Supplementary Materials, Part 2, for the Blueprint of the "Alignment Neighborhood Explorer" (ANEX) (tentatively named),

by Kiyoshi Ezawa

(Finished on August 4th, 2016; TOC edited on August 14th, 2020)

Table of Contents

Supplementary Results & Discussion

SR-1. Characterization of "purge"-like errors

To characterize the "purge"-like MSA errors, we used three sets of simulated MSAs: simulated MSAs among 12 primates **(set 3P)**, those among 15 mammals **(set 3M)**, and those among 9 fast-evolving mammals **(set 3F)**. Then we compared the actual "purge"-like errors made by a state-of-the-art aligner, PRANK (Loytynoja and Goldman, 2008), with the regions that were correctly reconstructed (control MSAs). The comparisons were made in terms of the P-value of the substitutional difference (defined as Eq.(A1-8) in the main Appendix), the branch length, the number of sequences 'involved', and the size of the block defining the "purge" (or control). We also examined the distribution of position-shifts caused by the "purge"-like errors. We found the following trends.

1. The frequency of (the absolute value of) the position-shift rapidly decreases as the position-shift increases (Table SS1).
2. The frequency of the block-size is nearly uniform when it is 1, 2, 3 and 4. Then, the frequency gradually decreases as the block-size increases (Figure SS1).
3. The frequency of "purge"s depends quite heavily on the length of the branch separating the sequences involved in the "purge"; the "purge"s occur more frequently along longer branches (Table SS2). The frequency also seems to vary depending on the number of sequences 'involved' (data not shown). However, such dependence seems originated mostly from the dependence on the branch length (data not shown).
4. The P-values are not distributed uniformly among controls (Table SS3). For example, the cumulative relative frequencies of controls with P-value < 0.20 were: 0.056 for set 3P, 0.057 for set 3M, and 0.038 for set 3F. (This may be because regions with less substitutional differences are more likely to be reconstructed correctly.) In contrast, the "purge"-like errors tend to have smaller P-values (Table SS3). For example, the cumulative relative frequencies of "purge"s with P-value < 0.20 were: 0.925 for set 3P, 0.78 for set 3M, and 0.55 for set 3F. (In general, the power to identify true "purge"s seems to decrease as the average branch length (or the total branch length) increases. However, this might not be such a serious problem, because set 3F is probably beyond the scope where ANEX is applicable.)
5. The P-value distribution also varies depending on the block size (Table SS4) and the branch length (Table SS5). In short, for both subjects and controls, the cumulative distribution shifts towards P-value = 0 as the block size increases and as the branch length increases. Thus, it might be a good idea to let the threshold P-value depend on the block

size and/or the branch length.

SR-2. Examining "Shift + Shift" error pairs

When examining the types of MSA errors in (Ezawa 2016a), we encountered quite a few pairs of errors classified as "Shift + Shift" (see Table S5 of *ibid*.). Here, we further examined such "Shift + Shift" error pairs. First, we sub-classified the error pairs more finely. As shown in Table SS6, an overwhelming majority of the pairs belonged to either "Shift + Shift(???)" or "Shift + Shift(gap-aligned)(?)." Thus, we examined some sample errors belonging to these two sub-classes (in Set 3M). We found the following (see Figure SS2).

(1) Most of "Shift + Shift(???)" were actually simpler errors (usually a simple "Shift" or two simple "Shift"s) that were mis-annotated because the associated position-shift-blocks were of somewhat complex shapes and thus mis-partitioned (Figure SS2, panels A & B). (Or, quite often, the interpretation gets complicated because of the *opposite positioning of independent insertions* in the reference and reconstructed MSAs (Figure SS2, panels B & C).)

(2) Most of "Shift + Shift(gap-aligned)(?)" were actually simple (nearly) independent combinations, each of a simple shift and a gap-aligned shift. But the latter was mis-annotated as "Complex" at the 1st stage because the involved block shared one end with the aligned gap (Figure SS2, panel D). (Currently, our prototype script cannot correctly handle such cases.)

Therefore, both of these erroneous annotations are expected to be rectified if (A) we ***correctly*** identify the position-shift blocks, and if (B) we consider the ***serial* effects** of the moves of the blocks, paying attention to the resulting **change(s)** in the inference of indels ***by the move of each single block*,** instead of considering the effects of parallel moves of the blocks.

However, some of these error pairs, especially those belonging to "Shift + Shift(gap-aligned)(?)," were quite long and involved many (> 4) inferred indels (and thus blocks) that need be taken into account. Such cases seem **beyond the scope** of the initial version of ANEX.

SR-3. Characterizing regions with "Complex" errors **[NEWLY ADDED (1)]**

One of the key points for the success of ANEX would be how accurately we can identify regions of the input reconstructed MSA that are likely to contain "Complex" errors, which are hard to rectify via the exploration of "neighboring" MSAs that ANEX is supposed to attempt.

Thus, we compared the **"Complex" errors** (including the errors that were "too

long" to be examined by ComplLiMment (Ezawa, 2016a)) with the non-complex errors and "correct" segments (as a composite **control**), in terms of such properties as the number of columns, the number of inferred indels, and the size of each gapless segment separating gapped segments (of particular types), etc. In the following, I will report on the results of such analyses.

SR-3.1. Properties of individual gapped segments

First, we compared individual gapped segments belonging to "Complex" errors and those belonging to the control. Major findings are as follows. (See "complex1.xls" for details.)

(1) Segment size (= the number of columns constituting the segment). The gapped segments in "Complex" errors tended to be larger than those in the control (Table SS7). However, the difference was not so remarkable. For example, up to the size corresponding to 90 percentile (from bottom) of the control gapped segments, as much as 67-75% of "Complex" gapped segments were distributed, regardless of the simulated MSA sets (Table SS7).

(2) Number of indels. The segments in "Complex" errors tended to show a larger number of indels (inferred via a Dollo parsimony method) than the control gapped segments. However, the difference was not so remarkable for this quantity, either (Table SS8). For example in set 3M (15 mammals), about 94% of control gapped segments showed 1 or 2 indels, whereas about 76% of "Complex" gapped segments did so.

(3) Total length of indels. This property exhibited almost the same tendency as the segment size (Table SS9).

(4) The number of insertions vs the number of deletions. The "Complex" gapped segments seem to be richer in insertions than the control gapped segments, though the bias is not so large (data not shown).

In summary, although the individual gapped segments of the two categories (control vs "Complex") showed different tendencies (or distributions), **the difference was not so clear as to sharply separate the two categories,** in terms of any of these properties.

SR-3.2. Entire "Complex" errors vs entire control segments

Second, we examined the properties of the segment of each "Complex" error as a whole, and compared with the properties of each control segment (also as a whole). We found the following. (See "complex2.xls" for details.)

(1) The number of gapped segments in each ("Complex" erroneous or control) segment. For set 3P (*i.e.*, 12 primates), a substantial fraction of "Complex" segments contain only one

gapped segment each (Table SS10). In contrast, for each of set 3M (*i.e.*, 15 mammals) and set 3F (*i.e.*, 9 fast-evolving mammals), the two categories show a marked difference in the distribution, so that it may be used to distinguish the categories (Table SS10).

(2) Total number of columns. Similarly to (1), the two categories show a marked difference in the distribution of this quantity for sets 3M and 3F, whereas the difference is relatively small for set 3P (Table SS11).

(3) Total number of indels. The two categories show an extremely conspicuous difference in the distribution of this quantity, especially for sets 3M and 3F. For example, only 0.9-2.7% and 7-38%, respectively, of "Complex" erroneous segments had <=1 and <= 2 inferred indels each, whereas the fractions were as much as 56-67% and 78-89%, respectively, among control segments (Table SS12).

(4) **Maximum number of indels** among the gapped segments in each "Complex" erroneous or control segment. The two categories show a marked difference in the distribution of this quantity, especially for sets 3M and 3F (Table SS13), although the difference is not as conspicuous as that in the total number of indels described in (3).

(5) Maximum number of columns. The difference in this quantity between the two categories is nearly as remarkable as that in the total number of columns described in (2) (data not shown).

(6) Maximum of the total length of indels. The two categories show nearly as remarkable difference in this quantity as that in the maximum number of columns described in (5) (data not shown).

(7) The number of insertions vs the number of deletions. The "Complex" erroneous segments are notably richer in insertions than the control segments (see, *e.g.*, the distributions of Max(#{ins}-#{del}) in Table SS14).

