Supplementary Materials for Birth and Death of LTR-Retrotransposons in *Aegilops tauschii*

# Criteria for Excluding Artificial Predictions

Intact LTR elements that satisfied all three characteristics, as described in the main text, were then inspected for artificial predictions. We first grouped LTR candidates by their LTR region, then we performed multiple sequence alignment (MSA) at both 5’ and 3’ ends. The sequences used for MSA were made of 25 bp LTR ends and 25 bp flanking sequences. The inspection was to check each alignment to make sure that (1) alignment score was high in the TE ends but low in flanking region, and (2) TSD and TG-CA sequences appeared at proper positions. Sometimes we found one or two sequences that gave strange MSA patterns, and these sequences were excluded.

# Definition of Distributions

The following list contains the probability mass function (PMF) for discrete distributions and the probability density functions (PDF) used in our analysis.

## Discrete distributions

Poisson:

Negative binomial NB:

where is the gamma function

Geometric:

## Continuous distributions

Gamma distribution :

Exponential:

# Maximum likelihood estimation (MLE)

The parameter estimates for the aforementioned distributions were obtained through MLE. For distributions without closed-form MLEs, numerical optimization methods such as Nelder-Mead and quasi-Newton were used. To fit a mixture distribution to the mismatch data, we consider the family of negative binomial mixture density

where and are the PMFs of Negative binomial and Negative binomial, respectively, and is the probability that an element is from the first negative binomial distribution. We then used MLE to estimate parameters Our estimation functions EstDynamics and EstDynamics2 are included in the R package TE.

# Supplementary Tables

**Table S1**. Summary for intact elements in LTR-RT families of Ae. tauschii that have 50 intact copies in the genome. For LTR divergence and length, the mean values are shown with corresponding standard deviations in parentheses.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Family | Highest match to known family | LTR divergence | LTR length (bp) | Copy (no) |
| *Gypsy 1* | *Fatima* | 0.04 (0.012) | 470 (14.0) | 4144 |
| *Gypsy 2* | *Ifis* | 0.02 (0.008) | 517 (8.6) | 2296 |
| *Copia 3* | *WIS/Angela* | 0.02 (0.009) | 1725 (90.4) | 2081 |
| *Copia 4* | *Angela* | 0.02 (0.010) | 1720 (40.0) | 1059 |
| *Gypsy 5* | *Ifis* | 0.02 (0.010) | 558 (57.9) | 867 |
| *Gypsy 7* | *Carmilla* | 0.03 (0.011) | 526 (11.3) | 693 |
| *Gypsy 8* | *Lisa* | 0.03 (0.009) | 1190 (26.6) | 631 |
| *Copia 9* | *Angela* | 0.03 (0.012) | 1717 (84.8) | 521 |
| *Gypsy 12* | *Cereba* | 0.02 (0.014) | 913 (12.6) | 311 |
| *Gypsy 13* | *Wilma* | 0.03 (0.009) | 1509 (62.3) | 277 |
| *Gypsy 14* | *Nusif* | 0.04 (0.009) | 892 (15.9) | 262 |
| *Gypsy 15* | *Nusif* | 0.05 (0.010) | 896 (17.9) | 238 |
| *Copia 16* | *WIS/Angela* | 0.02 (0.014) | 139 (0.9) | 216 |
| *Gypsy 17* | *Lisa/Laura* | 0.02 (0.006) | 1177 (24.9) | 212 |
| *Gypsy 19* | *Danae* | 0.04 (0.009) | 1173 (55.1) | 197 |
| *Gypsy 21* | *Fatima* | 0.04 (0.011) | 471 (12.5) | 177 |
| *Gypsy 24* | *Nusif* | 0.05 (0.010) | 893 (19.7) | 125 |
| *Gypsy 26* | *Fatima* | 0.04 (0.013) | 468 (15.0) | 119 |
| *Copia 27* | *Maximus* | 0.03 (0.008) | 1502 (24.6) | 116 |
| *Copia 29* | *Angela* | 0.03 (0.008) | 1324 (16.0) | 109 |
| *Gypsy 30* | *Lisa/Laura* | 0.03 (0.011) | 1184 (24.3) | 106 |
| *Gypsy 31* | *Ifis/Laura* | 0.01 (0.005) | 4049 (121.6) | 100 |
| *Gypsy 33* | *Jeli* | 0.03 (0.010) | 546 (11.9) | 93 |
| *Gypsy 34* | *Sabrina* | 0.03 (0.011) | 1611 (248.9) | 91 |
| *Gypsy 35* | *Quinta* | 0.03 (0.012) | 915 (11.3) | 90 |
| *Gypsy 36* | *Cassandra* | 0.03 (0.025) | 280 (86.0) | 84 |
| *Copia 38* | *Barbara* | 0.03 (0.007) | 1611 (40.4) | 83 |
| *Copia 39* | *Leojyg* | 0.01 (0.010) | 140 (0.5) | 82 |
| *Gypsy 40* | *Danae* | 0.05 (0.010) | 1173 (20.5) | 76 |
| *Gypsy 41* | *Jeli* | 0.03 (0.008) | 576 (9.4) | 64 |
| *Gypsy 42* | *Ifis* | 0.02 (0.007) | 517 (8.9) | 54 |
| *Gypsy 43* | *Ifis* | 0.02 (0.006) | 517 (8.0) | 54 |
| *Gypsy 44* | *Lisa* | 0.02 (0.014) | 1162 (18.2) | 52 |
| *Copia 45* | *WIS/ Angela* | 0.01 (0.008) | 2142 (38.4) | 51 |
| *Gypsy 46* | *Wilma* | 0.03 (0.008) | 1501 (79.9) | 50 |
| *Total* |  | 0.03 (0.014) | 930 (592.5) | 15,781 |

