### **High quality draft genome of Arogyapacha *(Trichopus zeylanicus*), an important medicinal plant endemic to Western Ghats of India**

### Biju Vadakkemukadiyil Chellappan, Shidhi PR, Sheethal Vijayan, Veena S Rajan, Anu Sasi, Achuthsankar S. Nair

Department of Computational Biology and Bioinformatics, University of Kerala, Thiruvananthapuram, Kerala, India

**Table S1**. ***Trichopus zeylanicus* whole genome sequencing read summary**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Library | Instrument | Insert size (bp) | Library ID | Read length | Raw data | | Clean data | |
| Number of reads (R1+R2) | Total size (Gb) | Number of reads (R1+R2) | Total size (Gb) |
| Illumina  paired-end | HiSeq 2500 | 300 | TZA1\_300BP | 150X2 | 335,834,172 | 50.3 | 335,632,386 | 48.9 |
| 500 | TZA1\_500BP | 150X2 | 330,505,382 | 49.5 | 329,998,742 | 47.9 |
| 800 | TZA1\_800BP | 150X2 | 289,589,646 | 43.4 | 289,519,092 | 39.6 |
| **Total** | | | | | **955,929,200** | **143.2** | **955,150,220** | **136.4** |
| SMRT cell | Pacbio-Sequel | 15 kb | SMRT1 | 7.2 kb (mean) |  |  | 321,090 | 2.7 |
| 15 kb | SMRT2 | 8.2 kb (mean) |  |  | 316,299 | 2.6 |
| 15 kb | SMRT3 | 8.9 kb (mean) |  |  | 417,866 | 3.7 |
| 15 kb | SMRT4 | 7.9 kb (mean) |  |  | 521,683 | 4.2 |
| 15 kb | SMRT5 | 8.1 kb (mean) |  |  | 504,583 | 4.1 |
| **Total** | | | | | | | **2,081,521** | **17.3** |

**Table S2. List of reference organelle genome used in this study**

|  |  |  |
| --- | --- | --- |
| Species | source |  |
| *Dioscorea rotundata* | Chloroplast | KJ490011.1 |
| *Dioscorea elephantipes* | Chloroplast | EF380353.1 |
| *Dioscorea zingiberensis* | Chloroplast | KP899622.1 |
| *Elaeis guineensis* | Chloroplast | JF274081.1 |
| *Phoenix dactylifera* | Chloroplast | GU811709.2 |
| Zey mays | Mitochondria | NC\_007982.1 |
| Oryza sativa | Mitochondria | NC\_011033.1 |
| Glycine Max | Mitochondria | NC\_020455.1 |

**Table S3. Repeat fraction of *T. zeylanicus* genome**

|  |  |  |
| --- | --- | --- |
| **Repeat Class** | **Total Base** | **% in the genome** |
|  |  |  |
| **Class I (Retro-transposons)** | **244792136** | **34.37** |
| Copia | 182826294 | 25.67 |
| Gypsy | 51141497 | 7.18 |
| Caulimovirus | 8498226 | 1.19 |
| Roo | 2325747 | 0.33 |
| Pao | 372 | 0 |
|  |  | 0 |
| **LINE** | **380952** | **0.05** |
| L1 | 380952 | 0.05 |
|  |  | 0 |
| **Class II (DNA transposons)** | **13462807** | **1.8** |
| hAT | 7174609 | 1.01 |
| MuDR | 2673081 | 0.37 |
| EnSpm | 1303132 | 0.19 |
| PIF-Harbinger | 1250761 | 0.18 |
| EnSpm | 619950 | 0.09 |
| Helitron | 31533 | 0 |
| Others DNA transposons | 409741 | 0.05 |
|  |  |  |
| **Unknown** | **39320652** | **5.52** |
|  |  |  |
| **Total interspersed elements** | **294941702** | **41.42** |
|  |  |  |
| Low\_complexity | 2324689 | 0.33 |
| Simple\_repeat | 43511910 | 6.11 |
|  |  |  |
| **Total Repeats** | **343793146** | **48.18** |

**Table S4. Non-coding RNAs in *T. zeylanicus***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Type** | **Copy number** | **Average length** | **Total length** | **% in the genome** |  |
| **tRNA** | **3644** | **75.10** | **273738** | **0.0383** |  |
| **rRNA** | **261** | **1599.46** | **417458** | **0.0584** |  |
| 18s | 45 | 1881.82 | 84682 | 0.0119 |  |
| 28s | 53 | 5931.89 | 314390 | 0.0440 |  |
| 8s | 163 | 112.80 | 18386 | 0.0026 |  |
| **miRNA** | **194** | **127.01** | **24639** | **0.0034** |  |
| **snRNA** | **448** | **386.50** | **55220** | **0.0077** |  |
| C/D-box | 283 | 115.77 | 32762 | 0.0046 |  |
| H/ACA-box | 67 | 131.42 | 8805 | 0.0012 |  |
| spliceosomal | 98 | 139.32 | 13653 | 0.0019 |  |

