Supplemental Files

Evidence for selfing in a vertebrate from whole-genome sequencing

Astrid Böhne, Zeynep Oğuzhan, Ioannis Chrysostomakis, Simon Vitt, Denis Meuthen, Sebastian Martin, Sandra Kukowka, Timo Thünken

This file contains
Supplemental Tables S1-S4
and
Supplemental Figures S1-S18

All code and auxiliary data to reproduce the analysis can be found in the files GenericCode.zip and GenomicData.zip. The whole genome sequencing data generated in this study can be accessed under the NCBI BioProject number PRJNA832128.

Supplemental Table S1

Genome estimates based on kmer analyses

Individual	Kmer size	Estimated genome size in bp	Estimated heterozygosity	Mean genome size in bp	Mean heterozygosity in %
wildI	20	819’370’356	0.04	817‘073‘328	0.0974
	30	816’988’795	0.0956		
	40	814’860’834	0.0927		
parent	20	813‘834‘447	0.156	811‘993‘447	0.1193
	30	812’108’800	0.104		
	40	810’037’093	0.098		
offspring1	20	801’613’321	0.0955	799‘791‘252	0.0838
	30	799’746’832	0.0811		
	40	798’013’603	0.0748		
offspring2	20	816’282’648	0.0903	814‘193‘387	0.0791
	30	814’136’320	0.0759		
	40	812’161’193	0.0712		

Supplemental Table S2

Heterozygosity across individuals

Variant sites					
Sample ID	Number of sites homozygous reference	Number of sites homozygous alternative	Number of sites heterozygous	Percentage heterozygosity	Expected number of heterozygous sites
wildI	73’821	69’243	225’389	61%	206’104
mother	72’623	68’819	227’011	62%	206’104
child1	109’950	105’698	152’805	41%	206’104
child2	109’906	106’070	152’477	41%	206’104
All accessible sites					
Sample ID	Number of sites homozygous reference	Number of sites homozygous alternative	Number of sites heterozygous	Percentage heterozygosity	
wildI	433’616’448	21’268’745	287’320	0.0631%	
mother	433’615’228	21’268’747	288’752	0.0634%	
child1	433’652’597	21’318’300	193’557	0.0425%	
child2	433’652’573	21’318’989	192’728	0.0423%	

Supplemental Table S3

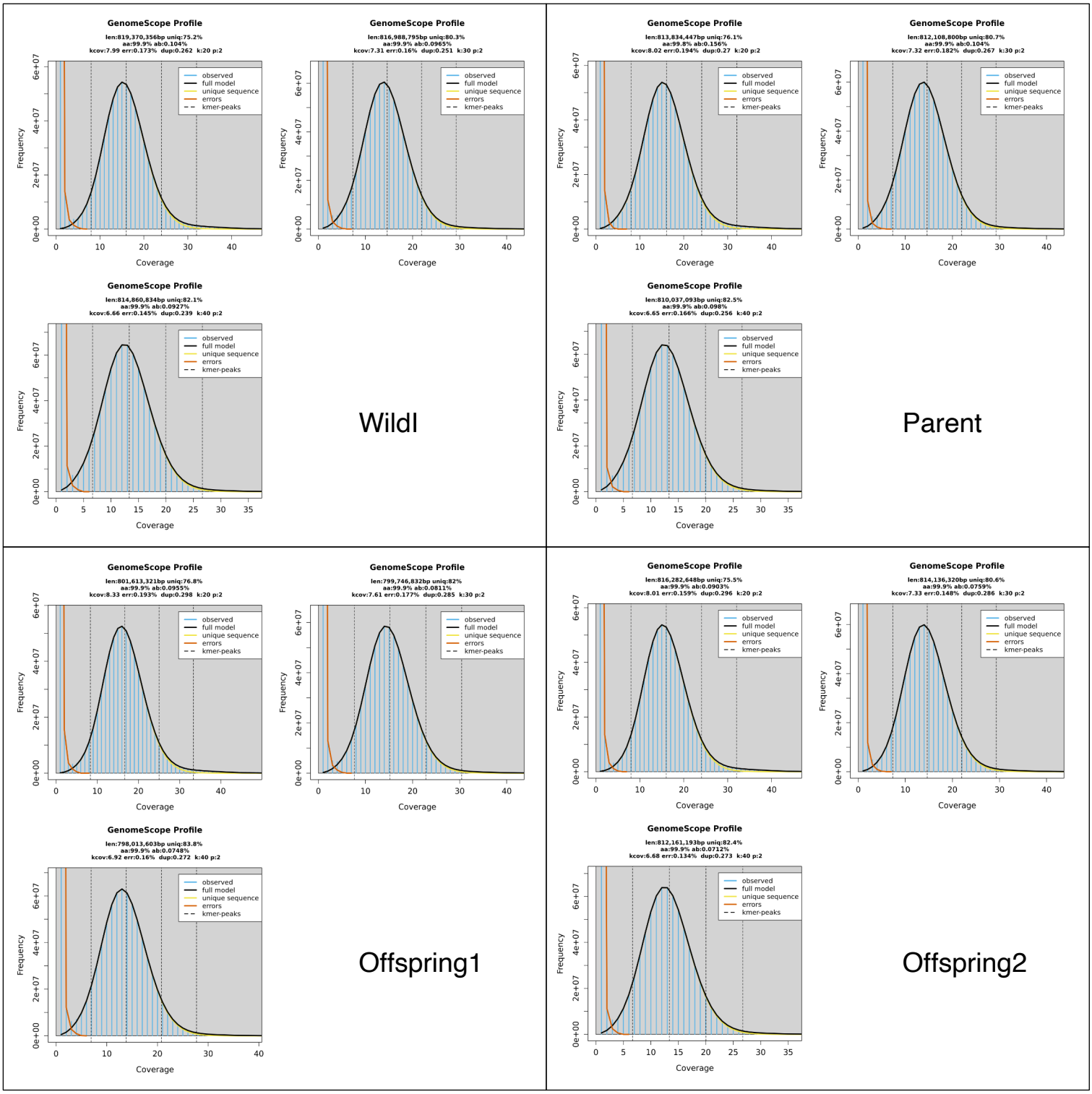
Retained heterozygosity in both offspring at heterozygous sites of the mother when applying different lower and upper limits to the individual depth and removing sites outside filter limits in any individual.

Filter settings	Overall sites with applied filter	Heterozygous sites in the parent with applied filter	Percentage heterozygous sites in the parent with applied filter	Offspring1 percentage retained heterozygous sites	Offspring2 percentage retained heterozygous sites
DP > 3	334’371	204’547	61.17%	59.05%	59.26%
DP > 8	302’417	183’075	60.53%	58.65%	58.92%
DP > 15	94’411	64’755	68.59%	70.28%	71.85%
DP > 3 < 20	175’690	101’999	58.06%	53.33%	52.72%
DP > 8 < 20	153’947	86’284	56.05%	51.40%	50.83%
DP > 15 < 20	29’933	16’768	56.02%	50.64%	52.88%
DP > 3 < 25	278’982	162’087	58.1%	53.50%	53.70%
DP > 8 < 25	254’709	144’584	56.76%	52.34%	52.60%
DP > 15 < 25	79’390	45’831	57.73%	53.92%	56.54%
DP > 3 < 30	310’088	182’882	58.98%	55.13%	55.39%
DP > 8 < 30	285’287	164’997	57.84%	54.26%	54.58%
DP > 15 < 20	99’312	59’893	60.31%	58.34%	60.60%
				mean (±SD): 55.90 ± 5.31 %	mean (±SD): 56.67 ± 5.67 %

Supplemental Table S4

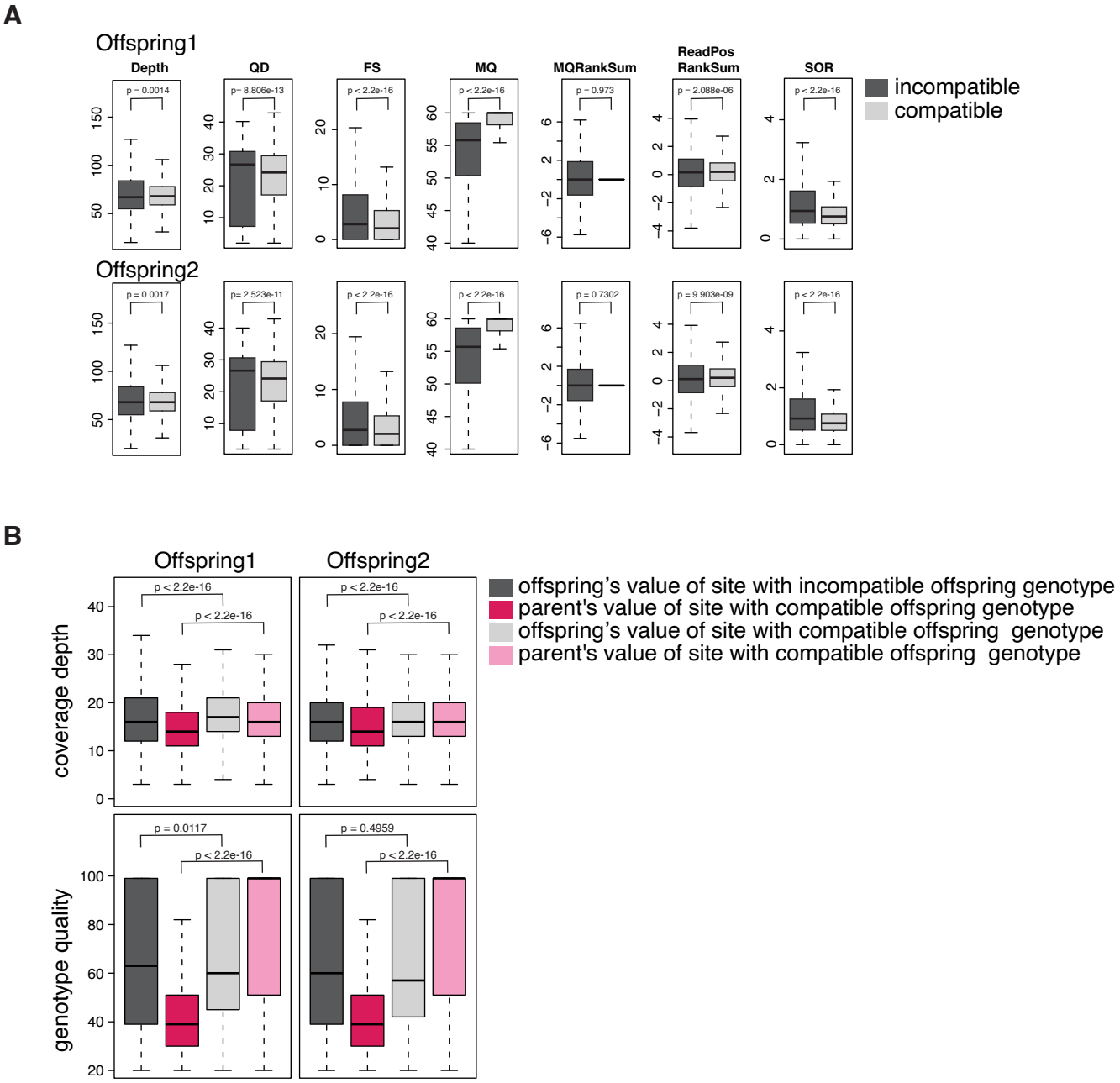
Genotypic state in both offspring at inferred centromeric regions based on Fig. 3 and Supplemental Fig. S18.

