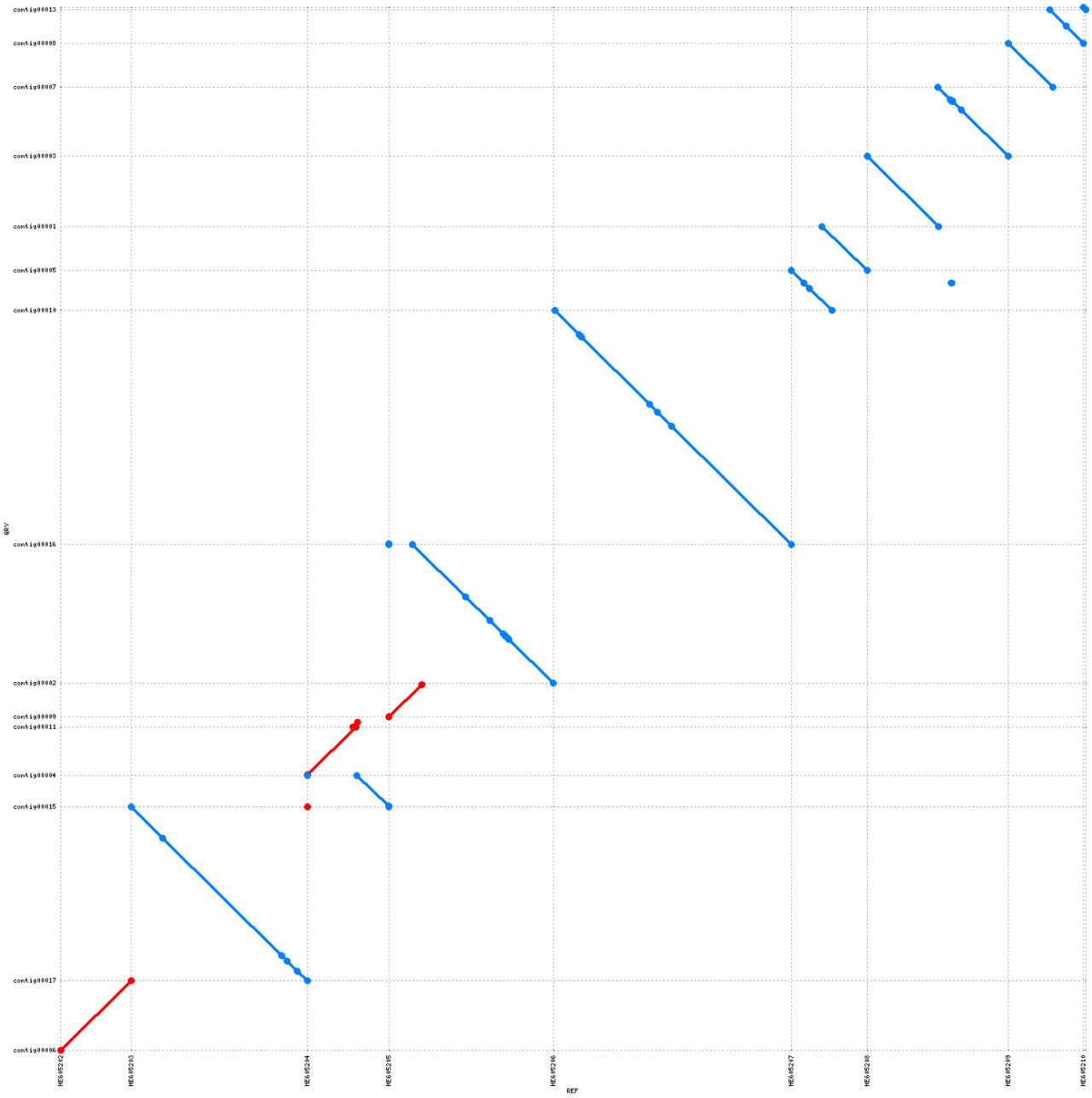
Supplementary figures and tables

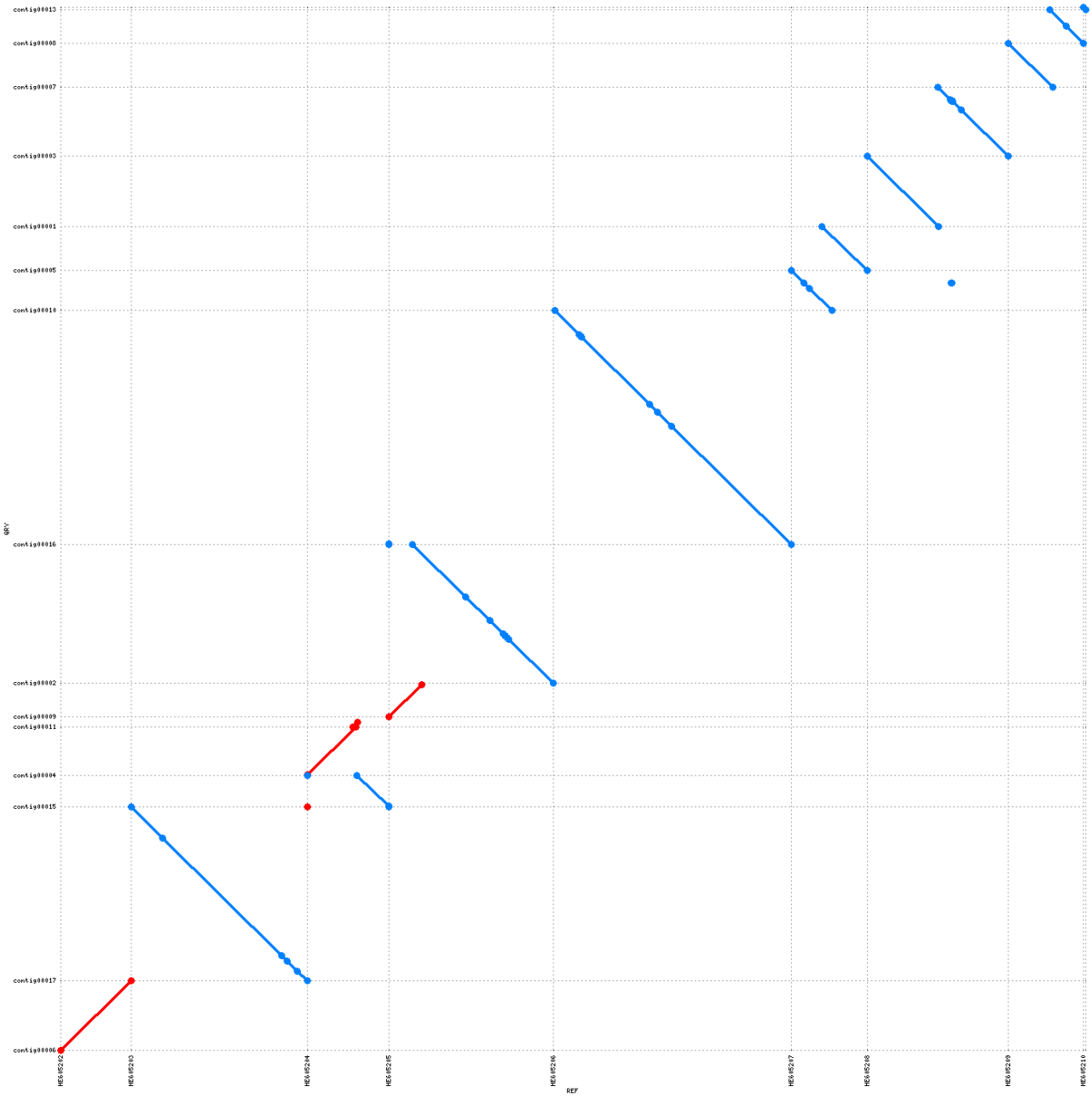
Supplementary figure S1: Alignments of assemblies from heterozygous genome assembly pipeline

Scaffolds returned by the heterozygous genome assembly pipeline for various level of LOH have been aligned onto *C. parapsilosis* CDC317 chromosomes. The reference chromosomes are denoted on X axis, while query contigs/scaffolds are denoted on Y axis. Best query-to-reference matches are denoted with dots, forward in red and reverse in blue. The regions of similarity spanning larger regions are denoted by lines. Subsequently, the alignments were scanned for potential rearrangements (marked by arrows on reference axis).