**Table S2**. Summary for intact elements in LTR-RT families of A. lyrata. The LTR regions of intact A. lyrata elements were compared to an A. thaliana TE database ([Lamesch et al. 2011](#_ENREF_1)) using BLAST to find known families using the criteria described in [Wicker et al. (2007)](#_ENREF_5). For LTR divergence and length, the mean values are shown with corresponding standard deviations in parentheses.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Family | Highest match *A. thaliana* families | LTR divergence | LTR length (bp) | Copy (no) |
| *Gypsy 1* | *N/A* | 0.01 (0.006) | 514 (5.5) | 183 |
| *Copia 2* | *N/A* | 0.01 (0.017) | 132 (1.0) | 58 |
| *Copia 3* | *N/A* | 0.00 (0.008) | 250 (4.3) | 27 |
| *Unknown 4* | *N/A* | 0.01 (0.013) | 132 (0.3) | 12 |
| *Gypsy 5* | *ATGP2* | 0.01 (0.005) | 1220 (18.6) | 9 |
| *Gypsy 6* | *N/A* | 0.01 (0.007) | 433 (1.6) | 7 |
| *Gypsy 7* | *ATGP2* | 0.01 (0.005) | 1293 (7.6) | 7 |
| *Unknown 8* | *ATHILA* | 0.01 (0.008) | 1639 (97.4) | 7 |
| *Gypsy 9* | *N/A* | 0.01 (0.004) | 515 (1.3) | 6 |
| *Copia 10* | *N/A* | 0.02 (0.016) | 237 (8.1) | 5 |
| *Gypsy 11* | *N/A* | 0.00 (0.002) | 535 (22.6) | 4 |
| *Copia 12* | *N/A* | 0.00 (0.004) | 252 (0.6) | 4 |
| *Unknown 13* | *ATGP2N* | 0.00 (0.004) | 2613 (31.5) | 4 |
| *Unknown 14* | *ATHILA4D* | 0.01 (0.005) | 1423 (8.5) | 4 |
| *Gypsy 15* | *N/A* | 0.01 (0.004) | 516 (1.5) | 4 |
| *Copia 16* | *N/A* | 0.01 (0.004) | 305 (4.0) | 4 |
| *Gypsy 17* | *ATHILA2* | 0.01 (0.009) | 1911 (47.9) | 3 |
| *Gypsy 18* | *ATHILA2* | 0.01 (0.004) | 1805 (67.3) | 3 |
| *Gypsy 19* | *N/A* | 0.01 (0.001) | 507 (14.2) | 3 |
| *Copia 20* | *N/A* | 0.00 (0.002) | 251 (0.0) | 3 |
| *Gypsy 21* | *N/A* | 0.00 (0.004) | 516 (1.5) | 3 |
| *Copia 22* | *N/A* | 0.00 (0.005) | 228 (1.2) | 3 |
| *Unknown 23* | *ATHILA2* | 0.01 (0.006) | 1866 (26.3) | 3 |
| *Copia 24* | *N/A* | 0.01 (0.013) | 133 (0.0) | 3 |
| *Gypsy 25* | *ATHILA4A* | 0.00 (0.005) | 1354 (65.1) | 2 |
| *Unknown 26* | *N/A* | 0.01 (0.001) | 1848 (1153.3) | 2 |
| *Gypsy 27* | *ATHILA4A/ATHILA4D* | 0.00 (0.001) | 1336 (38.9) | 2 |
| *Unknown 28* | *ATGP2N* | 0.00 (0.001) | 2528 (31.8) | 2 |
| *Unknown 29* | *ATHILA2* | 0.01 (0.001) | 1835 (25.5) | 2 |
| *Gypsy 30* | *N/A* | 0.01 (0.008) | 432 (0.7) | 2 |
| *Copia 31* | *N/A* | 0.02 (0.006) | 133 (0.0) | 2 |
| *Unknown 32* | *N/A* | 0.03 (0.001) | 792 (0.0) | 2 |
| *Gypsy 33* | *N/A* | 0.01 (0.001) | 432 (0.7) | 2 |
| *Unknown 34* | *ATGP7* | 0.01 (0.001) | 915 (14.1) | 2 |
| *Copia 35* | *N/A* | 0.02 (0.008) | 238 (14.8) | 2 |
| *Gypsy 36* | *N/A* | 0.01 (0.003) | 432 (0.7) | 2 |
| *Copia 37* | *N/A* | 0.04 (0.006) | 127 (0.0) | 2 |
| *Gypsy 38* | *N/A* | 0.01 (0.006) | 517 (0.0) | 2 |
| *Total* |  | 0.01 (0.010) | 552 (473.5) | 397 |

