Updated tomato genome assembly based on FISH results.

Using FISH results Steven Stack, the tomato genome scaffolds were re-ordered, re-oriented, and gap sizes between scaffolds were set. The citation for the FISH and optical mapping work is

Lindsay A. Shearer, Lorinda K. Anderson, Hans de Jong, Sandra Smit, José Luis Goicoechea, Bruce A. Roe, Axin Hua, James J. Giovannoni, and Stephen M. Stack. 2014. Fluorescence in situ hybridization and optical mapping to correct scaffold arrangement in the tomato genome.G3 g3.114.011197; doi:10.1534/g3.114.011197 (<http://www.g3journal.org/content/early/2014/05/30/g3.114.011197.abstract>)

Tiling Path Files (TPF) were modified using perl scripts in the Bio::GenomeUpdate package (authored by Surya Saha and Jeremy Edwards at SGN) available at <https://github.com/solgenomics/Bio-GenomeUpdate>. Accessioned Golden Path (AGP) and assembled chromosome and scaffold sequences were generated from the TPFs by tools available through the NCBI Genome Reference Consortium (GRC). Table 1 lists the order and orientation changes made. Table 2 lists the gap sizes that were set. Figures 1-12 show the dotplot alignments between the original build SL2.40 chromosome sequences and the new SL2.50 assemblies based on FISH results. The MUMmer version used is 3.23 and alignments of more than 1000bp with 100% identity were used to construct the dot plots.

Table 1. Order and orientation of scaffolds from FISH results.

|  |  |  |  |
| --- | --- | --- | --- |
| Chr. | Scaffold | Original Scaffold# | Reversed |
| 1 | SL2.40sc04133 | 1 |  |
| 1 | SL2.40sc04191 | 3 | yes |
| 1 | SL2.40sc03666 | 2 |  |
| 1 | SL2.40sc03594 | 4 |  |
| 1 | SL2.40sc05010 | 5 |  |
| 1 | SL2.40sc05941 | 6 |  |
| 1 | SL2.40sc06917 | 8 |  |
| 1 | SL2.40sc06903 | 7 |  |
| 1 | SL2.40sc04323 | 9 |  |
| 2 | SL2.40sc04732 | 4 | yes |
| 2 | SL2.40sc04208 | 6 | yes |
| 2 | SL2.40sc05776 | 5 |  |
| 2 | SL2.40sc06593 | 1 |  |
| 2 | SL2.40sc04142 | 2 |  |
| 2 | SL2.40sc03766 | 3 |  |
| 2 | SL2.40sc03665 | 7 |  |
| 3 | SL2.40sc04439 | 1 |  |
| 3 | SL2.40sc04696 | 4 |  |
| 3 | SL2.40sc05330 | 5 |  |
| 3 | SL2.40sc04126 | 6 |  |
| 3 | SL2.40sc04616 | 9 |  |
| 3 | SL2.40sc06725 | 7 |  |
| 3 | SL2.40sc04704 | 8 |  |
| 3 | SL2.40sc04822 | 2 | yes |
| 3 | SL2.40sc03721 | 12 |  |
| 3 | SL2.40sc03806 | 10 |  |
| 3 | SL2.40sc03796 | 11 |  |
| 3 | SL2.40sc06911 | 3 | yes |
| 3 | SL2.40sc03701 | 13 |  |
| 4 | SL2.40sc03604 | 1 |  |
| 4 | SL2.40sc05339 | 3 | yes |
| 4 | SL2.40sc03683 | 2 | yes |
| 4 | SL2.40sc06101 | 4 |  |
| 4 | SL2.40sc04680 | 5 |  |
| 4 | SL2.40sc04135 | 6 |  |
| 5 | SL2.40sc03726 | 1 |  |
| 5 | SL2.40sc06155 | 2 | yes |
| 5 | SL2.40sc03902 | 3 |  |
| 6 | SL2.40sc04474 | 1 |  |
| 6 | SL2.40sc06140 | 3 | yes |
| 6 | SL2.40sc05383 | 2 |  |
| 6 | SL2.40sc04279 | 4 |  |
| 6 | SL2.40sc05188 | 5 | yes |
| 6 | SL2.40sc05732 | 6 |  |
| 6 | SL2.40sc05054 | 7 |  |
| 6 | SL2.40sc03622 | 8 |  |
| 7 | SL2.40sc03731 | 1 |  |
| 7 | SL2.40sc05397 | 2 |  |
| 7 | SL2.40sc03685 | 3 |  |
| 7 | SL2.40sc04626 | 4 |  |
| 8 | SL2.40sc04813 | 1 |  |
| 8 | SL2.40sc03770 | 2 |  |
| 8 | SL2.40sc04167 | 7 |  |
| 8 | SL2.40sc03749 | 5 | yes |
| 8 | SL2.40sc04236 | 6 | yes |
| 8 | SL2.40sc03835 | 3 | yes |
| 8 | SL2.40sc04701 | 4 | yes |
| 8 | SL2.40sc04948 | 8 |  |
| 8 | SL2.40sc03923 | 9 |  |
| 9 | SL2.40sc03771 | 1 |  |
| 9 | SL2.40sc06916 | 2 |  |
| 9 | SL2.40sc04008 | 4 |  |
| 9 | SL2.40sc04950 | 3 | yes |
| 9 | SL2.40sc04785 | 5 |  |
| 9 | SL2.40sc04777 | 6 |  |
| 9 | SL2.40sc05269 | 7 |  |
| 9 | SL2.40sc03852 | 8 |  |
| 9 | SL2.40sc04828 | 9 |  |
| 9 | SL2.40sc06214 | 10 |  |
| 10 | SL2.40sc05925 | 1 |  |
| 10 | SL2.40sc03798 | 2 |  |
| 10 | SL2.40sc04872 | 4 |  |
| 10 | SL2.40sc05632 | 3 |  |
| 10 | SL2.40sc04534 | 5 |  |
| 10 | SL2.40sc04199 | 6 |  |
| 11 | SL2.40sc03748 | 1 |  |
| 11 | SL2.40sc06763 | 4 |  |
| 11 | SL2.40sc04054 | 2 |  |
| 11 | SL2.40sc03752 | 3 | yes |
| 11 | SL2.40sc06137 | 5 |  |
| 11 | SL2.40sc03876 | 6 |  |
| 12 | SL2.40sc04607 | 1 |  |
| 12 | SL2.40sc06147 | 2 |  |
| 12 | SL2.40sc04039 | 3 | yes |
| 12 | SL2.40sc04878 | 4 |  |
| 12 | SL2.40sc04266 | 5 |  |
| 12 | SL2.40sc04757 | 6 |  |
| 12 | SL2.40sc04057 | 7 |  |
| 12 | SL2.40sc04915 | 8 |  |
| 12 | SL2.40sc05611 | 9 |  |
| 12 | SL2.40sc05380 | 10 |  |

Table 2. Gap sizes between scaffolds determined by FISH.

