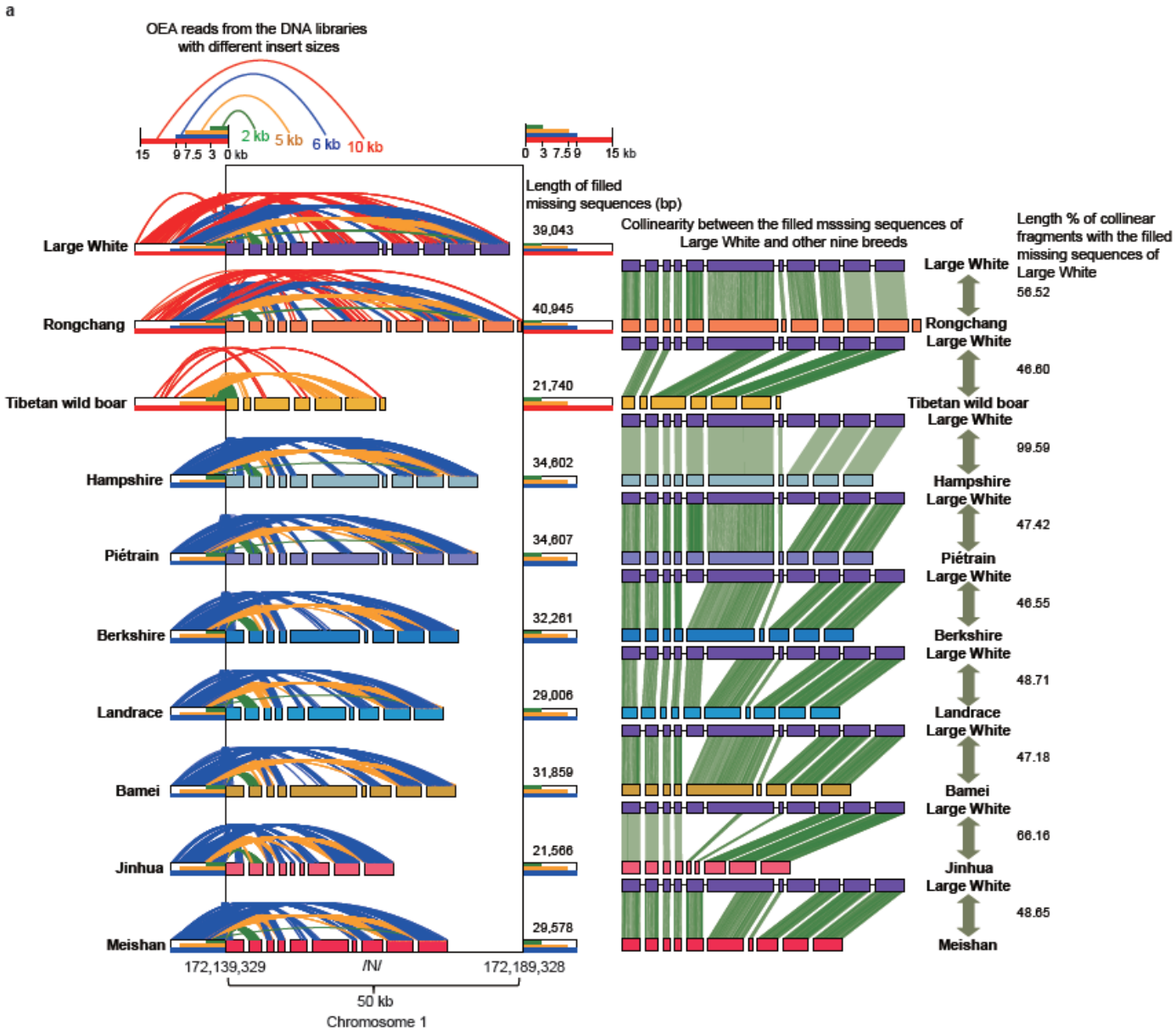
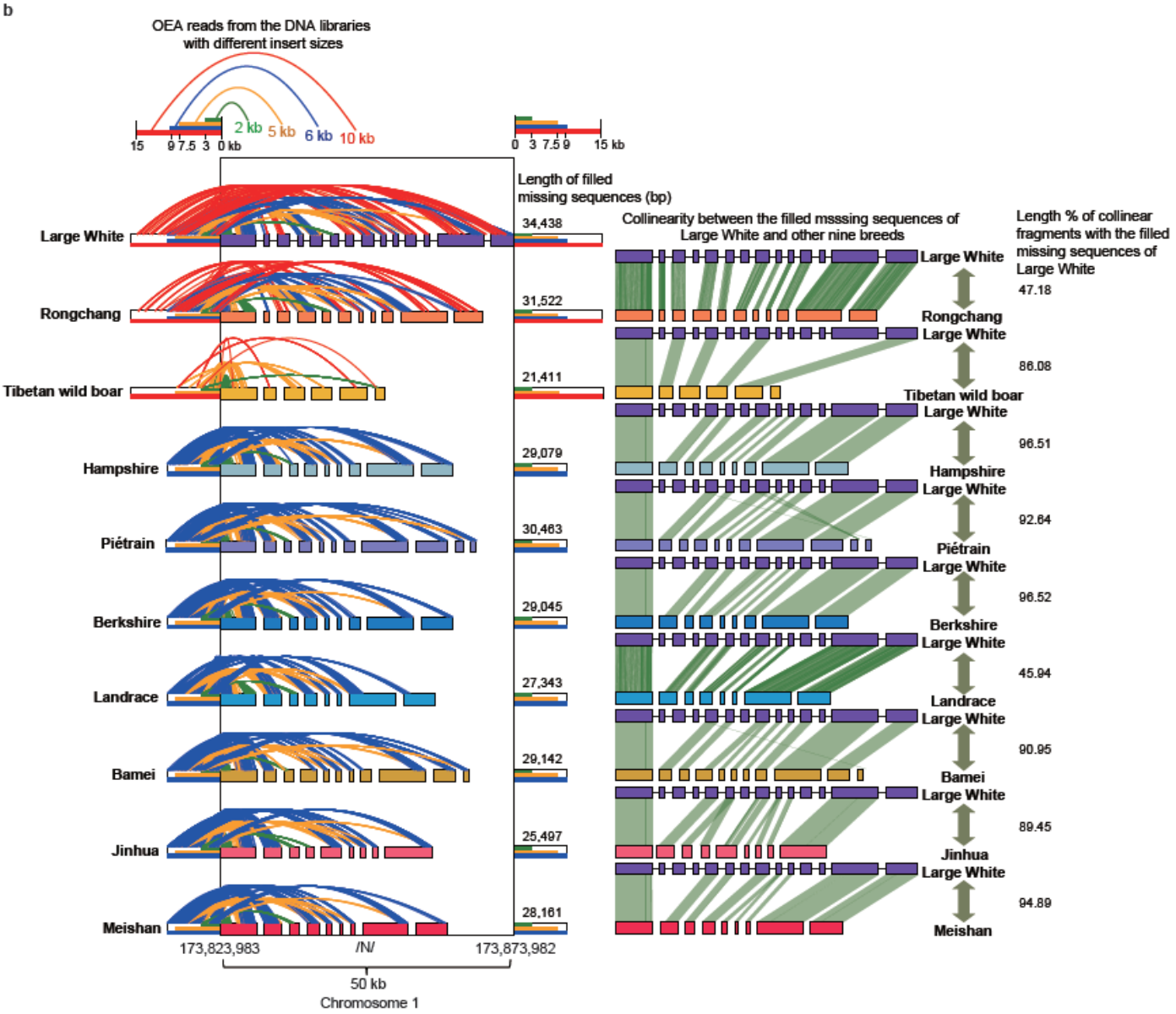
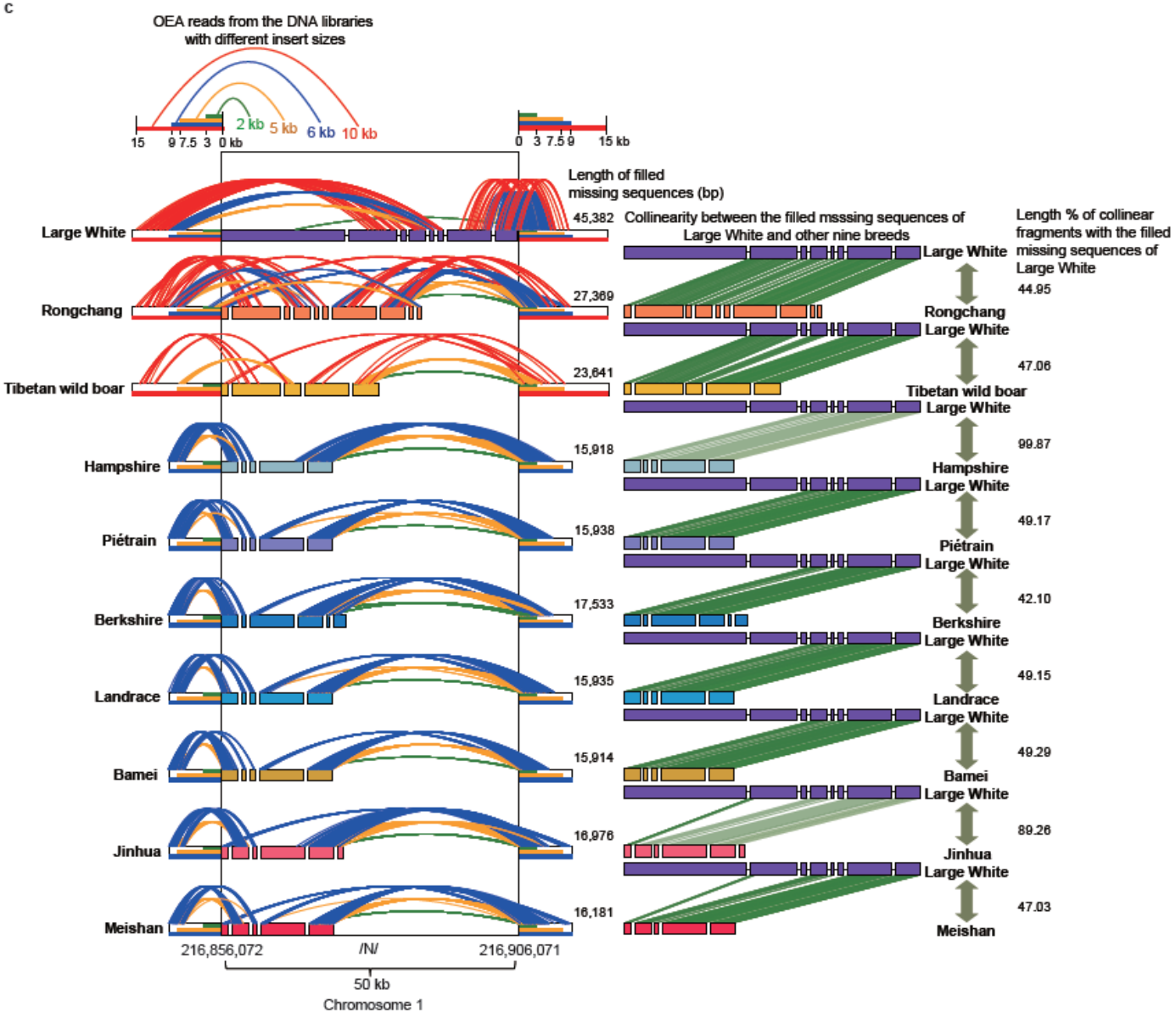
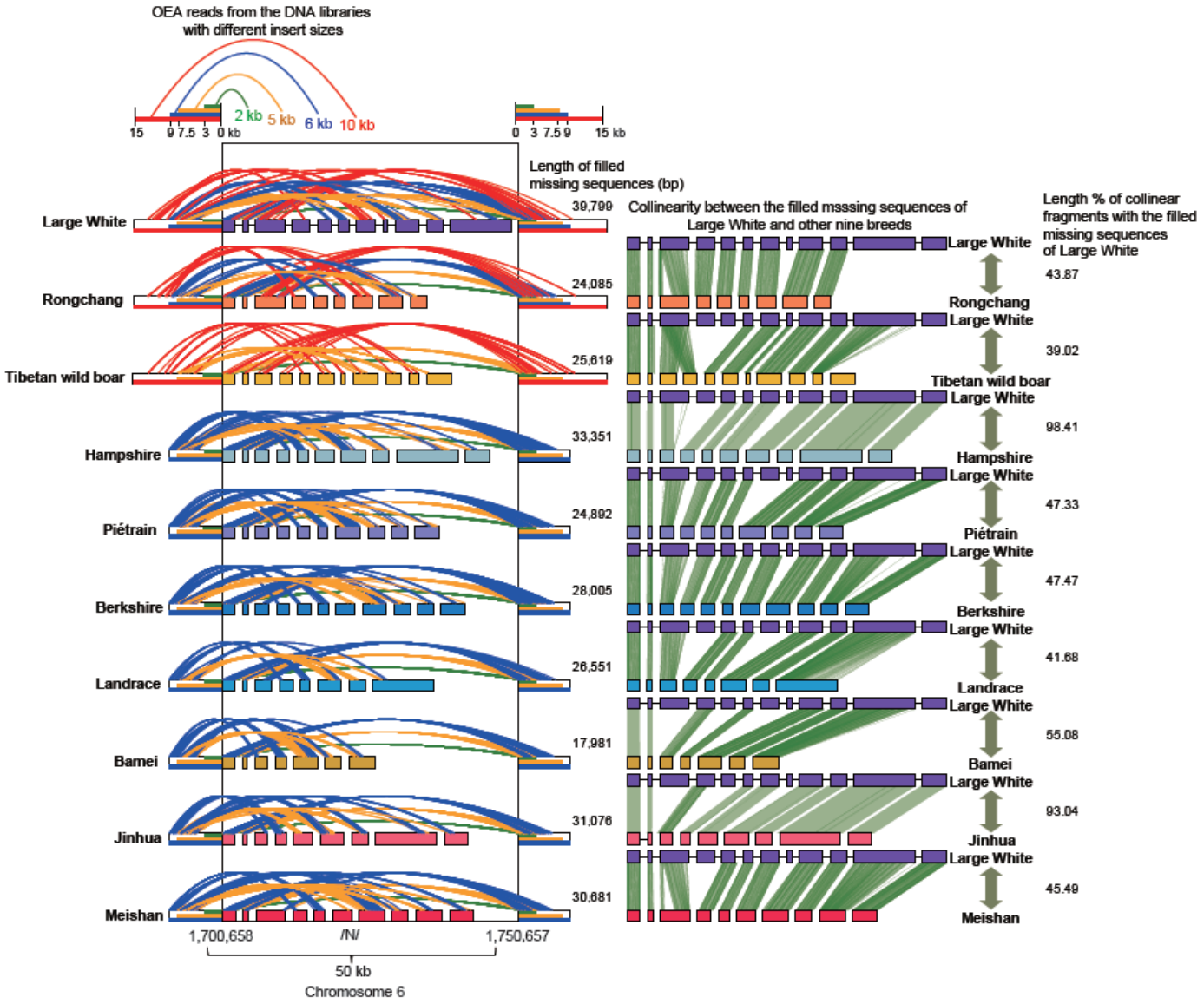

