

# Supplementary Information

## Diploid genomic architecture of *Nitzschia hildebrandi*, an elite biomass production diatom

Aaron Oliver, Sheila Podell, Agnieszka Pinowska, Jesse C. Traller, Sarah R. Smith, Ryan McClure, Alex Beliaev, Pavlo Bohutsky, Eric Hill, Ariel Rabines, Hong Zheng, Lisa Zeigler Allen, Igor V. Grigoriev, Alan Kuo, David Hazlebeck, Eric E. Allen

### Contents

|                                |    |
|--------------------------------|----|
| Supplementary Table S1 .....   | 2  |
| Supplementary Table S2 .....   | 3  |
| Supplementary Table S3 .....   | 4  |
| Supplementary Table S4 .....   | 5  |
| Supplementary Figure S1 .....  | 8  |
| Supplementary Figure S2 .....  | 22 |
| Supplementary Figure S3 .....  | 23 |
| Supplementary Figure S4 .....  | 24 |
| Supplementary Figures S5 ..... | 25 |
| Supplementary Figures S6 ..... | 26 |

**Supp. Table S1.** Assembly properties of syntenic paired contigs in the GAI-293 diploid assembly.

| pair number | Ref ctg len | Mate ctg(s)                                                                 | ref ctg length | ref ctg assembly coverage | mate ctg(s) assembly coverage        | average nucleotide identity | MUMmer aln len ref ctg | MUMmer pct aln ref ctg | MUMmer aln len mate ctg(s) | MUMmer pct aln mate ctg(s) |
|-------------|-------------|-----------------------------------------------------------------------------|----------------|---------------------------|--------------------------------------|-----------------------------|------------------------|------------------------|----------------------------|----------------------------|
| pair 1      | tig00000111 | tig00000107                                                                 | 3,679,959      | 60.3                      | 60.2                                 | 94%                         | 2,790,740              | 76%                    | 2,808,708                  | 78%                        |
| pair 2      | tig00044023 | tig00000310                                                                 | 2,278,761      | 60.1                      | 56.8                                 | 93%                         | 1,481,604              | 65%                    | 1,474,187                  | 77%                        |
| pair 3      | tig00000176 | tig00000241                                                                 | 2,972,054      | 62.5                      | 56.3                                 | 93%                         | 1,998,077              | 67%                    | 2,026,657                  | 78%                        |
| pair 4      | tig00000204 | tig00000247                                                                 | 3,108,069      | 62.7                      | 58.5                                 | 93%                         | 2,184,523              | 70%                    | 2,154,154                  | 80%                        |
| pair 5      | tig00007565 | tig00044005                                                                 | 2,584,008      | 60.2                      | 61.9                                 | 93%                         | 1,910,698              | 74%                    | 2,071,516                  | 66%                        |
| pair 6      | tig00044003 | tig00000121                                                                 | 4,242,913      | 61.1                      | 61.4                                 | 94%                         | 3,030,608              | 71%                    | 3,007,305                  | 85%                        |
| pair 7      | tig00043997 | tig00000071                                                                 | 5,376,619      | 67.3                      | 60.7                                 | 94%                         | 3,656,905              | 68%                    | 3,734,145                  | 86%                        |
| pair 8      | tig00007555 | tig00007557                                                                 | 292,286        | 55.7                      | 47.8                                 | 100%                        | 246,033                | 84%                    | 239,662                    | 99%                        |
| pair 9      | tig00044033 | tig00000034,<br>tig00000040                                                 | 4,610,928      | 64.4                      | 60.8,<br>49.7                        | 94%                         | 3,559,477              | 77%                    | 3,573,060                  | 87%                        |
| pair 10     | tig00007564 | tig00000069,<br>tig00000068                                                 | 3,170,983      | 77.8                      | 63.2,<br>46.6                        | 94%                         | 1,906,200              | 60%                    | 1,897,237                  | 85%                        |
| pair 11     | tig00007560 | tig00000376,<br>tig00000375,<br>tig00007561,<br>tig00007567                 | 900,008        | 62.5                      | 47.8,<br>47.9, 13.9,<br>51.2         | 100%                        | 838,710                | 93%                    | 904,674                    | 98%                        |
| pair 12     | tig00007562 | tig00000158,<br>tig00044009,<br>tig00000168,<br>tig00000162,<br>tig00007563 | 1,282,887      | 104.3                     | 33.0,<br>6.0,<br>10.5, 38.7,<br>39.7 | 100%                        | 474,708                | 37%                    | 459,813                    | 99%                        |
| pair 13     | tig00044028 | tig00000004,<br>tig00000074,<br>tig00000008                                 | 6,574,884      | 60.8                      | 62.1, 61.1,<br>54.7                  | 94%                         | 3,009,567              | 47%                    | 3,263,325                  | 50%                        |
| pair 14     | tig00000120 | tig00000301,<br>tig00000012                                                 | 3,900,614      | 61.1                      | 59.4,<br>61.3                        | 93%                         | 1,425,792              | 37%                    | 1,405,244                  | 69%                        |

**Supp. Table S2.** BUSCO completeness statistics for sequenced diatom genomes. Analyses were based on a Stramenopile model, using amino acid sequences for all predicted proteins from each genome.

| Species                                        | type    | num<br>seqs | pct<br>complete | single-<br>copy | dup-<br>licated | frag-<br>mented | pct<br>missing | Groups<br>searched |
|------------------------------------------------|---------|-------------|-----------------|-----------------|-----------------|-----------------|----------------|--------------------|
| <i>Nitzschia hildebrandi</i> GAI-293           | protein | 38,601      | 100%            | 8%              | 92%             | 0%              | 0%             | 100                |
| <i>Fragilariopsis cylindrus</i> CCMP1102       | protein | 18,121      | 95%             | 95%             | 0%              | 4%              | 1%             | 100                |
| <i>Pseudo-nitzschia multistriata</i> str. B856 | protein | 12,055      | 86%             | 85%             | 1%              | 1%              | 13%            | 100                |
| <i>Phaeodactylum tricornutum</i> v2.0          | protein | 10,198      | 94%             | 93%             | 1%              | 3%              | 3%             | 100                |
| <i>Seminavis robusta</i> D6                    | protein | 37,718      | 99%             | 92%             | 7%              | 0%              | 1%             | 100                |
| <i>Fistulifera solaris</i> JPCC DA0580         | protein | 20,429      | 97%             | 14%             | 83%             | 1%              | 2%             | 100                |
| <i>Thalassiosira pseudonana</i> v3.0           | protein | 10,608      | 94%             | 93%             | 1%              | 3%              | 3%             | 100                |

**Supp. Table S3.** Culture conditions used to obtain *Nitzschia hildebrandi* str. GAI-293 samples for transcriptome analysis.

| Condition Name     | Condition Details                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Control            | Global Algae standard growth media, 5mL of <i>N. hildebrandi</i> (at ~ O.D. of 1.0) grown in a 50 mL conical tube with closed lid at 24 C with shaking at 130 rpm (tubes were horizontal in the shaker to maximize agitation) in incubator in MCDL, exposed to ~20 of light for 4 hours before RNA samples collected.                                                                                                                                          |
| Low Nitrate        | Same as control but with 0.05 g/L (0.589 mM) of NaNO <sub>3</sub>                                                                                                                                                                                                                                                                                                                                                                                              |
| No Nitrate         | Same as control but lacking NaNO <sub>3</sub>                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Dark               | Same as control but tubes were wrapped in tinfoil                                                                                                                                                                                                                                                                                                                                                                                                              |
| Bioreactor         | Continuous cultivations in a custom-built photobioreactor <sup>*</sup> equipped with constant LED-based di-chromatic irradiation system. O.D. was held at 0.08 and the light intensity was 10:10E m <sup>-2</sup> sec <sup>-1</sup> at the 630:680 nm linear incident irradiance. PBR was operated under turbidostat mode at pH 9.2, temperature of 25°C, 5.5-L culture volume under 250 rpm agitation, sparging at a 1 L min <sup>-1</sup> rate with nitrogen |
| PBR Simulated Pond | Continuous cultivations photobioreactor simulating pond growth                                                                                                                                                                                                                                                                                                                                                                                                 |

<sup>\*</sup>Bohutskyi P, McClure R, Charania M, Hill E, Chrisler W, Lindemann S, Alex Beliaev (2019) Metabolic effects of Vitamin B<sub>12</sub> on *Cyanobacterium stanieri* gene expression, physiology and stress resistance. **Algal Research**. 42:101580.

**Suppl Table S4.** Telomere sequences on paired contig ends. Contig pair number refers to table S1. Sequences at contig 5' ends are shown as reverse complements, to facilitate comparison with 3' ends.

