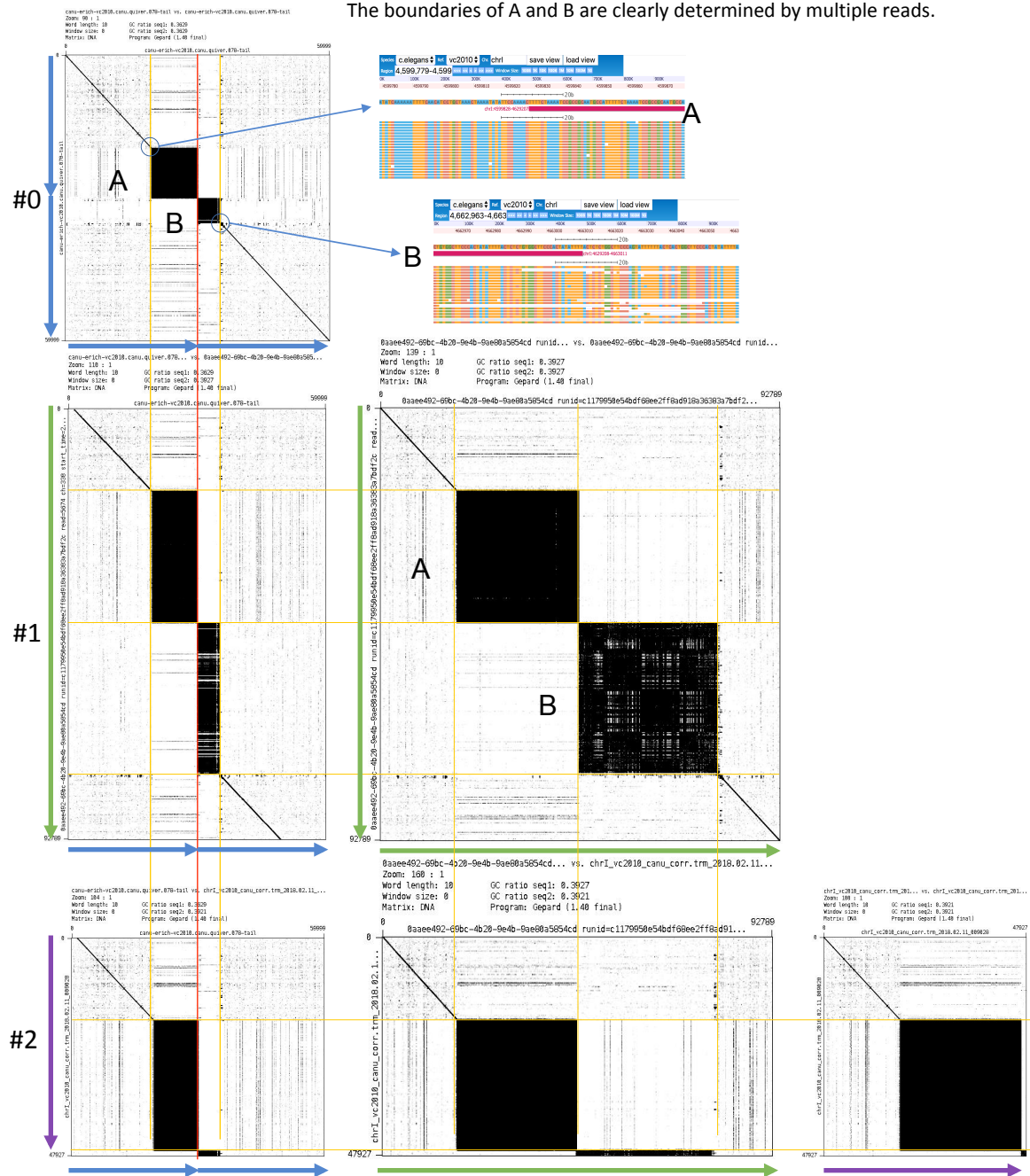


The boundaries of A and B are clearly determined by multiple reads.