a
OEA reads from the DNA libraries
with different insert sizes
2 kb
5 kb
6 kb
10 kb
15 9 7.5 3 0 kb
0 3 7.5 9 15 kb
Length of filled
missing sequences (bp)
Collinearity between the filled msssing sequences of
Large White and other nine breeds
Length % of collinear
fragments with the filled
missing sequences of
Large White
Large White
Rongchang
Tibetan wild boar
Hampshire
Piétrain
Berkshire
Landrace
Bamei
Jinhua
Meishan
39,043
40,945
21,740
34,602
34,607
32,261
29,006
31,859
21,566
29,578
Large White
Rongchang
Large White
Tibetan wild boar
Large White
Hampshire
Large White
Piétrain
Large White
Berkshire
Large White
Landrace
Large White
Bamei
Large White
Jinhua
Large White
Meishan
56.52
46.60
99.59
47.42
46.55
48.71
47.18
66.16
48.65
172,139,329
/N/
172,189,328
50 kb
Chromosome 1


b
OEA reads from the DNA libraries with different insert sizes
2 kb
5 kb
6 kb
10 kb
15 9 7.5 3 0 kb
0 3 7.5 9 15 kb
Length of filled missing sequences (bp)
Large White
Rongchang
Tibetan wild boar
Hampshire
Piétrain
Berkshire
Landrace
Bamei
Jinhua
Meishan
34,438
31,522
21,411
29,079
30,463
29,045
27,343
29,142
25,497
28,161
173,823,983
/N/
173,873,982
50 kb
Chromosome 1
Collinearity between the filled msssing sequences of Large White and other nine breeds
Large White
Rongchang
Large White
Tibetan wild boar
Large White
Hampshire
Large White
Piétrain
Large White
Berkshire
Large White
Landrace
Large White
Bamei
Large White
Jinhua
Large White
Meishan
Length % of collinear fragments with the filled missing sequences of Large White
47.18
86.08
96.51
92.64
96.52
45.94
90.95
89.45
94.89

