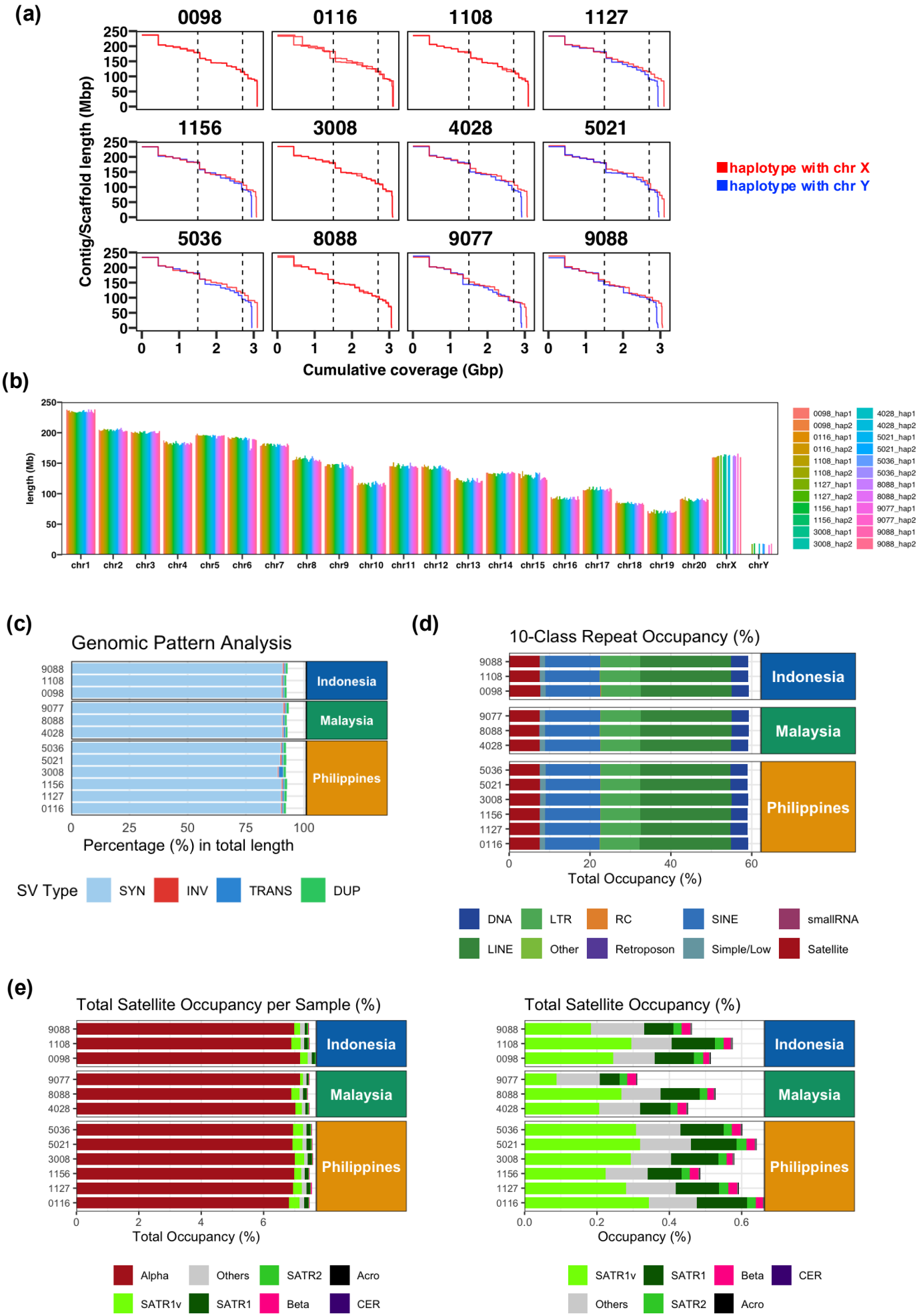
Supplementary Figure.1



Supplementary Figure.1

a, Individual cumulative coverage profiles of de novo haplotype-resolved contig assemblies. Cumulative coverage curves are shown separately for each individual, prior to reference-guided scaffolding. Contigs derived from haplotype-resolved de novo assembly (Verkko) were ordered by decreasing length, and cumulative assembly span (Gbp) was plotted against contig length (Mbp). Red and blue curves represent haplotypes containing chromosome X or chromosome Y, respectively. NG50 and NG90 were calculated relative to the haploid genome size of T2T-MFA8v1.1 (3,060,038,958 bp). These profiles demonstrate that chromosome-scale contiguity was achieved through reference-independent assembly across all individuals.

b, Chromosome-level scaffold length consistency across haplotypes following reference-guided scaffolding. Chromosome-scale scaffold lengths are shown for all 24 haplotypes after RagTag-based reference-guided scaffolding using T2T-MFA8v1.1 as a structural guide. For each chromosome, scaffold lengths are plotted across individuals and haplotypes. The limited variance in chromosome-scale scaffold lengths indicates consistent chromosomal representation across assemblies and supports that reference-guided scaffolding standardized ordering and orientation without introducing large-scale structural distortion.