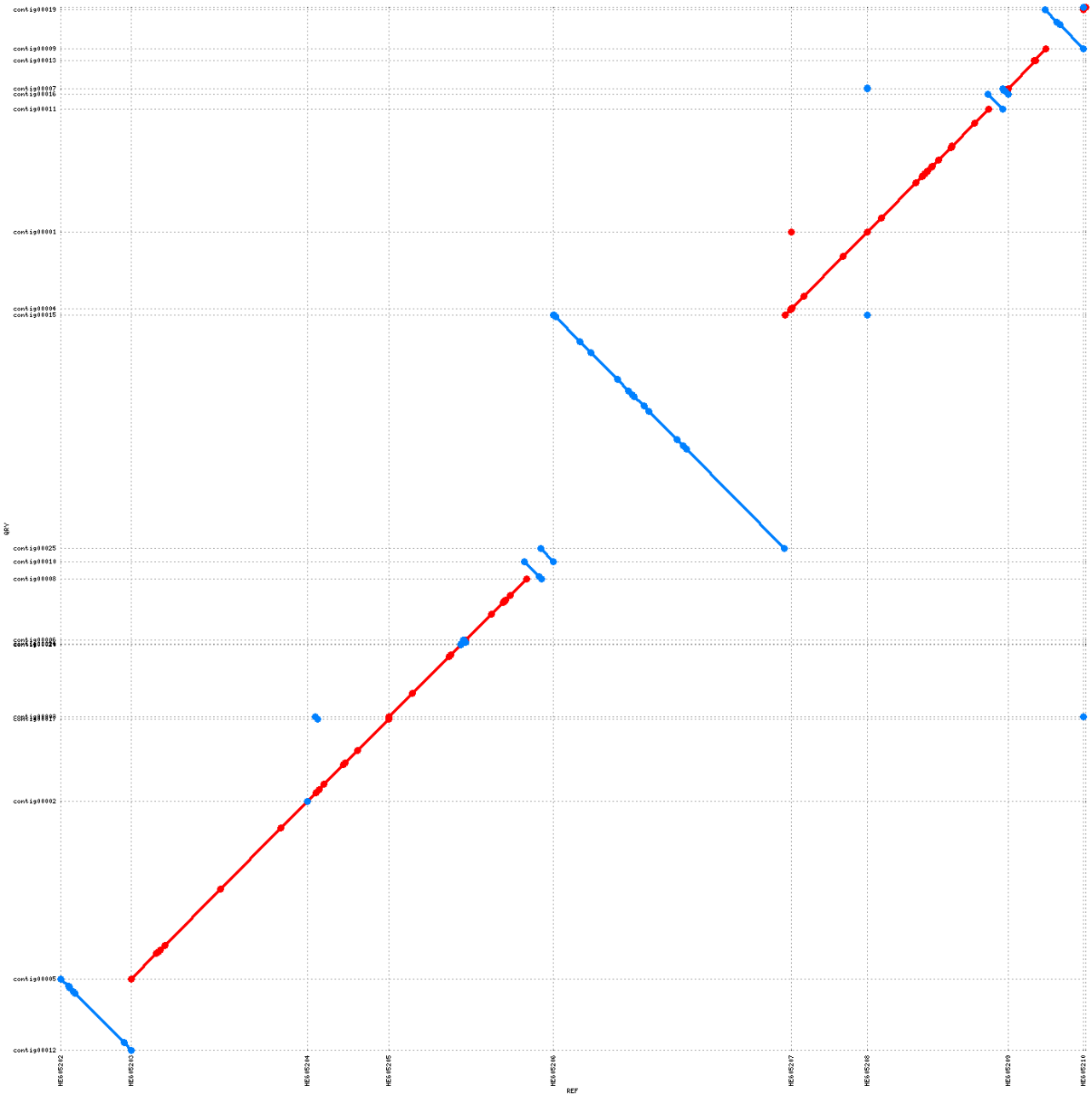
0% LOH started from contigs



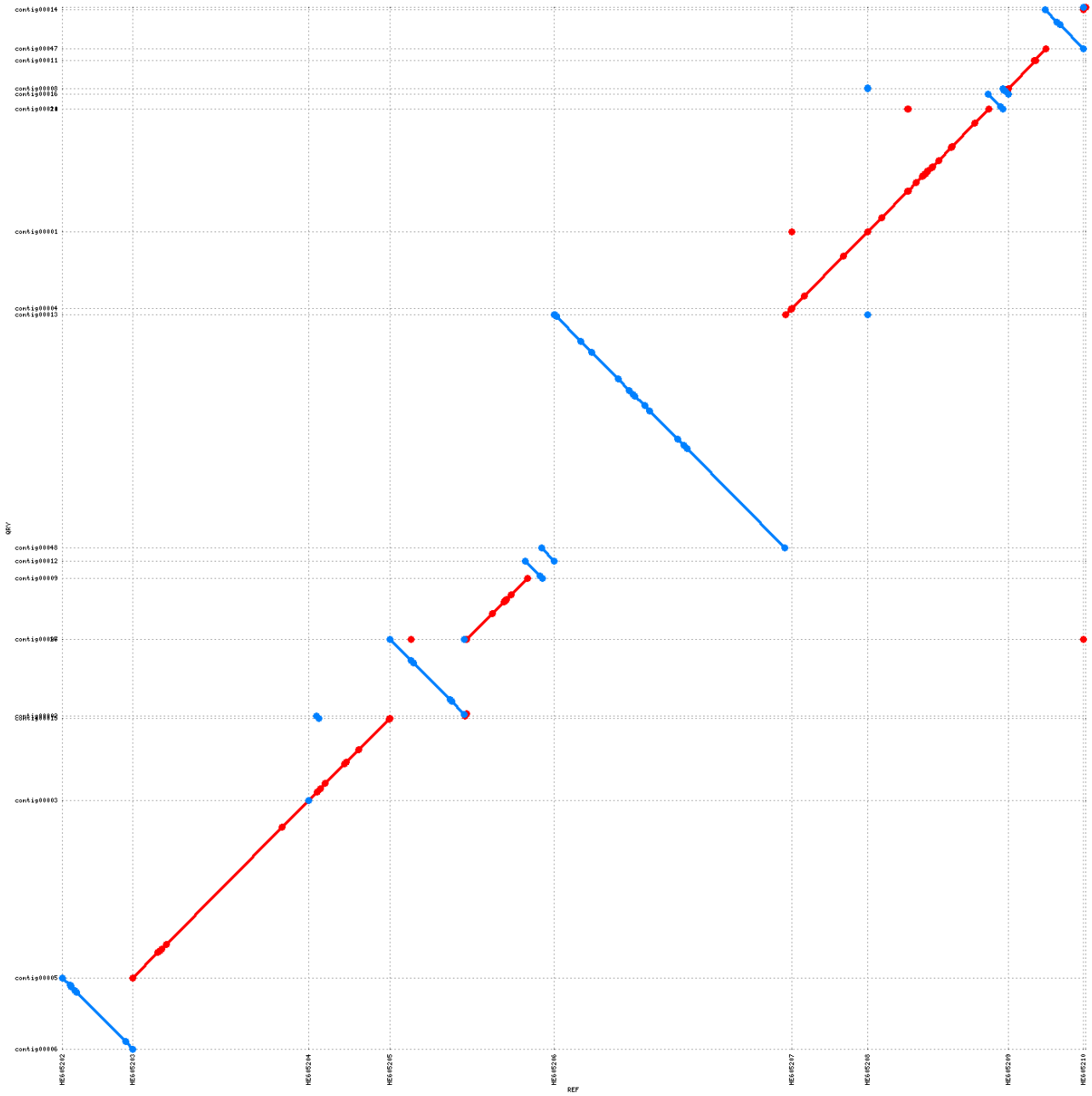
0% LOH started from scaffolds



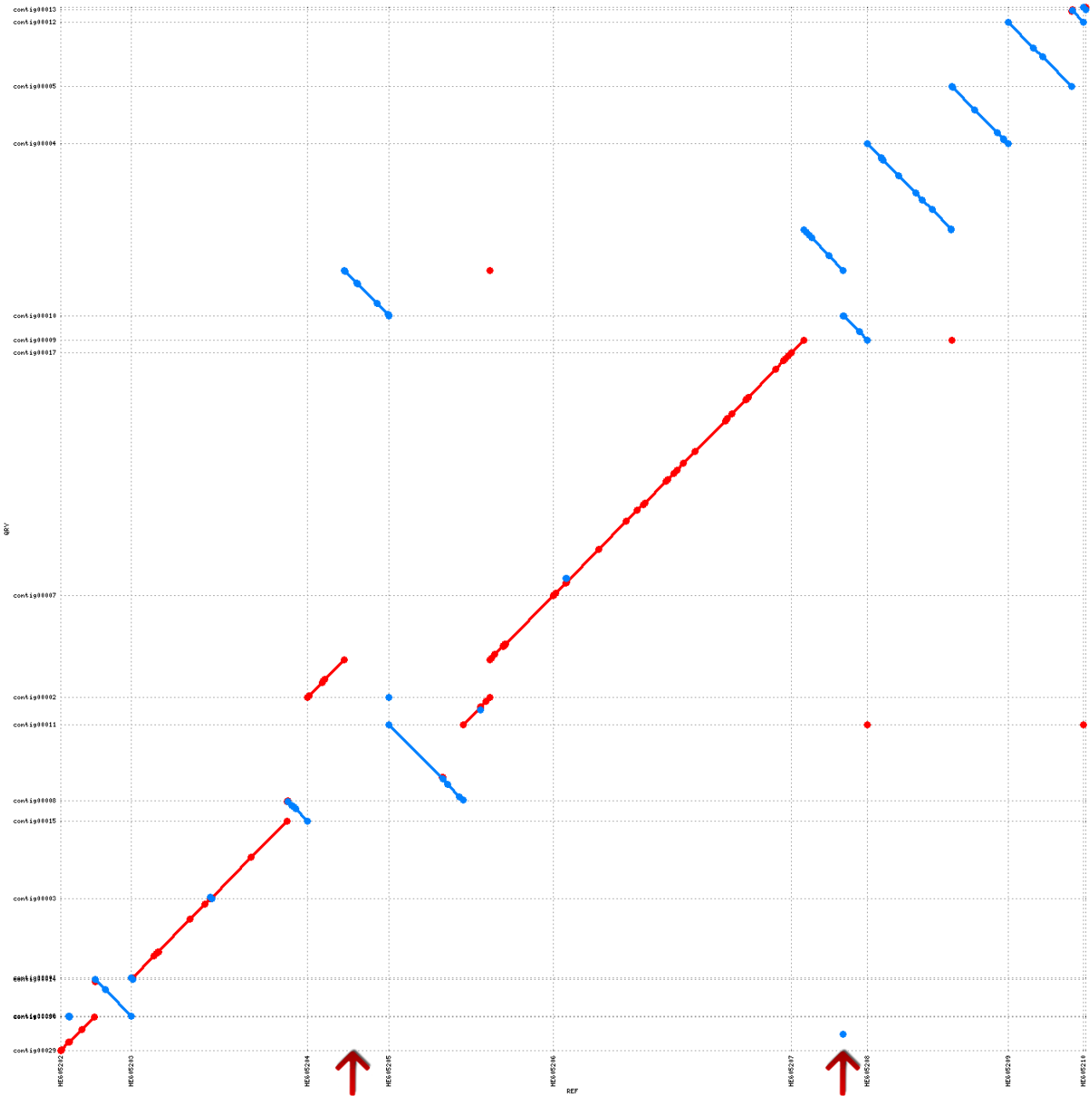
20% LOH started from contigs



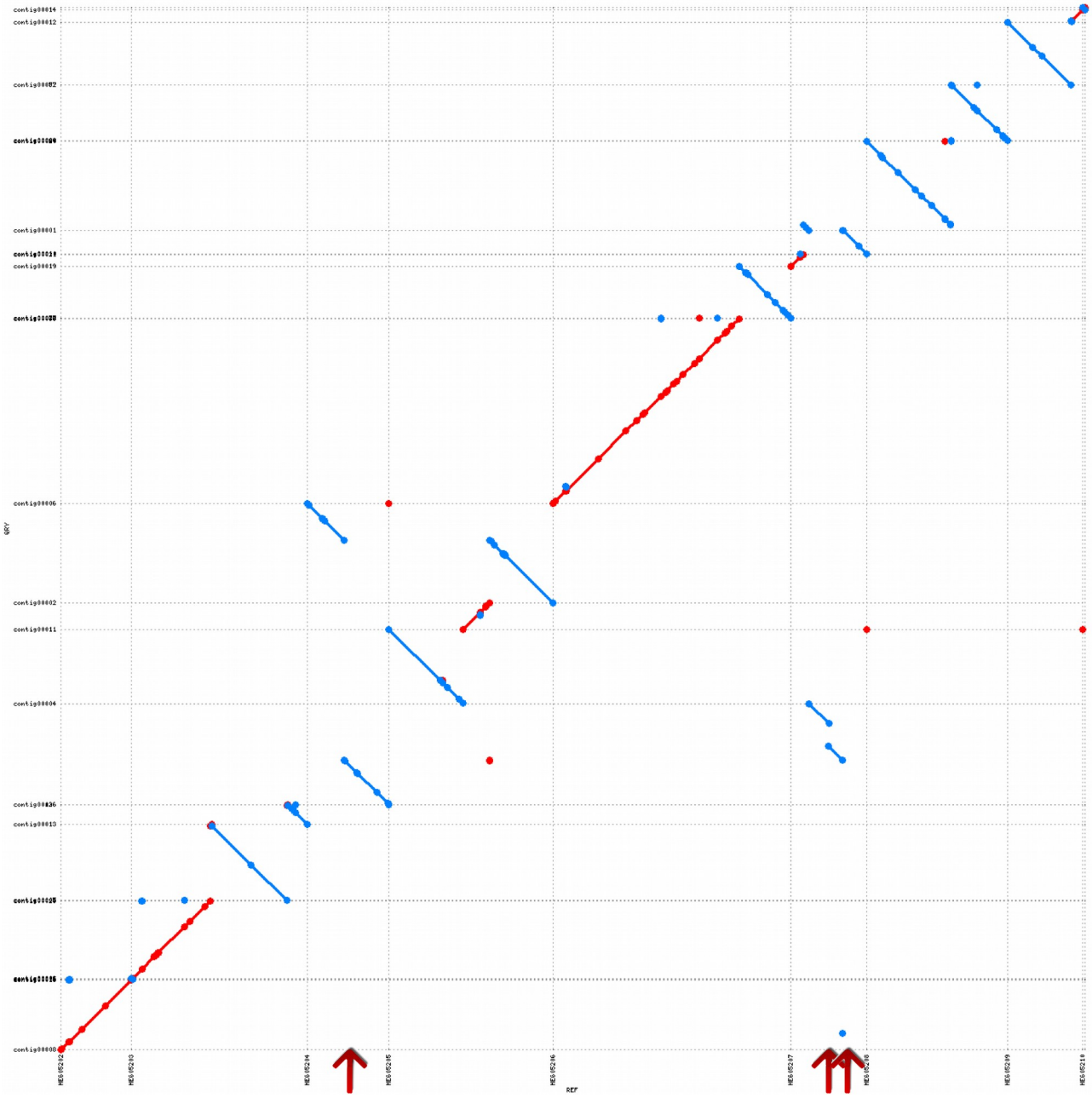
20% LOH started from scaffolds



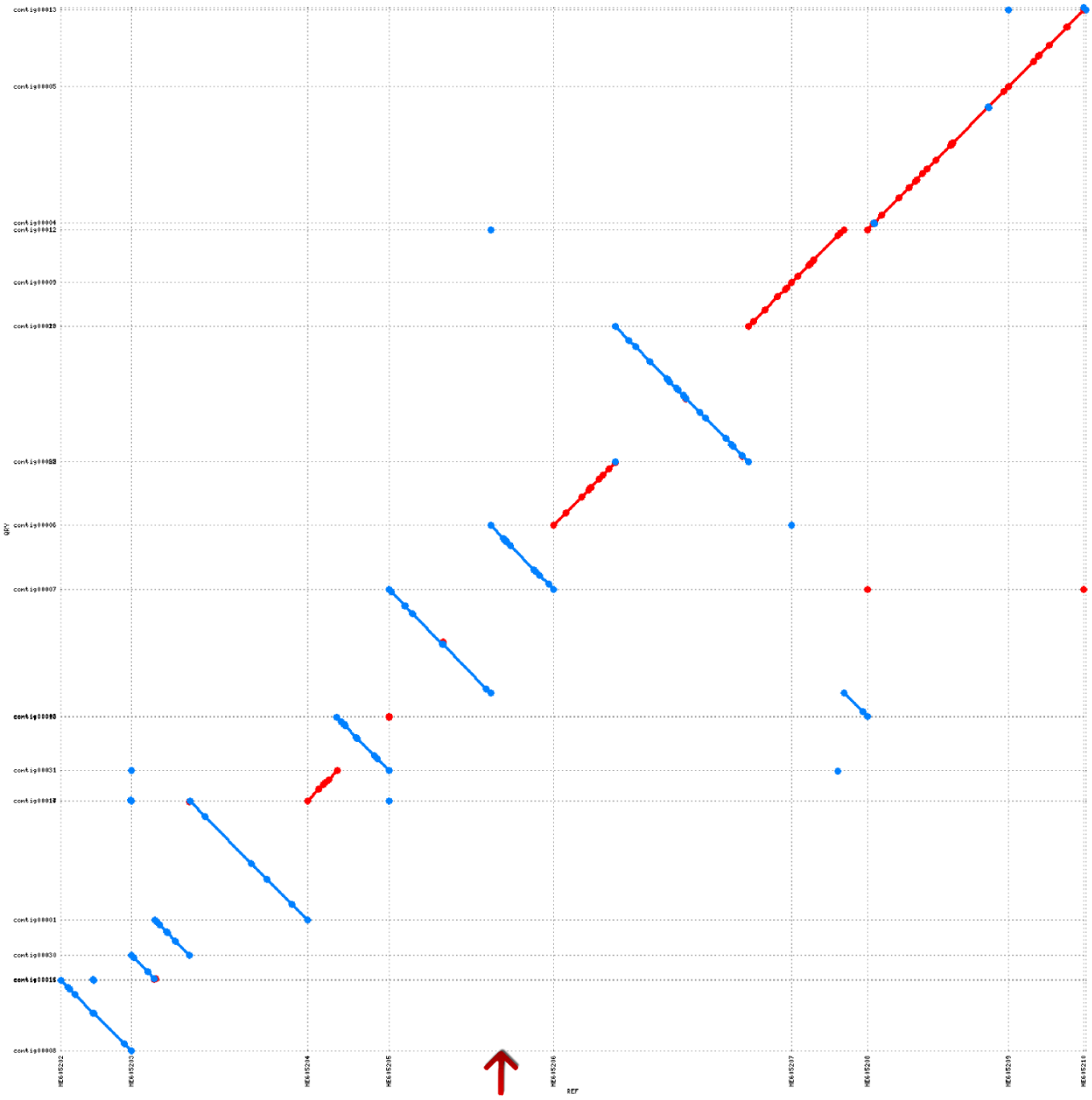
40% LOH started from contigs



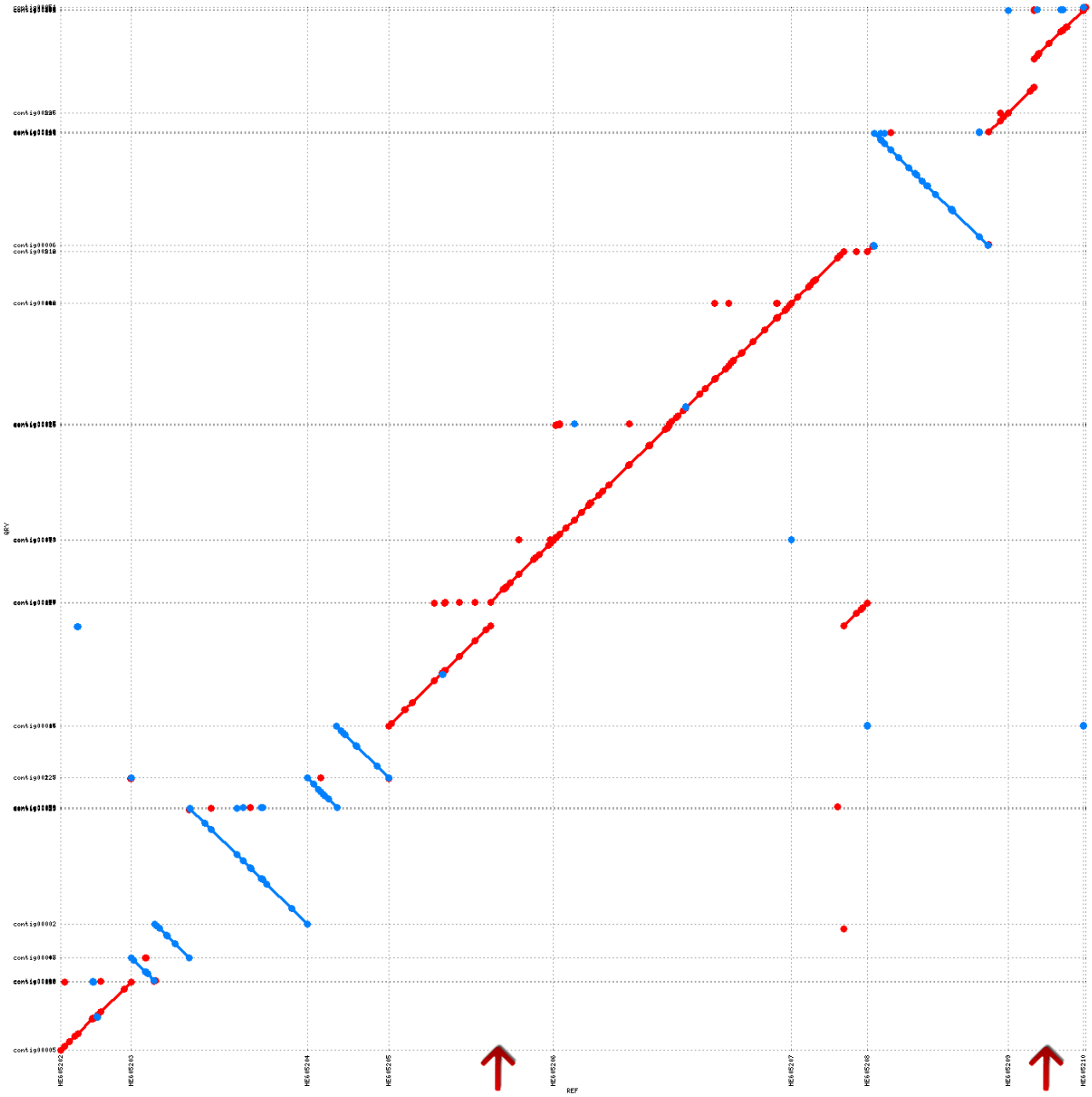
40% LOH started from scaffolds



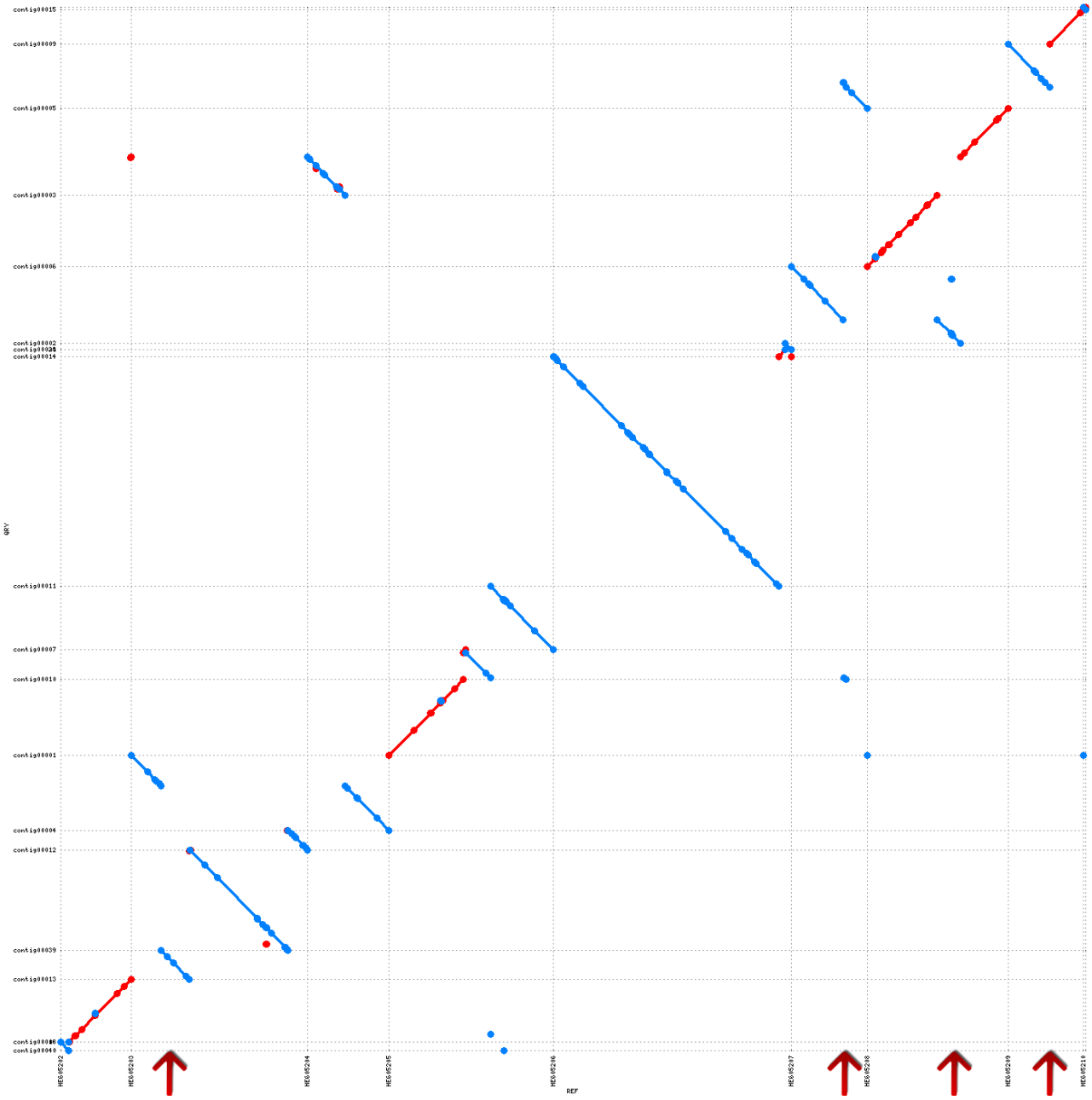
60% LOH started from contigs



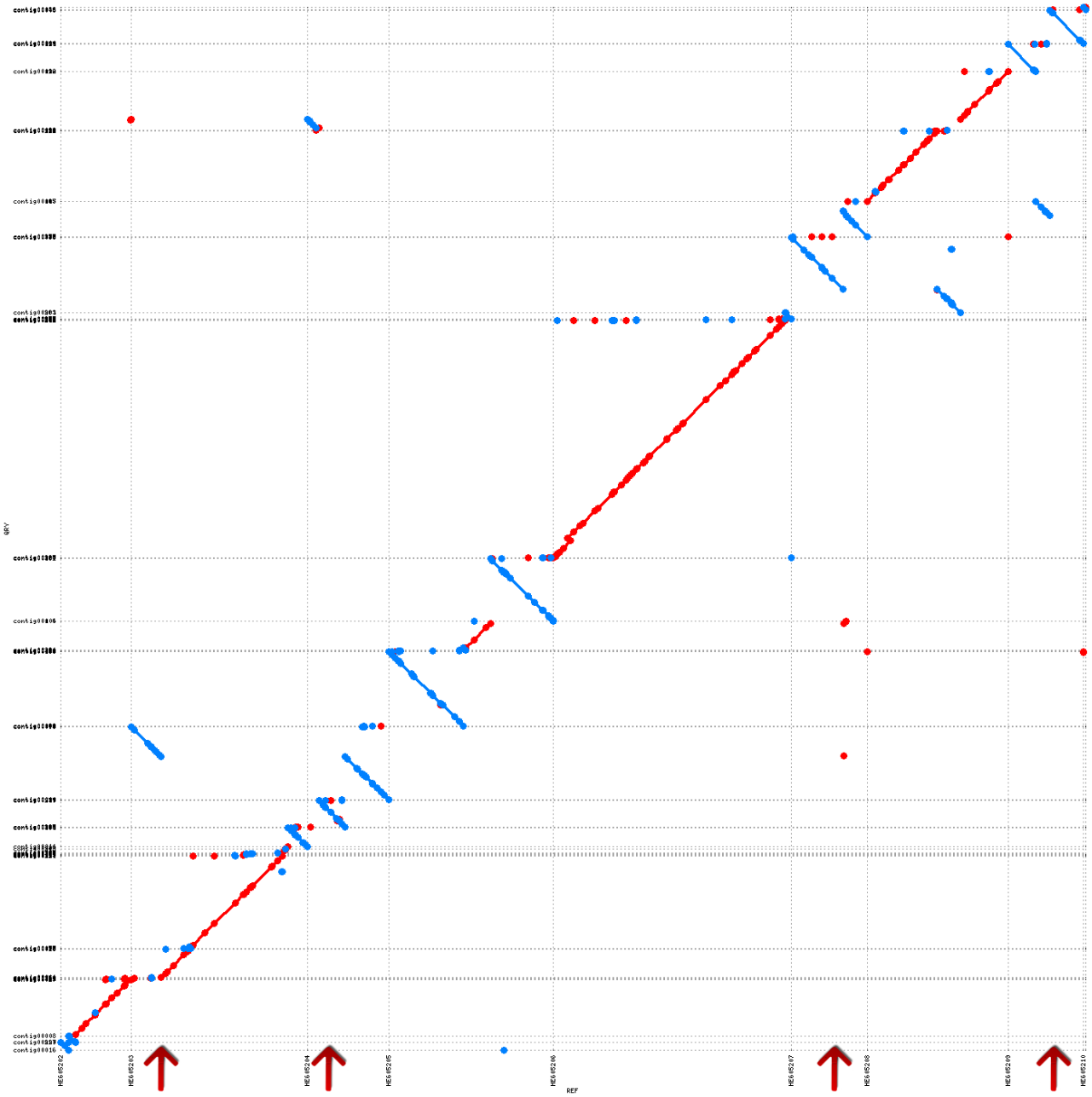
60% LOH started from scaffolds



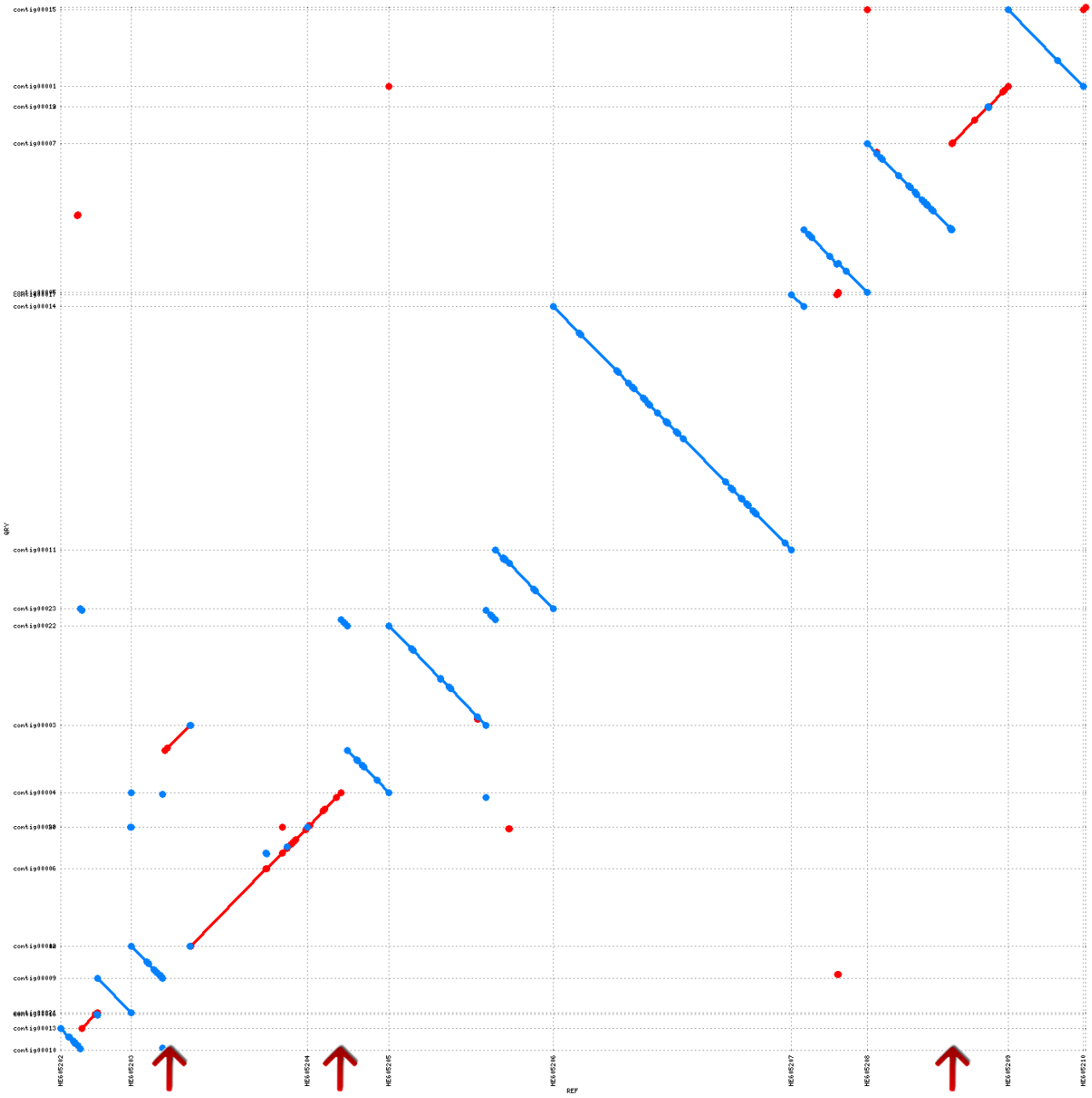
80% LOH started from contigs



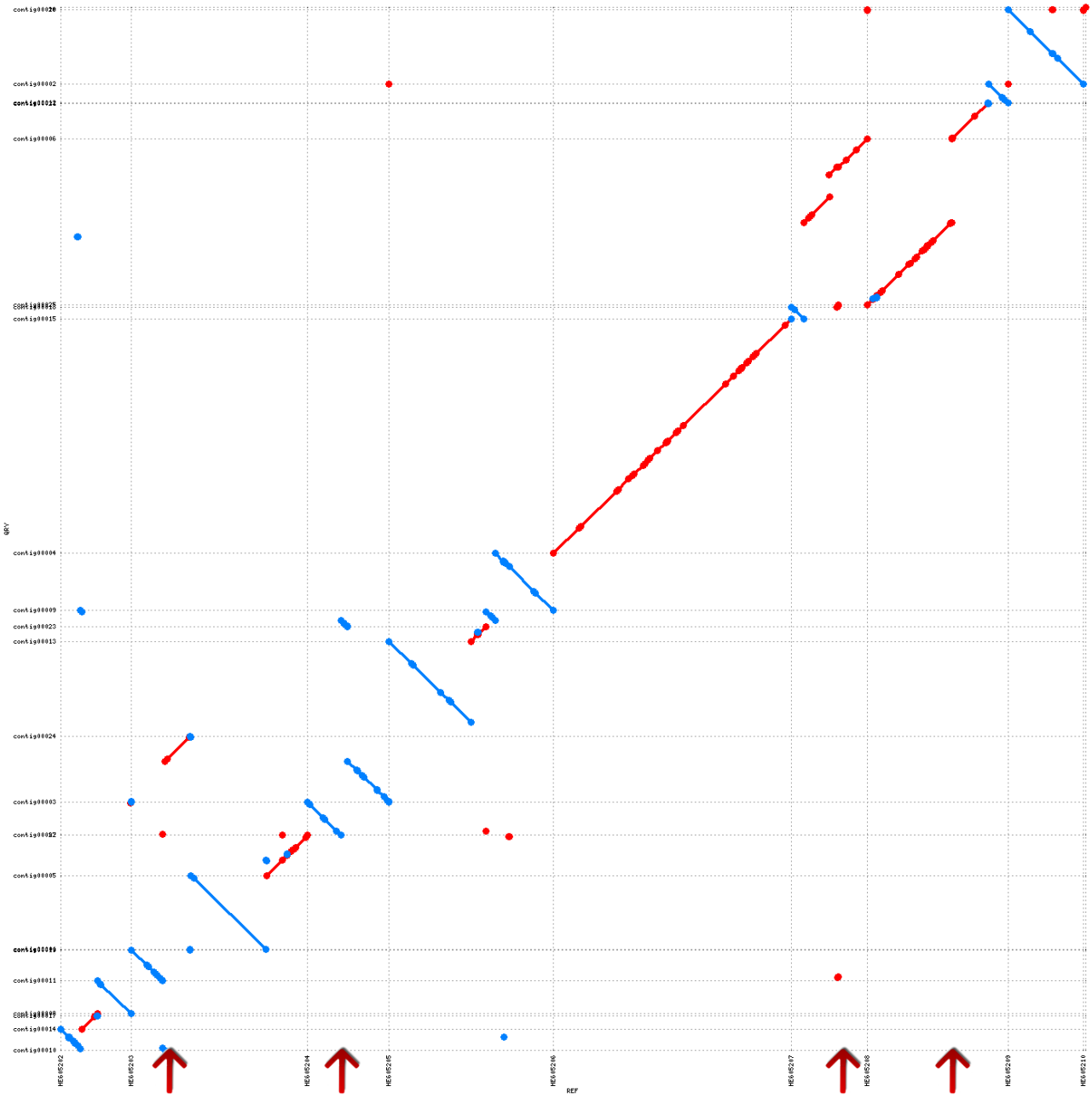
80% LOH started from scaffolds



