

## **SUPPLEMENTAL MATERIAL**

### **A butterfly pan-genome reveals that a large amount of structural variation underlies the evolution of chromatin accessibility**

#### **Authors**

Angelo A. Ruggieri<sup>1</sup>, Luca Livraghi<sup>2,3</sup>, James J. Lewis<sup>4</sup>, Elizabeth Evans<sup>1</sup>, Francesco Cicconardi<sup>5</sup>, Laura Hebberecht<sup>5</sup>, Yadira Ortiz-Ruiz<sup>1,6</sup>, Stephen H. Montgomery<sup>5</sup>, Alfredo Ghezzi<sup>1</sup>, José Arcadio Rodríguez-Martínez<sup>1</sup>, Chris D. Jiggins<sup>7</sup>, W. Owen McMillan<sup>3</sup>, Brian A. Counterman<sup>8</sup>, Riccardo Papa<sup>1,6</sup> & Steven M. Van Belleghem<sup>1,9</sup>

#### **Affiliations**

<sup>1</sup>Department of Biology, University of Puerto Rico, Rio Piedras, Puerto Rico.

<sup>2</sup>Department of Biological Sciences, The George Washington University, Washington DC, USA.

<sup>3</sup>Smithsonian Tropical Research Institute, Republic of Panama.

<sup>4</sup>Department of Zoology, University of British Columbia, Vancouver, BC, Canada.

<sup>5</sup>School of Biological Sciences, Bristol University, UK.

<sup>6</sup>Molecular Sciences and Research Center, University of Puerto Rico, San Juan, PR.

<sup>7</sup>Department of Zoology, University of Cambridge, Cambridge, UK.

<sup>8</sup>Department of Biological Sciences, Auburn University, Auburn, Alabama, USA.

<sup>9</sup>Ecology, Evolution and Conservation Biology, Biology Department, KU Leuven, Leuven, Belgium.

## TABLE OF CONTENTS

|                                                                                                                                                                                                                      |           |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>SUPPLEMENTAL TABLES .....</b>                                                                                                                                                                                     | <b>3</b>  |
| Supplemental Table S1. Genome assembly sizes, N50 and BUSCO completeness. ....                                                                                                                                       | 3         |
| Supplemental Table S2. Lineage-specific sequence composition. ....                                                                                                                                                   | 3         |
| Supplemental Table S3. Genes identified in <i>H. erato</i> that are absent in <i>H. charithonia</i> and/or <i>H. melpomene</i> . ....                                                                                | 4         |
| Supplemental Table S4. Genes with an ATAC-seq peak within 500 bp from the transcription start sites that is within a lineage-specific transposable element (TE) insertion. ....                                      | 10        |
| Supplemental Table S5. Genes near lineage-specific ATAC-seq peaks with head-specific accessibility and that associate with lineage-specific TE insertions. ....                                                      | 13        |
| Supplemental Table S6. ATAC-seq samples, tissue, origin, and GenBank accession numbers. ....                                                                                                                         | 14        |
| <b>SUPPLEMENTAL FIGURES .....</b>                                                                                                                                                                                    | <b>15</b> |
| Supplemental Fig. S1. Lineage specific transposable element activity. ....                                                                                                                                           | 15        |
| Supplemental Fig. S2. Patterns of lineage-specific sequence distribution and chromosome lengths for SNPs, 1 bp indels, SVs between 2 and 50 bp, SVs larger than 1,000 bp, SVs characterized as TEs and all SVs. .... | 16        |
| Supplemental Fig. S3. Ranked SweepFinder2 Composite Likelihood Ratio (CLR) values for ATAC-seq peaks with and without lineage-specific Transposable Element (TE). ....                                               | 17        |
| Supplemental Fig. S4. Completeness of Transposable Elements (TEs) for TEs associated with lineage-specific ATAC-peaks and TEs with no ATAC-seq peak association. ....                                                | 17        |
| Supplemental Fig. S5. Transposable Element (TE) composition of lineage-specific (i.e., compared to all other species) ATAC-seq peaks within TEs. ....                                                                | 18        |
| Supplemental Fig. S6. Transcription factor (TF) binding motif enrichment among lineage-specific TE insertions with ATAC-seq peaks. ....                                                                              | 19        |
| Supplemental Fig. S7. Example intervals of the pan-genome assembly. ....                                                                                                                                             | 20        |
| <b>SUPPLEMENTAL CODE.....</b>                                                                                                                                                                                        | <b>21</b> |
| Supplemental Code S1. Build pan genome using seq-seq-pan ( <a href="https://gitlab.com/chrjan/seq-seq-pan">https://gitlab.com/chrjan/seq-seq-pan</a> ). ....                                                         | 21        |
| Supplemental Code S2. Python script to create .bed files with sequence blocks present in pan genome for each genome. ....                                                                                            | 22        |
| Supplemental Code S3. Bash commands for merging and intersecting block .bed with bedtools. ....                                                                                                                      | 24        |
| Supplemental Code S4. Calculate pairwise nucleotide differences in homologous sequence between pairs of genomes in the pan genome alignment. ....                                                                    | 26        |
| Supplemental Code S5. Transform .xmfa file to fasta. ....                                                                                                                                                            | 29        |
| Supplemental Code S6. ATAC-seq read mapping, peak calling and counting. ....                                                                                                                                         | 30        |
| Supplemental Code S7. Python code to transform scaffold positions to genome positions for mapping with seq-seq-pan map. ....                                                                                         | 35        |
| Supplemental Code S8. ATAC-seq interval mapping to pan-genome. This code can be modified to map any feature such as TEs to the pan genome as well. ....                                                              | 36        |
| Supplemental Code S9. ATAC-seq .bedgraph mapping to pan genome. ....                                                                                                                                                 | 38        |
| Supplemental Code S10. R code to plot any interval of the pan genome with ATAC-seq signals and lineage-specific features. The current example is for the gene <i>chiffon</i> . ....                                  | 42        |

## SUPPLEMENTAL TABLES

**Supplemental Table S1. Genome assembly sizes, N50 and BUSCO completeness.** Busco: C = complete, S = complete and single-copy, D = complete and duplicated, F = fragmented, M = missing.

| Species                       | Assembly size (Mb) | N50 (kb) | BUSCO completeness                        |
|-------------------------------|--------------------|----------|-------------------------------------------|
| <i>H. melpomene</i>           | 275.3              | 14,309   | C:99.3% [S:98.3%, D:1.0%], F:0.3%, M:0.4% |
| <i>H. erato</i>               | 382.8              | 10,689   | C:98.8% [S:97.6%, D:1.2%], F:0.2%, M:1.0% |
| <i>H. charithonia</i> (hap 1) | 355.2              | 11,106   | C:98.9% [S:97.9%, D:1.0%], F:0.8%, M:0.4% |
| <i>H. charithonia</i> (hap 2) | 361.5              | 11,118   | C:98.2% [S:97.4%, D:0.8%], F:0.8%, M:1.0% |
| Homologous (pan-genome)       | 138.6              | NA       | C:94.9% [S:94.8%, D:0.1%], F:1.4%, M:3.7% |

**Supplemental Table S2. Lineage-specific sequence composition.** Lineage-specific (unique) sequences were characterized as being 1 bp indels, shorter than 50 bp (< 50 bp), larger than 1,000 bp, Transposable Elements (TE), or duplicated. % genome = percentage of total length of genome, % count = percentage of unique sequence count, % unique = percentage of unique sequence base pairs. Total identified includes SVs <50 bp, TEs and duplicated sequences.

|                                                                | Unique<br>sequence in<br>Mb<br>(% genome) | 1 bp in Mb<br>(% genome / %<br>count / % unique) | < 50 bp in Mb<br>(% genome / %<br>count / % unique) | > 1000 bp in Mb<br>(% genome / %<br>count / % unique) | TE<br>(%<br>unique) | Dupl.<br>(%<br>unique) | Total<br>identified<br>(%<br>unique) |
|----------------------------------------------------------------|-------------------------------------------|--------------------------------------------------|-----------------------------------------------------|-------------------------------------------------------|---------------------|------------------------|--------------------------------------|
| <i>H. charithonia</i><br>(indels)                              | 72.7<br>(18.0)                            | 0.74<br>(0.18 / 25.67 / 1.02)                    | 10.30<br>(2.56 / 98.09 / 14.26)                     | 54.24<br>(13.45 / 0.21 / 74.56)                       | 27.73               | 5.21                   | 47.20                                |
| <i>H. charithonia</i><br>(compared to<br><i>H. erato</i> )     | 175.2<br>(43.45)                          | 0.16<br>(0.04 / 5.76 / 0.09)                     | 18.28<br>(4.53 / 89.80 / 10.41)                     | 105.34<br>(26.12 / 0.42 / 59.99)                      | 27.34               | 6.17                   | 43.92                                |
| <i>H. erato</i><br>(compared to<br><i>H. charithonia</i> )     | 151.21<br>(39.49)                         | 0.21<br>(0.05 / 7.11 / 0.14)                     | 20.59<br>(5.44 / 87.68 / 13.62)                     | 63.98<br>(16.71 / 0.71 / 42.31)                       | 57.74               | 10.66                  | 82.02                                |
| <i>H. charithonia</i><br>(compared to<br><i>H. melpomene</i> ) | 252.7<br>(62.67)                          | 0.27<br>(0.07 / 6.90 / 0.11)                     | 28.61<br>(7.10 / 79.51 / 11.32)                     | 106.30<br>(26.36 / 0.31 / 42.06)                      | 44.81               | 9.10                   | 65.23                                |
| <i>H. erato</i><br>(compared to<br><i>H. melpomene</i> )       | 222.12<br>(58.02)                         | 0.17<br>(0.04 / 4.48 / 0.08)                     | 29.90<br>(7.89 / 78.02 / 13.46)                     | 65.40<br>(17.26 / 0.56 / 29.44)                       | 54.87               | 7.12                   | 75.54                                |
| <i>H. melpomene</i><br>(compared to<br><i>H. charithonia</i> ) | 113.5<br>(41.24)                          | 0.17<br>(0.06 / 6.14 / 0.15)                     | 18.80<br>(7.12 / 85.49 / 16.57)                     | 29.22<br>(11.07 / 0.44 / 25.74)                       | 47.59               | 5.29                   | 69.45                                |
| <i>H. melpomene</i><br>(compared to<br><i>H. erato</i> )       | 106.83<br>(38.82)                         | 0.13<br>(0.05 / 4.93 / 0.12)                     | 16.37<br>(6.20 / 86.20 / 15.32)                     | 29.01<br>(10.99 / 0.46 / 27.15)                       | 48.45               | 5.29                   | 69.06                                |

**Supplemental Table S3. Genes identified in *H. erato* that are absent in *H. charithonia* and/or *H. melpomene*.** In the Function column, 'TE' indicates genes related to transposable elements, '?' are genes with a question mark are uncharacterized, 'Chemo' indicates genes with a chemosensing function, 'Neuro' indicates genes related to neural activity, and 'TE repr' indicates genes with a potential function in TE repression. Gene IDs are relative to the *H. erato* genome assembly.

| Gene ID                  | Absent in <i>H. charithonia</i> | Absent in <i>H. melpomene</i> | Gene Description                                                            | Function |
|--------------------------|---------------------------------|-------------------------------|-----------------------------------------------------------------------------|----------|
| evm.model.Herato0101.210 | X                               | X                             | vacuolar protein sorting-associated protein 11 homolog                      |          |
| evm.model.Herato0101.211 | X                               | X                             | uncharacterized protein LOC113391951 isoform X2                             | ?        |
| evm.model.Herato0101.243 | X                               | X                             | papilin-like isoform X5                                                     |          |
| evm.model.Herato0101.251 | X                               | X                             | Periodic tryptophan protein 1-like                                          |          |
| evm.model.Herato0101.389 | X                               | X                             | endonuclease-reverse transcriptase HmRTE-e01                                | TE       |
| evm.model.Herato0101.573 | X                               | X                             | phospholipase A1-like                                                       |          |
| evm.model.Herato0101.574 | X                               | X                             | actin-related protein 2/3 complex subunit 5                                 |          |
| evm.model.Herato0101.576 | X                               | X                             | probable chitinase 2                                                        |          |
| evm.model.Herato0101.675 | X                               | X                             | Cyclic nucleotide-gated cation channel alpha-3                              |          |
| evm.model.Herato0101.677 | X                               | X                             | cysteine protease ATG4B                                                     |          |
| evm.model.Herato0101.697 | X                               | X                             | transferrin-like                                                            |          |
| evm.model.Herato0101.698 | X                               | X                             | trypsin, alkaline C-like                                                    |          |
| evm.model.Herato0101.855 | X                               | X                             | uncharacterized protein LOC111360937                                        | ?        |
| evm.model.Herato0204.28  | X                               | X                             | ---NA---                                                                    | ?        |
| evm.model.Herato0206.40  | X                               | X                             | programmed cell death protein 10                                            |          |
| evm.model.Herato0211.18  | X                               | X                             | uricase                                                                     |          |
| evm.model.Herato0211.99  | X                               | X                             | RNA-directed DNA polymerase from mobile element jockey-like isoform X1      | TE       |
| evm.model.Herato0212.1   | X                               | X                             | Retrovirus-related Pol polyprotein from type-1 retrotransposable element R1 | TE       |
| evm.model.Herato0212.9   | X                               | X                             | ---NA---                                                                    | ?        |
| evm.model.Herato0215.21  | X                               | X                             | aldo-keto reductase AKR2E4-like                                             | Chemo    |
| evm.model.Herato0215.77  | X                               | X                             | protein ANTAGONIST OF LIKE HETEROCHROMATIN PROTEIN 1-like                   |          |
| evm.model.Herato0215.78  | X                               | X                             | uncharacterized protein LOC113394144                                        | ?        |
| evm.model.Herato0215.80  | X                               | X                             | uncharacterized protein LOC113394144                                        | ?        |
| evm.model.Herato0301.123 | X                               | X                             | hypothetical protein EVAR_26063_1                                           | ?        |
| evm.model.Herato0301.94  | X                               | X                             | dual specificity mitogen-activated protein kinase kinase 1-like             |          |
| evm.model.Herato0304.1   | X                               | X                             | Retrovirus-related Pol polyprotein                                          | TE       |
| evm.model.Herato0309.1   | X                               | X                             | uricase                                                                     |          |
| evm.model.Herato0502.4   | X                               | X                             | axoneme-associated protein mst101(2)-like                                   |          |
| evm.model.Herato0503.287 | X                               | X                             | uncharacterized protein LOC111354683                                        | ?        |
| evm.model.Herato0503.339 | X                               | X                             | zinc finger BED domain-containing protein 1-like                            | TE repr? |
| evm.model.Herato0503.73  | X                               | X                             | beta-1,4-N-acetylgalactosaminyltransferase bre-4-like                       |          |
| evm.model.Herato0508.22  | X                               | X                             | Protein YIF1B-B                                                             | Neuro    |
| evm.model.Herato0508.59  | X                               | X                             | ---NA---                                                                    | ?        |
| evm.model.Herato0509.3   | X                               | X                             | delta(14)-sterol reductase                                                  |          |
| evm.model.Herato0511.4   | X                               | X                             | uncharacterized protein LOC112054750                                        | ?        |
| evm.model.Herato0601.1   | X                               | X                             | Kv channel-interacting protein 4-like                                       |          |
| evm.model.Herato0601.103 | X                               | X                             | RNA-directed DNA polymerase from mobile element jockey-like                 | TE       |
| evm.model.Herato0606.265 | X                               | X                             | Uncharacterized protein OBRU01_23876                                        | ?        |
| evm.model.Herato0606.3   | X                               | X                             | Nucleic-acid-binding protein from transposon X-element                      | TE       |
| evm.model.Herato0606.306 | X                               | X                             | RND multidrug efflux transporter                                            |          |
| evm.model.Herato0606.443 | X                               | X                             | enoyl-CoA delta isomerase 1, mitochondrial-like                             |          |
| evm.model.Herato0608.1   | X                               | X                             | programmed cell death protein 10                                            |          |
| evm.model.Herato0608.4   | X                               | X                             | ---NA---                                                                    | ?        |
| evm.model.Herato0609.30  | X                               | X                             | ---NA---                                                                    | ?        |
| evm.model.Herato0701.11  | X                               | X                             | Putative 115 kDa protein in type-1 retrotransposable element R1DM           | TE       |
| evm.model.Herato0701.134 | X                               | X                             | beta-1,4-N-acetylgalactosaminyltransferase bre-4-like                       |          |
| evm.model.Herato0701.241 | X                               | X                             | ---NA---                                                                    | ?        |
| evm.model.Herato0701.311 | X                               | X                             | uncharacterized protein LOC115444529 isoform X1                             | ?        |
| evm.model.Herato0701.371 | X                               | X                             | RNA-directed DNA polymerase from mobile element jockey-like isoform X1      | TE       |
| evm.model.Herato0701.415 | X                               | X                             | E3 ubiquitin-protein ligase MGRN1                                           |          |
| evm.model.Herato0701.724 | X                               | X                             | unnamed protein product                                                     | ?        |

|                          |   |   |                                                                   |         |
|--------------------------|---|---|-------------------------------------------------------------------|---------|
| evm.model.Herato0701.84  | X | X | ---NA---                                                          | ?       |
| evm.model.Herato0801.221 | X | X | endonuclease-reverse transcriptase                                | TE      |
| evm.model.Herato0801.62  | X | X | nuclear pore complex protein Nup153 isoform X1                    |         |
| evm.model.Herato0802.1   | X | X | Gag-like protein                                                  | TE      |
| evm.model.Herato0821.52  | X | X | ATP synthase subunit e, mitochondrial                             |         |
| evm.model.Herato0901.156 | X | X | hypothetical protein                                              | ?       |
| evm.model.Herato0901.170 | X | X | RNA-directed DNA polymerase from mobile element jockey-like       | TE      |
| evm.model.Herato0901.209 | X | X | tigger transposable element-derived protein 2-like                | TE      |
| evm.model.Herato0901.254 | X | X | Dihydrouridine synthase domain containing protein                 |         |
| evm.model.Herato0901.304 | X | X | Retrovirus-related Pol polyprotein LINE-1                         | TE      |
| evm.model.Herato0901.64  | X | X | general odorant-binding protein 56a-like                          | Chemo   |
| evm.model.Herato0902.1   | X | X | Putative non-capsid protein NS-1                                  | TE      |
| evm.model.Herato0902.3   | X | X | NIF3-like protein 1 isoform X3                                    |         |
| evm.model.Herato1001.128 | X | X | phospholipase A1-like                                             |         |
| evm.model.Herato1001.134 | X | X | uncharacterized protein LOC110385166                              | ?       |
| evm.model.Herato1001.164 | X | X | reverse transcriptase                                             | TE      |
| evm.model.Herato1003.178 | X | X | hypothetical protein EVAR_58729_1                                 | ?       |
| evm.model.Herato1003.242 | X | X | Gag protein                                                       | TE      |
| evm.model.Herato1003.31  | X | X | chymotrypsin-2-like                                               |         |
| evm.model.Herato1003.61  | X | X | endonuclease-reverse transcriptase HmRTE-e01                      | TE      |
| evm.model.Herato1003.97  | X | X | RNA-directed DNA polymerase from mobile element jockey-like       | TE      |
| evm.model.Herato1007.208 | X | X | RNA-directed DNA polymerase from mobile element jockey            | TE      |
| evm.model.Herato1108.1   | X | X | Gag-like protein                                                  | TE      |
| evm.model.Herato1108.407 | X | X | ---NA---                                                          | ?       |
| evm.model.Herato1108.411 | X | X | general transcription factor 3C polypeptide 6-like isoform X2     | TE repr |
| evm.model.Herato1108.414 | X | X | general transcription factor 3C polypeptide 6-like isoform X2     | TE repr |
| evm.model.Herato1116.2   | X | X | unnamed protein product                                           | ?       |
| evm.model.Herato1202.117 | X | X | ATP-dependent DNA helicase pif1                                   |         |
| evm.model.Herato1202.232 | X | X | Replicase large subunit                                           | TE      |
| evm.model.Herato1202.362 | X | X | ---NA---                                                          | ?       |
| evm.model.Herato1301.1   | X | X | chymotrypsin-like elastase family member 2A                       |         |
| evm.model.Herato1301.126 | X | X | ---NA---                                                          | ?       |
| evm.model.Herato1301.724 | X | X | ---NA---                                                          | ?       |
| evm.model.Herato1301.9   | X | X | ---NA---                                                          | ?       |
| evm.model.Herato1401.5   | X | X | Transposon Tf2-6 polyprotein                                      | TE      |
| evm.model.Herato1411.195 | X | X | Gag-like protein                                                  | TE      |
| evm.model.Herato1411.231 | X | X | endonuclease-reverse transcriptase                                | TE      |
| evm.model.Herato1504.1   | X | X | CLUMA_CG001710, isoform A                                         |         |
| evm.model.Herato1504.4   | X | X | uncharacterized protein LOC111363830                              | ?       |
| evm.model.Herato1506.1   | X | X | Nucleic-acid-binding protein from transposon X-element            | TE      |
| evm.model.Herato1507.140 | X | X | hypothetical protein                                              | ?       |
| evm.model.Herato1507.312 | X | X | putative dihydrouridine synthase domain containing protein        |         |
| evm.model.Herato1507.341 | X | X | uncharacterized protein LOC111364243, partial                     | ?       |
| evm.model.Herato1507.55  | X | X | uncharacterized protein LOC116776120                              | ?       |
| evm.model.Herato1507.60  | X | X | protein D2-like isoform X2                                        |         |
| evm.model.Herato1601.123 | X | X | RNA-directed DNA polymerase from mobile element jockey-like       | TE      |
| evm.model.Herato1601.71  | X | X | RNA-directed DNA polymerase from mobile element jockey            | TE      |
| evm.model.Herato1603.3   | X | X | nuclear pore complex protein Nup153 isoform X1                    |         |
| evm.model.Herato1605.113 | X | X | carosine N-methyltransferase                                      |         |
| evm.model.Herato1605.169 | X | X | Putative 115 kDa protein in type-1 retrotransposable element R1DM | TE      |
| evm.model.Herato1701.28  | X | X | ---NA---                                                          | ?       |
| evm.model.Herato1701.31  | X | X | Replicase large subunit                                           | TE      |
| evm.model.Herato1701.6   | X | X | hypothetical protein evm_015023, partial                          | ?       |
| evm.model.Herato1705.37  | X | X | Protein YIF1B-B                                                   | Neuro   |
| evm.model.Herato1705.52  | X | X | arginine--tRNA ligase                                             |         |
| evm.model.Herato1705.53  | X | X | uncharacterized protein LOC113393653 isoform X1                   | ?       |
| evm.model.Herato1705.54  | X | X | ATP synthase subunit e, mitochondrial                             |         |
| evm.model.Herato1706.8   | X | X | Dihydrouridine synthase domain containing protein                 |         |

