

Supplemental Material S1

4-Lineage Insertion Likelihood test

1. Mathematical model

Considering neutral insertions of retroelements in generation t (t = time measured in generations), we take into account ten different events for four lineages: $\omega_{i,j}$ ($1 \leq i \leq j \leq 4$):

$\omega_{i,i}$ – a retroelement in the orthologous locus is absent in lineage A_i , and present in the other three lineages;

$\omega_{i,j}$ – a retroelement in the orthologous locus is absent in lineage A_i and A_j ($i \neq j$), and present in the other two lineages.

We denote $p_{i,j}(t)$ as the probability of these events; $\mu_{i,j}(t)$ describes the numbers of events and $\nu(t)$ the number of all insertions of retroelements in generation t .

Taking into account that the probability of new insertions for each individual of a population ($\alpha(t)$) is small, we assume that $\nu(t)$ is a Poisson distributed random variable with a mean $n_0 = n_0(t)$ proportional to the effective population size at time t , with $N_e(t)$: $n_0 = N_e(t) \cdot \alpha(t)$.

Then, considering independent fixation probabilities for each retrotransposon insertion, we can denote that $\mu_{i,j}(t)$ is a Poisson distributed random variable with parameters: $b_{i,j} = b_{i,j}(t) = n_0(t)p_{i,j}(t)$.

Now we consider all possible generations with potential retrotransposon insertions that later become phylogenetically informative. The set of corresponding values of time t we have denoted by T . Then the total number of retrotransposon insertions with properties $\omega_{i,j}$ equal to the sums:

$$\xi_{i,j} = \sum_{t \in T} \mu_{i,j}(t),$$

are independent random Poisson distributed variables with parameters:

$$a_{i,j} = \sum_{t \in T} b_{i,j}(t) = \sum_{t \in T} n_0(t)p_{i,j}(t). \quad (1)$$

We first distinguish two simple speciation variants (tree topologies) excluding hybridization.

If we consider t_0 as a first branching point (common for all following tree topologies), then in Tree topology 1 the two generated branches split at time points $t_1 = t_0 + T_1$ and $t_2 = t_0 + T_2$, respectively, forming together the four branches $A_1 - A_4$ (Figure 1). Tree topology 2 represents successive divisions of the ancestral branch at time points $t_1 = t_0 + T_1$ and $t_2 = t_0 + T_1 + T_2$, forming four branches $A_1 - A_4$ (Figure 2).

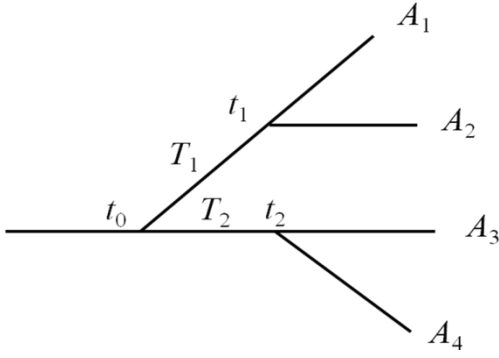


Fig. 1. Tree topology 1

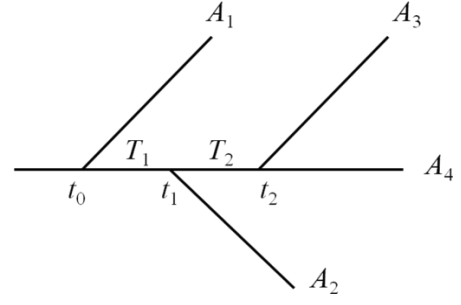


Fig. 2. Tree topology 2

For these topologies we denote the special cases of partial and full polytomies (Figures 3-5).

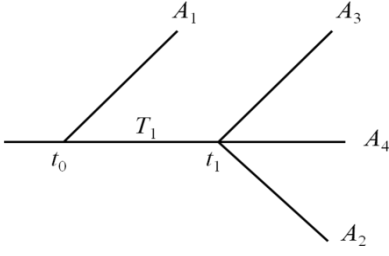


Fig. 3. Partial polytomy 1

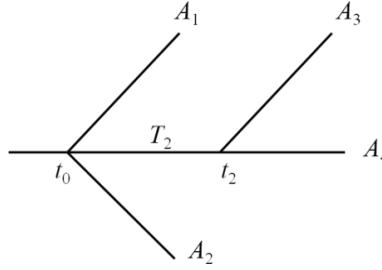


Fig. 4. Partial polytomy 2

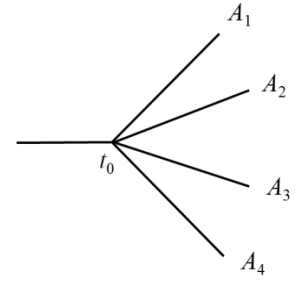


Fig. 5. Full polytomy

Now we introduce three models of tree topologies involving hybridization events (Fig. 6-8).