100% LOH started from contigs



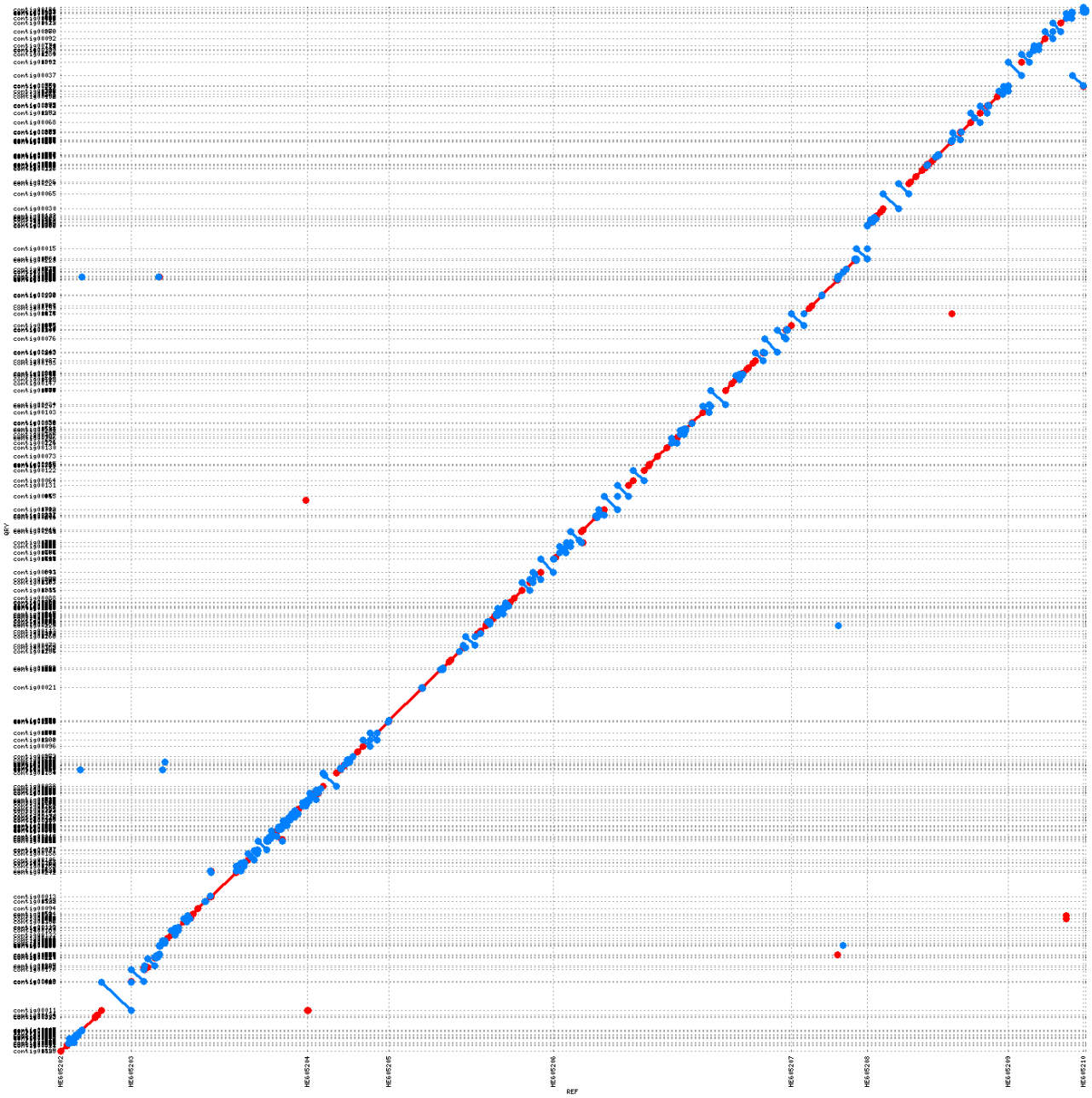
100% LOH started from scaffolds



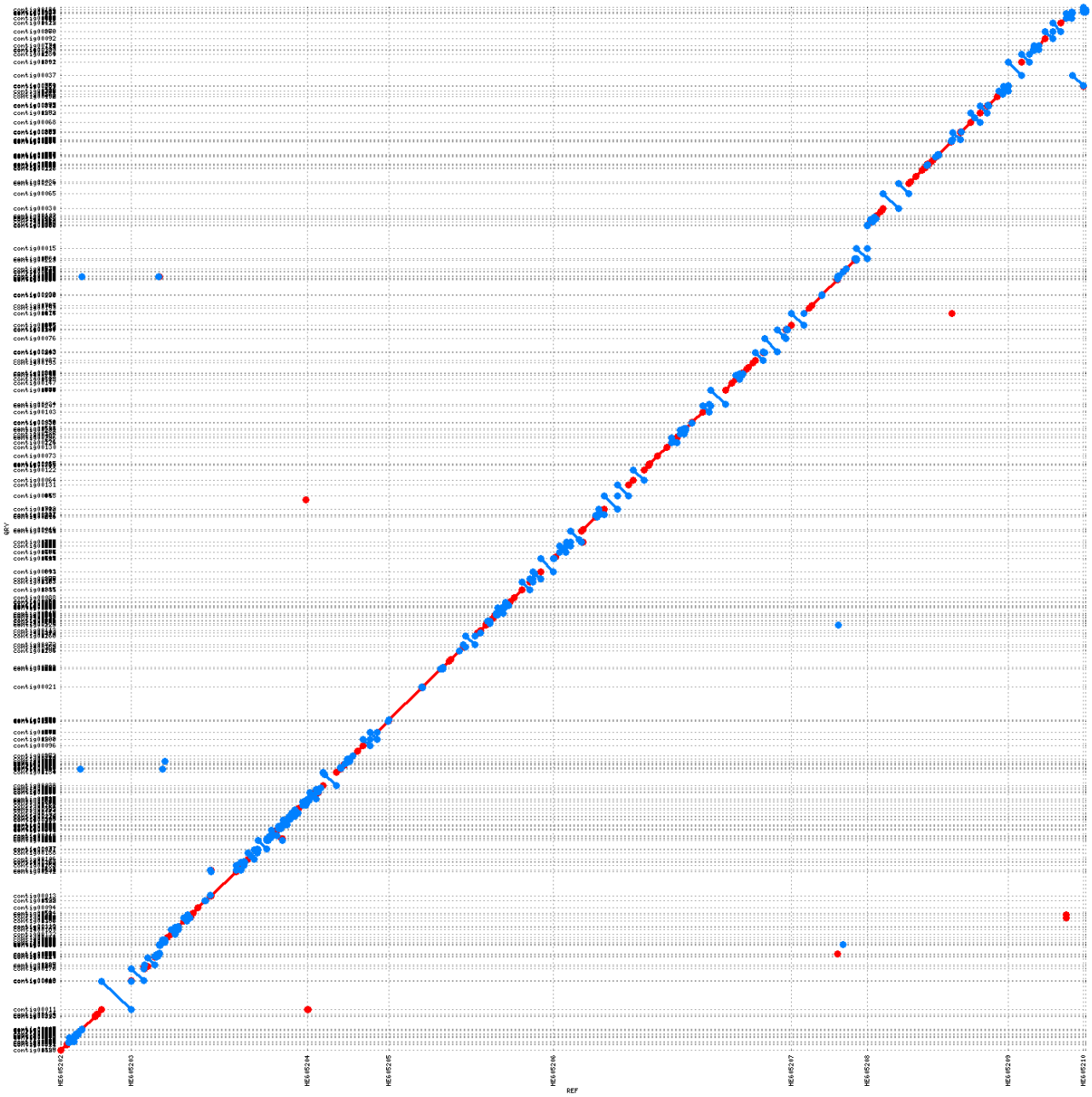
Supplementary figure S2: Alignments of assemblies from SPAdes

Contigs or scaffolds returned by the SPAdes assembler for various level of LOH have been aligned onto *C. parapsilosis* CDC317 chromosomes. The reference chromosomes are denoted on X axis, while query contigs/scaffolds are denoted on Y axis. Best query-to-reference matches are denoted with dots, forward in red and reverse in blue. The regions of similarity spanning larger regions are denoted by lines. Subsequently, the alignments were scanned for potential rearrangements (marked by arrows on reference axis).

