Oxford Nanopore Technologies

Automated single-platform telomere-to-telomere *de novo* genome assembly for human, plant, and animal genomes

Advances in Oxford Nanopore read accuracy, as well as read correction and assembly algorithms, enable routine T2T assembly for platinum-standard references and population-scale pangenome analysis

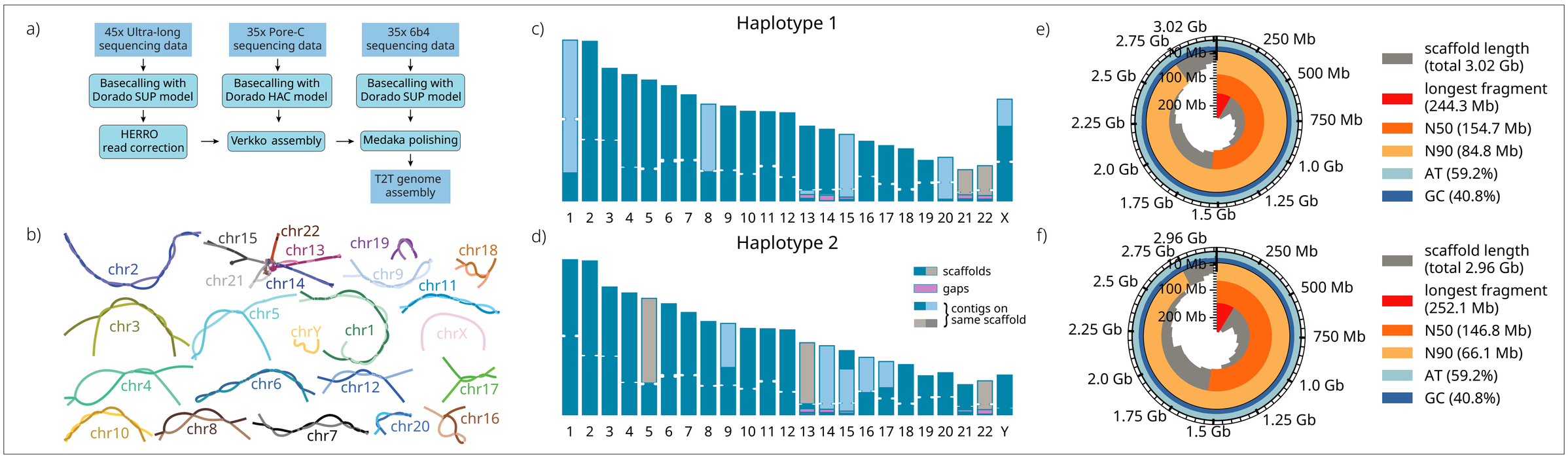

Fig. 1 a) Telomere-to-telomere assembly pipeline b) bandage plot of Verrko assembly graph c), d) karyoplots of haplotypes 1 and 2 e), f) assembly statistics for haplotypes 1 and 2.

Analysis of ultra-long and Pore-C sequencing data with Dorado correct (HERRO) and Verkko algorithms yields the most complete and contiguous automated human genome assembly yet produced

Telomere-to-telomere (T2T) assemblies provide the most information possible about an organism's genetic code, allowing for analysis of previously inaccessible regions of the genome such as centromeric satellites and other large/high copy repeat regions. The previous state-of-the-art technique for generating T2T assemblies required both highly accurate reads like Oxford Nanopore duplex reads as well as ultra-long simplex reads; however, advances in raw-read accuracy and the development of the HERRO correction algorithm (implemented in Dorado correct) bring the corrected accuracy of simplex reads on par with corrected duplex reads. Assembly of the GIAB HG002 reference with 50x of HERRO-corrected ultra-long reads (N50 = 110 kb) using the Verkko assembler yielded 20 large graph components corresponding to the 19 metacentric human chromosomes plus the acrocentric chromosomes (**Fig. 1a, b**). Haplotype reconstruction and scaffolding of this assembly graph with 35x of Pore-C data using the native Pore-C phasing and scaffolding module within Verkko yielded 28 T2T contigs and 14 additional T2T scaffolds (**Fig. 1a, c, d**). Haplotype 1 contained 3.02 Gb assembled sequence, with a scaffold N50 of 154.7 Mb (**Fig.1e**), while haplotype 2 contained 2.96 Gb assembled sequence, with a scaffold N50 of 146.8 Mb (**Fig. 1f**).

