SI Appendix

Homoeologous exchanges occur through intragenic recombination generating novel transcripts and proteins in wheat and other polyploids

Zhibin Zhang, Xiaowan Gou, Hongwei Xun, Yao Bian, Xintong Ma, Juzuo Li, Ning Li, Lei Gong, Moshe Feldman, Bao Liu* and Avraham A. Levy*

*Corresponding authors: e-mail: baoliu@nenu.edu.cn (BL), avi.levy@weizmann.ac.il (AL)

This PDF file includes:

- SI Materials and Methods
- Figs. S1 to S18
- Tables S1 to S9

www.pnas.org/cgi/doi/10.1073/pnas.2003505117

SI Materials and Methods

Plant materials, library construction and sequencing. The tetraploid wheat (DD) was produced by intergeneric hybridization between *T. urartu* (AA) and *A. tauchii* (DD) followed by colchicine-mediated genome doubling (1). A single S2 euploid individual seed was used as the founder to construct the tetraploid wheat population up to S12 (2). A total of 11 AADD individuals, including six euploid plants from S6, S9 and S12 generations and 5 compensated aneuploid plants with large-scale homoeologous exchanges (HEs) from S9 generation, were sampled and used in this study (Fig. S1). The 11 individual plants were karyotyped based on fluorescence *in situ* hybridization (FISH) and genomic *in situ* hybridization (GISH). All AADD plants and corresponding diploid parents (AA and DD) were grown in a glasshouse under normal conditions (25/20°C, 16/8 h, day/night). Young leaves collected from the trefoil stage plants were immediately frozen in liquid nitrogen. Genomic DNA was extracted by the CTAB methods. Total RNA was isolated using Trizol (Invitrogen) based on standard protocol. Library construction was carried out with standard protocols (Illumina, San Diego, CA, USA) for both DNA and RNA samples. Cluster generation and sequencing were performed on Illumina 2500 system with pair-end 150-bp read length. Clean data have been deposited in the SRA database (http://www.ncbi.nlm.nih.gov/sra/) with accession number PRJNA608801.

Publicly available data of other allopolyploid species used in this study. All publicly available sequencing data used in this study were downloaded from National Center for Biotechnology Information (NCBI) Sequence Read Archive collection (SRA). For diploid wheat progenitor, *Aegilops tauschii*, DNA methylation sequencing data by targeted capture (chromosome 3DL, ERR3149846, ERR3149847 and ERR3149848) and ATAC-seq data (ERR2927671, ERR2927672 and ERR2927673) were obtained.

For Brassica, whole-genomic re-sequencing data from *B. oleracea* (SRR3202040), *B. rapa* (SRR3203680) and 7 *B. napus* accessions (SRR1266094, SRR1266095, SRR1266021, ERR475359, ERR475361, ERR457868, ERR457756, ERR457816 and ERR457887) and RNA-seq data from *B. oleracea* (SRR8633040) and *B. rapa* (SRR8633039) and 2 *B. napus* accessions “Darmor-bsh” and “Yudal” (PRJNA362706), were obtained.

For banana, whole-genomic re-sequencing data from *M. balbisiana* (SRR6996488), *B. rapa* (SRR6996492) and 3 *Musa* triploid accession including “Pelipita” (SRR6996486, ABB), “FenJiao” (SRR6996487, AAB) and “Kamaramasenge” (SRR6996494, AAB) and RNA-seq data from *M. balbisiana* (SRR7006127), *M. acuminata* (SRR7006128) and “FenJiao” (SRR7006136), were also obtained.

For Arabidopsis, whole-genomic re-sequencing data from *A. arenosa* (SRR2040811) 15 *A. suecica* accessions (SRR3123759, SRR3123760, SRR3123761, SRR3123762, SRR3123763, SRR3123764, SRR3123765, SRR3123766, SRR3123767, SRR3123768,

SRR3123769, SRR3123770, SRR2084154, SRR2084157, SRR2084158), were also obtained.

For rice, whole-genomic re-sequencing data and RNA-seq data from diploid parents (Nipponbare and 9311), diploid hybrids (N9 and 9N) and four segmental tetraploids, were download from SRA database with accession no. PRJNA514100.

For peanut (*Arachis hypogaea*), SNP information between A and B sub-genomes from 39 different varieties or accessions were downloaded from PeanutBase (https://peanutbase.org/data/public/Arachis_hypogaea/Tifrunner.esm.TVDM/).

Genome resequencing reads alignment and HE detection in synthetic tetraploid wheat (AADD). A total of 1.46 Tb (~17x per AADD sample) re-sequencing data was generated. Before reads alignment, an *in silico* tetraploid control dataset was constructed by mixing the diploid parental whole-genome sequencing data based on the ratio of the reference genomes size between A (*Triticum urartu*) and D (*Aegilops tauchii*). After removing low-quality reads (reads with Phred quality less than 10) using Trimmomatic (Bolger, et al. 2014), sequencing reads from both real and *in silico* tetraploid individuals were aligned to the combined AADD reference genome (merging the *T. urartu* and *Ae. tauchii* reference genomes (3, 4)) by Minimap2 (5) with default parameters. Considering that more than 85% genomic regions of AADD genome were TE-related regions, only uniquely mapped reads were retained for identification of HE events. PicardTools (v2.18.26) were used for sorting data, marking PCR duplication and adding RQ tags. Indel realignment was performed by GATK3 (https://software.broadinstitute.org/gatk/).

HE junctions were detected for each individual as follows. First, candidate HE junctions were detected by calculating the average depth of each 100 kb window at whole-genome level. Read depths were normalized by the whole-genome averaged depth. A slide-interval approach with 50 intervals (20 kb) with step-size of one interval was used to judge the depth difference between real and *in silico* individuals by Student's t-test. Intervals associated with significant difference of depth (p-value < 0.05) were flagged as "diff", the remaining intervals were flagged as "same". Intervals with identical flag were merged together and the joint of interval with different flag were treated as candidate HE junctions. Then, CNVkit (v0.9.6) (6) was performed to compare the copy number between *in silico* and real tetraploid individuals by scanning the candidate HE regions with 1kb-window size to shrink the HE junction size. Finally, each candidate HE junction was manually checked to confirm the exact joint. For HE junctions which were located in low reads-mappable region or when it was difficult to measure the exact location, HEs closest flanking regions that possessed different copy numbers were considered as the border of HE junctions.

For Brassica and banana, above method was used to identify original HE junctions.

Validation of HE junction based on SNPs-based methods. To validate HE junctions located in D genome of the AADD wheat tetraploid, first, re-sequencing reads from diploid *A.*

tauschii were aligned to D reference genome by Minimap2 with default parameter. PicardTools (v2.18.26) was used for sorting data, marking PCR duplication and adding RQ tags. Indel realignment was performed by GATK3 (https://software.broadinstitute.org/gatk/). SNP calling was performed by BCFtools. Only SNPs with mapping quality > 30, base quality >30, depth > 4 and the genotype equal to "1/1" were retained to generate the SNP list marked as "list 1", which indicated the inherent difference between our D genome and reference D genome. Second, sequencing reads from diploid *T. urartu* were aligned to D reference genome and performed SNP calling. This list was marked as "list 2", indicating the difference between our A genome and reference D genome. We removed SNPs occurring in "list 2" which also belonged to "list1", to generate a SNPs list, "list 3", which represented the difference between our DD and AA diploids. Then, sequencing reads from the *in silico* mix tetraploid were aligned to D reference genome and SAMtools mpileup was performed to assign the A-derived and D-derived reads based on list 3. In list 3, the SNP sites were kept if both A-derived and D-derived read depth > 4 and both A- and D-allele frequency belonged to the interval of [0.1, 0.9]. The filtered SNP list was marked as "list 4", representing the high-confident differences between our DD and AA diploids, including 5,870,921 SNP sites. Finally, sequencing reads from tetraploid individuals were aligned to D reference genome and samtools mpileup was performed to assign the A-derived and D-derived reads based on list 4. For each SNP site: (i) if all reads derived from A allele and the depth > 4, this site was flagged as "A-derived"; (ii) if all reads derived from D allele and the depth > 4, this site was flagged as "D-derived" and (iii) if the depth of both allele >4, this site was flagged as "heterozygous". Customized Python script was used to merge SNPs with same flag to generate different types of intervals (including AA-derive, DD-derived and heterozygous intervals, respectively). The junction between two adjacent intervals with different type was defined as HE junction. To avoid false positive derived from small interval and sequence elimination, HE junction that satisfied the following three conditions were retained: (1) The intervals on both sides of HE junction were large than 100kb; (2) The intervals on both sides of HE junction included at least 100 SNP sites and (3) the average sequencing depth of SNP sites in HE junction-related AA- and/or DD-derived intervals were larger than 1.5 fold of that in AA and/or DD subgenomes at whole genome level.

For Brassica, first, re-sequencing reads from diploid *B. oleracea* were aligned to *B. oleracea* (CC) reference genome to call SNP "list 1". Second, sequencing reads from diploid *B. rapa* (AA) were aligned to CC reference genome and performed call SNP "list 2". SNPs occurred in "list 2" that also belonged to "list1" were removed to generate SNPs "list 3". Then, sequencing reads from *B. napus* "Farmor-bzh" were aligned to CC reference genome and SAMtools mpileup was performed to assign the A-derived and C-derived reads based on SNP "list 3". Similar filtering criteria as for wheat were used to identify high-confident SNP "list 4" between AA and CC diploids (also between A and D subgenomes in *B. napus*). In total, we identified 2,063,999 SNP sites belonging to "list 4". Finally, sequencing reads from *B. napus* of different accessions (including “H165”, “Bristol”, “Aburamasari”, “Aviso”, “Yudal” and “Kale”) were aligned to CC reference genome and samtools mpileup was performed to assign the A-derived and C-derived reads based on list 4. Similar filtering criteria with wheat were used to flag SNP sites in “list 4” and customized Python script was performed to identify HE junctions.

For banana, first, re-sequencing reads from diploid *M. balbisiana* was aligned to *M. balbisiana* (BB) reference genome to call SNP “list 1”. Second, sequencing reads from diploid *M. acuminata* (AA) were aligned to BB reference genome and performed call SNP “list 2”. SNPs occurred in “list 2” that also belonged to “list1” were removed to generate SNPs “list 3”. In total, we identified 1,513,073 SNP sites belonging to “list 3”. Finally, sequencing reads from 3 triploid accessions were aligned to BB reference genome and samtools mpileup was performed to assign the A-derived and B-derived reads based on list 3. Similar filtering criteria with wheat were used to flag SNP sites in “list 3” and customized Python script was performed to identify HE junctions.

For Arabidopsis, first, re-sequencing reads from diploid *A. arenose* was aligned to *A. thaliana* reference genome to call SNP “list 1” (homozygous SNP). Then re-sequencing reads from tetraploid *A. suecica* including SRR2084154, SRR2084158 and SRR3123766 was aligned to *A. thaliana* reference genome. SNPs with mapping quality > 30, base quality >30, depth > 8 and the genotype equal to “0/1” were retained to generate the SNP list marked as “list 2”, “list 3” and “list 4” (heterozygous SNP). SNP shared among all the four lists were generated to the final SNP site in “list 5”. In total, we identified 271,627 SNP sites belonging to “list 5”. Finally, sequencing reads from the remaining 12 *A. suecica* accessions were aligned to *A. thaliana* reference genome and identified HE junctions based on SNP “list 5”. For Arabidopsis, the intervals on both sides of HE junction were large than 50kb.

For rice, the method descripted previously (7) were used to call SNP list between cv. Nipponbare (*japonica* rice) and cv. 9311 (*indica* rice). In total, 1,457,488 SNP sites between Nipponbare and 9311 were used for HE junction calling in synthetic rice tetraploid individuals.

For peanut (*Arachis hypogaea*), 825,960 SNP sites between A and B sub-genomes from 39 different varieties or accessions (8) were used to identify HE junctions.

For each species, the Bam files of each HE junction and 100 kb flanking region were deposited in the SRA database (http://www.ncbi.nlm.nih.gov/sra/) with accession number PRJNA625880.

Judgement of chromosome rearrangement flanking HE junctions based on comparative genomic analysis in the synthetic tetraploid wheat (AADD). Chromosome rearrangement flanking HE junctions occurred in A sub-genome. To estimate if the rearrangements were attributed to HE or mis-assembly of A reference genome, chromosome sequences related to rearrangement flanking HE junctions in A genome were extracted and aligned to reference genomes of A sub-genome of wild emmer wheat (*T. turgidum* ssp. *dicoccoides*) and durum wheat (*T. turgidum* ssp. *durum*) and to D reference genome (*A.*

tauchii) by Minimap2 with “asm20” parameter. Only alignment hits with length large than 1000bp and alignment score large than 10 were retained. Dot plots between A genome and A sub-genome of emmer wheat, A sub-genome of durum wheat and D genome were performed and retrieved for structural variations, respectively.

Motif analysis. The 37 HE junctions from D subgenome were used for motif enrichment analysis by MEME Suite with classic mode (9). The candidate motif wide was set from 12 to 22 base pair. Motifs with E-value < 10e-10 were considered as candidate motifs. Genomic locations of motif were determined by FIMO (10) with q-value < 0.05. 3 kb flanking regions of each HE site was binned to 200bp non-overlapping windows to calculate the occurrences of each motif. Similar, as control, 37 random genomic regions with same size distribution were resampled with 1000 replicates using customized Python script. Bedtools coverage (11) was used to count motifs which overlapping with different genomic regions including promoter (upstream 2kb from gene region), gene and TE. For Brassica (104 HE junctions), rice (166 HE junctions) and peanut (26 junctions), similar methods were used to search HE-enriched motif and determine their genomic locations.

Identification of homoeologous gene pairs. Homoeolog pairs were identified by jcvi utility libraries (https://github.com/tanghaibao/jcvi) based on sequence similarity and genomic position of genes from homoeologous chromosomes. For wheat AADD segmental tetraploid, 13468 homoeolog pairs were identified between A and D reference genomes. For brassica, 25022 homoeolog pairs were identified between A and C reference genomes. For banana, 24497 homoeolog pairs were identified between A and B reference genomes. For rice, 24362 homoeolog pairs were identified between Nipponbare and 9311 reference genomes.

Validation of gene fusion on transcript level. Although not useful for validating HE sites in non-transcript genomic regions, RNA-seq data could be used for (i) testing gene fusion event on transcript level and (ii) narrowing down the resolution of HE events in transcript genomic regions (because of higher sequencing depth). Therefore, we also validated the HE interval using RNA-seq data based on the SNP list re-sequencing data. To test whether fusion transcripts occurred at the transcriptional level, Integrative genomics viewer (IGV) was performed to check if A-specific and D-specific SNP sites co-occurred in a single RNA-seq read or a single pair of RNA-seq reads (fusion-supporting reads). Furthermore, RT-PCR and Sanger sequencing were performed to obtain full-length fusion transcripts. For each fusion gene, 20 positive clones from diploid parents, control tetraploid and HE tetraploid were chosen for Sanger sequencing. MEGA7 was used for the alignment the results of Sanger sequencing. ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/) was used to predict the open reading frame of fusion transcripts.

For Brassica, banana and rice, above RNA-seq-based method was also used to validate gene fusion event at transcriptional level.

DNA Methylation data and ATAC-seq data analysis. After removing adaptor and low-quality reads, bisulphite-treated methylome reads were aligned to *A. tauschii* genome (for wheat data) or *B. oleracea* genome (for Brassica data) using Bismark software (12). Cytosine sites covered by at least 4 reads were retained further analysis. Methylation level of HE junctions and flanking regions (2Kb around) were calculated by CGmap Tools (13).

After removing adaptor and low-quality reads, ATAC-seq reads were aligned to *B. oleracea* genome using Bowtie2 (14) with 1-mismatch. Only uniquely mapped reads were retained for further analysis. Sequencing depth of HE junctions and flanking regions (2Kb around) were calculated by Bedtools (11).

1. Shaked H, Kashkush K, Ozkan H, Feldman M, & Levy AA (2001) Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *The Plant Cell* 13(8):1749-1759.
2. Gou X*, et al.* (2018) Transgenerationally precipitated meiotic chromosome instability fuels rapid karyotypic evolution and phenotypic diversity in an artificially constructed allotetraploid wheat (AADD). *Molecular biology and evolution* 35(5):1078-1091.
3. Ling HQ, Ma B, Shi X, Liu H, & Liang C (2018) Genome sequence of the progenitor of wheat A subgenome Triticum urartu. *Nature* 557(7705).
4. Luo M-C*, et al.* (2017) Genome sequence of the progenitor of the wheat D genome Aegilops tauschii. *Nature* 551(7681):498.
5. Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34(18):3094-3100.
6. Talevich E, Shain AH, Botton T, & Bastian BC (2016) CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS computational biology* 12(4):e1004873.
7. Li N*, et al.* (2019) DNA methylation repatterning accompanying hybridization, whole genome doubling and homoeolog exchange in nascent segmental rice allotetraploids. *New Phytologist*.
8. Bertioli DJ*, et al.* (2019) The genome sequence of segmental allotetraploid peanut Arachis hypogaea. *Nature genetics* 51(5):877.
9. Bailey TL*, et al.* (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic acids research* 37(suppl_2):W202-W208.
10. Grant CE, Bailey TL, & Noble WS (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27(7):1017-1018.
11. Quinlan AR & Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841-842.
12. Krueger F & Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *bioinformatics* 27(11):1571-1572.
13. Guo W*, et al.* (2017) CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics* 34(3):381-387.
14. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature*

methods 9(4):357.

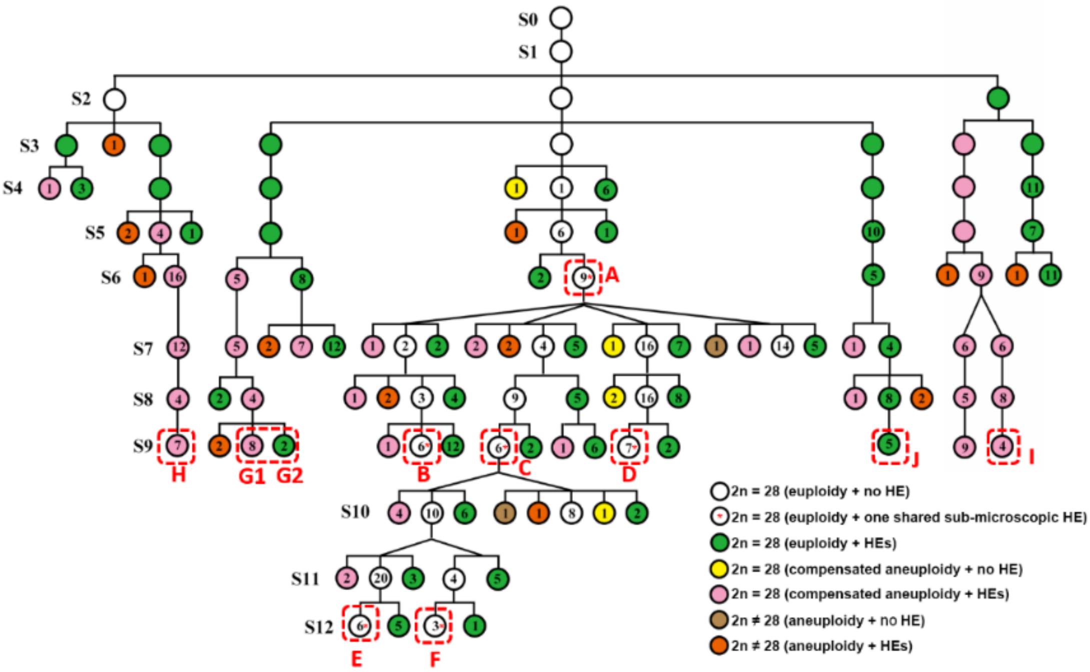


Fig. S1. Karyotypic composition of progenies of bona fide euploidy-descendants (being derived from one or more euploid parental individuals as indicated by the number of branches) in the synthetic tetraploid wheat (AADD) population from S0 to S12 generations. The seven different circles represent seven distinct karyotypic groups with numbers in the circles referring to individual numbers within a given karyotypic group. Circles within red dotted boxes denote the samples used in this study. Plants A-F (white circles with a red star) represent euploids at the cytological level, i.e., they did not contain numerical or structural variations including homoeologous exchange (HE). Sequencing of their genomes, however, showed that they all harbor a single sub-microscopic translocation between chromosomes 4A and 5D, and which most likely occurred in the S1 generation and then being vertically transmitted to, and shared by, all descendent plants. Because this translocation event occurred in regions including an ancient translocation between the long arms of chromosomes 4A and 5A in *T. urartu* (3), it was also a HE by definition. These plants (A-F), being sharing the same single sub-microscopic translocation, were thus represent ideal negative controls for the analysis of additional HEs in other plants.

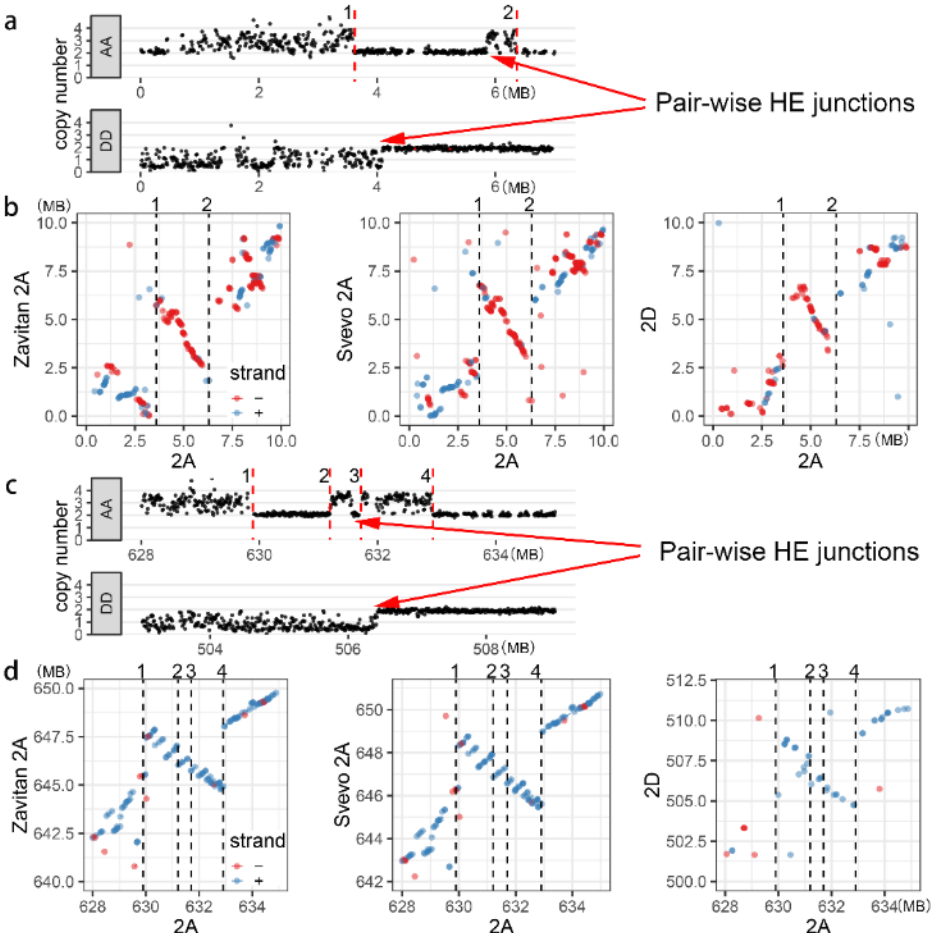


Fig. S2. Identification of HE-junction-related chromosome structural variation based on chromosome alignment among related wheat species. (a) and (c), two examples of candidate HE junction in A and D subgenomes. Regions between red dashed lines indicate the mis-ordered genomic regions (large inversion) in A subgenome. Red arrows represent confident HE junctions in both chromosomes. (b) and (d), chromosome alignment among wheat species of corresponding regions in (a) and (c). Alignments between regions of A subgenome and syntenic regions in A subgenome of cv. Zavitan (wild emmer wheat, left), in A subgenome of cv. Svevo (durum wheat, middle) and D genome from *Ae. tauschii* (right) were shown. Red and blue dots indicate alignment hits in forward and reverse orientations, respectively. The black dashed lines and corresponding digits above them are identical with which occurred in (a) and (c). In each case, if the three comparisons in shared similar sequence-rearrangement patterns, it will be inferred that the reference sequence of *T. urartu* contains specific ‘structure rearrangement’, which is attributed to mis-assembly other than HE-induced structural variation.

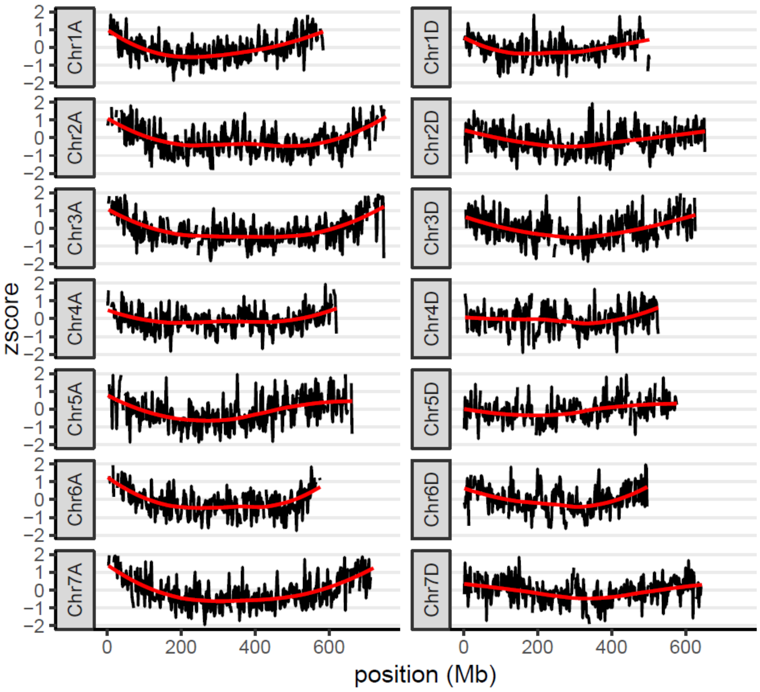


Fig. S3. Distribution of CCN motif in AADD genome. Genome was divided to 2Mb bins to count the number of CCN motif. X-axis indicates the chromosome position and y-axis indicates the Z-score transformed density of motif in each 2Mb bins. Red curves for each chromosome represent the fitting curves based on loess regression.

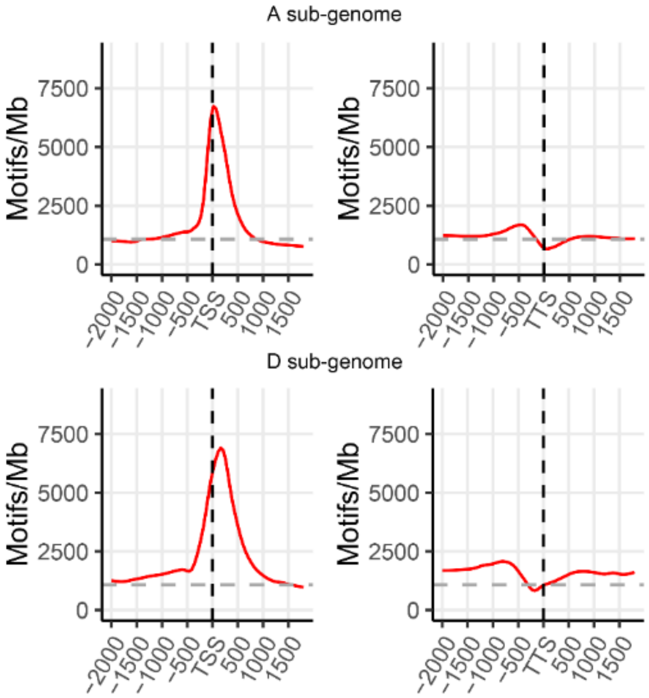


Fig. S4. Distribution of CCN motif in gene-related regions in A (a) and D (b) subgenomes. For each sub-genome, left panel represents the motif density flanking transcript start site (TSS) (2kb flanking regions), whereas right panel represents the motif density flanking transcript terminal site (TTS) (2kb flanking regions).

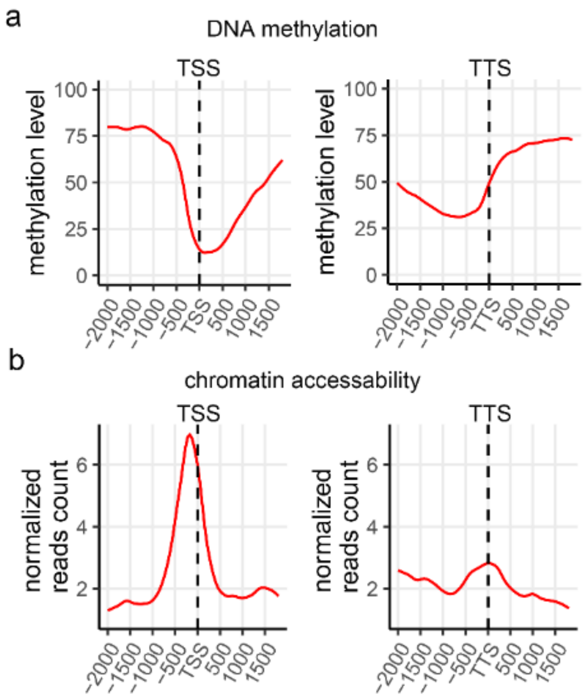
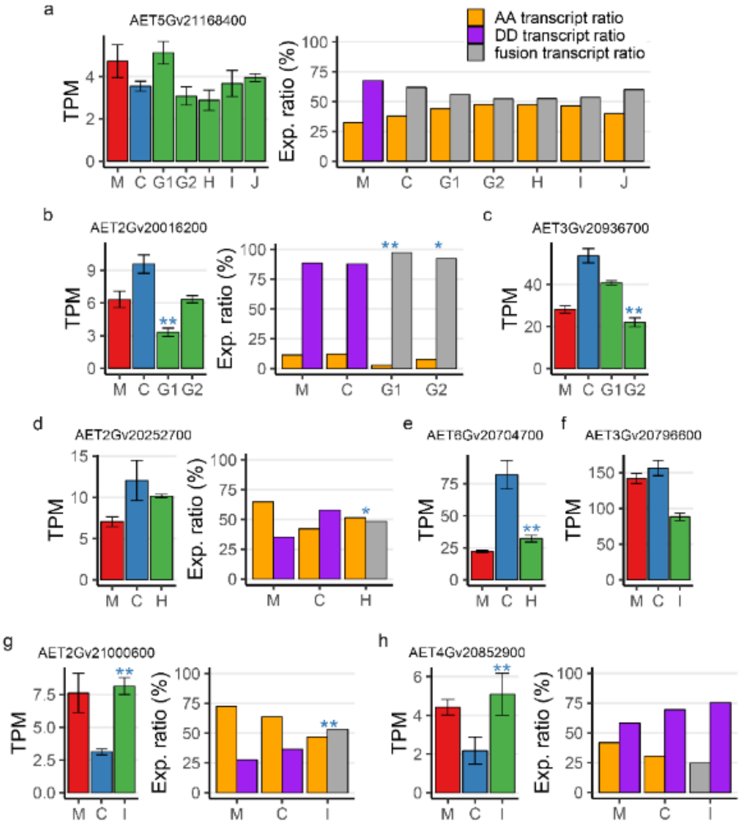


Fig. S5. Distribution of DNA methylation and chromatin accessibility in gene-related regions in *A. tauschii*. For both (a) DNA methylation and (b) chromatin accessibility, left panel represents the feature density flanking transcript start site (TSS) (2kb flanking regions), whereas right panel represents the feature density flanking transcript terminal site (TTS) (2kb flanking regions).



* Significance of expression difference or expression ratio difference between HE tetraploid and control tetraploid

Fig. S6. Expression patterns of eight fusion genes in HE-related tetraploids. RNA-seq data of *in silico* mix (M), control tetraploid (C) and HE tetraploids (G1, G2, H, I and J) were aligned to *Ae. tauschii* reference genome to calculated total expression level. Meanwhile, SNP list between A and D subgenome were used to divide A-specific and D-specific reads to calculate relative expression ratio of A-derived transcript, D-derived transcript and fusion transcript in different samples. For each gene, left panel represents the total expression level and right panel represents the relative expression ratio of different kinds of transcript in M, C and HE tetraploids. Blue asterisks indicate different expression (left panel, based on the results of DEseq2) or different relative expression ratio (right panel, based on Chi-squared test) compared with control sample (*: p-value < 0.05; **: p-value < 0.01). For (c), three types of transcripts (namely A-derived transcript, D-derived transcript and fusion transcript) generated simultaneously, calculating the relative expression level is not possible. For (e) and (f), only fusion transcript generated, namely the relative expression ratio of fusion transcript is equal to one.

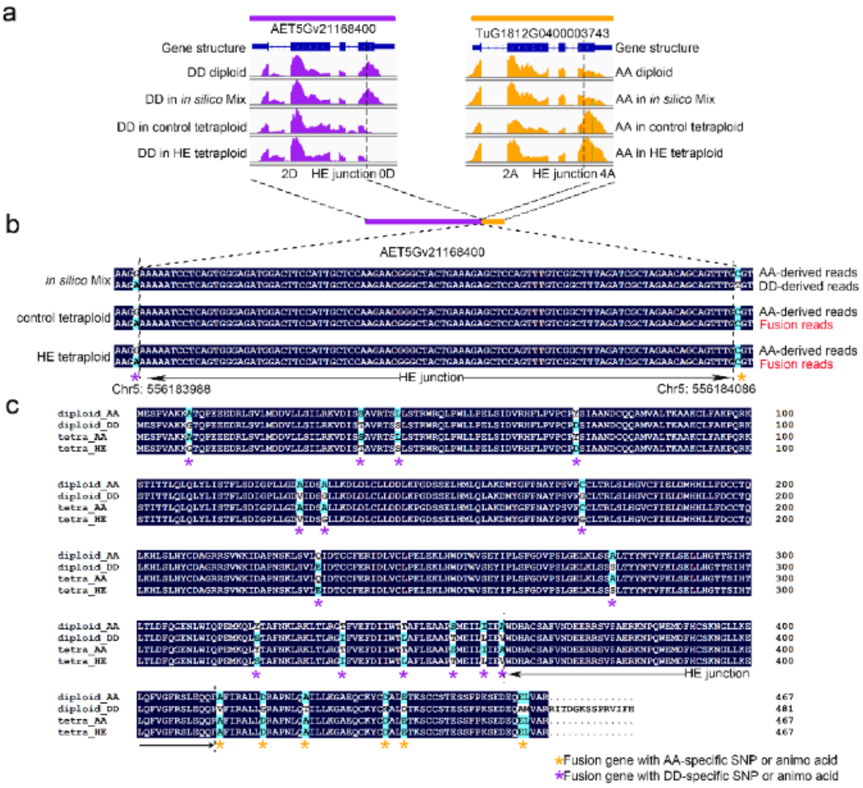


Fig. S7. Gene fusion event between HE-related homoeologous gene pairs. (a) IGV illustration of a candidate homoeolog fusion event based on RNA-seq reads depth. Each track (from up to down) represents the transcript structure and gene expression profile in diploid, in silico tetraploid, control tetraploid and HE-related tetraploid, respectively. Orange and purple indicate the profile of A and D subgenome genes, respectively. The position of HE junction were marked as dashed lines. The combination of number and letter flanking the HE junction represent the copy number of a given sub-genome in HE-related tetraploid (2A indicating 2 copies of A subgenome). (b) Validation of fusion transcript based on diploid-specific SNPs method. The tracks represent the consensus sequencing reads in *in silico* mix tetraploid, control tetraploid and HE-related tetraploid. Nucleotides without black background indicate high-confident SNPs between AA and DD diploid and used for genotyping the sequence reads. The asterisks in different colors under SNP positions indicate the origin of SNP (orange for A subgenome and purple for D subgenome) in fusion read. Region between adjacent asterisks of different colors represents HE junction, which position was also shown. (c) Predicted amino acid sequence of fusion transcript by ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/) based on Sanger sequencing. Similar with SNPs in (b), amino acids without black background indicate high-confident variants between AA and DD diploid. The asterisks in different colors under variants positions indicate the origin of amino acid (orange for A subgenome and purple for D subgenome) in predicted fusion protein. Region between adjacent asterisks of different colors represents HE junction.

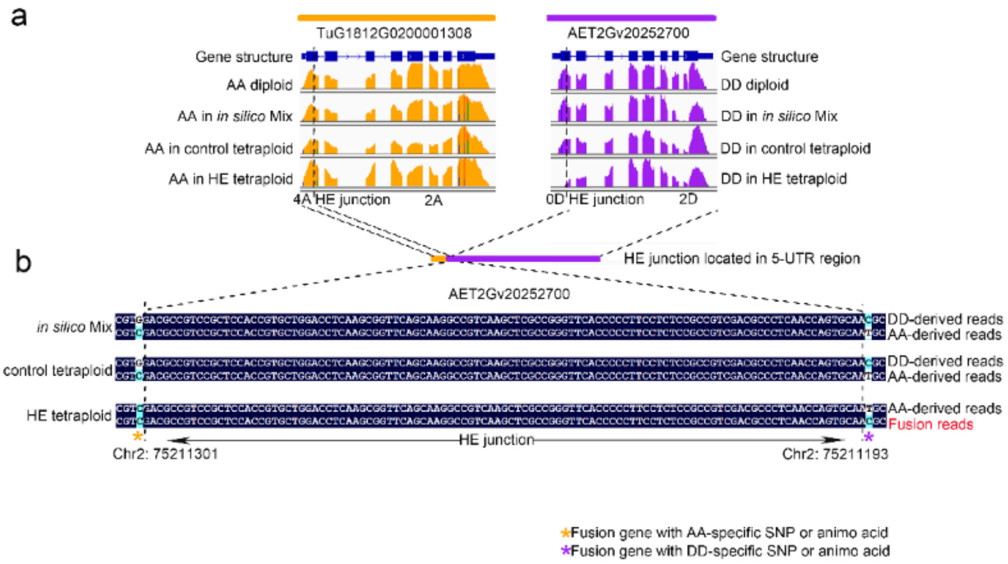


Fig. S7. Gene fusion event 3.

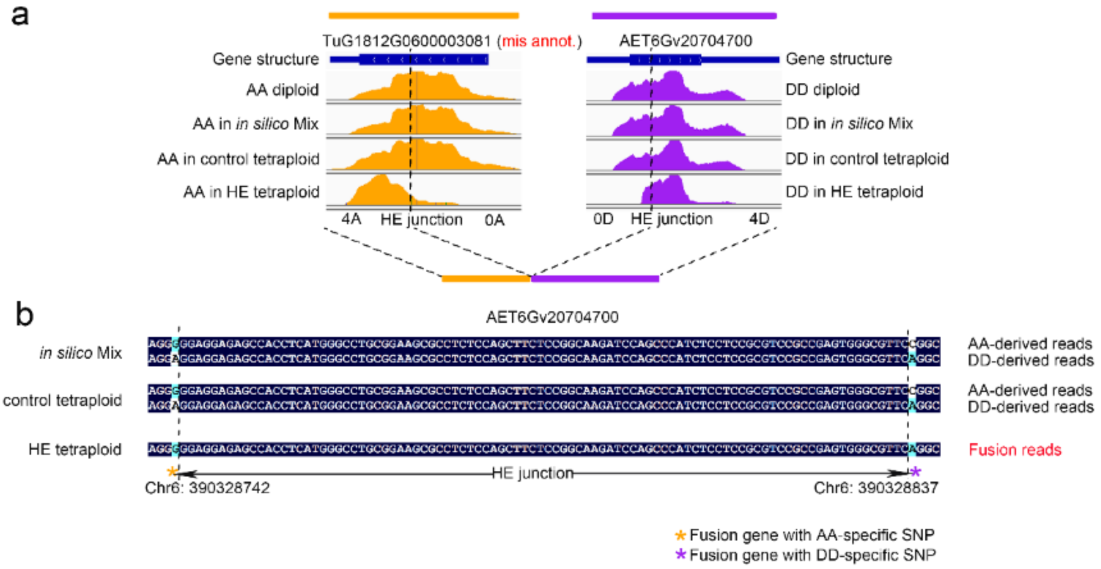


Fig. S7. Gene fusion event 4.

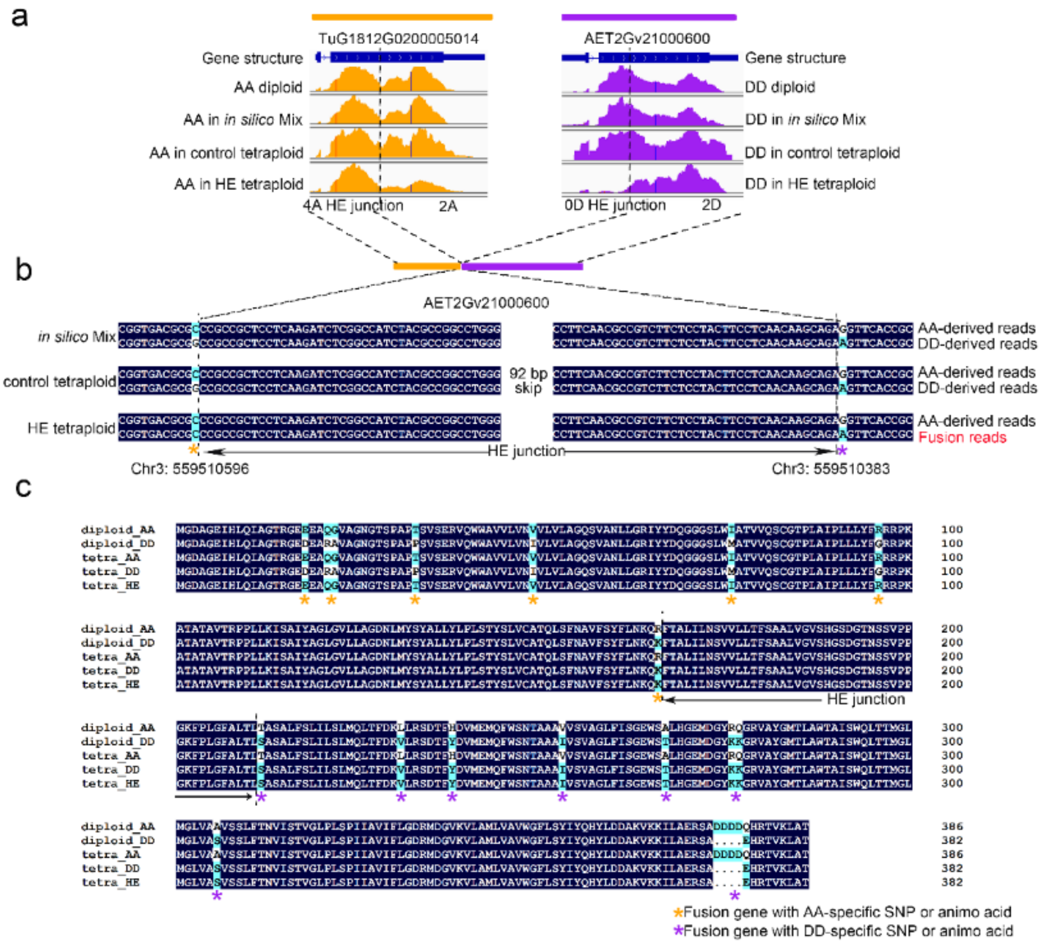


Fig. S7. Gene fusion event 5.

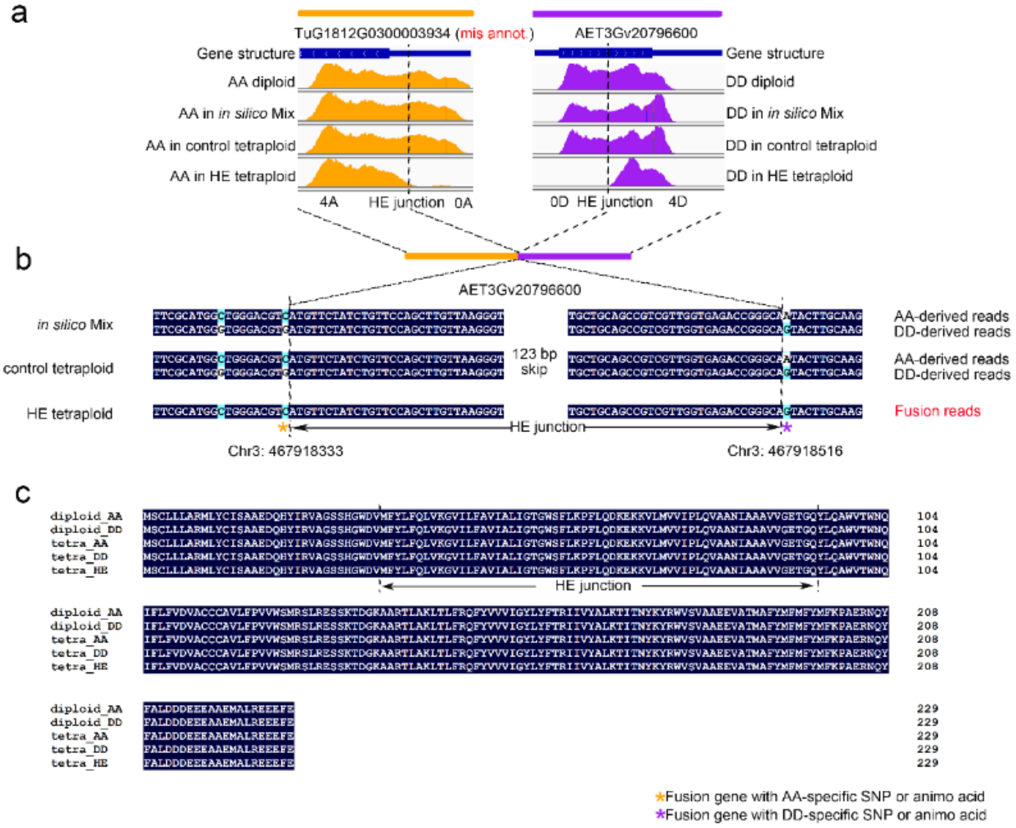


Fig. S7. Gene fusion event 6.

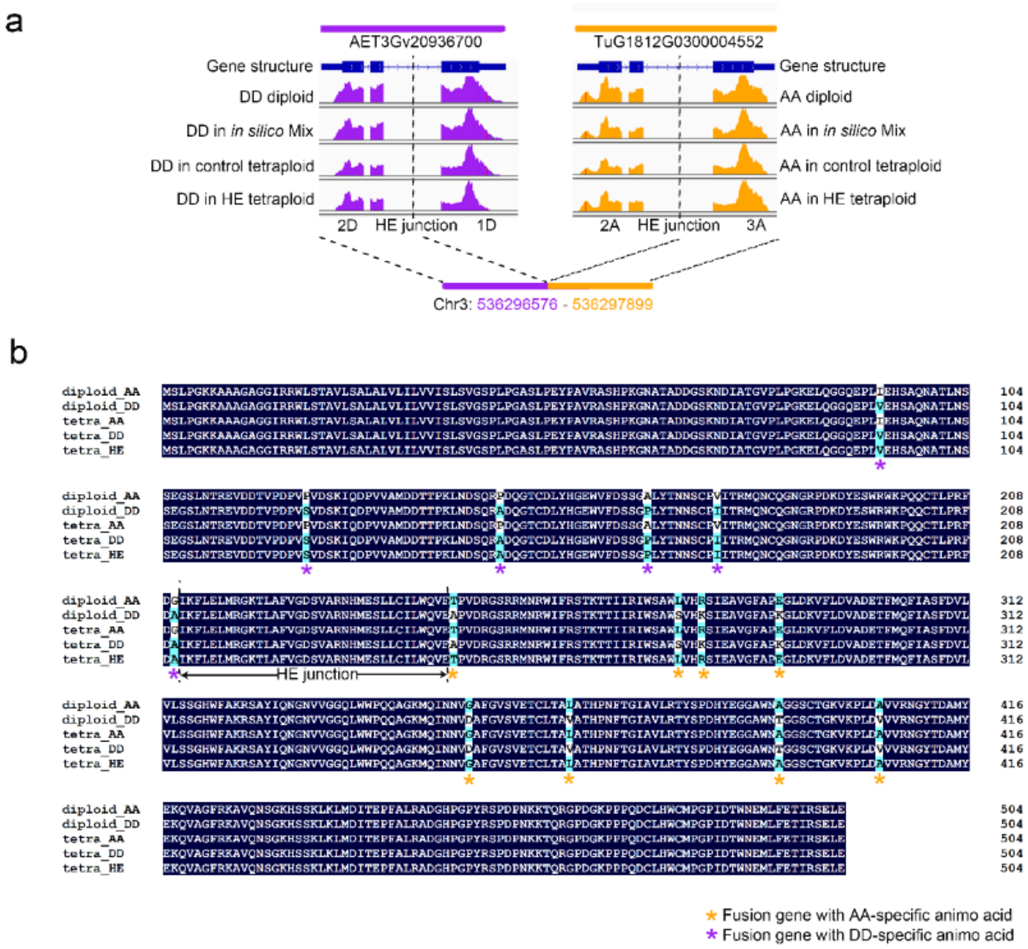


Fig. S7. Gene fusion event 7.

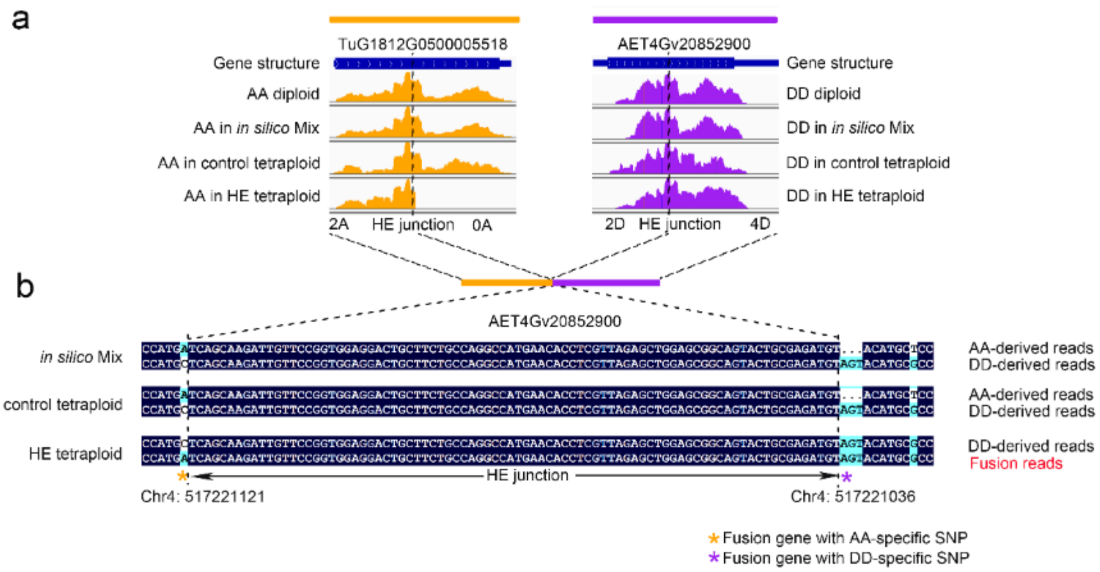


Fig. S7. Gene fusion event 8.

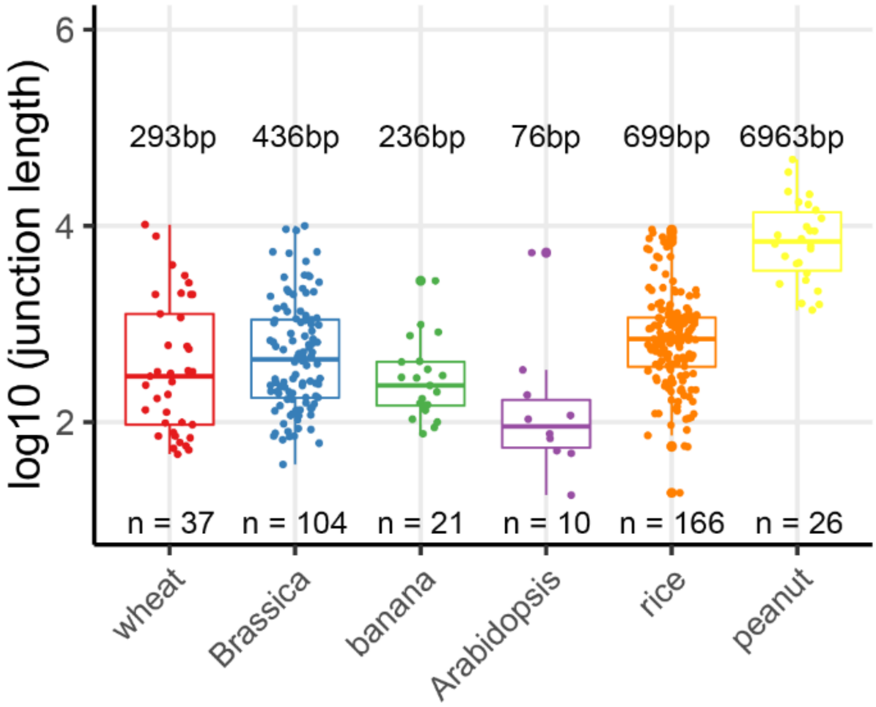


Fig. S8. Resolution of HE junctions based on diploid-specific SNP method in wheat, brassica and peanut. The number of HE junctions are shown above boxes. The median length of HE junctions are shown above of boxes.

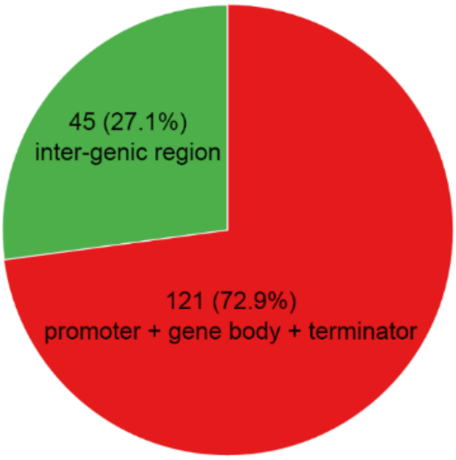


Fig. S9. Proportion of HE junctions including gene related regions (including promoter and terminator regions) and intergenic regions in the Nipponbare genome.

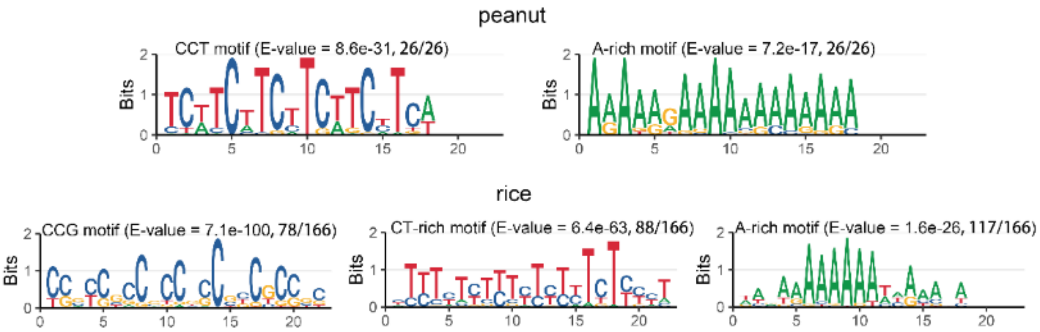


Fig. S10. HE-enriched motif logos in peanut and rice. Similar HE-enriched motif logos in Brassica was shown in Fig. 6b.

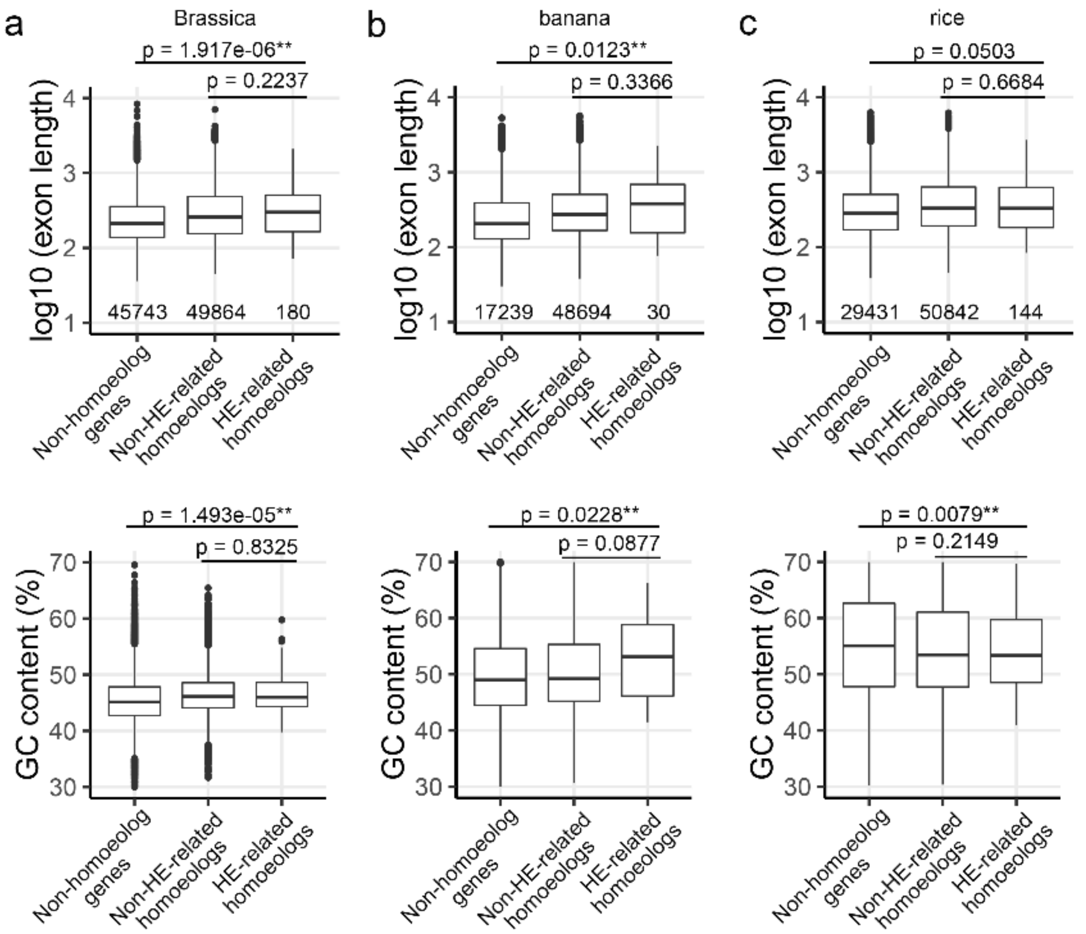


Fig. S11. Comparison of genomic features of HE-related homoeologs with non-HE-related homoeologs and non-homoeologous genes in (a) Brassica, (b) banana and (c) rice. For each species, up panel and bottom panel represent exon length and GC content, respectively, among three types of gene sets. The asterisks indicate significant differences between HE-related homoeologs and non-HE-related homoeologs or non-homoeologous genes (Mann-Whitney-U test; *: p-value < 0.05; **: p-value < 0.01).

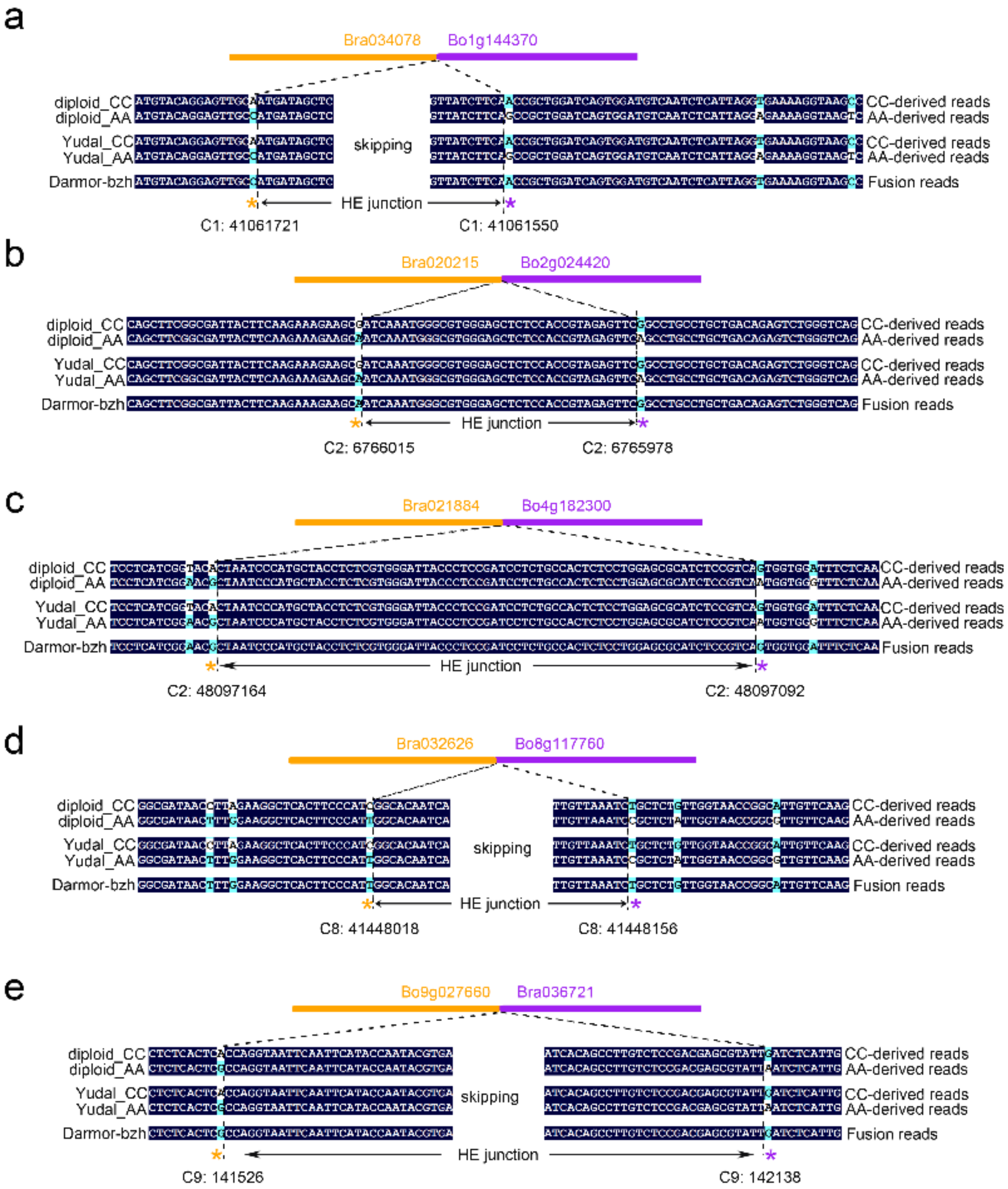


Fig. S12. Validation of gene fusion events based on RNA-seq data in Brassica. For (a) and (b), tracks represent the consensus sequencing reads in *B.oleracea* (CC) and *Brassica rapa* (AA), C and A subgenome in *B.napus* “Yudal” (control tetraploid) and HE-related tetraploid “Darmor-bzh”. Nucleotides without black background indicate high-confident SNPs between AA and CC diploid and used for genotyping the sequence reads. The asterisks in different colors under SNP positions indicate the origin of SNP (orange for A subgenome and purple for C subgenome) in fusion read. Region between adjacent asterisks with different colors represents HE junction, which position was also shown. Other similar cases of HE-related gene fusion events in Brassica were shown in Fig. 6e.

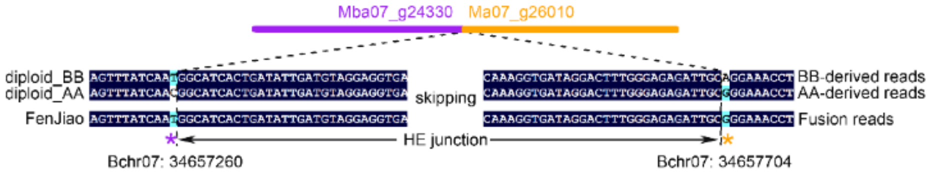


Fig. S13. Validation of gene fusion events based on RNA-seq data in banana. For (a) and (b), tracks represent the consensus sequencing reads in *M.balbisiana* (BB) and *M.acuminata* (AA), B and A subgenome in HE-related triploid “Fenjiao” (control tetraploid). Nucleotides without black background indicate high-confident SNPs between AA and BB diploid and used for genotyping the sequence reads. The asterisks in different colors under SNP positions indicate the origin of SNP (orange for A subgenome and purple for B subgenome) in fusion read. Region between adjacent asterisks with different colors represents HE junction, which position was also shown. Other similar cases of HE-related gene fusion events in banana were shown in Fig. 6e.

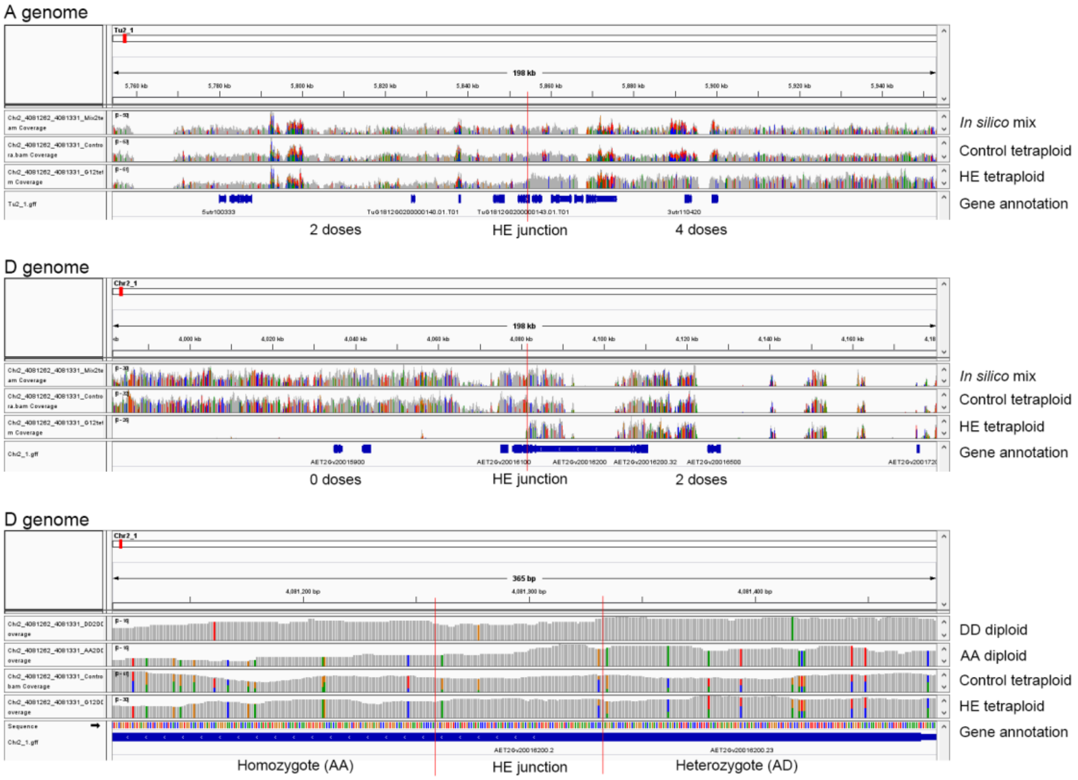
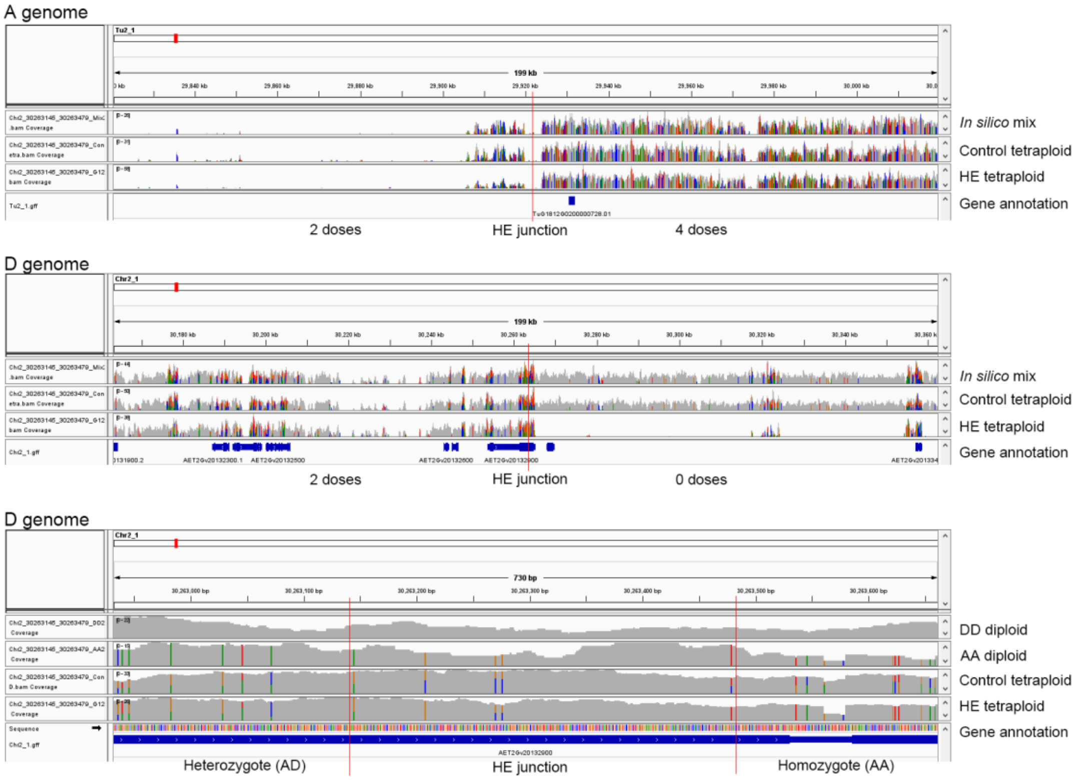
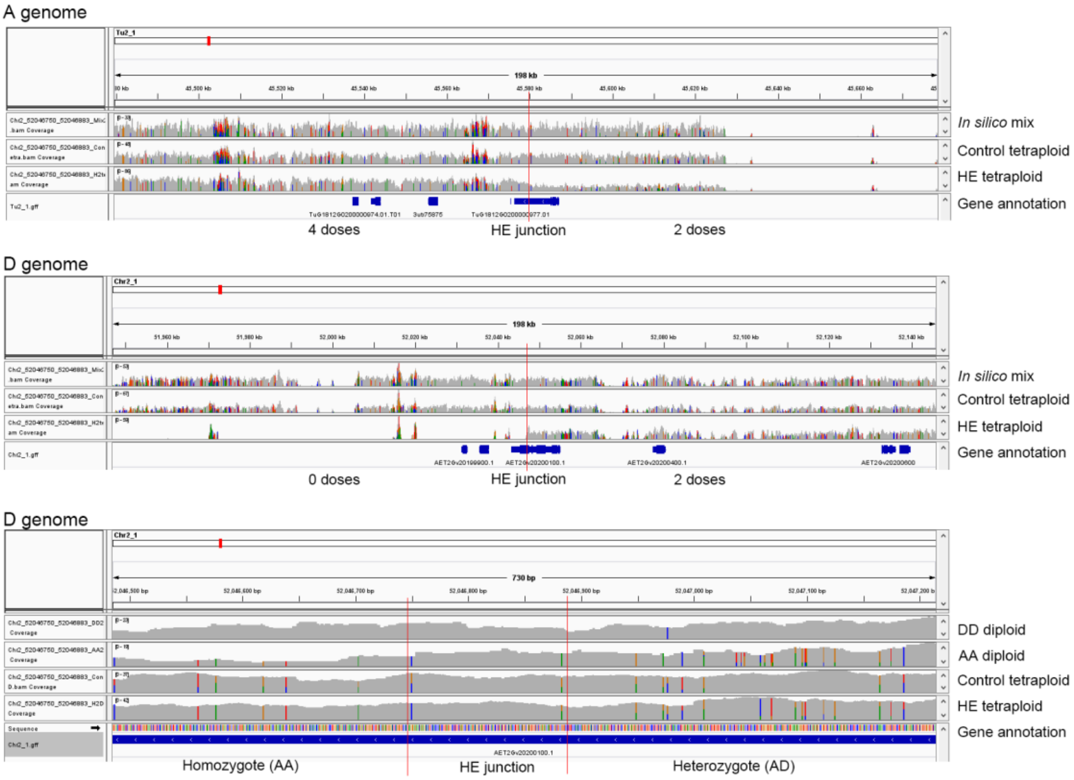


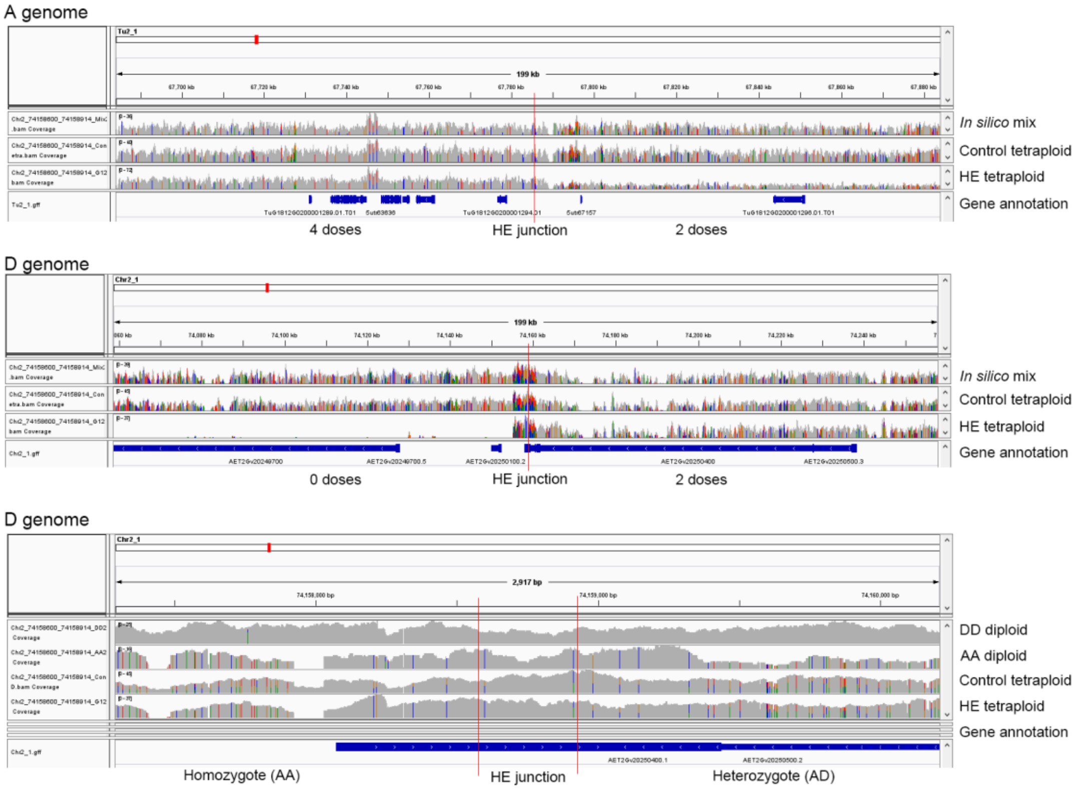
Fig. S14. IGV panels of HE junction regions in AADD tetraploid wheat (37 junctions). The top and middle panels show the sequencing depth around HE junctions in both A and D subgenomes, respectively. Red lines indicate the position of HE junctions and genome dosages are shown on both sides of HE junction. The bottom panel shows the genotypes (homozygote or heterozygote) around HE junction (region between red lines) based on SNP genotyping.



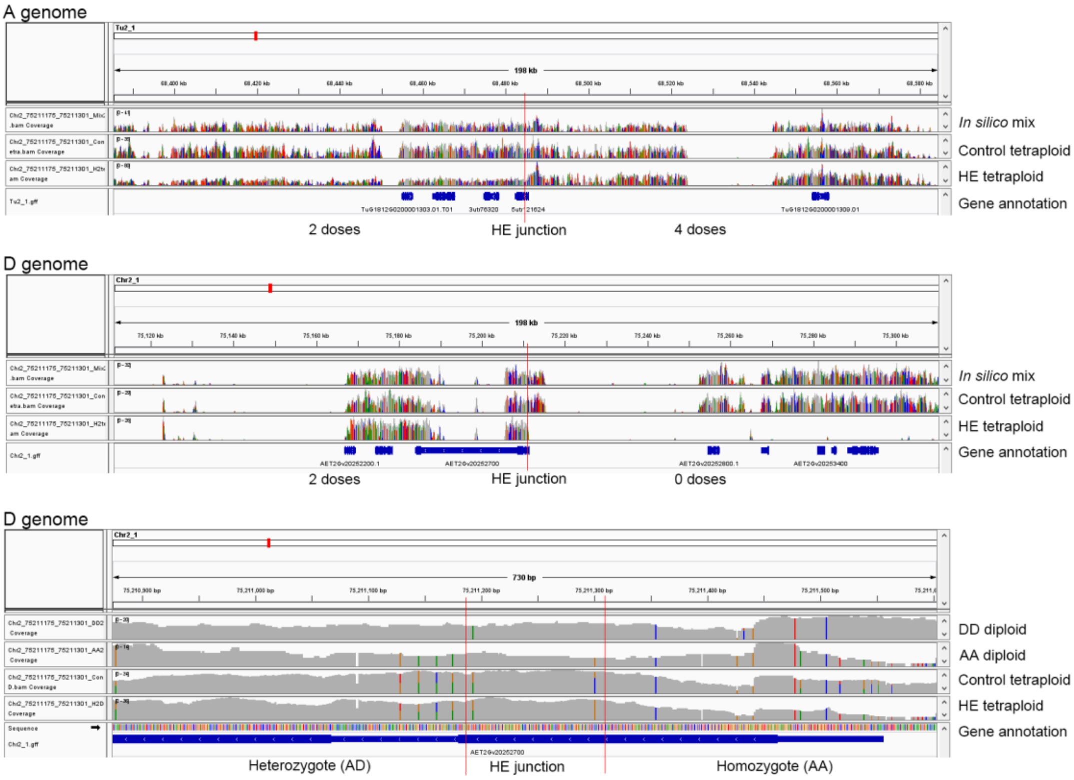
HE junction 2.



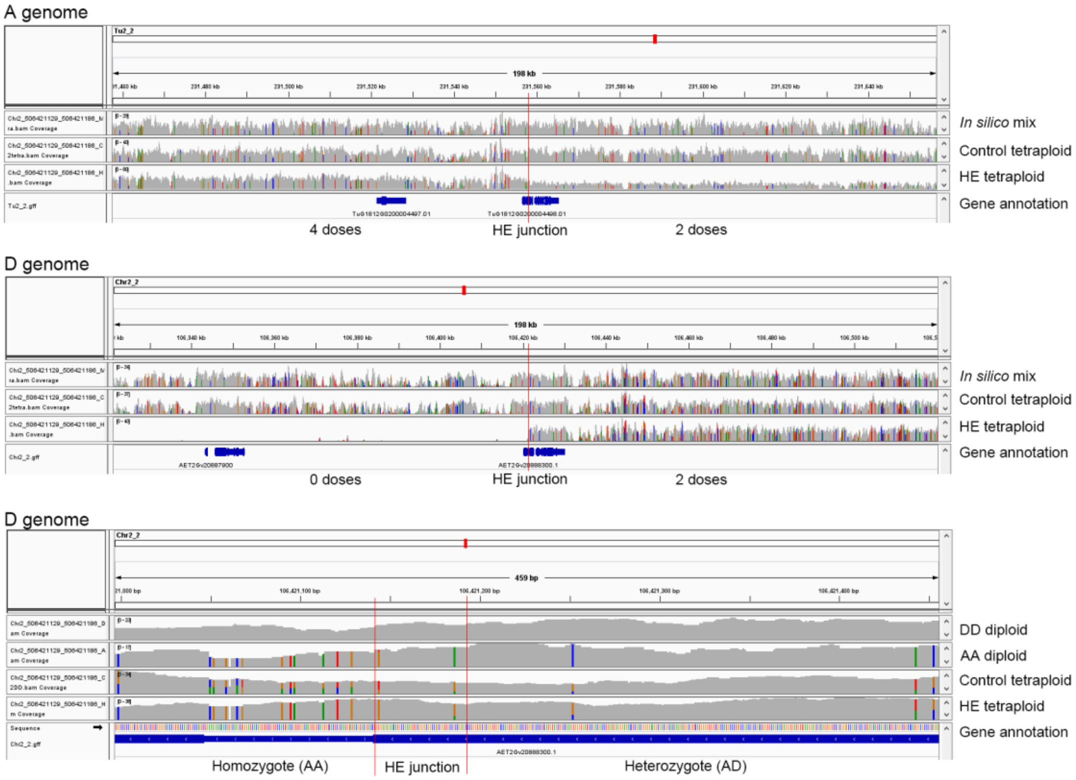
HE junction 3.