|                               |   |   |                                                                                |          |
|-------------------------------|---|---|--------------------------------------------------------------------------------|----------|
| evm.model.Herato1707.16       | X | X | hypothetical protein RR48_01322                                                | ?        |
| evm.model.Herato1707.17       | X | X | protein ALP1-like                                                              | Neuro    |
| evm.model.Herato1708.1        | X | X | snRNA-activating protein complex subunit 4                                     |          |
| evm.model.Herato1708.109      | X | X | zinc finger protein 26-like                                                    | TE repr? |
| evm.model.Herato1708.110      | X | X | GDNF-inducible zinc finger protein 1-like                                      | TE repr  |
| evm.model.Herato1708.111      | X | X | zinc finger protein 277                                                        | TE repr? |
| evm.model.Herato1708.112      | X | X | GDNF-inducible zinc finger protein 1-like                                      | TE repr  |
| evm.model.Herato1708.113      | X | X | zinc finger and SCAN domain-containing protein 2-like                          | TE repr? |
| evm.model.Herato1711.3        | X | X | RNA-directed DNA polymerase from mobile element jockey                         | TE       |
| evm.model.Herato1712.3        | X | X | unnamed protein product, partial                                               | ?        |
| evm.model.Herato1712.4        | X | X | ---NA---                                                                       | ?        |
| evm.model.Herato1712.6        | X | X | piggyBac transposable element-derived protein 2-like                           | TE       |
| evm.model.Herato1713.9        | X | X | endonuclease/exonuclease/phosphatase family protein                            | TE       |
| evm.model.Herato1801.127      | X | X | glucose dehydrogenase [FAD, quinone]-like                                      |          |
| evm.model.Herato1801.128      | X | X | cysteine protease ATG4B                                                        |          |
| evm.model.Herato1801.129      | X | X | glutathione S-transferase 1-1-like isoform X5                                  |          |
| evm.model.Herato1804.2        | X | X | putative inorganic phosphate cotransporter                                     |          |
| evm.model.Herato1805.64       | X | X | hypothetical protein evm_015023, partial                                       | ?        |
| evm.model.Herato1901.24       | X | X | hypothetical protein KGM_200003                                                | ?        |
| evm.model.Herato1901.3        | X | X | R2DM Retrovirus-related Pol polyprotein from type II retrotransposable element | TE       |
| evm.model.Herato1901.31       | X | X | reverse transcriptase                                                          | TE       |
| evm.model.Herato1901.44       | X | X | Retrovirus-related Pol polyprotein from transposon opus-like Protein           | TE       |
| evm.model.Herato1904.1        | X | X | 5-hydroxytryptamine receptor 1-like                                            |          |
| evm.model.Herato1904.11       | X | X | pyridoxal-dependent decarboxylase domain-containing protein 1                  |          |
| evm.model.Herato1904.99       | X | X | vegetative cell wall protein gp1-like                                          |          |
| evm.model.Herato1908.58       | X | X | thioredoxin-2                                                                  |          |
| evm.model.Herato1910.138      | X | X | facilitated trehalose transporter Tret1-like                                   |          |
| evm.model.Herato1910.140      | X | X | putative odorant receptor OR30                                                 | Chemo    |
| evm.model.Herato1910.157      | X | X | neurofilament medium polypeptide-like                                          | Neuro    |
| evm.model.Herato1910.182      | X | X | ---NA---                                                                       | ?        |
| evm.model.Herato1910.229      | X | X | facilitated trehalose transporter Tret1-like                                   |          |
| evm.model.Herato2001.262      | X | X | lymphocyte expansion molecule-like                                             |          |
| evm.model.Herato2001.295      | X | X | uncharacterized protein LOC113399997 isoform X2                                | ?        |
| evm.model.Herato2001.4        | X | X | hypothetical protein                                                           | ?        |
| evm.model.Herato2001.487      | X | X | neurofilament medium polypeptide-like                                          | Neuro    |
| evm.model.Herato2001.492      | X | X | putative RNA-directed DNA polymerase from transposon X-element                 | TE       |
| evm.model.Herato2001.497      | X | X | Dyak/GE14720-PA-like protein                                                   |          |
| evm.model.Herato2001.736      | X | X | uncharacterized protein LOC115444529 isoform X1                                | ?        |
| evm.model.Herato2101.112      | X | X | distal membrane-arm assembly complex protein 2                                 |          |
| evm.model.Herato2101.343      | X | X | nipped-B protein                                                               |          |
| evm.model.Herato2101.344      | X | X | Kv channel-interacting protein 4-like                                          |          |
| evm.model.Herato2101.346      | X | X | NPC intracellular cholesterol transporter 1                                    |          |
| evm.model.Herato2101.437      | X | X | ---NA---                                                                       | ?        |
| novel_gene_124                | X | X | ---NA---                                                                       | ?        |
| novel_gene_143                | X | X | ---NA---                                                                       | ?        |
| novel_gene_155                | X | X | ---NA---                                                                       | ?        |
| novel_gene_17                 | X | X | ---NA---                                                                       | ?        |
| novel_gene_173 novel_gene_175 | X | X | ---NA---                                                                       | ?        |
| novel_gene_211                | X | X | ---NA---                                                                       | ?        |
| novel_gene_228                | X | X | ---NA---                                                                       | ?        |
| novel_gene_234                | X | X | ---NA---                                                                       | ?        |
| novel_gene_240                | X | X | ---NA---                                                                       | ?        |
| novel_gene_255                | X | X | ---NA---                                                                       | ?        |
| novel_gene_79                 | X | X | ---NA---                                                                       | ?        |
| novel_gene_86                 | X | X | ---NA---                                                                       | ?        |
| novel_gene_96                 | X | X | ---NA---                                                                       | ?        |
| evm.model.Herato0503.326      | X |   | zinc finger protein 567                                                        | TE repr? |
| evm.model.Herato1505.2        | X |   | zinc finger protein 567                                                        | TE repr? |
| evm.model.Herato0101.166      |   | X | hypothetical protein evm_001344                                                | ?        |
| evm.model.Herato0101.575      |   | X | Actin-related protein 2/3 complex subunit 5                                    |          |
| evm.model.Herato0101.578      |   | X | Actin-related protein 2/3 complex subunit 5                                    |          |

|                          |  |   |                                                                                |          |
|--------------------------|--|---|--------------------------------------------------------------------------------|----------|
| evm.model.Herato0101.727 |  | X | integrin beta-6-like                                                           |          |
| evm.model.Herato0101.803 |  | X | Replicase large subunit                                                        | TE       |
| evm.model.Herato0209.15  |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato0211.83  |  | X | Retrovirus-related Pol polyprotein from transposon TNT 1-94                    | TE       |
| evm.model.Herato0212.10  |  | X | transient receptor potential cation channel trpm-like isoform X2               |          |
| evm.model.Herato0213.5   |  | X | Gag-like protein                                                               | TE       |
| evm.model.Herato0214.60  |  | X | gag polyprotein - red flour beetle retrotransposon Woot                        | TE       |
| evm.model.Herato0215.125 |  | X | Replicase large subunit                                                        | TE       |
| evm.model.Herato0215.14  |  | X | unnamed protein product                                                        | ?        |
| evm.model.Herato0215.79  |  | X | putative dihydrouridine synthase domain containing protein                     |          |
| evm.model.Herato0301.121 |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato0302.1   |  | X | cilia- and flagella-associated protein 45-like                                 |          |
| evm.model.Herato0405.26  |  | X | zinc finger BED domain-containing protein RICESLEEPER 1-like isoform X1        | TE repr? |
| evm.model.Herato0405.38  |  | X | mitochondrial inner membrane protein OXA1L                                     |          |
| evm.model.Herato0410.1   |  | X | Indole-3-acetaldehyde oxidase                                                  |          |
| evm.model.Herato0411.134 |  | X | TP53-binding protein 1-like                                                    |          |
| evm.model.Herato0411.141 |  | X | uncharacterized protein K02A2.6-like                                           | ?        |
| evm.model.Herato0411.143 |  | X | HEAT repeat-containing protein 5B-like                                         |          |
| evm.model.Herato0411.3   |  | X | furin-like protease 1 isoform X1                                               |          |
| evm.model.Herato0419.56  |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato0501.1   |  | X | unnamed protein product                                                        | ?        |
| evm.model.Herato0503.284 |  | X | protein ANTAGONIST OF LIKE HETEROCHROMATIN PROTEIN 1-like                      |          |
| evm.model.Herato0503.292 |  | X | enhancer of split mbeta protein-like                                           |          |
| evm.model.Herato0503.293 |  | X | enhancer of split mbeta protein-like                                           |          |
| evm.model.Herato0508.87  |  | X | Protein YIF1B-B                                                                | Neuro    |
| evm.model.Herato0508.89  |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato0605.1   |  | X | nucleic-acid-binding protein from mobile element jockey-like                   | TE       |
| evm.model.Herato0606.154 |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato0606.19  |  | X | reverse transcriptase                                                          | TE       |
| evm.model.Herato0606.220 |  | X | angiominin-like protein 2                                                      |          |
| evm.model.Herato0606.232 |  | X | larval cuticle protein LCP-30-like                                             |          |
| evm.model.Herato0606.242 |  | X | axoneme-associated protein mst101(2)-like                                      |          |
| evm.model.Herato0606.374 |  | X | Replicase large subunit                                                        | TE       |
| evm.model.Herato0606.376 |  | X | hypothetical protein RR48_01857                                                | ?        |
| evm.model.Herato0608.3   |  | X | uncharacterized protein LOC111002208, partial                                  | ?        |
| evm.model.Herato0609.39  |  | X | R2DM Retrovirus-related Pol polyprotein from type II retrotransposable element | TE       |
| evm.model.Herato0701.310 |  | X | S-antigen protein-like                                                         |          |
| evm.model.Herato0701.343 |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato0701.414 |  | X | craniofacial development protein 2-like                                        |          |
| evm.model.Herato0701.76  |  | X | Retrovirus-related Pol polyprotein LINE-1                                      | TE       |
| evm.model.Herato0801.103 |  | X | tigger transposable element-derived protein 6-like isoform X2                  | TE       |
| evm.model.Herato0801.136 |  | X | ubiquitin carboxyl-terminal hydrolase 34                                       |          |
| evm.model.Herato0810.2   |  | X | glucose dehydrogenase [FAD, quinone]-like                                      |          |
| evm.model.Herato0901.163 |  | X | Glycerophosphodiester phosphodiesterase 1                                      |          |
| evm.model.Herato0901.213 |  | X | reverse transcriptase                                                          | TE       |
| evm.model.Herato0901.258 |  | X | reverse transcriptase family protein                                           | TE       |
| evm.model.Herato0901.348 |  | X | reverse transcriptase family protein                                           | TE       |
| evm.model.Herato0902.5   |  | X | putative nuclease HARB11                                                       | TE       |
| evm.model.Herato1001.1   |  | X | unnamed protein product                                                        | ?        |
| evm.model.Herato1001.127 |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato1001.2   |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato1003.116 |  | X | DNA polymerase epsilon catalytic subunit A                                     |          |
| evm.model.Herato1003.117 |  | X | R2DM Retrovirus-related Pol polyprotein from type II retrotransposable element | TE       |
| evm.model.Herato1003.131 |  | X | ataxia telangiectasia mutated family protein                                   |          |
| evm.model.Herato1003.177 |  | X | trypsin CFT-1-like                                                             |          |
| evm.model.Herato1003.200 |  | X | trypsin, alkaline C-like                                                       |          |
| evm.model.Herato1003.201 |  | X | trypsin, alkaline C-like                                                       |          |

|                          |  |   |                                                                                |          |
|--------------------------|--|---|--------------------------------------------------------------------------------|----------|
| evm.model.Herato1003.33  |  | X | R2DM Retrovirus-related Pol polyprotein from type II retrotransposable element | TE       |
| evm.model.Herato1003.51  |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato1003.57  |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato1005.95  |  | X | zinc finger BED domain-containing protein 1-like                               | TE repr? |
| evm.model.Herato1007.17  |  | X | Pheromone-binding protein-related protein 1                                    | Chemo    |
| evm.model.Herato1103.1   |  | X | zinc finger DNA binding protein                                                | TE repr? |
| evm.model.Herato1109.2   |  | X | Gag-like protein                                                               | TE       |
| evm.model.Herato1201.18  |  | X | hypothetical protein evm_015023, partial                                       | ?        |
| evm.model.Herato1202.260 |  | X | PiggyBac transposable element-derived protein 4                                | TE       |
| evm.model.Herato1202.576 |  | X | piggyBac transposable element-derived protein 4-like                           | TE       |
| evm.model.Herato1202.830 |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato1301.125 |  | X | piggyBac transposable element-derived protein 3-like                           | TE       |
| evm.model.Herato1301.360 |  | X | cytochrome c-type heme lyase                                                   |          |
| evm.model.Herato1301.371 |  | X | Retrovirus-related Pol polyprotein LINE-1                                      | TE       |
| evm.model.Herato1301.372 |  | X | craniofacial development protein 2-like                                        |          |
| evm.model.Herato1301.482 |  | X | Replicase large subunit                                                        | TE       |
| evm.model.Herato1301.498 |  | X | endonuclease-reverse transcriptase                                             | TE       |
| evm.model.Herato1301.526 |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato1301.558 |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato1301.665 |  | X | Nucleic-acid-binding protein from transposon X-element                         | TE       |
| evm.model.Herato1301.709 |  | X | glucose dehydrogenase [FAD, quinone]-like                                      |          |
| evm.model.Herato1301.809 |  | X | hypothetical protein EVAR_2411_1                                               | ?        |
| evm.model.Herato1408.111 |  | X | centromere protein X-like                                                      |          |
| evm.model.Herato1408.113 |  | X | NADH dehydrogenase [ubiquinone] iron-sulfur protein 5-like                     |          |
| evm.model.Herato1408.114 |  | X | centromere protein X-like                                                      |          |
| evm.model.Herato1411.232 |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato1411.248 |  | X | ubiquitin-like modifier-activating enzyme ATG7                                 |          |
| evm.model.Herato1411.91  |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato1507.141 |  | X | probable tubulin polyglutamylase TTL1                                          |          |
| evm.model.Herato1507.276 |  | X | R2DM Retrovirus-related Pol polyprotein from type II retrotransposable element | TE       |
| evm.model.Herato1507.313 |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato1507.334 |  | X | myb/SANT-like DNA-binding domain-containing protein 4                          |          |
| evm.model.Herato1512.1   |  | X | mucin-6                                                                        |          |
| evm.model.Herato1512.4   |  | X | uncharacterized protein LOC113395066                                           | ?        |
| evm.model.Herato1515.1   |  | X | hypothetical protein                                                           | ?        |
| evm.model.Herato1601.55  |  | X | Putative 115 kDa protein in type-1 retrotransposable element R1DM              | TE       |
| evm.model.Herato1603.1   |  | X | uncharacterized protein LOC116776805                                           | ?        |
| evm.model.Herato1603.34  |  | X | serine/arginine repetitive matrix protein 2-like                               |          |
| evm.model.Herato1603.39  |  | X | serine/arginine repetitive matrix protein 2-like                               |          |
| evm.model.Herato1603.82  |  | X | uncharacterized protein LOC112056053                                           | ?        |
| evm.model.Herato1604.1   |  | X | polyhomeotic-like protein 1                                                    |          |
| evm.model.Herato1605.104 |  | X | R2DM Retrovirus-related Pol polyprotein from type II retrotransposable element | TE       |
| evm.model.Herato1605.26  |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato1702.4   |  | X | uncharacterized protein LOC113397096                                           | ?        |
| evm.model.Herato1703.42  |  | X | Retrovirus-related Pol polyprotein from type-1 retrotransposable element R1_4  | TE       |
| evm.model.Herato1704.20  |  | X | reverse transcriptase                                                          | TE       |
| evm.model.Herato1705.11  |  | X | Gag-like protein                                                               | TE       |
| evm.model.Herato1705.62  |  | X | protein unc-13 homolog C-like                                                  |          |
| evm.model.Herato1707.1   |  | X | BESS domain-containing protein                                                 |          |
| evm.model.Herato1708.29  |  | X | transcription factor Adf-1-like                                                |          |
| evm.model.Herato1710.4   |  | X | uncharacterized protein LOC115878124                                           | ?        |
| evm.model.Herato1710.5   |  | X | ATP-binding cassette sub-family A member 3                                     |          |
| evm.model.Herato1710.6   |  | X | ATP-binding cassette sub-family A member 3                                     |          |
| evm.model.Herato1713.8   |  | X | RNA-directed DNA polymerase from mobile element jockey                         | TE       |
| evm.model.Herato1715.1   |  | X | SW15-dependent HO expression protein 3-like                                    |          |
| evm.model.Herato1716.2   |  | X | ---NA---                                                                       | ?        |
| evm.model.Herato1804.6   |  | X | nuclear receptor-binding factor 2                                              |          |
| evm.model.Herato1804.7   |  | X | phosphatidylethanolamine-binding protein 4-like                                |          |
| evm.model.Herato1804.8   |  | X | nuclear receptor-binding factor 2                                              |          |

|                               |  |   |                                                             |          |
|-------------------------------|--|---|-------------------------------------------------------------|----------|
| evm.model.Herato1804.9        |  | X | phosphatidylethanolamine-binding protein 4-like             |          |
| evm.model.Herato1805.276      |  | X | uncharacterized protein LOC110373945                        | ?        |
| evm.model.Herato1807.103      |  | X | uncharacterized protein LOC113494301                        | ?        |
| evm.model.Herato1901.38       |  | X | hypothetical protein EVAR_5591_1                            | ?        |
| evm.model.Herato1901.43       |  | X | uncharacterized protein K02A2.6-like                        | ?        |
| evm.model.Herato1904.127      |  | X | ---NA---                                                    | ?        |
| evm.model.Herato1904.211      |  | X | RNA-directed DNA polymerase from mobile element jockey-like | TE       |
| evm.model.Herato1908.57       |  | X | trypsin, alkaline C-like                                    |          |
| evm.model.Herato1910.1        |  | X | neurofilament medium polypeptide-like                       | Neuro    |
| evm.model.Herato1910.183      |  | X | ---NA---                                                    | ?        |
| evm.model.Herato1910.190      |  | X | RNA-directed DNA polymerase from mobile element jockey-like | TE       |
| evm.model.Herato2001.491      |  | X | Nucleic-acid-binding protein from transposon X-element      | TE       |
| evm.model.Herato2001.589      |  | X | ---NA---                                                    | ?        |
| evm.model.Herato2101.144      |  | X | Pre-mRNA-splicing factor ISY1 protein                       |          |
| evm.model.Herato2101.145      |  | X | uncharacterized protein LOC111000646                        | ?        |
| evm.model.Herato2101.146      |  | X | Pre-mRNA-splicing factor ISY1 protein                       |          |
| evm.model.Herato2101.290      |  | X | serine/threonine-protein kinase tricornet isoform 3         |          |
| evm.model.Herato2101.345      |  | X | peptidyl-prolyl cis-trans isomerase-like 1                  |          |
| evm.model.Herato2101.347      |  | X | lymphocyte expansion molecule-like                          |          |
| evm.model.Herato2101.398      |  | X | UNC93-like protein                                          |          |
| evm.model.Herato2101.516      |  | X | zinc finger protein 729-like                                | TE repr? |
| novel_gene_104                |  | X | ---NA---                                                    | ?        |
| novel_gene_120                |  | X | ---NA---                                                    | ?        |
| novel_gene_149                |  | X | ---NA---                                                    | ?        |
| novel_gene_169                |  | X | ---NA---                                                    | ?        |
| novel_gene_174                |  | X | ---NA---                                                    | ?        |
| novel_gene_198 novel_gene_199 |  | X | ---NA---                                                    | ?        |
| novel_gene_200                |  | X | ---NA---                                                    | ?        |
| novel_gene_24                 |  | X | ---NA---                                                    | ?        |
| novel_gene_61                 |  | X | ---NA---                                                    | ?        |
| novel_gene_64                 |  | X | ---NA---                                                    | ?        |