**Table S3**. A summary of distribution parameters, death rate as estimated from solo-LTRs, p-value for constant insertion rate, and lack-of-fit for the 9 Copia families in Ae. tauschii with elements each.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Negativebinomial | Gammafit | Geometric | Survivaltime | *p*-valuefor  | Lack-of-fit |
| Family |  |  |  |  |  |  |  | Half-life My |  |  |
| *Copia 3 (WIS/Angela)* | 6.09 | 0.15 | 6.09 | 8.15 | 0.03 |  | 0.91 |  |  | 0.014 |
| *Copia 4 (Angela)*  | 6.87 | 0.15 | 6.87 | 7.75 | 0.02 |  | 0.78 |  |  | 0.003 |
| *Copia 9 (Angela)* | 5.39 | 0.11 | 5.39 | 5.59 | 0.02 |  | 0.71 |  |  | 0.011 |
| *Copia 16 (WIS/Angela)* | 4.67 | 0.64 | 4.67 | 6.39 | 0.27 |  | 0.01 |  |  | 0.006 |
| *Copia 27 (Maximus)*  | 11.51 | 0.23 | 11.51 | 11.35 | 0.02 |  | 0.93 |  |  | 0.040 |
| *Copia 29 (Angela)*  | 15.32 | 0.28 | 15.32 | 13.56 | 0.03 |  | 0.52 |  |  | 0.021 |
| *Copia 38 (Barbara)* | 23.20 | 0.34 | 23.20 | 21.57 | 0.02 |  | 0.72 |  |  | 0.032 |
| *Copia 39 (Leojyg)* | 6.76 | 0.83 | 6.76 | 17.41 | 0.41 |  | 0.06 |  |  | 0.022 |
| *Copia 45 (WIS/ Angela)* | 4.52 | 0.13 | 4.52 | 8.19 | 0.03 |  | 1.03 |  |  | 0.027 |

