

## Server manual

This server is prepared for the analysis of alignment file, produced by BLAT (.psl file). By providing the reference genome file and .psl file, produced by the blat software after alignment of contigs to reference genome, user can evaluate the performance of assembled genome by an assembler. With the help of this software, user can evaluate the quality of assembly. This software will provide the alignment result in form of N 50 contig length, Total genome covered by the contigs (% of the genome), Error rate(% of the assembly), Total contig matches(% of the total contig size sum) and Error rate(% of the total contig size sum).

### Results file (Example)

This is the analysis report:-

**N50 Contig length = 6038116**

**Total number of contigs present in Contigs.fa file:- 1**

**Contigs Unalienable to reference genome: - 0**

**Contigs aligned to reference genome: - 1**

**Total genome covered (Nucleotides):- 5999308**

**Total genome covered (% of the genome):- 99.9884666666667**

**Total contig matches (Nucleotides):- 5992513**

**Total contig matches (% of the total contig size sum):- 99.2447478650625**

**Error in the contig alignment:-**

**Mismatches = 0**

**Query gaps = 45603**

**Hit gaps = 6794**

**Ns = 0**

**Total Errors = 52397**

**Error rate (% of the total contig size sum) = 0**

**Total incorrect nucleotide alignment:-**

**Unalienable count = 0**

**Unalienable contig base count = 0**

**Mismatches = 0**

**Total incorrect = 52397**

**Error rate of total assembly (% of the assembly) = 0.867770675488845**

### Explanation of Result file

1. **N 50 contig length:** - The contig length such that 50% of the denovo assembled genome lies in blocks of this size or larger.
2. **Contigs present:** - Number of contigs produced by the genome assembler.
3. **Contigs Unalienable to reference genome:** - Number of contigs not aligned to reference genome.
4. **Contigs aligned to reference genome:** - Total number of contigs aligned to reference genome.
5. **Total genome covered (Nucleotides):-** Total number of nucleotides of genome on which contigs aligned.
6. **Total genome covered (% of the genome):-**  $\frac{\text{Total genome covered (Nucleotides)} * 100}{\text{Total reference genome size}}$
7. **Total contig matches (Nucleotides):-** Total number of nucleotides of contigs aligned to reference genome.
8. **Total contig matches (% of the total contig size sum):-**  $\frac{\text{Total nucleotides of contigs matches to reference genome} * 100}{\text{Total contig size sum}}$

### Error in the contig alignment

- **Mismatches:** - Total number of mismatches between contigs and reference genome.
- **Query gaps:-** Sum of number of inserts in query sequences.
- **Hit gaps:-** Sum of number of inserts in target sequence e.g reference genome.
- **Ns:** - Any nucleotide.
- **Total Errors:-** Mismatches + Query gaps + Hit gaps + Ns
- **Error rate (% of the total contig size sum):-**  $\frac{\text{Mismatches sum} * 100}{\text{Total contig size sum}}$

### Total incorrect nucleotide alignment

- **Unalienable count :-** Query size - (Matches + Ns + Mismatches + Query gaps)
- **Unalienable contig base count:-** Total nucleotide sum of contigs not aligned to reference genome.
- **Mismatches:-** Sum of nucleotides in alignment that don't match.
- **Total incorrect:-** Unalignable count + Unalienable contig base count + Mismatches
- **Error rate of total assembly (% of the assembly):-**  $\frac{(\text{Unalienable count} + \text{Total errors} + \text{Unalienable contig base count}) * 100}{\text{Total genome size}}$