**Supplemental Table S4. Genes with an ATAC-seq peak within 500 bp from the transcription start sites that is within a lineage-specific transposable element (TE) insertion.** Gene IDs are relative to the *H. erato* genome assembly.

| Gene ID                                           | Gene description                                                           |
|---------------------------------------------------|----------------------------------------------------------------------------|
| evm.model.Herato0101.133                          | Golgi pH regulator                                                         |
| evm.model.Herato0101.339                          | JmjC domain-containing protein 4                                           |
| evm.model.Herato0101.371                          | pre-mRNA-splicing factor SYF1                                              |
| evm.model.Herato0101.391                          | alpha-tocopherol transfer protein-like isoform X1                          |
| evm.model.Herato0101.42                           | DNA polymerase epsilon subunit 3                                           |
| evm.model.Herato0101.44                           | CCAAT/enhancer-binding protein zeta                                        |
| evm.model.Herato0101.484                          | trifunctional enzyme subunit alpha, mitochondrial                          |
| evm.model.Herato0101.511                          | exonuclease mut-7 homolog                                                  |
| evm.model.Herato0101.564                          | katanin p80 WD40 repeat-containing subunit B1 isoform X1                   |
| evm.model.Herato0101.641                          | alpha-tocopherol transfer protein-like                                     |
| evm.model.Herato0101.654                          | long-chain fatty acid transport protein 4-like                             |
| evm.model.Herato0101.692                          | abhydrolase domain-containing protein 2                                    |
| evm.model.Herato0101.749                          | eukaryotic translation initiation factor 3 subunit H                       |
| evm.model.Herato0101.768                          | RNA pseudouridylation synthase domain-containing protein 2-like isoform X3 |
| evm.model.Herato0101.77                           | general transcriptional corepressor trfA-like                              |
| evm.model.Herato0101.807                          | type II inositol 1,4,5-trisphosphate 5-phosphatase                         |
| evm.model.Herato0101.820                          | probable H/ACA ribonucleoprotein complex subunit 1                         |
| evm.model.Herato0204.24                           | PHD finger protein 12                                                      |
| evm.model.Herato0204.47                           | cellular tumor antigen p53                                                 |
| evm.model.Herato0211.10                           | eukaryotic translation initiation factor 3 subunit C                       |
| evm.model.Herato0211.12                           | histone deacetylase 8-like                                                 |
| evm.model.Herato0214.2 evm.model.Herato0214.3     | nardilysin-like                                                            |
| evm.model.Herato0214.30                           | leucine-zipper-like transcriptional regulator 1                            |
| evm.model.Herato0215.99                           | reverse transcriptase                                                      |
| evm.model.Herato0301.115                          | chromobox protein homolog 1-like                                           |
| evm.model.Herato0301.70                           | uncharacterized protein LOC112050341                                       |
| evm.model.Herato0310.106                          | protein-lysine methyltransferase METTL21D-like                             |
| evm.model.Herato0310.147                          | prefoldin subunit 6                                                        |
| evm.model.Herato0310.257                          | protein Peter pan                                                          |
| evm.model.Herato0310.271                          | AP-1 complex subunit sigma-2 isoform X2                                    |
| evm.model.Herato0310.53                           | probable ATP-dependent RNA helicase pitchoun                               |
| evm.model.Herato0401.5                            | zwei Ig domain protein zig-8-like                                          |
| evm.model.Herato0405.45                           | alpha-tocopherol transfer protein-like                                     |
| evm.model.Herato0411.3                            | furin-like protease 1 isoform X1                                           |
| evm.model.Herato0501.33                           | transcription initiation factor TFIID subunit 9                            |
| evm.model.Herato0503.212                          | protein split ends                                                         |
| evm.model.Herato0503.247                          | polypeptide N-acetylglucosaminyltransferase 5                              |
| evm.model.Herato0508.53                           | ATPase family AAA domain-containing protein 1-like isoform X2              |
| evm.model.Herato0511.7                            | hypothetical protein B5V51_8296                                            |
| evm.model.Herato0606.136                          | cell division cycle and apoptosis regulator protein 1-like                 |
| evm.model.Herato0606.155                          | transcription initiation factor TFIID subunit 13 isoform X1                |
| evm.model.Herato0606.255                          | eukaryotic translation initiation factor 4 gamma-like                      |
| evm.model.Herato0606.372                          | ---NA---                                                                   |
| evm.model.Herato0606.394 evm.model.Herato0606.395 | bromodomain-containing protein 8-like                                      |
| evm.model.Herato0606.397                          | Erg28-domain containing protein                                            |
| evm.model.Herato0606.399                          | solute carrier family 25 member 44                                         |
| evm.model.Herato0606.43                           | ATP-dependent DNA helicase DDX11                                           |
| evm.model.Herato0606.462                          | uncharacterized protein LOC113391373                                       |
| evm.model.Herato0606.462                          | uncharacterized protein LOC113391373                                       |
| evm.model.Herato0606.481                          | PR domain zinc finger protein 5-like                                       |
| evm.model.Herato0609.19                           | UPF0547 protein C16orf87-like                                              |
| evm.model.Herato0609.6                            | COMM domain-containing protein 5-like                                      |
| evm.model.Herato0609.92                           | ---NA---                                                                   |
| evm.model.Herato0701.128                          | actin-related protein 6                                                    |
| evm.model.Herato0701.129                          | T-complex protein 1 subunit epsilon                                        |
| evm.model.Herato0701.143                          | Peptidyl-prolyl cis-trans isomerase CWC27-like                             |
| evm.model.Herato0701.213                          | steroidogenic acute regulatory protein-like                                |
| evm.model.Herato0701.315                          | glutamic acid-rich protein-like                                            |
| evm.model.Herato0701.353                          | protein lifeguard 1-like isoform X2                                        |
| evm.model.Herato0701.377                          | folliculin                                                                 |
| evm.model.Herato0701.382                          | condensin complex subunit 2                                                |
| evm.model.Herato0701.440                          | mpv17-like protein 2                                                       |
| evm.model.Herato0701.441                          | nuclear receptor coactivator 5 isoform X2                                  |
| evm.model.Herato0701.451                          | protein TSSC4                                                              |
| evm.model.Herato0701.514                          | ganglioside-induced differentiation-associated protein 1                   |

|                                                 |                                                                    |
|-------------------------------------------------|--------------------------------------------------------------------|
| evm.model.Herato0701.536                        | sodium/hydrogen exchanger 8                                        |
| evm.model.Herato0701.592                        | ---NA---                                                           |
| evm.model.Herato0701.660                        | p21-activated protein kinase-interacting protein 1-like            |
| evm.model.Herato0701.716                        | putative leucine-rich repeat-containing protein DDB G0290503       |
| evm.model.Herato0701.87                         | histidine triad nucleotide-binding protein 2, mitochondrial-like   |
| evm.model.Herato0801.150                        | uncharacterized protein LOC117990928                               |
| evm.model.Herato0801.67                         | olfactory receptor 10                                              |
| evm.model.Herato0801.67                         | olfactory receptor 10                                              |
| evm.model.Herato0821.120                        | protein YIF1B                                                      |
| evm.model.Herato0901.144                        | D-tyrosyl-tRNA(Tyr) deacylase 1                                    |
| evm.model.Herato0901.197                        | centrosomal protein of 152 kDa-like                                |
| evm.model.Herato0901.26                         | uncharacterized protein LOC113391445 isoform X2                    |
| evm.model.Herato0901.292                        | DNA replication licensing factor Mcm2                              |
| evm.model.Herato0901.53                         | E3 ubiquitin-protein ligase TRAIP-like                             |
| evm.model.Herato0901.85                         | general transcription factor IIH subunit 1                         |
| evm.model.Herato0901.88                         | putative glutathione-specific gamma-glutamylcyclotransferase 2     |
| evm.model.Herato1001.178                        | checkpoint protein HUS1                                            |
| evm.model.Herato1001.196                        | UDP-glucuronosyltransferase 2B19-like                              |
| evm.model.Herato1001.80                         | alpha-N-acetylgalactosaminidase-like                               |
| evm.model.Herato1003.157                        | low molecular weight phosphotyrosine protein phosphatase-like      |
| evm.model.Herato1003.160                        | chromatin accessibility complex protein 1                          |
| evm.model.Herato1003.166                        | histone deacetylase 6 isoform X1                                   |
| evm.model.Herato1003.177                        | trypsin CFT-1-like                                                 |
| evm.model.Herato1003.253                        | sorting nexin-17                                                   |
| evm.model.Herato1005.11                         | probable RNA helicase armi                                         |
| evm.model.Herato1005.139                        | glycylpeptide N-tetradecanoyltransferase 2                         |
| evm.model.Herato1005.167                        | regulator of microtubule dynamics protein 1-like                   |
| evm.model.Herato1005.207                        | phosphatidate phosphatase LPIN3 isoform X1                         |
| evm.model.Herato1005.217                        | sodium/hydrogen exchanger 7 isoform X1                             |
| evm.model.Herato1005.57                         | nicotinate phosphoribosyltransferase isoform X1                    |
| evm.model.Herato1005.63                         | glutaminase liver isoform, mitochondrial isoform X2                |
| evm.model.Herato1007.125                        | deoxynucleoside kinase                                             |
| evm.model.Herato1007.149                        | protein RTF2 homolog                                               |
| evm.model.Herato1007.166                        | 2-(3-amino-3-carboxypropyl)histidine synthase subunit 2            |
| evm.model.Herato1007.208                        | RNA-directed DNA polymerase from mobile element jockey             |
| evm.model.Herato1007.99                         | conserved oligomeric Golgi complex subunit 3                       |
| evm.model.Herato1108.31                         | kelch domain-containing protein 4                                  |
| evm.model.Herato1108.346                        | b(0,+)-type amino acid transporter 1-like isoform X2               |
| evm.model.Herato1108.536                        | probable uridine-cytidine kinase isoform X2                        |
| evm.model.Herato1108.600                        | PAXIP1-associated glutamate-rich protein 1                         |
| evm.model.Herato1202.158                        | dysbindin protein homolog                                          |
| evm.model.Herato1202.240                        | exportin-2                                                         |
| evm.model.Herato1202.498                        | putative inorganic phosphate cotransporter                         |
| evm.model.Herato1202.516                        | mediator of RNA polymerase II transcription subunit 1              |
| evm.model.Herato1202.565                        | WD repeat-containing protein 75                                    |
| evm.model.Herato1202.641                        | Phosphonates-binding periplasmic protein                           |
| evm.model.Herato1202.718                        | pre-mRNA-processing-splicing factor 8                              |
| evm.model.Herato1202.754                        | 60S ribosomal protein L27                                          |
| evm.model.Herato1202.757                        | CBP80/20-dependent translation initiation factor isoform X1        |
| evm.model.Herato1301.470                        | probable rRNA-processing protein EBP2 homolog                      |
| evm.model.Herato1301.514                        | E3 SUMO-protein ligase NSE2-like                                   |
| evm.model.Herato1301.541                        | putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX16 |
| evm.model.Herato1301.634                        | protein tumorous imaginal discs, mitochondrial-like isoform X2     |
| evm.model.Herato1301.694                        | histone acetyltransferase KAT2A                                    |
| evm.model.Herato1301.76                         | DNA polymerase epsilon subunit 4                                   |
| evm.model.Herato1408.233                        | putative glycerol kinase 5                                         |
| evm.model.Herato1408.47                         | translation elongation factor 2                                    |
| evm.model.Herato1411.158                        | putative deoxyribonuclease TATDN1                                  |
| evm.model.Herato1411.33                         | centromere protein L                                               |
| evm.model.Herato1411.67                         | hepatocyte nuclear factor 4-gamma isoform X3                       |
| evm.model.Herato1505.10                         | probable phosphomannomutase                                        |
| evm.model.Herato1505.42                         | inositol monophosphatase 1                                         |
| evm.model.Herato1507.114                        | lysophospholipid acyltransferase 2                                 |
| evm.model.Herato1507.325                        | transcription factor Adf-1-like                                    |
| evm.model.Herato1601.102                        | V-type proton ATPase subunit D                                     |
| evm.model.Herato1601.20                         | terminal uridylyltransferase Tailor-like                           |
| evm.model.Herato1603.14                         | ATP-uridyl cassette sub-family B member 10, mitochondrial-like     |
| evm.model.Herato1603.25 evm.model.Herato1603.27 | histone-lysine N-methyltransferase setd3                           |
| evm.model.Herato1605.122                        | breast cancer metastasis-suppressor 1-like protein                 |

|                          |                                                                           |
|--------------------------|---------------------------------------------------------------------------|
| evm.model.Herato1701.156 | uncharacterized protein LOC105841373                                      |
| evm.model.Herato1701.216 | alpha-methylacyl-CoA racemase                                             |
| evm.model.Herato1701.307 | cytoplasmic dynein 2 heavy chain 1                                        |
| evm.model.Herato1701.326 | synembryn-A                                                               |
| evm.model.Herato1701.49  | nitric oxide synthase-interacting protein homolog                         |
| evm.model.Herato1701.67  | programmed cell death protein 6 isoform X1                                |
| evm.model.Herato1701.71  | hypothetical protein X975_16339, partial                                  |
| evm.model.Herato1704.19  | uncharacterized protein LOC117983159                                      |
| evm.model.Herato1705.24  | suppressor of fused homolog                                               |
| evm.model.Herato1705.29  | galectin-4-like                                                           |
| evm.model.Herato1708.118 | uncharacterized protein LOC113395896, partial                             |
| evm.model.Herato1708.129 | phenoloxidase-activating factor 2-like                                    |
| evm.model.Herato1708.59  | pyrimidodiazepine synthase-like                                           |
| evm.model.Herato1801.125 | iron-sulfur cluster assembly enzyme ISCU, mitochondrial                   |
| evm.model.Herato1801.52  | THAP domain-containing protein 1-like                                     |
| evm.model.Herato1805.37  | Hermansky-Pudlak syndrome 1 protein homolog                               |
| evm.model.Herato1807.117 | uncharacterized protein LOC114247813                                      |
| evm.model.Herato1807.58  | protein THEM6-like                                                        |
| evm.model.Herato1904.10  | snRNA-activating protein complex subunit 4                                |
| evm.model.Herato1904.109 | uncharacterized protein LOC113396016                                      |
| evm.model.Herato1904.196 | N-acetylgalactosaminyltransferase 7                                       |
| evm.model.Herato1904.245 | protein croquemort-like                                                   |
| evm.model.Herato1904.9   | DNA mismatch repair protein Mlh1                                          |
| evm.model.Herato1908.40  | L-2-hydroxyglutarate dehydrogenase, mitochondrial                         |
| evm.model.Herato1908.97  | transforming growth factor beta regulator 1                               |
| evm.model.Herato1910.10  | transcription initiation factor TFIID subunit 7-like                      |
| evm.model.Herato1910.12  | UDP-N-acetylglucosamine--dolichyl-phosphate N-                            |
| evm.model.Herato1910.184 | GRIP and coiled-coil domain-containing protein 1 isoform X1               |
| evm.model.Herato1910.30  | magnesium transporter NIPA2                                               |
| evm.model.Herato2001.162 | mitochondrial inner membrane protease ATP23 homolog                       |
| evm.model.Herato2001.165 | multiple inositol polyphosphate phosphatase 1-like                        |
| evm.model.Herato2001.181 | peptidyl-prolyl cis-trans isomerase H                                     |
| evm.model.Herato2001.408 | WD repeat-containing protein 46                                           |
| evm.model.Herato2001.477 | 23 kDa integral membrane protein-like                                     |
| evm.model.Herato2001.526 | DNA oxidative demethylase ALKBH2-like                                     |
| evm.model.Herato2001.550 | U11/U12 small nuclear ribonucleoprotein 48 kDa protein-like               |
| evm.model.Herato2001.555 | uncharacterized protein LOC113404712 isoform X2                           |
| evm.model.Herato2001.618 | myb/SANT-like DNA-binding domain-containing protein 3                     |
| evm.model.Herato2001.686 | probable multidrug resistance-associated protein lethal(2)03659           |
| evm.model.Herato2101.136 | pyruvate dehydrogenase [acetyl-transferring]-phosphatase 1, mitochondrial |
| evm.model.Herato2101.34  | hypothetical protein KGM_202987                                           |
| evm.model.Herato2101.370 | neuralized-like protein 4                                                 |
| evm.model.Herato2101.375 | dynein heavy chain 6, axonemal                                            |
| evm.model.Herato2101.379 | prefoldin subunit 1                                                       |
| evm.model.Herato2101.420 | transcription initiation factor TFIID subunit 7-like                      |
| evm.model.Herato2101.97  | protein Jumonji                                                           |

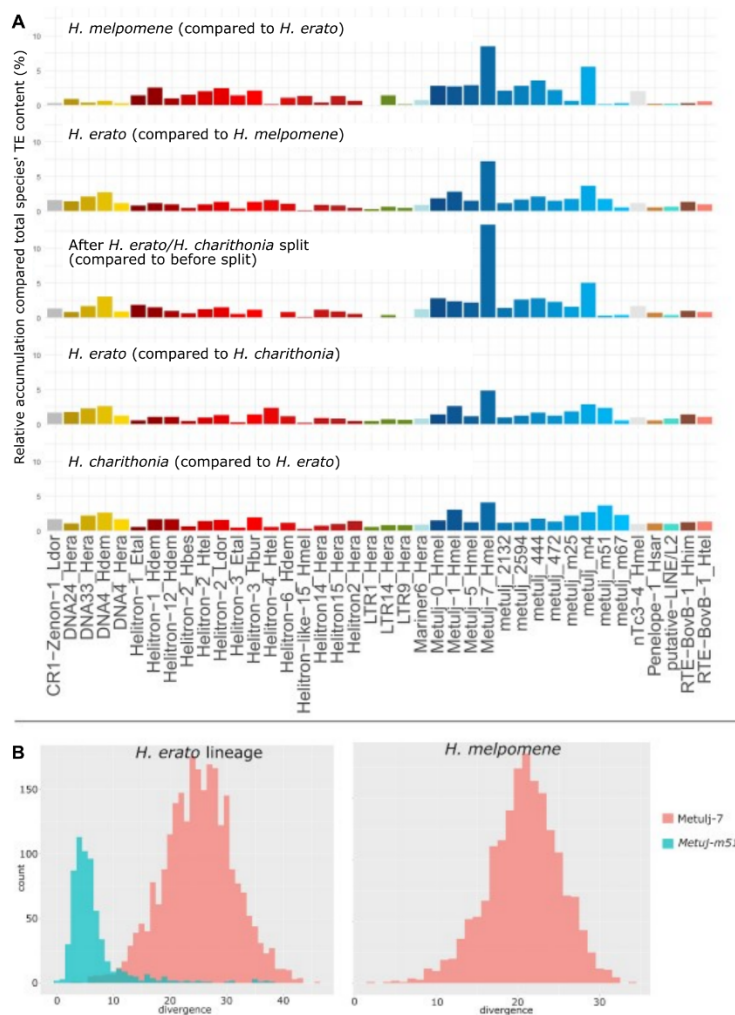
**Supplemental Table S5. Genes near lineage-specific ATAC-seq peaks with head-specific accessibility and that associate with lineage-specific TE insertions.** Rows indicated with an asterisk (\*) are peaks near genes with a known function related to eye or neuron development. Gene IDs are relative to the *H. erato* genome assembly. Note that origin relative to their split can only be presented for *H. erato* and *H. charithonia*.