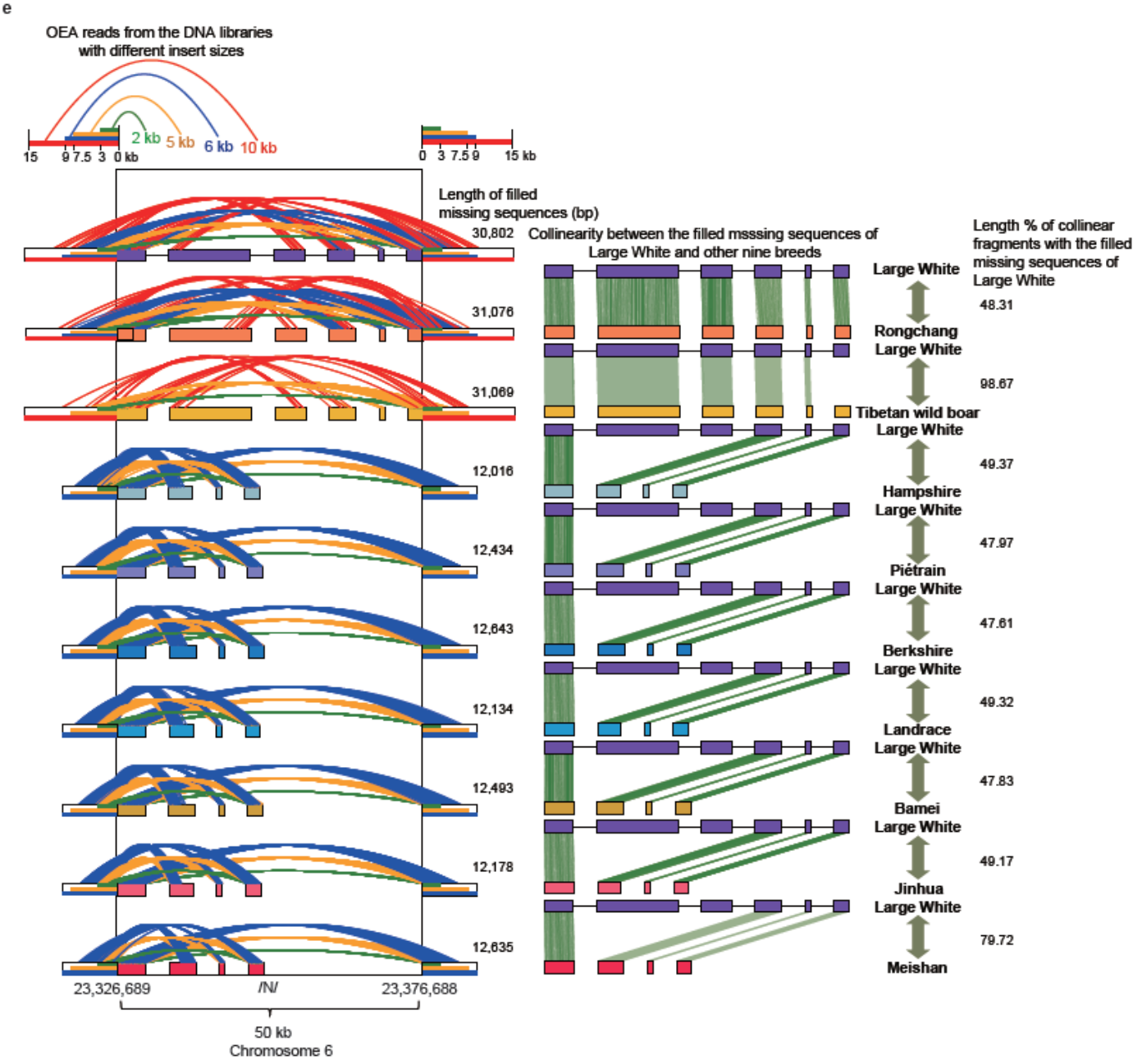
c


OEA reads from the DNA libraries with different insert sizes
2 kb
5 kb
6 kb
10 kb
15 9 7.5 3 0 kb
0 3 7.5 9 15 kb
Length of filled missing sequences (bp)
Large White
45,382
Rongchang
27,369
Tibetan wild boar
23,641
Hampshire
15,918
Piétrain
15,938
Berkshire
17,533
Landrace
15,935
Bamei
15,914
Jinhua
16,976
Meishan
16,181
216,856,072
/N/
216,906,071
50 kb
Chromosome 1
Collinearity between the filled msssing sequences of Large White and other nine breeds
Large White
Rongchang Large White
Tibetan wild boar Large White
Hampshire Large White
Piétrain Large White
Berkshire Large White
Landrace Large White
Bamei Large White
Jinhua Large White
Meishan
Length % of collinear fragments with the filled missing sequences of Large White
44.95
47.06
99.87
49.17
42.10
49.15
49.29
89.26
47.03

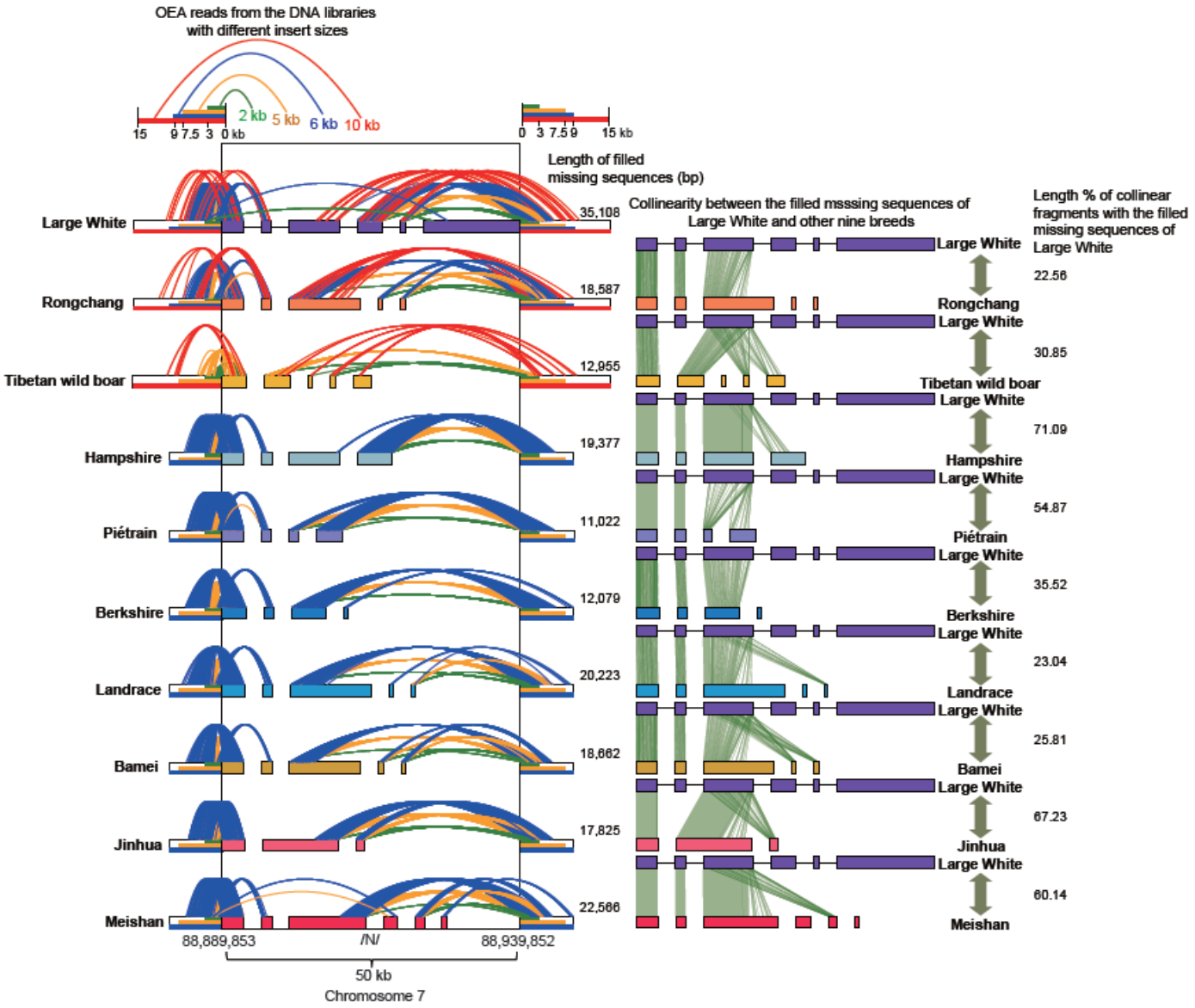
d


OEA reads from the DNA libraries with different insert sizes
2 kb
5 kb
6 kb
10 kb
15 9 7.5 3 0 kb
0 3 7.5 9 15 kb
Length of filled missing sequences (bp)
Large White
39,799
Rongchang
24,085
Tibetan wild boar
25,619
Hampshire
33,351
Piétrain
24,892
Berkshire
28,005
Landrace
26,551
Bamei
17,981
Jinhua
31,076
Meishan
30,681
1,700,658
/N/
1,750,657
50 kb
Chromosome 6
Collinearity between the filled msssing sequences of Large White and other nine breeds
Large White
Rongchang
Large White
Tibetan wild boar
Large White
Hampshire
Large White
Piétrain
Large White
Berkshire
Large White
Landrace
Large White
Bamei
Large White
Jinhua
Large White
Meishan
Length % of collinear fragments with the filled missing sequences of Large White
43.87
39.02
98.41
47.33
47.47
41.68
55.08
93.04
45.49