Chromosome	Offspring1 centromeric state	Offspring2 centromeric state
LG03	homozygous	homozygous
LG04	unclear (missing data)	unclear (missing data)
LG05	heterozygous	homozygous
LG06	probably heterozygous (missing data)	probably heterozygous (missing data)
LG07	heterozygous	homozygous
LG08	unclear (missing data)	unclear (missing data)
LG09	at recombination breakpoint	at recombination breakpoint
LG11	unclear (missing data)	unclear (missing data)
LG13	heterozygous	heterozygous
LG14	heterozygous	homozygous
LG15	heterozygous	heterozygous
LG16	probably heterozygous (missing data)	probably heterozygous (missing data)
LG17	heterozygous	heterozygous
LG18	homozygous	heterozygous
LG19	probably heterozygous (missing data & at chromosome end)	probably heterozygous (missing data & at chromosome end)
LG23	homozygous (missing data & at chromosome end)	heterozygous (missing data & at chromosome end)



Supplemental Figure S1

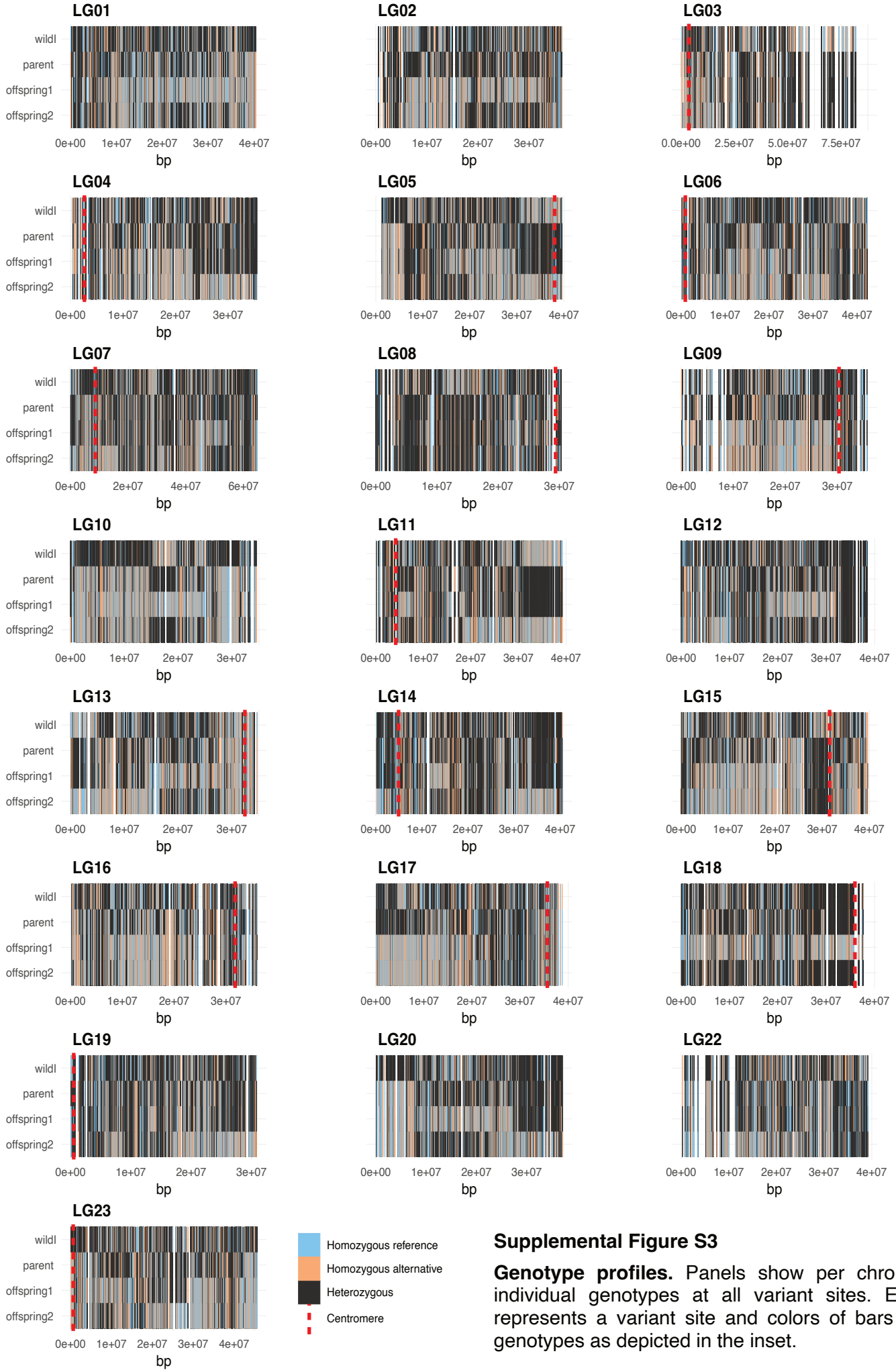
Genome profiles Kmer count profiles of raw sequencing data for all individuals and kmer sizes 20, 30, and 40.



Supplemental Figure S2

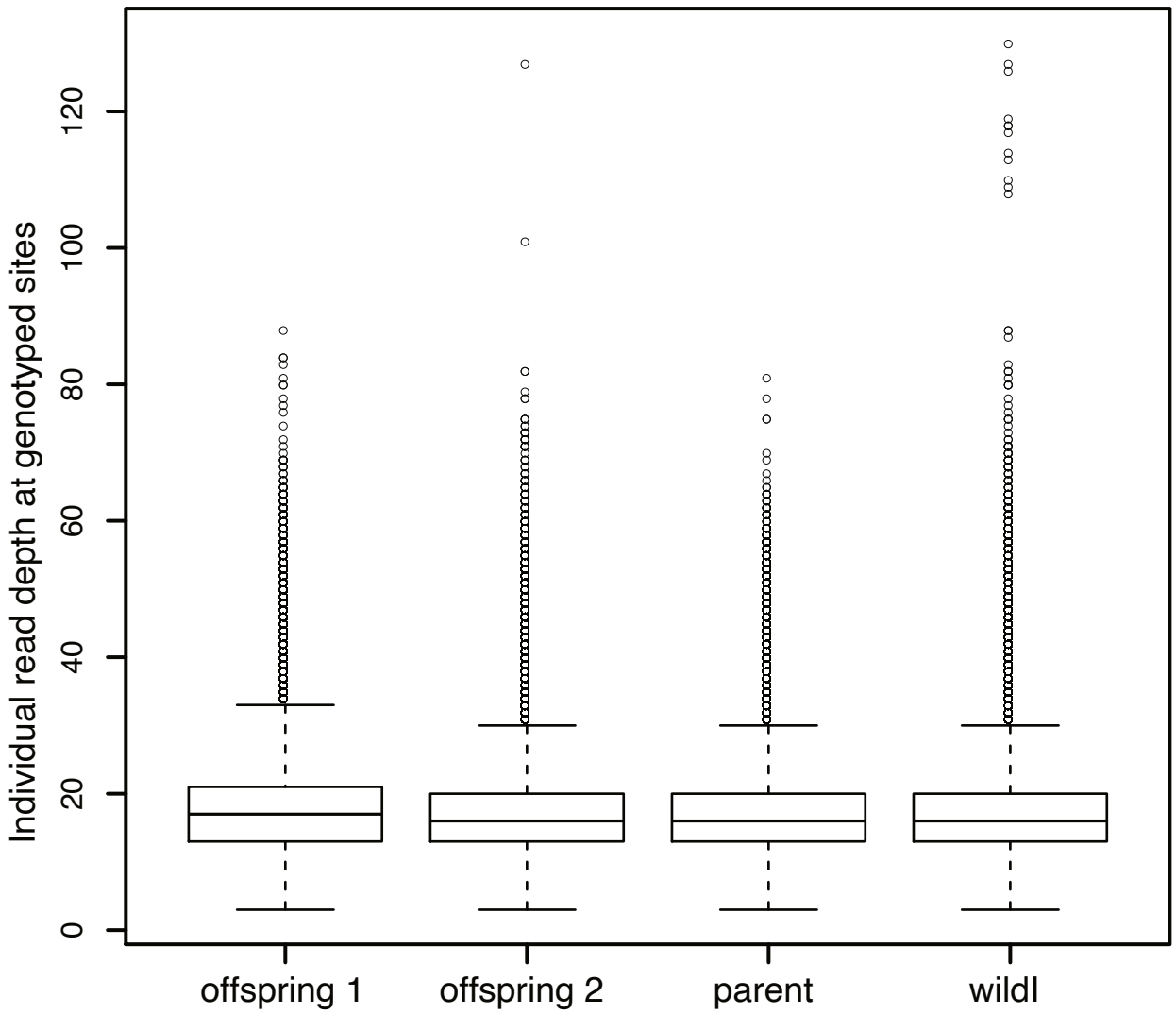
Genotype statistics A) Statistics per site across all individuals are shown for sites with alleles that are incompatible (dark grey) and compatible (light grey) with mono-parental origin in offspring1 (upper panel) and offspring2 (lower panel). Depth - Depth of coverage over all samples; QD - Variant confidence normalized by unfiltered depth of variant samples; ExcessHet - Phred-scaled p-value for exact test of excess heterozygosity; FS - Strand bias estimated using Fisher's exact test; MQ – RMS mapping quality; MQRankSum - Rank sum test for mapping qualities of reference allele *versus* alternative allele reads; ReadPosRankSum - Rank sum test for relative positioning of reference *versus* alternative alleles within reads; SOR - allele-specific strand bias estimated by the symmetric odds ratio test. **B)** Individual coverage depth (upper row) and genotype quality (lower row) for sites in offspring1 (left site) and offspring2 (right site) with genotypes in the respective offspring that are incompatible (dark colors) and compatible (light colors) with mono-parental origin.