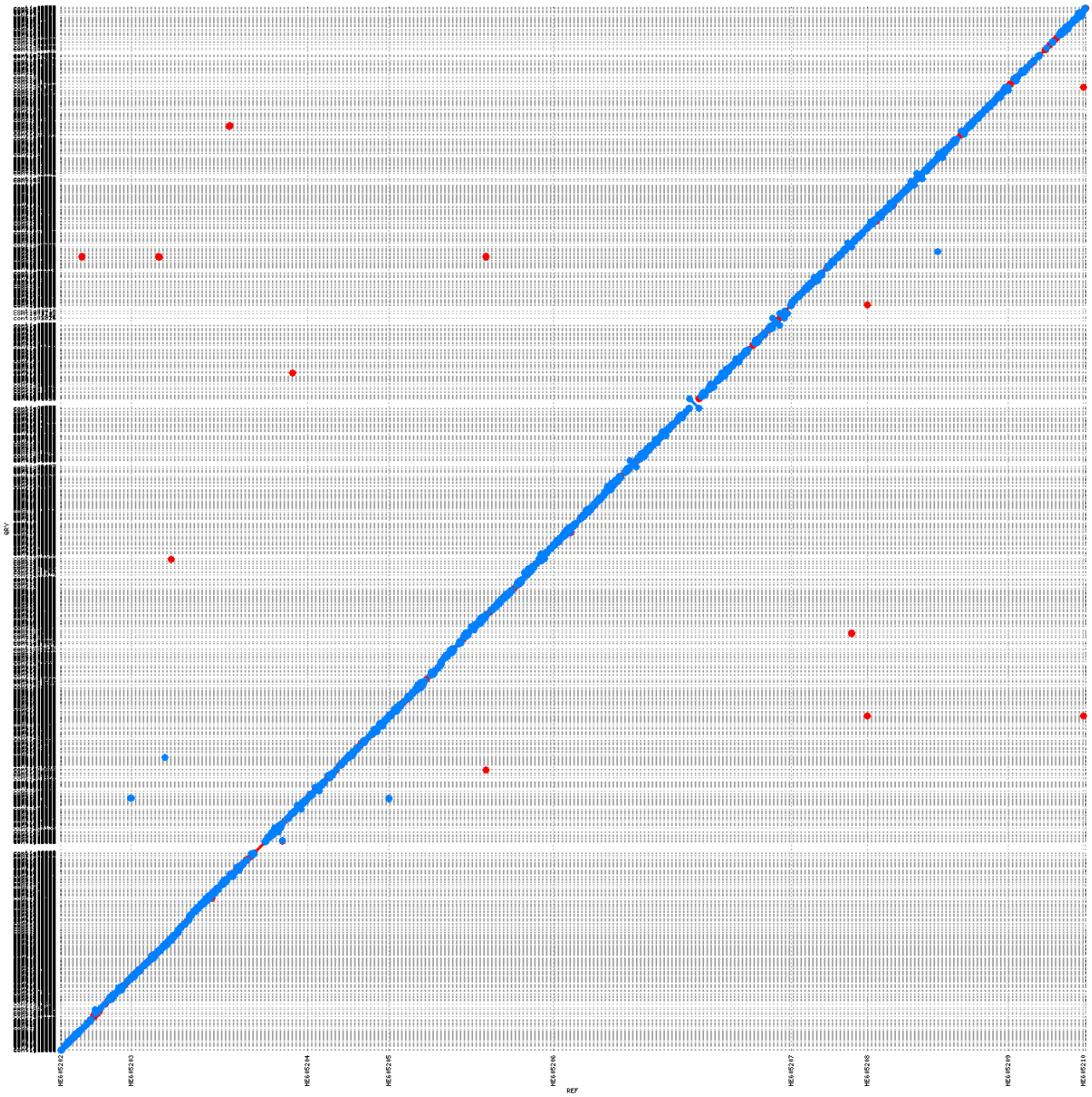
0% LOH started from contigs



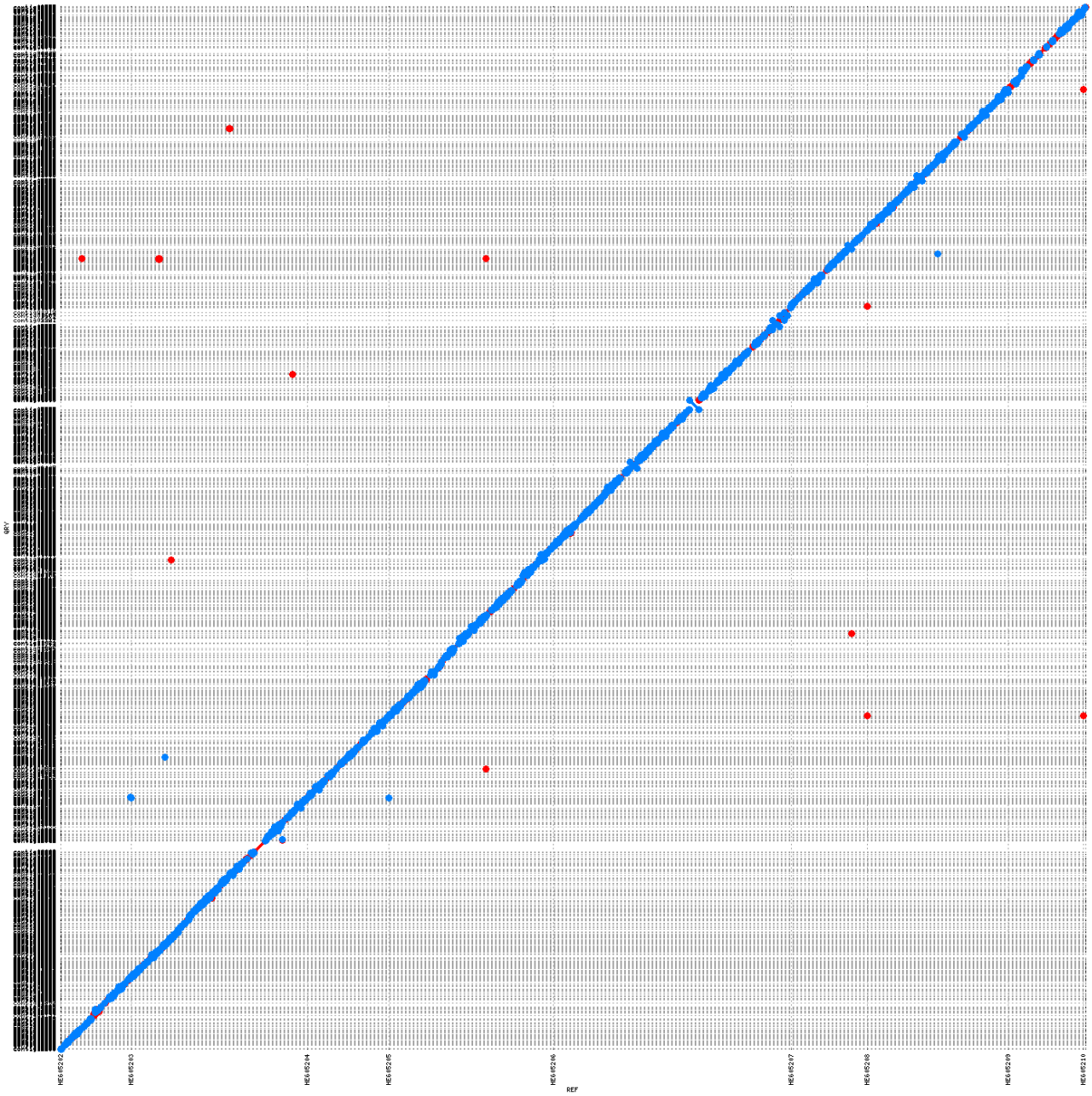
0% LOH started from scaffolds



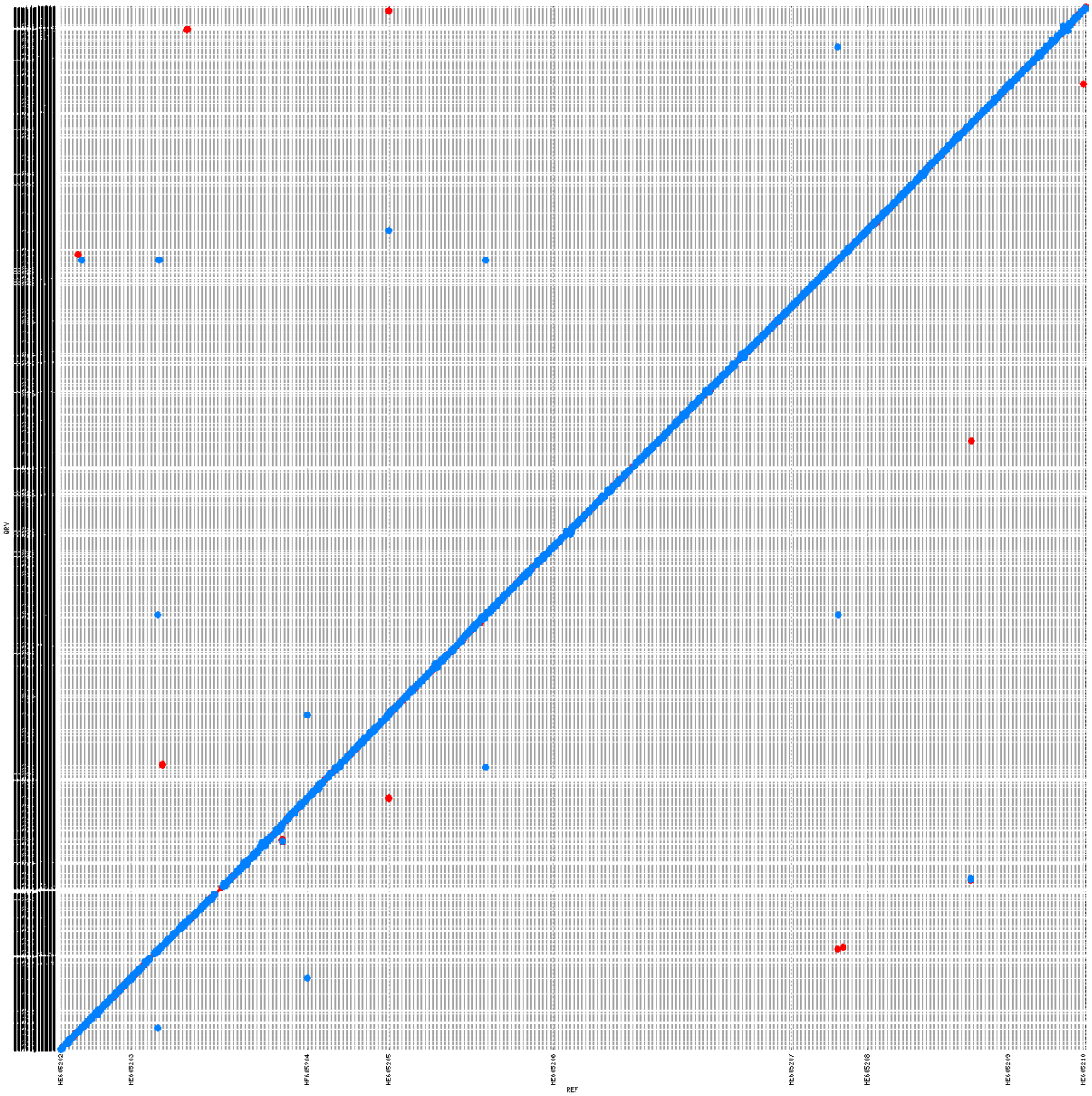
20% LOH started from contigs



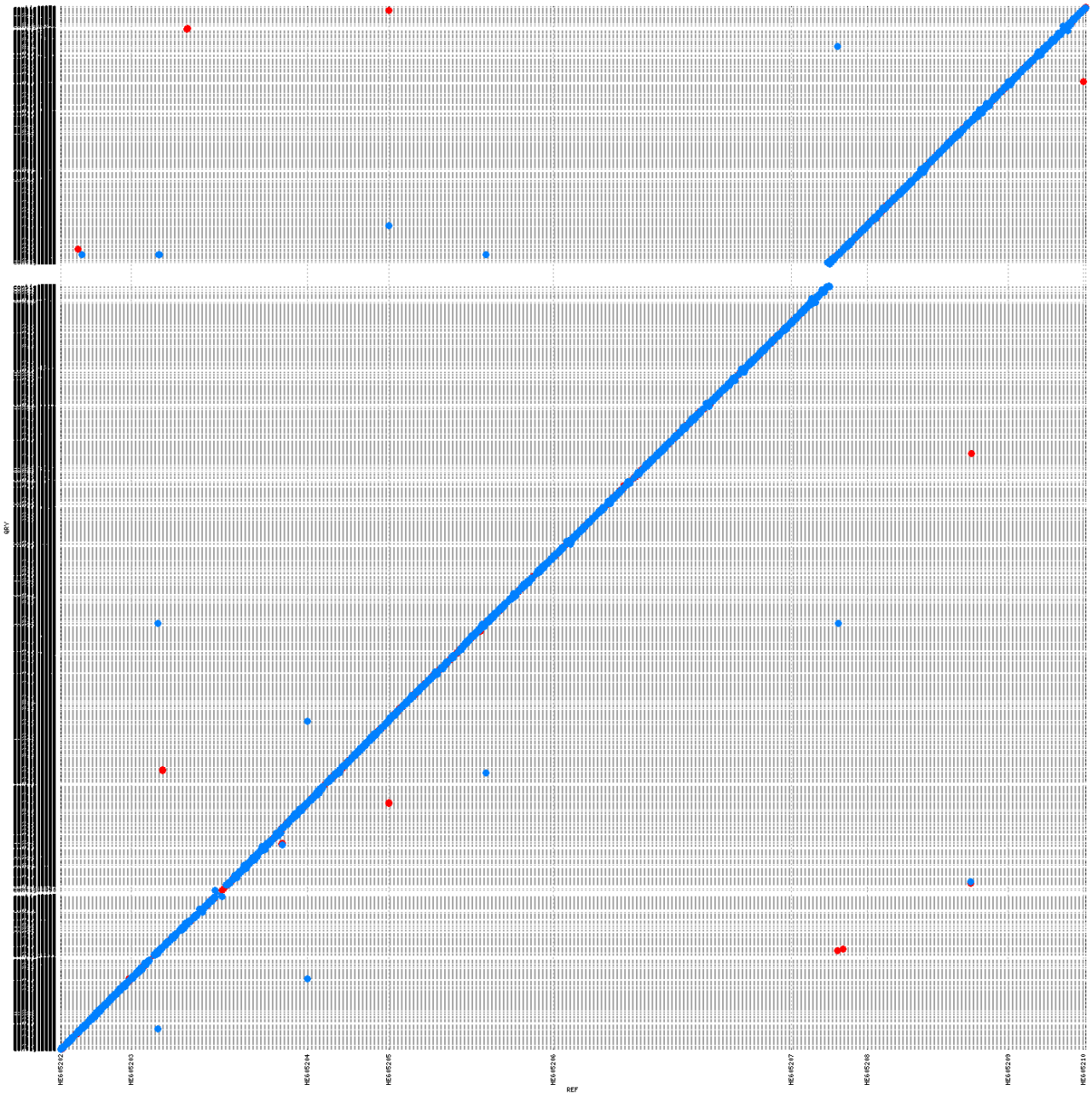
20% LOH started from scaffolds



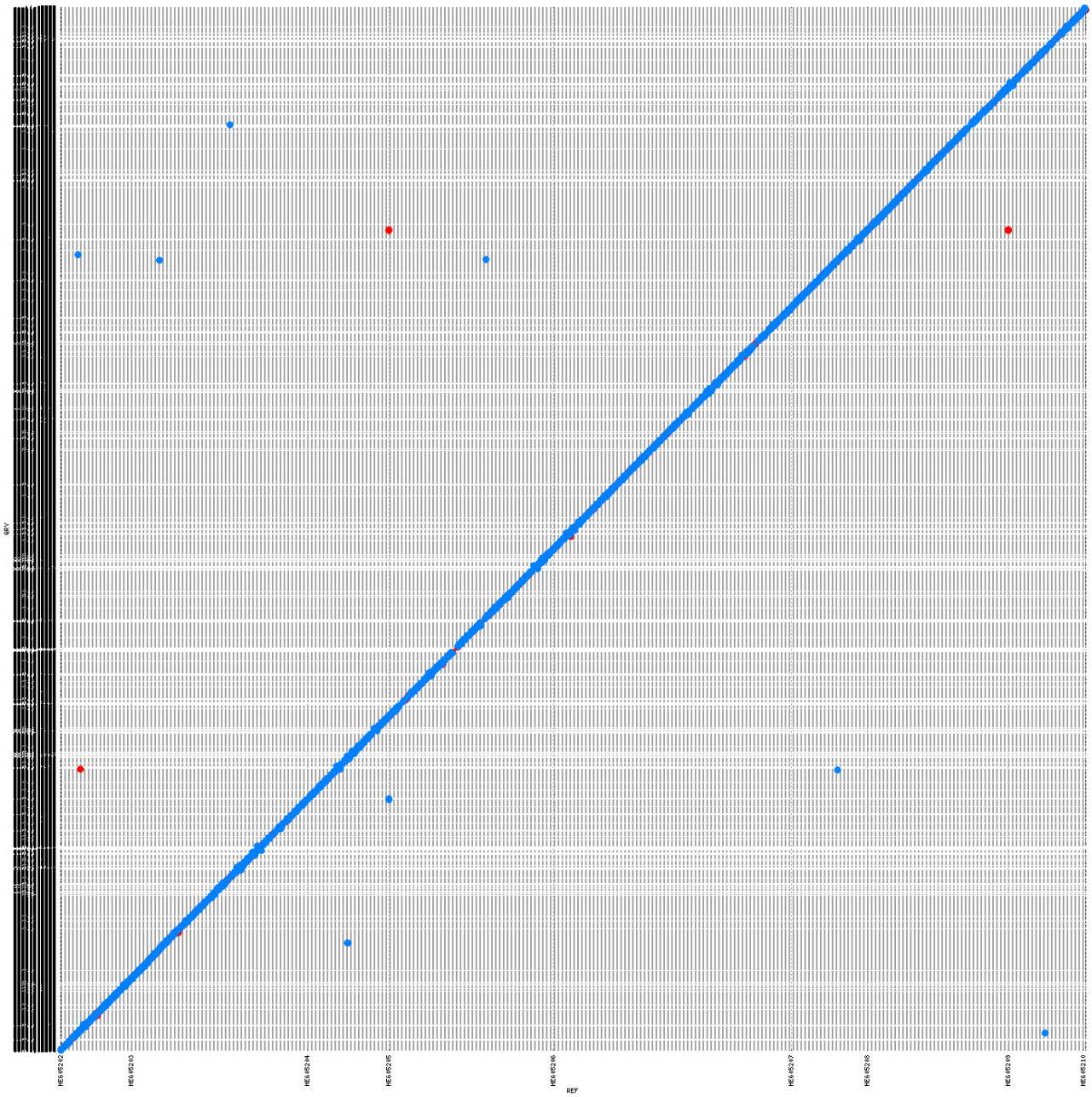
40% LOH started from contigs



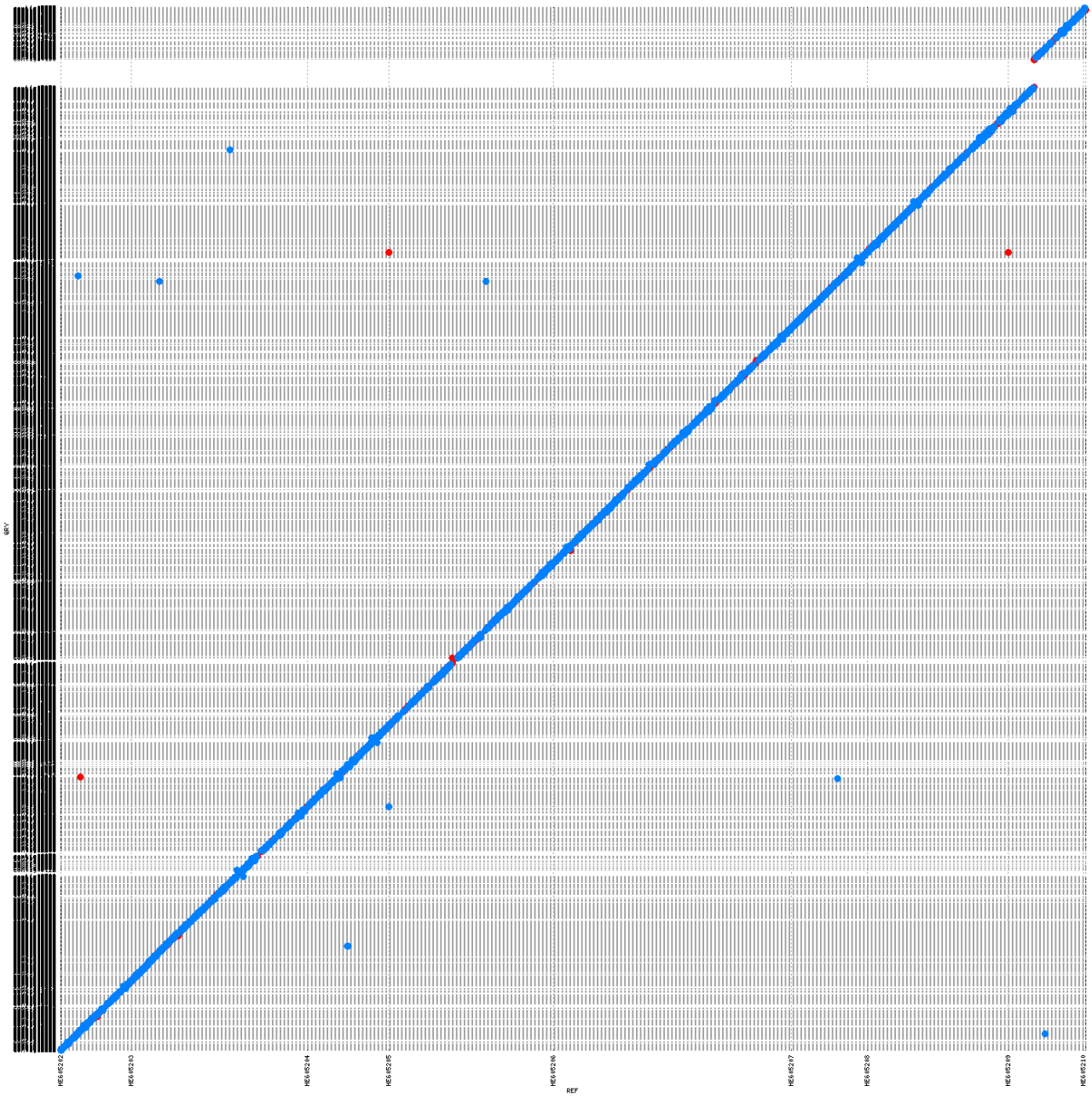
40% LOH started from scaffolds



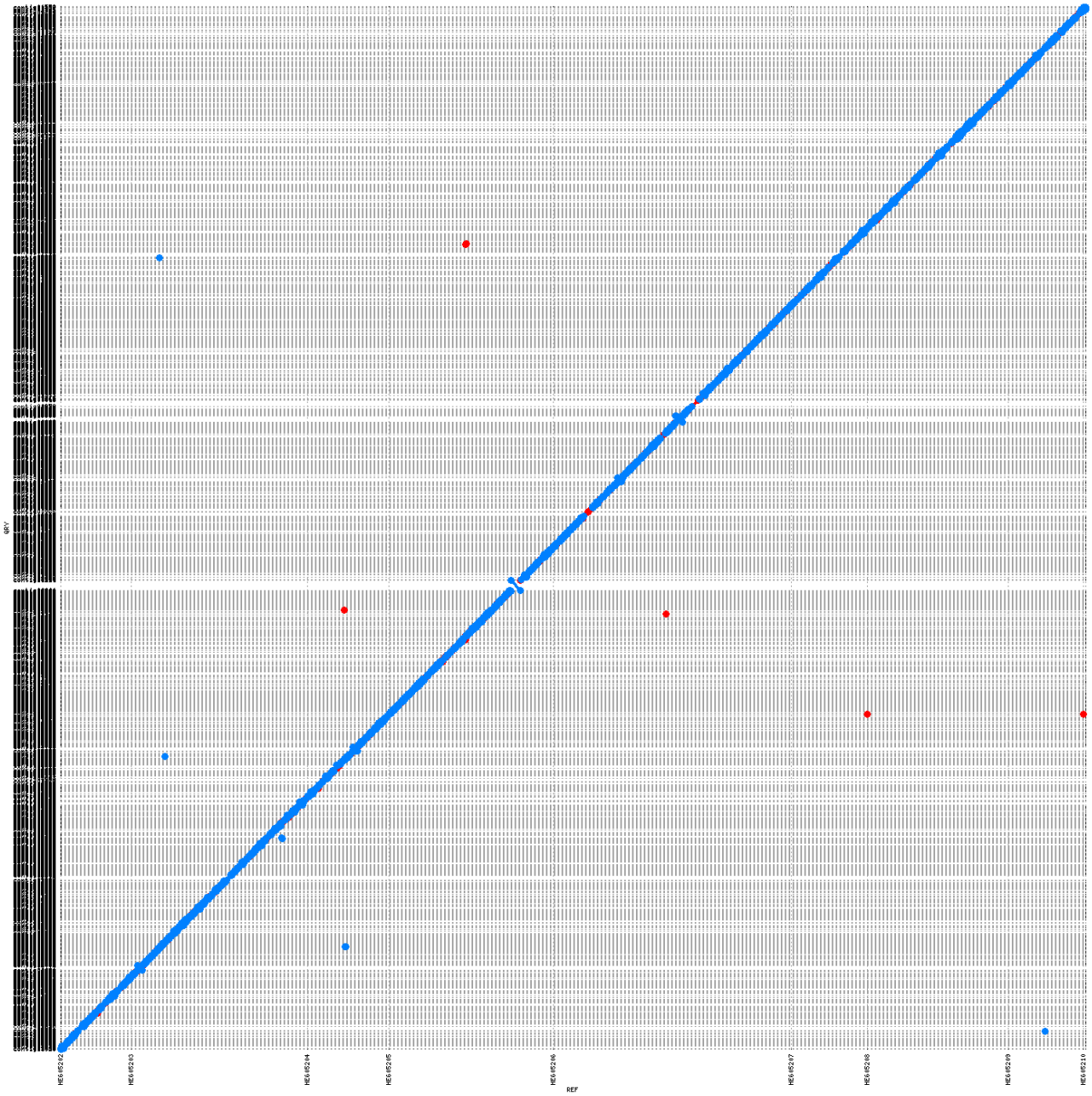
60% LOH started from contigs



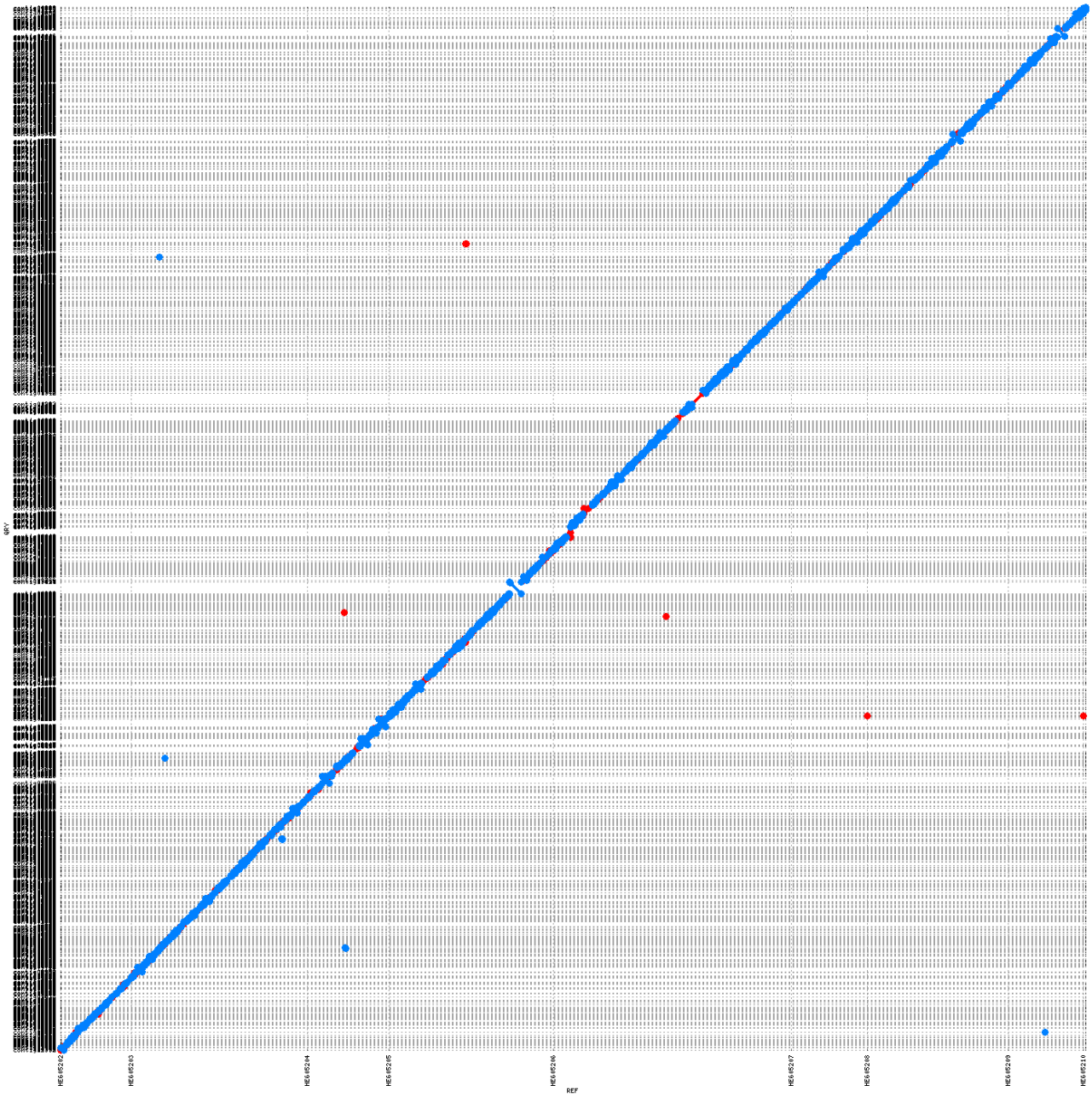
60% LOH started from scaffolds



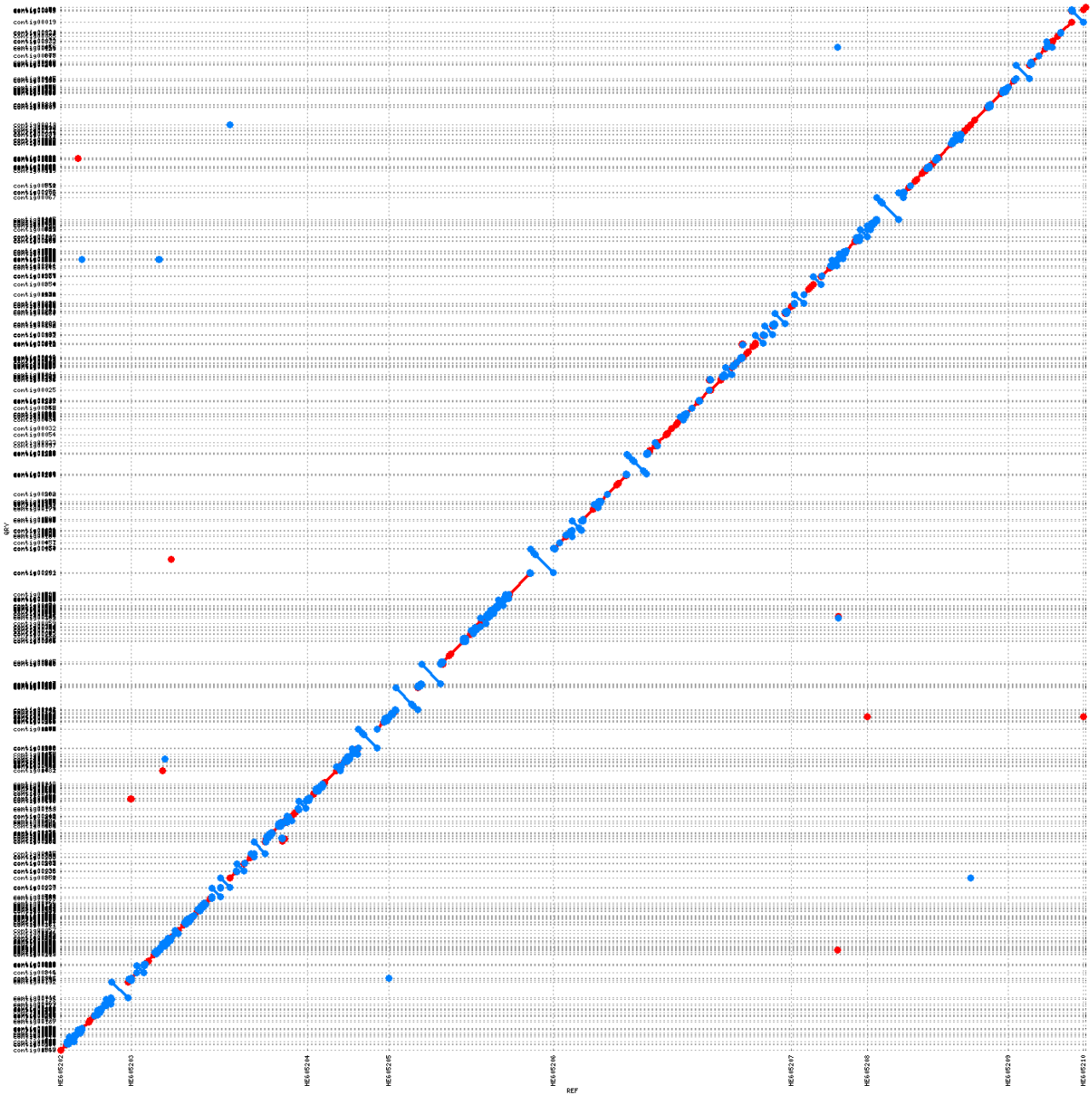
80% LOH started from contigs