Supp. Fig. S4A: Closing the gap in Chromosome I.



Read ID	length (b)	Base caller
#1	92,790	Albacore v2.1.7
#2	47,928	Albacore v2.1.7 and Canu's error-correction

repeat type	read ID	unit length (b)	number of units	repeat length (b)	Similarity among units
A	#0	26	405	10048	95.3%
	#1	25	1152	25392	85.8%
	#2	26	1130	27244	92.4%
B	#0	27	172	4303	89.6%
	#1	27	1252	25272	70.9%

type unit sequence in PacBio contigs

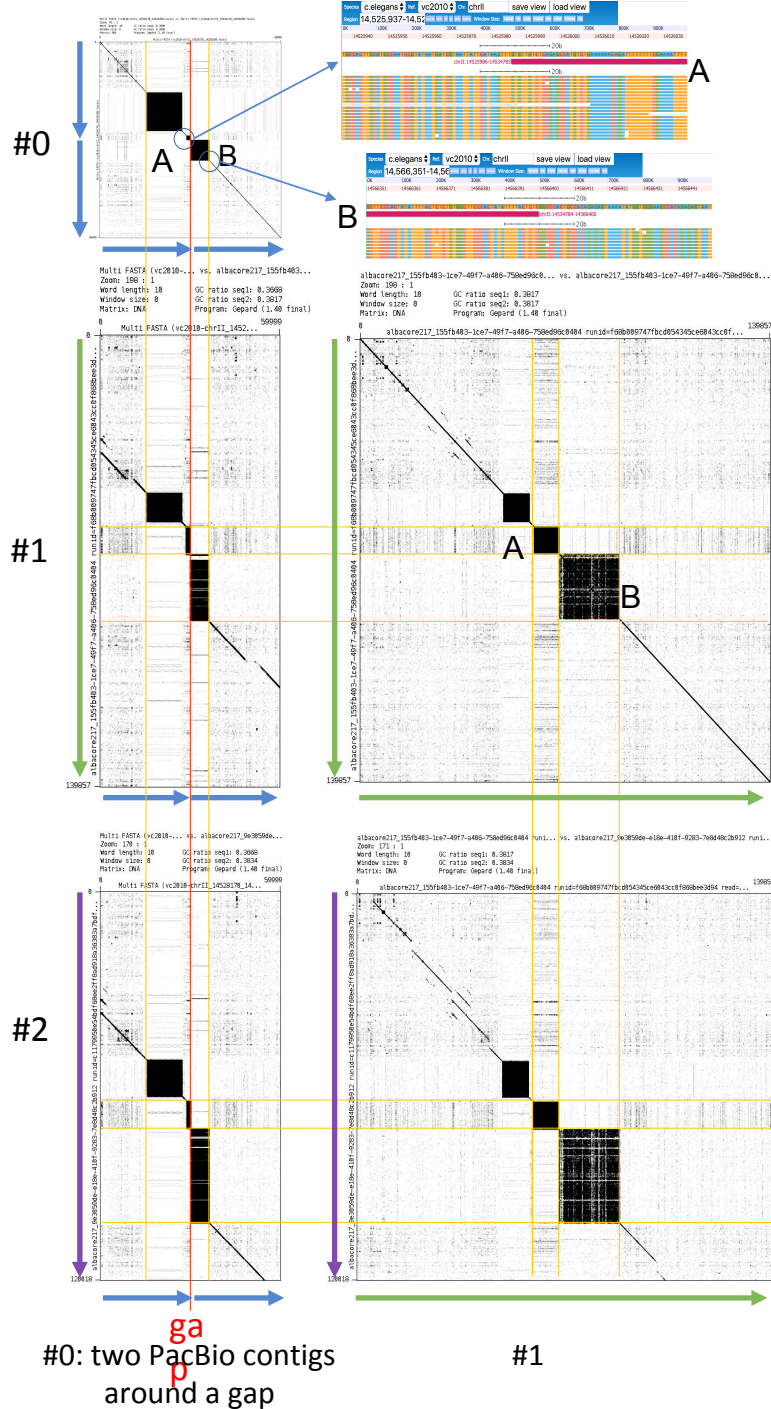
A TTTTCTAAAAATCCGCCGCAATGCCA

B ACTCTCTGTGGCTTCCCACTATATTTT

#0: two PacBio contigs around a gap

#1

#2



The boundaries
of A and B are
clear.