| Species               | Origin relative to split               | PAN start | PAN end   | Distance to TSS | Nearest Gene             | Blast description                                                            |   |
|-----------------------|----------------------------------------|-----------|-----------|-----------------|--------------------------|------------------------------------------------------------------------------|---|
| <i>H. melpomene</i>   |                                        | 43439207  | 43439817  | -53871          | evm.model.Herato0211.54  | nardilysin-like isoform X2                                                   |   |
|                       |                                        | 53556044  | 53557026  | -52466          | evm.model.Herato0215.115 | ATP-dependent RNA helicase WM6                                               |   |
|                       |                                        | 81976962  | 81977176  | 70377           | evm.model.Herato0401.20  | unnamed protein product                                                      |   |
|                       |                                        | 113559102 | 113559581 | 13883           | evm.model.Herato0503.140 | arylphorin subunit alpha-like                                                |   |
|                       |                                        | 141307229 | 141307607 | 4610            | evm.model.Herato0606.195 | odorant receptor 13a-like                                                    |   |
|                       |                                        | 244127509 | 244127706 | 5262            | evm.model.Herato1003.231 | F-box/WD repeat-containing protein 9-like                                    |   |
|                       |                                        | 249090365 | 249090829 | 6302            | evm.model.Herato1005.147 | nuclear speckle splicing regulatory protein 1                                |   |
|                       |                                        | 316442193 | 316442554 | -12875          | evm.model.Herato1202.784 | protein chiffon                                                              | * |
|                       |                                        | 337485755 | 337486339 | -18528          | evm.model.Herato1301.495 | elongation of very long chain fatty acids protein AAEL008004-like isoform X1 |   |
|                       |                                        | 422443342 | 422443560 | -70418          | evm.model.Herato1605.106 | HMG box transcription factor BBX isoform X1                                  |   |
|                       |                                        | 441652772 | 441653477 | 3975            | evm.model.Herato1701.397 | RNA transcription, translation and transport factor protein                  |   |
|                       |                                        | 442836249 | 442836683 | 225826          | evm.model.Herato1703.4   | nose resistant to fluoxetine protein 6-like                                  |   |
|                       |                                        | 472668250 | 472669760 | -63199          | evm.model.Herato1805.136 | serine/threonine-protein kinase PAK 3                                        | * |
|                       |                                        | 501603158 | 501603392 | -41438          | evm.model.Herato1904.173 | flybase fork head domain transcription factor slp2-like                      | * |
|                       |                                        | 506155200 | 506155751 | 62769           | evm.model.Herato1904.273 | zwei Ig domain protein zig-8-like                                            | * |
|                       |                                        | 521680935 | 521681435 | 14258           | evm.model.Herato1910.196 | DNA replication complex GINS protein PSF1-like                               |   |
| <i>H. erato</i>       |                                        | 550702397 | 550702548 | -126489         | evm.model.Herato2001.741 | uncharacterized protein LOC117985494                                         |   |
|                       |                                        | 559001222 | 559002513 | 7196            | novel_gene_269           | unnamed protein product                                                      |   |
|                       | After split from <i>H. charithonia</i> | 492945962 | 492946224 | 30444           | evm.model.Herato1901.113 | MICOS complex subunit Mic60-like                                             |   |
|                       | After split from <i>H. charithonia</i> | 492946674 | 492947610 | 31493           | evm.model.Herato1901.113 | MICOS complex subunit Mic60-like                                             |   |
| <i>H. charithonia</i> | After split from <i>H. erato</i>       | 526831481 | 526832240 | 1266            | novel_gene_251           | unnamed protein product                                                      |   |
|                       | After split from <i>H. erato</i>       | 526832505 | 526833450 | 2383            | novel_gene_251           | unnamed protein product                                                      |   |
| <i>H. charithonia</i> | Before split from <i>H. erato</i>      | 549632384 | 549632712 | 18655           | evm.model.Herato2001.719 | importin subunit alpha-1-like                                                |   |
|                       | After split from <i>H. erato</i>       | 143139789 | 143140178 | 19714           | evm.model.Herato0606.238 | tropomodulin-1 isoform X5                                                    | * |

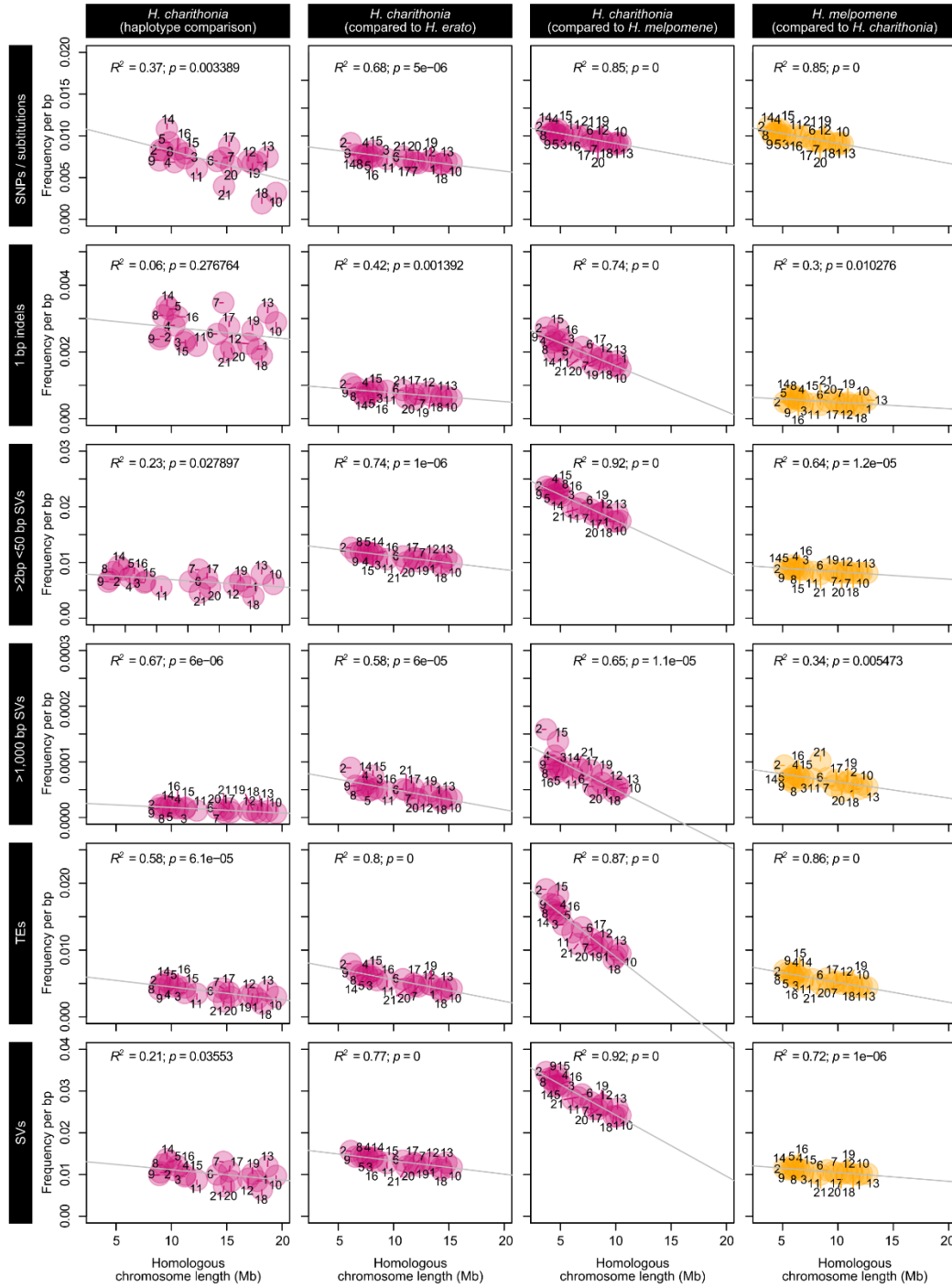
**Supplemental Table S6. ATAC-seq samples, tissue, origin, and GenBank accession numbers.** Number of reads are quality filtered and properly paired.

| <b>Sample ID</b> | <b>Tissue</b> | <b>Species</b>             | <b>Locality</b> | <b>Accession #</b> | <b># Reads</b> |
|------------------|---------------|----------------------------|-----------------|--------------------|----------------|
| CH_head2         | Head          | <i>H. charithonia</i>      | Puerto Rico     | SAMN24689923       | 14,045,582     |
| CH_head6         | Head          | <i>H. charithonia</i>      | Puerto Rico     | SAMN24689924       | 36,694,766     |
| BR12             | Head          | <i>H. erato demophoon</i>  | Panama          | SAMN24689925       | 18,019,744     |
| BR14             | Head          | <i>H. erato demophoon</i>  | Panama          | SAMN24689926       | 57,988,466     |
| BR11             | Head          | <i>H. melpomene rosina</i> | Panama          | SAMN24689927       | 10,214,746     |
| M4_head          | Head          | <i>H. melpomene rosina</i> | Panama          | SAMN24689928       | 45,921,776     |
| CH_FW2           | Forewing      | <i>H. charithonia</i>      | Puerto Rico     | SAMN24689929       | 8,605,992      |
| CH_FW7           | Forewing      | <i>H. charithonia</i>      | Puerto Rico     | SAMN24689930       | 23,397,866     |
| LI7_demophoon_FW | Forewing      | <i>H. erato demophoon</i>  | Panama          | SAMN24689931       | 17,294,536     |
| FW-pboy          | Forewing      | <i>H. erato demophoon</i>  | Panama          | SAMN24689932       | 49,257,240     |
| LI14_rosina_FW   | Forewing      | <i>H. melpomene rosina</i> | Panama          | SAMN24689933       | 23,449,240     |
| M4_FW            | Forewing      | <i>H. melpomene rosina</i> | Panama          | SAMN24689934       | 24,353,808     |
| CH_HW2           | Hindwing      | <i>H. charithonia</i>      | Puerto Rico     | SAMN24689935       | 26,910,064     |
| CH_HW7           | Hindwing      | <i>H. charithonia</i>      | Puerto Rico     | SAMN24689936       | 21,029,072     |
| LI7_demophoon_HW | Hindwing      | <i>H. erato demophoon</i>  | Panama          | SAMN24689937       | 14,631,856     |
| LB_42            | Hindwing      | <i>H. erato demophoon</i>  | Panama          | SAMN24689938       | 26,635,566     |
| LI14_rosina_HW   | Hindwing      | <i>H. melpomene rosina</i> | Panama          | SAMN24689939       | 23,176,622     |
| M4_HW            | Hindwing      | <i>H. melpomene rosina</i> | Panama          | SAMN24689940       | 36,959,108     |

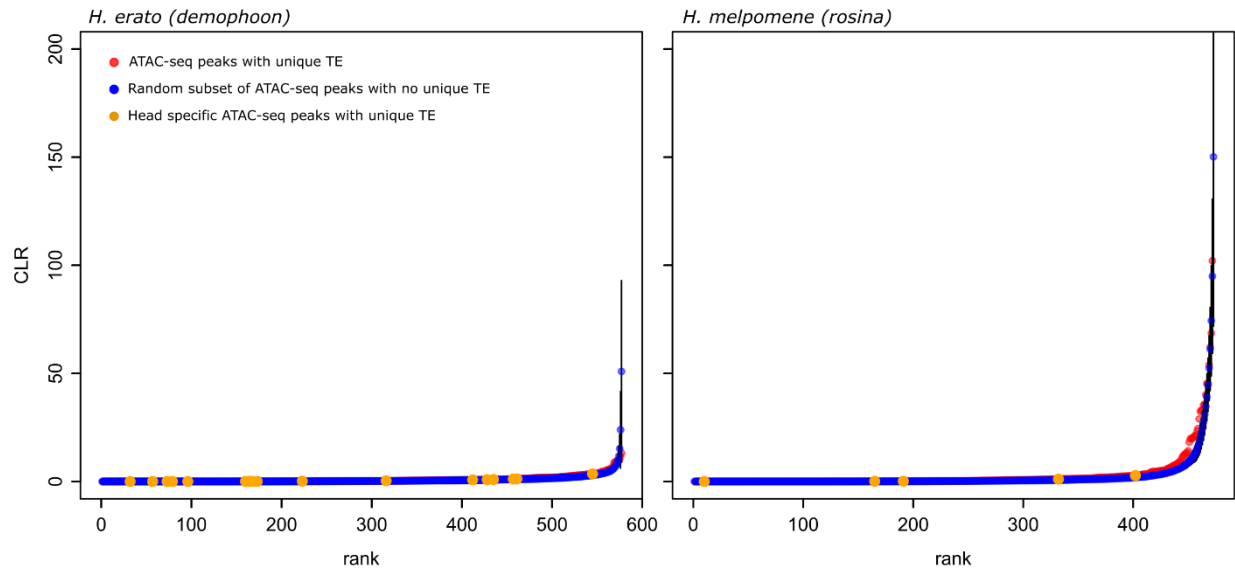
## SUPPLEMENTAL FIGURES



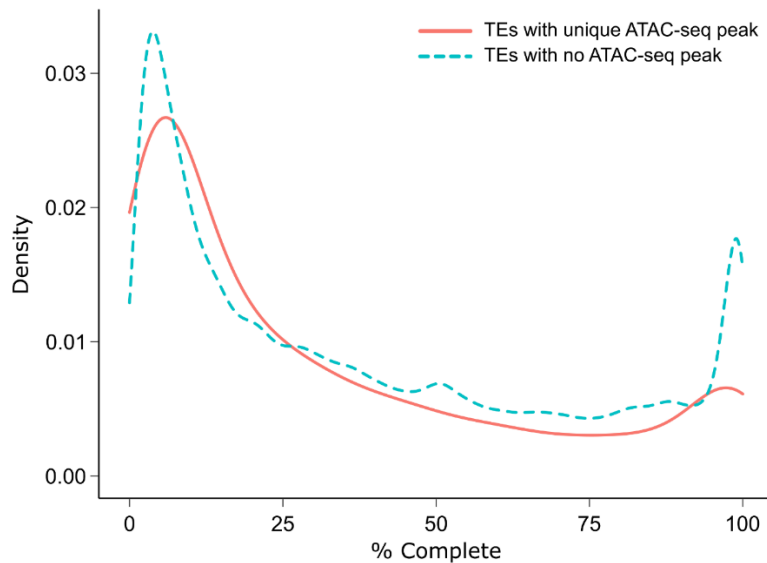
**Supplemental Fig. S1. Lineage specific transposable element activity. (A.)** Accumulation of the 40 most active TE families in each lineage (values are percentages among the total lineage-specific TEs in the species). These families were obtained by grouping the 20 most active families for each lineage comparison (see Fig. 3). **(B.)** Divergence distribution plots of Metulj-7 and Metulj-m51 represent times of activity in *H. erato* and *H. melpomene*. The x-axis represents the divergence (percentage) from the reference sequence corrected with the Jukes-Cantor model, the y-axis represents the sequence count.



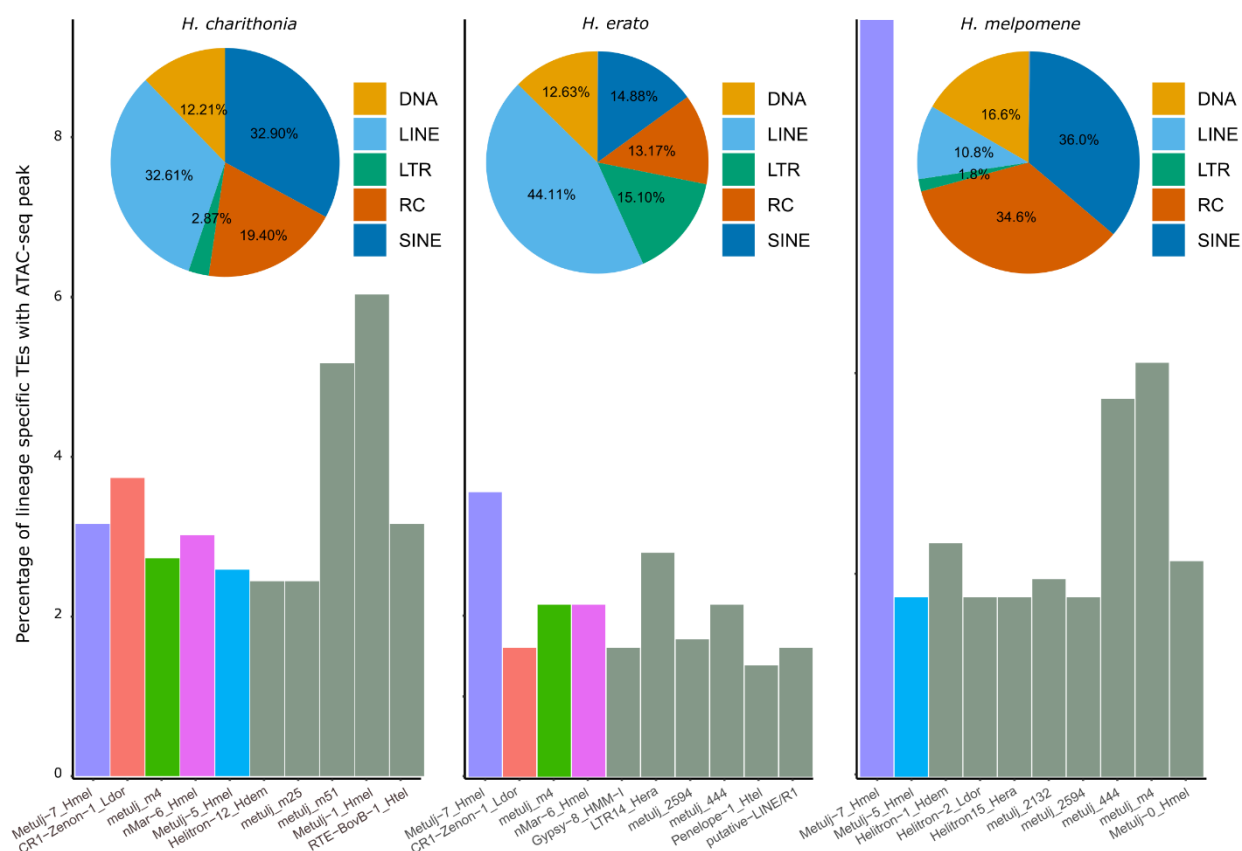
**Supplemental Fig. S2. Patterns of lineage-specific sequence distribution and chromosome lengths for SNPs, 1 bp indels, SVs between 2 and 50 bp, SVs larger than 1,000 bp, SVs characterized as TEs and all SVs. First column:** Correlation between homologous chromosome lengths and frequency of indels in the chromosomes of *H. charithonia*. **Second column:** Correlation between homologous chromosome lengths and frequency of lineage-specific sequences in the chromosomes of *H. charithonia* compared to *H. erato*. **Third column:** Correlation between homologous chromosome lengths and frequency of lineage-specific sequences in the chromosomes of *H. charithonia* compared to *H. melpomene*. **Fourth column:** Correlation between homologous chromosome lengths and frequency of lineage-specific sequences in the chromosomes of *H. melpomene* compared to *H. charithonia*. Lines indicate regression fit. Numbers indicate chromosome numbers. Colors refer to sequences specific to *H. charithonia* (pink), and *H. melpomene* (orange).



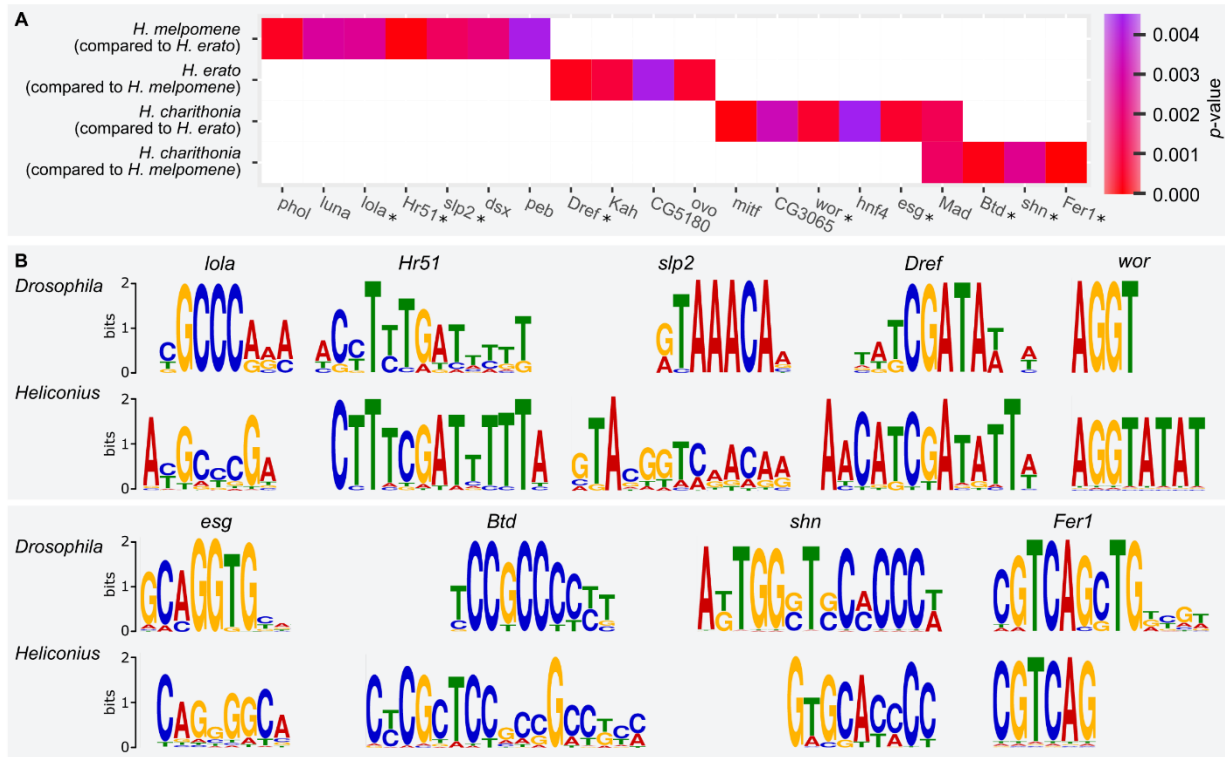
**Supplemental Fig. S3. Ranked SweepFinder2 Composite Likelihood Ratio (CLR) values for ATAC-seq peaks with and without lineage-specific Transposable Element (TE).** ATAC-seq peaks with lineage-specific TE (red) do not show greater signatures of selective sweeps than expected from random subsets of ATAC-seq peaks with no lineage-specific TE (blue) in both populations of *H. erato demophoon* (left) and *H. melpomene rosina* (right). Head-specific ATAC-seq peaks with a lineage-specific TE are indicated in orange.



