

Zea_mays
MBS847

Contig assembly

Wtdbg2 assembly
73X ONT

Flye assembly
73X ONT

Total size : 3 886 168 405

% of assumed genome size : 161.9 %

N50 contig length : 81 093

Total size : 2 182 167 593

% of assumed genome size : 90.9 %

N50 contig length : 3 122 117

Polishing

73X ONT
65X 10X Chromium

10X HiFi
(1e smrtCell)

Total size : 2 184 316 260

% of assumed genome size : 90.9 %

N50 contig length : 3 571 534

Busco : 94.0%[S:83.3%,D:10.7%]
F:1.9%,M:4.1%

Total size : 2 184 399 715

% of assumed genome size : 91 %

N50 contig length : 3 125 047

Busco : 97.3%[S:91.7%,D:5.6%]
F:1.1%,M:1.6%

Juicer stats

	MBS847_123	MBS847_11/02	MBS847_28/02	MBS847_05/06	MBS847_11/02 VS NRGene
Sequenced Read Pairs					
Normal Paired	316 260 404	2 535 954	108 725 802	206 127 921	2 535 954
Chimeric Paired	222 026 529 (70.20%)	1 943 694 (76.65%)	83 501 291 (76.80%)	136 210 474 (66.08%)	1 945 168 (76.70%)
Chimeric Ambiguous	36 989 807 (11.70%)	229 919 (9.07%)	9 669 251 (8.89%)	26 941 241 (13.07%)	247 364 (9.75%)
Unmapped	54 832 761 (17.34%)	350 380 (13.82%)	14 966 057 (13.76%)	41 175 340 (19.98%)	330 561 (13.03%)
Ligation Motif Present	2 411 307 (0.76%)	11 961 (0.47%)	589 203 (0.54%)	1 800 866 (0.87%)	12 893 (0.51%)
Alignable (Normal+Chimeric Paired)	48 529 733 (15.34%)	98 927 (3.90%)	4 252 498 (3.91%)	44 178 308 (21.43%)	98 927 (3.90%)
Unique Reads	259 016 336 (81.90%)	2 173 613 (85.71%)	93 170 542 (85.69%)	163 151 715 (79.15%)	2 192 532 (86.46%)
PCR Duplicates	228 966 469	2 170 738	84 024 147	143 843 684	2 190 592
Optical Duplicates	27,196,060	2,333	7,526,575	18,068,515	1,385
Library Complexity Estimate	2,853,807	542	1,619,820	1,239,516	555
Intra-fragment Reads	1,119,420,279	1,011,328,143	525,841,611	670,409,536	1,733,840,804
Below MAPQ Threshold	24,113,266 (7.62% / 10.53%)	360,573 (14.22% / 16.61%)	13,053,845 (12.01% / 15.54%)	10,920,483 (5.30% / 7.59%)	422,241 (16.65% / 19.28%)
Hi-C Contacts	103,632,634 (32.77% / 45.26%)	867,855 (34.22% / 39.98%)	35,476,630 (32.63% / 42.22%)	67,860,214 (32.92% / 47.18%)	817,512 (32.24% / 37.32%)
Ligation Motif Present	101,220,569 (32.01% / 44.21%)	942,310 (37.16% / 43.41%)	35,493,672 (32.65% / 42.24%)	65,062,987 (31.56% / 45.23%)	950,839 (37.49% / 43.41%)
3' Bias (Long Range)	10,347,569 (3.27% / 4.52%)	27,327 (1.08% / 1.26%)	1,025,697 (0.94% / 1.22%)	9,311,972 (4.52% / 6.47%)	28,293 (1.12% / 1.29%)
Pair Type % (L-I-O-R)	50% - 50%	44% - 56%	44% - 56%	52% - 48%	45% - 55%
Inter-chromosomal	25% - 25% - 25% - 25%	25% - 26% - 25% - 24%	24% - 26% - 26% - 24%	25% - 25% - 25% - 25%	25% - 25% - 25% - 25%
Intra-chromosomal	36,995,453 (11.70% / 16.16%)	299,943 (11.83% / 13.82%)	11,428,480 (10.51% / 13.60%)	25,278,778 (12.26% / 17.57%)	152,977 (6.03% / 6.98%)
Short Range (<20Kb)	64,225,116 (20.31% / 28.05%)	642,367 (25.33% / 29.59%)	24,065,192 (22.13% / 28.64%)	39,784,209 (19.30% / 27.66%)	797,862 (31.46% / 36.42%)
Long Range (>20Kb)	53,013,737 (16.76% / 23.15%)	593,146 (23.39% / 27.32%)	22,209,401 (20.43% / 26.43%)	30,494,478 (14.79% / 21.20%)	665,364 (26.24% / 30.37%)
	11,205,961 (3.54% / 4.89%)	49,175 (1.94% / 2.27%)	1,853,626 (1.70% / 2.21%)	9,286,530 (4.51% / 6.46%)	132,436 (5.22% / 6.05%)