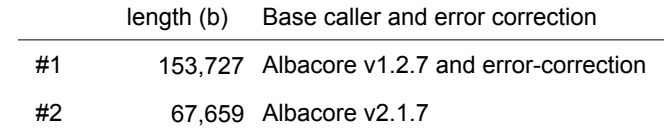
Supp. Fig. S4B: Closing the gap in Chromosome II.

	length (b)	Base caller
#1	139,858	Albacore v2.1.7
#2	120,019	Albacore v2.1.7

repeat type	read ID	unit length (b)	number of units	repeat length (b)	Similarity among units
A	#0	166	8	1378	96.8%
	#1	160	53	7793	85.9%
	#2	160	52	7534	85.1%
B	#0	27	223	5769	94.7%
	#1	29	760	16061	69.0%
	#2	26	1171	23015	71.3%

type	unit sequence in PacBio contigs
A	GTGTTGTCTCAGAGCCATGTCTGAAAAAA CAGATTTTTTTTTCGAAAATTTTAAAAAGT TCGTAAAAATTGGATTTTTTGACAAATCTA TGCTCTCTTTTTTTGTTTGAAATTTAGTT CAGTGACATACAATTTCAAAAATATAATG TTCTGAATTATTCATAAGTTT
B	TGTGGCTTCCCACTATATTTTACTCTC

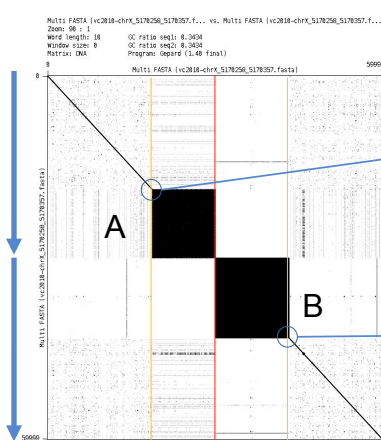
Supp. Fig. S4C: Closing the gap in Chromosome X (nt 4,149,383-4,226,837).



repeat type	read ID	unit length (b)	number of units	repeat length (b)	Similarity among units
A	#0	25	269	6506	96.4%
	#1	25	1023	22730	87.3%
	#2	25	197	3822	74.5%
B	#0	20	826	15200	91.1%
	#1	20	684	11447	82.6%
	#2	20	2594	42344	79.0%

type	unit sequence in PacBio contigs
A	CTAATGTTCTGCCAATTTATTTACT
B	ATATCGTATCGTAAATAGCT

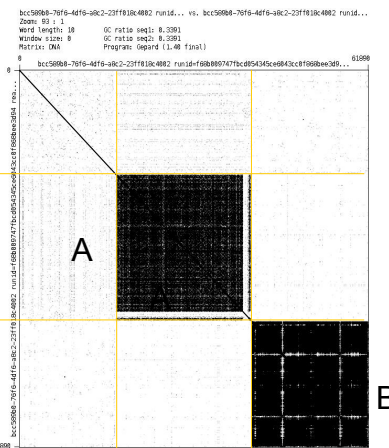
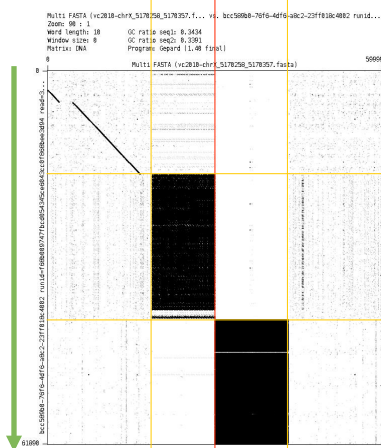
#0



