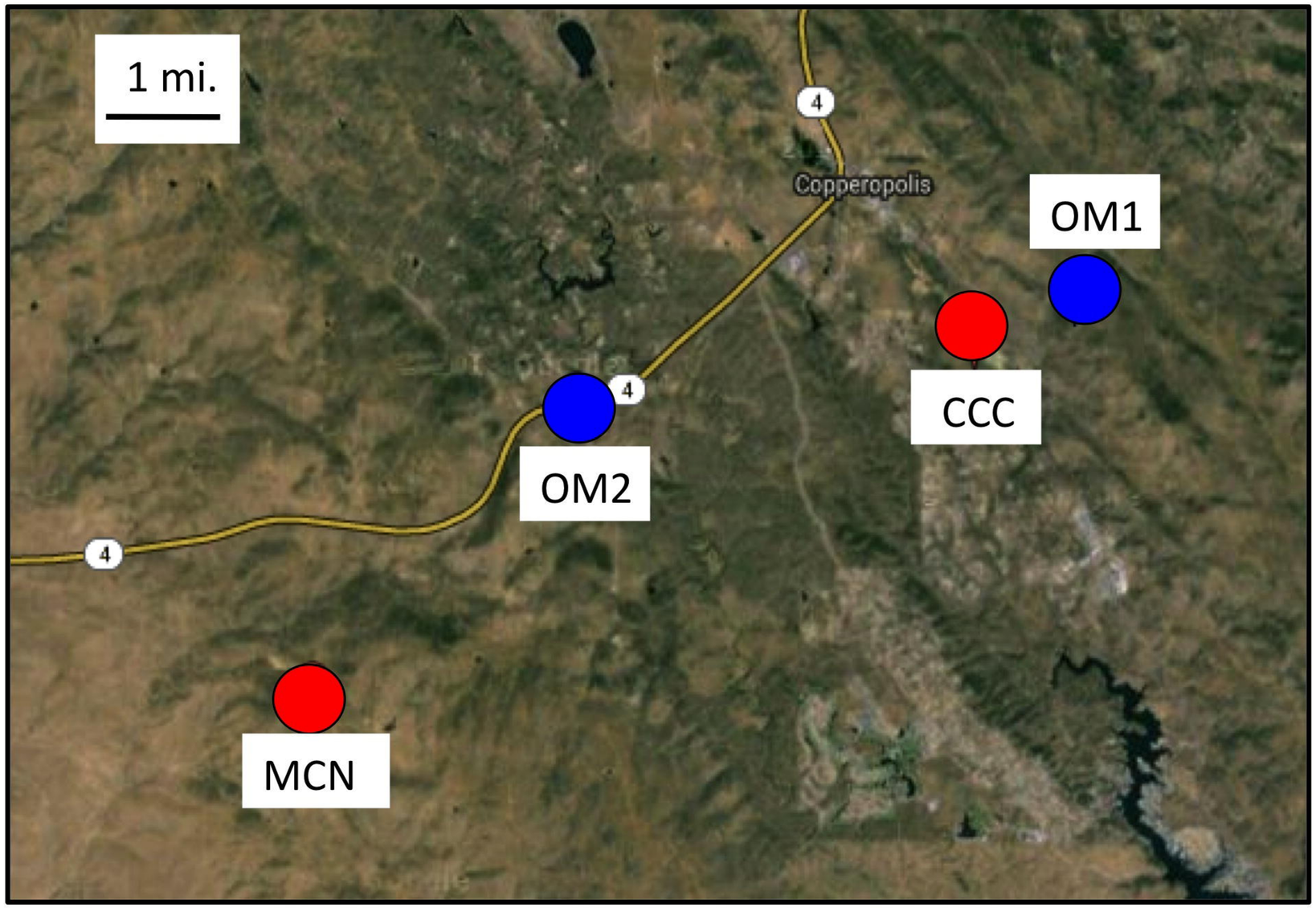
A.



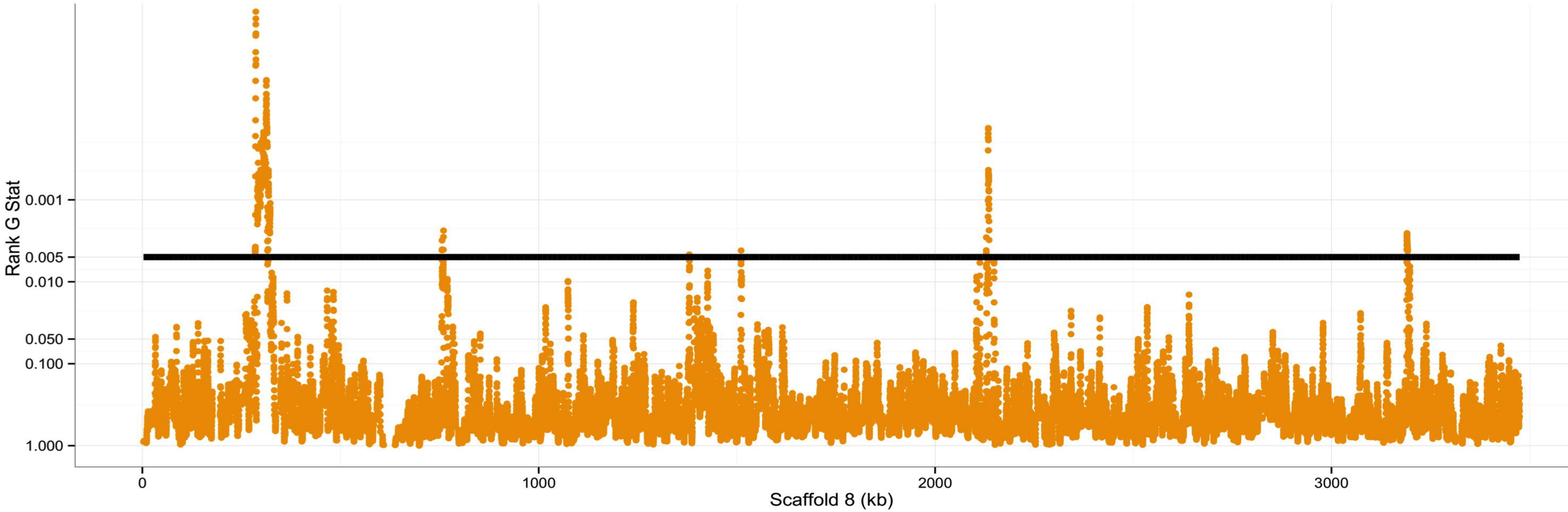
B.