Outliers are not shown, box plot centre lines represent the median, box limits represent the upper and lower quartiles, and whiskers represent 1.5× the interquartile range. Resulting p-values of Wilcoxon tests are indicated for each comparison.



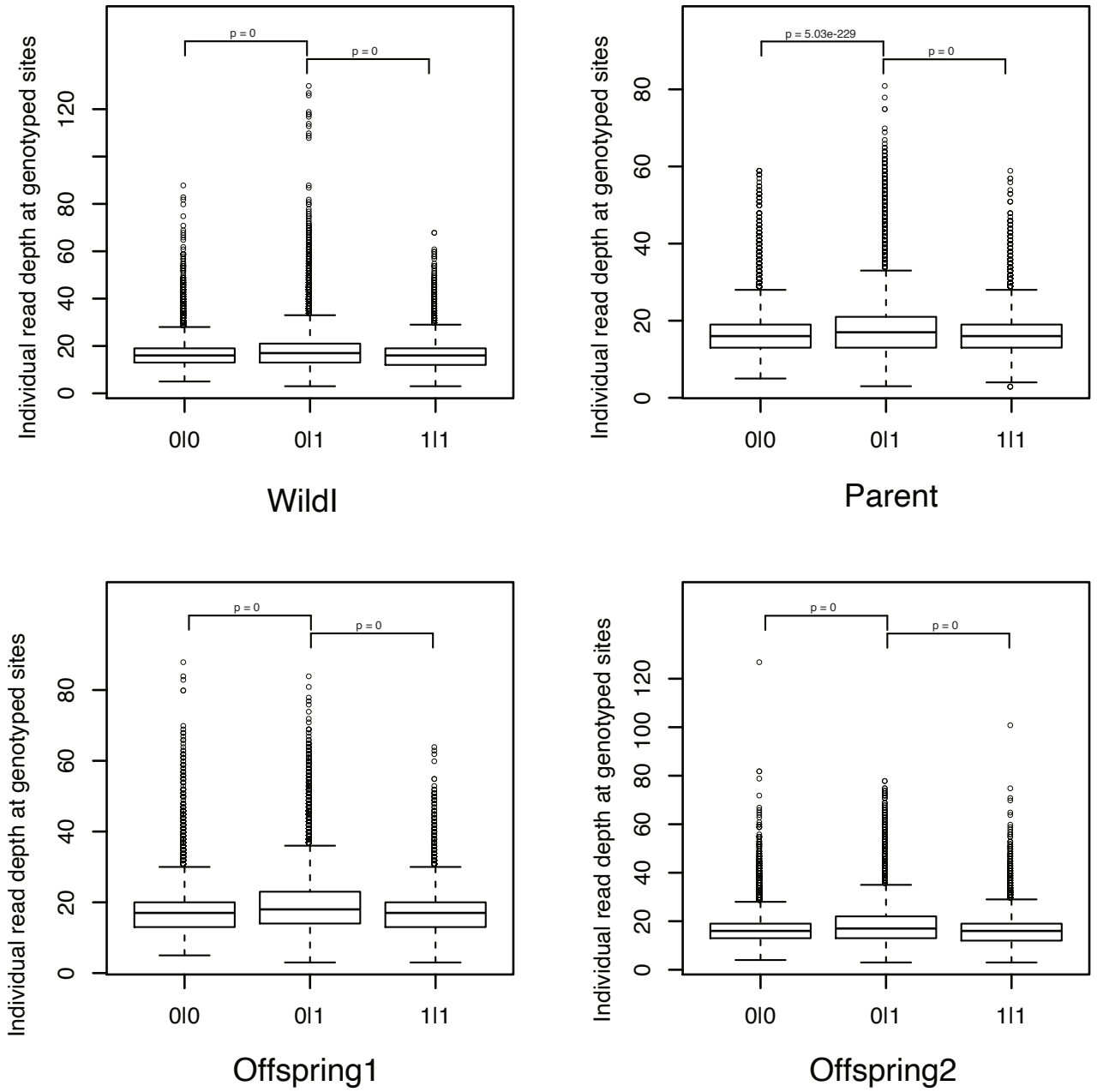
Supplemental Figure S3

Genotype profiles. Panels show per chromosome individual genotypes at all variant sites. Each bar represents a variant site and colors of bars indicate genotypes as depicted in the inset.



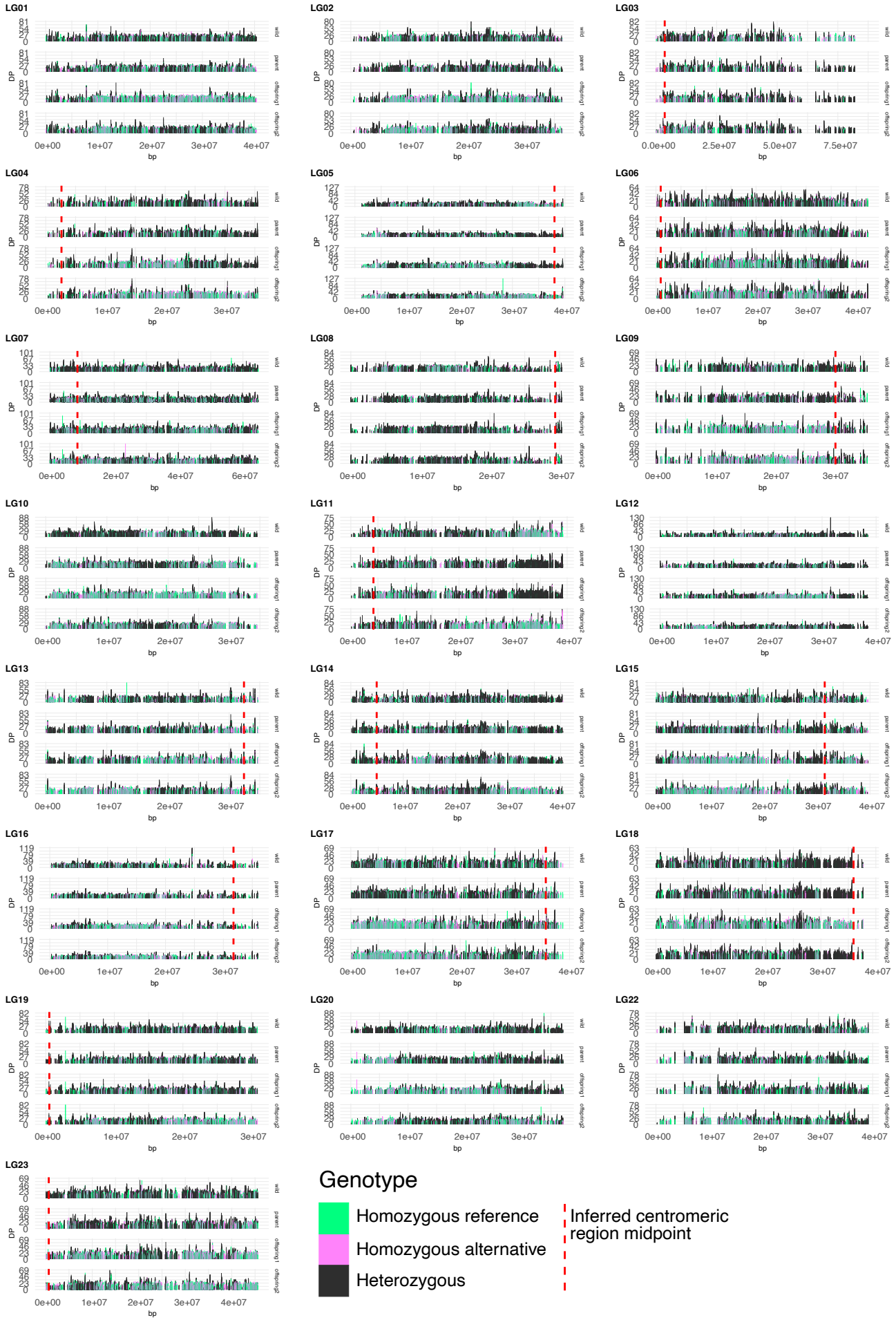
Supplemental Figure S4

Depth profiles Per individual read coverage of variant sites.