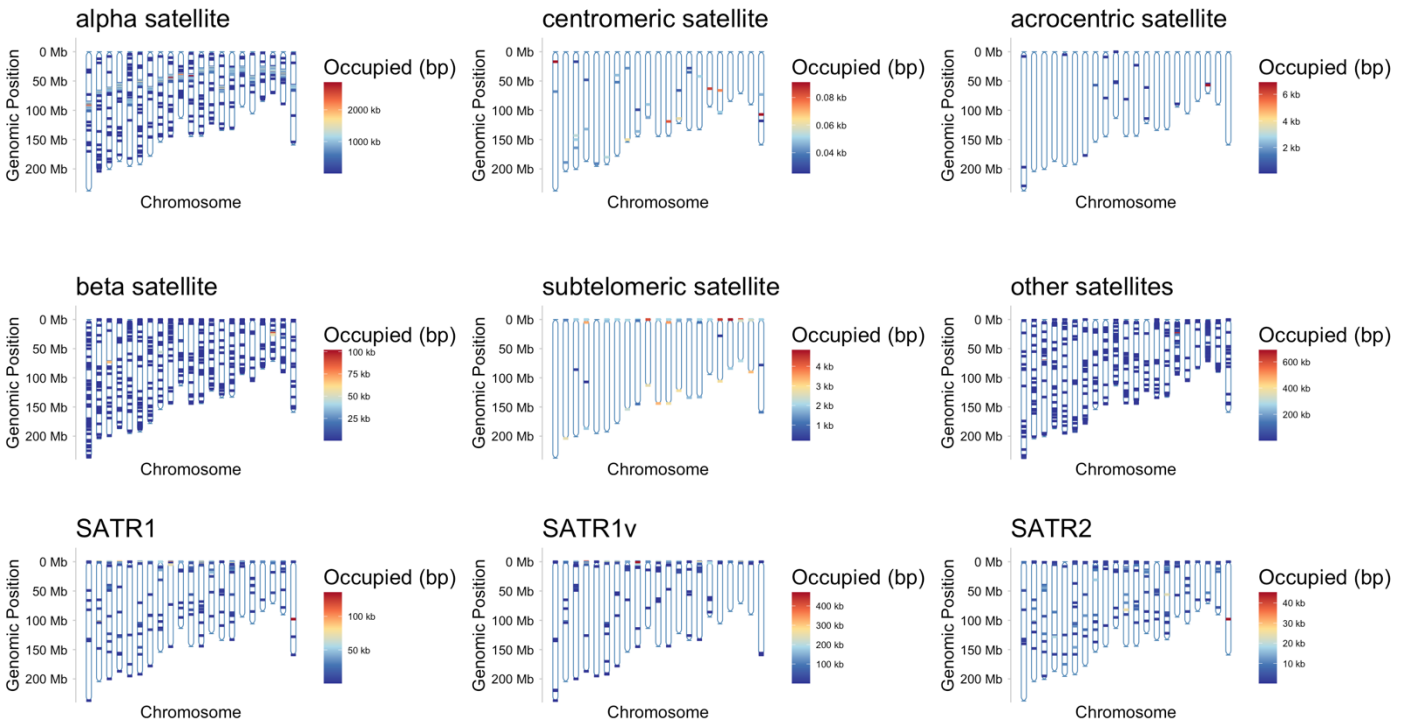
c, Genome-wide structural concordance between haplotype assemblies and T2T-MFA8v1.1.Whole-genome structural comparisons were performed using SyRI between each haplotype assembly and the T2T-MFA8v1.1 reference genome. The proportion of aligned sequence classified as syntenic (SYN), inversion (INV), translocation (TRANS), or duplication (DUP) is shown as a percentage of total aligned length. Across all haplotypes and geographic origins, the majority of genomic sequence is classified as syntenic, indicating high large-scale structural concordance while preserving haplotype-specific structural variation.

d, Genome-wide repeat composition across haplotypes.Repeat content was annotated using RepeatMasker and aggregated into ten major repeat classes (DNA transposons, LINEs, SINEs, LTR elements, RC elements, satellites, simple/low complexity repeats, small RNAs, and other categories). The proportion of each repeat class relative to total genome length is shown for all 24 haplotypes grouped by geographic origin. Overall repeat class composition is broadly consistent across individuals and populations, indicating that genome-wide repeat content is stable and does not exhibit large-scale compositional bias between assemblies.

e. Genome-wide satellite repeat occupancy and subtype composition across haplotypes. Left, total satellite repeat occupancy per haplotype as a percentage of total genome length, subdivided into major satellite categories (Alpha, SATR1v, SATR1, SATR2, Beta, Acro, CER, and others). Right, relative contribution of major SATR1-related subclasses within the satellite fraction for each haplotype. Values are shown for all individuals grouped by geographic origin. Overall satellite occupancy remains broadly consistent across individuals, whereas subtype composition shows moderate variation without clear population-level separation.