|  |  |  |  |
| --- | --- | --- | --- |
| Chr. | Gap start | Gap end | Size (Mb) |
| 1 | SL2.40sc04133 | SL2.40sc04191 | 2.28 |
| 1 | SL2.40sc04191 | SL2.40sc03666 | 2.13 |
| 1 | SL2.40sc03666 | SL2.40sc03594 | 0.57 |
| 1 | SL2.40sc03594 | SL2.40sc05010 | 2.12 |
| 1 | SL2.40sc05010 | SL2.40sc05941 | 0.51 |
| 1 | SL2.40sc05941 | SL2.40sc06917 | 0.25 |
| 1 | SL2.40sc06917 | SL2.40sc06903 | 0.17 |
| 1 | SL2.40sc06903 | SL2.40sc04323 | 0.21 |
| 2 | SL2.40sc04732 | SL2.40sc04208 | 0.34 |
| 2 | SL2.40sc04208 | SL2.40sc05776 | 0.10 |
| 2 | SL2.40sc05776 | SL2.40sc06593 | 0.10 |
| 2 | SL2.40sc06593 | SL2.40sc04142 | 3.05 |
| 2 | SL2.40sc04142 | SL2.40sc03766 | 0.49 |
| 2 | SL2.40sc03766 | SL2.40sc03665 | 1.34 |
| 3 | SL2.40sc04439 | SL2.40sc04696 | 0.09 |
| 3 | SL2.40sc04696 | SL2.40sc05330 | 0.10 |
| 3 | SL2.40sc05330 | SL2.40sc04126 | 0.32 |
| 3 | SL2.40sc04126 | SL2.40sc04616 | 2.58 |
| 3 | SL2.40sc04616 | SL2.40sc06725 | 0.16 |
| 3 | SL2.40sc06725 | SL2.40sc04704 | 0.74 |
| 3 | SL2.40sc04704 | SL2.40sc04822 | 1.09 |
| 3 | SL2.40sc04822 | SL2.40sc03806 | -0.37 |
| 3 | SL2.40sc03806 | SL2.40sc03796 | 0.10 |
| 3 | SL2.40sc03796 | SL2.40sc06911 | 0.07 |
| 3 | SL2.40sc06911 | SL2.40sc03701 | 0.08 |
| 4 | SL2.40sc03604 | SL2.40sc05339 | 0.30 |
| 4 | SL2.40sc05339 | SL2.40sc03683 | 0.47 |
| 4 | SL2.40sc03683 | SL2.40sc06101 | 0.08 |
| 4 | SL2.40sc06101 | SL2.40sc04680 | 0.03 |
| 4 | SL2.40sc04680 | SL2.40sc04135 | 1.53 |
| 5 | SL2.40sc03726 | SL2.40sc06155 | 0.85 |
| 5 | SL2.40sc06155 | SL2.40sc05339 | 0.75 |
| 6 | SL2.40sc04474 | SL2.40sc06140 | 2.34 |
| 6 | SL2.40sc06140 | SL2.40sc05383 | 0.61 |
| 6 | SL2.40sc05383 | SL2.40sc04279 | 0.34 |
| 6 | SL2.40sc04279 | SL2.40sc05188 | 1 |
| 6 | SL2.40sc05188 | SL2.40sc05732 | 1 |
| 6 | SL2.40sc05732 | SL2.40sc05054 | 1 |
| 6 | SL2.40sc05054 | SL2.40sc03622 | 1 |
| 7 | SL2.40sc03731 | SL2.40sc05397 | 0.39 |
| 7 | SL2.40sc05397 | SL2.40sc03685 | 2.29 |
| 7 | SL2.40sc03685 | SL2.40sc04626 | 0.10 |
| 8 | SL2.40sc04813 | SL2.40sc03770 | 0.10 |
| 8 | SL2.40sc03770 | SL2.40sc04167 | 0.52 |
| 8 | SL2.40sc04167 | SL2.40sc03749 | 0.26 |
| 8 | SL2.40sc03749 | SL2.40sc04236 | 0.10 |
| 8 | SL2.40sc04236 | SL2.40sc03835 | 0.04 |
| 8 | SL2.40sc03835 | SL2.40sc04701 | 0.97 |
| 8 | SL2.40sc04701 | SL2.40sc04948 | 0.80 |
| 8 | SL2.40sc04948 | SL2.40sc03923 | 0.04 |
| 9 | SL2.40sc03771 | SL2.40sc04008 | 1.46 |
| 9 | SL2.40sc04008 | SL2.40sc04950 | 0.72 |
| 9 | SL2.40sc04950 | SL2.40sc04785 | 1.05 |
| 9 | SL2.40sc04785 | SL2.40sc04777 | 1.25 |
| 9 | SL2.40sc04777 | SL2.40sc05269 | 1 |
| 9 | SL2.40sc05269 | SL2.40sc03852 | 1 |
| 9 | SL2.40sc03852 | SL2.40sc04828 | 0.04 |
| 9 | SL2.40sc04828 | SL2.40sc06214 | 0.07 |
| 10 | SL2.40sc05925 | SL2.40sc03798 | 0.10 |
| 10 | SL2.40sc03798 | SL2.40sc04872 | 0.39 |
| 10 | SL2.40sc04872 | SL2.40sc05632 | 0.10 |
| 10 | SL2.40sc05632 | SL2.40sc04199 | 0.10 |
| 11 | SL2.40sc03748 | SL2.40sc06763 | 0.35 |
| 11 | SL2.40sc06763 | SL2.40sc04054 | 0.53 |
| 11 | SL2.40sc04054 | SL2.40sc03752 | 1.31 |
| 11 | SL2.40sc03752 | SL2.40sc06137 | 0.07 |
| 11 | SL2.40sc06137 | SL2.40sc03876 | 0.65 |
| 12 | SL2.40sc04607 | SL2.40sc06147 | 0.17 |
| 12 | SL2.40sc06147 | SL2.40sc04039 | 0.01 |
| 12 | SL2.40sc04039 | SL2.40sc04878 | 0.08 |
| 12 | SL2.40sc04878 | SL2.40sc04266 | 0.55 |
| 12 | SL2.40sc04266 | SL2.40sc04757 | 0.47 |
| 12 | SL2.40sc04057 | SL2.40sc04915 | 0.01 |
| 12 | SL2.40sc04915 | SL2.40sc05611 | 0.07 |
| 12 | SL2.40sc05611 | SL2.40sc05380 | 0.01 |

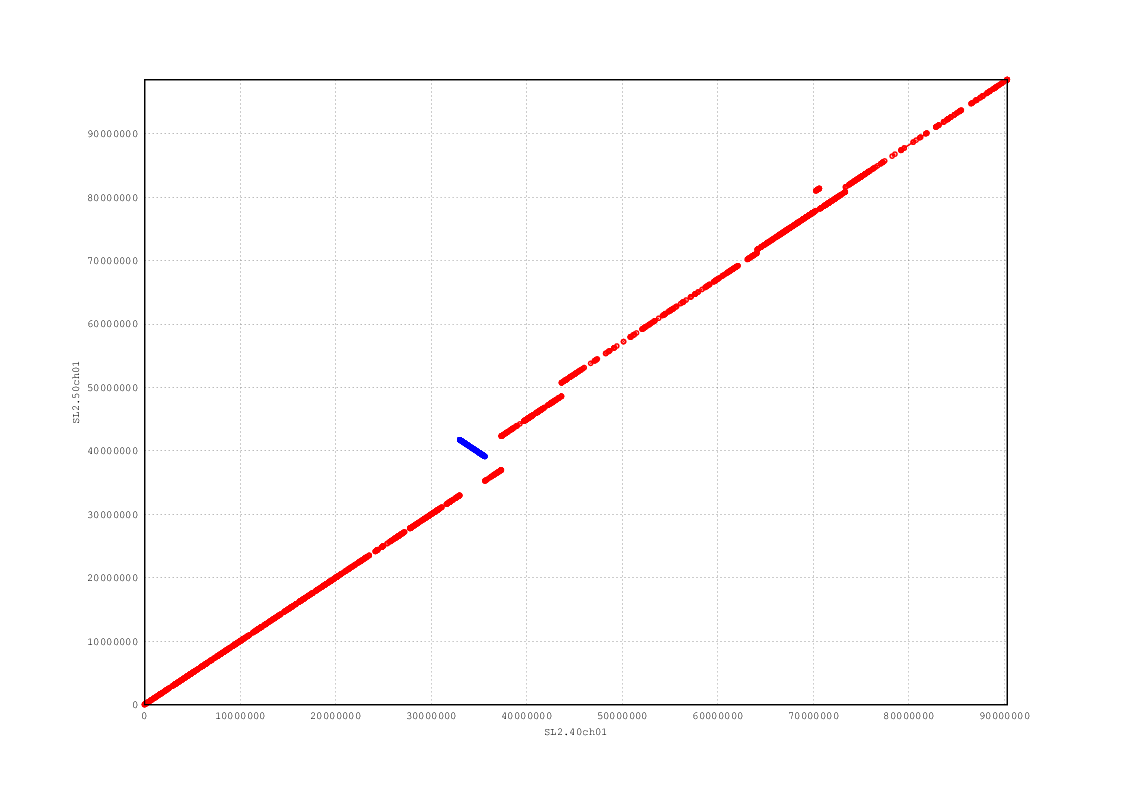
Figure 1. Chromosome 1: Dotplot of original vs FISH-modified chromosome sequence. 

Figure 2. Chromosome 2: Dotplot of original vs FISH-modified chromosome sequence.