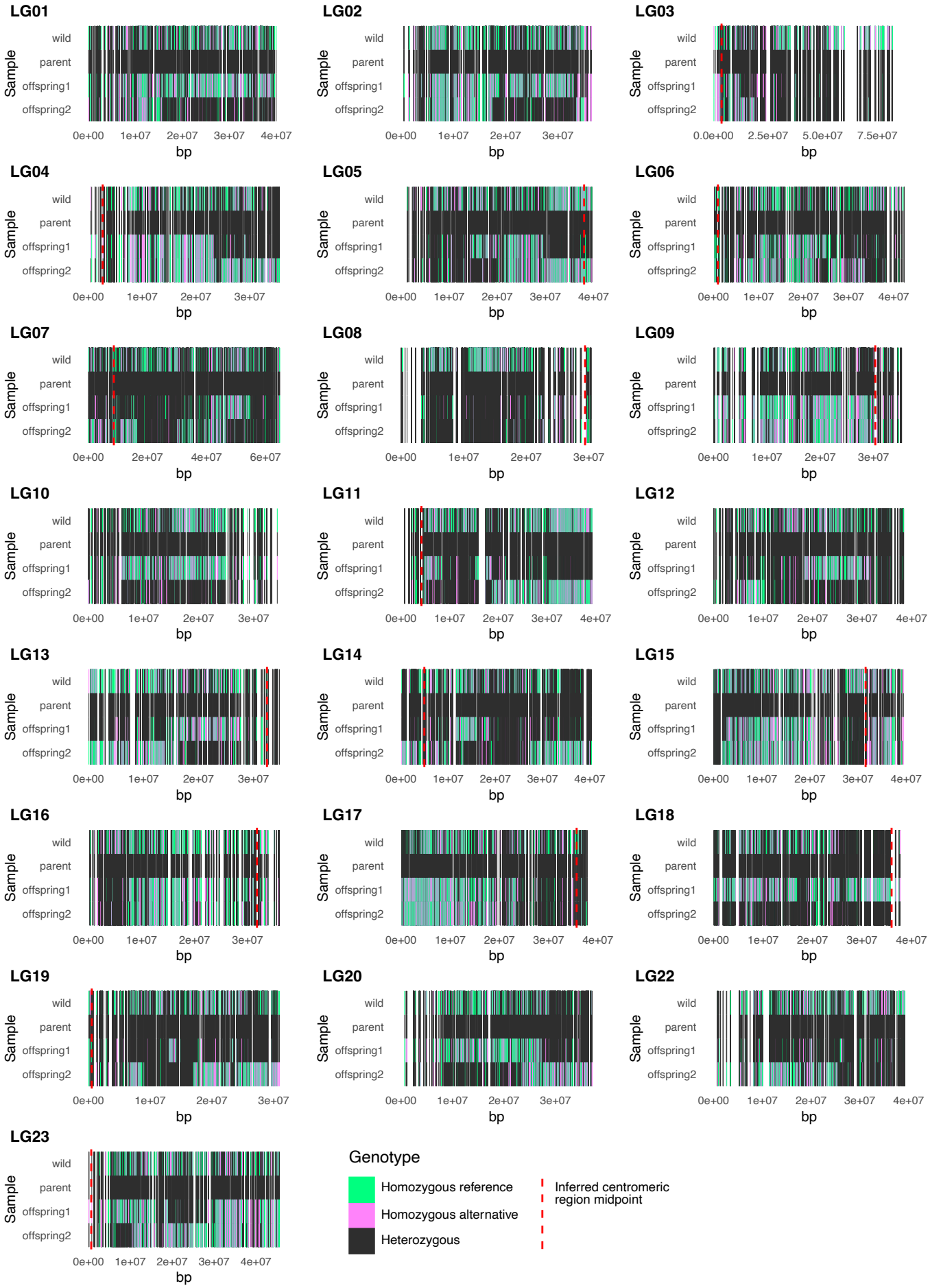
Supplemental Figure S5

Depth profiles Per individual read coverage at different genotypes of variant sites. Shown adjusted p-values are derived from pairwise Dunn's tests of multiple comparisons as post-hoc tests following group applied Kruskal-Wallis rank sum tests.



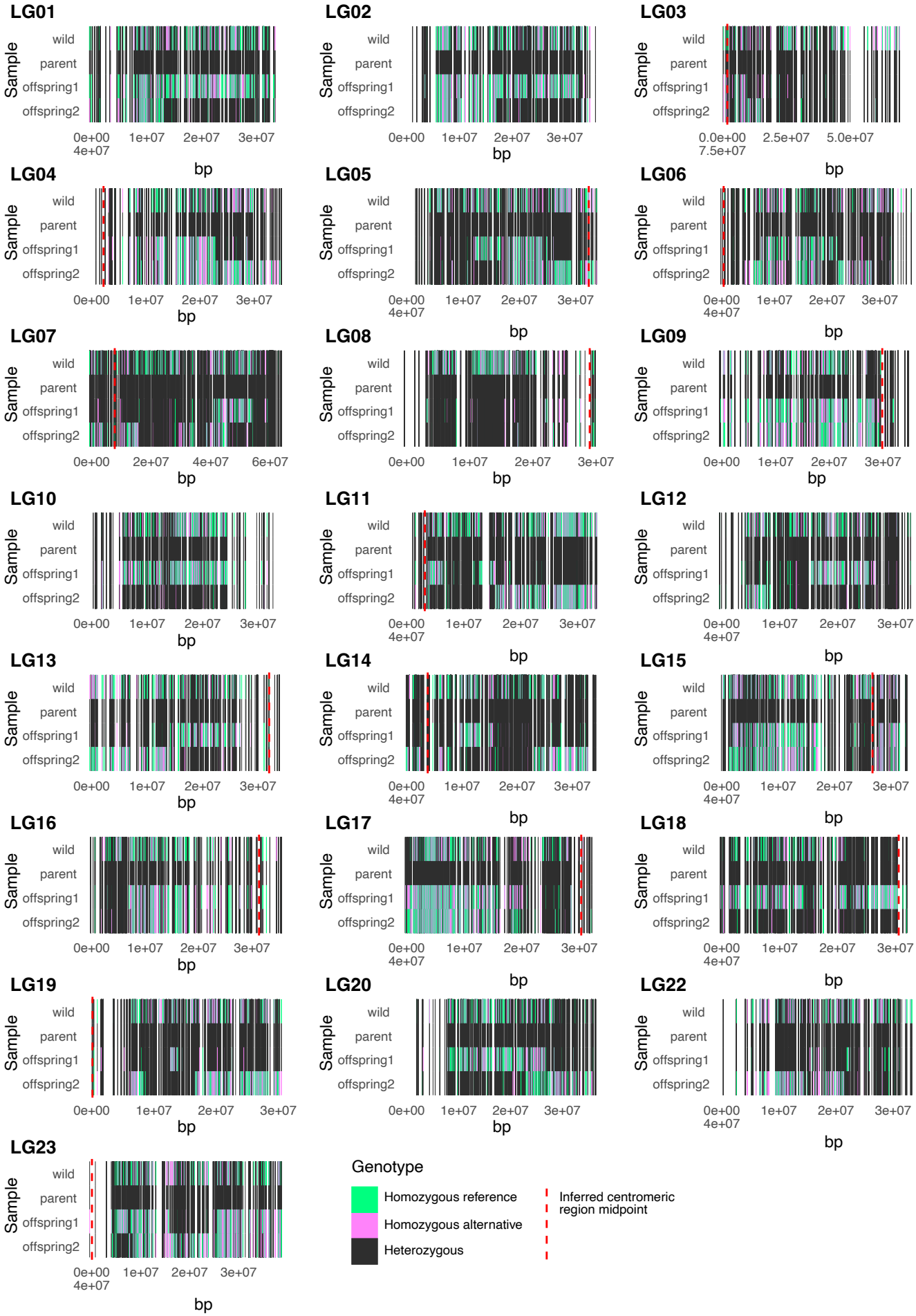
Supplemental Figure S6

Depth profiles Panels show per individual read coverage along chromosomes. Each bar represents a variant site and colors of bars indicate genotypes as depicted in the inset.