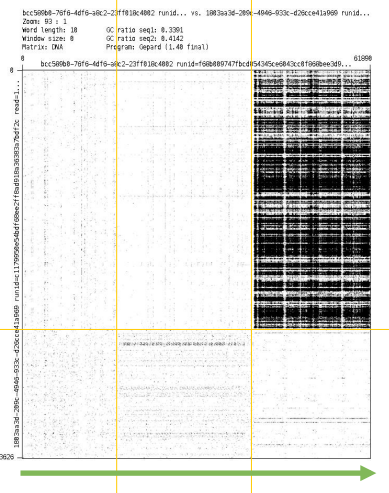
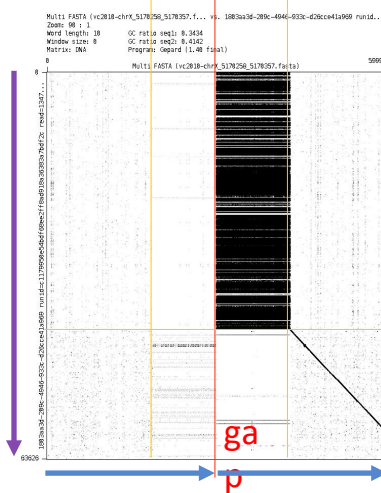
The boundaries of A and B are clearly determined by multiple reads.



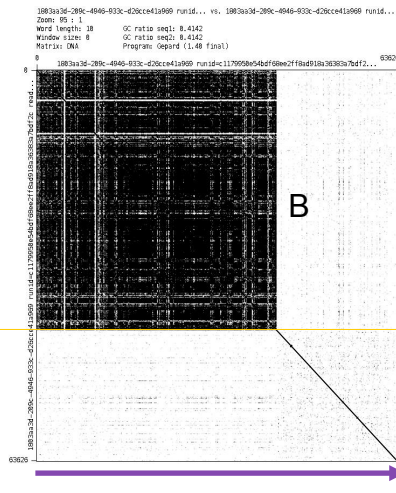
#1



#2



B



Supp. Fig. S4D: Partially closing the gap in Chromosome X (nt 5,215,995-5,284,061).

	length (b)	Base caller
#1	61,891	MinKNOW
#2	63,627	Albacore v2.1.7

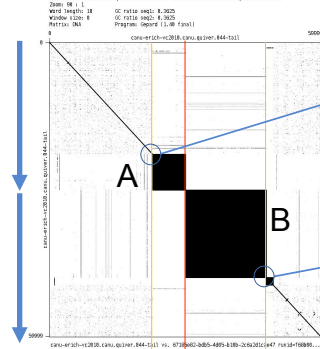
repeat type	read ID	unit length (b)	number of units	repeat length (b)	Similarity among units
A	#0	226	49	10981	97.6%
	#1	228	98	20676	89.0%
B	#0	27	497	13102	97.5%
	#1	24	882	17477	78.1%
	#2	26	1697	34319	74.0%

type	unit sequence in PacBio contigs
A	AAAACTACCATTTATAGACAAAAATAGATGAATTTTCCAAAC TTTGAAAATTCATAAATCTCTTCAAAGTAACTTTTTTATAAA TGTCTTTTCAGAAACTTTGTAGTAAATTTTAAGCTCTTTCTGA ATATATTAACAATATTCCAGTAGGTACAAGAAGCTTCACGTA GTTACAGAAATAGTACATTTTCAGCCCTACCTTTTAGTGCCT ATTTTATTATTAAATG
B	TATTTTACTCTCTGTGGCTTCCCACTA