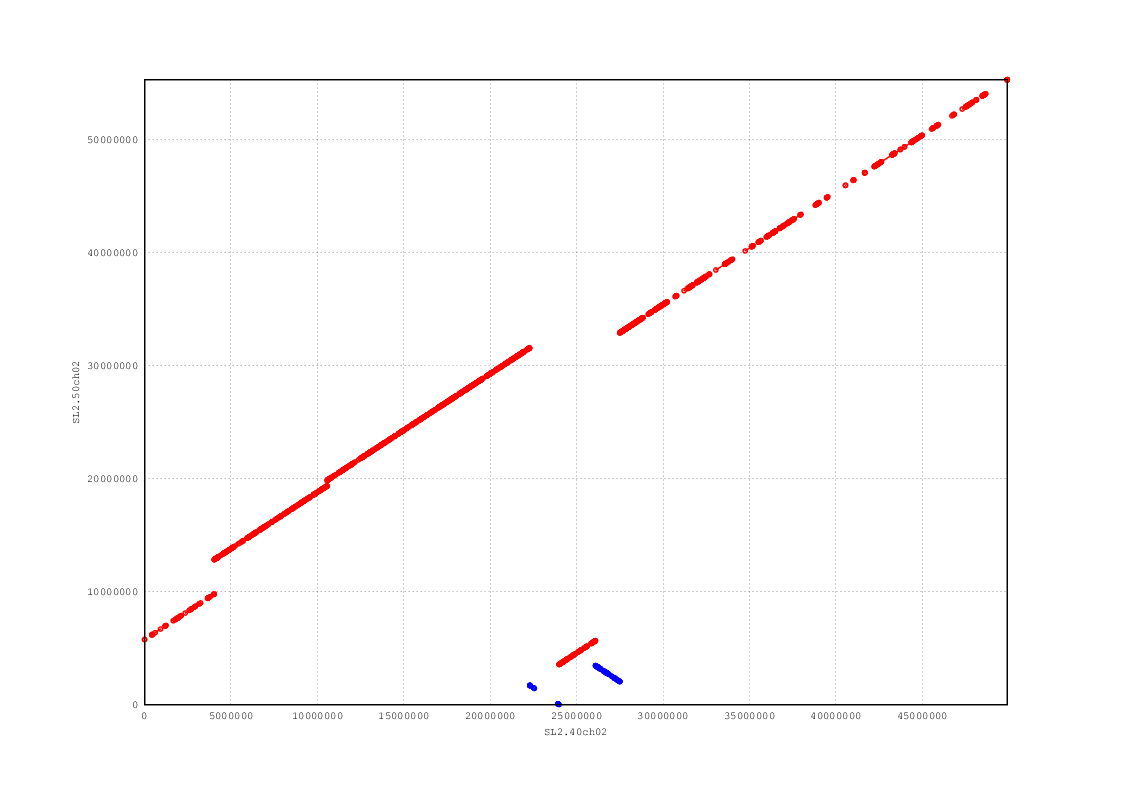


Figure 3. Chromosome 3: Dotplot of original vs FISH-modified chromosome sequence.

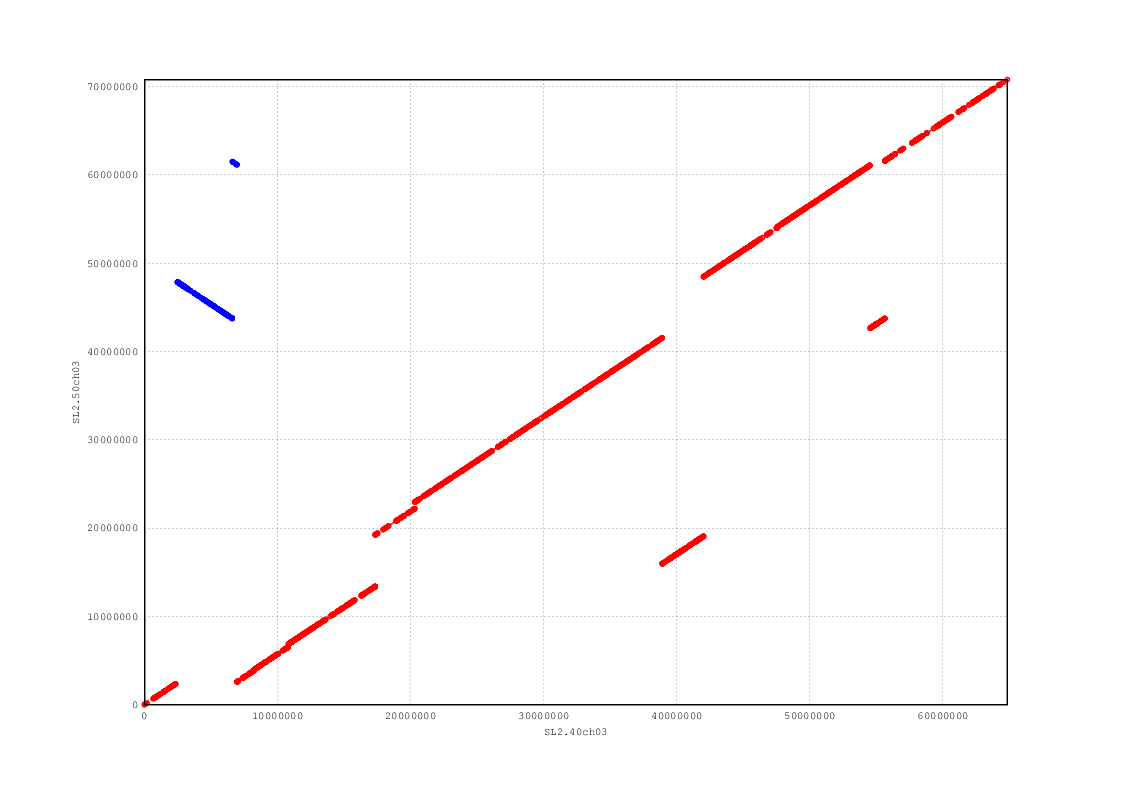


Figure 4. Chromosome 4: Dotplot of original vs FISH-modified chromosome sequence.

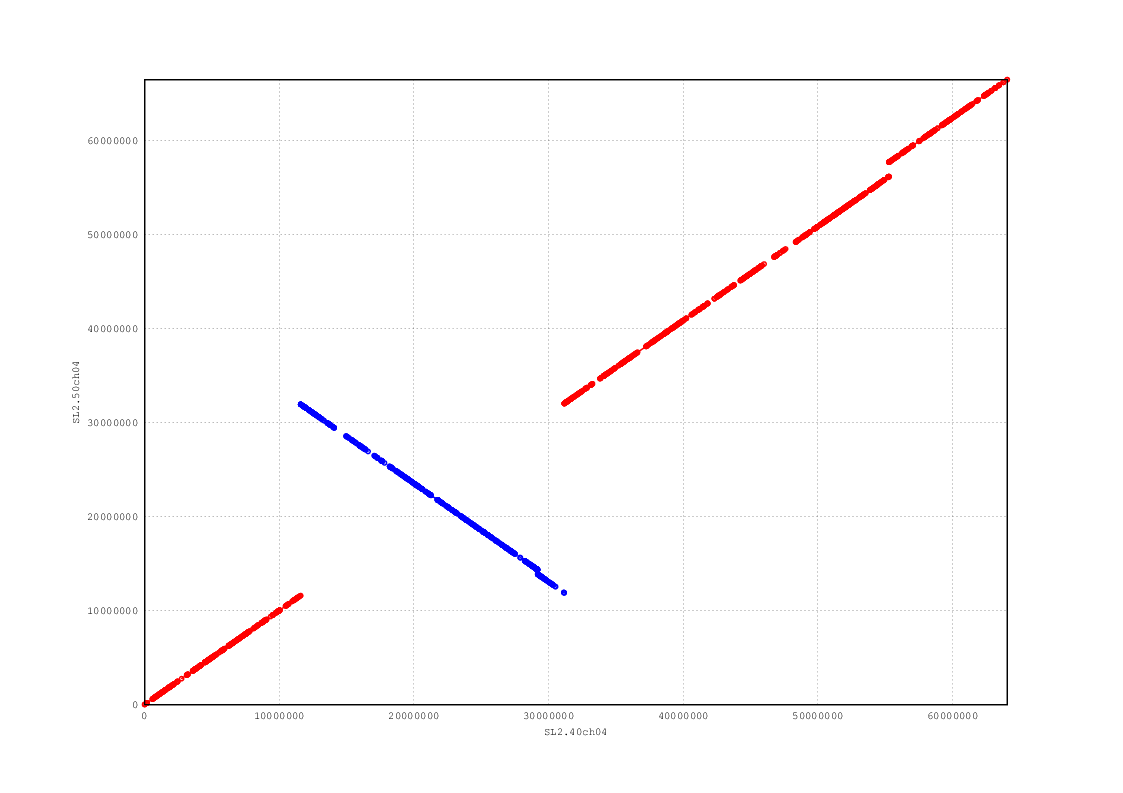


Figure 5. Chromosome 5: Dotplot of original vs FISH-modified chromosome sequence.