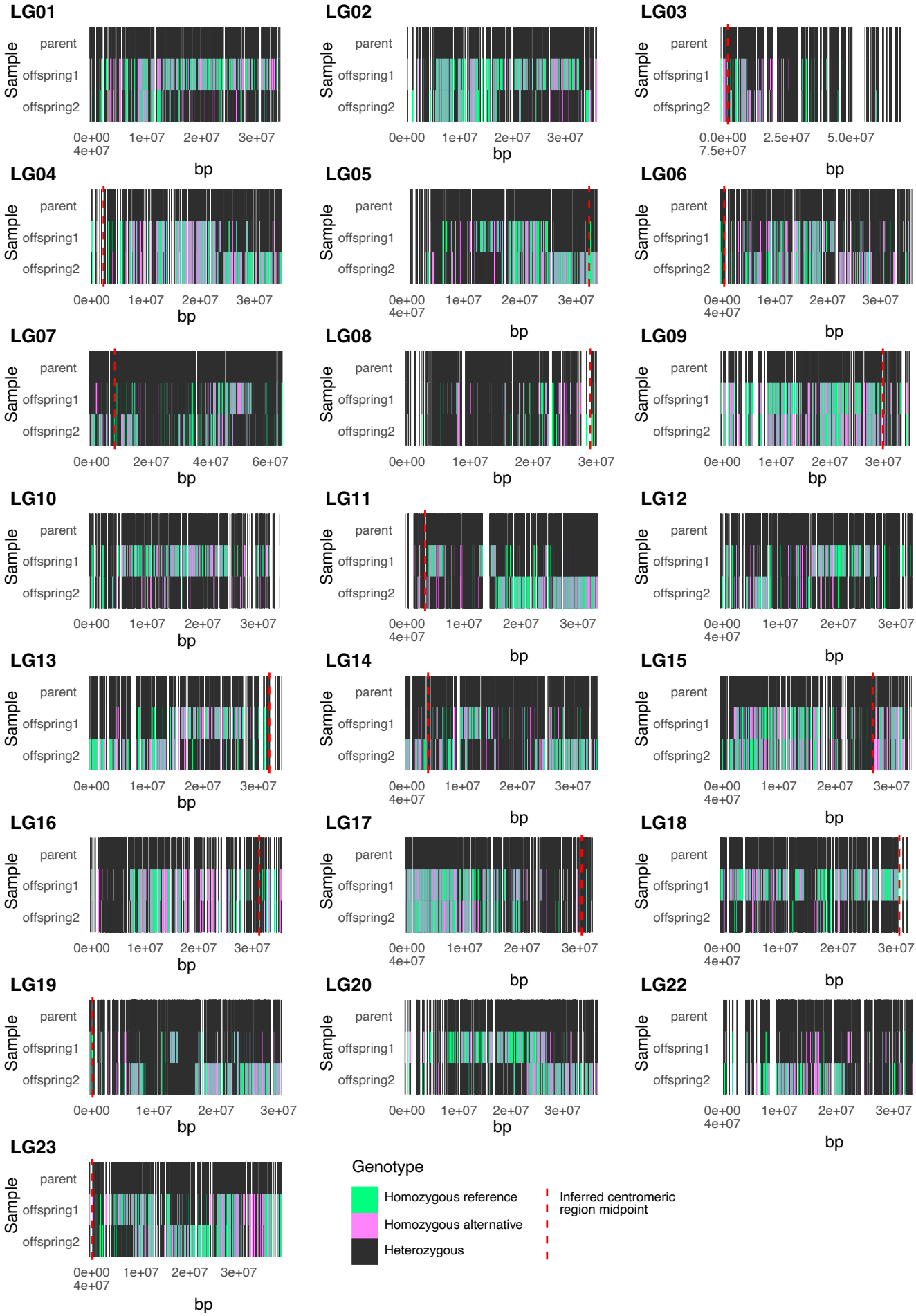
Supplemental Figure S7

Genotype profiles Panels show per chromosome individual genotypes at heterozygous sites of the parent with an individual depth filter of DP > 8. Each bar represents a variant site and colors of bars indicate genotypes as depicted in the inset.



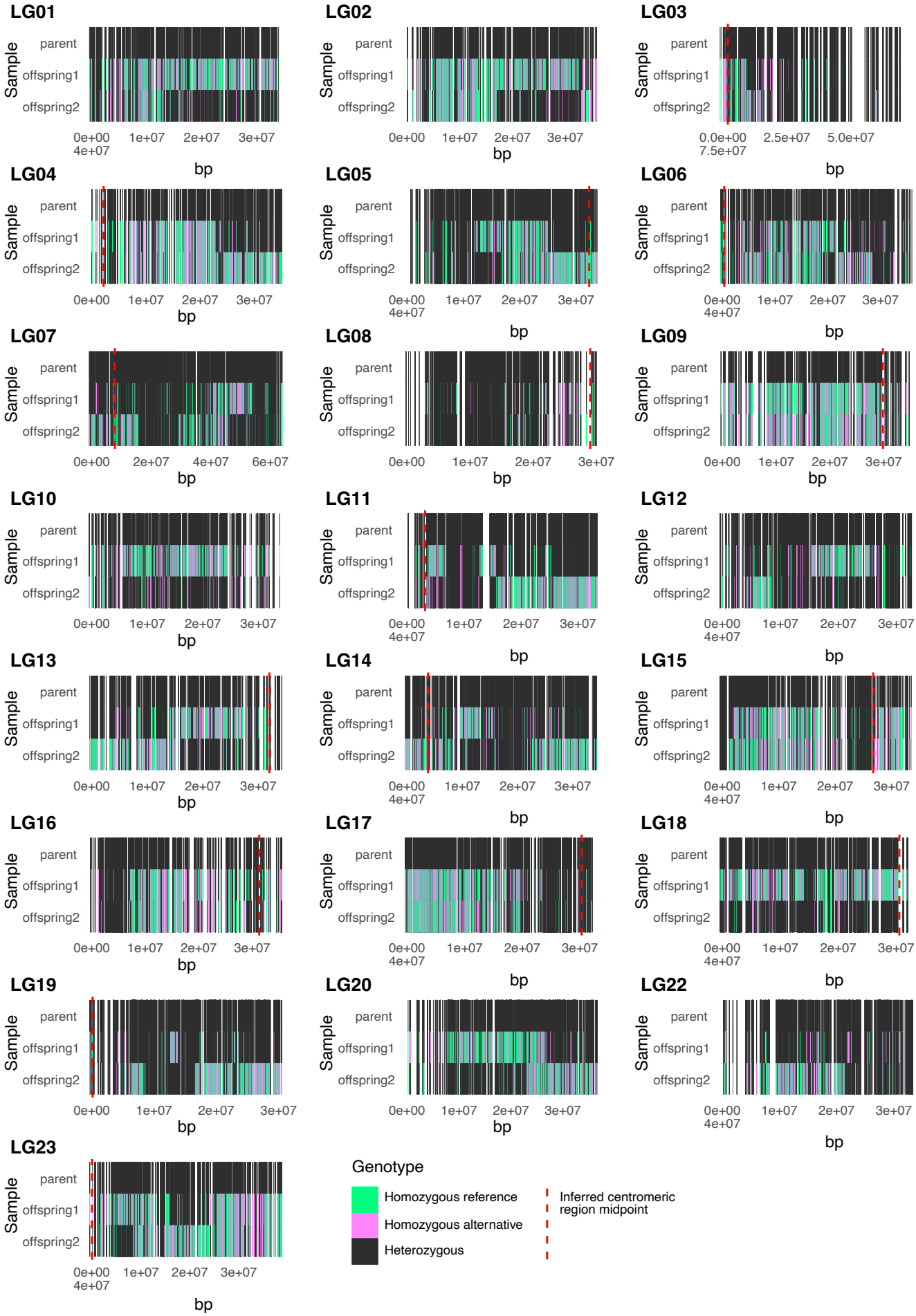
Supplemental Figure S8

Genotype profiles Panels show per chromosome individual genotypes at heterozygous sites of the parent with an individual depth filter of DP > 15. Each bar represents a variant site and colors of bars indicate genotypes as depicted in the inset.



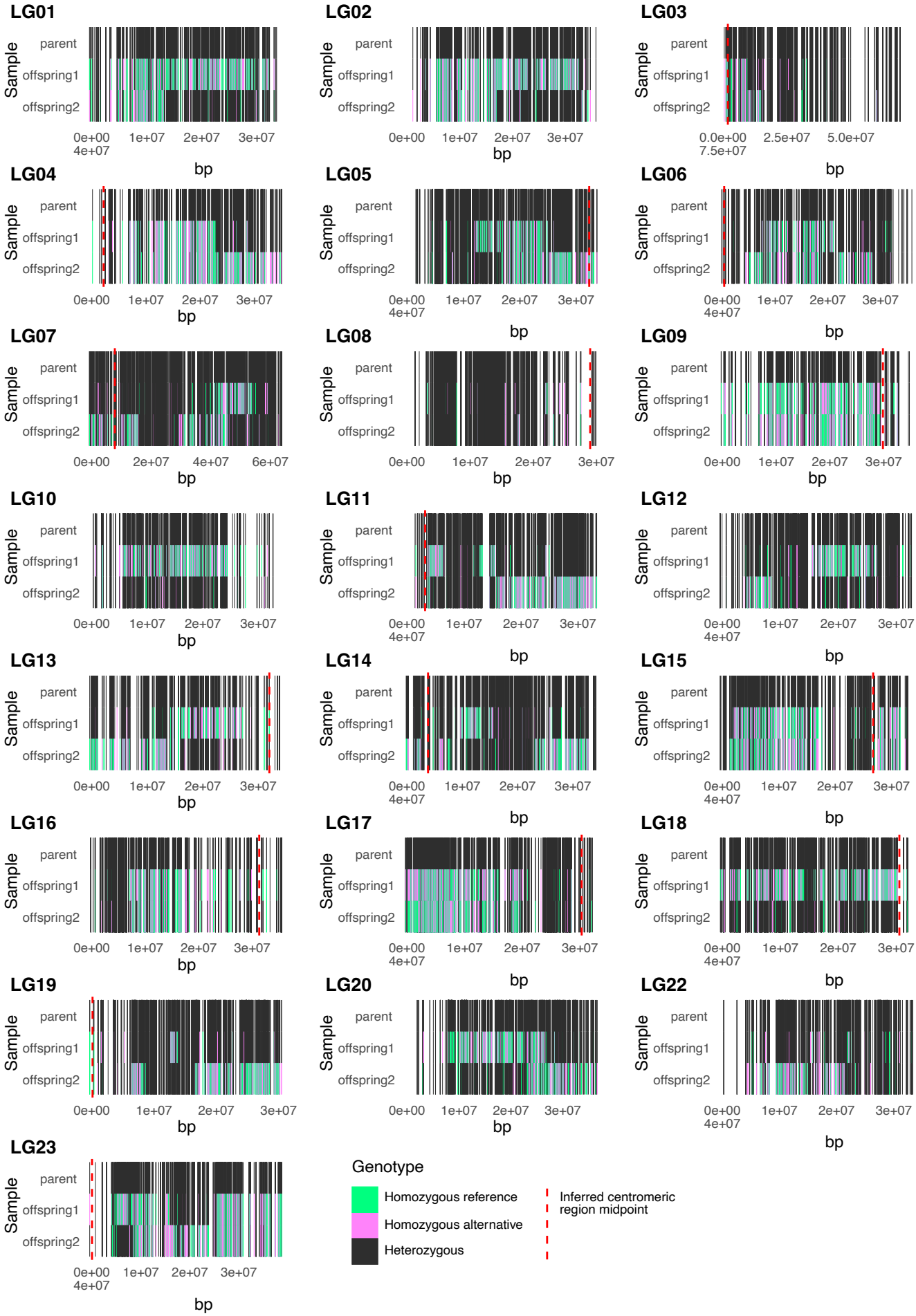
Supplemental Figure S9

Genotype profiles Panels show per chromosome individual genotypes at heterozygous sites of the parent with an individual depth filter of DP > 3 and DP < 30. Each bar represents a variant site and colors of bars indicate genotypes as depicted in the inset.