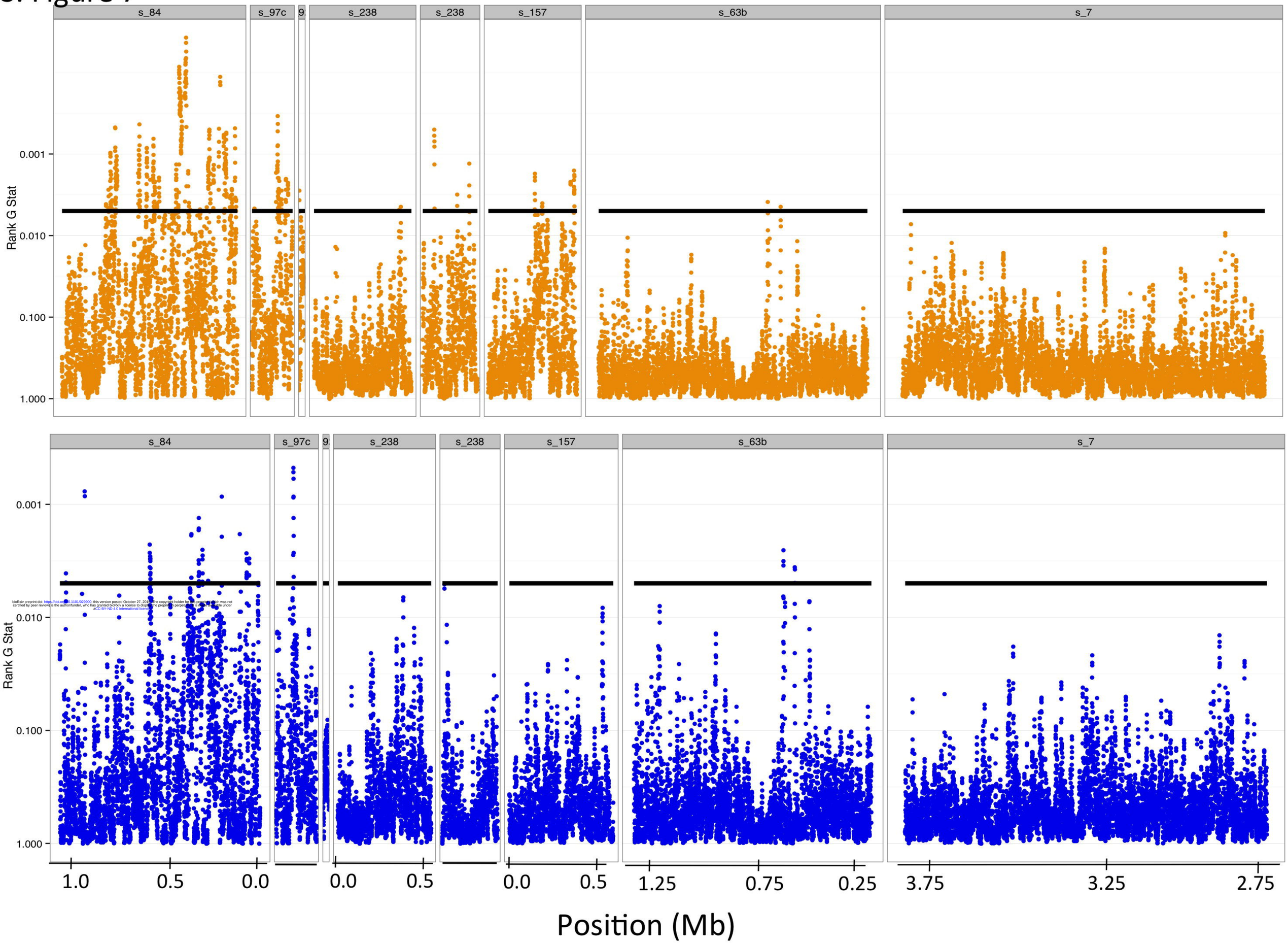
S Figure 5



S Figure 6

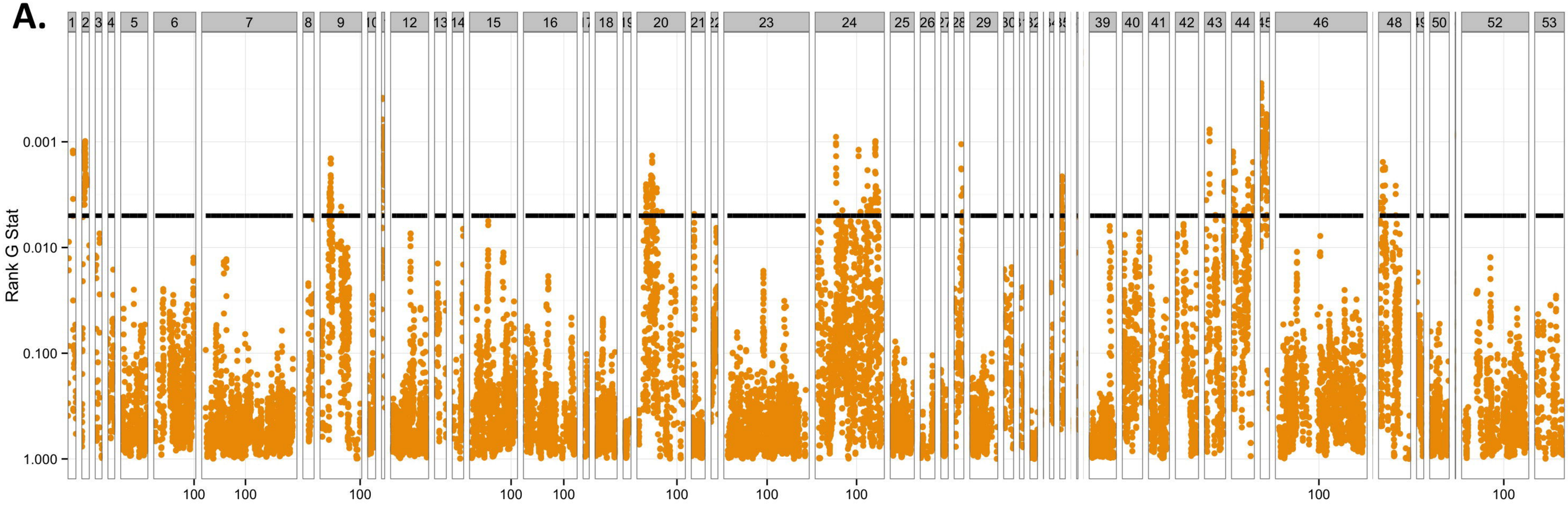