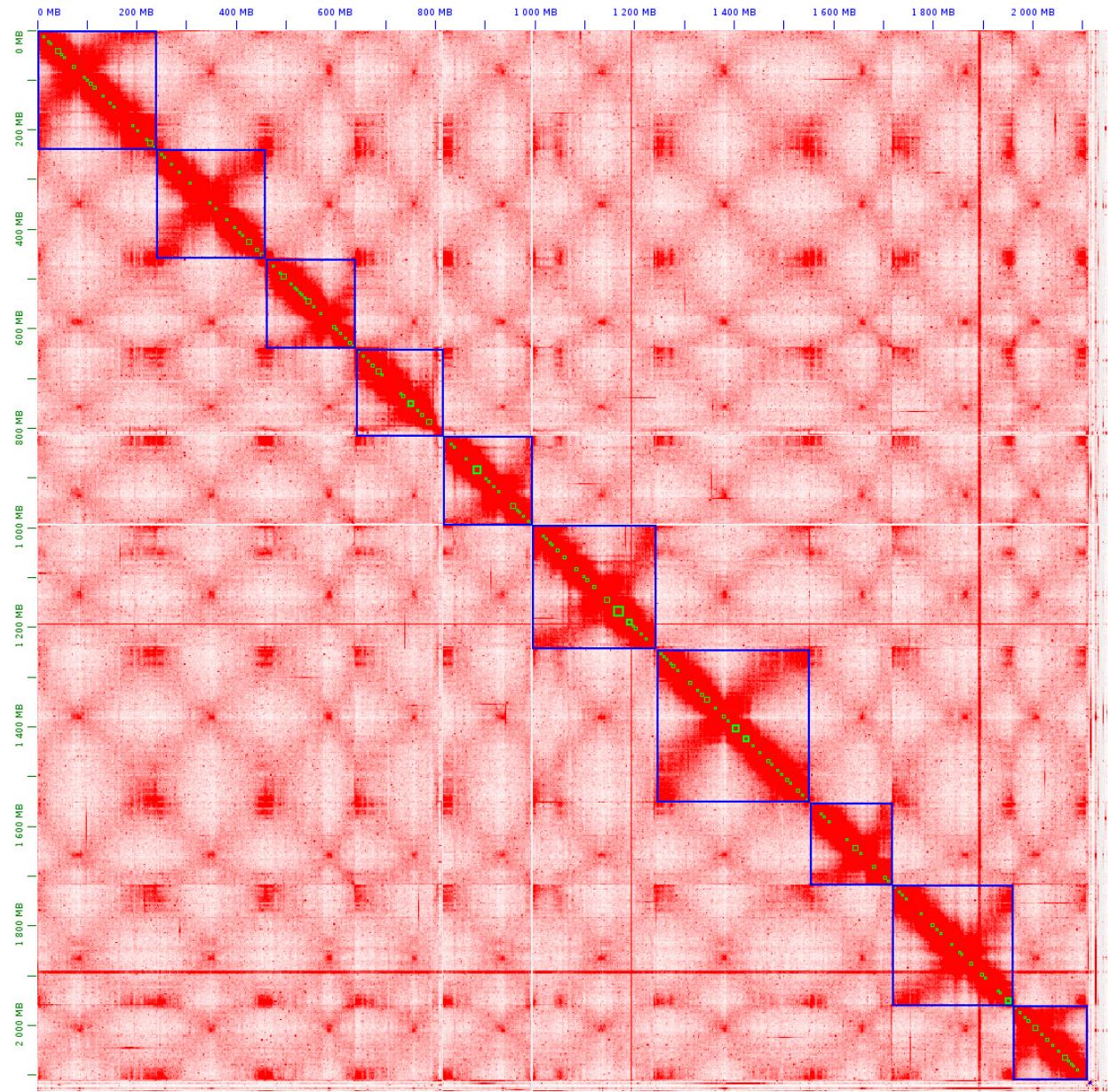
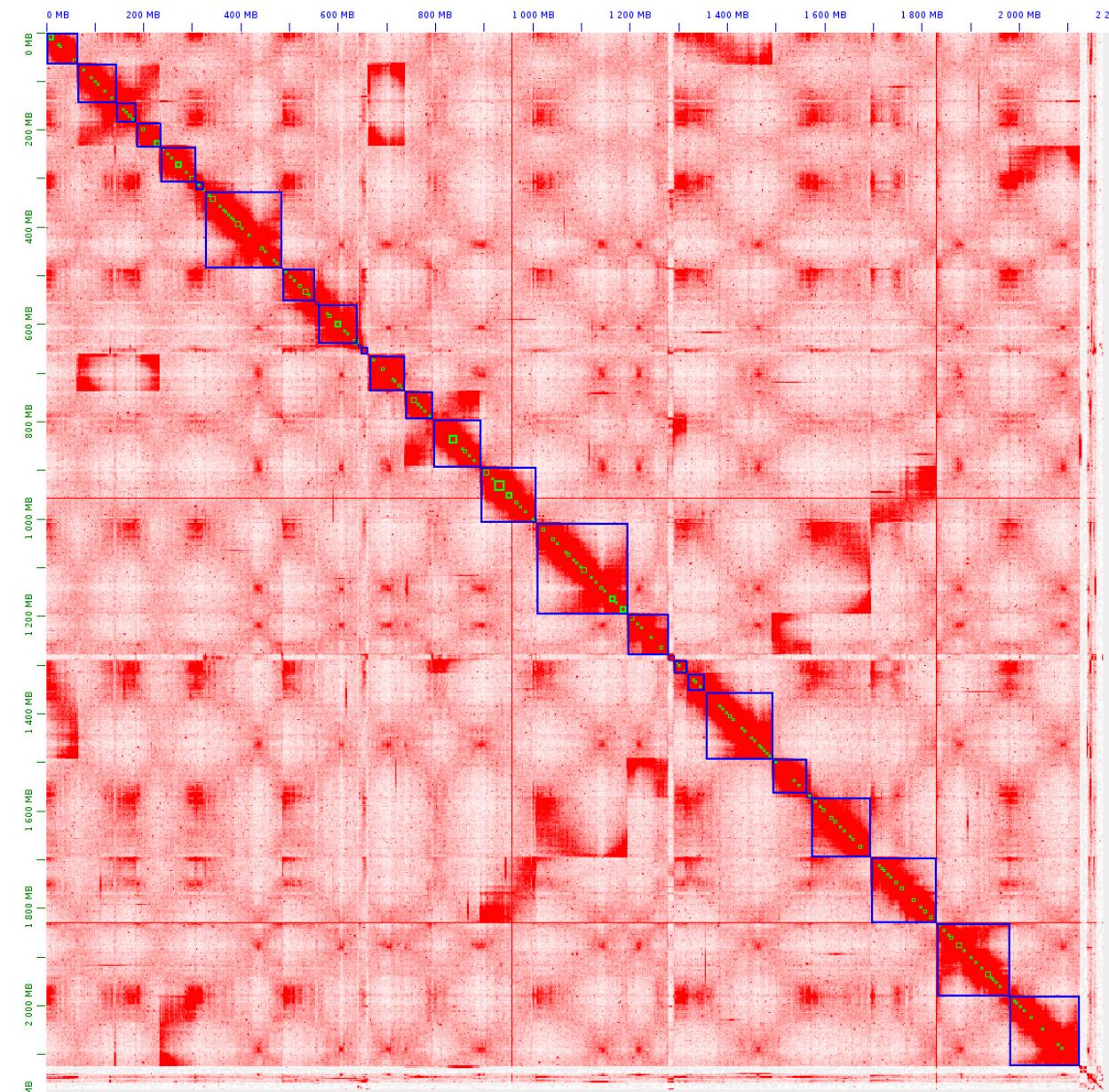
- Around 80% reads alignable
- Low HiC contact due to repetitive regions
- Low Intra chromosomal and Long Range
- Fragmented assembly