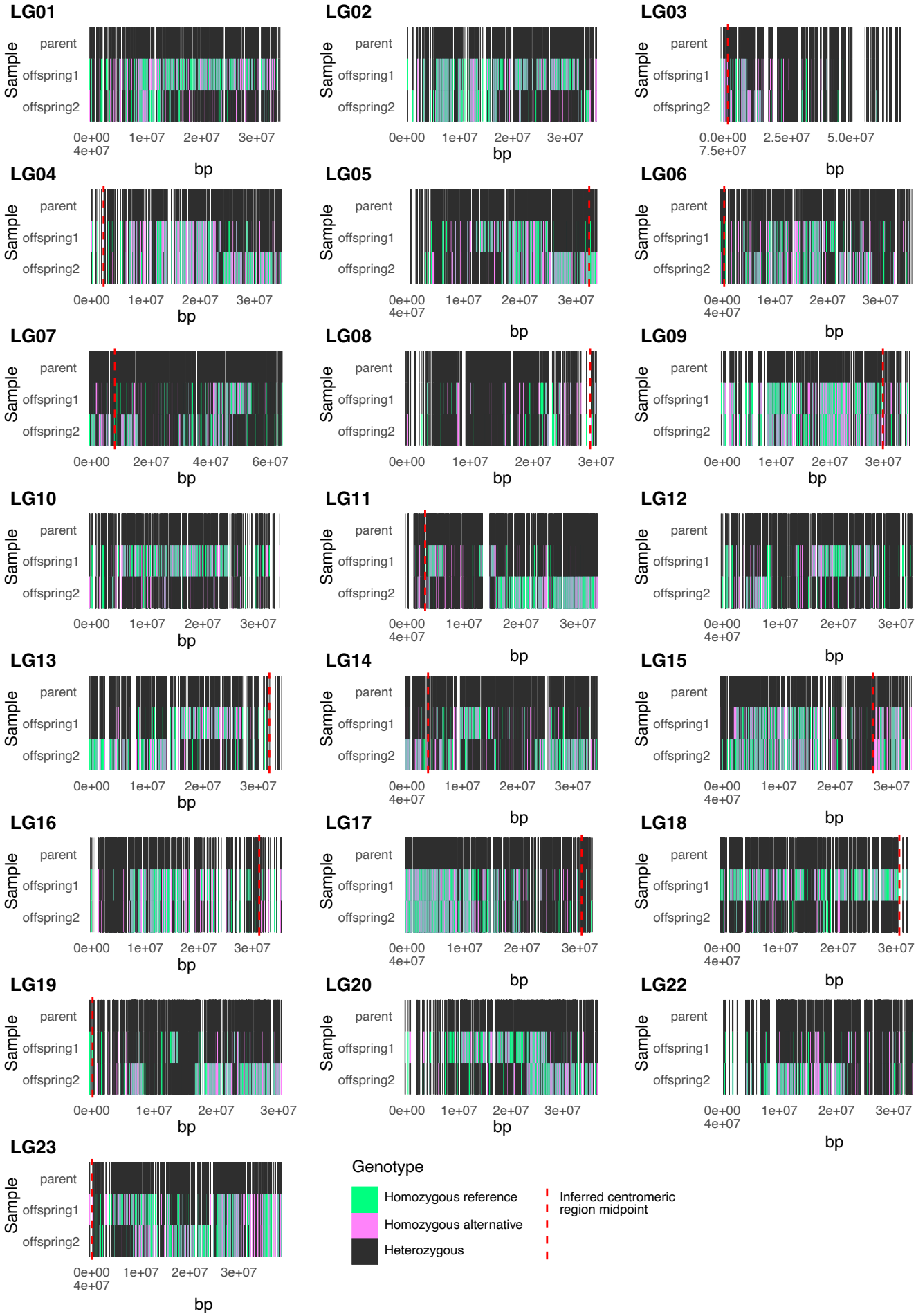
Supplemental Figure S10

Genotype profiles Panels show per chromosome individual genotypes at heterozygous sites of the parent with an individual depth filter of DP > 8 and DP < 30. Each bar represents a variant site and colors of bars indicate genotypes as depicted in the inset.



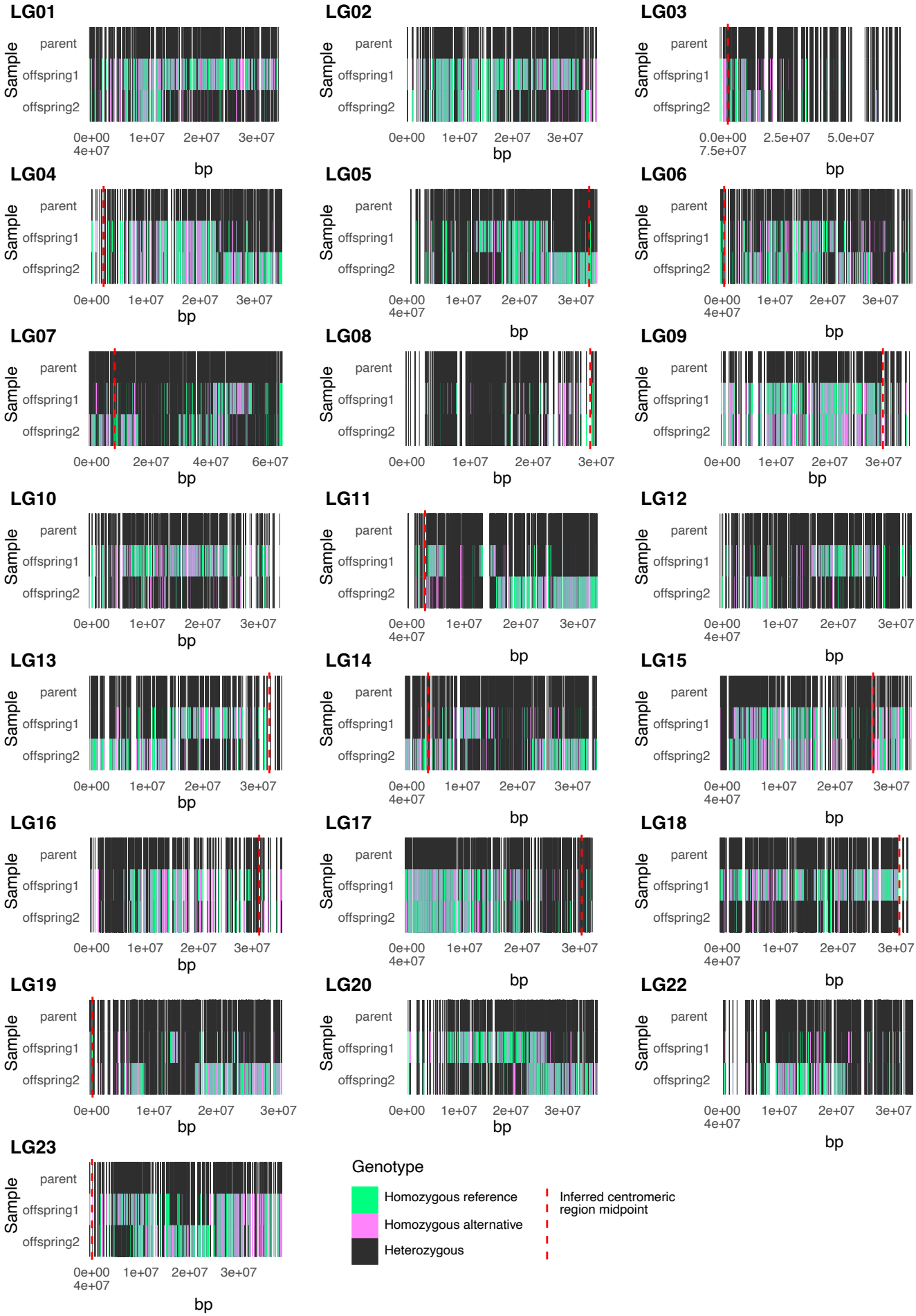
Supplemental Figure S11

Genotype profiles Panels show per chromosome individual genotypes at heterozygous sites of the parent with an individual depth filter of DP > 15 and DP < 30. Each bar represents a variant site and colors of bars indicate genotypes as depicted in the inset.