Supplementary Figure.2

(a)

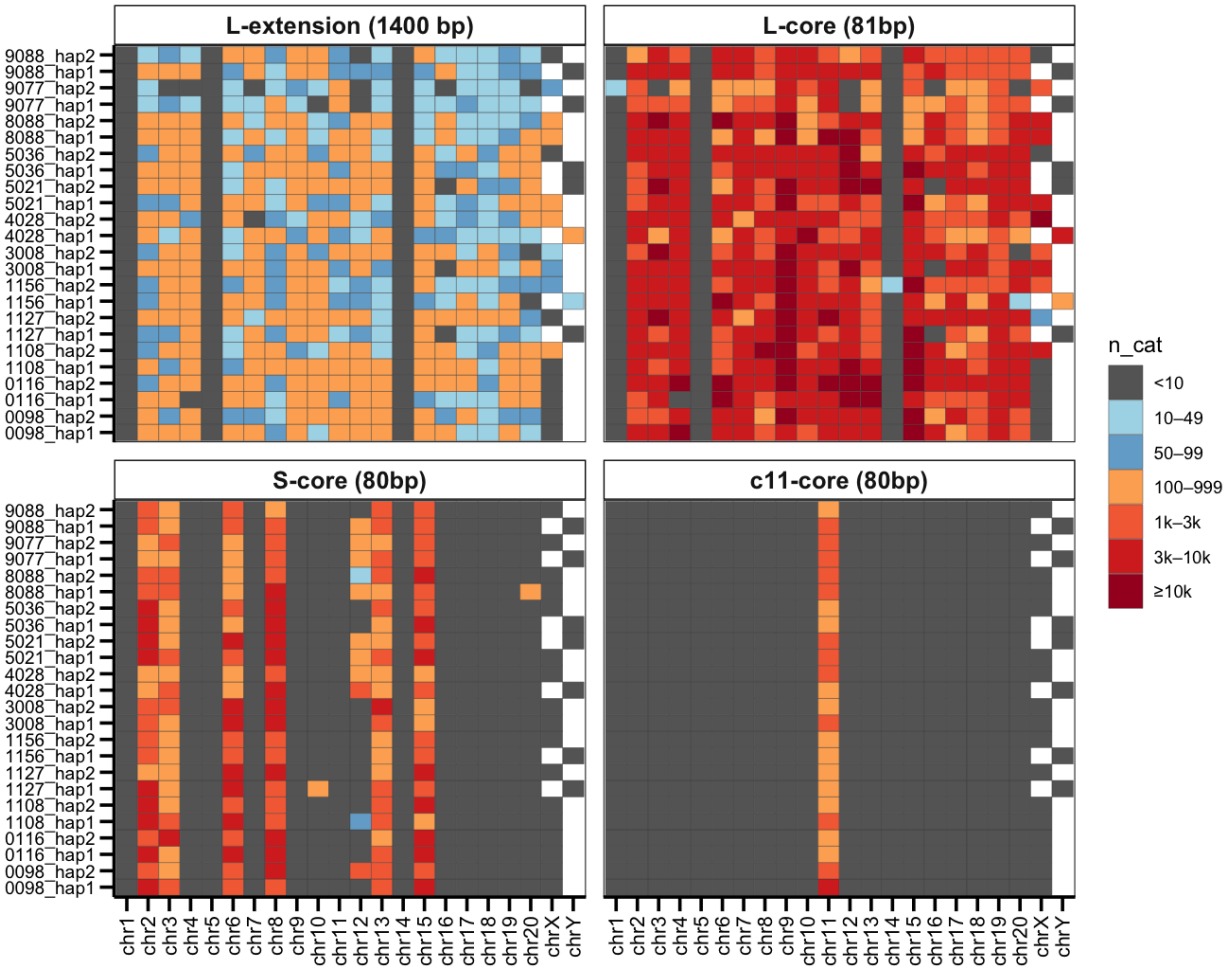


Supplementary Figure.2

a, Chromosome-scale spatial distribution of major satellite repeat classes. Genome-wide spatial distributions of major satellite repeat subclasses are shown for a representative haplotype (0098_hap1). Genomic occupancy was calculated in non-overlapping 1-Mb bins across each chromosome, and the amount of occupied sequence (bp) per bin is indicated by colour intensity. Distinct satellite subclasses exhibit non-uniform chromosomal distributions, with enrichment patterns that differ between repeat types. These spatial profiles provide a genome-wide context for subsequent analyses of hierarchical satellite architectures.