Juicer stats

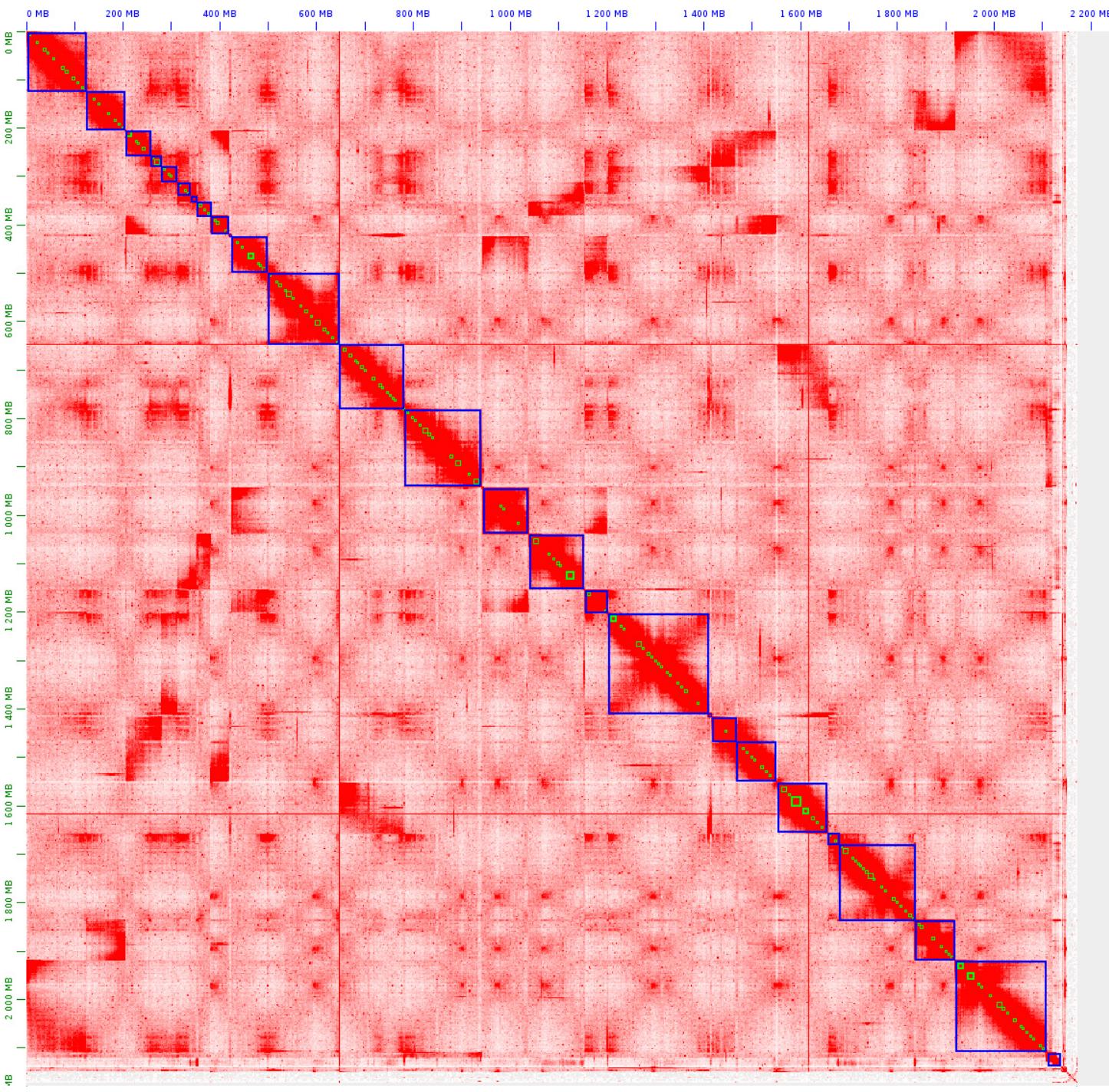
	MBS847_123	MBS847_11/02	MBS847_28/02	MBS847_05/06	MBS847_11/02 VS NRGene
Sequenced Read Pairs					
Normal Paired	316 260 404	2 535 954	108 725 802	206 127 921	2 535 954
Chimeric Paired	222 026 529 (70.20%)	1 943 694 (76.65%)	83 501 291 (76.80%)	136 210 474 (66.08%)	1 945 168 (76.70%)
Chimeric Ambiguous	36 989 807 (11.70%)	229 919 (9.07%)	9 669 251 (8.89%)	26 941 241 (13.07%)	247 364 (9.75%)
Unmapped	54 832 761 (17.34%)	350 380 (13.82%)	14 966 057 (13.76%)	41 175 340 (19.98%)	330 561 (13.03%)
Ligation Motif Present	2 411 307 (0.76%)	11 961 (0.47%)	589 203 (0.54%)	1 800 866 (0.87%)	12 893 (0.51%)
Alignable (Normal+Chimeric Paired)	48 529 733 (15.34%)	98 927 (3.90%)	4 252 498 (3.91%)	44 178 308 (21.43%)	98 927 (3.90%)
Unique Reads	259 016 336 (81.90%)	2 173 613 (85.71%)	93 170 542 (85.69%)	163 151 715 (79.15%)	2 192 532 (86.46%)
PCR Duplicates	228 966 469	2 170 738	84 024 147	143 843 684	2 190 592
Optical Duplicates	27,196,060	2,333	7,526,575	18,068,515	1,385
Library Complexity Estimate	2,853,807	542	1,619,820	1,239,516	555
Intra-fragment Reads	1,119,420,279	1,011,328,143	525,841,611	670,409,536	1,733,840,804
Below MAPQ Threshold	24,113,266 (7.62% / 10.53%)	360,573 (14.22% / 16.61%)	13,053,845 (12.01% / 15.54%)	10,920,483 (5.30% / 7.59%)	422,241 (16.65% / 19.28%)
Hi-C Contacts	103,632,634 (32.77% / 45.26%)	867,855 (34.22% / 39.98%)	35,476,630 (32.63% / 42.22%)	67,860,214 (32.92% / 47.18%)	817,512 (32.24% / 37.32%)
Hi-C Contacts	101,220,569 (32.01% / 44.21%)	942,310 (37.16% / 43.41%)	35,493,672 (32.65% / 42.24%)	65,062,987 (31.56% / 45.23%)	950,839 (37.49% / 43.41%)
Ligation Motif Present	10,347,569 (3.27% / 4.52%)	27,327 (1.08% / 1.26%)	1,025,697 (0.94% / 1.22%)	9,311,972 (4.52% / 6.47%)	28,293 (1.12% / 1.29%)
3' Bias (Long Range)	50% - 50%	44% - 56%	44% - 56%	52% - 48%	45% - 55%
Pair Type % (L-I-O-R)	25% - 25% - 25% - 25%	25% - 26% - 25% - 24%	24% - 26% - 26% - 24%	25% - 25% - 25% - 25%	25% - 25% - 25% - 25%
Inter-chromosomal	36,995,453 (11.70% / 16.16%)	299,943 (11.83% / 13.82%)	11,428,480 (10.51% / 13.60%)	25,278,778 (12.26% / 17.57%)	152,977 (6.03% / 6.98%)
Intra-chromosomal	64,225,116 (20.31% / 28.05%)	642,367 (25.33% / 29.59%)	24,065,192 (22.13% / 28.64%)	39,784,209 (19.30% / 27.66%)	797,862 (31.46% / 36.42%)
Short Range (<20Kb)	53,013,737 (16.76% / 23.15%)	593,146 (23.39% / 27.32%)	22,209,401 (20.43% / 26.43%)	30,494,478 (14.79% / 21.20%)	665,364 (26.24% / 30.37%)
Long Range (>20Kb)	11,205,961 (3.54% / 4.89%)	49,175 (1.94% / 2.27%)	1,853,626 (1.70% / 2.21%)	9,286,530 (4.51% / 6.46%)	132,436 (5.22% / 6.05%)

- Around 80% reads alignable
- Low HiC contact due to repetitive regions
- Low Intra chromosomal and Long Range
- Fragmented assembly