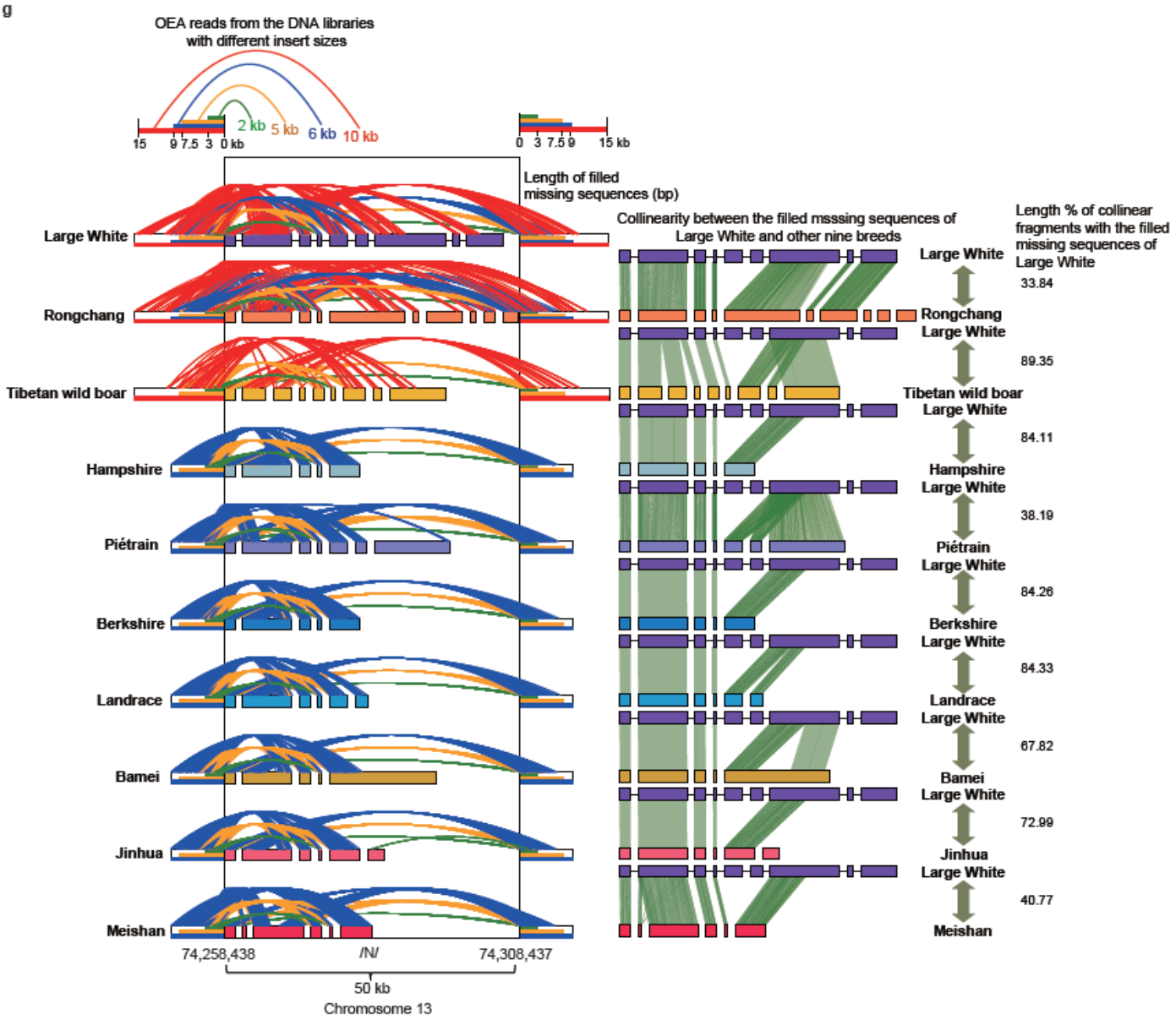
e


OEA reads from the DNA libraries with different insert sizes
2 kb
5 kb
6 kb
10 kb
15 9 7.5 3 0 kb
0 3 7.5 9 15 kb
Length of filled missing sequences (bp)
30,802
31,076
31,069
12,016
12,434
12,643
12,134
12,493
12,178
12,635
23,326,689
/N/
23,376,688
50 kb
Chromosome 6
Collinearity between the filled msssing sequences of Large White and other nine breeds
Large White
Rongchang
Large White
Tibetan wild boar
Large White
Hampshire
Large White
Piétrain
Large White
Berkshire
Large White
Landrace
Large White
Bamei
Large White
Jinhua
Large White
Meishan
Length % of collinear fragments with the filled missing sequences of Large White
48.31
98.67
49.37
47.97
47.61
49.32
47.83
49.17
79.72

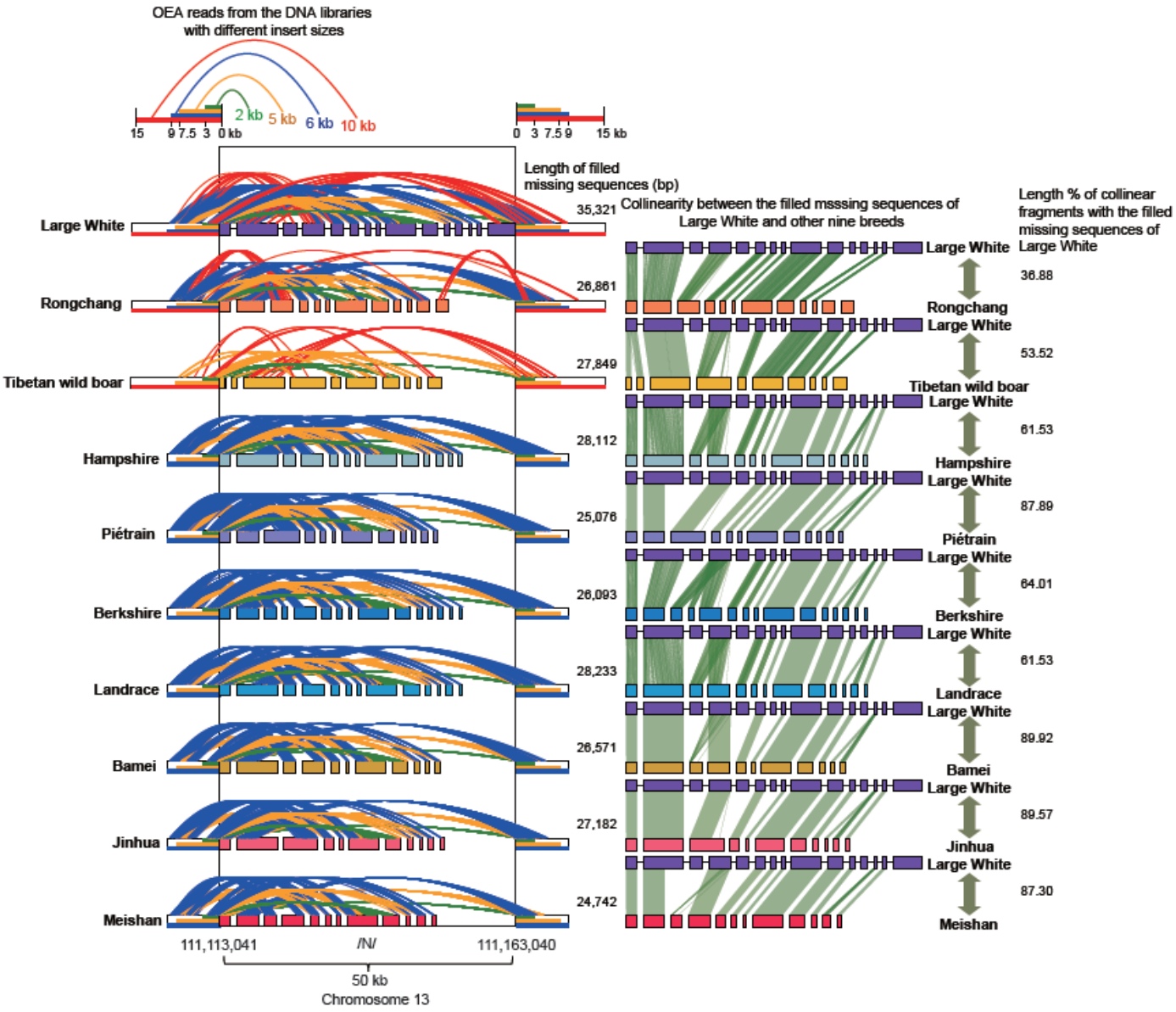
f


OEA reads from the DNA libraries with different insert sizes
2 kb
5 kb
6 kb
10 kb
15 9 7.5 3 0 kb
0 3 7.5 9 15 kb
Length of filled missing sequences (bp)
Collinearity between the filled msssing sequences of Large White and other nine breeds
Length % of collinear fragments with the filled missing sequences of Large White
Large White
35,108
Rongchang
18,587
Tibetan wild boar
12,955
Hampshire
19,377
Piétrain
11,022
Berkshire
12,079
Landrace
20,223
Bamei
18,662
Jinhua
17,825
Meishan
22,566
Large White
Rongchang
Large White
22.56
Tibetan wild boar
Large White
30.85
Hampshire
Large White
71.09
Piétrain
Large White
54.87
Berkshire
Large White
35.52
Landrace
Large White
23.04
Bamei
Large White
25.81
Jinhua
Large White
67.23
Meishan
60.14
88,889,853
/N/
88,939,852
50 kb
Chromosome 7