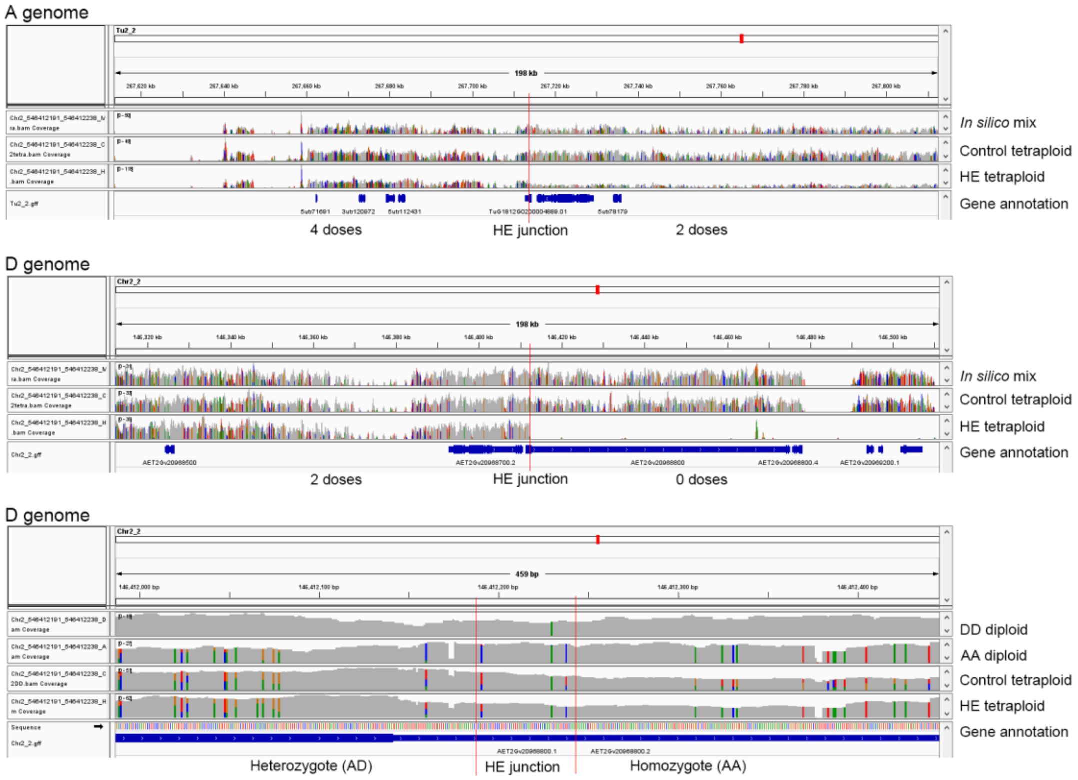
HE junction 4.



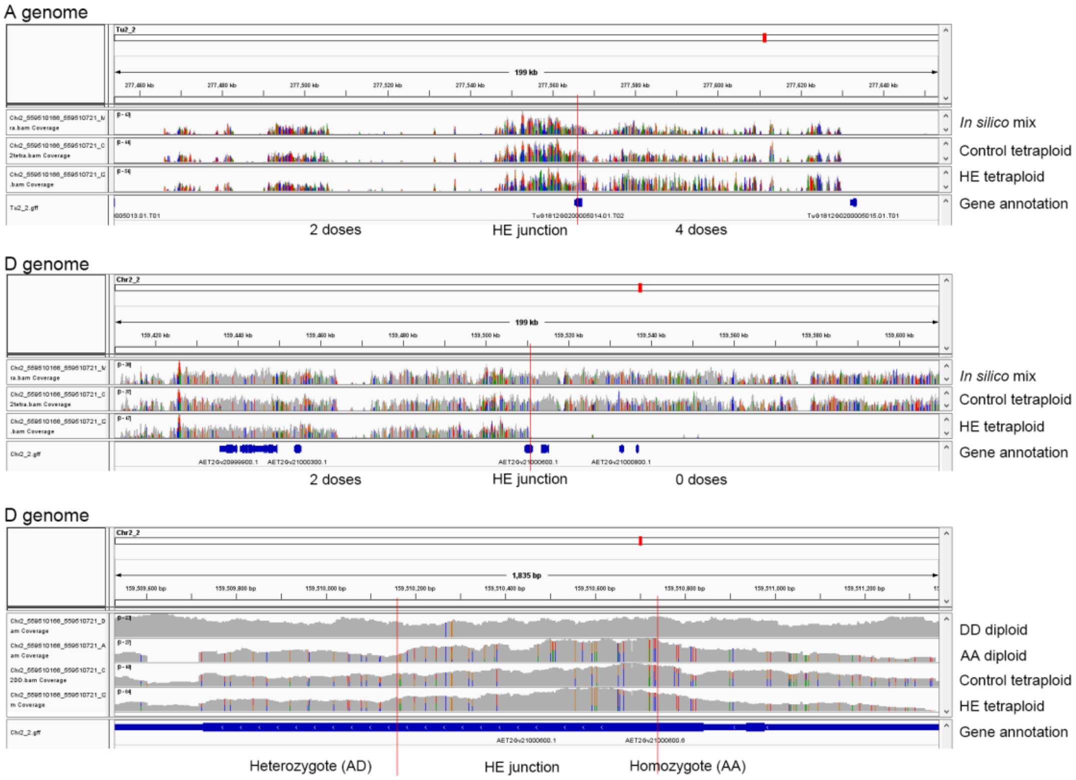
HE junction 5.



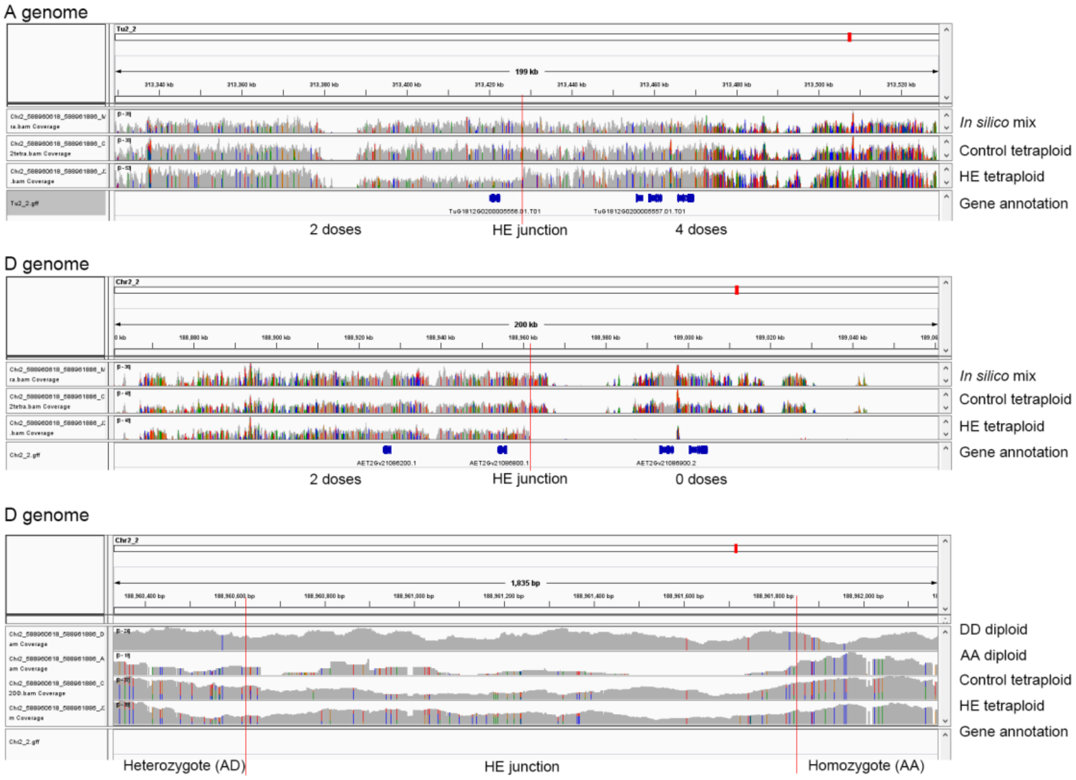
HE junction 6.



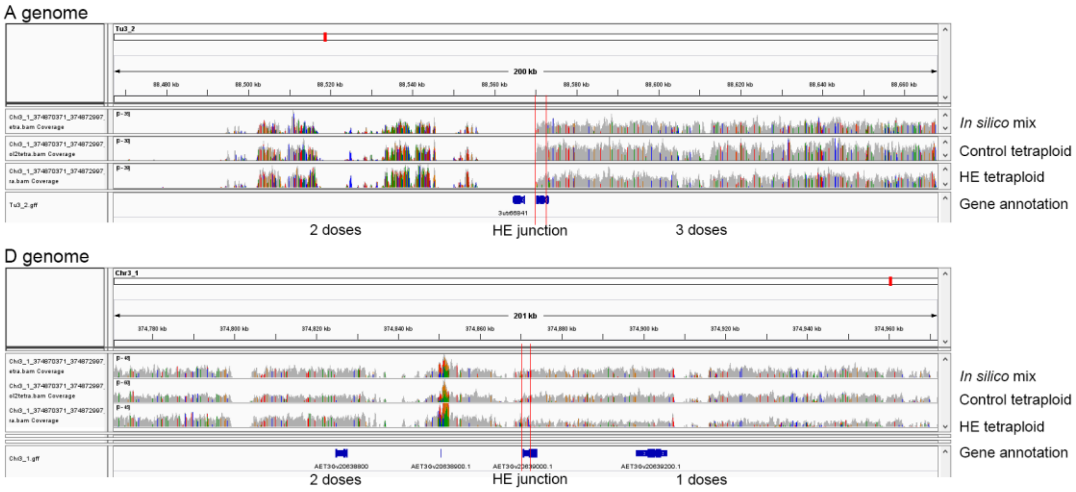
HE junction 7.



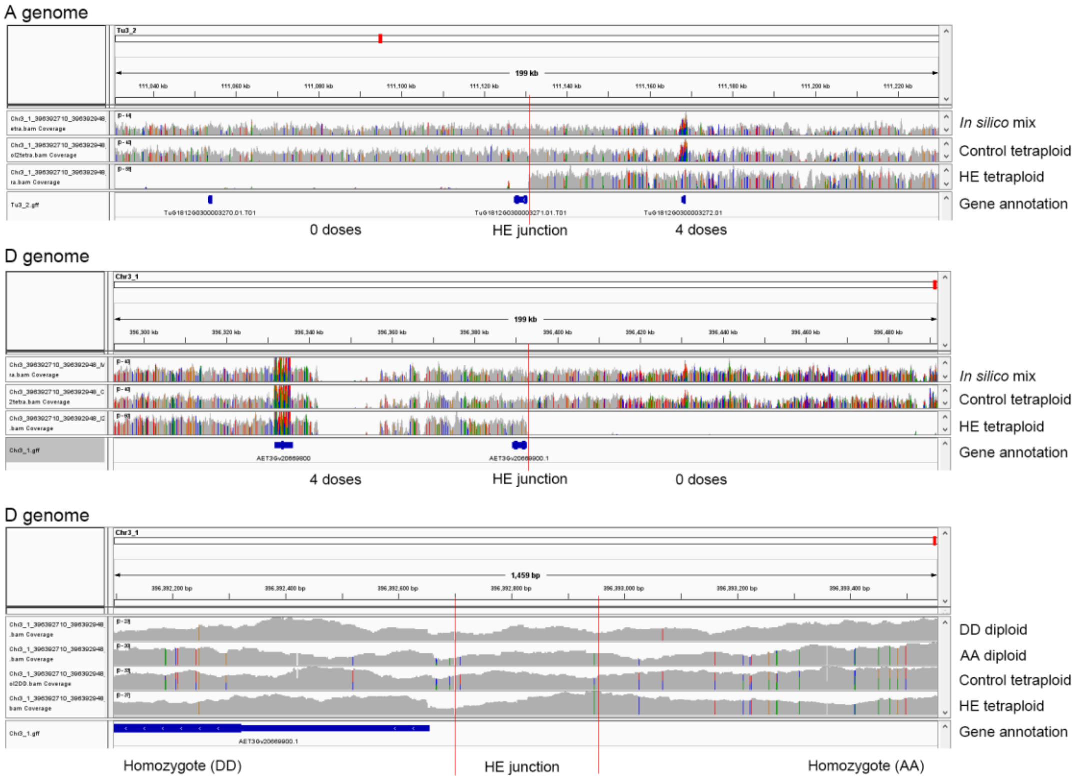
HE junction 8.



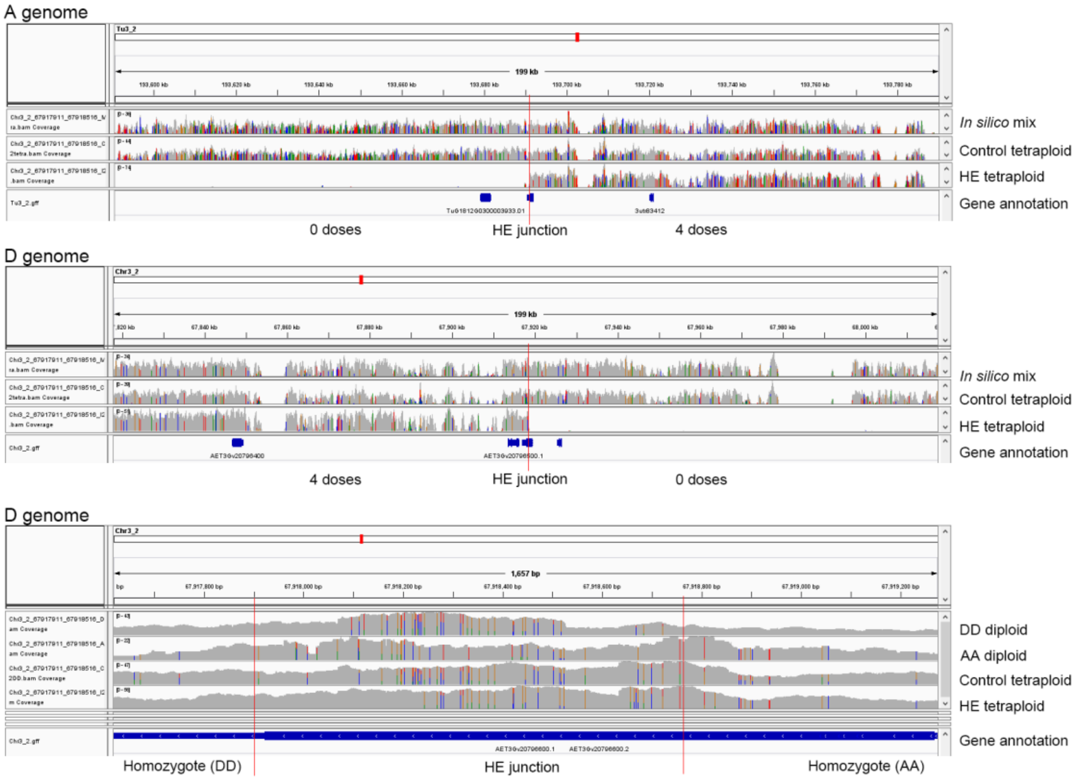
HE junction 9.



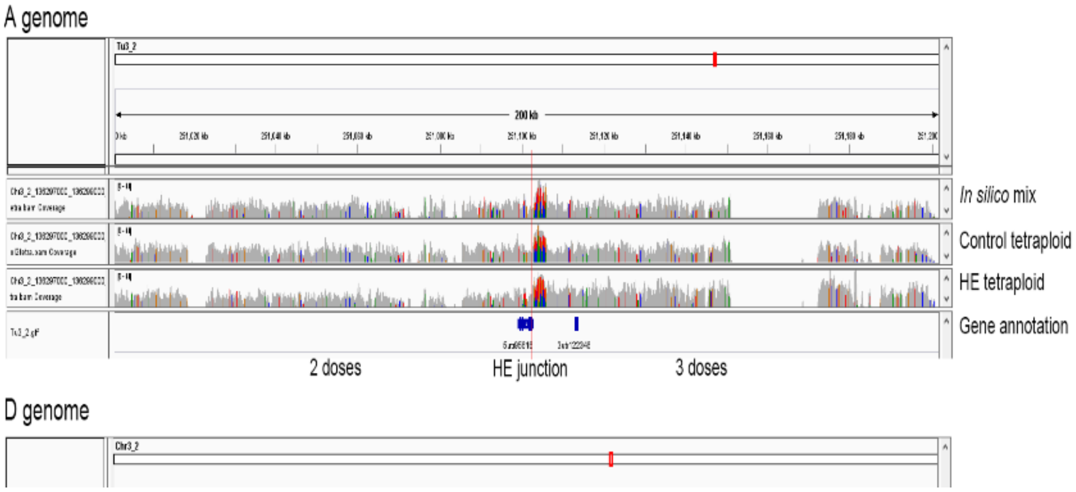
HE junction 10.



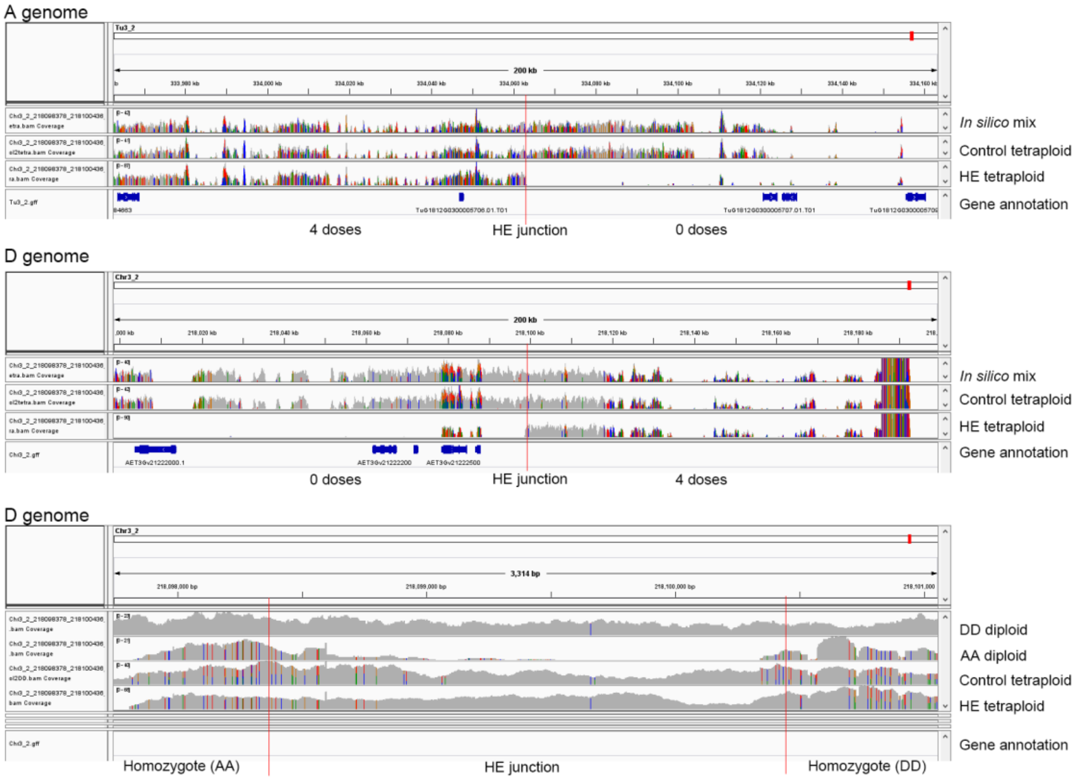
HE junction 11.



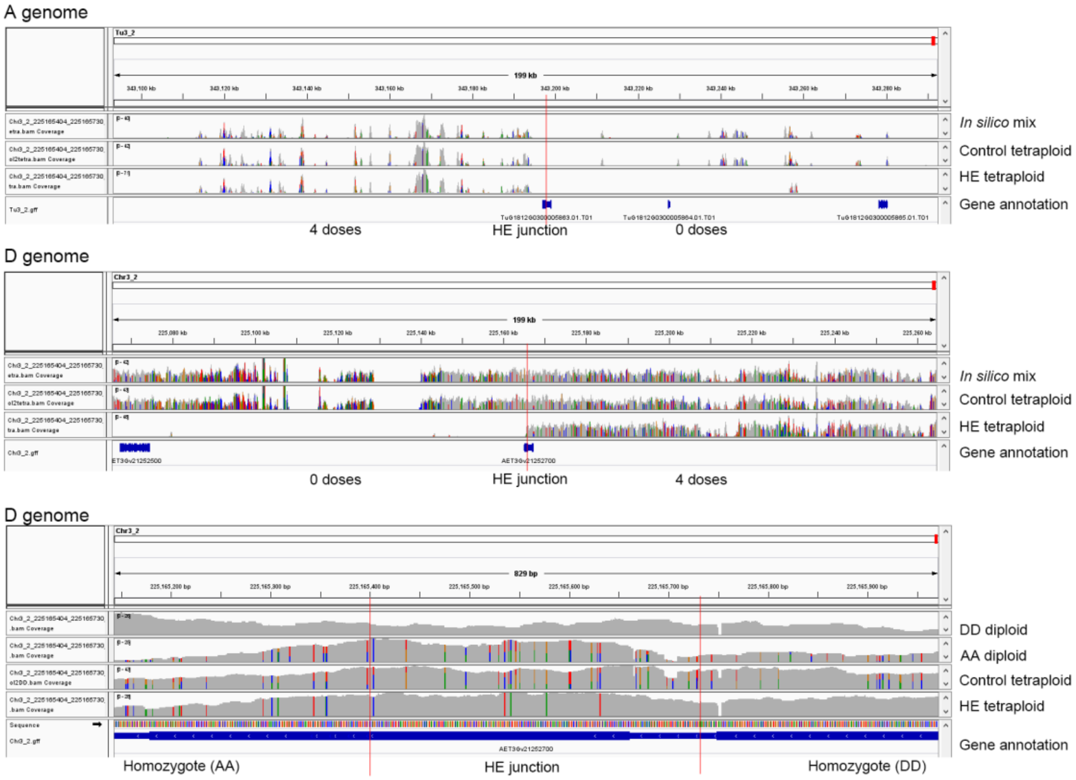
HE junction 12.



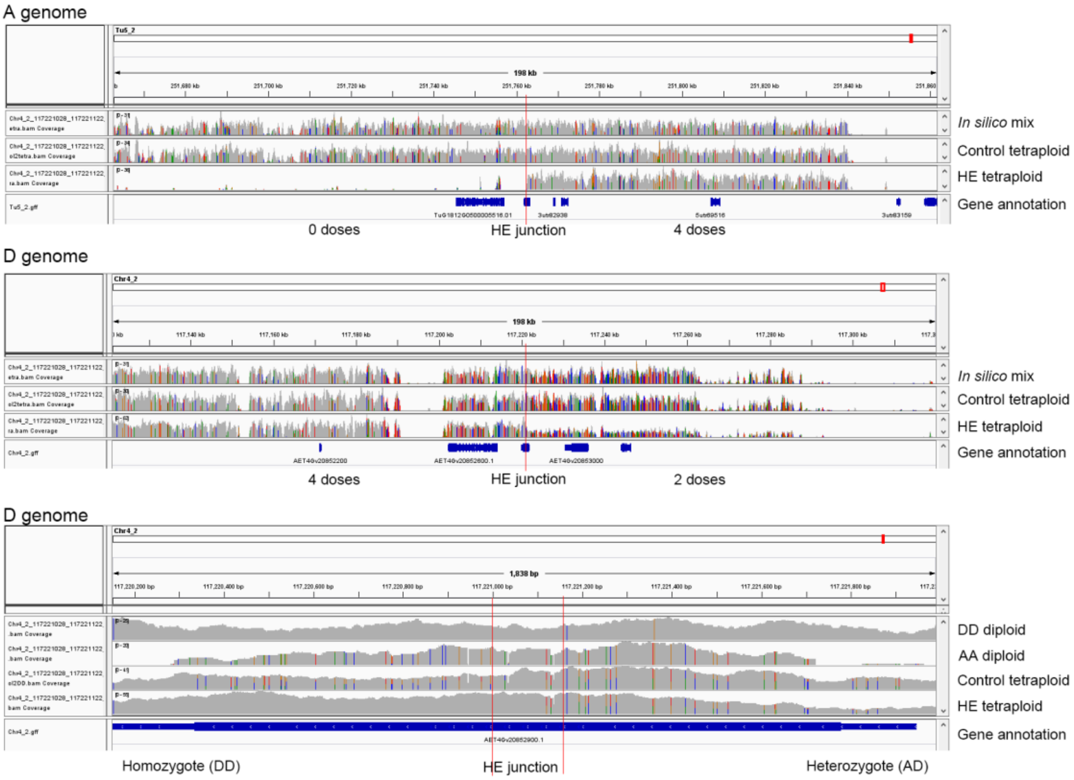
HE junction 13.



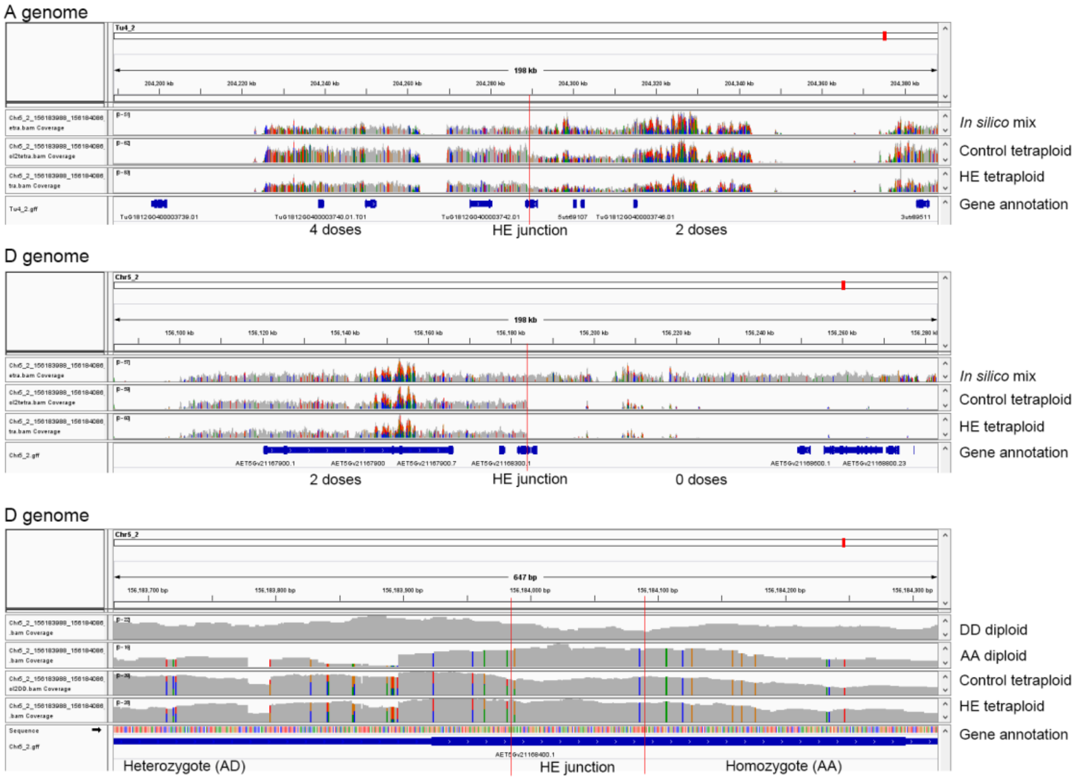
HE junction 14.



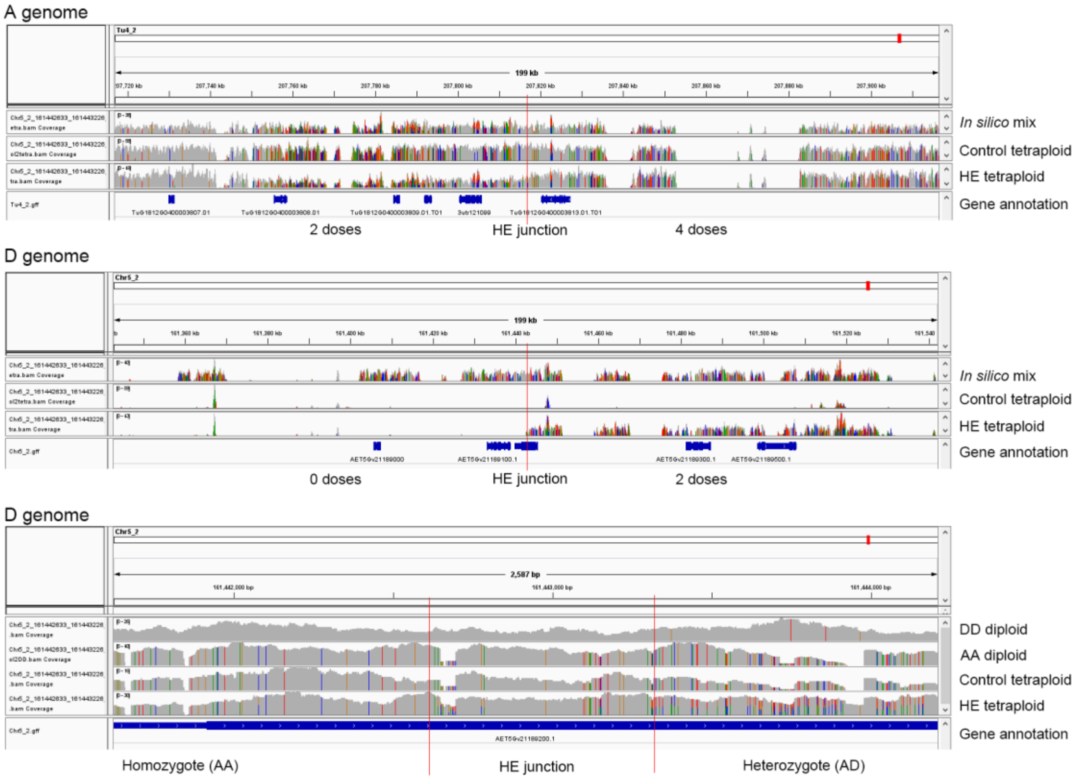
HE junction 15.



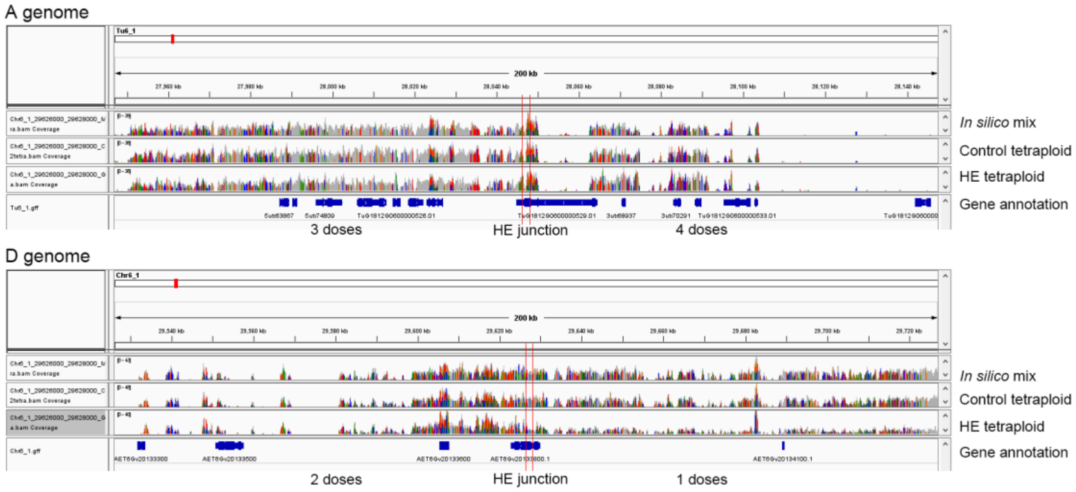
HE junction 16.



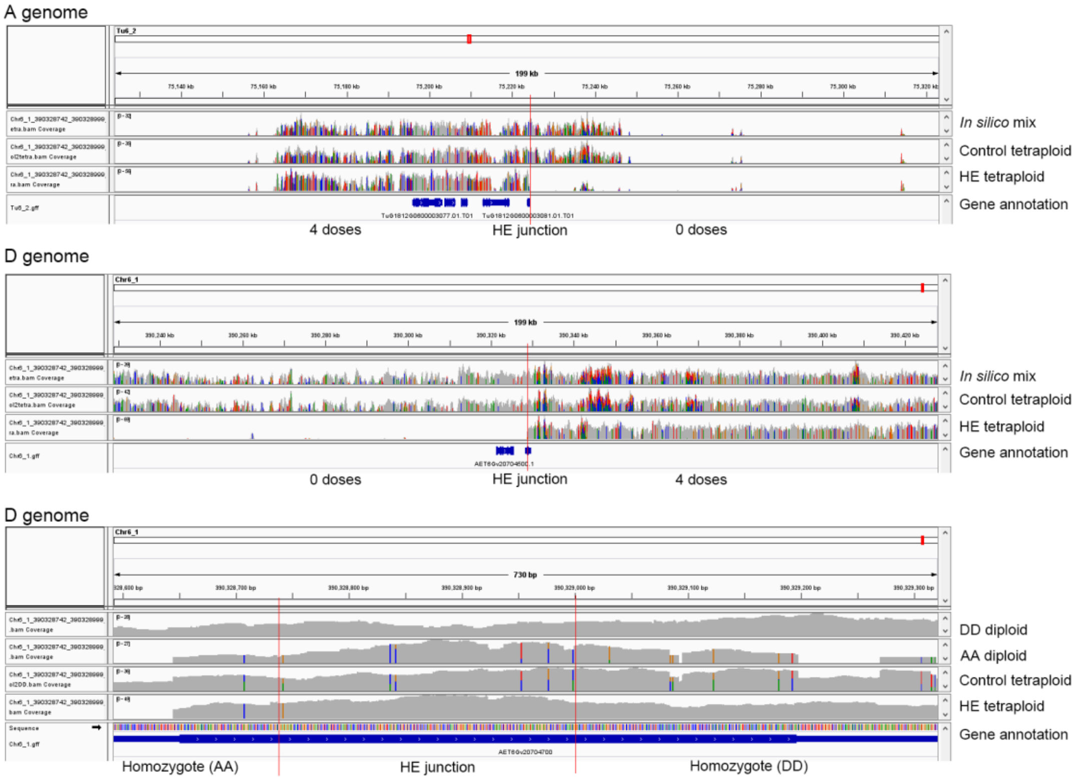
HE junction 17.



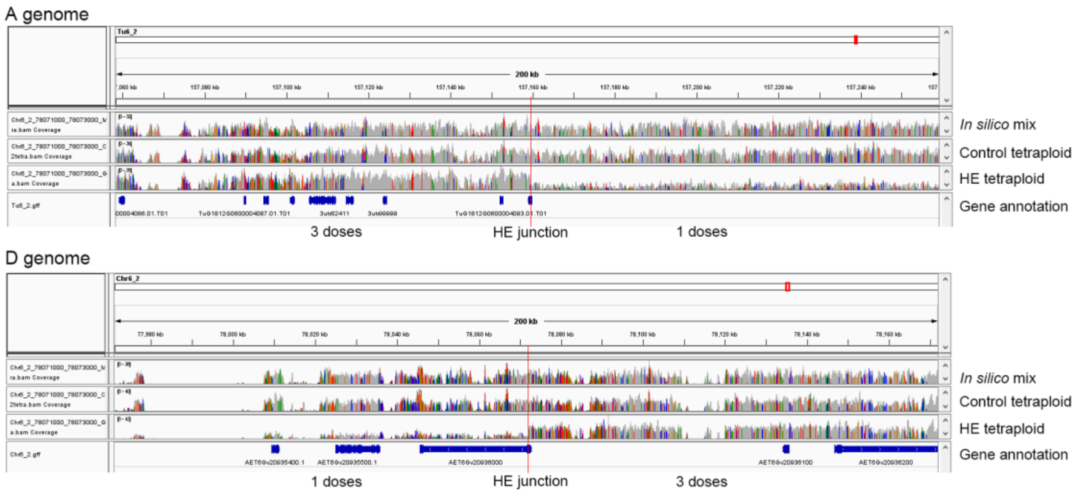
HE junction 18.



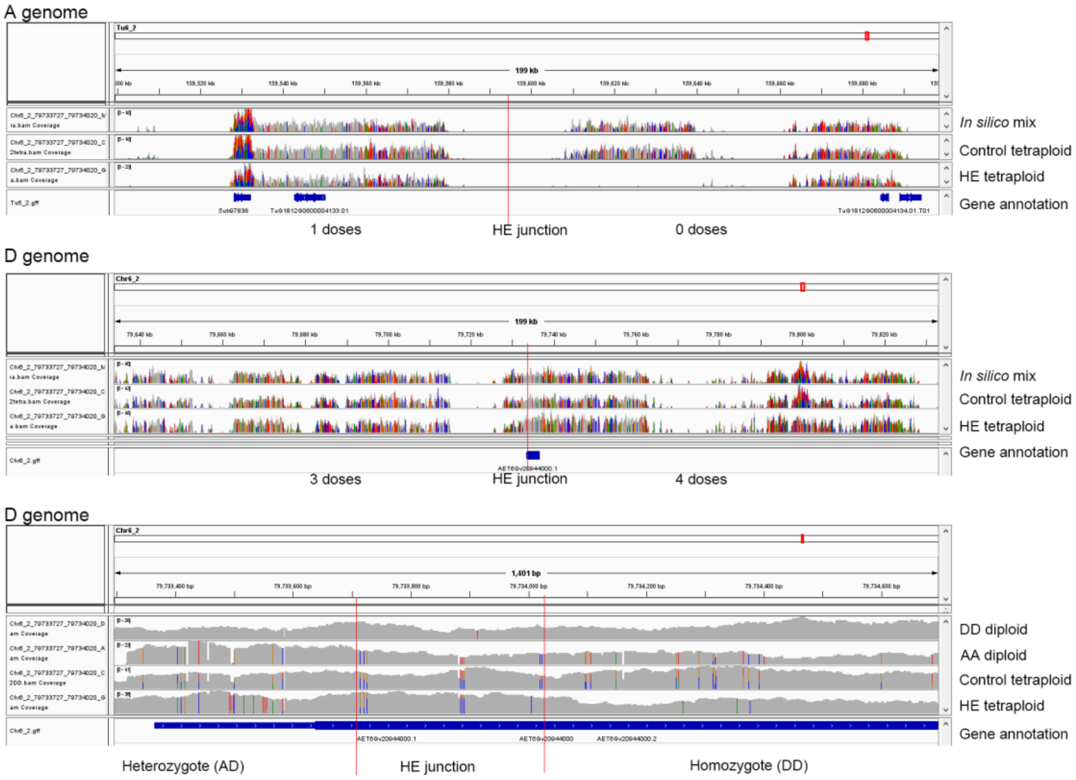
HE junction 19.



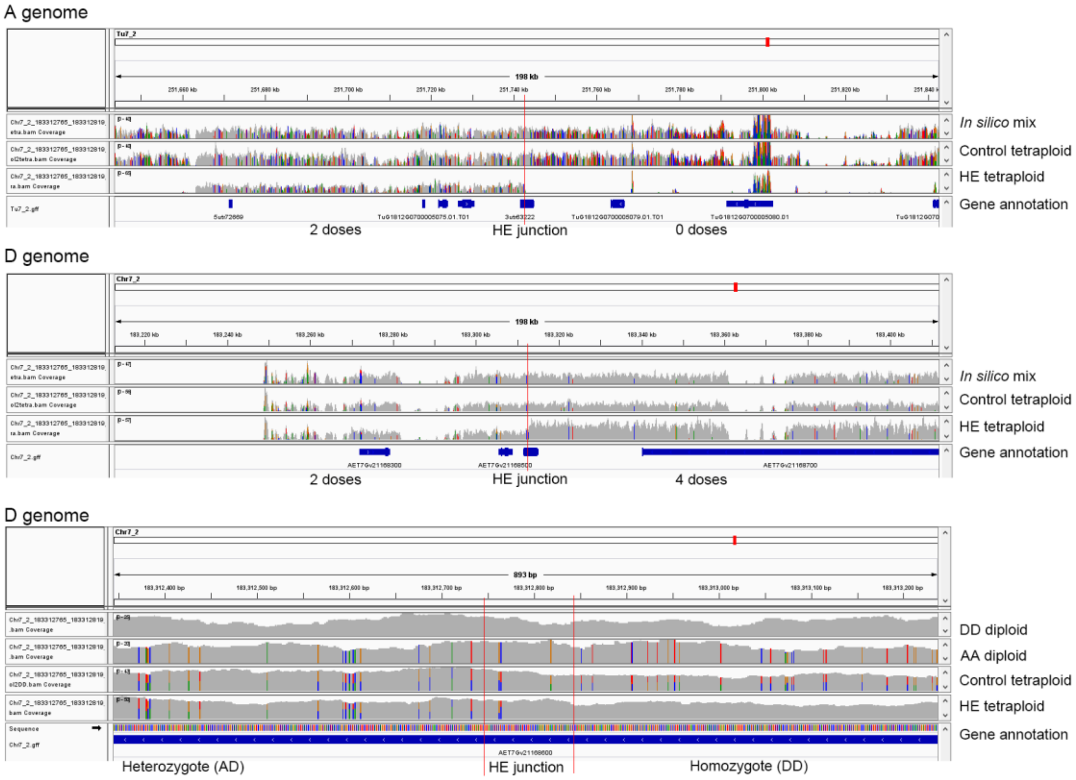
HE junction 20.



HE junction 21.

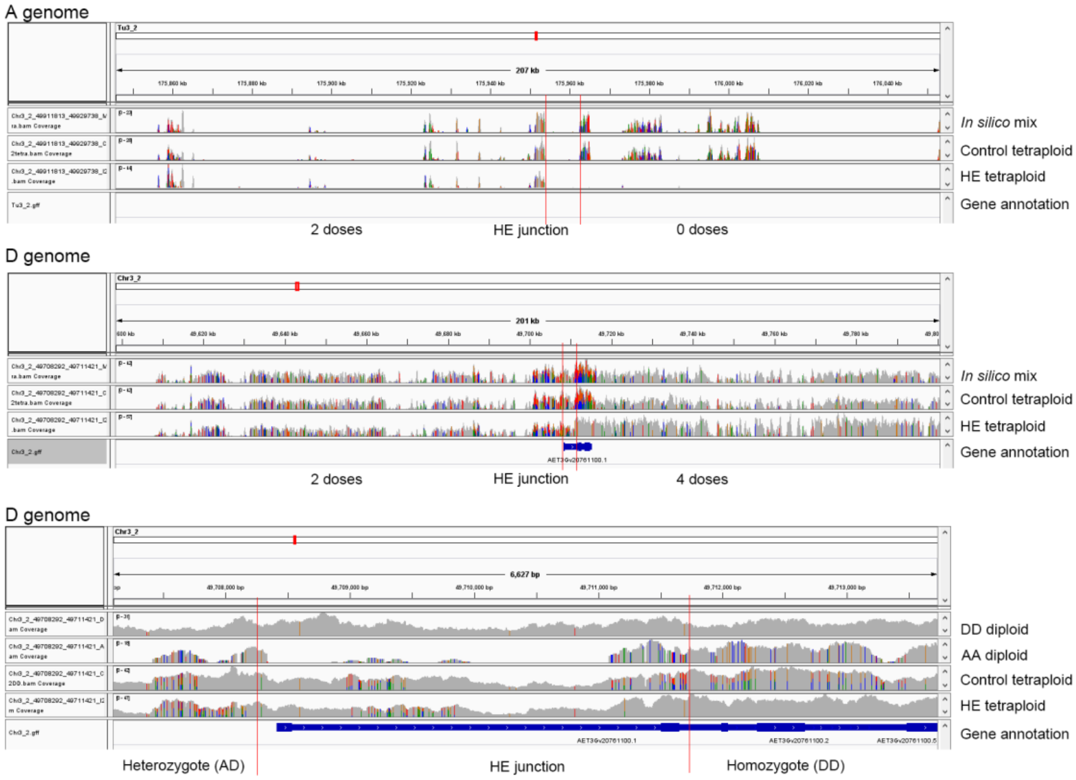


HE junction 22.

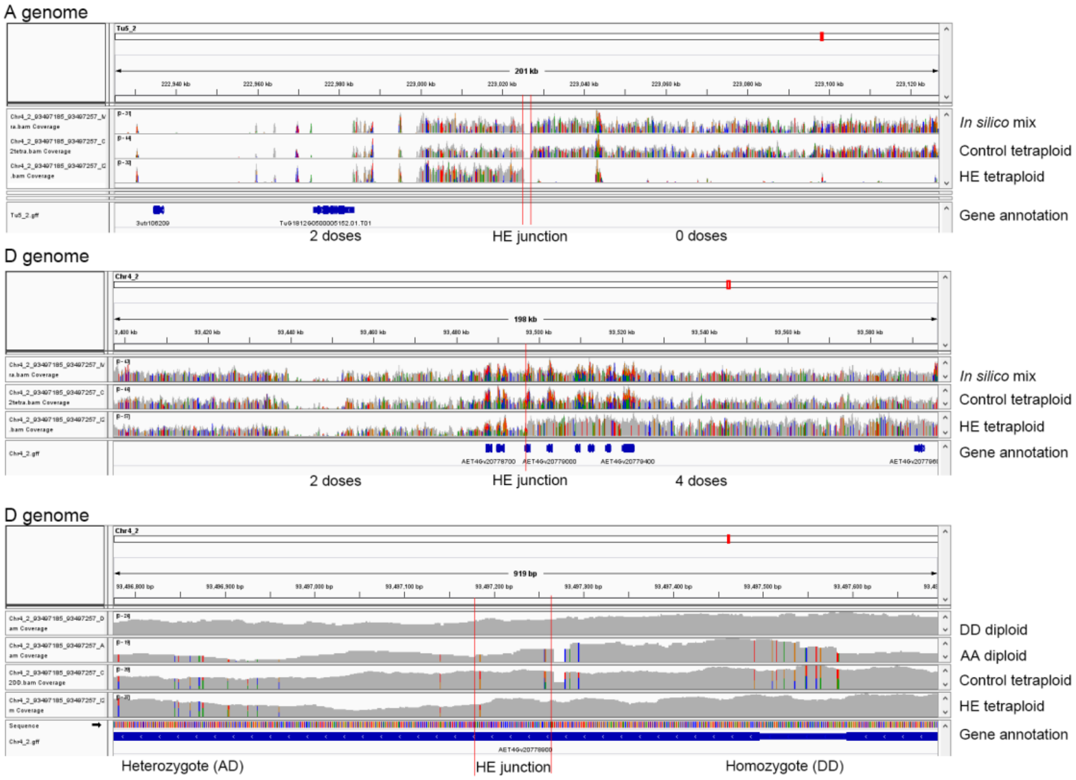


HE junction 23.

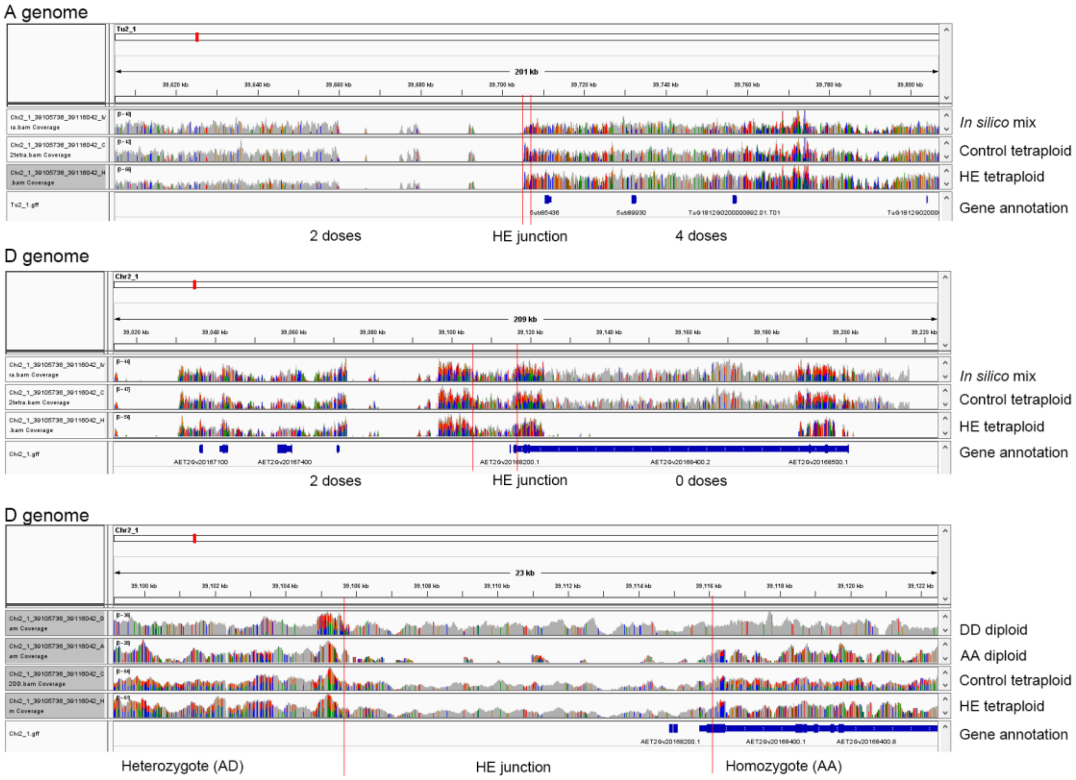
HE junction 24.



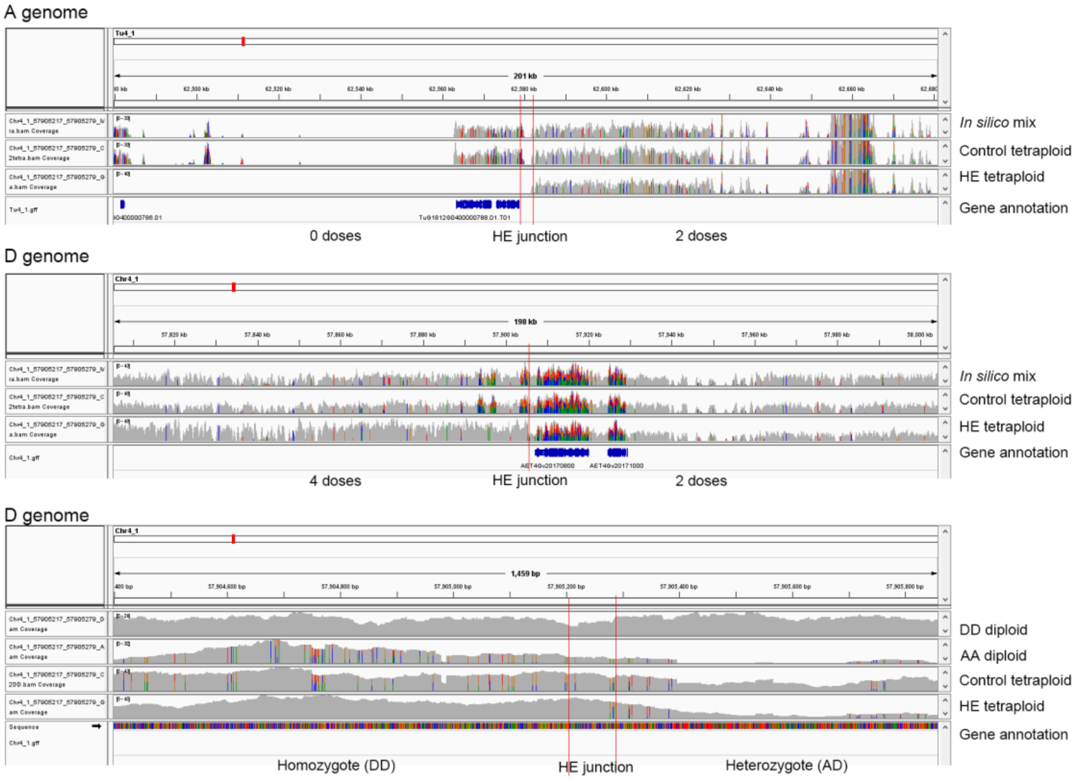
HE junction 25



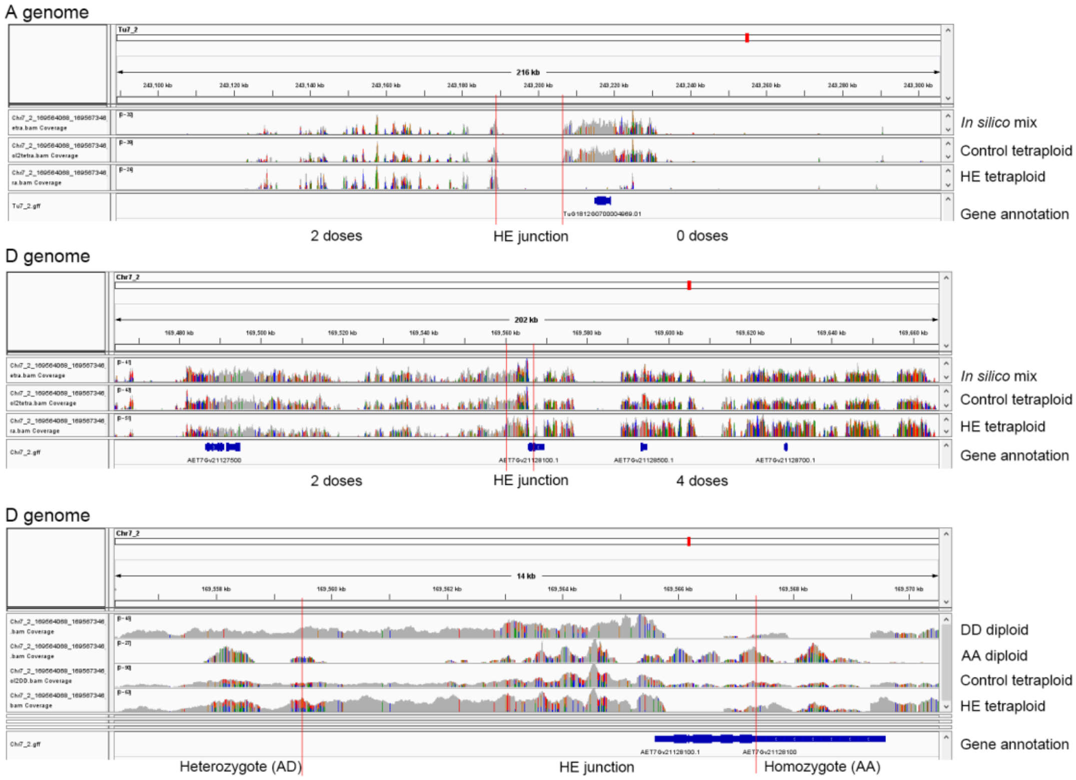
HE junction 26.



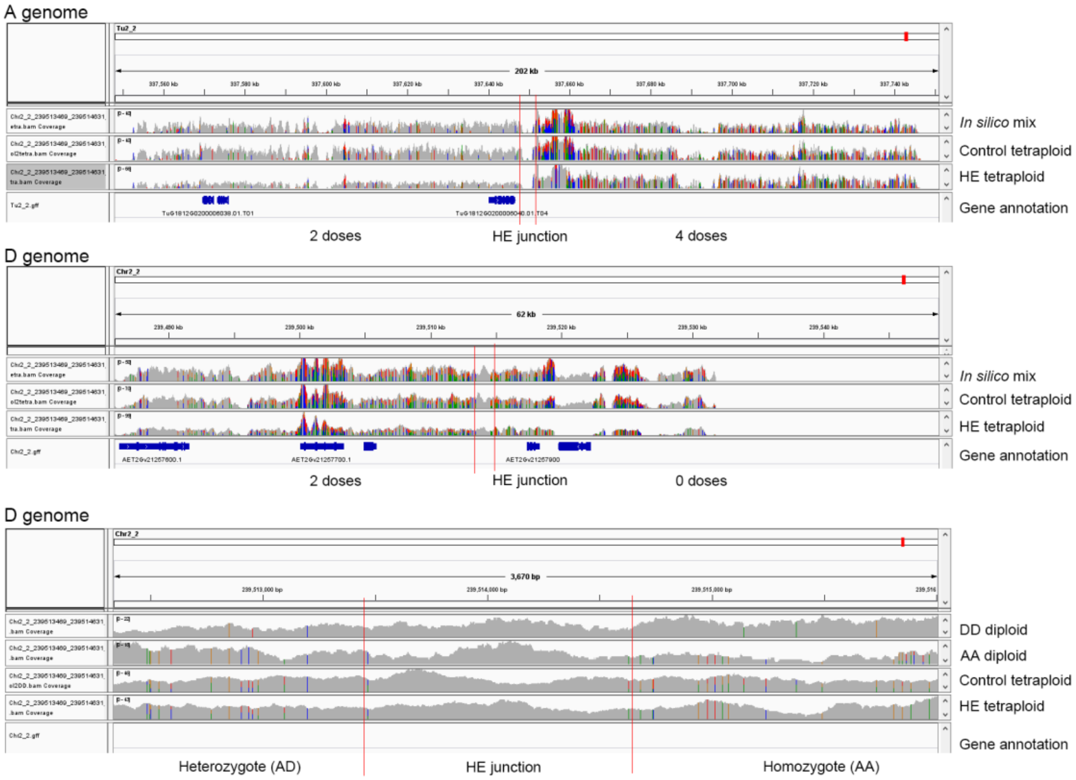
HE junction 27.



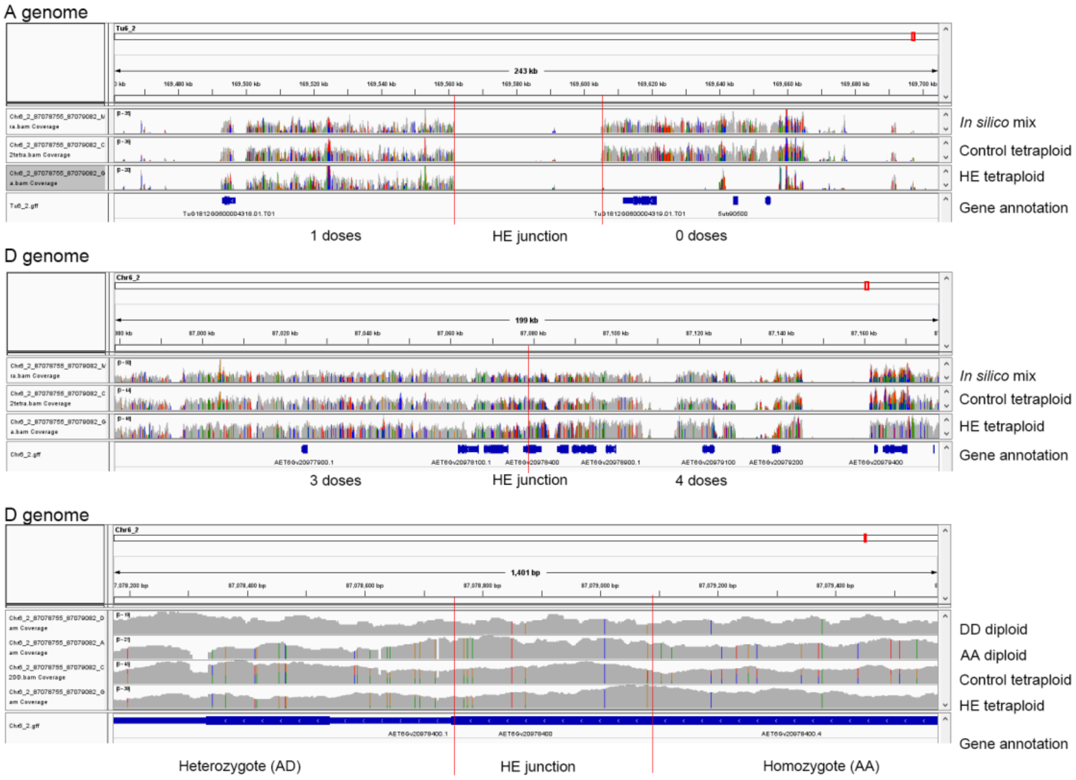
HE junction 28.



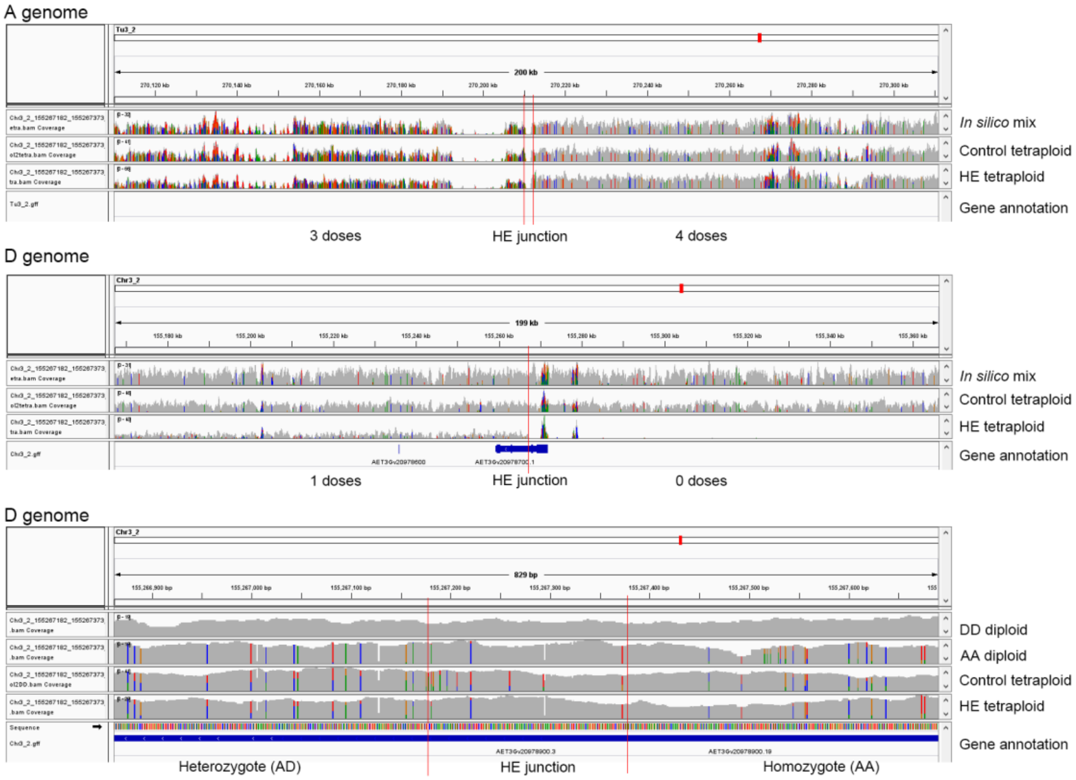
HE junction 29.



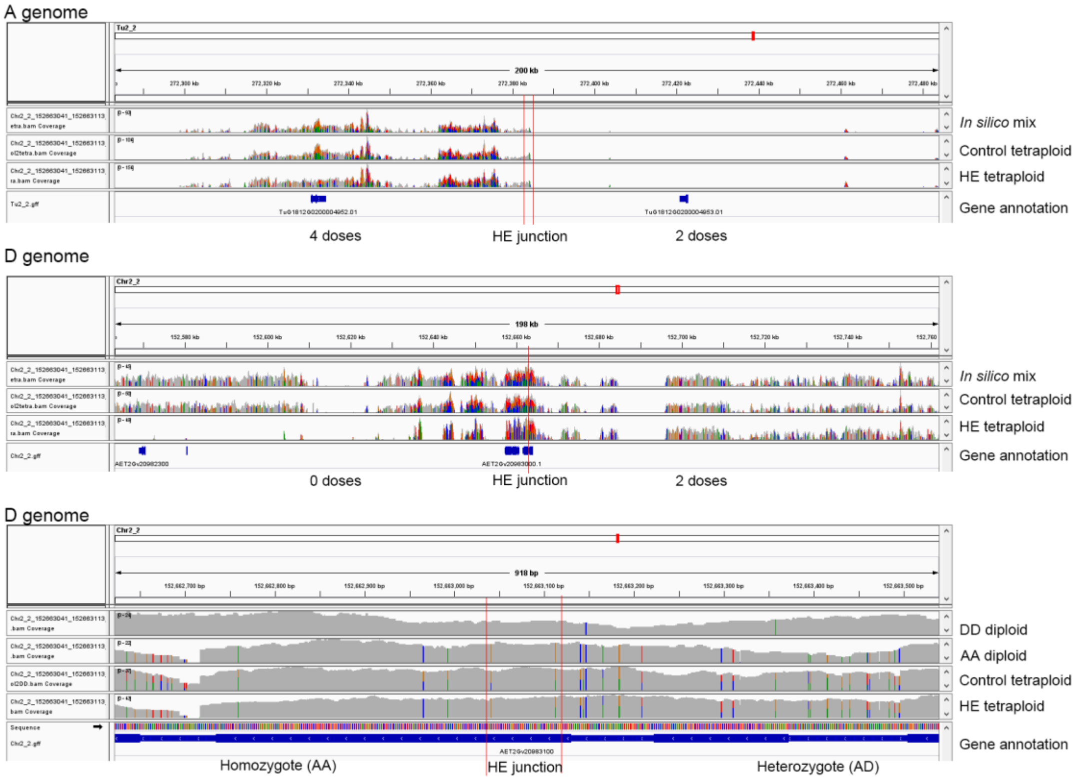
HE junction 30.



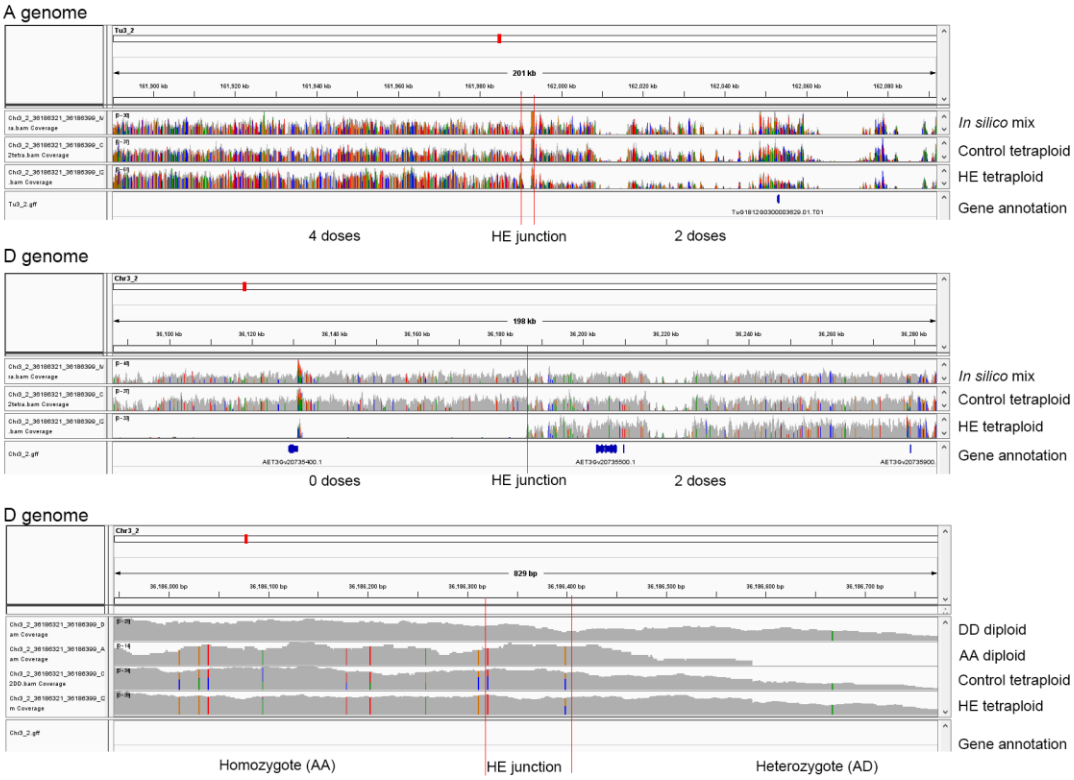
HE junction 31.



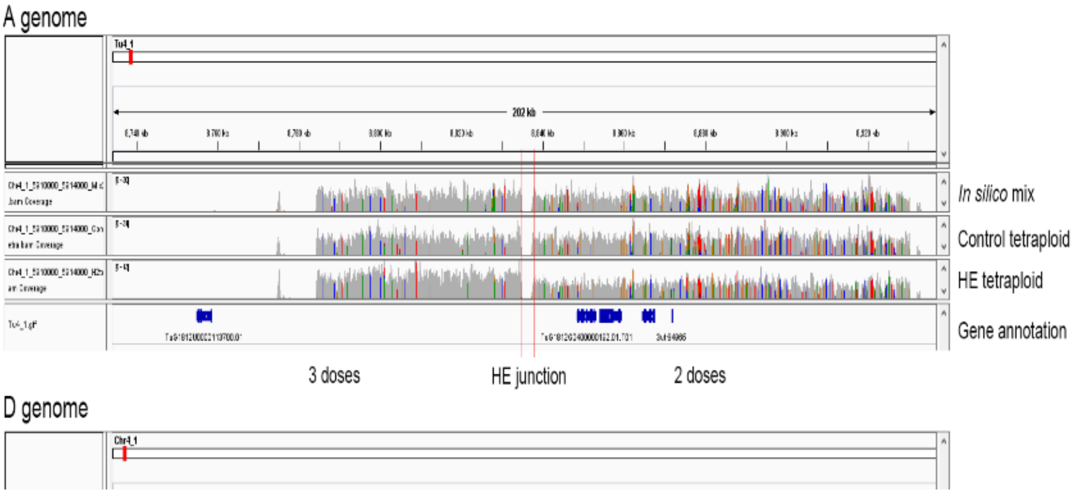
HE junction 32.



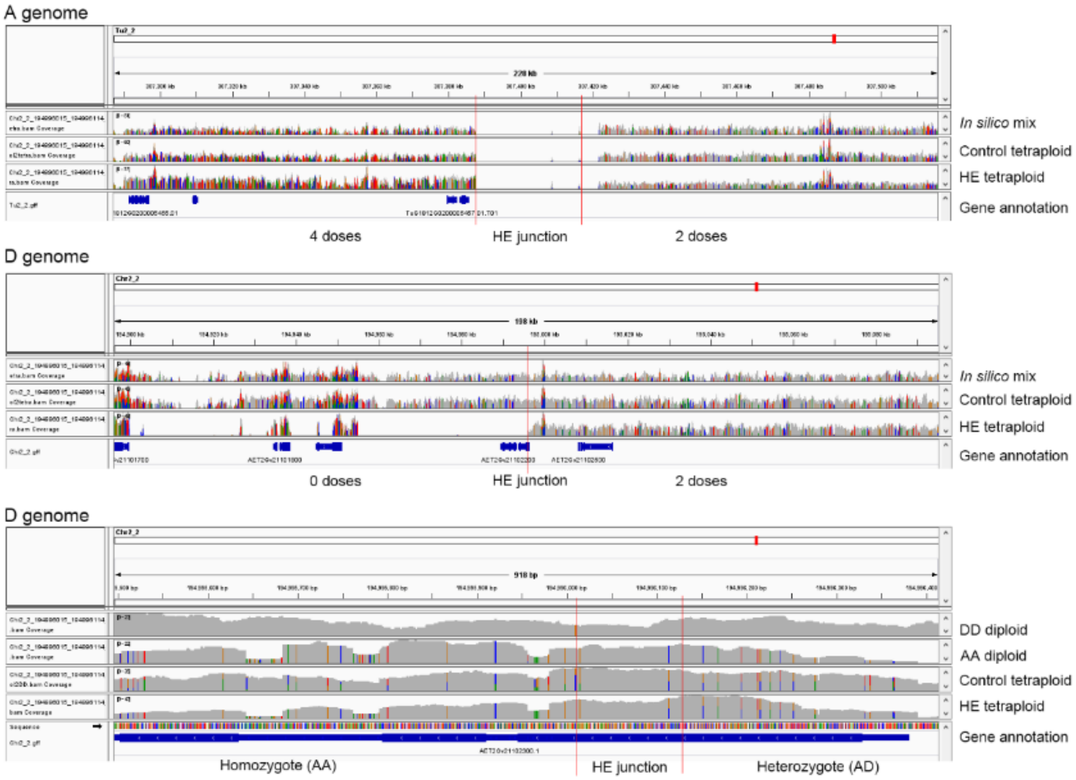
HE junction 33.



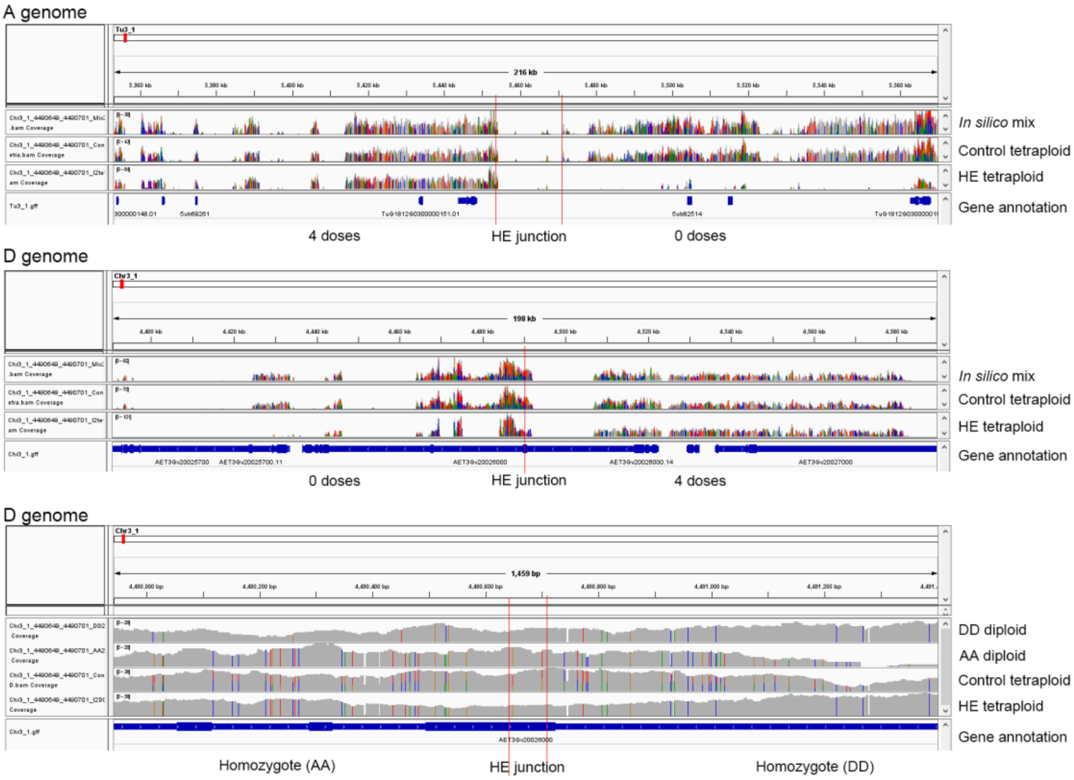
HE junction 34.



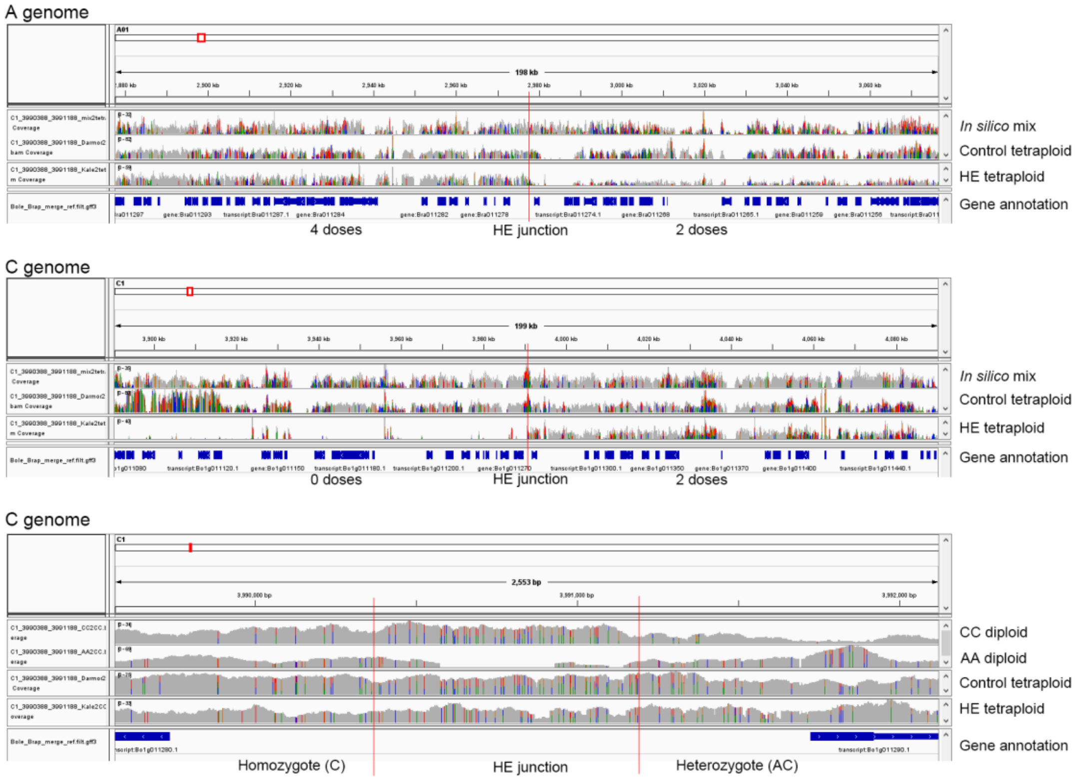
HE junction 35.



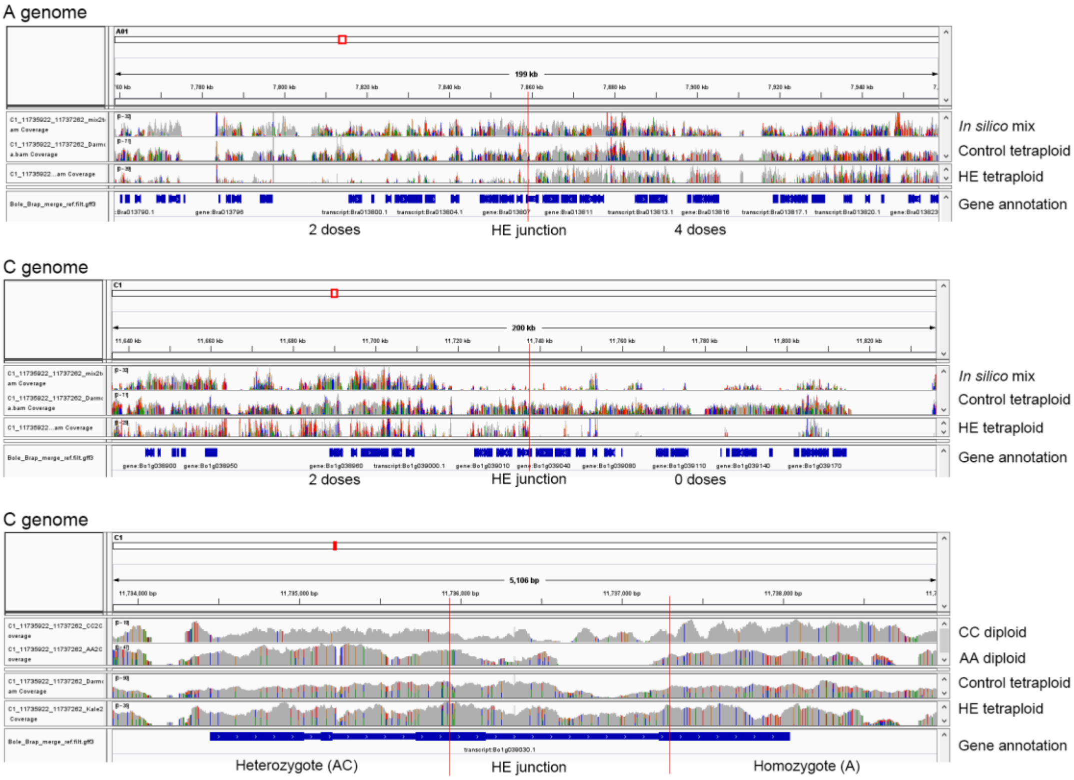
HE junction 36.