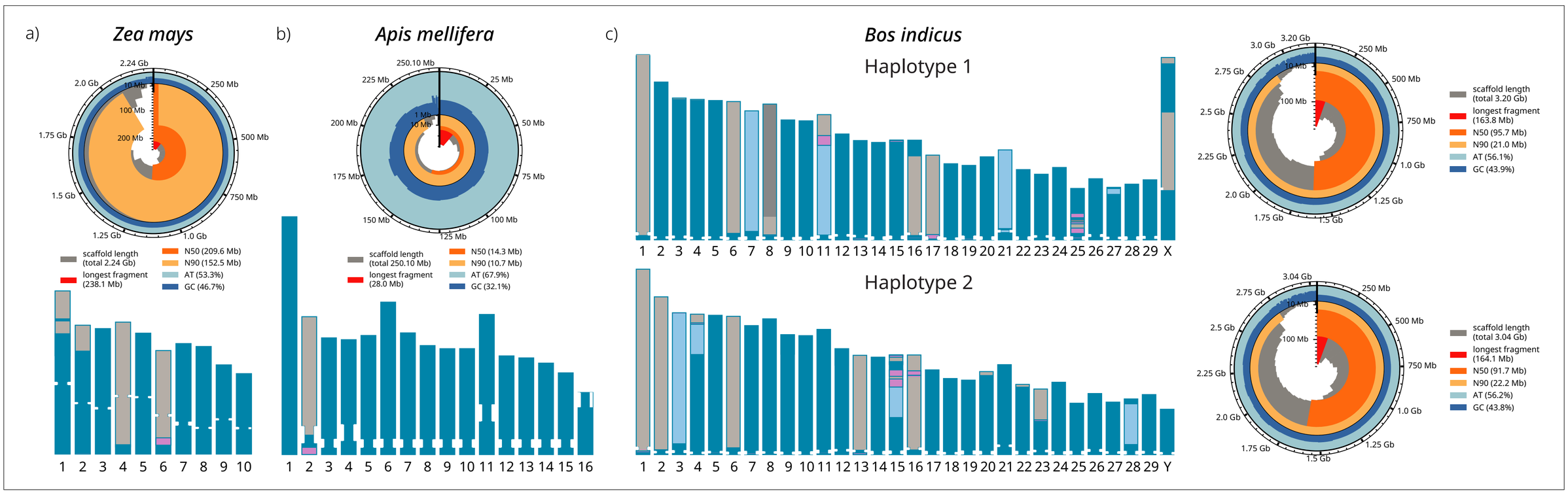

Fig. 2 a) Assembly statistics and karyoplot of *Zea mays* b) assembly statistics and karyoplot of *Apis mellifera* c) assembly statistics and karyoplots for each haplotype of *Bos indicus.*

Accurate and complete assemblies for both small and large genomes from a diverse range of organisms with highly scalable T2T assembly workflows

Gapless T2T assemblies can benefit plant and animal genomic applications by providing nearly perfect reference assemblies, as well as facilitating the discovery of the most complete genomic variants. However, obtaining ultra-long reads (60-100 kb N50) from such samples can be challenging due to sample quantity and quality limitations. The new Hifiasm (ONT) assembler is highly effective on Oxford Nanopore datasets with read length N50s of 10–30 kb, and using this data alone, it can output T2T assemblies for haploid/effectively haploid organisms and near-T2T, partially phased (a.k.a. 'dual') assemblies for diploid or polyploid organisms. We assembled 118 Gb of data from maize (*Zea mays*) with the Hifiasm (ONT) assembler, generating an assembly with six of ten chromosomes assembled T2T (**Fig. 2a**). Both the honeybee (*Apis mellifera*) and zebu cattle (*Bos indicus*) have primarily or exclusively acrocentric chromosomes, respectively, limiting assembly of most chromosomes. Nonetheless, we assembled seven of the 16 honeybee chromosomes T2T using 20 Gb of simplex reads (N50 = 26 kb) with Hifiasm (ONT) (**Fig. 2b**). For zebu cattle, we assembled 170 Gb of simplex reads (N50 = 27 kb) and 100 Gb of Pore-C reads using Hifiasm (ONT), yielding a highly complete and contiguous assembly with five chromosomes assembled as T2T contigs and one additional T2T scaffold (**Fig. 2c**). For both the honeybee and the zebu cattle, the majority of the remaining chromosomes were assembled telomere-to-centromere.