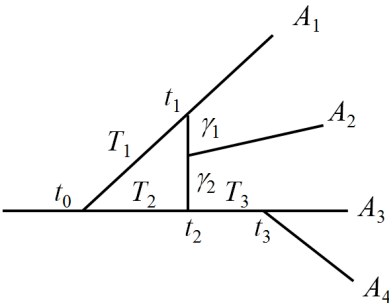


Fig. 6. Hybridization model 1

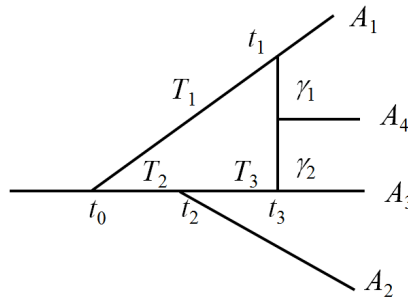


Fig. 7. Hybridization model 2

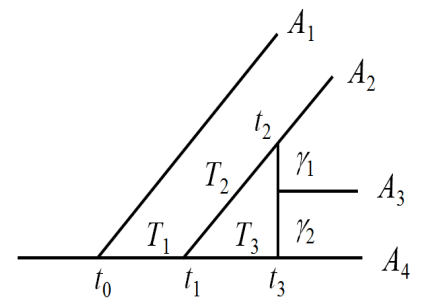


Fig. 8. Hybridization model 3

The proportions of the two fused sub-populations forming a new population are denoted by γ_1 and γ_2 ($\gamma_1 + \gamma_2 = 1$).

It should be noted, that for each hybridization-affected tree topology we can have 24 rearrangements of the four species A_1, A_2, A_3, A_4 . However, for the tree topology in Figure 6, due to symmetry of branches A_3 and A_4 against time point t_3 , the number of different rearrangements is reduced to 12. The same holds for the symmetry between species A_2 and A_4 against the time point t_1 in Figure 8, with an additional exchange of (γ_1, γ_2) and (T_2, T_3) .

In the case of a tree topology including hybridization, we denote the frequency of some neutral insertions of retroelements at time t as $X(t)$. Then, using the standard Wright-Fisher coalescent model (Fisher 1922, Wright 1931) we can consider $X(t)$ as a Markov process with the transition function $u(s, p, t, x)$, reflecting the conditional probability density of $X(t)$ for the condition $X(s) = p$. Under diffusion approximation, this transition function follows the Kolmogorov's equation:

$$\frac{\partial u}{\partial t} = \frac{1}{4N_e(t)} \frac{\partial^2}{\partial x^2} [x(1-x)u],$$

with the initial condition $u(s, p, s, x) = \delta(x - p)$, where $\delta(x - p)$ is a Dirac delta function, and $N_e(t)$ denotes the effective population size at time t (Kimura 1955a).

Let us introduce a neutral mutation with a frequency in a finite population p at time $t=s$. Then, as it follows from the neutral theory of molecular evolution (Kimura 1955a, Kimura 1955b), the conditional probability that this mutation will be fixed in the population at time $t \rightarrow \infty$ tends to p (whereby the probability of losing the mutation tends to $1-p$).

For retrotransposon insertions that take place at $t < t_0$, their frequencies in the respective lineages X_0, X_1, X_2, X_3 follow the probability densities of the random vector $X = (X_0, X_1, X_2, X_3)$. Then, introducing the vector $x = (x_0, x_1, x_2, x_3)$, where x_i is a random Poisson distributed variable we can write:

$$f(x_0, x_1, x_2, x_3) = u_0(t, p, t_0, x_0) \cdot u_1(t_0, x_0, t_1, x_1) \cdot u_2(t_2, x_2, t_3, x_3) \cdot u_3(x_0, x_1, x_2, x_3), \quad (2)$$

where $p = (2N_e(t))^{-1}$, and the index of the function u indicates the number of the tree branches.

Using the full probability formula, we obtain:

$$p_{i,j}(t) = P(\omega_{i,j}) = \int_0^1 \int_0^1 \int_0^1 \int_0^1 P(\omega_{i,j} | X = x) f(x) dx_0 dx_1 dx_2 dx_3. \quad (3)$$

By then replacing the summation by the integration in (1) we obtain:

$$a_{i,j}^0 = \int_{-\infty}^{t_0} p_{i,j}(t) dt. \quad (4)$$

Taking into account Equation 3 and all possible tree topology rearrangements exposed to Hybridization model 1 (see Figure 6), the conditional probabilities are:

$$\begin{aligned}
P(\omega_{1,1} | X = x) &= (1 - x_1)(\gamma_1 x_1 + \gamma_2 x_2) x_3^2 \\
P(\omega_{1,2} | X = x) &= (1 - x_1)(1 - \gamma_1 x_1 - \gamma_2 x_2) x_3^2 \\
P(\omega_{1,3} | X = x) &= P(\omega_{1,4} | X = x) = (1 - x_1)(\gamma_1 x_1 + \gamma_2 x_2)(1 - x_3) x_3 \\
P(\omega_{2,2} | X = x) &= x_1(1 - \gamma_1 x_1 - \gamma_2 x_2) x_3^2 \\
P(\omega_{2,3} | X = x) &= P(\omega_{2,4} | X = x) = x_1(1 - \gamma_1 x_1 - \gamma_2 x_2)(1 - x_3) x_3 \\
P(\omega_{3,3} | X = x) &= P(\omega_{4,4} | X = x) = x_1(\gamma_1 x_1 + \gamma_2 x_2)(1 - x_3) x_3 \\
P(\omega_{3,4} | X = x) &= x_1(\gamma_1 x_1 + \gamma_2 x_2)(1 - x_3)^2
\end{aligned}$$