Run 05/06



All runs



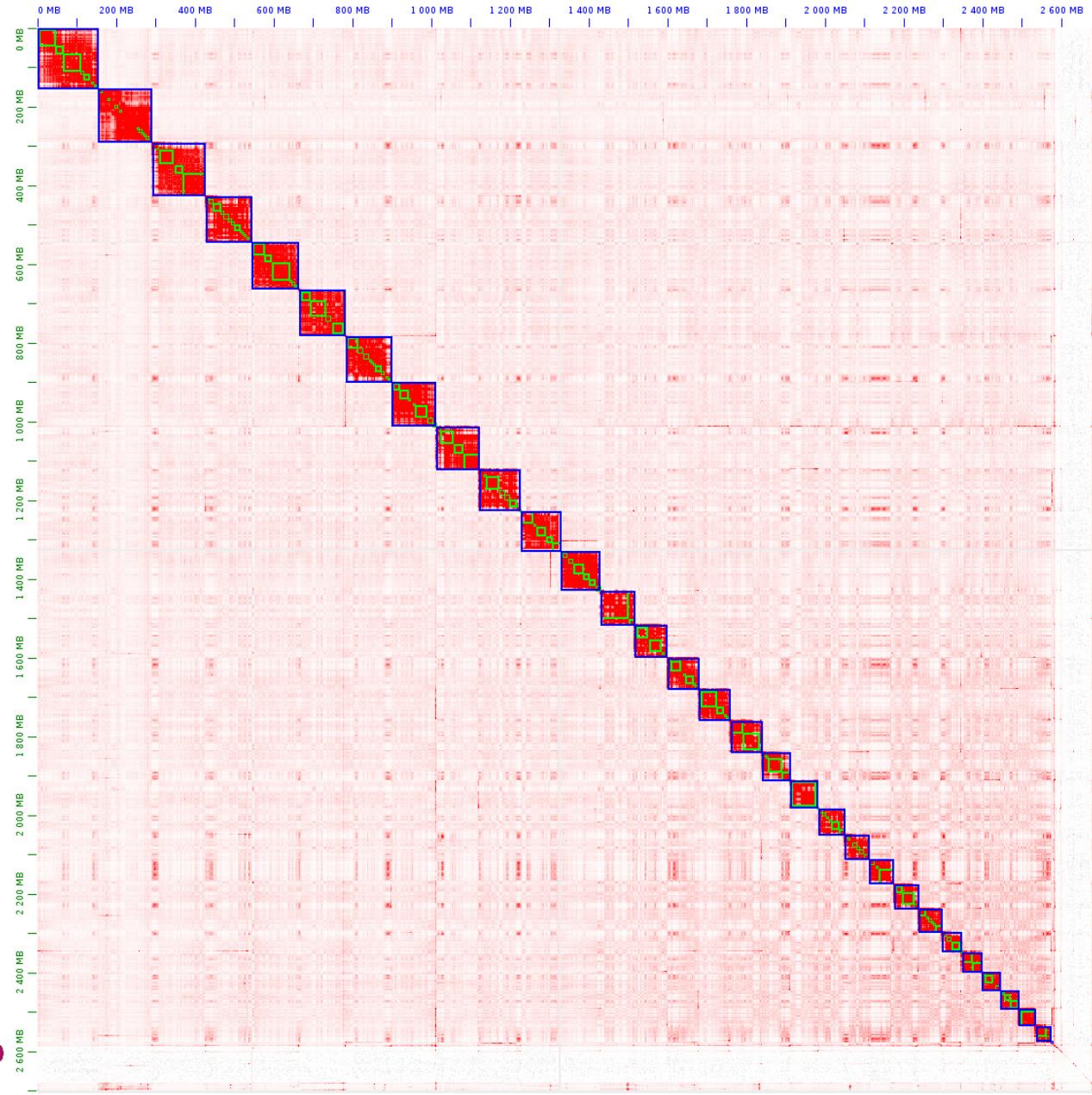
Offspring 2 - 37160



Assembly protocol

- 1- Wtdbg2 contig assembly with 58X ONT reads
- 2- Polishing Racon with 58X ONT reads
- 3- Polishing Pilon with 102X 10X reads
- 4- 3Ddna scaffolding with 28X HiC reads (Maison)
- 5- Manual modification of HiC map
- 6- Polishing Pilon with 102X 10X reads