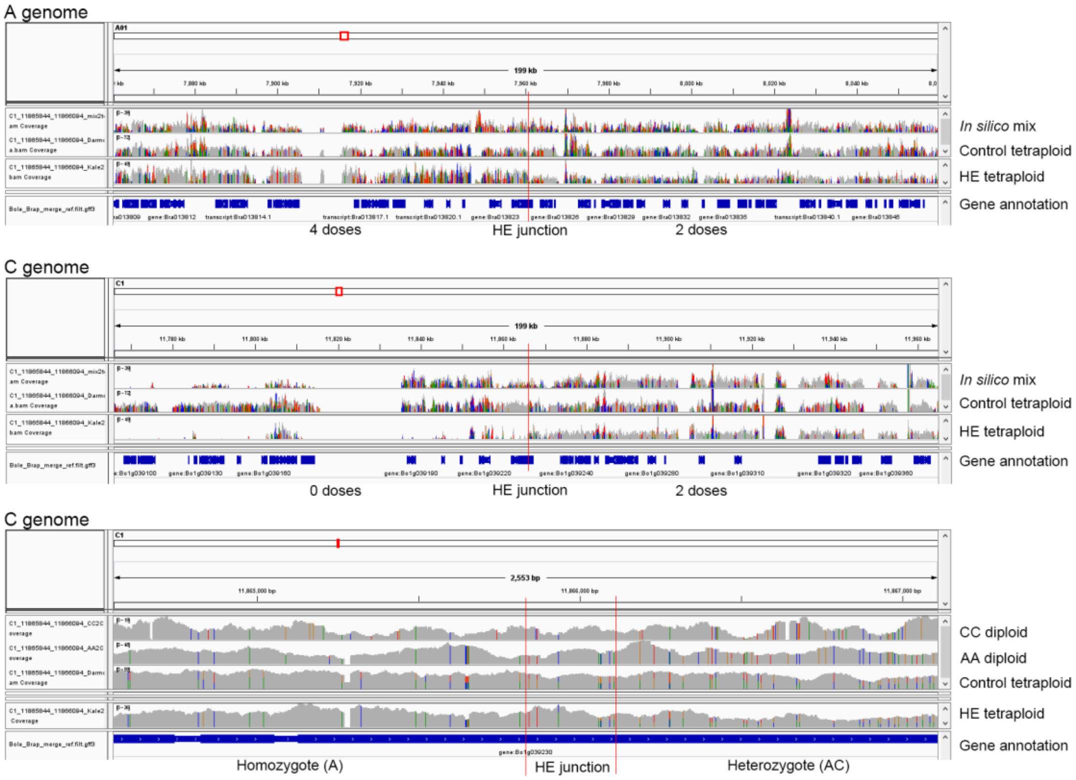
HE junction 37.



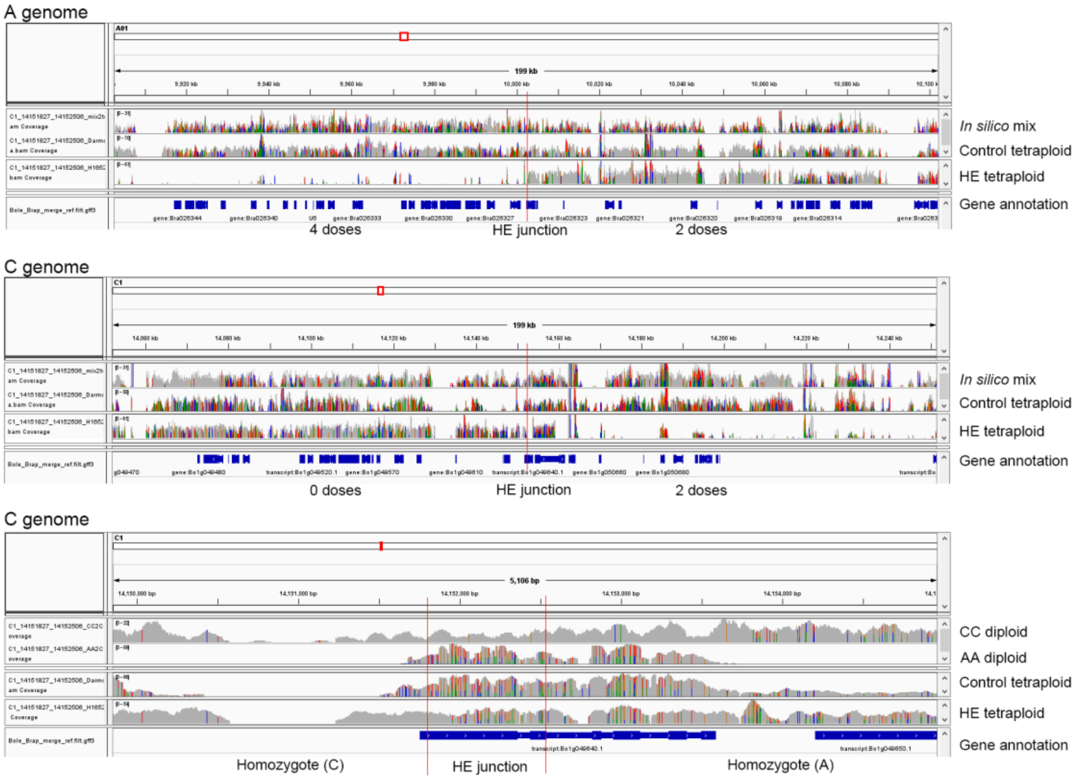
IGV panels of HE junction regions in *Brassica napus* (104 junctions). The top and middle panels show the sequencing depth around HE junctions in both *B. oleracea* (C) and *B. rapa* (A) genomes, respectively. Red lines indicate the position of HE junctions and genome dosages are shown on both sides of HE junction. The bottom panel shows the genotypes (homozygote or heterozygote) around HE junction (region between red lines) based on SNP genotyping.



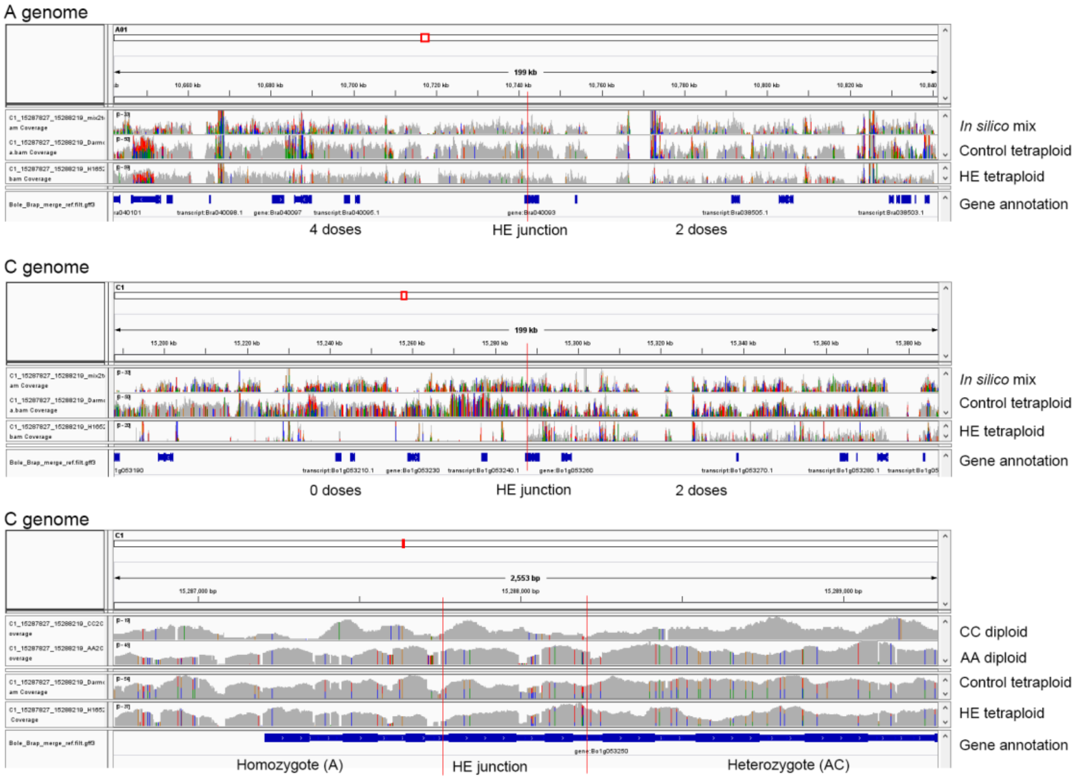
HE junction 2.



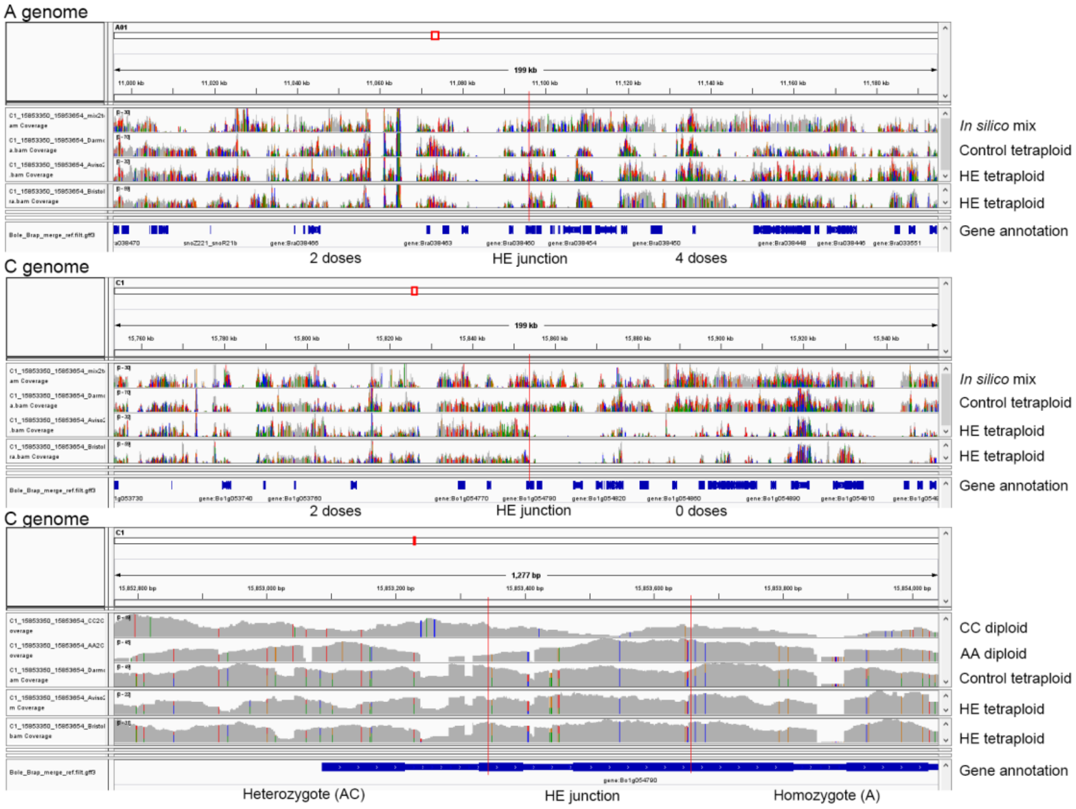
HE junction 3.



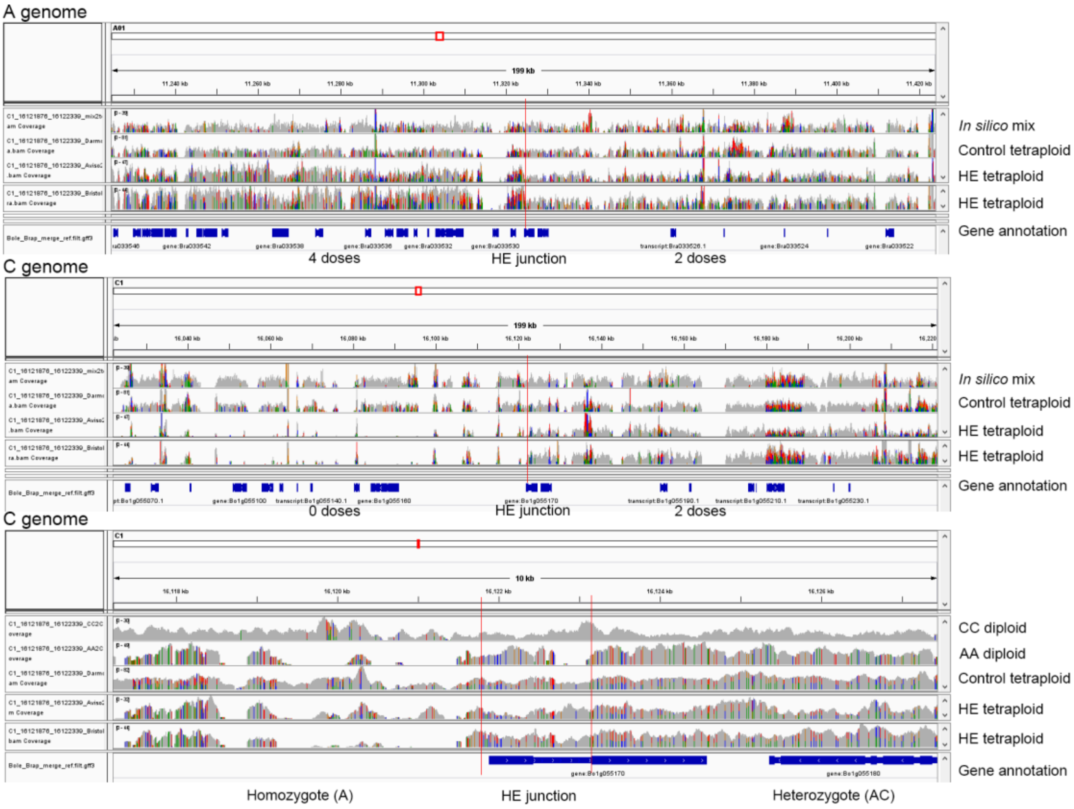
HE junction 4.



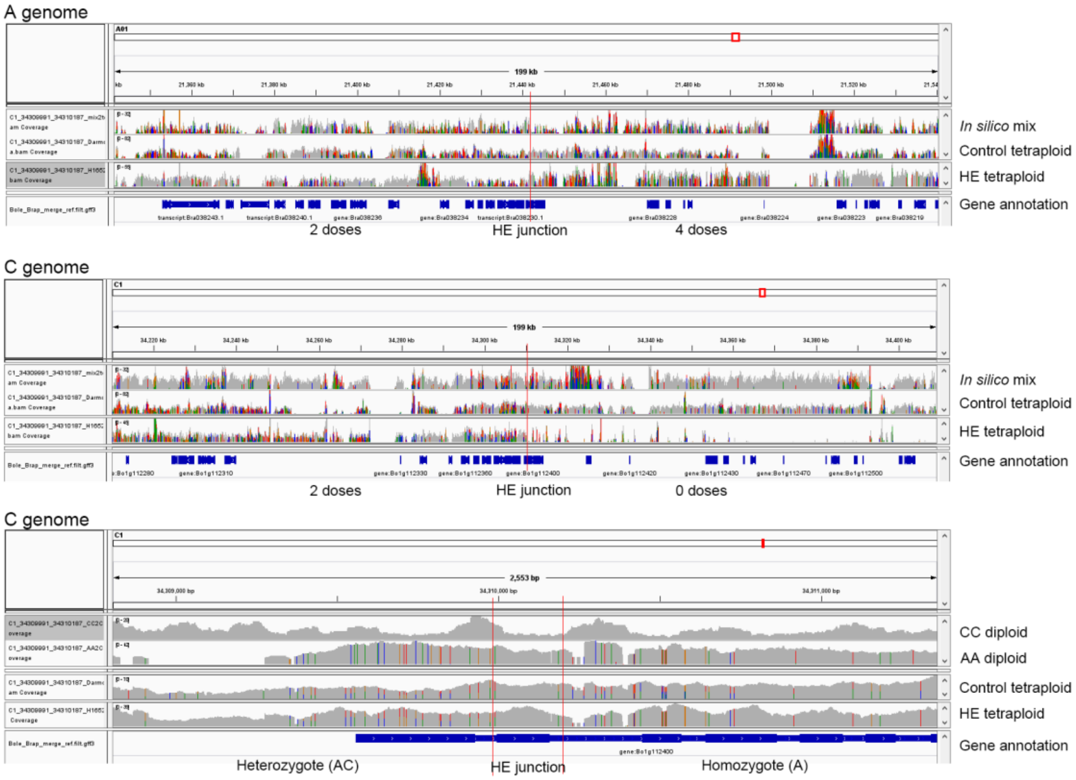
HE junction 5.



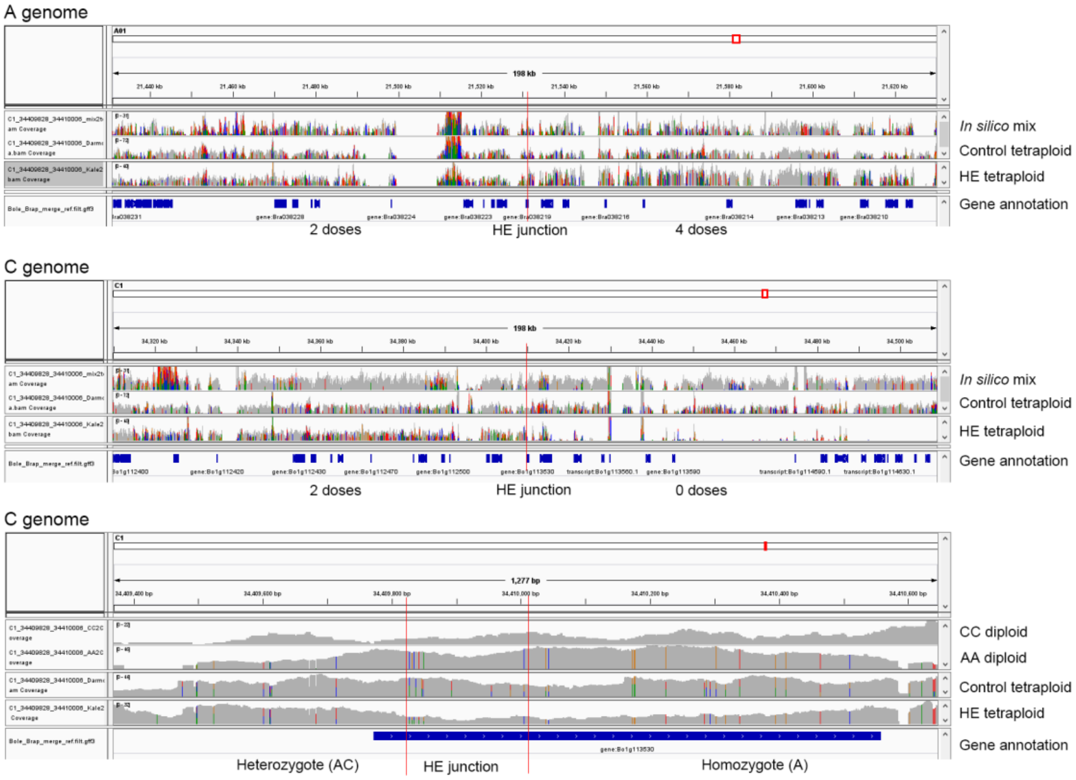
HE junction 6.



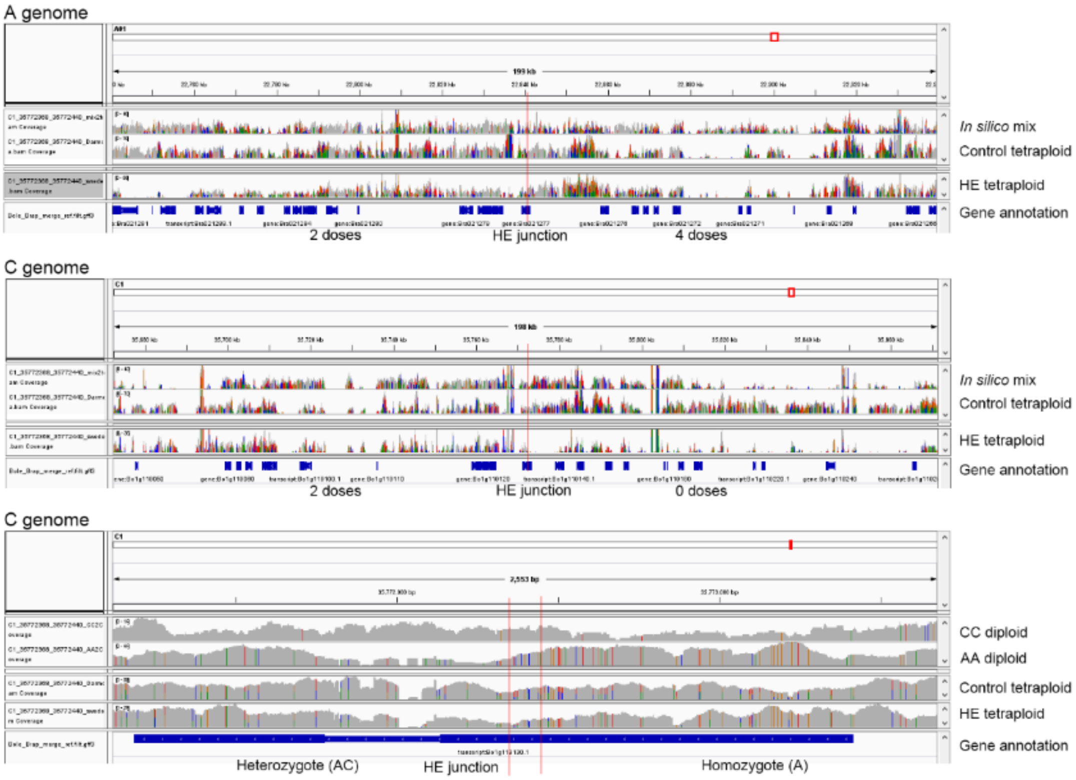
HE junction 7.



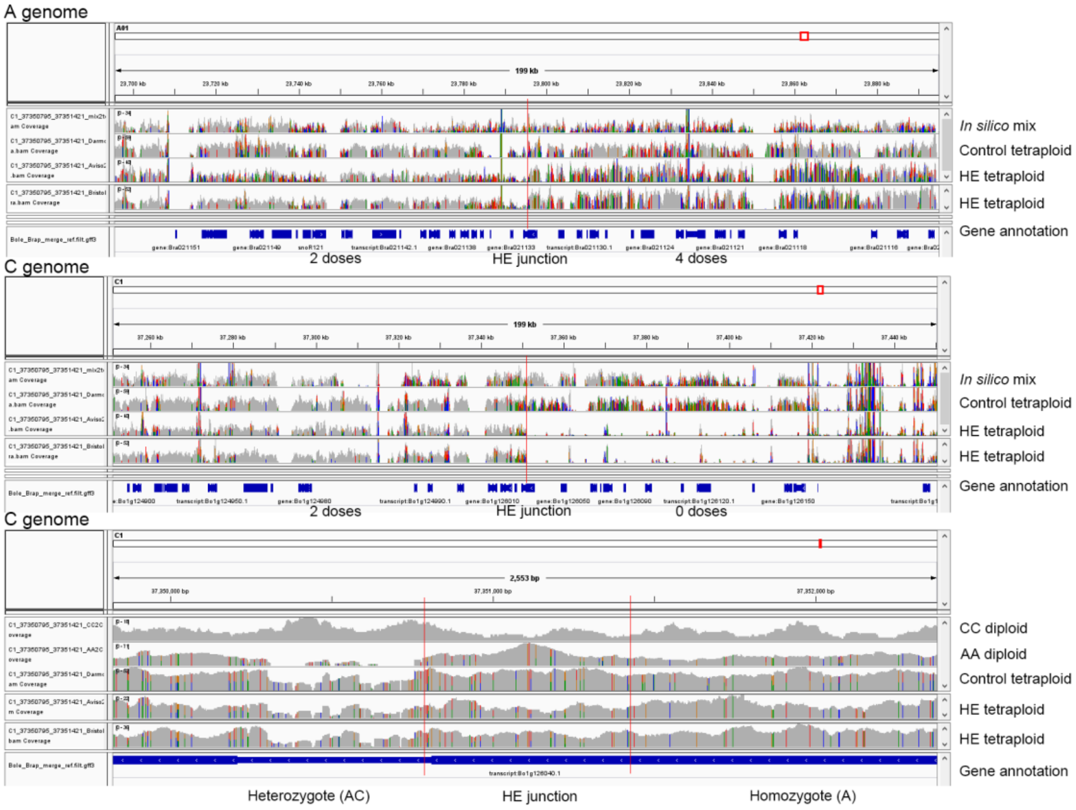
HE junction 8.



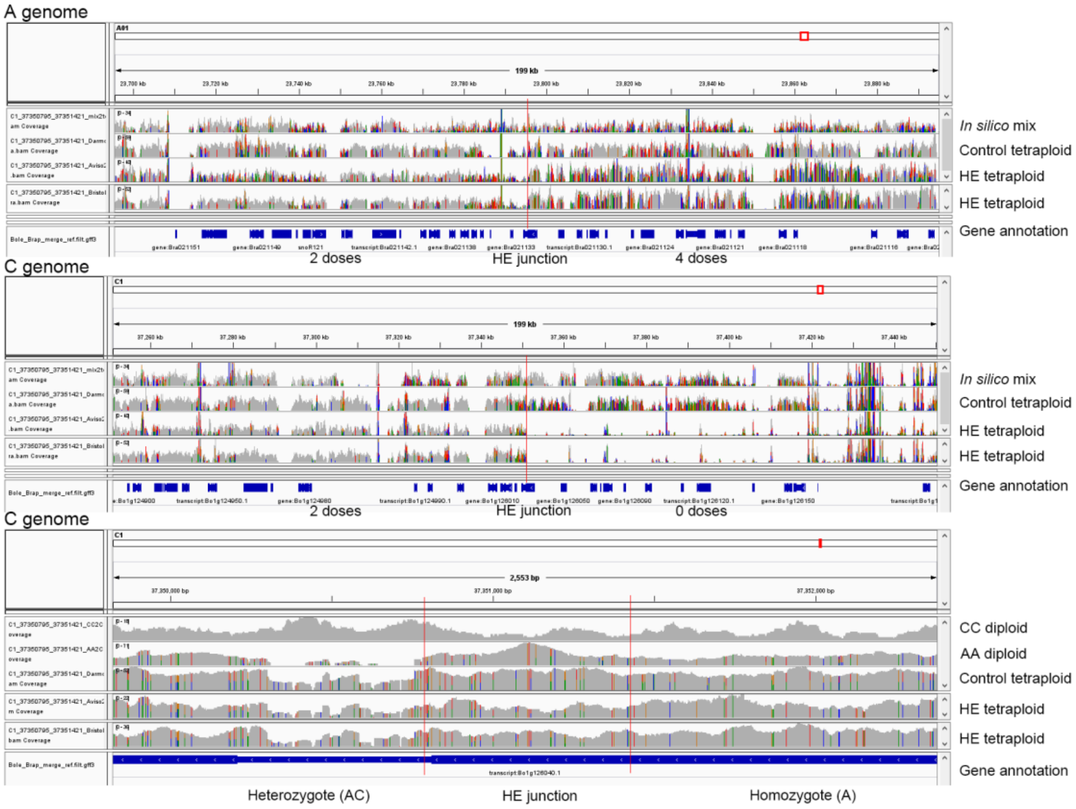
HE junction 9.



HE junction 10.



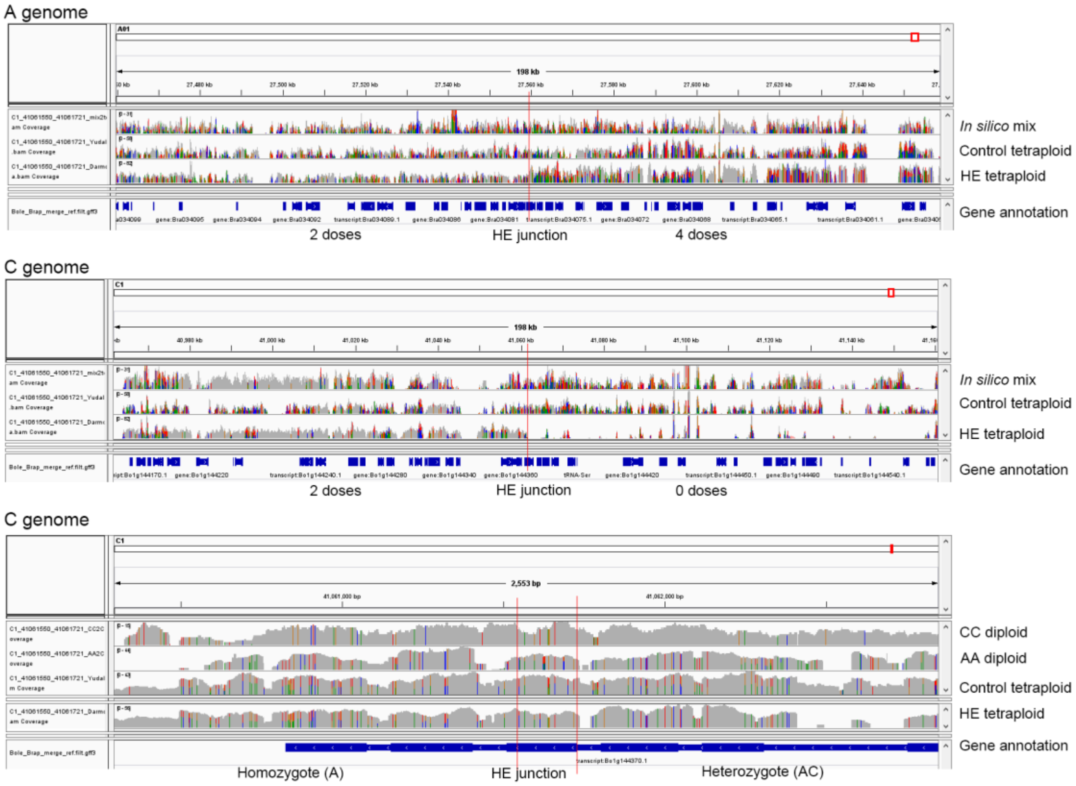
HE junction 11.



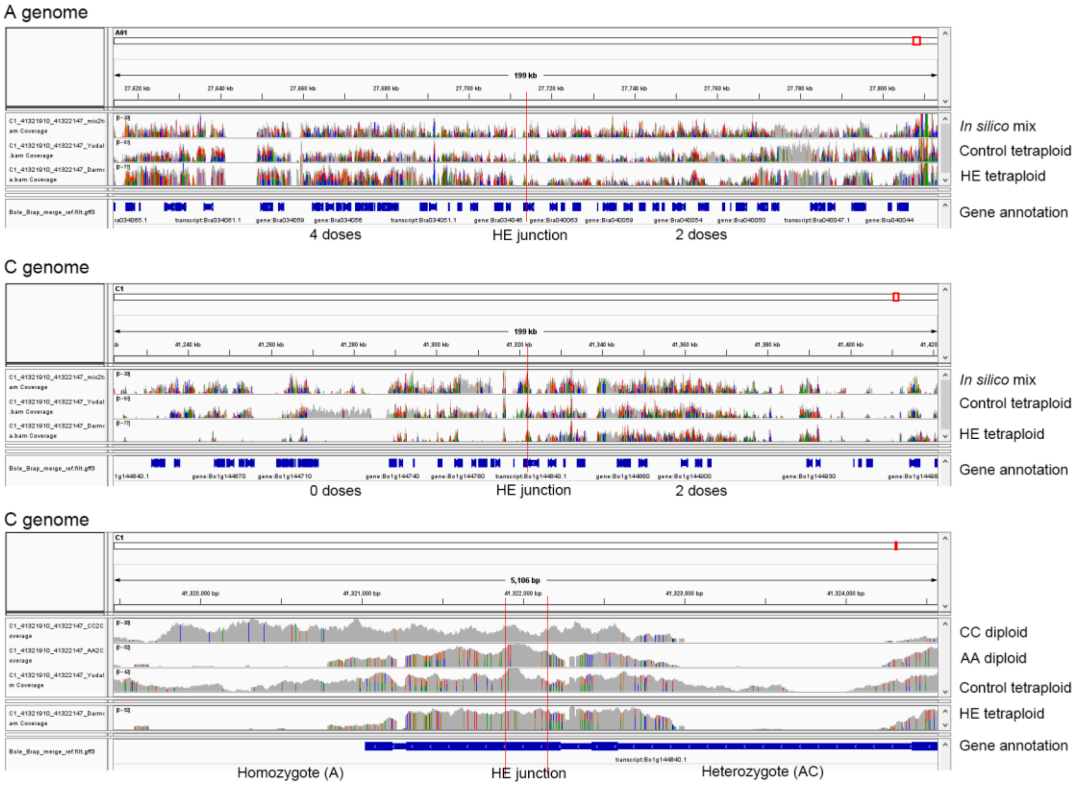
HE junction 12.

HE junction 13.

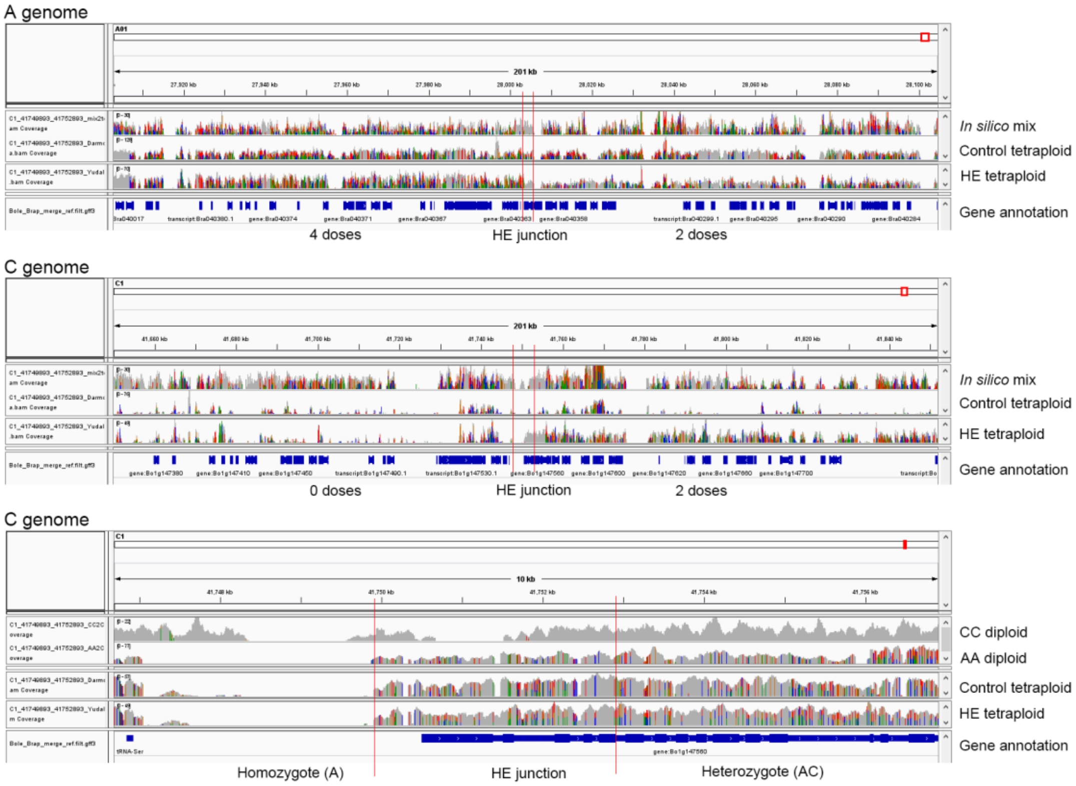
HE junction 14.



HE junction 15.



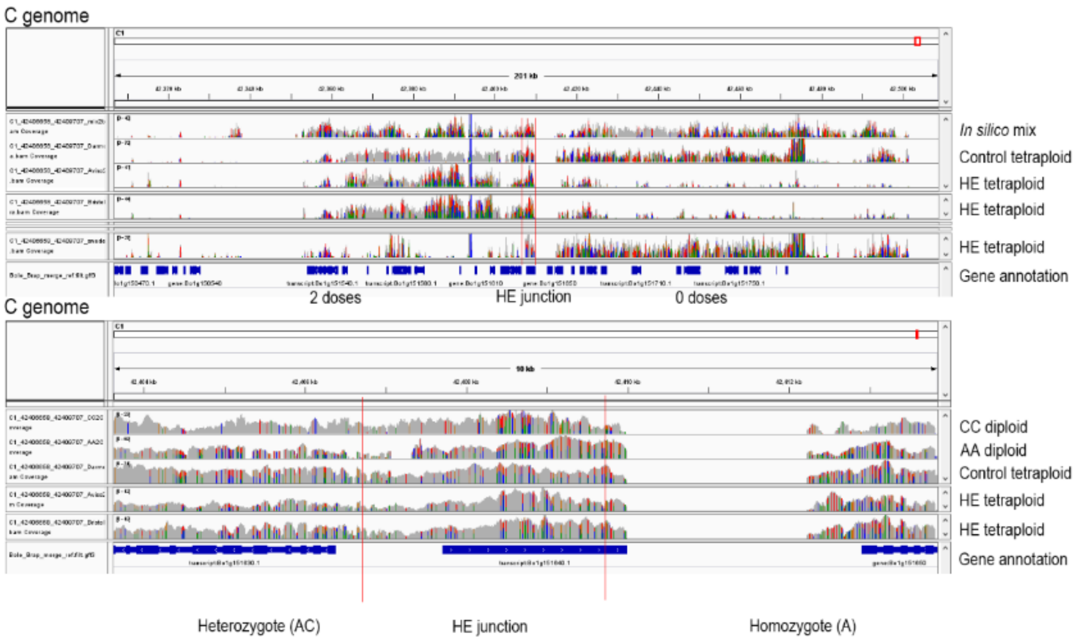
HE junction 16.


A genome

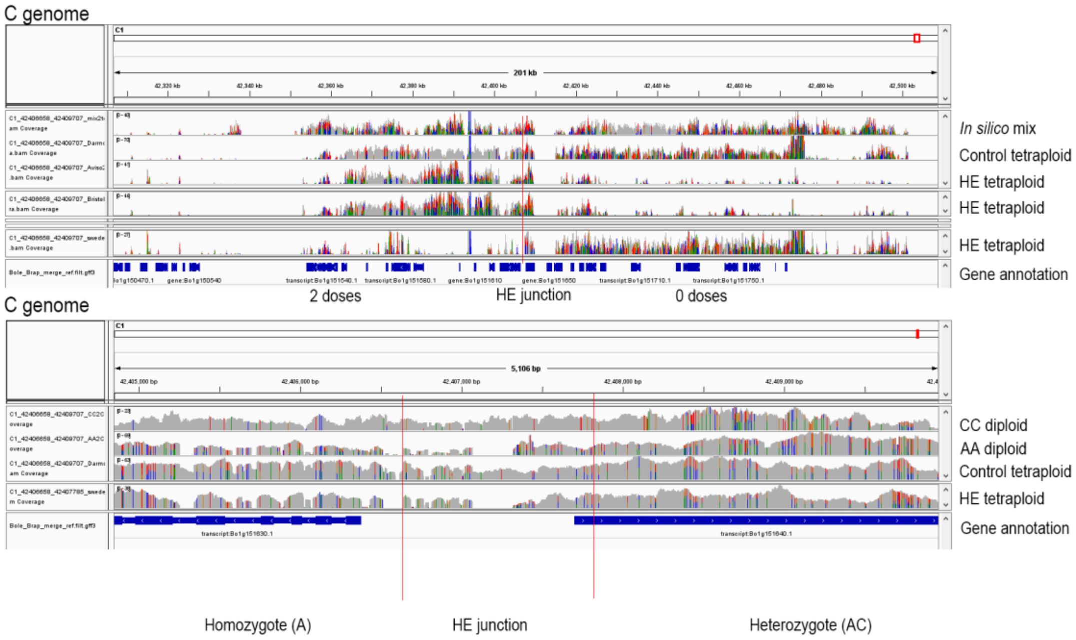
C genome

C genome

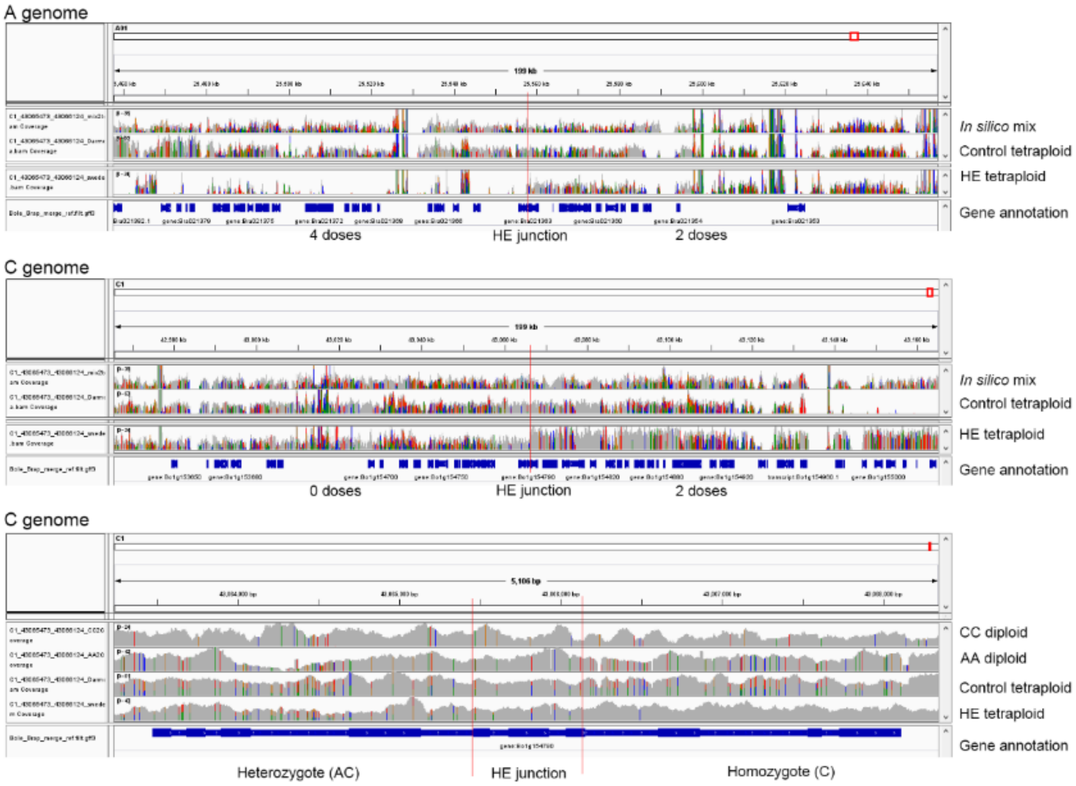
HE junction 17.



HE junction 18.

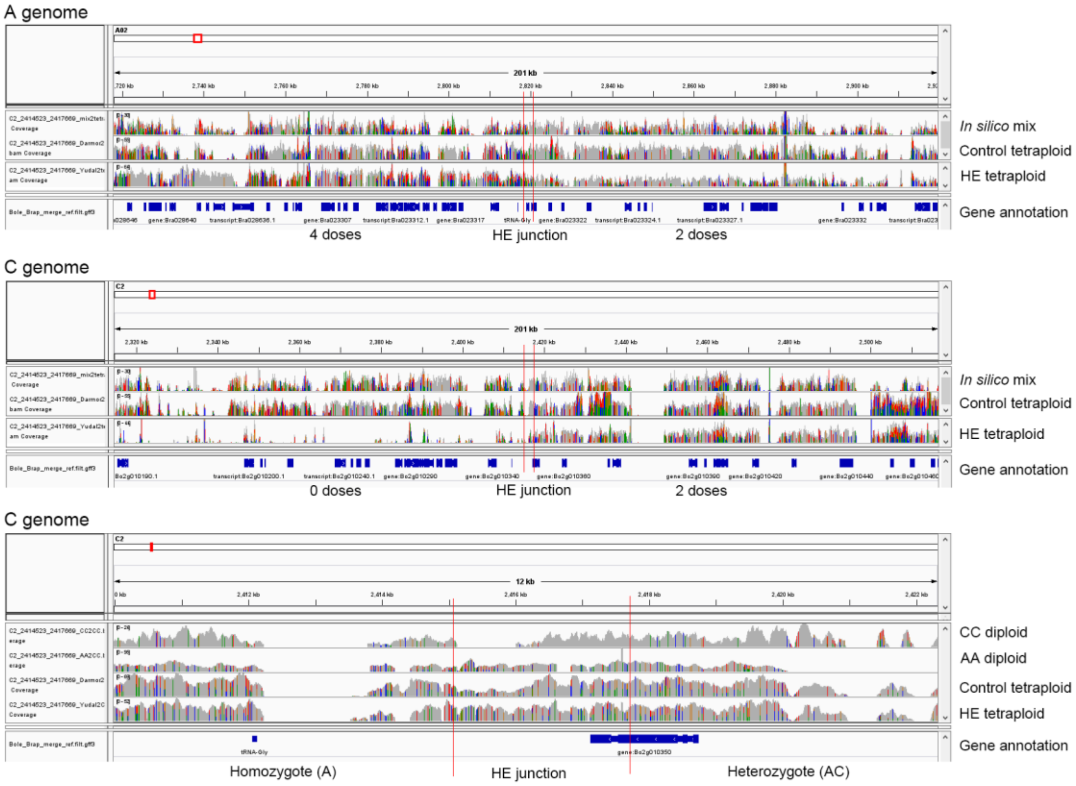


HE junction 19.

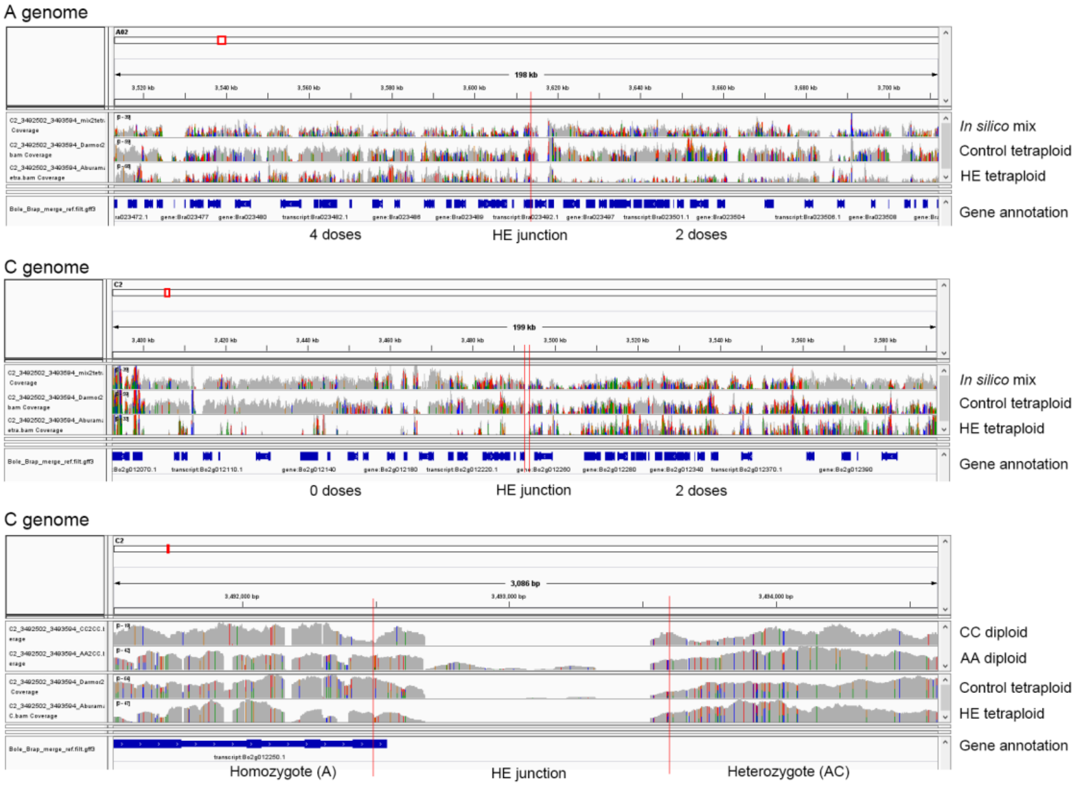


HE junction 20.

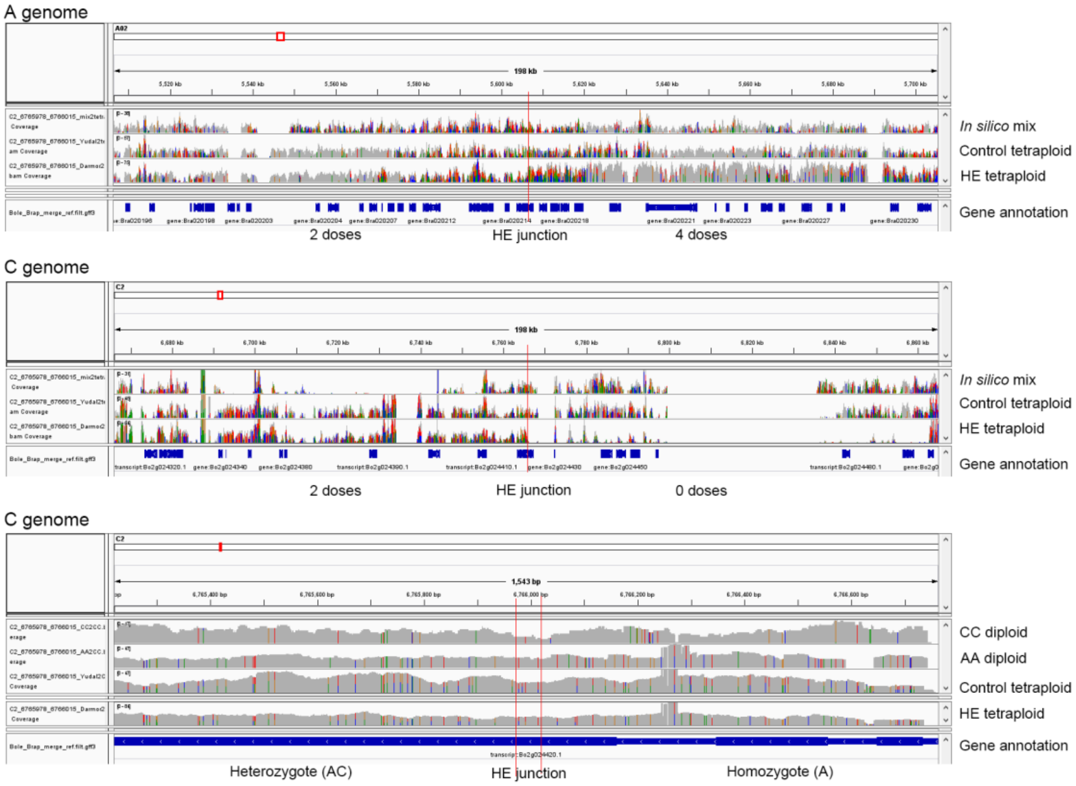
HE junction 21.



HE junction 22.

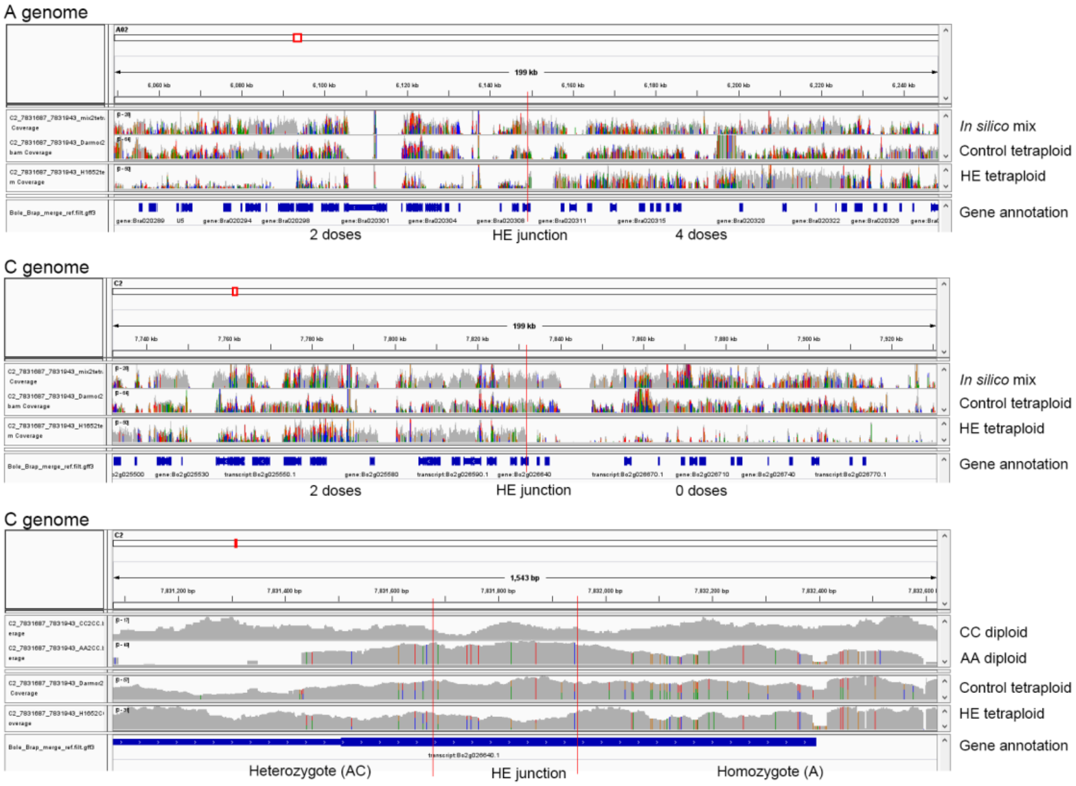


HE junction 23.

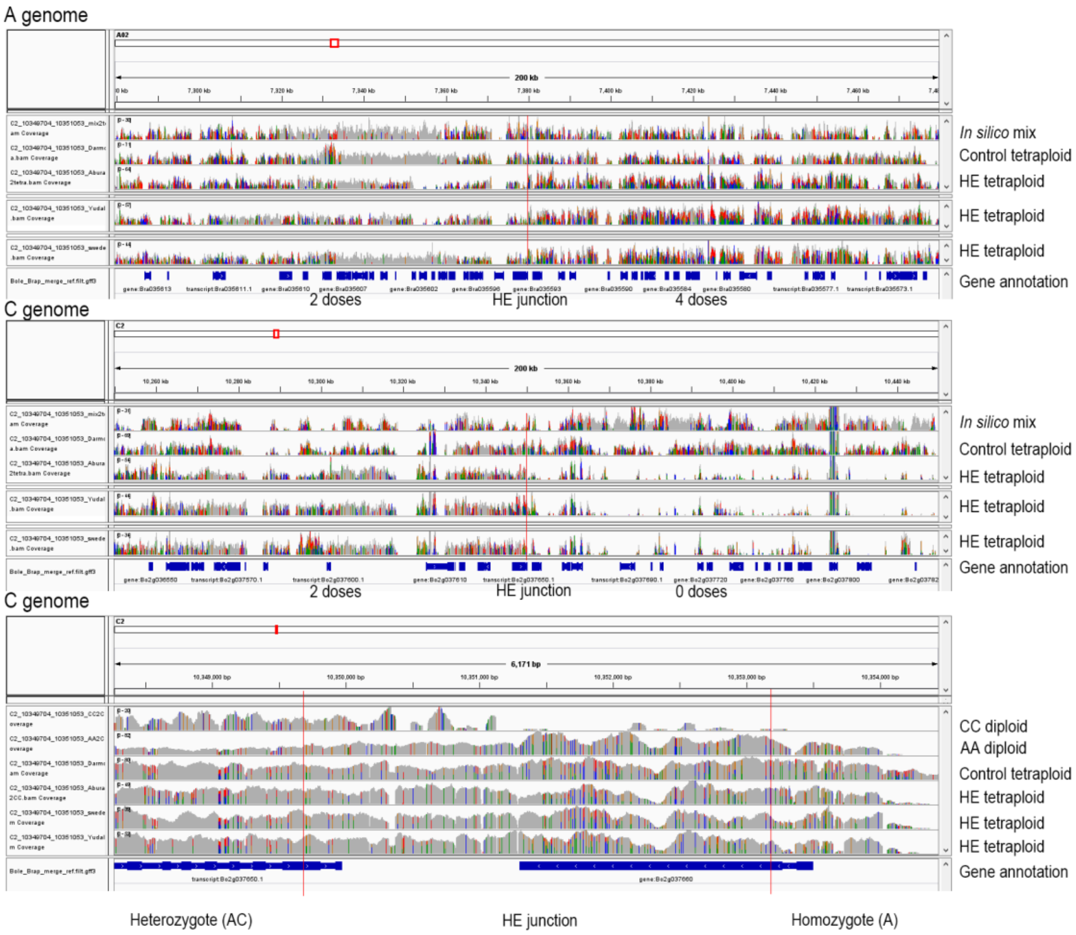


HE junction 24.

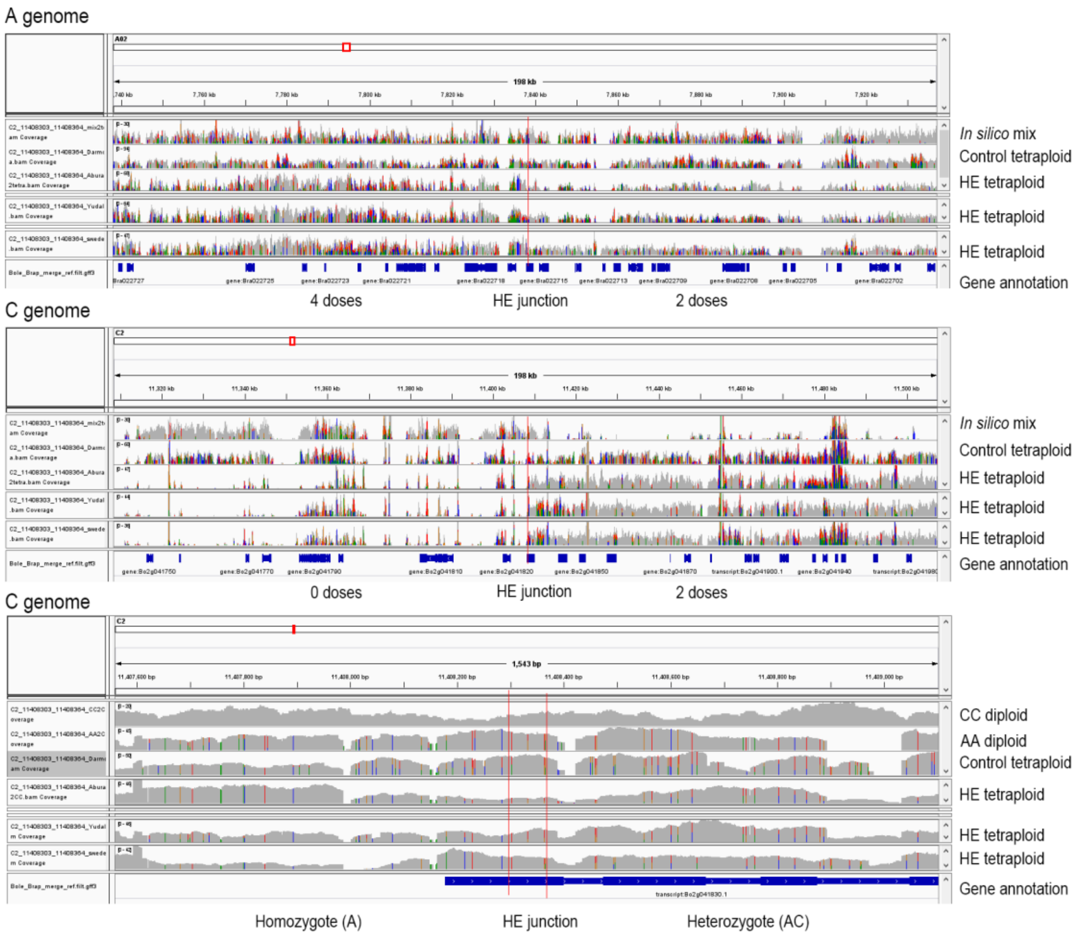
HE junction 25.



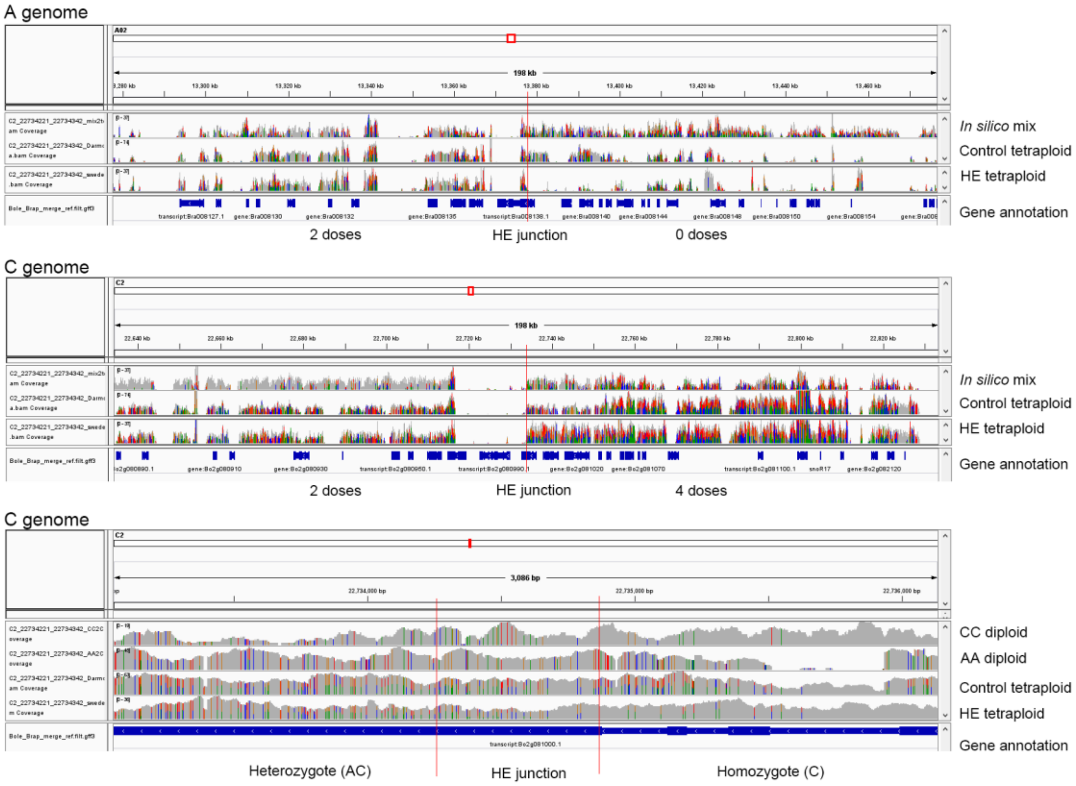
HE junction 26.



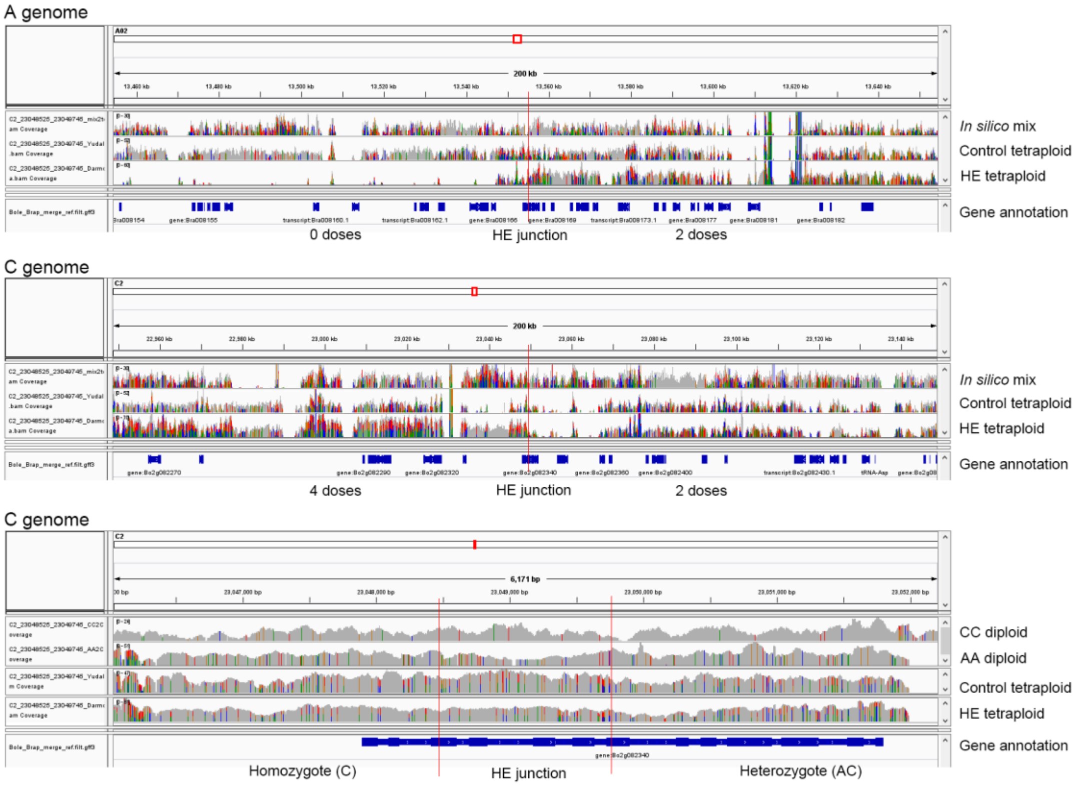
HE junction 27.



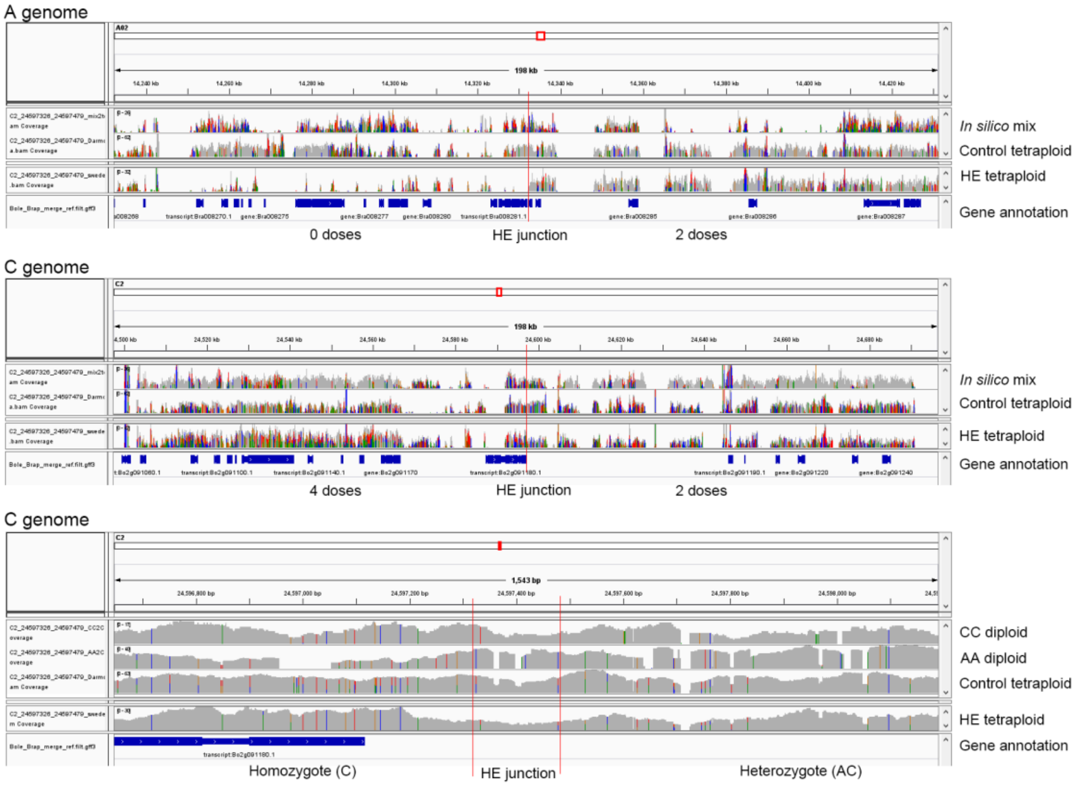
HE junction 28.



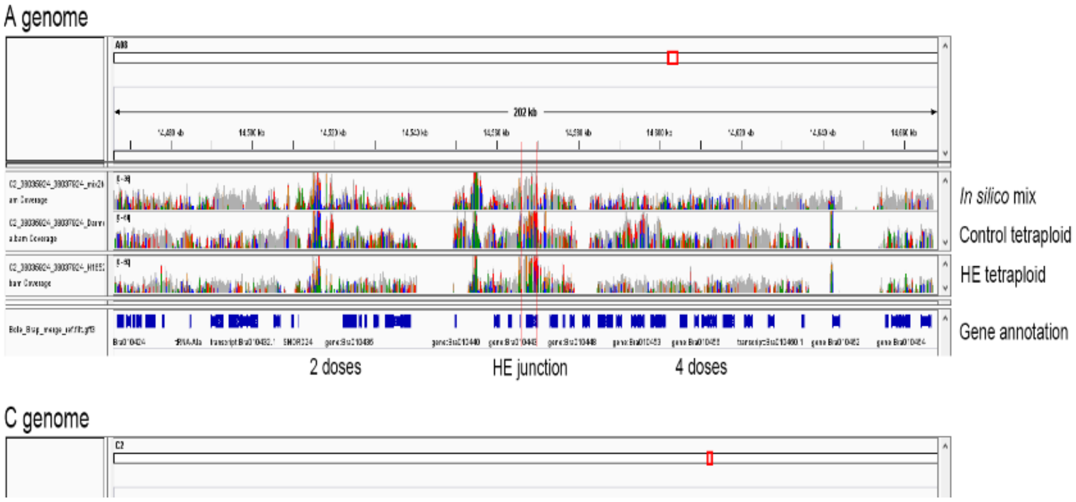
HE junction 29.



HE junction 30.

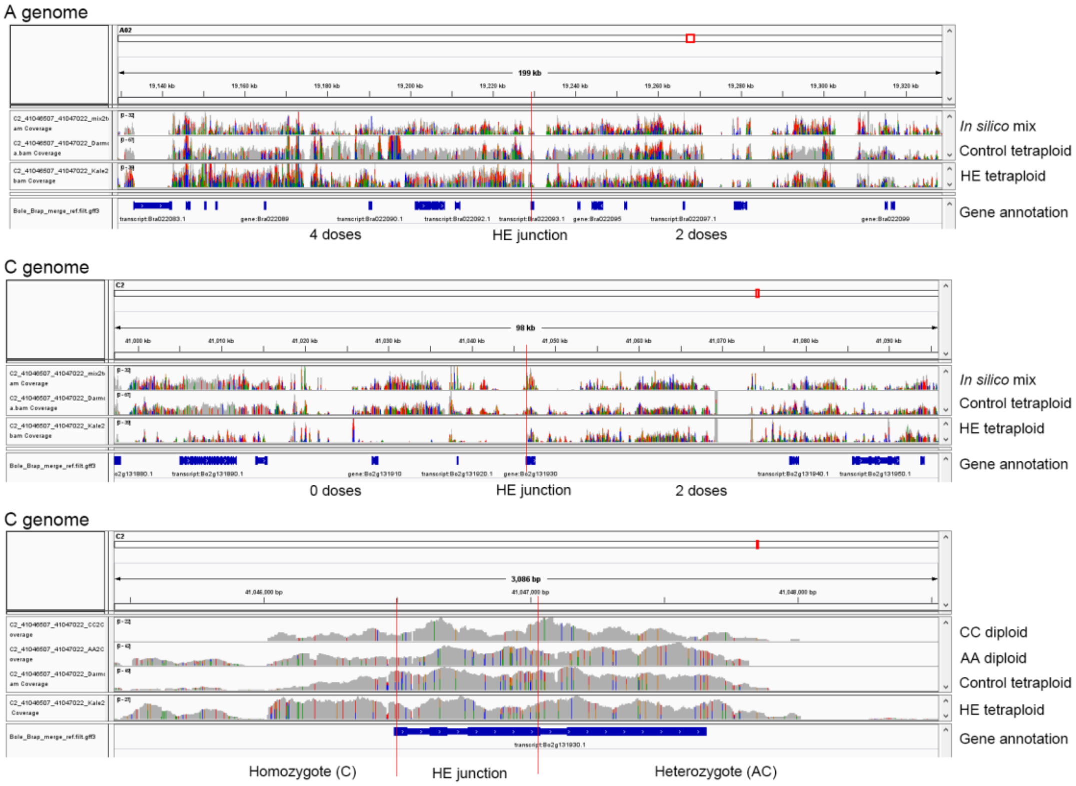


HE junction 31.

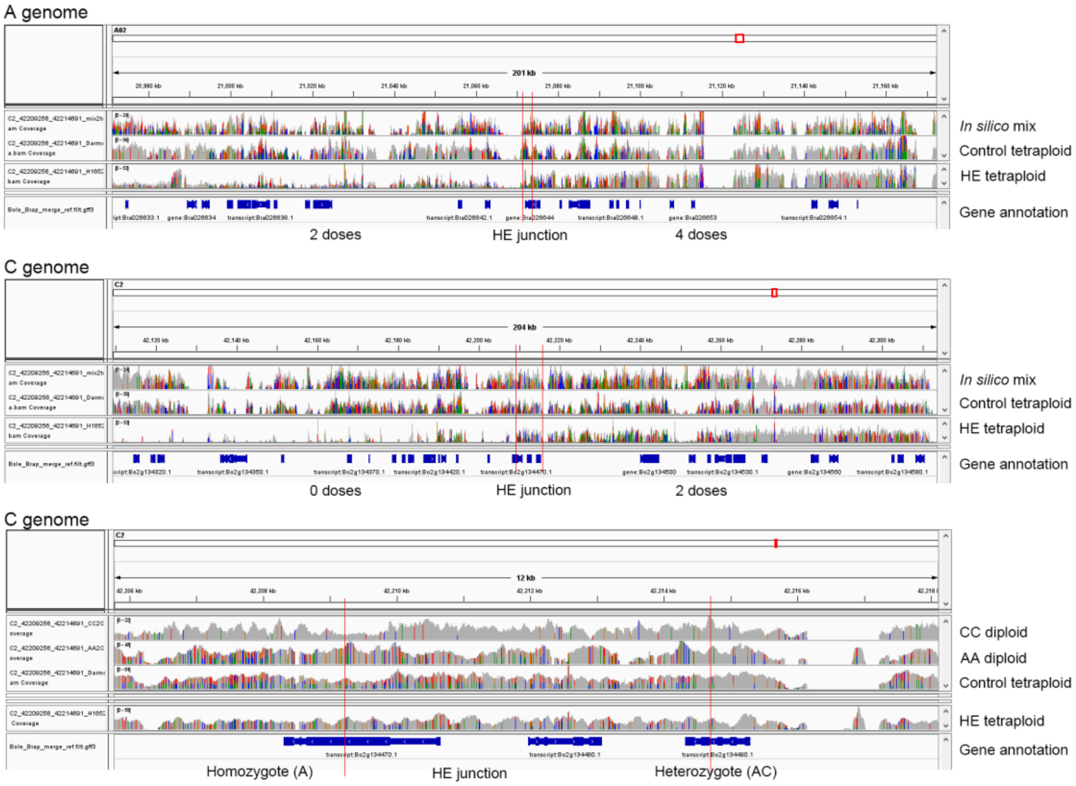


HE junction 32.

HE junction 33.

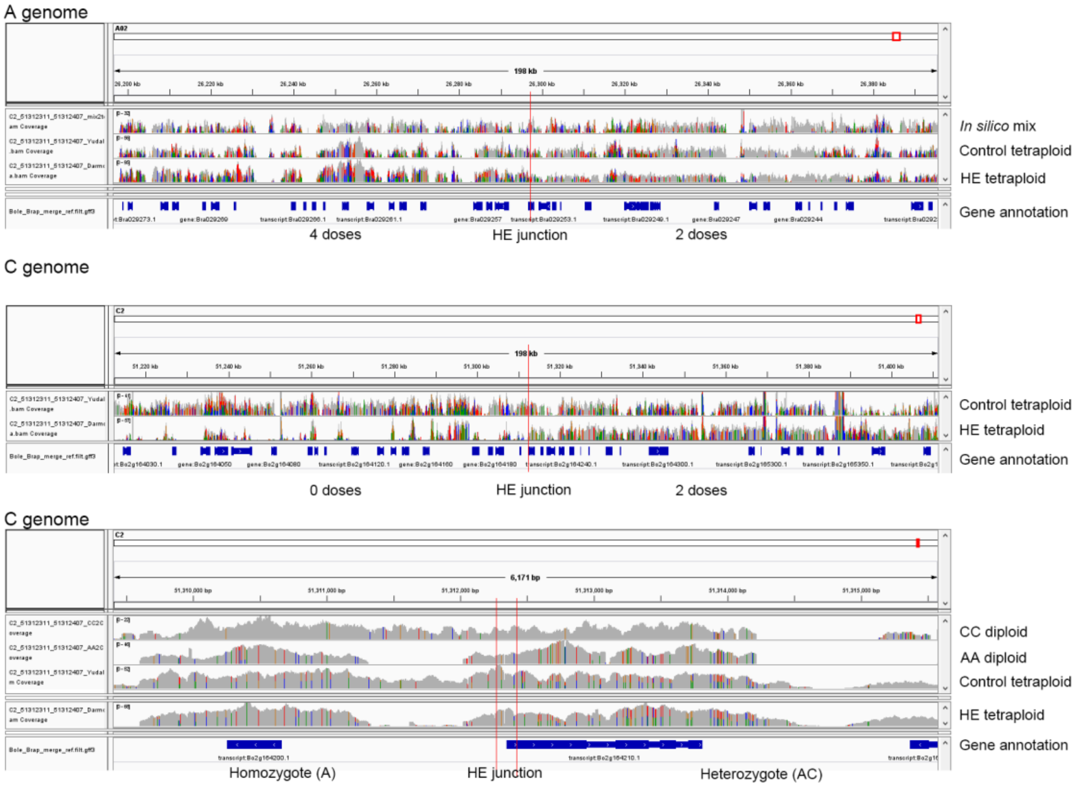


HE junction 34.

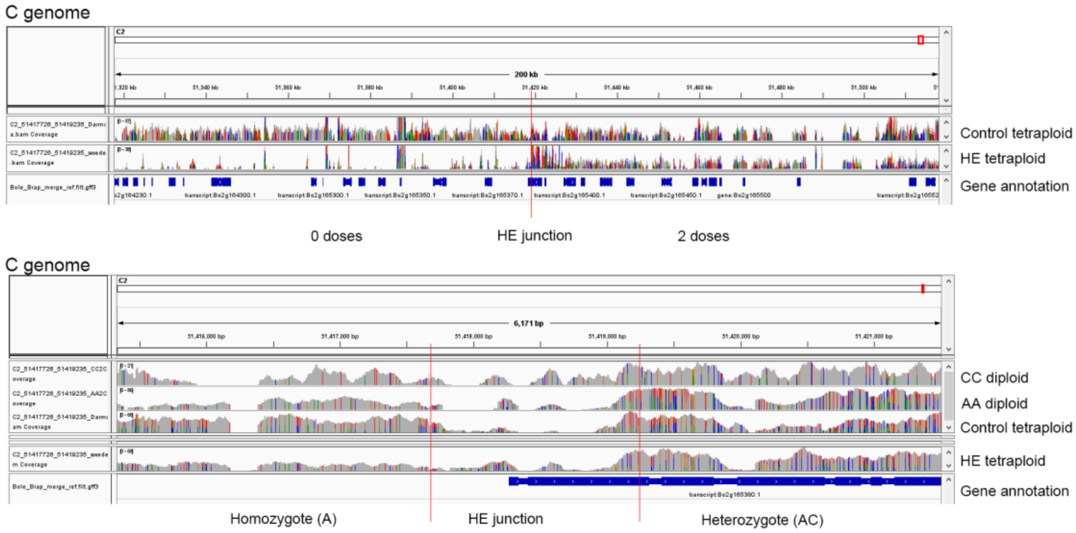


HE junction 35.