For Hybridization model 2 (see Figure 7) we obtain the conditional probabilities:

$$\begin{aligned}
P(\omega_{1,1} | X = x) &= (1 - x_1) x_2 x_3 (x_1 \gamma_1 + x_3 \gamma_2) \\
P(\omega_{1,2} | X = x) &= (1 - x_1)(1 - x_2) x_3 (x_1 \gamma_1 + x_3 \gamma_2) \\
P(\omega_{1,3} | X = x) &= (1 - x_1) x_2 (1 - x_3) (x_1 \gamma_1 + x_3 \gamma_2) \\
P(\omega_{1,4} | X = x) &= (1 - x_1) x_2 x_3 (1 - x_1 \gamma_1 - x_3 \gamma_2) \\
P(\omega_{2,2} | X = x) &= x_1 (1 - x_2) x_3 (x_1 \gamma_1 + x_3 \gamma_2) \\
P(\omega_{2,3} | X = x) &= x_1 (1 - x_2)(1 - x_3) (x_1 \gamma_1 + x_3 \gamma_2) \\
P(\omega_{2,4} | X = x) &= x_1 (1 - x_2) x_3 (1 - x_1 \gamma_1 - x_3 \gamma_2) \\
P(\omega_{3,3} | X = x) &= x_1 x_2 (1 - x_3) (x_1 \gamma_1 + x_3 \gamma_2) \\
P(\omega_{3,4} | X = x) &= x_1 x_2 (1 - x_3) (1 - x_1 \gamma_1 - x_3 \gamma_2) \\
P(\omega_{4,4} | X = x) &= x_1 x_2 x_3 (1 - x_1 \gamma_1 - x_3 \gamma_2)
\end{aligned}$$

For Hybridization model 3 (see Figure 8) the conditional probabilities are:

$$\begin{aligned}
P(\omega_{1,1} | X = x) &= (1 - x_0) x_2 x_3 (x_2 \gamma_1 + x_3 \gamma_2) \\
P(\omega_{1,2} | X = x) &= (1 - x_0)(1 - x_2) x_3 (x_2 \gamma_1 + x_3 \gamma_2) \\
P(\omega_{1,3} | X = x) &= (1 - x_0) x_2 x_3 (1 - x_2 \gamma_1 - x_3 \gamma_2) \\
P(\omega_{1,4} | X = x) &= (1 - x_0) x_2 (1 - x_3) (x_2 \gamma_1 + x_3 \gamma_2) \\
P(\omega_{2,2} | X = x) &= x_0 (1 - x_2) x_3 (x_2 \gamma_1 + x_3 \gamma_2) \\
P(\omega_{4,4} | X = x) &= x_0 x_2 (1 - x_3) (x_2 \gamma_1 + x_3 \gamma_2) \\
P(\omega_{2,3} | X = x) &= x_0 (1 - x_2) x_3 (1 - x_2 \gamma_1 - x_3 \gamma_2) \\
P(\omega_{2,4} | X = x) &= x_0 (1 - x_2)(1 - x_3) (x_2 \gamma_1 + x_3 \gamma_2) \\
P(\omega_{3,3} | X = x) &= x_0 x_2 x_3 (1 - x_2 \gamma_1 - x_3 \gamma_2) \\
P(\omega_{3,4} | X = x) &= x_0 x_2 (1 - x_3).
\end{aligned}$$

Then, following Kuritzin et al. (2016), taking into account that $N_e(t) \gg 1$ (this assumption provides the basis for the use of a diffusion model), and ignoring the terms with order p^2 and higher, by expanding Equation 4 we obtain the parameters for Hybridization model 1:

$$\begin{aligned}
a_{1,1}^0 &= n_0 \left(\frac{e^{-\tau_1}}{3} - \frac{1}{6} e^{-\tau_1 - \tau_2 - \tau_3} \right) \gamma_1 + n_0 \left(1 - \frac{2e^{-\tau_2}}{3} + \frac{1}{6} e^{-3\tau_2 - \tau_3} - \frac{1}{3} e^{-\tau_2 - \tau_3} \right) \gamma_2 \\
a_{1,2}^0 &= n_0 \left(1 - \frac{e^{-\tau_1}}{3} - \frac{2}{3} e^{-\tau_2 - \tau_3} + \frac{1}{6} e^{-\tau_1 - \tau_2 - \tau_3} \right) \gamma_1 + n_0 \left(\frac{2e^{-\tau_2}}{3} - \frac{1}{6} e^{-3\tau_2 - \tau_3} - \frac{1}{3} e^{-\tau_2 - \tau_3} \right) \gamma_2 \\
a_{1,3}^0 &= a_{1,4}^0 = n_0 \frac{1}{6} e^{-\tau_1 - \tau_2 - \tau_3} \gamma_1 + n_0 \left(\frac{-1}{6} e^{-3\tau_2 - \tau_3} + \frac{1}{3} e^{-\tau_2 - \tau_3} \right) \gamma_2 \\
a_{2,2}^0 &= n_0 \left(\frac{e^{-\tau_1}}{3} - \frac{1}{6} e^{-\tau_1 - \tau_2 - \tau_3} \right) \gamma_1 + n_0 \left(\frac{e^{-\tau_2}}{3} - \frac{1}{6} e^{-\tau_2 - \tau_3} \right) \gamma_2 \\
a_{2,3}^0 &= a_{2,4}^0 = n_0 \frac{1}{6} e^{-\tau_1 - \tau_2 - \tau_3} \gamma_1 + n_0 \frac{1}{6} e^{-\tau_2 - \tau_3} \gamma_2 \\
a_{3,3}^0 &= a_{4,4}^0 = n_0 \left(\frac{1}{3} e^{-\tau_2 - \tau_3} - \frac{1}{6} e^{-\tau_1 - \tau_2 - \tau_3} \right) \gamma_1 + n_0 \frac{1}{6} e^{-\tau_2 - \tau_3} \gamma_2 \\
a_{3,4}^0 &= n_0 \left(1 - \frac{2e^{-\tau_1}}{3} - \frac{1}{3} e^{-\tau_2 - \tau_3} + \frac{1}{6} e^{-\tau_1 - \tau_2 - \tau_3} \right) \gamma_1 + n_0 \left(\frac{e^{-\tau_2}}{3} - \frac{1}{6} e^{-\tau_2 - \tau_3} \right) \gamma_2
\end{aligned}$$