HiC Map

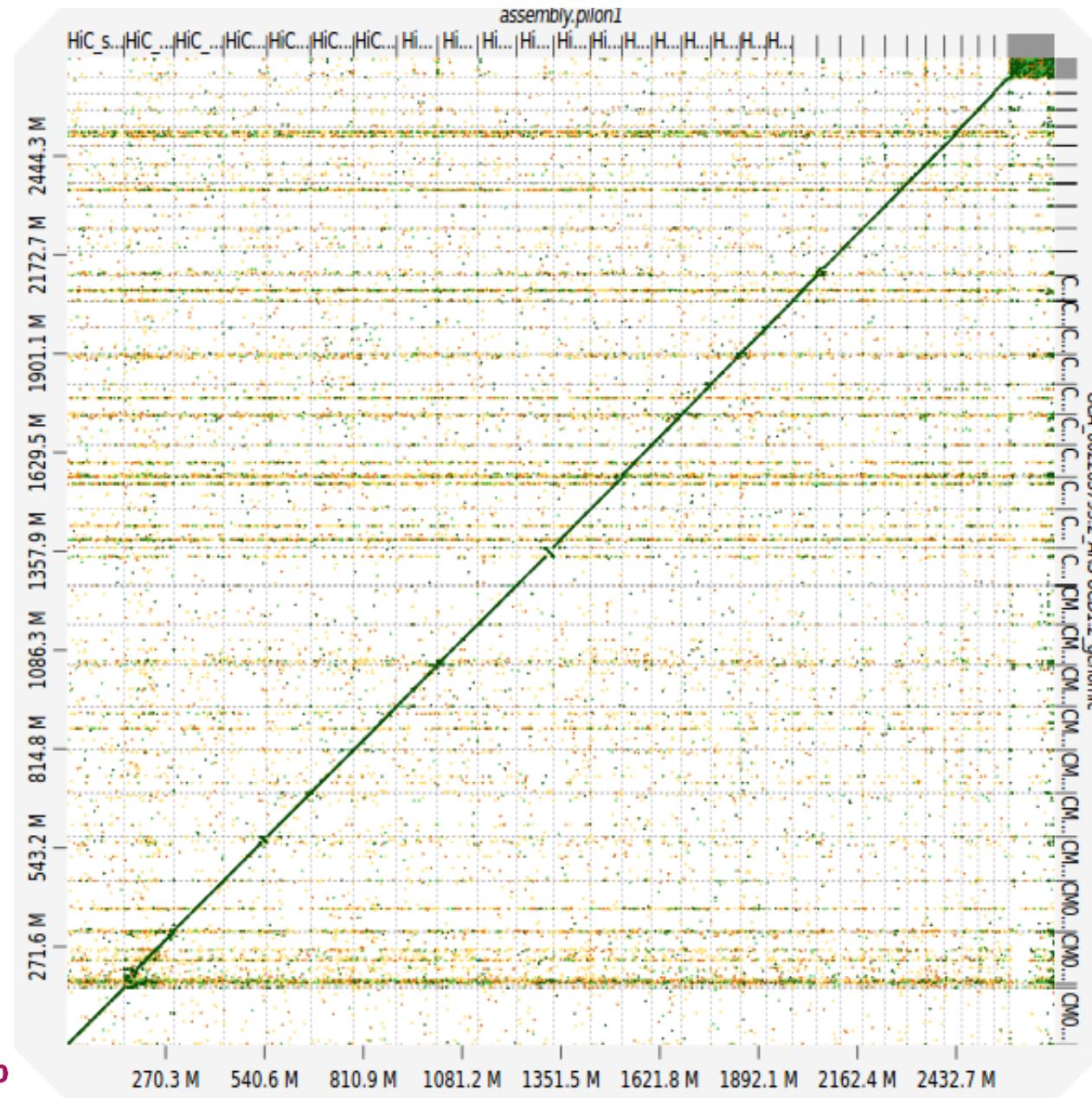


Final stats



Total size of scaffolds	Longest scaffold	N50/NG50 scaffold length	Busco
2 702 956 655	156 379 558	101 146 902 101 146 902	C:94.2% [S:93.0%,D:1.2%] F:3.1% M:2.7%

Alignment vs Bovine reference



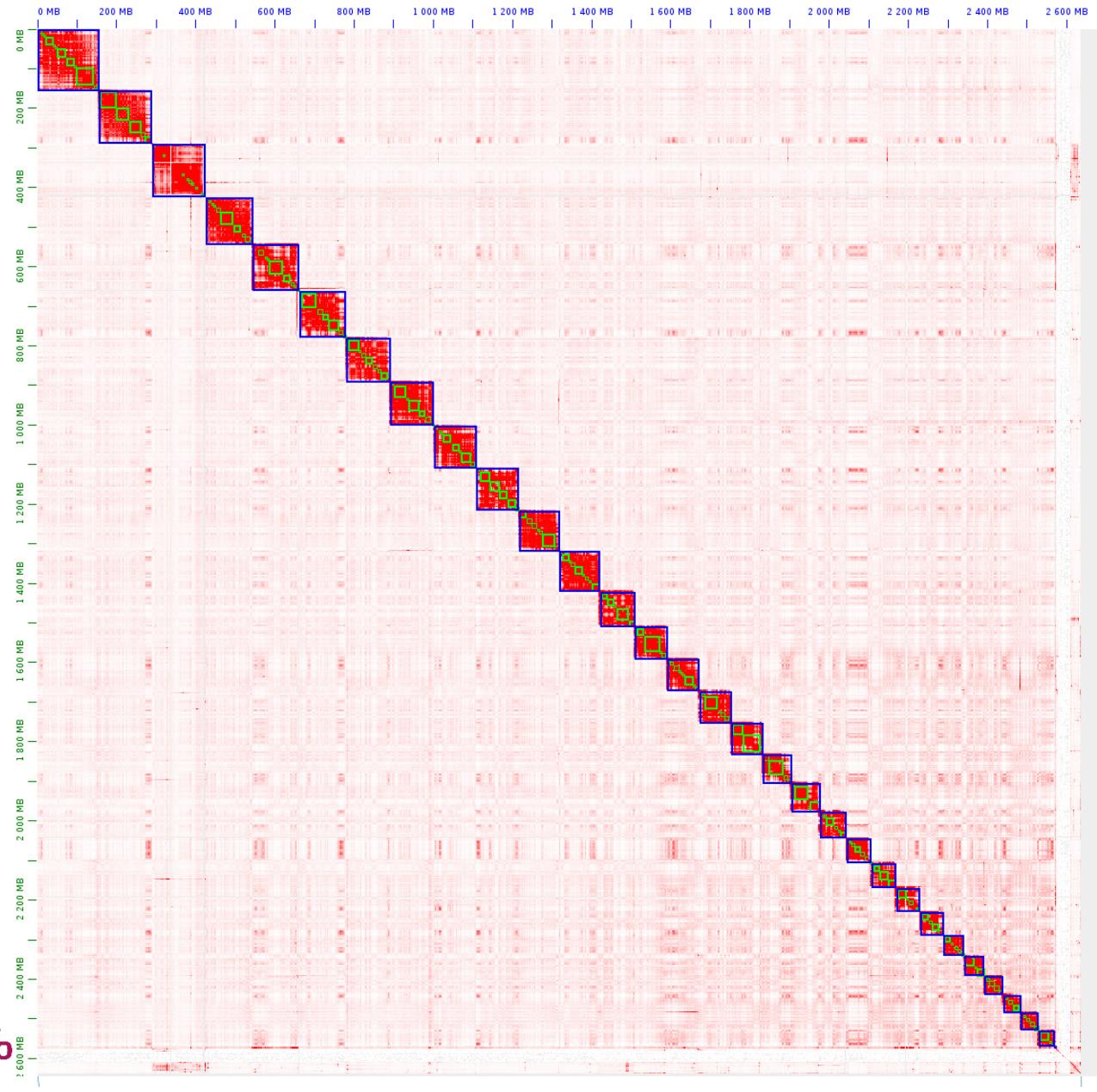
Mother 1 - 37165



Assembly protocol

- 1- Wtdbg2 contig assembly with 58X ONT reads
- 2- Polishing Racon with 58X ONT reads
- 3- Polishing Pilon with 84X 10X reads
- 4- 3Ddna scaffolding with 15X HiC reads (Maison)
- 5- Manual modification of HiC map
- 6- Gap-Filling TGSGapFiller with 58X ONT reads
- 7- Polishing Pilon with 84X 10X reads