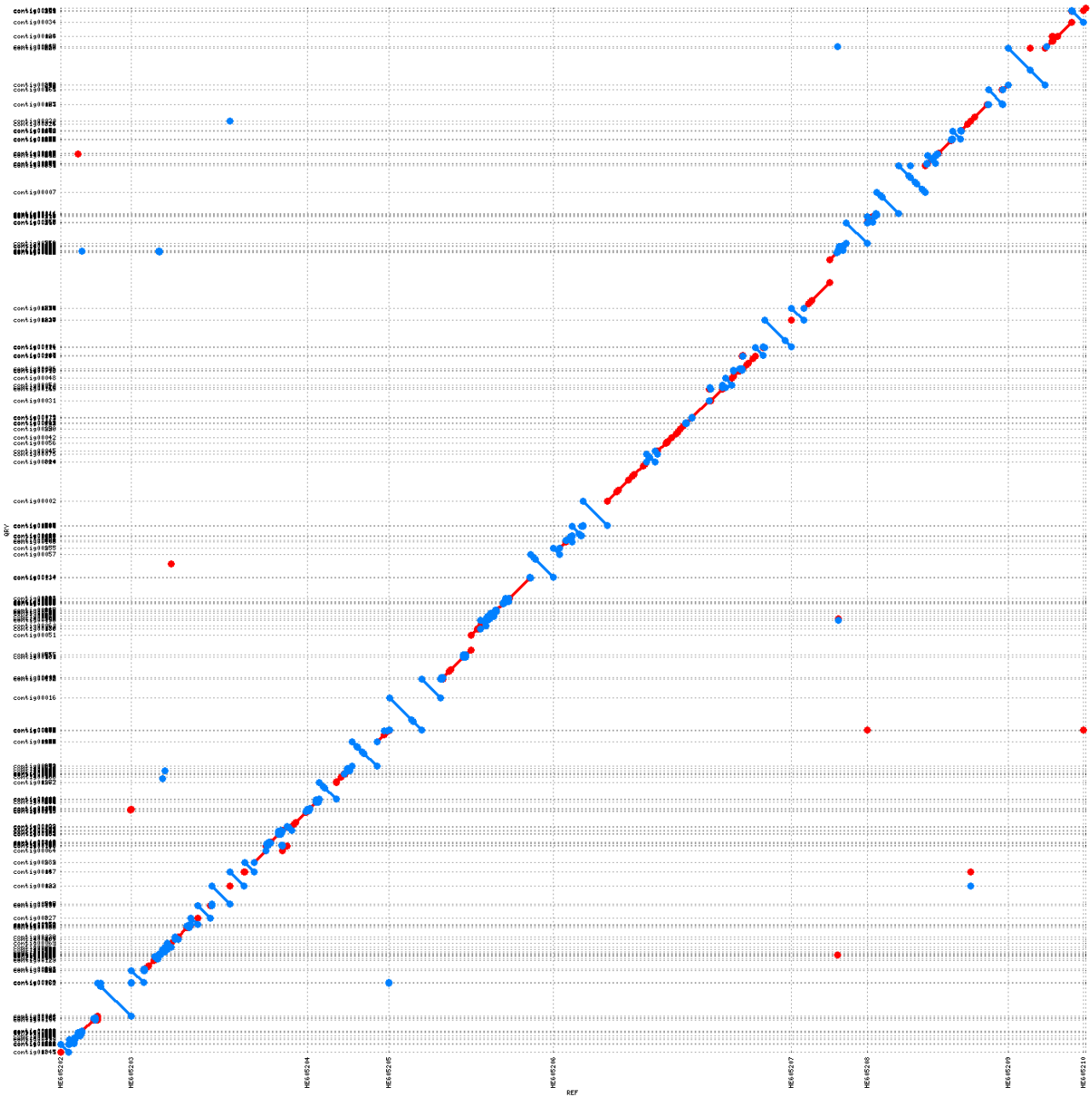
80% LOH started from scaffolds



100% LOH started from contigs



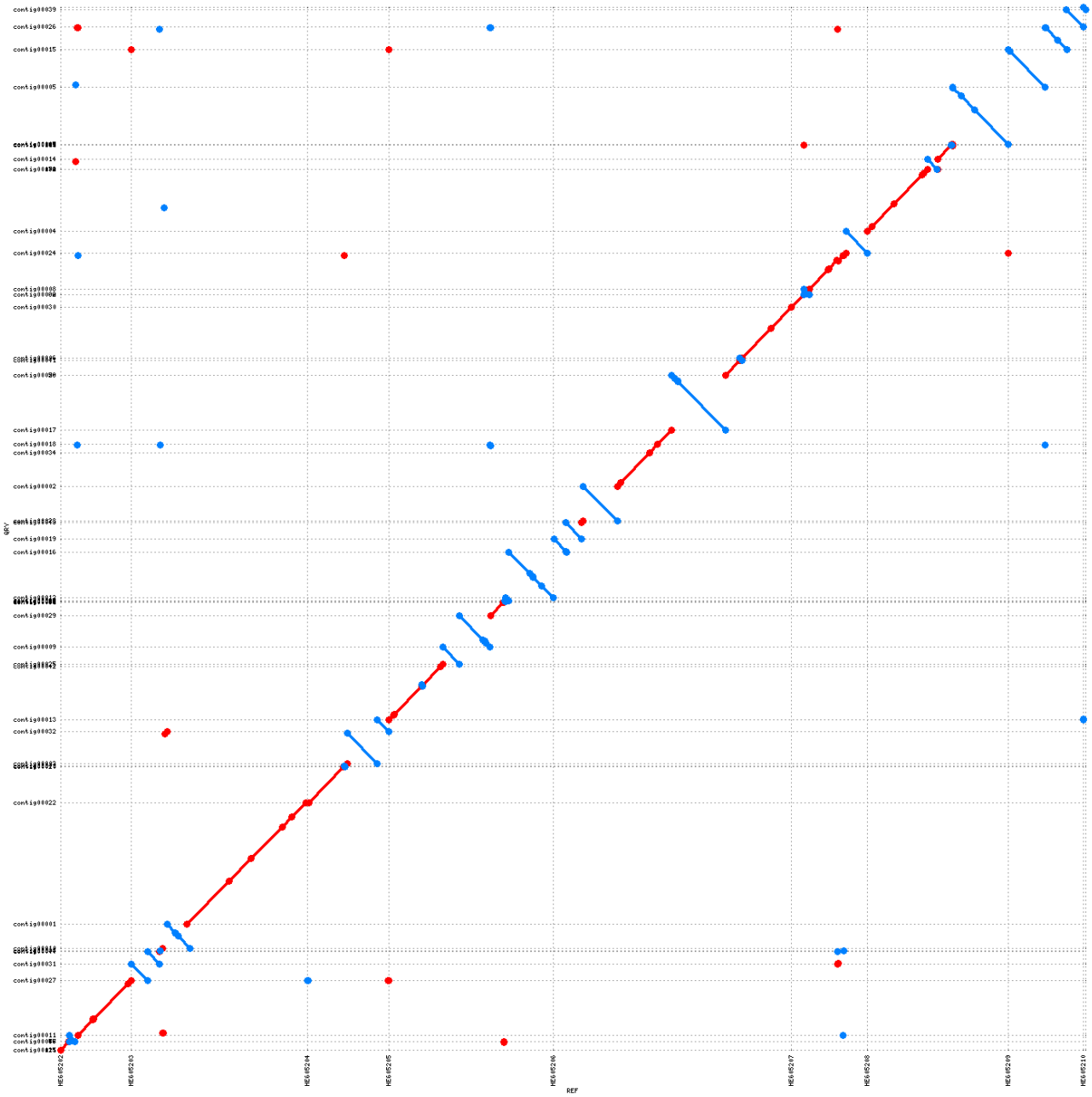
100% LOH started from scaffolds


QRY
REF
HE605202
HE605203
HE605204
HE605205
HE605206
HE605207
HE605208
HE605209
HE605210

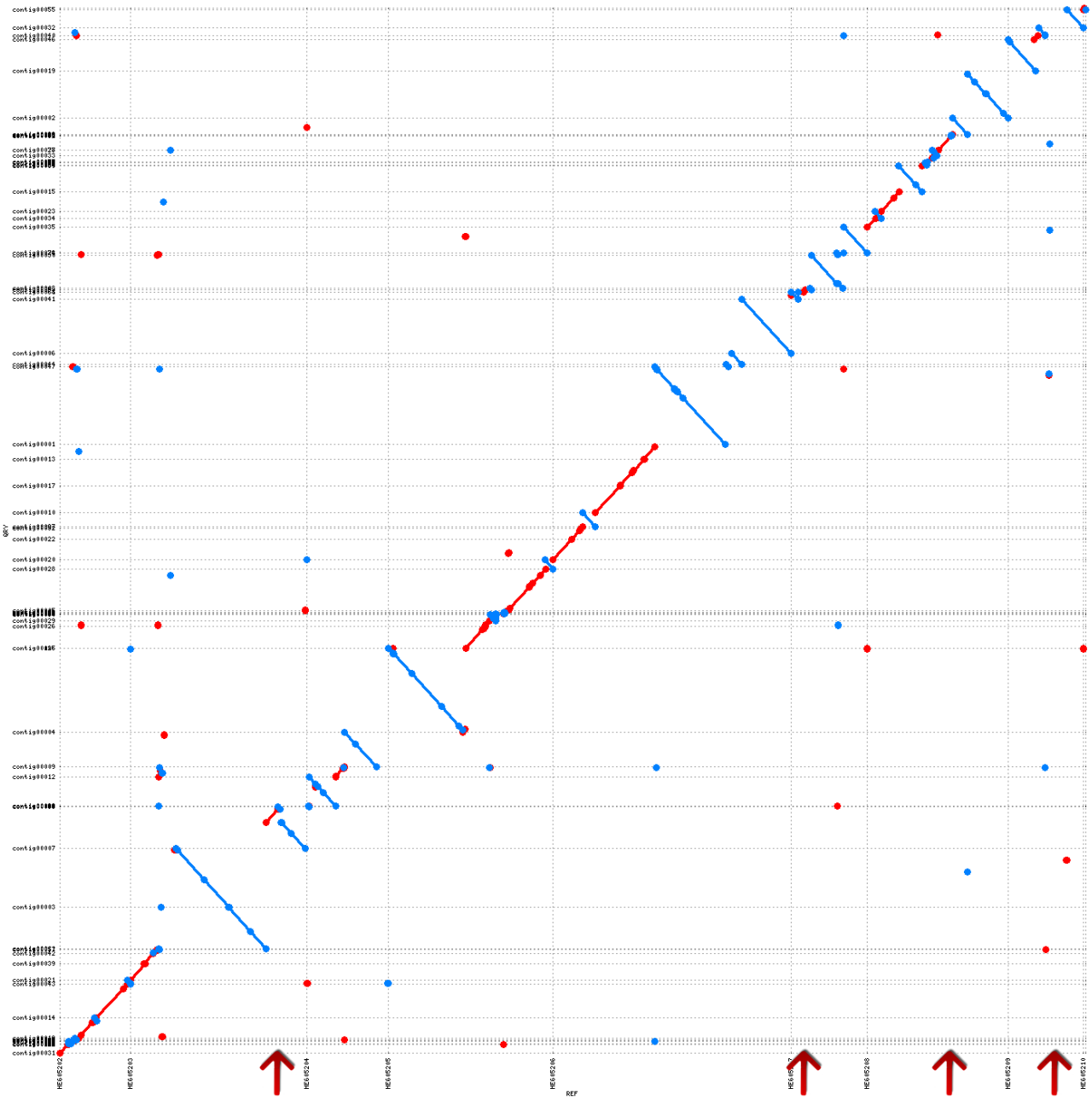
Supplementary figure S3: Alignments of assemblies from dipSPAdes

Consensus contigs returned by the dipSPAdes polymorphic genome assembly pipeline for various level of LOH have been aligned onto *C. parapsilosis* CDC317 chromosomes. The reference chromosomes are denoted on X axis, while query contigs/scaffolds are denoted on Y axis. Best query-to-reference matches are