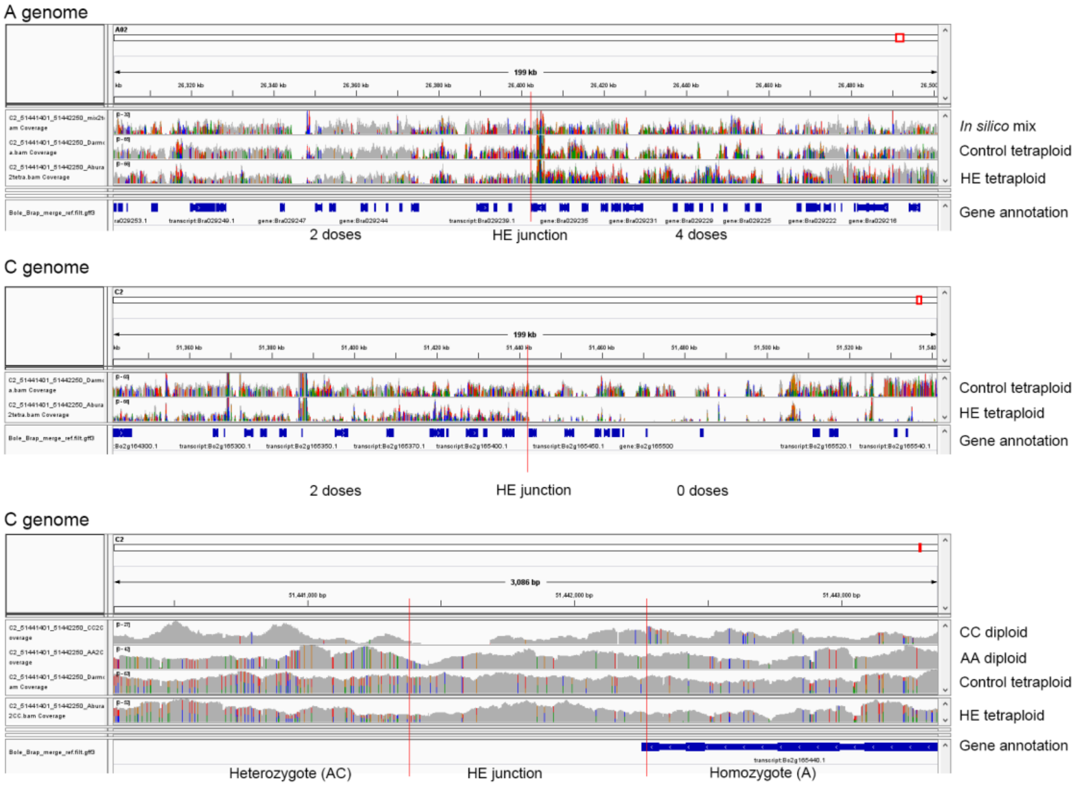
HE junction 36.



HE junction 37.

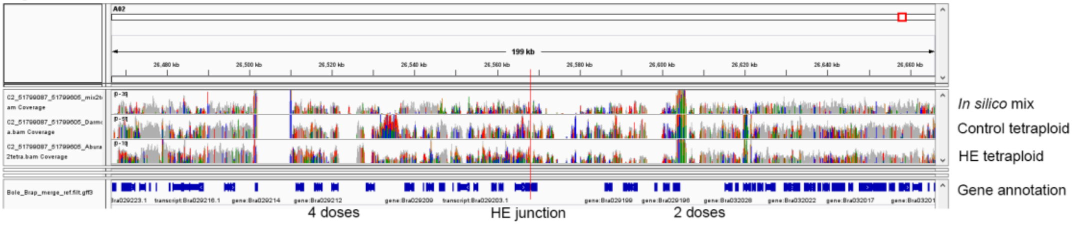


HE junction 38.

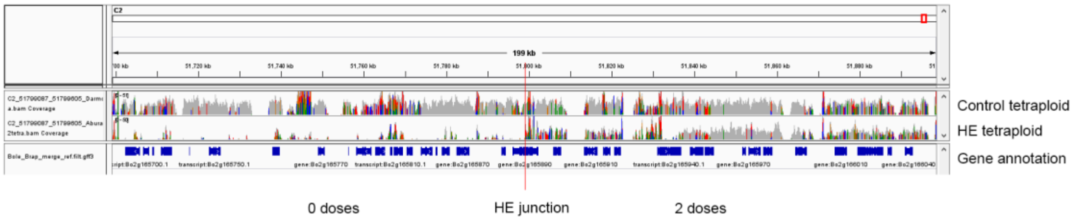


HE junction 39.

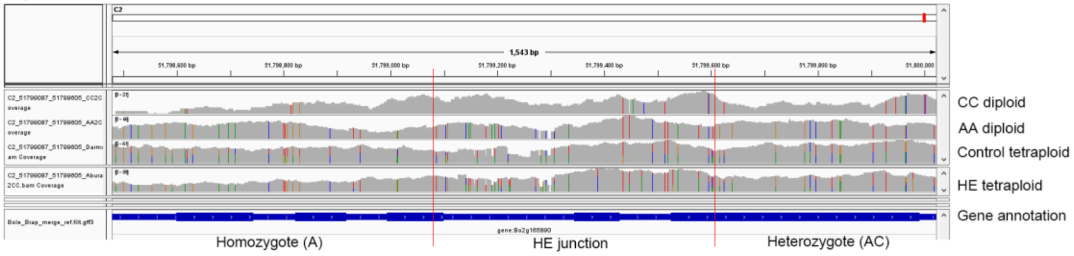
A genome



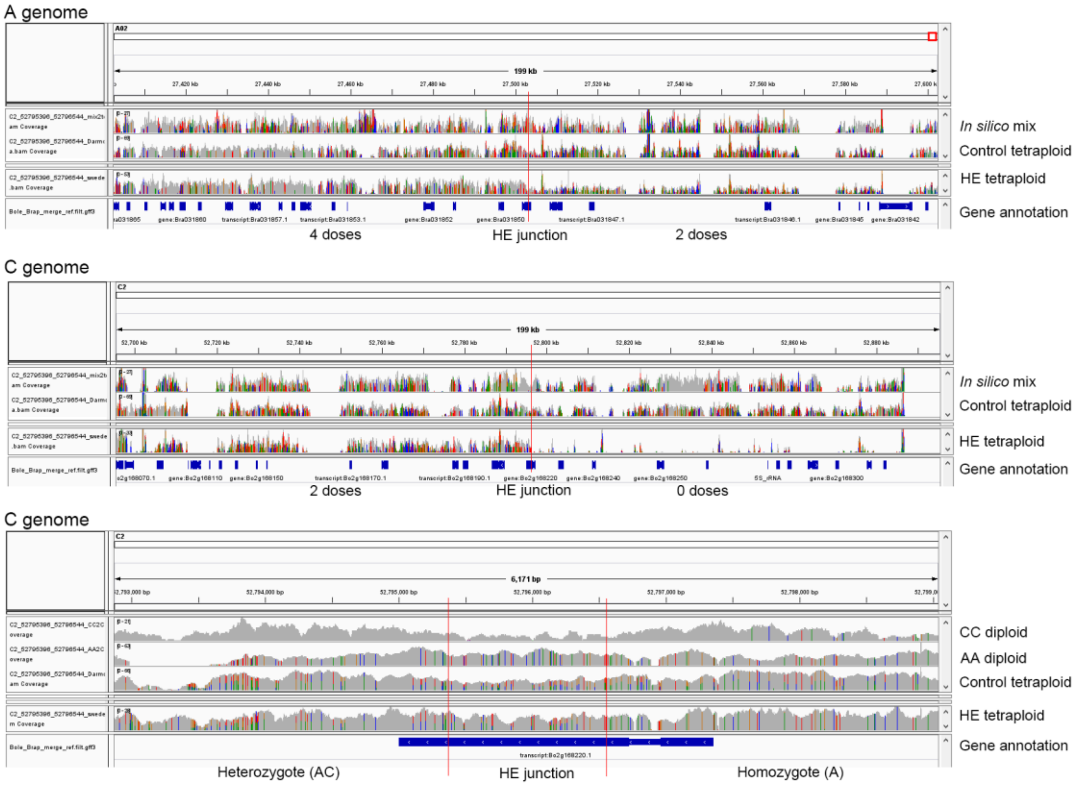
C genome



C genome

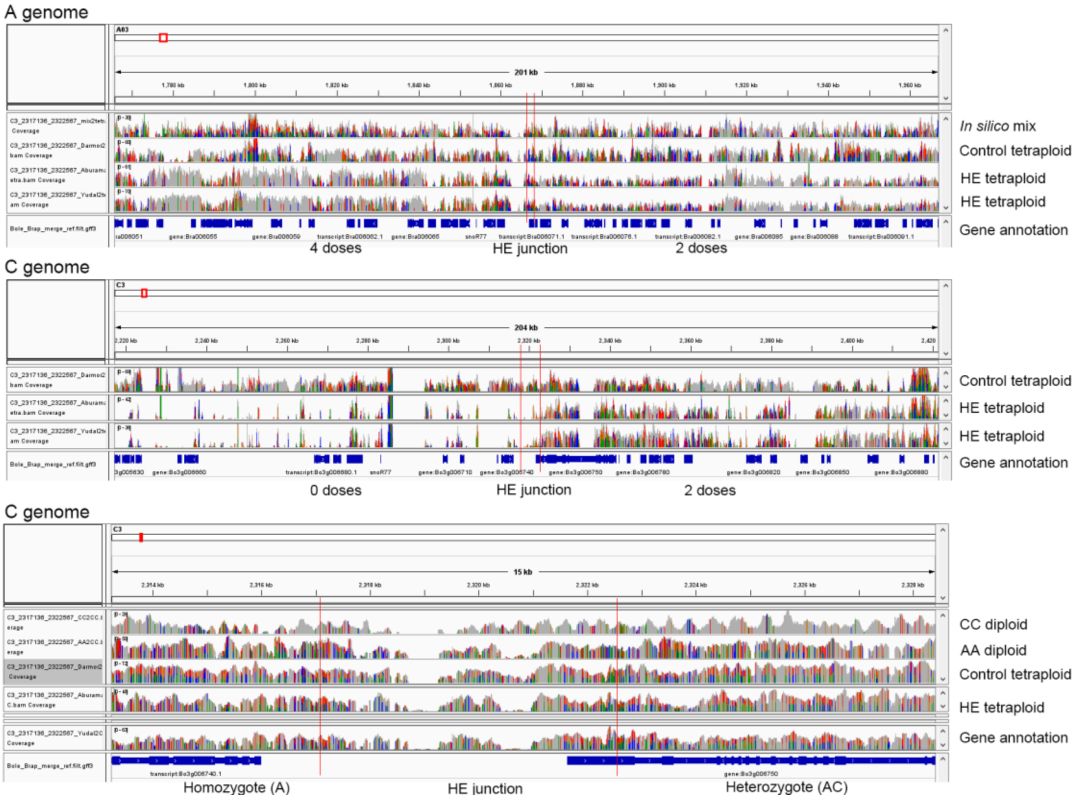


HE junction 40.

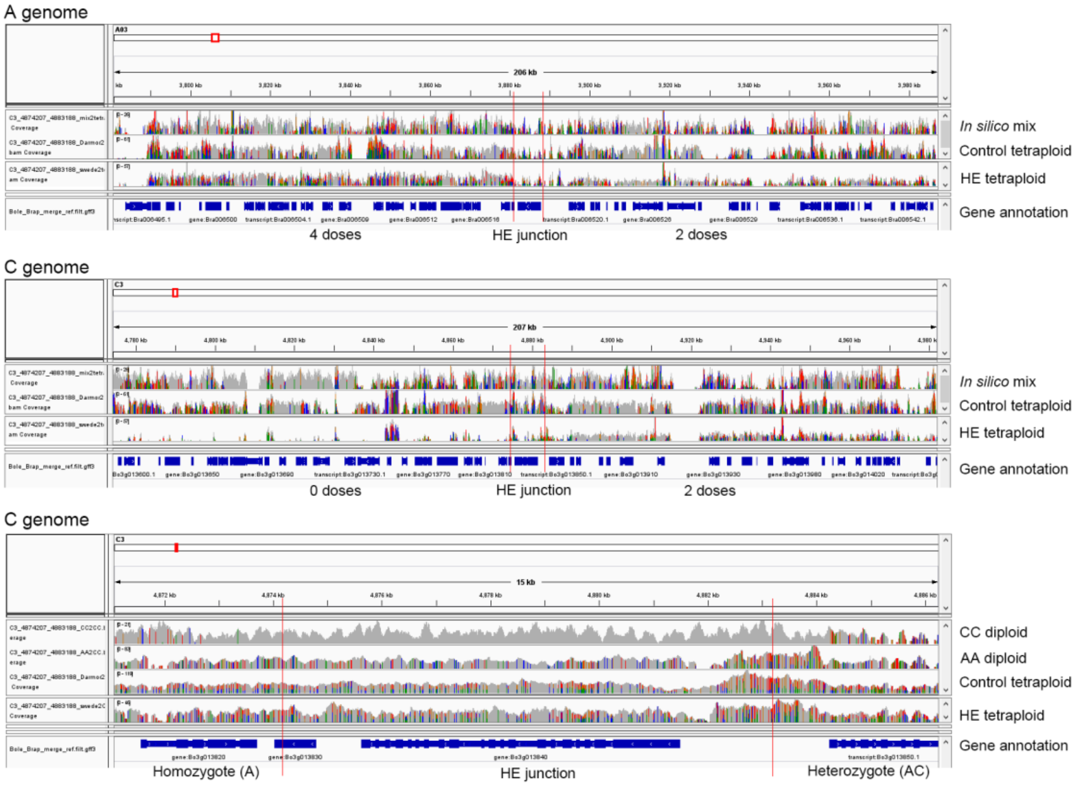


HE junction 41.

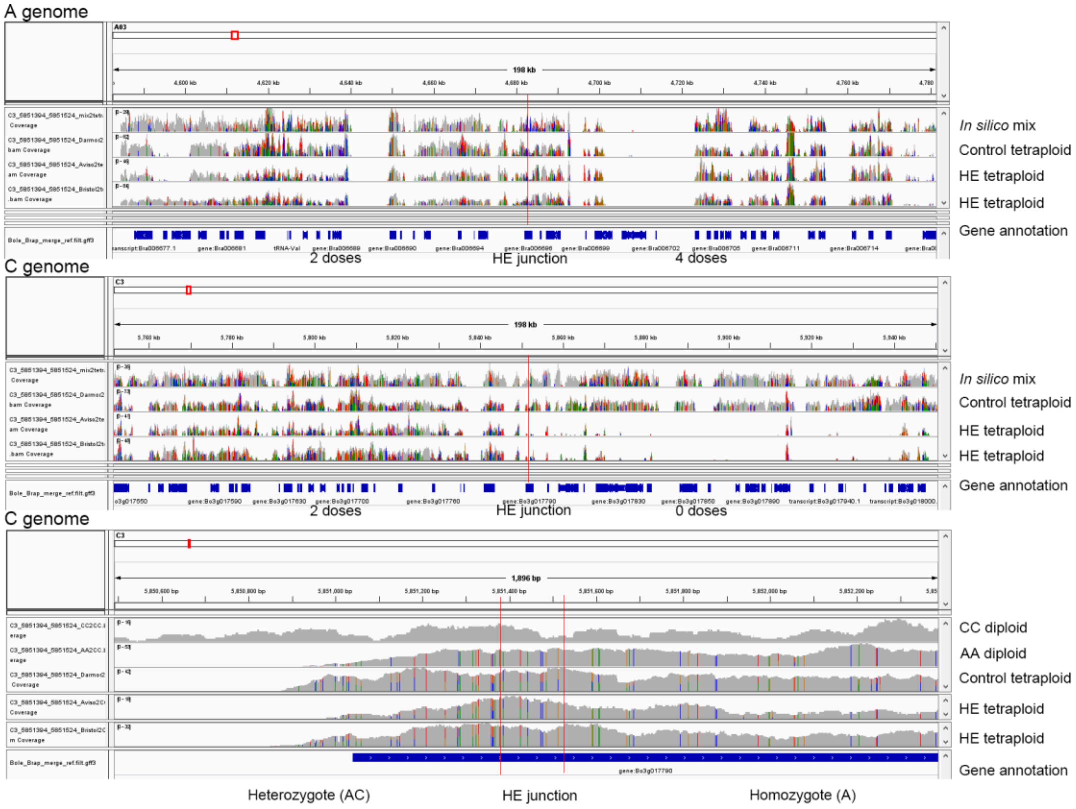
HE junction 42.



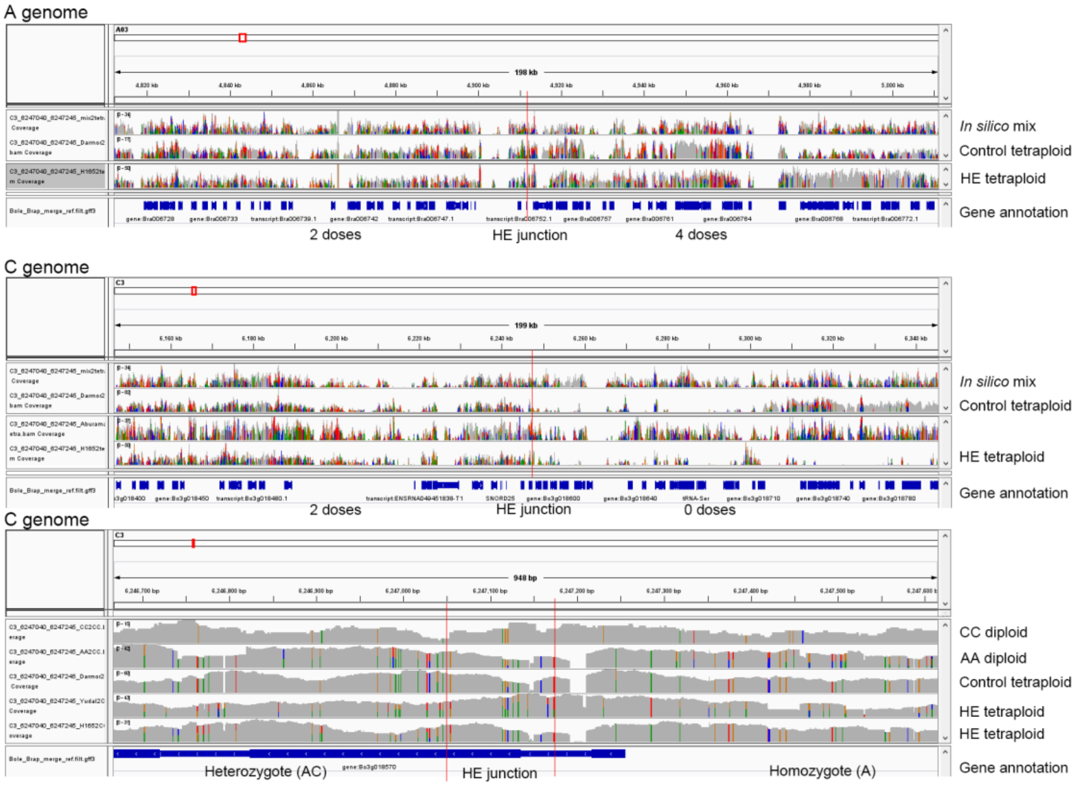
HE junction 43.



HE junction 44.



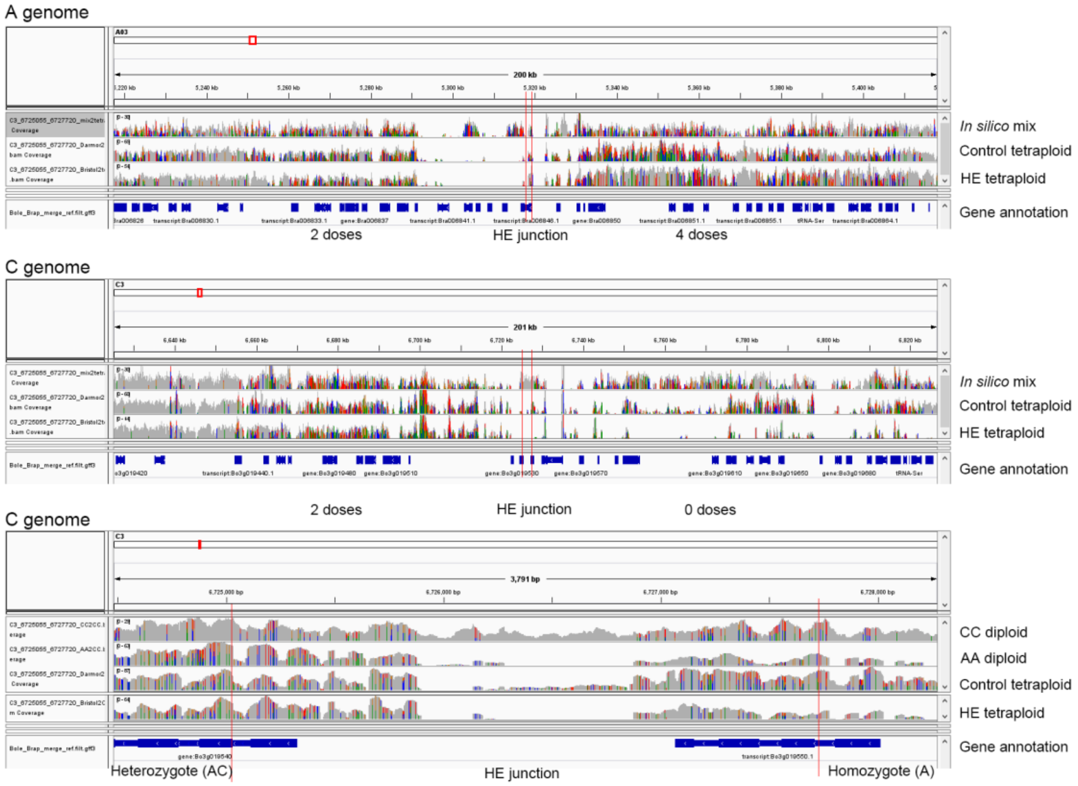
HE junction 45.



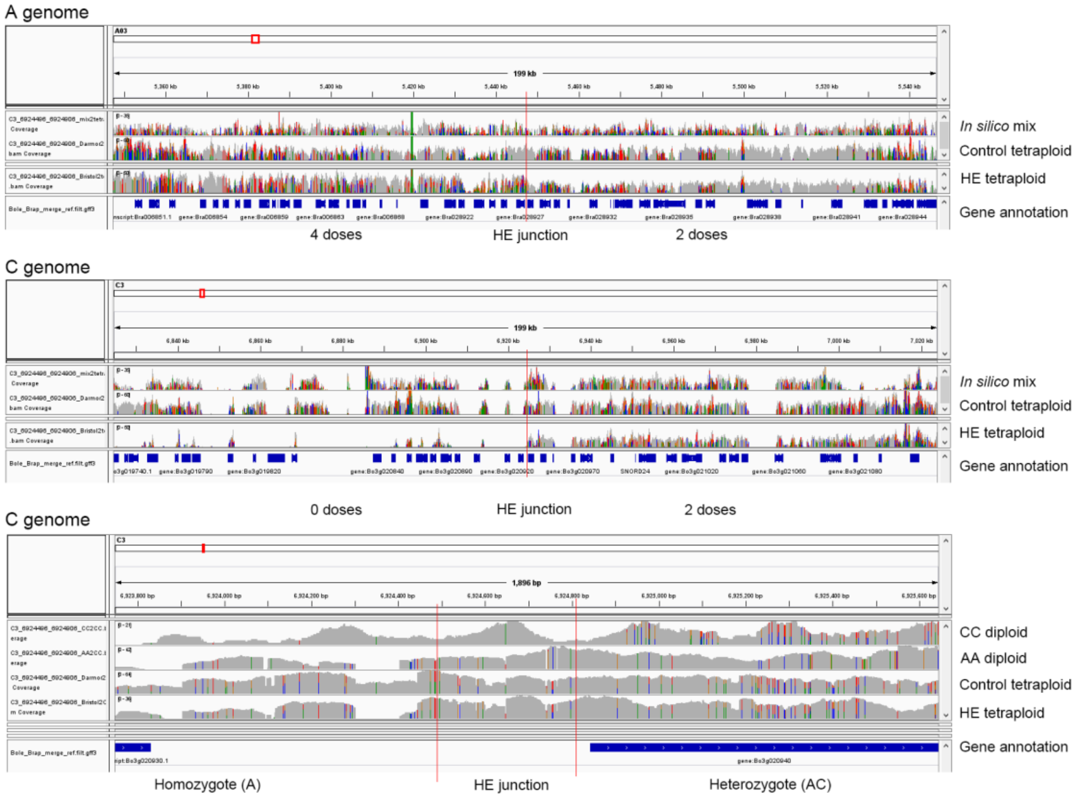
HE junction 46.

HE junction 47.

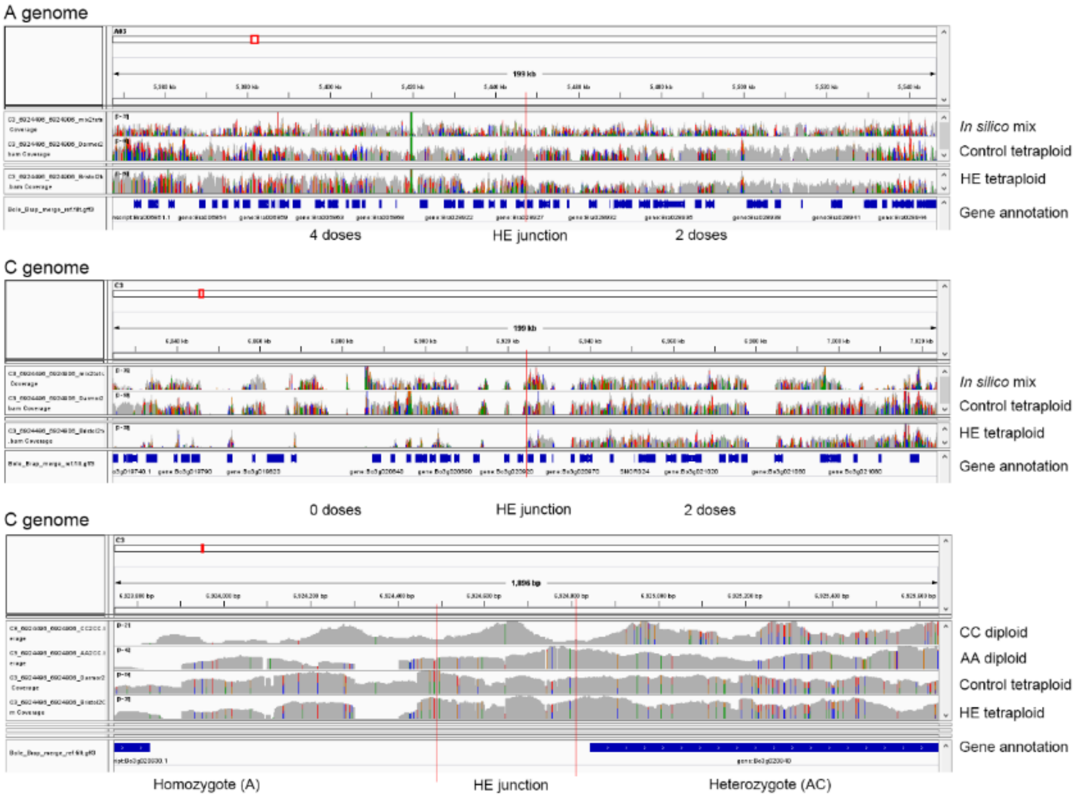
HE junction 48.



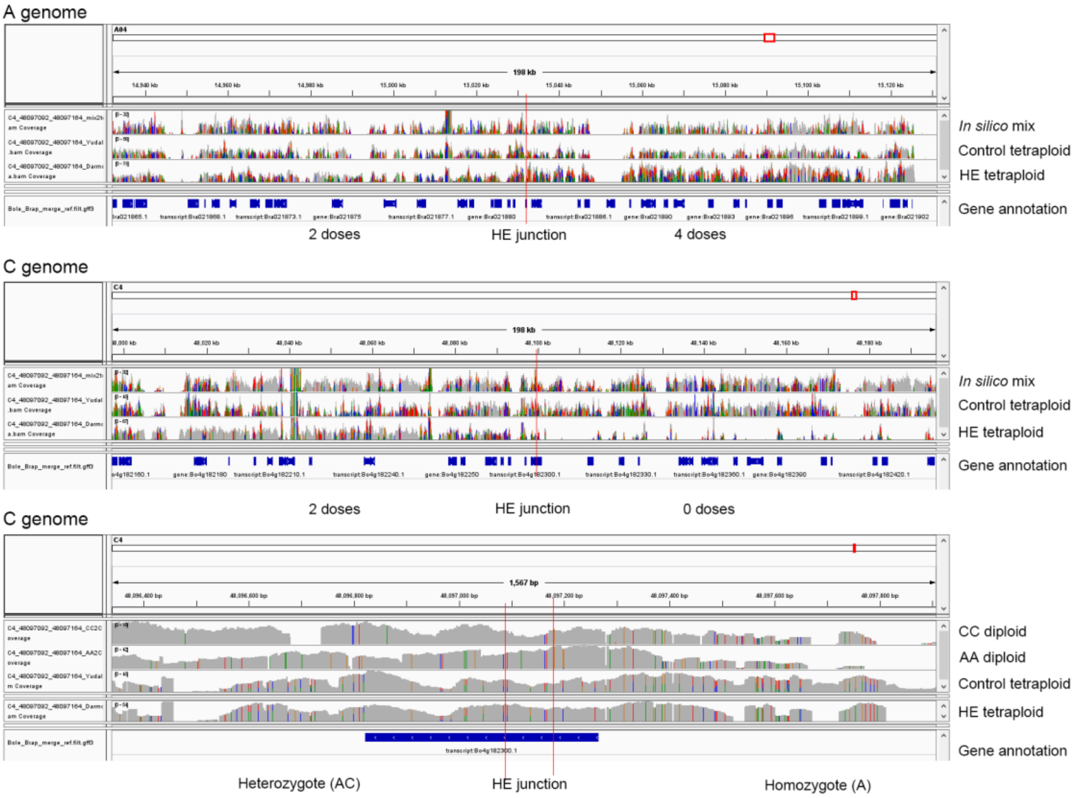
HE junction 49.



HE junction 50.

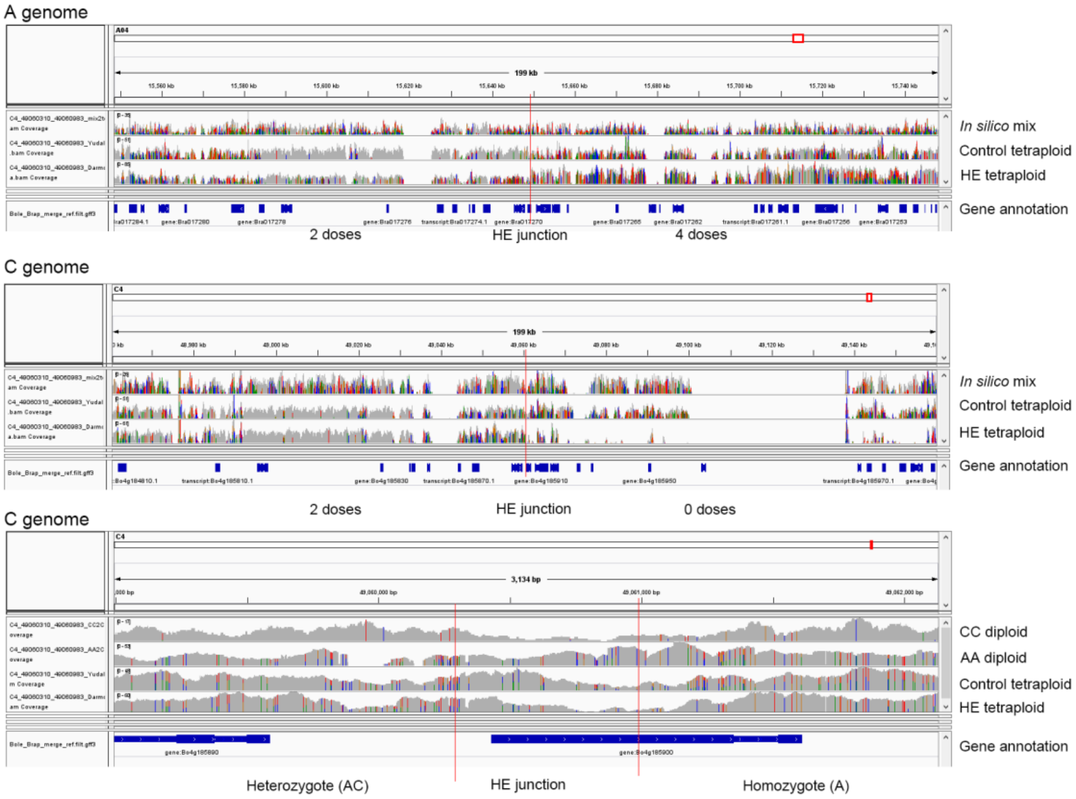


HE junction 51.

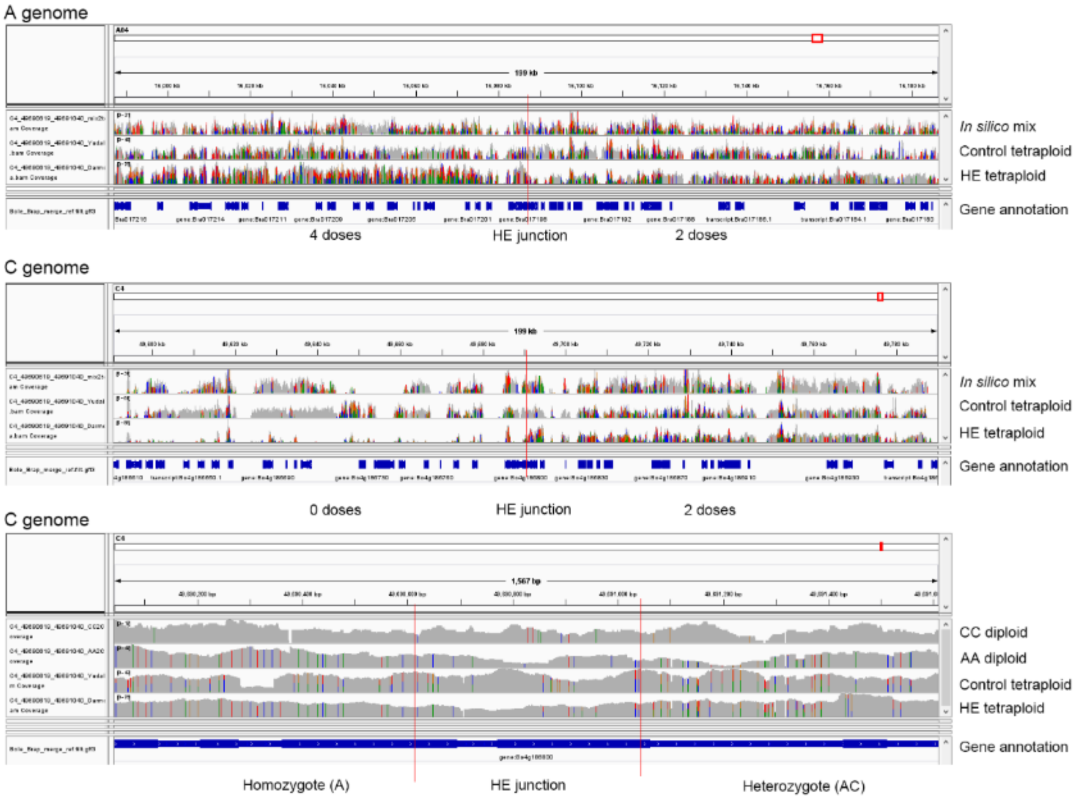


HE junction 52.

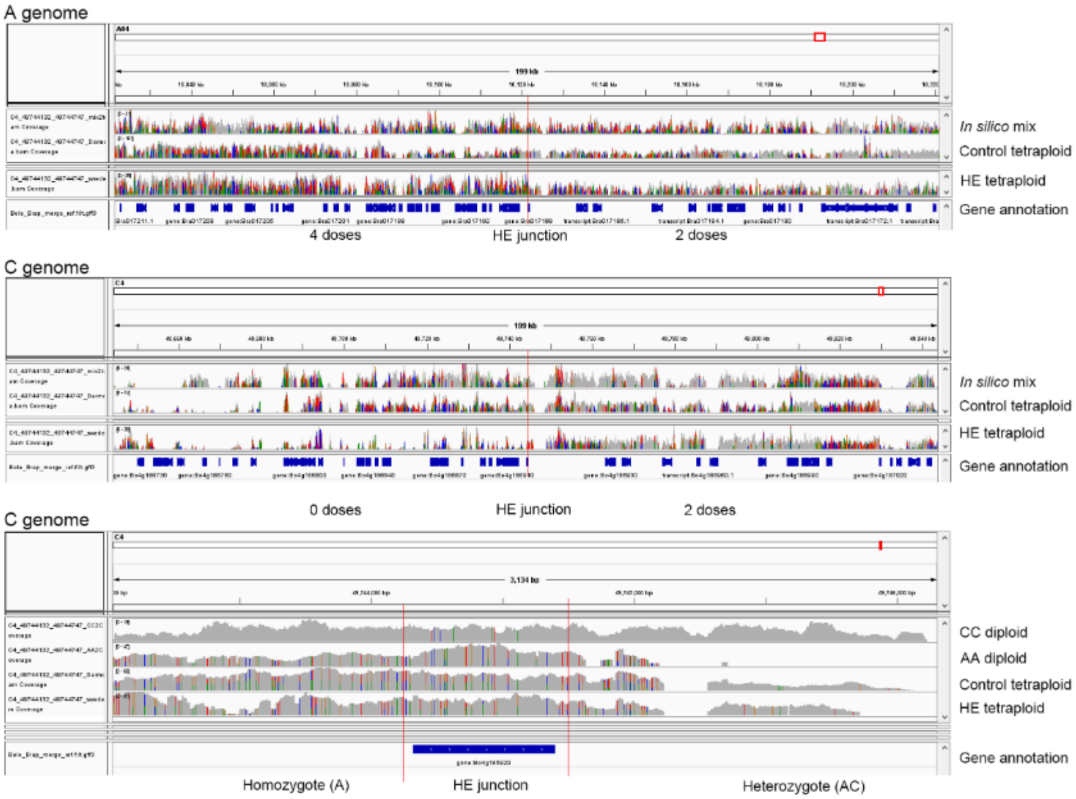
HE junction 53.



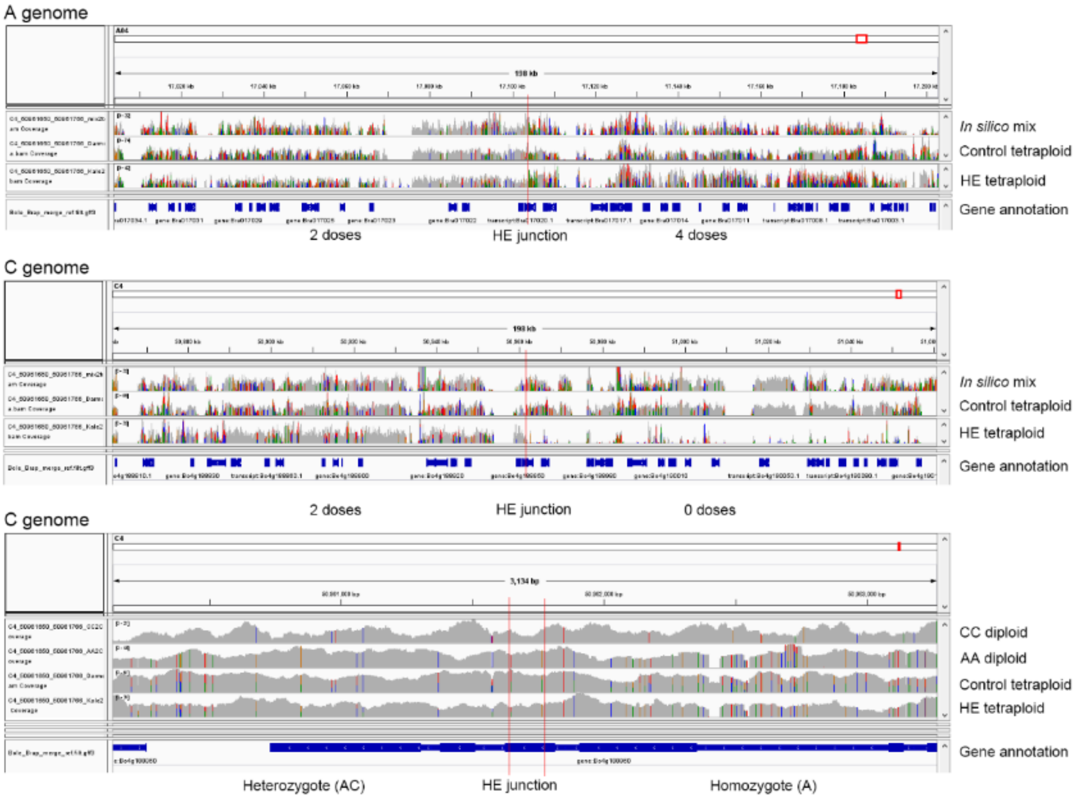
HE junction 54.



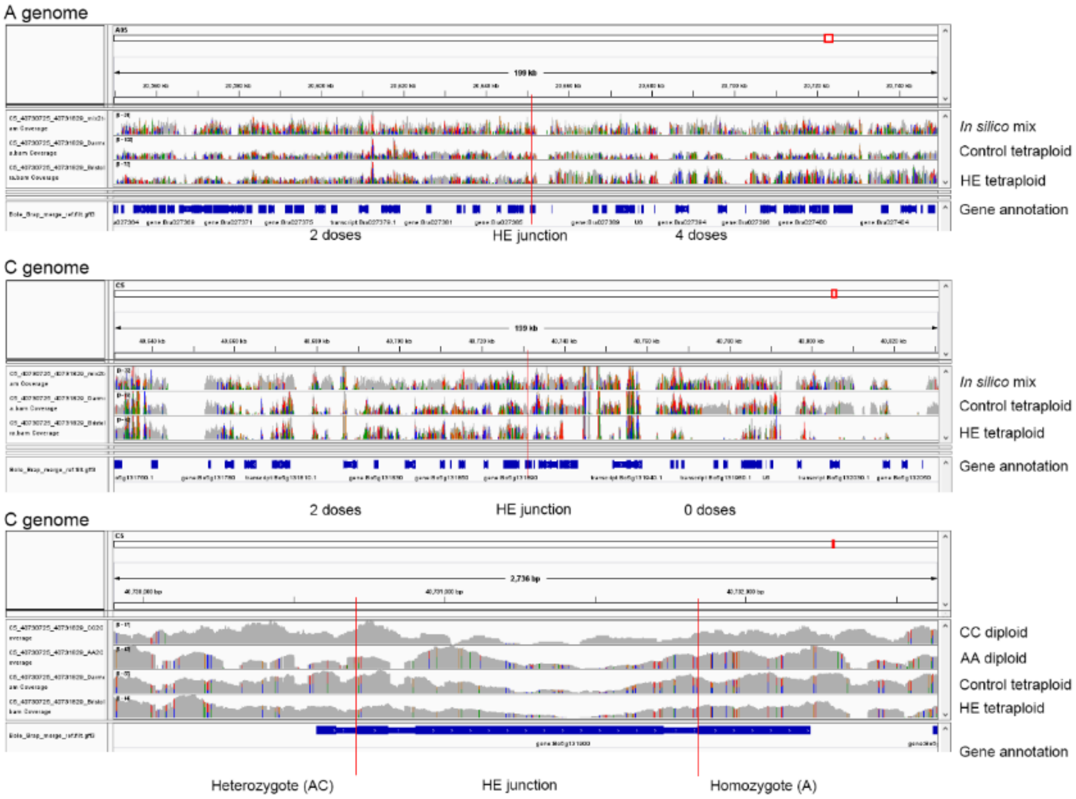
HE junction 55.



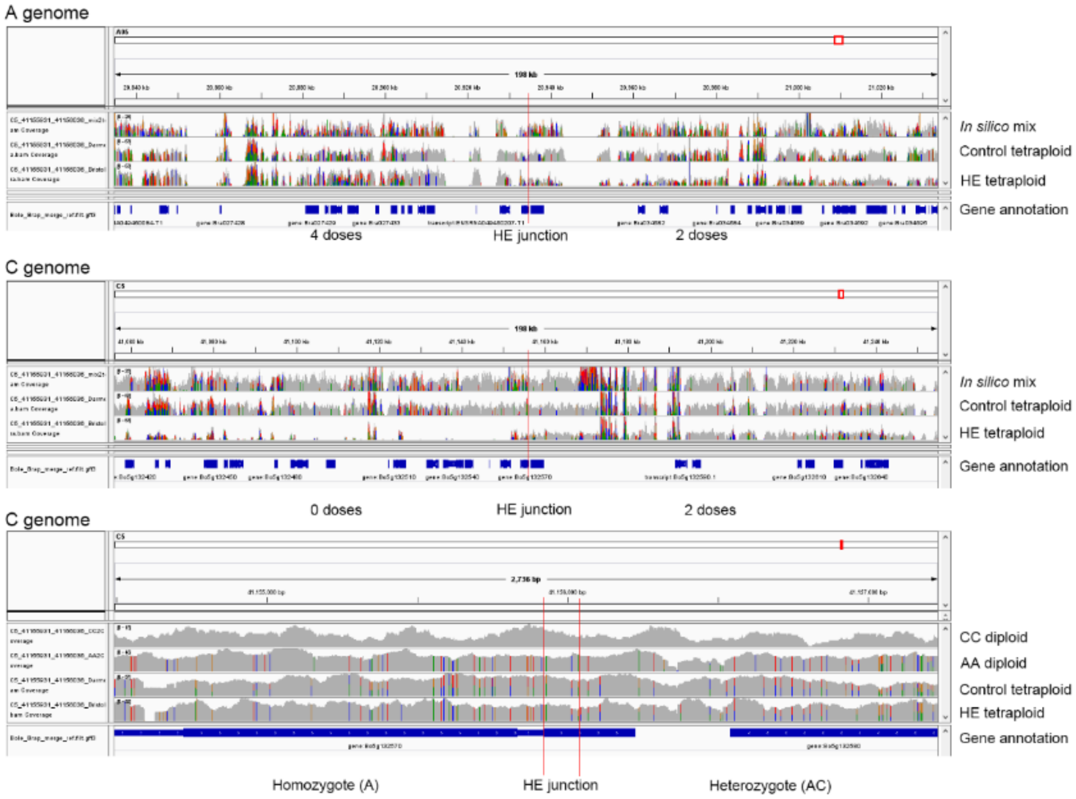
HE junction 56.



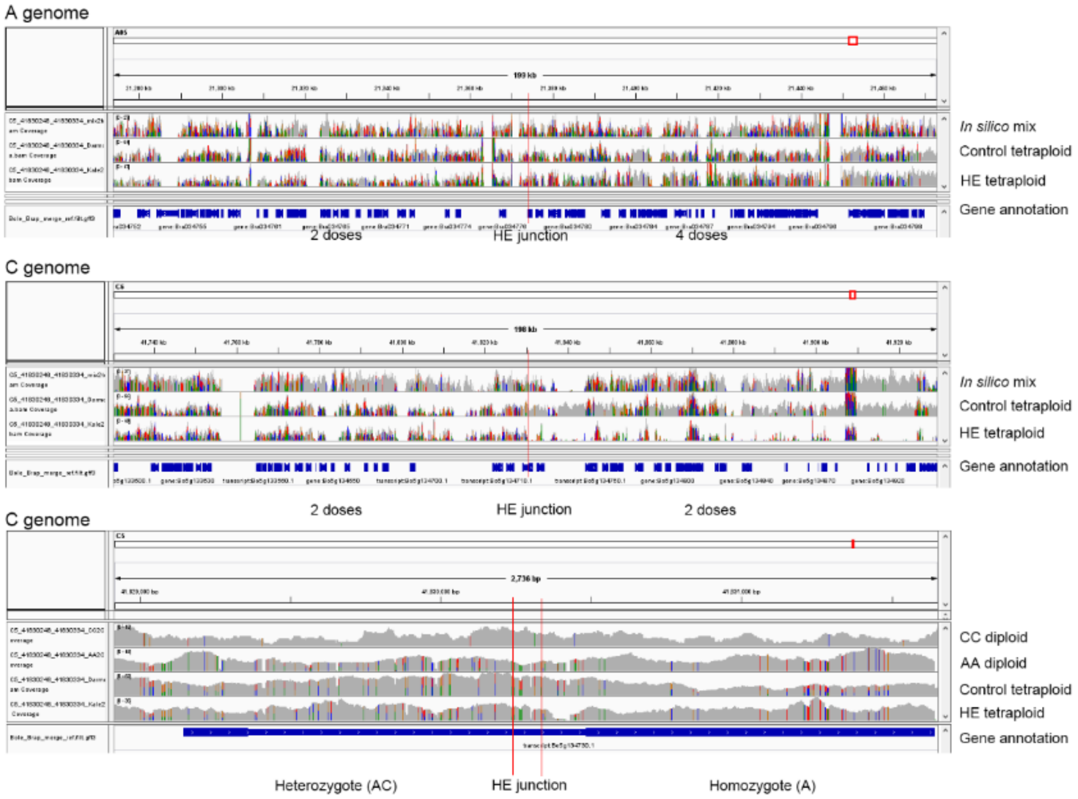
HE junction 57.



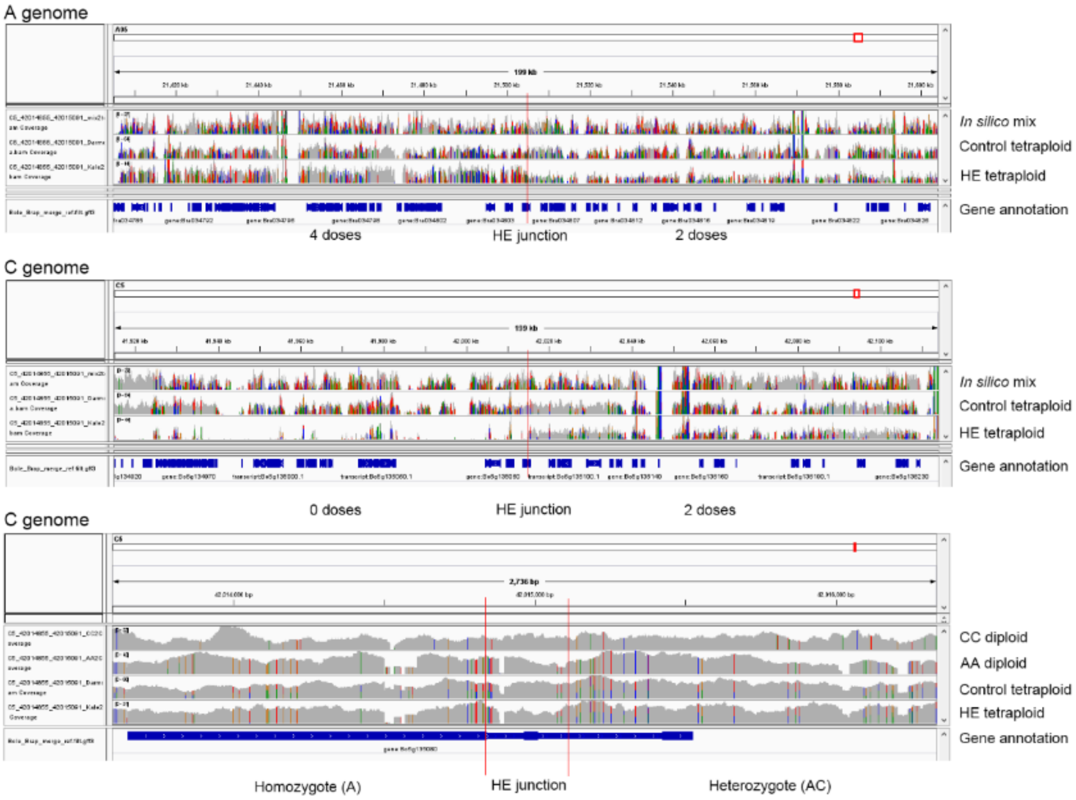
HE junction 58.



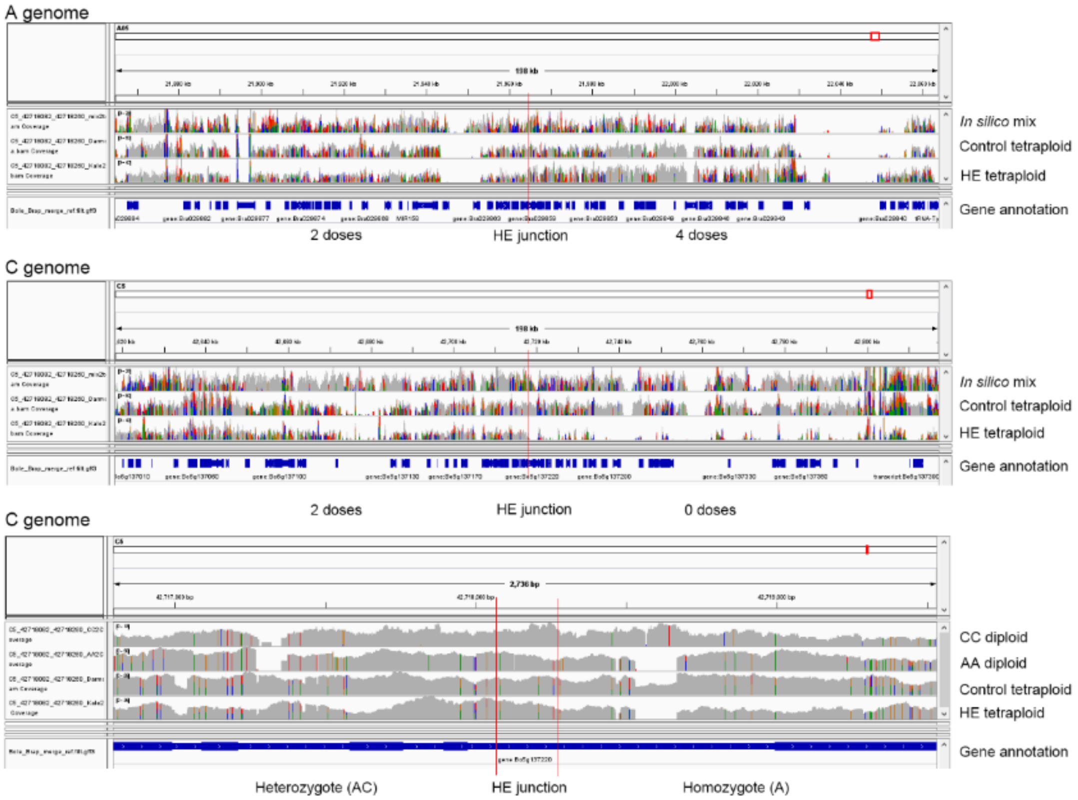
HE junction 59.



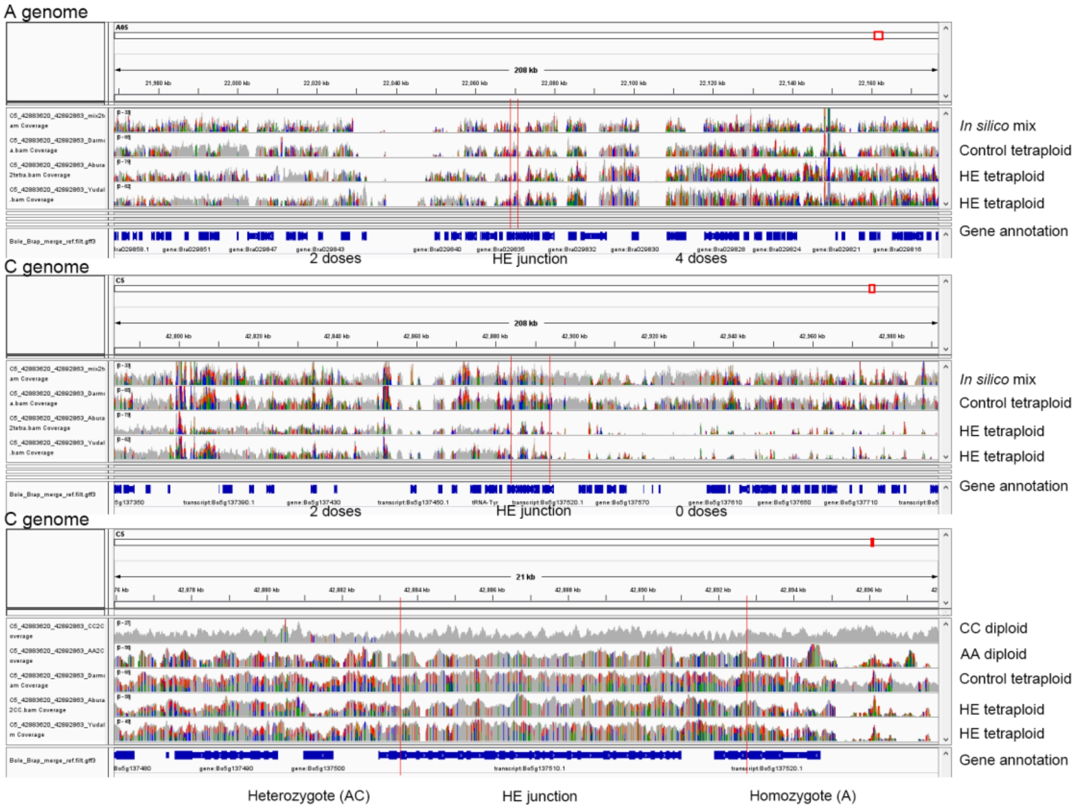
HE junction 60.



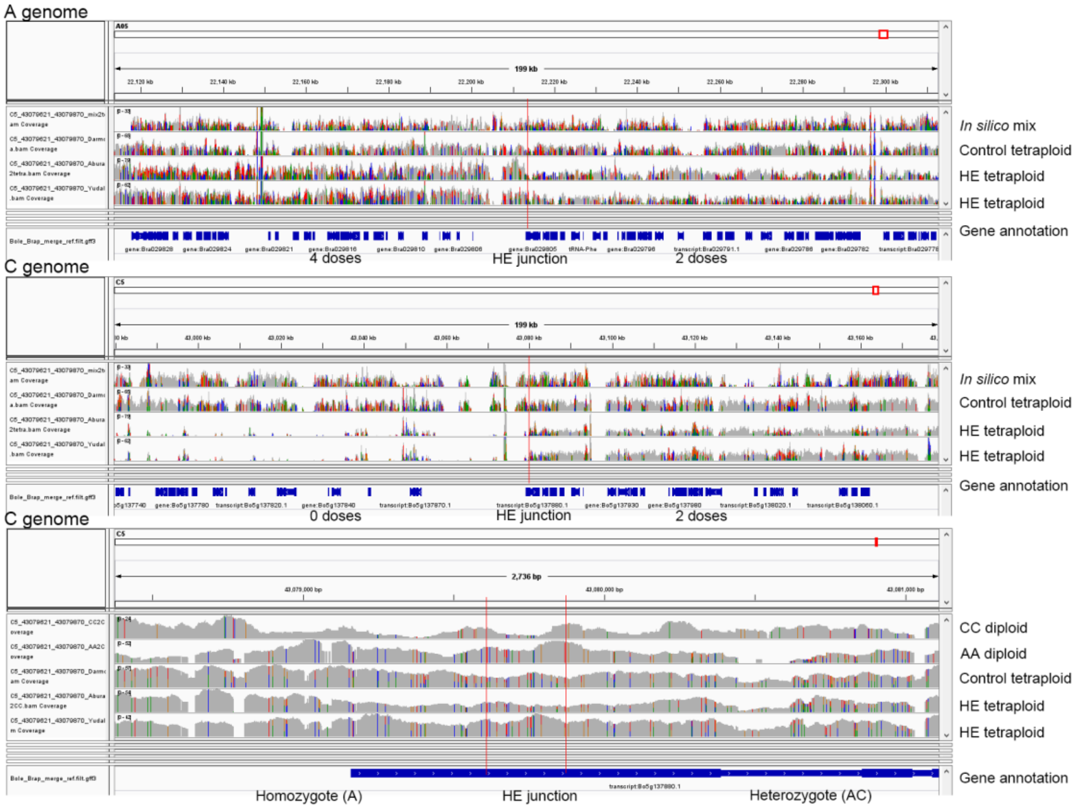
HE junction 61.



HE junction 62.

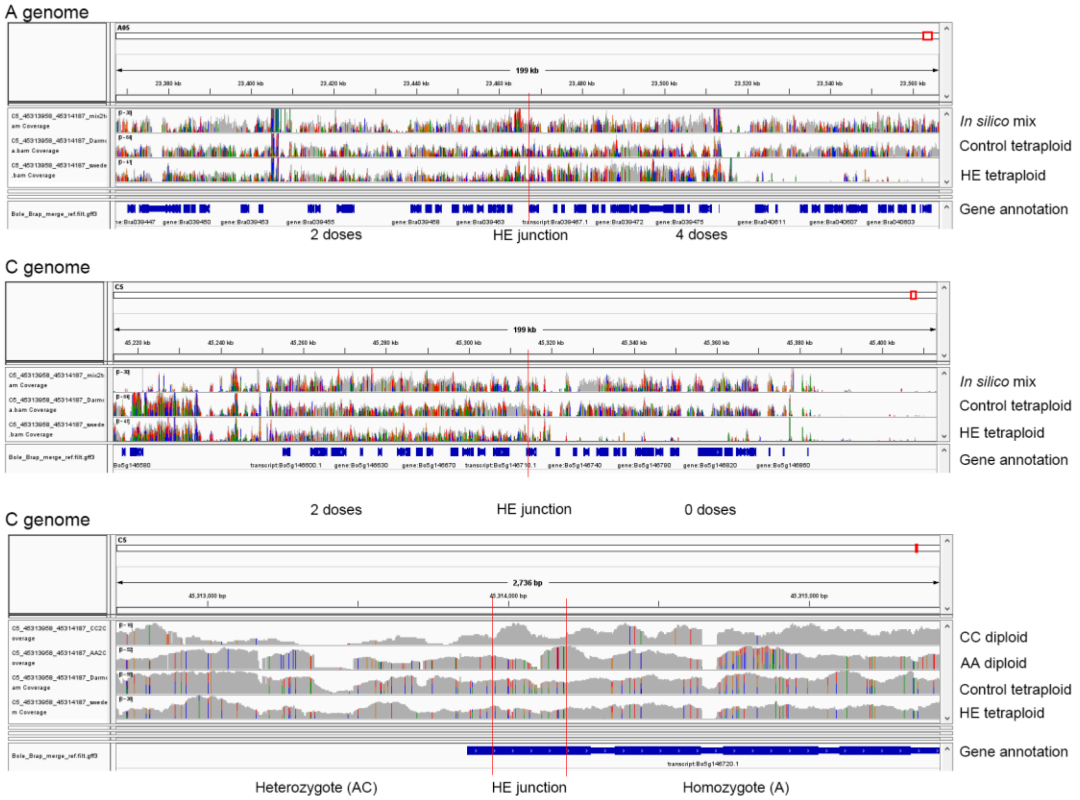


HE junction 63.



HE junction 64.

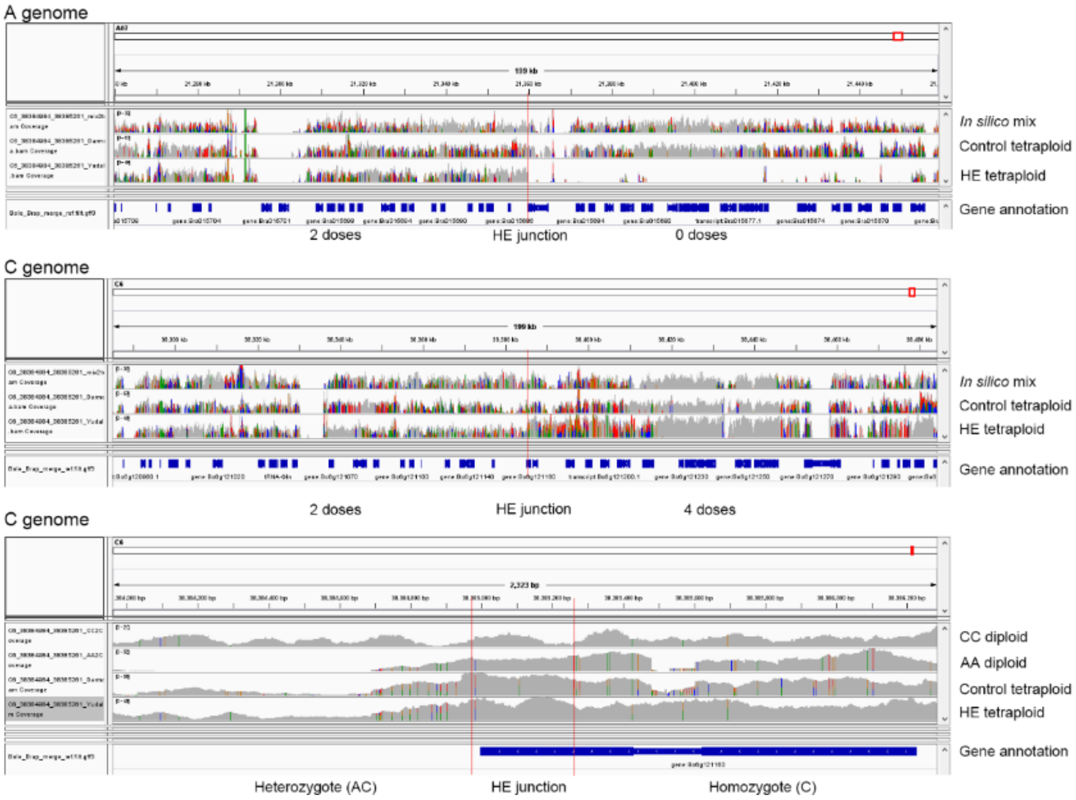
HE junction 65.



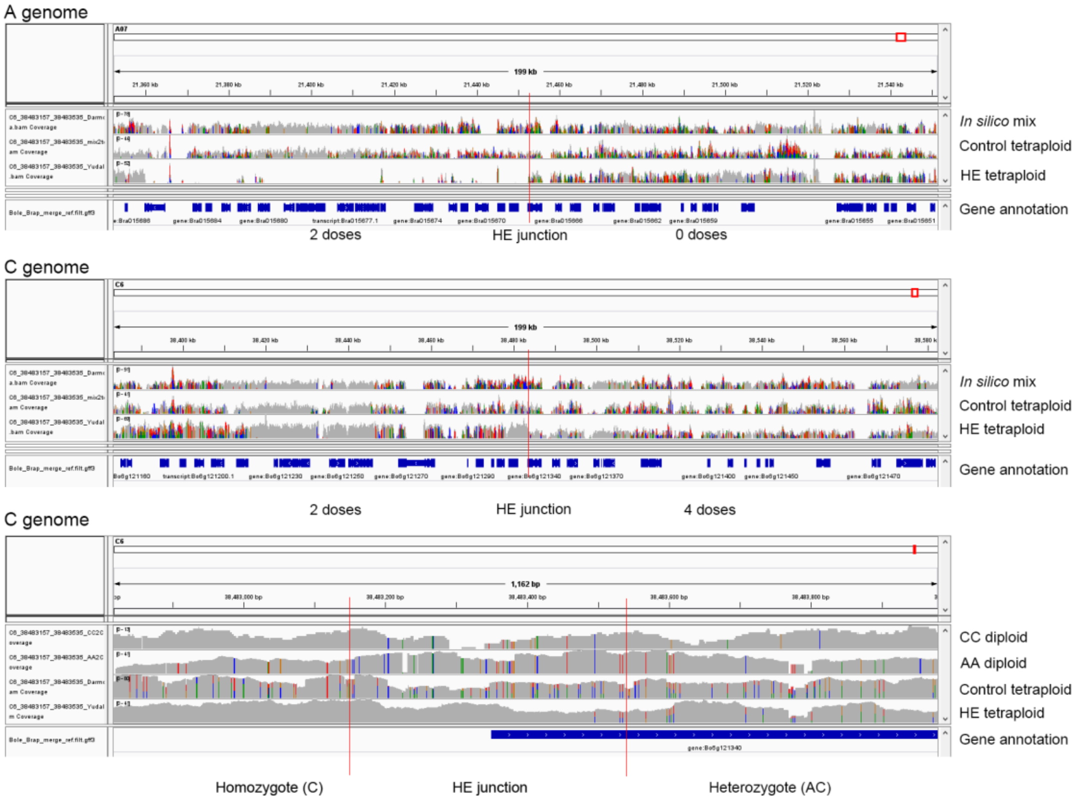
HE junction 66.

HE junction 67.

HE junction 68.



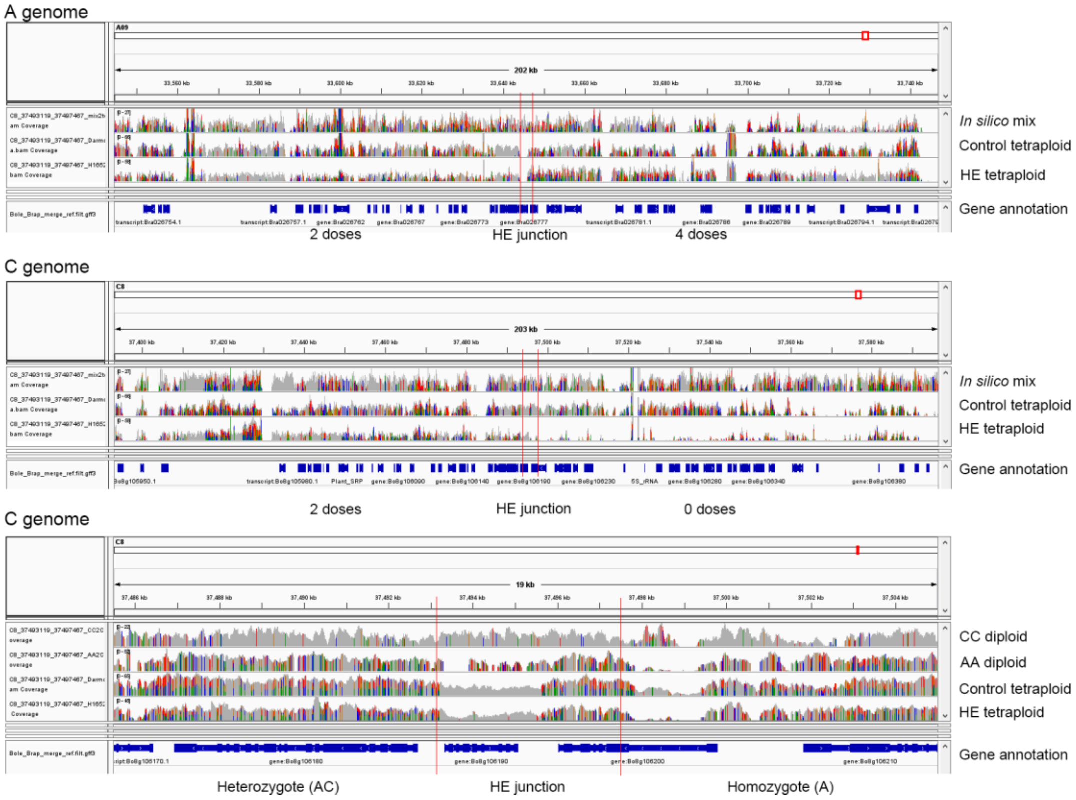
HE junction 69.



HE junction 70.

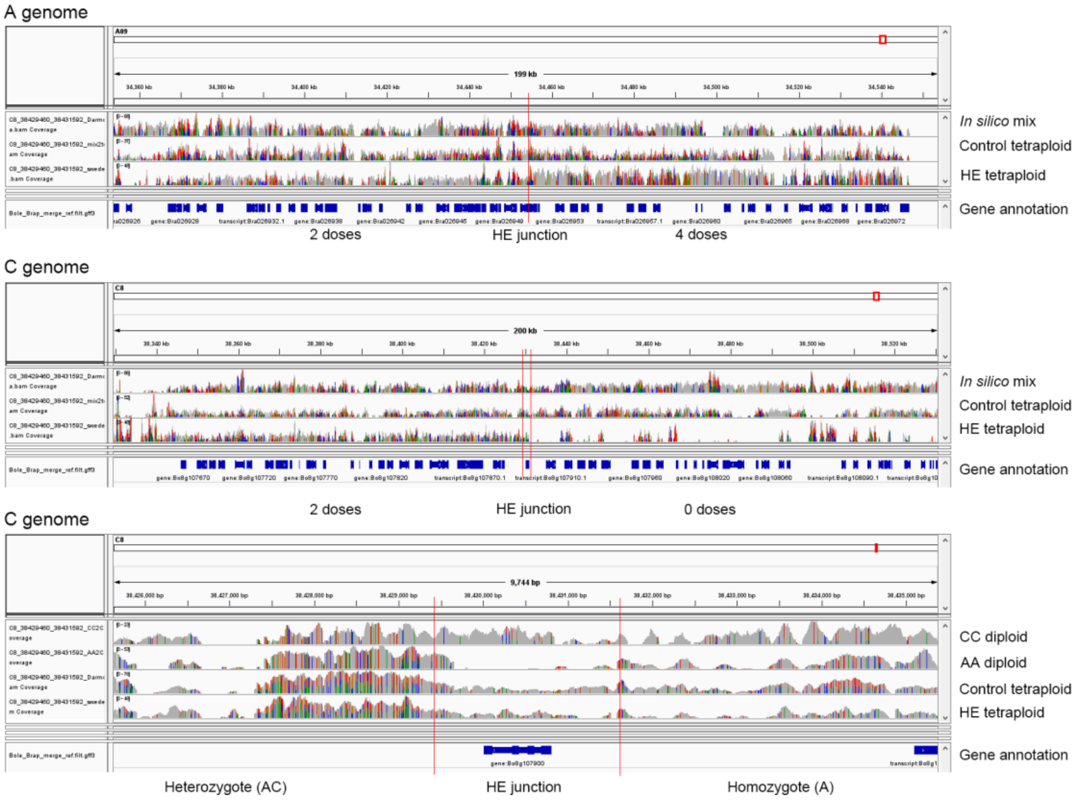
HE junction 71.

HE junction 72.

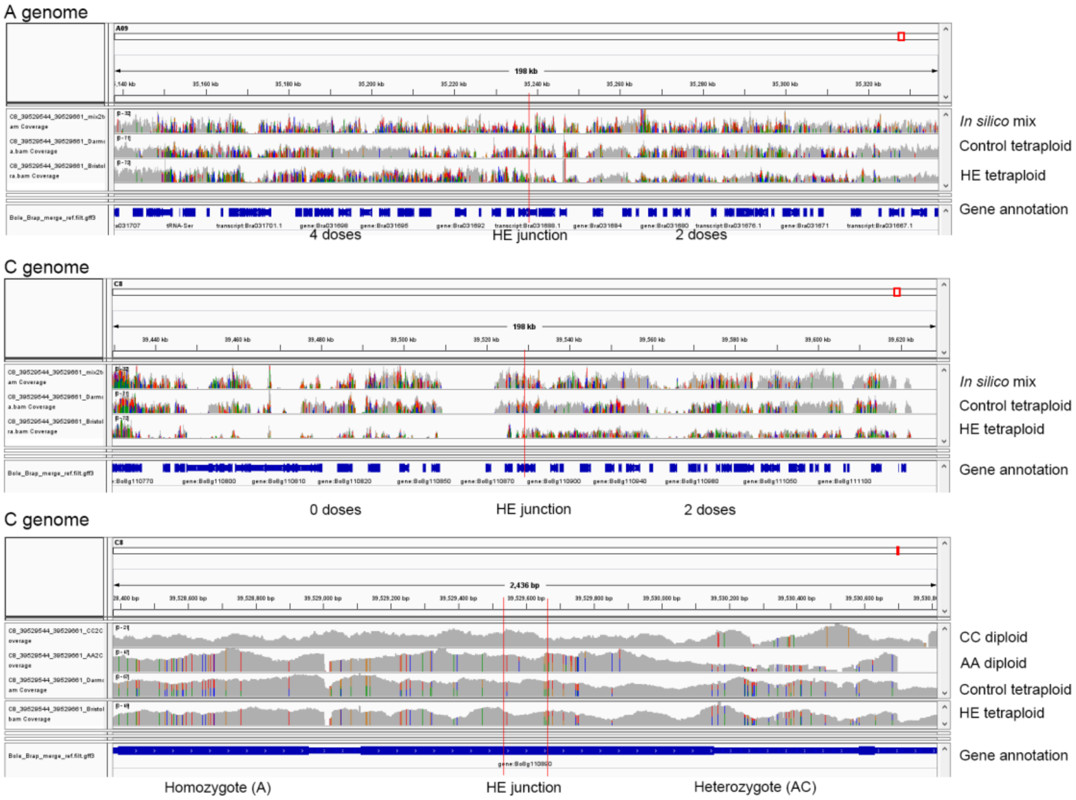


HE junction 73.

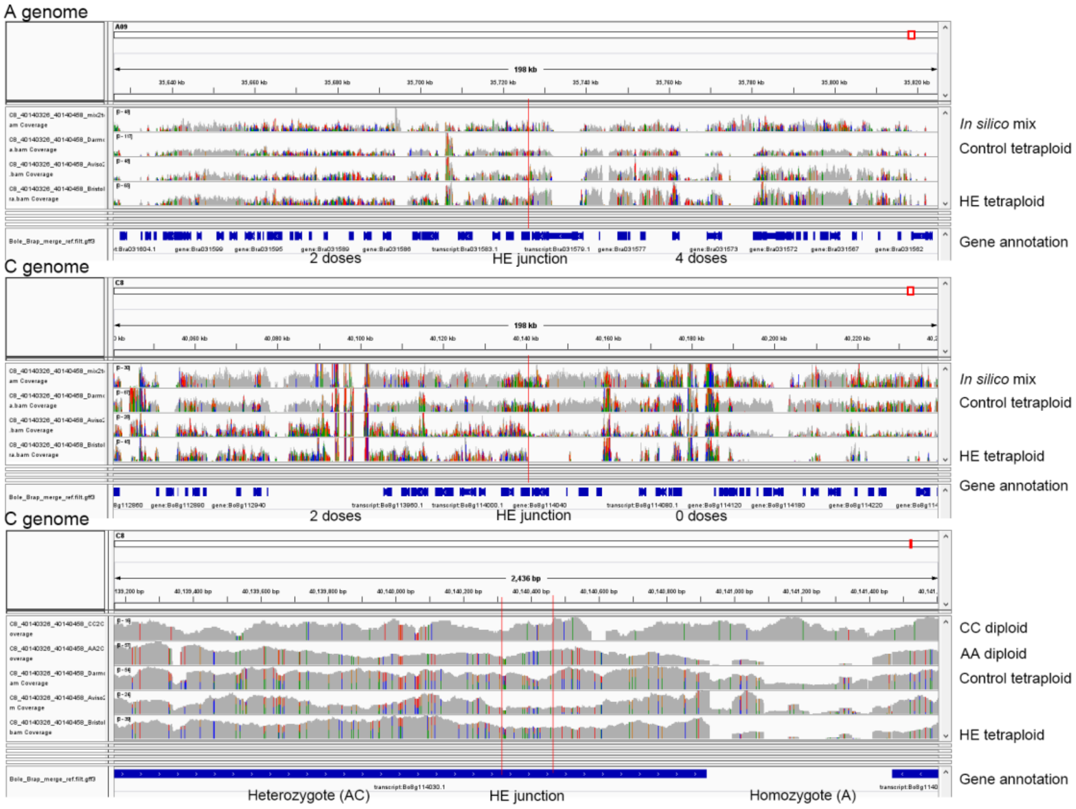
HE junction 74.