**Table S4**. A summary of distribution parameters, death rate as estimated from solo-LTRs, p-value for constant insertion rate, and lack-of-fit for the 26 Gypsy families in Ae. tauschii with elements each.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Negativebinomial | Gammafit | Geometric | Survivaltime | *p*-valuefor  | Lack-of-fit |
| Family |  |  |  |  |  |  |  | Half-life My |  |
| *Gypsy 1 (Fatima)*  | 18.21 | 0.52 | 18.21 | 13.44 | 0.06 |  |  |  |  | 0.001 |
| *Gypsy 2 (Ifis)*  | 9.02 | 0.51 | 9.02 | 13.91 | 0.10 |  |  |  |  | 0.002 |
| *Gypsy 5 (Ifis)*  | 3.65 | 0.28 | 3.65 | 5.72 | 0.10 |  |  |  |  | 0.004 |
| *Gypsy 7 (Carmilla)*  | 23.49 | 0.57 | 23.49 | 18.13 | 0.05 |  |  |  |  | 0.008 |
| *Gypsy 8 (Lisa)*  | 11.31 | 0.27 | 11.31 | 11.59 | 0.03 |  |  |  |  | 0.003 |
| *Gypsy 12 (Cereba)*  | 3.53 | 0.14 | 3.53 | 3.94 | 0.04 |  |  |  |  | 0.011 |
| *Gypsy 13 (Wilma)*  | 13.40 | 0.23 | 13.40 | 11.84 | 0.02 |  |  |  |  | 0.013 |
| *Gypsy 14 (Nusif)*  | 52.76 | 0.59 | 52.76 | 33.43 | 0.03 |  |  |  |  | 0.014 |
| *Gypsy 15 (Nusif)*  | 48.92 | 0.53 | 48.92 | 26.38 | 0.02 |  |  |  |  | 0.010 |
| *Gypsy 17 (Lisa/Laura)* | 10.46 | 0.36 | 10.46 | 17.59 | 0.05 |  |  |  |  | 0.016 |
| *Gypsy 19 (Danae)*  | 33.11 | 0.40 | 33.11 | 20.4 | 0.02 |  |  |  |  | 0.016 |
| *Gypsy 21 (Fatima)*  | 23.78 | 0.58 | 23.78 | 16.96 | 0.06 |  |  |  |  | 0.010 |
| *Gypsy 24 (Nusif)*  | 45.98 | 0.53 | 45.98 | 25.77 | 0.02 |  |  |  |  | 0.025 |
| *Gypsy 26 (Fatima)*  | 16.52 | 0.49 | 16.52 | 11.6 | 0.05 |  |  |  |  | 0.019 |
| *Gypsy 30 (Lisa/Laura)* | 8.39 | 0.19 | 8.39 | 7.37 | 0.03 |  |  |  |  | 0.017 |
| *Gypsy 31 (Ifis/Laura)* | 11.73 | 0.16 | 11.73 | 20.69 | 0.02 |  |  |  |  | 0.019 |
| *Gypsy 33 (Jeli)* | 24.25 | 0.59 | 24.25 | 20.35 | 0.06 |  |  |  |  | 0.006 |
| *Gypsy 34 (Sabrina)* | 7.75 | 0.14 | 7.75 | 6.98 | 0.02 |  |  |  |  | 0.023 |
| *Gypsy 35 (Quinta)* | 8.87 | 0.23 | 8.87 | 6.89 | 0.03 |  |  |  |  | 0.046 |
| *Gypsy 36 (Cassandra)* | 2.05 | 0.18 | 2.05 | 1.62 | 0.10 |  |  |  |  | 0.033 |
| *Gypsy 40 (Danae)* | 49.49 | 0.46 | 49.49 | 25.69 | 0.02 |  |  |  |  | 0.036 |
| *Gypsy 41 (Jeli)* | 31.64 | 0.67 | 31.64 | 30.59 | 0.06 |  |  |  |  | 0.023 |
| *Gypsy 42 (Ifis)* | 38.13 | 0.80 | 38.13 | 53.21 | 0.09 |  |  |  |  | 0.018 |
| *Gypsy 43 (Ifis)* | 38.36 | 0.81 | 38.36 | 58.6 | 0.10 |  |  |  |  | 0.019 |
| *Gypsy 44 (Lisa)* | 2.67 | 0.11 | 2.67 | 3.73 | 0.04 |  |  |  |   | 0.589 |
| *Gypsy 46 (Wilma)* | 18.94 | 0.30 | 18.94 | 17.08 | 0.02 |  |  |  |  | 0.020 |