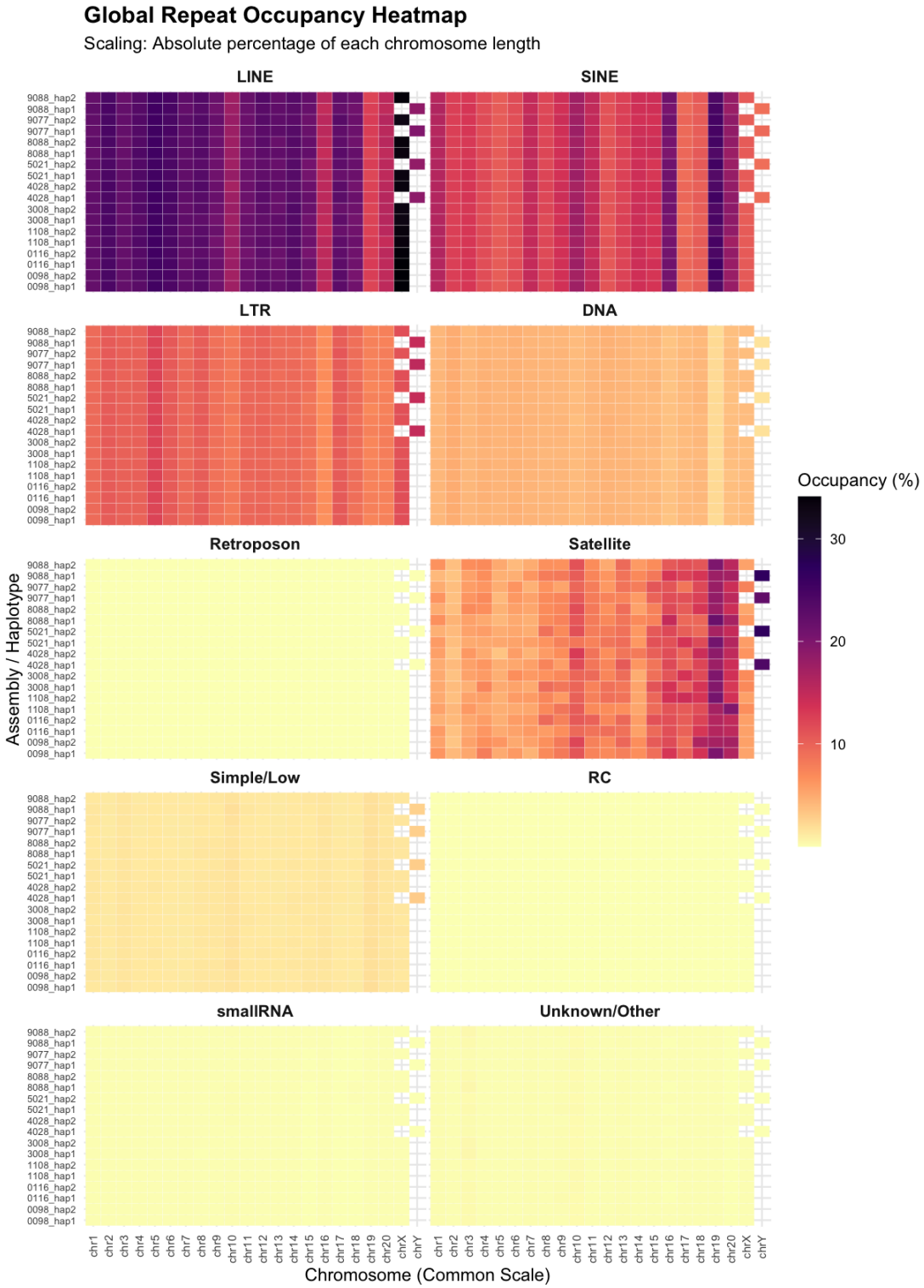
Supplementary Figure.3

(a)

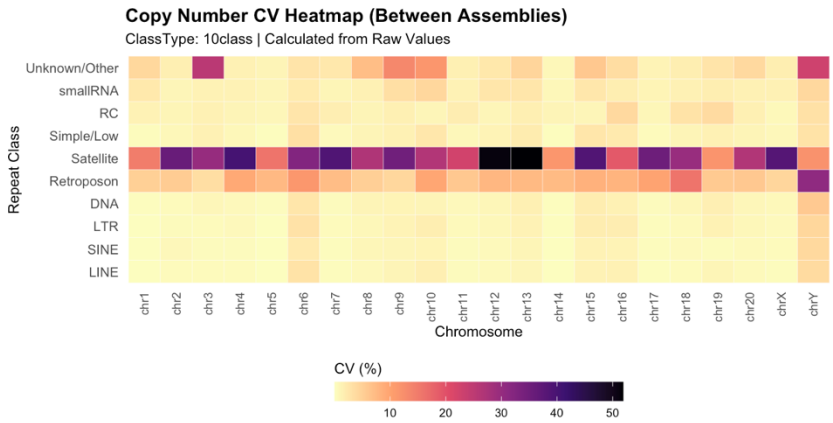


(b)