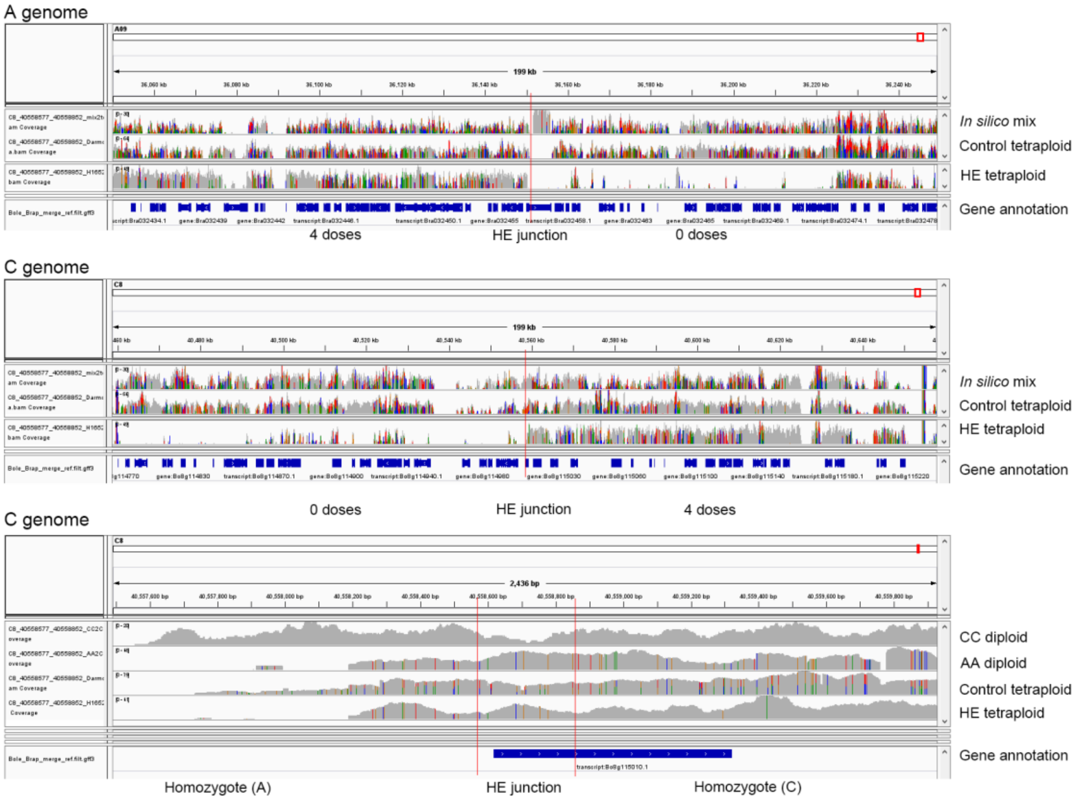
HE junction 75.



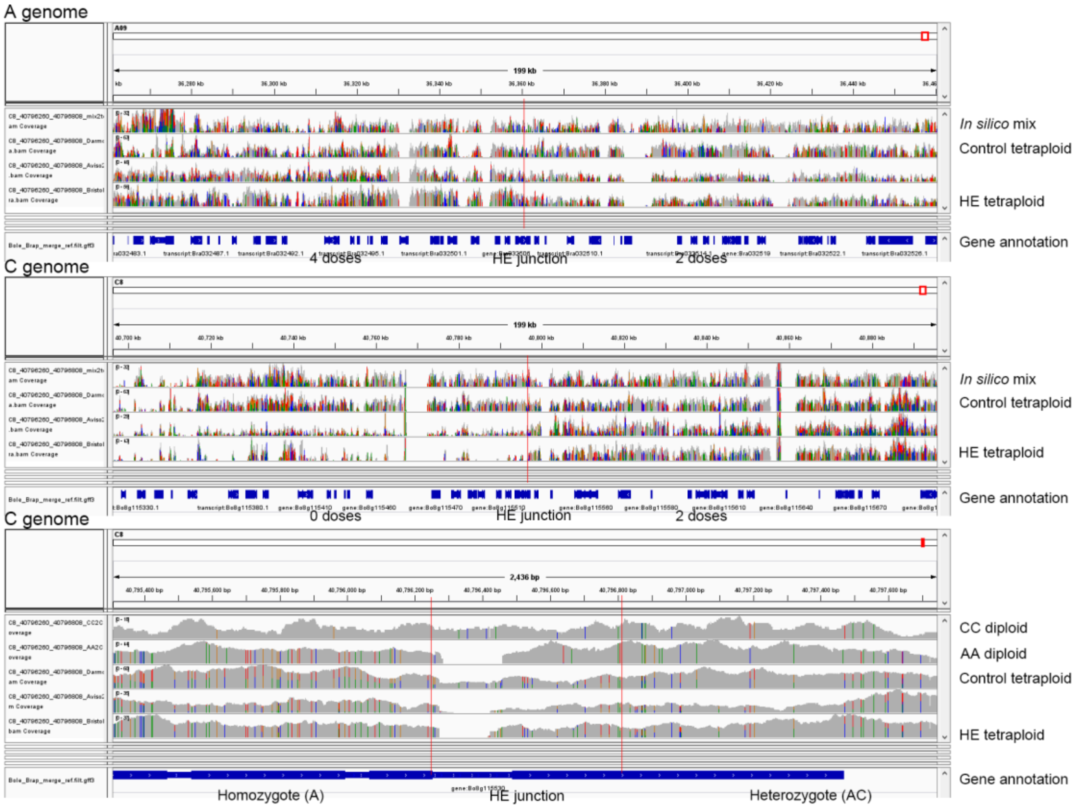
HE junction 76.



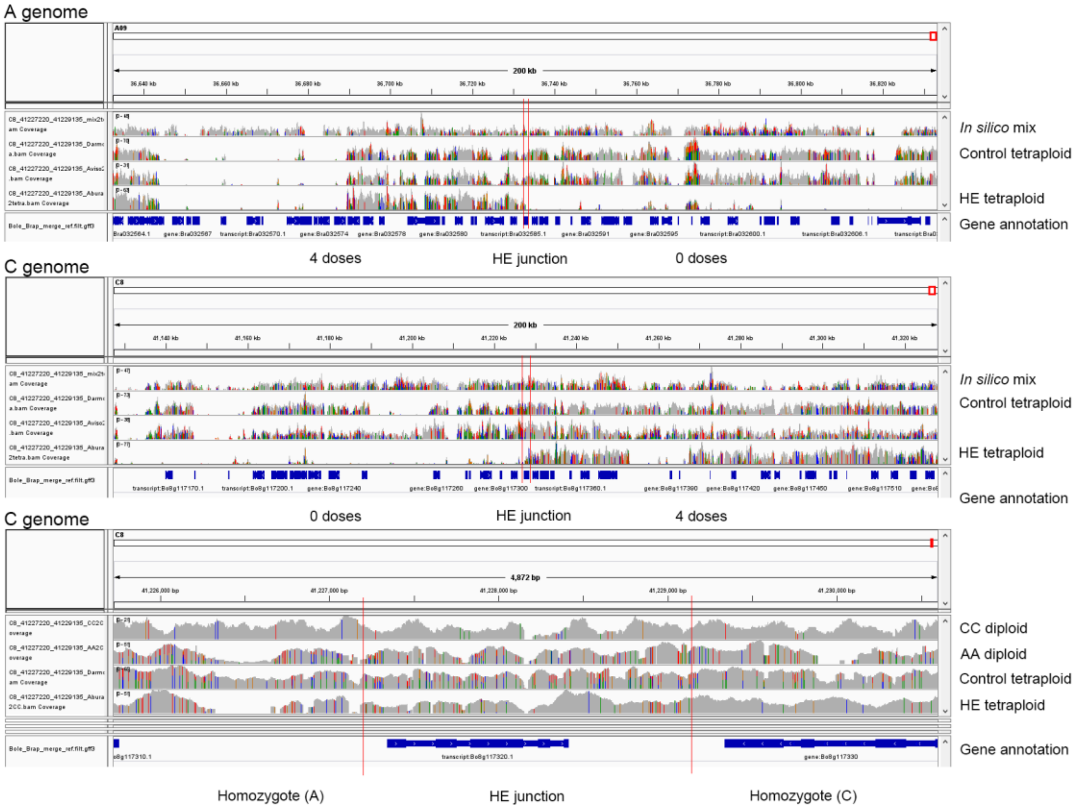
HE junction 77.



HE junction 78.

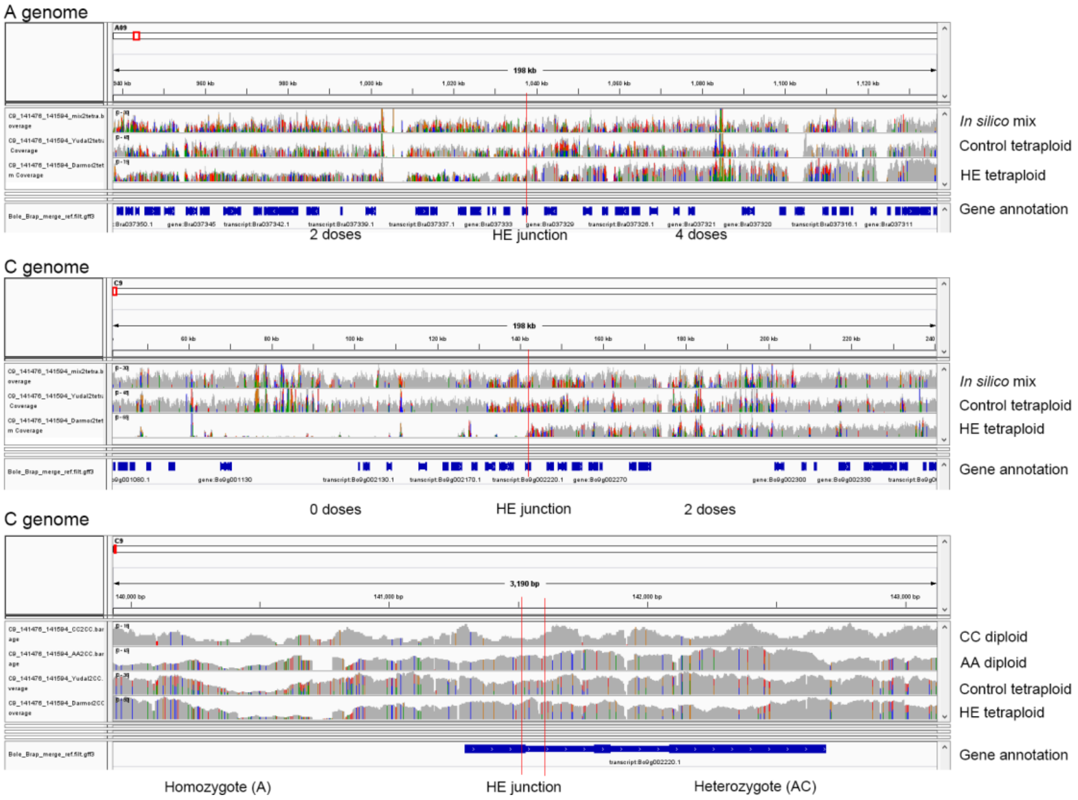


HE junction 79.



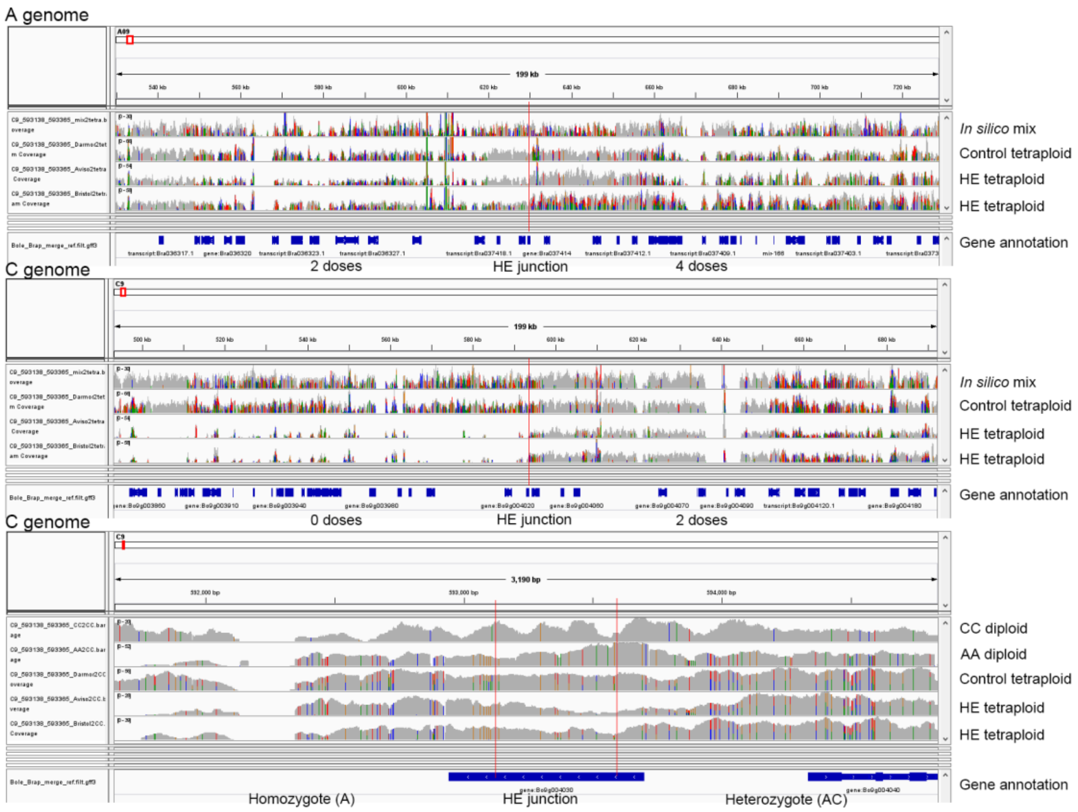
HE junction 80.

HE junction 81.



HE junction 82.

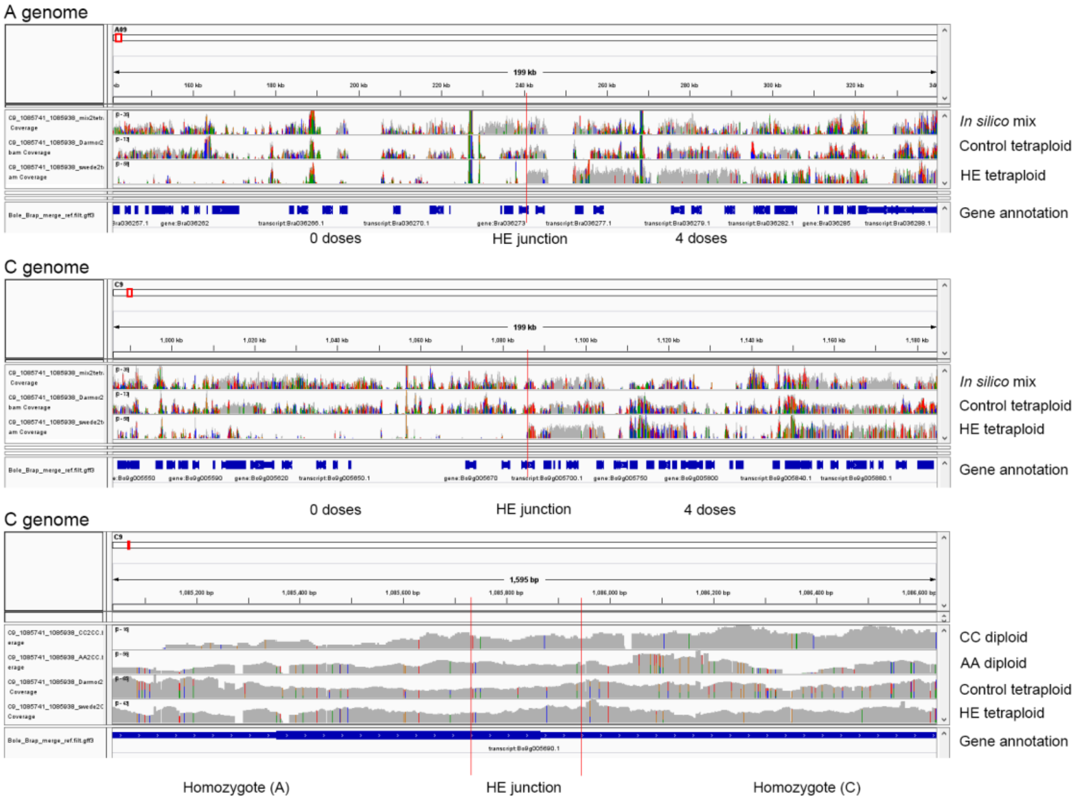
HE junction 83.



HE junction 84.

HE junction 85.

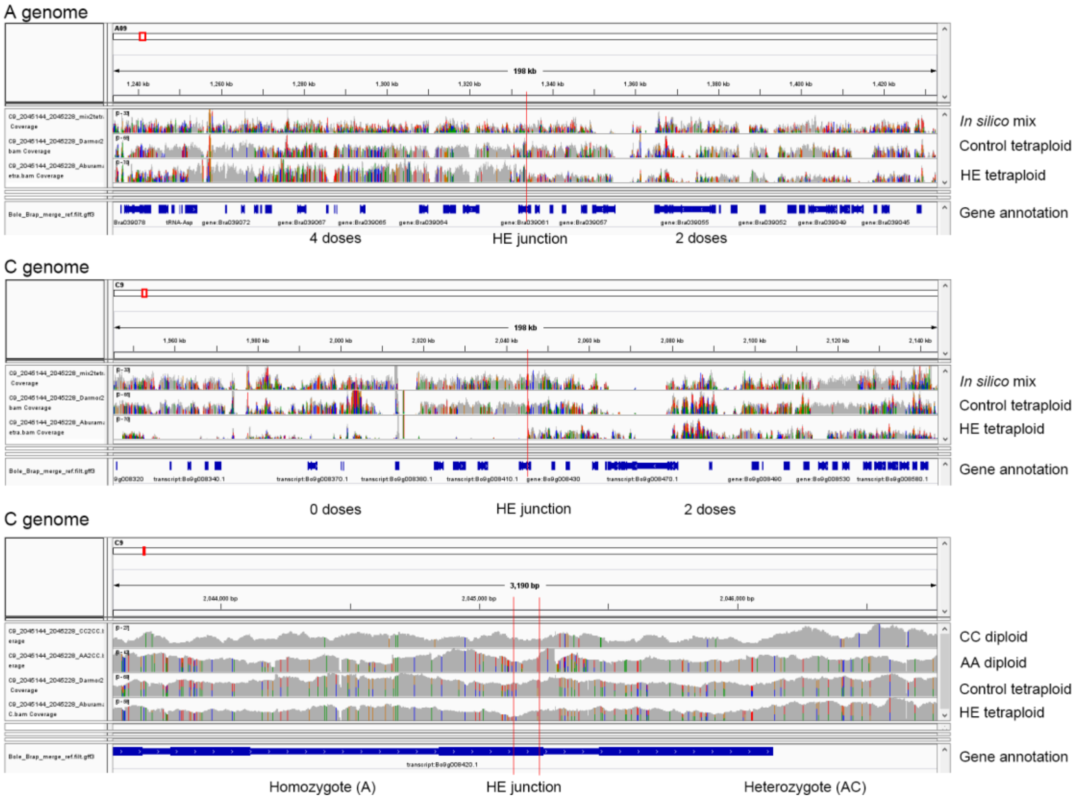
HE junction 86.



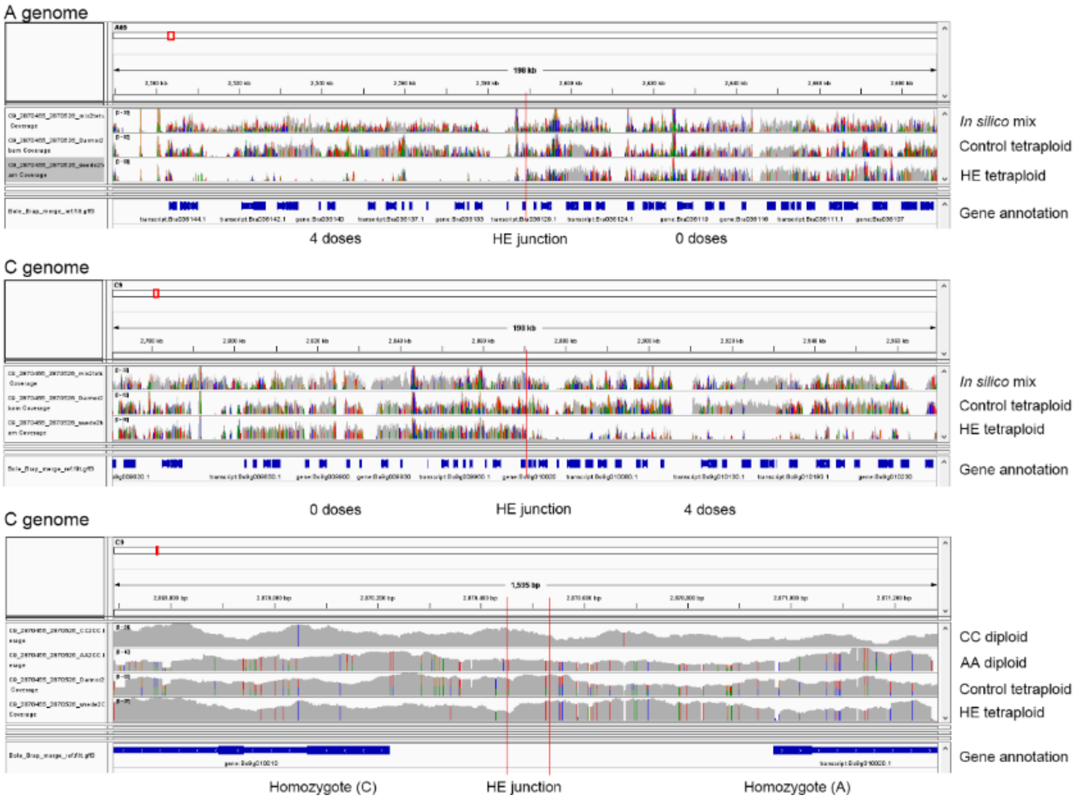
HE junction 87.

HE junction 88.

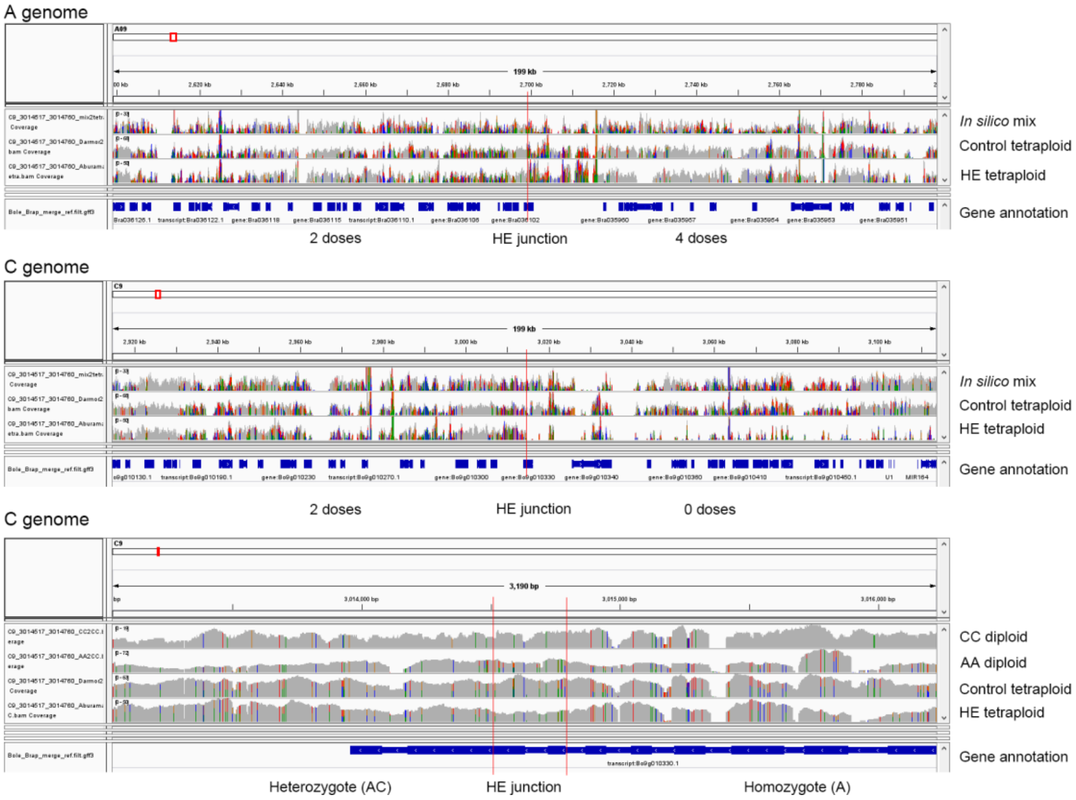
HE junction 89.



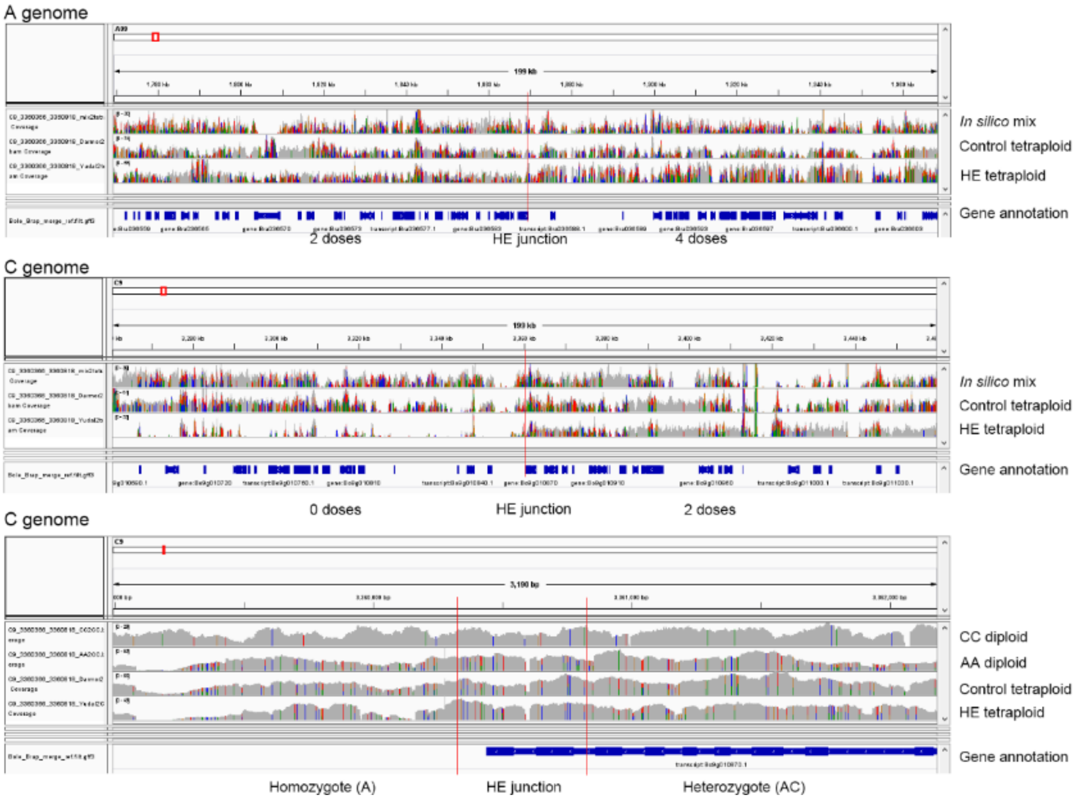
HE junction 90.



HE junction 91.

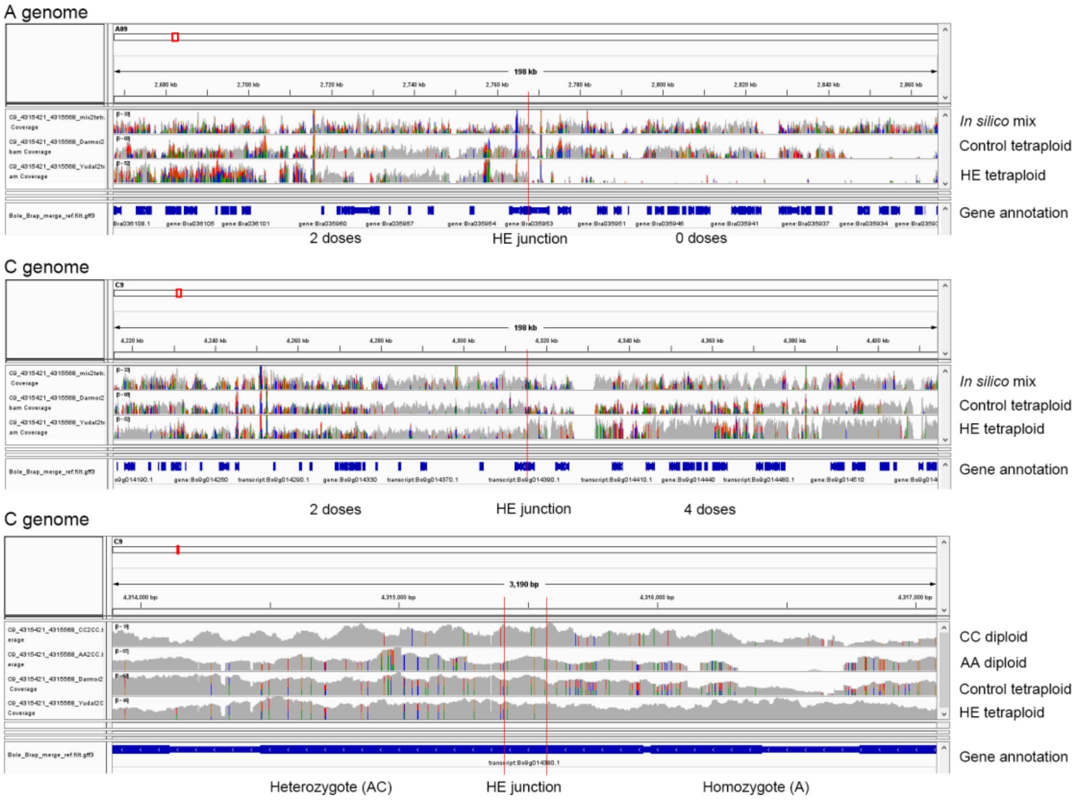


HE junction 92.

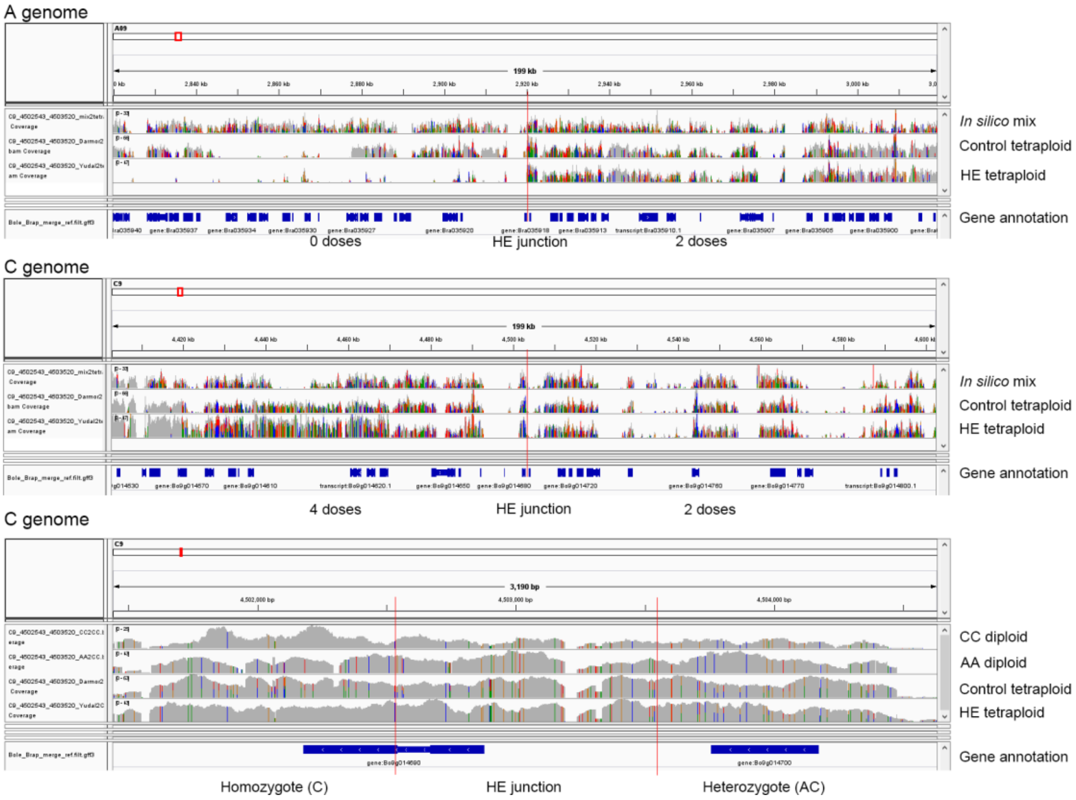


HE junction 93.

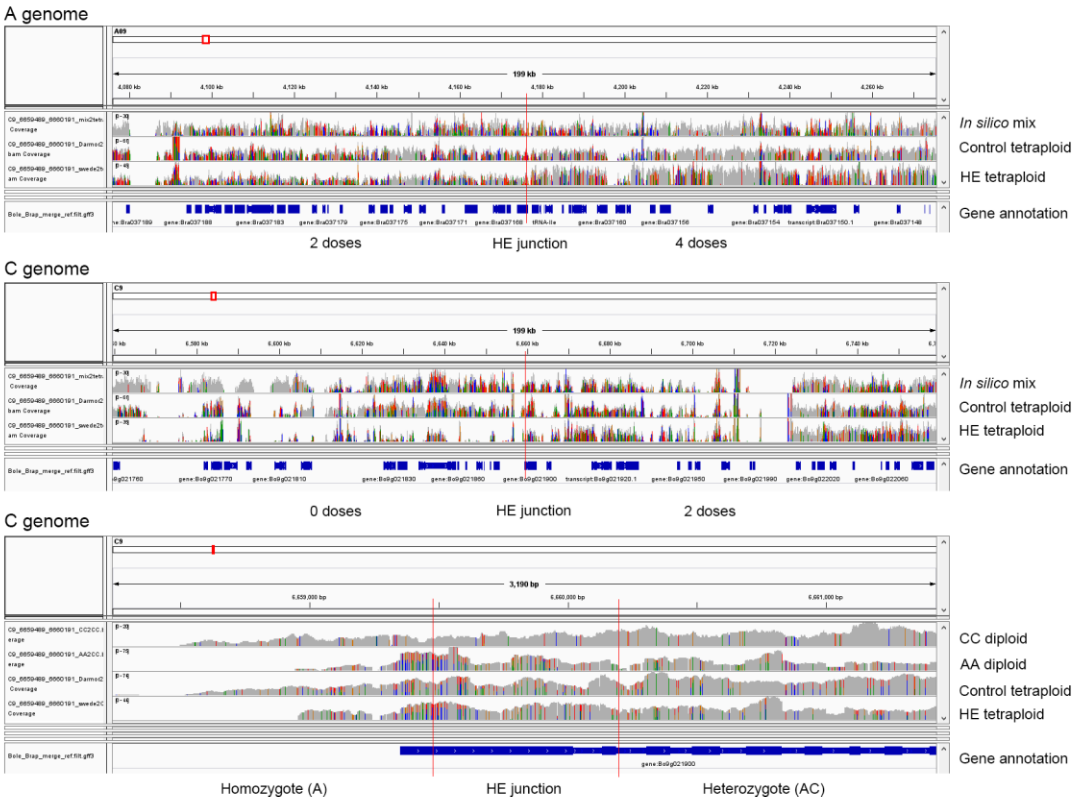
HE junction 94.



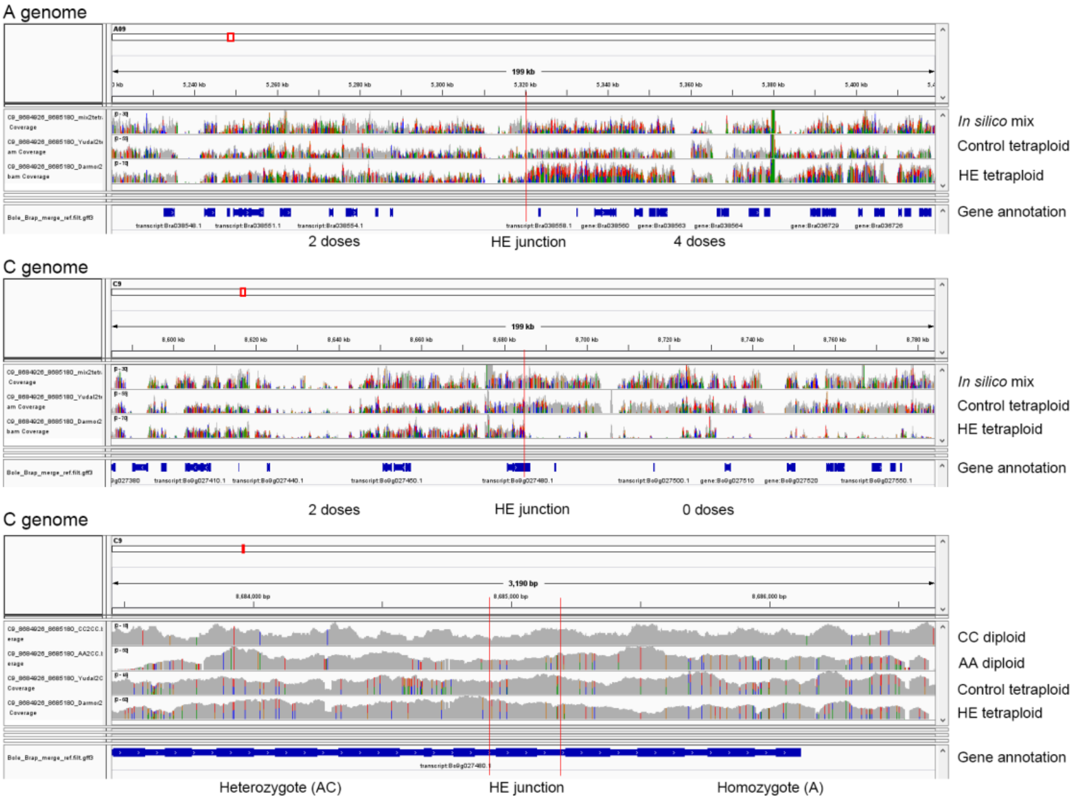
HE junction 95.



HE junction 96.

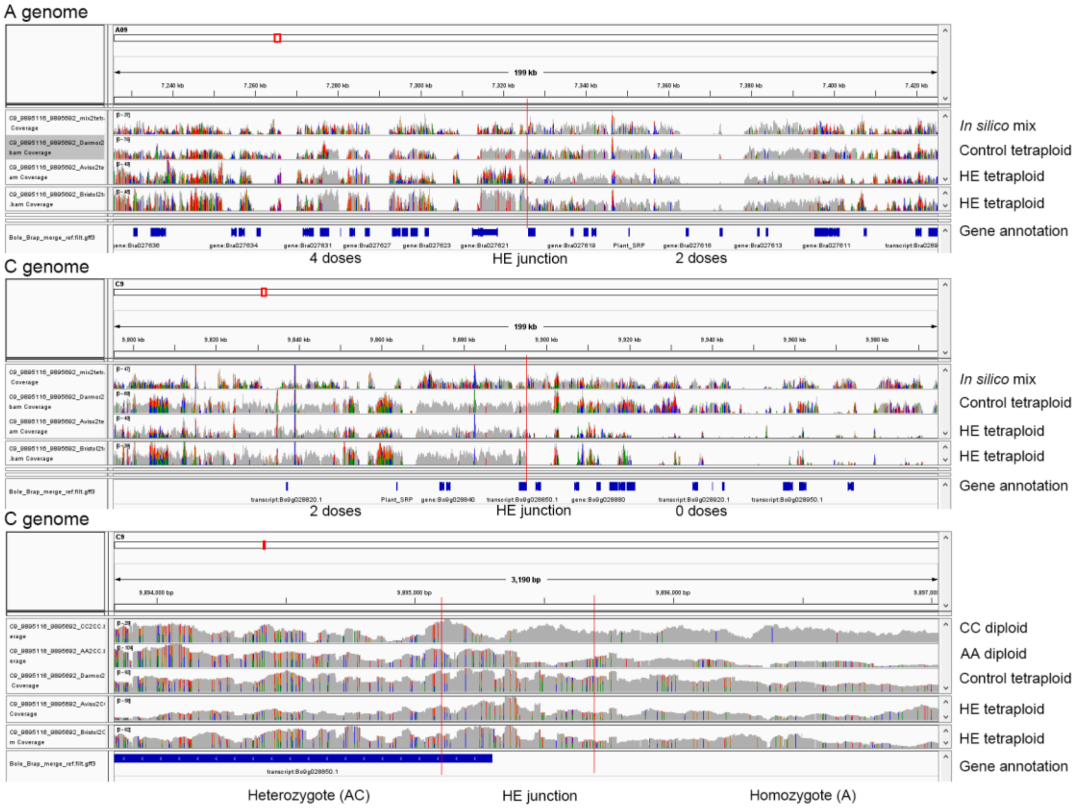


HE junction 97.

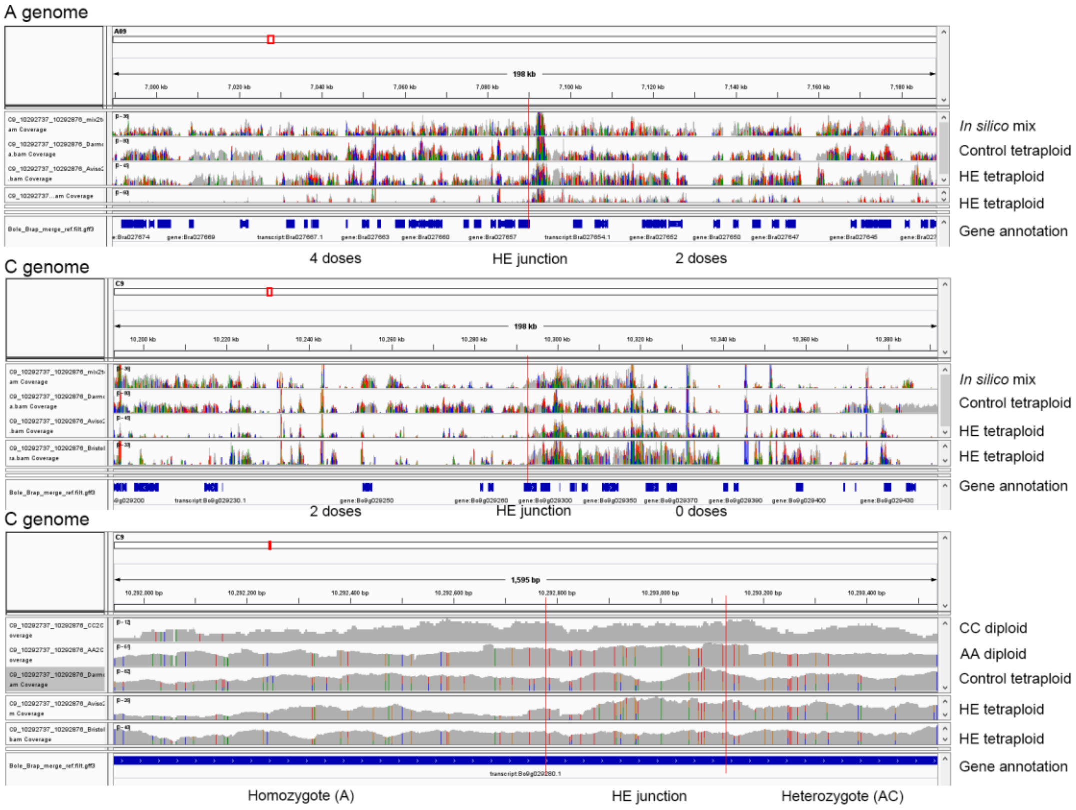


HE junction 98.

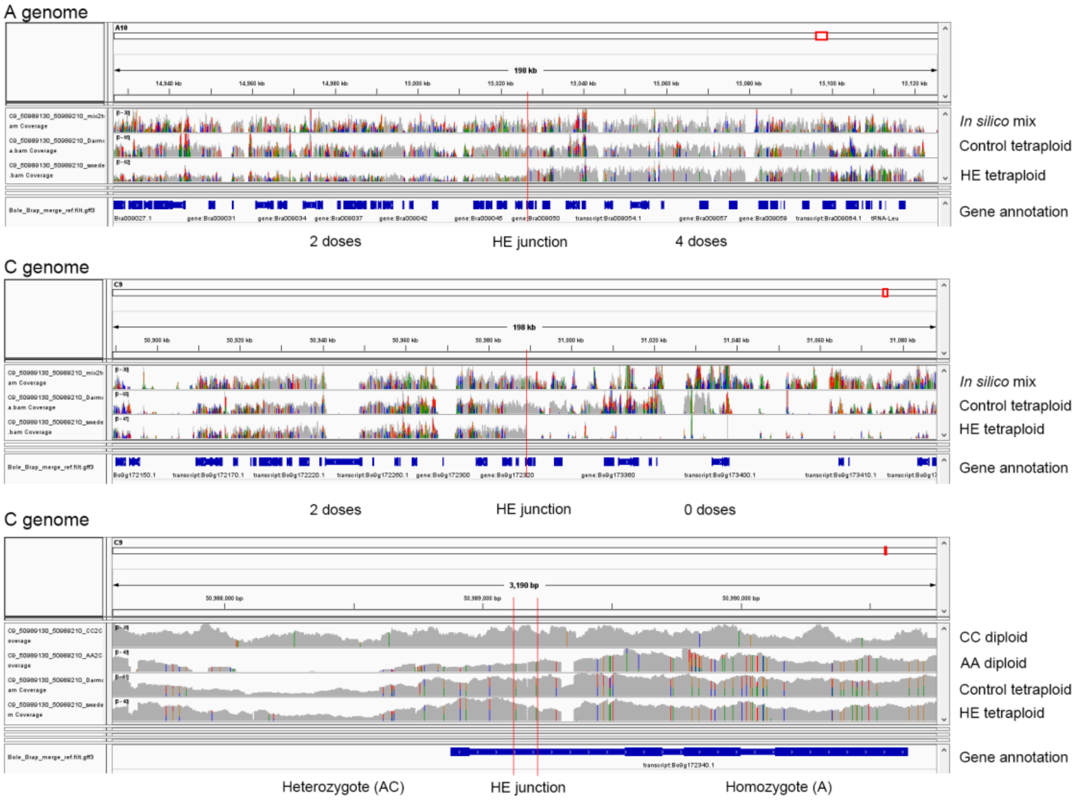
HE junction 99.



HE junction 100.



HE junction 101.



HE junction 102.

HE junction 103.

HE junction 104.

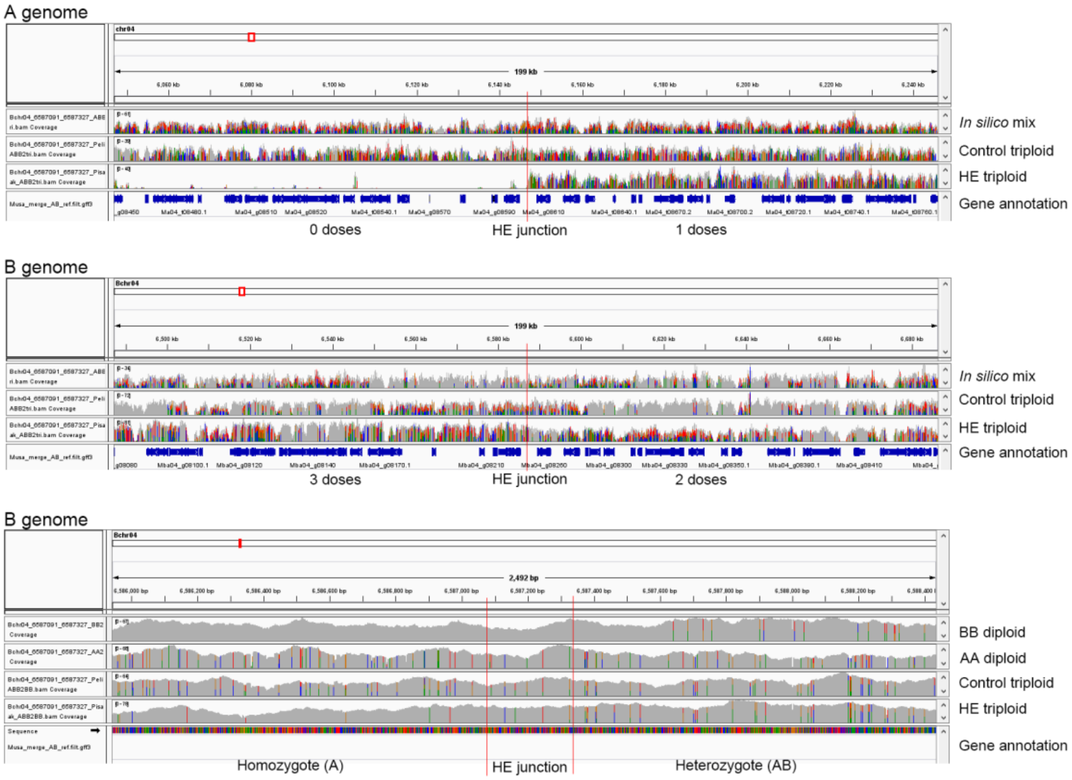
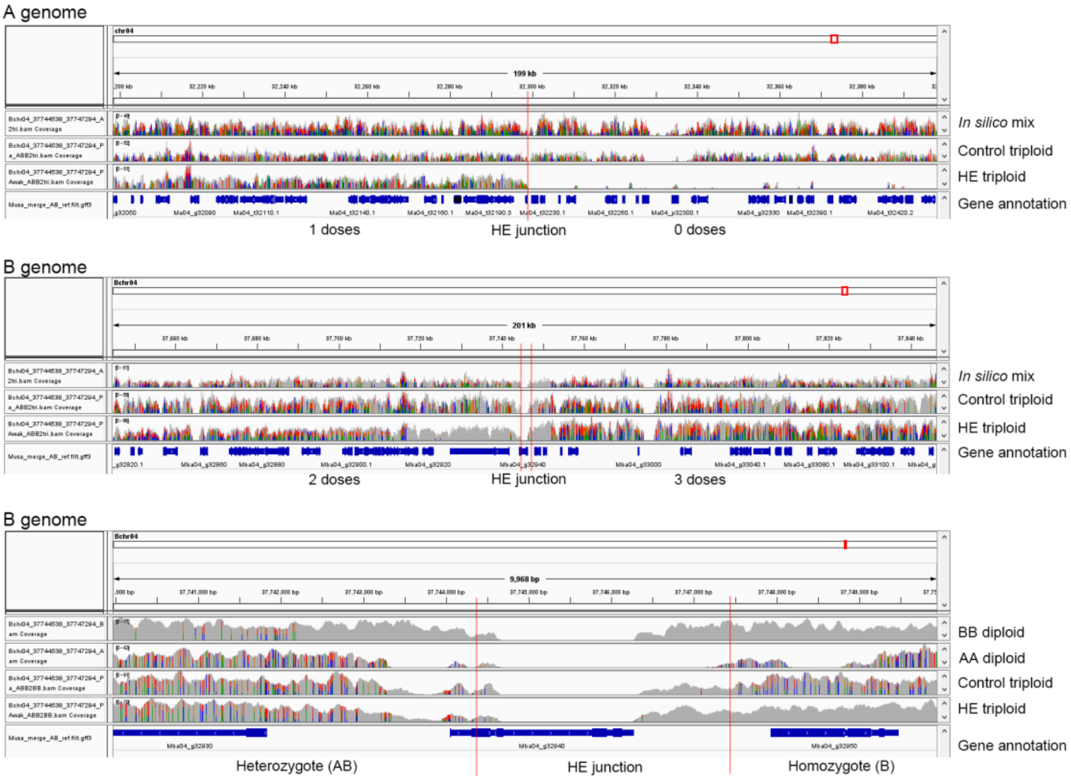
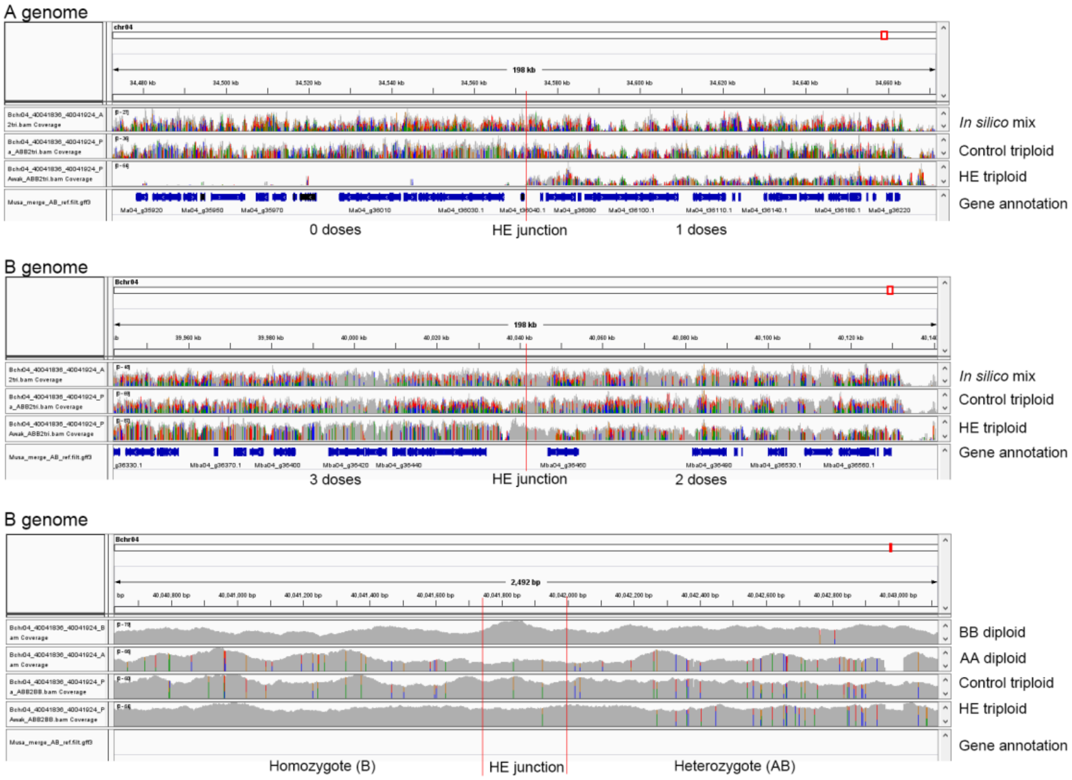


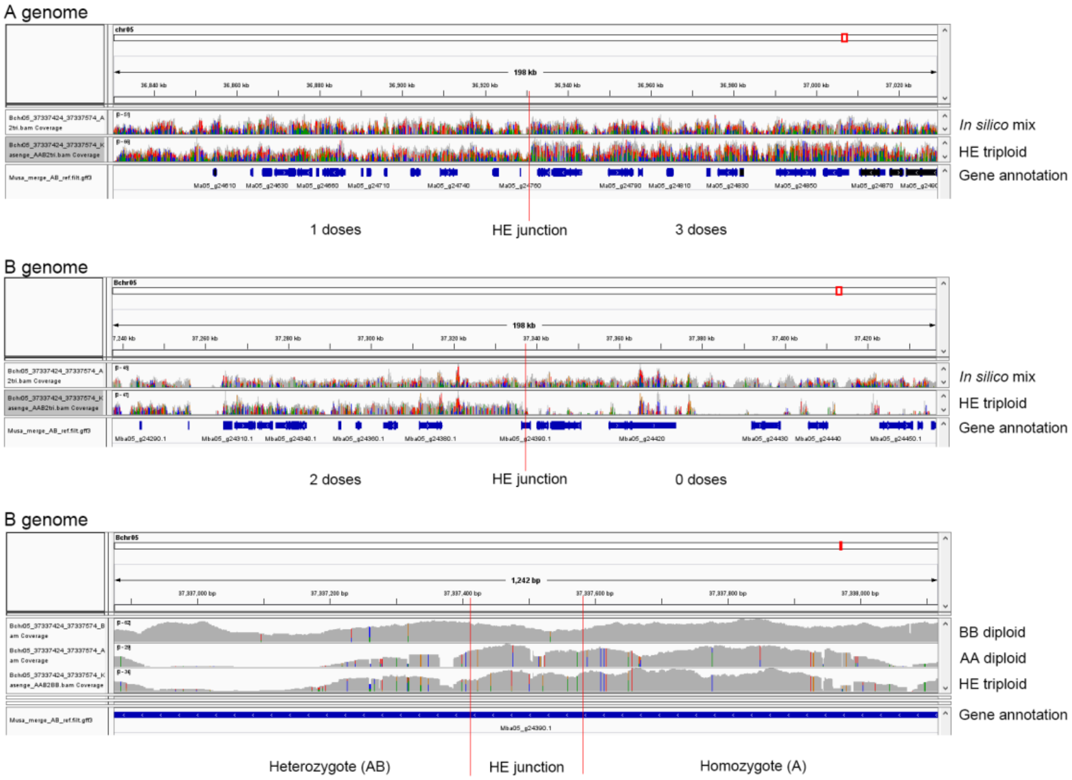
Fig. S16. IGV panels of HE junction regions in triploid banana (AAB and ABB, 21 junctions). The top and middle panels show the sequencing depth around HE junctions in both A and B subgenomes, respectively. Red lines indicate the position of HE junctions and genome dosages are shown on both sides of HE junction. The bottom panel shows the genotypes (homozygote or heterozygote) around HE junction (region between red lines) based on SNP genotyping.



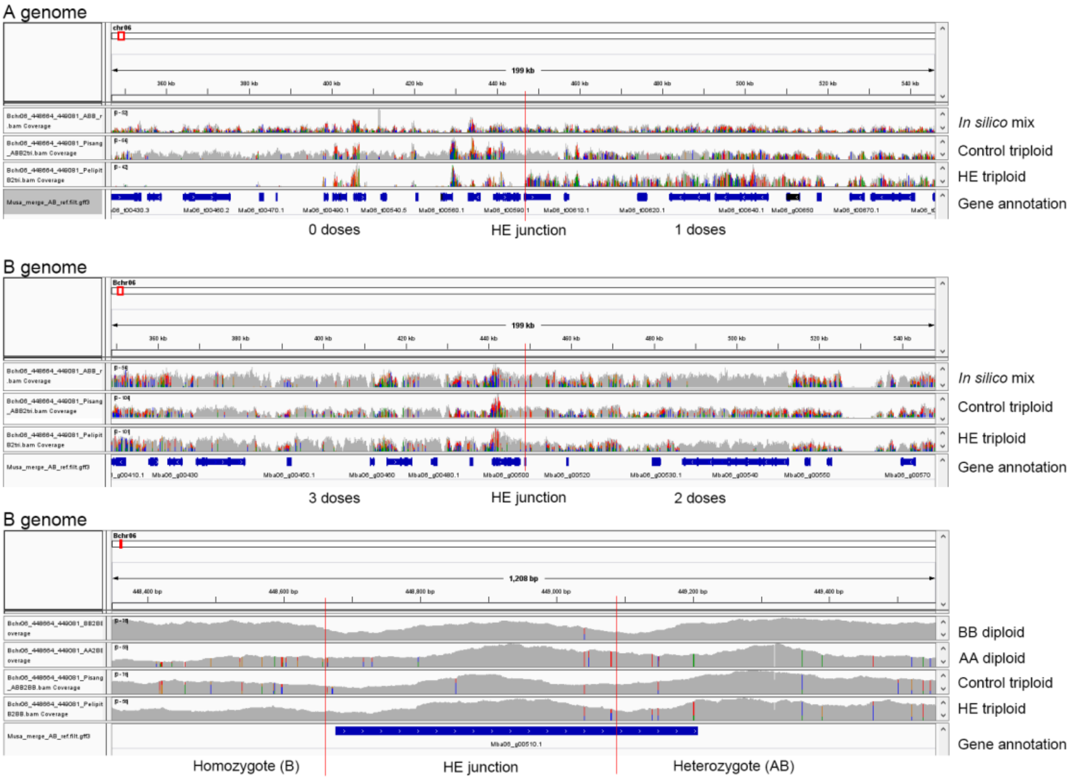
HE junction 2.



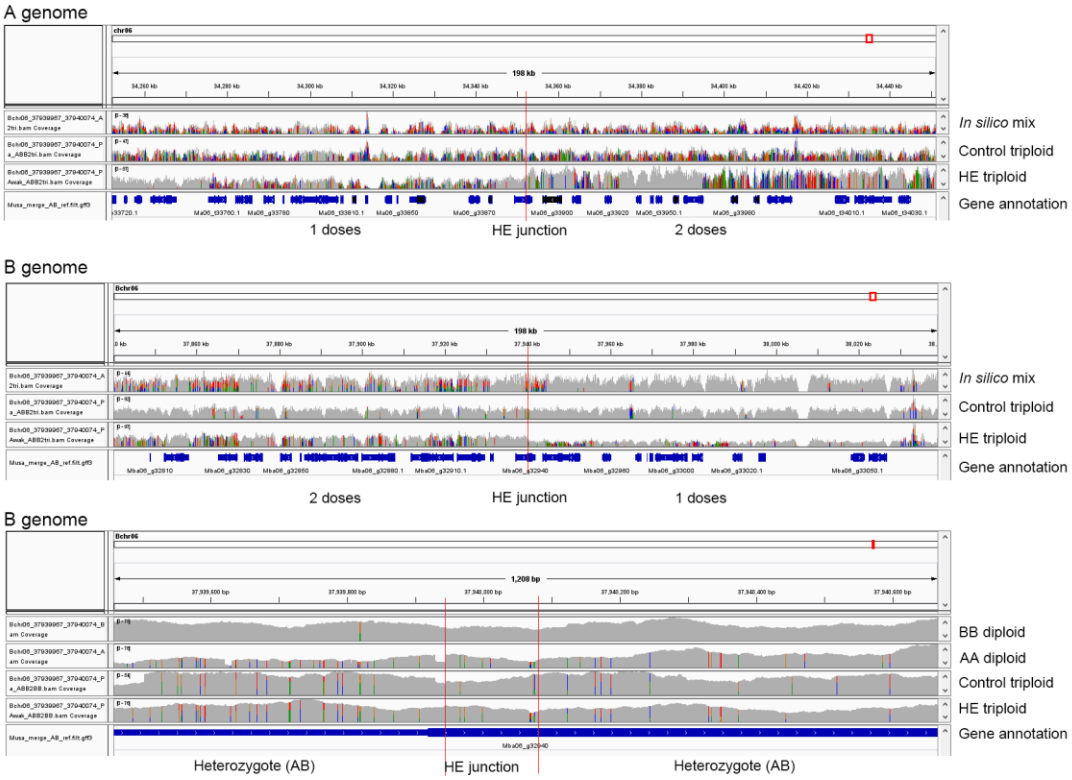
HE junction 3.



HE junction 4.



HE junction 5.

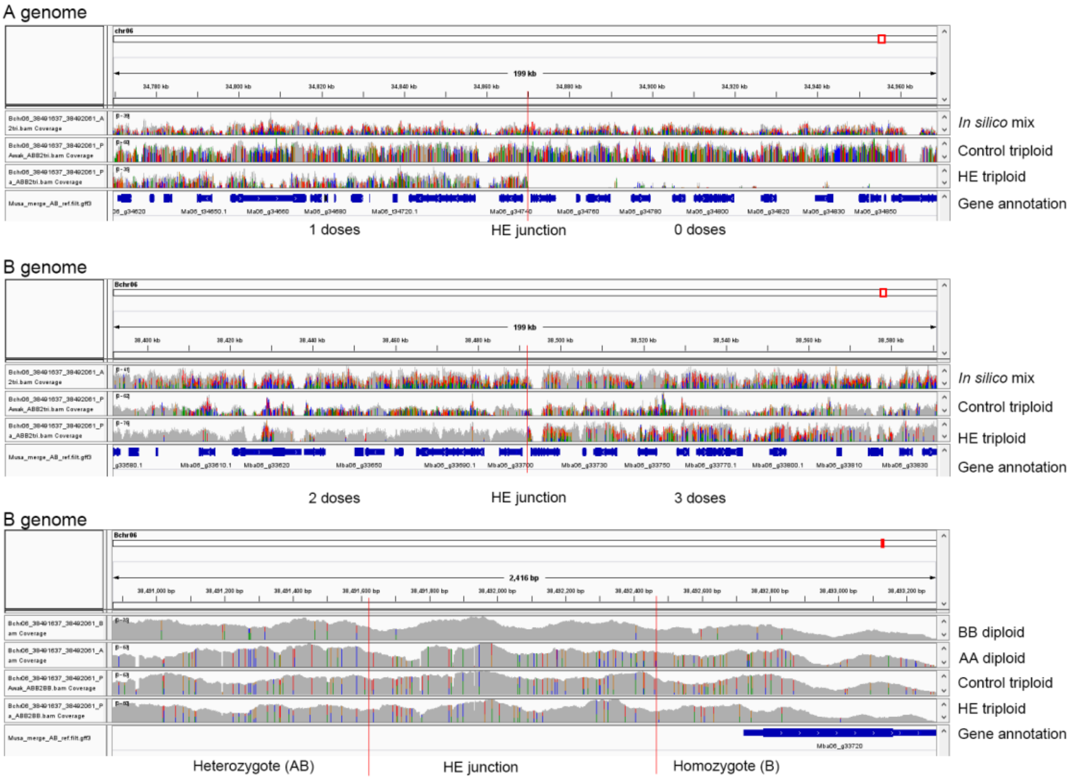


A genome

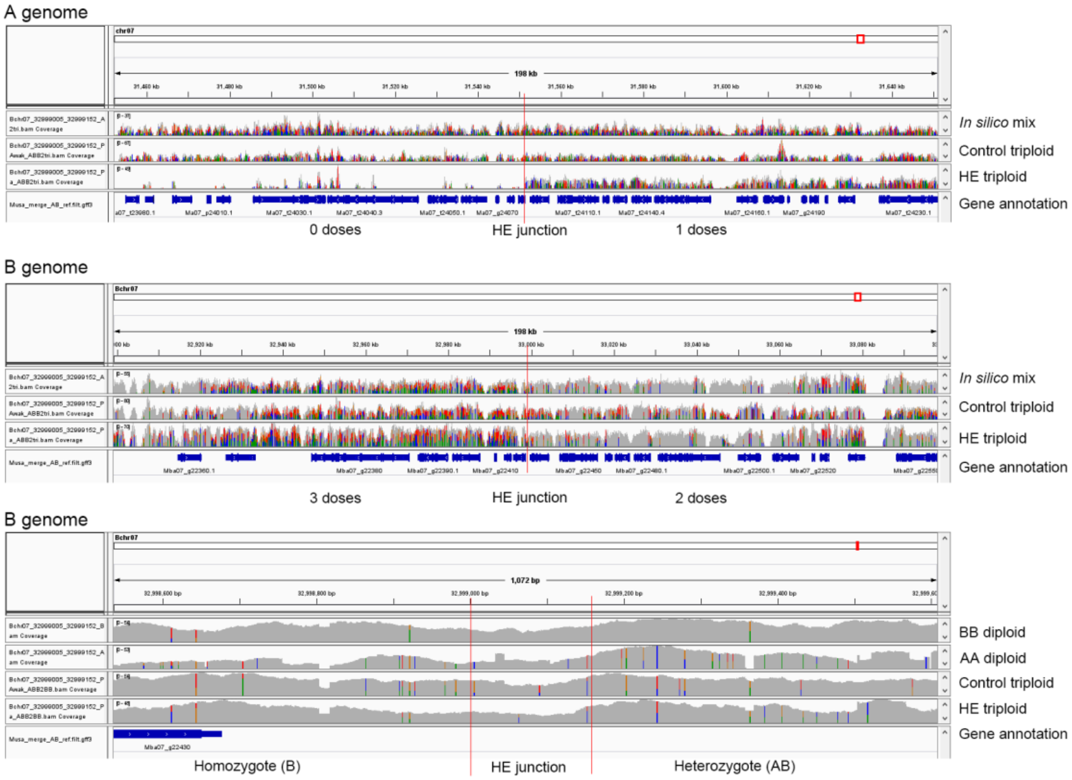
B genome

B genome

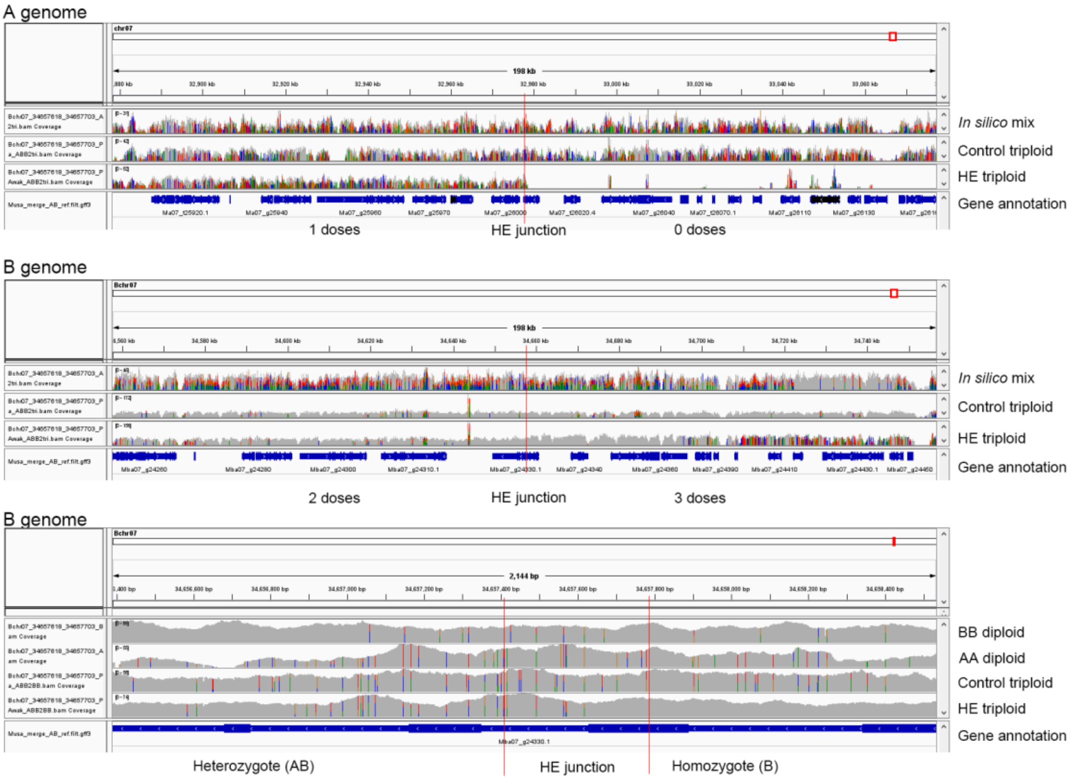
HE junction 6.



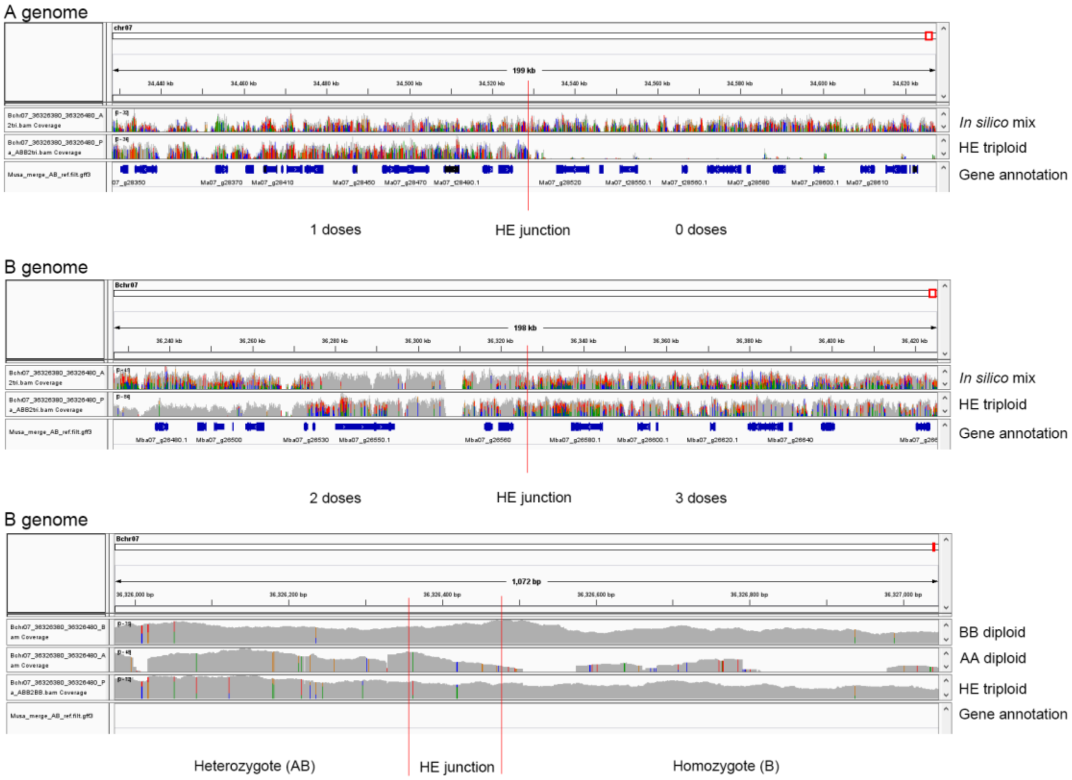
HE junction 7.



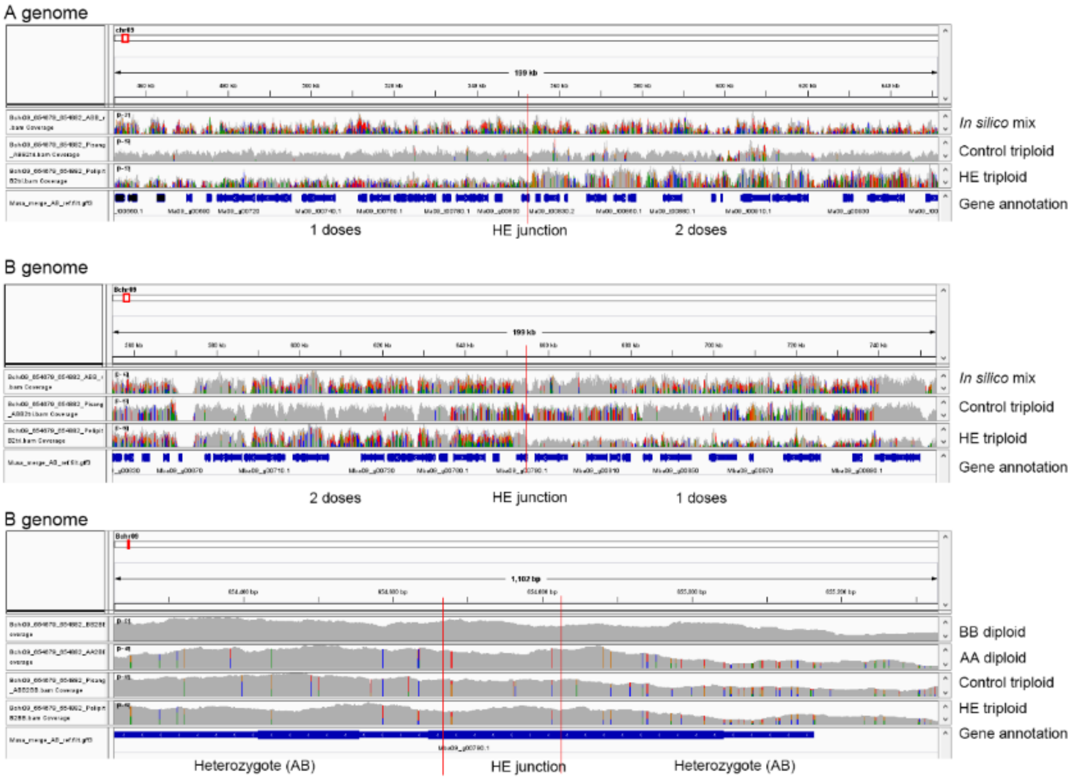
HE junction 8.



HE junction 9.



HE junction 10.

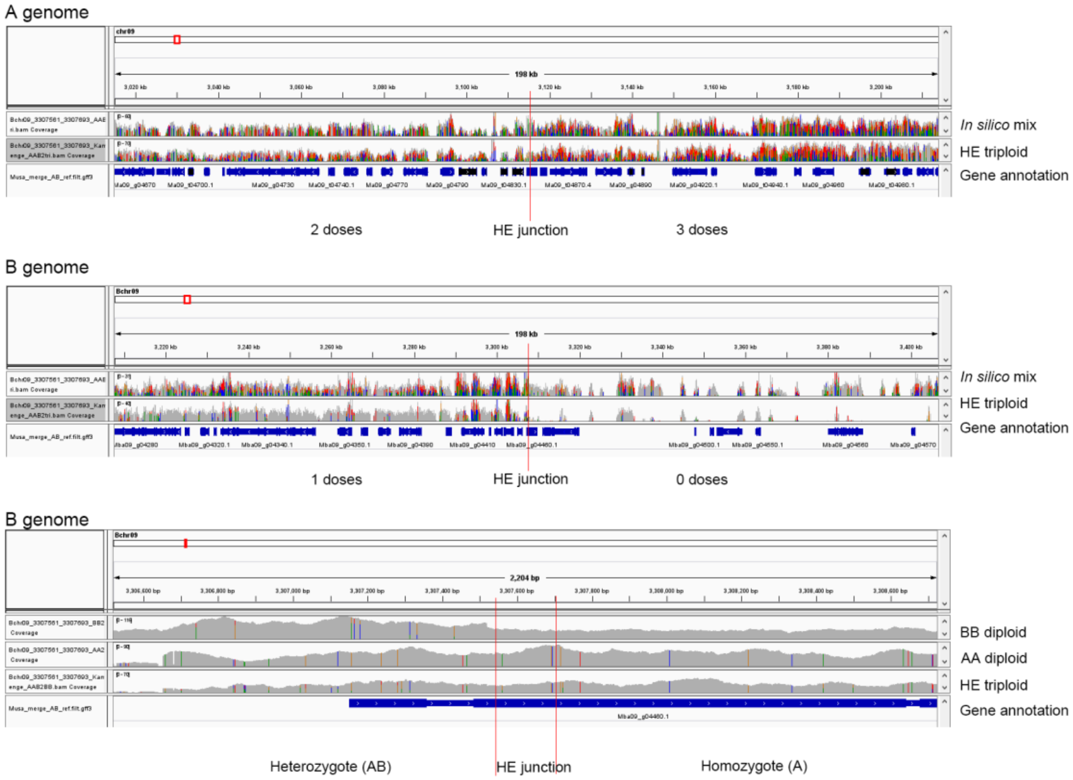


A genome

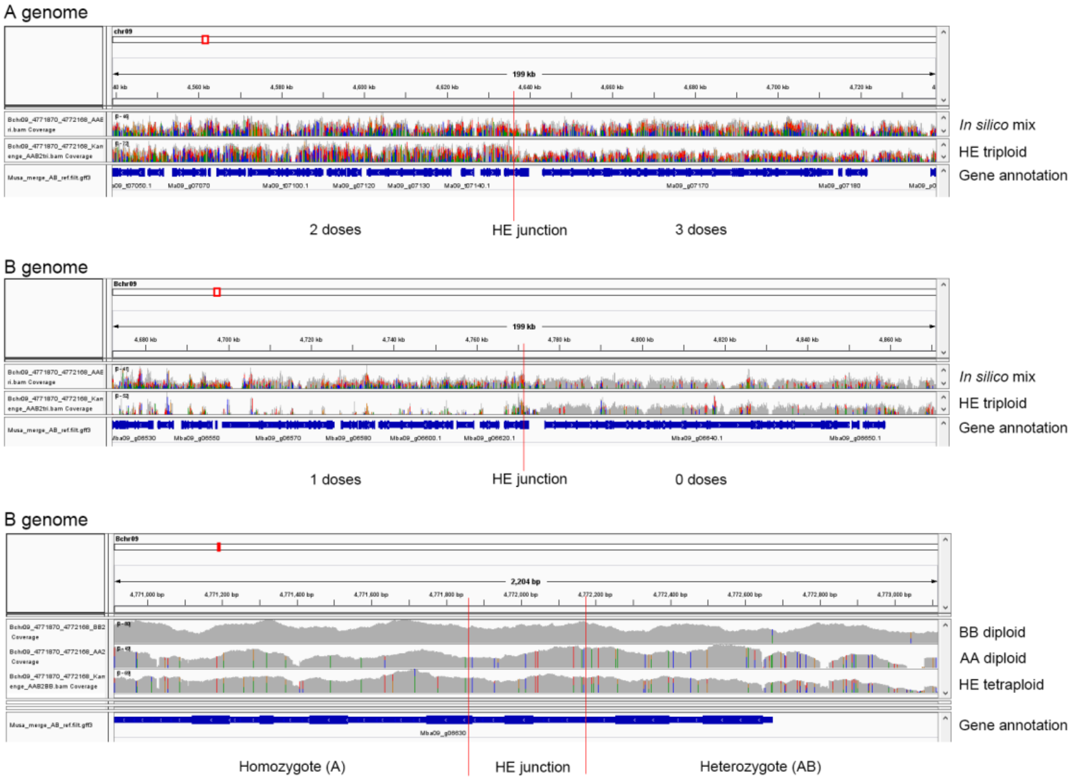
B genome

B genome

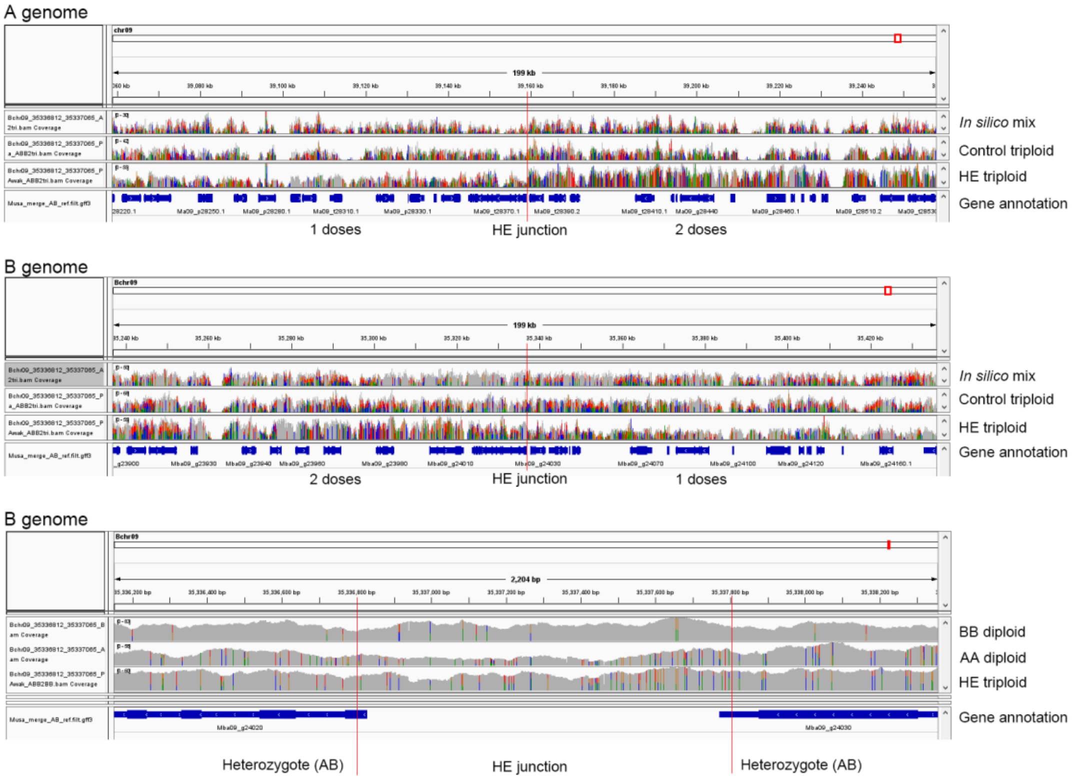
HE junction 11.