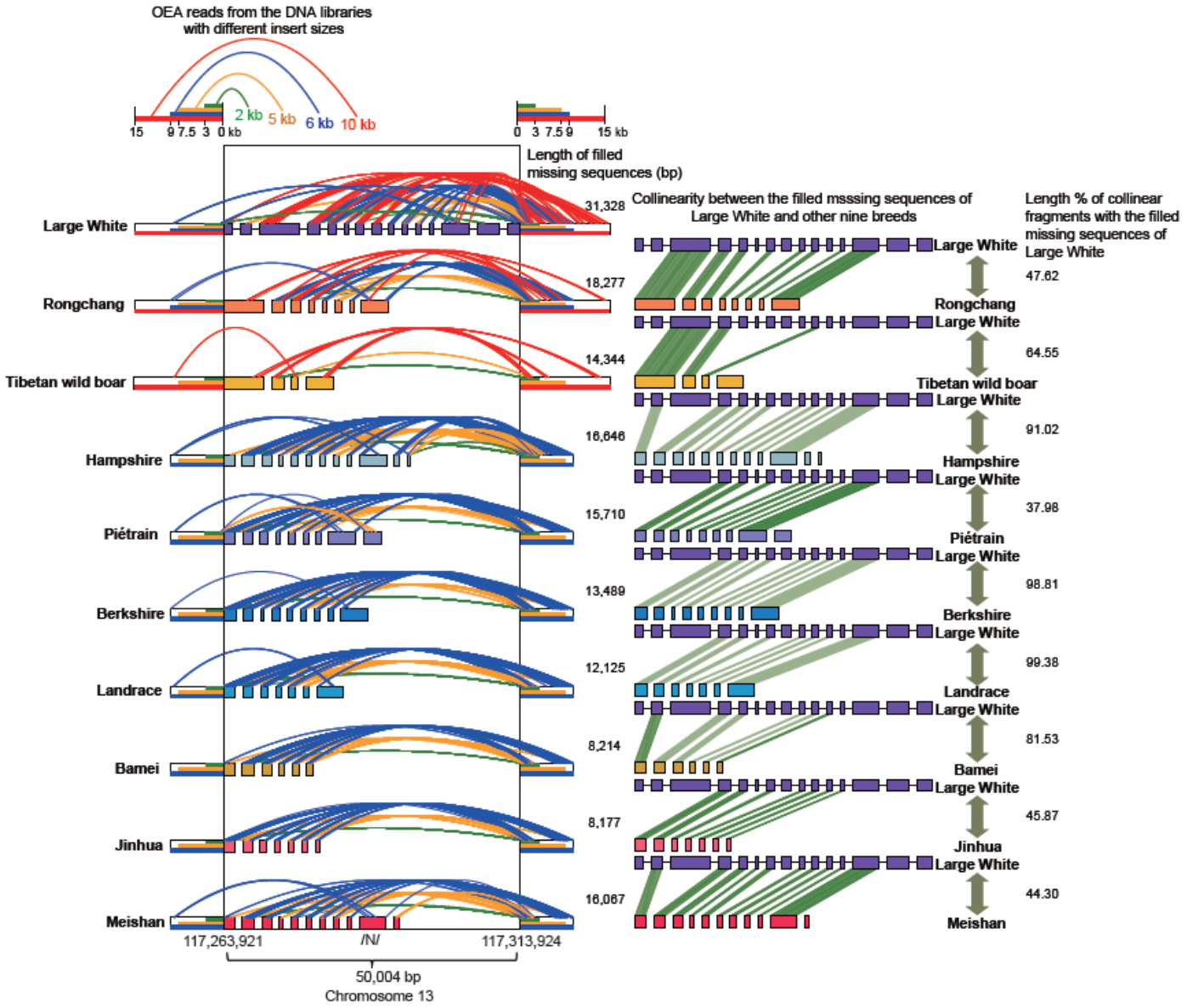
g


OEA reads from the DNA libraries with different insert sizes
2 kb
5 kb
6 kb
10 kb
15 9 7.5 3 0 kb
0 3 7.5 9 15 kb
Length of filled missing sequences (bp)
Collinearity between the filled msssing sequences of Large White and other nine breeds
Length % of collinear fragments with the filled missing sequences of Large White
Large White
Rongchang
Tibetan wild boar
Hampshire
Piétrain
Berkshire
Landrace
Bamei
Jinhua
Meishan
33.84
89.35
84.11
38.19
84.26
84.33
67.82
72.99
40.77
74,258,438
/N/
74,308,437
50 kb
Chromosome 13

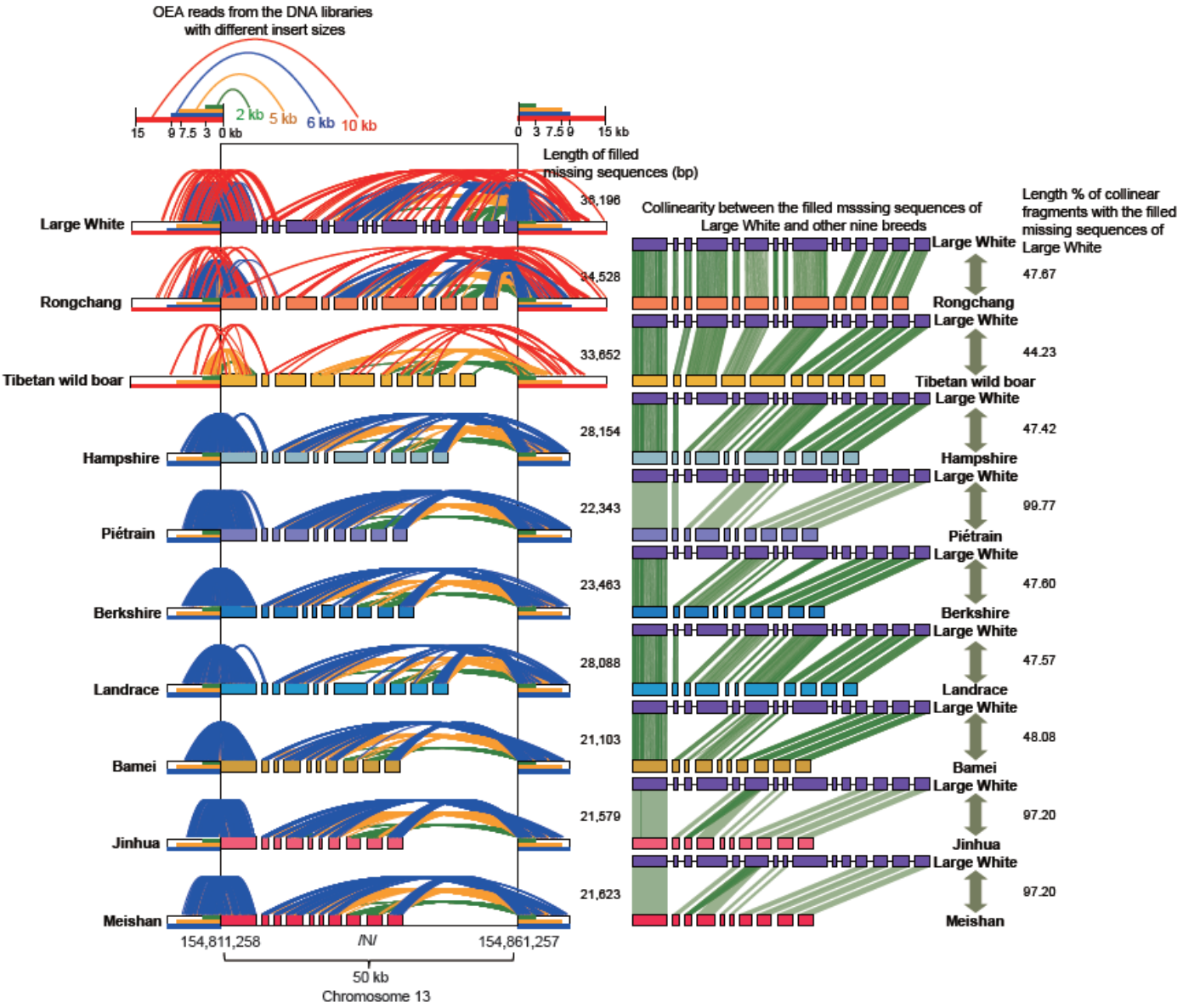
h


OEA reads from the DNA libraries with different insert sizes
2 kb
5 kb
6 kb
10 kb
15 9 7.5 3 0 kb
0 3 7.5 9 15 kb
Length of filled missing sequences (bp)
Large White
35,321
Rongchang
26,861
Tibetan wild boar
27,849
Hampshire
28,112
Piétrain
25,076
Berkshire
26,093
Landrace
28,233
Bamei
26,571
Jinhua
27,182
Meishan
24,742
111,113,041
/N/
111,163,040
50 kb
Chromosome 13
Collinearity between the filled msssing sequences of Large White and other nine breeds
Large White
Rongchang
Large White
Tibetan wild boar
Large White
Hampshire
Large White
Piétrain
Large White
Berkshire
Large White
Landrace
Large White
Bamei
Large White
Jinhua
Large White
Meishan
Length % of collinear fragments with the filled missing sequences of Large White
36.88
53.52
61.53
87.89
64.01
61.53
89.92
89.57
87.30

