**Supplementary Materials**

***The genome sequence of the avian vampire fly (Philornis downsi), an invasive nest parasite of Darwin’s finches***

**Supplementary Table 1:** *Genome Statistics of P. downsi estimated from short-sequencing reads*

***Property min max***

Heterozygosity 1.40879% 1.42194%

Genome Haploid Length 796,797,589 bp 799,050,413 bp

Genome Repeat Length 232,405,721 bp 233,062,812 bp

Genome Unique Length 564,391,869 bp 565,987,600 bp

Model Fit 95.7713% 98.7012%

Read Error Rate 1.228% 1.228%

**Supplementary Table 2:** *Busco statistics of P. downsi genome*

|  |  |  |
| --- | --- | --- |
| **Categories** | **Number** | **% of total** |
| Complete BUSCOs | 3147 | 95.80 |
| Complete and single-copy BUSCOs | 3059 | 93.12 |
| Complete and duplicated BUSCOs | 88 | 2.68 |
| Fragmented BUSCOs | 71 | 2.17 |
| Missing BUSCOs | 67 | 2.03 |
| Total BUSCO groups searched | 3285 | 100% |

**Supplementary Table 2:** *Repeat statistics in P. downsi genome*

sequences: 41,176

total length: 971,6346,46 bp

GC level: 35.10 %

bases masked: 502,200,627 bp (51.69 %)

==================================================

number of length percentage

elements\* occupied of sequence

--------------------------------------------------

Retroelements 224657 90253146 bp 9.29 %

**SINEs: 58 6761 bp 0.00 %**

Penelope 3187 661022 bp 0.07 %

**LINEs: 196268 74550945 bp 7.67 %**

CRE/SLACS 0 0 bp 0.00 %

L2/CR1/Rex 43733 18550153 bp 1.91 %

R1/LOA/Jockey 15852 7502835 bp 0.77 %

R2/R4/NeSL 834 673762 bp 0.07 %

RTE/Bov-B 102395 29493616 bp 3.04 %

L1/CIN4 1415 94740 bp 0.01 %

**LTR elements: 28331 15695440 bp 1.62 %**

BEL/Pao 12565 7084057 bp 0.73 %

Ty1/Copia 1561 775084 bp 0.08 %

Gypsy/DIRS1 9951 7481940 bp 0.77 %

Retroviral 2481 143387 bp 0.01 %

**DNA transposons** 621139 227262774 bp 23.39 %

hobo-Activator 8749 1542960 bp 0.16 %

Tc1-IS630-Pogo 576750 212960682 bp 21.92 %

En-Spm 0 0 bp 0.00 %

MuDR-IS905 0 0 bp 0.00 %

PiggyBac 1280 395680 bp 0.04 %

Tourist/Harbinger 4060 1924044 bp 0.20 %

Other 275 11451 bp 0.00 %

Rolling-circles 19653 4597029 bp 0.47 %

**Unclassified: 828801 166508339 bp 17.14 %**

Total interspersed repeats: 484024259 bp 49.82 %

Small RNA: 7148 2907047 bp 0.30 %

Satellites: 796 124775 bp 0.01 %

Simple repeats: 213281 8909635 bp 0.92 %

Low complexity: 34278 1638267 bp 0.17 %

**Supplementary Table 3:** *Number of shared orthogroups among various dipteran species and their outgroup*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *H.melpomene* | *A.aegypti* | *C.capitata* | *S.calcitrans* | *M.domestica* | *P.downsi* | *G.morsitans* | *D.melanogaster* |
| *H.melpomene* |  | 6,978 | 6,874 | 6,883 | 6,910 | 6,800 | 6,337 | 6,859 |
| *A.aegypti* | 6,978 |  | 8,015 | 8,008 | 8,018 | 7,796 | 7,228 | 7,995 |
| *C.capitata* | 6,874 | 8,015 |  | 9,123 | 9,186 | 8,855 | 8,014 | 9,096 |
| *S.calcitrans* | 6,883 | 8,008 | 9,123 |  | 9,650 | 9,197 | 8,087 | 9,041 |
| *M.domestica* | 6,910 | 8,018 | 9,186 | 9,650 |  | 9,295 | 8,144 | 9,117 |
| *P.downsi* | 6,800 | 7,796 | 8,855 | 9,197 | 9,295 |  | 8,103 | 8,770 |
| *G.morsitans* | 6,337 | 7,228 | 8,014 | 8,087 | 8,144 | 8,103 |  | 7,940 |
| *D.melanogaster* | 6,859 | 7,995 | 9,096 | 9,041 | 9,117 | 8,770 | 7,940 |  |