These analyses indicated that, *once correctly partitioned*, the "Complex" erroneous segments could be separated from the control segments with high accuracies, taking advantage of the conspicuous differences in the distributions of the properties of the entire segment.
A serious problem is that, *without the true MSA*, we have no means to correctly partition a reconstructed MSA into erroneous and correct segments. In order to take advantage of the aforementioned conspicuous differences, we need to come up with a method to quite accurately cluster the gapped segments into *approximate* erroneous and correct segments.

SR-3.3. Size of spacer

The size of a **spacer**, *i.e.*, a gapless segment between a pair of gapped segments, may differ depending on the categories of the flanking gapped segments, and thus it may be used to infer

regions containing "Complex" errors. (See "complex3.xls" and "complex3p.xls" for detailed results.)

First, we measured the sizes of two spacers flanking each gapped segment, and took two distributions of each of the average size, the larger size and the smaller size; one distribution is for gapped segments belonging to "Complex" erroneous segments, and the other is for those belonging to control segments. For any of the three types of the spacer size, spacers flanking the "Complex" gapped segments clearly tended to be smaller than than those flanking the control gapped segments (Table SS15). However, the distribution difference was not large enough to sharply separate the "Complex" gapped segments from the control gapped segments.

Next, we examined the size of each spacer, and classified the spacer by the categories of its flanking gapped segments. The spacer flanked by gapped segments belonging to two different erroneous or correct segments is expected to be longer than the spacer flanked by gapped segments belonging to the same erroneous or correct segment. Thus, they were put into separate categories. In total, there were 5 categories (illustrated in Figure SS3): 'A:A_in', 'B:B_in', 'A:A_ex', 'A:B_ex', and 'B:B_ex'. Here, 'A' and 'B' stands for a "Complex" gapped segment and a control gapped segment, respectively. Regarding the subscripts, '_in' indicates that the two gapped segments belong to the same erroneous or correct segment, and '_ex' indicates that the two gapped segments belong to different erroneous or correct segments. The inspection of the distributions for set 3M revealed a notable difference in the distribution between the broad categories, '_in' and '_ex' (Table SS16), as expected above. However, the differences between the categories within the same broad category were not so conspicuous (Table SS16) (, although they also were conspicuous for set 3P (data not shown)). This analysis indicates that the spacer size difference could be exploited at least to *artificially* cluster the gapped segments into likely erroneous or correct segments, with some (moderate) accuracy.

SR-3.4. Artificial clusters of gapped segments

As mentioned in *SR-3.2*, without the true MSA, there is no way of partitioning a reconstructed MSA into correct and erroneous segments. Thus, we need to come up with a method to *artificially* cluster the gapped segments, in order to enhance the accuracy of identifying regions that contain "Complex" errors. We tried three different methods. (Figure SS4 schematically illustrates these methods.)

In a simple clustering method **(Method I)**, two nearest-neighboring gapped segments are clustered if the spacer between them is shorter than a threshold value, *regardless of the gap*

patterns of the gapped segments (panel B of Figure SS4). **Method II** and **Method III** take account of gap patterns as well. **Method III** clusters two quite close gapped segments if they undergo indels along the same branch or along two phylogenetically nearest-neighboring branches (panel C). **Method II** is similar to Method III, but the requirement on the gap pattern is stricter and more complex; more precisely, **Method II** clusters two (not necessarily nearest-neighboring) gapped segments, as well as all segments in between them, if they are quite close to each other and either if they undergo indels along the same branch or if all three branches sharing an internal node undergo indels in these gapped segments and yet another neighboring segment (panel D).
See SM-3.2 for more details on these methods.

In our analyses, Method III performed the best and Method I performed the worst, and Method II performed slightly less efficiently than Method III, under respective optimum combinations of the spacer-size threshold and the spacer-size upper-bound (Table SS17).
[For details, see "complex6p_xx.xls", "complex6pp_xx.xls", "complex6p_xx.wo_ud4.xls", "complex6pp_xx.wo_ud4.xls", and "complex6pp_xx.wo_ud4.part2.xls" (with "_xx" = "", "_aug", "_aug_len").]

Thus, we will use Method III as the default artificial clustering method for ANEX.

[END of "NEWLY ADDED (1)"]

Supplementary Methods

SM-1. Characterizing "purge"-like errors

We performed a set of control analyses. The subjects are erroneously reconstructed segments each of which consists only of a single "purge." We examined the following statistics of each purge: (1) the size of the involved block, (2) (the absolute value of) the position-shift that the block underwent, (3) the length of the branch separating the involved block, (4) the number of "descendant" extant sequences of the branch, and (5) the P-value regarding substitutional differences along the branch (computed using Eq.(A1-8)). The controls were extracted from correctly reconstructed segments. Along each branch and for each window size (among 1, 2, …, 10), we prepared three windows if possible, from the left end, the center, and the right end of the segment excluding single columns at both ends. The window was excluded from the examination if it is horizontally interfered by at least an indel along the "purge"-separating branch or along one of its neighboring branches.

SM-2. Examining "Shift + Shift" error pairs

First, we selected those erroneous segments each of which contains at least one error pair of the "Shift + Shift" type. Then, we excluded those segments containing "Complex" errors. Then, we sub-classified the pairs according to the more detailed annotations of individual components constituting each "Shift + Shift" pair. (Some examples are: "Shift," "Shift(?),"Shift(???)," and "Shift(gap-aligned)(?).") Then, the sub-classes were tallied in each of the simulated MSA sets, 3P, 3M and 3F. Finally, we manually inspected some sample erroneous segments whose error pairs belongs to either of the two commonest sub-classes, "Shift + Shift(???)" and "Shift + Shift(gap-aligned)(?)," in the Set 3M (i.e., 15 mammals).

SM-3. Examining "Complex" errors [NEWLY ADDED (2)]

SM-3.1. Properties of "Complex" erroneous segments

…Maybe necessary later…

SM-3.2. Artificial clusters of gapped segments

In order to artificially cluster the gapped segments, we tried three different methods:

Method I: Two nearest-neighboring gapped segments are clustered if the spacer between them is shorter than **{spacer-size threshold}** (Figure SS4, panel B).

Method III: Two (not necessarily nearest-neighboring) gapped segments, as well as all

segments in between them, are artificially clustered if the following conditions are satisfied:

(a) EITHER [the (composite-)spacer between them is <= **{spacer-size threshold}**], OR [the (composite-)spacer is <= **{spacer-size upper-bound}**, AND {size of smaller gapped segment} >= {size of (composite-)spacer} /2, AND {size of larger gapped segment} >= 2 *{size of (composite-)spacer}];

(b) [the two gapped segments undergo at least one indel each along the same branch] OR [the two gapped segments undergo indels along two phylogenetically nearest-neighboring branches] (Figure SS4, panel C).

Method II: Similar to Method III, but the latter half of condition (b) is replaced as: [the two gapped segments, and yet another neighboring gapped segment undergo indels along three branches sharing an internal node] (Figure SS4, panel D).

Then, the artificial clusters were classified according to (#{insertions}, #{deletions}), and the total number of "Complex" gapped segments and control gapped segments were counted in each cluster. The clusters were then sorted in ascending order of the ratio:
#{"Complex" gapped segments} / #{control gapped segments}.
Finally, from the cluster with the smallest ratio, the clusters were chosen until #{control gapped segments} reaches a specified fraction (80% or 90%) of the total number of the control gapped segments.
We tried various combinations of the two parameters, namely, the threshold value and the upper-bound, both of the spacer size.

Before this analysis, each "Complex" error was re-classified as "non-complex" if it is inferred to result from only less than 4 indels, for each of the reconstructed and reference MSAs.
[END of "NEWLY ADDED (2)"]

Supplementary Tables SS1-SS??

Table SS1. Distribution of absolute values of position-shifts among "purge"s

|Shift|	Set 3P	Set 3M	Set 3F
1	0.840	0.841	0.838
2	0.099	0.104	0.103
3	0.023	0.028	0.030
4	0.017	0.009	0.017
5	0.010	0.006	0.005
6	0.006	0.003	0.005
7	0.001	0.002	0.002
8	0.001	0.002	0
9	0	0.001	0
>= 10	0.001	0.002	0
#{"Purge"s examined}	694	4652	593

NOTE: Set 3P, 3M and 3F consist of MSAs of fictitious DNA sequences simulated along the trees of 12 primates, 15 mammals and 9 fast-evolving mammals, respectively. Shown in each cell is the relative frequency of the specified absolute value of the position-shift (row) in a specified MSA set (column).