Similarly, for Hybridization model 2 we have:

$$\begin{aligned}
a_{1,1}^0 &= n_0 \left(\frac{e^{-\tau_1}}{3} - \frac{1}{6} e^{-\tau_1 - \tau_2} \right) \gamma_1 + n_0 \left(1 - e^{-\tau_2} + \frac{1}{6} e^{-3\tau_2} \right) \gamma_2 \\
a_{1,2}^0 &= n_0 \left(1 - \frac{e^{-\tau_1}}{3} - \frac{2}{3} e^{-\tau_2} + \frac{1}{6} e^{-\tau_1 - \tau_2} \right) \gamma_1 + n_0 \left(\frac{1}{3} e^{-\tau_2} - \frac{1}{6} e^{-3\tau_2} \right) \gamma_2 \\
a_{1,3}^0 &= a_{1,4}^0 = n_0 \frac{1}{6} e^{-\tau_1 - \tau_2} \gamma_1 + n_0 \left(\frac{1}{3} e^{-\tau_2} - \frac{1}{6} e^{-3\tau_2} \right) \gamma_2 \\
a_{2,2}^0 &= n_0 \left(\frac{e^{-\tau_1}}{3} - \frac{1}{6} e^{-\tau_1 - \tau_2} \right) \gamma_1 + n_0 \frac{1}{6} e^{-\tau_2} \gamma_2 \\
a_{2,3}^0 &= a_{2,4}^0 = n_0 \frac{1}{6} e^{-\tau_1 - \tau_2} \gamma_1 + n_0 \frac{1}{6} e^{-\tau_2} \gamma_2 \\
a_{3,3}^0 &= a_{4,4}^0 = n_0 \left(\frac{1}{3} e^{-\tau_2} - \frac{1}{6} e^{-\tau_1 - \tau_2} \right) \gamma_1 + n_0 \frac{1}{6} e^{-\tau_2} \gamma_2 \\
a_{3,4}^0 &= n_0 \left(1 - \frac{2 \cdot e^{-\tau_1}}{3} - \frac{1}{3} e^{-\tau_2} + \frac{1}{6} e^{-\tau_1 - \tau_2} \right) \gamma_1 + n_0 \frac{1}{6} e^{-\tau_2} \gamma_2
\end{aligned}$$

For Hybridization model 3:

$$\begin{aligned}
a_{1,1}^0 &= n_0 \left(\frac{e^{-\tau_1}}{3} - \frac{1}{6} e^{-\tau_1 - \tau_2} \right) \gamma_1 + n_0 \left(1 - \frac{2 \cdot e^{-\tau_2}}{3} + \frac{1}{6} e^{-3\tau_2 - \tau_3} - \frac{1}{3} e^{-\tau_2 - \tau_3} \right) \gamma_2 \\
a_{1,2}^0 &= \frac{n_0}{6} e^{-\tau_1 - \tau_2} \gamma_1 + n_0 \left(\frac{2 \cdot e^{-\tau_2}}{3} - \frac{1}{6} e^{-3\tau_2 - \tau_3} - \frac{1}{3} e^{-\tau_2 - \tau_3} \right) \gamma_2 \\
a_{1,3}^0 &= \frac{n_0}{6} e^{-\tau_1 - \tau_2} \gamma_1 + n_0 \left(\frac{1}{3} e^{-\tau_2 - \tau_3} - \frac{1}{6} e^{-3\tau_2 - \tau_3} \right) \gamma_2 \\
a_{1,4}^0 &= n_0 \left(1 - \frac{e^{-\tau_1}}{3} + \frac{1}{6} e^{-\tau_1 - \tau_2} - \frac{2 \cdot e^{-\tau_2}}{3} \right) \gamma_1 + n_0 \left(\frac{1}{3} e^{-\tau_2 - \tau_3} - \frac{1}{6} e^{-3\tau_2 - \tau_3} \right) \gamma_2 \\
a_{2,2}^0 &= n_0 \left(\frac{e^{-\tau_2}}{3} - \frac{1}{6} e^{-\tau_1 - \tau_2} \right) \gamma_1 + n_0 \left(\frac{e^{-\tau_2}}{3} - \frac{1}{6} e^{-\tau_2 - \tau_3} \right) \gamma_2 \\
a_{2,3}^0 &= n_0 \left(1 - \frac{2 \cdot e^{-\tau_1}}{3} + \frac{1}{6} e^{-\tau_1 - \tau_2} - \frac{e^{-\tau_2}}{3} \right) \gamma_1 + \frac{n_0}{6} e^{-\tau_2 - \tau_3} \gamma_2 \\
a_{2,4}^0 &= \frac{n_0}{6} e^{-\tau_1 - \tau_2} \gamma_1 + \frac{n_0}{6} e^{-\tau_2 - \tau_3} \gamma_2 \\
a_{3,3}^0 &= n_0 \left(\frac{e^{-\tau_2}}{3} - \frac{1}{6} e^{-\tau_1 - \tau_2} \right) \gamma_1 + \frac{n_0}{6} e^{-\tau_2 - \tau_3} \gamma_2 \\
a_{3,4}^0 &= \frac{n_0}{6} e^{-\tau_1 - \tau_2} \gamma_1 + n_0 \left(\frac{e^{-\tau_2}}{3} - \frac{1}{6} e^{-\tau_2 - \tau_3} \right) \gamma_2 \\
a_{4,4}^0 &= n_0 \left(\frac{e^{-\tau_1}}{3} - \frac{1}{6} e^{-\tau_1 - \tau_2} \right) \gamma_1 + \frac{n_0}{6} e^{-\tau_2 - \tau_3} \gamma_2
\end{aligned}$$

where $\tau_j = \int_{t_0}^{t_0+T_j} (2N_j(t))^{-1} dt = (2N_j)^{-1} T_j$ is the “drift time”, according to Waxman (2011), and N_j

denotes the average effective population size for the j -th branch.

If retrotransposon insertions occur in branch T_1 ($t_0 < t < t_1$), we have the following estimations:

For Hybridization model 1 (Figure 6) and Hybridization model 2 (Figure 7) we expect that all probabilities $P(\omega_{i,j})$, except $P(\omega_{3,4})$, are equal to 0. For $P(\omega_{3,4} | X = x) = \gamma_1 x_1^2$, using the full

probability formula we obtain: $p_{3,4}(t) = P(\omega_{3,4}) = \gamma_1 \int_0^1 x_1^2 u_1(t, p, t_1, x_1) dx_1$.

Similar to the previous formulation (4) we can denote:

$$a_{3,4}^1 = \int_{t_0}^{t_0+T_1} p_{3,4}(t) dt = n_1 \gamma_1 \Phi(\tau_1), \quad (5)$$

where n_1 is the average number of retrotransposon insertions in branch T_1 , and $\Phi(\tau) = \tau - 1 + e^{-\tau}$.

For Hybridization model 3 (Figure 8) in branch T_1 we obtain:

$$\begin{aligned} P(\omega_{1,1} | X = x) &= x_2 x_3 (x_2 \gamma_1 + x_3 \gamma_2) \\ P(\omega_{1,2} | X = x) &= (1 - x_2) x_3 (x_2 \gamma_1 + x_3 \gamma_2) \\ P(\omega_{1,3} | X = x) &= x_2 x_3 (1 - x_2 \gamma_1 - x_3 \gamma_2) \\ P(\omega_{1,4} | X = x) &= x_2 (1 - x_3) (x_2 \gamma_1 + x_3 \gamma_2) \end{aligned}$$

and all other $P(\omega_{i,j})$ are equal to 0. Using a full probability formula, we obtain:

$$p_{1,j}(t) = P(\omega_{1,j}) = \int_0^1 \int_0^1 \int_0^1 u_1(t, p, t_1, x_1) u_2(t_1, x_1, t_2, x_2) u_3(t_1, x_1, t_3, x_3) dx_1 dx_2 dx_3.$$

Then, similar to (4), we can write $a_{i,j}^1 = \int_{t_0}^{t_0+T_1} p_{i,j}(t) dt$, and after expansion we get:

$$\begin{aligned} a_{1,1}^1 &= n_1 \Phi(\tau_1) - \frac{n_1}{6} e^{-\tau_2} (1 - e^{-\tau_1})^2 (2 + e^{-\tau_1}) \gamma_1 - \frac{n_1}{6} e^{-\tau_3} (1 - e^{-\tau_1})^2 (2 + e^{-\tau_1}) \gamma_2 \\ a_{1,2}^1 &= \frac{n_1}{6} e^{-\tau_2} (1 - e^{-\tau_1})^2 (2 + e^{-\tau_1}) \gamma_1 + n_1 (1 - e^{-\tau_1}) \left(1 - \frac{1}{6} (e^{-2\tau_1} + e^{-\tau_1} + 4) e^{-\tau_3} \right) \gamma_2 \\ a_{1,3}^1 &= \frac{n_1}{6} e^{-\tau_2} (1 - e^{-\tau_1})^2 (2 + e^{-\tau_1}) \gamma_1 + \frac{n_1}{6} e^{-\tau_3} (1 - e^{-\tau_1})^2 (2 + e^{-\tau_1}) \gamma_2 \\ a_{1,4}^1 &= n_1 (1 - e^{-\tau_1}) \left(1 - \frac{1}{6} (e^{-2\tau_1} + e^{-\tau_1} + 4) e^{-\tau_3} \right) \gamma_1 + \frac{n_1}{6} e^{-\tau_3} (1 - e^{-\tau_1})^2 (2 + e^{-\tau_1}) \gamma_2. \end{aligned}$$

We then estimate parameters for retrotransposon insertions taking place in branch T_2 ($t_1 < t < t_2$).