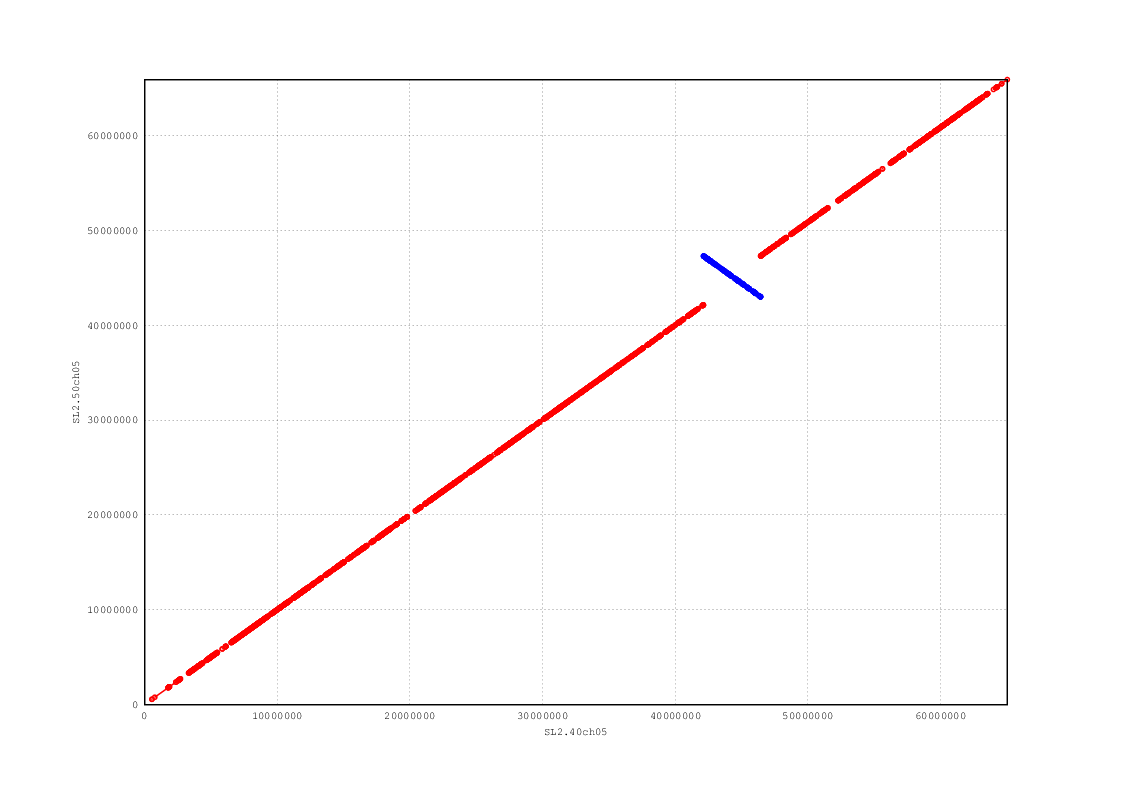


Figure 6. Chromosome 6: Dotplot of original vs FISH-modified chromosome sequence.

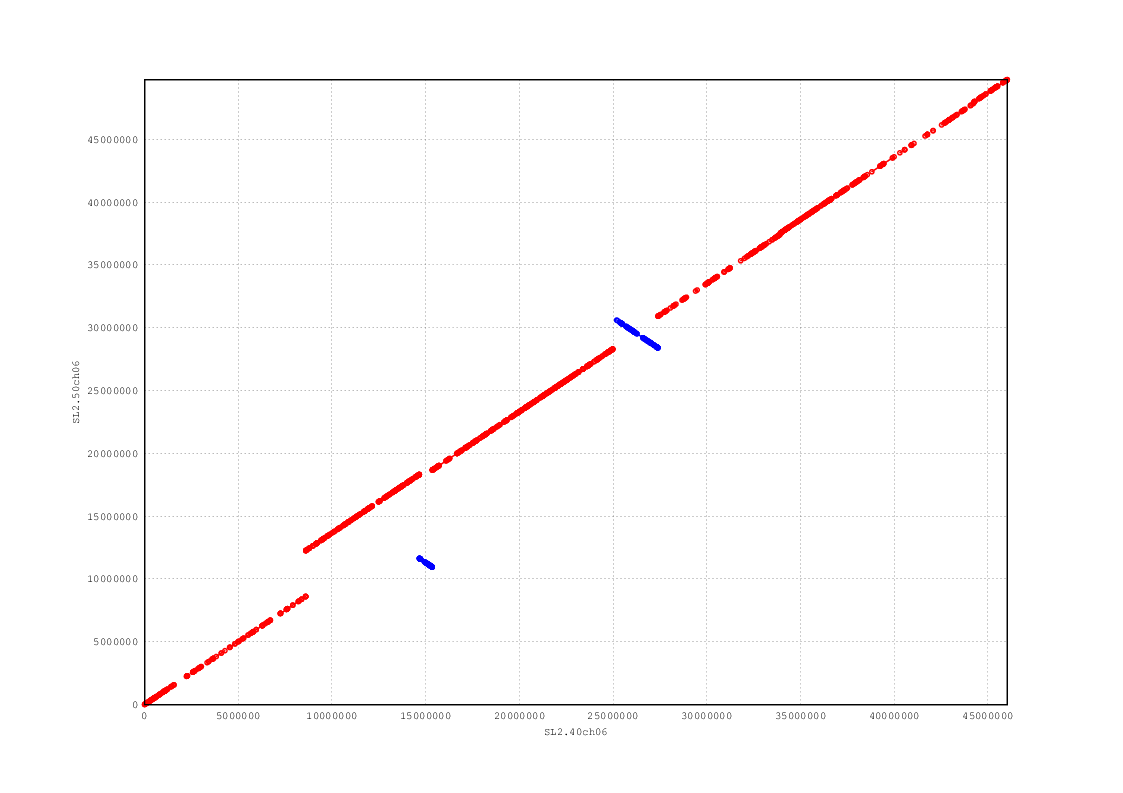


Figure 7. Chromosome 7: Dotplot of original vs FISH-modified chromosome sequence.

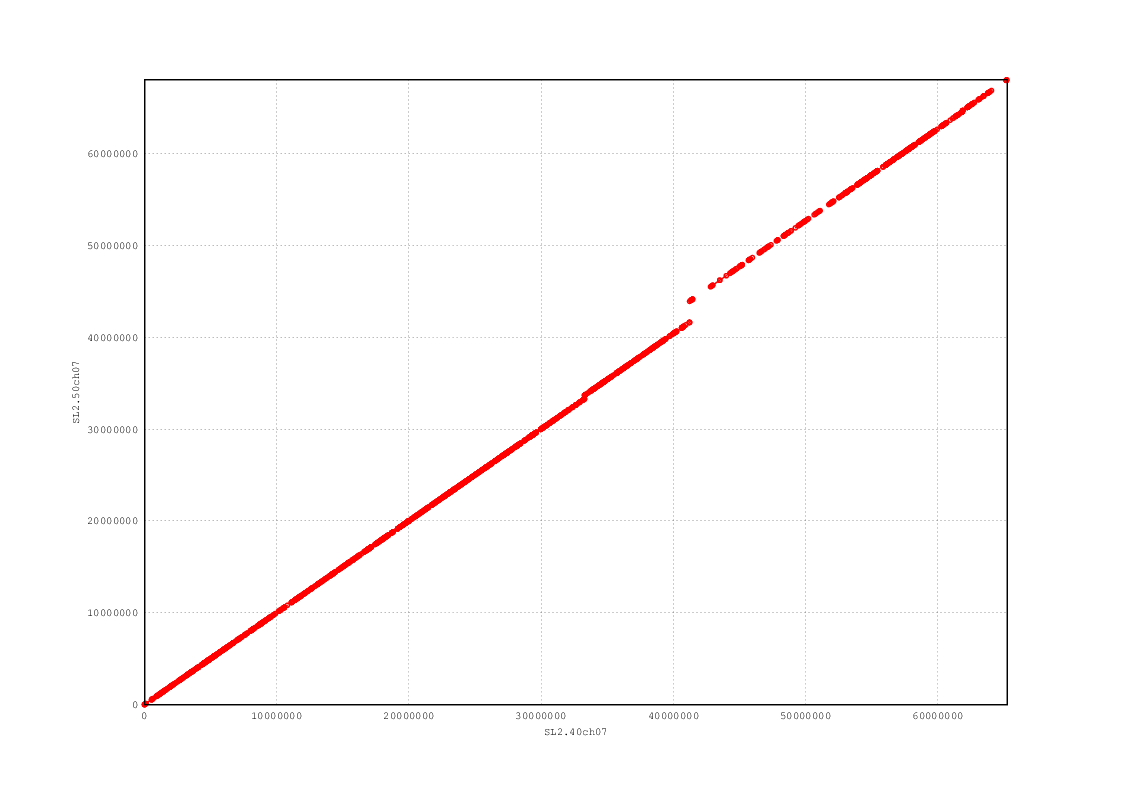


Figure 8. Chromosome 8: Dotplot of original vs FISH-modified chromosome sequence.