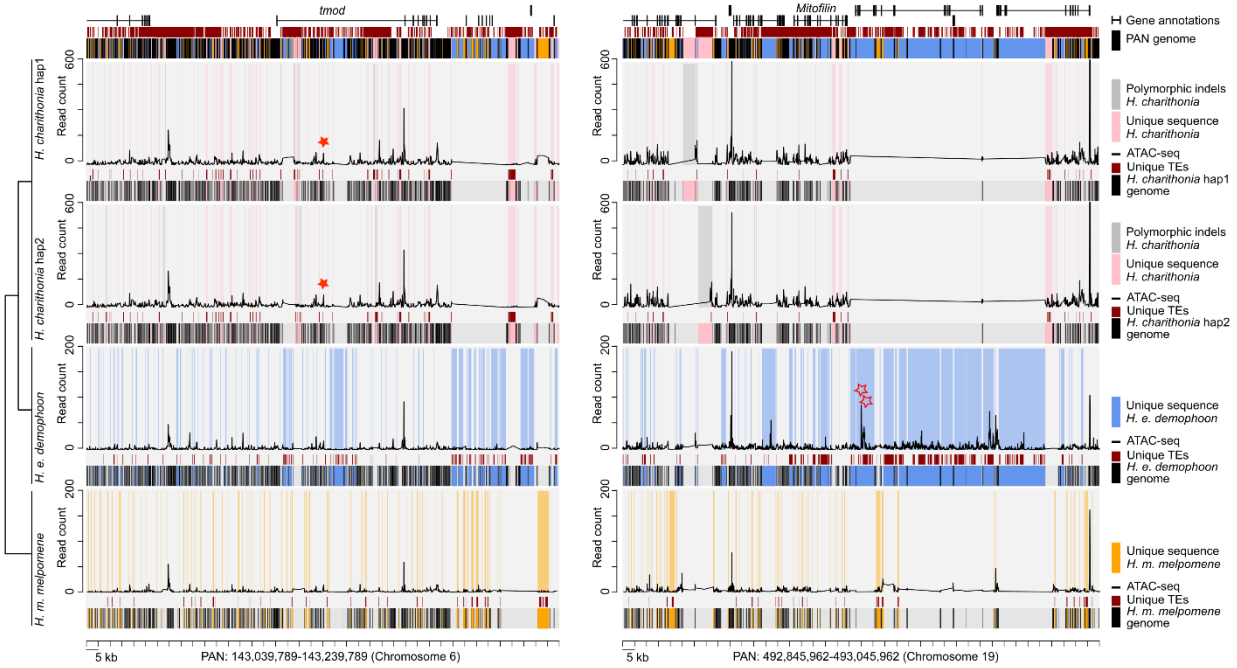
**Supplemental Fig. S4. Completeness of Transposable Elements (TEs) for TEs associated with lineage-specific ATAC-peaks and TEs with no ATAC-seq peak association.** TE completeness was calculated based on the total length of each element recovered from the reference library.



**Supplemental Fig. S5. Transposable Element (TE) composition of lineage-specific (i.e., compared to all other species) ATAC-seq peaks within TEs.** Pie charts show the TE family distribution. Bar plots show the percentage of the ten most represented TEs in each lineage relative to the total number lineage-specific ATAC-seq peaks within TEs.



**Supplemental Fig. S6. Transcription factor (TF) binding motif enrichment among lineage-specific TE insertions with ATAC-seq peaks. (A.)** Lineage-specific ATAC-seq peaks within TEs in *Heliconius melpomene* (compared to *H. erato*) were significantly enriched for binding sites with significant homology to flybase *D. melanogaster* TF binding of *pleiohomeotic like* (*phol*; FBgn0035997), *luna* (*luna*; FBgn0040765), *longitudinals lacking* (*lola*; FBgn0283521), *Hormone receptor 51* (*Hr51*; FBgn0034012), *sloppy paired 2* (*slp2*; FBgn0004567), *doublesex* (*dsx*; FBgn0000504), and *pebbled* (*peb*; FBgn0003053). *Heliconius erato* (compared to *H. melpomene*) was significantly enriched for binding sites for DNA replication-related element factor (*Dref*; FBgn0015664), *Kahuli* (*Kah*; FBgn0035144), *CG5180* (*CG5180*; FBgn0043457), and *ovo* (*ovo*; FBgn0003028). *H. charithonia* (compared to *H. erato*) was enriched in binding sites for *Mitf* (*Mitf*; FBgn0263112), *CG3065* (*CG3065*; FBgn0034946), *worniu* (*wor*; FBgn0001983), *hnf4* (*hnf4*; FBgn0004914), *escargot* (*esg*; FBgn0287768), and *Mothers against dpp* (*Mad*; FBgn0011648). *H. charithonia* (compared to *H. erato*) was enriched in binding sites for *Mothers against dpp* (*Mad*; FBgn0011648), *buttonhead* (*btd*; FBgn0000233), *schnurri* (*shn*; FBgn0003396), and *48 related 1* (*Fer1*; FBgn0037475). TFs with known functions in neuron development are indicated with an asterisk. P-value is the significance of enrichment as obtained by the MEME analysis. (B.) Homology of identified enriched motifs to *D. melanogaster* motifs of TFs with known functions in neuron development.



**Supplemental Fig. S7. Example intervals of the pan-genome assembly of *H. charithonia* (pink), *H. erato* (blue) and *H. melpomene* (orange) with alignment of lineage-specific genome sequences, transposable element (TE) annotations and ATAC-seq profiles in the pan-genome coordinate space for *tropomodulin-1* (*tmod*) and *MICOS complex subunit Mic60* (*Mic60*).** The plots show an illustrative interval of the pan-genome assembly near the gene *tmod* and *Mic60* that highlights sequences present in each of the genomes relative to the pan-genome (black shading underneath each of the graphs), lineage-specific sequences in each of the genomes (pink, blue and orange shading in graphs), TEs that overlap with lineage-specific sequences (dark red), and ATAC-seq profiles for head tissue (average of two biological replicates). Gray shading in the *H. charithonia* haplotype 2 (hap2) graph indicates an indel in the genome of a single *H. charithonia* individual. Red filled star indicates an ATAC-seq peak with head-specific accessibility (compared to wing tissue) that intersect with a lineage-specific TE insertion. Red open stars indicate ATAC-seq peaks that intersect with a lineage-specific TE insertion but do not have head-specific accessibility (compared to wing tissue).

## SUPPLEMENTAL CODE

Resource files have been deposited at <https://github.com/StevenVB12/Genomics>.

### Supplemental Code S1. Build pan genome using seq-seq-pan (<https://gitlab.com/chrjan/seq-seq-pan>).

```
seq-seq-pan-wga --config genomefile=genome_list.txt outfilename=SeqSeqPan_erato_cha1_cha2_melp

## genome_list.txt includes the names of each genome fasta file
# Herato_final.fasta
# Hcha.assembly.v1.1.fasta
# H_charithonia_10X.2.fasta
# Hmel2.fa

## Remove newline from xmfa file for downstream processing

perl -pe 'chomp if /^[ATCGNSBDHVMRWYKtcgnsbdhvmrwyk-]/' SeqSeqPan_erato_cha1_cha2_melp.xmfa | sed
's/\=/\n\=/g' | sed 's/>\n>/g' | sed '/^$/d' > SeqSeqPan_erato_cha1_cha2_melp_noNewline.xmfa
```

**Supplemental Code S2. Python script to create .bed files with sequence blocks present in pan genome for each genome.**

```
#!/usr/bin/env python

# Use as: python seq-seq-pan_blocks_intervals.py -I SeqSeqPan_erato_cha1_cha2_melp_noNewline.xmfa -g
1,2,3,4

from __future__ import division
from Bio.Seq import Seq
import argparse, re

parser = argparse.ArgumentParser()
parser.add_argument("-I", "--infile", help="xmfa file with no newlines from seq-seq-pan", required = True)
parser.add_argument("-g", "--genomeList", help="comma separated list of genomes in order of seq-seq-pan
genome", required = True)
args = parser.parse_args()

infile = args.infile
genomes = args.genomeList.split(",")

count = 0

for genome in genomes:

    identified = 0
    sequence = str()
    sequenceOr = str()

    f= open(genome + '_blocks_intervals.txt','a+')
    f.write("\t".join(("start","end","\n")))
    count +=1
    fileIn = open(infile, "r")
    fileLine = fileIn.readline()
    fileLine = fileLine.strip()

    print('searching genome ' + str(count) + ' ' + str(genome))

    while fileLine:

        if '>' + str(genome) + ':' in fileLine:

            identified += 1

            fileLine = fileIn.readline()
            fileLine = fileLine.strip()
            seq = Seq(fileLine)

            sequence = sequence + seq

        if '=' not in fileLine and '>' not in fileLine:
            lenSeq = len(fileLine)

        if '=' in fileLine and identified == 0:
            sequence = sequence + '-' * lenSeq

        if '=' in fileLine and identified == 1:
            identified = 0

        fileLine = fileIn.readline()
        fileLine = fileLine.strip()

    print('total length: ' + str(len(sequence)))
    start = 0
    for i in range(len(sequence)):

        if i > 0 and i < len(sequence)-1:
            if sequence[i-1] == '-':
```

```
        if sequence[i] != '-':
            start = i
        if sequence[i+1] == '-':
            if sequence[i] != '-':
                end = i
                f.write("\t".join((str(start),str(end),"\n")))

sequence = str()
sequenceOr = str()
```

### Supplemental Code S3. Bash commands for merging and intersecting block .bed with bedtools.

```
# Remove first line
sed -i -n -e '2,$p' 1_blocks_intervals.txt
sed -i -n -e '2,$p' 2_blocks_intervals.txt
sed -i -n -e '2,$p' 3_blocks_intervals.txt
sed -i -n -e '2,$p' 4_blocks_intervals.txt

# Remove space from end of line
sed -i 's/[[:space:]]*$//' 1_blocks_intervals.txt
sed -i 's/[[:space:]]*$//' 2_blocks_intervals.txt
sed -i 's/[[:space:]]*$//' 3_blocks_intervals.txt
sed -i 's/[[:space:]]*$//' 4_blocks_intervals.txt

# Add 'pan' as scaffold in the .bed file
lineN=$(< 1_blocks_intervals.txt wc -l)
pr -mts <(yes "pan" | head -n $lineN) <(cut -d$'\t' -f 1,2 1_blocks_intervals.txt) >
1_blocks_intervals.corr.bed
lineN=$(< 2_blocks_intervals.txt wc -l)
pr -mts <(yes "pan" | head -n $lineN) <(cut -d$'\t' -f 1,2 2_blocks_intervals.txt) >
2_blocks_intervals.corr.bed
lineN=$(< 3_blocks_intervals.txt wc -l)
pr -mts <(yes "pan" | head -n $lineN) <(cut -d$'\t' -f 1,2 3_blocks_intervals.txt) >
3_blocks_intervals.corr.bed
lineN=$(< 4_blocks_intervals.txt wc -l)
pr -mts <(yes "pan" | head -n $lineN) <(cut -d$'\t' -f 1,2 4_blocks_intervals.txt) >
4_blocks_intervals.corr.bed

# Merge block files for lineage comparisons
cat 1_blocks_intervals.corr.bed 2_blocks_intervals.corr.bed | sort -g -k2,2 | bedtools merge >
blocks_merged_1_2.txt
cat 1_blocks_intervals.corr.bed 3_blocks_intervals.corr.bed | sort -g -k2,2 | bedtools merge >
blocks_merged_1_3.txt
cat 1_blocks_intervals.corr.bed 4_blocks_intervals.corr.bed | sort -g -k2,2 | bedtools merge >
blocks_merged_1_4.txt
cat 2_blocks_intervals.corr.bed 3_blocks_intervals.corr.bed | sort -g -k2,2 | bedtools merge >
blocks_merged_2_3.txt
cat 2_blocks_intervals.corr.bed 4_blocks_intervals.corr.bed | sort -g -k2,2 | bedtools merge >
blocks_merged_2_4.txt
cat 3_blocks_intervals.corr.bed 4_blocks_intervals.corr.bed | sort -g -k2,2 | bedtools merge >
blocks_merged_3_4.txt

cat 1_blocks_intervals.corr.bed 2_blocks_intervals.corr.bed 3_blocks_intervals.corr.bed | sort -g -k2,2 |
bedtools merge > blocks_merged_1_2_3.txt
cat 1_blocks_intervals.corr.bed 2_blocks_intervals.corr.bed 4_blocks_intervals.corr.bed | sort -g -k2,2 |
bedtools merge > blocks_merged_1_2_4.txt
cat 1_blocks_intervals.corr.bed 3_blocks_intervals.corr.bed 4_blocks_intervals.corr.bed | sort -g -k2,2 |
bedtools merge > blocks_merged_1_3_4.txt
cat 2_blocks_intervals.corr.bed 3_blocks_intervals.corr.bed 4_blocks_intervals.corr.bed | sort -g -k2,2 |
bedtools merge > blocks_merged_2_3_4.txt

# Subtract blocks to get unique sequences
bedtools subtract -sorted -a 1_blocks_intervals.corr.bed -b blocks_merged_2_3_4.txt > blocks_unique_1.txt
bedtools subtract -sorted -a 2_blocks_intervals.corr.bed -b blocks_merged_1_3_4.txt > blocks_unique_2.txt
bedtools subtract -sorted -a 3_blocks_intervals.corr.bed -b blocks_merged_1_2_4.txt > blocks_unique_3.txt
bedtools subtract -sorted -a 4_blocks_intervals.corr.bed -b blocks_merged_1_2_3.txt > blocks_unique_4.txt
```

```

bedtools subtract -sorted -a 1_blocks_intervals.corr.bed -b 4_blocks_intervals.corr.bed >
blocks_unique_1_vs_4.txt
bedtools subtract -sorted -a 4_blocks_intervals.corr.bed -b 1_blocks_intervals.corr.bed >
blocks_unique_4_vs_1.txt

bedtools subtract -sorted -a 2_blocks_intervals.corr.bed -b 1_blocks_intervals.corr.bed >
blocks_unique_2_vs_1.txt
bedtools subtract -sorted -a 3_blocks_intervals.corr.bed -b 1_blocks_intervals.corr.bed >
blocks_unique_3_vs_1.txt

bedtools subtract -sorted -a 1_blocks_intervals.corr.bed -b blocks_merged_2_3.txt >
blocks_unique_1_vs_2_3.txt
bedtools subtract -sorted -a 4_blocks_intervals.corr.bed -b blocks_merged_2_3.txt >
blocks_unique_4_vs_2_3.txt

bedtools subtract -sorted -a 4_blocks_intervals.corr.bed -b 2_blocks_intervals.corr.bed >
blocks_unique_4_vs_2.txt

bedtools subtract -sorted -a 2_blocks_intervals.corr.bed -b 3_blocks_intervals.corr.bed >
blocks_unique_2_vs_3.txt
bedtools subtract -sorted -a 3_blocks_intervals.corr.bed -b 2_blocks_intervals.corr.bed >
blocks_unique_3_vs_2.txt

bedtools intersect -sorted -a 2_blocks_intervals.corr.bed -b 3_blocks_intervals.corr.bed >
blocks_shared_2_3.txt

bedtools subtract -sorted -a 4_blocks_intervals.corr.bed -b blocks_merged_1_2_3.txt >
blocks_unique_4_vs_1_2_3.txt

```

**Supplemental Code S4. Calculate pairwise nucleotide differences in homologous sequence between pairs of genomes in the pan genome alignment.**

```
#!/usr/bin/env python

# Use as: python SeqSeqPan_dxy.py -I SeqSeqPan_erato_cha1_cha2_melp_noNewline.xmfa -w 100000 -g 1,2 -o
dxy_1_2

from __future__ import division
import argparse, re

parser = argparse.ArgumentParser()
parser.add_argument("-I", "--infile", help=" xmfa file with no newlines from seq-seq-pan", required =
True)
parser.add_argument("-g", "--genomeList", help="comma separated list of genomes in order of seq-seq-pan",
required = True)
parser.add_argument("-w", "--windowSize", help="number", required = True)
parser.add_argument("-o", "--outfile", help="txt", required = True)
args = parser.parse_args()

infile = args.infile
genomes = args.genomeList.split(",")
window = args.windowSize
outfile = args.outfile

count = 0

genome1 = genomes[0]
genome2 = genomes[1]

identified = 0
sequence1 = str()

count +=1
fileIn = open(infile, "r")
fileLine = fileIn.readline()
fileLine = fileLine.strip()

print('searching genome ' + str(count) + ' ' + str(genome1))

while fileLine:

    if '>' + str(genome1) + ':' in fileLine:

        identified += 1

        fileLine = fileIn.readline()
        fileLine = fileLine.strip()
        seq = fileLine

        sequence1 = sequence1 + seq

    if '=' not in fileLine and '>' not in fileLine:
        lenSeq = len(fileLine)

    if '=' in fileLine and identified == 0:
        sequence1 = sequence1 + '-' * lenSeq

    if '=' in fileLine and identified == 1:
        identified = 0

    fileLine = fileIn.readline()
    fileLine = fileLine.strip()

print('total length: ' + str(len(sequence1)))

identified = 0
```

```

sequence2 = str()

count +=1
fileIn = open(infile, "r")
fileLine = fileIn.readline()
fileLine = fileLine.strip()

print('searching genome ' + str(count) + ' ' + str(genome2))

###
while fileLine:

    if '>' + str(genome2) + ':' in fileLine:

        identified += 1

        fileLine = fileIn.readline()
        fileLine = fileLine.strip()
        seq = fileLine

        sequence2 = sequence2 + seq

    if '=' not in fileLine and '>' not in fileLine:
        lenSeq = len(fileLine)

    if '=' in fileLine and identified == 0:
        sequence2 = sequence2 + '-' * lenSeq

    if '=' in fileLine and identified == 1:
        identified = 0

    fileLine = fileIn.readline()
    fileLine = fileLine.strip()

print('total length: ' + str(len(sequence2)))

f = open(outfile + '.txt','a+')
f.write("\t".join(("start", "end", "good_sites", "nonSNP", "SNP", "\n")))

start = 0
end = int(window)
windcount = 0

while int(end) <= len(sequence1):

    windcount += 1

    if windcount % 100 == 0:
        print(str(windcount) + ' windows done')

    seq1 = sequence1[start:end]
    seq2 = sequence2[start:end]

    good = 0
    countSame = 0
    countDiff = 0
    for i in range(0,len(seq1)):

        base1 = seq1[i]
        base2 = seq2[i]

        if base1 != '-' and base1 != 'N' and base2 != '-' and base2 != 'N':
            good += 1

            if base1 == base2:
                countSame +=1
            if base1 != base2:
                countDiff +=1

```

```
f.write("\t".join((str(start),
                  str(end),
                  str(good),
                  str(countSame),
                  str(countDiff), "\n")))

start += int(window)
end += int(window)

print('all done')
f.close()
```

## Supplemental Code S5. Transform .xmfa file to fasta.

```
#!/usr/bin/env python

# Use as: python seq-seq-pan_toFasta.py -I SeqSeqPan_erato_cha1_cha2_melp_noNewline.xmfa -g 1,2,3,4

from __future__ import division
from Bio.Seq import Seq
import argparse, re

parser = argparse.ArgumentParser()
parser.add_argument("-I", "--infile", help=" xmfa file with no newlines from seq-seq-pan", required =
True)
parser.add_argument("-g", "--genomeList", help="comma seperated list of genomes in order of seq-seq-pan",
required = True)
args = parser.parse_args()

infile = args.infile
genomes = args.genomeList.split(",")

count = 0

for genome in genomes:

    identified = 0
    sequence = str()

    f= open(genome + '_fasta.fasta',"a+")
    count +=1
    fileIn = open(infile, "r")
    fileLine = fileIn.readline()
    fileLine = fileLine.strip()

    print('searching genome ' + str(count) + ' ' + str(genome))

    while fileLine:

        if '> ' + str(genome) + ':' in fileLine:

            identified += 1

            fileLine = fileIn.readline()
            fileLine = fileLine.strip()
            #seq = Seq(fileLine)

            sequence = sequence + fileLine

        if '=' not in fileLine and '>' not in fileLine:
            lenSeq = len(fileLine)

        if '=' in fileLine and identified == 0:
            sequence = sequence + '-' * lenSeq

        if '=' in fileLine and identified == 1:
            identified = 0

        fileLine = fileIn.readline()
        fileLine = fileLine.strip()

    print('total length: ' + str(len(sequence)))

    f.write("".join((">genome_"+str(genome),"\n")))
    f.write("".join((sequence,"\n")))

    sequence = str()
```

