

Table S2. Terminologies used to assess the quality of the genome [1].

Matrices	Definition
Genome fraction (%)	The ratio of total aligned bases to the reference genome
Largest contig	The length of the largest contig in the assembly
Nx (N50/N75)	Shortest contig length needed to cover x% of the genome
NAx (NA50/NA75)	Shortest aligned blocks needed to cover x% of the genome. Aligned blocks are obtained by breaking contigs in misassembly events and removing all unaligned bases.
Lx (L50/L75)	The minimum number of contigs that make up x% of genome size
LAx (LA50/LA75)	The minimum number of aligned contigs that make up x% of genome
Total length	The total number of bases in the assembly
Largest alignment	The largest uninterrupted alignment in the assembly
Mismatches per 100 kbp	The average number of mismatches per 10 kbp aligned bases
Contigs >= 1000 bp	The number of contigs that are >= 1 kb
Number of Contigs	The total number of contigs in the assembly
Indels per 100 kbp	The average number of single nucleotide insertions or deletions per 100 kbp aligned bases
Total aligned length	The total aligned bases of in the assembly
Misassemblies	The number of misassembly events

References

1. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QCAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29:1072–5.