S. Figure 7

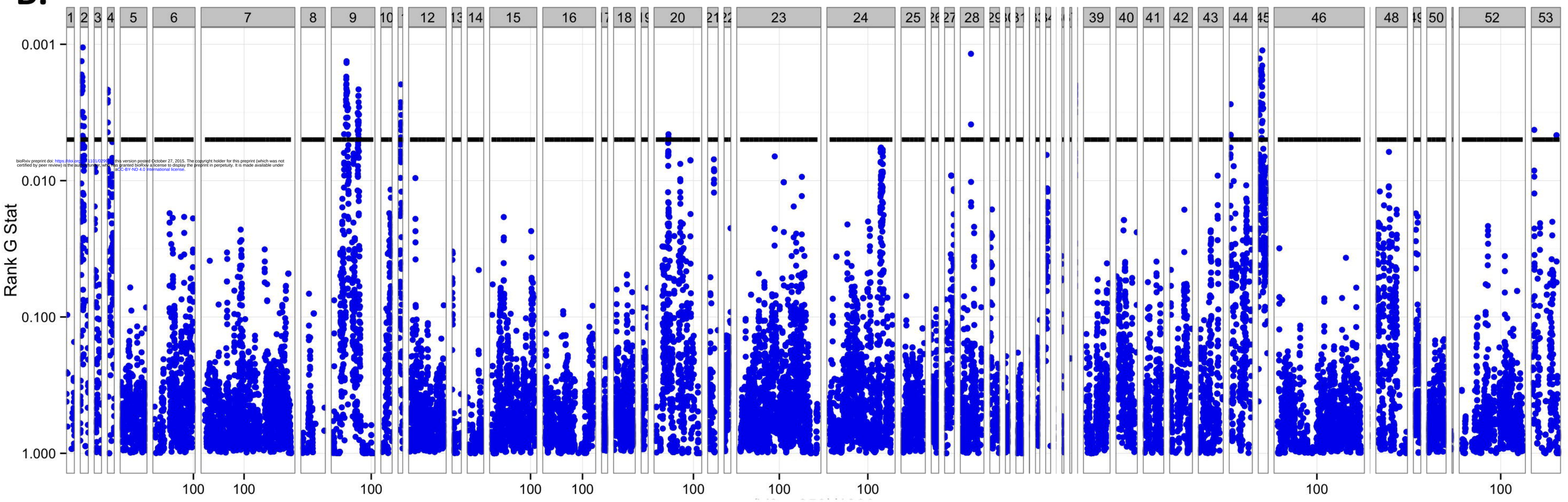


S. Figure 8

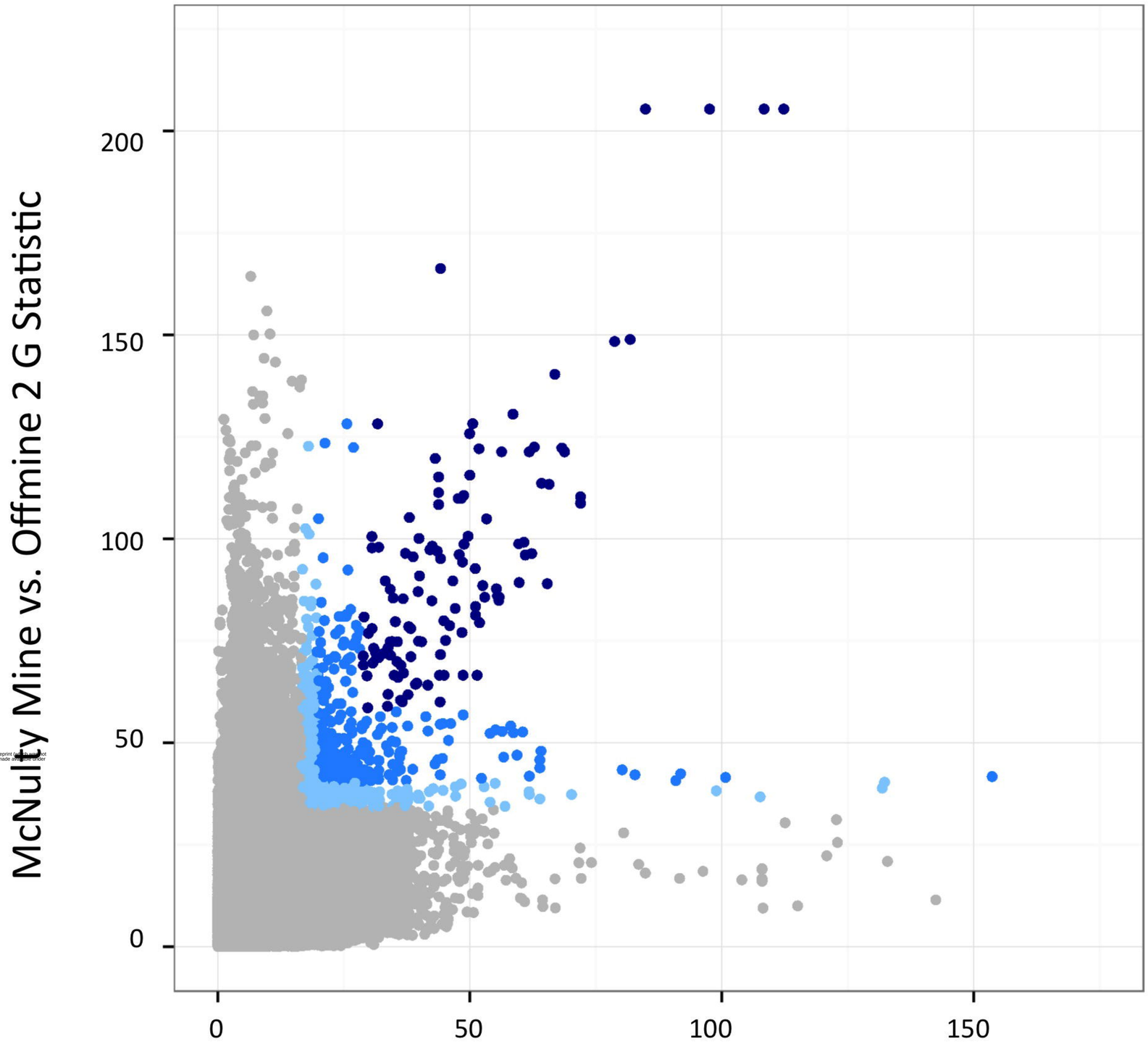
A.



B.