## Supplemental Code S6. ATAC-seq read mapping, peak calling and counting.

```
## Note that this bash code is written for a slurm array job and would still require to add slurm job
submission parameters

# Array ID

indID=$((SLURM_ARRAY_TASK_ID -1))

## samples

sampleList1=(BR10_Demophoon_Brain BR12_Demophoon_Brain BR15_Demophoon_Brain CH_head2 E4-Head
BR11_Rosina_Brain BR14_Demophoon_Brain BR5_Rosina_Brain CH_head6 M4-Head CH_FW2 CH_FW3 CH_FW7 CH_HW2
CH_HW3 CH_HW7)

## Trim reads

java -jar trimmomatic-0.39.jar PE -threads 1 -phred33 \
reads_ATACseq/$(echo "${sampleList1[indID]}")_R1.fastq.gz reads_ATACseq/$(echo
"${sampleList1[indID]}")_R2.fastq.gz \
trimmed/$(echo "${sampleList1[indID]}")_trim_R1.fastq.gz / trimmed/$(echo
"${sampleList1[indID]}")_trim_unpair_R1.fastq.gz \
trimmed/$(echo "${sampleList1[indID]}")_trim_R2.fastq.gz trimmed/$(echo
"${sampleList1[indID]}")_trim_unpair_R2.fastq.gz \
ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:keepBothReads LEADING:3 TRAILING:3 MINLEN:32

## Map reads to respective reference genome (bowtie reference)

REF=H_erato_dem/Herato_final_bowtie
#REF=Hmel2/Hmel2_bowtie
#REF=H_charithonia_PR/Hcha.assembly.v1.1_bowtie
#REF=H_charithonia_PR/H_charithonia_10X.2_bowtie

REFNAME=Herato
#REFNAME=Hmel2
#REFNAME=Hcha1
#REFNAME=Hcha2

## Example create size file using samtools
#samtools faidx H_erato_dem/Herato_final.fasta; cut -f1,2 H_erato_dem/Herato_final.fasta >
H_erato_dem/Herato_final.fasta.sizes

SIZES=H_erato_dem/Herato_final.fasta.sizes
#SIZES=Hmel2/Hmel2.fa.sizes
#SIZES=H_charithonia_PR/Hcha.assembly.v1.1.fasta.sizes
#SIZES=H_charithonia_PR/H_charithonia_10X.2.fasta.sizes

## Run bowtie

gunzip trimmed/$(echo "${samples[ID]}")_trim_R1.fastq.gz
gunzip trimmed/$(echo "${samples[ID]}")_trim_R2.fastq.gz

bowtie2 -t -k 2 -p 8 --local -x $REF \
-1 trimmed/$(echo "${samples[ID]}")_trim_R1.fastq \
-2 trimmed/$(echo "${samples[ID]}")_trim_R2.fastq |\
samtools view -bS - > BAM/$(echo "${samples[ID]}")_${REFNAME}.trim.bam

gzip trimmed/$(echo "${samples[ID]}")_trim_R1.fastq
gzip trimmed/$(echo "${samples[ID]}")_trim_R2.fastq

## Filter bam files
```

```

samtools view -f 0x02 -q 20 -b BAM/$(echo "${samples[ID]}")_$_REFNAME.trim.bam > BAM/$(echo
"${samples[ID]}")_$_REFNAME.trim.filtered.bam

## Sort bam files

samtools sort BAM/$(echo "${samples[ID]}")_$_REFNAME.trim.filtered.bam BAM/$(echo
"${samples[ID]}")_$_REFNAME.trim.filtered.sorted

## Remove PCR duplicates

java -jar picard.jar MarkDuplicates \
I=BAM/$(echo "${samples[ID]}")_$_REFNAME.trim.filtered.sorted.bam \
O=BAM/$(echo "${samples[ID]}")_$_REFNAME.trim.filtered.sorted.nd.bam \
Remove_Duplicates=true M=BAM/$(echo "${samples[ID]}")_dup_metrics.txt ASSUME_SORTED=true

## Remove intermediate files

rm BAM/$(echo "${samples[ID]}")_$_REFNAME.trim.filtered.bam
rm BAM/$(echo "${samples[ID]}")_$_REFNAME.trim.filtered.sorted.bam

## Run MACS2 for peak calling

macs2 callpeak \
-t BAM/$(echo "${samples[ID]}")_Herato.trim.filtered.sorted.nd.bam \
-n $(echo "${samples[ID]}") --outdir MACS2_sub/ \
-f BAMPE -g 382844248 --nomodel --shift -100 --extsize 200

#macs2 callpeak \
#-t BAM/$(echo "${samples[ID]}")_Hmel2.trim.filtered.sorted.nd.bam \
#-n $(echo "${samples[ID]}") --outdir MACS2_sub/ \
#-f BAMPE -g 275198613 --nomodel --shift -100 --extsize 200

#macs2 callpeak \
#-t BAM/$(echo "${samples[ID]}")_Hcha1.trim.filtered.sorted.nd.bam \
#-n $(echo "${samples[ID]}")_Hcha1 --outdir MACS2_sub/ \
#-f BAMPE -g 355179777 --nomodel --shift -100 --extsize 200

#macs2 callpeak \
#-t BAM/$(echo "${samples[ID]}")_Hcha2.trim.filtered.sorted.nd.bam \
#-n $(echo "${samples[ID]}")_Hcha2 --outdir MACS2_sub/ \
#-f BAMPE -g 357058320 --nomodel --shift -100 --extsize 200

## Merge MACS2 peaks for samples of each species

# Make sample list command
sampleList1=(BR14_Demophoon_Brain BR12_Demophoon_Brain)
sampleList2=(BR11_Rosina_Brain M4-Head)
sampleList3=(CH_head2 CH_head6)

ALL_LIST=""
for FILE in ${sampleList1[*]}
do
ALL_LIST="$ALL_LIST $FILE"_peaks.narrowPeak"
done

eval command=\${$(echo ALL_LIST)}

# cat bed files
cat $(echo $command) > H_e_dem_peaks.bed

ALL_LIST=""
for FILE in ${sampleList2[*]}
do
ALL_LIST="$ALL_LIST $FILE"_peaks.narrowPeak"
done

eval command=\${$(echo ALL_LIST)}

```

```

# cat bed files
cat $(echo $command) > H_m_ros_peaks.bed

ALL_LIST=""
for FILE in ${sampleList3[*]}
do
ALL_LIST="$ALL_LIST $FILE"_Hcha1_peaks.narrowPeak""
done

eval command=\${$(echo ALL_LIST)}

# cat bed files
cat $(echo $command) > H_cha1_peaks.bed

ALL_LIST=""
for FILE in ${sampleList3[*]}
do
ALL_LIST="$ALL_LIST $FILE"_Hcha2_peaks.narrowPeak""
done

eval command=\${$(echo ALL_LIST)}

# cat bed files
cat $(echo $command) > H_cha2_peaks.bed

## Sort bed files

sort -k1,1 -k2,2n H_e_dem_peaks.bed > H_e_dem_peaks.sort.bed
sort -k1,1 -k2,2n H_m_ros_peaks.bed > H_m_ros_peaks.sort.bed
sort -k1,1 -k2,2n H_cha1_peaks.bed > H_cha1_peaks.sort.bed
sort -k1,1 -k2,2n H_cha2_peaks.bed > H_cha2_peaks.sort.bed

## Retain only peaks that were found in both head samples with at least 25% reciprocal overlap

# intersect 25%

bedtools intersect -wa -f 0.25 -a BR14_Demophoon_Brain_peaks.narrowPeak -b
BR12_Demophoon_Brain_peaks.narrowPeak > H_e_dem_peaks_intersect_25_A.bed
bedtools intersect -wa -f 0.25 -a BR12_Demophoon_Brain_peaks.narrowPeak -b
BR14_Demophoon_Brain_peaks.narrowPeak > H_e_dem_peaks_intersect_25_B.bed

bedtools intersect -wa -f 0.25 -a BR11_Rosina_Brain_peaks.narrowPeak -b M4-Head_peaks.narrowPeak >
H_m_ros_peaks_intersect_25_A.bed
bedtools intersect -wa -f 0.25 -a M4-Head_peaks.narrowPeak -b BR11_Rosina_Brain_peaks.narrowPeak >
H_m_ros_peaks_intersect_25_B.bed

bedtools intersect -wa -f 0.25 -a CH_head2_Hcha1_peaks.narrowPeak -b CH_head6_Hcha1_peaks.narrowPeak >
H_cha1_peaks_intersect_25_A.bed
bedtools intersect -wa -f 0.25 -a CH_head6_Hcha1_peaks.narrowPeak -b CH_head2_Hcha1_peaks.narrowPeak >
H_cha1_peaks_intersect_25_B.bed

bedtools intersect -wa -f 0.25 -a CH_head2_Hcha2_peaks.narrowPeak -b CH_head6_Hcha2_peaks.narrowPeak >
H_cha2_peaks_intersect_25_A.bed
bedtools intersect -wa -f 0.25 -a CH_head6_Hcha2_peaks.narrowPeak -b CH_head2_Hcha2_peaks.narrowPeak >
H_cha2_peaks_intersect_25_B.bed

cat H_e_dem_peaks_intersect_25_A.bed H_e_dem_peaks_intersect_25_B.bed | sort -k1,1 -k2,2n | bedtools merge
-i - > H_e_dem_peaks_intersect_25.merged.sort.bed
cat H_m_ros_peaks_intersect_25_A.bed H_m_ros_peaks_intersect_25_B.bed | sort -k1,1 -k2,2n | bedtools merge
-i - > H_m_ros_peaks_intersect_25.merged.sort.bed
cat H_cha1_peaks_intersect_25_A.bed H_cha1_peaks_intersect_25_B.bed | sort -k1,1 -k2,2n | bedtools merge -
i - > H_cha1_peaks_intersect_25.merged.sort.bed
cat H_cha2_peaks_intersect_25_A.bed H_cha2_peaks_intersect_25_B.bed | sort -k1,1 -k2,2n | bedtools merge -
i - > H_cha2_peaks_intersect_25.merged.sort.bed

```

```

## Make bigwig (.bw) files from read coverage. Make sure to set the correct $REFNAME and $SIZES

bedtools genomecov \
-ibam BAM/$(echo "${samples[ID]}")_$REFNAME.trim.filtered.sorted.nd.bam -bg \
> bw/$(echo "${samples[ID]}")_$REFNAME.trim.filtered.sorted.nd.bdg

LC_COLLATE=C sort -k1,1 -k2,2n bw/$(echo "${samples[ID]}")_$REFNAME.trim.filtered.sorted.nd.bdg \
> bw/$(echo "${samples[ID]}")_$REFNAME.trim.filtered.sorted.nd.collate.bdg

rm bw/$(echo "${samples[ID]}")_$REFNAME.trim.filtered.sorted.nd.bdg

bedGraphToBigWig \
bw/$(echo "${samples[ID]}")_$REFNAME.trim.filtered.sorted.nd.collate.bdg \
$SIZES \
bw/$(echo "${samples[ID]}")_$REFNAME.trim.filtered.sorted.nd.collate.bw

## Make .bw files that represent scaled means of replicates (scaling factors can be obtained from DeSeq2
from the counts)

# Transform .bam file to bedgraph format using scaling factor

bedtools genomecov -ibam BAM/BR10_Demophoon_Brain_Herato.trim.filtered.sorted.nd.bam -bg -scale 0.4983282
-g H_erato_dem/Herato_final.fasta.sizes > bg/BR10_Demophoon_Brain_Herato_normalised.bg
bedtools genomecov -ibam BAM/BR12_Demophoon_Brain_Herato.trim.filtered.sorted.nd.bam -bg -scale 0.6296673
-g H_erato_dem/Herato_final.fasta.sizes > bg/BR12_Demophoon_Brain_Herato_normalised.bg
bedtools genomecov -ibam BAM/BR15_Demophoon_Brain_Herato.trim.filtered.sorted.nd.bam -bg -scale 0.8011250
-g H_erato_dem/Herato_final.fasta.sizes > bg/BR15_Demophoon_Brain_Herato_normalised.bg
bedtools genomecov -ibam BAM/BR14_Demophoon_Brain_Herato.trim.filtered.sorted.nd.bam -bg -scale 2.7586922
-g H_erato_dem/Herato_final.fasta.sizes > bg/BR14_Demophoon_Brain_Herato_normalised.bg
bedtools genomecov -ibam BAM/E4-Head_Herato.trim.filtered.sorted.nd.bam -bg -scale 1.4405454 -g
H_erato_dem/Herato_final.fasta.sizes > bg/E4-Head_Herato_normalised.bg
bedtools genomecov -ibam BAM/BR11_Rosina_Brain_Hmel2.trim.filtered.sorted.nd.bam -bg -scale 0.7861225 -g
Hmel2/Hmel2.fa.sizes > bg/BR11_Rosina_Brain_Hmel2_normalised.bg
bedtools genomecov -ibam BAM/BR5_Rosina_Brain_Hmel2.trim.filtered.sorted.nd.bam -bg -scale 0.5333467 -g
Hmel2/Hmel2.fa.sizes > bg/BR5_Rosina_Brain_Hmel2_normalised.bg
bedtools genomecov -ibam BAM/M4-Head_Hmel2.trim.filtered.sorted.nd.bam -bg -scale 2.6462821 -g
Hmel2/Hmel2.fa.sizes > bg/M4-Head_Hmel2_normalised.bg
bedtools genomecov -ibam BAM/CH_head2_Hcha1.trim.filtered.sorted.nd.bam -bg -scale 0.6003002 -g
H_charithonia_PR/Hcha.assembly.v1.1_bowtie > bg/CH_head2_Hcha1_normalised.bg
bedtools genomecov -ibam BAM/CH_head6_Hcha1.trim.filtered.sorted.nd.bam -bg -scale 1.6658331 -g
H_charithonia_PR/Hcha.assembly.v1.1_bowtie > bg/CH_head6_Hcha1_normalised.bg
bedtools genomecov -ibam BAM/CH_head2_Hcha2.trim.filtered.sorted.nd.bam -bg -scale 0.6011422 -g
H_charithonia_PR/Hcha.assembly.v1.1_bowtie > bg/CH_head2_Hcha2_normalised.bg
bedtools genomecov -ibam BAM/CH_head6_Hcha2.trim.filtered.sorted.nd.bam -bg -scale 1.6634998 -g
H_charithonia_PR/Hcha.assembly.v1.1_bowtie > bg/CH_head6_Hcha2_normalised.bg

# Combine bedgraph files and calculate mean

wiggletools write_bg brain_H_e_dem_normalized_mean.bg mean BR10_Demophoon_Brain_Herato_normalised.bg
BR12_Demophoon_Brain_Herato_normalised.bg BR14_Demophoon_Brain_Herato_normalised.bg
BR15_Demophoon_Brain_Herato_normalised.bg E4-Head_Herato_normalised.bg
wiggletools write_bg brain_H_e_ros_normalized_mean.bg mean BR11_Rosina_Brain_Hmel2_normalised.bg
BR5_Rosina_Brain_Hmel2_normalised.bg M4-Head_Hmel2_normalised.bg
wiggletools write_bg brain_H_cha1_normalized_mean.bg mean CH_head2_Hcha1_normalised.bg
CH_head6_Hcha1_normalised.bg
wiggletools write_bg brain_H_cha2_normalized_mean.bg mean CH_head2_Hcha2_normalised.bg
CH_head6_Hcha2_normalised.bg

## Count reads in ATAC-seq peaks

# Index .bam files

samtools index BAM/$(echo "${samples[ID]}")_$REFNAME.trim.filtered.sorted.nd.bam

# make sample list command
ALL_LIST1=""
for FILE in ${sampleList1[*]}

```

```

do
ALL_LIST1="$ALL_LIST1 BAM/${FILE}_Herato.trim.filtered.sorted.nd.bam"
done
echo $ALL_LIST1

## make sample list command
ALL_LIST2=""
for FILE in ${sampleList2[*]}
do
ALL_LIST2="$ALL_LIST2 BAM/${FILE}_Hmel2.trim.filtered.sorted.nd.bam"
done
echo $ALL_LIST2

## make sample list command
ALL_LIST3=""
for FILE in ${sampleList3[*]}
do
ALL_LIST3="$ALL_LIST3 BAM/${FILE}_Hcha1.trim.filtered.sorted.nd.bam"
done
echo $ALL_LIST3

## make sample list command
ALL_LIST4=""
for FILE in ${sampleList3[*]}
do
ALL_LIST4="$ALL_LIST4 BAM/${FILE}_Hcha2.trim.filtered.sorted.nd.bam"
done
echo $ALL_LIST4

eval command1=\$(echo ALL_LIST1)
eval command2=\$(echo ALL_LIST2)
eval command3=\$(echo ALL_LIST3)
eval command4=\$(echo ALL_LIST4)

# count reads in peaks

bedtools multicov -bams $(echo $command1) -bed H_e_dem_peaks_intersect_25.merged.sort.bed >
H_e_dem_peaks_intersect_25.counts
#bedtools multicov -bams $(echo $command2) -bed H_m_ros_peaks_intersect_25.merged.sort.bed >
H_m_ros_peaks_intersect_25.counts
bedtools multicov -bams $(echo $command3) -bed H_cha1_peaks_ALL_m1.merged.sort.bed >
H_cha1_peaks_ALL.counts
#bedtools multicov -bams $(echo $command4) -bed H_cha2_peaks_intersect_25.merged.sort.bed >
H_cha2_peaks_intersect_25.counts

```

**Supplemental Code S7. Python code to transform scaffold positions to genome positions for mapping with seq-seq-pan map.**

```
#!/usr/bin/env python

# Use as: se-seq-pan_bedgraph_chrompos.py -I start_end/H_e_dem_peaks.bed -g 1 -p genome_pos_H_e_dem.txt -o
start_end/H_e_dem_peaks

# genome_pos_H_e_dem.txt
# Scaf          Length      CummStart
# Herato0101     22325789    0
# Herato0201     28654      22325789
# Herato0202     186807     22354443
# ...           ...      ...

from __future__ import division
import argparse, re

parser = argparse.ArgumentParser()
parser.add_argument("-I", "--infile", help="bed file", required = True)
parser.add_argument("-g", "--genomeN", help="genome number in pan genome", required = True)
parser.add_argument("-p", "--scafgenomepos", help="table with cumulative genome positions of scaffolds",
required = True)
parser.add_argument("-o", "--outfile", help="Output txt file name", action = "store")
args = parser.parse_args()

infile = args.infile
outfile = args.outfile
genomeNumber = args.genomeN
scafpos = args.scafgenomepos

s= open(outfile + '_start.txt',"a+")
s.write("\t".join((genomeNumber,"c","\n")))

e= open(outfile + '_end.txt',"a+")
e.write("\t".join((genomeNumber,"c","\n")))

posIn = open(scafpos, "r")
posLine = posIn.readline()

posDic = {}
while posline:

    lineElem = posLine.split('\t')
    posDic[lineElem[0]] = lineElem[2]
    posLine = posIn.readline()

fileIn = open(infile, "r")
fileLine = fileIn.readline()

n = 1
while fileLine:
    lineElem = fileLine.split('\t')

    if(n ==1):
        start = int(lineElem[1]) + int(posDic[lineElem[0]]) + 1
    else:
        start = int(lineElem[1]) + int(posDic[lineElem[0]])

    end = int(lineElem[2]) + int(posDic[lineElem[0]])

    s.write("\t".join((str(start),"\n")))
    e.write("\t".join((str(end),"\n")))

    fileLine = fileIn.readline()

    n+=1
```