**Table S5. List of domains distributed in *T. zeylanicus* proteins sequences**

|  |  |
| --- | --- |
| **IPS Domain** | **Number** |
| Protein kinase domain | 1061 |
| Serine-threonine/tyrosine-protein kinase, catalytic domain | 324 |
| Zinc finger, RING-type | 322 |
| RNA recognition motif domain | 295 |
| SANT/Myb domain | 285 |
| Myb domain | 234 |
| WD40-repeat-containing domain | 206 |
| Myc-type, basic helix-loop-helix (bHLH domain) | 185 |
| EF-hand domain | 179 |
| Leucine-rich repeat-containing N-terminal, plant-type | 172 |
| Major facilitator superfamily domain | 170 |
| Zinc finger C2H2-type | 159 |
| F-box domain | 154 |
| DYW domain | 153 |
| AP2/ERF domain | 150 |
| NAC domain | 140 |
| Zinc finger, CCHC-type | 131 |
| ABC transporter-like | 129 |
| Tetratricopeptide repeat-containing domain | 125 |

**Table S6. Secondary metabolite pathway in *T. zeylanicus***

|  |  |  |  |
| --- | --- | --- | --- |
| **KEGG pathway** | **KEGG ID** | **Enzyme class** | **No. of genes** |
| **Metabolism of terpenoids and polyketides** | | | |
| Terpenoid backbone biosynthesis | map00900 | 23 | 65 |
| Carotenoid biosynthesis | map00906 | 6 | 26 |
| Diterpenoid biosynthesis | map00904 | 5 | 26 |
| Zeatin biosynthesis | map00908 | 3 | 20 |
| Limonene and pinene degradation | map00903 | 2 | 16 |
| Geraniol degradation | map00281 | 5 | 15 |
| Polyketide sugar unit biosynthesis | map00523 | 3 | 13 |
| Insect hormone biosynthesis | map00981 | 2 | 12 |
| Biosynthesis of vancomycin group antibiotics | map01055 | 1 | 9 |
| Sesquiterpenoid and triterpenoid biosynthesis | map00909 | 3 | 8 |
| Biosynthesis of ansamycins | map01051 | 1 | 3 |
| **Biosynthesis of other secondary metabolites** | | | |
| Phenylpropanoid biosynthesis | map00940 | 11 | 233 |
| Caffeine metabolism | map00232 | 4 | 60 |
| Streptomycin biosynthesis | map00521 | 9 | 45 |
| Flavonoid biosynthesis | map00941 | 7 | 44 |
| Isoquinoline alkaloid biosynthesis | map00950 | 10 | 33 |
| Indole alkaloid biosynthesis | map00901 | 2 | 32 |
| Neomycin, kanamycin and gentamicin biosynthesis | map00524 | 2 | 24 |
| Tropane, piperidine and pyridine alkaloid biosynthesis | map00960 | 5 | 19 |
| Monobactam biosynthesis | map00261 | 5 | 14 |
| Glucosinolate biosynthesis | map00966 | 2 | 11 |
| Novobiocin biosynthesis | map00401 | 3 | 10 |
| Acarbose and validamycin biosynthesis | map00525 | 1 | 9 |
| Betalain biosynthesis | map00965 | 3 | 5 |
| Aflatoxin biosynthesis | map00254 | 1 | 5 |
| Carbapenem biosynthesis | map00332 | 2 | 5 |
| Stilbenoid, diarylheptanoid and gingerol biosynthesis | map00945 | 1 | 4 |
| Phenazine biosynthesis | map00405 | 1 | 4 |
| Flavone and flavonol biosynthesis | map00944 | 2 | 2 |
| **Xenobiotics biodegradation and metabolism** | | | |
| Aminobenzoate degradation | map00627 | 6 | 380 |
| Drug metabolism - other enzymes | map00983 | 19 | 331 |
| Drug metabolism - cytochrome P450 | map00982 | 7 | 108 |
| Metabolism of xenobiotics by cytochrome P450 | map00980 | 7 | 86 |
| Nitrotoluene degradation | map00633 | 1 | 52 |
| Chloroalkane and chloroalkene degradation | map00625 | 4 | 33 |
| Caprolactam degradation | map00930 | 4 | 23 |
| Benzoate degradation | map00362 | 5 | 16 |
| Naphthalene degradation | map00626 | 2 | 14 |
| Styrene degradation | map00643 | 4 | 10 |
| Toluene degradation | map00623 | 1 | 4 |
| Ethylbenzene degradation | map00642 | 1 | 3 |
| Chlorocyclohexane and chlorobenzene degradation | map00361 | 2 | 3 |