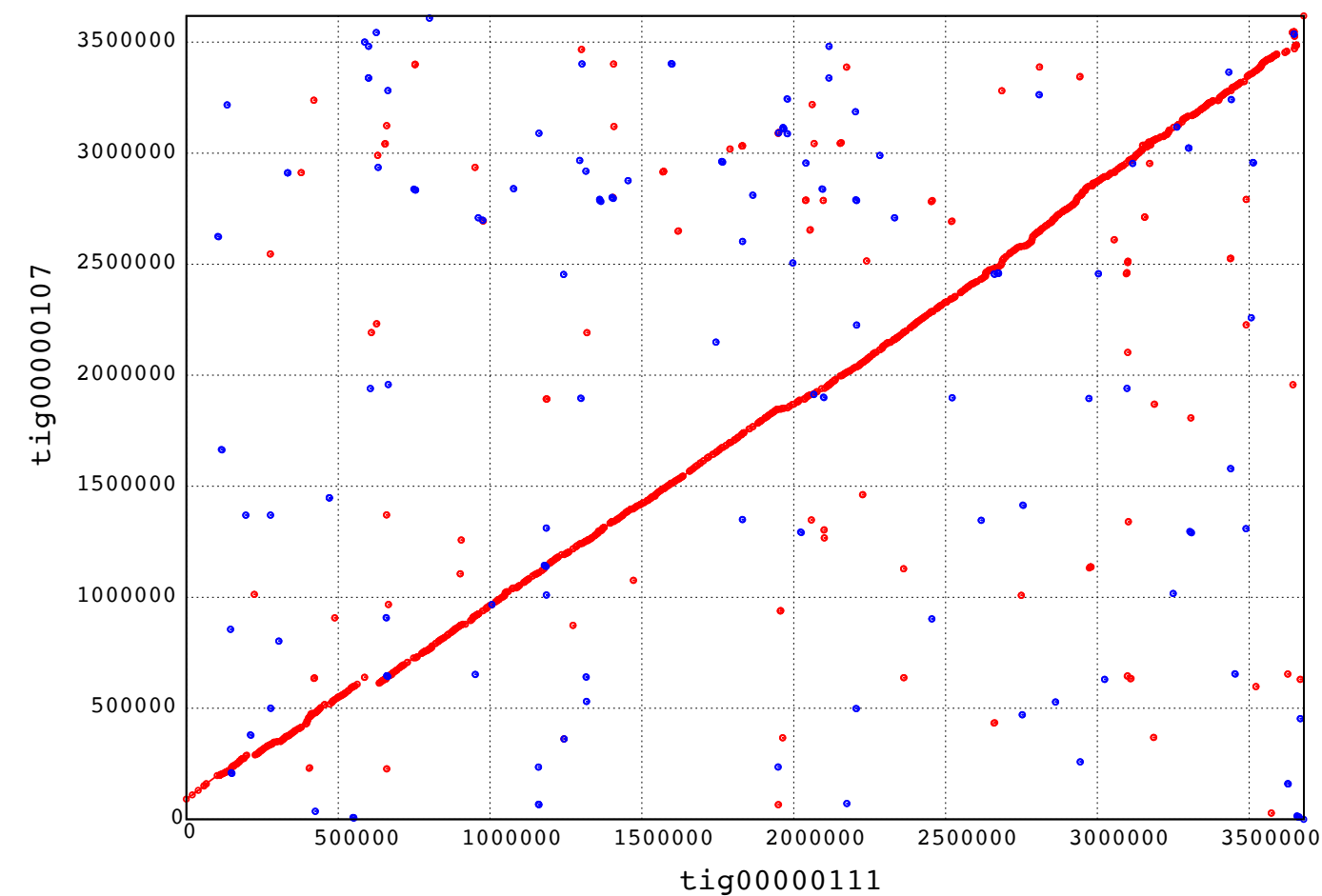
[illegible]



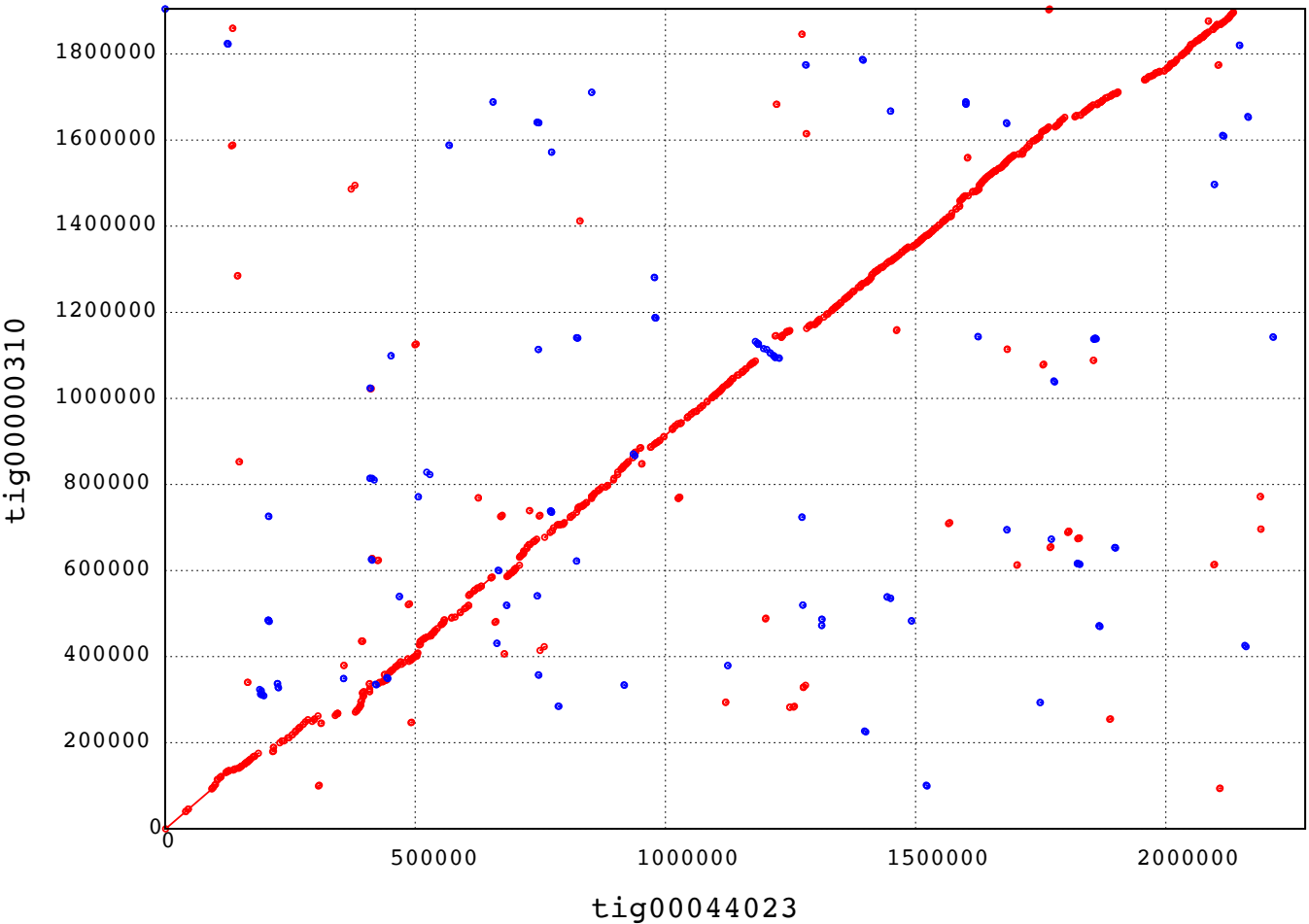


| telomere seq id | contig pair num. | telomer position | telo- mere len | telo- mere read cover | assembled telomere sequence                                                                                                                                                                                                                                                                                                                                                                                   |
|-----------------|------------------|------------------|----------------|-----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| tig00000120_3   | 14               | 3 prime          | 399            | 23.9                  | AGGGTTAGGGTTGGGGTTAGGGTTAGGGTTGGGTTAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTAGGGTTGGG GTTAGGGTTGGGGTTGGGGTTGGGGTTAGGGTTGGGGTTAGGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGG GTTGGGGTTGGGGTTGGGGTTGGGGTTAGGGTTGGGGTTTGGGTTAGGGTTAGGGTTGGGGTTAGGGT TGGGGTTTGGGTTAGGGTTAGGGTTGGGTTGGGGTTAGGGTTGGGGTTAGGGTTGGGTTAGGGTTGGGGTTAGG GTTGGGGTTGGGTTAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTGGGGTTAGGGTTGGGGTTAGGGTTGGGGT TGGGGTTAGGGTTGGGTTAGGGTTGGGGTT |
| tig00000301_3   | 14               | 3 prime          | 379            | 3.2                   | TAGGGTTAGGTTAGGGTTAGGGTTAGGGTTGGGGTTGGGGTTGAGGGTTAGGGTTAGGGTTGGGGTTGGGGTTA GGGTTGGGGTTAGGGTTGGGGTTAGGGTTAGGGTTGGGGTTGGGGTTAGGGTTAGGGTTGGGGTTAGGGTTGGG TTAGGGTTAGGTTAGGGTTGGGGTTTAGGGTTGGGGTTGGGGTTAGGGTTGGGGTTAGGGTTGGGGTTTAGGGTT AGGGTTAGGGTTGGGTTAGGGTTGGGGTTGGGGTTAGGGTTGGGGTTAGGGTTAGGGTTGGGGTTAGGGTTGGGT TAGGGTTGGTGTAGGGTTGGGGTTAGGGTTGGGGTTAGGGTTGGGTTGGGGTTGGGGTTAGGGTTGGGGTTA GGGTTGGG               |
| tig00000012_3   | 14               | 3 prime          | 338            | 3.2                   | TTAGGGTTAGGGTTGGGGTTAGGGTTGGGGTTAGGGTTAGGGTTAGGTTGGGGTTAGGGTTGGGGTTAGGGTTGG GGTAGGGTTAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTGGGGTTAGGGTTGGGTTAGGGTTGGGGTTAGGGTT AGGGTTGGGGTTTAGGGTTAGGGTTGGGGTTAGGGTTGGGGTTGGGGTTAGGGTTGGGGTTAGGGTTGGGGTTA GGGTTGGGGTTTAGGGTTGGGGTTAGGGTTGGGGTTAGGGTTGGGGTTAGGGTTGGGGTTAGGGTTGGGGTTAGG GTTGGGTTAGGGTTGGGTTGGGGTTAGGGTTGGGGTTGGG                                                        |