Table SS2. Richness of "purge"s along branches with various lengths

Branch length (=X)	12 primates	15 mammals	9 fe-mammals
0.00 <= X < 0.02	1.82E-06	4.74E-07	3.03E-06
0.02 <= X < 0.04	1.72E-05	4.78E-06	0
0.04 <= X < 0.06	5.96E-05		
0.06 <= X < 0.08		2.69E-05	1.53E-05
0.08 <= X < 0.10		3.99E-05	2.43E-05
0.10 <= X < 0.12			4.83E-05
0.12 <= X < 0.14		8.80E-05	
0.14 <= X < 0.16		1.33E-04	
0.16 <= X < 0.18		1.56E-04	6.94E-05
0.18 <= X < 0.20		1.85E-04	1.53E-04
0.20 <= X < 0.25			1.83E-04
0.25 <= X < 0.30			2.61E-04
0.30 <= X			4.00E-04
Overall average	9.67E-06	3.67E-05	1.19E-04

NOTE: Shown in each cell is the richness of the "purge"s (defined as #{"purges"}/#{controls}), in a specified MSA set (column), along the branches whose lengths are in a specified range (row).

Table SS3. Cumulative relative distributions of the P-values in different simulated MSA sets

	Subjects			Controls		
P-value (=X)	**12 primates**	**15 mammals**	**9 FE-mammals**	**12 primates**	**15 mammals**	**9 FE-mammals**
X < 0.001	0.363	0.085	0.022	1.2E-04	5.9E-05	1.0E-05
X < 0.005	0.571	0.196	0.056	7.7E-04	4.8E-04	1.2E-04
X < 0.010	0.618	0.263	0.088	1.4E-03	9.8E-04	3.4E-04
X < 0.025	0.68	0.38	0.169	4.1E-03	3.3E-03	1.5E-03
X < 0.050	0.762	0.495	0.258	0.01	8.8E-03	4.0E-03
X < 0.100	0.874	0.618	0.39	0.026	0.023	0.012
X < 0.200	0.925	0.78	0.545	0.056	0.057	0.038
X < 0.500	0.938	0.919	0.862	0.085	0.131	0.177
X < 0.800	0.941	0.938	0.909	0.086	0.184	0.331
X < 0.900	0.942	0.942	0.924	0.088	0.195	0.389
X < 0.950	0.942	0.944	0.927	0.092	0.211	0.432
X < 0.975	0.944	0.945	0.931	0.099	0.236	0.493
X < 0.990	0.944	0.951	0.938	0.105	0.296	0.586
X < 0.995	0.945	0.954	0.948	0.115	0.349	0.642
X < 0.999	0.951	0.959	0.949	0.176	0.444	0.74
X < 1.000	1	1	1	0.237	0.492	0.843

NOTE: Shown in each cell is the relative frequency, in a specified simulated MSA set (column), of "purge"-involved blocks (in the subjects) or windows (in the controls) with P-values less than a specified value (row).

Table SS4. Cumulative distributions of the P-values for various sizes of "purge"-involved blocks (for set 3M, *i.e.*, 15 mammals)

A. For subjects

Block size	[1,2)	[2,3)	[3,4)	[4,5)	[5,6)	[6,7)	[7,8)	[8,9)	[9,10)	[10,12)	Total
(P-value) < 0.001	0	1.1E-03	0.022	0.097	0.121	0.192	0.180	0.235	0.312	0.370	0.085
(P-value) < 0.005	0	0.020	0.124	0.241	0.337	0.351	0.391	0.512	0.514	0.647	0.196
(P-value) < 0.010	0	0.034	0.241	0.332	0.428	0.476	0.480	0.641	0.587	0.681	0.263
(P-value) < 0.025	3.8E-03	0.170	0.326	0.532	0.566	0.610	0.684	0.771	0.780	0.798	0.380
(P-value) < 0.050	0.027	0.395	0.405	0.633	0.732	0.756	0.746	0.859	0.908	0.849	0.495
(P-value) < 0.100	0.070	0.483	0.723	0.744	0.806	0.878	0.887	0.929	0.963	0.924	0.618
(P-value) < 0.200	0.619	0.541	0.793	0.907	0.933	0.930	0.938	0.971	0.991	0.958	0.780
(P-value) < 0.500	0.722	0.901	0.966	0.976	0.970	0.988	0.977	1	1	0.983	0.919
(P-value) < 0.800	0.753	0.927	0.977	0.990	0.998	0.997	0.996	1	1	1	0.938
(P-value) < 0.900	0.758	0.934	0.984	0.994	0.998	0.997	0.996	1	1	1	0.942
(P-value) < 0.950	0.763	0.938	0.984	0.995	0.998	0.997	0.996	1	1	1	0.944
(P-value) < 0.975	0.767	0.939	0.986	0.995	0.998	0.997	0.996	1	1	1	0.945
(P-value) < 0.990	0.791	0.944	0.988	0.995	1	0.997	0.996	1	1	1	0.951
(P-value) < 0.995	0.795	0.950	0.992	0.997	1	0.997	0.996	1	1	1	0.954
(P-value) < 0.999	0.811	0.960	0.993	0.998	1	0.997	1	1	1	1	0.959
(P-value) < 1.000	1	1	1	1	1	1	1	1	1	1	1

B. For controls

Block size	[1,2)	[2,3)	[3,4)	[4,5)	[5,6)	[6,7)	[7,8)	[8,9)	[9,10)	[10,12)	Total
(P-value) < 0.001	0	2.4E-05	4.6E-05	8.4E-05	7.1E-05	1.2E-04	9.0E-05	7.8E-05	9.1E-05	1.2E-04	5.9E-05
(P-value) < 0.005	0	2.0E-04	6.5E-04	4.0E-04	5.9E-04	5.3E-04	6.9E-04	9.7E-04	9.1E-04	9.5E-04	4.8E-04
(P-value) < 0.010	0	3.1E-04	1.0E-03	1.4E-03	1.3E-03	1.4E-03	1.3E-03	1.5E-03	1.7E-03	2.0E-03	9.8E-04
(P-value) < 0.025	1.9E-03	3.2E-03	1.6E-03	3.0E-03	3.5E-03	4.3E-03	5.6E-03	5.0E-03	4.3E-03	4.5E-03	3.3E-03
(P-value) < 0.050	7.1E-03	8.2E-03	7.4E-03	8.3E-03	9.6E-03	8.6E-03	9.8E-03	0.010	0.013	0.013	8.8E-03
(P-value) < 0.100	0.014	0.022	0.025	0.022	0.021	0.025	0.029	0.030	0.029	0.029	0.023
(P-value) < 0.200	0.041	0.035	0.048	0.062	0.078	0.074	0.067	0.066	0.072	0.069	0.057
(P-value) < 0.500	0.049	0.092	0.131	0.149	0.137	0.149	0.173	0.193	0.210	0.220	0.131
(P-value) < 0.800	0.055	0.101	0.145	0.185	0.221	0.253	0.281	0.307	0.328	0.326	0.184
(P-value) < 0.900	0.057	0.109	0.155	0.196	0.234	0.264	0.293	0.320	0.345	0.367	0.195
(P-value) < 0.950	0.063	0.119	0.170	0.214	0.254	0.288	0.318	0.346	0.370	0.392	0.211
(P-value) < 0.975	0.068	0.130	0.187	0.238	0.284	0.325	0.361	0.392	0.420	0.445	0.236
(P-value) < 0.990	0.099	0.181	0.249	0.307	0.354	0.394	0.430	0.461	0.490	0.515	0.296
(P-value) < 0.995	0.124	0.225	0.305	0.371	0.421	0.458	0.495	0.527	0.554	0.576	0.349
(P-value) < 0.999	0.169	0.301	0.400	0.477	0.536	0.582	0.620	0.648	0.673	0.695	0.444
(P-value) < 1.000	0.201	0.348	0.456	0.531	0.591	0.634	0.667	0.695	0.718	0.737	0.492

NOTE: Each column gives the cumulative distribution of P-values for a given range of the size of the blocks involved in "purge"s (subjects) or control blocks (controls).