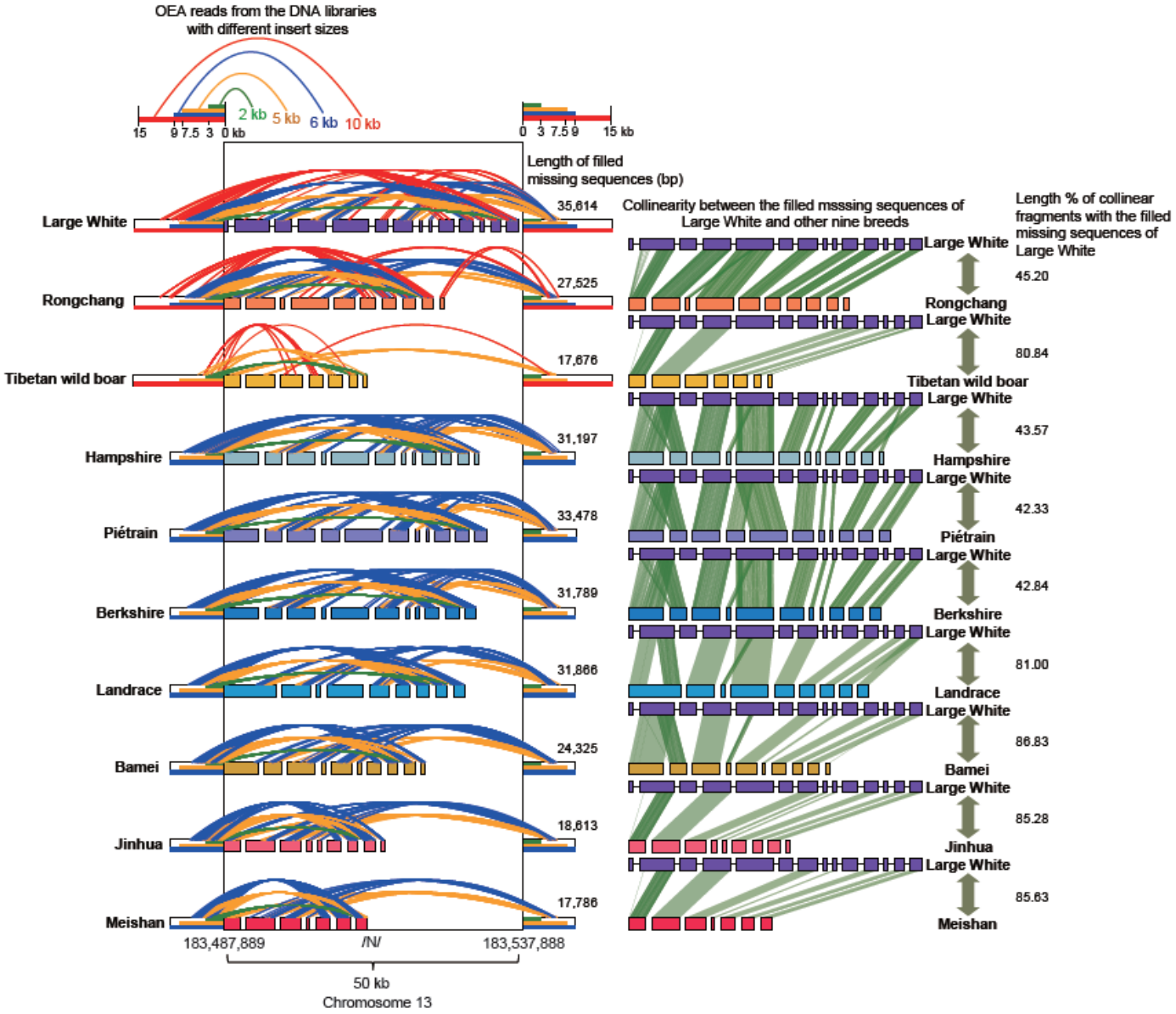
i


OEA reads from the DNA libraries with different insert sizes
2 kb
5 kb
6 kb
10 kb
15 9 7.5 3 0 kb
0 3 7.5 9 15 kb
Length of filled missing sequences (bp)
Large White
31,328
Rongchang
18,277
Tibetan wild boar
14,344
Hampshire
16,646
Piétrain
15,710
Berkshire
13,489
Landrace
12,125
Bamei
8,214
Jinhua
8,177
Meishan
16,067
117,263,921
/N/
117,313,924
50,004 bp
Chromosome 13
Collinearity between the filled msssing sequences of Large White and other nine breeds
Large White
Rongchang
Large White
Tibetan wild boar
Large White
Hampshire
Large White
Piétrain
Large White
Berkshire
Large White
Landrace
Large White
Bamei
Large White
Jinhua
Large White
Meishan
Length % of collinear fragments with the filled missing sequences of Large White
47.62
64.55
91.02
37.98
98.81
99.38
81.53
45.87
44.30

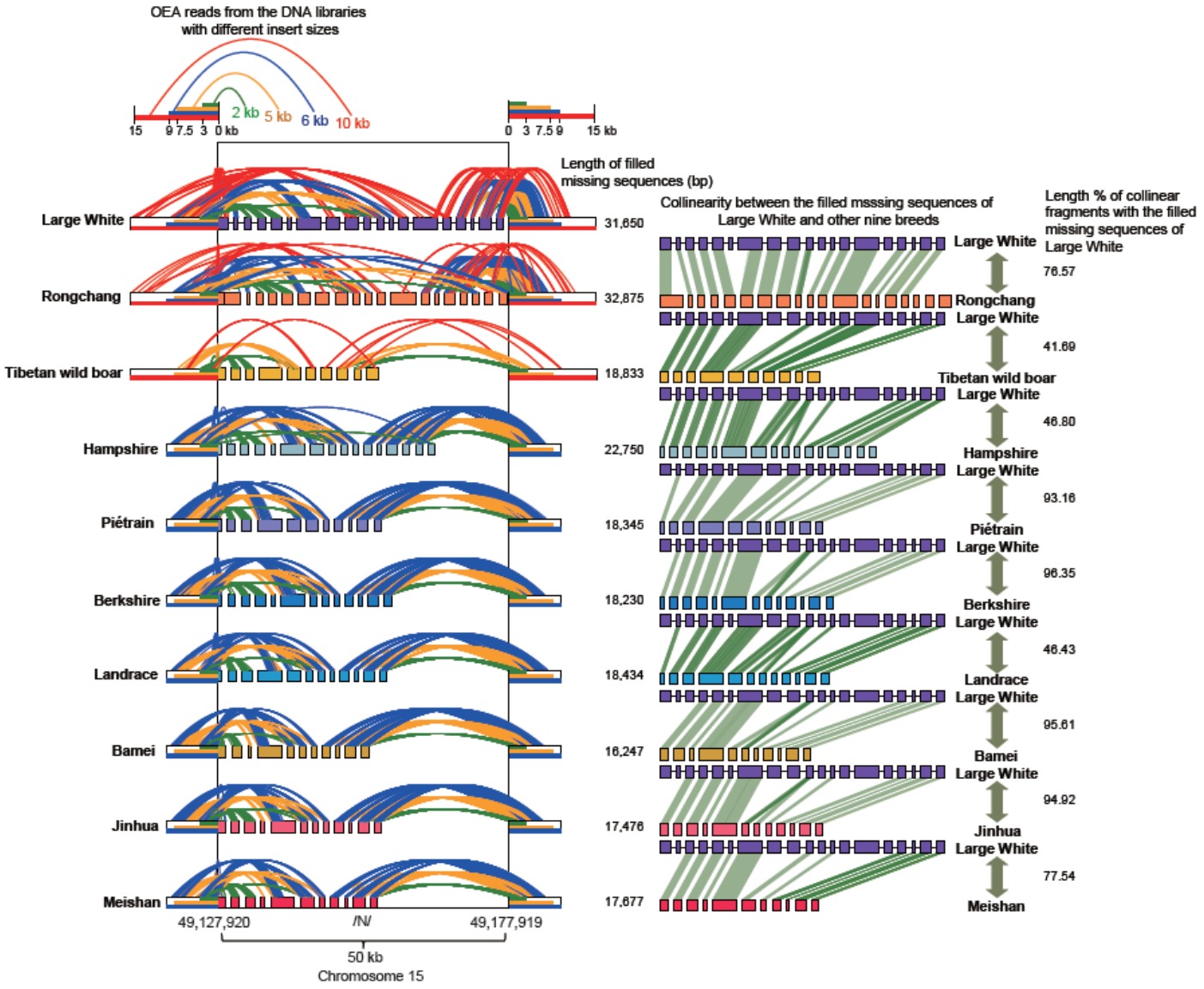
j


OEA reads from the DNA libraries with different insert sizes
2 kb
5 kb
6 kb
10 kb
15 9 7.5 3 0 kb
0 3 7.5 9 15 kb
Length of filled missing sequences (bp)
Large White
38,196
Rongchang
34,528
Tibetan wild boar
33,652
Hampshire
28,154
Piétrain
22,343
Berkshire
23,463
Landrace
28,088
Bamei
21,103
Jinhua
21,579
Meishan
21,623
154,811,258
/N/
154,861,257
50 kb
Chromosome 13
Collinearity between the filled msssing sequences of Large White and other nine breeds
Large White
Rongchang
Large White
Tibetan wild boar
Large White
Hampshire
Large White
Piétrain
Large White
Berkshire
Large White
Landrace
Large White
Bamei
Large White
Jinhua
Large White
Meishan
Length % of collinear fragments with the filled missing sequences of Large White
47.67
44.23
47.42
99.77
47.60
47.57
48.08
97.20
97.20