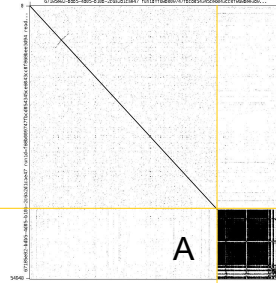
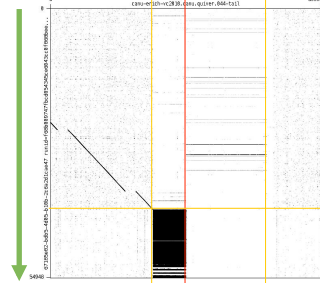
The boundaries of A and B are clearly determined by multiple reads.

Supp. Fig. S4E: Partially closing the gap in Chromosome III.

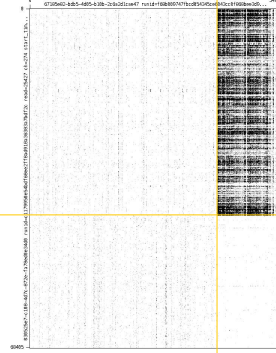
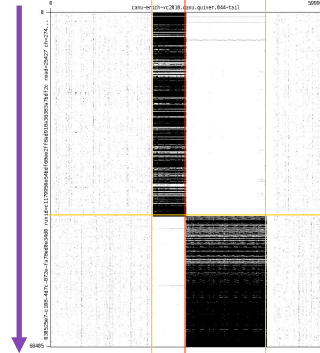
#0



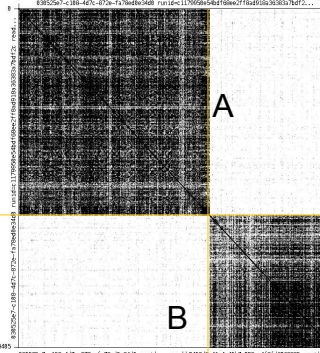
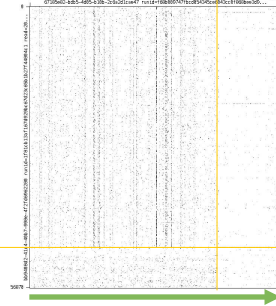
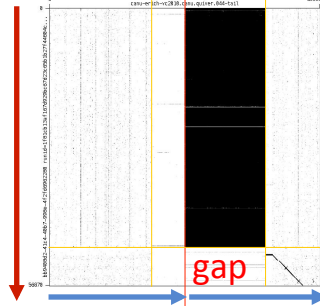
#1



#2



#3

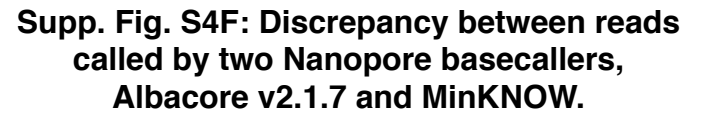


	length (b)	Base caller
#1	54,949	MinKNOW
#2	68,406	MinKNOW
#3	56,871	Albacore v2.1.7

repeat type	read ID	unit length (b)	number of units	repeat length (b)	Similarity among units
A	#0	27	279	7135	94.5%
	#1	24	611	11391	73.7%
	#2	26	1660	32015	69.3%
B	#0	94	192	17577	97.0%
	#2	88	306	21864	75.9%
	#3	93	539	43406	82.6%

type	unit sequence in PacBio contigs
A	AGAGTAAATATAGTGGGAAGCCACAG
B	AAAGCACCCAATATTTAGAGAACAGAAGA TTTTGAGAATTACTGCCTCCAGAAATTGA TGATTTTCCCATTTGATTGTCTACATAGG GCATCGA

#1



repeat type	read ID	unit length (b)	number of units	repeat length (b)	Similarity among units
A	#1	25	1152	25392	85.8%
	#2	25	1130	25157	86.4%
B	#1	27	1252	25272	70.9%
	#2	22	1095	16831	65.6%