Table SS5. Cumulative distributions of the P-values for various lengths of "purge"-involved branches (for set 3M, *i.e.*, 15 mammals)

A. For subjects

Branch length	[0.00,0.02)	[0.02,0.04)	[0.06,0.08)	[0.08,0.10)	[0.12,0.14)	[0.14,0.16)	[0.16,0.18)	[0.18,0.20)	Total
(P-value) < 0.001	0.333	0.191	0.144	0.167	0.102	0.100	0.060	0.068	0.085
(P-value) < 0.005	0.444	0.341	0.272	0.312	0.234	0.235	0.153	0.175	0.196
(P-value) < 0.010	0.444	0.423	0.392	0.392	0.300	0.296	0.217	0.243	0.263
(P-value) < 0.025	0.667	0.484	0.472	0.548	0.441	0.480	0.319	0.346	0.380
(P-value) < 0.050	0.722	0.630	0.528	0.634	0.554	0.536	0.443	0.491	0.495
(P-value) < 0.100	0.722	0.752	0.672	0.763	0.632	0.657	0.576	0.613	0.618
(P-value) < 0.200	0.778	0.829	0.848	0.882	0.795	0.805	0.744	0.797	0.780
(P-value) < 0.500	0.778	0.894	0.904	0.952	0.951	0.918	0.912	0.925	0.919
(P-value) < 0.800	0.833	0.943	0.904	0.962	0.959	0.932	0.932	0.950	0.938
(P-value) < 0.900	0.833	0.947	0.912	0.962	0.959	0.936	0.937	0.953	0.942
(P-value) < 0.950	0.833	0.947	0.912	0.968	0.966	0.939	0.939	0.953	0.944
(P-value) < 0.975	0.833	0.959	0.920	0.968	0.966	0.940	0.939	0.953	0.945
(P-value) < 0.990	0.833	0.972	0.944	0.968	0.968	0.945	0.945	0.955	0.951
(P-value) < 0.995	0.833	0.972	0.944	0.978	0.968	0.947	0.949	0.956	0.954
(P-value) < 0.999	0.833	0.972	0.944	0.978	0.971	0.950	0.956	0.965	0.959
(P-value) < 1.000	1	1	1	1	1	1	1	1	1

B. For Controls

Branch lenghth	[0.00,0.02)	[0.02,0.04)	[0.06,0.08)	[0.08,0.10)	[0.12,0.14)	[0.14,0.16)	[0.16,0.18)	[0.18,0.20)	Total
(P-value) < 0.001	6.9E-05	5.7E-05	5.5E-05	3.2E-05	5.7E-05	7.9E-05	4.3E-05	5.2E-05	5.9E-05
(P-value) < 0.005	4.8E-04	4.8E-04	4.2E-04	3.6E-04	5.7E-04	6.7E-04	3.9E-04	5.5E-04	4.8E-04
(P-value) < 0.010	7.6E-04	1.0E-03	1.2E-03	9.0E-04	1.3E-03	1.5E-03	1.1E-03	1.3E-03	9.8E-04
(P-value) < 0.025	2.9E-03	3.0E-03	3.8E-03	3.2E-03	5.0E-03	5.9E-03	3.7E-03	3.6E-03	3.3E-03
(P-value) < 0.050	9.0E-03	7.7E-03	9.0E-03	9.6E-03	0.013	0.011	0.010	0.011	8.8E-03
(P-value) < 0.100	0.025	0.019	0.027	0.030	0.026	0.029	0.024	0.027	0.023
(P-value) < 0.200	0.035	0.058	0.071	0.071	0.069	0.081	0.079	0.083	0.057
(P-value) < 0.500	0.037	0.113	0.217	0.225	0.251	0.254	0.271	0.249	0.131
(P-value) < 0.800	0.044	0.137	0.252	0.290	0.381	0.422	0.441	0.453	0.184
(P-value) < 0.900	0.046	0.145	0.262	0.301	0.387	0.436	0.481	0.487	0.195
(P-value) < 0.950	0.051	0.161	0.287	0.369	0.434	0.454	0.503	0.499	0.211
(P-value) < 0.975	0.055	0.203	0.359	0.394	0.467	0.469	0.513	0.508	0.236
(P-value) < 0.990	0.068	0.300	0.495	0.429	0.542	0.512	0.551	0.538	0.296
(P-value) < 0.995	0.095	0.368	0.539	0.551	0.578	0.537	0.627	0.562	0.349
(P-value) < 0.999	0.234	0.457	0.567	0.618	0.609	0.568	0.702	0.668	0.444
(P-value) < 1.000	0.266	0.498	0.649	0.682	0.674	0.682	0.752	0.760	0.492

NOTE: Each column shows the cumulative distribution of P-values for a given range of the lengths of branches involved in "purge"s (subjects) or separating the control blocks (controls).

Table SS6. Sub-classification of "Shift + Shift" error pairs

Sub-class	Set 3P	Set 3M	Set 3F
Shift + Shift(???)	41	877	87
Shift + Shift(gap-aligned)(?)	39	163	13
Shift + Shift	9	94	3
Others	3	20	2
Total	92	1154	105

NOTE: The number in each cell is the absolute frequency of the pairs belonging to a specified sub-class (row) in a specified MSA set (column). Those pairs that co-exist with "complex" errors were excluded. Sets 3P, 3M and 3F consist of fictitious MSAs simulated along the trees of 12 primates, 15 mammals and 9 fast-evolving mammals, respectively. (See (Ezawa 2016a) for more details on the simulations.)

[NEWLY ADDED (3)]

Table SS7. Cumulative distributions of sizes of individual gapped segments

	Complex (+Too_long) (cum%)			Control [= Error(non-complex)+Correct] (cum%)		
Segment size	12 primates	15 mammals	9 fe-mammals	12 primates	15 mammals	9 fe-mammals
1 – 4 (omitted)						
5	49.58%	57.16%	63.42%	76.16%	78.34%	87.00%
6	53.42%	60.86%	67.43%	78.87%	81.20%	89.50%
7	56.60%	63.86%	70.78%	80.95%	83.42%	91.11%
8	59.31%	66.41%	73.58%	82.64%	85.23%	92.47%
9	61.63%	68.59%	75.89%	84.06%	86.72%	93.68%
10	63.52%	70.50%	77.90%	85.24%	87.95%	94.50%
11, 12	67.18%	73.72%	81.14%	87.18%	89.88%	95.90%
13, 14	70.14%	76.29%	83.69%	88.71%	91.33%	96.69%
15, 16	72.34%	78.44%	85.72%	88.92%	92.44%	97.28%
17, 18	74.29%	80.23%	87.42%	90.94%	93.37%	97.73%
19, 20	76.08%	81.82%	88.83%	91.76%	94.12%	98.08%
21 – (omitted)						

NOTE: The number in each cell is the cumulative percentage of individual gapped segments up to (and including) the specified size (row), in the specified category ("Complex" or "Control") and the specified dataset (12 primates, 15 mammals, or 9 fe-mammals) (column). The percentage highlighted in yellow is closest to 90% of the "Control" gapped segments in each dataset, and the blue percentage of the "Complex" gapped segments corresponds to it. The size of each segment is the number of columns that it consists of.

Table SS8. Cumulative distributions of numbers of indels in individual gapped segments

	Complex (+Too_long) (cum%)			Control [= Error(non-complex)+Correct] (cum%)		
#[indels]	12 primates	15 mammals	9 fe-mammals	12 primates	15 mammals	9 fe-mammals
1	56.81%	58.53%	61.90%	89.62%	81.20%	85.92%
2	79.22%	75.92%	79.76%	97.14%	93.77%	96.94%
3	88.59%	83.61%	87.26%	98.87%	97.17%	99.10%
4	93.12%	88.01%	91.33%	99.49%	98.49%	99.66%
5	95.80%	90.80%	93.84%	99.78%	99.11%	99.88%
6 – (omitted)						

NOTE: Notes similar to those (but the last one) for Table SS7 apply also to this table. The indel(s) responsible for each individual gapped segment was/were inferred via the Dollo parsimony principle.

Table SS9. Cumulative distributions of total lengths of indels in individual gapped segments

	Complex (+Too_long) (cum%)			Control [= Error(non-complex)+Correct] (cum%)		
Totlen{indels}	12 primates	15 mammals	9 fe-mammals	12 primates	15 mammals	9 fe-mammals
1 – 4 (omitted)						
5	49.15%	56.29%	62.40%	75.98%	77.80%	86.47%
6	52.98%	59.94%	66.37%	78.69%	80.65%	89.03%
7	56.07%	62.92%	69.63%	80.76%	82.89%	90.70%
8	58.81%	65.42%	72.37%	82.45%	84.71%	92.05%
9	61.14%	67.56%	74.66%	83.87%	86.22%	93.23%
10	63.01%	69.44%	76.65%	85.06%	87.45%	94.16%
11, 12	66.66%	72.60%	79.87%	87.01%	89.40%	95.62%
13, 14	69.49%	75.14%	82.39%	88.54%	90.87%	96.45%
15, 16	71.84%	77.29%	84.43%	89.75%	92.04%	97.10%
17, 18	73.84%	79.06%	86.14%	90.77%	92.98%	97.56%
19, 20	75.64%	80.61%	87.56%	91.60%	93.74%	97.93%
21 – (omitted)						

NOTE: All notes similar to those for Table SS8 apply also to this table.