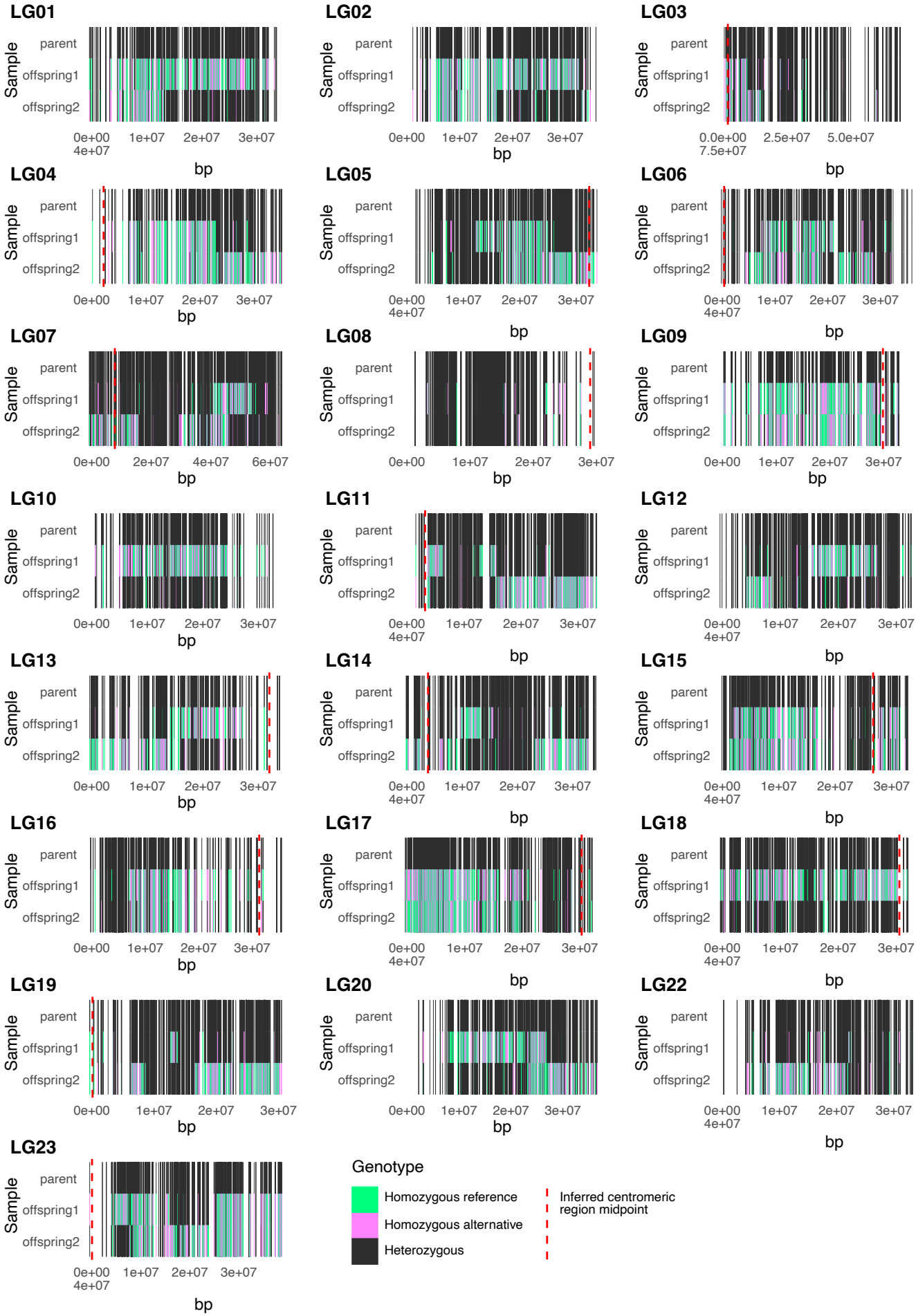
Supplemental Figure S12

Genotype profiles Panels show per chromosome individual genotypes at heterozygous sites of the parent with an individual depth filter of DP > 3 and DP < 25. Each bar represents a variant site and colors of bars indicate genotypes as depicted in the inset.



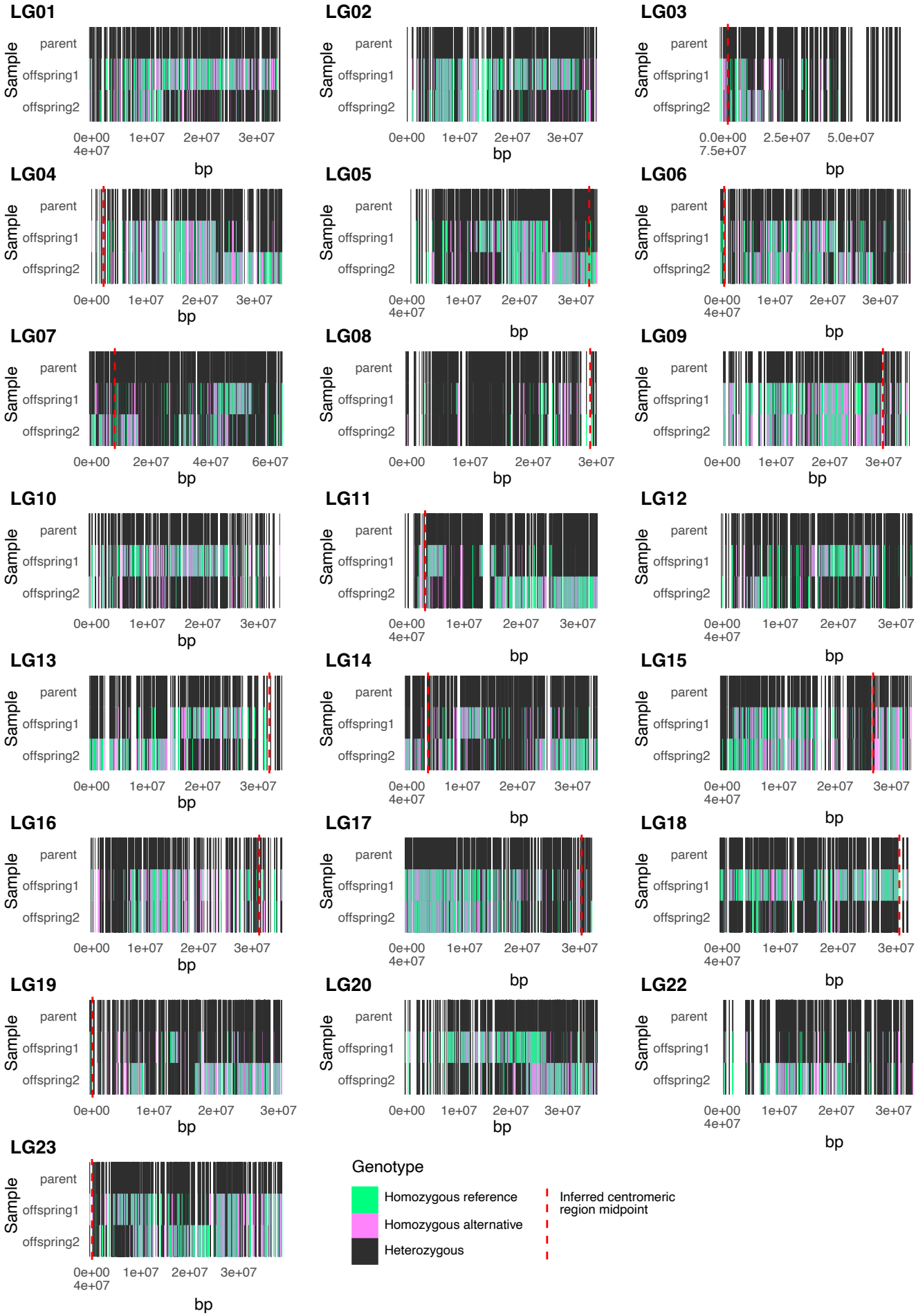
Supplemental Figure S13

Genotype profiles Panels show per chromosome individual genotypes at heterozygous sites of the parent with an individual depth filter of DP > 8 and DP < 25. Each bar represents a variant site and colors of bars indicate genotypes as depicted in the inset.



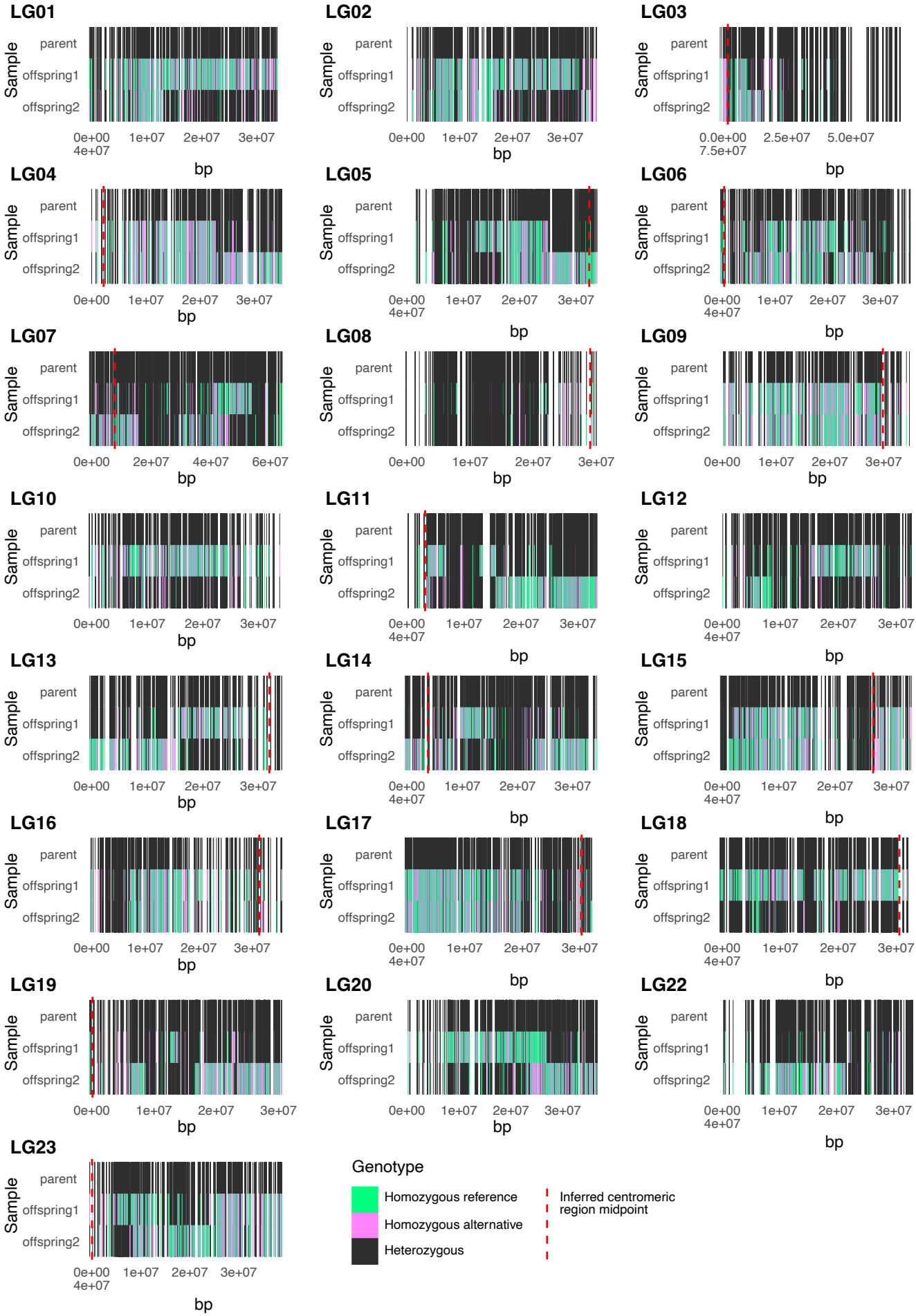
Supplemental Figure S14

Genotype profiles Panels show per chromosome individual genotypes at heterozygous sites of the parent with an individual depth filter of DP > 15 and DP < 25. Each bar represents a variant site and colors of bars indicate genotypes as depicted in the inset.