**Table S5**. A summary of the parameter estimates for the age distributions of LTR-RT families in *Ae. tauschii,* implementing the methods of [Promislow *et al.* (1999)](#_ENREF_4) and [Marchani *et al.* (2009)](#_ENREF_3). : Insertion rate; : Excision rate; : Population size expansion rate; : Ratio of elements in two adjacent classes i and i-1; age1: Age of the system assuming model 1 ( > 1); age2: Age of the system assuming model 2 (burst of insertions followed by a stasis lambda=1); ad hoc: Master gene age estimated by 2 times the average element age; MLE: Master gene age estimated by MLE; Average age (in myrs): mean divergence / .

|  |  |  |  |
| --- | --- | --- | --- |
|  | ([Promislow *et al.* 1999](#_ENREF_4)) | ([Marchani *et al.* 2009](#_ENREF_3)) |  |
| Family |  |  |  |  | age1 | age2 | ad hoc | MLE | Average age |
| *Gypsy 1 (Fatima)*  | 0.74 | 0.07 | 1.67 | 0.94 | 16.31 | 0.15 | 2.71 | 2.17 | 1.35 |
| *Gypsy 2 (Ifis)*  | 1.54 | 0.09 | 2.45 | 0.90 | 8.62 | 0.04 | 1.29 | 1.13 | 0.65 |
| *Copia 3 (WIS/Angela)* | 1.34 | 0.72 | 1.62 | 0.97 | 15.83 | 0.87 | 1.49 | 1.45 | 0.75 |
| *Copia 4 (Angela)*  | 1.13 | 0.61 | 1.52 | 0.98 | 16.67 | 1.05 | 1.77 | 1.60 | 0.88 |
| *Gypsy 5 (Ifis)*  | 1.57 | 0.19 | 2.38 | 0.90 | 7.79 | 0.09 | 1.27 | 1.26 | 0.64 |
| *Gypsy 7 (Carmilla)*  | 0.77 | 0.11 | 1.66 | 0.95 | 12.89 | 0.22 | 2.59 | 2.04 | 1.30 |
| *Gypsy 8 (Lisa)*  | 1.03 | 0.30 | 1.73 | 0.97 | 11.73 | 0.39 | 1.95 | 1.59 | 0.97 |
| *Copia 9 (Angela)* | 1.04 | 0.65 | 1.39 | 0.98 | 18.94 | 1.60 | 1.92 | 1.97 | 0.96 |
| *Gypsy 12 (Cereba)*  | 1.06 | 0.55 | 1.51 | 0.96 | 13.21 | 1.00 | 1.89 | 1.97 | 0.94 |
| *Gypsy 13 (Wilma)*  | 0.88 | 0.59 | 1.29 | 0.98 | 22.11 | 2.32 | 2.26 | 1.79 | 1.13 |
| *Gypsy 14 (Nusif)*  | 0.63 | 0.22 | 1.41 | 0.97 | 16.25 | 0.87 | 3.16 | 2.26 | 1.58 |
| *Gypsy 15 (Nusif)*  | 0.54 | 0.28 | 1.26 | 0.98 | 23.38 | 1.94 | 3.71 | 2.63 | 1.85 |
| *Copia 16 (WIS/Angela)* | 1.37 | 0.01 | 2.35 | 0.73 | 6.28 | 0.01 | 1.46 | 1.41 | 0.73 |
| *Gypsy 17 (Lisa/Laura)* | 1.68 | 0.50 | 2.18 | 0.95 | 6.88 | 0.25 | 1.19 | 1.00 | 0.60 |
| *Gypsy 19 (Danae)*  | 0.62 | 0.33 | 1.28 | 0.98 | 21.25 | 1.92 | 3.25 | 2.32 | 1.62 |
| *Gypsy 21 (Fatima)*  | 0.71 | 0.06 | 1.65 | 0.94 | 10.31 | 0.14 | 2.81 | 2.20 | 1.40 |
| *Gypsy 24 (Nusif)*  | 0.56 | 0.23 | 1.33 | 0.98 | 16.98 | 1.26 | 3.57 | 2.55 | 1.78 |
| *Gypsy 26 (Fatima)*  | 0.70 | 0.03 | 1.67 | 0.95 | 9.27 | 0.06 | 2.85 | 2.32 | 1.42 |
| *Copia 27 (Maximus)*  | 0.99 | 0.66 | 1.33 | 0.98 | 16.83 | 2.06 | 2.02 | 1.57 | 1.01 |
| *Copia 29 (Angela)*  | 0.89 | 0.46 | 1.42 | 0.97 | 13.29 | 1.23 | 2.26 | 1.72 | 1.13 |
| *Gypsy 30 (Lisa/Laura)* | 0.88 | 0.31 | 1.57 | 0.97 | 10.33 | 0.61 | 2.28 | 1.91 | 1.14 |
| *Gypsy 31 (Ifis/Laura)* | 1.76 | 1.52 | 1.25 | 0.98 | 20.73 | 3.45 | 1.13 | 0.90 | 0.57 |
| *Gypsy 33 (Jeli)* | 0.84 | 0.08 | 1.76 | 0.94 | 8.01 | 0.13 | 2.38 | 1.87 | 1.19 |
| *Gypsy 34 (Sabrina)* | 0.90 | 0.79 | 1.11 | 0.98 | 42.83 | 7.89 | 2.22 | 1.92 | 1.11 |
| *Gypsy 35 (Quinta)* | 0.76 | 0.37 | 1.39 | 0.97 | 13.56 | 1.27 | 2.63 | 2.13 | 1.31 |
| *Gypsy 36 (Cassandra)* | 0.79 | 0.35 | 1.43 | 0.90 | 12.29 | 1.04 | 2.54 | 2.64 | 1.27 |
| *Copia 38 (Barbara)* | 0.93 | 0.61 | 1.32 | 0.98 | 15.97 | 2.06 | 2.15 | 1.57 | 1.08 |
| *Copia 39 (Leojyg)* | 2.61 | 0.06 | 3.55 | 0.58 | 3.47 | 0.01 | 0.77 | 0.75 | 0.38 |
| *Gypsy 40 (Danae)* | 0.52 | 0.34 | 1.18 | 0.98 | 26.14 | 3.62 | 3.85 | 2.72 | 1.93 |
| *Gypsy 41 (Jeli)* | 0.97 | 0.18 | 1.79 | 0.94 | 7.10 | 0.23 | 2.07 | 1.61 | 1.03 |
| *Gypsy 42 (Ifis)* | 1.40 | 0.10 | 2.30 | 0.91 | 4.79 | 0.05 | 1.43 | 1.17 | 0.72 |
| *Gypsy 43 (Ifis)* | 1.53 | 0.08 | 2.45 | 0.90 | 4.44 | 0.04 | 1.31 | 1.07 | 0.65 |
| *Gypsy 44 (Lisa)* | 1.40 | 0.44 | 1.96 | 0.96 | 5.89 | 0.33 | 1.43 | 2.00 | 0.72 |
| *Copia 45 (WIS/ Angela)* | 1.81 | 0.82 | 1.99 | 0.97 | 5.70 | 0.45 | 1.10 | 1.15 | 0.55 |
| *Gypsy 46 (Wilma)* | 0.90 | 0.51 | 1.39 | 0.98 | 11.83 | 1.44 | 2.22 | 1.67 | 1.11 |