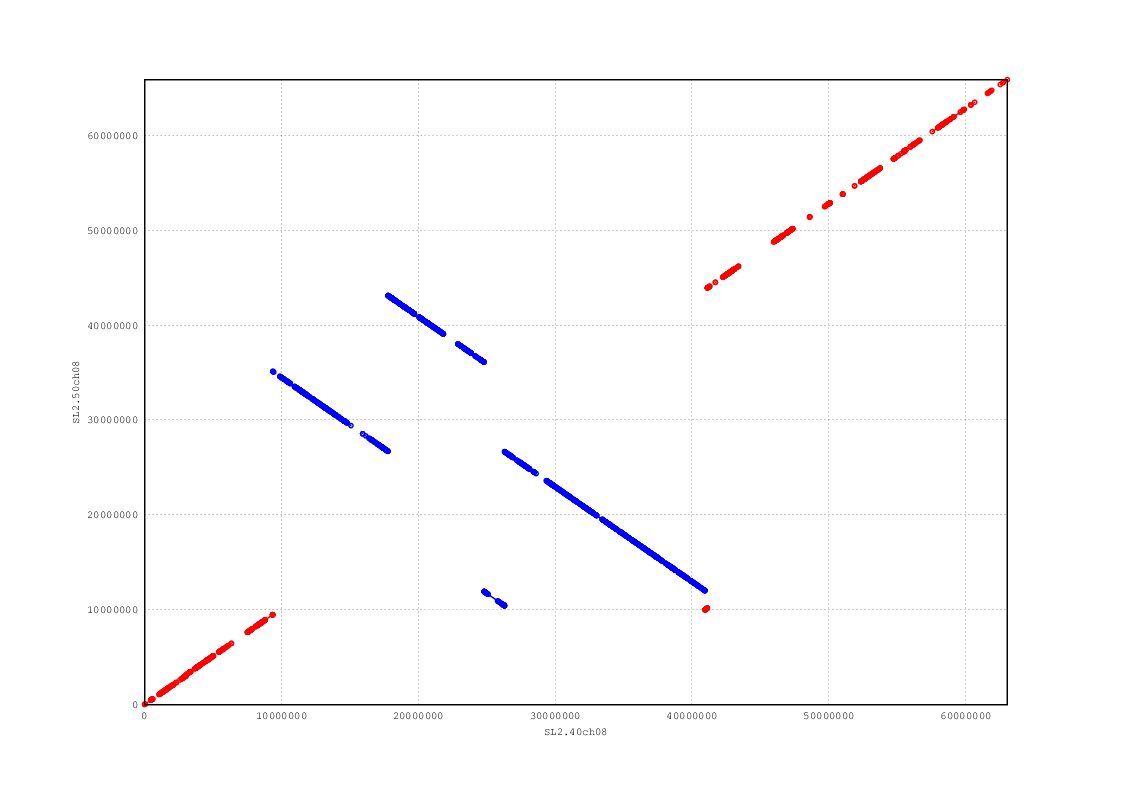


Figure 9. Chromosome 9: Dotplot of original vs FISH-modified chromosome sequence.

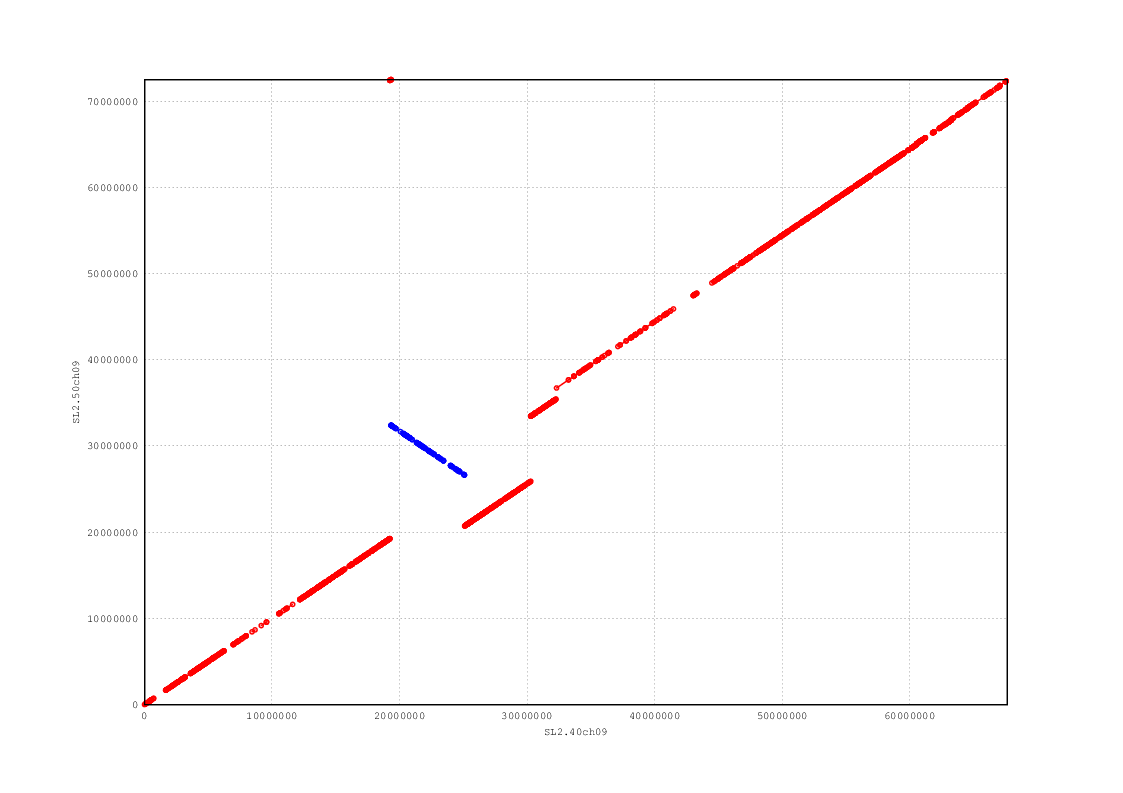


Figure 10. Chromosome 10: Dotplot of original vs FISH-modified chromosome sequence.