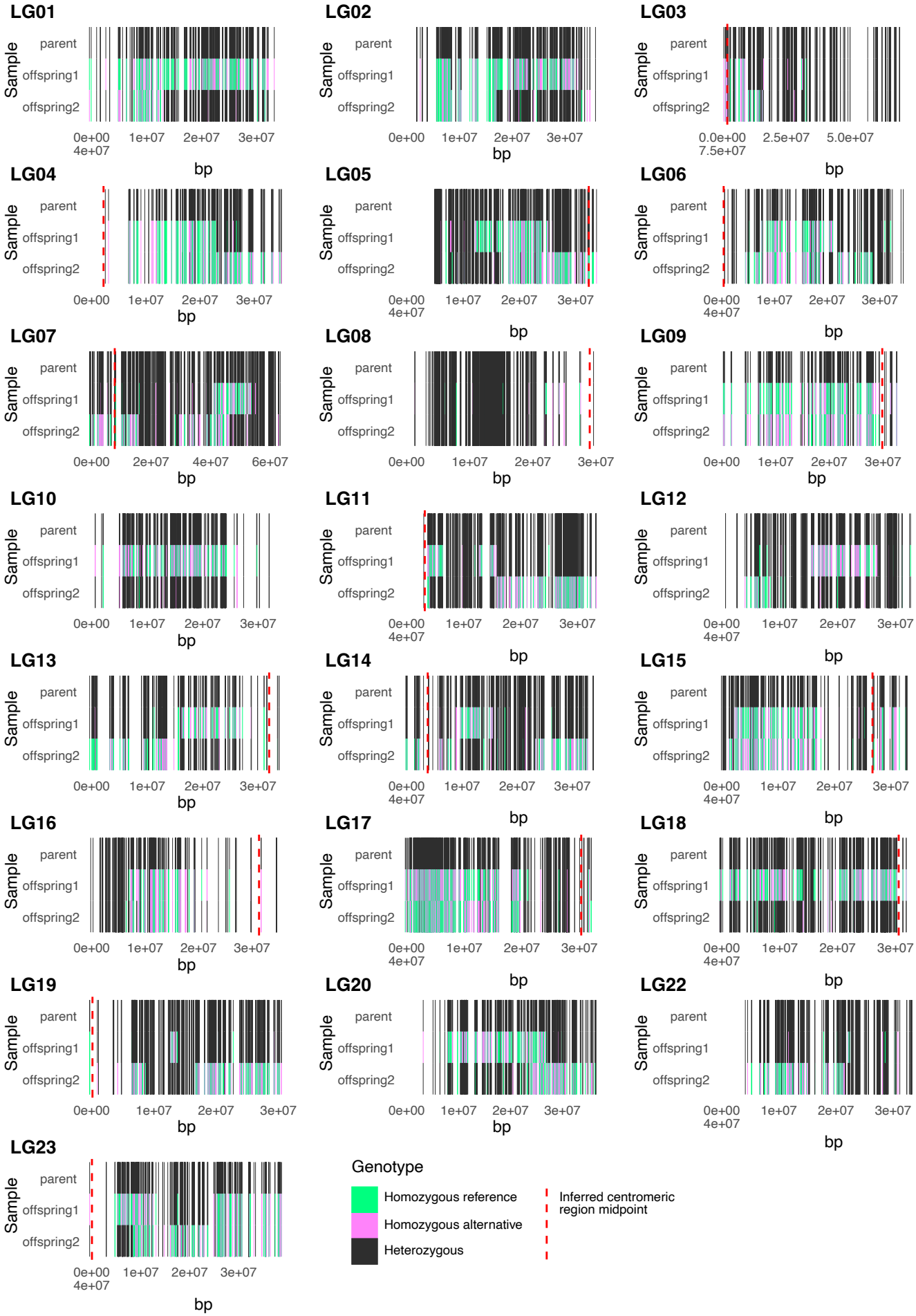
Supplemental Figure S15

Genotype profiles Panels show per chromosome individual genotypes at heterozygous sites of the parent with an individual depth filter of DP > 3 and DP < 20. Each bar represents a variant site and colors of bars indicate genotypes as depicted in the inset.



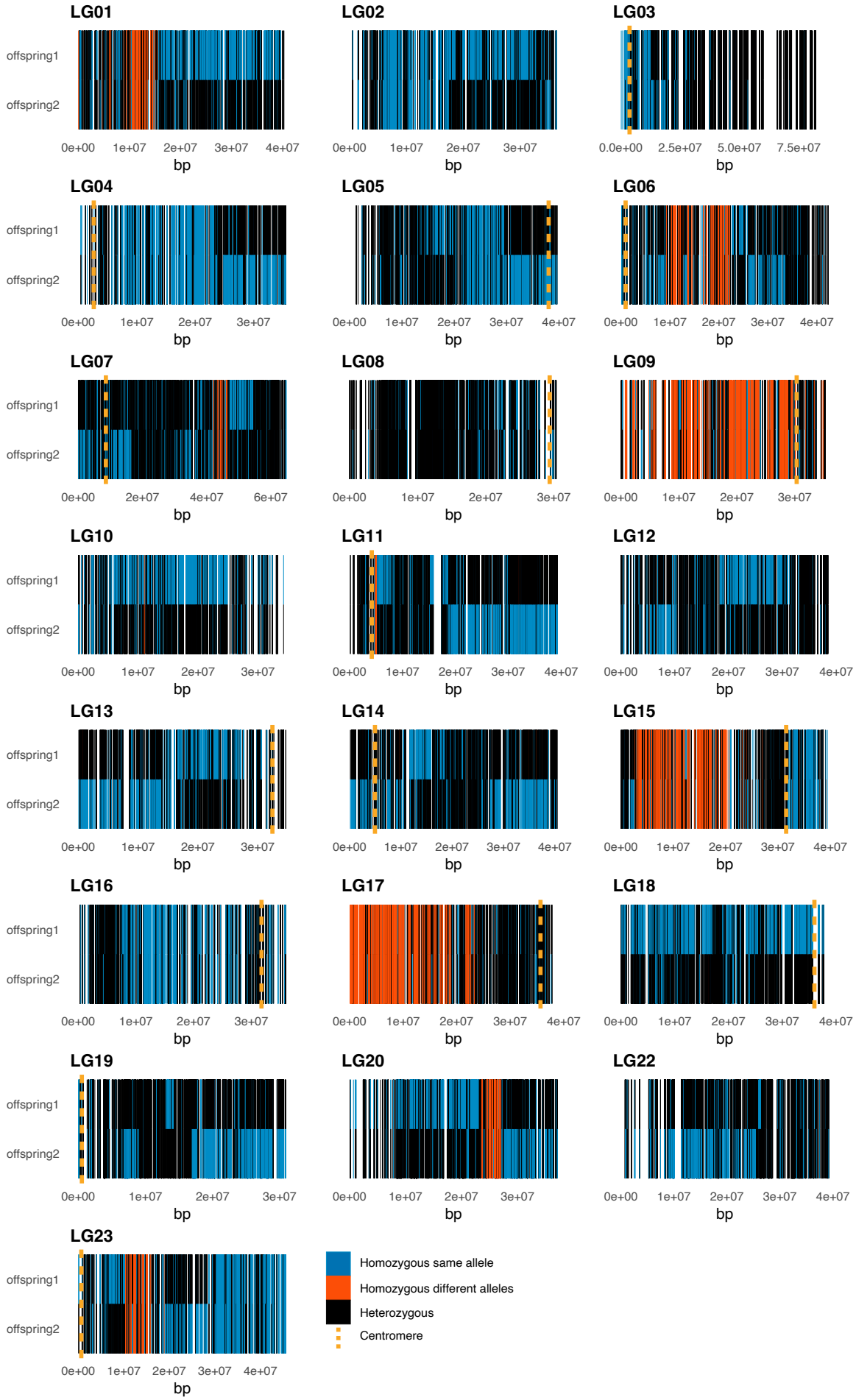
Supplemental Figure S16

Genotype profiles Panels show per chromosome individual genotypes at heterozygous sites of the parent with an individual depth filter of DP > 8 and DP < 20. Each bar represents a variant site and colors of bars indicate genotypes as depicted in the inset.



Supplemental Figure S17

Genotype profiles Panels show per chromosome individual genotypes at heterozygous sites of the parent with an individual depth filter of DP > 15 and DP < 20. Each bar represents a variant site and colors of bars indicate genotypes as depicted in the inset.



Supplemental Figure S18

Genotype profiles Panels show per chromosome individual genotypes of the two offspring at heterozygous sites of the parent. Each bar represents a variant site and colors of bars indicate genotypes as depicted in the inset.