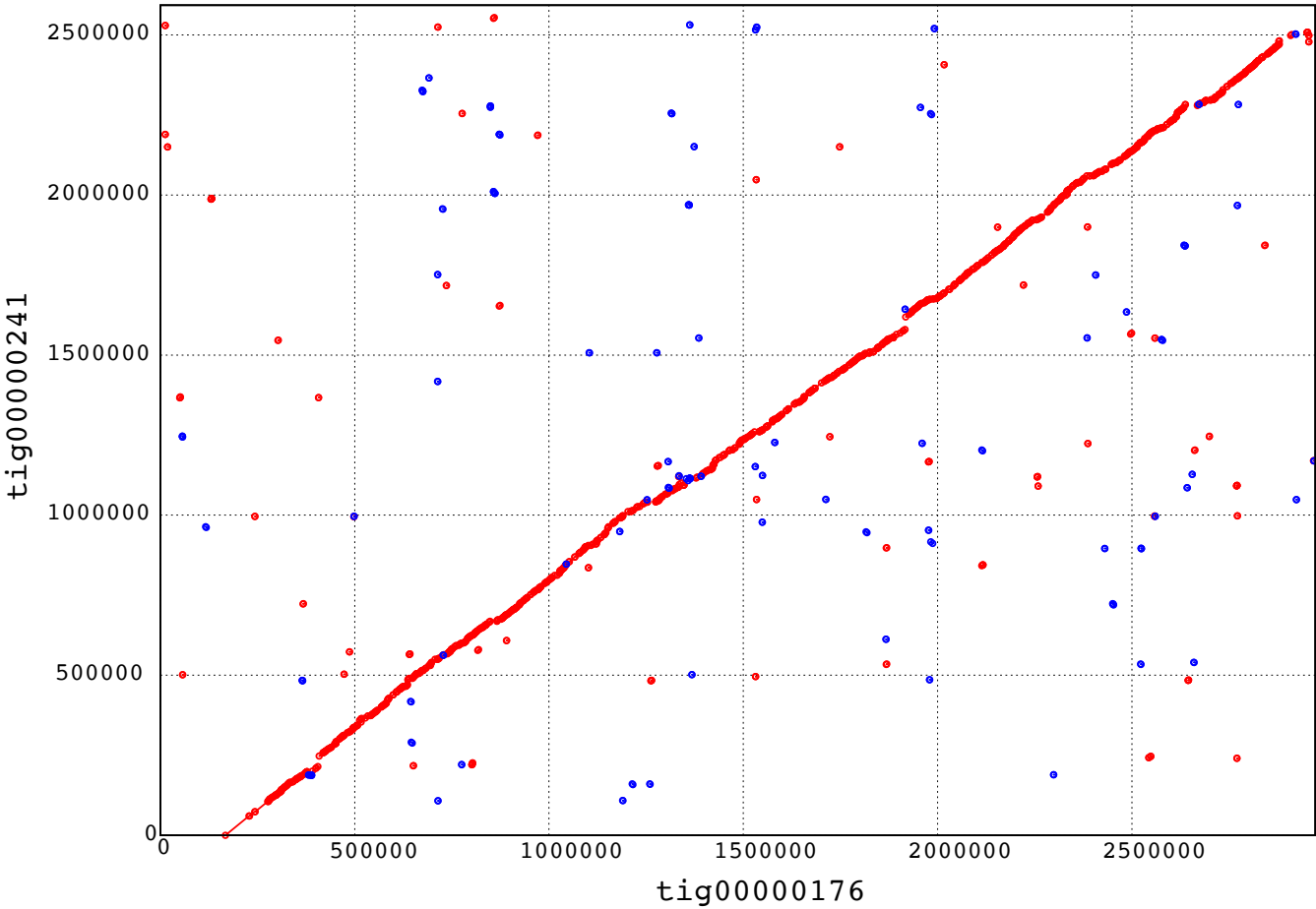
Supplementary Figure 1a. MUMmer dot plot of contig pair #1



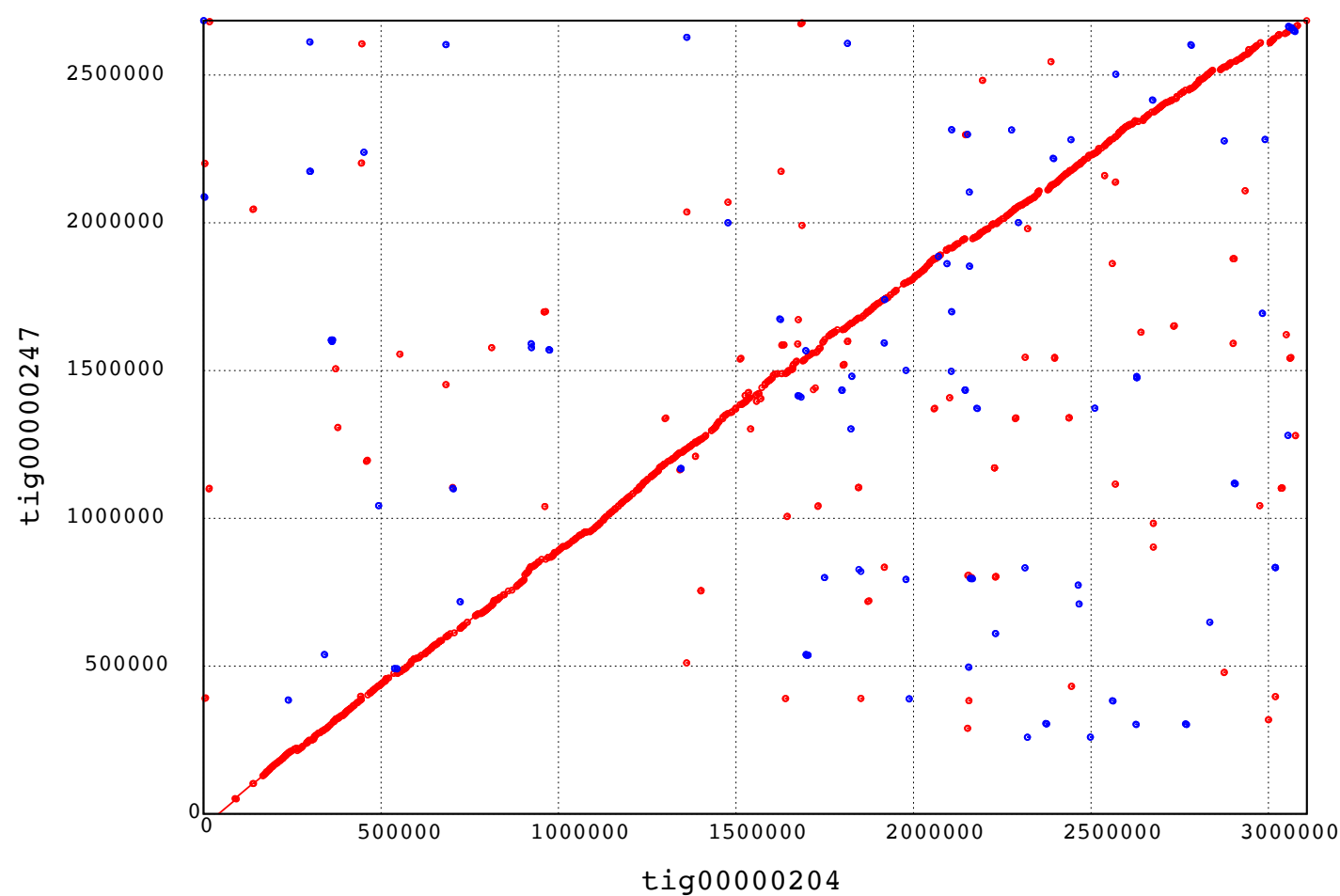
**Supplementary Figure 1b.** MUMmer dot plot of contig pair #2



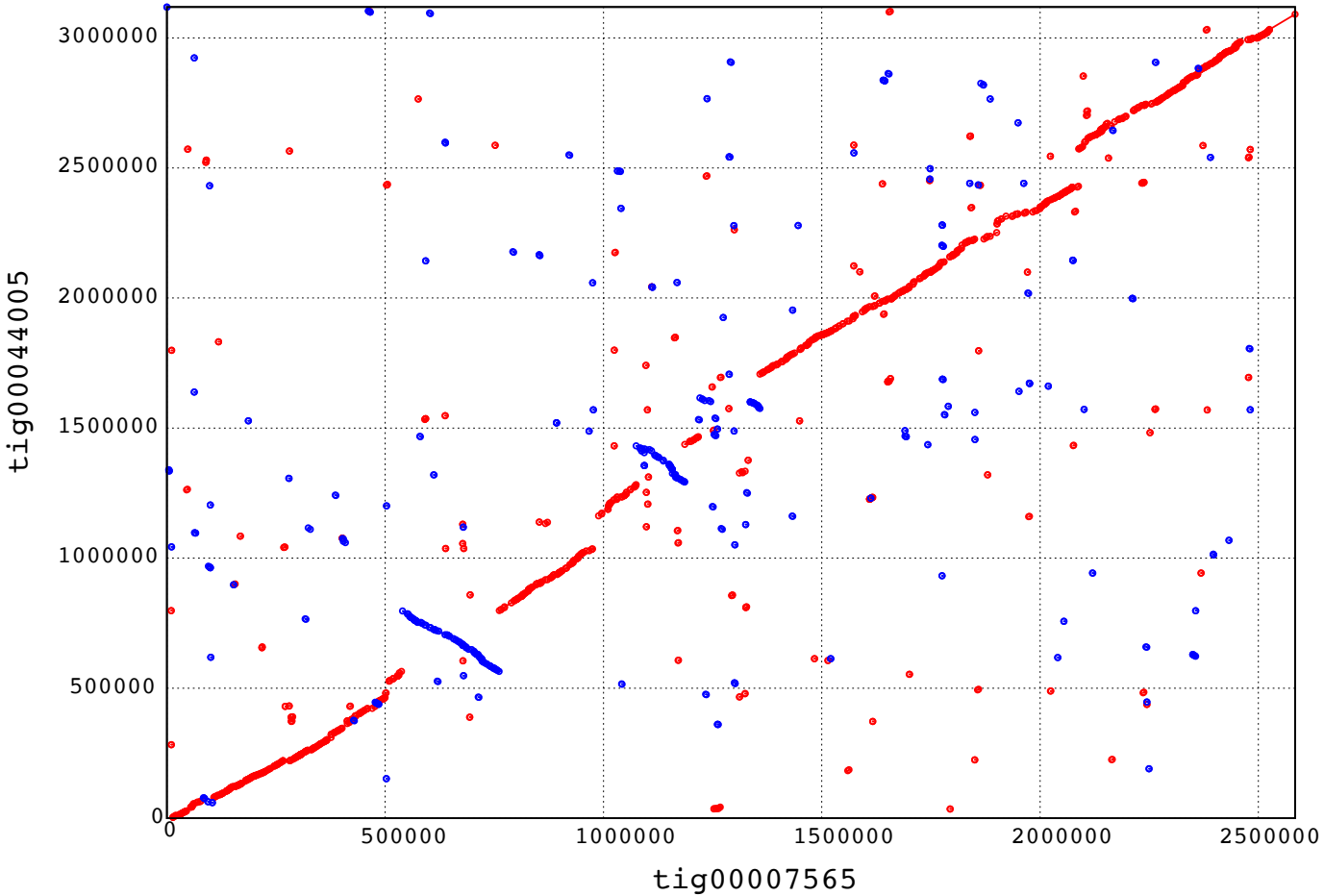
**Supplementary Figure 1c.** MUMmer dot plot of contig pair #3