**Table S6.** Regression coefficient estimates for the age of an LTR-RT as response and various genomic factors as predictors, namely chromosome membership (Chr2—7, with Chr1 as baseline), adjacent HR rate, proximity to the 5’ end of a gene (with proximity to the 3’ end as baseline), colinearity of the closest gene with the homologous gene on a homoeologous pseudomolecule in Brachypodium distachyon ([Luo et al. 2017](#_ENREF_2)) (with baseline that the gene is next to a non-colinear gene), log distance (bp) to the nearest gene, and LTR-RT superfamily membership (with Copia as baseline). The regression coefficients for the various predictors correspond to: For Chr2, the difference between the mean age on Chr2 and that on Chr1; analogously for Chr3—7; for Near 5’ end, the difference between the mean age of LTR-RTs near the 5’ end of a gene and those near the 3’ end; for Near colinear gene, the difference between the mean age of LTR-RTs next to colinear genes and those next to noncolinear genes; for Gypsy superfamily, the difference in mean age between Gypsy and Copia elements; for Intercept: the mean age in the reference level, consisting of Copia elements on Chr1 closest to a noncolinear gene and its 3’ end.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Predictor  | Regression coefficient | Std. error | t-value | p-value |
| Intercept | 0.224 | 0.043 | 5.176 | 0.000 |
| Chr2 | 0.025 | 0.011 | 2.258 | 0.024 |
| Chr3 | 0.011 | 0.011 | 0.981 | 0.327 |
| Chr4 | 0.026 | 0.012 | 2.246 | 0.025 |
| Chr5 | 0.013 | 0.011 | 1.135 | 0.256 |
| Chr6 | 0.025 | 0.012 | 2.079 | 0.038 |
| Chr7 | 0.043 | 0.011 | 3.800 | 0.000 |
| HR  | -0.025 | 0.007 | -3.677 | 0.000 |
| Near 5’ end | -0.014 | 0.006 | -2.274 | 0.023 |
| Near colinear gene | -0.039 | 0.006 | -6.498 | 0.000 |
| Log distance to a gene | 0.072 | 0.003 | 26.139 | 0.000 |
| Gypsy superfamily | 0.095 | 0.045 | 2.089 | 0.037 |