**Supplementary Table 4:** *Gene families unique only to P. downsi*

|  |  |
| --- | --- |
| EGT47331 | hypothetical protein CAEBREN\_08836 [Caenorhabditis brenneri] |
| GBO13109 | hypothetical protein AVEN\_233885-1 [Araneus ventricosus] |
| GBP44057 | PiggyBac transposable element-derived protein 4 [Eumeta japonica] |
| KMQ83311 | transposable element tc3 transposase [Lasius niger] |
| KNC20639 | hypothetical protein FF38\_06891, partial [Lucilia cuprina] |
| KNC20966 | hypothetical protein FF38\_09018 [Lucilia cuprina] |
| KNC26906 | hypothetical protein FF38\_10930 [Lucilia cuprina] |
| KXJ68891 | hypothetical protein RP20\_CCG001210 [Aedes albopictus] |
| OAF68034 | hypothetical protein A3Q56\_04227 [Intoshia linei] |
| PCG68510 | hypothetical protein B5V51\_5157, partial [Heliothis virescens] |
| XP\_011295145 | PREDICTED: tigger transposable element-derived protein 6-like isoform X2 [Musca domestica] |
| XP\_011295374 | PREDICTED: tigger transposable element-derived protein 6 [Musca domestica] |
| XP\_013109704 | PREDICTED: uncharacterized protein LOC106088638 [Stomoxys calcitrans] |
| XP\_017475229 | PREDICTED: uncharacterized protein LOC108365650 [Rhagoletis zephyria] |
| XP\_017478109 | PREDICTED: uncharacterized protein LOC108367917 [Rhagoletis zephyria] |
| XP\_017478991 | PREDICTED: uncharacterized protein LOC108368617 [Rhagoletis zephyria] |
| XP\_017479715 | PREDICTED: uncharacterized protein LOC108369194 [Rhagoletis zephyria] |
| XP\_017481195 | PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like [Rhagoletis zephyria] |
| XP\_019891578 | PREDICTED: ATP-binding cassette sub-family A member 3-like [Musca domestica] |
| XP\_019894716 | PREDICTED: uncharacterized protein LOC105262305 isoform X1 [Musca domestica] |
| XP\_021704105 | protein ALP1-like [Aedes aegypti] |
| XP\_022823959 | piggyBac transposable element-derived protein 4-like [Spodoptera litura] |
| XP\_022834134 | uncharacterized protein LOC111361914 [Spodoptera litura] |
| XP\_033325321 | uncharacterized protein LOC117219890 [Megalopta genalis] |
| XP\_036214104 | trypsin zeta-like [Bactrocera oleae] |
| XP\_036337749 | uncharacterized protein LOC118747737 isoform X3 [Rhagoletis pomonella] |
| XP\_036342696 | uncharacterized protein LOC118751975 [Rhagoletis pomonella] |

**Supplementary Table 5:** Number of genes in *Fibrinogen C-Terminal Domain-Containing* gene family and *SCP domain-containing* gene family

|  |  |  |
| --- | --- | --- |
| Species | Fibrinogen C-Terminal Domain-Containing gene family | SCP domain-containing gene family |
| *H.melpomene* | 1 | 2 |
| *A.aegypti* | 33 | 20 |
| *C.capitata* | 10 | 6 |
| *S.calcitrans* | 45 | 17 |
| *M.domestica* | 32 | 25 |
| *P.downsi* | 5 | 7 |
| *G.morsitans* | 7 | 7 |
| *D.melanogaster* | 11 | 15 |

Chart, histogram

Description automatically generated

**Supplementary Figure 1**: *k-mer spectrum and fitted modelling used for estimating genome parameters of P. downsi from short sequencing reads*

Chart, histogram

Description automatically generated

**Supplementary Figure 2:** *Distribution of the total coverage of the k-mer pair (y-axis) against relative minor k-mer coverage (x-axis) providing evidence of diploidy in P. downsi.*