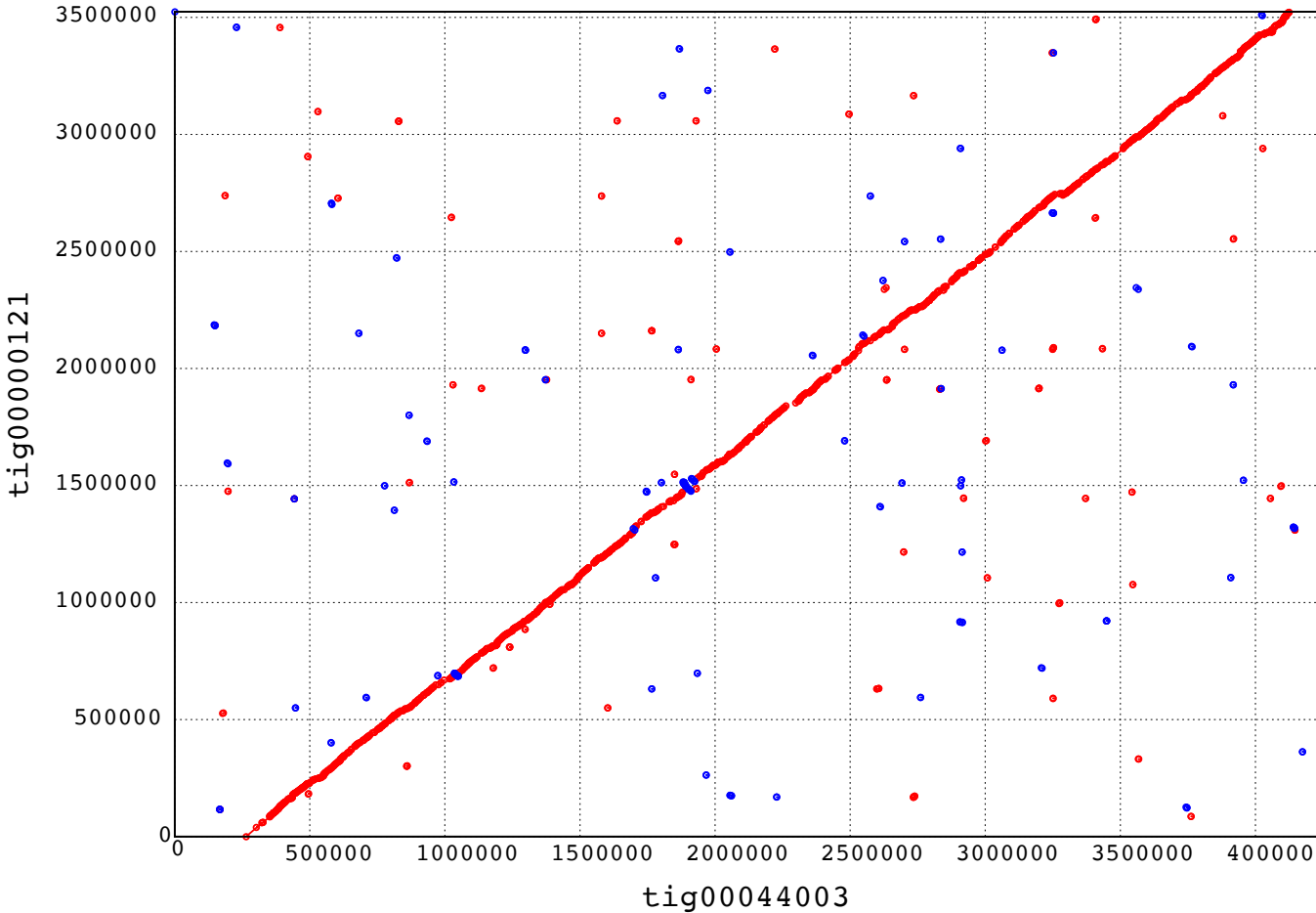
Supplementary Figure 1d. MUMmer dot plot of contig pair #4



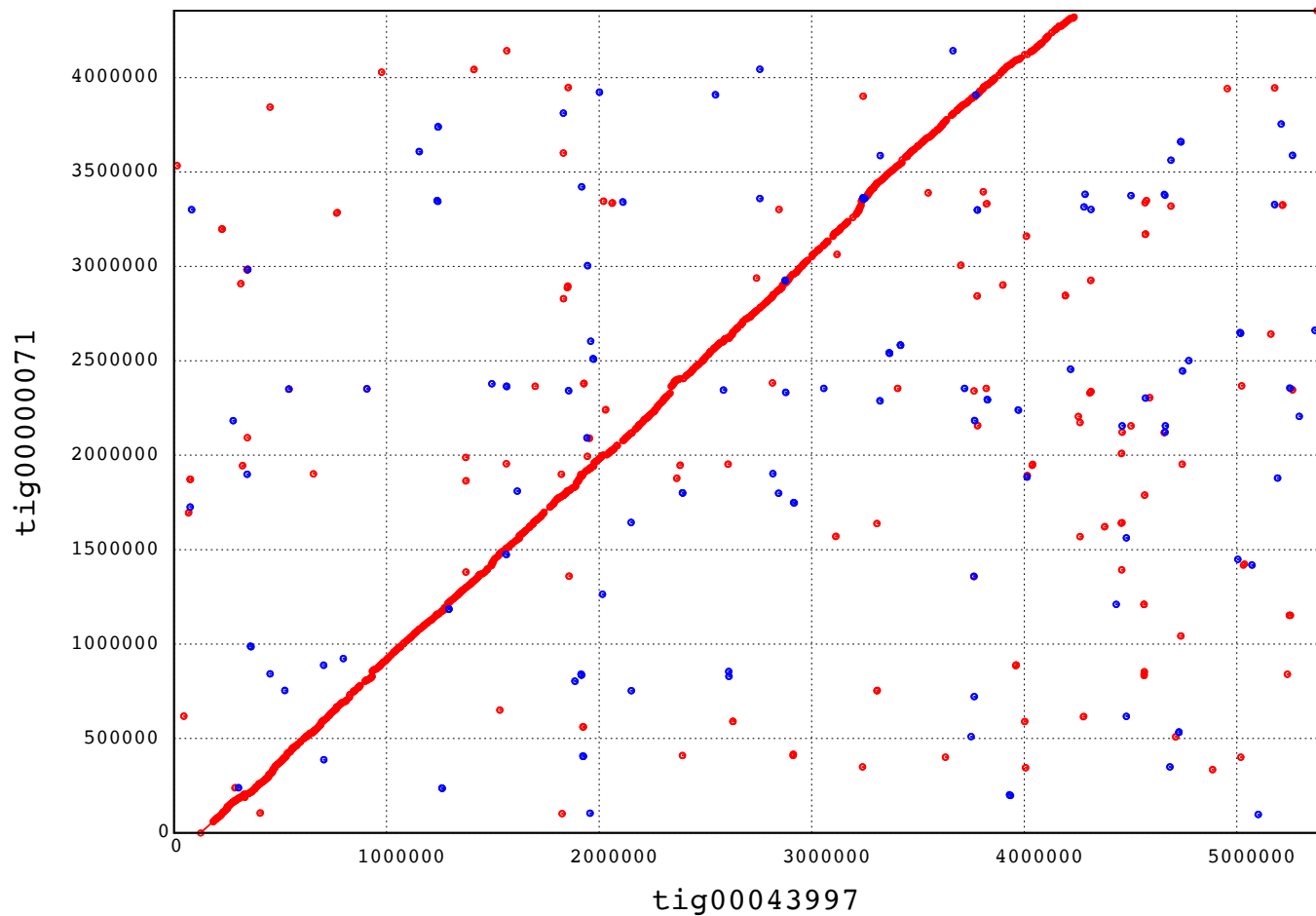
Supplementary Figure 1e. MUMmer dot plot of contig pair #5