denoted with dots, forward in red and reverse in blue. The regions of similarity spanning larger regions are denoted by lines. Subsequently, the alignments were scanned for potential rearrangements and missing reference regions (marked by arrows on reference axis).

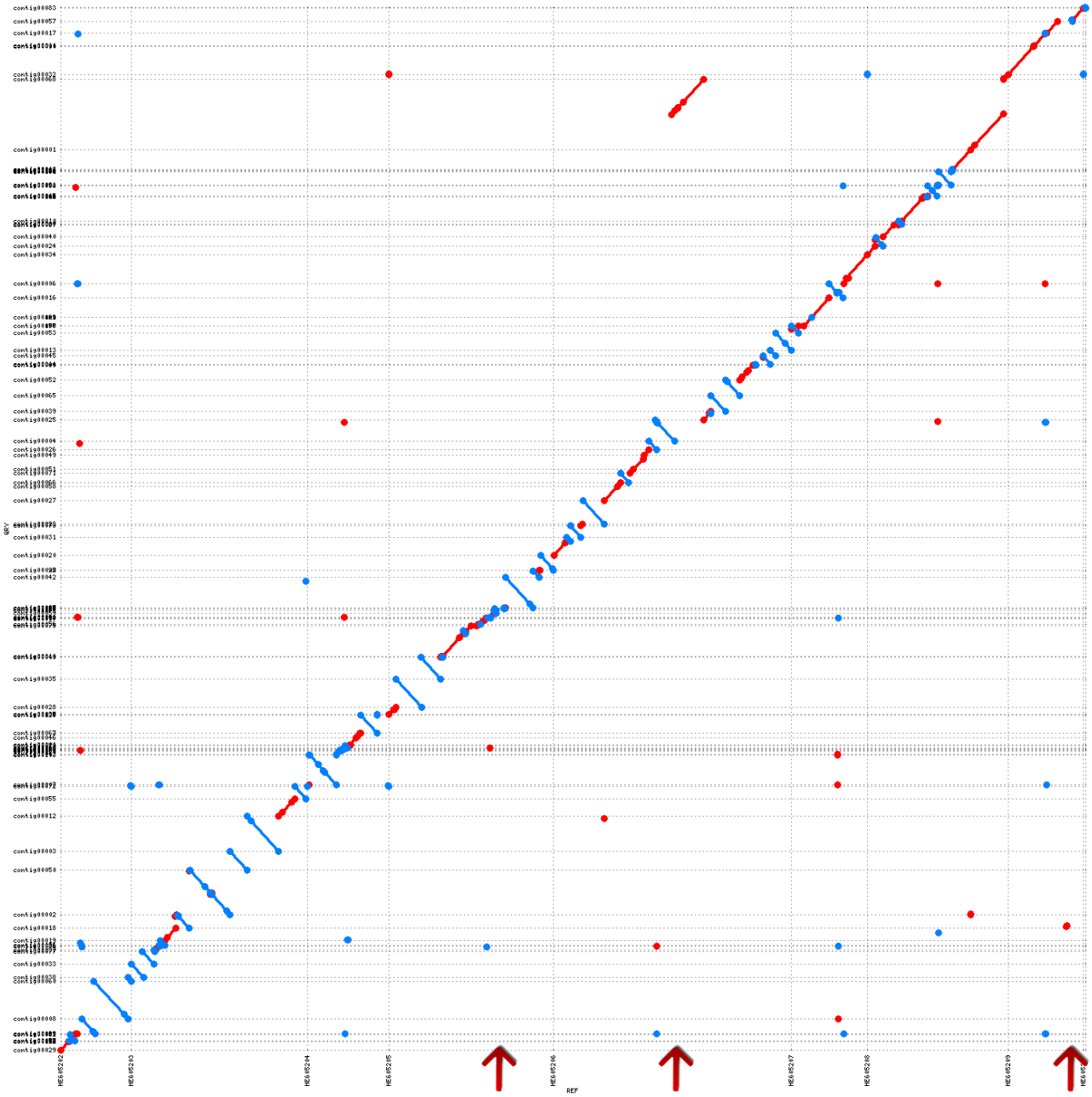
0% LOH


contig00039
contig00026
contig00015
contig00005
contig00014
contig00004
contig00024
contig00030
contig00017
contig00018
contig00034
contig00002
contig00019
contig00016
contig00029
contig00009
contig00013
contig00032
contig00022
contig00001
contig00031
contig00027
contig00011
HE605202
HE605203
HE605204
HE605205
HE605206
HE605207
HE605208
HE605209
HE605210
REF