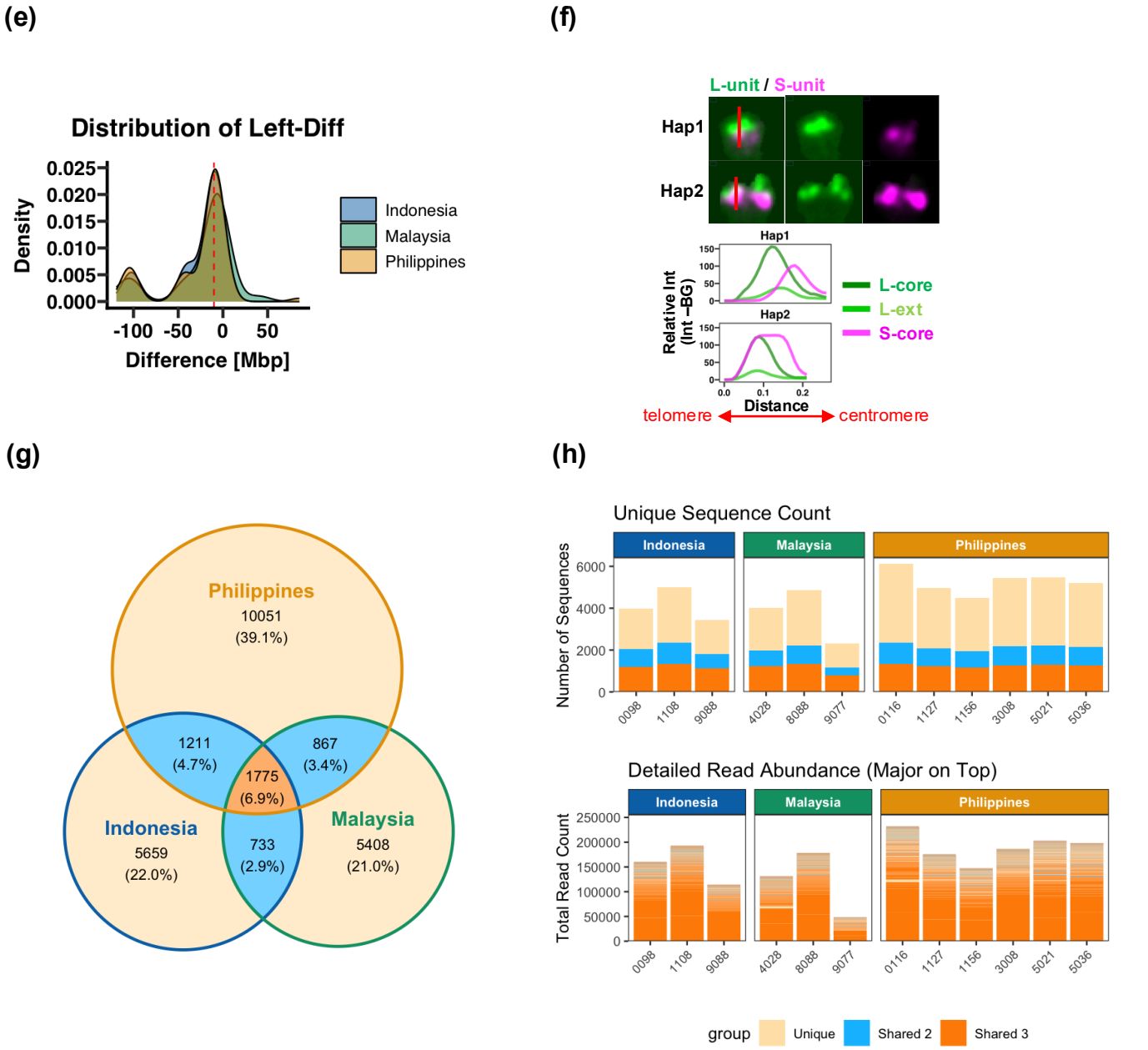
(c)


Global Repeat Occupancy Heatmap
Scaling: Absolute percentage of each chromosome length
LINE
SINE
LTR
DNA
Retroposon
Satellite
Simple/Low
RC
smallRNA
Unknown/Other
Assembly / Haplotype
9088_hap2
9088_hap1
9077_hap2
9077_hap1
8088_hap2
8088_hap1
5021_hap2
5021_hap1
4028_hap2
4028_hap1
3008_hap2
3008_hap1
1108_hap2
1108_hap1
0116_hap2
0116_hap1
0098_hap2
0098_hap1
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chrX
chrY
Chromosome (Common Scale)
Occupancy (%)
30
20
10

(d)


Copy Number CV Heatmap (Between Assemblies)
ClassType: 10class | Calculated from Raw Values
Repeat Class
Unknown/Other
smallRNA
RC
Simple/Low
Satellite
Retroposon
DNA
LTR
SINE
LINE
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chrX
chrY
Chromosome
CV (%)
10
20
30
40
50

Supplementary Figure.3



Supplementary Figure.3

a, Chromosome-level distribution of tangled-ball repeat units across haplotypes.

Heatmaps showing chromosome-level copy number categories of tangled-ball repeat subclasses across all haplotypes. Rows represent individual haplotypes and columns represent chromosomes. Repeat subclasses are shown separately: L-extension (~1.4 kb), L-core (81 bp;), S-core (80 bp), and the chromosome 11–specific subtype (80bp). Color intensity denotes raw copy number per chromosome. Distinct chromosome-level enrichment patterns are observed for each repeat subclass. L-extension and L-core units are distributed across multiple chromosomes, whereas the chr11-specific subtype is confined exclusively to chromosome 11. Overall distribution patterns are broadly conserved across individuals, consistent with species-level architectural organization of tangled-ball repeats. Copy numbers were derived from genome-wide repeat searches allowing up to five mismatches and represent raw counts without normalization by chromosome length.

b. Conservation of the chromosome 11 inversion and gene order across haplotypes.

Comparative gene-order alignment of the chromosome 11 region encompassing the previously described human–macaque inversion boundary. Gene orientation and genomic positions are shown for representative references (CHM13v2.0, MMU8v2.0, MFA8v1.1) and all haplotypes generated in this study. The inversion configuration relative to the human reference is consistently observed across macaque haplotypes, with preserved local gene order surrounding DDX11. These data support the stability of the chromosome 11 inversion and its conserved genomic context across individuals.

c, Chromosome-level repeat landscape across haplotypes. Heatmaps showing the absolute percentage occupancy of major repeat classes per chromosome across all haplotypes. Repeat annotations were obtained using RepeatMasker and grouped into ten major categories. Values represent the fraction of each chromosome occupied by the indicated repeat class. Overall repeat landscapes are broadly consistent across individuals, with no obvious population-level separation in global repeat composition.

d, Inter-haplotype variability of repeat copy number across chromosomes.

Heatmap showing the coefficient of variation (CV, %) of repeat copy number across haplotypes for each chromosome and repeat class. CV values were calculated from raw copy number counts without normalization. Elevated variability is primarily observed in satellite repeats, whereas other repeat classes (LINE, SINE, LTR, DNA transposons, retroposons, and simple/low-complexity repeats) exhibit comparatively low inter-haplotype variance. These results indicate that structural heterogeneity is disproportionately concentrated within satellite repeat classes.

e, Distribution of the directionality index (L-core minus S-core start position).