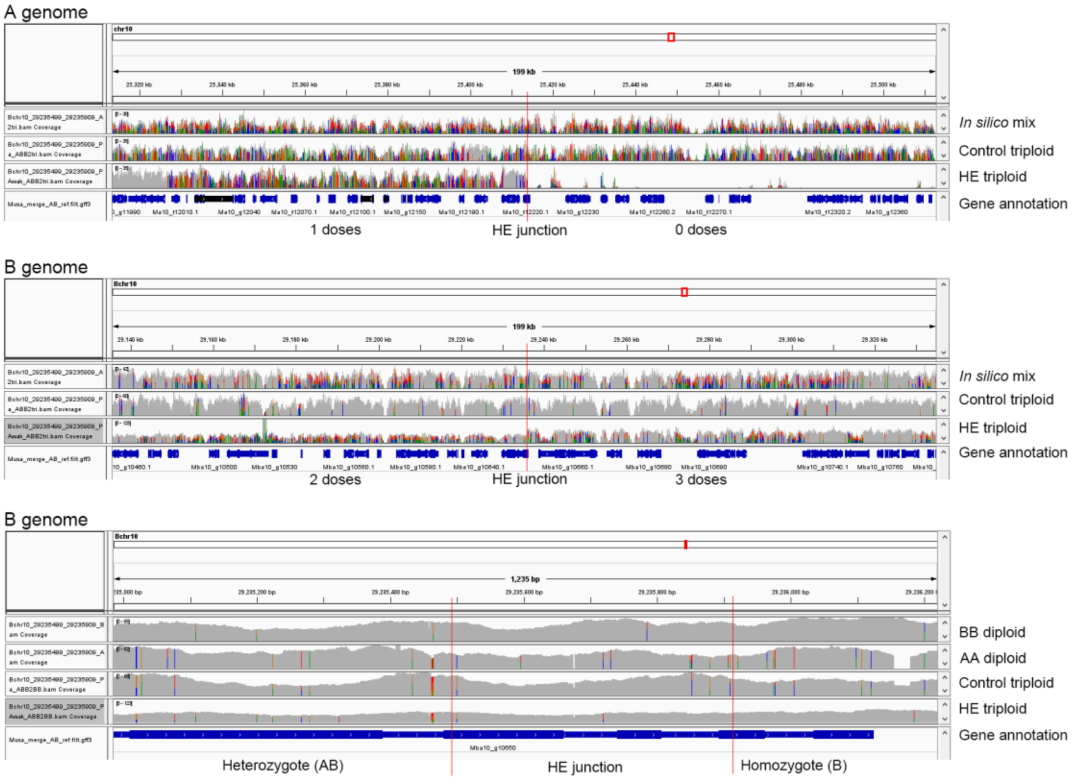
HE junction 12.



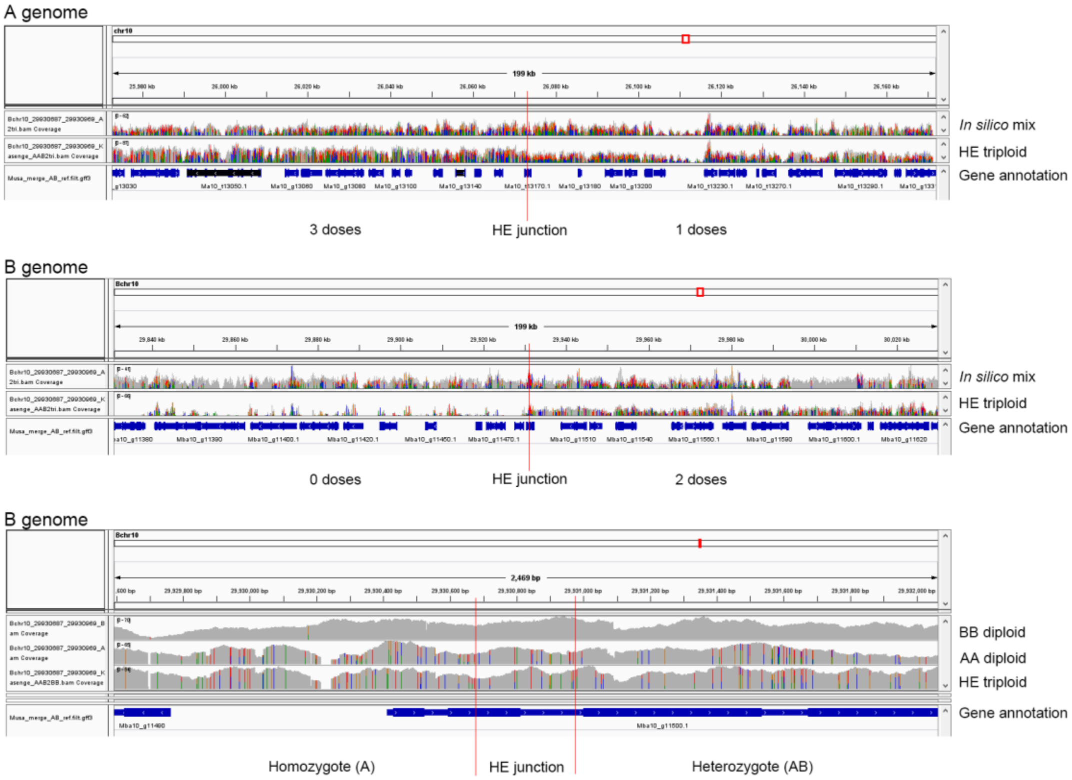
HE junction 13.



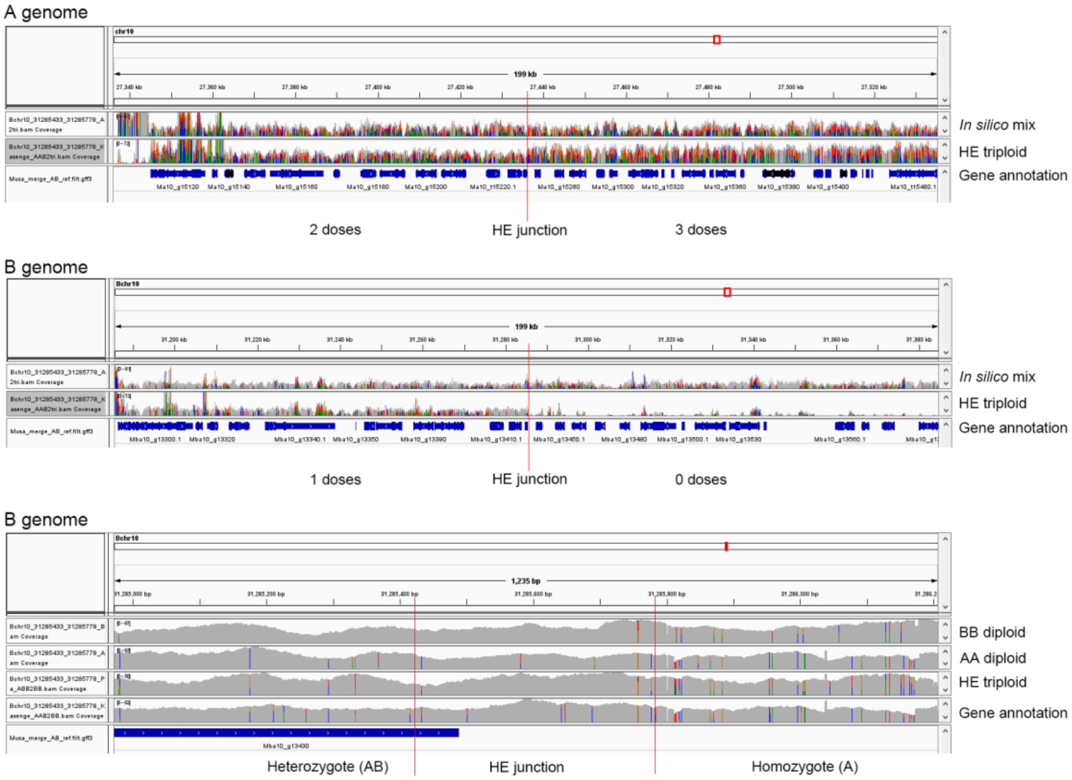
HE junction 14.



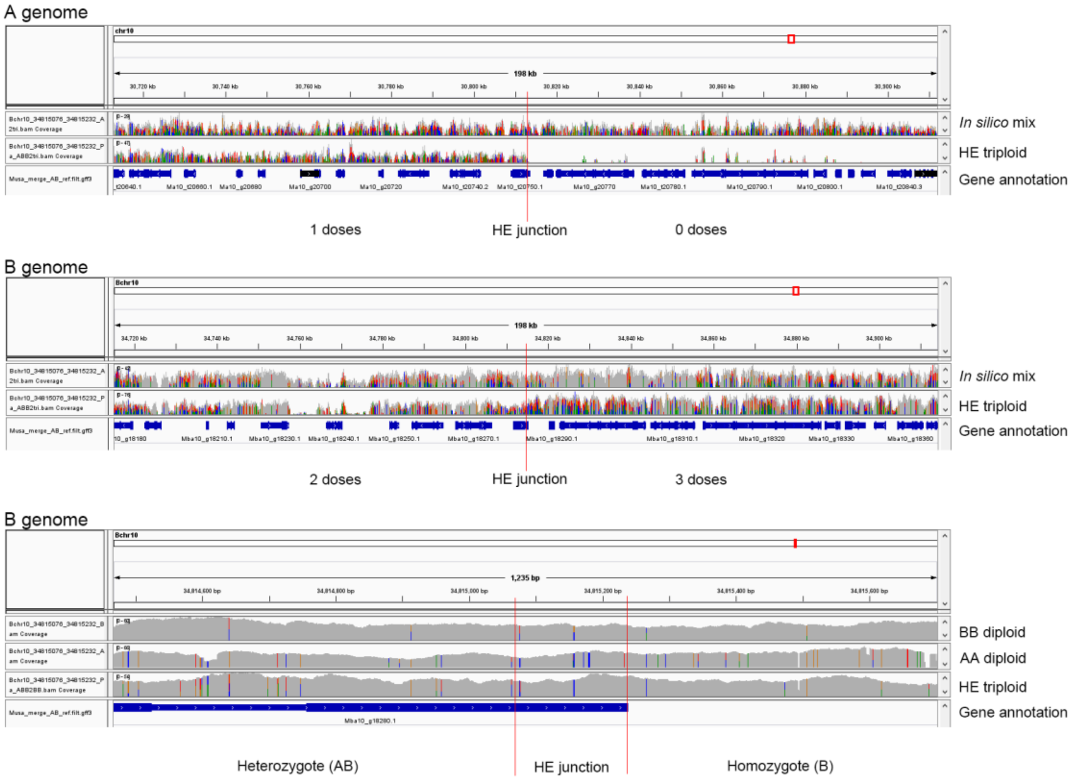
HE junction 15.



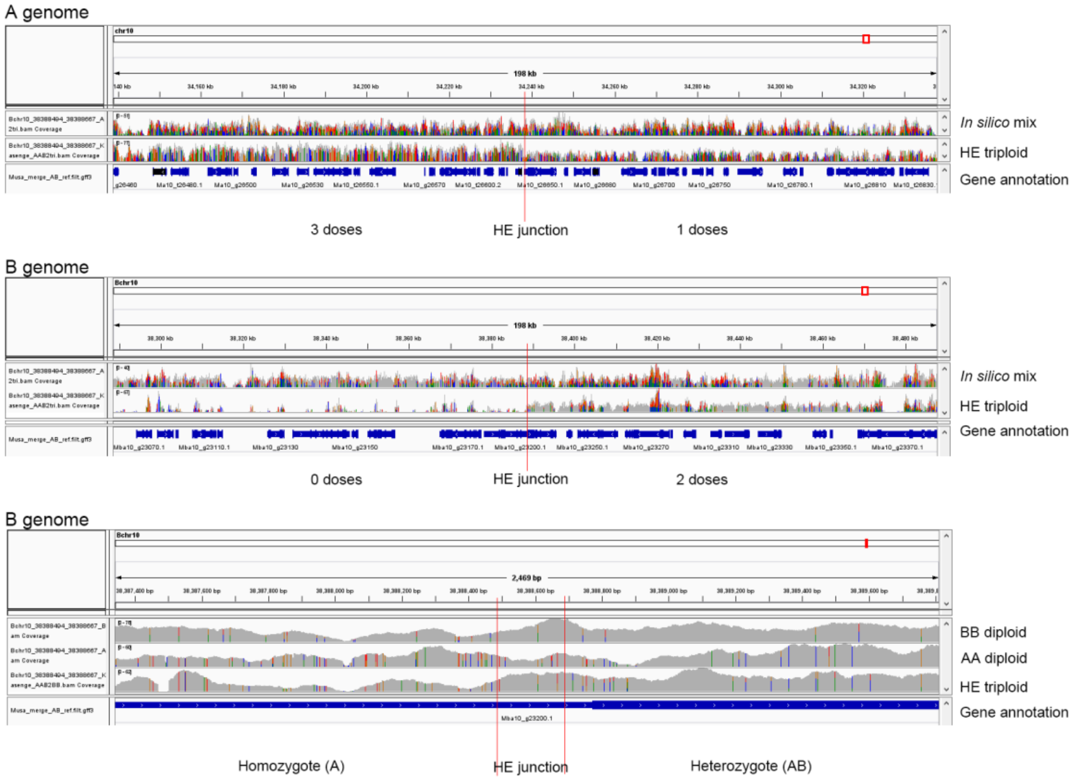
HE junction 16.



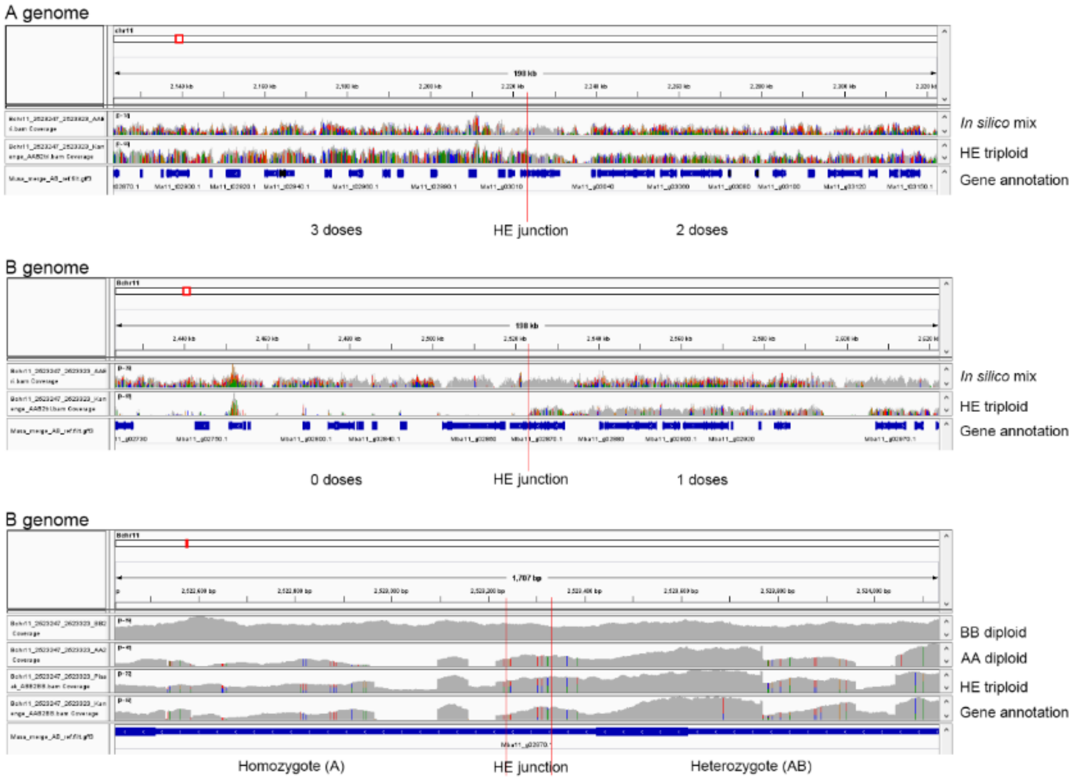
HE junction 17.



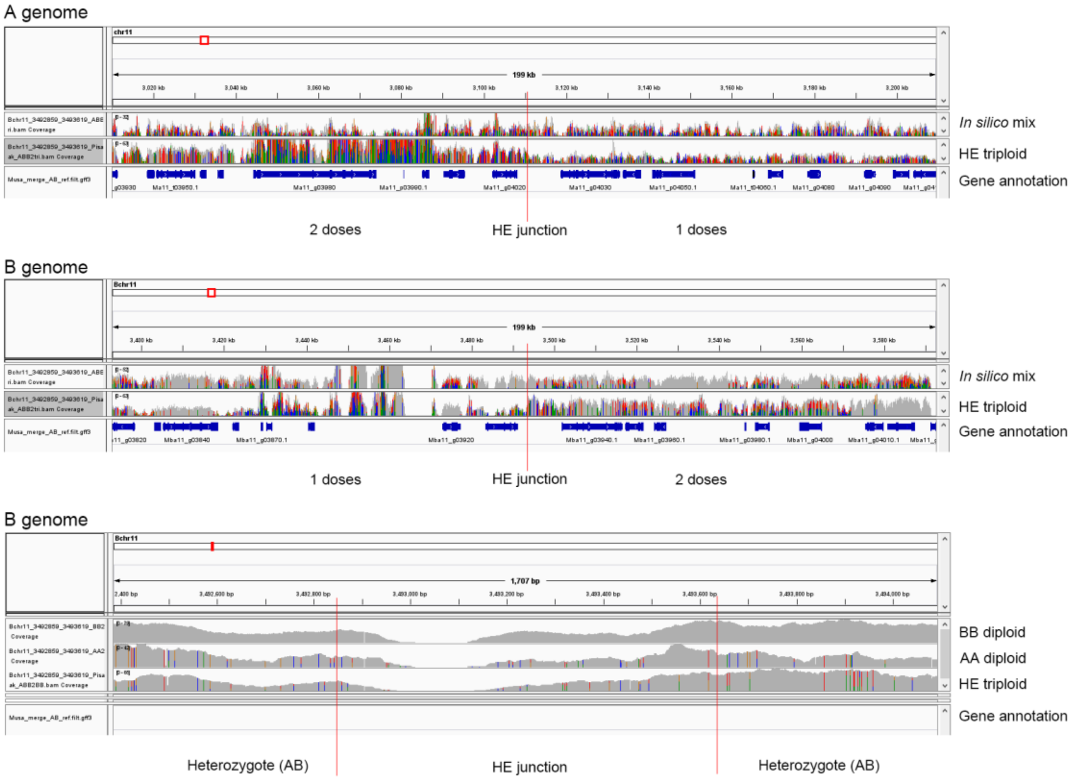
HE junction 18.



HE junction 19.



HE junction 20.



HE junction 21.

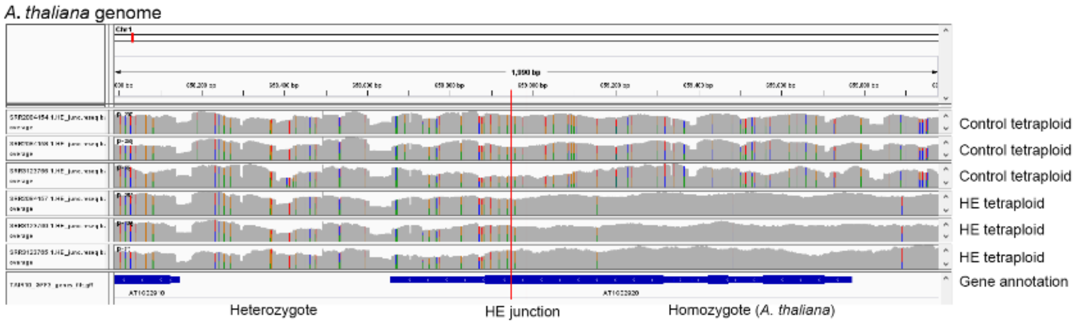
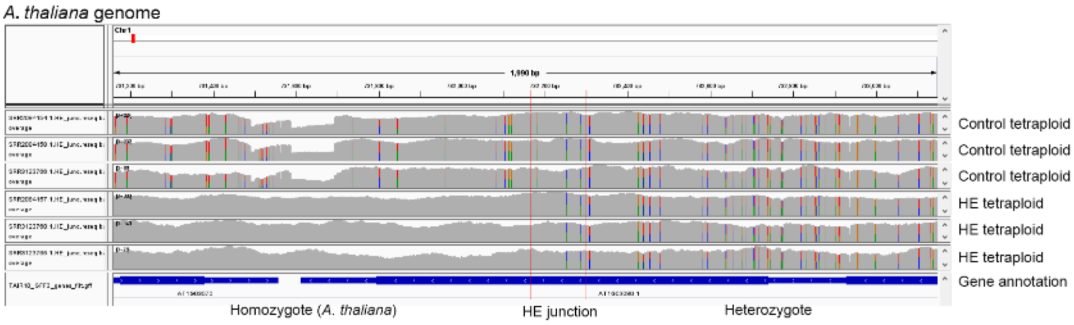
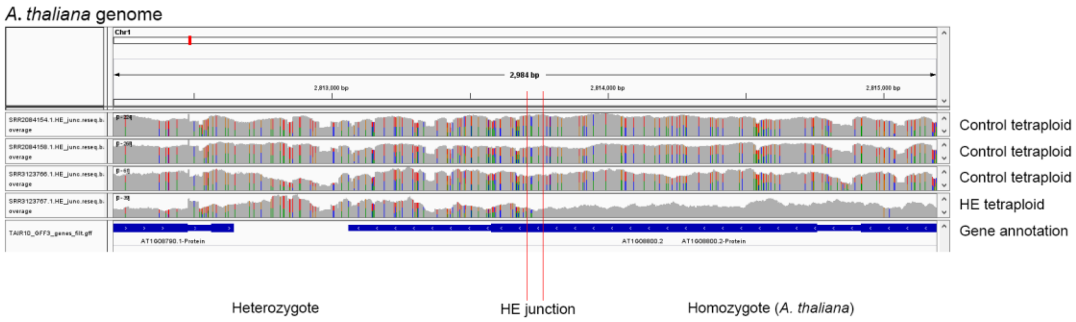


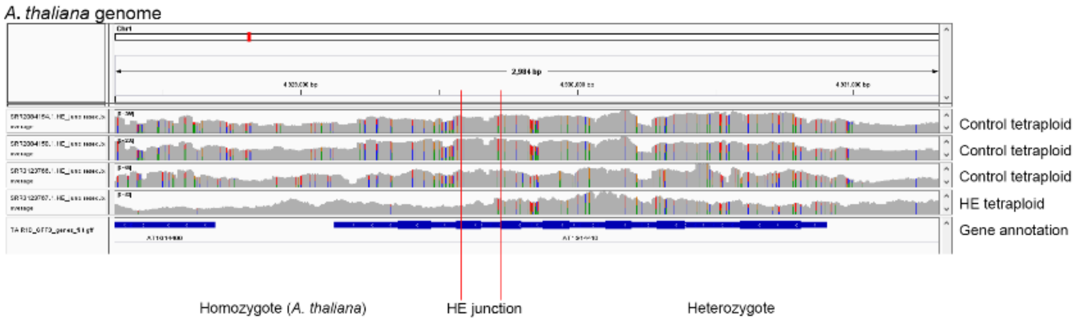
Fig. S17. IGV panel of HE junction regions in *Arabidopsis suecica* (9 HE junctions). The panel shows the genotypes (homozygote or heterozygote) around HE junction (region between red lines) based on SNP genotyping.



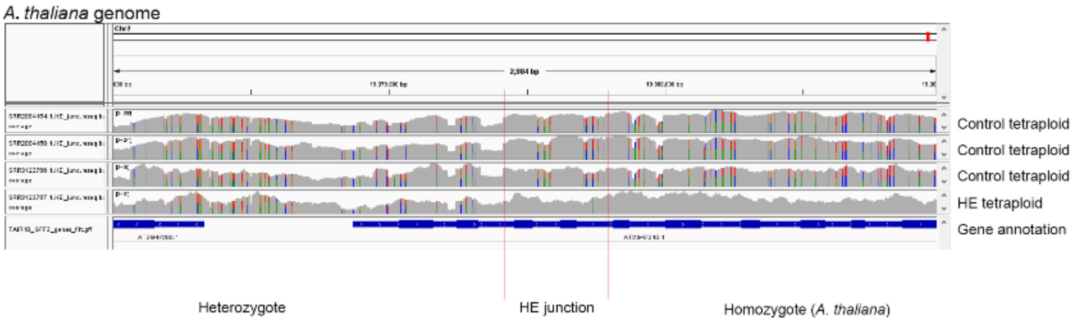
HE junction 2.



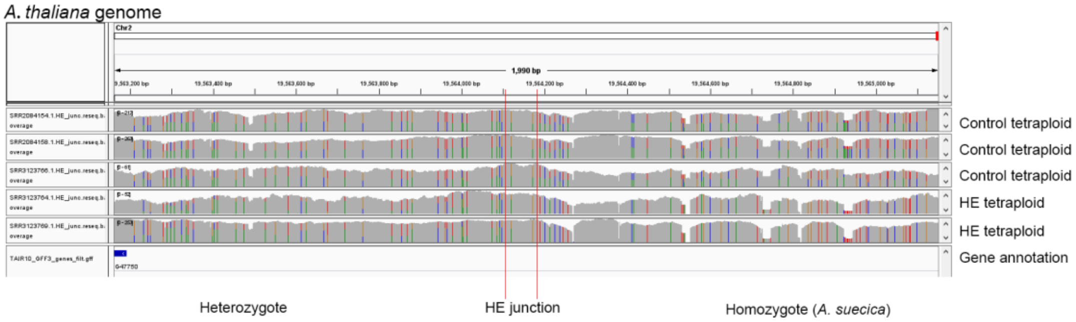
HE junction 3.



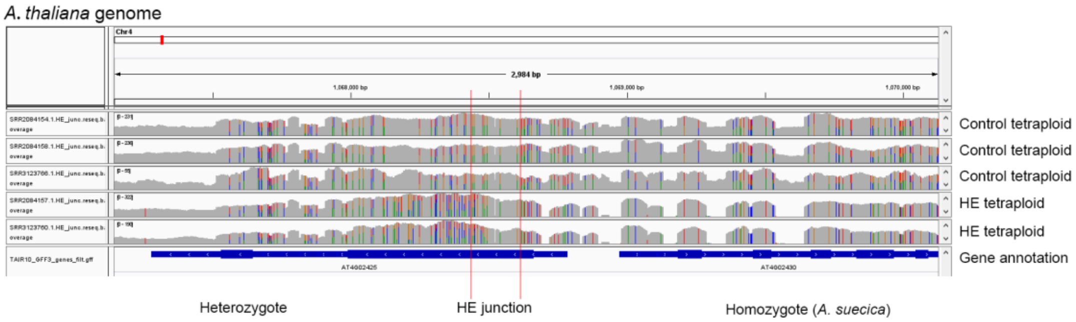
HE junction 4.



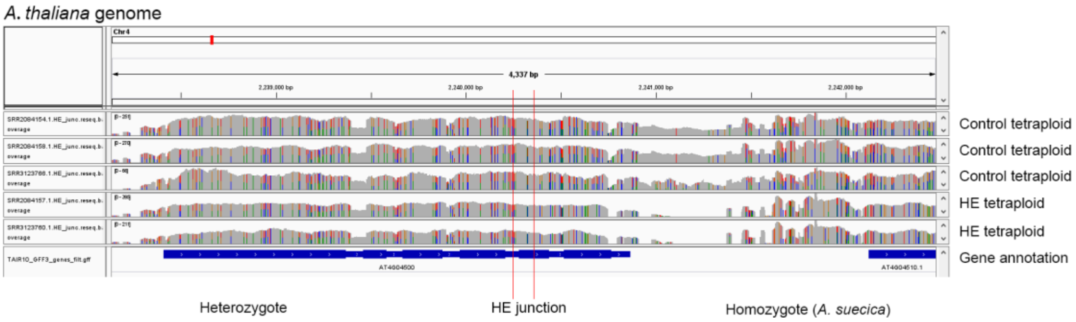
HE junction 5.



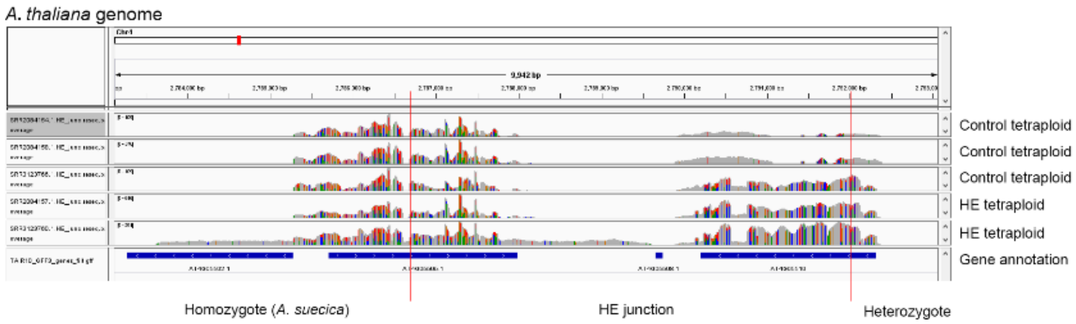
HE junction 6.



HE junction 7.



HE junction 8.



HE junction 9.

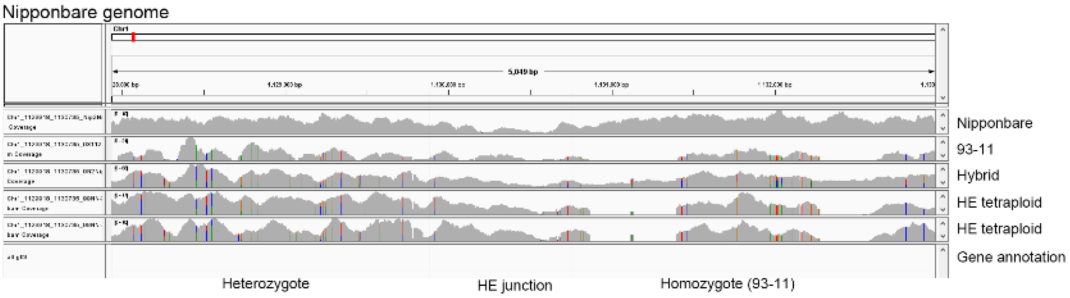
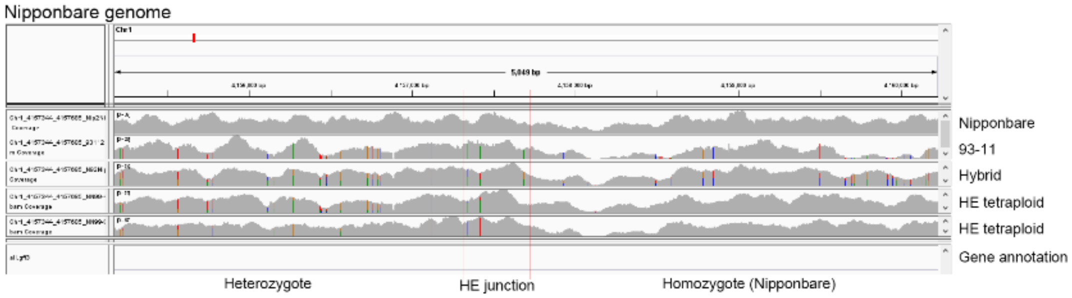
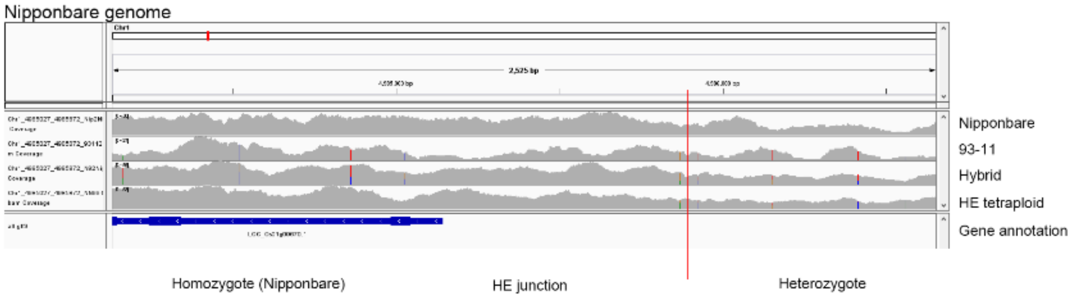


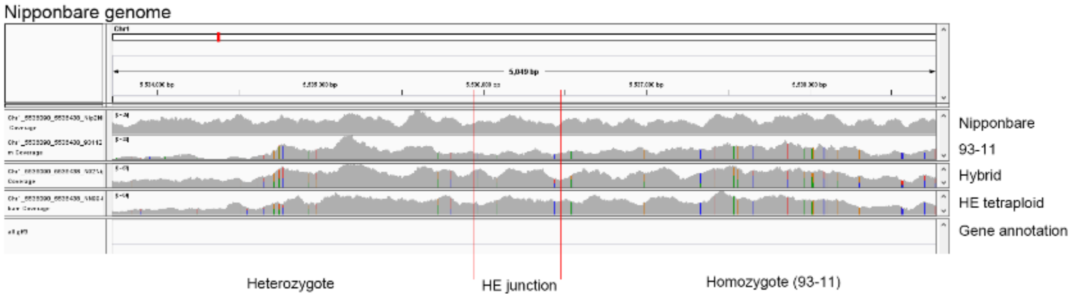
Fig. S18. IGV panel of HE junction regions in synthetic tetraploid rice (166 HE junctions). The panel shows the genotypes (homozygote or heterozygote) around HE junction (region between red lines) based on SNP genotyping between (Nipponbare and 93-11).



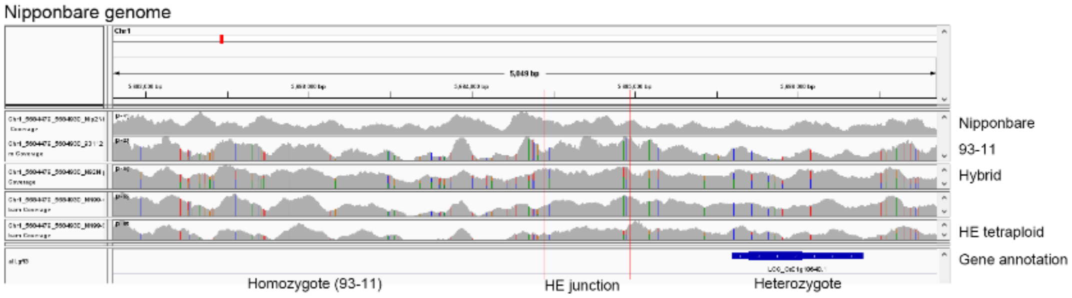
HE junction 2.



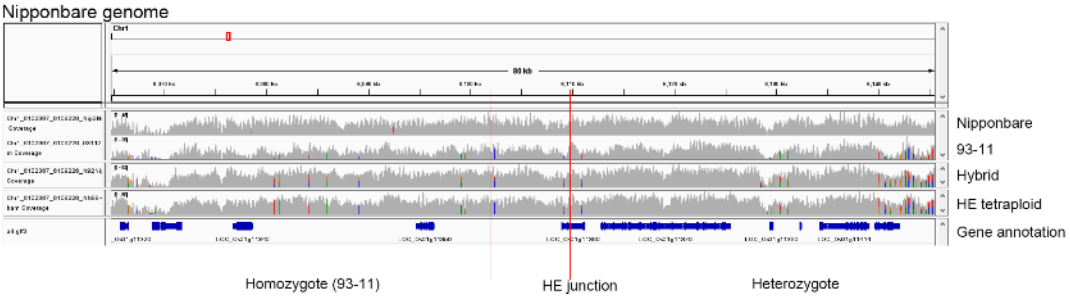
HE junction 3.



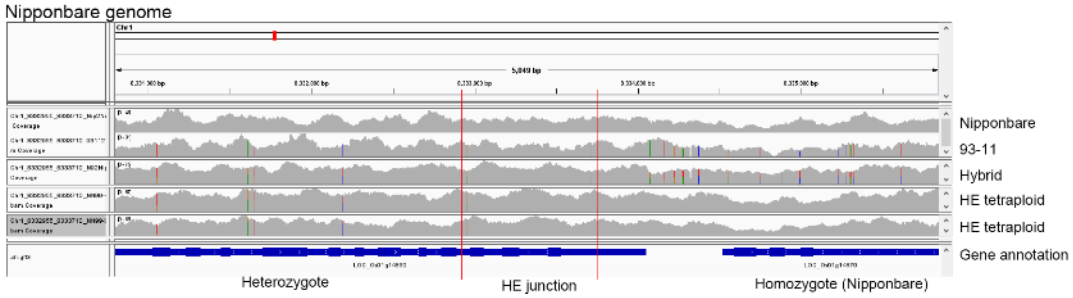
HE junction 4.



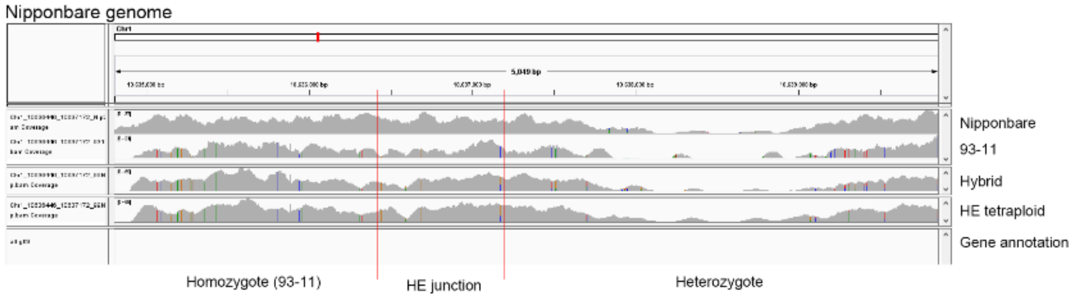
HE junction 5.



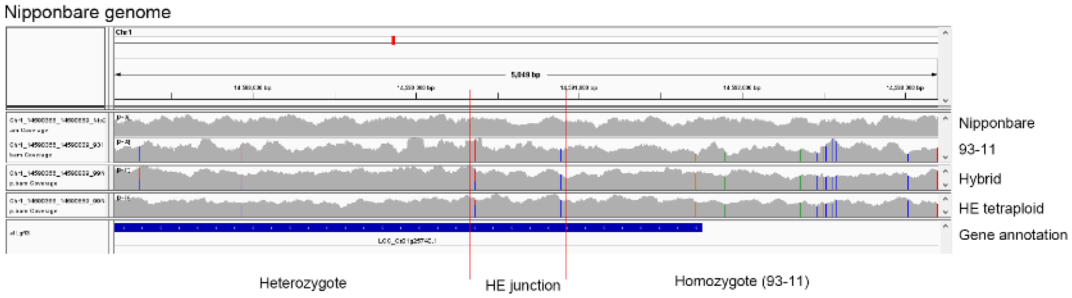
HE junction 6.



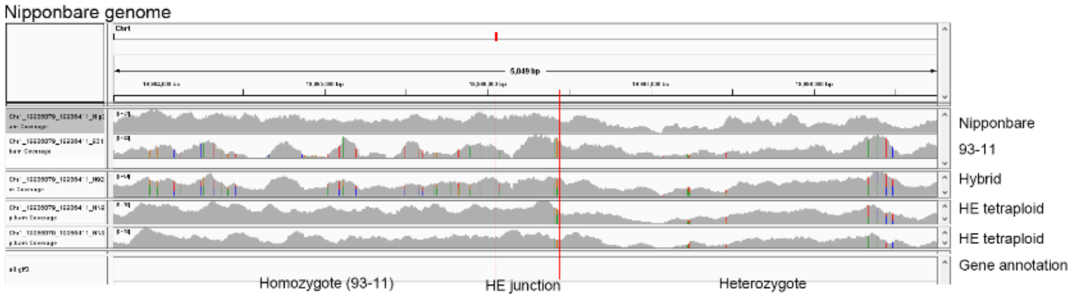
HE junction 7.



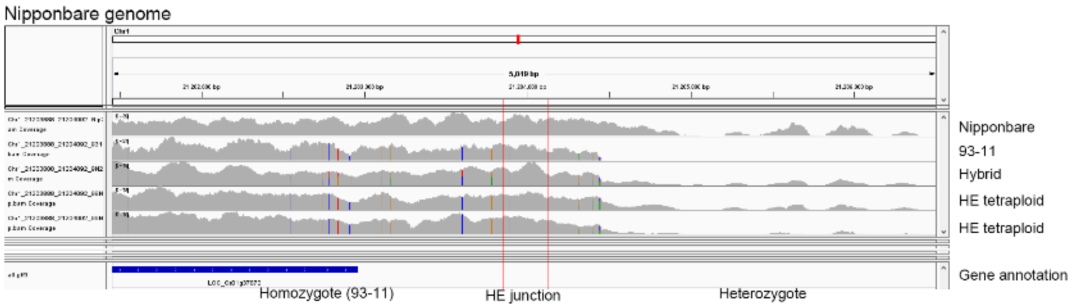
HE junction 8.



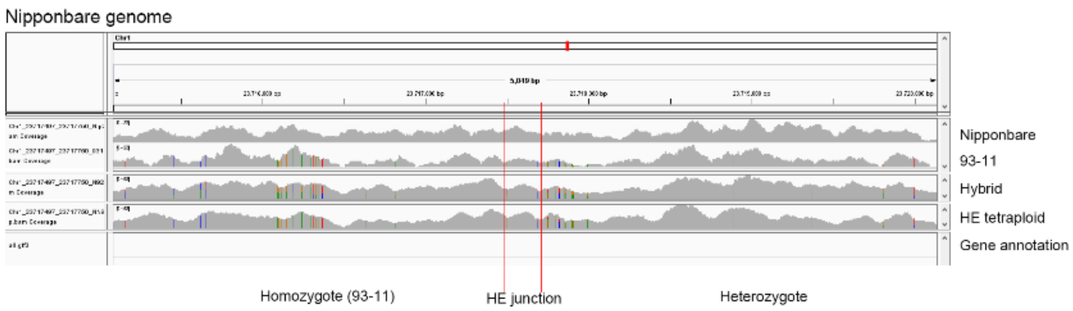
HE junction 9.



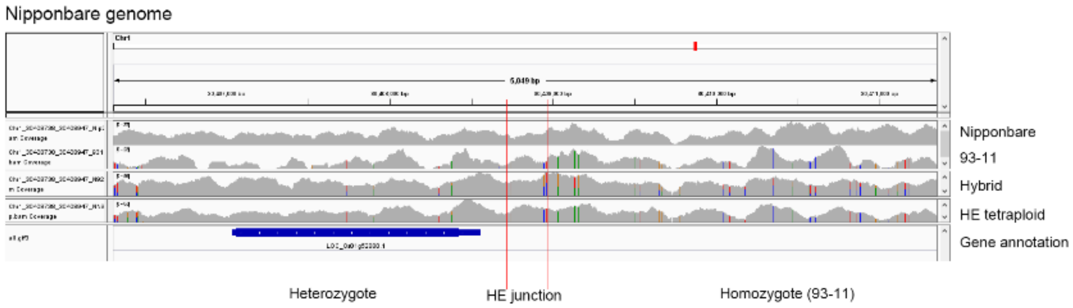
HE junction 10.



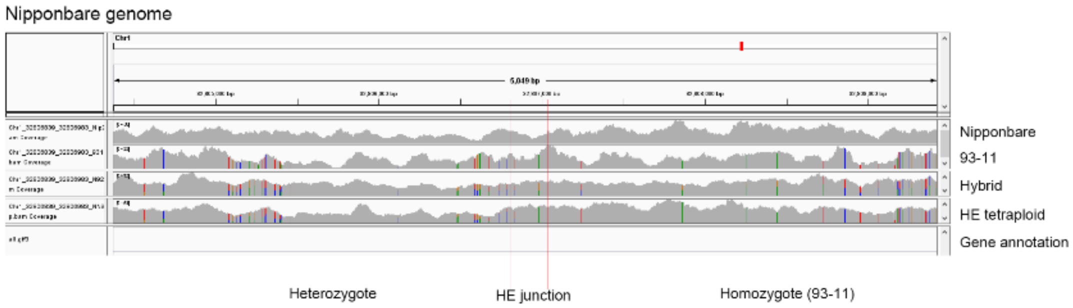
HE junction 11.



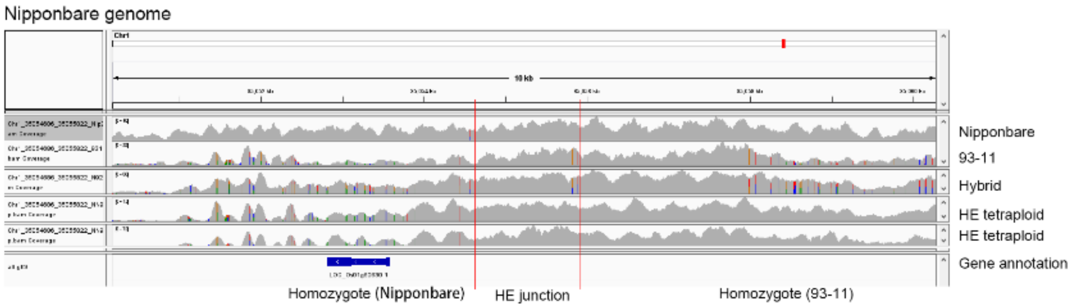
HE junction 12.



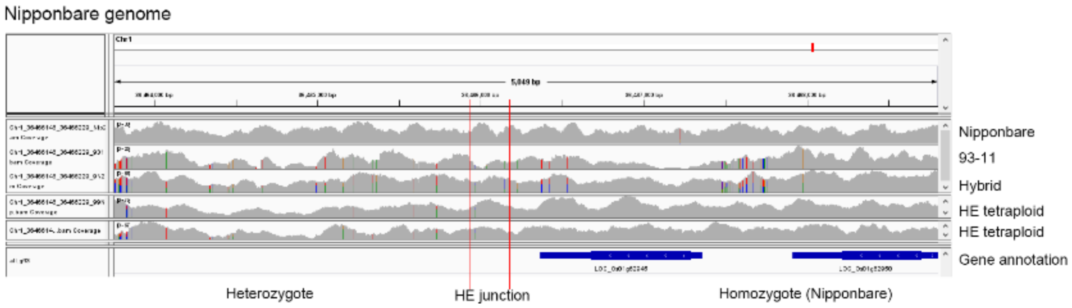
HE junction 13.



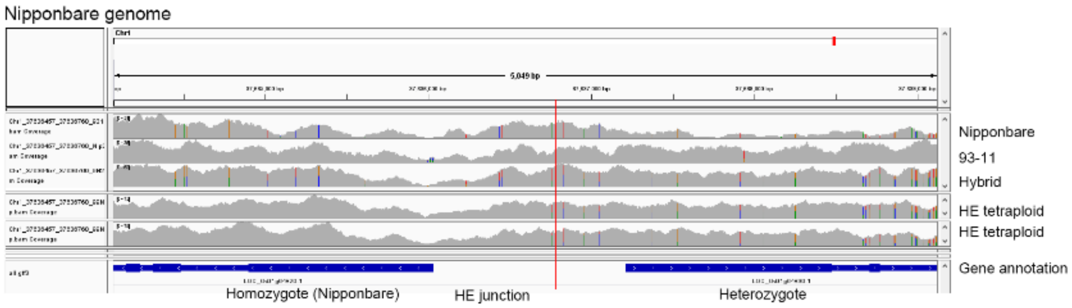
HE junction 14.



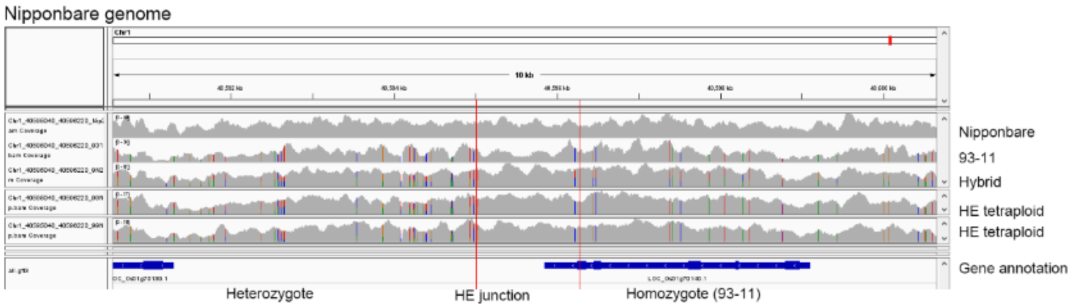
HE junction 15.



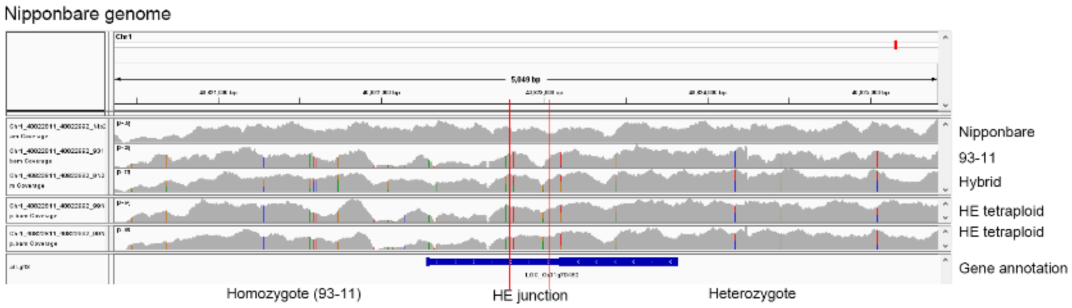
HE junction 16.



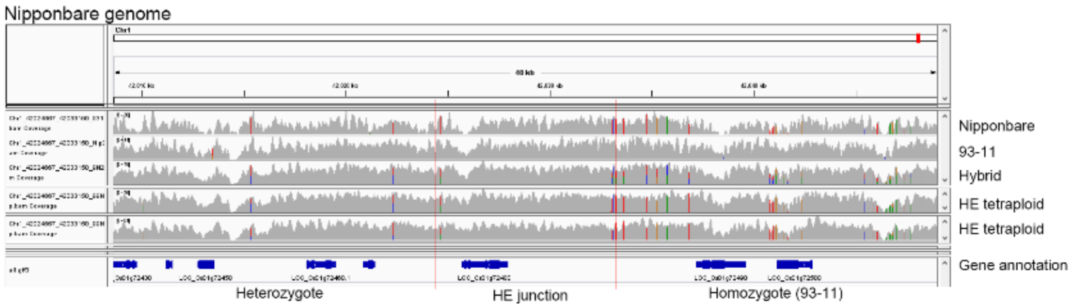
HE junction 17.



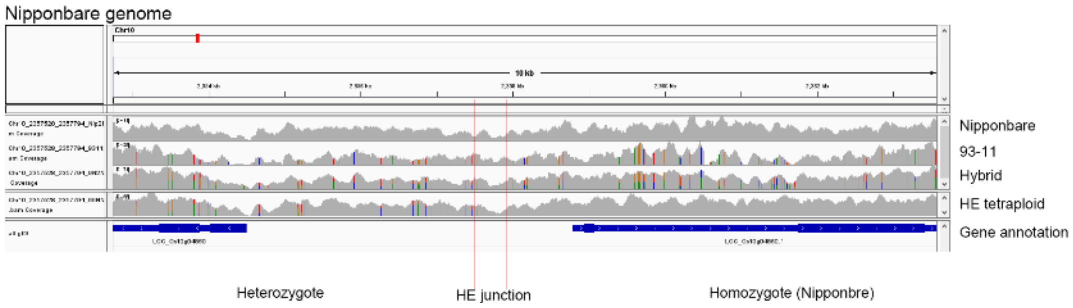
HE junction 18.



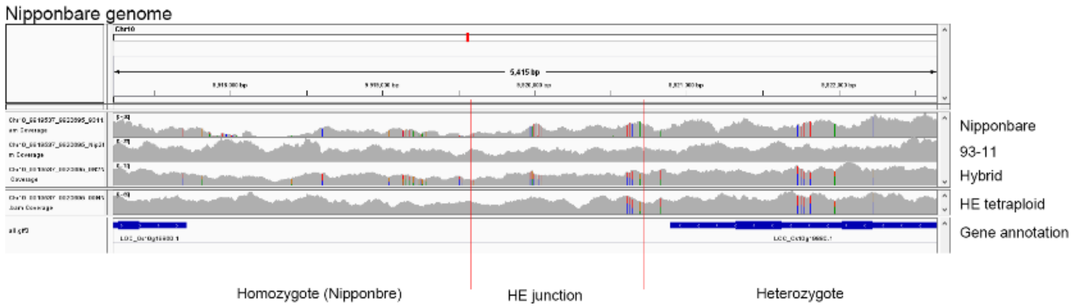
HE junction 19.



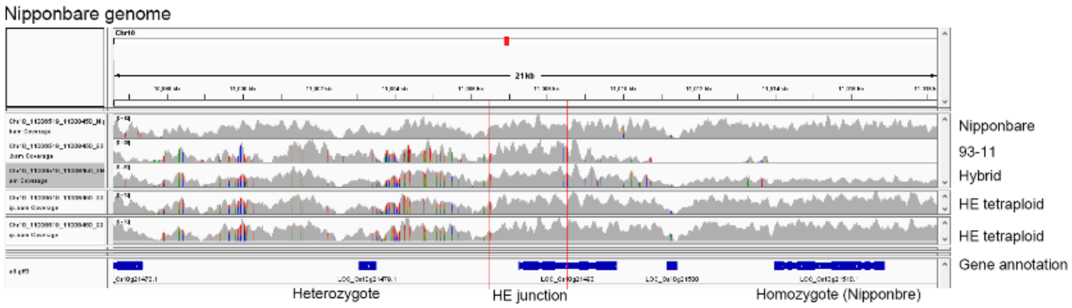
HE junction 20.



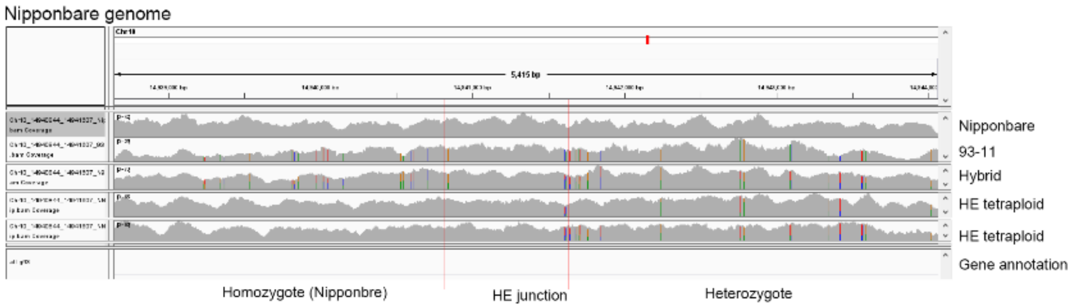
HE junction 21.



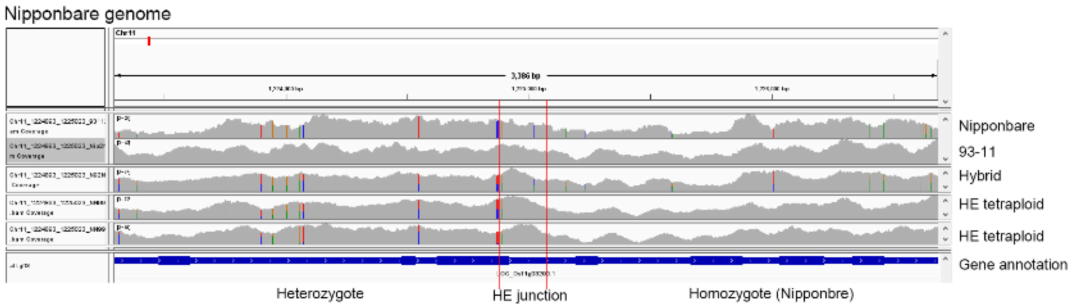
HE junction 22.



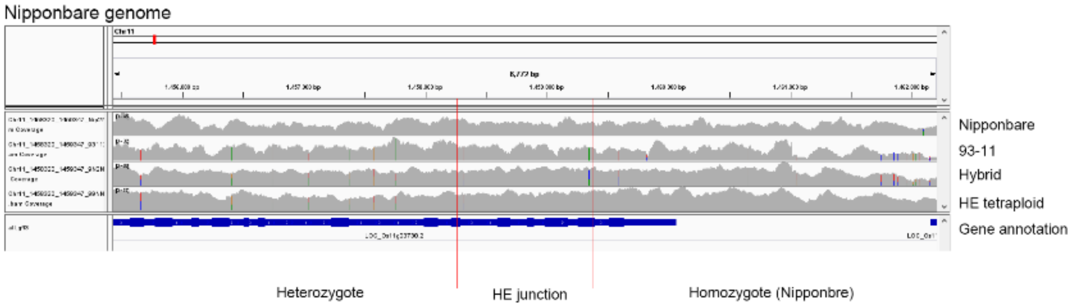
HE junction 23.



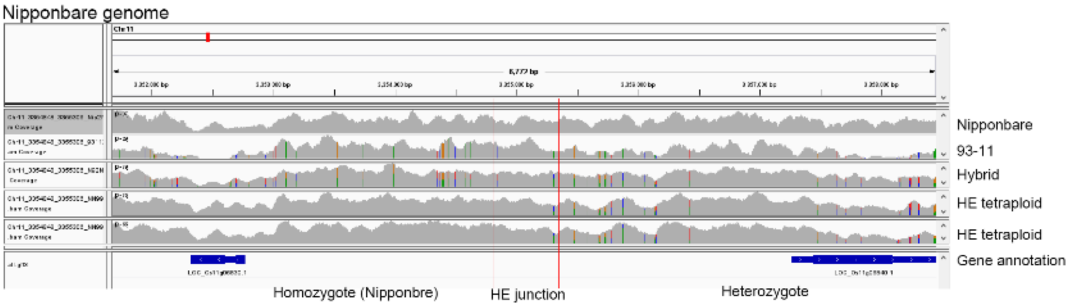
HE junction 24.



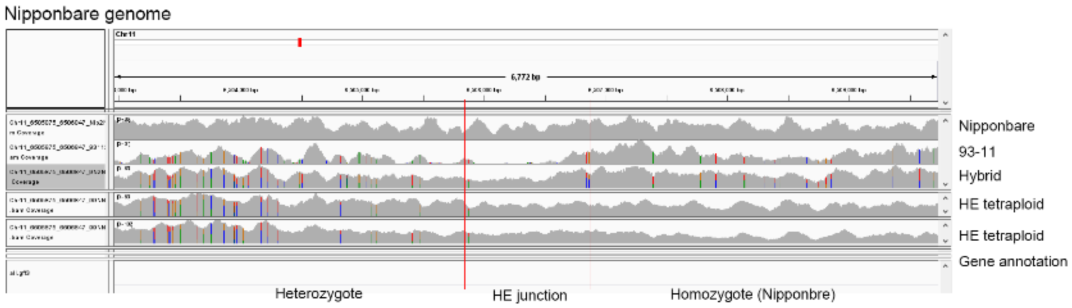
HE junction 25.



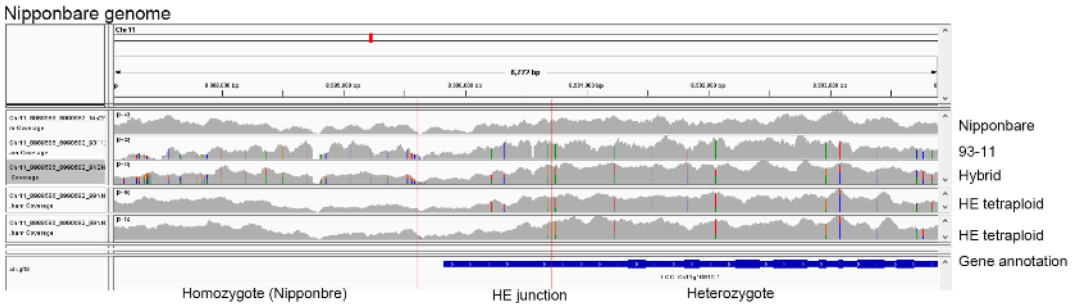
HE junction 26.



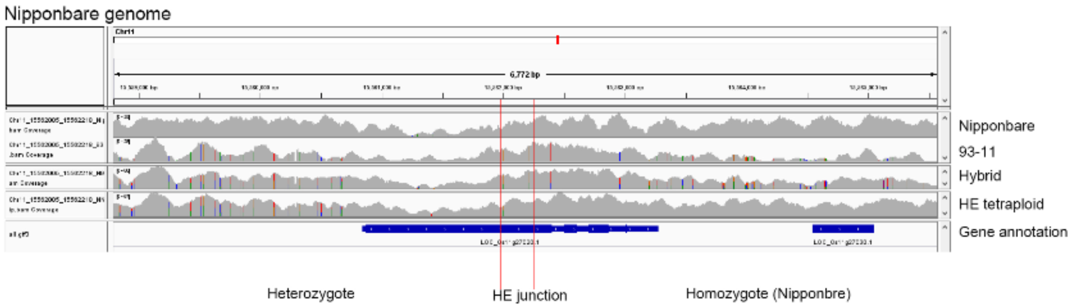
HE junction 27.



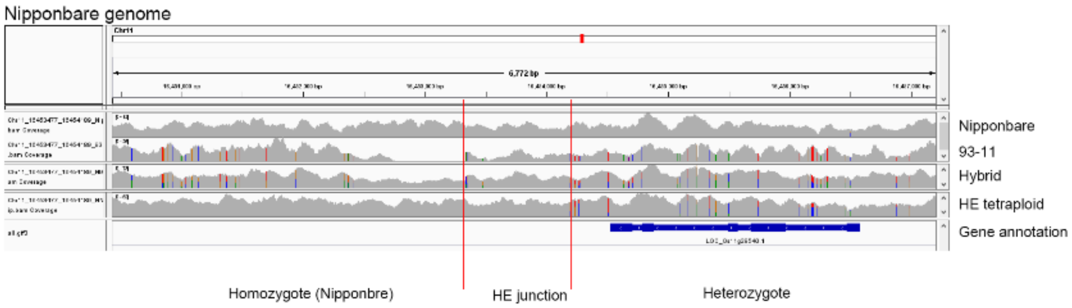
HE junction 28.



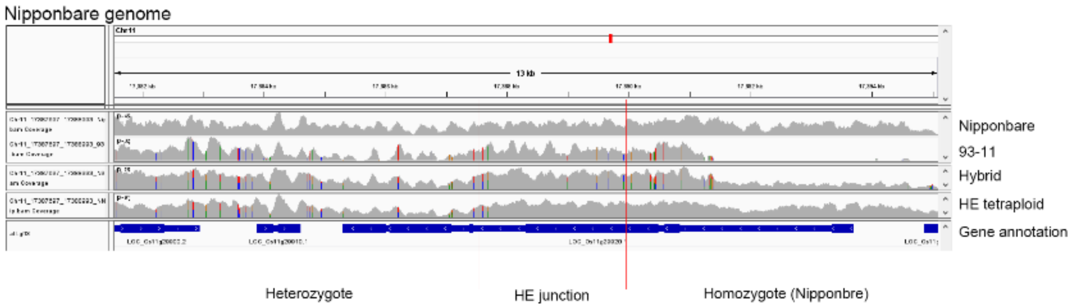
HE junction 29.



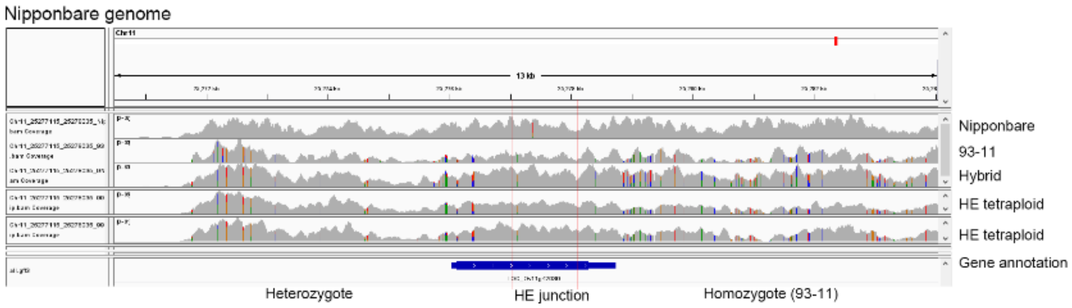
HE junction 30.



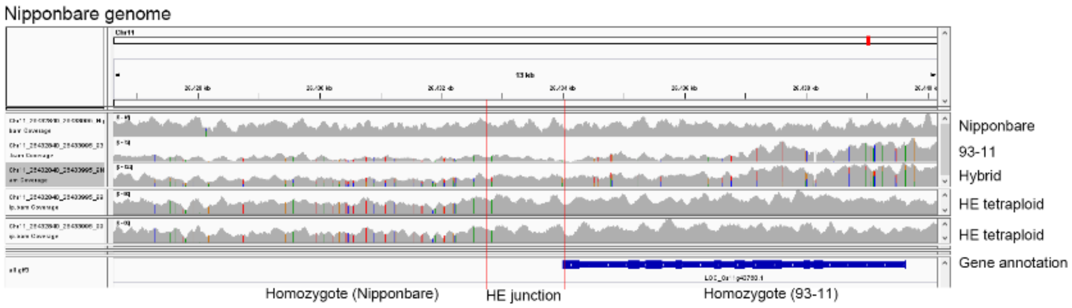
HE junction 31.



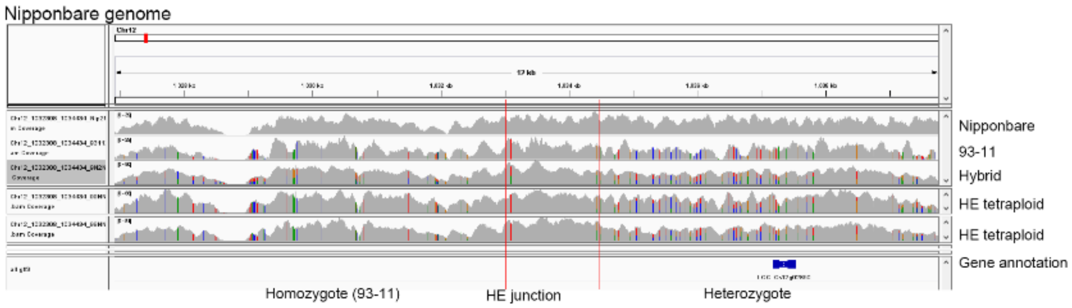
HE junction 32.



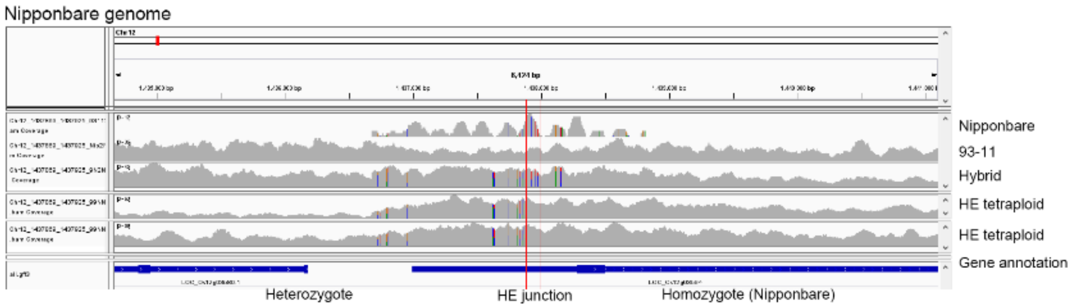
HE junction 33.



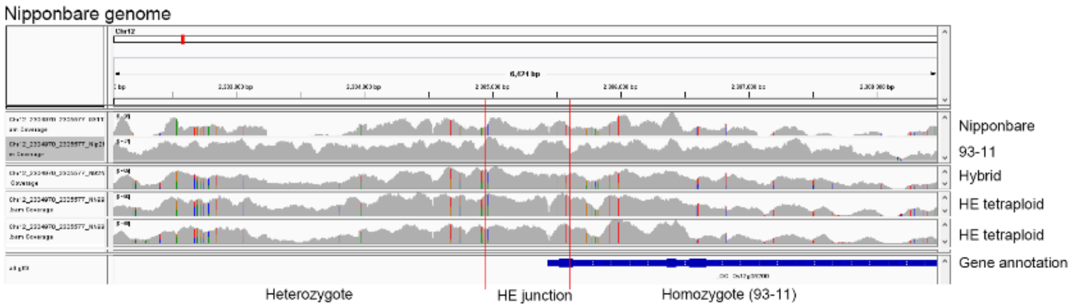
HE junction 34.



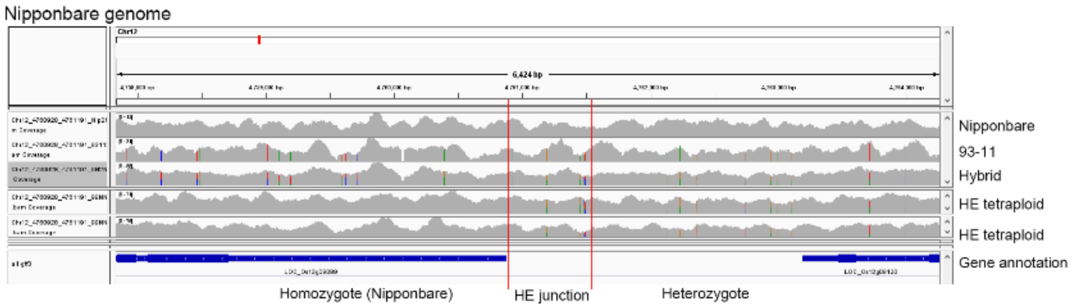
HE junction 35.



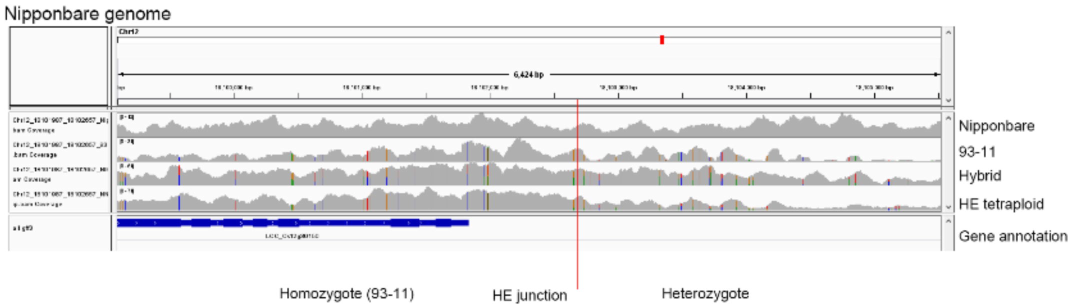
HE junction 36.



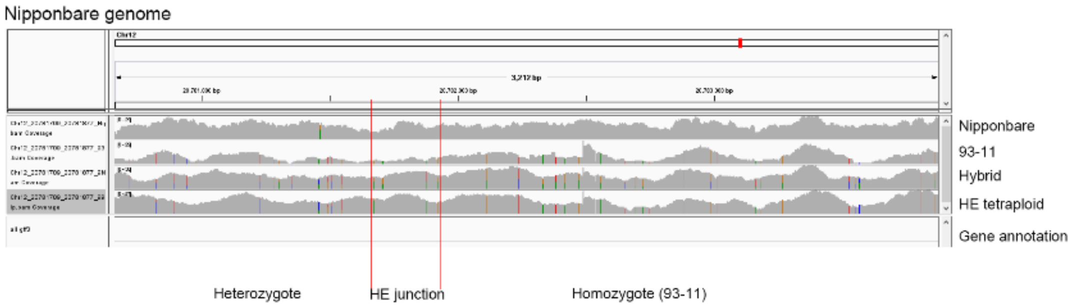
HE junction 37.



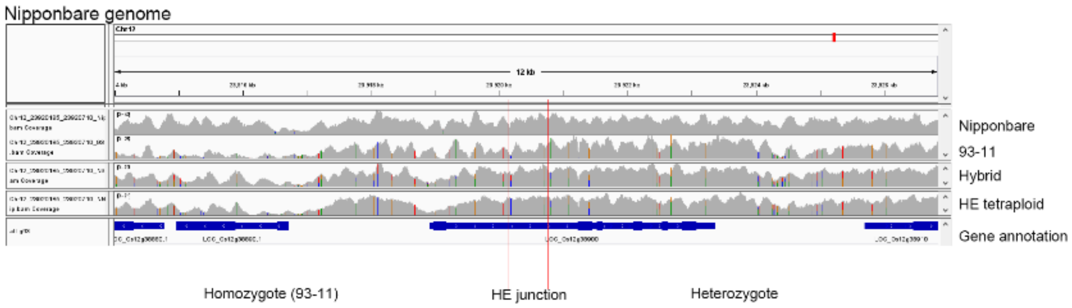
HE junction 38.



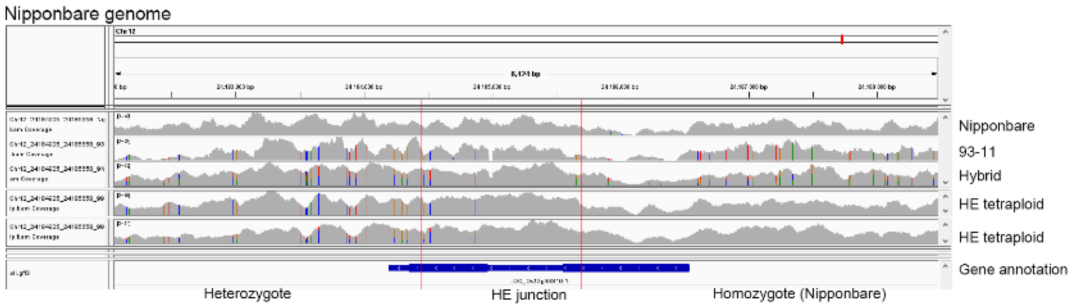
HE junction 39.



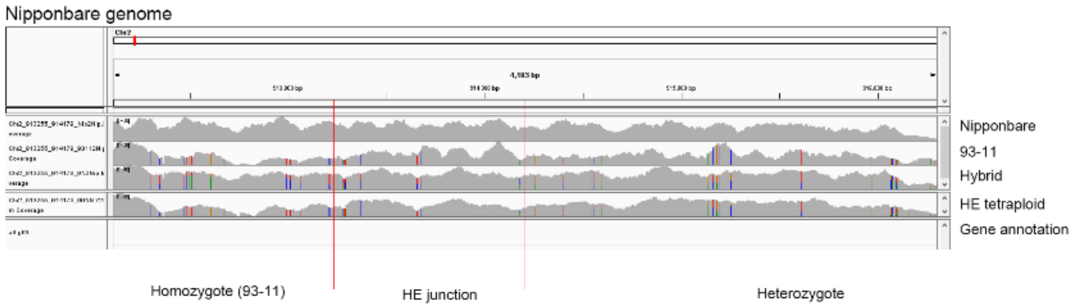
HE junction 40.



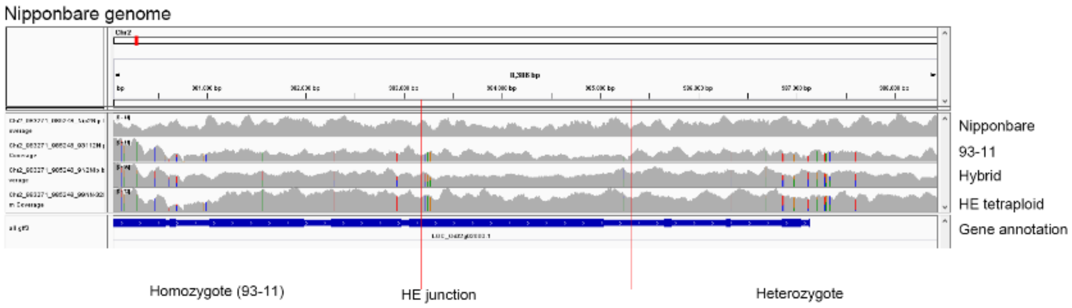
HE junction 41.



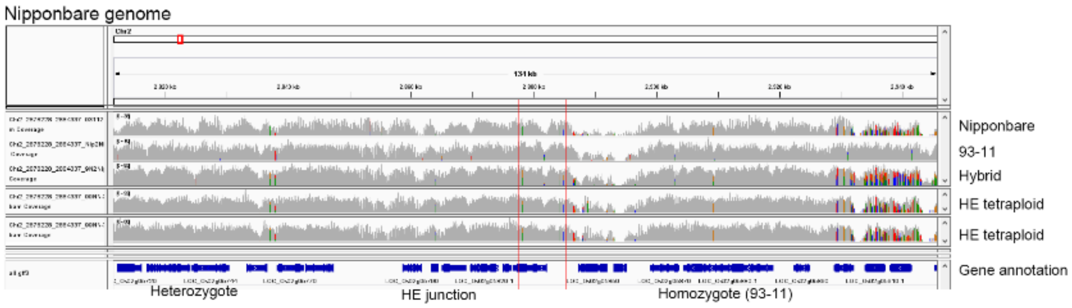
HE junction 42.