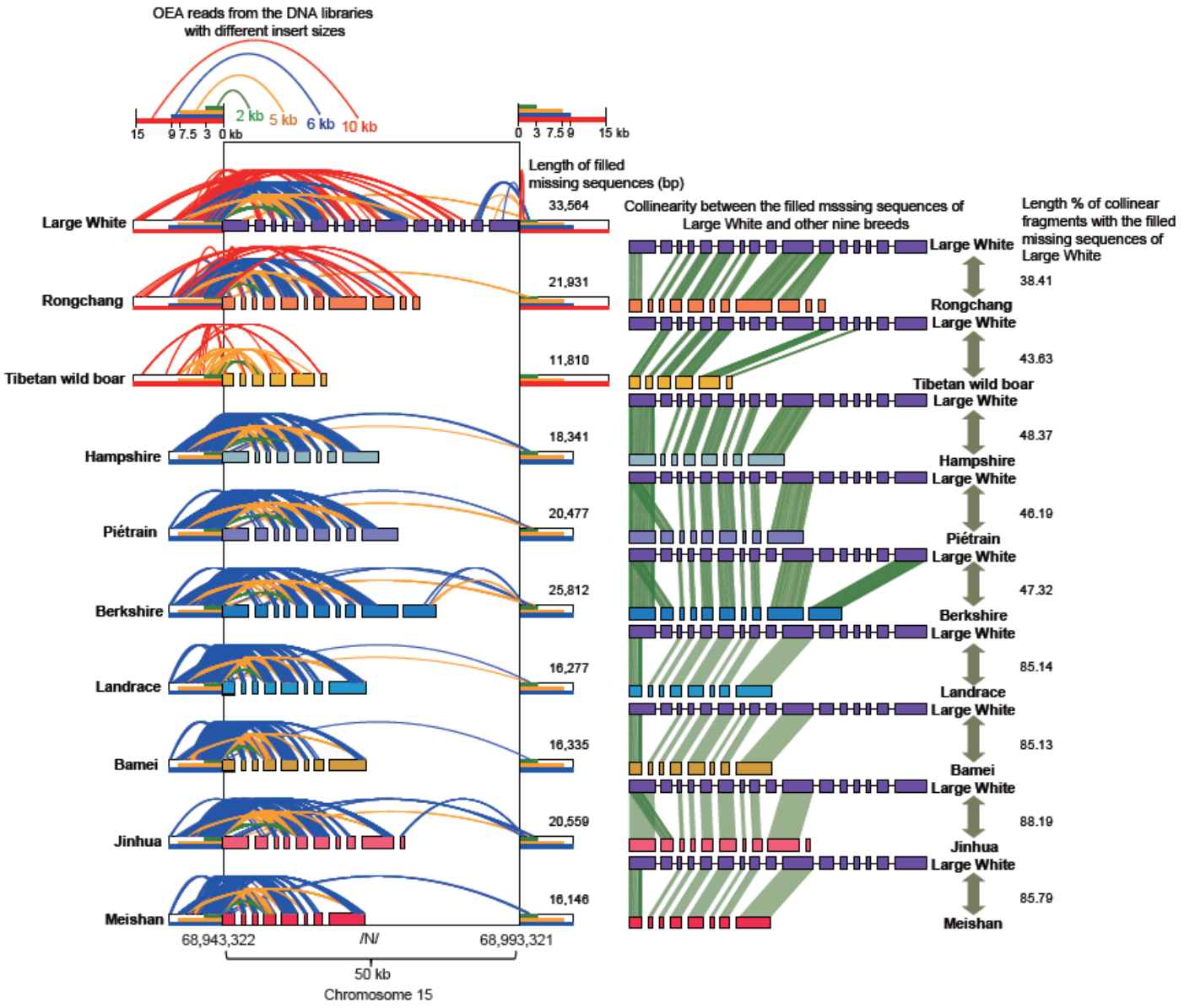
k


OEA reads from the DNA libraries with different insert sizes
2 kb
5 kb
6 kb
10 kb
15
9 7.5 3 0 kb
0 3 7.5 9 15 kb
Length of filled missing sequences (bp)
Large White
35,614
Rongchang
27,525
Tibetan wild boar
17,676
Hampshire
31,197
Piétrain
33,478
Berkshire
31,789
Landrace
31,866
Bamei
24,325
Jinhua
18,613
Meishan
17,786
183,487,889
/N/
183,537,888
50 kb
Chromosome 13
Collinearity between the filled msssing sequences of Large White and other nine breeds
Large White
Rongchang
Large White
Tibetan wild boar
Large White
Hampshire
Large White
Piétrain
Large White
Berkshire
Large White
Landrace
Large White
Bamei
Large White
Jinhua
Large White
Meishan
Length % of collinear fragments with the filled missing sequences of Large White
45.20
80.84
43.57
42.33
42.84
81.00
86.83
85.28
85.63