Density distributions of the directionality index across haplotypes, stratified by geographic origin. The index was defined as the genomic start position of the L-core array minus that of the S-core array along each chromosome arm. Negative values indicate the canonical telomere-proximal configuration (L-core → S-core), whereas positive values indicate the reversed configuration. The distributions are centered near zero with extended tails on both sides, consistent with the presence of orientation polymorphism without large-scale population-specific directional shifts.

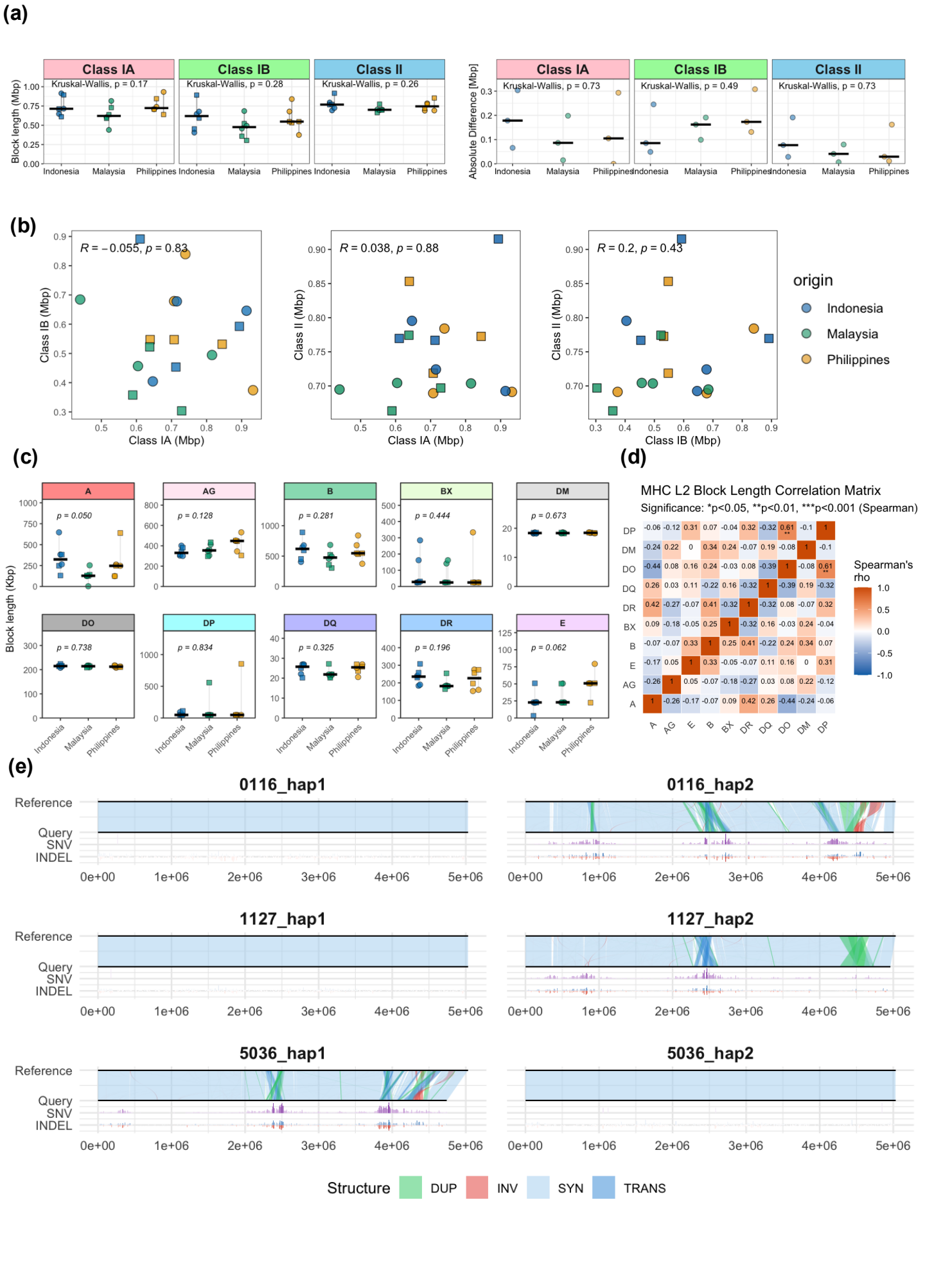
Supplementary Figure.3

f. Cytogenetic validation of haplotype-specific orientation polymorphism. Top, representative dual-color fluorescence in situ hybridization (FISH) images showing localization of L-core (green) and S-core (magenta) repeat units on homologous chromosomes. Merged images highlight spatial separation between the two repeat classes. Bottom, fluorescence intensity profiles plotted along the telomere-to-centromere axis for representative haplotypes (Hap1 and Hap2). Relative positioning of L-core, L-extension, and S-core signals demonstrates reversed spatial configurations between homologous chromosomes, consistent with the orientation polymorphism inferred from diploid T2T assemblies.