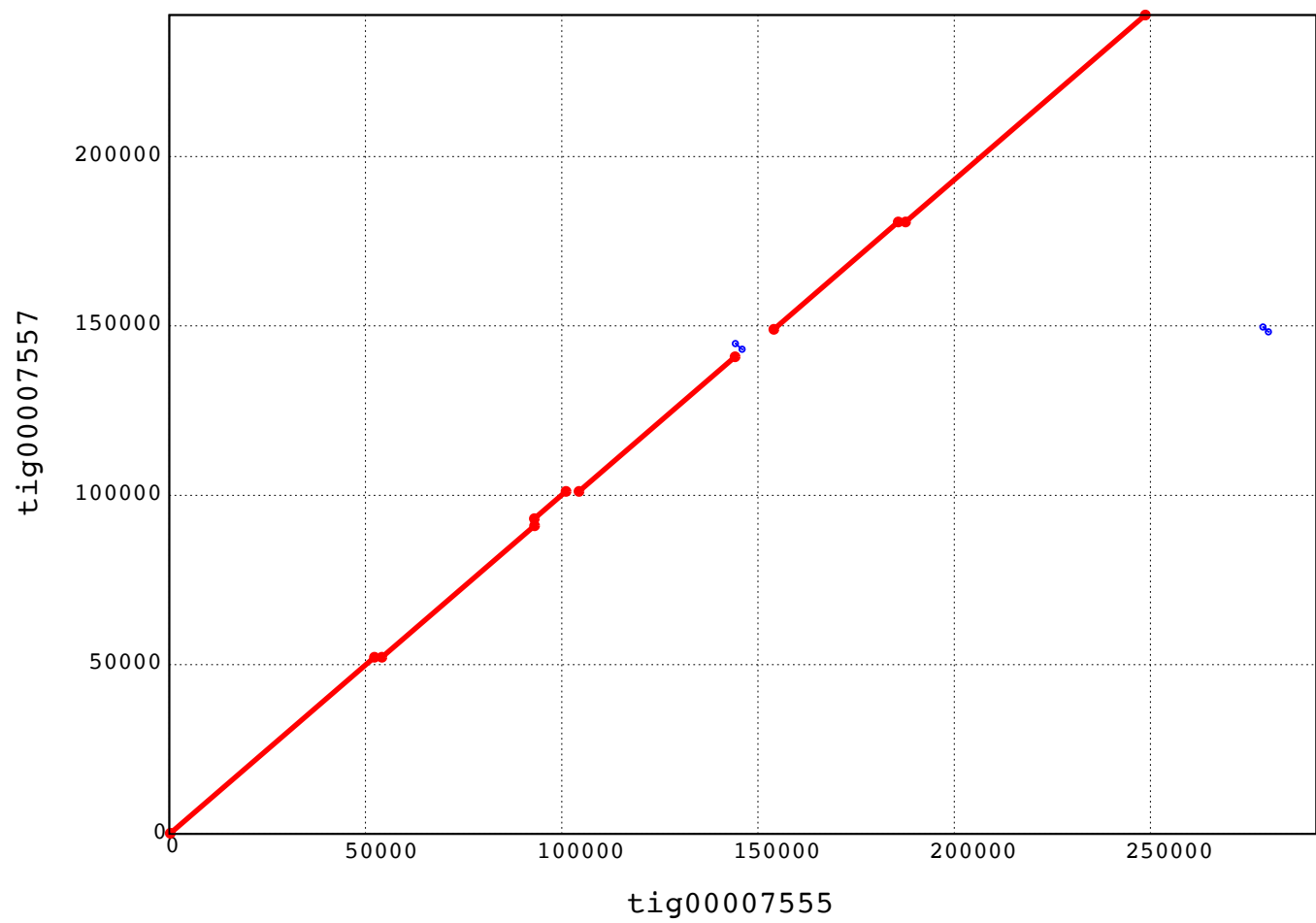
**Supplementary Figure 1f.** MUMmer dot plot of contig pair #6



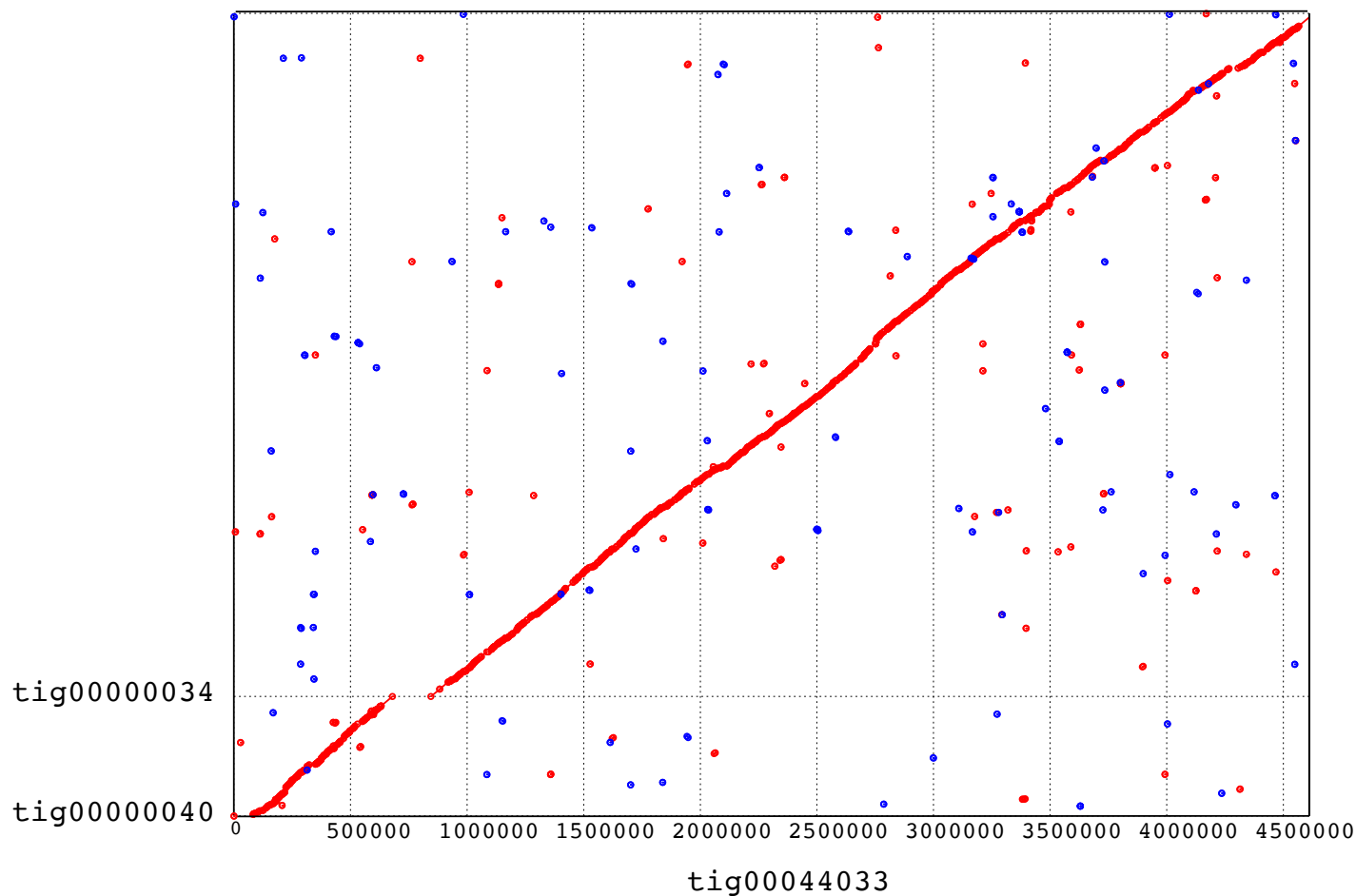
Supplementary Figure 1g. MUMmer dot plot of contig pair #7



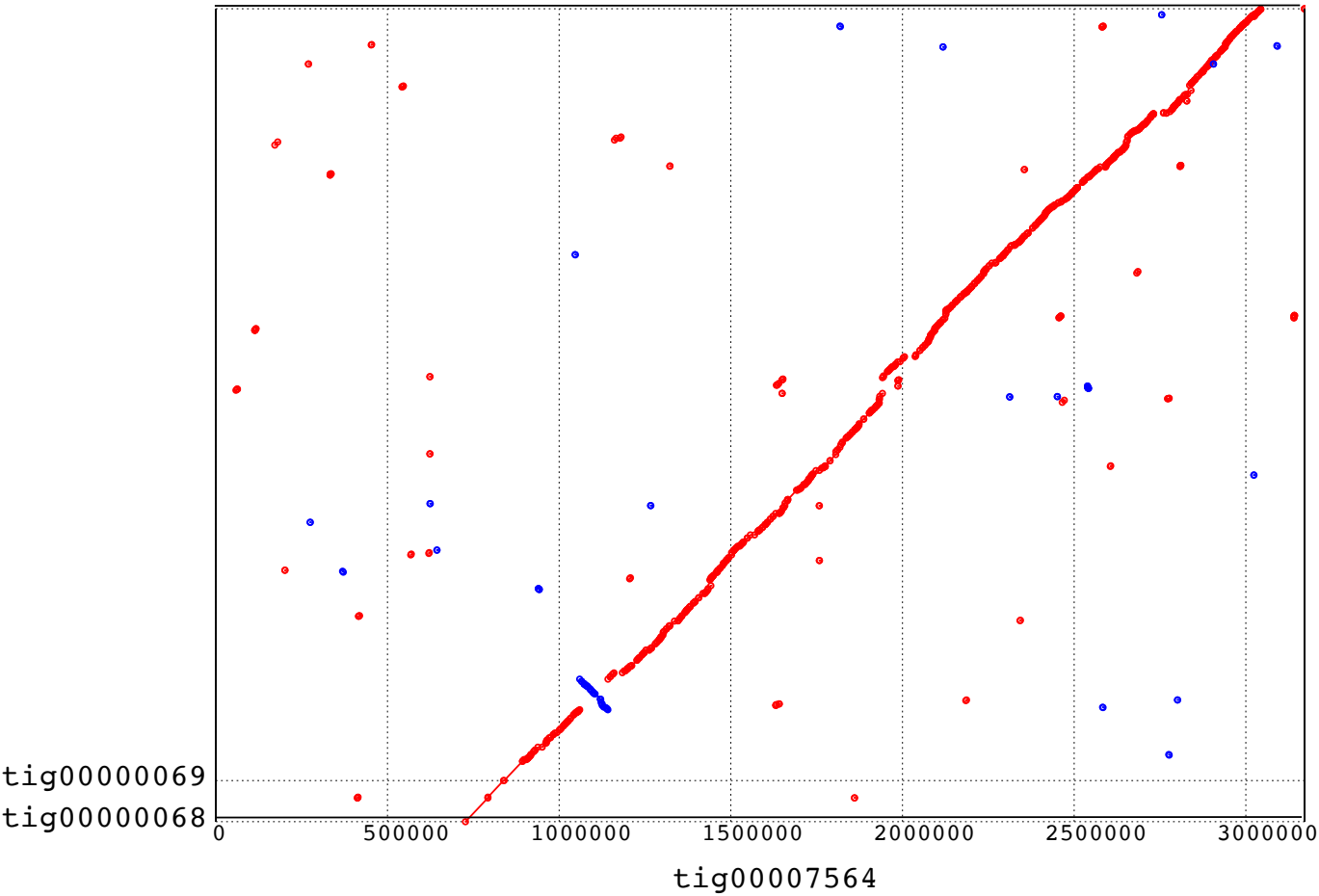
**Supplementary Figure 1h.** MUMmer dot plot of contig pair #8