Table SS10. Cumulative distributions of numbers of gapped segments in individual correct or erroneous segments

	Complex (+Too_long) (cum%)			Control [= Error(non-complex)+Correct] (cum%)		
#{gapped segments}	12 primates	15 mammals	9 fe-mammals	12 primates	15 mammals	9 fe-mammals
1	45.95%	20.12%	6.56%	71.02%	72.62%	78.72%
2	82.99%	47.07%	19.23%	83.67%	92.55%	96.26%
3	93.10%	66.17%	30.53%	89.78%	97.57%	99.34%
4	96.54%	77.26%	39.33%	93.51%	99.14%	99.93%
(others … omitted)						

NOTE: The number in each cell is the cumulative percentage of individual (erroneous or correct) segments each of which has up to (and including) the specified number of gapped segments (row), in the specified category ("Complex" or "Control") and the specified dataset (12 primates, 15 mammals, or 9 fe-mammals) (column). The percentage highlighted in yellow is closest to 90% of the "Control" (correct or erroneous) segments in each dataset, and the blue percentage of the "Complex"-erroneous segments corresponds to it. The percentage highlighted in light green is that of "Complex"-erroneous segments consisting only of one gapped segment (in each dataset).

Table SS11. Cumulative distributions of total numbers of columns in individual correct or erroneous segments

	Complex (+Too_long) (cum%)			Control [= Error(non-complex)+Correct] (cum%)		
Tot#[columns]	12 primates	15 mammals	9 fe-mammals	12 primates	15 mammals	9 fe-mammals
1 – 4 (omitted)						
5	25.33%	(omitted)	(omitted)	60.58%	(omitted)	(omitted)
6, 7, 8 (omitted)						
9	37.67%	23.82%	16.88%	71.20%	79.90%	90.73%
10	39.94%	25.85%	18.65%	72.96%	81.68%	91.95%
11, 12	43.97%	29.72%	21.56%	75.98%	84.52%	93.98%
13, 14	47.34%	33.15%	24.34%	78.42%	86.73%	95.23%
15, 16	50.10%	36.21%	26.65%	80.45%	88.41%	96.07%
17, 18	52.42%	38.87%	28.82%	82.16%	89.83%	96.76%
19, 20	54.66%	41.41%	30.92%	83.62%	90.97%	97.20%
21 – 25	59.59%	46.84%	35.43%	86.59%	93.09%	98.05%
26 – 30	64.07%	(omitted)	(omitted)	88.77%	(omitted)	(omitted)
31 – 40	70.90%	(omitted)	(omitted)	91.89%	(omitted)	(omitted)
41 – (omitted)						

NOTE: Notes similar to those of Table SS10 apply also to this table. Here, each row specifies the total number of columns in the gapped segments in each correct or erroneous segment. In this table, the percentages of "Complex"-erroneous segments that are near or less than 25% are highlighted in light green.

Table SS12. Cumulative distributions of total numbers of indels in individual correct or erroneous segments

	Complex (+Too_long) (cum%)			Control [= Error(non-complex)+Correct] (cum%)		
Tot#[indels]	12 primates	15 mammals	9 fe-mammals	12 primates	15 mammals	9 fe-mammals
1	2.65%	1.02%	0.88%	60.10%	56.06%	67.22%
2	38.47%	12.10%	6.78%	78.15%	80.76%	89.47%
3	63.76%	25.78%	14.19%	86.58%	90.94%	96.35%
4	(omitted)	37.82%	21.48%	(omitted)	95.43%	98.62%
5	(omitted)	47.24%	(omitted)	(omitted)	97.55%	(omitted)
6 – (omitted)						

NOTE: Notes similar to those of Table SS10 apply also to this table. Here, each row specifies the total number of indels inferred (via Dollo parsimony) to have resulted in each correct or erroneous segment. In this table, the cumulative percentages of "Complex"-erroneous segments with up to 1 and 2 indels each are highlighted in light green.

Table SS13. Cumulative distributions of maximum numbers of indels (per gapped segment) in individual correct or erroneous segments

	Complex (+Too_long) (cum%)			Control [= Error(non-complex)+Correct] (cum%)		
Max#{indels}	12 primates	15 mammals	9 fe-mammals	12 primates	15 mammals	9 fe-mammals
1	27.29%	15.80%	10.88%	82.78%	75.07%	82.92%
2	62.44%	38.89%	29.26%	95.08%	91.49%	96.18%
3	79.09%	53.61%	42.57%	98.05%	96.10%	98.88%
4	87.35%	63.55%	52.07%	99.12%	97.91%	99.57%
5 – (omitted)						

NOTE: Notes similar to those of Table SS10 apply also to this table. Here, each row specifies the **maximum** among the **numbers of indels** inferred (via Dollo parsimony) to have resulted in individual gapped segments belonging to each correct or erroneous segment. In this table, the percentages of "Complex"-erroneous segments whose maximum indel numbers are one are highlighted in light green.

Table SS14. Distributions of Max(#{ins}-#{del}) in individual correct or erroneous segments

	A = Complex (+Too_long) (%)			B = Control [= Error(non-complex)+Correct] (%)			A/B		
Max(#{ins} – #{del})	12 primates	15 mammals	9 fe-mammals	12 primates	15 mammals	9 fe-mammals	12 primates	15 mammals	9 fe-mammals
x <= −7 (omitted)									
−6	0.08%	0.11%	0.02%	0.00%	0.02%	0.02%	0.556	2.074	2
−5	0.11%	0.20%	0.05%	0.02%	0.04%	0.04%	0.212	1.512	1.667
−4	0.41%	0.43%	0.10%	0.08%	0.16%	0.08%	0.148	0.818	1.667
−3	1.10%	0.90%	0.29%	0.38%	0.57%	0.24%	0.0865	0.494	0.699
−2	3.34%	2.04%	0.81%	2.24%	3.16%	0.67%	0.0443	0.198	0.263
−1	10.88%	6.27%	3.49%	26.63%	24.99%	2.87%	0.0122	0.075	0.11
0	16.62%	9.43%	4.81%	4.99%	7.78%	3.95%	0.0992	0.362	0.721
1	44.19%	50.59%	47.01%	62.68%	58.66%	39.40%	0.021	0.259	0.734
2	17.93%	19.69%	27.92%	2.67%	3.97%	26.48%	0.2	1.556	12.546
3	3.54%	5.72%	9.19%	0.25%	0.47%	11.95%	0.415	4.238	106.741
x >= 4 (omitted)									
Total	100%	100%	100%	100%	100%	100%	0.0298	0.314	1.106

NOTE: The number in each cell (in the 2nd – 7th columns) is the percentage of correct or erroneous segments with the specified value of Max(#{ins}-#{del}) (, which is the maximum of the difference of the number of insertions from the number of deletions over the gapped segments belonging to each correct or erroneous segment,) (row), in the specified category ("Complex" or "Control") and the specified dataset (12 primates, 15 mammals, or 9 fe-mammals) (column). The percentages for Max(#{ins}-#{del}) = 0 are in boldface. The percentages over 10% are highlighted in light green. The number in each cell in the 8-10 th columns (labeled "A/B") is the ratio of the number of "Complex"-erroneous segments to the number of "Control" segments, both with the specified value of Max(#{ins}-#{del}) (row).