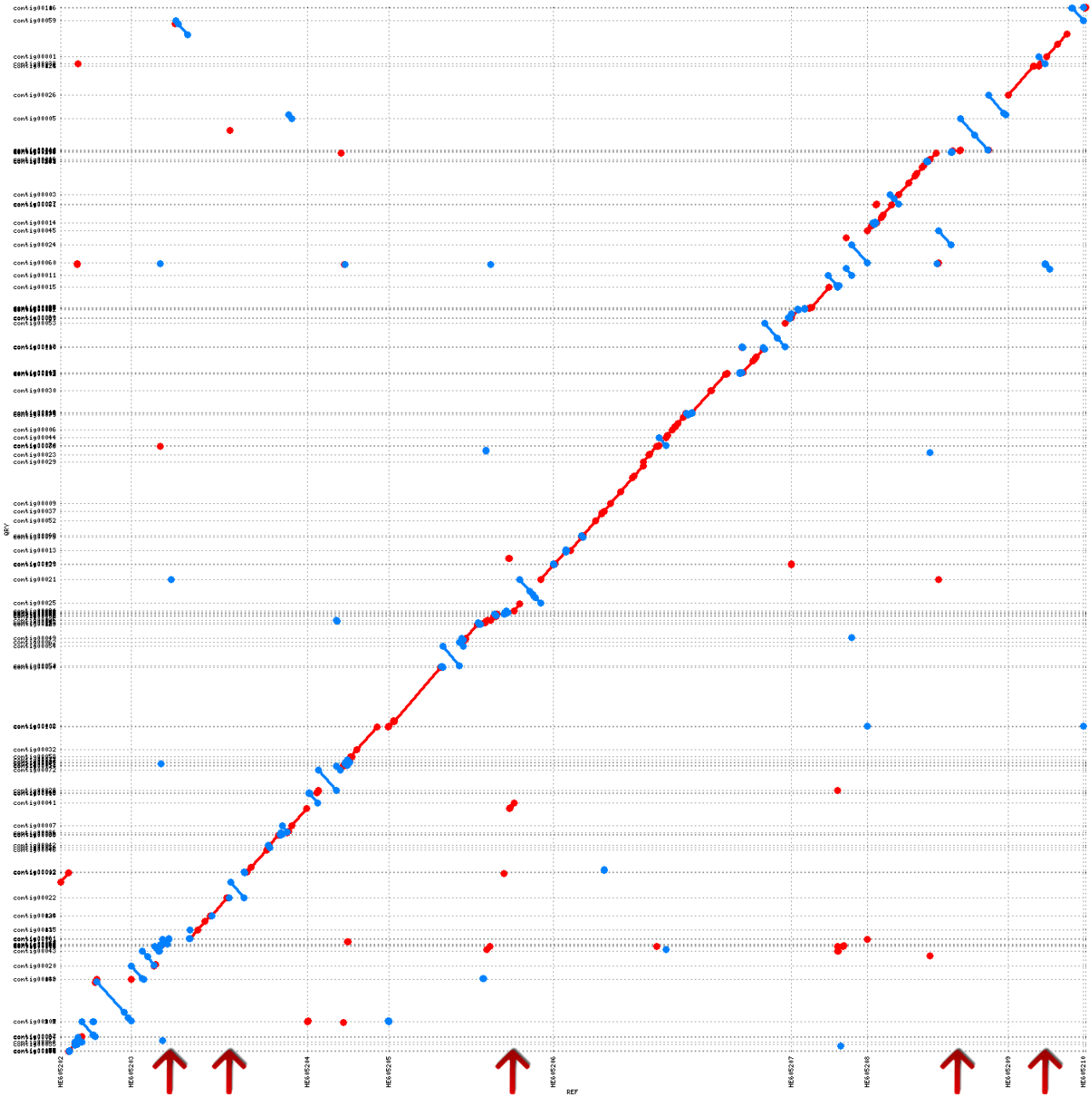
20% LOH



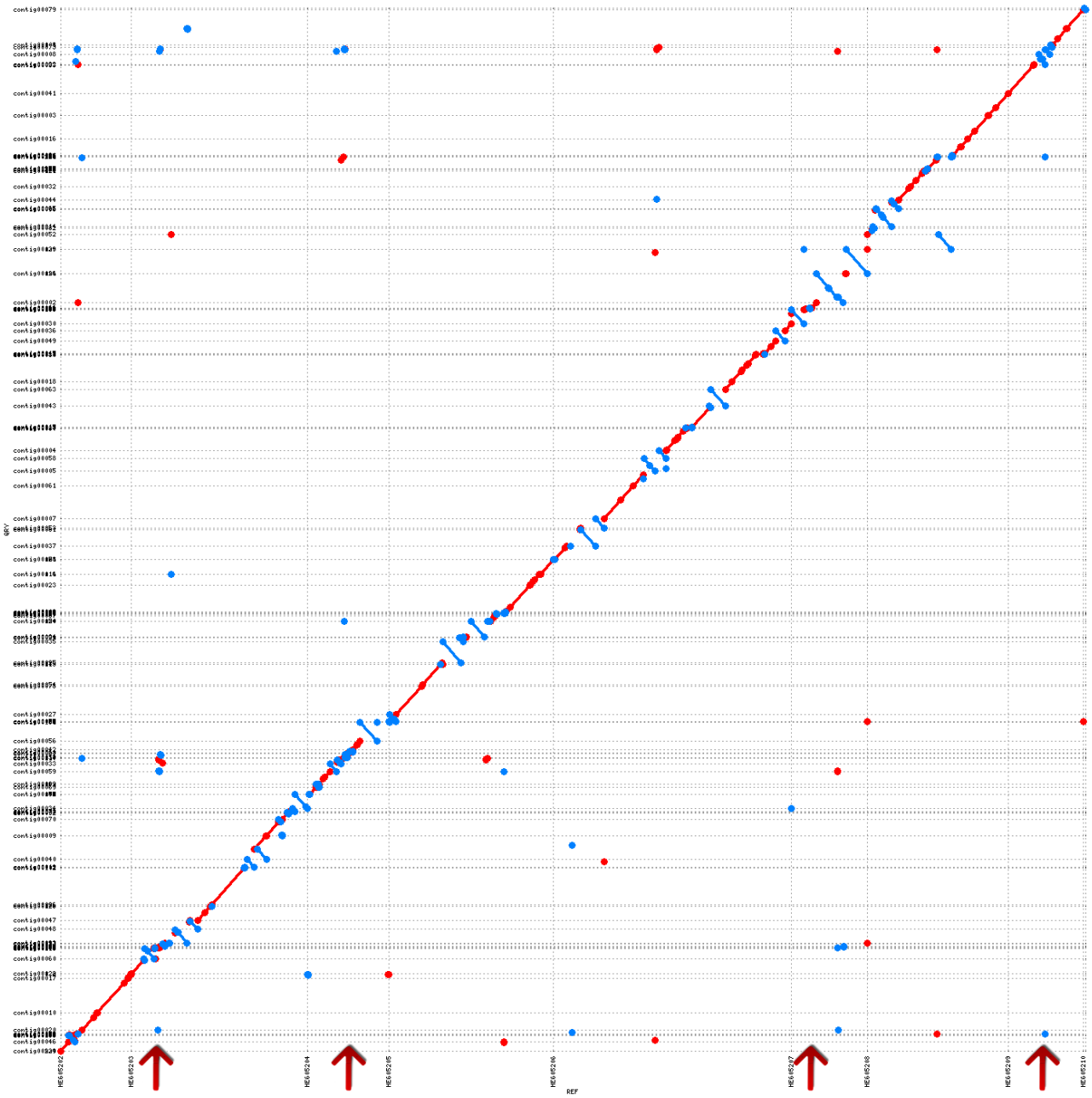
40% LOH



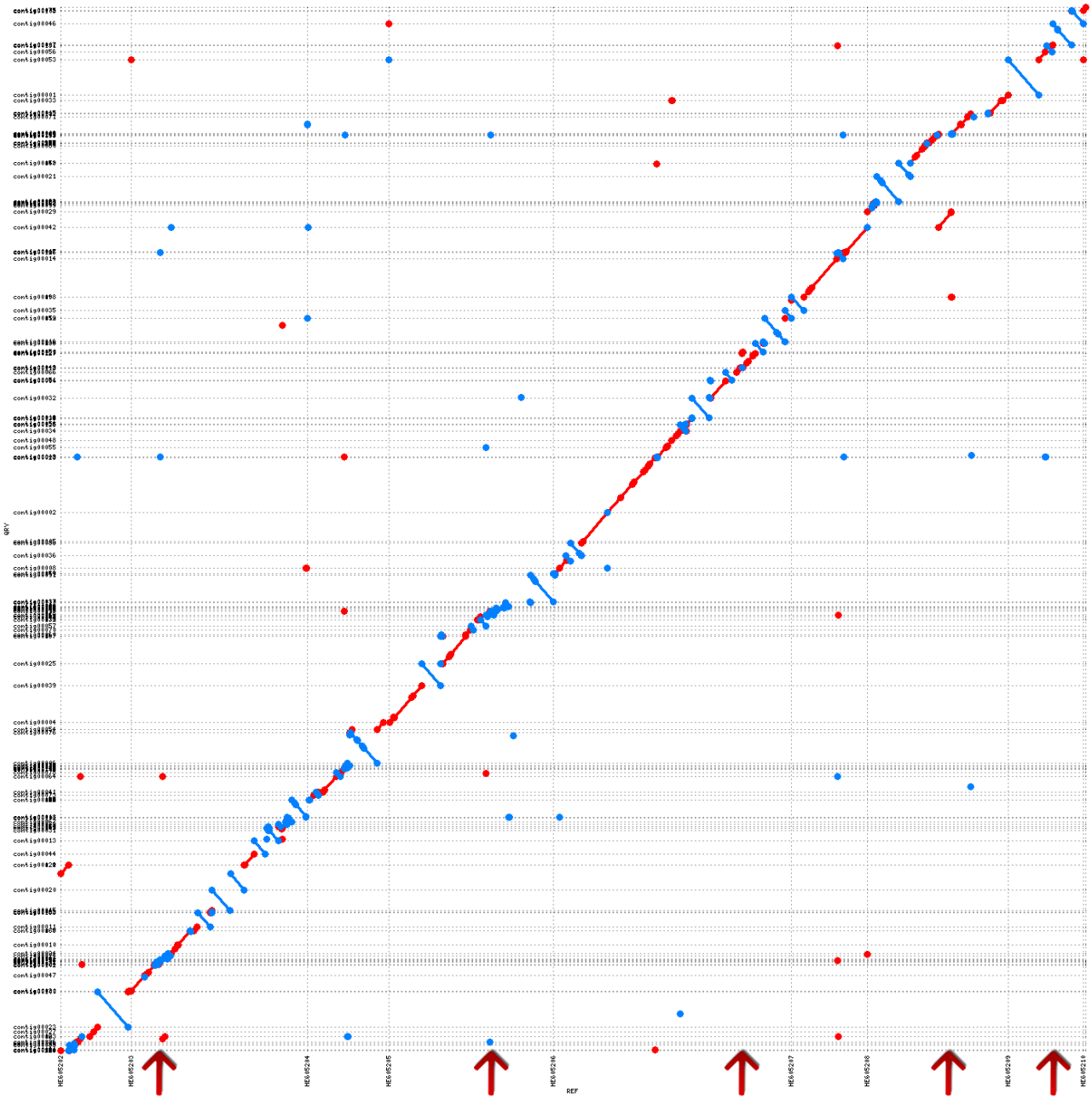
60% LOH


QRY
REF
HE6#5242
HE6#5243
HE6#5244
HE6#5245
HE6#5246
HE6#5247
HE6#5248
HE6#5249
HE6#5250

80% LOH



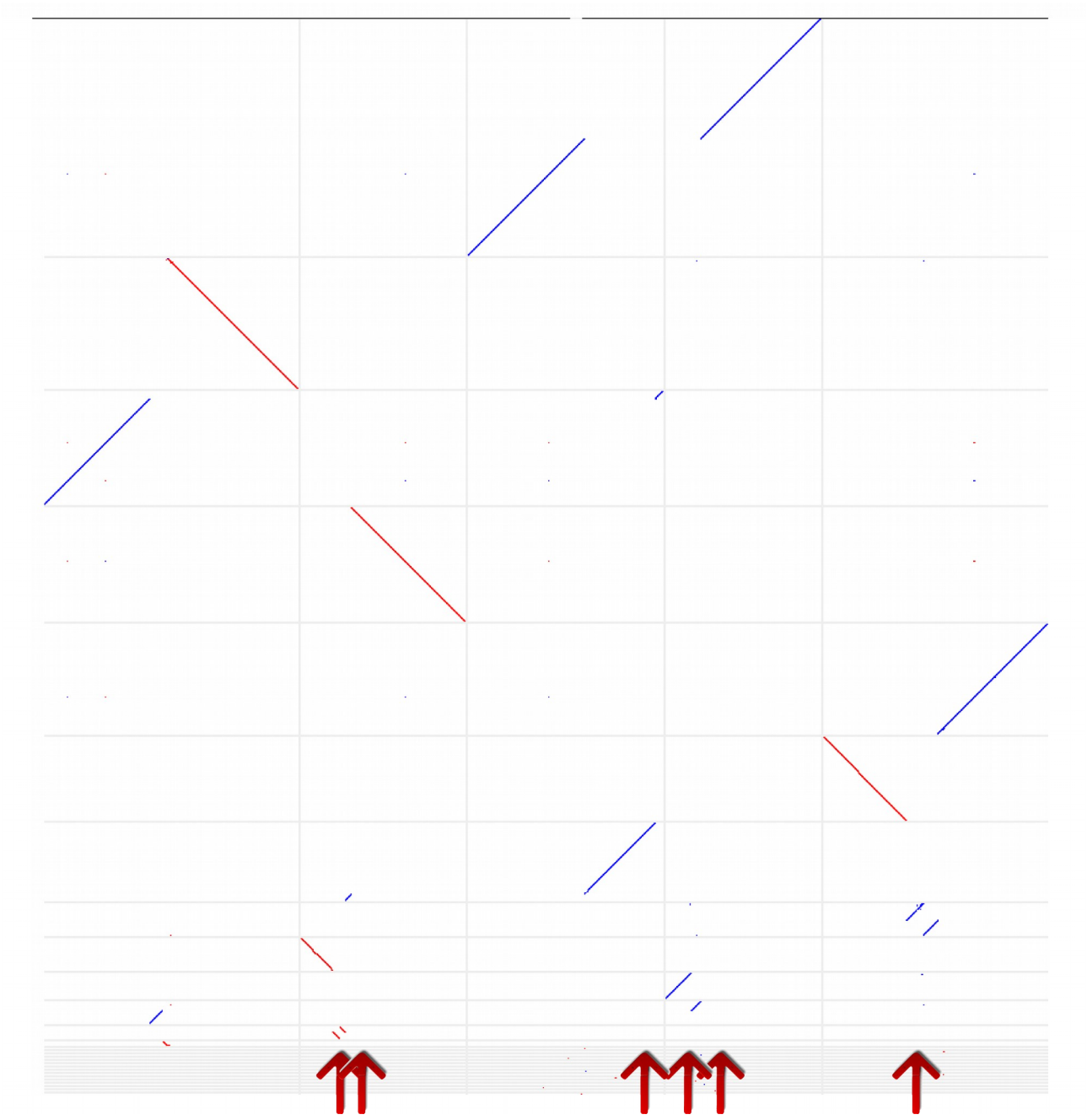
100% LOH



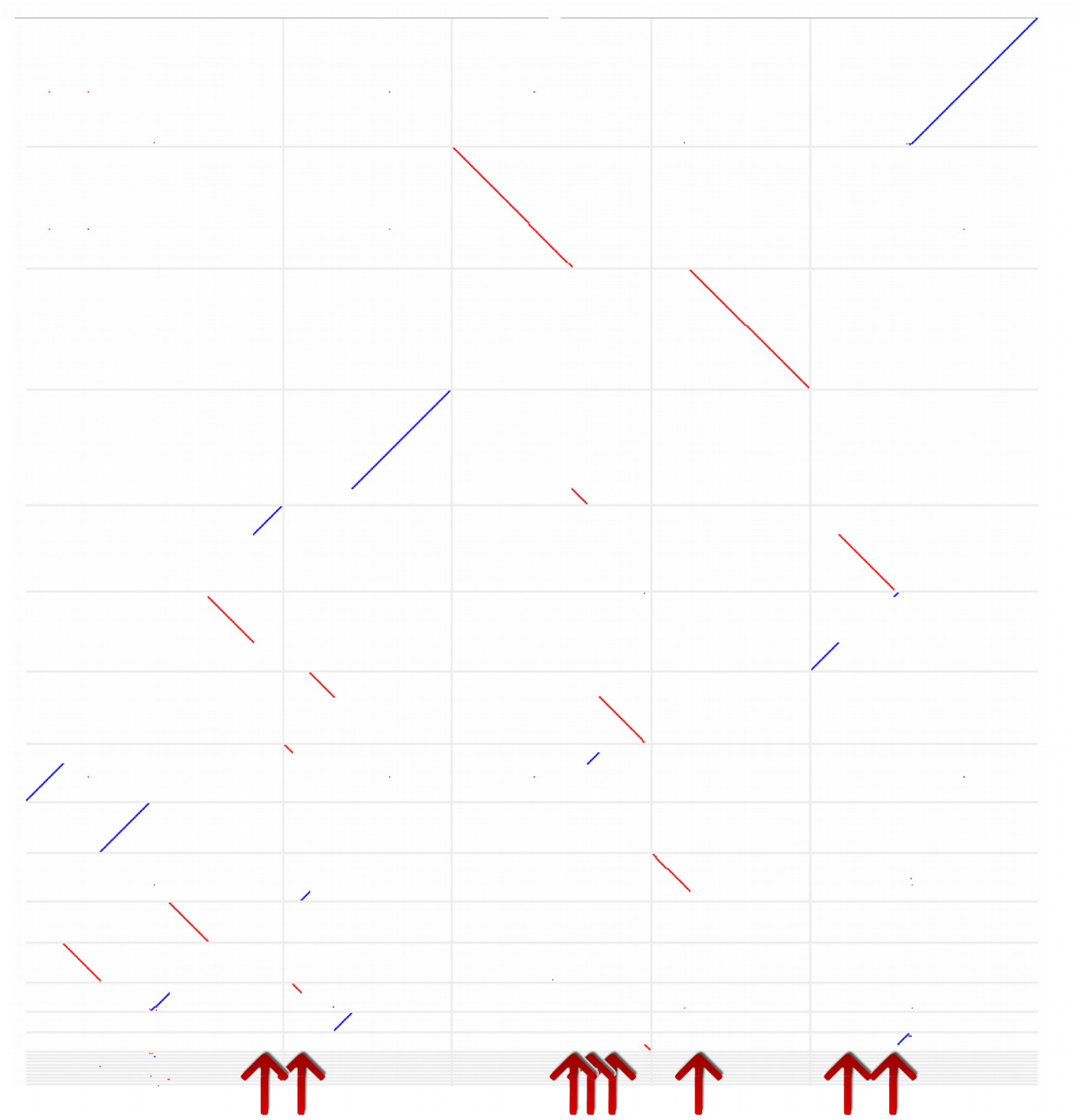
Supplementary figure S4: Alignments of assemblies from simulated plant genomes

Scaffolds returned by Redundas for various level of LOH have been aligned onto *A. thaliana* chromosomes using LAST aligner. The reference chromosomes are denoted on X axis, while query contigs/scaffolds are denoted on Y axis. Query-to-reference matches are denoted with dots, forward in blue and reverse in red. The regions of similarity spanning regions larger than 10 kb are denoted by lines. Subsequently, the alignments were scanned for potential rearrangements and missing reference regions (marked by arrows on reference axis).