HiC Map

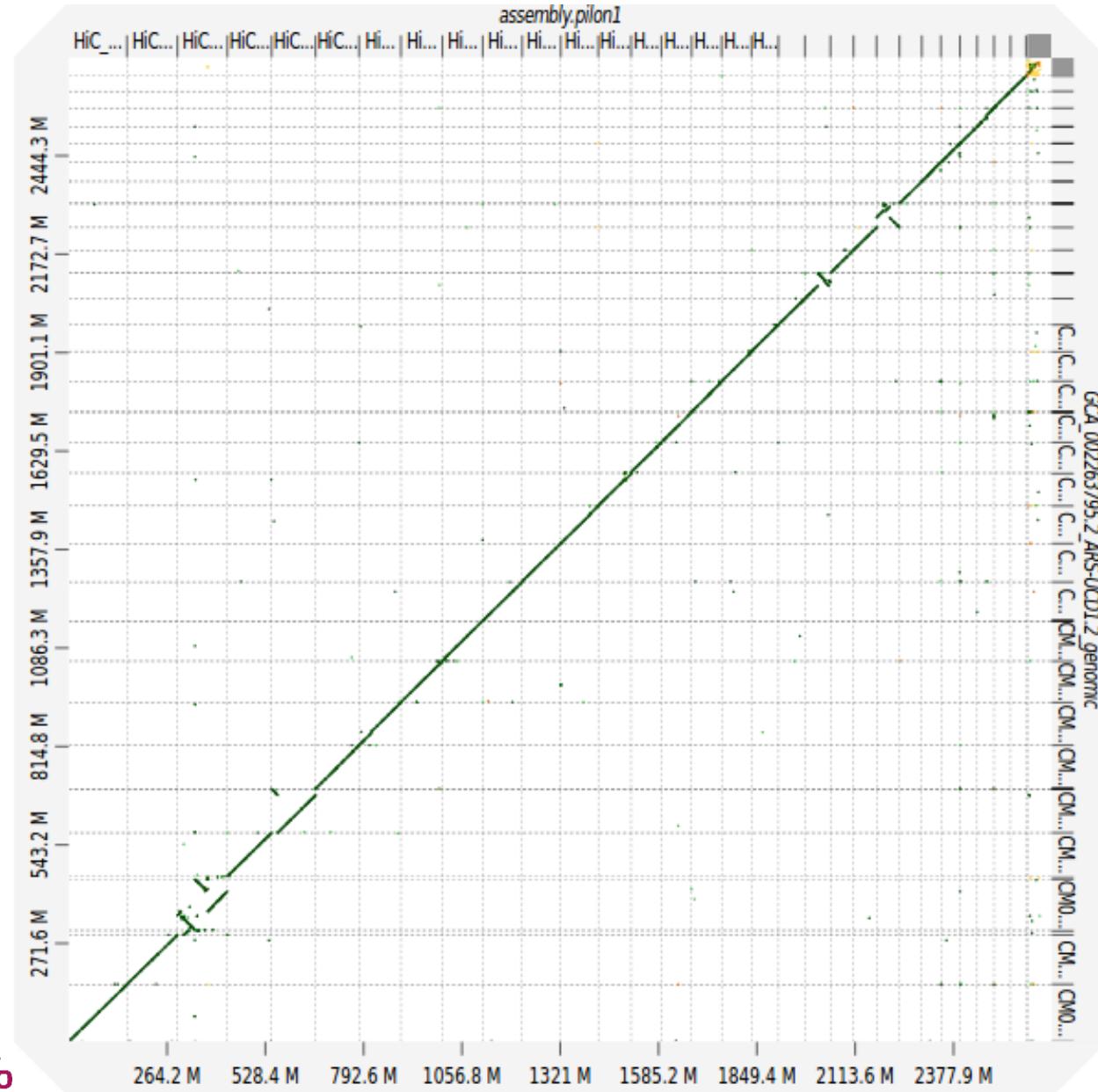


Final stats



Total size of scaffolds	Longest scaffold	N50/NG50 scaffold length	Busco
2 642 056 172	156 510 258	103 769 945 100 772 962	C:94.5% [S:93.3%,D:1.2%] F:3.1% M:2.4%

Alignment vs Bovine reference





Individu	Assembly	Polishing LR	Splitting	Polishing SR	Scaffolding	Gap Filling	Final Polishing
Mother1 37165							
Father1 37164							
Offspring1 37163							
Mother2 37162							
Father2 37161							
Offspring2 37160							
<u>Sus scrofa</u>							
<u>Coturnix japonica</u>							
Mais F2							
Mais MBS847							
Mais F252							
Zea Mais F4							

HiFi

Zea_mays :

Data :

MBS847 - 20X
F252 - 33 X

Bos_taurus :

Offspring2 - 15 X

Results :

Polishing MBS847 - 10X
- 7,4 time faster
- BUSCO 97.3%

Polishing - 8X
- 12.3 time faster
- BUSCO 94.3%

In progress :