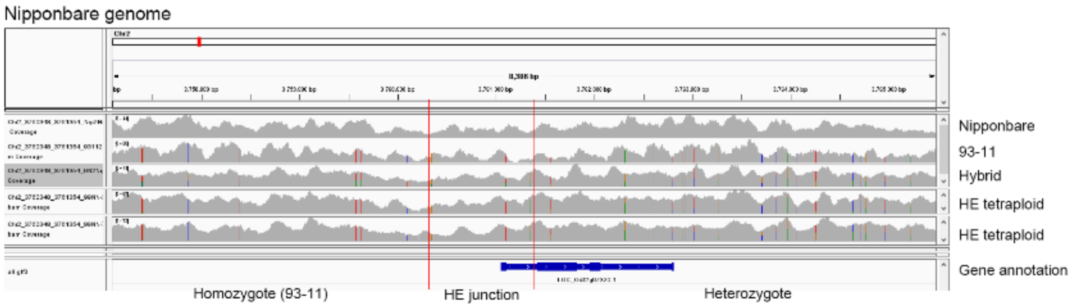
HE junction 43.



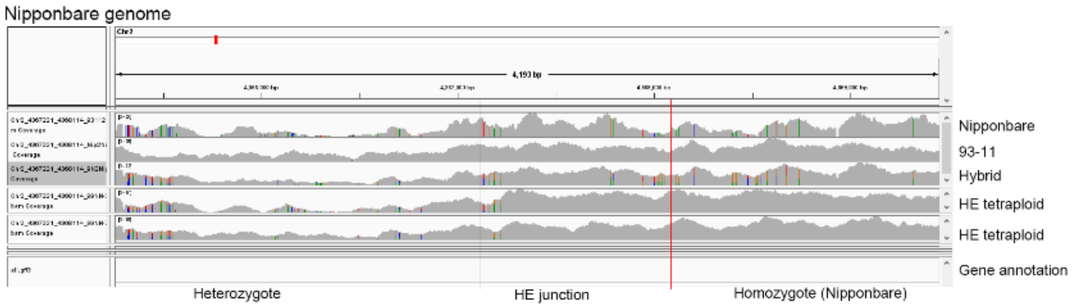
HE junction 44.



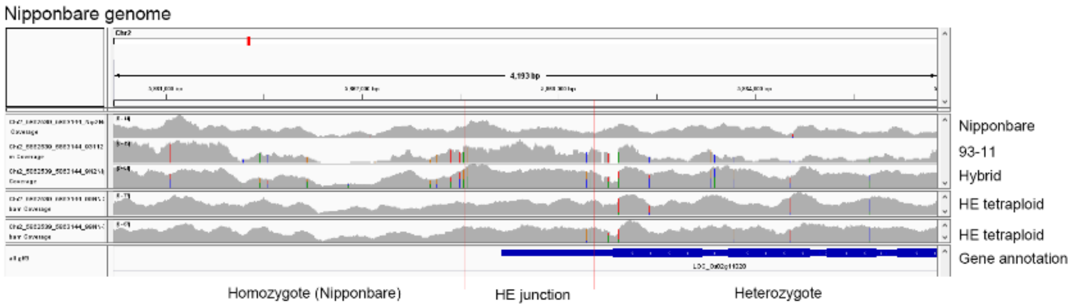
HE junction 45.



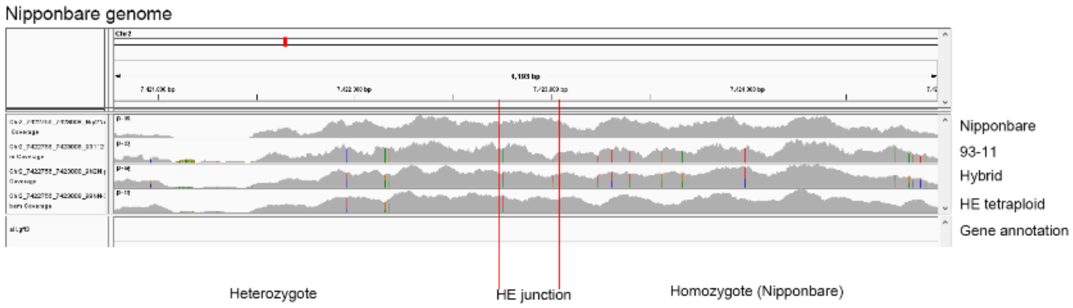
HE junction 46.



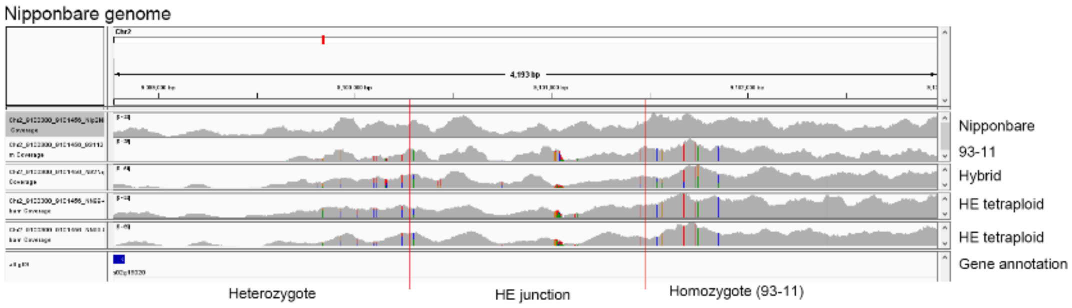
HE junction 47.



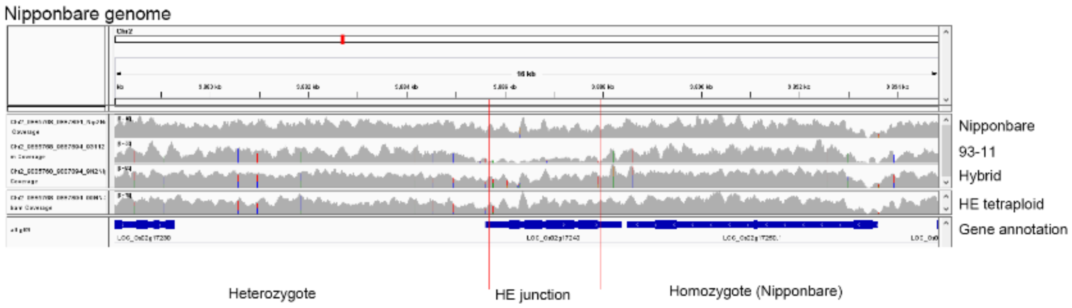
HE junction 48.



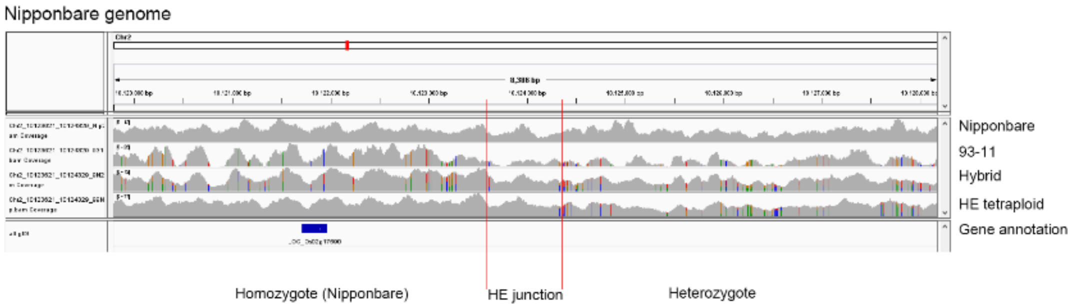
HE junction 49.



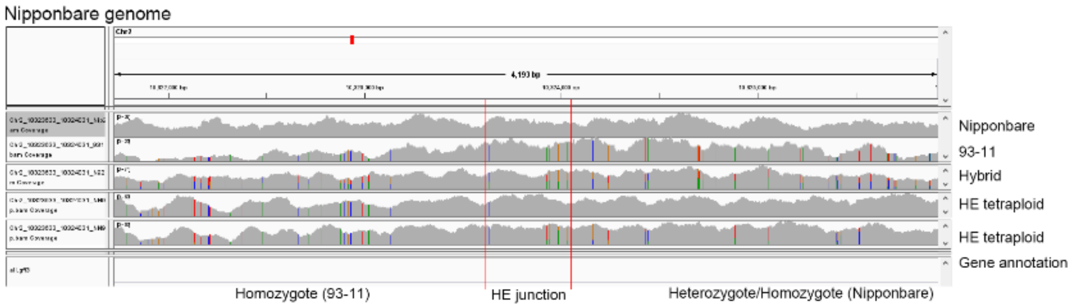
HE junction 50.



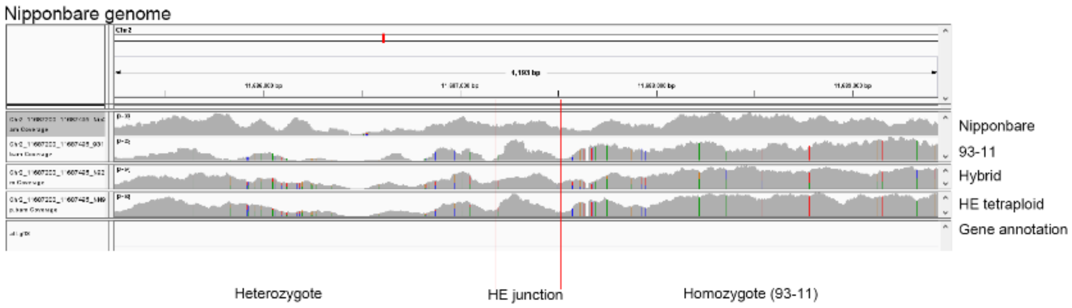
HE junction 51.



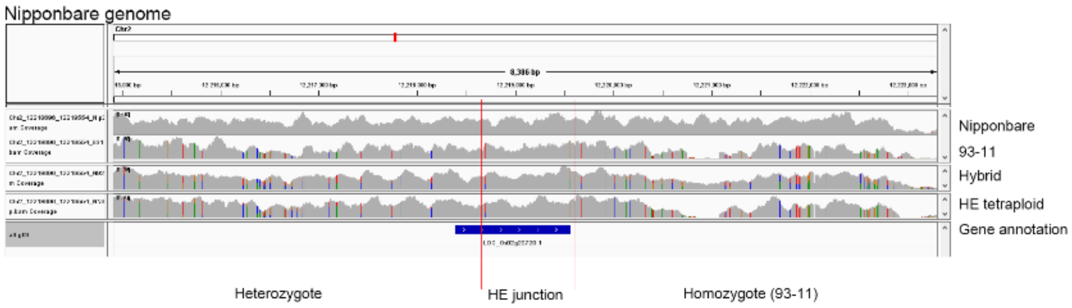
HE junction 52.



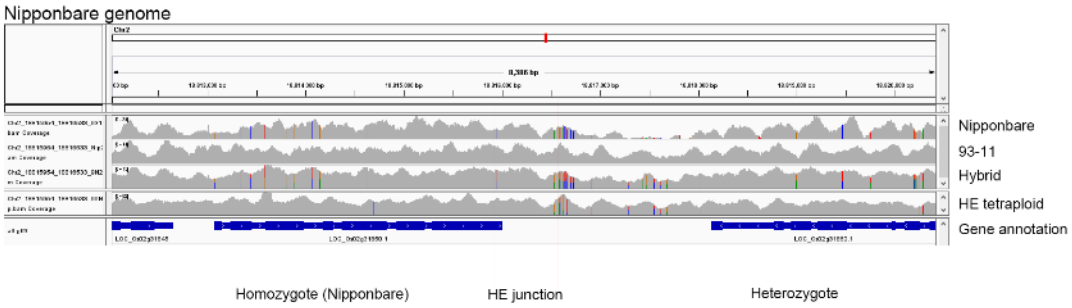
HE junction 53.



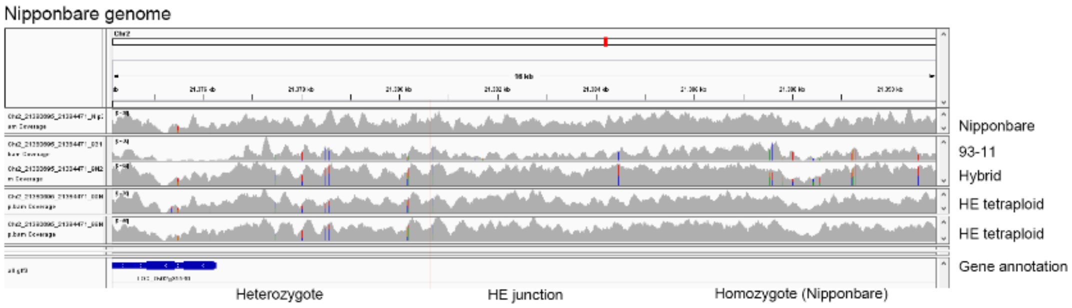
HE junction 54.



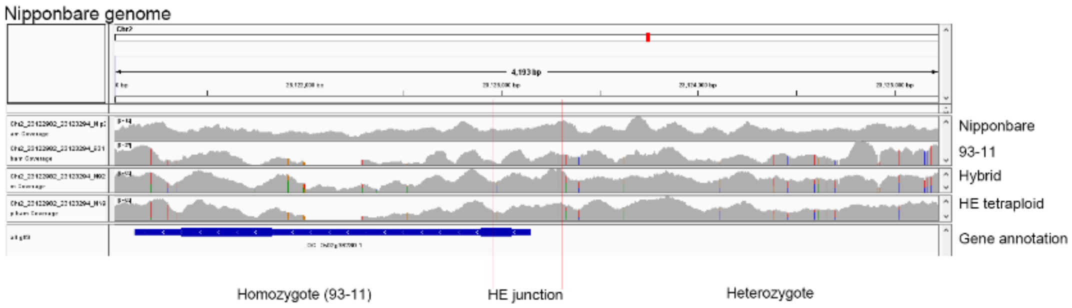
HE junction 55.



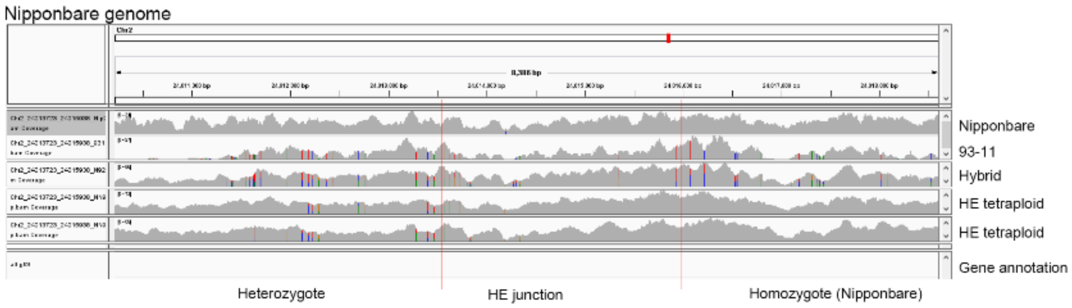
HE junction 56.



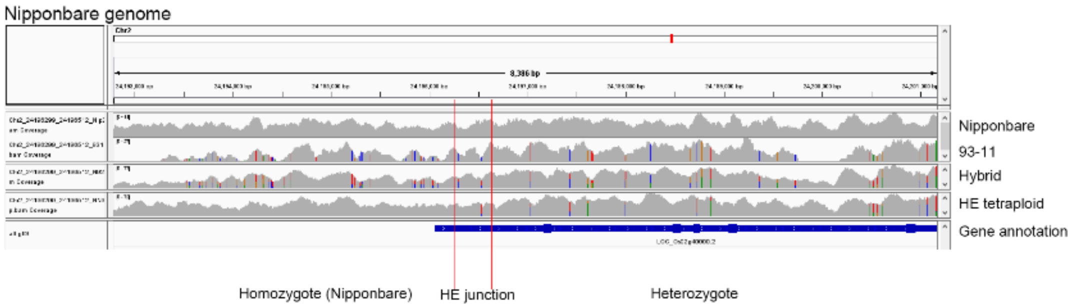
HE junction 57.



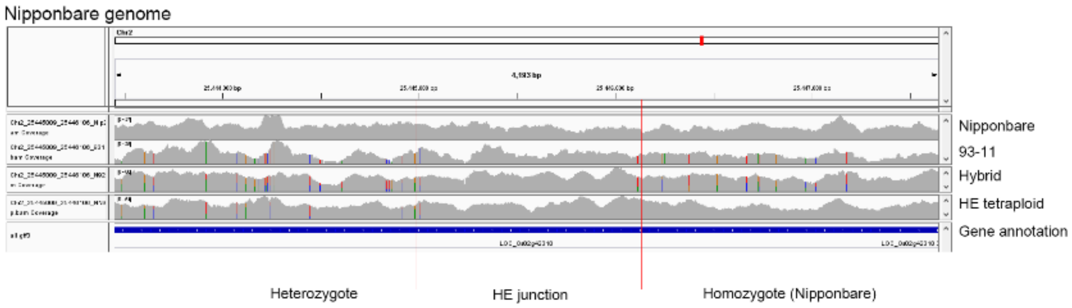
HE junction 58.



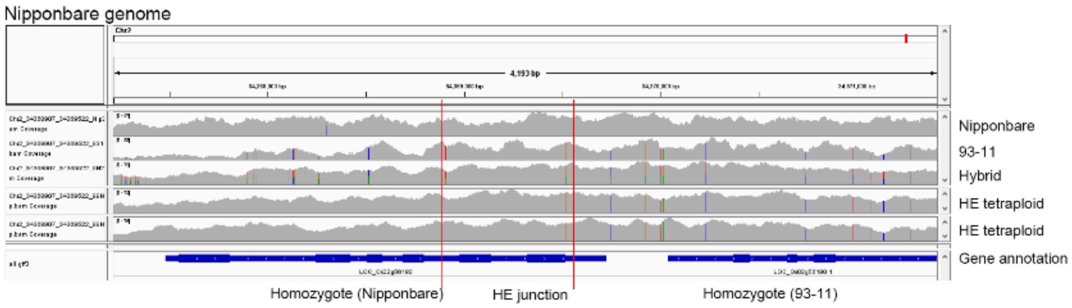
HE junction 59.



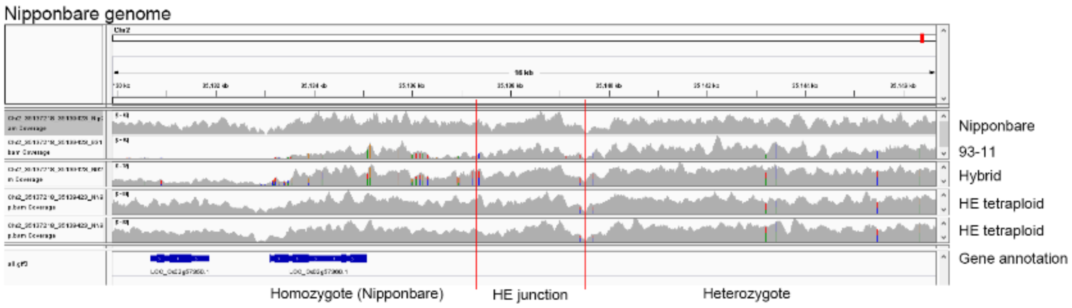
HE junction 60.



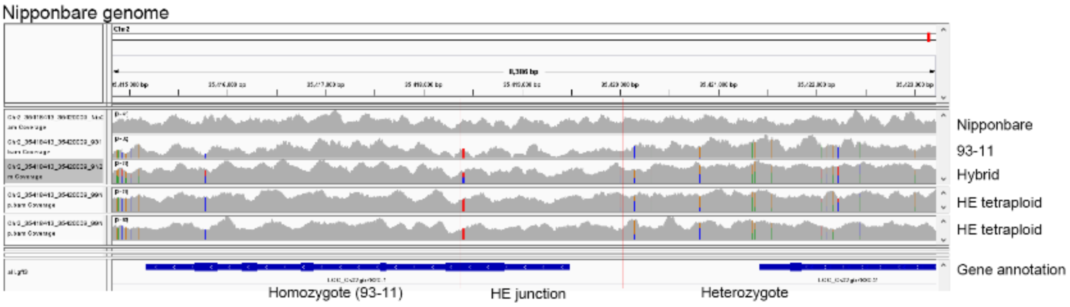
HE junction 61.



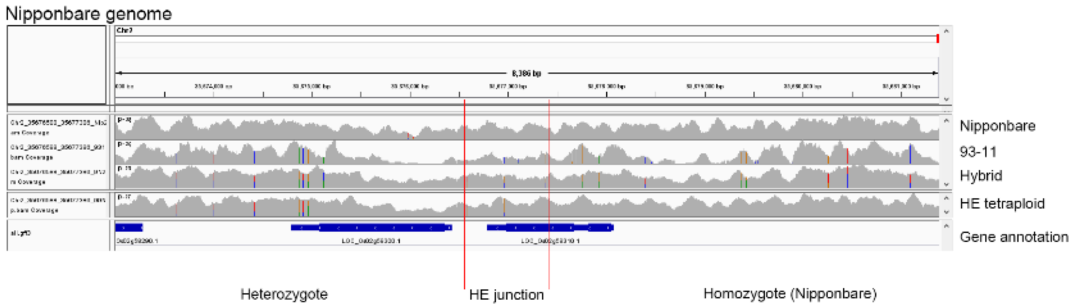
HE junction 62.



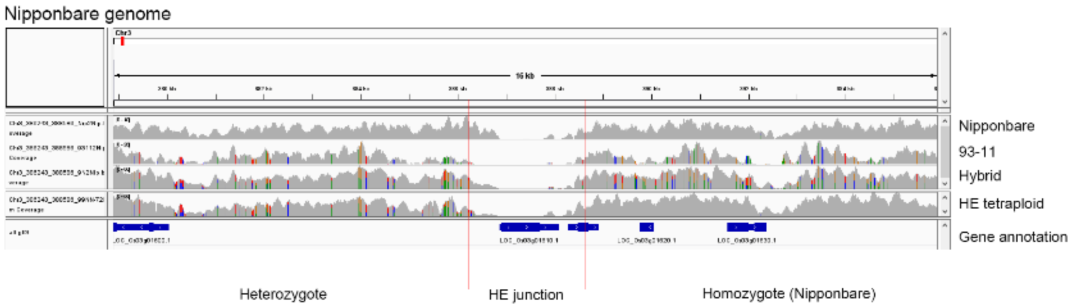
HE junction 63.



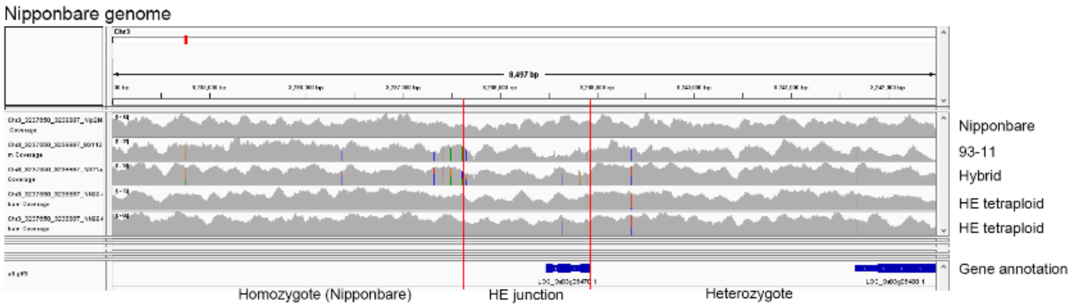
HE junction 64.



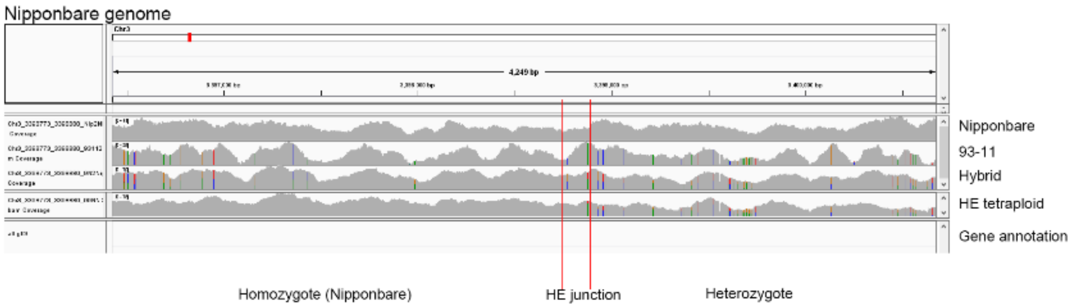
HE junction 65.



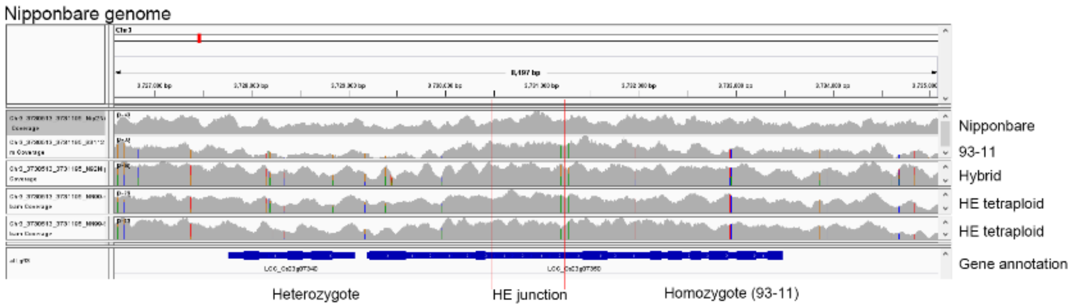
HE junction 66.



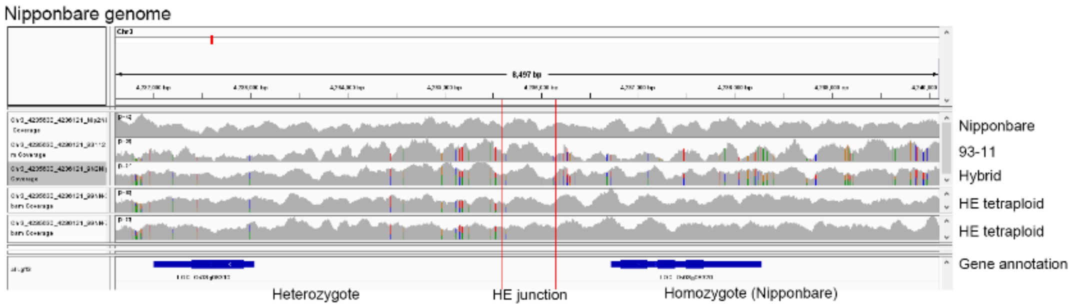
HE junction 67.



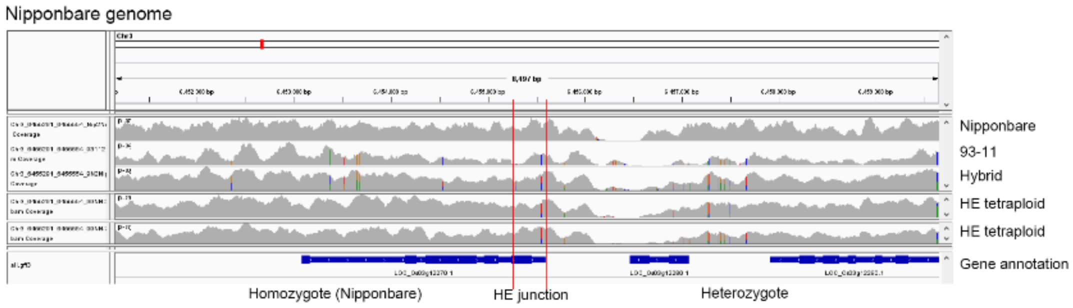
HE junction 68.



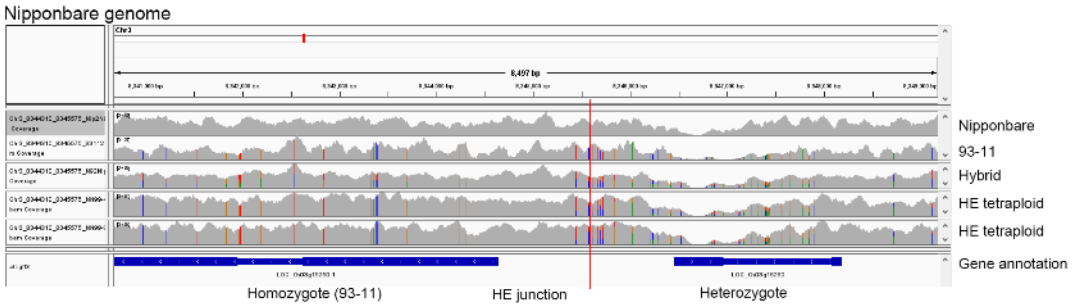
HE junction 69.



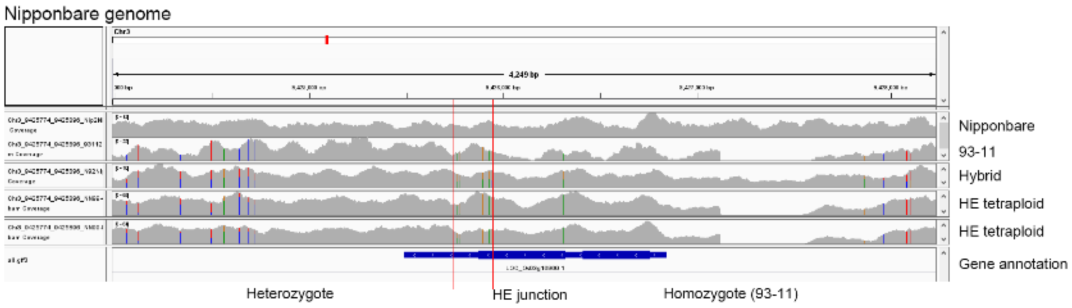
HE junction 70.



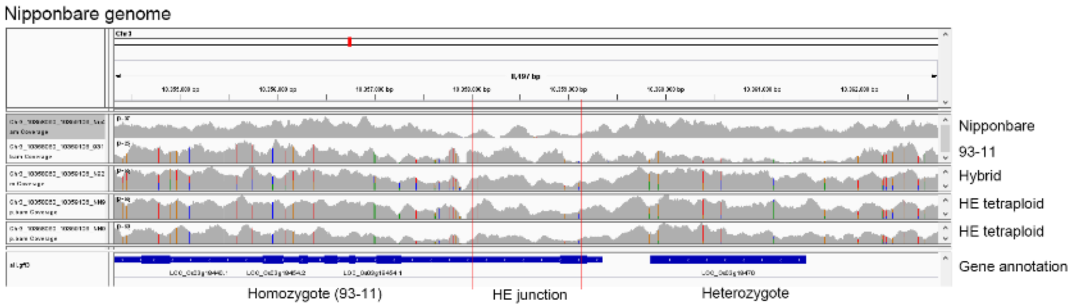
HE junction 71.



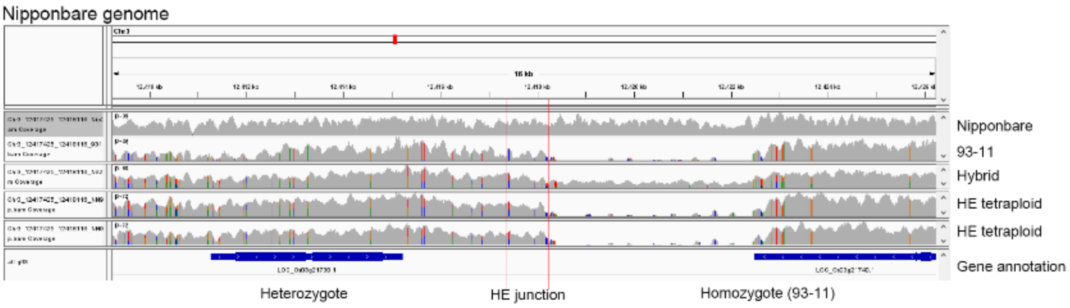
HE junction 72.



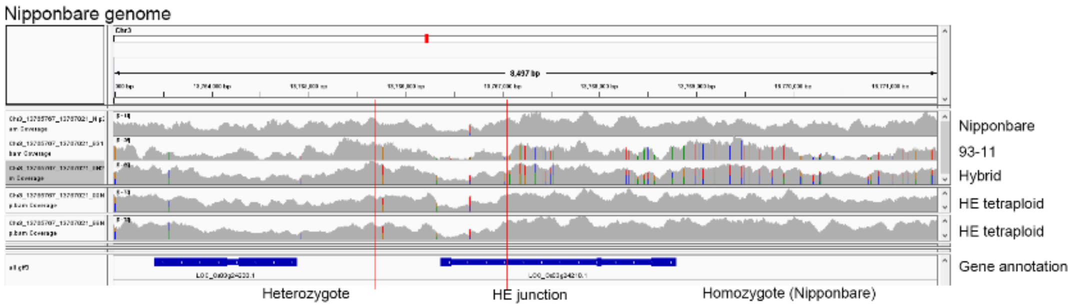
HE junction 73.



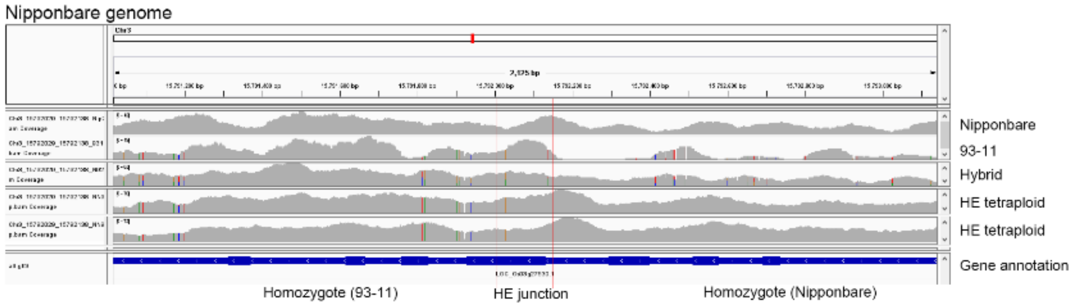
HE junction 74.



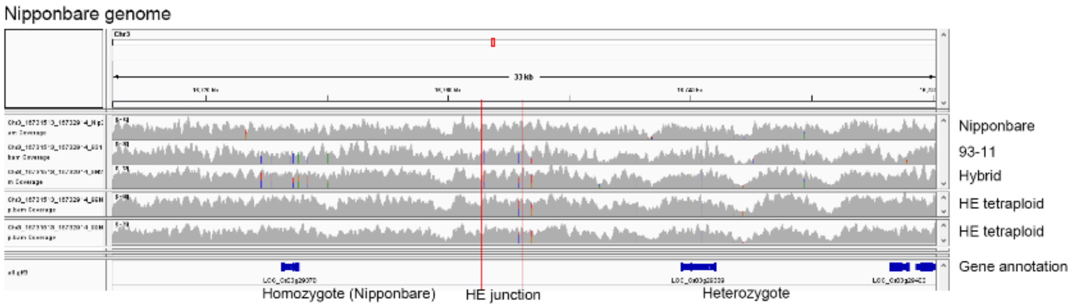
HE junction 75.



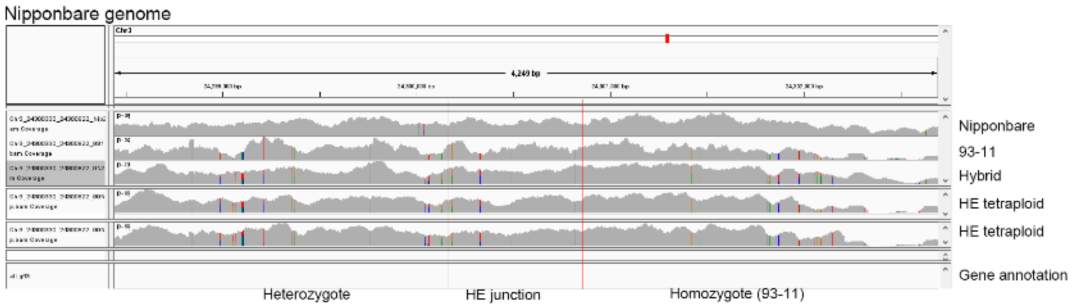
HE junction 76.



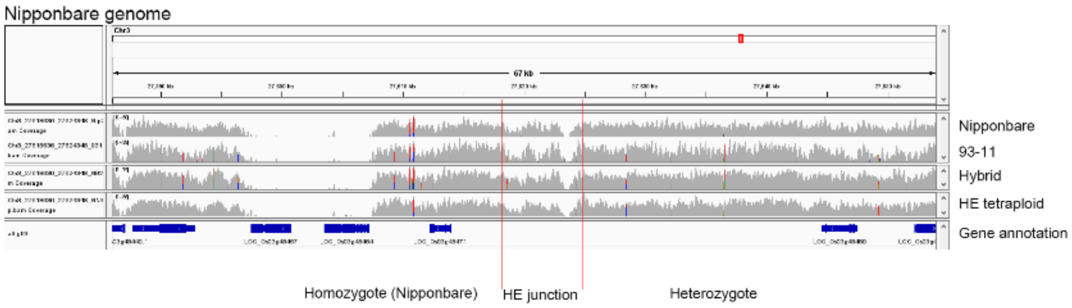
HE junction 77.



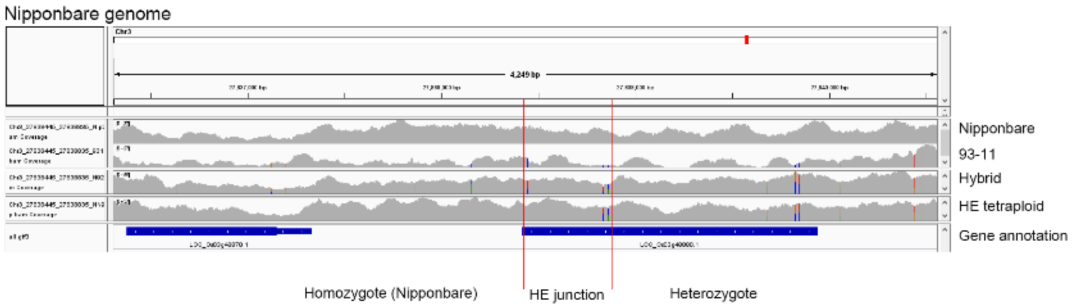
HE junction 78.



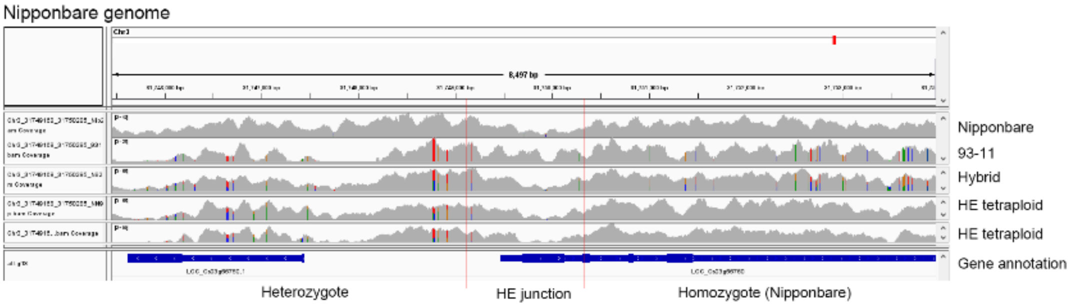
HE junction 79.



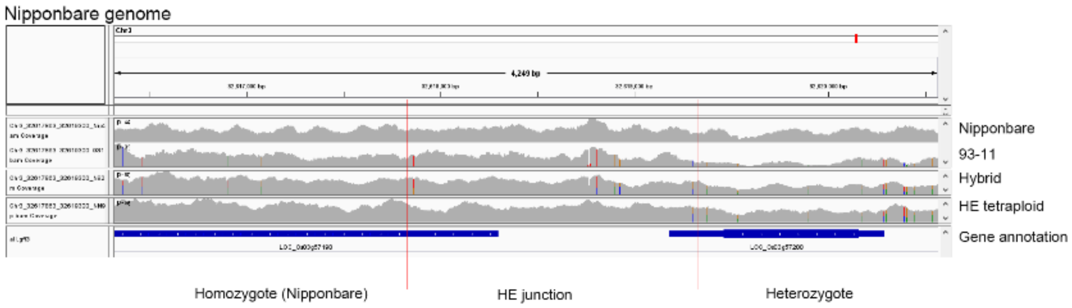
HE junction 80.



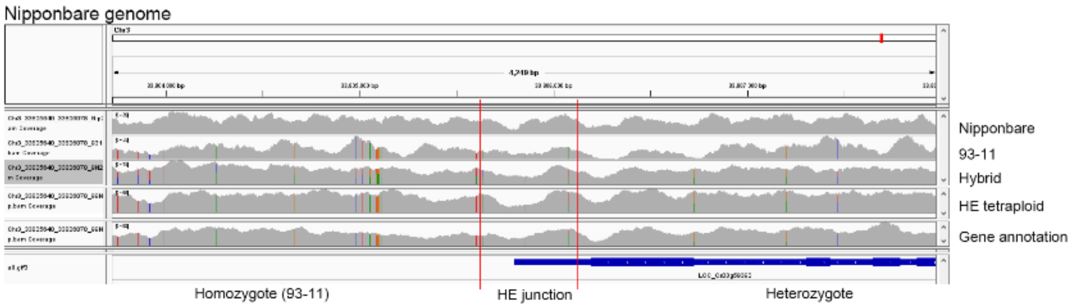
HE junction 81.



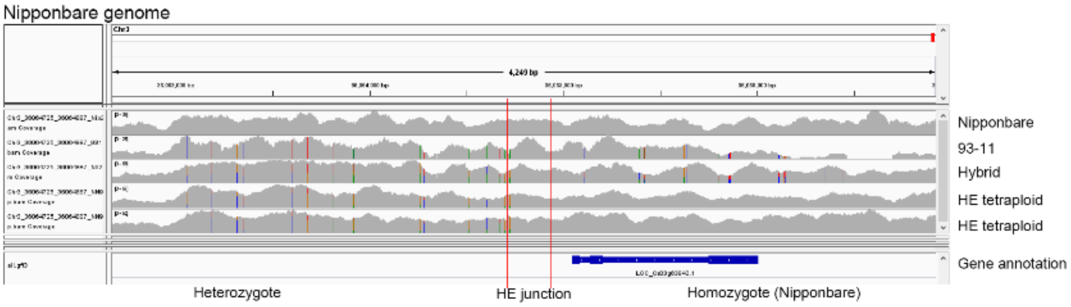
HE junction 82.



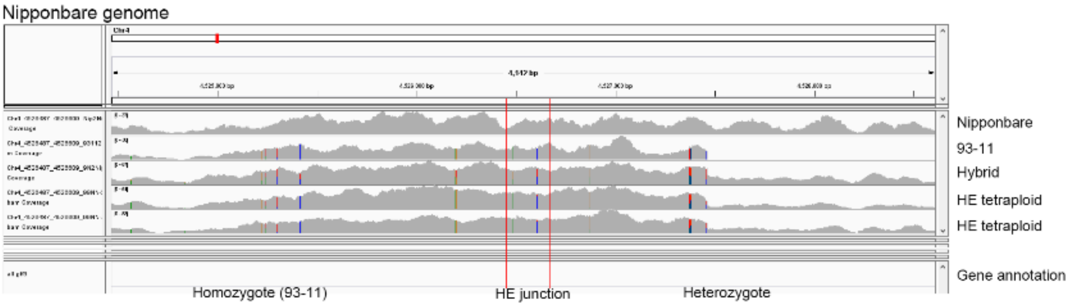
HE junction 83.



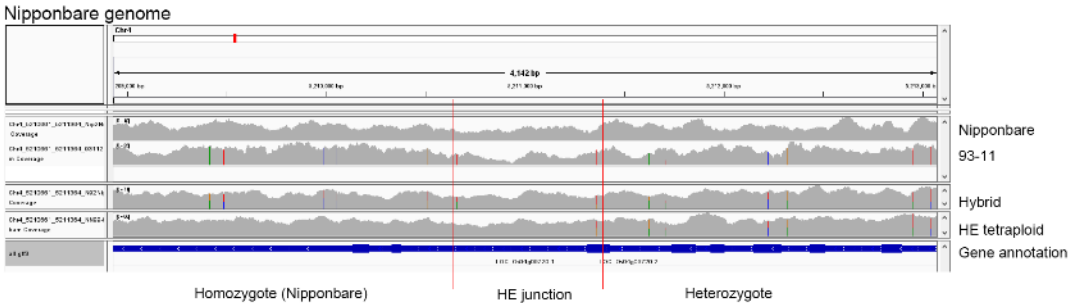
HE junction 84.



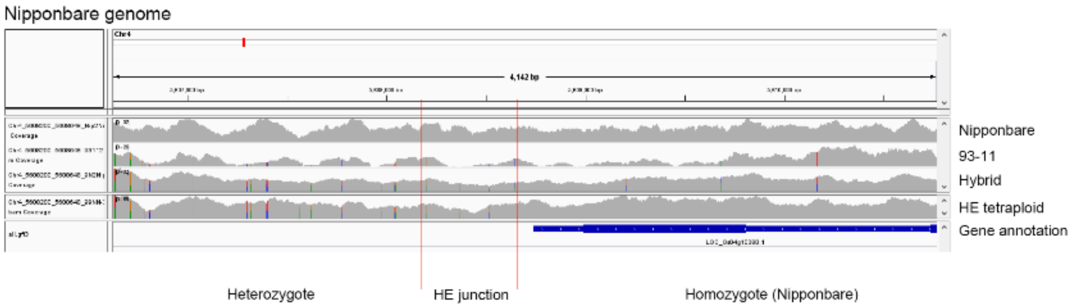
HE junction 85.



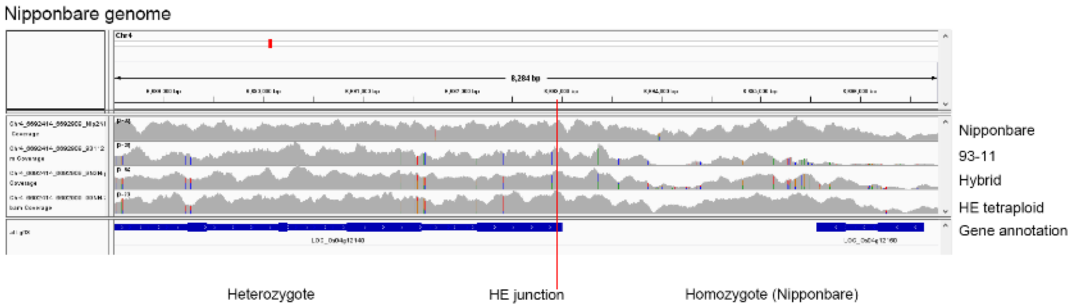
HE junction 87.



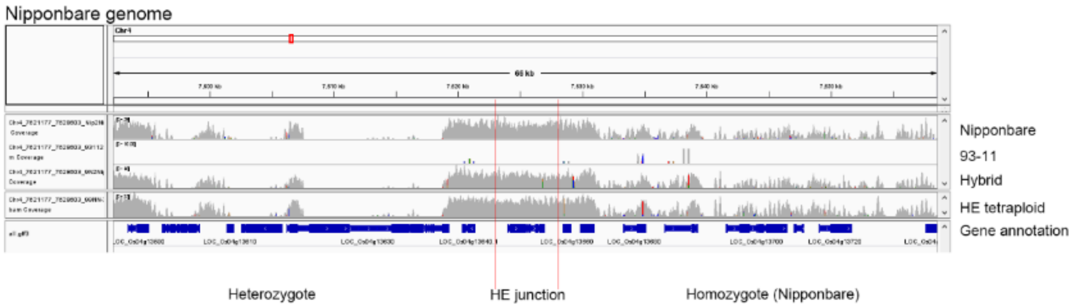
HE junction 88.



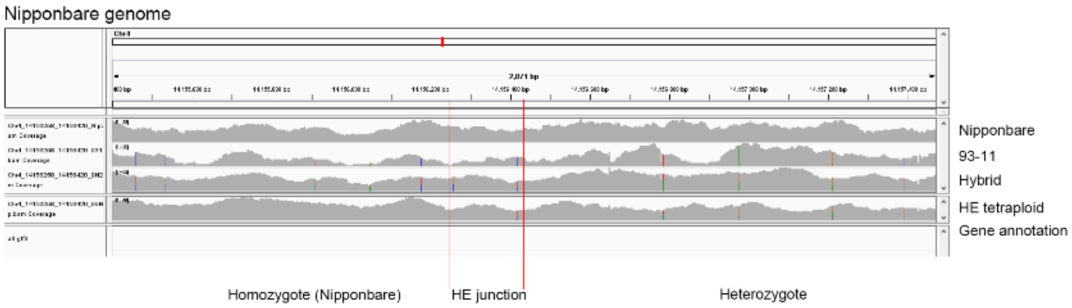
HE junction 89.



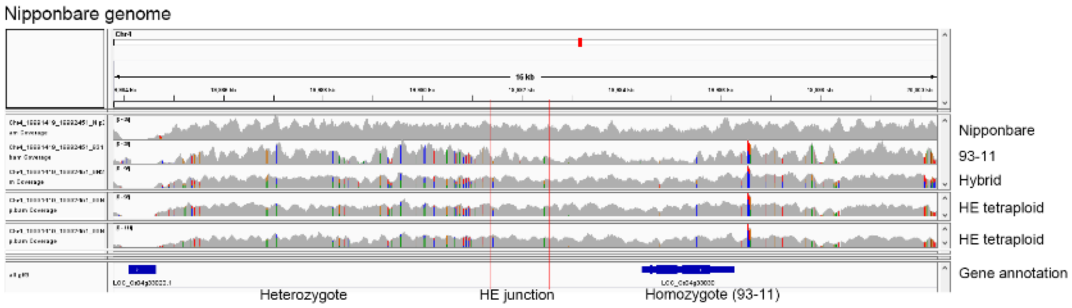
HE junction 90.



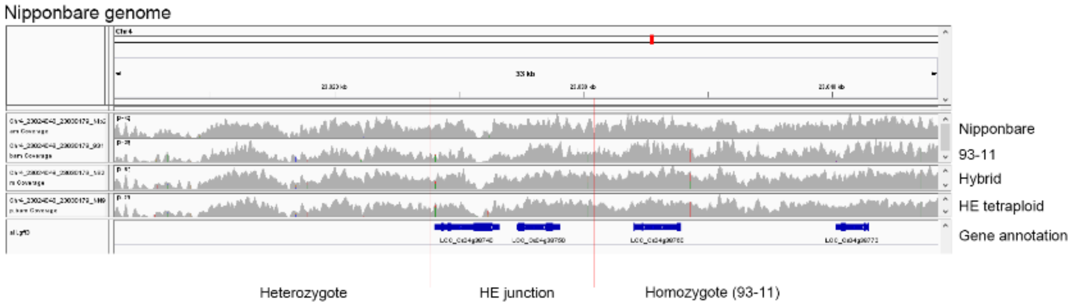
HE junction 91.



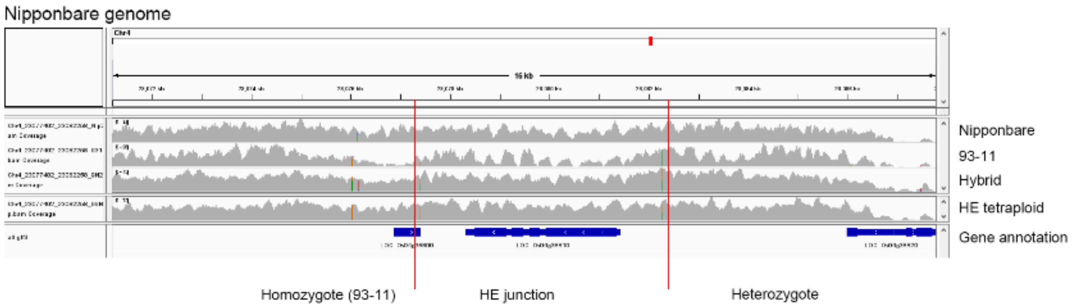
HE junction 92.



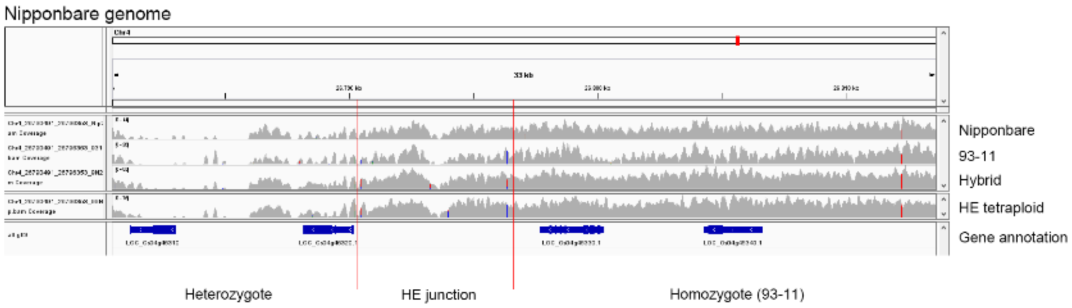
HE junction 93.



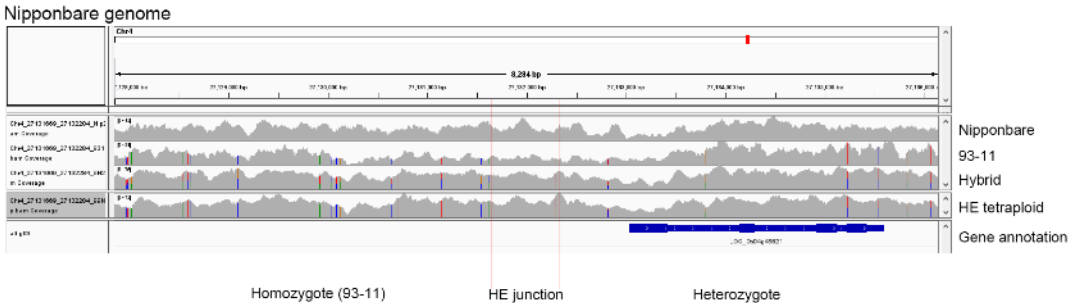
HE junction 94.



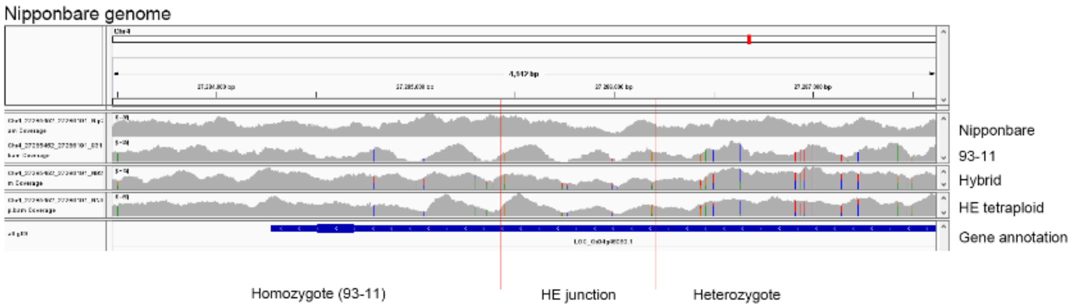
HE junction 95.



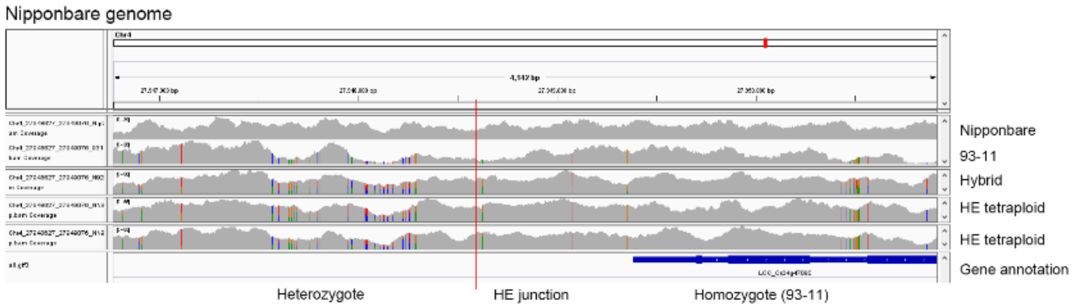
HE junction 96.



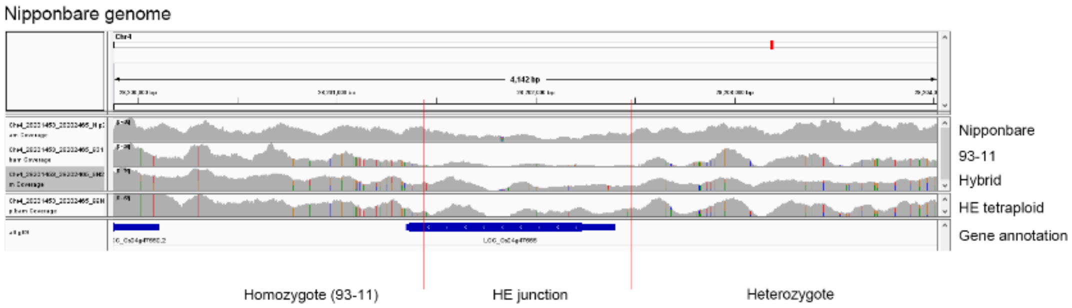
HE junction 97.



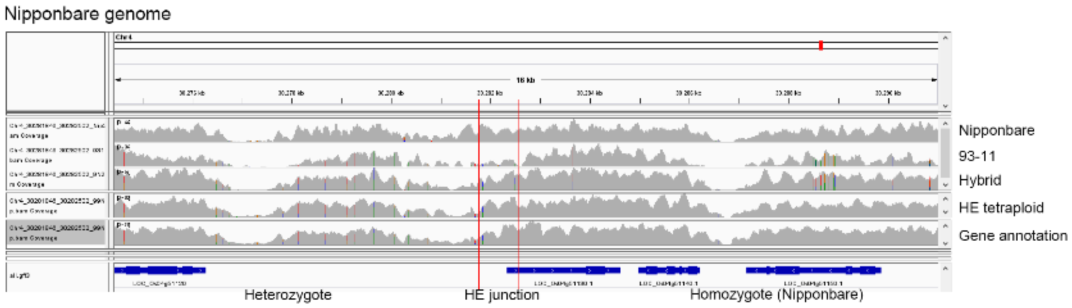
HE junction 98.



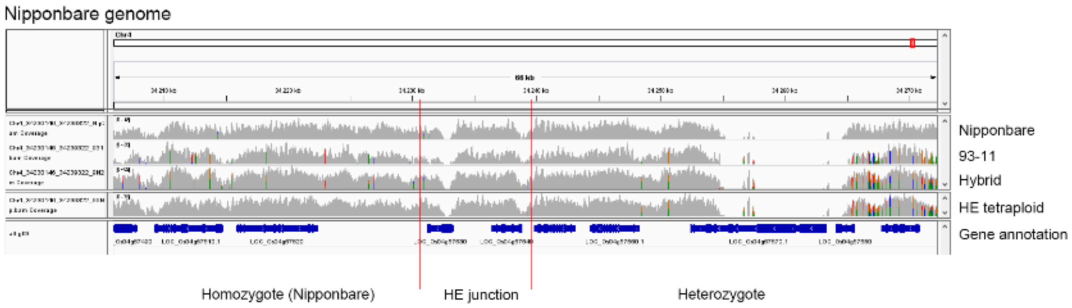
HE junction 99.



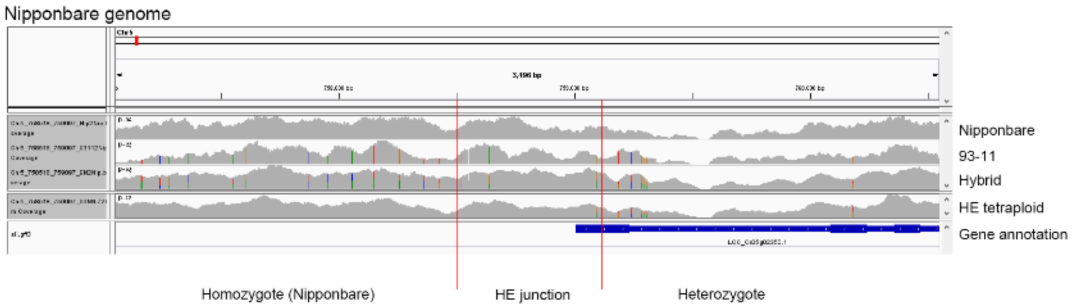
HE junction 100.



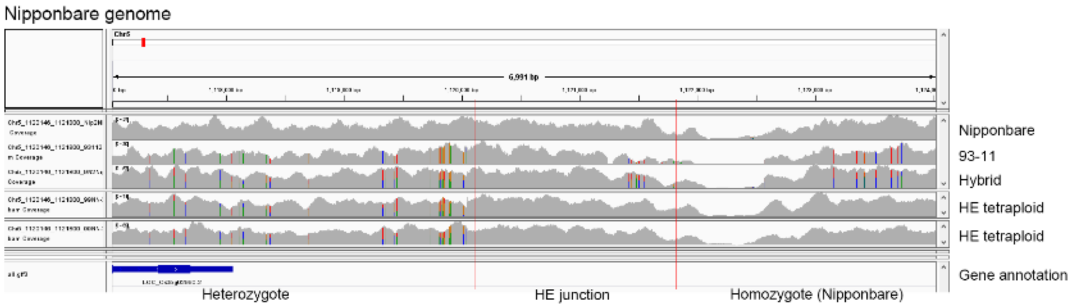
HE junction 101.



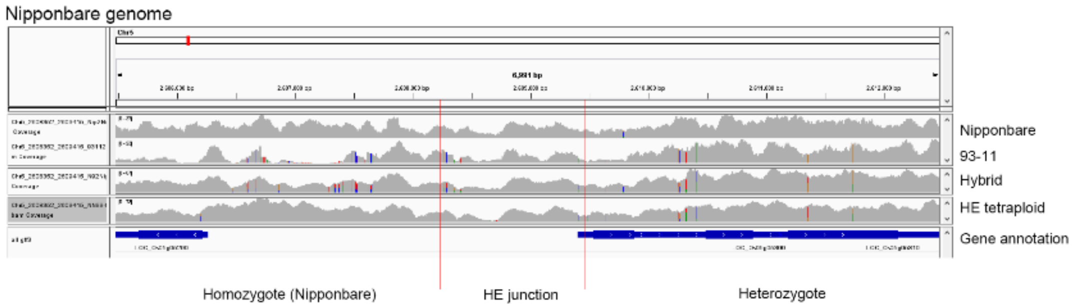
HE junction 102.



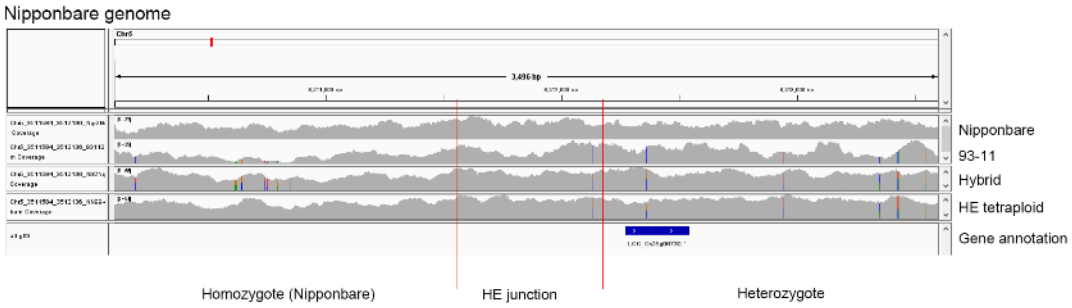
HE junction 103.



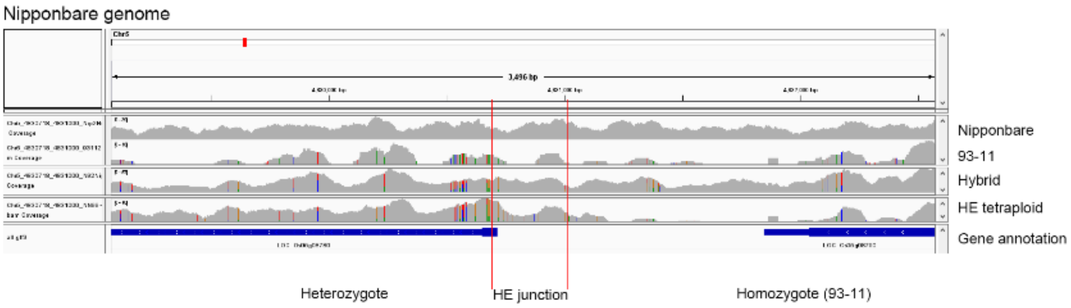
HE junction 104.



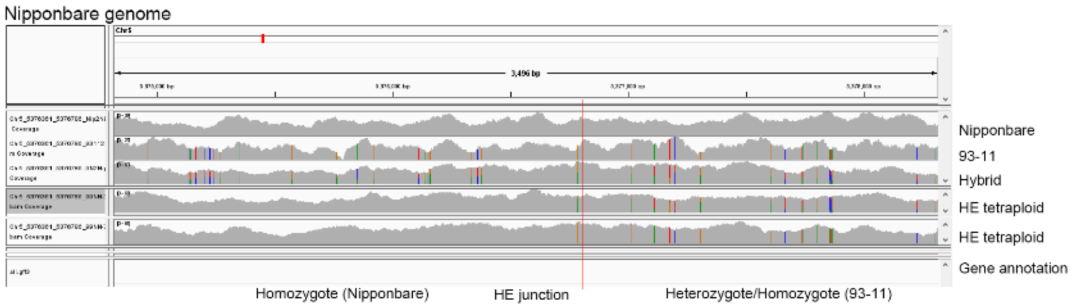
HE junction 105.



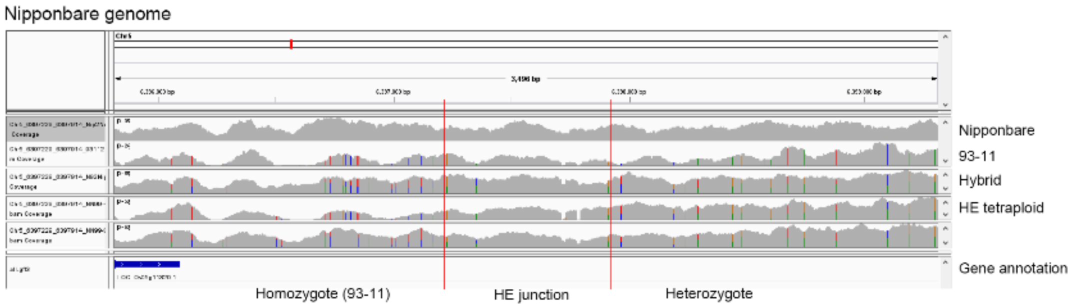
HE junction 106.



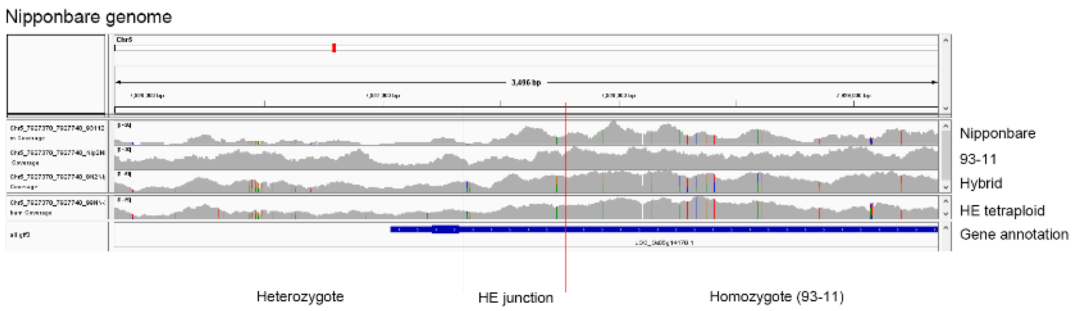
HE junction 107.



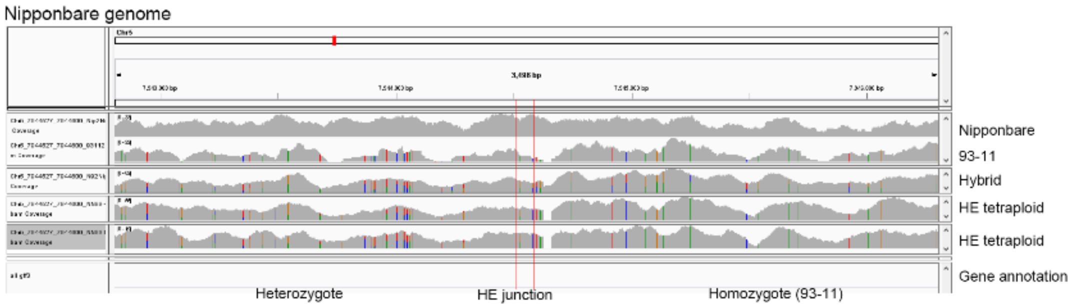
HE junction 108.



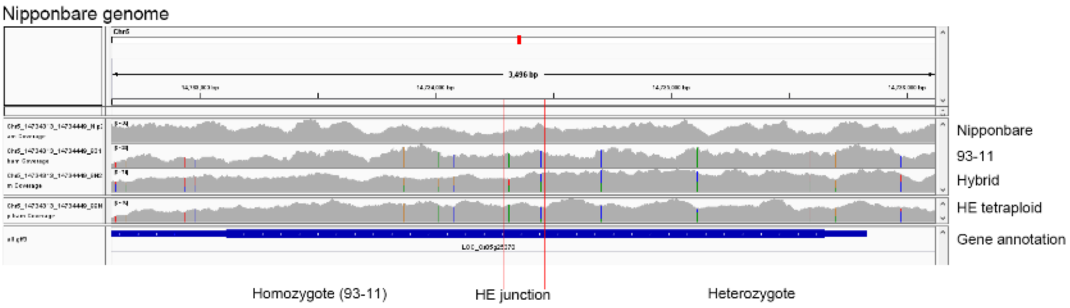
HE junction 109.



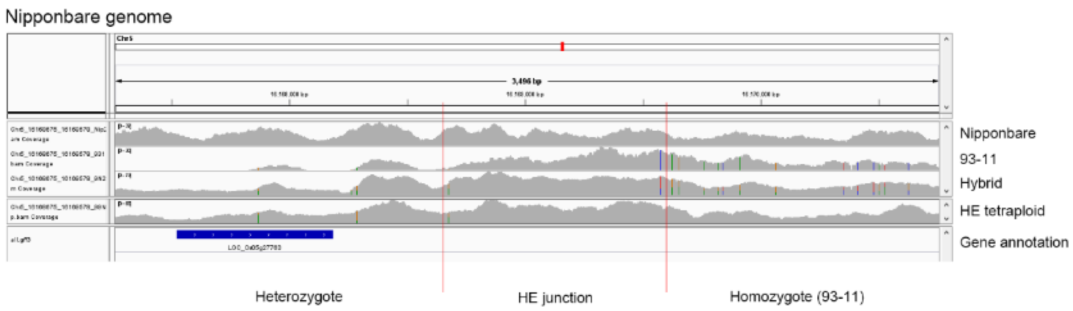
HE junction 110.



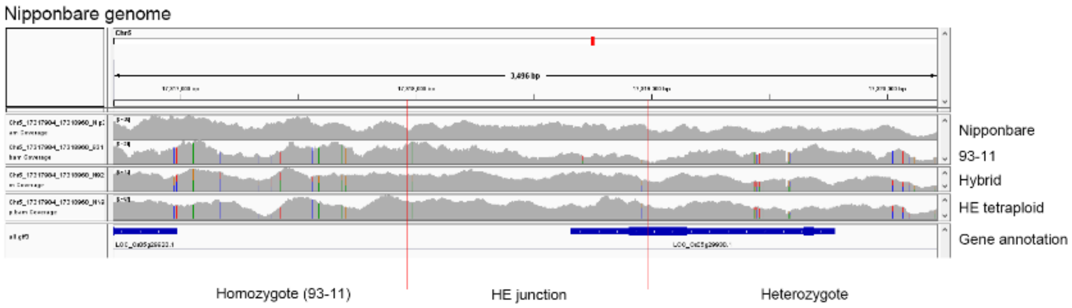
HE junction 111.



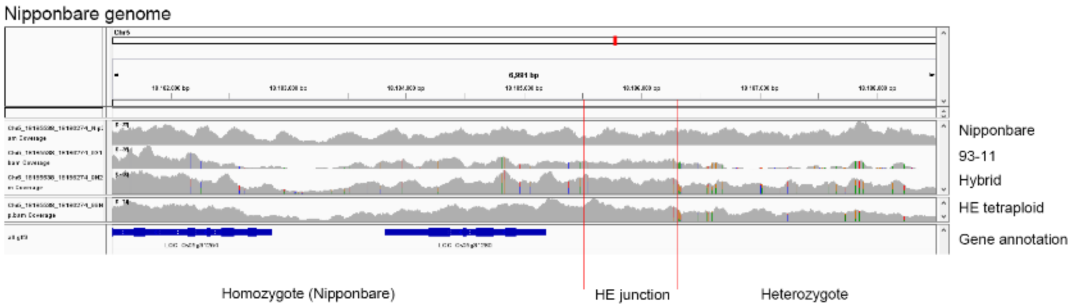
HE junction 112.



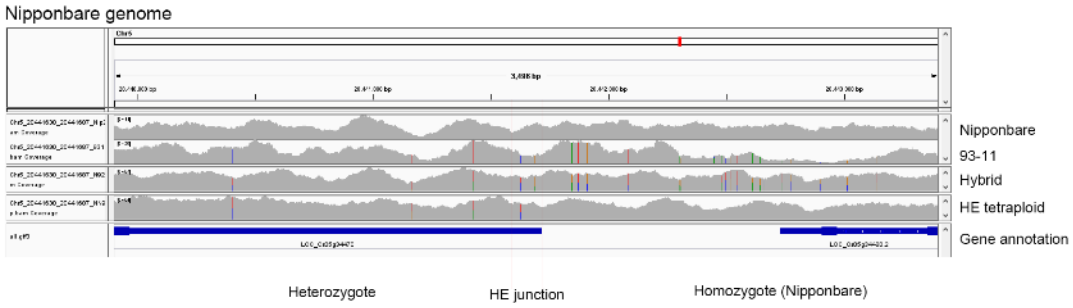
HE junction 113.



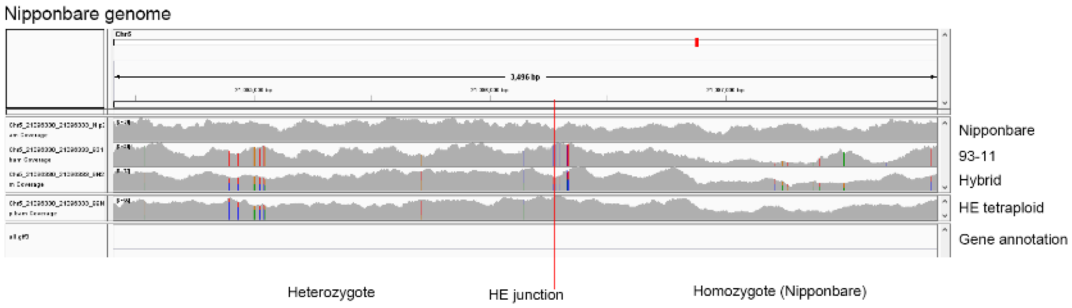
HE junction 114.



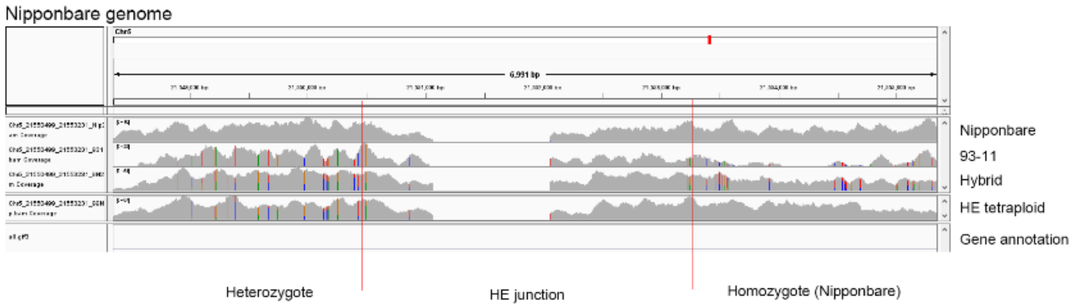
HE junction 115.



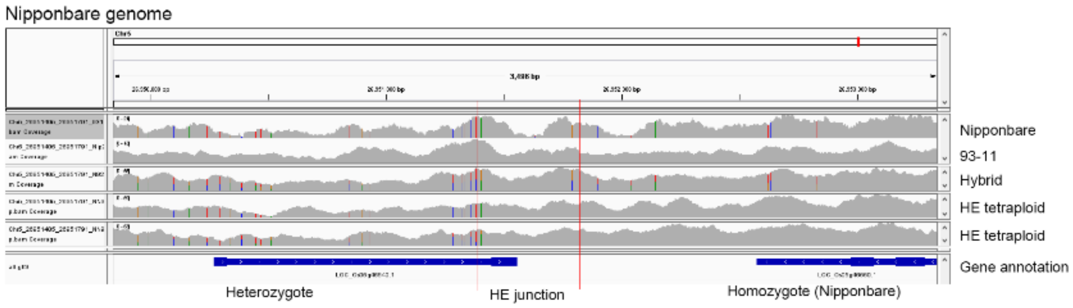
HE junction 116.



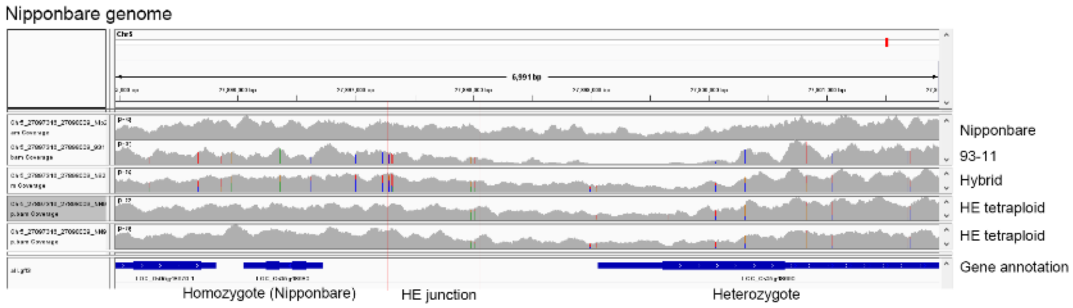
HE junction 117.



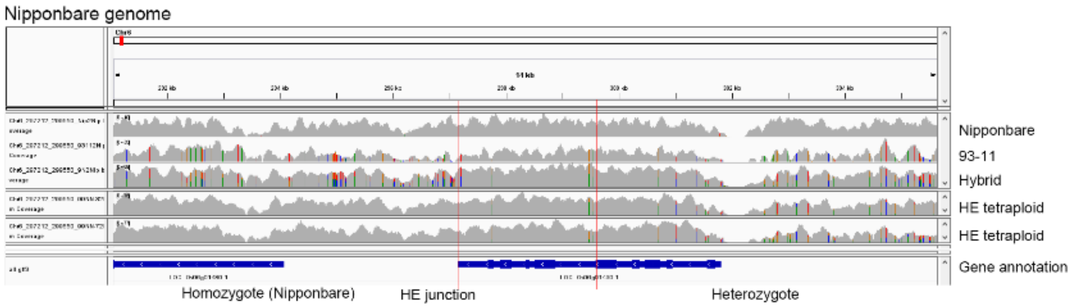
HE junction 118.



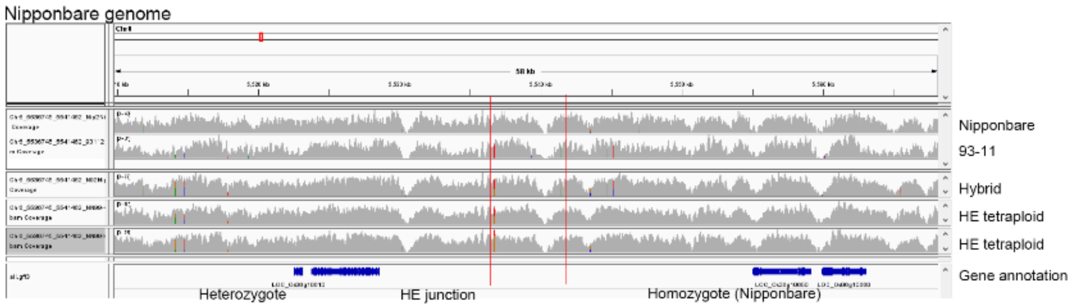
HE junction 119.