Table SS15. Cumulative distributions of representative sizes of spacers flanking individual gapped segments

A. Average spacer-size

	A= Complex (+Too_long) (cum%)			B = Control [= Error(non-complex)+Correct] (cum%)		
Ave(spacer size)	12 primates	15 mammals	9 fe-mammals	12 primates	15 mammals	9 fe-mammals
1	3.63%	11.28%	13.51%	0.49%	4.12%	3.51%
2	10.33%	28.66%	34.08%	1.66%	13.00%	12.17%
3	17.58%	45.79%	53.53%	3.46%	24.69%	24.51%
4	24.59%	60.02%	68.91%	5.75%	37.10%	38.64%
5	31.52%	71.08%	79.89%	8.46%	48.85%	52.36%
6	37.20%	79.34%	87.31%	11.46%	59.40%	64.06%
7	42.40%	85.32%	92.14%	14.75%	68.34%	73.72%
8	47.16%	89.64%	95.18%	18.15%	75.58%	81.33%
9	51.64%	92.73%	97.08%	21.66%	81.43%	86.89%
10-	100%	100%	100%	100%	100%	100%

B. Maximum spacer-size

	A= Complex (+Too_long) (cum%)			B = Control [= Error(non-complex)+Correct] (cum%)		
Max(Spacer size)	12 primates	15 mammals	9 fe-mammals	12 primates	15 mammals	9 fe-mammals
1	1.36%	4.26%	5.09%	0.17%	1.37%	1.05%
2	4.52%	14.04%	16.81%	0.65%	5.50%	4.97%
3	9.33%	26.00%	30.91%	1.51%	11.80%	11.22%
4	14.56%	37.77%	44.72%	2.68%	19.59%	19.25%
5	19.46%	48.51%	56.74%	4.09%	28.02%	29.03%
6	23.85%	57.80%	66.84%	5.78%	36.54%	38.76%
7	28.15%	65.55%	74.77%	7.67%	44.72%	48.15%
8	32.05%	71.90%	80.93%	9.69%	52.32%	57.09%
9	35.60%	81.35%	85.70%	11.81%	59.16%	64.76%
10	38.81%	87.61%	89.24%	14.04%	65.34%	71.28%
11-	100%	100%	100%	100%	100%	100%

C. Minimum spacer-size

	A= Complex (+Too_long) (cum%)			B = Control [= Error(non-complex)+Correct] (cum%)		
Min(Spacer size)	12 primates	15 mammals	9 fe-mammals	12 primates	15 mammals	9 fe-mammals
1	26.32%	38.83%	41.10%	7.96%	23.13%	21.82%
2	44.78%	62.58%	66.16%	15.75%	42.66%	42.00%
3	59.57%	77.53%	80.98%	23.25%	58.23%	58.19%
4	68.68%	86.43%	89.49%	30.15%	69.92%	70.62%
5	75.15%	91.84%	94.21%	36.53%	78.56%	80.15%
6	80.04%	95.05%	96.85%	42.42%	84.88%	86.81%
7-	100%	100%	100%	100%	100%	100%

NOTE: This table consists of three sub-tables, which are cumulative distributions of representative sizes (**A** for average, **B** for maximum, and **C** for minimum) between spacers (i.e., gapless segments) flanking each of gapped segments. Each row specifies the value of each representative size, and each column specifies the category of the gapped segments

("Complex" or "Control") and the dataset (12 primates, 15 mammals, or 9 fe-mammals). The cumulative percentages of "Control" gapped segments closest to 10% are highlighted in yellow. The blue cumulative percentages of "Complex" gapped segments correspond to the yellow-highlighted ones.

Table SS16. Cumulative distributions of sizes of individual spacers (for 15 mammals)

Spacer size	A:A_in (cum%)	A:A_ex (cum%)	A:B_ex (cum%)	B:B_in (cum%)	B:B_ex (cum%)
1	27.00%	4.58%	6.46%	25.88%	7.02%
2	46.80%	11.04%	15.06%	42.97%	17.17%
3	61.86%	18.42%	24.34%	55.35%	27.90%
4	72.66%	26.44%	33.63%	64.48%	38.11%
5	80.50%	34.49%	42.54%	71.46%	47.33%
6	86.07%	42.68%	50.72%	76.97%	55.49%
7	90.01%	50.10%	58.03%	81.28%	62.59%
8	92.80%	56.74%	64.40%	84.75%	68.70%
9	94.79%	62.82%	70.06%	87.49%	73.81%
10	96.20%	68.39%	74.83%	89.80%	78.23%
11, 12	97.99%	77.33%	82.32%	93.07%	84.95%
13, 14	98.91%	83.73%	87.74%	95.31%	89.60%
15, 16	99.41%	88.63%	91.53%	96.79%	92.86%
17, 18	99.66%	91.87%	94.13%	97.80%	95.10%
19, 20	99.81%	94.42%	95.96%	98.52%	96.63%
21– (omitted)					

NOTE: The number in each cell is the cumulative percentage of individual spacers (i.e., gapless segments) whose sizes are up to (and including) the specified number (row), and which are flanked by pairs of gapped segments of a specified category (column). Figure SS3 illustrates the categories of the pairs of gapped segment. In the "_in" categories, the percentages near and less than 25% are highlighted in yellow. In the "_ex" categories, the percentages near and less than 15% are highlighted in light green. The percentages closest to 50%, 80% and 90% (in each category) are colored blue, green, and red, respectively.

Table SS17. Performances of methods to artificially cluster gapped segments

Spacer-size upper-bound	Spacer-size threshold	Tot# {pure-B}	Tot# {pure-A}	Tot# {mixed}	Target %{in B}	Target %{in A}	Target#{in A} / Target#{in B}
Method I							
---	2	558793	378425	68389	76.25%	41.31%	0.3833
					82.03%	47.25%	0.4076
					88.42%	56.74%	0.4540
					90.22%	60.59%	0.4752
---	3	534389	343086	128312	78.36%	41.23%	0.3723
					81.15%	45.27%	0.3947
					89.53%	60.46%	0.4778
					90.49%	62.73%	0.4905
Method II							
10	5				74.97%	38.97%	0.3678
					81.01%	43.63%	0.3811
					89.51%	53.39%	0.4220
					90.16%	54.61%	0.4286
10	6				78.79%	40.63%	0.3649
					82.01%	43.50%	0.3753
					89.95%	54.38%	0.4278
					90.54%	55.74%	0.4356
15	7				78.48%	38.61%	0.3481
					80.49%	40.45%	0.3556
					89.78%	55.94%	0.4409
					90.09%	56.79%	0.4460
20	7				79.70%	39.57%	0.3513
					80.50%	40.34%	0.3546
					89.93%	57.72%	0.4541
					90.08%	58.19%	0.4571
Method III							
10	2	556980	369788	79019	73.51%	36.07%	0.3472
					80.14%	40.76%	0.3599
					89.44%	52.61%	0.4162

					90.30%	54.60%	0.4278
10	3	553751	364853	87183	77.81%	37.74%	0.3432
					81.15%	40.78%	0.3556
					89.44%	52.65%	0.4165
					90.06%	54.08%	0.4249
10	6	533101	333460	139226	79.84%	36.92%	0.3272
					81.01%	38.42%	0.3356
					89.68%	56.10%	0.4426
					90.04%	57.12%	0.4489
15	5	533538	333275	138974	77.50%	35.10%	0.3205
					80.28%	37.66%	0.3319
					89.90%	51.11%	0.4495
					90.26%	58.16%	0.4559
15	6	524762	320232	160793	79.65%	36.73%	0.3263
					81.28%	39.05%	0.3399
					89.82%	58.17%	0.4582
					90.08%	58.98%	0.4633
20	5	526973	323550	155264	79.32%	36.78%	0.3280
					80.34%	38.02%	0.3348
					89.64%	58.31%	0.4603
					90.00%	59.47%	0.4675

NOTE: For each of the three methods, only the results with the optimum and the near-optimum combinations of the parameters (spacer-size threshold and spacer-size upper-bound) are shown here. Regardless of the clustering methods, the total number of gapped segments in category B (*i.e.*, control) is 589021, the total number of gapped segments in category A (*i.e.*, "Complex" errors) is 416766, and the ratio of the former to the latter is 0.707557. Before this analysis, a "Complex" erroneous segment was re-classified as "non-complex" if only less than 4 indels were inferred from each of the reconstructed and reference MSAs. The best-performing results are highlighted in red, and relatively well-performing results are highlighted in yellow.

KEY:

Tot#{pure-A/B} = the total number of gapped segments belonging to the artificial clusters consisting only of gapped segments of category A/B.

Tot#{mixed} = the total number of gapped segments belonging to the artificial clusters each

of which is a mixture of gapped segments of categories A and B.

Target%{in A} = the cumulative percentages of category-A gapped segments (compared to their total number) that are the closest to the target % (80% and 90% in this analysis), on each of the upper- and lower-side.

Target%{in B} = the cumulative percentages of category-B gapped segments (compared to their total number) corresponding the Target%{in A}.

Target#{in A/B} = the total number of category-A/B gapped segments corresponding to the above Target%{in A/B}.

[END of "NEWLY ADDED (3)"]

Supplementary Figures SS1-SS???