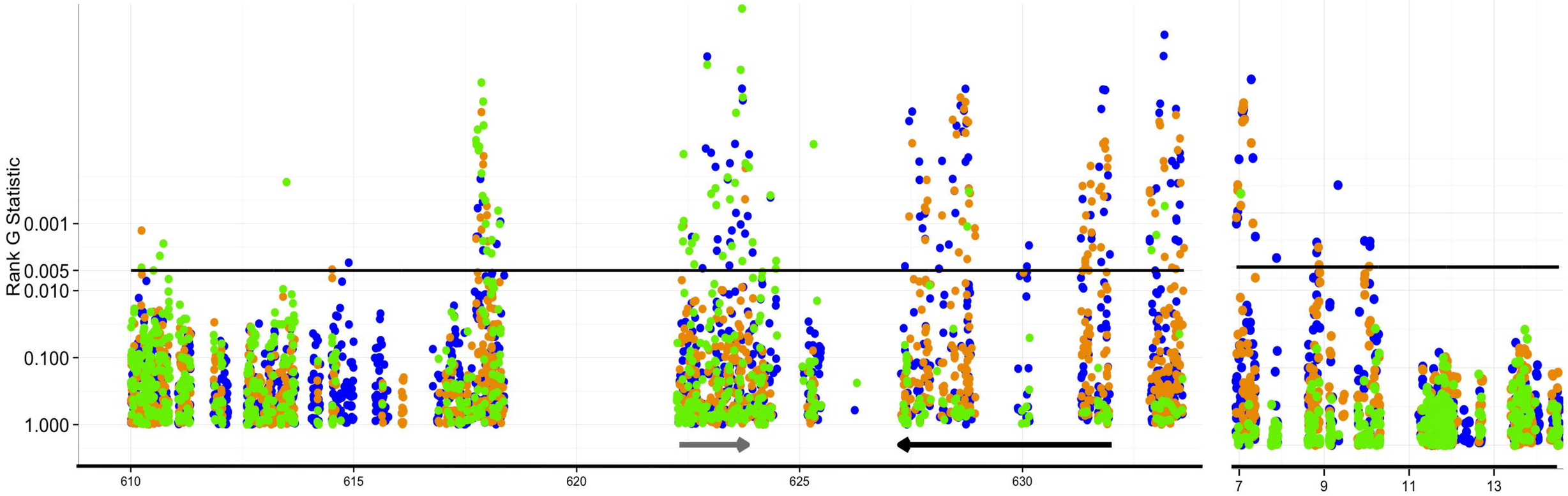


S Figure 9

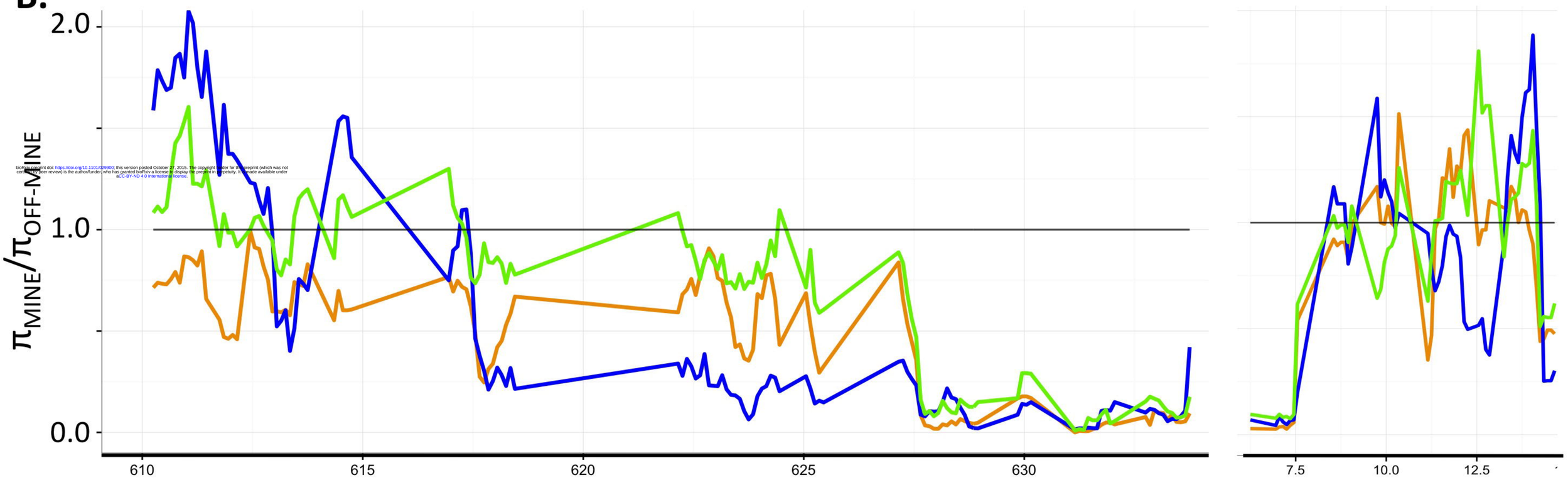


S Figure 10

A.



B.