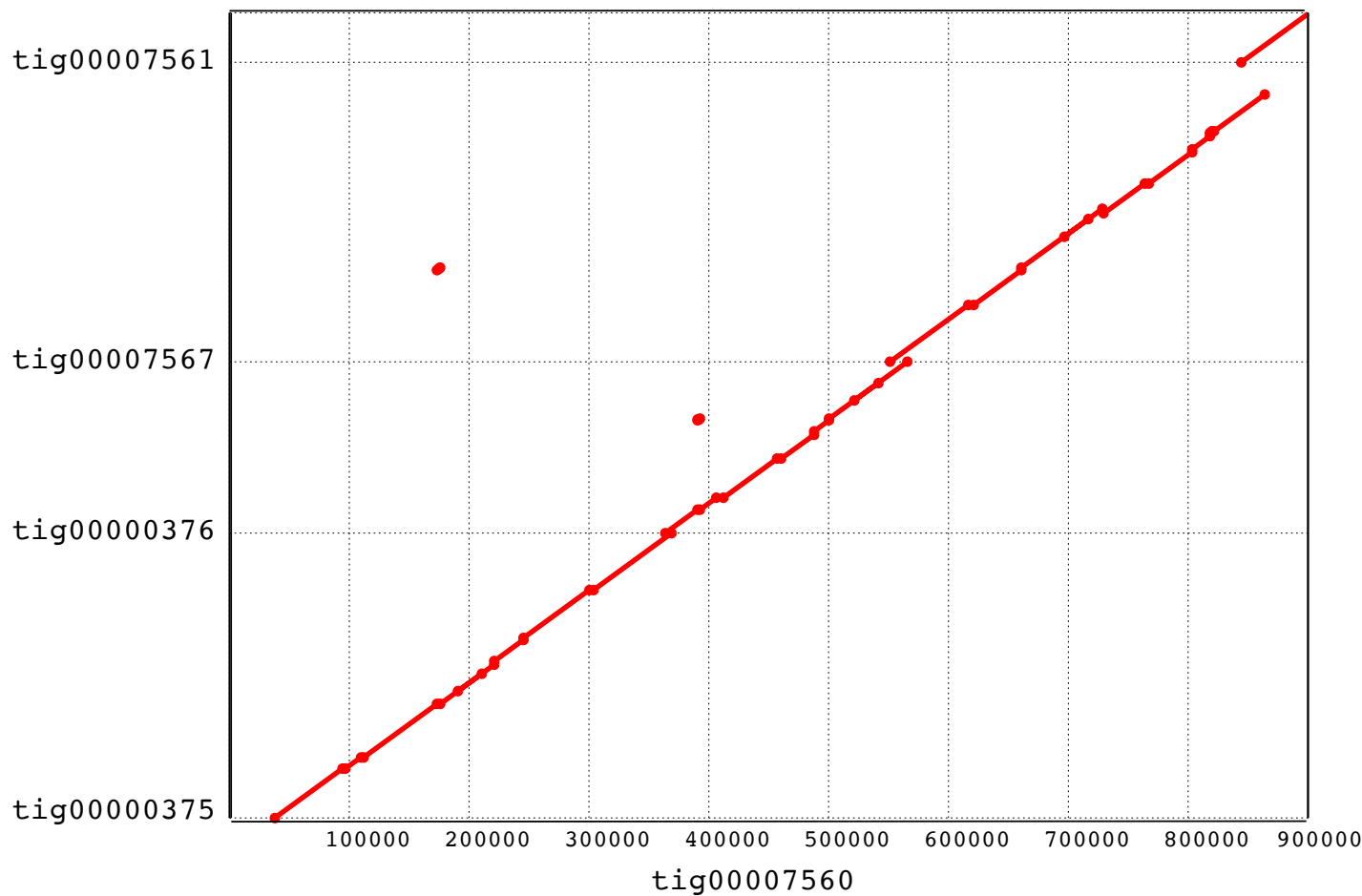
**Supplementary Figure 1i.** MUMmer dot plot of contig pair #9



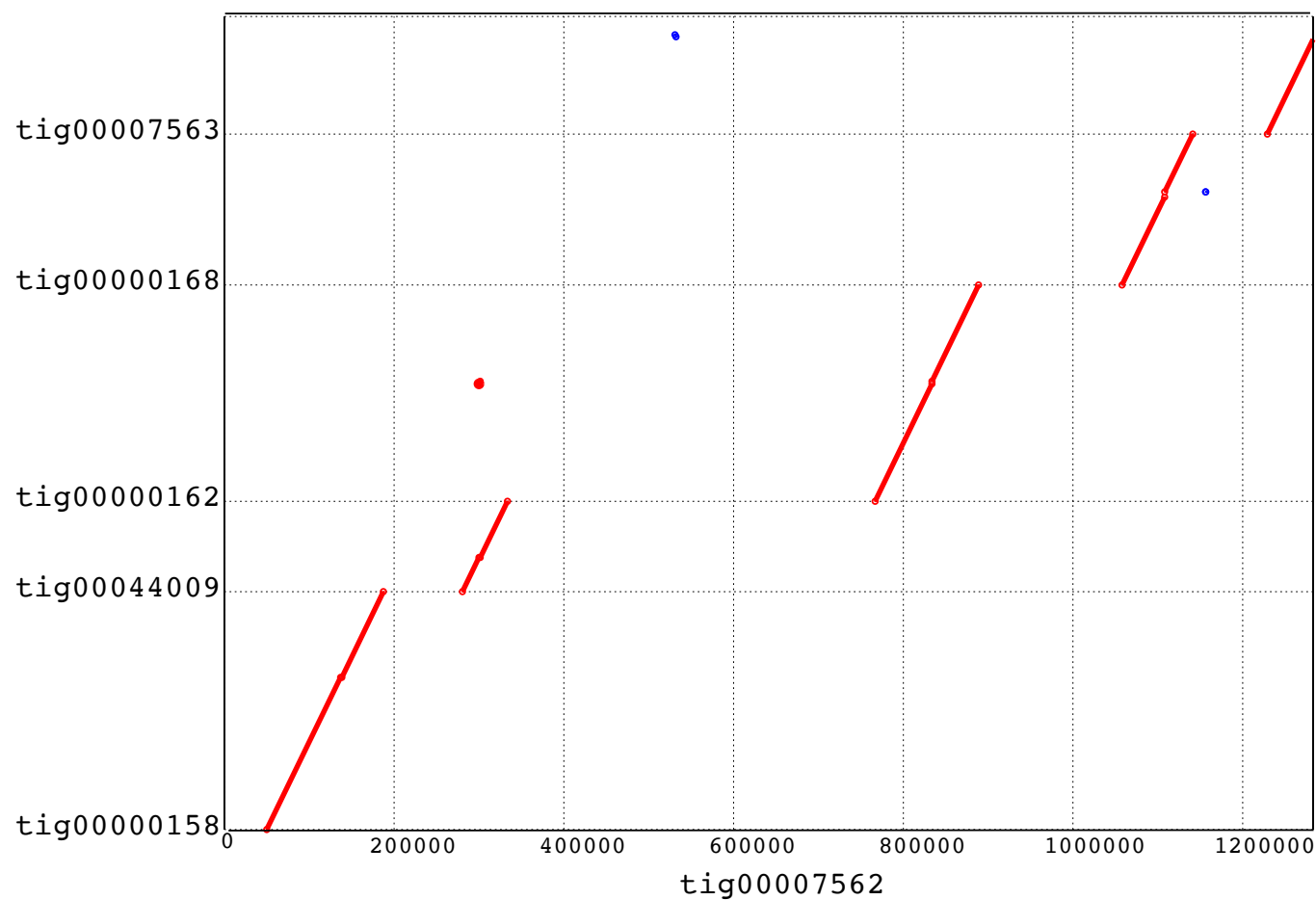
**Supplementary Figure 1j.** MUMmer dot plot of contig pair #10