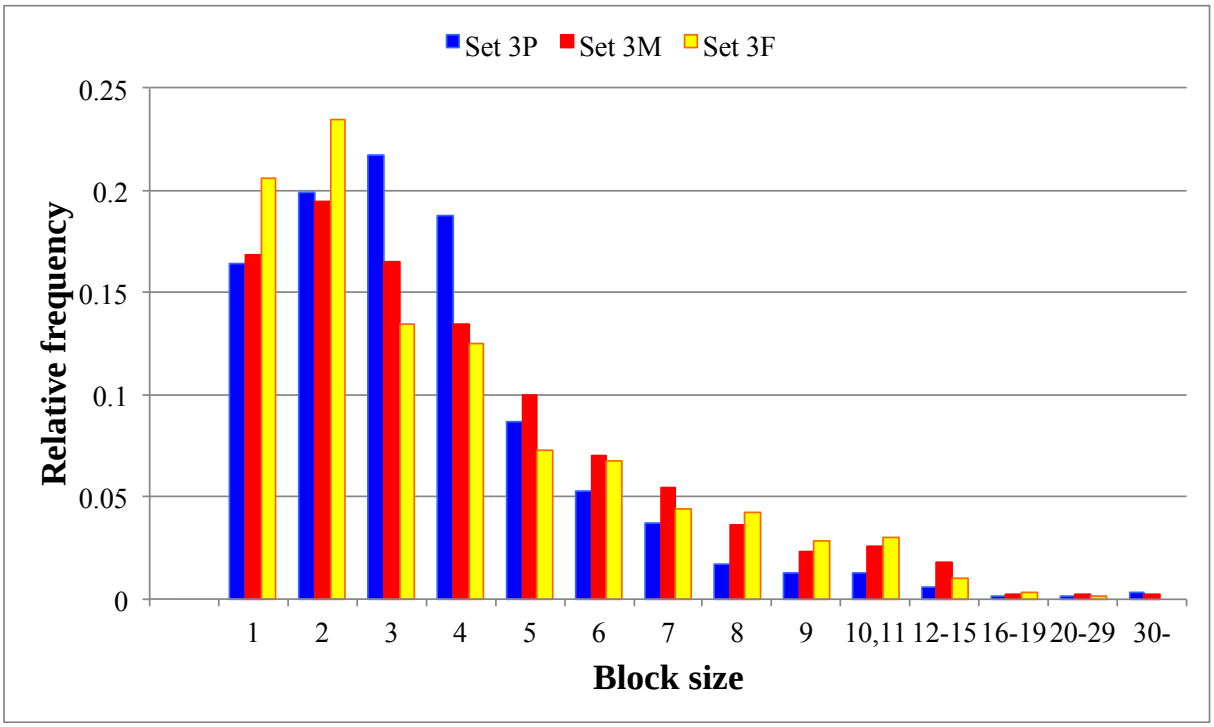

Figure SS1. Distributions of block-sizes of "purge"s.

The abscissa is the size of the block involved in the "purge"s, and the ordinate is the relative frequency among "purge"s. Sets 3P, 3M and 3F consist of MSAs of DNA sequences simulated along the trees of 12 primates, 15 mammals and 9 fast-evolving mammals, respectively.

A. "Shift + Shift(???)" [class0014/lclid0002736/segment13, original]

Seq_ID	0	1	2	3	4	5	6	7	8	9
seq0000	-	-	0	0	0	-2	-2	-	-	0
seq0001	-	-	0	0	0	-2	-2	-	-	0
seq0002	-	-	0	0	0	-2	-2	-	-	0
seq0003	0	0	0	0	0	-2	-2	-	-	0
seq0004	-	-	0	0	0	-2	-2	-	-	0
seq0005	-	-	0	0	0	0	0	0	0	0
seq0006	-	-	0	0	0	0	0	0	0	0
seq0007	-	-	0	0	-	-	-	0	0	0
seq0008	-	-	0	0	0	-2	-2	-	-	0
seq0009	-	-	0	0	0	-2	-2	-	-	0
seq0013	-	-	0	0	0	-2	-2	-	-	0
seq0014	-	-	0	0	0	-2	-2	-	-	0
seq0015	-	-	-	-	2	2	-2	-	-	0
seq0018	-	-	0	0	-	-	-	-	-	-
seq0020	-	-	0	0	0	-2	-2	-	-	0

B. "Shift + Shift(???)" [class0004/lclid0000752/segment25, original]

Seq_ID	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
seq0000	0	0	-5	-5	-5	-5	-5	-5	-	-	-5	-	-	-	-	-
seq0001	0	0	-5	-5	-5	-5	-5	-5	-	-	-5	-	-	-	-	-
seq0002	0	0	-5	-5	-5	-5	-5	-5	-	-	-5	-	-	-	-	-
seq0003	0	0	-5	-5	-5	-5	-5	-5	-	-	-5	-	-	-	-	-
seq0004	0	0	-5	-5	-5	-5	-5	-5	-	-	-5	-	-	-	-	-
seq0005	0	0	-5	-5	-5	-5	-5	-5	-	-	-5	-	-	-	-	-
seq0006	0	0	-5	-5	-5	-5	-5	-5	-5	-5	-5	-	-	-	-	-
seq0007	0	0	-5	-5	-5	-5	-5	-5	-5	-5	-5	-	-	-	-	-
seq0008	-	-	-	-	-	-	-	7	-	-	9	-	-	-	-	-
seq0009	-	-	-	-	-	-	-	7	-	-	9	-	-	-	-	-
seq0013	-	-	-	-	-	-	-	7	-	-	9	-	-	-	-	-
seq0014	-	-	-	-	-	-	-	7	-	-	9	9	9	9	9	9
seq0015	-	-	-	-	-	-	-	7	-	-	9	-	-	-	-	-
seq0018	-	-	-	-	-	-	-	7	-	-	9	-	-	-	-	-
seq0020	-	-	-	-	-	-	-	7	-	-	9	-	-	-	-	-

C. The same as panel B, with an alternative representation of the reference MSA [class0004/lclid0000752/segment25, modified]

Seq_ID	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
seq0000	0	0	0	0	0	0	0	0	-	-	0	-	-	-	-	-
seq0001	0	0	0	0	0	0	0	0	-	-	0	-	-	-	-	-
seq0002	0	0	0	0	0	0	0	0	-	-	0	-	-	-	-	-
seq0003	0	0	0	0	0	0	0	0	-	-	0	-	-	-	-	-
seq0004	0	0	0	0	0	0	0	0	-	-	0	-	-	-	-	-
seq0005	0	0	0	0	0	0	0	0	-	-	0	-	-	-	-	-
seq0006	0	0	0	0	0	0	0	0	0	0	0	-	-	-	-	-
seq0007	0	0	0	0	0	0	0	0	0	0	0	-	-	-	-	-
seq0008	-	-	-	-	-	-	-	7	-	-	9	-	-	-	-	-
seq0009	-	-	-	-	-	-	-	7	-	-	9	-	-	-	-	-
seq0013	-	-	-	-	-	-	-	7	-	-	9	-	-	-	-	-
seq0014	-	-	-	-	-	-	-	7	-	-	9	0	0	0	0	0
seq0015	-	-	-	-	-	-	-	7	-	-	9	-	-	-	-	-
seq0018	-	-	-	-	-	-	-	7	-	-	9	-	-	-	-	-
seq0020	-	-	-	-	-	-	-	7	-	-	9	-	-	-	-	-

D. "Shift + Shift(gap-aligned)(?)" [class0001/lclid0000171/segment30]

Seq_ID	0	1	2	3	4	5	6	7
seq0000	0	-	-	-	3	0	0	0
seq0001	0	-	-	-	3	0	0	0
seq0002	0	-	-	-	3	0	0	0
seq0003	0	-	-	-	3	0	0	0
seq0004	0	-	-	-	-	0	0	0
seq0005	0	-	-	-	3	0	0	0
seq0006	0	-	-	-	3	0	0	0
seq0007	0	-	-	-	3	0	0	0
seq0008	0	0	0	0	0	0	0	0
seq0009	-	-	-	-	-	-	-	7
seq0013	0	-	-	-	3	0	0	0
seq0014	0	-	-	-	3	0	0	0
seq0015	0	-	-	-	3	0	0	0
seq0018	0	-	-	-	3	0	0	0
seq0020	0	-	-	-	3	0	0	0

Figure SS2. Simple, illustrating examples of "Shift + Shift" error pairs.

Each example is represented with the position-shift map on an erroneous segment of a reconstructed MSA. **A.** An example of "Shift + Shift(???)." Although our prototype script partitioned this segment into three position-shift blocks (blue, red and purple), the red and purple blocks should actually be merged together to define a single composite-block. Once we correctly identify the blocks, it is straightforward to recognize that the entire error could result from the successive actions of two shifts with a fixed temporal order. **B.** A second example of "Shift + Shift(???)." Although our prototype script partitioned this segment into four position-shift blocks (green, blue, red and purple), the red and purple blocks should actually be merged together to define a single composite-block. And the moves of the green block (with shift = -5) and the composite (red and purple) block should be coordinated with each other, in order to preserve the independence of the two corresponding insertions (one into seq000[0-7] and the other into seq0014). If an alternative representation of the independent insertions were used in the reference MSA, the position-shift map is considerably simplified as in panel **C**. **D.** An example of "Shift + Shift(gap-aligned)(?)." The blue block is associated with a simple shift, and the red block is associated with a gap-aligned shift. Although these shifts could be exerted independently of each other, the gap-aligned shift was at first annotated as "Complex," because the red block shares the left-end with the gap (in the reference MSA); such cases cannot be correctly handled by our current prototype script.