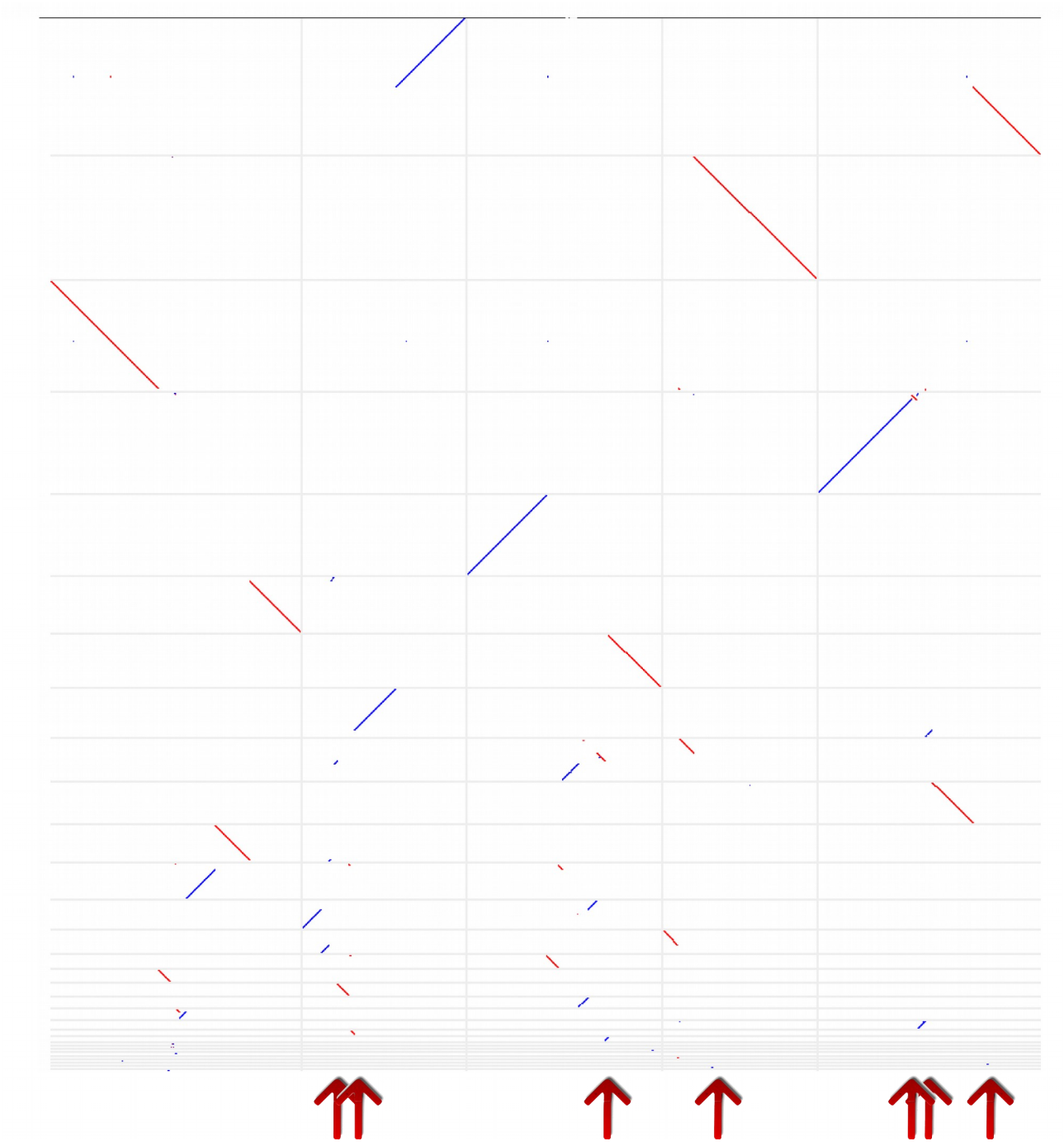
0% LOH



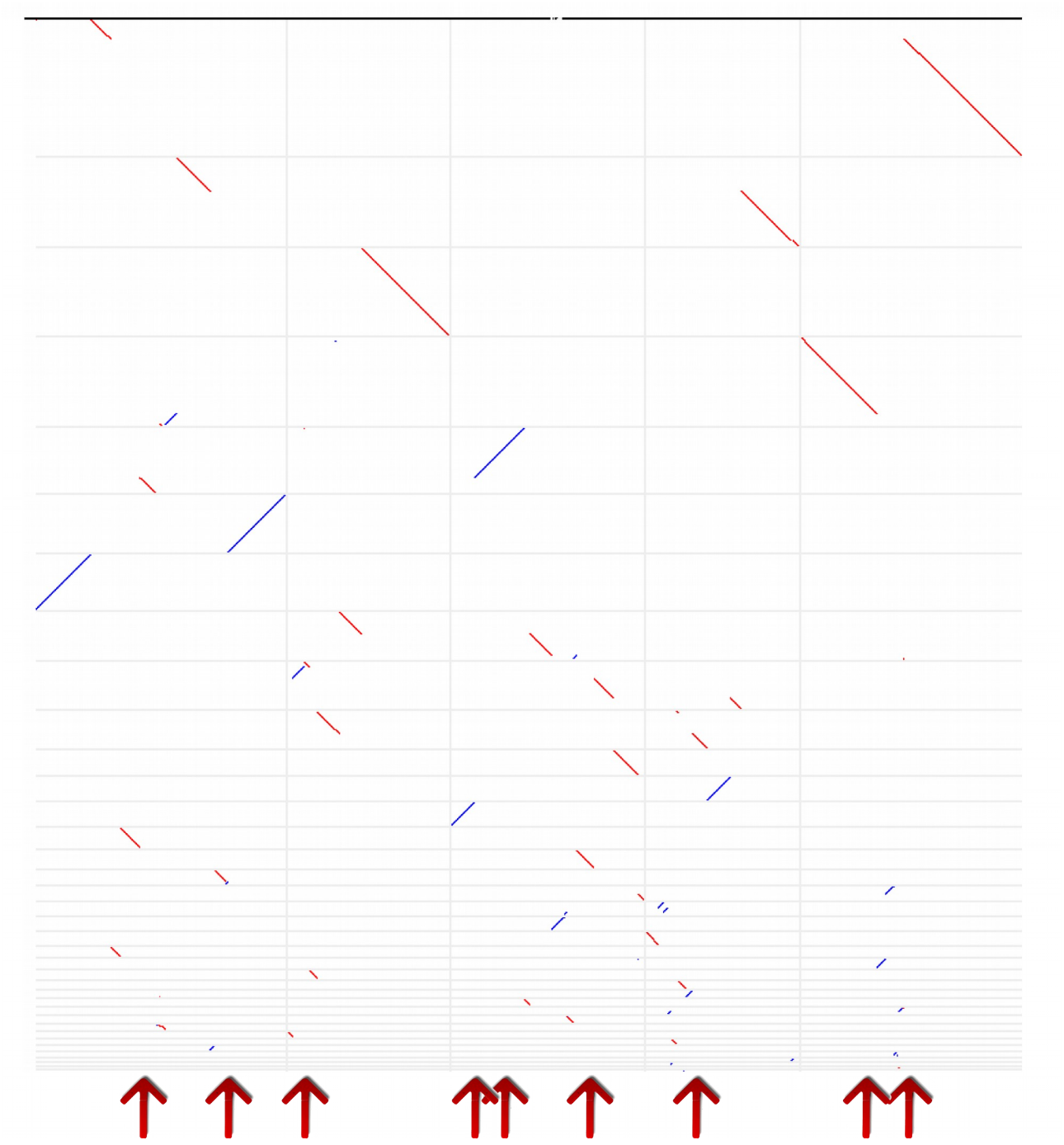
20% LOH



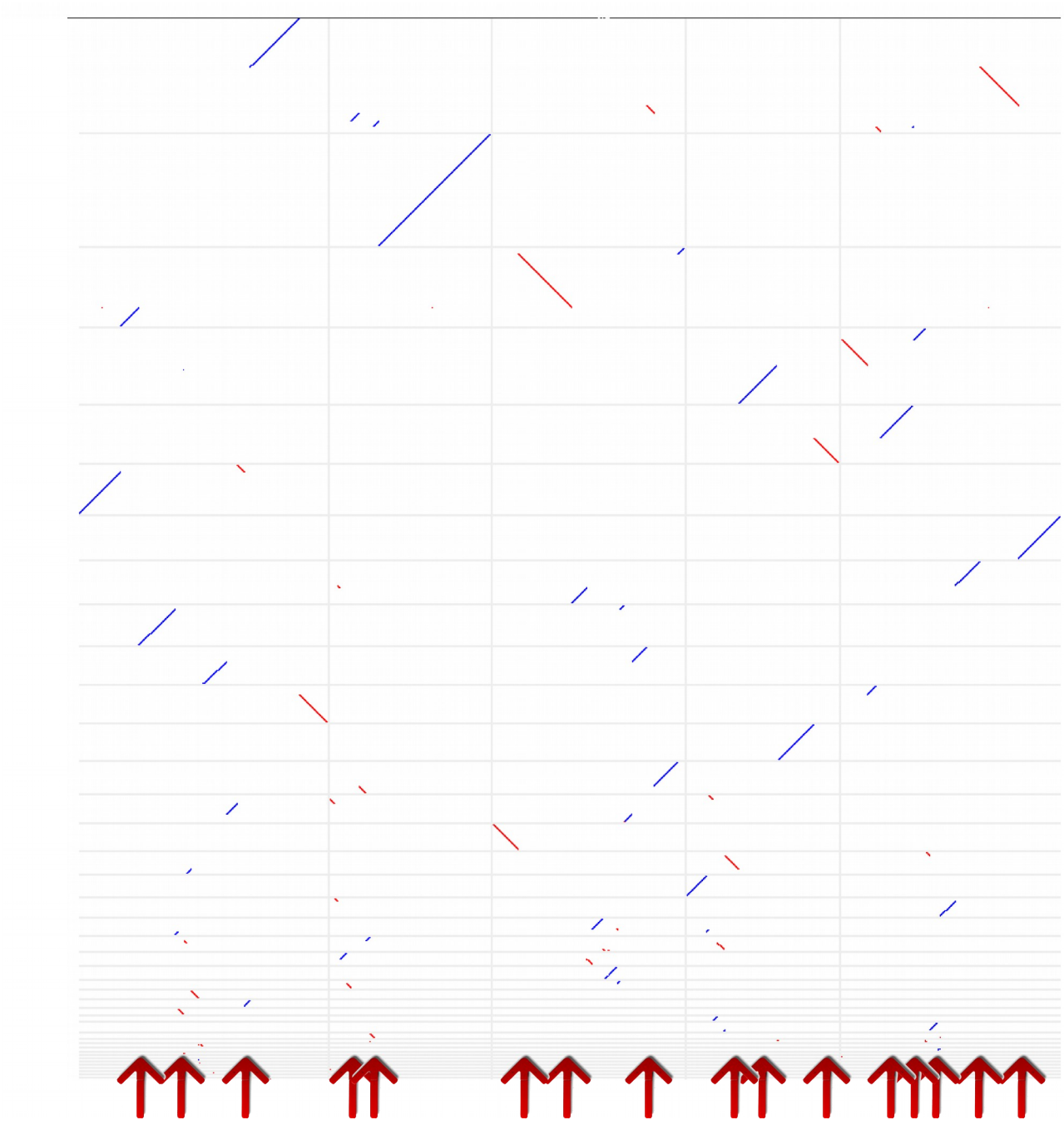
40% LOH



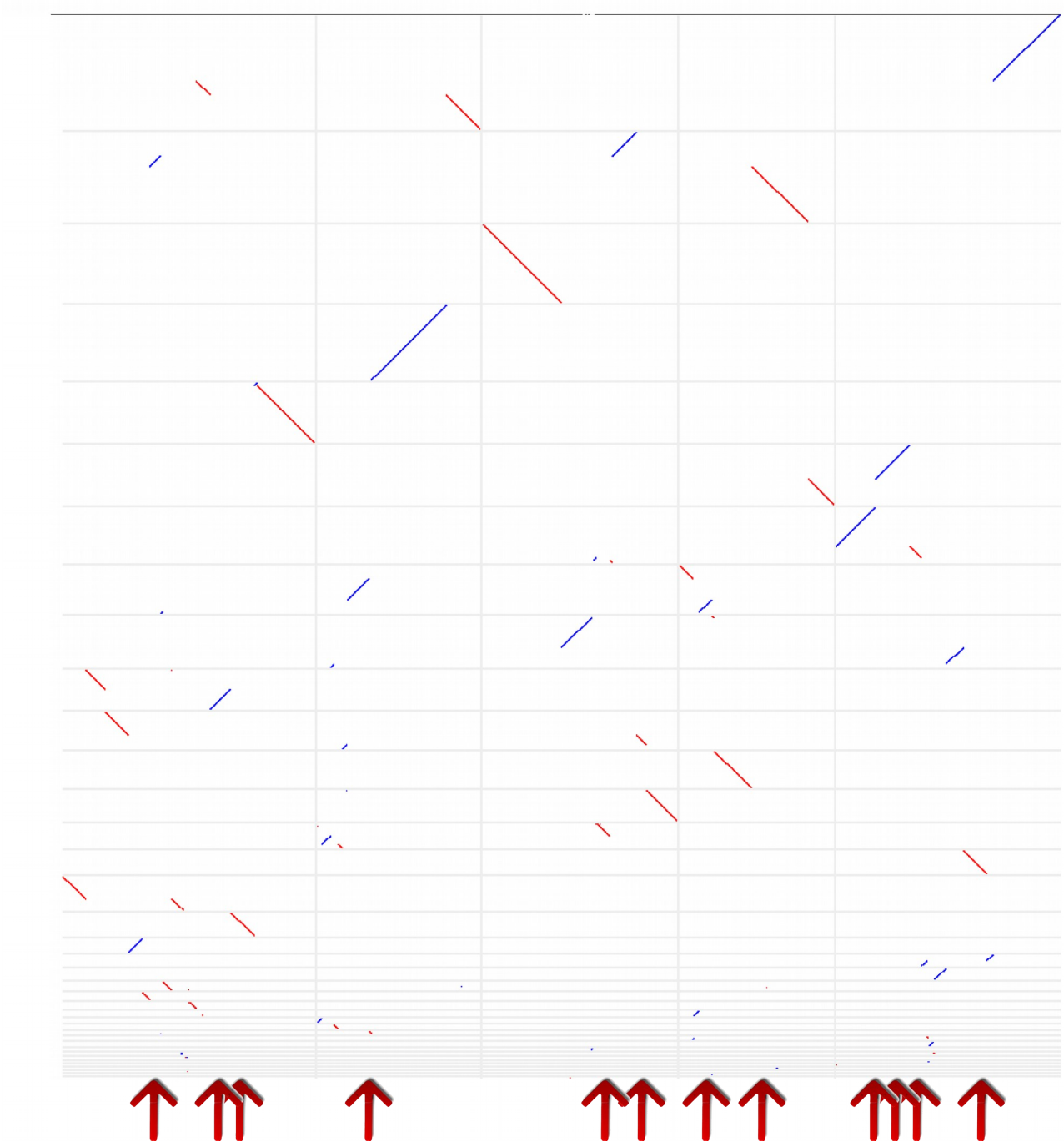
60% LOH



80% LOH

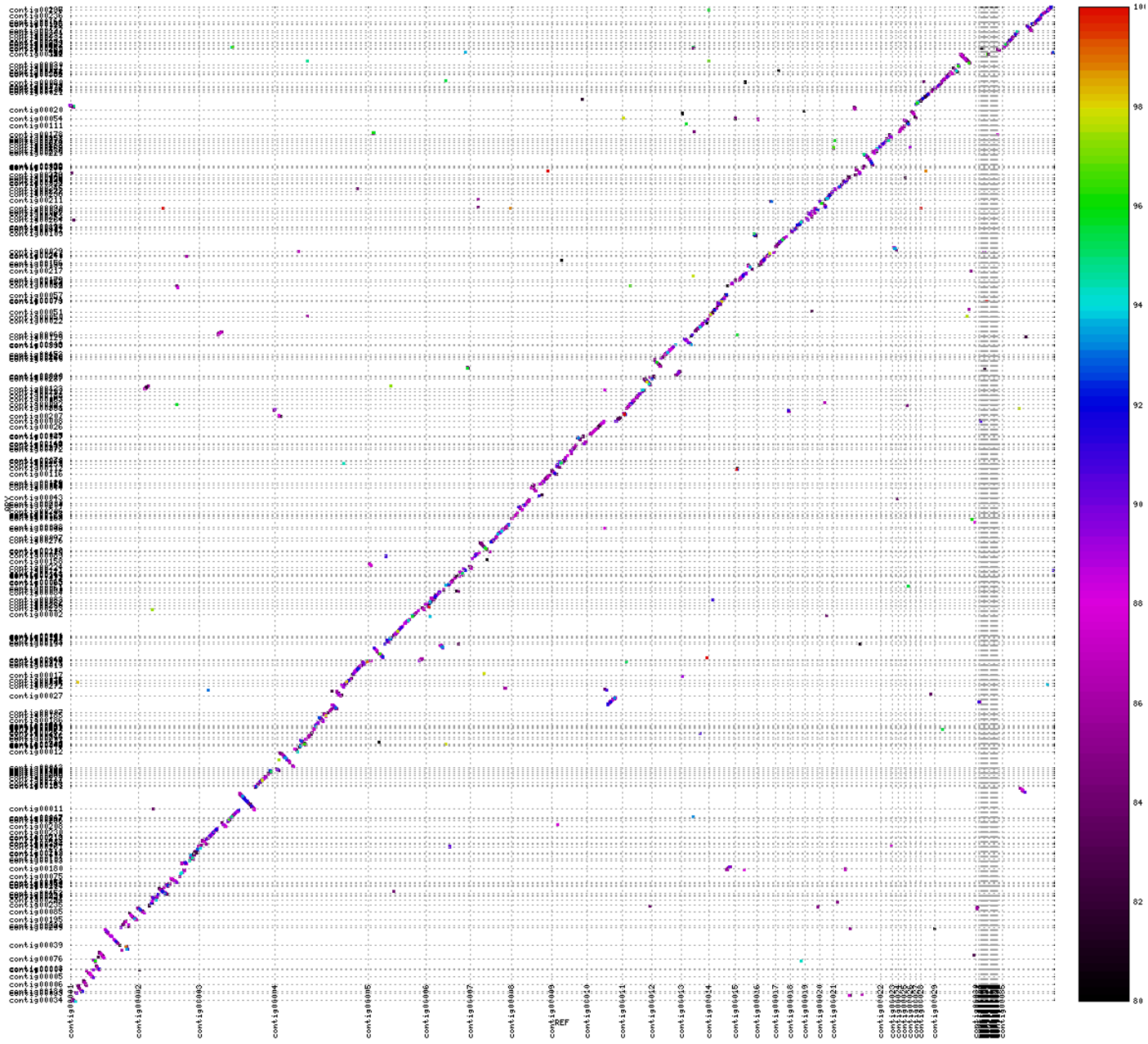


100% LOH



Supplementary figure S5: An alignment of *W. anomalus* and *W. ciferrii*

Scaffolds from homozygous *W. ciferrii* genome (Y axis) were aligned against *W. anomalus* scaffolds (X axis) reconstructed by heterozygous genome assembly pipeline. Synteny blocks have been coloured accordingly to the identity level between pair of query and target sequences. Noteworthy, the heterozygous genome of *W. ciferri* reconstructed by Redundans is less fragmented that the assembly for closely-related *W. anomalus*.



Supplementary table S1

Examples of heterozygous and homozygous genome assemblies retrieved from GenBank. For each analysed assembly, the table provides: species name, accession with the link to GenBank, type (contigs or scaffolds), size, number of contigs/scaffolds, cumulative size and number of identified heterozygous contigs/scaffolds, and cumulative size and number of non-redundant contigs/scaffolds.

Supplementary table S2

Simulated heterozygous genomes with various level of loss of heterozygosity were assembled using SPAdes, SOAPdevono, dipSPAdes and platanus. For each assembly, the table provides: the tool and parameters used, assembly type (contigs or scaffolds), simulated loss of heterozygosity level, cumulative size, number of contigs/scaffolds, cumulative size and number of identified heterozygous contigs/scaffolds, cumulative size and number of non-redundant contigs/scaffolds. Finally, the ratio of observed versus expected size is given as percentage for each assembly.

Supplementary table S3

Basic assembly statistics for simulated heterozygous genomes recovered by heterozygous genome assembly pipeline. The reconstructions were started from contigs produced by SPAdes. Number of contigs, cumulative assembly size, percentage of GC content, number of contigs longer than 1 kb and the cumulative size of these contigs, N50, N90, the cumulative size of gaps and the length of the longest contigs are given for each step and iteration of heterozygous genome assembly pipeline. Finally, the ratio of observed versus expected size (percentage), runtime, number of CPU cores and peak memory usage are given for each step of heterozygous genome assembly pipeline.

Supplementary table S4

Basic assembly statistics for simulated heterozygous genomes recovered by heterozygous genome assembly pipeline. The reconstructions were started from scaffolds produced by SPAdes. Number of contigs, cumulative assembly size, percentage of GC content, number of contigs longer than 1 kb and the cumulative size of these contigs, N50, N90, the cumulative size of gaps and the length of the longest contigs are given for each step and iteration of heterozygous genome assembly pipeline. Finally, the ratio of observed versus expected size is given as percentage for each assembly.

Supplementary table S5

Basic assembly statistics for simulated heterozygous genomes recovered by heterozygous genome assembly pipeline. The reconstructions were started from contigs and scaffolds produced by SPAdes. Number of contigs, cumulative assembly size, percentage of GC content, number of contigs longer than 1 kb and the cumulative size of these contigs, N50, N90, the cumulative size of gaps, the length of the longest contigs and the ratio of observed versus expected size are given as

percentage for each assembly. The assemblies returned by dipSPAdes and Platanus are also given for comparison purposes.

Supplementary table S6

Heterozygous genome assembly pipeline was applied to *Wickerhamomyces anomalus* contigs and scaffolds (AEGI01). Number of contigs, cumulative assembly size, percentage of GC content, number of contigs longer than 1 kb and the cumulative size of these contigs, N50, N90, the cumulative size of gaps and the length of the longest contigs are given for each step and iteration of heterozygous genome assembly pipeline. Finally, the ratio of observed versus expected size of each assembly is given as percentage. The assembly size of closely related homozygous genome of *Wickerhamomyces ciferrii* is taken as expected size.

Supplementary table S7

Basic assembly statistics for simulated heterozygous plant genomes recovered by heterozygous genome assembly pipeline. The reconstructions were started from contigs and scaffolds produced by SPAdes and Platanus. Number of contigs, cumulative assembly size, percentage of GC content, number of contigs longer than 1 kb and the cumulative size of these contigs, N50, N90, the cumulative size of gaps, the length of the longest contigs and the ratio of observed versus expected size are given as percentage for each assembly. The assemblies returned by dipSPAdes and Platanus are also given for comparison purposes.

Supplementary table S8

Basic assembly statistics for simulated heterozygous fungal genomes after heterozygous contig reduction with Haplomerger and Redundans (fasta2homozyous.py). The reduction was performed on contigs produced by SPAdes. Number of contigs, cumulative assembly size, percentage of GC content, number of contigs longer than 1 kb and the cumulative size of these contigs, N50, N90, the cumulative size of gaps, the length of the longest contigs and the ratio of observed versus expected size are given as percentage for each assembly. At the bottom, we provide the final assemblies after performing scaffolding and gap closing as implemented in Redundans (reduction was skipped with --noreduction parameter).
Note, Haplomerger was executed three times for each simulated assembly, while Redundans reduction step (fasta2homozyous.py) was executed only once.

Supplementary table S9

Basic assembly statistics simulated heterozygous plant genomes after heterozygous contig reduction with Haplomerger and Redundans (fasta2homozyous.py). The reduction was performed on contigs produced by SPAdes. Number of contigs, cumulative assembly size, percentage of GC content, number of contigs longer than 1 kb and the cumulative size of these contigs, N50, N90, the cumulative size of gaps, the length of the longest contigs and the ratio of observed versus expected size are given as percentage for each assembly. At the bottom, we provide the final assemblies after

performing scaffolding and gap closing as implemented in Redundans (reduction was skipped with --noreduction parameter).
Note, Haplomerger was executed three times for each simulated assembly, while Redundans reduction step (fasta2homozyous.py) was executed only once. Haplomerger crashed for 9 out of 12 genomes, which are marked with ‘Error’ message.