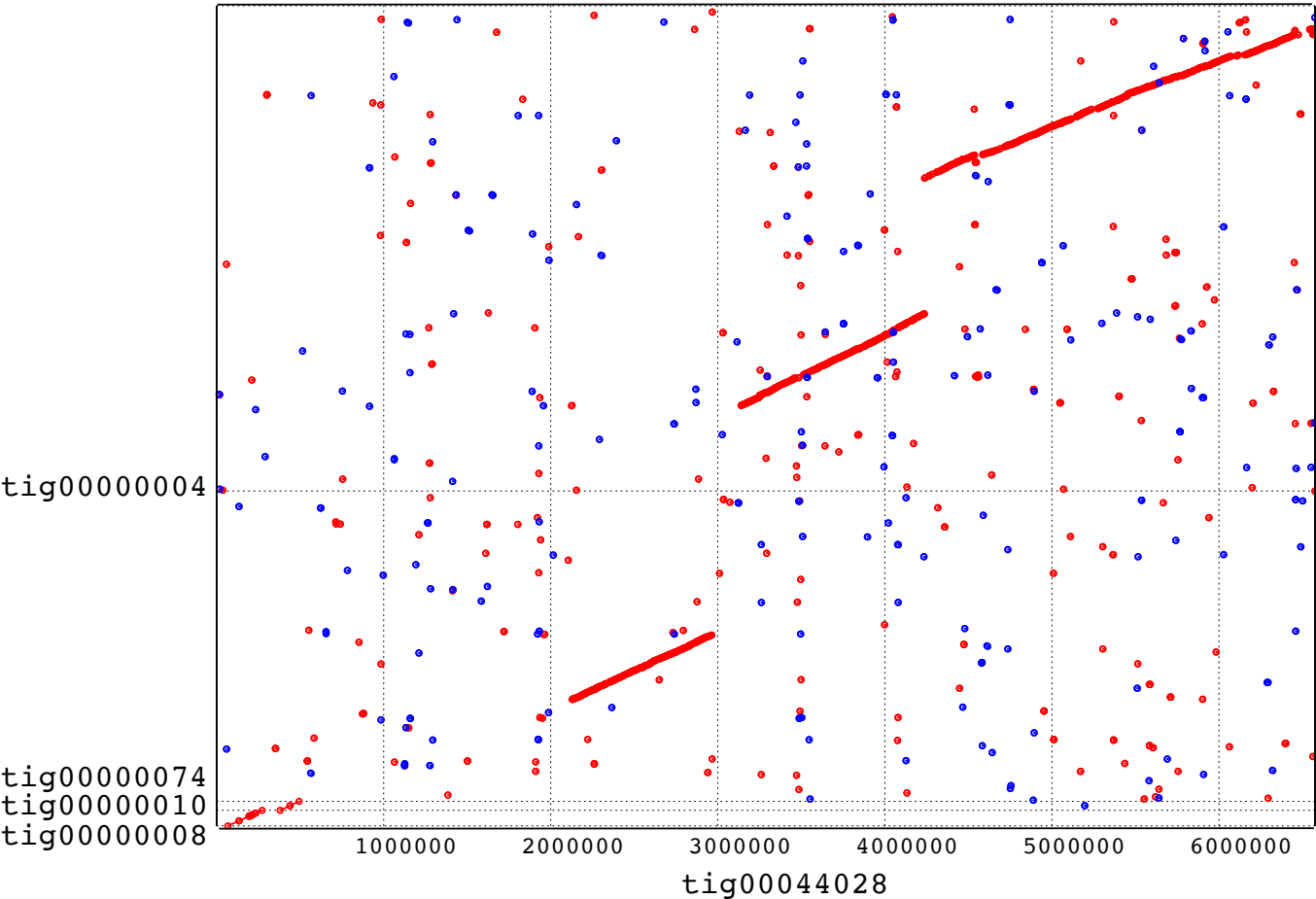
Supplementary Figure 1k. MUMmer dot plot of contig pair #11



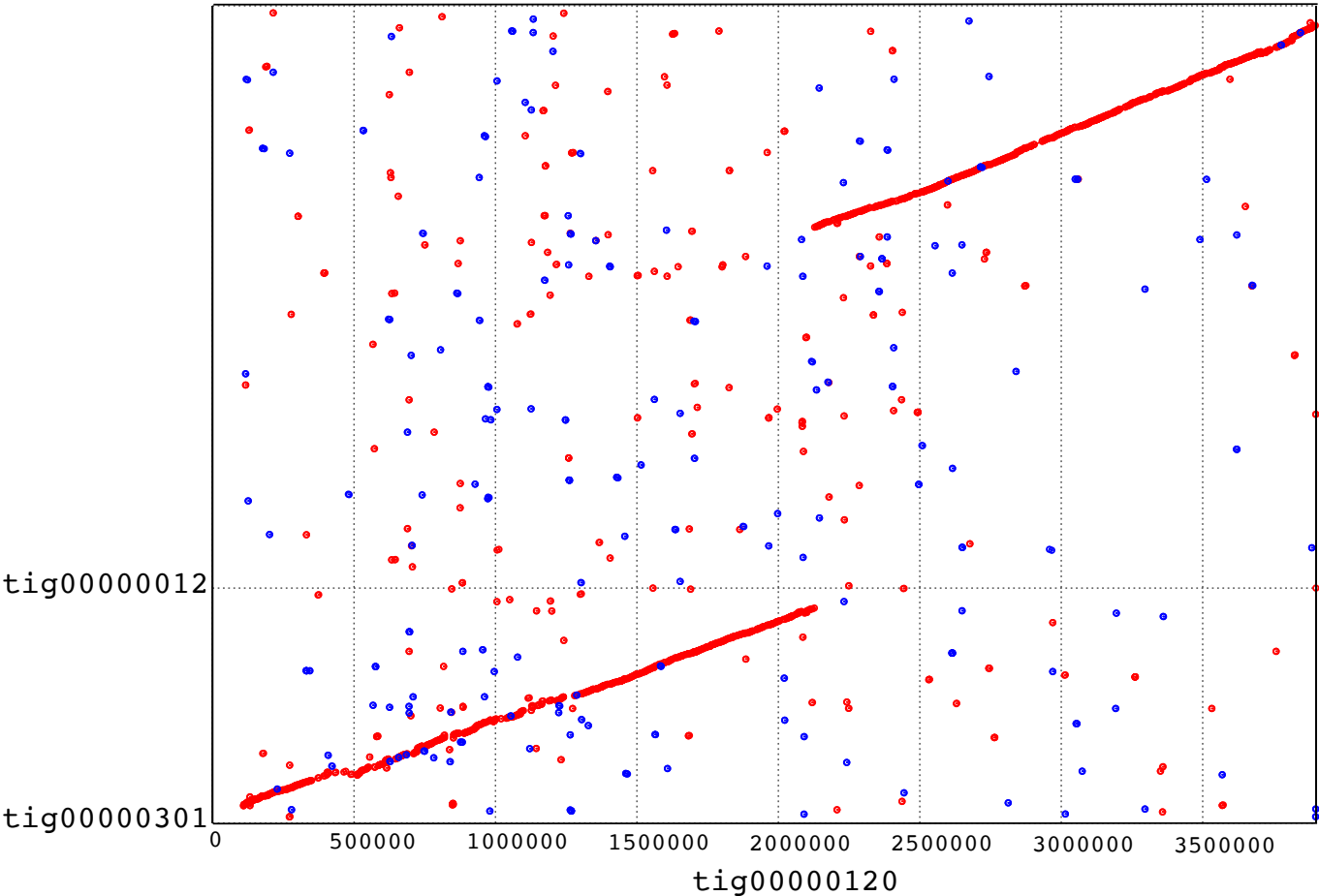
Supplementary Figure 1L. MUMmer dot plot of contig pair #12