For Hybridization model 1 (Figure 6) we have:

$$\begin{aligned} P(\omega_{1,1} | X = x) &= x_2 x_3^2 \gamma_2 \\ P(\omega_{1,2} | X = x) &= x_3^2 (1 - x_2 \gamma_2) \\ P(\omega_{1,3} | X = x) &= P(\omega_{1,4} | X = x) = x_2 (1 - x_3) x_3 \gamma_2 \end{aligned}$$

and all other $P(\omega_{i,j})$ are equal to 0. Using a full probability formula, we obtain:

$$p_{1,j}(t) = P(\omega_{1,j}) = \int_0^1 \int_0^1 u_2(t, p, t_2, x_2) u_3(t_2, x_2, t_3, x_3) dx_2 dx_3.$$

Then, similar to above we get $a_{i,j}^2 = \int_{t_0}^{t_0+T_2} p_{i,j}(t) dt$, and expand this to:

$$\begin{aligned} a_{1,1}^2 &= n_2 \left(\Phi(\tau_2) - \frac{1}{6} e^{-\tau_3} (1 - e^{-\tau_2})^2 (2 + e^{-\tau_2}) \right) \gamma_2 \\ a_{1,2}^2 &= n_2 \left(\left(\Phi(\tau_2) + (1 - e^{-\tau_2}) (1 - e^{-\tau_3}) \right) \gamma_1 + (1 - e^{-\tau_2}) \left(1 - \frac{1}{6} (e^{-2\tau_2} + e^{-\tau_2} + 4) e^{-\tau_3} \right) \gamma_2 \right) \\ a_{1,3}^2 &= a_{1,4}^2 = n_2 \frac{1}{6} e^{-\tau_3} (1 - e^{-\tau_2})^2 (2 + e^{-\tau_2}) \gamma_2. \end{aligned}$$

For Hybridization model 2 (Figure 7) we obtain:

$$\begin{aligned} P(\omega_{1,1} | X = x) &= x_2 x_3^2 \gamma_2 \\ P(\omega_{1,2} | X = x) &= (1 - x_2) x_3^2 \gamma_2 \\ P(\omega_{1,3} | X = x) &= x_2 (1 - x_3) x_3 \gamma_2 \\ P(\omega_{1,4} | X = x) &= x_2 x_3 (1 - x_3 \gamma_2) \end{aligned}$$

and all other $P(\omega_{i,j})$ are equal to 0. Similar to above we get:

$$\begin{aligned} a_{1,1}^2 &= n_2 \left(\Phi(\tau_2) - \frac{1}{6} (1 - e^{-\tau_2})^2 (2 + e^{-\tau_2}) e^{-\tau_3} \right) \gamma_2 \\ a_{1,2}^2 &= n_2 (1 - e^{-\tau_2}) \left(1 - \frac{1}{6} (4 + e^{-\tau_2} + e^{-2\tau_2}) e^{-\tau_3} \right) \gamma_2 \\ a_{1,3}^2 &= \frac{n_2}{6} (1 - e^{-\tau_2})^2 (2 + e^{-\tau_2}) \cdot e^{-\tau_3} \gamma_2 \\ a_{1,4}^2 &= n_2 \left(\Phi(\tau_2) \gamma_1 + \frac{1}{6} (1 - e^{-\tau_2})^2 (2 + e^{-\tau_2}) e^{-\tau_3} \gamma_2 \right) \end{aligned}$$

For Hybridization model 3 (Figure 8) all probabilities $P(\omega_{i,j})$, except $P(\omega_{3,4})$, are equal to 0. For

$$P(\omega_{3,4} | X = x) = \gamma_1 x_1^2, \text{ similar to above, we obtain: } a_{1,4}^2 = n_2 \gamma_1 \Phi(\tau_2).$$

If retrotransposon insertions occur in branch T_3 ($t_2 < t < t_3$), then for Hybridization model 1

(Figure 6) we have: $a_{1,2}^3 = \int_{t_2}^{t_2+T_3} p_{1,2}(t) dt = n_3 \Phi(\tau_3)$. Similarly, for the Hybridization models 2 and 3

we obtain: $a_{1,2}^3 = n_3 \gamma_2 \Phi(\tau_3)$.

The final result for each of the hybridization models is obtained by summarizing:

$$a_{i,j} = a_{i,j}^0 + a_{i,j}^1 + a_{i,j}^2 + a_{i,j}^3 \quad (6)$$

according to the type of hybridization.