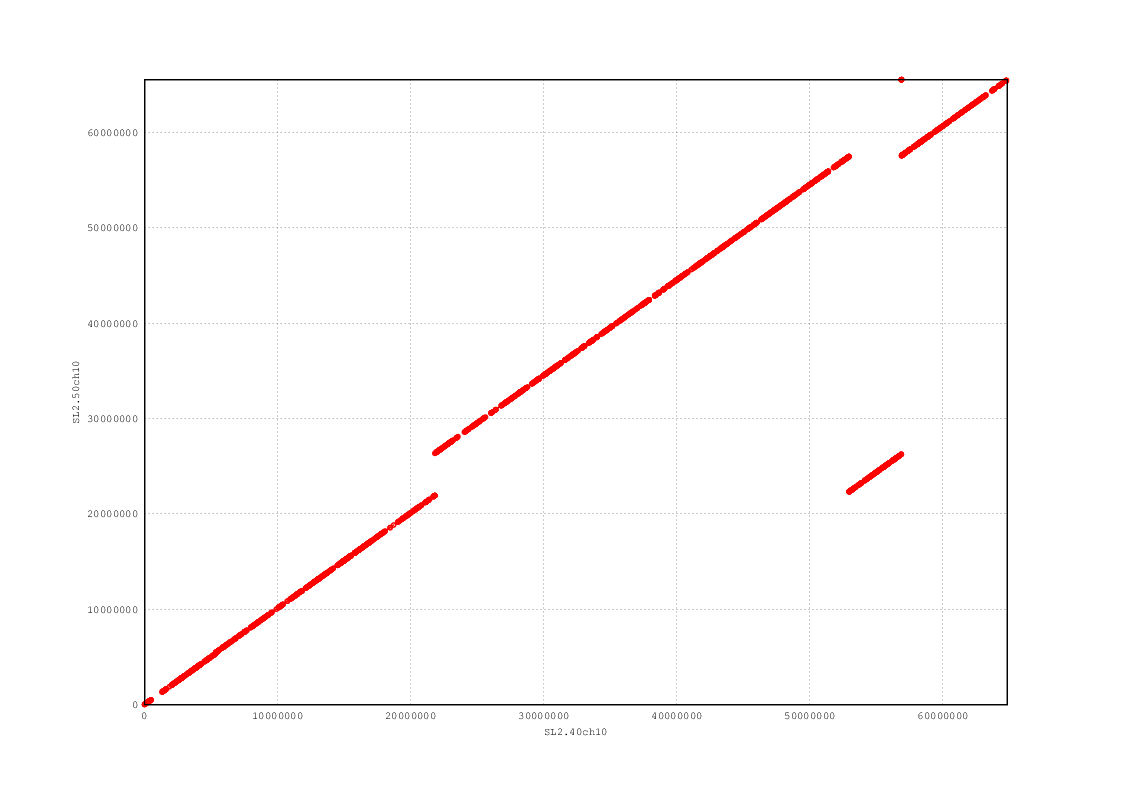


Figure 11. Chromosome 11: Dotplot of original vs FISH-modified chromosome sequence.

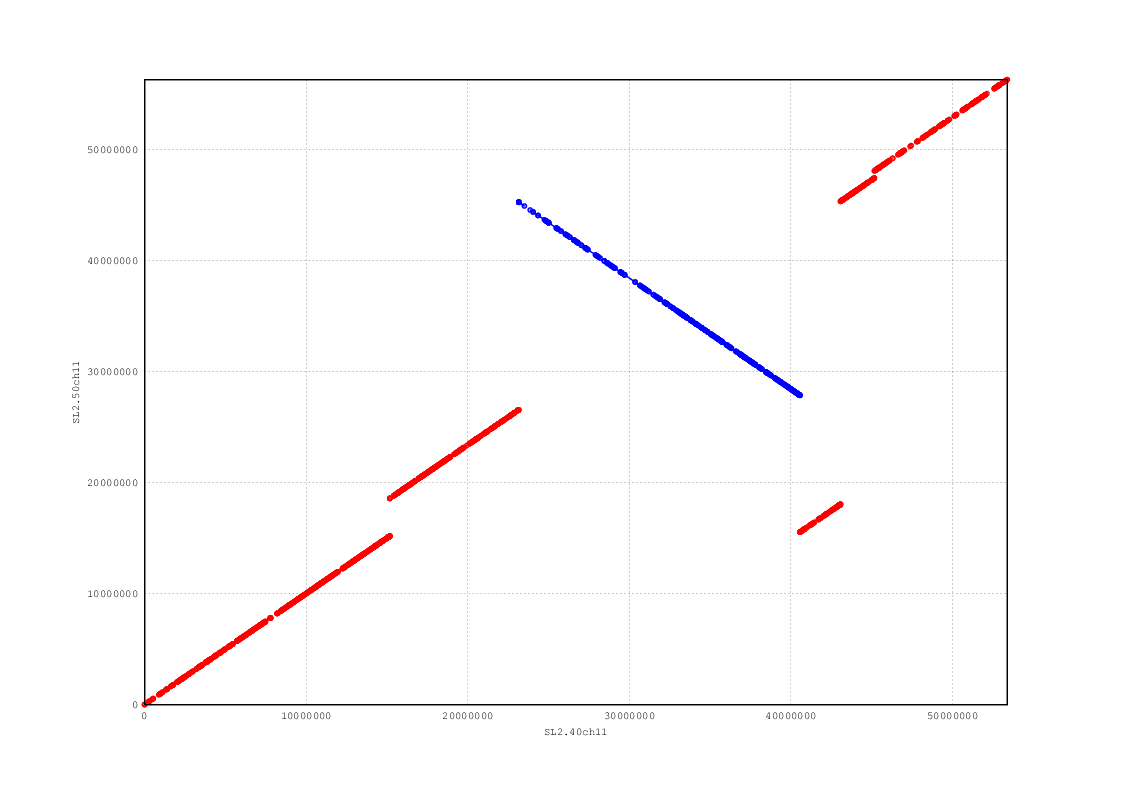
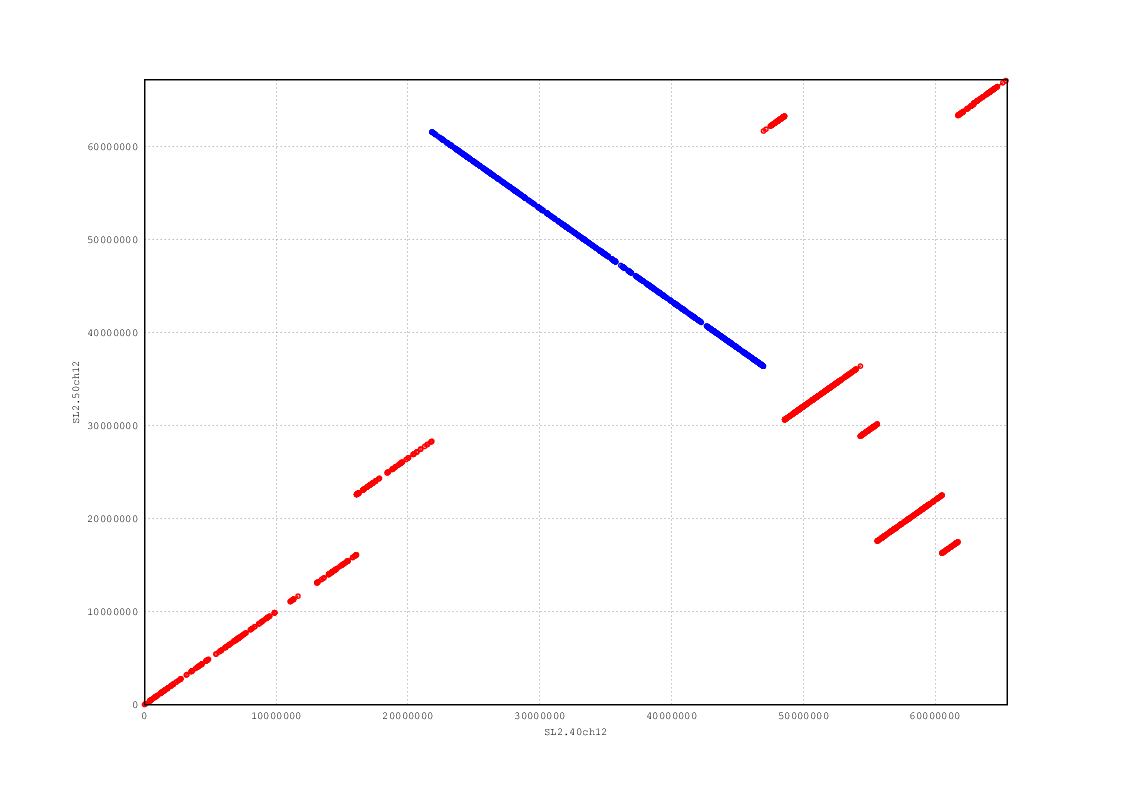


Figure 12. Chromosome 12: Dotplot of original vs FISH-modified chromosome sequence.