[NEWLY ADDED (4)]

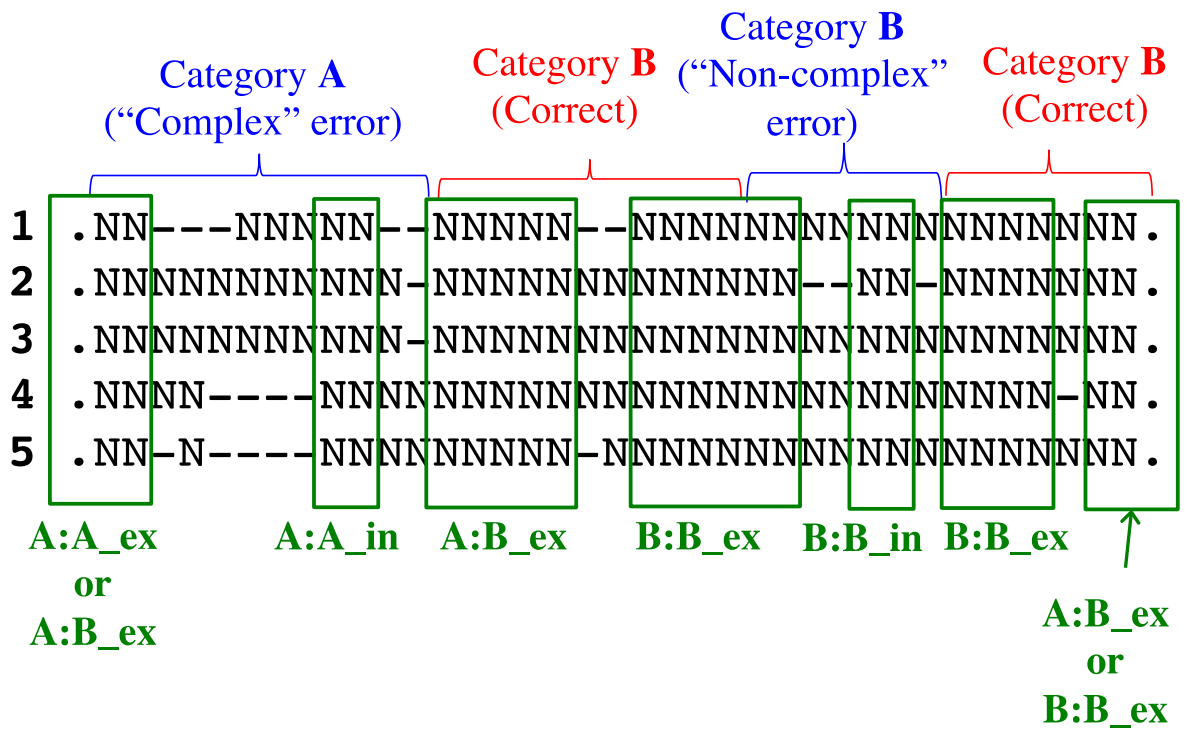

Figure SS3. Categories of spacers.

Each blue top curly bracket indicates an erroneous segment, and each red one indicates a correct segment.

Each open green rectangular box encloses a gapped segment, below which is its category.

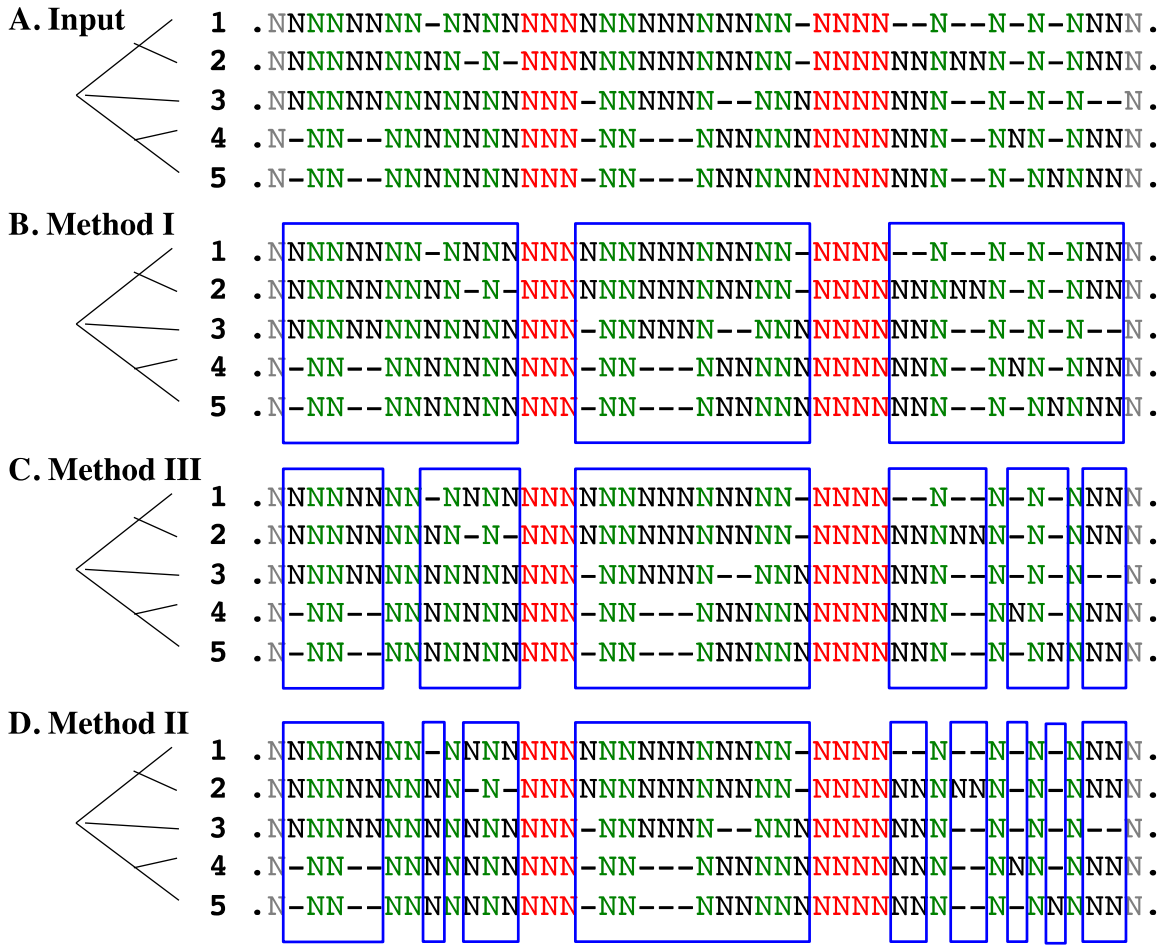

Figure SS4. Methods to artificially cluster gapped segments

This figure illustrates the results of different clustering methods applied to the same fictitious MSA (panel **A**, on the right), which resulted from an indel history along a given tree (panel **A**, on the left).

B. Result of Method I. This simple method does not care about gap patterns.

C. Result of Method III. This method examines the gap patterns of neighboring gapped segments. More precisely, it examines whether or not the neighboring gapped segments undergo indels along the same branch or along phylogenetically neighboring branches.

D. Result of Method II. This method also examines the gap patterns of neighboring gapped segments, in a more complex manner. More precisely, it examines whether or not there is a "trio" of branches that share the same internal node and all of which undergo indels in three neighboring gapped segments.

To focus on the topological issue, we assume that {spacer-size threshold} = {spacer-size upper-bound}. Then, the 2nd half of condition (a) for Methods II & III can be ignored, hence the condition on the spacer-size becomes identical to that for Method I. Here, for illustration,

the {spacer-size threshold} is assumed to be 2.
In each panel, the spacers (i.e., gapless segments) are colored; red indicates that the spacer-size is larger than the threshold, and green indicates that the size is smaller than or equal to the threshold. Each blue rectangular box encloses an artificial cluster of gapped segments (and intervening spacers).

[End of "NEWLY ADDED (4)"]