2. Special cases

The binary tree topologies can be derived from hybridization models using specific parameter values. For example, in Hybridization models 1 and 2 (Figures 6 and 7, respectively), setting $\gamma_1 = 0$ (and, respectively $\gamma_2 = 1$) results in Tree topology 1 (Figure 1), while setting $\gamma_1 = 1$ in these two hybridization models results in Tree topology 2 (Figure 2). In Hybridization model 3 (Figure 8), setting $\gamma_1 = 0$ or $\gamma_1 = 1$ both result in Tree topology 2 (Figure 2).

Partial polytomy 1 (Figure 3) can be derived from all three hybridization models with the following parameter settings: $\gamma_1 = 0$, $\tau_3 = 0$. Partial polytomy 2 (Figure 4) can be derived from Hybridization models 1 and 2 (Figures 6 and 7, respectively) by the settings $\gamma_1 = 0$, $\tau_2 = 0$. In all

the described models, the settings $\tau_1 = 0$, $\tau_2 = 0$, and $\tau_3 = 0$ lead to Full polytomy (Figure 5), where the value of γ is indefinite. In Hybridization models 1 and 2, the settings $0 < \gamma_1 < 1$, $\tau_3 = 0$ lead to an additional mixed hybridization-polytomy configuration (Figure 9).

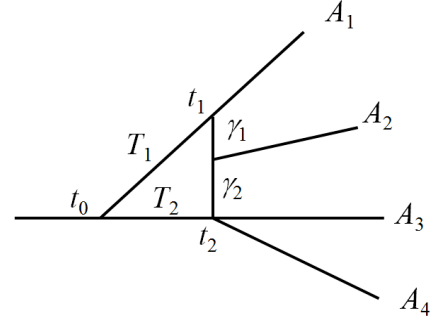


Fig. 9. Mixed hybridization-polytomy

3. Likelihood approach

To estimate the tree topology for real presence/absence patterns obtained from genomic screenings, we denote the number of markers $\eta_{i,j}$ with condition $\omega_{i,j}$ detected in the screening. For these informative markers we denote the probability β that they were detected. Then $\eta_{i,j}$ is a Poisson distributed random variable with parameters $a_{i,j}$, defined earlier (6), where n_0, n_1, n_2, n_3 are replaced by $\beta n_0, \beta n_1, \beta n_2, \beta n_3$.

Denoting $\eta = \{\eta_{i,j}\}_{1 \leq i \leq j \leq 4}$, $y = \{y_{i,j}\}_{1 \leq i \leq j \leq 4}$, where $y_{i,j}$ are arbitrary non-negative values, we can write:

$$P(\eta = y) = \prod_{1 \leq i \leq j \leq 4} \frac{a_{i,j}^{y_{i,j}}}{y_{i,j}!} e^{-a_{i,j}}.$$

Estimates of the parameters for the corresponding model are derived by maximizing the log-likelihood function:

$$L(y, \theta) = \sum_{1 \leq i \leq j \leq 4} (y_{i,j} \cdot \ln(a_{i,j}) - a_{i,j}), \quad (7)$$

where $\theta = \{n_0, n_1, n_2, n_3, \tau_1, \tau_2, \tau_3, \gamma\}$ is the model parameters vector, and functions $a_{i,j} = a_{i,j}(\theta)$ are defined earlier (6) (here $\gamma_1 = \gamma$, and $\gamma_2 = 1 - \gamma$).

Note, that for Full polytomy (Figure 5) ($\tau_1 = \tau_2 = \tau_3 = 0$) all $a_{i,j} = a = \frac{n_0}{6}$, and the log-likelihood function $L_0(y, a) = \sum_{1 \leq i \leq j \leq 4} y_{i,j} \cdot \ln(a) - 10a$. Equating to zero the derivative of $L_0(y, a)$ of a :

$$\frac{1}{a} \sum_{1 \leq i \leq j \leq 4} y_{i,j} - 10 = 0, \text{ provides an estimation of maximum likelihood } \hat{a} = \frac{1}{10} \sum_{1 \leq i \leq j \leq 4} y_{i,j}.$$

Thus the maximum value of log-likelihood function in case of Full polytomy is equal to:

$$L_0(y, \hat{a}) = 10\hat{a}(\ln(\hat{a}) - 1).$$

4. Testing laurasiatherian data

According to the requirements of our mathematical diffusion model, we included all phylogenetically informative markers, with those that were unfixed in the common ancestor of the four investigated laurasiatherian lineages (Chiroptera, Perissodactyla, Cetartiodactyla, Carnivora) or differently represented in the investigated orders (see Table 1). Due to the increasing complexity involved in including additional lineages we did not consider the fifth lineage Eulipotyphla that is clearly located outside the remaining laurasiatherians. Under our model conditions the moment of fixation of retroelements rather than the moment of their insertion is relevant. Our diffusion model is tolerant to any "outside of the focused lineages" conflicts. Therefore, the dataset for the 4-Lineage Insertion Likelihood test consists of 102 diagnostic markers (with no conflicts in the Eulipotyphla and Pholidota groups) plus 60 markers where Eulipotyphla sequence information was absent or Eulipotyphla or Pholidota showed conflicts (the additional 60 markers see in the Supplemental_Table_S1d, e and Supplemental_Material_S2b, c).