Supplementary Figure 1M. MUMmer dot plot of contig pair #13

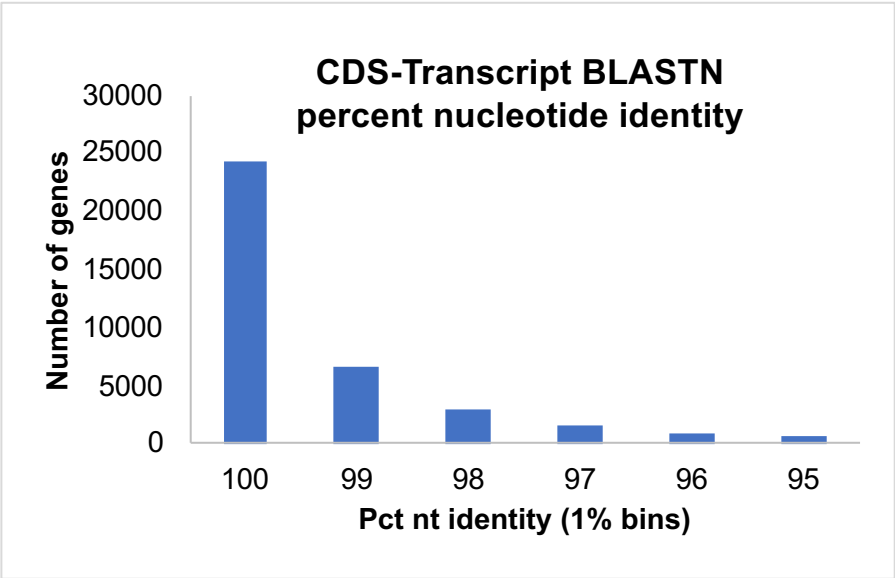


Supplementary Figure 1N. MUMmer dot plot of contig pair #14

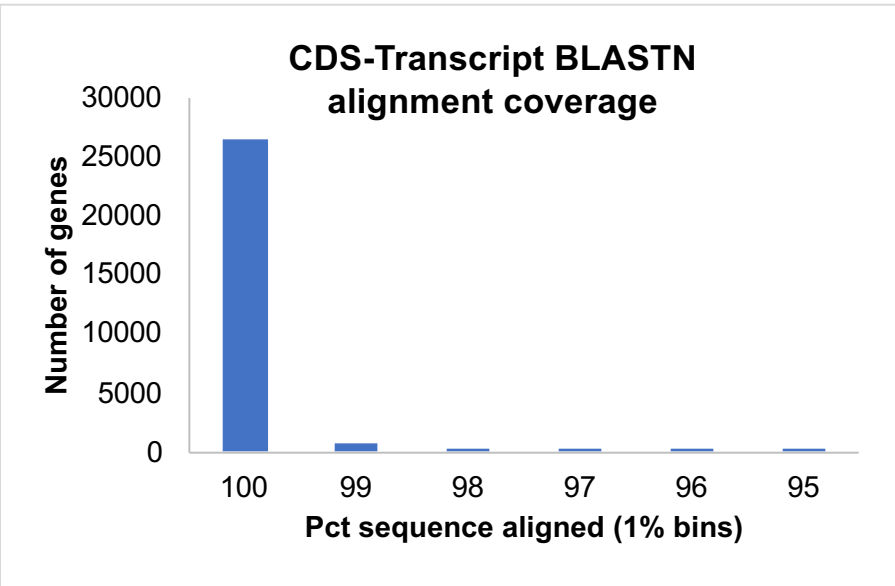


**Suppl Figure S2.** Transcriptome coverage of predicted coding sequences

**A**

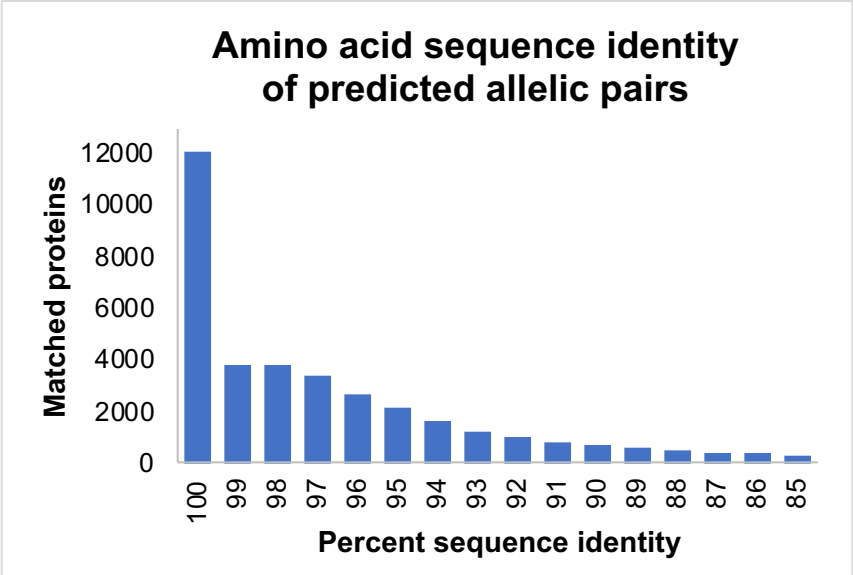


**B**

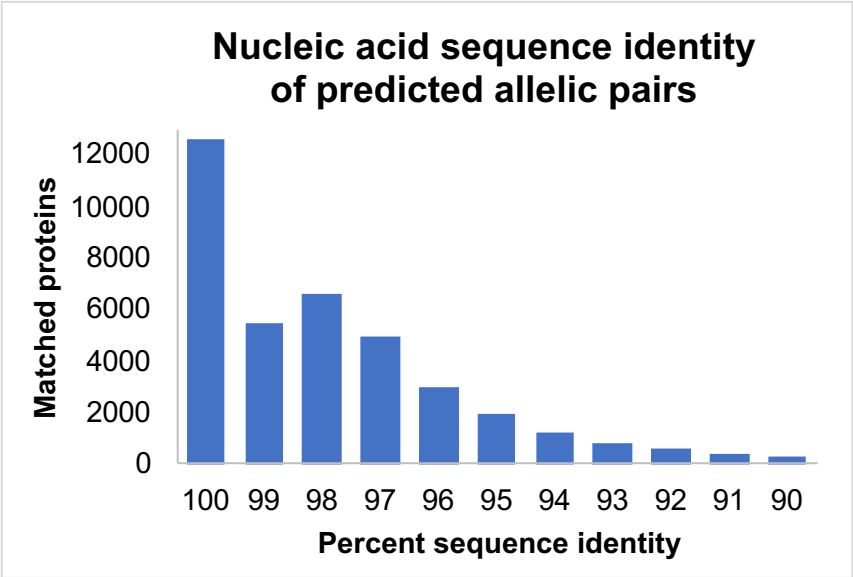


**Suppl Figure S3.** Sequence heterozygosity in diploid alleles of predicted proteins.

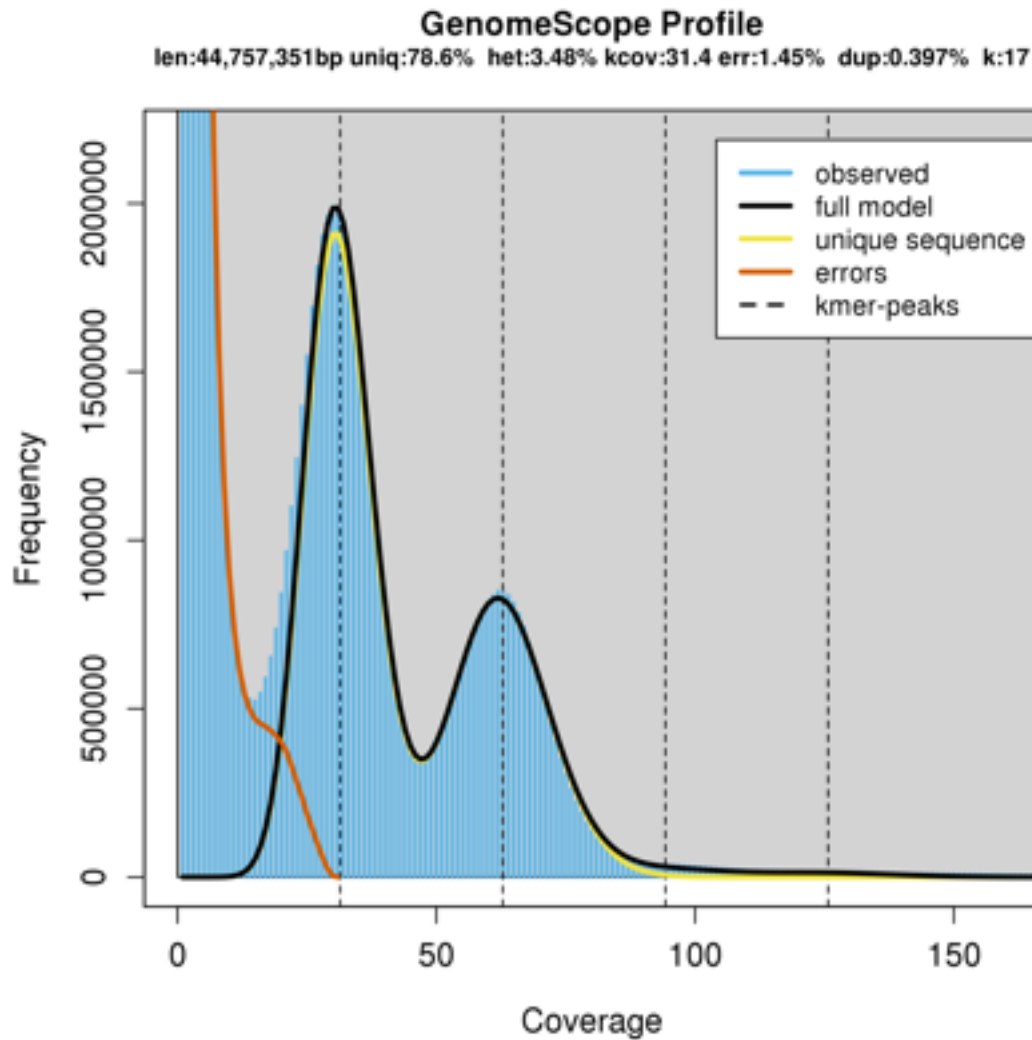
**A**



**B**



**Supplementary Figure S4.** Genome-wide heterozygosity based on 17-mer pattern frequency in unassembled sequencing reads. Heterozygosity was estimated from the relative heights of peaks A and B using an empirical formula derived from well-established reference genomes (Kajitani et al. *Genome Res.* 24(8):1384-95. 2014; Vurture et al. *Bioinformatics*, 33(14), 2202-2204. 2016).



**A**

[illegible]

B

26

**Suppl Figure S6.** Maximum-likelihood phylogeny of putative bicarbonate transporters from diatom genomes. Circles on branches indicate monophyly at the species level, X at the branch tips indicate outgroups. Centric diatom species are colored grey, pennate species are colored black (with the exception of outgroups), and haploid consensus *Nitzschia hildebrandi* sequences are indicated in dark red. Id number prefixes refer to Phycocosm genome catalogs for the following diatom species: Phatr2, *Phaeodactylum tricornutum*; Fisso1, *Fistulifera solaris*; Semro1, *Seminavis robusta*; Nithil2, *Nitzschia hildebrandi*; Psemu1, *Pseudo-nitzschia multiseriata*; Fracy1, *Fragilariopsis cylindrus*; Thaoce1, *Thalassiosira oceanica*; Mintr2, *Minidiscus variabilis*; Thaps3, *Thalassiosira pseudoana*; Cyccr1, *Cyclotella cryptica*.