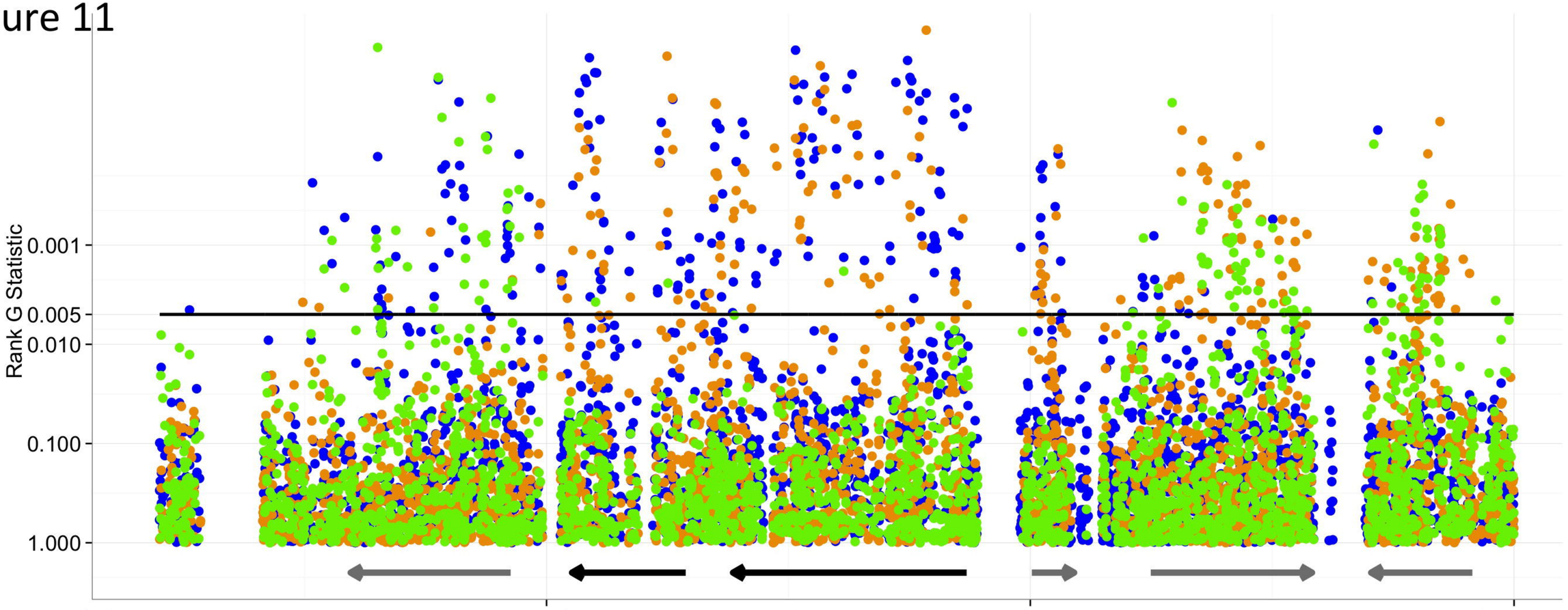


Scaffold 148

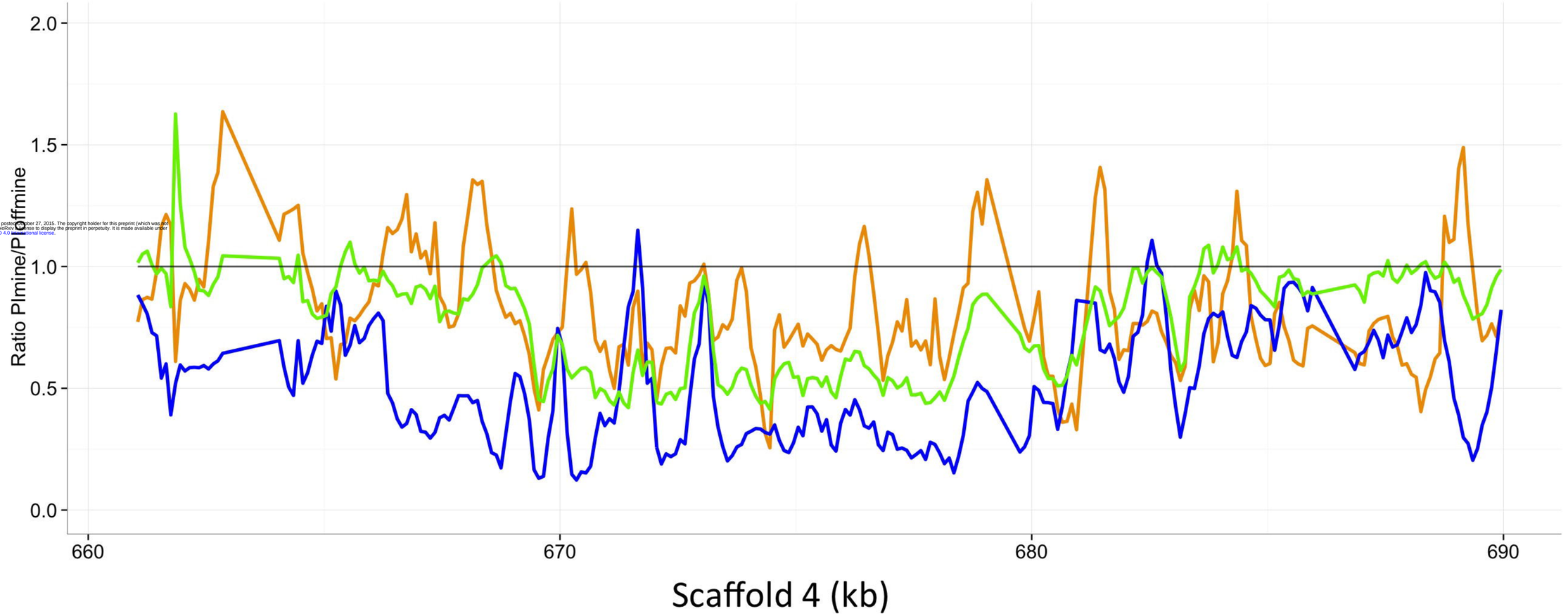