Commands used:

#create working directory

mkdir ~/genome\_update

cd ~/genome\_update

#install Bio-GenomeUpdate perl package from github repository

git clone <https://github.com/solgenomics/Bio-GenomeUpdate>

#add Bio-GenomeUpdate libraries to the PERL5LIB

export PERL5LIB="$HOME/genome\_update/Bio-GenomeUpdate/lib":$PERL5LIB

#add Bio-GenomeUpdate scripts to the path

export PATH="$HOME/genome\_update/Bio-GenomeUpdate/":$PATH

#get files specifying new scaffold order and gap sizes (based on FISH results) from:

#<https://github.com/solgenomics/sgn-home/tree/master/jeremy/tomato_genome_update/fish>

#files specifying order and orientation

wget <https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/order/chr01_order.txt>

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/order/chr02\_order.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/order/chr01_order.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/order/chr03\_order.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/order/chr01_order.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/order/chr04\_order.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/order/chr01_order.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/order/chr05\_order.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/order/chr01_order.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/order/chr06\_order.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/order/chr01_order.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/order/chr07\_order.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/order/chr01_order.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/order/chr08\_order.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/order/chr01_order.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/order/chr09\_order.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/order/chr01_order.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/order/chr10\_order.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/order/chr01_order.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/order/chr11\_order.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/order/chr01_order.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/order/chr12\_order.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/order/chr01_order.txt)

#files specifying gap sizes (chromosomes 6, 9, and 12 currently don’t have gap size information)

wget <https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/gaps/chr01_gaps.txt>

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/gaps/chr02\_gaps.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/gaps/chr01_gaps.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/gaps/chr03\_gaps.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/gaps/chr01_gaps.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/gaps/chr04\_gaps.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/gaps/chr01_gaps.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/gaps/chr05\_gaps.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/gaps/chr01_gaps.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/gaps/chr06\_gaps.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/gaps/chr01_gaps.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/gaps/chr07\_gaps.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/gaps/chr01_gaps.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/gaps/chr08\_gaps.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/gaps/chr01_gaps.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/gaps/chr09\_gaps.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/gaps/chr01_gaps.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/gaps/chr10\_gaps.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/gaps/chr01_gaps.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/gaps/chr11\_gaps.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/gaps/chr01_gaps.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/gaps/chr12\_gaps.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/gaps/chr01_gaps.txt)

#get TPF files for current tomato genome assembly (2.40)

wget <https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/tpf_from_2.40/tpf_4081_chr1.chr.txt>

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/tpf\_from\_2.40/tpf\_4081\_chr2.chr.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/tpf_from_2.40/tpf_4081_chr1.chr.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/tpf\_from\_2.40/tpf\_4081\_chr3.chr.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/tpf_from_2.40/tpf_4081_chr1.chr.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/tpf\_from\_2.40/tpf\_4081\_chr4.chr.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/tpf_from_2.40/tpf_4081_chr1.chr.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/tpf\_from\_2.40/tpf\_4081\_chr5.chr.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/tpf_from_2.40/tpf_4081_chr1.chr.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/tpf\_from\_2.40/tpf\_4081\_chr6.chr.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/tpf_from_2.40/tpf_4081_chr1.chr.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/tpf\_from\_2.40/tpf\_4081\_chr7.chr.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/tpf_from_2.40/tpf_4081_chr1.chr.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/tpf\_from\_2.40/tpf\_4081\_chr8.chr.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/tpf_from_2.40/tpf_4081_chr1.chr.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/tpf\_from\_2.40/tpf\_4081\_chr9.chr.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/tpf_from_2.40/tpf_4081_chr1.chr.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/tpf\_from\_2.40/tpf\_4081\_chr10.chr.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/tpf_from_2.40/tpf_4081_chr1.chr.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/tpf\_from\_2.40/tpf\_4081\_chr11.chr.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/tpf_from_2.40/tpf_4081_chr1.chr.txt)

wget [https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato\_genome\_update/fish/tpf\_from\_2.40/tpf\_4081\_chr12.chr.txt](https://raw.github.com/solgenomics/sgn-home/master/jeremy/tomato_genome_update/fish/tpf_from_2.40/tpf_4081_chr1.chr.txt)

#run Bio-GenomeUpdate scripts to change the order and orientation of scaffolds

reorder\_scaffolds.pl -t tpf\_4081\_chr1.chr.txt -l chr01\_order.txt > chr01\_fish.txt

reorder\_scaffolds.pl -t tpf\_4081\_chr2.chr.txt -l chr02\_order.txt > chr02\_fish.txt

reorder\_scaffolds.pl -t tpf\_4081\_chr3.chr.txt -l chr03\_order.txt > chr03\_fish.txt

reorder\_scaffolds.pl -t tpf\_4081\_chr4.chr.txt -l chr04\_order.txt > chr04\_fish.txt

reorder\_scaffolds.pl -t tpf\_4081\_chr5.chr.txt -l chr05\_order.txt > chr05\_fish.txt

reorder\_scaffolds.pl -t tpf\_4081\_chr6.chr.txt -l chr06\_order.txt > chr06\_fish.txt

reorder\_scaffolds.pl -t tpf\_4081\_chr7.chr.txt -l chr07\_order.txt > chr07\_fish.txt

reorder\_scaffolds.pl -t tpf\_4081\_chr8.chr.txt -l chr08\_order.txt > chr08\_fish.txt

reorder\_scaffolds.pl -t tpf\_4081\_chr9.chr.txt -l chr09\_order.txt > chr09\_fish.txt

reorder\_scaffolds.pl -t tpf\_4081\_chr10.chr.txt -l chr10\_order.txt > chr10\_fish.txt

reorder\_scaffolds.pl -t tpf\_4081\_chr11.chr.txt -l chr11\_order.txt > chr11\_fish.txt

reorder\_scaffolds.pl -t tpf\_4081\_chr12.chr.txt -l chr12\_order.txt > chr12\_fish.txt

#run Bio-GenomeUpdate scripts to change gap sizes

set\_scaffold\_gap\_sizes.pl -t chr01\_fish.txt -l chr01\_gaps.txt > chr01\_fish2\_gaps.txt

set\_scaffold\_gap\_sizes.pl -t chr02\_fish.txt -l chr02\_gaps.txt > chr02\_fish2\_gaps.txt

set\_scaffold\_gap\_sizes.pl -t chr03\_fish.txt -l chr03\_gaps.txt > chr03\_fish2\_gaps.txt

set\_scaffold\_gap\_sizes.pl -t chr04\_fish.txt -l chr04\_gaps.txt > chr04\_fish2\_gaps.txt

set\_scaffold\_gap\_sizes.pl -t chr05\_fish.txt -l chr05\_gaps.txt > chr05\_fish2\_gaps.txt

set\_scaffold\_gap\_sizes.pl -t chr06\_fish.txt -l chr06\_gaps.txt > chr06\_fish2\_gaps.txt

set\_scaffold\_gap\_sizes.pl -t chr07\_fish.txt -l chr07\_gaps.txt > chr07\_fish2\_gaps.txt

set\_scaffold\_gap\_sizes.pl -t chr08\_fish.txt -l chr08\_gaps.txt > chr08\_fish2\_gaps.txt

set\_scaffold\_gap\_sizes.pl -t chr09\_fish.txt -l chr09\_gaps.txt > chr09\_fish2\_gaps.txt

set\_scaffold\_gap\_sizes.pl -t chr10\_fish.txt -l chr10\_gaps.txt > chr10\_fish2\_gaps.txt

set\_scaffold\_gap\_sizes.pl -t chr11\_fish.txt -l chr11\_gaps.txt > chr11\_fish2\_gaps.txt

set\_scaffold\_gap\_sizes.pl -t chr12\_fish.txt -l chr12\_gaps.txt > chr12\_fish2\_gaps.txt

#upload resulting files (\*\_fish\_gaps.txt ) to The Genome Reference Consortium (GRC)

#<http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/>

#GRC Test TPF Submission

#<http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/CuratDat/tpf_solo.cgi>

#upload file, select all boxes including Chomosome FASTA and Scaffold FASTA and submit.

#download new chromosome assemblies from ftp-private.ncbi.nih.gov with FTP

ftp ftp-private.ncbi.nih.gov

#enter username

#enter password

#passive mode required

passive

cd TPFSOLO

mget \*fish2\_gaps.zip

quit

#unzip

find \*.zip -exec unzip {} \;

#get current tomato genome chromosome assemblies

wget ftp://[ftp.solgenomics.net/tomato\_genome/assembly/build\_2.40/S\_lycopersicum\_chromosomes.2.40.fa.gz](http://ftp.solgenomics.net/tomato_genome/assembly/build_2.40/S_lycopersicum_chromosomes.2.40.fa.gz)

gunzip S\_lycopersicum\_chromosomes.2.40.fa.gz

#get seq\_from\_fasta.pl from github to extract individual files from a multi-FASTA file

wget <https://raw.github.com/solgenomics/sgn-home/master/jeremy/scripts/seq/seq_from_fasta.pl>

chmod +x seq\_from\_fasta.pl

#extract each chromosome sequence in its own .fasta file.