**Supplemental Code S8. ATAC-seq interval mapping to pan-genome. This code can be modified to map any feature such as TEs to the pan genome as well.**

```
# Make bedfile from counts file
cut -d$'\t' -f 1-3 H_e_dem_peaks_intersect_25.counts > start_end/H_e_dem_peaks.bed
cut -d$'\t' -f 1-3 H_m_ros_peaks_intersect_25.counts > start_end/H_m_ros_peaks.bed
cut -d$'\t' -f 1-3 H_cha1_peaks_intersect_25.counts > start_end/H_cha1_peaks.bed
cut -d$'\t' -f 1-3 H_cha2_peaks_intersect_25.counts > start_end/H_cha2_peaks.bed

# Transform coordinates to genome coordinates
python se-seq-pan_bedgraph_chrompos.py -I start_end/H_e_dem_peaks.bed -g 1 -p genome_pos_H_e_dem.txt -o
start_end/H_e_dem_peaks
python se-seq-pan_bedgraph_chrompos.py -I start_end/H_m_ros_peaks.bed -g 4 -p genome_pos_H_m_melp.txt -o
start_end/H_m_ros_peaks
python se-seq-pan_bedgraph_chrompos.py -I start_end/H_cha1_peaks.bed -g 2 -p genome_pos_H_cha1.txt -o
start_end/H_cha1_peaks
python se-seq-pan_bedgraph_chrompos.py -I start_end/H_cha2_peaks.bed -g 3 -p genome_pos_H_cha2.txt -o
start_end/H_cha2_peaks

# Map coords
seq-seq-pan map -c SeqSeqPan_erato_melp_char_consensus.fasta -p ./ -n start_end/H_e_dem_peaks_start_pan -i
start_end/H_e_dem_peaks_start.txt
seq-seq-pan map -c SeqSeqPan_erato_melp_char_consensus.fasta -p ./ -n start_end/H_m_ros_peaks_start_pan -i
start_end/H_m_ros_peaks_start.txt
seq-seq-pan map -c SeqSeqPan_erato_melp_char_consensus.fasta -p ./ -n start_end/H_cha1_peaks_start_pan -i
start_end/H_cha1_peaks_start.txt
seq-seq-pan map -c SeqSeqPan_erato_melp_char_consensus.fasta -p ./ -n start_end/H_cha2_peaks_start_pan -i
start_end/H_cha2_peaks_start.txt

seq-seq-pan map -c SeqSeqPan_erato_melp_char_consensus.fasta -p ./ -n start_end/H_e_dem_peaks_end_pan -i
start_end/H_e_dem_peaks_end.txt
seq-seq-pan map -c SeqSeqPan_erato_melp_char_consensus.fasta -p ./ -n start_end/H_m_ros_peaks_end_pan -i
start_end/H_m_ros_peaks_end.txt
seq-seq-pan map -c SeqSeqPan_erato_melp_char_consensus.fasta -p ./ -n start_end/H_cha1_peaks_end_pan -i
start_end/H_cha1_peaks_end.txt
seq-seq-pan map -c SeqSeqPan_erato_melp_char_consensus.fasta -p ./ -n start_end/H_cha2_peaks_end_pan -i
start_end/H_cha2_peaks_end.txt

# Combine transformed start and end position and gave each the scaffold name 'pan'
lineN=$(  
start_end/H_e_dem_peaks_start.txt wc -l)  
pr -mts <(yes "pan" | head -n $lineN) <(cut -d$'\t' -f 2 start_end/H_e_dem_peaks_start_pan.txt) <(cut -  
d$'\t' -f 2 start_end/H_e_dem_peaks_end_pan.txt) > start_end/H_e_dem_peaks_start_end_pan.txt

lineN=$(  
start_end/H_m_ros_peaks_start.txt wc -l)  
pr -mts <(yes "pan" | head -n $lineN) <(cut -d$'\t' -f 2 start_end/H_m_ros_peaks_start_pan.txt) <(cut -  
d$'\t' -f 2 start_end/H_m_ros_peaks_end_pan.txt) > start_end/H_m_ros_peaks_start_end_pan.txt

lineN=$(  
start_end/H_cha1_peaks_start.txt wc -l)  
pr -mts <(yes "pan" | head -n $lineN) <(cut -d$'\t' -f 2 start_end/H_cha1_peaks_start_pan.txt) <(cut -  
d$'\t' -f 2 start_end/H_cha1_peaks_end_pan.txt) > start_end/H_cha1_peaks_start_end_pan.txt

lineN=$(  
start_end/H_cha2_peaks_start.txt wc -l)  
pr -mts <(yes "pan" | head -n $lineN) <(cut -d$'\t' -f 2 start_end/H_cha2_peaks_start_pan.txt) <(cut -  
d$'\t' -f 2 start_end/H_cha2_peaks_end_pan.txt) > start_end/H_cha2_peaks_start_end_pan.txt

# Sort rows (so that intervals have the start < end, which is sometimes not the case because of inversions  
and it crashes bedtools later)  
lineN=$(  
start_end/H_e_dem_peaks_start_end_pan.txt wc -l)  
pr -mts <(yes "pan" | head -n $lineN) <(awk ' {split( $0, a, " " ); asort( a ); for( i = 1; i <=  
length(a); i++ ) printf( "%s\t", a[i] ); printf( "\n" ); }' <(cut -d$'\t' -f 2,3  
start_end/H_e_dem_peaks_start_end_pan.txt)) > start_end/H_e_dem_peaks_start_end_pan.rowsort.txt
```

```

lineN=$(< start_end/H_m_ros_peaks_start_end_pan.txt wc -l)
pr -mts <(yes "pan" | head -n $lineN) <(awk ' {split( $0, a, " "); asort( a ); for( i = 1; i <=
length(a); i++ ) printf( "%s\t", a[i] ); printf( "\n" ); }' <(cut -d$'\t' -f 2,3
start_end/H_m_ros_peaks_start_end_pan.txt)) > start_end/H_m_ros_peaks_start_end_pan.rowsort.txt

lineN=$(< start_end/H_cha1_peaks_start_end_pan.txt wc -l)
pr -mts <(yes "pan" | head -n $lineN) <(awk ' {split( $0, a, " "); asort( a ); for( i = 1; i <=
length(a); i++ ) printf( "%s\t", a[i] ); printf( "\n" ); }' <(cut -d$'\t' -f 2,3
start_end/H_cha1_peaks_start_end_pan.txt)) > start_end/H_cha1_peaks_start_end_pan.rowsort.txt

lineN=$(< start_end/H_cha1_peaks_start_end_pan.txt wc -l)
pr -mts <(yes "pan" | head -n $lineN) <(awk ' {split( $0, a, " "); asort( a ); for( i = 1; i <=
length(a); i++ ) printf( "%s\t", a[i] ); printf( "\n" ); }' <(cut -d$'\t' -f 2,3
start_end/H_cha2_peaks_start_end_pan.txt)) > start_end/H_cha2_peaks_start_end_pan.rowsort.txt

# Remove tabs at the end of each row (otherwise this crashes bedtools)
sed -i 's/[[:space:]]*$//' start_end/H_e_dem_peaks_start_end_pan.rowsort.txt
sed -i 's/[[:space:]]*$//' start_end/H_m_ros_peaks_start_end_pan.rowsort.txt
sed -i 's/[[:space:]]*$//' start_end/H_cha1_peaks_start_end_pan.rowsort.txt
sed -i 's/[[:space:]]*$//' start_end/H_cha2_peaks_start_end_pan.rowsort.txt

```

## Supplemental Code S9. ATAC-seq .bedgraph mapping to pan genome.

```
## Note that this bash code is written for a slurm array job and would still require to add slurm job
submission parameters. Code in gray are pre- and postprocessing steps.

# Bin bedgraph files in 30 bp windows
bedtools makewindows -g Herato_final.fasta.sizes -w 30 | bedtools map -a - -b
bg/brain_H_e_dem_normalized_mean.bg -c 4 -o mean > bg/brain_H_e_dem_normalized_mean.w30s0bin.bg
bedtools makewindows -g Hcha.assembly.v1.1.fasta.sizes -w 30 | bedtools map -a - -b
bg/brain_H_cha1_normalized_mean.bg -c 4 -o mean > bg/brain_H_cha1_normalized_mean.w30s0bin.bg
bedtools makewindows -g H_charithonia_10X.2.fasta.sizes -w 30 | bedtools map -a - -b
bg/brain_H_cha2_normalized_mean.bg -c 4 -o mean > bg/brain_H_cha2_normalized_mean.w30s0bin.bg
bedtools makewindows -g Hmel2.fa.sizes -w 30 | bedtools map -a - -b bg/brain_H_m_ros_normalized_mean.bg -c
4 -o mean > bg/brain_H_m_ros_normalized_mean.w30s0bin.bg

# Modify bedgraph to genome positions
python se-seq-pan_bedgraph_chrompos.py -I bg/brain_H_e_dem_normalized_mean.w30s0bin.bg -g 1 -p
genome_pos_H_e_dem.txt -o bg/start_end/brain_H_e_dem_normalized_mean
python se-seq-pan_bedgraph_chrompos.py -I bg/brain_H_cha1_normalized_mean.w30s0bin.bg -g 2 -p
genome_pos_H_cha1.txt -o bg/start_end/brain_H_cha1_normalized_mean
python se-seq-pan_bedgraph_chrompos.py -I bg/brain_H_cha2_normalized_mean.w30s0bin.bg -g 3 -p
genome_pos_H_cha2.txt -o bg/start_end/brain_H_cha2_normalized_mean
python se-seq-pan_bedgraph_chrompos.py -I bg/brain_H_m_ros_normalized_mean.w30s0bin.bg -g 4 -p
genome_pos_H_m_melp.txt -o bg/start_end/brain_H_m_ros_normalized_mean

sort -u bg/start_end/brain_H_e_dem_normalized_mean_start.txt | wc -l
sort -u bg/start_end/brain_H_cha1_normalized_mean_start.txt | wc -l
sort -u bg/start_end/brain_H_cha2_normalized_mean_start.txt | wc -l
sort -u bg/start_end/brain_H_m_ros_normalized_mean_start.txt | wc -l

sort -u bg/start_end/brain_H_e_dem_normalized_mean_end.txt | wc -l
sort -u bg/start_end/brain_H_cha1_normalized_mean_end.txt | wc -l
sort -u bg/start_end/brain_H_cha2_normalized_mean_end.txt | wc -l
sort -u bg/start_end/brain_H_m_ros_normalized_mean_end.txt | wc -l

# Remove first line
sed -i -n -e '2,$p' bg/start_end/brain_H_e_dem_normalized_mean_start.txt
sed -i -n -e '2,$p' bg/start_end/brain_H_e_dem_normalized_mean_end.txt

sed -i -n -e '2,$p' bg/start_end/brain_H_cha1_normalized_mean_start.txt
sed -i -n -e '2,$p' bg/start_end/brain_H_cha1_normalized_mean_end.txt

sed -i -n -e '2,$p' bg/start_end/brain_H_cha2_normalized_mean_start.txt
sed -i -n -e '2,$p' bg/start_end/brain_H_cha2_normalized_mean_end.txt

sed -i -n -e '2,$p' bg/start_end/brain_H_m_ros_normalized_mean_start.txt
sed -i -n -e '2,$p' bg/start_end/brain_H_m_ros_normalized_mean_end.txt

# Split files with each 100000 lines of positions to transform
mkdir bg/start_end/split
split -l 100000 -a 3 -d bg/start_end/brain_H_e_dem_normalized_mean_start.txt
bg/start_end/split/brain_H_e_dem_normalized_mean_start_
split -l 100000 -a 3 -d bg/start_end/brain_H_e_dem_normalized_mean_end.txt
bg/start_end/split/brain_H_e_dem_normalized_mean_end_

split -l 100000 -a 3 -d bg/start_end/brain_H_cha1_normalized_mean_start.txt
bg/start_end/split/brain_H_cha1_normalized_mean_start_
split -l 100000 -a 3 -d bg/start_end/brain_H_cha1_normalized_mean_end.txt
bg/start_end/split/brain_H_cha1_normalized_mean_end_

split -l 100000 -a 3 -d bg/start_end/brain_H_cha2_normalized_mean_start.txt
bg/start_end/split/brain_H_cha2_normalized_mean_start_
split -l 100000 -a 3 -d bg/start_end/brain_H_cha2_normalized_mean_end.txt
bg/start_end/split/brain_H_cha2_normalized_mean_end_

split -l 100000 -a 3 -d bg/start_end/brain_H_m_ros_normalized_mean_start.txt
bg/start_end/split/brain_H_m_ros_normalized_mean_start_
```

```

split -l 100000 -a 3 -d bg/start_end/brain_H_m_ros_normalized_mean_end.txt
bg/start_end/split/brain_H_m_ros_normalized_mean_end_

# Array ID

indID=$((SLURM_ARRAY_TASK_ID -1))

# List used for filename numbering
number=(000 001 002 003 004 005 006 007 008 009 010 011 012 013 014 015 016 017 018 019 020 021 022 023
024 025 026 027 028 029 030 031 032 033 034 035 036 037 038 039 040 041 042 043 044 045 046 047 048 049
050 051 052 053 054 055 056 057 058 059 060 061 062 063 064 065 066 067 068 069 070 071 072 073 074 075
076 077 078 079 080 081 082 083 084 085 086 087 088 089 090 091 092 093 094 095 096 097 098 099 100 101
102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127
128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153
154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179
180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205
206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231
232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257
258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283
284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309
310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335
336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361
362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387
388 389 390 391 392 393 394 395 396 397 398 399 400)

# Run 128 jobs for H_e_dem
cat bg/head_1.txt bg/start_end/split/brain_H_e_dem_normalized_mean_start_${echo "${number[ID]}"} >
bg/start_end/split/brain_H_e_dem_normalized_mean_start_${echo "${number[ID]}"}_head
cat bg/head_1.txt bg/start_end/split/brain_H_e_dem_normalized_mean_end_${echo "${number[ID]}"} >
bg/start_end/split/brain_H_e_dem_normalized_mean_end_${echo "${number[ID]}"}_head

seq-seq-pan map -c SeqSeqPan_erato_melp_char_consensus.fasta -p ./ -n
bg/start_end/split/brain_H_e_dem_normalized_mean_start_pan_${echo "${number[ID]}"} -i
bg/start_end/split/brain_H_e_dem_normalized_mean_start_${echo "${number[ID]}"}_head
seq-seq-pan map -c SeqSeqPan_erato_melp_char_consensus.fasta -p ./ -n
bg/start_end/split/brain_H_e_dem_normalized_mean_end_pan_${echo "${number[ID]}"} -i
bg/start_end/split/brain_H_e_dem_normalized_mean_end_${echo "${number[ID]}"}_head

# Run 119 jobs for H_cha1
cat bg/head_2.txt bg/start_end/split/brain_H_cha1_normalized_mean_start_${echo "${number[ID]}"} >
bg/start_end/split/brain_H_cha1_normalized_mean_start_${echo "${number[ID]}"}_head
cat bg/head_2.txt bg/start_end/split/brain_H_cha1_normalized_mean_end_${echo "${number[ID]}"} >
bg/start_end/split/brain_H_cha1_normalized_mean_end_${echo "${number[ID]}"}_head

seq-seq-pan map -c SeqSeqPan_erato_melp_char_consensus.fasta -p ./ -n
bg/start_end/split/brain_H_cha1_normalized_mean_start_pan_${echo "${number[ID]}"} -i
bg/start_end/split/brain_H_cha1_normalized_mean_start_${echo "${number[ID]}"}_head
seq-seq-pan map -c SeqSeqPan_erato_melp_char_consensus.fasta -p ./ -n
bg/start_end/split/brain_H_cha1_normalized_mean_end_pan_${echo "${number[ID]}"} -i
bg/start_end/split/brain_H_cha1_normalized_mean_end_${echo "${number[ID]}"}_head

# Run 120 jobs for H_cha2
cat bg/head_3.txt bg/start_end/split/brain_H_cha2_normalized_mean_start_${echo "${number[ID]}"} >
bg/start_end/split/brain_H_cha2_normalized_mean_start_${echo "${number[ID]}"}_head
cat bg/head_3.txt bg/start_end/split/brain_H_cha2_normalized_mean_end_${echo "${number[ID]}"} >
bg/start_end/split/brain_H_cha2_normalized_mean_end_${echo "${number[ID]}"}_head

seq-seq-pan map -c SeqSeqPan_erato_melp_char_consensus.fasta -p ./ -n
bg/start_end/split/brain_H_cha2_normalized_mean_start_pan_${echo "${number[ID]}"} -i
bg/start_end/split/brain_H_cha2_normalized_mean_start_${echo "${number[ID]}"}_head
seq-seq-pan map -c SeqSeqPan_erato_melp_char_consensus.fasta -p ./ -n
bg/start_end/split/brain_H_cha2_normalized_mean_end_pan_${echo "${number[ID]}"} -i
bg/start_end/split/brain_H_cha2_normalized_mean_end_${echo "${number[ID]}"}_head

# Run 92 jobs for H_melp

```

```

cat bg/head_4.txt bg/start_end/split/brain_H_m_ros_normalized_mean_start_$(echo "${number[ID]}") >
bg/start_end/split/brain_H_m_ros_normalized_mean_start_$(echo "${number[ID]}")_head
cat bg/head_4.txt bg/start_end/split/brain_H_m_ros_normalized_mean_end_$(echo "${number[ID]}") >
bg/start_end/split/brain_H_m_ros_normalized_mean_end_$(echo "${number[ID]}")_head

seq-seq-pan map -c SeqSeqPan_erato_melp_char_consensus.fasta -p ./ -n
bg/start_end/split/brain_H_m_ros_normalized_mean_start_pan_$(echo "${number[ID]}") -i
bg/start_end/split/brain_H_m_ros_normalized_mean_start_$(echo "${number[ID]}")_head
seq-seq-pan map -c SeqSeqPan_erato_melp_char_consensus.fasta -p ./ -n
bg/start_end/split/brain_H_m_ros_normalized_mean_end_pan_$(echo "${number[ID]}") -i
bg/start_end/split/brain_H_m_ros_normalized_mean_end_$(echo "${number[ID]}")_head

# Combine mapped files
tail -n +2 bg/start_end/split/brain_H_e_dem_normalized_mean_start_pan_*.txt | grep -v '==>' | grep . | cut
-d$'\t' -f 2 > bg/start_end/brain_H_e_dem_normalized_mean_start_pan.txt
tail -n +2 bg/start_end/split/brain_H_e_dem_normalized_mean_end_pan_*.txt | grep -v '==>' | grep . | cut -
d$'\t' -f 2 > bg/start_end/brain_H_e_dem_normalized_mean_end_pan.txt

tail -n +2 bg/start_end/split/brain_H_cha1_normalized_mean_start_pan_*.txt | grep -v '==>' | grep . | cut
-d$'\t' -f 2 > bg/start_end/brain_H_cha1_normalized_mean_start_pan.txt
tail -n +2 bg/start_end/split/brain_H_cha1_normalized_mean_end_pan_*.txt | grep -v '==>' | grep . | cut -
d$'\t' -f 2 > bg/start_end/brain_H_cha1_normalized_mean_end_pan.txt

tail -n +2 bg/start_end/split/brain_H_cha2_normalized_mean_start_pan_*.txt | grep -v '==>' | grep . | cut
-d$'\t' -f 2 > bg/start_end/brain_H_cha2_normalized_mean_start_pan.txt
tail -n +2 bg/start_end/split/brain_H_cha2_normalized_mean_end_pan_*.txt | grep -v '==>' | grep . | cut -
d$'\t' -f 2 > bg/start_end/brain_H_cha2_normalized_mean_end_pan.txt

tail -n +2 bg/start_end/split/brain_H_m_ros_normalized_mean_start_pan_*.txt | grep -v '==>' | grep . | cut
-d$'\t' -f 2 > bg/start_end/brain_H_m_ros_normalized_mean_start_pan.txt
tail -n +2 bg/start_end/split/brain_H_m_ros_normalized_mean_end_pan_*.txt | grep -v '==>' | grep . | cut -
d$'\t' -f 2 > bg/start_end/brain_H_m_ros_normalized_mean_end_pan.txt

# Combine start and end columns
lineN=$(  
bg/start_end/brain_H_e_dem_normalized_mean_start_pan.txt wc -l)  
pr -mts <(yes "pan" | head -n $lineN) bg/start_end/brain_H_e_dem_normalized_mean_start_pan.txt  
bg/start_end/brain_H_e_dem_normalized_mean_end_pan.txt <(cut -d$'\t' -f 4  
bg/brain_H_e_dem_normalized_mean.w30s0bin.bg) >  
bg/start_end/brain_H_e_dem_normalized_mean_start_end_pan.bg

lineN=$(  
bg/start_end/brain_H_cha1_normalized_mean_start_pan.txt wc -l)  
pr -mts <(yes "pan" | head -n $lineN) bg/start_end/brain_H_cha1_normalized_mean_start_pan.txt  
bg/start_end/brain_H_cha1_normalized_mean_end_pan.txt <(cut -d$'\t' -f 4  
bg/brain_H_cha1_normalized_mean.w30s0bin.bg) > bg/start_end/brain_H_cha1_normalized_mean_start_end_pan.bg

lineN=$(  
bg/start_end/brain_H_cha2_normalized_mean_start_pan.txt wc -l)  
pr -mts <(yes "pan" | head -n $lineN) bg/start_end/brain_H_cha2_normalized_mean_start_pan.txt  
bg/start_end/brain_H_cha2_normalized_mean_end_pan.txt <(cut -d$'\t' -f 4  
bg/brain_H_cha2_normalized_mean.w30s0bin.bg) > bg/start_end/brain_H_cha2_normalized_mean_start_end_pan.bg

lineN=$(  
bg/start_end/brain_H_m_ros_normalized_mean_start_pan.txt wc -l)  
pr -mts <(yes "pan" | head -n $lineN) bg/start_end/brain_H_m_ros_normalized_mean_start_pan.txt  
bg/start_end/brain_H_m_ros_normalized_mean_end_pan.txt <(cut -d$'\t' -f 4  
bg/brain_H_m_ros_normalized_mean.w30s0bin.bg) >  
bg/start_end/brain_H_m_ros_normalized_mean_start_end_pan.bg

# Sort rows
lineN=$(  
bg/start_end/brain_H_e_dem_normalized_mean_start_pan.txt wc -l)  
pr -mts <(yes "pan" | head -n $lineN) <(awk ' {split( $0, a, " " ); asort( a ); for( i = 1; i <=  
length(a); i++ ) printf( "%s\t", a[i] ); printf( "\n" ); }' <(cut -d$'\t' -f 2,3  
bg/start_end/brain_H_e_dem_normalized_mean_start_end_pan.bg)) >  
bg/start_end/brain_H_e_dem_normalized_mean_start_end_pan.rowsort.bg  
sed -i 's/[[:space:]]*$// ' bg/start_end/brain_H_e_dem_normalized_mean_start_end_pan.rowsort.bg

lineN=$(  
bg/start_end/brain_H_cha1_normalized_mean_start_pan.txt wc -l)  
pr -mts <(yes "pan" | head -n $lineN) <(awk ' {split( $0, a, " " ); asort( a ); for( i = 1; i <=  
length(a); i++ ) printf( "%s\t", a[i] ); printf( "\n" ); }' <(cut -d$'\t' -f 2,3