Scaffold 198

S Figure 11

A.

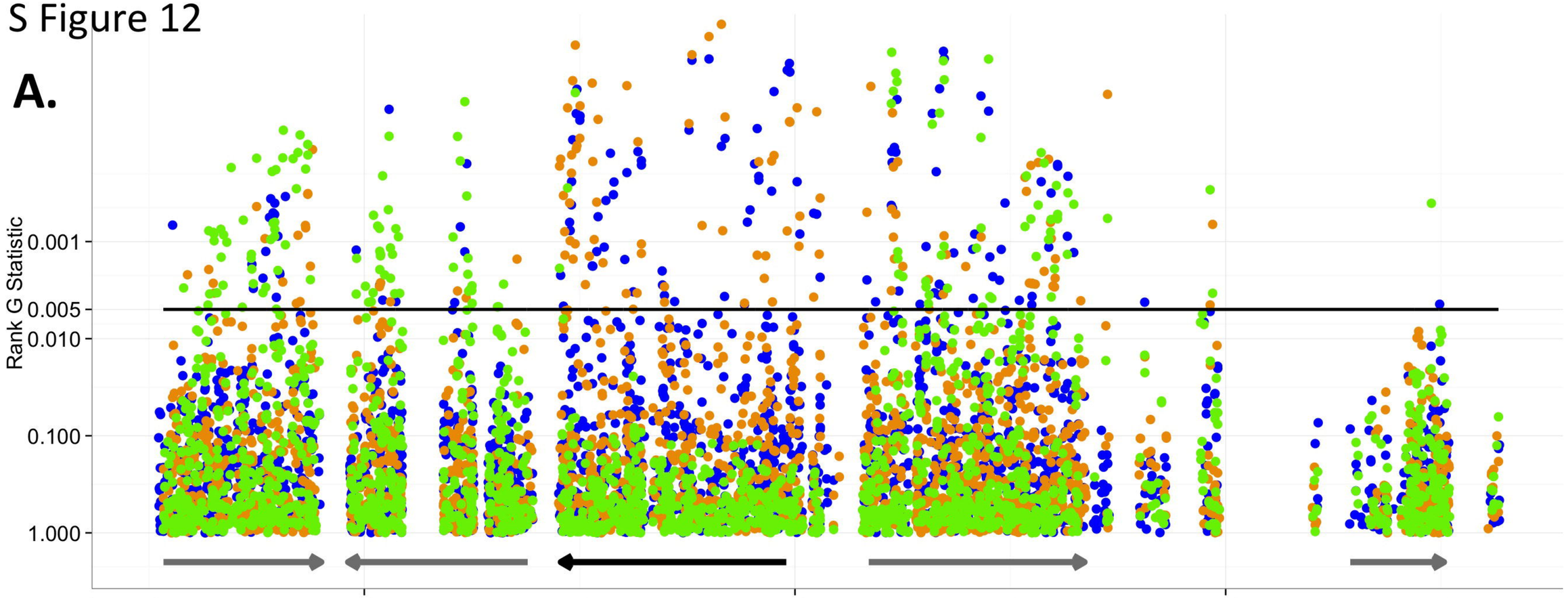


B.