./seq\_from\_fasta.pl -i S\_lycopersicum\_chromosomes.2.40.fa -n SL2.40ch00 > chr00.2.40.fa

./seq\_from\_fasta.pl -i S\_lycopersicum\_chromosomes.2.40.fa -n SL2.40ch01 > chr01.2.40.fa

./seq\_from\_fasta.pl -i S\_lycopersicum\_chromosomes.2.40.fa -n SL2.40ch02 > chr02.2.40.fa

./seq\_from\_fasta.pl -i S\_lycopersicum\_chromosomes.2.40.fa -n SL2.40ch03 > chr03.2.40.fa

./seq\_from\_fasta.pl -i S\_lycopersicum\_chromosomes.2.40.fa -n SL2.40ch04 > chr04.2.40.fa

./seq\_from\_fasta.pl -i S\_lycopersicum\_chromosomes.2.40.fa -n SL2.40ch05 > chr05.2.40.fa

./seq\_from\_fasta.pl -i S\_lycopersicum\_chromosomes.2.40.fa -n SL2.40ch06 > chr06.2.40.fa

./seq\_from\_fasta.pl -i S\_lycopersicum\_chromosomes.2.40.fa -n SL2.40ch07 > chr07.2.40.fa

./seq\_from\_fasta.pl -i S\_lycopersicum\_chromosomes.2.40.fa -n SL2.40ch08 > chr08.2.40.fa

./seq\_from\_fasta.pl -i S\_lycopersicum\_chromosomes.2.40.fa -n SL2.40ch09 > chr09.2.40.fa

./seq\_from\_fasta.pl -i S\_lycopersicum\_chromosomes.2.40.fa -n SL2.40ch10 > chr10.2.40.fa

./seq\_from\_fasta.pl -i S\_lycopersicum\_chromosomes.2.40.fa -n SL2.40ch11 > chr11.2.40.fa

./seq\_from\_fasta.pl -i S\_lycopersicum\_chromosomes.2.40.fa -n SL2.40ch12 > chr12.2.40.fa

#align current assemblies with nucmer (requires MUMmer, <http://mummer.sourceforge.net/>)

nucmer -p chr01\_fish\_gaps -o --noextend chr01.2.40.fa chr01\_fish2\_gaps.chr.fasta

nucmer -p chr02\_fish\_gaps -o --noextend chr02.2.40.fa chr02\_fish2\_gaps.chr.fasta

nucmer -p chr03\_fish\_gaps -o --noextend chr03.2.40.fa chr03\_fish2\_gaps.chr.fasta

nucmer -p chr04\_fish\_gaps -o --noextend chr04.2.40.fa chr04\_fish2\_gaps.chr.fasta

nucmer -p chr05\_fish\_gaps -o --noextend chr05.2.40.fa chr05\_fish2\_gaps.chr.fasta

nucmer -p chr06\_fish\_gaps -o --noextend chr06.2.40.fa chr06\_fish2\_gaps.chr.fasta

nucmer -p chr07\_fish\_gaps -o --noextend chr07.2.40.fa chr07\_fish2\_gaps.chr.fasta

nucmer -p chr08\_fish\_gaps -o --noextend chr08.2.40.fa chr08\_fish2\_gaps.chr.fasta

nucmer -p chr09\_fish\_gaps -o --noextend chr09.2.40.fa chr09\_fish2\_gaps.chr.fasta

nucmer -p chr10\_fish\_gaps -o --noextend chr10.2.40.fa chr10\_fish2\_gaps.chr.fasta

nucmer -p chr11\_fish\_gaps -o --noextend chr11.2.40.fa chr11\_fish2\_gaps.chr.fasta

nucmer -p chr12\_fish\_gaps -o --noextend chr12.2.40.fa chr12\_fish2\_gaps.chr.fasta

#filter the alignments for minimum length 1000 and minimum uniqueness of 100%

delta-filter -l 1000 -u 100 chr01\_fish\_gaps.delta > chr01\_fish\_gaps.filtered.delta

delta-filter -l 1000 -u 100 chr02\_fish\_gaps.delta > chr02\_fish\_gaps.filtered.delta

delta-filter -l 1000 -u 100 chr03\_fish\_gaps.delta > chr03\_fish\_gaps.filtered.delta

delta-filter -l 1000 -u 100 chr04\_fish\_gaps.delta > chr04\_fish\_gaps.filtered.delta

delta-filter -l 1000 -u 100 chr05\_fish\_gaps.delta > chr05\_fish\_gaps.filtered.delta

delta-filter -l 1000 -u 100 chr06\_fish\_gaps.delta > chr06\_fish\_gaps.filtered.delta

delta-filter -l 1000 -u 100 chr07\_fish\_gaps.delta > chr07\_fish\_gaps.filtered.delta

delta-filter -l 1000 -u 100 chr08\_fish\_gaps.delta > chr08\_fish\_gaps.filtered.delta

delta-filter -l 1000 -u 100 chr09\_fish\_gaps.delta > chr09\_fish\_gaps.filtered.delta

delta-filter -l 1000 -u 100 chr10\_fish\_gaps.delta > chr10\_fish\_gaps.filtered.delta

delta-filter -l 1000 -u 100 chr11\_fish\_gaps.delta > chr11\_fish\_gaps.filtered.delta

delta-filter -l 1000 -u 100 chr12\_fish\_gaps.delta > chr12\_fish\_gaps.filtered.delta

#produce DNA dotplots from the alignments requires MUMmer and GNUPlot)

mummerplot --postscript -p chr01\_fish\_gaps chr01\_fish\_gaps.filtered.delta

mummerplot --postscript -p chr02\_fish\_gaps chr02\_fish\_gaps.filtered.delta

mummerplot --postscript -p chr03\_fish\_gaps chr03\_fish\_gaps.filtered.delta

mummerplot --postscript -p chr04\_fish\_gaps chr04\_fish\_gaps.filtered.delta

mummerplot --postscript -p chr05\_fish\_gaps chr05\_fish\_gaps.filtered.delta

mummerplot --postscript -p chr06\_fish\_gaps chr06\_fish\_gaps.filtered.delta

mummerplot --postscript -p chr07\_fish\_gaps chr07\_fish\_gaps.filtered.delta

mummerplot --postscript -p chr08\_fish\_gaps chr08\_fish\_gaps.filtered.delta

mummerplot --postscript -p chr09\_fish\_gaps chr09\_fish\_gaps.filtered.delta

mummerplot --postscript -p chr10\_fish\_gaps chr10\_fish\_gaps.filtered.delta

mummerplot --postscript -p chr11\_fish\_gaps chr11\_fish\_gaps.filtered.delta

mummerplot --postscript -p chr12\_fish\_gaps chr12\_fish\_gaps.filtered.delta

#generate updated chromosome assembly FASTA file

cat chr00.2.40.fa > S\_lycopersicum\_chromosomes\_fish\_v\_1.0.fa

cat chr01\_fish2\_gaps.chr.fasta >> S\_lycopersicum\_chromosomes\_fish\_v\_1.0.fa

cat chr02\_fish2\_gaps.chr.fasta >> S\_lycopersicum\_chromosomes\_fish\_v\_1.0.fa

cat chr03\_fish2\_gaps.chr.fasta >> S\_lycopersicum\_chromosomes\_fish\_v\_1.0.fa

cat chr04\_fish2\_gaps.chr.fasta >> S\_lycopersicum\_chromosomes\_fish\_v\_1.0.fa

cat chr05\_fish2\_gaps.chr.fasta >> S\_lycopersicum\_chromosomes\_fish\_v\_1.0.fa

cat chr06\_fish2\_gaps.chr.fasta >> S\_lycopersicum\_chromosomes\_fish\_v\_1.0.fa

cat chr07\_fish2\_gaps.chr.fasta >> S\_lycopersicum\_chromosomes\_fish\_v\_1.0.fa

cat chr08\_fish2\_gaps.chr.fasta >> S\_lycopersicum\_chromosomes\_fish\_v\_1.0.fa

cat chr09\_fish2\_gaps.chr.fasta >> S\_lycopersicum\_chromosomes\_fish\_v\_1.0.fa

cat chr10\_fish2\_gaps.chr.fasta >> S\_lycopersicum\_chromosomes\_fish\_v\_1.0.fa

cat chr11\_fish2\_gaps.chr.fasta >> S\_lycopersicum\_chromosomes\_fish\_v\_1.0.fa

cat chr12\_fish2\_gaps.chr.fasta >> S\_lycopersicum\_chromosomes\_fish\_v\_1.0.fa