# Supplementary Figure Captions

**Figure S1.** The abundance of solo-LTRs vs intact LTR-RTs in *Ae. tauschii*. The number in the top right corner of each panel reports the ratio of solo to intact elements.

**Figure S2.** Regressions of the proportion of solo-LTRs on the logarithm-transformed mean LTR length for LTR-RT families in *Ae. tauschii* with intact elements. Each point corresponds to an LTR-RT family, where the y-value is formed by the number of solo-LTRs divided by the total number of intact and solo elements. LTR-RT families with longer mean LTR length tended to have larger proportions of solo-LTRs. The increase in the proportion of solo elements as the mean LTR length increased was larger for *Gypsy* families than for *Copia* families.

**Figure S3.** (A) The age distributions and (B) normalized insertion rates of the *Copia 3*, *Gypsy 31*, and *Gypsy 40* families in *Ae. tauschii*. For the age distribution of each of these families, the first peak corresponds to complete elements and the second peak to solo elements.

**Figure S4.** A comparison between the age distributions of (A) *Gypsy 1* and (B) *Copia 4* in *Ae. tauschii* estimated by the histogram, our proposed negative binomial fit , a matrix population model ([Promislow *et al.* 1999](#_ENREF_4)), and a master gene model ([Marchani *et al.* 2009](#_ENREF_3)). The matrix population model and the master gene model produced an exponential and a uniform fit, respectively, which did not reflect the peaks in the age distributions.

**Figure S5.** Regression of age residuals on the log distance to the nearest gene (bp). Overall, the distance to the nearest gene had a positive association with the mean LTR-RT age (blue curve), showing a plateau when the distance to the nearest gene is smaller than 22 kb (or log distance = 10, red dot) and a linear increase with an increase in the log distance after that. The construction of the nonlinear regression is detailed in Methods.

**Figure S6**. Distributional fits and recovered age distribution of (A) *Gypsy* family 1 and (B) *Copia* family 2 in *A. lyrata.* The recovered age distribution for *Gypsy* 1 had a peak that corresponds to 0.125 mya, while that for *Copia* 2 peaked at the present time. The constant insertion rate hypothesis was rejected for *Gypsy* 1 (), but not rejected for *Copia* family 2 ().

**Figure S7.** (A) Age distributions and (B) normalized insertion rates in the two largest LTR-RT families of *A. lyrata*. A recent burst of insertion activity was found for *Gypsy* 1, while the amplification activity of *Copia* 2 has been overall steady for the past 3 myr, showing different insertion dynamics from those in *Ae. tauschii.*

**References**

Lamesch, P., T. Z. Berardini, D. Li, D. Swarbreck, C. Wilks *et al.*, 2011 The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. Nucleic acids research 40**:** D1202-D1210.

Luo, M. C., Y. Q. Gu, D. Puiu, H. Wang, S. O. Twardziok *et al.*, 2017 Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. Nature 551**:** 498-502.

Marchani, E. E., J. Xing, D. J. Witherspoon, L. B. Jorde and A. R. Rogers, 2009 Estimating the age of retrotransposon subfamilies using maximum likelihood. Genomics 94**:** 78-82.

Promislow, D. E. L., I. K. Jordan and J. E. McDonald, 1999 Genomic demography: a life-history analysis of transposable element evolution. Proceedings of the Royal Society B: Biological Sciences 266**:** 1555-1560.

Wicker, T., F. Sabot, A. Hua-Van, J. L. Bennetzen, P. Capy *et al.*, 2007 A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics 8**:** 973-982.