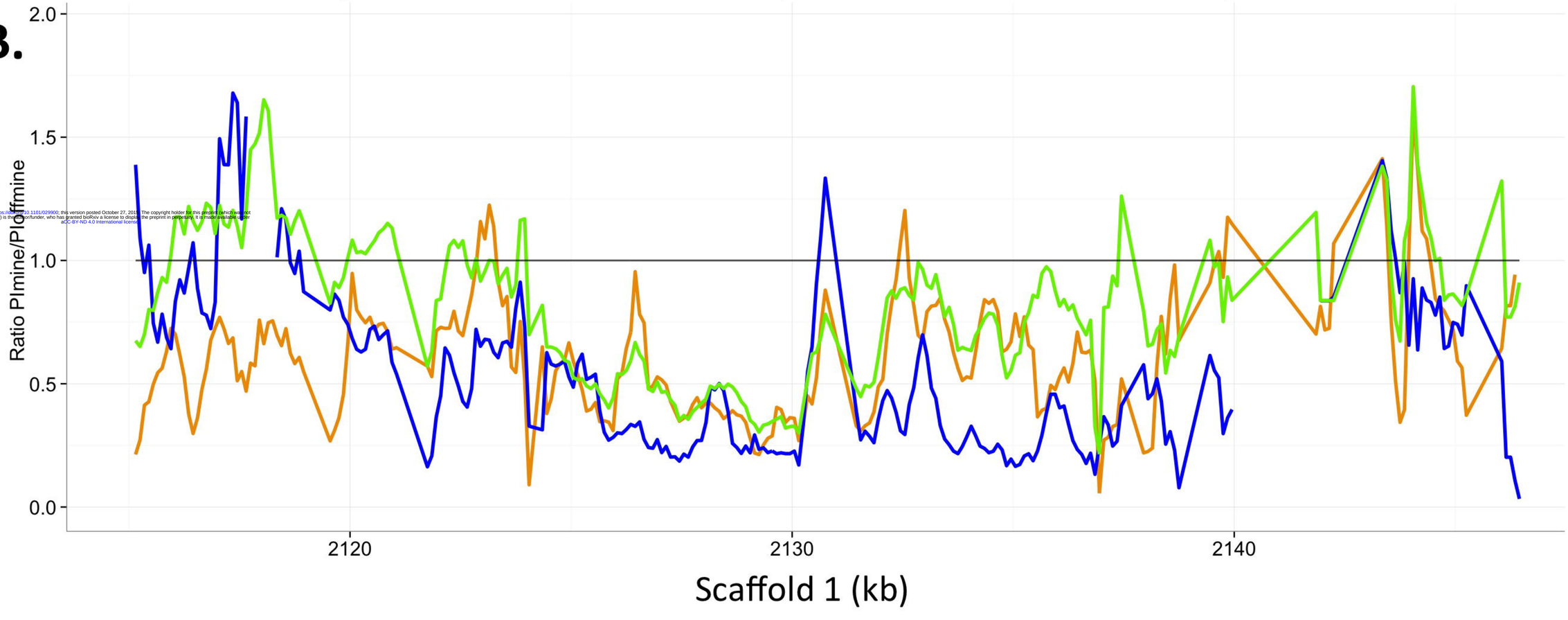


S Figure 12

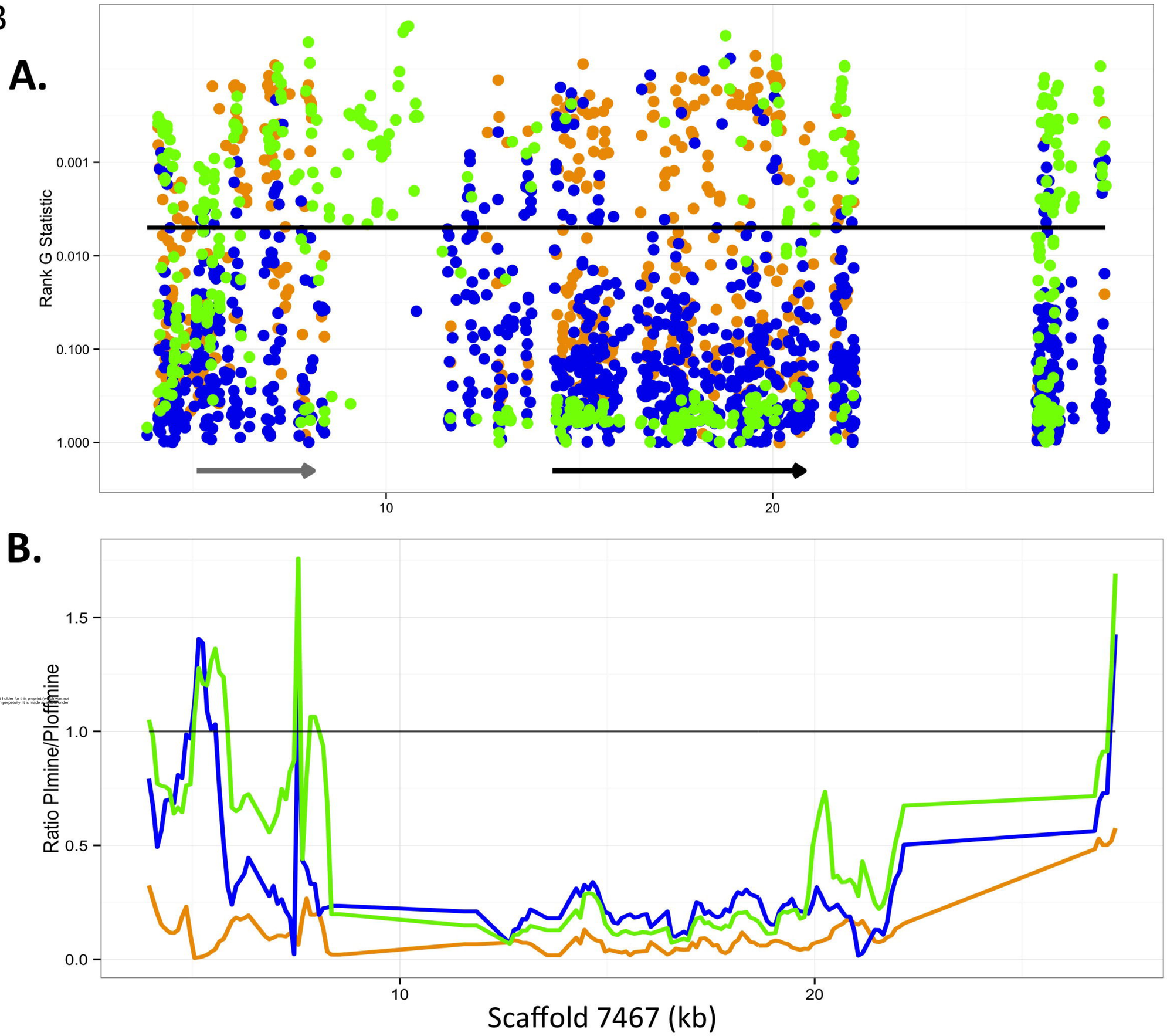
A.