Here $\sum_{1 \leq i \leq j \leq 4} y_{i,j} = 162$, and accordingly $\hat{a} = 16.2$, hence $\max_{H_0} L(y, \theta) = L_0(y, \hat{a}) = 289,1718$.

Table 1. Distribution of presence/absence patterns for different combinations of the four investigated laurasiatherian branches used in the 4-Lineage Insertion Likelihood test.

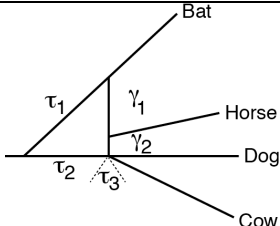
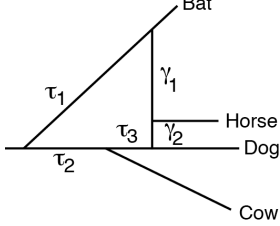
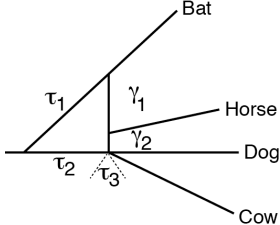
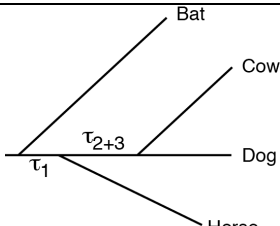
Dog	Cow	Horse	Bat	y_{ij}
+	+	-	-	17
+	-	+	-	18
+	-	-	+	12
-	+	+	-	11
-	+	-	+	7
-	-	+	+	21
+	+	+	-	24
+	+	-	+	16
+	-	+	+	14
-	+	+	+	22

Pluses and minuses denote presence and absence of individual retrotransposon insertions, respectively, and (y_{ij}) indicates the total number of insertions for each pattern.

We obtained maximum likelihood values for all presented hybridization models (taking rearrangements into account) using the FindMaximum function of the software package Wolfram Mathematica 10 (Version Number 10.4.0.0). For simplification we took $n_0 = n_1 = n_2 = n_3$. Our model conditions are additionally restricted to $\tau_1 \geq 0$; $\tau_2 \geq 0$; $\tau_3 \geq 0$; $0 \leq \gamma \leq 1$. The maximum log-

likelihood ratio was calculated as $d(y) = 2 \left(\max_{\theta} L(y, \theta) - L_0(y, \hat{a}) \right)$. According to Waddell et al. (2001), we approximated this difference by a χ -square distribution with two degrees of freedom. The best results for different hybridization models are presented in Table 2.

Table 2. Maximum log-likelihood ratios for different hybridization models.

Model configurations	Resulting Tree topologies	log-likelihood ratios (*significant)	τ_1	τ_2	τ_3	γ ($\gamma = \gamma_1$)	p-value
Hybridization model 1		6.986*	0.498	0.145	0	0.236	p<0.031
Hybridization model 2		7.057*	0.424	0.144	0.027	0.284	p<0.029
		6.986*	0.498	0.145	0	0.236	p<0.031
Hybridization model 3		3.630	0.063	0.065	0.078	0	p>0.1

Hybridization model 1 assigns a maximum log-likelihood value for a hybridization-polytomy topology (Figure 9), where the horse is a result of fusion of the dog-cow and the bat ancestors with a somehow stronger connection to the dog-cow ancestor (here $\gamma = \gamma_1$, hence $\gamma_2 = 1 - \gamma \approx 0.76$).

Hybridization model 2 represents two significant tree topologies. The first shows the maximum log-likelihood value representing hybridization by fusion between the dog and bat ancestors resulting in the horse lineage. This case shows a very short branch of dog ancestor before fusion, indicated by $\tau_3 = 0.027$ and other parameters are similar to the hybridization-

polytomy topology. The second tree topology is identical to the resulting *Tree* topology shown for Hybridization model 1.

The maximum log-likelihood ratio for Hybridization model 3 was not significant, but does indicate a resulting *Tree* topology without hybridization and a very short branch of the common ancestor of dog, cow, and horse ($\tau_1 = 0.063$; $\tau_{2+3} = 0.143$), identical to the tree topology we received with Dollop in Phylip (see Figure 2 here, also Figure 2 in the main text).

References

- Fisher RA. 1922. On the dominance ratio. *Proc Roy Soc, Edinb* **42**: 321-341.
- Kimura M. 1955a. Stochastic processes and distribution of gene frequencies under natural selection. *Cold Spring Harb Symp Quant Biol* **20**: 33–53.
- Kimura M. 1955b. Solution of a process of random genetic drift with a continuous model. *Proc Natl Acad Sci U S A* **41**: 144-150.
- Kuritzin A, Kischka T, Schmitz J, Churakov G. 2016. Incomplete lineage sorting and hybridization statistics for large-scale retroposon insertion data. *PLoS Comput Biol* **12**: e1004812. doi:10.1371/journal.pcbi.1004812.
- Waddell PJ, Kishino H, Ota R. 2001. A phylogenetic foundation for comparative mammalian genomics. *Genome Inform* **12**: 141-154.
- Waxman DA. 2011. Unified treatment of the probability of fixation when population size and the strength of selection change over time. *Genetics* **188**: 907–913.
- Wright S. 1931. Evolution in Mendelian Populations. *Genetics* **16**: 97-159.