Polishing MBS847 - 20X
Polishing F252 - 33X
Assembly MBS847 - 20X
Assembly F252 - 33X

Polishing - 15X
Assembly - 15 X



Juicer stats

	MBS847_123	MBS847_11/02	MBS847_28/02	MBS847_05/06	MBS847_11/02 VS NRGene
Sequenced Read Pairs	316 260 404	2 535 954	108 725 802	206 127 921	2 535 954
Normal Paired	222 026 529 (70.20%)	1 943 694 (76.65%)	83 501 291 (76.80%)	136 210 474 (66.08%)	1 945 168 (76.70%)
Chimeric Paired	36 989 807 (11.70%)	229 919 (9.07%)	9 669 251 (8.89%)	26 941 241 (13.07%)	247 364 (9.75%)
Chimeric Ambiguous	54 832 761 (17.34%)	350 380 (13.82%)	14 966 057 (13.76%)	41 175 340 (19.98%)	330 561 (13.03%)
Unmapped	2 411 307 (0.76%)	11 961 (0.47%)	589 203 (0.54%)	1 800 866 (0.87%)	12 893 (0.51%)
Ligation Motif Present	48 529 733 (15.34%)	98 927 (3.90%)	4 252 498 (3.91%)	44 178 308 (21.43%)	98 927 (3.90%)
Alignable (Normal+Chimeric Paired)	259 016 336 (81.90%)	2 173 613 (85.71%)	93 170 542 (85.69%)	163 151 715 (79.15%)	2 192 532 (86.46%)
Unique Reads	228 966 469	2 170 738	84 024 147	143 843 684	2 190 592
PCR Duplicates	27,196,060	2,333	7,526,575	18,068,515	1,385
Optical Duplicates	2,853,807	542	1,619,820	1,239,516	555
Library Complexity Estimate	1,119,420,279	1,011,328,143	525,841,611	670,409,536	1,733,840,804
Intra-fragment Reads	24,113,266 (7.62% / 10.53%)	360,573 (14.22% / 16.61%)	13,053,845 (12.01% / 15.54%)	10,920,483 (5.30% / 7.59%)	422,241 (16.65% / 19.28%)
Below MAPQ Threshold	103,632,634 (32.77% / 45.26%)	867,855 (34.22% / 39.98%)	35,476,630 (32.63% / 42.22%)	67,860,214 (32.92% / 47.18%)	817,512 (32.24% / 37.32%)
Hi-C Contacts	101,220,569 (32.01% / 44.21%)	942,310 (37.16% / 43.41%)	35,493,672 (32.65% / 42.24%)	65,062,987 (31.56% / 45.23%)	950,839 (37.49% / 43.41%)
Ligation Motif Present	10,347,569 (3.27% / 4.52%)	27,327 (1.08% / 1.26%)	1,025,697 (0.94% / 1.22%)	9,311,972 (4.52% / 6.47%)	28,293 (1.12% / 1.29%)
3' Bias (Long Range)	50% - 50%	44% - 56%	44% - 56%	52% - 48%	45% - 55%
Pair Type % (L-I-O-R)	25% - 25% - 25% - 25%	25% - 26% - 25% - 24%	24% - 26% - 26% - 24%	25% - 25% - 25% - 25%	25% - 25% - 25% - 25%
Inter-chromosomal	36,995,453 (11.70% / 16.16%)	299,943 (11.83% / 13.82%)	11,428,480 (10.51% / 13.60%)	25,278,778 (12.26% / 17.57%)	152,977 (6.03% / 6.98%)
Intra-chromosomal	64,225,116 (20.31% / 28.05%)	642,367 (25.33% / 29.59%)	24,065,192 (22.13% / 28.64%)	39,784,209 (19.30% / 27.66%)	797,862 (31.46% / 36.42%)
Short Range (<20Kb)	53,013,737 (16.76% / 23.15%)	593,146 (23.39% / 27.32%)	22,209,401 (20.43% / 26.43%)	30,494,478 (14.79% / 21.20%)	665,364 (26.24% / 30.37%)
Long Range (>20Kb)	11,205,961 (3.54% / 4.89%)	49,175 (1.94% / 2.27%)	1,853,626 (1.70% / 2.21%)	9,286,530 (4.51% / 6.46%)	132,436 (5.22% / 6.05%)

	Arima	Maison	Dovetail
Sequenced Read Pairs:	76 522 432	134 706 615	108 635 633
Normal Paired:	24 428 963	31.92%	75 149 462
Chimeric Paired:	34 531 814	45.13%	52 702 521
Chimeric Ambiguous:	16 968 467	22.17%	6 013 780
Ligation Motif Present:	35 449 898	46.33%	81 996 045
Alignable (Normal+Chimeric Paired):	58 960 777	77.05%	127 851 983
Unique Reads:	52 482 944	68.58 %	108 537 586
Intra-fragment Reads:	426 299	0.56% / 0.81%	5 487 648
Hi-C Contacts:	37 717 044	49.29% / 71.87%	93 428 392
Ligation Motif Present:	10 573 286	13.82% / 20.15%	39 584 527
Inter-chromosomal:	8 476 814	11.08% / 16.15%	32 342 195
Intra-chromosomal:	29 240 230	38.21% / 55.71%	61 086 197
Short Range (<20Kb):	14 679 645	19.18% / 27.97%	18 377 756
Long Range (>20Kb):	14 560 489	19.03% / 27.74%	42 708 399
		31.70% / 39.35%	31.70% / 39.35%
		15 519 997	14.29% / 22.43%