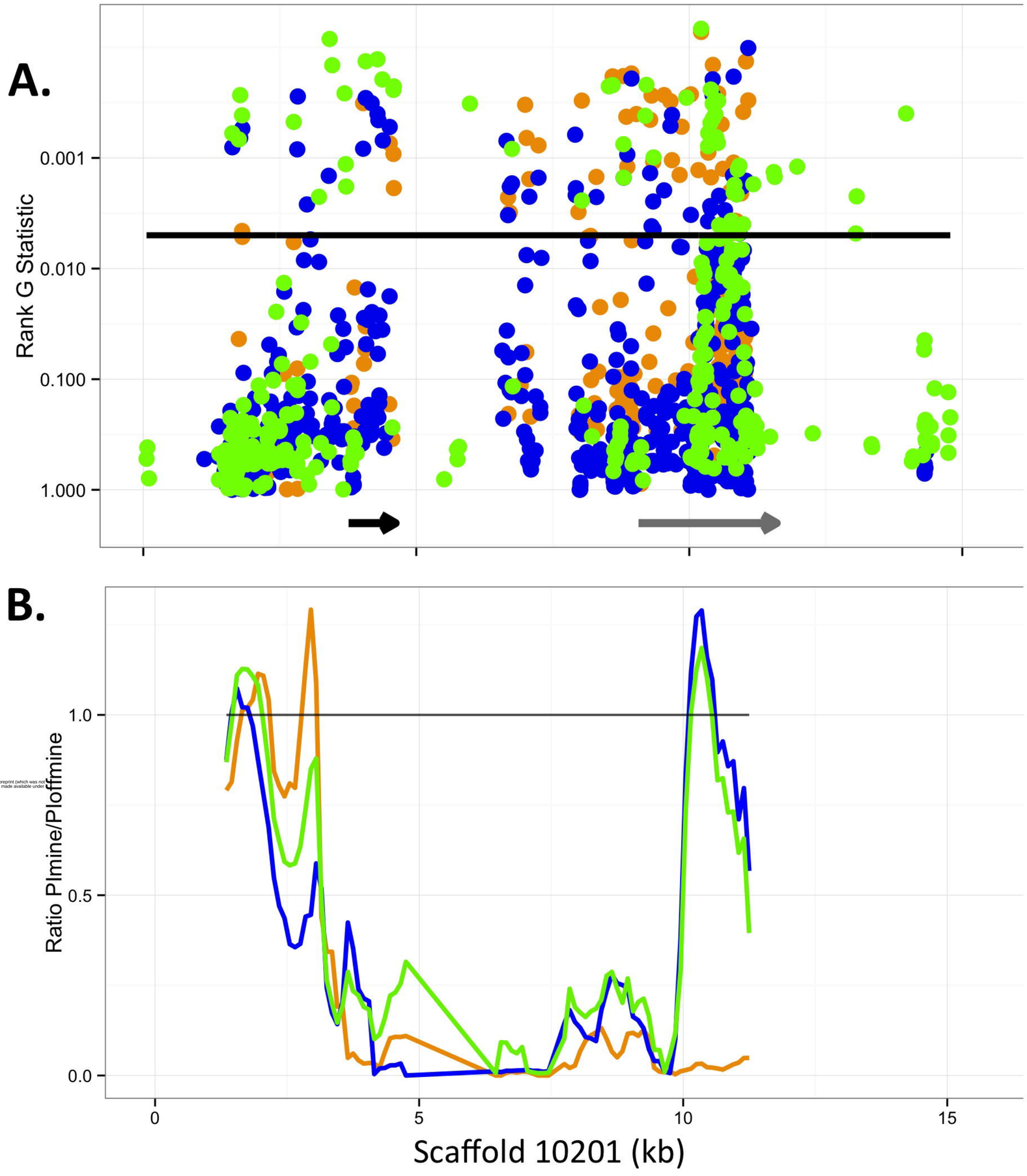
B.



S Figure 13



S Figure 14



S Figure 15