**Table S4.** High throughput sequencing results. JM: *Juglans mandshurica*; QM: *Quercus mongolica*; CB: Conifer-broadleaf forest; LG: *Larix gmelinii*; PK: *Pinus koraiensis*. Note: N (%) is the ratio of fuzzy bases to the total number of bases; GC (%) is the GC content, that is, the ratio of G bases and C bases to the total number of bases; Q20 (%) and Q30 (%) The ratio of the bases whose measurement accuracy is 99% and 99.9% or more to the total number of bases, respectively. Statistics of the *de novo* assembly for different revegetation forest soil metagenomes. JM: *Juglans mandshurica*; QM: *Quercus mongolica*; CB: Conifer-broadleaf forest; LG: *Larix gmelinii*; PK: *Pinus koraiensis*.

|  |  |  |  |
| --- | --- | --- | --- |
| Samples | File | Contigs | Scaffolds |
| QM | Min sequence length | 200 | 200 |
| Max sequence length | 8240 | 9020 |
| Total sequence number | 915504 | 908527 |
| N20 (bp) | 621 | 630 |
| N50 (bp) | 417 | 418 |
| N90 (bp) | 317 | 317 |
| N number: | 0 | 1563 |
| N rate | 0 | 3.91E-06 |
| Total sequence length | 401328651 | 400237355 |
| GC Number | 0.61751437 | 0.617582627 |
| Sequences greater than 1kb | 18141 | 19329 |
| CB | Min sequence length | 200 | 200 |
| Max sequence length | 8737 | 10899 |
| Total sequence number | 1157992 | 1147321 |
| N20 (bp) | 675 | 689 |
| N50 (bp) | 438 | 440 |
| N90 (bp) | 319 | 319 |
| N number | 0 | 2624 |
| N rate | 0 | 4.99E-06 |
| Total sequence length | 527671943 | 526041760 |
| GC Number | 0.612721365 | 0.612750684 |
| Sequences greater than 1kb | 30733 | 32933 |
| PK | Min sequence length | 200 | 200 |
| Max sequence length | 12714 | 15669 |
| Total sequence number | 1369973 | 1355172 |
| N20 (bp) | 646 | 660 |
| N50 (bp) | 433 | 435 |
| N90 (bp) | 319 | 319 |
| N number | 0 | 5413 |
| N rate | 0 | 8.82E-06 |
| Total sequence length | 615583093 | 613573735 |
| GC Number | 0.61386935 | 0.613889333 |
| Sequences greater than 1kb | 30228 | 33020 |
| LG | Min sequence length | 200 | 200 |
| Max sequence length | 35249 | 35249 |
| Total sequence number | 1468972 | 1452211 |
| N20 (bp) | 708 | 728 |
| N50 (bp) | 450 | 452 |
| N90 (bp) | 320 | 320 |
| N number | 0 | 11643 |
| N rate | 0 | 1.71E-05 |
| Total sequence length | 683483919 | 680952808 |
| GC Number | 0.609236306 | 0.609263874 |
| Sequences greater than 1kb | 44392 | 48422 |
| JM | Min sequence length | 200 | 200 |
| Max sequence length | 10951 | 13184 |
| Total sequence number | 942545 | 934630 |
| N20 (bp) | 605 | 614 |
| N50 (bp) | 415 | 417 |
| N90 (bp) | 317 | 317 |
| N number | 0 | 1625 |
| N rate | 0 | 3.98E-06 |
| Total sequence length | 409558557 | 408427356 |
| GC Number | 0.625613946 | 0.62566193 |
| Sequences greater than 1kb | 16145 | 17387 |

Note: The length of N20/N50/N90 refers to the corresponding length of the last sequence when the length of the addition reaches 20%/50%/90% of the total length of the Contigs/Scaffolds sequence, which obtained by splicing all the assemblies according to the length from long to short, and added in order from long to short. The longer the length of N20/N50/N90, the better the assembly and splicing effect, and the more complete the obtained metagenomic sequence is. Among them, the length of N50 is particularly important, which is one of the main indicators to measure the effect of metagenomic assembly and splicing.