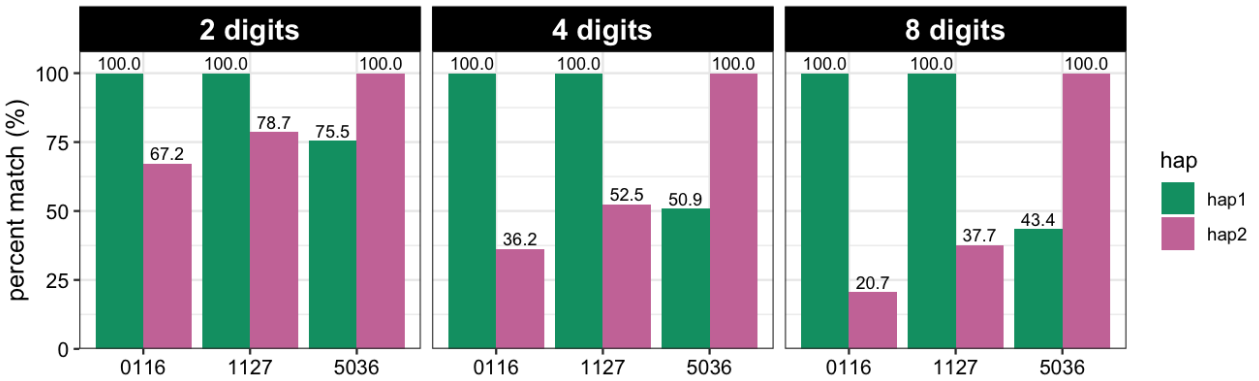
g. Geographic partitioning of L-core sequence variants. Venn diagram showing the distribution of L-core sequence variants across individuals from Indonesia, Malaysia, and the Philippines. Variants were defined based on clustering of L-core sequences allowing up to five mismatches, including insertions and deletions, corresponding to approximately 6% sequence divergence. Numbers indicate the count and percentage of total detected variants belonging to each region-specific or shared category. A substantial fraction of variants is shared among populations, whereas region-specific variants are also observed, particularly within the Philippine cohort. These results indicate sequence-level geographic stratification superimposed on a broadly conserved architectural framework.

h. Regional distribution of L-core variant counts and read abundance. Top, number of unique L-core variant sequences detected in each individual, partitioned into region-specific variants (Unique), variants shared between two regions (Shared 2), and variants shared among all three regions (Shared 3). Bottom, corresponding total read counts supporting each variant category, with major categories shown on top of the stacked bars. Individuals are grouped by geographic origin. Although region-specific variants are present in all populations, a substantial proportion of variants and read abundance derives from sequences shared across regions. These data indicate that geographic stratification occurs at the sequence level while preserving a common core repertoire across macaque populations.

Supplementary Figure.4



(f)


2 digits
4 digits
8 digits
percent match (%)
100
75
50
25
0
100.0
67.2
100.0
78.7
75.5
100.0
100.0
36.2
100.0
52.5
50.9
100.0
100.0
20.7
100.0
37.7
43.4
100.0
0116
1127
5036
hap
hap1
hap2

Supplementary Figure.4

a. Geographic comparison of MHC class block lengths and intra-individual differences.

Left: Block lengths of MHC Class IA, IB, and II regions across haplotypes grouped by geographic origin (cohort 1; N = 9 individuals, 18 haplotypes). Haplotypes from the same individual are connected by lines. Right: Absolute intra-individual differences in block length between paired haplotypes. Statistical comparisons were performed using the Kruskal–Wallis test. No significant differences were detected for either block lengths or intra-individual differences.

b. Correlation analysis of MHC class block lengths.

Scatter plots showing pairwise correlations of block lengths among MHC Class IA, Class IB, and Class II regions across haplotypes (cohort 1; N = 18 haplotypes). Left: Class IA vs Class IB. Middle: Class IA vs Class II. Right: Class IB vs Class II. Spearman's rank correlation coefficients (R) and corresponding p-values are indicated in each panel. No significant correlations were observed among the three class regions, supporting their structural independence. Only cohort 1 was included to avoid relatedness bias from the pedigree samples.

c. Geographic comparison of gene-block lengths within the MHC region.

Block lengths of eight MHC gene blocks (A, AG, B, BX, DM, DO, DP, DQ, DR, and E) were compared among haplotypes grouped by geographic origin (Indonesia, Malaysia, Philippines; cohort 1, N = 9 individuals; 18 haplotypes). Each point represents one haplotype. Statistical comparisons across geographic groups were performed using the Kruskal–Wallis test, and corresponding p-values are indicated in each panel. No statistically significant differences were detected among geographic groups for any block.

d. Spearman correlation matrix of MHC gene-block lengths.

Heatmap showing pairwise Spearman correlation coefficients (ρ) for block lengths among MHC gene blocks (A, AG, B, BX, DR, DQ, DO, DM, DP) across haplotypes (cohort 1; N = 18 haplotypes). Color intensity represents the magnitude and direction of correlation (ρ). Statistical significance thresholds are indicated as follows: *p < 0.05, **p < 0.01, ***p < 0.001. Overall, correlations among gene blocks were generally weak and inconsistent, supporting largely independent structural variation across MHC blocks.

e. Structural variation within the MHC region relative to 1156 hap1.