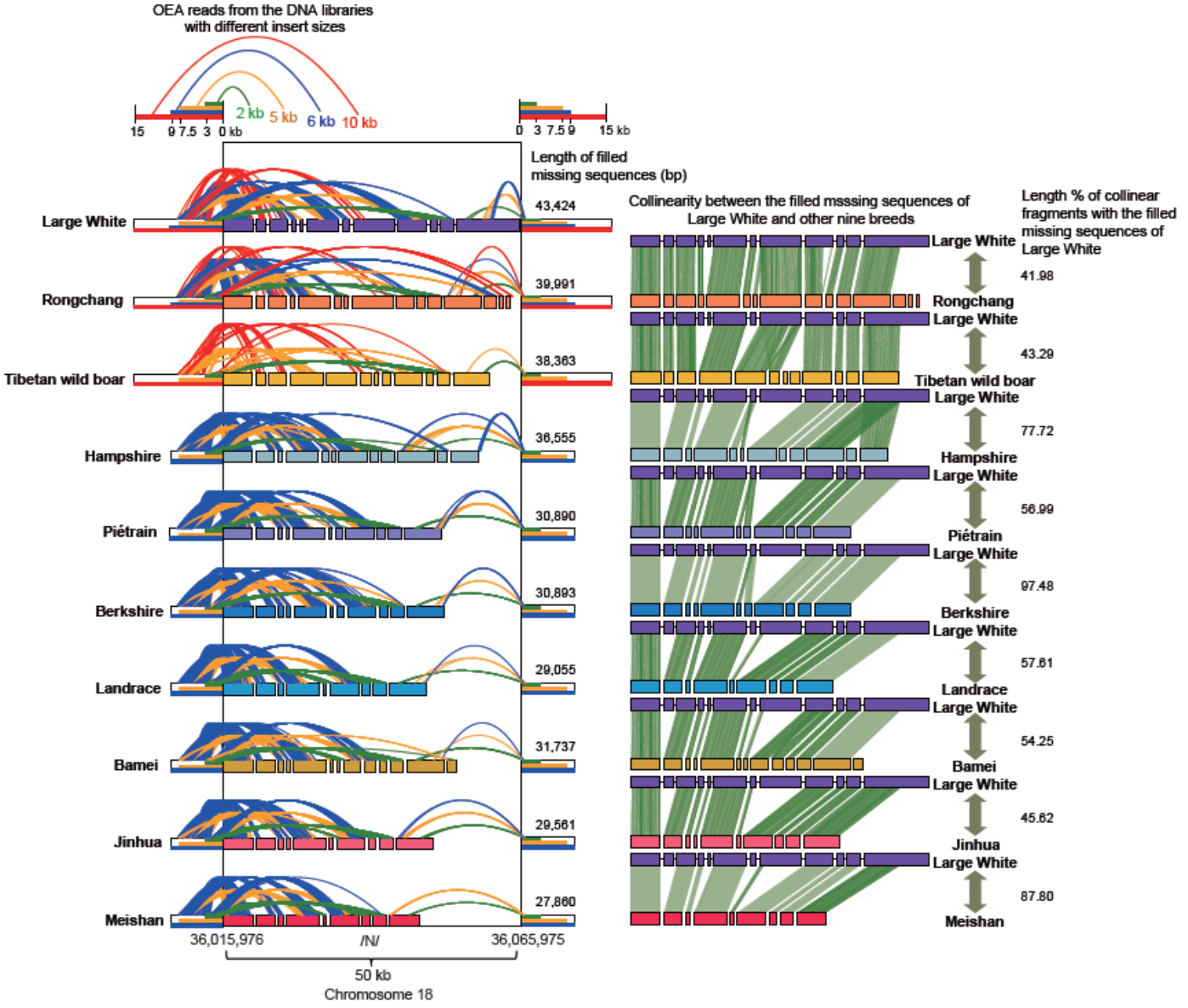
I

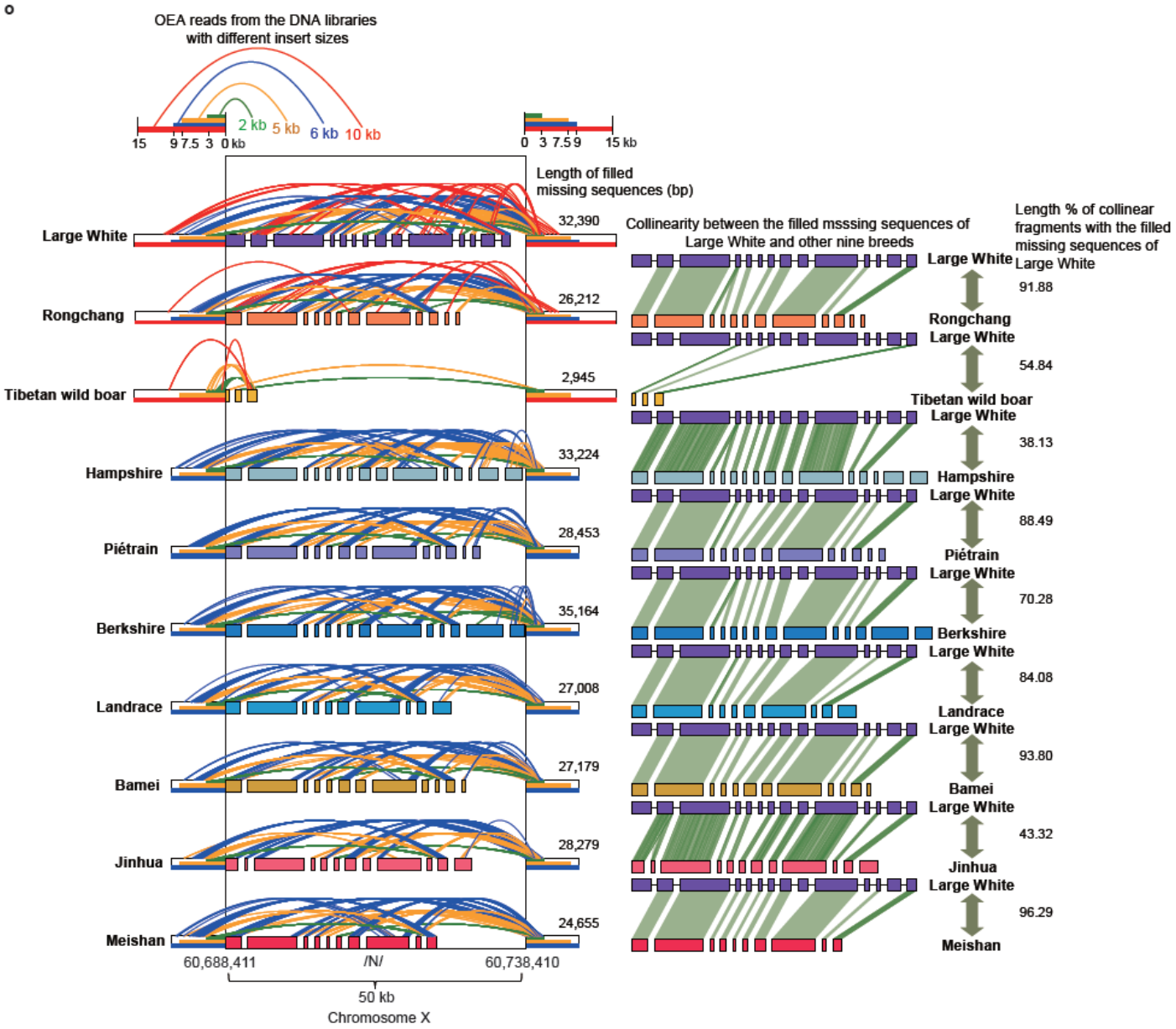

OEA reads from the DNA libraries with different insert sizes
2 kb
5 kb
6 kb
10 kb
15
9 7.5 3 0 kb
0 3 7.5 9 15 kb
Length of filled missing sequences (bp)
Large White
Rongchang
Tibetan wild boar
Hampshire
Piétrain
Berkshire
Landrace
Bamei
Jinhua
Meishan
31,650
32,875
18,833
22,750
18,345
18,230
18,434
16,247
17,476
17,677
49,127,920
/N/
49,177,919
50 kb
Chromosome 15
Collinearity between the filled msssing sequences of Large White and other nine breeds
Length % of collinear fragments with the filled missing sequences of Large White
Large White
Rongchang Large White
76.57
Tibetan wild boar Large White
41.69
Hampshire Large White
46.80
Piétrain Large White
93.16
Berkshire Large White
96.35
Landrace Large White
46.43
Bamei Large White
95.61
Jinhua Large White
94.92
Meishan
77.54

m


OEA reads from the DNA libraries with different insert sizes
2 kb
5 kb
6 kb
10 kb
15 9 7.5 3 0 kb
0 3 7.5 9 15 kb
Length of filled missing sequences (bp)
Large White
33,564
Rongchang
21,931
Tibetan wild boar
11,810
Hampshire
18,341
Piétrain
20,477
Berkshire
25,812
Landrace
16,277
Bamei
16,335
Jinhua
20,559
Meishan
16,146
68,943,322
/N/
68,993,321
50 kb
Chromosome 15
Collinearity between the filled msssing sequences of Large White and other nine breeds
Large White
Rongchang
Large White
Tibetan wild boar
Large White
Hampshire
Large White
Piétrain
Large White
Berkshire
Large White
Landrace
Large White
Bamei
Large White
Jinhua
Large White
Meishan
Length % of collinear fragments with the filled missing sequences of Large White
38.41
43.63
48.37
46.19
47.32
85.14
85.13
88.19
85.79

n


OEA reads from the DNA libraries with different insert sizes
2 kb
5 kb
6 kb
10 kb
15 9 7.5 3 0 kb
0 3 7.5 9 15 kb
Length of filled missing sequences (bp)
Large White
43,424
Rongchang
39,991
Tibetan wild boar
38,363
Hampshire
36,555
Piétrain
30,890
Berkshire
30,893
Landrace
29,055
Bamei
31,737
Jinhua
29,561
Meishan
27,860
36,015,976
/N/
36,065,975
50 kb
Chromosome 18
Collinearity between the filled msssing sequences of Large White and other nine breeds
Length % of collinear fragments with the filled missing sequences of Large White
Large White
41.98
Rongchang
Large White
43.29
Tibetan wild boar
Large White
77.72
Hampshire
Large White
56.99
Piétrain
Large White
97.48
Berkshire
Large White
57.61
Landrace
Large White
54.25
Bamei
Large White
45.62
Jinhua
Large White
87.80
Meishan



Supplemental Fig. S32. Examples of missing sequences filled in existing reference assembly gaps (Sscrofa10.2). **(a)** chromosome 1 (172,139,329-172,189,328); **(b)** chromosome 1 (173,823,983-173,873,982); **(c)** chromosome 1 (216,856,072-216,906,071); **(d)** chromosome 6 (1,700,658-1,750,657); **(e)** chromosome 6 (23,326,689-23,376,688); **(f)** chromosome 7 (88,889,853-88,939,852); **(g)** chromosome 13 (74,258,438-74,308,437); **(h)** chromosome 13 (111,113,041-111,163,040); **(i)** chromosome 13 (117,263,921-117,313,924); **(j)** chromosome 13 (154,811,258-154,861,257); **(k)** chromosome 13 (183,487,889-183,537,888); **(l)** chromosome 15 (49,127,920-49,177,919); **(m)** chromosome 15 (68,943,322-68,993,321); **(n)** chromosome 18 (36,015,976-36,065,975); and **(o)** chromosome X (60,688,411-60,738,410). Left panels: OEA reads against the reference genome from DNA libraries with long-insert sizes of 2 kb (green curves), 5 kb (golden curves), 6 kb (blue curves) and 10 kb (red curves) for each breed were aligned to their respective missing sequences and the reference assembly gaps. We filled a missing sequence to a gap (≥ 50 kb) that was anchored by the most of effectively anchored OEA reads (≥ 4), where one end of the pair was anchored to the missing sequences, and the other one was anchored to the flanking sequence at either end of a gap (maximal 1.5 folds of the library insert size, i.e., 3, 7.5, 9 and 15 kb in length for 2, 5, 6 and 10 kb insert size libraries, respectively). The missing sequences of Large White assembly from a scaffold were linked by black lines. Right panels: Multiple sequence alignment of missing sequences filled in gaps across ten breeds by an all-against-all BLASTn with the e-value of 1e-5. The fragmental

alignments were conjoined by Solar. The length percentage (%) of the collinear fragments with the filled missing sequences of Large White assembly were also listed for other nine breeds.