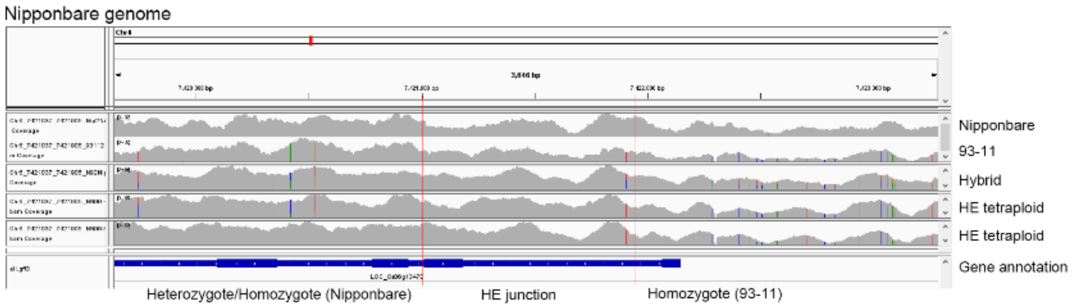
HE junction 120.



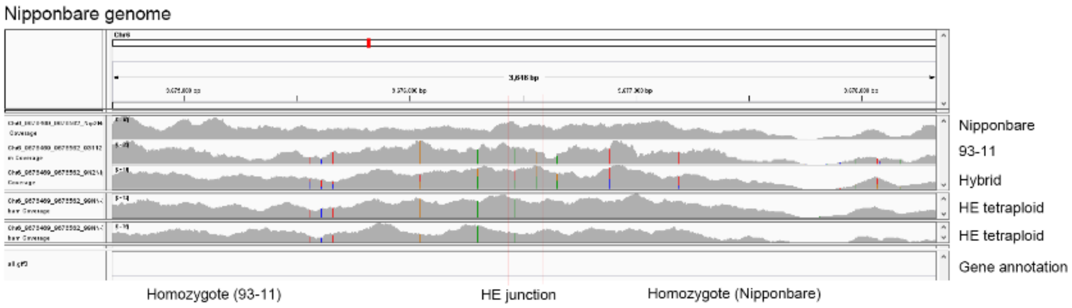
HE junction 121.



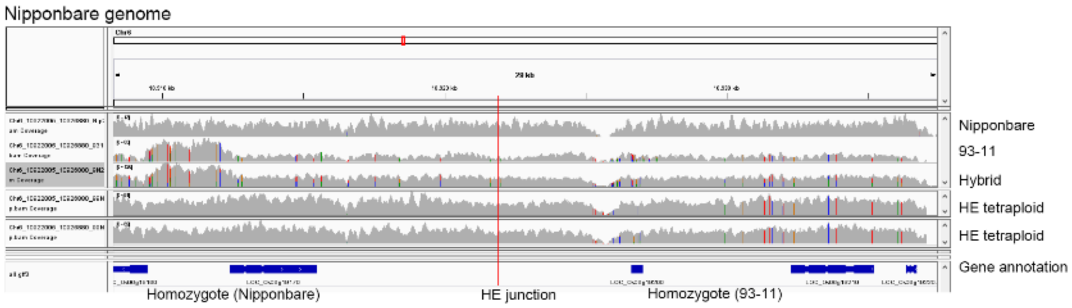
HE junction 122.



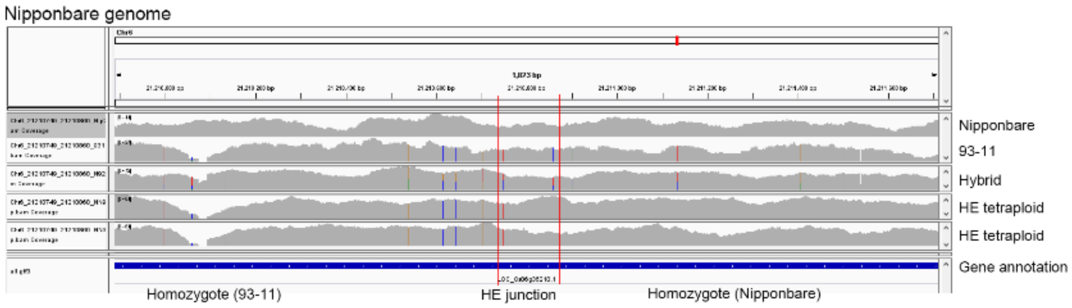
HE junction 123.



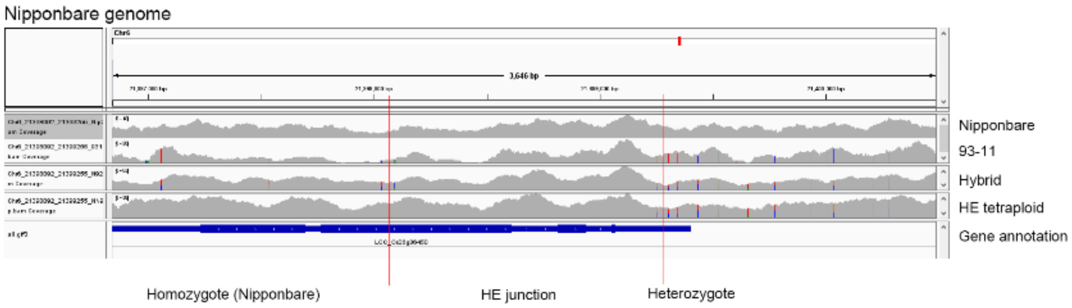
HE junction 124.



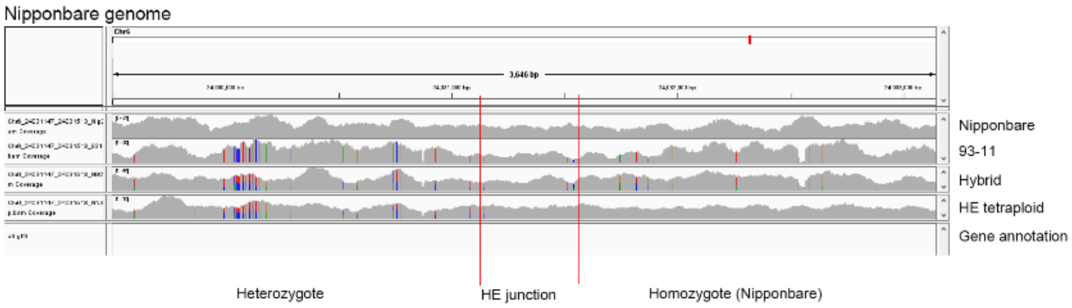
HE junction 125.



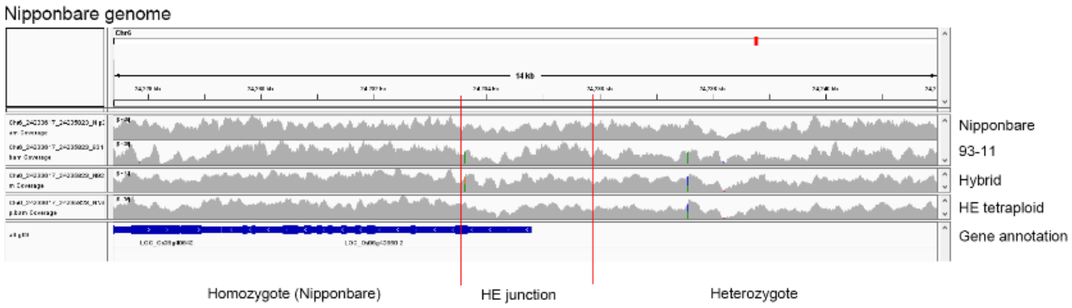
HE junction 126.



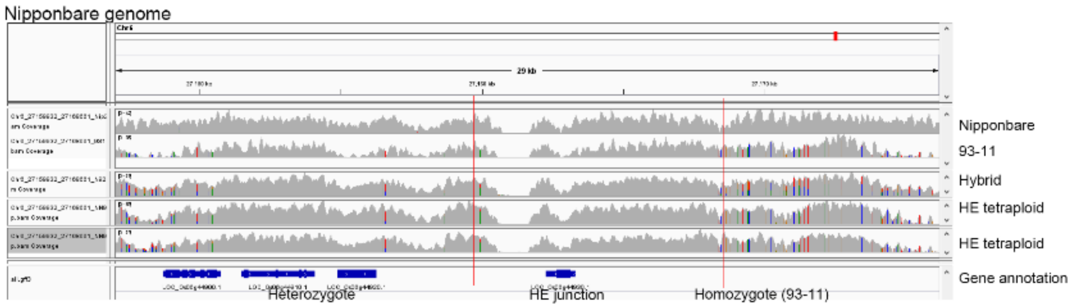
HE junction 127.



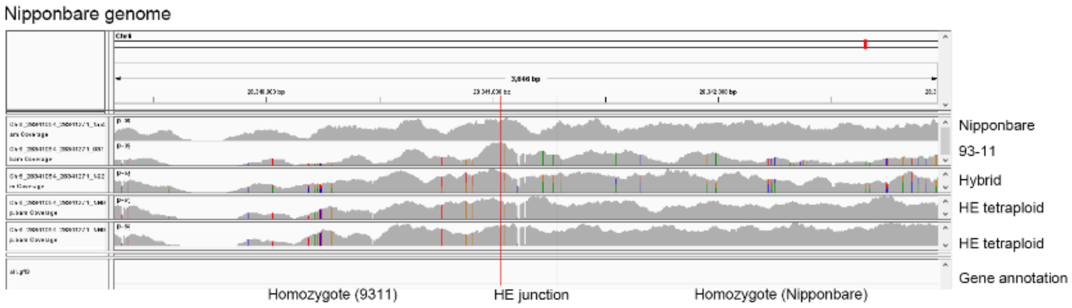
HE junction 128.



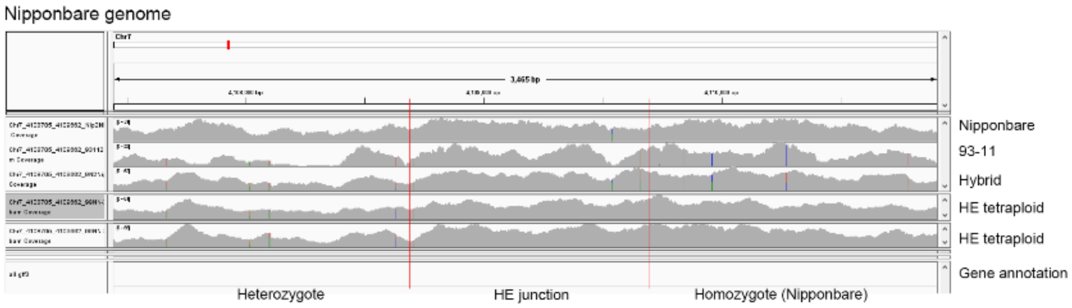
HE junction 129.



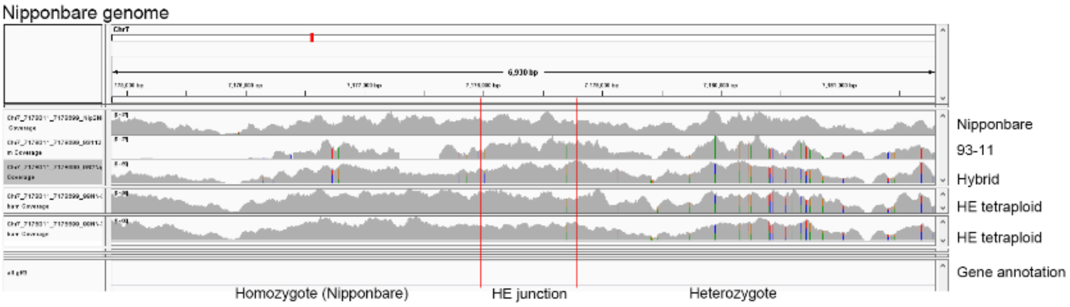
HE junction 130.



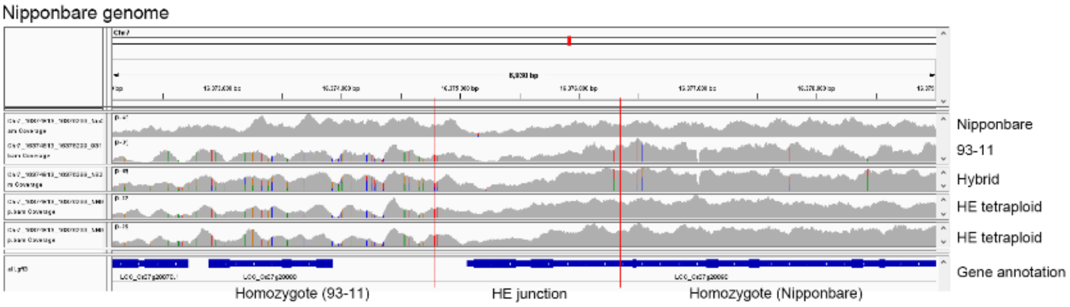
HE junction 131.



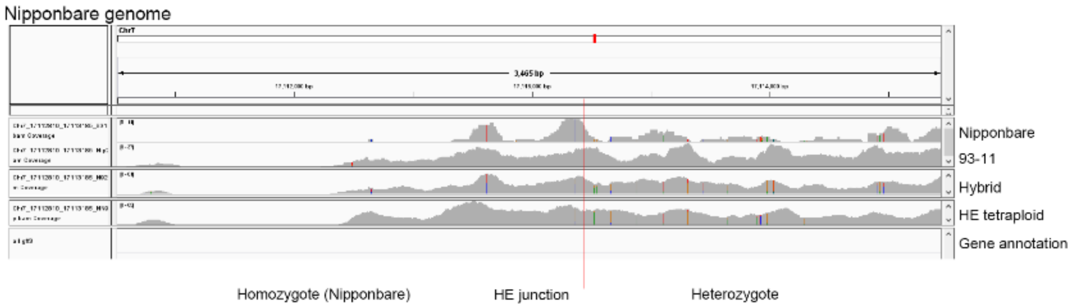
HE junction 132.



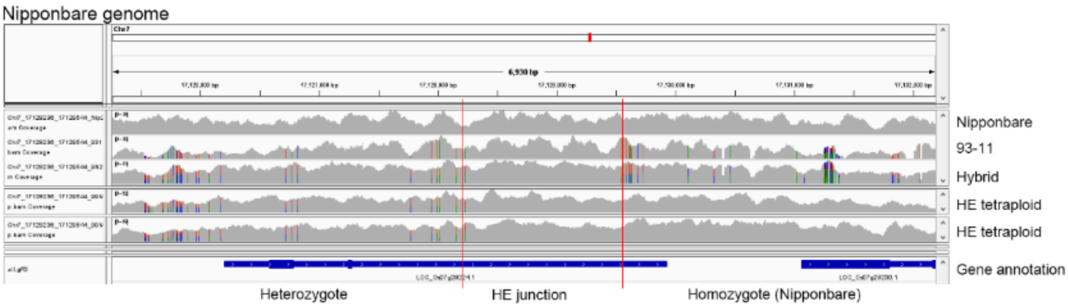
HE junction 134.



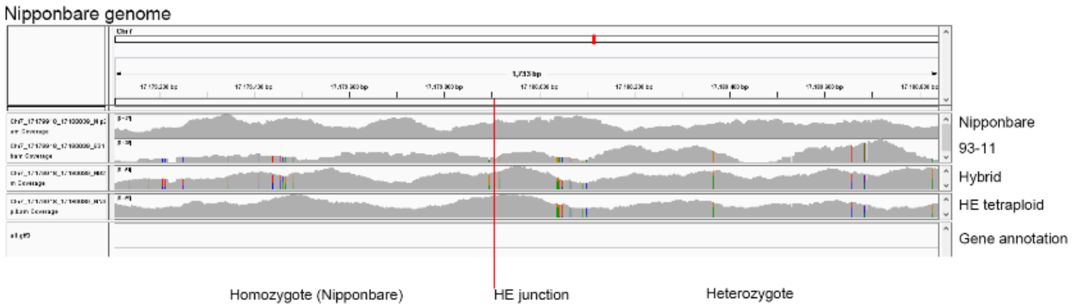
HE junction 135.



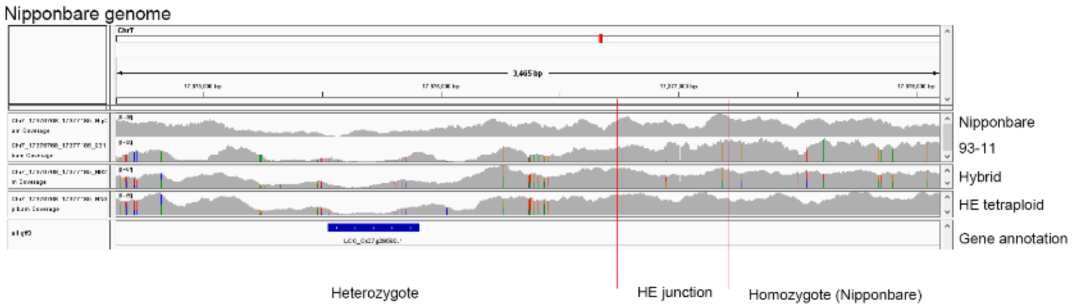
HE junction 136.



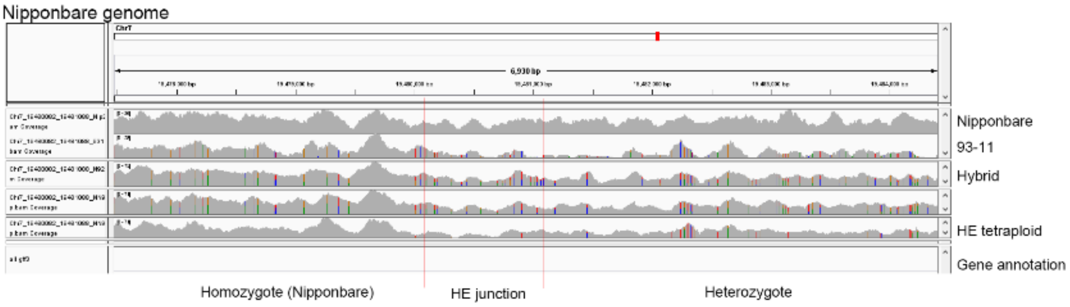
HE junction 137.



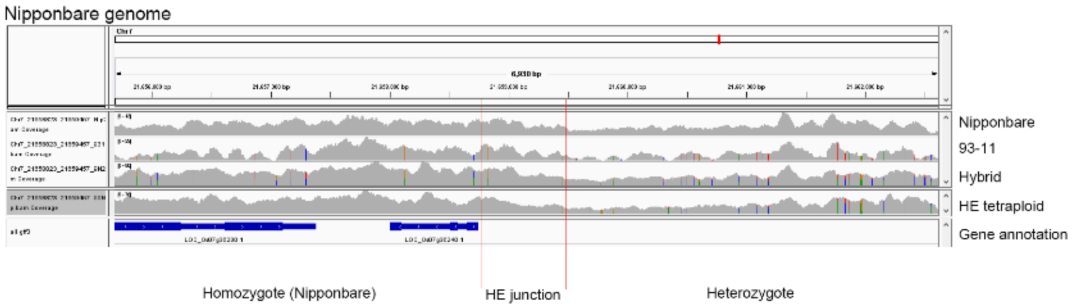
HE junction 138.



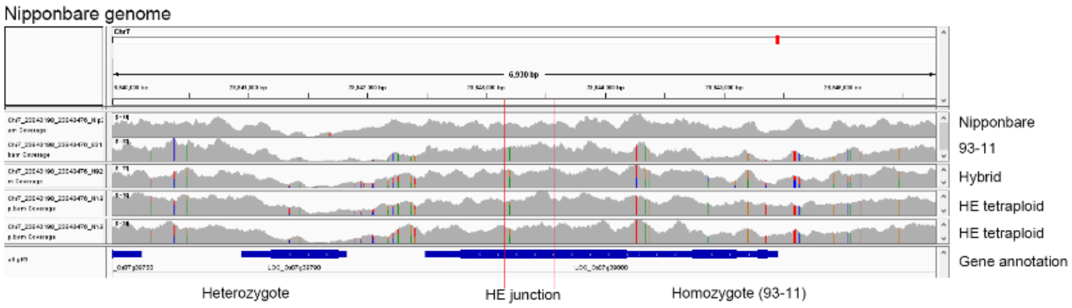
HE junction 139.



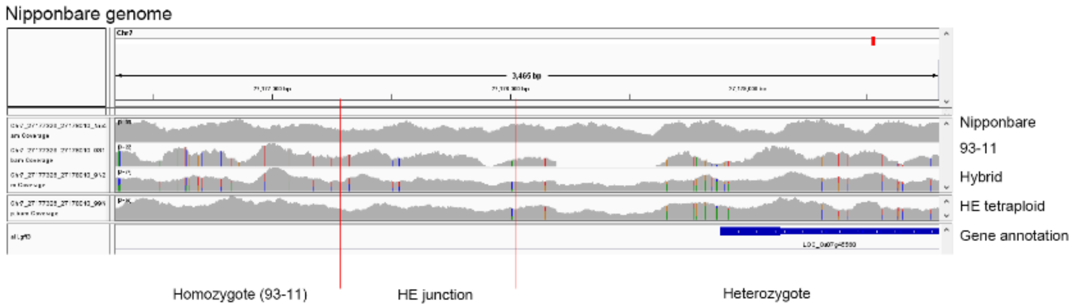
HE junction 140.



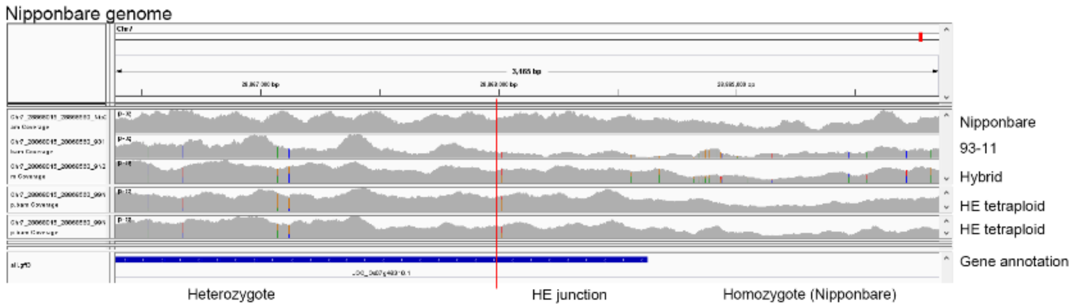
HE junction 141.



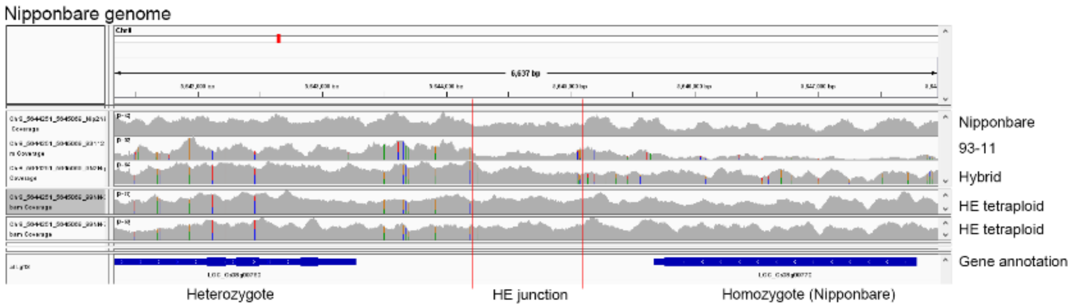
HE junction 142.



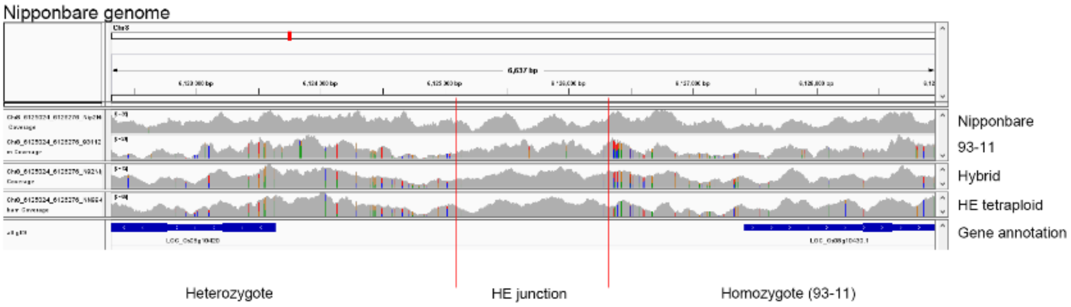
HE junction 143.



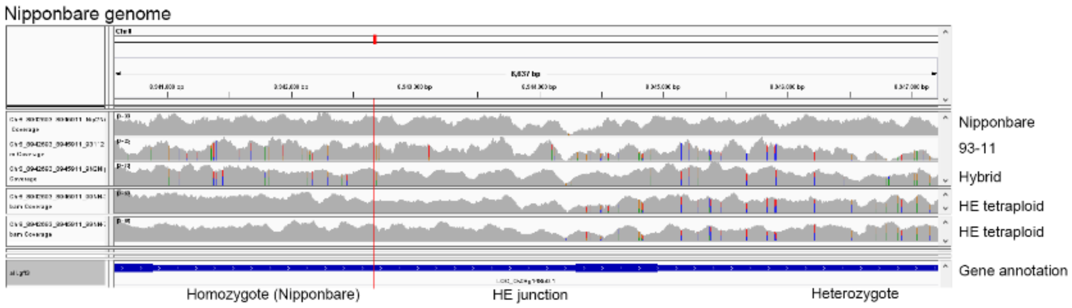
HE junction 144.



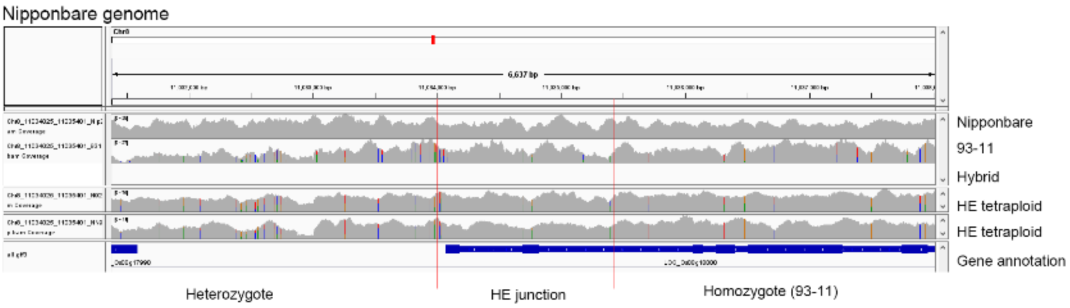
HE junction 145.



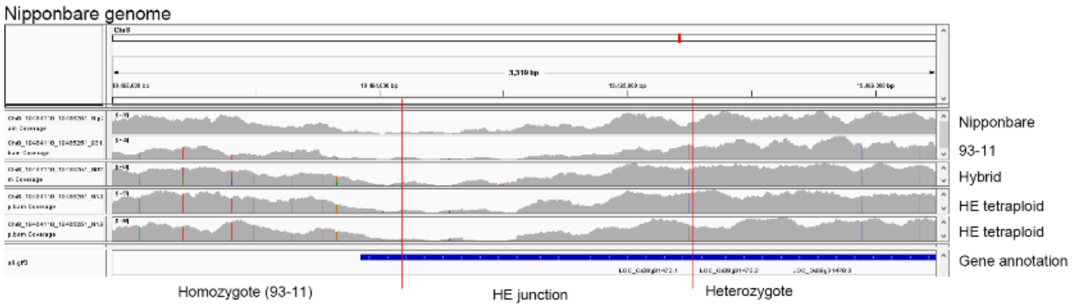
HE junction 146.



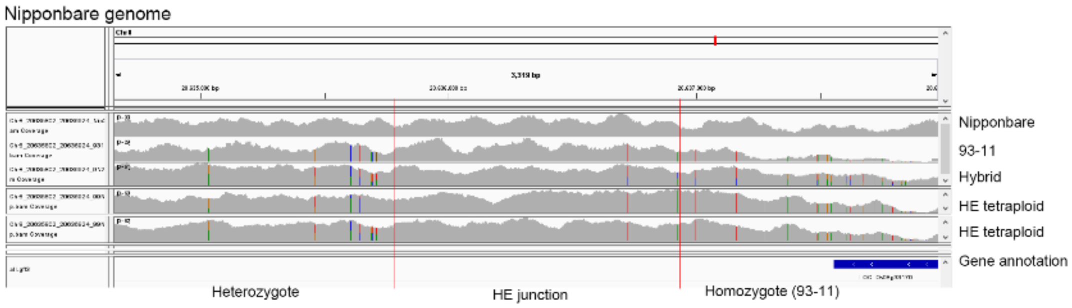
HE junction 147.



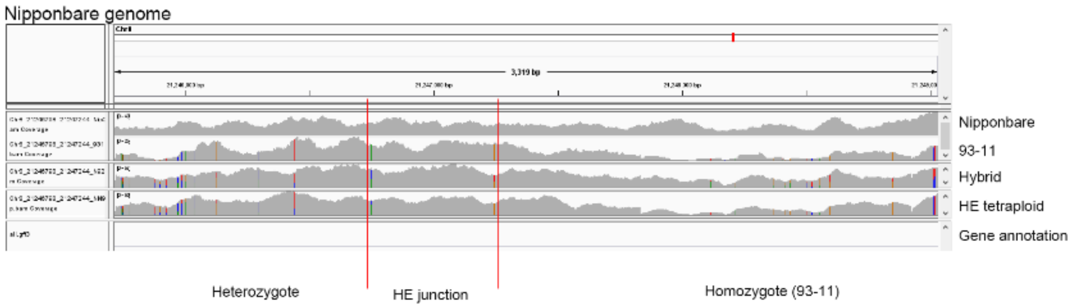
HE junction 148.



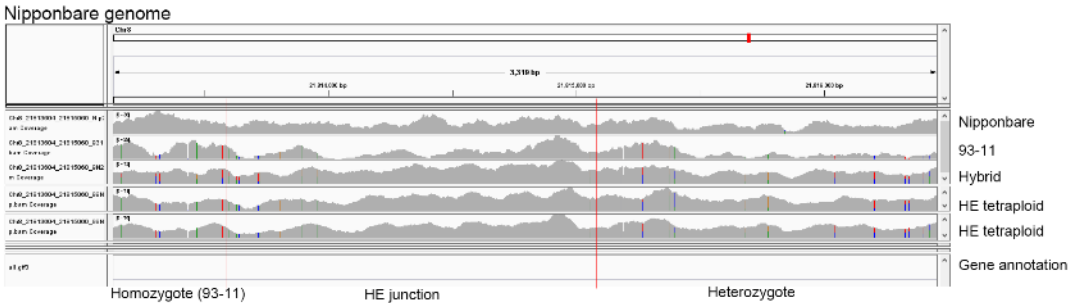
HE junction 149.



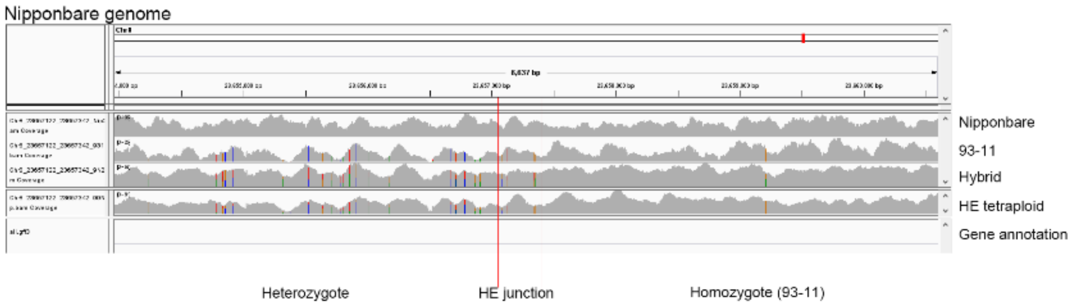
HE junction 150.



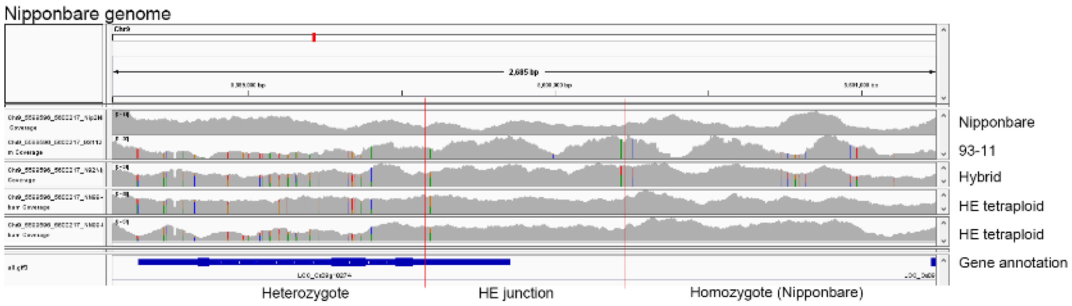
HE junction 151.



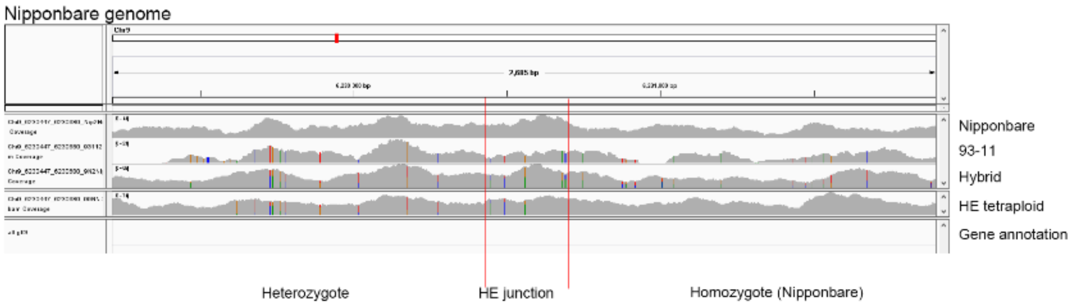
HE junction 152.



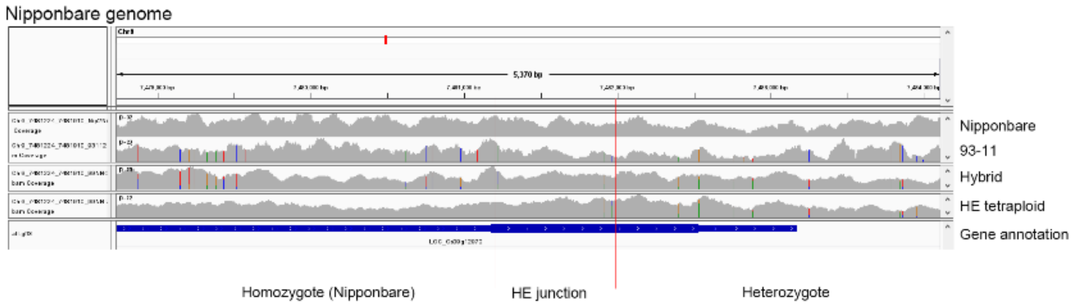
HE junction 153.



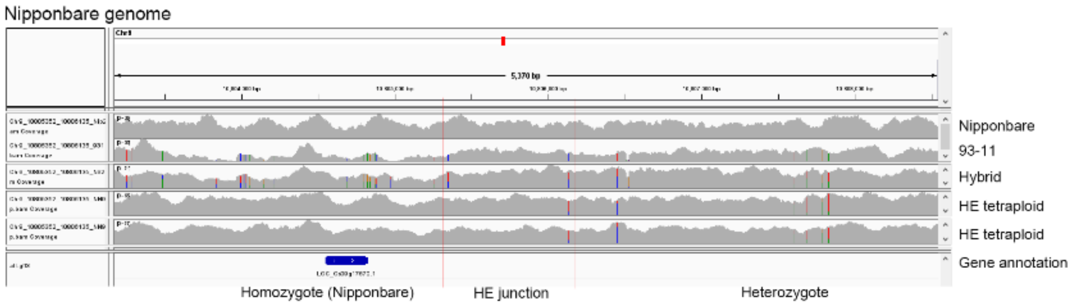
HE junction 154.



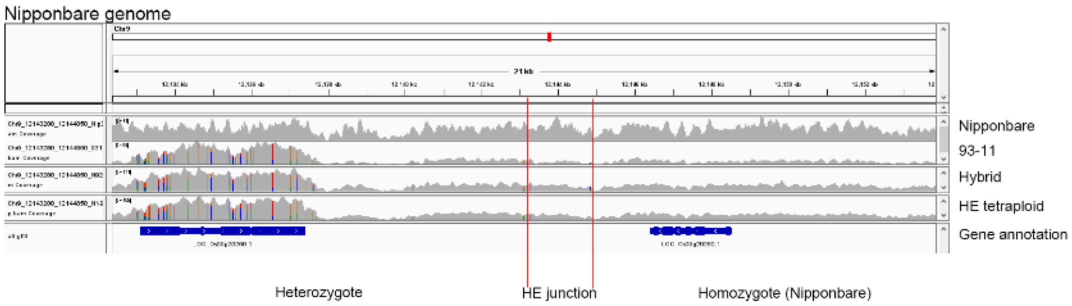
HE junction 155.



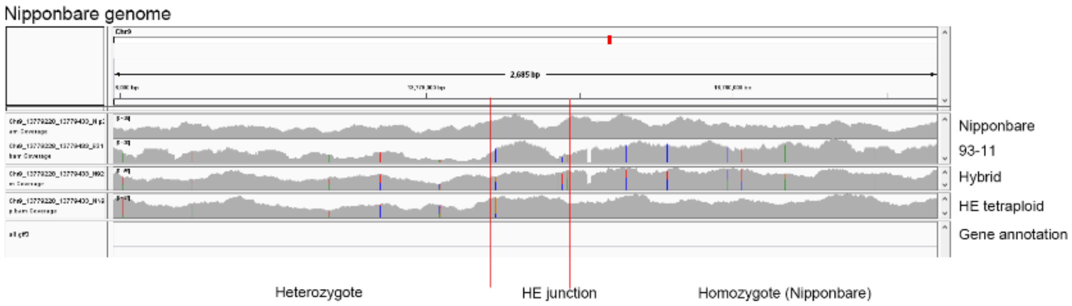
HE junction 156.



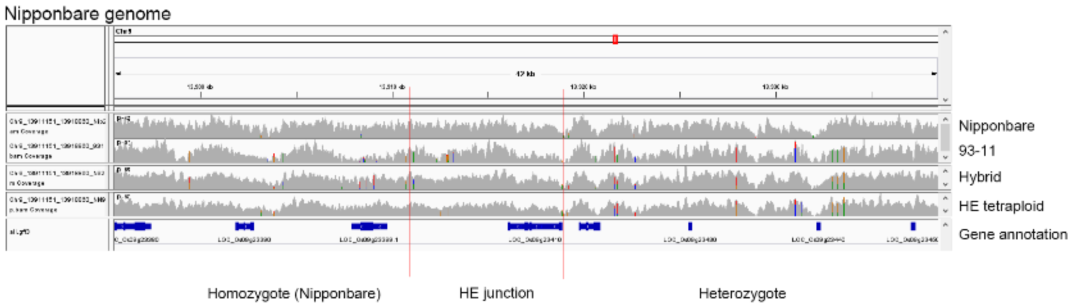
HE junction 157.



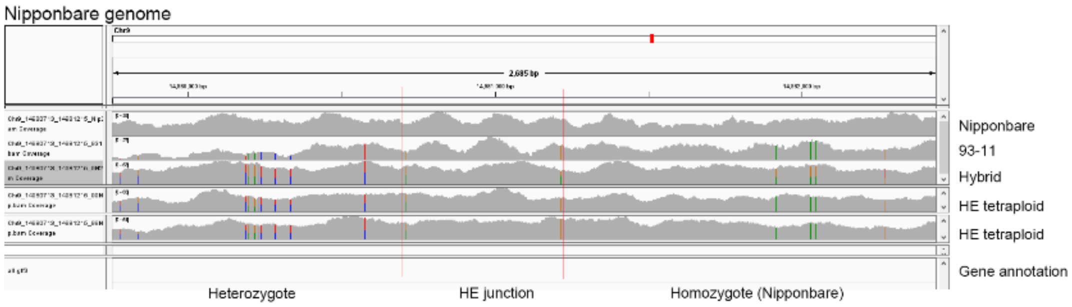
HE junction 158.



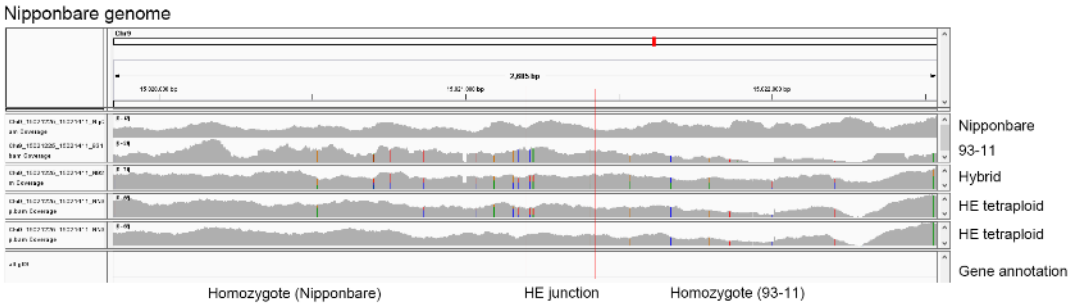
HE junction 159.



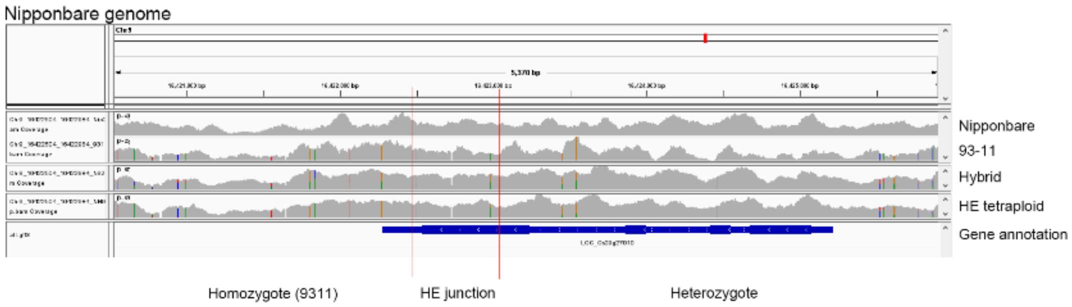
HE junction 160.



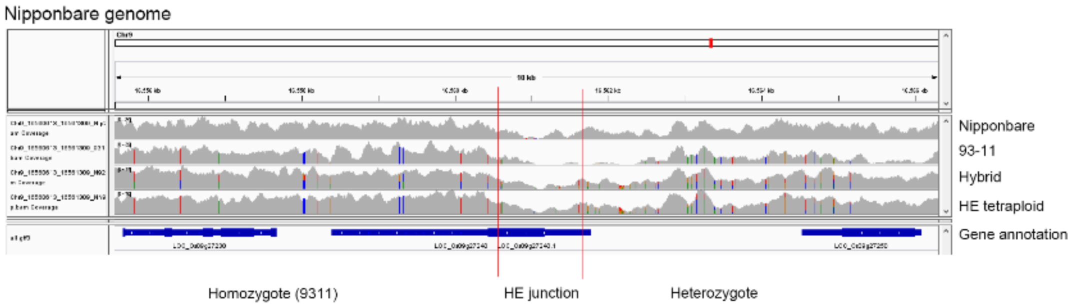
HE junction 161.



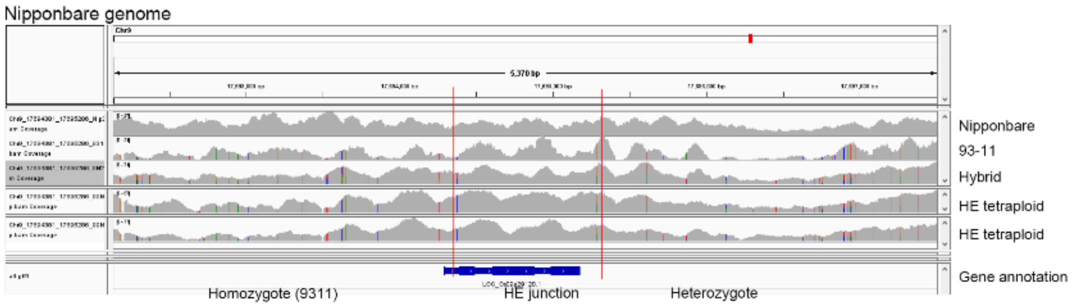
HE junction 162.



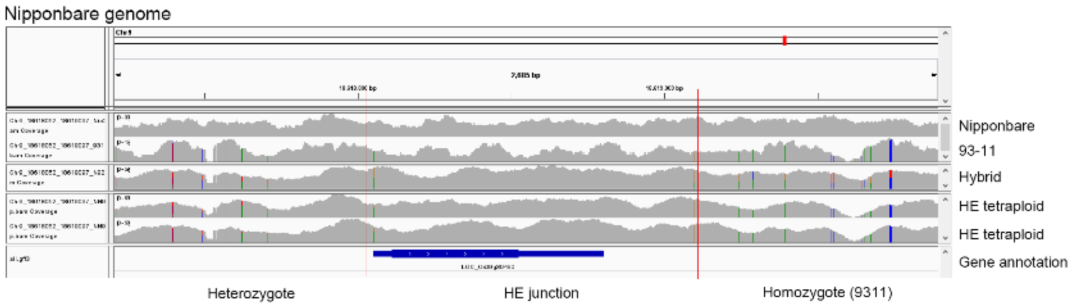
HE junction 163.



HE junction 164.



HE junction 165.



HE junction 166.

Table S1. Chromosome positions of HE junction in wheat A and D subgenomes.

Chr. D	Start D	End D	Chr. A	Start A	End A	Samples	Type[1]	Rank
Chr2	4081262	4081331	Tu2	5854113	5854181	G1 G2	match	pair1
Chr2	30263145	30263479	Tu2	29920319	29920639	G1 G2	match	pair2
Chr2	52046750	52046883	Tu2	45579512	45579643	H	match	pair3
Chr2	74158600	74158914	Tu2	67784015	67785007	G1 G2	match	pair4
Chr2	75211175	75211301	Tu2	68485323	68485447	H	match	pair5
Chr2	506421129	506421186	Tu2	631557463	631557518	H	match	pair6
Chr2	546412191	546412238	Tu2	667713487	667713532	H	match	pair7
Chr2	559510166	559510721	Tu2	677566210	677566761	I	match	pair8
Chr2	588960618	588961886	Tu2	713427396	713428374	J	match	pair9
Chr3	374870371	374872997	Tu3	488565520	488567696	J	match	pair10
Chr3	396392710	396392948	Tu3	511130697	511130934	I	match	pair11
Chr3	467917911	467918516	Tu3	593690522	593691126	I	match	pair12
Chr3	536297000	536299000	Tu3	651100682	651102829	G1 G2	match	pair13
Chr3	618098378	618100436	Tu3	734062498	734064366	I	match	pair14
Chr3	625165404	625165730	Tu3	743197205	743197529	G1 G2	match	pair15
Chr4	517221028	517221122	Tu5	651762614	651762704	I	match	pair16
Chr5	556183988	556184086	Tu4	604288955	604289052	all	match	pair17
Chr5	561442633	561443226	Tu4	607815658	607816345	G1	match	pair18
Chr6	29626000	29628000	Tu6	28046771	28048534	G2	match	pair19
Chr6	390328742	390328999	Tu6	475223901	475224157	H	match	pair20
Chr6	478071000	478073000	Tu6	557158220	557160221	G1 G2	match	pair21
Chr6	479733727	479734020	Tu6	559593480	559593770	G2	match	pair22
Chr7	583312765	583312819	Tu7	651742736	651742789	H	match	pair23
Chr2	643692861	643693035	Tu2	744609000	744619000	I	no match	pair24
Chr3	449708292	449711421	Tu3	575953875	575962554	I	low res.	pair25
Chr4	493497185	493497257	Tu5	623025000	623028000	I	no match	pair26
Chr2	39105736	39116042	Tu2	39705000	39708000	H	low res.	pair27
Chr4	57905217	57905279	Tu4	62579600	62582000	G1 G2	no match	pair28
Chr7	569559483	569567346	Tu7	643189000	643207000	J	shift(12.7Mb)	pair29
Chr2	639513469	639514631	Tu2	737648000	737652000	G1 G2	no match	pair30
Chr6	487078755	487079082	Tu6	569561000	569606000	G1	shift (387kb)	pair31
Chr3	555267182	555267373	Tu3	670210000	670212000	G1 G2	shift (427kb)	pair32
Chr2	552663041	552663113	Tu2	672383000	672385000	H	no match	pair33
Chr3	436186321	436186399	Tu3	561990000	561993000	I	shift (49Mb)	pair34
Chr4	5910000	5914000	Tu4	8834000	8838000	H	shift (77kb)	pair35
Chr2	594996015	594996114	Tu2	707387000	707417000	J	shift (94Mb)	pair36
Chr3	4490649	4490701	Tu3	5453000	5471113	I	no match.	pair37

[1] indicates the match types of HE junctions between A and D sugbenome. match: HE junction pairs were well aligned; shift: HE junctions from D subgenome were not aligned to those from A subgenome but in the same chromosome, content in brackets represent the genomic distance between the alignment hit and HE junction of in A subgenome; no match: HE

junctions from D subgenome were not aligned to A subgenome; low res.: HE junction with low resolution and not used for alignment.

Table S2. Number of HE junctions in different homoeologous groups in AADD tetraploids

	Total HE junctions	Well-aligned HE junctions
Group 1	0	0
Group 2	14	9
Group 3	10	6
Group 4	2	0
Group 5	0	0
Group 6	5	4
Group 7	2	1
4A and 5D	2	1
5A and 4D	2	2
All	37	23

Table S3. Primer sequences for validating full-length fusion transcripts by Sanger sequencing

Primers	Sequences
AET5Gv21168400-F	CACTCCGTGAGGAACCG
AET5Gv21168400-R	GAGGATTTGCCATCTGTGA
AET2Gv20016200-F	CAGGATGGGGATTCAAGA
AET2Gv20016200-R	ACTGTAACAGATTCAAGGCACT
AET2Gv20252700-F	CGCTCTACCTGCTCTTCG
AET2Gv20252700-R	CAGATACGCCGCTTCAGT
AET6Gv20704700-F	GGCCACCATCCTGGAGAACAT
AET6Gv20704700-R	GCTCTGGGTGTAGGCGAACTGC
AET2Gv21000600-F	GGGCGATGCGGGTGAAA
AET2Gv21000600-R	TCAAGGTAGTGCTGGTAGATGTAGGA
AET3Gv20796600-F	CCGACGATCTACGCCTTCT
AET3Gv20796600-R	CCCTCGCAGCAACACTCTA
AET3Gv20936700-F	ATGCGGCCATGAGTTTG
AET3Gv20936700-R	CTCCCTCTAGTTCGCTTCG
AET4Gv20852900-F	GTCCTCATCTTCCCCTTCTTCG
AET4Gv20852900-R	AGTTCTTCTTGGCTCTTGCTTCC

Table S4. Chromosome positions of 104 HE junctions in eight *Brassica napus* accessions (based on the position in C genome of *B. oleracea* and A genome of *B. rapa*).

Chr. C	Start C	End C	Chr. A	Start A	End A	Samples	Type[1]
C1	3990388	3991188	A01	2977924	2978111	Kale	match
C1	11735922	11737262	A01	7858737	7859586	Kale	match
C1	11865844	11866094	A01	7960324	7960579	Kale	match
C1	14151827	14152506	A01	10002456	10003132	H165	match
C1	15287827	15288219	A01	10741934	10742323	H165	match
C1	15853350	15853654	A01	11095573	11095878	Aviso, Bristol	match
C1	16121876	16122339	A01	11324456	11324918	Aviso, Bristol	match
C1	34309991	34310187	A01	21441244	21441437	H165	match
C1	34409828	34410006	A01	21530833	21531013	Kale	match
C1	35772368	35772440	A01	22840197	22840467	Swede	match
C1	37350795	37351421	A01	23795409	23796028	Aviso, Bristol	match
C1	39617623	39618660	Scaffold000164	42911	43947	Bristol	other_chr
C1	39966798	39969099	A01	26526111	26528203	Darmor	match
C1	40849674	40849952	A01	27396486	27396777	Darmor	match
C1	41061550	41061721	A01	27559695	27559865	Darmor	match
C1	41321910	41322147	A01	27714087	27714320	Darmor	shift (3kb)
C1	41749893	41752893	A01	28002805	28005720	Yudal	match
C1	42406658	42409707	Scaffold000229	48896	51211	Aviso, Bristol	other_chr
C1	42406658	42407785	Scaffold000229	50818	51211	Swede	other_chr
C1	43065473	43066124	A01	25557481	25558127	Swede	match
C2	494539	495291	A02	1700901	1701616	H165	match
C2	2414523	2417669	A02	2817850	2820774	Yudal	match
C2	3492502	3493594	A02	3612871	3613053	Aburamasari	match
C2	6765978	6766015	A02	5606300	5606417	Darmor	match
C2	7698382	7698643	A02	6055332	6055592	Darmor	match
C2	7831687	7831943	A02	6149124	6149379	H165	match
C2	10349704	10351053	A02	7379461	7380768	Aburamasari, swede, Yudal	match
C2	11408303	11408364	A02	7837983	7838043	Aburamasari, swede, Yudal	match
C2	22734221	22734342	A02	13377573	13377691	Swede	match
C2	23048525	23049745	A02	13554306	13555525	Darmor	shift (6kb)
C2	24597326	24597479	A02	14332176	14332306	Swede	match
C2	38035924	38037924	A08	14565897	14569316	H165	other_chr
C2	39469355	39469526	A02	18187047	18187225	Kale	match
C2	41046507	41047022	A02	19229213	19229718	Kale	match
C2	42209256	42214691	A02	21071121	21073404	H165	match
C2	50793193	50794281	A02	25881931	25883040	Darmor	match
C2	51312311	51312407	A02	26296470	26296560	Darmor	match
C2	51417726	51419235	Scaffold000490	5076	6337	Swede	other_chr
C2	51441401	51442250	A02	26401130	26401979	Aburamasari, Yudal	match
C2	51799087	51799605	A02	26567316	26567863	Aburamasari	match
C2	52795396	52796544	A02	27502393	27503540	Swede	match

C3	2006236	2008479	A03	1672744	1673744	Aburamasari	match
C3	2317136	2322567	A03	1865672	1867972	Aburamasari, Yudal	match
C3	4874207	4883188	A03	3880693	3888275	Swede	match
C3	5851394	5851524	A03	4682354	4682487	Aviso, Bristol	match
C3	6247040	6247245	A03	4911997	4912159	H165	match
C3	6367698	6367904	A03	5017476	5017680	Aviso	match
C3	6432816	6434256	A03	5068737	5070176	H165	match
C3	6725055	6727720	A03	5317319	5319125	Bristol	match
C3	6924496	6924906	A03	5447284	5447687	Bristol	match
C4	332560	334644	A05	244419	246516	Kale	match
C4	48097092	48097164	A04	15032101	15032172	Darmor	match
C4	48671789	48672442	A04	15394477	15395129	Darmor	match
C4	49060310	49060983	A04	15648440	15649120	Darmor	match
C4	49690619	49691040	A04	16087023	16087443	Darmor	match
C4	49744132	49744747	A04	16121435	16122058	Swede	match
C4	50961650	50961766	A04	17103791	17103906	Kale	match
C5	40730725	40731829	A05	20650702	20651411	Bristol	match
C5	41155931	41156036	A05	20934550	20934654	Bristol	match
C5	41830248	41830334	A05	21373620	21373882	Kale	match
C5	42014855	42015091	A05	21504892	21505111	Kale	match
C5	42718082	42718260	A05	21964626	21964802	Kale	match
C5	42883620	42892863	A05	22068769	22078022	Yudal, Aburamasari	match
C5	43079621	43079870	A05	22213585	22213833	Aburamasari, Yudal	match
C5	44628034	44628532	A02	23231824	23232375	Aviso, Bristol	other_chr
C5	45313958	45314187	A05	23467203	23467419	Swede	match
C5	46030229	46032365	A02	26802762	26804822	Swede	other_chr
C5	46416349	46416426	A05	23623171	23623446	Yudal	match
C6	38384984	38385261	A07	21359722	21359998	Yudal	match
C6	38483157	38483535	A07	21452115	21452484	Yudal	match
C8	35753612	35756743	A09	32062424	32062911	H165	match
C8	35853527	35854117	A09	32171396	32171965	H165	match
C8	37493119	37497467	A09	33644447	33647886	H165	match
C8	38028713	38029713	A09	34119912	34120709	Bristol	match
C8	38429460	38431592	A09	34453620	34454711	Swede	match
C8	39529544	39529661	A09	35237812	35237928	Bristol	match
C8	40140326	40140458	A09	35725845	35725976	Aviso, Bristol	match
C8	40558577	40558852	A09	36150008	36150282	H165	match
C8	40796260	40796808	A09	36360984	36361401	Aviso, Bristol	match
C8	41227220	41229135	A09	36732348	36734330	Aburamasari	match
C8	41448018	41448156	A09	36914218	36914355	Darmor	match
C9	141476	141594	A09	1037680	1037803	Darmor	match
C9	300095	300568	A09	901176	901509	Kale	match
C9	593138	593365	A09	629747	629974	Aviso, Bristol	match
C9	908000	918000	A09	397099	403042	swede, Yudal	match

C9	1080870	1086136	A09	240441	242307	Yudal	match
C9	1085741	1085938	A09	240637	240833	Swede	match
C9	1087374	1087558	A09	238966	239153	H165	match
C9	2043382	2043448	A09	1332047	1332112	Yudal	match
C9	2045144	2045228	A09	1333816	1333899	Aburamasari	match
C9	2870455	2870526	A09	2589483	2589553	Swede	match
C9	3014517	3014760	A09	2698977	2699219	Aburamasari	match
C9	3360366	3360818	A09	1869180	1869618	Yudal	match
C9	3361156	3362184	A09	1867829	1868824	Aburamasari	match
C9	4315421	4315568	A09	2767361	2767507	Yudal	match
C9	4502543	4503520	A09	2919760	2920408	Yudal	match
C9	6659489	6660191	A09	4175893	4176566	Swede	match
C9	8684926	8685180	A09	5319971	5320224	Darmor	match
C9	8857111	8857268	A09	5431010	5431189	Darmor	match
C9	9895116	9895692	A09	7325629	7326202	Aviso, Bristol	match
C9	10292737	10292876	A09	7089358	7089494	Aviso, Bristol	match
C9	50989130	50989210	A10	15026481	15026556	Swede	match
C9	51295644	51295847	A10	16468542	16468744	Swede	match
C9	54391835	54391968	A10	17410729	17410860	H165	match

Table S5. Chromosome positions of 21 HE junctions in three banana triploid accessions (based on the position in B genome of *M. balbisiana* and A genome of *M. acuminata*).

Chr. B	Start B	End B	Chr. A	Start A	End A	Samples	Type[1]
Bchr04	6587091	6587327	chr04	6146789	6147023	FenJiao	match
Bchr04	37744538	37747294	chr04	32298270	32299199	FenJiao	match
Bchr04	40041836	40041924	chr04	34572494	34572581	FenJiao	match
Bchr05	37337424	37337574	chr05	36930151	36930295	Kamaramasenge	match
Bchr06	448664	449081	chr06	446637	447053	Pelipita	match
Bchr06	37939967	37940074	chr06	34352094	34352195	FenJiao	match
Bchr06	38491637	38492463	chr06	34869521	34870340	Pelipita	match
Bchr07	32999005	32999152	chr07	31552034	31552163	Pelipita	match
Bchr07	34657417	34657703	chr07	32977979	32978265	FenJiao	match
Bchr07	36326380	36326480	chr07	34528141	34528384	Pelipita	match
Bchr09	654679	654882	chr09	552362	552565	Pelipita	match
Bchr09	3307561	3307693	chr09	3114861	3114991	Kamaramasenge	match
Bchr09	4771870	4772168	chr09	4639024	4639321	Kamaramasenge	match
Bchr09	35336812	35337797	chr09	39158712	39159668	FenJiao	match
Bchr10	29235499	29235909	chr10	25413320	25413728	FenJiao	match
Bchr10	29930687	29930969	chr10	26072695	26072976	Kamaramasenge	match
Bchr10	31285433	31285778	chr10	27435983	27436327	Kamaramasenge	match
Bchr10	34815076	34815232	chr10	30812846	30813001	Pelipita	match
Bchr10	38388494	38388667	chr10	34238777	34238949	Kamaramasenge	match
Bchr11	2523247	2523323	chr11	2223384	2223459	Kamaramasenge	match
Bchr11	3492859	3493619	chr11	3110017	3110517	FenJiao	match

Table S6. Chromosome positions of 166 HE junctions in four synthetic rice tetraploid individuals (based on the position in genome of Nipponbare and 9311).

Chr. Nip	Start Nip	End Nip	Chr. 9311	Start 9311	End 9311	Samples	Type[1]
Chr1	1129918	1130735	9311_chr01	1184035	1184831	99NN-3, 99NN-7	match
Chr1	4157344	4157685	9311_chr01	4353397	4353736	NN99-4, NN99-8	match
Chr1	4985027	4985872	9311_chr01	5108472	5109321	NN99-8	match
Chr1	5536090	5536438	9311_chr01	5670451	5670798	NN99-8	match
Chr1	5684479	5684930	9311_chr01	5825188	5825638	NN99-8	match
Chr1	6102367	6109226	9311_chr01	6291653	6298511	NN99-4	match
Chr1	8332955	8333710	9311_chr01	8809221	8809975	NN99-4, NN99-8	match
Chr1	10636446	10637172	9311_chr01	11272652	11273337	99NN-3, 99NN-7	match
Chr1	14590366	14590889	9311_chr01	15554627	15555148	99NN-3, 99NN-7	match
Chr1	19986079	19986411	9311_chr01	21398260	21398592	NN99-4, NN99-8	match
Chr1	21203888	21204092	9311_chr01	22665214	22665417	99NN-3, 99NN-7	match
Chr1	23717497	23717750	9311_chr01	25599562	25599814	NN99-8	match
Chr1	30408738	30408947	9311_chr01	32855527	32855737	NN99-8	match
Chr1	32806839	32806983	9311_chr01	35290081	35290224	NN99-4	match
Chr1	35054686	35055822	9311_chr01	37642338	37643472	NN99-4, NN99-8	match
Chr1	36466146	36466229	9311_chr01	39053787	39053869	99NN-3, 99NN-7	match
Chr1	37636457	37636760	9311_chr01	40166663	40166965	99NN-3, 99NN-7	match
Chr1	40595040	40596220	9311_chr01	43331519	43332693	99NN-3, 99NN-7	match
Chr1	40822811	40822992	9311_chr01	43620665	43620843	99NN-3, 99NN-7	match
Chr1	42024667	42033150	9311_chr01	44896269	44903327	99NN-3, 99NN-7	match
Chr10	2357528	2357794	9311_chr10	2433505	2433770	99NN-7	match
Chr10	9919537	9920695	9311_chr10	9719676	9720818	99NN-7	match
Chr10	11006519	11008450	9311_chr10	10799253	10801183	99NN-3, 99NN-7	match
Chr10	14940844	14941607	9311_chr10	15439859	15440620	NN99-4, NN99-8	match
Chr11	1224893	1225023	9311_chr11	1311212	1311341	NN99-4, NN99-8	match
Chr11	1458320	1459347	9311_chr11	1563533	1564559	99NN-7	match
Chr11	3354848	3355306	9311_chr11	3615392	3615850	NN99-4, NN99-8	match
Chr11	6505875	6506847	9311_chr11	6764486	6765358	99NN-3	match
Chr11	8989586	8990682	9311_chr11	8803805	8804892	99NN-3, 99NN-7	match
Chr11	15562005	15562218	9311_chr11	15894881	15895093	NN99-4	match
Chr11	16453477	16454189	9311_chr11	16976700	16977412	NN99-4	match
Chr11	17387697	17388993	9311_chr11	18070010	18071305	NN99-4, NN99-8	match
Chr11	25277115	25278035	9311_chr11	26875583	26876499	99NN-3, 99NN-7	match
Chr11	26432840	26433995	9311_chr11	27929386	27930533	99NN-3, 99NN-7	match
Chr12	1032308	1034434	9311_chr12	1021608	1023736	99NN-3, 99NN-7	match
Chr12	1437869	1437925	9311_chr12	1454193	1454248	99NN-3, 99NN-7	match
Chr12	2304970	2305577	9311_chr12	2268447	2269037	NN99-4, NN99-8	match
Chr12	4760928	4761191	9311_chr12	4714751	4715012	99NN-3, 99NN-7	match
Chr12	18101987	18102657	9311_chr12	17023965	17024636	NN99-4	match
Chr12	20781709	20781877	9311_chr12	19238131	19238298	99NN-3	match
Chr12	23920185	23920710	9311_chr12	22333373	22333894	NN99-8	match

Chr12	24184905	24185658	9311_chr12	22588876	22589592	99NN-3, 99NN-7	match
Chr2	913255	914179	9311_chr02	944414	945336	99NN-7	match
Chr2	983271	985248	9311_chr02	1018721	1020700	99NN-3	match
Chr2	2878228	2884337	9311_chr02	3115332	3121447	99NN-3, 99NN-7	match
Chr2	3760348	3761354	9311_chr02	4018044	4019050	99NN-3, 99NN-7	match
Chr2	4367221	4368114	9311_chr02	4715322	4716217	99NN-3, 99NN-7	match
Chr2	5862539	5863144	9311_chr02	6429265	6429868	99NN-3, 99NN-7	match
Chr2	7422756	7423008	9311_chr02	8012608	8012859	99NN-3	match
Chr2	9100300	9101456	9311_chr02	9802304	9804123	NN99-4, NN99-8	match
Chr2	9885768	9887894	9311_chr02	10422434	10424560	99NN-7	match
Chr2	10123621	10124329	9311_chr02	10673088	10673795	99NN-7	match
Chr2	10323633	10324031	9311_chr02	10854561	10854957	NN99-4, NN99-8	match
Chr2	11687200	11687495	9311_chr02	12306927	12307221	NN99-4	match
Chr2	12218696	12219554	9311_chr02	12766873	12767730	NN99-8	match
Chr2	18815954	18816533	9311_chr02	19304579	19305157	99NN-3	match
Chr2	21380695	21384471	9311_chr02	21958751	21962526	99NN-3, 99NN-7	match
Chr2	23122982	23123294	9311_chr02	23685187	23685498	NN99-4	match
Chr2	24013723	24015938	9311_chr02	24582986	24585195	NN99-4, NN99-8	match
Chr2	24196299	24196512	9311_chr02	24759135	24759347	NN99-8	match
Chr2	25445009	25446106	9311_chr02	25960455	25961543	NN99-8	match
Chr2	34368907	34369522	9311_chr02	34805513	34806126	99NN-3, 99NN-7	match
Chr2	35137218	35139423	9311_chr02	35625106	35627309	NN99-4, NN99-8	match
Chr2	35418413	35420009	9311_chr02	35929576	35931191	99NN-3, 99NN-7	match
Chr2	35676588	35677386	9311_chr01	423747	424522	99NN-7	other_chr
Chr3	386243	388586	9311_chr03	449557	450974	99NN-3, 99NN-7	match
Chr3	3237650	3238887	9311_chr03	3344116	3345722	NN99-4, NN99-8	match
Chr3	3398773	3398880	9311_chr03	3517353	3517459	99NN-3, 99NN-7	match
Chr3	3730513	3731195	9311_chr03	3849500	3850181	NN99-4, NN99-8	match
Chr3	4235630	4236121	9311_chr03	4325267	4325758	99NN-3, 99NN-7	match
Chr3	6455291	6455554	9311_chr03	6600626	6600888	99NN-3, 99NN-7	match
Chr3	8344310	8345575	9311_chr03	8521968	8523233	NN99-4, NN99-8	match
Chr3	9425774	9425896	9311_chr03	9713531	9713652	NN99-4, NN99-8	match
Chr3	10358060	10359106	9311_chr03	10825823	10826846	NN99-4	match
Chr3	12417425	12418116	9311_chr03	13329537	13330218	NN99-4, NN99-8	match
Chr3	13765767	13767021	9311_chr03	13768839	13769747	99NN-3, 99NN-7	match
Chr3	15792029	15792138	9311_chr03	17059628	17059736	NN99-4, NN99-8	match
Chr3	16731513	16732914	9311_chr03	18213603	18215003	99NN-3, 99NN-7	match
Chr3	24300330	24300822	9311_chr03	27056630	27057121	99NN-3, 99NN-7	match
Chr3	27618636	27624348	9311_chr03	30587227	30590573	NN99-8	match
Chr3	27838445	27838835	9311_chr03	30792106	30792495	NN99-4	match
Chr3	31749168	31750285	9311_chr03	34957957	34959073	NN99-4, NN99-8	match
Chr3	32617863	32619300	9311_chr03	35821512	35822935	NN99-4, NN99-8	match
Chr3	33805640	33806078	9311_chr03	37179699	37180136	99NN-3, 99NN-7	match
Chr3	36064725	36064887	9311_chr03	39508442	39508603	NN99-4, NN99-8	match

Chr4	1795488	1795507	9311_chr04	2268764	2269081	NN99-4	match
Chr4	4526487	4526609	9311_chr04	3849208	3849329	99NN-7	match
Chr4	5210661	5211364	9311_chr04	4579387	4580091	NN99-8	match
Chr4	5608200	5608648	9311_chr04	4952638	4953076	99NN-3	match
Chr4	6692414	6692909	9311_chr04	6031861	6032355	99NN-7	match
Chr4	7621177	7628603	9311_chr04	6847456	6848963	99NN-3	match
Chr4	14156258	14156420	9311_chr04	13343773	13343934	99NN-3	match
Chr4	19991419	19992451	9311_chr04	19475440	19476470	99NN-3, 99NN-7	match
Chr4	23024048	23030179	9311_chr04	22968677	22974808	NN99-8	match
Chr4	23077402	23082258	9311_chr04	23022044	23026900	99NN-3	match
Chr4	26790491	26796353	9311_chr04	27285169	27291029	99NN-3	match
Chr4	27131669	27132284	9311_chr04	27643140	27643754	99NN-7	match
Chr4	27285452	27286191	9311_chr04	27824211	27824950	NN99-8	match
Chr4	27948627	27949076	9311_chr04	28516860	28517308	NN99-4, NN99-8	match
Chr4	28201453	28202465	9311_chr04	28757241	28758248	99NN-3	match
Chr4	30281846	30282502	9311_chr04	30965072	30965728	99NN-3	match
Chr4	34230146	34239322	9311_chr04	35223316	35232493	99NN-3, 99NN-7	match
Chr5	758518	759097	9311_chr05	777958	778531	99NN-7	match
Chr5	1120146	1121800	9311_chr05	1196244	1197860	99NN-3, 99NN-7	match
Chr5	2608352	2609415	9311_chr05	2810677	2811737	NN99-8	match
Chr5	3511584	3512136	9311_chr05	3646067	3646618	NN99-4	match
Chr5	4830718	4831000	9311_chr05	4911035	4911323	NN99-4	match
Chr5	5376361	5376786	9311_chr05	5497807	5498232	99NN-3, 99NN-7	match
Chr5	6397229	6397914	9311_chr05	6462184	6462835	NN99-4	match
Chr5	7927370	7927740	9311_chr05	8189957	8190326	99NN-3	match
Chr5	7944527	7944600	9311_chr05	8206516	8206688	NN99-4, NN99-8	match
Chr5	14734313	14734449	9311_chr05	14980352	14980487	99NN-7	match
Chr5	16168675	16169578	9311_chr05	16285485	16286381	99NN-7	match
Chr5	17317984	17318960	9311_chr05	17603786	17604763	NN99-4	match
Chr5	18185538	18186274	9311_chr05	18592228	18592963	99NN-7	match
Chr5	20441630	20441687	9311_chr05	20972731	20972787	NN99-4, NN99-8	match
Chr5	21096330	21096333	9311_chr03	33155295	33155353	99NN-3	other_chr
Chr5	21550499	21553231	9311_chr05	22216186	22217353	99NN-7	match
Chr5	26951405	26951791	9311_chr05	27655586	27655971	NN99-4, NN99-8	match
Chr5	27897316	27898009	9311_chr05	28509997	28510691	NN99-4, NN99-8	match
Chr6	297212	299550	9311_chr06	227091	229427	99NN-3, 99NN-7	match
Chr6	5536745	5541462	9311_chr06	5667184	5669712	NN99-4, NN99-8	match
Chr6	7421037	7421905	9311_chr06	7739798	7740665	NN99-4, NN99-8	match
Chr6	9676469	9676562	9311_chr06	9865994	9866086	99NN-3, 99NN-7	match
Chr6	10922005	10926880	9311_chr06	10963059	10967934	99NN-3, 99NN-7	match
Chr6	21210749	21210860	9311_chr06	21241968	21242078	NN99-4	match
Chr6	21398092	21399255	9311_chr06	21424364	21425526	NN99-4	match
Chr6	24031147	24031513	9311_chr06	24084373	24084738	NN99-4	match
Chr6	24233617	24235823	9311_chr06	24279539	24281744	NN99-4	match

Chr6	27159932	27168661	9311_chr06	27621283	27628557	NN99-4, NN99-8	match
Chr6	28341054	28341271	9311_chr06	28884301	28884509	NN99-4, NN99-8	match
Chr7	4108705	4109662	9311_chr07	3859853	3860813	99NN-3, 99NN-7	match
Chr7	6042827	6044024	9311_chr07	5875674	5876869	NN99-4, NN99-8	match
Chr7	7178011	7178699	9311_chr07	6920679	6921365	99NN-3, 99NN-7	match
Chr7	16374813	16376299	9311_chr07	15444918	15446444	NN99-4, NN99-8	match
Chr7	17112810	17113185	9311_chr07	16119189	16119579	NN99-4	match
Chr7	17128236	17129544	9311_chr07	16141812	16143120	99NN-3, 99NN-7	match
Chr7	17179918	17180039	9311_chr07	16191719	16191839	NN99-8	match
Chr7	17376768	17377185	9311_chr07	16395612	16396024	NN99-8	match
Chr7	19480082	19481088	9311_chr07	18271054	18272071	NN99-8	match
Chr7	21658823	21659457	9311_chr07	20296530	20297163	99NN-3	match
Chr7	23843198	23843476	9311_chr07	22453708	22453985	NN99-4, NN99-8	match
Chr7	27177326	27178010	9311_chr07	25754806	25755493	99NN-7	match
Chr7	28868015	28868560	9311_chr07	27372995	27373539	99NN-3, 99NN-7	match
Chr8	5644251	5645069	9311_chr08	5626399	5627210	99NN-7	match
Chr8	6125024	6126276	9311_chr08	6245419	6246670	NN99-8	match
Chr8	8942693	8945911	9311_chr08	9445282	9448476	99NN-3, 99NN-7	match
Chr8	11034025	11035401	9311_chr08	12250322	12251696	NN99-8	match
Chr8	19464110	19465251	9311_chr08	21584746	21585534	NN99-4, NN99-8	match
Chr8	20635802	20636924	9311_chr08	22728137	22729259	99NN-3, 99NN-7	match
Chr8	21246798	21247244	9311_chr08	23338474	23338919	NN99-8	match
Chr8	21813604	21815060	9311_chr08	23934617	23936072	99NN-3, 99NN-7	match
Chr8	23657122	23657342	9311_chr08	25990865	25991084	99NN-7	match
Chr9	5599596	5600217	9311_chr09	5365432	5366052	NN99-4, NN99-8	match
Chr9	6230447	6230680	9311_chr09	6089856	6090089	99NN-7	match
Chr9	7481224	7481910	9311_chr09	7359392	7360077	99NN-7	match
Chr9	10805352	10806135	9311_chr09	10045490	10046272	NN99-4, NN99-8	match
Chr9	12143280	12144850	9311_chr09	11361572	11363141	NN99-4	match
Chr9	13779228	13779433	9311_chr09	13085281	13085485	NN99-8	match
Chr9	13911151	13918860	9311_chr09	13221169	13228876	NN99-4	match
Chr9	14980713	14981215	9311_chr09	14177249	14177752	99NN-3, 99NN-7	match
Chr9	15021225	15021411	9311_chr09	14212809	14212994	NN99-4, NN99-8	match
Chr9	16422504	16422984	9311_chr09	15655571	15656038	NN99-4	match
Chr9	16560613	16561309	9311_chr09	15781313	15782009	NN99-8	match
Chr9	17694381	17695286	9311_chr09	16989111	16990015	99NN-3, 99NN-7	match
Chr9	18618052	18619097	9311_chr09	17933762	17934806	NN99-4, NN99-8	match

Table S7. Chromosome positions of 10 HE junctions in *Arabidopsis suecica* individuals (based on the position in genome of *A. thaliana*).

Chr	Start	End	Samples
Chr1	658958	658976	SRR2084157, SRR3123760, SRR3123765
Chr1	732181	732288	SRR2084157, SRR3123760, SRR3123765
Chr1	2813704	2813752	SRR3123767
Chr1	4929662	4929738	SRR3123767
Chr2	19379436	19379776	SRR3123767
Chr2	19564121	19564172	SRR3123764, SRR3123769
Chr4	1068487	1068604	SRR3123760, SRR2084157
Chr4	2240257	2240325	SRR2084157, SRR3123760
Chr4	2786695	2792022	SRR2084157, SRR3123760

Table S8. Chromosome positions of 26 HE junctions in *Arachis hypogaea* individuals (based on the position in B genome of *A. hypogaea*).

Chr. B	Start B	End B	Samples
Arahy.12	117668999	117704310	Amonticola
Arahy.12	117817965	117822170	NC3033, N423, N421, N464, N692, N705, N693, N689, N717, N734, N726, N725, N711, N745, N709, N713
Arahy.12	118256778	118258164	TR, N744, N657, N625, N683, N675, N665, N663, N431, N741, N435, N733, N704, N702, N743, N700, NM, N730, N690, N740, N723, N751
Arahy.13	146343047	146347951	All samples
Arahy.14	77162579	77168419	N431, N741, N423, N421, N435
Arahy.14	90393195	90440728	N431, N741, N423, N421, N435
Arahy.14	137403639	137406947	N744, N709
Arahy.14	138862257	138863838	Amonticola
Arahy.14	139404845	139413753	Amonticola
Arahy.14	139899365	139905900	N733, N745
Arahy.14	140466186	140468348	N733
Arahy.14	140572534	140582341	N657, N700, N717, N725, N711, N745
Arahy.14	140582345	140589736	NC3033, N625, N683, N665, N663, N704, N702, N693, N713
Arahy.14	141430181	141442130	NC3033, N657, N625, N683, N675, N665, N663, N704, N702, N693, N700, N717, N713
Arahy.15	6306066	6308629	All samples
Arahy.16	3202449	3210486	TR, N683, N689, N700, N717, N734, N690
Arahy.16	146245083	146267448	NC3033, N683, N665, N431, N741, N423, N692, N743, N713
Arahy.16	151628012	151636887	NC3033
Arahy.16	152284902	152286523	TR, N713
Arahy.16	152723849	152730064	Amonticola
Arahy.16	152819251	152836728	TR, N657, N625, N683, N675, N665, N663, N464, N704, N705, N743, N689, N700, N734, N713, N723
Arahy.16	153003860	153007978	Amonticola, N744, N431, N741, N423, N421, N435, N692, N733, N702, N693, N717, NM, N726, N725, N711, N730, N745, N709, N690, N740, N751
Arahy.17	1438055	1452535	All samples
Arahy.17	1839088	1855636	N733
Arahy.17	2291088	2312052	N733
Arahy.18	1127952	1130739	All samples

Table S9. Genomic feature of HE-junction-related motifs in wheat, Brassica and rice.

A-rich motif

	Promoter (up-stream 2kb of TSS)	UTR and intron	Exon	Intergenic region
Brassica	10 (28%)	14 (40%)	5 (14%)	6 (18%)
Rice	26 (22%)	24 (21%)	11 (9%)	56 (49%)

CTT- or CT-repeat motif

	Promoter (up-stream 2kb of TSS)	UTR and intron	Exon	Intergenic region
Brassica	13 (13%)	19 (19%)	61 (61%)	9 (8%)
Rice	21 (24%)	25 (28%)	5 (6%)	37 (42%)

CCN-repeat motif

	Promoter (up-stream 2kb of TSS)	UTR and intron	Exon	Intergenic region
Wheat	6 (19%)	4 (13%)	17 (55%)	4 (13%)
Rice	14 (18%）	13 (17%)	18 (23%)	32 (42%)