SyRI-based structural rearrangement analysis of the MHC region is shown for each haplotype assembly compared against 1156 hap1 (reference), which corresponds to the paternal haplotype of the pedigree samples and therefore serves as a biologically meaningful reference for inheritance comparisons. For each sample and haplotype, the top track represents the reference (1156 hap1), and the lower tracks indicate query alignments and detected variants across the MHC interval (~5 Mb). Structural categories are color-coded as follows: duplications (DUP, green), inversions (INV, red), syntenic regions (SYN, light blue), and translocations (TRANS, blue). SNVs and small INDELs are displayed in the lower panels. All coordinates correspond to the MHC locus only.

f. **Haplotype-resolved allele matching relative to sample 1156 hap1.**

Bar plots show the percentage of alleles matched (percent match (%), y-axis) for each sample (ID, x-axis), comparing haplotype 1 (hap1; teal) and haplotype 2 (hap2; purple). Match rates were calculated relative to the 1156 hap1 reference haplotype, with total 56 *Mafa* genes assessed per comparison. Results are presented at three MHC resolution thresholds—2 digits, 4 digits, and 8 digits (facets). Values above bars indicate percentages

Table S1 | A summary of macaque background and sequencing data

ID	origin	age	sex	Cohort	Pedigree
0098	Indonesia	5	F	Cohort 1	
1108	Indonesia	5	F	Cohort 1	
9088	Indonesia	8	M	Cohort 1	
8088	Malaysia	5	F	Cohort 1	
9077	Malaysia	10	M	Cohort 1	
4028	Malaysia	7	M	Cohort 1	
3008	Philippines	7	F	Cohort 1	
0116	Philippines	6	F	Cohort 1/2	Paternal half-sib
5021	Philippines	11	M	Cohort 1	
5036	Philippines	8	M	Cohort 2	Paternal half-sib
1156	Philippines	21	M	Cohort 2	Sier
1127	Philippines	7	M	Cohort 2	Paternal half-sib

Table S2 | Sequencing statistics for Oxford Nanopore ultra-long, PacBio HiFi, and Pore-C datasets used for diploid genome assembly

ID	ONT UL bases (Gb)	ONT UL coverage (Gb)	ONT UL N50 (kb)	ONT UL Longest read (Mb)	ONT UL mean read length (kb)	ONT UL Q20 (%)	ONT UL Q30 (%)
0098	209.8	68.6	66.4	1.9	19.9	90.4	85.1
1108	234.8	76.7	74.4	1.5	12.7	91.0	85.9
9088	210.1	68.6	47.9	1.9	14.7	90.8	85.8
8088	213.2	69.7	80.5	1.7	35.6	82.4	72.8
9077	260.8	85.2	22.5	1.7	9.3	90.0	84.7
4028	272.8	89.1	32.7	1.9	12.7	90.0	84.8
3008	185.2	60.5	88.3	1.8	12.9	90.3	85.2
0116	226.3	74.0	77.1	1.3	53.9	93.5	88.6
5021	222.5	72.7	90.9	2.3	38.3	82.9	73.8
5036	191.9	62.7	82.9	2.0	16.3	91.7	86.8
1156	274.5	89.7	67.7	1.8	11.5	89.5	83.2
1127	158.5	51.8	78.6	2.1	31.7	89.4	83.8

ID	PacBio HiFi bases (Gb)	PacBio HiFi coverage (Gb)	PacBio HiFi N50 (kb)	PacBio HiFi Longest read (kb)	PacBio HiFi mean read length (kb)	PacBio HiFi Q20 (%)	PacBio HiFi Q30 (%)
0098	73.6	24.1	22.2	57.7	20.5	95.8	89.8
1108	89.9	29.4	19.4	60.0	18.4	96.0	90.4
9088	60.1	19.6	20.7	53.7	19.5	93.4	84.9
8088	60.3	19.7	22.5	56.1	20.4	93.1	84.2
9077	70.1	22.9	20.7	51.9	19.8	94.6	87.4
4028	59.8	19.5	19.6	56.9	12.8	94.4	87.1
3008	74.6	24.4	24.7	64.8	23.0	95.1	88.1
0116	96.0	31.4	19.6	62.2	18.8	96.1	90.7
5021	79.7	26.1	24.3	71.9	21.9	94.8	87.7
5036	169.5	55.4	22,0	61.8	20.4	95.8	89.9
1156	102.9	33.6	21.9	71.4	20.8	96.7	91.9
1127	96.2	31.4	23.6	63.8	21.7	96.2	90.6

ID	Pore-C bases (Gb)	Pore-C coverage (Gb)	Pore-C N50 (kb)	Pore-C Longest read (kb)	Pore-C mean read length (kb)
0098	58.5	19.1	7.4	156.4	3.0
1108	93.7	30.6	8.6	492.2	5.5
9088	90.5	29.6	8.0	564.2	5.0
8088	93.0	30.4	9.4	137.7	5.4
9077	113.0	36.9	7.8	385.3	4.8
4028	107.5	35.1	9.3	502.8	5.7
3008	87.2	28.5	8.7	219.6	5.6
0116	94.9	31.0	8.4	297.3	5.4
5021	96.1	31.4	9.2	309.1	5.5
5036	54.8	17.9	8.6	470.0	3.3
1156	86.3	28.2	7.8	204.1	3.3
1127	110.5	36.1	8.6	212.4	5.1

Values represent total sequencing yield and read statistics for each dataset. Quality metrics (Q20 and Q30) are reported for ONT and PacBio HiFi reads.