```

```

bg/start_end/brain_H_cha1_normalized_mean_start_end_pan.bg)) >
bg/start_end/brain_H_cha1_normalized_mean_start_end_pan.rowsort.bg
sed -i 's/[[:space:]]*$/' bg/start_end/brain_H_cha1_normalized_mean_start_end_pan.rowsort.bg

lineN=$(  
pr -mts <(yes "pan" | head -n $lineN) <(awk ' {split( $0, a, " " ); asort( a ); for( i = 1; i <=  
length(a); i++ ) printf( "%s\t", a[i] ); printf( "\n" ); }' <(cut -d$'\t' -f 2,3  
bg/start_end/brain_H_cha2_normalized_mean_start_end_pan.bg)) >  
bg/start_end/brain_H_cha2_normalized_mean_start_end_pan.rowsort.bg  
sed -i 's/[[:space:]]*$/' bg/start_end/brain_H_cha2_normalized_mean_start_end_pan.rowsort.bg

lineN=$(  
pr -mts <(yes "pan" | head -n $lineN) <(awk ' {split( $0, a, " " ); asort( a ); for( i = 1; i <=  
length(a); i++ ) printf( "%s\t", a[i] ); printf( "\n" ); }' <(cut -d$'\t' -f 2,3  
bg/start_end/brain_H_m_ros_normalized_mean_start_end_pan.bg)) >  
bg/start_end/brain_H_m_ros_normalized_mean_start_end_pan.rowsort.bg  
sed -i 's/[[:space:]]*$/' bg/start_end/brain_H_m_ros_normalized_mean_start_end_pan.rowsort.bg

# Add coverage data of original .bedgraph files  
pr -mts bg/start_end/brain_H_e_dem_normalized_mean_start_end_pan.rowsort.bg <(cut -d$'\t' -f 4  
bg/brain_H_e_dem_normalized_mean.w30s0bin.bg) >  
bg/start_end/brain_H_e_dem_normalized_mean_start_end_pan.rowsort.cov.bg  
pr -mts bg/start_end/brain_H_cha1_normalized_mean_start_end_pan.rowsort.bg <(cut -d$'\t' -f 4  
bg/brain_H_cha1_normalized_mean.w30s0bin.bg) >  
bg/start_end/brain_H_cha1_normalized_mean_start_end_pan.rowsort.cov.bg  
pr -mts bg/start_end/brain_H_cha2_normalized_mean_start_end_pan.rowsort.bg <(cut -d$'\t' -f 4  
bg/brain_H_cha2_normalized_mean.w30s0bin.bg) >  
bg/start_end/brain_H_cha2_normalized_mean_start_end_pan.rowsort.cov.bg  
pr -mts bg/start_end/brain_H_m_ros_normalized_mean_start_end_pan.rowsort.bg <(cut -d$'\t' -f 4  
bg/brain_H_m_ros_normalized_mean.w30s0bin.bg) >  
bg/start_end/brain_H_m_ros_normalized_mean_start_end_pan.rowsort.cov.bg

# clean  
awk '{ if ($2 < $3 && $3-$2 < 10000 && $2!=$3) print $0}'  
bg/start_end/brain_H_e_dem_normalized_mean_start_end_pan.rowsort.cov.bg | sort -g -k2,2 >  
bg/start_end/brain_H_e_dem_normalized_mean_start_end_pan.rowsort.clean.bg  
awk '{ if ($2 < $3 && $3-$2 < 10000 && $2!=$3) print $0}'  
bg/start_end/brain_H_cha1_normalized_mean_start_end_pan.rowsort.cov.bg | sort -g -k2,2 >  
bg/start_end/brain_H_cha1_normalized_mean_start_end_pan.rowsort.clean.bg  
awk '{ if ($2 < $3 && $3-$2 < 10000 && $2!=$3) print $0}'  
bg/start_end/brain_H_cha2_normalized_mean_start_end_pan.rowsort.cov.bg | sort -g -k2,2 >  
bg/start_end/brain_H_cha2_normalized_mean_start_end_pan.rowsort.clean.bg  
awk '{ if ($2 < $3 && $3-$2 < 10000 && $2!=$3) print $0}'  
bg/start_end/brain_H_m_ros_normalized_mean_start_end_pan.rowsort.cov.bg | sort -g -k2,2 >  
bg/start_end/brain_H_m_ros_normalized_mean_start_end_pan.rowsort.clean.bg

grep -v '\.$' bg/start_end/brain_H_e_dem_normalized_mean_start_end_pan.rowsort.clean.bg | bedtools merge  
-c 4 -d -2 -o mean > bg/start_end/brain_H_e_dem_normalized_mean_start_end_pan.rowsort.clean.nodot.bg  
grep -v '\.$' bg/start_end/brain_H_cha1_normalized_mean_start_end_pan.rowsort.clean.bg | bedtools merge -  
c 4 -d -2 -o mean > bg/start_end/brain_H_cha1_normalized_mean_start_end_pan.rowsort.clean.nodot.bg  
grep -v '\.$' bg/start_end/brain_H_cha2_normalized_mean_start_end_pan.rowsort.clean.bg | bedtools merge -  
c 4 -d -2 -o mean > bg/start_end/brain_H_cha2_normalized_mean_start_end_pan.rowsort.clean.nodot.bg  
grep -v '\.$' bg/start_end/brain_H_m_ros_normalized_mean_start_end_pan.rowsort.clean.bg | bedtools merge  
-c 4 -d -2 -o mean > bg/start_end/brain_H_m_ros_normalized_mean_start_end_pan.rowsort.clean.nodot.bg

# Bedgraph to bigwig  
bedGraphToBigWig bg/start_end/brain_H_e_dem_normalized_mean_start_end_pan.rowsort.clean.nodot.bg  
pan_sizes.txt bg/start_end/brain_H_e_dem_normalized_mean_start_end_pan.rowsort.clean.bw  
bedGraphToBigWig bg/start_end/brain_H_cha1_normalized_mean_start_end_pan.rowsort.clean.nodot.bg  
pan_sizes.txt bg/start_end/brain_H_cha1_normalized_mean_start_end_pan.rowsort.clean.bw  
bedGraphToBigWig bg/start_end/brain_H_cha2_normalized_mean_start_end_pan.rowsort.clean.nodot.bg  
pan_sizes.txt bg/start_end/brain_H_cha2_normalized_mean_start_end_pan.rowsort.clean.bw  
bedGraphToBigWig bg/start_end/brain_H_m_ros_normalized_mean_start_end_pan.rowsort.clean.nodot.bg  
pan_sizes.txt bg/start_end/brain_H_m_ros_normalized_mean_start_end_pan.rowsort.clean.bw

```

**Supplemental Code S10. R code to plot any interval of the pan genome with ATAC-seq signals and lineage-specific features. The current example is for the gene *chiffon*.**

```
# Interval to plot
PANpos = 316442193 # Midpoint (i.e., position of interest)

startP = -30000 # Left from midpoint
endP = 30000 # Right from midpoint

# SV plotting parameters
minLength = 80
minLengthTE = 80
maxLengthTE = 1000000

# Colors and names
cols <- c('pink','pink','cornflowerblue','orange')
names <- c('H. e. cha1', 'H. e. cha2', 'H. e. dem', 'H. m. ros')

# Numbering of genomes in PAN genome
genomes <- c(2,3,1,4)

###
# Load ATAC data
###

library(rtracklayer)

range <- paste('pan:', PANpos + startP, '-', PANpos + endP, sep='')

BRAIN_dem <- import.bw("bw/brain_H_e_dem_normalized_mean_start_end_pan.rowsort.clean.bw", which =
GRanges(range))
BRAIN_cha1 <- import.bw("bw/brain_H_cha1_normalized_mean_start_end_pan.rowsort.clean.bw", which =
GRanges(range))
BRAIN_cha2 <- import.bw("bw/brain_H_cha2_normalized_mean_start_end_pan.rowsort.clean.bw", which =
GRanges(range))
BRAIN_ros <- import.bw("bw/brain_H_m_ros_normalized_mean_start_end_pan.rowsort.clean.bw", which =
GRanges(range))

bw_list <- list(BRAIN_cha1, BRAIN_cha2, BRAIN_dem, BRAIN_ros)

###
# Load Lineage-specific interval data
###

unique_dem <- read.table('blocks_unique_1.txt')
unique_cha1_poly <- read.table('blocks_unique_2and3.txt')
unique_cha2_poly <- read.table('blocks_unique_2and3.txt')
unique_melp <- read.table('blocks_unique_4.txt')

uniqueList <- list(unique_cha1_poly, unique_cha2_poly, unique_dem, unique_melp)

###
# Load Lineage-specific TE interval data and total TE set
###

TEdem <- read.table('uniqueTE/erato.uniqueTE.unfiltered.bed')
TEcha1 <- read.table('uniqueTE/cha1.uniqueTE.unfiltered.bed')
TEcha2 <- read.table('uniqueTE/cha2.uniqueTE.unfiltered.bed')
TEmelp <- read.table('uniqueTE/melp.uniqueTE.unfiltered.bed')

colnames(TEdem) <- c('pan','start','end')
colnames(TEcha1) <- c('pan','start','end')
colnames(TEcha2) <- c('pan','start','end')
colnames(TEmelp) <- c('pan','start','end')

TE_list <- list(TEcha1, TEcha2, TEdem, TEmelp)
```

```

TEs <- read.table("total.TE.bed")

###
# Output file and figure arrangement
###

pdf('PAN_chiffon.pdf', width=30, height=50)

layout(matrix(c(1:16), nrow=16, byrow=TRUE), height =
c(0.2,0.1,0.2,1,0.1,0.2,1,0.1,0.2,1,0.1,0.2,1,0.1,0.2,1))
# layout.show(n=25)

par(mar = c(0,4,0,4), oma=c(2,0,2,0))

###
# Plot gene annotations
###
plot(NULL, xlim=c(PANpos + startP, PANpos + endP), ylim = c(0,1), axes=FALSE, ann=FALSE)

genesPAN <- read.table('PAN_H_charithonia_genes_eratoTransfer.gff', h=F, sep= '\t')

genesPAN <- subset(genesPAN, (genesPAN$V5-genesPAN$V4 >= minLength) & (genesPAN$V4 > PANpos + startP) &
(genesPAN$V5 < PANpos + endP))

colnames(genesPAN) <- c("contig", "HGC", "type", "con_start", "con_end", "dot", "str", "unk", "descr")

for (g in 1:nrow(genesPAN)){
  if (genesPAN$type[g] == "gene" && genesPAN$str[g] == "-")
    rect(genesPAN$con_start[g], 0.75, genesPAN$con_end[g], 0.75, col = NULL, border = "black")
  if (genesPAN$type[g] == "exon" && genesPAN$str[g] == "-")
    rect(genesPAN$con_start[g], 0.5, genesPAN$con_end[g], 1, col = "black", border = "black", lwd=0.3)
  if (genesPAN$type[g] == "gene" && genesPAN$str[g] == "+")
    rect(genesPAN$con_start[g], 0.25, genesPAN$con_end[g], 0.25, col = NULL, border = "black")
  if (genesPAN$type[g] == "exon" && genesPAN$str[g] == "+")
    rect(genesPAN$con_start[g], 0, genesPAN$con_end[g], 0.5, col = "black", border = "black", lwd=0.3)
}

###
# Plot TEs
###

plot(NULL, xlim=c(PANpos + startP, PANpos + endP), ylim = c(0,1), axes=FALSE, ann=FALSE)

colnames(TEs) <- c('scaf', 'start', 'end')
TEs <- subset(TEs, (TEs$end-TEs$start >= minLengthTE) & (TEs$end-TEs$start <= maxLengthTE) & (TEs$start >
PANpos + startP) & (TEs$end < PANpos + endP))

for(i in 1:nrow(TEs)){
  rect(TEs$start[i], 0, TEs$end[i], 1, col = 'darkred', border = NA)
}

###
# Plot PAN genome with lineage-specific intervals
###
plot(NULL, xlim=c(PANpos + startP, PANpos + endP), ylim = c(0,1), axes=FALSE, ann=FALSE)

top = 1
bot = 0

rect(PANpos + startP, 0, PANpos + endP, 1, col = "black", border = NA)

for(e in 1:length(uniqueList)){
  unique <- uniqueList[[e]]
  colnames(unique) <- c('scaf', 'start', 'end')
  unique <- subset(unique, (unique$end-unique$start >= minLength) & (unique$start > PANpos + startP) &
(unique$end < PANpos + endP))

```

```

if(nrow(unique) > 0){
  for(i in 1:nrow(unique)){
    rect(unique$start[i], bot, unique$end[i], top, col = cols[e], border = NA)
  }
}

if(e == 1){
  unique <- read.table('blocks_unique_2_vs_3.txt')
  colnames(unique) <- c('scaf', 'start', 'end')
  unique <- subset(unique, (unique$end-unique$start >= minLength) & (unique$start > PANpos + startP) &
(unique$end < PANpos + endP))

  if(nrow(unique) > 0){
    for(i in 1:nrow(unique)){
      rect(unique$start[i], bot, unique$end[i], top, col = adjustcolor('pink', alpha = 1), border = NA)
    }
  }
}

if(e == 2){
  unique <- read.table('blocks_unique_3_vs_2.txt')
  colnames(unique) <- c('scaf', 'start', 'end')
  unique <- subset(unique, (unique$end-unique$start >= minLength) & (unique$start > PANpos + startP) &
(unique$end < PANpos + endP))

  if(nrow(unique) > 0){
    for(i in 1:nrow(unique)){
      rect(unique$start[i], bot, unique$end[i], top, col = adjustcolor('pink', alpha = 1), border = NA)
    }
  }
}

###
# Plot PAN genome with ATAC-seq data and TE features
###

bot2 = 0
top2 = c(600,600,200,200) # Adjust these values to change the y-axis range

for(e in 1:length(bw_list)){

  # Plot lineage-specific sequences
  unique <- uniqueList[[e]]
  colnames(unique) <- c('scaf', 'start', 'end')
  unique <- subset(unique, (unique$end - unique$start >= minLength) & (unique$start > PANpos + startP) &
(unique$end < PANpos + endP))

  plot(NULL, xlim=c(PANpos + startP,PANpos + endP), ylim = c(bot2,top2[e]), axes=FALSE, ann=FALSE)

  if(nrow(unique) > 0){
    for(i in 1:nrow(unique)){
      rect(unique$start[i], bot2, unique$end[i], top2[e], col = adjustcolor(cols[e], alpha = 0.5), border
= NA)
    }
  }

  if(e == 1){
    unique <- read.table('blocks_unique_2_vs_3.txt')
    colnames(unique) <- c('scaf', 'start', 'end')
    unique <- subset(unique, (unique$end-unique$start >= minLength) & (unique$start > PANpos + startP) &
(unique$end < PANpos + endP))

    if(nrow(unique) > 0){
      for(i in 1:nrow(unique)){
        rect(unique$start[i], bot2, unique$end[i], top2[e], col = adjustcolor('gray', alpha = 0.5), border
= NA)
      }
    }
  }
}

```

```

    }
  }
}

if(e == 2){
  unique <- read.table('blocks_unique_3_vs_2.txt')
  colnames(unique) <- c('scaf', 'start', 'end')
  unique <- subset(unique, (unique$end-unique$start >= minLength) & (unique$start > PANpos + startP) &
(unique$end < PANpos + endP))

  if(nrow(unique) > 0){
    for(i in 1:nrow(unique)){
      rect(unique$start[i], bot2, unique$end[i], top2[e], col = adjustcolor('gray', alpha = 0.5), border
= NA)
    }
  }
}

# Plot ATAC-seq data
par(new=T)
plot(0.5*(start(bw_list[[e]]) + end(bw_list[[e]])), bw_list[[e]]$score, type='l', xlim = c(PANpos +
startP, PANpos + endP), ylim = c(bot2,top2[e]), ylab = "", yaxt = "n", lwd = 5, xlab = "", xaxt = "n",
main = "", bty='none')

  mtext(names[e], side = 1, cex=5, padj = -2, las = 1, adj=0) # comment if you don't want the names to be
printed

  axis(2, at = seq(bot2,top2[e], by = 50), labels = NA, lwd = 3, lwd.ticks = 3, tck = -0.05)

# Plot TE features
plot(NULL, xlim=c(PANpos + startP, PANpos + endP), ylim = c(0,1), axes=FALSE, ann=FALSE)

TE_spec <- TE_list[[e]]

TE_spec <- subset(TE_spec, (TE_spec$end - TE_spec$start >= minLengthTE) & (TE_spec$end - TE_spec$start
<= maxLengthTE) & (TE_spec$start > PANpos + startP) & (TE_spec$end < PANpos + endP))

if(nrow(TE_spec) > 0){
  for(i in 1:nrow(TE_spec)){
    rect(TE_spec$start[i], 0, TE_spec$end[i], 1, col = 'darkred', border = NA)
  }
}

# Plot pan genome track
blocks <- read.table(paste(genomes[e], "_blocks_intervals.corr.bed", sep=""), h = F)
colnames(blocks) <- c('scaf', 'start', 'end')
blocks <- subset(blocks, (blocks$end-blocks$start >= 20) & (blocks$start > PANpos + startP) &
(blocks$end < PANpos + endP))

plot(NULL, xlim=c(PANpos + startP, PANpos + endP), ylim = c(bot,top), axes=FALSE, ann=FALSE)
rect(PANpos + startP, 0, PANpos + endP, 1, col = "grey90", border = NA)

if(nrow(blocks) > 0){
  for(i in 1:nrow(blocks)){
    rect(blocks$start[i], bot, blocks$end[i], top, col = 'black', border = NA)
  }
}

unique <- uniqueList[[e]]
colnames(unique) <- c('scaf', 'start', 'end')
unique <- subset(unique, (unique$end - unique$start >= minLength) & (unique$start > PANpos + startP) &
(unique$end < PANpos + endP))

if(nrow(unique) > 0){
  for(i in 1:nrow(unique)){
    rect(unique$start[i], bot, unique$end[i], top, col = cols[e], border = NA)
  }
}

```

```

if(e == 1){
  unique <- read.table('blocks_unique_2_vs_3.txt')
  colnames(unique) <- c('scaf', 'start', 'end')
  unique <- subset(unique, (unique$end - unique$start >= minLemgth) & (unique$start > PANpos + startP) &
(unique$end < PANpos + endP))

  if(nrow(unique) > 0){
    for(i in 1:nrow(unique)){
      rect(unique$start[i], bot, unique$end[i], top, col = cols[e], border = NA)
    }
  }
}

if(e == 2){
  unique <- read.table('blocks_unique_3_vs_2.txt')
  colnames(unique) <- c('scaf', 'start', 'end')
  unique <- subset(unique, (unique$end - unique$start >= minLemgth) & (unique$start > PANpos + startP) &
(unique$end < PANpos + endP))

  if(nrow(unique) > 0){
    for(i in 1:nrow(unique)){
      rect(unique$start[i], bot, unique$end[i], top, col = cols[e], border = NA)
    }
  }
}

# Plot x-axis
plot(NULL, xlim=c(startP,endP), ylim = c(0,1), axes=FALSE, ann=FALSE)
axis(1, at = seq(startP, endP, by=5000), labels = NA, line =-50, lwd = 3, lwd.ticks = 3, tck = -0.05)
axis(1, at = seq(startP, endP, by=10000), labels = seq(startP/1000000, endP/1000000, by=0.01), line =-40,
cex.axis = 6, lwd=0)

dev.off()

```